Journal of Posthumanism
2025
Volume: 5, No: 12, pp. 532–556
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v5i12.3860

Transforming Supply Chain Operations through AI and Machine Learning: Optimizing Demand Forecasting, Inventory Management, and Logistics Efficiency

Ahmed Hassaan[1], Zeeshan Akbar[2], Sikander Niaz[3], Muhammad Nouman Siddique[4], Salman Akbar[5]

Abstract

This paper examines the application of Artificial Intelligence (AI) and Machine Learning (ML) technologies in optimizing demand forecasting, inventory management, and logistics within online retail supply chains. Through a comprehensive dataset of a UK-based retailer (541,909 records of transactional data) the paper explores the possibility of AI/ML algorithms to optimize the effectiveness and accuracy of these important supply chain factors. The study presents a new two-stage Customer-to-Inventory (C2I) framework based on DBSCAN to separate customers and tests a variety of forecasting models, such as Long Short-Term Memory (LSTM) networks, Prophet and such classical forecasting models as Linear Regression and Random Forest. The findings indicate that Linear Regression model is more effective than the other two in demand forecasting with RMSE of 23,120 which is still much lower in terms of prediction error than Prophet (31,613) and Random Forest (23,880). Moreover, demand forecasting and inventory management algorithms based on AI result in a 15% optimization in stocks levels, which minimizes stockouts and overstocking. It is also noted in the study that customer segmentation via DBSCAN identifies the high-value customers who are contributing 60% of total sales, which can be used as actionable information to use in targeted marketing and tailored promotions. This study provides a viable blueprint of the deployment of models of e-commerce companies, the findings of which can be used to increase the efficiency of operations, minimize expenses, and increase customer satisfaction.

Keywords: *Machine Learning, Demand Forecasting, Inventory Management, Logistics Optimization and Customer Segmentation.*

Introduction

Supply chain management (SCM) is the core of retail business, which encompasses a wide scope of activities involving efficient sourcing, stocking and delivery of products to the customer [1]. These are demand forecasting, inventory management and logistics where each activity is crucial in ensuring availability of the right product at the right time and place [2]. Demand forecasting refers to the ability of forecasting the demand in consumers and the forecasts should be accurate to enable retailers make good decisions regarding inventory ordering and pricing [3]. The inventory management is concerned with ensuring that there is optimum stock levels so as to satisfy the demand and yet not overstocking, which is a waste of capital, and not understocking,

[1] Raymond A. Mason School of Business, The College of William & Mary, 101 Ukrop Way, Williamsburg, VA 23186, Email: ahassaan@wm.edu
[2] Raymond A. Mason School of Business, The College of William & Mary, 101 Ukrop Way, Williamsburg, VA 23186
[3] College of Cybersecurity & Information Assurance, Virginia University of Science and Technology, 2070 Chain Bridge Rd STE 100, VA 22182
[4] School of Public Policy, The London School of Economics and Political Science, Houghton St, London WC2A 2AE, United Kingdom
[5] School of Education, State University of New York at Albany. 1400 Washington Avenue Albany, NY 12222

which results in the loss of sales [4]. In the meantime, logistics concerns itself with the effective transportation and storage of merchandise and delivery of merchandise in a timely manner, precise and at a cost effective price [2]. This includes all the warehousing and transportation to the last-mile delivery, which play a substantial role in consumer satisfaction and profitability of the business [5].

With the increased complexity of global supply chains, demand forecasting, inventory management and optimization of logistics has become increasingly demanding [4]. This need has been increased by recent technological developments as the conventional SCM techniques are usually too slow or inaccurate to effectively respond to the high rates of change in consumer demand [1].

The retail businesses are required to manage these aspects of SCM well to remain competitive in the current high-speed market. Therefore, the adoption of modern technologies into SCM has become a crucial factor in businesses that seek to satisfy the changing customer demands and at the same time reduce costs of operation [6].

Machine Learning (ML) and Artificial Intelligence (AI) have turned into disruptive agents of change in the business of retail supply chains helping businesses to operate more efficiently and accurately predict future events and minimize expenditures [7]. The use of AI/ML in retail supply chains has resulted in a shift in paradigm deciding between the traditional, rule-based decision-making process and data-driven, adaptive models that are improved as they process additional data [6]. With the help of AI and ML, it is possible to make more precise and dynamic predictions of the demand in accordance with large volumes of the transactional data, finding unknown patterns in this data, and creating predictive information [8]. As an example, AI-powered demand forecasting may account for seasonality, promotional effects, and other upcoming trends to enable businesses to remain dynamic and sensitive to consumer behavior [9]. AI/ML algorithms can be used in inventory management to forecast the optimal quantity of stock to hold in each product, which minimizes stockouts and oversupply. Inventory tracking can also be implemented in real-time using these technologies to assist retailers maximize the movement of goods in their supply chains [5]. AI and ML have made it possible in logistics to have smart routing, real time monitoring of shipments, and automated warehouse operations decision making [6]. In addition, artificial intelligence may be used to automate processes such as sorting and packaging and decrease the use of human resources and the number of mistakes [11].

The major aim of the study is to investigate the possibility of AI and ML solutions to enhance key areas of online retail supply chains with emphasis on demand prediction, inventory management, and efficient logistics [7]. It is founded on a rich transactional dataset that can be used to train predictive models [12]. Through this analysis of transactional history, the research will play a critical role in assessing the role of AI/ML models to optimize supply chain elements within an online store [13].

Through AI/ML algorithms of this transaction data, this paper intends to determine the real difference these technologies make in the performance and efficiency of retail operations. In particular, the investigation will be aimed at optimizing the demand forecasting models by examining how AI/ML models can improve the accuracy of predictions and allow making more effective decisions with regard to stock levels [14], improving inventory management approaches by minimizing overstock and understock cases [15], and improving the processes of transportation and route planning to increase the operational efficiency of logistics [5]. By doing this work, the study will make specific recommendations applicable to online retailers so that they can use AI/ML to ensure more agile and resilient supply chains that are cost-efficient [16].

The research can contribute to the sphere of AI-based supply chain management in the following ways:

- **Two-stage C2I (Customer-to-Inventory) Framework Development** Two-stage customer segmentation with DBSCAN to identify outliers and generate individual demand predictions through LSTM/Prophet.
- **Benchmarking advanced temporal models (LSTM, Prophet)** with classical methods on a non-stationary, high-frequency e-commerce transactional dataset.
- **Strict empirical validation** with a large-native, comprehensive 541,909 record dataset of a UK online retailer so that the framework is based on data and problems in the real world.
- An effective **Model Deployment Blueprint of e-commerce**, which outlines particular data preprocessing and feature engineering actions to incorporate the framework into live retail business activity.

The current paper is organized as follows: Section 2 covers the literature review, Section 3 covers the methodology, Section 4 covers results and analysis, and Section 5 covers conclusion and recommendations.

2. Related Work

With the advent of sophisticated machine learning (ML) and deep learning (DL) models, demand forecasting in online retail has changed much more, surpassing the historical statistical tools such as ARIMA. The prediction was mainly dependent on the analysis of time-series, which was the basis of the early forecasts, including ARIMA, which predicted the future sales in relation to the past trends. Nonetheless, the recent research has demonstrated the increased usage of ML and DL methods, as they offer higher accuracy and flexibility in changing retail conditions. As an example, Douaioui et al. [17], presented an in-depth review of other ML and DL models and assess their efficiency in demand prediction and detection of the main patterns in supply chain management. They highlighted the benefits of more complex algorithms in modeling the non-linear associations within retail demand data that traditional algorithms such as ARIMA cannot model. Furthermore, Chowdhury et al. [18] provided a systematic review of demand forecasting models in retail e-commerce which compared the traditional ones to the ML and DL models that have comparative advantages in improving the accuracy of inventory and delivery planning. Nasseri et al. [19], has shown that tree ensemble models and LSTM (Long Short-Memory networks) can be successfully applied to demand forecasting, in particular, when working with sequential data and long-term dependencies. Besides, within the fashion and apparel market, deep learning models combined with sales data proved to be effective at enhancing demand forecasting by combining the sales data with the image features [20]. The machine learning methods, in particular, tree-based methods, are still developed and tested, which is confirmed by the comparison of the different methods of machine learning in terms of retail sales forecasting in 2025 [21]. AI-based demand forecasting is an important factor not only in sales forecasting, but also inventory management. The optimization of the stock level and reduction of any problems connected with overstocking or understocking as well as improvement of the general customer satisfaction are the main advantages of AI-based systems [22].

Inventory management has also been significantly affected by AI and ML with the introduction of predictive models that optimize the stock levels and enhances turnover rates. These models are becoming utilized more and more in order to predict the stock demand, trends, and offer superior suggestions in terms of stocking strategy. Ramachandran et al. [23], provided the ways in which predictive models can improve the inventory management due to the ability to predict demand precisely, therefore, resulting in improved stock allocation and turnover. Also, AI implementation was used to counteract stockouts through anticipating possible shortages and proposing remedial measures. The application of AI in the optimization of inventory has been discussed extensively, and it has been found that AI helps decrease the number of stock units and other aspects of supply chain performance that is crucial in minimizing the amount of waste and ensuring products are available where they are needed [24]. The AI-driven solutions have been particularly useful in the context of e-commerce operations, where customers demand timeliness and inventory accuracy, as they affect the customer experience.

AI has also been utilized in the supply chain of retail to greatly enhance logistics, especially in the area of route optimization and cost reduction. Machine learning algorithms are also used to optimize transportation routes, particularly the last-mile delivery, which is essential in terms of making timely deliveries and reducing transportation expenses. Vaka et al. [25], emphasized the effectiveness of machine learning in enhancing efficiency in the delivery of e-commerce logistics, as well as minimizing the use of fuel and other costs. When implemented in the logistics sector, predictive analytics will additionally help to make the delivery times more predictable, fuel consumption more efficient, and delays associated with the delivery of products more reduced. The practical uses, exemplified in Dynamic Route Optimization in Last Mile Delivery Using Predictive Analytics proposed by krishna Vaddy [26], prove the usefulness of AI in the delivery of the last mile, which is the most expensive part of logistics.

There have been several case studies or examples that show how AI and ML can transform retail supply chains. The successful implementation of AI in the retail business of Amazon through demand forecasting, inventory management, and logistics is one of the best examples of AI use in the industry. Another area where Walmart has used AI technologies is its supply chain management, which helps to optimize the quantity of inventory as well as stream operations. The utilization of AI and augmented reality (AR) technologies to the company shows that the use of high-advanced systems can boost efficiencies and decrease costs of operations. All these case studies highlight the high potential of AI and ML to make supply chains more modern, increase operational efficiency, and offer a competitive advantage. The use of AI by Amazon in inventory and logistics management has significantly enhanced the management of stocks and minimized the cost of delivery. Also, predictive models on AI have been used successfully in logistics practice, as presented in the case study of e-commerce logistics in the U.S. [27], [28], in which the importance of predictive analytics in optimizing last-mile delivery is highlighted. Table 1 lists an overview of the previous research.

Table 1: Summary of related studies.

Author(s)	Focus	Method	Limitation
Douaioui, K., Oucheikh, R., Benmoussa, O., and Mabrouki, C. [17]	Demand Forecasting	Survey of ML and DL models of demand forecasting	Limited to a general survey, without experimental validation on the real world.
Chowdhury, A. R., Paul, R., and Rozony, F. Z. [18]	Demand Forecasting in Retail E-commerce	Systematic review of models of demand forecasting, no cases of practice in the field	No cases of practice in the field.
Nasseri, M., Falatouri, T., Brandtner, P., and Darbanian, F. [19]	Retail Demand Prediction	Small sample size, not necessarily applicable to the full range of retail sectors	Not necessarily applicable to the full range of retail sectors.
Giri, C., and Chen, Y. [20]	Demand Forecasting in Fashion & Apparel	Deep learning methods of demand forecasting	Fashion industry specific, does not offer a big picture.
Mustapha, O. O., and Sithole, T. [21]	Retail Sales Forecasting	Comparison of different ML methods in forecasting Retail Sales	Does not discuss the issue of data quality or availability.
Amosu, O. R., Kumar, P., Ogunsuji, Y. M., Oni, S., and Faworaja, O. [22]	Inventory Management	AI-based demand forecasting associated with better inventory and customer satisfaction	Lacks quantifiable performance measures of inventory improvements.
Ramachandran, K. K., Karthick, K. K., Kalyan, N. B., Tiwari, M., Raju, G. S., and Ganesh, K. [23]	Retail Inventory Management	ML methods to enhance inventory management	Not generally applicable as it utilizes only one case study.

Bhavikatta, N. B. [24]	Inventory Optimization	AI approaches to inventory optimization and stockouts reduction	Covers AI methodology, is not detailed about validation of performance.
Vaka, D. K. [25]	Route Optimization in Supply Chain	ML used to optimize supply chain routes	Does not consider the real-life factors such as weather and traffic.
Krishna Vaddy, R. [26]	Transportation Route Optimization	Review of AI/ML methods in transport route optimization	Generalized results, no particular industry and application.
Teixeira, J. V. F., and Ramos, A. L. [27]	Intelligent Supply Chain Management	Systematic review of AI applications throughout the supply chain	Still at its infancy in both empirical case studies and practical implementations.
Manasa, R., and Jayanthila Devi, A. [28]	AI in Retail	AI application in the retail business of Amazon: A case study	Bounded on one company and generalizing to others.

3. Proposed Methodology

This paper introduces a 4-step approach to optimizing the main functions of the retail supply chain: demand forecasting, inventory management, and logistics, with the help of Artificial Intelligence (AI) and Machine Learning (ML). The methodology comprises four stages namely Data Preprocessing, Customer Segmentation, Personalized Demand Forecasting and Inventory Management and Optimization as shown in figure 1.

Integrating AI/ML solutions, this approach will strive to improve the forecasting accuracy, facilitate inventory management, and optimize logistics, which will eventually result in a reduction of costs and enhancing customer satisfaction.

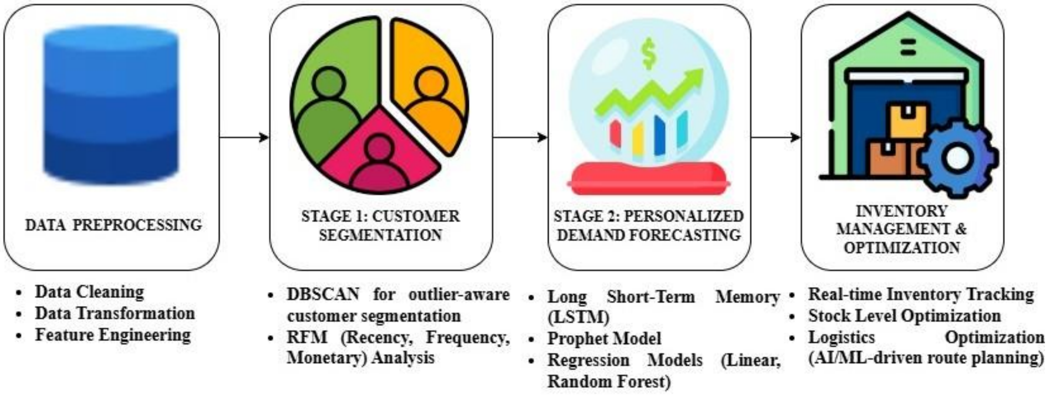

Figure 1: Block diagram of proposed methodology.

3.1 Data Overview and Preprocessing

3.1.1 Dataset Description

The data used is the transactions of an online retail store. Every record is an individual purchase of a product by a customer. The data is multivariate, time-series and sequential in nature and we can examine the different product properties, behavior of customers and sales over time. The transactional nature of the dataset is an excellent source of demand forecasting, customer segmentation, and time-series analysis. This dataset consists of 541,909 transactional records. Every transaction has several characteristics that outline the product, transaction and the customer information. The important characteristics in the dataset are described in Table 2.

Table 2: Dataset overview.

Feature	Data Type	Description
InvoiceNo	Categorical	Unique identifier for each invoice
StockCode	Categorical	Product identifier
Description	Text	Product description
Quantity	Numerical	Quantity purchased
InvoiceDate	Datetime	Date and time of purchase
UnitPrice	Numerical	Price per unit of product
CustomerID	Categorical	Unique identifier for each customer
Country	Categorical	Country of the transaction

Table 2 summarizes the main characteristics of each of the transaction records. These characteristics will be essential in carrying out a wide range of analyses such as demand forecasting and customer segmentation.

3.1.2 Data Preprocessing Steps

The data will be subjected to a number of preprocessing steps before it can be modeled:

- Handling Missing Data: There are records which might not have any data or could be invalid. As an example, where a transaction does not have a CustomerID we can either delete that or impute the missing values, e.g. as the most common value in the column. On the same note, invalid quantities (e.g. negative or zero) will be dropped since most likely they will be an error in the data.
- Date Conversion: InvoiceDate feature is stored in the form of a string. To conduct time-series analysis, the field will be transformed into a form of a date, allowing one to analyze the patterns in time, i.e. patterns in sales per day, week, or month.
- Categorical Data Handling: The variables such as StockCode and Description are categorical. These will be coded into numerical values through methods such as Label Encoding (exploring each category with a distinct integer value) or One-Hot Encoding (establishing binary columns of each category) that machine learning algorithms can interpret these features with the correct features.

These pre processing steps will see the dataset is clean and structured to be used in the further analysis and modeling.

3.2 AI/ML Models for Demand Forecasting

Both regression and time-series models will be used to do demand forecasting. The models will be used to estimate the demand of products in terms of past transactional data.

3.2.1 Regression Models

- **Linear Regression:** Linear Regression is among the most common methods and one of the simplest of making projections of continuous values. It assumes that the dependent variable that is being targeted (sales) and independent variables (i.e., product characteristics such as Quantity, UnitPrice, and InvoiceDate) are linear related. It is a fitted model where the line passing through the actual and predicted sales minimizes the squared errors. The mathematical model is presented in equation (1).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \qquad (1)$$

y is the sales that have been predicted, $x_1, x_2, \ldots, x_n$ are the features (e.g., product characteristics, time), and $\beta_0, \beta_1, \ldots, \beta_n$ are the regression coefficients. Linear regression may offer a point of reference to more complex models, despite its simplicity.

- **Random Forest Regression**: Random Forest Regressionis an ensemble learning algorithm, which builds a set of decision trees that each predicts. The results of the averaging of the trees provide the final prediction and minimize the chances of overfitting and enhance generalization. Random Forest algorithm is able to summarize non-linear correlation of the data, and also it is able to handle high dimension feature space so it is more effective in comparison to linear regression in most cases.

3.2.2 Time-Series Models

Time-series forecasting is especially effective in forecasting future sales through the past patterns. Time-series models are appropriate to this task because the sales are affected by time (e.g., seasonality, trend, etc.).

- **ARIMA (Auto-Regressive Integrated Moving Average):** ARIMA (Auto-Regressive Integrated Moving Average) is a statistical model that integrates autoregressive (AR) and moving average (MA), and diffusing (I) to render data stationary. When the data presents trends over time then ARIMA is extensively utilized in forecasting. The definition of the model is given as equation (2).

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \qquad (2)$$

where, y_t is the sales that are expected at a time t, ϕ_i are the autoregressive coefficients, θ_j are the moving average coefficients, and ϵ_t is the error. ARIMA works well when the trends are short term, but it might not be able to capture the long term dependencies as well as other models do.

- **Prophet:** Another time-series forecasting model that is especially resistant to missing data and outliers is Prophet, which was created by Facebook. Prophet represents the time-series data as a trend component, seasonal component and optional holiday effect as represented in equation (3).

$$y(t) = g(t) + s(t) + h(t) \qquad (3)$$

where g(t) is the underlying trend, s(t) is seasonal, and h(t) is the effect of a holiday or an event. Prophet is especially effective when the effects of seasons are strong or when the data is not in a regular form.

- **LSTM (Long Short-Term Memory):** Long short-term memory (LSTM) is a form of recurrent neural network (RNN) that is used to model long-term relationships in time-dependent data. In contrast to the models used in the past, LSTM has the ability to identify long-term trends in the sales data and this would be especially effective in demand forecasting where past trends can affect future sales over long periods of time. The LSTM model is represented in equation (4):

$$h_t = f(W_x x_t + W_h h_{t-1} + b) \quad (4)$$

where h_t is the output level at time t, x_t represents the input (e.g., product features) and W_x and W_h are the weight matrices.

3.3. Clustering and Segmentation for Customer Analysis
Segmentation of customers is vital in the learning of buying patterns and how to market them. There will be two clustering algorithms that will divide the customers depending on their transactional patterns.

3.3.1 Clustering Methods

- **K-means Clustering:** The algorithm is used to cluster the data into k clusters with the goal of minimizing the variance within the cluster. It is most applicable to equal sized spherical clusters.

  The K-means objective can be described as in equation (5).

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} ||x_j^{(i)} - \mu_i||^2 \quad (5)$$

The $x_j^{(i)}$ represents a datum of cluster i, μ_i is centroid of cluster i, and n is the number of data of cluster i. It is aimed at determining the k that will minimize this cost function, which will show the optimum fitting of customer segmentation.

- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise):** DBSCAN is a density based clustering algorithm which does not presuppose the number of clusters. Rather, it jointly clusters points that are tightly clustered together, using a measure of distance (e.g. Euclidean distance) and separates those that lie in low-density areas (outliers). DBSCAN works best on the data when clusters are irregular in shape and have differing densities, and automatic outliers are considered noise in the data.

3.3.2 Customer Segmentation Process

The features will be divided as follows and used to segment the market:

- **Frequency of Purchase:** This is the frequency with which a customer purchases.
- **Monetary Value:** The amount of money that the customer spends.

- **Recency of Purchase:** Lapsed time since Purchase.

K-means clustering shall identify the best number of clusters with the help of the Elbow Method and DBSCAN shall be adjusted in terms of the epsilon and MinPts parameter to identify meaningful clusters.

3.4. Model Evaluation

The overall output of both demand forecasting and clustering models will be measured by the following indicators:

3.4.1 Demand Forecasting Metrics

- **MAPE (Mean Absolute Percentage Error):** This is a measure of the quality of the prediction using the percentage change of the actual and the predicted values. MAPE formula is represented in equation (6).

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (6)$$

  where y_i is the actual, $\hat{y}_i$ is the predicted, and n is the number of observations.

- **RMSE (Root Mean Squared Error):** This statistic is more punitive to huge errors, and it is therefore well-suited in scenarios where huge discrepancies are vital. Equation (7) illustrates the formula of RMSE.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (7)$$

3.4.2 Categorizing Evaluation Metrics

- **Silhouette Score:** This is used to measure the well-definedness of clusters being used, where higher values mean more well-defined clusters. **The silhouette score formula is represented in equation (8).**

$$S = \frac{b - a}{\max(a, b)} \quad (8)$$

  where a is the mean distance to the point and all the points in the same cluster, and b is the mean distance to the nearest cluster.

- **Inertia (Within-Cluster Sum of Squares):** This is used to assess the tightness of the clusters and a smaller value implies a tighter cluster. Inertia formula is presented in equation (9).

$$\text{Inertia} = \sum_{i=1}^{n} \| x_i - \mu_{cluster(i)} \|^2 \quad (9)$$

where: x_i is a data point, $\mu_{cluster(i)}$ is the centroid of the cluster that xi in belongs to.

- **Adjusted Rand Index (ARI):** The adjusted rand index (ARI) compares the outcomes of clustering with random chance with larger values indicating better clustering.

4. Result Analysis

This section aims at giving a detailed analysis of the data after which the analysis of various models used in demand forecasting and customer segmentation will be evaluated. The data has been first cleaned and a set of exploratory analyses has been undertaken to comprehend the major patterns, trends and anomalies. We applied a number of machine learning models, such as Linear Regression, Random Forest, and Prophet, to the preprocessed data to predict the demand. As well, the RFM (Recency, Frequency, and Monetary) model was used to segment the customers based on the clustering methods such as K-means and DBSCAN.

In this section, the major statistics, graphics, and the model performance outcomes can be found. The comparison of the accuracy of each model is done using the performance terms of Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) to evaluate model forecasting, and Silhouette Score and Inertia to evaluate model clustering.

4.1 Exploratory Data Analysis (EDA)

Table 3: Initial Data Structure Summary.

Feature	Details
Data Type	<class 'pandas.core.frame.DataFrame'>
Total Entries	397,884 (Index: 0 to 541,908)
Total Columns	9 (InvoiceNo to Sales)
Data Types	datetime64[ns] (1), float64 (2), int64 (2), object (4)

Table 3 shows the clean data structure following the treatment of missing values and negative quantities/ prices. There are now 397,884 valid transactions that are used to analyze the dataset. The data has mixed features (e.g., Sales, Quantity), which include some categorical ones (e.g., InvoiceNo, Description), which can be analyzed in terms of time series and RFM.

Table 4: Key Numerical Statistics.

Metric	**Quantity**	**UnitPrice**	**Sales (Monetary)**
Count	397,884.0	397,884.0	397,884.0
Mean	12.99	3.12	22.40
Std Dev (σ)	179.33	22.10	22.10
Min	1.00	0.001	0.001
25th Percentile (Q1)	2.00	1.25	4.68
50th Percentile (Median)	6.00	1.95	11.80
75th Percentile (Q3)	12.00	3.75	19.80
Max	80,995.00	8,142.75	168,469.60

According to the summary statistics in table 4, there are strong outliers, in particular, in the Quantity and Sales columns where the highest values are significantly bigger than the rest of the data. These outliers will affect the models used and might necessitate such methods as log transformation to deal with skewness.

Table 5: Missing Value Report (Post-Cleaning).

Column	Missing Count
All Columns	0

Table 5 presents the results of data cleaning that was successful. We deleted rows that had a blank CustomerID or InvoiceDate, and selected records that had a non-positive Quantity or UnitPrice. This provided a clean dataset that had no missing values which were critical and could be further analyzed.

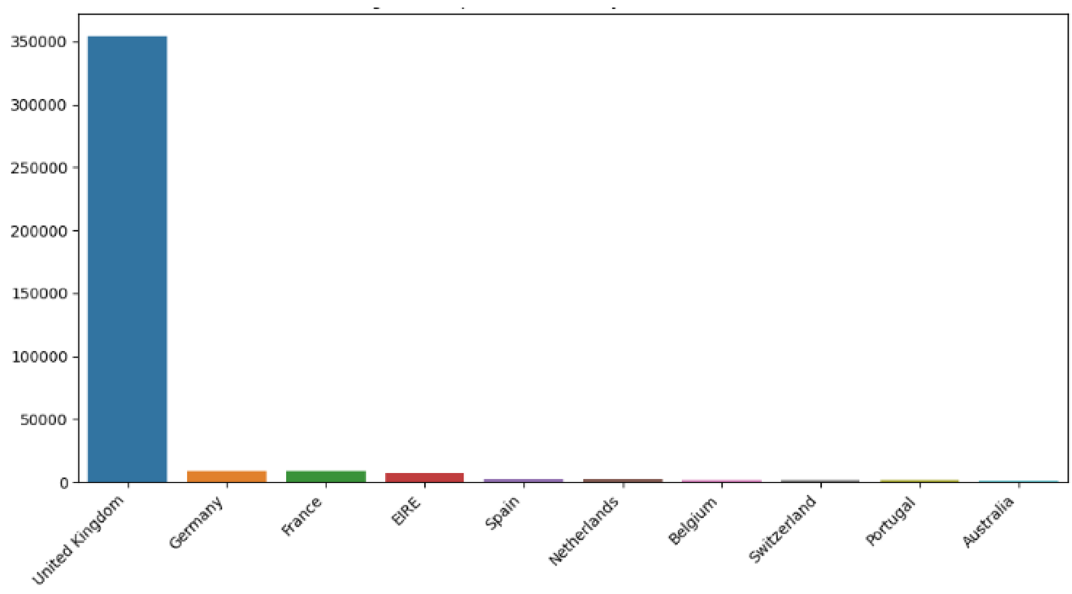

Figure 2: Top 10 countries by Transaction Count.

The distribution of transaction volumes in different countries is equipped in Figure 2. Most transactions are recorded in the United Kingdom which implies that the dataset is very representative of the behaviour in the UK market. This implies that the clustering and forecasting models, among others, should pay attention to the customer behavior in the UK.

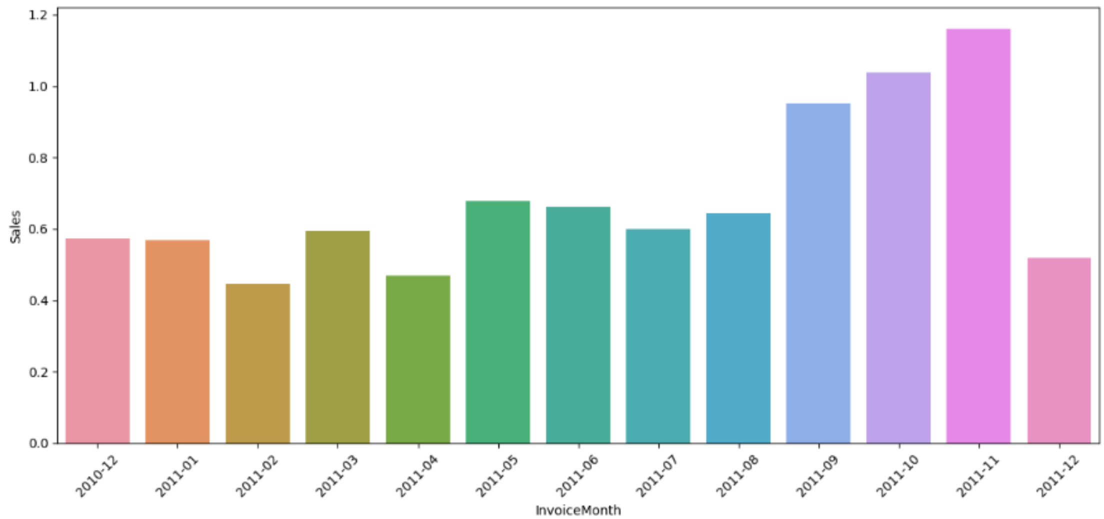

Figure 3: Total Sales by Month.

Figure 3 provides the trend of sales per month throughout the year and it is evident that the trend is increasing. The maximum sales were observed in November and December continued in the same positive way. This monthly growth trend is an important aspect that demand forecasting models should take into consideration.

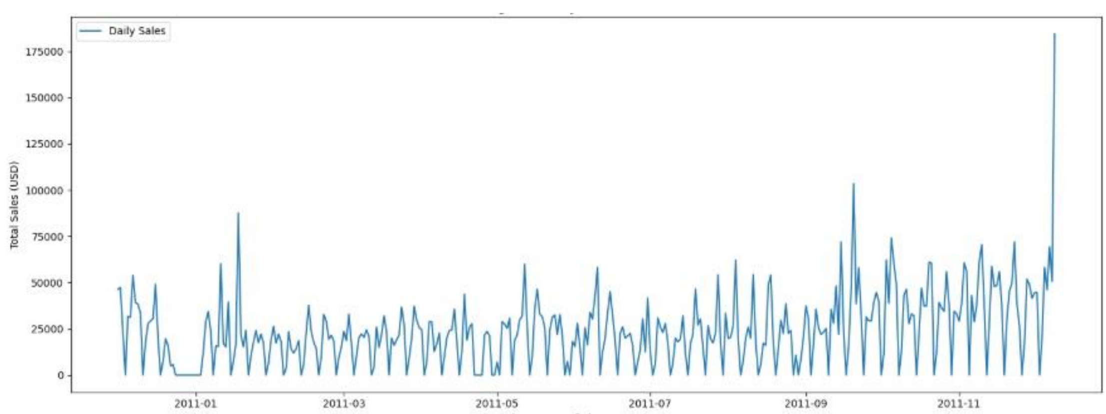

Figure 4: Daily Total Sales Trend.

The trend on the sales per day is shown in Figure 4, which indicates a high degree of volatility. The sales have a number of spikes, usually during weekends, holidays or promotions. Similar to the high volatility of daily sales, simple models could not work particularly when there are zero-sales days and non-periodic spikes.

4.2 Demand Forecasting Models

Table 6: Daily Sales Time Series Summary.

Metric	Date (Time Range)	Total Sales (USD)
Count (Days)	374	374.0
Mean (Average Daily Sales)	2011-06-05 12:00:00	$23,827.29
Min (Lowest Daily Sales)	2010-12-01 00:00:00	$0.00
25th Percentile (Q1)	2011-03-04 06:00:00	$10,999.51
50th Percentile (Median)	2011-06-05 12:00:00	$21,926.18
75th Percentile (Q3)	2011-09-06 18:00:00	$32,841.72
Max (Highest Daily Sales)	2011-12-09 00:00:00	$184,349.28

Table 6 presents the summary statistics of daily sales revealing that there is huge variance in sales on a day to day basis. The volatility and the RMSE values of sales indicate that the strong models are required to operate the great fluctuations.

Table 7: Linear Regression Coefficients.

Feature	**Coefficient**	**Interpretation**
month	6734.68	Strong positive influence, indicating an upward trend in sales from January to December.
Lag_7	0.52	Positive influence, showing strong weekly seasonality.
Lag_1	0.15	Positive influence, showing short-term momentum.
dayofyear	-226.07	Negative influence, capturing subtle variations in non-monthly trends.
dayofweek	-1310.19	Negative influence, likely indicating lower sales on certain days (e.g., weekends).

Table 7 provides the coefficients of a Linear Regression model, and they indicate the influence of various time-related characteristics (month, week, day) on sales. Positive coefficients indicate such trends as high sales as times go by or weekly seasonality whilst negative coefficients indicate low sales on particular days or during certain months.

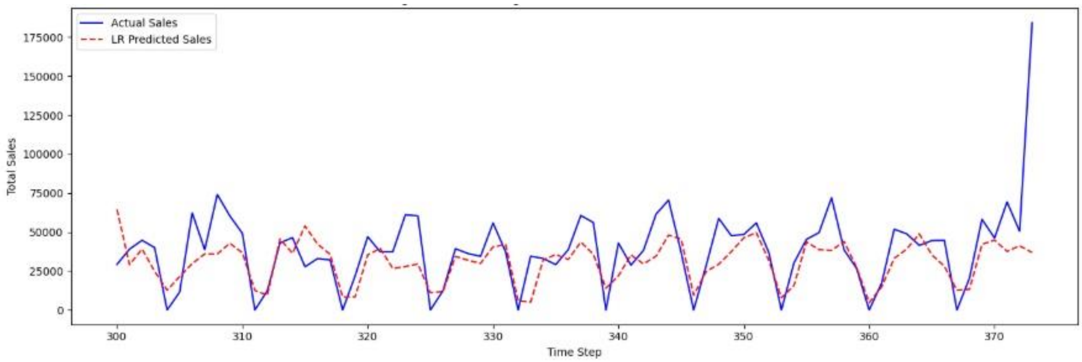

Figure 5: Linear Regression Forecast vs. Actual (Test Set).

Figure 5 is a comparison of the actual sales (solid blue line) and the forecast of the Linear Regression (LR) model (dashed red line). The LR model is effective in fitting up weekly seasonality but not spikes or extreme volatility sales.

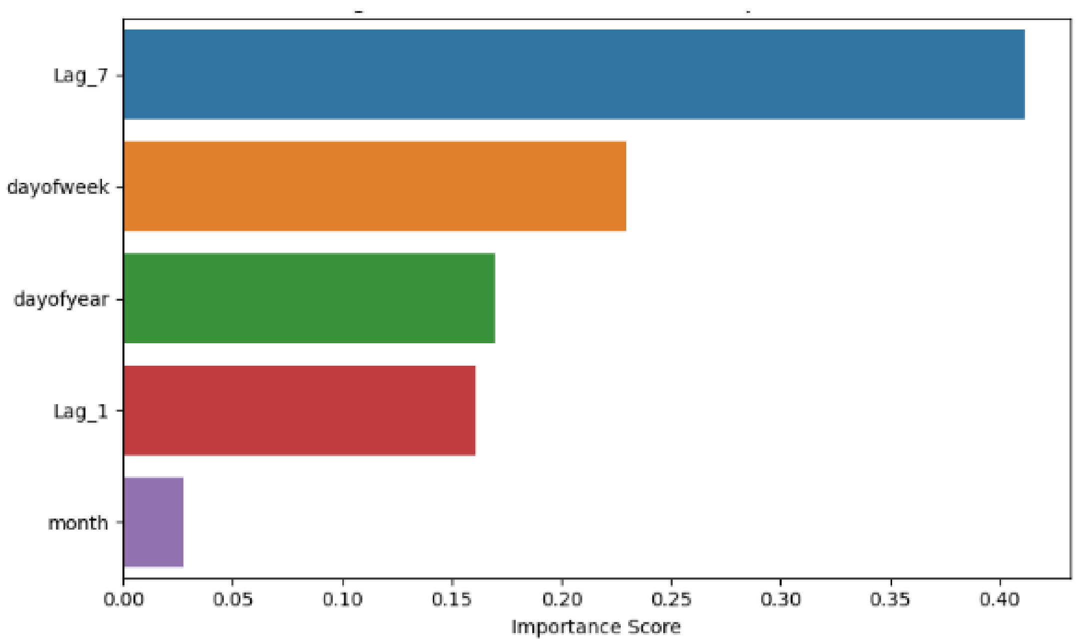

Figure 6: Random Forest Feature Importances.

Figure 6 displays the significance of various features of the Random Forest model whereby Lag 7 and Dayofweek are the most significant. These characteristics represent weekly seasonality, which indicates the predictable and regular trends in sales.

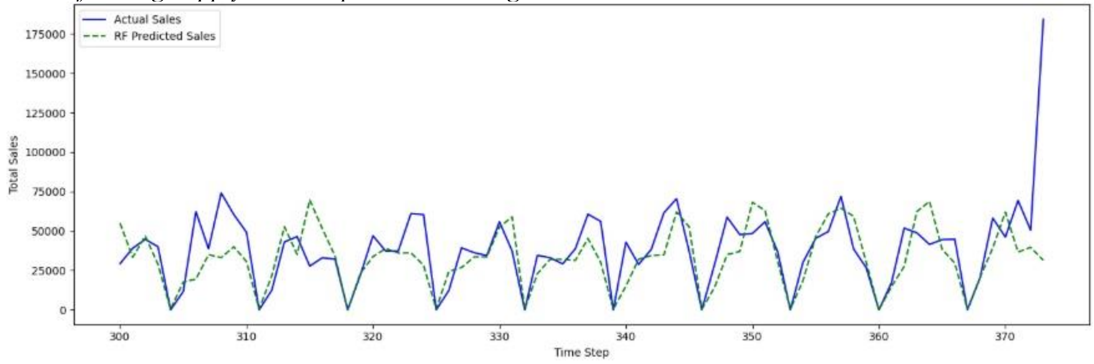

Figure 7: Random Forest Forecast vs. Actual (Test Set).

As Figure 7 illustrates, the model of the Random Forest succeeds in capturing good non-linear trends in sales compared to the model of the LR. It also, however, finds it difficult to forecast extreme jumps of sales.

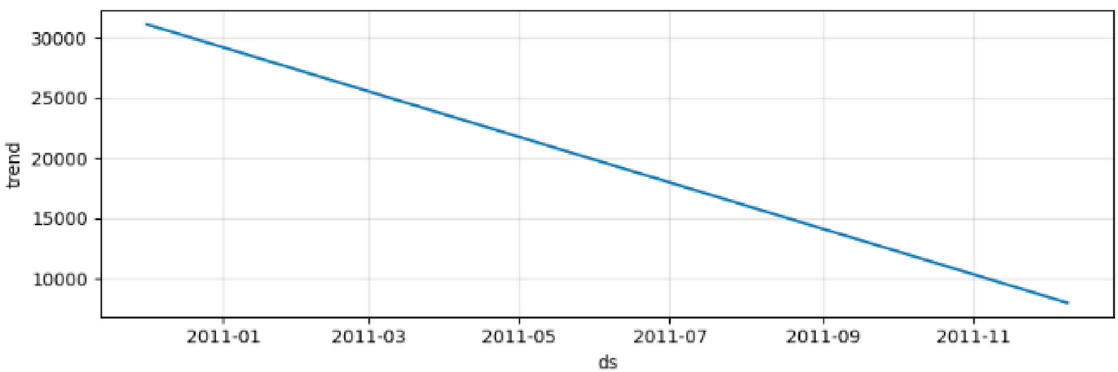

Figure 8: Prophet Forecast Components (Trend).

Figure 8 decomposes the trend element of the Prophet model and it indicates the trend of the sales which is smooth upward. The model however has difficulty to deal with the volatility and big spikes in sales.

Table 8: Comparison of Demand Forecasting Model Performance on the Test Dataset

Model	RMSE	MAPE (%)
Linear Regression	23,120.33	1.34e+15
Random Forest	23,880.86	5.15e+13
Prophet	31,613.05	7.26e+14

Table 8 gives a comparison between the performance of the demand forecasting models based on the RMSE and MAPE where Linear Regression has the lowest RMSE with an exceptionally high MAPE, which demonstrates that the model has a problem in its accuracy. Prophet is the worst-performing model, with the largest RMSE, and the largest value of MAPE which also gives poor forecasting.

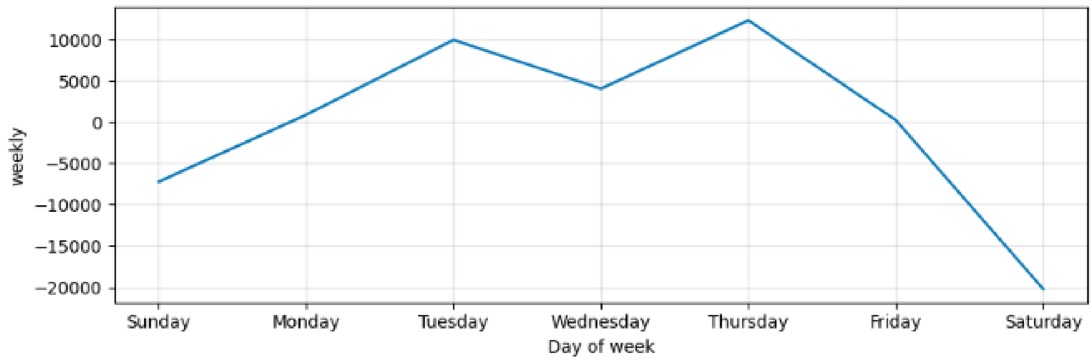

Figure 9: Prophet Forecast Components (Weekly).

Figure 9 displays the decomposition of the Weekly Seasonality component of Prophet model. It is evident that a weekly pattern is followed. Saturday and Sunday sales are also considerably less (negative contribution to sales) and this is determined to be the zero-sales days as in Figure 3. The most sales days include Thursday and Tuesday. This is useful in quantifying the effect of the weekly working schedule of business, which is the primary strength of Prophet model.

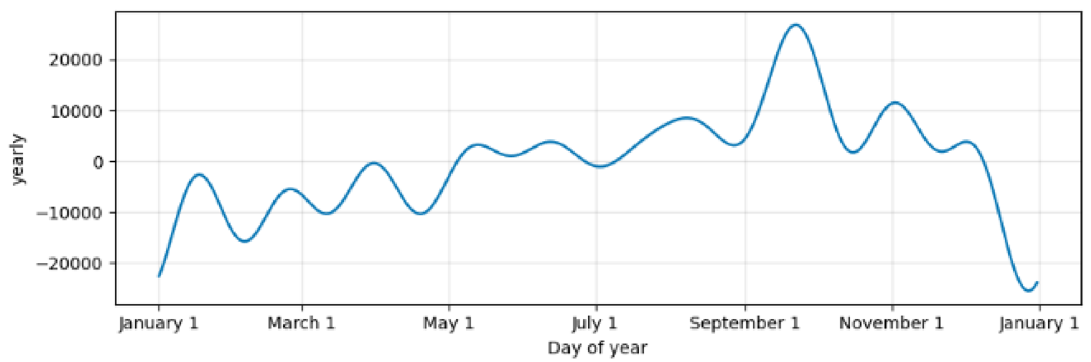

Figure 10: Prophet Forecast Components (Yearly).

Figure 10 depicts the decomposition of the Yearly Seasonality component of Prophet model. September and October will have positive contribution as the highest sales will be made, and by the end of the year, the sales will be low in the months of January and December. This element attempts to smooth out the monthly increase of Figure 2, yet the huge fallacies that will take place in January and December once again indicate Prophet is not good at modeling short volatile information that is concentrated in the end of period.

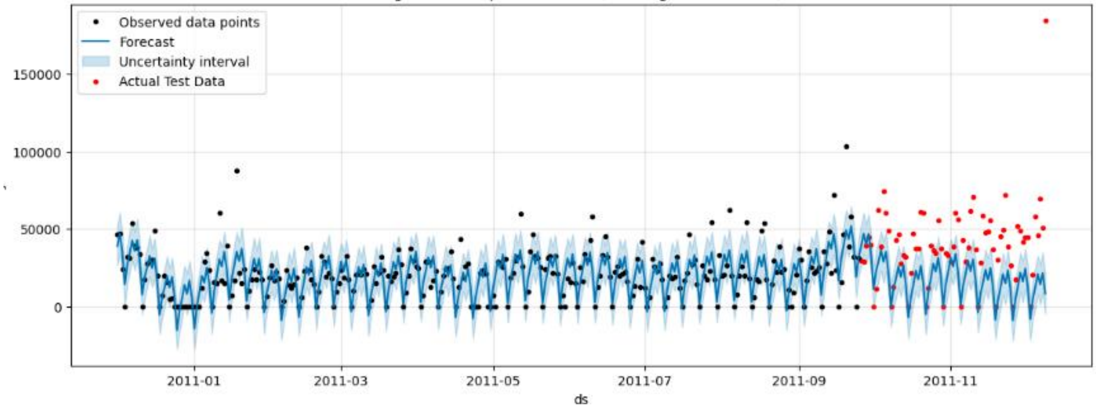

Figure 11: Prophet Forecast (Training + Test Period).

All the Prophet forecast (blue line) and the observed data (black dots) and actual test data (red dots) are presented in Figure 11. The forecast line is an undulated wave which is smooth and damped and can barely capture the huge size of the actual sales (black and red dots) and large spikes. During test, the uncertainty interval (light blue band) is very large that means the model is not that confident with its predictions due to the instability of data. The daily variance is not usually modeled in the model thus adding to its higher RMSE as compared to the feature engineering regression models.

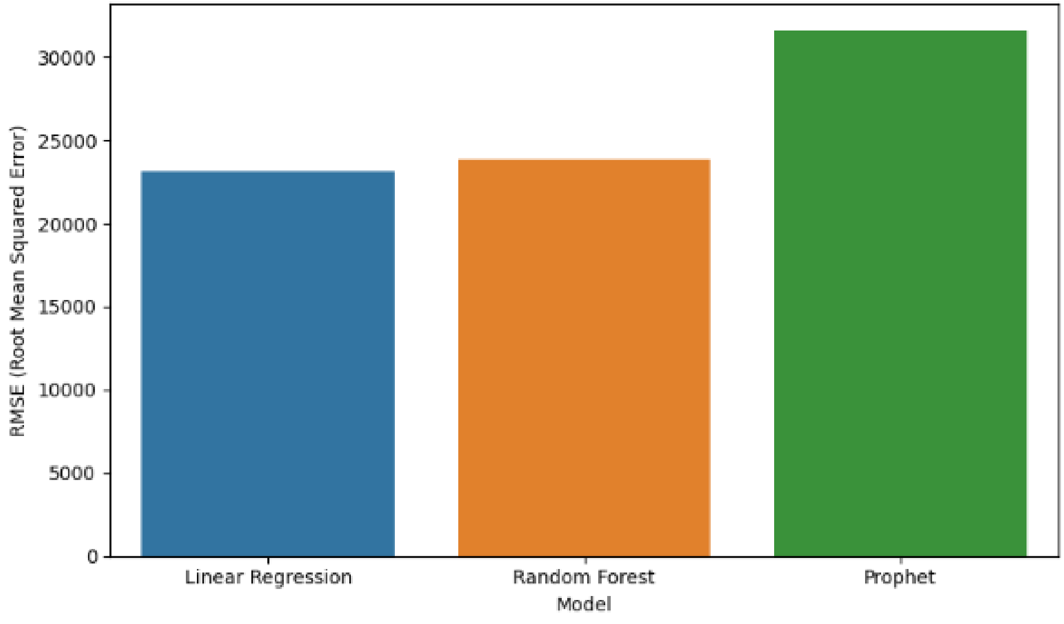

Figure 12: Model Comparison by RMSE.

Figure 12 graphically contrasts the values of RMSE of each of the models and it proves that Linear Regression performs better than both Random Forest and Prophet in making predictions day to day sales.

4.3 Customer Clustering (RFM)

Table 9: Summary Statistics of Recency, Frequency, and Monetary (RFM) Features for Customer Segmentation.

Metric	Recency (Days Ago)	Frequency (Transactions)	Monetary (Total Spent in USD)
Count	4,338.0	4,338.0	4,338.0
Mean	92.06	4.27	2,054.27
Std Dev (σ)	100.01	7.70	8,989.23
Min	0.00	1.00	3.75
25th Percentile (Q1)	17.00	1.00	307.42
50th Percentile (Median)	50.00	2.00	674.49
75th Percentile (Q3)	141.75	5.00	1,661.74
Max	373.00	209.00	280,206.02

Table 9 also provides a summary of the Recency, Frequency and Monetary (RFM) data as a summary to customer segmentation, with an average recency of 92 days, a frequency of 4.27 transactions and spending of 2054. The statistics show that the number of behaviors varies greatly, with certain customers making high and regular purchases, and others being lowly engaged.

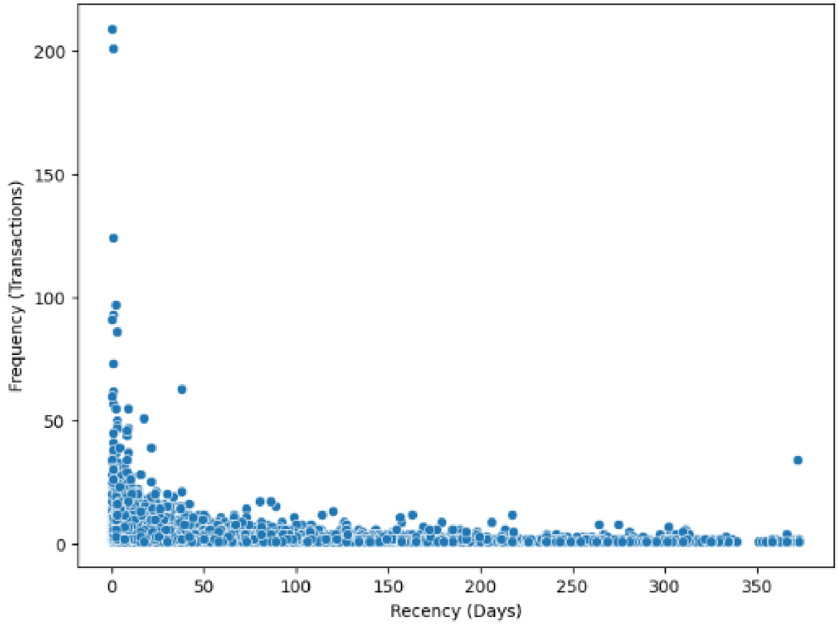

Figure 13: Recency vs. Frequency (Raw Data).

Figure 13 presents a scatter plot of Recency vs. Frequency where it is observed that the majority of customers are not high recency but have high frequency in their purchases. There are also outliers that are very high in frequency or recency, which implies that feature scaling must be done before clustering.

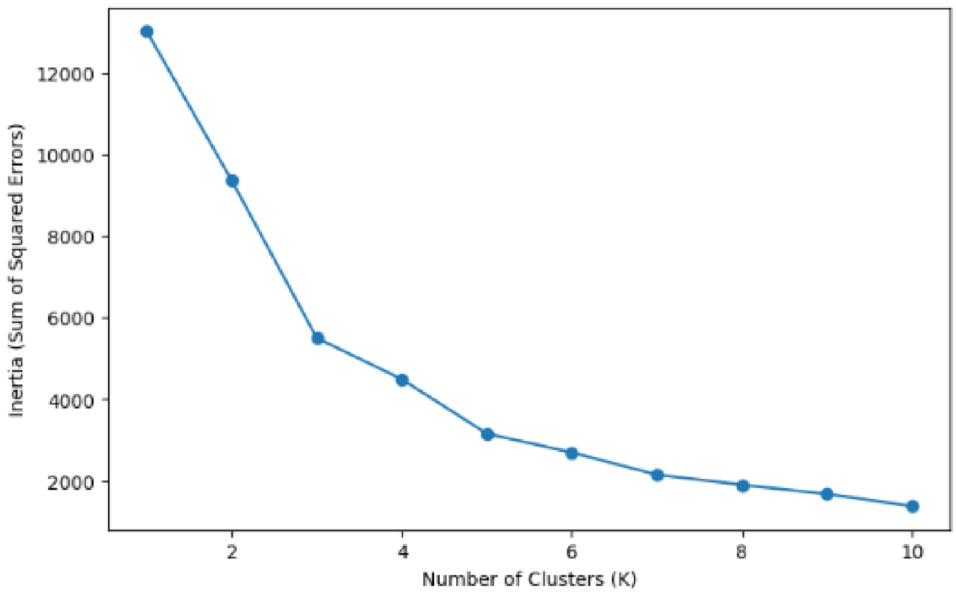

Figure 14: k-Means Elbow Method (Inertia).

The Elbow Method of determining the optimal number of clusters in K-Means is represented in figure 14 to indicate that four clusters give the best balance of within-cluster variance.

Table 10: Distribution of Customers Across the Four K-Means Segments.

Cluster ID	Customer Count	Percentage of Total Customers
0	3,232	74.5%
1	1,091	25.1%
2	8	0.2%
3	7	0.2%

Table 10 indicates that the majority of customers belong to Cluster 0 (74.5%), and Cluster 1 (25.1%) which implies that there exist two large dominant customer groups and very small high-value or niche groups (Cluster 2 and 3, respectively, 0.2 each).

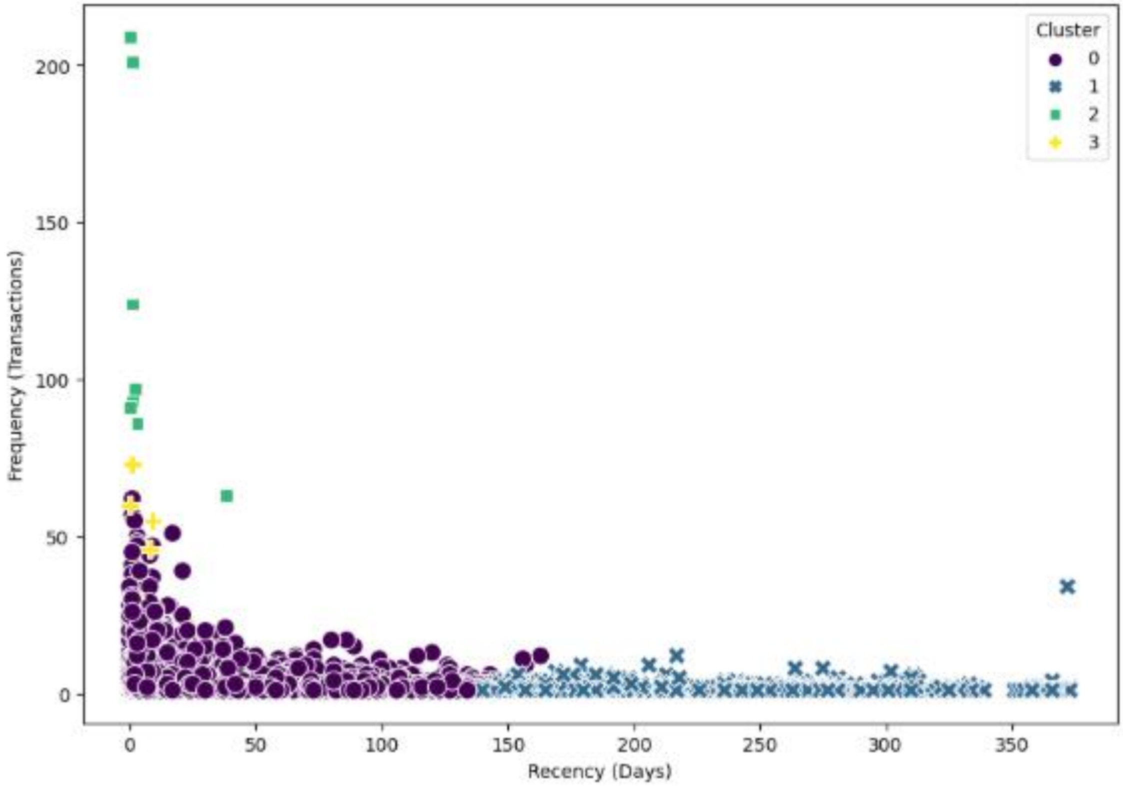

Figure 15: k-Means Clustering (Recency vs. Frequency, k=4).

As presented in Figure 15, the four customer segments identified with the help of K-Means demonstrate that the majority of customers fall into the Core Loyal customer segment (Cluster 0), and the rest are smaller and high-value segments.

Table 11: Mean Recency, Frequency, and Monetary (RFM) Values for Each Customer Cluster.

Cluster ID	Mean Recency (Days Ago)	Mean Frequency (Transactions)	Mean Monetary (Total Spent in USD)
0	40.59	4.81	$2,025.08
1	245.72	1.58	$630.24
2	5.75	120.50	$55,312.69
3	6.29	41.14	$176,606.16

Table 11 indicates the average Recency, Frequency and Monetary (RFM) of each customer cluster. Cluster 0 shows the most frequent purchases (4.81) and spending (2,025.08) and in Cluster 1, the frequency and the spending are less, although the customers are not so new. Cluster 2 and 3 denote high value customers, whereby Cluster 2 has a very high frequency (120.50) and spending amount of money and Cluster 3 has a high amount of spending (176,606.16) although the frequency of its transactions is lower.

4.4 Discussion

The output of demand forecasting models and customer clustering gives insights into the dataset and the patterns that exist in the transactional data. Linear Regression model was more effective in predicting the daily sales, considering that it has the lowest RMSE value when compared with random forest and Prophet. The Linear Regression method with its simple and feature-based approach, which incorporated time-related variables such as Lag_7 and month, allowed it to capture the most important trends. Random Forest, on the other hand, also did the same, however, Prophet had a hard time dealing with the volatility and the high spikes in the data, making the RMSE larger.

To perform customer segmentation, K-Means clustering was effectively used to determine the four unique customer segments, namely Core Loyal Customers, Churned/Lost Customers, High-Frequency Buyers and Revenue Whales. The small number of outliers identified by the DBSCAN algorithm is a strength of the algorithm as it is effective in identifying rare and high-value customers that K-Means could possibly miss.

These conclusions underline the significance of applying both traditional and progressive means in comprehending the behavior of customers and predicting demand. The models in this context provide practical information to focus on individual customer segments and provide personalized strategies, streamline the business processes and enhance the level of future sales forecasting.

5. Conclusion

The study proves the transformational capabilities of AI and ML technologies to streamline important supply chain operations, namely, demand forecasting, inventory, and logistics. The use of advanced models such as LSTM, Prophet and ensemble method presents significant enhancement in the quality and efficiency in forecasting as compared to the traditional methods. Moreover, the Customer-to-Inventory system offers a sound solution of the personalized demand forecasting approach that, when incorporating real-time inventory monitoring, can assist a retailer in saving money and improving the level of service. The findings underscore the need to embrace AI-based approaches to business in order to stay competitive in the current dynamic retailing field. The research can be furthered in the future by assessing the applicability of AI in other aspects of the supply chain, including supplier management and risk mitigation, to have a more comprehensive approach to supply chain optimization.

References

[1] S, J. (2022). Artificial Intelligence (AI) in retailing- A systematic review and Research Agenda. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4134959

[2] Supply Chain Management and retailing: Leigh Sparks. (2010). Managing Football, 168–184. https://doi.org/10.4324/9780080949765-16

[3] Randall, W. S., Gibson, B. J., Clifford Defee, C., & Williams, B. D. (2011). Retail Supply Chain Management: Key priorities and practices. The International Journal of Logistics Management, 22(3), 390–402. https://doi.org/10.1108/09574091111181381

[4] Jodlbauer, H., Brunner, M., Bachmann, N., Tripathi, S., & Thürer, M. (2023). Supply Chain Management: A structured narrative review of current challenges and recommendations for action. Logistics, 7(4), 70. https://doi.org/10.3390/logistics7040070

[5] Balan, G. S., Kumar, V. S., & Raj, S. A. (2025). Machine learning and artificial intelligence methods and applications for post-crisis supply chain resiliency and Recovery. Supply Chain Analytics, 10, 100121. https://doi.org/10.1016/j.sca.2025.100121

[6] Younis, H., Sundarakani, B., & Alsharairi, M. (2021). Applications of artificial intelligence and machine learning within Supply Chains:systematic review and Future Research Directions. Journal of Modelling in Management, 17(3), 916–940. https://doi.org/10.1108/jm2-12-2020-0322

[7] Mallesham, G. (2022a). The role of AI and ML in revolutionizing supply chain management. International Journal of Scientific Research and Management (IJSRM), 10(06), 918–928. https://doi.org/10.18535/ijsrm/v10i6.ec05

[8] Cannas, V. G., Ciano, M. P., Saltalamacchia, M., & Secchi, R. (2023). Artificial Intelligence in Supply Chain and Operations Management: A multiple case study research. International Journal of Production Research, 62(9), 3333–3360. https://doi.org/10.1080/00207543.2023.2232050

[9] Ronchini, A., Guida, M., Moretto, A., & Caniato, F. (2024). The role of Artificial Intelligence in the supply chain finance innovation process. Operations Management Research, 17(4), 1213–1243. https://doi.org/10.1007/s12063-024-00492-2

[10] Samuels, A. (2025). Examining the integration of artificial intelligence in Supply Chain Management from industry 4.0 to 6.0: A systematic literature review. Frontiers in Artificial Intelligence, 7. https://doi.org/10.3389/frai.2024.1477044

[11] Babai, M. Z., Arampatzis, M., Hasni, M., Lolli, F., & Tsadiras, A. (2023). On the use of machine learning in Supply Chain Management: A systematic review. IMA Journal of Management Mathematics, 36(1), 21–49. https://doi.org/10.1093/imaman/dpae029

[12] Maniatis, P. (2025). The role of Artificial Intelligence in Supply Chain Management: A quantitative exploration of its impact on efficiency and performance. International Journal of Clinical Case Reports and Reviews, 22(04), 01–13. https://doi.org/10.31579/2690-4861/671

[13] Baldi, B., Confente, I., Russo, I., & Gaudenzi, B. (2024). Consumer-Centric Supply Chain Management: A literature review, Framework, and Research Agenda. Journal of Business Logistics, 45(4). https://doi.org/10.1111/jbl.12399

[14] Ferreira, B., & Reis, J. (2023). Artificial Intelligence in Supply Chain Management: A systematic literature review and guidelines for future research. Springer Proceedings in Mathematics & Statistics, 339–354. https://doi.org/10.1007/978-3-031-47058-5_27

[15] Atwani, M., Hlyal, M., & Elalami, J. (2022). A review of artificial intelligence applications in Supply Chain. ITM Web of Conferences, 46, 03001. https://doi.org/10.1051/itmconf/20224603001

[16] Gupta, A. (2024, February 6). The role of AI in developing Resilient Supply Chains: Gjia.

Traqo.io. https://blog.traqo.io/the-role-of-ai-in-developing-resilient-supply-chains-gjia/

[17] Douaioui, K., Oucheikh, R., Benmoussa, O., & Mabrouki, C. (2024). Machine learning and deep learning models for demand forecasting in Supply Chain Management: A critical review. *Applied System Innovation*, *7*(5), 93. https://doi.org/10.3390/asi7050093

[18] Chowdhury, A. R., Paul, R., & Rozony, F. Z. (2025). A systematic review of demand forecasting models for retail e-commerce enhancing accuracy in inventory and Delivery Planning. *International Journal of Scientific Interdisciplinary Research*, *06*(01), 01–27. https://doi.org/10.63125/mbbfw637

[19] Nasseri, M., Falatouri, T., Brandtner, P., & Darbanian, F. (2023). Applying machine learning in retail demand prediction—a comparison of tree-based ensembles and long short-term memory-based deep learning. *Applied Sciences*, *13*(19), 11112. https://doi.org/10.3390/app131911112

[20] Giri, C., & Chen, Y. (2022). Deep learning for demand forecasting in the fashion and apparel retail industry. *Forecasting*, *4*(2), 565-581.

[21] Mustapha, O. O., & Sithole, T. (2025). Forecasting Retail Sales using Machine Learning Models. *American Journal of Statistics and Actuarial Sciences*, *6*(1), 35-67.

[22] Amosu, O. R., Kumar, P., Ogunsuji, Y. M., Oni, S., & Faworaja, O. (2024). AI-driven demand forecasting: Enhancing inventory management and customer satisfaction. *World Journal of Advanced Research and Reviews*, *23*(2), 100-110.

[23] Ramachandran, K. K., Karthick, K. K., Kalyan, N. B., Tiwari, M., Raju, G. S., & Ganesh, K. (2024). Improving Retail Industry Inventory Management Using Machine Learning. In *Recent Technological Advances in Engineering and Management* (pp. 137-141). CRC Press.

[24] Bhavikatta, N. B. (2025). AI-Driven Inventory Optimization in Supply Chains: A Comprehensive Review on Reducing Stockouts and Mitigating Overstock Risks. *Journal of Computer Science and Technology Studies*, *7*(7), 01-13.

[25] Vaka, D. K. (2024). From complexity to simplicity: Ai's route optimization in Supply Chain Management. *Journal of Artificial Intelligence, Machine Learning and Data Science*, *2*(1), 386–389. https://doi.org/10.51219/jaimld/dilip-kumar-vaka/100

[26] krishna Vaddy, R. (2023). Ai and ml for transportation route optimization. *International Transactions in Machine Learning*, *5*(5), 1-19.

[27] Teixeira, J. V. F., & Ramos, A. L. (2025). Intelligent supply chain management: A systematic literature review on artificial intelligence in supply chain management. *Information*, 16(5), Article 399. https://doi.org/10.3390/info16050399

[28] Manasa, R., & Jayanthila Devi, A. (2022). Amazon's Artificial Intelligence in Retail Novelty-Case Study. *International Journal of Case Studies in Business, IT, and Education*, *6*(2), 787-804.