DESIGN SCIENCE RESEARCH: DEVELOPING AND EVALUATING A FINANCIAL AID ANALYTICS SOFTWARE APPLICATION

A Dissertation Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of Computer Science

By

CHANTEL A. PERRY

Colorado Technical University

March 2018

Committee

Fonda Na'Desh, EdD, Chair

Glenn Stout, PhD, Committee Member

Caroline Howard, PhD, Committee Member

March 15, 2018

Abstract

This study aims to develop a list of technological enhancements to a financial aid analytics application (F3A) designed to assist higher education leaders in gaining actionable insights for decision-making with grants optimization. This research utilizes the Design Science Research framework which highlights the design and evaluation of an IT artifact. A focus group of eight participants, mostly leadership at the Kentucky Higher Education Assistance Authority, gathered to view a demonstration of the software application and provide feedback. The study yielded a framework for designing, developing, and enhancing a financial aid optimization application. The analysis of the focus group feedback revealed areas for developers to consider when creating tools, like the F3A, which could progressively be used by higher education administrators in the financial aid sector. The key findings include; a simple and more intuitive user interface, customization of the tool's functionality based on the student population, the ability for the user to select student profiles, and a system with the underlying capability for obtaining new, or refreshing, data.

Keywords: design science research, financial aid; focus group; analytics; software development

Table of Contents

List of Tables

List of Figures

CHAPTER ONE

The demand for financial aid outweighs state funds allotted for education purposes, causing leaders to become increasingly concerned with how to best delegate resources to prevent the liquidation of financial aid programs (Narozhnaya, 2015; Ragland, 2016; Thanh & Haddawy, 2007). A growing number of higher education institutions have utilized analytical tools to substantiate their need for additional funding and to find self-sustaining solutions (Pomeroy, 2014) but, many postsecondary education agencies have struggled with the successful adoption of these methods (Roscorla, 2015) and are still in the beginning stages of using them for strategic planning (Ferreira & Andrade, 2016). A previous survey revealed that less than 10% of college administrators utilize academic analytics for grants management regarding pre- and post- award performance as well as triggering award availability (Goldstein, 2005; Pomeroy, 2014).

The goal of this study was to develop and evaluate a software tool to assist financial aid managers and policymakers in the successful use of analytics for strategic award allocation. To address this goal, a Design Science Research (DSR) methodology was used to conduct this study (Hevner & Chatterjee, 2010). This methodology served as a guide for developing the software tool, assessing the design of the tool, and determining the value of the tool from both an information systems and business use standpoint (Hevner & Chatterjee, 2010).

Due to the rise in college tuition costs, students are incurring an average debt of $35,200 upon graduation (Sparks, 2011). It has been projected that more than 55 million jobs will require a post-secondary education by year 2020 (Carnevale, Smith, & Strohl, 2014), but the escalating amount of debt is a deterrent to many potential college-goers (Sparks, 2011). In an effort to counter the increase in college expenses, state financial assistance programs were created to assist potential student with college affordability, accessibility, and degree attainment (Sjoquist

& Winters, 2012). However, more research needs to be conducted to understand the actual implications these financial aid programs have on postsecondary degree completion (Narozhnaya, 2015) and to assist state aid policymakers in the awards management process (Ragland, 2016).

Chapter 1 introduces the concept of enterprise analytics and describes the challenges encountered with adoption of technology in higher education administration and financial aid management. The research proposes an evaluation of an analytical software tool developed to assist in improving the grants allocation process. This chapter also details the Design Science Research (DSR) methodology selected for the study as well as the assumptions, limitations, and essential terms.

Problem Statement

The problem addressed in this study is the lack of analytics used for strategic decision-making in higher education (Ferreira & Andrade, 2016; Macfadyen, Dawson, Pardo, & Gasevic, 2014; Roscorla, 2015), particularly in state financial aid resource allocation (Goldstein, 2005; Pomeroy, 2014; Thanh & Haddawy, 2007). The intent of developing state monetary aid programs was to lower college costs and thus promote student postsecondary degree attainment (Sjoquist & Winters, 2012), however, research on whether these programs have met this objective is scarce (Narozhnaya, 2015; Ragland, 2016). State-supported grant programs cost billions of dollars to implement, representing approximately twenty-five percent of state education funds (Delaney, 2011), making them one of the most significant resource allocation concerns for higher education leaders (Narozhnaya, 2015; Ragland, 2016; Thanh & Haddawy, 2007).

The use of business intelligence (BI) and data science (DS) has become increasingly important in the creation of analytical solutions (Gartner, Inc., 2012), especially for decision-makers in the higher education administration sector (Business-Higher Education Forum, 2014; National Commission on Higher Education Attainment, 2013). Many organizations are reliant on these tools for increased revenue (Hsinchun, Chiang, & Storey, 2012), strategic planning (Lublinsky, Smith, & Yakubovich, 2013), and daily business operations (Evans, 2015). However, higher education agencies have been slow to adopt analytical solutions (Roscorla, 2015) and are just starting to utilize them for decision-making (Ferreira & Andrade, 2016). Although demonstrating recent signs of advancement, many universities still rely on traditional decision support tools to track student's behaviors, identify at-risk students, and develop means for intervention (Dziuban, Moskal, Cavanagh, & Watts, 2012). Additionally, Pomeroy (2014) concluded there was not widespread use of analytics tools in higher education because administrators were not able or willing to invest the needed time or money. Analytics could assist these leaders in gaining actionable insight (Baer & Duin, 2014; National Commission on Higher Education Attainment, 2013) into grant distribution impacts on postsecondary degree attainment (Pomeroy, 2014).

Purpose Statement

The purpose of this DSR study was to develop a list of technological enhancements to a financial aid analytics application (F3A). As an initial step in the DSR method, the F3A software application was developed. The application was designed to assist higher education leaders in gaining actionable insights for decision-making with grants management by analyzing the impact of financial aid awards on student achievement. To address the problem, a focus group was conducted to evaluate the application with emphasis placed on identifying enhancements required to ensure that the F3A was effective in financial aid analysis. The group was comprised of a panel of Kentucky state government employees and post-secondary education leaders who support financial aid decisions at their respective institution. Participants assembled at the

Kentucky Higher Education Assistance Authority (KHEAA) facility, the moderator demonstrated the functionality of the tool, and the participants responded to a series of questions and provided an assessment on the usability and effectiveness of the F3A.

Research Question

The central research question proposed for the study is, "What technological enhancements should be made to the financial aid analytics application (F3A) to assist higher education leaders in decision-making regarding financial aid resource allocation?"

Propositions

A proposition is a claim made by the researcher or a general theory to test in a research study (Creswell, 2014). This study proposed that the analytical software application, and its outputs, help financial aid policymakers gain more insight into the effects of their current grant programs and could assist them in enhancing their ongoing subsidy management process.

Conceptual Framework

Analytics can be utilized for behavioral analysis (Chintagunta, Hanssens, & Hauser, 2016; Gessner & Scott, 2009), cost optimization (Soni, Ansari, Sharma, & Soni, 2011; Teffeteller & Kish, 2012), and risk management (Chan, Fan, Prodromidis, & Stolfo, 1999). However, many higher education administrators have experienced difficulties with acquiring and adopting essential technology and techniques (Ferreira & Andrade, 2016; Roscorla, 2015). The goal of this research study was to evaluate the F3A and document technological improvements, which could assist in the development of future applications targeting higher education leaders. Higher education administrators in the state of Kentucky, specifically those who support the development of financial aid programs, participated in a focus group interview session. The group focused on identifying the benefits and pitfalls of the F3A for assessing the impacts of scholarships on postsecondary degree completion, as well as, exploring possible avenues for utilizing analytics in grants management decisions.

Assumptions

The first assumption was that administrative authorities, who support the development of financial aid programs, can provide feedback about a grants management software application (Ye, Zhan, Li, Huang, & Jiang, 2016). The second assumption is that financial aid managers want to enhance their current methods of evaluating their programs with analytics. Recent financial aid studies have emphasized the need to include technology in selecting recipients and assessing grant programs (Purba & Sembiring, 2016; Ye et al., 2016). Additionally, in 2013, the National Commission on Higher Education Attainment (NCHEA) issued a letter to college and university leaders requesting they make it a priority to harness their resources in an effort to increase postsecondary enrollment and assist students in obtaining advanced degrees (Baer & Norris, 2016).

Significance of the Study

This study adds value to decision-making in financial aid allocation through analytics and contributes knowledge to assist information technologists and data analysts to better serve the needs of their stakeholders (i.e. students, higher education administrators). Many recent subsidy distribution studies cover institution-based aid using a quantitative or qualitative approach (Latumakulita, Purnama, Usagawa, Paturusi, & Prima, 2016; Ma, Yue, Zhang, Cui, & Qu, 2009; Purba & Sembiring, 2016; Wei, Han, Kong, & Xia, 2016; Ye et al., 2016), whereas this study focused on an analytical tool designed for state aid programs and used the DSR methodology. The Kentucky Higher Education Assistance Authority (KHEAA) research department plans to use the application evaluated in this study to assist with financial aid budgeting. Because it is configurable, other colleges, universities, and state education agencies may utilize this tool to improve or understand the effects of their current aid programs. Additionally, this study expects that providing solutions for grants optimization may assist in measuring the effectiveness and improving the allocation of grant resources making college less costly and more attainable for prospective college students.

Delimitations

The purpose of the DSR methodology is to evaluate an IT artifact, which contributes to both the business and IT sectors (Hevner & Chatterjee, 2010). This study aimed to receive feedback about the F3A for future enhancements but refrains from analyzing and evaluating the results produced by the tool. Therefore, information regarding the results was not summarized, and feedback pertaining to those outcomes was limited to the opinions of the study participants. Also, the study was limited to input from a targeted audience, specifically Kentucky state employees who support financial aid policy development. The KHEAA is the provider of the states' three main postsecondary financial aid programs, which support Kentucky residential students in continuing education (Kentucky Higher Education Assistance Authority, 2015). The KHEAA (2015) disbursed more than $208 million dollars to a combined total (i.e., may include duplicates) of 123,190 recipients.

Limitations

The qualitative approach is one method of evaluation used in the DSR methodology, (Hevner & Chatterjee, 2010) to analyze the focus group data. This design typically uses small sample sizes; thus, results were limited to the group but may be theoretically generalizable (Bryman & Bell, 2011). The data collection procedure consisted of conducting a focus group for the evaluation of the F3A. The feedback was limited to the authentic responses of the focus group participants. One final limitation was that the IT artifact was developed in a 35-day timeline using RStudio, a free open source software tool (O'Leary, 2017) and populated with synthetic data. Therefore, the technology and scope of the project may restrict the features of the F3A.

Definition of Terms

The following is a list of terms commonly used throughout this dissertation study:

- Business Analytics: The incorporation of both DS and BI tools and techniques to capture and process data to assist non-technical users with decision-making (Harper, 2014).
- Business Intelligence: The use of business-related information (i.e., reports, dashboards) to understand historical data (Mohanty, 2008).
- Big Data: The challenges of collecting, storing, and processing large amounts of data which demonstrates issues with one or more the following: volume, velocity, or variety (Chintagunta et al., 2016).
- Data Mining: The application of sophisticated tools, statistical algorithms, and data modeling to find hidden characteristics in structured and unstructured big data (BD) problems (Coenen, 2004).
- Data Science: A field that combines computer science, statistical techniques, and technology for solving big data problems (Cleveland, 2001).
- Financial Aid: Monetary assistance provided to students to help cover college expenses (i.e., tuition, books, room, and board) (University of Hawaii Community Colleges, 2017).
- Graduation Rate: The total number of students in a cohort who graduate divided by the total number of students in the group (National Center for Education Statistics, 2016).

General Overview of the Research Design

Research methodologies help guide the study by providing the procedure for data collection, evaluation, and interpretation (Creswell, 2014). Popular research methods include mixed methods, qualitative, and quantitative (Creswell), but study results yield more descriptive information rather than practical applications or solutions when one of the aforementioned methods is employed. (Peffers et al., 2006). This research followed the DSR methodology, which is a framework used to evaluate an IT artifact created to solve a real-world problem (Hevner & Chatterjee, 2010; Vaishnavi & Kuechler, 2008). DSR is significant to Information Systems (IS) research because it addresses the value of a new IT product as well as its application in the field in which it will be utilized (Hevner & Chatterjee). This methodology was selected because it serves as a guide for creating an IT artifact (the software tool), for assessing the design of the artifact, and for determining the value of the artifact from both an Information Systems and business use standpoint (Hevner & Chatterjee, 2010). The first step was to develop the IT artifact, which for this study was the F3A. Once developed, the F3A was evaluated by a focus group to gather suitable data from a target population that could potentially benefit from using the tool. (Bryman & Bell, 2011; Kreuger & Casey, 2008). The focus group was comprised of administrators who support the state of Kentucky's financial aid programs. The participants viewed a demo of the F3A and responded to questions about the technology with the objective of obtaining a list of technological enhancements that were applicable for use in meeting their business objectives.

Summary of Chapter One

Chapter 1 began with a brief overview of the increased use of and organizational dependence on business analytics across different industries (Business-Higher Education Forum, 2014; Evans, 2015; Gartner, Inc., 2012; Hsinchun et al., 2012; Lublinsky et al., 2013; National Commission on Higher Education Attainment, 2013). This section also detailed the adoption of technology challenges in higher education, particularly those regarding financial aid delegations (Goldstein, 2005; Pomeroy, 2014; Thanh & Haddawy, 2007). The problem statement referenced the literature gap of an analytical solution for state financial aid resources (Purba & Sembiring, 2016; Ye et al., 2016). This chapter also included a description of the DSR methodology that was used to guide the creation and evaluation of the F3A (Hevner & Chatterjee, 2010; Vaishnavi & Kuechler, 2008). Next, the conceptual framework which detailed the design of the study and the constraints of the research were defined. Chapter 2 summarizes past studies, which examined the need for further research of analytics in higher education as well as previous techniques used for analyzing financial aid programs. Additionally, the design of the F3A artifact applied the findings from the literature review, in Chapter 2, as a foundation for development.

Organization of Dissertation

This study utilized a process model that was developed to distinguish Design Science Research from simple practice. This framework includes the following six components (a) Problem Identification, (b) Solution Objectives, (c) Design, (d) Demonstration, (e) Evaluation, and (f) Communication (Peffers et al., 2006). In Chapter 1, the problem was identified, the concepts of analytics introduced, and an overview of the study provided. Chapter 2 contains a literature review of past and present analytical tools and techniques as well as technology used and analysis performed in postsecondary grants management. Chapter 3 details the solution's objectives and the design of the artifact. Chapter 4 explains the research design used to evaluate the F3A. Chapter 5 presents the data collected from a focus group interview in which the participants evaluated the F3A. Lastly, Chapter 6 covers significant findings determined from this research study as well as recommendations for further research.

CHAPTER TWO

The problem addressed in this study is the lack of analytics used for strategic decision-making regarding financial aid resource allocation (Goldstein, 2005; Pomeroy, 2014; Thanh & Haddawy, 2007). The higher education industry has struggled with the adoption of technology (Roscorla, 2015) leaving the literature fragmented on the topic of financial aid management systems for statewide assistance programs (Goldstein, 2005; Pomeroy, 2014; Thanh & Haddawy, 2007; Witten, Eibe, & Hall, 2011). The purpose of creating the F3A was to contribute to technology in higher education while assisting financial aid policymakers in strategic planning. This study contributes to the literature by providing insight into the design, implementation, and evaluation of the F3A.

This chapter examines the supporting literature by summarizing and comparing views from past research. The opening topics covered in this chapter include an overview of strengths and weaknesses when utilizing Business Intelligence, Data Science, and Data Mining techniques for analytical solutions. These topics are important in understanding whether the feedback from focus group participants was consistent with previous research and to influence the development of the F3A. Next, the Growth and Adoption of Analytics and Challenges of Analytics in Higher Education were detailed to identify a gap in the literature. Lastly, the subjects of Financial Aid Program Structures and Frameworks for Financial Aid Management Analytics were summarized to present findings from past studies similar to this one and to provide guidance for the development of the F3A. This chapter also includes the problem identification phase of the DSR methodology and briefly covers a system overview of the F3A. This chapter concludes with the conceptual framework and a description of the research design.

Business Intelligence

Mohanty (2008) defined BI as the use of business-related information to achieve company objectives. The range of business applications and their strategic implications are vast (Mohanty, 2008). The term *analytics* is referenced throughout this study and defined as the use of information technology to capture and process data for decision-making. Analytics encompasses the following actions:

- Developing infrastructures which support timely feedback (Ferreira & Andrade, 2016);
- Monitoring systems after development (Ferreira & Andrade, 2016); and
- Utilizing the output results for continuous improvement on the current information extraction process (He, 2014).

Although the origins of BI trace back to the 1500s, Dedijer sparked the *intelligence revolution* by utilizing war tactics, previously used for espionage during WWII, to compete with other businesses (Marren, 2004). These techniques led to a concept called *competitive intelligence*, or the *intelligent business,* in which companies spied on other agencies to gain intel for their sales strategies and in turn increase profits (Marren, 2004). Over time, these tactics transformed into BI, in which agencies also monitor their organization's and customer behaviors to support organizational objectives, enhance strategic development, increase customer satisfaction, decrease criminal activity, and improve employee retention (Marren, 2004).

Kavur (2009) wrote that BI solutions assisted organizations in making faster, more informed decisions causing their implementation to span across various industries. Over time, many organizations became dependent on BI solutions, leaving some to question the continued survival of companies that do not take advantage of BI. In 2010, survey results published in Bloomberg Newsweek revealed that companies generating over $100 million in revenue used some form of BI (Hsinchun et al., 2012). In 2015, The American Management Association further supported this claim by conducting a pre- and post- intervention survey of 800 business executives (Evans, 2015). The survey questioned organizational leaders on the significance of BI tools on their company's operations (Evans, 2015). The results of the pre-survey stated 58 % of managers deemed BI tools crucial, while the post-survey - conducted on the same executives five years later - yielded an increase of 82% in positive responses (Evans, 2015).

BI applications use statistics, predictive modeling, and data mining methods to traverse through and analyze data (Pomeroy, 2014). These techniques detect patterns or trends that may warrant more in-depth investigation into scenarios, such as identifying anomalies in bank transactions for fraud detection or categorizing at-risk student characteristics to place them in suitable, academic courses (Venter, 2009). The successful use of BI & Analytics (BI & A) can provide companies with a competitive advantage by assisting with the identification of when, who, and how to obtain and retain customers (Minkara, 2012), learn more about competitors, and become more intelligible about business strategies. However, there are many challenges, which can occur when attempting to gain intelligence from data (Venter, 2009).

There are two main categories of BI, strategic and operational (Bataweel, 2015). Businesses have utilized strategic BI for advanced concepts, in-depth research, and solving issues, which could affect an organization in the long term (Bataweel, 2015). Strategic BI involves the use of designated professionals such as a data analyst who understands how to manipulate data, develop statistical algorithms, and interpret their output (Bataweel, 2015). Conversely, operational BI is used for ad-hoc requests and is essential for day-to-day operations and reporting (Bataweel, 2015). Leadership and staff use popular applications such as dashboards and reports for quick decision-making. In general, BI has been known to have several beneficial characteristics including:

- Fast data processing;
- Reliable analysis produced by mathematical functions;
- Management of diverse data to assist varying departments in an organization; and
- Continuous data collection, which allows real-time analysis (Bataweel, 2015).

Companies typically have a few essential elements in place to utilize data successfully such as a data collection system, a quality control process, a retrieval method, and tools to analyze and visualize the information (Bataweel, 2015). Key factors which have influenced the effectiveness of BI include:

- Ease of using BI tools;
- Data quality;
- Information accessibility;
- Agility in decision making; and
- External variables such as marketing, product pricing, and organizational threats (Mohanty, 2008).

Due to the growing popularity of BI solutions, there has been an influx in the purchase of BI applications since 2009 (Chuah & Wong, 2011). However, 60 to 70% of companies fail to utilize BI successfully due to organizational challenges including lack of a supportive culture, finance, and infrastructure (Chuah & Wong, 2011). Another reason BI solutions fail is the challenging of obtaining the right data, presenting it to the right people, and showing it at the right time (DeVoe & Neal, 2005). Putting these elements in place can be a difficult task, especially for companies with a critical need for making complex decisions with little resources in a short timeframe (DeVoe & Neal, 2005). Though organizations have many challenges when implementing BI solutions, there are options to combat them. Solutions that have been identified to promote the efficient use of BI tools include:

- Developing data quality control processes;
- Training the technically challenged and hiring experienced professionals;
- Involving relevant people in the BI solution selection process; and
- Customizing data and applications down to the unit, departments, or employee level (DeVoe & Neal, 2005).

Additionally, BI has grown over time to become more scalable, affordable, and accessible thus minimizing the pitfalls in obtaining and using these tools (DeVoe & Neal, 2005).

Data Science

Though analytical BI solutions may include advanced components such as dashboards, summary statistics, and reports, these illustrations only cover a subset of available methods (EMC Education Services, 2015). DS is another technical subject, which encompasses a combination of sophisticated statistical and computing algorithms used for mining knowledge from data (EMC Education Services, 2015). The DS field was founded in 2001 to gain more knowledge from data through the improvement of data retrieval and analytics (Cleveland, 2001). Cleveland, a statistical researcher for Bell Labs, proposed DS as a solution to the constraints of advanced mathematical modeling. These constraints stemmed from the challenges of obtaining and utilizing practical resources for data exploration as well as other external factors that influence data quality, variety, and availability (Cleveland, 2001). Cleveland's objective for the field of DS was to provide an optimal computing environment in which models and methods were compatible with useful data analysis. Cleveland (2001) argued that computer scientists face challenges when approaching data related issues from an analytical perspective. Conversely, statisticians had limited computing knowledge and faced similar obstacles (i.e., data models and infrastructure). Thus, DS provided a partnership between the two subject matter areas (Cleveland, 2001). Therefore, the central principle of DS was to enhance the data analysis process by combining critical aspects of statistics and computer science (Cleveland, 2001). More specifically, the field aims to the combine and optimize various aspects of statistical modeling with hardware and software performance for better storage, transporting, accessing, manipulating, and modeling trends in data (Cleveland, 2001).

With the growing volume, velocity, and variety of data, data scientists have shown increasing interest in big data (McAfee, 2012). The data gathered, stored, and managed since 2012, was estimated to be 2.5 Exabytes of information each day (McAfee, 2012). A concept called big data (BD) defines the challenges of collecting, storing, and processing large amounts of data (Chintagunta et al., 2016). BD issues occur when institutions have one or more of the following problems with their data:

- Volume: The amount of data begins to outgrow the traditional infrastructure which provides challenges for storage and analysis;
- Velocity: The data processing speed increases and cannot support problems in need of quick solutions; and
- Variety: The data originates from a variety of sources and may be structured, unstructured, or semi-structured (Sagiroglu & Sinanc, 2013).

To harness the power of data, data scientists utilize a variety of big data analytics (BDA) technology and tools (EMC Education Services, 2015). DS frameworks such as Hadoop and Spark store, transport, and process structured and unstructured data types (McAfee, 2012). Applications such as R, SAS, Python, and Tableau manipulate and analyze data. DS research methods and instruments allow researchers to utilize text-mining, machine learning, forecasting, and network analysis (Ohri, 2016).

Data Mining

Data Mining (DM) is the application of tools and algorithms to find hidden characteristics in data (Coenen, 2004). DM is categorized under the subject area of computer science and firmly established as a method within the realm of machine learning (ML) and statistics (Coenen, 2004). In the 1990s, DM, along with data preparation and visualization, was known to be a sub process of Knowledge Discovery in Databases ([KDD]; Coenen, 2004). Fayyad (1996) - KDD and DM pioneer - defined KDD as a nontrivial framework for understanding patterns in data. DM was initially used in studying relational data, but with the growing number of unique data forms, such as images, text, networks, and documents, a significant focus in DM has shifted to investigating unstructured, semi-, and quasi- structured data forms (Coenen, 2004).

Though DM is under the umbrella of ML, there are some features ML maintain which DM does not (Coenen, 2004). DM focuses solely on the structure and analysis of various data types, while ML extends past studying characteristics of data and encompasses discovering the best methods for replicating, automating, and learning computational processes (Coenen, 2004). ML benefits include improved performance and higher accuracy, in comparison to stand-alone statistical models which can be automatically retrained (Information Resources Management Association, 2012).

There are many DM theories and techniques, but the two most common categories are supervised and unsupervised learning (EMC Education Services, 2015). In supervised learning, an analyst is provided with a set of predictor data points, each with a given outcome, and then asked to find a function which can predict these points (EMC Education Services, 2015). One of the significant issues with this type of learning is prediction error because the predictive models are trained on a set of values (Coenen, 2004). When models are tested using values outside of the training dataset, the prediction error increases (EMC Education Services, 2015). Two examples of supervised learning techniques are classification and regression (Coenen, 2004). Unsupervised learning occurs when the dataset in question contains no predictors or corresponding outcome variables, but the researcher seeks to identify patterns or trends (EMC Education Services, 2015). There are many methods utilized for unsupervised learning including clustering, density estimation, and dimensionality reduction (EMC Education Services, 2015). Numerous industries have utilized DM for a broad range of benefits such as,

- Studying consumer behavior to provide customers with a personalized shopping experience (Chintagunta et al., 2016);
- Identifying factors which influence specific medical diagnosis for illness prevention (Soni et al., 2011); and
- Detecting credit card fraud in banking (Chan et al., 1999).

Data Science versus Business Intelligence

Data analytics methodologies and usage may fall into two categories, BI or DS (EMC Education Services, 2015). BI typically defines analytics for creating reports, queries, or dashboards which assist stakeholders with questions about the past or current status of a behavior of a group or an event (EMC Education Services, 2015). BI tends to answer questions with the connotation of *when* and *where* (EMC Education Services, 2015). Also, the process of producing information using BI methods typically involves performing aggregates on historical data in a structured format such as spreadsheets, relational databases, or CSV (EMC Education Services, 2015).

Contrarily, DS is usually employed when there is a need for insight into the future. DS methods utilize more complex data mining algorithms to analyze disaggregated data and typically deal with various structured and unstructured data types (EMC Education Services, 2015). DS involves exploratory analysis which tends to answer questions of a broader tone such as *why* and *how* (EMC Education Services, 2015). DS models may be used to predict likely events or identify anomalies in data (EMC Education Services, 2015).

DM differs from summarized statistics - typically utilized in BI - because the process not only involves utilizing historical information but constructing logical patterns to form an assumption about future events (Han, Kamber, & Pei, 2012). Summarized statistics are restricted to the historical value of the dataset, whereas DM seeks to predict the future, which lies beyond the scope of the data (Han et al., 2012). DM provides a means for finding hidden patterns in large datasets (Han et al., 2012), which could not be detected using standard reporting techniques, as well as assisting in effective planning and decision making for the future (Nisbet, Elder, & Miner, 2009). Various DM methods and models exist including logistic regression, association, clustering, and decision trees (EMC Education Services, 2015). Selecting an adequate application depends on the nature of the data and the research problem (Larose, 2015).

Growth and Adoption of Analytics

Various types of analytics have grown in popularity across different industries with claims of improved strategic planning, efficiency, and cost-savings (Banerjee, Bandyopadhyay, & Acharya, 2014; Hackathorn & Margolis, 2016; Picciano, 2012; Wills, 2014). Though the reasons for the growth in analytics differ, many enterprises struggle with adopting these techniques and tools for decision support (Banerjee et al., 2014; Hackathorn & Margolis, 2016; Picciano, 2012; Wills, 2014).

Wills (2014) discussed the importance of Small Data Analytics (SDA) for better understanding small populations such as those with rare health issues. Picciano (2012) stated BDA was not the solution for all data analysis problems but covers their advantages for improving college student retention. Banerjee et al. (2014) described the evolution of business analytics - an all-encompassing term for BI, BDA, and DS - using examples from the banking, healthcare, manufacturing, and retail industries. Hackathorn and Margolis (2016) introduced immersive analytics, which blends analytical reasoning with virtual reality, for problem solving and simulation in both the corporate world and academia.

Wills (2014) claimed the increased interest in data-driven decision-making inspired the growth in data collection and analytics. Conversely, Picciano (2012) explained that the steady influx of the data collected by organizations sparked popularity in data analytics for decision support. Banerjee et al. (2014) and Hackathorn and Margolis (2016) claimed the growth in analytics stemmed from advances in technology as well as the increased accessibility to new data sources.

Lack of training, or expertise, for proper implementation and utilization of the analytical technology, was a recurring theme in the adoption of analytics (Banerjee et al., 2014; Hackathorn & Margolis, 2016; Picciano, 2012; Wills, 2014). Picciano (2012) and Banerjee et al. (2014) addressed the challenges of establishing a supportive analytics culture including data privacy, policy, and information management issues. Additionally, Hackathorn and Margolis (2016), Picciano (2012), and Wills (2014) discussed the complexity of developing functional hardware and software systems for analytics.

Challenges of Analytics in Higher Education

Academic analytics is a term coined in research performed by Goldstein (2005), which encompasses the utilization of BI methods and tools to solve problems in academia. Academic analytics has allowed higher education institutions to utilize their data for assessing and analyzing program, departmental, and individual performance (Goldstein, 2005). The process entails using analytics to influence actions such as policy change, learning interventions, technological improvements, and other areas of decision-making (Ferreira & Andrade, 2016). Academic analytics have typically been used by central finance, institutional research, and admissions offices within higher education institutions (Goldstein, 2005), but compared to other industries postsecondary education agencies are still in the beginning stages of using BI to solve problems and make decisions (Ferreira & Andrade, 2016; Macfadyen et al., 2014; Roscorla, 2015).

Colleges and universities have often been required to report their student achievement measures to assessment, accreditation, federal, and state agencies (Saupe, 1990) such as the Southern Association of Colleges and Schools, the National Center for Education Statistics, and the Kentucky Council of Postsecondary Education. Though reporting agencies may vary depending on the institution, most schools report student, faculty, staff, and institutional data for gaining institutional funding or ensuring these institutions are fulfilling their goal of providing quality education and environment for their students (Saupe, 1990). Consequently, universities have significant amounts of data at hand (i.e., related to student achievement, business operations, and technology), which could potentially be an excellent source for program and system enhancements, but this information is rarely mined (Pomeroy, 2014). Many colleges and universities have yet to adopt these tools for several reasons including policies, culture, inability to obtain tools, and lack of training (Pomeroy, 2014). Ignoring this abundance of information has caused higher education institutions and supporting organizations to fall behind in meeting their academic goals (Macfadyen et al., 2014). In particular, financial aid management tools and techniques have received little attention in former academic and practitioner research, likely because BI tools were rarely utilized for grant delegation purposes (Goldstein, 2005; Pomeroy, 2014; Thanh & Haddawy, 2007).

In previous years, many schools have experienced reduced funding leading faculty and administration to seek other sources of revenue and focus more attention on how limited funds are delegated (Narozhnaya, 2015; Ragland, 2016; Thanh & Haddawy, 2007). A growing number of higher education institutions have utilized data mining solutions to prove their need for more funding and to find self-sustaining solutions (Pomeroy, 2014). Past survey research indicated, less than ten percent of respondents in higher education used analytics for grants management regarding pre- and post-award performance as well as triggering award availability (Goldstein, 2005). These studies suggest that a need exists for more analytics regarding the distribution of financial aid monies and the effectiveness of aid programs in fulfilling their intended purpose.

Financial Aid Program Structures

In the United States a few entities, including the (a) federal government, (b) state government, (c) institution of attendance, (d) public organization, or (d) a private business, historically sponsored college monetary assistance programs (Delaney, 2011; Hossler, 2002). Many financial aid programs funded by the state and federal government were developed to promote postsecondary schooling amongst low-income families, (Delaney, 2011; Hossler, 2002), to supply job market demands, and to assist states in retaining residents (Groen, 2011). Over the years, there have been many different sources for obtaining financial aid and various types of monetary assistance, which fall under two primary categories: need-based and merit-based aid (Peterson's CollegeQuest, 2014). Individual students or student's whose family fell into the low-income category have eligibility to receive need-based aid (Peterson's CollegeQuest, 2014). Students with proven high academic performance (i.e. high-grade point average or test scores) typically apply for merit-based aid (Peterson's CollegeQuest, 2014). Some subcategories of need-based and merit-based aid include the following (Peterson's CollegeQuest, 2014):

- Loans: Monetary aid in which the recipient must pay back an agreed amount to the provider. The amount borrowed usually collects interest over time until the provider has been fully reimbursed (Citizens Financial Group Incorporated, 2016).
- Grants: A need-based award in which the awarded student's personal or family income falls under a specified a threshold, typically defined by the provider (i.e., low-income). The award does not require repayment to the sponsor (The fundamentals of financial aid award letters, 2001).
- Work-Study: Need-based monetary assistance, much like a grant, given to students who agree to work for the aid provider, usually the institution of attendance, over a specified period in return for tuition assistance (Center for Analysis of Postsecondary Education and Employment, 2015).
- Scholarships: Merit- or need-based aid in which the student must meet and maintain a set of eligibility requirements over the period of award disbursement (Toby, 2010).

Over a 25-year period, loans went from making up 30% of financial aid to comprise 60% of award types, resulting in more debt for college goers and their families (Creech, 1998). Some government agencies have tried to minimize student debt by replacing loans with grants and implementing loan repayment plans (Goodman, 2008). The option of donations was typically more attractive to students, yielding increased enrollment and student retention (Goodman, 2008).

The high number of award sources, programs, and objectives can result in many different combinations of outcomes. Therefore, it is essential to consider many factors when creating financial aid programs including student demographics, the cost of attendance, scholarship publication, and simplicity which represents a few among many other variables. For example, programs not widely publicized, or with complicated application processes, were shown to be less effective in increasing student retention and degree completion rates (Long, 2009). Other historical research trends reveal that need-based aid yields a larger impact on increasing student accessibility to college and student achievement measures (Long, 2009). Conversely, merit - based aid has shown to have a smaller impact than need-based aid on student success (Long, 2009). Past research has also shown most students who receive merit-based aid were usually academically high performing in high school and were from families who could afford to pay for college without the use of outside monetary assistance (Long, 2009). Therefore, students who typically receive merit-based awards would have chosen to attend college regardless of monetary aid (Long, 2009). Another study found financial aid was more effective when offered to stop-out students (DesJardin & McCall, 2010). Many students who decided to leave school came back when funding was available which increased re-enrollment as well as degree completion numbers (DesJardin & McCall, 2010). Researchers have studied the success of popular financial aid programs such as the Pell Grant, Georgia HOPE scholarship, and the federal loan service (Sjoquist & Winters, 2012) in meeting their objectives, but there remain many programs whose value have yet to be analyzed with the use of data mining tools and techniques (Goldstein, 2005; Pomeroy, 2014; Thanh & Haddawy, 2007).

Naturally, financial institutions have historically utilized high-level statistical models and applications for finance optimization (Cevizci, 2016; Gilli & Schumann, 2012; Leibfritz & Maruhn, 2009). Therefore, much of this research surrounds market trends, portfolio and risk management, and other monetary related topics. Although grants management falls into a finance-related category, there has not been much scholarly research performed on grants management (Goldstein, 2005; Pomeroy, 2014; Thanh & Haddawy, 2007). The following two paragraphs cover two scenarios for optimizing financial aid.

Georgia State University (GSU) utilized data mining techniques for financial aid delegation (Selingo, 2013). A significant number of students were dropping out of college because of financial deficits owed to the school (Selingo, 2013). Data mining techniques were used to identify students who were in default of amounts as low as $500 and were forced to drop out of classes, some shortly before graduating (Selingo, 2013). GSU administrators found their students' average default amount revealed a correlation between unmet need amount and student performance (Selingo, 2013). As the student’s default amount to the institution rose, student performance decreased as did the student's likelihood of attrition (Selingo, 2013). GSU leadership acted on these results by performing an experiment in which they gave small grants to 200 dropout students who left the institution due to nonpayment (Selingo, 2013). Those students re-enrolled and continued their coursework, which produced more than $660,000 in tuition and fees revenue (Selingo, 2013). Due to the outcome of this experiment, GSU developed the Panther Retention Grant, a financial aid program that strategically identifies students with unmet need and provides them with grant money (Selingo, 2013). This study shows the potential outcomes of utilizing data mining to detect anomalies, however, the report did not contain a detailed methodology or framework.

Another research study, conducted at a college located in Chile, examined the extent to which financial aid amounts and types influenced student dropout risks (Horn, Santelices, & Avendaño, 2014). Researchers elected to use a survival analysis model, which incorporated the institution's historical data (Horn et al., 2014). Survival analysis is a statistical technique used to predict the amount of time it will take before an event occurs (Despa, n.d.). The model applied in this study evaluated the effects of different variables, such as funding sources, student demographics, and their academic performance on student retention (Horn et al., 2014). Results revealed that need-based financial aid programs, regardless of whether they were grants or loans, were most prominent in student decisions to attend and remain enrolled in school (Horn et al., 2014). The researcher further recommended similar studies that assess the effects of financial aid programs on outcomes other than retention, such as graduation or workforce entry (Horn et al., 2014). Many financial aid studies provide feedback about the impact of financial aid programs on student behaviors, but these studies rarely make suggestions for a tool or address a BD challenge. Although this study does not address monetary optimization, the results provide useful information for the development of a practical solution.

Frameworks for Financial Aid Optimization

Financial aid could potentially have a positive impact on student decisions and performance, but the probability of these outcomes varies by award recipient and other relative factors (Sjoquist & Winters, 2012). Due to the unique types and number of possible student success scenarios, an accurate depiction of how monetary awards impact student achievement is crucial for financial aid optimization (Thanh & Haddawy, 2007). Thanh and Haddawy (2007) utilized a combination of data mining and transforming algorithms to build a framework. The purpose of the project was to find a solution, which maximized tuition revenue by detecting the minimal financial aid package needed for a student to enroll in a university (Thanh & Haddawy, 2007).

The researchers utilized Bayesian Networks (BN) to determine the probability of an applicant's enrollment given a certain amount of financial aid (Thanh & Haddawy, 2007). Bayesian Networks are a graphical approach to model and analyze conditional probabilities between variables (Darwiche, 2009). The longitudinal data used in the study included 4 years of graduate school admissions data, which contained a total of 7,788 applicants. Nine predictors were input into the network model including “age, marital status, nationality, the institution of the previous degree, GPA of the previous degree, school, degree program (master or doctor) to which the applicant is applying, financial aid” (Thanh & Haddawy, 2007, p. 8). The output factors were binary indicating *0* for *Not Enrolled* and *1* for *Enrolled* (Thanh & Haddawy, 2007).

The model was validated using the K-fold Cross Validation method (KCV) (Thanh & Haddawy, 2007) in which the data is split into *1,…,k* subsets, the model is trained on the other k-1 subsets and tested on the kth subset to minimize prediction error (Hastie & Tibshirani, 2009). The dataset was hugely imbalanced because 82% of applicants enrolled while only 18% of students did not (Thanh & Haddawy, 2007). The data was split into 60:40 partitions and two models were created (Thanh & Haddawy, 2007). One model was trained and tested on data which was biased towards positive responses (i.e., 60% *Enrolled* and 40% *Not Enrolled*) while the other model was biased towards negative responses (i.e., 60% *Not Enrolled* and 40% *Enrolled*; Thanh & Haddawy, 2007). Both models resulted in over 90% predictive accuracy rates, in the categories they were biased towards, but yielded less than 50% in the unbiased categories (Thanh & Haddawy, 2007). As a solution, Ensemble Model (EM) methods were utilized to combine the strengths of both models (Thanh & Haddawy, 2007).

The final EM yielded a 65.8% true positive rate and an 86.5 % true negative rate, which balances out the results better than the individual models. The Expected Tuition Revenue (ETR) is a function, which multiplies the probability a student would enroll by the amount of income obtained at a specific aid level (Thanh & Haddawy, 2007). Once the probabilities were calculated, different optimization scenarios were viewed by adjusting the students and viewing the outcomes of the ETF function (Thanh & Haddawy, 2007).

Thanh and Haddawy's (2007) study sampled graduate students from one institution yielding a small sample size, which leaves an opportunity to address the use of a Bayesian Networks in modeling a BD problem. In addition, six out of the seven predictors chosen to fit this model were discrete, aside from financial aid, which could mean the addition of other variable types may yield different model performance. Although there were some limitations with the dataset used in this study, the framework satisfied the business objectives, which could yield valuable information for the IT artifact developed in this study.

Boshardt, Lichtenstein, Palumbo, and Zaporowski (2010) developed a series of theoretical models for optimizing financial aid spending for the goal of institutional profit maximization. The student population was split into two groups, residents, and commuters, both analyzed separately, and the results were later combined (Boshardt et al., 2010). The first equation accounted for the cost of delivering education to the student body and each additional student after that (Boshardt et al., 2010). The second function calculated the probability of a prospective student accepting a specific tuition amount and factored in the influence of an effective tuition amount (i.e., scholarships minus tuition; Boshardt et al., 2010). As expected, when the sufficient tuition amount decreased, the probability of the student accepting the tuition amount increased (Boshardt et al., 2010). A third model was created to capture the students' likelihood of matriculation given demographic and circumstantial variables (Boshardt et al., 2010). The expected profit for a student was based on the combination of three equations modeled as; a function of the probability of matriculation, the likelihood of accepting a tuition package, and the cost of education (Boshardt et al., 2010). The student data was extracted from one institution and contained an estimated 5,000 observations (Boshardt et al., 2010). The results of this study displayed possible student segment scenarios and outcomes, but no accuracy measures were provided.

Problem Identification

The DSR methodology involves the exploration of an IT artifact, which solves a real-world problem. *Problem Identification* is the first step in the DSR process of creating an artifact for evaluation (Peffers et al., 2006). The problem was derived from the principal business question which inquires, *how can financial aid be distributed in a way which maximizes student success and minimizes costs?* This question spawned from a business request from the Kentucky Higher Education Assistance Authority (KHEAA), a state government agency within the Finance and Administration Cabinet (Kentucky Higher Education Assistance Authority, 2017) and was supported by the literature. The problem was split into four categories:

- Limited Financial Aid Funds;
- Pressure to Raise Student Success Metrics;
- BD Challenges for Large Aid Programs; and
- High Costs of Software/Hardware.

Therefore, the solution's objective was to *provide an inexpensive tool which demonstrates to the end user how to maximize student success with the minimal amount of aid needed and also has the capacity to either minimize the amount of data processing expended or possess the ability to process large amounts of data*.

KHEAA Business Request

The Kentucky Higher Education Assistance Authority (KHEAA) is a state government agency within the Finance and Administration Cabinet (Kentucky Higher Education Assistance Authority, 2015). The KHEAA is tasked with the responsibilities of carrying out the administration of state-funded postsecondary student financial aid programs and the disbursement of program funds according to statutes and regulations (Kentucky Higher Education Assistance Authority, 2015). The principal legislative intent of state-funded postsecondary student financial aid is to help individuals and families finance their postsecondary education by providing financial assistance for designated higher education opportunities (Kentucky Higher Education Assistance Authority, 2015). In an effort to prevent the potential liquidation of the Kentucky Educational Excellence Scholarship (KEES), the Kentucky Tuition Grant (KTG), and College Access Program (CAP) grant, the director of research at the KHEAA requested a grants management solution.

The three-leading postsecondary student financial aid programs created by the Kentucky General Assembly and administered by KHEAA for the benefit of Kentucky residential students are; the CAP, the KTG, and the KEES (Spalding, 2014). These three programs annually receive dedicated funding from the net income of Kentucky lottery tickets sales, less $3 million per year which is dedicated to state literacy programs (Ky. Act of 2005). Approximately 45% of lottery-funded financial aid money was allocated to KEES while 55% of funds are split between the CAP and KTG (Spalding, 2014). The average scholarship amount awarded to students for fiscal year (FY) 2015 was approximate $1,545 per student, which resulted in an annual total contribution of $107,716,000 (Kentucky Higher Education Assistance Authority, 2015). The Legislative Research Commission plans to start allowing high school students enrolled in dual credit classes to utilize the KEES award towards their coursework (Kentucky Higher Education Assistance Authority, 2015). The extension of KEES eligibility is set to begin in FY 2017-2018, which will result in a higher demand for this scholarship (Ky. Act of 2015). The demand is expected to require over $17,000,000 annually, which could result in the reduction, and possible liquidation of the KTG and CAP grant (Ky. Act of 2015).

The KHEAA collects and stores Kentucky high school and college student data such as Free Application for Federal Student Aid (FAFSA) information, student high school GPA each year in school, student demographic characteristics, standardized test scores, institutional information, and more (Kentucky Higher Education Assistance Authority, 2015; Spalding, 2014). The state of Kentucky had an estimated average of 192,265 public high school enrollees each fall between the years of 1990 and 2015 (National Center for Education Statistics, 2015). This value equates to approximately 3,076,234 possible collected observations, which does not include home-schooled individuals, private school attendees, and those who obtained a General Education Development (GED) certificate (National Center for Education Statistics, 2015). The KHEAA also keeps information of postsecondary school attendees who are also recipients of one or more of their monetary awards. Between 1990 and 2009, the organization awarded an estimated 715,959 students a sum of $1,033,494,377 (Kentucky Higher Education Assistance Authority, 2010). The mass amount and complexity of high school and college student information collected by the KHEAA could result in a BD problem. Thus, BD handling computational algorithms should be utilized for modeling data, analyzing the effects of grants on student success outcomes, and data visualization. This research provides the design and evaluation of an analytical tool that accounts for some potential BD concerns in financial aid delegation optimization.

Conceptual Framework

Employees of the Kentucky state government and leaders in the higher education administration, who directly influence decisions in the development of financial aid policies, were interviewed in a focus group setting. Emphasis was placed on the value of the financial aid software tool in helping the administrators understand the impact of grant programs on student success. The purpose was to identify specific enhancements displayed information produced by the software as well as the software itself. Figure 1 provides a high-level overview of the F3A structure. The visual diagram depicts the flow of data through a proposed system that utilizes BI and DS methods and technologies. The process consists of the following steps:

1. The end user provides an initial input dataset consisting of historical information at the student level which includes population characteristics, financial aid details (i.e., type, amount), institution of attendance information (i.e., size, cost of attendance), and graduation outcomes.
2. The data is mined and modeled using the software application and processed on one or more servers. Data mining computing algorithms must work together with hardware resources for more efficient data processing (EMC Education Services, 2015).
3. The information produced by the software is formed into a visual representation, such as reports or dashboards, and presented through a Graphical User Interface (GUI).
4. The end user then provides new student scenarios for analysis. Steps 2 through 4 are reiterated until the end user has their desired results.

Evaluating the software tool for technical enhancements, regarding state monetary aid management, starts with asking influential leaders and administrators about their experience with the tool (Baer & Duin, 2014; National Commission on Higher Education Attainment, 2013; Pomeroy, 2014). The study identified potential benefits not currently realized when using analytics to examine financial aid resource allocation. Analytics can be utilized for predicting behaviors (Chintagunta et al., 2016; Gessner & Scott, 2009), reducing costs while making improvements (Soni et al., 2011; Teffeteller & Kish, 2012), and evaluating risks to prevent undesirable outcomes (Chan et al., 1999). Unfortunately, much of the higher education industry has been slow to adopt analytics (Ferreira & Andrade, 2016; Roscorla, 2015). The research question proposes what software enhancements should be made to the F3A to increase end-user benefits and encourage analytics for decision-making in grants allocation. Additionally, the study focused on creating a list of technological improvements for future analytics applications. Figure 2 displays a graphic representation of the conceptual framework of this research study.

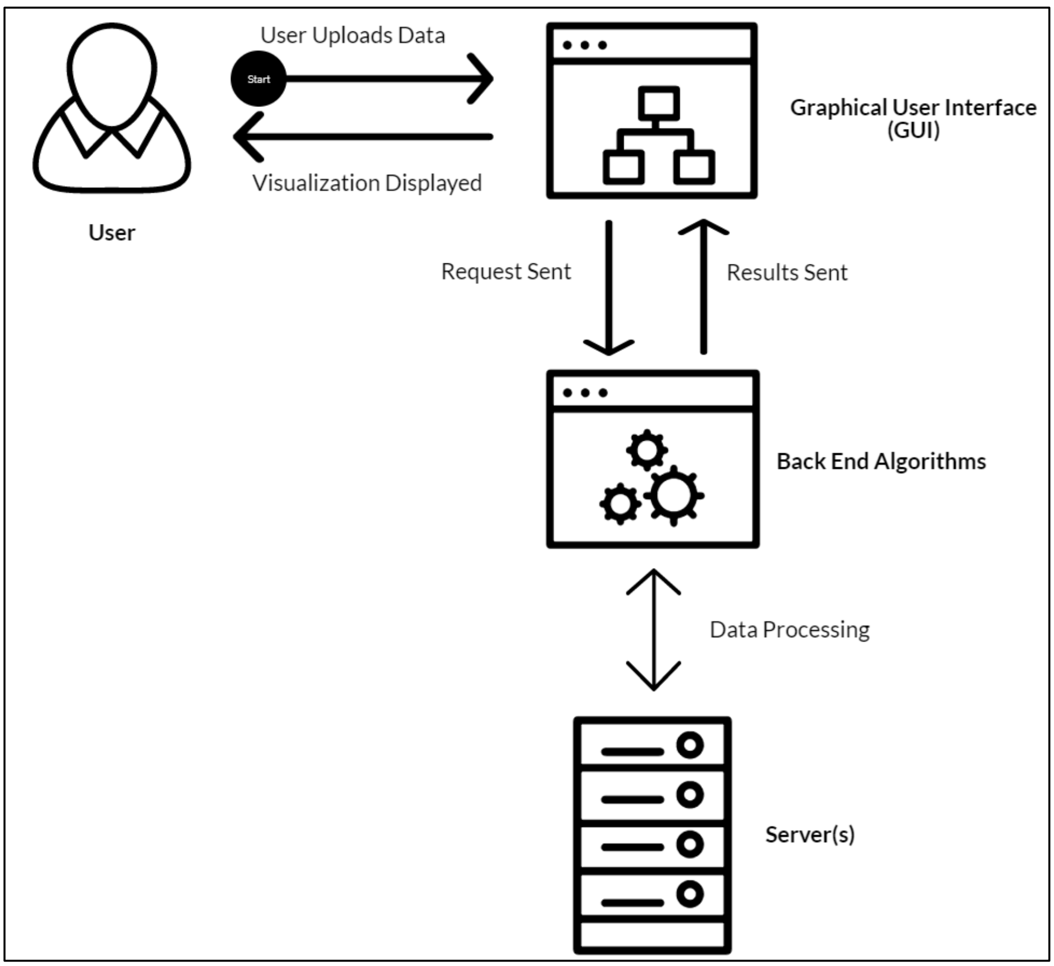


Figure 1. Overview of the financial aid analytics application system design.

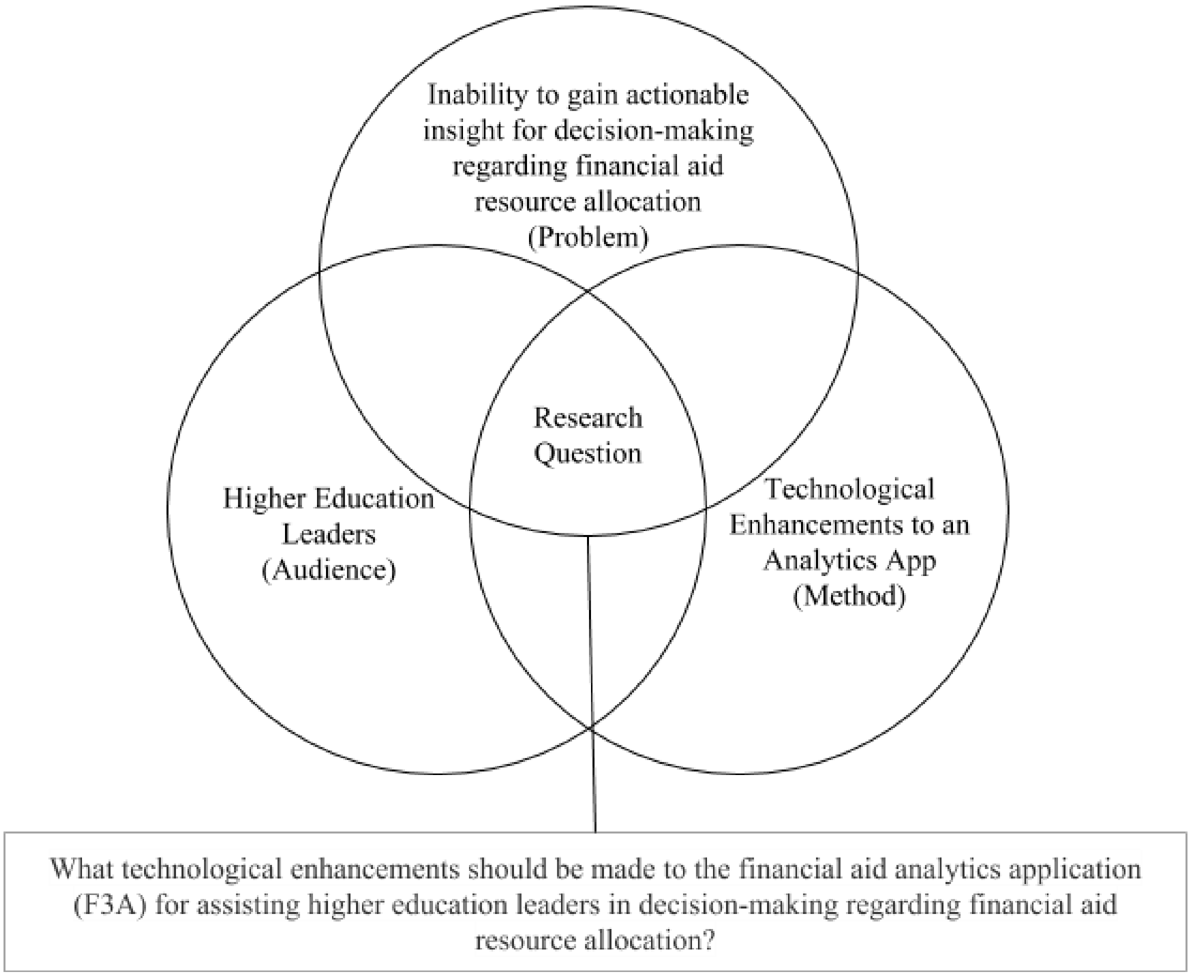


Figure 2. Conceptual framework for evaluation of a financial aid analytics application

Summary of Literature Review

This chapter first discussed the key benefits and shortcomings of BD and DS as well as their differences. Next, the chapter addressed issues with adopting technology in higher education with a focus on financial aid resource allocation. Another section described various aid programs, types, and previous analytical methods used for evaluating financial aid impacts on student success. This section also addressed the *Problem Identification* phase of the DSR methodology. The chapter concluded with an overview of the conceptual framework stemming from the challenges of using sophisticated analytics in higher education and the lack of research performed on financial aid delegation. Chapter 3 reveals the detailed design of the F3A.

CHAPTER THREE

This chapter details the design used for the development of a financial aid analysis application (F3A) created to address the issue of the lack of analytics in financial aid resource allocation. In Chapter 1 and Chapter 2, the research problem was identified as shown in the Design Science Research (DSR) methodology. After problem identification, the next phase of the DSR framework is to derive solution objectives for creating an artifact and to provide a corresponding design. This chapter details the solution objectives which guided the design and the development of the F3A as well as a detailed description of the tool's functionality. The design inspirations, diagrams, program code (Appendix A), and photos of the user interface (UI) (Appendix B) for the F3A design are discussed in detail.

Design Inspiration

The Cross Industry Standard Process for Data Mining (CRISP-DM), a framework for data mining projects that has been used by novices and experts alike (Shearer, 2000), inspired the majority of the F3A design. CRISP-DM was created by three Veterans in 1996 and successfully utilized and adopted by longtime data mining solutions provider SPSS Inc. (Chapman et al., 2000; Shearer, 2000). Over the years, the process has grown in popularity and was claimed to be the most widely used data mining model across various industries (Hiltbrand, 2013; Mariscal, Marban, & Fernandez, 2010). The CRISP-DM consists of six phases: (a) Business Understanding, (b) Data Understanding, (c) Data Preparation, (d) Modeling, (e) Evaluation, and (f) Deployment (Chapman et al., 2000; Hiltbrand, 2013; Kalgotra & Sharda, 2016). Each step contains multiple sub-phases, which guide the data analyst through the steps required to complete the process (Chapman et al., 2000; Hiltbrand, 2013; Kalgotra & Sharda, 2016; Shearer, 2000).

Business Understanding

The Business Understanding phase is likely the most significant part of the data mining process (Shearer, 2000). During this step, business needs and objectives are defined and explored to develop a detailed plan for valuation and create data mining goals for the project (Shearer, 2000). This phase has four key components (Chapman et al., 2000): (a) Determine Business Objectives, (b) Assess Situation, (c) Determine Data Mining Goals, and (d) Produce Project Plan. These steps help the data mining practitioner with identifying the needed data sources and possible modeling solutions (Shearer, 2000).

Determine Business Objectives

The first phase of the CRISP-DM process is to understand the business objectives and identify primary stakeholders (Mariscal et al., 2010; Shearer, 2000). The purpose of this research study is to analyze the performance of state financial aid awards in achieving postsecondary degree completion. Since state monetary assistance programs were developed to make college more accessible and increase degree attainment (Sjoquist & Winters, 2012), the primary business goal was derived as, *increase college completion rate for state financial aid recipients*. The primary stakeholders identified were financial aid policy makers and grant managers.

Another component of the Business Understanding phase was to specify a success metric and define the business success criteria (Mariscal et al., 2010; Shearer, 2000). Thanh and Haddawy (2007) developed a framework for estimating the minimal financial aid package needed for a student to enroll in a university. The measurement in their study yielded the projected probability a student would enroll given varying tuition assistance amounts (Thanh & Haddawy, 2007). The measurement utilized in this study closely mirrored that of Thanh and Haddawy. The defined measurement is the forecasted likelihood a student would complete college given varying financial aid packages. The tool also displays an overall, or totals, comparison of observed versus predicted values so users can evaluate the accuracy of the model (Piñeiro, Perelman, Guerschman, & Paruelo, 2008). The target measurement, graduation probability, was the outcome variable utilized in building the data mining model.

Assess Situation

The next step in the Business Understanding phase of the CRISP-DM process is to evaluate the situation by first outlining the resources used to develop the data mining software application (i.e., personnel, technology; Shearer, 2000). This research is a requirement of a doctoral dissertation study, which is an independent project authored by one person. Therefore, this project utilized one primary person, who currently works as a BI and DS consultant, possessing an educational background in statistics, mathematics, and computer science. Other contributing personnel include various Colorado Technical University doctoral faculty who supervised the dissertation study. Essential technology resources utilized for this study include a Hewlett Packard laptop with a Linux Operating System (OS), RStudio software, and the Shiny R package. RStudio is a powerful statistical computing application used by researchers and practitioners for data mining projects. RStudio has various BD processing packages (O'Leary, 2017). Shiny R is a web application framework, designed to be compatible with RStudio, which allows for user interaction via a graphical user interface ([GUI]; Gan, 2016). The next phase of designing the data mining application includes a statement of requirements, assumptions, and constraints of the project (Chapman et al., 2000).

The SPSS guidelines for topics that should be included in the requirements section are a completion schedule, a description of the quality of results, as well as security and the legal issues encountered by this project (Chapman et al., 2000; Shearer, 2000). Table 1 contains an estimated completion schedule, derived from industry standard timelines (Shearer, 2000), allocating a period of 5 weeks or 35 days duration.

Table 1

Completion schedule derived from CRISP-DM industry standard timelines

CRISP-DM Phase	Industry Standard Timeline	
	%	Count
Business Understanding	10 - 20%	4 to 7 days
Modeling	20 - 30%	6 to 9 days
Evaluation	50 - 70%	16 to 22 days
Data Understanding	20 - 30%	7 to 11 days
Data Preparation	50 - 70%	18 to 25 days
Deployment	5 - 10%	2 to 4 days
Total Allocated Time		35 days

Data mining models require adequate information to fulfill the research objective (Shearer, 2000). The dataset utilized in this design must contain the following:

- A minimum of 20 observations per variable (Ogundimu, Altman, & Collins, 2016); and
- A minimum of 100 financial aid recipients, of one or more state-funded aid packages, who have o graduated within a six-year window (Holmes & Jain, 2011).

Additional constraints include memory and processing capabilities of the laptop server. The HP laptop has a 1.60 GHz processor with 8GB of memory. In the event these specifications do not meet the needs of the software application, utilization of multi-threading independent processes or cloud computing servers for support was an option. The next section addresses probable risk factors associated with completing this project as well as contingency plans (Chapman et al., 2000) to mitigate these risks.

The three primary risks identified in the development of the data mining application include; utilizing synthetic data (SD) to create and test a tool intended for business use, a shorter period allocated for development, and limited hardware specifications. The timeframe allocated for the Data Preparation and Data Understanding phases were reduced to allow for the development of the tool and to permit the inclusion of new information that prevented the delay or failure of the project. In the event the development of the IT artifact was limited because of the hardware specifications of the laptop server, Amazon Web Services cloud computing and parallel job processing were elected as a backup plan. Amazon Web Services is a cloud services platform which offers scalable, inexpensive options for hosting web applications and providing information security that meets the requirements of the Family Education Rights and Privacy Act ([FERPA]; Amazon Web Services [AWS], 2015). The application could be transferred to AWS distributed systems which have the capacity to handle the processing and memory requirements of the project.

The next section in the CRISP-DM phase is to establish a glossary of business and data mining terms used in this project (Chapman et al., 2000). Relevant definitions for understanding this software development project are below:

- Accuracy: The rate at which outcomes are predicted correctly. This term used to define the overall success of a model (EMC Education Services, 2015).
- Area Under the Curve (AUC): The measure for the area under the plot of a model's true positive rate versus the false positive rate (EMC Education Services, 2015).

- Confusion Matrix: A table which displays how many of the actual outcomes matched the predicted outcomes (National Center for Education Statistics, 2016).
- Financial Aid Optimization: Making the most practical use of monetary aid resources (Boshardt et al., 2010) in a manner which assists college students in obtaining a degree.

A costs-benefit analysis (CBA) is usually performed to ensure project benefits exceed the project costs (Cellini & Kee, 2010). For projects related to the public-sector, such as this study, the population for which the tool is intended determines the value (Cellini & Kee, 2010). Chapter 4 summarizes the methodology used to evaluate the usefulness, or benefits, of the tool developed in this study.

Determine Data Mining Goals

As aforementioned, the business goal was to increase the college completion rate for state financial aid recipients. This objective must be defined in technical terms to derive the data mining goal ([DMG]; Chapman et al., 2000). The DMG for this study are defined as follows; predict how many full-time FTIC students will complete a postsecondary undergraduate program given academic history, demographic information, characteristics of their institution of attendance, and financial aid award package. The data mining success criteria for predictive analytics depends on the need for the model (Abbott, 2014). Due to the use of SD in building the model for this prototype application, a defined success criterion was not established.

Architecture Design

The F3A is a collection of five dashboards consolidated under one application to perform the main functions needed in the predictive analytics process. The dashboards are located on the same page and separated by tab panels (Appendix B). The names of each page include Upload, Describe, Verify, Model, and Predict. Figure 3 provides an architectural overview that depicts how the user should flow through each tab and the types of available functionality. When the user starts the program, they must first upload a file for analysis. After the file uploads successfully, the user will be able to view detailed field descriptions and select input variables to build a prediction model. Once the variables are input, the user can create a model and assess its performance. Lastly, the user can select a student profile and find the optimal aid amounts required to help the student achieve a specified probability of success.

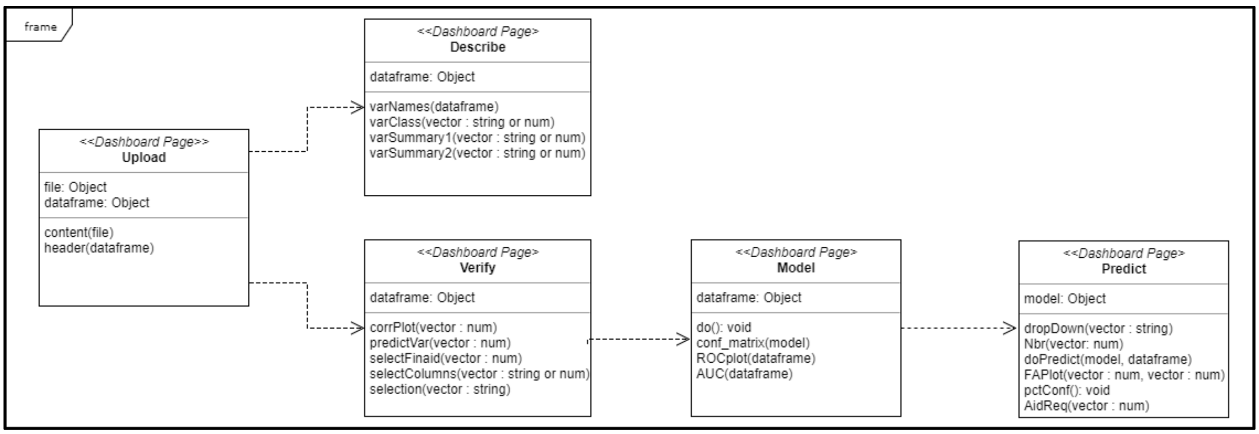


Figure 3. Financial aid analytics application architectural design.

Upload Page

In many BDA projects, the assembled data could include aggregate-level, structured, unstructured, and raw data (EMC Education Services, 2015). This project only accounts for the input of a relational dataset into the F3A. The test dataset used for this project was derived from Titanic passenger data in which the survivors were categorized as "graduated" and the deceased as "not graduated" (i.e. GRAD = 1 and GRAD = 0).

The data collection phase consisted of merely loading the dataset into the data mining tool (Chapman et al., 2000). The data importing procedure is a crucial step in the collection process and should be well standardized (Kramer, Müller, & Turowski, 2014). Comma-separated-value (CSV) files are common in various studies related to data import processes and performance (Eichinski & Roe, 2016; Kramer et al., 2014; Nirmala, Roopa, & Kumar, 2015).

Thus, this tool allows the user to choose a (CSV) file stored on an existing computer and upload it to the application. This design utilized the *read.csv()* function to import CSV data sets into the R Studio environment (EMC Education Services, 2015), but also provides the option to upload tab and semicolon separated files. The section below describes the *Upload* dashboard tab panel design followed by a sequence diagram (Figure 4) which details the interaction between the user and the system.

Name: Data Upload Tab

Type: Dashboard

Description:

Operations -

Name: contents()

Arguments: File location of a text file, delimiter type, and string format

Returns: A data frame

Pre-condition: The *shiny* package, a library of files which allows the user to develop a shiny app, is installed

Post-condition: The file is imported and the data is displayed on the screen

Exceptions: No

Flow of Events:

1. The user specifies whether the desired upload file contains fields delimited by a comma, tab, or semicolon.
2. The user specifies whether the desired upload file specifies strings via single quotes, double quotes, or no quotes.
3. The user clicks the "Browse…" button.

4. The user locates the text file to upload.
5. The user selects the text file.
6. If the file is valid, it is imported and a “file successfully uploaded” message is displayed to the user.
7. The file is transformed into a data frame object.
8. The data frame is displayed to the user in table format on the Upload tab.

Name: header()

Arguments: True or False

Returns: If the header box contains a check, the first few rows of a data frame are displayed. If the header box does not contain a check, the whole data frame is displayed.

Pre-condition: A text file is successfully uploaded.

Post-condition: The data is displayed as specified by the user.

Exceptions: No

Flow of Events:

1. The user has successfully imports a text file.
2. The user checks the header box.
3. The first few rows of the data frame are displayed.
4. The user unchecks the header box.
5. The whole data frame is displayed.

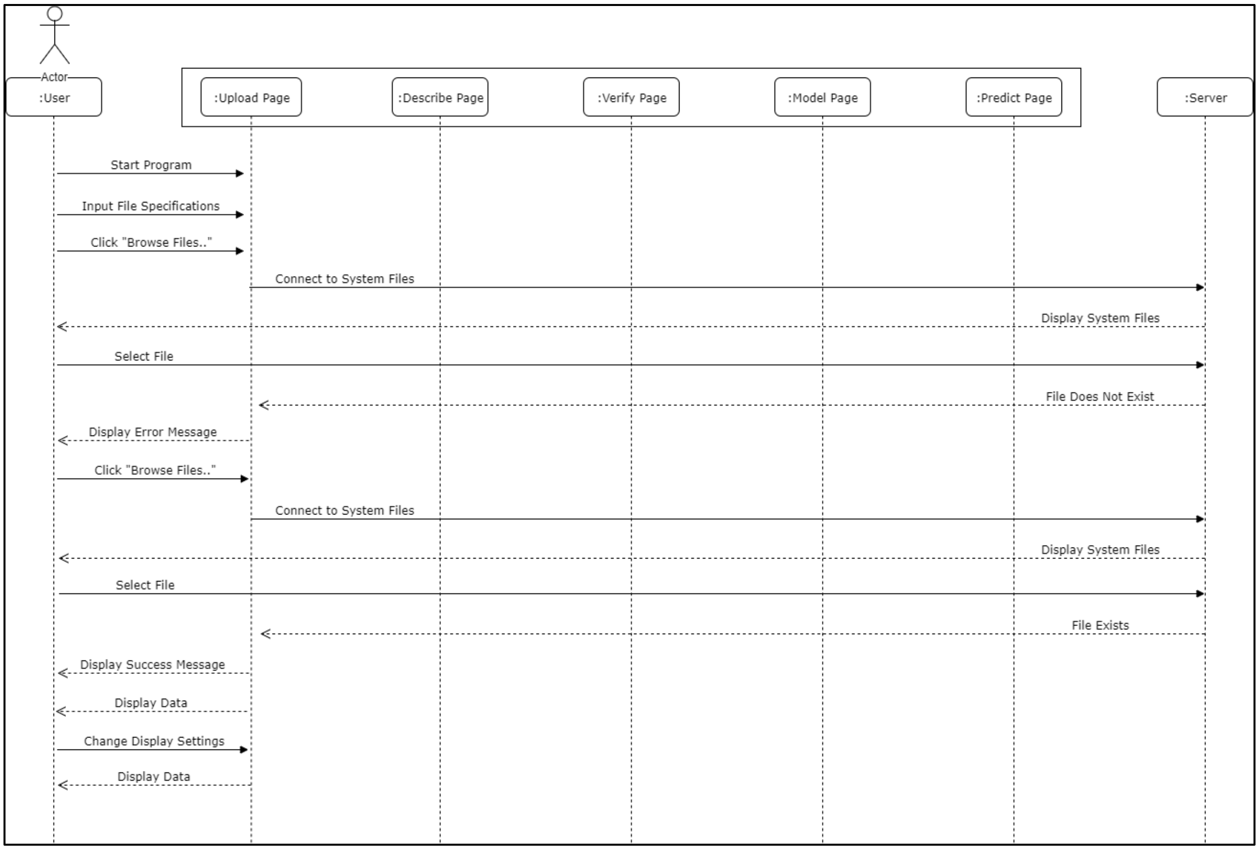


Figure 4. Upload page sequence diagram.

Describe Page

The Data Description phase involves the use of descriptive statistics to provide an overview report of the data so that the user can determine when there is enough information to continue further analysis (Chapman et al., 2000). The F3A produces a summary report from the imported data using BI techniques. The *summary()* function, which is part of the R language *faraway* package, provides univariate information about the data (Faraway, 2005). During this phase, the user assesses the report for accuracy (Chapman et al., 2000; Shearer, 2000). The section below illustrates the *Describe* dashboard tab panel design, followed by a sequence diagram (Figure 5) which details the interaction between the user and the system.

Name: Data Field Description Tab

Type: Dashboard

Description: Once data is successfully imported, the user moves to the Describe tab.

Operations -

Name: varNames()

Arguments: The names of each field in a data frame.

Returns: A drop-down menu with data frame field names.

Pre-condition: A data frame exists, and the user has selected the Describe tab.

Post-condition: The user can select a field name from a drop-down menu.

Exceptions: No

Flow of Events:

1. The user has successfully imported a text file.
2. The user selects the Describe tab.
3. A drop-down menu containing field names displays on the page.

Name: varClass()

Arguments: One field name from the imported dataset.

Returns: The field's class type (i.e., numeric, integer, string, date).

Pre-condition: A data frame exists, and the user has selected a field from the drop-down **menu**.

Post-condition: The variable’s class type is displayed to the user.

Exceptions: No

Flow of Events:

1. The user selects a field name from the drop-down menu.
2. The field’s class type is displayed.

Name: varSummary1()

Arguments: One field name from the imported dataset.

Returns: A statistical summary dependent on the variable type. Numeric variable returns the total count, the number of missing observations, the number of distinct variables, an information statistic, average, sum, and distance measurement. Character variables return a frequency table for each factor level.

Note: The R describe() function is used for this operation.

Pre-condition: A data frame exists, the user has selected a field from the drop-down menu, and the Hmisc package is installed.

Post-condition: Summary statistics for the selected field are displayed to the user.

Exceptions: No

Flow of Events:

1. The user selects a field name from the drop-down menu.
2. Summary statistics for the selected field are displayed.

Name: varSummary2()

Arguments: One field name from the imported dataset.

Returns: A statistical summary dependent on the variable type. Numeric variable returns minimum, average, median, maximum, and other quartile values. Character variables return a frequency table for each factor level.

Note: The R summary() function is used for this operation.

Pre-condition: A data frame exists, and the user has selected a field from the drop-down menu.

Post-condition: Summary statistics for the selected field are displayed to the user.

Exceptions: No

Flow of Events:

1. The user selects a field name from the drop-down menu.
2. Summary statistics for the selected field are displayed.

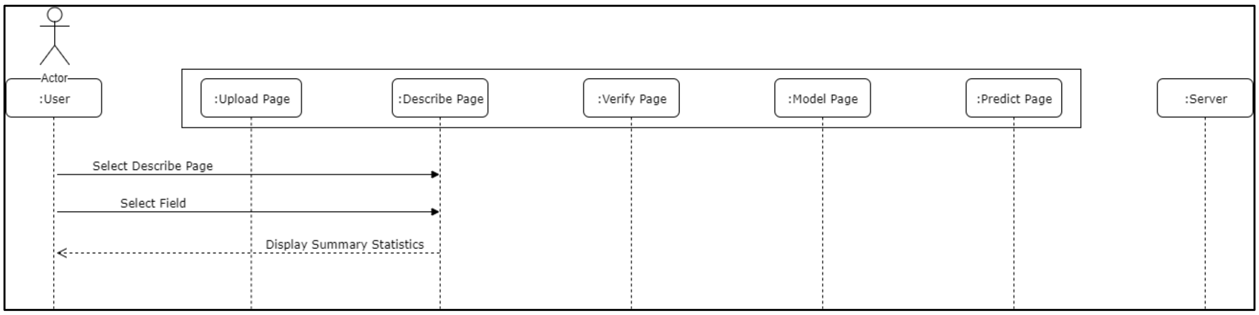


Figure 5. Describe page sequence diagram.

Verify Page

Data Selection is the first stage of the Data Preparation phase in which the user selects the desired fields and rows needed for a predictive model (Chapman et al., 2000; Shearer, 2000). The F3A provides a manual and automated feature selection option. The column selection page displays a custom input control function that is provided in the RStudio Shiny package. The user can select one or more column names and move them into the final dataset by clicking on the check box to the left of the fieldname. To reverse a selection, the user unchecks the box next to the fieldname. The user must select the dependent variable and financial aid factors for the construction of the predictive model from separate dropdown menus. The section below describes the *Verify* dashboard tab panel design and is followed by a sequence diagram (Figure 6), which details the interaction between the user and the system.

Name: User Variable Verification Tab

Type: Dashboard

Description: This page allows the user to specify the independent (i.e., predictors) variables, a financial aid variable, and a dependent (i.e., predicted) variable for a projection model. The page contains a correlation plot display, so the user can view the relationship between numeric variables in the dataset.

Operations -

Name: corrPlot()

Arguments: Two or more numeric vectors of the same length.

Returns: A correlation plot of the specified numeric variables.

Pre-condition: The corrPlot package has been installed, and a data frame exists.

Post-condition: A correlation plot is displayed to the user.

Exceptions: No

Flow of Events:

1. The user has successfully imported a text file.
2. The user clicks the Verify tab.
3. A correlation plot of all the numeric fields is displayed.

Name: predictVar()

Arguments: One field name from the imported dataset.

Returns: None

Pre-condition: A data frame exists, and the selected variable is a numeric binary indicator.

Post-condition: Stores the dependent variable as a single element vector.

Exceptions: No

Flow of Events:

1. The user expands the drop-down menu.
2. The user selects the field representing the dependent variable.

Name: selectFinAid()

Arguments: One field name from the imported dataset.

Returns: None

Pre-condition: A data frame exists, and the selected variable is numeric.

Post-condition: Stores the financial aid variable as a single element vector.

Exceptions: No

Flow of Events:

1. The user expands the drop-down menu.
2. The user selects the field representing the financial aid variable.

Name: selectColumns()

Arguments: One or more field names from the imported dataset.

Returns: None

Pre-condition: A data frame exists.

Post-condition: Stores the independent variables in a vector.

Exceptions: No

Flow of Events:

1. The user selects the check box next to the desired input variables.
2. The user deselects the check box next to the unwanted variables.

Name: selection()

Arguments: User-defined dependent variable, financial aid indicator, and independent variables.

Returns: A vector of all variables selected by the user.

Pre-condition: A data frame exists.

Post-condition: Displays selected variables to the user.

Exceptions: No

Flow of Events:

1. All fields are displayed to the user.
2. The user selects a binary success indicator.
3. The indicator is displayed to the user.
4. The user selects the financial aid field.
5. The field is displayed to the user.
6. The user selects the checkbox next to the desired input variable.
7. The input variable is displayed to the user.
8. The user deselects the checkbox next to one of the fields.
9. The field is removed from the user's display.

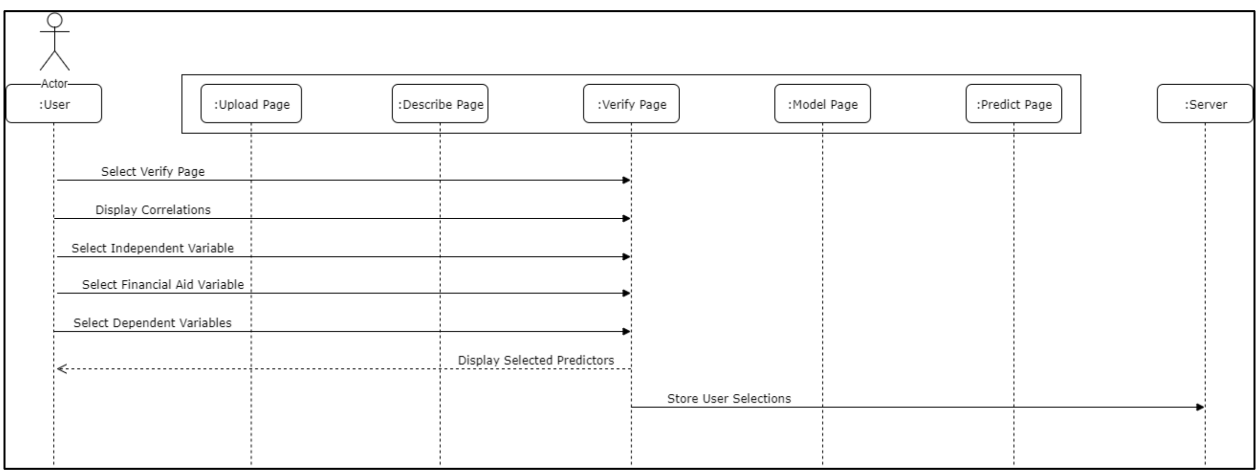


Figure 6. Verify page sequence diagram.

Model Page

Naïve Bayes is an easy to implement data mining technique used for prediction of classification problems (EMC Education Services, 2015). Advantages of Naïve Bayes classifiers include their ability to handle missing data and independent variables with little to no impact

(EMC Education Services, 2015). The Naïve Bayes approach utilizes conditional probabilities to determine the likelihood that a situation will occur (Witten et al., 2011). For example, if the probability of graduation for a female is known and the probability of graduation for someone who attended a community college is also known, then the conditional probability of graduation for a female who also attended a community college can be determined using this information. Another benefit of the Naïve Bayes classifier is its computational efficiency when analyzing large datasets with many combinations of conditional probabilities (EMC Education Services, 2015). Previous financial aid problems have also employed this method. Thanh and Haddawy (2007) presented a Bayesian Network financial optimization model at the 37th Annual Frontiers in Education Conference for predicting college applicant's probability of enrollment given a certain amount of financial aid. After attempting various network approaches, they found that the Naïve Bayes method yielded the best results (Thanh & Haddawy, 2007). Wei et al. (2016) utilized the Naïve Bayes algorithm to forecast whether students would receive a scholarship based on their academic and demographic information. However, one weaknesses of this technique is the assumption that each factor contributes equally and independently to prediction outcomes (EMC Education Services, 2015; Witten et al., 2011). This concept means correlated variables may deteriorate the accuracy of the model (EMC Education Services, 2015). This disadvantage was mitigated in the design of the software tool by providing the user with a correlation plot to use during the variable selection step. The Naïve Bayes algorithm used in the development of the F3A was obtained from the R package *e1071* package. The subsequent section describes the *Model* dashboard tab panel design, followed by a sequence diagram (Figure 7) which details the interaction between the user and the system.

Name: Model Performance Tab

Type: Dashboard

Description: This page allows the user to make a predictive model given their selection of variables from the Verify tab. A Naive Bayes classifier is used for the underlying model. The performance of the statistical model displays to the user with a confusion matrix, a ROC curve, an AUC metric, and a corresponding accuracy guide.

Operations -

Name: do()

Arguments: A user click signal.

Returns: None.

Pre-condition: Data frame and user-defined variable selections.

Post-condition: A Naive Bayes model object is created and stored.

Exceptions: No

Flow of Events:

1. The user has successfully imported a text file.
2. The user has specified input variables.
3. The user selects the Model tab.
4. The user selects the "Start Model" button.
5. A Naive Bayes model is produced and stored.

Name: conf_matrix()

Arguments: A statistical model object.

Returns: A confusion matrix data frame.

Pre-condition: A valid model object must be available.

Post-condition: A confusion matrix is stored in a data frame.

Exceptions: No

Flow of Events:

1. The user selects the “Start Model” button.
2. A confusion matrix displays on the screen.

Name: ROCplot()

Arguments: A data frame with the model's predicted values and a data frame with actual values.

Returns: An ROC curve plot.

Pre-condition: Two data frames available, user-defined variable selections, and pROC package installed.

Post-condition: An ROC curve plot object.

Exceptions: No

Flow of Events:

1. The user selects the “Start Model” button.
2. The model's ROC plot is displayed on the screen.

Name: AUC()

Arguments: A data frame with the model's predicted values and a data frame with actual values.

Returns: An AUC metric.

Pre-condition: Two data frames available, user-defined variable selections, and pROC package installed.

Post-condition: A single vector containing the AUC metric.

Exceptions: No

Flow of Events:

1. The user selects the “Start Model” button.
2. The model AUC displays on the screen.

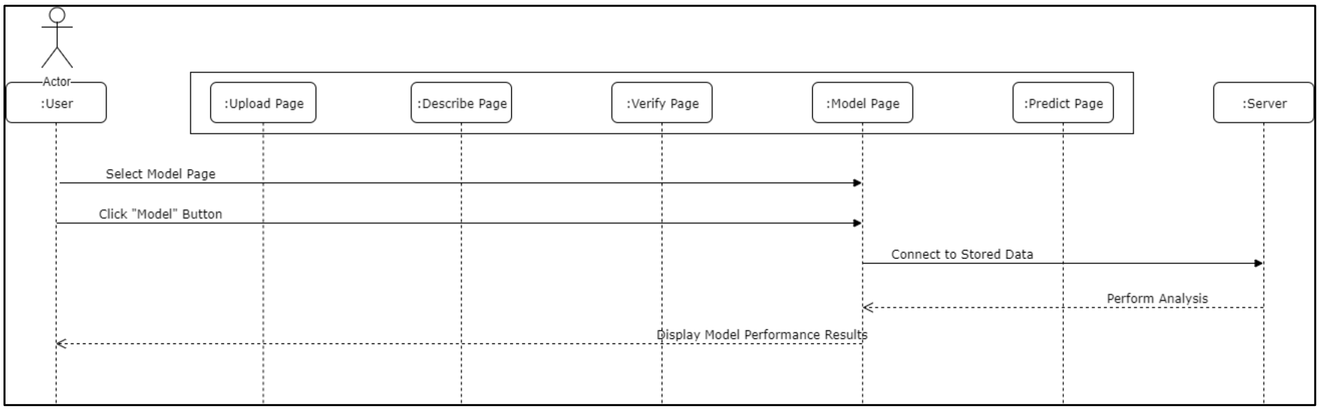


Figure 7. Model page sequence diagram.

Predict Page

User input widgets were created for the subsequent Prediction phase. In the Prediction step, the user can select a combination of student features and view their likelihood of graduating given varying financial aid amounts. The user is also able to obtain the minimum costs of a student graduating at a specific probability threshold. The section below describes the *Predict* dashboard tab panel design, followed by a sequence diagram (Figure 8) which details the interaction between the user and the system.

Name: Predict Student Success Outcomes Tab

Type: Dashboard

Description: This page allows the user to input student demographics and view the student's probability of success given varying amounts of financial aid. The user is provided two options to input student demographics: drop-down menus for character variables and free-form text fields for numeric variables. This page also has a slider for the user to input a desired probability of success and returns the minimum financial aid amount predicted to achieve success.

Operations -

Name: dropDown()

Arguments: A data frame.

Returns: A drop-down menu providing the distinct values of a character field in a data frame.

Pre-condition: A data frame is available, and the shiny package installed.

Post-condition: Dropdown selection menus for each character field are displayed to the user.

Exceptions: No

Flow of Events:

1. The user selects the Predict tab.
2. Dropdown selection menus are displayed to the user.

Name: Nbr()

Arguments: A data frame.

Returns: A free-form text field for inputting numeric values.

Pre-condition: A data frame is available, and the shiny package installed.

Post-condition: A series of free-form text fields representing the numeric variables in the data frame are available to the user for data entry.

Exceptions: No

Flow of Events:

1. The user selects the Predict tab.

2. Free form text fields are presented to the user.

Name: doPredict()

Arguments: A user click signal.

Returns: None

Pre-condition: A data frame is available, a model object is stored, and the e1071 package is installed.

Post-condition: A prediction set is created and stored in a data frame.

Exceptions: No

Flow of Events:

1. The user defines the student characteristics.
2. The user selects the "Predict Student Outcome" button.
3. The prediction dataset is stored in a data frame.

Name: FAPlot()

Arguments: One data frame containing predictions and another data frame containing financial aid amounts.

Returns: Plot of financial aid amount versus probability of success.

Pre-condition: Two data frames are available, and the shiny package is installed.

Post-condition: Exceptions: No

Flow of Events:

1. The user defines the student characteristics.
2. The user selects the "Predict Student Outcome" button.
3. A financial aid amount in conjunction with the corresponding probability of success is displayed to the user.

Name: pctConf()

Arguments: A probability the user specifies using a slider.

Returns: A probability value.

Pre-condition: None.

Post-condition: The probability is stored in a single element vector.

Exceptions: No

Flow of Events:

1. The user defines the student characteristics.
2. The user selects the "Predict Student Outcome" button.
3. The user selects the desired student success probability.
4. The probability value is stored.

Name: AidReq()

Arguments: A probability value

Returns: A financial aid value.

Pre-condition: A data frame containing probabilities and corresponding financial aid amounts.

Post-condition: A data frame containing the specified probability with the corresponding minimum aid amount estimated to achieve the probability.

Exceptions: No

Flow of Events:

1. The user defines the student characteristics.
2. The user selects the "Predict Student Outcome" button.
3. The user selects the desired student success probability.

4. The minimum amount of aid estimated to achieve the probability is displayed.

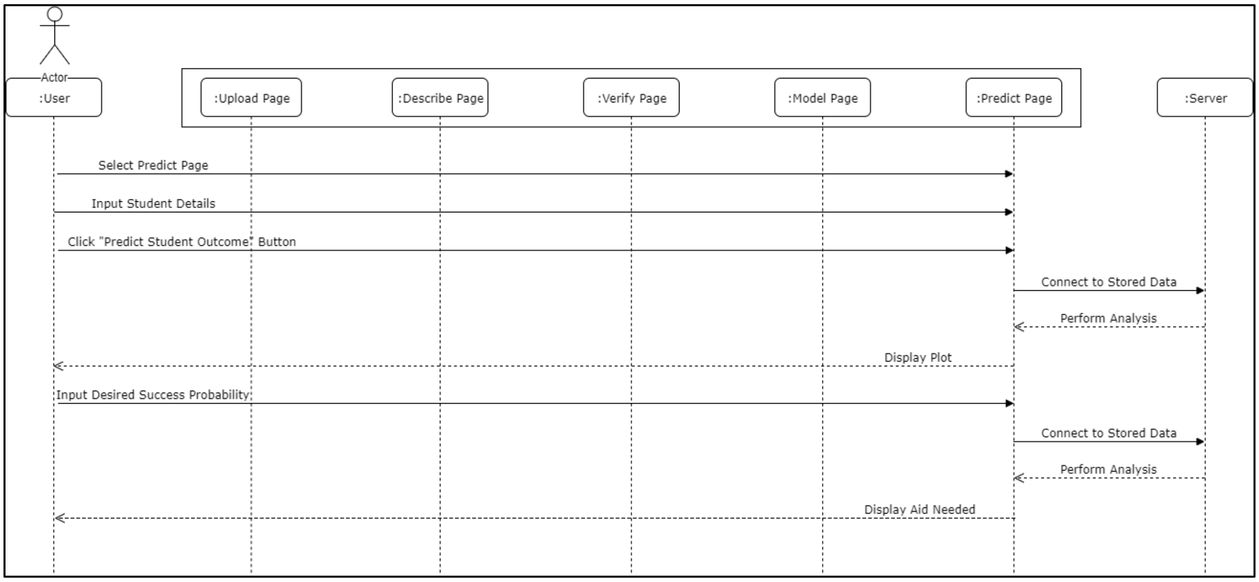


Figure 8. Predict page sequence diagram.

Summary of Chapter Three

This chapter explained the project's goal which was to design and develop an analytical software tool which displays the minimal amount of financial aid needed to achieve student success. Next, the project scope and constraints were discussed which includes a 35-day time limit for development utilizing R programming language, and other features such as the Shiny package used for data visualization. The chapter also elaborated on the software's design influences such as CRISP-DM, past studies addressing BD concerns, data mining techniques, and BI reporting techniques. Chapter 4 concluded with a detailed view of the F3A system and the steps for proper use of the tool.

CHAPTER FOUR

The issue presented in this study is the lack of *analytics* used for strategic decision-making in higher education (Ferreira & Andrade, 2016; Macfadyen, et al., 2014; Roscorla, 2015), explicitly in state financial aid resource allocation (Goldstein, 2005; Pomeroy, 2014; Thanh & Haddawy, 2007). The previous chapter described the design of the F3A tool in depth. This chapter covers the demonstration and evaluation of the application, which are the fourth and fifth steps of the DSR methodology (Hevner & Chatterjee, 2010). The purpose of this study is to evaluate a new software tool designed to ascertain the performance of state financial aid awards in postsecondary student success. This chapter examines the Design Science Research (DSR) methodology used in the evaluation of the financial aid analytics application (F3A). The following sections provides a description of the data collection method utilized for evaluating the IT artifact, the means for replicating the study, data analysis, and the steps taken to protect the identity of the participants who participated in this study. Study participants were previous and current employees of the Kentucky state government and higher education administration who provided insight for postsecondary education grants management decisions.

Research Tradition

Research, defined as the study of a phenomenon, historically focused on the natural sciences or events which occur naturally (Vaishnavi & Kuechler, 2008). Methodologies are used in research to guide the study by providing a procedure for data collection, evaluation, and interpretation (Creswell, 2014). The study of natural science includes modern traditions such as qualitative, quantitative, and mixed methods (Creswell, 2014). Conversely, the study of human-made inventions or phenomena has grown more significant which lead to the establishment of DSR (Vaishnavi & Kuechler, 2008). Mature theory encompasses quantitative methods for understanding relationships by analyzing data using statistical analysis (Edmondson & McManus, 2007). Many previous studies which address the impacts of financial aid on student success use a quantitative approach (Goldstein, 2005; Narozhnaya, 2015; Pomeroy, 2014; Ragland, 2016; Thanh & Haddawy, 2007). This research involved evaluating a tool intended for assessing the relationship between financial aid and student achievement, consequently the quantitative approach alone would not provide a framework for developing the software tool (Creswell, 2014). Similarly, the qualitative method could be used to understand the interaction users had with the IT artifact, but again the guidance for developing the tool does not exist within the qualitative methodologies (Creswell, 2014). Instead, the DSR methodology provides a blueprint for identifying a problem, setting an objective, and designing a solution (Vaishnavi & Kuechler, 2008). After the artifact has been developed, the following step in the DSR approach is to evaluate the tool using quantitative or qualitative methods (Vaishnavi & Kuechler, 2008).

The study applied the DSR methodology which guides the design and evaluation of an artifact promoting Information Technology (IT) to solve real-world problems (Hevner & Chatterjee, 2010). DSR is significant to Information Systems (IS) research because it addresses the importance of the IT artifact as well as its application in the respective field in which it will be practiced (Hevner & Chatterjee). DSR uses a six-step process distinguishing this method from simple practice which incorporates the following: (a) Problem Identification, (b) Solution Objectives, (c) Design, (d) Demonstration, (e) Evaluation, and (f) Communication (Peffers et al., 2006).

The intent of this study is to provide a grants management tool for financial aid managers; the DSR methodology was suitable for this study. DSR requires the development of an innovative framework, software, or hardware solution that is professionally relevant for both the IT and Business sectors (Hevner & Chatterjee, 2010). A research design is used to help the researcher identify the correct inquiry approaches within each methodology and guide the research process (Creswell, 2014). Designs assist the researcher in understanding which tools are appropriate for data collection as well as how to best retrieve the information needed to answer the research question (Rubin & Rubin, 2012). Additionally, research designs provide a plan to assure the required time, expertise, and money is available for the study (Rubin & Rubin, 2012). Though the DSR methodology differs from other methods, the evaluation of the artifact typically follows the research design using either a quantitative or a qualitative method (Vaishnavi & Kuechler, 2008). This study used a qualitative research technique for the evaluation of the software application.

Research Questions and Propositions

The central research question for the study is: What technological enhancements should be made to the financial aid analytics application (F3A) for assisting higher education leaders in decision-making regarding financial aid resource allocation? The research proposed an evaluation of the IT artifact for understanding conceivable effects of grants on postsecondary degree completion and assisting in strategic monetary aid allocation decisions.

Research Design

This research followed the DSR methodology, which is used to evaluate an IT artifact created to solve a real-world problem (Hevner & Chatterjee, 2010; Vaishnavi & Kuechler, 2008). The IT artifact is evaluated using a focus group, which focuses on gathering new data from a target population (Bryman & Bell, 2011; Kreuger & Casey, 2008). A focus group was held with analysts and leaders who support the state of Kentucky's major financial aid programs. The focus group was comprised of eight participants. The participants observed a demonstration of the F3A and responded to questions about the technology with the objective of obtaining a list of technological enhancements.

Population and Sample

A population narrows the research study by defining a group of individuals with similar characteristics designated for examination (Lofland, Snow, Anderson, & Lofland, 2006). The participants in this study were administrators who currently, or have previously, provided higher education leaders with expert knowledge regarding the development of financial aid programs (Long, 2009; Macfadyen et al., 2014). The population used to conduct the DSR study was a group of state government researchers in Kentucky who have explored data to assist financial aid policymakers in decision-making. The Bureau of Labor Statistics (2017) reported that there were approximately 3,283 Kentucky state government employees in 2016. The national percentage of state employees who work under financial operations is 12.07% while only 3.54% hold occupations relative to computers and mathematics (Bureau of Labor Statistics, 2016). Therefore, approximately 511 Kentucky state employees worked to support finance operations or in analytics related positions. The Kentucky Commonwealth advertises roughly 226 state agencies with only one bureau designated responsibility for financial aid program development. This yields an estimated population size of 2-3 employees who work to support financial aid program development. This exemplification is appropriate because it adequately represents those who provide analytics for postsecondary grants management in Kentucky. Since it is time-consuming and resource exhaustive to study a whole population, researchers utilized sampling groups to represent the population (Bryman & Bell, 2011; Creswell, 2014). This study utilized a focus group, in which the recommended number of individuals per group is two to ten (Bryman & Bell, 2011; Kreuger & Casey, 2008). In order to reach saturation Kreuger and Casey (2008) recommend conducting at least three to four focus groups. Conversely, the focus group evaluation method recommends purposive sampling for selecting participants, which appropriately aligns with the research topic (Bryman & Bell, 2011). Since the population size of state financial aid analysts was extremely limited within each state and the time and expense of gathering data from multiple state agencies was beyond the scope of this study, a single focus group was conducted with an established minimum number of at least two participants.

Sampling Procedure

The purpose of the sampling method, determined by the research methodology, is to ensure the researcher has appropriately chosen a group to be studied, which is representative of their population (Bryman & Bell, 2011). This research study utilized purposive sampling. Purposive sampling is a non-random strategic form of selecting participants who appropriately align with the research topic (Bryman & Bell, 2011). Purposeful sampling was applicable for this study because of the small size of the population (Lofland et al., 2006).

The focus group participants must fit the criteria of the subjects addressed in the research question (Bryman & Bell, 2011; Lofland et al., 2006). In this case, prospective contributors were higher education administrators who directly influence financial aid allocation decisions in the state of Kentucky. The contributors were asked to volunteer their time and were required to complete and return a consent form (Appendix C). Afterward, the respondents were sent a detailed invitation stipulating details for participation in the focus group.

Qualitative interviews are usually recorded with an audio device and transcribed (Bryman & Bell, 2011). Therefore, the focus group session was recorded with a video camera and an audio recorder. The audio file was sent to Rev.com and transcribed. The focus group schedule consisted of a 20-minute software demonstration and a 60-minute participant interview session. The transcribed data was saved in a password-protected Microsoft Word document on a secure flash drive.

Instrumentation

In a qualitative study, the researcher is the primary instrument for analyzing the participant's feedback (Lofland et al., 2006). Qualitative researchers gather data by questioning the participants about an experience or event and inferring themes from the information (Rubin & Rubin, 2012). Other instruments intended for use in this study include: (a) a focus group checklist, (b) an interviewer form, (c) an agenda, (d) a focus group protocol (Lofland et al., 2006) and (e) a focus group participant packet. The focus group protocol included open-ended, semi-structured questions, which encouraged contributors to elaborate on their experiences concerning the research topic (Rubin & Rubin, 2012). The interview also included probing questions as they can be helpful in gaining additional clarification or insight into the topic (Lofland et al., 2006). The first question was broad with the objective of encouraging the focus group members to speak freely and engage in discussion. The second and third questions were confined to specific aspects of the application. The remaining questions were much more restricted to the usefulness of the tool's decision-making component, which is the focus of this research study. Each of the questions aimed to elicit additional information to acquire a list of technological improvements for the F3A. The following were the questions were presented to the focus group participants:

1. What is your initial impression of the software tool?
2. Are enhancements necessary or helpful to make this tool more useful, more effective, and/or more efficient?
3. In general, what would you take away from this tool?
4. In general, what additions would you make to this tool?
5. In what ways did this tool help you understand how financial aid impacts college student degree completion?

6. What additions would you make to this tool in helping you understand how financial aid impacts college student degree completion?
7. How do the tools you currently use for understanding financial aid impacts on Kentucky college student degree completion compare to this tool?
8. What benefits does this tool provide in gaining insight into financial aid resource allocation?
9. What pitfalls does this tool have for gaining insight into financial aid resource allocation?
    a. What might you add or take away from this tool to overcome these pitfalls?
10. Do you think it may be appropriate to use this tool making financial aid decisions? If so, how, and when in the process might it be applied?

Each participant was provided the questions before the start of the presentation to prepare the respondents, redirect their attention, and encourage interaction for insightful discussion. The member responses were recorded with two devices including a video camera and audio recorder. Pencils and paper were distributed for note taking. The interview guides, notes, audio recording, and transcripts were labeled with the date and time.

Validity

Qualitative validity occurs when the researcher has taken measures to ensure the accuracy of the study results (Gibbs, 2007). The researcher, participant, and the reader view validity from different perspectives (Creswell & Miller, 2000). The researcher utilizes validity strategies (Creswell, 2014) to ensure the vital defining principles - trustworthiness, authenticity, and credibility – are met from all viewpoints (Creswell & Miller, 2000). Confirmability exist when the participants of a research study were accurately portrayed (Creswell, 2014). Creswell (2014) suggests using one or more strategies to determine if the study results are accurate. This research study included two strategies: member checking and presentation of discrepant data (Creswell, 2014; Riege, 2003). Member checking allows the participants to review the transcripts, data analysis, and results to determine and approve their accuracy (Creswell, 2014; Yin, 1994). Presenting discrepant information contradicts the themes defined by the researcher and increases the validity of the research study (Creswell, 2014). Credibility is significant to the trustworthiness of a research study because it ensures realistic findings through the utilization of established research methods (Shenton, 2004). Although the DSR methodology is the least popular in traditional research methodologies, the assessment of the IT artifact - required in DSR (Vaishnavi & Kuechler, 2008) – was performed via a focus group which is known to be a well-established technique for data collection (Bryman & Bell, 2011; Kreuger & Casey, 2008).

Reliability

Qualitative reliability demonstrates that the study can be carried out consistently across, or repeated in, other studies with different researchers and yield similar results (Gibbs, 2007; Riege, 2003). Dependability, like the concept of reliability, may be achieved by describing the research plan step-by-step and the actual outcome or execution of the plan (Shenton, 2004). Yin (2009) recommended meticulous documentation of all steps taken during the research process to ensure reliability. This study used the following strategies for documentation:

- Posed semi-structured questions during the interview process (Yin, 1994);
- Recorded data using an audio device (Nair & Riege, 1995); and
- Ensured code definition integrity remained consistent (Creswell, 2014).

The use of multiple strategies to prevent discrepancies in future research ensures increased reliability and dependability of the study (Creswell, 2014). Transferability is essential for the replication of a study, as a result the following details of the data collection process were recorded after data collection was complete: the actual number of participants and timeframe of the focus group (Shenton, 2004). Other areas, such as the population sample designated for study and the methodology, also support the transferability of the study (Shenton, 2004).

Data Collection

The central research question centered on a target population of state employees who support the development of grant programs for understanding the financial aid management tool. The data collection method for this study was to conduct one focus group with a minimum of two participants. Although three to four focus groups were recommended to achieve data saturation and extract common themes (Bryman & Bell, 2011; Kreuger & Casey, 2008), the population size of candidates was minimal and conducting more focus group sessions would exceed the scope of the study. Focus groups involve interviewing participants from the target population (Rubin & Rubin, 2012) in groups two to ten people (Bryman & Bell, 2011; Kreuger & Casey, 2008). Focus groups are performed when the researcher wants to understand the attitudes and ideas people have on a specific topic or wants to pilot-test new concepts (Bryman & Bell, 2011; Kreuger & Casey, 2008), such as a software tool.

The focus group conducted for this study was comprised of eight participants and the protocol consisted of open-ended semi-structured questions. Bryman and Bell (2011) explained the benefits of open inquiries which include: allowing for unique and unexpected responses, freedom in response selections, and introducing the researcher to new areas where they lack knowledge. Conversely, open-ended questions permit the opportunity for participants to share their experiences, encourage variability between participant responses, and engage respondents as participant researchers (Bryman & Bell, 2011). By providing semi-structured questions with probing statements, the researcher can aim the questions at a topic and extract more information from the participants when required (Lofland et al., 2006; see Appendix D). The focus group occurred at the KHEAA facility on December 17, 2017, from 11:00 AM to 1:00 PM. The participants were contacted via email to schedule and confirm the focus group time and location. Following the interview, participants were contacted by means of a follow-up email thanking them for their cooperation and providing them an analysis of their feedback for review.

The focus group followed the process below:

1. Developed rapport or trust with each participant by providing consistent, clear communication concerning the intent of the study and valuing each participant for their individual contribution, as detailed by Lofland et al. (2006);
2. Stated the risks of the study and informed the participants their cooperation was voluntary and confidential, as suggested by Kreuger and Casey (2008);
3. Acquired a sign-in consent form with each informant's signature (Appendix C);
4. Ensured the adequate materials were available for participants before they arrived, as discussed by Kreuger and Casey (2008);
5. Informed the interviewees that the focus group will be recorded to ensure validity, as stated by Bryman and Bell (2011);
6. Used the prepared interview guide which contains information regarding the interview process;
7. Thanked informants for their participation and provided them with a timeline for the transcription, as explained by Rubin and Rubin (2012).

The interviews were transcribed through a vendor and saved in a password protected Microsoft Word file. The file was stored on a password protected flash drive adding an additional layer of security. The data was transcribed using a service provided on Rev.com, which ensures confidentiality by guaranteeing the nondisclosure of participant information unless requested by a legal or government entity.

Data Analysis

The study utilized the DSR methodology with a qualitative approach for evaluating the software tool. The data analysis process takes the information gathered in a qualitative study and converts it into results that could be used to resolve the research question (Lofland et al., 2006; Rubin & Rubin, 2012). Rabiee (2004) stated that reducing the amount of information and focusing on the purpose of the study aides in understanding, handling, and cleaning the data in the analysis process. The goal of conducting the focus group for this research study was to retrieve a list of technological enhancements for the F3A. In many qualitative studies, the data is coded to extract themes, events, or concepts from the information (Creswell, 2014; Lofland et al., 2006; Rubin & Rubin, 2012). However, this technique will not allow for the retrieval of a list of improvements for the F3A. Rabiee (2004) explained that the information received from a focus group session could be easily presented with simple language and quotes from the participants. Therefore, the list of enhancements was created using the exact quotes of the participants. The following process was used to analyze the data collected:

1. Transcribed the data using a transcription service called Rev.com;
2. Read the transcription multiple times to obtain a clear understanding of the member's feedback, as suggested by Rabiee (2004);
3. Substituted names, locations, and events revealed during the focus group session with pseudonyms, as detailed by Creswell (2014) and Lofland et al. (2006);
4. Organized the quoted statements made by the participants about improving the tool and presented them in in the form of a list, as described by Rabiee (2004);

5. Used data reduction to eliminate, or remove, excessive information, as explained by Rabiee (2004), to develop a finalized list; and
6. Provided participants with a copy of the list for comments and corrections, as discussed by Rubin and Rubin (2012).

After receiving the feedback from the participants, a list of improvements for the F3A was derived using the transcription and amendments from the participants.

Ethical Considerations

An important aspect of a researcher's responsibility is to conduct research ethically, with full honesty, respect, and care for the well-being of those who participate in the research (Bromley, Mikesell, Jones, & Khodyakov, 2015). This study was designed to uphold that obligation. Prior to starting data collection and analysis, a permission form was obtained from the KHEAA site manager requesting authorization to perform a research study at their facility (Appendix E). Additionally, a consent form was signed by each participant before the focus group commenced (Appendix C).

The *Belmont Report* protocol was maintained as the ethical standard while conducting this study. The Belmont Report primarily covers the protection of human subjects involved in research studies (Department of Health, Education, & Welfare, 1974). The protocol requires that research participants are notified of possible risks during the study, that their information will be kept secured to ensure confidentiality and that informed consent must be provided to the researcher before participation begins (Department of Health, Education, & Welfare, 1974). Additionally, the core values of the Belmont Report – respect, beneficence, and justice (Vollmer & Howard, 2010) – were applied throughout this research study. Rubin and Rubin (2012) recommend research interviews be conducted at a place where participants feel comfortable, have privacy, and few interruptions. The focus group was conducted in a meeting room at the KHEAA facility, located in Frankfurt, KY. Participants were required to sign a consent form before participating in the focus groups (Appendix C). Research subjects may share their personal feelings and opinions during a study, so it is vital to ensure participant's' protection by providing confidentiality (Creswell, 2014; Rubin & Rubin, 2012). The participants are offering their time and experience, consequently the researcher owes them confidentiality in return (Rubin & Rubin, 2012). The participants were informed that their participation was voluntary and that they could terminate the study at any time (Kreuger & Casey, 2008). The consent form also informed the participants of these alternatives. (Kreuger & Casey, 2008). The research study ensured the privacy of the informants by using pseudonyms for names, locations, and events revealed in the study (Creswell, 2014; Lofland et al., 2006). Additionally, participants were provided with a copy of the data analysis and permitted to edit or remove any comments they provided (Rubin & Rubin, 2012).

Data and records were stored in a locked file cabinet and on a password-protected hard drive to keep participants' data safe. These materials were only accessible to the primary researcher and transcriptionists. Data transport was also limited to minimize the risk of losing information. At the conclusion of the study, the media device was cleared by overwriting the data file and then performing a manufacturer reset of the instrument.

Summary of Chapter Four

Chapter 4 discussed the DSR methodology used in this study for evaluating a financial aid analytics application. Next, the preference of conducting a focus group was explained as the selected method of artifact evaluation. Also, the process of data collection, focus group interview questions, and data analysis techniques were detailed. Finally, the chapter described the measures taken to protect the research participants' privacy. Chapter 5 includes the data collected from this research study, demographics for the participant population, and a list of technological enhancements for the F3A, which was derived from the focus group feedback.

CHAPTER FIVE

This DSR study involved the evaluation of a software application by posing the question "what technological enhancements should be made to the financial aid analytics application (F3A) for assisting higher education leaders in decision-making regarding financial aid resource allocation?" The issue addressed in this study was the lack of analytics used for strategic decision-making in higher education, particularly in state financial aid resource allocation. The purpose of this investigation was to develop a list of technological enhancements for the F3A. Chapter 5 presents the data collected after conducting a focus group to evaluate the software application. The results exhibited in this chapter include the participant demographics, direct responses to the interview questions, the list of enhancements for the F3A, and the researcher's findings.

Participant Demographics

The research participants in this study included Kentucky state government and college leadership and those who support leadership in making financial aid delegation decisions. All focus group members were contacted via emails relaying information about the purpose of the research study, the location and time of the study, the focus group agenda, and the associated questions. For individuals who agreed to participate in the study, a follow-up email was sent to thank them for their input and responses to any outstanding questions. All participants' line of work involved creating, evaluating the impact of, and delegating aid monies to potential college-goers and attendees of post-secondary education institutions in Kentucky. Table 2 discloses each focus group member's gender and job title retained at their respective organization. There were a total of eight participants; five males and three females. All participants agreed to the electronic taping and filming of the focus group interview. Participant quotes were not associated to any one person and remained anonymous. The request for participants was emailed to a total of 21 people located at eight different institutions in Kentucky. One organization provided seven focus group members, and another organization accounted for the final member of the group member.

Table 2

Participant demographics

#	Gender	Position
1	Male	Director of Research
2	Male	Executive Director
3	Female	Financial Aid Director
4	Male	Vice President of Operations
5	Female	Student Aid Director
6	Female	General Counsel
7	Male	Chief Executive Officer
8	Male	Research Analyst

Presentation of the Data

The eight focus group participants were asked 10 questions over a 90-minute collective interview session. The questions for this study were as follows:

1. What is your initial impression of the software tool?
2. Are enhancements necessary or helpful to make this tool more useful, more effective, and/or more efficient?
3. In general, what would you take away from this tool?

4. In general, what additions would you make to this tool?
5. In what ways did this tool help you understand how financial aid impacts college student degree completion?
6. What additions would you make to this tool in helping you understand how financial aid impacts college student degree completion?
7. How do the tools you currently use for understanding financial aid impacts on Kentucky college student degree completion compare to this tool?
8. What benefits does this tool provide in gaining insight into financial aid resource allocation?
9. What pitfalls does this tool have for gaining insight into financial aid resource allocation?
    a. What might you add or take away from this tool to overcome these pitfalls?
10. Do you think it may be appropriate to use this tool making financial aid decisions? If so, how, and when in the process might it be applied?

The focus group participants were employees of Kentucky state and college leadership who support leadership in making financial aid delegation decisions. A list of eligible participants, along with their professions, was provided to the interviewer by the Kentucky Higher Education Assistance Authority. Eight participants attended the focus group. The number of respondents was greater than the minimum number required to hold a focus group and less than the maximum number which could result in negative impacts on the effectiveness of the group responses. The focus group session schedule was provided to the participants. The software demonstration and group interview session were recorded via video camera and audio recorder with permission from all participants. The recordings were saved to a password-protected flash drive and transcribed using a third-party service Rev.com. Member checking and presentation of discrepant data were used to ensure confirmability and credibility of this study. The following sections detail the analysis of the data collected.

Interview Question 1

What is your initial impression of the software tool?

As shown in Table 3, the answer to this question was similar for many of the participants. In summary, they desired to see the *application of real-world data*. The participants responded with positive feedback regarding the potential of the tool, but as mentioned in the Design section of this study, the tool was developed using synthetic data. The focus group members were concerned with how the F3A might perform given real student demographics, success indicators, and outcomes.

Table 3

Participant responses to question 1

#	Responses
1	For academic measures this would be great!
2	I think it's promising. I would like to see it with actual data.
3	We understand vertical relationships which are probable correlations across all your variables but if we saw real data I think that's going to be much more of a tell as to what [the tool] could reveal in terms of the relationship.
4	It would be really nice to see the real data but, I think it's got a lot of potential.

Interview Question 2

Are enhancements necessary or helpful to make this tool more useful, more effective, and/or more efficient?

The responses to this question yielded the need for improvements to the graphical user interface (GUI). The feedback shown in Table 4 revealed the respondents desired that the front-end should be more intuitive, easier to operate for non-technical users, and more visually pleasing. They also explained their desire to create custom student segments or profiles. Another recommendation was to develop a system to provide the end user with the most up-to-date information. One suggestion was to manually update the data and adjust if necessary, while the other process involved an automated approach by connecting to an institution's database and scheduling a data refresh cycle.

Table 4

Participant responses to question 2

#	Responses
1	At certain point you're going to have to refine kind of the GUI frontend.
2	…make it prettier and more intuitive
3	The end-user is not going to be a computer programmer, but is going to be relatively sophisticated in their understanding. So, they understand what the variables are, but they want to make this, this and this into a profile. And then know what's going to happen to their uploaded 150 applicants this semester. That would make the biggest difference in who eventually would decide to use this or not. Again, I hate to say this, but most people really don't want to look behind the green curtain.
4	I think you need to think about how to update the bottom-line databases. If I take it to my

	university and I put it in there and it works great this year and it probably work okay next year but not as great as the first year because my data is getting older. So, you need to kind of think about how to oh, well, go to this website and click here and it will automatically update the whole system. By that it means probably managing to get an institution's data and format it appropriately and install in our place new data with old data.
5	A refresh of your historical data and once you put that new data in all of the calculations in the R models would automatically be updated and so that it would be self-updated.

Interview Question 3

In general, what would you take away from this tool?

Most of the participants fell silent when asked which components should be removed from the tool. Two participants responded to the questions and explained that they did not want to take any of the functionality away from the tool but wanted to add more (Table 5). Thus, the responses to most of the questions address additions to the F3A.

Table 5

Participant responses to question 3

#	Responses
1	I think it's more about adding a couple of things to it.
2	I agree.

Interview Question 4

What additions would you make to this tool?

***Probe Question* - What are some of the reasons for adding the mentioned component to this tool?**

Table 6 includes the recurring response to this question was associated with the addition of more variables. The group was interested in learning more about dual credit, transfer, and returning students since these students typically have hours when they enter a program. One participant suggested adding the program length as a factor to the predictive model since certifications, associates, bachelors, and graduate degrees differ in length. Student segmentation was mentioned with a more detailed approach for the F3A to include preset profiles for the end user to select and view possible student success outcomes.

Table 6

Participant responses to question 4

#	Responses
1	More variables.
2	More variables.
3	Just variables.
4	Could you have a major bar? An academic major bar?
5	The transfer cohort.
6	We’re having such a huge emphasis on dual a credit, how many credit hours are they starting with at the trans-institution which is kind of like the transfer thing though.
7	I think another variable that would be helpful is if you put the credential length. So we knew how many years, if you're talking about predicting certificates, you're looking at something short term. If you're looking at associates or bachelor's degrees, or even graduate degrees, longer terms on those models. I don't know if enrollment intensity is

	something you're getting there, but that impacts it very heavily.
8	I’m thinking large scale profiles. I want full-time business majors or I want part-time English majors that are over 25 years old returning to school. What if you had just like three or four or five different profiles, it could be a quick select and essentially, they’re macro defined. You could establish the individual variables in a profile so that you create a profile like that. I think that would make it much smoother, faster, quicker.

Interview Question 5

In what ways did this tool help you understand how financial aid impacts college student degree completion?

Though one of the participants was able to see how they could use this tool to request more money for aid delegation, another participant explained they would need to see more data to answer this question accurately. As shown in Table 7, the synthetic data used for the software demonstration was a repeated challenge in receiving feedback from the respondents.

Table 7

Participant responses to question 5

#	Responses
1	If you were trying to sell to the administration that you wanted more financial aid budget, then you could say, "Okay, if you're only going to give me this much, this is what you can expect.”
2	I think you can see the problem side. I think you would have to see it with data.

Interview Question 6

What additions would you make to this tool in helping you understand how financial aid impacts college student degree completion?

There were two similar yet differing responses, shown in Table 8, for the adding components to this tool to help understand financial aid impacts. One participant was interested in understanding how multiple programs affected student degree completion. At one point, they stated that the different aid programs have different impacts and thus wanted to be able to toggle the amounts from various programs. Conversely, the other participant only wanted to see the total amount of aid needed for the student. In the event these different aid types have varying impacts on student success, it may be difficult to model the outcome of a student given the total amount of aid awarded to them.

Table 8

Participant responses to question 6

#	Responses
1	Rather than using a single financial aid variable, could you use multiple programs? So, you could have a KEES bar, a CAP bar, an institutional bar, so you could change all of those.
2	The total aid, if they qualify for CAP and KTG, and KEES, additional aid.

Interview Question 7

How does this tool compare with other tools you may currently use for understanding financial aid impacts on your institution's student degree completion?

Table 9 shows that both respondents who had employed the use of some software application for financial aid management had constructed the tool themselves. One participant was able to look at the impact of financial aid on different student types but had not factored in academic programs. The other participant had attempted to find a predictive analytics tool on the market, but they were unsuccessful and built their tool using Microsoft Excel and Access.

Table 9

Participant responses to question 7

#	Responses
1	Participant A: …I use some predictive modeling right now. Participant B: You are already? Participant A: Yes. Participant B: Does it work anything like this? Participant A: Yes, a little bit. Only, we're not looking at specific programs. For example, you mentioned the five bars. I'm looking more at a total… I haven't built in academic makers.
2	I built [a tool] over several years. It's been built. It is between Excel and Access and not all people have the skills to use Excel or Access either one, but I couldn't find anything out there to be able to have a … I mean I understand they're all predictive tools that are available to us, but none to get it to a level that would impact us and be a tool that I could use. I just had to build it.

Interview Question 8

What benefits does this tool provide in gaining insight into financial aid resource allocation at your institution or college in general?

The group agreed that this question would be best answered by the Financial Aid Director who stated that the financial aid department at their institution would be able to serve their students. As stated in Table 10, the respondent was also the same participant who claimed there were no adequate tools on the market to assist with financial aid management.

Table 10

Participant responses to question 8

#	Responses
1	Mine is at the most elementary level. Be able to do what we need to do at our institution.

Interview Question 9

What pitfalls does this tool have for gaining insight into financial aid resource allocation at your institution or college in general?

***Probe Question* - What might you add or take away from this tool to overcome these pitfalls?**

This question spawned much concern around the delegation of money to low-income students and student debt management. The focus group members wanted to make sure loans were highlighted differently from other financial aid sources shown in the tool, as stated in Table 11. They believed loans impact student success, especially that of low-income students, since they result in student debt. The end user may be able to assist in managing student debt and provide students with other aid sources when necessary with a loan indicator and family income levels. Another concern was the misuse of this tool for providing unequal funding to certain types of students over others. The discussion of the F3A misuse led the group to the decision that the user should be shown a warning message which explains the intent of the tool and discourages abuse.

Table 11

Participant responses to question 9

#	Responses
1	For debt management.
2	Do you have student loans built in there anywhere?... Do you have family income in there?... I think the research shows, if not correct me, but that low-income students are risk adverse and don't want to take out loans.
3	I understand that, from a perspective that, if a student's family income is $20,000, they don't want to take out a $10,000 loan. That's 50% of their family income. That makes sense, but at the same time, if you had that as a variable that would help you know, "Okay, this one needs this much financial aid and I can give so much institutional money, but how much can I put as loan that they're likely to take and still graduate?"
4	Well, my dissertation was very similar to this. …I did come up with a tool that used financial aid to predict six-year graduation rate and my concern when I was working on that and also with this is a philosophical one which is if you use this tool to get to college coaching, you say you have somebody with particular demographic information and their graduation rate is going to be 50% and you say we want to bump that up. Okay, you intervene with them. But if you look at it you say I'm going to use this to adjust the level of financial aid, then

	you get into a situation which you have three students. Two of those students you can give $1000 reward to each and get a 70% graduation rate. The third one you can give $2000 each and you only have $2000. So, do you give that money to those two students that have a higher chance of graduating, or do you award it on the basis of using financial aid to level the playing field of access, which would be different from using of the variables. That was the concern when I put my tool out there and that's my concern with this tool, is that it can be misused in that way.

Interview Question 10

Do you think it may be appropriate to use this tool making financial aid decisions?

***Probe Question* - If so, how, and when in the process might it be applied?**

Table 12 shows that One respondent reiterated the importance of the tool for understanding the effects of academic programs on student success rather than financial aid. Another participant explained they would use this tool to optimize their fund delegation to ensure the maximum student success with the minimal amount of aid.

Table 12

Participant responses to question 10

#	Responses
1	From my perspective I would use it more on a program type of a basis. However, I can see where this would be extremely helpful on an individual basis as far as persistence is concerned. I think we all know that, and I have argued this for a long time, is that the money sometimes is not the factor in persistence and in graduation rates. If you want to increase those, it's not always that factor.

2 I would use this from my perspective in programs, in impacting enrollment as well as persistence in graduation rates, how to get the most bang for my buck.

Presentation and Discussion of Findings

The data was analyzed, and a concise list compiled from the participant's feedback. Each item was developed from the direct quotes of the participants without considering if one person mentioned the item or if it was agreed upon by the group. There were six main enhancements to the application which were as follows:

- A more customized application constructed using real-world student data.
- A more intuitive, easy to use, visually pleasing front-end.
- Custom student segments, or profiles, including but not limited to the following student demographics:
  - Dual Credit Status
  - Transfer Status
  - Program Type (i.e., certifications, associates, bachelor's, and graduate degrees)
  - Academic Major
  - Total Semester Credit Hours
  - Family Contribution Levels
- A system for refreshing underlying data that provides the end user with the most up-to-date information.
- The ability to toggle various aid types, especially student loans, to view how they impact student success.
- A warning message which explains the intent of the tool and discourages possible abuse.

After the data was transcribed, the information was reviewed multiple times to ensure that the researcher had a clear understanding of the participant's feedback. If the focus group member had revealed names, locations, or events they were replaced with pseudonyms in the data. Once the initial analysis was complete, the quoted statements made by the participants about improving the tool were sorted and developed into a concise list. Data reduction was used to eliminate or remove unnecessary information from the drafted list. Participants were provided with a copy of the list for comments and corrections. Upon receipt of the approval from the participants, a finalized list of enhancements for the F3A was derived.

This analysis revealed the need for additional elements to the F3A for widespread use of this application. The participants explained that potential users of this tool would likely be non-technical and would be more apt to use analytical software given a cleaner, more simplistic frontend interface. During the focus group, participants discussed the use of this tool beyond the needs of financial aid resource allocation and grant program impacts. Some of the respondents were interested in using this tool for enrollment management, debt management, and to better understand other factors which impact student achievement outside of financial aid. Overall, the primary concern was the lack of real-world data tested on this tool which left the participants skeptical about the student outcomes presented. However, most of the respondents claimed to see potential in the tool and stated they would participate in future focus groups which included actual data and other enhancements.

Summary of Chapter Five

Chapter 5 encompassed a description of the focus group demographics, data analysis, and a list of enhancements to the F3A derived from the responses of the participants. The focus group had eight participants of which the majority were leaders at the Kentucky Higher

Education Assistance Authority. The direct quotes from the respondents identified the tool usability, visualizations, and student profiles choices as areas for improvement. Chapter 6 discusses the findings in depth and recommends future studies for the promotion of analytics in financial aid resource allocation.

CHAPTER SIX

This Design Science Research (DSR) study addressed the issue of the lack of analytics used for strategic decision-making in state financial aid resource allocation. The research included the design, development, and evaluation of an analytical grants management application. The research question observed needed technological enhancements to the financial aid analytics application (F3A) which could help provide better insight to grant managers and financial aid policymakers. The DSR methodology requires the creation of an IT artifact and an evaluation of the artifact using quantitative or qualitative methods (Vaishnavi & Kuechler, 2008). This study used a focus group which an emphasis on gathering new data from a target population (Bryman & Bell, 2011; Kreuger & Casey, 2008). A focus group, consisting of a panel of state government employees who assist in aid program development, was held to determine the need for improvements to the application.

This chapter includes a discussion of the findings obtained from the focus group interview data shown in Chapter 5, a list of additions to the software application, and feedback for possible future implementation. The remaining sections in this chapter contain the limitations of the study, implications for practice, and recommendations for future research.

Findings and Conclusions

To better understand and discuss the results of the study, the research question is stated below:

> What technological enhancements should be made to the F3A for assisting higher education leaders in decision-making regarding financial aid resource allocation?

Findings

The following list of technological enhancements was derived from the focus group data to address the research question:

- A more customized application constructed using real-world student data.
- A more intuitive, easy to use, visually pleasing front-end.
- Custom student segments, or profiles, including but not limited to the following student demographics:
    - Dual Credit Status
    - Transfer Status
    - Program Type (i.e., certifications, associates, bachelor's, and graduate degrees)
    - Academic Major
    - Total Semester Credit Hours
    - Family Contribution Levels
- A system for refreshing underlying data that provides the end user with the most up-to-date information.
- The ability to toggle various aid types, particularly student loans, to view how they impact student success.
- A warning message which explains the intent of the tool and discourages possible abuse.

The previous literature states that the successful use of an analytics tool depends on various factors including:

- Handling of diverse data to assist varying departments in an organization;
- Continuous data collection which allows for real-time analysis (Bataweel, 2015);
- Ease of use;

- Data quality and data quality control (Mohanty, 2008);
- Involving relevant people in the solution selection process; and
- Customizing data and applications for various user types (DeVoe & Neal, 2005).

The focus group members had similar responses to describe the enhancements required for the widespread use of the F3A. The general recommendations for the software were consistent across the focus group members, but at times the implementation and more granular details differed.

Conclusions

Based on the findings the following conclusions are presented with the intention of promoting the use of analytics in financial aid delegation. These conclusions stem from the literature review and the focus group responses from the eight analysts and leaders who support the state of Kentucky's major financial aid programs.

Technological enhancement 1: Actual data. The first task on the list referred to the development of a more customized application developed using real-world student data. The tool necessitates the addition of this data to provide the end user with accurate insight into their specific population of interest. Although the participants were able to provide feedback about the look and feel of the tool, the use of synthetic data made it difficult for them to determine the overall usefulness of the software for real-world decision-making related to grant program management. This finding was consistent with previous research referencing the effectiveness of BI tools which indicate the challenges and possible causes of failure in BI solutions lacking data quality, data validity, and information accessibility (DeVoe & Neal, 2005; Mohanty, 2008). The focus group evaluation revealed the tool could not assist the intended audience in making a valid aid allocation decision because the data was not relevant to their organizations' financial aid programs or students.

Technological enhancement 2: User interface. The next improvement documented was that of a more intuitive, easy to use, and visually pleasing frontend interface. The focus group participants had a firm grasp on the concepts of predictive modeling but expressed concerned for non-technical users. They pointed out that many financial aid directors, and other higher education administrators interested in using decision support systems, typically had non-technical backgrounds. Thus, the participants recommended a cleaner, more self-explanatory, user interface. Mohanty (2008) stated that one of the critical factors which influence the practical use of Business Intelligence (BI) was the ease of using the tool while Bataweel (2015) highlights the need for data visualization to understand and utilize information successfully. DeVoe and Neal (2005) suggest training nontechnical users or hiring experts and even customizing the application down the unit, departments, or employee level to promote use of analytical tools. They also discussed involving relevant people in the BI solution selection process (DeVoe & Neal, 2005). Given the feedback from the focus group and past research findings, it is evident that conducting additional focus groups with non-technical financial managers and assistants may provide better insight into the specific changes needed for the F3A’s frontend. Further data collection may also assist the developers in better understanding the types of visualizations the user would require for decision making.

Specific advanced statistics shown to the user, such as the performance of the predictive model, could be excluded from the application. If the required data was available, the tool could be customized, and the predictive model refined in advance. However, when new population data is present, as would be the case after each semester of student attendance, then the underlying model would likely have to be adjusted or replaced. Based on the responses from the research participants, this tool falls under operational BI in which dashboards and reports provide the end user with the information for quick, easy decision-making (Bataweel, 2015). It may also be helpful to reduce the various summary statistics to common types (i.e., average, total, minimum, and maximum) along with their definitions.

Technological enhancement 3: Student profiles. The need for more custom student segments, or profiles, was a recurring response from the respondents. The focus group members were very interested in understanding the magnitude at which different financial aid types impacted certain groups of students. Past research has shown that student's income levels and academic history modifications can lower or heighten the level of impact monetary aid has on college success (DesJardin & McCall, 2010; Long, 2009; Sjoquist & Winters, 2012). Feedback on how to portray student profiles to the end user was mixed. While some participants wanted a few predefined student segments, others sought the freedom of being able to create their desired profile combinations. Allowing the user to create student profiles is a possibility, but the number of available category options (i.e., age, income level, and GPA) should be limited to achieve a cleaner visual appearance. Restricting the number of selections also helps to reduce the risk of exposing information when filtering on smaller populations.

Technological enhancement 4: Data refresh. Bataweel (2015), Mohanty (2008), DeVoe & Neal (2005) emphasize the importance of updating information for the successful use of analytical tools. Participants agreed that there was a necessity for the addition of a component to refresh underlying data to provide the end user with more up-to-date information. However, the focus group panel communicated diverse views for executing this task. One of the respondents suggested connecting the application to the end user’s database and allowing for a scheduled refresh of the data. Contrarily, another participant explained the need for a manual data refresh process due to potential issues with data quality which could negatively impact the application, particularly the accuracy of the underlying statistical model. In the event an automated process was implemented, it would be beneficial to inform the users of any data requirements to properly operate the tool. Mohanty (2008) stated that data quality was necessary to assist in the effectiveness of BI solutions required for timely, accurate decision-making.

Technological enhancement 5: Financial aid types. The literature review covered the effects of different financial aid types on student achievement. Past research shows that different aid programs (i.e., grants, loans, and scholarships) have varying magnitudes of impact on student success measures (Goodman, 2008; Long, 2009). Some of the focus group participants wanted the ability to toggle various aid types, particularly student loans, so that the impact could be observed. Conversely, others wanted to see the total monetary effects on these achievement measures. The ability to observe the difference in program effects was important for respondents who make decisions on discontinuing or increasing funding for those programs.

Technological enhancement 6: Technology misuse. Lastly, the topic of misusing the F3A was raised for debate and resulted in the task of appending a warning message to explain the intent of the tool and discourages possible abuse. Some participants were worried that findings presented by tool would influence decision-makers to favor supporting students with higher probabilities of success and cause indirect discrimination towards other potential college-goers. A few of the respondents agreed but discussed the possibility of using this tool to detect and place at-risk students in academic assistance programs which could increase their likelihood of success. The goal of this study is to provide grant managers with an analytical solution to assists them in improving the delegation of aid in the hopes that more potential students who lack funding are provided the opportunity to obtain an education. Also, the tool is not 100% accurate and excludes many external factors (i.e., social life, past trauma) which could affect student outcomes. The Forum Guide to Data Ethics states "…people may be tempted to engage in unethical behavior—to knowingly manipulate or misrepresent statistics to make a point; misuse data for personal gain; or convince themselves that privacy and confidentiality requirements don't need to be observed" (National Forum on Education Statistics, 2010, p. 2). Procedures should be established to prevent these kinds of issues from occurring. Though the end user's intentions cannot be controlled, the purpose and capabilities of this tool should be communicated to avoid misuse.

The F3A has considerable potential for use by Kentucky state government administrators who support financial aid program development. Obtaining data from a state agency, which includes similar information from various colleges and universities in the region, could yield a more scalable and easily implemented solution due to the uniform inputs. However, for agencies with unique student demographics and differing data, the underlying predictive model would need to be customized to yield more accurate, valuable feedback.

Limitations of the Study

Since the purpose of this study was to gather details about possible future improvements to the F3A, the results of the predictive model were not an aspect and not addressed in the research questions. However, respondents were interested in the prediction outcomes which could not be provided due to the lack of actual student data. Since the outputs were solely present for a visual presentation rather than to produce an accurate analysis, participant feedback regarding the output statistics was not summarized.

Other limitations that occurred within this research study included the unavailability of actual student data, the restricted scope of the project, and the diminutive population size of administrators who support state aid programs. Data was requested from the Kentucky Center for Education and Workforce Statistics (KCEWS) prior to the development of the F3A but was not received in sufficient time to modify the software and prepare the focus group interview. This information could have been used to model the effects of the College Access Program (CAP), Kentucky Tuition Grant (KTG), and the Kentucky Educational Excellence Scholarship (KEES) on the graduation rates of Kentucky college-goers. Furthermore, the project was limited in scope, time, and budget. The time allocated for the software design and development was 35 days, which left the F3A lacking in multiple areas. However, the limited timeframe yielded a simplistic application design making it easier to provide future customization for various users. Also, there was not a project budget for the development of the application. The lack of budget did not pose issues related to functionality since there are many free solutions for software development but problems could arise in the future with maintenance and improvement costs.

Another limitation to the scope of the project involved inadequate human resources; restricted man-hours contributed from a single person placed constraints on the design and creation of the tool. Lastly, the small size of the population resulted in holding one focus group. Thus, the feedback was biased towards the opinions of the members of the focus group panel.

Implications for Practice

The problem addressed in this study is the lack of analytics used for strategic decision-making in higher education (Ferreira & Andrade, 2016; Macfadyen et al., 2014; Roscorla, 2015), particularly in state financial aid resource allocation (Goldstein, 2005; Pomeroy, 2014; Thanh & Haddawy, 2007). The design and creation of the F3A provides a simple decision support application for financial aid program initiators which can be tailored to different student populations and customized to fit the needs of the end user. The design could be used as a framework for the development of other grants management tools as it provides the necessary basics requirements needed to understand financial impacts on student outcomes. These components include the ability to,

- Import and view the raw data;
- Analyze the summary statistics of each input;
- Select the financial aid programs as well as other desired variables;
- Assess the accuracy of the underlying predictive model; and
- View the likelihood of student outcomes through the selection of various student types.

The evaluation of the software application also provides software developers with areas in which to focus when creating solutions to support financial aid managers in gaining insight into their programs. Pomeroy (2014) explained there was not widespread use of analytics tools in higher education because administrators were not able or willing to invest the needed time or money. Developing a more straightforward system and cleaner interface could lower the time required to learn this tool which could increase its use by these administrators. Although the F3A was created using R programming language, the framework extracted from the design of the F3A could be recreated with little to no cost using Java or Python.

Implications of Study and Recommendations for Future Research

The results of this study suggested six main improvements to the F3A derived from the focus group feedback. While the tool had potential use for grant managers in understanding aid impacts, there were limitations in this study which restricted the development and usefulness of the tool. This section of the study includes recommendations for future research determined from the design, creation, and evaluation of the F3A.

Future Research: Use of Real-World Data

Participant feedback emphasized the importance of using actual student data to assist financial aid administrators in more efficient delegation of funds and improved program development. Information regarding student demographics, financial aid received, and academic success measures may be retrieved from the agency of interest as well as state and federal entities that collect this information. Typically, schools have an institutional research office or IT department that can provide this data. A Public Information Act request is required for a school or agency that receives government support or an institution's formal request procedure is commonly undertaken to solicit this type data. Acquiring the information in advance and using it to develop the tool could yield a greater understanding of aid impacts on student outcomes.

Future Research: Software Development Framework

The initial design and development of the F3A provided the user with each phase of the CRISP-DM framework. This design resulted in a complicated user interface with too many additional phases for non-technical users (i.e., data cleansing, model selection). It is beneficial to use a software development methodology, such as Agile when enhancing the current tool. The agile method is a sequential iterative approach which allows planners, designers, and testers to easily implement changes to a product over time. The goal of this approach is to create a working product for the customer in a short amount of time. This method is advantageous because it provides the opportunity for fast delivery and continuous improvement (Bowes, 2014). Researching different techniques for software development could contribute to understanding the requirements for creating financial aid related decision support systems.

Future Research: Choice of Programming Languages

This study used the R programming language which is widely used for big data analytics and statistical modeling projects. However, there are other free languages with useful data mining packages such as Python and Java. Python is another language used for data science projects and it has various analytical packages including *Pandas*, *NumPy*, *SciPy*, and *SciKi-Learn*. Furthermore, Java has robust data mining and manipulation packages such as *WEKA* which may provide additional functionality not offered in R.

Future Research: Testing of Various Statistical Models

This study used the Naïve Bayes model to understand financial aid impacts on student success, which assumes that all variables are equal and independently distributed. This assumption rarely occurs and could result in low model performance, particularly with social science problems. For example, past research shows that low-income strongly correlates with the poor academic performance of college students (Delaney, 2011; Hossler, 2002). The independence assumption is one reason it is not a good idea to assume the two variables are independent. More complex Bayesian Network Models may be used to model these dependencies, but are difficult to implement using R. However, there are many other model types (i.e., decision trees, logistic regression, and neural networks), variable interactions, and ensemble choices which could be used for the underlying analysis of student outcomes in the F3A.

Future Research: Software Solutions for Enrollment Management

A few of the focus group respondents were interested in the F3A beyond grants management. There was a desire to view the attributes of potential college-goers, or applicants, and their likelihood of student success for enrollment management. Even though the use cases of the tool are very similar, the amount of data provided by applicants is typically not as rich as the information accrued by students who are already enrolled. For instance, an alumnus or previous enrollee has academic history on file which could not be provided by an applicant thus making it more difficult to determine the outcome of the student. as this could prove challenging as far as data availability, the focus group members conveyed interest in an enrollment management software application. Therefore, a future DSR study focusing on a decision support system for enrolling students is a potential research opportunity.

Future Research: More Focus Groups for Evaluation

Since the audience for this study consisted of state financial aid program supporters, the population size was quite small. If this research were extended to include more program types, such as institutional, federal, or private, then the population size would increase tremendously. This expansion could provide researchers with the opportunity to hold more focus group interviews yielding richer data that could be applied to solutions for varying audiences. Further research with focus groups could also provide the researcher with the ability to develop theories due to data saturation.

Future Research: Collaborating with Other Developers

One person completed the design and development of the IT artifact, but more developers may have improved the final product. A team of three resources including a front-end, back-end, and predictive modeling developer may properly implement the enhancements to the F3A. Therefore, the continued development of the F3A with additional qualified personnel resources is recommended for future studies.

Conclusion

The goal of the research study was to gather a list of technological improvements for the F3A to help grants managers gain insight as to how aid programs impact student success and optimize the distribution of financial aid monies. The evaluation of the F3A, was achieved by conducting a focus group, which demonstrated the tool's potential for widespread use after making improvements to the user interface, using appropriate underlying data, including additional student factors, allowing for varying aid types, and implementing a warning message to prevent misuse of the application.

In summary, this study provided a basic framework for designing, developing, and enhancing a grants management tool for the improved understanding of the impacts of aid programs on, and allocation to, students and potential college-goers. The findings also provided areas of focus for developers to consider when creating analytical financial aid delegation software to encourage use by financial aid policymakers and managers. The focus group participants stated that aid management tool developers should keep in mind that many potential end users are nontechnical. A major takeaway from this study was that the F3A should be customized using a real-world dataset relating to the potential client or business to aide in understanding if the tool could assist in decision-making.

REFERENCES

Abbott, D. (2014). Applied predictive analytics: *Principles and techniques for the professional data analyst*. Indianapolis, IN: John Wiley & Sons.

Amazon Web Services, Inc. [AWS]. (2015). *FERPA compliance on AWS*. Retrieved from http://d0.awsstatic.com/whitepapers/AWS_FERPA_Whitepaper.pdf

Baer, L., & Duin, A. H. (2014). Retain your students: The analytics, policies and politics of reinvention strategies. *Planning for Higher Education*, *42(3), 30-41*. Retrieved from ProQuest database.

Baer, L. L., & Norris, D. M. (2016). A call to action for student success analytics. *Planning for Higher Education*, 44(4), 1-10.

Banerjee, A., Bandyopadhyay, T., & Acharya, P. (2014). Data analytics: Hyped up aspirations or true potential. *Vikalpa: The Journal For Decision Makers*, *39(1), 1-11*.

Bataweel, D. S. (2015). *Business intelligence: Evolution and future trends* (Doctoral dissertation, North Carolina Agricultural and Technical State University). Retrieved from ProQuest Dissertations and Theses database. (Order No. 1590701)

Bowes, J. (2014). *Agile vs Waterfall: Comparing project management methods*. Retrieved from https://manifesto.co.uk/agile-vs-waterfall-comparing-project-management-methodologies/

Bureau of Labor Statistics. (2016). *Quarterly census of employment and wages* [Data file]. Retrieved from https://data.bls.gov/cew

Bureau of Labor Statistics. (2017). *Occupational Employment Statistics* [Data file]. Retrieved from https://data.bls.gov/oes

Boshardt, D., Lichtenstein, L., Palumbo, G., & Zaporowski, M. P. (2010). Optimization techniques for college managers. *Journal of Student Financial Aid*, *40*(3), 39-57.

Bryman, A., & Bell, E. (2011). *Business Research Methods (3rd ed.).* Oxford, NY: Oxford University Press.

Bromley, E., Mikesell, L., Jones, F., & Khodyakov, D. (2015). From subject to participant: ethics and the evolving role of community in health research. *American Journal of Public Health, 105(5), 900-908.* doi:10.2105/AJPH.2014.302403

Business-Higher Education Forum. (2014). *Data Science emerges as essential tool for decision making and innovation across industry sectors*. Retrieved from ERIC database. (ERIC No. ED559640)

Carnevale, A. P., Smith, N., & Strohl, J. (2014). *Recovery: Job growth and education requirements through 2020*. Retrieved from the Georgetown University Center on Education and the Workforce website: https://cew.georgetown.edu/wp-content/uploads/2014/11/Recovery2020.FR_.Web_.pdf

Cellini, S. R., & Kee, J. E. (2010).Cost-effectiveness and cost-benefit analysis. In Wholey, Hatrey H. P., & Newcomer K. E., *Handbook of Practical Program Evaluation (pp. 493-530).* San Francisco, CA: Jossey-Bass.

Center for Analysis of Postsecondary Education and Employment. (2015). *The Federal Work-Study Program: Impacts on Academic Outcomes and Employment.* Retrieved from http://files.eric.ed.gov/fulltext/ED555433.pdf

Cevizci, A. (2016). A comparison of optimal portfolio performances of three optimization methods. *International Journal of Commerce and Finance*, *2*(1), 137-146.

Chan, P., Fan, W., Prodromidis, A., & Stolfo, S. (1999). Distributed data mining in credit card fraud detection. *IEEE Intelligent Systems, 14*(6), 67-74. doi:10.1109/5254.809570

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: A step-by-step data mining guide*. Retrieved from https://www.the-modeling-agency.com/crisp-dm.pdf

Chintagunta, P., Hanssens, D. M., & Hauser, J. R. (2016). Marketing and data science: Together the future is ours. *Gfk-Marketing Intelligence Review, 8*(2), 18-23. doi:10.1515/gfkmir-2016-0011

Chuah, M., & Wong, K. (2011). A review of business intelligence and its maturity models. *African Journal of Business Management, 5*(9), 3424-3428. doi:10.5897/AJBM10.1564

Citizens Financial Group Incorporated (2016). *Parents find student loan financial aid process more stressful than applying for college.* Retrieved from http://www.businesswire.com/news/home/20161013005820/en/CORRECTING-REPLACING-GRAPHIC-Students-Parents-Find-Student

Cleveland, W. S. (2001). Data Science: An action plan for expanding the technical areas of the field of statistics. *International Statistical Review, 69*(1), 21-26. doi:10.1111/j.1751-5823.2001.tb00477.x

Coenen, F. (2004). Data mining: Past, present and future. *The Knowledge Engineering Review,* 00(0), 1-24. Retrieved from http://cgi.csc.liv.ac.uk/~frans/PostScriptFiles/kerDataMining_2010-8-19.pdf

Creech, J. D. (1998). *State-funded merit scholarship programs: Why are they popular? Can they increase participation in higher education*. Atlanta, GA: Southern Regional Education

Board. Retrieved from Institute of Education Sciences website: http://files.eric.ed.gov/fulltext/ED426653.pdf

Creswell, J. W., & Miller, D. (2000). Determining validity in qualitative inquiry. *Theory into Practice, 39*(3), 124-130.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.).* Thousand Oaks, California: SAGE.

Darwiche, A. (2009). *Modeling and reasoning with Bayesian networks*. New York, NY: Cambridge University Press.

Delaney, J. A. (2011, February 10). State merit-based aid and enrolling in graduate study: Evidence from the scholarship. *Journal of Student Financial Aid*, *41*(2), 6-21. Retrieved from the National Association of Student Financial Aid Administrators website: http://publications.nasfaa.org/

Department of Health, Education, & Welfare. (1974). *The Belmont Report Ethical Principles and Guidelines for the Protection of Human Subjects of Research, 2*. Retrieved from the National Institutes of Health website: https://videocast.nih.gov/pdf/ohrp_appendix_belmont_report_vol_2.pdf

DesJardin, S. L., & McCall, B. P. (2010). Simulating the effects of financial aid packages on college student stop out, reenrollment spells, and graduation chances. *Review of Higher Education, 33*(4), 513–541.

Despa, S. (n.d.). What is survival analysis. *Statnews*. Retrieved from the Cornell University Statistical Consulting Unit website: https://www.cscu.cornell.edu/news/statnews/stnews78.pdf

DeVoe, L., & Neal, K. (2005). When business intelligence equals business value. *Business Intelligence Journal*, *10*, 57-63.

Dziuban, C., Moskal, P., Cavanagh, T., & Watts, A. (2012). Analytics that inform the university: Using data you already have. *Journal of Asynchronous Learning Networks, 16*(3), 21-38.

Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of Management Review, 32*(4), 1155-1179.

Eichinski, P., & Roe, P. (2016). *Datatrack: An R package for managing data in a multi-stage experimental workflow data versioning and provenance considerations in interactive scripting.* Presented at the 2016 IEEE 12th International Conference on e-Science (e-Science), Baltimore, MD. doi:10.1109/eScience.2016.7870895

EMC Education Services. (2015). *Data science & big data analytics: Discovering, analyzing, visualizing and presenting data*. doi:10.1002/9781119183686

Evans, J. R. (2015). Modern analytics and the future of quality and performance excellence. *Quality Management Journal, 22*(4), 6-17.

Faraway, J. F. (2005). *Linear models with R*. Boca Raton, FL: Chapman & Hill/CRC.

Fayyad, U., Piatetsky-Shapiro, H., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM, 39*(11), 27 - 34.

Ferreira, S., & Andrade, A. (2016). Academic analytics: Anatomy of an exploratory essay. *Education and Information Technologies*, *21(1), 229-243.* doi:10.1007/s10639-014-9317-9

The fundamentals of financial aid award letters. (2001). *Hudson Valley Business Journal*, *12*(2), 15.

Gan, E. (2016). *Using the R package shiny to create web applications that facilitate quantitative gene expression analysis.* Available from ProQuest Dissertations & Theses Global. (Order No. 10138582).

Gartner, Inc. (2012). Gartner says Big Data creates big jobs: 4.4 million IT jobs globally to support Big Data by 2015. *Gartner Newsroom*. Retrieved from http://www.gartner.com/newsroom/id/2207915

Gessner, G., & Scott, R. A. (2009). Using business intelligence tools to help manage costs and effectiveness of business-to-business inside-sales programs. *Information Systems Management*, *26*(2), 199-208. doi:10.1080/10580530902797623

Gibbs, G. R. (2007). Analyzing qualitative data. In U. Flick (Ed.), *The Sage qualitative research kit*. Thousand Oaks, CA: Sage.

Gilli, M., & Schumann, E. (2012). Heuristic optimisation in financial modelling. *Annals of Operations Research*, *193*(1), 129-158. doi:10.1007/s10479-011-0862-y

Goldstein, P. (2005). *Academic analytics: The uses of management information and technology in higher education.* Retrieved from the EDUCASE Center for Applied Research website: https://net.educause.edu/ir/library/pdf/ecar_so/ers/ERS0508/ekf0508.pdf

Goodman, J. (2008). Who merits financial aid: Massachusetts' Adams Scholarship. *Journal of Public Economics*, *92*(10), 2121-2131. Retrieved from the John F. Kennedy School of Government website: http://hks.harvard.edu/fs/jgoodma1/papers/meritaid

Groen, J. A. (2011). Do merit-aid programs help states build skilled workforces. *Change: The Magazine of Higher Learning*, *43*(6), 33-37. doi:10.1080/00091383.2011.618080

Hackathorn, R., & Margolis, T. (2016). *Immersive analytics: Building virtual data worlds for collaborative decision support*. Presented at the 2016 Workshop on Immersive Analytics (IA), Greenville, SC. doi:10.1109/IMMERSIVE.2016.7932382

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques*. Waltham, MA: Morgan Kaufmann.

Harper, J. (2014). *Distinguishing analytics, business intelligence, data science*. Retrieved from http://www.dataversity.net/distinguishing-analytics-business-intelligence-data-science/

Hastie, T., & Tibshirani, R. (2009). *Cross-validation and bootstrap*. Retrieved from https://lagunita.stanford.edu/c4x/HumanitiesScience/StatLearning/asset/cv_boot.pdf

He, X. J. (2014). Business intelligence and big data analytics: An overview. *Communications of the IIMA*, 14(3), 1-11. Retrieved from ProQuest database.

Hevner, A., & Chatterjee, S. (2010). Design science research in information systems. *In Design research in information systems: Theory and practice, 1*, 9-22. doi:10.1007/978-1-4419-5653-8

Hiltbrand, T. (2013). Behavior-Based Budget Management Using Predictive Analytics. *Business Intelligence Journal*, *18*(1), 25-32.

Holmes, D. E., & Jain, L. C. (2011). *Data mining: foundations and intelligent paradigms: Volume 1: Clustering, association, and classification.* Berlin: Germany. Springer.

Horn, C., Santelices, M. V., & Avendaño, X. C. (2014). Modeling the impacts of national and institutional financial aid opportunities on persistence at an elite Chilean university. *Higher Education*, 68(3), 471-488. doi:10.1007/s10734-014-9723-3

Hossler, D. (2002, December 14). The role of financial aid in enrollment management. *New Directions for Student Services*, *2000*(89), 77-90. doi:10.1002/ss.8906

Hsinchun, C., Chiang, R. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36*(4), 1165-1188.*

Information Resources Management Association. (2012). *Machine learning: Concepts, methodologies, tools and applications*. Hershey, PA: Information Science Reference.

Kalgotra, P., & Sharda, R. (2016). *Progression analysis of signals: Extending CRISP-DM to stream analytics*. Presented at the 2016 IEEE International Conference on Big Data (Big Data), Washington, DC. doi:10.1109/BigData.2016.7840937

Kavur, J. (2009). SAS isn't fazed by the competition. *ComputerWorld Canada, 25*(18), 6. Retrieved from ProQuest database. (Order No. 610971294)

Kentucky Higher Education Assistance Authority. (2010). *Kentucky Student Aid Aggregate Summary, 1990-91 to 2008-09* [Table]. Retrieved from https://www.kheaa.com/

Kentucky Higher Education Assistance Authority. (2017). *About us*. Retrieved from https://www.kheaa.com/website/kheaa/aboutkheaa?main=0

Kentucky Higher Education Assistance Authority. (2015). *Service and commitment: 2015 annual report*. Retrieved from https://www.kheaa.com/pdf/annual_2015.pdf

Kramer, F., Müller, H., & Turowski, K. (2014). *Acceleration of single inserts for columnar databases: An experiment on data import performance using SAP HANA.* Paper presented at the Tenth International Conference on Signal-Image Technology and Internet-Based Systems, Marrakech, pp. 672-676. doi:10.1109/SITIS.2014.45

Kreuger, R. A., & Casey, M. A. (2008). *Focus groups: A practical guide to applied research.* Thousand Oaks, CA: Sage.

Deposit of moneys 154A K.R.S. § 130 (1988 & Suppl. 2005).

Eligibility for educational excellence scholarship and supplemental awards 164 K.R.S. § 7881 (1998 & Suppl. 2015).

Larose, D. T. (2015). *Data mining methods and models*. Hoboken, NJ: Wiley-Interscience.

Latumakulita, L. A., Purnama, F., Usagawa, T., Paturusi, S., & Prima, D. A. (2016). *Indonesia scholarship selection framework using fuzzy inferences system approach: Case study: "Bidik Misi" scholarship selection*. Paper presented at the 2016 International Conference on Information & Communication Technology and Systems (ICTS), Surabaya. doi:10.1109/ICTS.2016.7910282

Leibfritz, F., & Maruhn, J. H. (2009). A successive SDP-NSDP approach to a robust optimization problem in finance. *Computational Optimization & Applications*, 44(3), 443-466. doi:10.1007/s10589-007-9163-4

Lofland, J., Snow, D. A., Anderson, L., & Lofland, L. H. (2006). *Analyzing social settings: A guide to qualitative observation and analysis (4th ed.).*

Long, B. T. (2009). Making college affordable by improving aid policy. *Issues in Science and Technology*. Retrieved from http://www.issues.org/26.4/long.html

Lublinsky, B., Smith, K. T., & Yakubovich, A. (2013). *Professional hadoop solutions*. Indianapolis, IN: John Wiley and Sons.

Ma, X., Yue, F., Zhang, Q., Cui, Y., & Qu, J. (2009). *Evaluation analysis of university student scholarship*. Paper presented at the 2009 International Conference on Test and Measurement, Hong Kong. doi:10.1109/ICTM.2009.5413089

Macfadyen, L. P., Dawson, S., Pardo, A., & Gasevic, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment, 9*. Retrieved from ProQuest database.

Mariscal, G., Marban, O., & Fernandez, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review, 25*(2), 137–166. doi:10.1017/S026988891000003

Marren, P. (2004). The father of business intelligence. *The Journal of Business Strategy, 25*(6), 5-7. Retrieved from ProQuest database.

McAfee, A. (2012). Big data: The management revolution. *Harvard Business Review*. Retrieved from https://hbr.org/2012/10/big-data-the-management-revolution

Minkara, O. (2012). The middle ground. *Marketing Research, 24*(4), 22-29.

Mohanty, S. (2008). Measuring the value of intelligence in business intelligence. *DM Review, 18*(12), 20. Retrieved from ProQuest database.

Nair, G. S., & Riege, A. M. (1995). *Convergent interviewing to refine the research problem of a PhD thesis.* Paper presented at the Marketing Educators' and Researchers' International Conference, Gold Coast.

Narozhnaya, Y. (2015). *Exploring the relationship between institutional revenue patterns, student financial aid packages, and degree completion from a multilevel perspective* (Doctoral dissertation, University of Maryland, College Park). Retrieved from ProQuest Dissertations & Theses database. (Order No. 3711765)

National Center for Education Statistics. (2015). *Public school enrollment in grades 9 through 12, by region, state, and jurisdiction: Selected years, fall 1990 through fall 2024* [Table]. Retrieved from http://nces.ed.gov/

National Center for Education Statistics. (2016). *2016-2017 survey materials: Glossary results*. Retrieved from https://surveys.nces.ed.gov/ipeds/VisGlossaryAll.aspx

National Commission on Higher Education Attainment. (2013). *An Open Letter to College and University Leaders: College Completion Must Be Our Priority.* Retrieved from www.acenet.edu/ news-room/Documents/An-Open-Letter-to-College-and-UniversityLeaders.pdf

National Forum on Education Statistics. (2010). *Forum guide to data ethics (801).* Retrieved from the Washington, DC: Government Printing Office website: https://nces.ed.gov/pubs2010/2010801.pdf

Nirmala, C. R., Roopa, G. M., & Kumar, K. R. N. (2015). *Twitter data analysis for unemployment crisis.* Paper presented at the 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT).

Nisbet, R., Elder, J., & Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. San Diego: Academic Press.

Ohri, A. (2016). *Tools and techniques for data science*. ITS Engineering College, India. Retrieved from https://www.slideshare.net/ajayohri/tools-and-techniques-for-data-science

Ogundimu, E., Altman, D. G., & Collins, G. (2016). Adequate sample size for developing prediction models is not simply related to events per variable. *Journal of Clinical Epidemiology*, *76*, 175-182. doi:10.1016/j.jclinepi.2016.02.031

O'Leary, B. (2017). Big data, big decisions. *Electrical Apparatus*, *70*(7), 34-35.

Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). *The Design Science research process: A model for producing and presenting Information Systems Research.* Retrieved from the Pennsylvania State University College of Information Sciences and Technology website: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.469.2936&rep=rep1&type=pdf

Peterson's CollegeQuest. (2014). *What are the different types of financial aid for college students*. Retrieved from http://www.collegequest.com/different-types-of-financial-aid-for-college.aspx

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, *16*(3), 9-20.

Piñeiro, G., Perelman, S., Guerschman, J., & Paruelo, J. (2008). How to evaluate models: Observed vs. predicted or predicted vs. observed. *Ecological Modelling, 216*, 316-322. doi:10.1016/j.ecolmodel.2008.05.006.

Pomeroy, W. L. (2014). *Academic analytics in higher education: Barriers to adoption* (Doctoral dissertation, Walden University). Retrieved from ProQuest Dissertations & Theses database. (Order No. 3666435)

Purba, R. A., & Sembiring, J. (2016). *Selection of scholarship recipients by using Promethee method in Polytechnic Unggul LP3M Medan.* Paper presented at the 2016 International Seminar on Application for Technology of Information and Communication (ISemantic), Semarang. doi:10.1109/ISEMANTIC.2016.7873815

Rabiee, F. (2004). Focus-group interview and data analysis. *Proceedings of the Nutrition Society, 63*, 655-660.

Ragland, S. E. (2016). *The effect of state financial aid policies on college completion* (Doctoral dissertation, Walden University). Retrieved from ProQuest Dissertations & Theses database. (Order No. 10011627).

Riege, A. M. (2003). Validity and reliability tests in case study research: A literature review with "hands-on" applications for each research phase. *Qualitative Market Research, 6*(2), 75-86. Retrieved from ProQuest database.

Roscorla, T. (2015, Nov 02). The top 10 higher ed IT issues of 2016. *TCA Regional News*. Retrieved from ProQuest database.

Rubin, H. J., & Rubin, I. S. (2012). *Qualitative interviewing: The art of hearing data (3rd ed.).* Thousand Oaks, CA: Sage.

Sagiroglu, S., & Sinanc, D. (2013) *Big data: A review*. Paper presented at the Collaboration Technologies and Systems International Conference, San Diego, CA. doi:10.1109/CTS.2013.6567202

Saupe, J. L. (1990). *Functions of institutional research*. Retrieved from the Association of Institutional Research website: http://www.airweb.org/EducationAndEvents/Publications/Pages/FunctionsofIR.aspx

Selingo, J. (2013). How a little data can solve one of higher education's biggest problems. The *Chronicle of Higher Education*, Retrieved from ProQuest database.

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing*, *5*(4), 13-22. Retrieved from https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information, 22*, 63–75.

Sjoquist, D. L., & Winters, J. V. (2012). *State merit-based financial aid programs and college attainment.* Retrieved from http://ftp.iza.org/

Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. *International Journal of Computer Applications*, *17*(8), 43-48.

Spalding, A. (2014). New commitment to need-Based financial aid critical in context of rising tuition. *Kentucky Center for Economic Policy*.

Sparks, R. J. (2011). A value-added model to measure higher education returns on government investment. *Contemporary Issues in Education Research*, *4*(2), 15-22. doi:10.19030/cier.v4i2.4078

Teffeteller, S., & Kish, T. (2012). Mining internal data to reduce clinical costs. HFM (Healthcare Financial Management), *66*(12), 80-85.

Thanh, L. V., & Haddawy, P. (2007). *Deriving financial aid optimization models from admissions data*. Paper present at the Annual Frontiers in Education Conference in Milwaukee, WI. doi:10.1109/fie.2007.4417959

Toby, J. (2010). How scholarships morphed into financial aid. *Academic Questions, 23*(3), 298-310. doi:10.1007/s12129-010-9174-y

University at Albany (n.d.*). Institutional Review Board electronic data management policy*. Retrieved from http://www.albany.edu/orrc/assets/Institutional_Review_Board_Data_Management_Policy_v_1_0.pdf

University of Hawaii Community Colleges (2017). *Paying for college: What is financial aid*. Retrieved from http://uhcc.hawaii.edu/financial/

Vaishnavi, V., & Kuechler, B. (2008). *Design science research in information systems*. Retrieved from http://desrist.org/

Venter, J. (2009). What is BI. *Accountancy SA*, 36-37. Retrieved from ProQuest database.

Vollmer, S., & Howard, G. (2010). Statistical power, the Belmont Report, and the ethics of clinical trials. *Science and Engineering Ethics*, *16*(4), 675-691. doi:https://doi.org/10.1007/s11948-010-9244-0

Wei, W., Han, J., Kong, J., & Xia, H. (2016). *Prediction of the scholarship using comprehensive development.* Paper presented at the 4th International Conference on Enterprise Systems (ES), Melbourne, VIC. doi:10.1109/ES.2016.30

Wills, M. J. (2014). Decisions through data: Analytics in healthcare. *Journal of Healthcare Management*, *59*(4), 254-262.

Witten, I. H., Eibe, F., & Hall, M. A. (2011*). Data mining: Practical machine learning tools and techniques (3rd Ed.).* Morgan Kaufmann Publishers.

Ye, H. J., Zhan, D. C., Li, X., Huang, C., & Jiang, Y. (2016). *College student scholarships and subsidies granting: A multi-modal multi-label approach.* Paper presented at the 2016

IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain. doi:10.1109/ICDM.2016.0067

Yin, R. K. (1994). *Case study research: Design and methods, applied social research methods (5th Ed.)*. Newbury Park, CA: Sage.

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA: Sage.

APPENDIX A: FINANCIAL AID ANALYTICS APPLICATION PROGRAM CODE

```
## Chantel Perry
## October 6, 2017
## The purpose of this program is to provide minimum viable product for a financial aid
optimization
## tool

# Import needed libraries
library(shiny)
library(e1071)
library(taRifx)
library(ROCR)
library(pROC)
library(Hmisc)
library(corrplot)

################################################################################
#########################
# Beginning of the user interface code

ui <- fluidPage(
  # Set tabs for each dashboard
  tabsetPanel(
#----------------------------------------------------------------------------------------------------#
#--------------------------------------------DATA IMPORT TAB-------------------------------------#
#----------------------------------------------------------------------------------------------------#
  tabPanel(
    # Dashboard Title
    titlePanel("Upload"),

    # Add a sidebar with the following options
    sidebarLayout(
      sidebarPanel(

        # Add file import option
        fileInput("file1", "Choose CSV File",
                  multiple = TRUE,
                  accept = c("text/csv",
                             "text/comma-separated-values,text/plain",
                             ".csv")),

        # Add a horizontal line
        tags$hr(),

        # Allow user to specify whether CSV file has a header
```

```
      checkboxInput("header", "Header", TRUE),

      # Allow user to choose file from comma, semicolon, or tab delimited input data
      radioButtons("sep", "Separator",
                   choices = c(Comma = ",",
                               Semicolon = ";",
                               Tab = "\t"),
                   selected = ","),

      # Allow user to specify format of string fields
      radioButtons("quote", "Quote",
                   choices = c(None = "",
                               "Double Quote" = '"',
                               "Single Quote" = "'"),
                   selected = '"'),

      # Add a horizontal line
      tags$hr(),

      # Upon successful data import allow the user to see either the head of the data or whole
      # dataset
      radioButtons("disp", "Display",
                   choices = c(Head = "head",
                               All = "all"),
                   selected = "head")

    ),

    # Start a main panel in the center of the screen
    mainPanel(

      # Output dataset view to the main panel
      tableOutput("contents")

    )

  )),

 #----------------------------------------------------------------------------------------------------#
 #----------------------------------------DATA DESCRIPTION TAB------------------------------------
---#
 #----------------------------------------------------------------------------------------------------#

 tabPanel(
   # Dashboard Title
   titlePanel("Describe"),
```

```
  # Add a sidebar with the following options
  sidebarLayout(
    sidebarPanel(

      #Display a list of variable or field names to choose from a drop down list
      uiOutput("varNames")

    ),

    # Start a main panel in the center of the screen
    mainPanel(

      #Output variable class
      verbatimTextOutput("varClass"),

      #Output variable descrption
      verbatimTextOutput("varSummary1"),

      #Output variable summary information
      verbatimTextOutput("varSummary2")

    )

  )),

#-----------------------------------------------------------------------------------------------------#
#---------------------------------------VARIABLE SELECTION TAB-----------------------------------
---#
#-----------------------------------------------------------------------------------------------------#

    tabPanel(

      # Dashboard Title
      titlePanel("Verify"),

      # Add a sidebar with the following options
      sidebarLayout(
        sidebarPanel(

          # Allow the user to select the variable to be predicted
          uiOutput("predictVar"),

          # Allow the user to select the financial aid variable
          uiOutput("selectFinAid"),
```

```
        # Allow the user to select the predictor variables
        uiOutput("selectColumns"),

        #Allow the user to select all variables from the list to be input into the model
        actionLink("selectall","Select All/De-select All")
      ),

      # Start a main panel in the center of the screen
      mainPanel(

        # Display the list of dependent and independent variables chosen by the user
        verbatimTextOutput("selection"),

        # Show a correlation plot of variables
        plotOutput("corrPlot")
      )

    )),

#-----------------------------------------------------------------------------------------------------#
#-------------------------------------------MODEL BUILDING TAB--------------------------------------
-#
#-----------------------------------------------------------------------------------------------------#

   tabPanel(

    # Dashboard Title
    titlePanel("Model"),

    # Add a sidebar with the following options
    sidebarLayout(
      sidebarPanel(

        # Provide a button for user to activate the model building process
        actionButton("do", "Start Model"),

        # Display a confusion matrix so user can view model prediction performance
        tableOutput("conf_matrix"),

       #Add an Area Under the Curve Measurement with model performace guide
       h3("Accuracy (AUC) Measurement Guide"),
       h4(".90-1 = excellent (A)"),
       h4(".80-.90 = good (B)"),
       h4(".70-.80 = fair (C)"),
       h4(".60-.70 = poor (D)"),
       h4(".50-.60 = fail (F)")
```

```
      ),

      # Start a main panel in the center of the screen
      mainPanel(

        # Display an ROC plot
        plotOutput("ROCplot"),

        # Output the AUC metric
        verbatimTextOutput("AUC")

      )

    )),

#--------------------------------------------------------------------------------------------------#
#---------------------------------STUDENT OUTCOME PREDICTION TAB---------------------------
-----------#
#--------------------------------------------------------------------------------------------------#

  tabPanel(

    # Dashboard Title
    titlePanel("Predict"),

    # Add a sidebar with the following options
    sidebarLayout(
      sidebarPanel(

        # Show all factor variable choices in a drop down menu
        uiOutput("dropDown"),

        # Allow free form text entry for numeric variables
        uiOutput("Nbr"),

        # Allow user to enter the maximum amount of aid a student can receive
        # Default is equal to 0
        numericInput("MaxAid","What is the maximum amount of financial aid a student could
receive?",
                     value=0),

        # Provide a button which calculates the outcome of the student profile selected above
        actionButton("doPredict", "Predict Student Outcome")

      ),
```

```
    # Start a main panel in the center of the screen
    mainPanel(

      #tableOutput("testdf"),

      # Plots the amount of aid vs. the probability for the specified student
      plotOutput("FAPlot"),

      # Provide a slider for user to adjust financial aid amounts
      uiOutput("pctConf"),

      # Display th outcome probability and financial aid amount
      verbatimTextOutput("AidReq")
    )

  ))
 )
)

# End of the user interface code
##############################################################################
########################
# Beginning of the underlying functionality code

server <- function(input, output) {

 #Import Dataset
 output$contents <- renderTable({

  # input$file1 will be NULL initially. After the user selects
  # and uploads a file, head of that data file by default,
  # or all rows if selected, will be shown.
  req(input$file1)

  # Store data input in a dataframe
  df <<- read.csv(input$file1$datapath,
                 header = input$header,
                 sep = input$sep,
                 quote = input$quote)

  # If the user selects "head" once the data is imported
  # then display the top five rows of the dataset
  # otherwise display the whole dataset
  if(input$disp == "head") {
   return(head(df))
  }
```

```
  else {
    return(df)
  }

})

#Get a list of input variables
output$varNames = renderUI({

  # Make a selection list from the names in the dataframe
  selectInput('variables2', 'Variables', names(df))
})

# Save the variable class type
output$varClass = renderPrint(lapply(df[input$variables2], class))

# Save the description summary information for each variable
output$varSummary1 = renderPrint(describe(df[input$variables2]))
output$varSummary2 = renderPrint(summary(df[input$variables2]))

# Correlation prep and plot
output$corrPlot <- renderPlot({

  # Store all numeric variable names in a dataframe
  nums <- sapply(df, is.numeric)

  # Pull all numeric variables from the imported data
  dat <- df[ , nums]

  # Store the correlation information
  M < -cor(dat)

  # Plot the correlation information
  corrplot(M, method="circle")

})

#Provide a drop down menu for the user to select the prediction variable
output$predictVar = renderUI({

  # Provide possible prediction variables
  selectInput("predictorVar", 'Please select the dependent (i.e. predicted) variable:', names(df))

})

# Output a list of numeric variables which could possible used as financial aid variables
```

```
  output$selectFinAid <- renderUI({

    # Provide possible financial aid variables
    selectInput("selectFinAid","Please select the financial aid variable:",
names(which(sapply(df,is.numeric))))

  })

  #Output a list of column names for the user to select
  output$selectColumns <- renderUI({

    # Provide a list of possible independent variables to build the model
    checkboxGroupInput("selectData","Check the independent (i.e predictor) variables you would
like to keep in the model:", names(df))

  })

  observe({
    #If the user clicks Select All then check all boxes to select all variables
    if(input$selectall == 0) return(NULL)
    else if (input$selectall%%2 == 0)
    {
      updateCheckboxGroupInput(session,"selectData","Choose variables you would like to
keep:", choices=names(df))
    }
    #If the user clicks Select All again then unhcheck all the boxes
    else
    {
      updateCheckboxGroupInput(session,"selectData","Choose variables you would like to
keep:",choices=names(df),selected=names(df))
    }
  })

  output$selection <- renderPrint(

    # Display the selected variables to the user
    c(input$predictorVar,input$selectFinAid, input$selectData)

  )

  # Run Model
  observeEvent(input$do, {

    # Ensure the dataframe has not been transformed by turning it into a dataframe
    df <<- as.data.frame(df)
```

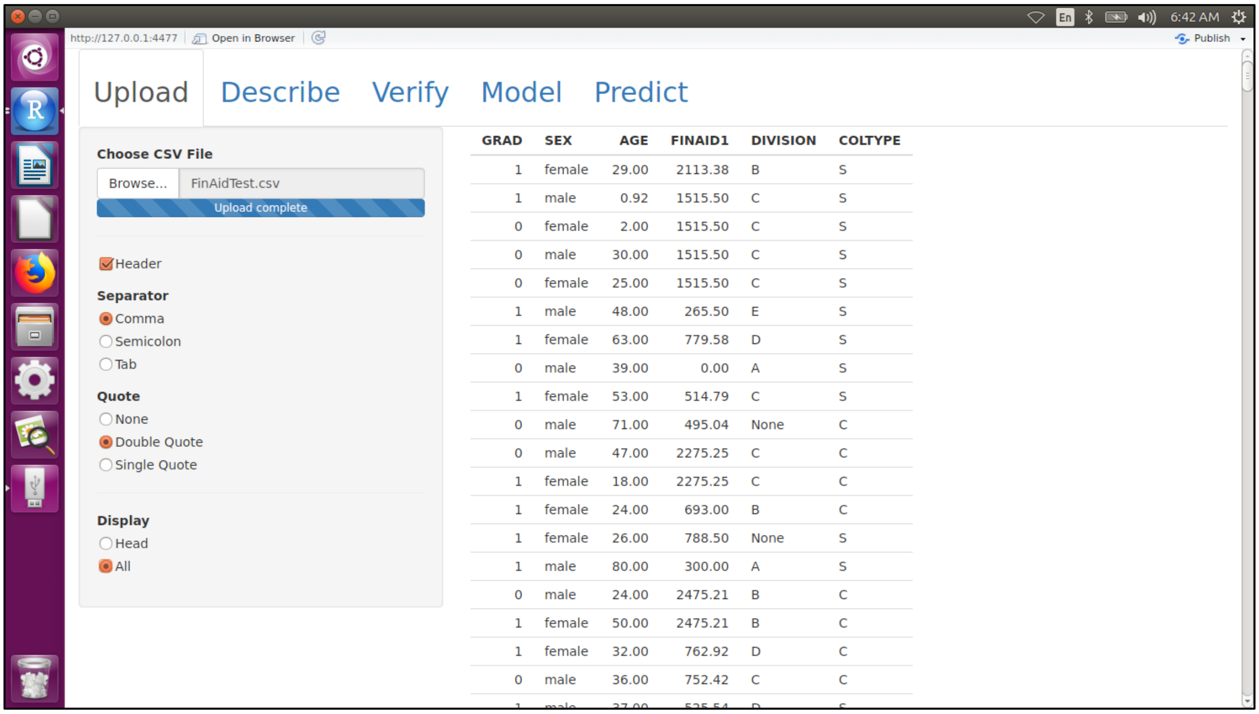
```
  # Naive Bayes formula
  NBFormula <- reactive({
    # Once the Model Button has been selected put this equation into the Naive Bayes model
    as.formula(paste(input$predictorVar, '~ .'))

  })

  # Ensure the prediction variable is a factor
  df[,input$predictorVar] <- as.factor(df[,input$predictorVar])

  # Save the model
  model <- naiveBayes(NBFormula(), data = df)

  # Predict all possible outcomes using
  preds <- predict(model, newdata = df)

  # Output a confusion matrix
  output$conf_matrix <- renderTable({

    # Save the confusion matrix table as a dataframe
    conf_matrix <- table(preds, df[,input$predictorVar])
    conf_table <- as.data.frame(conf_matrix)

    #Set the headers for the matrix
    names(conf_table) <- c("Predicted", "Actual", "Frequency")

    return(conf_table)

    }
  )

  # Model Accuracy
  output$ROCplot <- renderPlot(

    # Plot an ROC Curve
    plot(roc(as.numeric(preds),as.numeric(df[,input$predictorVar]), direction="<"),
col="yellow", lwd=3, ylim = c(0,1), xlim = c(1,0))

    )

  output$AUC <- renderPrint(

    # Output the AUC metric
    roc(as.numeric(preds),as.numeric(df[,input$predictorVar]), direction="<")[9]

  )
```

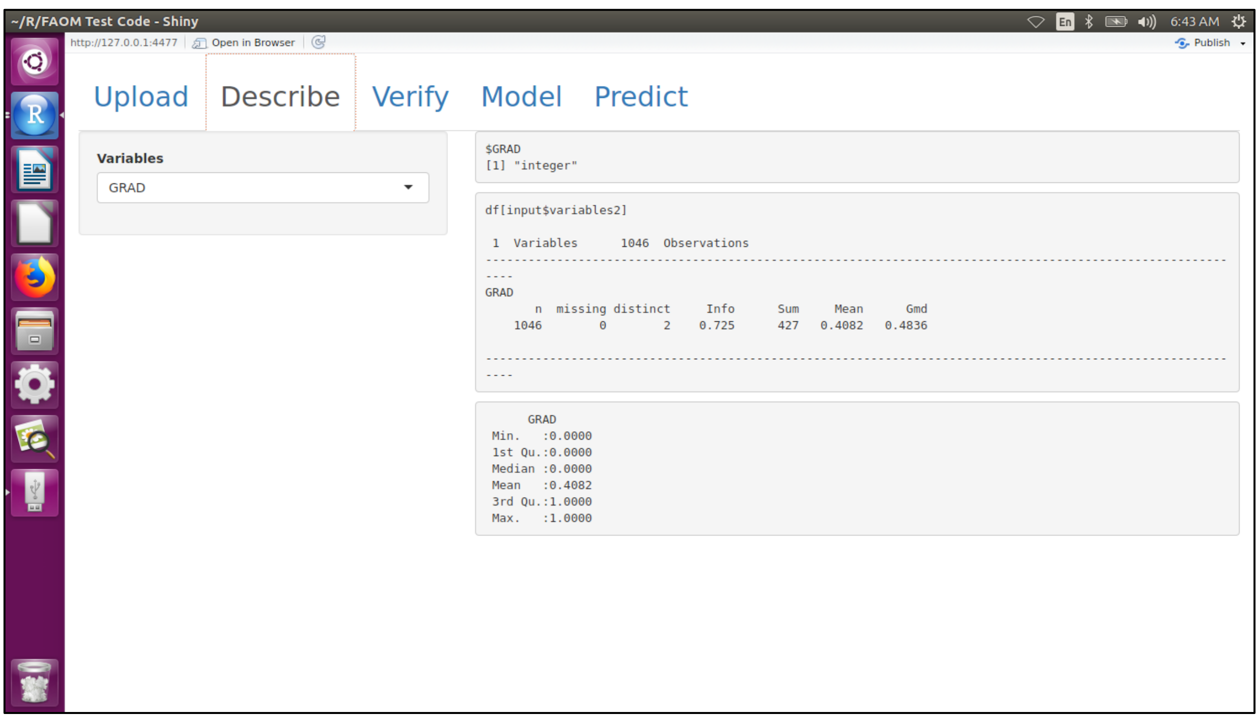
```
# Initialize list of inputs
inputTagList1 <- list()
inputTagList2 <- list()

output$dropDown <- renderUI({
  for(i in 1:length(df)){

    #If the predictor or the financial aid variable then do not make a user input selection
    if (names(df[i]) == input$selectFinAid || names(df[i]) == input$predictorVar) {}

    #If the variable is a character then make a dropdown menu of choices
    else if (is.character(df[,i]) || is.factor(df[,i])){
    newInputId <- paste0("input", names(df[i]))
    newInputLabel <- paste("Input", names(df[i]))
    newInput <- selectInput(newInputId,newInputLabel, choices = unique(df[i]))
    inputTagList1 <<- tagAppendChild(inputTagList1, newInput)
    }

    #If the variable is a numeric value then make user enter a value
    else {}
  }
  return(inputTagList1)
})

output$Nbr <- renderUI({
  for(i in 1:length(df)){

    #If the predictor or the financial aid variable then do not make a user input selection
    if (names(df[i]) == input$selectFinAid || names(df[i]) == input$predictorVar) {}

    #If the variable is a character then make a dropdown menu of choices
    else if (is.integer(df[,i]) || is.numeric(df[,i]) || is.double(df[,i])){
      newInputId <- paste0("input", names(df[i]))
      newInputLabel <- paste("Input", names(df[i]))
      newInput <- numericInput(newInputId,newInputLabel,value=1)
      inputTagList2 <<- tagAppendChild(inputTagList2, newInput)
    }

    #If the variable is a numeric value then make user enter a value
    else {        }
  }
  return(inputTagList2)
})

#Get user input for preductions
```

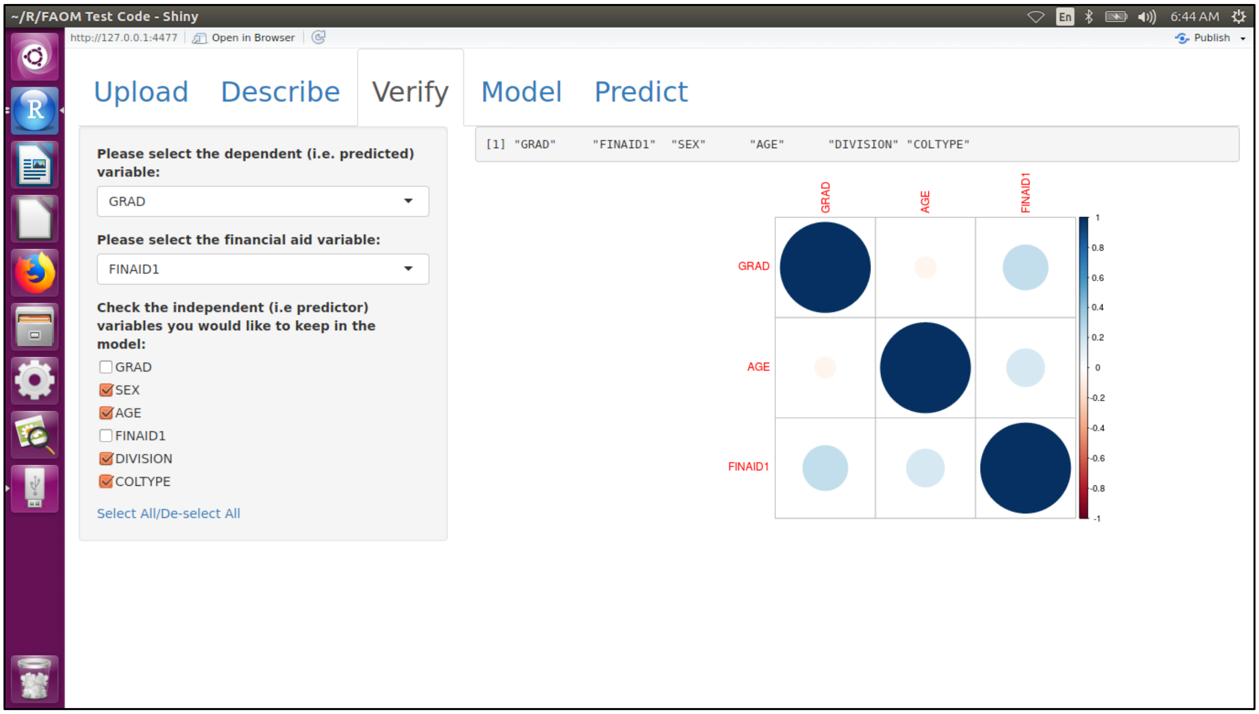
```
  observeEvent(input$doPredict,{

    # We need a new dataset for predictions
    newdf <- data.frame(matrix(ncol=ncol(df),nrow=1000))
    colnames(newdf) <- colnames(df)

    # Need to make a prediction dataset
    for(i in 1:length(newdf)){

      #If the predictor or the financial aid variable then do not make a user input selection
      if (names(newdf[i]) == input$selectFinAid || names(newdf[i]) == input$predictorVar) {}

      # Replicate the character variable 1000 times
      else if (is.character(df[,i]) || is.factor(df[,i])){
        userInput <- paste0("input$input", names(df[i]))
        newdf[i] <- rep(userInput,1000)
      }

      # Replicate the numeric variable 1000 times
      else {
        userInput <- paste0("input$input", names(df[i]))
        newdf[i] <- rep(userInput,1000)
      }

    }


    #output$testdf <- renderDataTable(newdf)

    # Make a column consisting of various financial aid amounts for prediction
    newdf[,input$selectFinAid] <- seq(0, input$MaxAid, by=input$MaxAid/999)

    # Predict the probability outcomes given the different amounts
    preds2 <- predict(model, newdata = newdf, type="raw")

    # Plot the aid amount against the outcome probability
    output$FAPlot <- renderPlot(
      plot(newdf[,input$selectFinAid],preds2[,2], xlab = "Financial Aid Amount", ylab =
"Probability of Success")
    )

    # Output the financial aid amount along with the associated outcome likelihood
    dfout <- data.frame(matrix(ncol=2,nrow=1000))
    names(dfout) <- c("FA", "Score")
    dfout$FA <- newdf[,input$selectFinAid]
    dfout$Score <- preds2[,2]
```

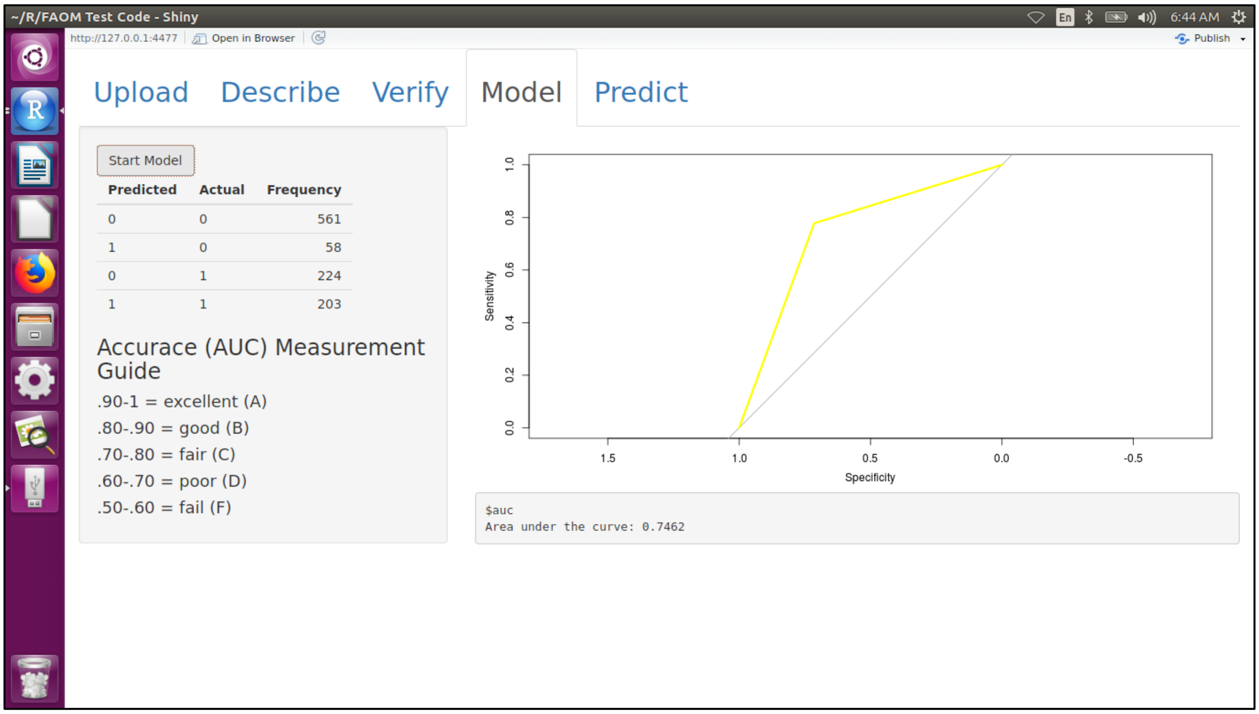
```
    # Provider a slider for user to input success likelihood they are seeking
    output$pctConf <- renderUI(
      sliderInput("obs", "Select the probability of success to find the needed amount of financial
aid:", min = 0, max = 1, value = .5)
    )

    # Returns minimum value of financial aid for the users specified probability
    output$AidReq <- renderPrint(
      dfout[min(which(dfout$Score >= input$obs)),]
      )
  })

 })

}

##############################################################################################################
# End of the underlying functionality code

# Activate the Shiny application feature
shinyApp(ui = ui, server = server)
```

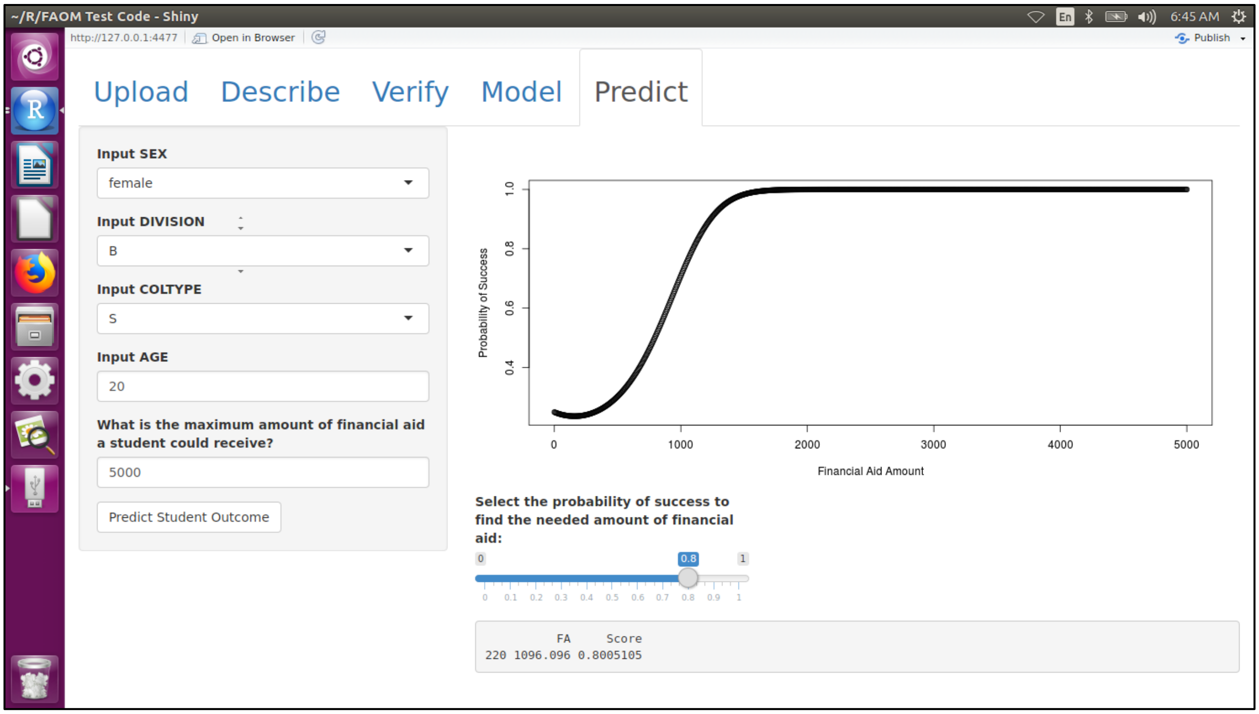
APPENDIX B: FINANCIAL AID ANALTYICS APPLICATION USER INTERFACE






~/R/FAOM Test Code - Shiny
http://127.0.0.1:4477
Open in Browser
Publish
Upload
Describe
Verify
Model
Predict
Please select the dependent (i.e. predicted) variable:
GRAD
Please select the financial aid variable:
FINAID1
Check the independent (i.e predictor) variables you would like to keep in the model:
GRAD
SEX
AGE
FINAID1
DIVISION
COLTYPE
Select All/De-select All
[1] "GRAD" "FINAID1" "SEX" "AGE" "DIVISION" "COLTYPE"
GRAD
AGE
FINAID1


~/R/FAOM Test Code - Shiny
http://127.0.0.1:4477
Open in Browser
Publish
Upload
Describe
Verify
Model
Predict
Start Model
Predicted Actual Frequency
0 0 561
1 0 58
0 1 224
1 1 203
Accurace (AUC) Measurement Guide
.90-1 = excellent (A)
.80-.90 = good (B)
.70-.80 = fair (C)
.60-.70 = poor (D)
.50-.60 = fail (F)
Sensitivity
Specificity
$auc
Area under the curve: 0.7462


~/R/FAOM Test Code - Shiny
http://127.0.0.1:4477
Open in Browser
Publish
Upload
Describe
Verify
Model
Predict
Input SEX
female
Input DIVISION
B
Input COLTYPE
S
Input AGE
20
What is the maximum amount of financial aid a student could receive?
5000
Predict Student Outcome
Probability of Success
Financial Aid Amount
Select the probability of success to find the needed amount of financial aid:
FA Score
220 1096.096 0.8005105

APPENDIX C: INFORMED CONSENT

Title of Study: Utilizing Business Analytics for Finacial Aid Management: A Design Science Research Study

Investigator: Chantel Perry

Contact Information: contact@chantelperry.com; 304-906-3316

Purpose of the Study

You are invited to participate in a research study. The purpose of the research study is to evaluate a new software tool designed to understand the performance of state financial aid awards in achieving postsecondary degree completion. The intent of the study is to provide higher education leaders with actionable insights for decision-making with grants management.

Participants

You are being asked to participate in the study because you have experience, knowledge, and insights in supporting the state government in financial aid decision making through data analytics.

Procedures

If you volunteer to participate in this study, you will be asked to do the following:

- View a software demonstration (i.e. 20 minutes)
- Provide your open and honest opinion about the sofware and its outputs (i.e. 60 minutes)
- Hand in your notes about the software

Benefits of Participation

There may/may not be direct benefits to you as a participant in this study. However, we hope to learn more about methods and techniques which do and do not work for understanding state aid programs and how they effect student outcomes.

Risks of Participation

There are risks involved in all research studies. This study is estimated to involve minimal risk. An example of this risk is “feeling uncomfortable with providing an opinion about the current methods of financial aid management at your agency”.

Cost/Compensation

This will be no financial cost to you to participate in this study. The study will approximately 90 minutes. You will not be compensated for your time, but food and beverages will be provided. *Colorado Technical University will not provide compensation or free medical care for an unanticipated injury sustained as a result of participating in this research study.*

Contact Information

If you have any questions or concerns about the study, you may contact Chantel Perry and Dr. Livingood, RLivingood@coloradotech.edu, 520-296-4695. For questions regard the rights of research subjects, any complaints or comments regarding the manner in which the study is being conducted, you may contact Colorado Technical University – Doctoral Programs at 719-598-0200.

Voluntary Participation

Your participation in this study is voluntary. You may refuse to participate in this study or in any part of this study. You may withdraw at any time without prejudice. You are encouraged to ask questions about this study at the beginning or at any time during the research study.

Confidentiality

Participants may provide information which could be considered sensitive or used negatively. It is the responsibility of the researcher to minimize risks of abuse against informants during the research process. Data and records will be stored in a locked filing cabinet and on a password protected hard drive to keep participants’ data safe. These materials will only be accessible only to the primary researcher and transcriptionists. Data movement will also be limited to minimize the risk of losing information. At the end of the study, the media device will be first be cleared by overwriting the data file and then by performing a manufacturer reset of the instrument. The device will then be purged through a degaussing process as recommended by the hard drive manufacturer. Finally, the device will be either disintegrated or pulverized at a licensed facility. A certificate will be obtained upon the destruction of the instrument.

Participant Consent

I have read the above information and agree to participate in this study. I am at least 18 years of age. A copy of this form has been given to me.

Signature of Participant Date

Participant Name (Please Print)

APPENDIX D: FOCUS GROUP QUESTIONNAIRE

1. What is your initial impression of the software tool?
2. Are enhancements necessary or helpful to make this tool more useful, more effective, and/or more efficient?
3. In general, what would you take away from this tool?
4. In general, what additions would you make to this tool?
5. In what ways did this tool help you understand how financial aid impacts college student degree completion?
6. What additions would you make to this tool in helping you understand how financial aid impacts college student degree completion?
7. How do the tools you currently use for understanding financial aid impacts on Kentucky college student degree completion compare to this tool?
8. What benefits does this tool provide in gaining insight into financial aid resource allocation?
9. What pitfalls does this tool have for gaining insight into financial aid resource allocation?
    a. What might you add or take away from this tool to overcome these pitfalls?
10. Do you think it may be appropriate to use this tool making financial aid decisions? If so, how, and when in the process might it be applied?

APPENDIX E: SITE PERMISSION LETTER

FINANCE AND ADMINISTRATION CABINET
KENTUCKY HIGHER EDUCATION ASSISTANCE AUTHORITY

Matthew G. Bevin
Governor

P.O. Box 798
Frankfort, Kentucky 40602-0798
Phone: 1.800.928.8926
Fax: 502.696.7373
www.kheaa.com

William M. Landrum III
Secretary

Carl P. Rollins
Executive Director

11 October 2017

Ms. Chantel Perry
6039 Whitby Road, #707
San Antonio, TX 78240

Subject: Doctoral dissertation research at Colorado Technical University.

Dear Ms. Perry:

The Kentucky Higher Education Assistance Authority (KHEAA) welcomes opportunities to collaborate with scholars, such as you, to further academic research evaluating the outcomes of postsecondary student financial aid. KHEAA, the agency in the Commonwealth of Kentucky charged with administering the preponderance of state funded postsecondary student financial aid programs, will provide ancillary technical assistance and consultation for your research entitled, *Utilizing Business Analytics for Financial Aid Management: A Design Science Research Study*, to the extent permissible within agency administrative policies, procedures and priorities, and subject to existing federal and state statutes.

KHEAA will assist you in developing personal contacts within the postsecondary education financial aid community in Kentucky and among other individuals professionally associated with education. Also, you will be able to use one of KHEAA's public conference rooms, subject to established reservation procedures, to hold a focus group meeting of education professionals.

KHEAA understands that you will complete the Kentucky Center for Education and Workforce Statistics (KCEWS) data use and confidentiality agreements as required in order to obtain access to that agency's public data as the primary data for your research. To the extent the terms of that agreement authorize use of KCEWS' aggregated public access longitudinal state data system for academic analysis and reporting on postsecondary student financial aid policy and outcomes, and you fully comply with those contractual terms for that access, no further permission or authorization of any type is required for you to perform your analysis. The KCEWS data use

KentuckyUnbridledSpirit.com

Kentucky UNBRIDLED SPIRIT

An Equal Opportunity Employer M/F/D

agreement contains a section on data confidentiality. Those contractual terms shall govern, at all times and in all circumstances, without exception, any and all issues or questions concerning KCEWS data confidentiality, data access, and use of data from that source.

I have asked Dr. Melvin E. Letteer to be your point of contact and for him to assist you with your dissertation research. Please feel free to call on him directly. He may be reached at (502) 696-7471 or by email at MLetteer@KHEAA.com.

We are pleased to be of assistance and are very interested to see the outcome of your research. I wish you the best of luck with your dissertation.

Sincerely,

Carl P. Rollins

Carl P. Rollins, Ph.D.
Executive Director
Kentucky Higher Education Assistance Authority
Kentucky Higher Education Student Loan Corporation