

In this issue:

- 4. Are Tech Savvy Students Tech Literate? Digital and Data Literacy Skills of First-Year College Students**  
Kevin Mentzer, Nichols College  
Mark Frydenberg, Bentley University  
Adam Patterson, Nichols College
- 25. Teaching Case**  
**Teaching Data Literacy Using Titanic Survival Factors**  
Mark Sena, Xavier University  
Thilini Ariyachandra, Xavier University
- 36. Ethics-Driven Education: Integrating AI Responsibly for Academic Excellence**  
Chukwuemeka Ihekweazu, Sam Houston State University  
Bing Zhou, Sam Houston State University  
Elizabeth Adepeju Adelowo, St. Cloud State University
- 47. Cheating Better with ChatGPT: A Framework for Teaching Students When to Use ChatGPT and other Generative AI Bots**  
David R. Firth, University of Montana  
Mason Derendinger, University of Montana  
Jason Triche, University of Montana
- 61. Teaching Case**  
**How Popular is Your Name? Finding the Popularity of US Names Using Big Data Visualization**  
Frank Lee, Georgia State University  
Alex Algarra, Georgia State University
- 76. Data Analytics Position Description Analysis: Skills Review and Implications for Data Analytics Curricula**  
Queen E. Booker, Metropolitan State University  
Carl M. Rebman Jr., University of San Diego  
Hayden Wimmer, Georgia Southern University  
Steve Levkoff, University of San Diego  
Loreen Powell, Marywood University  
Jennifer Breese, Pennsylvania State University

The **Information Systems Education Journal** (ISEDJ) is a double-blind peer-reviewed academic journal published by **ISCAP** (Information Systems and Computing Academic Professionals). Publishing frequency is five times per year. The first year of publication was 2003.

ISEDJ is published online (<https://isedj.org>). Our sister publication, the Proceedings of the ISCAP Conference (<https://iscap.us/proceedings>) features all papers, abstracts, panels, workshops, and presentations from the conference.

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the ISCAP conference. All papers, whether award-winners or not, are invited to resubmit for journal consideration after applying feedback from the Conference presentation. Award winning papers are assured of a publication slot; however, all re-submitted papers including award winners are subjected to a second round of three blind peer reviews to improve quality and make final accept/reject decisions. Those papers that are deemed of sufficient quality are accepted for publication in the ISEDJ journal. Currently the target acceptance rate for the journal is under 35%.

Information Systems Education Journal is pleased to be listed in the Cabell's Directory of Publishing Opportunities in Educational Technology and Library Science, in both the electronic and printed editions. Questions should be addressed to the editor at [editor@isedj.org](mailto:editor@isedj.org) or the publisher at [publisher@isedj.org](mailto:publisher@isedj.org). Special thanks to volunteer members of ISCAP who perform the editorial and review processes for ISEDJ.

### 2024 ISCAP Board of Directors

Jeff Cummings  
Univ of NC Wilmington  
President

Amy Connolly  
James Madison University  
Vice President

Eric Breimer  
Siena College  
Past President

Jennifer Breese  
Penn State University  
Director

David Gomillion  
Texas A&M University  
Director

Leigh Mutchler  
James Madison University  
Director/Secretary

RJ Podeschi  
Millikin University  
Director/Treasurer

David Woods  
Miami University  
Director

Jeffry Babb  
West Texas A&M University  
Director/Curricular Items Chair

Tom Janicki  
Univ of NC Wilmington  
Director/Meeting Facilitator

Paul Witman  
California Lutheran University  
Director/2024 Conf Chair

Xihui "Paul" Zhang  
University of North Alabama  
Director/JISE Editor

Copyright © 2024 by Information Systems and Computing Academic Professionals (ISCAP). Permission to make digital or hard copies of all or part of this journal for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial use. All copies must bear this notice and full citation. Permission from the Editor is required to post to servers, redistribute to lists, or utilize in a for-profit or commercial use. Permission requests should be sent to Paul Witman, Editor, [editor@isedj.org](mailto:editor@isedj.org).

# INFORMATION SYSTEMS EDUCATION JOURNAL

## Editors

---

**Paul Witman**  
Editor  
California Lutheran  
University

**Thomas Janicki**  
Publisher  
U of North Carolina  
Wilmington

**Dana Schwieger**  
Associate Editor  
Southeast Missouri  
State University

**Kevin Mentzer**  
Associate Editor  
Nichols College

**Ira Goldstein**  
Teaching Cases & Exercises  
Co-Editor  
Siena College

**Michelle Louch**  
Teaching Cases & Exercises  
Co-Editor  
Duquesne University

**Donald Colton**  
Emeritus Editor  
Brigham Young University  
Hawaii

**Jeffry Babb**  
Emeritus Editor  
West Texas A&M  
University

# Are Tech Savvy Students Tech Literate? Digital and Data Literacy Skills of First-Year College Students

Kevin Mentzer  
Kevin.Mentzer@nichols.edu  
Nichols College  
Dudley, MA USA

Mark Frydenberg  
mfrydenberg@bentley.edu  
Bentley University  
Waltham, MA USA

Adam Patterson  
Adam.Patterson@nichols.edu  
Nichols College  
Dudley, MA USA

## Abstract

Much has been written on the skills and capabilities of a new generation raised with social media, smartphones, and immediate access to data and information via the Internet. Today's college students grew up using the Internet, where they both generate and consume data. But do incoming college students have the skills necessary to thrive in a digital world that requires the ability to generate, analyze, and share insights from data? This paper presents a study, performed at two small New England institutions each with a business focus, which examines the digital skill sets of first-year college students in relation to the skills they have developed before entering college. The authors also consider whether there is a "digital divide" among first-year college students in relation to their previous technology skills. When applying the Databilities framework for evaluating data literacy competencies, results show that teaching data literacy skills to first-year college students will be critical to their academic success as future information technology workers.

**Keywords:** Data Literacy, Digital Literacy, Digital Skills, Digital Native, First-Year College Students.

**Recommended Citation:** Mentzer, K., Frydenberg, M., Patterson, A., (2024). Are Tech Savvy Students Tech Literate? Digital and Data Literacy Skills of First-Year College Students. *Information Systems Education Journal*, 22(3), pp.4-24. <https://doi.org/10.62273/MFSD3135>.

# Are Tech Savvy Students Tech Literate? Digital and Data Literacy Skills of First-Year College Students

*Kevin Mentzer, Mark Frydenberg and Adam Patterson*

## 1. INTRODUCTION

Students today are adept at using mobile devices to manage their lives, but often lack the skills to generate, evaluate, and share the data that they interact with professionally or informally every day. (McCarron & Frydenberg, 2023; Tugend, 2023) This study, performed at two small New England institutions each with a business focus, aims to evaluate the digital and data literacy skills of incoming college students, and ascertain which skills they find critical to their success academically and as future information technology workers. Employers need data-literate employees who can interact with data at all stages throughout a project. Today's job market requires proficiency in gathering and collecting data, managing and processing data from multiple sources, analyzing data to identify patterns and trends, and communicating these insights through visualization tools to inform stakeholders in their decision-making process (Leon-Urrutia et al., 2022). As the demand for technology professionals who possess basic data literacy skills continues to increase (Hartzel & Ozturk, 2022), for the availability of increased instruction in these areas will prepare students for lucrative future careers as information workers and knowledge professionals in fields relying on data-driven decision-making skills.

A survey of 432 college students conducted in 2023 by the Chronicle of Higher Education (Tugend, 2023) found that 52% of those surveyed self-identified as "highly proficient", 45% as "somewhat proficient", and 3% as "somewhat deficient" when asked to rate their proficiency with digital technology and digital skills. These students, whom Prensky (2001) coined digital natives because they grew up in an era where devices and technology were ubiquitous, assessed their overall digital proficiency in terms of awareness, comprehension, and user experience. Many attributed their technology proficiency to high school experiences that require the use of productivity tools (writing articles for the school newspaper, designing the yearbook, maintaining budgets, etc.)

While many college students claim to adeptly use technology in their free time for personal and social purposes, the Chronicle study concludes that they often do not have the skills to make sense of the information and data that they create and interact with regularly.

## 2. FRAMEWORKS FOR EVALUATING DIGITAL AND DATA LITERACY

Digital literacy refers to the ability to use digital technologies for communication, locating information online, navigating an operating system, and other computer literacy tasks (McCarron & Frydenberg, 2023). Achieving data literacy requires the digital literacy skills necessary to obtain and interact with large sets of data, the critical thinking skills required to evaluate and make sense of it, and the communication skills needed to share insights to inform decision making in a business context. Students also need to know why and when to use those skills along with how to select the appropriate data to be used (Frank et al., 2016; Leon-Urrutia et al., 2022).

"Digital intelligence" (DQ) is an emerging umbrella term that includes both digital and data skills, competencies, readiness, and literacy (IEEE, 2021). The IEEE's DQ Framework identifies twenty-four competencies including personal and social identities in digital spaces, use of devices and digital media, personal and professional online communication and collaboration, skills for navigating digital life and careers, safety and security concerns, emotional and relational impacts of technology, and upholding human rights in a technology-enabled world (IEEE, 2021, p. 14).

The IEEE Digital Intelligence Quotient provides a framework for classifying digital skills and assessing competencies required in a world driven by technology and developing these skills to further digital intelligence. Students must think creatively and possess the digital skills necessary to succeed in the workplace. Digital literacy includes the technical skills necessary to

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
At this level of competency, an individual can complete <b>simple tasks with instruction.</b>	At this level of competency, an individual can complete <b>simple tasks on their own, with guidance where needed.</b>	At this level of competency, an individual can complete <b>well defined tasks on their own.</b>	At this level of competency, an individual can complete <b>well defined problems and tasks on their own.</b>	At this level of competency, an individual can <b>assist others</b> to complete <b>simple tasks and problems.</b>	At this level of competency, an individual can <b>teach and assist</b> others to complete <b>complex problems and tasks.</b>
Activities to support this level of competency should focus on <b>remembering.</b>	Activities to support this level of competency should focus on <b>remembering.</b>	Activities to support this level of competency should focus on <b>understanding.</b>	Activities to support this level of competency should focus on <b>understanding.</b>	Activities to support this level of competency should focus on <b>applying skills.</b>	Activities to support this level of competency should focus on <b>applying skills.</b>

**Figure 1. Levels in the Databilities Framework, adapted from (Data to the People, 2018)**

participate responsibly in a technology-driven society. Digital literacy combines technical, procedural, cognitive and emotional-social skills, for example, creating a document or spreadsheet involves technical skills (knowing how to use an office productivity tool to accomplish a task), procedural skills (opening, locating, and saving documents in an operating system), and cognitive skills (navigating menus and visual cues in a user interface) (Aviram, Aharon & Eshet-Alkalai, Yoram, 2006, p. 1).

IEEE’s DQ Framework identifies data literacy as one of the competencies necessary to foster competitiveness in the digital economy through solving global challenges. IEEE defines data literacy as “finding, reading, evaluating, synthesizing, creating, adapting and sharing information, media and technology”(IEEE, 2021, p. 15) Similarly, “data literacy is our ability to read, write and comprehend data. More than that, it’s the ability to derive meaningful and useful information from data and apply this to create better products, services, and experiences” (Crofts, 2022).

While the IEEE DQ Framework encompasses a wider range of digital and data literacy skills, the Databilities framework (Data to the People, 2022) focuses on developing and measuring specific skills and knowledge required for interacting with and making decision informed by data. It includes four components: data awareness, data analysis, data interpretation, and data ethics.

The Databilities framework assesses fifteen core competencies across the dimensions of reading, writing, and comprehension to assess individual

data literacy competencies. Databilities defines data literacy as the “ability to read, write, and comprehend” data with different levels of guidance, ranging from accomplishing simple tasks with instruction or guidance, to individually, and being able to assist or teach others how to perform tasks and solve problems of varying difficulty. Figure 1 describes the different levels of competency in the Databilities Framework.

### 3. METHODOLOGY

This descriptive research study aims to better understand the digital literacy skills and higher-level data literacy competencies of incoming first-year college students at two different New England institutions which have a business focus. This study extends the work of McCarron and Frydenberg (2023) which focused on digital skills of first-year students, and limiting this study to first-year students provides insights into the data and digital literacy skills they have upon entering college. The paper analyzes students’ digital and data skills given a self-assessment of student’s tech savviness, as presented in Figure 2. The authors define tech savviness as the extent to which one is informed or proficient about the use of digital technologies and devices.

	Skills	Tech Savvy
<b>Digital Literacy</b>	RQ1a	RQ1b
<b>Data Literacy</b>	RQ2a	RQ2b

**Figure 2. Factors guiding research questions**

This matrix inspires the research questions which guide this study:

- RQ1a: What digital literacy skills do first-year students possess?
- RQ1b: Is there a discrepancy between students' digital literacy skills based on self-assessment of their own tech savviness?
- RQ2a: What higher level data literacy skills do first-year students possess?
- RQ2b: Is there a discrepancy between students' data literacy skills based on self-assessment of their own tech savviness?

### Data Collection

To evaluate the research questions, the authors created an online survey using Qualtrics (see Appendix A) which asked students to provide the following information:

- Demographics including age, gender, ethnicity, first-generation status, location of high school
- Hardware information including college and high school computers used, how often devices are used, which devices are used for which tasks
- High school information including IT topics learned as part of a course, programming languages studied
- Self-ranking of whether respondents consider themselves "tech-savvy"
- Self-ranking of their ability to complete specific tasks using a computer or common productivity applications (word processing, spreadsheets, presentation software, databases, email, operating system, cloud storage, web browser, digital media, other tasks)
- Self-ranking of data literacy framework competencies (discovery, quality, collection, management, analysis, interpretation, visualization, presentation, and decision making)
- Identification of the most important technology skills that students feel they need for success in college and their careers

The survey included multiple choice, multiple answer, open-ended questions, and scaled responses using Likert-type scales. Competency-based questions were reviewed by several IT professors and authors of IT textbooks to affirm that the researchers' assessments of beginner, intermediate, and expert tasks were accurate.

Digital literacy competency questions were created after reviewing the literature and standards created by JISC (2014), IEEE (2021) and information literacy value rubrics by AACU

(2013).

Data literacy questions and corresponding factor level responses directly followed the Dataabilities Framework (Data to the People, 2022). While the framework assesses skills across 15 core competencies, several of these were advanced skills that had no means of measuring lower-level skills and therefore were excluded from this study. As a result, we surveyed students based on 10 of the 15 core competencies. These 10 competencies broken out along their three categories included:

- Reading
  - Data Discovery
  - Quality – Trustworthiness
  - Quality – Errors
- Writing
  - Collection
  - Management
- Comprehension
  - Analysis
  - Interpretation
  - Visualization
  - Presentation
  - Decision Making

### Sample

Students enrolled in either CS100 "Introduction to Information Technology" (University A) or "Introduction to Data Literacy" (University B) were offered this survey. Participating students completed this survey during the first weeks of either the Fall 2022 or Spring 2023 semester (n=1164). Both courses primarily serve first-year and transfer students (96.44%). Students self-selected into the survey by agreeing to participate (n=1132). Respondents received no remuneration for participation. Survey results for the Dataabilities questions are shown in Appendix B.

To remove missing values resulting from survey fatigue, the sample was further reduced to include students that spent greater than 6 minutes of response time (n= 982). In addition, two respondents under the age of 18 were excluded to abide by Institutional Review Board requirements. Finally, the sample was subset to remove 120 transfer students from analysis as we are interested in the skills of incoming first-year students (n=860, 74%). After eliminating incomplete surveys, the final count was n=860. 63.3% of respondents identify as male while 36.7% identify as female. 62% of respondents identified as White, 11.6% Asian or Asian Indian, 6.5% Hispanic, Latino, or Spanish, 5.1% Black or African American, 2.2% Middle Eastern or Northern Africa, and 12.6% as multiracial or

other.

### Data Analysis

Data was downloaded from Qualtrics in csv format and uploaded into R Studio for analysis. Manipulation and cleansing were performed to prepare a desired variable format, most noticeably compressing digital literacy survey questions to an average for the six specific tasks surveyed per skill. Principal Component Analysis (PCA) (Hotelling, 1933; Pearson, 1901), using Varimax rotation, was performed as a form of clustering to uncover important features within the data.

PCA is a statistical method used to obtain a reduced set of orthogonal linear projections given original correlated variables, where the projections, or components, are ordered by decreasing variances. We perform PCA to better understand RQ1a and RQ2a while examining if literacy skills group together and if so, which skills have been self-reported as most similar.

We look to identify if a subset of skills will emerge as technologically "strong" or "weak". This can be interpreted as a proxy for a group of skills being taught, or lack thereof, at the High School level. Orthogonalization allows us to determine that skill groups, components, are uncorrelated from each other. Given this objective, PCA is an appropriate method to identify unique skill sets belonging to students.

Each component consists of factor loadings, for all variables, indicating their relationship to the component. According to Hair et. al. (2006), factor loading coefficients of at least 0.5 specify that the variable is related to the component. Greater coefficients indicate larger amounts of feature variation retained in the component. We use 0.5 as our coefficient cutoff for variable retention in every component. Using this threshold, we follow conventional reporting techniques of PCA in removing values less than the cutoff to emphasize the underlying structure of each component (Hair, 2006).

A limitation of PCA is that it does not handle outliers very well. We feel comfortable using this technique given the construction of our instrument in that all survey response levels are homogenous. We employed two thresholds to determine the optimal number of components (i) eigenvalues greater than approximately 1 and (ii) cumulative proportion of variance criterion.

The first approach only keeps components with approximately one or more variables worth of

information in it, the second states that at least 90% of the variance in the sample must be accounted for in the components. Given these guidelines, we found the suitable number of components to be 9 for digital skills and 7 for data literacy skills.

## 4. RESULTS

Our analysis aims to provide a more nuanced understanding of digital and data literacy skills of first-year college students by dividing the group into those who consider themselves "tech savvy" and those who do not.

### Tech Savviness

While the Chronicle of Higher Education survey (Tugend, 2023) found 97% of students considered themselves to be highly or somewhat digital proficient, a similar question in this study, in which students evaluated the extent to which they agreed with the statement "I consider myself to be tech-savvy," had vastly different results. Our results showed that only 32.6% agreed or strongly agreed with the statement, 40.4% were neutral, and 26.9% either disagreed or strongly disagreed (see Figure 3). This suggests that the wording of questions related to understanding technology skills is critical. While students may have bought into the notion of being digital natives, this notion is quickly challenged when students are asked more pointed questions.

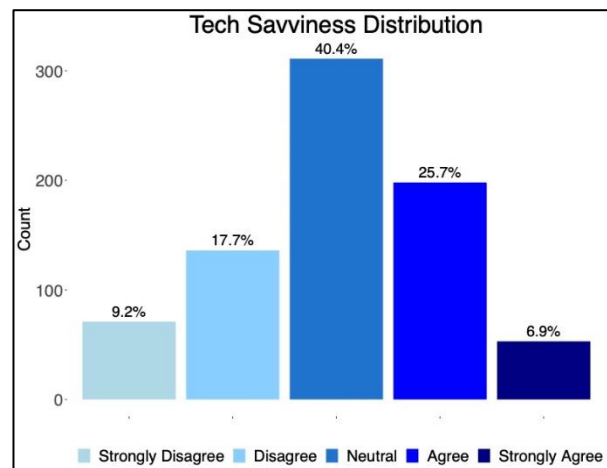


Figure 3. Self-Reported Tech Savviness

Throughout the rest of the analysis, we classified those who responded "agreed" or "strongly agreed" as being "tech savvy", those that responded "disagreed" and "strongly disagreed" as not being "tech savvy", and the remainder were kept "neutral" as shown in Figure 3. When comparing the "high tech savvy" versus the "low



tech savvy” groups, the “neutral” group was ignored.

**Analysis of Digital Literacy Skills and Data Literacy Skills of Tech Savvy Students**

We begin analyzing the technology related skills by looking to see if there is a correlation between self-reported tech savviness and our measures for both the digital skills and data literacy (See the correlation matrix in Figure 4).

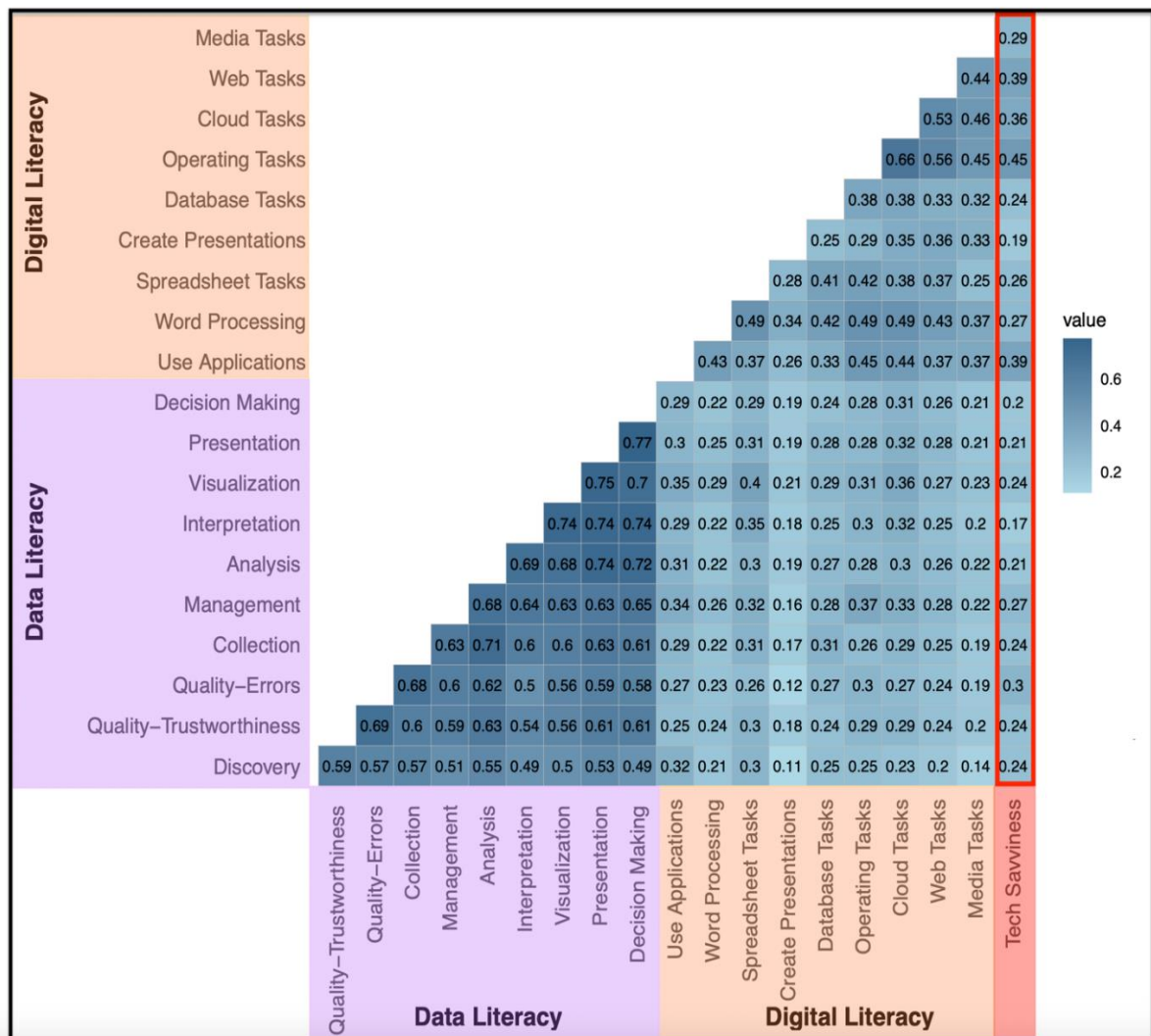
First, we notice that there are stronger correlations between the data literacy components (shaded purple) than there are among the digital literacy components (shaded orange).

Second, it is interesting to note that there is not

a strong correlation between any specific digital literacy skills. We may have expected to see high correlations between the “office suite” of skills (word processing, spreadsheets, presentations), but this was not the case.

Finally, in looking at the skills and their relationship to our tech savviness indicator (last column boxed in red), unsurprisingly, all measures are positive indicating that those who consider themselves tech savvy also report higher scores across the entire question dataset for both digital literacy and data literacy skills. However, no common suite of skills stands out amongst the tech savvy, as all correlations rate either negligible (0.0 to 0.30) or low (0.30 to 0.50) using the interpretation set forth by Hinkle and Wiersma (2003).

**Correlation Matrix**



**Figure 4. Correlation Matrix of Digital Literacy and Data Literacy Skills**

	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6	Comp 7	Comp 8	Comp 9
Use Applications		0.93							
Word Processing							0.89		
Spreadsheet Tasks					0.93				
Create Presentations	0.96								
Database Tasks			0.94						
Operating Tasks									0.85
Cloud Tasks								0.86	
Web Tasks						0.90			
Media Tasks				0.93					

**Figure 5. Principal Component Analysis (Digital Literacy)**

### Digital Literacy Skills

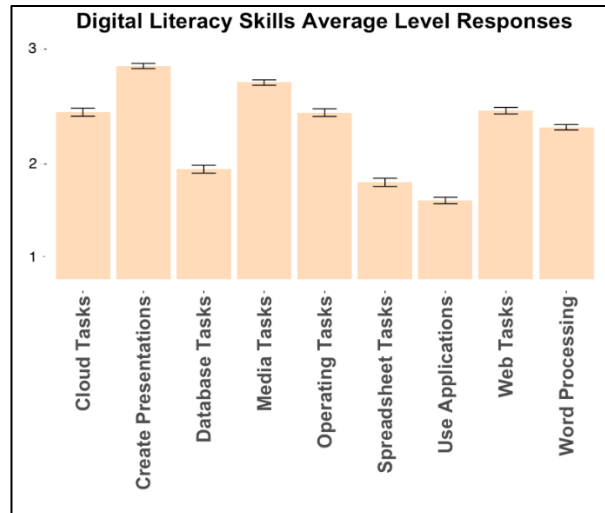
We next wanted to know whether there was a core set of digital literacy skills that all students brought with them to college. Understanding this would allow us to focus on skills often taught in first-year technology courses that needed more attention versus those that all or most students possessed before coming to college.

To reduce the number of dimensions in the data, we performed a principal component analysis to find relationships between sets of digital literacy skills that all students possess. Figure 5 shows the result of that analysis and, surprisingly, we found that there were no core set of digital literacy skills that merged into one principal component.

The first component, containing the largest sample variation, is comprised mostly of the create presentations skill. All components form singletons, meaning that they are approximately built upon the original set of features. The singleton structure confirms the results that we obtain in the correlation matrix which shows low correlation between digital literacy skills.

Looking at the average response for the digital literacy skills, as shown in Figure 6, the digital skills that students were most comfortable with include creating presentations and performing media tasks, while they were least comfortable with using applications, performing spreadsheet tasks, and performing database tasks.

We next consider the digital literacy skills broken out by tech savviness (Figure 7). Both groups appear comfortable with creating presentations, but that is the only digital literacy skill that is positive for both groups. For the tech savvy group, a majority are comfortable with media, operating (system), cloud, and web tasks, while their responses related to database tasks, using applications, spreadsheet, and word processing tasks are neutral.

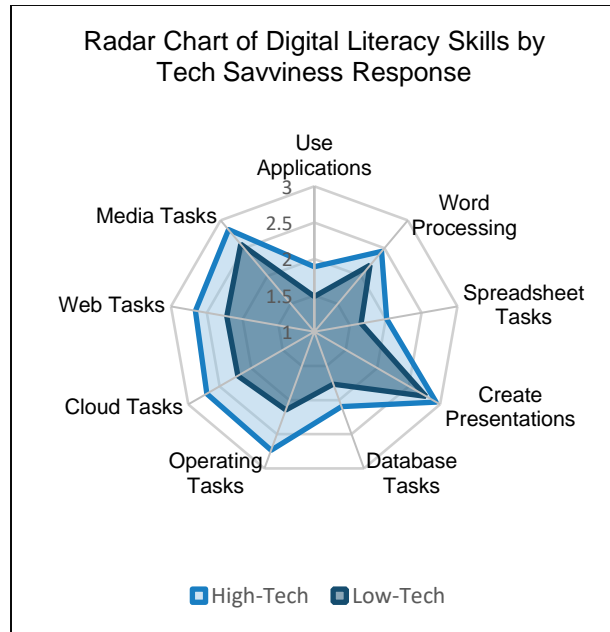


**Figure 6. Digital Literacy Skills**

For those who do not consider themselves tech savvy the only skills the majority of the students feel comfortable with are presentations and media tasks. The one skill this group clearly felt uncomfortable with was using applications. On all other skills the respondents were neutral.

A general pattern emerges across all digital literacy skills measured; there is a clear gap between the skills of the two groups. The smallest difference between the two groups was in creating presentations, the digital skill in which both groups also scored highest. The weakest skills for both groups were in using applications, spreadsheet tasks, and database tasks. Operating (system) tasks, cloud tasks, and using applications are the areas that represent the greatest skill differences between the two groups.

The concurrent trending of analysis, presentation, and decision making suggest that students are aware that they need to review and evaluate data before they can share it with others or understand it well enough to make informed decisions based on it.



**Figure 7. Radar Chart of tech skills for high- and low-tech savviness**

**Data Literacy Skills**

We followed the same approach used to examine the digital skills when evaluating the data literacy questions and responses. We began by performing the principal component analysis to better understand whether certain data literacies correlated with each other. However, unlike the digital skills, we did have principal components emerge with more than one data literacy competency (Figure 8).

Our first principal component showed that analysis, presentation, and decision making trended together. Our second principal component grouped interpretation and visualization.

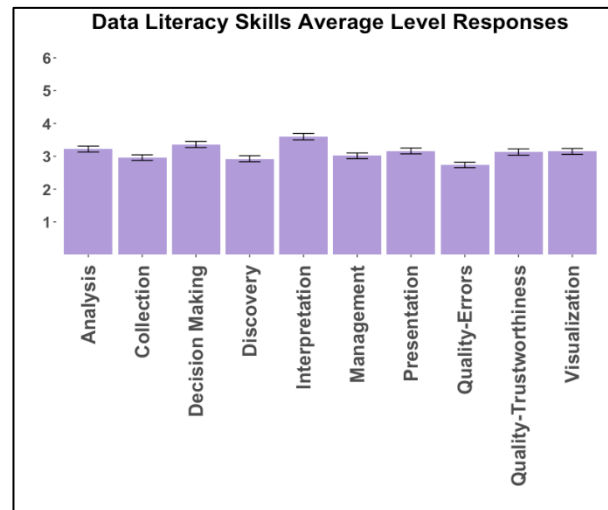
No other skills combined under a single principal

	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6	Comp 7
Discovery				0.89			
Quality-Trustworthiness						0.82	
Quality-Errors							0.78
Collection			0.79				
Management					0.82		
Analysis	0.63						
Interpretation		0.70					
Visualization		0.80					
Presentation	0.66						
Decision Making	0.74						

**Figure 8. Principal Component Analysis (Databilities Data Literacy Skills)**

component meaning that first-year students do not have a common skillset that they bring with them to college.

Figure 9 details the average level of data skills for students and serves to determine if the ability to comprehend data is being 'taught or forgot'. Students claimed to be better at analysis, decision making, and interpreting data, whereas respondents lacked skills in determining the quality of data, discovering new data, and collecting data.



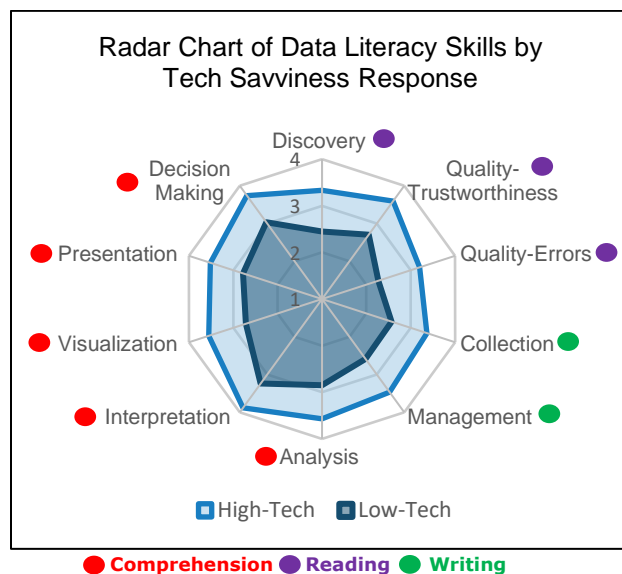
**Figure 9. Data Literacy Skills**

We next consider the data skills broken down by those who are less and more tech savvy. Similar to the digital skills findings, students with higher tech savviness consistently score at higher levels across all categories than students with self-reported lower tech savviness. With both groups, the weakest skills appear to be in the areas of Discovery and Quality – Errors.

These results suggest that while both groups have data comprehension skills, many performing at higher levels, even the majority of tech savvy students can improve their data literacy skills in these areas.

Additionally, we observe that these five clustered competencies all belong to the Comprehension branch of data literacy, as defined by Databilities, and shown in red in the radar chart in Figure 10. The grouped components represent topics often taught in many secondary schools.

Outcomes are corroborated in the radar chart (Figure 10). Datalabels are color coded to match the three branches of data literacy: purple (reading), green (writing), red (comprehension). We observe the chart trend toward the direction of the red colored labels, indicating outperformance from both tech savvy groups in the areas of decision making, presentation, visualization, interpretation, and analysis.



**Figure 10. Radar Chart of Databilities skills for high- and low-tech savviness**

## 5. DISCUSSION AND CONCLUSIONS

The results of this study suggest that while only 32% of those surveyed at both participating institutions agreed or strongly agreed that they considered themselves to be tech savvy, incoming first-year college students need to further develop their data literacy skills, especially in the areas of collecting, managing, and evaluating the quality of data. Most students were performing at the lowest levels of the digital literacy framework in all areas except for interpretation, meaning that they can complete

simple tasks with instruction or some guidance. We hypothesize that their higher levels of assessment regarding data visualization is due to prior study of Excel or other spreadsheet software, where students are often introduced to making different types of charts and graphs.

Our data did not show a strong correlation between any specific digital literacy skills. We anticipated high correlations between office productivity tools because many students come to college having at least basic knowledge of them. Many productivity tools incorporate “wizards” or artificial intelligence capabilities to guide the user into creating impressive presentations or complex documents, even though their prior experience is beginner level at best.

While we expected that more specialized skills such as operating (system) tasks, database tasks, and cloud tasks were not uniformly learned, we did expect basic productivity skills related to word processing, spreadsheets, and presentations to trend together. The assumption that students who knew how to use one productivity tool were more likely to know how to use the others was incorrect. This suggests that there is no common core of skills being taught at high schools where our first-year students attended, and that introductory technology classes in college cannot assume that knowledge in one skill suggests knowledge in another skill.

Although digital natives are adept with using technology for personal pursuits, the misconception remains that they are also qualified to complete well defined tasks and problems using data independently or show others how to complete simple or complex problems and tasks.

Our findings suggest that students are trained to comprehend data rather than read or write it. Insights can be developed into proper course preparation at the collegiate level by spending additional training within reading and writing data. To truly comprehend data, one must be able to collect and manage data, and evaluate its quality.

The authors acknowledge limitations of this study: first, the study was based on students’ self-assessment, rather than an objective assessment of their abilities to complete various tasks that demonstrate their data and digital literacy skills, and second; the study was conducted at two business-focused institutions. Results may vary among populations of students

at liberal arts colleges, community colleges, or other types of institutions of higher education.

The authors recommend that institutions take advantage of the fact that students believe productivity skills are important (spreadsheets, programming, word processing), and further introduce them to tools and techniques for presenting, visualizing, interpreting, analyzing, managing, collecting, and evaluating the quality of data so that they will develop skills necessary for success in the workplace.

The data literacy competencies that students come in with when they begin their college careers are generally not sufficient to succeed in college or in the workplace. Additional instruction will be necessary to make up the areas where skills are deficient and strengthen those areas where they have some prior experience.

Students who develop data literacy skills will know what questions to ask, how to make informed decisions, and what to say when sharing their results with others. Developing data literacy skills is necessary to make informed decisions based on evaluating data that will position both college students and their future employers to succeed.

This work contributes to the growing body of knowledge around our understanding of tech-focused literacies by beginning to explore the differences found in data literacy versus digital literacy. With the rise in mass adoption of artificial intelligence tools such as ChatGPT, assessing underlying literacies of technology users will be instrumental in evaluating their capabilities and adoption rates of current technologies such as AI.

## 6. REFERENCES

American Association of Colleges and Universities. (2013, July). *VALUE Rubrics*. AAC&U.  
<https://www.aacu.org/initiatives/value-initiative/value-rubrics/value-rubrics-information-literacy>

Aviram, A. & Eshet-Alkalai, Y. (2006). Towards a theory of digital literacy: Three scenarios for the next steps. *European Journal of Open, Distance, and E-Learning*, 9(1), 11.

Crofts, J. (2022). *2022 Global Data Literacy Benchmark*.

Data to the People. (2022). *MyDataBilities | your personal data literacy assessment*. MyDataBilities.  
<https://www.mydatabilities.com>

Frank, M., Walker, J., Attard, J., & Tygel, A. (2016). Data Literacy—What is it and how can we make it happen? *The Journal of Community Informatics*, 12(3), Article 3.  
<https://doi.org/10.15353/joci.v12i3.3274>

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Pearson University Press.

Hartzel, K. S., & Ozturk, P. (2022). Analytics in the Business School: Insights from the Literature. *Journal of Information Systems Education*, 33(2), 169-181.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417-441.  
<https://doi.org/10.1037/h0071325>

IEEE. (2021). IEEE Standard for Digital Intelligence (DQ)—Framework for Digital Literacy, Skills, and Readiness. *IEEE Std 3527.1-2020*, 1-47.  
<https://doi.org/10.1109/IEEESTD.2021.9321783>

Joint Information Systems Commission. (2014, March 6). *Developing digital literacies*. Jisc.  
<https://www.jisc.ac.uk/guides/developing-digital-literacies>

Leon-Urrutia, M., Taibi, D., Pospelova, V., Splendore, S., Urbsiene, L., & Marjanovic, U. (2022). Data Literacy: An Essential Skill for the Industry. In B. Lalic, D. Gracanin, N. Tasic, & N. Simeunović (Eds.), *Proceedings on 18th International Conference on Industrial Systems - IS'20* (pp. 326-331). Springer International Publishing.  
[https://doi.org/10.1007/978-3-030-97947-8\\_43](https://doi.org/10.1007/978-3-030-97947-8_43)

McCarron, E., & Frydenberg, M. (2023). Digitally Prepared for Success? Technology Skills of Incoming First-Year College Students. *Information Systems Education Journal* 21(3), 70-90.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The*

*London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572.  
<https://doi.org/10.1080/14786440109462720>

Tugend, A. (2023). The Myth of the Digital Native: How Colleges Are Dealing With Student Digital-Literacy Gaps. *The Chronicle of Higher Education*.

Prensky, M. (2001). Digital Natives, Digital Immigrants part 2: Do they really think differently? *On the Horizon*, 9(6), 1-6.

**Editor's Note:**

*This paper was selected for inclusion in the journal as a 2023 ISCAP Conference Distinguished Paper. The acceptance rate is typically 7% for this category of paper based on blind reviews from six or more peers including three or more former best papers authors who did not submit a paper in 2023.*

## Appendix A. Survey Instrument

### Willingness to Participate

You are invited to participate in a research study on the use of technology as it relates to IT education. You were selected as a possible participant because you are enrolled in CS100 or DS 101. Please read this form and ask any questions you may have before agreeing to be in the study.

This study asks about your experiences learning about and using technology from high school to the present. If you agree to be in this study, you will be asked to answer several questions about your technology experiences. Your responses will be recorded and downloaded for analysis. The survey should take approximately 15 minutes to complete.

Click YES to participate. Click NO to withdraw from this survey.

- Yes, I will participate.
- No, I will not participate.

Skip To: End of Survey If Click YES to participate.

### Demographics

In which course and section are you enrolled?

▼ (drop-down list of sections and instructors)

How old are you?

- Under 18
- 18-20
- 21 or over

Skip To: End of Survey If How old are you? = Under 18

With which gender identity do you most identify?

- Please choose one
- female
- male
- transgender female
- transgender male
- gender variant/non-conforming
- Not listed
- Prefer not to answer

What is your race or ethnicity (please select all that apply):

- White
- Hispanic, Latino, or Spanish origin
- Black or African American
- Asian or Asian Indian
- Native American or Alaska Native
- Middle Eastern or North African
- Native Hawaiian or Other Pacific Islander
- Another race, ethnicity or origin

Are you a:

- First-Year Student
- Second, Third, or Fourth-Year Student
- Transfer Student

*Display This Question:*

*If Are you a: = Transfer Student*

Did you take a technology class at your previous college?

- Yes
- No

*Display This Question:*

*If Did you take a technology class at your previous college? = Yes*

Did you receive credit for your technology course at a previous college?

- Yes
- No

The computer that you currently use for school is a:

- Mac
- Windows Computer
- Chromebook
- Other

The computer that you currently use for college was:

- Purchased through the college / university
- Purchased on my own
- One you had at home and/or used in high school

Where did you attend high school?

- New England (MA, NH, CT, RI, VT, ME)
- Elsewhere in the US
- Outside of the US

Are you the first person in your family to go to college?

- Yes
- No

What is your major?

---

## **High School Experience**

Where would you most often use a computer when completing homework assignments in high school?

- A computer in a lab at school
- A computer at the public library
- A computer at home

Which of these topics, if any, did you learn as part of a course in high school?

- Programming
- Web Design / Making Websites
- Digital photography
- Digital video production
- Digital audio production
- Computer Science
- Word Processing
- Spreadsheets
- Presentation Software
- Email
- Databases
- How to use the World Wide Web



- Social Media

*Display This Question:*

*If Which of these topics, if any, did you learn as part of a course in high school? = Programming*

Which programming languages or environments did you learn in high school? (Check all that apply)

- Alice
- Scratch
- Java
- JavaScript
- C or C#
- Python
- HTML
- Visual Basic
- Other

Which computer or device did you use most in high school?

- Google Chromebook (Desktop or Laptop)
- Windows Computer (Desktop or Laptop)
- Mac Computer (Desktop or Laptop)
- Linux Computer (Desktop or Laptop)
- iPad (41)
- Android Tablet
- Other
- 

Did you take an AP Computer Science Test?

- Yes
- No

Did you work or volunteer as a help desk, computer lab assistant, or technology tutor while in high school?

- Yes
- No

### **Your Current Use of Computers and Devices**

How often do you use these devices?

	Rarely/Never	Weekly	Daily
Iphone	0	0	0
Android phone	0	0	0
Other mobile phone	0	0	0
Smart Watch	0	0	0
Ipad or Tablet	0	0	0
Raspberry Pi	0	0	0
Laptop	0	0	0
Gaming Device	0	0	0
VR Headset	0	0	0
Smart Speaker (Alexa/Google Home)	0	0	0

Which device(s) do you use to perform the following tasks?

	Laptop/Desktop	Tablet	Mobile Phone
Watch a video	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Send email	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Send a text message	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Visit websites	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Use a search engine	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Edit a document	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Make an online purchase	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Listen to music	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Take a photo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Edit a photo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Post a photo to social media	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

How many times per day do you currently use the following social media apps?

	Never	1 to 5 times	6 to 10 times	More than 10
Facebook	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Twitter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Instagram	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Snapchat	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
LinkedIn	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### Your Tech Skills

Keep up the good work! You're about half-way there! This section has 11 questions related to tasks you might know how to complete with various software applications.

Rate your ability to use each of these *applications* before you came to college.

	Beginner	Intermediate	Expert
Word Processing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Spreadsheets	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Presentation Software	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Database Software	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Email Software	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Your computer's operating system (Windows or Mac OS)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cloud Storage (Google Drive, Dropbox, OneDrive, etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Web Browsers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Digital Media Editing (photos, videos, music)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online and Social Collaboration	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online calendar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Rate your ability to perform these *word processing* tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Format text using bold, underline, or different fonts and sizes to change the appearance of a document	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Add page numbers to a document	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Perform a mail merge	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Use the references feature to create a bibliography	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Add a picture to a document	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Use the format painter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Rate your ability to perform these *spreadsheet* tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Create a basic pie, line, or bar chart	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Write formulas in excel using references such as \$A\$1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Create a pivot table	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Write formulas to look up values in a table	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Write a formula to calculate the sum of two numbers if they are equal, and calculate their product if not equal	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Write a formula to calculate the sum of a range of cells	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Rate your ability to perform these *presentation software* tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Use fonts and styles to change the appearance of a slide	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Modify a slide's background	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Apply slide transitions and animations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Apply a theme to your presentation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Insert shapes or images	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Modify a slide's layout pattern	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Rate your ability to perform these *database* tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Create a table to store data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Add data to a table	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Create a report	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Add a primary key	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Relate information in two different tables	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Find records that match a specific condition (such as all employees from Connecticut)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Rate your ability to perform these *e-mail* tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Send a message	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Send a message with an attachment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Send a message encrypted for security	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Organize messages in folders	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reply to everyone who received a message	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Delete messages from your inbox	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Rate your ability to perform these *operating system* tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Create a new folder	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Zip or compress a file	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
See how much free storage remains on your hard drive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Install or uninstall a program	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Copy a file from one folder to another	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Back up files on your computer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Rate your ability to perform *these cloud storage service* (such as OneDrive, Google Drive, or Dropbox) tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Upload your files to the cloud from your computer or mobile device	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Access your files in the cloud from your computer or mobile device	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Share a link to a document stored in the cloud with someone else	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Synchronize your files stored in the cloud across multiple devices	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Edit a document stored in the cloud collaboratively at the same time as someone else	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Specify folders on your computer to store on the cloud	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Rate your ability to perform these *web browser* tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Accept cookies	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bookmark a website	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
View HTML source code of a website	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Clear your browser's cache	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Build a search query to limit search results using modifiers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Get directions using an online mapping service	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Rate your ability to perform these *digital media* tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Crop or resize a photo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cut a section from a video clip	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Flip or rotate an image	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Add a title or credits to a video	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Publish a video to YouTube	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Make a podcast	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Rate your ability to perform these *social media* tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Write a post to a blog or social media site	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Set up a blog or website using a content management tool	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Write a comment on a blog or social media post	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Make a video call or participate in a video conference	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Create an appointment on an online calendar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Specify which friends or groups of friends can see your posts to Facebook	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### Important to Know

To what extent do you agree with each of these statements?

Coding is a valuable skill to have.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am concerned about my privacy online.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would like to build mobile apps.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I consider myself to be tech-savvy.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I understand how technology works and know how to use it responsibly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can explain good practices for selecting a strong password for my accounts online.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can explain different ways to protect my technology systems and information from unethical users.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can explain basic practices that contribute to a website's accessibility to people with disabilities.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

What three technology skills do you think are the most important that will help prepare you for your college education then into your future career?

---

---

---

---

### Your Data Skills

Which of these statements best describe you?

- With guidance, I can open and use datasets provided to me
- I can open and use datasets provided to me
- I can identify and access data I need from a range of sources provided to me
- I can identify, locate, and access data I need from a variety of sources
- I can assist others to identify and access data they need from a range of provided sources
- None of these describe me

Which of these statements best describe you?

- With guidance, I can identify whether the data provided to me is trustworthy
- I can identify whether the data provided to me is trustworthy
- I can identify whether data I use is trustworthy
- I can identify whether data I use is trustworthy and locate alternative sources if required
- I can assist others to identify whether data they use is trustworthy
- None of these describe me

Which of these statements best describe you?

- With guidance, I can determine whether data provided to me contains any errors or problems
- I can determine whether data provided to me contains any errors or problems
- I can determine whether data contains any errors or problems
- I can identify and take actions to correct any errors or problems in a range of data sources
- I can assist others to identify and errors or problems in a range of data sources
- None of these describe me

Which of these statements best describe you?

- With guidance, I can collect simple data in a format provided to me
- I can collect simple data in a format provided to me
- I can collect data in simple and more complex forms
- I can collect data in a variety of forms to support my needs
- I can assist others to collect data in forms to support their needs
- None of these describe me

Which of these statements best describe you?

- With guidance, I can navigate structured organization systems (including folders, directories, and file naming) to find data I need
- I can navigate structured organization systems (including folders, directories, and file naming) to find data I need
- I can navigate database models to locate data related to my needs
- I can navigate database models to locate and retrieve data related to my needs
- I can assist others to navigate organizational systems and database models to locate and retrieve data related to their needs
- None of these describe me

Which of these statements best describe you?

- With guidance, I can ask and answer basic questions with data provided to me
- I can ask and answer basic questions with data provided to me
- I can ask and answer a range of questions with data provided to me
- I can ask and answer a range of questions using a variety of data sources
- I can assist others to ask and answer a range of questions using a variety of data sources
- None of these describe me

Which of these statements best describe you?

- With guidance, I can read and understand simple tables, charts, and graphs provided to me
- I can read and understand simple tables, charts, and graphs provided to me
- I can read and understand simple tables, charts, and graphs
- I can read and understand a range of tables, charts, and graphs
- I can assist others to read and understand a range of tables, charts, and graphs
- None of these describe me

Which of these statements best describe you?

- With guidance, I can use templates to create simple tables and charts to visually present data provided to me
- I can use templates to create meaningful tables and charts to visually present data provided to me
- I can create meaningful tables and charts to visually present simple datasets
- I can create meaningful tables and charts to visually present a range of data sources
- I can assist others to create meaningful tables and charts to visually present simple data sources
- None of these describe me

Which of these statements best describe you?

- With guidance, I can verbally describe key points of interest in data provided to me
- I can verbally describe key points of interest in data provided to me
- I can verbally describe simple datasets and data visualizations
- I can verbally describe a range of data sources and data visualizations
- I can assist others to verbally describe provided data sources and data visualizations
- None of these describe me

Which of these statements best describe you?

- With guidance, I can use data provided to me to support my decision-making process
- I can use data provided to me to support my decision-making process
- I can use data to inform my decision-making process
- I can use data to analyze and inform my data-making process
- I can assist others to use data to analyze and inform their decision-making process
- None of these describe me

End of Survey

## Appendix B. Survey Results of Databilities Questions

Discovery			Analysis		
	%	Count		%	Count
With guidance, I can open and use datasets provided to me	43.27%	518	With guidance, I can ask and answer basic questions with data provided to me	27.30%	323
I can open and use datasets provided to me	12.45%	149	I can ask and answer basic questions with data provided to me	25.95%	307
I can identify and access data I need from a range of sources provided to me	14.29%	171	I can ask and answer a range of questions with data provided to me	18.09%	214
I can identify, locate, and access data I need from a variety of sources	16.37%	196	I can ask and answer a range of questions using a variety of data sources	15.72%	186
I can assist others to identify and access data they need from a range of provided sources	4.51%	54	I can assist others to ask and answer a range of questions using a variety of data sources	6.51%	77
None of these describe me	9.11%	109	None of these describe me	6.42%	76
Total	100%	1197	Total	100%	1183
Quality-Trustworthiness			Interpretation		
	%	Count		%	Count
With guidance, I can identify whether the data provided to me is trustworthy	34.23%	407	With guidance, I can read and understand simple tables, charts, and graphs provided to me	20.92%	246
I can identify whether the data provided to me is trustworthy	19.68%	234	I can read and understand simple tables, charts, and graphs	22.36%	263
I can identify whether data I use is trustworthy	15.73%	187	I can read and understand simple tables, charts, and graphs	15.48%	182
I can identify whether data I use is trustworthy and locate alternative sources if required	16.23%	193	I can read and understand a range of tables, charts, and graphs	22.79%	268
I can assist others to identify whether data they use is trustworthy	6.73%	80	I can assist others to read and understand a range of tables, charts, and graphs	13.18%	155
None of these describe me	7.40%	88	None of these describe me	5.27%	62
Total	100%	1189	Total	100%	1176
Quality-Errors			Visualization		
	%	Count		%	Count
With guidance, I can determine whether data provided to me contains any errors or problems	44.01%	522	With guidance, I can use templates to create simple tables and charts to visually present data provided to me	31.13%	368
I can determine whether data provided to me contains any errors or problems	18.30%	217	I can use templates to create meaningful tables and charts to visually present data provided to me	23.86%	282
I can determine whether data contains any errors or problems	15.68%	186	I can create meaningful tables and charts to visually present a range of data sources	15.06%	178
I can identify and take actions to correct any errors or problems in a range of data sources	9.95%	118	I can create meaningful tables and charts to visually present simple datasets	17.34%	205
I can assist others to identify and errors or problems in a range of data sources	2.87%	34	I can assist others to create meaningful tables and charts to visually present simple data sources	6.26%	74
None of these describe me	9.19%	109	None of these describe me	6.35%	75
Total	100%	1186	Total	100%	1182
Collection			Presentation		
	%	Count		%	Count
With guidance, I can collect simple data in a format provided to me	33.73%	401	With guidance, I can verbally describe key points of interest in data provided to me	30.50%	362
I can collect simple data in a format provided to me	28.34%	337	I can verbally describe key points of interest in data provided to me	23.59%	280
I can collect data in simple and more complex forms	11.44%	136	I can verbally describe simple datasets and data visualizations	17.44%	207
I can collect data in a variety of forms to support my needs	13.79%	164	I can verbally describe a range of data sources and data visualizations	15.92%	189
I can assist others to collect data in forms to support their needs	4.79%	57	I can assist others to verbally describe provided data sources and data visualizations	5.64%	67
None of these describe me	7.91%	94	None of these describe me	6.91%	82
Total	100%	1189	Total	100%	1187
Management			Decision Making		
	%	Count		%	Count
With guidance, I can navigate structured organization systems (including folders, directories, and file naming) to find data I need	32.04%	380	With guidance, I can use data provided to me to support my decision-making process	25.95%	308
I can navigate structured organization systems (including folders, directories, and file naming) to find data I need	29.17%	346	I can use data provided to me to support my decision-making process	23.50%	279
I can navigate database models to locate data related to my needs	13.24%	157	I can use data to inform my decision-making process	18.37%	218
I can navigate database models to locate and retrieve data related to my needs	12.48%	148	I can use data to analyze and inform my data-making process	19.97%	237
I can assist others to navigate organizational systems and database models to locate and retrieve data related to their needs	5.90%	70	I can assist others to use data to analyze and inform their decision-making process	6.49%	77
None of these describe me	7.17%	85	None of these describe me	5.73%	68
Total	100%	1186	Total	100%	1187



## Teaching Case

# Teaching Data Literacy Using Titanic Survival Factors

Mark Sena  
Xavier University  
sena@xavier.edu

Thilini Ariyachandra  
Xavier University  
ariyachandrat@@xavier.edu

Xavier University  
Cincinnati, OH 45207

### Abstract

The Titanic disaster is a topic that continues to fascinate. As the importance of analytics continues to grow in industry, data literacy skills have become increasingly important in business education. This project allows students to use the passenger data from the Titanic to build their data literacy skills using an engaging, experiential topic. The project requires students to extract, transform, describe, analyze, and draw conclusions regarding the factors that impacted survival on the Titanic. The project can be deployed using various application software and tools. We describe how the assignment can be completed using Excel, Tableau, and Python using the Pandas library.

**Keywords:** experiential learning, analytics, data literacy, Tableau, Python, Pandas, Excel

**Recommended Citation:** Sena, M., Ariyachandra, T., (2024). Teaching Data Literacy Using Titanic Survival Factors. *Information Systems Education Journal*, 22(3), pp.25-35.  
<https://doi.org/10.62273/VLIH4928>

# Teaching Data Literacy Using Titanic Survival Factors

*Mark Sena and Thilini Ariyachandra*

## 1. INTRODUCTION

The world today is characterized by a data revolution, with an unparalleled volume of data being generated, disseminated, and scrutinized every day. Data has emerged as a vital asset in the digital age, influencing individuals, businesses, and society as a whole. It offers valuable insights that can steer decision-making, enhance performance, and foster innovation (Mandinach & Gummer, 2016). For individuals in society, data is instrumental in improving personal decision-making and lifestyle management. As per a Pew Research Center survey, 69% of U.S. adults monitor at least one health, fitness, or diet metric, underscoring the significance of data in personal health management (Fox & Duggan, 2013). Additionally, a study by the National Bureau of Economic Research revealed that personal finance apps, which utilize data to offer financial insights, can help decrease spending by 15.7% (Baker, Meyer, Pagel, & Yannelis, 2018).

In the business sphere, data is crucial for strategic planning, operational efficiency, and gaining a competitive edge. Companies that base their decisions on data and analytics, rather than intuition or gut feeling, perform better across almost all metrics. A study conducted by Harvard Business School and MIT Sloan School of Management on over 2,000 public companies across various industries found that data-driven companies had superior financial performance, were more likely to survive, and were more innovative (Garg, 2023). A recent study by McKinsey discovered that data-driven companies outperform their competitors by up to 20% (Garg 2023). At the societal level, data can inform policy-making, enhance public services, and drive social change. The United Nations has highlighted the importance of data for achieving the Sustainable Development Goals, with a call for a "data revolution" to tackle global challenges such as poverty, inequality, and climate change (United Nations, 2014). Furthermore, a Pew Research Center study found that 65% of Americans believe that government use of data to make decisions can improve public services (Rainie & Anderson, 2017).

Given its impact on the individual, business and society, the ability to understand and utilize data effectively has become a highly sought after skill in industry. As society has grown more dependent on data, it has become imperative to ensure that all individuals possess the necessary skills to be data literate. Adopting data literacy throughout the organization has begun to gain attention as a means of ramping up data driven decision making in day to day as well as strategic operations in companies. All data-related roles are growing at a rapid pace with data scientists roles expected to grow at 36 percent from 2021 to 2031 according to the U.S. Bureau of Labor Statistics (2023). According to a recent study of expert interviews and survey responses from over 1,200 global C-level executives and 6,000 employees conducted by Qlik (2023), data literacy will be the most sought-after skill by 2030. Furthermore, 85 percent of executives predict that data literacy will be as crucial in the future as computer literacy is in the present.

As the demand for data literacy skills continues to grow it falls upon institutions of higher education to arm future employees with the necessary data savvy required in organizations. This data savvy, also known as data literacy, empowers individuals to collect, manage, evaluate, and apply data competently to decision making (Mandinach & Gummer, 2016). Teaching data literacy skills has its own challenges. Many students express concerns and face challenges in acquiring data literacy skills. One of the primary issues students face is a lack of confidence in their abilities to interpret and analyze data. According to a study by Prinsloo and Slade (2017), students often feel overwhelmed by the complexity of data and the technical skills required to manipulate and interpret it. This lack of confidence can lead to a fear of data and a reluctance to engage with it. Another common issue is the lack of understanding of the relevance of data literacy to their future careers. Many students fail to see how data literacy skills will be applicable in their chosen fields, leading to a lack of motivation to acquire these skills (Koltay, 2015). The project described in this paper attempts to introduce data literacy skills in an interesting and entertaining way that can ease students into the world of data manipulation.

## 2. OVERVIEW OF DATA LITERACY

In recent years, various definitions of data literacy have emerged. It is loosely considered the ability to understand and use data (Frank, Walker, Attard & Tygel, 2016). According to UNESCO (2004) "literacy" is the ability to define, understand, interpret, create, and calculate through relevant written and printed sources. More specifically, data literacy can be described as the ability to read, work with, analyze, and argue with data (Koltay, 2015). Some include the need for statistical analysis as part of data literacy. Xu (2023) describes it as knowing how to access data in different ways, asking questions, and making basic statistical analysis. Past literature has also described it in terms of actions as well as in terms of specific competencies that an individual should acquire. Pothier and Condon (2019) identified seven baseline data literacy competencies to help prepare business students for the workforce: (1) data organization and storage, (2) understanding data used in business contexts, (3) evaluating the quality of data sources, (4) interpreting data, (5) data-driven decision making, (6) communicating and presenting effectively with data, and (7) data ethics and security.

Upon investigating multiple definitions of data literacy, Wolff, Gooch, Cavero Montaner, Rashid and Kortuem (2016) developed a comprehensive definition for data literacy as "the ability to ask and answer real-world questions from large and small data sets through an inquiry process, with consideration of ethical use of data. It is based on core practical and creative skills, with the ability to extend knowledge of specialist data handling skills according to goals. These include the abilities to select, clean, analyze, visualize, critique and interpret data, as well as to communicate stories from data and to use data as part of a design process" (pg. 23). The definition describes seven abilities that involve data literacy: (1) selecting, (2) cleaning, (3) analyzing, (4) visualizing, (5) critiquing, (6) interpreting data, as well as (7) communicate stories. The following definition by D'Ignazio and Bhargava (2015) expresses data literacy in terms of the ability to carry out four specific actions related to data: reading data, working with data, analyzing data, and arguing with data. It is widely used in widely as a concise and easy definition that can be incorporated into business practices. The following project enables students to gain experience in nearly all of the preceding foundational abilities.

## 3. PROJECT DESCRIPTION

### Objectives and Questions for the Analysis

To analyze the Titanic data set, we are interested in learning what factors were associated with survival. These include such questions as:

- 1) Were females more likely than males to survive?
- 2) Were children or senior citizens more likely to survive than adults?
- 3) Were first or second-class passengers more likely to survive than third class passengers?
- 4) Were passengers whose cabin was located in certain locations more likely to survive?
- 5) Were passengers who paid a higher fare more likely to survive than those who paid a lower fare?
- 6) In the movie Titanic, the main character Rose survived without a lifeboat by clinging on to a board. Could she have been based on a real passenger?

### Extracting the Data

The Titanic data is widely available on the internet and has been used for analytical projects for many years. Kaggle.com (kaggle, 2023) is one common source. Note that there are various subsets of the data for projects that may differ in terms of number of records and in file type. The data set that we use in our project includes 1309 passengers. The dataset includes fourteen fields as depicted in Table 1.

For a dataset that is over one hundred years old, it is surprisingly rich. However, there are various missing data points throughout the data. The age and cabin locations of passengers is more complete for first and second class passengers than for third class passengers.

### Transforming the Data Set

To conduct the analyses in Table 1, we need to first create some new columns that will allow us to create charts and statistics. These include:

- 1) creating a numeric equivalent of the gender field.
- 2) creating a label equivalent of the survived field (so that charts can show "survived" or "died" instead of 0, 1 as labels)
- 3) creating a variable that groups age into Blank (equal to ""), children (under 18), adults, senior citizens (over 64)
- 4) creating a one letter (LEFT() function) cabin location variable

Field	Description
Survival (1, 0)	1 = Survived, 0 = Died
PClass (1,2,or 3)	Passenger Class (first, second, or third class)
Name	Name of Passenger
Sex (male, female)	Gender of Passenger
Age (numeric or blank if unknown)	Age of Passenger (if known)
Sisp	Number of Siblings/Spouses aboard
Parch	Number of Parents/Children aboard
Ticket	Ticket Number
Cabin (blank if unknown)	Cabin Number
Embarked (C, Q, or S)	Port of Embarkment: Cherbourg, Queenstown, Southampton
Boat (blank if passenger not on lifeboat)	Lifeboat number or letter
Body (blank if not recovered)	Body Number if recovered
Fare (in Pounds)	Fare Paid
Home/Dest	Home town and destination of passenger

**Table 1: Fields in Titanic Data Set**

To make these transformations in Excel, we would use simple IF functions for the first two steps, a nested IF() function for the third, and a LEFT() function for step four. In Tableau, these transformations could be done in Excel or using various actions in Tableau itself (such as creating calculated or grouped fields). In Python, after reading the dataset, students can create list variables by looping through each data then creating new fields by setting the data frame field equal to the list variable.

### Demographics and Summary Data

Before exploring the relationships, students should summarize key passenger data to establish a base line. These include:

- 1) Showing the number of survivors vs deaths
- 2) Showing the number of each gender on board
- 3) Showing the number of children, adults, and senior citizens on board
- 4) Showing the number of passengers in each class (1,2,3)
- 5) Showing the number of passengers in various fare groups.
- 6) Showing the number of passengers in each cabin class

In Excel, students could create pivot tables for each column shown above and use Count as the

Value Field Setting. See Figure 1 in the Appendix for an example of the output in Excel. In Tableau, a similar approach could be used with worksheets showing the Count measure in a row or column and a field (such as Survived) in the row or column to produce a column or bar chart. All six summary charts could also be combined into a dashboard visualization. In Python, there are various summary code options in the Pandas library such as histogram plots (`df.hist('variable')`) or tabular results using code such as `.counts` or `.size` (`df.groupby('variable').size()`)

### Relationship with Survival

Now that students have a baseline of passenger characteristics, we can examine how each factor relates to survival. These include:

- 1) Showing the average survival rate by gender
- 2) Showing the average survival rate by age group
- 3) Showing the average survival rate by passenger class
- 4) Showing the average survival rate broken down by fare groups.
- 5) Showing the average survival rate by cabin level.

In Excel, students can create pivot tables to show tabular and visualizations by selecting the survived column with Average as the Value Field setting with other fields in the Row. See Figure 2 in the Appendix for an example. In Tableau, students would simply drag the Survived field in the row or column and the other field in the opposite row or column. Students would need to adjust the Measure option for Survived from the Sum default to Average. In Python, students could use the groupby option in the Pandas library for each combination. For example, the following code would show survival by gender: `print(df.groupby('gender')['survived'].mean())`

### Exploring Multiple Survival Factors

Since there are several possible combinations of factors that could be chosen, it would be preferable to develop an interactive visual to allow the user to explore relationships. For example, a user may wish to view survival rate of adult females in first class. To do so in Excel, students can create a pivot table with various slicers that can be selected to interactively filter the data. We ask students to show both the survival rate (average of Survived) and the number of passengers (count of Survived). See Figure 3 in the Appendix for an example. In Tableau, a similar approach could be used with Filters. In Python, it is more difficult in the Pandas library to create interactive statistics without

teaching students to develop a user interface. However, students can explore several combinations of options using the `groupby` and `mean`. For example, the following would show survival rate by gender and passenger class: `print(df.groupby(['gender', 'pclass'])['survived'].mean())`.

#### **Correlation Between Survived and Factors**

To better understand the survival factors, students can compute the correlation between survival and each numeric field in the data set to measure the strength and direction of the linear relationships. In Excel, this can be conducted using the `CORREL` function (and other possible methods). Students can also create a chart comparing the relationships. See Figure 4 in the Appendix for an example of correlations that use conditionally formatted bar charts. In Tableau, students can create Scatter Plots with a trend line added that shows both the correlation and *p*-value of the relationships. In Python, a simple `df.corr()` line of code will produce a table of correlations among all numeric fields in the data set.

#### **Exploring Whether “Rose” Could Be Real**

To engage the students, especially those who have seen the movie, we ask students to examine potential passenger characteristics that meet those of the character, Rose, who was a young adult female who survived without making it onto lifeboat. In Excel, students can convert the primary data into a table then filter the fields to meet the criteria. This would include filters such as Female, Age between 15 and 35, survived = 1 and boat equal to blank (or null). See Figure 4 in the Appendix for an example. In Tableau, the students can drag the Name field (and any other field of interest) into the Row, then uncheck Aggregate Measures, and add each relevant field into the Filter box and select appropriate ranges or allowable values. In Python, this would be most easily done using a for loop with an if statement that includes all combinations of appropriate criteria and an indented statement to print name and other fields. While this task may seem extraneous to the primary analytical purpose of examining survival factors in the data, the students surprisingly struggle with the solution (particularly since they are focused on pivot tables while this task requires filtering of raw rather than summarized data).

#### **Explaining Results and Findings**

To summarize and interpret the results of the analysis, we ask students to include a written narrative addressing the six questions previously identified as objectives and research questions.

These could be embedded in worksheets in Excel, as separate Word documents or as Comments in Python. See Figure 5 in the Appendix for an example of the Findings. An important skill of analytics is to communicate and explain findings. Instructors could include a formal storytelling requirement to this portion of the assignment.

#### **4. CONCLUSIONS AND LIMITATIONS**

Limitations of this assignment include the potential for academic dishonesty (including the possible use of AI tools) and the popularity of the data set in higher education. Because students are analyzing the same data with little possibility in variation of results, the potential for collaboration across students is high. AI resources such as ChatGPT (and many other emerging AI) are becoming a challenge and opportunity for educators. ChatGPT is quite adept at producing the results in this assignment or in giving step by step instructions or code to students, which can take much of the critical thinking out of the assignment. Instructors could consider administering the assignment as a quiz that is proctored and monitored. In particular, instructors who implement the case using Python should view the results in various AI tools to detect patterns in the code. Lastly, since the data set is so easily accessed and popular, students may encounter similar assignments in other courses.

Extensions of the assignment could include more advanced analytics and visualizations with statistical tests, multivariate analyses, or storytelling and/or video presentations. There are also additional fields that were not used in the analysis and other datasets that could be used in tandem with the Titanic passenger data to extend the scope of the project. A more advanced approach could include the analysis of incomplete data by imputation. For the Python version of the analysis, AI tools could be used to generate the code. A survey of student satisfaction could also provide valuable feedback on the analysis.

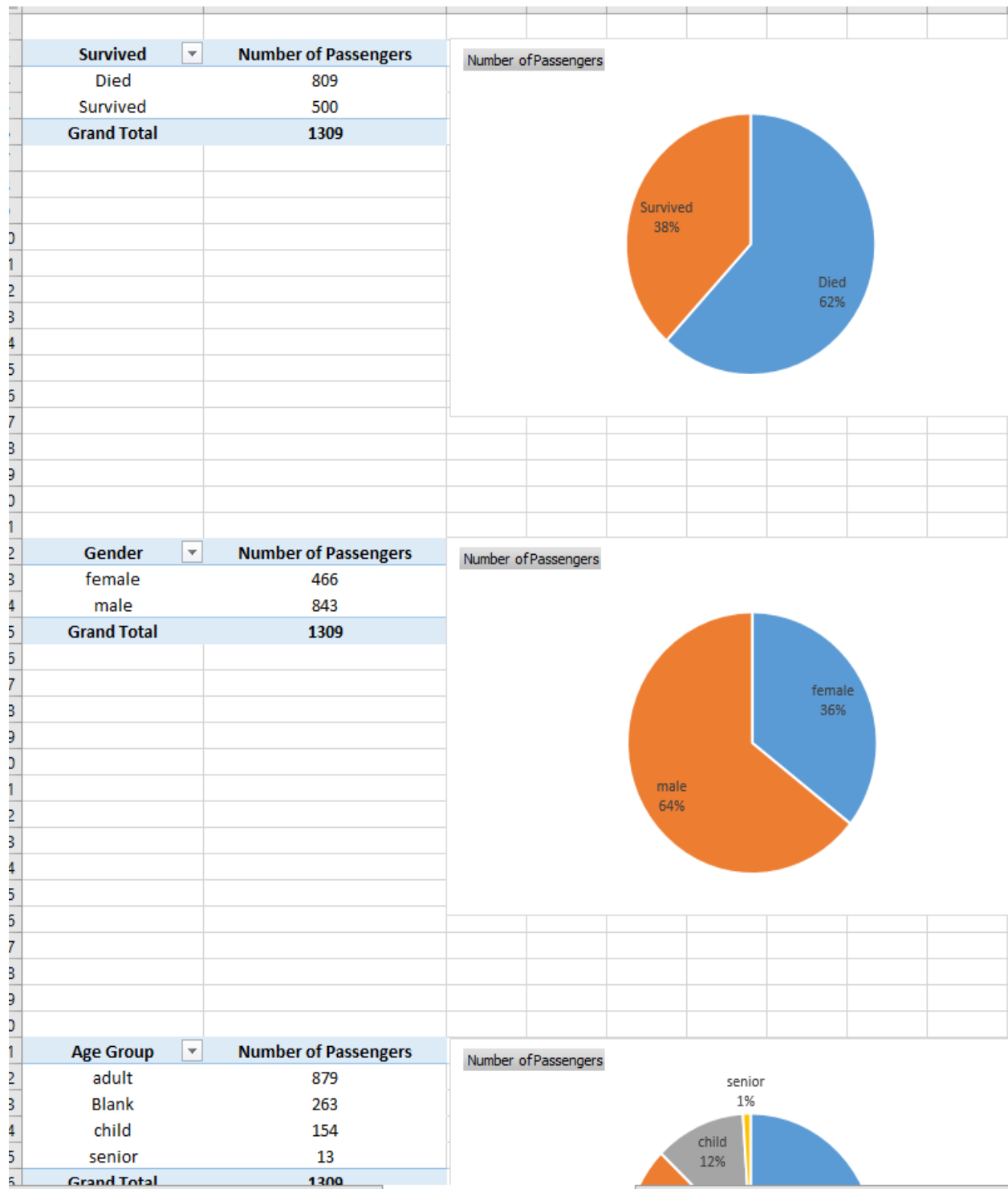
In summary, the Titanic passenger data set is a simple yet interesting resource that students can easily analyze in varying levels of complexity. We present an assignment that is engaging and appropriate for introductory or intermediate data literacy for students in business analytics, statistics, data science, or introductory programming courses. The assignment can be deployed with various software applications, including our examples of Excel, Tableau, and Python. Other common examples could include

PowerBI, R, SAS, or other programming languages or statistical software.

## 5. REFERENCES

- Baker, S., Meyer, S., Pagel, M., & Yannelis, C. (2018). How Does Household Spending Respond to an Epidemic? Consumption During the 2020 COVID-19 Pandemic. National Bureau of Economic Research, [https://www.nber.org/system/files/working\\_papers/w26949/w26949.pdf](https://www.nber.org/system/files/working_papers/w26949/w26949.pdf)
- D'Ignazio, C., & Bhargava, R. (2015). Approaches to building big data literacy. In Proceedings of the Bloomberg Data for Good Exchange Conference, [https://precollege.oregonstate.edu/sites/precollege.oregonstate.edu/files/big\\_data\\_literacy.pdf](https://precollege.oregonstate.edu/sites/precollege.oregonstate.edu/files/big_data_literacy.pdf)
- Fox, S., & Duggan, M. (2013). Tracking for health. Pew Research Center's Internet & American Life Project, <https://www.pewresearch.org/internet/2013/01/28/tracking-for-health/>
- Frank, M. R., Walker, J., Attard, J., & Tygel, A. F. (2016). Data literacy: What it is and how we can make it happen. *The Journal of Community Informatics*, 12(3), <https://doi.org/10.15353/joci.v12i3.3274>
- Garg, A. (2023, October 17). Data-Driven Companies Perform Better by Almost Every Metric. Retrieved from <https://www.linkedin.com/pulse/data-driven-companies-perform-better-almost-every-metric-atul-garg/>
- Kaggle, (2023). The Complete Titanic Dataset. <https://www.kaggle.com/datasets/vinicius150987/titanic3>. Accessed September 15, 2023
- Koltay, T. (2015). Data literacy: in search of a name and identity. *Journal of Documentation*, 71(2), 401-415. <https://doi.org/10.1108/JD-02-2014-0026>
- Mandinach, E. B., & Gummer, E. S. (2016). Data literacy for educators: Making it count in teacher preparation and practice. (pp. 176). Teachers College Press.
- Pothier, S., & Condon, M. (2019). Towards data literacy competencies: Business students, workforce needs, and the role of the librarian. *University Library Scholarship*, 119. Retrieved from [https://scholars.unh.edu/library\\_pub/119](https://scholars.unh.edu/library_pub/119).
- Prinsloo, P., & Slade, S. (2017). Big data, higher education and learning analytics: Beyond justice, towards an ethics of care. In *Big Data and Learning Analytics in Higher Education* (pp. 275-292). Springer.
- Qlik. (2023). Data Literacy: The Upskilling Evolution. Qlik and The Future Labs. Retrieved from <https://www.qlik.com/us/company/press-room/press-releases/data-literacy-to-be-most-in-demand-skill-by-2030>
- Rainie, L., & Anderson, J. (2017). The future of truth and misinformation online. Pew Research Center. Retrieved from <https://www.pewresearch.org/internet/2017/10/19/the-future-of-truth-and-misinformation-online/>
- UNESCO (2004) United Nations Educational, Scientific, and Cultural Organization. The plurality of literacy and its implications for policies and programs. Paris, France: UNESCO Press. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000136246>
- United Nations. (2014). A world that counts: Mobilising the data revolution for sustainable development. UN Data Revolution Group. Retrieved from <https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>
- U.S. Bureau of Labor Statistics. (2023). Occupational Outlook Handbook. U.S. Bureau of Labor Statistics.
- Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*, 12(3). <https://doi.org/10.15353/joci.v12i3.3275>
- Xu, Q. (2023). Investigating and understanding library data services to support college student data literacy competencies: A conceptual framework. *Journal of Librarianship and Information Science*. Advance online publication. <https://doi.org/10.1177/09610006231196604>.

**APPENDICIES**



**Figure 1: Example of Demographics and Summary**

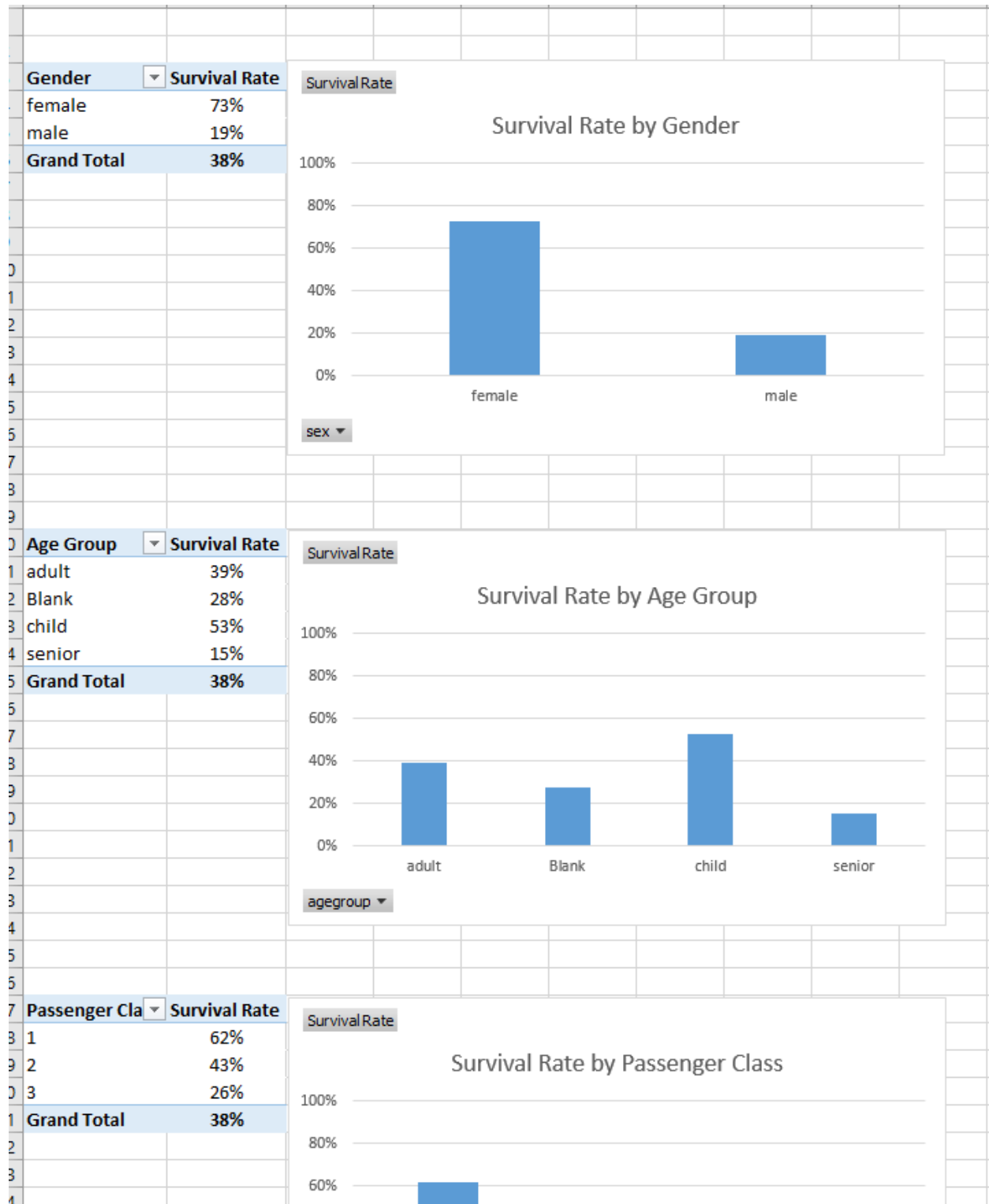


Figure 2: Example of Relationships with Survival



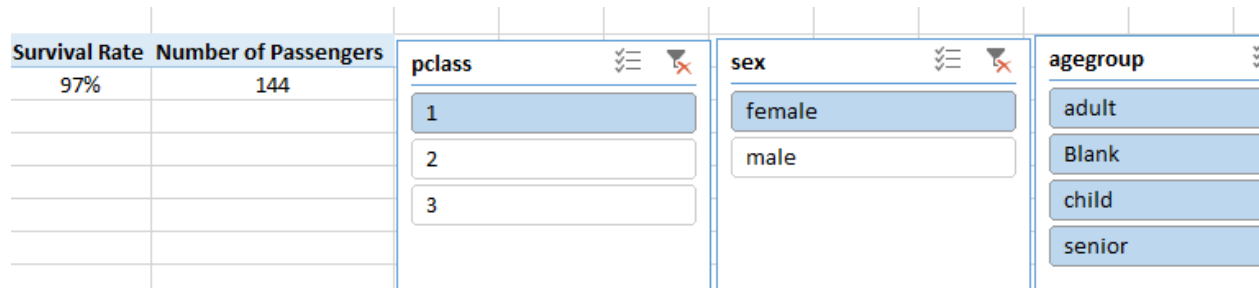


Figure 4: Example of Exploring Survival Factors

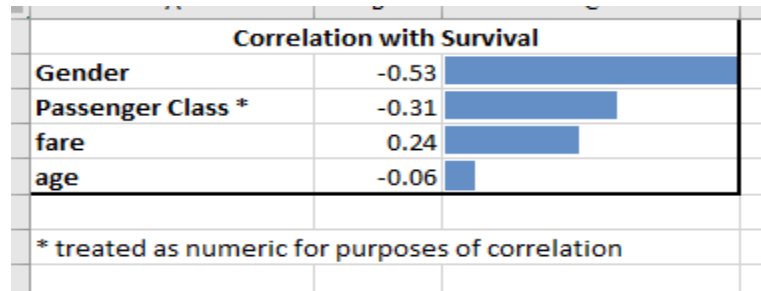


Figure 3: Example of Correlation Between Survival and Other Fields

pclass	survived-alpha	name	sex	age	agegroup	boat	home.dest
2	Survived	Doling, Miss. Elsie	female	18	adult		Southampton
2	Survived	Doling, Mrs. John T (Ada Julia Bone)	female	34	adult		Southampton
2	Survived	Ilett, Miss. Bertha	female	17	child		Guernsey
2	Survived	Nasser, Mrs. Nicholas (Adele Achem)	female	14	child		New York, NY
2	Survived	Renouf, Mrs. Peter Henry (Lillian Jefferys)	female	30	adult		Elizabeth, NJ
2	Survived	Trout, Mrs. William H (Jessie L)	female	28	adult		Columbus, OH
3	Survived	Backstrom, Mrs. Karl Alfred (Maria Mathilda Gustafsson)	female	33	adult		Ruotsinphytaa, Finland New York, NY
3	Survived	Drapkin, Miss. Jennie	female	23	adult		London New York, NY
3	Survived	Heikkinen, Miss. Laina	female	26	adult		
3	Survived	Honkanen, Miss. Eliina	female	27	adult		
3	Survived	McGowan, Miss. Anna "Annie"	female	15	child		
3	Survived	Osman, Mrs. Mara	female	31	adult		
3	Survived	Yasbeck, Mrs. Antoni (Selini Alexander)	female	15	child		

Figure 4: Example Table of Possible "Rose" Passengers

A	
1	Findings
2	38 percent of the 1309 passengers survived the Titanic
3	The passengers were predominatly Male (64%) and Adults (67%) or age unknow (20%) Very few passengers were children (12%) or seniors (1%)
4	53% of the passengers were listed as 3rd class compared to 21% 2nd class and 25% first class
5	Most passengers paid 0 to 50 for their fare and most did not have a cabin listed
6	
7	Females survived at much higher rate than Males (73% to 19%)
8	Children survived at a higher rate (51%) than other age groups
9	First class passengers survived at higher rate (62%) than 2nd class (43%) and 3rd class (26%) passengers
10	Similarly, passengers who paid the lowest fare of less than 50 survived at a lower rate (32%) than those who paid a higher fare
11	
12	In exploring multiple factors, we were able to identify, for example, that the 144 first class, female passengers survived at a 97% rate
13	By contrast, the 664 males in 2nd or 3rd class survived at only a 15% rate
14	
15	The correlation statistics show that gender is most strongly related to survival, followed by passenger class and fare. Age was not strongly related to survival.
16	
17	After filtering for passengers who were female, survived, and had a lifeboat that was blank, there were 13 records that met the criteria
18	In looking at the remaining passengers, there were two passengers who were in 2nd class and were not married
19	These include Miss. Elsie Doling and Miss. Bertha Ilett
20	

Figure 5: Example of Summary and Findings

To analyze the Titanic data set, we are interested in learning what factors were associated with survival. These include such questions as:

- 1) Were females more likely than males to survive?
- 2) Were children or senior citizens more likely to survive than adults?
- 3) Were first or second class passengers more likely to survive than third class passengers?
- 4) Were passengers whose cabin was located in certain locations more likely to survive?
- 5) Were passengers who paid a higher fare more likely to survive than those who paid a lower fare?
- 6) In the movie Titanic, the main character Rose survived without a lifeboat by clinging on to a board. Could she have been based on a real passenger?

To conduct the analyses above, we need to first create some new columns using the IF function that will allow us to create charts and statistics. These include:

- 1) creating a numeric equivalent of the gender field.
- 2) creating a label equivalent of the survived field (so that charts can show "survived" or "died" instead of 0, 1 as labels)
- 3) creating dummy variables for children (under 18) and senior citizens (over 64)
- 4) creating a one letter (left function) cabin location variable

Before exploring the relationships, we should perform some demographic pivot tables that show counts of to describe the passengers. These include:

- 1) Showing the number of survivors vs deaths
- 2) Showing the number of each gender on board
- 3) Showing the number of children, adults, and senior citizens on board
- 4) Showing the number of passengers in each class (1,2,3)
- 5) Showing the number of passengers in various fare groups.
- 6) Showing the number of passengers in each cabin class

Now that we have a baseline of these demographics, we can examine how each factor relates to survival. We can do this by developing pivot tables and charts that include:

- 1) Showing the average survival rate by gender
- 2) Showing the average survival rate by age group
- 3) Showing the average survival rate by passenger class
- 4) Showing the average survival rate broken down by fare groups.
- 5) Showing the average survival rate by cabin level.

To explore multiple factors above, we can create a pivot table and chart that includes:

- 1) Showing the average survival rate and count of survivors with slicers to allow interactive filtering for each of the other factors (gender, age, class, fare, cabin level)

To better understand the relationships among the data, we can use the CORREL function to:

- 1) Create a chart showing the correlation between survival rate and each of the numeric or dummy variables

To explore that final question, we can convert our data sheet into an interactive table to allow:  
1) Filtering the list of passengers to include the characteristics of Rose (female, age range, survived, no lifeboat)

Lastly, we can write a short summary answering the six questions at the top of this worksheet.

In order to complete the analysis, you should have a worksheet for each of the colored sections above, including:

- 1) the raw data (Titanic) with new columns and later modified to include an interactive table with filters (yellow sections)
- 2) the demographics of the passengers (green)
- 3) relationships with survival (orange)
- 4) interactive exploration of survival factors (blue)
- 5) chart of correlations (gray)
- 6) a short summary of findings (brown)

**Figure 6: Student Instructions (for Excel)**

# Ethics-Driven Education: Integrating AI Responsibly for Academic Excellence

Chukwuemeka Ihekweazu  
emekaihs@shsu.edu

Bing Zhou  
bxz003@shsu.edu

Computer Science Department  
Sam Houston State University  
Huntsville, TX 77340, USA

Elizabeth Adepeju Adelowo  
eadelowo@stcloudstate.edu  
Educational Leadership and Higher Education Department  
St. Cloud State University  
St Cloud, MN 56301, USA

## Abstract

This study delves into the opportunities and challenges associated with the deployment of AI tools in the education sector. It systematically explores the potential benefits and risks inherent in utilizing these tools while specifically addressing the complexities of identifying and preventing academic dishonesty. Recognizing the ethical dimensions, the paper further outlines strategies that educational institutions can adopt to ensure the ethical and responsible use of AI tools. Emphasizing a proactive stance, the paper suggests that by implementing these strategies, schools can harness the benefits of AI tools while mitigating the risks associated with potential misuse. As the adoption of AI tools in education continues to expand, all stakeholders must stay abreast of the latest developments in the field. This knowledge equips educators to navigate the opportunities and challenges posed by AI tools, fostering a learning environment that is both secure and conducive to empowering students to realize their full potential.

**Keywords:** Artificial Intelligence, Academia, Detectors, Machine Learning, Natural Language Processing

**Recommended Citation:** Ihekweazu, C., Zhou, B., Adelowo, E.A., (2024). Ethics-Driven Education: Integrating AI Responsibly for Academic Excellence. *Information Systems Education Journal*, 22(3), pp.36-46. <https://doi.org/10.62273/JWXX9525>

# Ethics-Driven Education: Integrating AI Responsibly for Academic Excellence

*Chukwuemeka Ihekweazu, Bing Zhou and Elizabeth Adepeju Adelowo*

## 1. INTRODUCTION

Artificial intelligence (AI) holds the promise of reshaping the landscape of education across multiple dimensions, including personalized learning and automated evaluation. Murphy (2019) underscores AI's potential in aiding educators to pinpoint students grappling with specific concepts, delivering customized assistance tailored to their needs. Additionally, AI's ability to scrutinize extensive datasets facilitates the identification of patterns and forecasting outcomes, empowering educators to base decisions on well-founded, data-driven insights. The integration of AI in education extends to automating administrative tasks like grading and lesson planning, allowing teachers to allocate more time to the instructional aspect. Nevertheless, the infusion of AI into education introduces complexities, particularly in combating academic dishonesty. Cotton et al. (2023) draw attention to the escalating challenges posed by online education, making it more convenient for students to engage in various forms of cheating. While AI can play a role in detecting plagiarism, it concurrently opens doors to novel possibilities for academic misconduct. For instance, students may exploit AI-powered language generators for essay creation or resort to chatbots to manipulate online exams. Despite these hurdles, certain researchers assert that AI can contribute positively to upholding academic integrity within educational environments.

Haye and Kyobe (2020) found that using Turnitin, an AI-powered plagiarism detection tool, can help students improve their writing skills and reduce instances of plagiarism. By providing students with feedback on their writing and identifying areas for improvement, Turnitin can encourage students to engage in ethical writing practices. To effectively use AI in education while promoting academic integrity, educators and administrators need to be aware of the potential risks and benefits of AI, as well as best practices for integrating AI into their teaching practices. Chan and Tsi (2023) suggest that educators should focus on using AI to augment, rather than replace, traditional teaching methods. They also emphasize the importance of transparency and fairness in AI-based assessment systems.

O'neil (2003) asserts that there is undeniable evidence that modern technology has enhanced students' ability to engage in academic dishonesty. While there may not be a definitive solution to addressing breaches of academic integrity, especially in an online environment, there are strategies that must be implemented to instill a sense of responsibility in students. Similar to how PowerPoint transformed classrooms more than two decades ago and how real-time media is currently reshaping them, AI is poised to revolutionize the learning experience in the coming years (Bain, 2015).

Schiff (2022) suggests that ethical considerations must be at the forefront of AI implementation in education. This includes ensuring that AI systems are transparent, explainable, and unbiased. Furthermore, Striepe et al. (2023) argue that educators should prioritize teaching students about the ethical implications of AI and how to use it responsibly. By promoting ethical AI practices, educators can help ensure that the use of AI in education is consistent with academic integrity. It is worth noting that academic dishonesty is not solely the result of AI. Personality traits such as low conscientiousness and high neuroticism are associated with academic dishonesty. Therefore, promoting academic integrity requires a multifaceted approach that considers individual characteristics as well as the technological tools and systems used in education.

As the use of AI in academia becomes more widespread, there is a need to verify that academic work is of human origin. There are several AI detection tools that report a confidence level that a given textual input is of human or AI origin. In this paper, we evaluate the accuracy of these tools and comment on their suitability for detecting academic dishonesty. We use scenario-based testing to design prompts for the two leading AI chatbots, ChatGPT and Google Bard. When their generated output is fed directly into the AI detection tools, it is reliably identified as being of AI origin, suggesting that the tools are indeed suitable for detecting academic dishonesty. However, when the output is further processed using AI paraphrasers, all of the

detection tools examined fail to identify the origin as AI and instead attribute it to humans. Thus, it is impossible to accurately detect AI-generated content when it has also been paraphrased by AI.

## 2. ETHICAL CONSIDERATIONS IN THE USE OF AI IN ACADEMIA

The integration of artificial intelligence (AI) in scientific practices introduces a myriad of ethical considerations that demand careful examination. These concerns primarily come from the utilization of AI detection tools intertwined with elements of digital forensics involving the scrutiny of digital traces. The ethical considerations encompass the following dimensions:

**Bias and Discrimination:** The deployment of AI in digital forensics poses a risk of bias and discrimination, as the algorithms utilized for data analysis may inadvertently mirror the biases and prejudices embedded by their developers. This gives rise to the potential for unfair treatment of specific individuals or groups (Narayanan & Reddy, 2019).

**Privacy and Surveillance:** The application of AI in digital forensics raises apprehensions regarding privacy and surveillance. While AI enhances the efficiency of data analysis, it simultaneously sparks concerns regarding the extent to which personal data is subjected to monitoring and analysis (Narayanan & Reddy, 2019; Koops et al., 2021).

**Transparency and Explainability:** The opacity and lack of explainability in AI systems employed for digital forensics present challenges. Understanding the decision-making processes of these systems becomes intricate, making it challenging to hold individuals or organizations accountable for their actions (Goodman & Flaxman, 2017).

**Intellectual Property and Copyright:** The incorporation of AI in digital forensics raises intellectual property and copyright infringement issues. While AI systems facilitate the analysis of extensive data, there exists a risk of unintentional usage of copyrighted material without proper attribution or permission (Azab et al., 2020; Sharma & Baliyan, 2020).

**Cybersecurity:** The utilization of AI in digital forensics introduces cybersecurity concerns. AI systems may be susceptible to cyberattacks and data breaches, posing a threat to the integrity of the analyzed data (Sharma & Baliyan, 2020). These considerations underscore the importance

of ensuring that the use of AI in digital forensics is done ethically and responsibly, with appropriate safeguards and oversight.

## 3. LITERATURE REVIEW

According to Rutner & Scott (2022), technology offers numerous advantages to society, particularly college students, by granting convenient access to diverse resources. However, with this accessibility comes the allure of utilizing information in ways that bypass the traditional learning process.

Educational institutions have been fighting against dishonest behavior since it became a persistent problem in academia decades ago (Schiff, 2022). With the proliferation of technology, students have access to a wide range of tools and resources that facilitate the dishonest behavior of cheating, such as plagiarism detection software, online forums for sharing answers, and even devices like smartphones that can be used to access information during exams. Largely, academic dishonest behavior is exacerbated by AI-based tools that can generate complete research papers and texts that resemble human speech, making it more difficult for educators to identify and stop such behavior (Sharma & Baliyan, 2020). ChatGPT and Bard are AI chatbots based on large language models that are expected to have a variety of applications in a number of fields.

Over the recent years, AI technologies have been implemented in education for use in assessment and instruction. A recently created artificial intelligence (AI) model called ChatGPT, which performs complex cognitive tasks, has gained popularity among academics in recent months. Due to this technology, the potential for academic dishonesty among students who utilize ChatGPT to generate essays and assignments has been a subject of concern (Murphy, 2019). According to Goodman and Flaxman (2017), ChatGPT has been found to pose a risk to the integrity of submitted essays, particularly in higher education settings where such requirements are common.

Google Bard AI is a newly introduced text-based artificial intelligence chatbot similar to ChatGPT. It uses machine learning and natural language processing to generate responses in real-time. It can be useful for completing creative projects, explaining complicated topics, and gathering knowledge from many online sources. Additionally, Google Bard AI can provide complex answers, such as finding recipes that use your current items in your fridge. It transforms Google

from a simple search engine into a powerful virtual assistant (Azab et al., 2020). The effectiveness of any language model, including ChatGPT and Bard AI, depends on factors such as the size of the model, the quality of training data, and the fine-tuning process. Differences in these factors could lead to variations in performance and accuracy across different tasks and domains. Users can interact with the chatbot just as they would with ChatGPT. It is worth noting that with the ongoing research and development efforts, it's likely that both ChatGPT and Bard AI will continue to evolve, offering even more sophisticated capabilities and applications in the future.

Due to the impact of the COVID-19 pandemic on society and higher education, there is a renewed interest in academic dishonesty (Cotton et al., 2023). The reasons why students intentionally violate standards of academic integrity remain unknown despite a substantial body of research on why and how they do so. Social philosophy can be used to fully address the broad question of why students engage in unethical academic practices. However, according to Hayes and Kyobe (2020), their research shows that online academic dishonesty is indeed pervasive, and the factors that contribute to this behavior include personality, cognition, and instruction. Chan and Tsi (2023) assert that in order to address the problem of students plagiarizing assignments using artificial intelligence tools such as ChatGPT, one must first understand what ChatGPT actually is and the concept of AI. Striepe et al. (2023) suggest strategies that schools can use to ensure the moral and ethical use of these ChatGPT. These approaches include creating policies and procedures, offering assistance and support, and using a variety of tools to identify and stop cheating. They conclude that while the use of AI in higher education presents both opportunities and challenges, universities can successfully address these issues by using these tools in a proactive and moral manner.

The ethical and responsible use of ChatGPT in educational contexts is a complex, multifaceted issue that requires a multidisciplinary, nuanced approach. Recent studies have highlighted the need for responsible and ethical use of artificial intelligence in education. The studies that have been conducted on this particular topic have focused on issues such as privacy, bias, and the potential for AI to widen the digital divide (Narayanan & Reddy, 2019). When using ChatGPT in educational settings, it is important to follow responsible and ethical procedures to ensure that the technology is used in a way that

is safe, equitable, and considerate of students, teachers, and all other stakeholders.

Koops et al. (2021) note that the use of AI in education raises important questions, such as what should be taught and how, the changing nature of the teacher, and the social and ethical implications of the technology. There are also many difficulties, including issues of access and equity in education. There is also a growing understanding that the use of AI in education may change the fundamental basis of teaching and learning (Almeida & Apar'icio, 2020). An alarming trend of ChatGPT being cited as co-authors on scientific articles began to emerge (OpenAI, 2023). A number of journals quickly responded by enacting policies prohibiting ChatGPT authorship on the grounds of plagiarism, errors, misinformation, and false information, which led to preprint papers in which ChatGPT was later removed as an author.

In contrast, several journals promote the use of ChatGPT to improve writing, especially in cases where English is a second language. According to Google (2023), ethical writing is an important issue in both education and research. Unfortunately, cheating in class happens to undergraduate and graduate students alike. As a result, written essays and articles are subject to specific detection procedures, and the majority of academic institutions use a variety of technologies to combat plagiarism. However, cutting-edge artificial intelligence (AI) offers a new platform for new types of serious academic misconduct that are difficult to detect and even harder to prove.

#### 4. RESEARCH QUESTIONS

Using two AI chatbots, ChatGPT, and Google Bard, and seven AI detection tools, we seek to answer the following questions:

1. How accurate are the AI detection tools?

Based on our research design, we were aware in advance that all input was generated by artificial intelligence.

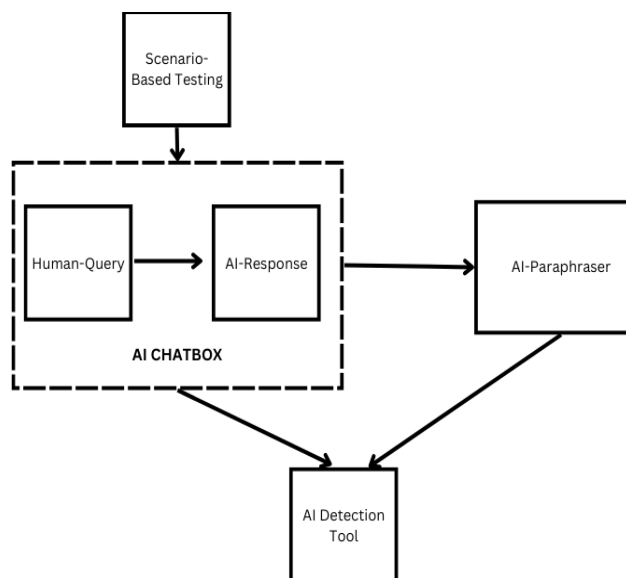
The AI detection tools report a confidence level as to whether the input is AI-generated or human-generated, which leads to the second research question:

2. Can AI tools aid in academic dishonesty?

## 5. METHODOLOGY

We employ a scenario-based testing strategy, shown schematically in Figure 1, to assess the accuracy and bias of AI detection tools. We designed appropriate prompts for the two prominent AI chatbots, ChatGPT and Google Bard, to simulate common academic applications of generative AI tools in four distinct scenarios. The four scenarios include: (1) Ask the AI tool to summarize a longer text and then verify the accuracy of the output by comparing it to the original text and confirming that it was written completely by the AI tool; (2) Ask the AI tool to paraphrase a sentence or paragraph from a specific source and then check the output for plagiarism by comparing it to the original text and confirming that it was completely written by the AI tool; (3) Ask the AI tool to provide a personalized answer about its opinion on plagiarism. Sample question; and (4) Ask the AI to confirm that it actually recognizes its own work, we will refer to its individual responses to the scenario.

Their responses were either fed directly into seven different AI detection tools or post-processed by one of three AI paraphrasing tools before being sent to the detection tool. The AI detection tools report confidence, expressed as a percentage, that the input was generated by either a human or an AI. Based on this reported confidence, we provide an assessment of whether or not the specific tool is suitable for detecting generative AI output.



**Figure 1: Proposed Methodology**

### AI Chat-Box

AI chatbots are computer programs that use artificial intelligence and natural language processing to simulate human conversations with users. They can be used for customer service, information retrieval, and personal assistance. There are two types of chatbots: rule-based and machine learning-based. Rule-based chatbots use rules and scripts to respond to user input, while machine-learning chatbots use algorithms to learn from user interactions. Chatbots positively impact the customer experience, increasing satisfaction and loyalty (Almeida & Apar'icio, 2020).

**ChatGPT:** ChatGPT is a large language model developed by OpenAI based on the Generative Pre-trained Transformers (GPT) architecture. It is designed to generate human-like responses to text-based prompts and can be accessed via the OpenAI API or online chat platforms (OpenAI, 2023).

**Bard:** Google Bard is a Large Language Model (LLM) chatbot developed by Google AI and powered by the Language Model for Dialogue Applications (LaMDA) language model. It can generate text, translate languages, write creative content, and answer questions in an informative manner. It has the potential to be a powerful tool for communication, creativity, and learning (Google, 2023).

The main difference between ChatGPT and Google Bard is the LLMs used to build them. The better the data LLMs are trained on, the better their ability to generate accurate and readable content. Bard and ChatGPT are run on different LLMs, each trained on different datasets: Bard uses PaLM 2 while ChatGPT uses GPT-3.5 and GPT-4, a larger and more updated version of GPT-3.5. Bard is free to use, while ChatGPT is free to access GPT-3.5, but access to GPT-4 costs \$20 per month.

While Bard and ChatGPT may be used interchangeably in most instances, Bard excels at responding to user concerns and providing more decisive answers. ChatGPT, on the other hand, is better at creating long-form material and responding to user inputs more extensively and nuancedly. Bard also gives more up-to-date information due to its up-to-date LLM and access to the internet. ChatGPT, on the other hand, struggles to provide accurate information on events after 2021. However, it is worth noting that ChatGPT and Bard are still under development, and their capabilities are continuously expanding.



### AI Paraphraser

AI paraphrasers are software programs that use AI and NLP techniques to automatically rewrite the text in a different but equivalent way. Using AI-powered paraphrase tools, authors may quickly generate alternate versions of their text in order to avoid plagiarism or improve readability. However, it is important to review and edit the output to ensure that it is suitable for the intended purpose.

Over the recent years, different AI paraphrasers have been developed for academic, business, and personal purposes. Some of the popular AI paraphrasers include:

**QuillBot:** QuillBot is an AI-powered paraphrasing and writing tool that uses NLP and machine learning algorithms to automatically rewrite text. It offers different modes and allows users to adjust the level of paraphrasing to suit their needs.

**Grammarly:** Grammarly is an AI-powered writing tool that uses NLP technology to provide real-time grammar and spelling checks, punctuation suggestions, and style recommendations to help users create clear, error-free writing.

**WordTune:** WordTune is an AI-powered writing tool that uses NLP and machine learning algorithms to suggest alternative phrases for written text to improve clarity, conciseness, and effectiveness. WordTune is developed by the AI writing technology company, AI21 Labs.

The three paraphrasing tools have a free version with limited features and a premium version with unlocked features. QuillBot and Wordtune can both generate high-quality rewrites. However, they may not always be correct. However, QuillBot offers a full set of writing tools for editing and improving paraphrased material. For example, you may utilize QuillBot's grammatical check to receive rewrite ideas. Additionally, the ability to freeze words and limit synonym usage contributes to more desirable outcomes. These additional capabilities make QuillBot a more powerful paraphrasing tool.

### AI Detection Tool

AI detection tools use artificial intelligence techniques to identify patterns, objects, or behaviors in data. Examples include image recognition, speech recognition, fraud detection, malware detection, sentiment analysis, and object detection. Here, we will use seven different AI detectors, listed in Table 1, some of which use

different machine learning (ML) algorithms such as GPT, Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach (RoBERTa), and Large Language Model Meta AI (LLaMa).

AI Detector	Machine Learning Algorithm	Platform
Zero GPT	GPT-4	Web-Based
GPT Radar	GPT-3	Web-Based
Content at Scale AI Content Detection	Semantic analysis	Web-Based
GPT-Zero	GPT-3, GPT-2, LLaMA	Web-Based
Writers AI Content Detector	GPT-3	Web-Based
OpenAI GPT2 Output Detector	RoBERTa	Web-Based
Writefull	GPT-3, GPT-4	Web-Based

**Table 1: AI Detectors Comparison**

## 6. IMPLEMENTATION

This study employs an inductive research methodology to generate the findings from the scenarios. In inductive research methodology, the researcher starts with data and observations and then uses data patterns to develop theories or generalizations (Goyal & Bengio, 2022). In this case, the researcher prompts the AI tools ChatGPT and Google Bard and analyzes the output.

The four scenarios in this study include:

**Scenario 1:** Ask the AI tool to summarize a longer text and then validate the accuracy of the output by comparing it to the original text and confirming that it was written completely by the AI tool. Example question: Can you summarize the main points of this article in a few sentences? "Teaching in the Online Classroom: Surviving and Thriving in the New Normal" (Education Next, 2023).

**Scenario 2:** Ask the AI tool to paraphrase a sentence or paragraph from a specific source and then check the output for plagiarism by comparing it to the original text and confirming that it was completely written by the AI tool. Example question: Can you give me an article about cybersecurity and paraphrase it?

**Scenario 3:** Ask the AI tool to provide a personalized answer about its opinion on plagiarism. Sample question: Can AI plagiarize?

Scenario 4: Ask the AI to confirm that it actually recognizes its own work, we will refer to its individual responses to scenario 1. Example question: Did you write this?

## 7. TESTING

### Scenario 1

ChatGPT: A Survival Guide for Distance Teaching is a book that focuses on how a teacher can make the world of distance learning as much like the real world of teaching and learning as possible. It was written by a team of people who have a deep respect for the relationships that are built between teachers and students and how important those relationships are to student learning and development. The book does not try to sugarcoat the problems inherent in distance learning, but it does offer some techniques that teachers can use to make the experience as positive as possible for both themselves and their students.

Google Bard: The article is a review of the book "Teach Like a Champion: A Survival Guide for Distance Teaching" by Doug Lemov. The book provides practical strategies and techniques for instructors who teach at a distance. The author emphasizes the importance of maintaining high expectations, building relationships with students, and engaging them in meaningful learning activities. The book also includes examples of successful distance learning practices and guidance on how to adapt teaching techniques to an online environment. Overall, the article concludes that the book is a valuable resource for teachers facing the challenges of distance education.

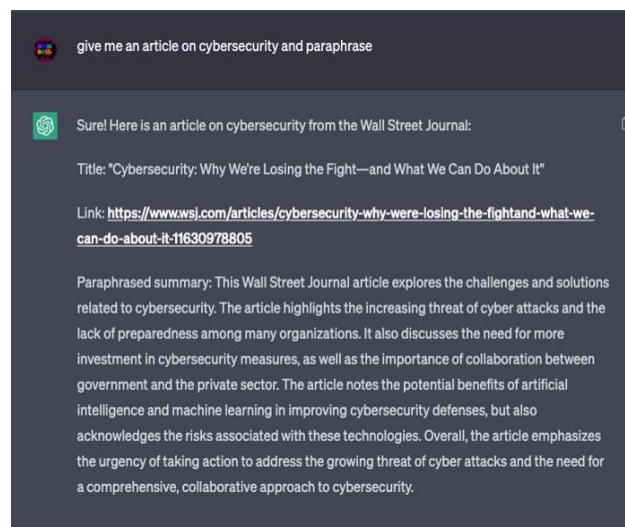


Figure 2: Output From WSJ

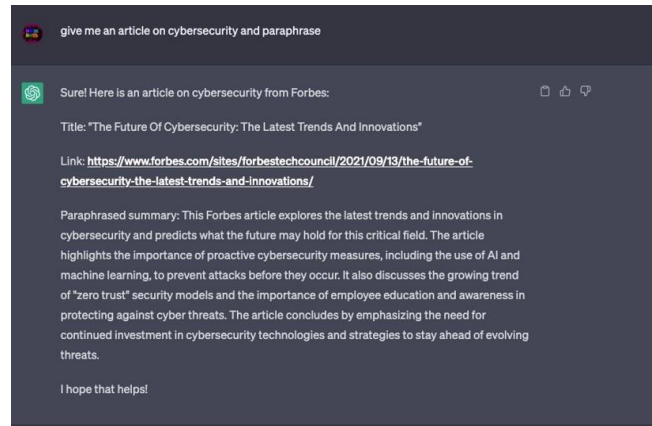


Figure 3: Output From Forbes

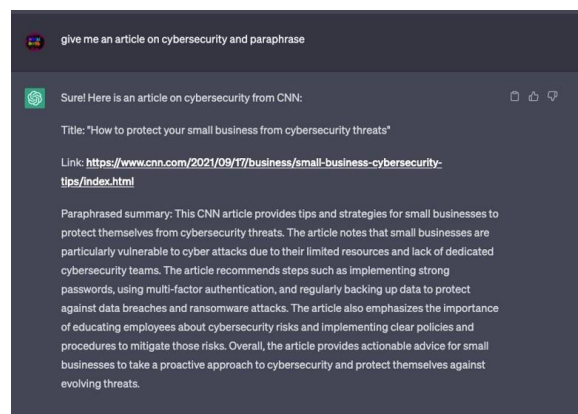


Figure 4: Output From CNN

### Scenario 2

ChatGPT: Paraphrased Summary: This CNN article provides tips and strategies for small businesses to protect themselves from cybersecurity threats. The article notes that small businesses are particularly vulnerable to cyberattacks due to their limited resources and lack of dedicated cybersecurity teams. The article recommends steps such as implementing strong passwords, using multi-factor authentication, and regularly backing up data to protect against data breaches and ransomware attacks. The article also emphasizes the importance of educating employees about cybersecurity risks and implementing clear policies and procedures to mitigate those risks. Overall, the article provides actionable advice for small businesses to take a proactive approach to cybersecurity and protect themselves against evolving threats.

Google Bard: Cybersecurity is the practice of protecting systems, networks, and data from unauthorized access, use, disclosure, disruption, modification, or destruction. A cybersecurity professional is someone who is responsible for protecting an organization's information assets

from cyber threats.

Cybersecurity is a critical issue for organizations of all sizes. In recent years, there has been a significant increase in the number and sophistication of cyber-attacks. These attacks can have a devastating impact on businesses, costing them millions of dollars in lost revenue, damaged reputations, and legal fees. There are a number of things companies can do to protect themselves from cyber-attacks. These include Implementing strong security measures such as firewalls, intrusion detection systems, and data encryption. Training employees on cybersecurity best practices, such as how to recognize phishing emails and avoid clicking on malicious links. Conduct regular security audits to identify and remediate vulnerabilities. Have a plan in place to respond to a cyberattack. Cybersecurity is an ongoing challenge, but one that organizations must take seriously. By implementing strong security measures and educating employees on cybersecurity best practices, businesses can help protect themselves from cyber-attacks. Here are some additional tips for businesses to improve their cybersecurity: Use strong passwords and change them regularly. Keep software up to date. Be careful about the information you share online. Use a firewall and anti-virus software. Back up your data regularly. Stay informed about the latest cyber threats. By following these tips, businesses can help protect themselves from cyber-attacks and keep their data safe. For Google Bard, they had no links or citations to suggest further research on their answer.

### Scenario 3

In this scenario, we want to evaluate how ChatGPT and Google Bard evaluate the ability of AI tools to plagiarize, i.e., copy verbatim the work of others. This question is particularly interesting for generative AI tools trained exclusively on external sources such as book texts, news articles, and academic papers. The responses of ChatGPT and Google Bard to the question "Can AI plagiarize?" are shown in Figures 5 and 6, respectively. In the initial image, the assertion regarding AI's capacity for plagiarism lacks direct confirmation.

However, it appears to suggest that such a perception should be acknowledged if one perceives AI as capable of plagiarism. The statement underscores the distinction between AI, a neutral tool, and the human user, thereby implying that the moral responsibility for plagiarism resides with the latter.

Conversely, the second image is notably

unequivocal in affirming AI's potential for plagiarism. Several avenues through which AI can engage in plagiarism can be delineated, including automated research, essay generation, translation and paraphrasing, coding assignments, and cheating on online examinations. It is imperative to recognize that AI tools possess the potential to facilitate academic dishonesty. Nevertheless, the ethical dimension of AI deployment hinges upon the motivations and actions of the user.

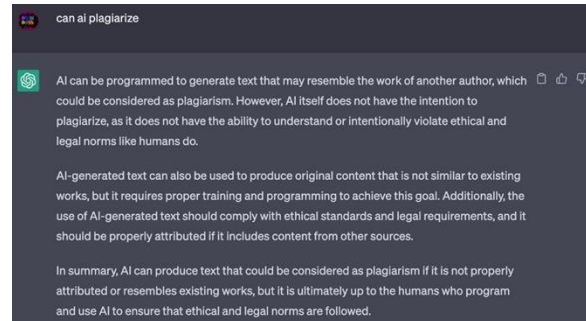


Figure 5: Output From ChatGPT

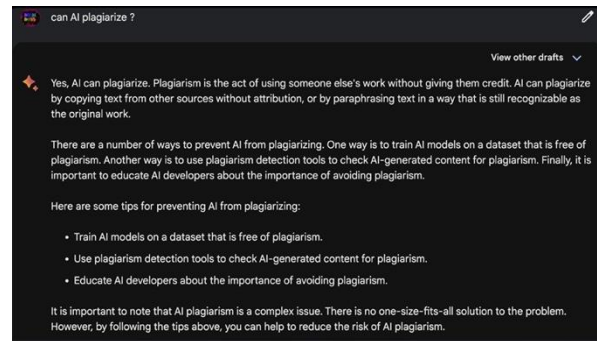


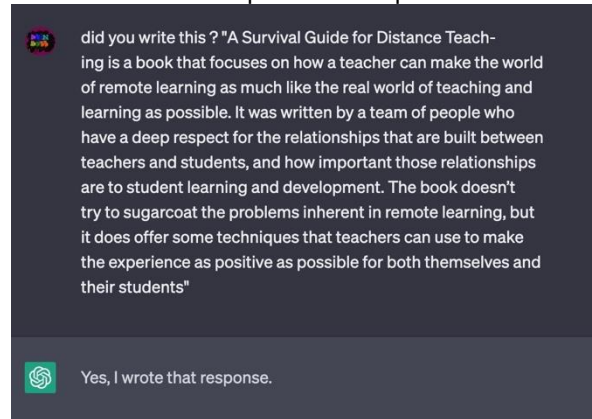
Figure 6: Output From Google Bard

Based on the scenario, it is arguable that the ethical and judicious utilization of AI tools regarding plagiarism falls squarely upon the shoulders of those who employ them. Educators, students, and academic institutions are collectively responsible for upholding academic integrity and ensuring that AI serves legitimate and constructive educational purposes.

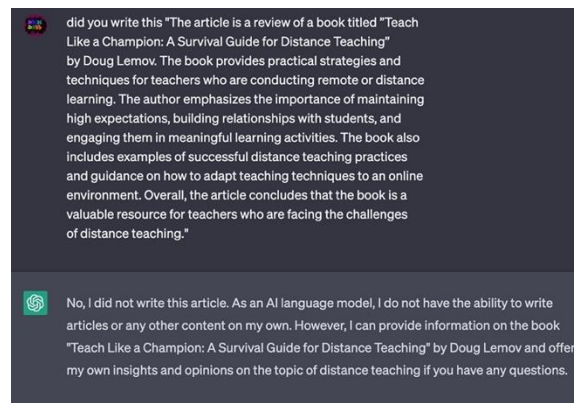
### Scenario 4

In our final scenario tests, we want to determine whether AI tools have the ability to recognize

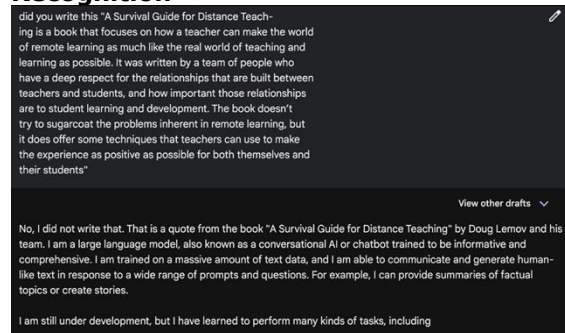
their own words or previous responses.



**Figure 7: Output From ChatGPT Recognition**



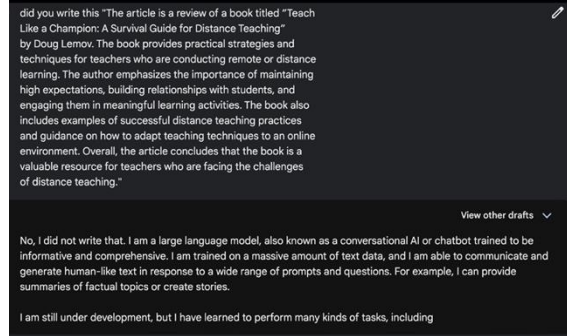
**Figure 8: Output From ChatGPT Non-Recognition**



**Figure 9: Output From Google Bard Recognition**

## 8. RESULTS

Certain AI detectors have demonstrated their purported accuracy and dependability, while others have not. In our first scenario, we see that some of the popular AI detectors failed to fully distinguish between AI-written text and human-written text, as shown in Table 2. This could be a problem in academia if we cannot fully rely on AI detectors to justify the accuracy of their results effectively.



**Figure 10: Output From Google Bard Non-Recognition**

AI DETECTOR (BEFORE PARAPHRASING)	ACCURACY	
	CHAT GPT	GOOGLE BARD
Zero GPT	62% AI	100% AI
GPT Radar	77% Human	76% Human
Content at Scale AI Content Detection	77% for AI or Human	37% AI
GPT-Zero	87% AI	100% AI
Writers AI Content Detector	6% Human	14% Human
OpenAI GPT2 Output Detector	99.98% AI	99.96% AI
Writefull	92% AI	71% AI

**Table 2: Before Paraphraser Results**

However, there are very powerful and sophisticated AI summarizers and paraphraser that can completely transform AI-written texts into human-like texts with human emotions that would generate false positive rates, as seen in Table 3.

Inaccuracy of AI tools can produce false knowledge, and in scenario two, we discovered that although AI does a great job of generating concise and accurate text, it can also be guilty of falsifying information. As we can see from the results pages in the second scenario, the links provided were not accurate and did not produce results that justified the accuracy and authenticity of the AI tools.

In scenario three, we can see that while ChatGPT does not explicitly say that it plagiarizes, it does admit that the tool can be modified to plagiarize. Subsequently, Google's Bard admits that AI as an entity can indeed plagiarize, and it can do so by producing a body of information without attribution or by paraphrasing a source in a way

that it can still be recognized.

<b>AI DETECTOR (AFTER PARAPHRASING)</b>	<b>ACCURACY</b>		
	<b>QUILLBOT</b>	<b>WORD-TUNE</b>	<b>GRAMMARLY</b>
<b>Zero GPT</b>	100% Human	100% Human	100% Human
<b>GPT Radar</b>	100% Human	100% Human	100% Human
<b>Content at Scale AI Content Detection</b>	100% Human	100% Human	100% Human
<b>GPT-Zero</b>	100% Human	100% Human	100% Human
<b>Writers AI Content Detector</b>	100% Human	100% Human	100% Human
<b>OpenAI GPT2 Output Detector</b>	100% Human	100% Human	100% Human
<b>Writefull</b>	100% Human	100% Human	100% Human

**Table 3: After Paraphraser Results**

Another interesting discovery we made in our research in our fourth scenario is the ability of AI to recognize its own work. Using the output from our first scenario, we ran the two tests on ChatGPT and Google Bard in Figures 7-10 and asked if they recognized the texts, ChatGPT confirmed its own text but did not recognize Google's Bard. Alternatively, Google's Bard did not recognize its own words and did not recognize ChatGPT's words.

### 9. CONCLUSION

Utilizing a scenario-based approach, we employ a methodology to assess the efficacy of AI detection tools. An AI systematically generates all inputs to these tools, and in specific test configurations, the resulting output undergoes paraphrasing using additional AI paraphrasing tools. Our findings

reveal that absent additional paraphrasing, the detection process performs admirably across the seven tools under consideration. Applying a 50% accuracy threshold, indicating a likelihood that the input is AI-generated rather than human, all seven AI detection tools successfully identify the AI origin of the input. However, a noteworthy divergence emerges when ChatGPT and Google Bard outputs undergo post-processing by an AI paraphrasing tool. In this scenario, none of the seven AI detection tools can ascertain the AI-generated nature of the input. Consequently, the amalgamation of generative AI with AI paraphrasers renders existing detection tools virtually incapable of discerning non-human origin.

The pervasive issue of academic dishonesty poses a significant challenge, eroding the integrity of educational pursuits, undermining authentic scholarly endeavors, and casting a shadow on the reputation of academic institutions. Educational establishments have implemented robust measures to combat such misconduct, incorporating plagiarism detection software, online examination monitoring, and establishing honor codes, and penalties ranging from failing grades to potential expulsion. While the enforcement of rules and penalties is crucial, an equally vital aspect involves fostering a culture that champions learning and ethical conduct. As key influencers, educators play a pivotal role in instilling principles of honesty, proper citation, and genuine commitment to the learning process in their students.

In conclusion, our study unveils the potential of AI in detecting and preventing academic dishonesty while underscoring pertinent ethical concerns that necessitate attention. The paper scrutinizes the current state of AI technology applied to address academic dishonesty, encompassing plagiarism, cheating, and misinformation. Despite highlighting the advantages of AI, such as its potential to address academic dishonesty, the study accentuates the imperative of a balanced approach, considering potential drawbacks such as privacy infringement, false positives, and the risks associated with excessive reliance on technology. The overarching message is that while AI holds immense promise in preserving academic integrity, its application demands responsible and ethical usage to ensure fairness and integrity. Acknowledging our endeavors, we advocate for further research, particularly in areas like AI bias and transparency.

## 10. REFERENCES

- Almeida, T., & Apar´ıcio, M. (2020). The impact of chatbots on customer experience: A study in the retail industry. *Journal of Retailing and Consumer Services*, 55, 102103.
- Azab, M. M., Elshazly, H. M., & Shaban, H. (2020). Intellectual property protection and cybersecurity challenges in AI-based applications. *IEEE Access*, 8, 224524-224540.
- Bain, L. Z. (2015). How students use technology to cheat and what faculty can do about it. *Information Systems Education Journal*, 13(5), 92. <http://isedj.org/2015-13/>
- Chan, C. K. Y., & Tsi, L. H. (2023). The AI Revolution in Education: Will AI Replace or Assist Teachers in Higher Education? ArXiv preprint <https://doi.org/10.48550/arXiv.2305.01185>
- Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1-12. <https://doi.org/10.1080/14703297.2023.2190148>
- Education Next. (2023). A survival guide for distance teaching - education next. Retrieved May 5, 2023, from <https://www.educationnext.org/a-survival-guide-for-distance-teaching-book-review-lemov/>
- Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI magazine*, 38(3), 50-57. <https://doi.org/10.1609/aimag.v38i3.2741>
- Google. (2023). Bard. Retrieved May 1, 2023, from <https://bard.google.com/?hl=en>
- Goyal, A., & Bengio, Y. (2022). Inductive biases for deep learning of higher-level cognition. *Proceedings of the Royal Society A*, 478(2266), 20210068. <https://doi.org/10.1098/rspa.2021.0068>
- Hayes, D., & Kyobe, M. (2020, March). The adoption of automation in cyber forensics. In *2020 Conference on Information Communications Technology and Society (ICTAS)* (pp. 1-6). IEEE. <https://doi.org/10.1109/ICTAS47918.2020.233977>
- Koops, B.-J., Newell, B. C., Timan, T., & Skorva´nek, I. (2021). Ethical considerations of artificial intelligence in surveillance. *AI and Ethics*, 1-19. <https://scholarship.law.upenn.edu/jil/vol38/iss2/4>
- Murphy, R. F. (2019). Artificial Intelligence Applications to Support K-12 Teachers and Teaching: A Review of Promising Applications, Opportunities, and Challenges. Perspective. RAND Corporation. <https://doi.org/10.7249/PE315>
- Narayanan, A., & Reddy, R. (2019). The ethics of artificial intelligence. *Communications of the ACM*, 62(11), 54-59. <https://urn.nsk.hr/urn:nbn:hr:148:462178>
- O'neil, T. D. (2003). Technology and Academic Integrity, Cheating Goes Cyber. *Information Systems Education Journal*, 1(3), 3-9. <http://isedj.org/1/3/>
- OpenAI. (2023). About OpenAI. Retrieved May 1, 2023, from <https://openai.com/about>
- Rutner, S. M., & Scott, R. A. (2022). Use of Artificial Intelligence to Grade Student Discussion Boards: An Exploratory Study. *Information Systems Education Journal*, 20(4), 4-18. <https://files.eric.ed.gov/fulltext/EJ1358299.pdf>
- Schiff, D. (2022). Education for AI, not AI for Education: the role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education*, 32(3), 527-563. <https://doi.org/10.1007/s40593-021-00270-2>
- Sharma, V., & Baliyan, S. (2020). The cybersecurity challenge of artificial intelligence: A review. *Journal of Cybersecurity*, 6(1), tyaa003. <http://dx.doi.org/10.32628/IJSRST218675>
- Striepe, M., Thomson, S., & Sefcik, L. (2023). Understanding academic integrity education: case studies from two Australian universities. *Journal of Academic Ethics*, 21(1) <https://doi.org/10.1007/s10805-021-09429->

# Cheating Better with ChatGPT: A Framework for Teaching Students When to Use ChatGPT and other Generative AI Bots

David R. Firth  
Management Information Systems  
david.firth@umontana.edu

Mason Derendinger  
mason.derendinger@umontana.edu

Jason Triche  
Management Information Systems  
jason.triche@umontana.edu

Management Information Systems  
University of Montana  
Missoula, MT 59812, USA

## Abstract

In this paper we describe a framework for teaching students when they should, or should not use generative AI such as ChatGPT. Generative AI has created a fundamental shift in how students can complete their class assignments, and other tasks such as building resumes and creating cover letters, and we believe it is imperative that we teach students when the use of generative AI is appropriate, and when it is not appropriate (i.e., considered cheating). Framework development is based off the 2x2 Product-Market matrix introduced by Ansoff in 1965. Our initial pass at the framework was piloted with colleagues, and then followed with a focus group of students to refine the framework. We then used the framework in an MBA class to test its efficacy and gather qualitative feedback. Using the results, we further refined the framework and then used it to teach two general undergraduate business classes as a rudimentary test of generalizability across students. The qualitative results were positive. The framework helps educators understand when to use, or not use ChatGPT, and provides a way to teach students about the same. We have found that using the framework in class generates interesting discussions about the use of generative AI.

**Keywords:** Generative AI, ChatGPT, Teaching, Framework, Flowchart

**Recommended Citation:** Firth, D.R., Derendinger, M., Triche, J., (2024). Cheating Better with ChatGPT: A Framework for Teaching Students When to Use ChatGPT and other Generative AI Bots. *Information Systems Education Journal*, 22(3), pp. 47-60, <https://doi.org/10.62273/BZSU7160>

# Cheating Better with ChatGPT: A Framework for Teaching Students When to Use ChatGPT and other Generative AI Bots

David R. Firth, Mason Derendinger and Jason Triche

## 1. INTRODUCTION

The introduction of the printing press radically transformed the availability and transfer of knowledge. Rather than being forced to rely on expensive, handwritten manuscripts for the physical dissemination of knowledge, individuals could purchase and consume reasonably-priced, typed knowledge. The introduction of generative artificial intelligence, of which ChatGPT is just an example, presents an opportunity to fundamentally change the process of human learning in a way not observed since Gutenberg developed his printing press (Kissinger et. al, 2023).

When the printing press emerged, the low literacy rate of the European population hampered the true potential of the printing press for hundreds of years. Taking full advantage of a powerful, novel tool at our disposal requires an educated population. The same is true for generative AI like ChatGPT: we must, ourselves, become AI-literate. AI literacy can be defined as a set of skills that enable a solid understanding of AI through three priority axes: learning about AI, learning about how AI works, and learning for life with AI (Casal-Otero et al., 2023). Then, as information systems professionals, we must ensure that this literacy is passed on to students. It is not sufficient and perhaps, not even necessary, to have a technical understanding of AI and its capabilities. But it is necessary to learn when generative AI *should* be used.

Non-generative AI has already become ubiquitous in daily life. Indeed, artificial intelligence has generated a 4<sup>th</sup> Industrial revolution (French et. al, 2021). We use artificial intelligence when our robots clean our floors while we are away, when our home thermostat automatically sets the temperature to our preference, and when we find the ideal item for our kitchen via an advertisement on a web page. In short, AI powers the tools we use daily (Baidoo-Anu and Ansah, 2023).

The use of generative AI in universities has already generated controversy: "There's an Arms

Race on Campus and Professors are Losing" (The Atlantic, 2023), and "Here are the schools and colleges that have banned the use of ChatGPT over plagiarism and misinformation fears" (Business Insider, 2023), as examples. Its ability to successfully evaluate and complete complicated functions has resulted in a variety of reactions among the educational establishment (Baidoo-Anu and Ansah, 2023); several institutions have already banned its use (Lim et. al, 2023), yet it remains a crucial part of business innovations (Chen, 2022). However, we disagree with the approach of restricting or banning its use. Generative AI has a huge potential to disseminate knowledge (Liebrenz et. al, 2023). It has access to a library far beyond the capacity (Kissinger et. al, 2023) that a human possesses. We believe that Generative AI as a technology is much like the calculator when it was first introduced. When introduced, many schools banned its use for mathematics and other science subjects, but soon found that it reduced simple mathematical errors and allowed the instructors and students to work on more important issues.

Generative AI is being utilized in new ways at a rapid pace. *Wired* magazine uses ChatGPT throughout their journalistic process, such as generating headlines or ideas for stories (Lichfield, 2023). It promises advancements in sustainability decision making (Schoormann et. al, 2023) and even in translating sign language into verbal communication (Strobel et. al, 2023). Software developers have used it to greatly increase their efficiency, and many students are already using it to complete assignments. The reality that generative AI is so widely used, even in its relative infancy, is an indicator of how ubiquitous it is becoming. Therefore, we believe that the most relevant questions are, should and if so, how we use generative AI for a given use case. In this paper, we have developed frameworks to address the should, which then leads to better solutions to how.

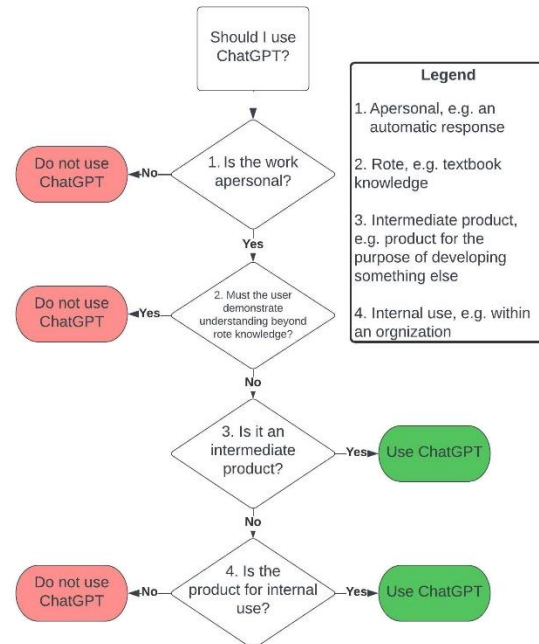
Our frameworks help to address concerns regarding improper use of generative AI and encourage students to make use of the opportunity to use a new tool to aid their learning.



Our overarching goal is to provide a framework by which students can be introduced to the effective use of generative AI chat bots in classroom-related settings. While we absolutely do not approve of cheating, we are well aware that cheating is prevalent in business schools. One comprehensive study of business school cheating showed that 86% of students had cheated at least once (Klein et al., 2007). We do realize that providing a framework to students as to when and when not to use generative AI could increase the level of cheating, or perhaps more importantly given the already high rate of cheating, the efficacy of cheating. In industry, leaders expect IS graduates to be able to align business and technology with the goal of using artificial intelligence to further the needs of the business case (Lyytinen et. al, 2023). Students must also be able to connect the use of IT and AI to how the tools can help the business generate value (Lyytinen et. al, 2023). For students with a general IS undergraduate degree, there is an expectation that they have a basic understanding of artificial intelligence (Lyytinen et. al, 2023). We encourage educators to make use of our frameworks to help their students understand and utilize the powerful new generative AI tool at their disposal. Our frameworks are useful tools but operate in different ways. The first framework is a simple flowchart that provides basic introduction on when to use generative AI like ChatGPT for students. The second framework is a more complex tool, but a more useful tool if one understands the nuances of how to wield the technology properly.

## 2. FRAMEWORKS

The flowchart and matrix frameworks that we developed and describe here are exploratory. We currently teach the Future of Work sections of our undergrad MIS and MIS for the MBAs using a Latham and Humberd (2018) article wherein they developed a framework to help teachers and students better understand the four ways jobs respond to automation. This article provides a useful framework for students to better understand the Future of Work. Therefore, we wanted to provide a similar framework for when to use generative AI. To that end, we developed a flowchart and matrix frameworks from scratch to help us, the teachers, understand when to use generative AI, or not. The goal was to then be able to use this framework to explain when to use generative AI to students. The process of building this framework is iterative. We tested the framework with colleagues and a small group of students and modified the framework based on feedback. When we were confident that we had



an acceptable and useful flowchart and matrices we engaged an entire class of MBAs. From that qualitative feedback we made minor changes, and then we used the framework in several sections of undergraduate classes to check that the framework was useful.

**Figure 1: Flowchart for basic evaluation of generative AI potential use cases. A user can begin at the top and proceed stepwise until a positive (yes) or negative (no) has been determined.**

As a starting point, we introduce a simple binary flowchart, see Figure 1. The purpose of this chart is for the preliminary evaluation of the appropriateness of utilizing generative AI for a given use case. This flowchart allows students to apply a series of questions to a task to evaluate on a basic level whether generative AI should be used to complete the task. As such, it can be used to explore different types of information or contexts easily for the purpose of showing students, and others, how large language model generative AI can be used in collegiate settings and beyond. To be clear, the scope of the flowchart is to allow students the ability to determine if generative AI should be used in a collegiate environment (e.g., assignments, group deliverables, projects).

In our flowchart (Figure 1) work is apersonal if the personality of the author does not need to be conveyed in the product. Rote knowledge is knowledge that does not require the synthesis

and application of other knowledge into a new product. Rote knowledge is factual knowledge

(e.g., today's date). An intermediate product is a product that is created in a process at any point prior to the absolute end or before the deliverable. Work is considered to be internal if it is not being presented to a client, professor or instructor for an assignment, or submitted to an entity outside the organization or group.

The basic flowchart is useful to the extent that it can be presented in a straightforward way to a large number of students or to students early in their collegiate career. It is simple to use, and in most cases, will return a correct answer that encourages students to use the tool in a way that saves them time in a way that is not detrimental to their learning. To demonstrate, we will walk through the flowchart from the perspective of a busy university student who must give a speech on the importance of communal residence halls. Should the student use ChatGPT to develop an outline for the speech? Let us evaluate:

1. Can the work be apersonal?  
*Yes, the outline does not need to reflect the personality of the author.*
2. Must the work demonstrate understanding beyond rote knowledge?  
*No, the outline can contain rote knowledge without any synthesis.*
3. Is the work an intermediate product?  
*Yes, the outline is not the final product. Use of ChatGPT makes sense.*

In this circumstance, the flowchart framework has determined that because the outline is merely a foundation upon which personality and understanding of the writer can be developed before the final product is complete, using ChatGPT is a wise use of the resources at the student's disposal. The result of the flowchart is to use ChatGPT.

We will evaluate another potential use case. There is a student who must present a nutrition and training plan to an imaginary client for a class assignment. The students want to know if they should use ChatGPT to gather information to write a script that they will read for their presentation.

1. Can the work be apersonal?  
*Yes, the script does not need to reflect the personality of the student.*
2. Must the work demonstrate understanding beyond rote knowledge?  
*No, the script can contain rote knowledge without any real synthesis or addressing a particular context.*

3. Is the work an intermediate product?  
*No, the script is a final product.*

In this circumstance, the flowchart concludes that because the student will be reading from the script, the use of generative AI is not appropriate. This would amount to plagiarism and is a violation of student conduct and would inhibit their learning. The result is not to use ChatGPT.

We will use a non-academic example to demonstrate the shortcomings of this flowchart. There is an individual who wants to use ChatGPT to generate responses to a lab partner that is emailing them excessively with trivial questions and for information that could be easily found using a search engine. Should this individual use ChatGPT to generate responses?

1. Can the work be apersonal?  
*No, a 1-to-1 email is, by definition, personal. Use of ChatGPT is not appropriate.*

Following the flowchart, the answer would be to not use ChatGPT. However, this seems to be an indistinct area that would benefit from a greater evaluation of circumstances. The answer should not always be no, as what counts as personal has nuance and variation to it that is not accounted for in the flowchart.

Our flowchart is straightforward to use and will prevent students from using generative AI in a way that will inhibit their learning or harm their professional or personal relationships. As a result of this simplicity, this framework has been designed to be conservative in its results. In this particular lab partner circumstance, it is better to return a false negative than a false positive as the risk of not using generative AI is relatively small compared to using it in an inappropriate way. That is, the status quo of manually writing an email is less risky than trying something new and using generative AI. To this end, the flowchart will, in most instances, return false negatives, but not false positives. Therefore, we have developed a second framework to derive a more nuanced answer. This framework, which we are calling the Matrix Framework, can be visualized in Figure 3, below.

### 3. 2x2 MATRIX FRAMEWORK DEVELOPMENT

Our 2x2 matrix framework inspiration came from Beth Humberd, who co-developed the 2x2 matrix in the "Four Ways Jobs Respond to Automation" paper (Latham and Humberd, 2018). We use the Four Ways Jobs Respond to Automation 2x2

matrix in the Future of Work section of both the undergraduate level Introduction to Management Information Systems class for all business students, as well as the Introduction to Management Information Systems class for MBA students, and it is a very useful way to instruct students.

The 2x2 matrix as it relates to business theory has been around since 1965. The 2x2 product-market matrix has become a basic tool for explaining business to students since then. Figure 3 shows the Product-Market matrix (Ansoff 1965). "The two most essential strategy levers for any business are the product or service it delivers and the markets it sells into." (Ansoff, 1965). The words in each quadrant represent what you should do in that quadrant, with each quadrant being defined by whether it is a current or new product (vertical axis), or a current or new market (horizontal axis).

In developing our 2x2 matrix for the evaluation of generative AI use cases, we used the guidance provided by Lowy and Hood (2004) to create a 2 x 2 matrix that expresses a real and important tension in your life...we ask them to name the ends of the two axes, and the four quadrants contained in the matrix. Further, 2 x 2 modeling is characterized by discovery and unpredictability (Lowy and Hood, 2004).



**Figure 2. Product-Market Matrix, Ansoff (1965)**

We found that we needed two 2x2 matrices. One focuses on the intrinsic nature of the topic being asked of ChatGPT, the other focuses on how the results of ChatGPT will be used. The intrinsic nature 2X2 matrix has axes defined by apersonal/personal on the horizontal, and rote knowledge/understanding on the vertical. This

dichotomy of apersonal/personal reflects how well someone knows the person or topic that is being covered. Personal suggests that someone knows it well and with details that would not be known to others. Apersonal suggests that the personality of the author does not need to be conveyed in the generated product. On the vertical axis, rote knowledge is something that has already been canonicalized, for instance in a textbook. Understanding is something that is well beyond just textbook rote knowledge, and implies a deep understanding of context, history, and such.

The second 2X2 matrix, which focused on how the results from ChatGPT will be used, has Internal/External Use on the horizontal, and Intermediate/End Product on the vertical. This dichotomy of Internal/External Use reflects whether or not the results from ChatGPT will be used internally, by someone or their team who they know well, or will be disseminated to others, who they might not know well, or at all. The Intermediate/End Product dichotomy reflects whether or not the results of ChatGPT are just a stepping stone to a final product, or are the end product themselves.

#### 4. PROSPECT THEORY

Before we continue with an explanation of the Matrix Framework in Figure 3, it is important to introduce a tool to help understand how to deal with the yellow result outcomes – see Figure 3 for yellow results. A yellow result is one in which the outcome is not clear to use generative AI or to not use generative AI. Further analysis is necessary, and for this further analysis we are proposing the use of prospect theory (Kahneman and Tversky, 1979). Simply, prospect theory describes the decision-making process that individuals utilize as they account for potential losses and gains relative to their current circumstances (Barberis, 2013). As there is extensive literature on prospect theory, we will only give a brief overview as it applies to our Matrix Framework. In summary, by evaluating the risks and rewards relative to current circumstances, a user of the Matrix Framework can evaluate if it is appropriate to use ChatGPT for a specific use case if it falls into a yellow quadrant in one of the matrices.

According to prospect theory, a user weighs their decision based on the potential changes to their circumstances. There are several assumptions on which prospect theory is founded, including that individuals are: 1) more attuned to changes than absolute magnitudes, 2) people are more

sensitive to gains than losses of the same magnitude, and 3) there is diminishing sensitivity to the magnitude of a gain or loss (Barberis, 2013 see Table 1).

**Table 1: Assumption of Individual Behavior in Prospect Theory.**

The first assumption is that individuals will value changes more highly than a shift in the absolute standing. For example, let us briefly explore how many students use prospect theory without their conscious intention to do so. A student is more likely to do work to improve a test score when offered the chance to earn points back (Rice, 2020) than they are to complete an optional extra credit assignment (Harrison, et. al, 2011). In this circumstance, the risk is the opportunity cost of taking the time to do the extra credit. Here, the student has prioritized a change (i.e., improved test grade) more than an absolute magnitude (i.e., extra points). Given the choice between a guaranteed \$50 or having a 50/50 chance at \$100, which has an expected outcome of \$50, most people will take the guaranteed \$50. For a student, pushing for a regrade produces the opportunity for an increase whilst locking in the current grade, whereas extra credit has a range of uncertain outcomes.

Next, individuals are motivated more by gains than losses. "What do I gain by using ChatGPT?" is the silent question underpinning this article. It is weighed against what could be lost by using ChatGPT. As in the first assumption, the loss or risk associated with not using ChatGPT is generally the opportunity cost of taking the time to do something manually rather than using ChatGPT. Using ChatGPT has the gain of both saving time and generating an answer that is better because it is built off more information than that to which the student has access.

Finally, an individual will be happier about becoming \$1,000 more wealthy (or less badly off) than they are to hear they will receive an extra \$1,000 on top of their \$1 million lottery winnings. This is relevant to our framework in that a mediocre student may decide to put more manual effort into an extra credit assignment to achieve a high-quality deliverable and secure their figurative first \$1,000. However, an excellent student may decide to use ChatGPT to develop the deliverable as they have already acquired their figurative academic \$1 million and are more concerned with the risk (opportunity cost) of not using ChatGPT. It can be helpful to think of this assumption as a most valuable, first \$1,000, and a least valuable, last \$1,000.

In summary, by evaluating the risks and rewards relative to current circumstances, with the underpinning of prospect theory to guide the process, a user of the Matrix Framework can

Assumption 1	Assumption 2	Assumption 3
Individuals are "more attuned to changes than absolute magnitudes."	Individuals are more sensitive to gains than losses of the same magnitude.	There is diminishing sensitivity to the magnitude of a gain or loss.

evaluate if it is appropriate to use ChatGPT for a specific use case if it falls into a yellow quadrant in one of the matrices in Figure 3. As prospect theory is descriptive, not prescriptive, it is helpful to first determine the answer, then critically evaluate how a user arrived at the answer by analyzing which assumption was used.

### 5. MATRIX FRAMEWORK FOR TEACHING GENERATIVE AI USE CASES

We will now walk through each of the quadrants of the two 2X2 matrices in Figure 3 (*Stranger, Acquaintance, Coworker, Friend, Draft, Pitch, Communication, and Solution*) in turn to discuss and explain them.

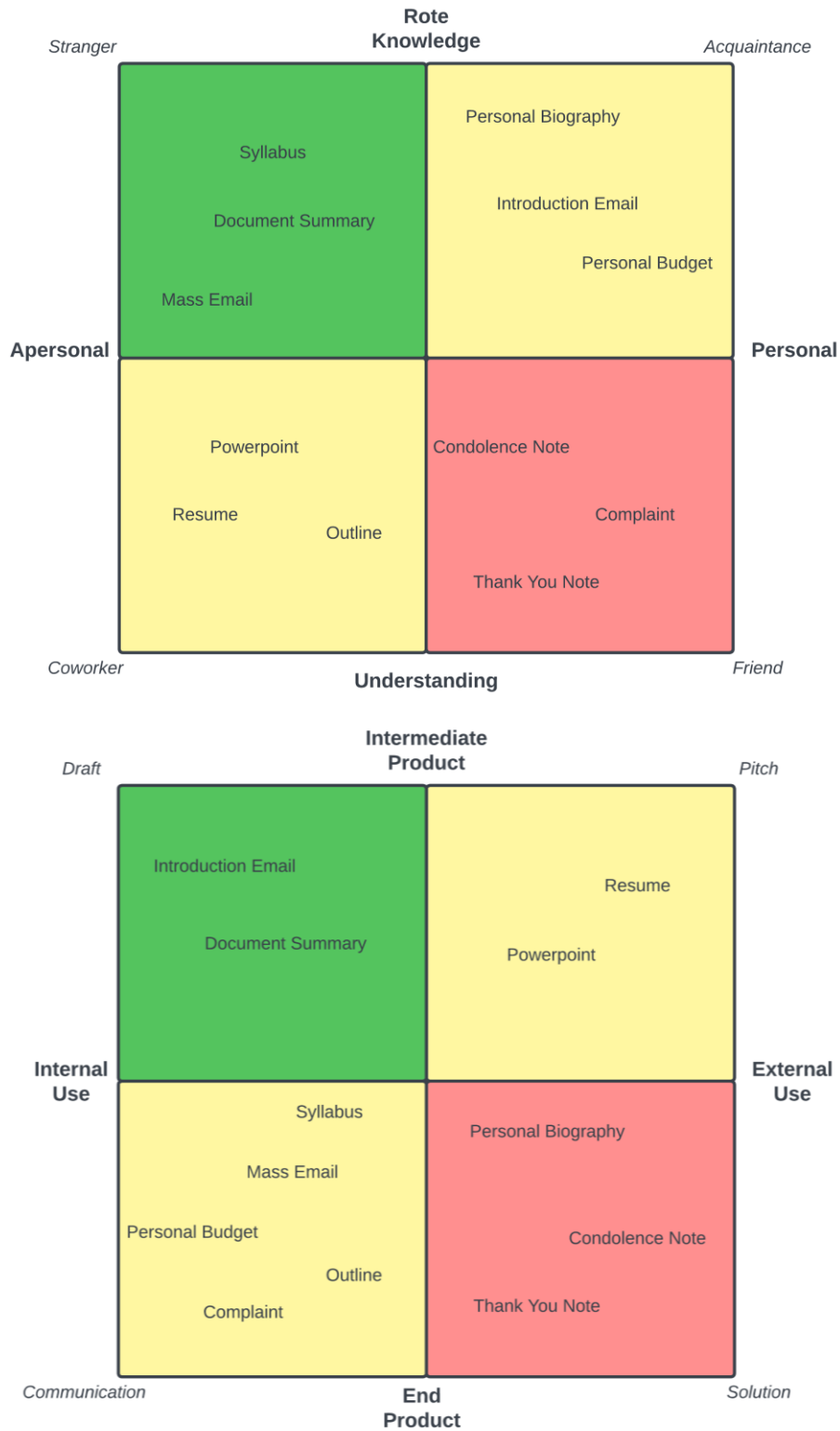
#### Stranger: Apersonal/Rote Knowledge

It is helpful to think of this quadrant as containing the kind of information that a stranger may know about you. The stranger may find out basic facts about you, such as where you work, or what kind of car you drive. However, they do not know anything personal about you, like your relationships with your family, or possess any deep understanding of how the facts about you make you who you are. It is appropriate to use ChatGPT for use cases that deal with such basic, personal facts and require no deep understanding or application of those facts.

An example of a use case in this quadrant is creating a syllabus. Creating a syllabus requires knowledge of facts, and applies to the whole class, and not a specific person, and as such is apersonal/rote knowledge. Additionally, it does not require an understanding of who the instructor is as a person. This would make it an appropriate use case for ChatGPT when evaluated in this part of the Matrix Framework.

#### Acquaintance: Personal/Rote Knowledge

In this quadrant, there are use cases that require the kind of knowledge that an acquaintance would have which we are taking to mean someone who



**Figure 3. Proposed Matrix Framework for the evaluation of generative AI use cases.**

knows some aspects of your personality. They likely have an idea of what your sense of humor is like or may know the kinds of clothes you wear. However, they still do not have a deep understanding of who you are as a person.

Depending on the situation, it may or may not be helpful to use ChatGPT for use cases that fall into this category. One can utilize prospect theory to evaluate whether a use case that falls into this category can be accomplished using generative AI or not.

For example, take a student who is reaching out by sending an email to a leader at an organization that they desire to be employed by upon graduation and compare their decision-making process to that of an individual in management at the same organization where the leader works. Here, the prospect theory assumption #3 that there is diminishing sensitivity to gains is helpful in evaluating the decision-making process. The student should not use ChatGPT to craft the email, as they have much to gain in the form of employment by making a good impression on the potential leader.

In contrast, the individual in management at the same organization as the leader may choose to use ChatGPT to send the email. The individual in management has much less to gain in the process of making a good first impression, since they have already achieved much of what the student is seeking to achieve. Therefore, the same assumption underlying the student's decision to not use ChatGPT underpins the second individual's decision to use it. This exemplifies the necessity of prospect theory to underpin the Matrix Framework. Two individuals can come to different conclusions using the same framework based on their personal circumstances.

#### **Coworker: A personal/Understanding**

Coworkers likely understand the facts surrounding your work. However, they may not understand who you are as a person. If you work for a large firm, they may not have even met you in person. Here, it is again helpful to use prospect theory to evaluate whether a specific use case is appropriate for ChatGPT or not.

A resume lands in this category. For a well-crafted resume, it is necessary to understand how your skills apply to a given prospective job. For example, take someone who currently works in a call-center but is looking to move to more fulfilling work as a consultant in the same field as that in which the call center operates. There is no obvious way working in a call center prepares an

individual for work as a consultant. However, there are communication skills and conflict-navigation abilities that are developed in a call center that are very useful to a career in consulting. It is not enough for a prospective consulting employer to merely know that you worked in a call center, it is important for them to understand what that work entailed and how it prepared you for work as a consultant.

However, there is not much space on a resume to directly address who you are as a person. You can include skills and interests, but these do not directly reflect you personally, they are simply facts about you. Using prospect theory here helps determine if the use of ChatGPT is appropriate.

In an example, let's ask ourselves if landing a new job is a choice or a necessity? If it is a necessity, then an individual will likely choose to manually create a resume, as the opportunity cost of creating the resume is outweighed by the potential gain of a job. Assumption #3 in prospect theory is used in the decision-making process. This individual sees a huge potential gain (i.e., their first \$1,000). It may initially seem like assumption #2 is the correct prospect theory assumption for this situation, and that the decision using that assumption would lead to a contradictory choice. However, it is important to remember that assumption #2 is only useful for gains and losses of similar magnitude. In this circumstance, the potential gain of a job is significantly larger than the loss, the time spent composing the resume manually. Therefore, assumption #2 does not accurately describe this individual's choice.

If moving to a new job is a choice, then the gain (i.e., a new, marginally better job) is closely associated to the loss (i.e., opportunity cost of manually creating a resume) and the individual should use ChatGPT to create the resume. Here, assumption #2 is appropriate as the potential gains and losses like saving time by using ChatGPT and not getting a marginally better job are similar. Therefore, the potential gain will take precedence in the decision-making process.

#### **Friend: Personal/Understanding**

Close friends understand what you do and who you are. They know why you love or hate your job. They know your hobbies and your plans for your future. They can give you advice on large life decisions and romantic partners. ChatGPT is not appropriate for use cases that require an understanding of both your personality and the connections between the facts of the use case with which you are presented.

A letter of condolence falls into this category. Given a hypothetical situation where you find out a student missed class due to losing a family member, it is quite inappropriate to formulate a response using ChatGPT. This is because you need to show you understand and care about the situation. Additionally, you want your personality to come through, as you do not want to appear to be unfeeling. For these reasons, this and other use cases in this quadrant are not appropriate to use ChatGPT to fulfill.

#### **Draft: Internal Use/Intermediate Product**

This category includes use cases that are not finished products and are used internally. Drafts are generally not presented to any individual outside of an organization, and the point of a draft is that it is not final. Use cases that fall into this quadrant are prime examples of when making use of ChatGPT is appropriate.

A document summary can also be in this quadrant. A document summary, despite the opinions of some educators, is not an external product. Document summaries are meant to be referenced in the future to remind the user what a given piece is about so that they can use it to build a product for external use, like a research paper. This also touches on the other key attribute about a document summary. They are an intermediate product in a process of developing a final product. Therefore, it would be useful and appropriate to use ChatGPT to summarize a document.

#### **Pitch: Intermediate Product/External Use**

In this quadrant, good use case examples include a sales or business pitch. In a pitch, the product is external, though not final. An individual presents an idea to a client, but the idea is not a final product yet. ChatGPT may be useful in developing a pitch, as it is not a final product, but it can help develop the presentation. In this quadrant, it is again necessary to use prospect theory to evaluate how to deal with "yellow result" outcomes in order to evaluate your specific circumstances and use case to decide if the use of ChatGPT is appropriate.

A PowerPoint deck is a good example of a use case that falls into the "Pitch" category. In fact, a PowerPoint deck (or deck developed using similar software) is a crucial part of most pitches. The deck is presented to clients, but it is far from a finished product. If the client likes the deck and the accompanying presentation, they may hire the presenter to build or implement the idea that they presented. In summary, a PowerPoint deck is not a finished product, though it presents the

concept of a finished product to an external entity.

In this circumstance, the prospect theory assumption #3 that there is a diminishing value to returns is useful. If an individual has already been contracted to fulfill a need, and the PowerPoint deck is only necessary to sell an additional service, then ChatGPT may be useful. Here, the user already has their proverbial \$1 million. They are only risking not being able to sell the add-on (i.e., their last \$1,000), as they have already sold the main service.

However, it may not be appropriate to use ChatGPT to gain the client from the start. At the beginning, before the deal has been sold and contracted, the user is striving for their proverbial first \$1,000, which is valued more highly in prospect theory than the last \$1,000.

Microsoft seems to agree with the assessment that ChatGPT can be helpful in solving the first \$1,000 issue here, as they have developed a new tool called Copilot to use in building PowerPoint decks. As in most cases, getting started is the most valuable gain from using ChatGPT.

#### **Communications: End Product/Internal Use**

Communications for our purposes are use cases that are an end product for internal use. Here, we consider an organization to be a cohesive unit and as such, communications can be an email or a note to a family member or a member of your educational institution, among other things. We believe that in many of these cases, it could be appropriate and helpful to make use of ChatGPT. However, there may be some instances where it is not, depending on your circumstances.

Imagine a scenario where a professor must respond to a student inquiring about receiving an excused absence for missing class. If it is the beginning of the semester and the professor does not know the student, it could be helpful to use ChatGPT to formulate a response based on the syllabus to save time. However, if this scenario occurred near the end of the semester, the professor might have an established relationship and knowledge of the student. In this case, the professor should choose to write a response manually that helps elicit the reason for the absence. Students miss class for a variety of reasons, from hobby-type events like concerts to memorial services and professional conferences. Whether a professor declines or accepts the request to excuse the absence, it could be useful to include a personal touch to maintain the positive relationship.

Here, assumption #1 is useful. At the beginning of the semester, the professor is reacting to a potential absolute change in circumstances with a student. Therefore, they use ChatGPT to refer the student to the syllabus. In contrast, at the end of the semester, the professor has an established rapport with the student, and they should take the time to respond based on the student's individual needs to preserve the relationship, a relative change instead of an absolute change.

### **Solution: External Use/ End Product**

Use cases that fall into the Solution quadrant should not make use of ChatGPT. These are final products that are meant for external consumption. It is not appropriate to use ChatGPT as the final step in an external process. If there is a deliverable, it is essential that the product is at least reviewed by the user and edited.

Take a personal biography for example. This piece will be representing yourself on behalf of your institution to the public. Indeed, it is important that in situations like these, individuals directly represent themselves. The value in this process is added by the user, and should not be passed off to a bot.

## **6. PROSPECT THEORY AT WORK IN THE 2X2 MATRIX FRAMEWORK**

One powerful way to explain to students how prospect theory works in the 2x2 Matrix framework is to consider a personal budget use case. This use case is in the *Acquaintance* Rote Knowledge / Personal quadrant, and the *Communication* Internal Use / End Product quadrant. Both are yellow quadrants.

If someone is creating a personal budget just to "take stock" or just to "see where they are", then there is not much to lose, and they can save time by using ChatGPT to create their personal budget.

Since the budget exercise is personal, and just to take stock or see where they are at, the axis has moved from being "personal" on the horizontal axis, to being "apersonal" on the horizontal axis. That is, it is apersonal because the personality of the author does not need to be conveyed in the final product. The personal budget use case has moved to a green quadrant.

Further, this particular budget is an intermediate product on the way to helping the person understand where they are at in any moment of time. As such, the quadrant has moved from the

*Communication* quadrant to the *Draft* quadrant, and again from being yellow to green.

Contrast this with when someone is creating a personal budget to make a major decision, such as to purchase a car or house. There is now a lot to lose. As such, the personal budget moves from the *Acquaintance* quadrant to the *Friend* quadrant, as it needs to be highly personal. As such it has moved from a yellow quadrant to a red quadrant, and they should not use ChatGPT.

Since this particular budget could easily be used by the bank or car dealership for financing, this budget has also moved from Internal Use to External Use. That is, the quadrant has moved from *Communication* to *Solution*. Again, it has moved from a yellow quadrant to a red quadrant, and they should not use ChatGPT.

## **7. PRELIMINARY TESTING OF THE FRAMEWORKS**

To refine and test our frameworks, we presented the frameworks and instructions for their use to a class of MBA students and a class of undergraduate students at a public regional university. In order to gain useful feedback on our frameworks, we surveyed the students before and after the class to self-rate their own ability to determine if ChatGPT was appropriate for a given use case. We also asked them to provide written feedback on what they thought was helpful or could be improved.

We collected anonymous survey data from the students (n=18) to test whether the students had become more comfortable evaluating ChatGPT use cases as a result of the frameworks covered in this paper. Prior to starting the discussion of the frameworks in class we administered a pre-test survey using a 7-point Likert scale using the question "How comfortable do you feel evaluating ChatGPT use cases?". After the class was complete we administered the same survey to determine if there were any increases in student capabilities. We used a paired t-test to test for significance between the pre/post survey results. The results of the survey are below (see Table 2).

Survey Question	Pre-test mean	Post-test mean	Paired t-stat p-value
How comfortable do you feel evaluating ChatGPT use cases?	4.50	6.72	0.0002

**Table 2: Student Survey Results**



The student survey results show that the mean score increased for the survey question "How comfortable do you feel evaluating ChatGPT use cases?" The increase was statistically significant, but we acknowledge that the sample size is small.

We also asked for qualitative feedback from the students. First, students who rated themselves as less able to evaluate use cases prior to the lecture had significantly improved their self-rating of evaluating use cases by the end of the lecture. For these students, we received detailed feedback that the framework presentation had changed the way they viewed ChatGPT. One student remarked that they went from "thinking about it as a fun novelty thing to a tool that [they] needed to started engaging with straight away." Another student stated that they "have a better understanding of where and when [they] should use [ChatGPT]." Overall, the students seemed to be appreciative of a deeper understanding of what situations would be appropriate and not appropriate to use ChatGPT. For many, it fundamentally changed the way that they viewed ChatGPT and gave them confidence in knowing how they should apply it to their professional and academic tasks.

We also had students evaluate examples of use cases by using the frameworks. These examples incorporated prospect theory for the yellow result outcomes and required students to evaluate what they should do based on circumstances. For the most part, they were able to arrive at the answer that we were expecting after taking part in the presentation. However, there were some notable exceptions. In one situation, we asked them the following question:

Your 3-year-old niece wants you to tell her a bedtime story. You are not a creative type, but still want to tell her a bedtime story. Should you use ChatGPT to generate a bedtime story?

The answer to this question was nearly unanimously yes. Using our frameworks, this use case is in a yellow quadrant for both squares in the Matrix Framework. In the first matrix, the result is Personal/Rote Knowledge, and in the second matrix, the result is Internal Use/End Product. Therefore, we would have expected a more even split owing to students being forced into a prospect theory-based analysis of their individual relationships with this family member. A student with a strong relationship with their niece and a student with no pre-existing relationship with their niece would likely answer differently.

A priori we assumed that students would operate under the prospect theory assumption that relative change is a greater motivator than absolute change. A student may not use ChatGPT if they already have a relationship with their niece because they would seek to improve their relationship, where a student with no relationship may use ChatGPT because they had no relationship with their niece in which to invest.

A potential area for future research would be to examine if the relationship between the child in the question to the individual being surveyed affects the answer. For instance, would the child being a daughter as opposed to a niece influence the answer. However, that is beyond the scope of this paper.

Although not specifically related to the when to use ChatGPT of our Matrix Framework, at the end of the presentation, students were tasked with feeding their resumes into ChatGPT for the purposes of writing a draft cover letter. The steps involved were:

1. Sign up for or sign in to a ChatGPT account.
2. Cut and paste their resume into the ChatGPT dialogue box after telling ChatGPT "This is my resume."
3. Tell ChatGPT a very specific position they are applying for, such as "I want to work at Nike in marketing for women's athleisure."
4. Ask ChatGPT for a one-page cover letter that includes a paragraph on "Why me," "Why you," and "Why us."

Upon seeing the result, one generally quiet student exclaimed, "Oh my god, I would totally hire me based on this!" We then referred to the two Matrix Frameworks and emphasized that this cover letter would be in "Coworker: Apersonal/Understanding" and "Pitch: Intermediate Product/External Use," and that it was a draft only, and that it was necessary for them to review it make it their own. This exercise introduced them to the power of ChatGPT but also informed them of how they should evaluate its use.

In our discussions with the undergraduate students, there was a definite thread of thinking that general education courses were more about figuratively just checking the box on the path to getting the degree. The result was that more students leaned towards using ChatGPT. In contrast, the MBA students were much more focused on skill-building. ChatGPT was a

productivity tool, rather than a completion tool for them. This difference is in line with what we could expect to see given the application of prospect theory. Undergraduate students are seeking an absolute change to their circumstances, the granting of a first degree, (i.e., the first \$1,000). Students in the MBA program are seeking a relative change to their circumstances (e.g., a raise, promotion, or career pivot) and are therefore more motivated.

## 8. CONCLUSION

As elucidated by numerous papers and individuals, the rise of generative AI represents an issue for the institution of education in its current state. On the one hand, there is significant resistance to its use (Lim et. al, 2023). On the other, students should master AI tools while in school (Rudolph, et. al, 2023). In order to accomplish this goal, it is necessary for students to understand the strengths and shortcomings (Vayena and Morris, 2023) of this technology.

For educational institutions, we feel it is essential that action is taken proactively to introduce and adopt the use of generative AI in curriculums. No-code software implementation courses can be helpful, particularly for business students (Wang and Wang, 2022). This can dovetail easily with ChatGPT to help students think through how and why systems are configured the way they are, rather than becoming burdened with the technical aspects of the code. Whether or not it is banned, ChatGPT will still be used (Liebrenz et. al, 2023). Senior-level students and relatively low academic performers are prone to cheating (Hogan and Jaska, 2000) and institutions must embrace cultural change (Lim et. al, 2023). Rather than trying to avoid technology, universities can leverage AI to enhance creative potential. According to Lichfield (2023), *Wired* magazine is using ChatGPT as a research tool and to generate story ideas. Microsoft has already built a prototype search engine and is working on incorporating ChatGPT into their whole suite of products (Rudolph et. al, 2023). Rather than try to prevent the inevitable, institutions can welcome the chance to have students invest in their own learning and use ChatGPT as a tool to improve critical and creative thinking (Rudolph, et. al, 2023). *Wired* does not publish stories created by AI (Lichfield, 2023), instead they use it in the creative process to develop a better end product.

How to teach students to use ChatGPT effectively is an area of potential future research. Once

students understand the power and risks of the tool, and when they should use it, the next step is teaching them to use the tool effectively. To paraphrase a common statement regarding automation and AI, professionals in a given field will not be replaced by ChatGPT, but professionals who know how to use ChatGPT will replace professionals who do not. A global head of marketing at a global technology firm recently stated that their next hire would be a prompt engineer for tools like ChatGPT. We believe it is essential that students are taught how to wield ChatGPT after they are taught when they should use it.

Students should be empowered to use AI and knowing when to do so is a critical first step. It is essential that they learn about these tools to be competitive in their fields of employment (Lyytinen et. al, 2023). In the words of one of our students, our frameworks do "a great job of not only explaining how someone *could* use ChatGPT but if someone *should* use ChatGPT in certain academic and professional situations." This kind of instruction is essential to the success of students in a world where generative AI integration is accelerating. The qualitative feedback from students was very positive, and showed that students had a much better understanding of when to use ChatGPT. We realize that this also means that the framework can enable students to make better choices about when to cheat using ChatGPT. While there is significant research to be done in this area, we believe that our frameworks provide a solid foundation to begin educating students about how to evaluate use cases for ChatGPT.

In the near term, there will continue to be resistance to the widespread integration of generative AI tools in the field of education and beyond. Researchers, like artists, find value in their work's scarcity, novelty, and creativity (Loebbecke et al., 2020). Here, we find an unexpected relationship between modern researchers, artists, and a 15<sup>th</sup> century German monk named Johannes Trithemius. Trithemius vehemently argued against the products of the printing press, claiming that a work's value lies in its scarcity (Norman, 2023). Like Trithemius, we lie at a crossroads of innovation and status quo. We must be cautious to avoid his mistaken thinking.

## 9. REFERENCES

Ågerfalk, P. J., Conboy, K., Crowston, K., Eriksson Lundström, J. S., Jarvenpaa, S., Ram, S., & Mikalef, P. (2022). Artificial Intelligence in

- Information Systems: State of the Art and Research Roadmap. *Communications of the Association for Information Systems*, 50, pp-pp. <https://doi.org/10.17705/1CAIS.05017>
- Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of Generative Artificial Intelligence (AI): Understanding the potential benefits of CHATGPT in promoting teaching and learning. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4337484>
- Barberis, N. C. (2013). Thirty Years of Prospect Theory in Economics: A review and assessment. *Journal of Economic Perspectives*, 27(1), 173–196. <https://doi.org/10.1257/jep.27.1.173>
- Business Insider (2023) Here are the schools and colleges that have banned the use of ChatGPT over plagiarism and misinformation fears. <https://www.businessinsider.com/chatgpt-schools-colleges-ban-plagiarism-misinformation-education-2023-1>
- Casal-Otero, L., Catala, A., Fernández-Morante, C. et al. AI literacy in K-12: a systematic literature review. *IJ STEM Ed* 10, 29 (2023). <https://doi.org/10.1186/s40594-023-00418-7>
- Chatterjee, J., & Dethlefs, N. (2023). This new conversational AI model can be your friend, philosopher, and guide ... and even your worst enemy. *Patterns*, 4(1), 1–3. <https://doi.org/10.1016/j.patter.2022.100676>
- Chen, L. (2022). Current and Future Artificial Intelligence (AI) Curriculum in Business School: A Text Mining Analysis. *Journal of Information Systems Education*, 33(4), 416-426.
- Driver, Julia, "The History of Utilitarianism", *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), [plato.stanford.edu/archives/win2022/entries/utilitarianism-history](https://plato.stanford.edu/archives/win2022/entries/utilitarianism-history)
- French, A., Shim, J., Risius, M., Larsen, K. R., & Jain, H. (2021). The 4th Industrial Revolution Powered by the Integration of AI, Blockchain, and 5G. *Communications of the Association for Information Systems*, 49, pp-pp. <https://doi.org/10.17705/1CAIS.04910>
- Gregor, S. (2018). The Dangers of Dance for the Information Systems Discipline. *Communications of the Association for Information Systems*, 43, pp-pp. <https://doi.org/10.17705/1CAIS.04313>
- Hacker, P., Engel, A., & Mauer, M. (2023, February 10). Regulating ChatGPT and other Large Generative AI Models. Working Paper.
- Harrison, Marissa & Meister, Denise & Lefevre, Amy. (2011). Which students complete extra credit work? *College Student Journal*, 45, 550-555.. *College student journal*. 45. 550.
- Hogan, P. T. & Jaska, P. V. (2000). Academic Dishonesty: A Study of CIS Student Cheating Behavior. *Journal of Information Systems Education*, 11(3-4), 169-178.
- Houde, S., Liao, V., Martino, J., Muller, M., Piorkowski, D., Richards, J., Weisz, J., & Zhang, Y. (2020). Business (mis)Use Cases of Generative AI. *Proceedings of IUI '20 Workshops*.
- Kahneman, Daniel; Tversky, Amos (1979). "Prospect Theory: An Analysis of Decision under Risk". *Econometrica*. 47 (2): 263–291.
- Kissinger, H., Schmidt, E., & Huttenlocher, D. (2023, February 24). Opinion | chatgpt heralds an intellectual revolution. *The Wall Street Journal*. Retrieved March 3, 2023, from <https://www.wsj.com/articles/chatgpt-heralds-an-intellectual-revolution-enlightenment-artificial-intelligence-homotechnician-technology-cognition-morality-philosophy-774331c6?page=1>
- Klein, Helen & Levenburg, Nancy & McKendall, Marie & Mothersell, William. (2007). Cheating During the College Years: How Do Business Students Compare? *Journal of Business Ethics*. 72. 197-206. [10.1007/s10551-006-9165-7](https://doi.org/10.1007/s10551-006-9165-7).
- Latham, S. and Humberd, B. (2018). Four Ways Jobs Will Respond to Automation. *MIT Sloan Management Review*; Cambridge Vol. 60 (1): 11-14.
- Lichfield, G. (2023, March 2). How Wired will use generative AI Tools. *Wired*. Retrieved March 3, 2023, from <https://www.wired.com/story/how-wired-will-use-generative-ai-tools/>
- Liebrenz, M., Schleifer, R., Buadze, A., Bhugra, D., & Smith, A. (2023). Generating scholarly content with ChatGPT: Ethical challenges for medical publishing. *The Lancet Digital Health*, 5(3), e105–e106. [https://doi.org/10.1016/s2589-7500\(23\)00019-5](https://doi.org/10.1016/s2589-7500(23)00019-5)
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from

- management educators. *The International Journal of Management Education*, 21(2), 100790.  
<https://doi.org/10.1016/j.ijme.2023.100790>
- Loebbecke, C., El Sawy, O. A., Kankanhalli, A., Markus, M., Te'eni, D., Wrobel, S., Rydén, P., & Obeng-Antwi, A. (2020). Artificial Intelligence Meets IS Researchers: Can It Replace Us?. *Communications of the Association for Information Systems*, 47, pp-pp. <https://doi.org/10.17705/1CAIS.04713>
- Lowy, A., & Hood, P. (2004). The power of the 2x2 Matrix, Jossey-Bass
- Lyytinen, K., Topi, H., & Tang, J. (2023). MaCuDE IS Task Force: Final Report and Recommendations. *Communications of the Association for Information Systems*, 52, 566-586. <https://doi.org/10.17705/1CAIS.05224>
- Mirbabaie, M., Brendel, A. B., & Hofeditz, L. (2022). Ethics and AI in Information Systems Research. *Communications of the Association for Information Systems*, 50, pp-pp. <https://doi.org/10.17705/1CAIS.05034>
- Norman, J. (2023, February 23). Trithemius favors vellum over paper for long term information storage. *Trithemius Favors Vellum over Paper for Long Term Information Storage : History of Information*. Retrieved March 13, 2023, from <https://www.historyofinformation.com/detail.php?id=337>
- Rice, B. S. (2020, June), How Extra Credit Quizzes and Test Corrections Improve Student Learning While Reducing Stress Paper presented at 2020 ASEE Virtual Annual Conference Content Access, Virtual On line . 10.18260/1-2--34732
- Rudolph, J., Tan, S., & Tan, S. (2023). Chatgpt: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1). <https://doi.org/10.37074/jalt.2023.6.1.9>
- Schoormann, T., Strobel, G., Möller, F., Petrik, D., & Zschech, P. (2023). Artificial Intelligence for Sustainability—A Systematic Review of Information Systems Literature. *Communications of the Association for Information Systems*, 52, pp-pp. <https://doi.org/10.17705/1CAIS.05209>
- Siemon, D., Strohmman, T., & Michalke, S. (2022). Creative potential through Artificial Intelligence: Recommendations for improving corporate and Entrepreneurial Innovation Activities. *Communications of the Association for Information Systems*, 50, 241–260. <https://doi.org/10.17705/1cais.05009>
- Strobel, G., Schoormann, T., Banh, L., & Möller, F. (in press). Artificial Intelligence for Sign Language Translation – A Design Science Research Study. *Communications of the Association for Information Systems*, 52, pp-pp. Retrieved from <https://aisel.aisnet.org/cais/vol52/iss1/33>
- Vayena, E., & Morris, A. (2023, February 28). A bioethicist and a professor of medicine on regulating AI in health care. *The Economist*. Retrieved March 3, 2023, from <https://www.economist.com/by-invitation/2023/02/28/a-bioethicist-and-a-professor-of-medicine-on-regulating-ai-in-health-care>
- The Atlantic (2023). The First Year of AI College Ends in Ruin <https://www.theatlantic.com/technology/archive/2023/05/chatbot-cheating-college-campuses/674073/>
- Wang, H., & Wang, S. (2022). Teaching Tip: Improving Student Performance by Introducing a No-Code Approach: A Course Unit of Decision Support Systems. *Journal of Information Systems Education*, 33(2), 127-134.

## Teaching Case

# How Popular is Your Name? Finding the Popularity of US Names Using Big Data Visualization

Frank Lee  
flee@gsu.edu

Alex Algarra  
Aalgarra1@student.gsu.edu

George State University  
Atlanta, Georgia 30302

### Abstract

Exploratory data analysis (EDA), data visualization, and visual analytics are essential for understanding and analyzing complex datasets. In this project, we explored these techniques and their applications in data analytics. The case discusses Tableau, a powerful data visualization tool, and Google BigQuery, a cloud-based data warehouse that enables users to store, query, and analyze large datasets. It also explored the benefits and applications of both tools and their integration with other platforms and services. The project offers an introductory insight into Tableau's functionalities, employing a data file from the US Census Bureau via Google BigQuery.

**Keywords:** Big Data, Visualization, BigQuery, Tableau

**Recommended Citation:** Lee, F., Algarra, A., (2024). How Popular is Your Name? Finding the Popularity of US Names Using Big Data Visualization. *Information Systems Education Journal*, 22(3), pp. 61-75. <https://doi.org/10.62273/XESK6929>

# How Popular is Your Name? Finding the Popularity of US Names Using Big Data Visualization

Frank Lee and Alex Algarra

## 1. INTRODUCTION

### Overview

Data visualization is an essential tool in data analytics and is used across various fields and industries to communicate insights and findings from data. Visual elements like charts, graphs, and maps represent data and information graphically in data visualization (Healy, 2018). The purpose of data visualization is to simplify complex data so it can be understood and easily interpreted. It can assist in identifying patterns, relationships, and trends in large amounts of data, making them more accessible and actionable. The best data visualizations incorporate design and communication principles, making them aesthetically pleasing, easy to read, and effectively conveying the intended message (Healy, 2018).

Visual analytics uses interactive and visual methods to analyze and understand complex data. Visual analytics allows you to explore and understand large and complex data sets flexibly and interactively by combining data mining, statistical analysis, and information visualization techniques. Visual analytics aims to enable decision-makers to gain insight and understanding of complex data sets rapidly and to support them in making informed decisions. Using visual analytics tools, users can explore and identify patterns in the data that may have yet to be apparent through other analysis methods (Keim et al., 2010).

Throughout the project, it is crucial to consider the limitations of the data, including missing values, data quality issues, and biases, and to take appropriate steps to address these issues. Communication and collaboration with stakeholders, including domain experts, are essential to ensure that the insights and recommendations are relevant and actionable.

### Tableau

Tableau is a powerful data visualization tool that allows users to create interactive and dynamic visualizations, dashboards, and reports. It is used for analyzing and presenting complex data simply and intuitively, making it easier for users to identify patterns, trends, and insights (Murray, 2018).

### Google BigQuery

Google BigQuery is a cloud-based data warehouse that enables users to store, query, and analyze large datasets. It is part of the Google Cloud Platform and offers a highly scalable and cost-effective solution for managing big data (Lakshmanan & Tigani, 2017). With BigQuery, users can quickly load and query massive amounts of data from various sources, including Google Cloud Storage, Google Drive, and third-party sources such as Amazon S3. The platform uses a columnar data storage format for faster query speeds and more efficient data processing.

### Tableau and Google BigQuery Combined

The BigQuery public datasets can be used with Tableau to create powerful data visualizations to help users gain insights and make data-driven decisions (Lakshmanan & Tigani, 2017). To get started, users can connect Tableau to their BigQuery account and access the public datasets available on the platform. Once the dataset is selected, users can create interactive visualizations, dashboards, and reports using Tableau's drag-and-drop interface and data visualization tools.

### Learning Objectives

By completing this assignment, you will be able to:

- Apply the fundamental concepts of data visualization to define a project in your field.
- Practice the core principles using widely available tools (e.g., Tableau).
- Demonstrate the best practice that presents your story in creating data visualization, including connecting to different data sources, assessing the data quality, and converting raw data into visualizations that provide actionable information.

## 2. CASE BACKGROUND

### Case Text

Alex Smith has decided to leverage the power of data analytics techniques to answer the question: how common is the name *Alex* within the United States compared to other names starting with *A*? After recently taking a data analytics class at the university, *Alex* learned of the ability to access Big Data (BigQuery) public datasets and how to analyze the data using a visualization tool

(Tableau), which has proven invaluable in answering this question. Based on his data analytics experience, *Alex* understands that finding how common his name is can be extended to any name, and the process used can give others the option of learning the same.

*Alex* began by accessing the Social Security Administration's dataset, including all names from Social Security card applications for births in the United States after 1879. He imports the Google BigQuery dataset into Tableau, where he can easily query and analyze the data. Using Tableau, he created filters only to include names starting with the letter A and then grouped the data by name to count the number of occurrences for each name. With this information, *Alex* creates a visualization using Tableau that shows the top 10 most common names starting with A and the number of occurrences for each name. He can see that the name Andrew is the most common, followed by Anthony and Arthur.

Wanting to dig deeper, *Alex* creates a map visualization of the United States with bubbles representing the number of people named *Alex* in each state. The size of each bubble corresponds to the number of people with the name *Alex* in that state. By connecting Tableau to the BigQuery dataset, he can easily map the number of *Alex*'s in each state.

Upon examining the map, *Alex* notices that the states with the highest number of people named *Alex* are California, Texas, and New York. However, *Alex* is curious to see if any regional trends or outliers may indicate other states where the name *Alex* is prevalent. Using Tableau's visualization tools, *Alex* breaks down the data by region and discovers that the Northeast and West Coast have the highest concentrations of people named *Alex*. He also notices that the name *Alex* is less common in the Midwest and the South. After completing this project, *Alex* shared the step-by-step guide and the code used to analyze the data with others interested in finding out how common their names are. Other people can follow the same steps but with their name of interest and use the same visualization tools to create maps and bar charts showing their name's frequency across different states and years.

To apply this project to other names, the user would need to modify the filters to include their name. They can then follow the steps outlined and use Tableau to create a map visualization or bar chart that shows the frequency of their name across different states and years. Overall, this project provides a framework and a set of tools

that can be used by anyone interested in exploring the frequency of their name in the United States. The step-by-step guide can help other users replicate the process and apply it to their own names. By leveraging the power of big data and visualization tools, anyone can now explore the frequency of their name and gain insights into the popularity of their name across different states and years.

### The Data Source

The project uses the public dataset created by the Social Security Administration hosted in Google BigQuery. The dataset contains all names from Social Security card applications for births in the United States after 1879. Many people born before 1937 never applied for a Social Security card, so their names are not included in this data. Records may not show the place of birth for others who did apply, and their names are not included in the data. All data are from a 100% sample of records on Social Security card applications as of the end of February 2015. Once you load the dataset into Tableau, you will begin filtering your name by the number of occurrences and geographical location.

## 3. PROJECT ACTIVITY

The purpose of this exercise is to become familiar with the process of analyzing a Google BigQuery dataset using the data visualization application Tableau. We will use the "USA Names" public dataset created by the Social Security Administration, which contains all names from Social Security card applications for births that occurred in the United States after 1879. The dataset will be used to explore how common your name is based on the given data.

The data was collected by the US Census Bureau. The instructions describe using some of the basic features of Tableau. After Connecting Your BigQuery Account to Tableau, please make sure to follow each of the steps in order.

### Creating a Project and Connecting to a Data Source

1. Select "Project" under the project tab and type "bigquery-public-data" (Appendix A, Figure 1). Once you have completed this step, you can access the "USA Names" dataset in the "bigquery-public-data" project and analyze it in Tableau.
2. After selecting the "bigquery-public-data" project, the next step is to choose the specific dataset you want to analyze (Appendix A, Figure 1). In this case, we use the "USA Names" public dataset the Social Security

Administration created. In the "Select a dataset" field, click the drop-down arrow and select "usa\_names." By selecting the "usa\_names" dataset, Tableau will load the data from this dataset into the visualization environment, allowing you to create visualizations and analyze this specific dataset.

3. After loading the data from the "usa\_names" dataset in Tableau, you will be directed to the worksheet area where you will see a list of available sheets. Click on the "Sheet 1" tab to create a new sheet to start building your visualization.

### Information Filtering

1. Once you have added the data source to the report, the next step is to create a visualization. First, let's look at the most common names beginning with the letter "A." The first step is to create a filter for the "Name" variable, allowing you to narrow down the dataset to show only the names that start with the letter "A".
2. When working with data in Tableau, filters are a powerful tool that can be used to narrow down the data to focus on specific subsets of interest (Appendix A, Figure 2). To create a filter in Tableau, you can select the "Add" button within the data source tab.
3. Next choose "Add" within the edit data source filters. After selecting the "Add" button in the Data Source tab, you will be directed to the "Edit Data Source Filters" dialog box. You can create a filter in this dialog box by selecting the "Add" button. Once you click "Add," you will see a list of all the variables in your data source. In this case, you will see a list of all the variables in the "USA Names" dataset.
4. Select the "Name" variable from this list, as this is the variable we want to filter on. Once you select the "Name" variable, you will see several filter options, including the ability to filter by individual values, ranges of values, or wildcard filters.
5. Click the "Wildcard" option in the filter dialog box (Appendix A, Figure 3). This option will open a new dialog box where you can select the type of wildcard you want to use. In this case, you want to filter the names that start with a specific letter, so you should choose the "Starts With" category. Then, you can insert the first letter of your first name in the text field provided. For example, if your name is *Alex*, you would insert the letter "A" in the

text field. It will filter the dataset only to show names that start with the letter "A".

6. After creating the filter for the "Name" variable in Tableau, the next step is to create a filter for the "Gender" variable. To do this, repeat step 5 for the "Gender" variable and select your gender. After following these steps, the dataset will be filtered only to include names matching the first letter of your name and gender.
7. After adding a filter in the data source tab or importing new data, you need to update the data set to ensure it reflects the changes made. To do this, click on the "Update Now" button which is located at the bottom of the data source tab in Tableau. It will refresh the data and apply any new filters or changes that have been made. It is important to note that failing to update the data set can result in inaccurate or incomplete visualizations, so it is always important to double-check that the data is up to date before proceeding with any analysis or visualization.

### Creating Two Basic Charts

1. Once in the Sheet view, locate the "Measures" section in the Data pane on the left-hand side of the screen (Appendix A, Figure 4). Find the measure called "usa\_1910\_current" and click and drag it over to the "Columns" shelf on the top right-hand side of the screen. Next, find the "Dimensions" pane below the "Measures" pane. Click and drag the "Name" variable over to the "Rows" shelf on the screen's right-hand side. Once both the "usa\_1910\_current" measure and the "Name" variable have been added, you should see a data table displayed in the Sheet's main view.
2. Because the Name category contains many names, we will add a filter for "Name" within Sheet 1 (Appendix A, Figure 5). To start, please click and drag the name category to the Filters section, and the Filter dialog box will appear.
3. Once in the filter, choose the "Top" tab, which allows you to filter the data based on the top values in a particular field (Appendix A, Figure 6). Once you have selected the "Top" tab, you can choose the "By field" option to select a field to determine the top values and then choose the "Top 10."
4. Choose "Sort Descending" to sort the names in descending order based on the count (Appendix A, Figure 7). You should now see a



bar chart that displays the top 10 names, starting with the first letter of your name and sorted by the number of occurrences in descending order. In the graph below, we have sorted the names descending by count and you quickly see that *Alex* is the 9th most common name in the dataset. To make the graph more visually appealing, we can edit the color palette by dragging "Name" and dropping it in the Color tab under Marks (Appendix A, Figure 8). Once you have done this, Tableau will use the Automatic color palette by default. If you want to change the colors, click on the Name legend on the right, as shown below. You can choose from various pre-built color palettes or create your own custom color palette by selecting "Palette" from the drop-down menu. To create a custom color palette, you can select individual colors by clicking on the color wheel or choosing a color scheme from the drop-down menu.

5. Once you have applied the filter and edited the color palette, the completed graph should be displayed in the "Sheet1" tab. The graph will display the top 10 names that start with the first letter of your name, based on the number of occurrences in the dataset. The bars on the graph represent the count of occurrences for each name, and the color of the bars corresponds to the color palette chosen. The x-axis displays the names, and the y-axis displays the count of occurrences. The final graph should look similar to the example provided in the appendix.
6. Tableau offers numerous types of charts that are great for data visualization. For example, Tableau can format the data into the following:
  - Bar charts: A common type of chart used to represent numerical data using rectangular bars, where the length of each bar represents the value of the data being displayed.
  - Line charts: Used to show trends in data over time, with data points connected by a line.
  - Pie charts: Circular graphs are divided into slices to represent the relative sizes of different categories of data.
  - Maps: Used to visualize geographic data and display information based on location.
  - Density maps: Use color to indicate the concentration of data points in specific areas.
  - Scatter plots: Used to visualize the

relationship between two variables and can help identify patterns in data.

- Gantt charts: Commonly used in project management to show the duration of tasks and their dependencies.
  - Bubble charts: Like scatter plots, but the size of the marker represents the value of a third variable.
  - Tree maps: Use nested rectangles to display hierarchical data, with larger rectangles representing higher-level categories and smaller rectangles representing subcategories.
7. For example, we will create a Bubble Chart, in a new sheet (Appendix A, Figure 9). First right click the Sheet 1 tab and rename the sheet to "Bar Chart" and then select new sheet and rename Sheet 2 as "Bubble Chart". After renaming the Sheet 1 tab to "Bar Chart" and creating a new "Bubble Chart" sheet, we can start building our Bubble Chart by selecting the desired variables. Once you have opened the new sheet we have created, to start, click and drag `usa_1910_current` (Count) to the Color and Size Tab in the Marks section. Click and drag Name to the Label Tab in the Marks section. After completing these steps, Tableau automatically analyzes what you are attempting to accomplish and automatically generates a bubble chart for you based on the specified criteria.

### Creating a Map Chart

We now have two good charts for this data visualization, and it is time to move on to find how many people in each state share your name. We need to create a new sheet to create a visualization showing how many people in each state share your name. First, right-click on any existing sheet, select "New Worksheet," and name it "Map."

1. Drag the "Name" variable to the "Filters" shelf, just like we did earlier in the Bar Chart sheet (Appendix A, Figure 10). In the "Filter Field" dialog box, choose the "Wildcard" option, select "Exactly matches," and enter your name in the "Match Value area." After that, click on the "Apply" button to filter the data by name.
2. Now that we have applied the filter, we can begin specifying the variables to its appropriate categories (Appendix A, Figure 11). First, click and drag the "latitude" variable to the "Rows" shelf. The "latitude" variable represents the geographic coordinate that specifies the north-south

- position of a point on the Earth's surface. By dragging "latitude" to the "Rows" shelf, we are telling Tableau to use this variable to define the vertical axis of the map. After dragging the latitude field to the Rows section, the next step is to click and drag the longitude field to the Columns section.
- Now, we will add the size and color of the data points. Here, we want to use the count of occurrences of the selected name as the size and color of the data points. To do this, click and drag the `usa_1910_current` variable to the "Size" and "Color" options in the "Marks" section. The "Size" option controls the size of each data point on the map, while the "Color" option specifies the color of each data point based on a chosen variable. In this case, we want to use the count of occurrences of the selected name for both the size and color so that we will drag and drop the same variable to both options. After dragging and dropping, you will see the map update with the selected name's count of occurrences in each state reflected in the data points' size and color. Finally, click and drag the "State" field from the "Data" pane to the Detail area on the Marks card. It tells Tableau to group the marks by state, so you can see the number of people with your name in each state.
  - Tableau has automatically created your Map. You may add a filter for the State measure to remove Alaska and Hawaii. You may also change the map's background to make it more visually appealing.
  - Now, we have created a Map visualization. Let's look at how visualizing this data looks like in a bar chart (Appendix A, Figure 12). In this case, we want to create a bar chart to show the top 20 states with the most names. We will need to use the same data source as our previous visualizations to do this. First, create a new sheet and rename it "Top 20 States with the Most Names". The next step is to click and drag the `usa_1910_current` measure to Columns and click and drag State to Rows and Colors. It will create a vertical bar chart where the height of the bar corresponds to the number of people with the given name in each state. In this chart, the x-axis represents the different states, and the y-axis represents the number of people with the given name in each state. The bars are colored according to the color palette selected, and you can hover over each bar to see the exact number of people in that state with the given name.
  - To create a filter for the Name measure, click on the "Name" dimension in the "Data" pane on the left side of the screen. Then, drag the "Name" dimension to the "Filters" shelf at the bottom of the screen. This will open the "Filter" dialog box, and choose the Wildcard option only to include your name.
  - To create a filter for the State measure and choose the top 20 states first click on the "State" dimension in the "Data" pane on the left side of the screen. Then, drag the "State" dimension to the "Filters" shelf at the bottom of the screen. In the "Filter" dialog box, select the "Top" tab. In the "Top" tab, select "By Field" and choose "`usa_1910_current`" from the drop-down menu of available fields. Set the "Top" filter to "20" (or any other desired number). Click "OK" to apply the filter. Next, change the color palette to one you think is the most effective.
- ### Dashboard and Export
- Finally, we export the visualizations to PowerPoint to create a more shareable output. Tip: To optimize a dashboard for PowerPoint, on the Dashboard tab, choose Size > Fixed Size > PowerPoint (1600 x 900). To begin, create a new Dashboard tab and rename it to `Your_Name_Data_Visualization`. You create a dashboard much like you create a new worksheet.
  - From the Sheets list left, drag views to your dashboard on the right. Now, we can click and drag the four visualizations into the dashboard area (Appendix A, Figure 13).
  - To export your new dashboard to PowerPoint, you must first click File on the top right side of the screen, next click Export as PowerPoint... In the following prompt click Select All, and finally, click Export. Tableau will then ask you to save the file to your device. Please save it under a file name of your choice and open the PowerPoint (Appendix A, Figure 14).
  - Once you open the PowerPoint, you should see your visualizations in individual slides, including the title, bar chart, bubble chart, map, and the top 20 states bar chart. From here, you can begin formatting your PowerPoint and adding the stories to your slides to make the data exploration come to life (Appendix A, Figure 15).

#### 4. PROJECT REPORT

The task in this assignment is to use Big Data (BigQuery) and a visualization tool (Tableau) to formulate and answer a (series of) specific question(s) about a data set of your choice and then write a story about the data. After answering the questions using the data, you need to create final visualizations that tell the story of the data. Finally, write a few paragraphs describing the story, the visualization, and how it answers your questions.

#### 5. REFERENCES

Healy, K. (2018). *Data Visualization: A Practical Introduction*. Princeton University Press.

Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2010). Visual Analytics: Definition, Process, and Challenges. In *Information Visualization* (pp. 154-175). Springer, Berlin, Heidelberg. [http://dx.doi.org/10.1007/978-3-540-70956-5\\_7](http://dx.doi.org/10.1007/978-3-540-70956-5_7)

Murray, R. (2018). *Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software*. Wiley.

Lakshmanan, V. & Tigani, J. (2017). *Google BigQuery: The Definitive Guide: Data Warehousing, Analytics, and Machine Learning at Scale*. O'Reilly Media, Inc.

### APPENDIX A Guided Step-by-Step Figures

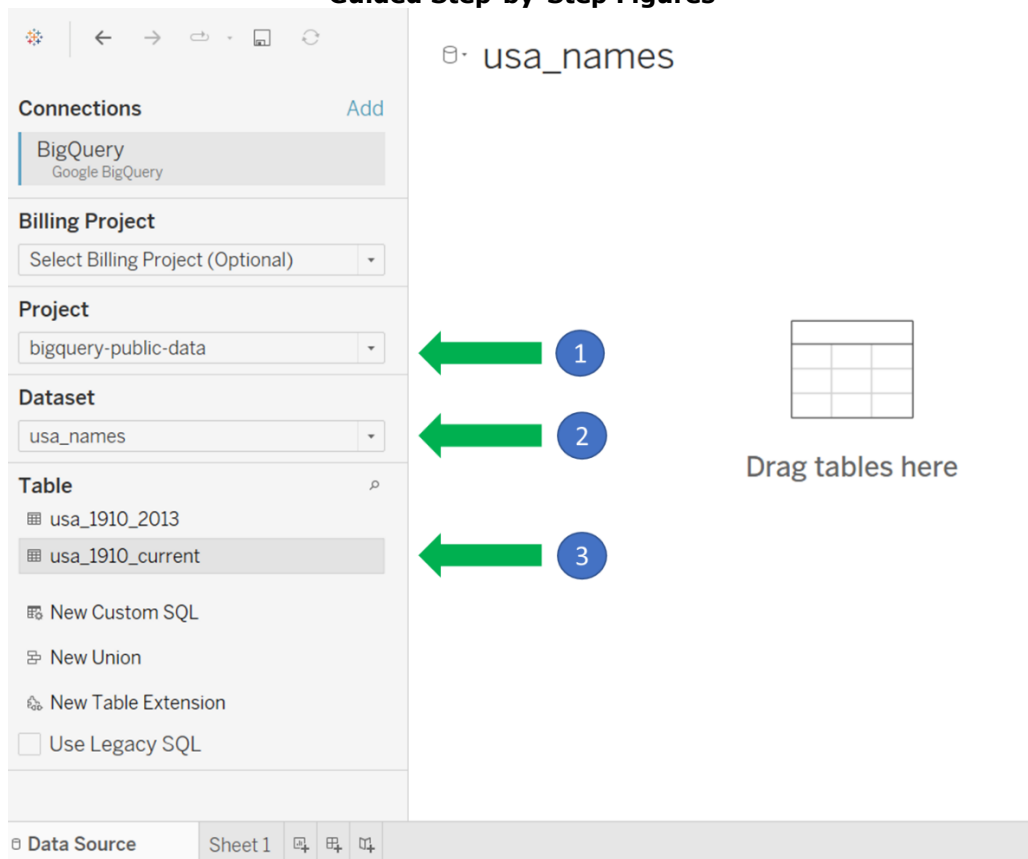


Figure 1. Loading Data Source

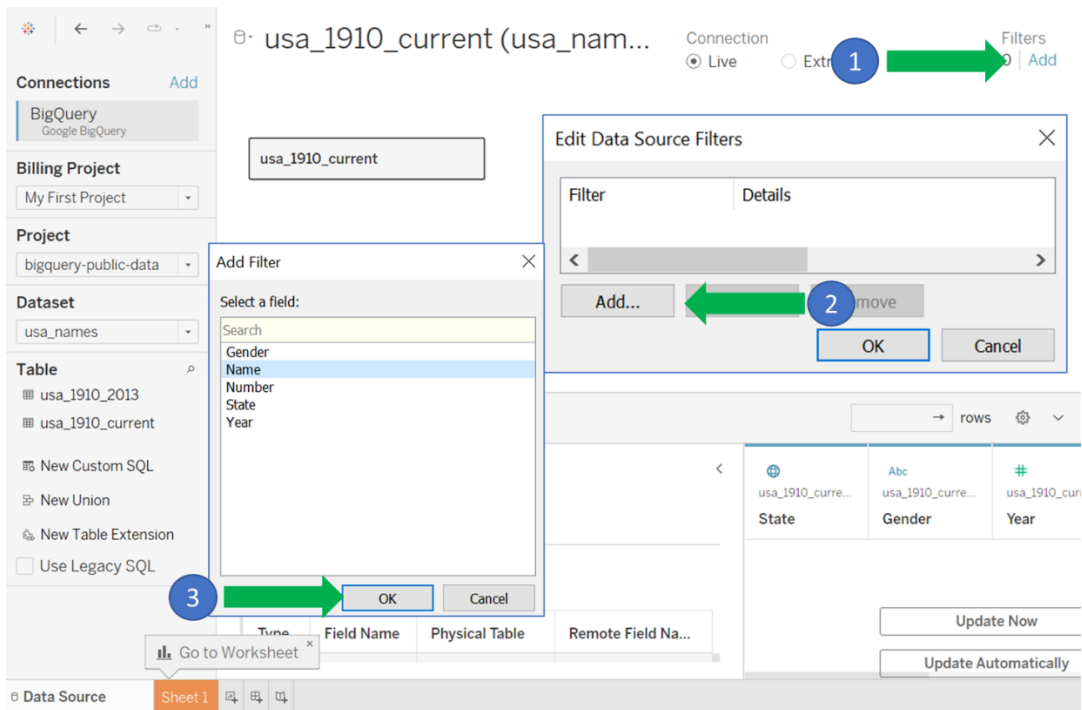


Figure 2. Filter by Variable

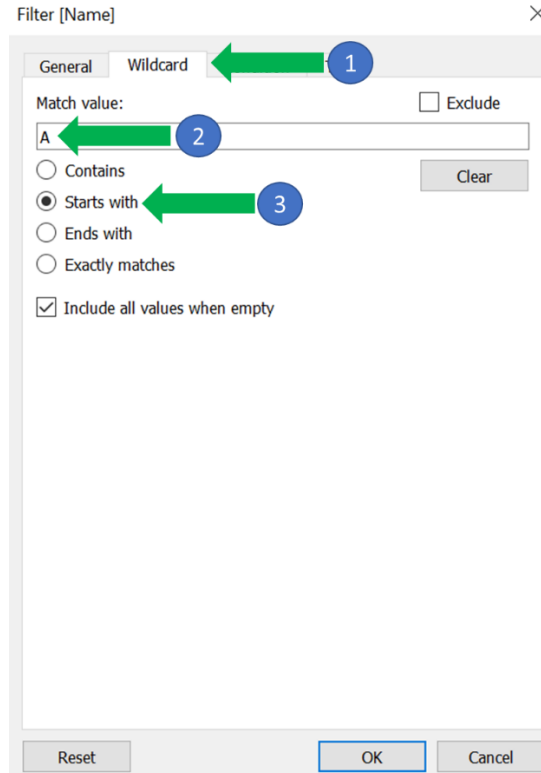


Figure 3. Filter by Wildcard

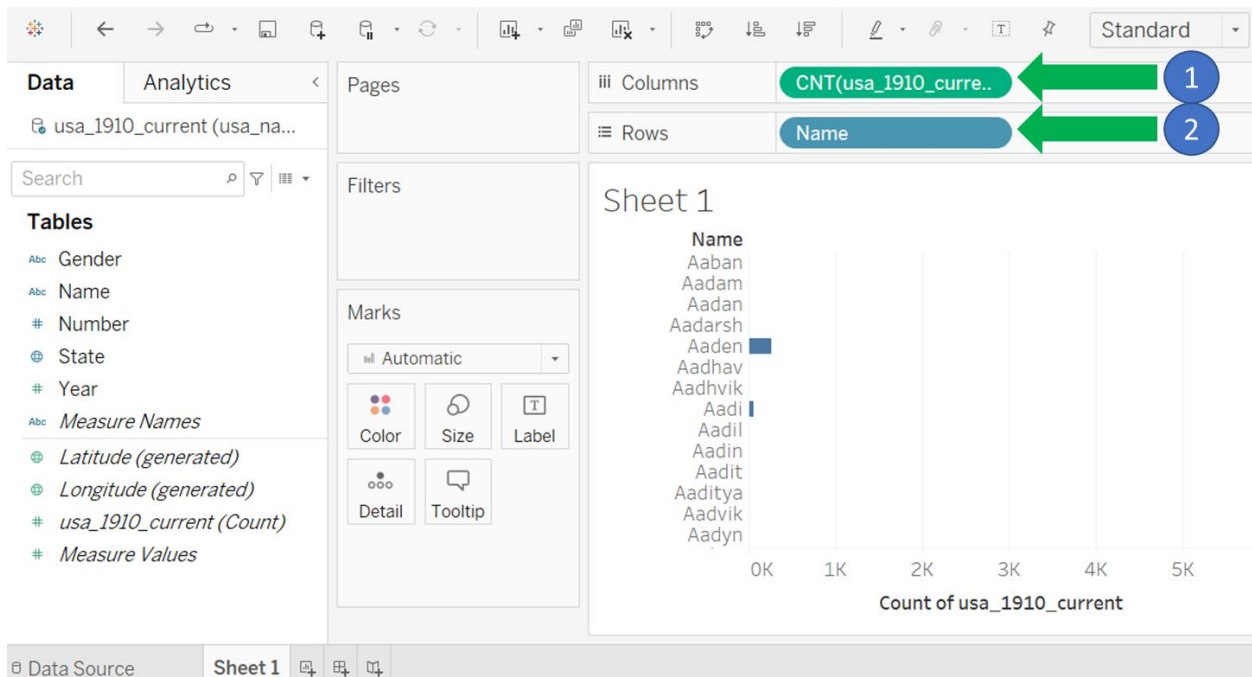


Figure 4. Inserting Data into Columns and Rows

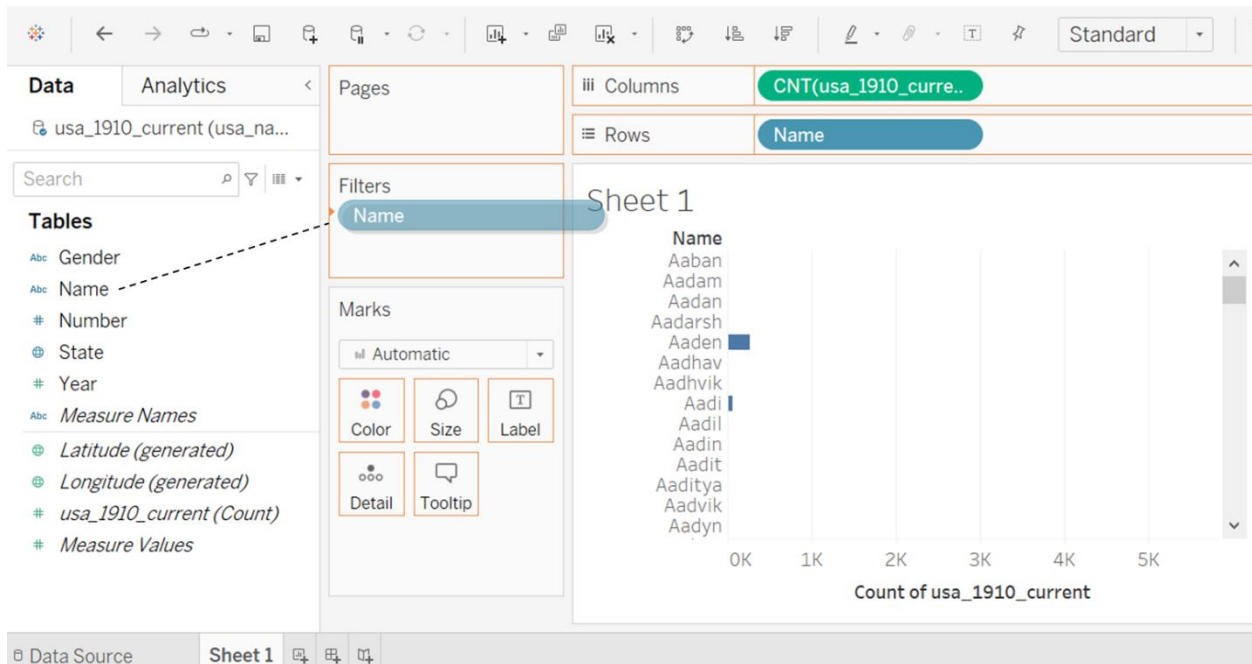


Figure 5. Applying Filter to Name

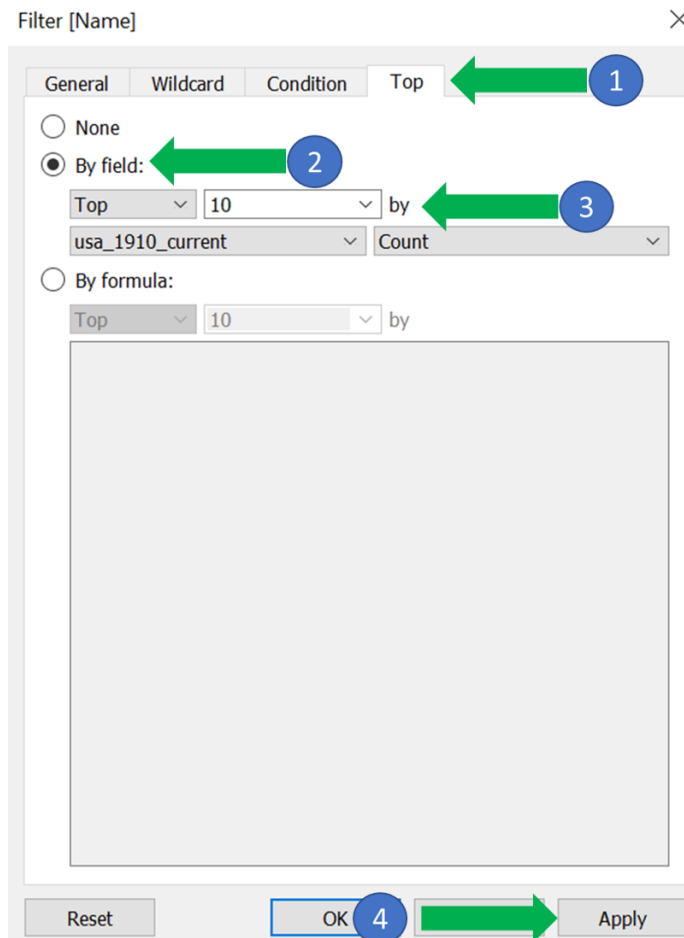


Figure 6. Top 10 names

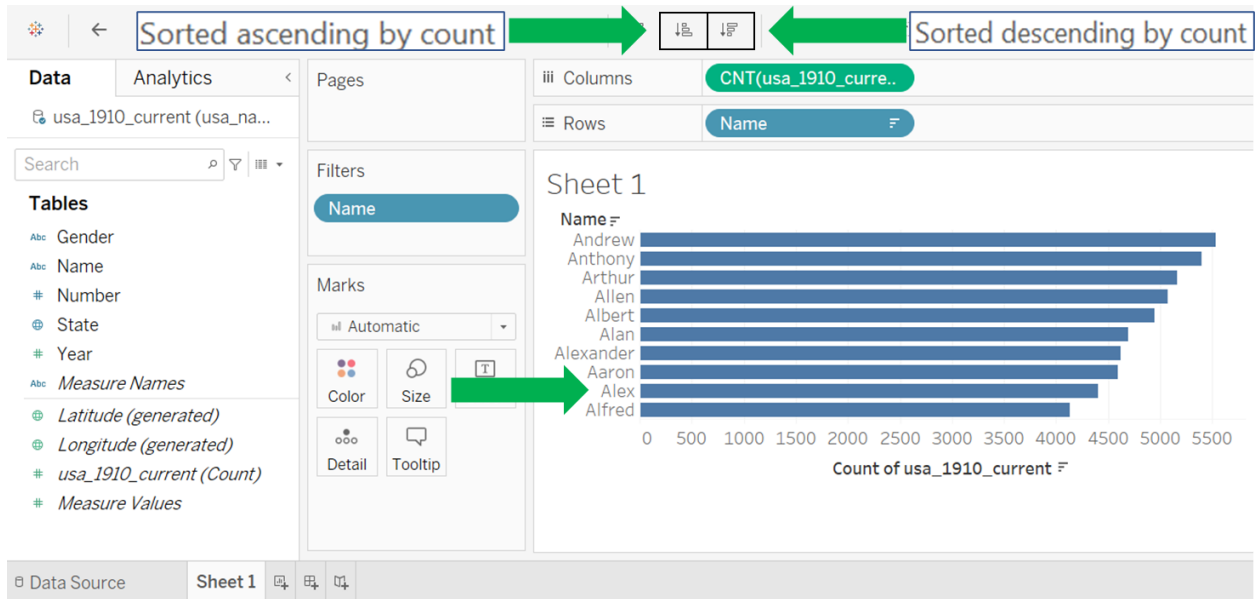


Figure 7. Sort Ascending and Descending by Count.

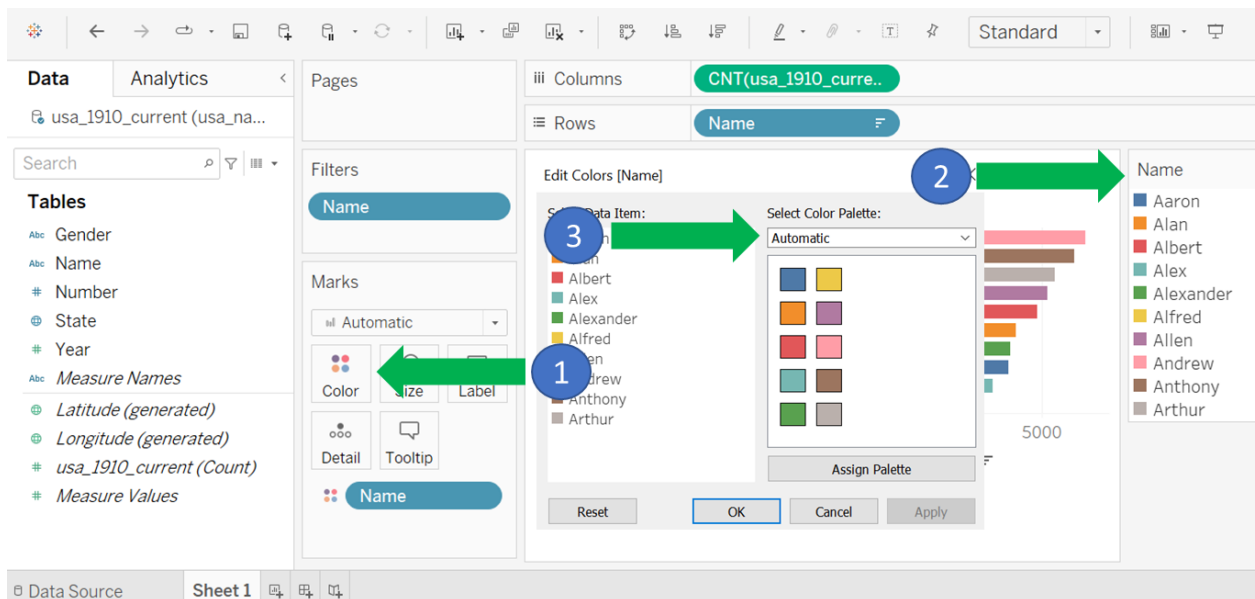


Figure 8. Edit Colors.

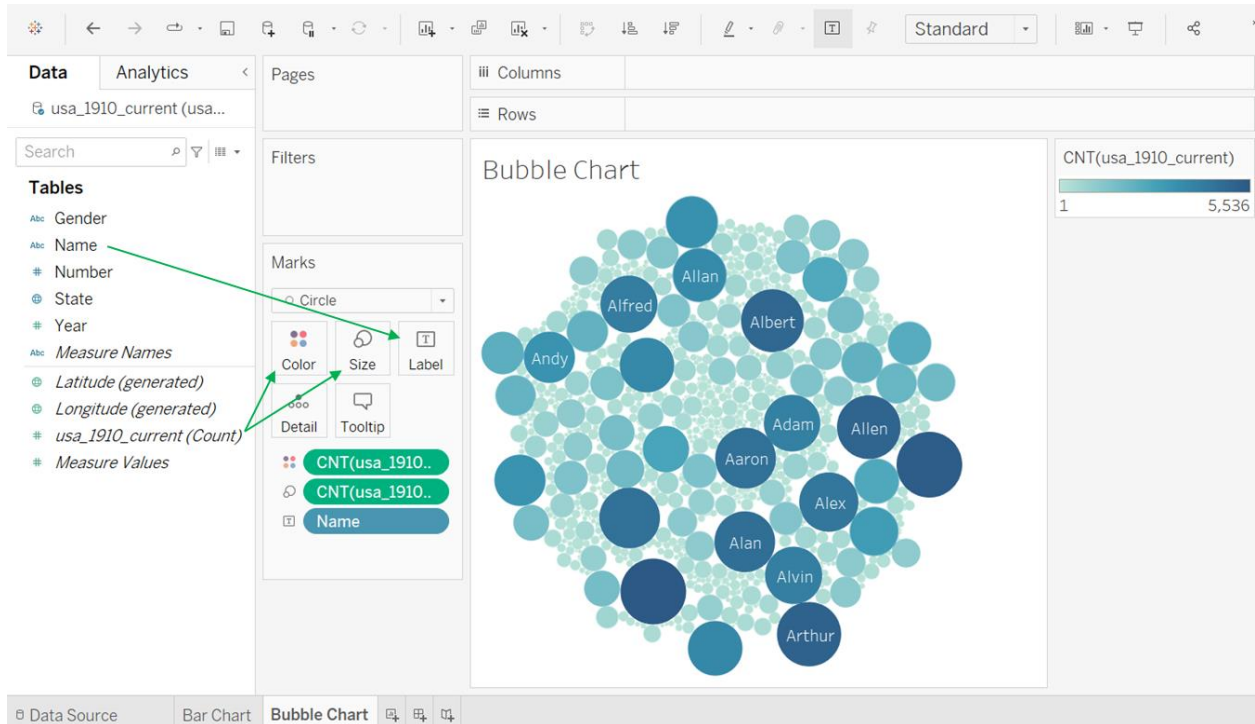


Figure 9. Bubble Chart

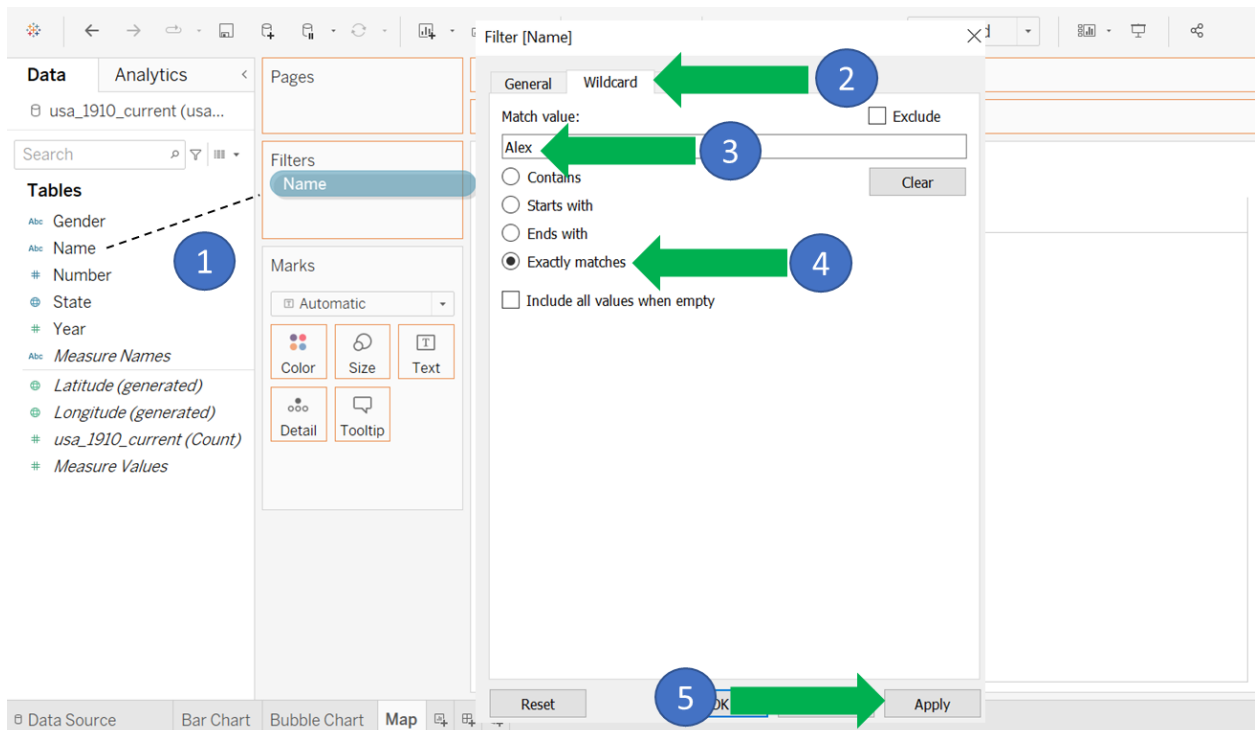
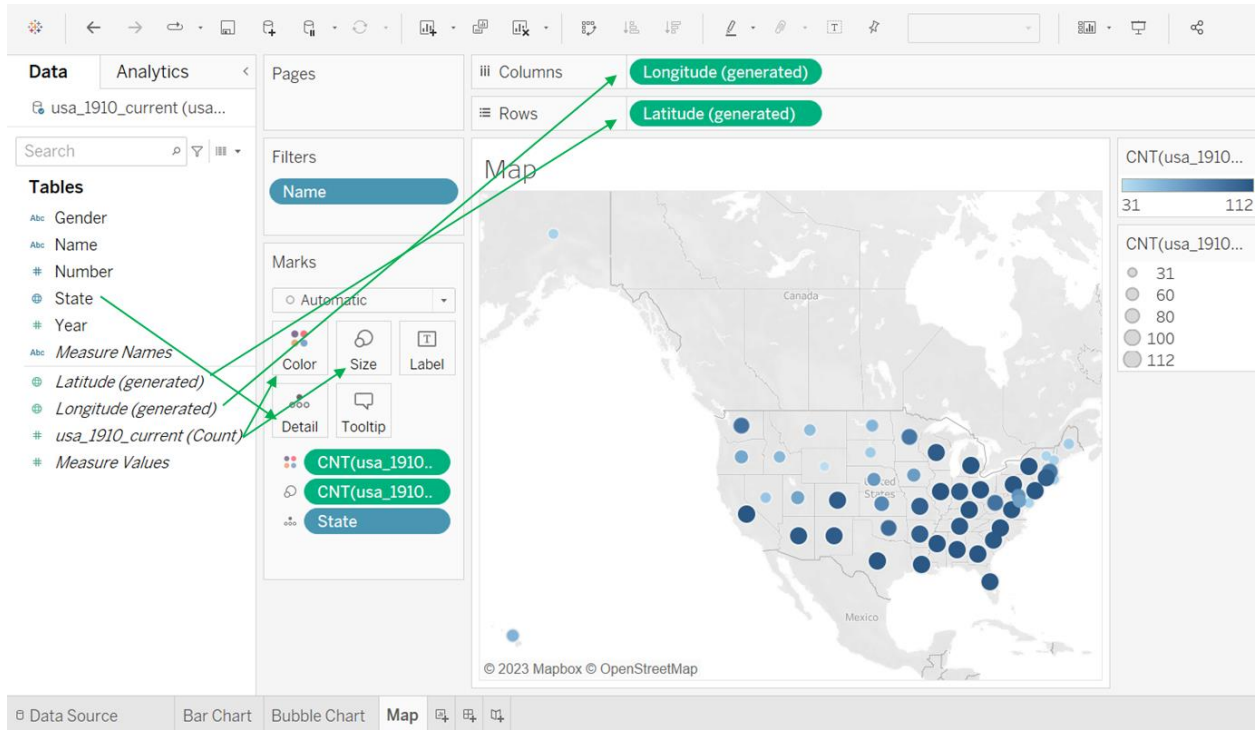
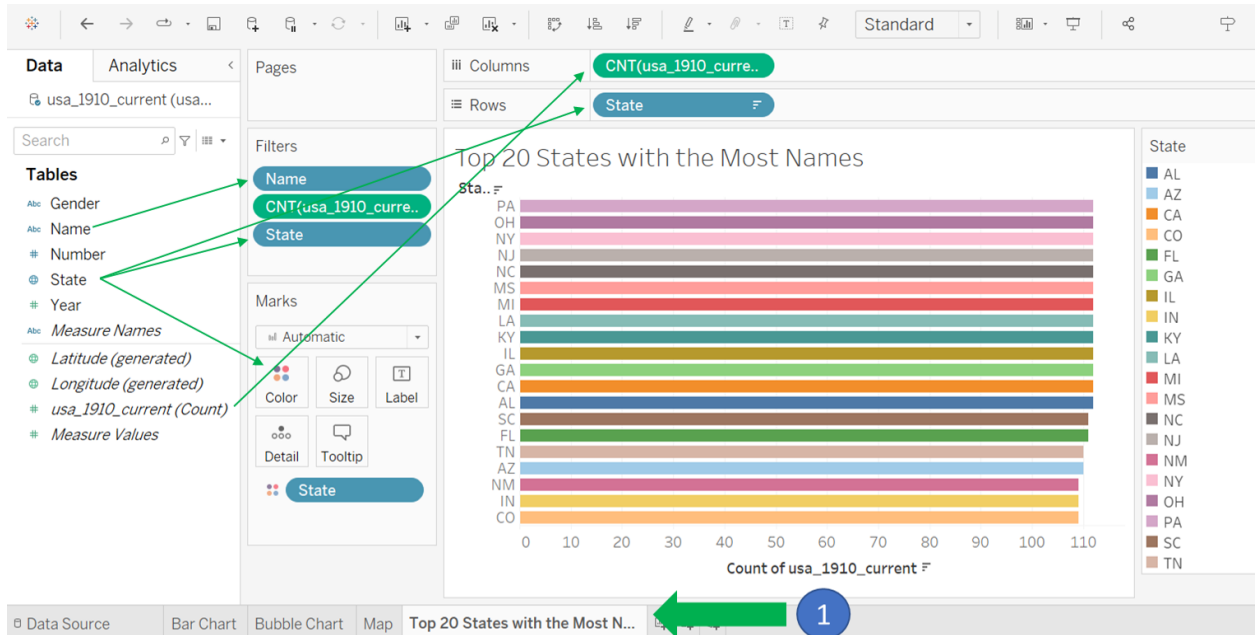


Figure 10. Name Filter for Map





**Figure 11.** Creating a Map Chart



**Figure 12.** Bar Chart Containing Total Name Count by State

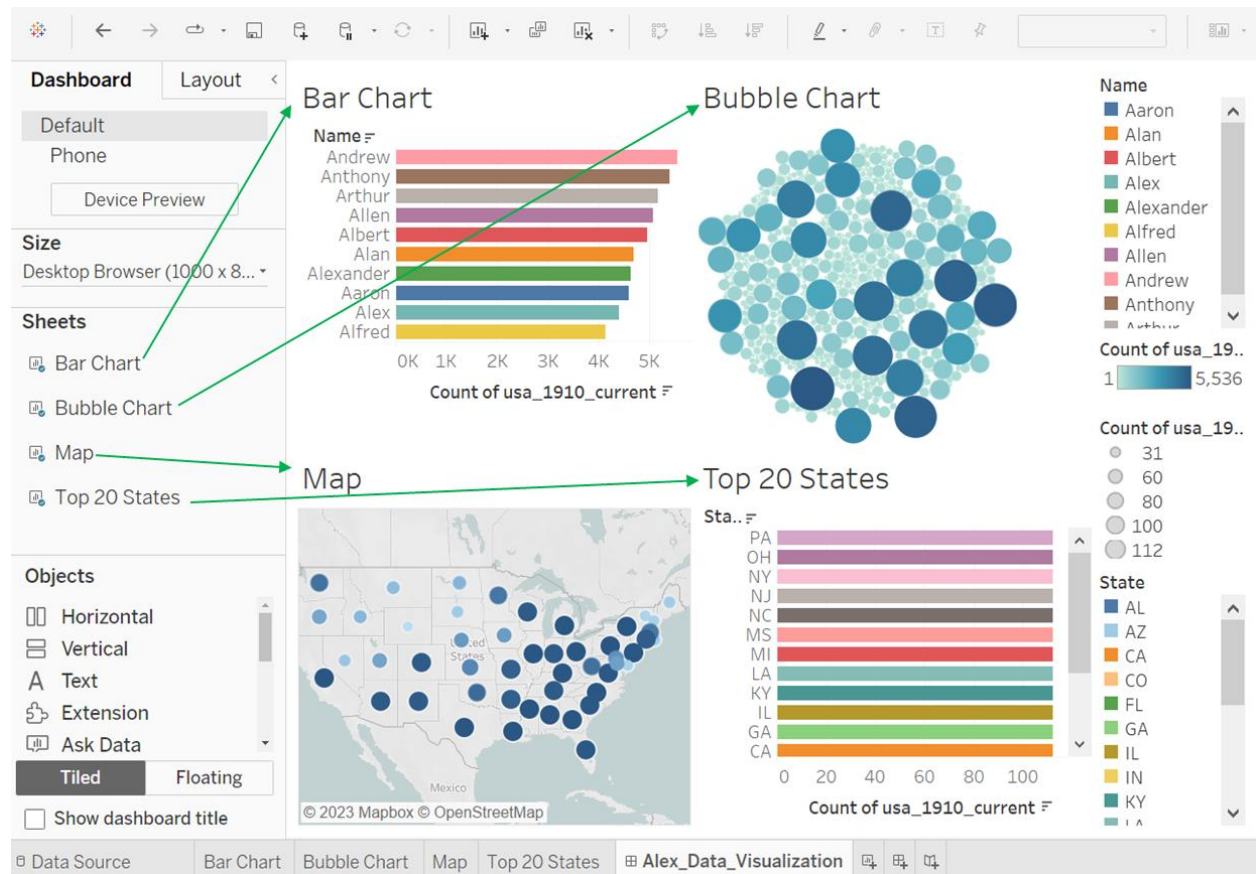


Figure 13. Visualizations Uploaded to Dashboard.

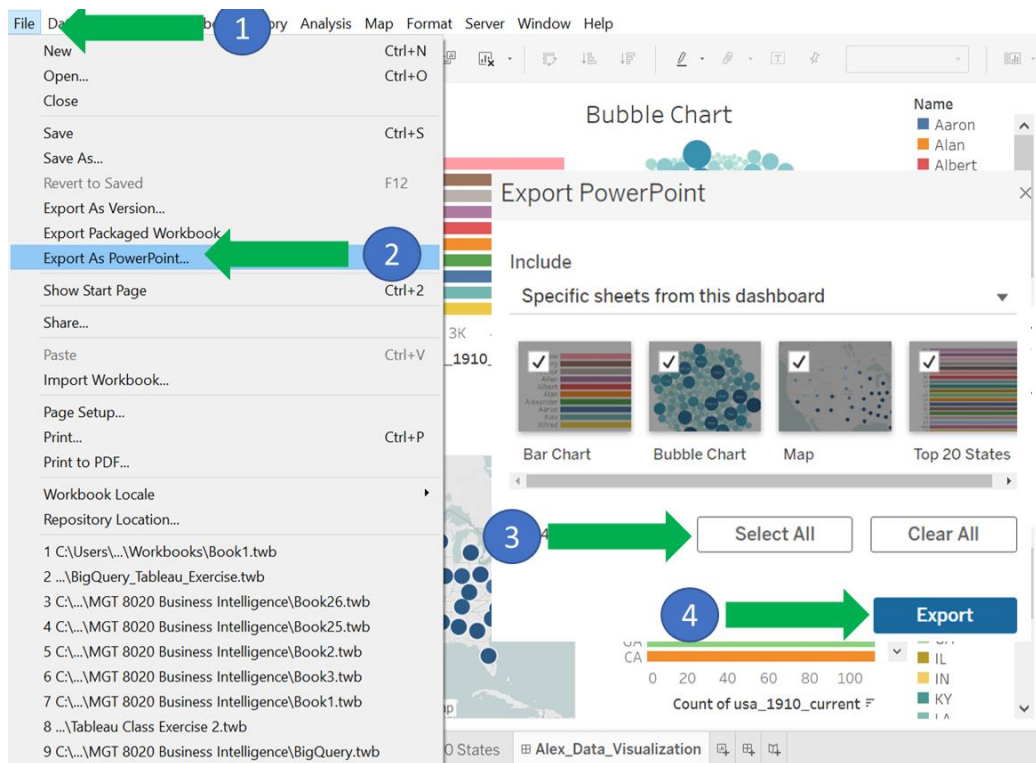
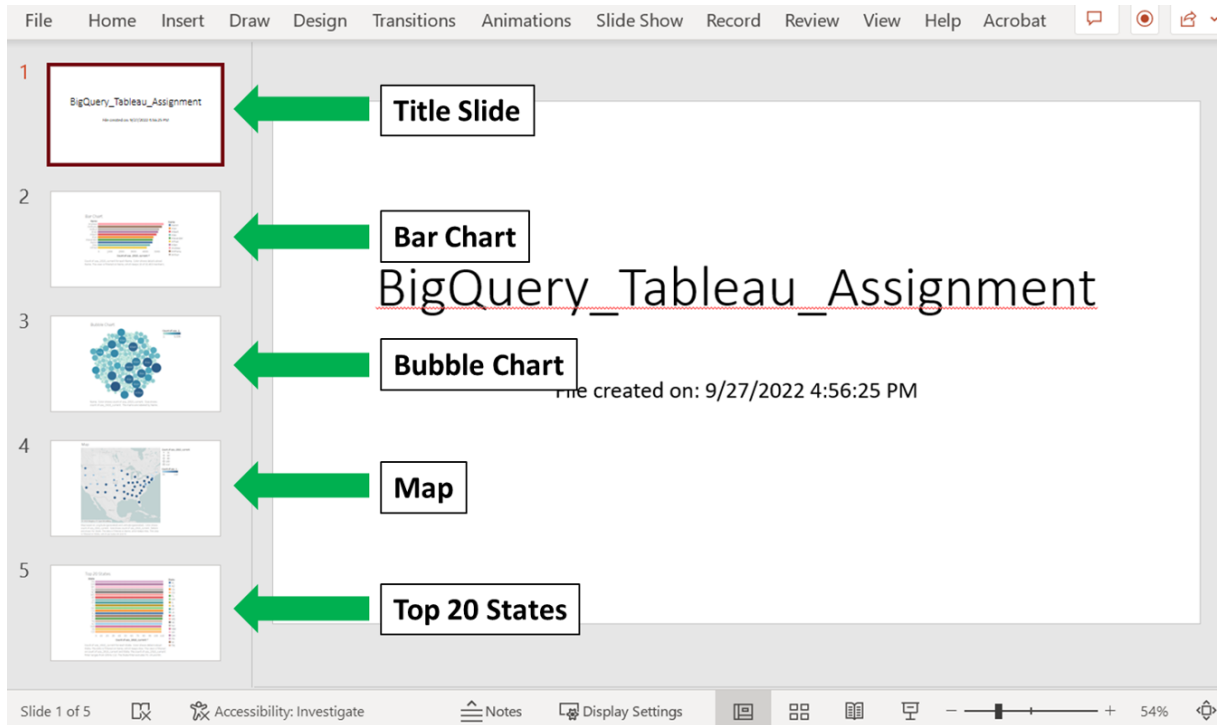


Figure 14. Export the New Dashboard



**Figure 15.** Exported PowerPoint

# Data Analytics Position Description Analysis: Skills Review and Implications for Data Analytics Curricula

Queen E. Booker  
queen.booker@metrostate.edu  
Department of Management Information Systems  
Metropolitan State University  
Minneapolis MN. 55403, USA

Carl M Rebman, Jr.  
carlr@sandiego.edu  
Knauss School of Business  
Department of Supply Chain, Operations, and Information Systems  
University of San Diego  
San Diego, CA 92110, USA

Hayden Wimmer  
hwimmer@georgiasouthern.edu  
Department of Information Technology  
Georgia Southern University  
Statesboro, GA 30460, USA

Steve Levkoff  
slevkoff@sandiego.edu  
Knauss School of Business  
Department of Economics  
University of San Diego  
San Diego, CA 92110, USA

Loreen Powell  
lpowell@maryu.marywood.edu  
School of Business and Global Innovation  
Marywood University  
Scranton, PA 18509, USA

Jennifer Breese  
jzb545@psu.edu  
IST and Cybersecurity Analytics and Operations  
Pennsylvania State University  
McKeesport, PA 15132

## Abstract

The focus of this study was to assess the skill requirements for data analytics positions and to understand data analysis employment expectations for new graduates. Furthermore, this work seeks to highlight issues relevant to curriculum management in university degree programs. 786 job postings were analyzed for domain-related, soft skills, as well as degree requirements. Soft skills, often referred to as people skills, comprised the largest part of the results (11 of the top 21 skills). Results revealed the most frequent soft skills were related to communication and teams or teamwork. The most frequent domain skills were related to visualization, data cleaning, data extraction and programming. Implications for curriculum based on results are discussed, and suggestions for future research are provided.

**Keywords:** curriculum, analytics, Information Systems, job description analysis, skills analysis

**Recommended Citation:** Booker, Q., Rebman Jr., C., Wimmer, H., Levkoff, S., Powell, L., Breese, J., (2024). Data Analytics Position Description Analysis: Skills Review and Implications for Data Analytics Curricula. *Information Systems Education Journal*, 22(3), pp.76-87. <https://doi.org/10.62273/SSQL7979>

### 1. INTRODUCTION

It is important that faculty be aware of the current skill requirements and industry demands. Faculty must prepare curricula that is appropriate and current to ensure employability for new graduates in the roles in the information systems and data analysis (Brooks, et al., 2018, Mills et al., 2016;). Data analysis and analytics are becoming more important roles in organizations as organizations seek ways to best manage the large amount of data generated on a daily basis. Hence, data analysis and analytics curricula should support the growing and changing needs to meet industry expectations (Brooks et al., 2018).

Currently, it is documented that a staggering 328.77 million terabytes of data, or 120 zettabytes, are being generated daily in 2023. Moreover, the data market is anticipated to experience substantial growth. Specifically, it is expected to increase by over 150% in 2025, hitting 181 zettabytes, which is five times the amount generated in 2019 (Duarte, 2023). These statistics highlight the importance of effective data analysis and data management. They also and provide reasons why data analytics curricula should be a key component of information systems curricula. As a result, it is important that new graduates possess the tools to help organizations ethically and competently analyze the data to support effective product development and introduction.

While research emphasizes the importance of data analysis skills, it also acknowledges the shortage of skilled individuals to fill much-needed roles in the field (White et al., 2013). The demand for skilled data analysis professionals and the shortage of individuals to meet that demand are

challenges faced in the data analysis workforce. Many of these positions will need to be filled with new entrants to the field, particularly new graduates or career changers. Thus, it is important to understand both the skills needed and expected by organizations and the role of educational institutions in developing and delivering curriculum to meet those expectations (Triche et al., 2016).

Curriculum development begins with determining the objectives and requirements for the curriculum (Cowan & Harding, 1986; Walker, 1971). Curriculum planning tends to follow either the scientific method to curriculum design and/or interactions generated by trusted professionals from the field in conjunction with faculty to produce an optimal learning experience (Knight, 2001). Curriculum developers need to know which topics, skills, and attitudes learners need to be successful in their chosen fields. Wolf (2007) presented a curriculum development process that demonstrated how universities could update their curricula through a data-informed process utilizing position descriptions and industry professional input to analyze the strengths and weaknesses of the current curriculum.

Because there is not a specific degree program that can prepare students with every skill desired by every employer, it is important the curricula program designers know the key or most frequent skills used or required. By assessing the current and projected employment environments, programs can better develop the most relevant curriculum to ensure student success in employment upon graduation. It is not uncommon for faculty to review recent industry position descriptions, survey employers and recent graduates, and ensure that the curriculum

meets current and future expectations from both student and employer perspectives (Rebman et al., 2023).

To expand upon the above research, this paper concentrated on analyzing job listings related to data analysis roles posted in May 2023 on a prominent U.S. national job board. Specifically, this research focused on positions that were open to individuals without prior experience. The goal was to understand the key skills new graduates or new entrants to the field of data analytics are expected to have.

Moreover, this study focused analysis on three areas: tools, domain-related skills, and soft skills to determine expectations of employers. The study is limited to undergraduate degrees and certificates. The rest of the paper is organized as follows. The next section provides the literature review followed by the methodology, results and limitations, and next steps.

## 2. LITERATURE REVIEW

Research has shown that prospective career opportunities can impact a student's choice of major, which is often based on ability to find employment once they graduate. Understanding the appropriate employment market and having a curriculum that supports the employment market can lead to higher employment opportunities and strong career placement for new graduates. The use of job postings to examine the desired qualifications is a well-established approach to support curriculum development in higher education which in turn supports market needs and career placement (Brooks et al., 2018, Debusse & Lawley, 2009; Harris et al., 2012; Lee & Han, 2008; Prabhakar et al., 2005; Todd et al., 1995;). Job postings enable prospective employees to know what skills are in demand. When organizations advertise job opportunities, they describe the qualifications and skills that they desire and expect in prospective employees, oftentimes distinguishing between required and preferred qualifications.

Using these job postings also provides faculty with a rich source of information to analyze how employer requirements change without having to continually request information from employers directly (Todd et al., 1995). It is widely known that industry heavily relies on colleges and universities to produce qualified individuals to support the day-to-day operations in all aspects of the organization. Information systems is not immune to this expectation. Therefore, it is incumbent on faculty to understand industry

expectations and be "in step" with needs to ensure that the investment students make in their education is consistent with the needs of hiring organizations (Brooks et al., 2018). This understanding can thus lead to effective decision making about degree programs as well as specific course and curricula designs.

Prior research in this area by Todd et al. (1995) analyzed job ads to investigate technical skills, alongside business, and interpersonal skills. Their findings indicated a gradual increase in the number of technical skills demanded or favored by employers in job advertisements over time. This shift placed greater importance on technical skills in the curricula of colleges and universities. Additional studies (Brooks, et al., 2018, Maier, Clark, & Remington, 1998; Maier et al., 2002) analyzed job postings to identify key information systems skills and qualifications and found that soft skills are becoming equally important to the domain and technical skills. Similarly, Gardiner et al. (2017) completed a similar study for big data and found that employers desired both traditional systems development skills along with analytical skills along with significant soft skills.

### Data Analytic Tools

Several data analytic tools are commonly used across various industries and research domains to extract insights from data. Microsoft Excel, a spreadsheet software, remains a ubiquitous choice for basic data analysis due to its user-friendly interface and familiarity. Excel allows users to manipulate, visualize, and summarize data efficiently. Tableau is another widely adopted tool known for its data visualization capabilities, enabling users to create interactive dashboards and explore data through visually appealing charts and graphs (Microsoft, 2021; Sharma, 2020; Tableau, 2021).

Moreover, IBM SPSS (Statistical Package for the Social Sciences) is a renowned tool for advanced statistical analysis. It provides a comprehensive set of statistical tests and procedures for data exploration, hypothesis testing, and predictive modeling. RapidMiner, an open-source data science platform, is known for its user-friendly interface and powerful machine learning capabilities, making it a preferred choice among data scientists for tasks ranging from data preprocessing to predictive modeling (IBM, 2021; RapidMiner, 2021).

Furthermore, R and Python have gained substantial popularity as versatile open-source data analytic tools. R is particularly favored for its

robust and accessible statistical modeling capabilities and extensive libraries for data analysis and visualization. Python, known for its readability and versatility in full-stack development, has emerged as a go-to language for data analysis, offering libraries like Pandas for data manipulation, NumPy for numerical computing, and scikit-learn for machine learning tasks. Python's flexibility extends to deep learning frameworks like TensorFlow and PyTorch, making it an all-encompassing choice for data scientists (Python Software Foundation, 2021, R Core Team, 2021).

The choice of data analytic tool often depends on the specific requirements of the analysis, the user's proficiency, and the scale of the data. Today, Excel, SPSS, Tableau, Python, SQL, PowerBI (Zang, 2020), Orange, RapidMiner, R (Wimmer & Powell, 2016), Qualtrics, and Access are some of the most common data analytic tools which aid and empower professionals and researchers to make data-driven decisions and gain insights from complex datasets across diverse fields.

### **Domain-Related Skills**

Domain-related skills are a crucial component of professional expertise in various fields, ranging from healthcare to technology and beyond. These skills encompass the specialized knowledge and competencies required to excel in a particular industry or discipline (Bransford et al. 2020). For instance, in healthcare, domain-related skills may include clinical diagnosis and treatment protocols, while in the technology sector, skills related to programming languages, network administration, and cybersecurity become paramount. These skills are often developed through formal education, on-the-job training, or professional certifications. Moreover, they are essential for individuals to navigate the intricacies of their chosen fields, make informed decisions, and contribute effectively to their organizations. As industries evolve, staying current with the latest developments in domain-related skills is essential for professionals to remain competitive and adaptable to emerging challenges (Davenport & Harris, 2007; Schwab, 2016).

Professionals in data analytics positions require a diverse skill set encompassing domain-specific and technical competencies. Data cleaning, the initial step in data preparation, involves identifying and rectifying errors, inconsistencies, and missing values in datasets. Data extraction is equally crucial, involving the retrieval of relevant information from various sources, such as

databases and APIs. Visualization skills enable professionals to communicate complex data insights effectively, using tools like Tableau or Power BI to create intuitive charts and graphs. Object-oriented programming (OOP) proficiency is invaluable for developing data-driven applications and automating data workflows, enhancing efficiency. Statistical knowledge is fundamental for conducting robust analyses, interpreting results, and making data-driven recommendations. Pivot tables are often used in Excel, which help to facilitate data summarization and cross-tabulation. Script programming, such as with Python or R, enables advanced data manipulation and modeling. All of these skills collectively empower data analysts to transform raw data into actionable insights, driving informed decision-making within organizations (Jones, 2017; Wickham & Grolemund, 2017). Furthermore, the significance and dynamic nature of many industries demands a continuous commitment to updating and expanding domain-related skills to address emerging trends, regulations, and technological advancements (Van Dijck; 2020, World Economic Forum, 2020).

Ultimately, domain knowledge coupled with technical competency empowers data analysts to not only analyze data but also offer strategic insights that drive innovation and progress within their industries. These multidimensional skills ensure that data analytics professionals are well-equipped to extract valuable information from data, leading to informed decision-making and industry advancements (Goodfellow et al., 2016; Few, 2009).

### **Soft Skills**

Today, soft skills are pivotal for data analytics professionals as these skills greatly enhance workplace performance and effectiveness. Specifically, effective communication skills, enable data analytics professionals to articulate their findings and insights clearly to both technical and non-technical stakeholders, ensuring a shared understanding of complex data concepts (Smith, 2017). Similarly, strong problem-solving abilities are also important soft skills for data analysts to navigate intricate datasets, identify patterns, derive meaningful conclusions and communicate with management (Brown & Jackson, 2018). Additionally, teamwork and collaboration skills aid in collaboration within multidisciplinary teams, which may amplify creativity and improve decision-making. Additionally, soft skills such as presentation skills, empathy, integrity, ethical conduct, and professional behavior are key for data analytics

professionals in building trust, maintaining ethical standards, and ensuring a positive professional image within the industry (Gupta & Sharma, 2018).

### 3. METHODOLOGY

This purpose of this study was to understand current domain, tool and soft skill requirements for new graduates in the field of data analytics.

The research questions for the study were:

1. *What are the skills organizations desire in new graduates in data analysis?*
2. *Is there a difference in skill requirements based on degree?*

This study followed the consensus pile-sort protocol methodology of Gardiner et al. (2017). R was used to strip the data of punctuation, special characters, and definite articles, parse the data, and produce a list of the most frequently occurring terms and phrases. The order of the terms remained as it was necessary to further analyze specific terms such as Excel which could be a stem for excellent or a description, e.g., "Excel at communicating." Following the consensus pile-sort protocol (Gardiner et al., 2017), two faculty and three IS industry professionals acted as informants.

The categories for the data were educational requirement, degree, tool, domain skills, and soft skills. Each term was placed into one of the agreed-upon categories. The informants reconciled any differences in the agreed upon categories.

This study follows the methodology described in previous studies described in the literature review regarding position analysis, specifically Brooks et al., 2018, Debuse & Lawley, 2009; Harris et al., 2012; Lee & Han, 2008; Prabhakar et al., 2005; Todd et al., 1995). Data was collected from a national web-based job posting site. To obtain relevant advertisements for the analysis, keywords used included "data analyst", "data analytics", "business analyst", "data science", "data engineer", "data visualization", and "information analyst". Each search queried only those job listings posted in the information technology category on the job-listing website and only positions that required no work experience.

The different types of job postings that were queried over a 30-day time period in May 2023 resulted in 1,021 job postings of which 786 were specific to undergraduate degrees as the

minimum educational requirement and no prior work experience. Data items obtained from each listing included job requirements, job title, and location. The removal of duplicate ads ensured a valid and reliable dataset.

The data was then placed in an Excel spreadsheet for examination and analysis. Anomaly analysis using count functions identified misspelled words which were corrected. Additional analysis for the term "Excel" was performed by manually reviewing a random selection of position descriptions. For example, Excel could refer to a tool as well as describe expectations of a domain or soft skill such as "Excel at communications." When Excel was referred to as a tool in the position, it was coded as a tool. When Excel was used as a descriptor, only the skill was coded, such as communication in the above example. This extra analysis ensured Excel was not over counted as a tool.

Analysis of the normalized data focused first on the frequency with which the various skills appeared in position descriptions. This analysis followed t-tests to compare the skills by degree to determine if there was a significant difference in skill requirements based on the degree and job title. Reporting of the frequencies and noted significant differences provide curriculum designers and faculty with a strong understanding of employer expectations and is consistent with previous research in the literature.

### 4. RESULTS

#### Data Overview

The number of original job positions that required no prior work experience was 1,021. Of the 1,021 positions, 786 required a bachelor's degree as the minimum educational attainment. 31 required a doctorate/PhD, 144 required a master's and 60 required a minimum of a high school diploma.

The analysis identified and provided for the review of several items categorized as a tool, domain-related skill, or soft skill. Education was used to select only those job positions for which a bachelor's degree was the minimum required education. The most frequent job titles are shown in Table 1.

Data Scientist was the highest reported job title from the search with 25% of the position titles. This was followed by Business Analyst at 17%. There were other position titles such as Information Visualist, End User Analyst, and Data Manager that appeared in the search but were



less than 1% of all titles and are not analyzed as part of this study.

Title	Number	Percent
Data Scientist	195	25%
Business Analyst	136	17%
Data Analyst	128	16%
Data Engineer	117	15%
Data Visualization Specialist	88	11%
Information Analyst	79	10%

**Table 1: Position Titles**

### Analysis of Majors and Job Titles

Types of major were analyzed because having the “correct” major on the resume helps when students are applying for positions, especially when graduate are using an electronic recruitment system. The top degrees are listed in Table 2. Data Analytics was the top major followed by Business Analytics, Computer Science, Data Science, and Management Information Systems. Math and Statistics followed at a significant distance.

The only other “major” to reach above 10% was “Any major with coursework in analytics”. This type of statement implies if students study any other discipline, if they minor in or complete a certificate or coursework in an analysis related field, they are potentially qualified to work in data analysis positions. Most position descriptions listed two or more acceptable majors.

Major	Number	Percent
Data Analytics	452	58%
Business Analytics	433	55%
Computer Science	428	54%
Data Science	408	52%
Management Information Systems	330	42%
Math	132	17%
Statistics	119	15%
Any with coursework in analytics	90	11%

**Table 2: Majors**

Table 3 (in Appendix A. Additional Tables) shows the crosstab of position titles and major. There is almost a one- to-one relationship between Data Science and Data Scientist, Business Analytics and Data Analyst, and, Management Information Systems and Business Analyst.

Table 4 (in Appendix A. Additional Tables) shows the correlation analysis between title and position. The correlation analysis was performed by using binary coding for each position and each time a degree was listed in the position description. This binary table was the basis for examining the correlations. The results show a strong correlation between data science majors and all position descriptions except information analyst. MIS majors showed a strong correlation between business analysts, data visualization specialists and information analysts.

Computer Science showed strong correlation with Data Scientist, Data Analyst, Data Engineer and Information Analyst. The only negative correlation was between any with coursework in analytics and Data Scientist indicating that students who are not majoring in one of the actual analyst positions are unlikely to gain employment as a data scientist. This is important for faculty as certificates and minors are popular methods of increasing enrollment in courses. But these certificates and minors may not be sufficient for employment as data scientists.

Given the strong correlation between desired majors in the position descriptions and the job title, the next step is to better understand the relationship between the expected major and the job title. From the position description and skills analysis, there were 31 broad categories with soft skills the largest general category, comprising 11 of the 31 skills with both domain and tools each sharing 10.

### Analysis of Soft Skills

Categorizations for analysis of soft skills included traditional soft skills such as communication, presentation, ethics, integrity, time management, and team skills. The percent of positions listing the top specific soft skills are listed in Table 5.

All positions listed communications as a desired or required skills, followed by problem solving and teamwork. Motivation, collaboration, presentation, compassion, integrity, and ethics form the second tier of most desired skills, followed distantly by professional behavior and independent learner/learning. While there were other soft skills such as time management,

respectful and comfortable in a diverse working environment, the percentage for those skills was less than 10% of the sample.

The next analysis was to determine which soft skills and tools were most often associated with each degree and job title. For soft skills, as stated earlier, communication was present in 100% of the position descriptions. Visualizations and presentation skills were present in at least 90% of position descriptions except for Information Analyst.

Professional behavior was present in over 90% of the Business Analysts and Data Analysts positions as well as the Data Scientists. Ethics and integrity were consistently in about 77% of the descriptions. A t-test between each position title and each soft skill revealed no significant difference between the means assuming equal variances.

This implies that regardless of the degree program, the curriculum should strongly emphasize soft skills with a particular emphasis on communication – both written and oral, professional behavior which includes team skills and working with diverse populations, and ethical behavior.

Soft Skills	Number	Percent
Communication	786	100%
Problem Solving	724	92%
Team/Teamwork	719	91%
Motivated	636	81%
Collaboration	627	80%
Presentation Skills	624	79%
Compassionate/ Empathy	618	79%
Integrity	618	79%
Ethical	605	77%
Professional Behavior	314	40%
Independent Learner	203	26%

**Table 5: Soft Skills**

**Analysis of Domain Skills**

Categorization of domain skills included terms such as visualization, pivot tables, statistics, programming and so forth. The results of the domain top skills are shown in Table 6.

Domain Skills	Number	Percent
---------------	--------	---------

Visualization	712	91%
Data Cleaning	695	88%
Data Extraction	611	78%
Object Oriented Programming	531	68%
Statistics	456	58%
Pivot Tables	438	56%
Script Programming	326	41%
Neural Networks	145	18%
Dashboards	127	16%

**Table 6: Domain Skills**

As shown in the table, visualization and data cleaning were the top domain skills followed by data extraction. Programming, statistics and pivot tables were the second tier, followed by neural networks and dashboard. Other skills such as software development methodologies (Agile, Waterfall, etc.) and specific algorithms for data mining were listed in the position descriptions but were all below 10% of the observed skills.

A deeper analysis showed that visualization, data cleaning and data extraction appeared in more than 90% of the positions for Data Science, Analyst and Data Engineer, 100% of the positions for Data Visualization Specialist, and less than 1% in the Information Analyst positions which strongly suggests majors related to those positions need to ensure the curriculum is strong in preparing students with visualization, data extraction and data cleaning skills. Programming appeared most often with Data Engineers, Data Scientists and Data Analysts, appearing in more than 70% of the positions. 24% of Business Analysts positions also required programming. When combining the positions where MIS was a stated required or preferred degree, MIS should be added to the list of majors where programming should be a critical component. However, the tool analysis demonstrates the difference in the specific programming expectations by positions which also impacts the type of expected programming skills by major.

**Analysis of Required and Desired Tool Knowledge/Experience**

Specific software, when listed as required or desired in the job postings, were categorized as tools. The study focused on tools for data mining, visualization, and statistics. The results are shown in Table 7.

Tool	Number	Percent
------	--------	---------

Excel	658	84%
Tableau	560	71%
SPSS/SPSS Modeler	554	70%
Python	549	70%
RapidMiner	540	69%
MySQL/SQL/ PostgresSQL/	535	68%
R	524	67%
PowerBI	479	61%
Orange	122	16%
Qualtrics	119	15%
Access	95	12%

**Table 7: Tools**

Microsoft Excel was the most required tool listed followed by Tableau, SPSS/SPSS Modeler, Python, RapidMiner and some form of query language, R and PowerBI. Orange, Qualtrics and Access completed the list of tools above 10%. There were several tools listed including ReactJS, GraphQL, MongoDB, AWS: Step/Lambda Functions, C / C++ / other compiled language experience, Github, Kibana, Oracle, Looker, GoogleData Studio, SSRS, REST, Clojure, Redis, Kafka, HashiCorp, Datorama, Segment, JSON, BASH, PowerShell, NoSQL, Allegrograph, JanusGraphy, and Neo4J which were all listed under 10% of the time. However, it is important to note the diversity of the tools actually listed.

For tools, Excel was the most mentioned tool for all positions, appearing more than 80% across all positions. The data scientist, data analyst and data visualization specialists' positions mention Tableau, PowerBI, Python, and/or R more than 45% of the time. Typically, most of these positions simply state preferred experience with visualization tools. What is surprising was the number of times SPSS and SPSS modeler were listed. A deeper analysis of SPSS and SPSS Modeler shows SPSS and SPSS Modeler are often mentioned in conjunction with other tools such as SQL, Tableau, SAS, JMP, Minitab, Knime and similar statistical software application but never as a standalone required or preferred skill. The diversity of applications for most positions seem to suggest employers are "casting a net" for software skills. However, the majority of the positions did require knowledge of programming but varied by position which suggests the type and level of programming emphasis in majors depends on the major. For example, for Data Analysts, Data Engineer and Data Scientists positions, at least two of SQL, R, and Python were listed as required knowledge whereas for other

positions, the skills were more preferred than required. This argues that having significant experience to SQL, R and/or Python or similar tools in the Data Science, Data Analytics and Computer Science curriculum is essentially required whereas in the MIS curriculum, exposure to the similar tools may be sufficient.

## 5. CONCLUSIONS, LIMITATIONS AND NEXT STEPS

This study examined position descriptions to determine the skills most in demand by organizations. It also attempts to better understand how hiring managers view preparation for data analysis employment for new graduates by comparing the most frequent skills and tools listed in position descriptions along with specific degrees.

Based on the analysis, specific degrees are expected to have expertise or knowledge regarding data extraction, cleaning, and visualization. Data scientists, data analysts and data engineers are expected to know some type of programming whether it is script, object oriented or some version of SQL. Further all positions expect new graduates to have excellent communication skills along with problem solving and teamwork. Though Excel is listed in most position descriptions, only the positions such as data scientist, data analysts, data visualization specialists and data engineers had high frequencies for visualization tools. The implications for curricula are multi-faceted.

However, this study is limited in that it only analyzed positions in May 2023 and for students graduating with a bachelor's degree. A more thorough analysis is needed to determine if skill or tool interests are changing. The study also only used data from one job position source and conducted the study on a national level. Third, though care was taken to verify the data used, it is possible that observations were missed that may influence some outcomes.

However, the study does show there are specific skills associated with specific position descriptions and degrees/majors. Faculty at regional institutions who do not have a national employment market may need to study their own regions to determine how well these results compare to their regional markets prior to using the results to adjust their curricula.

The next steps for this research are to examine more closely regional differences for position titles and degree titles for companies operating in

specific markets; examine and compare differences in skill requirements for experienced data analysts; examine and compare skills requirements for new graduate based on degree attained; and analyze position descriptions to compare those with remote work possibilities to those that are in-office positions. These additional studies would help faculty and curriculum designers improve curricula to better meet their needs based on location as well as degree program level. Further, the analysis of tools expected of experienced workers can help faculty adjust their curriculum to help students be prepared to perform better at entry level positions.

Finally, position analysis studies benefit from additional input from hiring managers. As stated in the study, this analysis was based on position analysis. While position analysis is a valid method of understanding what employers are looking for, it is still a process that looks at "as is" as opposed to "to be", meaning that the analysis does not help us to understand what employers are seeking in the future. An employer survey is in the process of being executed with a closing date of December 31, 2023. The survey will hopefully provide insight into how well what is being advertised matches what employers actually want.

## 6. REFERENCES

- Aasheim, C. L., Williams, S., & Butler, S. E. (2009). Knowledge and skill requirements for IT graduates. *Journal of Computer Information Systems*, 49(3), 48–53. <https://doi.org/10.1080/08874417.2017.1289354>
- Boud, D., & Walker, D. (1991). Experience and learning: Reflection at work. EAE600 adults learning in the workplace series, Part A. Melbourne, Australia: Deakin University.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How People Learn: Brain, Mind, Experience, and School*. National Academies Press.
- Brooks, N. G., Greer, T. H., & Morris, S. A. (2018). Information systems security job posting analysis: Skills review and implications for information systems curriculum. *Journal of Education for Business*, 93(5), 213–221. <https://doi.org/10.1080/08832323.2018.1446893>
- Brookshire, R. G. (2006). Letter from the editor: Strategies for fighting declining enrollments. *Information Technology, Learning, and Performance Journal*, 24, I–III.
- Brown, T. R., & Jackson, T. W. (2016). The information systems skill gap: Exploring the skills necessary for a thriving career in the 21st century. *Journal of Information Technology Education: Research*, 15, 263–287.
- Cowan, J., & Harding, A. G. (1986). A logical model for curriculum development. *British Journal of Educational Technology*, 17, 103–109. <https://doi.org/10.1111/j.1467-8535.1986.tb00500.x>
- Davenport, T. H., & Harris, J. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business Review Press.
- Debusse, J., & Lawley, M. (2009). Desirable ICT graduate attributes: Theory vs. Practice. *Journal of Information Systems Education*, 20, 313–323.
- Duarte, F. (2023). Amount of data created daily. <https://explodingtopics.com/blog/data-generated-per-day>
- Few, S. (2009). *Now you see it: Simple visualization techniques for quantitative analysis*. Analytics Press.
- Gardiner, A., Aasheim, C., Rutner, P., & Williams, S. (2017). Skill requirements in big data: A content analysis of job postings. *Journal of Computer Information Systems*, 57, 1–11. <https://doi.org/10.1080/08874417.2017.1289354>
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT press Cambridge. <https://doi.org/10.1007/s10710-017-9314-z>
- Gupta, B., & Sharma, S. K. (2018). Soft skills in software engineering and the relevance of team formation models. In *Proceedings of the 6th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)* 429–434. IEEE.
- Harris, A. H., Greer, T. H., Morris, S. A., & Clark, J. W. (2012). Information systems job market late 1970's-early 2010's. *Journal of Computer Information Systems*, 53, 72–79.
- IBM. (2021). IBM SPSS Statistics. <https://www.ibm.com/analytics/spss-statistics-software>
- Jones, M. (2017). *Python for Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media, Inc.

- Knight, P. T. (2001). Complexity and curriculum: A process approach to curriculum-making. *Teaching in Higher Education*, 6, 369–381. <https://doi.org/10.1080/13562510120061223>
- Kyootai, L., & Dinesh, M. (2010). Dynamics of the importance of IS/IT Skills. *Journal of Computer Information Systems*, 50, 67–78.
- Lee, D. K., & Han, J. (2008). Analysis of skills requirement for entry-level programmer/analysts in fortune 500 corporations. *Journal of Information Systems Education*, 19, 17–27.
- Lerouge, C., Newton, S., & Blanton, E. (2005). Exploring the systems analyst skill set: Perceptions, preferences, age, and gender. *Journal of Computer Information Systems*, 45, 12–23.
- Litecky, C. R., Arnett, K. P., & Prabhakar, B. (2004). The paradox of soft skills versus technical skills in IS hiring. *Journal of Computer Information Systems*, 45, 69–76.
- Maier, J. L., Clark, W. J., & Remington, W. S. Jr (1998). A longitudinal study of the management information systems (MIS) job market. *Journal of Computer Information Systems*, 39, 37–42.
- Maier, J. L., Greer, T. H., & Clark, W. J. (2002). The management information systems (MIS) job market late 1970s-late 1990s. *Journal of Computer Information Systems*, 42, 44–49.
- Malgwi, C. A., Howe, M. A., & Burnaby, P. A. (2005). Influences on students' choice of college major. *Journal of Education for Business*, 80, 275–283. <https://doi.org/10.3200/joeb.80.5.275-282>
- Marsick, V. J. (1988). Learning in the workplace: The case of reflectivity and critical reflectivity. *Adult Education Quarterly*, 38, 187–198. <https://doi.org/10.1177/0001848188038004001>
- Microsoft. (2021). Excel. <https://www.microsoft.com/en-us/microsoft-365/excel>
- Mills, R. J., Chudoba, K. M., & Olsen, D. H. (2016). IS programs responding to industry demands for data scientists: A comparison between 2011 – 2016. *Journal of Information Systems Education*, 27, 131–141.
- Prabhakar, B., Litecky, C. R., & Arnett, K. (2005). IT skills in a tough job market. *Communications of the ACM*, 48(10), 91–95. <https://doi.org/10.1145/1089107.1089110>
- Python Software Foundation. (2021). Welcome to Python.org. <https://www.python.org/>
- R Core Team. (2021). R: The R Project for Statistical Computing. <https://www.r-project.org/>
- RapidMiner. (2021). RapidMiner: Data Science Platform. <https://www.rapidminer.com/>
- Rebman Jr, C. M., Booker, Q. E., Wimmer, H., Levkoff, S., McMurtrey, M., & Powell, L. M. (2023). An Industry Survey of Analytics Spreadsheet Tools Adoption: Microsoft Excel vs Google Sheets. *Information Systems Education Journal*, 21, N5, 29-42.
- Schwab, K. (2016). *The Fourth Industrial Revolution*. Crown Business.
- Sharma, A.M. (2020), "Data Visualization", Kumari, S., Tripathy, K.K. & Kumbhar, V. (Ed.) *Data Science and Analytics*, Emerald Publishing Limited, Bingley, 1-22. <https://doi.org/10.1108/978-1-80043-876-720211002>
- Smith, M. T. (2017). Communication skills in the IT industry: Assessing the demand for oral and written communication skills in the IT workforce. *Journal of Technical Writing and Communication*, 47(4), 431-452.
- Stevens, D., Totaro, M., & Zhu, Z. (2011). Assessing IT critical skills and revising the MIS curriculum. *Journal of Computer Information Systems*, 51, 85–95.
- Tableau. (2021). Tableau: Business Intelligence and Analytics Software. <https://www.tableau.com/>
- Todd, P. A., McKeen, J. D., & Gallupe, R. B. (1995). The evolution of IS job skills: A content analysis of IS job postings from 1970 to 1990. *MIS Quarterly*, 19, 1–27. <https://doi.org/10.2307/249709>
- Triche, J. H., Firth, D., & Harrington, M. (2016). A comprehensive framework to enhance the effectiveness of the recruiting experience for data science graduates. *Communications of the Association for Information Systems*, 39(1), 1. <https://doi.org/10.17705/1cais.03901>
- Van Dijck, J. (2020). The Datafication of Health. *Culture, Medicine, and Psychiatry*, 44(2), 203-221.
- Walker, D. F. (1971). A naturalistic model for curriculum development. *The School Review*, 80, 51–65. <https://doi.org/10.1086/443014>

- Wickham, H., & Grolemund, G. (2017). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media, Inc.
- Wilkerson, J. W. (2012). An alumni assessment of MIS related job skill importance and skill gaps. *Journal of Information Systems Education*, 23, 85–98.
- Wimmer, H. & Powell, L. (2016). A comparison of opensource tools for data science. *Journal of Information Systems Applied Research*, 9(2), 4-12.
- Wolf, P. (2007). A model for facilitating curriculum development in higher education: A faculty-driven, data-informed, and educational developer-supported approach. *New Directions for Teaching and Learning*, 112, 15–20. <https://doi.org/10.1002/tl.294>
- World Economic Forum. (2020). The Future of Jobs Report 2020. <https://www.weforum.org/reports/the-future-of-jobs-report-2020>
- Zang, J. (2020). 10 Data Analysis Tools for Beginners and Experts. <https://towardsdatascience.com/10-data-analysis-tools-for-beginners-and-experts-2d083203b06e>

## Appendix A. Additional Tables

Title							
Degree/Major	Data Scientist	Business Analyst	Data Analyst	Data Engineer	Data Visualization Specialist	Information Analyst	Total
<b>Data Analytics</b>	121	63	103	67	53	45	452
<b>Business Analytics</b>	81	67	122	75	45	43	433
<b>Computer Science</b>	101	77	93	89	53	15	428
<b>Data Science</b>	195	34	37	117	12	13	408
<b>Management Information Systems</b>	43	116	99	57	9	6	330
<b>Math</b>	53	32	31	5	0	11	132
<b>Statistics</b>	73	17	0	0	0	29	119
<b>Any with coursework in analytics</b>	0	32	41	12	0	5	90

**Table 3: Position Titles and Degree/Major Counts**

Title							
	Data Scientist	Business Analyst	Data Analyst	Data Engineer	Data Visualization Specialist	Information Analyst	
<b>Data Analytics</b>	0.98	0.79	0.52	0.78	0.40	0.92	
<b>Business Analytics</b>	0.61	0.18	0.06	0.71	0.65	0.07	
<b>Computer Science</b>	0.73	0.25	0.90	0.77	0.32	0.57	
<b>Data Science</b>	0.94	0.82	0.81	0.72	0.94	0.48	
<b>Management Information Systems</b>	0.23	0.88	0.23	0.47	0.57	0.77	
<b>Math</b>	0.15	0.62	0.73	0.75	0.68	0.99	
<b>Statistics</b>	0.34	0.28	0.45	0.88	0.34	0.23	
<b>Any with coursework in analytics</b>	-0.03	0.17	0.44	0.23	0.08	0.13	

**Table 4: Correlation Matrix between Position Titles and Degree/Majors**