

CRITICAL SUCCESS FACTORS FOR MANAGING DATA SCIENCE PROJECTS WITHIN AGILE METHODOLOGY

Garusin Arachchige Imali Limesha

199116P

Master of Business Administration in Information Technology Specialized
in Business Analytics

Department of Computer Science Engineering

University of Moratuwa

Sri Lanka

July 2021

CRITICAL SUCCESS FACTORS FOR MANAGING DATA SCIENCE PROJECTS WITHIN AGILE METHODOLOGY

Garusin Arachchige Imali Limesha

199116P

Thesis Submitted in Partial Fulfilment of the Requirements for the Master
of Business Administration in Information Technology specialized in
Business Analytics.

Department of Computer Science Engineering

University of Moratuwa

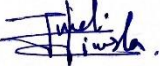
Sri Lanka

July 2021

DECLARATION

This thesis contains, as far as I am aware, no material previously submitted for a Master of Business Administration in any other university or institute of higher education, nor does it include, without recognition, material previously published or written by someone other than when the recognition is provided by another individual.

I also confer a non-exclusive right on University Moratuwa to reproduce and distribute my thesis in print, electronic or other media, wholesale or partial. I remain entitled to make full or part use of this material in future work (such as articles or books).

Signature: 

Date: 03rd July, 2021

The supervisor/s should certify the thesis/dissertation with the following Declaration.

The above candidate has carried out research for the MBA in Information Technology Dissertation under my supervision.

Name of the supervisor: Dr. Amal Shehan Perera

Signature of the supervisor:



Date: 03- 07- 2021

Abstract

Data science is an evolving discipline with a major emphasis on developing accessible data analysis techniques. But far less emphasis has been given on the factors which affect the success of data science projects under the Agile umbrella. In the field of software engineering agile approaches were initially developed and are characterized by their iterative software development approach. It is recommended to use process models or methodologies in literature to increase the success rate of data science projects however, organizations, which are perceived to be too restrictive and do not accept the traditional iteratives and transparent nature of data science projects, reluctant to use them. And there are some potential challenges which have been identified in the literature for using Agile methodologies in data science projects. The characteristics of possible critical success factor (CSF)s for Data Science projects have been established from the literature by updating Chow and Cao's list of success factors for agile software development projects in this research. The factors have been identified under five dimensions of organizational, people, process, technical and project. The findings of this study indicate team environment, team capacity, client engagement, project definition processes, agile software engineering techniques and project schedule as the factors that impact the success of data science projects within Agile methodology. Even though these factors were listed as important for managing data science projects within Agile methodology, the significance of these factors may vary according to the nature of the project that the team is involved in. Therefore, the team should always focus on these factors relative to the nature of the project.

Key words: Data Science, Project Management, Agile Methodology

Acknowledgement

I would like to thank all of those who helped me to carry out my research study on “Critical Success Factors for Managing Data Science Projects within Agile Methodology” successfully. First and foremost, I wish to thank my research supervisor Dr. Amal Shehan Perera Senior Lecturer of the Department of Computer Science and Engineering, University of Moratuwa, for the continuous support, encouragement and attention extended to me throughout my research.

And I would specially thank all the respondents who participated and spent their precious time giving their valuable responses to my online survey. Moreover, thanks to MBA batch mates I extend my commitment to the guidance and support for getting this study to a successful conclusion.

My dear husband, Shanaka Chathuranga, is most appreciative of all the motivation, care and unconditional help that I receive in good and bad times during this research study.

I would like to thank all those who helped me to carry this analysis to a successful conclusion.

Table of Contents

DECLARATION	iii
Abstract	iv
Acknowledgement	v
1.1 Introduction	1
1.2 Background and Motivation	2
1.3 Problem Statement	3
1.4 Objectives of Research	4
1.5 Outline	5
CHAPTER 2 - LITERATURE REVIEW	6
2.1 Introduction	6
2.2 Agile Software Development	6
2.2.1 Emergence of Agile	6
2.2.2 Drivers of Agile	7
2.3 Agile Frameworks used in Data Science Projects	9
2.3.1 SCRUM	9
2.3.2 CRISP-DM	9
2.3.3 Kanban	10
2.3.4 Team Data Science Process (TDSP)	11
2.3.5 SCRUM- DS	11
2.4 Challenges of Agile Frameworks for Data Science Projects	12
2.4.1 Challenges of SCRUM in Data Science projects	12
2.4.2 Challenges of Kanban in Data Science projects	12
2.4.3 Challenges of CRISP-DM in Data Science projects	13
2.4.4 Challenges of TDSP in Data Science projects	13
2.5 Critical Success Factors for Data Science Projects	13
2.6 Summary	16
CHAPTER 3 – Methodology	17
3.1 Introduction	17
3.2 Conceptual Framework	17
3.3 Variables on Relationships	18

3.3 Hypothesis Development	18
3.3.1 Organizational Factors	18
3.3.2 People Factors	20
3.3.3 Process Factors	20
3.3.4 Technical Factors	22
3.3.5 Project Factors	23
3.4 Questionnaire Instrument Development	24
3.5 Sample Selection	25
3.6 Method of Data Collection	26
3.7 Summary	26
CHAPTER 4 – Analysis and Interpretation	27
4.1 Introduction	27
4.2 Reliability Test	27
4.2.1 Preliminary Survey	27
4.2.1 Research Survey	29
4.3 Descriptive Statistics Analysis	30
4.3.1 Composition of the sample according to Age Group	30
4.3.2 Composition of the sample according to Gender	30
4.3.3 Composition of the sample according to Designation	30
4.3.4 Composition of the sample according to Working Experience.	33
4.3.5 Composition of the sample according to the used Agile approach.	34
4.4 Inferential Statistics – Inter Item Correlation Analysis	35
4.5 Hypothesis Testing – Pearson’s Correlation Analysis	39
4.6 ANOVA Testing	47
4.8 Discussion on Results	47
4.9 Summary	49
CHAPTER 5 – Conclusion and Recommendation	50
5.1 Introduction	50
5.2 Research Implications	50
5.3 Research Limitation and Future Works	52
References	53

APPENDIX A	56
Questionnaire Instrument	56
APPENDIX B	61
Descriptive Statistics	61

Table of Figures

Figure 2. 1: Agile Manifesto	7
Figure 2. 2: Agile learning Curve	8
Figure 3. 1: Conceptual Framework	17
Figure 4.1: Frequency Distribution of Age Group	28
Figure 4.2: Frequency Distribution of Gender	29
Figure 4.3: Distribution of Job Designation	30
Figure 4.4: Distribution of Working Experience	31
Figure 4.5: Distribution of Agile methodologies used	32

Table of Tables

Table 3. 1: Summary of measurements in independent variables	24
Table 3. 2: Summary of measurements in dependent variable	25
Table 3. 3: Summary of measurements in Demographic factors	25
Table 4. 1: Cronbach's Alpha Coefficient for Preliminary Survey	30
Table 4. 2: Item-wise Cronbach's Alpha Coefficients	31
Table 4. 3: Cronbach's Alpha Coefficient for Research Survey	32
Table 4. 4: Cronbach's Alpha Coefficient for Dependent Variable	32
Table 4. 5: Single Item Variables	32
Table 4. 6: Age Distribution	33
Table 4. 7: Gender Distribution	34
Table 4. 8: Distribution of Job Designation	34

Table 4. 9: Distribution of Working Experience	35
Table 4. 10: Distribution of Agile methodologies used.	36
Table 4. 11: Inter Item Correlation for Organizational Environment	37
Table 4. 12: Inter Item Correlation for Team Environment	37
Table 4. 13: Inter Item Correlation for Customer Involvement	37
Table 4. 14: Inter Item Correlation for Project Definition Process	38
Table 4. 15: Inter Item Correlation for Agile Software Engineering Techniques	38
Table 4. 16: Inter Item Correlation for Project Schedule	38
Table 4. 17: Inter Item Correlation for Success of the Data Science projects within the Agile Methodology.	39
Table 4. 18: Pearson’s Correlation- Management Commitment and Success of the Data Science projects within the Agile Methodology.	40
Table 4. 19: Pearson's Correlation- Organizational Environment and Success of the Data Science projects within the Agile Methodology.	40
Table 4. 20: Pearson’s Correlation- Team Environment and Success of the Data Science projects within the Agile Methodology.	41
Table 4. 21: Pearson's Correlation- Team Capability and Success of the Data Science projects within the Agile Methodology.	42
Table 4. 22: Pearson's Correlation- Customer Involvement and Success of the Data Science projects within the Agile Methodology.	42
Table 4. 23: Pearson's Correlation- Project Management Process and Success of the Data Science projects within the Agile Methodology.	43
Table 4. 24: Pearson's Correlation- Project Definition Process and Success of the Data Science projects within the Agile Methodology.	43
Table 4. 25: Pearson's Correlation- Improved Coordination and Success of the Data Science projects within the Agile Methodology.	44
Table 4. 26: Pearson's Correlation- Agile Software Engineering Techniques and Success of the Data Science projects within the Agile Methodology.	44
Table 4. 27: Pearson's Correlation- Delivery Strategy and Success of the Data Science projects within the Agile Methodology.	45
Table 4. 28: Pearson's Correlation- Project Type and Success of the Data Science projects within the Agile Methodology.	46

Table 4. 29: Pearson's Correlation- Project Schedule and Success of the Data Science projects within the Agile Methodology.	46
Table 4. 30: One-Way ANOVA- Agile methodologies used.	47
Table B. 1 Statistics: Management Commitment (MC) – Item No. 1	64
Table B. 2 Statistics: Management Commitment (MC) – Item No. 2	65
Table B. 3 Statistics: Organizational Environment (OE) – Item No. 3	66
Table B. 4 Statistics: Organizational Environment (OE) – Item No. 4	67
Table B. 5 Statistics: Team Environment (TE) – Item No. 5	67
Table B. 6 Statistics: Team Environment (TE) – Item No. 6	68
Table B. 7 Statistics: Team Capability (TC) – Item No. 7	69
Table B. 8 Statistics: Team Capability (TC) – Item No. 8	69
Table B. 9 Statistics: Customer Involvement (CI) – Item No. 9	70
Table B. 10 Statistics: Customer Involvement (CI) – Item No. 10	71
Table B. 11 Statistics: Project Management Process (PMP) – Item No. 11	71
Table B. 12 Statistics: Project Management Process (PMP) – Item No. 12	72
Table B. 13 Statistics: Project Definition Process (PDP) – Item No. 13	73
Table B. 14 Statistics: Project Definition Process (PDP) – Item No. 14	73
Table B. 15 Statistics: Improved Coordination (IC) – Item No. 15	73
Table B. 16 Statistics: Agile Software Engineering Techniques (ASE) – Item No. 16	74
Table B. 17 Statistics: Agile Software Engineering Techniques (ASE) – Item No. 17	74
Table B. 18 Statistics: Agile Software Engineering Techniques (ASE) – Item No. 18	75
Table B. 19 Statistics: Agile Software Engineering Techniques (ASE) – Item No. 19	75
Table B. 20 Statistics: Delivery Strategy (DS) – Item No. 20	75
Table B. 21 Statistics: Delivery Strategy (DS) – Item No. 21	76
Table B. 22 Statistics: Project Type (PT) – Item No. 22	77
Table B. 23 Statistics: Project Schedule (PS) – Item No. 23	77
Table B. 24 Statistics: Project Schedule (PS) – Item No. 24	78
Table B. 25 Statistics: Success of the Data Science projects within Agile methodology (SUCCESS) – Item No. 25	79

Table B. 26 Statistics: Success of the Data Science projects within Agile methodology (SUCCESS) – Item No. 26	79
Table B. 27 Statistics: Success of the Data Science projects within Agile methodology (SUCCESS) – Item No. 27	80

CHAPTER 1 - INTRODUCTION

1.1 Introduction

After vast amounts of data became available, companies in almost every industry focused on data discovery to gain a competitive advantage. In data science, the term "science" refers to the application of the scientific method. This scientific approach provides a learning loop in which the team will ask intriguing questions, which they will then research to come up with new ideas. Data Science represents a collection of foundational concepts that endorse and direct high-level data processing. Working in a data science team necessitates a completely different attitude. Most businesses still depend on planning, setting up goals, and specific deadlines. But the data science team must challenge, investigate, and respond.

In software development firms, it is seen that how challenging it is to get data science teams to operate inside a project management system. Traditional projects rely on scope, cost, and schedule. Data science is about experiments and exploration. It is a challenge to match the defined scope. It is also a problem to use traditional project management language. Data Science teams will not usually be able to find clear success criteria. It is not like a typical project. We cannot plan the work and then work the plan. There will not be necessarily have a set of clear objectives. Most of our team's experiments will just be dead ends. Just because traditional project management does not fit to our team it does not mean that teams must work in chaos. The absence of a plan should not mean the absence of intent or value. Since data science increases organizational knowledge. New insights can decrease the time-to-market, create new revenue, avoid costs, and even build good will.

For a data science team to be effective, companies must rewire their thinking about work, letting go of the confidence in planning and producing in favor of exploring and discovering. The aim of the study is to find what factors influence the progress of data science projects when handling the related tasks under the Agile software development

methodology. While the literature has identified many success factors for Agile software development projects, it has also been indicated that there are many issues in data science projects using Agile methodology. As a result, knowing all about data science ventures will help businesses maximize market value by leading, helping, and managing the team.

1.2 Background and Motivation

Data science is known as the entire lifecycle of specifications, data processing, preparing, research, displaying, managing, and maintaining vast sets of data, resulting in practical insights [1]. Given that the approach includes a wide variety of disciplines including computer science, math, knowledge processing, and, above all, large data engineering, the idea that data science is not just analysis is supported by this comprehensive approach.

The Data Science team will investigate new possibilities and work to make data more available to the rest of the company. They are trying to figure out how to get a better understanding of their customers. All these efforts would be difficult to incorporate into a traditional project management system. Hence, there is no defined scope. The method of project management is well-defined. Before starting, the team must have a thorough understanding of the deliverable. Data science is a method based on empirical evidence. Data scientists must be able to respond to the information they are given. If they want to learn new things, they must be prepared for the unexpected.

When data science teams want to work differently, they must use a different life cycle. When confronted with an obstacle, a life cycle is a sequence of steps that must be followed. There are two common types that most professionals will encounter in large organizations. The first is the software development life cycle or SDLC. Plan, evaluate, develop, code, test, and deploy are the six phases of this life cycle and it is known as the waterfall model. This is because each step must be completed before the next can begin. The program is planned, and the specifications are analyzed by teams. Then they start coding after creating the basic designs. The program will be reviewed by quality assurance

experts until the code is complete. When it passes all the checks, it is released for public use.

The second life cycle is the cross-industry standard approach for data mining or CRISP-DM. Instead of software, this process model is used for data. It is a little more adaptable than the traditional waterfall model. There are six stages in all. Business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Teams devote a significant amount of time to either the preparation and analysis process or the market comprehension phase of all life cycles. The aim is to collect as much data as possible before starting, then produce it all at once. Not much is communicated in terms of final deliverables before the final deployment phase. When it comes to data science, this is not always the right solution. Since data science is an exploratory and experimental area. To understand the data, the teams will take an analytical approach, idea is that teams do not fully understand what they need until start working on it.

These life cycles cannot be used if companies want their data science teams to be versatile and exploratory. Instead, they should pursue a more lightweight approach to providing insights, one that provides structure while still allowing them to respond to new ideas. Good project management has different goals and processes than data science. However, this does not rule out the possibility of it being used in data science. Someone would almost certainly push the data science team to use these project management concepts. Many organizations are so used to project management, that it is hard to imagine spending money without the familiar scope, cost, and schedule. Project management is the tool that they are used to control these dimensions. Therefore, it is important to find out the factors which affect the success of the data science projects within the Agile framework.

1.3 Problem Statement

The objective of this analysis is to examine the variables that could have an influence on the progress of Agile methodology data science projects. Since an Agile project approach is also proposed as useful for projects in the field of data science, the Agile methodology

focuses on data science projects. An agile project life cycle will include a testing, iterative approach to tackling uncertainty and help project stakeholders learn and develop a workable solution. Critical success factors for data science operations using Agile approaches have yet to be thoroughly investigated. “What are the critical success factors in handling data science activities in IT projects using Agile methodology?” is the research question of the study. Agile approaches have developed over the last decade, and the key reasons for software development project success have been identified under the Agile umbrella. Though success factors for managing data science activities within Agile methodologies are uncommon, most of the literature does not discuss testing of relationships between factors and dimensions of project success. As a result, the study will fill in the gap as to what the relationship is between an established critical success factor and the success of a data science project using Agile methodology. The literature review and questionnaire will be used to address the research question.

1.4 Objectives of Research

The purpose of this research is to identify critical success factors for managing data science projects in an Agile framework. To compare various ideas about success factors of Agile methodology in software development affect in data science projects. For that, it is important to research the latest Agile methodologies used in the tech industry for data science initiatives, as well as the difficulties faced by data science practitioners when conducting their data science projects using the current methodology. As a result, to ensure the performance of data science initiatives in the IT industry, I aim to establish guidelines for managing data science activities through an Agile framework.

- To determine important success factors for handling data science projects in Agile methodology.
- To identify the existing Agile methodologies followed in the Software Industry to manage data science projects.

- To determine the connection between each defined success factor and the success of data science projects within Agile methodology.
- To make recommendations for the IT industry on how to handle data science activities using Agile methodologies.

The achievement of the above objectives will be mostly based on a literature review and questionnaire.

1.5 Outline

This thesis is divided into five chapters, each of which is listed briefly below.

The motivation and background for the study, as well as the research problem and research objectives, are all covered in Chapter 1.

Chapter 2 will provide an overview of the various literature related to essential success factors for handling data science activities in Agile IT projects. It has also expanded on the results of other studies that have looked at Agile models for handling data science projects.

The research methodology used in this analysis is explained in detail in Chapter 3, which includes detailed explanations of the theoretical context, research variable definitions, variable relationships, questionnaire instrument development, survey approach, hypotheses development, and so on.

The fourth chapter will present statistically the findings and outcomes associated with the collected data. There will also be a thorough review and discussion of the conclusions and outcomes.

Chapter 5 draws conclusions on the overall research result, including the shortcomings of the current study as well as potential research directions, based on the review and interpretation of the results.

CHAPTER 2 - LITERATURE REVIEW

2.1 Introduction

This chapter presents the findings of related studies done by different authors on critical success factors for managing data science projects within Agile methodology. Section 2.2 will discuss Agile Software Development. Section 2.3 will discuss the Agile frameworks which have been used to manage data science projects, and then section 2.4 will describe the challenges which have faced with these Agile frameworks when managing data science projects. Finally, in section 2.5, the essential success factors from previous studies in this field of study will be discussed.

2.2 Agile Software Development

The Agile Life Cycle (SDLC) software development paradigm has been evolving over the last 15 years and it has become dominant across organizations and projects in the world. There is a more trend to use Agile over waterfall methods in SDLC. The rise in popularity of Agile and the application of its concepts to data science is evidence of how this trend has progressed over time.

2.2.1 Emergence of Agile

Professionals who previously worked in the software industry acknowledged the shortcomings of waterfall approaches and suggested solutions that are more derived into an iterative methodology by recognizing that processes should not be guided sequentially but iteratively. As per the publication of the Agile manifesto and principles [29], number of reasons have been highlighted as waterfall technique shortcomings and have led to the establishment of Agile practice in software. These factors can be grouped into three primary categories. They are,

1. Software Development Human Factors
2. Cost and Risk Factors Changes
3. Recognizing the characteristics of complex systems

MANIFESTO FOR AGILE SOFTWARE DEVELOPMENT

We are uncovering better ways of developing software by doing it and helping others do it.
Through this work we have come to value:

Individuals and interactions over processes and tools
Working software over comprehensive documentation
Customer collaboration over contract negotiation
Responding to change over following a plan

That is, while there is value in the items on the right, we value the items on the left more.

Kent Beck	James Grenning	Robert C. Martin
Mike Beedle	Jim Highsmith	Steve Mellor
Arie van Bennekum	Andrew Hunt	Ken Schwaber
Alistair Cockburn	Ron Jeffries	Jeff Sutherland
Ward Cunningham	Jon Kern	Dave Thomas
Martin Fowler	Brian Marick	

© 2001, the above authors
this declaration may be freely copied in any form,
but only in its entirety through this notice.

Figure 2. 1: Agile Manifesto [29]

2.2.2 Drivers of Agile

As mentioned in the above section there are drivers which were identified in [29] for the Agile Software Development. One of the key constraints in modern software development is the team's rate of learning capacity. Agile encourages regular incorporation and iterations, which help to sustain this human aspect. The learning curve from [30] in figure 2.2 shows that early in the project, the team shows fast growth in terms of understanding.

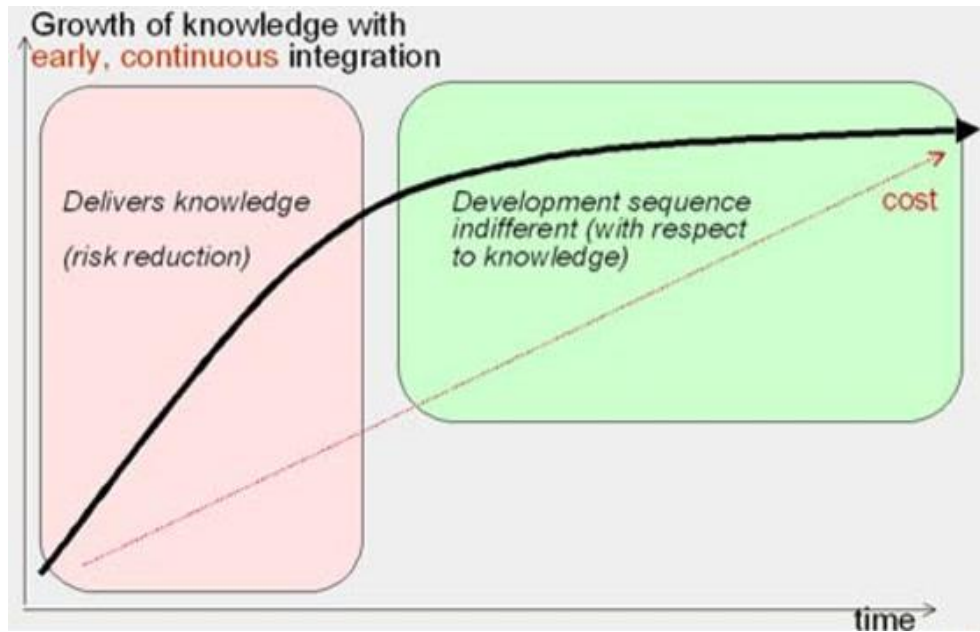


Figure 2. 2: Agile learning Curve [29]

Apart from these human factors, another aspect that drives the Agile approach in software development is the possibility of changes to specifications or design. Because of the rapid pace at which technology advances and the shortened expectations for product life cycle cycles, the cost of delaying product release is the greatest risk. There is an early Return on Investment and the opportunity to gain early reviews because of Agile's advice to generate usable applications early and regularly, which facilitates the delivery of the minimum viable product [31].

Many tasks fall into complicated and diverse realms when considering many software inventions. Interactions between humans and computers are unquestionably complicated. Agile teams' iterative delivery and emergent design practices are well-suited to complex domain projects. [31].

2.3 Agile Frameworks used in Data Science Projects

This section goes through the various Agile frameworks that a data science team might use to help them complete the project. This includes the various Agile methods, as well as how different teams can collaborate and what they can do to complete the project's data science tasks.

2.3.1 SCRUM

Scrum is a system for creating, delivering, and maintaining complex products that allows people to solve complex adaptive problems while producing high-value products in a constructive and innovative manner. [3].

It splits a big project into manageable smaller projects called "sprints," which is one to four weeks long, usually. The product owner, the development team, and the scrum master are the three positions that make up a Scrum team. The product owner addresses the most relevant things of the product backlog, an organized list of product design proposals, during the sprint planning meeting. The development team foresees the product requirements to be generated by the end of each iteration and then creates a sprint strategy to build an improvement in the product. The team works closely together during a sprint and holds regular standup meetings. At the end of the sprint, the team presents the stakeholders' newly produced project results and requests input during their sprint analysis. This improvement should be ready to be released and adhere to a pre-defined definition of completion. During the sprint retrospective, the team evaluates itself and prepares for how to progress in the next sprint. Across the cycle, the scrum master is a mentor as well as a coach, aiding others in successfully executing Scrum. [3]. Scrum is the most widely used agile methodology, with over 12 million professionals. [4]. Scrum is most often used by software companies, although it is often used by a wide range of other businesses.

2.3.2 CRISP-DM

Data mining, or the "usage of data pattern search techniques," is a general term for the method of finding knowledge in data," is also known as knowledge discovery in databases

(KDD)[5]. After its inception in the 1990s, CRISP-DM (Cross Industry Standard Process for Data Mining) has been consistently used for KDD projects, and more often, within a data science background. Business awareness, data comprehension, data management, modeling, assessment, and implementation are the six main iterative phases described by CRISP-DM, with some high-level, between points, iteration is permitted. [6]. The team usually goes through the steps as they see fit by using this framework. The team can reconnect to a previous step (for example, further data preparation) if required, and can identify milestones in general. In several ways, CRISP-DM can be compared to a waterfall paradigm for data mining. [6].

2.3.3 Kanban

Lean thinking focuses on value maximization and waste reduction in production processes. The idea was born in the early days of Japan's automobile industry. Kanban is one way to put lean ideas into practice. Although agile practices like Scrum have a definite phase for work on the structure, Kanban does not have one, nor does it have any clearly defined roles. Kanban, on the other hand, is a collection of ideas that includes analyzing the process, minimizing work-in-progress, monitoring and controlling flow, establishing consistent process policies, and enhancing collectively / integrating feedback loops [7]. Any process system that supports or promotes these Kanban values is acceptable to each team. Kanban has two major advantages they are, (1) It depicts work on a Kanban screen, with work objects streaming across columns of growing successful completion. (i.e., On a Kanban board, work objects are visually illustrated, enabling all members of the team to see the progress of any mission at any time), usually start with a column 'to do' and finish with a column 'complete', and (2) It seeks to reduce the amount of work in progress by setting WIP limits. WIP minimization allows for a lean approach (to reduce the time it takes to complete a job or a customer account) and also, since any time a new task starts, future tasks are re-prioritized, it allows for agility. Proponents of Kanban argue that it improves project awareness, software efficiency, team morale, coordination, and teamwork. [8]. According to a study of Kanban software development professionals, Kanban is simple to understand and implement. [9]. Kanban has many perceived

advantages, according to the respondents, including increasing job exposure, reducing work in progress, enhancing production flow, improving team cooperation, and promoting teamwork.

2.3.4 Team Data Science Process (TDSP)

TDSP was introduced by Microsoft as "an agile, iterative data science approach for quickly delivering predictive analytics technologies and intelligent applications." It has a CRISP-DM-like central project lifecycle of five iterative phases. [10],

1. Gain a better understanding of the market challenges by framing it, defining priorities, and identifying data sources.
2. Data Ingestion to Interpretation: ingest data to see how it responds to the question.
3. Train models and engineer features as part of the modeling process.
4. Deployment: Position the application in a development system.
5. Consumer acceptance: Allow customers and test the system and see how it suits their needs.

The method based on CRISP is combined with Agile concepts, which divide the bigger project in iterative phases, stories, and tasks. Each cycle begins with sprint preparation and is structured in a scrum-like manner. [11]. In addition to CRISP-DM, the system defines four distinct roles group manager, team manager, the leading project, and contributor. During each step of the project lifecycle, the process clearly describes its roles. [10]. Microsoft makes tools available on Github that project planning model, flow diagrams, documentation methods, model reporting, calculation of the performance model, and technical stack guidance are included within TDSP's wider project management concepts and processes [10].

2.3.5 SCRUM- DS

Jeroen Baijens, Remko Helms, and Deniz Iren have spent their time developing a full Scrum system for data science projects. Scrum-DS takes Scrum components, and the

CRISP-DM measures and applies them, evaluating the results by demonstrating them to data science team members [12]. Scrum-DS is a Scrum-based data science methodology which incorporates CRISP-DM factors. Scrum-DS has been improved by dividing the sprint into different sprints after the market and data comprehension phases. First, prepare data for sprint zero. The standard sprint for modeling to produce incremental value is the second stage. The problem they have run into is a lack of demonstration on a real-world experiment to see how the project process will do on a real-world data science project [12].

The authors explain that there is a need for a method in [18] [19] [20], that People, operation, and technologies, not just analytics, are becoming increasingly important in programs, and an approach would be beneficial. In terms of process methods, Agile-like methodologies have been found to have many benefits over conventional waterfall-like methodologies, including faster time-to-market, better teamwork, and early risk detection by iterative growth [20]

2.4 Challenges of Agile Frameworks for Data Science Projects

There are several possible challenges that teams may face when using Agile methodology in a data science project, as described by various authors in related studies.

2.4.1 Challenges of SCRUM in Data Science projects

As noted in [23] Unreliable task estimations are one of the major issues that arise by using SCRUM in the field of data science. In SCRUM, the team must normally settle on the task length for the sprint, so if the team cannot correctly predict the task duration, task execution would be troublesome in the sprint. Aside from the inaccurate estimates, another issue that has arisen in SCRUM is the problem of set sprint time. Because of this, the exploratory research results would be delayed by the data science team.

2.4.2 Challenges of Kanban in Data Science projects

Besides the many advantages of using Kanban in data science programs, [9] recognized possible difficulties such as a lack of organizational resources, a lack of experience, and a misconception

of core concepts. As described in the previous section, Kanban does not specify the functions or procedures of a project. Therefore, this freedom has itself become a challenge in implementing Kanban. Since it lacks processes and role definition, data science teams need to define its own roles and processes when using Kanban. The Kanban framework's lack of process specification highlights the need for additional practices to complement it [24]. As Ahmad et al. mentioned that “Kanban necessitates the introduction of current Agile techniques, which can be difficult, costly, and take time” [25].

2.4.3 Challenges of CRISP-DM in Data Science projects

CRISP-DM provides a comprehensive collection of instructions for completing data science tasks, and it works well with the Agile system to organize how the team investigates and iterates. However, these mechanisms do not consider team coordination, collaboration, or prioritization [26]. This is useful for determining what to do, but not for determining how to do it. Despite the idea of "looping back" to iterate the various CRISP-DM phases, the team has no clear mechanism for knowing when to iterate. As a result, several data science teams are opting for other Agile systems instead of CRISP-DM [27].

2.4.4 Challenges of TDSP in Data Science projects

The SCRUM methodology is being used by TDSP. Therefore, sprint challenges from SCRUM can be used in TDSP as well. And TDSP, like CRISP-DM, lacks a definition for when to iterate. As a result, while TDSP is beneficial in some ways, it does not address how a team can experiment and iterate. [28].

By recognizing the potential issues of various Agile Frameworks in the context of data science, these distinct approaches present benefits and drawbacks.

2.5 Critical Success Factors for Data Science Projects

With the growth of software development, most of the studies have applied Critical Success Factors (CSFs), because of various research areas. In the studied literature, there were precise definitions for their critical success factors. In this research, critical success factors are defined as factors which influence the success of data science projects in Agile methodology. Critical Success Factors (CSFs) were introduced in [32] as a tool to assist

managers in determining which data is most important for achieving organizational goals and objectives.

Most of the studies reveal that the article written by the authors in [2] has established twelve key performance factors affecting Agile software development projects. Those variables were divided into five categories: organizational, people, process, technical, and project factors. This study also examined the effects that influence the output of data science projects in an agile environment, along with the variables presented in [2].

Chow and Cao put their model to the test by sending a cross-sectional survey to Agile Alliance participants [2]. 6 of the 12 at least one dimension of success was associated with success factors, according to responses from 109 people: delivery approach, agile software engineering techniques, team capacity, project management process, team setting, and customer engagement. As the writers point out, there may be a shortage of non-Agile supporters, a poor response rate, and the comparatively inexperienced state of the Agile approach now. As the thesis by Chow and Cao was published more than a decade ago, this paper conducted a literature review that aimed at updating Chow and Cao's possible success factors and characteristics, as well as research into the data-scientific investigation in projects.

According to the literature, organizational, people, process, technical, and project aspects all play an important role in Agile-based ventures in the IT industry. These dimensions are expected to be critical success drivers in any project management process, not just Agile project management. According to Chow and Cao's model, there are variables for each of the above dimensions that are essential for the effectiveness of Agile software project management. Moreover, twelve factors for the success of Agile development projects in terms of consistency, scale, time, and expense were discovered [2].

Furthermore, the authors of [13] published a report that has consequences for sustainable social impact in organizations so they would be able to better appreciate the Critical Success Factors of agile software development programs. It has helped companies to implement techniques to enhance agile software development programs and cost savings,

resulting in increased software project viability and productivity. And, in addition to the factors introduced in Chow and Cao's model, the authors in [13] have identified factors under the dimension of performance expectancy in their research. The factors that influence agile software development programs are determined by performance expectancy, which includes expected usefulness. According to the authors' results in [14], there were 33 critical success drivers for implementing Big Data projects. Data, governance, process, objectives, teams, and tools are the six main categories of key performance factors defined by them. There are currently no accepted standards for Big Data projects, so it is beneficial to adopt a better method strategy. An advanced series of success factors have been introduced to help professionals to complete big data projects. There are thirty-three success factors in total, which are organized by the six main project characteristics [14] However, there were several areas of research that were not included in the study. There are no trials, for example, to compare the efficacy of various methodologies used by teams.

According to the authors in [15], data context, analytical context, team context, and organizational context are the four categories of efforts in data science that they have identified. And They used 14 characteristics to characterize the projects that were completed by various organizations. This has created a standardized framework for identifying various data science initiatives. The model has introduced two main attributes: (1) discovery standard about the concerns of the project, and (2) The emphasis of the project's technology, as a data science project, needs a lot of computation [15].

Adapted Agile analytics success factors and attributes were elaborated more in [16] by examining two analytics initiatives with varying levels of project success. The same larger financial institution conducted both the projects, but by separate project teams and in different departments. The authors identified six characteristics in these two ventures. Specifically, psychological safety inside the team, team potency, team happiness, team engagement, complexity requirements, and process uncertainty. And there were some

distinctions between the two programs. They have given a comparison of two projects in terms of the critical success factors outlined in the Chow and Cao model [2].

The authors of [14] defined multiple features of an enterprise with mature big data programs by the combination of recommendations from different works of literature. They are Data, Governance, Process, Objectives, Team, and Tools.

Moreover, these CSFs need to be prioritized, refined, or validated, and they must be rated in order of importance to the project's success.

2.6 Summary

The Success Factors from [2] have been considered in this research to carry out the analysis to identify the CSFs when managing data science projects within Agile methodology. These factors have been divided into four dimensions by Chow and Cao: organizational, people, method, and technical. In addition to these dimensions, there are success factors related to the project dimension. Authors have identified four attributes which are related to project dimension as Quality which focuses on delivering of good product or project as the outcome. And Time which means the project to be delivered on time. And Scope which focuses on achieving all the requirements and objectives. And Cost is about delivering the project within estimated time and effort.

In this research, all these consolidated factors have been used to analyze and measure the level of success of data science projects.

CHAPTER 3 – Methodology

3.1 Introduction

The purpose of this research is to identify the critical success factors for managing data science projects within Agile methodology in the IT industry. The professionals who worked on data science projects in the IT industry are the study's target audience. This chapter on methodology discusses the conceptual research context, theories proposed, data collection techniques, population and sampling, and the methodology used.

3.2 Conceptual Framework

The proposed conceptual framework for the research is presented in figure 3.1 below.

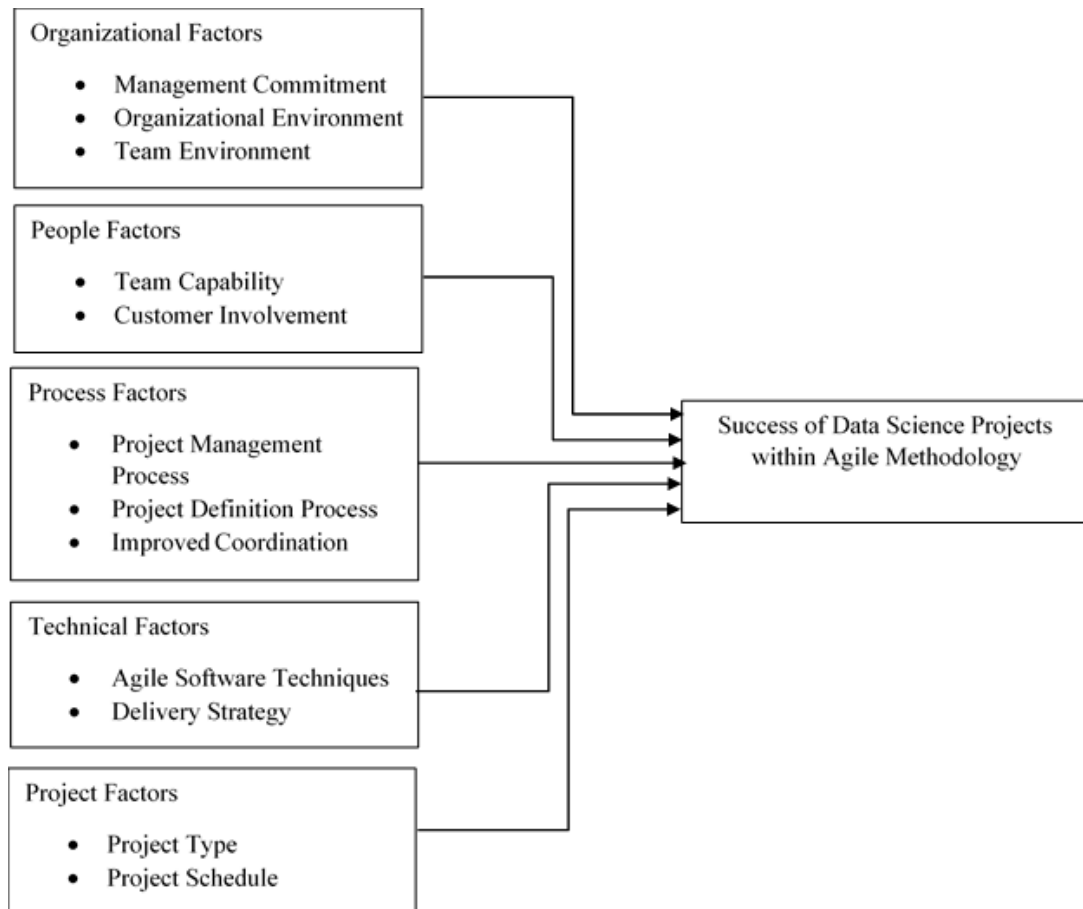


Figure 3. 1: Conceptual Framework

3.3 Variables on Relationships

The success of the Data Science projects within the Agile Methodology is the dependent variable of the study. Management Commitment, Organizational environment, Team environment, Team Capability, Customer Involvement, Project Management Process, Project Definition Process, Improved Coordination, Agile Software Engineering Techniques, Delivery Strategy, Project Type, and Project Schedule are the study's independent variables.

The organizational dimension, People dimension, Process dimension, technical dimension, and Project dimension are the five dimensions that these independent variables are divided into.

3.3 Hypothesis Development

Based on the above conceptual framework, the study presents a number of hypotheses that represent different relationships among the variables identified.

Let,

H_A – Alternate Hypothesis

H_0 – Null Hypothesis

3.3.1 Organizational Factors

Relationship between management commitment and the success of data science projects within Agile Methodology

According to the literature, strong management commitment is critical in Agile development projects; further, the commitment of the company's management for a project can drive substantial changes in the client's approval or resistance to a project or product. Following hypothesis, H1 was developed since literature has established management engagement as a significant success factor in general for the project, and for Agile project success when managing data science projects.

H1_A: The presence of strong management commitment has a relationship with the success of managing data science projects within Agile methodology.

H1₀: The presence of a strong management commitment has no relationship to the success of managing data science projects within Agile methodology.

Relationship between organizational environment and the success of data science projects within Agile Methodology

Following hypothesis, H2 was established because the previous literature reinforces the idea of an efficient Agile project implementation influenced by the organization's environment and culture.

H2_A: The presence of an agile and welcoming business atmosphere has a connection with the performance of data science operations.

H2₀: The presence of an agile and welcoming business atmosphere has no connection with the performance of data science operations.

Relationship between team environment and the success of data science projects within Agile Methodology

Data science teams must be able to effectively deliver the solution by sharing knowledge, experience, and information for successful project implementation and agile work. It was important to create a collaborative team atmosphere that would enable colleagues to share creativity and inspire new ideas. Since previous research has shown that an Agile-friendly team atmosphere is beneficial to the success of Agile projects, hypothesis H3 was created.

H3_A: The existence of an agile team environment is linked to the performance of the management of data science projects.

H3₀: The existence of an agile team environment has no linkage to the performance of the management of data science projects.

3.3.2 People Factors

Relationship between team capability and the success of data science projects within Agile Methodology

Competence plays an important role in project performance because a highly skilled team can produce a high-quality product quickly. Any project will fail if it does not have the right people on board. Following hypothesis, H4 has been established on the basis of previous literature, which promotes the team's competence effect on the progress of the project.

H4_A: Having a top-notch team has a relationship with the success of managing data science projects within Agile methodology.

H4₀: Having a top-notch team has no relationship with the success of managing data science projects within Agile methodology.

Relationship between customer involvement and the success of data science projects within Agile Methodology

To effectively execute any project, the need for client participation has become increasingly significant. As a result, it was critical to deciding if the project's clients had been identified. Hypothesis H5 was derived from past literature, indicating that customer involvement in software development projects and project success is positive.

H5_A: Having a high level of customer engagement has a relationship with the success of managing data science projects within Agile methodology.

H5₀: Having a high level of customer engagement has no relationship with the success of managing data science projects within Agile methodology.

3.3.3 Process Factors

Relationship between project management process and the success of data science projects within Agile Methodology

The success of software development programs is aided by good project management practices. Customers were more satisfied with project management activities in Agile-driven approaches because the project was more competitive. As a result, a literature-based hypothesis H6 was developed to support the Agile-focused project management techniques and the performance of software development data science management activities.

H6_A: Agile project management practice has a relationship with the success of managing data science projects.

H6₀: Agile project management practice has no relationship with the success of managing data science projects.

Relationship between project definition process and the success of data science projects within Agile Methodology

The project scope and objectives must be clearly defined and signed with the customer beforehand to ensure the success of Agile software development and data science projects. As a result, previous research suggests that a consistent and precise description of the whole project is critical for effective software project execution. Following hypothesis, H7 was established based on this claim.

H7_A: The process of defining a project in a methodical manner has a relationship to the success of managing data science projects within Agile methodology.

H7₀: The process of defining a project in a methodical manner has no relationship to the success of managing data science projects within Agile methodology.

Relationship between improved coordination and the success of data science projects within Agile Methodology

The procedure aided the team in keeping track of work, distributing the workload among team members, and ensuring proper coordination. According to the literature, better

teamwork has aided the team in staying on board with where they wanted and needed to be. It has also aided in the identification of the project's main areas of focus. Planning and execution became more apparent because of the increased teamwork. The following hypothesis, H8, was established because of this.

H8_A: The improved coordination has a relationship on the success of managing data science projects.

H8₀: The improved coordination has no relationship with the success of managing data science projects.

3.3.4 Technical Factors

Relationship between Agile software engineering techniques and the success of data science projects within Agile Methodology

Agile strategies included applying clear coding concepts, codebase sharing between all or most data engineers /programmers, adequate documentation, proper integration testing, project feature delivery, and technical training for the team, among other things. These beneficial activities contribute to the success of software development and data science programs. Due to the results indicated that Agile technology and the software project link to the success of the data science project, the H9 hypothesis has been created.

H9_A: Agile software techniques practice has a relationship with the success of managing data science projects.

H9₀: Agile software techniques practice has no relationship with the success of managing data science projects.

Relationship between correct delivery strategy and the success of data science projects within Agile Methodology

The gradual development of working software or prototypes for client assessment and optimization was emphasized by agile methods. Hypothesis H10 was developed based on

previous literature that indicated the impact of software product delivery strategy and data science project deliverables on project results.

H10_A: The proper implementation of a delivery strategy has a relationship with the success of managing data science projects within Agile methodology.

H10₀: The proper implementation of a delivery strategy has no relationship to the success of managing data science projects within Agile methodology.

3.3.5 Project Factors

Relationship between project type and the success of data science projects within Agile Methodology

Unlike conventional methodologies, Agile approaches also use a dynamic project scope instead of a defined scope. Projects with a flexible nature and emergent criteria contributed to the overall performance of Agile projects. As a result, hypothesis H11 was created.

H11_A: Restricting only dynamic scope projects with emerging requirements has a relationship with the success of managing data science projects within Agile methodology.

H11₀: Restricting only dynamic scope projects with emerging requirements has no relationship on the success of managing data science projects within Agile methodology.

Relationship between dynamic schedule and the success of data science projects within Agile Methodology

Since the project scope is variable in the Agile approach, the variable and complex project program is important, too. As a result, follow complex plans that can accommodate changes in scope and operation, just as you would for any Agile system. As a result, Hypothesis H12 was developed in relation to the effect of complex schedules on the output of software development and data science projects derived from previous research findings.

H12_A: Projects with flexible schedules have a relationship with the success of managing data science projects within Agile methodology.

H12₀: Projects with flexible schedules have no relationship with the success of managing data science projects within Agile methodology.

Relationship between the Agile method practiced in the organization and the success of data science projects within Agile Methodology.

Practicing Agile methodology is important in the organization to drive the projects in a methodical manner, therefore below specified H13 was built.

H13_A: The Agile methodology practiced in the organization has a relationship with the success of managing data science projects within Agile methodology.

H13₀: The Agile methodology practiced in the organization has no relationship to the success of managing data science projects within Agile methodology.

3.4 Questionnaire Instrument Development

A questionnaire tool (Annex A) was developed to evaluate the variables mentioned previously as well as to capture the understanding of the participants of the crucial success factors for managing data science activities in IT projects using Agile methodology.

Tables 3.1, 3.2, and 3.3 illustrate a summary of the measures used.

Dimension	Variable	No. of Items	Scale
Organizational	Management Commitment	2	Five Point Likert
	Organizational Environment	2	Five Point Likert
	Team Environment	2	Five Point Likert
People	Team Capability	2	Five Point Likert
	Customer Involvement	2	Five Point Likert
Process	Project Management Process	2	Five Point Likert
	Project Definition Process	2	Five Point Likert

	Improved Coordination	1	Five Point Likert
Technical	Agile Software Techniques	4	Five Point Likert
	Delivery Strategy	2	Five Point Likert
Project	Project Type	1	Five Point Likert
	Project Schedule	2	Five Point Likert

Table 3. 1: Summary of measurements in independent variables

Variable	No. of Items	Scale
Success of the Data Science projects within the Agile Methodology	3	Five Point Likert

Table 3. 2: Summary of measurements independent variable

Demographic Factor	No. of Items	Scale
Age Group	1	Ratio
Gender	1	Ratio
Designation	1	Nominal
Level of Experience	1	Ratio
Agile methodologies practiced by data science professionals	1	Nominal

Table 3. 3: Summary of measurements in Demographic factors

3.5 Sample Selection

The study's total population represents all data science practitioners in the IT industry who works on a project at least that employs the Agile methodology as a project management tool. The Cochran formula is used to determine the sample size for the research. [22] Reasonable sample size has been calculated using a Z score, which is a numerical measurement used in statistics. The Z score can be determined by considering the normal

deviation set at 95% confidence level (1.96) (Z), and the percentage of responses at 20% (P), and the confidence interval (0.05) (C).

$$n = \frac{Z^2(P)(1 - P)}{C^2}$$
$$n = \frac{1.96^2(0.2)(1 - 0.2)}{0.05^2}$$
$$n = 245.83$$

The sample size is obtained as 246 according to the above equation.

3.6 Method of Data Collection

The required data for the study will be collected through an online questionnaire. The questionnaire is designed as per the previous literature items adopted by the researchers and will be distributed among the data science professionals in the IT industry who work in an Agile project management context.

3.7 Summary

Data was gathered using an online survey that was sent to the target population of Project Managers, Data Scientists, Data Analysts, Data Engineers, Software Architects, and Business Analysts who work on data science projects. The survey was divided into four parts. The first section dealt with demographic information. Sections 2 and 3 dealt with success factors and perceptions of success, respectively. Section 4 was set aside for supplementary remarks. A 5-point Likert scale was used to assess the value of performance factors and perceptions of success. The survey was prolonged for three months, yielding 240 complete answers.

CHAPTER 4 – Analysis and Interpretation

4.1 Introduction

This chapter has been analyzed and interpreted the data gathered during the data collection process in such a way that facilitating the provision of a basic understanding to the reader to ease the reading of the next few chapters. Participant's age group, gender, designation, working experience were primarily analyzed to get a basic idea about the selected sample. For that purpose, graphical tools were used to present the data gathered from the study. This chapter is dedicated to evaluating the data to test the hypothesis presented in Chapter 3. For this purpose, many statistical studies have been performed to accept or reject the hypothesis of nullity and to develop further information to explain the presence and connection of the study variables.

4.2 Reliability Test

The reliability test was carried out to determine the instrument's quality, which was then used to determine the degree to which the metrics were free of bias. The study used reliability testing to see whether the intervention was internal and whether the issues raised in each area were mutually supportive. For this purpose, Cronbach's Alpha Coefficient, used for scalable multipoint artifacts [33].

4.2.1 Preliminary Survey

A preliminary survey of 60 participants was conducted. The calculations were done question by question. All 27 items were checked for Cronbach's Alpha Coefficient using SPSS version 1.0.0-1447, and the results are shown in tables 4.1 and 4.2. Cronbach's Alpha Coefficient values above 0.7 are generally considered suitable [33].

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
------------------	--	------------

.850	.856	27
------	------	----

Table 4. 1: Cronbach's Alpha Coefficient for Preliminary Survey

	Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Q05	99.21	95.236	.151	.	.853
Q06	99.05	93.253	.264	.	.849
Q07	98.95	90.993	.492	.	.842
Q08	99.02	91.497	.431	.	.844
Q09	98.79	90.716	.484	.	.842
Q10	98.59	88.961	.577	.	.839
Q11	98.75	92.780	.424	.	.844
Q12	98.75	92.780	.424	.	.844
Q13	98.89	91.540	.363	.	.846
Q14	98.68	91.807	.473	.	.843
Q16	99.23	88.466	.483	.	.842
Q17	99.05	92.176	.403	.	.845
Q18	99.18	92.919	.476	.	.844
Q19	98.91	89.266	.595	.	.839
Q20	99.23	87.892	.552	.	.839
Q21	99.09	89.050	.562	.	.839
Q22	99.41	88.781	.472	.	.842
Q25	98.86	88.661	.588	.	.838
Q26	99.11	91.827	.343	.	.847
Q27	99.29	95.272	.158	.	.852
Q28	99.05	98.239	-.020	.	.857
Q29	99.80	98.311	-.043	.	.862
Q23	99.09	90.593	.490	.	.842
Q24	98.95	91.208	.415	.	.844
Q30	99.02	93.650	.334	.	.847
Q31	99.34	94.252	.239	.	.850
Q32	99.00	90.009	.477	.	.842

Table 4. 2: Item-wise Cronbach's Alpha Coefficients

The overall Cronbach's alpha calculated for the above 27 items is 0.850 which is acceptable reliability for the instruments used in the study. According to the item total statistics, all the items have acceptable reliabilities as per table 4.2. Since all Cronbach's alpha coefficients nearly vary around 0.850 none of the items was removed from the questionnaire.

4.2.1 Research Survey

In addition, 240 participants examined the reliability of the analysis for the entire sample. In the research survey, there were 27 items that were included for reliability testing. Cronbach's Alpha Coefficient was tested for the item level as well as for the dimension level of the variables. The below tables will illustrate the relevant Cronbach's Alpha for each dimension.

For 240 respondents, the Cronbach's Alpha Coefficient is 0.831. All of the dimensions and items have passed the reliability test with a Cronbach's Alpha Coefficient of greater than 0.7.

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.831	.838	27

Table 4. 3: Cronbach's Alpha Coefficient for Research Survey

Reliability test for Dependent variable for 240 respondents.

Cronbach's Alpha Coefficient value for the dependent variable which is the Success of the Data Science projects within the Agile Methodology is 0.751 which has acceptable reliability.

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.751	.761	3

Table 4. 4: Cronbach's Alpha Coefficient for Dependent Variable

Reliability test for independent variables for 240 respondents.

The reliability test for independent variables was done in dimension wise of Organizational, People, Process, Technical and Project, except for the single item variables.

Dimension	Variable	No. of Items	Cronbach's Alpha Coefficient Value
Process	Improved Coordination	1	N/A
Project	Project Type	1	N/A

Table 4. 5: Single Item Variables

4.3 Descriptive Statistics Analysis

The sample was drawn from data science practitioners employed in software companies in the IT industry. The entire survey was conducted online using an online questionnaire.

4.3.1 Composition of the sample according to Age Group

Age Group	Frequency	Percentage (%)
20 -30 years	156	65
31-40 years	68	28
41-50 years	4	2
Above 51 years	12	5
Total	240	100

Table 4. 6: Age Distribution

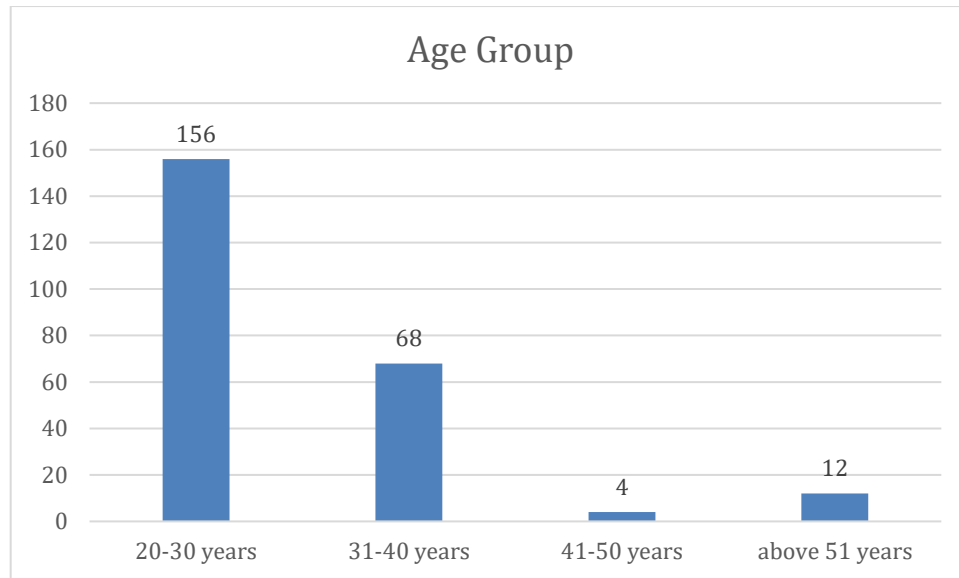


Figure 4. 1: Frequency Distribution of Age Group

Ages were categorized into four levels extending from 20 years of age. Most respondents were in the age level of 20-30 where the percentage is 65%. Age levels of 31-40 are at the moderate level with the representing of participants of 28%. Twelve participants represent the age level of above 51 and only four participants have responded to the survey where the age level is 41-50.

4.3.2 Composition of the sample according to Gender

Gender	Frequency	Percentage (%)
Male	172	72
Female	68	28
Total	240	100

Table 4. 7: Gender Distribution

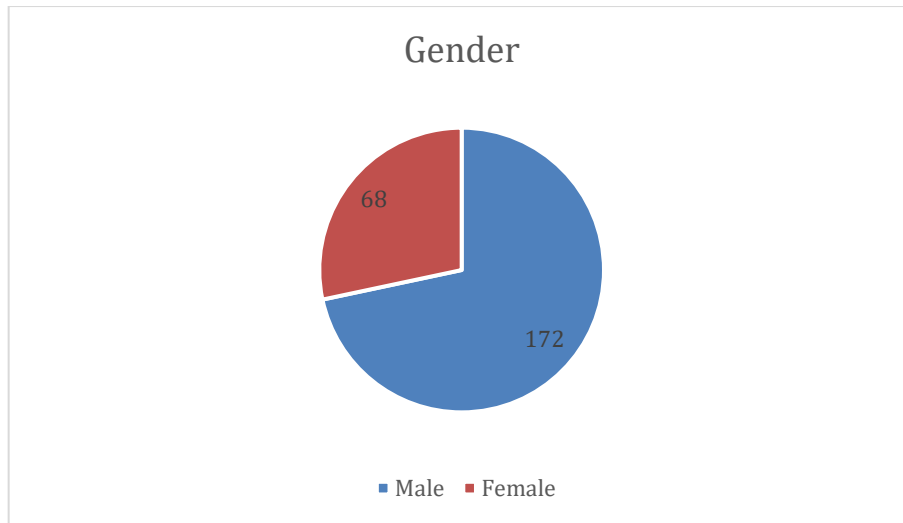


Figure 4. 2: Frequency Distribution of Gender

Male and female respondents account for 72 percent and 28 percent of the 240 respondents, respectively, as shown in table 4.7 and figure 4.2

4.3.3 Composition of the sample according to Designation

Designation	Frequency	Percentage (%)
Data Scientist	28	11.67
Data Engineer	68	28.33
Data Analyst	28	11.67
Architect	16	6.67
Big Data Engineer	8	3.33
Project Manager	28	11.67
Business Analyst	32	13.33
Software Engineer	24	10
Database Administrator	4	1.67
Marketing Researcher	4	1.67
Total	240	100

Table 4. 8: Distribution of Job Designation

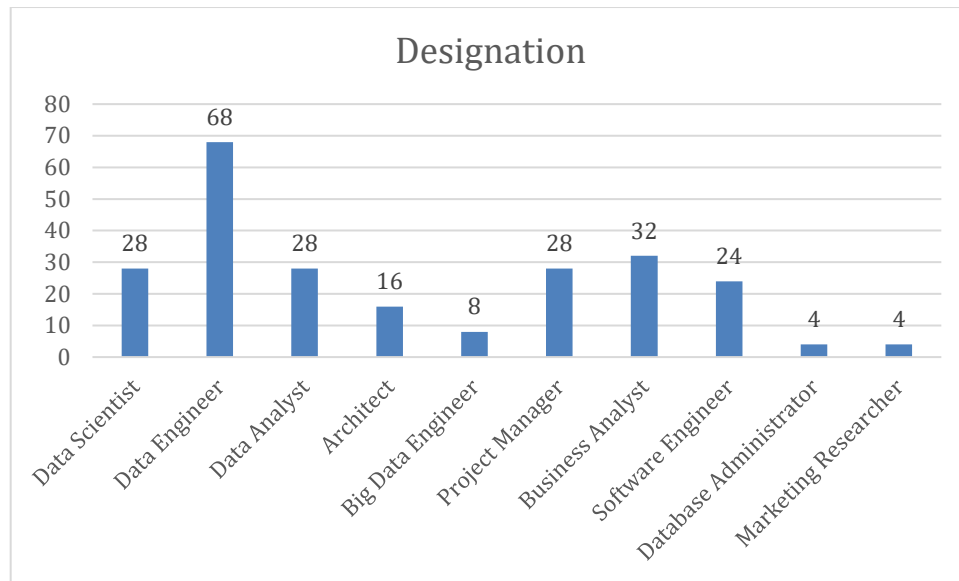


Figure 4. 3: Distribution of Job Designation

Designations were mainly categorized according to the jobs related to data science domain. As shown in above table 3 and figure 5, 28.33% of the respondents are from Data Engineer job category. And 13.33% of the respondents are Business Analysts. From Data Scientist, Data Analyst and Project Manager job categories there is 28% each.

4.3.4 Composition of the sample according to Working Experience.

Level of Experience	Frequency	Percentage (%)
Lesser than 1 year	24	10
1-3 years	84	35
3-5 years	68	28.33
More than 5 years	64	26.67
Total	240	100

Table 4. 9: Distribution of Working Experience

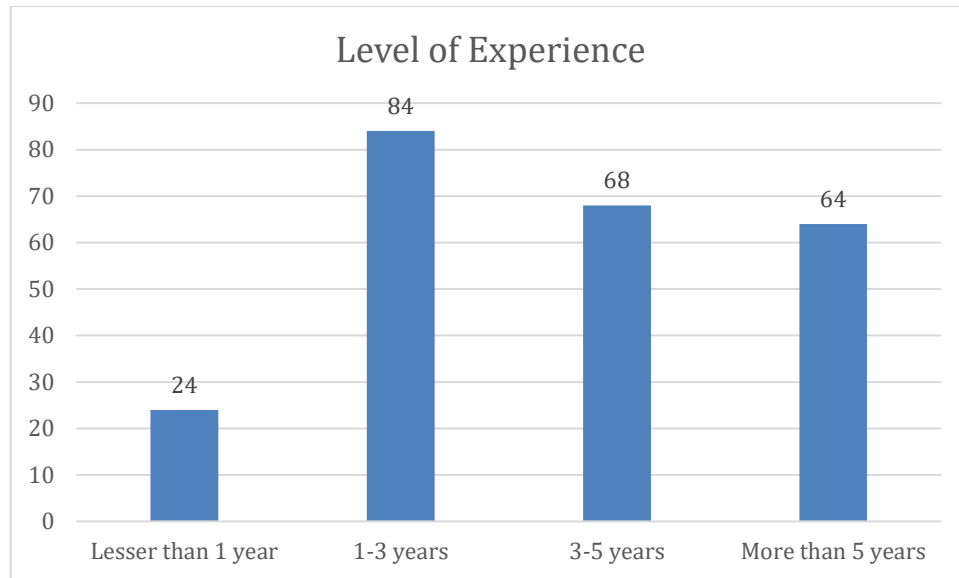


Figure 4. 4: Distribution of Working Experience

According to the results, 10% of participants have less than one year of work experience in their profession and 35% have between one and three years of experience. In that sense, 45% of the sample represents young workers who have experience of fewer than 3 years. Apart from that, 28.33% have an experience level between 3-5 while 26.67% represent more than 5 years of experience.

4.3.5 Composition of the sample according to the used Agile approach.

Agile Approach	Frequency	Percentage (%)
SCRUM	144	60
Kanban	16	6.67
SCRUM, Kanban	48	20
Kanban, SCRUM- DS	8	3.33
CRISP-DM, TDSP	4	1.67
SCRUM, SCRUM- DS	8	3.33
TDSP	4	1.67
SCRUM- DS	8	3.33
Total	240	100

Table 4. 10: Distribution of Agile methodologies used.

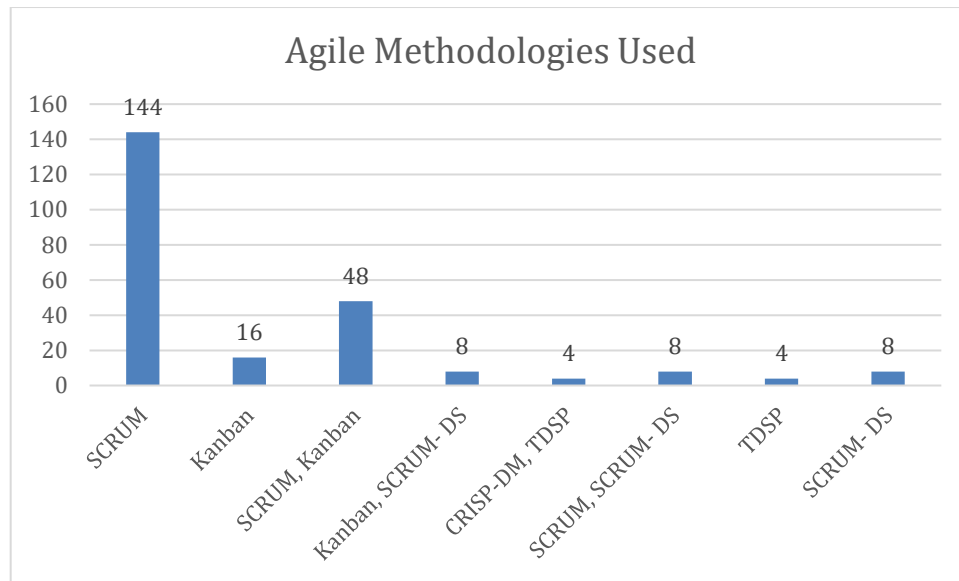


Figure 4. 5: Distribution of Agile methodologies used.

According to the findings, it is observed that 60% of participants are using only SCRUM as the Agile methodology in data science in their organizations. 20% of respondents are using both SCRUM and Kanban were only 6.67% of the participants are using only Kanban in their projects. SCRUM- DS, CRISP-DM, and TDSP Agile methodologies usage is 13.33% which is less compared to SCRUM and Kanban.

In this study, the critical success factor in the management of data science activities in IT projects using the agile approach was used in a five-point Likert scale from "strongly disagree" (valued as "1") to "Strongly agree" (valued as "5"). The tables of Appendix B present the summary information for each variables' descriptive statistics.

4.4 Inferential Statistics – Inter Item Correlation Analysis

Inferential statistics are used to check the inter-item correlation given with the Pearson Correlation matrix. This test was carried out with each variable to see whether the inter-items of each variable were related.

Organizational Environment, Team Environment, Customer Involvement, Project Definition Process, Agile Software Engineering Techniques, Project Schedule in the respective independent variable, items had positively correlated each other.

Inter Item Correlation for Organizational Environment

		Q07	Q08
Q07	Pearson Correlation	1	.661**
	Sig. (2-tailed)		.000
	N	240	240
Q08	Pearson Correlation	.661**	1
	Sig. (2-tailed)	.000	
	N	240	240

** . Correlation is significant at the 0.01 level (2-tailed).

Table 4. 11: Inter Item Correlation for Organizational Environment

Inter Item Correlation for Team Environment

		Q09	Q10
Q09	Pearson Correlation	1	.289**
	Sig. (2-tailed)		.000
	N	240	240
Q10	Pearson Correlation	.289**	1
	Sig. (2-tailed)	.000	
	N	240	240

** . Correlation is significant at the 0.01 level (2-tailed).

Table 4. 12: Inter Item Correlation for Team Environment

Inter Item Correlation for Customer Involvement

		Q13	Q14
Q13	Pearson Correlation	1	.540**
	Sig. (2-tailed)		.000
	N	240	240
Q14	Pearson Correlation	.540**	1

Sig. (2-tailed)	.000	
N	240	240

** . Correlation is significant at the 0.01 level (2-tailed).

Table 4. 13: Inter Item Correlation for Customer Involvement

Inter Item Correlation for Project Definition Process

		Q18	Q19
Q18	Pearson Correlation	1	.318**
	Sig. (2-tailed)		.000
	N	240	240
Q19	Pearson Correlation	.318**	1
	Sig. (2-tailed)	.000	
	N	240	240

** . Correlation is significant at the 0.01 level (2-tailed).

Table 4. 14: Inter Item Correlation for Project Definition Process

Inter Item Correlation for Agile Software Engineering Techniques

		Q21	Q22	Q23	Q24
Q21	Pearson Correlation	1	.588**	.497**	.428**
	Sig. (2-tailed)		.000	.000	.000
	N	240	240	240	240
Q22	Pearson Correlation	.588**	1	.366**	.276**
	Sig. (2-tailed)	.000		.000	.000
	N	240	240	240	240
Q23	Pearson Correlation	.497**	.366**	1	.347**
	Sig. (2-tailed)	.000	.000		.000
	N	240	240	240	240
Q24	Pearson Correlation	.428**	.276**	.347**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	240	240	240	240

** . Correlation is significant at the 0.01 level (2-tailed).

Table 4. 15: Inter Item Correlation for Agile Software Engineering Techniques

Inter Item Correlation for Project Schedule

		Q28	Q29
Q28	Pearson Correlation	1	.710**
	Sig. (2-tailed)		.000
	N	240	240
Q29	Pearson Correlation	.710**	1
	Sig. (2-tailed)	.000	
	N	240	240

** . Correlation is significant at the 0.01 level (2-tailed).

Table 4. 16: Inter Item Correlation for Project Schedule

Items were positively correlated within respective variables, inter-item correlation checks for dependent variables. The items within the Organizational Environment, Customer Involvement, and Project Schedule showed a high correlation according to Tables 4.11, 4.13, and 4.16, respectively. Items in Team Environment, Project Definition process, and Agile Software Engineering Techniques showed a moderate correlation according to Tables 4.12, 4.14, and 4.15, respectively.

Improved Coordination and Project Type each had one item in the current report. It was impossible to find the associated values for these two variables due to a single component.

Inter Item Correlation for Success of the Data Science projects within the Agile Methodology

		Q30	Q31	Q32
Q30	Pearson Correlation	1	.699**	.423**
	Sig. (2-tailed)		.000	.000
	N	240	240	240
Q31	Pearson Correlation	.699**	1	.420**
	Sig. (2-tailed)	.000		.000
	N	240	240	240
Q32	Pearson Correlation	.423**	.420**	1
	Sig. (2-tailed)	.000	.000	
	N	240	240	240

** . Correlation is significant at the 0.01 level (2-tailed).

Table 4. 17: Inter Item Correlation for Success of the Data Science projects within the Agile Methodology.

According to Table 4.17, the items in the independent variable showed a high correlation within the respective variable.

4.5 Hypothesis Testing – Pearson’s Correlation Analysis

This is the method for determining the relationship between two variables. The relationship between two variables is explained using correlation analysis. The correlation coefficient shows the intensity of the relationship, and the direction of the relationship can be calculated using a formula that considers both sets of figures [33]. To evaluate the study's reported hypothesis, the relationship between independent variables and depending on variables was analyzed using correlation analysis.

Inferential statistics are used to test the relationship between dependent and independent variables from 1 to 12. The Pearson Correlation Matrix was used to evaluate the importance of each variable using standard averaging.

Hypothesis 1:

Management Commitment and Success of the Data Science projects within the Agile Methodology

		Management Commitment	Success
Management Commitment	Pearson Correlation	1	-.120
	Sig. (2-tailed)		.063
	N	240	240
Success	Pearson Correlation	-.120	1
	Sig. (2-tailed)	.063	
	N	240	240

Table 4. 18: Pearson’s Correlation- Management Commitment and Success of the Data Science projects within the Agile Methodology.

As shown in Table 4.18 the relationship between Management Commitment and Success of the Data Science projects within the Agile Methodology is not significant. The null hypothesis (H_{10}) is therefore justified, and the alternative (H_{1A}) is dismissed. Therefore,

there is no linear relationship. Hence, the existence of a strong management commitment has no impact on the success of managing data science projects within Agile methodology.

Hypothesis 2:

Organizational Environment and Success of the Data Science projects within the Agile Methodology

		Organizational Environment	Success
Organizational Environment	Pearson Correlation	1	-.046
	Sig. (2-tailed)		.474
	N	240	240
Success	Pearson Correlation	-.046	1
	Sig. (2-tailed)	.474	
	N	240	240

Table 4. 19: Pearson's Correlation- Organizational Environment and Success of the Data Science projects within the Agile Methodology.

As shown in Table 4.19 the relationship between the Organizational Environment and Success of the Data Science projects within the Agile Methodology is not significant. Therefore, the null hypothesis (H2₀) is justified, and the alternate hypothesis (H2_A) is dismissed. Therefore, there is no linear relationship. Hence, the presence of Agile friendly organizational environment has no relationship on the success of managing data science projects.

Hypothesis 3:

Team Environment and Success of the Data Science projects within the Agile Methodology

		Team Environment	Success
Team Environment	Pearson Correlation	1	.189**
	Sig. (2-tailed)		.003
	N	240	240
Success	Pearson Correlation	.189**	1

	Sig. (2-tailed)	.003	
	N	240	240

** . Correlation is significant at the 0.01 level (2-tailed).

Table 4. 20: Pearson's Correlation- Team Environment and Success of the Data Science projects within the Agile Methodology.

As shown in Table 4.20 the relationship between Team Environment and Success of the Data Science projects within the Agile Methodology is significantly positively correlated. Therefore, the null hypothesis (H3₀) is dismissed, and the alternate hypothesis (H3_A) is justified. Therefore, there is a linear relationship. Hence, the existence of an Agile-friendly project team environment has a relationship on the success of managing data science projects.

Hypothesis 4:

Team Capability and Success of the Data Science projects within the Agile Methodology

		Team Capability	Success
Team Capability	Pearson Correlation	1	.171**
	Sig. (2-tailed)		.008
	N	240	240
Success	Pearson Correlation	.171**	1
	Sig. (2-tailed)	.008	
	N	240	240

** . Correlation is significant at the 0.01 level (2-tailed).

Table 4. 21: Pearson's Correlation- Team Capability and Success of the Data Science projects within the Agile Methodology.

As shown in Table 4.21 the relationship between Team Capability and Success of the Data Science projects within the Agile Methodology is significantly positively correlated. Therefore, the null hypothesis (H4₀) is dismissed, and the alternate hypothesis (H4_A) is

justified. Therefore, there is a linear relationship. Hence, a team of high competence has a relationship with the success of managing data science projects within Agile methodology.

Hypothesis 5:

Customer Involvement and Success of the Data Science projects within the Agile Methodology

		Customer Involvement	Success
Customer Involvement	Pearson Correlation	1	.134*
	Sig. (2-tailed)		.039
	N	240	240
Success	Pearson Correlation	.134*	1
	Sig. (2-tailed)	.039	
	N	240	240

*. Correlation is significant at the 0.05 level (2-tailed).

Table 4. 22: Pearson's Correlation- Customer Involvement and Success of the Data Science projects within the Agile Methodology.

As shown in Table 4.22 the relationship between Customer Involvement and Success of the Data Science projects within the Agile Methodology is substantially strongly correlated. Therefore, the null hypothesis (H5₀) is dismissed, and the alternate hypothesis (H5_A) is justified. Therefore, there is a linear relationship. Hence, having a strong customer involvement has a relationship with the success of managing data science projects within Agile methodology.

Hypothesis 6:

Project Management Process and Success of the Data Science projects within the Agile Methodology

		Project Management Process	Success
Project Management Process	Pearson Correlation	1	.120
	Sig. (2-tailed)		.063
	N	240	240

Success	Pearson Correlation	.120	1
	Sig. (2-tailed)	.063	
	N	240	240

Table 4. 23: Pearson's Correlation- Project Management Process and Success of the Data Science projects within the Agile Methodology.

As shown in Table 4.23 the relationship between the Project Management Process and the Success of the Data Science projects within the Agile Methodology is not significant. Therefore, the null hypothesis (H_{0}) is justified, and the alternate hypothesis (H_{A}) is dismissed. Therefore, there is no linear relationship. Hence, the practice of the Agile project management process has no relationship with the success of managing data science projects.

Hypothesis 7:

Project Definition Process and Success of the Data Science projects within the Agile Methodology

		Project Definition Process	Success
Project Definition Process	Pearson Correlation	1	.191**
	Sig. (2-tailed)		.003
	N	240	240
Success	Pearson Correlation	.191**	1
	Sig. (2-tailed)	.003	
	N	240	240

** . Correlation is significant at the 0.01 level (2-tailed).

Table 4. 24: Pearson's Correlation- Project Definition Process and Success of the Data Science projects within the Agile Methodology.

As shown in Table 4.24 the relationship between the Project Definition Process and the Success of the Data Science projects within the Agile Methodology is significantly positively correlated. Therefore, the null hypothesis (H_{0}) is dismissed, and the alternate hypothesis (H_{A}) is justified. Therefore, there is a linear relationship. Hence, the practice

of the systematical project definition process has a relationship with the success of managing data science projects within Agile methodology.

Hypothesis 8:

Improved Coordination and Success of the Data Science projects within the Agile Methodology

		Improved Coordination	Success
Improved Coordination	Pearson Correlation	1	.043
	Sig. (2-tailed)		.510
	N	240	240
Success	Pearson Correlation	.043	1
	Sig. (2-tailed)	.510	
	N	240	240

Table 4. 25: Pearson's Correlation- Improved Coordination and Success of the Data Science projects within the Agile Methodology.

As shown in Table 4.25 the relationship between Improved Coordination and Success of the Data Science projects within the Agile Methodology is not significant. Therefore, the null hypothesis (H_{8_0}) is justified, and the alternate hypothesis (H_{8_A}) is dismissed. Therefore, there is no linear relationship. Hence, the improved coordination has no relationship with the success of managing data science projects.

Hypothesis 9:

Agile Software Engineering Techniques and Success of the Data Science projects within the Agile Methodology

		Agile Software Engineering Techniques	Success
Agile Software Engineering Techniques	Pearson Correlation	1	.357**
	Sig. (2-tailed)		.000
	N	240	240
Success	Pearson Correlation	.357**	1
	Sig. (2-tailed)	.000	
	N	240	240

** . Correlation is significant at the 0.01 level (2-tailed).

Table 4. 26: Pearson's Correlation- Agile Software Engineering Techniques and Success of the Data Science projects within the Agile Methodology.

As shown in Table 4.26 the relationship between Agile Software Engineering Techniques and Success of the Data Science projects within the Agile Methodology is significantly positively correlated. Therefore, the null hypothesis (H9₀) is dismissed, and the alternate hypothesis (H9_A) is justified. Therefore, there is a linear relationship. Hence, the practice of Agile software techniques has a relationship on the success of managing data science projects.

Hypothesis 10:

Delivery Strategy and Success of the Data Science projects within the Agile Methodology

		Delivery Strategy	Success
Delivery Strategy	Pearson Correlation	1	.229**
	Sig. (2-tailed)		.000
	N	240	240
Success	Pearson Correlation	.229**	1
	Sig. (2-tailed)	.000	
	N	240	240

** . Correlation is significant at the 0.01 level (2-tailed).

Table 4. 27: Pearson's Correlation- Delivery Strategy and Success of the Data Science projects within the Agile Methodology.

As shown in Table 4.27 the relationship between Delivery Strategy and Success of the Data Science projects within the Agile Methodology is significantly positively correlated. Therefore, the null hypothesis (H10₀) is dismissed, and the alternate hypothesis (H10_A) is justified. Therefore, there is a linear relationship. Hence, the proper implementation of a delivery strategy has a relationship with the success of managing data science projects within Agile methodology.

Hypothesis 11:

Project Type and Success of the Data Science projects within the Agile Methodology

		Project Type	Success
Project Type	Pearson Correlation	1	-.029
	Sig. (2-tailed)		.653
	N	240	240
Success	Pearson Correlation	-.029	1
	Sig. (2-tailed)	.653	
	N	240	240

Table 4. 28: Pearson's Correlation- Project Type and Success of the Data Science projects within the Agile Methodology.

As shown in Table 4.28 the relationship between Project Type and Success of the Data Science projects within the Agile Methodology is not significant. Therefore, the null hypothesis ($H11_0$) is justified, and the alternate hypothesis ($H11_A$) is dismissed. Therefore, there is no linear relationship. Hence, restricting only dynamic scope projects with emerging requirements has no relationship on the success of managing data science projects within Agile methodology.

Hypothesis 12:

Project Schedule and Success of the Data Science projects within the Agile Methodology

		Project Schedule	Success
Project Schedule	Pearson Correlation	1	.278**
	Sig. (2-tailed)		.000
	N	240	240
Success	Pearson Correlation	.278**	1
	Sig. (2-tailed)	.000	
	N	240	240

** . Correlation is significant at the 0.01 level (2-tailed).

Table 4. 29: Pearson's Correlation- Project Schedule and Success of the Data Science projects within the Agile Methodology.

As shown in Table 4.29 the relationship between Project Schedule and Success of the Data Science projects within the Agile Methodology is significantly positively correlated. Therefore, the null hypothesis (H_{12_0}) is dismissed, and the alternate hypothesis (H_{12_A}) is justified. Therefore, there is a linear relationship. Hence, projects with dynamic schedules have a relationship on the success of managing data science projects within Agile methodology.

4.6 ANOVA Testing

ANOVA is suitable to test hypothesis 13 since there are more than two classes of Agile methods practiced in organizations. Therefore, a one-way ANOVA test was performed to test hypothesis 13.

Hypothesis 13:

ANOVA					
Success of the Data Science projects within the Agile Methodology					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.859	7	.408	1.047	.399
Within Groups	90.472	232	.390		
Total	93.331	239			

Table 4. 30: One-Way ANOVA- Agile methodologies used.

The F values of 1.047 are not important at level 0.399 as shown in Table 4.30. The null hypothesis (H_{13_0}) is thus supported, and the alternative (H_{13_A}) hypothesis is rejected. Therefore, there is no connection between the Agile methodology practiced in the organization and success of the Data Science projects within the Agile Methodology.

4.8 Discussion on Results

Team Environment, Team Capability, Customer Involvement, Project Definition Process, Agile Software Engineering Techniques, and Project Schedule showed substantial

correlations with the success of Data Science projects under the Agile Methodology, according to Pearson's Correlations. Team Environment was positively correlated to the success of data science projects indicating that the existence of Agile friendly project team environment with committed and motivated team members always drives the data science projects to success.

Team Capability was positively correlated to the success of data science projects implying that having a high caliber team is an advantage for the organization to achieve the success of its projects. In other words, a team with high technical competence and expertise in subject matter, as well as competence in problem-solving, will drive to the success of the project while providing the appropriate technical training through the project to the team members.

Customer Involvement was positively correlated to the success of the data science project indicating that having a strong customer involvement in the project is driving its success, as it will strengthen the relationship between the customer and the team. And having a good customer commitment is always supportive to the team with regards to understanding the business requirements and provide a good solution.

Project Definition Process was positively correlated to the success of data science projects showing that practicing of methodical project definition process is needed for a data science project to make it a success. This emphasis on the well-defined scope and objectives of the project is essential and the project should follow a well-developed Agile-oriented requirement process to achieve success.

The adoption of Agile Software Engineering Techniques in data science projects by data science professionals with proper Agile-focused configuration management processes of using effective management of the version control or source code was positively associated with the performance of data science projects. And the project will use a well-defined coding standard from the start, as well as appropriate platforms and resources for Agile practice.

Delivery Strategy was positively correlated to the success of data science projects implying that execution of correct delivery strategy is critical for a project to become a success in the data science domain. The delivery strategy consisting of delivering a working software within a short period of time as well as delivering the most important features first which adhere to the Agile manifesto.

The relationship between the Agile methodology used by different project teams and the success of the project was tested by using One- Way ANOVA test, and the Agile methodology used was not significant. That means there is no relationship between the Agile methodology used and the success of data science projects. The analysis shows that there is no relationship to the success of the data science project either the project used SCRUM, Kanban, SCRUM-DS, CRISP-DM, TDSP.

4.9 Summary

In the present study, test results appear to be compatible with previous research. Most of the previous research has found that Management Commitment, Project Management Process, Agile Software Engineering Techniques, Delivery Strategy, Project Type, and Project Schedule are critical success factors that sign indicating the success of IT projects within Agile methodology.

As per the results obtained from the study, Team Environment, Team Capability, Customer Involvement, Project Definition Process, Agile Software Engineering Techniques, and Project Schedule showed a significance with the success of Data Science projects under the Agile Methodology, according to Pearson's Correlations.

CHAPTER 5 – Conclusion and Recommendation

5.1 Introduction

This chapter contains a description of the entire analysis as well as the research's conclusion based on the results mentioned in the previous chapter. In addition, this chapter covers the study's recommendations, limitations, and suggestions for potential research topics.

5.2 Research Implications

This study evaluates the critical success factors of managing data science projects within Agile methodology. As per the results obtained from the research instrument, Team Environment, Team Capability, Customer Involvement, Project Definition Process, Agile Software Engineering Techniques, and Project Schedule are the most significant factors to consider when managing data science projects within Agile methodology.

Team Environment is one of the important factors that the organization should take into consideration. There are few aspects that must be focused on when considering the team environment. The location of the team is one of the aspects to look at because the whole data science team should be in the same office space where they share the same workspace. Even though the members of the team are working remotely due to some reasons like a pandemic situation, organizations must facilitate the members of the same team with proper communication media and effective measures to practice and collaborate on project performance.

Team Capability is one of the important factors for the success of data science projects within Agile methodology. Therefore, data science teams should focus on the appropriate diversity of skills and knowledge of their members. And the Projects should enlist the assistance of qualified individuals who are committed to the project's success. Team members with great motivation is always a plus factor for the success of a project which builds commitment and confidence in their teams.

Under Agile methodology, good customer relationships and high customer loyalty are considered, along with the people factor on customer engagement as another important factor for managing data science projects. The project team should always have a good relationship with the customer. Since most customers are motivated to participate in projects and share business knowledge with team members because they expect the team to have successful and productive solutions to achieve their objectives and priorities, most customers are motivated to participate in projects and share business knowledge with team members. There should be a welcoming environment for the customers and obtain the customer competencies as well as they are willing to dedicate to the project to make it a success. On the other hand, project team will get the benefits of getting to know the customer's business well and sometimes, the project team will be able to access and explore the data with the level of authority that the customer is having.

The project Definition Process emphasizes the clarity of the goals set by the project team. The team members should have a clear understanding of the goal which is trying to achieve by the team. Therefore, the project managers should give a clear understanding of the project goals and the process that the team needs to follow to achieve the goals. At first, the business case should be developed and analyze the risk and cost and then select the process to follow. This is normally done during sprint zero in SCRUM methodology. Since this is considered a significant factor for managing the data science projects within Agile methodology, it is important to decide on the process and setting up the goals at the very beginning of the project.

Practicing Agile Software Engineering Techniques is one of the important factors that need to consider when managing data science projects within Agile methodology. When it comes to data science projects it is always a high priority factor on data quality. The early stages of a project are usually spent learning about and exploring the available data in preparation for the next steps. Therefore, Agile techniques can be utilized to explore and analyze data to identify and minimize the data quality issues if any. The project teams should demonstrate superior analytics techniques and handle the solution in a way that

built customer's understanding and satisfaction. Maintaining good version control techniques and tools in the project needs to be considered for data science projects as well to track the project versions and changes which helps the project in the long run as that will give a proper understanding and a controlling mechanism to the project.

Project Schedule should be considered when managing the data science projects within Agile methodology. Since most of the IT projects are time-boxed, the data science team must consider the given project time frames when planning their tasks. There can be different challenges face with the timelines due to data uncertainties by the data science team. But it is important to plan accordingly by understanding the nature of the project to proceed with the tasks within the project schedule.

Factors were listed as important for managing data science projects within Agile methodology. But the significance of these factors may vary according to the nature of the project that the team is involved in. Therefore, the team should always focus on these factors relative to the nature of the project.

5.3 Research Limitation and Future Works

In this study, the factors have been updated from the Chow and Cao model [2], and the additional attributes were added as per the reviewed literature. When considering the findings of this study the analyses of the inter-relationship of the factors have not been conducted. Even though the study shows the findings on the relationship of the used Agile methodology to the success of the project, it will be better to conduct the analyses to showcase the relationship between the independent factors as well.

Apart from the above dimensions of organizational, people, process, technical, and project another aspect that impacts the success of data science projects is 'Business and Marketing' which can be considered as an improvement to this study. Therefore, the factors which cover the business and marketing aspect could also be included into future research work. I hope that future studies will consider further improvements and examine interdependencies as well when considering various projects of various types.

References

- [1] J. Saltz, J. Stanton, “An Introduction to Data Science,” *SAGE Publications*, 2017.
- [2] T. Chow and D.-B. Cao, “A survey study of critical success factors in agile software projects,” *Journal of Systems and Software*, vol. 81, no. 6, pp. 961–971, 2008.
- [3] J. Sutherland and K. Schwaber, “The Scrum Guide,” scrumguides.org, Nov-2017.
- [4] D. West, “Scrum Guide Update”, *Scrum.org*, 2017.
- [5] N. Grady, “KDD meets Big Data,” *IEEE International Conference on Big Data (Big Data)*, pp. 1603-1608, *IEEE*, 2016
- [6] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, W. Rudiger, “CRISP-DM 1,” *Retrieved from The Modeling Agency*, 2000.
- [7] D. Anderson, “Kanban: Successful Evolutionary Change for Your Technology Business”. *Sequim, WA: Blue Hole Press*, 2010.
- [8] M. Ahmad, P. Kuvaja, M. Oivo, J. Markkula, “Transition of software maintenance teams from Scrum to Kanban,” *49th Hawaii International Conference on System Sciences (HICSS)*, January, 2016
- [9] M. Ahmad, J. Markkula, M. Oivo, “Insights into the Perceived Benefits of Kanban in Software Companies: Practitioners’ Views”, *International Conference on Agile Software Development (pp. 156-168)*, 2016.
- [10] Microsoft, “Team Data Science Process from Microsoft,” *github*, 2017
- [11] J. Spoelstra, H. Zhang, G. Kumar, “Data Science Doesn't Just Happen, It Takes a process. Learn about Ours...,” *Channel 9*, 2016, <https://channel9.msdn.com/Events/Machine-Learning-andData-Sciences-Conference/Data-Science-Summit-2016/MSDSS23>
- [12] J. Baijens, R. Helms, D. Ire, “Applying Scrum in Data Science Projects,” *IEEE Xplore*, August, 2020, doi: 10.1109/CBI49978.2020.00011
- [13] T. Chiyangwa, E. Mnkandla, “Modelling the critical success factors of agile software development projects in South Africa”, *South African Journal of Information Management* 19(1), a838, 2017, doi: <https://doi.org/10.4102/sajim.v19i1.838>
- [14] J. Saltz, I. Shamshurin, “Big Data Team Process Methodologies: A Literature Review and the Identification of Key Factors for a Project’s Success,” *IEEE International Conference on Big Data (Big Data)*, 2016

- [15] J. Saltz, I. Shamshurin, C. Connors, “Predicting Data Science Sociotechnical Execution Challenges by Categorizing Data Science Projects,” *Journal of the Association for Information Science and Technology*, 68(12):2720–2728, 2017
- [16] M. Tsoy, D. Staples, “Exploring Critical Success Factors in Agile Analytics Projects,” *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 2020
- [17] J. Saltz, I. Shamshurin, “Exploring the process of doing data science via an ethnographic study of a media advertising company”, *IEEE International Conference on Big Data (Big Data)*, 2015
- [18] J. Gao, A. Koronios, S. Selle, “Towards A Process View on Critical Success Factors in Big Data Analytics Projects”, *Twenty-first Americas Conference on Information Systems (AMCIS), Puerto Rico*, 2015
- [19] N. Grady, M. Underwood, A. Roy, W. Chang, “Big Data: Challenges, practices and technologies: NIST Big Data Public Working Group workshop”, *IEEE International Conference on Big Data (Big Data)*, 2014
- [20] D. Larson, V. Chang, “A review and future direction of agile, business intelligence, analytics, and data science,” *International Journal of Information Management*, 36(5), 700–710, 2016
- [21] D. Dutta and I. Bose, "Managing a Big Data project: The case of Ramco Cements Limited", 2015. doi:10.1016/j.ijpe.2014.12.032
- [22] W. Cochran, “Sampling Techniques”, 2nd Ed., New York: John Wiley and Sons, Inc., 1963.
- [23] J. Saltz, I. Shamshurin, K. Crowston, “Comparing data science project management methodologies via a controlled experiment”, *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.
- [24] M. Ikonen, E. Pirinen, F. Fagerholm, P. Kettunen, P. Abrahamsson, “On the impact of Kanban on software project work: An empirical case study investigation”, *Engineering of Complex Computer Systems (ICECCS), 16th IEEE International Conference on (pp. 305-314), IEEE*, 2011.
- [25] B. Ahmed, T. Dannhauser, N. Philip, “A Lean Design Thinking Methodology (LDTM) for Machine Learning and Modern Data Projects”, *10th Computer Science and Electronic Engineering (CEECE) (pp. 11-14), IEEE*, 2018.

- [26] J. Saltz, "The need for new processes, methodologies, and tools to support big data teams and improve big data project effectiveness", *IEEE International Conference on Big Data (Big Data)* (pp. 2066-2071), 2015.
- [27] G. Piatetsky, "CRISP-DM, Still the top methodology for analytics, data mining, or data science projects", KDnuggets, 2014. [Online]. Available: <https://www.kdnuggets.com/2014/10/index.html>.
- [28] J. Saltz and A. Sutherland, "SKI: An Agile Framework for Data Science", *IEEE International Conference on Big Data (Big Data)*, 2019.
- [29] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, D. Thomas, "Manifesto for Agile Software Development", 2001.
- [30] A. Cockburn, Design as Knowledge Acquisition, 2008, [online] Available: <http://a.cockburn.us/1973>
- [31] N. W. Grady, J. A. Payne and H. Parker, "Agile big data analytics: AnalyticsOps for data science," *2017 IEEE International Conference on Big Data (Big Data)*, 2017, pp. 2331-2339, doi: 10.1109/BigData.2017.8258187.
- [32] C. Bullen and J. Rockart, "A Primer on critical success factors, Information Systems Research", 1981.
- [33] U. Sekaran, Research Methods for Business -A Skill Building Approach, 4th ed. John Wiley & Sons., 2006.

APPENDIX A

Questionnaire Instrument

CRITICAL SUCCESS FACTORS FOR MANAGING DATA SCIENCE PROJECTS WITHIN AGILE METHODOLOGY

This survey includes the possible success factors of data science projects in software development using Agile methods, which had been compiled and consolidated from the academic and professional literature.

This survey is conducted for the research study as the partial fulfillment of the requirements for the

Master of Business Administration in Information Technology specialized in Business Analytics, Department of Computer Science and Engineering in University of Moratuwa

All the information you provide will remain completely confidential and will be used solely for academic purposes only.

Thank You,

Imali Limesha.

limesha.19@cse.mrt.ac.lk

Demographic Information

1. Age group
 - a. 20-30 years
 - b. 31-40 years
 - c. 41-50 years
 - d. Above 51 years
2. Gender
 - a. Male
 - b. Female

3. Designation
 - a. Data Scientist
 - b. Data Engineer
 - c. Data Analyst
 - d. Architect
 - e. Project Manager
 - f. Business Analyst
 - g. Other: -----

4. Level of experience in the above role
 - a. Lesser than 1 year
 - b. 1-3 years
 - c. 3-5 years
 - d. More than 5 years

Scale: Strongly Disagree 1

Disagree 2

Neutral 3

Agree 4

Strongly Agree 5

No.	Question	1	2	3	4	5
Organizational Dimension						
5	The Project receives strong executive support. Executive may mean the whole board of directors or the CEO, CTO, CIO, etc. who influenced the decision making.					
6	The organization has a cooperative culture instead of hierarchical one. A cooperative culture is one that					

	fosters ad-hoc teams driven by the needs of the job at hand. While hierarchical culture is one that has clear divisions of responsibility and authority.					
7	Agile methodology was universally accepted in the organization.					
8	The data science team is facilitated with a proper Agile-style work environment.					
9	Team members have great motivation and are committed to the project's success.					
10	The project team works in a coherent, self-organizing work manner and adapt to changing conditions					
People Dimension						
11	The selected project team members have high technical competence and expertise. (Problem Solving, Subject matter)					
12	The project provides appropriate technical training to the team including training on the subject matter and Agile processes.					
13	Project management has a good relationship with the customer					
14	The project has strong customer commitment and presence.					
Process Dimension						

15. What is the Agile methodology/ies practice in your organization to drive data science projects?
- a. SCRUM
 - b. CRISP-DM (Cross Industry Standard Process for Data Mining)
 - c. Kanban

- d. TDSP (Team Data Science Process)
- e. SCRUM- DS
- f. Other- -----

16	Project management is knowledgeable in Agile principles and processes					
17	The project follows Agile project management style (Eg: Plans generally, not being documented in detail, and deviations and changes being readily accepted and incorporated into the project plan)					
18	The project scope and objectives were well defined.					
19	The project follows Agile oriented requirement process (Eg: Specifying initial requirements at a very high level, leaving much room for interpretation adaptation as the project progress.)					
20	The project processes have developed effective communication and coordination within the team members					
Technical Dimension						
21	The project follows Agile oriented configuration management process. (Eg: Employing good version control or source code management processes)					
22	The project imposes a well-defined coding standard upfront.					
23	The project maintains the right amount of documentation for Agile purpose					
24	The project employs proper platforms, technologies, and tools suitable for Agile practice.					

25	The project supports delivering working software regularly within a short period of time.					
26	The project delivers the most important features first					
Project Dimension						
27	The project type is of variable scope with emergent requirements					
28	The project follows continuous and rigorous unit and integration testing for each iteration					
29	The project has a dynamic, accelerated schedule.					
Perception of success of the Data Science projects within Agile methodology						
30	The project is successful in terms of the scope and requirements of the project is met.					
31	The project is successful in terms of timeliness of project completion					
32	The project is successful in terms of the quality of its outcome.					

33. Please enter any additional comments or thoughts you have on what other critical factors affect the success of managing data science projects within Agile methodology.

APPENDIX B

Descriptive Statistics

Statistics: Management Commitment (MC) – Item No. 1

Question 05: The Project receives strong executive support. Executive may mean the whole board of directors or the CEO, CTO, CIO, etc. who influenced the decision making.

		Organizational-MC			Cumulative Percent
		Frequency	Percent	Valid Percent	
Valid	Strongly Disagree	12	5.0	5.0	5.0
	Disagree	22	9.2	9.2	14.2
	Neutral	73	30.4	30.4	44.6
	Agree	110	45.8	45.8	90.4
	Strongly Agree	23	9.6	9.6	100.0
Total		240	100.0	100.0	

Statistics

Organizational-MC		
N	Valid	240
	Missing	0
Mean		3.46
Std. Deviation		.963
Variance		.927
Minimum		1
Maximum		5

Table B. 1 Statistics: Management Commitment (MC) – Item No. 1

Statistics: Management Commitment (MC) – Item No. 2

Question 06: The organization has a cooperative culture instead of hierarchical one. A cooperative culture is one that fosters ad-hoc teams driven by the needs of the job at hand. While hierarchical culture is one that has clear divisions of responsibility and authority.

Organizational-MC

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	5	2.1	2.1	2.1
	Disagree	20	8.3	8.3	10.4
	Neutral	67	27.9	27.9	38.3
	Agree	97	40.4	40.4	78.8
	Strongly Agree	51	21.3	21.3	100.0
	Total	240	100.0	100.0	

Statistics

Organizational-MC

N	Valid	240
	Missing	0
Mean		3.70
Std. Deviation		.964
Variance		.929
Minimum		1
Maximum		5

Table B. 2 Statistics: Management Commitment (MC) – Item No. 2

Statistics: Organizational Environment (OE) – Item No. 3

Question 07: Agile methodology was universally accepted in the organization.

Organizational-OE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	5	2.1	2.1	2.1
	Disagree	15	6.3	6.3	8.3
	Neutral	34	14.2	14.2	22.5
	Agree	135	56.3	56.3	78.8
	Strongly Agree	51	21.3	21.3	100.0
	Total	240	100.0	100.0	

Statistics

Organizational-OE		
N	Valid	240
	Missing	0
Mean		3.88
Std. Deviation		.884
Variance		.781
Minimum		1
Maximum		5

Table B. 3 Statistics: Organizational Environment (OE) – Item No. 3

Statistics: Organizational Environment (OE) – Item No. 4

Question 08: The data science team is facilitated with a proper Agile-style work environment.

Organizational-OE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	8	3.3	3.3	3.3
	Disagree	7	2.9	2.9	6.3
	Neutral	63	26.3	26.3	32.5
	Agree	119	49.6	49.6	82.1
	Strongly Agree	43	17.9	17.9	100.0
	Total		240	100.0	100.0

Statistics

Organizational-OE		
N	Valid	240
	Missing	0
Mean		3.76
Std. Deviation		.896
Variance		.803
Minimum		1
Maximum		5

Table B. 4 Statistics: Organizational Environment (OE) – Item No. 4

Statistics: Team Environment (TE) – Item No. 5

Question 09: Team members have great motivation and are committed to the project’s success.

		Organizational- TE			Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Disagree	13	5.4	5.4	5.4
	Neutral	12	5.0	5.0	10.4
	Agree	112	46.7	46.7	57.1
	Strongly Agree	103	42.9	42.9	100.0
	Total	240	100.0	100.0	

Statistics

Organizational- TE		
N	Valid	240
	Missing	0
Mean		4.27
Std. Deviation		.791
Variance		.625
Minimum		2
Maximum		5

Table B. 5 Statistics: Team Environment (TE) – Item No. 5

Statistics: Team Environment (TE) – Item No. 6

Question 10: Team members have great motivation and are committed to the project’s success.

Organizational-TE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	3	1.3	1.3	1.3
	Neutral	25	10.4	10.4	11.7
	Agree	152	63.3	63.3	75.0
	Strongly Agree	60	25.0	25.0	100.0
	Total	240	100.0	100.0	

Statistics

Organizational-TE

N	Valid	240
	Missing	0
Mean		4.12
Std. Deviation		.625
Variance		.391
Minimum		2
Maximum		5

Table B. 6 Statistics: Team Environment (TE) – Item No. 6

Statistics: Team Capability (TC) – Item No. 7

Question 11: The selected project team members have high technical competence and expertise. (Problem Solving, Subject matter)

People- TC

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	12	5.0	5.0	5.0
	Neutral	32	13.3	13.3	18.3
	Agree	125	52.1	52.1	70.4
	Strongly Agree	71	29.6	29.6	100.0
	Total	240	100.0	100.0	

Statistics

People- TC

N	Valid	240
	Missing	0
Mean		4.06
Std. Deviation		.792
Variance		.628
Minimum		2
Maximum		5

Table B. 7 Statistics: Team Capability (TC) – Item No. 7

Statistics: Team Capability (TC) – Item No. 8

Question 12: The project provides appropriate technical training to the team including training on the subject matter and Agile processes.

		People-TC			Cumulative Percent
		Frequency	Percent	Valid Percent	
Valid	Disagree	20	8.3	8.3	8.3
	Neutral	55	22.9	22.9	31.3
	Agree	109	45.4	45.4	76.7
	Strongly Agree	56	23.3	23.3	100.0
	Total	240	100.0	100.0	

Statistics

People-TC		
N	Valid	240
	Missing	0
Mean		3.84
Std. Deviation		.879
Variance		.773
Minimum		2
Maximum		5

Table B. 8 Statistics: Team Capability (TC) – Item No. 8

Statistics: Customer Involvement (CI) – Item No. 9

Question 13: Project management has a good relationship with the customer

People-CI

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	3	1.3	1.3	1.3
	Disagree	8	3.3	3.3	4.6
	Neutral	38	15.8	15.8	20.4
	Agree	120	50.0	50.0	70.4
	Strongly Agree	71	29.6	29.6	100.0
Total		240	100.0	100.0	

Statistics

People-CI		
N	Valid	240
	Missing	0
Mean		4.03
Std. Deviation		.838
Variance		.702
Minimum		1
Maximum		5

Table B. 9 Statistics: Customer Involvement (CI) – Item No. 9

Statistics: Customer Involvement (CI) – Item No. 10

Question 14: Project has strong customer commitment and presence

People-CI

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	5	2.1	2.1	2.1
	Neutral	19	7.9	7.9	10.0
	Agree	138	57.5	57.5	67.5
	Strongly Agree	78	32.5	32.5	100.0
Total		240	100.0	100.0	

Statistics

People-CI		
N	Valid	240
	Missing	0

Mean	4.20
Std. Deviation	.669
Variance	.448
Minimum	2
Maximum	5

Table B. 10 Statistics: Customer Involvement (CI) – Item No. 10

Statistics: Project Management Process (PMP) – Item No. 11

Question 16: Project management is knowledgeable in Agile principles and processes

		Process-PMP			Cumulative Percent
		Frequency	Percent	Valid Percent	
Valid	Disagree	3	1.3	1.3	1.3
	Neutral	25	10.4	10.4	11.7
	Agree	152	63.3	63.3	75.0
	Strongly Agree	60	25.0	25.0	100.0
	Total	240	100.0	100.0	

Statistics

Process-PMP		
N	Valid	240
	Missing	0
Mean		4.12
Std. Deviation		.625
Variance		.391
Minimum		2
Maximum		5

Table B. 11 Statistics: Project Management Process (PMP) – Item No. 11

Statistics: Project Management Process (PMP) – Item No. 12

Question 17: The project follows the Agile project management style (Eg: Plans generally, not being documented in detail, and deviations and changes being readily accepted and incorporated into the project plan)

		Process-PMP			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	3	1.3	1.3	1.3
	Neutral	65	27.1	27.1	28.3
	Agree	160	66.7	66.7	95.0
	Strongly Agree	12	5.0	5.0	100.0
	Total	240	100.0	100.0	

Statistics

Process-PMP		
N	Valid	240
	Missing	0
Mean		3.75
Std. Deviation		.558
Variance		.312
Minimum		2
Maximum		5

Table B. 12 Statistics: Project Management Process (PMP) – Item No. 12

Statistics: Project Definition Process (PDP) – Item No. 13

Question 18: The project scope and objectives were well defined.

		Process-PDP			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	46	19.2	19.2	19.2
	Neutral	44	18.3	18.3	37.5
	Agree	102	42.5	42.5	80.0
	Strongly Agree	48	20.0	20.0	100.0
	Total	240	100.0	100.0	

Statistics

Process-PDP		
N	Valid	240
	Missing	0
Mean		3.63
Std. Deviation		1.010
Variance		1.020
Minimum		2
Maximum		5

Table B. 13 Statistics: Project Definition Process (PDP) – Item No. 13

Statistics: Project Definition Process (PDP) – Item No. 14

Question 19: The project follows Agile oriented requirement process (Eg: Specifying initial requirements at a very high level, leaving much room for interpretation adaptation as the project progress.)

		Process-PDP			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	3	1.3	1.3	1.3
	Disagree	2	.8	.9	2.1
	Neutral	45	18.8	19.2	21.4
	Agree	158	65.8	67.5	88.9
	Strongly Agree	26	10.8	11.1	100.0
	Total	234	97.5	100.0	
Missing	System	6	2.5		
Total		240	100.0		

Statistics

Process-PDP		
N	Valid	234

Missing	6
Mean	3.86
Std. Deviation	.660
Variance	.436
Minimum	1
Maximum	5

Table B. 14 Statistics: Project Definition Process (PDP) – Item No. 14

Statistics: Improved Coordination (IC) – Item No. 15

Question 20: The project processes have developed effective communication and coordination within the team members

		Process-IC			Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Strongly Disagree	18	7.5	7.5	7.5
	Disagree	51	21.3	21.3	28.7
	Neutral	60	25.0	25.0	53.8
	Agree	103	42.9	42.9	96.7
	Strongly Agree	8	3.3	3.3	100.0
	Total	240	100.0	100.0	

Statistics

Process-IC		
N	Valid	240
	Missing	0
Mean		3.13
Std. Deviation		1.030
Variance		1.062
Minimum		1
Maximum		5

Table B. 15 Statistics: Improved Coordination (IC) – Item No. 15

Statistics: Agile Software Engineering Techniques (ASE) – Item No. 16

Question 21: The project follows Agile oriented configuration management process. (Eg: Employing good version control or source code management processes)

		Technical-ASE			Cumulative Percent
		Frequency	Percent	Valid Percent	
Valid	Disagree	9	3.8	3.8	3.8
	Neutral	42	17.5	17.5	21.3
	Agree	134	55.8	55.8	77.1
	Strongly Agree	55	22.9	22.9	100.0
	Total	240	100.0	100.0	

Statistics

Technical-ASE		
N	Valid	240
	Missing	0
Mean		3.98
Std. Deviation		.746
Variance		.556
Minimum		2
Maximum		5

Table B. 16 Statistics: Agile Software Engineering Techniques (ASE) – Item No. 16

Statistics: Agile Software Engineering Techniques (ASE) – Item No. 17

Question 22: The project imposes a well-defined coding standard upfront.

		Technical-ASE			Cumulative Percent
		Frequency	Percent	Valid Percent	
Valid	Disagree	46	19.2	19.2	19.2
	Neutral	27	11.3	11.3	30.4
	Agree	125	52.1	52.1	82.5
	Strongly Agree	42	17.5	17.5	100.0
	Total	240	100.0	100.0	

Statistics

Technical-ASE		
N	Valid	240
	Missing	0
Mean		3.68
Std. Deviation		.977
Variance		.955
Minimum		2
Maximum		5

Table B. 17 Statistics: Agile Software Engineering Techniques (ASE) – Item No. 17

Statistics: Agile Software Engineering Techniques (ASE) – Item No. 18

Question 23: The project maintains the right amount of documentation for Agile purpose

		Technical-ASE			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	22	9.2	9.4	9.4
	Neutral	46	19.2	19.6	28.9
	Agree	124	51.7	52.8	81.7
	Strongly Agree	43	17.9	18.3	100.0
	Total	235	97.9	100.0	
Missing	System	5	2.1		
Total		240	100.0		

Statistics

Technical-ASE		
N	Valid	235
	Missing	5
Mean		3.80
Std. Deviation		.846
Variance		.716
Minimum		2

Maximum	5
---------	---

Table B. 18 Statistics: Agile Software Engineering Techniques (ASE) – Item No. 18

Statistics: Agile Software Engineering Techniques (ASE) – Item No. 19

Question 24: The project employs proper platforms, technologies, and tools suitable for Agile practice.

		Technical-ASE			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	7	2.9	2.9	2.9
	Neutral	35	14.6	14.6	17.5
	Agree	122	50.8	50.8	68.3
	Strongly Agree	76	31.7	31.7	100.0
	Total	240	100.0	100.0	

Statistics

Technical-ASE		
N	Valid	240
	Missing	0
Mean		4.11
Std. Deviation		.754
Variance		.569
Minimum		2
Maximum		5

Table B. 19 Statistics: Agile Software Engineering Techniques (ASE) – Item No. 19

Statistics: Delivery Strategy (DS) – Item No. 20

Question 25: The project supports to deliver working software regularly within a short period of time

		Technical-DS			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	22	9.2	9.2	9.2
	Neutral	73	30.4	30.4	39.6

Agree	113	47.1	47.1	86.7
Strongly Agree	32	13.3	13.3	100.0
Total	240	100.0	100.0	

Statistics

Technical-DS		
N	Valid	240
	Missing	0
Mean		3.65
Std. Deviation		.826
Variance		.682
Minimum		2
Maximum		5

Table B. 20 Statistics: Delivery Strategy (DS) – Item No. 20

Statistics: Delivery Strategy (DS) – Item No. 21

Question 26: The project delivers the most important features first

Technical-DS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	9	3.8	3.8	3.8
	Disagree	37	15.4	15.4	19.2
	Neutral	65	27.1	27.1	46.3
	Agree	100	41.7	41.7	87.9
	Strongly Agree	29	12.1	12.1	100.0
	Total	240	100.0	100.0	

Statistics

Technical-DS		
N	Valid	240
	Missing	0
Mean		3.43
Std. Deviation		1.012
Variance		1.024

Minimum	1
Maximum	5

Table B. 21 Statistics: Delivery Strategy (DS) – Item No. 21

Statistics: Project Type (PT) – Item No. 22

Question 27: The project type is of variable scope with emergent requirements.

		Project-PT			Cumulative Percent
		Frequency	Percent	Valid Percent	
Valid	Disagree	16	6.7	6.7	6.7
	Neutral	43	17.9	17.9	24.6
	Agree	150	62.5	62.5	87.1
	Strongly Agree	31	12.9	12.9	100.0
	Total	240	100.0	100.0	

Statistics

Project-PT		
N	Valid	240
	Missing	0
Mean		3.82
Std. Deviation		.737
Variance		.544
Minimum		2
Maximum		5

Table B. 22 Statistics: Project Type (PT) – Item No. 22

Statistics: Project Schedule (PS) – Item No. 23

Question 28: The project follows continuous and rigorous unit and integration testing for each iteration

Project-PS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	24	10.0	10.0	10.0
	Neutral	38	15.8	15.8	25.8
	Agree	147	61.3	61.3	87.1
	Strongly Agree	31	12.9	12.9	100.0
	Total	240	100.0	100.0	

Statistics

Project-PS		
N	Valid	240
	Missing	0
Mean		3.77
Std. Deviation		.799
Variance		.638
Minimum		2
Maximum		5

Table B. 23 Statistics: Project Schedule (PS) – Item No. 23

Statistics: Project Schedule (PS) – Item No. 24

Question 29: The project has a dynamic, accelerated schedule.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	18	7.5	7.5	7.5
	Neutral	35	14.6	14.6	22.1
	Agree	131	54.6	54.6	76.7
	Strongly Agree	56	23.3	23.3	100.0
	Total	240	100.0	100.0	

Statistics

Project-PS

N	Valid	240
	Missing	0
Mean		3.94
Std. Deviation		.823
Variance		.678
Minimum		2
Maximum		5

Table B. 24 Statistics: Project Schedule (PS) – Item No. 24

Statistics: Success of the Data Science projects within Agile methodology (SUCCESS) – Item No. 25

Question 30: The project is successful in terms of the scope and requirements of the project is met.

SUCCESS					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	10	4.2	4.2	4.2
	Neutral	51	21.3	21.3	25.4
	Agree	154	64.2	64.2	89.6
	Strongly Agree	25	10.4	10.4	100.0
	Total	240	100.0	100.0	

Statistics		
SUCCESS		
N	Valid	240
	Missing	0
Mean		3.81
Std. Deviation		.670
Variance		.448
Minimum		2

Maximum	5
---------	---

Table B. 25 Statistics: Success of the Data Science projects within Agile methodology (SUCCESS) – Item No. 25

Statistics: Success of the Data Science projects within Agile methodology (SUCCESS) – Item No. 26

Question 31: The project is successful in terms of timeliness of project completion

		SUCCESS			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	31	12.9	12.9	12.9
	Neutral	62	25.8	25.8	38.8
	Agree	132	55.0	55.0	93.8
	Strongly Agree	15	6.3	6.3	100.0
	Total	240	100.0	100.0	

Statistics

SUCCESS		
N	Valid	240
	Missing	0
Mean		3.55
Std. Deviation		.796
Variance		.634
Minimum		2
Maximum		5

Table B. 26 Statistics: Success of the Data Science projects within Agile methodology (SUCCESS) – Item No. 26

Statistics: Success of the Data Science projects within Agile methodology (SUCCESS) – Item No. 27

Question 32: The project is successful in terms of the quality of its outcome.

SUCCESS					
		Frequency	Percent	Valid Percent	Cumulative Percent

Valid	Strongly Disagree	3	1.3	1.3	1.3
	Disagree	15	6.3	6.3	7.5
	Neutral	30	12.5	12.5	20.0
	Agree	146	60.8	60.8	80.8
	Strongly Agree	46	19.2	19.2	100.0
	Total	240	100.0	100.0	

Statistics

SUCCESS		
N	Valid	240
	Missing	0
Mean		3.90
Std. Deviation		.820
Variance		.673
Minimum		1
Maximum		5

Table B. 27 Statistics: Success of the Data Science projects within Agile methodology (SUCCESS) – Item No. 27