Utilizing Predictive Analytics to Aid Project Continuity Decision Making

by Mohammed Altaf Hossain

B.S. in Information Technology, May 2003, Rochester Institute of Technology
M.B.A. in Finance, May 2009, Marymount University

A Praxis submitted to

The Faculty of
The School of Engineering and Applied Science
of the George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Engineering

May 19, 2019

Praxis directed by

John Fossaceca
Professorial Lecturer in Systems Engineering

Muhammad Islam
Professorial Lecturer in Systems Engineering

The School of Engineering and Applied Science of The George Washington University certifies that Mohammed Altaf Hossain has passed the Final Examination for the degree of Doctor of Engineering as of date of March 01, 2019. This is the final and approved form of the Praxis.

Utilizing Predictive Analytics to Aid Project Continuity Decision Making

by Mohammed Altaf Hossain

Praxis Research Committee:

John Fossaceca, Professorial Lecturer in Systems Engineering, Praxis Co-Director,

Muhammad Islam, Professorial Lecturer in Systems Engineering, Praxis Co-Director.

Amirhossein Etemadi, Assistant Professor of Engineering and Applied Science, Committee Member.

Ebrahim Malalla, Visiting Associate Professor of Engineering and Applied Science, Committee Member.

Dedication

To Allah (God): who taught men what he knew not (Quran 96:5).

To Prophet Mohammed (pbuh): who said to seek knowledge from the cradle to the grave.

To my family: without their unconditional support this journey would have been challenging.

Acknowledgements

There are people in this world who live a selfless life and live to help others. I would like to acknowledge a few of those people from my life. The following people have influenced my decisions that shaped my life:

Mrs. Sally Mae Izetta Alston Cooper

Mrs. Ann Neg Tran Shultis

Mr. Victor Del'Mastro

Mr. Mohammed Harunur Rashid

Mr. Mohammed Elias Ullah

I would like to thank my research committee for their ongoing support in completing this praxis. Thank you!

Abstract of Praxis

Utilizing Predictive Analytics to Aid Project Continuity Decision Making

Although corporations collect project information such as performance indicators, they tend to use this information only to estimate the cost of future projects rather than as a way of predicting potential project failures. By not using performance indicators from prior projects to predict future project success or failure, organizations are not taking advantage of information that can be used to avoid spending time, money, and resources on projects that are likely to fail. If companies were able to accurately predict project failures, they could avoid incurring the opportunity cost of carrying out projects that should be cancelled early in their life cycle. This praxis utilizes predictive analytics based on project performance indicators as inputs to identify potential project failure candidates allowing managers to stop project work and redirect project resources to other potentially successful projects. In this study we demonstrate that predicting project failure based on past performance of similar projects could enable organizations to make scientific, data-driven, and evidence-based decisions on whether a project should continue. The proposed model recommended in this praxis yields an average 98.56% prediction accuracy with 0% False Positive Rate (FPR). Data used in this praxis is from a large size organization with employees numbering in the tens of thousands. By predicting projects likely to failure after phase two based on past project performance data, this study found that the organization would save about $300,000 on average, per project.

Table of Contents

Chapter 1: Introduction

1.1. Overview

The aim of this study is to establish utilization of predictive analytics in everyday projects to aid leadership decision-making in the allocation of financial resources. Traditionally, project continuity decisions are made within the constraints of budget, schedule, resources, and expert recommendations. This praxis proposes the use of past project performance data in order to make predictions about the potential for success or failure for current projects. Traditionally, predictive analytics have been used decision-making in sales and marketing (Halper, 2014). Because today's computational and analytical resources and algorithms are readily accessible, along with vast collections of project performance data, it is possible for organizations to make predictions about project outcomes. In the past, keeping detailed records of completed projects were tedious and costly, but record-keeping has become less expensive in recent years as a result of automated data collection tools and cloud storage capabilities (McKemmish, 2013). Today, organizations hoard more data than ever before, but in some cases, hardly utilize the saved data. There is no organizational objective standard for storing these data other than organizations "wanting to" store data. Many organizations lack plans for using these data, formatting the data, or identifying critical data attributes while storing the data (Wiewiora & Murphy, 2015). By understanding an organization's specific vision for growth and formatting these data as they are stored with a plan for future use, organizations would obtain a competitive advantage in their respective industries. Most organizations usually undertake similar projects or projects within their respective domains. The problem this praxis addresses is that organizations do not use their past project performance data to minimize project failure while advancing in their domains. Indeed, organizations should use their past project performance data to predict the future of a current project by utilizing all available project attributes. This prediction analysis will help organizations identify potential project failure candidates, suspend current project work, and redirect the current project's resources toward more projects with a higher likelihood of success.

1.2. Statement of the Problem

Past project data is not being used to identify projects that are likely to fail and therefore companies are optimizing use of their resources. Although corporations collect project information such as performance indicators, they tend to use this information only to estimate the cost of future projects rather than as a way of predicting potential project failures. By not using performance indicators from prior projects to predict future project success or failure, organizations are not taking advantage of information that can be used to avoid spending time, money, and resources on projects that are likely to fail. These performance indicators are project type, end user type, technology type, internal review, and time spent in phases, etc. Project failure in this praxis defined as project failed to be monetized. If companies were able to accurately predict project failures, they could avoid incurring the opportunity cost of carrying out projects that should be cancelled early in their life cycle. This praxis addresses the need to use predictive analytics and past project performance in project management practice to stop current project work ahead of project completion to save time, money, and project resources. This waste of time, money, higher rates of unsuccessful projects, less successful projects undertaken in a given year, inefficient use of resources, increased attrition rate, a poor industry reputation, and increased cost of re-staffing are all vital consequences of failing to predict project failure. Many studies were completed in project success criteria analysis and identifying root cause of project failure, but overall project performance has not been evaluated based on past project performance indicators (Costantino & Nonino, 2015). In this quantitative research, some theoretical constructs were investigated. First, how to identify candidates for potential project failure from project performance indicators? Organizations would identify common elements within projects which lead to project failure. Factor Analysis (FA) is used to examine projects element contribution in factor creation, and the uniqueness of the project elements. Like FA, Principal Component Analysis (PCA) was also used to examine the contributing factors in determining the most important components. An iterative process was used to select project elements from factor one and/or principal component one to perform project failure prediction. This is repeated with project elements from factor two and/or from principal component two to compile the final list of project elements to attain highest accuracy and lowest false positive rate. Four different models were then used to understand and identify relationships between project performance indicators with respect to project success and failure, to perform failure prediction from the weighted average of the past project performance indicators on a current project being worked. These models are Decision Tree (DT), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Support Vector Machine (SVM). While all four models accurately predicted the project failure, model with the highest accuracy and lowest False Positive Rate (FPR) was recommended for identifying the project failure candidates.

1.3. Hypotheses and Approach

As part of the data analysis and sense-making between recorded project performance indicators and their interrelationships, this praxis utilized several approaches. One such approach was to find indicators that were true for all projects that either passed or failed. A list of 29 performance indicators out of 79 were then selected. Next the elements from the PCA and FA were used to create the list of variables to calibrate the models, and that list of variables was further refined based on the performance indicators' uniqueness and relevance to the hypothesis. Once these indicators successfully yielded high accuracy these results are validated in reverse with the findings by selecting these indicators to see if all identified projects either passed or failed. During this process all 29 identified project performance indicators would be evaluated to create the final list of 12 performance indicators for this praxis. Periodic internal reviews, the type of project, the type of end user, time spent completing phase one and phase two work, role the company plays and the partnership it held in undertaking these projects should be weighted in this selection process to support the hypothesis. Finally, this praxis seeks to find whether a prediction can be made with identified indicators.

1.3.1. Hypotheses

H1: Project success or failure can be forecasted from project performance indicators. Accept.

1.3.2. Null Hypotheses

$H1_0$: Project success or failure cannot be forecasted from project performance indicators.

1.3.3. Approaches

This praxis utilized project data from a large organization with known outcomes. Initially, a study was completed of the projects with a known state of pass or failed. Next, this praxis identified pass or fail relationship with the type of projects and exploration of "driving factors" was studied with questions such as "Were any end users for these projects?", "Did projects fail after the gate one review or gate two review?, "What partnerships had the company undertaken with other organizations for these projects?", "How much time was spent in each of the phases of these projects?", "Were these projects were for domestic customers or international customers?", "Was the project assigned any technical lead?".

To define and categorize, *type of project* consists of domestic, international, classified, and commercial. *Gate one review* where projects probability to go is weighted (see **Appendix B**), and *gate two review* where project "probability to win" is weighted (see **Appendix C**). An *end user* is determined based on the project type and who is funding the project. A *technical lead* consists of Chief Engineer (CE), Principal Engineer (PE), Content Manager (CM), Business Development Lead (BDL), and Project Manager (PM). These individuals provide technical guidance and connect project resources to various internal and external expertise. Internal or external resources include cross-mission area experts within and outside the organization. Usually these leads are partially allocated to various projects, and for larger organizations (those with thousands of employees), most resources are usually found within the organization. In exceptional situations, when a product under development requires field testing to obtain field data, external resources are utilized. Having an internal technical lead assigned to the project helps coordinate all project resource needs, including projects with end users and/or end user requirements. Usually, the problem, or the pain the end user community experiences, produces these requirements that leads to the inception of many of these projects. Cost of these projects are measured by the hours spent in project phases in the project lifecycle. The results are discussed in Chapter 4 of this praxis.

1.4. Purpose and Research Questions

The overall purpose of this study is to aid leadership decision-making processes in decisions of project continuity. Specifically, if a project's leadership team utilized the models this praxis presents and already knew the health of a given project, would the leadership team suspend current project work knowing the project will fail based on the performance data of past projects? <u>Research questions include: how can organizations identify potential candidates for project failure from current projects based on project performance indicators from prior project performance data? Can a predictive model be built based on prior project performance data that can be applied to currently running projects?</u>

1.5. Statement of Potential Significance

The idea of utilizing past project performance data via a statistical tool to aid project continuity decision-making would change not only how project data are collected or how projects are evaluated during their lifecycle but, would also change how projects are selected and managed from beginning to end. This praxis proposes a change in project risk management to perform an ongoing evaluation of project health. This change would add value to which project data are being collected, and how the data are stored. This concept would also evaluate the importance of which project attributes enable management teams in assessing the health of a future project. As a result, this concept would affect project planning, the use of different types of project management tools, and methods for project reporting. Once adopted, organizational processes for planning projects would change, and forecasting the potential success or failure of a project would be weighted even before projects begin—especially for research, development, proof-of-concept (POC), and prototype type projects. Many studies were conducted to identify project critical success factors (CSFs), and how to cluster and weight CSFs to make project selection decisions, as well as how to improve the management of future projects. However, no study has utilized past project performance data to derive project continuity decisions via predictive analytics, nor have previous studies used the artificial intelligence (AI) based data feed to make project planning decisions. Combining past project data and AI would alter the landscape of project management considerably soon, and significantly for years to come. Whether the project is a construction project in China (Zhao, 2018) or an IT project in India (Mukerjee, 2017), all relevant geographic, weather, socio-political sentiment, and regulatory information will form the AI capable of enabling a PM tool to identify critical tasks, incorporate safety and regulatory risks while developing project plans and project schedules, and predict overall project success criteria or possibilities of failure. The study in this praxis is one positive step toward this reality.

1.6. Theoretical Foundation

The theoretical foundation in this study shows how predicting project failure before or during project work, based on performance data of similar projects in the past, would help project stakeholders make better decisions. This study focuses on similar projects from a large organization based on similar attributes to predict if a similar current project will succeed or fail. Stakeholders and decision-makers would then have the necessary information to ask difficult questions: should the current project continue to completion, or should management terminate project work upon current phase completion and redirect resources to the next project? If the project has an extrinsic value to the organization, then a tradeoff should be assessed based on the total value, which is not part of this study, but can be part of future studies.

1.7. Summary of the Methodology

First, a review of the current literature presented the need for a wholesome approach to understand how various project performance indicators work together to project success, or project failure. The following is a summary methodology of this praxis:

1) Identify trends in data and analyze data relationship in full, applicable comprehension of past project performance data;

2) Perform data summary for initial data cleanup by eliminating missing and unrelated data attributes;

3) Perform Factor Analysis (FA) to understand the weights of these project performance indicators and the p-value;

4) Identify the uniqueness of these project performance indicators as highly unique values signifies less correlation with the other performance indicators or data value;

5) Mask company proprietary data and perform various data classifications for prediction;

6) Take three years data to train the model and use the fourth-year data to validate the models;

7) Perform t-test to establish average time spent in each of the project phases and perform time value money to convert those hours into amount of money spent in each of these phases;

8) Forecast possible savings if one or more of these project phases were not completed because the project was identified as failure candidate and stopped prior to completing that and/or subsequent phase;

9) Lastly, utilize the past project prediction to predict the current project pass or failure based on the given project attributes to demonstrate that a project with given attributes could be predicted failure ahead of project being completed.

1.8. Limitations and Delimitations

A limitation of this praxis is that it focuses on quantitative analysis, and not qualitative—an analytical approach which should be studied in future work. This praxis focuses on overall project failure based on past project performance activities and not on qualitative information from project decision makers. Model accuracy are based on the company data used for this praxis; therefore, the result will differ depending upon the type of organization and the project data being studied utilizing this model. While the overall prediction accuracy is important, this praxis focuses on the FPR for accurate identification of the failed project. The value of this model accuracies is presented purely in dollar amounts saved, but do not take any extrinsic value of the project to the company into consideration. Only time and money are part of this study, and future work could include the effects of a failed project on a work environment and a workforce. In addition, cost of attrition and negative extrinsic effect on the company are not part of this study. Nevertheless, this praxis focuses on evaluating the success or failure of current projects, and not on the selection of future projects. Savings will be based on suspending or terminating current project work, and not whether the project should be pursued differently with identified CSFs corrected. Many types of research, development, proof-of-concept (POC), and prototype projects carry soft and extrinsic value to their organizations, as they highlight each organization’s capabilities. Therefore, these projects will not be affected by this study, as these projects will continue to completion irrespective of the outcome of the models used in this praxis. Modification of project management tools to simplify project planning based on past projects, utilizing AI data based on the project logistical plan, supply-chain management, geo-political sentiment, and time feeding into the project plan, and predictability of project success or failure with project planning could form the focus of future studies.

Chapter 2: Literature Review

2.1. Introduction: Topics, Purposes, and Methods of the Literature Review

The topic for this literature review is project management with predictive analytics as well as project failure with predictive analytics. The purpose of this literature review is to see if any previous studies have utilized predictive analytics either to predict project failure or to derive a decision based on predictive analytics. Key data elements from past studies were used to better align this study with past research. These data elements are critical success factors (CSF), team performance and the importance of working together (TP), understanding requirements and customer needs (UR), time spent in Phase 1 and the subsequent phases (TS), stakeholder involvement (SI), and utilizing past project performance data to predict the future of the current project. Each of these elements play a critical role in project success and would collectively help decide the fate of the project. Stakeholders at an individual level would learn how to make better project continuity decisions and learn when to stop a project, while the organization at large would benefit from more successful projects by eliminating projects with a high possibility of failure. The limitations of this literature review remain with the data elements used in this study. Based on the organization, the organization's industry or line of business, and respective data elements from past projects, additional research could possibly cover specific project management activities or key areas for better project success or utilized to eliminate risky projects with higher probabilities of failure. Focus of this praxis remains in identifying project management activities utilizing predictive analytics. Additional project management topics could be part of future research.

2.2. Data Collection and Utilization

As organizations are create "data lakes" by gathering as much data as possible, but it is important to make sense of the data, create use cases purpose behind collecting the data, and study the data to identify trends in data like the one in this praxis. Not just for data analysts, but all across the organizational hierarchy. It is especially important for senior management to understand, support, and push for the effective use of data.

Mayhew, Saleh, and Williams (2016) suggest eight critical elements when working with data: ask the right questions, think small but big, embrace taboos, connect the dots, run loops not lines, make output usable, build a multiskilled team, and adopt the deliverables. These elements include expert's opinion and data driven mathematical output, which will give data meanings and purposes for the organizations' success.

Roger (2009) in his book titled Software Engineering – a practitioner's approach, he recommends collecting data from past projects and using those data to forecast as much possible of the current project. Forecast cost, schedule, resources, errors to failure, to successfully undertake software engineering projects. He also discusses how some things can never be predicted, like ever changing customer requirements, changing technologies, design, process, and the human element of a project.

Pickett and Elliott (2007) discuss in their research the use of company historical data in favor of the company's strategic advantage. They argue that only a few companies has the mechanisms in place to collect meaningful data. Company past project undertaking can help decide which project would align with the company future and project data can help decide if the project is on the right track. However, they did not discuss predicting company performance and future project nor the project success or failure criteria utilizing these historical data to aid their current decision-making process, which is the gap this praxis is bridging.

2.3. Project Failures and Management

According to McGrath and Martin (2017), 50% of projects fail due to poor overall visibility of the project management process. The authors discussed how project managers seldom learn from past project failures and recommend establishing an office of project management. They further recommend an establishment of governance, transparency, quality assurance, eliminating redundancy, and management of historical details about the projects, yet do not clarify the importance of these historical details about past projects, or how these details might be utilized to reduce project failure or to aid the process of successful project management. Laumer, Maier, and Weitz (2017) discuss the importance of having structured data for information to work and creating an Enterprise Content Management (ECM) department with the end user keeping organizational objectives in mind. As organizations collect and store data, they should always consider the use of these data. If data is structured for future use while being stored, then future usage of the data would be more effective. By 2020, Pecorino (2018) speculates that there will be around 50 billion active devices leading to new research behind big data streams and fog computing. While Pecorino (2018) discusses future learning management systems (LMS) and fog computing, this praxis focuses on structuring data for meaningful future use.

McGrath and Martin (2017) identify various reasons why projects fail, but do not consider those failures in relation to time, money, and resources used for projects. Instead, they recommend process improvement by implementing lean, six sigma, or agile project development processes. While these standards of development processes work if implemented and practiced correctly, implementing, following, and sustaining these standard processes comes with a cost. It is important for organizations to continuously measure the effectiveness of these standards and evolve to maximize the benefits because, without realizing the benefits of these practices and reasons for following them, practitioners could become unproductive. As a result, the cost of following these standards would outweigh the benefits with increasing tensions and constraint in project success. The authors do not discuss any cost value analysis, return on investment, or breakeven point if these standards were implemented and followed. Missing from their study is an assessment of the risk tolerance of organizations based on organization size. It is often more cost effective to have a non-standard development process with ongoing predictive decision-making concerning project performance, and whether project work should continue.

2.4. Project Failures and Expectations

In the last decade, organizations and governments spent an estimated $1 trillion on IT hardware, software, and services worldwide (Charette, 2005). Of the IT projects started, 5% to 15% were abandoned before or shortly after delivery. Many other projects were completed late as a result of exceeding the project budget or the need for massive reworking. It is critical to note that software failure is predictable and avoidable, but the most common hurdle for predictive analysis of potential project failures is the understanding and correct usage of past project performance data. Engle (2016) suggests using project failures as a resource for improving project deliverables. Project management should use continuous improvement processes to mitigate risk management.

Although it is critical to project success, organizations continue to fail when it comes to prioritizing tasks which lead to preventing project failures. This is primarily due to ill-defined system requirements, poor estimates of required resources, weak communication methods, unprepared staff working on complex projects, minimum reporting standards, lack of project management, and stakeholder politics. More than half (50-90%) of information technology projects have been reported to not achieve their project goals within schedule (Engle, 2011). Roger (2009) suggests predicting project failure from project defects data. He also recommends bringing various best practices to prevent project failures and execute more successful projects. Others have implemented scenario planning to help avoid project failures (Derbyshire & Wright, 2016; Wach, 1985a,b). Scenario planning puts an emphasis on finding and sorting trends which lead to project uncertainties.

2.5. Project Failures and Team Dynamics

Successful projects must have strong teams. Chen and Schiele (2017) indicated small and medium enterprises (SMEs) positioning strategies may vary depending on firm size in reference to team effort versus lone effort. Putting the right group of people together would influence and impact project performance. Effective leadership and engaged stakeholders drive many challenging projects to succeed. Althonayan and Althonayan (2017) illustrate how stakeholder performance and the performance of the project are inextricably connected and mutually influential. One of the hypotheses in this praxis will illustrate how a project of the subject organization failed because the project lacked both a target end user and a chief engineer. Studies show that the right set of resources will lead to project success. That is true for this study, where complete sets of required resources lead to project success, including projects without end user or end user requirements. However, while organizations may believe a fixed set of individuals can lead projects to success, the reality would be difficult to prove true. Data used in this study rejects the hypothesis that a fixed set of individuals can always lead to project success.

Hodge, Turner, and Sanders (2017) in the *Journal of Behavioral Health Services and Research* discuss Evidence Based Program (EBP) implementation to enhance organizational capacity. The authors evaluated various components, such as increasing program benefits, lowering program burden, increasing workplace support and cohesion, and implementing positive leadership style, and how sustaining such implementation over three years of post-training can enhance an organizations' capacity.

In general, all projects would share degrees of commonality, like phases, team players, and activities, among others. Ellinas, Allan, Durugbo, and Johansson (2015) state that a project can be seen as a large interactive system in which the complexity of the system emerges at a number of dimensions. In their study, the authors focus on identifying project work from a single task that causes catastrophic failure, and a cascading effect that leads a large-scale project to fail as a direct result of this complexity. Just because tasks are running smoothly during the beginning stages of a project, is not predictive of the final project stages. The remaining project steps, which are incomplete, can still influence the remaining project cycle. Ellinas, Allan, Durugbo, and Johansson (2015) provide a procedure to sort tasks adept of such impact, understands and underlines the process and approach, and draws parallels with natural science. Their study also analyzes the project and activities from the perspective of critical path, activity on node, evaluation, and review of project activities to add weight to each of the activities. This methodology reveals how one local task can impact the project at the global level. Node betweenness, another popular metric, does a dissatisfactory job predicting project failure when it comes to detecting the size of failure. This is analogous to Lawyer's (1959) concerns of low accuracy when using such metrics. Critics still encourage using insight form a network perspective rather than older traditional methods. This insight could swiftly allocate tasks which are critical and at the same time prevent reproduction. However, the study conducted by Ellinas, Allan, Durugbo, and Johansson (2015) does not compare similar projects or similar activities from past projects to validate the accuracy of the weight, nor does it weigh time and money in preventing such activity from failing. The study therefore does not consider predictive analytics as part of project management to assist in project continuity decision-making. The authors conclude that further work is needed to identify various restrictions like resource availabilities and cash flow. If past project activities, restrictions, and critical tasks were collected, then possible predictions could identify these critical tasks. These can aid decision-making by drawing similitude from past projects before initiating the current project.

2.6. Defining Factors of Project Failure

Many researchers conducted studies in the areas of identifying Critical Success Factors (CSF). While every project contains multiple CSFs, these vary from project to project. Maghsoodi and Khalilzadeh (2018) surveyed experts in the construction field to determine the percentage of importance given to critical success factors including time, cost, quality, and safety. Because of the high costs of failure in the construction industry, there is an increased need to identify critical success factors as well as rank these CSFs in order of priority. Maghsoodi and Khalilzadeh (2018) use one of the Multiple Criteria

Decision Making (MCDM) methods, known as Fuzzy Technique for Order Preference by Similarity to Ideal Solution (Fuzzy TOPSIS), to rank critical success factors. According to their research, the level of the effect of each critical factor on the success of each project will be provided. Although they have made various predictions of project risk, they did not produce a conclusion with respect to overall project success or failure based on those risks. The authors have predicted risks—not project failures based on those risks. They have begun to prioritize these success factors within the strategic planning phase with more accuracy, as the study discussed using prediction to derive appropriate budgets. However, Maghsoodi and Khalilzadeh (2018) did not cover success factors from past projects, nor did they use similar past projects to predict the value of these CSFs.

Costantino and Nonino (2015) adopted a more innovative approach in selecting projects based on CSFs, project manager experience, and competitions. They used 150 projects, their CSFs, and key performance indicators (KPI), to create their project implementation profile (PIP); they then used an artificial neural network (ANN) to assess projects before selection. A criterion is "a principle or standard by which anything can be judged," while a factor can be described as "any circumstance, fact, or influence which contributes to a result" (Lim and Mohamed, 1999). While their research focuses on selecting the best possible project for success, the research did not discuss false positives and the cost of project failures. The Constantino and Nonino (2015) model does not consider the value of saving time and money. As for future work, they acknowledged that a bigger sample is needed to generalize the result, which means subjective decisions have been made in selection of these projects, which added biases to their study. In this praxis, organizations establish which projects to pursue and how many projects to pursue each year. This process provides the decision maker with predictive failure with respect to possible savings if management terminates work on a given project and redirects resources to the next project. This praxis will demonstrate that, when project work stops in a timely manner, organizations could save a minimum of $100,000 that could be used for other or subsequent projects.

CSFs are also studied in the construction industry. Projects developing roads, bridges, and buildings are measured as a factor of a country's economic development. Alzahrani and Emsley (2013) used a survey based on CSFs in clusters and grouped these under an umbrella of nine categories. Since the construction industry has not adopted critical success criteria, it remains subjective to one's study. The CSFs used in the Alzahrani and Emsley study were a combination of qualitative and quantitative. The study also discussed project success from the contractor's perspective and the general project management prospective. The surveys elicited from customer, contractor, and consultant were then compared with an industry regulatory body, like the Occupational Safety and Health Administration Incidence Rate (OSHAIR), for health, safety, and quality matrices collected from the survey. Nevertheless, the accuracy of these measures, or lack thereof, were not assigned a dollar value, nor did the study calculate the cost of poor project planning and execution.

2.7. Project Failures in the IT Industry

Tesch, Kloppenborg, and Frolick (2007) assessed project risk factors in the IT industry. The assessment discovered that, throughout the years, the IT industry constantly sees systems failures.

Earlier software development data showed that only 34% of their projects would succeed and has gone to just 28% project success now (Standish Group, 2004). This could be due to unclear project scope, unqualified staff, project distractions, misunderstood vendor promises, and lack of proper support. There has been an ongoing trend in increased project failures for the information systems industry (Hughes, Rana, & Simintiras, 2017c). First, the authors (Hughes, Rana, & Simintiras, 2017c) evaluate what components can lead to project failure and examine how these major areas can impede project success. The authors then look into the facets which greatly effect project performance. They express issues found within the organizations in order to integrate change. Future research needs to delve into Factor Analysis, in order to find a quantitative technique to develop causal relationships between failure factors.

Due to continuous high failure rates for Information System (IS) projects, there is a greater need for further exploration. This praxis highlights gaps in research which need to be further studied to reduce the high rate of failures. This literature supports the intent of this praxis to present a new outlook for IS research in order to connect the gaps in literature.

IT projects continue to fail despite an abundance of literature found on risk mitigation. In an Asian logistics firm, Lim, Sia, and Yeow (2011) evaluated the route of back to back crises. Their studies highlight the barriers of building collectiveness in risk construction, the requirement for risk managers to check the influence of social structures and being prepared to deal with sudden changes when working on complex IT projects. McManus and Wood-Harper (2003) demonstrated that only 12.5% of IT projects (1 out of 8) are completely successful. They define failure as projects which fail to meet the original deadline, fail to follow cost, and fail to meet quality criteria by the owner. Although there is an increased awareness of failure, the IT industry continues to show increasing failures, which it then largely writes off as expenses. For instance, The European Union wrote off 142 billion Euros in 2004. Risks of failure remain high in projects generally, and even higher in technology projects (Mukerjee & Prasad, 2017). Success factors and reasons for failures are sparse. As IT consultants play a critical role, their perspective and definition of success is crucial. This dynamic of project success is a still untouched area in IT/project management literature.

For decades, practitioners have viewed product life cycles in three stages of development: first, the idea stage; second, the product development, design, and engineering stage; and third, the marketing stage (New and Schlacter, 1979). In order to create a more successful and integrated framework for a product cycle, research gaps should be bridged between these three stages. By doing so, project cycles could cease before advancing to later stages or becoming too expensive before they ultimately fail. It is especially critical for technology-based companies to remain mindful of these stages. The blank concept must be used early in the development cycle to predict project success or failure. Research shows that about 50% of projects fail because they were not planned with the entire project management process in mind. Without open communication and transparency, there is no visibility in the work cycle. McGrath and Martin (2017) refer to this problem as no “single version of the truth.” All stakeholders need to understand the long-term rewards of the immediate benefits of project management. *Information Week* found that the failure rate for technology projects is anywhere from 37% to 75% (Meshing, 2013). Drawing on two decades of experience, *Information Week* found four reasons IT projects fail as often as they do: technology Return on Investment (ROI) numbers are mostly fiction; ROI rarely drives technology investment decisions; long-term accountability is rare in technology; and finally, detailed plans are the enemy.

It is unfortunate, but very common, for projects to finish behind schedule, exceed their budget, and fail to meet their owner's expectations. Most often, these projects are failing long before the project manager completes the assignment for delivery. Aiyer, Rajkumar, and Havelka (2005) describe a process for recovering such projects from troubled situations. They outline a generic framework for recovering and rehabilitating troubled IS software development projects and details each of its four stages (recognition, immediate recovery, sustained recovery, maturity) and its twelve steps. The 15th IPMA World Congress reviewed paper proposals, analyzing their reasoning given for project failures. The UK Industrial Society showed that 77% of projects in the UK fail, while 83% fail in the US. Rudolf Burkhard, a senior consultant at Dupont Consulting Systems, argues that the rush for productivity leads to ineffective management, which then leads to a lot of multi-tasking and multiple needs for the same resources. Due to the demand for critical resources, all projects begin to compete for resources which become scarce. In the *International Management Review*, authors investigate the root causes of IT project failures. They categorize failures into six generic root causes based on two major areas: common factors of failure, and specific causes existing, which generalize the taxonomy. The study concludes that any type of IT project failures is related to one of six categories (Al-Ahmad, et al., 2009).

Project management must recognize early warning signs before projects fail. Early warning signs (EWSs) provide signals for project management to identify and are helpful for future projects. Keil and Montealegre (2014) recommend managers to ask themselves if any "red flags" appear serious enough to warrant project termination or significant redirection. The EWSs of IT project failures include lack of support, weak project management, minimum stakeholder involvement, lacking commitment from project team, members lack required expertise, and the subject matter experts are overscheduled. The six process-related EWSs of IT project failure are 1) lack of documented requirements; 2) no change control process; 3) ineffective schedule planning or management; 4) communication breakdown among stakeholders; 5) resources assigned to a higher priority project; and 6) no business case for the project (Kappelman, McKeeman, & Zhang, 2007).

2.8. Project Failures in the Construction Industry

Construction is one of the least safe industries in the world with life-threatening stories. The industry continuously struggles to find methods for safe, complete project deliverables. Tingshen, Kazemi, Wen, and Miao (2018) studied the crux of construction safety management failures in China by analyzing the status quo of safety management and the "last mile" problem. They defined the "last mile" problem as the failure to implement the extensive legal and regulatory systems on the construction site. These safety factors were extracted from a 34-item questionnaire. Through factor analysis and ranking correlation, the study found five human factors the most challenging. According to the National Bureau of Statistics (2015), mechanical injuries are one of five major types of accidents. Tingshen, Kazemi, Wen, and Miao (2018) proposed and tested a new safety management framework for the Wuhan-Shenzhen highway project. Through expert auditing, they confirmed that the framework could improve construction safety and bridge the "last mile." These cruxes of safety were collected from the survey but collected without historical record or monetary value. Since historical record and monetary value factored into the project management failures, they could be used to predict future project risks. Once consistently collected, these factors would become part of the organizational project and safety management.

The construction industry faces significant risks beyond those of the IT industry. Al Sabah (2014) uses Relative Importance Index (RII), Significance Score (SS), Risk Mapping Matrix (RMM), and Principal Component Analysis (PCA) to evaluate performance metrics. In an effort to lessen the chances of project failure, project managers need to create more effective risk mitigation plans.

Ansah, Sorooshian, Mustafa, and Oludapo (2017) created a risk priority number (RPN) to evaluate risks found in the construction industry. The RPN measured the number of occurrences along with the level of risk severity. Their research found that the most common risks include financial issues, lack of proper coordination by contractors, impractical requests for changes, defective work, incorrect materials management, poor communication, unobtainable resources, flawed construction site layout, inaccurate budgeting, and delays in getting approvals.

Due to increasing world globalization, cultural influence plays a large part in project success or failure (Nguyen and Watanabe, 2017). Despite the abundance of literature on why projects fail, there are ongoing studies to better understand what influence project success. Project critical success factors (CSFs) include project manager experience, types of work environments, company's objectives, and the entire project selection process. These CSFs are essential to detect project failures. Constantino, Di Gravio, and Nonino

(2015) propose a new way to assist project leads in order to evaluate projects during the early selection process. The authors use artificial neural network (ANN), along with CSFs, to assess the degree of a project's risks by evaluating the project manager's experience with past projects.

In another study, Amoatey and Hayivor (2017) researched the critical success factors (CSFs) for effective project stakeholder management in Ghana. They used data from questionnaires given to key project stakeholders for identifying and ranking CSFs. The top five CSFs identified by the authors were 1) communicating with and engaging stakeholders; 2) identifying stakeholders properly; 3) formulating a clear project mission statement; 4) keeping and promoting good relationships; and 5) analyzing stakeholder conflicts and coalitions. Their results indicated a need for future research. Iyer and Jha (2005) define critical success as "project manager confidence; top management support; project manager's coordinating and leadership skill; monitoring and feedback by the participants; coordination among project participants; owners' competence and favorable climatic conditions." (pg 1). The authors also examined, in detail, the cost performances of projects: "conflict among project participants; ignorance and lack of knowledge; presence of poor project specific attributes and nonexistence of cooperation; hostile socio economic and climatic condition; reluctance in timely decision; aggressive competition at tender stage; and short bid preparation time" (Iyer and Jha, 2005).

Critical success factors in the construction industry play a major role in the success and failure of their projects. Maghsoodi & Khalilzadeh (2018) created a questionnaire based on time, cost, quality, and safety, which they used as metrics to gauge the success and failure of projects. Professionals evaluated these factors, ranking the importance of each and assigning a percentage of approval. Fuzzy TOPSIS multiple criteria decision-making methods were used to rank the critical success factors of the construction projects. Boyton, Ayscough, Kaveri and Chiong (2015) examine the failures of business intelligence (BI) implementations to understand why projects fail and what actions can lead to project success. The study identified four areas of BI projects to measure success: 1) return on investment, 2) non-concrete measures, 3) project management measures, and 4) user satisfaction. The study concludes by providing insights into what factors contribute to the success of a BI implementation, and what factors contribute to its failure.

Brookfield, Fischbacher, Smith, Mohd Rahim, & Boussabaine (2014a) demonstrate how key risk management activities can be embedded within a carefully developed risk construct, how mitigation strategies are likely to be used in practice, and over which risk areas of the risk construct; and report a new finding concerning the strategic responses to technical failure in projects. Dos Santos, Flávio Roberto Souza, & Cabral (2008b) present a risk management tool based on two well-known sets of concepts: FMEA (Failure Mode and Effect Analysis) and PMBOK (Project Management Body of Knowledge). A study of the Brazilian mail service shows that the proposed model was largely successful because it identified and classified risks. Furthermore, the model helped to document the strategies and action plans needed to respond to these risks. Duris (2002) highlighted the ten primary reasons warehouse management systems failed: 1) unrealistic deadlines; 2) company politics; 3) automating a bad process; 4) trying to do too much at one time; and 5) management lost their commitment to the project. Research showing that large-scale project failures undermine a project team

’s ability to successfully deliver them is still in its infancy (Ellinas, Allan, Durugbo, & Johansson (2015). A single task failure can lead to a cascading process. A significant number of tasks induce no cascades, while a handful of tasks are capable of triggering surprisingly large cascades.

The mining industry also has similar project failure complications. Ferguson, Clinch, and Kean (2011) found project failures to occur regardless of whether deposits were underground or open pits. Ghiasi, Kaivan, Arzjani, and Arzjani (2017c) used a failure mode and effects analysis (FMEA) along with fuzzy risk priority number (FRPN) to find resolutions to prevent project completion delays and explore possible reasons which lead to the delays. They find a connection between early warning signs and project and performance measurement. They present an innovative way to used performance management to detect project delays.

A great proportion of projects in the IT industry involve a vast number of stakeholders with impractical expectations. Standing, Standing, and Kordt (2016) worked with project managers in reference to their attributional style and level of seniority. They used attribution theory explain project outcomes. In the *Engineering Management Journal*, Purvis, McCray, and Roberts (2004) focused on project management in order to attain the required project specifications in a timely manner. Biased specifications were proven to go towards project failure.

Jørgensen (2014a) found the probability of project failure to correlate with the actual number of failures. The study used a binary logistic regression model to find 74% project failures. Furthermore, Joslin and Müller (2016c) used contingency theory to find a positive relationship between project methodology and project success components.

Kealey, Protheroe, MacDonald, and Vulpe (2006b) highlight the challenge of performance improvement for global projects. The authors state "the most frequent causes of failure in international projects are: 1) failure of motivation; 2) failure of selecting of partners; 3) failure of setting and delivering on goals; 4) failure of setting realistic goals; 5) failure of the governance issue; 6) failure of consultation and consensus building; 7) failure of commitment and support; 8) failure of the project's political and socioeconomic environment; 9) failure of the challenge of culture; and 10) failure of trust" (Kealey et al., 2006b).

Kutsch, Browning, and Hall (2014c) found that new product development projects (NPD) have high risks especially because they are not able to stay within budget, meet the delivery deadlines, or achieve the desired project outcomes. According to agency theory, contracts between project managers and system developers are designed to lessen conflicts and more likely to lead to project success. Mahaney and Lederer (2011a) found that monitoring reduces privately held information which in turn increases the likelihood of that success. This leads researchers to believe that monitoring systems developers might work better than providing them with incentives for project success

All projects carry defined beginnings, but it is often unclear when projects should terminate. Even with the presence of risk, exactly when is project termination prudent or necessary? Marsh (2010a) described project termination to not just be the last part of the cycle but phased out from organizational procedures. Marthandam, Calvo-Amodio, and Ng (2014) present an economic model as a tool to guide the decision maker to determine when to terminate a project and gain insight if the termination would hurt the company's market value.

In today's world, firms compete in a fast "internet time." Park, Im, and Keil (2008a) discuss how IT projects fail due to responsibility as well as time urgency. Often times projects do not fail even though they experience obstacles and challenges. Parkes & Davern (2011) provide a comprehensible explanation for how project drivers interact with context. Project management techniques are used more frequently today as companies recognize the added benefits. Indeed, without sufficient training and procedures, projects are bound to fail.

Pinto and Kharbanda (1996) found reasons that led to project failure such as using new technology too soon, neglecting the project environment, not creating back-up options, and letting new ideas go stagnant.

Souder and Chakrabarti (1979) identify four categories within their collected data. They grouped challenges as threats. They defined *tactical opportunities* as project need factors. Planned schemes were sorted into *strategic expansions*, and *line protectors* represented developmental changes of existing product lines. The findings suggest that framing of variability for categorized risk factors in enterprise resource projects (ERP) is not necessarily culturally bound. The relationships between risk and success factors can impact the success of the ERP (Ojiako, Papadopoulos, Thumborisuthi, & Yang, 2012b).

Chapter 3: Methodology

3.1. Introduction

This chapter covers, in detail, how the problems of this praxis were studied and how the proposed hypotheses were validated. This chapter also contains several complex concepts, and for the reader to understand the thought process presented in the study, it is best to begin with a discussion of the data. Once the reader understands the data, the chapter will make sense of the data relationships, interpret the data elements and their significance, and lastly perform statistical analysis and interpretation of those results. Again, this praxis asserts that corporations collect project information such as performance indicators, they tend to use this information only to estimate the cost of future projects rather than as a way of predicting potential project failures. By not using performance indicators from prior projects to predict future project success or failure, organizations are not taking advantage of information that can be used to avoid spending time, money, and resources on projects that are likely to fail. If companies were able to accurately predict project failures, they could avoid incurring the opportunity cost of carrying out projects that should be cancelled early in their life cycle.

3.2. Research Questions and Hypothesis

The primary research questions this praxis presents is, how can organizations identify the potential project failure candidate from project performance indicators? In order to understand the research question and attempt to answer it, it is important to understand these project performance indicators, their origin, and the process of their measurement. Not all project performance indicators were adopted to measure and evaluate project health and success. Indeed, most were collected initially for the sole sake of record keeping—highlighting the first argument of this research, that organizations should develop plans to reuse project performance data while collecting and storing them. In other words, organizations should develop an intelligent data collection aligned with the successful achievement of organizational objectives. Although many organizations collect these project performance data for annual financial and investment decisions for reporting, they should also develop and deploy plans to utilize these data to gain a competitive edge in their respective industries.

Sample data of this praxis are found in **Appendix D**. Table 1 presents a data summary for various analyses and understanding. Table 2 collects the key sets of final project performance indicators, compiled after multiple evaluations and two stages of cleanups. Table 2 defines these performance indicators, which are then discussed in greater detail thereafter.

The hypotheses of this research are:

H1: Project success or failure can be forecasted from project performance indicators. Accept.

These statements were evaluated by multi-stage analysis. First removing all projects with significant missing data, then take projects that completed, lastly understand the weighted relationship between performance indicators to select a mix set of performance indicators to forecast accurate results. Sorting every project which failed, followed by a critical analysis of the project performance data elements, which concluded that all failed projects did not have a target end user. Because projects are tailored to satisfy end user requirements to meet the needs of customers, it is critical for projects to always focus on the target end user. After evaluation, the study revealed that all of the failed projects had not included a technical lead, which appears to be a driving factor critical to project success. A technical lead constructs and perceives a vision for how project elements work together to achieve the project's desired objective.

Furthermore, developing countries are more likely to have less restrictive regulatory and environmental laws, which makes overall international projects more cost effective than the domestic project. International labors are also more affordable than the domestic labor, hence per project, domestic project cost 50% more than the international projects. After examining the data, this study concluded that failed projects at a given phase cost more than projects that successfully passed, and project failed after gate one review cost more than the project failed after gate two review. More detailed findings are in Chapter 4 of this praxis under section *Trends in Data*.

3.3. Quantitative Research

This study is based on the project data from the research, development, proof-of-concept (POC), and prototype type projects of a large organization. Actual project data were used in this study hence there are a lot of incomplete and unique data which were removed to increase accuracy of the models. All projects irrespective of their sizes have similar approaches and attributes, but larger projects would have more variables and repeated activities than the smaller projects. Some of those critical details for these large projects were not recorded by the organization. Therefore, only project records with the most completed data used limited the selection of variables. While calculating the hours worked in each phase, the range of dates were converted into hours and called phase-hours, where any paid and unpaid time off were unknown and not taken in account. These are some of the limitations and biases in this quantitative research.

3.4. False Positive

False positive (FP) in this praxis studied from four different models and only the model with lowest FPR was recommended for use. Accuracies ranges from 91.67% to 100%, it includes False Negative (FN) and FP. Since the problem this praxis is addressing is to identify potential project failure candidate, then the lowest rate of FP would be utilized to select the right model. FPR ranges from 29.51% to 0% and the goal is to select the model with the lowest PFR or PFR closest to 0%. While lowest FPR rates represents model rightly predicting project failure candidate, the accuracy rate represents the total saving if the recommended model is used. Interpretation of the missing accuracy in this praxis means projects identified success or failure by the model inaccurately. FP for the prediction of a current project need to be weighted with stakeholders' expert opinions. It is recommended to run the models a few times with various data variables to better understand the outcome as it would show the percentage being predicted, and then factor in expert opinion based on the project's current health already known to make the project continuity decision.

3.5. Overview of Methodology

The project data used in this praxis came from a large organization, henceforth referred to as Company. In addition, all the data will be masked to maintain confidentiality. The projects from Company were used in coordination with the methodology in this praxis to predict project failure following the first two phases of a project life cycle while taking the weights of the other project performance indicators into consideration. The percentage of accuracy calculated by all four models used in this praxis ranges from 91.67% to 100%, which means, if organizations terminate a project by following the lowest accuracy model, organization could save a minimum of 91.67% and a maximum of 100% if the highest accuracy model is selected for the remaining cost of the project.

While the accuracy is important for overall saving, it is important to know what this accuracy consist of. Models' accuracy consists of True Positive (TP), False Positive (FP), True Negative, and False Negative (FN). TP and TN are the results of models' correct categorizations of the projects outcome. FP and FN, also known as Type I and Type II Errors, respectively. Since the objective of this praxis is to forecast failed project then it is important to investigate the False Positive Rate (FPR) or the Type I Error rate. FPR represents models' incorrect categorization of successful projects as failure. This is important because organization would lose money by cancelling projects that could be successful, as the selected model identified them as failure. Therefore, it is important to select a model, as recommended in this praxis, with the lowest FPR. Currently FPR between all four models ranges from 29.51% to 0%, and it is recommended to select a model with FPR close to 0%.

Similarly, False Negative Rate (FNR), known as Type II Error, represents models' incorrect categorization of failed projects as successful. Too many of these could drain organizational resources as organizations would complete projects that will ultimately fail. FNR between all four models ranges from 4.74% to 0%, and it is recommended to select a model with FNR close to 0%. Detailed results shown in Chapter 4 of this praxis in table 12.

Prediction represents accuracy of categorization by the models tried in this praxis. Scoring a higher percentage will denote models were able to accurately categorize past projects results used to train and test the models. These results of prediction percentages will be used to aid the decision to continue the given project through the final two phases, or the decision to end the project and save time, money, and resources.

These models output a true average for the total hours used in each of the four phases. Average time spent in Phase 1 (P1) is 1,204.63 hours, Phase 2 (P2) 1,521.38 hours, Phase 3 (P3) is 645.36 hours, and Phase 4 (P4) is 521.47 hours. Phases 3 and 4 of the project lifecycles—645.36 hours spent during Phase 3 and 521.47 hours spent during Phrase 4. When combined, about 1,166.83 hours of work at the rate of $255 per hour (the Company proprietary rate) could be saved if the project stopped before beginning work on Phases 3 and 4. These model as discussed earlier takes the weights of project performance indicators, as their combined weight outputs the accuracy of the prediction. In other words, based on the Company's past performance on similar projects, utilizing the recommended model Company will save 91.67% to 100% of the remaining project cost, which could lead to (1166.83*255*.9167 or) $ 272,756.43 to (1166.83*255*1 or) $297,541.65. The goal for this study was to save a minimum of $100,000 per project if stopped.

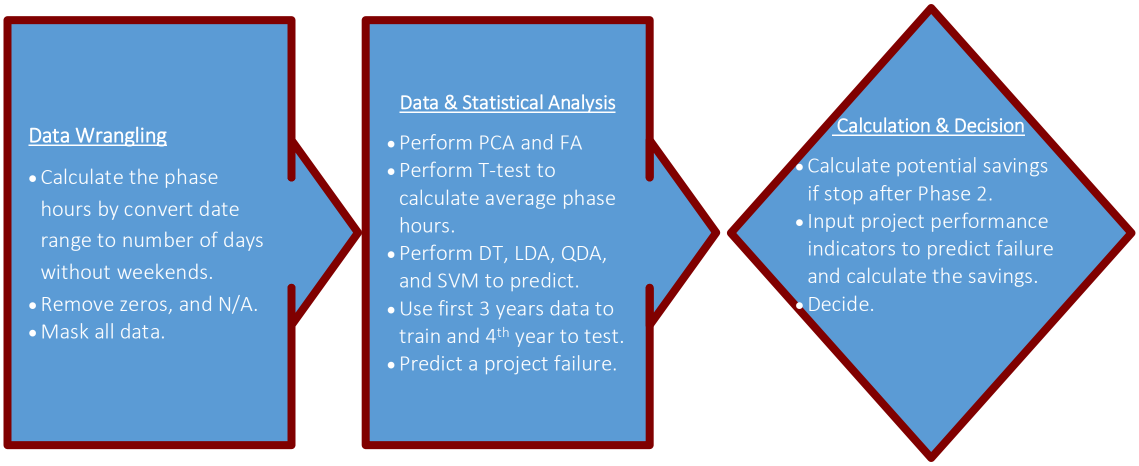

Figure 1: A graphical representation of the methodology used in this praxis.

Figure 1 above is a graphical representation of the methodology used in this praxis. It consists of three phases. The first phase is Data Wrangling, the second, Data and Statistical Analysis, and the third, Calculation and Decision. R-Studio will be used as the statistical analysis tool for this model, but there are other statistical analytics tools which could also be used to draw similar conclusions. Factor Analysis performed to understand relationships between project failure with other project performance indicators. T-test performed to find the true mean of each phase hours. DT, LDA, QDA, and SVM are performed to predict project failure. Calculations for potential savings performed to stop project work and save resources. The order in which these techniques and methodologies are used is a required process because each technique becomes the input of the next technique. Consequently, errors may occur if proper steps are not followed.

3.6. Data Wrangling

The Company recorded the data of over 21,648 projects with 79 data attributes, collected over the last 16 years. Part of this specific data wrangling process included cleaning up the data by removing mostly missing, unknown, and unique data attributes, without affecting the performance of the project. In this study, emphases are given to the project performance indicators including phase data to establish weighted relationships, as well as true average hours spent in these phases. The objective of this study is to predict failure of a project after the first two phases or to stop project work for phases 3 and 4 to save project resources. These resources can be redirected to the next project to give the next project an early start and to undertake more successful projects over time, while saving money by stopping the current project. There are many projects with incomplete project data and projects missing submission dates. Those projects are not included in this study. The goal of this study is to use a few hundred projects with submission dates and the most complete project variables. Projects without a recorded submission dates are impairing phase hours calculations and considered incomplete. The submission date is used to calculate Phase 4 hours. Once collected, projects with submission dates were grouped by year, and only years with a minimum of 25 or more completed projects were selected for this study. 4,753 projects completed with submission years between 2007 and 2018 were found. A sample of the actual past project data from the Company with some masking due to data propriety are found in **Appendix D**. Some column names displayed are abbreviated for the same reason, but their definitions, descriptions, data values, and the justifications for elimination for future use are listed under *Data Definition* section.

Table 1: Data Summary

Column Names		Data Summary						
1	YR	Min.:2007	1stQu.:2012	Median:2014	Mean:2014	3rdQu.:2016	Max.:2018	NA's:16895
2	Project.ID	Min.:8	1stQu.:4939	Median:8778	Mean:9335	3rdQu.:13941	Max.:19486	NA
3	Booking.ID	BKG-11888:1	BKG-11935:1	BKG-11936:1	BKG-11937:1	BKG-12016:1	BKG-12017:1	(Other): 21642
4	.	Min.:82333	1stQu.:103128	Median:104196	Mean:104042	3rdQu.:106785	Max.:111458	NA's:11720
5	UID.Booking..and.Budget.Type	102376BKG-49603:1	102376NotConverted-100356:1	102377BKG-48556:1	102377BKG-48557:1	102377BKG-48558:1	102377BKG-48559:1	(Other): 21642
6	X	Min.:42928	1stQu.:42937	Median:42955	Mean:42960	3rdQu.:42987	Max.:42996	NA's:21045
7	Booking.Type	Unknown:21640	Base/Initial:1	Base/Initial(GatedSeparately):1	OriginallyProposedOption/Increase(GatedwithBase):1	Other:1	PostAwardOption/Increase(GatedSeparately):1	(Other): 3

8	MA	Unknown:21610	ACT:1	ACT(IIS):1	AMS:1	AT:1	AUS:1	(Other):33
9	SubMA	Unknown:21567	ATCOTS:1	ATS:1	BD-BUSMGR:1	BUSMGR(ACT):1	BUSMGR(BD):1	(Other):76
10	Booking.Name	TEMPTVRBALANCE:254	Base:87	JPSProduct:16	TCSTrustedThinClient:15	SCRS(AMCOMEXPRESS):12	Base/Initial:10	(Other):21254
11	ERPM.ID	Min.:82333	1stQu.:103128	Median:104196	Mean:104042	3rdQu.:106785	Max.:111458	NA's:11720
12	Booking.ID.1	BKG-11888:1	BKG-11935:1	BKG-11936:1	BKG-11937:1	BKG-12016:1	BKG-12017:1	(Other):21642
13	Win_Loss	Incomplete:11106	Won:2339	InProgress-sister:2245	InProgress:1904	Incomplete-sister:1212	Lost:750	(Other):2092
14	Booking.Status	Dropped:6417	PotentialAward:4899	Terminated:3712	Cancelled:2523	Won:1708	Awarded:1256	(Other):1133
15	Opportunity.Status	Dropped:6417	InPursuit:4426	Won:3385	NoBid:2702	Cancelled:2523	Lost:994	(Other):1201
16	Passed.Failed	InPursuit:4426	Lost:10705	NoBid:2702	Protest:21	Unknown:409	Won:3385	NA
17	Last.Gate	Gate4:6010	Lead:4212	Pre-Gate:3316	Gate0:2898	Gate1:2525	Gate2:1715	(Other):972
18	Sales.Type	DCS:935	Dom:8645	FMS:250	N:10931	Unknown:48	Y:839	NA
19	TVR	Min.:0	1stQu.:2430	Median:10500	Mean:91769	3rdQu.:47550	Max.:300000000	NA's:1513
20	G1...PGO	Min.:1.00	1stQu.:50.00	Median:75.00	Mean:71.29	3rdQu.:90.00	Max.:100.00	NA's:1106
21	G2..PWIN	Min.:1.0	1stQu.:45.0	Median:50.0	Mean:63.2	3rdQu.:95.0	Max.:100.0	NA's:1110
22	G1...PGO.YN	No:1106	Yes:20542	NA	NA	NA	NA	NA
23	G2..PWIN.YN	No:1110	Yes:20538	NA	NA	NA	NA	NA
24	RFP.Issue	No:2111	Yes:19537	NA	NA	NA	NA	NA
25	Contract.Award.Date	No:896	Yes:20752	NA	NA	NA	NA	NA
26	Opp..Type	Non-Routine:525	Routine:19559	Strategic:127	TBD:73	Unknown:1364	NA	NA
27	Comp..Type	New:5814	Competitive:NewProgram:4764	Recompete:1910	SoleSource:Other:1821	SoleSource:1662	Competitive:TakeAway:1473	(Other):4204
28	RTN.Role	JV:14	Prime:14118	Sub:1868	Sub(IOT):423	Sub(orIOT):3316	TBD:695	Unknown:1214
29	Prime	Raytheon:14600	Unknown:1916	PraxisEngineeringTechnologiesInc.:1199	LockheedMartin:760	TBD:265	RTN(IDS):179	(Other):2729
30	Contract.Type	Unknown:8123	CPFF:3494	TBD:2935	FFP:2591	CPAF:1906	T&M:683	(Other):1916
31	Project.Type	INT:1894	NTL:19754	NA	NA	NA	NA	NA
32	End.User.L1.YN	No:502	Yes:21146	NA	NA	NA	NA	NA
33	End.User.L1	ClassifiedL1:2953	ProprietaryL1:1593	USAF:1342	Raytheon:764	USArmy:691	NSA:675	(Other):13630
34	Sub.Project.Type	Classified:2999	USAF:2661	Proprietary:1942	USAR:1721	USNA:1597	ONTL:924	(Other):9804
35	PL_G1	No:15843	Yes:5805	NA	NA	NA	NA	NA
36	PL_G2	No:17248	Yes:4400	NA	NA	NA	NA	NA
37	PL_G3	No:17213	Yes:4435	NA	NA	NA	NA	NA
38	PL_G4	No:16365	Yes:5283	NA	NA	NA	NA	NA
39	Proposal.Submitted	No:16883	Yes:4765	NA	NA	NA	NA	NA
40	Proposal.Due	No:4555	Yes:17093	NA	NA	NA	NA	NA
41	P1	Min.:5.71	1stQu.:674.05	Median:1525.71	Mean:2659.30	3rdQu.:3560.00	Max.:23937.12	NA's:17184
42	P2	Min.:5.71	1stQu.:480.00	Median:1417.14	Mean:2855.31	3rdQu.:3468.57	Max.:26428.55	NA's:18123
43	P3	Min.:5.71	1stQu.:97.14	Median:171.43	Mean:456.84	3rdQu.:365.71	Max.:23748.55	NA's:17734
44	P4	Min.:1.19	1stQu.:17.14	Median:51.43	Mean:616.49	3rdQu.:251.43	Max.:231096.91	NA's:17417
45	THC	Min.:1.19	1stQu.:360.00	Median:1900.00	Mean:3208.15	3rdQu.:4407.14	Max.:231096.91	NA's:13440
46	BH	No:21093	Yes:555	NA	NA	NA	NA	NA
47	BT1	No:21209	Yes:439	NA	NA	NA	NA	NA
48	BT2	No:21257	Yes:391	NA	NA	NA	NA	NA
49	Pink	No:21034	Yes:614	NA	NA	NA	NA	NA
50	Red	No:20914	Yes:734	NA	NA	NA	NA	NA
51	Gold	No:21091	Yes:557	NA	NA	NA	NA	NA
52	BD.Lead	No:20980	Yes:668	NA	NA	NA	NA	NA
53	Capture.Manager	No:20293	Yes:1355	NA	NA	NA	NA	NA
54	Proposal.Manager	No:20565	Yes:1083	NA	NA	NA	NA	NA
55	CE	No:20564	Yes:1084	NA	NA	NA	NA	NA
56	PE	No:21131	Yes:517	NA	NA	NA	NA	NA
57	BD.Lead.YN	Unknown:20980	AaronPlamondon:1	AaronRenner:1	AaronWatts:1	AdamP.SMITH(1106975):1	AdrienneRivera(HITP6399):1	(Other):663
58	Capture.Manager.YN	Unknown:20293	AaronAllen:1	AaronPlamondon:1	AaronRenner:1	AaronWatts:1	AdamDavidson:1	(Other):1350
59	Proposal.Manager.YN	Unknown:20565	AaronAllen:1	AaronPlamondon:1	AaronRenner:1	AdamFisher:1	AdamFraser:1	(Other):1078

60	CE.YN	Unknown:20564	AaronAllen:1	AaronRenner:1	AdamFraser:1	AdrienneRivera(HITP6399):1	AimeeDavid:1	(Other):1079
61	PE.YN	Unknown:21131	AaronRenner:1	AdamSmith(1035647):1	AkhileshRathore(HITJ3640):1	AlFelts(HMGN5462):1	AlJost(68059):1	(Other):512
62	Booking.Value...K.	Min.:0	1stQu.:1000	Median:4000	Mean:35462	3rdQu.:12000	Max.:300000000	NA's:3082
63	PTW...K.	Min.:5700	1stQu.:27775	Median:62000	Mean:112448	3rdQu.:143975	Max.:1445100	NA's:21576
64	RTN.Bid...K.	Min.:5900	1stQu.:26805	Median:72404	Mean:116704	3rdQu.:139363	Max.:1422100	NA's:21576
65	Bid.Pwin	Min.:0.240	1stQu.:0.555	Median:0.745	Mean:0.701	3rdQu.:0.840	Max.:0.990	NA's:21576
66	Gate.4.Rate	Min.:0.030	1stQu.:0.070	Median:0.090	Mean:0.088	3rdQu.:0.108	Max.:0.130	NA's:21626
67	Current.Booking.Rate	Min.:0.030	1stQu.:0.050	Median:0.070	Mean:0.082	3rdQu.:0.108	Max.:0.160	NA's:21626
68	EAC.Period	Q12016:1	Q12017:8	Q32016:4	Q42016:9	Unknown:21626	NA	NA
69	Competitive.Advantage.Rating	Adv:31	Disadv:18	Neut:39	Unknown:21560	NA	NA	NA
70	X.1	,0,0,0:17252	,0,0,3:40	,0,2,0:2254	,0,2,3:1	,1,0,0:1283	,1,2,0:818	NA
71	Total.B.P.and.Selling.Hours	Min.:0.5	1stQu.:58.0	Median:191.2	Mean:1149.6	3rdQu.:690.9	Max.:73091.7	NA's:17303
72	B.P.Hours	Min.:0.02	1stQu.:63.27	Median:198.27	Mean:901.65	3rdQu.:622.03	Max.:43643.80	NA's:19552
73	Selling.Hours	Min.:0.5	1stQu.:52.1	Median:183.1	Mean:1012.5	3rdQu.:725.5	Max.:73091.7	NA's:18581
74	IRAD.Hours	Min.:191.1	1stQu.:2376.1	Median:3833.0	Mean:5316.0	3rdQu.:5410.1	Max.:33867.4	NA's:21607
75	Total.BOE.Hours.for.Completed.Proposals..Last.Gate...4.	Min.:2.00	1stQu.:54.17	Median:209.03	Mean:525.34	3rdQu.:553.47	Max.:4128.07	NA's:21590
76	Total.ENG.BOE.Hours.for.Completed.Proposals..Last.Gate...4.	Min.:2.0	1stQu.:60.1	Median:266.5	Mean:448.1	3rdQu.:519.0	Max.:2705.8	NA's:21599
77	Total.ENG.B.P.Hours.for.Completed.Proposals..Last.Gate...4.	Min.:0.60	1stQu.:29.12	Median:91.00	Mean:503.30	3rdQu.:326.85	Max.:12872.60	NA's:21249
78	TVR.Efficiency	Min.:0.0	1stQu.:7.7	Median:30.4	Mean:1041.3	3rdQu.:166.6	Max.:325000.0	NA's:18089
79	Factored.TVR.Efficiency	Min.:0.01	1stQu.:3.88	Median:14.50	Mean:398.70	3rdQu.:67.03	Max.:136521.74	NA's:18139
80	Competitors	Unknown:21448	None:22	LM:15	NotFound:12	Many:4	NG:4	(Other):143
81	Cust.Engmnt..fr.Gate.or.BH.Pkgs.	NF:11	Unknown:21597	Yes:40	NA	NA	NA	NA
82	Cust.Engmnt.Freq..fr.Gate.or.BH.Pkgs.	<1:7	>4:7	1:06:00AM	>1:3	4:03:00AM	(Other):3	NA's:21619
83	Potential.Risk.Cost.impact...K...fr.Gate.pkg.	Min.:10	1stQu.:457	Median:883	Mean:3085	3rdQu.:3599	Max.:15819	NA's:21611
84	Risk.Prob...fr.Gate.pkg.	Min.:0.080	1stQu.:0.150	Median:0.220	Mean:0.221	3rdQu.:0.280	Max.:0.460	NA's:21611
85	Factored.Risk.Cost.Impact..fr.Gate.pkg.	Min.:2.0	1stQu.:86.0	Median:246.0	Mean:639.6	3rdQu.:775.0	Max.:3335.0	NA's:21611
86	Potential.Opp.Cost.impact...K.	Min.:10	1stQu.:283	Median:1000	Mean:4240	3rdQu.:3119	Max.:64164	NA's:21611
87	Opp.prob...fr.Gate.pkg.	Min.:0.100	1stQu.:0.180	Median:0.210	Mean:0.252	3rdQu.:0.300	Max.:0.690	NA's:21611
88	Factored.Opp.cost.Impact..fr.Gate.pkg.	Min.:1	1stQu.:68	Median:292	Mean:1333	3rdQu.:915	Max.:28166	NA's:21611
89	LM	No:21588	Yes:60	NA	NA	NA	NA	NA
90	NG	No:21619	Yes:29	NA	NA	NA	NA	NA
91	BAE	No:21618	Yes:30	NA	NA	NA	NA	NA
92	SAIC	No:21637	Yes:11	NA	NA	NA	NA	NA
93	GD	No:21623	Yes:25	NA	NA	NA	NA	NA
94	IBM	No:21643	Yes:5	NA	NA	NA	NA	NA
95	Boeing	No:21636	Yes:12	NA	NA	NA	NA	NA
96	BAH	No:21628	Yes:20	NA	NA	NA	NA	NA

Table 1 includes the data summary prior to any data cleanup. An additional 17 columns were added to the 79 original data columns due to calculations of the phase hours, gate reviews, technical leads, separating international and domestic projects, and grouping by projects lost, and years, for a total of 96 project attributes or columns of data per project.

These 96 attributes then decreased to 29 after removing the unknown, n/a, missing data values, and uniqueness in data attributes.

3.6.1. Data Definitions

The following is a list of project data variables, their definitions, an explanation for their selections or otherwise, and their conversions into numeric values for statistical analysis, when necessary:

1. **YR: Year.** The year in which the project was completed. There were 21,648 projects undertaken between year 2007 and 2017 and only 4,753 projects completed. This is a unique data attribute and needed for the next round of analysis. Data values for this column are 2007, 2008, and 2009, etc. Since these are numeric values, they were used as is.
2. **Project ID (PID): Project Identifier.** Generated by corporate research and development (R&D) to utilize corporate annual budget allocated for R&D projects. It is unique to each project, sequentially assigned by the company to track project activities, progress, resources, expenses, and funding. Since this is unique to each project, it would not be used in next round of analysis. Data value for this column are the incremental numbers 1, 2, and 3, etc. Since these are numeric values, they were uses as is.
3. **Booking ID (BID):** Generated by the mission area for the project to associate with the sub mission area funding. Since this is unique to each project, it would not be used in next round of analysis. Data values for this column are BKG for booking followed by a five-digit number representing funding string, BKG-12345, BKG-12346, and BKG-12347, and so on. Since these are not all numeric

values, they were converted into numbers by replacing all identical strings with a numeric value in increments of 1, 2, and 3, etc. For instance, all "BKG-12345" would be replaced with numeric value 1, and "BKG-12346" would be replaced with numeric value of 2, etc.

4. **[Unknown Column Name] (UNK1): Budget Type**. Internal to the Company to identify the source of funding for the project, which then became the prefix of the Booking ID to create budget type. Since this is unique to each project and missing data value for 11,720 projects out of 21,648, it would not be used in next round of analysis. Data values for this column are all numeric, and five-digit values 12345, 12346, and 12347, etc. Since these are all numeric values, they were used as is.
5. **UID-Booking, and Budget Type (BID2): Budget Identifier**. Unique to each project, sequentially assigned by the company to track funding, combined with division and subdivision funding buckets. Since this is unique to each project, it would not be used in the next round of analysis. Data values for this column are alpha numeric, ranging from twelve to fifteen digits, like 12345ABC-12345, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical strings with a numeric value in increments of 1, 2, and 3, etc. For instance, all "12345ABC-12345" would be replaced with numeric value 1, and "12346ABC-12346" would be replaced with numeric value of 2, etc.
6. **[Unknown Column Name] (UNK2): A date with undetermined representation**. There are 21,045 project missing data value out of 21,648, therefore this data attribute would not be used in the next round of analysis. Data values for this column are alpha numeric: date with month, day, year, hour, and

minutes, like 11/11/2011 11:11am, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical strings with a numeric value in increments of 1, 2, and 3, etc. For instance, all "11/11/2011 11:11am" would be replaced with numeric value 1, and "11/11/2011 11:11pm" would be replaced with numeric value of 2, and so forth.

7. **Booking Type (BT): Project type**. Based on the idea of what is being built and the targeted end user these projects categorized. It ranges from an add-on feature, enhanced functionalities, to new product or sub-products. If the product and service were new and not developed before, then it simply received a category of new, or otherwise it inherited a name from the current project. This would be used in next round of analysis to study relationships with other data variables. Data values for this column are strings of text with wild characters, like "initial/base", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical strings with a numeric value in increments of 1, 2, and 3, etc. For instance, all "12345ABC-12345" would be replaced with numeric value 1, and "12346ABC-12346" would be replaced with numeric value of 2, etc.
8. **MA: Mission Area**. Internal to the company based on the group who provided the idea for funding and oversight of the project. This would be used in the next round of analysis to study relationships with other data variables. Data values for this column are strings of text with wild characters, like "MSM" or "IIS", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical strings with a numeric value in increments of 1, 2, and 3,

etc. For instance, all “12345ABC-12345” would be replaced with numeric value 1, and “12346ABC-12346” would be replaced with numeric value of 2, etc.

9. **SubMA (SMA): Sub-Mission Area**. Internal to the Company, one level lower than the MA. This would be used in the next round of analysis to study relationships with other data variables. Data values for this column are strings of text with wild characters, like “MSM-IRAD” or “IIS-ICP”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical strings with a numeric value in increments of 1, 2, and 3, etc. For instance, all “MSM-IRAD” would be replaced with numeric value of 1, and “IIS-ICP” would be replaced with numeric value of 2, etc.
10. **Booking Name (BN): Project name**. Internal to the Company; adopted based on the mission area, product, service, and targeted end user. It is unique and would not be used in the next round of analysis. Data values for this column are strings of text with wild characters, like “GMA – KSA Base” or “Base Year -EICCS”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical strings with numeric values in increments of 1, 2, and 3, etc. For instance, all “GMA – KSA Base” would be replaced with numeric value of 1, and “Base Year -EICCS” would be replaced with numeric value of 2, etc.
11. **ERPM ID (EID): Enterprise resource planning and management**. Internal to the company, which is identical to budget type for most projects. This would not be used in the next round of analysis. Data values for this column are five-digit numbers, like 12345, or 12346, etc. Since these are all numeric values, they would be used as is.

12. **Booking ID (BID3): Duplicate column of information of column 2 above**. This would not be used in the next round of analysis. Data values for this column are BKG for booking followed by a five-digit number representing funding string, BKG-12345, BKG-12346, and BKG-12347, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical strings with a numeric value in increments of 1, 2, and 3, etc. For instance, all "BKG-12345" would be replaced with numeric value 1, and "BKG-12346" would be replaced with numeric value of 2, etc.
13. **WL: Win_Loss**. Internal assessment prior to customer feedback, somewhat like booking status and opportunity status. Since these are internal assessments of the project health, this would be used for the next round of analysis. Data values for this column are strings of text with wild characters, like "Lost" or "Won", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical strings with a numeric value in increments of 1, 2, and 3, etc. For instance, all "Lost" would be replaced with numeric value 1, and "Won" would be replaced with numeric value of 2, etc.
14. **Booking Status (BS)**: Internal assessment based on potential customer feedback. Since these are internal assessments of project health, they would be used for the next round of analysis. Data values for this column are strings of text, like "Cancelled" or "Dropped," etc. Since these are not all numeric values, they were converted into numbers by replacing all identical strings with a numeric value in increments of 1, 2, and 3, etc. For instance, all "Cancelled" would be replaced

with numeric value 1, and “Dropped” would be replaced with numeric value of 2, etc.

15. **Opportunity Status (OS):** Final outcome of the project recorded for accountability, reporting, and future projects. Success if there is a buyer for the product and service. Failure if otherwise. Since this is the final outcome of the project, this would be used for the next round of analysis. Data values for this column are strings of text with wild characters, like “Lost” or “Won”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical strings with a numeric value in increments of 1, 2, and 3, etc. For instance, all “Lost” would be replaced with numeric value 1, and “Won” would be replaced with numeric value of 2, etc.
16. **Project pass fail (PF)**: Created by grouping projects that customer cancelled, company cancelled, company dropped, and lost into failed and won project as passed. There were 10,705 projects failed, 3,385 projects passed, and 409 projects with unknown status.
17. **Last Gate: Last phase (LP)**. Last phase recorded for project work, out of 4 phases internal to the Company project. This would be used for the next round of analysis. Data values for this column are alpha numeric, like “Gate 1” or “Gate 2”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical strings with a numeric value in increments of 1, 2, and 3, etc. For instance, all “Gate 1” would be replaced with numeric value 1, and “Gate 2” would be replaced with numeric value of 2, etc.

18. **Sales Type (ST)**: Whether the potential customer of the project is domestic, foreign, or internal. Since is identifies potential targeted customer for the project, this would be used for the next round of analysis. Data values for this column are strings of text, like “Dom” or “FMS”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “Dom” would be replaced with numeric value 1, and “FMS” would be replaced with numeric value of 2, etc.
19. **TVR (TV): Total value in return**. Internal to the company, a forecasted value in return for the project work. Since this defines potential value expected from the project, this would be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
20. **PGO (PG): Probability of go**. Internal to the company, gate 1 review after Phase 1 work; every project is evaluated for probability of go or no-go, if the project work should continue. Standard evaluation form is used for this gate evaluation of the project health – see **Appendix B**. This would be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
21. **PWIN (PW): Probability of win.** Internal assessment of the project, gate 2 review after phase 2 work would yield a sale or attract a potential customer. Standard evaluation form is used for this gate evaluation of the project health –

see **Appendix C**. Since this defines likelihood of potential customer, this would be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.

22. **PGO YN (PG12):** Original value is from 1% to 100%, see #20 and #21. This column was created to give data a different meaning of yes or no to reflect if this review for a project was completed or not. Therefore, the data values for this column are strings of text, like "Yes" or "No". Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, and 2. For instance, all "Yes" would be replaced with numeric value 1, and "No" would be replaced with numeric value of 2.
23. **PWIN YN (PW12):** Original value is from 1% to 100%, see #20 and #21. This column was created to give data a different meaning of yes or no to reflect if this review for a project was completed or not. Therefore, the data values for this column are strings of text, like "Yes" or "No". Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, and 2. For instance, all "Yes" would be replaced with numeric value 1, and "No" would be replaced with numeric value of 2.
24. **RFP Issue (RI): Request for proposal dat**e. If there is an official RFP issued for the project is being worked on. Usually a product is ready before RFP is issued if the customer challenge is known to the company. This would not be used for the

next round of analysis. Data values for this column are all alpha numeric values with wild characters, like “1/29/2014 6:00:00 AM” or “1/29/2014 6:00:00 PM”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “1/29/2014 6:00:00 AM” would be replaced with numeric value 1, and “1/29/2014 6:00:00 PM” would be replaced with numeric value of 2, etc.

25. **Contract Award Date (CA): Buyer secured.** An actual purchase order from a buyer for the product or service. This would not be used for the next round of analysis. Data values for this column are all alpha numeric values with wild characters, like “1/29/2014 6:00:00 AM” or “1/29/2014 6:00:00 PM”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “1/29/2014 6:00:00 AM” would be replaced with numeric value 1, and “1/29/2014 6:00:00 PM” would be replaced with numeric value of 2, etc.
26. **Opp. Type (OT): Opportunity typ**e. Does the project bring a strategic advantage to the organization or just a routine pursuit of new technologies? This would be used for the next round of analysis. Data values for this column are strings of text with wild characters, like “Routine” or “Non-Routine”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “Routine” would be replaced with numeric value 1, and “Non-Routine” would be replaced with numeric value of 2, etc.

27. **Comp. Type (CMT): Competition type**. How does the project aid our current customer is it an enhancement to our sole-source project, is it to give us competitive edge on future re-competition or is this a completely new pursuit? This would be used for the next round of analysis. Data values for this column are strings of text with wild characters, like “Competitive” or “New”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “Competitive” would be replaced with numeric value 1, and “New” would be replaced with numeric value of 2, etc.
28. **RTN Role (RR): Organizational/Company role**. Would the future work be in partnership with existing technologies or some other organizations, or it would be our technologies? This would be used for the next round of analysis. Data values for this column are strings of text with wild characters, like “Prime” or “Sub (or IOT)”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “Prime” would be replaced with numeric value 1, and “Sub (or IOT)” would be replaced with numeric value of 2, etc.
29. **Prime (PRM): Primary organization**. Who own the current technology and if our work can either partner with them or if we can run our own. This would be used for the next round of analysis. Data values for this column are alpha numeric, like “21CT” or “22nd Century”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “21CT” would be

replaced with numeric value 1, and "22nd Century" would be replaced with numeric value of 2, etc.

30. **Contract Type (CNT): Based on federal government contracting type**. Time and Money (T&M), Firm Fixed Price (FFP), and Cost-Plus Fixed Fee (CPFF). This would be used for the next round of analysis. Data values for this column are strings of text with wild characters, like "T&M" or "FFP/Rates", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all "T&M" would be replaced with numeric value 1, and "FFP/Rates" would be replaced with numeric value of 2, etc.
31. **Project Type (PT)**: This column was created to identify and separate international projects from domestic projects. There are 1,894 international and 19,754 domestic projects recorded.
32. **End User L1 (EU): Targeted end user of the project**. Who would use this product or service and who is being targeted by the subject project. This would be used for the next round of analysis. Data values for this column are alpha numeric, like "2021 LLC" or "211.ORG", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all "2021 LLC" would be replaced with numeric value 1, and "211.ORG" would be replaced with numeric value of 2, etc.
33. **End User L1 YN (EU12):** Original values are the names of the targeted user of a project. This column was created to give data a different meaning of yes or no to

reflect if there is a targeted project user or not. Therefore, the data values for this column are strings of text, like “Yes” or “No”. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, and 2. For instance, all “Yes” would be replaced with numeric value 1, and “No” would be replaced with numeric value of 2.

34. **Subproject Type (SPT):** This column was created to distinguish subset of projects types like classified, air force, navy, army, KSA, and UK, etc.
35. **PL_G1 (G1): Gate one.** Internal to the organization and signifies the start date for gate one. Every gate has certain constrains and checks applies to every project as it successfully moves through these gates. This would not be used for the next round of analysis. Data values for this column are all alpha numeric values with wild characters, like “1/29/2014 6:00:00 AM” or “1/29/2014 6:00:00 PM”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “1/29/2014 6:00:00 AM” would be replaced with numeric value 1, and “1/29/2014 6:00:00 PM” would be replaced with numeric value of 2, etc.
36. **PL_G2 (G2): Gate two**. Internal to the organization and signifies the start date for gate two. This would not be used for the next round of analysis. Data values for this column are all alpha numeric values with wild characters, like “1/29/2014 6:00:00 AM” or “1/29/2014 6:00:00 PM”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “1/29/2014

6:00:00 AM" would be replaced with numeric value 1, and "1/29/2014 6:00:00 PM" would be replaced with numeric value of 2, etc.

37. **PL_G3 (G3): Gate three**. Internal to the organization and signifies the start date for gate three. This would not be used for the next round of analysis. Data values for this column are all alpha numeric values with wild characters, like "1/29/2014 6:00:00 AM" or "1/29/2014 6:00:00 PM", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all "1/29/2014 6:00:00 AM" would be replaced with numeric value 1, and "1/29/2014 6:00:00 PM" would be replaced with numeric value of 2, etc.
38. **PL_G4 (G4): Gate four**. Internal to the organization and signifies the start date for gate four. This would not be used for the next round of analysis. Data values for this column are all alpha numeric values with wild characters, like "1/29/2014 6:00:00 AM" or "1/29/2014 6:00:00 PM", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all "1/29/2014 6:00:00 AM" would be replaced with numeric value 1, and "1/29/2014 6:00:00 PM" would be replaced with numeric value of 2, etc.
39. **Proposal Submitted (PS): Proposal submission date.** Internal to the organization when the proposal for the project was submitted. This would not be used for the next round of analysis. Data values for this column are all alpha numeric values with wild characters, like "1/29/2014 6:00:00 AM" or "1/29/2014 6:00:00 PM", etc. Since these are not all numeric values, they were converted into

numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “1/29/2014 6:00:00 AM” would be replaced with numeric value 1, and “1/29/2014 6:00:00 PM” would be replaced with numeric value of 2, etc.

40. **Proposal Due (PD): Proposal due date**. Internal to the organization, from the initial ideation a date is set for the proposal to be submitted for funding and other internal logistical requirements. This would not be used for the next round of analysis. Data values for this column are all alpha numeric values with wild characters, like “1/29/2014 6:00:00 AM” or “1/29/2014 6:00:00 PM”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “1/29/2014 6:00:00 AM” would be replaced with numeric value 1, and “1/29/2014 6:00:00 PM” would be replaced with numeric value of 2, etc.
41. **P1: Phase 1 hours**. Calculated by subtracting gate one start date from gate two starts date and multiplied by 5.7 work hours per day. If there was no gate one start date, then Phase 1 got zero hours. If a subsequent gate start date is missing, then the next gate start date was used to calculate the Phase 1 hours. If there were no subsequent gate start date, then the submission date was used to calculate the Phase 1 hours. This would be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.

42. **P2: Phase 2 hours**. Calculated by subtracting gate two start date from gate three starts date and multiplied by 5.7 work hours per day. If there was no gate two start date, then Phase 2 got zero hours. If a subsequent gate start date is missing, then the next gate start date was used to calculate the Phase 2 hours. If there were no subsequent gate start date, then the submission date was used to calculate the Phase 2 hours. This would be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
43. **P3: Phase 3 hours.** Calculated by subtracting gate three start date from gate four starts date and multiplied by 5.7 work hours per day. If there was no gate three start date, then Phase 3 got zero hours. If a subsequent gate start date is missing, then the next gate start date was used to calculate the Phase 3 hours. If there were no subsequent gate start date, then the submission date was used to calculate the Phase 3 hours. This would be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
44. **P4: Phase 4 hour**s. Calculated by subtracting gate four start date from submission date and multiplied by 5.7 work hours per day. If there was no gate four start date, then Phase 4 got zero hours. Since there are no subsequent gate then the submission date was used to calculate the Phase 4 hours. This would be used for the next round of analysis. Data values for this column are numeric and

therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.

45. **THC: Total hours calculated**. Represents the total numbers of hours spent on the project by adding all four Phase hours. This would be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.

46. **BH: Blue team**. Part of organizational business development team receives the project briefing to assess if the project has a potential customer and if it can be monetized. This would not be used for the next round of analysis. Data values for this column are all alpha numeric values with wild characters, like “1/29/2014 6:00:00 AM” or “1/29/2014 6:00:00 PM”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “1/29/2014 6:00:00 AM” would be replaced with numeric value 1, and “1/29/2014 6:00:00 PM” would be replaced with numeric value of 2, etc.

47. **BT1: Blue team one**. Part of the organizational business development team receives the project briefing to review the end user requirements and possible applicability. This would not be used for the next round of analysis. Data values for this column are all alpha numeric values with wild characters, like “1/29/2014 6:00:00 AM” or “1/29/2014 6:00:00 PM”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “1/29/2014

6:00:00 AM" would be replaced with numeric value 1, and "1/29/2014 6:00:00 PM" would be replaced with numeric value of 2, etc.

48. **BT2: Blue team two**. Part of the organizational business development team receives the project briefing to prepare sells proposals for the end user and possible buyer. This would not be used for the next round of analysis. Data values for this column are all alpha numeric values with wild characters, like "1/29/2014 6:00:00 AM" or "1/29/2014 6:00:00 PM", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all "1/29/2014 6:00:00 AM" would be replaced with numeric value 1, and "1/29/2014 6:00:00 PM" would be replaced with numeric value of 2, etc.
49. **Pink (PNK): Pink team**. Part of the organizational business development team receives the project briefing and responsible to complete 60-70 percent of the project proposal documentations. This would not be used for the next round of analysis. Data values for this column are all alpha numeric values with wild characters, like "1/29/2014 6:00:00 AM" or "1/29/2014 6:00:00 PM", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all "1/29/2014 6:00:00 AM" would be replaced with numeric value 1, and "1/29/2014 6:00:00 PM" would be replaced with numeric value of 2, etc.
50. **Red: Red team**. Part of the organizational business development team receives the project briefing and responsible to complete 80-90 percent of the project proposal documentations. This would not be used for the next round of analysis.

Data values for this column are all alpha numeric values with wild characters, like "1/29/2014 6:00:00 AM" or "1/29/2014 6:00:00 PM", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all "1/29/2014 6:00:00 AM" would be replaced with numeric value 1, and "1/29/2014 6:00:00 PM" would be replaced with numeric value of 2, etc.

51. **Gold (GLD): Gold team**. Part of the organizational business development team receives the project briefing and responsible to review all pricing, project documentations, and present to end users and customers. This would not be used for the next round of analysis. Data values for this column are all alpha numeric values with wild characters, like "1/29/2014 6:00:00 AM" or "1/29/2014 6:00:00 PM", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all "1/29/2014 6:00:00 AM" would be replaced with numeric value 1, and "1/29/2014 6:00:00 PM" would be replaced with numeric value of 2, etc.
52. **BD Lead (BDL): Business development lead**. Responsible for supervising the proposal development activities when from the initial briefing to making a sell. This would be used for the next round of analysis. Data values for this column are alpha numeric with wild characters, like "Brad Long (1061671)" or "Sunshine Baker (1110511)", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all "Brad Long (1061671)" would be

replaced with numeric value 1, and “Sunshine Baker (1110511)” would be replaced with numeric value of 2, etc.

53. **Capture Manager (CM)**: Provides content for the project based on the possible end user and buyer. This would be used for the next round of analysis. Data values for this column are alpha numeric with wild characters, like “Brad Long (1061671)” or “Sunshine Baker (1110511)”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “Brad Long (1061671)” would be replaced with numeric value 1, and “Sunshine Baker (1110511)” would be replaced with numeric value of 2, etc.
54. **PM: Proposal manager**. Responsible to supervise overall proposal activities, logistics, and schedule. This would be used for the next round of analysis. Data values for this column are alpha numeric with wild characters, like “Brad Long (1061671)” or “Sunshine Baker (1110511)”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “Brad Long (1061671)” would be replaced with numeric value 1, and “Sunshine Baker (1110511)” would be replaced with numeric value of 2, etc.
55. **CE: Chief engineer**. Responsible to provide engineering planning and artifacts development. This would be used for the next round of analysis. Data values for this column are alpha numeric with wild characters, like “Brad Long (1061671)” or “Sunshine Baker (1110511)”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value

in increments of 1, 2, and 3, etc. For instance, all "Brad Long (1061671)" would be replaced with numeric value 1, and "Sunshine Baker (1110511)" would be replaced with numeric value of 2, etc.

56. **PE: Principle engineer**. Responsible to provide engineering support for planning and artifacts development. This would be used for the next round of analysis. Data values for this column are alpha numeric with wild characters, like "Brad Long (1061671)" or "Sunshine Baker (1110511)", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all "Brad Long (1061671)" would be replaced with numeric value 1, and "Sunshine Baker (1110511)" would be replaced with numeric value of 2, etc.
57. **BDL12**: Original values are the names of the business development lead for the project. This column was created to give data a different meaning of yes or no to reflect if there is a business development lead for the project or not. Therefore, the data values for this column are strings of text, like "Yes" or "No". Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, and 2. For instance, all "Yes" would be replaced with numeric value 1, and "No" would be replaced with numeric value of 2.
58. **CM12**: Original values are the names of the capture manager for the project. This column was created to give data a different meaning of yes or no to reflect if there is a capture manager for the project or not. Therefore, the data values for this column are strings of text, like "Yes" or "No". Since these are not all numeric

values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, and 2. For instance, all "Yes" would be replaced with numeric value 1, and "No" would be replaced with numeric value of 2.

59. **PM12**: Original values are the names of the proposal manager for the project. This column was created to give data a different meaning of yes or no to reflect if there is a proposal manager for the project or not. Therefore, the data values for this column are strings of text, like "Yes" or "No". Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, and 2. For instance, all "Yes" would be replaced with numeric value 1, and "No" would be replaced with numeric value of 2.

60. **CE12**: Original values are the names of the chief 60um60unte for the project. This column was created to give data a different meaning of yes or no to reflect if there is a chief engineer for the project or not. Therefore, the data values for this column are strings of text, like "Yes" or "No". Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, and 2. For instance, all "Yes" would be replaced with numeric value 1, and "No" would be replaced with numeric value of 2.

61. **PE12**: Original values are the names of the principle engineer for the project. This column was created to give data a different meaning of yes or no to reflect if there is a principle engineer for the project or not. Therefore, the data values for this

column are strings of text, like “Yes” or “No”. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, and 2. For instance, all “Yes” would be replaced with numeric value 1, and “No” would be replaced with numeric value of 2.

62. **Booking Value (BV)**: **Booking value**. Possible valuation of the project and possible contact if a buyer is identified. This would be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
63. **PTW: Price to win**. Strategic price for the project to win the possible contract over the competition. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
64. **RTN Bid (RB): Organizational bid**. Price proposed to the end user and customers for the product and services. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
65. **Bid Pwin (BP): Bid probability to win**. At what or a given price what are the probability of wining the contract. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is.

However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.

66. **Gate 4 Rate (G4R): Phase 4 rate**. Internal to the organization, evaluation of the project at the Phase 4. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
67. **Current Booking Rate (CBR): Current booking rate**. Internal to the organization as it rates the booking of the current project. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
68. **EAC Period (EAC): Estimate at completion period**. Internal to the organization as it estimates the quarter in which a given project would be completed. This would not be used for the next round of analysis. Data values for this column are alpha numeric with wild characters, like “Q1 2016” or “Q2 2015”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “Q1 2016” would be replaced with numeric value 1, and “Q2 2015” would be replaced with numeric value of 2, etc.
69. **Competitive Advantage Rating (CAR): Competitive advantage rating**. Internal to the organization with respect to the industry competition. It rates in three categories, neutral, advantage, and disadvantage. A measure to help make

various decisions on the project and about the project. If the project gives organization a competitive advantage then organization would dedicate more resources to that project in development, marketing, and sales. This would not be used for the next round of analysis. Data values for this column are strings of text, like “Adv” or “Neut”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “Adv” would be replaced with numeric value 1, and “Neut” would be replaced with numeric value of 2, etc.

70. **[Unknown Column Name] (UNK3): Unknown column name**. Project attributes are unknown. This would not be used for the next round of analysis. Data values for this column are numeric with wild characters, like “,0,0,0” or “,0,2,0”, etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all “,0,0,0” would be replaced with numeric value 1, and “,0,2,0” would be replaced with numeric value of 2, etc.
71. **Total B&P and Selling Hours (TBPSH): A set of total hours**. Collected from bidding, proposal, and selling activities. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
72. **B&P Hours (BPH): Hour spent in bidding and proposal activities**. Internal to the organization. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros

(0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.

73. **Selling Hours (SH): Hours spent in selling activities**. Internal to the organization. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
74. **IRAD Hours (IRDH): Hours spent IRAD activities**. Internal to the organization. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
75. **Total BOE Hours for Completed Proposals (Last Gate = 4) (TBOEH): Hours spent in BOE activities of the proposal work**. Internal to the organization. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
76. **Total ENG BOE Hours for Completed Proposals (Last Gate = 4) (TENGBOEH): Hours spent in engineering activities of the proposal work**. Internal to the organization. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.

77. **Total ENG B&P Hours for Completed Proposals (Last Gate = 4) (TENGBPH): Hours spent in engineering, bidding activities of the proposal**. Internal to the organization. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
78. **TVR Efficiency (TVRE): Total value in return**. Internal to the organization, a total value in return efficiency calculation. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
79. **Factored TVR Efficiency (FTVRE): Factored tvr efficiency**. Internal to the organization. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
80. **Competitors (CMPTRS): Competitors**. Possible competitors from current or future products and services. This would not be used for the next round of analysis. Data values for this column are strings of text with wild characters, like "Pure Services: ABC" or "CSRA (CSC)", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all "Pure Services:

ABC" would be replaced with numeric value 1, and "CSRA (CSC)" would be replaced with numeric value of 2, etc.

81. **Cust Engmnt (fr Gate or BH Pkgs) (CEBHP): Customer engagement**. For gate or bh package, with possible response of yes, no, or not found. This would not be used for the next round of analysis. Data values for this column are strings of text, like "NF" or "Yes", etc. Since these are not all numeric values, they were converted into numbers by replacing all identical string with a numeric value in increments of 1, 2, and 3, etc. For instance, all "NF" would be replaced with numeric value 1, and "Yes" would be replaced with numeric value of 2, etc.
82. **Cust Engmnt Freq (fr Gate or BH Pkgs) (CEFBHP): Customer engagement**. For gate or bh package at one of the four gates. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
83. **Potential Risk Cost impact ($K) (fr Gate pkg) (PRC): Potential cost impact**. Potential amount at risk for hours spent for gate package. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
84. **Risk Prob. (fr Gate pkg) (RP): Risk probability**. Internal to the organization, risk percentage for gate package. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is.

However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.

85. **Factored Risk Cost Impact (fr Gate pkg) (FRC): Factored risk and cost impact for gate package**. Internal to the organization. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
86. **Potential Opp Cost impact ($K) (OCI): Opportunity cost impact**. Potential opportunity cost impact, internal to the organization. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
87. **Opp prob. (fr Gate pkg) (OPP): Opportunity probability**. Internal to the organization, probability based on the gate package. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
88. **Factored Opp cost Impact (fr Gate pkg) (FOC): Factored opportunity cost impact**. Internal to the organization, for gate package. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.

89. **LM: Partnering company**. Various form of partnership based on the products and services. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
90. **NG: Partnering company**. Various form of partnership based on the products and services. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
91. **BAE: Partnering company**. Various form of partnership based on the products and services. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
92. **SAIC: Partnering company**. Various form of partnership based on the products and services. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
93. **GD: Partnering company**. Various form of partnership based on the products and services. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0)

would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.

94. **IBM: Partnering company**. Various form of partnership based on the products and services. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
95. **Boeing: Partnering company**. Various form of partnership based on the products and services. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.
96. **BAH: Partnering company**. Various form of partnership based on the products and services. This would not be used for the next round of analysis. Data values for this column are numeric and therefore, used as is. However, all zeros (0) would be replaced with 999 as R, the statistical tool used in this research cannot process 0 values.

3.6.2. Round Two Analysis

The remaining 29 columns are YR, BT, MA, SMA, WL, BS, PF, LG, ST, TVR, PG, PW, OT, CMT, RR, PRM, CNT, EU, P1, P2, P3, P4, TH, BDL, CM, PM, CE, PE, and BV. Based on the uniqueness of these 29 project performance indicators, below are sets of performance indicators which were selected for project performance prediction.

The uniqueness of the 29 project performance indicators are: YR (0.005), BT (0.438), MA (0.005), SMA (0.614), WL (0.005), BS (0.531), PF (0.005), LG (0.976), ST (0.005), TVR (0.005), PG (0.366), PW (0.005), OT (0.975), CMT (0.637), RR (0.585), PRM (0.587), CNT (0.882), EU (0.946), P1 (0.005), P2 (0.005), P3 (0.005), P4 (0.005), TH (0.005), BDL (0.368), CM (0.396), PM (0.305), CE (0.204), PE (0.005), and BV (0.615). Based on these uniqueness values, over .5 or 50% of the project attributes were removed, unless it is needed to validate a hypothesis. Human elements also, like PM, CE, and PE, would be removed, as they carry unique properties and these properties need to be studied as a whole personal and professional background. However, to validate a hypothesis they might be used. This praxis does not consider studying human elements. That being said, these individuals as project performance attributes would require understanding their internal and external, personal and professional relationships. Lastly, project performance indicators like BT were removed, because 21,639 out of 21,648 projects were missing the data value, Table 2 contains the final list of 16 selected data attributes or project performance indicators.

Table 2: Data Variables and Descriptions

#	Abbreviation	Name	Description
1	WL	Win Loss	Internal Gate 3 assessment of win or loss.
2	PF	Pass or Fail	End result of a project if passed or failed.
3	ST	Sales Type	Internal categorization of project based on sale.
4	PG	Probability Go	Gate 1 probability go
5	PW	Probability Win	Gate 2 probability win
6	CMT	Competition Type	Internal to company, like new, extension, or
7	RR	Company Role	Company Role in the project (Prime or Sub)
8	PRM	Prime	Company's partnership on the project with others
9	CNT	Contract Type	CPAF, CPFF, FFP, IDIQ, and T&M
10	EU	End User	End user of the project or targeted end user(s).
11	SPT	Sub Project Type	Internal category of project type.
12	P1	Phase 1	Calculated Phase 1 hours from gate 2 minus gate 1.
13	P2	Phase 2	Calculated Phase 2 hours from gate 3 minus gate 2.

Table 2 shows the final 13 data attributes after the second iteration of analysis during which more of the unique attributes were eliminated, including highly unique variables as a project attribute. These are the variables being studied, because these data attributes weights to a higher prediction accuracy and lowest FPR and FNR that would be used to predict the future of the current project. One such an indicator is project Probability-To-Go, which is decided in a collective setting of experts and project stakeholders. It is part of Phase 1 activities before the project is approved to go to Phase 2. Similarly, Probability-To-Win is a Phase 2 review before projects is approved to go to Phase 3. Examples of these review templates are in **Appendix B** and **Appendix C**.

3.6.3. Data Cleaning

The following is the data wrangling process by which the project data were evaluated and selected for this study:

1. The study began by calculating all phase hours from the original dataset:
    a. If a phase start date was not available, then that phase received 0 hours; this is true for all other phases also.
    b. Phase were calculated by the duration between two phases of the project. Phase 1 hours equal to phase 2 start date minus Phase 1 start date; Phase 3 start date minus Phase 2; and for Phase 4 hours, project submission date minus Phase 4 start date.
    c. If Phase 1 start dates were available, but Phase 2 were not, then Phase 3 start date was used to calculate Phase 1 hours and Phase 2 gets zero hours. Similarly, if Phase 1 date was available and Phase 2, 3, and 4 data was not available then project submission date was used to calculate Phase 1 hours. This does not

change the total hours worked in the project, but does add bias in the data for Phase 1 hours. This is also true for all other phases, and a bias acknowledged in the model.

d. Calculation: [Phase 2 – Phase 1 * {(40hours*52weeks)/365days}] equals to 5.7 work hours per day for 365 days (40*52/365=5.7).

e. Any paid and unpaid time off or holidays taken during the project were not included in this calculation. This is another bias acknowledged in this model. There were not enough data other than project phase date. Knowing the number of people who worked on the project with known ten federal holidays, more accurate calculations were possible.

f. Replacing project completion date with the due date would have added more projects to the study, since many projects were completed without a recorded submission date and most submission dates are the same as project due dates. This would have added more bias to the data and therefore not done for this study.

2. Selected the projects with end results of passed and failed, which brought the list of projects from 21,648 to 4,753.
3. Selected the projects with recorded end results as passed or failed, and this brought the number of projects from 4,753 to 4,032.
4. Added a column (preferable at the beginning, column A) called year (YR for short). Sorted the projects based on the submission date and added the respective years. Only selected years with at least 25 or more projects. For this praxis projects used between years 2007 to 2017, as all other project years listed fewer than 25 projects.

5. Copied and pasted these projects into a new spreadsheet, or new tab of a current spreadsheet. Then, exported the spreadsheet to save as a .csv file.
6. Imported the spreadsheet into R and verified that all projects were tagged with a year from 2007 to 2017.
    a. Removed the unknown and missing data columns as part of initial clean up.
    b. Renamed the title of all data columns.
    c. Masked the data by converting text to numeric values. Goal was to replace text with numbers in a chronological order, and all identical text would get same numeric value. For example, for the text series “Mohammed”, “Altaf”, and “Hossain”, each time the text “Mohammed” appeared, it would be replaced with numeric value 1, because it was the first text found, then text “Altaf” would be replaced with numerical value 2, and the text “Hossain” would be replaced with numeric value 3. The logic would hold true for all other text. All identical text (e.g. “Mohammed”) received the same numeric value during this process.
    d. Below is the sample of data after initial clean up and data masking hours.

Table 3: Cleaned and masked data (Note: for the data definition and numeric value representation, see **Data Definitions**)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	YR	BT	MA	SMA	WL	PS	PF	LP	PT	TV	PG	PW	OT	CMT	CR	PRM	CNT	PU	P1	P2	P3	P4	THC	BL	CL	PM	CE	PE	PV
2	2014	2	11	5	6	3	2	6	3	33912	20	100	3	4	3	139	3	162	0	0	0	0	0	20	37	50	17	30	$711
3	2014	1	4	8	6	3	2	6	3	33912	90	100	3	4	3	139	3	162	0	0	0	0	0	45	68	90	16	7	$711
4	2014	1	3	15	6	3	2	6	3	33912	90	100	3	4	3	139	3	162	0	0	0	0	0	27	140	63	53	5	$711
5	2016	1	9	3	6	3	2	6	3	281	90	60	3	4	4	162	3	359	0	0	0	0	0	71	38	40	30	3	$82
6	2016	1	7	6	6	3	2	6	3	281	90	60	3	4	4	162	3	359	0	0	0	0	0	4	36	94	69	18	$84
7	2016	1	8	7	6	3	2	6	3	281	90	60	3	4	4	162	3	359	0	0	0	0	0	52	80	58	82	31	$86
8	2016	1	2	9	6	3	2	6	3	8523	90	89	3	4	4	119	16	126	0	0	142.86	0	142.86	62	46	22	7	20	$1,289
9	2016	1	6	2	6	3	2	6	3	8523	90	89	3	4	4	119	16	126	0	0	142.86	0	142.86	49	90	13	48	56	$1,378
10	2016	1	5	14	6	3	2	6	3	8523	90	89	3	4	4	119	16	126	0	0	142.86	0	142.86	57	126	48	93	37	$1,396
11	2016	1	10	11	6	3	2	6	3	8523	90	89	3	4	4	119	16	126	0	0	142.86	0	142.86	63	9	8	74	8	$1,555
12	2017	1	1	13	6	3	2	6	3	20000	90	100	3	10	3	139	3	314	1240	445.71	0	199.05	1884.76	11	10	76	39	22	$4,000
13	2017	1	1	4	6	3	2	6	3	20000	90	100	3	10	3	139	3	314	1240	445.71	0	199.05	1884.76	40	109	54	70	10	$4,000
14	2017	1	1	12	6	3	2	6	3	20000	90	100	3	10	3	139	3	314	1240	445.71	0	199.05	1884.76	33	14	21	19	48	$4,000
15	2017	1	1	10	6	3	2	6	3	20000	90	100	3	10	3	139	3	314	1240	445.71	0	199.05	1884.76	80	40	44	91	49	$4,000
16	2015	1	1	1	6	3	2	6	3	12091	75	100	3	12	3	139	6	253	0	0	0	17.14	17.14	19	73	77	106	16	$1,500
17	2015	1	1	1	6	3	2	6	3	2200	50	75	3	12	4	164	1	1	0	0	34.29	62.86	97.14	56	69	9	115	38	$400
18	2015	1	1	1	6	3	2	6	3	2200	50	75	3	12	4	164	1	1	0	0	34.29	62.86	97.14	12	4	47	26	57	$600
19	2015	1	1	1	6	3	2	6	3	2200	50	75	3	12	4	164	1	1	0	0	34.29	62.86	97.14	73	49	23	60	40	$600
20	2017	1	1	1	5	3	2	6	3	6600	60	60	3	6	4	80	1	48	1137.14	1400.24	34.29	6988.09	9559.75	78	107	112	6	47	$2,200

e. Next, I found and replaced all zeros with 999, and made sure all columns were numeric in value (e.g., column AC or PV has dollar sign, $), and searched for any non-numeric value.

f. Saved the .csv file with a different name or overwrote the same file.

7. Imported the spreadsheet back into the R to perform t-test and LDA.

Table 4: Final Data for Models' Classifications and Predictions

WL	PF	ST	PG	PW	CMT	RR	PRM	CNT	EU	SPT	P1	P2
12	2	2	88	100	11	2	162	17	107	3	3228.33	120.00
12	2	2	88	100	11	2	162	17	107	3	3228.33	120.00
12	2	2	88	100	11	2	162	17	107	3	3228.33	120.00
12	2	2	85	100	3	2	162	2	290	43	2525.47	1108.57
11	2	2	50	100	3	2	162	2	21	40	2525.47	1108.57
11	2	2	55	90	3	2	162	2	290	43	2525.47	1108.57
11	2	2	99	100	3	2	162	2	290	43	2525.47	1108.57
12	2	2	90	100	10	2	162	1	107	3	554.05	1594.52
12	2	2	90	100	10	2	162	1	107	3	554.05	1594.52
12	2	2	90	100	10	2	162	1	107	3	554.05	1594.52
12	2	1	90	100	10	2	162	17	414	38	228.57	108.57
12	2	1	90	100	10	2	162	17	414	38	228.57	108.57

Note for Table 4 – for the data definition and numeric value representation, see **Data Definitions**.

3.7. Data and Statistical Analysis

The statistical analysis tool R was used to perform various statistical analyses for this model. First, perform factor analysis to identify the unique variables.

Table 5: Factor Analysis 29 Variables, After 1st Round of Cleanup

```
Uniquenesses:
   YR    BT    MA   SMA    WL    BS    PF    LG    ST   TVR    PG    PW    OT   CMT    CR   PRM   CNT    EU    P1    P2    P3    P4   TH
BDL    CM    PM    CE    PE
0.005 0.438 0.005 0.614 0.005 0.531 0.005 0.976 0.005 0.005 0.366 0.005 0.975 0.637 0.585 0.587 0.882 0.946 0.005 0.005 0.005 0.005 0.005 0.3
68 0.396 0.305 0.204 0.005
   BV
0.615

Loadings:
```

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8	Factor9	Factor10	Factor11	Factor12	Factor13	Factor14	Factor15	Factor16
YR						-0.362				0.876						0.231
BT							0.198							0.722		
MA	0.121						0.873							0.464		
SMA	0.191						0.577					0.120				
WL		0.973				0.102										-0.143
BS		0.251				0.540				-0.146					-0.254	0.144
PF		0.948											0.106		0.169	0.214
LG																-0.144
ST	-0.176					0.898				-0.337						-0.171
TVR					0.991											
PG				0.792												
PW				0.991												
OT																
CMT	0.102	0.172				-0.136									0.540	
CR						0.273				-0.153			-0.543			
PRM						0.123							0.619			
CNT						0.106				-0.214			-0.229			
EU						-0.140							0.117			
P1								0.158	0.984							
P2			0.918					0.364	-0.143							
P3			-0.947					0.296								
P4											0.973					0.193
TH								0.980	0.172							
BDL	0.642						0.278					0.369				
CM	0.766															
PM	0.826															
CE	0.884															
PE	0.502						0.184					0.836				
BV					0.615											

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8	Factor9	Factor10	Factor11	Factor12	Factor13	Factor14	Factor15	Factor16
SS loadings	2.835	1.961	1.753	1.623	1.398	1.398	1.268	1.216	1.026	0.991	0.968	0.860	0.784	0.749	0.428	0.254
Proportion Var	0.098	0.068	0.060	0.056	0.048	0.048	0.044	0.042	0.035	0.034	0.033	0.030	0.027	0.026	0.015	0.009
Cumulative Var	0.098	0.165	0.226	0.282	0.330	0.378	0.422	0.464	0.499	0.533	0.567	0.596	0.624	0.649	0.664	0.673

Test of the hypothesis that 16 factors are sufficient.
The chi square statistic is 3131.13 on 62 degrees of freedom.
The p-value is 0

Factor Analysis (FA): the goal is to examine the project data variables with respect to their uniqueness and their relationships. The top of the table shows the uniqueness of all 29 data variables. This is helpful to identify highly unique data variables, and as stated above highly unique variables were not used for discriminant analysis. The Loadings section shows how the data variables cluster together, making various factors. The most influential variables create the first factor, followed by the second more influential variables create second factor and so on. These data variables are correlated either negatively or positively.

3.8. Recommended Model – Decision Tree (DT)

The choice of using DT over other models is based on the understanding that DT yields higher accuracy predictions with the lowest FPR. Predicting functions show the probability of the project variables and where the variables originate. Based on historical data, project pass or fail will be predicted utilizing DT. The accuracy of the prediction then can be multiplied by the accuracy of the model from model validation. This process will be run using a set of the past three years of project data to test the model, and current year project data to validate the model. For instance, year 2007, 2008, and 2009 data were used to train the model; and year 2010 data to validate or to test the model. After running data through DT, there is 98.56% to 100% accuracy for the past three years project for test and the current projects for validation. FPR is 0% and FNR ranges 2.09% to 0%. The approach taken in this model shows the value of an organization taking the last three years of project data (according to this model), to perform a prediction on their current data. Lastly, perform prediction using any one of these validated models with an example of a set of project performance indicators to obtain a project pass or fail prediction.

3.9. Calculation and Decision

Based on the data and statistical analysis above, once highly unique variables and human factors are removed from the data, predictions were made and validated for project pass and fail. Based on the accuracy of the various validations, predictions can be made on current projects.

Chapter 4: Results

This chapter covers the interpretations of the data, findings, results, based on the research questions, hypotheses, and methodologies. There were total of 4,753 projects completed between 2007 and 2018. In it 2,570 projects passed, and 1,462 projects failed, totaling 4,032 and remaining 721 project status are in-pursuit, no bid, or in protest. Here onward, the total number of completed projects used in this praxis is 4,032. Therefore, taking 4,032 projects as 100%, based on the data recorded, Company has 63.74% success and 36.26% failure rate. Which could be interpreted as on average, one in every three projects has failed from year 2007 to 2017. Project failure is defined by the organization as – project that failed to be monetized.

4.1. Trends in Data

As stated in the intro based on all completed projects between 2007 and 2017, one in every three projects fails. Which led to further investigation of all projects data from the inception of project data collection. Followings are some of those findings or trends in data identified:

1. There are 21,648 projects recorded in the data from the inception of the project data collection.

1.1. 4,426 projects recorded with status In-pursuit (I) and will not be counted toward passed or failed in this study because the final outcome of these projects are unknown.

1.2. 409 projects recorded with status Unknown (U) and will not be counted toward passed or failed in this study because the final status of these projects are unknown.

1.3. 2,523 projects recorded with status Cancelled I by Company and will count toward the total failed projects.

1.4. 771 projects recorded with status Cancelled by Customer (CC) and will count toward the total failed projects.

1.5. 6,417 projects recorded with status Dropped (D) and will not be counted toward passed or failed in this study because the final outcome of these projects are unknown.

1.6. 994 projects recorded with status Failed (F) and will count toward the total failed projects.

1.7. 2,702 projects recorded with status No Bid (N) and will not be counted toward passed or failed in this study because the final outcome of these projects are unknown.

1.8. 21 projects recorded with status Protested (P) and will not be counted toward passed or failed in this study because the final outcome of these projects are unknown.

1.9. 3,385 projects recorded with status Passed (S) and will count toward the total passed projects.

2. That brings the total number of projects failed to (2523+771+6417+994=) 10,705 and total number of projects passed to 3,385, with a total of 14,090 projects.

3. 4,753 projects recorded completed between 2007 and 2018.

3.1. 2,570 projects out of 4,753 recorded passed.

3.2. 1,462 projects out of 4,753 recorded failed.

3.3. 4,032 projects in total recorded as complete with pass and failed, and used in this praxis.

3.4. 721 projects recorded in-pursuit, no bid, or in protest, and not used in this praxis.

4. 16,985 projects recorded without a technical lead.

4.1. From 16,985 projects, 400 with unknown (U) status, 2523 with canceled by company I status, 747 with cancelled by customer (CC) status, 6,417 with dropped by company (D) status, 992 with failed (F) status, 2,669 with no bid (N) status, 21 with protest (Pt) status, and 3,216 with passed (P) status.

4.2. That totals the number of projects failed without a tech lead (2523+747+6417+992=) 10,679 and total number of projects passed without a tech lead are 3,217. That's about 23.15% pass and 76.85% fail for projects without a tech lead.

4.3. Also to note, total number of project recorded failed is 994 and 992 of those recorded failed projects had no technical leads.

5. 13,015 projects recorded with no completion date, includes:

5.1. From 13,015 projects, 409-U, 2,311-C, 672-CC, 6,122-D, 133-F, 2,558-N, 3-Pt, and 807-P.

5.2. That totals the number of projects failed without a completion date to (2311+672+6122+133=) 9,238 and total number of projects passed is 807. That's about 8.03% pass and 91.97% fail for projects without a completion date.

6. 4,207 projects recorded with completion date, includes:

6.1. From 4,207 projects, 0-U, 212-C, 99-CC, 295-D, 861-F, 144-N, 18-Pt, and 2,578-P.

6.2. That totals the number of projects failed with a completion date to (212+99+295+861=) 1,467 and total number of projects passed is 2,578. That's about 63.73% pass and 36.27% fail for projects with a completion date.

7. 12,200 projects recorded with no Gate 1 review:

7.1. From 12,200 projects, 409-U, 1,726-C, 629-CC, 4,720-D, 479-F, 2,026-N, 2-Pt, and 2,209-P.

7.2. That totals the number of projects failed without gate 1 review to (1726+629+4720+479=) 7,534 and total number of projects passed is 2,209. That's about 22.67% pass and 77.33% fail for projects without gate 1 review.

8. 13,441 projects recorded with no Gate 2 review:

8.1. From 13,441 projects, 409-U, 1,910-C, 671-CC, 5,630-D, 432-F, 2,257-N, 3-Pt, and 2,129-P.

8.2. That totals the number of projects failed without gate 2 review to (1910+671+5630+432=) 8,643 and total number of projects passed is 2,129. That's about 19.76% pass and 80.24% fail for projects without gate 2 review.

9. 11,472 projects recorded with no Gate 1 and Gate 2 reviews:

9.1. From 11,472 projects, 409-U, 1,592-C, 605-CC, 4,578-D, 382-F, 1,936-N, 2-Pt, and 1,968-P

9.2. That totals the number of projects failed without gate 1 and gate 2 reviews to (1592+605+4578+382=) 7,157, and total number of projects passed is 1,968. That's about 21.57% pass and 78.43% fail for projects without gate 1 and gate 2 reviews.

10. 409 projects recorded with status Unknown (U).

10.1. All projects with U status are domestic projects, all never got completed, all had no technical lead, and all had no gate review.

11. 2,523 projects recorded with status Cancelled I by Company.

11.1. All projects with C status had no technical lead (TL).

11.2. 2,311 without submission date, 1,726 without Gate 1 (G1) review, 1,910 without Gate 2 (G2) review, and 1,592 without G1 and G2 reviews.

12. 771 projects recorded with status Cancelled by Customer (CC).

12.1. 685 domestic projects, 747 without tech lead, 672 without submission date, 629 without G1, 671 without G2, 605 without G1 and G2, and 584 without TL, G1, and G2.

12.2. 600 domestic projects without tech leads were CC status.

13. 6,417 projects recorded with status Dropped (D).

13.1. All projects with D status had no tech lead.

13.2. 6,122 without submission date, 4,720 without G1, 5,630 without G2, and 4,578 without G1 and G2 reviews.

14. 994 projects recorded with status Failed (F).

14.1. Almost all (992) projects with F status had no tech lead.

14.2. 133 without submission date, 479 without G1, 432 without G2, and 382 without G1 and G2 reviews.

15. 2,702 projects recorded with status No Bid (N).

15.1. 2,558 without submission date, 2,669 without tech lead, 2,026 without G1, 2,257 without G2, 1936 without G1 and G2, and 1,915 without TL, G1, and G2 reviews.

16. 21 projects recorded with status Protested (Pt).

16.1. 3 without submission date, 21 without tech lead, 2 without G1, 3 without G2, 2 without G1 and G2, and 2 without TL, G1, and G2 reviews.

17. 3,385 projects recorded with status Passed (P).

17.1. 807 without submission date, 3,216 without tech lead, 2,209 without G1, 2,129 without G2, 1,968 without G1 and G2, and 1,902 without TL, G1, and G2 reviews.

18. 271 international projects are recorded with gate one review with average of 1,664.69 hours spent in phase one.

18.1. 63 cancelled by the company with average of 1,840.54 hours spent in Phase one.

18.2. 15 cancelled by the customers with average of 1,774.86 hours spent in Phase one.

18.3. 121 dropped by the company with average of 2,715.69 hours spent in Phase one.

18.4. 33 failed with average of 1,108.22 hours spent in Phase one.

18.5. 39 passed with average of 884.12 hours spent in Phase one.

19. 200 international projects are recorded with gate two review with average of 883.74 hours spent in phase two.

19.1. 47 cancelled by the company with average of 1,932.52 hours spent in Phase two.

19.2. 13 cancelled by the customers with average of 426.50 hours spent in Phase two.

19.3. 55 dropped by the company with average of 305.11 hours spent in Phase two.

19.4. 40 failed with average of 620.39 hours spent in Phase two.

19.5. 45 passed with average of 1,134.18 hours spent in Phase two.

20. 4,022 domestic projects are recorded with gate one review with average of 1,744.54 hours spent is phase one.

20.1. 732 cancelled by the company with average of 1,840.54 hours spent in Phase one.

20.2. 127 cancelled by the customers with average of 1,774.86 hours spent in Phase one.

20.3. 1552 dropped by the company with average of 2,715.69 hours spent in Phase one.

20.4. 482 failed with average of 1,108.22 hours spent in Phase one.

20.5. 1129 passed with average of 884.12 hours spent in Phase one.

21. 3095 domestic projects are recorded with gate two review with average of 2,340.71 hours spent is phase two.

21.1. 565 cancelled by the company with average of 1,840.54 hours spent in Phase one.

21.2. 87 cancelled by the customers with average of 1,774.86 hours spent in Phase one.

21.3. 722 dropped by the company with average of 2,715.69 hours spent in Phase one.

21.4. 522 failed with average of 1,108.22 hours spent in Phase one.

21.5. 1,199 passed with average of 884.12 hours spent in Phase one.

22. Average hours spent in each of the project phases are followings:

22.1. Phase one: 2,659.30 hours over the 4,464 projects.

22.2. Phase two: 2,855.31 hours over the 3,525 projects.

22.3. Phase three: 456.84 hours over the 3,914 projects.

22.4. Phase four: 616.49 hours over the 4,231 projects.

4.2. Hypotheses

H1: Project success or failure can be forecasted from project performance indicators. Accept.

$H1_0$: Project success or failure cannot be forecasted from project performance indicators.

According to the methodologies discussed in chapter 3 above, all four models used in this praxis predicted project outcome accurately. Accuracy ranges from 91.67% to 100%, FPR ranges from 29.51% to 0%, and FNR ranges from 4.74% to 0%. In Table 6 below, an example of a hypothetical project is being predicted based on the models trained and validated in this praxis. All four models predicts the project to be a failure. Therefore, there are enough evidence to reject the null hypothesis.

Table 6. Example of Making Predictions on a Current Project by All Four Models:

Independent Variables (Ivs): WL, ST, PG, PW, CMT, RR, PRM, CNT, EU, SPT, P1, and P2.		
Dependent Variable (DV): PF	**1 = Failed** and **2 = Passed**	
Example: A current project that is being predicted by the models, where: WL=6, ST=5, PG=80, PW=70, CMT=5, RR=4, PRM=1, CNT=3, EU=10, SPT=5, P1=1000, and P2=2000.		
DT	**R Code:**	**Result**

Prediction:	`> predict(lda1416, newdata = data.frame(WL=6,ST=5,PG=80,PW=70,CMT=5,RR=4,PRM=1,CNT=3,EU=10,SPT=5,P1=1000,P2=2000))`
	`  Failed Passed`
	`1      1      0`
		Failed

LDA	**R Code:**	**Result**

Prediction:	`> predict(lda1416, newdata = data.frame(WL=6,ST=5,PG=80,PW=70,CMT=5,RR=4,PRM=1,CNT=3,EU=10,SPT=5,P1=1000,P2=2000))`
	`$class`
	`[1] 1`
	`Levels: 1 2`
	
	`$posterior`
	`5. 2`
	`1 0.9999754 2.456537e-05`
	
	`$x`
	`       LD1`
	`1 -3.03811`
		Failed

QDA	**R Code:**	**Result**

Prediction:	`> predict(lda1416, newdata = data.frame(WL=6,ST=5,PG=80,PW=70,CMT=5,RR=4,PRM=1,CNT=3,EU=10,SPT=5,P1=1000,P2=2000))`
	`$class`
	`[1] 1`
	`Levels: 1 2`
	
	`$posterior`
	`6. 2`
	`1 1 1.537988e-08`
		Failed

SVM	**R Code:**	**Result**

Prediction:	`> predict(lda1416, newdata = data.frame(WL=6,ST=5,PG=80,PW=70,CMT=5,RR=4,PRM=1,CNT=3,EU=10,SPT=5,P1=1000,P2=2000))`
	`     1`
	`Failed`
	`Levels: Failed Passed`
		Failed

According to **Table 6** above, all four models used in this praxis yield exact same prediction results for a given hypothetical project. Also, **Table 8** below show some of the very similar output of accuracy rate, FPR, and FNR while validating the models. Therefore, there is not enough evidence to reject the null hypothesis.

4.3. Method Selection for Data Classification

There are over 80 data classifications methods available. It is recommended to try multiple methods in order to draw conclusions based on the model that yields high accuracy and low false positive. Four methods are used for this study. They are Decision Tree (DT), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Support Vector Machine (SVM). For classification and prediction model training and validation eight (8) different data sets were used. First three years data was used to train the model and forth year data to validate the model. Table below shows the test data and validation data:

Table 7. Test and Validation Datasets

Datasets	Test Data by Year	Validation Data
1.	2007, 2008, 2009	2010
2.	2008, 2009, 2010	2011
3.	2009, 2010, 2011	2012
4.	2010, 2011, 2012	2013
5.	2011, 2012, 2013	2014
6.	2012, 2013, 2014	2015
7.	2013, 2014, 2015	2016
8.	2014, 2015, 2016	2017

For data classification Decision Tree (DT) evaluates every influential attribute to make a decision. It does so by splitting these attributes into groups. However, while attributes are unique, it is a fixed variation and repeated. For analysis, actual project performance indicators were used instead of their ranges for DT prediction. In other words, performance indicators like phase hours or pgo and pwin could have been taken in range them from 0-100, 101-200 and in doing so, inaccuracy would have been added to the results. The goal was to keep data at it original state as much possible while giving data meanings where needed for accurate data classification.

LDA and QDA are used interchangeably to attain higher accuracy, lower false positives, and false negatives. Objectives of Discriminant Analysis is to create clusters and to be able to group the projects data based on their characteristics. While LDA assumes a linear relationship between variables, QDA does not assume that relationships between variables. In other words, LDA assumes covariance between variables being identical, where QDA does not assume covariance being identical. LDA if used for regression then assumes linear relationships between variables, where QDA assume non-linear relationships between variations.

Similarly, Support Vector Machine (SVM) groups and separates the data points using lines known as hyperplanes. These hyperplanes are multidimensional and locates the best dimensions to separate the data points in space. It identifies the largest separations between data points. Project performance indicators combined can create a project unique in space, therefore perfect for hyperplanes. The objective is not to find the perfect distances between the two sets of projects, rather to predict which group of project performance indicators would yield highest prediction accuracy with lowest false positive. Also, SVM does not provide a percentage with its prediction. Therefore, when it predicts a project being on the failed side or passed side without a percentage of accuracy, it would add risk to decision making. Not knowing the exact percentage of failure would not add much value in project continuity decision making.

Table 8. Failed Projects Accuracy Based on Various Models [FP = (FP/(FP+TN))]

Independent Variables (IVs): WL, ST, PG, PW, CMT, RR, PRM, CNT, EU, SPT, P1, and P2.																
Dependent Variable (DV): PF																
	DT				**LDA**				**QDA**				**SVM**			
2010	Accuracy:		**100.00%**		Accuracy:		**98.31%**		Accuracy:		**99.66%**		Accuracy:		**97.29%**	
	TN=	85	FP=	0	TN=	85	FP=	0	TN=	84	FP=	1	TN=	77	FP=	8
	FN=	0	TP=	210	FN=	5	TP=	205	FN=	0	TP=	210	FN=	0	TP=	210
	FNR=	0.00%	**FPR=**	**0.00%**	FNR=	2.38%	**FPR=**	**0.00%**	FNR=	0.00%	**FPR=**	**1.18%**	FNR=	0.00%	**FPR=**	**9.41%**
2011	Accuracy:		**100.00%**		Accuracy:		**100.00%**		Accuracy:		**99.41%**		Accuracy:		**98.24%**	
	TN=	137	FP=	0	TN=	137	FP=	0	TN=	137	FP=	0	TN=	131	FP=	6
	FN=	0	TP=	204	FN=	0	TP=	204	FN=	2	TP=	202	FN=	0	TP=	204
	FNR=	0.00%	**FPR=**	**0.00%**	FNR=	0.00%	**FPR=**	**0.00%**	FNR=	0.98%	**FPR=**	**0.00%**	FNR=	0.00%	**FPR=**	**4.38%**
2012	Accuracy:		**100.00%**		Accuracy:		**100.00%**		Accuracy:		**97.14%**		Accuracy:		**99.52%**	
	TN=	167	FP=	0	TN=	167	FP=	0	TN=	167	FP=	0	TN=	165	FP=	2
	FN=	0	TP=	253	FN=	0	TP=	253	FN=	12	TP=	241	FN=	0	TP=	253
	FNR=	0.00%	**FPR=**	**0.00%**	FNR=	0.00%	**FPR=**	**0.00%**	FNR=	4.74%	**FPR=**	**0.00%**	FNR=	0.00%	**FPR=**	**1.20%**
2013	Accuracy:		**100.00%**		Accuracy:		**100.00%**		Accuracy:		**100.00%**		Accuracy:		**99.78%**	
	TN=	172	FP=	0	TN=	172	FP=	0	TN=	172	FP=	0	TN=	171	FP=	1
	FN=	0	TP=	288	FN=	0	TP=	288	FN=	0	TP=	288	FN=	0	TP=	288
	FNR=	0.00%	**FPR=**	**0.00%**	FNR=	0.00%	**FPR=**	**0.00%**	FNR=	0.00%	**FPR=**	**0.00%**	FNR=	0.00%	**FPR=**	**0.58%**
2014	Accuracy:		**98.87%**		Accuracy:		**98.49%**		Accuracy:		**98.87%**		Accuracy:		**94.15%**	
	TN=	237	FP=	0	TN=	237	FP=	6	TN=	237	FP=	6	TN=	212	FP=	31
	FN=	6	TP=	287	FN=	2	TP=	285	FN=	0	TP=	287	FN=	0	TP=	287
	FNR=	2.05%	**FPR=**	**0.00%**	FNR=	0.70%	**FPR=**	**2.47%**	FNR=	0.00%	**FPR=**	**2.47%**	FNR=	0.00%	**FPR=**	**12.76%**
2015	Accuracy:		**99.74%**		Accuracy:		**99.60%**		Accuracy:		**99.74%**		Accuracy:		**98.01%**	
	TN=	281	FP=	0	TN=	281	FP=	2	TN=	281	FP=	2	TN=	277	FP=	6
	FN=	2	TP=	472	FN=	1	TP=	471	FN=	0	TP=	472	FN=	9	TP=	463
	FNR=	0.42%	**FPR=**	**0.00%**	FNR=	0.21%	**FPR=**	**0.71%**	FNR=	0.00%	**FPR=**	**0.71%**	FNR=	1.91%	**FPR=**	**2.12%**
2016	Accuracy:		**98.56%**		Accuracy:		**97.99%**		Accuracy:		**95.55%**		Accuracy:		**95.26%**	
	TN=	217	FP=	0	TN=	218	FP=	9	TN=	201	FP=	26	TN=	206	FP=	21
	FN=	10	TP=	469	FN=	5	TP=	464	FN=	5	TP=	464	FN=	12	TP=	457
	FNR=	2.09%	**FPR=**	**0.00%**	FNR=	1.07%	**FPR=**	**3.96%**	FNR=	1.07%	**FPR=**	**11.45%**	FNR=	2.56%	**FPR=**	**9.25%**
2017	Accuracy:		**98.61%**		Accuracy:		**99.07%**		Accuracy:		**91.67%**		Accuracy:		**96.76%**	
	TN=	58	FP=	0	TN=	59	FP=	2	TN=	43	FP=	18	TN=	59	FP=	2
	FN=	3	TP=	155	FN=	0	TP=	155	FN=	0	TP=	155	FN=	5	TP=	150
	FNR=	1.90%	**FPR=**	**0.00%**	FNR=	0.00%	**FPR=**	**3.28%**	FNR=	0.00%	**FPR=**	**29.51%**	FNR=	3.23%	**FPR=**	**3.28%**
	DT				**LDA**				**QDA**				**SVM**			
FA's P-Value:	0.0347				0.0347				0.0347				0.0347			
Example of a current project that is being predicted by the models, where: WL=6, ST=5, PG=80, PW=70, CMT=5, RR=4, PRM=1, CNT=3, EU=10, SPT=5, P1=1000, and P2=2000.																
Prediction Results:	Failed				Failed				Failed				Failed			

Average time spent in Phase 1 (P1) is 1,204.63 hours, Phase 2 (P2) 1,521.38 hours, Phase 3 (P3) is 645.36 hours, and Phase 4 (P4) is 521.47 hours. Phases 3 and 4 of the project lifecycles—645.36 hours spent during Phase 3 and 521.47 hours spent during Phrase 4. When combined P4 and P4 hours, about 1,166.83 hours of work at the rate of $255 per hour (the Company proprietary rate) could be saved if the project stopped before beginning work on Phases 3 and 4. These model as discussed earlier takes the weights of project performance indicators, as their combined weight outputs the accuracy of the prediction. In other words, based on the Company's past performance on similar projects, utilizing the recommended model Company will save 91.67% to 100% of the remaining project cost, which could lead to (1166.83*255*.9167 or) $ 272,756.43 to (1166.83*255*1 or) $297,541.65. The goal for this study was to save a minimum of $100,000 per project if stopped. The p-values in the bottom of Table 8 are from the factor analysis of the final 12 IVs, where confidence level was set to 95%. Therefore, goal was to have a p-value less than .05, and as expected, all four models yield p-value less than .05%.

Table 9. A comparison of the averages of the Phase Hours.

Total projects undertaken:	14,090	TVM (255/hr.)
Average Phase 1 hours:	550.91	$ 140,482.05
Average Phase 2 hours:	525.66	$ 134,043.30
Average Phase 3 hours:	96.72	$ 24,663.60
Average Phase 4 hours:	137.77	$ 35,131.35
Phase 3 and 4 hours:	234.49	$ 59,794.95
Total completed projects:	4,032	TVM (255/hr.)
Average Phase 1 hours:	1204.63	$ 307,180.65
Average Phase 2 hours:	1521.38	$ 387,951.90
Average Phase 3 hours:	645.36	$ 164,566.80

Average Phase 4 hours:	521.47	$ 132,974.85
Phase 3 and 4 hours:	1,166.83	$ 297,541.65

As shown above in Table the average hours spent in project phases are different in average of all completed projects with missing phase hours and final list of projects selected for this study. Goal of this study was to save organizations a minimum of $100,000 if the project work stopped after first two phases.

4.4. Decision-Making

Since the goal of this study is to predict project failure in order to aid decision-making, the use of the lowest prediction represents minimum savings. Models used to predict possible outcomes of a current project after Phases 1 and 2, given that the other elements of the project were known and remain the same. These predictions use the Company's past performance of similar projects, and if the project work terminated after the first two phases, then resources are saved from stopping the last two phases. This applies to all projects for evaluation and prediction. Models make predictions after Phase 2 because stopping a project after Phase 1 might be considered too premature a decision. Similarly, for certain special projects, work may continue until Phase 3 before prediction analysis to see if the project would pass or fail before continuing to Phase 4. With the true average for Phase 4 hours being 521.47 hours, stopping project prior to Phase 4 would save organization over one hundred thousand (521.47*255=$132,974.85).

On average, closed to three hundred thousand dollars could have been saved if any of these failed projects were terminated earlier, after Phase 2. Certain projects might never stop for various reasons, however. For instance, an organizational annual goal to complete a set number of projects, project importance level, project sponsorship, organizational size, reason for the project, and reputation of the organization influence whether a project is terminated or completed. In other words, there are soft values (non-monetary) to every project just as much as there are hard values (monetary). Although it is subjective to the organization with regards to how it decides to stop a project, this study is an additional tool to help an organization make that decision. This study would be an excellent asset to organizations because the collected data reveals their already-stored internal stakeholder predictions on probability to succeed and probability to terminate. These statistical predictive models, in conjunction with other tools and methods, would aid organizational decision-making. Bigger organizations may have higher risk tolerance than smaller, but similar, organizations. Large size organizations keep larger budgets for research, development, proof-of-concept (POC), and prototype projects to stay competitive, enter new markets, or to undertake projects just to market their capabilities to end users. Although this praxis does not cover those problems, future research can answer questions pertaining to them.

Chapter 5: Interpretations, Conclusions, and Recommendations

5.1. Opening

Project performance indicators can forecast project failure and utilizing this prediction model can help decision-makers terminate projects and save organizational resources. The study above shows that performing Factor Analysis to understand the relationships between project success or failure with project performance indicators and which factors lead to project success or failure. Followed by performing various classification methods on a current project based on past project success or failure can predict the probability of failure weighted to dollar value, and that terminating project work after the first few (recommended two) phases will potentially save organizations a significant amount of money and allow resources to be deployed more efficiently. For the organization studied, this proposed model would have resulted in about three hundred thousand dollars of savings per project cancelled early in its lifecycle.

5.2. Interpretation and Conclusions

Interpretation of the results found in this study shows that project continuity decisions will benefit by utilizing this study. Although the accuracy of this model is subjective to the data used in this study, the practice of predictive analytics in everyday project management will influence how decisions are made. The results of this study are not meant as the only approach to introduce predictive analytics to project management but are meant to take the accuracy percentage and use it to calculate potential savings by terminating project work and, therefore, make better decisions. Four classification models used in this study yields similar results, thus concluded that any data classification model can be used to reach same conclusion. The literature reviewed in this study shows predictive analytics is not being used currently for project management and project continuity decisions making. Organizations are more actively storing data, yet do not have a plan to use these data. Studies in the area of critical activities may demonstrate how an overall project fails, root cause analysis of the project failure, and preventive measures to avoid project failure. What is being proposed in this study is only the beginning of a world of potential for the impact of predictive analytics on project management, because once organizational leadership begins to adopt predictive analytics, the dynamics of project management will change. New tools will emerge, tools which would aid predictability across all four phases of the project lifecycle and alter traditional project management and, more specifically, project continuity decision-making.

To conclude, predictive analytics is not being used to aid either project continuity decision-making or overall project management. As a result, many projects fail, run behind schedule, and most incur costs which exceed original planning estimates. Many studies pursued answers concerning the identification of critical tasks, root cause analysis, and establishing the life span of mechanical equipment. Many other areas of project management have been studied, but no previous study has investigated predictive analysis of project failure.

5.3. Recommendations

It is recommended that organizations maintain clear and actionable objectives and develop more objective-focused data storage and archiving plans. Studies show organizations are more focused on identifying why a project fails after it has failed to prevent future projects from failing. Organizations do not have a plan for using their past project performance data to study the contributing factors into these project failures and how these could be predicted. It is also recommended to curve a project management solution utilizing historical project data as input, utilize the current and relevant data available that would affect the given project performance, and use predictive analytics in all area of project planning, execution, and reporting.

5.4. Future Research

In future research it recommended to apply predictive analytics in all area of project management from planning, execution, and to reporting. It is also recommended that organizations maintain clear and actionable objectives and develop more objective-focused data storage and archiving plans. Next, it would be nice to see possible benefits of objective focused data in combination of artificial intelligence.

References

Adler, P. S. (2005). The evolving object of software development. *Organization*, 12(3), 401-435. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/218634728?accountid=11243

Al Sehaimi, A. O., Patricia, T. F., & Koskela, L. (2014). Improving construction management practice with the last planner system: A case study. *Engineering, Construction and Architectural Management*, 21(1), 51-64. Doi:http://dx.doi.org/10.1108/ECAM-03-2012-0032

Althonayan, M., & Althonayan, A. (2017). E-government system evaluation. *Transforming Government: People, Process and Policy*, 11(3), 306-342. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1954348864?accountid=11243

Alzahrani, J.I. and Emsley, M.W. (2013). The impact of contractors' attributes on construction project success: a post construction evaluation. *International Journal of Project Management*, Vol. 31 No. 2, pp. 313 – 322 .

Appan, R., & Browne, G. J. (2010). Investigating retrieval-induced forgetting during information requirements determination. *Journal of the Association for Information Systems*, 11(5), 250-275. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/366333794?accountid=11243

Askar, M., Imam, S., & Prabhaker, P. R. (2009). Business metrics: A key to competitive advantage. *Advances in Competitiveness Research*, 17(1), 91-109. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/211371112?accountid=11243

Aßländer, M.,S., Roloff, J., & Nayir, D. Z. (2016). Suppliers as stewards? Managing social standards in first- and second-tier suppliers. *Journal of Business Ethics*, 139(4), 661-683. Doi:http://dx.doi.org/10.1007/s10551-016-3148- 0

Attaran, M., PhD. (2017). Cloud computing technology: Leveraging the power of the internet to improve business performance. *Journal of International Technology and Information Management*, 26(1), 112-137. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1938528845?accountid=11243

Bala, H., & Venkatesh, V. (2017). Employees' reactions to IT-enabled process innovations in the age of data analytics in healthcare. *Business Process Management Journal*, 23(3), 671-702. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1913344548?accountid=11243

Beauregard, Y., Bhuiyan, N., & Thomson, V. (2011). Post-certification Engineering Taxonomy and task value optimization in the aerospace industry. *Engineering Management Journal*, 23(1), 86-100. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/862561384?accountid=11243

Bigliardi, B., Galati, F., & Petroni, A. (2014). How to effectively manage knowledge in the construction industry. *Measuring Business Excellence*, 18(3), 57-72. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1661349142?accountid=11243

Briggs, R. O., Reinig, B. A., & Vreede, G. d. (2008). The yield shift theory of satisfaction and its application to the IS/IT domain*. *Journal of the Association for Information Systems*, 9(5), 267-293. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/198766300?accountid=11243

Chan, A. P. C., & Chan, A. P. L. (2004). Key performance indicators for measuring construction success. *Benchmarking*, 11(2), 203-221. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/217369294?accountid=11243

Aiyer, J., Rajkumar, T. M., & Havelka, D. (2005a). A staged framework for the recovery and rehabilitation of troubled IS development projects. *Project Management Journal*, 36(4), 32-43. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/218772103?accountid=11243

Al-Ahmad, W., Al-Fagih, K., Khanfar, K., Alsamara, K., Abuleil, S., & Abu-Salem, H. (2009b). A taxonomy of an IT project failure: Root causes. *International Management Review*, 5(1), 104,106. Retrieved from

http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/195546967?accountid=11243

Al Sabah, R. (2014b). *Evaluating the impact of construction risks on project success in the Arabian gulf region (AGR) construction industry from the perspective of multinational firms (Ph.D.)*. Available from ProQuest Dissertations & Theses Global. (1526021023). Retrieved from https://search-proquest-com.proxygw.wrlc.org/docview/1526021023?accountid=11243

Amoatey, C., & Hayibor, M. V. K. (2017a). Critical success factors for local government project stakeholder management. *Built Environment Project and Asset Management*, 7(2), 143-156. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1895059480?accountid=11243

Ansah, R. H., Sorooshian, S., Mustafa, S. B., & Oludapo, O. S. (2017). Constructions project management risks' framework. *Calitatea*, 18(158), 90-95. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1912585467?accountid=11243

Boyton, J., Ayscough, P., Kaveri, D., & Chiong, R. (2015). Suboptimal business intelligence implementations: Understanding and addressing the problems. *Journal of Systems and Information Technology*, 17(3), 307-320. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1697924261?accountid=11243

Brookfield, D., Fischbacher-smith, D., Mohd-rahim, F., & Boussabaine, H. (2014a). Conceptualising and responding to risk in IT projects. *Risk Management*, 16(3), 195-230. Doi://dx.doi.org.proxygw.wrlc.org/10.1057/rm.2014.10

Charette, R. N. (2005a). Why software FAILS. *IEEE Spectrum*, 42(9), 42-49. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/196663847?accountid=11243

Chen, S., & Schiele, K. (2017). Positioning SMES with a lone designer: A qualitative study of small design firms. *Journal of Marketing Theory and Practice*, 25(3), 213-233. Doi:http://dx.doi.org/10.1080/10696679.2017.1311217

Chou, H. (2015). Multiple-technique approach for improving a performance measurement and management system: Action research in a mining company. *Engineering Management Journal*, 27(4), 203-217. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1761611070?accountid=11243

Cifalino, A., & Baraldi, S. (2009). Training programs and performance measurement. *Journal of HRCA : Human Resource Costing & Accounting*, 13(4), 294-315. Doi:http://dx.doi.org/10.1108/14013381011010141

Costantino, Francesco, Giulio Di Gravio, and Fabio Nonino. (2015). Project selection in project portfolio management: An artificial neural network model based on critical success factors. *International Journal of Project Management,* 33, (8) (11): 1744, http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1728271430?accountid=11243 (accessed October 20, 2018).

Conference report. (2001a). *The Journal of Management Development*, 20(9/10), 865-871. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/216350581?accountid=11243

Costantino, F., Di Gravio, G., & Nonino, F. (2015). Project selection in project portfolio management: An artificial neural network model based on critical success factors. *International Journal of Project Management*, 33(8), 1744. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1728271430?accountid=11243

Cwikla, M., & Jalocha, B. (2015). Unspread wings. *International Journal of Managing Projects in Business*, 8(4), 626-648. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1732530290?accountid=11243

Damian, D., & Chisan, J. (2006). An empirical study of the complex relationships between requirements engineering processes and other processes that lead to payoffs in productivity, quality, and risk management. *IEEE Transactions on Software Engineering*, 32(7), 433-453. Doi:http://dx.doi.org/10.1109/TSE.2006.61

dos Santos, Flávio Roberto Souza, & Cabral, S. (2008b). FMEA and PMBOK applied to project risk management. *Journal of Information Systems and Technology Management*: JISTEM, 5(2), 347-364. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/233667737?accountid=11243

Duris, R. (2002). Top 10 reasons why WMS projects fail. *Frontline Solutions*, 3(4), 62. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/217794870?accountid=11243

Ellinas, C., Allan, N., Durugbo, C., & Johansson, A. (2015). How robust is your project? From local failures to global catastrophes: A complex networks approach to project systemic risk. *PloS One*, 10(11), e0142469. Doi://dx.doi.org.proxygw.wrlc.org/10.1371/journal.pone.0142469

Engle, P. (2016). Buffers, a tale told of false security. *ISE: Industrial and Systems Engineering at Work*, 48(5), 20. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1792128785?accountid=11243

Engle, P. (2011). Crunch time. *Industrial Engineer*, 43(6), 20. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/873256307?accountid=11243

Farris, J. A., van Aken, E.,M., Letens, G., Chearksul, P., & Coleman, G. (2011). Improving the performance review process. *International Journal of Operations & Production Management*, 31(4), 376-404. Doi:http://dx.doi.org/10.1108/01443571111119524

Ferguson, A., Clinch, G., & Kean, S. (2011c). Predicting the failure of developmental gold mining projects. Australian Accounting Review, 21(1), 44-53. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/858796978?accountid=11243

Folkerd, C., & Spinelli, G. (2009). User exclusion and fragmented requirements capture in publicly-funded IS projects. *Transforming Government: People, Process and Policy*, 3(1), 32-49. Doi:http://dx.doi.org/10.1108/17506160910940722

Ghiasi, V., Kaivan, E., Arzjani, N., & Arzjani, D. (2017c). Analyzing the causes of delay in development projects by fuzzy analysis. *The International Journal of Quality & Reliability Management*, 34(9), 1412-1430. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1987383829?accountid=11243

Gleirscher, M., Golubitskiy, D., Irlbeck, M., & Wagner, S. (2014). Introduction of static quality analysis in small- and medium-sized software enterprises: Experiences from technology transfer. *Software Quality Journal*, 22(3), 499- 542. Doi:http://dx.doi.org/10.1007/s11219-013-9217-z

Goh, J. C., Pan, S. L., & Zuo, M. (2013). Developing the agile IS development practices in large-scale IT projects: The trust-mediated organizational controls and IT project team capabilities perspectives. *Journal of the Association for Information Systems*, 14(12), 722-756. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1477881060?accountid=11243

Gospodarevskaya, E., & Churilov, L. (2011). Process performance indicators in redesigning the patient care process. *Business Process Management Journal*, 17(6), 1012-1038. Doi:http://dx.doi.org/10.1108/14637151111182738

Haji-Kazemi, S., Anderson, B. (2013). Application of performance measurement as an early warning system: A case study in the oil and gas industry. *International*

Journal of Managing Projects in Business, Vol. 6 Issue: 4, pp.714-738, https://doi.org/10.1108/IJMPB-04-2012-0015

Halper, F. (2014). Predictive analytics for business advantage. *TDWI Research*, 1-32.

Harry K.H. Chow, Choy, K. L., & Lee, W. B. (2007). Knowledge management approach in build-to-order supply chains. *Industrial Management & Data Systems*, 107(6), 882. Doi:http://dx.doi.org/10.1108/02635570710758770

Hawkins, T. G., & Muir, W. A. (2014). An exploration of knowledge-based factors affecting procurement compliance. *Journal of Public Procurement*, 14(1), 1-32. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1536852251?accountid=11243

Hodge, L. M., B.Sc, Turner, K. M., T., P., Sanders, M. R., PhD., & Filus, A., PhD. (2017). Sustained implementation support scale: Validation of a measure of program characteristics and workplace functioning for sustained program implementation. *The Journal of Behavioral Health Services & Research*, 44(3), 442-464. Doi:http://dx.doi.org/10.1007/s11414-016-9505-z

Hosseini, M. R., Chileshe, N., Baroudi, B., Zuo, J., & Mills, A. (2016). Factors affecting perceived level of virtuality in hybrid construction project teams (HCPTs). *Construction Innovation*, 16(4), 460-482. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1844297065?accountid=11243

Hsu, C., Backhouse, J., & Silva, L. (2014). Institutionalizing operational risk management: An empirical study. *Journal of Information Technology*, 29(1), 59-72. Doi:http://dx.doi.org/10.1057/jit.2013.15

Hughes, D. L., Rana, N. P., & Simintiras, A. C. (2017c). The changing landscape of IS project failure: An examination of the key factors. *Journal of Enterprise Information Management*, 30(1), 142-165. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1871543158?accountid=11243

Iyer, K.C., & Jha, K.N. (2005). Factors affecting cost performance: evidence from Indian construction projects. *International Journal of Project Management*, (23) 4. 283-295.

Jørgensen, M. (2014a). Failure factors of small software projects at a global outsourcing marketplace. *The Journal of Systems and Software*, 92, 157. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1518986390?accountid=11243

Joslin, R., & Müller, R. (2016c). The impact of project methodologies on project success in different project environments. *International Journal of Managing Projects in Business*, 9(2), 364-388. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1774537976?accountid=11243

Joshi, H., Chawla, D., & Farooquie, J.,A. (2014). Segmenting knowledge management (KM) practitioners and its relationship to performance variation: Some empirical evidence. *Journal of Knowledge Management*, 18(3), 469-493. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1660768340?accountid=11243

Kang, Y., & Stasko, J. (2014). Characterizing the intelligence analysis process through a longitudinal field study: Implications for visual analytics. *Information Visualization*, 13(2), 134-158. Doi:http://dx.doi.org/10.1177/1473871612468877

Kappelman, L. A., McKeeman, R., & Zhang, L. (2007). Early warning signs of IT project failure: The dominant dozen. *Edpacs*, 35(1), 1-10. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/234908047?accountid=11243

Kealey, D. J., Protheroe, D. R., MacDonald, D., & Vulpe, T. (2006b). International projects: Some lessons on avoiding failure and maximizing success. *Performance Improvement*, 45(3), 38-46. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/237245395?accountid=11243

Kearns, G. S. (2004). A multi-objective, multi-criteria approach for evaluating IT investments: Results from two case studies. *Information Resources Management Journal*, 17(1), 37-62. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/215883620?accountid=11243

Keil, M., Smith, H. J., Iacovou, C. L., & Thompson, R. L. (2014). The dynamics of IT project status reporting: A self-reinforcing cycle of distrust. *Journal of the Association for Information Systems*, 15(12), 879-912. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1642336745?accountid=11243

Kessler, E. H., & Chakrabarti, A. K. (1996). Innovation speed: A conceptual model of context, antecedents, and outcomes. *Academy of Management: the Academy of Management Review*, 21(4), 1143. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/210973172?accountid=11243

Kloppenborg, T. J., Stubblebine, P. C., & Tesch, D. (2007). Project manager vs. executive perceptions of sponsor behaviors. *Management Research News*, 30(11), 803-815. Doi:http://dx.doi.org/10.1108/01409170710832241

Kutsch, E., Browning, T. R., & Hall, M. (2014c). Bridging the risk gap: The failure of risk management in information systems projects. *Research Technology Management*, 57(2), 26-32. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1507797846?accountid=11243

Laumer, S., Maier, C., & Weitzel, T. (2017). Information quality, user satisfaction, and the manifestation of workarounds: A qualitative and quantitative study of enterprise content management system users. *European Journal of Information Systems*, 26(4), 333-360. Doi:http://dx.doi.org/10.1057/s41303-016-0029-7

Lempinen, H., & Rajala, R. (2014). Exploring multi-actor value creation in IT service processes. *Journal of Information Technology*, 29(2), 170-185. Doi:http://dx.doi.org/10.1057/jit.2014.1

Lessing, J., & Brege, S. (2015). Business models for product-oriented house-building companies: Experience from two Swedish case studies. *Construction Innovation*, 15(4), 449-472. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1709708069?accountid=11243

Letens, G., Farris, J. A., & Van Aken, E.,M. (2011). A multilevel framework for lean product development system design. *Engineering Management Journal*, 23(1), 69-85. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/862561357?accountid=11243

Levina, N. (2005). Collaborating on multiparty information systems development projects: A collective reflection-in-action view. *Information Systems Research*, 16(2), 109-130. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/208178705?accountid=11243

Lim, W., Sia, S. K., & Yeow, A. (2011). Managing risks in a failing IT project: A social constructionist view*. *Journal of the Association for Information Systems*, 12(6), 374-400. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/876278470?accountid=11243

Lin, H., & Su, J. (2014). A case study on adoptive management innovation in China. *Journal of Organizational Change Management*, 27(1), 83-114. Doi:http://dx.doi.org/10.1108/JOCM-07-2012-0112

Loye, R. (2008). Requirement for knowledge management: Business driving information technology. *Journal of Knowledge Management*, 12(3), 156-168. Doi:http://dx.doi.org/10.1108/13673270810875930

Nguyen, Luong H.; Watanabe, Tsunemi. (2017). The Impact of Project Organizational Culture on the Performance of Construction Projects. Sustainability 2017, 9(5), 781; https://doi.org/10.3390/su9050781

Maghsoodi, A. I., & Khalilzadeh, M. (2018). Identification and evaluation of construction projects' critical success factors employing fuzzy-TOPSIS approach. *KSCE Journal of Civil Engineering*, 22(5), 1593-1605. Doi://dx.doi.org.proxygw.wrlc.org/10.1007/s12205-017-1970-2

Mahaney, R. C., & Lederer, A. L. (2011a). An agency theory explanation of project success. *The Journal of Computer Information Systems*, 51(4), 102-113. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/884811293?accountid=11243

Mainga, W. (2017). Examining project learning, project management competencies, and project efficiency in project-based firms (PBFs). *International Journal of Managing Projects in Business*, 10(3), 454-504. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1907223744?accountid=11243

Marsh, D. J. (2010a). How to save a failing project: Chaos to control. *Project Management Journal*, 41(1), 89. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/218776663?accountid=11243

Marthandam, N. S., Calvo-Amodio, J., PhD, & Ng, E. H., PhD. (2014). Terminating a failed project without affecting the market value of the firm. *IIE Annual Conference. Proceedings*, 3552-3563. Retrieved from

http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1622308755?accountid=11243

Mayhew, H., Saleh, T., & Williams, S. (2016). Making data analytics work for you – instead of the other way around. *McKinsey Quarterly*.

McGrath, J., & Martin, P. (2017). Why we want project failure. *Manufacturing Business Technology*, n/a. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1931470023?accountid=11243

McKemmish, S. (2013). Recordkeeping and archiving in the cloud. Is there a silver lining?. Actas Infuture, 17-29.

McManus, J., & Wood-Harper, T. (2003). Information systems project management: The price of failure. *Management Services*, 47(5), 16. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/234336458?accountid=11243

Meshing, C. (2013). Why tech projects fail: 5 unspoken reasons. *InformationWeek*, (1363), 8,10. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1355717174?accountid=11243

Mpofu, B., Ochieng, E. G., Moobela, C., & Pretorius, A. (2017). Profiling causative factors leading to construction project delays in the United Arab Emirates. *Engineering, Construction and Architectural Management*, 24(2), 346-376. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1879020476?accountid=11243

Mukerjee, H. S., & Prasad, U. D. (2017c). Definitions of project success in implementation of customer relationship management (CRM) information technology (IT) solutions: Perspectives of consultants from India †. *South Asian Journal of Management*, 24(4), 142-157. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/2018570876?accountid=11243

New, D. E., & Schlacter, J. L. (1979). Abandon bad R&D projects with earlier marketing appraisals. *Industrial Marketing Management*, 8(4), 274. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/204490804?accountid=11243

Pickett, Todd W., C.C.C., and Bruce G. Elliott C.C.C. 2007. Transforming historical project data into useful information. AACE International Transactions: ES21-ES26, http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/208198904?accountid=11243

Pai, D. R., Subramanian, G. H., & Pendharkar, P. C. (2015). Benchmarking software development productivity of CMMI level 5 projects. *Information Technology and Management*, 16(3), 235-251. Doi:http://dx.doi.org/10.1007/s10799-015-0234-4

Park, C., Im, G., & Keil, M. (2008a). Overcoming the mum effect in IT project reporting: Impacts of fault responsibility and time urgency*. *Journal of the Association for Information Systems*, 9(7), 409-431. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/198832295?accountid=11243

Parkes, A., & Davern, M. (2011). A challenging success: A process audit perspective on change. *Business Process Management Journal*, 17(6), 876-897. Doi://dx.doi.org.proxygw.wrlc.org/10.1108/14637151111182675

Pinto, J. K., & Kharbanda, O. P. (1996). How to fail in project management (without really trying). *Business Horizons*, 39(4), 45. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/195349795?accountid=11243

Pecori, R. (2018). A virtual learning architecture enhanced by fog computing and big data streams. *Future Internet*, 10(1), 4. Doi:http://dx.doi.org/10.3390/fi10010004

Purvis, R. L., McCray, G. E., & Roberts, T. L. (2004). Heuristics and biases in information systems project management. *Engineering Management Journal*, 16(2), 19-27. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/208955814?accountid=11243

Rahmana, N., Kumarb, N., & Rutzc, D. (2016). Managing application compatibility during ETL tools and environment upgrades. Journal of Decision Systems, 25(2), 136-150. Doi:http://dx.doi.org/10.1080/12460125.2016.1138392

Roger, S. P. (2009). “Software Engineering – A Practioner’s Approach Seventh Edition“, McGraw Hill International Edition, p.698.

Rouibah, K., & Al-Rafee, S. (2009). Requirement engineering elicitation methods. Information Management & Computer Security, 17(3), 192-217. Doi:http://dx.doi.org/10.1108/09685220910978086

Schoenmaker, R., & de Bruijn, H. (2016). Embracing complexity in performance-based contracts for road maintenance. *International Journal of Productivity and Performance Management*, 65(1), 4-24. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1748546855?accountid=11243

Seth, F. P., Mustonen-ollila, E., Taipale, O., & Smolander, K. (2015). Software quality construction in 11 companies: An empirical study using the grounded theory. *Software Quality Journal*, 23(4), 627-660. Doi:http://dx.doi.org/10.1007/s11219-014-9246-2

Sheard, A. G., & Kakabadse, A. P. (2004). A process perspective on leadership and team development. *The Journal of Management Development*, 23(1), 7-11,13-41,43-79,81-106. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/216365855?accountid=11243

Siemieniuch, C. E., & Sinclair, M. A. (2004). A framework for 111um111unted111d111al readiness for knowledge management. *International Journal of Operations & Production Management*, 24(1), 79-98. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/232346905?accountid=11243

Simon, S. J. (2005). Balanced scorecard: A tool to improve IS department planning and evaluation. *Journal of Information Technology Case and Application Research*, 7(4), 7-29. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/21489658 7?accountid=11243

Sohel, S. M., & Quader, M. S. (2017). Transforming IT from a cost centre to a value centre perspective: A case study on the British Standards Institute. *Journal of Services Research*, 17(1), 71-105. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/19880239 89?accountid=11243

Souder, W. E., & Chakrabarti, A. K. (1979). Industrial innovations: A demographical analysis. *IEEE Transactions on Engineering Management*, EM26(4), 101. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/195641963?accountid=11243

Standing, O., Standing, S., & Kordt, E. (2016). Explaining attribution in information technology projects. *Journal of Systems and Information Technology*, 18(2), 216-227. Retrieved from http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1823117075?accountid=11243

Stayton, J., & Mangematin, V. (2016). Startup time, innovation and organizational emergence: A study of USA-based international technology ventures. *Journal of International Entrepreneurship*, 14(3), 373-409. Doi:http://dx.doi.org/10.1007/s10843-016-0183-y

Tesch, D., Kloppenborg, T. J., & Frolick, M. N. (2007). It project risk factors: The project management professionals perspective. *The Journal of Computer Information Systems*, 47(4), 61-69. Retrieved from

http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/232575798?accountid=11243

Terouhid, S. A., & Ries, R. (2016). People capability. *Journal of Modelling in Management*, 11(3), 811-841. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/1841766132?accountid=11243

Tichacek, R. L. (2006). Effective cost management: Back to basics. *Cost Engineering*, 48(3), 27-33. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/220469055?accountid=11243

Tingshen Zhao, Seyed Ebrahim Kazemi, Wen Liu, and Miao Zhang. (2018). The Last Mile: Safety Management Implementation in Construction Sites. *Advances in Civil Engineering*, vol. 2018, Article ID 4901707, 15 pages, 2018. https://doi.org/10.1155/2018/4901707.

Udechukwu Ojiako, Papadopoulos, T., Chonnikarn Thumborisuthi, & Yun Fan Yang. (2012b). Perception variability for 113um113unted113d risk factors. *Industrial Management & Data Systems*, 112(4), 600-618. Doi://dx.doi.org.proxygw.wrlc.org/10.1108/02635571211225503

Vallejo, C., Romero, D., & Molina, A. (2017). Implementation of best manufacturing practices using logic models and system dynamics: Project design and project assessment views. *Information Systems and eBusiness Management*, 15(2), 535-575. Doi:http://dx.doi.org/10.1007/s10257-016-0327-6

Wiewiora, Anna, and Glen Murphy. (2015). Unpacking ‘lessons learned’: Investigating failures and considering alternative solutions. *Knowledge Management Research & Practice* 13, (1) (02): 17-30, http://proxygw.wrlc.org/login?url=https://search-proquest-com.proxygw.wrlc.org/docview/1652312158?accountid=11243 (accessed February 22, 2019).

Wilkin, C. L., & Riddett, J. (2009). IT governance challenges in a large not-for-profit healthcare organization: The role of intranets. *Electronic Commerce Research*, 9(4), 351-374. Doi:http://dx.doi.org/10.1007/s10660-009-9038-0

Wu, D., & Passerini, K. (2013). Uncovering knowledge-based time management practices. *International Journal of Managing Projects in Business*, 6(2), 332-348. Doi:http://dx.doi.org/10.1108/17538371311319052

Yeoh, W., Koronios, A., & Gao, J. (2008). Managing the implementation of business intelligence systems: A critical success factors framework. *International Journal of Enterprise Information Systems*, 4(3), 79-94. Retrieved from http://proxygw.wrlc.org/login?url=https://search.proquest.com/docview/222678639?accountid=11243

Appendix A – R Script

```
# Mohammed Altaf Hossain's Praxis script            #
# Email: altaf@gwu.edu & altaf@altaf.com             #
# Phone: 703.597.0000                                #

rm(list = ls())
gc()
library(corrplot)
library(car)
library(MASS)
# library(caret)

setwd("C:/Users/Altaf/Google Drive/D. Eng. Class of 2018/Fall 2018 (FINAL)")
Data <- read.csv(file = "MSDB final.csv", head=TRUE)

colnames(Data) <- c("YR","PID","BID","UNK1","BID2","UNK2","BT","MA","SMA","BN","EID",
          "BID3","WL","BS","OS","PF","LG","ST","TVR","PG12","PW12","PG","PW","RI",
          "CA","OT","CMT","RR","PRM","CNT","PT","EU12","EU","SPT","G1","G2",
          "G3","G4","PS","PD","P1","P2","P3","P4","TH","BH","BT1","BT2","PNK",
          "RED","GLD","BDL12","CM12","PM12","CE12","PE12","BDL","CM","PM","CE",
          "PE","BV","PTW","RB","BP","G4R","CBR","EAC","CAR","UNK3","TBPSH",
          "BPH","SH","IRDH","TBOEH","TENGBOEH","TENGBPH","TVRE","FTVRE","CMPTRS",
          "CEBHP","CEFBHP","PRC","RP","FRC","POC","OPP","FOC","LM","NG","BAE",
          "SAIC","GD","IBM","BOE","BAH")

summary(Data)
data.summ <- summary(Data)
write.csv(data.summ, file = "summary of data final.csv", row.names=TRUE)

# unique(Data$YR) #outputs all unique values.
# length(unique(Data$YR)) #output total numbe of unique value.
# table(Data$YR) #outputs all unique values with number it appears.

Rm(list = ls())
gc()
library(corrplot)
library(car)
library(MASS)
# library(caret)

setwd("C:/Users/Altaf/Google Drive/D. Eng. Class of 2018/Fall 2018 (FINAL)")
Data <- read.csv(file = "MSDB final 07-18.csv", head=TRUE)

colnames(Data) <- c("YR","PID","BID","UNK1","BID2","UNK2","BT","MA","SMA","BN","EID",
          "BID3","WL","BS","OS","PF","LG","ST","TVR","PG12","PW12","PG","PW","RI",
          "CA","OT","CMT","RR","PRM","CNT","PT","EU12","EU","SPT","G1","G2",
          "G3","G4","PS","PD","P1","P2","P3","P4","TH","BH","BT1","BT2","PNK",
          "RED","GLD","BDL12","CM12","PM12","CE12","PE12","BDL","CM","PM","CE",
          "PE","BV","PTW","RB","BP","G4R","CBR","EAC","CAR","UNK3","TBPSH",
          "BPH","SH","IRDH","TBOEH","TENGBOEH","TENGBPH","TVRE","FTVRE","CMPTRS",
          "CEBHP","CEFBHP","PRC","RP","FRC","POC","OPP","FOC","LM","NG","BAE",
          "SAIC","GD","IBM","BOE","BAH")
summary(Data)
data.summ <- summary(Data)
write.csv(data.summ, file = "summary of data final 07-18.csv", row.names=TRUE)

# unique(Data$YR) #outputs all unique values.
# length(unique(Data$YR)) #output total numbe of unique value.
# table(Data$YR) #outputs all unique values with number it appears.

# BID,BID2,BT,MA,SMA,BN,BID3,WL,BS,PF,LG,ST,RI,CA,OT,CMT,RR,PRM,CNT,PT,EU,SPT,
# G1,G2,G3,G4,PS,PD,BH,BT1,BT2,PNK,RED,GLD,BDL12,CM12,PM12,CE12,PE12,BDL,CM,PM,
# CE,PE,EAC,CAR,UNK3,CMPTRS,CEBHP,CEFBHP,LM,NG,BAE,SAIC,GD,IBM,BOE,BAH
```

```
Data$BID <- as.numeric(Data$BID)
Data$BID2 <- as.numeric(Data$BID2)
Data$BT <- as.numeric(Data$BT)
Data$MA <- as.numeric(Data$MA)
Data$SMA <- as.numeric(Data$SMA)
Data$BN <- as.numeric(Data$BN)
Data$BID3 <- as.numeric(Data$BID3)
Data$WL <- as.numeric(Data$WL)
Data$BS <- as.numeric(Data$BS)
Data$OS <- as.numeric(Data$OS)
Data$PF <- as.numeric(Data$PF)
Data$LG <- as.numeric(Data$LG)
Data$ST <- as.numeric(Data$ST)
Data$PG12 <- as.numeric(Data$PG12)
Data$PW12 <- as.numeric(Data$PW12)
Data$RI <- as.numeric(Data$RI)
Data$CA <- as.numeric(Data$CA)
Data$OT <- as.numeric(Data$OT)
Data$CMT <- as.numeric(Data$CMT)
Data$RR <- as.numeric(Data$RR)
Data$PRM <- as.numeric(Data$PRM)
Data$CNT <- as.numeric(Data$CNT)
Data$PT <- as.numeric(Data$PT)
Data$EU12 <- as.numeric(Data$EU12)
Data$EU <- as.numeric(Data$EU)
Data$SPT <- as.numeric(Data$SPT)
Data$G1 <- as.numeric(Data$G1)
Data$G2 <- as.numeric(Data$G2)
Data$G3 <- as.numeric(Data$G3)
Data$G4 <- as.numeric(Data$G4)
Data$PS <- as.numeric(Data$PS)
Data$PD <- as.numeric(Data$PD)
Data$BH <- as.numeric(Data$BH)
Data$BT1 <- as.numeric(Data$BT1)
Data$BT2 <- as.numeric(Data$BT2)
Data$PNK <- as.numeric(Data$PNK)
Data$RED <- as.numeric(Data$RED)
Data$GLD <- as.numeric(Data$GLD)
Data$BDL12 <- as.numeric(Data$BDL12)
Data$CM12 <- as.numeric(Data$CM12)
Data$PM12 <- as.numeric(Data$PM12)
Data$CE12 <- as.numeric(Data$CE12)
Data$PE12 <- as.numeric(Data$PE12)
Data$BDL <- as.numeric(Data$BDL)
Data$CM <- as.numeric(Data$CM)
Data$PM <- as.numeric(Data$PM)
Data$CE <- as.numeric(Data$CE)
Data$PE <- as.numeric(Data$PE)
Data$EAC <- as.numeric(Data$EAC)
Data$CAR <- as.numeric(Data$CAR)
Data$UNK3 <- as.numeric(Data$UNK3)
Data$CMPTRS <- as.numeric(Data$CMPTRS)
Data$CEBHP <- as.numeric(Data$CEBHP)
Data$CEFBHP <- as.numeric(Data$CEFBHP)
Data$LM <- as.numeric(Data$LG)
Data$NG <- as.numeric(Data$NG)
Data$BAE <- as.numeric(Data$BAE)
Data$SAIC <- as.numeric(Data$SAIC)
Data$GD <- as.numeric(Data$GD)
Data$IBM <- as.numeric(Data$IBM)
Data$BOE <- as.numeric(Data$BOE)
Data$BAH <- as.numeric(Data$BAH)

write.csv(Data, file = "MSDB Final – numeric 07-18.csv", row.names=TRUE)


# DT                                                                                          #
===============================================================================================
rm(list = ls())
gc()
```

```
library(corrplot)
library(car)
library(MASS)
library(caret)
library(rpart)
library(rpart.plot)
# library(party)
#
# library(corrplot)
# library(car)
# library(MASS)
# library(caret)

setwd("C:/Users/Altaf/Google Drive/D. Eng. Class of 2018/Fall 2018 (FINAL)")
Data.raw <- read.csv(file = "MSDB Final – numeric 07-18.csv", head=TRUE)

DataP <- Data.raw[Data.raw$PF==2,]
DataF <- Data.raw[Data.raw$PF==1,]

DataP$PF[DataP$PF==2] <- "Passed"
DataF$PF[DataF$PF==1] <- "Failed"
Data <- rbind(DataP, DataF)

Data07 <- Data[Data$YR==2007,]
Data08 <- Data[Data$YR==2008,]
Data09 <- Data[Data$YR==2009,]
Data10 <- Data[Data$YR==2010,]
Data11 <- Data[Data$YR==2011,]
Data12 <- Data[Data$YR==2012,]
Data13 <- Data[Data$YR==2013,]
Data14 <- Data[Data$YR==2014,]
Data15 <- Data[Data$YR==2015,]
Data16 <- Data[Data$YR==2016,]
Data17 <- Data[Data$YR==2017,]

Data07 <- Data07[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data08 <- Data08[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data09 <- Data09[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data10 <- Data10[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data11 <- Data11[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data12 <- Data12[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data13 <- Data13[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data14 <- Data14[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data15 <- Data15[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data16 <- Data16[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data17 <- Data17[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]

Data79 <- rbind(Data07, Data08, Data09)
Data810 <- rbind(Data08, Data09, Data10)
Data911 <- rbind(Data09, Data10, Data11)
Data1012 <- rbind(Data10, Data11, Data12)
Data1113 <- rbind(Data11, Data12, Data13)
Data1214 <- rbind(Data12, Data13, Data14)
Data1315 <- rbind(Data13, Data14, Data15)
Data1416 <- rbind(Data14, Data15, Data16)

lda79 <- rpart(PF ~ ., Data79)
lda810 <- rpart(PF ~ ., Data810)
lda911 <- rpart(PF ~ ., Data911)
lda1012 <- rpart(PF ~ ., Data1012)
lda1113 <- rpart(PF ~ ., Data1113)
lda1214 <- rpart(PF ~ ., Data1214)
lda1315 <- rpart(PF ~ ., Data1315)
lda1416 <- rpart(PF ~ ., Data1416)

rpart.plot(lda79, type=4, extra = 101, cex = .5)
rpart.plot(lda810, type=4, extra = 101, cex = .5)
rpart.plot(lda911, type=4, extra = 101, cex = .5)
rpart.plot(lda1012, type=4, extra = 101, cex = .5)
rpart.plot(lda1113, type=4, extra = 101, cex = .5)
```

```
rpart.plot(lda1214, type=4, extra = 101, cex = .5)
rpart.plot(lda1315, type=4, extra = 101, cex = .5)
rpart.plot(lda1416, type=4, extra = 101, cex = .5)

P10 <- predict(lda79, Data10, type = “class”)
P11 <- predict(lda810, Data11, type = “class”)
P12 <- predict(lda911, Data12, type = “class”)
P13 <- predict(lda1012, Data13, type = “class”)
P14 <- predict(lda1113, Data14, type = “class”)
P15 <- predict(lda1214, Data15, type = “class”)
P16 <- predict(lda1315, Data16, type = “class”)
P17 <- predict(lda1416, Data17, type = “class”)

T10 <- table(prediction = P10, actual = Data10$PF)
T11 <- table(prediction = P11, actual = Data11$PF)
T12 <- table(prediction = P12, actual = Data12$PF)
T13 <- table(prediction = P13, actual = Data13$PF)
T14 <- table(prediction = P14, actual = Data14$PF)
T15 <- table(prediction = P15, actual = Data15$PF)
T16 <- table(prediction = P16, actual = Data16$PF)
T17 <- table(prediction = P17, actual = Data17$PF)

T10
T11
T12
T13
T14
T15
T16
T17

# Accuracy metric
sum(diag(T10)/sum(T10))
sum(diag(T11)/sum(T11))
sum(diag(T12)/sum(T12))
sum(diag(T13)/sum(T13))
sum(diag(T14)/sum(T14))
sum(diag(T15)/sum(T15))
sum(diag(T16)/sum(T16))
sum(diag(T17)/sum(T17))

DataP <- Data.raw[Data.raw$PF==2,]
DataF <- Data.raw[Data.raw$PF==1,]
Data <- rbind(DataP, DataF)

Data <- Data[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]#14,17,19,23,24,28,29,30,31,34,35,42,43)]

pca <-  prcomp(Data, scale = TRUE)
summary(pca)
screeplot(pca)

efa <- factanal(Data,factors = 8, scores = “regression”)
efa

predict(lda1416, newdata =
data.frame(WL=6,ST=5,PG=80,PW=70,CMT=5,RR=4,PRM=1,CNT=3,EU=10,SPT=5,P1=1000,P2=2000))


# DT – END                                                                    #
# LDA                                                                         #
```

```
rm(list = ls())
gc()
library(corrplot)
library(car)
library(MASS)
library(caret)

setwd(“C:/Users/Altaf/Google Drive/D. Eng. Class of 2018/Fall 2018 (FINAL)”)
Data.raw <- read.csv(file = “MSDB Final – numeric 07-18.csv”, head=TRUE)
```

```
DataP <- Data.raw[Data.raw$PF==2,]
DataF <- Data.raw[Data.raw$PF==1,]
Data <- rbind(DataP, DataF)

Data07 <- Data[Data$YR==2007,]
Data08 <- Data[Data$YR==2008,]
Data09 <- Data[Data$YR==2009,]
Data10 <- Data[Data$YR==2010,]
Data11 <- Data[Data$YR==2011,]
Data12 <- Data[Data$YR==2012,]
Data13 <- Data[Data$YR==2013,]
Data14 <- Data[Data$YR==2014,]
Data15 <- Data[Data$YR==2015,]
Data16 <- Data[Data$YR==2016,]
Data17 <- Data[Data$YR==2017,]

Data07 <- Data07[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data08 <- Data08[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data09 <- Data09[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data10 <- Data10[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data11 <- Data11[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data12 <- Data12[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data13 <- Data13[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data14 <- Data14[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data15 <- Data15[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data16 <- Data16[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data17 <- Data17[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]

Data79 <- rbind(Data07, Data08, Data09)
Data810 <- rbind(Data08, Data09, Data10)
Data911 <- rbind(Data09, Data10, Data11)
Data1012 <- rbind(Data10, Data11, Data12)
Data1113 <- rbind(Data11, Data12, Data13)
Data1214 <- rbind(Data12, Data13, Data14)
Data1315 <- rbind(Data13, Data14, Data15)
Data1416 <- rbind(Data14, Data15, Data16)

lda79 <- lda(PF ~ ., Data79)
lda810 <- lda(PF ~ ., Data810)
lda911 <- lda(PF ~ ., Data911)
lda1012 <- lda(PF ~ ., Data1012)
lda1113 <- lda(PF ~ ., Data1113)
lda1214 <- lda(PF ~ ., Data1214)
lda1315 <- lda(PF ~ ., Data1315)
lda1416 <- lda(PF ~ ., Data1416)

P10 <- predict(lda79, Data10)
P11 <- predict(lda810, Data11)
P12 <- predict(lda911, Data12)
P13 <- predict(lda1012, Data13)
P14 <- predict(lda1113, Data14)
P15 <- predict(lda1214, Data15)
P16 <- predict(lda1315, Data16)
P17 <- predict(lda1416, Data17)

mean(P10$class==Data10$PF)
mean(P11$class==Data11$PF)
mean(P12$class==Data12$PF)
mean(P13$class==Data13$PF)
mean(P14$class==Data14$PF)
mean(P15$class==Data15$PF)
mean(P16$class==Data16$PF)
mean(P17$class==Data17$PF)

T10 <- table(Actual = Data10$PF, Prediction = P10$class)
T11 <- table(Actual = Data11$PF, Prediction = P11$class)
T12 <- table(Actual = Data12$PF, Prediction = P12$class)
T13 <- table(Actual = Data13$PF, Prediction = P13$class)
T14 <- table(Actual = Data14$PF, Prediction = P14$class)
```

```
T15 <- table(Actual = Data15$PF, Prediction = P15$class)
T16 <- table(Actual = Data16$PF, Prediction = P16$class)
T17 <- table(Actual = Data17$PF, Prediction = P17$class)

T10
T11
T12
T13
T14
T15
T16
T17
# Accuracy metric
sum(diag(T10)/sum(T10))
sum(diag(T11)/sum(T11))
sum(diag(T12)/sum(T12))
sum(diag(T13)/sum(T13))
sum(diag(T14)/sum(T14))
sum(diag(T15)/sum(T15))
sum(diag(T16)/sum(T16))
sum(diag(T17)/sum(T17))

Data.bkup <- Data
Data <- Data[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
# write.csv(Data, file = "for spss final 07-18.csv", row.names=TRUE)
pca <-  prcomp(Data, scale = TRUE)
summary(pca)
screeplot(pca)

efa <- factanal(Data,factors = 8, scores = "regression")
efa

predict(lda1416, newdata =
data.frame(WL=6,ST=5,PG=80,PW=70,CMT=5,RR=4,PRM=1,CNT=3,EU=10,SPT=5,P1=1000,P2=2000))

t.test(Data.bkup$P1, Data.bkup$P2, conf.level = 0.95)
t.test(Data.bkup$P1, Data.bkup$P3, conf.level = 0.95)
t.test(Data.bkup$P1, Data.bkup$P4, conf.level = 0.95)
t.test(Data.bkup$P2, Data.bkup$P1, conf.level = 0.95)
t.test(Data.bkup$P2, Data.bkup$P3, conf.level = 0.95)
t.test(Data.bkup$P2, Data.bkup$P4, conf.level = 0.95)
t.test(Data.bkup$P3, Data.bkup$P1, conf.level = 0.95)
t.test(Data.bkup$P3, Data.bkup$P2, conf.level = 0.95)
t.test(Data.bkup$P3, Data.bkup$P4, conf.level = 0.95)
t.test(Data.bkup$P4, Data.bkup$P1, conf.level = 0.95)
t.test(Data.bkup$P4, Data.bkup$P2, conf.level = 0.95)
t.test(Data.bkup$P4, Data.bkup$P3, conf.level = 0.95)


# LDA – END                                                                   #
# QDA                                                                         #
```

```
rm(list = ls())
gc()
library(corrplot)
library(car)
library(MASS)

setwd("C:/Users/Altaf/Google Drive/D. Eng. Class of 2018/Fall 2018 (FINAL)")
Data.raw <- read.csv(file = "MSDB Final – numeric 07-18.csv", head=TRUE)

DataP <- Data.raw[Data.raw$PF==2,]
DataF <- Data.raw[Data.raw$PF==1,]
Data <- rbind(DataP, DataF)

Data07 <- Data[Data$YR==2007,]
Data08 <- Data[Data$YR==2008,]
Data09 <- Data[Data$YR==2009,]
Data10 <- Data[Data$YR==2010,]
Data11 <- Data[Data$YR==2011,]
```

```
Data12 <- Data[Data$YR==2012,]
Data13 <- Data[Data$YR==2013,]
Data14 <- Data[Data$YR==2014,]
Data15 <- Data[Data$YR==2015,]
Data16 <- Data[Data$YR==2016,]
Data17 <- Data[Data$YR==2017,]

Data07 <- Data07[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data08 <- Data08[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data09 <- Data09[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data10 <- Data10[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data11 <- Data11[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data12 <- Data12[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data13 <- Data13[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data14 <- Data14[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data15 <- Data15[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data16 <- Data16[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data17 <- Data17[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]

Data79 <- rbind(Data07, Data08, Data09)
Data810 <- rbind(Data08, Data09, Data10)
Data911 <- rbind(Data09, Data10, Data11)
Data1012 <- rbind(Data10, Data11, Data12)
Data1113 <- rbind(Data11, Data12, Data13)
Data1214 <- rbind(Data12, Data13, Data14)
Data1315 <- rbind(Data13, Data14, Data15)
Data1416 <- rbind(Data14, Data15, Data16)

lda79 <- qda(PF ~ ., Data79)
lda810 <- qda(PF ~ ., Data810)
lda911 <- qda(PF ~ ., Data911)
lda1012 <- qda(PF ~ ., Data1012)
lda1113 <- qda(PF ~ ., Data1113)
lda1214 <- qda(PF ~ ., Data1214)
lda1315 <- qda(PF ~ ., Data1315)
lda1416 <- qda(PF ~ ., Data1416)

P10 <- predict(lda79, Data10)
P11 <- predict(lda810, Data11)
P12 <- predict(lda911, Data12)
P13 <- predict(lda1012, Data13)
P14 <- predict(lda1113, Data14)
P15 <- predict(lda1214, Data15)
P16 <- predict(lda1315, Data16)
P17 <- predict(lda1416, Data17)

mean(P10$class==Data10$PF)
mean(P11$class==Data11$PF)
mean(P12$class==Data12$PF)
mean(P13$class==Data13$PF)
mean(P14$class==Data14$PF)
mean(P15$class==Data15$PF)
mean(P16$class==Data16$PF)
mean(P17$class==Data17$PF)

T10 <- table(Actual = Data10$PF, Prediction = P10$class)
T11 <- table(Actual = Data11$PF, Prediction = P11$class)
T12 <- table(Actual = Data12$PF, Prediction = P12$class)
T13 <- table(Actual = Data13$PF, Prediction = P13$class)
T14 <- table(Actual = Data14$PF, Prediction = P14$class)
T15 <- table(Actual = Data15$PF, Prediction = P15$class)
T16 <- table(Actual = Data16$PF, Prediction = P16$class)
T17 <- table(Actual = Data17$PF, Prediction = P17$class)

T10
T11
T12
T13
T14
T15
```

```
T16
T17

Data.bkup <- Data
Data <- Data[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]

pca <-  prcomp(Data, scale = TRUE)
summary(pca)
screeplot(pca)

efa <- factanal(Data,factors = 8, scores = "regression")
efa

predict(lda1416, newdata =
data.frame(WL=6,ST=5,PG=80,PW=70,CMT=5,RR=4,PRM=1,CNT=3,EU=10,SPT=5,P1=1000,P2=2000))


# QDA – END                                                                 #
# SVM                                                                       #
```

```
rm(list = ls())
gc()
library(e1071)
library(ggplot2)

setwd("C:/Users/Altaf/Google Drive/D. Eng. Class of 2018/Fall 2018 (FINAL)")
Data.raw <- read.csv(file = "MSDB Final – numeric 07-18.csv", head=TRUE)

DataP <- Data.raw[Data.raw$PF==2,]
DataF <- Data.raw[Data.raw$PF==1,]
DataP$PF[DataP$PF==2] <- "Passed"
DataF$PF[DataF$PF==1] <- "Failed"
Data <- rbind(DataP, DataF)

Data07 <- Data[Data$YR==2007,]
Data08 <- Data[Data$YR==2008,]
Data09 <- Data[Data$YR==2009,]
Data10 <- Data[Data$YR==2010,]
Data11 <- Data[Data$YR==2011,]
Data12 <- Data[Data$YR==2012,]
Data13 <- Data[Data$YR==2013,]
Data14 <- Data[Data$YR==2014,]
Data15 <- Data[Data$YR==2015,]
Data16 <- Data[Data$YR==2016,]
Data17 <- Data[Data$YR==2017,]

Data07 <- Data07[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data08 <- Data08[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data09 <- Data09[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data10 <- Data10[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data11 <- Data11[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data12 <- Data12[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data13 <- Data13[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data14 <- Data14[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data15 <- Data15[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data16 <- Data16[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]
Data17 <- Data17[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]

Data79 <- rbind(Data07, Data08, Data09)
Data810 <- rbind(Data08, Data09, Data10)
Data911 <- rbind(Data09, Data10, Data11)
Data1012 <- rbind(Data10, Data11, Data12)
Data1113 <- rbind(Data11, Data12, Data13)
Data1214 <- rbind(Data12, Data13, Data14)
Data1315 <- rbind(Data13, Data14, Data15)
Data1416 <- rbind(Data14, Data15, Data16)

lda79 <- svm(PF ~ ., data=Data79, type="C")
lda810 <- svm(PF ~ ., data=Data810, type="C")
lda911 <- svm(PF ~ ., data=Data911, type="C")
```

```
lda1012 <- svm(PF ~ ., data=Data1012, type="C")
lda1113 <- svm(PF ~ ., data=Data1113, type="C")
lda1214 <- svm(PF ~ ., data=Data1214, type="C")
lda1315 <- svm(PF ~ ., data=Data1315, type="C")
lda1416 <- svm(PF ~ ., data=Data1416, type="C")

P10 <- predict(lda79, Data10)
P11 <- predict(lda810, Data11)
P12 <- predict(lda911, Data12)
P13 <- predict(lda1012, Data13)
P14 <- predict(lda1113, Data14)
P15 <- predict(lda1214, Data15)
P16 <- predict(lda1315, Data16)
P17 <- predict(lda1416, Data17)

T10 <- table(Actual = Data10$PF, Prediction = P10)
T11 <- table(Actual = Data11$PF, Prediction = P11)
T12 <- table(Actual = Data12$PF, Prediction = P12)
T13 <- table(Actual = Data13$PF, Prediction = P13)
T14 <- table(Actual = Data14$PF, Prediction = P14)
T15 <- table(Actual = Data15$PF, Prediction = P15)
T16 <- table(Actual = Data16$PF, Prediction = P16)
T17 <- table(Actual = Data17$PF, Prediction = P17)

T10
T11
T12
T13
T14
T15
T16
T17
# Accuracy metric
sum(diag(T10)/sum(T10))
sum(diag(T11)/sum(T11))
sum(diag(T12)/sum(T12))
sum(diag(T13)/sum(T13))
sum(diag(T14)/sum(T14))
sum(diag(T15)/sum(T15))
sum(diag(T16)/sum(T16))
sum(diag(T17)/sum(T17))

DataP <- Data.raw[Data.raw$PF==2,]
DataF <- Data.raw[Data.raw$PF==1,]
Data <- rbind(DataP, DataF)

Data <- Data[c(14,17,19,23,24,28,29,30,31,34,35,42,43)]

pca <-  prcomp(Data, scale = TRUE)
summary(pca)
screeplot(pca)

efa <- factanal(Data,factors = 8, scores = "regression")
efa

predict(lda1416, newdata =
data.frame(WL=6,ST=5,PG=80,PW=70,CMT=5,RR=4,PRM=1,CNT=3,EU=10,SPT=5,P1=1000,P2=2000))


# SVM – END                                                                  #
```

Appendix B – Gate 1 Evaluation Form

Organizational Gate 1 – Evaluation Form (PGO)		
Team Leads	**Justifications**	**Scores**
1. Business Development Lead	<Likelihood of project success or failure>	<Probability of Go>
2. Requirement Lead	<Likelihood of project success or failure>	<Probability of Go>
3. User interface Lead	<Likelihood of project success or failure>	<Probability of Go>
4. Design Lead	<Likelihood of project success or failure>	<Probability of Go>
5. Software Development Lead	<Likelihood of project success or failure>	<Probability of Go>
6. Quality Assurance Lead	<Likelihood of project success or failure>	<Probability of Go>
7. Test Lead	<Likelihood of project success or failure>	<Probability of Go>
8. Principle Engineer	<Likelihood of project success or failure>	<Probability of Go>
9. Chief Engineer	<Likelihood of project success or failure>	<Probability of Go>
10. Project Sponsor	<Likelihood of project success or failure>	<Probability of Go>
	Weighted Score:	<Average Score>
<Final Decision>	<Justifications>	<Go or No-Go>
Note:		
Date:		

Acknowledgement:

1. Business Development Lead: ____________________
2. Requirement Lead: ____________________
3. User interface Lead: ____________________
4. Design Lead: ____________________
5. Software Development Lead: ____________________
6. Quality Assurance Lead: ____________________
7. Test Lead: ____________________
8. Principle Engineer: ____________________
9. Chief Engineer: ____________________
10. Project Sponsor: ____________________

Appendix C – Gate 2 Evaluation Form

Organizational Gate 2 – Evaluation Form (PWIN)		
Team Leads	**Justifications**	**Scores**
1. Business Development Lead	<Likelihood of project success or failure>	<Probability of WIN>
2. Requirement Lead	<Likelihood of project success or failure>	<Probability of WIN >
3. User interface Lead	<Likelihood of project success or failure>	<Probability of WIN >
4. Design Lead	<Likelihood of project success or failure>	<Probability of WIN >
5. Software Development Lead	<Likelihood of project success or failure>	<Probability of WIN >
6. Quality Assurance Lead	<Likelihood of project success or failure>	<Probability of WIN >
7. Test Lead	<Likelihood of project success or failure>	<Probability of WIN >
8. Principle Engineer	<Likelihood of project success or failure>	<Probability of WIN >
9. Chief Engineer	<Likelihood of project success or failure>	<Probability of WIN >
10. Project Sponsor	<Likelihood of project success or failure>	<Probability of WIN >
	Weighted Score:	<Average Score>
<Final Decision>	<Justifications>	<Go or No-Go>
Note:		
Date:		

Acknowledgement:

1. Business Development Lead: ______________________________
2. Requirement Lead: ______________________________
3. User interface Lead: ______________________________
4. Design Lead: ______________________________
5. Software Development Lead: ______________________________
6. Quality Assurance Lead: ______________________________
7. Test Lead: ______________________________
8. Principle Engineer: ______________________________
9. Chief Engineer: ______________________________
10. Project Sponsor: ______________________________

Appendix D – Sample of Raw Data (Past Project Data from the Company)

Column Names	A few Data Sets							
1. YEAR				2014				
2. Project ID	8	9	10	11	12	13	14	15
3. Booking ID	BKG-14203	BKG-18504	BKG-18608	BKG-18616	BKG-18700	BKG-18701	BKG-18702	BKG-18728
4. [Unknown]	86209	86129	85758	86348	86569	86570	86571	86602
5. UID-Booking, and Budget Type	86209BKG-14203	86129BKG-18504	85758BKG-18608	86348BKG-18616	86569BKG-18700	86570BKG-18701	86571BKG-18702	86602BKG-18728
6. [Unknown]								
7. Booking Type	Base/Initial	Originally Proposed Option/Increase (Gated with Base)	Post Award Option/Increase (Gated Separately)	Base/Initial (Gated Separately)	Other	Post Award Option/Increase (Growth not requiring a Gate)	Sustainment	0
8. MA	SCS	0	CSM	MSM	AT	TSS	GIS (IIS)	GTS
9. SubMA	0	MS	MS (MSM)	C2A (deactivated)	PS	IOM	GCS	IPX
10. ooking Name	SNEP IDIQ	Option Year 4	CTTR OY4	Data Integration OY1	Base – WHCA Eng Sppt	Base – WHCA-CMS	WHCA-EASPS	Base
11. RPM ID	86209	86129	85758	86348	86569	86570	86571	86602
12. ooking ID	BKG-14203	BKG-18504	BKG-18608	BKG-18616	BKG-18700	BKG-18701	BKG-18702	BKG-18728
13. in_Loss	In Progress	UNKNOWN-bad Opp status-sister	UNKNOWN-bad Opp status-sister	UNKNOWN-bad data-sister	In Progress	In Progress	UNKNOWN-bad data	UNKNOWN-bad Opp status
14. ooking Status	Potential Award	Potential Award	Potential Award	Potential Award	Potential Award	Potential Award	Potential Award	Potential Award
15. pportunity Status	In Pursuit	0	0	Won	In Pursuit	In Pursuit	Customer Cancelled	0
16. ast Gate	Gate 0	Gate 4	Gate 4	Gate 4	Gate 0	Gate 0	Gate 0	Gate 1
17. ales Type	Dom	Dom	Dom	Dom	Dom	Dom	Dom	Dom
18. VR	0	556057	9270	33912	3500	37500	1000	400
19. GO	0	100	100	20	25	40	50	0
20. WIN	0	100	100	100	40	50	50	0
21. FP Issue				41668.3	43059.2	43224	43059.2	
22. ontract Award Date	42795.2			41939.2	43151.2	43313	43151.2	

23. pp. Type	0	0	0	Routine	0	0	0	0
24. omp. Type	0	Competitive: Recompete	Competitive: New Program	Competitive: New Program	Competitive: Take Away	Competitive: Take Away	Competitive: Take Away	0
25. TN Role	Prime	0	0	Prime	Prime	Prime	Prime	0
26. rime	Raytheon	0	0	Raytheon	Raytheon	Raytheon	Raytheon	0
27. ontract Type	0	0	0	CPFF	FFP	FFP	FFP	0
28. nd User L1	US NAVY	USAF	US NAVY	DTRA	DISA	DISA	DISA	US NAVY
29. L_G1								
30. L_G2								
31. L_G3								
32. L_G4								
33. roposal Submitted				41697.3				
34. roposal Due				41697.3		43259		
35. 1	0	0	0	0	0	0	0	0
36. 2	0	0	0	0	0	0	0	0
37. 3	0	0	0	0	0	0	0	0
38. 4	0	0	0	0	0	0	0	0
39. HC	0	0	0	0	0	0	0	0
40. H								
41. T1								
42. T2								
43. ink								
44. ed								
45. old				41694.3				
46. D Lead	Bruce Pawelczyk (46804)	0	Chris Peterson (1015136)	Dan Conn (1036441)	Dave Altman (1010714)	Mike Spann (1074131)	Monica Bal (1069741)	Ben Krug (HTSP1794)
47. apture Manager	Bruce Pawelczyk (46804)	0	Chris Peterson (1015136)	Dan Conn (1036441)	Mike Willoughby (1047781)	Dennis Mclean (1029359)	Kevin Frazier (HTSP0741)	John Tobey (HTSC7417)
48. roposal Manager	0	John Tobey (HTSC7417)	Dennis Mclean (1029359)	James Burton (1012896)	Mark Thorwart (776227)	Karl Heuple (345950)	Bredt Martin (NRP0203325)	Svetlana May (1036236)

49. E	0	Michael Williams (400364)	Ken (Cyber) LAMKIN (1103920)	Chris Hohne (1074076)	Gabriel Comi (1022309)	Byron Thompson (131975)	Karen Casey (HITJ9814)	Jim Negro (HTSP018 5)
50. E	0	Michael Williams (400364)	Bob Mojazza (1051002)	Jim Negro (HTSP018 5)	Jason Horn (1117494)	Karen Casey (HITJ9814)	Jeff Driskell (1041772)	David Walker (135500)
51. ookin g Value ($K)	0	1000	1899	711	700	7500	1000	400
52. TW ($K)								
53. TN Bid ($K)								
54. id Pwin								
55. ate 4 Rate								
56. urrent Booki ng Rate								
57. AC Period								
58. ompet itive Adva ntage Ratin g								
59. Ukno wn]	,0,0,0	,0,0,0	,0,0,0	,0,0,0	,0,0,0	,0,0,0	,0,0,0	,0,0,0
60. otal B&P and Sellin g Hours	0	0	0	0	0	0	0	0
61. &P Hours	0	0	0	0	0	0	0	0
62. elling Hours	0	0	0	0	0	0	0	0
63. RAD Hours	0	0	0	0	0	0	0	0
64. otal BOE Hours for Comp leted Propo sals (Last		0	0	0				

Gate = 4)								
65. otal ENG BOE Hours for Completed Proposals (Last Gate = 4)		0	0	0				
66. otal ENG B&P Hours for Completed Proposals (Last Gate = 4)		0	0	0				
67. VR Efficiency	No TVR	No Hrs	No Hrs	No Hrs	No Hrs	No Hrs	No Hrs	No Hrs
68. actored TVR Efficiency	No TVR	No Hrs	No Hrs	No Hrs	No Hrs	No Hrs	No Hrs	No Hrs
69. ompetitors								
70. ust Engmnt (fr Gate or BH Pkgs)								
71. ust Engmnt Freq (fr Gate or BH Pkgs)								
72. otential Risk Cost impact ($K) (fr Gate pkg)								
73. isk								

Prob. (fr Gate pkg)								
74. actored Risk Cost Impact (fr Gate pkg)								
75. otential Opp Cost impact ($K)								
76. pp prob. (fr Gate pkg)								
77. actored Opp cost Impact (fr Gate pkg)								
78. M								
79. G								
80. AE								
81. AIC								
82. D								
83. BM								
84. oeing								

Appendix E – Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
BDL	Business Development Lead
BI	Business Intelligence
BKG	Booking
BS	Booking Status
BT	Budge Type
BV	Book Value
CC	Cancelled by Customer
CE	Chief Engineer
CM	Capture Manager
CMT	Competition Type
CNT	Contract Type
CSFs	Critical Success Factors
DT	Decision Tree
EBP	Evidence Based Program
ECM	Enterprise Content Management
ERP	Enterprise Resource Projects
ERPM ID	Enterprise Resource Planning and Management Identifier
EU	End User
EWSs	Early Warning Signs
FA	Factor Analysis
FP	False Positive
FPR	False Positive Rate
FMEA	Failure Mode and Effect Analysis
FN	False Negative
FNR	False Negative Rate
Fuzzy TOPSIS	Fuzzy Technique for Order Preference by Similarity to Ideal Solution
ID	Identifier
IS	Information System
IT	Information Technology
KPI	Key Performance Indicators
KS	Kolmogorov-Smirnov
LDA	Linear Discriminant Analysis
LG	Last Gate
LMS	Learning Management Systems
MA	Mission Area
MCDM	Multiple Criteria Decision Making
NA	Not Applicable/Available
NPD	New Product Development
OSHAIR	Occupational Safety and Health Administration Incidence Rate

OT	Opportunity Type
PCA	Principal Component Analysis
P1	Phase One
P2	Phase Two
P3	Phase Three
P4	Phase Four
PE	Principle Engineer
PF	Pass or Fail
PG	See PGO
PGO	Probability to Go
PIP	Project Implementation Profile
PM	Proposal Manager
PMBOK	Project Management Body of Knowledge
POC	Proof of Concept
PRM	Prime
PWIN	Probability to Win
PW	See PWIN
QDA	Quadratic Discriminant Analysis
RII	Relative Importance Index
RMM	Risk Mapping Matrix
ROI	Return on Investment
RPN	Risk Priority Number
RR	Company Role
SI	Stakeholder Involvement
SMA	Sub Mission Area
SMEs	Small and Medium Enterprises
SPT	Sub Project Type
SS	Significance Score
ST	Sales Type
SubMA	Sub Mission Area
SVM	Support Vector Machine
SW	Shapiro-Wilk
TH	See THC
THC	Total Hours Calculated
TL	Technical Lead
TP	Team Performance
TS	Time Spent
TVM	Time Value Money
TVR	Total Value in Return
UR	User Requirement
WL	Win Loss
YR	Year