DSCI 5260 Business Process Analytics Group 3

Final Project Report

Optimizing Inventory and Fulfilment Operations: Analyzing Warehouse Efficiency, Customer Satisfaction and Financial Performance

Rohan Mamidi
Kavya Venkatesh
Lakshmi Ravi Chandra Talluri
Yeswanth Kiran Kumar Godavarthi

Table of Contents:

List of Figures:

List of Tables:

Acknowledgement

We extend our deepest gratitude to Dr. Maryam Khatami, whose expert guidance and unwavering support were instrumental throughout this project. Your insights and encouragement have been invaluable to our team.

We are also thankful to the faculty and staff of the G. Brint Ryan College of Business at The University of North Texas for their assistance and for providing the necessary resources to successfully complete this project.

Our sincere appreciation goes to our peers and colleagues for their constructive feedback and camaraderie during this academic journey. Their perspectives have significantly enriched our work.
Special thanks to our families and friends for their patience, understanding, and support. Your belief in our abilities has been a constant source of motivation and strength.

Lastly, we acknowledge each member of our team for their dedication and hard work. This project was not only a test of our academic knowledge but also of our ability to collaborate effectively. Each team member brought unique skills and perspectives, making our collective effort truly rewarding.

Group 3

Abstract

Inventory Management is the process of maintaining goods in systematic manner by exploring the optimization of inventory and fulfillment operations, accelerating the shift from traditional to automated warehousing to advance and improve in various aspects like efficiency of the organization, customer satisfaction and financial performance. Inventory methods that are used in traditional ways are rely more on manual processes, which may lead to error-prone stock handling and inadequate handling of real-time data because of the constraints around them. Whereas these automated warehousing systems indulge in advanced technologies like the Internet of Things and analytical techniques that deliver substantial enhancements in dealing with real-time data with more accuracy and precision, more efficiency in utilizing space and enhanced capabilities of decision-making. This study assesses the importance of inventory management and its impact on the satisfaction of customers, illustrating that in availability of products and timely delivery, which are the factors that directly influence customer satisfaction and loyalty, are very crucial in inventory management. We analyze the relationship between financial performance and inventory management practices, identifying a beneficial correlation between advanced inventory techniques and the stronghold financial outcomes with the help of statistical techniques and analysis, using the knowledge of predictive analytics and performing models like ARIMA and Implementation of machine learning techniques like logistic regression, decision tree and random forest methods for enhancing the optimization of stock levels and anticipate the customer behaviors. Finally, these all factors combine to enhance the financial performance and give us optimized and efficient warehouse management.

1. Introduction

Inventory is the main component and essentials that are needed for every business. Without an inventory, it's difficult for an organization to commence. It helps in the determination of the company's financial health, operational efficiency and plays a key part in helping the business to run smoothly. It is considered an asset for the company, reflects the turnover of the company, and ensures the business operations go seamlessly. The efficiency of inventory makes the business more sustainable, and the accuracy of the inventory makes the customer happy. Considering the volatile nature of the market, proper inventory management is key in supply chain optimization.

Inventory management is the most common way of requesting, putting away, utilizing, and selling the organization's items. It combines the processing of items alongside the warehousing and management of raw materials, components, and finished products. Having an inventory management system is good for companies as these systems help identify waste, turnover and theft [1]. To achieve effective inventory management, good control over the inventory is needed. Management should ensure that the stock levels are accurate, no unnecessary money is invested in inventory, the warehouse capacity is both economical and efficient and the goods are properly preserved. To achieve this, a focus on the correct evaluation, identification, classification, quantification, retrieval, and security of goods would provide a clear and accurate view [2].

Traditionally, inventory management means manual tracking, physical counts, paper record-keeping, manual data entry, processing of manual orders, and reaction for the foreseeable future [3]. Frequently facing challenges are overstocking or stockouts because of a lack of vision for the future. Therefore, traditional methods of inventory management affect warehouses by reducing their operational efficiency through improper maintenance of stock levels. Delay in product delivery and inability to meet customer demands on time leads to customer dissatisfaction and decreased profit.

To increase the performance of the warehouse, it is necessary to identify Key Performance Indicators (KPIs). Each warehouse will have different KPIs. Identifying the correct KPIs that match the business needs will help managers to take corrective actions efficiently and effectively. Popular inventory KPIs include inventory accuracy, shrinkage, carrying cost of inventory, inventory turnover, and inventory-to-sales ratio [4]. The explanation for KPIs is as follows, Inventory turnover which measures how efficiently inventory is managed by calculating the number of items sold and replaced over a period. Inventory accuracy explains how accurately inventory records match the actual physical inventory; higher accuracy is better. Inventory carrying explains the costs associated with storing and maintaining inventory levels, lower costs are best. Customer satisfaction exhibits customer happiness with order fulfillment and availability of products, higher satisfaction is worth. Inventory-to-sales ratio compares inventory levels to sales to assess if the right amount of inventory is held, a lower ratio is generally better.

Investigating the relationship between inventory management policies and financial performance, we can expect to see that these factors indirectly affect each other. Financial performance is the first indirect measure of warehouse performance, and concerns how a firm uses its resources to achieve its highest productivity without any issues and meet the expectations of stakeholders and customers [5]. The coalition of these factors (warehouse efficiency, customer satisfaction, financial performance) takes inventory control to guarantee materials and availability of items, making sure the customer has access to the best services and achieves a competitive advantage. Problems that arise if the inventory is not properly monitored include inconsistency and high costs [6]. It is crucial for the people who have authority over the inventory in the warehouse to constantly be objective of satisfying the customer and to keep inventory costs at a minimal level [7].

The efficiency of inventory and fulfillment operations determines the rise and fall of a company, and it's not just merely a logistical concern. The warehouse emerges as a crucial nexus that balances the adjoining performances of supply chain efficiency, customer satisfaction, and financial performance. Effective inventory management is a critical component of operational costs with direct consequences for profitability. Moreover, it molds customers' experience where we constantly meet their expectations, thus creating a loyal customer base and good customer lifetime value.

Research Questions:

This research will apply statistical analysis and predictive modeling to identify patterns, discern trends, and propose optimization strategies to analyze a comprehensive warehouse inventory dataset and its fundamental questions including-

- What stock levels strike the perfect balance between availability and cost-effectiveness for various product categories?
- How do we quantify the influence of seasonal trends and external shocks on inventory requirements?
- Can predictive analytics tools be harnessed to anticipate future stock needs more accurately?
- How do packaging and delivery ratings influence customer satisfaction and, by extension, the propensity for repeat purchases?
- What are the correlations between the variables?

2. Literature Review

This research brings warehouse-operated inventory stocks and inventory management involving findings, experiences, and theories. The important topics under consideration are inventory maintenance, stock handling, customer satisfaction and its challenges, warehouse efficiency, financial performance, and automated warehousing.

Inventory management is an action that balances product availability in the warehouse, optimizing cost and enhancing customer service. This balance is pivotal for maintaining the operational and financial health of an organization. Inventory management aims to satisfy the customers with product quantity and quality. It helps to predict current and future stocks for all inventory types by eliminating overstocking, product delays and minimizing the costs by optimizing it to maintain efficiency. This optimization is crucial for internal operations, facilitating smoother workflows and reducing complexities associated with stock-level management [8].

Ancient inventory management relies on manual processes, and paper documentation represents several challenges in the modern company environment. These difficulties are encountered as there are limitations of hand-picking processes that are prone to human mistakes that can affect inventory tracking on a large scale. These mistakes have an impact on the stock levels of warehouses, which can cause damage to the company. Conventional inventory techniques are flat and uphill by nature as these are significantly dependent on manual processes that will affect productivity and critical decision-making [9]. As companies grow, the effectiveness of traditional approaches becomes complicated. As operations grow, simple manual tasks become difficult, impacting the efficiency and effectiveness of inventory management as they become huge in scale [10]. From this we get to know that traditional inventory management systems suffer from the limitation of not having access to real-time data. It prevents a business from responding quickly to changes in market demand. It makes it hard to seize sales opportunities or effectively manage inventory levels, consequently, resulting in surplus or shortages.

Additionally, although traditional inventory management techniques may seem economical, especially for small businesses, they tend to increase operational costs over time, leading to increased operating expenses. These costs arise from the inefficiencies inherent in manual processes, including the labor involved in stock counts and data entry, the potential for stock discrepancies and missed opportunities for optimization. The thorough perspective of these challenges highlights the requirement for businesses to progress beyond conventional inventory handling techniques [10]. Therefore, adopting technological innovations provides a path to overpowering these difficulties, improving the inventory precision, Operational effectiveness, and adaptability of the business.

To enhance warehouse efficiency having barcode scanning, RFID (Radio frequency identification) tags, and cloud-native systems is a must in every warehouse. A case study highlighting the successful transition of a client from traditional Excel spreadsheet-based inventory management to a cloud-native WMS (Warehouse Management Software) [11]. The client experienced benefits such as real-time inventory tracking, streamlined operations, and improved employee morale. It also stressed upon providing training to staff on recently introduced technologies, conducting regular audits of inventory is a key success and every organization having these transformative tools will greatly enhance the accuracy and efficiency of inventory management. To obtain actionable insights, to predict inventory accuracy and to make data collection and analysis easy, the shift from traditional warehousing to automated warehousing in the field of inventory management is needed, that is the adoption of smart warehouse technology, is an area that provide many advantages compared to traditional warehousing, including the provision of real-time information and instant update as well as effective space utilization [12]. However, to adopt smart warehouse technology, businesses need to clearly understand the current state of their inventory which is what our research is providing.

Inventory management has also been shown to impact on firms' financial performance. The role of inventory management on the financial performance in some selected manufacturing companies in Mogadishu [13]. They explained that manufacturing companies use inventory management and examined the relationship between inventory management and financial performance and found a positive relationship with a result of $r = 0.683$. The effects of inventory management on the financial performance of firms funded by government venture capital in Kenya [14]. They found a positive correlation between these two variables, with a Pearson correlation coefficient of 0.759 and a r^2 of 0.577. By conducting an in-depth analysis of inventory in our research, we are confident that this will help businesses better manage their inventory which will, in turn, result in better financial performance.

A study examines the role of inventory management in improving customer satisfaction [15]. The study highlights that retailers are increasingly implementing "perfect order" systems to fulfill customer needs. This system emphasizes complete and on-time deliveries, which are key factors in achieving customer satisfaction. The study used a qualitative customer satisfaction performance survey to assess the relationship. However, a limitation of this approach was the low participation rate. Despite this, the results indicated that effective inventory management systems helped small businesses improve their customer service. Hence, to know the impact of inventory management on customer satisfaction we will be using classification models.

From a business point of view to be competitive and financially strong, allocating the budgets and investing time in modern-day technologies, techniques and practices have utmost importance besides their benefits. Adaptation to these could be so beneficial in business objectives like customer satisfaction and financial stability with their technologies because they not only streamline but also give us strategic planning and approaches. Finally, to achieve the ideal balance between product availability, cost-effectiveness, and customer satisfaction, ultimately leading to better financial performance, optimizing inventory and fulfillment operations is crucial. Our research aims to contribute to this goal by leveraging statistical analysis, predictive modeling techniques and utilizing ARIMA models and logistic regression, respectively.

3. Data

3.1. Data Description

The dataset is a real time dataset obtained from ‘Texas Boys Wholesale’ from its Zoho inventory from 2015- 2024 was chosen for this research and includes extensive stock and order data from operational sub warehouses activities. It contains a broad range of factors that are essential for examining metrics related to customer satisfaction, order processing effectiveness and inventory management.

This dataset is a collection of transactional and operational data, illustrating the everyday struggles in warehouse management and making it relevant to our project subject. We can know how well warehouses are managing supply and demand by seeing patterns in stock turnover and identifying problems with overstocking or stockouts by comparing inventory levels to order numbers. The incorporation of customer satisfaction ratings establishes a clear connection between operational effectiveness and customer experience, enabling us to investigate how variables like order fulfillment durations and product accessibility impact consumer attitudes and buying habits.

A comprehensive picture of warehouse operations is also provided by the variables in the dataset about order priority, buying channels (online vs offline) and warehouse locations. We will be able to comprehend the challenges of managing various product lines across various storage facilities and distribution networks by analyzing these factors. Additionally, it will enable us to investigate how logistical choices affect customer service and operational effectiveness.

Dataset size: 28179 rows and 26 columns

3.2. Columns Description:

Table (1) Data - Columns Description

Variables	Description of the variables
Item ID	The unique number that identifies the item in the warehouse which is used to track and locate products.
List of Items	The description of the item for pulling the product differentiated by flavors, grams etc. Here we cannot disclose the item names and hence renamed as products 1, 2, 3 and others.
SKU	The universal way to find a product that is scanned while selling or purchasing.
Warehouse Name	There are 4 sub warehouses which each have different items to sell according to the packaging, time limit and restriction issues.
Item Type	The product is categorized by item names like drinks, chips, clothes etc.
Purchase Invoice Channel	The invoice channel for a particular item is for offline/online purchasing and taking orders for products.

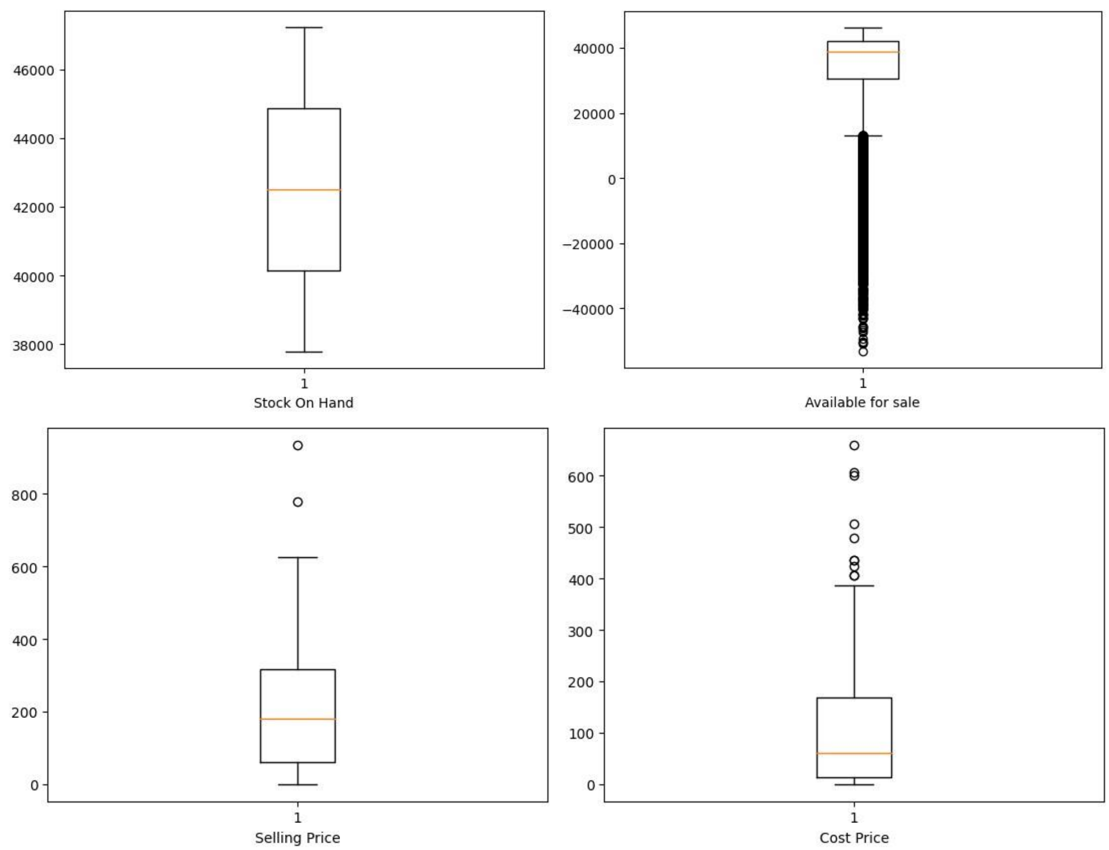
Order Priority	The priority of purchase orders and invoice orders in a sequential manner according to urgency requirements for refilling the warehouse and fulfillment of taken orders.
Order Date	The date on which the product was ordered manually and automatically by following the economic order quantity of the company offline and online.
Order ID	The unique code like invoice number which is useful in identification and for future reference.
Qty Ordered	The quantity that is ordered to manufacturers manually and automatically refilling of stocks according to the requirement in the entire warehouse for the fulfillment of present and future orders.
Qty IN	The quantity that was received against the quantity ordered which is crucial for updating stock levels and monitoring supply.
Qty OUT	The stock that is sold to the customer product wise.
Stock on Hand	The stock that is remaining in the warehouse to be sold for future orders.
Committed Stock	The stock that has been allocated to orders may be in available quantity or non-availability.
Available for Sale	The quantity of items that are available for selling after unsold and available for customers to purchase.
Status	The presence of an item in the market whether it is active or inactive stage.
Unit	The measurement scale of an item such as a piece or box.
Selling Price	The price at which the product is sold to the customer.
Cost Price	The price to purchase the product.
Created By	The employee who created the product in the system for entry or record in the inventory.
Total Money on Hold	The total money that is on hold is calculated based on unsold products in the warehouse.
Item Availability Rating	The rating given by the customer for the availability or presence of a variety of products in the warehouse according to his requirements.
Offline Rating	The rating given by the customer based on the physical shopping method and order fulfillment.
Online Booking Rating	The rating given by the customers based on online shopping for availability of products and easiness of shopping.
Packaging Rating	The rating given by the customer for packing items that are packed for his order like glasses.
Delivery Rating	The rating given by the customer based on the delivery of products to the customer.

3.3. Data Handling:

- Removed non relevant columns.
- Identified outliers and kept it like that only without solving it because we needed those outliers for future demand forecasting.
- Imputation is done with median for all the rating columns.
- Changed the future orders date to March 10th, 2024.

3.4. Outliers:

For efficient and effective warehouse operations, it is important to investigate whether the outliers are occurring due to exceptional cases like data entry errors, stock control problems, price issues and irregularities in sale patterns.



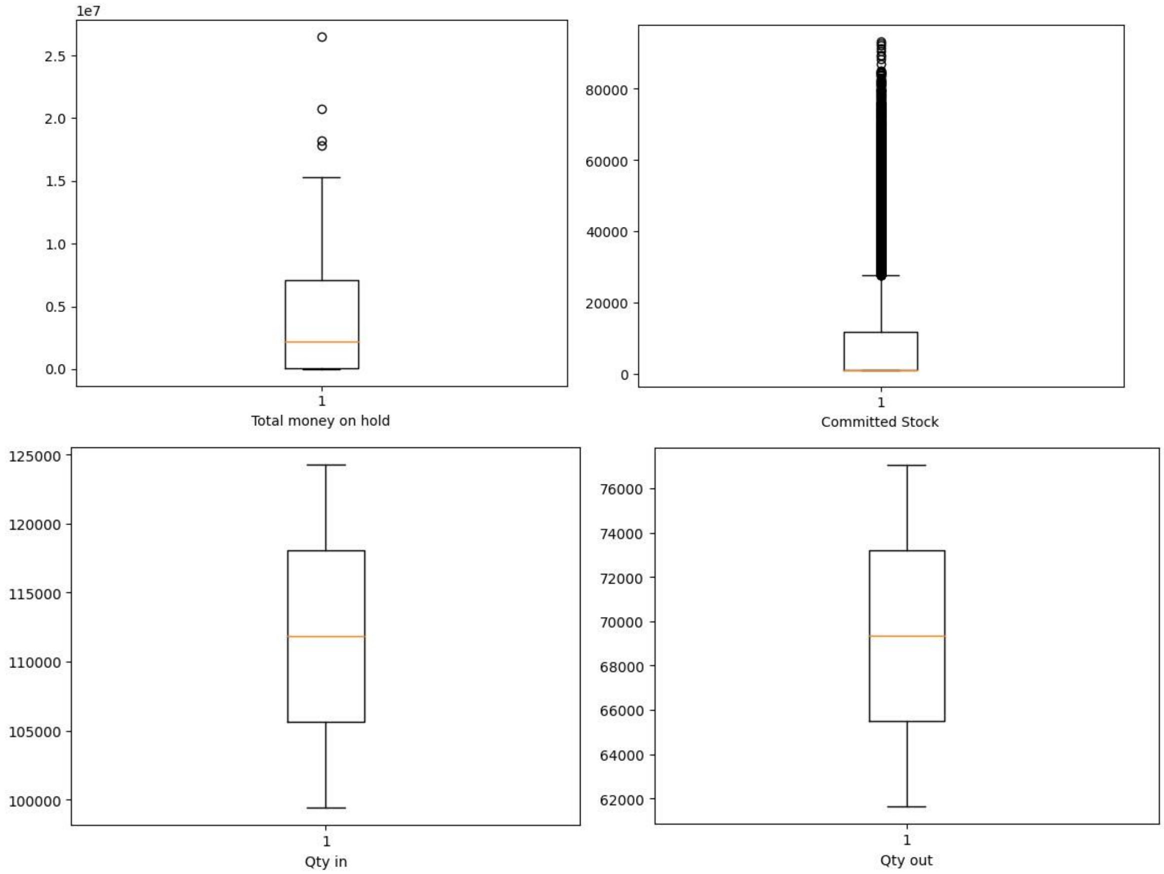


Figure (1) Outliers for Operational and Financial Variables

The cost price, total money on hold and selling price have the presence of outliers higher than the upper whisker indicating higher cost price for some items, unsold items lead to holding money for the company, the quantity of items moving out is higher than usual and some items are sold at a significantly higher price point than most of the others.

The Available for sale variable outlier is below the lower whisker indicating the products are shortage or stock shortage. Stock on hand and plot of quantity-in has a symmetric distribution of data with no visible outliers. On the other hand, the committed stock variable has outliers on the upper whisker, indicating that some items have a higher level of committed stock compared to the rest. This implies that certain products have more commitments, such as pending orders or reservations, which could affect their availability for sale.

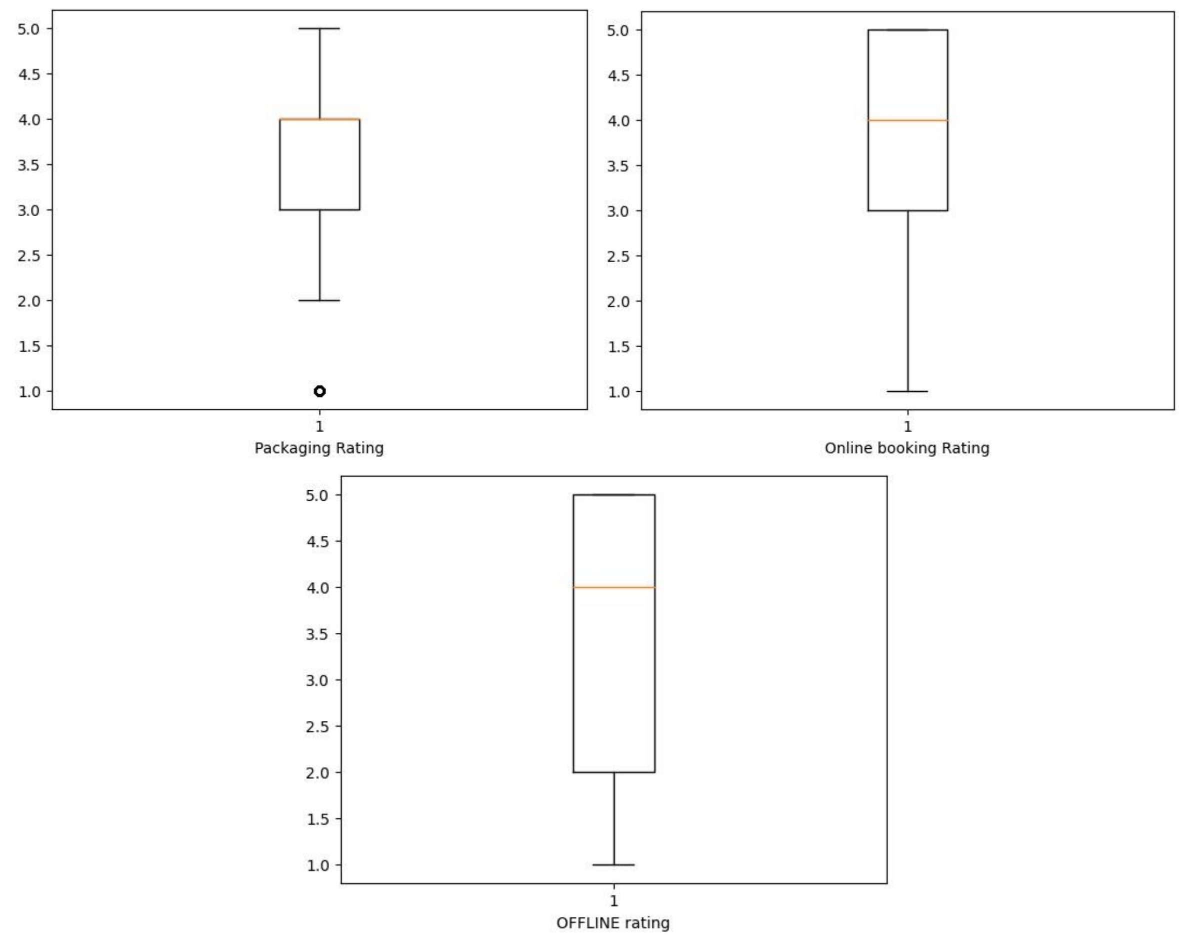


Figure (2) Outliers for Ratings

For customer satisfaction Figure ratings, there are no outliers in offline rating, online rating, and packaging rating except one outlier depicting consistent satisfaction among customers. The purchases made by customers online and offline are most satisfied by the company's streamlined operations and customer service provided.

3.5. Descriptive Statistics:

Table (2) Descriptive Statistics

	Order ID	Qty Ordered	Qty in	Qty out	Stock On Hand	Committed Stock	Available for sale
count	2.82E+04	28179	28179	28179	28179	28179	28179
mean	5.52E+07	134988.1257	111851.0686	69347.66156	42503.40658	9044.607474	33458.79911
std	2.60E+07	8663.006096	7178.160407	4450.45724	2727.699111	14040.33136	14319.41599
min	1.00E+07	120002	99434	61649	37785	1000	-53144
25%	3.26E+07	127456	105610	65478	40132	1000	30590.5
50%	5.55E+07	134989	111852	69348	42504	1000	38680
75%	7.79E+07	142476	118055.5	73194.5	44861	11690	42209.5
max	1.00E+08	149999	124289	77059	47230	93295	46230

	Selling Price	Cost Price	Total money on hold	Item Availability rating	OFFLINE rating	Online booking Rating	Packaging Rating	Delivery Rating
count	28179	28179	2.82E+04	28172	28179	28179	28179	23941
mean	193.387409	95.442977	3.81E+06	2.995882	3.506476	3.491927	3.549487	3.496846
std	141.303082	91.233916	4.05E+06	1.487104	1.355786	1.295489	1.221622	1.294404
min	0	0	0.00E+00	0	1	1	1	0
25%	61.45	13	6.49E+04	2	2	3	3	3
50%	179.76	60.13	2.19E+06	3	4	4	4	4
75%	317.69	168.17	7.11E+06	4	5	5	4	5
max	935	660	2.65E+07	5	5	5	5	5

From the above table we get to know the statistical analysis of variables (mean, min, max, standard deviation, count, and so on) to give a basic overview of the dataset and it gives numerical insights into the data.

3.6. Exploratory Data Analysis

We have used different kinds of visualization techniques to get more insights visually that have more impact and could reveal more interesting patterns, insights and more viable knowledge on the data that could be more helpful to research and give us more focus on the areas where we concentrate. We used bar plots, pie charts, heat maps and line plots for the visual representations.

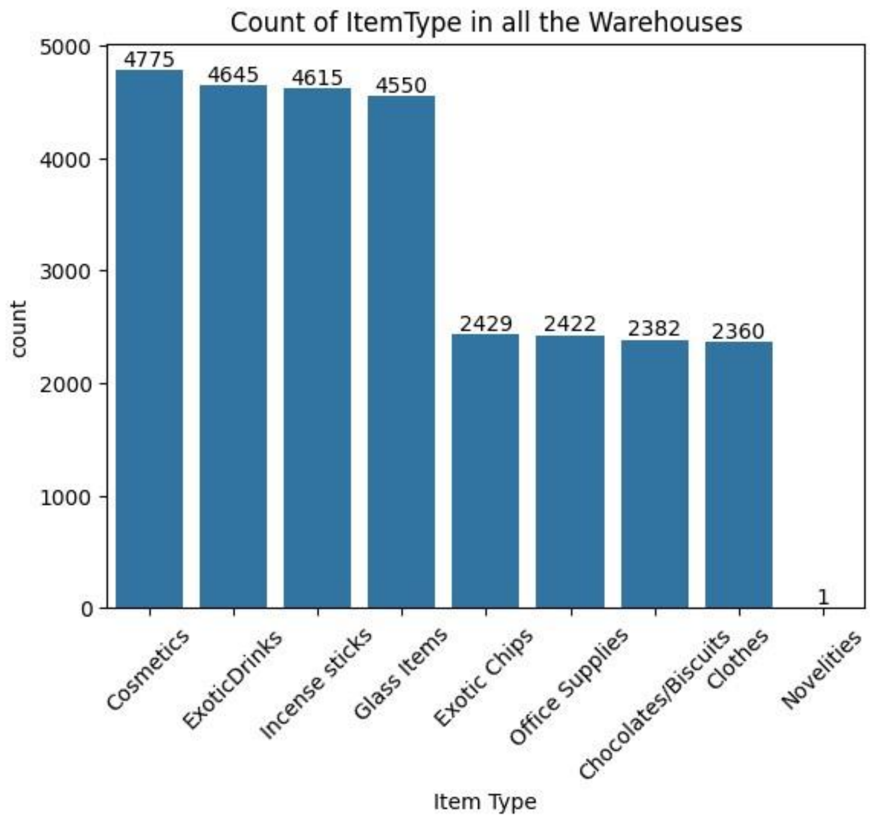


Figure (3) Count of item type in all the warehouses

From the figure in operational conditions, it's apparent that the warehouse holds a greater quantity of cosmetics compared to other items. This suggests that cosmetics are one of the most abundant products in the inventory. Following cosmetics, Exotic Chips, Office Supplies, and Chocolates/Biscuits are within a similar range of availability. These items seem to have comparable quantities in stock.

On the other hand, novelties appear to be the least stocked item in the warehouse, indicating that they have the lowest availability among the nine types of items.

In summary, the bar graph provides insights into the distribution of items in the warehouse, showing that cosmetics dominate the inventory while novelties have the lowest representation. This understanding can be crucial for inventory management and resource allocation within the warehouse.

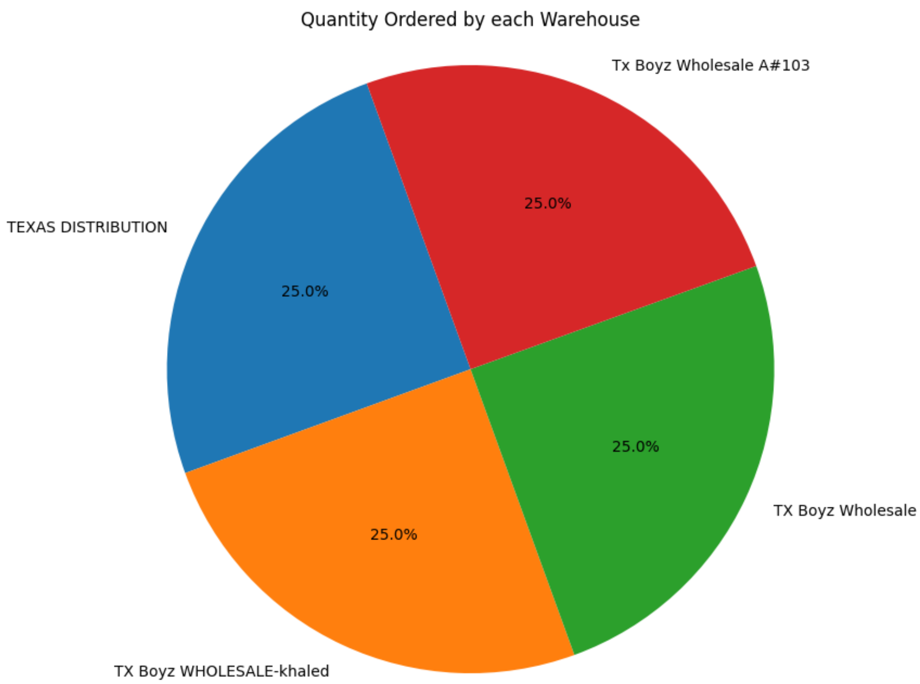


Figure (4) Quantity ordered by each Warehouse.

From the figure we can observe that every order is equally shared between the warehouses, which maintains balance in inventory management.

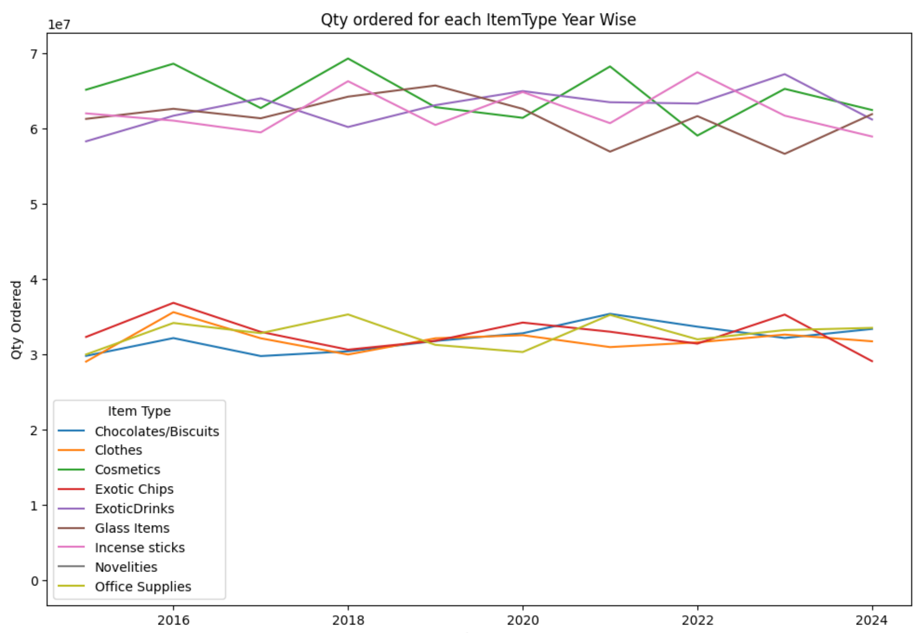


Figure (5) Quantity ordered for each Item type by year.

The above line graph figure shows quantity ordered for each Item type yearly. Cosmetics, incense sticks, exotic drinks, and glass items are the most selling item types, and they re-ordered those items. All the categories show some variations in the quantities year to year.

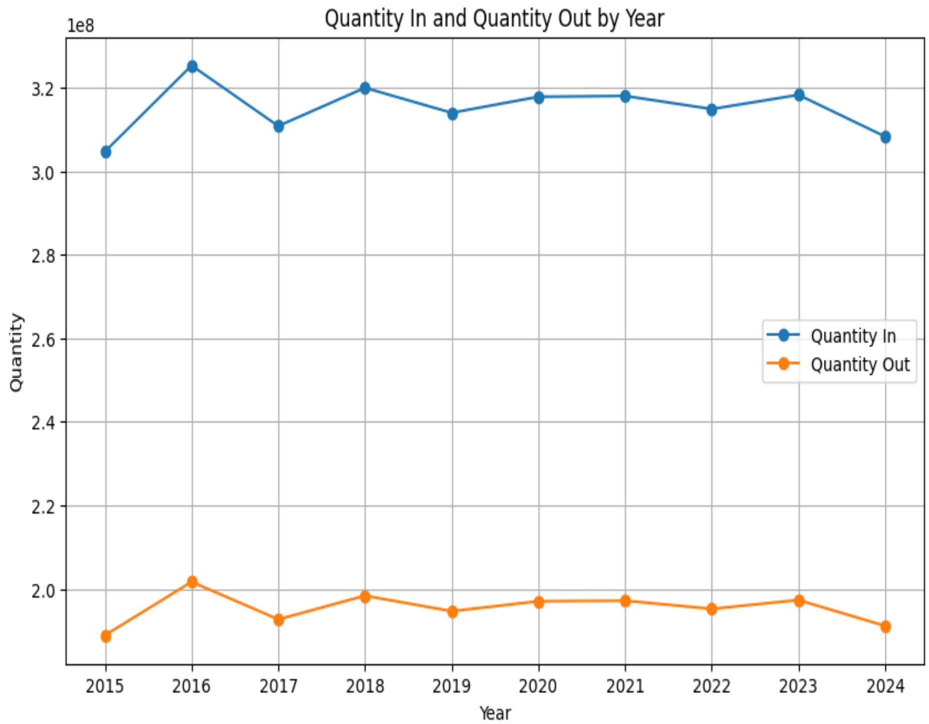


Figure (6) Quantity In and Quantity Out by Year.

Figure depicts that the quantity coming into inventory is always greater than the quantity going out of the inventory. This could lead to miscarriage of the expired products. It is almost 1.6 times the difference between the quantity in and quantity out. Our most intake and outgoing products were from inventory in 2016. But constantly we can see that whenever we take a certain amount of quantity approximately the same amount of the quantity of the products are going out. We can see that quantity in, and quantity out are directly dependent on each other. The least goes to the year 2015. And again, there is dip in the year 2024.

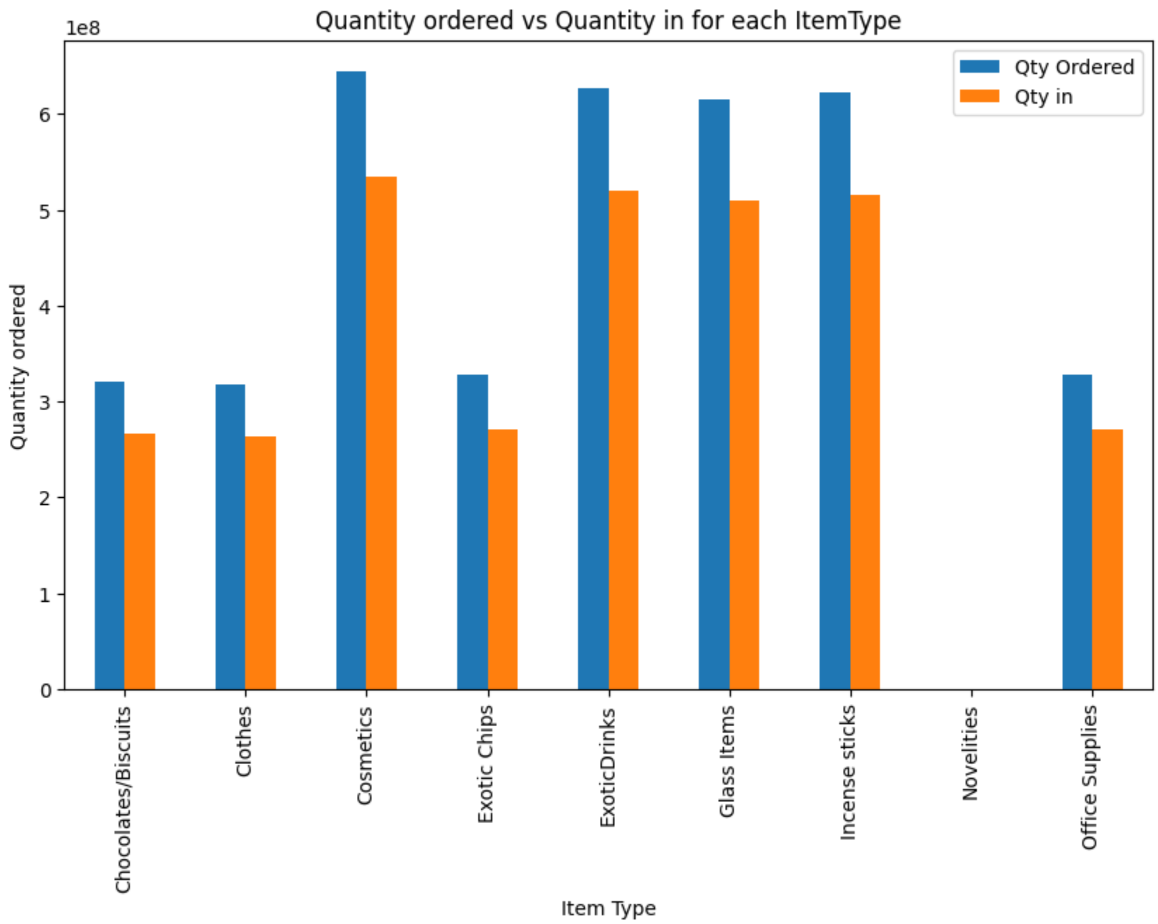


Figure (7) Quantity ordered vs Quantity In for each Item Type

After grouping each item type based on quantity ordered and quantity in, we observed varying quantities for each item. In the dimensions mentioned above Figure e^8 = 10 ^8 (Chocolates/Biscuits sum of quantity ordered in is nearly 3.2e^8= 321 million and so on for other item types).

We have a total of 9 different items. The blue color indicates the quantity ordered, and the orange color indicates the quantity in. We can observe that for every product the quantity coming to the inventory is less than the quantity ordered. This indicated there is some gap between the fulfillments of the requirements. This may sometimes lead to the failure to reach the requirements of the customer, and this affects customer satisfaction. This also suggests that the warehouses are not utilized to their complete potential due to the bridge gap between the quantity ordered and quantity coming to the inventory.

Coming to the products, the negligible item ordered is novelties and the highest ordered is cosmetics following up that position the exotic drinks and so on. Cosmetics, exotic drinks, glass items, and incense sticks fall under the high-ordered category. Chocolates/biscuits, clothes, exotic chips, and office supplies come under the low-ordered category due to their huge difference in the margin of ordering.

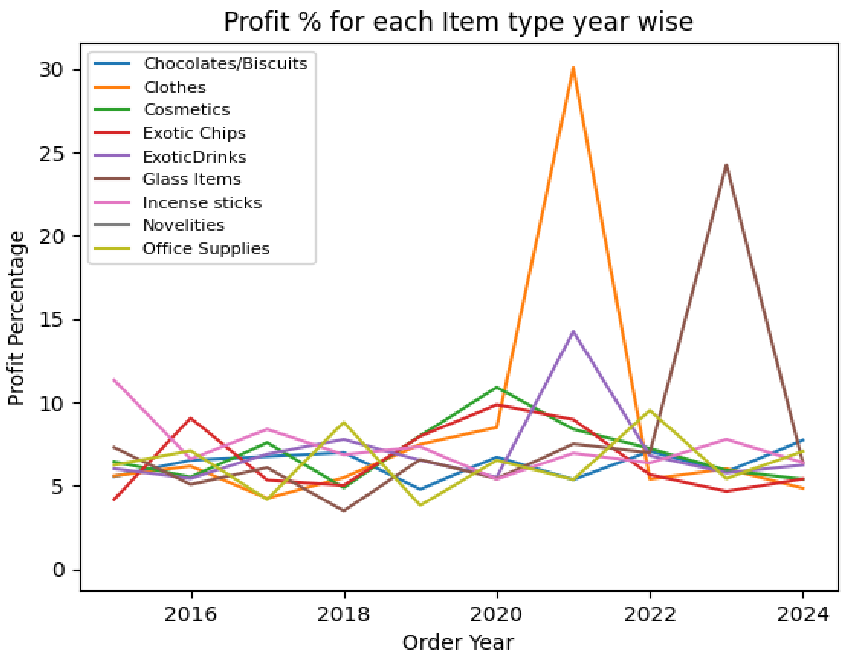


Figure (8) Profit % for each item type by year.

Figure shows the profit percentage for each Item year wise. Item types were sold every year, and each item type was facing ups and downs. In the initial year incense stick was the major profit contribution item. The clothes initially were getting lower profit% but in the year 2021, it made the highest profit %, contributing 30%, and declined in the upcoming year, from the general analysis, the US clothing got rise in 2021. This rise was due to the demand recovery after COVID - 19 pandemic. Glass items that made the least profit in the year 2018 drastically improved their profit by 2023 approximately contributing 25%. The item novelties never contributed to the profit % because there was only 1 item sold.

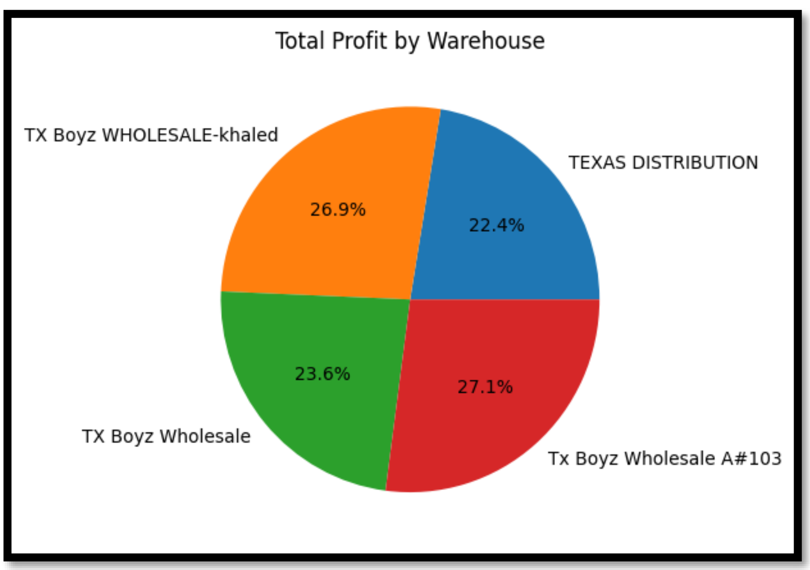


Figure (9) Total profit earned by each warehouse.

The above pie chart Figure shows the profit earned by each warehouse. Tx Boyz Wholesale A#103 is the most profitable and TEXAS DISTRIBUTION is the least profitable among the 4 warehouses.

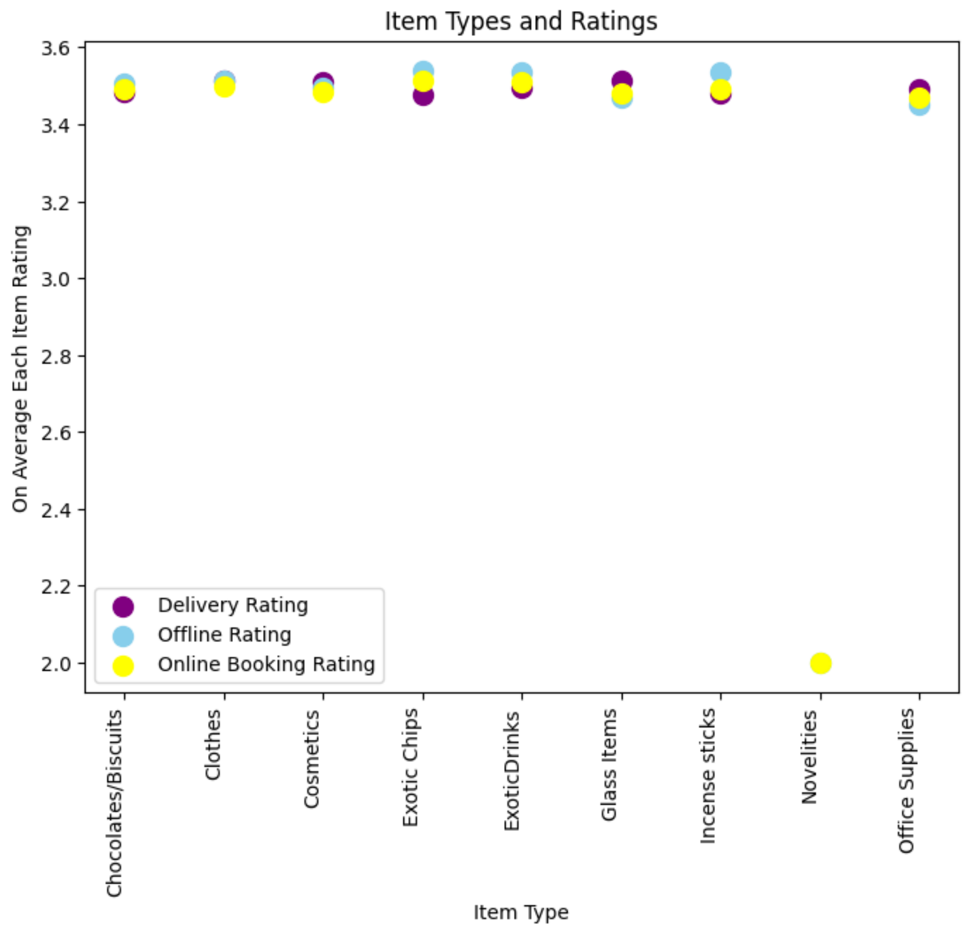


Figure (10) Item Types and Ratings.

Figure represents the scatter plot for the ratings and products in the inventory. Here our most ratings are between 3.4 to 3.6. Our approximation of the average rate of the products will be 3.5.

Coming to each product, for Chocolates/biscuits we got slightly more ratings in the offline market and delivery rating is the lowest for this product. In clothes, all the ratings are almost equal. In cosmetics, the delivery rating has the upper hand and the rest two are almost equal. For the item exotic chips, the offline rating is more, followed by the online booking rating and last delivery rating. Exotic drinks take the path of exotic chips but with less margin. Glass items have more delivery ratings and the rest two are almost equal. Incense sticks are more inclined toward the offline ratings compared to the other two ratings. Novelties have only an online market and are doing quite poorly than the others. Office Supplies have more delivery ratings and decent offline and online ratings. Except for the office supplies excluding the novelties, every other product has more offline ratings than online ratings.

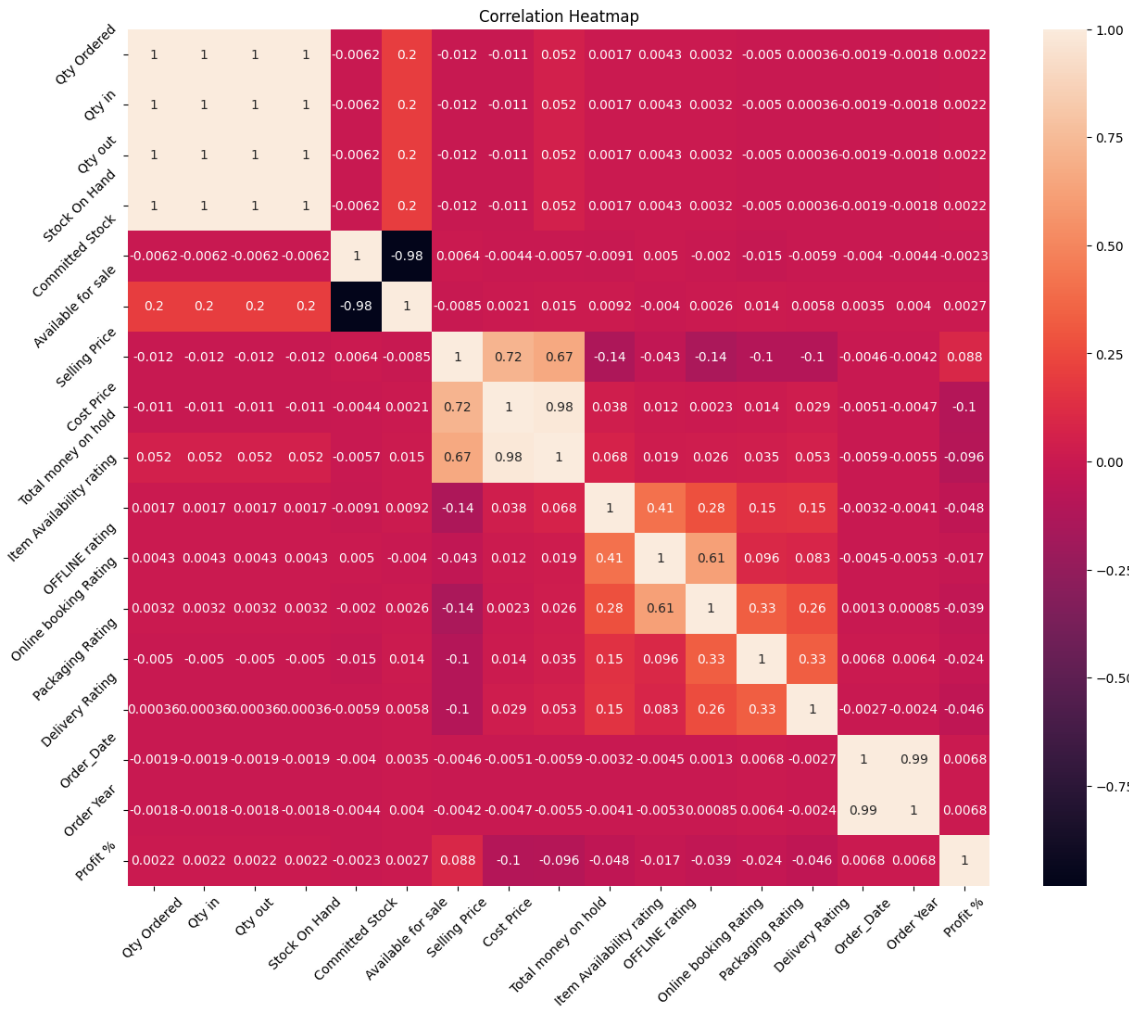


Figure (11) Correlation Heatmap

Dividing the correlation between the variables based on Pearson Coefficient:

Highly correlated variables: The variable “total money on hold” is highly correlated with the variable “cost price” (0.98). That employs the fact that total money on hold will be dependent on the cost price we are setting for the product.

Above Moderately correlated variables: The variable “cost price” is more moderately correlated with the selling price. (0.72) Because of that selling price might be decided based on the cost price but it is not entirely dependent on that, and it might include other factors as well. One more that subsequently comes under this category is the correlation between the variable “total money on hold” and “selling price” as we know that total money on hold had some high influence on the cost price and cost price is related to the selling price. So, the variable of the total money on hold also had some indirect influence which is reasonably noticeable in the selling price variable (0.67). One more relation that falls into this area is the correlation between the variable “online booking rating” and the variable "Offline rating”. It indicates that online and offline ratings are reasonably effective on each other.

Lower Moderately correlated variables: In this area of the section relativity is moderately correlated quite lower. So, the variables “offline rating” and “item availability rating” have partial effects on each other. We can presume that customer satisfaction might have a chance depending on when that product is available in the warehouse when he visits the warehouse which finally gives us the Offline rating (0.42).

Low Correlated Variables: In this category, the correlation between the variables is low but it still might affect each other. The variables “online book rating”, “packaging rating” and “delivery rating” are correlated with each other (0.33). A possible reason could be that a warehouse might receive online orders and for the delivery of the products we should have proper packing. In our thoughts, some ripple correlations are prevailing here (0.33). The variable “item availability rating” and the variable “online rating” have the same kind of effect as mentioned above. The reason might be that online bookings take some time from order to delivery. Sometimes product availability could be managed during that buffer period if we don’t have but we can’t say that would be the condition for all times. This could imply why the correlation is low. Two more pairs take place in this category, the first pair is the variable “packing rating” and the variable “item availability rating” and the second pair is the variable “delivery rating” and the variable “item availability rating”. These pairs suggest to us low correlation and minor influence on each other.

4. Methodology

4.1. ARIMA Model:

ARIMA is a Autoregressive Integrated Moving Average, a time forecasting model which uses the past data to predict a series patterns of future trend. It combines autoregressive (AR) and Moving Average (MA) components with differencing (I).

The Model starts by creating a required data frame 'Demand' by constituting the 'Available for sale' and 'order date' from the original data frame that is our dataset. From the order date column of the demand data frame, we extracted the Month and year. After that, Feature Engineering was done to create a Demand Column from the available-for-sale column by mapping the values of Available for sale, if they are positive as 0 and if they are negative as positive as demand. Create the Final Demand data frame. Grouped it by the Month and year with the aggregated sum of Demand column. Plotted a seasonal trend on the Final demand Data frame to check the demand data points over these years of period.

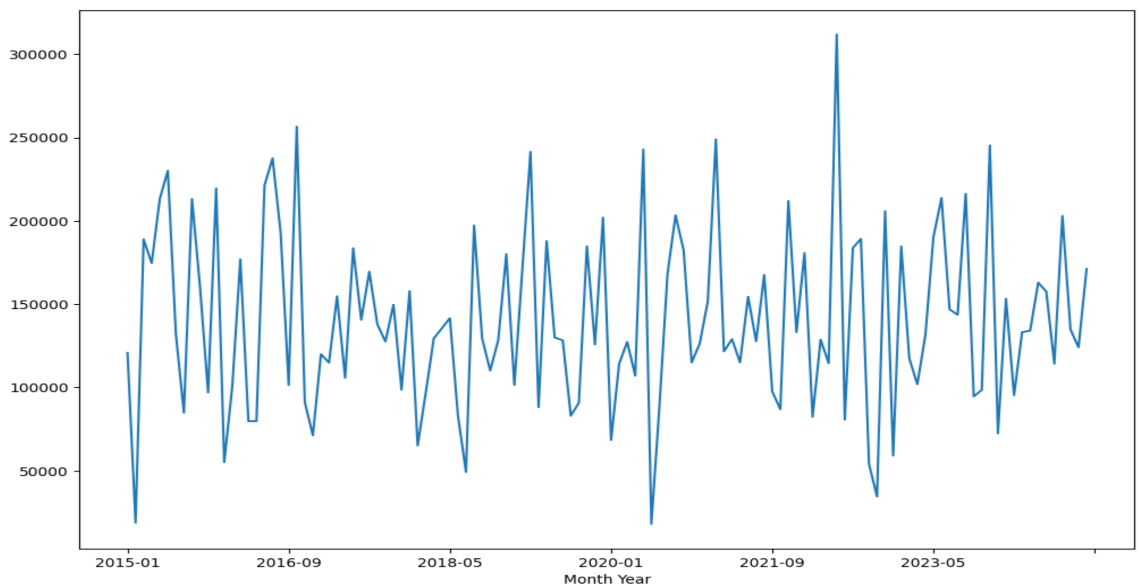


Figure (12) Stationary Graph

From this Plot, we understand that demand data points during the time are stationary. To check the stationarity of the above time series we can do it by ADF (Augmented Dickey-Fuller) Test, ACF (autocorrelation Function) and PACF (partial autocorrelation function).

ADF (Augmented Dickey-Fuller) Test: In this test, we can check the stationarity of the demand time series. Stationarity is the prime factor for the time series models like ARIMA. If the time series is non-stationary, we need to perform some transformations to make it stationary. After running the ADF Test we found that the ADF Statistic is -13.043747 and the p-value is 0.0000. With the given information we can interpret that the p-value is less than 0.05 indicating that the data is stationary.

ACF (autocorrelation Function) and PACF (partial autocorrelation function): These plots are used to identify the lag values and provide insights into the underlying structure of the time series. These structures help to determine the appropriate orders of AR and MA terms in the ARIMA Model.

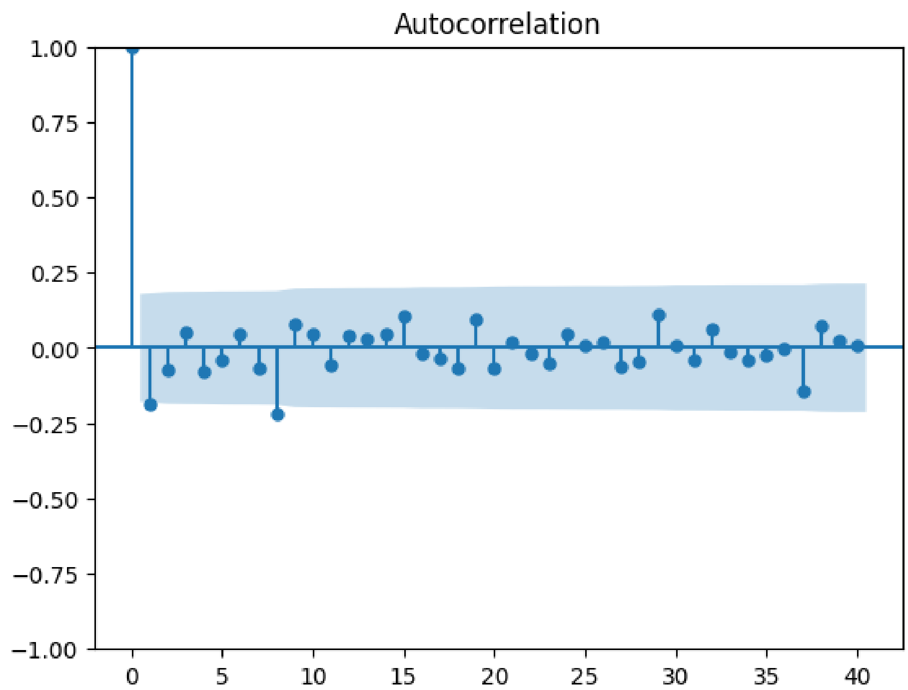


Figure (13) Autocorrelation

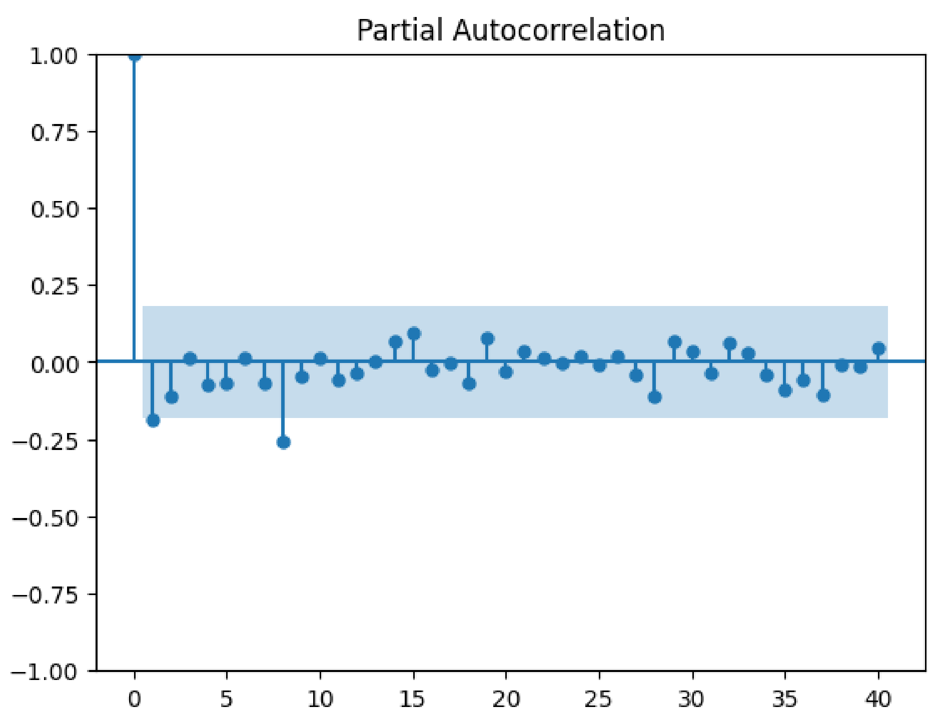


Figure (14) Partial Autocorrelation

The above plots show us that less than 5% lags cross the confidence interval of 95%. It can be justified that the data is stationary statistically.

On this basis, we are using the Arima Model for creating the forecast as the demand time series is stationary. After we ran the Arima Model we got this result.

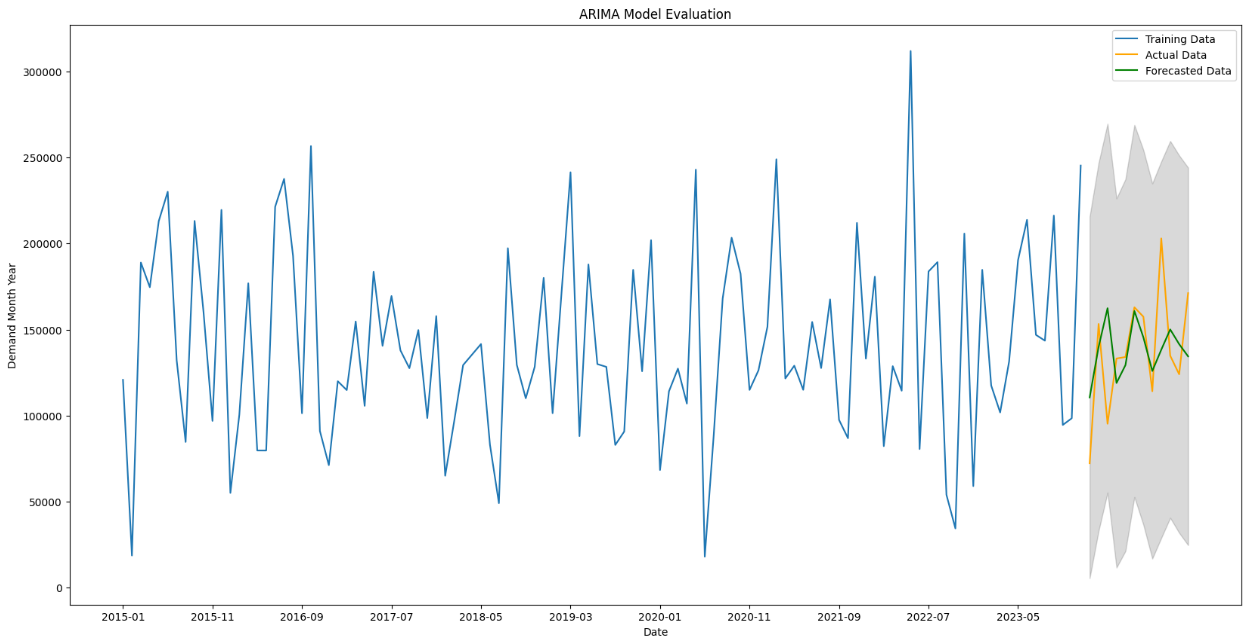


Figure (15) ARIMA Model Evaluation

From this model prediction, we got the RMSE score is 35290.40712222344 and in the forecasted data there is a dip of forecasted data than the actual data. It indicates a loss in demand for the products. Whenever there is a sudden dip or loss in actual data it might be affected by factors like economic shifts, market changes, new regulations, or other non-recurring events like COVID-19.

4.2 Classification:

We are using three classifications: logistic regression, Decision Tree classifier and Random Forest classifier.

Logistic regression: It is a statistical method that models the probability of discrete outcomes being given an input variable. It is mainly used for prediction or classification. It gives the relationship between the predictor variables and a categorical response. There are mostly three kinds of logistic regression, those are binary logistic regression, Nominal logistic Regression and ordinal logistic regression. In these logistic regressions, we use binary logistic regression which gives us True/False or yes/no or 1/0 these kinds of discrete values.

Decision Tree classifier: It is a machine learning algorithm that runs the model to predict the target variables based on feature lives by creating a tree-like structure under a supervised learning approach. In this tree structure, consists of root nodes, Internal nodes, branches and leaf nodes. This model is popular for its interpretation that can handle both numerical and categorical values.

Random forest classifier: It is an algorithm that runs on ensemble learning. It uses multiple decision trees in the training period and combines them to give a better prediction at the end. It is one of the most used algorithms due to its simplicity and diversity. It is popular for its high feature selection, near to precise accuracy and robustness.

Pre-Working

First, we need consider to splitting the data into two categories consisting of 30% test data and 70% train data and dealing with the categorical values. These categorical values are being encoded ordinally. Created dummy variables for the required variables.

Created a column called "Customer Satisfaction" in the data frame using the sum of these variables' column based on the "Item Availability rating", "offline rating", "Online booking rating", "Packaging rating", "Delivery rating" and all the other independent variables featuring into 'final_df' data frame. Now we set the pipelines for each model. Pipelines are used for maintaining order when the chain of steps is involved, where each step may be involved in the transformation of the data. Each pipeline has two main components which are the standard scaler (which helps us in standardizing the features) and a Model for each pipeline, where we have our classifiers (Logistic regression, random forest, and Decision tree classifier) with random seed of reproducibility.

Hyper-Parameters

Created the hyperparameters for the classifiers which have their complete order details incorporated for the parameters of each model. After setting the details we performed a fitting and tuning of the models using GridSearchCV, which is used for hyperparameters tuning by performing a search over a specified parameter grid to seek the best combination of hyperparameters.

In the results, every model has been fitted and run the model to get the Accuracy Score for both test data and training data and received the accuracy scores as follows:

Classification Models Results

Table (3) Classification Models Results

Type of Data/ Classifiers	**Logistic Regression(lr)**	**Random forest (rf)**	**Decision Tree (DT)**
Train data	0.659962871	0.697091584	0.661355198
Test data	0.665904716	0.677935515	0.659287777
difference	0.005941845	-0.019156069	-0.002067421

When you observe the table there is a minute difference between the Train data and the test data that can be seen as negligible and by this interpretation, we can say that our models are fitting properly, and they are neither overfitting nor underfitting. The random forest classifier is the most accurate.

Performance Metric Models

Plotted our models in a bar plot with accuracy, specificity, and sensitivity metrics. These metrics give us the correct measurements of models. First, we need to create the list for these metrics and create a confusion matrix to learn about the predicting values that give us these values TP (true positive), TN (true negative) and so on for the models we ran.

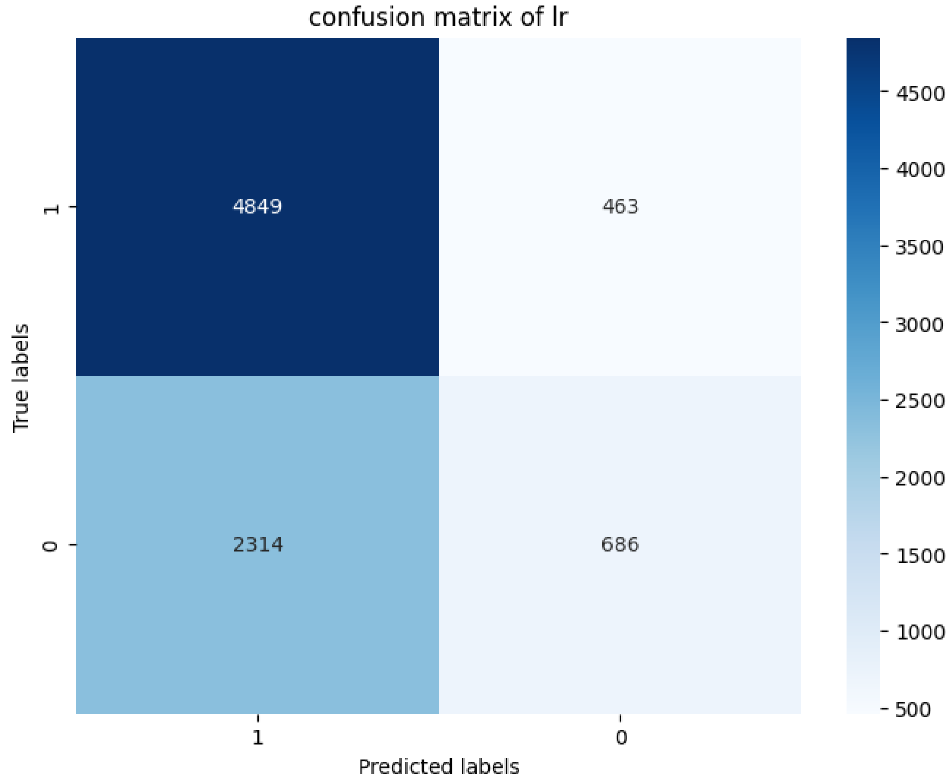


Figure (16) Confusion Matrix of lr

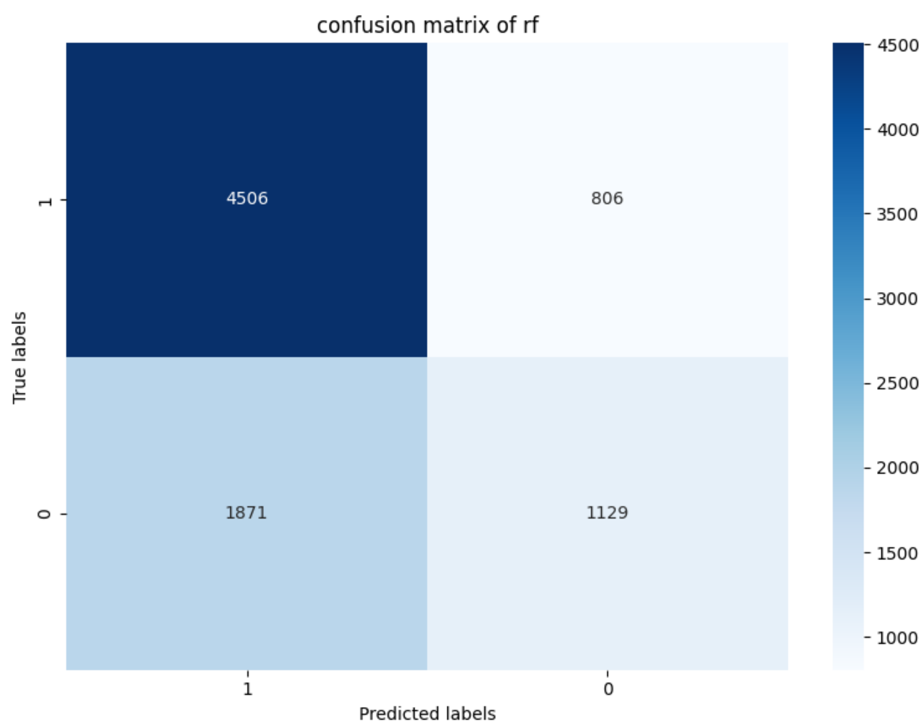


Figure (17) Confusion Matrix of rf

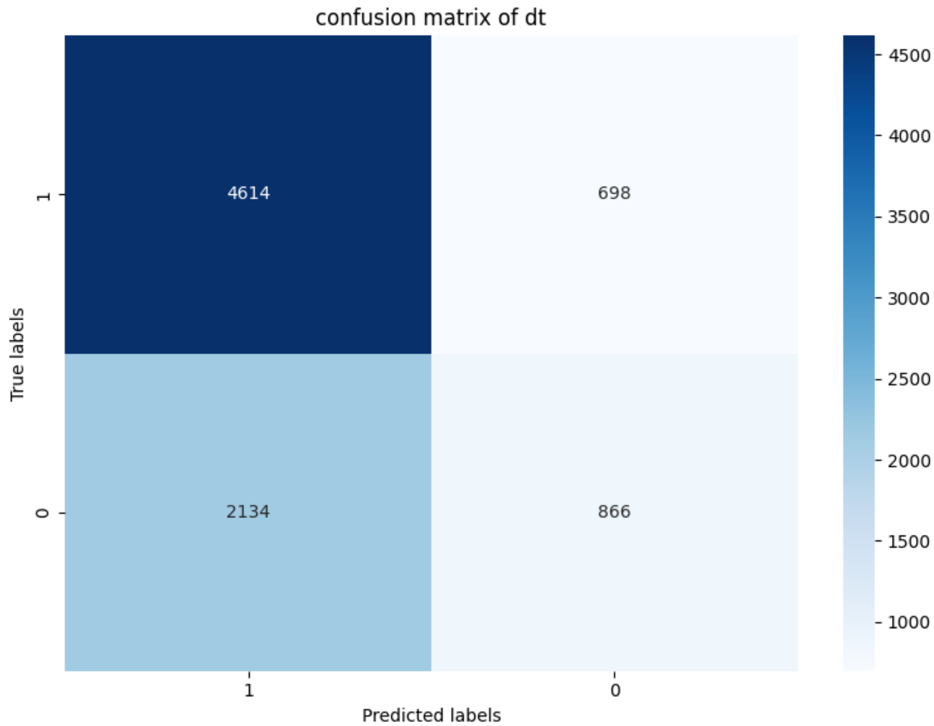


Figure (18) Confusion Matrix of dt

With this confusion matrix of the models, we can know how values are transformed during the prediction which gives us accuracy, sensitivity, and specificity at the end.

Using the metrics list we plot a bar graph for a better knowledge of these.

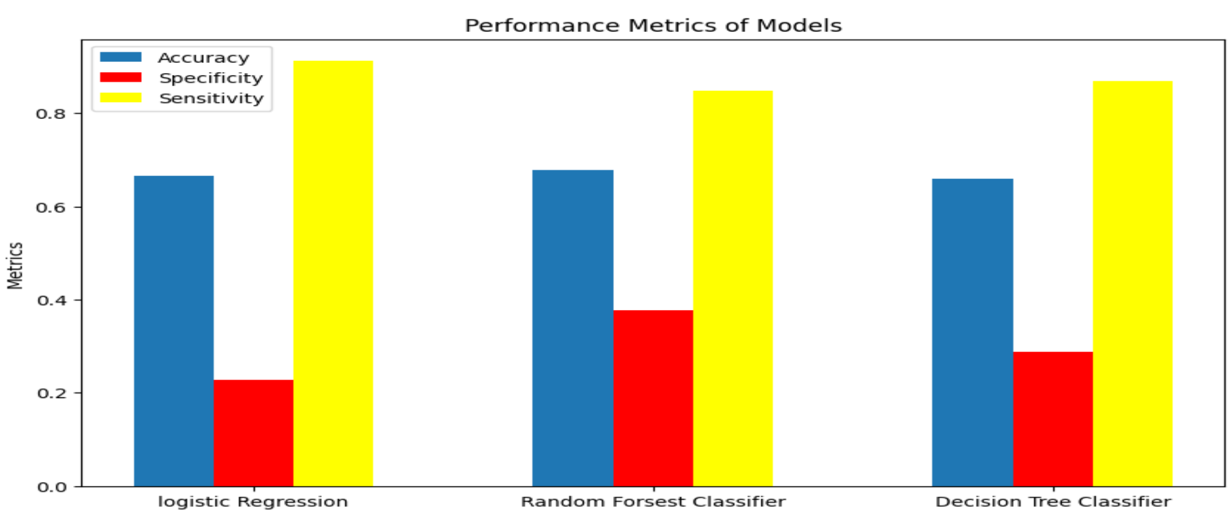


Figure (19) Performance Metrics of Models

From the bar graph, we can understand that our place of interest is given by the Random Forest classifier. Specificity is also more in the random forest classifier which lets us know that it is giving us the truer positive values in the prediction that we can rely on, and sensitivity is quite opposite to the specificity which gives more true negative values where logistic regression takes top place. Randomness can also be affected in these models.

4.3 Churn

Churn is the rate of decreasing of the number of items to the total number of items available. We are going to check how much quantity is decreasing on average for each item type for every quarter. This implies how much quantity could be reordered in every quarter for each item type. Taking the feature variables for determining the average churn and for that, we need to churn variable column and run the model on logistic regression as a regressor. We got an accuracy of 0.914094 and ran the model to test for the fitting. The average churn for each item type on quarterly basis is plotted.

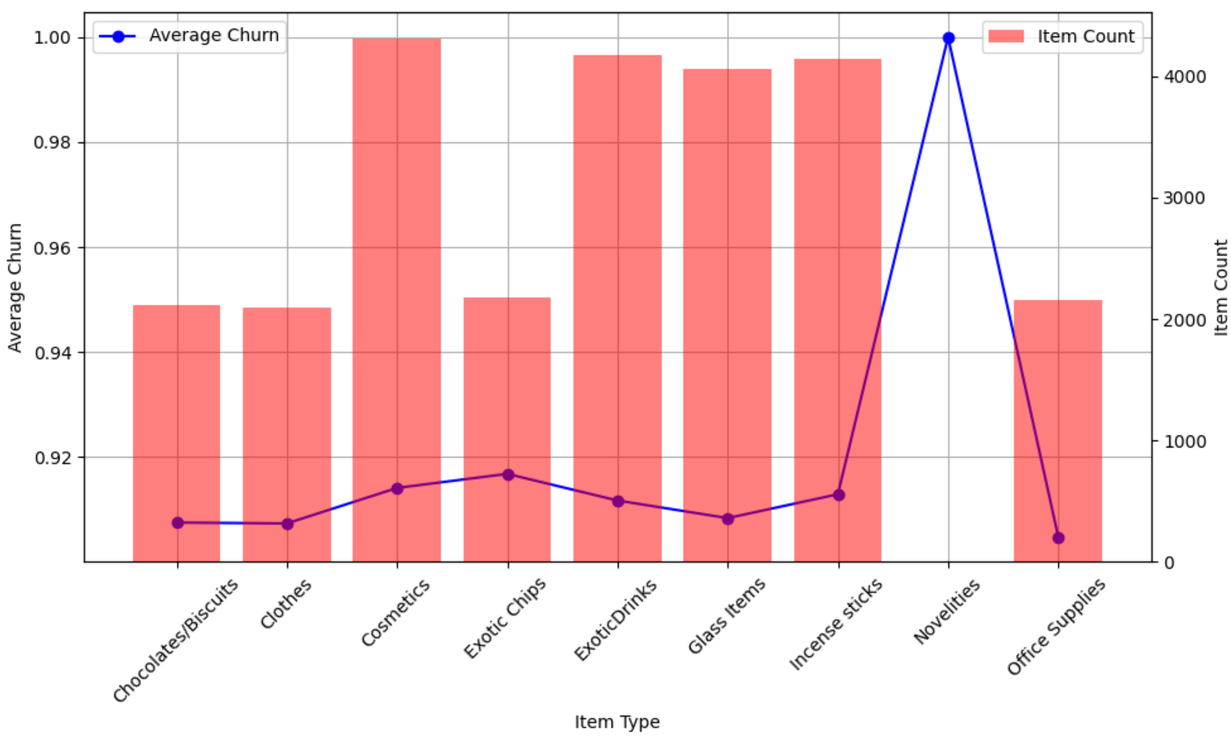


Figure (20) Average Churn Rate of Each Item Type for every quarter

The average outgoing of the orders is very poor for every item type but for novelties, since it is a one item it could be gone easily. That's why there is a high rate of churn for that in the plot. The reason for failing to send out the items or sell them involves either poor customer orders or stocking them than necessary for future purposes that may increase the holding cost and high difficulty for the storage space in the Inventory.

4.4 Correlation between the Quantity ordered and item type.

The below visualization demonstrates the quantity ordered for each item type. We can infer that except the novelties every other item type is almost have same ordered quantity with minute differences. This could establish a fact that all the item types are accommodated in the inventory equally. The quantity ordered depends on each item type sales and profitability.

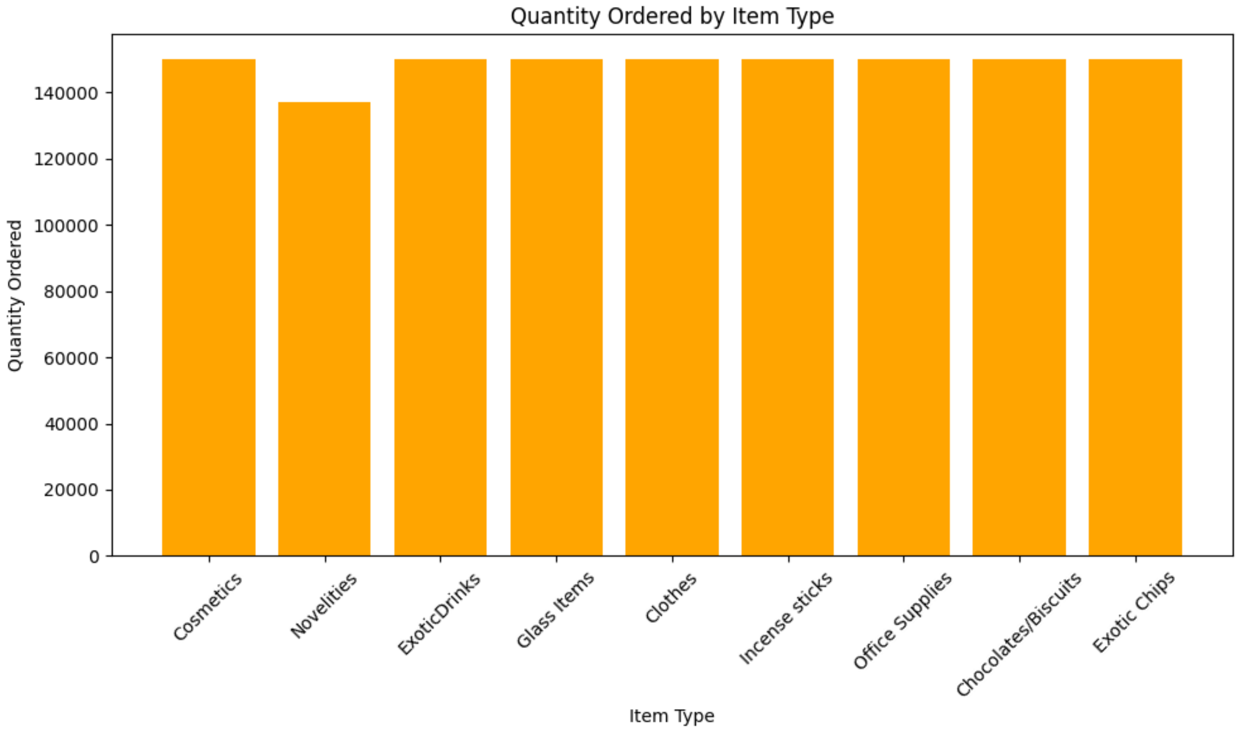


Figure (21) Quantity ordered vs Item Type

5. Results

In our research we unlatched the insights and patterns by answering our challenging research questions using our analytical skills and predictive models from a complex dataset. In this expedition, we did statistical analysis to gain knowledge on how the data set is, data-visualization to seek some of the hidden patterns and extend the advancement on variables, ARIMA models paving the path for seasonal trends insights and potential external shocks on inventory along with that brings things into limelight of a solid fit with historical data and highlighting obstacles in precision demand forecasting. The RMSE score, p-value and ACF scores of this model signifies that there is a deviation from the actual demand and implies for the lost demand. Logistic regression, Decision tree and random forest variables are the classification models to deepen our knowledge on customer satisfaction. These classification models using their pivotal techniques and methods embarked a valuable information on revealing the relationship between featuring variables and customer satisfaction ratings, where every model is completely fitted and random forest is the most accurate. While correlation analysis and its heart map discovered the relationship between the variables, in which these variables 'total money on hold' and 'cost price' having high correlation. In addition to that we did the churn analysis which succumbed the average decrease rating of each items for every quarter in a year and indicated the outrageous decrease of the novelties and made realizing of its short quantities.

Research questions and answers

- What stock levels strike the perfect balance between availability and cost-effectiveness for various product categories?

From the EDA analysis the stock levels have been maintained well to provide a variety of products to the customers. The availability of products was excellent, but the cost effectiveness has issues since the availability of stock is more by 2-3 times the selling were holding cost increases affecting the financial performance. The main reasons may be due to credit of some products which is added to inventory showing the holding cost high and year- to-year expiry replacement by companies of certain products.

- How do we quantify the influence of seasonal trends and external factors on inventory requirements?

We have a given trend chart on the profit% for each item type during these years. As a result, we can quantify the more profitable items by raising their ordering quantity which yields the needs of the inventory and there is high peak in the clothes (item_type) due to external factors like COVID-19 had a great influence during that year. Sometimes these types of external factors can make a direct influence inventory requirement.

- Can predictive analytics tools be harnessed to anticipate future stock needs more accurately?

The future seasonal trends obtained from the ARIMA model was good where actual data followed the forecasted data which means stock was more available to customers meeting the demand leading to availability and customer satisfaction.

- How do packaging and delivery ratings influence customer satisfaction and, by extension, the propensity for repeat purchases?

We got a certain measure to judge customer satisfaction by feature engineering all the customer rating variables into this and with inventory operations predictors. Using the classifiers with feature selection variables we ran the models and got almost 70% accuracy with the best model. This indicates how ratings are influencing customer satisfaction.

- What are the correlations between the variables?

The correlation of the variables is determined by the heatmap and considering the Pearson coefficient (R-value) we can determine the statistical measure of relationship between the variables and how they are corresponding to each other. The variable ‘Total money on hold’ and ‘cost price’ are highly correlated. These relations explored with help of these relations.

6. Conclusion

The analysis of the inventory management dataset was by using various data science techniques with the help of machine learning techniques like linear Regression, logistic regression, random forest and decision tree. Also the predictive time series analysis was used to predict the forecasting demand of products with historical data. All these methods were implemented for research purposes and the results have been reliable and thus the exercise was rewarding in applying these diverse techniques to real time world data projects. The classification methods helped in identifying accuracy of customer satisfaction with the help of other predictors that involved in inventory operations within the data. The ARIMA model provided insights into time series pattern for the data mainly in understanding trends over time for forecasted demand.

There is much more scope inspite of the progress made, the future exploration is on integration of external factors like transportation, financial depression and calamities like COVID etc. The correlations can also be focused on interplay between ratings and management practices.

References

[1] W. Kenton, "What Is Inventory? Definition, Types, and Examples" investopedia.com. https://www.investopedia.com/terms/i/inventory.asp (accessed March 27, 2024)

[2] D. N. Chorafas, Warehousing: Planning Organising and Controlling the Storage and Distribution of Goods. Macmillan Press Ltd, London, 1974.

[3] "Traditional vs. modern methods of inventory management" subscribed.fyi https://subscribed.fyi/blog/traditional-vs-modern-methods-of-inventory-management/ (accessed March 27, 2024)

[4] E. Kusrini et al. "Determining key performance indicators for warehouse performance measurement – a case study in construction materials warehouse.", in MATEC Web of Conferences. 154, 01058, 2018.

[5] S. Roumiantsev and S. Netessine, "Should inventory policy be lean or responsive? An Empirical Analysis," Working Paper, University of Pennsylvania, 2005.

[6] D. O. Esther and W. Elizabeth, "Effect of Inventory Management Practices on Organizational Performance in Public Health Institutions in Kenya: A Case Study of Kenyatta National Hospital," International Journal of Education Research, vol. 3, pp. 703-714, 2015.

[7] C. Drury, "Management and Cost Accounting" 6th ed. Thomson Learning, London, 2005

[8] Sarah Amsler, "inventory management" techtarget.com https://www.techtarget.com/searcherp/definition/inventory-management (accessed on March 27, 2024)

[9] David Ingram, "Traditional Inventory Systems", bizfluent.com https://bizfluent.com/traditional-inventory-systems.html (accessed on March 27, 2024)

[10] "Advantage Of Traditional Inventory Management Techniques", businesscoachphil.com https://www.businesscoachphil.com/advantage-of-traditional-inventory-management-techniques (accessed on March 27, 2024)

[11] "How to Solve Traditional Inventory Management Problems", claruswms.co.uk https://claruswms.co.uk/challenges-for-traditional-inventory-management/ (accessed on March 27, 2024)

[12] Maarten van Geest et al. "Design of a reference architecture for developing smart warehouses in industry 4.0", computers in industry, 2021a;124:103343.

[13] Abdikani SA, Mahad MA, Mohamed HD, "The Role of Inventory Management on Financial Performance in some Selected Manufacturing Companies in Mogadishu". Int J Account Res 6: 179, 2018.

[14] Kilonzo JM, Memba FS, Njeru D, "Effect of Inventory Management on Financial Performance of Firms Funded by Government Venture Capital in Kenya". Eur J Bus Manage 8: 35-41), 2016. https://core.ac.uk/download/pdf/234627094.pdf

[15] S. G. Eckert, "Inventory Management and Its Effects on Customer Satisfaction," Journal of Business and Public Policy, vol 1, no. 3, pp. 1-13, 2007.