

**BCIS 5110**  
**Programming Languages for Business Analytics**  
**Fall 2024**

**Project Proposal**  
**On**

**Strengthening Operations in Warehouses and Customer Fulfilment: An  
Analysis on enhancing Efficiency, Client satisfaction and Economic Outcomes**

**Group Members:**

## TABLE OF CONTENTS

SI No.	TOPIC	PAGE No.
1	EXECUTIVE SUMMARY	3
2	PROJECT MOTIVATION/BACKGROUND	3
3	LITERATURE REVIEW	3
4	DATA DESCRIPTION	5
5	DATA PREPARATION	6
6	EXPLORATORY DATA ANALYSIS	9
7	METHODOLOGY - MODELS AND ANALYSIS	16
8	RESULTS	23
9	CONCLUSION	26
10	REFERENCES	27

## **1. EXECUTIVE SUMMARY**

The primary component of any firm is inventory. Without an inventory, an organization finds hard to get started because of its important role in supply chain industry. It is important for operations of company which helps to find the financial results and success of the organization. By making good operation processes, it is best viewed as a resource by any company for their revenue cycle. Also, Good inventory helps business expand and satisfy customers. Effective inventory management is essential to supply chain optimization due to market trends.

The main aim of this project is to study XYZ Wholesale Company historical and present data with a focus on operational efficiencies, stock performance, cost enhancements and customer happiness based on ratings. This study will also employ classification models and graphs to obtain conclusions from the data. The XYZ Company was founded in 2015 provide excellent service in product availability to clients. It has four divisions throughout XYZ to provide high-quality items and services at the most reasonable costs. Convenience stores, beer and wine shops, dollar stores, tobacco outlets and smoke hubs are sold.

## **2. PROJECT MOTIVATION/BACKGROUND**

Granting funds and more time for new technologies, following good procedures and methods is best for company competency and financial improvement. Adapting to the strategies will be beneficial in terms of meeting the customer satisfaction and financial growth through careful planning and technical methods. Finally, maintaining inventory and best operations is best for attaining the best balance of product availability, cost-effectiveness and customer satisfaction which leads to profits. Our research aims to contribute to this goal by conducting statistical analysis using logistic regression and predictive analytics such as ARIMA models.

## **3. LITERATURE REVIEW**

Internal operations depend on optimisation since it reduces complexity and increases efficiency in stock-level management procedures [1]. The products are rebranded and the orders are manually selected and distributed by brand in traditional warehousing. Furthermore, traditional warehouses typically use human tracking and recording techniques to manage inventory [2]. However, sophisticated automation and collaborative options are absent from the antiquated method [3]. Traditional inventory methods rely on manual procedures, which have an impact on output and judgement. Customer satisfaction is of a high calibre. Statistical analysis and sampling are required. Customer happiness is impacted by product unavailability. Customer satisfaction is correlated with completed orders and on-time deliveries for the deliveries made [4]. The following questions to be answered at the end of this project.

**Descriptive Questions:**

1. What is most available item type in the warehouses? Which item type has the least stock in warehouses?
2. How quantities are ordered distributed across the different warehouses?
3. What is annual trend of quantities ordered for each item type?
4. What insights are found between quantities in and out of the inventory?
5. Which item type showed the highest profit percentage in any given year and what might be the reason?
6. What is overall average rating of inventory items and how do ratings differ across item types? How do packaging, delivery and offline ratings vary for different products?
7. How cost price and total money on hold correlated and what do you find?

**Predictive Answers:**

- 1) Can the future stock requirements be predicted using ARIMA models? What impact showed by changing market trends on future Stock storages?
- 2) Can customer satisfaction be predicted scores based on packaging, delivery and availability ratings? Can customer ratings predict the repeated purchases made?
- 3) What are the likely trends in item churn rates for different product varieties?
- 4) Can seasonal trends help forecast peak demand periods for specific item types?
- 5) How do changes in cost prices influence the forecasted selling prices over time?

#### 4. DATA DESCRIPTION

The dataset is a real-time dataset gathered from XYZ Wholesale from Zoho inventory between 2015 and 2024 which includes substantial stock and order data from operational sub warehouse activities. It includes all variables that are critical for analysing key parameters such as customer satisfaction, order processing performance and stock requirements of the products that are kept for sale.

**Dataset Name:** Inventory from XYZ Wholesale

**Source:** Dataset Is Taken from Zoho App of XYZ Wholesale

**Dataset Details:** 28k Rows and 26 Columns.

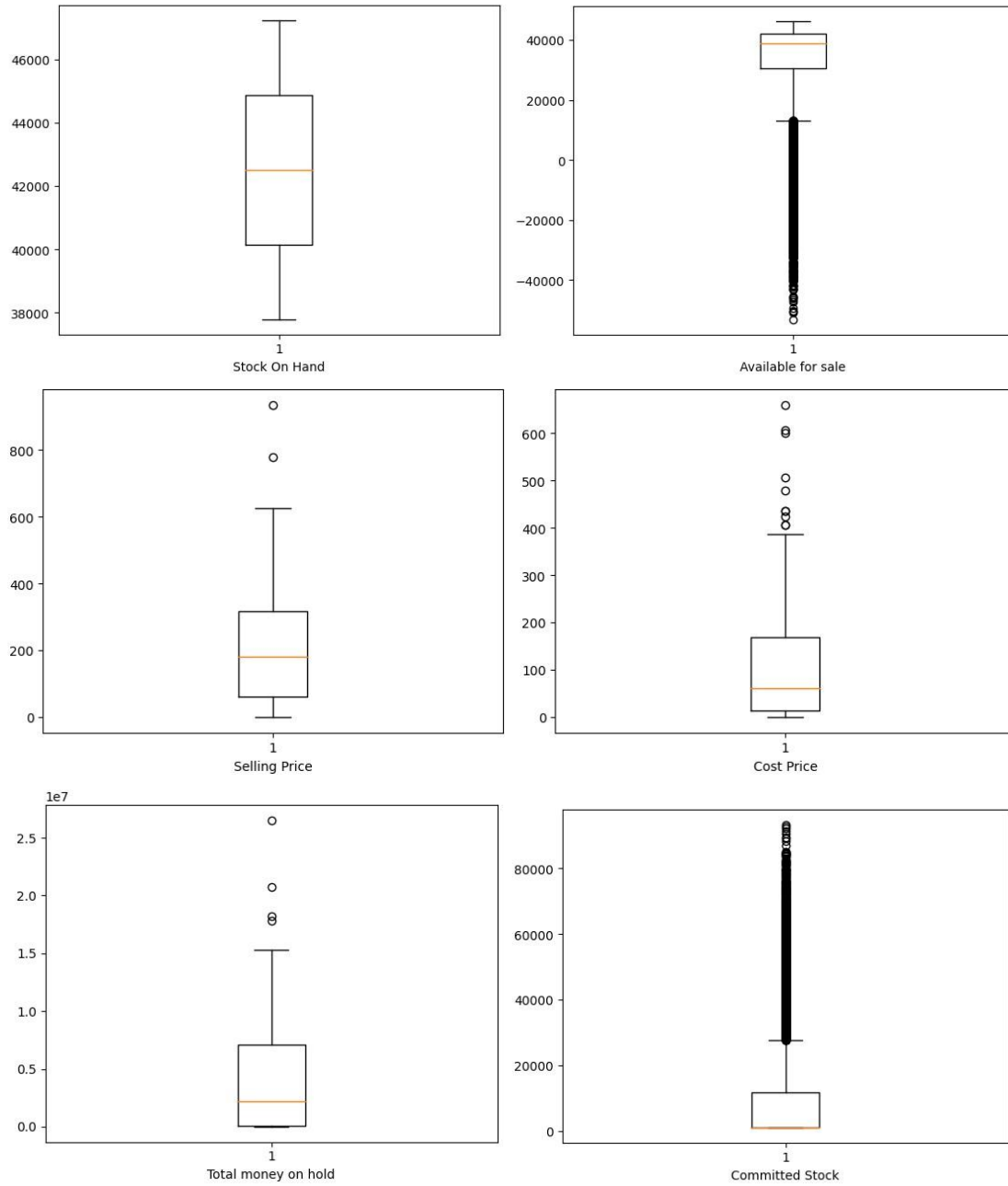
##### Column Description:

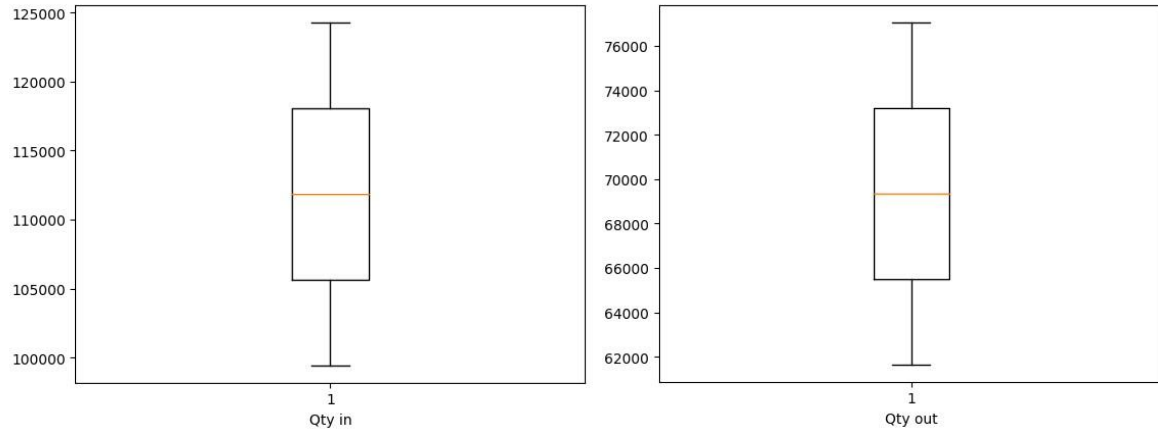
Variables	Description of the variables
Item ID	A unique automatic number generated that is assigned to the new item
List of Items	Description of items that are in the warehouse
SKU	Bar code for scanning a product.
Warehouse Name	Four places of different warehouses present in USA
Item Type	Product of each variety of item
Purchase Invoice Channel	Invoice purchase channel via online/offline
Order Priority	Priority of order packaging
Order Date	Order date made by the customer
Order ID	Identification code or number for the order
Qty Ordered	Quantity of item ordered by the clients
Qty IN	The Quantity that came to warehouse of different items.
Qty OUT	The items that are sold to the customer
Stock on Hand	The items that are to be sold or left out in warehouse
Committed Stock	The items committed by company to clients which is available or not available.
Available for Sale	The items left after checking committed stock
Status	The item is active or inactive

Unit	Number presence of each item
Selling Price	Price sold to client
Cost Price	Price brought from manufacturer
Created By	Item creation when entered into the market
Total Money on Hold	The money that is withheld in form of stock
Item Availability Rating	Rating by client for the item or entire order.
Offline Rating	Rating by client in-store purchase
Online Booking Rating	Rating by client on online purchase
Packaging Rating	Rating by packaging given by client
Delivery Rating	Rating by the client for delivery

## 5. DATA PREPARATION

The data is prepared by using difficult procedures. For ensuring data cleaning of the dataset, data was handled carefully for values missed. Four columns were found missing like Delivery Rating and others. Missing values were imputed using the median rating for Delivery Rating and Purchase Invoice Channel computed individually for online and offline channels. This strategy ensured that the imputed values reflected the trends of the appropriate channel. For efficient and effective warehouse operations, it is critical to determine whether outliers are caused by situations like data entry errors, stock control concerns, price issues and anomalies in sales patterns.





*Figure (1) Outliers - Operational and Financial Variables*

The cost of buying, total money on hold and selling price all have outliers that are higher than the upper whisker indicating that some items have a higher cost price, unsold items cost the company money, quantity of items moving out is higher than usual and that some items are sold at a significantly higher price point than the majority of others. The Available for Sale variable outlier is located below the lower whisker indicating a product or stock deficit. The stock on hand and quantity-in plots show a symmetric distribution of data with no evident outliers. On the other hand, the committed stock variable shows outliers on the upper whisker indicating that some goods have a higher level of committed stock than others. This means that certain products have greater commitments such as outstanding orders or reservations.

### **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.



## 6. EXPLORATORY DATA ANALYSIS

We employed several visualization techniques to gain more visual insights that have a greater impact and can disclose more intriguing patterns, insights and viable knowledge on the data which may be more useful to research and offer us more emphasis on the areas where we concentrate. We utilized bar graphs, pie charts, heat maps, and line charts to analyse the data.

Summary 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		33458.7991
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

Figure (2) Summary statistics

The figure 2 contains statistical analyses of variables (mean, min, max, standard deviation, count, and so on) that provide a basic overview of the dataset and numerical insights of data.

The warehouse has a bigger quantity of cosmetics than other commodities. This shows that cosmetics are among the most abundant products in the inventory. Following cosmetics, Exotic Chips, Office Supplies and Chocolates/Biscuits are all available in similar quantities. These items appear to have comparable quantities in stock. Novelties appear to be the least stocked item in the warehouse indicating that they are the least available of the nine commodities.

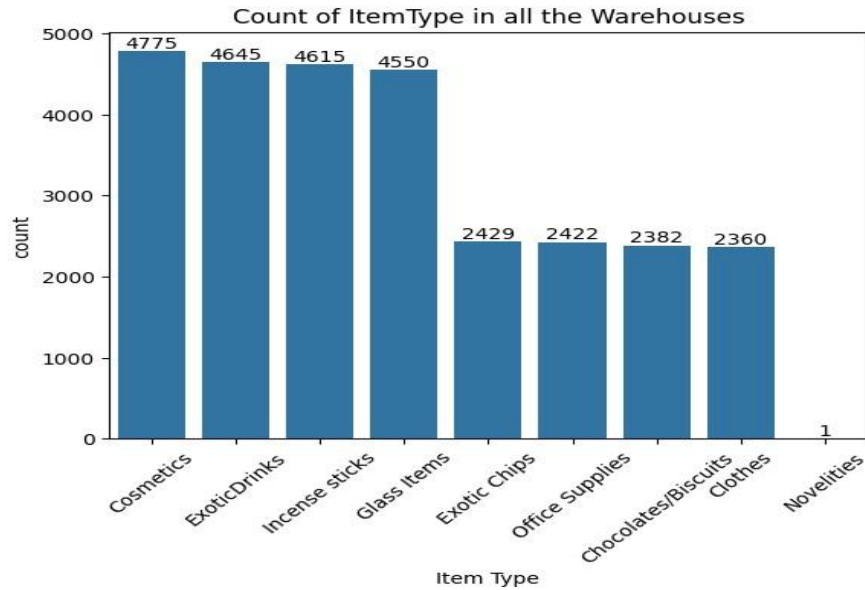


Figure (3) Count of item type in all the warehouses

In summary, the bar graph offers information about the distribution of commodities in the warehouse demonstrating that cosmetics dominated the inventory while novelties have the lowest representation. This knowledge can be critical for inventory management and resource allocation within the warehouse.

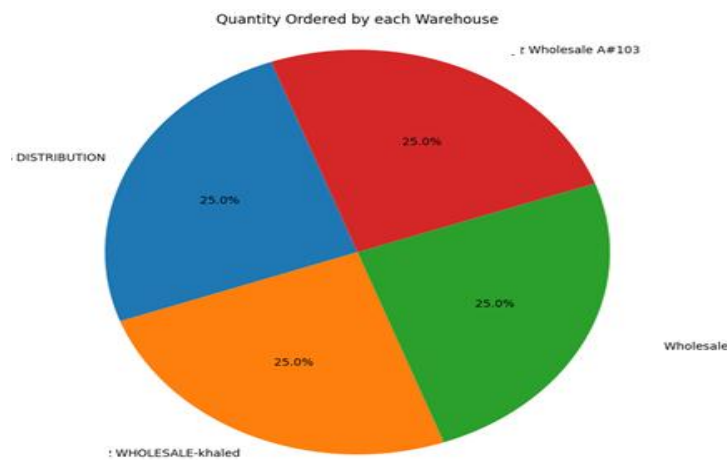
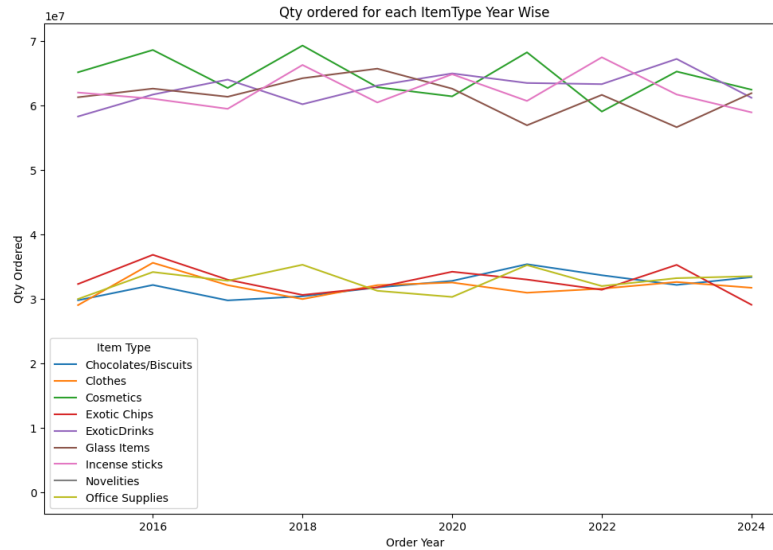


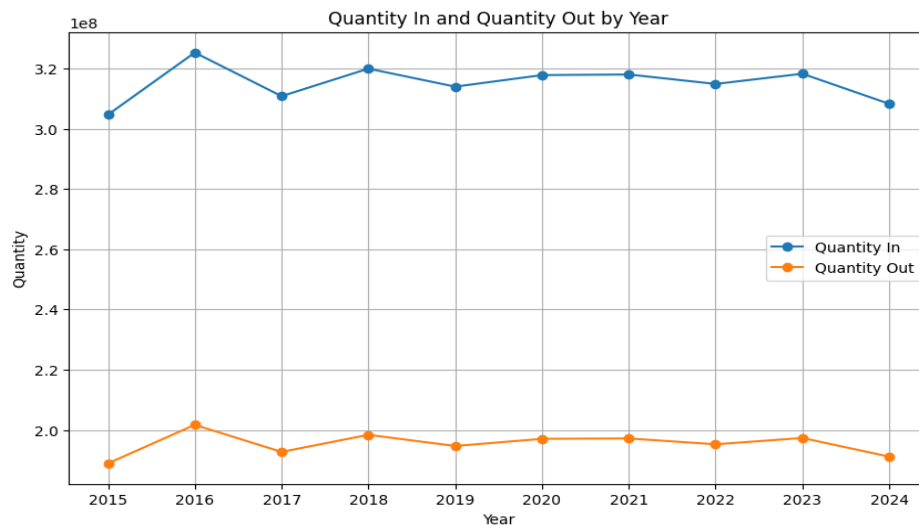
Figure (4) Quantity ordered by each Warehouse.

The figure shows that every order is divided equally throughout the warehouses ensuring inventory management balance.



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

The line graph above depicts the yearly quantity ordered for each item type. The most popular item types are cosmetics, incense sticks, exotic drinks and glass products, which they re-ordered. All of the categories' quantities vary from year to year.



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

Figure demonstrates that the amount entering inventory is always greater than the quantity leaving inventory. This could result in a miscarriage using expired products. It is nearly 1.6 times the difference between the quantities in and out. In 2016, inventories accounted for the majority of our inflow and leaving products. However, we can consistently see that if we take a certain amount of quantity nearly the same amount of product is released. Quantity in and quantity out are clearly linked. The least goes to 2015. Again, there is a decrease in 2024 in ordering.

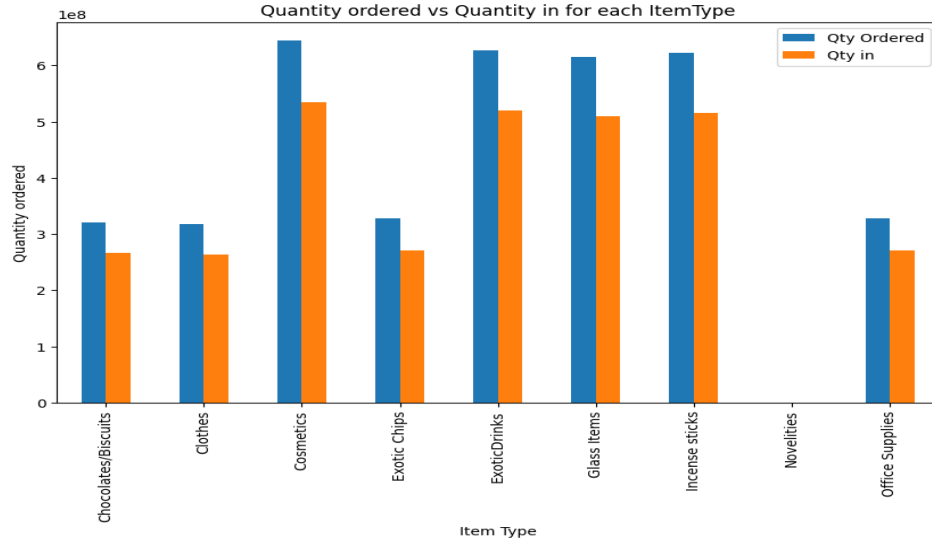


Figure (7) Quantity ordered vs Quantity In for each Item Type

We noticed variable quantities for each item category after categorizing them by quantity ordered and quantity in. In the dimensions indicated above, Figure  $e^8 = 10^8$  (for Chocolates/Biscuits, the total number ordered is about  $3.2e^8 = 321$  million, and so on for other item kinds). We have a total of nine different products. The blue hue represents the quantity ordered while the orange colour represents the quantity received. We can see that for each product, the quantity received is less than the quantity purchased. This indicates a discrepancy in the accomplishment of the conditions. This may sometimes result in a failure to meet the customer's requirements which lowers customer happiness. This also implies that warehouses are not being used to their full potential due to the difference between the quantity ordered and the quantity received into inventory. In terms of products, novelties are the least ordered while cosmetics are the most popular followed by unusual drinks and so on. The high-ordered category includes cosmetics, exotic drinks, glassware and incense sticks. Chocolates/biscuits, clothing, exotic chips, and office supplies are classified as low-ordered due to the significant variation in order margins.

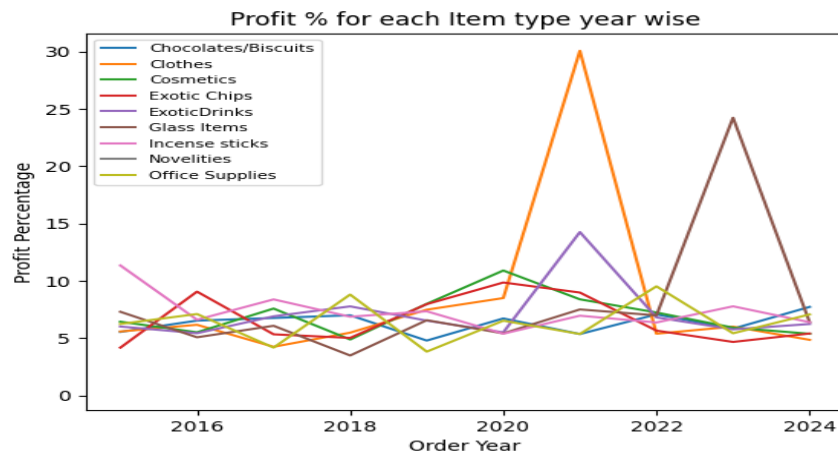


Figure (8) Profit % for each item type by year.

The figure depicts the profit % for each item over the course of a year. Every year, different item types were sold with each having its own ups and downs. In the first year, incense sticks were the most profitable item. The garments initially had a smaller profit percentage initially they had the largest profit percentage contributing 30% by 2021and small downfall in the following year. According to general study, the US clothing industry increased in 2021. This increase was attributable to a surge in demand following the COVID-19 epidemic. Glass items that generated the least profit in 2018 significantly increased their profit by 2023 contributing almost 25%. Because only one item was sold, the item novelty had no impact on the profit percentage.

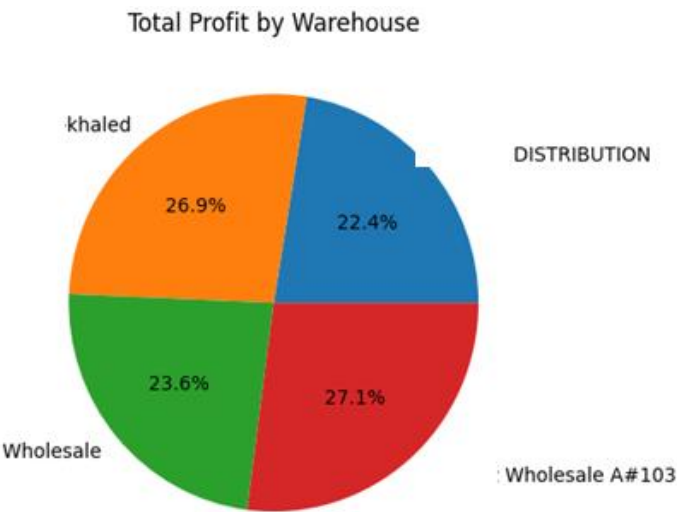


Figure (9) Total profit earned by each warehouse.

The above pie chart depicts the profit earned by each warehouse. XYZ Wholesale A#103 is the most profitable, while XYZ Distribution is the least profitable of the four warehouses.

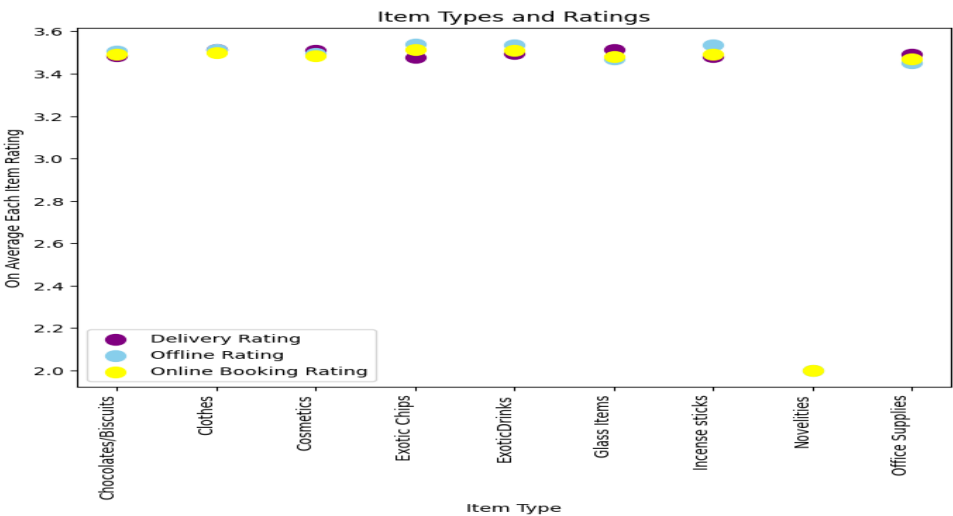


Figure (10) Item Types and Ratings.

The scatter plot depicts the ratings and products in the inventory. Our most common ratings range from 3.4 to 3.6. We estimate that the average rate of the items will be 3.5. Regarding each product, Chocolates/biscuits received somewhat higher ratings in the offline market, while the delivery rating was the lowest for this product. All clothing ratings are nearly equal. In cosmetics, the delivery rating has the upper hand while the other two are about equal. For exotic chips, the offline rating is higher, followed by the online booking rating and the latest delivery rating. Exotic drinks follow the same pattern as exotic chips, albeit with a smaller margin. Glass items have higher delivery scores, but the remaining two are practically equal. In comparison to the other two ratings, incense sticks are more likely to be rated offline. Novelties only have an internet market and are performing poorly compared to the others. Office Supplies have higher delivery ratings and better offline and online reviews. Except for office supplies and novelties, every other product gets higher offline ratings than online ratings.

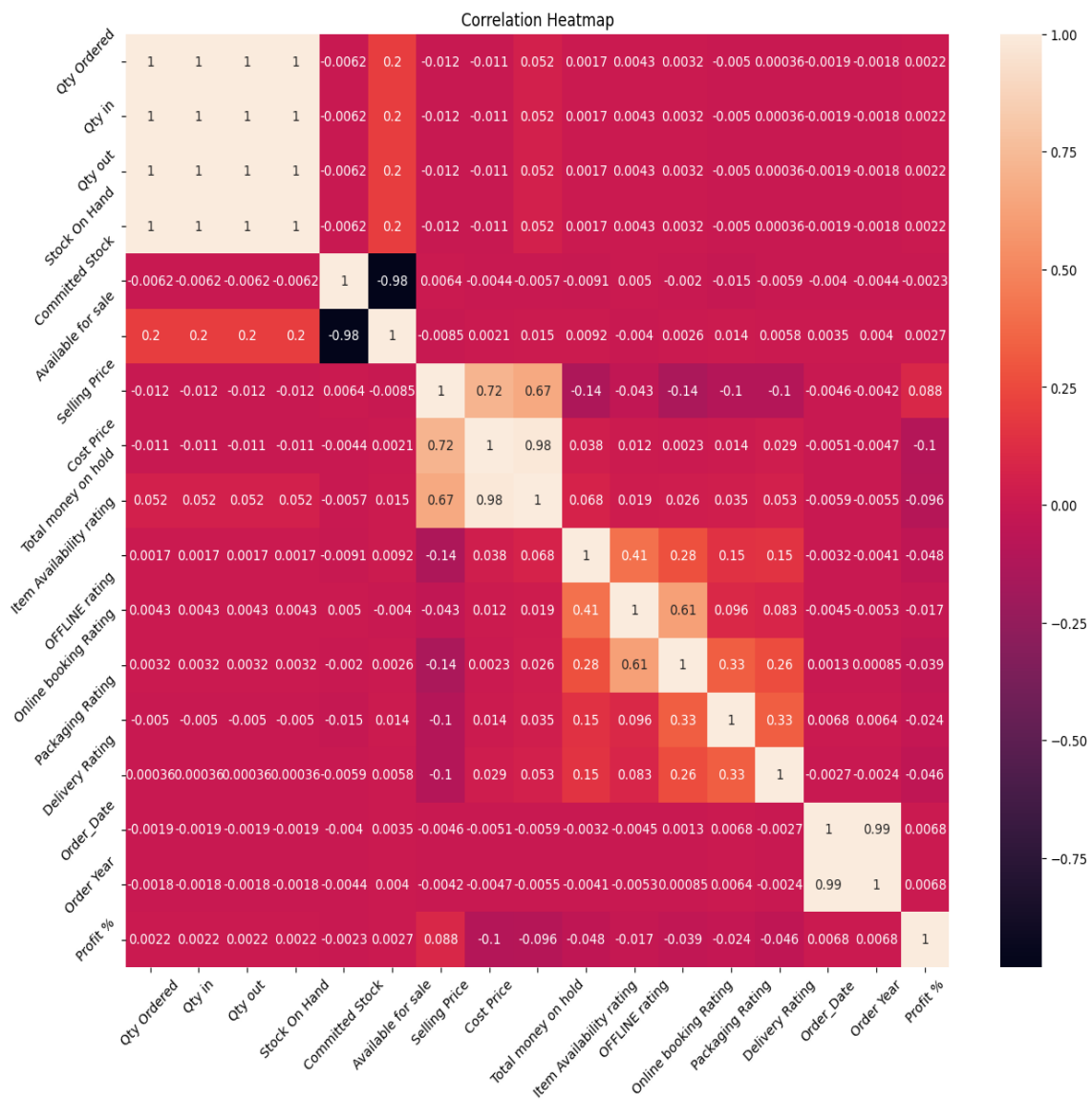


Figure (11) Correlation Heatmap

### **Dividing the correlation between the variables based on their relatability:**

**Highly correlated variables:** The variable "total money on hold" has a strong correlation with the variable "cost price" (0.98). This takes use of the fact that the total amount of money on hold is determined by the product's cost price.

**Above Moderately correlated variables:** The variable "cost price" has a more moderate correlation with the selling price. (0.72). As a result, the selling price may be determined based on the cost price, but it may also contain other elements. One further variable that falls into this category is the correlation between the variables "total money on hold" and "selling price," since we know that total money on hold has a significant influence on the cost price, and cost price is related to the selling price. So, the variable of total money on hold had an indirect influence on the selling price variable (0.67).

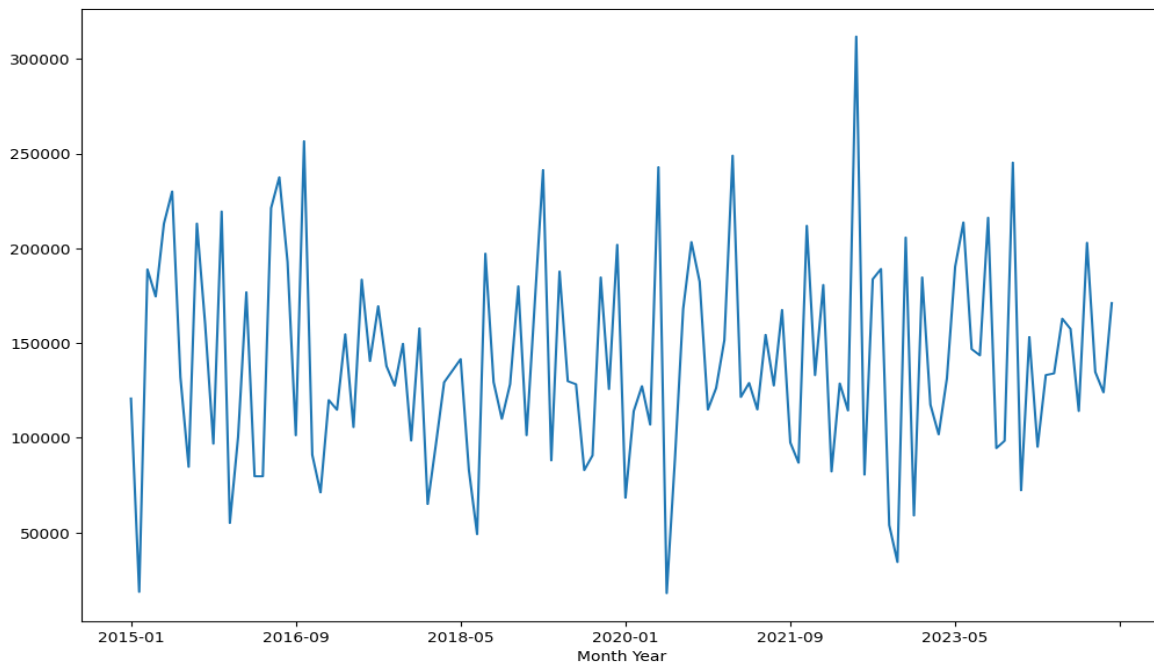
**Lower moderately correlated variables:** In this region, relativity is marginally correlated but significantly lower. As a result, the variables "offline rating" and "item availability rating" have only a partial effect on one another. We can assume that customer happiness may be affected by when the product is available in the warehouse when he arrives, resulting in an Offline rating (0.42).

**Low Correlated Variables:** In this category, the variables have a low correlation but may still effect one another. The variables "online book rating," "packaging rating," and "delivery rating" have a correlation coefficient of 0.33. One probable reason is that a warehouse may handle online orders, and suitable packaging for product delivery is required. In our opinion, some ripple correlations exist here (0.33). The cause could be that internet bookings take time from order to delivery. If we don't have the product, we can manage it throughout the buffer period, but we can't say that will always be the case. This could explain why the connection is poor.

## 7. METHODOLOGY - MODELS AND ANALYSIS

### 7.1 ARIMA Model:

ARIMA stands for Autoregressive Integrated Moving Average. It's a time forecasting model. The Model makes predictions in a series based on previous data. It incorporates the autoregressive (AR) and moving average (MA) components, as well as differencing. The Model begins by generating a required data frame 'Demand' by combining the 'Available for sale' and 'order date' from the original data frame, which is our dataset. We extracted the month and year from the order date column in the demand data frame. Following that, Feature Engineering was used to build a Demand Column from the Available for Sale column by mapping the Available for Sale values to 0 if they are positive and positive if they are negative. A seasonal graph was plotted on the Final demand Data frame to check the demand data points over these years of period.

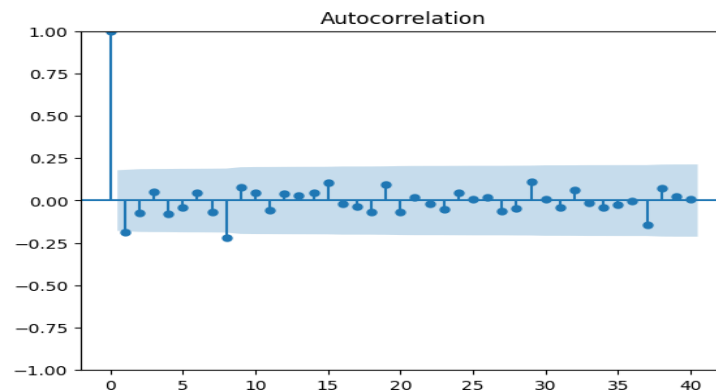


*Figure (12) Stationary Graph*

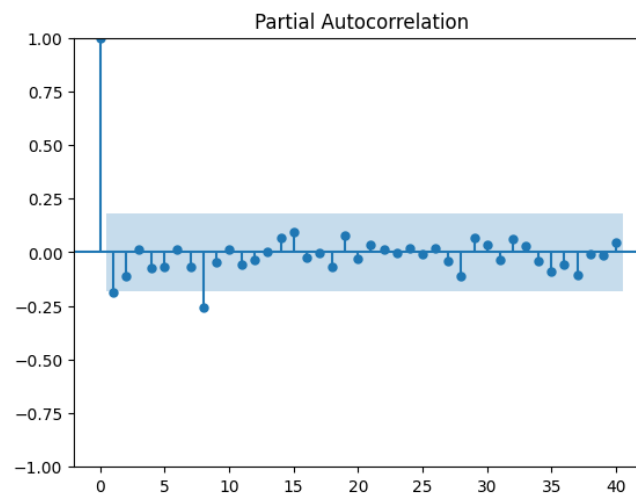
This plot shows that demand data points throughout time are stationary. The ADF (Augmented Dickey-Fuller) Test, ACF (autocorrelation Function) and PACF (partial autocorrelation function) can together used to determine the stationarity of the above time series. ADF (Augmented Dickey-Fuller) Test: This test determines the stationarity of a demand time series. Stationarity is a key element for time series models such as ARIMA. If the time series is non-stationary, we must apply several transformations to make it stationarity. After conducting the ADF Test, we discovered that the ADF Statistic is -13.043747 with a p-value of 0.0000. With the available information, we may conclude that the p-value is less than 0.05 implying that the data is steady.



ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function): These plots are used to identify lag values and reveal the underlying structure of the time series. These structures aid in determining the right ordering of AR and MA terms in the ARIMA Model.



*Figure (13) Autocorrelation*



*Figure (14) Partial Autocorrelation*

The charts above show that less than 5% of lags breach the 95% confidence interval. The data's statistical stationarity can be justified. We use the Arima Model to create the forecast because the demand time series is stationary. This is the result of running the Arima Model.

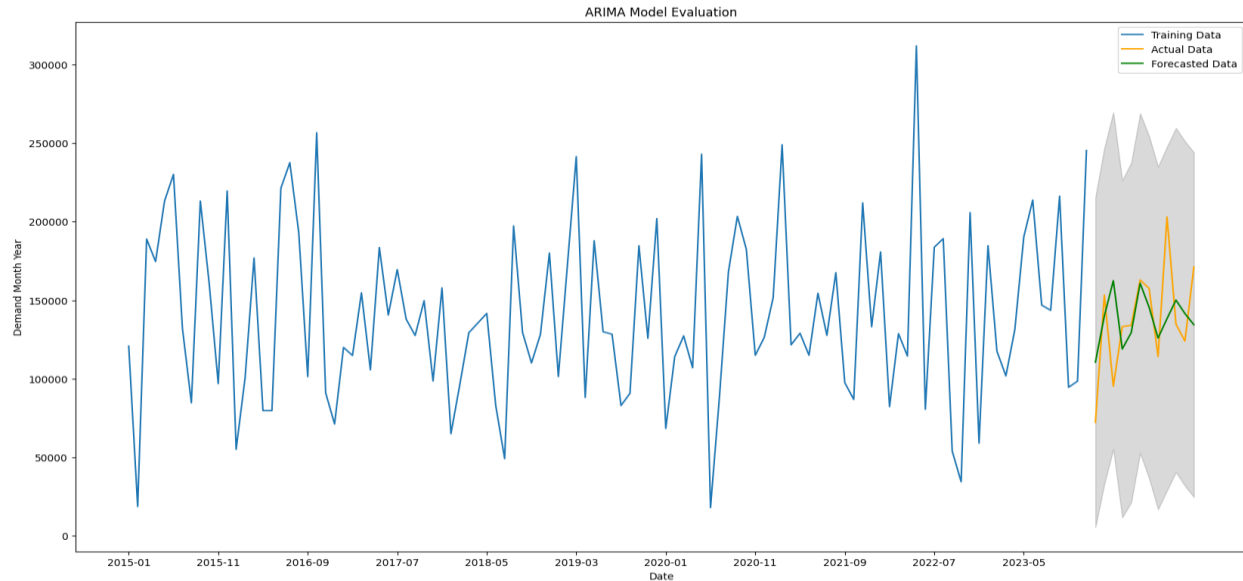


Figure (15) ARIMA Model Evaluation

We obtained an RMSE score of 35290.4 from this model prediction and the predicted data shows a decrease compared to the actual data. It suggests that there is less of a market for the goods. Any abrupt decline or loss in actual data could be caused by a variety of circumstances such as changes in the market economy, new rules or other occurrences like COVID-19.

## 7.2 Classification:

**Logistic regression:** Given an input variable, this models brings the likelihood of discrete outcomes. Prediction and classification are its primary uses. It provides the correlation between a category response and the predictor variables. Binary logistic regression, nominal logistic regression and ordinal logistic regression are the three main types of logistic regression. Binary logistic regression which provides True/False, yes/no, or 1/0 discrete values is used in these logistic regressions.

**Decision Tree classifier:** It is a technique that uses a supervised learning approach to build a tree-like structure that predicts the target variable based on features. This tree structure is made up of internal nodes, branches, leaf node and root nodes. Because it can deal numerical and categorical values, this model is widely used.

**Random forest classifier:** This algorithm uses ensemble learning to function. In order to provide a more accurate forecast at the end, it employs decision trees throughout the training phase. Its simplicity and variety make it one of the most used algorithms. Its perfect precision and large feature choices make it well-liked.

**Pre-Working:** Prior to addressing the categorical values, we must think about dividing the data into two forms with 30% of the data being test data and 70% being train data. Ordinal encoding is being used for these categorical variables. For the necessary columns , dummy variables were created.

Using the sum of these variables' columns based on the "Item Availability rating," "offline rating," "Online booking rating," "Packaging rating," "Delivery rating" and all the other independent variables appearing in the "final\_df" data frame, a column named "Customer Satisfaction" was created in the data frame.

The pipelines for every model are now configured. When there is a chain of processes involved, where each step may be engaged in the transformation of the data, pipelines are utilized to keep things organized. The standard scaler which aids in standardizing the features and a model which houses our classifiers with random seeds of reproducibility are the two primary parts of each pipeline.

### **Hyper-Parameters:**

The classifiers' hyper parameters were developed, incorporating all of the model's parameters' order details. Following the setup of the parameters, we fitted and tuned the models using GridSearchCV which is a tool for tuning hyper parameters that searches a parameter grid for the optimal collection of hyper parameters. Each model is fitted and ran to determine the accuracy score for both training and test data. The accuracy scores obtained are as follows:

### **Classification Models Results:**

Types	Logistic Regression(lr)	Random forest (rf)	Decision Tree (DT)
Train	0.66	0.70	0.66
Test	0.67	0.68	0.66
difference	0.01	-0.02	0.00

According to this view, we can state that our models are fitting correctly and are neither overfitting nor underfitting because of the small discrepancy between the train and test data that is seen in the table. The most accurate classifier is the random forest one. We used measures for accuracy, specificity, and sensitivity to plot our models in a bar plot. These measures provide us with accurate model measurements. To learn about the predictive values that provide us with these values TP (true positive), TN (true negative), and so forth for the models we tested, we must first establish a list for these metrics and a confusion matrix.

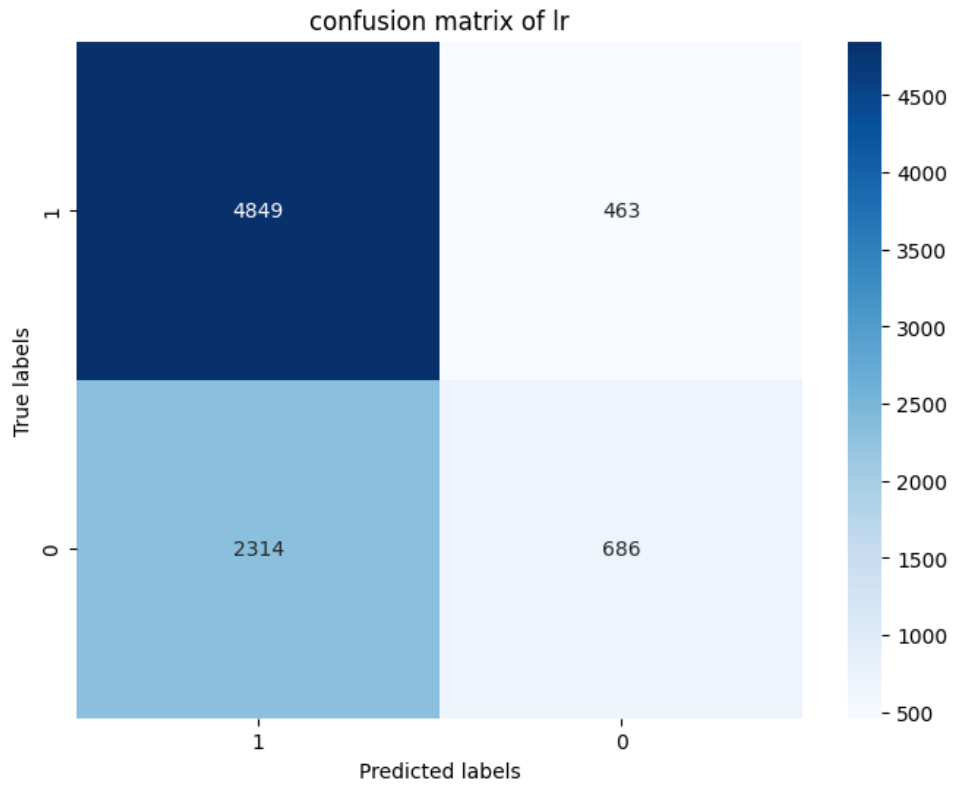


Figure (16) Confusion Matrix of lr

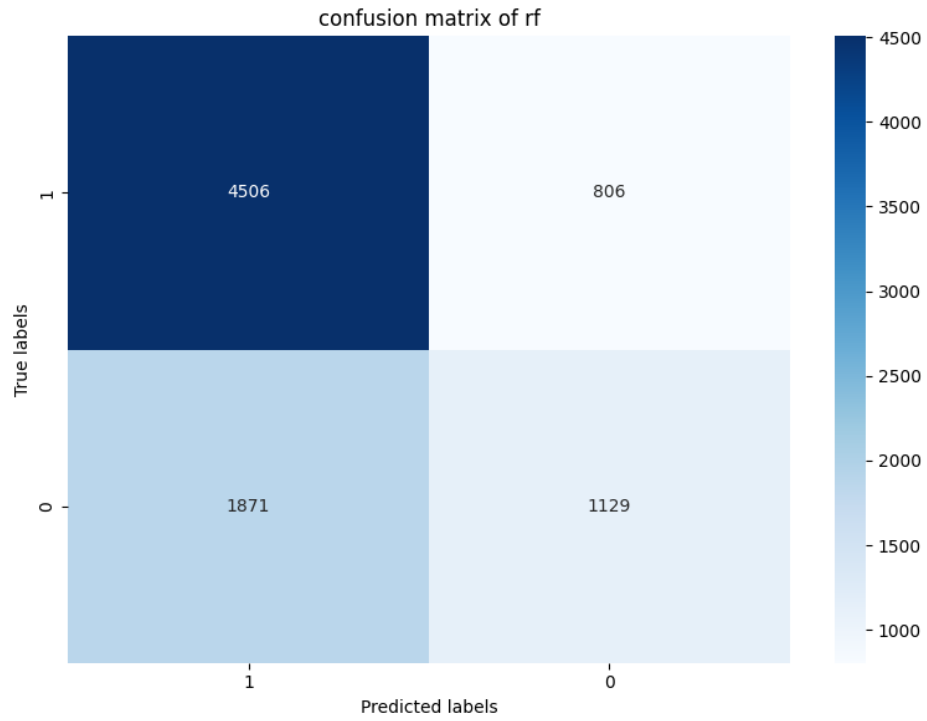


Figure (17) Confusion Matrix of rf

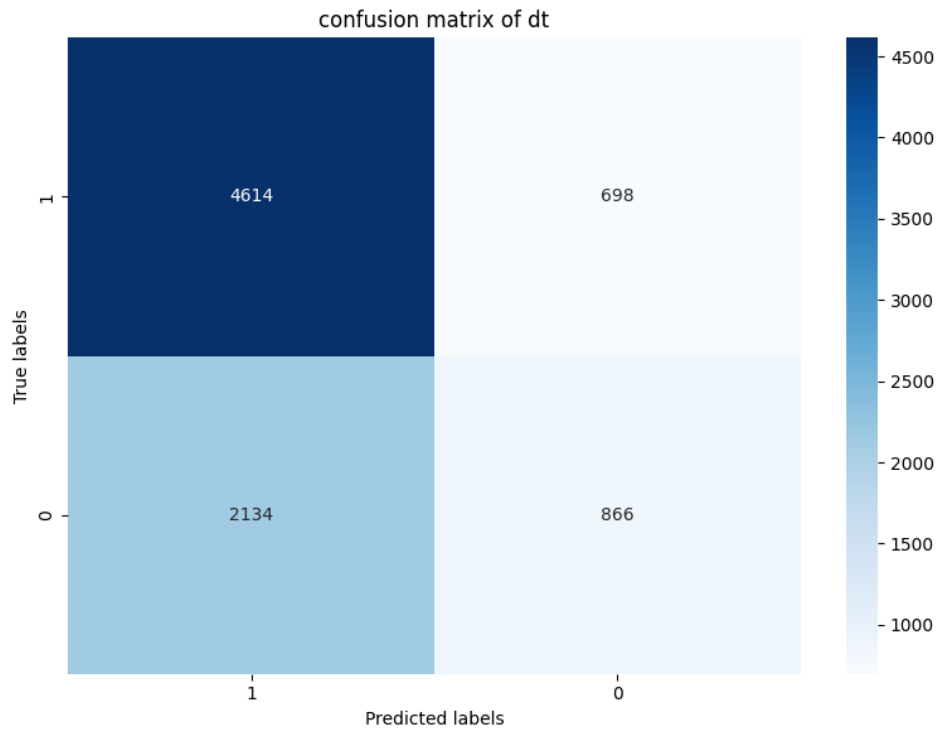


Figure (18) Confusion Matrix of dt

This model's confusion matrix allows us to understand how values change throughout the prediction process, ultimately providing us with accuracy, sensitivity, and specificity.

To gain a better understanding of these, we construct a bar graph using the metrics list.

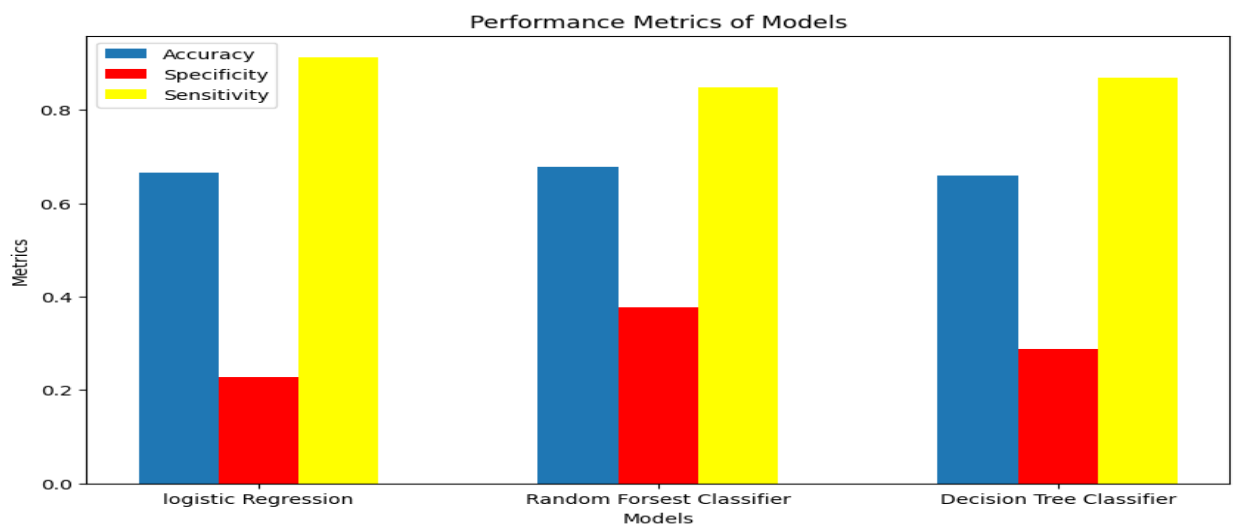


Figure (19) Performance Metrics of Models

We can infer from the bar graph that the Random Forest classifier provides our area of interest. The random forest classifier has higher specificity, which indicates that it is providing us with more reliable positive values in the prediction. In contrast, logistic regression has higher sensitivity, which provides more real negative values. In these models, randomness may also be impacted.

### 7.3 Churn

The rate at which the quantity of items available decreases relative to the total number of items is known as churn. Every quarter, we will examine the average amount of quantity that is declining for each category of item. This suggests how much of each item category could be reordered every three months. Using the logistic regression model as a regressor, we must take the feature variables and use the churn variable column to get the average churn. We tested the model for fitting and obtained an accuracy of 0.91. Plotting of the average quarterly turnover for each item category is done. Every item type has extremely low average outgoing orders, although novelty items could easily run out of stock because they are single items. Because of this, the plot has a high rate of churn. Poor customer orders or overstocking for future uses could result in higher holding costs and more challenges for inventory storage space, which is the reason why the things aren't being sent or sold.

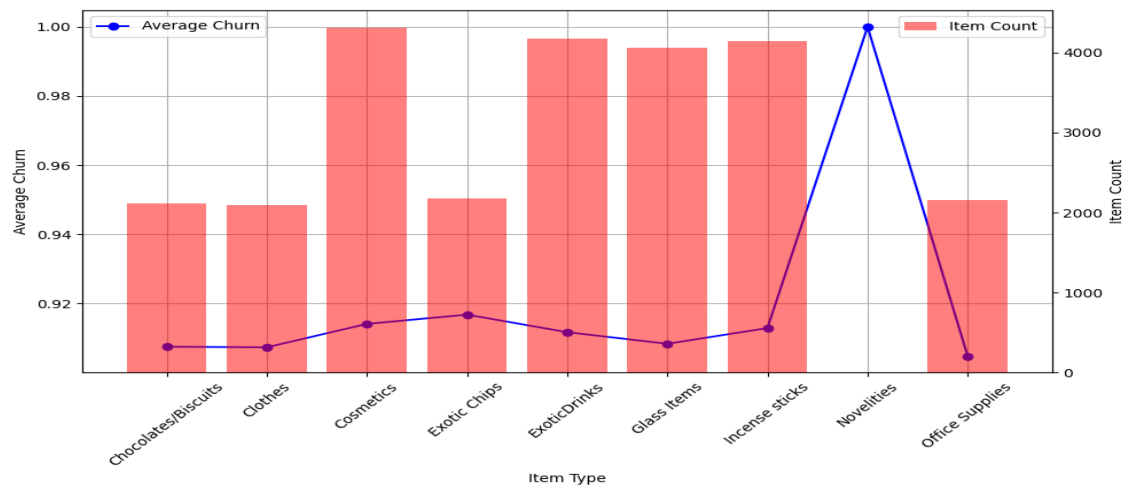
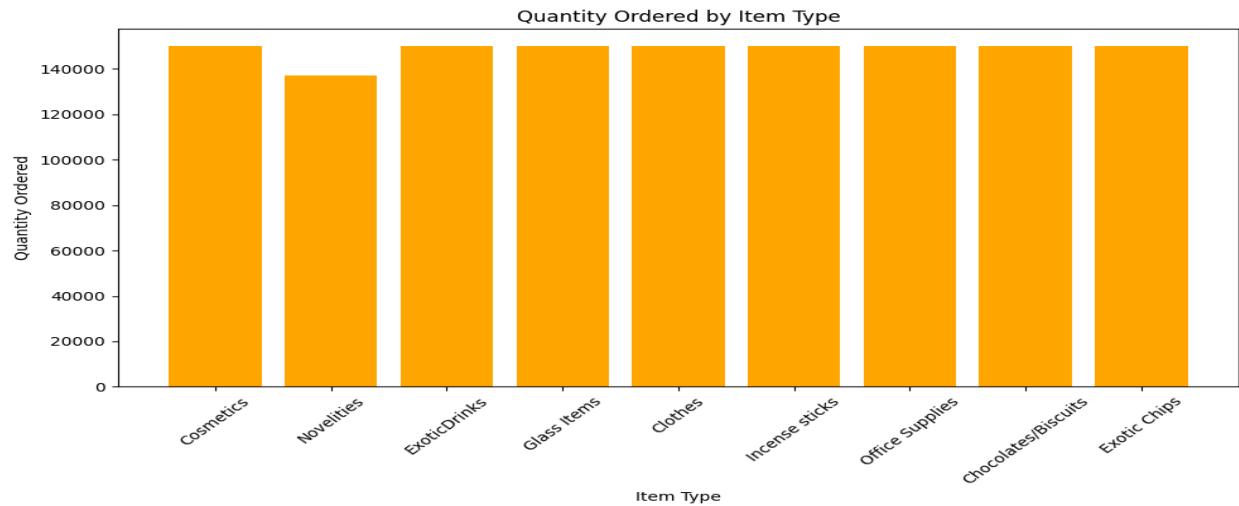


Figure (20) Average Churn Rate of Each Item Type for every quarter

Every item type has extremely low average outgoing orders, although novelty items could easily run out of stock because they are single items. Because of this, the plot has a high rate of churn. Poor customer orders or overstocking for future uses might result in holding costs and more challenges for inventory storage space which is the reason why the things aren't being sent or sold.

## 7.4 Correlation between the Quantities ordered and item type.

The quantity ordered for each item category is shown in the graphic below. We can conclude that, with the exception of novelties, every other item class has a nearly same ordered amount with just slight variations. This could prove that every kind of item is evenly represented in the inventory. The sales and profitability of each item category determine how much is ordered.



## 8. RESULTS

The analytical abilities and prediction models to a complicated dataset for the patterns and insights in our research by addressing our descriptive and predictive questions. Statistical analysis was used in this expedition to learn more about the data set, data visualization was used to bring some insight patterns and further progress on variables and ARIMA models were used to highlight challenges in precision demand forecasting and pave the way for insights into seasonal trends and possible external shocks on inventory. This model's RMSE, p-value and ACF scores indicate that the actual demand is not being met and suggest that demand has been lost. The categorization methods used to increase our understanding of customer satisfaction are logistic regression, decision trees and random forest variables. Every model is fully fitted with random forest being the most accurate and these classification models with key methods and procedures started to expose the relationship between feature variables and satisfaction ratings. However, correlation analysis and its heat map revealed the association between the variables like "cost price" and "total money on hold" showing a strong correlation. Furthermore, we conducted a churn analysis that revealed the average decline in rating for each item over the course of a year as well as the absurd decline in novelties and increased awareness of their limited supply.

## **Descriptive Answers:**

- 1. What is the most available item type in the warehouses? Which item type has the least stock in warehouses?**
  - From figure-7, Cosmetics is the most available item type to customers crossing 600 million. Novelties is least item type which is very less compared to others.
- 2. How quantities are ordered, distributed across the different warehouses?**
  - From figure-4, Orders are equally distributed across all four warehouses maintaining inventory balance with 25% each for all four warehouses which can be seen in figure4.
- 3. What is the annual trend of quantities ordered for each item type?**
  - From figure-5, Cosmetics, incense sticks, and exotic drinks consistently show higher orders annually, reflecting their popularity.
- 4. What patterns seen in the relationship between quantities in and out of the inventory?**
  - From figure-6, the quantity coming into inventory is consistently higher than the quantity going out, potentially leading to overstock and increased holding costs.
- 5. Which item type showed the highest profit percentage in any given year and what might be the reason?**
  - From figure-8, Clothes had the highest profit percentage in 2021, likely due to increased demand post-COVID-19.
- 6. What is the overall average rating of inventory items and how do ratings differ across item types? How do packaging, delivery and offline ratings vary for different product categories?**
  - From figure-10, most inventory items have an average rating around 3.5, with offline ratings typically outperforming online ratings. Delivery ratings are generally higher for cosmetics and glass items, while exotic chips and drinks score higher for offline ratings.
- 7. How cost price and total money on hold correlated and what do you find?**
  - From figure-11, There is a strong positive correlation (0.98), indicating that higher cost prices are associated with higher amounts of money on hold.

## **Predictive Answers:**

- 1. Can the future stock requirements be predicted using ARIMA models? What impact showed by changing market trends on future stock storages?**
  - Yes, with the help of from Figure-12 previous trends from "Available for Sale" and "Order Date" variables, ARIMA model provide reliable predictions of future stock needs. The stationary nature of the demand data and autocorrelation patterns suggest strong model accuracy for forecasting. Market trends such as increased demand for clothes in 2021 due to post-pandemic recovery demonstrate that external shocks impact the storages.



**2. Can customer satisfaction be predicted scores based on packaging, delivery and availability ratings? Can customer ratings predict the repeated purchases?**

- Customer satisfaction scores can be accurately predicted by using classification models like Random Forest and Logistic Regression. These models achieved approximately 70% accuracy showing the importance of rating variables on satisfaction.
- Yes, feature-engineered customer ratings (offline, online, packaging, delivery) combined with classification models provide insights into repeat purchase behaviour. High offline ratings are strongly predictive for finding loyalty customers.

**3. What are the likely trends in item churn rates for different product categories?**

- From Figure-20, Churn analysis shows that novelties have the highest churn due to limited stock and lower demand. Logistic regression predicts these trends quarterly with high accuracy (91%).

**4. Can seasonal trends help forecast peak demand periods for specific item types?**

- Seasonal trends are observed in profit percentages and sales volumes for clothing sales in 2021 indicating that certain items experience predictable highs. ARIMA forecasts can model these seasonal variations effectively.

**5. How do changes in cost prices influence the forecasted selling prices over time?**

- The strong correlation between cost prices and selling prices ( $r = 0.72$ ) indicates that changes in cost prices directly affect future pricing strategies. Predictive models are used for these relationships to estimate selling price trends.

**Findings and Managerial Implications:**

**Findings**

1. Item Stock Optimization
2. Demand Forecasting by predictive analysis for stock
3. Product variety and Inventory Holdings
4. Customer Satisfaction
5. Operations Management

**Managerial Implications**

1. Strategic Inventory Planning
2. Technology Integration
3. Enhanced Customer Experience
4. Category-Specific Strategies
5. Sustainability and Cost Management

## 9. CONCLUSION

A variety of data science methods including classification models like logistic regression, random forest, decision trees and linear regression were used to analyse the inventory management information. Predictive time series analysis was used to forecast product demand using past data. It was a satisfying exercise to apply these many strategies to real-time world data projects because all of these methods were used for research reasons and the outcomes were trustworthy. Other variables engaged in inventory operations within the data were used to aid the classification techniques identify the accuracy of customer satisfaction. In order to comprehend changes over time for predicted demand, the ARIMA model primarily offered insights into the data's time series pattern.

Despite the advancements, there is still much more room for investigation into the integration of outside variables like transportation, the financial crisis and natural disasters like COVID.

## 10. REFERENCES

- 1) **Liu, J., Li, Z., Sun, L., Wang, J., & Cao, N. (2018).** Research on the application of RFID technology in logistics warehousing management system. *Algorithms and Architectures for Parallel Processing*. Volume 11338, Springer, pp. 235-245
- 2) **Puica, E. (2023).** Improving supply chain management by integrating RFID with IoT shared database: Proposing a system architecture. *\*IFIP Advances in Informat*
- 3) **Ali, I., & Phan, H. M. (2022).** Industry 4.0 technologies and sustainable warehousing: A systematic literature review and future research agenda. *The International Journal of Logistics Management*, 33, 644–662.
- 4) **Cheng, S. W., Choy, K. L., & Lam, H. Y. (2015).** A work low decision support system for achieving customer satisfaction in warehouses serving machinery industry. *IFAC-PapersOnLine*, 48, 1714–1719.