Submitted in part fulfilment of the requirements for the degree of Master of Science in Business Analytics

UNIVERSITY OF SURREY

SURREY BUSINESS SCHOOL

MSC BUSINESS ANALYTICS

Academic year: - 2023-24

Inventory Optimisation using ABC XYZ Classification

by

Nimish Datkhile

URN: - 6794120

Word Count: - 13,594

1 Executive Summary

This dissertation investigates the optimisation of inventory management through the integration of the ABC-XYZ classification method with K-Means clustering. The ABC-XYZ classification is a well-established technique that categorizes inventory based on value (ABC) and demand variability (XYZ). Traditionally, this method has been effective in enhancing inventory control, reducing holding costs, and improving service levels. However, this research expands on the traditional approach by incorporating K-Means clustering to refine the classification process, offering a more precise and dynamic inventory management strategy.

The study confirmed that the ABC-XYZ classification method, when enhanced with K-Means clustering, significantly improves inventory management. The integration allows for more accurate categorization of inventory items, aligning them better with appropriate Make-to-Order (MTO) and Make-to-Stock (MTS) strategies. The results showed that high-value, stable-demand items (AX) could be effectively managed with MTS strategies, ensuring their availability, while low-value, high-variability items (CZ) benefit from MTO strategies, minimizing inventory costs and reducing the risk of obsolescence.

The research employed a structured methodology using the CRISP-DM framework, which guided the data mining and analysis process. The dataset, comprising inventory transaction records from a UK-based online retailer, was cleaned, normalized, and subjected to K-Means clustering. The clustering process categorized inventory items into nine distinct groups based on their sales patterns and demand variability, corresponding to the ABC-XYZ classification. The study used advanced data analysis techniques, including the calculation of the coefficient of variation (CoV) to assess demand variability, which was crucial in refining the XYZ classification.

The findings of this research have significant implications for inventory management. The enhanced ABC-XYZ classification, combined with K-Means clustering, provides a more nuanced understanding of inventory dynamics, leading to more informed decision-making. By aligning inventory strategies with the specific characteristics of each item, businesses can optimise stock levels, reduce holding costs, and improve overall operational efficiency. However, the research also identified limitations, including the heavy reliance on historical sales data, which may not account for future trends or unexpected changes in demand.

While the study demonstrates the benefits of integrating K-Means clustering with the ABC-XYZ classification, it also highlights several limitations. The model's reliance on historical data may lead to inaccuracies in classification, especially in rapidly changing markets. The lack of real-time data integration could result in delays in responding to changes in demand or supply chain disruptions. Additionally, the scalability of the K-Means model may be limited when applied to larger datasets, and the sensitivity of the clustering process to initial centroid selection could introduce bias.

Further research should focus on incorporating real-time data into the classification process to enhance the model's adaptability and responsiveness. Exploring alternative clustering techniques, such as hierarchical clustering or density-based clustering, could also provide better scalability and stability for large-scale applications. Moreover, validating the proposed methods in different industrial contexts or with larger datasets would help assess their generalizability and robustness across various inventory management scenarios. Integrating external factors such as market trends, seasonality, and economic shifts into the model could further improve demand forecasting and inventory optimisation strategies.

This dissertation provides a comprehensive analysis of how integrating ABC-XYZ classification with K-Means clustering can optimise inventory management. The research confirms that this approach enhances inventory control, reduces costs, and improves service levels, offering a powerful tool for businesses seeking to optimise their inventory management practices. However, addressing the identified limitations through further research and development is crucial for maximizing the model's effectiveness in diverse real-world scenarios.

2 Declaration of Originality

I hereby declare that this dissertation 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.*

Name: - Nimish Arvind Datkhile

Date: - 08 / 09 / 2024

Signature: -

3 Acknowledgement

I would like to express my sincere gratitude to my supervisor, Dr. Wolfgang Garn for their invaluable guidance and support throughout this dissertation. Their expertise in data analytics, machine learning, and inventory management was instrumental in shaping the research methodology and enhancing the analysis. The integration of K-Means clustering with ABC XYZ classification, and the thoughtful approach to Make-to-Order (MTO) and Make-to-Stock (MTS) strategies, was possible due to their insightful advice. Their encouragement and feedback were essential in ensuring the successful completion of this project.

Furthermore, I am deeply grateful to Dr. Colin Fu, the program director, whose profound ideas and guidance have helped me to embark on this academic journey. The informative workshops held under his supervision proved useful.

I'd like to convey my heartfelt gratitude to these distinguished individuals for their insightful contributions for this dissertation. Their advice has extended my perspective and enriched my journey. Overall, the assistance and wisdom conferred to me by these remarkable individuals have been crucial in influencing the course of my studies, for which I am deeply thankful.

Table of Contents

List of Tables

Table of Figures

4 Introduction

4.1 Background

Inventory Optimisation is a fundamental aspect of supply chain management that involves determining the optimal quantity and timing of inventory to align supply with demand while minimising costs and maximising service levels. Effective inventory optimisation ensures that companies can meet customer demand without overstocking or understocking, which can lead to lost sales, increased holding costs, and reduced profitability (Garn, 2020, pp.199–203).

According to Yeung (2023), there are several techniques involved in Inventory Optimisation: -

- Demand Forecasting: - Using historical data, market trends, and statistical models to predict future demand accurately. This helps in planning inventory levels more effectively.
- Economic Order Quantity (EOQ): - A formula that determines the optimal order quantity minimising the total cost of ordering and holding inventory.
- Safety Stock Calculation: - Keeping additional stock to mitigate the risk of stockouts caused by demand variability and lead time uncertainties.
- ABC Analysis: - Categorizing inventory into three categories (A, B, and C) based on their importance, with 'A' items being the most valuable. This helps prioritize inventory management efforts.
- Just-In-Time (JIT): - A strategy to increase efficiency by receiving goods only as they are needed in the production process, reducing inventory holding costs.
- Inventory Turnover Ratio: - Monitoring this ratio helps assess how efficiently inventory is being used and replenished.

4.2 Research Gap

There are several challenges that occurs in inventory optimisation. According to Ross (2015), some of the challenges faced by inventory optimisation are as follows: -

Challenges in Inventory Optimisation: -

- Overstocking and Understocking: - Balancing inventory levels is crucial yet challenging. Overstocking ties up capital and increases the holding costs, while understocking can lead to stockouts, lost sales, and customer dissatisfaction due to unavailability.
- Demand Variability: - Fluctuating demand patterns make it difficult to predict the right amount of inventory to keep, leading to either excess inventory or insufficient stock to meet customer needs.
- SKU Proliferation: - With an increasing number of Stock Keeping Units (SKUs), managing and optimising inventory becomes more complex, requiring more granular and tighter control over each item.
- Inefficient Resource Allocation: - Without clear prioritization, resources may be evenly spread across all inventory items, leading to inefficient management and focus on low-value or low-impact items.
- Complex Forecasting: - Different items may have varying demand patterns, making it challenging to apply a one-size-fits-all forecasting method. Some items require sophisticated forecasting techniques due to high demand variability, while others may need simpler methods.

The optimisation of inventory management through the ABC-XYZ classification method presents a significant research gap, particularly in the integration of demand forecasting and the application of advanced analytical techniques. While the ABC-XYZ classification has been recognized for its ability to enhance inventory control by categorizing items based on their value and demand variability, existing papers indicates that there are still unexplored dimensions that could further improve its efficacy.

One notable gap is the integration of demand forecasts into the ABC-XYZ analysis. Scholz-Reiter et al. (2012) findings suggested that by incorporating demand forecasts, one can enhance item classification performance, yet the practical implications of such integration remain underexplored. This indicates a need for further research into how demand forecasting can be systematically integrated into the ABC-XYZ framework to optimise inventory levels.

Moreover, the limitations of the ABC classification alone have been highlighted by (Stojanovic & Regodić, 2017), who argue that the absence of a demand analysis can hinder effective inventory management. By merging ABC with XYZ classifications, a more comprehensive multi-criteria approach can be developed. However, the literature lacks detailed methodologies on how to effectively implement this integrated approach in various industries, particularly in sectors with fluctuating demand patterns.

Additionally, the application of machine learning in refining the ABC-XYZ classification presents another significant research opportunity. Recent studies, such as those by (Qaffas et al., 2023) suggests that advanced analytical frameworks can enhance decision-making processes in inventory classification. However, the practical application of these technologies in real-world scenarios, particularly in dynamic environments, remains largely unaddressed. There is a need for empirical studies that explore how these technologies can be effectively utilized to improve inventory classification and management.

And lastly, as we know that the ABC-XYZ classification method is widely recognized for its effectiveness in inventory management, particularly in optimising storage costs and improving demand forecasting (Herlambang & Parung, 2021). However, the existing research primarily focuses on the application of these classifications in isolation, without adequately addressing how they can be integrated into MTO and MTS strategies. For instance, while we discuss the implementation of workload control in MTO-MTS environments, they do not delve into how ABC-XYZ classifications can enhance this process (Kundu et al., 2022). This indicates a need for further exploration into how these classifications can be adapted to support the unique demands of MTO and MTS systems.

4.3 Research Contribution

The research contributes to the academic understanding of inventory management by offering a data-driven methodology that enhances the precision of inventory strategies, leading to improved operational efficiency. The findings are valuable for both theoretical advancements and practical applications, particularly in optimising inventory management systems across various industries. According to Jenkins (2022), inventory optimisation helps in following ways: -

Benefits of Inventory Optimisation: -

- Cost Reduction: - Optimising inventory levels help minimize holding costs, reduce waste, and lower the capital tied up in stock, leading to overall cost savings.
- Improved Service Levels: - Ensuring that the right products are available at the right time enhances customer satisfaction and loyalty.
- Enhanced Efficiency: - Streamlined inventory management processes reduce waste, improve productivity, and enhance overall supply chain efficiency.
- Better Decision-Making: -With accurate data and insights, businesses can make informed decisions about purchasing, production, and inventory policies, aligning inventory levels with market demand.
- Competitive Advantage: - Companies with optimised inventory systems can respond more quickly to market changes and customer demands, gaining a competitive edge.

This dissertation caters to these challenges of Inventory Optimisation by culminating ABC XYZ classification method with MTS (Make – To – Stock) and MTO (Make – To – Order) strategies. a more focused and strategic approach to managing inventory, resulting in cost savings, improved service levels, enhanced cash flow, better decision-making, and increased operational efficiency and enables a more focused and strategic approach to managing inventory, resulting in cost savings, improved service levels, enhanced cash flow, better decision-making, and increased operational efficiency (Pandya and Thakkar, 2016).

Inventory optimisation is a critical aspect of supply chain management, ensuring that the right products are available at the right time while minimizing costs. The ABC-XYZ classification method has emerged as one of the most effective approaches for optimising inventory, combining the principles of the Pareto Principle (ABC analysis) and demand variability (XYZ analysis). ABC analysis categorizes inventory into three classes: A items (high value, low quantity), B items (moderate value, moderate quantity), and C items (low value, high quantity). XYZ analysis, on the other hand, classifies items based on demand variability: X items (steady demand), Y items (moderate variability), and Z items (high variability. By integrating these two methods, inventory managers can develop tailored strategies for each category, thereby improving forecasting accuracy, optimising stock levels, reducing holding costs, and enhancing overall supply chain efficiency. This integrated approach ensures that critical items are consistently available while minimizing excess inventory and associated costs for less critical items (Pandya and Thakkar, 2016).

To further optimise inventory management, this research explores the integration of ABC-XYZ classification with Make-to-Order (MTO) and Make-to-Stock (MTS) strategies. MTO and MTS are two common production strategies that determine how customer orders are fulfilled. MTO involves producing goods only after receiving a specific order, while MTS involves producing goods in advance and storing them until an order is placed (Hayes, 2020). By aligning MTO and MTS strategies with ABC-XYZ classifications, a hybrid inventory policy can be developed. For instance, items categorized as AX or BX should be managed with an MTS approach due to their predictable demand and higher value, while items categorized as CZ or BZ should be managed with an MTO approach due to their unpredictable demand and lower value. This combined strategy ensures that inventory management remains responsive to changing demand patterns and market conditions, thereby reducing inefficiencies and enhancing overall performance (Roser, 2021).

The contributions of this research are threefold. First, it investigates the effectiveness of the ABC-XYZ classification method in categorizing inventory items based on their value and demand variability, offering a nuanced understanding of how this classification can be applied in practice. Second, it proposes an integration of ABC-XYZ classification with MTO and MTS strategies, providing a framework for developing a hybrid inventory policy that aligns production strategies with inventory characteristics. Third, it examines the potential of machine learning techniques, such as clustering, to enhance inventory classification, comparing these methods with traditional ABC-XYZ classification in terms of accuracy and efficiency.

4.4 Aim and Objectives

i. To check whether ABC and XYZ classification can be effectively applied to categorize inventory items based on their value and demand variability respectively.
ii. To further classify ABC XYZ classification in terms of MTO (Make – to – Order) and MTS (Make – to – Stock) strategies for an efficient inventory optimisation.
iii. To examine effective usage of a machine learning technique like Clustering to classify the inventory.

4.5 Research Questions

i. How effective is the ABC-XYZ classification method in categorizing inventory items based on their value and demand variability, and what are the impacts on inventory management performance?
ii. How can ABC-XYZ classification be integrated with MTO (Make-to-Order) and MTS (Make-to-Stock) strategies to optimise inventory management, and what are the potential benefits of this integration?
iii. Can machine learning techniques, such as Clustering, enhance the classification of inventory items, and how do these methods compare with traditional ABC-XYZ classification in terms of accuracy and efficiency?

4.6 Dissertation Structure

The structure of this research is as follows: Chapter 2 provides a comprehensive literature review, highlighting the theoretical foundations of ABC-XYZ classification, MTO and MTS strategies, and the use of machine learning in inventory management. Chapter 3 outlines the research methodology, detailing the data collection process, classification methods, and the proposed integration framework. Chapter 4 presents the results of the classification and integration, followed by a discussion of the findings in Chapter 5. Finally, Chapter 6 concludes the paper with recommendations for future research and practical implications for inventory management.

5 Literature Review

5.1 Overview

Inventory classification is a critical aspect of inventory management, aiming to optimise stock levels and improve operational efficiency. The ABC analysis classifies inventory based on annual consumption value, identifying high-priority items that require close monitoring (A items), mid-priority items (B items), and low-priority items (C items) (Scholz-Reiter et al, 2012). Complementing this, the XYZ analysis categorizes items based on demand variability, with stable items classified as X, moderately variable items as Y, and highly variable items as Z (Scholz-Reiter et al, 2012). ABC XYZ classification matrix can be seen in *Figure 1.*

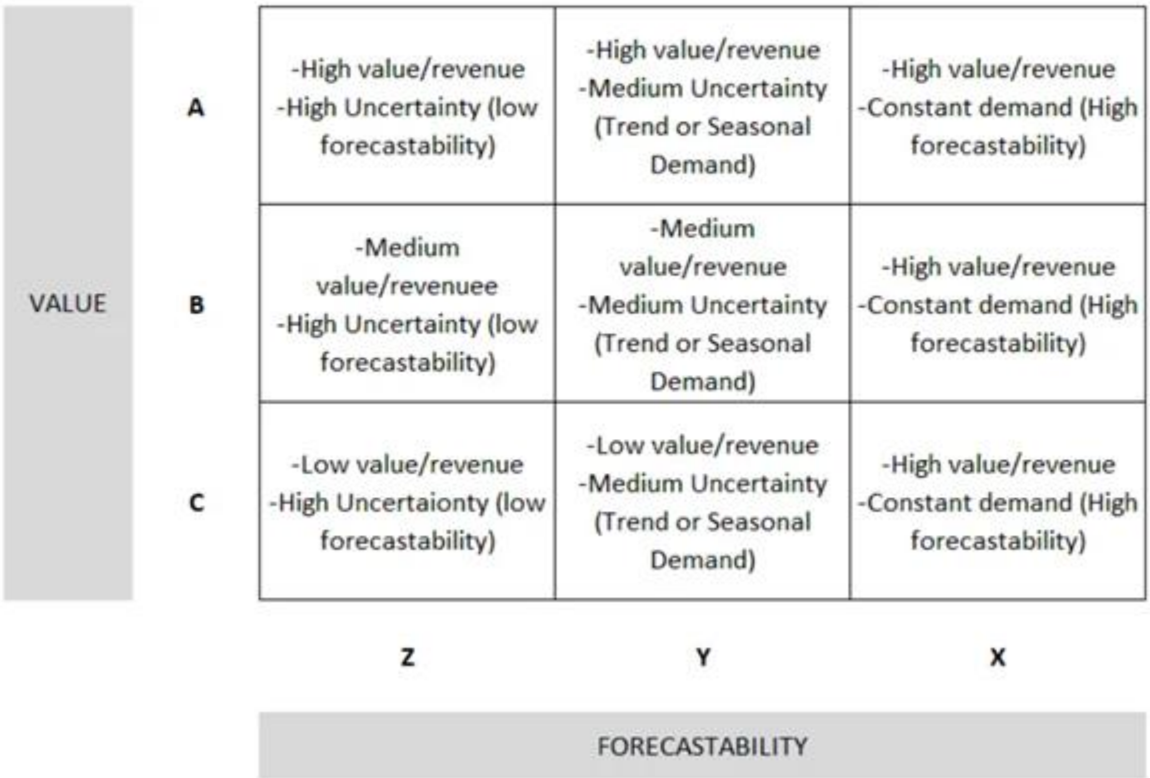


Figure 5.1: - ABC XYZ Classification Matrix (Yilmaz, 2023)

ABC and XYZ classification provide a multi-dimensional approach to inventory management, enabling more precise control over stock levels and resource allocation. This combined methodology is shown to be effective in enhancing service levels and reducing holding costs (Pandya and Thakkar, 2016). For instance, AX items (high value, stable demand) require different inventory strategies compared to CZ items (low value, high variability), allowing for tailored management practices (Scholz-Reiter et al, 2012).

The ABC-XYZ analysis includes fluctuations. The ABC parts simply follow Pareto principle thereby grouping it into three groups. Sometimes it is alternatively the product value that you want to give more expensive products a higher priority than cheap products, but this would be not so good for the decision between make-to-order (MTO) or make-to-stock (MTS) (Khan, Dong and Yu, 2017).

The XYZ part encapsulates the fluctuations which are also grouped into three groups. It basically captures the uncertainty or the stability. X-items fluctuate less, Y-items slightly more, and Z-items fluctuate the most. Technically, these fluctuations are the standard deviations of the demand (i.e., the standard deviation of how many products have been sold or daily / weekly basis). However, since product comparison is done for different absolute quantities, the coefficient of variation (the standard deviation divided by the mean) should be considered (Kabanova, Malakhova and Zenkova, 2022).

Another potential measure that comes into the picture is the number of orders i.e. orders for a specified period. A high number of orders indicates a stable demand, whereas a low number of orders showcases an unstable demand. This is easier to measure but is not as accurate as that of coefficient of variation. For example, Umbrellas are ordered around Fall and Winter as a result these orders are all clustered around the October – January thereby resulting in a lot of fluctuations (Krzyżaniak, 2017). The coefficient of variation could serve the purpose over here but for seasonal fluctuations one needs to switch between Make-to-Stock (in season) and Make-to-Order (out of season) anyway. Overall, the paper is trying to measure how stable the demand is (Khan, Dong and Yu, 2017).

Inventory classification plays a pivotal role in inventory management by aiming to optimise stock levels and enhance operational efficiency. Effective inventory classification ensures that resources are allocated appropriately, reducing unnecessary costs and improving service levels. The ABC classification, one of the most widely used classification methods, segments inventory based on annual consumption value. This method identifies high-priority items that require close monitoring (A items), mid-priority items (B items), and low-priority items (C items). This classification helps stakeholders focus their attention and resources on the most critical items, ensuring that high-value items are always available and that overstocking of low-value items is avoided (Scholz-Reiter et al, 2012).

Complementing the ABC classification is the XYZ classification, which categorizes items based on demand variability. This method classifies items with stable demand as X, those with moderately variable demand as Y, and items with highly variable demand as Z. The XYZ classification helps in understanding the predictability of demand for each item, allowing for better planning and forecasting. Items with stable demand (X items) can be stocked in larger quantities with less risk of overstocking, while items with highly variable demand (Z items) require more careful management to avoid stockouts and excess inventory (Scholz-Reiter et al, 2012).

The integration of ABC and XYZ classification provides a multi-dimensional approach to inventory management. This combined methodology allows for more precise control over stock levels and resource allocation, tailoring inventory strategies to the specific characteristics of each item. For example, AX items (high value, stable demand) are crucial and should be kept in stock to avoid any disruptions in operations. On the other hand, CZ items (low value, high variability) may be managed with a Make-to-Order strategy to minimise holding costs and reduce the risk of excess inventory (Scholz-Reiter et al, 2012).

This combined ABC-XYZ classification has been shown to be effective in enhancing service levels and reducing holding costs. By categorizing items based on both value and demand variability, managers can develop tailored inventory strategies that meet the specific needs of each item category. For instance, AX items, which are both high-value and have stable demand, can be stocked in larger quantities to ensure they are always available. In contrast, CZ items, which are low-value and have high demand variability, may be produced only when there is a confirmed order, minimising the risk of overstocking (Pandya and Thakkar, 2016).

The ABC-XYZ classification also considers demand fluctuations. The ABC classification follows the Pareto principle, grouping items into three categories based on their value and impact on overall inventory costs. This method helps prioritize expensive products, but it may not be as effective for deciding between Make-to-Order (MTO) or Make-to-Stock (MTS) strategies. The XYZ classification, on the other hand, focuses on demand variability, grouping items based on their demand fluctuations. X-items, with less fluctuation, are easier to manage, while Z-items, with high fluctuation, pose more challenges in inventory management (Khan, Dong and Yu, 2017).

Technically, these fluctuations are measured by the standard deviation of demand, indicating how much the demand varies from the average. Since products have different absolute quantities, it is important to use the coefficient of variation (the standard deviation divided by the mean) to compare them accurately. This measure provides a relative standard deviation, allowing for a more accurate comparison of demand variability across different products (Kabanova, Malakhova and Zenkova, 2022). *Figure 2* provides an overview about the MTO MTS strategies.

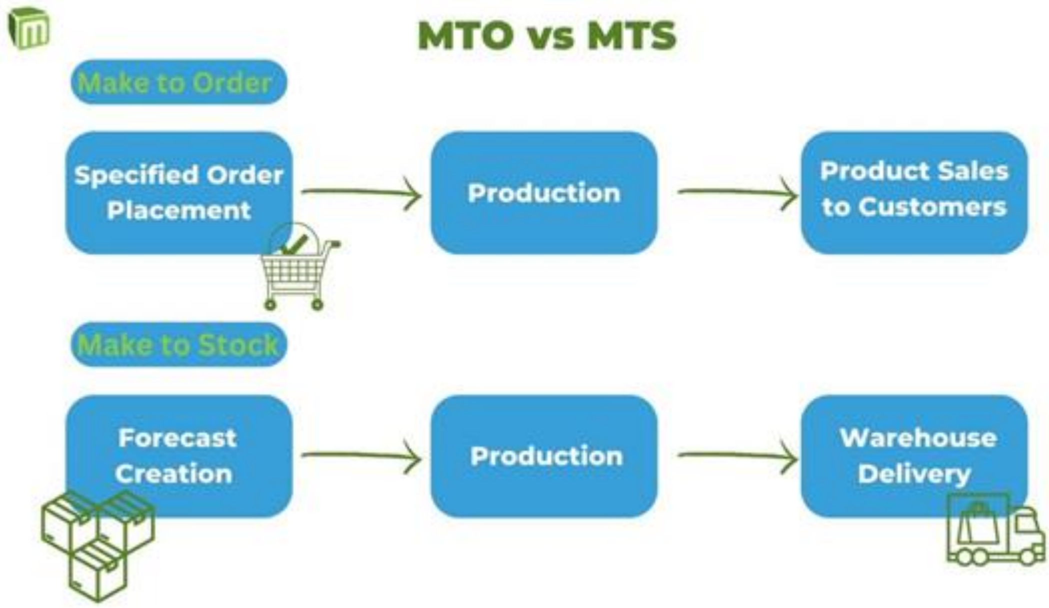


Figure 5.2: - MTO MTS Strategies (Spyridoula Karkani, 2023)

Another measure that can be considered is the number of orders within a specified period. A high number of orders indicates stable demand, while a low number of orders suggests unstable demand. This measure is easier to track but may not be as accurate as the coefficient of variation. For example, seasonal items like umbrellas see clustered orders around certain times of the year, leading to significant fluctuations in demand (Krzyżaniak, 2017). For such seasonal items, switching between Make-to-Stock (in season) and Make-to-Order (out of season) strategies can be effective in managing inventory levels and meeting demand.

Overall, the aim is to measure demand stability and determine the appropriate inventory management strategy. By integrating ABC and XYZ classifications, businesses can better understand their inventory dynamics and implement strategies that optimise stock levels, reduce costs, and improve service levels (Khan, Dong, and Yu, 2017).

5.2 Detailed Example and Applications

To further elaborate, an example of the manufacturing company producing electronic components is provided to illustrate the practical application of the ABC-XYZ classification method in real-world inventory management. By categorizing inventory items based on their value and demand variability, and then aligning these categories with appropriate production strategies (MTS for AX items and MTO for CZ items), the example demonstrates how companies can optimise inventory levels, reduce costs, and improve customer satisfaction. This tailored approach highlights the benefits of combining ABC-XYZ classification with production strategies, making the theoretical concepts more tangible and relevant to practical inventory management scenarios. Consider an example of a manufacturing company that produces a range of electronic components. Using the ABC classification, the company identifies that high-value microprocessors fall under the A category, while resistors and capacitors, being low-value, fall under the C category. Complementing this with XYZ analysis, the company finds that the demand for microprocessors (A items) is relatively stable (X), while the demand for certain capacitors (C items) is highly variable (Z) (Zamazii, 2022).

For AX items (high value, stable demand), adoption of Make-to-Stock (MTS) strategy is required. By maintaining a higher inventory level of these critical components, the company ensures they are readily available for production, reducing the risk of delays and enhancing customer satisfaction. The stable demand pattern justifies the cost of holding inventory, as the likelihood of stockouts is minimised (Wolff et al., 2021).

In contrast, for CZ items (low value, high variability), the company opts for a Make-to-Order (MTO) strategy. Producing these items only upon receiving confirmed orders helps the company avoid the costs associated with holding excess inventory and reduces the risk of obsolescence. The high variability in demand makes it impractical to stock large quantities, as predicting future demand accurately is challenging (Wolff et al., 2021).

This tailored approach extends to other combinations as well. For instance, BY items (moderate value, moderate variability) might employ a mixed strategy, where a base stock is maintained to cover regular demand, and additional quantities are produced as needed to handle demand spikes. This flexibility also allows to balance inventory costs with service level requirements effectively (Wolff et al., 2021).

5.3 Clustering for ABC XYZ Classification

This section builds on the earlier example by introducing K-Means clustering as a sophisticated tool to enhance ABC-XYZ classification in inventory management. By addressing the limitations of traditional methods, such as static classification and limited dimensionality, K-Means clustering offers more precise inventory segmentation and dynamic reclassification. This is crucial for developing tailored inventory strategies that adapt to changing market conditions, ultimately improving demand forecasting, reducing costs, and enhancing service levels. The integration of clustering methods highlights the importance of leveraging advanced techniques to optimise inventory management in complex, real-world scenario (Porter and Michele, 2018). Before moving into K-Means clustering, let understand what exactly clustering is. Clustering is a machine learning technique that groups data points into clusters based on similarity. In inventory management, clustering can identify patterns in inventory data, facilitating more informed decision-making (Rodriguez et al., 2019). This technique is particularly useful for enhancing ABC XYZ classification by providing a more nuanced understanding of inventory characteristics for better optimisation (Evdokimova, 2021). While ABC XYZ classification provides a structured approach to inventory management, several researchers have identified certain limitations. Teunter et al. (2010) argued that the traditional cutoff points (e.g., 80% for A items) are arbitrary and may not be optimal for all businesses. Bhattacharya et al. (2007) noted that ABC XYZ classification only considers two dimensions, potentially overlooking other important factors showcasing it limited dimensionality. Boylan et al. (2008) highlighted that traditional classification methods are often static and failed to adapt to the changing market conditions. These limitations led researchers to explore more sophisticated approaches, including the integration of machine learning techniques like K-Means clustering. Clustering can be done using several methods like K – Means, DBSCAN, Hierarchical etc.

Here, the method used is K – Means Clustering since according to the recent study conducted by Porter and Michele (2018) states that K-Means clustering algorithm can be used to partition data into distinct clusters based on similarity. It aims to minimise the variance within each cluster and maximise the variance between clusters. Integrating K-Means clustering with ABC XYZ classification offers a powerful approach to inventory optimisation. By leveraging the strengths of both methods, businesses can achieve more precise inventory segmentation, improved demand forecasting, and optimised inventory levels. This approach enhances inventory management's efficiency and effectiveness, leading to cost savings, improved service levels, and better alignment with business objectives. K-Means clustering provides more detailed segmentation of inventory items, allowing for the identification of specific patterns and relationships that traditional ABC XYZ classification might miss. By identifying clusters that do not fit neatly into traditional categories, businesses can develop tailored strategies for each cluster, optimising inventory management. Clustering also reveals hidden patterns and trends in demand variability, improving the accuracy of demand forecasts. Accurate demand forecasts enable businesses to maintain optimal inventory levels, reducing the risk of stockouts and excess inventory. K-Means clustering specifically supports dynamic reclassification of inventory items based on changing demand patterns and market conditions. This adaptability ensures that inventory strategies remain aligned with current business needs and customer demands. K-Means clustering when applied to segment inventory items into distinct groups based on similar attributes helps to identify patterns and relationships. After creation of clusters, they can be analysed to understand their characteristics and can be mapped to appropriate inventory strategies. For example, high-cost items with long lead times might require different management strategies than low-cost items with short lead times. According to Srinivasan et al. (2011) study, K-Means clustering can be used for multi-criteria inventory classification, incorporating factors beyond just value and variability.

Several studies have highlighted the benefits of integrating K-Means clustering with ABC XYZ classification. A study by Evdokimova (2021) demonstrated that K-Means clustering provides more detailed segmentation, allowing for the identification of patterns that traditional methods might miss whereas Kourentzes et al. (2014) showed that clustering-based approaches can reveal hidden patterns in demand variability, improving forecast accuracy. The ability of clustering-based methods to support dynamic reclassification of items as market conditions change were emphasized by Boylan et al. (2008). A study by Teunter, Babai and Syntetos, (2009) demonstrated improved service levels and customer satisfaction through more precise inventory control. Apart from the benefits there are some limitations that are discussed subsequently. A study by Porter and Michele (2018) stated that the downside of k-means is that it cannot consider characteristics or traits that are not value based, therefore, it cannot cluster on names, or identification numbers. Having said that there are different ways to cluster other than k-means, such as manually by what the part does, or when it will be needed, Hierarchical, K-Nearest Neighbours, etc. However, due to time constraints these are not explored in this dissertation.

According to a study by Karaca (2018), in the manufacturing industry, K-Means clustering was applied to segment products based on demand variability and value. This integration led to an optimal reduction in inventory holding costs and improved service levels. By identifying specific clusters, the company could develop tailored strategies for managing high-value and high-variability items, ensuring optimal inventory levels and reducing the risk of stockouts whereas in fashion industry K-Means clustering is used to identify seasonal demand patterns and optimise inventory levels for fashion products. This approach reduced stockouts considerably and enhanced customer satisfaction (Kandemir, 2022). The clustering analysis revealed hidden patterns in demand variability, allowing the company to adjust inventory levels proactively and meet customer expectations more effectively.

5.4 Segregating ABC XYZ Classification into MTS (Make – to – Stock) and MTO (Make – to – Order) Strategies.

After clustering the inventory into ABC and XYZ classes, it is further segregated into MTS and MTO strategies. (Wolff et al., 2021) research proposed a framework for matching product characteristics with manufacturing strategies. They suggested that high-value, low-variability items (AX) are often suitable for MTS, while low-value, high-variability items (CZ) may be better suited for MTO. Zaerpour et al. (2008) developed a decision support system for choosing between MTO and MTS strategies based on ABC classification. Their model considered factors such as setup times, holding costs, and service level requirements. Kampen et al. (2012) conducted an empirical study on the application of ABC XYZ classification in MTO/MTS decisions. They found that while the classification provided valuable insights, other factors such as product life cycle and production capacity also played crucial roles in strategy selection. Although, this research mostly relies on revenue / holding costs, quantities / product lifecycle and variation in time for demand recognition but other factors like service level requirements and product lifecycle can be use in future study. The segmentation of ABC XYZ into MTO and MTS can be observed in *Figure 3.*

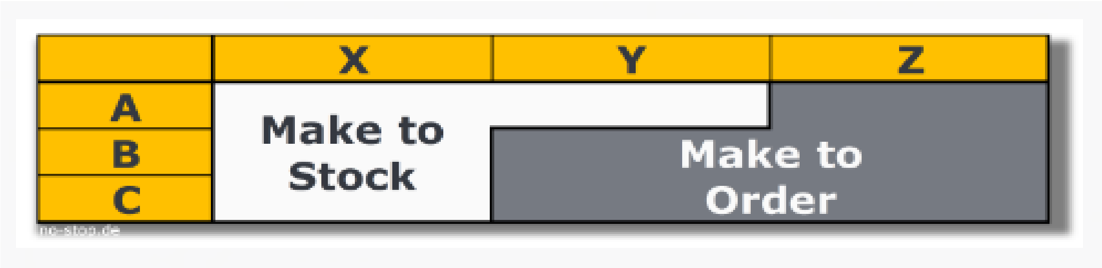


Figure 5.3: - Integration of ABC XYZ Classification with MTO MTS Strategies (Noll, 2020)

Implementing this combined ABC-XYZ methodology requires a systematic approach. First, inventory items need to be classified based on their value (ABC) and demand variability (XYZ). This involves analysing historical sales data to determine the annual consumption value and demand patterns for each item (Pandya and Thakkar, 2016). Advanced data analytics tools and inventory management software can facilitate this process by providing accurate and timely data insights (Scholz-Reiter et al, 2012).

Once the classification is complete, inventory policies and strategies can be developed for each category. For example, AX items may be assigned higher safety stock levels and more frequent review cycles to ensure their availability. In contrast, CZ items might be reviewed less frequently, with production scheduled based on actual orders received (Pandya and Thakkar, 2016).

Integrating this methodology into the company's Inventory Management System (IMS) is crucial for its success. The IMS should support dynamic reclassification and strategy adjustments based on real-time data. Regular audits and reviews of inventory performance are necessary to ensure that the classification remains accurate and that the strategies are effective in meeting business goals.

Training and awareness programs for Inventory classification is essential for optimising stock levels and improving operational efficiency. The ABC analysis sorts inventory based on annual consumption value, identifying high-priority items (A), mid-priority items (B), and low-priority items (C) (Scholz-Reiter et al, 2012). Complementing this, XYZ analysis categorizes items by demand variability: X for stable demand, Y for moderate variability, and Z for high variability (Scholz-Reiter et al, 2012).

Integrating ABC and XYZ classifications provides a multi-dimensional approach to inventory management, enhancing service levels and reducing holding costs (Pandya and Thakkar, 2016). For instance, AX items (high value, stable demand) require different inventory strategies compared to CZ items (low value, high variability), allowing tailored management practices (Scholz-Reiter et al, 2012).

ABC-XYZ analysis incorporates demand fluctuations. ABC classification follows the Pareto principle, sometimes prioritizing more expensive products over cheaper ones (Khan, Dong, and Yu, 2017). XYZ classification captures demand stability or instability, grouping items by their standard deviation of demand. Since comparisons are across different quantities, the coefficient of variation (standard deviation divided by the mean) should be considered (Kabanova, Malakhova, and Zenkova, 2022).

Another measure is the number of orders within a period, indicating stable or unstable demand. For example, seasonal items like umbrellas have clustered orders, causing demand fluctuations (Krzyżaniak, 2017). The coefficient of variation is useful here, but seasonal items may require switching between MTS (in season) and MTO (out of season) strategies. The aim is to measure demand stability (Khan, Dong, and Yu, 2017).

Inventory classification optimises stock levels and operational efficiency by combining ABC and XYZ analyses. ABC analysis segments inventory by annual consumption value, identifying high-priority items requiring close monitoring (A items), mid-priority items (B items), and low-priority items (C items) (Scholz-Reiter et al, 2012). XYZ analysis categorizes items based on demand variability: X for stable demand, Y for moderate variability, and Z for high variability (Scholz-Reiter et al, 2012). This multi-dimensional approach enhances service levels and reduces holding costs (Pandya and Thakkar, 2016).

Implementing the combined ABC-XYZ methodology involves several steps. First, classify inventory items based on value (ABC) and demand variability (XYZ) by analysing historical sales data. Use advanced data analytics tools and inventory management software for accurate data insights. Develop inventory policies for each category: AX items may have higher safety stock levels and frequent review cycles, while CZ items might be reviewed less frequently with production based on actual orders.

Integrating the methodology into the inventory management system (IMS) for dynamic reclassification and strategy adjustments based on real-time data, regular auditing and reviews ensure classification accuracy and strategy effectiveness in meeting business goals

(Scholz-Reiter et al, 2012). By integrating ABC and XYZ classifications, businesses can optimise stock levels, reduce costs, and improve service levels. AX items (high value, stable demand) benefit from MTS strategies to ensure availability and minimise stockouts. CZ items (low value, high variability) benefit from MTO strategies to avoid excess inventory and reduce holding costs. This tailored approach balances inventory costs with service level requirements, enhancing overall operational efficiency (Wolff et al., 2021).

5.5 Summary

The chapter of literature review highlights the significance of inventory classification through ABC and XYZ analyses in optimising stock levels and operational efficiency. ABC analysis prioritizes items based on value (A, B, C), while XYZ analysis assesses demand variability (X, Y, Z). Combining these methods offers a nuanced approach to inventory management, leading to tailored strategies like Make-to-Stock (MTS) for high-value, stable demand items (AX) and Make-to-Order (MTO) for low-value, variable demand items (CZ). Integrating K-Means clustering can enhance this approach by addressing limitations of traditional methods, such as static classification and limited dimensionality, thereby improving demand forecasting and inventory optimisation. This review links to research questions by exploring how advanced techniques can address these limitations and enhance inventory strategies.

6 Research Methodology

6.1 Overview

Research Methodology is built upon the theoretical frameworks and key concepts discussed in the literature review, applying them to the specific research context. The methodologies chosen are directly informed by the gaps and best practices identified in the literature, ensuring that the research design is both relevant and robust. This approach allows for a systematic exploration of the research questions, grounded in established knowledge while addressing the identified gaps.

This section of dissertation will outline the approach, methods, and techniques used to collect and analyse data. It includes the research design (qualitative, quantitative, or mixed methods), methodology (CRISP – DM), data collection methods (e.g., surveys, interviews, secondary data), sampling strategy, and data analysis techniques. It also addresses validity, reliability, ethical considerations, and limitations of the study. This section ensures the research process is systematic and transparent, allowing for the evaluation of the study's credibility. It basically aligns the research approach with the objectives thereby providing a clear pathway from research questions to findings (Dissanayake, E V A., 2023).

6.2 Saunders Research Model

According to Melnikovas (2018), research methodology begins with delineating the fundamental philosophy, selecting methodologies, methods, and tactics, and defining time constraints, which together lead the research logic to the study design - the key techniques and procedures of data collecting and analysis which is shown in the *Figure 4* below.

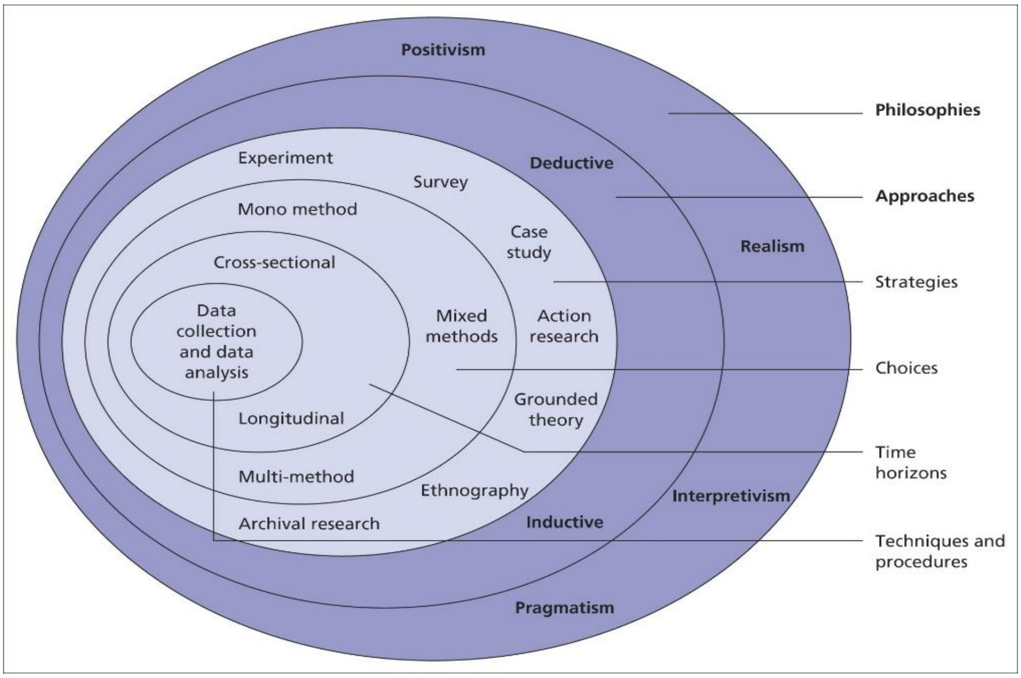


Figure 6.1: - Saunder's Research Model (Saunders et al., 2019)

The research onion consists of six main layers: -

- Research Philosophy

  Positivism: - This project adopts a positivist approach. It basically focuses on objective reality and uses quantitative data to test hypotheses. This is suitable for employing machine learning techniques like K-Means clustering which requires statistical analysis of large datasets.

- Research Approach

  Deductive Approach: - This project follows a deductive approach. It deals with existing theories and frameworks, such as ABC XYZ classification, and applies K-Means clustering to test hypotheses related to inventory segmentation and optimisation.

- Research Strategy

  Archival Research: - The strategy used over here is archival research since this strategy uses existing data (the dataset provided) to perform the analysis and here a dataset has been procured from UCI which contains inventory transaction records from a UK-based online retail store spanning from 2009 to 2011.

- Research Choices
  Quantitative Methods: - The project primarily employs quantitative methods, including statistical analysis and machine learning algorithms, to analyse inventory data and segment it into clusters. Quantitative data is essential for applying K-Means clustering and measuring its impact on inventory management.

- Time Horizons
  Cross-sectional: - The research adopts a cross-sectional time horizon, analysing inventory data at a specific point in time. The analysis is performed on a dataset for a specific period without considering how the data changes over time.

- Techniques and Procedures
  Data Collection: - Inventory data is collected from has been procured from UCI which contains inventory transaction records from a UK-based online retail store spanning from 2009 to 2011. It includes fields such as *'InvoiceNo'*, *'StockCode'*, *'Description'*, *'Quantity'*, *'InvoiceDate'*, *'UnitPrice'*, *'CustomerID'*, and *'Country'*.
  Data Preparation: - The collected data is cleaned and normalized to ensure accuracy and comparability across different attributes. The cleaned and normalized data is subjected to K-Means clustering to identify distinct groups of inventory items based on value and demand variability.

6.3 CRISP – DM Methodology

Data mining is a creative process that takes a variety of talents and expertise. There is a need of a common framework for carrying out data mining operations. The success of a data mining project depends on the individual or team responsible, and good practices may not be replicated throughout the company. A conventional method to data mining requires translating business challenges into tasks, recommending relevant data transformations and techniques, and evaluating and recording findings as a result CRISP – DM methodology came into the picture. The CRISP-DM (Cross industrial Standard Process for Data Mining) methodology addresses some of these issues by creating a process model for data mining projects that is independent of industrial sector and technology (Wirth and Hipp, 2000).

A schematic version of CRSIP – DM can be seen in *Figure 5.*

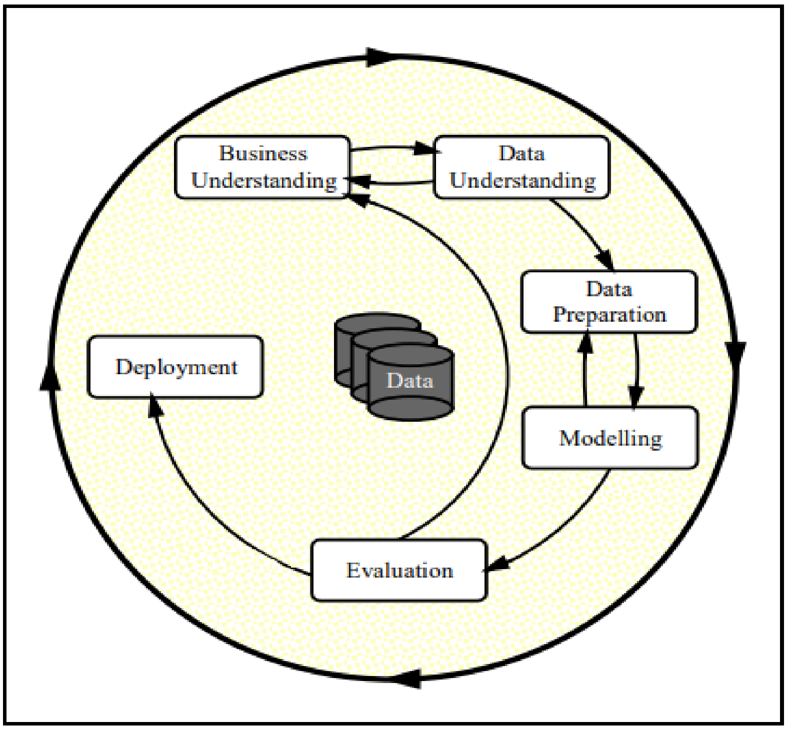


Figure 6.2: - CRISP-DM process model in data mining (Wirth and Hipp, 2000)

The CRISP-DM data mining reference model provides an overview of the project's life cycle. It contains the project's phases, tasks, and outputs. The life cycle of a data mining project is divided into six parts, as illustrated in *Figure 5*. The order of the phases does not follow a hard and fast rule. The arrows highlight simply the most essential and frequent relationships between phases; nonetheless, in a specific project, the next phase, or specific task of a phase, is determined by the outcome of each phase. The outer circle in *Figure 5* represents the cyclical nature of data mining itself. Data mining does not end once a solution is implemented. Lessons learnt during the process and from the deployed solution can prompt new, more specific business issues. Subsequent data mining methods will profit from earlier ones' experiences (Wirth and Hipp, 2000).

The six iterative phases of CRISP-DM methodology are as follows: -

- Business Understanding: - This phase particularly focuses on the project objectives and requirements from a business perspective and tries to interpret it into a data mining problem definition to achieve the objectives (Wirth and Hipp, 2000). In this dissertation, the primary objective is to optimise inventory management using K-Means clustering to enhance the traditional ABC XYZ classification method. This involves understanding how clustering can improve inventory segmentation, demand forecasting, and overall stock management.

- Data Understanding: - This phase focuses more on activities to get familiar with the data, it also identifies data quality problems, tries to discover insights into the data and detects patterns to form hypotheses (Wirth and Hipp, 2000). In this case, Data Understanding deals with gathering historical inventory data which consists of attributes such as annual consumption value, demand variability, lead time etc. Data sources may include ERP systems, sales records, and inventory management software

- Data Preparation: - This phase covers all activities to construct the final dataset (data that will be used for modelling) from the initial raw data. This phase includes tasks that are likely to be performed numerous times without following any prescribed order. Tasks include table, record and feature selection, data cleaning, feature extraction and transformation of data for modelling tools etc (Wirth and Hipp, 2000). In this dissertation as a part of Data Preparation, data cleaning is done by addressing missing values, outliers, and inconsistencies in the data. For example, filling in missing values, normalizing data to ensure consistency across different scales, and removing or correcting outliers. For feature selection and feature extraction, specific features can be selected, and additional features can be extracted that might be useful for clustering, such as the coefficient of variation for demand, or a combined score for lead time and cost. Furthermore, data has also been normalized or standardized to ensure all attributes contribute equally to the clustering process as a part of data transformation. This step is crucial for K-Means, which is sensitive to the scale of the data.

- <u>Modelling:</u> - This phase deals with selecting and applying various modelling techniques and their parameters are calibrated to optimal values. Several techniques can be used depending upon data mining problem type. Some techniques require specific data formats (Wirth and Hipp, 2000). In this dissertation as a part of model selection, K-Means Clustering has been selected and has been applied to segment the inventory data into clusters based on similarity in value and demand variability. Hyperparameter tuning is done thereafter to determine the optimal number of clusters (k) using methods like the elbow method or silhouette score. Model is executed and the clustering algorithm is run to generate clusters that categorize inventory items into meaningful segments. Thereafter, mapping to ABC XYZ classification takes place after analysing the characteristics of each cluster and mapping them to the appropriate ABC XYZ categories. This helps in aligning the clusters with business objectives, such as prioritizing high-value, high-demand variability items.

- <u>Evaluation:</u> - Before proceeding to final deployment of the model, the final model must be evaluated with the steps employed to execute that model, this is done to ascertain whether the model achieves the business objectives. A key objective is to also determine whether any important business issue has been neglected or not. After the end of this phase, a decision on the use of the final model should be reached (Wirth and Hipp, 2000). For this project, evaluation phase looks after evaluating the quality of clusters by assessing intra-cluster similarity and inter-cluster dissimilarity. Assessing the effectiveness of the clustering model by comparing it to traditional ABC XYZ classification methods. From business impact point of view, calculation of potential impact of the new classification on inventory costs, stockouts, and service levels must be done. If necessary, clustering model must be refined by revisiting data preparation or adjusting the number of clusters.

- <u>Deployment:</u> - Creating the model is typically not the end of the project. Knowledge should be organised and provided in a user-friendly format for customers. Depending on the objectives, the deployment step can range from as easy as generating a report to as difficult as implementing a repeatable data mining process. In most circumstances, the deployment procedures will be carried out by the user rather than the data analyst. In any event, it is critical to understand ahead of time what activities would be required to really use the produced models (Wirth and Hipp, 2000). In this project, deployment phase includes integrating the clustering results into the existing inventory management system, developing dashboards or reports that visualize the new inventory segments for decision-makers, continuously monitoring the performance of the new inventory strategy and updating the clustering model periodically for adapting to the changes in inventory data and business needs.

6.4 Dataset Description

For this dissertation, a secondary dataset has been procured from UCI which contains inventory transaction records from a UK-based online retail store spanning from 2009 to 2011. The dataset used is structured and consists of numerical and categorical variables. It includes fields such as *‘InvoiceNo’*, *‘StockCode’*, *‘Description’*, *‘Quantity’*, *‘InvoiceDate’*, *‘UnitPrice’*, *‘CustomerID’*, and *‘Country’*. Quantitative analysis can be employed on such type of data which involves finding patterns, trends, relationships and insights (Taylor, 2022).

The dataset comprises a total of 8 attributes, which represent various aspects of the retail transactions. Each row corresponds to a single transaction. This dataset can be used for academic research and practical applications in inventory management, demand forecasting, and customer segmentation. It serves as a basis for performing ABC and XYZ analyses to optimise stock levels and improve inventory strategies. The data supports the application of various statistical and machine learning techniques for analysing retail performance and enhancing operational efficiency.

The description for various data fields are as follows: -

Data fields	Type	Description
Invoice	Categorical (String)	A unique identifier assigned to each transaction. It is a combination of numbers and letters.
StockCode	Categorical (String)	A unique code representing each product item.
Description	Text (String)	A textual description of the product associated with the *StockCode*.
Quantity	Numeric (Integer)	The number of units of the product purchased in the transaction. Positive values indicate sales, while negative values may indicate returns or adjustments.
InvoiceDate	DateTime	The exact date and time when the transaction occurred.
Price	Numeric (Float)	The unit price of the product in GBP (British Pounds).
Customer ID	Categorical (Integer)	A unique identifier assigned to each customer. This ID allows for the aggregation of transactions made by the same customer.
Country	Categorical (String)	The country from which the customer made the purchase. The dataset includes transactions primarily from the United Kingdom but also from other countries.

Table 6.1: - Data fields and description

6.5 Data Analysis and Methodological Framework

This section outlines the comprehensive process of data analysis and methodology used in this research. It encompasses data exploration, feature engineering, method selection, and model evaluation, ensuring that each step aligns with the research objectives. The approach is thoroughly documented and integrated into the overall research findings, providing a robust framework for deriving meaningful insights and conclusions.

- Data Exploration and Preprocessing: - The first step involves loading the dataset into the environment. This is typically done using libraries like *pandas* in Python, which allow for efficient data manipulation. Thereafter the data is explored to understand its structure, the types of variables (e.g., categorical, numerical), and any missing values or anomalies. For example, examining the first few rows of the dataset provides insight into its content and potential issues that needs to be addressed. Data cleaning is performed which involves handling missing values, removing duplicates, and converting data types where necessary. Standard scaling is also performed as a part of data preprocessing step to ensure that differences in scale does not unduly influence the clustering results.

- Feature Engineering: - Feature engineering involves creating new variables or modifying existing ones to better capture the underlying patterns in the data. For example, calculating *TotalRevenue* as *Quantity* * *Price* helps in understanding the economic impact of each product which in turn helps in ABC classification. Other derived features, like the standard deviation (*std_quantity*) and the coefficient of variation (*cv_quantity*), are computed to quantify demand variability, which is crucial for the XYZ analysis.

- Method Selection and Implementation: - Depending on the research question, appropriate clustering methods are chosen. In this case, K-Means clustering is applied to classify products into ABC categories based on revenue and quantity, and XYZ categories based on demand variability. K-Means was chosen for its simplicity and effectiveness in partitioning data into distinct clusters. For ABC analysis, it helps segment products by economic impact, while for XYZ analysis, it segments based on demand stability (Porter and Michele, 2018). Clustering algorithms like K-Means require careful selection of parameters (e.g., number of clusters). These choices are justified based on the research context, such as selecting 3 clusters for ABC (to classify into A, B, C categories) and XYZ analysis.

- Model Execution and Evaluation: - The code executes the clustering algorithms on the prepared data. This involves fitting the model to the data and predicting cluster assignments for each product. The resulting clusters are evaluated to ensure they align with theoretical expectations. For instance, in ABC classification, products in the 'A' category should have higher revenue and sales volumes compared to those in 'C'. The cluster centres are examined to interpret the characteristics of each group. This step ensures that the clustering results are meaningful and can be used to draw valid conclusions.

- Documentation and Visualization: - Documentation and visualization of the results have been done which includes plotting the cluster centres or the distribution of products across different clusters to illustrate the findings. The entire process, from data loading to final clustering is documented to ensure that the research is reproducible and transparent. Comments in the code explain the purpose of each step, aiding for the reviewing of the code.

- Result Integration and Reporting: - The results, including the classification into ABC-XYZ categories, which are further divided into MTO and MTS strategies, are incorporated into the research findings. These results are often compared with existing theories or models to validate the approach. The findings are reported in a structured format, often including tables, charts, and statistical summaries that convey the research outcomes effectively.

6.6 Proposed Model

K-Means clustering model has been proposed for this project. K-Means is a popular unsupervised learning algorithm used for partitioning a dataset into *K* distinct non-overlapping clusters. The algorithm aims to minimise the within-cluster variance by iteratively updating the cluster centroids and reassigning data points to the closest cluster (Sinaga and Yang, 2020).

First choose the number of clusters *K* and then randomly initialize *K* cluster centroids, μ_1, μ_2 ,…,μ_k. Thereafter each data point x_i has been assigned to the closest centroid, forming *K* clusters. The closest centroid is determined using a distance metric, typically the Euclidean distance:

$$C_i = argmin_k \, ||x_i - \mu_k||$$

Here:

C_i = Index of the cluster assigned to x_i.

μ_k = Centroid of the *k*th cluster.

$||x_i - \mu_k||$ = Euclidean distance between the data point x_i and centroid μ_k.

$argmin_k$ = Index k that minimises this distance (argument by which a function can be minimised).

After all points are assigned to clusters, centroids are updated by calculating the mean of all data points in each cluster.

The new centroid μ_k for each cluster is k is calculated as:

$$\mu_k = \frac{1}{|\, C_k \,|} \sum_{x_i \in C_k} x_i$$

Here:

C_k = Set of all points assigned to cluster *k*.

$|\, C_k \,|$ = Number of points in cluster *k*.

$\sum_{x_i \in C_k} x_i$ = sum of all points in cluster *k*.

Assignment and updating steps are to be repeated until the centroids no longer change significantly or until a maximum number of iterations is reached.

An objective function is created for the algorithm with an objective to minimise the sum of squared distances (inertia) between the data points and their corresponding cluster centroids.

$$J = \sum_{k=1}^{K} \sum_{x_i \in C_k} ||x_i - \mu_k||^2$$

Here:

J = Objective function (also known as inertia or distortion).

$||x_i - \mu_k||^2$ = Squared Euclidean distance between the data point x_i and centroid μ_k.

The aim is to minimise J to achieve tight and well-separated clusters.

The K-Means algorithm is guaranteed to converge, but it may converge to a local minimum rather than the global minimum of the objective function. The result can depend on the initial positions of the centroids, which is why K-Means is often iterated with different initializations, and the best result is chosen.

Optimal number of clusters can be determined by using Elbow Plot. It analyses the WCSS (Within Cluster Sum of Squares) which basically measures the sum of squared distances between data points and clusters. In this dissertation business rules and predefined thresholds are being used to categorize products, rather than trying to find natural clusters in the data. The number of categories is determined by these rules and not by an optimal clustering solution.

To summarise one can, say that K-Means clustering algorithm is a powerful and efficient tool for unsupervised learning which can reveal underlying patterns and structures within complex datasets. By grouping data points into well-defined clusters, it provides valuable insights that can drive decision-making and innovation across various fields. Its simplicity, scalability, and effectiveness make K-Means a cornerstone technique in data analysis which in turn enables users to uncover meaningful relationships and enhance their understanding of the data at hand. With careful implementation and thoughtful consideration of its limitations, K-Means can significantly contribute to successful outcomes in diverse applications (Sinaga and Yang, 2020).

6.7 Segmentation of ABC XYZ in MTO and MTS strategies

After clustering the ABC XYZ classes, they are further segmented into MTO and MTS strategies. AX, AY and AZ are categorized in MTO since these are high-value items with variable demand. MTO also helps in reducing inventory holding costs for these expensive items while accommodating demand uncertainty. CX is also categorized into MTO since it consists of low-value items with stable demand. Despite low value, stable demand allows for efficient production scheduling in an MTO system. MTS includes BX, BY, BZ, CY, and CZ categories, it deals with items of lower value items where the cost of stockouts is less critical, making MTS more suitable (Wolff et al., 2021).

6.8 Summary

In research methodology chapter, we saw the integration of the Saunders Research Onion model and CRISP-DM framework to systematically explore inventory management optimisation through K-Means clustering and further segmentation of ABC XYZ classification into MTO MTS strategies. The study adopts a positivist philosophy with a deductive approach, utilizing archival research and quantitative methods. Dataset based on a UK-based online retailer is cleaned, normalized, and clustered to improve inventory segmentation and forecasting. The methodology emphasizes rigorous data analysis, model evaluation, and refinement, ensuring alignment with business objectives, and contributes to enhancing the theoretical understanding of inventory management practices.

7 Analysis

7.1 Overview

This dissertation's analysis employed K-Means clustering to enhance the ABC XYZ classification thereby offering a data-driven approach to segment inventory based on value and demand variability. The methodology chapter provided the framework, ensuring systematic data preparation, clustering, and evaluation processes. By aligning these methods with business objectives, the research achieved more accurate inventory segmentation, which was crucial for optimising stock levels and improving forecasting accuracy. The methodology's structured approach directly informed the analytical procedures, ensuring coherence and validity in the research findings.

7.2 Data Pre-processing

Data preprocessing refers to the steps taken to clean, transform, and prepare raw data for analysis. It involves tasks like handling missing values, normalizing data, encoding categorical variables, and splitting the dataset into training and testing sets with supervised machine learning point of view, since this dissertation deals with an unsupervised machine learning algorithm there's no need to split the dataset. Proper preprocessing ensures that the data is in the right format and condition for the machine learning models to perform effectively. It can also involve feature selection, where irrelevant or redundant data is removed to improve the model's performance (Chithra et al., 2022).

First step of preprocessing was to check whether are any missing values present in the dataset. From *Table 2* it becomes quite evident that there are certain percentage of missing values in columns like '*Description*' and '*Customer ID*'. '*Description*' column has 0.26 % of missing values whereas '*Customer ID*' has 24.96 % of missing values.

Data fields	Missing values	Percentage (%)	After treating the missing values	Data fields	Missing values	Percentage (%)
Invoice	0	0		*Invoice*	0	0
StockCode	0	0		*StockCode*	0	0
Description	1454	0.268310		*Description*	0	0
Quantity	0	0		*Quantity*	0	0
InvoiceDate	0	0		*InvoiceDate*	0	0
Price	0	0		*Price*	0	0
Customer ID	135080	24.926648		*Customer ID*	0	0
Country	0	0		*Country*	0	0

Table 7.1: - Missing values analysis

Since, the missing values were present in the variables that weren't used in the modelling, rows with these missing values were removed to make further analysis more efficient and robust.

Also, numerical variables like *'Price'* and *'Quantity'* had negative values which indicate special cases or data anomalies that may not be relevant to the primary analysis as a result those records are removed from the dataset as well as part of preprocessing stage. Categorical variables like '*Description*' and '*Customer ID*' are being removed for the same.

7.3 Summary Statistics

Summary statistics provide key insights into a dataset, including measures of central tendency (mean, median, mode), variability (standard deviation, variance), distribution shape (skewness, kurtosis), and range (minimum, maximum). It helps in understanding the overall characteristics of the data, identifying trends, and detecting anomalies (Islam and Al-Shiha, 2018).

Data fields	Count	Mean	STD	Min	25%	50%	75%	Max
Quantity	397885	12.98821	179.3316	1	2	6	12	80995
Price	397885	3.116525	22.09786	0.001	1.25	1.95	3.75	8142.75

Table 7.2: - Summary statistics for numerical variables

Table 3 presents summary statistics for the numerical variables *'Quantity'* and *'Price'* in the dataset. According to the statistics, the average quantity sold is approximately 13 units, with a wide standard deviation of 179, indicating high variability. The quantities range from 1 to 80,995, with the majority falling between 2 and 12 units whereas the average price is about £ 3.12, with a significant standard deviation of 22.10, suggesting considerable price variation. Prices range from £ 0.001 to £ 8,142.75, with most items priced between £1.25 and £3.75. These statistics indicate significant variability in both quantity and price thereby highlighting the need for careful segmentation and inventory management strategies.

Statistics	*Invoice*	*StockCode*	*Description*	*InvoiceDate*	*Country*
Count	397885	397885	397885	397885	397885
Unique	18532	3665	3877	17282	37
Top	576339	85123A	WHITE HANGING HEART T-LIGHT HOLDER	14-11-2011 15:27	United Kingdom
Frequency	542	2035	2028	542	354321

Table 7.3: - Summary statistics for categorical variables

Table 4 provides a summary of categorical variables like *'Invoice'*, *'StockCode'*, *'Description'*, *'InvoiceDate'*, *'Country'* from the dataset. According to the statistics, there are 397,885 transactions, with 18,532 unique invoices for *'Invoice'*. The most frequent invoice appears 542 times. There are 3,665 unique *'StockCode'*, with "85123A" being the most common, appearing 2,035 times. There are 3,877 unique products *'Description'*, with "WHITE HANGING HEART T-LIGHT HOLDER" being the most frequent whereas for *'InvoiceDate',* there are 17,282 unique dates. *'Country'* variable shows that the transactions are executed from 37 countries, with the United Kingdom being the most common, accounting for 354,321 transactions.

7.4 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a crucial step where datasets are explored to uncover patterns, spot anomalies, test hypotheses, and check assumptions. This process involves using various statistical tools and data visualization techniques, such as histograms, box plots, scatter plots, and correlation matrices, to gain insights into the structure and relationships within the data. EDA helps in understanding the data's underlying distribution and guides subsequent data processing, model selection, and analysis (Albert and Rizzo, 2011).

For this project several new features / variables were derived to support the analysis. For ABC classification, *'TotalSales'* was derived which basically depicted the total revenue for each product and was used to classify items into A, B, and C categories. This new feature helped to identify high-value products that contribute significantly to overall revenue, allowing for prioritized inventory management and focused attention on key items.

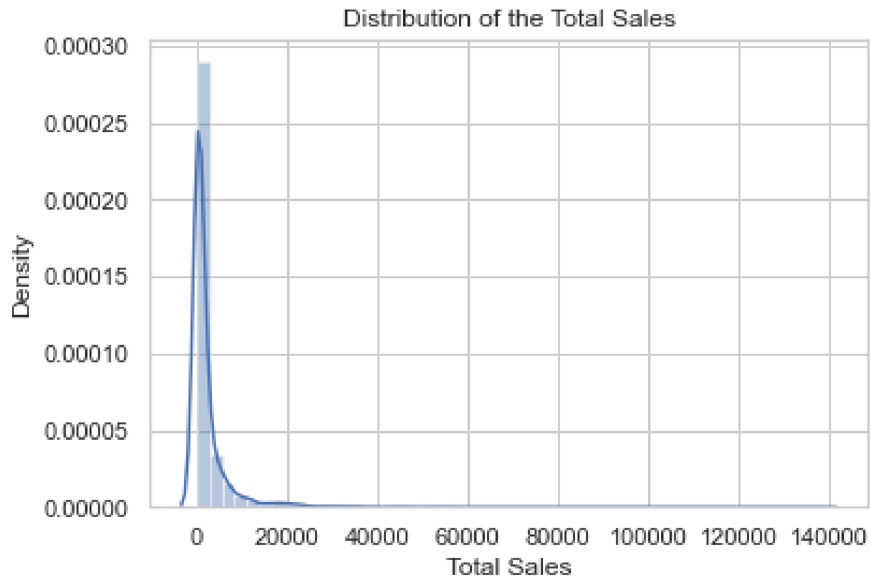


Figure 7.1: - Distribution of total sales

Figure 6 illustrates the distribution of total sales across products. The distribution is heavily right-skewed, indicating that most products generate low sales, while a small number of products contribute to significantly higher sales. This pattern is consistent with the Pareto principle, where a small percentage of products (likely categorized as "A" items) account for most of the total sales. It is quite evident that 80 % of the sales is taking place between the range of £ 10,000 to £ 20,000 while the remaining products are classified as "B" and "C" items (10 % each) contribute relatively little to the overall revenue. This type of distribution is common in retail and inventory management scenarios.

To conduct XYZ classification, a new set of features / variables are derived. Coefficient of Variation (*'CoV'*) to measure demand variability. This new feature helps to categorize products based on their demand patterns, which is crucial for forecasting accuracy and inventory planning. The coefficient of variation (CV) is a statistical measure that expresses the extent of variability in relation to the mean of the population. It is defined as the ratio of the standard deviation (σ) to the mean (μ), often expressed as a percentage (Brown, 1998): -

$$CV = \frac{\sigma}{\mu}$$

Conventionally if CV is less than or equal to 0.5, than that product is classified as X, while if the CV is between 0.5 and 1, it can be assigned as Y and if the CV is greater than 1, which shows the most uncertainty in sales/demand pattern, it can be assigned as Z (Cook and Nauman, 2014). Same convention is followed for this research. Two new features are derived for calculating *'CoV'* which are *'AverageSales'* and *'std_dev'*.

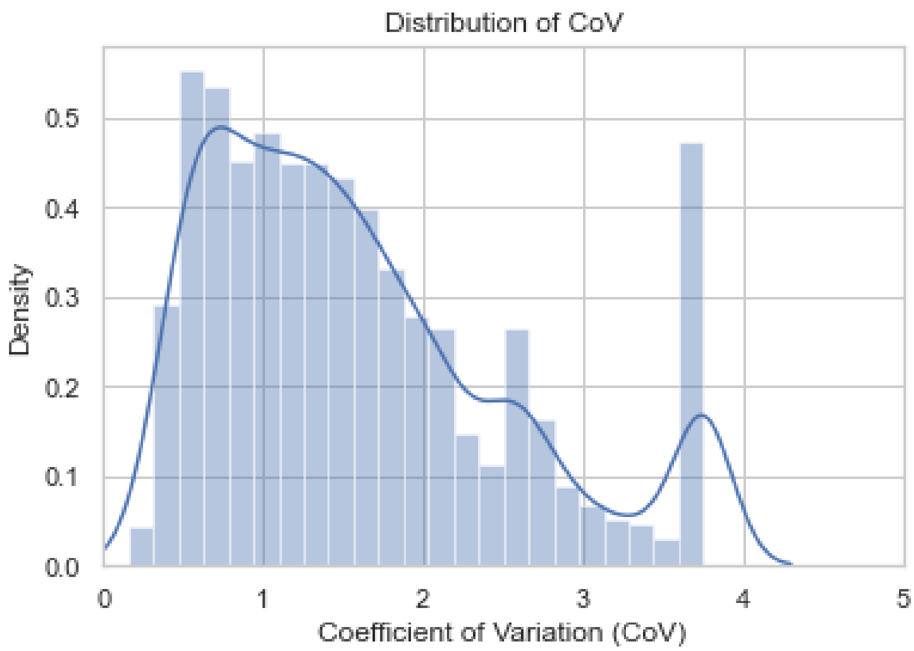


Figure 7.2: - Distribution of Coefficient of Variation (CoV)

Figure 7 demonstrates the distribution of the Coefficient of Variation (*'CoV'*) for a dataset, which measures the relative variability of data. The *'CoV'* values are spread across a range, with a noticeable peak around 1, indicating that many items have moderate variability. The distribution shows that most items have a *'CoV'* between 0 and 2, suggesting relatively stable demand, while there are fewer items with higher variability (*'CoV'* > 2). The distribution is somewhat right skewed, with occasional higher peaks, indicating some items with very high variability.

7.5 Data Modelling

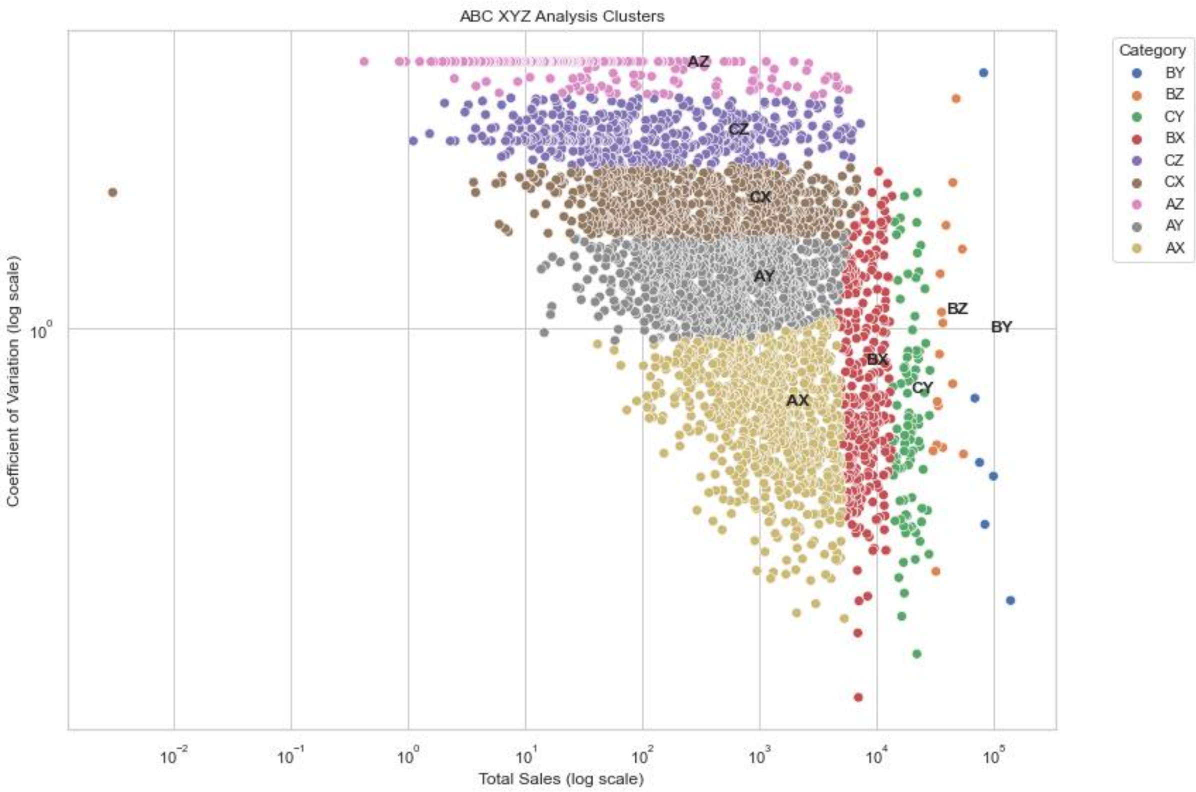
Data modelling involves creating a conceptual representation of the data structures and relationships within a system to analyse, visualize, and make predictions based on the data. It serves as a blueprint for organizing and using data effectively in various applications (Cooper, 2000). Clustering is used as a part of data modelling to group similar data points together based on certain characteristics. It helps in identifying patterns, segmenting data, and uncovering hidden structures within datasets which is essential for predictive modelling, customer segmentation, and other analytical tasks (Rodriguez et al., 2019).

K-Means clustering is used in this project to categorize inventory items into distinct groups based on their sales patterns and demand variability. The primary goal is to improve the traditional ABC XYZ classification by leveraging K-Means' ability to identify natural groupings in data. This enhances inventory management by enabling more precise segmentation, which is critical for developing tailored inventory strategies. K-Means is particularly suitable due to its efficiency in handling large datasets and its effectiveness in creating well-separated clusters that align with business objectives (Porter and Michele, 2018).

Initial step for clustering was to perform scaling on the concerned variables to normalize different features that may have vastly different ranges. In our case, we have features like *'TotalSales'* and Coefficient of Variation (*'CoV'*) which are likely on very different scales. Without scaling, the feature with larger values (probably *'TotalSales'*) would dominate the analysis. After the scaling process, K-Means clustering is performed with the value of *k* set to 9 highlighting the 9 categories of ABC XYZ classification on which the clustering is performed. This approach assumes that the clusters are ordered in a specific way, with 9 total clusters (3x3 grid for ABC-XYZ). The cluster numbers are expected to be in the range 0-8, where clusters 0-2 correspond to 'C' category, clusters 3-5 correspond to 'B' category, clusters 6-8 correspond to 'A' category and within each ABC category. Thereafter, clusters 0, 3, 6 correspond to 'X' category, clusters 1, 4, 7 correspond to 'Y' category and clusters 2, 5, 8 correspond to 'Z' category. This method provides a quick way to assign ABC-XYZ categories based on cluster numbers, assuming the clustering algorithm has grouped the products in a way that aligns with the ABC-XYZ classification logic.

Figure 7.3: - ABC XYZ cluster analysis

Figure 8 illustrates an ABC XYZ classification of products based on their total sales and



demand variability, offering insights into inventory management strategies. High-sales, low-variability items (AX) are critical, demanding consistent stock availability due to their stable demand and significant revenue contribution. Conversely, low-sales, high-variability items (CZ) are less predictable and contribute minimally to revenue, suggesting they should be stocked sparingly or managed with a make-to-order strategy which is discussed further. This clustering approach allows for tailored inventory strategies, optimising stock levels, reducing costs and enhancing service levels in line with business objectives.

Some of the key observations are most products fall into the middle categories (AY, BY, CY), there's a clear distinction between A, B, and C categories in terms of sales volume. The variability (X, Y, Z) shows more overlap between categories. Few products fall into extreme categories like AZ (high sales, high variability) or CX (low sales, low variability). Overall, this analysis helps in prioritizing inventory management strategies. For example, AX products might benefit from lean inventory practices, while CZ products might require more careful forecasting and potentially higher safety stocks as demonstrated in *Table 5*.

Category	Distribution	Cluster description
AY	911	High sales but with increasing variability
AX	831	High sales, low variability - ideal products
CX	715	Low sales with varying levels of demand stability
CZ	464	Low sales with varying levels of demand stability
AZ	339	High sales but with increasing variability
BX	288	Medium sales with varying levels of demand stability
CY	95	Low sales with varying levels of demand stability
BZ	16	Medium sales with varying levels of demand stability
BY	6	Medium sales with varying levels of demand stability

Table 7.4: - Category distribution and its attributes

Table 5 summarizes the distribution of products across different ABC XYZ categories based on sales and demand variability.

AX: - High sales, low variability; ideal for consistent stocking.

AY and AZ: - High sales but increasing variability, requiring flexible management.

CX and CZ: - Low sales with varying stability, suggesting minimal stocking.

BX, BY, BZ: - Medium sales with varying demand stability, needing moderate inventory attention.

This categorization helps tailor inventory strategies to optimise stock levels and improve overall efficiency.

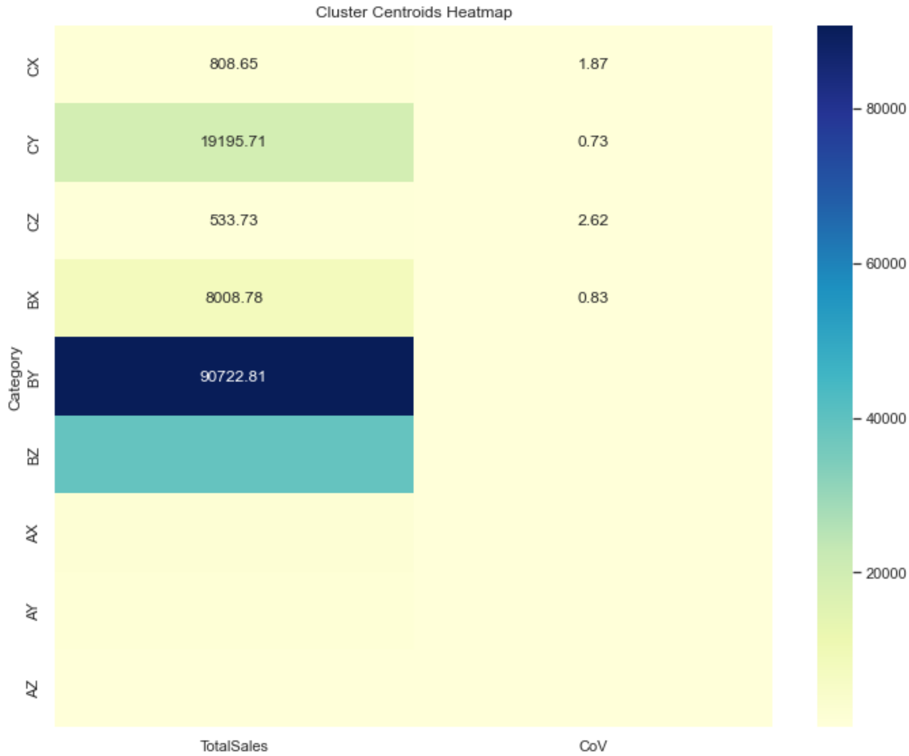


Figure 7.4: - Cluster centroid heatmap

Figure 9 illustrates a heatmap which visualizes the centroids of clusters in the ABC XYZ analysis based on two dimensions; *'TotalSales'* and the Coefficient of Variation (*'CoV'*). *'TotalSales'* indicates the average sales within each cluster, with dark colours representing higher sales. For example, the BY cluster has the highest average sales at £ 90,722.81. Coefficient of Variation (*'CoV'*) indicates demand variability with higher values meaning more variability. For instance, the CZ cluster has the highest '*CoV* 'at 2.62, indicating the most unpredictable demand. This visualization helps identify and differentiate clusters by both sales' performance and demand stability. A much more comprehensive illustration of the heatmap is presented below in *Table 6.*

Cluster	*TotalSales*	*CoV*	Category
0	808.653179	1.868223	CX
1	19195.70568	0.726976	CY
2	533.731379	2.619211	CZ
3	8008.775329	0.832378	BX
4	90722.81167	0.979979	BY
5	38746.3175	1.076594	BZ
6	1662.708161	0.680862	AX
7	880.627669	1.266235	AY
8	239.253481	3.667997	AZ

Table 7.5: - Cluster centroid description

7.6 Segmentation of ABC XYZ clusters into MTO MTS strategies

The MTO (Make-to-Order) and MTS (Make-to-Stock) segmentation performed with ABC-XYZ classification is a strategic approach to inventory and production management. It helps in optimise inventory levels, reduce costs, and improve customer service by aligning production and stocking strategies with item value and demand patterns. It allows companies to focus resources on the most critical items while efficiently managing less critical ones (Scholz-Reiter et al, 2012).

AX, AY, BX and BY are categorized as MTS since they are typically produced and stocked in advance, as they have high sales and/or stable demand (low variability). This strategy ensures that these high-priority items are readily available to meet customer demand.

Items under CX, CY, CZ, AZ, and BZ have either low sales, high demand variability, or both. These items are produced only when there is a confirmed order, minimizing inventory holding costs for less predictable or lower-priority items. Although, CX items can be sometimes segmented into MTS depending upon its variability and sale quotient. Similarly, BY can be categorized into MTO for the same. There is no rule of thumb that must be followed for the segmentation.

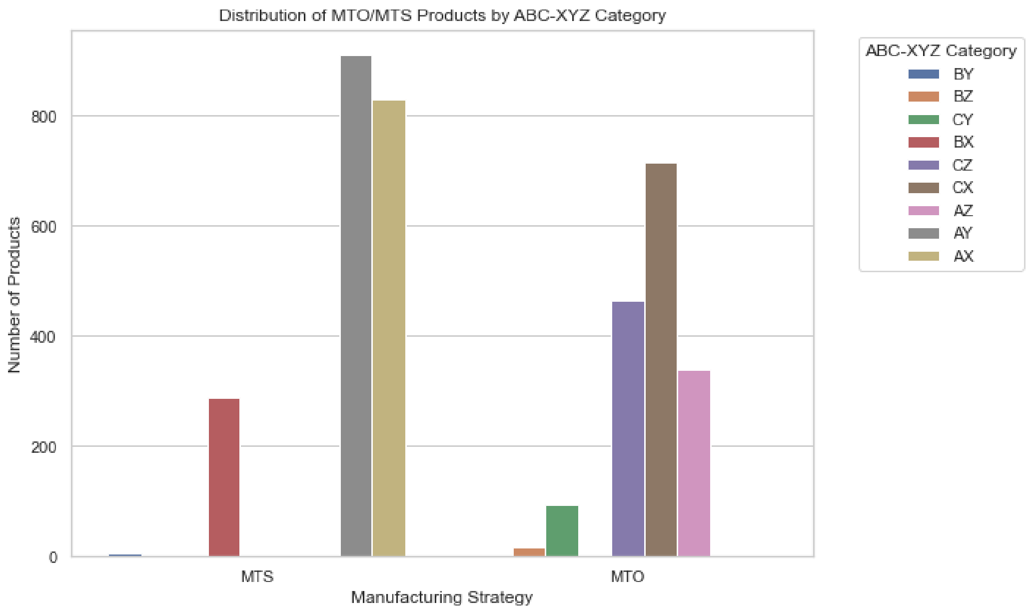


Figure 7.5: - Distribution of MTO / MTS products by ABC XYZ category

Figure 10 visualizes the distribution of products categorized under Make-to-Stock (MTS) and Make-to-Order (MTO) strategies according to their ABC XYZ classification.

- MTS: - Categories AX and AY dominate, indicating that many high-demand, stable products are pre-produced and stocked. BX also has a notable presence, suggesting moderate-demand items are similarly managed.
- MTO: - Categories CX and CZ are most prevalent, meaning many low-demand or highly variable items are produced only when ordered. The graph highlights how MTO is applied to manage less predictable or lower-priority items, minimizing inventory holding costs.

Categories	AX	AY	AZ	BX	BY	BZ	CX	CY	CZ
MTO	0	0	339	0	0	16	715	95	464
MTS	831	911	0	288	6	0	0	0	0

Table 7.6: - MTO / MTS distribution by ABC-XYZ category

Table 7 demonstrates the distribution of MTO / MTS by ABC-XYZ category with actual numbers.

7.7 Summary

Category	*StockCode*	*TotalSales*	*AverageSales*	*CoV*
BX	288	2309413	177647.2	0.833783
CY	95	1823592	140276.3	0.726976
AX	831	1387083	106698.7	0.6799
AY	911	801903.7	61684.9	1.26555
BZ	16	619941.1	47687.78	1.076594
CX	715	578187	44475.92	1.868223
BY	6	544336.9	41872.07	0.979979
CZ	464	247651.4	19050.1	2.619211
AZ	339	81106.93	6238.995	3.667997

Table 7.7: - Final summary statistics of the by ABC XYZ categories

Table 8 provides a detailed summary of the ABC XYZ classification after applying the MTO (Make-to-Order) and MTS (Make-to-Stock) segmentation strategies.

- High *'TotalSales'*, Low *'CoV'* (AX, AY, BX): - These categories, particularly AX and AY, show significant *'TotalSales'* and *'AverageSales'* with low demand variability (*'CoV'*). This confirms their suitability for the MTS strategy, ensuring that these high-priority items are always in stock to meet demand reliably.

- Low *'TotalSales'*, High *'CoV'* (CZ, AZ): - Categories like CZ and AZ have low *'TotalSales'* and high *'CoV'*, indicating high demand variability and unpredictability. These are ideal products for the MTO strategy, where products are only manufactured when there is an actual order, reducing the risk of overproduction and minimizing inventory holding costs.

- Moderate *'TotalSales'*, Moderate *'CoV'* (CX, CY): - Categories like CX and CY, with moderate *'TotalSales'* and variability, require a balanced approach. For instance, CX, which has a higher *'CoV'*, should be managed carefully under MTO to avoid excess inventory. Conversely, CY, with a relatively lower *'CoV'*, might still benefit from an MTS approach, but with cautious stock levels.

The combination of ABC classification with MTO/MTS segmentation allows for a tailored approach to inventory management. High-demand and low-variability items are stocked in advance, ensuring availability and meeting customer expectations, while low-demand and high-variability items are produced on order, minimizing unnecessary inventory. This method effectively balances inventory levels, reduces holding costs, and improves overall operational efficiency, making the supply chain more responsive and cost-effective.

8 Discussion and Critical Analysis

8.1 Overview

The Analysis chapter provided a comprehensive analysis of inventory management using an enhanced ABC XYZ classification integrated with K-Means clustering. This approach effectively segments inventory based on both sales volume and demand variability, enabling more accurate and data-driven strategies for inventory optimisation. The analysis revealed the distribution of products across high-value (A), medium-value (B), and low-value (C) categories. This segmentation allows for prioritized inventory management, focusing resources on the most valuable items that contribute significantly to overall revenue whereas by categorizing products based on their demand variability (X for stable, Y for moderate, Z for highly variable) offered insights into forecasting challenges and required safety stock levels for different product groups. The matrix resulting from combining ABC and XYZ classifications provided a nuanced view of product characteristics. For instance, AX items (high-value, stable demand) require different inventory strategies compared to CZ items (low-value, highly variable demand) and further segmentation into Make-to-Order (MTO) and Make-to-Stock (MTS) strategies based on the ABC-XYZ classification offered a strategic approach to production planning and inventory management.

8.2 Implications

The results procured in the previous chapter did implicate that the effectiveness of the ABC-XYZ classification method in categorizing inventory items based on value and demand variability is at par. The results demonstrated significant improvements in inventory management performance, particularly in optimising stock levels and reducing holding costs. It showed that the high-priority items (e.g., AX, AY) are identified for consistent availability, while low-priority items are managed more efficiently which basically answers one of the research questions of this dissertation which deals with effectiveness of ABC-XYZ classification and its impact on inventory management performance (Kabanova, Malakhova and Zenkova, 2022).

The successful integration of ABC-XYZ classification with MTO and MTS strategies, showed clear benefits in inventory management optimisation. MTS strategies were applied to high-demand, stable items, ensuring availability, while MTO strategies minimized inventory costs for items with unpredictable demand based on the results. The segmentation aligned well with the proposed classification, demonstrating potential benefits in operational efficiency (Kampen et al., 2012).

The analysis also leveraged K-Means clustering to enhance the ABC-XYZ classification, confirming that machine learning techniques can improve accuracy and efficiency. The clustering method provided a more nuanced segmentation of inventory, leading to more precise inventory strategies. This approach was found to be superior to traditional methods by offering deeper insights into demand patterns and variability, ultimately enhancing the overall classification process (Porter and Michele, 2018).

8.3 Research Limitations

From the analysis, it can be inferred that the analysis relies heavily on historical sales data, which may not accurately reflect future trends or account for unexpected changes in demand. This limitation can affect the reliability of the inventory classifications and subsequent management strategies. Since, the study does not incorporate real-time data, it could potentially lead to delays in responding to changes in demand patterns or supply chain disruptions. This affects the timeliness and responsiveness of inventory decisions (Akcay, Biller and Tayur, 2011).

Also, the assumption that the demand patterns might remain consistent overlooks potential fluctuations caused by external factors like market trends, seasonality, or economic shifts. This could result in misclassification of items and suboptimal inventory strategies, which also brings us to the supply chain variability, since the study does not account for supply chain risks or lead time variability, which are crucial factors in inventory management as suggested by Timonina-Farkas, Glogg and Seifert (2022) in their research. Sudden supply chain disruptions could render the MTO/MTS segmentation ineffective, leading to stockouts or excess inventory. For example, Covid - 19 wave in 2020.

While K-Means clustering enhances classification, the scalability of the model may be limited when applied to much larger datasets or more complex inventory systems. The computational complexity and resource requirements could increase significantly. Also, the use of K-Means clustering with a predefined number of clusters may lead to over-segmentation, where items are classified into too many categories, complicating inventory management rather than simplifying it. Having said that, since we know that clustering process is sensitive to the initial selection of centroids, which could eventually introduce bias and affect the consistency of the classification. Different runs of the algorithm may yield slightly different results, impacting the robustness of the inventory strategies developed (Thammano A and Kesisung P, 2013).

Lastly, one could point out that since the findings are based on a specific dataset and context, it might limit the generalizability of the results to other industries or types of inventory systems. The applicability of the proposed methods in different settings remains uncertain without further validation.

8.4 Summary

In summary, this research employed an enhanced ABC XYZ classification integrated with K-Means clustering to optimise inventory management by segmenting items based on sales volume and demand variability. This approach effectively categorized high-priority items for Make-to-Stock (MTS) and low-priority items for Make-to-Order (MTO), improving stock levels and reducing costs. The study confirmed that machine learning techniques like K-Means clustering enhance classification accuracy and efficiency. However, limitations include reliance on historical data, lack of real-time responsiveness, potential over-segmentation, and challenges in generalizing results across different industries or datasets. Further validation is required to ensure broader applicability and robustness in diverse contexts.

9 Conclusion and Further Research

9.1 Conclusion

This dissertation explored the application of the ABC-XYZ classification method integrated with K-Means clustering for optimising inventory management. Through a systematic approach of combining these methods, the research has demonstrated significant improvements in stock level optimisation and cost reduction. The results indicate that the ABC-XYZ classification is highly effective in categorizing inventory based on value and demand variability, leading to more precise inventory strategies and better resource allocation.

The integration of Make-to-Order (MTO) and Make-to-Stock (MTS) strategies with the ABC-XYZ classification further enhances inventory management. By aligning production strategies with inventory classifications, the research shows that businesses can achieve greater operational efficiency. High-demand, stable items are managed through MTS strategies, ensuring their consistent availability, while MTO strategies minimize inventory costs for items with unpredictable demand. This tailored approach to inventory management addresses the unique needs of different product categories, improving service levels and reducing the risk of stockouts or excess inventory (Wolff et al., 2021).

However, the research also highlights several limitations. The heavy reliance on historical sales data may not accurately reflect future trends or account for unexpected changes in demand. This limitation affects the reliability of inventory classifications and subsequent management strategies. The lack of real-time data integration could delay responses to changes in demand patterns or supply chain disruptions, reducing the timeliness and responsiveness of inventory decisions (Akcay, Biller and Tayur, 2011). Moreover, the assumption that demand patterns remain consistent overlooks potential fluctuations caused by external factors such as market trends, seasonality or economic shifts. These factors could lead to the misclassification of items and suboptimal inventory strategies (Timonina-Farkas, Glogg and Seifert, 2022).

The study also identifies potential challenges in the scalability of K-Means clustering when applied to larger datasets or more complex inventory systems. The computational complexity and resource requirements could increase significantly, posing limitations on the model's practicality for broader applications. Additionally, the sensitivity of the clustering process to the initial selection of centroids may introduce bias and affect the consistency of the classification results (Thammano A and Kesisung P, 2013).

9.2 Further Research

To address the limitations identified in this study, future research should focus on incorporating real-time data into the inventory classification process. This would enhance the responsiveness and adaptability of inventory strategies, allowing businesses to respond more effectively to sudden changes in demand or supply chain conditions. One of the primary benefits of utilizing real-time data in the ABC-XYZ classification is the improvement in demand forecasting accuracy. Traditional methods often rely on historical data, which may not accurately reflect current market conditions. Scholz-Reiter et al. (2012) emphasize that integrating demand forecasts with ABC-XYZ analysis can significantly enhance item classification performance, leading to better inventory control and reduced costs associated with overstocking or stockouts (Scholz-Reiter et al., 2012). This integration allows businesses to adjust their inventory strategies dynamically, ensuring that high-value items (A-class) are always available while minimizing the holding costs of lower-value items (C-class) (Gabriel & Nurcahyo, 2022).

Moreover, real-time data facilitates more effective inventory monitoring and management. For instance, the use of RFID technology for automatic inventory monitoring enables businesses to track inventory levels in real-time, providing critical data that can be fed into optimisation models (Erlangga et al., 2022). This real-time visibility allows for timely replenishment decisions, ensuring that inventory levels align with current demand patterns.

The proactive nature of real-time data management supports enhanced demand forecasting accuracy and reduces the likelihood of stockouts, thereby improving overall supply chain efficiency (Harahap, 2024).

Additionally, the application of advanced analytical techniques, such as machine learning and data analytics, further enhances inventory optimisation. An AI-based holistic model that utilizes real-time data to analyse vast amounts of information, identify patterns, and develop effective inventory strategies (Choudhary, 2024). This approach not only streamlines inventory management but also allows for the identification of trends that can inform future inventory decisions, thus optimising the entire inventory lifecycle.

Exploring alternative clustering techniques, such as hierarchical clustering or density-based clustering, may offer better scalability and stability, making them more suitable for large-scale applications (Irvan Prama Defindal and Nopriadi Saputra, 2023). Further validation of the proposed methods in different industrial contexts or with larger datasets is necessary to assess their generalizability and robustness across various inventory management scenarios. This would help determine the applicability of the ABC-XYZ classification and K-Means clustering integration in different settings, ensuring that the methods are effective across a wide range of industries.

Additionally, future research should consider integrating external factors such as market trends, seasonality, and economic shifts into the inventory classification model. By accounting for these factors, the model could provide more accurate demand forecasting and inventory optimisation strategies. This would improve the overall effectiveness of inventory management, helping businesses maintain optimal stock levels and reduce costs even in the face of unpredictable market conditions (Kavishka T, Rupasinghe S, 2019)

In conclusion, while the integration of ABC-XYZ classification with K-Means clustering offers a powerful approach to inventory management, addressing its limitations through further research and development is crucial. By enhancing the model's adaptability, scalability, and accuracy, future studies can contribute to more effective inventory management practices, ensuring that businesses can remain competitive and responsive in a dynamic market environment.

10 References

Akcay, A., Biller, B. and Tayur, S. (2011) 'Improved inventory targets in the presence of limited historical demand data,' *Manufacturing & Service Operations Management*, 13(3), pp. 297–309. https://doi.org/10.1287/msom.1100.0320.

Albert, J. and Rizzo, M. (2011) 'Exploratory data analysis,' in *Use R!*, pp. 133–151. https://doi.org/10.1007/978-1-4614-1365-3_5.

Arit, Thammano., Pannee, Kesisung. (2013). Enhancing K-means algorithm for solving classification problems. 1652-1656. Available from: 10.1109/ICMA.2013.6618163

Bhattacharya, A., Sarkar, B. and Mukherjee, S.K. (2007). Distance-based consensus method for ABC analysis. *International Journal of Production Research*, 45(15), pp.3405–3420. doi:https://doi.org/10.1080/00207540600847145.

Boylan, J.E., Syntetos, A.A. and Karakostas, G.C. (2008). Classification for forecasting and stock control: a case study. *Journal of the Operational Research Society*, 59(4), pp.473–481. doi:https://doi.org/10.1057/palgrave.jors.2602312.

Brown, C.E. (1998) 'Coefficient of variation,' in *Springer eBooks*, pp. 155–157. https://doi.org/10.1007/978-3-642-80328-4_13.

Chithra, Y., Prathibha, Kiran., Manoj, P, B. (2022). The Novel Method for Data Preprocessing CLI. 117-120. Available from: 10.53759/aist/978-9914-9946-1-2_21

Choudhary, S. (2024). Ai-based holistic model for decision-support system in inventory optimisation using deep learning. International Journal for Research in Applied Science and Engineering Technology, 12(3), 422-430. https://doi.org/10.22214/ijraset.2024.58733

Cook, D. and Nauman, E. (2014). *Journal of Biomechanics | ScienceDirect.com by Elsevier*. [online] www.sciencedirect.com. Available at: https://www.sciencedirect.com/journal/journal-of-biomechanics.

Cooper, R. (2000) 'Data Modelling: Introduction and state of the art,' in *Springer eBooks*, pp. 549–552. https://doi.org/10.1007/978-3-642-59623-0_48.

Dissanayake, E V A. (2023). Research Onion: A Systematic Approach for Designing Research Methodology. Part One. 10.13140/RG.2.2.24466.12489.

Erlangga, S., Yunita, A., & Satriana, S. (2022). Development of automatic real time inventory monitoring system using rfid technology in warehouse. Joiv International Journal on Informatics Visualization, 6(3), 636. https://doi.org/10.30630/joiv.6.3.1231

Evdokimova, S.A. (2021) 'Segmentation of store customers to increase sales using ABC-XYZ-analysis and clustering methods,' *Journal of Physics Conference Series*, 2032(1), p. 012117. https://doi.org/10.1088/1742-6596/2032/1/012117.

Gabriel, D. S. and Nurcahyo, R. (2022). Integration of abc-xyz analysis in inventory management optimisation: a case study in the health industry. Proceedings of the International Conference on Industrial Engineering and Operations Management. https://doi.org/10.46254/af03.20220070

Garn, W. (2020). *Introduction toManagement ScienceModelling, Optimisations and Probability*. 1st ed. [online] Smartana Ltd, pp.199–203. Available at: https://bibliu.com/app/#/view/books/9781527284234/pdf2htmlex/index.html#page_203.

Garn, W. and Aitken, J. (2015). *Splitting hybrid Make-To-Order and Make-To-Stock demand profiles*. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.1504.03594.

Harahap, A. (2024). Enhancing supply chain efficiency: implementation of vendor managed inventory in inventory routing problem. Information Management and Business Review, 16(2(I)S), 212-218. https://doi.org/10.22610/imbr.v16i2(i)s.3761

Hayes, A. (2020). *Make to order: Everything you need to know*. [online] Investopedia. Available at: https://www.investopedia.com/terms/m/make-to-order.asp.

Herlambang, C. and Parung, J. (2021). Information system design and inventory management on pharmacy business within abc-xyz analysis method. Airlangga Journal of Innovation Management, 2(2), 194. https://doi.org/10.20473/ajim.v2i2.31124

Irvan Prama Defindal and Nopriadi Saputra (2023). Applying Machine Learning on ABC-XYZ Inventory Model Using Multivariate and Hierarchical Clustering. *Advances in Social Science, Education and Humanities Research/Advances in social science, education and humanities research*, [online] pp.322–334. doi:https://doi.org/10.2991/978-2-38476-132-6_30.

Islam, M.A. and Al-Shiha, A. (2018) 'Basic Summary Statistics,' in *Springer eBooks*, pp. 39–72. https://doi.org/10.1007/978-981-10-8627-4_2.

Jenkins, A. (2022). *Inventory Management: 20 Challenges, 20 Solutions*. [online] Oracle NetSuite. Available at: https://www.netsuite.com/portal/resource/articles/inventory-management/inventory-management-challenges.shtml.

Kabanova, T., Malakhova, T. and Zenkova, Z. (2022). ABC-XYZ inventory analysis accounting for change points. *International Journal of Logistics Systems and Management*, 41(1/2), p.59. doi:https://doi.org/10.1504/ijlsm.2022.120981.

Kandemir, B. (2022) 'A Methodology for Clustering Items with Seasonal and Non-seasonal Demand Patterns for Inventory Management,' *Journal of Advanced Research in Natural and Applied Sciences*, 8(4), pp. 753–761. https://doi.org/10.28979/jarnas.1112146.

Karaca, Z. (2018) 'THE CLUSTER ANALYSIS IN THE MANUFACTURING INDUSTRY WITH K-MEANS METHOD: AN APPLICATION FOR TURKEY,' *Eurasian Journal of Economics and Finance*, 6(3), pp. 1–12. https://doi.org/10.15604/ejef.2018.06.03.001.

Khan, S.A.R., Dong, Q.L. and Yu, Z. (2017) 'Role of ABC analysis in the process of efficient order Fulfillment: case study,' *Advanced Engineering Forum*, 23, pp. 114–121. https://doi.org/10.4028/www.scientific.net/aef.23.114.

Kourentzes, N., Petropoulos, F. and Trapero, J.R. (2014) 'Improving forecasting by estimating time series structural components across multiple frequencies,' *International Journal of Forecasting*, 30(2), pp. 291–302. https://doi.org/10.1016/j.ijforecast.2013.09.006.

Krzyżaniak, S. (2017). *Influence of delivery quantity on service level in the stock replenishment system based on reorder level for irregular distributions of demand*. [online] typeset.io. Available at: https://typeset.io/papers/influence-of-delivery-quantity-on-service-level-in-the-stock-3jd8f46qme [Accessed 20 Jul. 2024].

Kundu, K., Rossini, M., & Costa, F. (2022). Implementing workload control through lean-based project in mto-mts environment. Journal of Manufacturing Technology Management, 33(5), 893-912. https://doi.org/10.1108/jmtm-05-2021-0183

LečiÄ‡-CvetkoviÄ‡, D., Atanasov, N. and BabarogiÄ‡, S. (2010). An Algorithm for Customer Order Fulfillment in a Make-to-Stock Manufacturing System. *INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL*, [online] 5(5), pp.783–791. Available at: https://univagora.ro/jour/index.php/ijccc/article/view/2238 [Accessed 29 Jul. 2024].

Melnikovas, A. (2018). *Towards an Explicit Research Methodology: Adapting Research Onion Model for Futures Studies * Journal of Futures Studies*. [online] Journal of Futures Studies. Available at: https://jfsdigital.org/articles-and-essays/2018-2/towards-an-explicit-research-methodology-adapting-research-onion-model-for-futures-studies/#:~:text=The%20research%20onion%2C%20proposed%20by.

Noll, A. (2020). *mts versus mto: wann make-to-stock für Ersatzteile einsetzen?* [online] Andreas Noll, Ihr Unternehmensberater für Ersatzteilmanagement. Available at: https://no-stop.de/make-to-stock-mts-fuer-ersatzteile/ [Accessed 19 Jul. 2024].

Pandya[1], B. and Thakkar, H., 2016. A review on inventory management control techniques: ABC-XYZ analysis. REST Journal on Emerging trends in Modelling and Manufacturing, 2(3), p.2016.

Porter and Michele, D. (2018). *Using K-means clustering to create cost and demand functions that decrease excess inventory and better manage inventory in defense*. [online] dspace.mit.edu. Available at: https://dspace.mit.edu/handle/1721.1/117798.

Qaffas, A., Hajkacem, M., Ncir, C., & Nasraoui, O. (2023). Interpretable multi-criteria abc analysis based on semi-supervised clustering and explainable artificial intelligence. Ieee Access, 11, 43778-43792. https://doi.org/10.1109/access.2023.3272403

Rodriguez, M.Z., Comin, C.H., Casanova, D., Bruno, O.M., Amancio, D.R., Costa, L. da F. and Rodrigues, F.A. (2019). Clustering algorithms: A comparative approach. *PLOS ONE*, [online] 14(1), p.e0210236. doi:https://doi.org/10.1371/journal.pone.0210236.

Roser, C. (2021). *Make-to-Order vs Make-to-Stock: The ABC XYZ Analysis | AllAboutLean.com*. [online] www.allaboutlean.com. Available at: https://www.allaboutlean.com/mto-vs-mts-2/ [Accessed 19 Jul. 2024].

Ross, D.F. (2015). Managing Supply Chain Inventories. In: Distribution Planning and Control. Springer, New York, NY. https://doi.org/10.1007/978-1-4899-7578-2_7

Saunders, M.N.K., Lewis, P., Thornhill, A. and Bristow, A. (2019). *Understanding research philosophy and approaches to theory development*. [online] oro.open.ac.uk. Available at: https://oro.open.ac.uk/53393/.

Scholz-Reiter, B. *et al.* (2012) 'Integration of demand forecasts in ABC-XYZ analysis: practical investigation at an industrial company,' *International Journal of Productivity and Performance Management*, 61(4), pp. 445–451. https://doi.org/10.1108/17410401211212689.

Sinaga, K.P. and Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm. *IEEE Access*, [online] 8, pp.80716–80727. doi:https://doi.org/10.1109/access.2020.2988796.

S.M., Hatefi., Seyed, Ali, Torabi. (2015). A Common Weight Linear Optimisation Approach for Multicriteria ABC Inventory Classification. 2015, 1-11. Available from: 10.1155/2015/645746

Spyridoula Karkani (2023). *Make to Order vs Make to Stock: A Comparative Study*. [online] The Megaventory Blog. Available at: https://blog.megaventory.com/2023/05/make-to-order-vs-make-to-stock-a-comparative-study/.

Stojanovic, M. and Regodić, D. (2017). The significance of the integrated multicriteria abc-xyz method for the inventory management process. Acta Polytechnica Hungarica, 14(5), 29-48. https://doi.org/10.12700/aph.14.5.2017.5.3

Taylor, S. (2022). *Quantitative Analysis*. [online] Corporate Finance Institute. Available at: https://corporatefinanceinstitute.com/resources/data-science/quantitative-analysis/.

Teunter, R.H., Babai, M.Z. and Syntetos, A.A. (2009). ABC Classification: Service Levels and Inventory Costs. *Production and Operations Management*, [online] 19(3), pp.343–352. doi:https://doi.org/10.1111/j.1937-5956.2009.01098.x.

Thieuleux, E. (2022). *ABC XYZ Analysis In Inventory Management: Example In Excel | AbcSupplyChain*. [online] AbcSupplyChain. Available at: https://abcsupplychain.com/abc-xyz-analysis/.

Thushan, Kavishka, Abeysekara., Sulochana, Rupasinghe. (2019). Analysis of Influential Factors for Inventory Forecasting Systems. 2019, 1-4. Available from: 10.1109/I2CT45611.2019.9033725

Timonina-Farkas, A., Glogg, R.Y. and Seifert, R.W. (2022) 'Limiting the impact of supply chain disruptions in the face of distributional uncertainty in demand,' *Production and Operations Management*, 31(10), pp. 3788–3805. https://doi.org/10.1111/poms.13781.

Van Kampen, T.J., Akkerman, R. and Van Donk, D.P. (2012) 'SKU classification: a literature review and conceptual framework,' *International Journal of Operations & Production Management*, 32(7), pp. 850–876. https://doi.org/10.1108/01443571211250112.

Wirth, R. and Hipp, J. (2000). *CRISP-DM: Towards a Standard Process Model for Data Mining*. [online] Available at: https://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf.

Wolff, K., Janine Tatjana Maier, Heuer, T., Nyhuis, P. and Schmidt, M. (2021). Effects Of Different Order Processing Strategies On Operating Curves Of Logistic Models: A Comparison Of Make-to-Order And Make-to-Stock. *Journal of Production Systems and Logistics*, 2021(1), p.18. doi:https://doi.org/10.15488/11532.

Yeung, A. (2023). *6 Effective Inventory Optimisation Techniques You Should Use*. [online] ThoughtSpot. Available at: https://www.thoughtspot.com/data-trends/analytics/inventory-optimisation-techniques.

Yilmaz, U. (2023). *ABC-XYZ Inventory Classification with Python*. [online] Medium. Available at: https://medium.com/@ulas_yilmaz/abc-xyz-inventory-classification-with-python-50ebee552fe4.

Zaerpour, N., Rabbani, M., Gharehgozli, A.H. and Tavakkoli-Moghaddam, R. (2008). Make-to-order or make-to-stock decision by a novel hybrid approach. *Advanced Engineering Informatics*, 22(2), pp.186–201. doi:https://doi.org/10.1016/j.aei.2007.10.002.

Zamazii, O. (2022). The relevance of using ABC inventory analysis in profitability management of the enterprise with the purpose of optimising working capital. *Economic Analysis*, 32(32(4)), pp.178–185. doi:https://doi.org/10.35774/econa2022.04.178.

11 Appendix

11.1 Code

```
# -*- coding: utf-8 -*-
"""
Created on Wed Aug 28 01:43:46 2024

@author: nimis
"""

# -*- coding: utf-8 -*-
"""
Created on Wed Aug 28 01:28:23 2024

@author: nimis
"""
#%%

# For suppressing the warnings
import warnings
warnings.filterwarnings("ignore")

# Loading necessaryt packages
import pandas as pd
import numpy as np
```

```
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

#%%

# Loading the dataset
df = pd.read_csv('C:\\Nimish\\Imp docs\\University of Surrey\\Business Analytics Dissertation\\online_retail_II.csv', parse_dates=['InvoiceDate'])



#%%

#%%

# Displaying the attribute names of the dataframe
print('\n\n\n----------------------------------------------------------------------')
print('Attribute Names of the Dataframe')
print('----------------------------------------------------------------------')
print(df.columns)
print('---------------------------------------------------------------------- \n\n\n')

# Displaying the top 5 observations of the DataFrame to verify that the data has been read correctly
print('---------------------------------------------------------------------------')
print('Top 5 observations of the DataFrame')
print('---------------------------------------------------------------------------')
print(df.head())
print('---------------------------------------------------------------------------\n\n\n')

# Displaying the last 5 observations of the DataFrame to verify that the data has been read correctly
print('---------------------------------------------------------------------------')
print('Last 5 observations of the DataFrame')
print('---------------------------------------------------------------------------')
```

```
print(df.tail())
print('---------------------------------------------------------------------------\n\n\n')

# Displaying the types and information about data
print('---------------------------------------------------------------------------')
print('Types and Information about DataFrame')
print('---------------------------------------------------------------------------')
print(df.info())
print('---------------------------------------------------------------------------\n\n\n')

#%%

#%%

# ---------------------------------------
# Missing value identification
# ---------------------------------------

# Missing Values Calculation
ms = df.isnull().sum()
# Calculating the percentage of missing values in each column
ms_percentage = (df.isnull().sum()/(len(df)))*100
# Combining the missing value information into one dataframe
Missing_Data_Info = pd.DataFrame({'Total Missings': ms, 'Percentage': ms_percentage})
# Printing them the missing value information on screen
print('---------------------------------------')
print('Missing Data Information')
print('---------------------------------------')
print(Missing_Data_Info)
print('---------------------------------------\n\n\n')

# Removing rows where 'Description' or 'Customer ID' is blank (NaN) and where 'Price' and 'Quantity' is negative (< = 0)
new_dataframe = df.dropna(subset=['Description', 'Customer ID'])
```

```
new_dataframe = new_dataframe.drop(new_dataframe[new_dataframe['Price'] <= 0].index)
new_dataframe = new_dataframe.drop(new_dataframe[new_dataframe['Quantity'] <= 0].index)

# Checking minimum value of 'Price' variable
min_price = new_dataframe['Price'].min()
print(min_price)

# Displaying the first few rows to confirm the operation
print(new_dataframe.head())

# --------------------------------------------------
# Repeating - Missing value identification
# --------------------------------------------------

# Missing Values Calculation
ms2 = new_dataframe.isnull().sum()
# Calculating the percentage of missing values in each column
ms_percentage2 = (new_dataframe.isnull().sum()/(len(new_dataframe)))*100
# Combining the missing value information into one dataframe
Missing_Data_Info2 = pd.DataFrame({'Total Missings': ms2, 'Percentage': ms_percentage2})
# Printing them the missing value information on screen
print('--------------------------------------------------')
print('Missing Data Information - After Dropping Attribute')
print('--------------------------------------------------')
print(Missing_Data_Info2)
print('--------------------------------------------\n\n\n')

#%%

#%%

# -------------------
# Data Statistics
# -------------------
```

```
# Statistical summary of data belonging to numerical datatype such as int, float
Data_Stat = new_dataframe.describe().T
print('---------------------------------------------------------------------------')
print('Data Summary')
print('---------------------------------------------------------------------------')
print(Data_Stat)
print('---------------------------------------------------------------------------\n\n\n')

# Dataframe without 'Customer ID' variable
new_dataframe_1 = new_dataframe.drop('Customer ID', axis = 1)
Data_Stat_1 = new_dataframe_1.describe().T
print('---------------------------------------------------------------------------')
print('Data Summary for Numerical Variables')
print('---------------------------------------------------------------------------')
print(Data_Stat_1)
print('---------------------------------------------------------------------------\n\n\n')

# Providing a statistical summary of all data, include object, category etc
Data_Stat_All = new_dataframe.describe(include='all').T
pd.set_option('display.max_columns', None)
print('---------------------------------------------------------------------------')
print('Data Summary for all variables')
print('---------------------------------------------------------------------------')
print(Data_Stat_All)
print('---------------------------------------------------------------------------\n\n\n')

# Providing a statistical summary of all categorical variables
#new_dataframe_2 = new_dataframe.drop('Quantity', 'Price', axis = 1)
Data_Stat_cat = new_dataframe.describe(include=['object', 'category'])
print('---------------------------------------------------------------------------')
print('Data Summary for Categorical Variables')
print('---------------------------------------------------------------------------')
print(Data_Stat_cat)
```

```
print('------------------------------------------------------------------------------\n\n\n')

#%%

#%%

# Deriving new features

# Extracting year and month from 'InvoiceDate'
new_dataframe['InvoiceDate'] = pd.to_datetime(new_dataframe['InvoiceDate'], dayfirst=True)
new_dataframe['year'] = pd.DatetimeIndex(new_dataframe['InvoiceDate']).year
new_dataframe['month'] = pd.DatetimeIndex(new_dataframe['InvoiceDate']).month

# Calculating Revenue for ABC analysis
new_dataframe['Revenue'] = new_dataframe['Quantity'] * new_dataframe['Price']

# Calculating product_metrics by grouping 'StockCode'
product_metrics = new_dataframe.groupby(['StockCode','year', 'month']).agg({
    'Revenue': ['sum']
}).reset_index()
product_metrics["month"] = product_metrics.month.map("{:02}".format)
product_metrics['year_month']=product_metrics['year'].map(str) + '-' +
product_metrics['month'].map(str)

product_metrics.columns = ['StockCode','year', 'month', 'TotalRevenue', 'year_month']
product_metrics = product_metrics.pivot(index='StockCode',columns='year_month',
values='TotalRevenue').reset_index().fillna(0)

# Calculating 'TotalSales'
product_metrics['TotalSales'] = product_metrics.iloc[:,1:13].sum(axis=1,numeric_only=True)

# Calculating 'AverageSales' on monthly basis
product_metrics['AverageSales'] = product_metrics['TotalSales']/13
```

```
# Calculating 'std_dev' for CoV calculation
product_metrics['std_dev'] = product_metrics.iloc[:,1:13].std(axis=1)

# Calculate Coefficient of Variation for XYZ analysis
product_metrics['CoV']=product_metrics['std_dev']/product_metrics['AverageSales']
product_metrics['CoV'] = product_metrics['CoV'].replace([np.inf, -np.inf], np.nan)

# Handle NaN and infinite values
product_metrics['CoV'] = product_metrics['CoV'].fillna(product_metrics['CoV'].max())
product_metrics['TotalSales'] = product_metrics['TotalSales'].replace([np.inf, -np.inf], np.nan)
product_metrics['TotalSales'] =
product_metrics['TotalSales'].fillna(product_metrics['TotalSales'].max())
product_metrics['AverageSales'] = product_metrics['AverageSales'].replace([np.inf, -np.inf], np.nan)
product_metrics['AverageSales'] =
product_metrics['AverageSales'].fillna(product_metrics['AverageSales'].max())

product_metrics.head()
#%%

#%%

# Visualising the 'TotalSales' graph

sns.distplot(product_metrics["TotalSales"], kde=True)
plt.title('Distribution of the Total Sales')
plt.xlabel('Total Sales')
plt.ylabel('Density')
plt.show()

#%%

#%%

# ABC Analysis using Pareto Principle
```

```
product_metrics = product_metrics.sort_values('TotalSales', ascending=False)
product_metrics['CumulativeSales'] = product_metrics['TotalSales'].cumsum()
product_metrics['SalesFraction'] = product_metrics['CumulativeSales'] /
product_metrics['TotalSales'].sum()

def condition_abc(x):
    if x <= 0.80:
        return "A"
    elif x <= 0.90:
        return "B"
    else:
        return 'C'

product_metrics['ABC'] = product_metrics['SalesFraction'].apply(condition_abc)

#%%

#%%

# Visualising the coefficient of variance graph
sns.distplot(product_metrics["CoV"])
plt.xlim(0, 5)  # Adjust this range based on the distribution
plt.title('Distribution of CoV')
plt.xlabel('Coefficient of Variation (CoV)')
plt.ylabel('Density')
plt.show()

#%%

#%%

# XYZ Analysis using Coefficient of Variation
def condition_xyz(x):
    if x <= 0.5:
```

```
        return "X"
    elif x>=0.5 and x <= 1:
        return "Y"
    else:
        return 'Z'

product_metrics['XYZ'] = product_metrics['CoV'].apply(condition_xyz)

#%%

#%%

# Clustering for ABC XYZ classification

# Prepare features for clustering
features = product_metrics[['TotalSales', 'CoV']]
# Normalize features
scaler = StandardScaler()
features_scaled = scaler.fit_transform(features)

# Perform K-means clustering
kmeans = KMeans(n_clusters=9, random_state=42)
product_metrics['Cluster'] = kmeans.fit_predict(features_scaled)

# Function to assign ABC XYZ categories
def assign_category(cluster):
    abc = ['C', 'B', 'A']
    xyz = ['X', 'Y', 'Z']
    return abc[cluster // 3] + xyz[cluster % 3]

product_metrics['Category'] = product_metrics['Cluster'].apply(assign_category)

# Display results
```

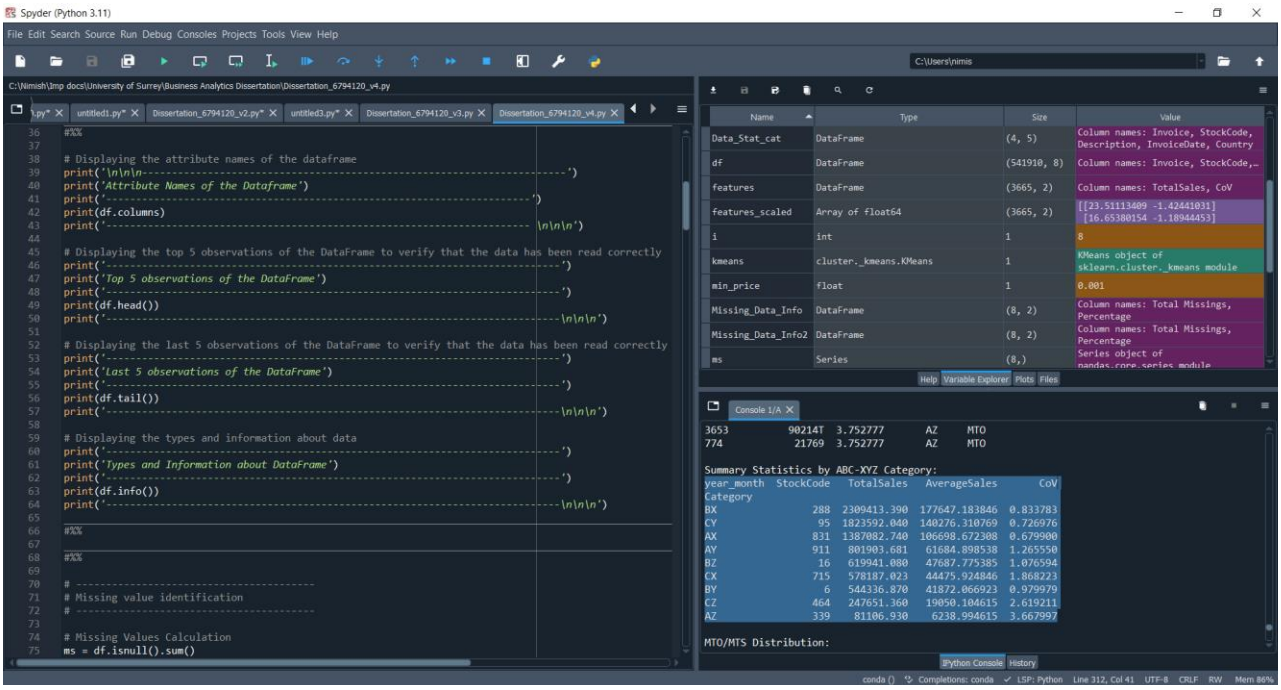
```
print(product_metrics[['StockCode', 'TotalSales', 'CoV', 'Category']].sort_values('TotalSales',
ascending=False).head(20))

# Summary statistics
print("\nCategory Distribution:")
print(product_metrics['Category'].value_counts())

#%%

#%%

# Cluster Visualisation

# Cluster centroids
centroids = scaler.inverse_transform(kmeans.cluster_centers_)
centroid_df = pd.DataFrame(centroids, columns=['TotalSales', 'CoV'])
centroid_df['Category'] = [assign_category(i) for i in range(9)]
print("\nCluster Centroids:")
print(centroid_df)

# Visualization
plt.figure(figsize=(12, 8))
sns.scatterplot(data=product_metrics, x='TotalSales', y='CoV', hue='Category', palette='deep', s=50)

# Add cluster centroids
centroids = scaler.inverse_transform(kmeans.cluster_centers_)
for i, centroid in enumerate(centroids):
    plt.annotate(assign_category(i), centroid, fontsize=12, fontweight='bold')

plt.xscale('log')
plt.yscale('log')
plt.xlabel('Total Sales (log scale)')
plt.ylabel('Coefficient of Variation (log scale)')
plt.title('ABC XYZ Analysis Clusters')
```

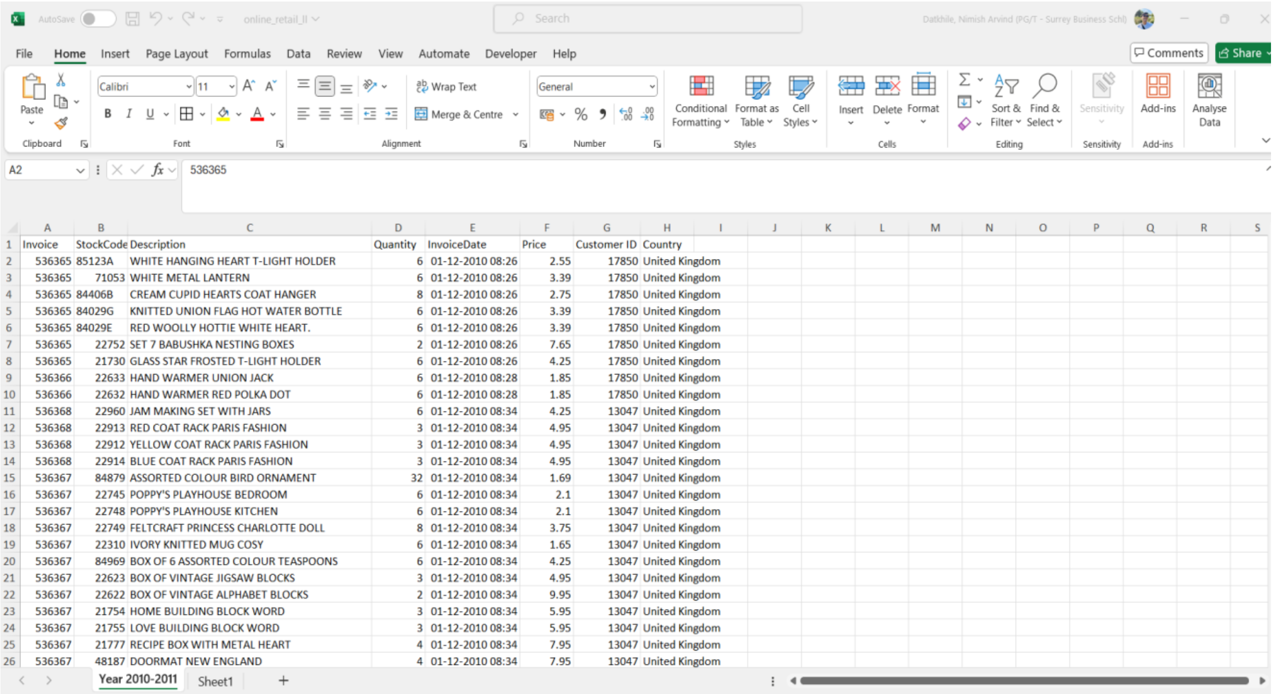
```
plt.legend(title='Category', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()

# Additional visualization: Heatmap of cluster centroids
centroid_df = pd.DataFrame(centroids, columns=['TotalSales', 'CoV'])
centroid_df['Category'] = [assign_category(i) for i in range(9)]
centroid_df = centroid_df.set_index('Category')

plt.figure(figsize=(10, 8))
sns.heatmap(centroid_df, annot=True, fmt='.2f', cmap='YlGnBu')
plt.title('Cluster Centroids Heatmap')
plt.tight_layout()
plt.show()

#%%

#%%

# Segmenting ABC XYZ into MTO and MTS Strategies

def mto_mts_category(abc_xyz):
    if abc_xyz in ['AX', 'AY', 'BX', 'BY']:
        return 'MTS'
    else:
        return 'MTO'

# Apply the categorization
product_metrics['MTO_MTS'] = product_metrics['Category'].apply(mto_mts_category)

# Visualize the MTO/MTS distribution
plt.figure(figsize=(10, 6))
sns.countplot(x='MTO_MTS', hue='Category', data=product_metrics)
plt.title('Distribution of MTO/MTS Products by ABC-XYZ Category')
```

```
plt.xlabel('Manufacturing Strategy')
plt.ylabel('Number of Products')
plt.legend(title='ABC-XYZ Category', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()

# Print summary statistics
print("\nMTO/MTS Distribution:")
print(product_metrics['MTO_MTS'].value_counts())

print("\nMTO/MTS Distribution by ABC-XYZ Category:")
print(product_metrics.groupby(['MTO_MTS', 'Category']).size().unstack(fill_value=0))

#%%

#%%

# Printing final summary statistics

print("\nABC-XYZ Distribution:")
print(product_metrics['Category'].value_counts())

print("\nMTO/MTS Distribution:")
print(product_metrics['MTO_MTS'].value_counts())

print("\nTop 10 products by sales:")
print(product_metrics[['StockCode', 'TotalSales', 'Category', 'MTO_MTS']].head(10))

print("\nTop 10 products by variability:")
print(product_metrics[['StockCode', 'CoV', 'Category', 'MTO_MTS']].sort_values('CoV',
ascending=False).head(10))

# Comparing the 'TotalSales' with 'AvgSales'
```

```
print("\nSummary Statistics by ABC-XYZ Category:")
print(product_metrics.groupby('Category').agg({
    'StockCode': 'count',
    'TotalSales': 'sum',
    'AverageSales': 'sum',
    'CoV': 'mean'
}).sort_values(by = 'TotalSales', ascending = False))

print("\nMTO/MTS Distribution:")
print(product_metrics['MTO_MTS'].value_counts())


#%%
```

11.2 Snippets





Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01-12-2010 08:26	2.55	17850	United Kingdom
536365	71053	WHITE METAL LANTERN	6	01-12-2010 08:26	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	01-12-2010 08:26	2.75	17850	United Kingdom
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	01-12-2010 08:26	3.39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	01-12-2010 08:26	3.39	17850	United Kingdom
536365	22752	SET 7 BABUSHKA NESTING BOXES	2	01-12-2010 08:26	7.65	17850	United Kingdom
536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	01-12-2010 08:26	4.25	17850	United Kingdom
536366	22633	HAND WARMER UNION JACK	6	01-12-2010 08:28	1.85	17850	United Kingdom
536366	22632	HAND WARMER RED POLKA DOT	6	01-12-2010 08:28	1.85	17850	United Kingdom
536368	22960	JAM MAKING SET WITH JARS	6	01-12-2010 08:34	4.25	13047	United Kingdom
536368	22913	RED COAT RACK PARIS FASHION	3	01-12-2010 08:34	4.95	13047	United Kingdom
536368	22912	YELLOW COAT RACK PARIS FASHION	3	01-12-2010 08:34	4.95	13047	United Kingdom
536368	22914	BLUE COAT RACK PARIS FASHION	3	01-12-2010 08:34	4.95	13047	United Kingdom
536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	01-12-2010 08:34	1.69	13047	United Kingdom
536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	01-12-2010 08:34	2.1	13047	United Kingdom
536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	01-12-2010 08:34	2.1	13047	United Kingdom
536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	01-12-2010 08:34	3.75	13047	United Kingdom
536367	22310	IVORY KNITTED MUG COSY	6	01-12-2010 08:34	1.65	13047	United Kingdom
536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	01-12-2010 08:34	4.25	13047	United Kingdom
536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	01-12-2010 08:34	4.95	13047	United Kingdom
536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	01-12-2010 08:34	9.95	13047	United Kingdom
536367	21754	HOME BUILDING BLOCK WORD	3	01-12-2010 08:34	5.95	13047	United Kingdom
536367	21755	LOVE BUILDING BLOCK WORD	3	01-12-2010 08:34	5.95	13047	United Kingdom
536367	21777	RECIPE BOX WITH METAL HEART	4	01-12-2010 08:34	7.95	13047	United Kingdom
536367	48187	DOORMAT NEW ENGLAND	4	01-12-2010 08:34	7.95	13047	United Kingdom