IMPROVING SOFTWARE PROJECT MANAGEMENT QUALITY THROUGH THE USE OF ANALYTICS ON PROJECT MANAGEMENT DATA

by

RUTENDO NGARIRA

submitted in accordance with the requirements for
the Degree of

MASTER OF SCIENCE

in the subject

COMPUTING

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: Professor Ernest Mnkandla

CO-SUPERVISOR: Professor Nehemiah Mavetera

December 2019

DECLARATION

Name: Rutendo Ngarira

Student number: 468-936-28

Degree: MSc in Computing

Exact wording of the title of the dissertation as appearing on the electronic copy submitted for examination:

Improving Software Project Management Quality Through the Use of Analytics on Project Management Data

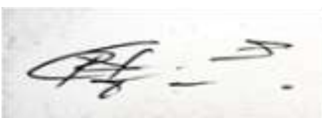
---

I declare that the above dissertation is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

I further declare that I submitted the dissertation to originality checking software and that it falls within the accepted requirements for originality.

I further declare that I have not previously submitted this work, or part of it, for examination at Unisa for another qualification or at any other higher education institution.

(*The dissertation will not be examined unless this statement has been submitted.*)



19 December 2019

SIGNATURE DATE

(Mr R NGARIRA)

ABSTRACT

Software project management has been less effective as a result of being focused on resource management and the completion of projects within allocated resources and other confines. There has not been much focus on improving software project management quality through improved decision-making, software project management standards and methodologies, hence the focus of this study to explore the possibility of using data analytics with project management standards and methodologies to improve software project management quality.

The main question to be addressed in this study is: *Can data analytics use in software project management improve decision-making and project management quality?* This study, therefore, explores and provides insight on data analytics use, by means of a survey that was completed by software project managers. A questionnaire was used to collect data from software project managers. The gathered data was captured and analysed using the Statistical Package for the Social Sciences (SPSS), and the analysed data was used for validity testing, while the reliability of the measurement items was tested using Cronbach's Alpha. A hypothesis was used to evaluate the effect of data analytics use on software project management quality. The research made use of the positivist research method.

The study established that data analytics has not yet been widely adopted by software project managers and organisations alike, as both the project managers and organisations have not done enough to promote the training in, and the adoption of data analytics. The research also established that data analytics can improve software project management quality through improved decision-making and in complementing software project management standards. The study findings will be beneficial to software project managers, researchers and organisations as it reveals the factors that are necessary to effectively use data analytics in software project management, as well as highlighting how data analytics improves software project management quality.

Keywords

Traditional project management; Quality; Analytics; Qualitative data, Quantitative data; Software Project Management

Acknowledgements

I give glory to GOD Almighty, who grants strength, wisdom and knowledge for enabling me to complete this study.

Special mention and thanks go to the following individuals and institutes for their support:

- My supervisor, Professor Ernest Mnkandla (University of South Africa), for his consistent guidance and advice, and my co-supervisor, Professor Nehemiah Mavetera (North West University), for his guidance and support. Their steadfast support, professional and academic guidance during this study are hereby acknowledged.
- Doctor Tawanda Blessing Chiyangwa for his dedication, support, encouragement and guidance with evaluating statistical analysis.
- All software project managers who took part in this study's survey, for your valued input.
- Mr. Victor Ramoseng William Mecoamere, for editing this dissertation
- My God-given son and daughter, Anesu Ngarira and Shalom Ngarira, for their encouragement and support.
- My mother, Mrs. Rita Siwunya and family, for their encouragement and love.
- My brother, Robert David Ngarira, for his support.
- My friends for their continued support.
- The University of South Africa, for giving me the opportunity to undertake the Master of Science in Computing Research.

Table of Contents

List of Figures

List of Tables

List Acronyms

PMBOK	Project Management Body of Knowledge
SWEBOK	Software Engineering Body of Knowledge
PRINCE 2	Projects in Controlled Environments
PMI	Project Management Institute
ETL	Extract, Transform and Load
CMMI	Capability Maturity Model Integration
ISO/IEC	International Organization for Standardization / International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
USA	United States of America
JCSE	Johannesburg Centre for Software Engineering
KPA	Key Process Areas
EFA	Exploratory Factor Analysis
SPSS	Statistical Package for the Social Sciences
IITPSA	Institute of Information Technology Professionals South Africa
CFA	Confirmatory Factor Analysis
ANOVA	Analysis of variance
PCA	Principal Component Analysis

Chapter 1: Introduction

1.0 Background

Project management has not only become many organisations' means for achieving set organisational goals, but is now also a way of managing change in business operations, and also being used to achieve competitive advantage (Andersen & Jessen 2003:457). Software Project Management is a sub-discipline of project management, dedicated to the planning, implementation, monitoring and control of software and web projects (Joyce 2006).

The functionality of software project management has however been largely limited to that of control and monitoring, meant among other things to facilitate the completion of projects within specified confines and resources (Stolovitsky 2014:1; Špundak 2014:941), which has limited it to being an aiding tool in tactical decision-making. Despite using this approach, many projects have experienced huge delays, and many more have failed, notwithstanding the use of enhancing tools in software project management (Buse & Zimmermann 2012:987). Such a phenomenon is what many, including Špundak (2014:942), concur as being the traditional project management's approach that is responsible for prescribing the same methods and techniques across all projects. This is in addition to projects having been treated in isolation of their environment, which has made software project management to appear highly inadequate and criticised amid growing eagerness to try the technological innovations which have positively impacted other industries, for example fraud detection in law enforcement as cited by Chaudhuri (2012:2). This has necessitated the enhancement of software project management approaches.

Due to the high levels of many projects' failure to meet the set goals and the high failure rate averaging 33% as indicated by Buse and Zimmermann (2010:79), software project management is now regarded by many as a complicated practice (Stamelos 2010:52). This high failure rate, as has been acknowledged by Maqsood and Javed (2007:471), is essentially due to effective project management not being practiced despite being a crucial aspect which influences the success or failure of a project. Effective software project management as indicated by Maqsood and Javed (2007:471) is realised through the application of theoretical and practical aspects of managing software projects which should be taught to software project managers rather than the current varying practices being advised by literature that maybe available. Therefore, blending the practical lessons learnt from project management with project management theory will equip project managers for effective software project management.

One of the important elements that is critical to software project management quality is the set of tools being used to achieve the desired goals or objectives. Some of the tools used including PROM and Hackystat, though they are capable of collecting data, monitoring and reporting on a number of software project statistics, they have not been widely adopted and used as they focus primarily on data collection and lack the ability to analyse large volumes of data from varied sources and presentation (Buse & Zimmermann 2010:79). In addition, Traditional database schemas for example are rigid and have made data analysis difficult as opposed to newer technologies like Hadoop MapReduce which offer redesigned analytics and makes it possible to use huge data sets from various sources and combine such data which was not possible with the rigid traditional database schemas (Ames & Sokol 2012:2).

The analytic systems, among other tools that are being used in traditional project management have failed to leverage on the growing unstructured volumes of data to extract actionable information for quick and sound decision-making which would enhance project management (Ames & Sokol 2012:1; Buse & Zimmermann 2010:77). An example of traditional analytic tools include basic Business Intelligence (BI) which as explained by Rouse (2017:1), focus on historical data as opposed to advanced analytics that focus on future forecasting and predictions based on data analyses. Experience has shown that traditional analytic systems and tools become challenged as data grows bigger quickly and becomes increasingly unstructured (Ames & Sokol 2012:1), as they cannot handle the huge data volumes from different sources to provide accurate and timely analysis in near real time. Project management effectiveness has therefore, been significantly reduced since such tools offer little support, even for simple tasks (Gopalkrishnan & Steier 2012:10), thereby making projects to remain risky to conduct, and with many failing or being delayed despite the abundance of data.

Due to some inferior analytic tools being used, projects have been treated in isolation, partly to handle the huge data volumes involved and to avoid complexities that would result from processing such data volumes (Gopalkrishnan & Steier 2012:10). The lack of data integration as suggested by Ames and Sokol (2012:6), is responsible for many organisations' reluctance to invest in recent analytic technology innovations which would improve decision-making and project management quality. In the absence of effective tools and methods, past experience and intuition has been sorely relied upon for critical decision-making, resulting in sub-optimal and less effective decision-making, leading to poor project quality or complete projects failure (Buse & Zimmermann 2012:987).

These traditional project management practices' shortcomings underpin the need to improve software project management, so that it becomes a strategic asset which provides insight for improved decision-making. This has resulted in a concerted advocacy to enhance software project management quality, which this study proposes can be realised through enhanced user requirements analysis, decision-making and improved project management standards, taking advantage of the new technological advances in analytics, as has also been suggested by Ames and Sokol (2012:1). Project Management Body of Knowledge (PMBOK) is one among sundry standards that are widely adopted and used in project management (PMI 2008:443) in a bid to advance project management and its increased acceptance suggests that, among other standards, if complemented it can significantly influence projects' success.

It is unequivocal, as has been argued by Hans and Mnkandla (2013:1), that project managers require tools which enable them to make effective project management decisions. This observation is further affirmed by Buse and Zimmermann (2010:79), who indicate that there is a substantial divide between the required information by software project managers for good decision-making, as opposed to what is being supplied by current analytic tools.

This study, therefore, looks at the possibility of incorporating data analytics into the software project management process to improve software project management quality. It seeks to establish if data analytics can assist software project managers to realise this by improving decision-making and complementing project management standards in software project management. To realise quality software project management, the study considers how data analytics would facilitate decision-making which improves the carrying out of project activities and, ultimately, the chances of consistent successful projects completion.

1.1 Definition of Key terms

Project: A temporary attempt to produce a unique service or product. (PMI 2008:5).

Project Management: The use of skills, tools, knowledge and techniques on project activities to be able to satisfy stakeholders' expectations from a project (PMI 2008:6).

Software Project Management: A sub-discipline of project management dedicated to the planning, implementation, monitoring and control of software and web projects (Joyce 2006)

Traditional project management: A projects management approach which treats projects as being fairly simple, linear and predictable, with defined boundaries, making it easy to have a detailed plan and to follow the project plan without much change (Špundak 2014:941).

Quality: Refers to the extent to which a process or system satisfies user expectations and the stated requirements (Al-Kilidar et al. 2005:126).

Analytics: The practice of using data to manage information and performance (Deloitte 2013:3).

Quantitative data: Refers to the evidence or data that is founded on numbers (Oates 2006:245).

Decision-making: A process of choosing a favourable option or plan of action from available alternatives, using a defined criteria or approaches (Wang, Liu & Ruhe 2004:124).

1.2 Research Purpose

This research mainly focuses on establishing the prospect of using data analytics to improve software project management quality through enhanced decision-making and project management standards. It has been observed in previous research including as noted by Chaudhuri (2012:2) the fact that use of data analytics has been successfully used and had even led to increased productivity and efficiency in other sectors, including fraud detection and law enforcement. However, its introduction into software project management is a concept that this study anticipates will improve software project management quality.

This research, therefore, focuses on determining the possibility of using data analytics in conjunction with project management standards, such as PMBOK, to improve software project management quality. This will be realised through questionnaires, which will be completed by software project managers in South Africa and reviewing randomly selected online documents.

1.3 Research Problem

Non-integration of data analytics into software project management, a major contributor to poor quality software project management.

Software projects continue being risky and difficult to envisage despite the abundance of data, widespread adoption and use of project management standards and methodologies by organisations (Buse & Zimmermann 2012:987). Organisations are still experiencing budget overruns and projects failure despite engaging different tools, project management standards and methodologies. It would appear that the issue lies with ineffective project management practices which Maqsood and Javed (2007:471); Barber and Warn (2005:1033) concur are now required to evolve and become proactive, as project managers improve their technology management skills. For project management to become proactive, it is critical that vital information should be extracted from data to improve software project management decision-making and realise optimum results (Lavalle, Lesser, Shockley, Hopkins & Kruschwitz 2015:2).

Data analytics use is anticipated to improve project management quality through improved data analysis and effective decision-making, as Barber and Warn (2005:1033) opine that software project management has continuously benefited from other methods and innovative technical advances, which have also sought to control potentially chaotic events. The role of technological advances, including this study's proposal of using data analytics to improve software project management quality, is underscored by Hans and Mnkandla (2013:2) as being what software project managers require, among other project management tools to enable them to improve decision-making processes and project management capabilities. This study reveals how data analytics use in software projects' decision-making and project management standards can help improve software project management quality.

1.4 Significance of the Study

This research aligns with the observation by Graham and Englund (2004:10) that project management is increasingly becoming an important part of management organisations world-wide and its significance and how the frequency of use continues to increase. In light of this, the study focuses on using data analytics in software projects' decision-making and project management standards, thereby enhancing software project management quality. Maqsood and Javed (2007:471) acknowledge that projects have continued to fail due to the traditional project management methods being limited in the massaging of data to get actionable information to make timely decisions in software projects.

Apart from the limitation in data analysis by the traditional project management methods, project managers to an extent, compound the problem. Maqsood and Javed (2007:472) highlight that this is due to their limited skills in the use of technology, hence in certain instances their reluctance to embrace new innovations which would advance software project management. This study is, therefore, significant as it brings value to software project managers and researchers alike by revealing how data analytics can be taken advantage of in decision-making and in complementing project management standards in the endeavour to enhance software project management quality.

Much of the consulted literature indicates that this relatively new field of data analytics has been successfully used in other industry sectors and this study contributes to this knowledge by exploring how to leverage on this fairly new technology to improve software project management quality.

1.5 Research Questions

The main research question for this dissertation is:

How can data analytics be used in software project management to improve software project management quality?

The sub questions that will be addressed are;

i. What challenges and obstacles necessitate using data analytics to improve software project management quality? This question is meant to establish the problems currently being experienced by software project managers, which necessitates using data analytics in software project management.

ii. What is the level of preparedness of software project managers to adopt and use data analytics for the improvement of software project management practices? The purpose of this question is to determine the willingness of software project managers to use data analytics to improve software project management and the project managers' level of knowledge and experience in data analytics.

iii. What are the necessary factors required to effectively use data analytics in software project management? This question seeks to establish the necessary factors for data analytics use and the preparedness of companies to incorporate data analytics use in their project management operational processes.

iv. What are the dimensions of software project management quality, and does data analytics use in software project management improve the quality dimensions? This question establishes whether data analytics use leads to an improvement in software project management quality dimensions.

1.6 Research Objectives

This research was conducted in the South African software project management context. The cited increasing rate of projects' failure to meet budgets, timelines, set objectives and, worse still, their cancellation despite the adoption and use of project management methodologies and standards, necessitates a relook at software project management for improvement.

This study, more precisely has the following objectives:

i. To investigate the challenges necessitating data analytics, use to improve software project management quality. This objective was realised by means of a questionnaire.

ii. To establish the current level of data analytics knowledge among software project managers. This objective was realised through a questionnaire.

iii. To ascertain the basic factors that are required to use data analytics to improve software project management quality. This objective was realised by means of a questionnaire.

iv. To examine whether data analytics use in software project management improves software project management quality dimensions. This objective was realised through literature review and a questionnaire.

1.7 Research Methodology

This research is guided by the objectivism epistemology with its underlying premise being that things are in existence as meaningful objects from which research can get unbiased facts and meaning (Crotty 1998:5). Objectivism guides this research since it underscores the significance of researching the nature of relationships among elements in their constituents (Bahari 2010: 25) and, in the context of this study, assists in establishing the relationship between data analytics use and software project management quality.

The study takes a positivist theoretical perspective which Oates (2006:286) indicates as calling for the neutrality of the researcher while remaining objective in discovering facts. Positivism is used in this study in line with the argument by Bahari (2010:23) that it enables the researcher to search for patterns and causal relationships amongst elements.

In addition, positivism helps to identify causal explanations and the important laws that explain uniformities (Smith, Thorpe & Lowe 2002:1) which assist this study to establish the effect of data analytics use on project management quality. Furthermore, positivism allows this study to use a hypothesis to empirically test the theory that data analytics use improves software project management quality.

The survey research strategy employing questionnaires was used in this study. Questionnaires were used in this study as they are relatively economical and easy to administer (Oates 2006:229; Auriat & Siniscalco 2005:3). Furthermore, a questionnaire survey enabled this study to get data from many people in a standardised and systematic way, as indicated by Oates (2006:220). In addition the survey methodology assists this study in confirming an association of elements, an aspect explained by MacDonald and Headlam (2009:35) which enabled this study to establish the depth of understanding on data analytics use in software project management.

This study employs the survey strategy which Oates (2006:103) states is commonly used in computing for evaluation of software systems and in the investigation of project managers' practices and views about an information system aspect. The questionnaire survey strategy has been used as Kelley *et al.* (2003:262) indicates that this enables the accessibility of many people across the country as questionnaires can also be emailed to and from the respondents. Furthermore, using a survey questionnaire offers a broad and all-encompassing coverage of people which is likely more representative of the broader population as agued by Oates (2006:104) and Kelley *et al.* (2003:261). Purposive sampling is used in conjunction with Snowball Sampling, as the study targeted software project managers; some with data analytics knowledge and experience and focused on potential respondents from the initial target respondents.

1.8 Assumptions, limitations and delineations

This study uses the survey approach to get the important information from project managers through a questionnaire and documents review. Some assumptions, limitations and delineations are being made, and these are summarised as follows:

1.8.1 Assumptions

The following assumptions are made:

- Questionnaires are properly structured to collect relevant data to effectively establish project managers' views across South Africa.
- The questionnaire survey offers a decent representation of the population under consideration.

- The respondents do not misinterpret the questions in the questionnaire and that the responses that are given are sincere.
- Enough respondents with data analytics knowledge or experience are interviewed, providing enough relevant data to arrive at a representative conclusion of the population under study.
- Questions are appropriately structured to gather the required data for an analysis of the effects of data analytics on project management quality.

1.8.2 Limitations

This study has the following limitation:

- Questionnaires were limited to South Africa but are not evenly distributed across the country for a true representative survey population.
- The study was conducted in 36 months, with no financial sponsors, which restricted the study's survey to South Africa.

1.8.3 Delineations

- The study does not gather data from any other population except project management professionals in South Africa.
- The research does not look at the impact of data analytics on decision-making in software projects across the whole world.

1.9 Outline of the research report

This study report outline is given as follows:

- Chapter 1 gives an introduction and an overview of the study.
- Chapter 2 contains the study's theoretical foundation, based on the literature analysis.
- Chapter 3 details the research methodology.
- Chapter 4 presents data collection and analysis of results.
- Chapter 5 contains the results obtained and the findings made.
- Chapter 6 gives the conclusions and recommendations of the research based on the obtained results.

1.10 Chapter Conclusion

This chapter provided background to the study, research questions, objectives and the significance of the study. The chapter also provided an outline of the research. The following chapter presents literature review which will introduce the theoretical background for this research.

Chapter 2: Literature Review

2.0 Introduction

There is a noticeable rise in project management use in software projects and data analytics use in general (Fauser, Schmidthuysen & Scheffold 2016:67), but there has not been a positive association established between data analytics use in software projects and projects management quality.

The use of various methods, techniques and standards spanning across all aspects of project management show a progression in project management and has led to software project management now being globally accepted as a basis for professional expertise practice (White & Fortune 2002:1). However, despite the acknowledgement of its significance, Nayebi, Ruhe, Mota and Mufti (2015:18), in agreement with Ika (2009: 309), highlight that software project management is still lacking in maturity and that it is essential to regulate this crucial innovation among the range of fast evolving technologies. This suggests that the current software project management quality levels beg for improvement.

The widespread use of different project management standards including the Software Engineering Body of Knowledge (SWEBOK), Project Management Body of Knowledge (PMBOK) and methodologies such as Projects in Controlled Environments (PRINCE 2) in traditional project management practices is indicative of a concerted effort to improve software project management quality. Despite these progressive innovations, the way in which data is analysed in software project management brings the effectiveness of the traditional data analysis tools into question amid continued projects failure, with authors including Guillaume-Joseph and Wasek (2015:26) acknowledging that software projects failure are pervasive and persistent in spite of research being continually conducted.

The effect of the inadequacy of data analysis tools, as expressed by Buse and Zimmermann (2012:987), is that ineffective data analysis is hampering good decision-making in project management and remains responsible for software projects' inconsistency in attaining successful software projects delivery. Therefore, it is highly anticipated that improving projects data analysis in project management will improve data quality, decision-making and project management standards which enhances software projects management quality.

To effectively investigate the relationship between data analytics use in software project management and software project management quality, the meaning of these constructs is explained.

2.1 Project Management Quality

The element of quality is very critical in software projects since it is the basic integrative component, which exerts influence on the quality project management concept (Bobera & Trninić 2006:45). It is commendable that the significance of software project management is highlighted in the works of several researchers, including Sangeeta and Sharma (2016:3589) as well as Nayebi *et al*. (2015:18), making it essential that this study should contextualise the aspect of software project management quality.

There is a common understanding among authors on the concept of quality, with Al-Kilidar, Cox and Kitchenham (2005:126) expressing quality as; "the degree to which a system, component, or process meets specified requirements and customer or user needs or expectations". Likewise, Jamsutkar, Patil and Chawan (2012:686) define quality as a product or services' totality of features with the capacity of satisfying the specified or inferred needs. Similarly, Newton describes quality in the project management context as the process of establishing an expected quality level when a project starts, then maintaining that quality standard throughout the project (Newton 2015:4). Therefore, software project management quality in this study refers to the application of knowledge, skills and tools in entirety during planned project activities to realise set objectives and to consistently deliver successful software projects.

2.1.1 Assessing project management quality

The criteria to measure software projects management quality has not yet been universally agreed upon, and set as a defined measurement standard guidance, as indicated by Liberatore and Pollack-Johnson (2013:518) and Ika (2009:322). Issac, Rajendran and Anantharaman (2004:334) acknowledge that; "the quality of software is very difficult to define, and its measurement is quite cumbersome". In light of this, Agarwal and Rathod (2006:359) concur with Grobler and Steyn (2006:151) that software projects' quality is linked and measured by the success of the software produced. Therefore, to measure software project management quality, it is vital to assess the success of the produced software and quality-contributing standards.

However, measuring software projects' success has been a highly contentious topic, with the measuring criteria being traditionally limited to the time, cost and quality factors (Ika 2009:7; Atkinson 1999:339). Nonetheless, a more inclusive measure of software projects' success with which this study aligns, indicates a project's level of quality as measurable by business success, impact on customers and preparing for the future (Atkinson 1999:340; Carvalho, Patah & Bido 2015:1510).

The Impact on Customer factor is also singled out by several other sources, including Liberatore and Pollack-Johnson (2013:519), Ika (2009:316)**,** as well as Agarwal and Rathod (2006:359) as being a key factor that is used to measure projects quality. Customer satisfaction by the system that meets user requirements and delivers what it has been designed for essentially reflects on project management quality. To effectively outline the quality concept in software project management, related quality issues are also taken into consideration.

2.1.2 Project Quality Management Issues

The quality of software project management essentially hinges on the quality of data involved and the processes followed. The low levels of quality in software project management has been influenced by several factors, including lack of business requirements documentation or lack of maintenance upon their change (Piprani & Ernst 2008:3). Piprani and Ernst in addition, state that these undocumented software definitions, coupled with the inadequacy in the auditing or monitoring of changes to the architecture and gathered data are among the common traits observed in failing software projects (Piprani & Ernst 2008:3). These cited observations by Piprani and Ernst (2008:3) suggest the need to step up quality management during user requirements gathering, related documentation and the project management processes.

Quality assessment is one way of identifying the extent of data quality, which Veregin (1999:178) indicates as being achieved by determining the levels of reliability, completeness, accuracy, consistency, validity and the uniqueness of data. The Project

Management Institute (PMI 2008:293), in agreement with Piprani and Ernst (2008:4), also indicate that quality assessment is another way which can be conducted as either detection tests that assess quality, and is used to determine risk mitigation efforts or as faulty data that is then used in a system to help identify system deficiencies and contribute towards improved quality. It is therefore imperative to concentrate on ensuring the quality of data at the source equally as during analysis, which is one of this study's suggested ways of eliminating the flaws that are contributory to ineffective decision-making and projects' failure. Software project management data sources have been looked at to establish how data gathering and its subsequent processing can be improved by data analytics use.

2. 2 Data Analytics use

"Data analytics is a science of exploring raw data and elicitation of the useful information and hidden pattern" (Dwivedi, Kasliwal & Soni 2016:1). There are essentially three forms of data analytics, namely descriptive, predictive and prescriptive analytics. This study, by advocating for the improvement of data analysis, decision-making and project management standards, proposes the use of predictive analytics, which Sanjay and Alamma (2016:3) state as predicting future behaviours based on past events using statistical models.

The use of data analytics is expected to improve data analysis and decision-making, which in turn improves software project management. Software development as a tools-driven process, involves some automation, with many potential breakdown points which lead to poor quality software (Kaulgud & Sharma 2016:10). Since analytics increases the focus on execution and promises low tolerance on projects' failure (Fauser, Schmidthuysen & Scheffold 2016:67), its introduction in software project management improves processing and projects' success rate. It counters a major limitation with traditional data analysis which Chawda and Thakur (2013:2) indicate moves data from one system, transforms it into relational data form and is passed to another system for processing, whereas with data analytics the data is stored in one place and processed with much accuracy and speed.

This limitation, as echoed by Rao, Gudivana and Raghavan (2015:2654) epitomises the challenge of the traditional Extract, Transform and Load (ETL) data quality management tools, which have contributed to the persistent experience of missing data, realising inconsistent or incorrect data, and the non-identification of associated data and failure to link the data for analysis. The application of predictive analytics in project management data analysis and processes promises to counter these traditional data analysis tools' limitations. This includes enabling consistent and faster decision-making based on data analysis, allowing detecting and forecasting trends in data which enable anticipation for change and allows resources allocation where they are needed most (Buse & Zimmermann 2012:989; Singh 2015:2).

2. 2.1 Predictive Analytics

Predictive analytics will potentially improve decision-making and judgements of possible future events' outcomes in software projects. This is due to predictive analytics' ability to establish probable future outcomes of projects events or chances of situations occurring by its automatic analysis of multi-variable high volume data (Mishra & Silakari 2012:4434). The quick analysis of high volume data enables data-based decision-making, which, according to Mishra and Silakari (2012:4435), indicates what predictive analytics offers by its suggestion of the actions to be taken, along with the timing and production of insights, for strategic decision-making.

The ability of predictive analytics to leverage projects data and predict possible future events' outcomes with a greater degree of accuracy, which complements project management as indicated by Fauser *et al.* (2016:67), makes it ideal for use in this study. As data processing and forecasting improves, it enhances the maturity of projects' processes. In addition Fauser *et al*. (2016:67) highlights that one of the main benefits of predictive analytics over other technologies is that it is forward-looking, rather than reporting or being analytical of past events. This introduces more projects' insight and allows adequate reactions to upcoming challenges.

Predictive analytics' reliance on considering the associations between the predicted and explanatory variables from previous experiences, as cited by Mishra and Silakari (2012:4435), makes it instrumental in forecasting the future outcomes of projects' decisions. Predictive analytics, being data-driven is preferred in this study's approach to project management quality improvement, as Fauser *et al*. (2016:68) note that it has not only been successfully implemented in the financial and retail sectors, but more so for its provision of empirical and fact-based justification. The sources above suggest that decision-making and the forecasting of future outcomes in project management may improve from data analytics use.

Predictive analytics use therefore, will enhance project management quality, as it is forward-looking, it can improve decision-making, making the project management process proactive to any coming challenges.

2.3 Decision-making contribution to quality

Project management, as highlighted by White and Fortune (2002:1), has significantly advanced with the entire project management processes now being covered. However, a major cause for concern as contended by White and Fortune (2002:1) is that over 50% of software projects undertaken globally continue to fail. Many authors as indicated in the sources that follow, concur that these failures are partly attributed to poor decision-making. Therefore, decision-making is key for software project management's success, and its significance is highlighted by Nayebi *et al*. (2015) who acknowledge that software project management is a process which is decision-intensive and that the failure or success of projects is extremely reliant on the decisions made.

An analysis by Vasconcellos, Silva, Cunha and Moura (2016:26) reveals that 47% of projects fail as a direct result of poor decision-making. This high failure rate is being experienced despite the availability of the tools and techniques like decision models and traditional decision analysis that are meant to assist in decision-making. It, therefore, can be contended that effective decision-making is not solely dependent on the aiding tools that are being used, but is also reliant on the quality of the data that is being used. In addition, Vasconcellos *et al*. (2016:27) are of the view that what makes decision-making even more difficult is that there is no systematic approach of how software project managers make decisions in various contexts. This indicates that good project management decision-making is neither dependent on one factor, nor can it be prescribed, making it a complex process that relies on data quality as a fundamental element for its improvement.

The need to improve decision-making as a quality improving factor in project management is supported by the identified weaknesses in the current practice. One of the weaknesses with traditional software project management, as stated by Mcavoy and Butler (2009:372), is that decision-making is less effective since the responsibility solely lies with the project manager, who operates from a position of command and control. This management approach is inhibiting, as Singh (2015:4) acknowledge that the majority of project managers due to their lack of training in embracing the knowledge and use of analytical tools, technologies and processes are reluctant to move away from the subjective legacy approach of decision-making in project management. In situations where some analytical tools are being used, Bose (2009:170) contends that the existing software packages lack enough support for directing analytic processes and are limited to a somewhat narrow set of techniques. In light of this limitation, Delen and Demirkan (2013:360) highlight that, to be effective, decision-making requires tools and systems that are proficient in providing the correct and appropriate information to consistently make the right decisions.

The decision-making improvement process is hinged upon projects data quality. The significance of quality data in decision-making is emphasised by Haug, Arlbjorn, Zachariassen and Schlichter (2013:238), who state that 88% of projects involving data integration fail or go over budget, with less than 50% of the companies regarding themselves to be confident in their own data quality. At the heart of this dilemma appears to be the poor quality of data to support good decision-making. This sentiment is supported by Singh and Singh (2010:41), who argue that the abundance of data alone or the availability of methodologies, although very critical, are in themselves insufficient to warrant good decision-making and to provide consistent successful software projects delivery.

Likewise, White and Fortune (2002:7) state that despite the majority of project managers using a combination of tools, methods and methodologies in projects, a considerably high percentage of 52% of project managers do not use any decision-making techniques. Against this backdrop, Haug *et al*. (2013:237) established that poor quality data has largely led to inefficient decision-making processes, causing lack of confidence on the data and resistance to initiatives founded on such data.

This observation highlights the necessity of improving software projects’ data quality and decision-making processes, of which data analytics is one initiative that is being explored by this study to induce such a difference. In agreement, Stamelos (2010:57) points out that traditional software project management practices have gaps and appear insufficient as more than half of the Information Technology (IT) projects continue to fail, indicating the need for an improvement. Given these observations of inferior data quality and ineffective decision-making practices amounting to poor project management quality, Sanjay and Alamma (2016:3) suggest that data analytics' aggregation and analyses of information in real-time adequately assists in decision-making. This supports the notion that data analytics holds the promise to improve project management decision-making and to enhance project management quality. The sources which feed the decision-making process and the rest of the project management processes with data are therefore considered.

2.4 Software Projects Data sources

User requirements data gathering being a very important source of data in software development projects establishes what the users require of the software to deliver (PMI 2008:105). The process of data gathering, therefore, is of interest if the quality of data is to be enhanced. Traditional methods, including interviews, observations and studying documentation, as highlighted by the PMI (2008:107–109), are arguably the most common means through which data is gathered. The process of data gathering and the traditional way of documenting user requirements which provide a requirements checklist, constituting an agreement between developers and the project sponsor, requires improvement to ensure a high degree of data accuracy and consistency.

There are essentially three data sources, namely primary, secondary or tertiary sources, and their characteristics influence data quality. Primary data is acquired from a source and is used without the processing it goes through, thus changing its elements (Bruce 2014:20). Primary sources of data are the least potential error sources since their data is acquired first hand. Shankaranarayanan and Cai (2006:308) however contend that such data may, nonetheless, undergo formatting or inspection, even though not being changed but resulting in component data.

Secondary data source, as explained by Bruce (2014:20), provides data which originates from a source and gets transformed into another format before it is reported on. The process of compiling user requirements data into a single user requirements list forms the secondary source, which in turn is used for coding. This compilation process exposes data to potential errors during transfer.

Tertiary data is the third category, which is made up of secondary sources' processed information that is transformed into appropriate and easily readable forms (Bruce 2014:20). The processing and transformation of data creates an opportunity for data spoiling, hence the need for data monitoring and analysis, as data quality threats exist at source and during transformation.

2.4.1 User Requirements Specification data quality

In the software development cycle, the requirements gathering phase is highly critical as the quality of data gathered influences the chances of success in software projects. However, an unfavourable situation exists, as has been pointed out by Young (2002:11), reflecting that an estimated 85% of defects in developed software have their roots in the requirements' specification. Alshazly, Elfatatry and Abougabal (2014:514) along with Anuar and Ahmad (2015:102), concur that the quality of software requirements specification is plagued by inaccurate, incomplete, inconsistent, ambiguous and duplicate data among other defects as a result of human, processes and documentation errors. In addition, Guillaume-Joseph and Wasek (2015:40) state that changing user requirements, poor requirements management and vague requirements are among the leading causes of software projects failures.

The inadequacy of technology in traditional project management data analysis is also identified as a cause for projects failure from poor quality data processing. Walia and Carver (2009:1094) state that some of the user requirements' quality issues are due to ineffective technology use during analysis and inadequate skills. Some of the technologies currently in use include the famous spreadsheets and relational databases requirements repositories. These are, as argued by Anuar and Ahmad (2015:102), insufficient as they make it difficult to sort, query or maintain the requirements due to their incapability to manage critical detail, including the source, priority, status and type of the requirements' data. Furthermore, the traditional spreadsheets and paper requirements lists have a limitation of abstracting all the requirements, thereby offering little context and making it difficult to identify the most important requirements (Dorfman 2010:3). In addition, Dorfman (2010:3) indicate that due to the resulting generalisation, different people get different versions of the system when they read such requirements. The existence of such a challenge during user requirements data gathering puts the eventual quality of any project's end-product in an uncertain position.

In light of the importance of ensuring quality in data at the source and during processing, Young (2002:11) suggests that improved validation and verification of the requirements data will ensure traceability, consistency and eliminate any ambiguities. In agreement, Anuar and Ahmad (2015:102), assert that insufficient validation and verification, coupled with poor tools, constitute a major contributor to poor quality requirements and data management. The process of validation requires accurate data analysis, which calls for the use of newer technologies, including data analytics. Using data analytics will therefore assist in quick data validation and processing, eliminate redundancy, and improve data consistency, even during requirements gathering. Such agility may be made possible through the data analytics' features, which in summary, as indicated by Herodotou *et al*. (2011:1), offer appeal, swiftness and depth to the users.

2.5 The significance of Software projects data quality

Any data set's level of quality attributes determines the degree of quality passed onto the final software product that would have been produced. To deliver successful software projects consistently, it is imperative that the gathered user requirements data be of high quality, to facilitate relevant and timely decision-making. This is acknowledged by various authors, including Rodríguez and Riveill (2010:1), who concur that highly relevant data provision presents a greater possibility of correct information processing, interpretation and appropriate interventions. In addition, Lucas (2010:3) further qualifies that data is regarded to be of high quality when it is suitable for its intended purpose in planning, operations and decision-making. Similarly, Bruce (2014:3) explains the concept of data quality as the value or correctness of the collected information to meet requirements, resulting from a high standard of data capturing, verifying and analysis.

To enhance data quality, emphasis is placed on the role and significance of validation coupled with verification during data capturing and processing. In support of this, AHIMA (2009:1) state that a commonly agreed-upon practice to enhance data quality is by continuously subjecting data to optimal quality management, which is a continuous improvement process. Likewise, the North Carolina Department of Public Instruction (2010:1) by stating that the data quality management process, which includes the use of procedures and policies regarding the collection, maintenance and reporting of data contribute towards data quality demonstrates the significance of validation and verification towards attaining the highest form of data quality.

The quality of data in software projects is inarguably critical for successful software delivery and in sustaining quality in project management. However, in spite of the awareness of the importance of data quality in software projects, Haug *et al*. (2013:235) established that, in practice, poor quality data is among the main challenges that companies continue to face to this day. It is this persistent challenge of poor data quality that has rendered the objective to deliver quality software projects to appear to be far-fetched. Therefore, it is imperative that what constitutes data quality in this study be put into perspective.

Data quality refers to the degree to which data characteristics fulfil set requirements (Laranjeiro, Soydemir & Bernardino 2015:179). Different authors on different scales agree on categorising data quality dimensions as a means to manage data quality. Haug *et al*. (2013:236) postulate that these data quality dimensions are summarised into at least 26 categories. Likewise, Bruce (2014:5) and Yeo *et al*. (1999:2) affirm that data quality is commonly assessed on a broader scale, in terms of timeliness, validity, reliability, confidentiality, precision, completeness, integrity and ethics. On the contrary, Shankaranarayanan and Cai (2006:303) as well as Laranjeiro *et al.* (2015:179) nonetheless choose to narrow down the main and commonly-addressed qualities of data to accuracy, completeness, timeliness and consistency.

Further to the attributes earlier alluded to, data is generally considered to be of high quality, hence being relevant for effective decision-making and being contributory to project management quality if it exhibits some of the following common attributes:

i. Completeness - Singh and Singh (2010:41) state that it is the extent to which an entity has all the expected attribute values. Shankaranarayanan and Cai (2006:304) reiterate that completeness is a significant data quality element, as it measures the extent to which there are no missing values. Equally important are the end users who have to be identified to warrant complete data collection for the application and to clarify how the data will be used (AHIMA 2009:1).

ii. Consistency is another important attribute which, according to Singh and Singh (2010:41), denotes the extent of an information object's consistent occurrence in the same compatible format as the other similar information objects. AHIMA (2009:1) argues that it calls for data value to be identical across applications and systems for it to be reliable.

iii. Accuracy - The level that data qualities appropriately represent the correct value of the planned object, referred to as accuracy, is also a commonly identified as a significant attribute. Butt *et al.* (2013:4569) state that accuracy denotes the preciseness of software in giving the correct results. In addition, Bruce (2014:11) states that for data to be precise, only relevant information should be collected and in situations where the data is sample-based, an error margin should be tolerable for the obtained data.

iv. Accessibility is an attribute of quality data, which refers to how easily available data is in a certain context of use. Maydanchik (2007:212) states that accessibility is the level of the easy obtainability of data and how quickly reachable it is. It is essential to take note, as indicated by AHIMA (2009:1), that despite data being easily accessible, it should nevertheless be legally obtainable, and within the reasonable legal and financial confines, and applications.

v. Currency of data shows the extent to which data possesses the right age attributes and timeliness of the results from the frequency of collection, which has to be recent (Bruce 2014:11). Peralta (2008:7) postulates that some surveys and studies conducted have shown that data currency is directly related to the success of information systems.

vi. Validity refers to how data adequately represents performance and is measured in terms of face validity, which denotes the sound relationship between an activity and what is to be measured (Bruce 2014:3).

vii. Validity takes three different forms, which according to Bruce (2014:3) are measurement validity, where data measurement tools and procedures are used to limit the potential for errors, such as sampling and non-sampling errors. Transcription validity, on the other hand, depends on sound data entry and, collation procedures, where steps are taken to ensure that data was transcribed, entered and tallied correctly (Bruce 2014:9).

viii. Data reliability, as expressed by Veregin (1999:177), is produced by stable and consistent processes and instruments that are used over time to collect data and use procedures to deal with any missing data. Internal quality controls are essential to guide periodic data collection, maintenance and assessment and, to realise reliability, there has to be transparency which is achievable by means of data gathering, cleaning, inspection, reporting and quality valuation processes (Veregin 1999:182).

ix. The Integrity of data is also very important, where data must be accurate and be free from human or technologically-introduced errors (Pipino *et al*. 2002:212). There is a need for measures to be put in place to warrant that no manipulation or prejudice is introduced in the data. Bruce (2014:10) hints that there is need to guard against manipulation which may unconsciously arise from rushed data collection or entry, and wilful integrity issues, such as intentionally providing false data in a survey due to pressure or non-supervision.

x. Maydanchik (2007:212) cites the appropriate amount of data as one of the requisite qualities which expresses the appropriateness of the data volume in relation to the task to be performed.

2.6 Data Quality Issues and their impact

Data quality issues currently being experienced date back as far as the early days of computing, with issues such as missing data, inconsistent data, non-identification of linked data from multiple sources and incorrect data being among the focus areas of data quality and transformation investigations (Rao, Gudivada & Raghavan 2015:2654). These data quality issues have been compounded by contributory factors like the introduction of the internet and the data availed by the internet in the early 1960's (Rao *et al*. 2015:2654).

Notwithstanding that it is readily available, data from the internet adds to the quality woes as Rao *et al*. (2015:2654) state that it adds to the bulk of big data whose emergence in the last five years has magnified the data quality problems and also brought in several new research challenges.

One of the reasons of information systems' failure to deliver is attributable to poor data quality (Piprani & Ernst (2008:1). This poor data quality being suffered is, as indicated by Haug *et al.* (2013:237), due to insufficient data analysis techniques and ineffective decision-making among other main causes. Rodríguez *et al.* (2010:4) classify these data quality issues as relating to technology, such as the erroneous manipulation of data, human intervention, input errors and some to data-processing errors. The ineffective traditional data analysis techniques and the errors due to human intervention are among the most conspicuous shortcomings which data analytics is anticipated to curb in projects data processing.

Rao *et al.* (2015:2655) identifies one of the initial data quality issues experienced to date, as record linking that requires determining if more than one data object is linked to the same entity without a key identification value. Such situations require an intervention to conduct a search to pair the matched records, which increases the chances of data corruption. Quality issues also emanate from faulty data creation and the transformations it goes through, which necessitates embedding data quality measures into queries to deal with propagated and accumulated quality errors (Rao *et al.* 2015:2655).

2.6.1 Data Quality classifications

Data qualities can be categorised as inherent and pragmatic qualities. Laranjeiro *et al*. (2015:181) describe inherent qualities as having static quality characteristics, which include data completeness and conformance to business rules, whereas pragmatic quality refers to how clear and understandable data allows users to achieve their goals. Data qualities can be further classified into fifteen dimensions under four categories, as explained by Haug *et al.* (2013:236) are;

- **Intrinsic**: Addresses the concept of believability, objectivity, accuracy and reputation.
- **The contextual category** focuses on the timeliness, completeness, relevance and suitable amount of data.
- **Accessibility** looks at access security and how easy data can be accessed.

- **Representational** refers to how easy it is to understand, concise representation, consistency in representation and how easily interpretable data is.

Conversely, Sidi *et al*. (2012:301) classify these data quality issues as single or multi source issues having the schema level and Instance level categories. The single source schema level problem is characterized by substandard schema design and lack of referential integrity, whereas the instance level issues include data-capturing errors, redundant duplicates and inconsistent values (Sidi *et al*. 2012:301). Multi-source schema level problems, on the other hand, as expressed by Sidi *et al*. (2012:301), include diverse models of data and schema design, as well as conflicting names, while the instance level problems include contradictory, overlapping and inconsistent data.

A different categorisation by Laranjeiro *et al*. (2015:185) classifies these issues into a hierarchy of four levels, as follows;

- Multiple source data problems, which include the heterogeneity of syntaxes, measure of units and representation.
- Multiple relation level issues comprise of referential integrity violations.
- Single relations level issues consist of approximate and inconsistent duplicate tuples and attributes.
- Tuple or attribute level problems include missing values, syntax violations or misspellings, with missing values identified as being the most common problems which hinder completeness and impairs accuracy from the incorrect values.

The classifications cited above, despite their different approaches, all focus and place an emphasis on data accuracy, completeness and consistency which is pivotal to effective decision-making. The classifications, therefore, reveal the areas of potential improvement in data quality which, with the introduction of data analytics, is anticipated to improve project management quality. In the same way, focusing on the sources of data quality issues also reveals areas where data analytics will potentially improve the data quality in project management's data-gathering processes.

2.6.2 Sources of data quality issues

Issues of data quality essentially come up at any point in the data cycle, which can be at the point of the data sources, during data profiling or integration, during data staging stage, as well as at database modelling. Singh and Singh (2010:42) state that data quality issues emanate from poor data management processes, migration errors in between systems, externally-acquired data not fitting with company data standards and the inability to keep to data recording and maintenance processes. Data quality problems, as summarised by Singh and Singh (2010:43), emanate from four common sources, namely:

Data source – Due to the diverse methods of storing data, some sources offer unsecured access, which leads to unreliable and poor-quality data. Issues such as lack of validation routines, missing values, misspelled data and inconsistent data formatting, especially from legacy systems, are the identified common causes at data source.

Data profiling – Through sources assessment, data quality problems including the propagation of poor data quality due to no data evaluation before integration, unidentified relationships in data, inadequate data analysis and poor quality from manually derived information have been found to be common.

Data grooming - This is done at data staging once data is collected from source systems, basically involving the validation of data quality and tracing the data problems. The causes of data quality issues include poor system conversions, migration, reengineering and consolidation; incorrect data extraction to the relevant fields, and issues caused by different business rules on handling data from different sources.

Data modelling – Focuses on schema design issues such as the quality of information which depends on data quality, application programmes and the database schema. Some causes of the quality issues of schema design include incorrect requirements analysis resulting in substandard schema design, incorrect identification of facts or tables relationships leading to poor data quality and insufficient integrity and validation rules in the schema contributing to poor data quality. These sources of data quality problems have to be analysed for data to drive enterprise-wide decision-making. For this to be achievable, there should be support for the resources needed to maintain quality which, as has been asserted by the North Carolina Department of Public Instruction (2010:1), includes ensuring that all the data-handling staff should understand the importance of data accuracy and become trained. In addition, Buse and Zimmermann (2010:79) suggest that validation checks through analytic tools in data collection systems, profiling and audit reports also have to be built into local source systems to ensure data quality processing.

2.6.3 Summary of factors affecting software project management quality

Factors that affect data quality, decision-making and subsequently software project management quality are briefly categorised as follows:

Table 2.1: Factors affecting software project management quality

Factor Category	Factor Description	Source
Data Source and Quality Factors	User requirements are not clearly captured when collected at source or keep changing resulting in different interpretations.	Singh and Singh (2010:42- 48)
	Transferring of user requirements from different sources into a single user requirements list result in some incorrect data modification.	
	Lack of validation and verification at data source and during processing result in missing and inconsistent data.	
	Poor data quality with regards to accuracy, timeliness, completeness and consistency emanating from legacy software data.	
	Lack of training in data collection and processing procedures.	
	Lack of reviews during data entry.	
	Inconsistent data and naming conflicts from multisource data.	
	Lack of validity checks result in incorrectly entered, transcribed and tallied data.	
	Lack of integrated data storage and refreshing.	
	Lack of training in the use of analytic tools	
Data Analytic Tools Factors	Lack of effective tools to track, handle and measure variables in decision-making.	Singh (2015:4)
	Incapable and slow analytic methods to provide relevant information.	White and Fortune (2002:5)
	Lack of data analytics use to extract actionable information.	Dwivedi *et al*. (2016:1)
	Data not analysed quickly enough for quick decision-making.	Buse and Zimmermann (2010:77)
	Traditional tools depend on slow relational databases for query execution and data storage.	Sanjay and Alamma (2016:2)
	Models such as Garbage Can model are not fully incorporated in project management to be able to improve decision-making.	Ularu *et al*. (2012:7)
Research Methodologies,	More than 52% of project managers are not using any decision-making techniques resulting in projects failure.	Judgev and Thomas (2002:6)

Models and Standards Factors	Models and standards including PMBOK are not all inclusive or complete making them inadequate to guarantee project quality management.	White and Fortune (2002:7)
	PMBOK focuses on project progress monitoring and appears not to be a strategic tool for decision-making.	Stamelos (2010:57) White and Fortune (2002:1)
	SWEBOK knowledge areas are insufficient to adequately capture user requirements.	Singh (2018:4)
	Conventional data quality management approaches including data cleansing and statistical process control do not offer a methodical approach to data quality management.	Bourque *et al.* (1999:41)
	Software projects are under-staffed of project management experts hence the general lack of informed decision-making in projects.	Shankaranarayanan and Cai (2006:303)
Human Resource factors	Faulty data creation, capture and transformation, resulting in poor quality.	Eriksson and Brannemo (2011:6)
	Poor system conversions, migration and consolidation of data.	Rao *et al.* (2015:2655)
	Lack of quality assessment to determine levels of accuracy in data consistency, reliability and data validity.	Singh and Singh (2010:43)
	Business requirements documentation is either non-existent or not maintained when changed.	Haug *et al*. (2013: 240)
	Unclearly specified business rules specification.	Piprani and Ernst (2008:3)
		Piprani and Ernst (2008:3)

2.7 Data quality impact on decision-making

Poor quality data negatively impacts the quality of software projects as it operationally causes customer dissatisfaction and increases costs due to more resources being required to correct errors (Laranjeiro *et al*. 2015:179). The effect of inferior quality data on projects' decision-making, strategic planning and project execution is highlighted by a number of authors. Laranjeiro *et al*. (2015:179) argue that poor quality data tactically affects decision-making resulting in mistrust and strategically makes it difficult to define and execute an organisation's strategies. Rao *et al*. (2015:2654) correspondingly acknowledge that high data quality is critical in decision-making, as well as in strategic planning.

Likewise, Haug *et al*. (2013:237) state that inefficient decision-making processes and increased operating costs being experienced result from poor quality data. In agreement, the North Carolina Department of Public Instruction (2010:1) also emphasises that data is a critical assert to business decisions and that it must be valid in one system to be useful in another system. The mechanisms currently in use to improve data quality are therefore looked at.

2.8 Data quality improvement mechanisms

Different approaches and strategies are being used in an effort to improve data quality in software projects. Sidi *et al*. (2012:300) cite data-driven and process-driven strategies as the two commonly-adopted approaches for improving data quality, where the former improves data quality by among other means directly adjusting data values, error correction and record linkage, as well as improving data integration. On the other hand, process-driven strategy as explained by Glowalla and Sunyaev (2014:1), redesigns the produced processes or processed data to enhance its quality. The process-driven strategy eliminates the bulk of the causes of the quality problems as it makes use of the process control technique which checks and manages data, while process redesign eliminates the low quality causes and controls formatting of data before storage (Glowalla & Sunyaev 2014:2; Sidi *et al*. 2012:300).

These employed strategies are to an extent managing data quality by analysing factors that may influence its quality at any data cycle phase and they try to safeguard the data against being polluted, which improves data quality. To further enhance data quality, Bruce (2014:19-20) proposes some processes, which include:

- Building data entry checks into the data collection process.
- Improving data integrity by presenting relevant and accurate data.
- Making relevant data easily accessible to key decision-makers for data based decision-making.

The three proposals can be enhanced by making use of data analytics to improve automated data entry checks, which further enhances data accuracy and decision-making. Quality data which is required for effective decision-making is also enhanced by the interdependency of the independent dimensions which include completeness, consistency, accuracy, quality and the dependent currency variables (Sidi *et al.* 2012:303; Yeo *et al*. 1999:5).

The interdependences illustrate that the data quality dimensions identified do not work in isolation, but are rather complementary and indicate that they would greatly benefit from technological advancement, including the use of data analytics. Peralta (2008:3) concurs that in most cases these attributes are treated independently, yet most of them are intensely interrelated, which makes their separate evaluation challenging.

Data entry checks, including human spot checks or software checks, have to some extent assisted in introducing some degree of cleanliness in the collected data (Bruce 2014:21). However, to further enhance data quality, Bruce (2014:21) suggests that people should be trained in data collection methods and procedures, that data entry reviews be conducted to check for errors and data collection processes, including collection instruments and the consistent use of sampling. It is worth noting that these recommendations show a heavy reliance on the human element in the data collection, collating, and processing chain, which highlights the need to use data analytics to reduce human error, improve on data accuracy and quality during collection and processing. To effectively manage data quality, the classification of data quality issues for improved assessment and the use of quality models need to be complemented with the use of data analytics.

2.9 Quality Management Models and Techniques' contribution

The effort to deliver successful software projects consistently has seen quality management models and techniques being widely adopted by project managers. In this attempt to improve data quality, Bruce (2014:6) states that project managers often use quality management plans, which essentially define how to manage data standards and evaluate data quality through assessments. The introduction of data quality management models is meant to continually improve the quality of data as the models include data collection, analysis and warehousing mechanisms (AHIMA 1998:1).

In an attempt to improve quality in software projects a three-staged process has often been used, which includes planning and defining quality, undertaking quality assurance and controlling quality. Quality management planning involves defining the level of quality required before the project begins, defining how quality will be measured, performing quality assurance, explaining how the quality plan is going to be executed and controlling quality by correcting problems as soon as they are identified (Kloppenborg & Petrick 2004:5). A number of other techniques, which include benchmarking, cost of quality and cost benefit analysis are also used in the quality planning process (Kloppenborg & Petrick 2004:14).

As part of the greater effort to improve software quality, Butt *et al*. (2013:4568) indicate that organisations dealing with software projects use quality assurance models including the Capability Maturity Model Integration (CMMI) and the ISO/IEC 90003 standards to improve software quality. Similarly, Khalid (2008:304) confirms that the principles from these models introduce appropriate standards and defined procedures to be followed, which brings some level of management quality. These models introduce planned and systematic approaches, which bring some conformance of the software or processes to established principles, processes and measures (Khalid 2008:304). Since these models have introduced some level of quality, some authors have argued that the low levels of management quality are in fact as a result of improper implementation. In agreement, Khalid (2008:304) argues that, despite the abundance of the software quality models, it is the improper implementation rather than the inadequacy of these models and the high costs involved which is leading to poor quality in software projects.

2.10 Project management standards and methodologies

The project management discipline though improved, however continues to show weak results since the conventional view of project success is still being limited to time, cost and objectives fulfilment only (Carvalho, Patah & Bido 2015:1519). Project management methodologies including PRINCE 2 and the PMBOK body of knowledge still account for 22.3% of limitations, which directly impacts on the quality of projects and the probability of the projects' success (Joslin & Müller 2015:1378). This points to an underlying threat to the projects' quality management and is an indication that software project management standards and methodologies should be improved if they are to become more effective.

Some level of quality has been brought into software project management by the introduction of these standards and methodologies. These have assisted by dividing projects into partial phases for better project control; for instance, PRINCE 2 introduces control by dividing a project into five sequential phases which are concept, definition, implementation, handover and closeout (Kostalova, Tetrevova & Svedik 2015:97). Likewise, PMBOK body of knowledge also divides a project into nine knowledge areas to improve project management. Due to the improvements it has brought so far, PMBOK body of knowledge has become one of the generally-accepted and widely-used project management standards, and its increased acceptance suggests that, if enhanced, it can significantly influence a project's success (Ghosh, Forrest, DiNetta, Wolfe & Lambert 2012:2; PMI 2008:4).

2.10.1 Project management body of knowledge (PMBOK)

The PMBOK standard focuses on nine knowledge areas, which are Integration, Scope, Time, Cost, Quality, Human Resource, Communications, Risk and Procurement Management taking care of the different facets and processes in individual projects (PMI 2008:14).

A brief preview of the nine PMBOK knowledge areas:

Table 2.2: PMBOK Project Management knowledge areas (PMI 2008)

Knowledge Area	Function
Integration	Focuses on proper coordination of various project elements by means of project plan and execution control.
Scope Management	Covers the work required to successfully complete a project.
Time Management	Ensures timely project completion through scheduling and sequencing of activities.
Cost Management	Caters for resource planning and cost control to warrant completion of a project within the set budget.
Quality Management	Ensures quality is achieved and that the project meets the objectives it was set for. It involves: i. Quality planning - explains the formulated quality policy implementation. ii. Quality assurance - ensures a project meets the set quality values. iii. Quality control - responsible for project results monitoring against set quality standards and seeks to remove the sources of substandard results.
Human Resource Management	Sets out the processes to acquire and effectively use people in a project.
Communications Management	Handles the generation, dissemination and disposal of project information.

Risk Management	Identifies risks associated with a project and how to avert or manage the risks.
Procurement Management	Deals with the sourcing of goods and the handling of contracts to their logical conclusion.

PMBOK Limitations

Despite bringing some project management control, PMBOK's limitation is that, if used in its original state in traditional project management, it is not all-inclusive nor complete (Ghosh *et al.* 2012:2). PMBOK's insufficiency makes it a potential improvement area where data analytics can be used to enhance its integration, scope management and quality management.

A major limitation with PMBOK's knowledge areas is that they do not turn a project into a strategic tool which would influence decision-making as they simply monitor a project against development parameters (Stolovitsky 2011:1). Ghosh *et al*. (2012:3) concur that despite PMBOK being the most accepted and used standard, an average of 28% of projects continue to fail. Furthermore, Singh (2015:4) states that PMBOK, despite being considered a global standard for project management processes, is not sufficient in providing an analytics focused approach. Therefore, in addressing some of the concerns leading to projects failure, Ghosh *et al*. (2012:3) concur that PMBOK also needs to be improved if it is to enhance project management. It is upon this supposition that this study anticipates that using data analytics in project management will complement standards such as PMBOK and other methodologies to improve data processing and analysis in areas such as scope management and quality management.

2.10.2 Software Engineering Body of Knowledge (SWEBOK)

This guide makes use of knowledge areas to assist in quality improvement in software projects. One of the knowledge areas, the software requirements analysis knowledge area through its sub areas contribute towards quality in the following manner (Bourque, Dupuis & Abran 1999:42; IEEE Computer Society 2004:1–4):

- The requirements elicitation subarea focuses on where the requirements come from during software development, and on the collection method.

- The requirements analysis subarea evaluates user requirements and seeks to detect and resolve conflicts in the requirements.
- Requirements validation subarea checks if there are omissions, ambiguities or conflicts, and guarantees that requirements are in line with the set standards of quality. This process ensures that requirements are sufficiently described to avoid misinterpretation before the resources are devoted to addressing the requirements.
- The Requirements management sub area caters for change management and accurately maintains the requirements, to mirror the software to be developed.

Data Analytics will therefore improve the SWEBOK's requirements analysis knowledge area by improving accuracy in requirements validation, requirements analysis and track the changes applied to the captured requirements.

Besides the requirements analysis knowledge area, the software quality analysis knowledge area also endeavours to improve software quality by focusing on user requirements specification. It concentrates on software quality assurance, verification and validation activities to try and avoid requirements specifications miss-representing the customer needs; for the developed software to avoid failing to fulfil any or related requirements, and to ensure that elusive errors get detected (Bourque *et al*. 1999:41). The software quality assurance component includes actions that attempt to offer confidence that a product follows the specified technical requirement and verification, as well as whether the validation processes deliver an objective valuation of software throughout the software's life cycle (IEEE Computer Society 2004:1–6).

However, despite SWEBOK best software engineering processes being in place, the requirements specification may still potentially not meet customer requirements or the code can still fail to fulfil user specified requirements (Bourque *et al*. 1999:41). In addition, Shankaranarayanan and Cai (2006:303) state that these conventional quality management approaches, including data cleansing, data tracking and process control even though being useful to an extent these however do not give a methodical approach for the management of data quality.

2.11 Quality management models and decision-making

Decision-making is influenced by various interrelated aspects and models, such as the Garbage Can Model and the Mixed Scanning Model, which are based on the concept that rational decision-making processes lead to better results (Eriksson & Brannemo 2011:1). In spite of the complexity of most of the software projects, Marques *et al*. (2010:1058) state that modelling has played an important role in supporting decision-making. However, Eriksson and Brannemo (2011:1) on the contrary argue that there are very few projects which are a direct result of the rational decision-making process. This shows that the element of decision-making has not been comprehensively exploited using quality management models to improve software project management.

It is one of the core functions of decision-making to focus or refocus projects during implementation. In this regard, decision-making in relation to projects' deviation and redirection has received little attention, resulting in 55% of the critical projects decisions being made on an ad-hoc basis (Eriksson & Brannemo 2011:2). The ad-hoc decision-making shows lack of planned direction or control of many of the projects. To emphasise on the significance of decision-making in guiding project management, Marques *et al*. (2010:1058) state that decision-making determines the progression of projects from the current status to the set objectives. The two prior arguments point to the fact that, regardless of the stage at which a project may be, decision-making is central for the focussing or refocussing of a project, based on the initial objectives. However, contrary to what is being practiced, Goff (2011:2) observes that, in practice, decision-making is being executed based on non-existent facts or on slow and not yet discovered facts. This underlines the importance of having accurate and quick processing of data for timely decision-making which directs or redirects projects.

A good decision occurs when a situation is matched by an independent remedy in a satisfying way in line with the set goals (Eriksson & Brannemo 2011:10). To be able to make good decisions, project managers have to consider a number of variables in their activities and measure these, while evaluating the project's progress in all aspects (Marques *et al*. 2010:1068). Goff (2011:3) emphasise that excellent decision-making emanates from planning and tracing scope as a leading project indicator.

To be able to track and evaluate the variables in project management, effective tools are required and this study focuses on data analytics as one such a critical tool.

Besides the necessity to have appropriate tools for effective decision-making, the need for project management experts is also equally crucial to be able to explore and effectively use the available data analytics technology. Eriksson and Brannemo (2011:6) underscore the need for project management experts to counter poor decision-making by stating that the shortage of project management experts, which is resulting in the understaffing of projects, has been one of the reasons for lack of informed decision-making leading to *trial and error* decisions being made.

2.12 Using data analytics in decision-making

Introducing data analytics into project management improves the existing project management methodologies and standards and assists in strategic decision-making (Bose 2009:167). As has been alluded to earlier, data analytics provides the means to sieve through the huge volumes of data in an organisation to come up with actionable information from the data (Buse & Zimmermann 2010:77; Talia 2013:98). The use of data analytics to extract meaningful information will thus turn the project management process from just being a process monitoring practice and translating operational data into strategic information, which helps to influence decision-making (Bose 2009:171; Buse & Zimmermann 2012:988).

Data analytics being part of big data, comprises of analytics and business intelligence, of which big data refers to unstructured, voluminous data produced by various applications that requires further analysis to extract actionable information (Hansmann & Niemeyer 2014:44; Cuzzocrea *et al.* 2011:101). Big data leverages on both the structured and unstructured data and has successfully been used for predictive analysis in such fields as crime prevention, law enforcement and the medical fraternity (Tene & Polonetsky 2013:247; Bose 2009:162; Chaudhuri 2012:2). Consequently, this has influenced the premise of this study, that data analytics can also be used to get actionable information out of projects data, improve decision-making and ultimately attain satisfactory software project management quality.

The proposal to use data analytics for decision-making improvement is beneficial, as has been echoed by Delen and Demirkan (2013:361), who state that data analytics facilitates the foreseeing of future problems or opportunities and allows optimisation of processes, thereby enhancing performance. This makes data analytics a crucial element for extracting actionable information from data to influence decision-making. In agreement, Zhang *et al.* (2011:1) state that data analytics facilitates data-driven tasks, achieves appropriate decision-making and helps anticipate future problems. In addition, the statement by Singh (2015:1) that data analytics uses analytics-based metrics suggests that project managers will make rational project decisions with analytical certainty.

Singh (2015:8) further indicates that good decision-making is made possible through predictive analytic models which can be used to analyse the available factors before making rational decisions to effectively manage a risk. Likewise, Bose (2009:155) in agreement with Singh (2015:9) further states that the handling of statistical analysis which assists in predicting certain processes or systems' behaviours or projects in environments of uncertainity is essential for data-driven decision-making which becomes quicker and more accurate. These data analytics functions will potentially improve decision-making and assist in directing or redirecting projects. It has however not yet been substantially established if data analytics directly influences the quality of software project management. This is despite a survey finding by the Massachusetts Institute of Technology, as indicated by Tene and Polonetsky (2013:243), which confirmed that data-directed decision-making assists in information analysis and strategic planning, which resulted in a six percent, 6% increase in productivity.

This increase in productivity suggests why data analytics, as has also been observed by Bose (2009:155), is increasingly being focused on across all industrial sectors for better operational, tactical and strategic decision-making. Therefore, it can be inferred that incorporating data analytics into software project management is anticipated to induce strategic planning and improve the analysis of projects' user requirements data. This will facilitate extracting what forms the clear objectives a system under development will deliver from the gathered user requirements. The study, therefore, establishes the effect data analytics use by software project managers has on software project management quality.

Introducing data analytics into user requirements gathering and processing of data is potentially beneficial in many ways. This is particularly so as Ularu *et al*. (2012:7) indicate that data analytics handles structured and unstructured data, thus offering flexibility in the processing or analysis of the data, leading to quick and improved decision-making. In addition, the ability to handle highly unstructured data allows the user requirements gathered through interviews, documentation and observations, which are not in the same format at the point of capture, to be processed easily and accurately. The fact that data analytics offers deeper data analysis and insights by using new analytical methods as indicated by Ularu *et al*. (2012:6) and Chaudhuri (2012:2), makes user requirements and other projects data analyses quicker, while bringing better prospects of improved decision-making accuracy.

Furthermore, Ularu *et al*. (2012:7) assert that data analytics combines the best tools as opposed to the traditional analysis tools that depend on relational databases for query execution and data storage. The use of better analytical tools indicates that data processing speed, accuracy and availability will subsequently improve when using data analytics. Likewise, Buse and Zimmermann (2012:988) also anticipate that data analytics would revolutionise decision-making by offering project managers flexibility in decision-making, using dedicated analytic tools.

Similarly, Stolovitsky (2011:2), in agreement states that effectively incorporating data analytics into project management would allow data-driven decision-making, thereby turning project management into a strategic asset for effective decision-making. In support of this notion, the Analytics Advantage Survey results published by Deloitte (2013:3) indicate that a significant forty-nine percent, 49% of the respondents are of the view that the main benefit of analytics use is that it is a critical factor in better decision-making, which works by influencing decision-related cultures. Due to these contributions, data analytics therefore, can be viewed as being potentially key to improved data quality and decision-making, with its role becoming increasingly significant (Bose 2009:171), which this study suggests it can also be capitalised on to improve software project management decision-making.

2.13 Complementing project management standards with data analytics

The significance of data analytics in software project management lies in its potential to reinvigorate data analysis, project management standards and methodologies. Data analytic systems, by means of information flow evaluation, enable data and other project management issues to be identified in almost real time and to be corrected quickly (Williams *et al*. 2014:315). This is beneficial during quality control when using standards such as Project Management Body of Knowledge (PMBOK), and the use of data analytic tools will enhance the quality control's inspection or "walkthrough" process.

Data analytics makes such techniques as inspection or walkthroughs to become quick and more useful. The outcomes from these tools get to be implemented before a project's completion for corrective action, rather than to be used for review purposes after the project's completion (Williams *et al*. 2014:313). That brings greater project control as analytics can be used during the process to break down a project's processes and systems to predict their behaviour and outcomes (Singh 2015:4). The realised greater project control makes it possible for strategic decision-making to keep projects on course and within budget, based on the predictive information. The ability to use analytics to evaluate unstructured data previously not possible creates an opportunity for the new analysis methods to allow deeper insights in project management methodologies (Williams *et al*. 2014:313). Therefore, complementing project management standards and methodologies with data analytics will bring deeper data analysis and flexibility in decision-making during project management.

2.13.1 Enhancing quality control's Inspection technique with data analytics

Quality management complements modern project management as it emphasises on prevention and not mere inspection (PMI 2008:190). As part of ensuring quality during user requirements gathering, Anda *et al*. (2002:128) state that a quality control's inspection, which is an evaluation technique, assists by examining the user requirements and design to detect violations on the set standards. The handling and comparison of data in the inspection process needs accuracy, which is inherent in data analytics.

As expressed by PMI (2008:124), inspection, which is an element of project management quality control, includes measuring, examining and testing and is conducted to ascertain if the project results conform to requirements. Inspections such as walk-throughs, reviews or audits which require improvement may, as indicated by PMI (2008:213), be conducted at any stage of the project to measure individual activity results or at the completion of a project, measuring the deliverables or the final product.

Anda and Sjøberg (2002:134) point out that, with different stakeholders involved in a project, inspection becomes a useful means to improve on quality as part of the review and checklists measure against the adopted quality policy elements. Data analytics would aid in detecting trends in past projects to assist in identifying potential risks and making data-based decisions when using the inspection techniques. On the contrary, Anda and Sjøberg (2002:128) however highlight that assessing data using walkthrough, checklist or review has some limitations. Such limitations are the gap which data analytics is expected to address.

2.14 Theoretical Framework

Theories and models assist with the investigation of the factors which influence people to use computers, applications and their acceptance behaviour (Taiwo & Downe 2013:48). Imenda (2014:189) defines a theoretical framework as the use of a philosophy or a collection of ideas acquired from a theory to give an account for an event, or to clarify a research problem. An established theoretical framework guides this study since in its absence a research lacks proper direction. A theoretical framework therefore, gives the researcher a specific perspective to discover, understand and explain events of the area under study (Imenda 2014:188). This research uses a quantitative study approach and makes use of the Capability Maturity Model Integration (CMMI).

2.14.1 Capability Maturity Model Integration (CMMI)

CMMI is a model which improves projects' quality through its organised view of refining processes in an institution and providing direction on quality procedures, and offering a benchmark for evaluating existing practices (Modeling 2005:17). This model is instrumental in improving quality in projects as outlined by Modeling (2005:21) through its use of systems engineering rules in the development of software in addition to its increased focus on the requirements development, management and systems development. Judgev and Thomas (2002:6) concur that using CMMI has enabled software development organisations to deliver systems with predictable results by evolving the affected, benefitting or related organisations' project management, from immature stages to the solid structures, accompanied by the relevant supporting infrastructure.

Khalid (2008:305) further states that this level of transformation is due to the CMMI providing guidance for quality processes by organising these processes into five levels of maturity to support and guide process-enhancement and to serve as a reference point for appraising current methods. The five CMMI levels of maturity responsible for the transformation stages towards quality in an organisation's projects are Level 1 - Initial, Level 2 -Managed, Level 3 - Defined, Level 4 - Quantitatively Managed and Level 5 - Optimised (Thomas & Jugdev 2002:6).The CMMI (2001:11–13) summarises these levels as follows:

Initial phase	Is characterised by poorly controlled, reactive and chaotic processes, with the organisation's success mainly dependent on individual competence, and not on proven organisational processes. Projects at this level, though, produce working products that usually exceed budget and schedules, with processes usually being abandoned and there being a failure to repeat past successes.
Managed level	Involves projects being performed according to the documented plans, with managed requirements, planned and controlled processes. Work products and services meet their specified requirements and standards.
Defined level	Organisations have established set standards, which creates consistency across the organisation, with more detailed process descriptions tailored for the organisation. Processes are described in more detail and are proactively managed, with a better understanding of the processes' interrelationships.
Quantitatively Managed	Is characterised by organisations that have realised all the goals from the previous three stages and focuses on a process' performance and quality improvement, with an enhanced performance of the processes' predictability. Quality and process performance measures are used to support future fact-based decision-making.
Optimising level	Organisations focus on the continuous improvement of processes through innovative incremental technological

improvements. Processes focus on correcting the common sources of process variation to achieve set objectives.

Despite the alleged inflexibility of CMMI by some including Thomas and Jugdev (2002:6) and its supposed ignoring of people focussing on processes (Bach 1994:2), the model's processes have a transformative capacity to improve projects' maturity and quality. The quality management levels introduce quality measures as it has been established by Ika (2009:314) that attaining higher organisational levels of process maturity lead to higher quality being attained.

2.15 Hypothesis and Conceptual Framework

2.15.1 Hypothesis

Given all the above, the following four hypotheses are constructed to establish the impact of data analytics use on software project management quality as assessed through a project's success, Impact on Customer, preparation for future project management and project management efficiency:

Hypothesis H1: Data analytics use in software project management is positively related to projects' successes.

Hypothesis H2: Data analytics use in software project management is directly related to a project's Impact on Customer.

Hypothesis H3: Data analytics use in software project management is directly related to organisational success.

Hypothesis H4: Data analytics use in project management is directly related to improved project management efficiency.

Hypothesis H5: Data analytics use with software projects management models and standards prepares organisations for future project management.

2.15.2 Conceptual Framework

Conceptual Framework Diagram

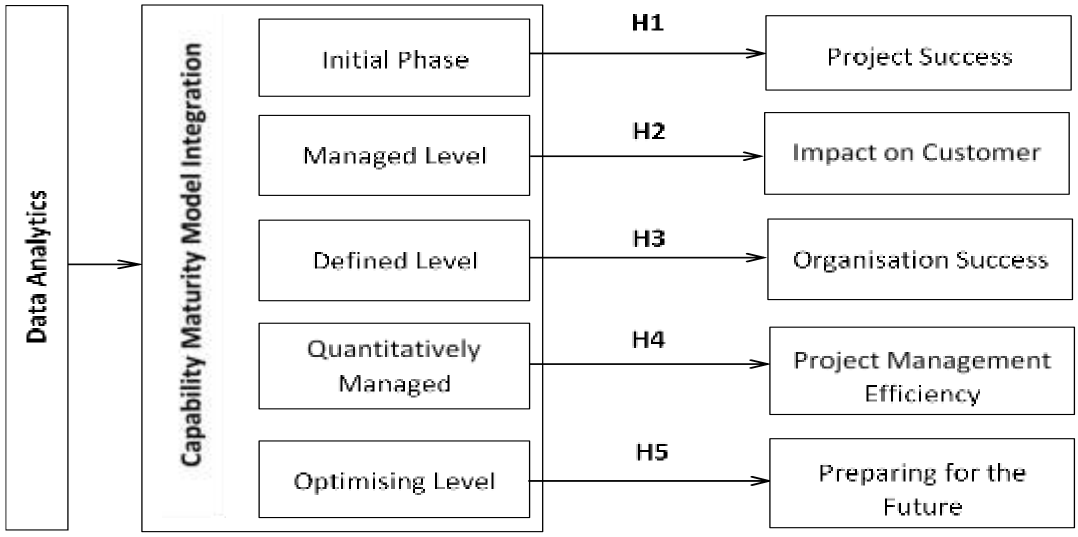


Figure 2.1: Conceptual Framework

The purpose of this study is to establish if data analytics use in software project management improves project management quality. The conceptual framework section of this study defines and explains the key concepts, factors and variables of this study and how these are related. According to Tamene (2016:51), conceptual framework is a structure of ideas, beliefs, expectations, assumptions and theories that support a research by linking the core components of the research design responding to the research question.

The CMMI model is one of the models being used to minimise risks in an endeavour to improve project management quality. Its levels of maturity's policies guide organisations on improving project management. It is extensively being used in the United States and Europe in countering risk when awarding tenders, but its application in South Africa has been limited due to high costs and complexity (Van de Groenendaal 2017). Powell-Morse (2017) acknowledge that, currently, there are only 5 000 companies in more than seven countries using CMMI as a process improvement standard. There has been an annual increase in the use of CMMI in China, the US and India (CMMI 2017). However, the Johannesburg Centre for Software Engineering (JCSE) at the University of Witwatersrand introduced CMMI to companies in South

Africa at a much-reduced cost (Van de Groenendaal 2017) to encourage its adoption and improve on project management quality. CMMI's levels are summarised as follows:

i. **Initial Phase and Project Success**

At this level much of the work is chaotic and ad-hoc and there are no set standards for processes in place and organisations do not have repeatable processes. According to the CMMI (2017), 42% of organisations do not have established standard planning processes, with 54% of the organisations not participating in any form of measuring.

ii. **Managed level and Impact on Customer**

The managed level is established by adhering to organisational policies, following establishes processes, descriptions and plans and monitoring and controlling processes performance against set plans and management requirements, among other activities (CMMI 2005). IBM SA is one of the organisations which attained the Managed level Key Process Areas (KPA) (IBM Africa Magazine Issue 08). In the initial stages of CMMI implementation, very few companies took up the implementation initiative. For an example, in July 2007, only 22 companies in Pakistan were implementing CMMI level 2, and only two of the companies attained the level 2 in that year (Shaa et. al 2012:1006)

iii. **Defined level and Organisational Success**

Organisations that attain this level will have well-defined engineering processes for process areas that include decision analysis, integrated project management and organisational process definition (Broadsword 2019). The majority of organisations working with the CMMI model are at this level. The JCSE has successfully worked with several companies in South Africa, with Bytes Group attaining CMMI's level three, having realised all the process' level two and level three attainment on several assigned goals (Van de Groenendaal 2017). Despite these successes, there is a perception that the CMMI model is now old-fashioned, too formal and process-centred, as opposed to the other models, including the Agile which the big banks in South Africa had switched to using, but have since started looking for a more formal process akin to the CMMI (Van de Groenendaal 2017).

iv. **Quantitatively Managed Level and Project Management Efficiency**
Organisations at this level are characterised by process performance predictability, with the projects processes controlled through the use of quantitative techniques. Organisations getting to this level believe in the investment based on the previous level that it would yield big returns by streamlining and automating the processes, eliminating the documentation burden and accelerating learning across projects (Alder, McGarry, Irion-Talbot and Binney 2005:221)

v. **Optimising Level and Preparing for the Future**
This is a level attained by very few organisations, with the current CMMI institute database revealing that only 344 organisations have been rated level 5 worldwide (IEEE COMPUTER SOCIETY 2014:80). Antoniol, Gradala and Venturi (2004:33) acknowledge that moving towards level 4 and 5 requires a radical change in the way projects' life cycle activities are approached. A high level of diverse expertise is required for an organisation to attain level 5, including quantitative measuring techniques, statistical analysis, change management and process optimisation (IEEE COMPUTER SOCIETY 2014:83). The 2019 updated list (Sharma 2014) indicate that there are 60 companies in India which have attained CMMI level 5.

Hypotheses H1, H2, H3, H4 and H5 were tested using the collected survey data to establish if data analytics use in software project management processes influence software project management quality.

Variables and Measures

The variables used in the study are from three constructs; data analytics use, project management processes maturity, and projects success quality.

Predictive data analytics use influence will be assessed based on the following:

i. Processes and data analysis improvement.
ii. Prediction of project event outcomes (Mishra & Silakari 2012:4434)
   - Financial Predictions and gain
   - Dealing with projects deviations or processes control

iii. Insight for strategic decision-making (Mishra & Silakari 2012:4438; Fauser *et al*. 2016:68)
iv. Resources Management (Fauser *et al*. 2016:72)
    - Resource planning and management

Project management process maturity's contribution to quality will be assessed through organisations' capability in progressing their project's management maturity levels (Initial, Managed, Defined, Quantitatively Managed and Optimised), and in improving time management, quality management, scope management and integration.

Project quality assessed through a project's success will be measured using the following criteria, which aligns with the indication by Agarwal and Rathod (2006:359):

i. Customer satisfaction, which is part of Impact on Customer;
    - Software delivered to the customer within planned time and cost.
    - Meeting software functionality goals, including reliability (Issac *et al*. 2004: 317).
    - Satisfying user operational requirements (Agarwal & Rathod 2006:368)**.**
ii. Meeting project goals, which is part of Project efficiency;
    - Achieving planned project goals (Agarwal & Rathod 2006:367).
    - Handling changing requirements (Agarwal & Rathod 2006:361).
iii. Contractor benefit, assessed through organisational projects' success;
    - Financial gain from the project.
    - Consistency in projects' completion according to set plans.
iv. Preparation for future;
    - Preparation for future quality technology use.
    - Contribution to project management body of knowledge.

In constructing this study's conceptual framework, consultation was made from different sources to inform this framework, as suggested by Tamene (2016:52).

There are, essentially, four potential sources of information that are used to build a conceptual framework. Maxwell (2012:223) outlines these information sources as being; experiential knowledge, existing theories and research, pilot and exploratory studies and thought experiments.

These are explained as follows:

Experiential knowledge - refers to the contribution one brings to the research from their background.

Existing theories and research - Includes published and unpublished work, dissertations and presentations, which help in organising data, testing or modifying personal theories. The established theories assist in developing personal theories and presenting alternatives, instead of using the already established theories. Established theories, however, have the potential to deform an argument if individual insights are fitted into the established theories.

Pilot and exploratory studies - Involves the designing of pilot studies to test an idea or method or to develop grounded theory. Pilot studies improve the understanding of concepts and philosophies held by other theorists.

Thought experiments - Combines the aspects from theory and experiments, as it focuses on the implications of the properties of a study area. It encourages creativity and exploration since it is used to test a theory for logical problems, and to generate new theoretical insights.

Both theoretical and conceptual frameworks provide a unified meaning of issues in a study area, enabling the addressing of a researcher's specific problem (Imenda 2014:189). As suggested by Imenda (2014:193), the variables and concepts for this study have been identified through establishing the study's conceptual framework. The experiential knowledge of the researcher on data analytics use in project management is very limited to be used as a source of concepts to formulate this study's conceptual framework. Experiential knowledge, therefore, could not be used to come up with a conceptual framework for this study.

Likewise, the pilot and exploratory method was also not considered a viable option in this study, since the researcher required established frameworks to guide this study from its conception. The thought experiments has also not been considered the ideal option for this study as this requires one's knowledge of experiments, drawing from experience, in which case the researcher has already acknowledged his limited experience in data analytics use in project management. The researcher has therefore used content analysis, by reviewing and analysing literature, including theories and research work, as suggested by Krippendorff (2004:84) and Jabareen (2009:52) to develop a conceptual frame work for this study.

2.16 Chapter Conclusion

The traditional project management methodologies and standards have been identified to be limited and requiring improvement to enhance software project management. There is potential in augmenting these standards and methodologies with data analytics, which would improve data quality, decision-making and software project management quality. This, as supported by the reviewed sources, is presumed attainable if the potential in data analytics is harnessed into the project management processes which improve on processing speed, accuracy and decision-making. Such exercises, including the quality control's inspection techniques can become even better and help to correct issues during a project rather than after the project as part of a review. Integral to this anticipated improvement is the need for data quality to also be improved through rigorous checks and validation during creation, and the processing of the requirements data, as good decision-making also depends on data quality.

Chapter 3: Research Methodology

3.0 Introduction

This chapter of the study outlines the research methodology, and the approach used in this study for the effective collection of data and effective analysis. Research methodology describes how the research is conducted, while outlining the work plan for the research towards a solution to the identified problem (Rajasekar, Chinnathambi & Philominathan 2013:5). Scotland (2012:9) describes research methodology as a strategy or action plan behind the use of chosen research methods. Therefore, this chapter includes the explanation of the philosophy used in the research, the study approach, data gathering methods and the analysis techniques used to realise the goals of this study. This chapter does not only discuss the methodology and techniques employed, based on the chosen philosophical approach, but as suggested by Kothari (2004:8), also considers the logic behind the chosen methods and techniques selected to answer the research question.

3.1 General research procedure outline

The research approach used in this study follows the Onion Research process as explained by Saunders, Lewis and Thornhill (2009:107). Saunders *et al*. (2009:107), by means of the Onion Research process, explains the research methodology approach as encompassing the research philosophy, research approach, research strategy, research choice, time horizon and data collection techniques and analysis procedures. The research procedure sections are illustrated in Figure 3.1

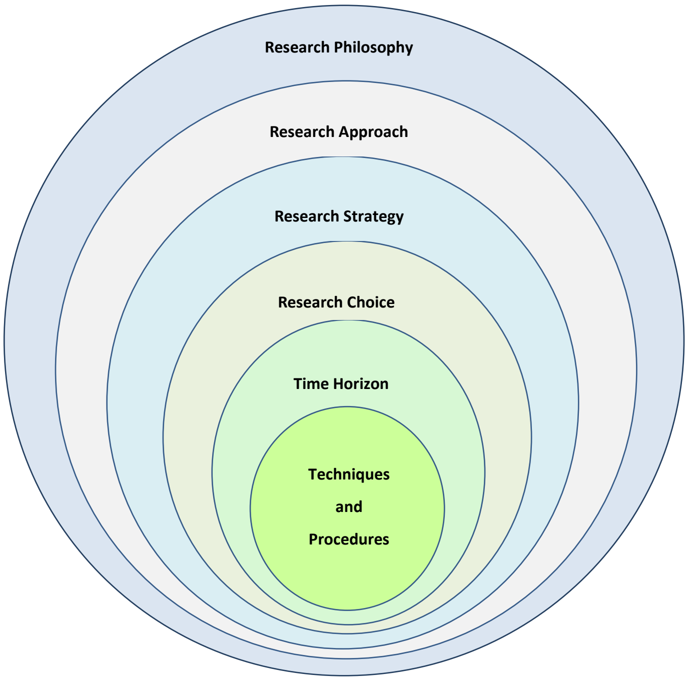


Figure 3.1: Research Process Onion (Saunders et al 2009:108)

3.2 Research Philosophies

A research philosophy outlines the philosophical basis of the study, which forms the basis of the link between the theoretical basis and the practical aspects of the study. As guided by Crotty (1998:3), this section further discusses what the research intended to achieve through a defined methodology and techniques based on the chosen philosophical approach set to achieve the research goals.

According to Dudovskiy (2019:1), there are mainly three research philosophies, namely; Positivism, Interpretivism and Critical Research which are briefly described as follows:

- Positivism makes the assumption that the world is orderly and that we can investigate how it works objectively, without interference from our personal feelings (Ormston, Spencer, Barnard & Snape 2013:10). Oates (2006:284) further states that positivism allows establishing cause and effect based on hypothesis and theories testing, with its findings based on repeatability, producing the same results, and is mainly associated with quantitative data analysis.

- Interpretivism refers to approaches focusing on how we can gain knowledge through interpreting and understanding meanings attached to human actions (Reilly 2009:2). The central tenet, as argued by Bryman and Liao (2004:1) is that the study of social phenomena requires an understanding of people's social world they live in, which they have created and have interpreted the meaning they produce.

  Similarly, Ritchie *et al.*(2013:12) also indicate that Interpretivism centres around producing knowledge by exploring and understanding people's social world, focusing on their meanings and interpretations. Aliyu *et al.* (2014:82) as well as MacDonald and Headlam (1999:8–9) concur that Interpretivism has a strong preference for qualitative data analysis, which provides rich and detailed results from mostly non-numeric data analysis. Oates (2006:295), in addition states that Interpretivism has to show descriptions, explanations and interpretations supported by data evidence generated from interviews, observations, questionnaires and documents.

  Interpretivism is considered as not being the most appropriate philosophy for this study, considering the view by Scotland (2012:12) that its methods explain the participants' actions and bring an understanding of their behaviour, but then disregard the external structural factors that are influencing behaviour. Although software project managers' behaviour determines the extent of data analytics adoption in project management, the focus of this research is not based entirely on adoption, but on its impact on project quality management. Since interpretive research findings represent the participants' sociological understanding, the results of this study would be greatly limited if Interpretivism is sorely used since issues of data use, ownership and the extent of control of findings by participants are not uncommon, as suggested by Scotland (2012:14).

- Critical research postulates that people create their social reality and that the social realities possess objective properties that dominate our experiences (Oates 2006:296). Oates (2006:296) further indicates that Critical Research focuses on trying to challenge and eradicate power relations, conflicts and differences in our society and organisations, as these cause alienation and domination. Critical research is not regarded as ideal for this study, as the research does not focus on challenging any power relations or elimination of any societal or organisational conflicts.

3.2.1 Research Philosophy for this research

This study uses the positivist research strategy, also known by many as objectivism or realism, which Aliyu, Bello, Kasim and Martin (2014:81) indicate as being centred upon the ontological principle that reality and truth are independent of the researcher. Aliyu *et al.* (2014:82) state that the concepts of impartiality, repeatability and objectivity are central to the positivist approach, which makes use of methodologies including quantitative analysis and laboratory experiments. In addition, positivism uses a quantitative approach in investigating phenomena.

Positivism is used in this study due to its principles, which Scotland (2012:10) highlights as allowing the researcher to impartially discover information about unbiased reality by maintaining the independence of the researcher from the study. Since positivism is focused on explaining relationships and finding what influences the outcomes, as indicated by Scotland (2012:10), it is particularly ideal for this study to be able to show the causative effect of data analytics use on improved projects decision-making and project management quality. As advocated by Saunders, Lewis and Thornhill (2009:114), this study by means of the positivist philosophy leverages on its structured methodology and uses statistical analysis on quantifiable observations. This assists in systematically gathering, analysing and interpreting the study's data, given that positivism, as indicated by Oates (2006:286), has a strong preference for mathematical modelling which provides a sound and unbiased way of analysing observations and results.

In addition, Scotland (2012:10) indicates that results from samples are attained through inferential statistics analysis, which is generalised to the populations. One other contribution that positivism brings even to this study, as further expressed by Oates (2006:286) and Scotland (2012:10) is that, with its methodology being neutral enables establishing patterns or indisputable facts from data, irrespective of the researcher's feeling, which as earlier alluded to, allows for the generalisation of its output. Part of the main considerations taken into account for the use of positivism in this study is, as argued by Scotland (2012:11), its implementation of high levels of rigor and methods, which attempt to yield commonly-accepted results.

The study comes up with a set of recommendations that will assist project management professionals to take advantage of data analytics to improve software project management quality. To achieve this, the research is guided by the earlier defined research questions, the use of hypothesis, and makes use of the positivist theoretical perspective.

3.3 Research approaches

There are different levels of research methods classification and description, with the most basic among these being philosophical (Crossan 2016:48). Levy (2006:374) defines a research method as the planned course of action behind the selected methods in a study, where the selection and the use of methods are linked to the expected outcome. There are, essentially two research approaches, which are qualitative and quantitative research methods. The Qualitative Research approach, according to Kothari (2004:5), focuses on the subjective assessment of attitudes, behaviour and opinions; and the approach generates results that are non-quantitative and cannot be quantitatively analysed. In addition, Dawson (2002:14) states that qualitative research seeks to get an in-depth opinion from participants, making use of such methods as interviews and focus groups.

The quantitative approach, on the contrary, and as explained by MacDonald and Headlam (1999:8), quantifies data, subjects this to rigorous, formal and quantitative analysis and generalises the results from a sample of the population that is being targeted. The quantitative study entails the researcher to remain objectively-separated from the study. Kothari (2004:5) suggests that quantitative research can be further sub-classified into three approaches as follows:

- Simulation – involves establishing a mock setup from which relevant information can be generated. This allows the monitoring of a system in a controlled environment and the building of models to understand future conditions.

- Inferential – This includes establishing a database from which the characteristics of a population will be inferred. It generally takes the survey form, where a sample population is studied, and it will then be inferred that the observed characteristics are representative of the population.
- Experimental – It involves better control of the study environment, with variables being changed in some instances to monitor how these impact other variables.

3.3.1 This study's research approach

This study uses the quantitative research approach, as it explains phenomena through the gathering of numerical data which is analysed using mathematically-centred procedures. Quantitative research is used as Dawson (2002:15) indicates that it makes use of the survey research, employing the questionnaire method in this study. Surveys allow for the gathering of data from a broader spectrum of people at a manageable cost and the findings from the survey are generalisable.

The fact that a quantitative approach requires the researcher to remain objectively-separate from the subject under study ensures unbiased outcomes from the research, making it ideal for this study. The study follows a deductive approach, using a hypothesis and the CMMI maturity levels to assess the impact of data analytics use on project management quality through improved decision-making. Since the deductive approach begins with an expected general pattern which is tested against observations (Morgan 2014:48), it allows this study's hypothesis to be tested against observations from the study. Morgan (2014:48) further states that the deductive procedure of transitioning from theory to observation is linked to specific targets, for instance, connecting causes to effects, which enables this research to establish the causal impact of data analytics use on project management quality.

3.4 Research Strategy

Oates (2006:25) defines a research strategy as a general approach in responding to the research question. Research strategies include a survey which is used in this study, ethnography, experiment, design and creation, action research and case study.

3.4.1 Strategy for this study

This study uses the survey research strategy, making use of a questionnaire. A survey was used as it allows the study to get the same type of data from many people in a homogenous and methodical way and confirms an association between the elements under study (Oates 2006:104). The survey questionnaire, as argued by MacDonald and Headlam (1999:35), enables the researcher to establish the depth of opinion on the subject being studied. The survey strategy is very much associated with positivism, as it seeks to establish patterns and generalisations (Oates 2006:93).

3.5 Time Horizon

This study was conducted over a period of 36 months. Given the time limitation of the study, data gathering is based on sample population of software project managers, which enabled the gathering and subsequent analysis of data in the available time. The actual gathering and analysis of the data was done in less than 24 months, but this time limitation does not suggest any compromised data gathering and analysis in this study.

3.6 Data Collection and research sample

A survey research methodology was chosen to gather data to test the research hypothesis. Questionnaires were sent to Software development organisations and to Project Management Institute members in South Africa, mostly in the Gauteng province. Since surveys allow for the acquisition of data from many people or events, using standardised and systematic means and establishes data patterns, as indicated by Oates (2006:93), this allows the related findings to be generalised to a much bigger population than the actual survey group. The questionnaire survey method has been used as Johnson (2011:2) indicates that it is relatively easy to administer and provides a wide range of information that can be easily categorised and is economical to conduct.

Questionnaires have been used to obtain data from software project managers in software development organisations. Data collected has been about the inadequacies of project management standards and methodologies, in addition to the project managers' experiences with data analytics and how that has influenced their software project management processes. A document review provided information on the traditional software project management shortcomings that may have been experienced and helped to enlighten how data analytics use may improve decision-making and software project management quality.

3.6.1 Sampling

This study was conducted over a period of 36 months with limited financial resources to gather data, therefore a sample population was used to collect data from the software project management professionals. A sample is a subgroup or representative of a population which may be probability-based for a required level of confidence in the collected data or non-probability based, which does not involve calculating and knowing the level of bias or error in the data (Latham 2007:1). The researcher's goal in relation to the data-collection from certified project managers and ensuring validity of the collected data, therefore influenced the sampling method used in this study.

3.6.1.1 Sampling Frame

The online directory of software project management companies in South Africa and registered and accredited software project managers from the sampling frame were used in this study. A sampling frame is expressed as a group of elements in a target population or documents, which can be contained within a survey from which a sample is selected and the related data is generalised to the entire population (Herek 2012:2).

3.6.1.2 Sampling Technique

The study used Purposive sampling and Snowball sampling to get the questionnaires to the respondents.

Purposive Sampling

Purposive sampling involves the researcher picking a sample, including instances that are likely to produce important data to realise the research purpose (Oates 2006:98). Purposive sampling is used in this study as the researcher identified and sampled software project managers who are accredited with the Project Management Institutions in South Africa and Software project managers from software development organisations, most of whom were based in Gauteng, some of these with experience in data analytics use.

Snowball Sampling

Snowball sampling has been used in conjunction with purposive sampling in this study to get questionnaires to the respondents. Oates (2006:98) states that Snowball sampling is used when the researcher gets one person from the target population and gets data from the person before asking for suggestions about more people who may be relevant to the research subject. Snowball sampling was chosen for this study as it made it easier for the researcher to contact respondents from organisations the researcher's company outsources from, members of the Institute of Information Technology Professionals of South Africa, the Project Management Institute, and got further referrals from these initial respondents.

3.6.1.3 Sample Size

The sample size, according to Glasow (2005:2–2), depends on the extent of the required precision, required statistical power or the researcher's ability to gain access to the study's subjects. This study uses a questionnaire for which there were responses from one hundred respondents, which was conducted in line with the suggestion by Oates (2006:100) that a sample of less than thirty responses produces unreliable analysis results. More than 300 questionnaire copies were issued, since it was anticipated that some of the given questionnaires would not be returned, as Oates (2006:100) indicates that a response rate of just ten per cent is not uncommon.

3.6.2 Data gathering techniques

Questionnaires have been used to gather data from software project managers on their current practices in Software project management and quality monitoring techniques. Furthermore, the questionnaires have also been used to establish the software project managers' experiences in project management and their level of knowledge with data analytics use in project management. A documents review was also used to collect data on the project management standards, methodologies and the developments in data analytics use, of which the use of project management standards in conjunction with data analytics use is part of this study's drive to establish the possibility of using the two to improve the quality of software project management.

3.6.2.1 Questionnaire design and distribution

a. Questionnaire design

A questionnaire as defined by Siniscalco and Auriat (2005:2) is a set of pre-defined questions, organised in a specific order that is designed to provide the researcher with data for analysis and interpretation. The questionnaires that were used included closed rating questions, which Glasow (2005:2–8) indicates as allowing for speedy completion, while keeping the participants' responses within certain summarising responses, thus making classification and analysis of the responses easier. In line with the suggestion by Oates (2006:222) not to confine respondents to predefined responses, the questionnaire also has open-ended questions requesting respondents to self-report on their project management practices and their perceived projects success. The questionnaire is divided into demographics, background Information and open-ended sections and also seeks to measure the projects that were carried out not more than three years back.

b. Research Variables and Measures

As had been alluded to earlier in the preceding chapter, the variables used in this research are from the constructs data analytics use, project management processes maturity and projects success and quality.

Predictive data analytics use influence will be assessed based on the following:

i. Processes and data analysis improvement.
ii. Prediction of project event outcomes (Mishra & Silakari 2012:4434)
    - Financial Predictions
    - Dealing with projects deviations or changes
iii. Insight for strategic decision-making (Mishra & Silakari 2012:4438; Fauser *et al*. 2016:68)
iv. Resources Management (Fauser *et al*. 2016:72)
    - Planning and resource allocation

Project management processes maturity's contribution to quality will be assessed through the organisations' capability in progressing in their projects' management maturity levels (Initial, Managed, Defined, Quantitatively Managed and Optimised);

improving time management, quality management, scope management and integration.

Project quality assessed through the projects' success will be measured using the following criteria, which aligns with the indication by Agarwal and Rathod (2006:359):

i. Customer satisfaction, which is part of Impact on Customer.
    - Software delivered to the customer within a planned time and cost.
    - Meeting software functionality goals, including reliability (Issac *et al*. 2004: 317).
    - Satisfying user operational requirements (Agarwal & Rathod 2006:368)**.**

ii. Meeting project goals – Project efficiency.
    - Achieving planned project goals (Agarwal & Rathod 2006:367).
    - Handling changing requirements (Agarwal & Rathod 2006:361).

iii. Contractor benefit assessed through organisational projects success
    - Financial gain from the project.
    - Consistency in projects completion according to plan.

v. Preparation for the future:
    - Preparation for future quality technology use.
    - Contribution to project management quality body of knowledge.

Further to the success measures described above, a factor measuring contribution of data analytics to future project management and contribution to project management quality is also included in the questionnaire.

c. Validation and the Reliability Assessment

The questionnaire was assessed for reliability using the Cronbach's Alpha Method, which Tavakol and Dennick (2011:53) suggest as enhancing the questionnaire's accuracy of assessment and evaluation. Tavakol and Dennick (2011:53) also indicate that validity focuses on the extent to which an instrument measure what it intends measuring, while reliability looks at the instrument's ability to measure consistency. To measure the internal consistency of a test, Santos (1999:1) suggests that Alpha may be used to describe the reliability of factors taken from multi-point formatted scales or questionnaires. Internal consistency ensures validity as it looks at how related items in a test are constituted, and describes the degree to which the elements in a test measure the same concept (Tavakol & Dennick 2011:53). Field (2009:675) highlights that by using the Alpha to measure the reliability of the questionnaire, the reliability estimations will show the amount of errors that can be found in a related measurement in a questionnaire test.

Using some closed ended questions in the questionnaire creates a level of similarity from the sample, which adds to reliability. To further check on reliability and minimise measurement errors, a pilot run of the questionnaire instrument was conducted as suggested by Kimberlin and Winterstein (2008:2277) to ensure consistency before the questionnaire was sent out to the respondents.

d. Questionnaire Distribution

Three hundred and twenty questionnaires were distributed to the respondents through LinkedIn and email of which one hundred and fifteen of them were completed and returned. One hundred complete responses were used for analysis as eleven were incomplete and four were returned without the participants' consent forms. Ninety-seven of the questionnaires were self-administered by the respondents, while eighteen were researcher-administered which Oates (2006:221) suggests encourages the completion of the questionnaire by respondents and increases the questionnaire response and return rate. A log was created to keep track of the questionnaires given out and reminders were sent fortnightly to achieve the speedy completion and return of the questionnaires. As suggested by Siniscalco and Auriat (1998:72) a questionnaire pilot was run with five, project managers to determine if there were any ambiguous questions, to establish whether the predefined responses covered all desired answers, and to find out the time it would take to complete the questionnaire.

A questionnaire was used in this research as it is relatively inexpensive, easy to administer and is not affected by geographic boundaries, having been distributed by email and in person. As warned by Siniscalco and Auriat (1998:24) and Oates (2006:222), the researcher anticipated that not all the questionnaires would be returned, and that the pre-specified responses in some questions would limit the respondents, where they could have provided their own responses.

3.6.2.2 Online Documents

Online documents are used as a source of information on the existing project management practices, standards and methodologies, as well as the role of data analytics. Oates (2006:235) defines a document as any written material or symbolic representation that can be recorded and retrieved for analysis. Online documents (articles accessed through the internet) have been used to explore recent innovations in data analytics use to device how it can improve decision-making in project management. Documents have been used since they are obtained quickly and conveniently, without making any appointments, and Oates (2006:241) further indicates that, with these being in the public domain assists in checking this research against the available online publications, thus improving on its credibility. However, as Oates (2006:241) highlights that access to some internet-based documents is temporary, a few become inaccessible on initial sites which delayed access to some much-needed information.

3.7 Data Analysis

This research is a quantitative study which collected quantitative data for analysis. Quantitative data analysis involves looking at the collected data to establish general trends emerging from the data (Field 2009:18). Oates (2006:245) states that quantitative data is analysed by techniques, including tables, charts or graphs, which enable the identification of patterns in data. Data Analysis is the process of going through a data set, looking at the parts separately and learn how the data sets are related to other samples of data while data analytics applies to the logic or techniques used to drive the analysis (Brown 2019). Data analytics involves using software tools and statistical techniques to assess data for decision making and improved future performance.

To establish the strength of the correlation between data analytics use and decision-making and the strength of correlation between decision-making and project management quality, Pearson's correlation was used. Gogtay and Thatte (2017:78) indicates that Correlation Analysis measures the relationship between quantitative variables. Data analysis methods employed in this study are as explained in section 3.7.1 and the software used to analyse the quantitative data is outlined in section 3.7.2.

3.7.1 Descriptive Demographics

Reliability

One of the main issues of reliability concerns a scale's internal consistency, such as the extent to which the elements of a scale measures the same underlying construct. Cronbach's alpha has been used to assess the reliability of the related variables. A reliability coefficient of 0.70 or higher is considered acceptable when using the Cronbach's alpha (Tavakol & Dennik 2011:54).

Exploratory Factor Analysis (EFA) Validity

Exploratory Factor Analysis (EFA) is a statistical approach for establishing a data set's variables correlation (Pallant 2013:188). EFA analysis allows the collection of variables based on strong correlations. Factor analysis functions on the conception that quantifiable and observable variables can be reduced to fewer latent variables that share a common variance, and that these are unobservable; known as reducing dimensionality (Yong & Pearce 2013:80). When conducting the EFA, factors are rotated to try and get to a factor solution equivalent to that which is found in the initial extraction but with a simpler interpretation. The two major forms of rotation are Orthogonal, which produce uncorrelated aspects; and Oblique rotations, which produces correlated factors. The best orthogonal rotation is regarded as Varimax, while the Oblique rotations popularly uses the Direct Quartimin, Promax, and Harris-Kaiser Orthoblique (DeCoster 1998:3).

The Mean

The Mean is one of the descriptive statistics used in data analysis in this study. The Mean gives a central tendency, providing an idea of where data seems to cluster around (Kalla 2009). The Mean has been used to calculate the average score for the variables, as part of the descriptive statistics. The Mean has also been used as it includes every value in the data set and produces the lowest amount of error from all the values in the data set (Lund & Lund 2018).

Regression

Multiple regression describes how a response variable linearly depends on a number of predictor variables (Brema 2012:18). Regression is used to identify the strength of the effect of independent variables on dependent variables, helps with the understanding of the change in dependent variable given the change in one or more independent variables and it is also used to predict trends and future values (Statistics Solutions 2013). Regression analysis being a statistical method used to estimate the relationships between independent and dependent variables; is an element of Data analytics which as stated by Wulff (2017), is the overarching discipline encompassing complete data management. Regression analysis has been used in this study, as it indicates if the variables have a significant relationship with a dependent variable and can indicate the relative strength of the different independent variables' effect on a dependent variable (Sarstedt & Mooi 2014:194). Regression has been used in this study to identify the impact of data analytics on:

- Project Success.
- Impact on Customer.
- Organisation Success.
- Project Management Efficiency.
- Preparing project management for the future.

The following regression assumptions have been applied (Sarstedt & Mooi 2014:20):

- The expected mean error of the regression model is zero.
- The variance of the errors is constant (homoskedasticity).
- The errors are independent (no autocorrelation).
- The regression model can be expressed in a linear way.

Multiple linear regression makes use of more than one independent variable as compared to simple linear regression that uses one independent variable (Ray 2015).

The study used the Statistical Package for the Social Sciences (SPSS) analysis software to analyse the quantitative data.

3.7.2 SPSS

Statistical Package for the Social Sciences (SPSS), is a data analysis software that allows for the entering data into a data editor through the data view and the definition of variables characteristics through the variable view (Field 2009:64). SPSS is used as it is a quantitative analysis programme that allows the analysis of data without having to write a complex command syntax (Crossman 2016:1).

The statistical measures observed from the collected data are grouped under descriptive statistics. Descriptive statistics provide summarised information about data, including Ordinal data, interval or ratio data and categorical data (Connolly 2011:7).

3.8 Ethical Considerations

The researcher obtained ethics clearance from the University of South Africa (UNISA) and abides by the Unisa research ethics policy. Ethical issues were considered in this research and this research was guided by the Unisa research ethics rules including:

i. Voluntary participation

Participants identified in this study were furnished with details about the researcher. The reason for the study and the respondents' involvement in the study were explained to the participants before the decisions whether or not to take part in the research were made.

ii. Confidentiality, anonymity and participants' privacy

Anonymity of the respondents and the confidentiality of the collected data was guaranteed to the participants. The respondents were assured that their names and addresses, or names of the organisations they were linked to would not form part of the research publication. It is also explained to the participants that the information gathered was purely for research purposes and would not be used in any way beyond the intended purpose, which violates their privacy.

3.9 Chapter Conclusion

This Chapter explained the research approaches in general and discussed the approach that was used to conduct the study. This chapter further discussed the questionnaire used, the data collection approach used and the data analysis. This chapter also discussed the Ethical considerations. Chapter 4 focuses on data analysis and the data analysis results are then discussed and summerised in Chapter 5.

Chapter 4: Experimentation

4.0 Introduction

This chapter focuses on data analysis and interpretation as described in the Methodology chapter. The analysis of the data was conducted using the Statistical Package for the Social Sciences (SPSS version 24.0). SPSS was used as it is a quantitative analysis programme that allows for the analysis of data without having to write a complex command syntax (Crossman, 2016:1).

The decision variables used in this research are multi-dimensional, and these are *Project Success, Impact on Customer, Organisation Success, Project Management Efficiency, Preparing for the Future*, *Initial Phase, Managed Level, Defined Level, Quantitatively Managed, and Optimising Level*; and they are made up of a number of factors. For the purpose of this research, the correlational, descriptive, and the reliability analysis comprise the respective factors of the variables, but regression was performed on the total scores of selected variables of this study.

4.1 Data Collection

Data collection is the process gathering of information from relevant sources which enables the researcher to answer research questions, test hypothesis and evaluate outcomes (Dudovskiy 2011). A questionnaire survey has been used to collect data from software project managers and the sample surveyed is outlined in section 4.1.1.

4.1.1 Sample

A total of three hundred and twenty questionnaires were sent out, of which one hundred and fifteen, were returned and the survey used one hundred complete responses for analysis and another eleven were not used as they were incomplete, and four were returned as these had come without the participants' consent forms. Questionnaires considered incomplete would have had mainly the Background Information section answered, basically dealing with the demographics which did not contribute much towards answering the research variable questions in Section B. These questionnaires which had less than 15% of total questions answered were considered incomplete and were left out. The questionnaire was distributed to software project Managers in South Africa, mostly in Gauteng. The questionnaires were distributed in person to respondents at the Project Management Institute of South Africa's Gauteng Branch conferences, some through email and the request for participation in the survey was also placed on the Institute of Information Technology Professionals South Africa (IITPSA) Facebook page, while some of the software project managers were also identified and contacted using LinkedIn.

The age of the surveyed respondents in relation to their age ranges and frequency is discussed under the following Age Statistics heading.

Age Statistics

Many of the participants (35%) are between the age of 30 and 39 years, 30 of the participants constituting (30%) were in the 40 to 49 years age range, 20 of the participants (20%) in the 20 to 29 years range and 15 participants (15%) were in the 50 to 59 years age group. No respondents were under the age of 20 or above the age of 59 years. The statistics, as illustrated in Figure 4.1 show that many of the respondents who participated in the survey are in the middle aged range, from 30 to 49 years, which is a group of people with project management experience, some of whom have used data analytics in their projects.

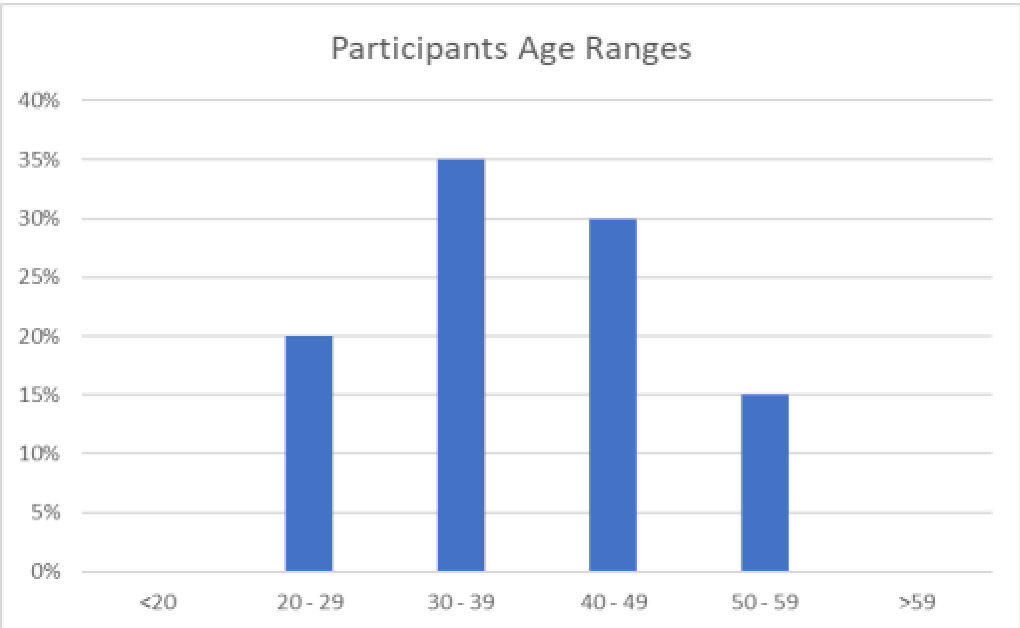


Figure 4.1: Participants Age Ranges

Gender Statistics

Table 4.1: Gender Statistics

Gender	Frequency	Percentage
Male	55	55%
Female	45	45%

The statistics as presented in Table 4.1 show that many of the participants in this research, 55 (55%) are males while 45 (45%) are females. The study received more input from males than females even though this does not have a bearing on the findings of this study as a separate study conducted by Rodríguez, Montequín, Morán and De Arriba (2017:463) established that gender does not have a significant effect on software project management performance even though slight differences were found on male and female leadership behaviours on project teams leadership.

Table 3.2: Project Management Position Statistics

Position	Frequency	Percentage
Inhouse Project Manager	55	55 %
Consulting Firm Consultant	35	35 %
Freelance Project Manager	10	10 %
Currently not managing software Projects	0	0 %

Most of the survey participants are practicing as in-house software project managers in organisations, constituting 55% (55 participants) while 35% (35 participants) are software project managers employed and practicing within consulting firms as consultants. Software project Managers working for consulting firms brought in their experience from a variety of projects they have handled in different operating environments where they have managed projects. Ten percent of the participants are practicing as freelance Software project managers, who at the time of the survey were not attached to any organisation. There are no software project managers who are currently not practising in software project management among the respondents.

Among the 55% of the project managers working in-house and the 35% working for consulting firms they have some knowledge on data analytics use and 27.78% (25) of the respondents indicated that the organisations they were working for in some form facilitate for their data analytics training, with 60% of them getting access to attend project management workshops. The surveyed project managers who are attached to organisations brought in valued contribution to this study from their experience, as they got further training and upskilling through the organisations they are attached to.

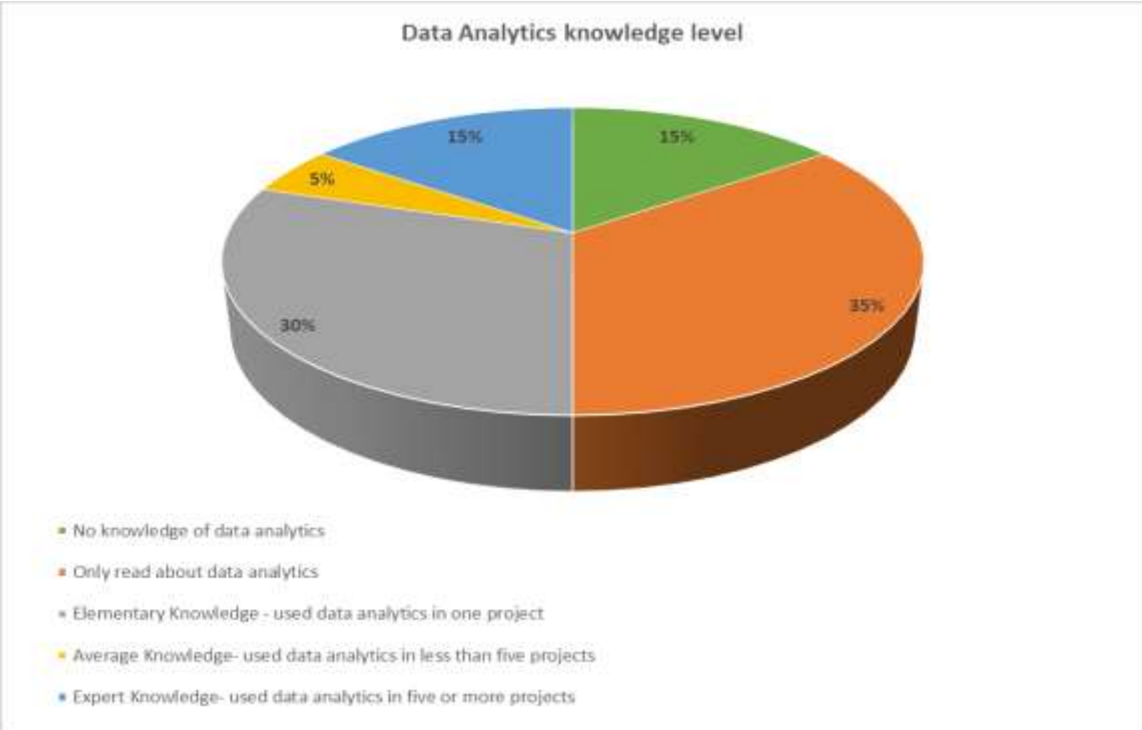


Figure 2.2: Data Analytics knowledge level Statistics

Responses indicate that 50% of the surveyed participants have some experience in data analytics use in project management with 15% of the users having used it in five or more projects, five percent, have used it in less than five projects while 30% of the respondents have used data analytics in only one project. Equally, 50% of the respondents have not used data analytics in project management with 35% only having read about it and 15% with no knowledge of data analytics. The input of project managers with data analytics experience brought into this study the much-needed input from experienced people for effective assessment of its impact. A considerably high percentage of Project managers surveyed, 43.75%, confirm that they had used data analytics use in their past software projects, and that this had enhanced their various projects' processes, and had helped to achieve set project goals

4.2 Descriptive Statistics of Variables and Frequency

This Section discuses and presents the item descriptive statistics for the predictor and predicted variables and starts by presenting statistics per item.

4.2 (a) Descriptive Statistics of Variables

Table 4.3 presents the statistics per item calculated using the SPSS system in which the Item Colum has the decision variables whose Mean and standard deviation is computed for N participants per Item (Decision Variable). Each variable question responses were coded and the participants responses per Variable question where captured using the value code into a dataset. The dataset was imported into the SPSS system to compute the Minimum, Maximum, Mean and Standard Deviation per item (Decision Variable) as shown in Table 4.3.

Table 4.3: Descriptive statistics per item

Item	N	Minimum	Maximum	Mean	Std. Deviation
IP	100	1.50	4.50	2.6250	0.93845
PS	100	1.67	4.33	2.9000	0.84154
MI	90	2.00	5.00	3.3056	1.08797
IC	85	2.00	5.00	3.5294	1.07003
DI	75	2.50	5.00	3.6000	0.86603
OS	80	2.50	5.00	3.7813	0.92382
QML	100	1.00	5.00	3.9000	1.26730
OL	95	1.00	3.67	2.1842	0.89801
PF	90	1.00	3.00	2.0000	0.74953
PME	40	3.00	5.00	4.3125	0.61694

Decision Variables or Items Description

IP = Initial Phase
PS = Project Success
MI = Managed Level
IC = Impact on Customer
DI = Defined Level
OS = Organisation Success
QML = Quantitatively Managed Level

OL = Optimising Level
PF = Preparing for the Future
PME = Project Management Efficiency

Project Management Efficiency, Quantitatively Managed and Organisation Success variables have the highest mean scores of 4.31, 3.9 and 3.78 respectively, with the standard deviations of 0.62, 1.27 and 0.92, respectively. The two decision variables with the least mean scores are; *Preparing for the Future, and Optimising Level,* with respective scores of 2 and 2.18 having standard deviations of 0.75 and 0.90, as shown in Table 4.3 .

4.2(b) Frequency

4.2.1 Quantitatively Managed level

The aim of the variable question was to determine if data analytics use influences decision-making and the effect of improved decision-making on software project management quality. This variable response has a high mean score of 3.9 and a standard deviation of 1.26, as shown in Table 4.3.

The respondents have indicated that data analytics use has an impact on decision-making, with 30.77% of the respondents to the question agreeing that the use of data analytics in their projects led to quicker and timely decision-making. Twenty-three percent (23.08%) of the respondents said that this had led to greater project control, while 15.38% had concurred that there was improved resources allocation and management. Seven percent (7.69%) of the respondents pointed out that data analytics had no effect on decision-making in their projects as they had not used data analytics in their projects. A total of 65 participants responded to the question as to whether or not data analytics use has an impact on decision-making.

Furthermore, many of the respondents affirm that improved decision-making directly impacts software project management. The majority of the respondents, 40%, strongly agree that improved decision-making has a direct impact on software project management quality. Thirty five percent (35%) of the respondents agree that improved decision-making directly impacts software project management quality, while 10% of the respondents somewhat agree. Only 10% of the respondents strongly disagree that improved decision-making directly impacts software project management quality.

In addition, 47.06% of the respondents indicated that in their last projects they had less than 25% delayed or incorrect decision-making due to ineffective data analysis, 23.53% of the respondents indicated they had between 25 and 49%, while 29.41% of them stated that they had experienced between 50% and 74% of delayed or incorrect decision-making due to ineffective data analysis. This strongly suggests that effective data analysis improves decision-making.

A number of software project managers have used more than one quality control technique in their project management, with ten managers indicating that they have used the Cause and Effect diagram, another ten have used the Inspection technique while five have used the flow chart. Twenty-five respondents indicated that at least one software quality dimension was improved by using data analytics, bearing witness to the fact that data analytics does have an impact on quality control tools. Performance, Accuracy, Completeness and Consistency were indicated to have improved by 25% of the participants who responded to the question, with 40% of them stating that performance had improved, another 20% said that consistency had improved, and 25% confirmed that project product serviceability had improved.

4.2.2 Organisation Success

The purpose of this variable question was to find out if software project managers attend ongoing trainings, and if projects stakeholders' involvement and data analytics use contributed to organisational success financially.

The variable has a high mean score of 3.78 given that the average mean for the variables is 3.21 and it has a standard deviation of 0.92, as shown in Table 4.3. The mean score means the majority of the respondents agreed or strongly agreed. Many of the respondents (60%) indicated that they attended regular on the job project management training workshops and courses, with 10% among them strongly agreeing. This shows that the majority of the surveyed project managers keep updating their project management knowledge and skills through workshops which contribute to software project management quality.

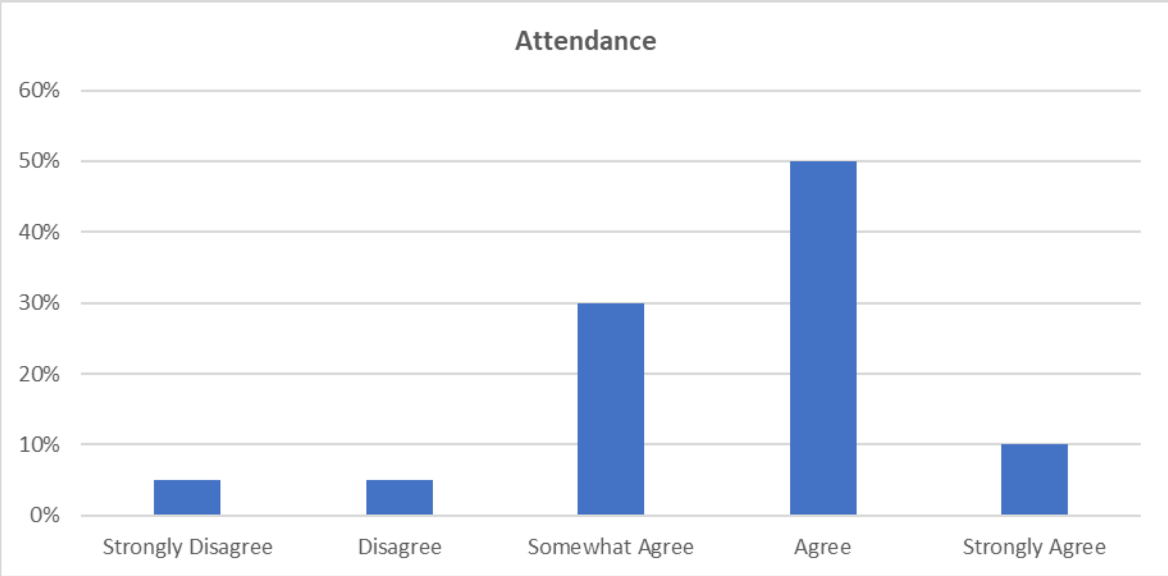


Figure 4.3: On job project management training workshops and courses attendance

Many of the software project managers who had used data analytics stated that data analytics use contributes towards organisational success through system reliability and efficiency. Of the 65 respondents to the question, 30% of the respondents agree, 10% of those who strongly agree that the use of data analytics in their projects led to organisational success from system reliability and efficiency, while 30 more respondents somewhat agree. Twenty-five (38.46%) of the respondents to the question strongly agree that where they have used data analytics, it has enhanced standardisation in their respective organisations' processes. Twenty (30.76%) of the respondents to the question agree, while 38.46% (25) somewhat agree. The results show that data analytics does contribute towards system reliability and efficiency and that it enhanced standardisation in organisational processes.

In relation to stakeholders' participation and its contribution towards projects success, 40% of the respondents agree, while 25% respondents strongly agree that stakeholders' participation led to their projects' success. The result does show that stakeholder participation leads to projects success as confirmed by the majority (65%) of the respondents.

4.2.3 Initial Phase

The purpose of this variable question was to establish if the non-use of data analytics in software projects results in ad hoc projects processes and activities; whether this impacted on measuring projects milestones, and how it affects repeating successful past projects processes.

Thirty percent of the respondents disagreed that the initial phase projects they worked on, which had no defined processes and standards in place, had ad hoc processes and activities, while an equal percentage (30%) of the respondents somewhat agree. Twenty percent of the participants agree that the initial phase projects they worked on had ad hoc processes and activities, another 15% of the respondents strongly agreed, while five percent, 5% strongly disagreed. This indicates, as shown by the 35% of the respondents, that lack of defined standards and processes in the projects, result in ad hoc processes and activities in projects.

Fifty five percent of the respondents disagreed that in their initial phase projects, the absence of data analytics had affected the repeating of their successful projects' processes. Only 20% of the respondents agree that the non-use of data analytics in their initial phase projects had affected the repeating of their past successful projects' processes. A total of 15% of the respondents agree that their projects' milestones were not easily measurable when data analytics was not used, while 55% of the respondents disagree that the milestones were not easily measurable when data analytics was not used in their projects' processes. The results show that to some extent measuring milestones is improved by data analytics use as supported by 15% of the respondents, however the majority of the respondents (55%) indicate that data analytics use does not affect measuring of projects milestones.

On the other hand, 65% of the respondents specified that at least one, but not more than five of their initial phase projects where data analytics was not used had failed due to poorly controlled processes. It therefore suggests that data analytics brings processes control, which contributes towards projects success.

4.2.4 Preparing for the Future

The purpose of the questions for this variable were intended to establish if companies facilitate data analytics training for software project managers and how data analytics contributes towards future project management. The surveyed project managers have used at least one development methodology, with 18.5% of these using the traditional development method, and a further 18.5% using the Agile Development Method, while

62.5% of the respondents used both the Traditional and Agile software development methodologies.

Seventy two percent (65) of the respondents indicated that the organisations they worked for do not facilitate or offer data analytics training. This indicates that many companies from the surveyed population are not investing in equipping the software project managers for the future possibility of project managers incorporating data analytics into project management. There are, however, some organisations that in some ways are facilitating for data analytics training, as 27.78% (25) of the respondents somewhat agree that the organisations they work for facilitated for their training. These results are an indication that more needs to be done by organisations in training software project managers in the use of data analytics for the improvement of software project management.

4.2.5 Optimising Level

The purpose of the questions for this variable were meant to establish if using data analytics assisted in achieving software design requirements; if using data analytics improved handling projects processes variations, and if project managers chose to use other automated data analysis tools besides data analytics.

While establishing *if software project managers used any other automated data analysis tools besides data analytics* in their last three projects, 82.35% (70) of the respondents said that they had not used any other automated data analysis tools besides data analytics, while 5.88% had used one form of automated data analysis tools in their projects, and 11.76% have used two forms of automated data analysis tools in their projects. This suggests there is a preference by some of the project managers to use other automated data analysis tools besides data analytics.

On the aspect of determining *if using data analytics assists in realising software design requirements*, 23.08% of the respondents disagree, while 38.46% of the respondents somewhat agree, 38.46% agree whereas 15.38% of the respondents strongly agree. These results show that data analytics use does assist in realising software design requirements as a total of 53.84% of the respondents indicated that it does. The respondents have further indicated that data analytics use improved handling projects processes variations, as 71.43% of the respondents agree, while 28.57% respondents disagree.

4.2.6 Project Success

The purpose of the questions for this variable were intended to determine the effect of non-use on data analytics on the projects' budgeted costs, the projects' planned time, user requirements analysis and project scope.

According to 50% of the respondents, the assertion that, based on their experience, the *non-use of data analytics does not lead to the failure to complete projects within the budgeted cost* is true. On the contrary, 20% of the respondents agree that the non-use of data analytics in their projects led to the failure to complete projects within the planned cost budgets, which is in addition to 30% of the respondents who strongly agree that not using data analytics in their projects resulted in the non-completion of projects within the planned cost budgets. In response to the question, if the *non-use of data analytics affected delivery of software projects within estimated time*, 36.84% of the respondents agree that this was so, while 26.32% somewhat agree, as opposed to 31.58% who disagree. The results show that not using data analytics does affect completion of projects within budgeted cost as indicated by 50% of the respondents who agree. The results further show that non-use of data analytics to some extent affect the delivery of software projects within estimated time as 36.84% of respondents agree that it does.

Not all of the participants responded to the question *whether the projects not being completed in time or at budgeted cost was as a result of ineffective user requirements analysis*. Almost thirty two percent (31.57%) of the participants agree that the non-completion of projects in time and on budget resulted from ineffective user requirements analysis, with a further 47.37% somewhat agreeing, suggesting that there are some of their projects where it was true, yet in some projects this was not the case. However, 21.05% of the respondents disagree that ineffective user requirements analysis had mainly led to overtime and over-budget regarding the completion of some of their projects. The results show that to some extent ineffective user requirements analysis does affect completion of projects within planned time and budgeted cost as indicated by 31.57% who agree, with a further 47.37% somewhat agreeing despite a considerable 21.05% indicating that it does not.

Failure to meet project scope as a result of scope creep is one of the challenges that project managers face, with 25% of the respondents in this survey agreeing that such is common where data analytics might not have been used. Twenty percent of the respondents somewhat agree, while 50% of the respondents disagree that in instances where no data analytics had been used, failing to meet projects scope was not common. While many (50%) of the respondents disagree that not using data analytics results in scope creep and projects failure, 25% of the participants stated that there is a portion of such related project failures. The results show that using data analytics does not entirely reduce scope creep in project management even though to some extent it does as shown by the 25% of the respondents who agree it does.

Fifty percent of the respondents have also indicated that more than 50% of their initial phase projects' success had depended on individual talent. Thirty five percent of the respondents have specified that less than 25% of their initial phase projects success had depended on individual talent. The results therefore show that many of the initial phase projects' success where data analytics was not used depended on individual talent.

4.2.7 Defined Level

The purpose of the questions for this variable was to establish if the use of data analytics in software projects leads to consistency in the completion of software projects as had originally been planned, if it enhances management of projects' interrelated processes and if the project managers comply with any software quality standards.

Thirty-eight percent (38.46%) of the participants who responded to the question, *if the use of data analytics in software projects leads to consistency in completion of software projects as planned*, do agree, with 15.38% of them strongly agreeing and a further 38.46% somewhat agreeing. Despite a high 38.46% of respondents agreeing and a further 38.46% somewhat agreeing, twenty three percent (23.08%) of the respondents disagree that using data analytics in projects leads to consistency in the completion of software projects. The results as indicated by the majority of the respondents (76.92%) show that data analytics use does to a great extent improve consistency in the completion of software projects as planned. The results also show that there are situations where it has not resulted in the consistency in the completion of software projects as planned as indicated by the 23% of the respondents.

Seventy five percent of the respondents to the question *whether data analytics use in software projects enhance management of interrelated processes or not* agree that it does, with a further 33.33% of the respondents somewhat agreeing, and only 6.67% of the respondents to the question disagreeing. The result as indicated by 75% of the respondents show that management of interrelated processes in software projects can be improved by data analytics use. However, the 6.67% of the respondents who do not agree is an indication that merely using data analytics does not in itself guarantee improvement of management of interrelated processes as that percentage show there are cases where it has not been the case.

However, 63.16% of software project managers surveyed do not comply with any software quality standards, while 36.84% of the software project managers comply with some software quality frameworks or standards. The results show that there is need for more software project managers to be complying with software quality standards for software project management quality to improve as only 36.84% of the surveyed project managers comply with some quality standards.

4.2.8 Managed Level

The objective of the questions for this variable was to establish *if data analytics use in software projects improved projects' processes and helped achieve set projects goals*.

Forty three percent (43.75%) of the respondents agree, with 25% of them strongly agreeing that data analytics use in their past software projects had enhanced the projects' processes which had helped to achieve set project goals. A further 25% also somewhat agree that data analytics improved projects' processes, thereby assisting in realising the set project goals. However, 31.25% of the respondents disagree that their past projects processes were enhanced by data analytics use, and that it did not assist in achieving the set project goals. The result show that though a considerable percentage (31.25%) of the respondents disagrees, data analytics use does improve projects process and helps achieve set project goals as indicated by 43.75% of the respondents.

Forty-one percent (41.17%) of the respondents to the question agree and a further 35.29% somewhat agree that in their past projects, projects monitoring, and control processes were more attainable by using data analytics. However, 23.53% of the respondents do not agree.

Thirty-six percent (36.36%) of the 55 respondents indicated that none of their projects' managed level processes were further enhanced by data analytics use, while 27.27% of the respondents said that Measurement and analysis had been improved; of these, 18.18% declared that Project planning had been improved, while, 9.09% had confirmed that Configuration Management and Processes and Product quality assurance were further enhanced by data analytics use. The results show that using data analytics does improve projects monitoring and control processes as indicated by 41.17% of the respondents who agree. The results further show that Measurement and analysis is also improved by data analytics use as indicated by 27.27% of the respondents.

4.2.9 Impact on Customer

The purpose of the questions for this variable was to assess if data analytics use in the respondents' projects influenced realising user functionality requirements and if this had improved the chances of realising software performance goals.

Eighty percent (80%) of the surveyed participants responded to the question; *if data analytics use in their projects influenced realising user functionality requirements*, of which 37.5% of the respondents agree that it does, with 18.75% of strongly agreeing. A further 37.5% somewhat agree, while 25% of the respondents disagree.

Seventy-five percent (75%) of the participants responded to the question; *if data analytics use in their projects enhanced chances of realising software performance goals*. Sixty percent of the respondents agree that data analytics use in their software projects enhanced chances of realising software performance goals, with 33.33% of the participants strongly agreeing and a further 33.33% somewhat agreeing. Only 6.67% of the respondents disagree, though, that data analytics use increased the chances of realising software performance goals. The results show that data analytics use improves chances of realising user functionality requirements as indicated by 37.5% of the respondents. The results also show that using data analytics in projects improves chances of realising software performance goals as indicated by 60% of the respondents.

The descriptive statistics and Cronbach's alpha coefficient for the questionnaire instrument is discussed as follows.

4.3 Reliability of Constructs

Table 4.4: Descriptive statistics and Cronbach's alpha coefficient for each variable

	Item	Mean	SD	Total Correlation	Cronbach Alpha / N Items
Initial Phase	IP_2	2.7000	0.818	0.793	0.877
	IP_3	2.5500	1.159	0.793	
Project Success	PS_1	8.8421	6.411	0.897	0.876
	PS_2	8.8421	6.198	0.646	
	PS_3	8.5263	7.273	0.735	
	PS_4	8.8421	6.305	0.720	
Managed Level	MI_1	3.4000	1.054	0.923	0.957
	MI_2	3.4667	1.333	0.923	
Impact on Customer	IC_1	3.7857	0.895	0.890	0.942
	IC_2	3.5000	0.978	0.890	
Defined Level	DI_2	3.6154	0.709	0.607	0.749
	DI_3	3.3077	0.998	0.607	
Organisation Success	OS_3	3.4615	0.877	0.567	0.721
	OS_4	3.5385	0.721	0.567	
Optimising Level	OL_2	3.2308	0.805	0.791	0.881
	OL_3	3.3077	0.998	0.791	
Project management Efficiency	PME_2	4.4286	0.546	0.600	0.750
	PME_3	4.2857	0.504	0.600	

When using Cronbach's alpha, a reliability coefficient of 0.70 or higher is considered acceptable (Tavakol & Dennik 2011:54). The alpha coefficients for all the variables in this study lie between 0.72 and 0.96, which means that the variables have high internal consistency and are acceptable, as they are above the recommended 0.70 score.

4.4 Validity

In order to establish if the questions contributed to their constructs in the questionnaire, factor analysis was carried out. Factor Analysis is a technique which is applied to a set of observed variables and seeks to find the underlying factors from which the observed variables were generated (DeCoster 1998:1). There are two forms of Factor analysis: Exploratory Factor Analysis (EFA), which tries to discover the nature of constructs influencing a set of responses; and Confirmatory Factor Analysis (CFA) which tests whether a specified set of constructs is influencing responses in a predicted way (DeCoster 1998:1). The study considered the Exploratory Factor Analysis as it is used to explore the possible underlying factor structure of a set of observed variables without imposing a preconceived structure on the outcome (Child, 1990). By performing EFA, the underlying factor structure is identified. EFA was considered given that this study makes use of eight variables being measured in relation to project management quality.

CFA is a theory oriented model which confirms the factor structure extracted by EFA and allows the testing of the hypothesis that a relationship between observed variables and their underlying constructs exists (Child 1990). CFA has not been considered as ideal as the researcher was not seeking to confirm or reject a measurement theory.

4.4.1 The extraction and rotation method

The Principal Component Analysis (PCA) was used for extraction and rotation before factors calculation. Principal Component Analysis is a data-reduction technique which reduces many variables into a smaller number of components yet still containing most of the information from the larger variables set (Yong & Pearce 2013:84). Reducing the data sets makes the analysis of data much faster and easier as there will be lesser dimensions of data to handle than in the original data set. The conversion from the larger data set of possibly correlated variables results into a set of values of linearly uncorrelated variables referred to as Principal Components (Tefas & Pitas 2016). When conducting PCA, Eigenvector referred to as Characteristic Vector is a nonzero vector that changes by a factor when the linear transformation is applied to it, while the corresponding factor by which the Eigenvector is scaled is called the Eigenvalue (Namboodiri 2011).

Rotation is an integral part of PCA which is used to get each variable load on as few factors as possible, allowing each factor to define a specific collection of interrelated variables for easier interpretation (Yong & Pearce 2013:84). Rotation takes two forms, Orthogonal rotation, where it is assumed that the factors are uncorrelated; and Oblique rotation, where the factors are considered as being correlated (Yong & Pearce 2013:84). It is preferable to use the Oblique rotation when exploring the correlation between components. Tables 4.6 and 4.7 show the principal component analysis.

4.4.2 Bartlett's test for Sphericity and the KMO

Bartlett's test for Sphericity was used to check if there is redundancy between the variables that can be summarised (Stephanie 2014) The test was performed before using the Principal component analysis to verify if the data reduction technique actually compresses the data

Table 4.5 illustrates the results of Bartlett's test for Sphericity and the KMO value. The KMO of 0.6 and the magnitude ($p<0.05$) of Bartlett's test indicates that the correlation structure is significantly strong for performing a factor analysis of the items.

4.4.3 Communalities

The communalities show the extent to which an individual item correlates with the other variables (StatWiki 2019). A value close to 1 indicates an item that correlates highly with the rest of the variables. Variables with low communalities (near 0.2) should be reconsidered. For the 19 variables, communalities are reasonable, as their extraction ranges from 0.71 to 0.93 (see Table 4.5). All the variables use correlate with the other variables as all their extraction values are much closer to 1. There are no variables that were reconsidered as no variables with Extraction values close to 0.2 as shown in Table 4.5

Table 4.5: Communalities

Question	Extraction	Question	Extraction	Question	Extraction
IP_2	0,737	IC_1	0,884	PME_2	0,926
IP_3	0,812	IC_2	0,750	PME_3	0,930
PS_1	0,887	DI_2	0,872	OL_1	0,753
PS_2	0,838	DI_3	0,830	OL_2	0,831
PS_3	0,717	OS_3	0,849	OL_3	0,806
MI_1	0,934	OS_4	0,923	PF_1	0,912
MI_2	0,888	QML_2	0,864		

4.4.4 Rotated component matrix and factor loading

The rotated component matrix converged after five iterations and each construct excluding IP_1, DI_1, OS_1, OS_2, OML_1 and PME_1 fits well. Since the researcher selected the principal component analysis as the method of extraction, the "Total", "%

of Variance", and "Cumulative %" columns are identical to those of the first two components in the "Extraction Sums of squared Loadings" and "Rotation Sums of Squared Loadings", see Tables 4.6 and 4.7. Table 4.6 shows the importance of each of the 19 components with only the first 5 having eigenvalues shown in the Total Colum of over 1.00, and these combined explain 84,7% of the total variability in the data (See Table 4.7). Table 4.6 shows a significant decline in variability after Component 5.

Table 4.6: Extraction Method - Principal Component Analysis

	Initial Eigenvalues		
Component	**Total**	**% of Variance**	**Cumulative %**
1	9,696	48,478	48,478
2	2,641	13,206	61,685
3	2,175	10,876	72,561
4	1,326	6,630	79,190
5	1,105	5,523	84,713
6	0,634	3,168	87,881
7	0,614	3,069	90,950
8	0,468	2,341	93,291
9	0,316	1,581	94,872
10	0,284	1,419	96,291
11	0,216	1,081	97,372
12	0,164	0,819	98,191
13	0,155	0,774	98,965
14	0,091	0,453	99,419
15	0,056	0,282	99,701
16	0,034	0,171	99,872
17	0,016	0,079	99,950
18	0,009	0,046	99,996
19	0,001	0,004	100,000

Table 4.7: Extraction Method - Principal Component Analysis

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	9,696	48,478	48,478	9,696	48,478	48,478
2	2,641	13,206	61,685	2,641	13,206	61,685
3	2,175	10,876	72,561	2,175	10,876	72,561
4	1,326	6,630	79,190	1,326	6,630	79,190
5	1,105	5,523	84,713	1,105	5,523	84,713

4.5 Correlation Analysis

This section focuses on the relationship between the decision variables, such as the *Initial Phase, Project Success, Managed Level, Impact on Customer, Defined Level, Organisation Success, Quantitatively Managed Level, Optimising Level, Preparing for the Future and Project Management Efficiency.*

The relationship between the constructs of this study, expressed by means of Pearson Correlations, are reported in Table 4.8.

4.5.1 Decision Variables' Correlations

Correlation refers to the relationship or association between two or more quantitative variables and is based on an assumption of a straight line (Gogtay & Thatte 2017:78). Correlation has been used as it allows measuring the extent of an association between variables. Correlation Analysis uses a correlation coefficient with values ranging from -1 to +1. A +1 denotes a perfect positive relation between two variables, where a 0 shows no linear relationship between the variables, and a -1 shows that variables are perfectly related in a negative way (Gogtay & Thatte 2017:78). Positive relationship between variables means two variables increase or decrease at the same time (McLeod 2018:1), implying high values on one variable are associated with high values on the other and low values on one variable are associated with low values on the other. A negative (inverse) relationship means high values on one variable is associated with a decrease in the other variable (McLeod 2018:1). Therefore, a perfect negative relationship is when the relationship between two variables is negative all the time (Picardo 2019:1) while a perfect positive relationship exists when variables' percentages move together at the same percentage and in the same direction all the time (Investopedia 2018:1)

The variables' relationships will be discussed with reference to Pearson's correlation and linear regression as was reported in Table 4.8. The variable *Initial Phase* correlates with the *Project and Organisation Success* variable with a score of 0.773, as shown in Table 4.8. Since the score is less than 1, it shows that the variables are positively linearly related. *Managed Level and Impact on Customer* variables are positively linearly related as they have a correlation of 0.717 and p value <0.05. *Defined Level and Organisation Success* variables have a correlation of 0.918, showing that the variables are almost perfectly positively related, as they have a score much closer to 1 and p value < 0.05.

Quantitatively Managed and Project management Efficiency variables have a correlation score of 0.179, which shows a low linear positive relation, as the score is not too far off from 0 and p value < 0.05. *Optimising Level* and Preparing for the future variables have a correlation score of -0.273, showing that the two variables are negatively linearly related and p value < 0.05.

Table 4.8: Correlation Matrix of the variables

	IP		PS		MI		IC		DI		OS		QML		OL		PF		PME	
	Correlation	Sig.	Correlation	Sig.	Correlation	Sig.	Correlation	Sig.	Correlation	Sig.	Correlation	Sig	Correlation	Sig	Correlation	Sig	Correlation	Sig	Correlation	Sig
IP	1		.773	0.00	.821	0.00	.641	0.00	.592	0.00	.273	0.014	.329	0.001	.364	0.00	.244	0.021	-0.114	0.484
PS	.773	0.000	1		.819	0.000	.628	0.000	.654	0.000	.389	0.00	.227	0.023	.447	0.000	-0.119	0.262	-0.181	0.264
MI	.821	0.000	.819	0.000	1		.717	0.000	.651	0.000	0.122	0.279	.588	0.000	.681	0.000	0.156	0.168	.365	0.021
IC	.641	0.000	.628	0.000	.717	0.000	1		.800	0.000	.328	0.003	.396	0.000	0.190	0.092	0.041	0.726	-.401	0.017
DI	.592	0.000	.654	0.000	.651	0.000	.800	0.000	1		.918	0.000	0.134	0.253	0.080	0.511	-0.016	0.902	-0.245	0.237
OS	.273	0.014	.389	0.000	0.122	0.279	.328	0.003	.918	0.000	1		-496	0.000	-0.060	0.612	-.273	0.022	-0.020	0.917
QML	.329	0.001	.227	0.023	.588	0.000	.396	0.000	0.134	0.253	-.496	0.000	1		.351	0.000	.286	0.006	0.179	0.270
OL	.364	0.000	.447	0.000	.681	0.000	0.190	0.092	0.080	0.511	-0.060	0.612	.351	0.000	1		.248	0.019	.680	0.000
PF	.244	0.021	-0.119	0.262	0.156	0.168	0.041	0.726	-0.016	0.902	-.273	0.022	.286	0.006	.248	0.019	1		-0.064	0.716
PME	-0.114	0.484	-0.181	0.264	.365	0.021	-.401	0.017	-0.245	0.237	-0.020	0.917	0.179	0.270	.680	0.000	-0.064	0.716	1	

4.5.2 Hypothesis testing

The Pearson correlation of continuous variables was considered besides the other correlations that exist for the ordinal variables. As illustrated in Table 4.8, the r values range from 0.014 to 0.918, which indicates a reasonably strong relationship between the variables, since the r value can range from -1 to 1. The relationship between the variables is stronger when the value of r is closer to 1. The p-values, smaller than 0.05, indicate a significant correlation between the decision variables.

Using Table 4.8, I now look at the following Hypotheses which address some of the questions posed in Chapter 1.

Alternate Hypothesis H1:	There is a significant correlation (association) between non-use of data analytics in the Initial *Phase projects* and *Project Success*.
Alternate Hypothesis H2:	There is a significant correlation (association) between data analytics use on processes monitoring and control on *Managed Level Projects* and *Impact on Customer*.
Alternate Hypothesis H3:	There is a significant correlation (association) between data analytics use on interrelated process management, completion of projects according to plan on *Defined Level projects,* and *Organisational financial Success*.
Alternate Hypothesis H4:	There is no significant correlation (association) between data analytics use on decision-making in *Quantitatively Managed Level* projects and *Project Management Efficiency*.
Alternate Hypothesis H5:	There is a significant correlation (association) between *Optimising Level* data analytics use and project managers' training and *Preparing for the Future.*

4.6 Analysis of variance (ANOVA) among constructs

An ANOVA was conducted on the constructs to establish the items homogeneity. The results are reported in Table 4.9.

Table 4.9: Test of homogeneity (ANOVA)

Constructs		Sum of squares	df	Mean square	F	Sig.
IP PS	Regression	41,866	1	41,866	145,258	.000[b]
	Residual	28,245	98	0,288		
	Total	70,111	99			
IC MI	Regression	49,490	1	49,490	87,986	.000[b]
	Residual	46,686	83	0,562		
	Total	96,176	84			
OS DI	Regression	42,535	1	42,535	365,580	.000[b]
	Residual	7,912	68	0,116		
	Total	50,446	69			
PME QML	Regression	0,474	1	0,474	1,253	.270[b]
	Residual	14,370	38	0,378		
	Total	14,844	39			
PF QL	Regression	3,068	1	3,068	5,753	.019[b]
	Residual	46,932	88	0,533		
	Total	50,000	89			

IP = Initial Phase
PS = Project Success
MI = Managed Level
IC = Impact on Customer
DI = Defined Level
OS = Organisation Success
QML = Quantitatively Managed Level
OL = Optimising Level
PF = Preparing for the Future
PME = Project Management Efficiency

Test for Homogeneity of Variances and Test of Equality

The assumption of homogeneity of variance needs to be tested when comparing three or more constructs with ANOVA. Homogeneity of Variances ensures the distributions of outcomes in each independent constructs are comparable or equal and to meet the assumption of homogeneity of variance, the *p-value* Levene's Test should be above .05 (Statistics Solutions 2019).

The p value of .270 as shown in Tables 4.10 and 4.11 for *Project Management Efficiency* and *Quantitatively Managed Level is* greater than 0.05, indicating that the variances do not differ significantly and that the assumption of equality of the variances is satisfactory. IP, PS, IC, MI, OS, DI, PF and OL all have sig values less than 0.05, indicating that the variances do differ.

Table 4.10: Test for Homogeneity of Variances and Test of Equality

Constructs		Statistic	df1	df2	Sig.
MI	IC	.717[a]	1	83	0,000
IP	PS	.773[a]	1	98	0,000
DI	OS	.918[a]	1	68	0,000
QML	PME	.179[a]	1	38	0,270
OL	PF	.248[a]	1	88	0,019

4.7 Regression Analysis

Regression analysis is a method of identifying the variables' impact on a topic. It is a statistical method that is used to determine the degree to which the independent variables influence the dependent variables (Foley 2018). Regression analysis has been used in this study as it indicates if the variables have a significant relationship with a dependent variable and can indicate the relative strength of the different independent variables' effect on a dependent variable (Sarstedt & Mooi 2014:194).

Regression uses a significance level of 0.05, with the statistical values greater than 0.05 being non-statistically significant (Frost 2019). Using Table 4.11, the variables' *Initial Phase* and *Project Success*; *Managed level* and *Impact on Customer*; *Defined*

Level and Organisation Success have a significant value of 0.000, showing that the variables are statistically significant with the *Organisation Level and Preparing for the Future* having a value of 0.019.

Quantitatively Managed Level and Project Management Efficiency have a value 0.270, which is not less than the usual 0.05 and, therefore not being significant. In regression, the dependant variable is the main factor trying to be predicted and the independent variable being hypothesised to have an impact on the dependent variable (Foley 2018).

Table 4.11: Regression Analysis

Variable	R	R square	Adjusted R square	Std. error of the estimate	R square change	F change	Sig. F change
a. Predictors: (Constant), IP							
b. Dependent Variable: PS	.773[a]	0,597	0,593	0,53686	0,597	145,258	0,000
a. Predictors: (Constant), MI							
b. Dependent Variable: IC	.717[a]	0,515	0,509	0,74999	0,515	87,986	0,000
a. Predictors: (Constant), DI							
b. Dependent Variable: OS	.918[a]	0,843	0,841	0,34110	0,843	365,580	0,000
a. Predictors: (Constant), QML							
b. Dependent Variable: PME	.179[a]	0,032	0,006	0,61494	0,032	1,253	0,270
a. Predictors: (Constant), OL							
b. Dependent Variable: PF	.248[a]	0,061	0,051	0,73028	0,061	5,753	0,019

4.8 Chapter Conclusion

This chapter explained the procedure that was followed in gathering data and the analyses of the data. The chapter explained the sampling procedure that was used, the descriptive statistics of the variables and frequency and how the reliability of the constructs and validity was measured. The chapter also explained the correlation analysis that was carried out, the hypothesis testing done and the analysis of the variance among the constructs. Regression analysis carried out to identify the variables impact on a topic was also explained.

Chapter 5: Discussion of the Results

5.1 Introduction

Chapter 5 discuses and summarises the findings of this study and attempts to answer the research questions presented in chapter one, based on the analysis of the results presented in Chapter 4. Guided by the research objectives and questions, the study made use of the questionnaire survey to obtain quantitative data from software project managers, which was used to establish if data analytics can be used in software project management to improve software project management quality. The survey allowed the researcher to establish the factors necessary for data analytics use in software project management.

5.2 Discussion of Findings

The data analysis section of the study has established some findings and provided some answers to the research objectives, questions and the hypotheses.

The study's investigation *into the challenges necessitating the use of data analytics to improve software project management quality*, revealed that, among other factors, ineffective data analysis and decision-making in project management where data analytics was not used compromised project management quality. Thirty percent (30.77%) of the respondents to this question indicated that data analytics use in their projects led to quicker and timely decision-making, indicating the gap that data analytics addresses with regards to decision-making in project management.

This aligns with the argument by Buse and Zimmerman (2010:1), that there is a substantial disconnect between the available information required by project managers to make good decisions and the information being provided by the existing analytic tools. The study's finding, that data analytics improves decision-making in project management also concurs with Singh (2015:3), who states that analytics can be used to analyse projects' multiple risk factors for effective decision-making and risk management.

The research further established that improved decision-making positively impacts project management quality, with 40% of the respondents based on their experience strongly agreeing that improved decision-making directly impacts software project management quality. The confirmation by 40% of the respondents shows that improving decision making though not entirely, still does improve software project management quality. This finding confirms the statement by Linders (2015:1) that quality in project management can also be driven by taking and communicating timely decisions.

The research further established that the projects processes control was not effective in projects were data analytics was not used with 41.17% of the respondents indicating that data analytics use led to greater projects processes control. Forty-One percent of the respondents is a considerably high percentage to acknowledge that data analytics does contribute towards software project management quality through improved projects processes control. This research finding corresponds with the argument by Singh (2015:3) that data analytics use improves the quality of projects' processes and the final project product by addressing the gap between the state of projects' processes and the desired processes' state.

Time and cost overruns are the other challenges which have been considered as necessitating the use of data analytics to improve software project management quality. The non-use of data analytics in software project management has been shown to have an effect on project time frames and costs, as indicated by 36.84% of the surveyed project managers, who specified that the non-use of data analytics affects the delivery of software projects within the planned time; while 50% of the respondents agree that this affects the delivery of software projects within the planned cost and makes managing scope creep difficult. While the result i.e. 50% of respondents affirms that using data analytics improves managing scope creep, it also show that there are cases where it has not been the case as not all the respondents agreed it does. The result, 36.84% of respondents also shows that though data analytics use improves projects time frames and costs management, it will not be the sole panacea to time and cost overruns as the remainder of the respondent to the question did not agree that data analytics use improves projects time frames and costs management.

The study findings on time and cost overruns affirm what Dhollander (2017:1) indicated, that data analytics use brings instant value through the saving of precious time and unlocking previously hidden opportunities for improvement. This was also echoed by Daddikar (2018:1), who stated that using data analytics helps project managers to forecast and manage tasks within the set time lines and budgeted costs, enabling the taking of proactive actions where early signs of slippage would have been detected.

However, the study also established from 50% of the respondents based on their experience that the non-use of data analytics in software projects does not lead to failure to complete projects within budget. This, however, can be due to the 50% of the respondents not having used data analytics in their projects.

The study survey established *the current level of data analytics knowledge in software project managers.* It has been revealed through the software project managers who participated in the survey that there is some level of data analytics knowledge among the software project managers, as 50% of the managers confirmed having used data analytics in at least one of their projects, as shown in Figure 4.2. This supports the argument by Jiwat (2017:2) that data analytics influence also stretches to project management, as it can bring improved decision-making changes. However, 15% of the surveyed project managers have no knowledge of data analytics, with 35% of the managers having reading knowledge only and having not used data analytics in any project. Though the 50% of respondents who have used data analytics show the existence of some data analytics knowledge among software project managers, the population using data analytics needs to increase drastically.

Regarding *ascertaining basic factors required to use data analytics in software project management to improve software project management quality*, the study also made some findings. The study has established through the survey that some of the basic factors necessary for data analytics to be successfully used in software project management include the need to have stakeholders' participation in project management. It is critical to have buy-in and participation from all projects' stakeholders, as their support with regards to adoption and the financing of data analytics contributes towards projects' success. This finding is in line with the argument by PMP (2018) that stakeholder support is required for a project's success. However, their varying needs equally need to be managed, lest these jeopardise the project. Such crucial stakeholders, including the project managers should be willing to adopt new technology in project management.

Likewise, organisations' support and willingness to invest in the adoption of data analytics is critical for the projects' success. The study's findings on the participants' companies offering or facilitating for data analytics training is concerning, as 72% of the participants' companies do not offer data analytics training; nor do the companies facilitate for the training. The result (72%) shows that many of the organisations are not yet making use of data analytics in their software project management probably due to financial constrains or some are just being conservative.

Another factor which emerged from the study is for software project managers to be trained in data analytics use to be able to effectively use leverage on data analytics. The training would not only bolster the adoption of data analytics' use, but also assist in improving on the interpretation of data from data analytics tools for effective software projects management. However, although some of the companies are facilitating for the adoption through the training as 27.78% of the respondents agree that the organisations they work for do facilitate for their training, more needs to be done to increase data analytics use. Twenty seven (27.78%) percent of the respondents' organisations is a low percentage to realise a significant effect on software project management quality through data analytics use and more organisations need to facilitate for data analytics use and incentivise their software project managers.

It has also emerged from the study that data analytics use becomes effective with the use of automated data analytic notification tools which send notifications on any variances to project managers. Considering that about 17.64% of the respondents preferred to use other automated data analytic tools besides data analytics, data analytics training may improve the project managers' willingness to adopt and use data analytics.

In as much as having data analytics tools and knowledge in place is good, collecting, integrating and preparing data must be done meticulously for the analysis to be effective. The respondents have indicated that processes followed in collecting and the analysis of data should be set out by means of clearly-defined policies and procedures.

The objective *to examine whether data analytics use in software project management improves software project management quality dimensions* has been addressed by the study findings. In relation to the performance quality dimension, the survey established through 43.75% of the respondents that data analytics brings an improvement to the projects' processes, which helps to attain set project goals. This corresponds with what Delen and Demirkan (2013:361) established, that data analytics allows the optimisation of processes, thereby enhancing performance. Likewise, Singh (2015:4) also established that data analytics brings greater projects control as analytics can be used during the process to break down a project's processes and systems to predict their behaviour and outcomes. The 43.75% result may imply that the other respondents may have used other tools besides data analytics to manage projects processes.

The improvements brought by data analytics complement the quality measures that standards including PMBOK may have introduced. This is confirmed by the survey findings that show that data analytics improve quality dimensions. Forty percent of the respondents on the quality dimensions improvement enquiry declared that performance had improved, 20% said that consistency improved and 25% revealed that the serviceability of the project product had improved. Despite the 40% confirming that data analytics does significantly complement other project management standards' quality measures which is a much needed contribution, it is concerning that the remaining greater percentage did not realise the quality dimensions improvement. It is further concerning that only 20% of the respondents could confirm that data analytics did improve consistency in projects management but the inconsistency maybe as a result of the project managers not having consistently used data analytics in all their projects. Jiwat (2017:1) concurs that data analytics can be used to develop quality standards, quality control procedures and monitor quality during a project's execution. The findings also show that the completeness and performance of the projects' output were also improved by using data analytics.

The study further made findings on the hypothesis, as illustrated in Table 4.8. The research established that a reasonably strong relationship exists between the variables, as the r values range from 0.014 to 0.918. Using the Pearson correlation of continuous variables and the feedback from the respondents, the following was arrived at on the Hypotheses;

H1: *Data analytics use in software project management is positively related to projects success.*

The results show that there is a significant correlation between data analytics use in Software project management and Projects Success. The hypothesis was accepted based on the results.

H2: *Data analytics use in software project management is directly related to project impact on customer.*

The research survey data analysis conducted shows that there is a significant association between data analytics use in software project management and project impact on customer. Based on the findings, the Hypothesis was accepted.

H3: *Data analytics use in software project management is directly related to organisational success.*

The results reveal that data analytics use on software project management has a direct relationship with organisational success.

H4: *Data analytics use in project management is directly related to improved project management efficiency.*

The results show that there is no significant association between data analytics use on project management's decision-making and improved project management efficiency.

H5: *Data analytics use with software projects management models and standards prepares organisations for future project management.*

The results reveal that there is a significant relationship between data analytic use in projects management and the preparation of organisations for future project management.

5.3 Limitations

Due to limited financial resources, the study was mainly restricted to Gauteng province in South Africa. The research was conducted in South Africa, with the majority of the respondents being from the Gauteng province, hence the findings are particular to project management in the South African context and may not be generalised across the world.

The questionnaire used in the survey was designed with the intention of the respondents to answer some of the questions, using their practical experience from data analytics use in their project management processes. However, a possibility that some may have answered using their theoretical knowledge of data analytics may not be ruled out.

Despite these limitations, the findings of the research are still valid and important for both software project managers for their application, and for academics for further research.

5.4 Chapter Conclusion

This chapter discussed the findings of the research in relation to the research objectives and the question set in Chapter 1. The findings on the hypothesis are also presented in this chapter, following the testing conducted and as explained in chapter 4. The limitations of this study were also discussed and presented in this chapter. The conclusion to this study and recommendations for future studies are presented and discussed in Chapter 6.

Chapter 6: Summary and Conclusions

6.1 Introduction

The aim of the study was to establish the prospect of using data analytics to improve software project management quality through enhanced decision-making and project management standards. To achieve this objective, the study used a questionnaire survey involving software project managers as the respondents. The research had four objectives as listed in Section 1.6 to realise the research aim. Section 6.2 summarises the study findings.

6.2 Summary

The study used a survey methodology were 100 completed questionnaires were analysed from the returned 115 questionnaires. Eleven questionnaires were incomplete and four did not have consent forms, so these could not be used.

The findings from the study show that ineffective data analysis and decision-making in software project management are among the major contributory factors to poor software project management quality, aligning with the observation of Elgendy and Elragal (2016:1083) that effective data analysis reveals hidden insights in data and enhances decision-making.

Time, cost overruns and projects' processes control were also identified by the study as being the other contributory factors which necessitate the use of data analytics in software project management for their betterment. The research showed that software project management quality can be improved by using data analytics to enhance decision-making, project management processes control and enable the completion of software projects' tasks within planned times and budget. Furthermore, the study revealed that the improvement of decision-making, processes control, time and costs management complement the project management standards' quality control measures.

The study revealed that there are few software project managers with data analytics knowledge, as 50% of the surveyed project managers do not have any knowledge of data analytics; this compounded with the insufficient data analytics training facilitated by organisations among other challenges, is hampering the speedy adoption and effective use of data analytics in software project management. Many organisations are still undecided on adopting data analytics due to a lack of understanding and experience (Kwon, Lee & Shin 2014:386).

Stakeholder participation has been found to be one of the key factors necessary for the adoption and effective use of data analytics in software project management. Kibera (2013:124) asserts that stakeholder participation raises support and ensures successful implementation and adoption. The study reveals that increased organisations' support in training of Software project managers can lead to an accelerated adoption and use of data analytics, thus resulting in the improvement of software project management quality. The other factors that emerged from the study include the use of automated data analytic notification tools to alert on variances and deviations and establishing clearly defined processes and procedures for data gathering and analysis in organisations.

The study has also established that data analytics does improve software project management quality dimensions. The study revealed that the main quality dimensions which data analytics improves include the performance quality dimension where, notably, projects processes control is greatly enhanced, and the consistency, completeness and serviceability of the project's finished product is also improved. The significant relationship established by the study's hypotheses between data analytic use and the postulation that this prepares organisations for future project management and its direct relationship with decision-making and project management efficiency, clarifies that data analytics ought to be used to improve software project management quality.

6.3 Conclusion

The research questions used in the study were shown as discussed in Chapter 4 section 4.3 and 4.4, to be reliable and valuable for future research. All the variables used had high internal consistency coefficients between 0.72 and 0.96 which is above the recommended 0.70 score. The study established that there is a direct relationship between effective data analysis and correct decision making as up to 74% of the delayed or incorrect decisions in the participants' last projects had been due to ineffective data analysis. The marked improvement in consistency as indicated by 20% of the participants in their projects is evident in the consistency in timely and correct decision making. The confirmation by 69.22% of the respondents that data analytics use in their projects improved standardisation of processes and in turn improved system efficiency and reliability further confirmed financial organisational success through data analytics use.

In terms of improving software project management quality dimension the study reveals the need for more project managers to comply with quality standards as 63.16% of software project managers surveyed do not comply with any quality standards, while 36.84% of the software project managers comply with some quality frameworks or standards. The study established that data analytics complement the quality measures that standards including PMBOK introduce.

The study further revealed that there is a direct correlation between improved processes control and project success as 65% of the participants experienced at least one project failure in their last five projects due to poorly controlled processes where data analytics was not used. It has also emerged that the majority (72%) of the surveyed project managers' companies need in terms of data analytics use, to do more towards future project management by facilitating for staff training in data analytics. This will affect the adoption and use of data analytics in software project management in future as only 27.78% of the surveyed participants' companies do offer or facilitate for training in data analytics.

Projects failure through inadequate user requirements analysis, scope creep and costs overruns are some of the challenges which data analytics has proved can mitigate. It has been established as confirmed by 50% of the respondents that improving user requirements analysis through data analytics use reduces projects failure from ballooned user requirements due to scope creep and cost overrun.

Organisational success due to system reliability and efficiency has also proved to be influenced by enabling factors for data analytics use. Stakeholder participation, software project managers’ ongoing training and Organisational support through training are some of the key factors identified as necessary for effective data analytics use.

6.4 Recommendations

The following recommendations are made based on the findings from this study.

It is recommended that Software project managers increase the rate of adopting and using data analytics in software project management, as 50% of the surveyed project managers indicated having no experience in data analytics, with 35% of the managers having read about data analytics only; and worse still 15% having no knowledge of data analytics at all. If these project managers adopt and use data analytics, this will improve software project management quality as 43.75% of the surveyed project managers confirmed that data analytics improved project processes’ control and assisted in achieving set project goals.

Companies should play a more active role in encouraging the adoption and use of data analytics, as the study revealed that 72% of the companies do not facilitate or offer data analytics-related training. Companies can incentivise employees by paying for their training expenses or by rewarding project managers who attend the training and adopt data analytics in their project management practices.

Software project management institutions and corporate sponsors should incentivise software project managers to adopt and use data analytics which would help to improve software project management by giving a discount on the members’ annual subscription fees to members who go for and complete training in data analytics.

6.5 Future Research

In light of the findings of this study, the following recommendations are made for future research:

This study provides researchers with important information for further studies to be conducted to establish how data analytics can also be used to improve software development methodologies. The feedback from 62.5% of the participants who have used both the traditional and agile software development methodologies, together with data analytics, provides the valuable basis for further research.

Future studies could be done to assess the impact of data analytics use on individual quality dimensions per study to devote more time and resources and have an in-depth understanding of the individual dimensions. The study revealed that Performance, Accuracy, Completeness and Consistency improved by the use of data analytics, but further research should be conducted for in-depth analysis of how these dimensions are individually improved and how the rest of the quality dimensions are also impacted by data analytics use.

This study was conducted in the South African context; hence a comparative study could be conducted in other technologically advanced countries to establish if the study conducted in other well-resourced environments would reach different findings and conclusions, especially in relation to the adoption and use of data analytics. It could be that project managers in other countries that are technologically ahead of South Africa may have embraced data analytics better and could lead to a different outcome of the same study.

The study has revealed that only 20% of the surveyed software project managers have used data analytics in at least two projects, with 30% more having used data analytics in only one of their projects. Future research should therefore be conducted to establish the preparedness of software project managers in South Africa to adopt and use new data analysis technology including data analytics and the impact of their state of preparedness on software project management quality improvement.

References

Agarwal, N. & Rathod, U. 2006, ‘Defining “success” for software projects: An exploratory revelation’, *International Journal of Project Management*, vol. 24, no. 4, pp. 358–70.

AHIMA 2009, ‘Practice brief: Data quality management model’, *Journal of AHIMA*, vol. 69, no. 6, pp. 2–7.

Ahmad, Z.A. & Yekta, Z.A. 2010, ‘Relationship between perceived organizational support, leadership behavior, and job satisfaction: An empirical study in Iran’, *Intangible Capital*, vol. 6, no. 2, pp. 162–84.

Al-Kilidar, H., Cox, K. & Kitchenham, B. 2005, ‘The use and usefulness of the ISO/IEC 9126 quality standard’, *2005 International Symposium on Empirical Software Engineering, ISESE 2005*, pp. 126–32.

Aliyu, A.A., Bello, M.U., Kasim, R. & Martin, D. 2014, *Positivist and Non-Positivist Paradigm in Social Science Research : Conflicting Paradigms or Perfect Partners ?*, vol. 4, no. 3, pp. 79–95.

Alshazly, A.A., Elfatatry, A.M. & Abougabal, M.S. 2014, ‘Detecting defects in software requirements specification’, *Alexandria Engineering Journal*, vol. 53, no. 3, pp. 513–27.

Anda, B. & Sjøberg, D.I.K. 2002, ‘Towards an inspection technique for use case models’, *Proceedings of the 14th international conference on Software engineering and knowledge engineering SEKE 02*, no. 1325, p. 127.

Andersen, E.S. & Jessen, S.A. 2003, ‘Project maturity in organisations’, *International Journal of Project Management*, vol. 21, no. 6, pp. 457–61.

Anuar, U. & Ahmad, S. 2015, ‘A Simplified Systematic Literature Review : Improving Software Requirements Specification Quality with Boilerplates’, *2015 9th Malaysian Software Engineering Conference (MySEC)*, pp. 99–105.

Atkinson, R. 1999, ‘Project management: cost time and quality two best guesses and a phenomenon, it’s time to accept other success criteria’, *International Journal of Project Management*, vol. 17, no. 6, pp. 337–42.

Bach, J. 1994, *The Immaturity of CMM*, *American Programmer*.

Bahari, S.F. 2010, *Qualitative versus quantitative research strategies: contrasting epistemological and ontological assumptions*, pp. 17–28.

Barber, E. & Warn, J. 2005, ‘Management Decision"Leadership in projects: what we know from the literature and new insights", Team Performance Management: An Leadership in project management: from firefighter to firelighter’, *Management Decision Industrial and Commercial Training Iss International Journal Iss International Journal of Productivity and Performance Management*, vol. 43, no. 2, pp. 1032–9, viewed 24 May 2016, <http://dx.doi.org/10.1108/00251740510610026>.

Bhattacharya, S., Porter, M., Harrild, K., Naji, A., Mollison, J., Van Teijlingen, E., Campbell, D.M., Hall, M.H. & Templeton, A. 2006, ‘Absence of conception after caesarean section: Voluntary or involuntary?’, *BJOG: An International Journal of*

Obstetrics and Gynaecology, vol. 113, no. 3, pp. 268–75.

Bose, R. 2009, 'Advanced analytics: opportunities and challenges', *Industrial Management & Data Systems*, vol. 109, no. 2, pp. 155–72.

Bourque, P., Dupuis, R. & Abran, A. 1999, 'The guide to the software engineering body of knowledge', *Software, IEEE*, vol. 16, no. 6, pp. 35–44.

Bronnert, J., Clark, J.S., Cassidy, B.S., Fenton, S., Hyde, L., Kallem, C. & Watzlaf, V. 2012, 'Data quality management model (updated).', *Journal of AHIMA / American Health Information Management Association*, vol. 83, no. 7, pp. 62–71.

Brown, T. 2019, *Data Analytics Vs Predictive Analytics*.

Bruce, K. 2014, *Field Guide for Data Quality Management Field Guide for Data Quality Management*, pp. 1–57.

Bryman, A. & Liao, T.F. 2004, *The SAGE Encyclopedia of Social Science Research Methods*.

Buse, R.P.L. & Zimmermann, T. 2012, 'Information needs for software development analytics', *Proceedings - International Conference on Software Engineering*, pp. 987–96.

Buse, R.R.P.L.L. & Zimmermann, T. 2010, 'Analytics for software development', *Proceedings of the FSE/SDP Workshop on the Future of Software Engineering Research, FoSER 2010*, pp. 77–80.

Butt, F.S., Shaukat, S., Nisar, M.W. & Munir, E.U. 2013, *Software Quality Assurance in Software Projects : A Study of Pakistan*, vol. 5, no. 18, pp. 4568–75.

de Carvalho, M.M., Patah, L.A. & de Souza Bido, D. 2015, 'Project management and its effects on project success: Cross-country and cross-industry comparisons', *International Journal of Project Management*, vol. 33, no. 7, pp. 1509–22.

Chaudhuri, S. 2012, 'What next?: a half-dozen data management research goals for big data and the cloud', *Proceedings of the 31st ACM SIGMOD-SIGACT-SIGAI*, viewed 22 May 2016, <http://dl.acm.org/citation.cfm?id=2213558>.

Chawda, R.K. & Thakur, G. 2013, 'Big Data and Advanced Analytics Technologies and Use Cases', *BI Research*, pp. 1–8.

Child, D. (1990). The essentials of factor analysis (2nd ed.). Cassell Educational.

CMMI 2001, *Capability Maturity Model ® Integration (CMMI)*.

Connolly, P. 2011, 'Quantitative Data Analysis using SPSS', *An International for Health and Social Science*, pp. 1–283.

Crossan, F. 2003, 'Research philosophy: towards an understanding.', *Nurse researcher*, pp. 46–55.

Crossman, A. 2016, 'A Review of Software Tools for Quantitative Data Analysis', *About Education Inc.*

Crotty, M. 1998, 'Introduction: the research process', *The foundations of social research: Meaning and perspective in the research process*, p. 256.

Cuzzocrea, A., Song, I.-Y. & Davis, K.C. 2011, 'Analytics over large-scale multidimensional data: the big data revolution!', *… 14th international workshop on Data …*, pp. 101–4.

Delen, D. & Demirkan, H. 2013, 'Data, information and analytics as services', *Decision Support Systems*, vol. 55, no. 1, pp. 359–63.

Deloitte 2013, *The Analytics Advantage. We ' re just getting started Dear readers ,* p. 21.

Dorfman, M. 1997, 'Requirements Engineering', *Software Requirements Engineering*, no. Second Edition, pp. 7–22.

Dwivedi, S., Kasliwal, P. & Soni, S. 2016, '(RapidMiner , Weka , R tool , Knime)', *Comprehensive Study of Data Analytics Tools*, pp. 1–8.

Eriksson, J. & Brannemo, A. 2011, 'Coping With Deviation and Decision-Making', *Proceedings of the 18th International Conference on Engineering Design (iced 11): Impacting Society Through Engineering Design, Vol 1: Design Processes*, vol. 1, no. August, pp. 429–40.

Fauser, A.J., Schmidthuysen, M. & Scheffold, B. 2016, *The Prediction of Success Predictive Project Analytics Projekt- management T raining*, no. 5, pp. 66–74.

Field, A. 2009, *Discovering statistics using SPSS*, Third Edit., SAGE Publications.

George, S.Z., Coronado, R. a, Beneciuk, J.M., Valencia, C., Werneke, M.W. & Hart, D.L. 2011, 'Depressive symptoms, anatomical region, and clinical outcomes for patients seeking outpatient physical therapy for musculoskeletal pain.', *Physical therapy*, vol. 91, no. 3, pp. 358–72.

Ghosh, B.S., Forrest, D., Dinetta, T., Wolfe, B. & Lambert, D.C. 2012, 'Enhance PMBOK ® by Comparing it with P2M , ICB , PRINCE2 , APM and Scrum Project Management Standards', *PM World Today*, vol. XIV, no. I, pp. 1–77.

Glasow, P.A. 2005, *Fundamentals of Survey Research*.

Glowalla, P. & Sunyaev, A. 2014, 'Process-Driven Data Quality Management: A Critical Review on the Application of Process Modeling Languages', *Journal of Data and Information Quality*, vol. 5, no. 1–2, pp. 1–30.

Gmbh, B.D. 2013, 'Contents :The theory of inter-hybridity: socio-political dimensions and migration experiences of post- communist western Balkan states', *Journal of Advanced Research in Organizational Psychology*, vol. II, no. 2, pp. 4–54.

Goff, S. 2011, 'Project Management Success and Decision-Making Under Difficult Contexts', *AMIP, IPMA-Mexico, 2nd International de la Associ- acion Mexicano de Ingeniera de Proyectos Symposio*, pp. 1–8.

Gopalkrishnan, V. & Steier, D. 2012, 'Big data, big business: bridging the gap', *BigMine '12: Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining*, pp. 7–11.

Grobler, P.J. & Steyn, H. 2006, *Project Management Maturity Models: Does One Size Fit All?*, no. June, pp. 149–57.

Guillaume-Joseph, G. & Wasek, J.S. 2015a, 'Improving software project outcomes through predictive analytics: Part 1', *IEEE Engineering Management Review*, vol. 43, no. 3, pp. 26–38.

Guillaume-Joseph, G. & Wasek, J.S. 2015b, 'Improving software project outcomes through predictive analytics: Part 2', *IEEE Engineering Management Review*, vol. 43, no. 3, pp. 39–49.

Hansmann, T. & Niemeyer, P. 2014, 'Big data - Characterizing an emerging research field using topic models', *Proceedings - 2014 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IAT 2014*, vol. 1, pp. 5228–35.

Haug, A., Stentoft, J., Frederik, A., Jakob, Z., Haug, A., Arlbjørn, J.S. & Zachariassen, F. 2013, *Master data quality barriers : an empirical investigation*.

Herek, G.M. 2012, *A Brief Introduction to Sampling*, pp. 1–7.

Herodotou, H., Lim, H., Luo, G., Borisov, N. & Dong, L. 2011, 'Starfish : A Self-tuning System for Big Data Analytics', *Cidr*, vol. 11, pp. 261–72.

IEEE Computer Society 2004, *Guide to the software engineering body of knowledge*, Bob Werner (ed.),.

Ika, L.A. 2009, 'The art of managing relationships in interorganizational collaboration', *Project Success as a Topic in Project Management Journals*, vol. 40, no. 4, pp. 6–19.

Investopedia 2018, 'Positive Correlation Definition', *Investopedia*.

Issac, G., Rajendran, C. & Anantharaman, R.N. 2004, 'A conceptual framework for Total Quality Management in software organizations', *Total Quality Management & Business Excellence*, vol. 15, no. 3, pp. 307–44.

Jamsutkar, K., Patil, V. & Chawan, P.M. 2012, *Software Project Quality Management*, vol. 2, no. 3, pp. 686–90.

John Dudovskiy 2011, 'Data Collection Methods - Research-Methodology', *Research-Methodology*.

Joslin, R. & M??ller, R. 2015, 'Relationships between a project management methodology and project success in different project governance contexts', *International Journal of Project Management*, vol. 33, no. 6, pp. 1377–92.

Joyce, J. 2006, 'Software project management, simplified', *Scientific Computing*, vol. 23, no. 9.

Judgev, K. & Thomas, J. 2002a, *19 - Project management maturity models: The silver bullets of competitive advantage*, vol. 33, no. 4, pp. 4–14.

Judgev, K. & Thomas, J. 2002b, *19 - Project management maturity models: The silver bullets of competitive advantage*, vol. 33, no. 4, pp. 4–14.

Kaulgud, V. & Sharma, V.S. 2016, 'Software development analytics: Experiences and the way forward', *Proceedings - 2015 30th IEEE/ACM International Conference on Automated Software Engineering Workshops, ASEW 2015*, pp. 10–3.

Kelley, K., Clark, B., Brown, V. & Sitzia, J. 2003, ‘Good practice in the conduct and reporting of survey research’, *International Journal for Quality in Health Care*, pp. 261–6.

Khalid, E. 2008, ‘A knowledge-based advisory system for software quality assurance’, *Journal of Chemical Information and Modeling*, vol. 5, no. 3, pp. 304–10.

Kimberlin, C.L. & Winterstein, A.G. 2008, ‘Validity and reliability of measurement instruments used in research’, *American Journal of Health-System Pharmacy*, vol. 65, no. 23, pp. 2276–84.

Kloppenborg, T.J. & Petrick, J.A. 2004, ‘Managing project quality’, *IEEE Engineering Management Review*, vol. 32, no. 4, pp. 86–90.

Kostalova, J., Tetrevova, L. & Svedik, J. 2015, ‘Support of Project Management Methods by Project Management Information System’, *Procedia - Social and Behavioral Sciences*, vol. 210, pp. 96–104.

Kothari, R. 2003, *Research methodology: methods and techniques*, 2nd edn, New Age International Publishers.

Krippendorff, K.H. 2004, *Content Analysis: An Introduction to Its Methodology*, Second Edi., SAGE Publications, London.

Laranjeiro, N., Soydemir, S.N. & Bernardino, J. 2016, ‘A Survey on Data Quality: Classifying Poor Data’, *Proceedings - 2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing, PRDC 2015*, pp. 179–88.

Latham, B. 2007, ‘Sampling : What is it ?’, *Quantitative Research Methods ENGL 5377*, vol. 1, pp. 1–12.

Lavalle, S., Lesser, E., Shockley, R., Hopkins, M.S. & Kruschwitz, N. 2015, *Big Data , Analytics and the Path*, pp. 1–14.

Levy, D. 2006, ‘Qualitative methodology and grounded theory in property research’, *Pacific Rim Property Research Journal*, vol. 12, no. 4, pp. 369–88.

Liberatore, M.J.. & Pollack-Johnson, B.. 2013a, ‘Improving project management decision making by modeling quality, time, and cost continuously’, *IEEE Transactions on Engineering Management*, vol. 60, no. 3, pp. 518–28.

Liberatore, M.J.. & Pollack-Johnson, B.. 2013b, ‘Improving project management decision making by modeling quality, time, and cost continuously’, *IEEE Transactions on Engineering Management*, vol. 60, no. 3, pp. 518–28.

Lucas, A. 210AD, *Corporate Data Quality Management - From Theory to Practice*, pp. 1–7.

MacDonald, S. & Headlam, N. 1999, *Research methods & statistics handbook*, *Statistics*.

Maqsood, M. & Javed, T. 2007, ‘Practicum in software project management: an endeavor to effective and pragmatic software project management education’, *Proceedings of the 4th WSEAS/IASME International Conference on Engineering Education*, pp. 103–12.

Marques, G., Gourc, D. & Lauras, M. 2011, ‘Multi-criteria performance analysis for decision making in project management’, *International Journal of Project Management*, vol. 29, no. 8, pp. 1057–69.

Maxwell, J.A. 2012, *Qualitative Study*, no. January 2012.

Maydanchik, A. 2007, ‘Data quality assessment’, *Communications of the ACM*, vol. 45, no. 4, pp. 211–8.

Mcavoy, J. & Butler, T. 2009, *The role of project management in ineffective decision making within Agile software development projects*, no. June, pp. 372–83.

McLeod, S. 2018, ‘Correlation | Simply Psychology’, *Simply Psychology*.

Mishra, N. & Silakari, S. 2012, ‘Predictive Analytics : A Survey , Trends , Applications ’, *International Journal of Computer Science and Information Technologies*, vol. 3, no. 3, pp. 4434–8.

Modeling, C.M. 2005, ‘Capability Maturity Model ® Integration (CMMI ®) Overview’, *Defense*, pp. 1–58.

Morgan, D.L. 2014, ‘Research design and research methods’, *Integrating Qualitative and Quantitative Methods: a pragmatic approach*, pp. 45–62.

Namboodiri, K. 2011, ‘Eigenvalues and Eigenvectors’, *Matrix Algebra*, pp. 75–94.

Nayebi, M., Ruhe, G., Mota, R.C. & Mufti, M. 2015, ‘Analytics for Software Project Management -- Where are We and Where do We Go?’, *2015 30th IEEE/ACM International Conference on Automated Software Engineering Workshop (ASEW)*, pp. 18–21.

Newton, P. 2015, ‘Managing Project Quality’, *Bookboon*, pp. 1–50.

North Carolina Department of Public Instruction 2010, *Data Quality Management : Best Practices*, pp. 1–2.

O’Reilly, K. 2012, ‘Key Concepts in Ethnography Ethics’, *Key Concepts in Ethnography*, pp. 34–8.

Peralta, V. 2008, *Data Quality Evaluation in Data Integration Systems*.

Picardo, E. 2019, ‘Negative Correlation Definition’, *Investopedia*.

Pipino, L.L., Lee, Y.W. & Wang, R.Y. 2002a, ‘Data quality assessment’, *Communications of the ACM*, vol. 45, no. 4, p. 211.

Pipino, L.L., Lee, Y.W. & Wang, R.Y. 2002b, ‘Data quality assessment’, *Communications of the ACM*, vol. 45, no. 4, p. 211.

Piprani, B. & Ernst, D. 2008, ‘A Model for Data Quality Assessment’, *On the Move to Meaningful Internet Systems: OTM 2008 Workshops*, pp. 750–9.

Project Management Institute 2008, *Project Human Resource Management*, *A Guide to the Project Management Body of Knowledge (PMBOK® Guide) - Fourth Edition*, 4th edn, Project Management Institute, Inc.

Rajasekar, S., Philominathan, P. & Chinnathambi, V. 2013, ‘Research Methodology’, *New Age International*, p. 53.

Rao, D., Gudivada, V. & Raghavan, V. 2015, *Data Quality Issues in Big Data*, pp. 2654–60.

Rao, D., Gudivana, N.V. & Raghavan, V.V. 2015, 'Data Quality Issues in Big Data', *2015 IEEE International Conference on Big Data (Big Data) Data*, IEEE, pp. 2654–60.

Ritchie, J., Lewis, J., Nicholls, C.M. & Ormston, R. 2013, 'The Foundation of Qualitative Research', *Qualitative Research Practice: A Guide for Social Science Students and Researchers*, pp. 0–25.

Rodríguez, C.C.G., Riveill, M. & Antipolis, S. 2010, *E-Health Monitoring Applications : What About Data Quality ?*, no. August.

Rodríguez, P., Montequín, V.R., Morán, H. & De Arriba, L. 2017, 'Gender influence in project management: Analysis of a case study based on master students', *Procedia Computer Science*, vol. 121, pp. 461–8.

Sangeeta & Sharma, K. 2016, *Quality Issues with Big data Analytics*, pp. 3589–91.

Sanjay, M. & Alamma, B.H. 2016, 'An insight into big data analytics — Methods and application', *2016 International Conference on Inventive Computation Technologies (ICICT)*, pp. 1–5.

Santos, J.R.A. 1999, 'Cronbach's alpha: A tool for assessing the reliability of scales', *Journal of Extension*, pp. 1–5.

Saunders, M., Lewis, P. & Thornhill, A. 2008, *Research Methods for Business Students*, *Research methods for business students*, 5th Editio., Prentice Hall.

Scotland, J. 2012, 'Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms', *English Language Teaching*, vol. 5, no. 9, pp. 9–16.

Shankaranarayanan, G. & Cai, Y. 2006, 'Supporting data quality management in decision-making', *Decision Support Systems*, vol. 42, no. 1, pp. 302–17.

Sidi, F., Shariat Panahy, P.H., Affendey, L.S., Jabar, M.A., Ibrahim, H. & Mustapha, A. 2012, 'Data quality: A survey of data quality dimensions', *Information Retrieval & Knowledge Management (CAMP), 2012 International Conference on*, pp. 300–4.

Singh, H. 2018, 'Introduction to Project Management Analytics: A Data-Driven Approach to Making Rational and Effective Project Decisions | What Is Analytics? | FT Press', *Financial Times Press*, viewed <http://www.ftpress.com/articles/article.aspx?p=2461067>.

Singh, R. & Singh, K. 2010, 'A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing', *IJCSI International Journal of Computer Science Issues*, vol. 7, no. 3, pp. 41–50.

Siniscalco, M.T. & Auriat, N. 1998, 'Questionnaire Design', *Quantitative research methods in educational planning*, vol. 46, no. 4, pp. 7–47.

Sokol, L. & Ames, R. n.d., *Analytics in a Big Data Environment*.

Špundak, M. 2014, ‘Mixed Agile/Traditional Project Management Methodology – Reality or Illusion?’, *Procedia - Social and Behavioral Sciences*, vol. 119, pp. 939–48.

Stamelos, I. 2010, ‘Software project management anti-patterns’, *Journal of Systems and Software*, vol. 83, no. 1, pp. 52–9.

Statistics Solutions 2019, ‘The Assumption of Homogeneity of Variance - Statistics Solutions’, *StatisticsSolutions*.

Stephanie 2014, *A Guide to Bartlett’s Test of Sphericity - Statology1*.

Stolovitsky, N. 2011, *The Role of Analytics in Projects Tunnel vision*, no. January 2011, pp. 1–3.

TAIWO, A.A. & DOWNE, A.G. 2013, ‘<THE THEORY OF USER ACCEPTANCE AND USE OF TECHNOLOGY (UTAUT)- A META-ANALYTIC REVIEW OF EMPIRICAL FINDINGS .pdf>’, *Journal of Theoretical and Applied Information Technology*, vol. 49, no. 1.

Talia, D. 2013, ‘Clouds for scalable big data analytics’, *Computer*, vol. 46, no. 5, pp. 98–101.

Tamene, E.H. 2016, ‘Theorizing conceptual framework’, *Asian Journal of Educational Research*, vol. 4, no. 2, pp. 50–6.

Tan, W. 2008, *Practical Research Methods*.

Tavakol, M. & Dennick, R. 2011a, ‘Making sense of Cronbach’s alpha’, *International Journal of Medical Education*, vol. 2, pp. 53–5.

Tavakol, M. & Dennick, R. 2011b, ‘Making sense of Cronbach’s alpha’, *International Journal of Medical Education*, vol. 2, pp. 53–5.

Tefas, A. & Pitas, I. 2016, ‘Principal component analysis’, *Intelligent Systems*.

Tene, O. & Polonetsky, J. 2013, ‘Big data for all: Privacy and user control in the age of analytics’, *Northwestern Journal of Technology and Intellectual Property Volume*, vol. 11, no. 5, pp. 240–73.

ULARU, E.G., PUICAN, F.C., APOSTU, A. & VELICANU, M. 2012, ‘Perspectives on Big Data and Big Data Analytics Elena’, *Perspectives on Big Data and Big Data Analytics Elena*, vol. III, no. September, pp. 1–79.

Vasconcellos, F.J.S. & Moura, H.P. De 2016, *Decision-Making in Software Project Management: A Qualitative Case Study of a Private Organization*, pp. 26–32.

Veregin, H. 1999, ‘Data quality parameters’, *Geographical information systems*, pp. 177–90.

Walia, G.S. & Carver, J.C. 2009, ‘A systematic literature review to identify and classify software requirement errors’, *Information and Software Technology*, vol. 51, no. 7, pp. 1087–109.

Wang, Y., Liu, D. & Ruhe, G. 2004, ‘Formal Description of the Cognitive Process of Decision Making’, *Strategies*.

White, D. & Fortune, J. 2002, ‘Current practice in project management — an empirical study', *International Journal of Project Management*, vol. 20, no. 1, pp. 1–11.

Williams, N., Ferdinand, N.P. & Croft, R. 2014, 'Project management maturity in the age of big data', *International Journal of Managing Projects in Business*, vol. 7, no. 2, pp. 311–7.

Wulff, N. 2017, 'What's The Difference Between Data Analytics And Data Analysis?', *Getsmarter*.

Yeo, E., Rudra, A. & Yeo, E. 1999, 'Key issues in achieving data quality and consistency in data warehousing among large organisations in Australia', *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, vol. Track7, no. c, p. 8.

Young, R.R. 2002, 'Recommended requirements gathering practices', *Crosstalk*, vol. 15, no. 4, pp. 9–12.

Zhang, D., Dang, Y., Lou, J.-G., Han, S., Zhang, H. & Xie, T. 2011, 'Software analytics as a learning case in practice', *Proceedings of the International Workshop on Machine Learning Technologies in Software Engineering - MALETS '11*, pp. 55–8.

Appendix A

Ethics Clearance Certificate

UNISA university of south africa

UNISA COLLEGE OF SCIENCE, ENGINEERING AND TECHNOLOGY'S (CSET) RESEARCH AND ETHICS COMMITTEE

08 December 2017

Ref #: 110/RN/2017/CSET_SOC
Name: Mr Rutendo Ngarira
Student #: 46893628

Dear Mr Rutendo Ngarira

Decision: Ethics Approval for 3 years (Humans involved)

Researcher: Mr Rutendo Ngarira
Unit 2 Valley Pride, 1005 Tanga Street, Strubens Valley, Code 1735, Roodepoort
46893628@mylife.unisa.ac.za / Ngarira@yahoo.com
+27 73 652 6314 / 083 702 0004

Supervisor(s): Prof E Mnkandla, mnkane@unisa.ac.za, +27 11 670 9059

Proposal: Improving software project management quality through the use of analytics on project management data

Qualification: MSc in Computing

Thank you for the application for research ethics clearance by the Unisa College of Science, Engineering and Technology's (CSET) Research and Ethics Committee for the above mentioned research. Ethics approval is granted for a period of three years, from 08 December 2017 to 08 December 2020.

1. The researcher will ensure that the research project adheres to the values and principles expressed in the UNISA Policy on Research Ethics.

University of South Africa
Preller Street, Muckleneuk Ridge, City of Tshwane
PO Box 392 UNISA 0003 South Africa
Telephone +27 12 429 3111 Facsimile +27 12 429 4150
www.unisa.ac.za

2. Any adverse circumstance arising in the undertaking of the research project that is relevant to the ethicality of the study, as well as changes in the methodology, should be communicated in writing to the Unisa College of Science, Engineering and Technology's (CSET) Research and Ethics Committee. An amended application could be requested if there are substantial changes from the existing proposal, especially if those changes affect any of the study-related risks for the research participants.
3. The researcher will ensure that the research project adheres to any applicable national legislation, professional codes of conduct, institutional guidelines and scientific standards relevant to the specific field of study.
4. Only de-identified research data may be used for secondary research purposes in future on condition that the research objectives are similar to those of the original research. Secondary use of identifiable human research data require additional ethics clearance.

Note:

The reference number 110/RN/2017/CSET_SOC *should be clearly indicated on all forms of communication with the intended research participants, as well as with the Unisa College of Science, Engineering and Technology's (CSET) Research and Ethics Committee.*

Yours sincerely

pp (Dr B Chimbo)

Dr. A Da Veiga

Chair: Ethics Sub-Committee School of Computing, CSET

Prof I. Osunmakinde

Director: School of Computing, CSET

Prof B. Mamba

Executive Dean: College of Science, Engineering and Technology (CSET)

University of South Africa
Preller Street, Muckleneuk Ridge, City of Tshwane
PO Box 392 UNISA 0003 South Africa
Telephone: +27 12 429 3111 Facsimile: +27 12 429 4150
www.unisa.ac.za

Appendix B

Questionnaire

The purpose of this research is to establish if using data analytics in software project management improves software project management quality through enhanced decision-making and project management standards.

Please complete the questions by selecting the appropriate answers or filling in the provided spaces.

The completed questionnaire should be emailed to Ngarira@yahoo.com or 46893628@mylife.unisa.ac.za

This Questionnaire forms part of research project in fulfilment of the requirements for the MSc in Computing (UNISA).

The study is being conducted by **Rutendo Ngarira** and there is no organisation or group sponsoring the research.

Please Note:

Completion of this questionnaire is completely voluntary, and the information provided through this questionnaire is solely for research purposes as mentioned above. The confidentiality of the provided information will be maintained.

Definition of Terms

Quality: Refers to the extent to which a system or process satisfies the specified requirements and user expectations (Al-Kilidar et al. 2005:126).

Project Management: Is the use of knowledge, skills, tools and techniques on project activities to be able to satisfy stakeholders' expectations from a project (PMI 2008:6).

Decision-making: Is a process of choosing a preferred option or a course of action from a set of available alternatives on the basis of given criteria or strategies (Wang, Liu & Ruhe 2004:124).

Name and Surname :

Name of Company :

Email Address :

Pease tick or provide the appropriate answers

SECTION A: Background Information

1. What is your Age range?

< 20	20-29	30-39	40-49	50-59	> 59

2. What is your Gender?

Male	**Female**

3. Which of the following best describes your current working project management position?

	✓
In-house project manager	
Works for a consulting firm	
Freelance project manager	
Currently not managing software projects	

4. How do you measure your level of data analytics knowledge in software project management?

Data Analytics knowledge level	✓
No knowledge of data analytics	
Only read about data analytics	
Elementary Knowledge - used data analytics in one project	
Average Knowledge- used data analytics in less than five projects	
Expert Knowledge- used data analytics in five or more projects	

SECTION B: RESEARCH VARIABLES

Initial Phase

5. With the initial phase organisations' software projects (without defined processes standards in place) you worked on, the projects' processes and activities were mostly Ad-hoc.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

6. Where data analytics has not been used in your past initial phase projects, the successful projects processes were not easily repeatable.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

7. Based on your past initial phase projects where data analytics was not used by the project team, the projects milestones were not easily measurable.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

8. With the initial phase software projects, you were involved in (which you did not use data analytics), how many would you consider having failed due to poorly controlled processes?

	✓
0	
1 - 5	
6 – 10	
> 10	

Project Success

9. In the projects where data analytics was not used, the projects failed to be completed within the budgeted cost.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

10. Based on your past projects where data analytics was not used, it affected the delivery of software projects within estimated time?

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

11. The projects you were part of which failed to be completed in the estimated time and cost was mainly due to ineffective analysis of user requirements.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

12. Projects failure in terms of the projects scope not being met due to scope creep where data analytics was not used in the projects was common.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

13. Based on your past projects, what percentage of the initial phase (with no prescribed organisational standards) projects' successes depended on individual talent?

Percentage (%)	✓
< 25	
25 – 49	
50 – 74	
75 – 100	

Managed level

14. Given your past projects, data analytics use enhanced projects processes which helped achieve set projects goals.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

15. Considering your past projects, the projects processes control (Project Monitoring and Control (PMC)) have been more attainable by using data analytics.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

16. Which of the following managed level process areas would you say where further enhanced by data analytics use in your projects?

	✓
None	
Project Planning (PP)	
Configuration Management (CM)	
Supplier Agreement Management (SAM)	
Measurement and Analysis (MA)	
Process and Product Quality Assurance (PPQA)	

Impact on Customer

17. The use of data analytics in your projects influenced realising projects user functionality requirements (Requirements Management (REQM))?

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

18. Given the projects you have been part of where data analytics was used, it enhanced chances of realising software performance goals.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

Defined level

19. Which of the following software quality frameworks or standards do you comply with?

	✓
ISO 9000	
CMMI	
CMM	
PMMM	
NONE	
Other (Please Specify)	

20. The use of data analytics in your projects led to consistency in software projects completion according to plan?

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

21. Based on your experience, data analytics use enhance management of projects interrelated processes?

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

Organisational Success

22. Your project team frequently attend on job project management training workshops and courses?

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

23. Projects stakeholders i.e. users, sponsors and management actively participated in the projects I worked on which led to the projects' success.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

24. Where data analytics was used in your projects it led to organisational success in terms of financial gain from system reliability and efficiency.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

25. In your projects where data analytics has been used it enhanced standardisation in organisations' processes.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

Quantitatively Managed Level

26. What effect if any did data analytics use have on decision-making in your projects?

Decision-Making Efficiency	✓
None	
Quicker and timely decision-making	
Greater project control	
Improved resource allocation and management	
Specify Other:	

27. Based on your past projects, improved decision-making has a direct impact on software project management quality.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

28. From your last project, what percentage of delayed or incorrect decision-making can you attribute to ineffective data analysis?

Percentage (%)	✓
< 25	
25 – 49	
50 – 74	
75 – 100	

29. Which of the following quality control techniques have you used in your projects?

	✓
Cause and Effect Diagram	
Inspection	
Flow Chart	
Control Chart	
Pareto Chart	
Histogram	
Run Chart	
Statistical Sampling	
Scatter Diagram	
Other (Specify)	

30. Which of the following software quality dimensions were improved in your last five projects by using data analytics?

	✓
Reliability	
Performance	
Accuracy	
Completeness	
Consistency	
Serviceability	
Integrity	

Project Management Efficiency

31. What software project management certification do you have?

Certification	✓
CAPM: Certified Association in Project Management (PMI)	
PMP: Project Management Professional (PMI)	
CSM: Certified Scrum Master (Scrum Alliance)	
CSSBB: Certified Six Sigma Black Belt (ASQ)	
CSSGB: Certified Six Sigma Green Belt (ASQ)	
Global Association for Quality Management	
Prince2	
CompTIA Project +	
Other (Specify) ______________________________	

32. Considering the projects you have used quality control's inspection technique, how does data analytics improve the technique or other quality control techniques you have used?

____
____
____

33. Based on your experience, how else can data analytics improve software project management quality?

____
____
____
____

Optimising Level

34. Have you used any automated data analysis tools besides data analytics in the last three software projects you were involved in?

	✓
Not at all	
In one of the projects	
In two of the projects	
In all three projects	

35. Where you have used data analytics, achieving software design requirements was easier?

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

36. Based on your past projects, data analytics use improved handling projects processes variations.

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

Preparing for the Future

37. Your company facilitate or offer data analytics training to software project managers?

	✓
Strongly Disagree	
Disagree	
Somewhat Agree	
Agree	
Strongly Agree	

38. Which software development methodology do you use in your projects? Traditional methodologies include Spiral, Waterfall, SDLC and Agile methodologies include Scrum, Extreme Programming (XP), Feature Driven Development (FDD), Adaptive Software Development (ASD) etc.

	✓
Traditional Software development Methodologies	
Agile Software Development Methodologies	
Both Methodologies	
Other (Specify)	

39. How does the use of data analytics improve the software development methodology that you have used, contributing towards enhanced future project management?

____

____

____

40. What factors did you observe to be necessary for data analytics to be successfully used in software project management?

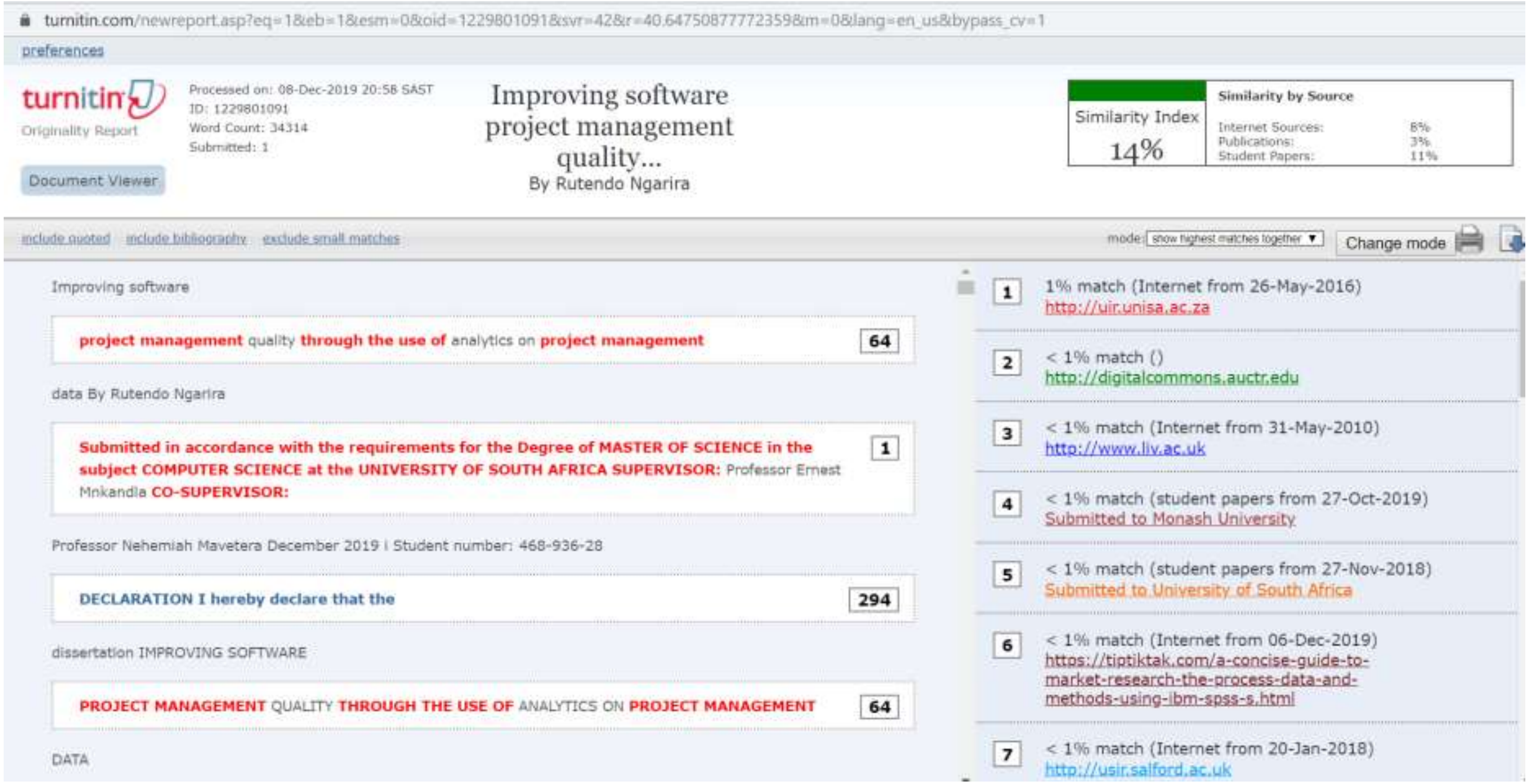
____

____

____

Appendix C

Turnitin Results



turnitin.com/newreport.asp?eq=1&eb=1&esm=0&oid=1229801091&svr=42&r=40.64750877772359&m=0&lang=en_us&bypass_cv=1

preferences

turnitin Originality Report

Processed on: 08-Dec-2019 20:58 SAST
ID: 1229801091
Word Count: 34314
Submitted: 1

Document Viewer

Improving software project management quality...
By Rutendo Ngarira

Similarity Index	Similarity by Source	
14%	Internet Sources:	8%
	Publications:	3%
	Student Papers:	11%

include quoted include bibliography exclude small matches

mode: show highest matches together ▼ Change mode

Improving software

project management quality through the use of analytics on project management 64

data By Rutendo Ngarira

Submitted in accordance with the requirements for the Degree of MASTER OF SCIENCE in the subject COMPUTER SCIENCE at the UNIVERSITY OF SOUTH AFRICA SUPERVISOR: Professor Ernest Mnkandla CO-SUPERVISOR: 1

Professor Nehemiah Mavetera December 2019 I Student number: 468-936-28

DECLARATION I hereby declare that the 294

dissertation IMPROVING SOFTWARE

PROJECT MANAGEMENT QUALITY THROUGH THE USE OF ANALYTICS ON PROJECT MANAGEMENT 64

DATA

1 1% match (Internet from 26-May-2016)
http://uir.unisa.ac.za

2 < 1% match ()
http://digitalcommons.auctr.edu

3 < 1% match (Internet from 31-May-2010)
http://www.liv.ac.uk

4 < 1% match (student papers from 27-Oct-2019)
Submitted to Monash University

5 < 1% match (student papers from 27-Nov-2018)
Submitted to University of South Africa

6 < 1% match (Internet from 06-Dec-2019)
https://tiptiktak.com/a-concise-guide-to-market-research-the-process-data-and-methods-using-ibm-spss-s.html

7 < 1% match (Internet from 20-Jan-2018)
http://usir.salford.ac.uk

Appendix D

Language-editing confirmation

CONFIRMATION OF EDITING

To whom it may concern at the University of South Africa.

I, Victor Ramoseng William Mecoamere, former Nation Building manager and news editor at Sowetan newspaper, now retired, hereby declare that I was responsible for the editing and proofreading of the dissertation by Mr Rutendo Ngarira, titled:

IMPROVING SOFTWARE PROJECT MANAGEMENT QUALITY THROUGH THE USE OF ANALYTICS ON PROJECT MANAGEMENT DATA.

SIGNED: ....

PLACE: EDENVALE, GAUTENG

DATE: 04-12-2019