Submitted in part fulfilment of the requirements for the degree of

Master of Science in Business Analytics

Utilizing Data Analytics and Power BI Visualization for Customer and Product Prioritization in the Manufacturing Industry

by

Siddhesh Suresh Jadhav

URN: 6761688

Faculty Of Arts and Social Science

University Of Surrey

September 2023

Word Count:

14755

Executive Summary

This study investigates the ways in which manufacturing companies might improve their prioritisation of customer and product enquiry using data analytics and Power BI visualisations. We used a dummy dataset consisting of 1542 enquiries from a manufacturer in order to reveal actionable insights. This dataset was analysed by using predictive clustering and interactive visualisations.

This research successfully created specific variables using characteristic data to gain deeper insights from the feature columns. This allowed us to understand trends in sales cycles over time and see monthly changes in inquiries and contract receipts. Further analysis revealed that k-means clustering was a better fit for our data than hierarchical clustering. This is because the k-means approach produced fewer, more understandable clusters, while the hierarchical method generated many clusters that weren't as useful for our dataset.

K-means clustering was used to divide clients into three categories with respect to their sales patterns. This revealed which clusters had superior overall performance and made it possible to design different sales strategies. A shorter best sales cycle duration is suggestive of a greater priority, and this metric has appeared as a useful enquiry prioritisation tool. Power BI visualisations allowed for multidimensional, interactive analysis, which was then used to find cluster-specific trends and make decisions based on the collected data.

Customers were categorised into clusters using clustering algorithms according to their enquiry and purchase patterns. Customers who fell into Clusters 1 and 2 proved much higher levels of engagement and conversion to contract than those who fell into Cluster 3. It is possible to develop sales strategies that are particular to each cluster by the characteristics and requirements of each group.

In clusters 1 and 2, the highest conversion rates were found with shorter enquiry cycles, specifically between 21 and 30 days. The sales cycle time for an enquiry was found to be a useful criterion for sales prioritisation.

Dashboards that enabled interaction supported a quick review of enquiries, PO received, and discounts offered by the cluster. Slicers made it possible to filter on important dimensions so that cluster-specific insights could be obtained. In comparison to the complex hierarchical clustering outputs, simplified visuals significantly improved the readability of the data. For

the purpose of prioritisation, Power BI transformed complex data into easily digestible and actionable insights.

One of the standout revelations from this study is the empowerment of managers through a strategically designed dashboard. This tool enables them to develop clear, actionable plans for each inquiry received daily. Given that inquiries are organized into clusters, sometimes overlapping in nature but always distinct in the combination of customer and product, managers have the insight to effectively categorize and prioritize them. Using metrics like sales rep performance and sales cycle duration, managers can strategically decide on inquiry deadlines, optimizing their approach to crafting techno-commercial offers.

Furthermore, this systematic approach ensures that follow-ups occur at the most opportune times, known as 'sweet spot days', maximizing the conversion potential of inquiries into purchase orders (POs). On particularly busy days, managers can harness this insight to prioritize. By sidestepping Cluster 3 inquiries and promoting those from Clusters 1 and 2, the company can maintain a profitable trajectory.

In simple terms, the findings of this research project proved how the combination of predictive analytics and interactive visualisations may enable manufacturing companies to realise their full sales potential by the data-driven prioritisation and control of their enquiry management. The findings of the study proved a model for achieving a competitive advantage using data analytics to support focused enquiry management strategies.

Declaration of Originality

I hereby declare that this thesis has been composed by myself and has not been presented or accepted in any previous application for a degree. The work, of which this is a record, has been carried out by me unless otherwise stated and where the work is mine, it reflects personal views and values. All quotations have been distinguished by quotation marks and all sources of information have been acknowledged by means of references including those of the Internet. I agree that the University has the right to submit my work to the plagiarism detection sources for originality checks.

12/09/2023

Date

Siddhesh Suresh Jadhav

Siddhesh Suresh Jadhav

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge my gratitude to Dr Andrew Hill my dissertation supervisor, for his guidance throughout the development of this dissertation. His sound advice, patience and humour were very motivating and much appreciated. Additionally, this project would have not been successful without the help of the Sunrise Process Equipment Pvt Ltd and Mr. N.V. Chaudhary who kindly provided additional dummy enquiry register data.

1 Introduction

The manufacturing industry is a complicated industry, and the business tactics that manufacturers employ play a significant part in how effectively their operations are carried out. Companies will typically choose between a make-to-stock (MTS) or make-to-order (MTO) strategy for their production (Akyildiz, et al., 2015). Manufacturers that follow the MTS technique develop products and hold them in stock, waiting for buyers to buy them (Arredondo & Martinez, 2010). On the other hand, manufacturers that use the MTO technique, such business that we are concentrating on for this research, create products in response to the orders placed by customers.

Because they take an MTO strategy, the organisation receives enquiry about variety of products. Each enquiry is unique because customers ask for a different type of equipment that is constructed from a different type of material and features a unique design and sizes. The manufacturing company need to supply a techno-commercial proposal in response to each of these enquiries. However, there is a difficulty. The number of enquiries that are sent to a company far exceeds the number of actual sales that are closed by that company. There is a lot that needs to be handled, but there are not enough resources, especially considering how few estimating or sales engineers they have.

In the day-to-day operations, the team leader assigns enquiries to engineers according to the dates on which these enquiries were received. This practice almost always results in the need to resolve backlogs, which creates difficulties for the sales staff. They routinely follow up on enquiries that, in some instances, have already been closed by the customers, with contracts being awarded to other manufacturing companies. This technique, while meticulous, may accidentally lead to inefficiencies and missed chances in the sales process, which is why a closer analysis is required.

The study suggests that utilising data analytics could be beneficial. It is possible for it to assist the company in determining which enquiries are more likely to result in sales if it is done correctly. However, there is yet another aspect to this, and that is visualisation. Even though data can supply the answers, it is visualisation tools like Power BI that can help make sense of this data. These tools can be useful for people who are not particularly knowledgeable about technology like machine learning.

However, very few companies in the manufacturing enquiry management sector have utilised data analytics to prioritise their customers or products enquiry. This is despite the fact that many other sectors have done so. Relatively few people have attempted to make decisions with the assistance of graphical representations of the data, such as those provided by Power BI.

As a result, the purpose of this research is to provide responses to the following important questions:

RQ1: How can data analytics, with the Aid of Power BI visualisation be utilised to prioritize customer and Product Enquiries?

RQ2: How does enquiry duration impact sales conversion, and can it be used as an effective metric for prioritisation?

RQ3: How do visualisation tools like Power BI influence decision-making processes in customer and product enquiry management?

In this research, we will build on earlier work done in the field of data analysis for the purpose of customer and product enquiry prioritisation. Specifically, we will investigate the appropriate method of data analysis and attempt to visualise the results using Power BI. The combination of data and graphics will be used with the intention of assisting businesses, in particular those working on the MTO model, in the process of making better decisions.

2. Literature Review

In the past ten years, we've noticed that manufacturing companies have been studying a variety of approaches to forecast sales and prioritise customers and products in order of value. What is their final goal? Increasing their revenue as well as their profitability. They are coming to rely more and more on technologies that help them analyse or visualise data in order to help them in making decisions. This change in strategy has resulted in some significant technological improvements in industry.

In order to acquire a full understanding, we carried out an in-depth review of the relevant literature. Throughout the course of our investigation, we made use of Google Scholar to do searches using terms such as "Data analytics in manufacturing," "Predictive analytics in manufacturing," "Customer prioritisation models," "Clustering for customer and product prioritisation," and "Power BI in decision-making." These searches assisted us in finding research papers that seemed relevant.

As we read more relevant literature that had been published on the topic, we became familiar with several repeating themes, most notably those that had a connection with the Prioritisation of Customers or Products. In addition, we are trying to discover a significant emphasis on the part that BI plays in the process of customer and product prioritisation.

In this study, we take a comprehensive look at the previous research that has been done on this subject. It is to observe the increasing significance of Power BI, which is a technology that has gained significant acceptance because of the outstanding data visualisation skills it possesses. Business Intelligence (BI) solutions, of which Power BI is a great example, are gaining an increasingly prominent importance in the changing landscape of the business world. Systems like this are extremely important in directing the processes that are necessary for decision-making.

2.1 Customer and Product Prioritization

The authors (Homburg, et al., 2008) discuss the critical importance of focusing on customers who bring the most value in their study. They advocate for businesses to adopt data-driven approaches to determine which customers to prioritize, regardless of whether they operate in Business-to-Business (B2B) or Business-to-Consumer (B2C) contexts. The core idea here is that companies should tailor their service levels and allocate resources based on the value each customer brings.

We found this idea further explored by (Kreuzer, et al., 2020) in their study. They approach business operations from a customer-centric perspective, emphasizing the significance of ensuring customer satisfaction. In essence, both studies advocate for using data to guide efforts in prioritizing customers and subsequently focusing on keeping these prioritized customers satisfied.

In contrast, (Huang, 2012) tackles a similar topic in his research, Instead of promoting complex models, he suggests that simpler methods may be more effective in determining which customers to prioritize. He cautions against overly intricate data analysis techniques, as they may not necessarily yield better results. In simpler terms, even when trying to predict future top clients using sophisticated models, simpler methods might actually be more accurate.

(Cao & Gruca, 2005) delve into how businesses can manage their relationships with customers in their study, they argue that a wise approach involves focusing on customers who generate profits. This concept aligns with (Huang, 2012), where businesses choose manufacturing requests to prioritize based on their profit potential, albeit from a sales perspective.

In another study by (Belhadi, et al., 2019) explain how Big Data Analytics (BDA) can provide a competitive edge to businesses by improving manufacturing processes.

A study by (Saleem, et al., 2019) primarily focused on e-commerce, underscores the notion that astute data utilization can benefit businesses across various industries. This shows the widespread relevance of data-driven decision-making.

The study by (Zhao & Keikhosrokiani, 2022) explores how predicting sales and recommending products can be influenced by observing customer behaviours, particularly in online shopping. It explains how understanding the duration customers calculates before making a purchase can aid in enhancing sales.

Furthermore, (Zhao & Keikhosrokiani, 2022) emphasize the impact of selecting the right customers, especially in online shopping contexts. In contrast, another study by (Mathur & Kumar, 2013), highlights the significance of the duration of a business's relationship with its customers in retaining them. Additionally, a research piece by (Yenaiaras & Kaya, 2021) discusses how choosing certain customers over others can affect job stress levels and customer service quality, especially in complex product scenarios and closely-knit business relationships.

These studies collectively underscore the importance of customer selection, but they approach it from various angles. For instance, (Zhao & Keikhosrokiani, 2022) delve into the use of technology for understanding online shoppers, while (Mathur & Kumar, 2013) focus on the historical relationships between businesses and their customers. This diversity highlights the existence of multiple strategies, some leveraging cutting-edge technology and others emphasizing long-term customer relationships.

The study (Haber & Fargnoli, 2019) investigates how businesses can determine the essential services their customers need, especially in uncertain circumstances. The researchers use various tools to address this issue and find that clear communication with customers is crucial for understanding their needs.

Even though a great number of studies have highlighted the significance of data-driven decision-making, there is still a lack of information in the published literature regarding how businesses, particularly those in the manufacturing of equipment, may successfully implement these ideas. Although predictive analytics have been mentioned as a method for predicting future trends, there is little information available on how effectively they can be applied in the actual world.

In addition, despite many research that examines the function of data analytics in decision-making, there is a shortage of work that focuses especially on the integration of visualisation tools such as Power BI in the manufacturing sector for the purpose of prioritising customers and products enquiries. Additionally, even though the significance of understanding the relationships with one's customers has been brought to light, the direct influence that visualisation tools have on moulding and improving these connections is still relatively uncharted territory.

After reviewing the current literature, we found that there are gaps in both the scope and the depth of the topic. Even though data analytics and the role it plays in decision-making have been at the centre of a great number of studies, specific tools like Power BI and their application in the prioritisation of customers and products within the manufacturing industry require further investigation. There is a lot of undiscovered territory when it comes to understanding how companies effectively use Power BI for prioritising customers and products enquiries. In addition, although predictive models show promise for identifying future patterns, there is still a need for further research into their applicability and the difficulties they present when applied to real-world circumstances.

2.2 Business Intelligence for Customer and Product Prioritization

As per (Negash, 2004) Business Intelligence can be best described as systems that bring together data gathering, data storage, knowledge management, and analysis. All these elements work in harmony to evaluate complex corporate and competitive information. Presenting this invaluable information to decision-makers with the aim of enhancing the timeliness and quality of decision-making (Anon., 2018).

Moving forward, let's explore the multifaceted nature of Business Intelligence. (Evelson, 2023) distinguishes between two primary definitions – the broad and the narrow. Our preference leans towards the broader definition, which views BI as a comprehensive set of methodologies, processes, architectures, and technologies. Transforming raw data into actionable insights, which, in turn, empower more effective strategic, tactical, and operational decision-making. On the other hand, the narrower definition focuses on the upper layers of the BI architecture, encompassing reporting, analytics, and dashboards.

In our research journey, we investigate the expansive realm of BI, not only exploring its broad applications but also its specific role within the manufacturing industry. This requires contextualizing recent research within the broader landscape of BI.

(Yiu, et al., 2021) in their studies emphasize that as BI gains traction, data becomes an even more invaluable asset for businesses. While their research primarily focuses on the tech sector, it's remarkably relevant to our study, which aims to decipher how BI can elevate operations in manufacturing companies.

However, there's a noteworthy shift towards placing greater emphasis on user experience (UX), as (Eriksson & Ferwerda, 2021) highlighted. This shift underscores the significance of not just raw data but also the usability and efficiency of BI tools' interfaces. A user-friendly interface can potentially lead to superior decision-making, cost savings, and heightened operational efficiency.

(Yiu, et al., 2020) suggest that a company's reliance on technology and its size can profoundly influence how it utilize BI tools and the benefits they extract from them. Adding another layer to the discussion is (Djerdjouri, 2020) with his research examines the real-world challenges companies face when attempting to implement advanced tools like Power BI. Especially for micro, small, and medium enterprises (MSMEs), these tools can pose complexities and financial constraints. Therefore, while BI tools hold immense potential, it's vital to comprehend the practical challenges associated with them.

(Zohuri & Moghaddam, 2020) emphasize the essential role of data, particularly in the domains of data analytics and BI. They discuss how tools like Power BI empower companies to gain deeper insights from their data, facilitating informed decision-making. Furthermore, (Zohuri & Moghaddam, 2020) delve into "Operational Business Intelligence," highlighting its role in aiding businesses in making instantaneous decisions, such as determining the importance of a product or customer.

Shifting our focus to the manufacturing sector, (Widjaja & Mauritsius, 2019) underscores Power BI's significance, especially in the automotive manufacturing industry. This application can be extend to data visualization and strategic decision-making.

The current state of the research landscape has, actually made contributions of BI tools more transparent. However, a lot of research tend to adopt a larger perspective, covering a variety of different businesses, without diving into particular technologies or their complex applicability in varied circumstances. Our research seeks to discover how Power BI is utilised in the manufacturing industry in order to fill this void in the knowledge base.

An attractive opportunity for further research exists in the application of Power BI in manufacturing, for the prioritisation of customers and products enquiry. Even though its benefits have been determined, there are not many studies that clearly indicate its practical application in this setting. In addition, the relationship that exists within the manufacturing industry between the prioritisation of customers and visualisation tools such as Power BI has not been thoroughly investigated.

In conclusion, the research we are conducting is to investigate the practical applications of Power BI within the manufacturing industry. In addition to bringing new ideas, it tries to improve our grasp of visualisation technology by combining it with approaches for prioritisation and syncing the two together.

Furthermore, although (Eriksson & Ferwerda, 2021) explore the user experience (UX) of BI systems, there are not any studies specifically delve into Power BI's UX within the manufacturing industry concerning customer and product prioritization. This research intends to fill that gap by concentrating on the design, usability, and effectiveness of Power BI, particularly as it relates to the prioritisation of customers and products enquiries.

The goal of our research is to give fresh insights, particularly about the ways in which Power BI may revolutionise operations in contexts associated with enquiry management in manufacturing industry. In order to properly design our research, we will make use of the information that we have gained from past research.

In simple terms, previous research has placed a strong emphasis on the importance of identifying key clients and the benefit of conducting data analysis. The majority of researchers believe that successful sales and a grasp of customer requirements are of the utmost importance. On the other hand, there is a lack of knowledge regarding the way tools such as Power BI make this possible, particularly in the manufacturing industry. This knowledge void is the focus of our research, which aims to fill it by highlighting the ways in which Power BI may improve decision-making in the context of customer and product enquiry prioritisation.

Previous research has demonstrated the critical importance of business intelligence (BI) technologies in a variety of business sectors. To summarise, business intelligence solutions like Power BI have evolved to the point where they are now essential for making well-informed decisions. Despite the fact that we recognise their worth, the primary objective of our research is to gain an understanding of the part that Power BI plays in the process of prioritising customer and product enquiries in manufacturing. We intend to provide a more comprehensive perspective on Power BI by investigating its user experience as well as its practical efficacy. In doing so, we will bridge the gaps that have been left by past studies.

2.3 Methodological Review

In this section, we will thoroughly examine the methodologies employed in five research papers.Our objective is to gain insights into the clustering techniques utilized in these papers and assess their applicability to our research.

We have categorized these articles based on their research methodologies. All five of them fall within the quantitative domain, involving extensive use of numbers and formulas. Notably, (Trappey, et al., 2010), (Syakur, et al., 2018), and (Frades & Mattiesen, 2009)employ the K-means clustering method, while (Soni & Bhardwaj, 2022) adopts the

Hierarchical clustering method for their data analysis. Additionally, (Frades & Mattiesen, 2009) introduces various other methodological techniques, as detailed in Table 2.1.

K-means clustering is a robust method for partitioning datasets into distinct groups, particularly suitable for handling large datasets. However, it requires a prior determination of the number of clusters and can face challenges with irregular datasets. It's worth noting that (Syakur, et al., 2018) addresses this challenge by using the Elbow method to determine the optimal number of clusters.

In contrast, Hierarchical Clustering constructs a hierarchy of clusters, aiding decision-makers in comprehending data relationships without the need to predefine cluster numbers. However, it demands more computational resources than K-means and is relatively slower.

Table 2. 1: Overview of Cluster Analysis Method (Frades & Mattiesen, 2009)

Pattern representation	**Similarity measure**	**Clustering algorithm**	**Assessment of the output**	**Representation of clusters**
Feature Selection	Euclidean distance	*Hierarchical* BIRCH, CURE, ROCK, DIANA, MONA,	*Clustering tendency*	*Graphs* Relevance networks
	Manhattan distance	*Partitional* *k*-means, ISODATA, PAM, CLARA, CLARANS, nearest Neighbor		*Partitions*
	Pearson's correlation coefficient	*Density-based* DBSCAN, DENCLUE	*Cluster validity* External, internal, and	*Classification trees*

			relative criteria. Validity indices (Dunn’s)	
	Vector angle distance	*Grid-based* WaveCluster and STING		
Feature extraction (Principal component analysis)	Squared Pearson’s correlation	*Fuzzy clustering* FCM	*Cluster stability* Bagging	*Dendrogram* Displaying the assessment of the uncertainty in hierarchical cluster analysis
	Inner product	Artificial neural networks for clustering. SOM , SOTA		
	Spearman’s rank correlation and Kendall’s Tau	*Evolutionary approaches for clustering* Genetic algorithms		
	Mutual information	Biclustering		

Table 2.1 provides an overview of various cluster analysis methods, encompassing feature selection, similarity measures, clustering algorithms, assessment of output, and cluster representation. Each of these techniques serves a unique purpose in data analysis, offering a wide array of options for researchers.

Upon careful examination of these articles, it becomes evident that K-means is the preferred choice among researchers. However, (EDWARDS & CAVALLI-SFORZA, 1965) and

(Syakur, et al., 2018) stand out for their comprehensive exploration of clustering methodologies. Notably, (Trappey, et al., 2010) and (Soni & Bhardwaj, 2022), while pursuing different objectives, both delve into customer prioritization within distinct contexts.

These two articles, (Trappey, et al., 2010) and (Soni & Bhardwaj, 2022), offer highly specific insights, though not all of their findings may be universally applicable. Both employ surveys to directly engage stakeholders for input. Interestingly, they do not extensively address the collection and analysis of existing data. In contrast, (Syakur, et al., 2018) focuses on customer profiling and segmentation, particularly in the context of segmenting bank customers based on characteristics such as age, gender, and residential area. This is vital for customer relationship management and targeted marketing. As this research talk about prioritisation of customer and product using clustering, we can use clustering method for our research. It's important to note that none of these articles utilize visualization tools like Power BI for cluster representation and decision-making.

After doing an exhaustive assessment of this body of research, it has become abundantly clear that K-means and Hierarchical Clustering are the approaches that are most frequently selected by researchers. The Elbow technique is a popular approach for K means clustering that can be used to estimate the best possible number of clusters to use. For the Hierarchical Cluster, a distance matrix was calculated with the Euclidean distance measure as the basis for the calculation. The Euclidean distance, which is the most widely employed metric, is a measurement that determines the linear distance between data points in a space that contains several dimensions. Nevertheless, we need to keep a watchful eye out for potential problems and think about different methods as we are carrying out this research. The purpose of our research is to improve the quality of the data analysis that is performed using these two analyses by introducing more powerful visualisation tools such as Power BI.

2.4 Hypothesis

In today's environment, which is driven mostly by data, visualisations play an important part in the decision-making process, which helps to improve decision-making capabilities. Power BI, which is recognised for the dynamic visualisation capabilities it possesses, is one of the tools that has received a lot of attention. Our research intends to explore the role that Power BI visualisations play in aiding the prioritisation of customer and product enquiries inside manufacturing companies. This is important to us given the potential impact that it could have.

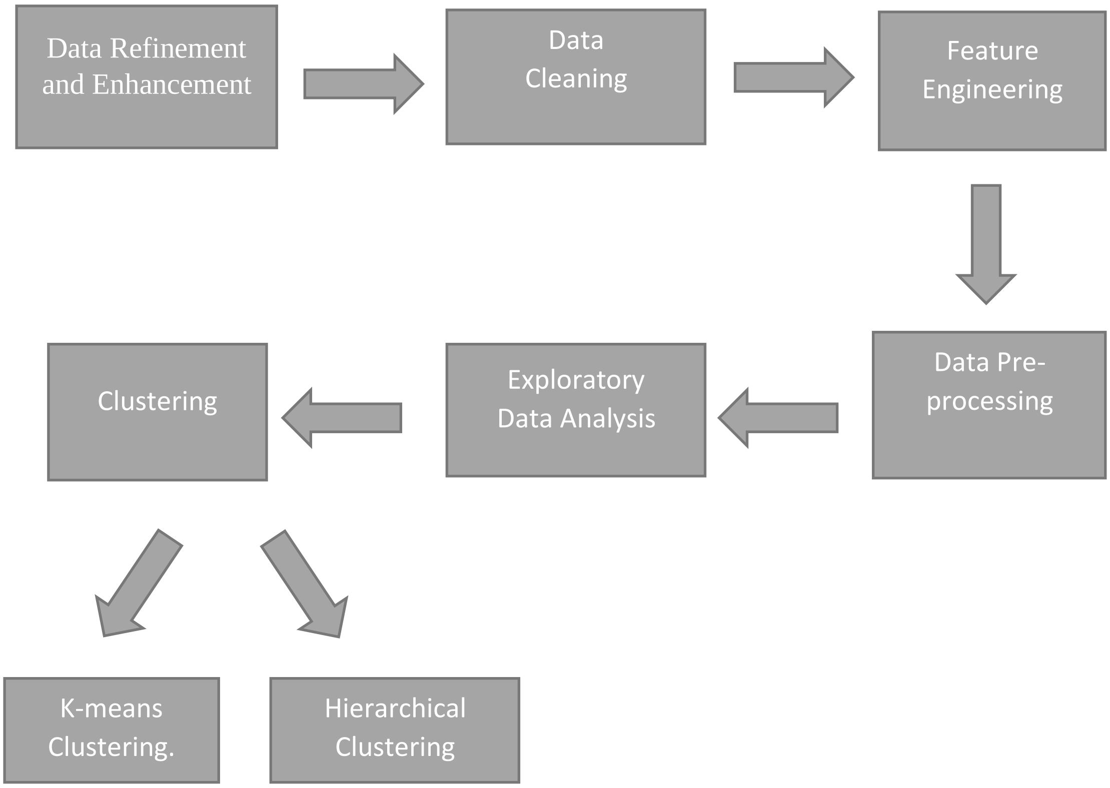
Null Hypothesis (H0): The utilization of Power BI visualizations does not significantly enhance the prioritization of customer and product enquiries in manufacturing companies.

Alternative Hypothesis (H1): The application of Power BI visualizations significantly improves the prioritization of customer and product enquiries in manufacturing companies.

3 Methodology

The research onion framework is used as a source of inspiration for our research technique, which takes a pragmatic point of view. Given the structured and data-driven nature of our study, we believe that the quantitative technique is the one that is best suited to yield statistically valid results, and as a result, this is the methodology that we have chosen to put into practise for our research. For the purpose of establishing whether Power BI can live up to its potential as a platform for data analysis and visualisation, this method is very important.

Regarding the development of our methodology, we follow to the steps that are listed in Figure 3.1.



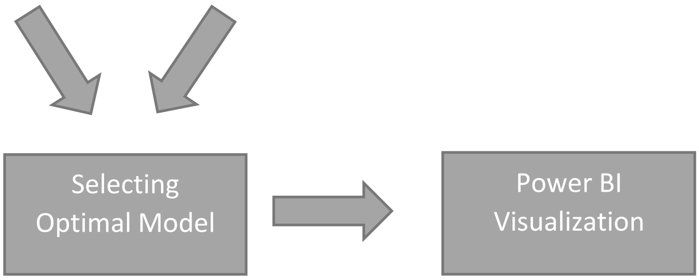


Figure 3. 1: Block diagram for Methodology

3.1 Data Refinement and Enhancement

Our data came from a "dummy enquiry register" this register features a variety of columns, including Sr No for sequential identification, Client Names to capture customer identities, Enquiry Specification for detailed enquiry information, Quotation Number for unique references, Dates for follow-ups, and the Status column to indicate enquiry outcomes.

We were aware that the initial data contained problems, the most significant of which was its inability to differentiate between different types of equipment based on an ambiguous column named "equipment specification." As a result, we started the process of modifying the data because we were aware of these flaws. We included a number of additional columns in Excel in order to remove any confusion and make it simpler and faster to identify the various types of equipment. These three new columns have been designated as Equipment Type, Equipment Material, and Equipment Specification respectively.

3.2 Data Cleaning, Feature Engineering, and Pre-processing

We were able to get new insights on the data kinds and initial values after loading the necessary RStudio packages and carrying out an initial analysis of the dataset. We began by applying the ‘lubridate’ package to the date columns in order to create greater consistency. Next, we added a column in which to keep records of follow-ups, and last, we standardised the date columns. Additional elements, such as the enquiry month and then calculated amount of time from the beginning to the end of the enquiry, were computed in order to provide a finer degree of data granularity. This was done in order to provide a more detailed level of information. During the process of transforming categorical variables into binary matrices, unncesssary columns were removed, and one-hot encoding was implemented. The data underwent a scaling operation in order to standardise the independent variables and make the clustering technique more equal. This was accomplished by reducing the variance of the data.

3.3 Time Study

The duration of sales enquiries from initial enquiry arrival to final purchase order or closed status was examined using a time study. The purpose of this test was to investigate whether or not the enquiry cycle time may serve as a useful statistic for the prioritisation of sales.

The enquiry dataset included date columns that captured the start date of the enquiry as well as the end date and the date the enquiry was closed. With the help of R studio algorithms, the time duration between these dates was determined in days for each enquiry. It was hypothesised that better priority and conversion potential would result from shorter enquiry cycles since customers would make decisions more quickly.

3.4 Exploratory Data Analysis (EDA) and Clustering

The dataset that we are analysing contains 1542 observations, each of which uniquely identifies an enquiry with a number that falls between the range of 1 to 1542 and is denoted by the notation 'SR.NO.' These entries cover a period of time beginning on April 1, 2021 and ending on March 21, 2022, which provides us with the chance to record a variety of clients within this time span. The number one spot goes to "Customer 14," who has 134 enquiries, followed by "Customer 87," who has 92 enquiries, and finally "Customer 42," who has 66 enquiries.As can be observed from the distribution, various customer enquiries occur at various frequencies.

We can see in the frequency distribution chart 3.2 below, the 'Enquiry Type' column classifies these enquiries as either 'new order' or ‘service,' with the new order category being the more common of the two.

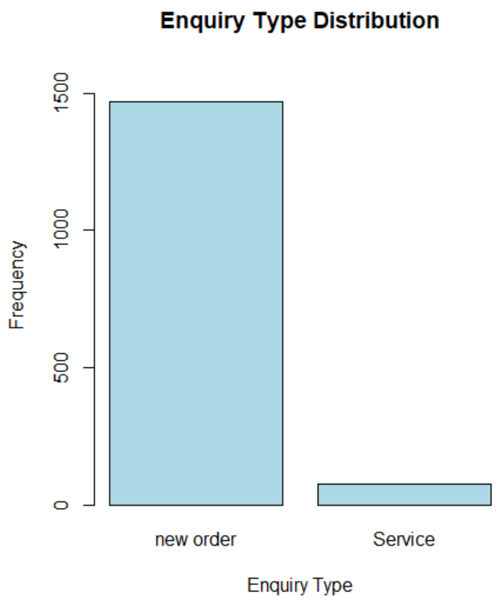


Figure 3. 2: Enquiry Type Distribution in Dataset

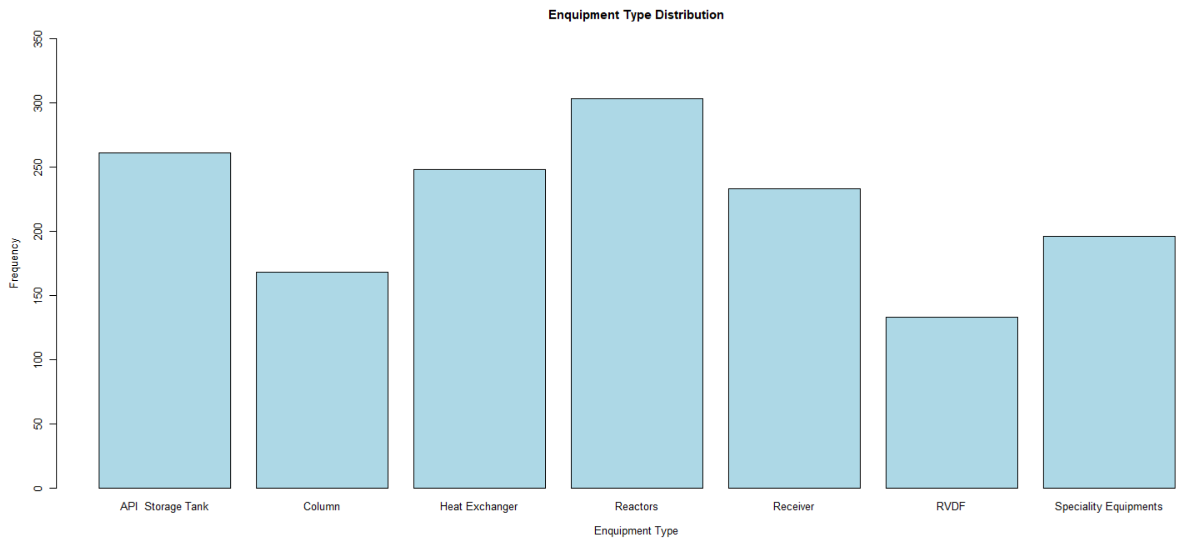


Figure 3. 3: Equipment Type Distribution in Dataset

Figure 3.3 offers some interesting insights into the distribution of enquiries based on the 'Equipment Type.' The term 'Reactors' receives the most enquiries, followed by 'API Storage,' 'Heat Exchanger,' and 'Receiver,' while the terms 'Speciality Equipment,' 'Exceeding Columns,' and 'RVDF' receive relatively fewer enquiries.

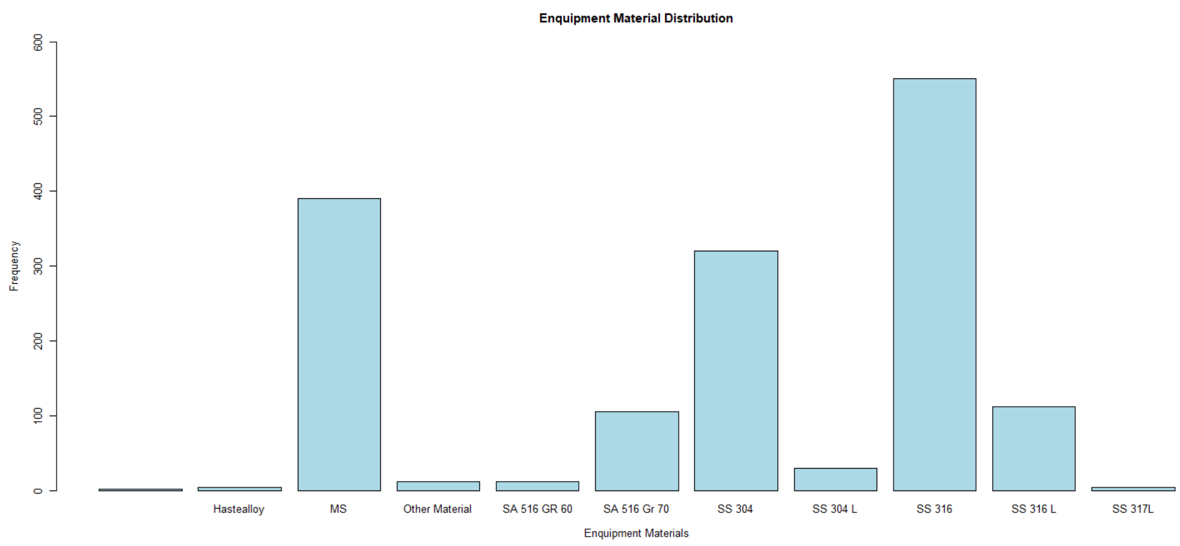


Figure 3. 4: Equipment Material Distribution in Dataset

The distribution of the 'Equipment Material' variable in the dataset is shown in Figure 3.4. 'SS 316' equipment leads enquiries, followed by 'MS' and 'SS 304,' suggesting their popularity among clients.

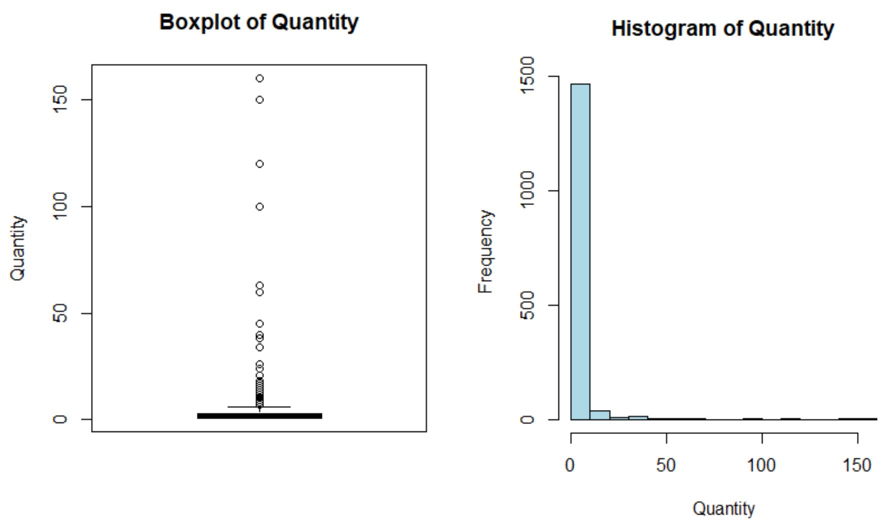


Figure 3. 5: Boxplot and Frequency distribution of Quantity

Customers have made enquiries concerning a quantity of equipment types, as depicted in figure 3.5. On average, customers indicate an interest in around 4 units of equipment. However, the distribution is not uniform; some customers are interested in acquiring just one unit of equipment, while others are interested in purchasing as many as 160 units of equipment. It is essential to consider the fact that the middle fifty percent of enquiries take place within the range of one to three pieces of equipment. There are certain inconsistencies in the data, such as occasions in which customers make enquiries for significantly larger quantities, possibly in connection with bulk orders or for big projects.

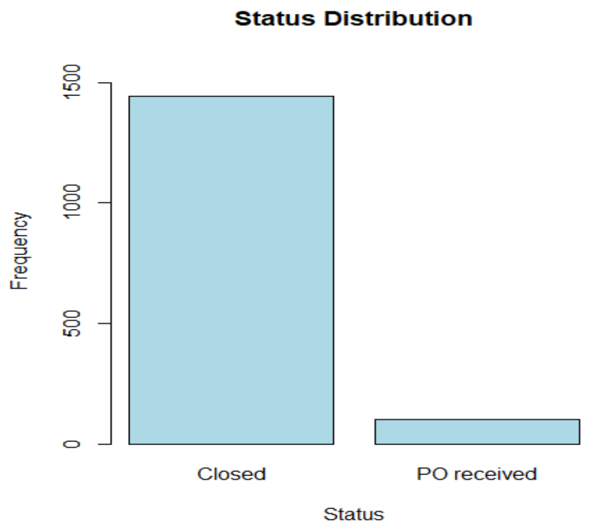


Figure 3. 6: Frequency Distribution of Status

Figure 3.6 is a graphical representation of the distribution of enquiries based on the 'Status' variable. It highlights the disparity between successful contracts (101 out of 1542 enquiries) and those that did not come to fruition.

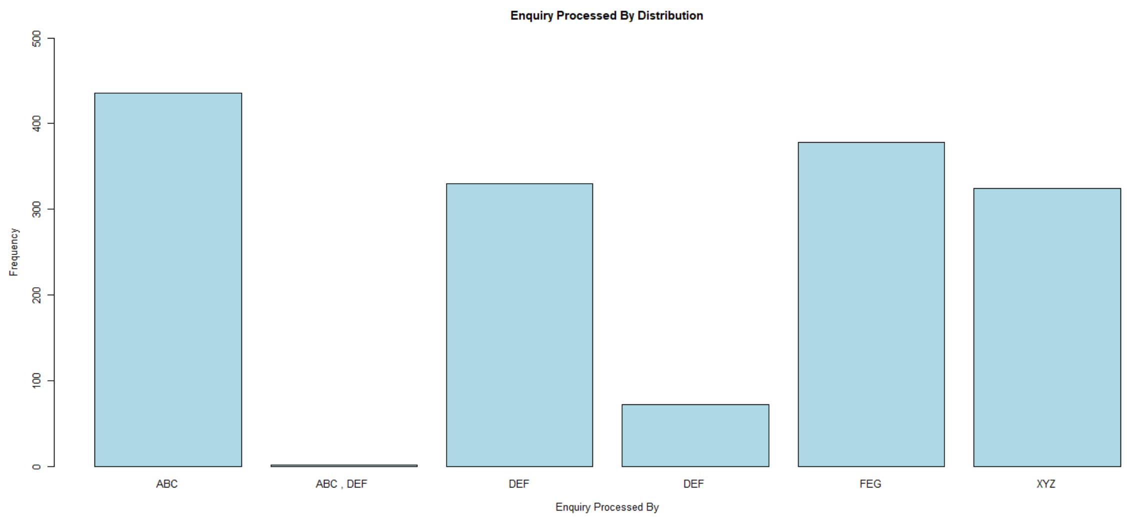


Figure 3. 7: Frequency distribution of Enquiry Processed By

Figure 3.7 presents the frequency distribution of the 'Enquiry Process By' variable, shedding light on the distribution of enquiries among company employees.

After the feature engineering of the dataset, we added three more variables to improve our knowledge. In December, there was significantly more activity than in the other months, as seen in Figure 3.8, which provides a monthly breakdown of enquiries. On the other hand, there are fewer enquiries during the months of March and April.

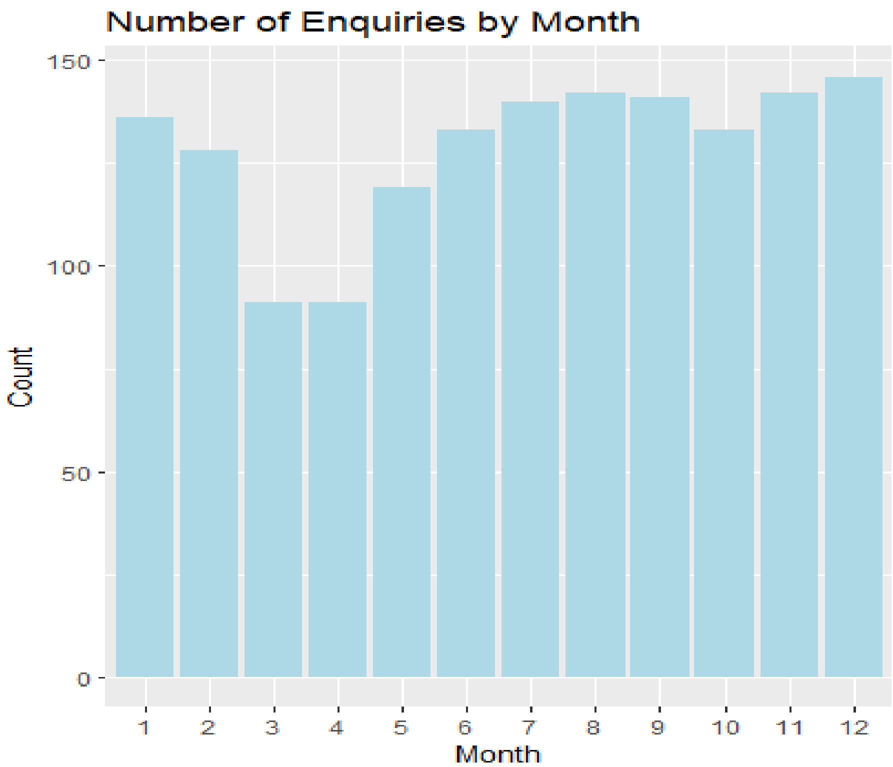


Figure 3. 8: Frequency Distribution of Enquiries by Month

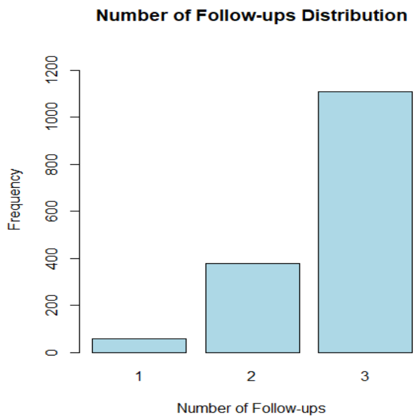


Figure 3. 9 : Frequency distribution of Number of Follow-ups

Figure 3.9 graphically depicts the frequency distribution of follow-ups for enquiries. It is evident from the chart that the company primarily conducted three follow-ups for most enquiries, surpassing the frequency of both one and two follow-ups.

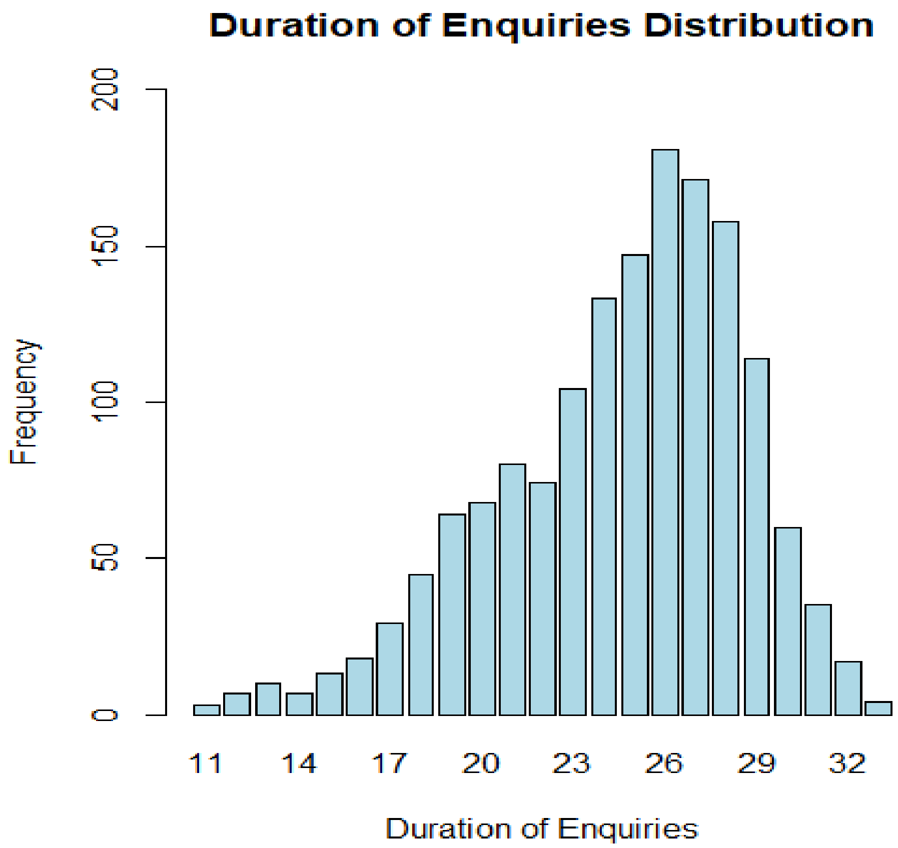


Figure 3. 10: Frequency distribution of Duration of Enquiries

Figure 3.10 is a thorough visualisation that provides an overview of the amount of time that passes between when a client gives an enquiry and when it is closed. Most enquiries got closed over a period of 24 to 28 days, although there are exceptions; some take as long as 38 days, while others are completed within 10 to 15 days.

3.5 Clustering

3.5.1 K-means Clustering

When performing k-means clustering, one of the most important decisions to make is deciding how many groups should be created. K-means is a method for clustering that is based on a centroid, and the goal of the method is to minimise the sum of squared distances that separate each point from the centre that is assigned to it. The total within-cluster sum of squares (SSE) is plotted against different cluster counts using the "elbow plot" technique to determine the optimal number of clusters.

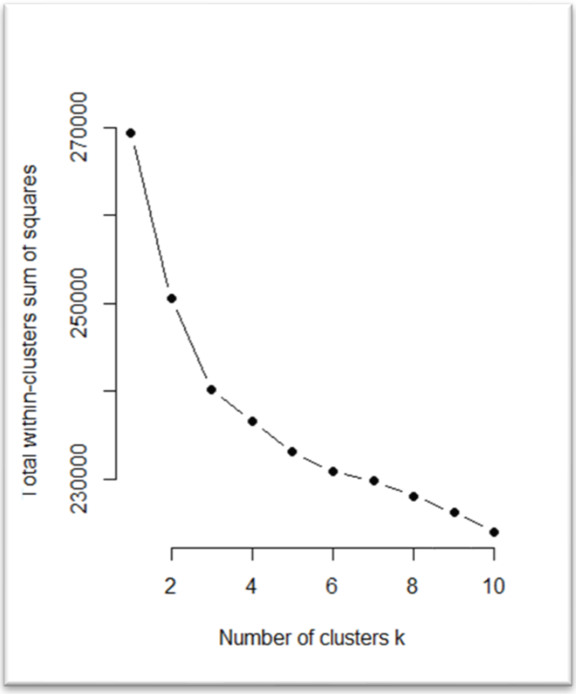


Figure 3. 11: Number of clusters using Elbow method.

As a result of our investigation into figure 3.11, we have come to the conclusion that three clusters (k=3) provide an equitable solution. This is due to the fact that the curves take a sharp bend at 3, which provides meaningful distinctions without adding unneeded complication. The data went through some pre-processing steps before the clustering phase began. The column means were used to fill in any gaps in the dataset with missing values. Infinite values were changed to not applicable (NA), and columns that had a variance of zero were found and eliminated because they did not provide any new information to the clustering process.

After counting the number of clusters, we move on to the next step of the process, which is k-means clustering with three centres. From among 25 possible random configurations, we choose the one that results in the most cohesive clustering. After the clustering was complete, the fviz_cluster function was used to visualise the clusters by making use of the principal components of the numeric variables. Additional 2D visualisations based on features such as 'Number Of Follow Up', 'ENQUIRY START TO END IN DAYS', 'ENQUIRY MONTH', and 'STATUS' were developed in order to study the structure and distribution of clusters in order to gain a more in-depth knowledge. The results are incorporated into the dataset, and each data point is subsequently assigned to the appropriate cluster.

The initial dataset underwent additional modifications in order to get it ready for additional analysis in Power BI. It was decided to build a binary representation of the STATUS feature, with the value 0 denoting "Closed" and 1 meaning "PO received." The 'Last Discount Offered' feature has been transformed to a numeric scale by eliminating the % sign and scaling the percentage values, which were formerly expressed as percentages, to fall between 0 and 1.

3.5.2Hierarchical Clustering

'hclust' is a model that is used to construct a tree of clusters in hierarchical clustering. This tree is used to classify the data into meaningful clusters. For the purpose of computing distance, we computed the distance matrix utilising the Euclidean distance measure. A multidimensional space can be analysed using the Euclidean method, which calculates the linear distance between each pair of data points

Plotting was done on the tree (dendrogram) that emerged as a result of the clustering process. This visually intuitive interpretation of the clustering was made possible by utilising the 'dendextend' package, which was used to colour and name branches according to the clusters that they belonged to. Because hierarchical clustering is memory-intensive, especially when applied to big datasets, a representative sample of 1000 observation was obtained from the scaled data for certain processes to ensure that they are computationally feasible. This sample was taken from the data.

Following the completion of the clustering process, a bar plot was utilised in order to visualise the distribution of data points across the resulting clusters. This offers an uncomplicated perspective on the manner in which the data points are dispersed among the clusters. For the purpose of delving deeper into the characteristics of the clusters, scatter plots were developed to visualise the ways in which the clusters varied with regard to particular aspects, such as the "Number of Follow Up" and the "Enquiry Duration (Days)."

3.6 Power BI Visualization

After completing the data analysis in RStudio, it was time to move on to the visualisation phase of the project. First, we took a step back to assess our options and determine which approach to clustering made the most intuitive sense to us. After deciding on one, we organised our data so that Power BI could present it in a visually. Instead of writing our R script from scratch, we imported it directly from RStudio into Power BI.

Our Power BI dashboard has been updated to include a number of interactive features in order to enhance both the user experience and the readability of the data. The crucial variables of "Client," "Products," "Quantity," "Status," "Cluster," "Sales Rep," and "Enquiry Month" were all specifically included in our slicers. Users can easily filter and delve into particular subsets of the data by utilising these slicers, which ultimately results in more precise insights being provided. Cards that indicated important metrics, such as 'Enquiries,' 'PO received,' and 'Discount Offered,' were added to our dashboard in order to further improve its functionality. Users were able to rapidly absorb the information that was most important to them because to the fact that these cards offered a concise breakdown of key performance indicators that could be seen immediately.

Our data visualisation strategy included a variety of chart types in order to cater to the various requirements for the portrayal of data. For instance, a "Pie chart" was utilised to illustrate the relationship between the number of customers and the number of enquiries, with the data being further segmented into clusters. This visualisation made it simple to dissect the client distribution across multiple clusters, drawing attention to any potential areas of interest or concern that may have been present.

In addition, a 'Clustered Column Chart' was crafted so that the sales cycle times for each individual client could be shown. With the 'Enquiry start to finish days' plotted on the x-axis, the 'number of enquiries' plotted on the y-axis, and additional categorization offered by the legend based on clusters, the visual brought to light patterns and trends in sales cycles across a variety of client groups.

In the end, we utilised a scatter plot to demonstrate the relationship between the number of enquiries received and the average discounts that were provided. The purpose of developing this scatter plot was to demonstrate, taking into account the various clusters, how the typical discounts related to the amount of enquiries that are received. A visualisation of this kind was needed in order to determine whether there was a propensity to grant more discounts in response to an increase in enquiries or whether some clusters tended to receive more discounts regardless of the quantity of enquiries that they submitted.

Using columns like "Equipment Type," "SR. NO.," and "Cluster," we created a "Equipment Demand by Cluster" visualisation for our analysis. We were able to interpret the distinct demands that existed across clusters for the various types of equipment by using a column chart that was grouped. In addition, we utilised a ‘treemap’ so that we could investigate the material preferences shared by each of these clusters. This graphical representation, which was constructed utilising the columns labelled "Equipment Material," "Cluster," and "SR. NO.," can be used to show the material breakdown by cluster and shedding light on distinct material tendencies within each group.

We decided to implement some fast measurements so that we could gain a better knowledge of our conversion numbers. To calculate the conversion percentage, we compared the total number of enquiries with the number of purchase orders (PO) that were received.

Specifically, we used the "divide" feature to do this comparison. This automated indicator painted a very clear picture of how effectively our enquiries were converted into actual purchase orders.

We decided to add a new column to our spreadsheet titled "Enquiry Status" so that we could gather more particular insights. The word "Status" was picked to serve as the construction criterion, and the value 1 was selected. The primary purpose of this article was to identify and gain an understanding of the essential components that make it possible for an enquiry to effectively become an approved PO. Through analysis of this data, we were able to determine the primary elements that contributed to the achievement of successful conversions.

To conduct a more in-depth analysis, we used "key influencers" tool. We were able to discover the essential components that affected the potential that an enquiry would evolve into a received PO by making use of this feature within Power BI. In other words, we were able to determine the likelihood that an enquiry would become a received PO. By analysing these influencers, we were able to gain a comprehensive understanding of the aspects that drive conversion, which facilitated more effective planning and led to the simplification of our processes.

The most important aspect of our work with Power BI went beyond merely visualising the data; rather, it focused on obtaining actionable insights from the information we gathered. In order to investigate the subtle patterns and aspects that form the foundation of our company's operations, we carefully manipulated the data and took advantage of the cutting-edge features that Power BI has to offer.

4. Findings

4.1 Clients

Our research is aimed at determining the extent to which data analytics, when paired with the visualisation capabilities of Power BI, utilised to rank prioritise customer and product enquiries within the manufacturing business. The findings that are associated with our primary research topics are detailed in this section.

Out of 1,542 enquiries, just 101 enquiries turned into purchase orders, according to a preliminary analysis of a dataset provided by the manufacturing company. This contrast highlights a opportunity for the company to boost the volume of their sales. These indicate that there is a large untapped market for the company's equipment.

Several major patterns that show the performance of our organisation arose from our in-depth analysis of customer interactions related equipment procurement.

Most of the customers didn’t buy equipment in compared to customers who have bought the equipment. For example, "Customer 1," "Customer 11," "Customer 14," and "Customer 45" all had a noticeably larger number of closed statuses in relation to the number of POs they have given.

The fact that 'Customer 14' had a notable level of interaction with the company was highlighted by the fact that company had received 6 contracts in contrast to losing 128 contracts. In a similar vein, "Customer 39" and "Customer 66" each received 8 and 7 contracts, respectively.

Some customers, such as "Customer 21," "Customer 36," "Customer 43," and "Customer 55," had a balanced relationship, with an equal number of contracts received and contracts missed.

"Customer 87" displayed a substantial inconsistency by having 90 missed contracts compared to 2 received contracts given, which suggests that there may have been difficulties in the negotiation or product fitment for this client.

Even if such encounters resulted in a contracts loss status, most of the clients on the list had constant interactions with the organisation. This might be an indication of a strong presence in the market, but it might also reveal areas of product, pricing, or sales approach that could be looked at again and improved.

In conclusion, even though the company has been successful in securing contracts with a variety of clients, there is a clear pattern of possibilities that have been lost. It is essential to conduct more research into the factors that led to the 'Closed' status in order to identify and address any potential weaknesses in the sales or product strategy of the company.

4.2 Equipment Type

Table 4. 1: Equipment type vs Status.

SR NO	EQUIPMENT.TYPE	STATUS	Count
1	API Storage Tank	Closed	247
2	API Storage Tank	PO received	14
3	Column	Closed	155
4	Column	PO received	13
5	Heat Exchanger	Closed	234
6	Heat Exchanger	PO received	14
7	RVDF	Closed	125
8	RVDF	PO received	8
9	Reactors	Closed	278
10	Reactors	PO received	25
11	Receiver	Closed	221
12	Receiver	PO received	12
13	Speciality Equipment’s	Closed	181
14	Speciality Equipment’s	PO received	15

Upon doing an analysis of Table 4.1, namely the company equipment column and the status columns, substantial insights regarding the sales status of each type of equipment emerged, including the following:

The company API Storage Tanks recorded a total of 14 contracts received (PO received), out of which 247 cases were recorded as 'Closed' because the company was unable to secure the contract. For Column equipment, the company was able to successfully secure contracts for 13 of the 155 enquiries it received.

Heat Exchangers had a more favourable ratio, as the company was successful in acquiring 14 contracts while failing to take advantage of 234 opportunities. RVDF equipment saw a decrease in the number of successful contracts received, which came to a total of 8, in comparison to the 125 closed instances. The Reactors business showed a significant departure from the overall trend, as only 25 contracts were successfully received compared to the 278 that were closed. In the case of Receivers, the company was successful in securing 12 contracts but failed to secure 221 others. In the end, Speciality Equipment’s was only able to get 15 successful contracts for the company, while 181 other attempts resulted in the company being unable to finalise the contract.

In conclusion, among the several types of equipment that were examined, reactors had the highest number of contracts that were received by the company, but they also had a significantly higher number of possibilities that were passed up. It is vital that any future initiatives take into consideration these insights in order to increase the percentage of successful sales across all types of equipment.

4.3 Discount Offered

Table 4. 2: Final Discount Offered vs Status.

Sr No	Final Discount Offered	STATUS	Count
1	0%	Closed	56
2	10%	Closed	4
3	2%	Closed	254
4	3%	Closed	120
5	4%	Closed	318
6	4%	PO received	2
7	5%	Closed	165
8	5%	PO received	7
9	6%	Closed	360
10	6%	PO received	35
11	7%	Closed	151
12	7%	PO received	53
13	8%	Closed	5
14	8%	PO received	4
15	9%	Closed	8

Following an examination of Table 4.2, in which the relationship between discounts offered and the conversion of possible contracts was investigated, several interesting trends became apparent. There were 56 missed opportunities when a no discounted approach was taken, demonstrating that a no discounted strategy may not be the most successful way to secure contracts. At the 6% discount level, the highest level of involvement was observed, with 360 contracts not being secured but an impressive 35 contracts being purchased successfully. This would imply that this discount range may be able to strike a balance between the client's desire for an attractive discount and the profitability of the discount for the company. The 7% discount, which resulted in 151 missed opportunities but secured 53 contracts, had the highest success rate among all discounts that were offered. Clients clearly prefer this discount range, as evidenced by this. Surprisingly, even greater discounts, such as 8% and 9%, did not indicate a continuous increase in the acquisition of contracts. In the case of the 8% discount, for example, the outcomes for closed and successfully received contracts were almost identical.

There is a predominant pattern of missed possibilities for discounts ranging from 2% to 5%, with the 2% discount revealing an astounding 254 contracts that were not secured.

In conclusion, the data reveal that although offering discounts can raise the likelihood of getting a contract, the relationship between the two is not a simple linear one. It appears that discounts in the middle of the range, particularly between 6% and 7%, are the most successful in contract conversions. However, better success rates might not necessarily be directly correlated to discounts that are either extremely high or extremely low. This highlights the significance of taking a strategic approach when offering discounts, taking into consideration both the clients' perceived value and the actual value that is supplied to them.

4.4 Estimation engineer Or Sales Rep

Table 4. 3: Enquiry Process By vs Status.

Sr No	ENQUIRY.PROCESSED.BY	STATUS	Count
1	ABC	Closed	396
2	ABC	PO received	40
3	ABC , DEF	Closed	2

4	DEF	Closed	317
5	DEF	PO received	13
6	DEF	Closed	57
7	DEF	PO received	15
8	FEG	Closed	369
9	FEG	PO received	9
10	XYZ	Closed	300
11	XYZ	PO received	24

The dataset indicates how effective certain agents are in processing enquiries, as well as their rate of success in converting these enquiries into actual contracts:

396 contracts were not secured as a result of "Agent ABC" processing a sizable quantity of enquiries. Despite this, ABC was able to secure forty contracts. In addition, when joint efforts with 'DEF' were attempted, there were two opportunities that were missed. This suggests that 'ABC' has a respectable conversion rate, but that it may make greater progress through collaborative efforts or alternate tactics.

A total of 374 enquiries were dealt with by "Agent DEF," both on its own and with an extra space in the name (perhaps an input inconsistency). 28 of these opportunities led to successfully secured contracts, while the remaining 346 were missed opportunities. This leads one to believe that even if 'DEF' processes a sizable number of contracts, there is potential for improvement in the conversion rate.

"Agent FEG," who oversaw managing 378 enquiries, was able to successfully secure 9 contracts. This represents a significantly lower conversion rate in comparison to that of some of the other agents, which suggests that there may be room for improvement in terms of either the sales methods or the training.

"Agent XYZ" processed a total of 324 enquiries, of which 24 were successfully turned into contracts. This may indicate a higher conversion rate when compared to some of the other agents, which may indicate efficient client management or a more successful approach to clients.

In conclusion, even though all agents have demonstrated the ability to secure contracts, there are clear differences in the rates of success that each agent achieves. While Agent DEF and FEG may need additional methods or support to achieve their goals, Agent ABC and XYZ appear to have significantly higher conversion rates. These findings highlight the significance of ongoing training, performance reviews, and even the development of collaborative initiatives as a means of increasing the percentage of contracts successfully acquired.

4.5 Comparing Hierarchical and K-means Clustering.

During the process of breaking down our dataset in order to recognise unique groups, we analysed and contrasted two widely used clustering methods: hierarchical and K-means clustering. Our findings are broken down into the following categories:

The characteristics of our data set were critical factors in determining the clustering technique that we went with. K-means clustering stood out when confronted with the complexity and scale of bigger datasets by effectively segmenting them into identifiable clusters. The Hierarchical technique excels at revealing granular, layered associations, which makes it particularly appropriate for more condensed datasets. On the other hand, the K-means clustering method excelled when faced with the complexity and scale of larger datasets.

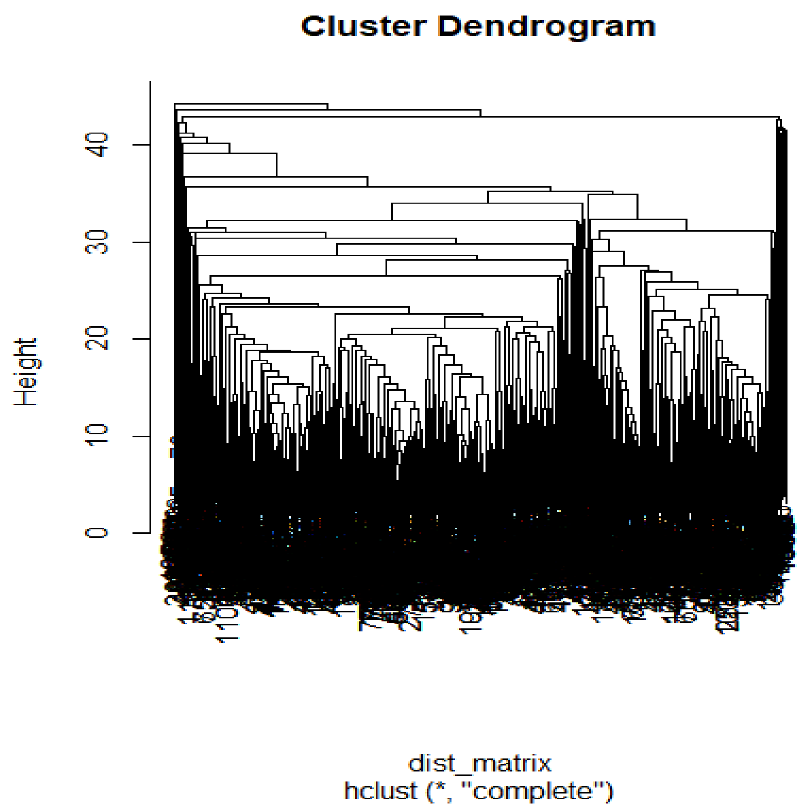


Figure 4. 1: Hierarchical clustering Tree Diagram

After performing what is known as "one hot encoding," we were presented with a massive dataset consisting of 157 variables that were spread out over 1542 observations. Each chapter in this volume posed its unique set of difficulties.

When we tried hierarchical clustering, we quickly realised that our dataset was too large for it to be effective, therefore we abandoned that approach. In order to make our dataset applicable to the Hierarchical model, we were forced to reduce it to a subset consisting of 1000 observations. Even after making this reduction, the final dendrogram (also known as a hierarchical tree) that was created from our distance matrix was quite complicated. The visual outputs were difficult to comprehend as a result of this intricacy, which was worsened by the sheer number of variables; as a result, they were probably unsuitable for the purposes of our research. The observation is made possible by referring to figure 4.1.

On the other hand, the K-means clustering method turned out to be the most reliable option for our extensive dataset. Not only did it take in the entirety of the dataset without requiring any reductions to be made, but it also made it easier to visualise the information. Because of this, we were able to make full use of our expanded dataset without having to make any concessions, which resulted in insights that were both deeper and more all-encompassing.

The cluster results that we merged with the original dataset in order to decide the cluster for enquiries, we try to display the results with the help of a bar chart in order to see the distribution of the clusters in the dataset. Both the K-means and the hierarchical clustering techniques can be applied here.

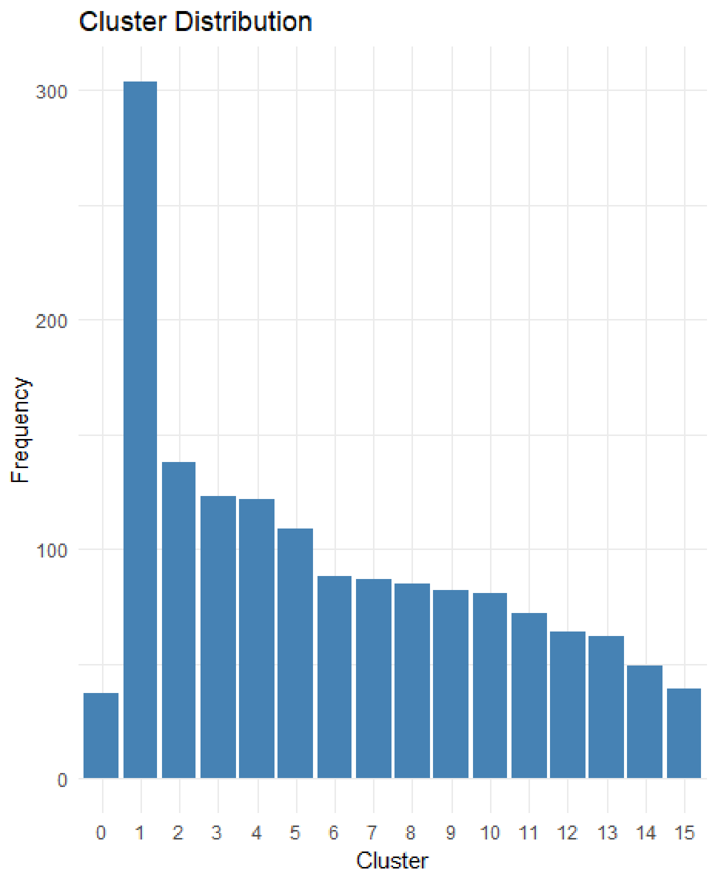


Figure 4. 2: frequency distribution of enquiries in clusters in Hierarchical clustering

The data visualisations show that hierarchical clustering and K-means clustering perform noticeably differently on our dataset, as can be shown in figures 4.2 and 4.3, respectively. These changes in performance are evident.

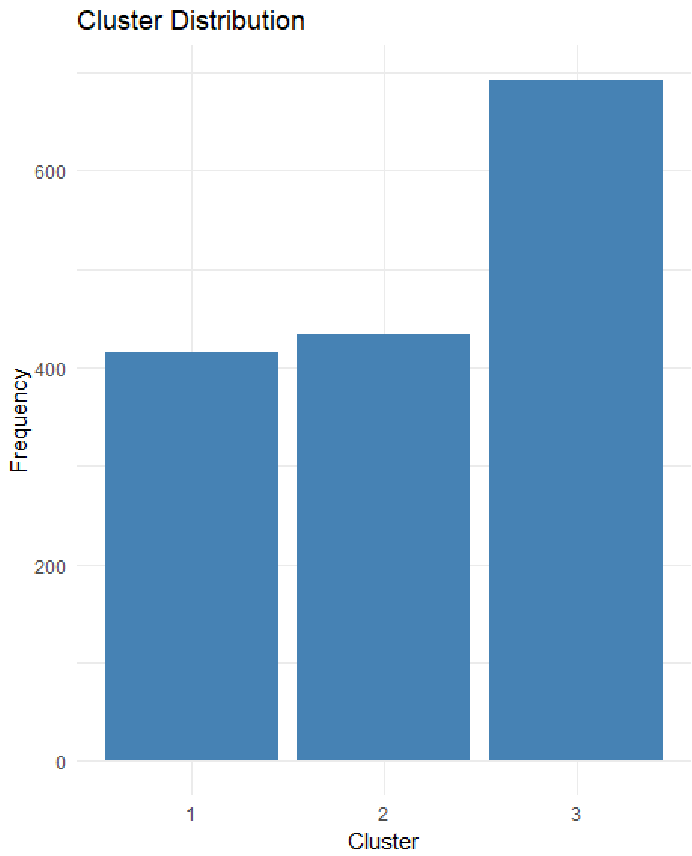


Figure 4. 3: Frequency distribution of enquiries in K means clustering.

The hierarchical clustering that was performed on our data has resulted in the formation of sixteen separate clusters, as shown in Figure 4.2. At first glance, this appears to be comprehensive; yet, there are issues presented by the distribution and frequency within these clusters. To be more specific, the spread that exists inside each cluster is not used on the complete dataset, which makes it impossible to infer unambiguous patterns or to establish interpretations that are robust. The many different clusters, in conjunction with their high frequency, make the visualisation a little bit difficult to understand. This makes it a difficult effort to grasp the essential reasons that are driving the formation of each cluster, which prevents meaningful insights from being gained.

On the other hand, when we look at Figure 4.3, which depicts the K-means clustering, we see a striking disparity between the two. Only three clearly identified clusters were produced by using this strategy on our dataset. The fact that the K-means model was run on the whole dataset, so ensuring that each observation was taken into consideration, is a considerable benefit in this scenario. The three clusters that were produced as a result are also considerably simpler to manage, which enables a more distinct visualisation and improved interpretation.

This helps in distinguishing the qualities that identify each cluster, which in turn facilitates a greater understanding of the underlying patterns that may be found in our data.

In a nutshell, hierarchical clustering provided a granulated picture with 16 clusters, but its applicability to real-world situations was restricted because of the complicated distribution and frequency of the data. In comparison, the K-means clustering provided a more meaningful and practical perspective of our dataset because to its three unique clusters that collectively covered all our observations.

4.6 Power BI visualisation

Figure 4. 4: Cluster 1 dashboard.

4.6.1 Cluster 1

During the in-depth analysis that was performed using the K-means clustering method, our dashboard proved to be an excellent tool for extracting meaningful insights from our dataset. The analysis of Figure 4.4 reveals several patterns to us. Cluster 1 is comprised of a notable 86 clients, which suggests that a major amount of the contacts and transactions are focused inside this group. When focusing on contracts that were successfully completed (POs received), the number of clients drops to 29.

The length of time it takes for a sale to go through this cluster's sales cycle typically falls somewhere in the range of 21 to 31 days. It is interesting to note that the sales cycle typically lasts between 23 and 29 days when focusing on successful contracts (PO received). This suggests that a shorter sales cycle, particularly inside this window, has a better possibility of culminating in a successful deal. This is especially true when compared to longer sales cycles.

The 'follow up 2' and 'follow up 3' sequences are the types of follow-up sequences that are most prominently involved in this cluster. This lends credence to the idea that persistent and regular follow-ups have the potential to play a crucial role in bringing about the desired outcomes of enquiries. The third follow-up seems to have a very significant influence, especially when one focuses on contracts that are successfully completed.

The range of discounts that are available throughout the entirety of Cluster 1 is anywhere from 2% to 10%, with the average discount being calculated at 5.26%. However, the prevalent discount range is rather more constrained for contracts that were successfully received, and it ranges anywhere from 5% to 8% off the original price. This points to a potential sweet spot in the discount strategy of the company, one that is likely to resonate more with clients.

The total number of enquiries in Cluster 1 is 538, and there were 45 purchase orders received. When broken down by the various types of equipment, a distinct hierarchy emerges: When it comes to successfully completed contracts, Specialty Equipment is at the head of the pack, followed by Columns and Heat Exchangers. This can imply that the company's Specialty Equipment offerings are very competitive or in-demand in the market.

Cluster 1 is comprised of equipment that is mostly made of four different materials: SS 316, MS, SS 304, and SS 316 L. Among these, SS 316 and MS are the predominate materials, accounting for a bigger percentage of the equipment that is contained inside this cluster.

In summary, the insights that were obtained from Cluster 1 shed light on certain customer behaviours, preferences, and operational factors that could shape the company's future strategies. The regularity of sales cycles, the success of follow-up sequences, and the disclosure of the most in-demand equipment and materials provide a road map for further improving sales and client engagement.

4.6.2 Cluster 2

Figure 4. 5: Cluster 2 Dashboard.

We gain some insights about the cluster 2 by examining Figure 4.5. Cluster 2 has a total of 87 clients, which places it in a position of minor superiority above Cluster 1 in terms of client count. This leads one to believe that these two clusters, when combined, account for a sizeable portion of the client interactions and sales efforts made by the company.

The length of time that the sales cycle took for Cluster 2 ranged from 22 to 33 days. A shift in the window of the ideal sales cycle has arisen, which is an intriguing finding in comparison to Cluster 1. When looking at only the contracts that were completed successfully (PO was received), the most common duration was between 24 and 30 days. This highlights the fact that a productive sales cycle that falls within this range relates to an increased likelihood of closing a successful deal.

Reiterating what was said in Cluster 1, the significance of maintaining active participation is emphasised once more in Cluster 2. Most interactions were comprised of 'follow up 2' and 'follow up 3'. Because this pattern holds true across all clusters, the significance of these follow-ups cannot be overstated when it comes to converting enquiries into successful contracts.

Cluster 2 represents 603 enquiries, of which 47 resulted in a purchase order being received. The typical discount range is anything from 5% to 8%, with the average discount rate coming in at 5.41%. This echoes the trends that were noticed in Cluster 1, pointing to a consistent discounting strategy that connects with a variety of customers subgroups.

Within Cluster 2, reactors are the most popular piece of equipment, which is more evidence of the demand they have in the market. The API Storage Tanks come in a close second, which further highlights the importance of their position in the product line-up. When we move our attention to successfully received contracts, we see that in addition to Receivers and Reactors, additional equipment types such as API Storage Tanks and Specialty Equipment emerge as top performers. This indicates that there is a diverse demand among clients who have successfully received contracts.

Cluster 2 is partial to the following four basic materials: stainless steel 316, stainless steel 304, MS, and stainless steel 316L. Among these, stainless steel type 316 stands out as the material that is requested the most, particularly in successfully converted contracts. The fact that this pattern holds true throughout both clusters is suggestive of a wider market preference for SS 316 in the fabrication of equipment.

In conclusion, the insights provided by Cluster 1 are complemented by the perspective that Cluster 2 brings to the table. Even though certain patterns, such as the significance of continuous follow-ups and the predominance of SS 316 as a preferred material, continue to be consistent, other nuances, particularly those surrounding the optimal sales cycle duration, offer differentiated insights. The prominence of Reactors and API Storage Tanks within this cluster offers a strategic direction for the product offerings of the company as well as the client interaction techniques that are being implemented.

4.6.3 Cluster 3

Figure 4. 6: Cluster 3 Dashboard

Our investigation into Cluster 3, which provided a different point of view. This was done by building on the insights gained from Clusters 1 and 2. We were able to obtain more information regarding Cluster 3 with the assistance of figure 4.6. Cluster 3 has a lower level of involvement compared to the other clusters, as evidenced by its 61 total client enquiries. It is interesting to note that although the number of enquiries has decreased, the percentage of contracts successfully received has remained quite low. Cluster 3 has only secured 9 contracts, making it the least successful of the three segments in terms of its overall performance.

Cluster 3 has a sales cycle that can last anywhere between 11 and 23 days. Notably, when looking at the purchased orders that were received, the bulk of them converged around the 20 to 22-day mark, with the 21st day being the best possible option. Given the limited amount of time that the sales cycle window covers, a fast response time is essential for this cluster.

'Follow up 1' through 'Follow up 3' are examples of engagement methods that often take a more passive approach than those found in Cluster 3. This may reflect the lower levels of involvement that have been observed in this cluster. This approach of discounting is likewise more conservative, with a range that goes from 0% to 7%.

At first look, the preferences for equipment and material trends in Cluster 3 are very similar to those in Cluster 2, even though there are some subtle differences. This shows that despite varying client habits, their equipment and material requirements are mostly the same.

In conclusion, Cluster 3 provides a unique perspective on the situation. Although it shares certain similarities with Clusters 1 and 2 in terms of equipment and material preferences, it presents a distinct set of issues due to its lower engagement and conversion rate. These insights, those involving trends in the sales cycle and engagement, might help in the process of designing customised strategies to revitalise this market segment.

4.6.4 Key Influencers for Receiving PO

Figure 4. 7: Key Influencers for Receiving PO

During our investigation into the dataset, which was supported by the use of key influencers, the distinct patterns shown in Figure 4.7 emerged. These patterns have the potential to greatly impact the conversion of an enquiry into a purchase order (PO).

Consistent engagement with clients is crucial, as shown by the data. It is important to note that after three follow-ups, the possibility of an enquiry turning into a purchase order is significantly increased. In fact, the conversion rate may be regarded as being practically limitless at this point. This highlights the need of maintaining open and timely contact with clients in order to nudge them towards making a positive purchasing decision.

Alterations to prices have a discernible impact on conversion rates. Offering discounts in the region of 6% to 8% has a significant impact on the likelihood that an enquiry will result in a purchase order. This optimal level of discounting can be used to boost conversion rates and sales strategies without having a substantial impact on profit margins. This "sweet spot" of discounting can be used to optimise sales methods.

The channel by which a client communicates with a company is also extremely important. In comparison to enquiries that are started through other channels, phone calls have a higher chance of developing into purchase orders. This may imply that more direct and personal engagement with a client, such as a phone conversation, can lead to more decisive and speedier decisions from that client.

Cluster 1 distinguishes out from the other clusters in terms of its capacity for conversion, as shown in Figure 4.7. When compared to enquiries that fit within other clusters, those that fall into this cluster have a significantly higher likelihood of developing into PO received. This lends credence to the notion that the qualities and behaviour patterns of clients in Cluster 1 are more in line with the company's offerings or sales strategy.

To summarise, our findings shed light on how critically important it is to engage people in ways that are both purposeful and well-informed. Conversion rates can be greatly increased by dramatically driving up the conversion rates, turning enquiries into profitable purchase orders. This can be done through regular follow-ups, effective discounting, capitalising on direct ways of communication, and understanding the intricacies of each cluster.

4.6.5 Comparison of all Clusters

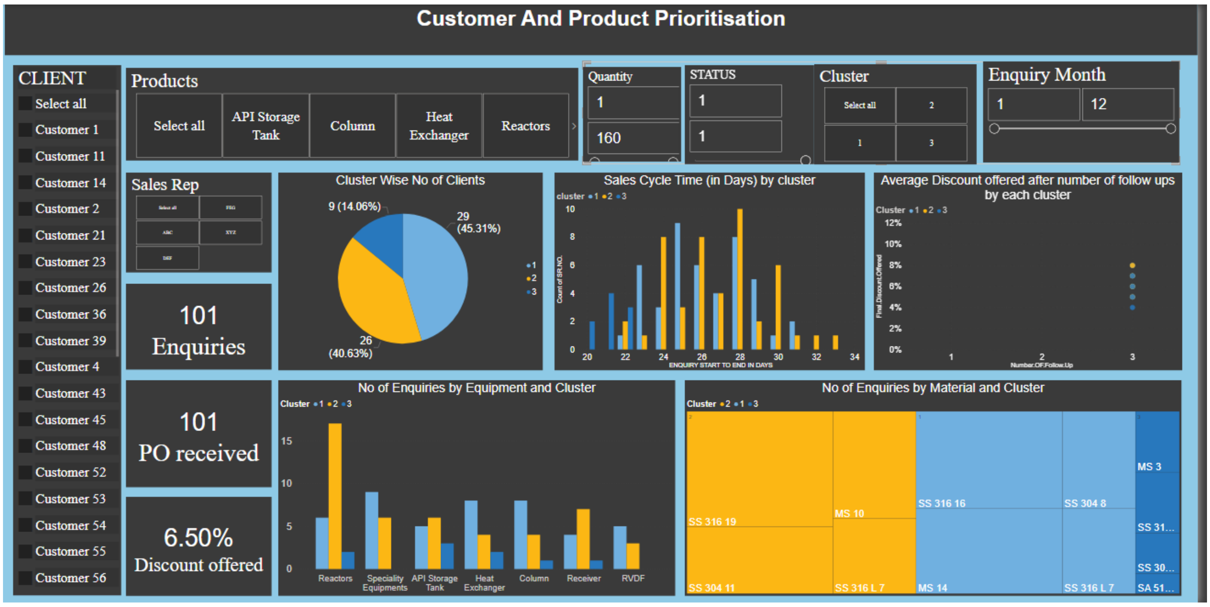


Figure 4. 8: the comparison of clusters.

The chart that can be found in Figure 4.8 is quite helpful in assisting us in better comprehending our research. It demonstrates rather clearly the distinctions that exist between the clusters based on the equipment type (product) that each one favours. This enables businesses to determine which product categories are more well-liked by customer segments.

The chart also demonstrates the significance of classifying customers in order to prioritise which kind of enquiries to address first. Simply put, some customers are more inclined to make a buy right away, while others might take a little bit more time. Having this information allows businesses to determine how quickly they should respond to each enquiry.

When it comes to making offers on sales, Figure 4.8 also provides insights on discounts. Different types of customers anticipate receiving various types of discounts. By considering the preferences of each demographic, businesses can improve the planning of their discount offerings. They might give unique discounts to categories of customers as an alternative to providing everybody with the same discount.

5. Discussion & Limitation

5.1 Discussion

In the context of the manufacturing industry, the purpose of this research was to investigate the ways in which data analytics, in integration with Power BI visualisations, might be utilised to prioritise customer and product enquiries. Based on an analysis of a dummy dataset provided by a manufacturing company, a number of important findings have arisen that provide answers to our research questions and throw light on the hypotheses.

RQ 1: How can data analytics, with the Aid of Power BI visualization be utilised to prioritize customer and Product Enquiries

The clustering analysis revealed separate groups of customers with distinct behavioural patterns, unique requirements, and varying conversion rates for contracts. The method to strategic prioritisation is built on top of this segmentation, which serves as the foundation. Customers who fell into Clusters 1 and 2 demonstrated higher levels of engagement and a larger possibility of contract conversion. A significant contribution was made by variables such as the sales cycle duration, follow-up patterns, and discounting technique. Cluster 3 customers, on the other hand, demonstrated lower engagement and contract conversion rates, which necessitated the development of customised strategies in order to revitalise this cluster.

The visualisations in Power BI gave crucial insights that allowed us to identify the features of these clusters and develop data-driven strategies for prioritising our customers' needs. The interactive dashboards made it possible to perform multidimensional analysis of sales indicators, which uncovered subtle differences between the different clusters. Users were given the ability to filter on important dimensions and gain insights relevant to a cluster thanks to slicers. There were charts that provided a summary of the various clusters' equipment need, material preferences, and follow-up effectiveness. The sales rep slicer can be used to gain a better understanding of the sales rep's specialism in terms of cluster, customer, and product enquiry. The sales representatives of a certain cluster could be designated as having a particular specialisation or as being skilled at addressing difficult client segments if they so want. This can maximise the cluster's transformation potential and transform the poor conversion rate. These visualisations took the raw data and turned it into insights that could be used to influence the prioritisation of customers.

RQ 2: How does enquiry duration impact sales conversion, and can it be used as an effective metric for prioritization?

According to the findings of the analysis, conversion rates were higher for shorter enquiry cycles, which ranged anywhere from 21 to 30 days in length depending on the cluster. The "sweet spot" for Cluster 1 was between 23 and 29 days, the "sweet spot" for Cluster 2 was between 24 and 30 days, and the "sweet spot" for Cluster 3 was merely 21 and 22 days. When only successful cluster conversions were analysed across all clusters, this tendency remained persistent. Because of this, the duration of the enquiry emerges as a candidate for an effective indicator for prioritisation, with shorter cycles suggesting a greater level of priority. In order to make the most of the conversion potential presented by short-cycle enquiries, sales techniques may include the provision of accelerated responses, quotes, and follow-ups.

RQ 3: How do visualization tools like Power BI influence decision-making processes in customer and product enquiry management?

As shown by the results of our research, the utilisation of Power BI visualisations allowed for the rapid comprehension of important metrics, patterns, and relationships included within the dataset. A multifaceted analysis was made possible thanks to interactive elements such as slicers. The use of charts, cards, and graphs together made it possible to recognise patterns in the data quickly and easily. When compared to the overwhelming results of hierarchical clustering, the comprehension of the data was greatly simplified as a result of the dataset being divided into just three clusters. Because of these benefits, the process of decision-making was facilitated by the transformation of complicated data into visuals that are easy to understand. Because of this, Power BI shows to be a very significant asset in the process of prioritising customer and product enquiry as well as more general decision-making.

A focused sales approach that is tailored to the demands of each cluster can be achieved by segmenting customers based on their enquiry and purchase patterns. Customers with high, medium, and low engagement levels can all benefit from different sales methods. Establishing a standard for accelerating answers and follow-ups can be accomplished by determining the ideal duration of the sales cycle for each market sector. Based on this cycle, sales representatives can be trained to recognise enquiries that have a high potential for conversion. Increasing conversion rates can be accomplished by standardising the number of follow-ups and discount percentages based on the preferences of the clusters. Business insights may be more easily shared throughout an organisation using interactive dashboards built with Power BI. To summarise, the sales performance of manufacturing companies may be considerably improved by utilising a focused sales approach that is founded on data-driven customer segmentation, the streamlining of sales cycles, the optimisation of pricing, and the deployment of BI tools.

5.2 Return on Investment

The monthly licencing cost for Power BI is around £8.2 per user for a small firm. This comes out to around $394 per year when divided among four users. If the initiatives, which are enabled by Power BI analytics, are successful in increasing annual revenues by even £15,000 or £20,000, the return on investment for Power BI would be greater than 50 times the initial expenditure. When an organisation has integrated analytics and interactive visuals, decision-making throughout the organisation can be improved, which creates secondary benefits like as cost and time savings in addition to the direct cost savings. The platform makes it possible to aggregate data from multiple sources into a single view in order to gain holistic insights. When compared to the analysis of several sources, this makes the value extraction process much more effective.

In addition, the analysis yielded enough evidence to refute the null hypothesis, which stated that the utilisation of Power BI visualisations does not considerably improve enquiry prioritisation. Instead, we discovered substantial support for the alternative hypothesis, which states that the prioritisation of customer and product enquiry data in manufacturing companies can be significantly improved by using Power BI visualisations. Granular and actionable insights on strategic prioritisation were made possible thanks to the interactive dashboards, simplified cluster visuals, and data manipulation capabilities.

5.3 Limitations

Although this research yielded some promising findings, it is important to remember that there are some limitations that need to be considered. First, the use of a dummy dataset and the cost of the equipment make it harder to confirm findings in the real world, even though these steps are necessary to protect confidential information. The insights could be better used in the real world if data-driven strategies were tested using live data from the manufacturing company.

Additionally, this analysis only used the data from the past enquiry. By integrating data in real time as new enquiries come in, it might be possible to enhance the prioritisation skills.

The data from the past shows a still picture of the market, which is always moving. Using a live dashboard that pulls in real-time enquiry data could possibly improve the value gained.

In the context of the developed model, a significant limitation emerges when confronted with enquiries from new customers representing new companies. Since these entities have no prior data in the system, the model is challenged in making accurate predictions or categorizing these enquiries into any predefined clusters. This lack of historical data may potentially cause missed opportunities or misjudgments. However, the company has the avenue to address this gap by prioritizing such enquiries. Drawing from past purchase orders received from other customers and leveraging historical data can be instrumental in understanding the nature and potential of the enquiry. By doing so, the company not only navigates the limitation of the model but also harnesses an opportunity to acquire and nurture a new customer relationship.

The number of clusters had to be known ahead of time when using the K-means clustering method. Even though we chose the three clusters in a methodical way, there are other segmentation strategies that would have given us more insights. Other, previously undiscovered patterns may be found by experimenting with a variety of cluster numbers.

5.4 Future Scope

Power BI has undoubtedly revolutionised the field of data visualisation by offering a user-friendly interface and the ability to show data in a way that is both more interactive and easier to understand. To keep up with the constantly changing needs of current data analysis, Power BI may need to be further advanced in some areas.

The integration of advanced analytics and artificial intelligence (AI) is an exciting possibility. The addition of capabilities for advanced AI-generated code could greatly increase Power BI's usefulness, even though it is already capable of producing effective visualisations. These capabilities can lead to both predictive and prescriptive insights, which help users not only understand past patterns but also predict future trends and find out what to do.

Also, the use of Natural Language Processing (NLP) and Natural Language Generation (NLG) could change how Power BI display users interact with them. A chatbot, which is made possible by these technologies, is one of the useful tools that can be used. This chatbot can quickly understand user questions and create visualisations based on them. This will make data analysis more democratic in the long run. Users who don't know much about technology will benefit a lot from this kind of feature because it will help them gain insights without having to use complicated analysis tools.

Power BI isn't as complicated as some other tools for data science, which is another of its flaws. Its current analytical capabilities are good enough for basic analysis, but it would be better if it could use more complex models. For basic analysis, its analytical capabilities are strong. By connecting data visualisation and in-depth data science, Power BI can usher in a new age of data analysis. Users can use intricate data methods and models right inside the software thanks to this potential.

Finally, Power BI still has a long way to go when it comes to real-time data integration, especially with Enterprise Resource Planning (ERP) systems. Decision-makers would always have access to the most up-to-date information if ERP integration were seamless, allowing for quick data retrieval and real-time analysis. For businesses moving towards decision making and running operations in real time, this integration is crucial.

In conclusion, Power BI's current features have made a big difference in data visualisation, but they haven't even come close to being used to their fullest potential. Through the integration of artificial intelligence, advanced analytics, natural language processing, natural language generation, and real-time data retrieval, it is possible to redefine its capabilities. This opens the door for a comprehensive analytical tool that can handle both basic and real-time data needs. We have a responsibility as experts to be able to use these new technologies and contribute to the direction of data analysis and visualisation in the future.

6. Conclusion

The goal of this research was to find out how data analytics and Power BI visualisations can be used to prioritise customer and product enquiries in the manufacturing industry. After performing a thorough analysis of a dummy data set provided by a manufacturing company, the study came to a number of important findings.

The most important findings show that clustering algorithms can effectively divide customers into groups based on their enquiry patterns, needs, and buying habits. This customer segmentation, which is based on data, is used to create sales strategies that are tailored to each customer group. Cluster 1 and 2 customers were more engaged and had higher contract conversion rates than Cluster 3 customers.

The duration of the enquiry time cycle is another measure that could be quite important for prioritising sales. The best sales cycle duration windows typically range from 21 to 30 days, but this could change based on the cluster, according to the research's findings. The best conversion probability was found during these shorter cycle enquiry cycles. This gives sales teams a way to focus on high-value enquiries that is based on data and is not subjective.

The visualisations in Power BI also made it possible to do an interactive, multidimensional analysis of the dataset. Combining clustering algorithms with the interaction of Power BI was what turned complex data into insights that were easy to understand and could be use. Features like slicers made it possible to look more closely at patterns that were only found in certain clusters.

In general, the methods shown in this research show how powerful the combination of predictive analytics and interactive visualisations can be for getting intelligence that can be used. Even though it has some flaws, the study gives manufacturing companies a plan for how to improve their sales by focusing on data-centric enquiry management.

The most important thing to learn from this is that manufacturing companies can no longer depend on their gut feelings or generic answers to customer enquiries. If they use data science and visualisation tools, they can get a big advantage over their competitors by developing strategies that are based on data-backed customer and product insights.

In conclusion, the findings of this research show the enormous potential that data-driven enquiry decision making has for decision management in the manufacturing sector. the integration of technology, particularly through live dashboards, offers substantial benefits in managing enquiries. By utilizing these dashboards, managers can strategically prioritize enquiries and direct them to cluster-specific specialists, leveraging their unique expertise. This proactive approach not only streamlines the follow-up process but also enables the company to plan discount strategies in advance. Consequently, this preparation provides ample room for negotiation, fostering better relationships with customers. Ultimately, such strategic use of technology can lead to securing more contracts, thus boosting profits for the company.

7. References

Akyildiz, B., Kadaifci, C. & Topcu, I., 2015. A decision framework proposal for customer order prioritization: A case study for a structural steel company. *Int. J. Production Economics,* Volume 169, pp. 21-30.

Anon., 2018. *Wikipedia.* [Online]
Available at: https://en.wikipedia.org/wiki/Business_intelligence#cite_note-11

Arredondo, F. & Martinez, E., 2010. Learning and adaptation of a policy for dynamic order acceptance in make-to-order manufacturing. *Comput. Ind. Eng.,* 58(1), pp. 70-83.

Belhadi, A. et al., 2019. Understanding Big Data Analytics for Manufacturing Processes: Insights. *Computers & Industrial Engineering,* Volume 137.

Cao, Y. & Gruca, T. S., 2005. Reducing Adverse Selection ThroughCustomer Relationship Management. 69(4), pp. 1-265.

Djerdjouri, M., 2020. Data and Business Intelligence Systems for Competitive Advantage: prospects,challenges, and real-world applications. *Mercados y Negocios.*

EDWARDS, A. W. F. & CAVALLI-SFORZA, L. L., 1965. A METHOD FOR CLUSTER ANALYSIS. *International Laboratory of Genetics and Biophysic,* 21(2), pp. 362-375.

Eriksson, M. & Ferwerda, B., 2021. Towards a User Experience Framework for Business Intelligence. *JOURNAL OF COMPUTER INFORMATION SYSTEMS,* 61(5), p. 428–437.

Evelson, B., 2023. *Forrester.* [Online]
Available at:
https://web.archive.org/web/20160806102752/http://blogs.forrester.com/boris_evelson/10-04-29-want_know_what_forresters_lead_data_analysts_are_thinking_about_bi_and_data_domain

Frades, I. & Mattiesen, R., 2009. Overview on Techniques in Cluster Analysis. *Bioinformatics Methods in Clinical Research,* Volume 593, pp. 81-107.

Haber, N. & Fargnoli, M., 2019. Prioritizing customer requirements in a product-service system (PSS) context. *The TQM Journal,* 31(2).

Homburg, C., Droll, M. & Totzek, D., 2008. Customer Prioritization: Does it Pay off, and how Should it be Implemented?. *Journal of Marketing,* 72(5), pp. 110-130.

Huang, C.-Y., 2012. To model, or not to model: Forecasting for customer prioritization. *International Journal of Forecasting,* Volume 28, pp. 497-506.

Kreuzer, T., Röglinger, M. & Rupprecht, L., 2020. Customer-centric prioritization of process improvement projects. *Decision Support Systems,* Volume 133.

Mathur, M. & Kumar, S., 2013. Customer Retention Through Prioritization: Integrating Time-Dependent Context of Relationship Dynamics. *Journal of International Consumer Marketing,* 25(5), pp. 323-343.

Mestry, S., Damodaran, P. & Chen, C.-S., 2011. A branch and price solution approach for order acceptance and capacity planning in make-to-order operations. *Eur. J. Oper. Res.,* 211(3), pp. 480-495.

Negash, S., 2004. Business Intelligence. *AIS Journal (Communications of the Association for Information Systems),* Volume 13.

Saleem, H. et al., 2019. Strategic Data Driven Approach to Improve Conversion. *International Journal of Scientific & Engineering Research,* 10(4), pp. 588-593.

Soni, P. & Bhardwaj, R., 2022. Examining Attitude of Indian Consumers toward Retail Loyalty Programs:A Cluster Analysis Approach. *International Journal of Economics and Business Reaserach.*

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S. & Satoto, B. D., 2018. Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering,* Issue 336.

Trappey, C. V., Trappey, A. J., Chang, A.-C. & Huang, A. Y., 2010. Clustering analysis prioritization. *Industrial Management & Data Systems,* 110(5), pp. 731-743.

Widjaja, S. & Mauritsius, T., 2019. THE DEVELOPMENT OF PERFORMANCE DASHBOARD VISUALIZATION WITH POWER BI AS PLATFORM. *International Journal of Mechanical Engineering and Technology (IJMET),* 10(05), pp. 235-249.

Yenaiaras, V. & Kaya, I., 2021. Customer prioritization, product complexity and business ties: implications for job stress and customer service performance. *Journal of Business & Industrial Marketing,* 37(2).

Yiu, L. D., Yeung, A. C. & Jong, A. P., 2020. Business intelligence systems and operational capability: an empirical analysis of high-tech sectors. *Industrial Management & Data Systems,* 120(6), pp. 1195-1215.

Yiu, L. M. D., Yeung, A. C. L. & Cheng, T. C. E., 2021. The impact of business intelligence systems on profitability and risks of firms. *International Journal of Production Research,* 59(13), pp. 3951-3974.

Zhao, X. & Keikhosrokiani, P., 2022. Sales Prediction and Product Recommendation Model Through. *Computers, Materials & Continua,* 70(2), pp. 3855-3874.

Zohuri, B. & Moghaddam, M., 2020. From Business Intelligence to Artificial Intelligence. *Journal of Material Sciences & Manufacturing Research,* 1(1), pp. 1-10.

8.Appendix

```
###############################################################################################################
###### K means Clustering #####################################
###############################################################################################################
# Compute total within-cluster sum of square
wss <- sapply(1:10, function(k){
  kmeans(data_clustred, centers = k, nstart = 25)$tot.withinss
})


# Plot the elbow curve
plot(1:10, wss, type="b", pch = 19, frame = FALSE,
     xlab="Number of clusters k",
     ylab="Total within-clusters sum of squares")

# with elbow curve we got to know k=3
colnames(data_clustred)
data_clustred[is.na(data_clustred)] <- sapply(data_clustred, mean, na.rm = TRUE)
data_clustred[sapply(data_clustred, is.infinite)] <- NA


# 2. Apply k-means clustering
kmeans_result <- kmeans(data_clustred, centers = 3, nstart = 25)  # Choose 3 centers, and try 25 random sets of initial centers
data_clustred$cluster <- as.factor(kmeans_result$cluster)  # Assign cluster results back to the data frame


# Identify columns with zero variance
zero_vars <- nearZeroVar(data_clustred, saveMetrics= TRUE)

# Select only the columns with variance
data_clustred <- data_clustred[, zero_vars$nzv == FALSE]


# Keeping only numeric columns from data_clustred
numeric_data <- data_clustred[, sapply(data_clustred, is.numeric)]

# Visualizing clusters
cluster2plot <- fviz_cluster(kmeans_result, data = numeric_data)
print(cluster2plot)
library(GGally)

ggpairs(data_clustred, columns = 1:10, aes(color = cluster))  # Assuming you're plotting the first four features.

# for specific cunctions
library(ggplot2)

str(data_clustred)
ggplot(data_clustred, aes(x = Number_OF_Follow_Up , y = ENQUIRY_START_TO_END_IN_DAYS, color = cluster)) +
  geom_point() +
  ggtitle("2D K-means Clustering") +
  theme_minimal()

ggplot(data_clustred, aes(x = ENQUIRY_MONTH , y = STATUS, color = cluster)) +
  geom_point() +
  ggtitle("2D K-means Clustering") +
  theme_minimal()
```

Appendix 1 : code snippet for K-means Clustering

```r
##########################################################################################################################
####  Hierarchical clustering  ##########################################################################################
##########################################################################################################################

data_clustred <- scaled_data
# Compute the distance matrix
dist_matrix <- dist(data_clustred, method = "euclidean")

# Hierarchical clustering using the complete linkage method
hclust_result <- hclust(dist_matrix, method = "complete")

# Plot the dendrogram
plot(hclust_result)

# Cut the dendrogram to create 3 clusters
clusters <- cutree(hclust_result, k = 3)

# Use the dendextend package to color and label branches
dend <- as.dendrogram(hclust_result)
dend <- color_branches(dend, k = 3) # Color branches by clusters

# Plot the enhanced dendrogram
plot(dend)

set.seed(123)  # for reproducibility
data_sample <- data_clustred[sample(1:nrow(data_clustred), 1000), ]  # Adjust the number based on your data size and memory capacity.
clusters <- cutreeDynamic(dend = hclust_result,method = "tree")



# Add clusters to data
data_clustred$hcluster_dynamic <- as.factor(clusters)


# Add clusters to data
data_clustred$hcluster_dynamic <- as.factor(clusters)
data_clustered_orig<- data
# Convert percentage column to numeric
data_clustered_orig$Final.Discount.Offered <- as.numeric(gsub("%", "", data_clustered_orig$Final.Discount.Offered)) / 100
data_clustered_orig$cluster <- data_clustred$hcluster_dynamic
barplot(table(data_clustered_orig$cluster), main="cluster Distribution", xlab="Cluster", ylab="Frequency")
library(ggplot2)


ggplot(data_clustered_orig, aes(x = cluster)) +
  geom_bar(fill = "steelblue") +
  labs(title = "Cluster Distribution", x = "Cluster", y = "Frequency") +
  theme_minimal()

ggplot(data_clustered_orig, aes(x = Number.OF.Follow.Up, y = ENQUIRY.START.TO.END.IN.DAYS, color = cluster)) +
  geom_point(alpha = 0.6, size = 3) +
  labs(title = "Visualization of Clusters", x = "Number of Follow Up", y = "Enquiry Duration (Days)") +
  theme_minimal()
```

Appendix 2: Code Snippet For Hierarchical Clustering

Dissertation Topic Name Change

I began my dissertation on "The Integration of AI-Generated Code in Business Intelligence Systems for Data Analysis and Decision-Making: Customer and Product Prioritizing in the Manufacturing Industry". I managed to write machine learning code and successfully integrated it into Power BI, just as I described in my dissertation. However, my primary goal of integrating NLP and NLG functions with Power BI wasn't achieved, as I faced challenges with that integration.