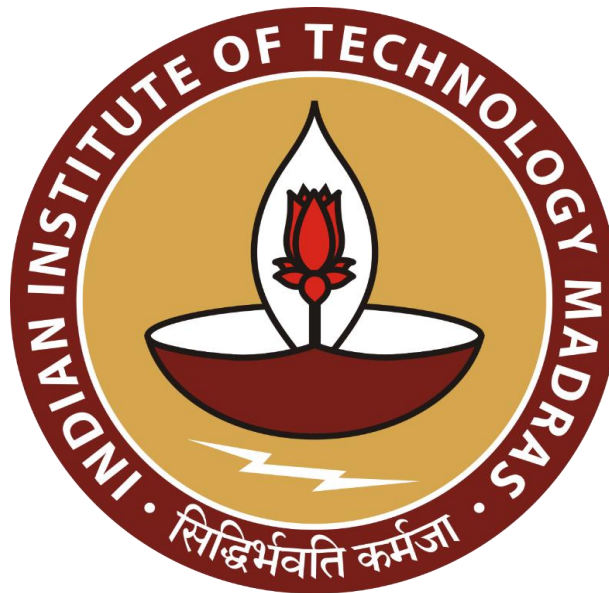# Improving Inventory Management and Customer Engagement Using Data Analysis for Big Basket

A Final report for the BDM capstone Project

Submitted by:

Arjita Singh Kushwaha

21f3000507

IITM Online BS Degree Program,

Indian Institute of Technology,

Madras, Chennai Tamil Nadu, India,

600036

Contents

1. Executive Summary

Big Basket is one of the India's online grocery platform which offers the wide range of products and serving millions of customers across the country. The company's extensive inventory and diverse customer base bongs both challenges and opportunities in order to maintain the optimal stock levels and ensuring the customer satisfaction.

This report is about the analysis done for Big Basket to address the key challenges identified related to inventory management and customer engagement. The objective is to find insights from data and able to improve the efficiency of business and optimize the levels of stocks and enhance the customer engagement. The analysis has been done including the several steps starting from data cleaning and prepossessing following by the extensive exploratory data analysis which is performed using various visualization techniques such as bar charts, line graphs etc. to identify the trends and product demands.

This report outlines the practical recommendation for the business to improve its inventory management, reduce overstocking and wastage and enhance the customer engagement. By implementing the suggested data driven strategies big basket can able to improve operational efficiency and drive the greater customer satisfaction and ultimately positioning itself for sustained growth in the competitive online grocery market.

2. Proof of Originality

The data used in this project was taken from Kaggle, which is a well-known platform provides datasets for analysis purpose.

The dataset includes customer purchase history which helps in understanding inventory issues and customer buying behaviourism from Big Basket.

The dataset can be accessed at:

https://www.kaggle.com/datasets/validmodel/bigbasket-customer-analytics

The data was only used for analysis and no changed was made except basic cleaning and preprocessing to make it suitable for this capstone project study and to find the valuable insights.

Colab link: https://colab.research.google.com/drive/1hDenPxHF7Jd-9Ma-q-YYks-bq0HupYbm?usp=sharing

3. Meta Data and Descriptive Statistics

3.1. Meta data

The detailed summary of the dataset which includes the key attributes, statistical insights and categorical distribution is described below. To understand these essential attributes and elements are necessary to analyze the inventory trends and customer purchasing patterns along with the product demand fluctuations

Overview of dataset:

The dataset used in this project contains the details of customer orders from the Big Basket and it includes the important information about the products, customers and order history.

The dataset contains total 62141 records and the fields in the data-set includes:

Member ID	A unique ID for each customer who placed an order.	Object
Order Number	A unique number assigned to each order.	Integer
SKU (Stock Keeping Unit)	A unique code for each product to track inventory.	Integer
Created On	A unique code for each product to track inventory.	Object
Description	A Textual representation of the product purchased.	Object

Table 3.1.1 : Data Set Description

This data will help in understanding the customer purchase patterns, identify the popular products among them and help in analyzing the inventory trends.

The dataset is well structured and has no missing values that ensures the data is complete and consistent for analysis.

3.2. Descriptive Statistics

To get an overview of the data, basic statistical analysis was done. Here are some key findings:

Total Unique Orders	8387
Total Unique Customers	106
Total Products Sold	1732
Most Ordered Product	Other Vegetables
Month with Highest Sales	September
Month with Lowest Sales	November
Average Orders per Customer	79

Table 3.2.1: Descriptive Stastics

Summary of numerical data

Coulmn	Order	SKU
Mean	7,642,313	17,743,230
Min	6,422,558	6,884,195
Standard Deviation	513,112	14,424,770
25%	7,457,967	15,668,380
50%	7,725,501	15,668,520
75%	8,006,749	15,669,870
Max	8,388,492	93,319,500

Table 3.2.2 : Numerical data Summary

Order distribution: the number of orders are sequential having the relatively small standard deviation which shows the consistent volume of transaction over time.

SKU variability: the values of SKU has a significant range having some products are being solve with higher number than others which indicated the difference in demand across the categories of products

The key observation includes:

The average SKU number is quite high which indicates that the dataset includes a mix of both high-volume and special products.

The order numbers follow a stable pattern and reflecting regular purchasing behavior across different customers.

The range in SKU values suggests that some products may be more frequently purchased than others leads to potential inventory imbalances.

Categorical data distribution

Customer buying behaviour: The most frequent customer having the member id M38622 has placed 1438 orders which shows that only few customers have a significant contribution in total sales.

There are total 106 unique customers.

Product demand distribution: The dataset contains 216 unique products which includes the variety od products available.

The most frequent purchased product is "Vegetables" which has been purchased 4604 times by the customers followed by other grocery essentials.

Order frequency trends: The dataset has total 8352 unique timestamps that means certain time periods have higher volume of orders

4. Detailed Explanation of Analysis Process/Method

To analyze the big basket's problems identified, I followed a structured approach using the data analysis technique to extract the meaningful insights

4.1. Data Cleaning and Preprocessing

First, I loaded the dataset and then checked for missing values or any incorrect values. I found out that there is no missing values in the dataset. Then Data types was standardized for consistency. This step is crucial to prepare the data for accurate analysis to be performed.

I format the data to make it more understandable and to drive the meaningful insights from it. The order dates has been converted into standardized date format.

4.2. Exploratory Data Analysis(EDA)

Once the data was cleaned and checked for any missing or incorrect in order values,then through data visualization techniques. EDA plays a crucial role in order to uncover the patterns, trends and valuable insights. I analyzed the different aspects of the business and find some insights from it and multiple visualization has been employed to highlight important insights, they are follows:

A. **Sales trends:** I plotted the graphs using the python library matplotlib to find patterns in sales trends and to see how sales fluctuated across the different seasons, months and days. From this I was able to identify the peak and low demand periods. Additionally, seasonal trend indicates the higher demand for certain product during the specific months.

B. **Product performance Analysis:** I analyzed that which products were selling well and which ones were under-performing. this helped me in identifying the low performing products and I was able to highlight the need for improving the inventory management.

This analysis show that some products have consistently high sales while others remained underutilized.

C. **Customer purchase behaviour:** I looked at how often customers purchasing the same products and weather they are exploring the new options or not. I found out that many customers are repeatedly purchasing the same set of products which is indication of the limited engagement with new or alternative product options.

D. **Correlation analysis:** to explore the relationship between numerical feature, a correlation heatmap has been created. This visual analysis revealed minimal correlation between SKU codes and order volume. The weak correlation suggested that SKU codes alone were insufficient in predicting sales trends. This finding highlighted the need to incorporate additional features such as seasonality, customer preferences, and product categories when forecasting demand.

4.3. Identifying Inventory Gaps and Customer Engagement Issues

After performing the EDA and correlation analysis i was able to identify the critical gaps in inventory planning and customer engagement strategies:

A. **Inventory gaps**

The weak correlation between SKU codes and sales indicated that inventory decisions were not effectively aligned with customer demand.

Seasonal peaks in December further highlighted the need for dynamic stock management to ensure sufficient supply during high-demand periods.

Additionally, the presence of numerous low-frequency purchases indicated potential stock misalignment, where less popular items might be taking up valuable inventory space.

B. **Customer Engagement Issues**

The histogram showing purchase frequency revealed a substantial group of customers with minimal engagement.

While repeat buyers represented a loyal customer base, the lack of engagement from certain segments suggested missed opportunities in promoting complementary or lesser-known products.

5. Result and Findings

After analyzing the data, I find out the several insights which is highlighting the trends in customer behaviour, seasonal fluctuation and product demands.

5.1. Distribution of order per customer

The data revealed a significant variation in customer buying pattern. Some customers placed frequent orders while there is also a substantial portion which had minimal engagement. This uneven distribution indicates that a smaller group of loyal customers contributed substantially to overall sales.
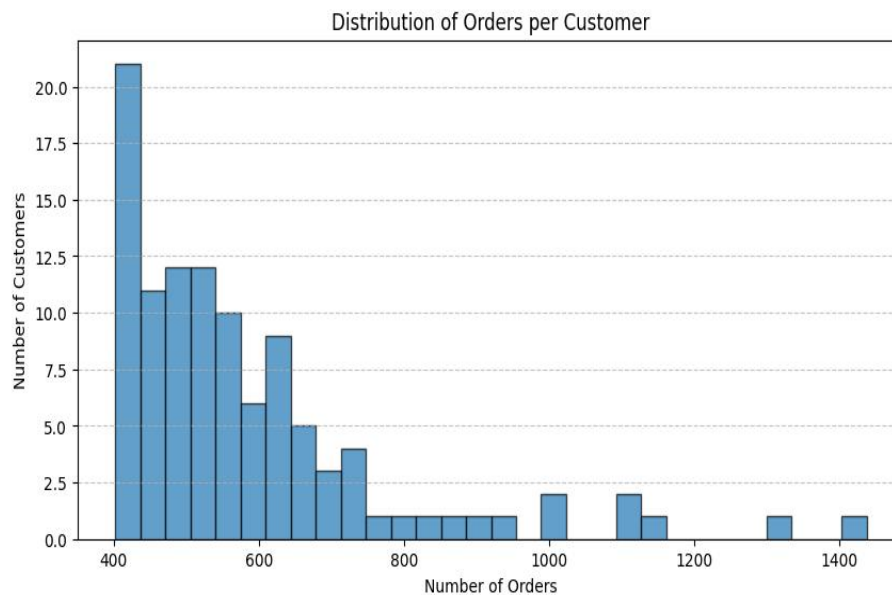

Fig 5.1: Distribution of orders per customers

Insight: The presence of infrequent shoppers highlights an opportunity to target this segment with personalized offers, discounts, or loyalty programs to encourage more frequent purchases.

5.2. Top 10 most ordered products

The most popular product categories included Other Vegetables, Fruits, and Dairy Products. These essential grocery items were consistently in high demand, reinforcing their importance in daily household needs. The dominance of staple food items suggests that maintaining steady stock levels for these categories is crucial to meeting customer expectations.
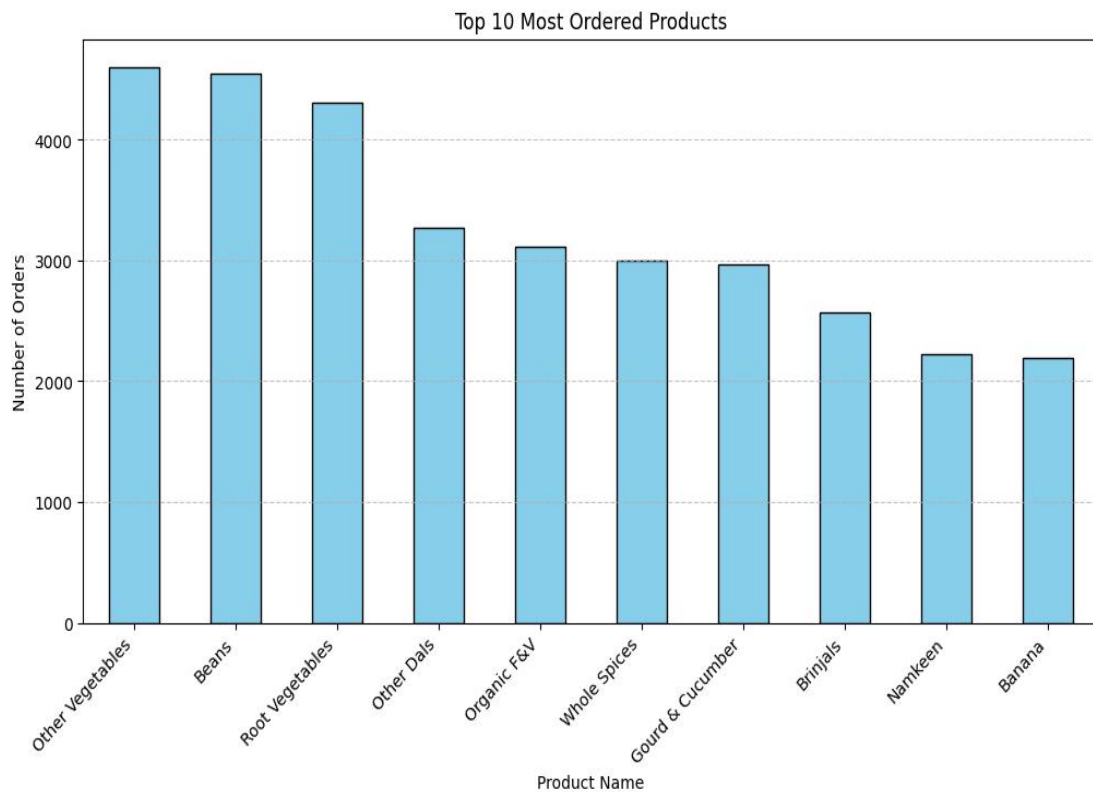
Fig 5.2: top 10 most ordered products

Insight: The dominance of staple food items indicates that maintaining steady stock levels for these essential products is crucial to meeting customer expectations and ensuring consistent sales.

5.3. Monthly Order Trends

An analysis of order trends over time showed that sales peaked in September, suggesting heightened demand during the festive season. Meanwhile, November experienced the lowest sales volume, indicating a post-holiday decline in consumer spending.
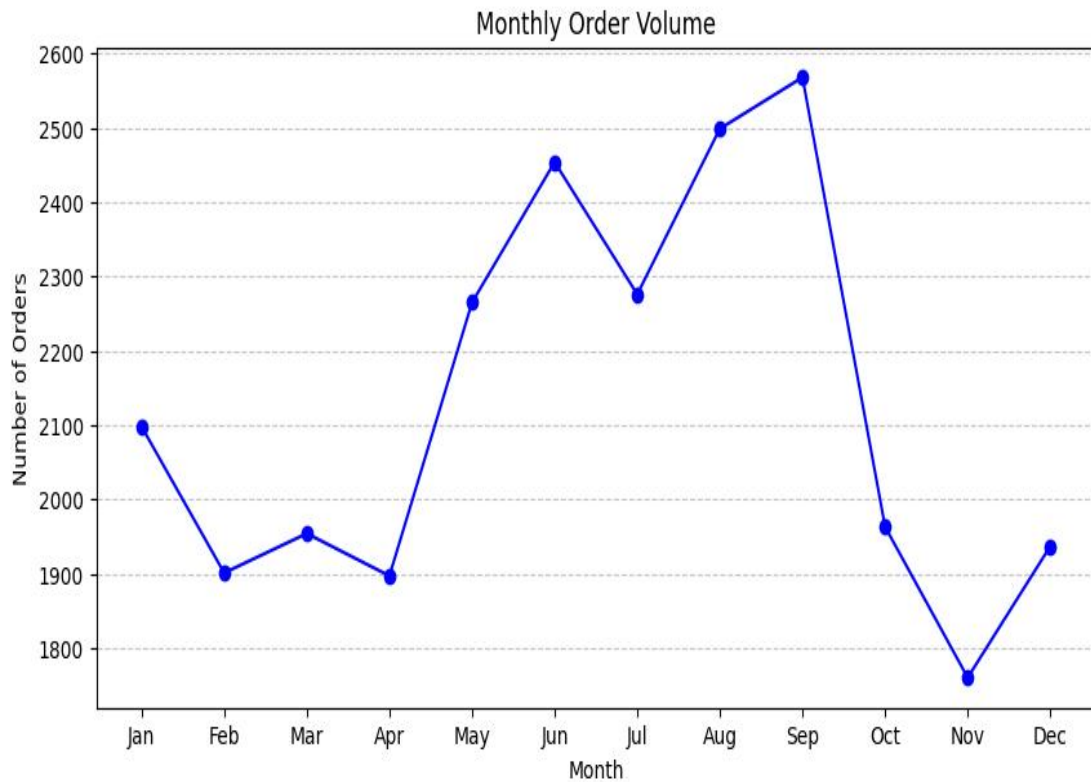
Fig 5.3: Montly order volume

Insight: This seasonal pattern underscores the importance of proactive stock management. Increasing inventory levels for popular items during festive months can help prevent stockouts, while adjusting stock levels in quieter months can reduce excess inventory.

5.4. Customer Repurchase Behavior Analysis

To understand customer purchasing patterns, a distribution plot was created to visualize how frequently customers repurchase the same products. The graph provides insights into customer loyalty and the tendency to stick with familiar items.
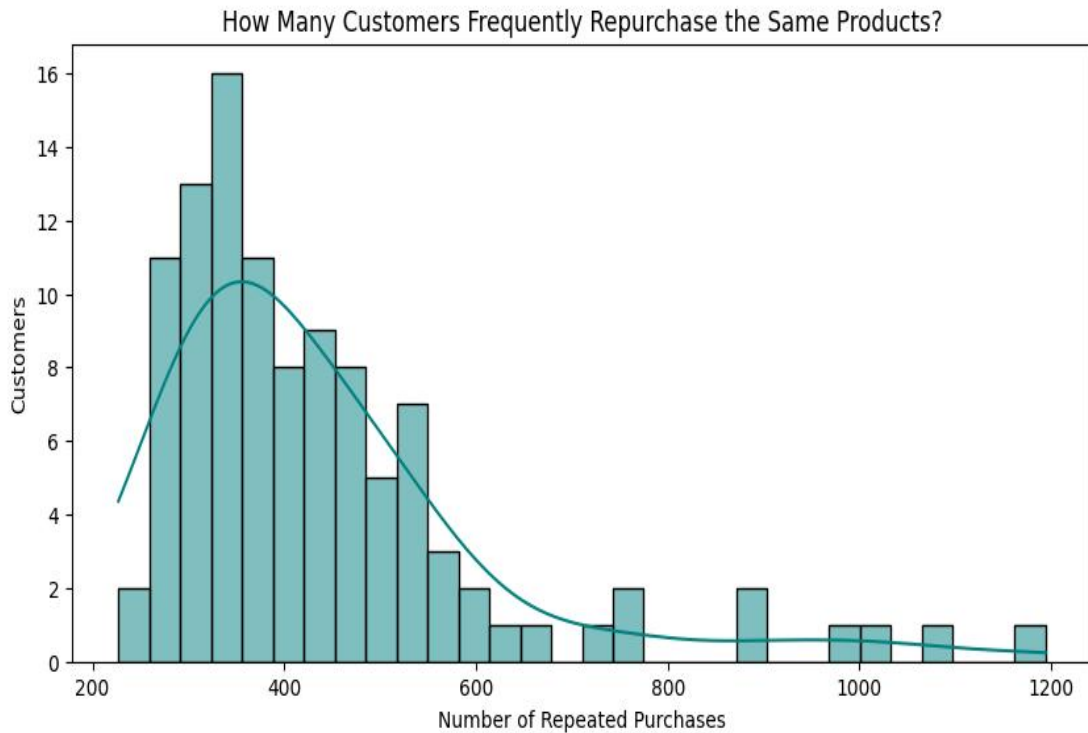
Fig 5.4: Customer Repurchase Behavior Analysis

Insight: The concentration of repeat purchases around 300 indicates that Big Basket's customers rely heavily on staple products, such as groceries and daily essentials. Customers with fewer repeat purchases may indicate potential churn risk, presenting an opportunity for targeted marketing campaigns to improve retention.

5.5. Distribution of Unique Products Bought Per Customer

The data reveals a noticeable variation in the number of unique products purchased by customers. While a majority of customers buy between 90 to 130 unique products, there are also segments that purchase significantly fewer or more items. This pattern suggests that some customers prefer a focused selection of products, while others explore a broader range.
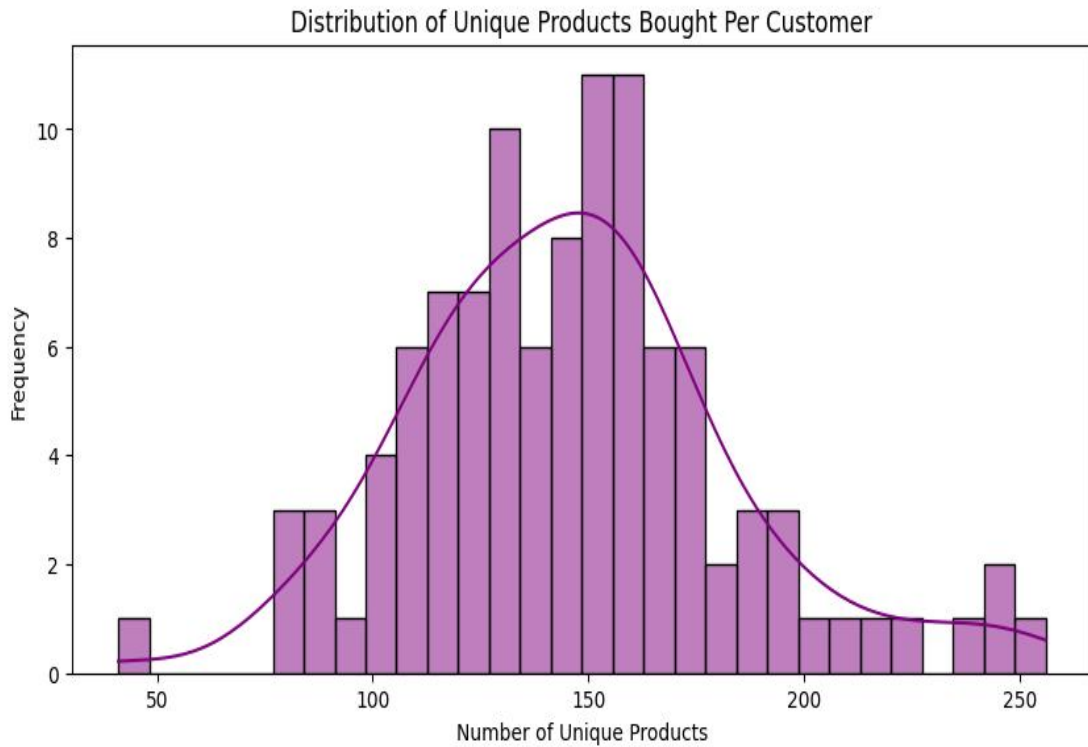
Fig 5.5: Distribution of Unique Products Bought Per Customer

Insight: The presence of customers purchasing fewer unique products presents an opportunity to promote complementary items or personalized recommendations to encourage greater product exploration. Similarly, frequent buyers with diverse shopping habits can be targeted with bundle offers or exclusive deals to enhance loyalty and maximize sales.

5.6. Orders Distribution by Time of Day

The analysis of order distribution across different times of the day reveals distinct customer behavior patterns. The data shows that most orders are placed in the morning, followed by the night, with the afternoon having the least number of orders.

Fig 5.6: Orders Distribution by Time of Day

Insight: The higher order volume in the morning suggests a key opportunity to promote breakfast items, fresh produce, and essential groceries. The dip in orders during the afternoon indicates potential for targeted campaigns, such as discounts or flash sales, to boost engagement during this period. The notable number of orders at night highlights an opportunity to promote convenience products or next-day delivery services to meet customer preferences.

5.7. Correlation Analysis

The heatmap visualizes the correlation between numerical features such as Order and SKU. The correlation values range from -1 (strong negative correlation) to 1 (strong positive correlation), with 0 indicating no correlation.
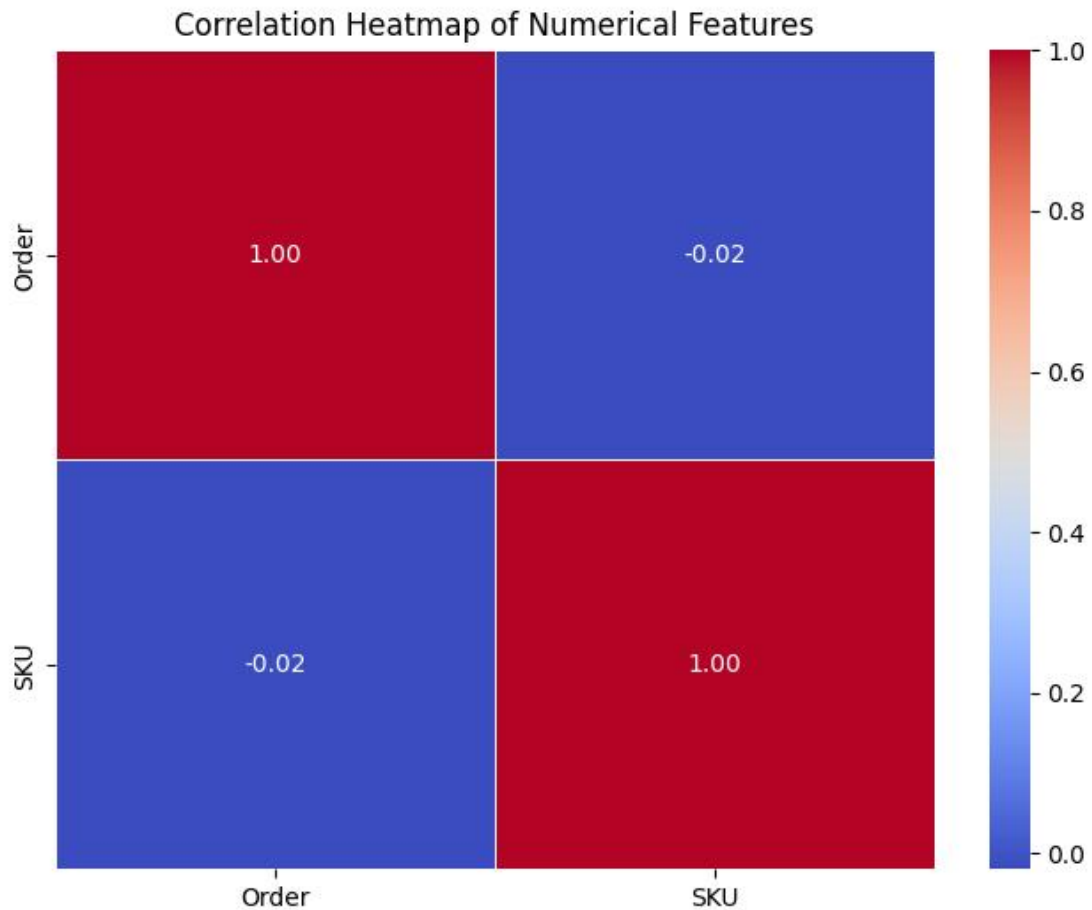
Fig 5.7: Heatmap numerical features

Insight: The weak correlation suggested that SKU codes alone were insufficient in predicting sales trends. This finding highlighted the need to incorporate additional features such as seasonality, customer preferences, and product categories when forecasting demand.

6. Interpretation of Results and Recommendations

On the basis of the analysis I have done, I was able to identify the key challenges in inventory management and customer engagement which is impacting the big basket's overall efficiency and profitability. The interpretation of the insights carried out along with recommendation to address these issues are listed below:

6.1 Inventory management challenges

● There is an significant stock imbalances that means some products are overstocked while frequency of other products which are essential go out of stock.

● Seasonal demand patterns were not properly accounted for which leads to wastage and missed sales opportunities.

● The fast moving essential items were not restocked on time which is causing the customer dissatisfaction.

Interpretation:

From all the above point the interpretation is made that the inventory system is not currently aligned with real-time purchase patterns which is leading to the inefficiency in stock allocation and holding the unnecessary costs.

6.2 Customer engagement

● There is a large percentage of customers who are purchasing the same products repeatedly without exploring or discovering the new products.

● There is also a noticeable drop in the sales during the off-peak days for example early weekdays due to the lack of promotional efforts.

● There is a missing opportunities of personalized product recommendations which is limiting the cross selling and upselling opportunities.

Interpretation:

The interpretation made from above insights is that customers are not encouraged to find or explore the new items because of lack of personalized engagement strategies which leads to missed opportunities of revenue.

6.3 Understanding Customer Ordering Patterns

● There is the distinct patterns in customer activity. The data shows that majority of orders are placed during the morning and evening hours having fewer order occurring in the afternoon time.

● While observing monthly trends I find out that customer orders quantity significantly increased during September. In contrast, there is a noticeable dip in order volumes during October and November.

Interpretation:

This seasonal variation suggests that Big Basket could implement strategic inventory adjustments to prepare for increased demand before September and consider targeted promotions or festive offers to encourage sales during the quieter months.

6.4 Exploring Customer Purchase Behavior

● Customer purchase data reveals interesting patterns regarding order frequency and product preferences. Most customers place between 400 to 600 orders has only a small segment exceeding 1,000 orders.

● Analysis indicates that repeat purchases are common with many customers repeatedly buying between 250 to 400 products.

Interpretation:

This shows that Big Basket has a dedicated group of high frequency shoppers who consistently engage with the platform.

Implementing automated restocking mechanisms for frequently purchased products can help minimize the stockout situations and improve order fulfillment efficiency.

6.5 Understanding Product Demand Trends

● The product analysis highlights that certain product such as 'Other Vegetables', 'Beans' and 'Root Vegetables' consistently rank among the most popular items.

● Many customers purchase between 100 to 130 unique products, relatively few explore a broader variety of options.

Interpretation:

This shows that while shoppers often stick to familiar products, there is untapped potential to introduce them to complementary or alternative items. By implementing personalized product recommendations based on individual purchase history, Big Basket can encourage product discovery, ultimately boosting sales and enhancing the shopping experience.

6.6 Recommendations

Based on the analysis performed and insights gathered the following recommendations can help Big Basket address its challenges and enhance business performance:

6.6.1 Optimizing Inventory Management

- **Forecasting Demand with Precision:** Leverage historical sales trends and seasonal patterns to build predictive models. For example, increase stock for dairy and produce in September and November to meet peak demand.

- **Dynamic Stock Replenishment:** Introduce dynamic inventory systems to automate restocking alerts for fast-moving items and flag slow-moving products.

- **Managing Surplus Inventory Through Bundling:** Combine slow-moving products (e.g., niche cooking sauces) with high-frequency items to encourage exploration and reduce stagnant stock.

6.6.2 Enhancing Customer Engagement

- **Personalized Recommendations:** Use customer purchase history to recommend complementary items. For instance, a buyer of pulses and spices might be suggested oils, condiments, or ready-to-cook kits.

- **Targeted Promotions and Discounts:** To improve the sales of less popular items, Big Basket can introduce targeted marketing campaigns. These campaigns can highlight slow-moving items as part of special offers, combo deals or discounts.

- **Loyalty Programs and Incentives:** Implementing a customer loyalty program can encourage repeat purchases. Rewarding frequent buyers with discounts, points or exclusive offers can significantly improve customer retention.

6.6.3 Sales Planning and Seasonal Adjustments

- **Seasonal Sales Forecasting:**Big Basket can align marketing strategies according to festive demands. For example, launching promotional campaigns in November can mitigate the sales dip, while additional stocking and offers during September can capitalize on high demand.

- **Flexible Pricing Strategies:** Introducing dynamic pricing models can help maximize sales during peak seasons.

6.6.4 Data-Driven Insights for Business Growth

- **Advanced Analytics for Demand Prediction:** Implementing machine learning models can enhance demand forecasting accuracy by analyzing variables such as seasonality, customer preferences and historical purchase patterns.

- **Customer Feedback Analysis:** Collecting and analyzing customer reviews can reveal valuable insights into product preferences, customer dissatisfaction or emerging trends. For example, feedback highlighting poor product quality or missing items can prompt Big Basket to adjust supplier partnerships or enhance quality checks.