

In this issue:

We're delighted to deliver Issue #4, with a strong focus on the intersection of Analytics and the teaching of Analytics. Our authors have studied Tableau as a supplement to Excel, LinkedIn Learning in Analytics courses, and Excel vs. Google Sheets. And our teaching tips and case studies include predictive analytics to predict acquittals at trial, and to predict the risk of falling in knee replacement patients, as well as a case using the timely topic of taxes.

- 4. Using LinkedIn Learning as a Component of Blended Learning in Two Separate Analytics Courses—Early Results**
James J. Pomykalski, Susquehanna University
- 15. An Examination of Tableau as a Supplement to Excel to Enhance Data Literacy Skills**
Mark P. Sena, Xavier University
Thilini Ariyachandra, Xavier University
- 23. Teaching Case:
Using Supervised Machine Learning and CRISP-DM to Predict an Acquittal Verdict**
Frank Lee, Georgia State University
Clinton Baxter, American Tire Distributors
- 37. Teaching Case:
Tax Time: An Interdisciplinary Accounting Analytics Experiential Learning Activity**
Joseph M. Woodside, Stetson University
Monica Mendoza, Stetson University
- 46. Teaching Case:
Health Care Management: Preventing Post-Surgical Falls after Hip or Knee Replacement Surgery through Predictive Analytics**
Richard McCarthy, Quinnipiac University
Wendy Ceccucci, Quinnipiac University
- 53. Teaching Case
Robotic Process Automation Overdue Collections Case**
Bryant Richards, Nichols College
Nicholas Kolodziejczak, Nichols College
Kevin Mentzer, Nichols College
Kerry Calnan, Springfield College

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 EDSIGCON (<https://proc.iscap.info>) 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 36%.

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 or the publisher at publisher@isedj.org. Special thanks to members of ISCAP who perform the editorial and review processes for ISEDJ.

2023 ISCAP Board of Directors

Jeff Cummings
Univ of NC Wilmington
President

Anthony Serapiglia
Saint Vincent College
Vice President

Eric Breimer
Siena College
Past President

Jennifer Breese
Penn State University
Director

Amy Connolly
James Madison University
Director

RJ Podeschi
Millikin University
Director/Treasurer

Michael Smith
Georgia Institute of Technology
Director/Secretary

David Woods
Miami University (Ohio)
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/2023 Conf Chair

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

Copyright © 2023 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.

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

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

Using LinkedIn Learning as a Component of Blended Learning in Two Separate Analytics Courses—Early Results

James J. Pomykalski
pomykalski@susqu.edu
Sigmund Weis School of Business
Susquehanna University
Selinsgrove, PA 17870, USA

Abstract

Blended learning, which is the “thoughtful fusion of face-to-face and online learning experiences” (Garrison & Vaughan, 2008), is a pedagogical paradigm used in courses across numerous disciplines. In this paper, the use of LinkedIn Learning modules as the primary technological component is described; the focus of this first-time use (Fall 2021 and Spring 2022) in two different analytics courses is on selection and assessment of the content used from LinkedIn Learning. Early feedback from students was also solicited and is presented as well. We conclude by discussing future course enrichment plans through blended learning and the selection and incorporation of additional LinkedIn Learning content.

Keywords: Blended Learning, Data Analytics, LinkedIn Learning, Assessment.

1. INTRODUCTION

Due to the COVID-19 pandemic, many instructors, who traditionally taught in a face-to-face environment, were “forced” into teaching online. This challenge for instructors created a new opportunity to “blend” the best of the online experience with the traditional face-to-face teaching environment.

While blended learning pre-dates the COVID-19 pandemic and the idea of combining face-to-face with online instruction in education is not new (Garrison and Kanuka, 2004, Graham, 2006, Osguthorpe and Graham, 2003). The reexamination of teaching pedagogies allows for the use of blended learning as a viable teaching paradigm. Blended learning has many different definitions. The simplest definition of blended learning is “thoughtful fusion of face-to-face and online learning experiences” (Garrison & Vaughan, 2008). In this paper the definition presented by Bluc, et al. (2007). “Blended learning describes learning activities that involve a systematic combination of co-present (face-to-

face) interactions and technologically mediated interactions between students, teachers, and learning resources” (p.234) is most applicable. One of the issues in developing a blended learning course is finding appropriate technology-based content to supplement and enhance the face-to-face course activities (Alford & Sweat, 2015). The focus of this paper is on the selection, use, and assessment of LinkedIn Learning (LIL) content as the primary technological component.

This paper is structured in the following manner. Section two describes the pros and cons of the three primary design approaches to blended learning courses. Section three examines the question of content selection as one of the major hinderances to the use of the blended learning approach. This section describes how LIL content is used as an effective and powerful technological component to enhance blended learning. Section four presents a brief description of the two separate analytics courses and the initial selection, use and assessment of the LIL modules. Section five highlights the student feedback from this initial pedagogical trial. The

final two sections describe the future directions of the use of blended learning and LIL and the conclusions of our work.

2. DESIGN APPROACHES TO BLENDED LEARNING

There are multiple approaches to designing a blended learning course. Blended learning can be implemented in an incremental fashion by simply adding one (or more) online components to an existing course. However, simply adding one (or more) component(s) does not mean that a successful blended learning experience has been developed (Jeffery, Milne, Suddaby, & Higgins, 2014). In all courses, the instructor, as designer and delivery agent, plays a pivotal role in the success of the learning experience; this is especially true in blended learning environments (Mayes & Morrison, 2008).

According to Alammary, et al. (2014), a blended learning course design approach can be viewed across a spectrum, ranging from low-impact to high-impact blend.

A low-impact blend approach entails adding extra online activities to an existing course, no other content is removed from the course; this is often referred to as the “course and half” approach. The medium-impact approach “redesigns the existing course to replace some face-to-face activities with online components” (n.p). In this way, the online components reflect a more effective means of reaching course objectives than previous face-to-face components. Finally, in the high-impact approach, the blended learning course is built from scratch; this represents a full redesign of the course in which each learning objective is examined to determine the best delivery option—face-to-face or online—to meet the objective. Alammary et al. (2014) describe the benefits of, challenges with, and a set of recommendations for each blended learning approach.

In a larger study of blended learning, Rasheed, Kamsim, and Abdullah (2020) highlight a series of challenges faced by both students and faculty in the blended learning environment. The technological infrastructure imposed by the learning management system (LMS) is also addressed.

Students face both personal (self-regulation and isolation) as well as technological (technological literacy, competency, sufficiency, and complexity) challenges. Faculty do not have the personal challenges but are faced with many of the same technological challenges. Faculty also

need to find sufficient online content which is another major challenge.

The student isolation challenges are not limited to the blended learning classroom but also exist in the traditional face-to-face classroom as well. This challenge can be met by providing the proper motivation and relevance for the content, provide a clear structure to guide the students through the challenging learning activities and monitor student engagement throughout the course by providing timely feedback and personal contact (Jeffery, Milne, Suddaby, & Higgins, 2014).

3. CONTENT SELECTION

As cited in Rasheed, et al. (2020), “the task of creating quality online video has been a top challenge for teachers”. Brown (Brown, 2016) and Long, et al. (2017) state that blended learning instructors spend too much time and effort creating online content. One solution is to employ a preexisting library of content such as LinkedIn Learning.

The benefits of embedding LinkedIn Learning in a University classroom are that students retained more information, and found that learning the material was both easier and more enjoyable (Petrone, 2020).

The author’s institution, Susquehanna University, adopted LIL in March 2020. Professional courses in both the Sigmund Weis School of Business (SWSB) and the Communications department incorporated LIL directly to the classroom and beyond (Petrone, 2021). SWSB Dean Matt Rousu said “LinkedIn Learning is a valuable platform at all stages of a career, so it is great that our students are active now using this resource” (Petrone, 2021).

3.1 LinkedIn Learning

LinkedIn Learning is an online learning platform with over 16,000 on-demand courses. The content on LIL can be used to enhance the learning experience of the students. LIL allows faculty to curate individual lessons and courses by tailoring the content to specific course learning goals and objectives. Faculty can create learning paths of multiple courses or videos to share with students or to embed within the course LMS to allow blended learning to be enabled.

Alford and Sweat (2015) surveyed 300 students about the desired qualities in videos used in a blended learning course; the LIL courses address many of these desired qualities.

Every LinkedIn Learning course has a consistent format. Each course is broken into smaller lessons that are generally no more than five minutes long allowing the students the ability to pause their learning activity and return to it later (Alford & Sweat, 2015). In addition, the modules can be viewed (from their laptop) numerous times or paused and restarted at the discretion of the student. The high-quality, professionally developed and delivered content in LIL is both easily accessible and digestible which overcomes two of the primary challenges for students (Rasheed, et al, 2020; Alford & Sweat, 2015).

One of the major issues discovered in the Alford and Sweat (2015) survey is that students reported needing the motivation of a graded incentive. For the courses described in the next section, the students are given a set of specific learning items known as “takeaways” (see Appendix for an example of the takeaways for Project Management LIL lessons). Takeaways are for the benefit of the student; they serve as highlights of the important concepts in the readings or LIL content; they are optional, not submitted, and not graded. Takeaways are used as questions on in-class quizzes which are generally given at the beginning of the next face-to-face class session; students may use their completed takeaways in taking the quiz. Spanjers, et al. (2015) found that quizzes had a positive impact of the effectiveness and attractiveness of blended learning. Finally, “takeaway” items often reappear on exams within each course.

One additional benefit to the use of LIL content is that students can easily build an online professional development presence by enabling the course completion certificates to be displayed on the LinkedIn professional profile (LinkedIn Learning, 2022). These completion certificates highlight, to future employers, the skills the student has acquired and their interest in building skills to aid their careers (Santa Maria, 2022).

4. COURSES

Blended learning, using LinkedIn Learning courses, was attempted in two different analytics related courses during the 2021-2022 academic year. In Fall 2021, a sophomore-level business foundation course, which serves as an introduction to business analytics, was taught using various LIL courses. In Spring 2022, senior level business foundation course in data analytics, which is focused on data visualization, was

redesigned using LIL courses. Both courses are discussed in this section.

4.1 Data Collection and Modeling

The Fall 2021 the sophomore level business foundation course, entitled Data Collection and Modeling, was the first attempt at incorporating individual modules from LIL courses and full LIL courses as the online component.

The course is described as (Susquehanna University, 2021-2022):

This is the first of three business analytics courses. This first course deals primarily with assignments and projects that collect and structure data as part of the preparation for data analysis and visualization. A major emphasis is on understanding data and issues in its collection, management, structuring, and use for the analysis and solving of business problems. Students will study project management, data collection/cleaning/structuring, data modeling, and Structured Query Language (SQL).

A low impact blended approach was used by simply adding the LIL content to the existing material in the course. Two significant modules (courses) from LIL were added to the course. The two courses correspond to the two primary learning modules (Project Management and SQL). Other smaller LIL courses were used as optional (“highly recommended”) modules to upgrade student Excel skills to work more effectively and efficiently on the term project.

4.1.1 Project Management

The Project Management section of the course provides a basic introduction into project management including the development and construction of a Gantt chart (with a baseline).

The students were required to view the first four modules of the LIL course, Project Management Foundations (Biafore, 2019). These four modules cover a basic introduction, as well as the initiation, planning and schedule development stages of project management lifecycle.

Through LIL modules and in-class activities students gain the skills needed to develop, build, and maintain a Gantt chart for a project. As part of the course team project, each student team is required to develop and maintain a Gantt chart. The teams must create the initial schedule, establish a baseline, and update their progress throughout the semester.

4.1.2 SQL

The SQL section of the course provides a basic introduction to developing SQL commands to retrieve data from a relational database. While the main support for this module has been and remains a textbook (Forta, 2013), this full LIL course (Thouin, 2021) serves as initial introduction and a skill development resource for learning SQL.

The final assessment on the SQL material covered in the course (both from the LIL course and face-to-face meetings) involves each student completing a set of in-class exercises to build SQL statements to retrieve data from a Microsoft Access database.

4.1.3 Excel

Since Excel is used extensively for data collection, formatting, cleaning, and analysis, basic Excel skills are expected. In addition, one of the objectives of SWSB business foundation curriculum is to enhance the Excel skills of all business students. Since this course serves as the first “technical” course in the business foundation, Excel skill development is key learning objective for the course.

While the Excel LIL course (Taylor, 2020) was not a required element, students were expected to have the necessary skills to perform Excel operations that are covered in this course. An additional, short LIL course on developing and using pivot tables (Ludwig, 2018) was also optional, but highly recommended for students to complete to assist in the analysis of their project data.

4.2 Data Insight and Visualization

A low-impact blended learning approach incorporating LIL modules and courses was also used in a Spring 2022 senior level business foundation course entitled Data Insight and Visualization.

The course is described as (Susquehanna University, 2021-2022):

This is the third of three business analytics courses. This four-semester-credit-hour is both writing-intensive and team-intensive and provides a framework for understanding the technologies associated with algorithmic analysis and data presentation for business decision-making. Students will study analysis techniques that enable insights and patterns to be drawn from descriptive, predictive, and prescriptive analytics. Tools that support these techniques will be researched and

presented by student teams, and they will also be investigated through individual research projects.

As the online component two LIL courses were utilized. The two courses correspond to two primary learning objectives of the course (Chart Development/Usage and Data Visualization). Other smaller LIL courses, like the ones used in the Fall 2021 course, were used as optional, but highly recommended, to upgrade student Excel skills to work more effectively and efficiently on the course project.

4.2.1 Chart Development/Usage

Since improving the student understanding and use of different chart types is a key learning objective, students are required to state their rationale for chart choices in telling their data story.

To begin to understand different chart types and how to choose between various options, the LIL course by Taylor (2019) is utilized. Taylor focuses on the QuickLayout option within Excel to give students insight to making the correct chart to display their data. Students are required to view the entire 55-minute course.

The primary visualization tool in the course is Tableau. Since Tableau only displays appropriate charts based on data types, the students still must select the chart that best tells the story of the data based on the business question.

A series of in-class exercises, using Excel data sets, are assigned in which the student must not only create a chart for the data (both in Excel and Tableau), given a listed objective, but they must justify their selection of the final chart type from the alternative charts they considered.

4.2.2 Data Visualization

Just prior to the beginning of the Spring 2022 semester, a new LIL course was discovered that covered multiple topics on data visualization (Shander, 2022). Shander is continuing to build this LIL course by adding additional lessons periodically on various data visualization topics—there are (as of February 2023) 39 lesson and listen sessions available—at the conclusion of the Spring semester course, there were only 29 lesson/listen sessions available.

Each lesson/listen section deals with Shander introducing a data visualization topic. Examples of the topics include: Data Literacy, The Data Visualization Process, Color in Data Visualization, Making Decisions with Data Viz, and Visualization

in the Real World. The second part of each lesson (the listen section) is an interview that Shander conducts with an "expert" on the topic.

The lesson on data literacy was used as part of the introduction to the course; it is listed as one of the learning objectives in each of the three analytics courses in the business foundation curriculum.

For the data literacy lesson/listen session the students were again given a set of takeaways, a quiz, and these takeaways were also used on exams later in the course.

In the Spring 2022 course, the students were asked to pick three additional topics—of their own choosing—and create a report that covered:

1. an introduction of the topic and a rationale for their choice of the topic,
2. a summary of both the lesson and listen sections, highlighting at least two noteworthy things they learned, and
3. two thoughtful lessons they learned from viewing the entire module.

The assessment of this LIL module was different because there was not sufficient time to examine each of the topics from the LIL course (Shander, 2022). Two advantages of this approach were that students were given the flexibility to examine topics of their choosing and the selection of the most interesting topics—as selected by the students—could be used as the focus of future versions of the course (see section six). This assignment is part of a number of individual assignments in the class that constitute 25% of the final grade. The assignment is shown in Appendix two.

4.2.3 Excel

While the primary visualization tool used in this course was Tableau, initially students used their familiarity with Excel to create visualizations with smaller data sets.

However, due to the course sequencing, some students had weaker Excel skills. For those students who had not yet examined the Taylor Excel LIL course (Taylor, 2020) used in the Fall course, they were given the option of accessing this course. Finally, each student could also access the short LIL course on developing and using pivot tables (Ludwig, 2018).

For the Spring 2022 course, no LIL courses dealing with Tableau were utilized, however,

several different courses are under consideration at the present time for Fall 2022.

5. STUDENT FEEDBACK

Student feedback on the use of the LinkedIn Learning modules was only solicited during the Spring semester (a shortcoming due to the instructor's lack of preparedness).

However, the feedback from the Spring 2022 course was generally very positive. The students enjoyed the LIL courses, especially the Shander course (2022).

There were 22 students enrolled in the Spring 2022 course. Of the 22, 15 provided feedback on the use of the LinkedIn Learning work. Specifically, the instructor asked the students to comment on the impact of the LinkedIn Learning assignments on their learning.

Of the 15 students, 14 were positive, only one student said "While I was able to gain some knowledge from them, I would not say the insights from the LinkedIn Learning courses were worth the equivalent amount of time and effort put into them".

Many students commented on the usefulness and benefits of adding the LinkedIn Learning content to expand and enhance their understanding of data visualization topics. One student said "The LinkedIn learning assignments were very beneficial because they explained difficult topics in a clear and concise manner, and included guest speakers who had real-world experience with the topics."

One student saw the benefit of using LinkedIn Learning to enhance their skills beyond the course as well when they commented "Another way Dr. Pomykalski helps students in professionally is through LinkedIn Learning assignments which help to build certain critical hard and soft skills we will need to care on later in life."

Given the very positive feedback generated from this first-time use of LIL lessons and courses further expansion and assessment of LIL as blended learning content is under consideration.

6. FUTURE INCORPORATION

Given the positive feedback gained from the students, future iterations of both courses will increasingly incorporate LIL content. In particular, the courses—both currently beginning re-design for Fall 2022—will transition to the

medium-blend approach (Alammary, Sheard, & Carbone, 2014).

For both courses, the lesson/listen module on data literacy (Shander, 2022) will be used in the future. Since data literacy is a shared course objective, this LIL module, along with additional readings will be used to reinforce this topic. The readings from the original course structure were reconfigured to make them shorter and for the student to rely more on the LIL content for their learning.

Since the students' Excel skills are generally weak in the Data Collection and Modeling course, the Excel LIL course (Taylor, 2020) will move from being optional to a mandatory element and completed within the first two weeks; a short exam on the content will be used as an assessment. Also, a better approach to incorporating and leveraging both the project management and SQL from the LIL material will be examined. The project management LIL material (Biafore, 2019) will provide more of the focus for the student work and a shorter additional reading on project management fundamentals will be assigned.

In the Data Insight and Visualization course, a glaring oversight for the Spring 2022 courses was the lack of blended learning content on the use of Tableau. For Fall 2022, the Tableau Essential Training course (Frye, 2021) is being used as a basic introduction to Tableau for the students.

Also, in the Data Insight and Visualization course, the decision has already been made on a different text to support the student's understanding of different chart types and other important data visualization topics. The Excel charts course (Taylor, 2019) will be integrated closely to the specific topics (chart types) within the textbook. The feedback gained from the Spring assignment in examining the different selected modules from the Shander (Shander, 2022) LIL course is being incorporated as individual topics (e.g., The Data Visualization Process and Color in Data Visualization) in the Fall semester and supported by the text.

7. CONCLUSIONS

The piloting of LinkedIn Learning courses as the online component to a blended learning environment has been successful and encourages the author to continue to move to a medium blend course for both courses in the future. As stated in the previous section, one of the main objectives will be to incorporate the online materials (primarily from LinkedIn Learning) into the face-

to-face materials in a seamless manner. Students need to be given time to complete the LIL content and therefore some consideration of fewer face-to-face meetings needs to be considered in a medium blend course. (Alammary, Sheard, & Carbone, 2014).

LinkedIn Learning is a valuable environment that will not only add to the depth of the learning experience but will help the students both professionally, through adding career skills, adding certificates to their LinkedIn profile to aid them professionally (LinkedIn Learning, 2022) and to create in each student a sense of life-long learning (Santa Maria, 2022).

8. ACKNOWLEDGEMENTS

The author would like to acknowledge the efforts of individuals for their contributions toward making this paper better. Dr. Peter Dadalt, Assistant Professor of Finance at Susquehanna University spent time reviewing the manuscript prior to the submission deadline. In addition, the thoughtful and pointed suggestions of the reviewers helped to refine and strengthen the paper.

9. REFERENCES

- Alammary, A., Sheard, J., & Carbone, A. (2014). Blended Learning in Higher Education: The Different Design Approaches. *Australasian Journal of Educational Technology*, 30(4). Retrieved June 26, 2022, from <https://ajet.org.au/index.php/AJET/article/view/693>
- Alford, K. L., & Sweat, A. R. (2015, November 15). Getting Started with Blended Learning Videos. *The Teaching Professor*. (M. Publications, Ed.) Madison, WI, USA. Retrieved July 8, 2022, from Teaching Professor: <https://www.teachingprofessor.com/topics/teaching-strategies/blended-flipped/getting-started-with-blended-learning-videos/>
- Biafore, B. (2019, June 28). Project Management Foundations. LinkedIn Learning. Retrieved September 7, 2021, from <https://www.linkedin.com/learning/project-management-foundations-4?u=86746674>
- Bliuc, A.-M., Goodyear, P., & Ellis, R. (2007). Research Focus and Methodological Choices in Studies into Students' Experiences of Blended Learning in Higher Education. *The Internet and*

- Higher Education*, 10, 231-244. Retrieved June 26, 2022
- Collection%3A6739922207503355904&u=86746674
- Brown, M. (2016). Blended Instructional Practice: A Review of the Empirical Literature on Instructor Adoption and Use of Online Tools in Face-to-Face Teaching. *The Internet and Higher Education*, 31, 1-10. Retrieved June 26, 2022
- Mayes, T., & Morrison, D. (2008). You Take the High Road: National Programmes for the Development of e-learning in Higher Education. *Reflecting Education*, 4(1), 6-16. Retrieved June 26, 2022
- Cleveland-Innes, M., & Wilton, D. (ND). *Guide to Blended Learning*. Commonwealth of Learning. Retrieved from <https://openbooks.col.org/blendedlearning/front-matter/introduction/>
- Petrone, P. (2020, July 14). What Happened When LinkedIn Learning was Embedded into a Classroom. LinkedIn. Retrieved July 12, 2022, from <https://www.linkedin.com/business/learning/blog/higher-education/what-happened-when-linkedin-learning-was-embedded-into-a-classroom?src=li-other&veh=business.linkedin.com%7Cli-other>
- Forta, B. (2013). *SAMS Teach Yourself SQL in 10 Minutes* (Fourth ed.). Indianapolis, IN, USA: Pearson Education, Inc.
- Petrone, P. (2021, March 16). How this University Drove Such Strong Adoption of LinkedIn Learning. (LinkedIn, Ed.) Retrieved July 12, 2022, from <https://www.linkedin.com/business/learning/blog/higher-education/susquehanna-university-linkedin-learning>
- Frye, C. (2021, June). Tableau Essential Training. LinkedIn Learning. Retrieved September 1, 2022, from <https://www.linkedin.com/learning/tableau-essential-training-2021?u=86746674>
- Garrison, D., & Vaughan, N. (2008). *Blended learning in higher education*. San Francisco, CA: Jossey-Bass.
- Rasheed, R. A., Kamsin, A., & Abdullah, N. A. (2020). Challenges in the Online Component of Blended Learning: A Systematic Review. *Computes & Education*, 144, 103701. Retrieved June 26, 2022
- Jeffery, L. M., Milne, J., Suddaby, G., & Higgins, A. (2014). Blended Learning: How Teachers Balance the Blend of Online and Classroom Components. *Journal of Information Technology Education*, 121-140. Retrieved June 26, 2022
- Santa Maria, A. (2022, January 10). How to Use Skills to Build Career Paths for Your Entire Workforce. Retrieved July 8, 2022
- LinkedIn Learning. (2022, May). *Add Learning Certificates of Completion and Skills to Your LinkedIn Profile*. Retrieved July 8, 2022, from LinkedIn.com: <https://www.linkedin.com/help/linkedin/answer/a704787/adding-learning-certificates-of-completion-and-skills-to-your-linkedin-profile?lang=en>
- Shander, B. (2022, June 14). Data Visualization: A Lesson and Listen Series. Retrieved July 8, 2022
- Long, T., Cummins, J., & Waugh, M. (2017). Use of the flipped classroom instructional model in higher education: instructors' perspectives. *Journal of Computing in Higher Education*, 29, 179-200. Retrieved June 26, 2022
- Spanjers, I. A., Konings, K. D., Leppink, J., Verstegen, D. M., de Jong, N., Czabanowska, K., & van Merrienboer, J. J. (2015). The Promised Land of Blended Learning: Quizzes as a Moderator. *Educational Research Review*, 15, pp. 59-74.
- Ludwig, D. (2018, February 16). Excel: PivotTables for Beginners. LinkedIn Learning. Retrieved from <https://www.linkedin.com/learning/excel-pivottables-for-beginners?contextUrn=urn%3Ali%3Alearning>
- Susquehanna University. (2021-2022). *Course Catalog*. Retrieved from Susquehanna University: <https://www.susqu.edu/live/files/639-coursecatalog2021-22-3pdf>
- Taylor, D. (2019, January 17). Excel: Introduction to Charts and Graphs. Retrieved July 8, 2022

Taylor, D. (2020, October 29). Excel Essential Training (Office 365/Microsoft 365). LinkedIn Learning. Retrieved from <https://www.linkedin.com/learning/excel-essential-training-office-365-microsoft-365?contextUrn=urn%3Ali%3AlearningCollection%3A6739922207503355904&u=86746674>

Thouin, J. (2021, January 12). SQL for Non-Programmers. LinkedIn Learning. Retrieved from <https://www.linkedin.com/learning/sql-for-non-programmers?contextUrn=urn%3Ali%3AlearningCollection%3A6734173490636722179&u=86746674>

APPENDIX ONE

Takeaways—LinkedIn Learning—Project Management Foundations

<https://www.linkedin.com/learning/project-management-foundations-4/deliver-successful-projects-3?u=86746674>

Get to Know Project Management

1. Define project.
2. Describe the three key elements in the definition of the project.
3. Describe a budget
4. Describe how a project differs from operations.
5. Describe/define project management
6. Describe the five questions of project management.
7. Describe the five skills need by a project manager. Which is most important and why?
8. Describe the five stages of the project management life cycle and what is accomplished in each stage.
9. Describe the agile project management approach and how it differs from the traditional (waterfall) approach.
10. Describe organizational culture. Describe how it impacts projects.

First Things First

1. Describe/define project initiation.
2. Describe what happens in the initiation phase.
3. Describe the project charter
4. Describe a stakeholder and the different types of stakeholders.
5. Describe what you need to know about each of the stakeholders and why that information is important.
6. Describe a project goal.
7. Why should you try to develop a problem statement? Describe the importance of asking why?
8. Describe the relationship between the problem statement and the project goal.
9. Describe what the acronym SMART means in terms of objectives.
10. Describe what project objectives provide.
11. Describe how to choose a strategy and the role brainstorming plays.
12. Describe requirements and the need for getting them complete and correct.
13. Describe the four techniques for gathering requirements.
14. Describe what project requirements provide.
15. Describe project deliverables and their role in the project.
16. Describe success criteria.
17. Describe the primary concern with assumptions and risks
18. Describe project scope, the problems of scope creep and a scope statement.
19. Describe the contents of a scope statement.
20. Describe the three possible outcomes of a project review.
21. Describe the project charter and its contents.

Develop a Project Plan

1. Describe the benefits of creating a Work Breakdown Schedule (WBS).
2. Describe the two types of tasks in a WBS.
3. Describe who is involved in building a WBS.
4. Describe the initial inputs to the WBS.
5. Describe a responsibility matrix. Define A,R,I, and C.
6. Describe the role of the responsibility matrix in communicating with stakeholders.
7. Describe the project organizational chart and its role.
8. Describe the skills matrix and its function.
9. Describe the steps in schedule development.

10. Describe a risk management plan.
11. Describe the purpose of a quality management plan.
12. Describe the purpose of change management plan.
13. Describe the three steps in the change management process.
14. Describe a change request log.

Build a Project Schedule

1. Describe a network diagram.
2. Describe a task dependency. Describe the four types of dependencies.
3. Describe a milestone. Describe what milestones indicate in terms of the project.
4. Describe/define the critical path. Why is the critical path important?
5. Describe the critical chain method for scheduling. Describe the focus of the critical chain method.
6. Describe baseline documents. Describe the purpose of the baseline documents.
7. Describe the relationship between the baseline documents and the change management process.

APPENDIX TWO

Topic Examination Data Visualization

SOURCE:

Shander, B. (2019). *Data Visualization: A Lesson and Listen Series*. LinkedIn Learning. <https://www.linkedin.com/learning/data-visualization-a-lesson-and-listen-series/introduction?autoAdvance=true&autoSkip=false&autoplay=true&contextUrn=urn%3Ali%3AlearningCollection%3A6734172245767270400&resume=false&u=86746674>.

OBJECTIVE:

Data visualization is a multiple faceted subject area. There are many different aspects of data visualization that one can explore, and, through this multiple part assignment, you will do so.

ACTIONS:

You will review and produce a short report on at least three different lessons from this LinkedIn Learning course. You can choose any of the lessons for this assignment, except the data literacy lesson, which we already examined at the beginning of the semester.

DELIVERABLE:

Once you have viewed each section you will create a report which includes:

1. (5 Pts) A 50–100-word introduction of the topic and why you selected it.
2. (15 Pts) A 300–450-word summary of the lesson section. You should highlight the key aspects (at least two) of the lesson and note their importance related to the topic.
3. (20 Pts) A 450–600-word summary of the listen section. You should (1) give a brief bio of the speaker and (2) highlight the key aspects (at least two) of the discussion and note their importance related to the topic.
4. (10 Pts) A brief (150–300-word) discussion on at least two “thoughtful” lessons that you learned from the reviewing this lesson.
5. (10 Pts) Grammar, spelling, punctuation, sentence structure, and paragraph formation.

SUBMISSIONS:

As stated above you will develop three deliverables. The due dates for these deliverables are Friday, March 4, 2022, Friday, April 1, 2022, and Friday, April 29, 2022.

An extra credit deliverable, **due 5/9/2022**, will be available to any student who did not perform well on any of the three previous deliverables; received a grade less than 80% on at least of the previous deliverables. To qualify for this extra credit opportunity, the student must gain permission to attempt the extra credit. Extra credit will not be granted if the student missed any of the three previous deliverables or plagiarized any of the previous deliverables. If a student does the extra credit the three highest scores for this assignment will be used in the computation of the final grade. The requirements for the extra credit will be the same as the first three deliverables.

GRADING:

Each deliverable will count for 60 points.

An Examination of Tableau as a Supplement to Excel to Enhance Data Literacy Skills

Mark P. Sena
sena@xavier.edu

Thilini Ariyachandra
ariyachandrat@xavier.edu

Business Analytics and Information Systems Department
Xavier University
Cincinnati, OH USA

Abstract

Data literacy has become a much sought after skill by organizations as the importance of data driven decision making continues to rise in importance. This study explores the teaching of Tableau as a supplement to Excel to enhance the data literacy skills of students. Twenty-eight students taking an introductory course in Business Analytics and Information Systems completed a series of five tasks in both Excel and Tableau and answered a two question survey on their perceptions of the relative ease of use and ability to produce correct and visually pleasing results for Tableau versus Excel. An assessment of the results showed that Tableau submissions produced higher scores in terms of correctness and more professionally formatted visualizations. Student survey responses also supported a significant preference for Tableau over Excel in terms of both ease of use and ability to produce correct results.

Keywords: Tableau, Excel, Data Literacy, Analytics, Data Visualization.

1. INTRODUCTION

Recent disruptive changes in the global scape have led organizations to realize the value of data driven decision making. Information is now a necessary ingredient to become nimble and thrive in disruptive business environments. In over 90 percent of organizations, information is explicitly recognized as a critical enterprise asset in corporate strategies (Rollings et al., 2022). Yet its potential remains untapped as an Accenture survey of over 9000 employees revealed that 75 percent are underprepared for the growing use of data in their roles (Vohra, 2020). Only 21 percent of survey respondents were confident in their data literacy skills. According to Gartner, one of the two main road blocks to the creation of a data driven business environment is poor data literacy within an organization (Boobier, 2021). As a

result of these gaps in data literacy, business leaders become blind to potential business opportunities as well as potential threats and weaknesses (Goasduff, 2020).

Today, data literacy is seen as a critical skill for those entering the workforce. A survey of 1200 global C-level executives and 6000 employees indicates that data literacy will be the most in-demand skill by 2030 (Qlik Technologies, Inc., 2022). Over 85 percent of executive respondents believed that data literacy will be vital to the future, akin to the ability to use a computer today. A recent Forrester survey further confirm evidence of the growing demand for these skills as survey respondents indicated a 40 percent rise in the heavy use of data by employees in the last four years (Blackborow, 2022). The demand for data literacy skills by employers suggests a shift

in the hiring requirements as well as a shift in the existing role requirements in organizations.

Consequently, institutions of higher education need to incorporate data literacy skills into their core curriculum to meet the needs of potential employers. Heeding the call for greater data literacy skills in future employees, recently academic institutions have begun using data visualization course work as a means to infuse data literacy into the curriculum (Batt et al., 2020; Schuff, 2018; Usova & Laws, 2021). However, it has not been a fluid integration (Hunt, 2005; Philip, Olivares-Pasillas & Rocha 2016; Wolff et al., 2016). The most common IT application used to teach data literacy skills is Microsoft Excel (Slayter & Higgins 2018).

The introductory IS course found in the typical undergraduate business curriculum generally contains a sizable share of data skills content taught to students using Microsoft Excel (Modaresnezhad & Schell 2019). More recently, an introduction to data visualization applications, such as Tableau and Power BI, have begun to emerge in the content presented in introductory IS courses. A tool that is used for the sole purpose of visualizing data such as Tableau offers an easy means of supplementing the core data skills students learn through MS Excel. This paper explores the teaching of Tableau as a supplement to Excel to enhance the data literacy skills of students. In so doing, the authors present options for instructors hoping to incorporate content to bolster the data literacy skills currently being taught in Introductory IS courses.

To do so, the paper first introduces the concepts of data literacy and its importance. Next, it describes the methodology used to explore the usefulness of Tableau as a supplement to MS Excel skills taught in the introductory IS course. This is followed by a detailed discussion of the study's findings. Finally, the conclusion of the study is presented along with limitations and future research options.

2. DATA LITERACY

While a relatively recent addition to the growing body of literacies (e.g., IT literacy and statistical literacy), the definition of data literacy is evolving as the extant universe of data and its characteristics keep growing. Similar to other types of literacy, a simple definition for data literacy is the ability to understand and use data (Frank et al., 2016). D'Ignazio and Bhargava (2015), express data literacy in terms of the ability to carry out four specific actions related to

data; specifically, reading data, working with data, analyzing data and arguing with data. They describe each of these actions as follows:

- "1) Reading data, which means understanding what data is and the aspects of the world it represents.
- 2) Working with data involves creating, acquiring, cleaning, and managing it.
- 3) Analyzing data involves filtering, sorting, aggregating, comparing, and performing other analytic operations on it.
- 4) Arguing with data involves using data to support a larger narrative that is intended to communicate some message or story to a particular audience." (D'Ignazio & Bhargava, 2015, p. 2).

Within the context of an organization, data literacy has been defined as the ability of employees to understand and work with data to the appropriate degree (Brown 2021). Tableau provides a more comprehensive description of data literacy. They define data literacy as "the ability to explore, understand and communicate with data using critical thinking skills to use, interpret and make decisions; then, convey its value and significance to others" (Tableau, 2022). While assumed in the Tableau definition, recent data literacy definitions make specific reference to data visualization. For example, Shreiner and Dykes (2021) state it as the ability to comprehend, analyze, and interpret data and data visualizations. Furthermore, an individual can achieve data literacy by gaining verbal, numerical and graphical literacy (Brown, 2021). Graphical literacy, the understanding of how to decode and interpret visual representations of meaning or data visualization (Roberts & Brugar, 2017) is a key means by which individuals engage in data literacy today.

The specific inclusion of graphical literacy as part of data literacy suggests that the introduction and use of data visualization applications such as Tableau in an introductory IS course can enhance the content currently being taught in the course. Next, the methodology section describes the process used to empirically examine the value of a data visualization tool to enhance data literacy in an introductory IS course.

3. METHODOLOGY

In order to explore the relative effectiveness of teaching Tableau as a compliment to Excel, two sections enrolled in an introductory undergraduate course in Business Analytics and Information Systems were asked to complete an ungraded test with identical questions using both

Excel and Tableau. Students were also asked to complete a two-question survey asking their opinion on the ease of use and effectiveness of using Excel vs Tableau for the test questions. The purpose of the assessment was to collect data for this study while providing students with an opportunity to test their skills in Tableau while relating them to skills previously learned in Excel.

The course is a 200-level class that is required of all business students and is comprised of mostly sophomores and juniors at a private university with approximately 6000 students located in the Midwest of the United States. All students take the course as a requirement of their major. Students are comprised of various business majors along with selected other majors (particularly health and sport management) throughout the university. Prior to the test, students had completed approximately seven weeks of Excel training and approximately two weeks of Tableau training, with the Tableau training coming immediately prior to the test. A major objective of the course is to acquire skills in Excel that are useful across the various majors of enrolled students. Most students have limited exposure to Excel prior to enrollment. After learning the basics and intermediate Excel skills, students are given approximately two weeks of training on data analysis skills (such as tables, pivot tables, and correlation functions). This is followed by approximately two weeks of Tableau training that covers similar content. As a result, students are prepared as equally as possible for the data literacy tasks measured in this study.

The test asked students to download an Excel file that is available on Kaggle (Kaggle, 2022). The file is a widely used training dataset with fields that include the gas mileage per gallon of various automobile models by year and number of engine cylinders. The questions included the following:

1. Create a chart that shows the average mpg by year grouped in 5-year intervals.
2. Create a chart that shows each car's mpg and cylinders in a scatter plot. Label the model.
3. Create a chart that shows the average mpg by Japanese automakers vs others. These include Datsun, Honda, Toyota, Subaru, and Mazda.
4. Show a list of the top ten cars with the highest mpg.
5. Show a list of models that have at least 6 cylinders and at least 30 mpg.

In the prior section on data literacy, we summarized four major skills of data literacy: reading data, working with data, analyzing data and arguing with data (D'Ignazio & Bhargava,

2015). In this study, our research questions focus primarily on the relative ability of students to use Tableau and Excel to perform the third set of skills. Because the sample data file was provided to students and did not require any manipulation or cleansing, the first two categories of data literacy were not measured in this study. Due to time constraints, students were not asked to write a narrative explaining their findings. Specifically, our study focuses on the following research questions:

Research Question 1: To what extent are students able to use Excel vs Tableau for filtering, sorting, aggregating, comparing, and performing other analytic operations on a set of five problems?

Research Question 2: To what extent are students able to use Excel vs Tableau to communicate the results of their analysis of the five problems to a particular audience in a clear and professionally formatted presentation?

Each test question was graded on a five-point scale for correctness with 5 representing completely correct and 1 representing completely incorrect or not attempted. The questions were also graded for clarity and professionalism of appearance with 5 representing very clear and professional in appearance and 1 representing very unclear or unprofessional in formatting or appearance. The second set of evaluation is intended to measure the extent to which Tableau allows for increased ease in formatting and creating visually pleasing visualizations.

In one course section, students completed the test using Excel first while the other section completed the test using Tableau before Excel. Students took the test in proctored classroom and were required to work individually but were allowed "open book" access to the Internet, prior training materials, etc. However, the test was timed and limited to fifty minutes. Twenty eight students successfully completed both sections of the test and were included in this analysis. There were a few students who failed to submit both sections or had submissions errors (did not attach the correct file) and were excluded from the dataset.

After completing the tasks in both Excel and Tableau, students were required to upload the files then were asked to complete the following two survey questions:

- Please select the response which best fits your opinion as to which tool was easier to use:

1. Tableau was much easier to use
 2. Tableau was somewhat easier to use
 3. Tableau and Excel were equally easy to use
 4. Excel was somewhat easier to use
 5. Excel was much easier to use
- Please select the response which best fits your opinion as to which tool produced the better results (correct answer, visually pleasing, etc.):
 1. Tableau produced much better results
 2. Tableau produced somewhat better results
 3. Tableau and Excel produced equal results
 4. Excel produced somewhat better results
 5. Excel produced much better results.

4. FINDINGS

1. *Create a chart that shows the average mpg by year grouped in 5-year intervals.*

In order to complete this task in Excel, students would insert a pivot table then select the Year field as a row and the MPG field as the Sum Value field. The Value Field setting for MPG would need to be changed from the default of Sum to Average. The Year field would need to be grouped in fixed intervals of five years. In Tableau (after loading the Excel data source), students would need to create a "Bin" using the Year field to allow for five year sets of groups. Afterwards, students would drag the two fields, Year Bin and MPG onto the row and column sections of the screen then would need to adjust the Measure setting of MPG to Average.

Question: Avg MPG by Year	
Average of Excel Correct	4.68
Average of Tableau Correct	4.86
P(T<=t) two-tail	0.31
Average of Excel Clarity	4.04
Average of Tableau Clarity	4.39
P(T<=t) two-tail	0.12

Table 1: Question 1 Results

As shown in Table 1, students generally performed well in both Excel and Tableau for this question. The moderately lower scores for Excel clarity are due to the inherent pivot table formatting where numbers need to be formatted, headings need to be replaced from generic ("Row Label") and chart headings need to be added. In Tableau, fields created using Bins default to

showing just the first entry in a grouped range (such as 1970 instead of 1970-1974) but otherwise Tableau creates visually appealing charts more easily than Excel pivot tables.

2. *Create a chart that shows each car's mpg and cylinders in a scatterplot. Label the model.*

In order to create the scatterplot in Excel, students simply needed to select the ranges for both MPG and Cylinders and select the chart option to insert a Scatterplot. Creating labels for the car model would require advanced skills to allow for models to show only when hovered (otherwise the labels would make the scatterplot difficult to read). In Tableau, students would drag to two fields onto the Row and Column but would need to unselect Aggregate Measures from the ribbon to allow for the plotting of individual records instead of summary values (the default would plot just one point). Students would also need to drag the Model field on to the Label option in the Marks section of the screen.

As shown in Table 2, students performed much better using Tableau for this task than using Excel. Many students assumed that the task should be completed using a pivot table (providing average MPG by cylinder), which does not produce the required result.

Question: MPG vs Cylinders Scatterplot	
Average of Excel Correct	2.46
Average of Tableau Correct	4.32
P(T<=t) two-tail	<.001
Average of Excel Clarity	2.36
Average of Tableau Clarity	4.29
P(T<=t) two-tail	<.001

Table 2: Question 2 Results

3. *Create a chart that shows the average mpg by Japanese automakers vs others. These include Datsun, Honda, Toyota, Suburu, and Mazda*

Similar to question 1, this task required students to create a pivot table in Excel with the Auto Maker field in the row and MPG (with a setting of Avg) as the Sum Value field. Students then needed to manually group the Auto Maker range and format the label of the two groups. The challenge with this grouping is that the required records were not in adjacent rows so students would need to hold the Control key (or Command key in Mac) to select multiple entries. In Tableau,

students would begin by creating a group field from the existing Auto Maker field. However, the grouping is conducted in a pop-up window that allows for easy selection of group entries and labels. The remaining task would simply be to drag the grouped Auto Maker field and MPG onto the screen in the row/column and again change MPG measure from sum to average.

As shown in Table 3, students struggled with the task of selecting the required groupings in Excel resulting in much lower scores for both correctness and clarity.

Question: Avg MPG for Japanese automakers vs. other automakers	
Average of Excel Correct	3.21
Average of Tableau Correct	4.46
P(T<=t) two-tail	<.001
Average of Excel Clarity	2.75
Average of Tableau Clarity	4.75
P(T<=t) two-tail	<.001

Table 3: Question 3 Results

4. *Show a list of the top ten cars with the highest mpg.*

In Excel, the preferred method of completing this task would be to insert a table then apply a top ten filter to the MPG field. Students could also simply create a copy of the raw data and apply a sort and manually select the first ten rows. Students often failed to include a heading for the audience. In Tableau, students would drag the Model and MPG (with measure of Avg) onto the row and column, then drag the Model into the Filter section of the screen and select the Top Ten option (which defaults to top ten largest MPG average).

Question: Top Ten Models with Highest Average MPG	
Average of Excel Correct	4.25
Average of Tableau Correct	4.25
P(T<=t) two-tail	1.00
Average of Excel Clarity	3.71
Average of Tableau Clarity	4.57
P(T<=t) two-tail	.013

Table 4: Question 4 Results

As shown in Table 4, students performed equally well in getting the correct solution (with the sort option counted as an alternative correct result)

but with less clarity as compared to the Tableau version. One modest formatting advantage in Tableau is that worksheet names are automatically used as chart headings so the audience can more clearly understand the result.

5. *Show a list of models that have at least 6 cylinders and at least 30 mpg.*

This was another task that could be completed in various ways. In Excel, ideally students would insert a table and apply numeric filters to limit the entries that met the requirements. Students could also apply a two level sort and manually select the records. Some students created a new column using either a nested IF or IF(AND) function. As in question 4, students often failed to label or include a heading indicating the purpose of the analysis. In Tableau, there were also multiple ways of achieving the result. The simplest approach was to apply two filters by dragging the MPG and Cylinder fields into the Filter area of the screen then applying the minimum filter options. Some students chose to create a new "Calculated" field where students enter a formula into a popup window. In this option, students would also need to drag the calculated field into the filter and limit the entries to "True".

As shown in Table 5, students struggled with this question in both tools, particularly in the clarity of the presentation. It is important to note that because the test was timed, some students may have had limited time to complete the last question.

Question: List of Models with 30+ MGP and 6+ Cylinders	
Average of Excel Correct	2.68
Average of Tableau Correct	3.36
P(T<=t) two-tail	0.07
Average of Excel Clarity	2.18
Average of Tableau Clarity	2.86
P(T<=t) two-tail	0.04

Table 5: Question 5 Results

Table 6 shows that, overall students produced more correct results using Tableau to complete the tasks and had substantially better clarity in terms of formatting, labels and professional looking visualizations. Students who completed the questions in Tableau prior to Excel actually had higher overall means for correctness in both Tableau and Excel. This may have been due to time constraints because the Excel solutions were

less structured and the recent training in Tableau may have allowed those who completed Tableau first to have more time remaining. It may also be a function of small sample sizes given the relatively small number (28) of students who completed both sections of the test.

	Excel	Tableau	Excel	Tableau
	Correct Solution		Clarity	
Overall Avg (n=28)	3.46	4.25 *	3.01	4.17
Excel First (n=18)	3.34	4.12	3.03	4.17
Tableau First (n=10)	3.66	4.48	2.96	4.18

Table 6: Overall Averages and Order of Tasks

* $P(T \leq t)$ two-tail < .001 for all comparisons

After completing the ungraded test, students were asked their opinion as to which tool produced the better result and which tool was easier to use. As shown in Figure 1, the majority of students found Tableau produced much better results and was much easier to use than Excel for the specific questions in the test.

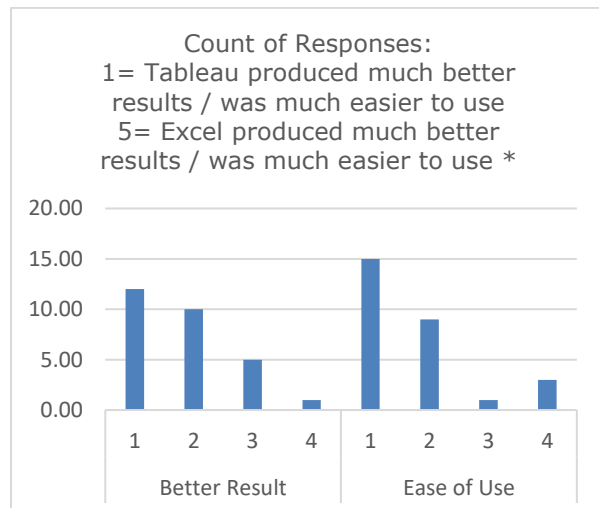


Figure 1: Student Opinions: Tableau vs Excel

* Note: no students selected "5" as a response

It is important to note that students were preparing for a graded skills test in Tableau and the recent training on Tableau could have resulted in some halo effect in the opinions. The questions in the test were appropriate for Tableau analysis. Clearly Tableau is not a replacement for all Excel features as Tableau is not designed for the wide spectrum of Excel uses. For example, students could not easily use Tableau to create a

budget or compute payments on a mortgage, etc., Thus, the findings should be limited to data literacy as they pertain to tasks that can be accomplished easily in Tableau.

5. CONCLUSIONS

As organizations strive to increase data literacy skills among employees, universities play an important role in providing students with a foundation. Software such as Tableau (along with similar tools such as Microsoft's Power BI) can play an important role in expanding the capability of students to manage, cleanse, interpret, and present data in a visual manner that students find intuitive and engaging. Similarly, Excel, which has long been the key application in introducing students to data literacy skills, remains an important component in business analytics education.

This study provides evidence in support of using Tableau for both research questions examined in the study. The results support the enhanced ability of using Tableau for filtering, sorting, aggregating, comparing, and performing other analytic operations and to better communicate the results of their analyses in a clear and professionally formatted presentation.

In addition to providing rationale for inclusion of data visualization tools in business analytics coursework, this study provides a foundation for future studies that compare data literacy tasks in an experimental setting. Future studies can augment the skill sets tested, the software used for comparisons, and test the findings against larger sample sizes.

Clearly, the implications of this study are limited by several factors. Our sample size of only 28 students on a small set of five problems provides only preliminary evidence. With problems that are unstructured and result in students providing various solutions, grading on a five point scale as we did in this study is inherently subjective. Our conclusions are also influenced by the specific questions chosen, the prior experience of the students and the manner in which the test was administered. Future studies could focus on establishing a more robust set of questions with more specific grading rubrics and a larger sample size. Particularly for the assessment of clear and professionally formatted presentations, a more descriptive set of criteria is needed to add rigor to the understanding of the importance of this aspect of data literacy.

The results of the study provide some initial evidence that could form the foundation for future studies in the application of data visualization tools for developing data literacy skills. The analysis of the results is also insightful to software makers to show examples of the challenges that users face in solving the sample questions in the test. Additional academic and pedagogical research should focus on the benefits and success factors of using data visualization software to enhance the data literacy of students and employees.

6. REFERENCES

- Batt, S., Grealis, T., Harmon, O., & Tomolonis, P. (2020). Learning Tableau: A data visualization tool. *The Journal of Economic Education*, 51(3-4), 317-328.
- Blackborow, J., (2022). Building Data Literacy. Forrester Consulting. https://www.tableau.com/sites/default/files/2022-03/Forrester_Building_Data_Literacy_Tableau_Mar2022.pdf
- Brown, S. (2021, February 9). How to build data literacy in your company. MIT Sloan. Retrieved July 12, 2022, from <https://mitsloan.mit.edu/ideas-made-to-matter/how-to-build-data-literacy-your-company>
- Boobier, T. (2021, July 7). Perhaps Execs don't need data literacy after all, just good judgement. Customer Insight Leader. Retrieved July 12, 2022, from <https://www.customerinsightleader.com/other/execs-dont-need-data-literacy/>
- D'Ignazio, C., & Bhargava, R. (2015, September). Approaches to building big data literacy. Proceedings of the Bloomberg Data for Good Exchange Conference, New York, NY, United States.
- 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).
- Goasduff, L. (2020, July 9). Avoid 5 Pitfalls When Building Data and Analytics Teams. Gartner. Retrieved July 12, 2022, from <https://www.gartner.com/smarterwithgartner/avoid-5-pitfalls-when-building-data-and-analytics-teams>
- Hunt, K. (2005). The challenges of integrating data literacy into the curriculum in an undergraduate institution. *IASSIST Quarterly*, 28(2-3), 12-12.
- Kaggle (2022). Auto-mpg dataset. Retrieved July 14, 2022 from <https://www.kaggle.com/datasets/uciml/autompkg-dataset>
- Modaresnezhad, M., & Schell, G. (2019). The soul of the introductory information systems course. *Information Systems Education Journal*, 17(5), 39.
- Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming racially literate about data and data-literate about race: Data visualizations in the classroom as a site of racial-ideological micro-contestations. *Cognition and Instruction*, 34(4), 361-388.
- QLIK Technologies, Inc. (2022, March 22). Data Literacy to be Most In-Demand Skill by 2030 as AI Transforms Global Workplaces. GlobeNewswire News Room. Retrieved July 12, 2022, from <https://www.globenewswire.com/en/news-release/2022/03/22/2407543/0/en/Data-Literacy-to-be-Most-In-Demand-Skill-by-2030-as-AI-Transforms-Global-Workplaces.html>
- Roberts, K. L., & Brugar, K. A. (2017). The view from here: Emergence of graphical literacy. *Reading Psychology*, 38(8), 733-777.
- Rollings, M., Duncan, A., & Logan, V. (2019). 10 Ways CDOs Can Succeed in Forging a Data-Driven Organization. Gartner (G00378249).
- Schuff, D. (2018). Data science for all: a university-wide course in data literacy. In *Analytics and data science* (pp. 281-297). Springer, Cham.
- Shreiner, T. L., & Dykes, B. M. (2021). Visualizing the teaching of data visualizations in social studies: A study of teachers' data literacy practices, beliefs, and knowledge. *Theory & Research in Social Education*, 49(2), 262-306.
- Slayter, E., & Higgins, L. M. (2018). Hands-On Learning: A Problem-Based Approach to Teaching Microsoft Excel. *College Teaching*, 66(1), 31-33.
- Tableau. (n.d.). Data Literacy. Retrieved July 13, 2022, from <https://www.tableau.com/why-tableau/data-literacy>
- Usova, T., & Laws, R. (2021). Teaching a one-credit course on data literacy and data visualisation. *Journal of Information Literacy*, 15(1).
- Vohra, S. (2020, January 16). The human impact

of data literacy. Accenture.
<https://www.accenture.com/us-en/insights/technology/human-impact-data-literacy>.

Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*, 12(3)

Teaching Case

Using Supervised Machine Learning and CRISP-DM to Predict an Acquittal Verdict

Frank Lee

flee@gsu.edu

J. Mack Robinson College of Business
Georgia State University
Atlanta, Georgia 30303, USA

Clinton Baxter

cjbaxter1989@gmail.com

American Tire Distributors
Huntersville, NC 28078

Abstract

The Cross Industry Standard Process for Data Mining (CRISP-DM) framework was developed in the 1990s and has been widely used as the most relevant and comprehensive leading principle for conducting analytics projects. Despite the wide acceptance and adoption of the CRISP-DM framework, the current business analytics discipline often focuses on the modeling phase and overlooks the interplays between the phases. Consequently, students often lack a comprehensive understanding of the entire analytics process. This teaching case is created to demonstrate the importance of the data analytics life cycle and how six phases collectively contribute to the success of analytics projects using R. This case collects real-life data and follows the six CRISP-DM phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. At the end of the project, students will learn the importance of the data analytics life cycle, especially the data understanding and preparation phases, which often receive minimal attention in business analytics projects. This project will also demonstrate the importance of storytelling, ensuring that critical insights are conveyed to the audience.

Keywords: Data Mining, Legal Analytics, CRISP-DM, Analytics Project, R

1. INTRODUCTION

Under the United States Justice system and the Constitution, defendants are afforded indestructible rights such as the right to remain silent, counsel, and speedy and public trial by an impartial jury. If these rights are breached or violated, this could trigger a cascading of events—generally in the form of appeals—that could result in the original verdict being overturned and the case being dismissed or retried. In a perfect system, these indelible rights are always upheld under the U.S. Constitution, and for the

remainder of this paper, you will assume they are and have been. Although indestructible, these rights are not binding, meaning defendants can choose to waive their rights at any time. For example, defendants may waive their right to remain silent and testify in a court of law or waive their right to a jury trial and plead the case. This project is designed to delve deeper into those cases that go to trial and their associated outcomes; guilty or not guilty (acquittal).

Your first mission is to understand whether surface-level variables like the age, sex, ethnicity

of the defendant (and plaintiff), and other variables play a significant role in a jury's verdict of guilty or not guilty. Electing for a jury trial can be very time-consuming, stressful (for all parties), and unpredictable. You need to provide a probability of receiving an acquittal verdict to a defendant on trial for murder in Cobb County, Georgia.

The Cross Industry Standard Process for Data Mining

The Cross Industry Standard Process for Data Mining (CRISP-DM) framework was developed in the 1990s and has been widely used as the most relevant and comprehensive leading principle for carrying out analytics projects (Wirth & Hipp, 2000). CRISP-DM consists of six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment, with arrows indicating the most important and frequent dependencies between phases. The sequence of the phases is flexible and can be customized easily. Project work can occur in several phases simultaneously, and the movement can be either forward or backward between phases, as necessary.

Learning Objectives

By completing this assignment, you will be able to:

- Describe the data analytics project lifecycle and critical elements of each phase
- Obtain sufficient relevant data and conduct data analytics using scientific methods
- Apply appropriate and powerful connections between quantitative analysis and real-world problems
- Present descriptive statistics and models in a business context and employ appropriate data visualizations
- Apply advanced techniques to conduct a thorough and insightful analysis
- Interpret the results correctly with detailed and valuable information

2. CASE BACKGROUND

Case Text

Sarah Brown, a defendant in an upcoming murder case in Cobb County, Georgia, has enlisted the services of the local law firm Confident Cases LLC. Clint Baxter will be representing her as legal counsel throughout her case. As an astute lawyer, Clint understands that a defendant enjoys the presumption of innocence. At the same time, the onus relies on the State's prosecution to provide

evidence beyond a reasonable doubt that the defendant is guilty of the charges. The choice is always up to the defendant, who ultimately must make the decision regardless of Clint's recommendations. Clint wants his defendant to make the best suitable decision for her (and her family), as these decisions will have enormous life-changing impacts. Clint knows prior cases are public records, allowing him to data mine and collate specific data points into a working data set. Using his prior statistical knowledge, Clint wants to give Sarah the probability of beating the case (being acquitted of murder), so Sarah can make the most informed decision.

The Data Source

You will use data from the Cobb County Clerk of the Superior Court query to look up individual murder cases in 15 years from 2004 to 2019 in the Cobb County area. The Cobb County Clerk of the Superior Court allows you to filter the court cases by charge type using a specific date range. Once you load the query for murder cases from 2004 to 2019, you will look through each case's sentencing documents to determine if the defendant had pleaded out their case or chose a jury trial. If they went to a jury trial, you will include their case in the model as a data point and gather all the case variables. If they plead out, their information can be disregarded from the model.

Assumptions

Jury selection is part of the judicial process but is highly time-consuming and outside the scope of this analysis. The critical assumption in analytics projects is that the future will continue to be like the past. MacCoun (1989) used Bayesian methodology to research mock juries to uncover any innate biases each juror may have before their selection. The study concluded that it is difficult to predict human behavior on such a vast scale with many variables. Therefore, we will not be diving into each juror's prior disposition but rather assume each juror is a "rational" person. While we understand those variables could undoubtedly play a role in each outcome, we only want to look at known variables. In addition, you should also assume all constitutional rights have been upheld, so you will not be looking at any future data.

3. PROJECT ACTIVITY

Business Understanding

This phase focuses on understanding the intentions and requirements of the project from the business perspective and converting this knowledge into an analytics problem. This phase

also determines the aim of the project and designs the analytics plan. This phase aims to understand whether surface-level variables like the age, sex, ethnicity of the defendant (and plaintiff), and other variables play a significant role in a jury's verdict of guilty or not guilty. To accomplish the objective, you need to provide a probability of receiving an acquittal verdict to a defendant on trial.

Data Understanding

This phase involves an initial data collection and proceeds to activities that help the participants become familiar with the data. This project intends to use individual murder cases from the last ten years from the Cobb County Clerk of the Superior Court in Georgia. To find the relevant data, students first look through each case's sentencing documents to determine if the defendant pleaded out their case or went to a jury trial.

The following steps will help guide you along during your data collection. Please make sure to follow each of the steps in order.

1. Visit <https://ctsearch.cobbssuperiorcourtclerk.com/CaseType> and filter by case type "criminal" for murder cases from 2004 to 2019. (Appendix A, Figure 1).
2. Click on the paper icon located in the "view" column next to the defendant's name. (Appendix A, Figure 2).
3. Go to pleadings and search for "jury list" to identify if it is a jury trial case. Note: If you do not find a jury list, the case was more than likely pleaded out. Keep in mind that the jury list will never be available to protect each juror's anonymity. (Appendix A, Figure 3).
4. You can also obtain the prosecutor's name, case I.D., and the defendant's name at the top of the "Case Details." (Appendix A, Figure 4).
5. Click on the "Attorneys" tab to identify the attorneys the defendant retained. If the status shows active, you assume the attorney represented the defendant until trial completion. If it says released, they are not counted as an attorney for the defendant in your model. In the example provided, there are four. (Appendix A, Figure 5).
6. Look under the "defendants" tab to see how many codefendants there are. In the example

provided, there are none. (Appendix A, Figure 6).

7. Look under the pleading tab for "list of witnesses." This pleading pdf will be locked (to protect the identity of the witnesses involved), but if you see it listed, you know the case involved eyewitness testimonies. (Appendix A, Figure 7).
8. Look under the pleadings tab once again for the indictment pdf. Once opened, scroll through the indictment to determine how the murder was committed, when the murder took place, and the total number of victims. For example, you may find that a murder took place with one victim on 11/05/2003 by firearm. You'll need to categorize the murder methods by "Firearm," "Stabbing," or "Other." (Appendix A, Figure 8).
9. Go to the "offenses" tab to identify how many charges were brought against the defendant(s). (Appendix A, Figure 9).
10. You'll need to obtain the verdict handed down by the jury. This can be found in the "verdict" document under the "pleadings" tab or in the "sentence" document if the verdict document is sealed. Remember, if they were acquitted of other counts but guilty of even one count of murder, the case is considered a loss for the defendant. In this case, we will record Guilty as Class 0 and Not Guilty as Class 1 (Appendix A, Figure 10).
11. The last two variables needed are prior criminal convictions and age at the time of the murder. They are readily available public information.

Visit

<http://www.dcor.state.ga.us/GDC/OffenderQuery/jsp/OffQryForm.jsp>

Use the search to locate the convicted inmate. You'll also find the convict's DOB as well as any other prior convictions (calculate the age of the defendant during the trial by subtracting the sentencing date from their DOB). More digging might be necessary to identify the remaining variables if the defendant was proven not guilty. (Appendix A, Figure 11).

Data Preparation

This phase selects a subset of the data, performs data cleansing, and prepares the data for analysis. You are looking for completeness, consistency, and accuracy in the data. You must

ensure all columns were filled appropriately with their corresponding values and spot-checked any inconsistencies before loading into R. (Appendix A, Figure 12).

Before data preparation and modeling, you must provide a few aggregated/ summarized statistics. The summary statistics allow you to identify patterns while improving your understanding of the data. During your data collection, you gathered various attributes, including the sex of the defendant, the defendant's age, and whether the defendant had any prior criminal convictions.

First, find and visualize the distribution of the defendants' age in your dataset. You may create age buckets such as <18, 18-25, 26-35, 35-50, and 50+. Discuss the findings.

Second, find and visualize the distribution of the method of the murder. Discuss the findings.

Third, visualize and discuss the distribution of "Guilty" vs. "Not Guilty" between sex.

Finally, create a visualization that clearly shows the relationship between the "Age" of the defendant and the "Number of Charges." Do you notice a pattern or any relationship? Discuss your findings.

Modeling

This phase involves the selection and development of analytics techniques and models. In addition, portions of a data set are often set aside for training and validating the model(s). This teaching uses the programming language R for illustration, but all the analytics tasks can be similarly completed with any other software such as Python or RapidMiner.

Decision Trees models are quite popular supervised models for various reasons: they are easy to implement and interpret, and the complexity of a full tree can be optimized by incorporating pruning (Kucheryavskiy, 2018). You may start your analysis with a Decision Tree that uses the Classification and Regression Trees (CART) algorithm and move to an Ensemble methodology (Bagging and Boosting) which can help with the overfitting problem seen with single decision trees.

While decision trees are a viral classification and prediction technique, they are sensitive to changes in the data. One solution to these problems is to use ensemble tree models, which combine multiple single-tree models into an ensemble model. Three common strategies are

used when creating ensemble models: bagging, boosting, and random forest. Bagging uses the aggregation technique to create multiple subsets by repeatedly sampling the original data with replacement (Zhang et al., 2010). The boosting strategy also uses repeated sampling, but it creates a variety of training data sets by random sampling with replacement from the primary training dataset (Zounemat-Kermani et al., 2021). The sequence of models is trained in this method. Random forest is an extension of the bagging strategy and implements repeated sampling of the training data and a random selection of a subset of predictor variables (Breiman, 2001).

Classification Tree

To run the decision tree model in R, you need to preprocess data by converting your categorical data into factors using the **as.factor()** function. (See R Codes Snippets in Appendix B).

Next, partition the data into a train and validation set using a 60/40 ratio to create the default tree.

Create the default classification tree using the **repart** function.

Next, create a full tree that you can prune appropriately based on the cp (complexity parameter) results. Find the best pruned three with the least complexity. To identify the cp value associated with the smallest cross-validated classification error, use the **printcp** function to display the complexity parameter table.

	CP	nsplit	rel error	xerror	xstd
1	0.545455	0	1.000000	1.00000	0.26629
2	0.090909	1	0.454545	0.45455	0.19285
3	0.045455	3	0.272727	0.81818	0.24697
4	0.022727	7	0.090909	1.09091	0.27454
5	0.000000	11	0.000000	1.18182	0.28196

Figure 1 Complexity Parameter Table

Here you can see the best-pruned tree with the least complexity is the second one with the lowest **xerror** score of 0.45455, which is still the lowest when factoring in the **xstd** score (0.45455+0.19285 = 0.6474).

Next, run the prediction and create the confusion matrix as well as the Lift, Decile-wise, and ROC charts. Evaluate the model using accuracy, sensitivity, and specificity.

```

Reference
Prediction 0 1
0 22 3
1 4 4

Accuracy : 0.7879
95% CI : (0.6109, 0.9102)
No Information Rate : 0.7879
P-Value [Acc > NIR] : 0.5994

Kappa : 0.3969

McNemar's Test P-Value : 1.0000

Sensitivity : 0.5714
Specificity : 0.8462
Pos Pred Value : 0.5000
Neg Pred Value : 0.8800
Prevalence : 0.2121
Detection Rate : 0.1212
Detection Prevalence : 0.2424
Balanced Accuracy : 0.7088

'Positive' Class : 1

```

Figure 2 R CART Confusion Matrix

The model has a decent accuracy at 0.7879 and a quite reasonable specificity at 0.8462. However, the model lacks the ability to correctly classify the target class, which in our case is a verdict of Not Guilty with a subpar score of 0.5714.

You still want to understand more about the model's overall performance and finish this model by completing the Lift, Decile-wise, and ROC charts.

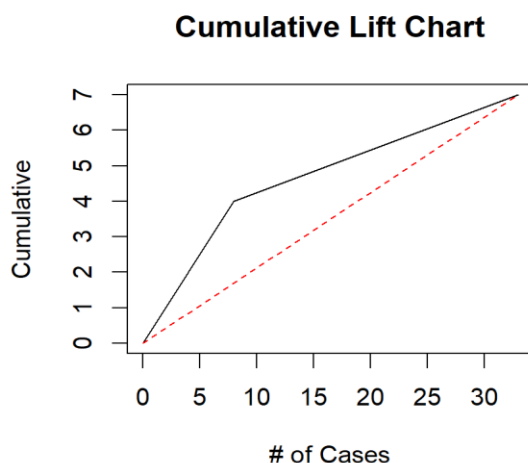


Figure 3 Cumulative Lift Chart

As you can see, though the model's sensitivity is not within our acceptable range, you know the model is better at predicting a Not Guilty verdict when compared to a random guess.

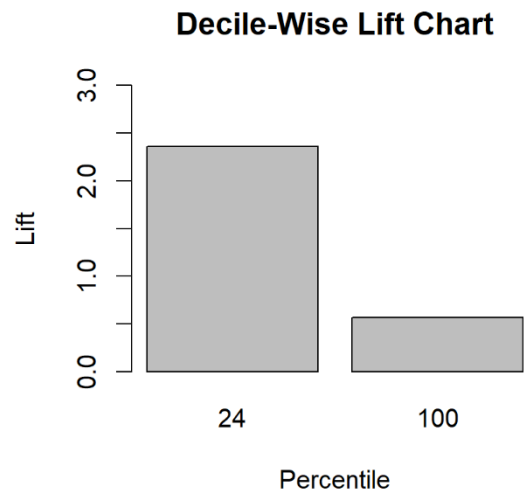


Figure 4 Decile-Wise Lift Chart

After reviewing the decile-wise lift chart, you can conclude that the model's top 24% of the observations contain 2.25 times as many Class 1 cases as the 24% of randomly selected observations.

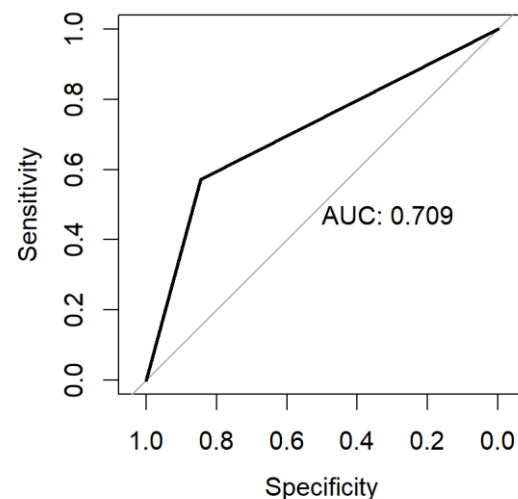


Figure 5 ROC Chart

You can see by viewing our ROC that while the sensitivity doesn't quite match up to what you want or expect, the specificity is still quite good. You can validate this by displaying the AUC score right around the 0.71 mark in this model's case.

Ensemble (Bagging)

We now want to know how well our ensemble models will perform. You will need to complete

the same minor preprocessing step as from our CART tree model.

Again, you will begin by splitting the data set into a train and validation set to maintain consistency across all models. You will use a 60/40 split. Once complete, run the model using the **randomForest()** function and specify the number of variables by setting the **mtry** option equal to 10—this tells the model to use a bagging strategy by using all ten predictor variables in the model.

While running the model, you also want to know how important each feature is to the model. Using the **varImpPlot()** function, you can visually identify which variables are essential for an average decrease in accuracy if they were omitted.

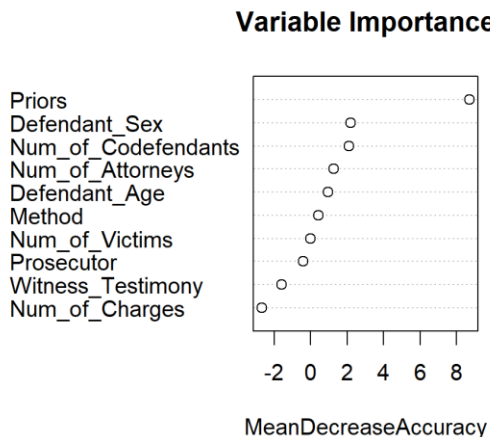


Figure 6 Variable Importance Chart

Not surprisingly, whether a defendant has any prior criminal convictions is extremely important to the model, meaning the model would suffer a tremendous decrease in accuracy if we dropped this variable. Conversely, we could drop the Num_of_charges variable and possibly notice a slight increase in accuracy—which makes sense as prior criminal history is often suppressed during a trial. After running the model, you want to view the confusion matrix, as shown below.

```

Reference
Prediction 0 1
0 20 2
1 6 5

Accuracy : 0.7576
95% CI : (0.5774, 0.8891)
No Information Rate : 0.7879
P-value [Acc > NIR] : 0.7462

Kappa : 0.4

Mcnemar's Test P-value : 0.2888

Sensitivity : 0.7143
Specificity : 0.7692
Pos Pred Value : 0.4545
Neg Pred Value : 0.9091
Prevalence : 0.2121
Detection Rate : 0.1515
Detection Prevalence : 0.3333
Balanced Accuracy : 0.7418

'Positive' Class : 1

```

Figure 7 R Bagging Confusion Matrix

This model delivers a much better sensitivity rating than the previous decision tree model. We do notice a slight tick down in the precision, meaning this bagging model might present clients with a false hope of beating the murder charges. This may occur as the model is generating false positives – classifying verdicts as Not Guilty that are, in reality, Guilty verdicts. You also should notice a degradation in the specificity meaning this model is not quite as good as classifying our non-target class (Guilty verdicts). Just as you did in the previous model, you will need to create cumulative lift, decile-wise, and ROC charts.

Ensemble (Boosting)

In your final model, you will use another ensemble method with a boosting strategy. You will again prepare the data using the same techniques as the previous model.

Setting **mfinal** equal to 100 tells the model to sample across multiple weak learner single trees repeatedly.

```

                Reference
Prediction 0 1
0 23 2
1 3 5

Accuracy : 0.8485
95% CI : (0.681, 0.9489)
No Information Rate : 0.7879
P-Value [Acc > NIR] : 0.2701

Kappa : 0.5692

McNemar's Test P-Value : 1.0000

Sensitivity : 0.7143
Specificity : 0.8846
Pos Pred Value : 0.6250
Neg Pred Value : 0.9200
Prevalence : 0.2121
Detection Rate : 0.1515
Detection Prevalence : 0.2424
Balanced Accuracy : 0.7995

'Positive' Class : 1

```

Figure 8 R Boosting Confusion Matrix

As you can see, the confusion matrix for the boosting model looks quite promising, excelling in each performance statistic well above the others. This model provides high accuracy, excellent sensitivity, precision, and specificity rates.

Evaluation and Deployment

This evaluation phase involves reviewing and interpreting the analysis results in the context of the business objectives and success criteria described in the first phase. Lastly, the deployment stage translates the knowledge gained from data analysis into a set of actionable recommendations.

4. PROJECT REPORT

You need to write a comprehensive project report. The project report should provide an executive

summary, introduction, data collection, data preparation, methodology, conclusion, reference, and appendix. Specifically, 1) after evaluating and running each model, you should be able to compare their results. 2) Your discussion should focus, in particular, on the results that are most interesting, surprising, or important. 3) Interpret the results with detailed and valuable information. It would be best if you also discussed the consequences or implications. 4) Finally, if the answers or findings are unexpected, see whether you can find an explanation for them, such as other factors that your analysis did not include.

2. REFERENCES

- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Kucheryavskiy, S. (2018). Analysis of NIR spectroscopic data using decision trees and their ensembles. *Journal of Analysis and Testing*, 2(3), 274-289.
- MacCoun, R. J. (1989). Experimental Research on Jury Decision-Making. *Science*, 244(4908), 1046-1050.
<http://www.jstor.org/stable/1703992>
- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*.
- Zhang, D., Zhou, X., Leung, S. C., & Zheng, J. (2010). Vertical bagging decision trees model for credit scoring. *Expert Systems with Applications*, 37(12), 7838-7843.
- Zounemat-Kermani, M., Batelaan, O., Fadaee, M., & Hinkelmann, R. (2021). Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology*, 598, 126266.

APPENDIX A
Case Figures

Cobb County Clerk of Superior Court
Connie Taylor

Index Data thru 09/07/2022

COURT RECORDS SEARCH BY CASE TYPE

Select either Civil or Criminal, then select the case/offense type. [? Help](#)


Search Civil or Criminal cases
☒ Civil ☐ Criminal

Case Type
MURDER

Date Range
Specific Date Range

[Back to Main Menu](#)

Figure 1. Cobb County Clerk of the Superior Court



Cobb

Index Data Through:09/07/2022

Prev Next Expand All Back to Main Menu Edit Search New Search Record


View	Defendants	Aliases	Offenses
	WALKER MACQUES ANTONIO [D]		MURDER

Figure 2. The view column next to the defendant's name

Print Case Details Close

Criminal Case Number: 04900203				Case Type: INDICTMENT			
STATE v MANZANO JESUS GUERRERO				Filing Date: 01/09/2004			
Judge: SCHUSTER				Prosecutor: EVANS,JESSE			




Defendants	Pleadings	Hearings	Attorneys	Offenses	Appeals	Bond Information	Sentences
 118	09/18/2009	12/16/2009	1	DISTRICT ATTORNEY	MOTION LIMINE		20090124601
 119	09/18/2009	12/16/2009	1	DISTRICT ATTORNEY	REQ TO CHARGE		20090124602
 120	09/18/2009	12/16/2009	1	COURT	JURY LIST		20090124603

Figure 3. Jury trial case

Print Case Details Close

Criminal Case Number: 04900203	Case Type: INDICTMENT
STATE v MANZANO JESUS GUERRERO	Filing Date: 01/09/2004
Judge: SCHUSTER	Prosecutor: EVANS,JESSE

Figure 4. Prosecutor Information

Print Case Details Close

Criminal Case Number: 04900203				Case Type: INDICTMENT			
STATE v MANZANO JESUS GUERRERO				Filing Date: 01/09/2004			
Judge: SCHUSTER				Prosecutor: EVANS,JESSE			

Defendants	Pleadings	Hearings	Attorneys	Offenses	Appeals	Bond Information	Sentences
Defendant #	Attorney Name	Stat	Attorney Address				
1	KEATON,REBECCA HULSEY	ACTIVE	615 ROSWELL STREET SUITE 400 MARIETTA,GA30060				
1	BERRY,JIMMY D	ACTIVE	236 WASHINGTON AVE MARIETTA,GA30060				
1	DURHAM,MITCH	ACTIVE	301 WASHINGTON AVENUE MARIETTA,GA30060				
1	IVAN,MICHAEL JOHN	ACTIVE	143 ATLANTA STREET MARIETTA,GA30066				

Figure 5. Attorney Information

Case Details									
Criminal Case Number: 04900203					Case Type: INDICTMENT				
STATE v MANZANO JESUS GUERRERO					Filing Date: 01/09/2004				
Judge: SCHUSTER					Prosecutor: EVANS, JESSE				
Defendants	Pleadings	Hearings	Attorneys	Offenses	Appeals	Bond Information	Sentences		
View	Def. #	Name	Address		OBTN	Warrant	DOB	SSN	Alias
	1	MANZANO JESUS GUERRERO	COBB COUNTY JAIL MARIETTA, GA 30060		108592643	03W0012069	N/A	N/A	N

Figure 6. Codefendants

Case Details									
Criminal Case Number: 04900203					Case Type: INDICTMENT				
STATE v MANZANO JESUS GUERRERO					Filing Date: 01/09/2004				
Judge: SCHUSTER					Prosecutor: EVANS, JESSE				
Defendants	Pleadings	Hearings	Attorneys	Offenses	Appeals	Bond Information	Sentences		
36	01/21/2005	01/24/2005	1	DEFENDANT	MOTION LIMINE		20050009766		
37	01/21/2005	01/24/2005	1	DEFENDANT	LIST OF WITNESSES		20050009768		

Figure 7. List of Witness

Pleadings By Defendant									
Criminal Case Number: 04900203					Defendant #1				
STATE v MANZANO JESUS GUERRERO					Defendant Name: MANZANO JESUS GUERRERO				
View	Pleading #	File Date	Add Date	Defendant	Filed By	Type	CCFN		
	1	01/09/2004	01/22/2004	1	DISTRICT ATTORNEY	INDICTMENT	20040003839		

2 / 3 100% + -

Christopher D. Weiner

Susan T. Szkoda

Barry D. McCollum

Julie Diane Wigley Barnes

Tammy L. Johnson

Gregory Robert Crosby

Janet S. Cardell

Denice J. Blasé-Laudani-Alt #1

Kelly Ann Connors

Sunny Brook Veldboom

David Scott Breckling

Takissia Reshae Davis

Max Milliam Shadmani

Su Hyun Crocker-Alt #2

in the name and behalf of the citizens of Georgia, charge and accuse JESUS GUERRERO MANZANO with the offense of MURDER for that the said accused, in the County of Cobb and State of Georgia, on and about the 5TH day of NOVEMBER, 2003, did unlawfully and with malice aforethought cause the death of Claudia Rodriguez, a human being, by shooting Claudia Rodriguez in her head with a .45 caliber firearm contrary to the laws of said state, the good order, peace and dignity thereof.

ID#

Figure 8. Murder Method

Print Case Details Close

Criminal Case Number: 04900203
STATE v MANZANO JESUS GUERRERO
Judge: SCHUSTER

Case Type: INDICTMENT
Filing Date: 01/09/2004
Prosecutor: EVANS, JESSE

Defendants	Pleadings	Hearings	Attorneys	Offenses	Appeals	Bond Information	Sentences	
Defendant #	Count	Offense	Severity	Disposition	Status	Judge	Dispo Date	Signed Date
1	1	MURDER	FELONY	VERDICT NOT GUILTY	JAIL	SCHUSTER	02/25/2005	02/25/2005
1	2	MURDER	FELONY	VERDICT GUILTY	JAIL	SCHUSTER	02/25/2005	02/25/2005

Figure 9. The number of Charges

Print Case Details Close

Criminal Case Number: 04900203
STATE v MANZANO JESUS GUERRERO
Judge: SCHUSTER

Case Type: INDICTMENT
Filing Date: 01/09/2004
Prosecutor: EVANS, JESSE

Defendants	Pleadings	Hearings	Attorneys	Offenses	Appeals	Bond Information	Sentences
47	02/25/2005 02/28/2005	1	COURT	SENTENCE		20050027192	
48	02/25/2005 02/28/2005	1	COURT	VERDICT		20050027193	

First Pleading Prev Pleading Next Pleading Last Pleading Close

1 / 1 100%

IN THE SUPERIOR COURT OF COBB COUNTY, GEORGIA
CRIMINAL ACTION NO. 04-9-203-48
WARRANT NO. 03612069

Filed In Office Feb-25-2005 18:48:47
ID# 2005-0027192-CR
Page 1

Jay C. Stephenson
Clerk of Superior Court Cobb County

The State
vs
Jesus Guerrero Manzano

OFFENSE(S)
ct.1) murder
ct.2) murder

☐ PLEA
☐ NEGOTIATED
☐ GUILTY ON COUNT(S)
☐ NOLO CONTENDERE ON COUNT(S)
☐ TO LESSER INCLUDED ON COUNT(S)
☒ VERDICT
☒ GUILTY ON COUNT(S) 2
☒ NOT GUILTY ON COUNT(S) 1
☐ GUILTY OF LESSER INCLUDED
☒ FELONY SENTENCE
☐ MISDEMEANOR SENTENCE

WHEREAS, the above-named defendant has been found guilty of the above-stated offense, WHEREUPON, it is ordered and adjudged by the Court that the said defendant hereby sentenced to...

Figure 10. Verdict

|| Georgia Department of Correct x +

← → ↻ ⚠ Not secure | dcor.state.ga.us/GDC/OffenderQuery/jsp/OffQryForm.jsp

Search by Name or Description

Last Name (matches partial): Manzano

First Name (matches partial):

Gender: All

Race: All

Age: Between And Years Old

Most Recent Institution: All

Submit Form Reset Form More Options

Search by ID or Case Number

Figure 11. Public Criminal Information

Case Id	Defendant Name	Num_of_Codefendants	Defendant_Age	Defendant_Sex	Num_of_Att
4900191	Antonio Walker	0	30	1	3
4900199	Andre Lawrence	1	16	1	2
4900203	Jesus Manzano	0	27	1	4
4900673	Stacey Humphreys	0	30	1	6
4901186	Rodgerick Swanson	0	42	1	2
4904319	Aaron Willis	0	21	1	3
4904974	Donovan Leger	0	33	1	2
6902974	Sonya Yvette Smith	1	39	0	4
6903363	Colton Williams	5	16	1	2
7900473	Eliot Ellerton Jeffers	1	30	1	4
7901329	William Brian Hughes	0	32	1	2
7901990	Natasha Wynetta Demery	0	41	0	2
7902004	Mario Hodges	0	40	1	2
7902824	Christian Javon Wormum	0	18	1	2

Figure 12. Data for Analysis

APPENDIX B R Code Snippets

```
suppressWarnings(RNGversion("3.5.3"))
install.packages(c("randomForest"))
install.packages("adabag")

library(caret)
library(gains)
library(rpart)
library(rpart.plot)
library(pROC)
library(randomForest)
library(readxl)
library(adabag)
myData_DT <- read_excel("Final_Project.xlsx", sheet = "Verdict_Data")

myData_DT$Verdict<- as.factor(myData_DT$Verdict)
myData_DT$Prosecutor <- as.factor(myData_DT$Prosecutor)
myData_DT$Priors <- as.factor(myData_DT$Priors)
myData_DT$Method <- as.factor(myData_DT$Method)
myData_DT$Witness_Testimony <- as.factor(myData_DT$Witness_Testimony)
myData_DT$Defendant_Sex <- as.factor(myData_DT$Defendant_Sex)

myData_DT <- myData_DT[, 3:13]
View(myData_DT)

set.seed(1)
myIndex <- createDataPartition(myData_DT$Verdict, p=0.6, list=FALSE)
trainSet <- myData_DT[myIndex,]
validationSet <- myData_DT[-myIndex,]
View(trainSet)

set.seed(1)
default_tree <- rpart(Verdict ~., data = trainSet, method="class")
summary(default_tree)
prp(default_tree, type=1, extra=1, under=TRUE)

data.frame(imp = default_tree$variable.importance)

set.seed(1)
full_tree <- rpart(Verdict ~., data = trainSet, method="class", cp=0, minsplit=2, minbucket=1)
prp(full_tree, type=1, extra=1, under=TRUE)
printcp(full_tree)

data.frame(imp = full_tree$variable.importance)

pruned_tree <- prune(full_tree, cp=0.545454)
prp(pruned_tree, type=1, extra=1, under=TRUE)

predicted_class <- predict(pruned_tree, validationSet, type="class")
confusionMatrix(predicted_class, validationSet$Verdict, positive="1")

data.frame(actual = validationSet$Verdict, predicted = predicted_class)

predicted_prob <- predict(pruned_tree, validationSet, type="prob")
validationSet$`Verdict (1=Not Guilty)` <- as.numeric(as.character(validationSet$Verdict))
```

```
validationSet$Verdict <- as.numeric(as.character(validationSet$Verdict))
gains_table_DT <- gains(validationSet$Verdict, predicted_prob[,2])
gains_table_DT

plot(c(0, gains_table_DT$cume.pct.of.total*sum(validationSet$Verdict)) ~ c(0,
gains_table_DT$cume.obs),
     xlab="# of Cases",
     ylab = "Cumulative",
     main="Cumulative Lift Chart",
     type="l")
lines(c(0, sum(validationSet$Verdict)) ~ c(0, dim(validationSet)[1]), col="red", lty=2)
barplot(gains_table_DT$mean.resp/mean(validationSet$`Verdict (1=Not Guilty)`),
names.arg=gains_table_DT$depth,
     xlab="Percentile",
     ylab="Lift",
     ylim=c(0,3),
     main="Decile-Wise Lift Chart")
roc_object_DT <- roc(validationSet$Verdict, predicted_prob[,2])
plot.roc(roc_object_DT, print.auc = TRUE)
auc(roc_object_DT)
```

Teaching Case

Tax Time: An Interdisciplinary Accounting Analytics Experiential Learning Activity

Joseph M. Woodside
joseph.m.woodside@gmail.com
Department of Business Systems and Analytics
Stetson University
DeLand, FL 32723

Monica Mendoza
monica.mendoza@stetson.edu
Department of Accounting
Stetson University
DeLand, FL 32723

Abstract

Given the importance of interdisciplinary learning and the growth of data analytics skillsets within the accounting domain, this manuscript describes a teaching case for IRS tax filing statistics to develop students' knowledge of analytics and connect current events in business and accounting. The experiential learning activity is developed in the context of an undergraduate upper-level course on descriptive and predictive analytics. The contributions of this teaching case are an experiential learning activity applied to a real-world current event and an interdisciplinary learning activity that allows students to apply and develop their curricular knowledge. The overall case objectives are to estimate the total number of individual IRS tax returns processed and total refunds for the current tax season and offer additional insights and recommendations based on the analysis.

Keywords: Accounting Analytics, Data Analytics, Forecasting, Decision Making, IRS, Tax Returns, Tax Refunds, Experiential Learning, Teaching Case

1. INTRODUCTION AND OBJECTIVES

Tax season is often the busiest time of year for tax preparers, employers, financial custodians, businesses, and accounting professionals. Each year in the U.S. individual taxpayers prepare their financial statements from the previous year and submit tax returns to the Internal Revenue Service (IRS). Each year analysts, reporters, and subject matter experts try to review various factors such as tax law changes and offer tax season predictions. Various headlines such as "A new report predicts huge swaths of Americans should expect bigger tax refunds", "Report suggests more taxpayers will owe tax due to

insufficient withholding", "Refunds are higher this year", or "Americans are worried about smaller tax refunds this year" are commonplace every tax season (Phillips Erb, 2018; Kiersz 2018, Picchi 2022; Picchi, 2022b; Buchwald, 2022).

Tax season is the period of time during which taxes must be filed and paid, typically running from January 1st through April 15th of each year unless that day is a weekend or holiday. The IRS announced the tax season for 2022, which began on January 24th when returns started being accepted, through the April 18th tax filing deadline due to the Emancipation Day holiday. Taxpayers were also able to request an extension until

October 17th. The IRS selected January 24th to allow enough time to complete programming and testing of IRS systems to ensure proper tax returns such as the 2021 Child Tax Credit and Recovery Rebate Credit. Individuals who were not required to file in prior years were required to file in 2021 and 2022 to claim new credits (IRS, 2022; Kagan, 2022; Safane, 2022).

In 2022, IRS Commissioner Chuck Rettig posted that the tax season was unlike any other with challenges from the COVID-19 pandemic, staffing shortages, and efforts to distribute stimulus checks and advanced Child Tax Credits. The tax administration's effectiveness has further suffered from underinvestments in human capital and information technology. Over the last decade, the IRS' budget decreased by more than 15 percent, IRS enforcement personnel decreased by 30 percent, and the filing population increased by 14 percent. An estimated 16 million tax returns, transactions, and accounts management cases were backlogged at the end of 2021. To address these issues along with employee retirements, up to 10,000 employees were planned to be hired in 2022. An estimated 52,000 employees will need to be hired in the next six years for the IRS to maintain current service levels. The fiscal year 2023 provides \$14.1 billion to the IRS with plans to improve taxpayer service, ensure fairness in the tax system, and modernize systems (Cohn, 2022; Retting, 2022). In 2022, the average individual refund was \$3,176, an increase of nearly 14% over 2021. According to tax experts 2023 tax refunds are expected to decrease. While some enhanced tax credits remain, other tax changes including no stimulus checks to claim and expanded child tax credit which ended in 2021 are anticipated to have an effect on tax refunds (Keshner, 2023). Considering the unprecedented pressures and challenges that occurred during the prior tax seasons, the overall objectives for this case are to predict the total number of individual income tax returns and the total dollar amount of income tax refunds processed by the IRS filing deadline.

2. BACKGROUND

IRS

The IRS is a bureau of the Department of the Treasury and tax administrator. During the fiscal year 2020, the IRS collected nearly \$3.5 trillion in revenue and nearly 250 million tax returns. The mission of the IRS is to provide top-quality service to taxpayers by helping them to understand their tax responsibilities and enforcing the law. The U.S. Congress is

responsible for passing tax laws, and the taxpayer is responsible for understanding and meeting their tax obligations. The IRS carries out the responsibilities of the Secretary of the Treasury under section 7801 of the Internal Revenue Code, and section 7803 of the Internal Revenue Code provides the appointment of a commissioner to oversee and apply the laws (IRS, 2022b)

Tax Filings - PESTLE

Tax filings and the number of returns and refunds may vary by year. A secondary analysis review is provided below using a Political, Economic, Social, Technical, Legal, and Environmental business analysis framework (PESTLE) to examine factors related to tax filings. The PESTLE and similar methods can be utilized for strategic analysis and developing a theoretically-informed understanding of the business environment for strategic deployment (Downey, 2007; Woodside, Augustine, & Giberson, 2017).

Political

Taxes are often a common political platform. Tax policies including who pays taxes, their effects on the economy, and the amount of revenue raised have been part of political discussions across centuries. Tax policies and the resulting economic impacts are extremely complex, and even simple questions lead to debates, for example, a common debate for policymakers focuses on whether tax increases or decreases will impact jobs, economic growth, deficits, revenue, and/or fiscal positions (Moss, Nunn, & Shambaugh, 2020).

Economic

Economic conditions can also impact tax revenues and filings. Typically, income taxes are more volatile than consumption-based taxes, for example during the Great Recession period 2008-2010, income taxes decreased by 16 percent (Walczak, 2020). More recently, The Tax Policy Center estimated that in 2020 there was a 38% increase in households with no federal income tax liability (Tax Policy Center, 2021; Watson, 2021). However, 2020 was an unusual year due to economic impact payments and expanded tax credits. Government funding and support were provided to address the COVID-19 pandemic's economic effects (Watson, 2021).

Social

Often individuals compare tax refunds to the national average or prior tax year refunds, although a larger refund isn't always best, since this is usually a result of overpaying taxes in the prior year. In the fiscal year 2020, the IRS issued over \$736 billion in refunds, money that could

have been received earlier on regular paychecks. In 2020, the IRS redesigned the W-4 form used for withholdings and the estimated tax payments sent to the IRS from each paycheck (Mercado, 2017; Doyle, 2022). Tax refunds may also vary by state, owing primarily to differences in income and income taxes paid in the prior year. In 2019 Wyoming received the largest average tax refund, followed by Connecticut and New York with averages of \$5,207, \$4,461, and \$4,444 respectively. The average U.S. tax refund in 2019 was \$3,651 (Picchi, 2022).

Technical

The IRS is undergoing a multi-year modernization plan which includes technological improvements such as cybersecurity, cloud, robotic process automation, digitization, and application programming interfaces. The plan runs from the fiscal year 2019 to the fiscal year 2024 with an estimated \$2.3-2.7 billion in costs to implement. One goal is to reduce call wait times with customer callback technology, online notice, and live online customer support (IRS, 2019; IRS, 2022c). During the 2022 tax season, the understaffed IRS had 1 person for every 16,000 calls received (Kaplan & Zeballos-Roig, 2022).

Legal

Taxes have existed since the beginning of civilization, with the earliest known individual tax implemented in Mesopotamia over 4500 years ago. In the U.S., taxes can be traced to the 1700s in Colonial America. In 1913, power was granted through Congress to tax personal income, as ratified by the 16th amendment to the Constitution. In 1914, the first income tax form was available, Form 1040. This is the same primary income tax form today, though it has been continuously modified over the years. By 1915 Congress and the public voiced concerns about the complexity of the tax form and difficulty filing returns. In 1916, The Revenue Act created sliding tax rates by income similar to those in use today. There have been several major tax revisions, with the most recent being in 2017 with the Tax Cuts and Jobs Act scheduled to run through 2025. The U.S. tax code has increased in length and complexity since the beginning; in 1913 the tax code could fit on a single page, whereas the same today can require 174 pages. In only the past decade, the tax code has been updated over 4,000 times (eFile, 2022).

Environmental

While the tax deadline is typically April 15th each year, environmental factors can affect the results. On March 17, 2021, the 2020 tax year filing date was extended until May 17, 2021 (IRS, 2021) due

to the COVID-19 pandemic (IRS, 2021). The emergence of the novel coronavirus and the resulting pandemic changed many aspects of society and economies (Witman & Prior, 2021; Oum, Kates, & Wexler, 2022). Further, in 2021, extensions were made until June 15, 2021, in Texas, Oklahoma, and Louisiana as a result of winter storms that occurred in February 2021 (IRS, 2021).

3. IRS TAX ACCOUNTING ANALYTICS EXPERIENTIAL LEARNING ACTIVITY

Accountants may be seen as being at the forefront of analytics for their organizations, as their positions require providing management with information used in strategic data-driven decision-making. Therefore, analytics skills and competencies are critical prerequisites for accountants (Kokina, Pachamanova, & Corbett, 2017; Woodside, et al., 2020). Nearly half of accounting firms are using or plan to use advanced and predictive analytics. Among data analytics and new technology adoption, only 25% have completed the implementation of predictive modeling as compared with 48% having implemented self-service reporting. Organizations are utilizing predictive models within the business processes that can be updated as new data sources are available continuously adapting and increasing accuracy based on market and customer changes (Maryville University, 2022).

While accounting is a vital and basic function of any organization, the role of accountants is transforming due to the increase in data and analytics. The American Institute of Certified Public Accountants and National Association of State Boards of Accountancy have identified the importance of integrating analytics within accounting core coursework (AICPA and NASBA, 2021; Losi, Isaacson, & Boyle, 2022). Organizations seek individuals that can leverage information and have a broad range of expertise, including knowledge of analytics such as data exploration, predictive modeling, and the use of analytical software tools (Woodside, Augustine, Chambers, & Mendoza, 2020). For example, in tax accounting, analytics can be utilized to review global compliance and transactional data in order to provide insights for reducing effective tax rates (Villanova, 2022).

In response to rapid technological transformation in the industry, university curricula must adapt to relevant and practical applications. The Association to Advance Collegiate Schools of Business (AACSB) Accounting Standards,

including Standard A5 (v. 2021), require real-world learning experiences and skills that integrate information technology in accounting and business. AACSB recommends an integrated, interdisciplinary curriculum in three primary areas of information systems, data analytics, and technology agility. These areas include data creation, data management, statistical techniques, data modeling, analysis, and predictive analytics (Woodside, et al., 2021).

Analytics Process

Your initial task is to follow a standard analytics process to generate a response for each step of the process. The CRoss Industry Standard Process for Data Mining (CRISP-DM) model is one of the most common standardized processes, and is a non-proprietary methodology intended to allow data mining projects to be more reliable, cost less, be repeatable, and improve management. The market benefits of a common model include customer satisfaction and establishing a specific data mining process (Wirth and Hippa, 2001). The CRISP-DM phases include business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Business understanding concerns outlining project objectives, business requirements, problem definition, and initial project plan. Data understanding begins with data collection to familiarize oneself with the data, locate data quality issues, and form hypotheses. The two sections of business understanding and data understanding are closely linked. Data preparation covers table creation, record selection, cleansing, deriving attributes, and the transfer of data into modeling tools. In the modeling phase, different techniques are chosen and applied, and data preparation and modeling are closely linked. In the evaluation phase, models have been constructed, and verify the model meets business requirements. Deployment, while the final phase does not indicate the end of a project, and may require additional steps by the data analytics or customer, and care should be taken to clearly define the setup of the working model (Wirth and Hippa, 2001; Woodside, 2016; Woodside, 2018).

IRS Dataset

As part of the data understanding and data preparation series of steps, you will be required to compile a dataset with IRS tax filing data. The data and various tax statistics are available in the public US Domain through the IRS.gov websites at [IRS.gov/statistics](https://irs.gov/statistics) and [IRS.gov/newsroom/filing-season-statistics-by-year](https://irs.gov/newsroom/filing-season-statistics-by-year) (IRS, 2022d; IRS, 2022e). The IRS provides an individual income tax returns bulletin article

and related statistical tables on a sample of individual tax returns (e.g. Forms 1040, 1040A, and 1040EZ). Data includes two versions with preliminary data estimates January-September, and complete year data January-December (IRS, 2022d). Tax season filing statistics are also available by year and by week, with data fields including total returns received, total returns processed, e-filing receipts by tax professionals and self-prepared, web usage visits to IRS.gov, the total number of refunds, the total amount of refunds, average refund amount, and number and amount of direct deposit refunds. A sample IRS dataset with individual fields available on the IRS tax statistics website is shown in Table 1 (IRS, 2022e).

Table 1: IRS Dataset Fields

IRS Category	Field Name
Individual Income Tax Returns	Total Receipts
	Total Processed
E-filing Receipts	E-Filing Total
	E-Filing Tax Professionals
	E-Filing Self-Prepared
Web Usage	Visits to IRS.gov
Total Refunds	Number
	Amount (Billions)
	Average refund
Direct Deposit Refunds	Number
	Amount (Billions)
	Average refund

Analytics Platform

The experiential learning activity (ELA) can be completed through a data analytics modeling platform including Excel, Knime, R, RapidMiner, SAS, or any equivalent platform. An ELA template has been provided in the Appendix following the CRISP-DM repeatable analytics process. The steps and objectives are listed within the template including developing a dataset, exploring and modeling the data, providing an evaluation summary of findings, insights, and recommendations following the analysis.

Response Prompts

1. Business Understanding – Provide an Executive Summary of the problem/objective. In this case it would be to predict the total number of individual

income tax returns and the total dollar amount of income tax refunds processed by the IRS filing deadline. Similarly, a prediction can be made of the total dollar amount of income tax collected using available data.

2. Data Understanding / Data Preparation - Develop a dataset using the IRS page <https://www.irs.gov/newsroom/filing-season-statistics-by-year> and generate a data taxonomy for the final dataset. This data requires some preparation by combining several years of data for our modeling and basic data cleaning to remove headings, non-numerical data and other formatting issues that may arise.
3. Data Exploration - Generate descriptive analytics using a data analytics toolset. Describe the output and review the data quality metrics for the dataset.
4. Data Modeling - Use R (or similar data analytics tools) and time series analysis to generate a prediction for the total number of individual tax returns processed and the total amount of refunds issued by the tax filing deadline.
5. Evaluation and Deployment - Describe how this data-driven result and confidence compared with your original prediction and confidence at the beginning of the course. Provide a summary of insights and recommendations based on your analysis.
6. Integration of Learning / Ethics - Describe your existing knowledge connections that helped the most in solving this problem, and how was learning adjusted or adapted. Describe the ethical considerations of teamwork, and how teamwork contributed to the success of this activity.

4. REFERENCES AND RESOURCES

- AACSB. 2018 Standards for Accounting Accreditation. (2021). Retrieved June 6 2022 from: https://www.aacsb.edu/-/media/documents/accreditation/accounting/standards-and-tables/accounting2018standards_2021.pdf?rev=080cc71d56d54f0aae30bc09274aa26d
- American Institute of Certified Public Accountants (AICPA) and National Association of State Boards of Accountancy (NASBA). (2021). CPA Evolution Model Curriculum. Retrieved December 28, 2022 from <https://www.thiswaytocpa.com/collectedmedia/files/cpa-evolution-model-curriculum-update.pdf>
- Buchwald, E. (2022). 2022 taxes: Refunds are higher thanks to economic stimulus checks, Child Tax Credit. USA Today. Retrieved June 28, 2022 from <https://www.usatoday.com/story/money/taxes/2022/04/13/irs-tax-2022-refund-delay/7289255001/?gnt-cfr=1>
- Cohn, M. (2022). IRS still hampered by backlog and staffing shortages. Accounting Today. Retrieved June 1, 2022 from <https://www.accountingtoday.com/news/irs-still-hampered-by-backlog-and-staffing-shortages>
- Downey, J. (2007). Strategic Analysis Tools. CIMA Global. Retrieved June 3, 2022 from http://www.cimaglobal.com/Documents/ImportedDocuments/cid_tg_strategic_analysis_tools_nov07.pdf
- Doyle, K. (2022). Why a Tiny Tax Refund Is Good for Your Finances — and How To Make It Happen. Retrieved June 3, 2022 from <https://www.gobankingrates.com/taxes/refunds/why-getting-a-large-tax-refund-is-bad/>
- eFile. (2022). Income Tax History. Retrieved June 2, 2022 from <https://www.efile.com/tax-history-and-the-tax-code/>
- Kagan, J. (2022). Tax Season. Investopedia. Retrieved June 1, 2021 from <https://www.investopedia.com/terms/t/tax-season.asp#:~:text=Tax%20season%20is%20when%20individuals,for%20the%202021%20tax%20year>
- Kaplan, J. & Zeballos-Roig, J. (2022). Business Insider. Retrieved June 3, 2022 from <https://www.businessinsider.com/irs-tax-season-problems-ahead-delayed-refunds-2022-1>
- Keshner, A. (2023). Brace yourself: Your tax refund could shrink in 2023. Here's Why. Retrieved January 7, 2023 from https://www.marketwatch.com/story/brace-yourself-your-tax-refund-could-shrink-in-2023-heres-why-11672778149?mod=mw_latestnews
- Kiersz, A. (2018). A new report predicts huge swaths of Americans should expect bigger tax refunds — but not everyone is set to benefit. Business Insider. Retrieved May 1, 2022 from <https://www.businessinsider.com/ubs-report-married-filers-bigger-federal-tax-refund-tax-day-2019-2018-12>
- Kokina, J., Pachamanova, D., & Corbett, A. (2017). The Role of Data Visualization and Analytics in Performance Management: Guiding Entrepreneurial Growth Decisions. *Journal of Accounting Education*, 38, 50-62.

- IRS. (2019). Integrated Modernization Business Plan. Department of the Treasury Internal Revenue Service. Retrieved June 8, 2022 from <https://www.irs.gov/pub/irs-pdf/p5336.pdf>
- IRS. (2021). Tax Day for individuals extended to May 17: Treasury, IRS extend filing and payment deadline. Retrieved June 22, 2022 from <https://www.irs.gov/newsroom/tax-day-for-individuals-extended-to-may-17-treasury-irs-extend-filing-and-payment-deadline>
- IRS. (2022). 2022 tax filing season begins Jan. 24; IRS outlines refund timing and what to expect in advance of April 18 tax deadline. Retrieved June 1, 2021 from <https://www.irs.gov/newsroom/2022-tax-filing-season-begins-jan-24-irs-outlines-refund-timing-and-what-to-expect-in-advance-of-april-18-tax-deadline>
- IRS. (2022b). The Agency, its Mission and Statutory Authority. Retrieved June 3, 2022 from <https://www.irs.gov/about-irs/the-agency-its-mission-and-statutory-authority>
- IRS. (2022c). IRS Modernization Plan provides plan to improve services for taxpayers, tax community. Retrieved June 8, 2022 from <https://www.irs.gov/newsroom/irs-modernization-plan-provides-plan-to-improve-services-for-taxpayers-tax-community>
- IRS. (2022d). SOI Tax Stats - Individual Income Tax Returns. Retrieved June 1, 2022 from <https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-returns>
- IRS. (2022e). Filing Season Statistics by Year Retrieved June 9, 2022 from <https://www.irs.gov/newsroom/filing-season-statistics-by-year>
- Losi, H.J., Isaacson, E.V., & Boyle, D.M.. (2022). Integrating Data Analytics into the Accounting Curriculum: Faculty Perceptions and Insights. *Issues in Accounting Education* 37(4):1.
- Maryville University. (2022). Uses of Data Analytics in Accounting and Finance. Retrieved June 28, 2022 from <https://online.maryville.edu/blog/data-analytics-in-accounting/>
- Mercardo, D. (2017). Here's Why That Big Tax Refund Isn't Great News For You. CNBC. Retrieved June 3, 2022 from <https://www.cnbc.com/2017/03/27/heres-why-that-big-tax-refund-isnt-great-news-for-you.html>
- Moss, E., Nunn, R., & Shambaugh, J. (2020). Tackling the tax code: Efficient and equitable ways to raise revenue. Brookings Institute. Retrieved June 8, 2022 from <https://www.brookings.edu/multi-chapter-report/tackling-the-tax-code-efficient-and-equitable-ways-to-raise-revenue/>
- Oum, S., Kates, J., & Wexler, A. (2022). Economic Impact of COVID-19 on PEPFAR Countries. KFF. Retrieved June 22, 2022 from <https://www.kff.org/global-health-policy/issue-brief/economic-impact-of-covid-19-on-pepfar-countries/>
- Phillips Erb, K. (2018). Report Suggests More Taxpayers Will Owe Tax In 2019 Due To Insufficient Withholding. Forbes. Retrieved May 1, 2022 from <https://www.forbes.com/sites/kellyphillipserb/2018/08/04/report-suggests-more-taxpayers-will-owe-tax-in-2019-due-to-withholding/?sh=2fdbcdaf5f0f>
- Picchi, A. (2022). Americans are worried about smaller tax refunds. They may have reason for concern. CBS News. Retrieved June 28, 2022 from <https://www.cbsnews.com/news/tax-refund-irs-smaller-refunds/>
- Picchi, A. (2022b). Tax refunds: Here's where Americans get the biggest IRS refunds. CBS News. Retrieved June 3, 2022 from <https://www.cbsnews.com/news/tax-refund-size-by-state-wyoming-connecticut-new-york/>
- Rettig, C. (2022). IRS Tax Day Message: Thank You for Filing. Retrieved June 1, 2021 from <https://www.irs.gov/about-irs/irs-tax-day-message-thank-you-for-filing>
- Safane, J. (2022). What Is Tax Season? The Balance. Retrieved June 1, 2022 from <https://www.thebalance.com/what-is-tax-season-5201643>
- Tax Policy Center. (2021). T21-0161 - Tax Units with Zero or Negative Income Tax, 2011-2031. Retrieved September 9, 2022 from <https://www.taxpolicycenter.org/model-estimates/tax-units-zero-or-negative-income-tax-liability-august-2021/t21-0161-tax-units-zero>
- Walczak, J. (2020). Income Taxes Are More Volatile Than Sales Taxes During an Economic Contraction. Retrieved June 3, 2022 from <https://taxfoundation.org/income-taxes-are-more-volatile-than-sales-taxes-during-an-economic-contraction/>

- more-volatile-than-sales-taxes-during-recession/
- Villanova. (2022). Data Analytics in Accounting. Retrieved June 28, 2022 from <https://taxandbusinessonline.villanova.edu/blog/data-analytics-in-accounting/>
- Watson, G. (2021). COVID-19 Tax Relief Added to Increasing Share of Households Paying No Income Tax. Tax Foundation. Retrieved June 22, 2022 from <https://taxfoundation.org/us-households-paying-no-income-tax/>
- Wirth, R., Hipp, J. (2001). CRISP-DM: Towards a Standard Process Model for Data
- Witman, P.D., & Prior, J. (2021). Vaccination Registration –What Could Possibly Go Wrong? 2021 Proceedings of the EDSIG Conference.
- Woodside, J. M. (2016). BEMO: A Parsimonious Big Data Mining Methodology . AJIT-e: Academic Journal of Information Technology , 7 (24) , 113-123 . DOI: 10.5824/1309-1581.2016.3.007.x
- Woodside, J. M. (2018). Applied Health Analytics and Informatics Using SAS. SAS Press.
- Woodside, J.M.; Augustine, F.K. Jr.; and Giberson, W. (2017). Blockchain Technology Adoption Status and Strategies. Journal of International Technology and Information Management, 26(2). Retrieved June 3, 2022 from <https://scholarworks.lib.csusb.edu/jitim/vol26/iss2/4>
- Woodside, J., Augustine, F., Chambers, V. and Mendoza, M., (2020). Integrative Learning and Interdisciplinary Information Systems Curriculum Development in Accounting Analytics. Journal of Information Systems Education, 31(2), 147-156.

Appendix

Appendix A: Experiential Learning Activity Template

Experiential Learning Activity

Name

- ELA Aligned AAC&U Essential Learning Outcomes:

AAC&U Essential Learning Outcomes	
1. Inquiry and Analysis	✓
2. Critical and Creative Thinking	✓
3. Written and Oral Communication	✓
4. Quantitative Literacy	✓
5. Information Literacy	✓
6. Teamwork and Problem Solving	✓
7. Ethical Reasoning and Action	✓
8. Foundations and Skills for Lifelong Learning	✓
9. Integrative and Applied Learning: Synthesis Across General and Specialized Studies	✓

- Employer-Valued Knowledge, Skills, and Abilities (KSAs) Gained:
 - Knowledge of Data-driven Decision Making
 - Knowledge of Data Mining Techniques and Processes
 - Skilled in a Data Analytics Toolset (R or similar)
 - Ability to Extract and Develop Structured Data Sources
 - Ability to Apply Analytics to Various Industry Domains
- Description:
 - Describe your Process Steps and Methodology
 - Describe your Knowledge Resources
 - Paste Screenshots for the outputs (PrtScn -> Paste, Mac: Cmd+Shift+4)
 - Determine the total individual income tax filing results through the tax filing deadline
 - [Provide a Reflection](#)

Business Understanding

Executive summary of the business problem/opportunity/objective:

Describe your project resources, requirements, and methodology.

Data Understanding, Data Preparation, and Data Integration

Generate a data taxonomy for the final dataset:

Prepare the source data in a spreadsheet for analysis (source hint: [IRS Data](#)):

Data Exploration and Data Quality Review

Generate descriptive analytics in a data analytics toolset. Describe the output and review the data quality metrics for the dataset (paste screenshots of each result):

Data Modeling and Data Mining

Use the above descriptive analytics outputs, data analytics tools, and data mining method to generate a prediction, provide screenshots and a business description of the results:

January 1st - April 18th, 2022 (or tax filing deadline)

Total Individual Income Tax Returns Received:

Confidence (1-100%):

\$ Total Individual Income Tax Returns Refunds:

Confidence (1-100%):

Evaluation, Deployment, and Operationalization

How did this data-driven decision and confidence compare with your original prediction and confidence level from the beginning of the course?

Provide a summary of insights and recommendations based on your overall analysis.

Integration of Learning / Ethics Reflection

Describe Integration of Learning. What existing knowledge did you connect that helped the most in solving this problem? How did you adjust or adapt your learning for this objective?

What are the ethical considerations of professional teamwork, and how did your teamwork contribute to the success of the activity? Describe one thing you learned from your teammate on the activity, and describe one thing your teammate learned from you on the activity. As part of your continuous improvement efforts, recommend one specific change the team could make to further progress everyone's task performance and learning for next time.

Teaching Case

Health Care Management: Preventing Post-Surgical Falls after Hip or Knee Replacement Surgery through Predictive Analytics

Richard McCarthy
richard.mccarthy@quinnipiac.edu

Wendy Ceccucci
wendy.ceccucci@quinnipiac.edu

Department of Business Analytics and Information Systems
Quinnipiac University
Hamden, CT 06518 USA

Abstract

The adult population in the United States is more physically active and are living longer than prior generations. Due to the advancement in surgical techniques and the increased number of active people, there has been a rise in the number of hip and knee replacement surgeries. This rise in the number of surgeries is expected to continue. Post-surgical care is a critical component to a successful patient recovery. After surgery, patients experience limited mobility while the muscles around the impacted joints need time for inflammation to subside. Physicians and other medical providers are concerned with making sure that patients do not experience falls during this time as it may lead to more serious injuries. A sample dataset of patients who underwent elective hip or knee surgery from January 2014 to March 2020 has been provided to analyze other medical conditions that may contribute to the likelihood of a patient falling. The goal is to identify important factors that can assist in predicting the probability of a patient falling after surgery.

Keywords: health care analytics, data science, predictive analytics, SAS Enterprise Miner™

1. CASE SUMMARY

Orthopedic surgeons are interested in understanding the factors that may increase the probability of a patient falling after hip or knee surgery. This is important for both patient health care and for medical cost control. A hospital from the northeastern United States has made over six years of patient data available. Personally identifiable data about specific patients is not included so as not to violate the patients' rights to privacy. The file contains data on the patients that

underwent elective hip or knee surgery between January 2014 to March 2020. In this case study, current machine learning techniques will be used to identify factors that can contribute to a patient falling after surgery.

Based upon the data provided, post-operative falls after hip or knee surgery occurred in approximately 1% of patients. In statistical terms this is considered a rare event. A rare event occurs in a binary outcome when the probability of the outcome is low. This case provides an

opportunity to use real life data to build multiple predictive models using a variety of tools (e.g., SAS, R, Python). It also provides an opportunity to work with rare event data.

2. BACKGROUND

Globally post operative falls and their subsequent injuries are a health care expenditure amongst the aging population. In 2014, patients over age 65 in the United States sustained 29 million falls, resulting in 7 million injuries, 800,000 hospitalizations, 27,000 deaths, and an estimated 31 billion dollars in annual Medicare expenditure (Bergen, Stevens & Burns, 2014). Recent literature suggests that patients who fall in the early post-operative period, after undergoing hip or knee replacement surgery, are more likely to experience a fracture.

Medicaid services, the single largest payor for hip or knee surgery in the United States, has identified in-hospital falls as a “preventable” acquired condition for which they would no longer cover in-hospital costs. Several studies have previously attempted to develop clinical tools to serve as predictors of inpatient falls (Conley, Schultz, & Selvin, 1999; Hendrich, Bender, & Nyhuis, 2003). Despite these interventions, post-operative joint replacement fall rates remain constant.

To better understand the factors that affect the likelihood of a patient falling, patient comorbidities, medications, and other factors will be analyzed. What is a *comorbidity*? When a patient has two or more health conditions at the same time, or if one condition occurs right after the other, this is considered a comorbidity. For example, a patient may have arthritis, a heart condition, and diabetes. Each of these would be considered a *comorbidity*.

The risk of falling increases significantly with age due to generalized osteoarthritis, tinnitus, cognitive impairment, and two or more comorbidities (Lastrucci, Lorini, Rinaldi, & Bonaccorsi, 2018). This risk is expected to continue to become a more prominent problem as the population continues to age worldwide. This is further compounded by the increase in implanted medical devices amongst younger patients due to hip or knee replacement surgery (Ong, Lau, Moore, & Heller, 2009). The most common types of falls observed from emergency departments (ED) are fractures (56%), superficial injuries (20.9%), and head injuries (8.7%), with the most common fractures being hip, wrist, and upper arm (Hartholt, van der

Velde, Looman, van Lieshout, Panneman, van Beeck, Tischa, & van der Cammen, 2010).

Drugs, which are a modifiable risk factor have also been found to contribute to risk of falling. Benzodiazepines, which are commonly used to treat anxiety and sleeping disorders have been found to be associated with falls (de Jong, Van der Elst, & Hartholt, 2013).

There is great interest in determining which patients are likely at risk for falling and possible actions that can be taken to prevent falls. Hendrich (2013) developed and subsequently modified the Hendrich Fall Risk Model, (see Appendix A), as a simplified approach to try to determine patients at greatest risk for falls. The Hendrich Fall Risk Model is a scoring model that assigns points based upon known causes of post-surgical falls for the purpose of identifying which patients are most likely at risk. When using this scorecard approach, any patient with a score of 5 or higher is considered to be at higher risk of falling and therefore should have a higher degree of monitoring to reduce the likelihood of falling.

However, with the explosion of analytic technologies, there is a need to determine other causations to further reduce risk of injury and to further contain medical costs. A more sophisticated analysis is necessary.

3. DATA DESCRIPTION

The patient data set for analysis consists of data on 17,275 patients who have had hip or knee replacement surgery from a single hospital located in the northeast United States. Patients with incomplete data were not included. The available data set is comprised of three Excel files:

1. Patient file
2. Medications by Category
3. Comorbidity Description.

Patient File

The patient file contains basic patient characteristics and surgery data. Many of the patient demographic information has been omitted due to HIPAA regulations in the United States and to ensure that the patients are not personally identifiable. The data dictionary containing a list and description of the variables is provided in Appendix B. The file contains information about the patient such as their gender, body mass index (BMI), type of procedure (knee or hip), if the patient had a post-operative fall, if they were injured when they fell, and subsequent data related to the fall. In

addition, patient comorbidities for each of the patients is available. There are a variable number of comorbidities per patient. Some comorbidities are very common across the patient data set, for example, osteoarthritis. There are other comorbidities however, which occur less frequently.

The key number is used as a sequential unique identifier for each patient. Key number was used to uniquely identify each patient without personally identifying any individual.

Medications by Category File

Medications are an important consideration. The medications by category file contains information on the medications that patients were taking after their surgery. This includes medications directly related to the surgery as well as other medications that the patient was taking due to other, pre-existing comorbidities. After hip or knee replacement surgery, there are typically several medications that patients are prescribed, ranging from vitamins and muscle relaxers to opioids for pain management. The data dictionary is presented in Appendix C. The file contains one row of data per patient per medicine. Since most patients are using many different medications, the resulting file contains 190,699 observations (rows).

Comorbidity by Description File

The comorbidity by description file contains a list of comorbidity codes along with their accompanying comorbidity name. There are 3,797 different comorbidity codes. As a result of changes to medical coding over time, some descriptions may appear to be duplicates or very similar. These are included for completeness of understanding of comorbidity coding.

4. THE ANALYSIS

The goal of your analysis is to determine what factors affect the probability of a patient falling after surgery. This requires an in-depth analysis of the patients, their underlying comorbidities, and their medications to determine additional factors that may cause a patient to fall. Complete the following steps in your analysis.

Step 1. Prepare your data. The data sets were extracted from a patient medical record database. The format is not necessarily a final format for analysis, so consider how best to define a data set for analytic evaluation.

Step 2. Analyze and Transform Variables

Check the distribution of each of the variables. Based on these results do any of the variables need to be transformed? If so, please indicate which variables were transformed and which approach was used. Review the number of categories in your categorical data. Do any of these require additional grouping or binning? If so, consider how to bin the data in the context of how it would be useful to orthopedic surgeons, nurses, physical therapists, and other post-operative care givers.

Are there variables that should be rejected and would not impact the likelihood of a patient falling after surgery? How are missing values treated? This is a rare event analysis. How does that impact how the data is analyzed?

Step 3. Model Selection

Determine four or more appropriate modelling techniques to be used for analysis. What are the possible measurement criteria that could be used for analysis? What measurement will be used to evaluate the models? Why did you select this measure?

Step 4. Develop and Train Models

Check for multicollinearity and any redundancies. Appropriately partition the data for analysis and then check the model performance. What is the impact of a Type I or Type II error?

Step 5. Validate and Test the Models

Compare the results of each model. Which model performed the best. What could be done to further optimize the model?

5. FINAL REPORT

Prepare a final report that will help orthopedic surgeons understand what factors affect the probability of a patient experiencing a post-operative fall. Some of the variables will be obvious but what are some of the not so obvious variables that may impact the likelihood of a patient falling?

In your final report, be sure to discuss each of the five steps, any actions taken and the results.

6. REFERENCES

Bergen, G., Stevens, M.R., & Burns, E.R. (2016). Falls and Fall Injuries Among Adults Aged ≥ 65 Years-United States-2014. *MMWR Morbidity and Mortality Weekly Report*, 65(37), 993-998.

- Conley, D., Schultz, A. A., & Selvin, R. (1999). The challenge of predicting patients at risk for falling: Development of the Conley Scale. *Medsurg Nursing*, 8(6), 348-354.
- de Jong, M., Van der Elst, M., & Hartholt, K. (2013). Drug-related falls in older patients: Implicated drugs, consequences, and possible prevention strategies, *Therapeutic Advances in Drug Safety*, 4(4), 147-154.
- Hartholt, K., van der Velde, N., Looman, C., van Lieshout, E., Panneman, M., van Beeck, E., Patka, P., Tischa, J., & van der Cammen, M. (2010). Trends in fall-related hospital admissions in older persons in the Netherlands, *Arch Intern Med*, 170(10), 905-911.
- Hendrich, A., (2013). Fall risk assessment for older adults: The Hendrich II Fall Risk Model, *The Hartford Institute for Geriatric Nursing*, 8, 1-2.
- Hendrich A., Bender, P.S., Nyhuis, A., (2003, February) Validation of the Hendrich II Fall Risk Model: a large concurrent case/control study of hospitalized patients. *Applied Nursing Research*, 16(1), 9-21.
- Lostrucci, V., Lorini, C., Rinaldi, G., & Bonaccorsi, G. (2018). Identification of fall predictors in the active elderly population from the routine medical records of general practitioners, *Primary Health Care Research & Development*, 19(2), 131-139.
- Ong, K., Lau, E., Moore, T., & Heller, M. (2009). Accidental falls involving medical implant re-operation, *Injury*, 40(10), 1088-1092.

APPENDIX A Hendrich Fall Risk Model

Hendrich II Fall Risk Model		
RISK FACTOR	RISK POINTS	SCORE
Confusion/Disorientation/Impulsivity	4	
Symptomatic Depression	2	
Altered Elimination	1	
Dizziness/Vertigo	1	
Gender (Male)	1	
Any Administered Antiepileptics (anticonvulsants): (Carbamazepine, Divalproex Sodium, Ethotoin, Ethosuximide, Felbamate, Fosphenytoin, Gabapentin, Lamotrigine, Mephenytoin, Methsuximide, Phenobarbital, Phenytoin, Primidone, Topiramate, Trimethadione, Valproic Acid) ¹	2	
Any Administered Benzodiazepines: ² (Alprazolam, Chloridiazepoxide, Clonazepam, Clorazepate Dipotassium, Diazepam, Flurazepam, Halazepam ³ , Lorazepam, Midazolam, Oxazepam, Temazepam, Triazolam)	1	
Get-Up-and-Go Test: "Rising from a Chair"		
If unable to assess, monitor for change in activity level, assess other risk factors, document both on patient chart with date and time.		
Ability to rise in single movement - No loss of balance with steps	0	
Pushes up, successful in one attempt	1	
Multiple attempts but successful	3	
Unable to rise without assistance during test	4	
If unable to assess, document this on the patient chart with the date and time.		
(A score of 5 or greater = High Risk)	TOTAL SCORE	
© 2012 AHI of Indiana, Inc. All rights reserved. United States Patent No. 7,282,031 and U.S. Patent No. 7,682,308. Reproduction of copyright and patented materials without authorization is a violation of federal law.		

On-going Medication Review Updates:

1. Levetiracetam (Keppra) was not assessed during the original research conducted to create the Hendrich Fall Risk Model. As an antiepileptic, levetiracetam does have a side effect of somnolence and dizziness which contributes to its fall risk and should be scored (effective June 2010).

2. The study did not include the effect of benzodiazepine-like drugs since they were not on the market at the time. However, due to their similarity in drug structure, mechanism of action and drug effects, they should also be scored (effective January 2010).

3. Halazepam was included in the study but is no longer available in the United States (effective June 2010).

APPENDIX B
Patient File Data Dictionary

Variable	Definition
Key Number	A unique identifier for each patient
Gender	Gender of the patient; valid values are Female or Male
BMI	is a measure of body fat based on height and weight that applies to adult men and women
Patient Age at Surgery	Chronological age of the patient in years at the time of the surgery
Surgery Year	The year the surgery was performed
Primary Procedure Name	The type of surgery that was performed (i.e., hip or knee replacement)
Falls	An indicator to denote if the patient fell after surgery
Injury	An indicator to denote if the patient was injured because of a post-surgical fall
Location	Indicates where the fall took place (Internal is hospital, External is other than hospital)
90 Day Complication	Indicates if complications occurred within 90 days after the surgery
Subsequent Readmission	Indicates if the patient was readmitted to a hospital after the surgery
Subsequent_ED_Visit	Indicates if the patient had a subsequent visit to an emergency department after the surgery
Comorbidity Code	A unique code that defines a specific condition that impacts a patient

APPENDIX C
Medicine by Category

Variable	Definition
Key Number	A unique identifier for each patient
Medication Name	Name of the medication
Dose	Amount of medication administered
Route	Indicates how the medication was administered
Category	Group that the medication name belongs within

Teaching Case

Robotic Process Automation Overdue Collections Case

Bryant Richards
Bryant.Richards@nichols.edu

Nicholas Kolodziejczak
Nicholas.Kolodziejczak@nichols.edu

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

Kerry Calnan
KCalnan@springfieldcollege.edu
Springfield College
Springfield, MA USA

Abstract

Organizations are keenly interested in employees trained in business analysis and robotic process automation. This is because those methods are growing in demand for improving efficiency in business processes. Business students exiting college must understand how to map out business processes and implement automation fundamentals to be competitive in the job market. In this hands-on project, students will analyze a real-world situation and utilize NICE Automation Studios to automate the process.

Keywords: Robotic Process Automation, RPA, Automation, Business Process Improvement, Business Process Reengineering

1. INTRODUCTION

Imagine you are on your way to your first job after college; you are not looking forward to the day because it is the day you have to send emails to customers who are overdue with their payments to your company. This is one of the mundane tasks you wish you did not have to do, which is why they give it to the new employee. But today, you will automate that task using RPA (Robotic Process Automation). RPA will complete the tasks with a few clicks, savings hours to be allocated on other essential activities.

The world is full of problems to be solved with technology. The future is full of opportunities for those with the right skills. According to *Deloitte Insights – Automation with Intelligence* (Watson, Schaefer, Wright, Witherick, Horton, Polner, & Telford, T.W, 2020), the leading skills needed within industry are uniquely human skills such as business process analysis, complex problem-solving, and cognitive abilities. These skills ranked significantly higher than traditional technology skills, such as system, content, and technical skills.

Startlingly, many technology implementations do not yield benefits. According to Forbes, 90% fail to achieve quantifiable returns. The top reason cited for these alarming statistics was the lack of talent available within an organization (Andriole, 2021).

This case study will allow you to practice those uniquely human skills: problem-solving and critical thinking. As with most business challenges, there are many possible solutions. You will be limited only by your available technology (NICE Automation Studio and Microsoft Office), your ability to analyze and understand the problem, and your creativity.

Consistent with real-world technology problems, you will find that this case requires process and data analysis work. These skills are critical to the success of any implementation.

Since automation solutions are typically delivered using an agile methodology, this case will expose you to the variety of skills needed for a successful implementation through rapid development cycles. On a personal note, you will find that this case provides you with specific insights into two different careers: automation development and business analysis. Both roles offer strong salaries and excellent career paths and are in high demand. In the United States, RPA developers and business analysis roles, as of June 2022, had average annual salaries of \$99,096 (*Glassdoor*, 2021b) and \$84,125 (*Glassdoor*, 2021a), respectively.

2. LEARNING OUTCOMES

Before starting this project, you should be familiar with the following: first, you should have a basic knowledge of Microsoft Excel and Outlook, and second, if you are new to the world of RPA, then you are encouraged to scan through and review some of these resources:

- Everest Group Blogs (*Everest Group Reports* - View, n.d.)
- NICE YouTube Channel (*NICE* - YouTube, n.d.)

Finally, if you have not used NICE Automation Studio before, complete the 10 practice modules found in Appendix C, Practice Modules. Pay particular attention to *Project 9: Generate Emails from Excel Data*, as it is similar to this case.

After completing this project, you will be able to:

- Demonstrate a basic understanding of business analysis as evidenced by creating:
 - A problem summary
 - A basic flowchart
 - A data dictionary
 - Instructions for developing an automation
- Solve a business problem using automation by:
 - Automating data collection across multiple Excel files
 - Automating sending customized email
 - Utilizing various rules and procedures to interpret data

3. PROJECT DESCRIPTION

In this project, you will utilize an RPA software, NICE Automation Studios, to automate the Overdue Collections process. You will analyze the case study, identify necessary data by compiling it into a data dictionary, create a flowchart of the To-Be state of the process, create instructions, and execute the solution. At the conclusion of this case, you will walk through your project to demonstrate your automation, highlighting your understanding of the entire business process and your automation of that process.

Deliverables

Each step of the project will have a specific deliverable, which will help advance you to the next step of the project. In addition, these deliverables represent realistic outcomes that could be expected of you performing this role within industry. The deliverables are as follows:

Step 1: One-page problem summary

Step 2: Data dictionary

Step 3: Flowchart of the future state

Step 4: Solution instructions

Step 5: Automation

Step 6: Video presentation of automated solution

As projects become more complex, the steps are even more critical, as incomplete or skipped steps can lead to significant downstream problems in solution development. While everyone is tempted to skip straight to problem-solving and automation development, avoid this temptation as it can lead to significant debugging challenges and missed requirements.

Discussion Questions

Each project step ends with questions for you to consider. Discuss the questions with your group, prepare responses for class discussion, or share your responses in a format as specified by your instructor. You will likely find the discussions an interesting and valuable exercise to help with overall comprehension. You will learn significantly from your peers while getting the opportunity to test new vocabulary and budding insights. This case doesn't have one "right" answer, and you will have the flexibility to customize your solution. Sharing ideas with your peers can lead to ideation resulting in a better end product.

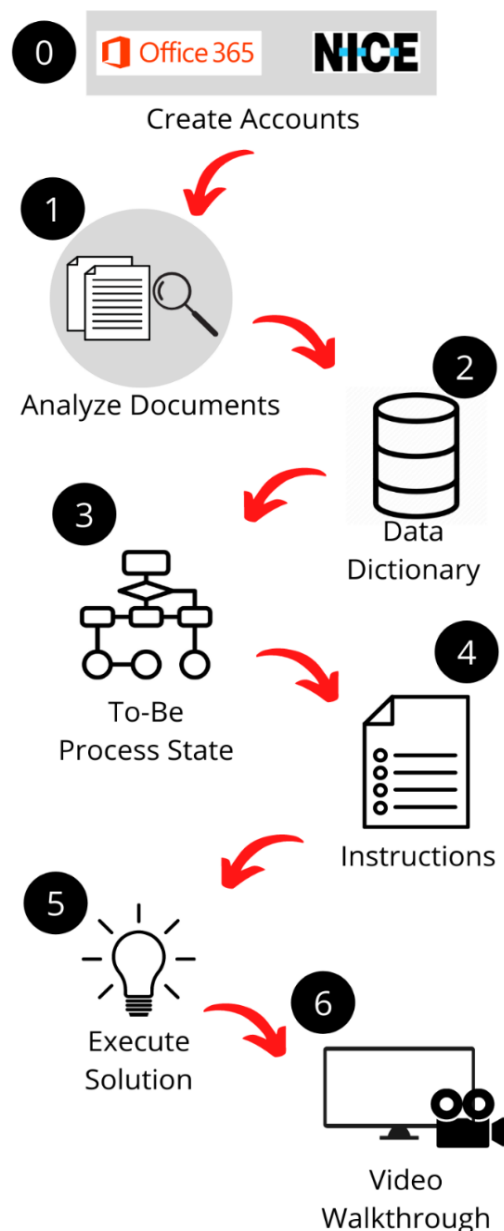


Figure 1. Workflow

Begin by familiarizing yourself with Figure 1. This provides an overview of the workflow and data flow for this project. Your solution will follow these steps:

Setup step (Step 0): Create applicable accounts if needed.

Step 1: Analyze the problem which, in this example, is the case document in Appendix A. Understand the client's current process and automation requests. Examine the Excel documents to gain familiarity with the data.

Step 2: Identify necessary data and compile it into a data dictionary.

Step 3: Document the high-level future state of the process.

Step 4: Create instructions on the proposed solution for the case.

Step 5: Execute the solution.

Step 6: Create a video walkthrough to demonstrate understanding of the project.

Now let us walk through each step in more detail.

Setup Step (Step 0 on Figure 1). Create Accounts

NICE Automation Studios primarily uses Office 365 applications in its functions. In this project, you will be extracting data from Excel, analyzing it, and using Outlook to send emails automatically. Both applications must be installed on the computer before building the automation.

NICE Automation Studios has a free trial version of its software. Rather than using an application on your desktop, open it through your web browser. Follow the steps on the website: <https://info.nice.com/RPA-Free-Trial.html> to begin your trial.

Things to watch out for

When this case was developed, Automation Studios did not support Mac or Chromebooks. If this is still the case and you use one of those devices, then you will need to set up a virtual desktop or use another device. If you do not have another device, talk with your instructor about whether there are suitable labs on your campus.

Discussion Questions

1. Looking at Automation Studios' free trial webpage, what useful features does this tool offer?

Name	Excel Sheet Name	Column Location	Description	Data Type	Variable Type
Customer Address	Customer Data	D	Where the customer lives	string	List of text
Customer Account Balance	Unpaid Account Data	A	The amount a customer still owes the business	number	List of numbers

Table 1. Data Dictionary Example

2. What steps or processes in your daily life, or current/past jobs, could this tool automate? In other words, how might it be used?

Step 1. Analyze the Problem (Appendix A RPA Overdue Collection Case)

Begin this step by reading through the case description (See Appendix A). When reading through the case description, note the company's current process. After reading the automation request, determine how you plan to meet each feature.

Next, examine the Excel workbook (Appendix B). Explore the columns and make sure you understand the data titles and what each column represents. Identify the data that will need to be extracted from each report (hint: you will need data from both reports to complete this case).

After reviewing the case and supporting documentation, write a concise summary of the business problem. Include the key details needed to solve the problem and a high-level assessment of the benefits to the stakeholders. This may feel a little redundant as much of what is necessary is already included in the case, but typically you will not be given such documentation. Instead, you would generate the documentation after talking with the business users. Take the time to highlight precisely what data and documents are needed.

The following is a recommended outline to help organize your problem summary.

- I. Problem statement
- II. Stakeholders
- III. Benefits/Opportunities
- IV. Risks/Challenges
- V. Data/Documentation needed.

Deliverable

Create and submit your completed problem summary.

Discussion Questions

1. What are the benefits and disadvantages of a business such as GE Appliances utilizing RPA?
2. What other stakeholders could be impacted that are not mentioned in the case? What might be the impact on them?
3. What risks exist to automating the process as requested? In what situation may these risks outweigh the benefits?

Step 2. Identify necessary data and compile it into a data dictionary.

A data dictionary (see Table 1 above) is a table that captures the critical data and describes it in a fashion that is useful when evaluating a process and designing an automation. After examining the RPA overdue collections case document and Excel workbook, explore the data in more detail to identify the format and type of data. For data type, determine whether the data is a string (i.e., alphanumeric data) or numeric (i.e., numbers only). Using the variable types from NICE Automation Studio, select the variable type you believe best matches your data.

Create a data dictionary similar to Table 1. Recommended column headers include, but are not limited to, Name, Excel Sheet Name, Column Location, Description, Data Type, and Variable type.

Things to watch out for

Not all of the data is necessary, but some data will enhance the quality of your solution. Data type matters in variable selection, so have a clear idea of each needed data type.

Deliverable

Create and submit your data dictionary. Ensure it includes only the data needed for the automation (Hint – not all the fields in both sheets are required for this case).

Discussion Questions

1. What data points may not be necessary for this particular solution but could be

helpful to solve other potential problems or create other valuable automations?

2. What is the impact of incorrectly capturing data in your process?
3. What could be done within the process that would result in a change to the data dictionary?

Step 3. Document future process state

A future process state is a flowchart showing the steps to solve a business problem or implement a new technology solution. In this step, you will create a flow chart that documents the future process state. It will often require that you start with an understanding of the current process and then add, change and/or delete steps to design a new process. A strong future state process should have all the steps needed for the solution but nothing extra. It will provide the organization with a roadmap for implementing the change and a guide for the developer on how best to create the automation instructions.

A variety of applications are available to create flow charts. Some high-quality applications include Microsoft Visio and Lucid Charts, but this can also be done by hand if you don't have access to either of those tools.

Before you begin, you must know the standard process flow notations.

Notations

Notations showcase steps within a flowchart. Different shapes represent different types of steps. Figure 2 illustrates some of the most common notations recommended for this project.

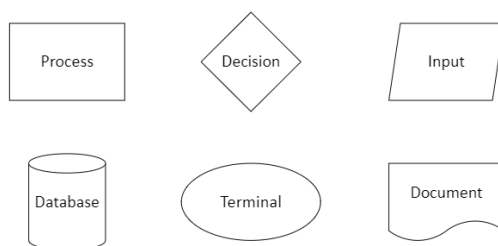


Figure 2. Standard Flowchart Notations

Organization

Flows should run left to right and top to bottom of the page. This ensures that another person can easily read and understand the flowchart. See Figure 3 for a sample flowchart.

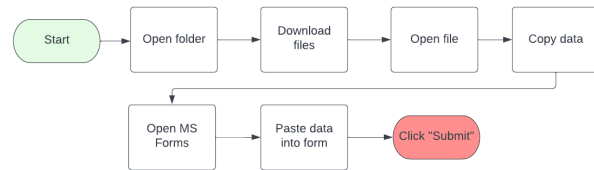


Figure 3. Flowchart Example

Labeling

When labeling each step, the best practice is to use the "verb-noun" format. For example, a step about moving data out of the original database and into Excel could be labeled "Export Data" or "Export Data to Excel". Labels should be as short as possible without losing the meaning of the step. See figure 3 for more label examples.

Process vs. Task Level Steps

A task is a piece of work needed to be completed, while a process is a group of tasks done for a particular purpose.

When choosing steps to include in your flowchart, it may help to begin by analyzing the task-level steps. Make a note of any instances where data moves or changes. Then summarize these instances from the process level in your flowchart.

For example, assume you copy columns A, B, and C from one Excel sheet. These are all instances where data is moved in the same manner. You can summarize these instances in one step rather than including three steps in your flowchart.

When unsure whether a step should be included in your flowchart, ask yourself, does someone reading my flowchart need this information to understand the flow? And then only include those that are necessary for comprehension.

Stakeholders

Stakeholders are often defined in terms of interest in, impact on, and influence over the change requested. They are grouped based on their relationship to the needs, changes, and solutions. Identifying stakeholders helps demonstrate value, context, and understanding.

Key questions to ask:

1. Whom is the stakeholder involved?
2. How are they involved?
3. What is their engagement going to be?
4. How are they needed to support this initiative?

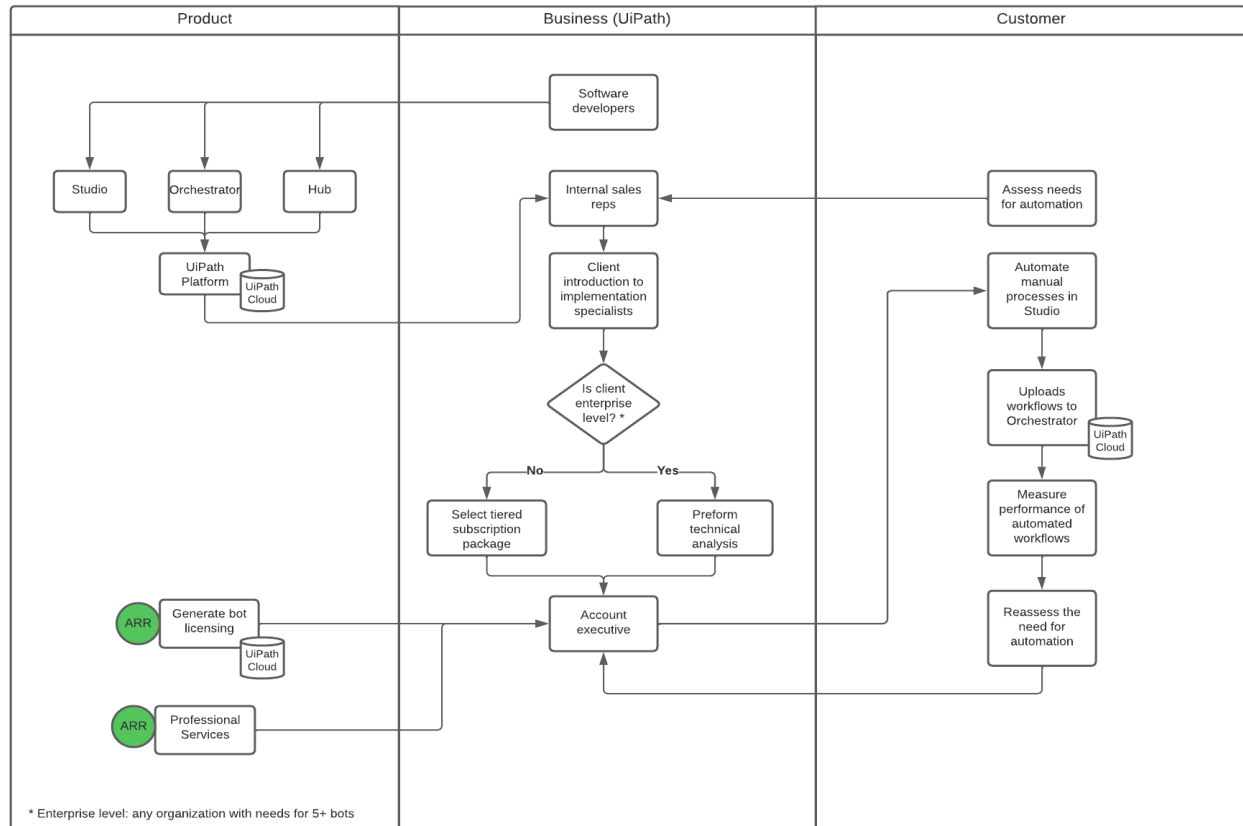


Figure 4. Swim Lanes

Swim Lanes

Swim lanes refers to cross-functional flowcharts. Swim lanes divide the flowchart into horizontal or vertical lanes, representing the tasks done by one role, function, or department; after identifying your stakeholders, label each swim lane column with the appropriate stakeholder. You will find many ways to do this effectively, depending on the process you are documenting.

They help provide more information on who is involved in the process. Their drawback is that they require more space in your charting application.

Things to watch out for

Figure 4 is an example of a completed solution. When documenting the future process state, ensure you analyze it from a high-level overview. You may feel tempted to add extra details, but more detailed information will be required for Step 4.

The flowchart is a tool to help explain your ideas to someone unfamiliar with the process. Be sure to keep your audience in mind.

Students unfamiliar with business process modeling notation (BPMN) including flowcharts and swimlanes should refer to White (2004).

Deliverable

Create and submit a future state flowchart.

Discussion Questions

1. There are multiple solutions to this case. Can you recommend one improvement to include in your solution? Why do you think your new solution is better than your first one?
2. What implementation challenges could arise from different solutions?

Step 4. Create solution instructions

Based on your flowchart, create a list of instructions. The flowchart focuses on the high-level overview of the process, while the instructions help capture the more intricate details.

One way to approach this step is to write the instructions thoroughly enough so another student can build the automation using your instructions.

An example set of solution instructions may look like the following, which are the first 6 steps of instructions for an automation to collect the price of Tesla's stock (TSLA) and chart the changes:

1. Start by creating a blank Excel file with pre-defined fields 'highest price' and 'lowest price.'
2. Open the stock market website (finance.yahoo.com)
3. Enter TSLA in the Quote Search field
4. The automation extracts the High and Low Value from the "Day's Range" and updates the Excel
5. The automation extracts values from the website after the market close and compares them with the values present in the existing values within Excel.

Things to watch out for

Remember to go back and modify your flowchart if you realize a step was incorrect or missing. Flowcharting is often an iterative process.

Deliverable

Create and submit the automation instructions. Submit a revised flowchart if changes were made to the flowchart.

Discussion Questions

1. What about creating your instructions helped you shape your solution?
2. Do your automation instructions tie perfectly to your future state flowchart? Why or why not? Which has more detail?
3. If you were hiring someone to automate something for you, would you want them to have more or less detail in their automation instructions? Why?

Step 5. Execute solution

Using your flowchart and instructions as a guide, create your automation within NICE Automation Studio. One method to approach this step is to develop and test your automation in small pieces. This methodology is a best practice as it minimizes the debugging process.

The Automation Studio Help Center can significantly help a new developer. Use the link below to access helpful resources.

<http://help.nice-automation.com/content/home.htm>

Things to watch out for

You may find that you must deviate from your instructions to build the automation successfully.

Be sure to note any changes you make and evaluate the implications to the rest of your automation before completing your design. As mentioned, frequent testing along the way will save you from more prolonged and more complex debugging later.

Also, you may find that your automation does not work completely. Imperfect automations can provide equally valuable learning experiences. Demonstrating the working components and isolating the issue(s) can help showcase the strength of your automation and deepen your understanding of the process.

Deliverable

Create and submit the automation files (both files .dproj and .resx). If changes were made to any of the other documents, then submit those revised documents as well.

Discussion Questions

1. What challenges did you encounter during the development of your automation? How did you handle these challenges?
2. What would you do differently to make your automation development easier?

Step 6. Video Walkthrough

Create a 5-8 minute video walkthrough of the automation following the outline below:

1. Introduction: Introduce yourself and your project
2. Problem description: Describe the problem that you intend to solve. Think back to your one-page summary.
3. Data: Show your data dictionary and talk about the data you needed from the process and where it was stored.
4. Process change: Using your future state flowchart, describe how your automation changed the process.
5. Instructions: Walk through the automation instructions to explain your design. Extra points for those who can do this while walking through the automation within the design panel of the Automation Studio.
6. Demonstration: Press run and record the automation working. Show the automation results, including the emails and any changes to the data.

When creating and submitting a video, use a software program that is compatible with your learning management system (LMS). Many programs work very well. Below are the steps if

using the Studio function, which is an option in Canvas. Many free screen capture tools are available if you do not have Canvas as an LMS including OBS Studio, Wondershare, and Screen-O-Matic. Before starting the steps below, make sure to have all screens open that you want to show.

1. Access the Canvas dashboard and find the TV-shaped icon. Placing your cursor over it will show you the "Studio" label (see Figure 5).
2. On the top right, you will see the record button (see Figure 6). Click once to begin.

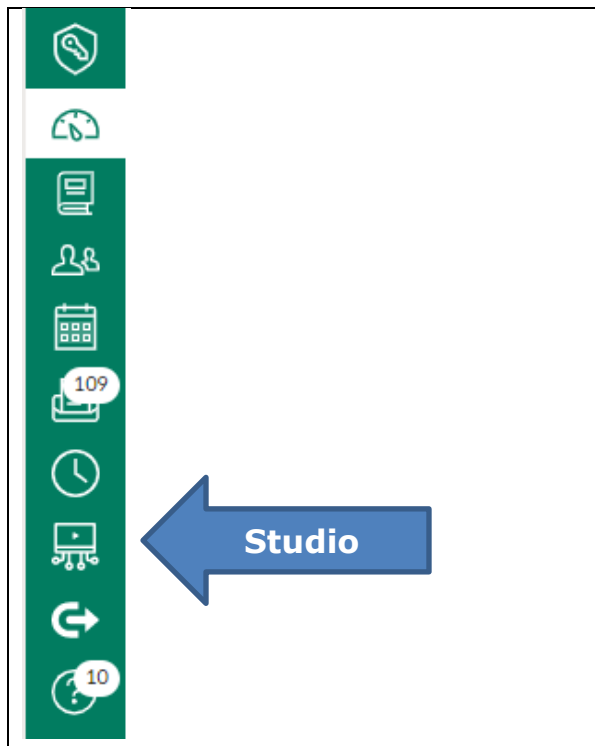


Figure 5. Canvas Studio

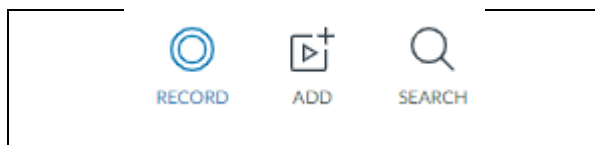


Figure 6. Canvas Studio Record

3. Once clicked, the Record button will offer you the choice of "Screen Capture" or "Webcam Capture." Select "Screen Capture" (see Figure 7).

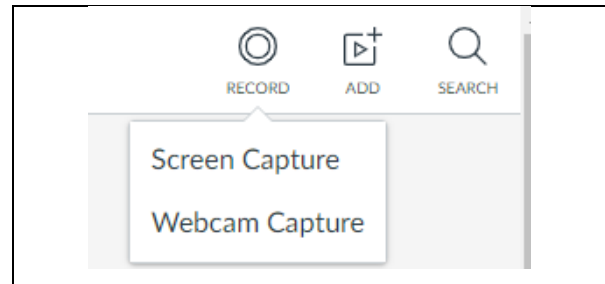


Figure 7. Canvas Studio Capture

4. Studio will open up a dashed box that you can adjust to your screen size. Make sure to start on the first screen you want to record.
5. Press Record (see Figure 8). Press the pause button when finished. Press the done button to complete the upload.
6. Your video can be uploaded in an assignment or shared from "My Library," which is the original screen you see when you first enter Studio in step 1.

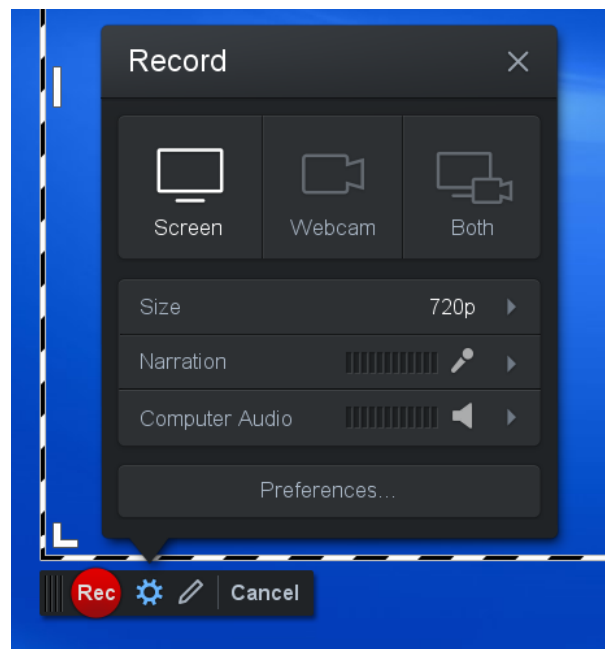


Figure 8. Canvas Studio Record

Things to watch out for

Keep your audience in mind. In your explanations, imagine you are talking to the client of the case study. Expect that they will not have a deep knowledge of Automation Studio.

Deliverable

Create and submit your video walkthrough recording.

Discussion Questions

1. Upon reflection, which steps were the most challenging during the project? Why?
2. Which steps are most important for convincing the company that your automation would solve their problem? Why? Given more time, what could you have done differently to make your presentation more compelling?

4. CONCLUSIONS AND NEXT STEPS

By completing this exercise, you have demonstrated readiness for a future role in automation. Your accomplishments, listed below, are worthy of discussion in job interviews and inclusion in your resume, as many employers are looking for this type of employee experience.

- Demonstrated a basic understanding of business analysis as evidenced by creating:
 - A problem summary
 - A basic flowchart
 - A data dictionary
 - Instructions for developing an automation
- Solved a business problem using automation by:
 - Automating data collection across multiple Excel files
 - Automating sending customized email
 - Utilizing various rules and procedures to interpret data

A solid foundation in business analysis and automation development can prepare you for roles specific to each or provide a deeper understanding of technology deployment challenges within industry.

For those looking to explore a deeper learning experience, there are three valuable areas to consider: 1) deeper analysis and development, 2) broader applicability of solutions, and 3) analysis of benefits.

Deeper analysis and development allow you to explore more levels of reality within the case by anticipating the needs of typical stakeholders and providing solutions. As you can imagine, this scenario is not realistic and is just the first step in fleshing out the entire collections process.

Considering that the case requirements stop short in many ways, which is often common within

industry, careful evaluation of business processes will allow you to flush out many more potential requirements and solutions. For example, you could go into more depth with the email solution of the automation by anticipating the best frequency (hourly, daily, weekly, etc.), the most stakeholder-friendly text within the emails, and what reporting could be helpful to develop and include in each email.

Broader applicability will allow you to explore different applications given what you know about the company and what other possible processes could benefit from the solution (or portions of the solution) developed in this case study. Companies are filled with processes that require data collection, rules-based comparison, or organization and dissemination via email.

Benefits analysis is helpful in understanding and communicating the value of technology solutions. The case provides little information that makes this easy for you to perform. However, simulating steps and estimating times will offer you baseline time studies. In addition, careful consideration of typical customer and employee behaviors will help you identify other key benefits of your solution. Some questions you might consider are as follows:

- Do employees like this task? Do they always perform it perfectly? What happens if an employee is out sick?
- Will customers be more likely to pay if reminded regularly?
- What benefits will the company receive by having a better collection process? Will they have less bad debt? Will the CFO make better decisions?

5. REFERENCES

Andriole, S. (2021, March 25). *3 Main Reasons Why Big Technology Projects Fail – & Why Many Companies Should Just Never Do Them*. Forbes. <https://www.forbes.com/sites/steveandriole/2021/03/25/3-main-reasons-why-big-technology-projects-fail---why-many-companies-should-just-never-do-them/>

Everest Group Reports—View. (n.d.). Retrieved June 14, 2022, from <https://www2.everestgrp.com/reportaction/EGR-2021-38-R-4655/Marketing?SearchTerms=Peak%20Matrix%20for%20RPA>

Glassdoor (2021a). Salary: Business Analyst.
https://www.glassdoor.com/Salaries/business-anlayst-salary-SRCH_K00,16.htm

Glassdoor (2021b). Salary: Rpa Developer.
https://www.glassdoor.com/Salaries/rpa-developer-salary-SRCH_K00,13.htm

NICE - YouTube. (n.d.). Retrieved June 14, 2022,
from <https://www.youtube.com/c/NICESystems>

Watson, J., Schaefer, G., Wright, D., Witherick, D., Horton, R., Polner, A., & Telford, T. (2020). *Automation with Intelligence, Pursuing organisation-wide reimagination* (p. 32). Deloitte. https://www2.deloitte.com/content/dam/insights/us/articles/73699-global-intelligent-automation-survey/DI_Automation-with-intelligence.pdf

White, S. (2004) *Introduction to BPMN*. IBM Corporation

.

APPENDIX A

Case Description

Case study description: RPA Overdue Collections

Background: GE Appliances is an American home appliance manufacturer based in Louisville, Kentucky. The company has been struggling to get customers to pay their unpaid balances in recent years. GE's current process to track customers is manually intensive. GE believes automation will solve their issues to save time and improve employee morale.

Current Process: The process begins when a customer purchases an appliance on credit. A sales representative logs the customer's contact and balance information in separate Excel files. Periodically, a sales representative checks the customer balance Excel file to look for unpaid accounts. Paid accounts are removed from the file. Suppose the account is unpaid and overdue by LESS than 90 days. In that case, the sales representative must get their contact information from the customer contact information Excel sheet, then send the customer an email regarding the unpaid account. Suppose the account is unpaid and overdue by OVER 90 days. In that case, the customer's account information is emailed to the Chief Financial Officer (CFO) to write the balance off as an allowance for doubtful accounts.

Automation: GE Appliances would like you to automate many of the following features.

- Extract data from Excel sheets into a UDT (User Defined Type)
- Determine if payment is overdue and by how many days (ex. payment due 7 days ago)
- Send customized emails to account holders based on how many days until payment is due/how many days payment is past due. (The more customized the email is, the better!)
- Forward the account information 90 days overdue to the Chief Financial Officer to write the account off as an allowance for doubtful accounts.

APPENDIX B

Excel Spreadsheets

Unpaid Balance Report

It contains information relating to customers with unpaid balances.

customer_name	balance_owed	Payment_Due_Date	Todays_Date	Days_Between_Dates
Benton, John B Jr	391	4/30/2022	4/29/2022	1
Sport En Art	432	3/30/2022	4/29/2022	-30
Professional Image Inc	587	5/6/2022	4/29/2022	7
Rangoni Of Florence	678	4/11/2022	4/29/2022	-18
Post Box Services Plus	682	4/28/2022	4/29/2022	-1
Century Communications	959	1/28/2022	4/29/2022	-91
Feiner Bros	1673	5/21/2022	4/29/2022	22

Customer Contact Report (Simulated Data)

It contains the contact information of all of GE Appliances' customers.

first_name	last_name	Full_Name	company_name	address	city	county	state	zip	City_ST_Zip	phone_1	phone_2	email	web
James	Butt	James Butt	Benton, John B Jr	6649 N Blue Gum St	New Orleans	Orleans	LA	70116	New Orleans, LA 70116	504-621-8927	504-845-1427	jbutt@gmail.com	http://www.bentonjohnbjr.com
Josephine	Darakjy	Josephine Darakjy	Chanay, Jeffrey A Esq	4 B Blue Ridge Blvd	Brighton	Livingston	MI	48116	Brighton, MI 48116	810-292-9388	810-374-9840	josephine_darakjy@darakjy.org	http://www.chanayjeffreyaesq.com
Art	Venere	Art Venere	Chemel, James L Cpa	8 W Cerritos Ave #54	Bridgeport	Gloucester	NJ	8014	Bridgeport, NJ 8014	856-636-8749	856-264-4130	art@venere.org	http://www.chemeljamescpa.com
Lenna	Paprocki	Lenna Paprocki	Feltz Printing Service	639 Main St	Anchorage	Anchorage	AK	99501	Anchorage, AK 99501	907-385-4412	907-921-2010	lpaprocki@hotmail.com	http://www.feltzprinting-service.com
Donette	Foller	Donette Foller	Printing Dimensions	34 Center St	Hamilton	Butler	OH	45011	Hamilton, OH 45011	513-570-1893	513-549-4561	donette.foller@cox.net	http://www.printingdimensions.com
Simona	Morasca	Simona Morasca	Chapman, Ross E Esq	3 Mcauley Dr	Ashland	Ashland	OH	44805	Ashland, OH 44805	419-503-2484	419-800-6759	simona@morasca.com	http://www.chapmanrosseeesq.com
Mitsue	Tollner	Mitsue Tollner	Morlong Associates	7 Eads St	Chicago	Cook	IL	60632	Chicago, IL 60632	773-573-6914	773-924-8565	mitsue_tollner@yahoo.com	http://www.morlongassociates.com
Leota	Dilliard	Leota Dilliard	Commercial Press	7 W Jackson Blvd	San Jose	Santa Clara	CA	95111	San Jose, CA 95111	408-752-3500	408-813-1105	leota@hotmail.com	http://www.commercialpress.com
Sage	Wieser	Sage Wieser	Truhlar And Truhlar Attys	5 Boston Ave #88	Sioux Falls	Minnehaha	SD	57105	Sioux Falls, SD 57105	605-414-2147	605-794-4895	sage_wieser@cox.net	http://www.truhlarandtruhlarattys.com
Kris	Marrier	Kris Marrier	King, Christopher A Esq	228 Runamuck Pl #2808	Baltimore	Baltimore	MD	21224	Baltimore, MD 21224	410-655-8723	410-804-4694	kris@gmail.com	http://www.kingchristopheraesq.com
Minna	Amigon	Minna Amigon	Dorl, James J Esq	2371 Jerrold Ave	Kulpsville	Montgomery	PA	19443	Kulpsville, PA 19443	215-874-1229	215-422-8694	minna_amigon@yahoo.com	http://www.dorljamesjesq.com

Abel	Macl ead	Abel Maclead	Rangoni Of Florence	37275 St Rt 17m M	Middle Island	Suffolk	N Y	11 95 3	Middle Island, NY 11953	631- 335- 3414	631- 677- 3675	amaclead@gmai l.com	http://www.rangonio fflorence.com
Kiley	Calda rera	Kiley Caldarera	Feiner Bros	25 E 75th St #69	Los Angeles	Los Angeles	CA	90 03 4	Los Angeles, CA 90034	310- 498- 5651	310- 254- 3084	kiley.caldarera@ aol.com	http://www.feinerbro s.com
Graci ela	Ruta	Graciela Ruta	Buckley Miller & Wright	98 Connecticut Ave Nw	Chagri n Falls	Geaug a	O H	44 02 3	Chagrin Falls, OH 44023	440- 780- 8425	440- 579- 7763	gruta@cox.net	http://www.buckleym illerwright.com
Cam my	Albar es	Cammy Albares	Rousseaux, Michael Esq	56 E Morehead St	Laredo	Webb	TX	78 04 5	Laredo, TX 78045	956- 537- 6195	956- 841- 7216	calbares@gmail. com	http://www.rousseau xmichael.esq.com
Matti e	Poqu ette	Mattie Poquette	Century Communicati ons	73 State Road 434 E	Phoeni x	Marico pa	AZ	85 01 3	Phoenix, AZ 85013	602- 277- 4385	602- 953- 6360	mattie@aol.com	http://www.centuryc ommunications.com
Meag han	Garuf i	Meaghan Garufi	Bolton, Wilbur Esq	69734 E Carrillo St	Mc Minnvi lle	Warre n	T N	37 11 0	Mc Minnville, TN 37110	931- 313- 9635	931- 235- 7959	meaghan@hotmail. com	http://www.boltonwil buresq.com
Gladys	Rim	Gladys Rim	T M Byxbee Company Pc	322 New Horizon Blvd	Milwa ukee	Milwau kee	W I	53 20 7	Milwaukee, WI 53207	414- 661- 9598	414- 377- 2880	gladys.rim@rim. org	http://www.tmbyxbe ecompanypc.com
Yuki	Who brey	Yuki Whobrey	Farmers Insurance Group	1 State Route 27 394	Taylor	Wayne	MI	48 18 0	Taylor, MI 48180	313- 288- 7937	313- 341- 4470	yuki_whobrey@ aol.com	http://www.farmersi nsurancegroup.com
Fletc her	Flosi	Fletcher Flosi	Post Box Services Plus	Manchester Blvd	Rockfo rd	Winne bago	IL	61 10 9	Rockford, IL 61109	815- 828- 2147	815- 426- 5657	fletcher.flosi@y ahoo.com	http://www.postboxs ervicesplus.com
Bette	Nicka	Bette Nicka	Sport En Art	6 S 33rd St	Aston	Delawa re	PA	19 01 4	Aston, PA 19014	610- 545- 3615	610- 492- 4643	bette_nicka@co x.net	http://www.sportena rt.com
Vero nika	Inouy e	Veronika Inouye	C 4 Network Inc	6 Greenleaf Ave	San Jose	Santa Clara	CA	95 11 1	San Jose, CA 95111	408- 540- 1785	408- 813- 4592	vinouye@aol.co m	http://www.cnetwork inc.com
Willard	Kolm etz	Willard Kolmetz	Ingalls, Donald R Esq	618 W Yakima Ave	Irving	Dallas	TX	75 06 2	Irving, TX 75062	972- 303- 9197	972- 896- 4882	willard@hotmail .com	http://www.ingallsdo naldresq.com
Mary ann	Royst er	Maryann Royster	Franklin, Peter L Esq	74 S Westgate St	Albany	Albany	N Y	12 20 4	Albany, NY 12204	518- 966- 7987	518- 448- 8982	mroyster@royst er.com	http://www.franklinp eterlesq.com

Alisha	Slusarski	Alisha Slusarski	Wtlz Power 107 Fm	3273 State St	Middlesex	Middlesex	NJ	88 46	Middlesex, NJ 8846	732-658-3154	732-635-3453	alisha@slusarski.com	http://www.wtlzpowerfm.com
Allen	Iturbide	Allene Iturbide	Ledecky, David Esq	1 Central Ave	Steven's Point	Portage	WI	54 48 1	Stevens Point, WI 54481	715-662-6764	715-530-9863	allene_iturbide@cox.net	http://www.ledeckydavidesq.com
Chanel	Caudy	Chanel Caudy	Professional Image Inc	86 Nw 66th St #8673	Shawnee	Johnson	KS	66 21 8	Shawnee, KS 66218	913-388-2079	913-899-1103	chanel.caudy@ccaudy.org	http://www.professionalimageinc.com
Ezekiel	Chui	Ezekiel Chui	Sider, Donald C Esq	2 Cedar Ave #84	Easton	Talbot	MD	21 60 1	Easton, MD 21601	410-669-1642	410-235-8738	ezeziel@chui.com	http://www.siderdonaldcesq.com
Willow	Kusko	Willow Kusko	U Pull It	90991 Thorburn Ave	New York	New York	NY	10 01 1	New York, NY 10011	212-582-4976	212-934-5167	wkusko@yahoo.com	http://www.upullit.com

APPENDIX C

Practice Modules

Completing each automation helps develop confidence and capability with NICE Automation Studio. Sample Project 9 is the one automation that could be most easily and directly applied to this case for creating a possible solution. Practice modules take between 15 minutes to 2 hours for most students to complete, with the average development time of about an hour per automation.

Module	Description	Link
Workflows and Variables	Manipulates the values of variables.	http://help.nice-automation.com/content/topics/gettingstarted.htm
Callouts and Event Handlers	Teaches you the basics around callouts and event handlers.	http://help.nice-automation.com/content/topics/gettingstarted.htm
Soccer Statistics	Copies and pastes soccer players' statistics from an internet browser and inputs the data into Excel. This automation teaches callouts and data entry to Excel. Creates macros within Excel that will organize the data and sorts the data alphabetically by the player's name. This automation teaches macro usage in Excel and how to run the macros in NICE Automation Studio.	https://www.youtube.com/watch?v=qLx2s2urGuU https://www.youtube.com/watch?v=R01krYJJKMo
Football Statistics	Collects NFL player statistics from an internet browser and inputs the data into Excel. Formats the data you collected from the NFL player statistics. The macros involved with this automation will involve conditional formatting, highlighting the data for any statistics over 10%, and creating graphs for each statistic.	https://www.youtube.com/watch?v=8dN8jBVmLjQ https://www.youtube.com/watch?v=q6I_sLmLXvE
YouTube Analytics	Allow you to track YouTube likes and views for 5 videos you choose based on personal preference. This automation will teach you to capture YouTube likes and views and append the data to the Excel file. Allow you to convert your text data into numbers and create graphs that will be updated each time you run the automation through macros within Excel. After the macros are created, you will run the macros within NICE Automation Studios.	https://www.youtube.com/watch?v=d_Xf4YqNF3I https://www.youtube.com/watch?v=k5AzuOwMGjY
Sample Project 1 - Save new email contents to a text file	The automation will continuously check the Outlook inbox until a new email is received. After the new email is received, the contents of the email are stored in a text file.	http://help.nice-automation.com/content/topics/howtoscenario-listofscenarios.htm

Sample Project 5 - Write Data from PDF Forms into Excel	First, it will collect all PDF forms located in a specific folder. Next, data will be extracted from each form by capturing the data found in specific fields (ex., Firstname, Lastname, phone, etc.). Lastly, this data is outputted to an Excel sheet.	http://help.nice-automation.com/content/topics/howto/scenario-listofscenarios.htm
Sample Project 9 - Generate Emails from Excel Data	Extracts data from columns in multiple Excel sheets. Next, this data is compiled into a User Defined Type (UDT). Lastly, the data in the UDT is used to compose individualized emails.	http://help.nice-automation.com/content/topics/howto/scenario-listofscenarios.htm
Sample Project 10 - Read Data from an XML File	First load an XML file into an XML document variable. The file is put here to extract all the contents from the file. Next, data from the XML file is separated into a list of nodes that hold our data in an unstructured or "messy" form. Lastly, these nodes are parsed/separated into a User Defined Type (UDT) or a list of lists to clean up the data.	http://help.nice-automation.com/content/topics/howto/scenario-listofscenarios.htm
Sample Project 11 - Extract Text Data Using Regular Expressions	The automation extracts data from a text file. This data looks messy, but there is a consistent pattern, so we'll parse through it using a regular expression. After the data is organized, we want to input it into a callout. Lastly, the callout buttons are linked to a user-defined function to make the callout interactive.	http://help.nice-automation.com/content/topics/howto/scenario-listofscenarios.htm