

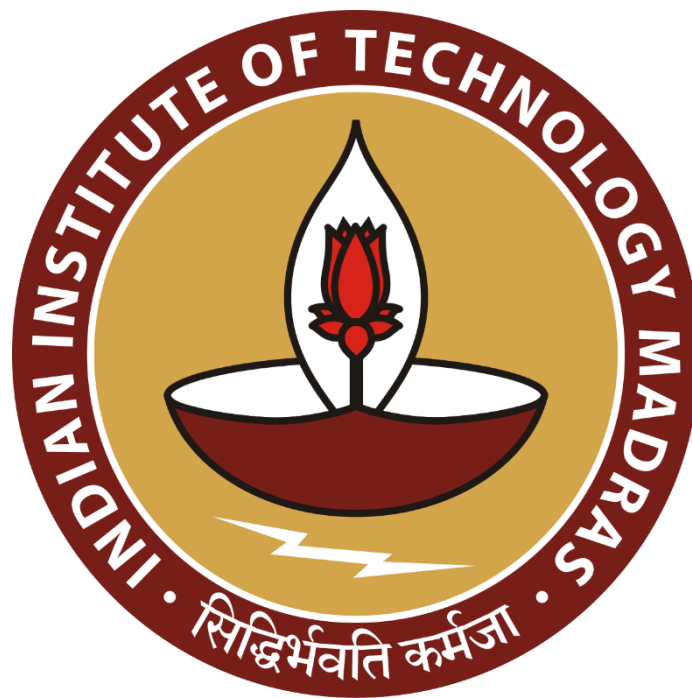
Inventory Management and Sales Forecasting of Walmart Stores

An End-Term report for the BDM capstone Project

Submitted by

Name: Shriya Singh

Roll number: 24f2008693



IITM Online BS Degree Program,
Indian Institute of Technology, Madras, Chennai
Tamil Nadu, India, 600036

Contents

| | | |
|---|--|----|
| 1 | Executive Summary | 3 |
| 2 | Proof of Originality | 4 |
| 3 | Meta Data and Descriptive statistics | 5 |
| 4 | Detailed Explanation for Analysis Process/Method | 8 |
| 5 | Results and Findings | 11 |
| 6 | Interpretation of Results and Recommendations | 20 |

1 Executive Summary

“Walmart Inc.”, an American multinational retail company that operates a chain of hypermarkets and departmental stores in the US and 23 other countries. The business primarily operates in B2C providing daily household items to customers. The business struggles with issues of proper management of inventory causing dissatisfaction in customers, wastage of product and increased storage, which is tackled by forecasting sales of stores to reduce stockouts and overstocking.

The data was collected from Github and Kaggle websites which are in csv format ranging from February 2010 to October 2012. These datasets were sourced from publicly available repositories, curated to include relevant variables, cleaned for consistency, and formatted to facilitate data analysis. The issues are resolved by predicting the sales for the upcoming weeks and keeping the required stocks in the inventory to fulfill the demand and supply of customers. A detailed analysis of the entire data is done using numpy and pandas, pivot tables for clear understanding of data and summarization of sales and finances, data visualization and interpretation of results using matplotlib and seaborn and forecasting the sales using Time-Series model.

A thorough analysis of sales, revenue, and efficient store operations is carried out. Sales forecasts indicate periodic seasonal spikes, particularly towards the year-end. The highest projected sales are observed in October 2012, likely influenced by seasonal trends or economic conditions. Observing the trends and seasonality patterns in customer behaviour to track the shopping goods for proper management to reduce loss.

Elaborate evaluation of sales, revenue and proper management of the store is done by monitoring the trends of customer behaviour, which items are sold more at which duration, look for product demands which increase and decrease at what interval. Through targeted marketing, focusing promotions based on gender preferences, analyzing gender imbalance and adjusting pricing or promotions accordingly. This will help the organization to manage the inventory with respect to the demand of the customers and will seek to maximize the profits and minimize the loss faced by the store.

2 Proof of Originality

Details of repositories from where the data is collected are kaggle and github. The dataset drive link is present along with the python code in google colab notebook.

WEBSITE- Kaggle, Github

DATASET SOURCE LINK-

<https://www.kaggle.com/datasets/shriyasingh900/walmart-dataset>

[https://github.com/manasik411/WalmartSalesAnalysis/blob/main/WalmartSalesData.csv%20\(1\).csv](https://github.com/manasik411/WalmartSalesAnalysis/blob/main/WalmartSalesData.csv%20(1).csv)

DATASET LINK-

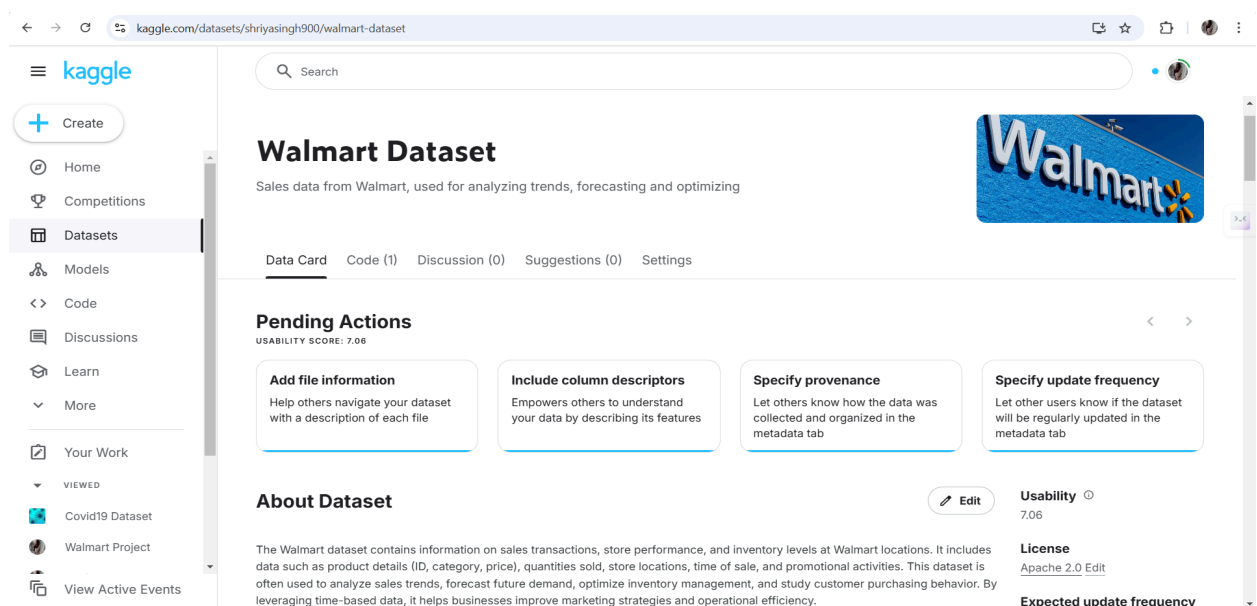
https://drive.google.com/file/d/1Vs7i5rTygdIhqEaXESTdBsD0i9hUiyQ3/view?usp=drive_link

https://drive.google.com/file/d/1UBZjtggjFiL_ReMehdmM0BRiEPImeJNi/view?usp=drive_link

COLAB NOTEBOOK-

<https://colab.research.google.com/drive/1B6fTs0GDzcND9gfZY1Nb0tO6Y1GUKhlR?usp=sharing>

REPOSITORY PROOF-



The screenshot shows the Kaggle dataset page for 'Walmart Dataset'. The page header includes the Kaggle logo and a search bar. The dataset title 'Walmart Dataset' is prominently displayed, followed by a description: 'Sales data from Walmart, used for analyzing trends, forecasting and optimizing'. Below this, there are tabs for 'Data Card', 'Code (1)', 'Discussion (0)', 'Suggestions (0)', and 'Settings'. The 'Data Card' tab is active, showing a 'Pending Actions' section with four tasks: 'Add file information', 'Include column descriptors', 'Specify provenance', and 'Specify update frequency'. Each task has a brief description of what it entails. Below the pending actions is an 'About Dataset' section, which provides a detailed description of the data, including product details, store performance, and inventory levels. The page also features a 'Usability Score' of 7.06 and a 'License' of Apache 2.0.

The screenshot shows a GitHub repository page for 'manasik411 / WalmartSalesAnalysis'. The file 'WalmartSalesData.csv (1).csv' is selected, and its preview is displayed. The preview shows a table with 15 columns: Invoice ID, Branch, City, Customer type, Gender, Product line, Unit price, Quantity, Tax 5%, Total, Date, Time, Payment, cogs, and gross margin percentage. The table contains 9 rows of data.

| | Invoice ID | Branch | City | Customer type | Gender | Product line | Unit price | Quantity | Tax 5% | Total | Date | Time | Payment | cogs | gross margin pe |
|---|-------------|--------|-----------|---------------|--------|------------------------|------------|----------|---------|----------|------------|----------|-------------|--------|-----------------|
| 1 | 750-67-8428 | A | Yangon | Member | Female | Health and beauty | 74.69 | 7 | 26.1415 | 548.9715 | 2019-01-05 | 13:08:00 | Ewallet | 522.83 | 4.761904762 |
| 2 | 226-31-3081 | C | Naypyitaw | Normal | Female | Electronic accessories | 15.28 | 5 | 3.82 | 80.22 | 2019-03-08 | 10:29:00 | Cash | 76.4 | 4.761904762 |
| 3 | 631-41-3108 | A | Yangon | Normal | Male | Home and lifestyle | 46.33 | 7 | 16.2155 | 340.5255 | 2019-03-03 | 13:23:00 | Credit card | 324.31 | 4.761904762 |
| 4 | 123-19-1176 | A | Yangon | Member | Male | Health and beauty | 58.22 | 8 | 23.288 | 489.048 | 2019-01-27 | 20:33:00 | Ewallet | 465.76 | 4.761904762 |
| 5 | 373-73-7910 | A | Yangon | Normal | Male | Sports and travel | 86.31 | 7 | 30.2085 | 634.3785 | 2019-02-08 | 10:37:00 | Ewallet | 604.17 | 4.761904762 |
| 6 | 699-14-3026 | C | Naypyitaw | Normal | Male | Electronic accessories | 85.39 | 7 | 29.8865 | 627.6165 | 2019-03-25 | 18:30:00 | Ewallet | 597.73 | 4.761904762 |
| 7 | 355-53-5943 | A | Yangon | Member | Female | Electronic accessories | 68.84 | 6 | 20.652 | 433.692 | 2019-02-25 | 14:36:00 | Ewallet | 413.04 | 4.761904762 |
| 8 | 315-22-5665 | C | Naypyitaw | Normal | Female | Home and lifestyle | 73.56 | 10 | 36.78 | 772.38 | 2019-02-24 | 11:38:00 | Ewallet | 735.6 | 4.761904762 |

3 Meta Data and Descriptive Statistics

Information about the Walmart Data

- Data Format: CSV (comma separated values)
- Range: February 2010- October 2012

Features of the Data 1

- Store- The serial number of the store as the data given is of 45 different stores
- Date- The particular sale is made on which date between Feb 2010 to Oct 2012, the data type of this column is datetime.
- Weekly Sales(\$)- The sales value of all the stores on a particular day of that week, which is in USD (\$) – United States Dollar and data type is float.
- Holiday Flag- If it was a holiday on that particular day or not, if it was then sales can increase accordingly and data type is integer.
- Temperature(Fahrenheit)- the temperature on that day which might affect the sales, the unit of measurement of temperature is Fahrenheit and data type is float.
- Fuel Price(\$/L)- Fuel price of that day which can influence the sales in a negative direction, the unit of measurement is USD per Liter and data type is float.
- CPI- Consumer Price Index is a measure of inflation of the economy, rising CPI indicates inflation, which affects sales by influencing consumer purchasing power.
- Unemployment- If the unemployment rate is high during that period of time, it reduces consumer purchasing power, leading to lower sales, data type is float.

| | Store | Date | Weekly_ Sales | Holiday _Flag | Tempe rature | Fuel_ Price | CPI | Unemploy ment |
|---|-------|------------|------------------|------------------|-----------------|----------------|--------|------------------|
| 0 | 1 | 05-02-2010 | 164369.9 | 0 | 42.31 | 2.572 | 211.09 | 8.106 |
| 1 | 1 | 12-02-2010 | 164195.7 | 1 | 38.51 | 2.548 | 211.24 | 8.106 |
| 2 | 1 | 19-02-2010 | 161196.7 | 0 | 39.93 | 2.514 | 211.28 | 8.106 |

Table 1 - Description of the dataset (6435 rows, 8 Feature columns)

Features of the Data 2

- Gender- Whether it's male or female buying the product, the gender imbalance can affect the sales of the product category and data type is string.
- Product line- the product belongs to which of the following categories and data type is string.
- Unit price(\$)- the price of the individual product of each category in USD, float dtype.
- Quantity- how many quantities of each product did the customer buy, integer dtype.
- Tax- gst tax applied on the product. Increases the final price of goods, leading to lower consumer demand and data type is float.
- Total- total amount of the product which customer pays at the end of their shopping.
- Payment- mode of payment of the product, whether it is UPI, cash, debit or credit card, and data type is string.
- Cogs- Cost of goods sold, higher cogs- Lower profit margins unless prices are increased and data type is float.
- Gross margin- how much of the revenue remains after deducting cogs value and the data type is float.
- Percentage- represents how much of the revenue is spent on the cost of goods sold
- Gross income- total revenue left after subtracting the cogs, data type is float.
- Rating- customer rating for each product of the category (1-10) and data type is float.

| | gender | product line | unit price | quan tity | TAX | Total | Pay | cogs | gross mar. | gross Inc. | rating |
|---|--------|------------------|---------------|--------------|------|-------|------|-----------|---------------|---------------|--------|
| 0 | female | health beauty | 74.69 | 7 | 26.4 | 548.9 | cash | 522. 8 | 4.761 | 26.14 | 9.1 |
| 1 | female | electron ic | 15.28 | 5 | 3.82 | 80.22 | upi | 76.4 0 | 4.761 | 3.820 | 9.6 |
| 2 | male | lifestyle | 46.33 | 7 | 16.2 | 340.5 | upi | 324 | 4.761 | 16.21 | 7.4 |

Table 2 - Description of the dataset (1000 rows, 11 Feature columns)

DESCRIPTIVE STATISTICS

FOR DATASET 1:

| | Store | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment |
|--------------|---------|--------------|--------------|-------------|------------|---------|--------------|
| Count | 5920.00 | 5.9203 | 5920.00 | 5920.00 | 5920.00 | 5920.00 | 5920.00 |
| Mean | 22.79 | 1.039 | 0.690 | 60.404 | 3.340 | 175.01 | 7.722 |
| STD | 13.092 | 5.528 | 0.253 | 18.427 | 0.458 | 39.029 | 1.243 |
| MIN | 1.00 | 2.099 | 0.00 | -2.069 | 2.472 | 126.062 | 4.308 |
| MAX | 45.00 | 2.685 | 1.00 | 100.14 | 4.461 | 227.237 | 10.926 |
| 25% | 11.00 | 5.526 | 0.00 | 46.966 | 2.891 | 132.766 | 6.891 |
| 50% | 22.00 | 9.469 | 0.00 | 62.615 | 3.145 | 190.068 | 7.852 |
| 75% | 34.00 | 1.427 | 0.00 | 74.724 | 3.728 | 213.799 | 8.494 |

Table 3

FOR DATASET 2:

| | Unit Price | Quantity | TAX | Total | COGS | Gross Margin %age | Gross Income | Rating |
|--------------|------------|----------|--------|--------|--------|-------------------|--------------|---------|
| Count | 1000.00 | 1000.00 | 1000.0 | 1000.0 | 1000.0 | 1.03 | 1000.00 | 1000.00 |
| Mean | 55.672 | 5.510 | 15.379 | 322.96 | 307.58 | 4.7611 | 15.379 | 6.972 |
| STD | 26.494 | 2.923 | 11.708 | 245.88 | 234.17 | 6.131 | 11.708 | 1.718 |
| MIN | 10.080 | 1.00 | 0.508 | 10.678 | 10.170 | 4.76 | 0.508 | 4.00 |
| MAX | 99.960 | 10.00 | 49.650 | 1042.6 | 993.00 | 4.76 | 49.650 | 10.00 |
| 25% | 32.875 | 3.00 | 5.924 | 124.42 | 118.49 | 4.76 | 5.924 | 5.500 |
| 50% | 55.230 | 5.00 | 10.088 | 253.84 | 241.76 | 4.76 | 12.088 | 7.00 |
| 75% | 99.960 | 8.00 | 22.445 | 471.35 | 448.90 | 4.76 | 22.445 | 10.00 |

Table 4

The descriptive statistics helps to summarize and interpret data for better decision-making, values like mean, count, standard deviation, minimum and maximum values of the columns. It helps in understanding the distribution, central tendency, and variability of the data further of all the feature columns of the dataset.

In Walmart data, these statistics assist in sales forecasting, inventory management, and pricing strategies.

- Mean helps to understand the typical sales volume, used for inventory planning to avoid stockouts and overstocking.
- IQR measures sales variability and plans for demand fluctuations. A high IQR indicates high demand fluctuations, requiring flexible inventory strategies.
- Standard deviation measures how much sales numbers fluctuate from the mean. Higher standard deviation indicates unpredictable sales, requiring advanced forecasting models.
- Range shows how much sales fluctuate over time. A large range indicates high variability, requiring dynamic inventory planning.
- Percentile helps Walmart identify high sales periods, useful for predicting peak demand and ensuring enough stock.
- Minimum and maximum values indicate slow sales periods and plan discount strategies and peak demand periods.

4 Detailed Explanation for Analysis Process/Method

- The analysis method of the project involves both qualitative and quantitative business analysis, each of them performed to address the particular aspect of the business problem. Using Microsoft Excel and its functions like pivot tables and graphs and charts for data summarization and visualizing, it provides for various necessary calculations for the analysis process. Tasks for calculating the total sales and revenue, to know for which particular date were effectively handled with excel functions.
- The analysis of the data which is collected through Kaggle has been done by the Time-Series analysis process. Since the data is thoroughly distributed over 2 years starting from February 2010 to October 2012 with weekly sales data, therefore the best way to analyze the data is by using Time-Series analysis.

The main steps for analyzing the data and following approach adopted is-

1. Exploratory data analysis
2. Statistical data analysis
3. Model building
4. Predictive analysis

By examining the trend patterns, seasonality and variations over the period of time, we get the statistical information and insights about the business performance. This model studies the trend patterns and seasonality through the past data of the store which is crucial for the model to learn the important parameters and give business decisions accurately. Python along with its data science libraries like numpy, pandas, matplotlib, seaborn were useful in exploratory data analysis and visualization of data.

The main objective of time-series analysis is to identify the sales of the stock based on the particular time period. The pandas and numpy library were useful for analysing and exploring the data, for outlier analysis, feature engineering and matplotlib and seaborn library were useful for representation and visualization of various features.

Exploratory Data Analysis

| Store | 01-10-2010 | 02-04-2010 | 02-07-2010 | 03-09-2010 | 03-12-2010 | 04-06-2010 | 05-02-2010 | 05-03-2010 | 05-11-2010 |
|-------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| 1 | 1453329.5 | 1594968.28 | 1492418.14 | 1540163.53 | 1548033.78 | 1615524.71 | 1643690.9 | 1554806.68 | 1551659.28 |
| 2 | 1827440.43 | 2066187.72 | 2003940.64 | 1904608.09 | 2015781.27 | 2102539.93 | 2136989.46 | 1991013.13 | 1939061.41 |
| 3 | 358784.1 | 423294.4 | 381151.72 | 366473.97 | 476420.77 | 396968.8 | 461622.22 | 415202.04 | 423175.56 |
| 4 | 1842821.02 | 1979247.12 | 1881337.21 | 1935857.58 | 2102530.17 | 1903290.58 | 2135143.87 | 1971057.44 | 2013115.79 |
| 5 | 283178.12 | 331406 | 305993.27 | 323798 | 344490.88 | 337825.89 | 317173.1 | 288855.71 | 325310.3 |
| 6 | 1328468.89 | 1770333.9 | 1759777.25 | 1510925.32 | 1677067.24 | 1857380.09 | 1652635.1 | 1601348.82 | 1505442.15 |
| 7 | 448998.73 | 561145.14 | 575570.77 | 532765.05 | 552811.62 | 509183.22 | 496725.44 | 491419.55 | 480512.44 |
| 8 | 804105.49 | 914500.91 | 852333.75 | 899036.47 | 952766.93 | 958225.41 | 1004137.09 | 881503.95 | 927266.34 |
| 9 | 495692.19 | 545206.32 | 528832.54 | 511049.06 | 578164.82 | 586061.46 | 549505.55 | 507297.88 | 517869.97 |
| 10 | 1645892.97 | 2138651.97 | 1845893.87 | 1766331.45 | 2251206.64 | 1931406.28 | 2193048.75 | 1987090.09 | 1832211.96 |
| 11 | 1182490.46 | 1446210.26 | 1302600.14 | 1303914.27 | 1380522.64 | 1396322.19 | 1528008.64 | 1426622.65 | 1332759.13 |
| 12 | 850936.26 | 1011822.3 | 951957.31 | 1075758.55 | 1069533.17 | 1049357.36 | 1100046.37 | 1077018.27 | 929690.71 |

Table 5 - Sum of Weekly Sales

EXPLANATION: Pivot Table used in the above image is the sum of weekly sales for that day of month for all of the 45 stores respectively.

The sum of weekly sales provides valuable insights for business decision-making, inventory planning, and financial analysis. Helps determine the total income generated from sales over a given period. Useful for evaluating business performance and setting financial targets. Comparing weekly sales across different time periods helps identify trends and seasonality.

The functionality of spreadsheet-like pivot tables were used for the aggregation of weekly sales of stores for each date, primarily used for analyzing, calculating and organizing data, manipulation of numerical and categorical data allowing for easy and efficient calculations and visualization of important metrics.

Statistical Data Analysis

- **OUTLIER ANALYSIS-** The handling of outliers in data is a very crucial step as the data may contain some noise and irrelevant information since the data is huge so it's good to remove the outliers in the beginning or else it will affect the performance of the model and result in low accuracy, which can lead to incorrect prediction of the sales value of business and ultimately it will compromise with the overall revenue and profit of the business.

OBSERVATIONS: Outliers were checked whether they are in the data or not and we found there were **outliers in sales and unemployment columns** which were found using boxplot in matplotlib library in python and then are handled by finding IQR (interquartile range) to find the range between which we want our data to be.

- **CORRELATION-** We also tried to find the correlation between weekly sales and all the feature columns for each of the stores in data. It's important to find relations between various features of the data before building the model so that we know which columns and feature values affect the sales of the business.

OBSERVATION: We found the correlation between weekly sales and unemployment rate to see if sales are affected by it or not, we found that since the **correlation between weekly sales and unemployment column is -0.07** which is close to 0, that implies it is hardly possible that sales are affected by unemployment.

- Profit is an important element obtained by deducting the loss, total cost and all the other expenses from the revenue. Every business wanted to achieve high revenue and sales so as to obtain a high profit amount. This provides a clear indication of business financial performance and if it's doing good in terms of revenue or not.

Mathematical formula used for calculating the profit:

$$\text{Net Profit} = \text{Total} - \text{COGS}$$

| A | B | C | D | E | F | G | H | I | J | K | L |
|--------|------------------|------------|----------|---------|----------|-------------|--------|------------------|--------------|--------|-------------|
| Gender | Product line | Unit price | Quantity | Tax 5% | Total | Payment | cogs | gross margin per | gross income | Rating | Profit (\$) |
| Female | Health and beau | 74.69 | 7 | 26.1415 | 548.9715 | Ewallet | 522.83 | 4.761904762 | 26.1415 | 9.1 | 26.1415 |
| Female | Electronic acces | 15.28 | 5 | 3.82 | 80.22 | Cash | 76.4 | 4.761904762 | 3.82 | 9.6 | 3.82 |
| Male | Home and lifesty | 46.33 | 7 | 16.2155 | 340.5255 | Credit card | 324.31 | 4.761904762 | 16.2155 | 7.4 | 16.2155 |
| Male | Health and beau | 58.22 | 8 | 23.288 | 489.048 | Ewallet | 465.76 | 4.761904762 | 23.288 | 8.4 | 23.288 |
| Male | Sports and trave | 86.31 | 7 | 30.2085 | 634.3785 | Ewallet | 604.17 | 4.761904762 | 30.2085 | 5.3 | 30.2085 |
| Male | Electronic acces | 85.39 | 7 | 29.8865 | 627.6165 | Ewallet | 597.73 | 4.761904762 | 29.8865 | 4.1 | 29.8865 |
| Female | Electronic acces | 68.84 | 6 | 20.652 | 433.692 | Ewallet | 413.04 | 4.761904762 | 20.652 | 5.8 | 20.652 |
| Female | Home and lifesty | 73.56 | 10 | 36.78 | 772.38 | Ewallet | 735.6 | 4.761904762 | 36.78 | 8 | 36.78 |
| Female | Health and beau | 36.26 | 2 | 3.626 | 76.146 | Credit card | 72.52 | 4.761904762 | 3.626 | 7.2 | 3.626 |
| Female | Food and beverc | 54.84 | 3 | 8.226 | 172.746 | Credit card | 164.52 | 4.761904762 | 8.226 | 5.9 | 8.226 |
| Female | Fashion accessc | 14.48 | 4 | 2.896 | 60.816 | Ewallet | 57.92 | 4.761904762 | 2.896 | 4.5 | 2.896 |
| Male | Electronic acces | 25.51 | 4 | 5.102 | 107.142 | Cash | 102.04 | 4.761904762 | 5.102 | 6.8 | 5.102 |

Table 6 - Net Profit Calculation

5 Results and Findings

Time Series Decomposition

In a time-series model, one of the biggest factors that affects are trends and seasonality patterns, it's very crucial to know about the trends and seasonality in any given data. Businesses use trend analysis to predict future demand and adjust inventory and seasonal forecasting to help businesses stock up on high-demand products before peak periods. Here it rises on some dates and on a few days it is exceptionally low.

We have computed trend patterns and seasonality for a particular store 22 to see how it affects the overall efficiency of the model.

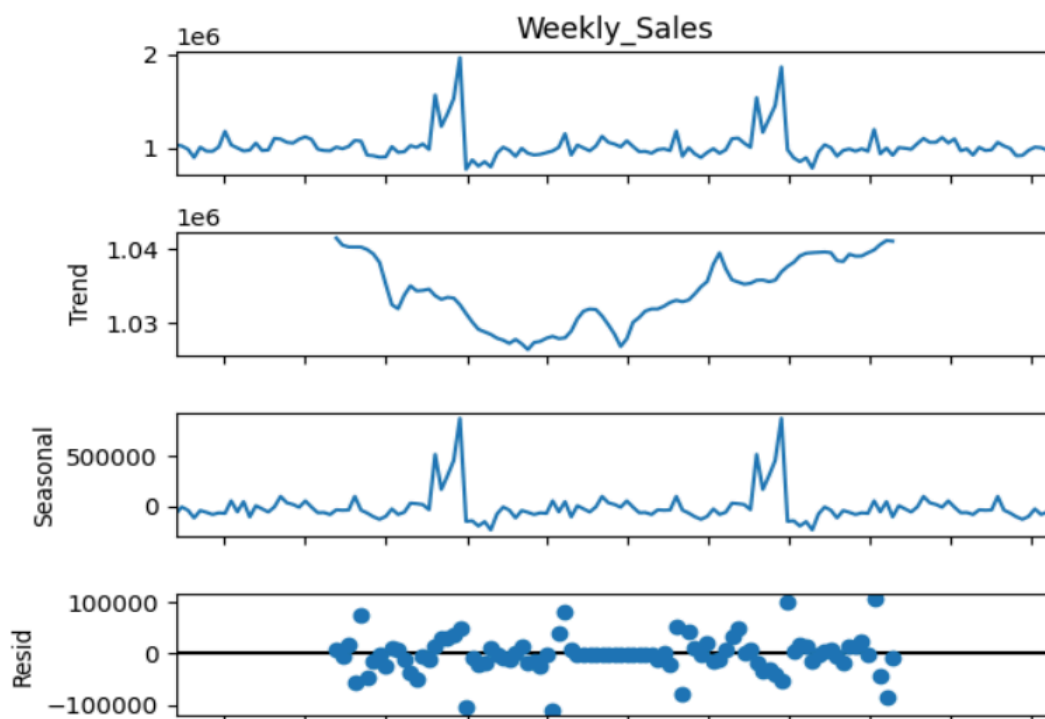


Figure 1- Trends and Seasonality of Store 22

OBSERVATIONS:

- There are noticeable spikes, indicating periods of higher sales. Shows fluctuations and possible trends over time.

- The trend shows a gradual decline followed by a rise, indicating possible seasonal effects or external factors affecting sales. Represents the underlying long-term pattern in the data.
- The presence of peaks suggests that sales increase at regular intervals, possibly due to seasonal demand. Captures repeating patterns in the data.
- The presence of large residual spikes may indicate anomalies or external events impacting sales. Ideally, residuals should appear as random noise. Represents the remaining variance after removing trend and seasonality.

Visualization of Dataset

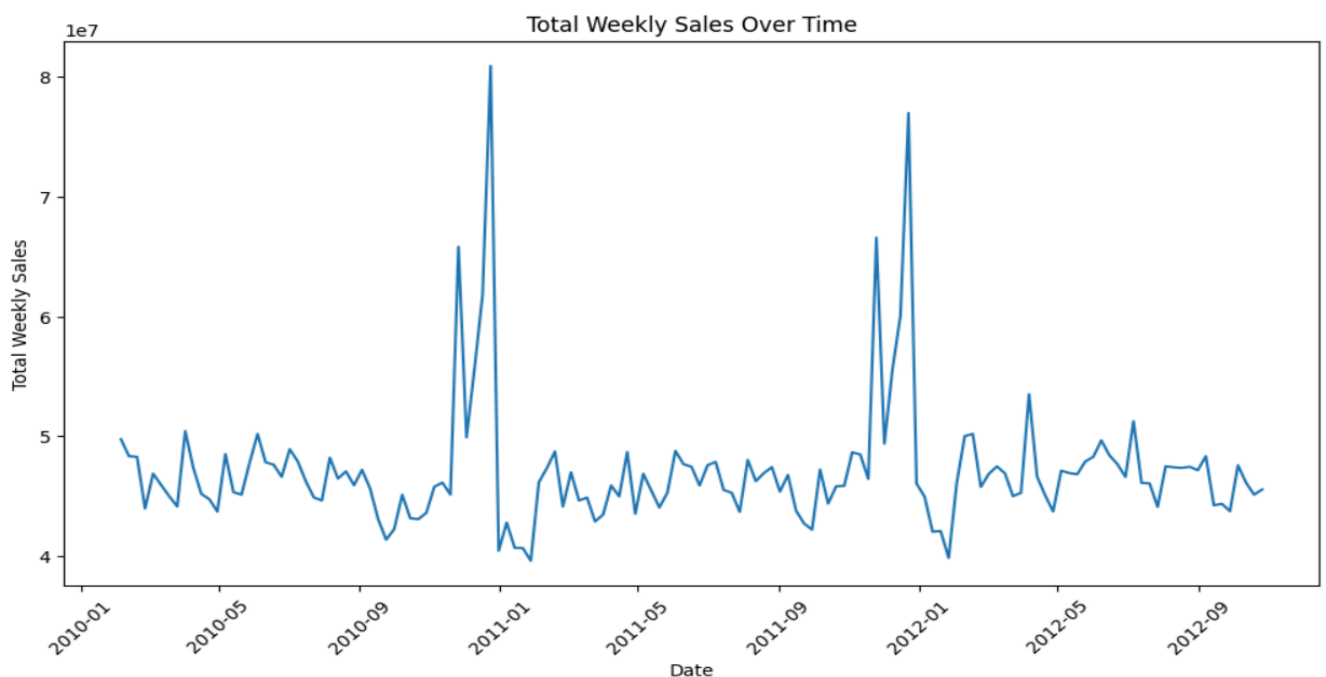


Figure 2

The line graph represents the total weekly sales over time from 2010 to 2012. The trend indicates relatively stable sales with periodic fluctuations, but there are two significant spikes, reaching peak sales above 80 million. After each peak, sales drop before stabilizing again at the baseline. The pattern highlights the impact of consumer behavior, external market conditions, and seasonal demand on overall sales performance.

OBSERVATIONS:-

- **Seasonal Spikes:** There are two major peaks around the beginning of 2011 and 2012. These spikes suggest a holiday season effect

- **Overall Trend:** Apart from the peaks, total weekly sales fluctuate around a steady range. There are minor ups and downs, indicating periodic variations in sales.
- **Periodic Declines:** Noticeable dips in sales occur at multiple points, possibly after peak seasons. These declines may be due to seasonal demand fluctuations or external economic factors.

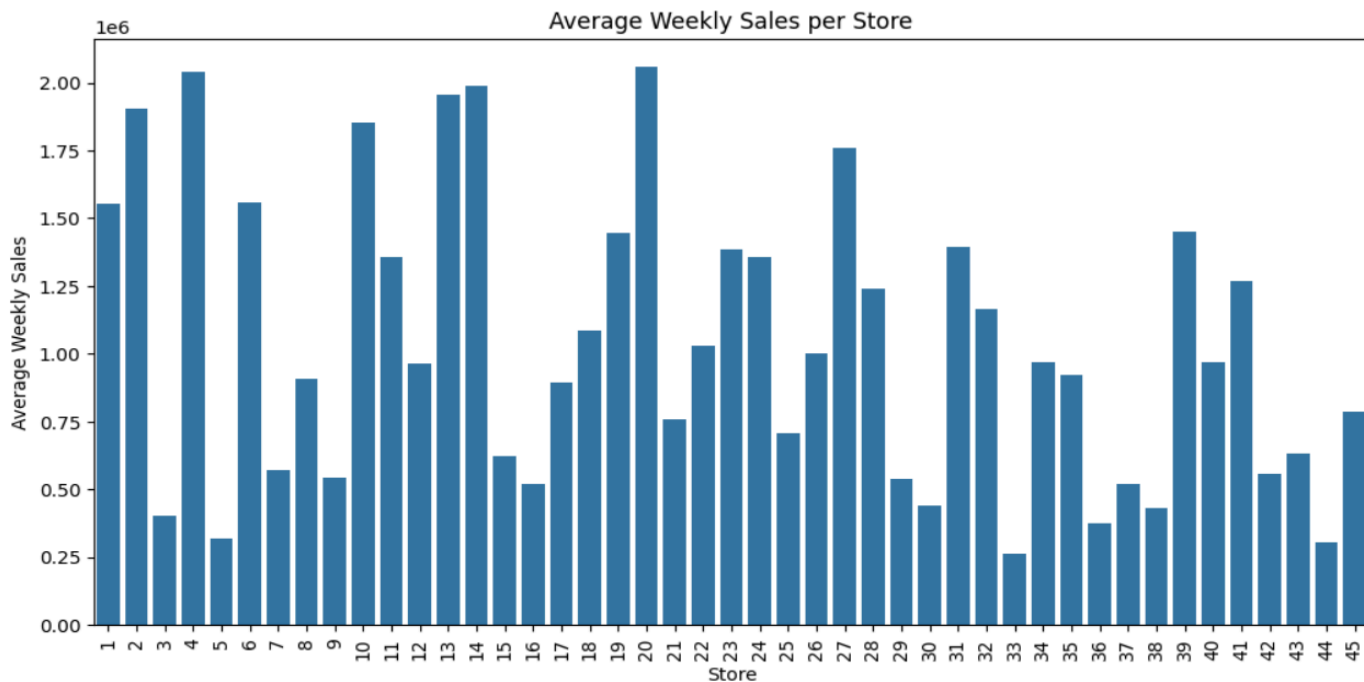


Figure 3 - Weekly sales value for 45 Stores

OBSERVATIONS:

The above graph is a visual representation of data of average weekly sales of all the stores across the country in the US. Shows how much weekly sales vary from the average.

- Stores 20, 14, and 4 have the highest average weekly sales, exceeding 2 million. Store 21 is also among the top-performing stores.
- There is a noticeable fluctuation in sales across stores. Some stores have consistently high sales, while others show low or moderate performance.
- Stores 20 and 14 show exceptionally high sales compared to the rest, which might indicate:
 - High demand in those locations.
 - Better marketing strategies.
 - Seasonal or promotional sales effects.

The results drawn from the analysis of the data are as following:

- After grouping by the stores to see the average of weekly sales and sorting the values, we saw Store number 20 makes the highest sales and revenue generation.
- The lowest performing store is store number 38 with the lowest sales value among all the stores.

| Weekly_Sales | |
|--------------|--------------|
| Store | |
| 20 | 2.800237e+08 |
| 14 | 2.761276e+08 |
| 2 | 2.687221e+08 |
| 13 | 2.682025e+08 |
| 10 | 2.556789e+08 |

Table 7 - Top 5 stores with the highest sales

The significant difference between highest and lowest sale is 272676355.92

Therefore, the stores with their assumed sales values should keep up with the stocks accordingly so that the issue of stockouts and overstocking is being minimized and the wastage of products of as less as possible because the wastage leads to less revenue and spending more on storage and handling of the items in each store.

Model Building

- During building a Time-Series model, we found the data is not stationary as there is seasonality and trends in the data which we checked using the ADfuller test. So we made the data stationary using first and second order differencing and logarithmic differencing. Once the p value is less than 0.05, that will indicate the data is stationary and we are good to go with the model building part.
- The parameters p,d,q we found came out to be (5,1,5) using the pmdarima library then we built a SARIMAX model for the store number 22 using the weekly sales feature column, for next 12 weeks, then we trained the model to make predictions.

Predictive Analysis

| | Predictions | Weekly_Sales |
|------------|-------------|--------------|
| Date | | |
| 2011-07-15 | 1233090 | 961954 |
| 2011-07-22 | 1224381 | 964684 |
| 2011-07-29 | 1256553 | 944959 |
| 2011-08-05 | 1301854 | 983233 |
| 2011-08-12 | 1294915 | 991779 |
| ... | ... | ... |
| 2012-09-28 | 1483318 | 976480 |
| 2012-10-05 | 1597960 | 1009887 |
| 2012-10-12 | 1534783 | 1004040 |

Table 8

- The above table represents the actual sales value and the predicted sales for the store number 22 for the data given to us.

The actual weekly sales and predicted sales generally follow a similar trend over time. Predicted sales are consistently higher than actual weekly sales, indicating an overestimation.

- Sales tend to increase gradually from July 2011 to October 2012. The highest predicted sales appear in October 2012, which might indicate a peak shopping period (potentially due to seasonal effects).
- There is a noticeable gap between predicted and actual sales. The difference varies but appears to be relatively stable across different weeks, meaning the model is consistently overestimating sales. This could indicate external factors affecting actual sales, such as economic conditions, promotions, or weather events.

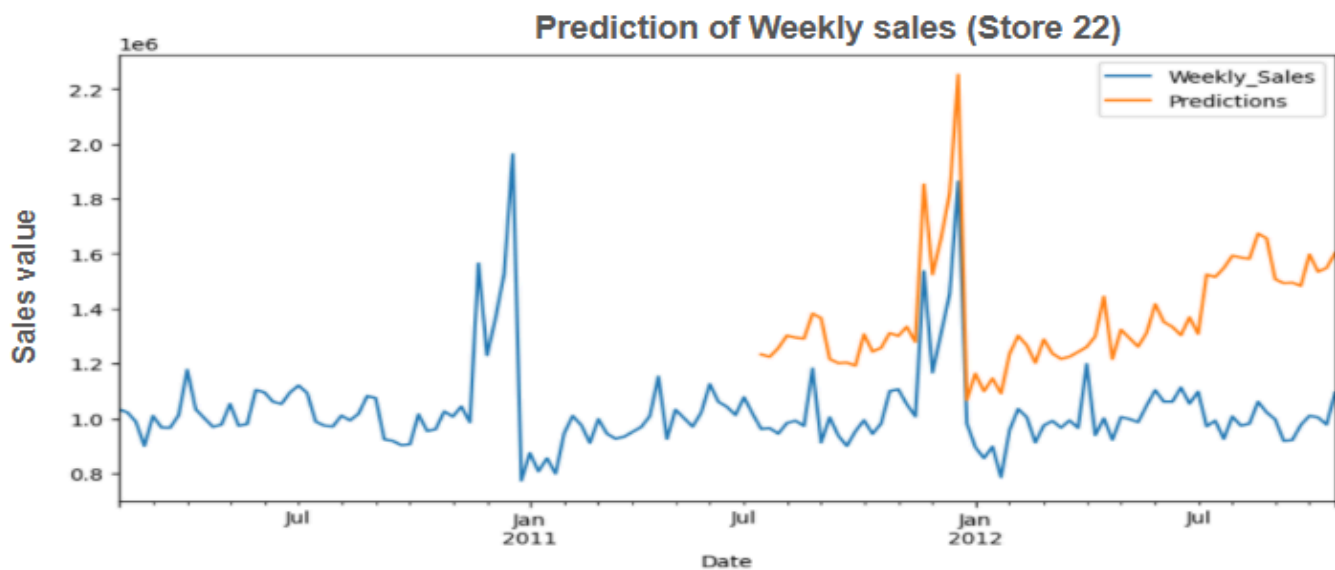


Figure 4

The predictions are generally higher than actual sales, especially in late 2011 and early 2012. This suggests that the forecasting model may be overestimating future sales trends. Both actual and predicted sales show sharp spikes at certain points, especially in early 2011 and late 2012. These could be due to seasonal sales events, promotions, or external factors .

Next we did the **model forecasting** to find the weekly sales value for next 12 weeks in order to achieve proper functioning of the store.

| | predicted_mean |
|------------|----------------|
| 2012-11-02 | 1.212285e+06 |
| 2012-11-09 | 1.176948e+06 |
| 2012-11-16 | 1.128439e+06 |
| 2012-11-23 | 1.672924e+06 |
| 2012-11-30 | 1.298890e+06 |
| 2012-12-07 | 1.438789e+06 |
| 2012-12-14 | 1.582330e+06 |
| 2012-12-21 | 1.999425e+06 |
| 2012-12-28 | 1.111048e+06 |
| 2013-01-04 | 1.026683e+06 |

Table 9

- The **visual representation of forecasted sales** made on the data of Store 22:

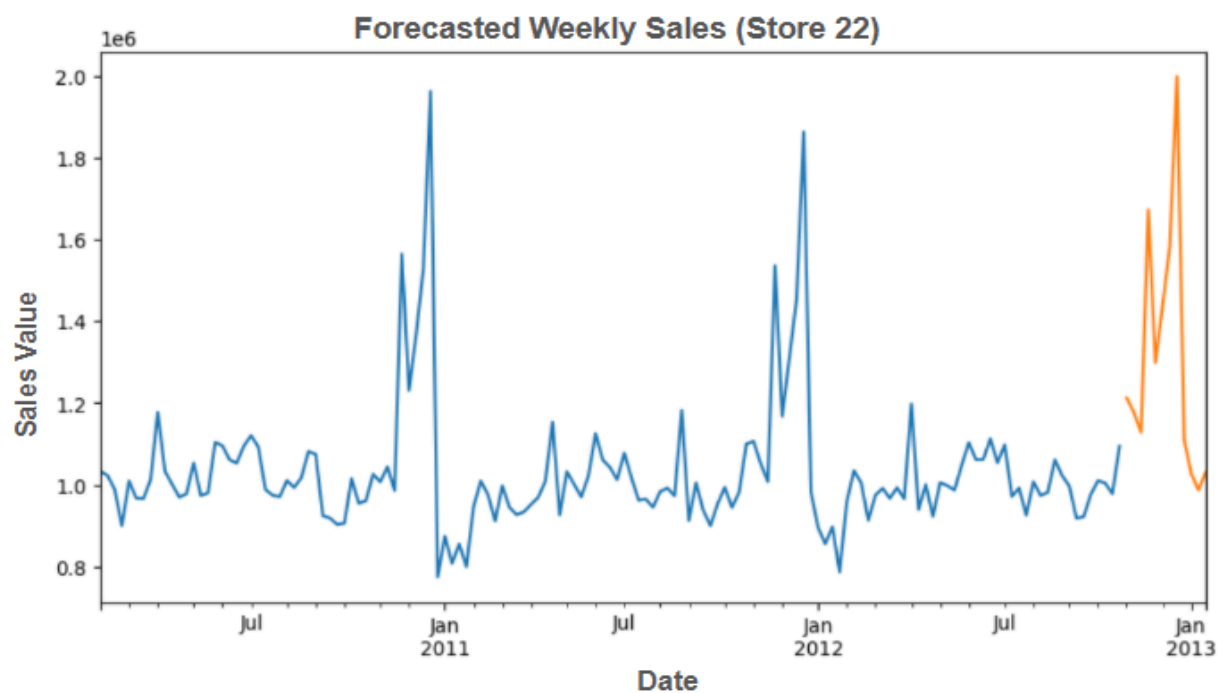


Figure 5

- Historical Trends Remain Consistent: The actual sales (blue line) still show seasonal spikes at regular intervals (around year-end). The fluctuations are generally stable, except for a few sharp peaks.
- Future Sales Predictions (Orange Line): The predicted values for early 2013 show an unexpected surge. The increase is abrupt, which is based on historical patterns.
- It may account for real-world business cycles like supply chain disruptions, economic factors, or company policies.

| | 2012-11-02 | 2012-11-09 | 2012-11-16 | 2012-11-23 | 2012-11-30 | 2012-12-07 | 2012-12-14 | 2012-12-21 | 2012-12-28 | 2013-01-04 | ... |
|----|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-----|
| 28 | 1.354139e+06 | 1.309411e+06 | 1.141846e+06 | 1.225175e+06 | 1.276162e+06 | 1.358286e+06 | 1.152711e+06 | 1.208250e+06 | 1.241359e+06 | 1.345351e+06 | |
| 29 | 5.365321e+05 | 5.119774e+05 | 5.299322e+05 | 5.393069e+05 | 5.311570e+05 | 5.260661e+05 | 5.312896e+05 | 5.345672e+05 | 5.311143e+05 | 5.293440e+05 | |
| 30 | 4.413775e+05 | 4.400999e+05 | 4.391165e+05 | 4.410540e+05 | 4.397714e+05 | 4.401879e+05 | 4.395332e+05 | 4.408696e+05 | 4.394191e+05 | 4.403025e+05 | |
| 31 | 1.376716e+06 | 1.392733e+06 | 1.368553e+06 | 1.371223e+06 | 1.379710e+06 | 1.385141e+06 | 1.380738e+06 | 1.374650e+06 | 1.377395e+06 | 1.381706e+06 | |
| 32 | 1.216259e+06 | 1.192397e+06 | 1.197412e+06 | 1.201023e+06 | 1.193555e+06 | 1.188666e+06 | 1.191051e+06 | 1.192440e+06 | 1.191421e+06 | 1.190992e+06 | |
| 33 | 2.591816e+05 | 2.919184e+05 | 2.781067e+05 | 2.544876e+05 | 2.547806e+05 | 2.805553e+05 | 2.904025e+05 | 2.582854e+05 | 2.539076e+05 | 2.676257e+05 | |
| 34 | 9.743347e+05 | 9.684254e+05 | 9.646098e+05 | 9.643752e+05 | 9.696137e+05 | 9.700410e+05 | 9.663676e+05 | 9.661878e+05 | 9.683921e+05 | 9.688764e+05 | |
| 35 | 8.895217e+05 | 9.048523e+05 | 8.623604e+05 | 8.570245e+05 | 8.501783e+05 | 8.559821e+05 | 8.573111e+05 | 8.593232e+05 | 8.587694e+05 | 8.586207e+05 | |
| 36 | 2.732442e+05 | 2.927310e+05 | 2.952894e+05 | 2.783999e+05 | 2.701801e+05 | 2.862074e+05 | 2.975943e+05 | 2.854268e+05 | 2.703340e+05 | 2.791569e+05 | |
| 37 | 5.355655e+05 | 5.189542e+05 | 5.322094e+05 | 5.383309e+05 | 5.327864e+05 | 5.286744e+05 | 5.218017e+05 | 5.409068e+05 | 5.290554e+05 | 5.364406e+05 | |
| 38 | 4.443304e+05 | 4.684608e+05 | 4.372671e+05 | 4.258701e+05 | 4.283218e+05 | 4.671437e+05 | 4.541600e+05 | 4.323352e+05 | 4.256221e+05 | 4.521034e+05 | |
| 39 | 1.614506e+06 | 1.573340e+06 | 1.575919e+06 | 1.583199e+06 | 1.573682e+06 | 1.567724e+06 | 1.569876e+06 | 1.569827e+06 | 1.568943e+06 | 1.569521e+06 | |

Table 10 - Forecasted sales of all stores for 12 weeks

The above image represents the forecasted sales of all the stores for next 12 weeks, to help the business identify the stock needs and prepare the stocks of each store accordingly.

Quantitative Analysis

The quantitative analysis made on product sales dataset of walmart using Pivot tables:

| Product line | SUM of Quantity |
|------------------------|-----------------|
| Electronic accessories | 971 |
| Fashion accessories | 902 |
| Food and beverages | 952 |
| Health and beauty | 854 |
| Home and lifestyle | 911 |
| Sports and travel | 920 |
| Grand Total | 5510 |

Table 11

- The most bought products are electronic accessories followed by food and beverages then sports and travel, it might be because of the peak holiday season.

| <i>Product line</i> | <i>AVERAGE of Rating</i> |
|------------------------|--------------------------|
| Electronic accessories | 6.424705882 |
| Fashion accessories | 7.829213483 |
| Food and beverages | 7.113218391 |
| Health and beauty | 7.503289474 |
| Home and lifestyle | 6.8375 |
| Sports and travel | 6.21626506 |
| Grand Total | 6.987365382 |

Table 12

- This shows us the rating of the product categories indicating Fashion accessories have the highest rating and Sports and Travel have the lowest rating so this product category must be improved in terms of quality to increase sales among all products. Also the other products should adjust quality and pricing to increase ratings.

AVERAGE of Rating vs. Product line

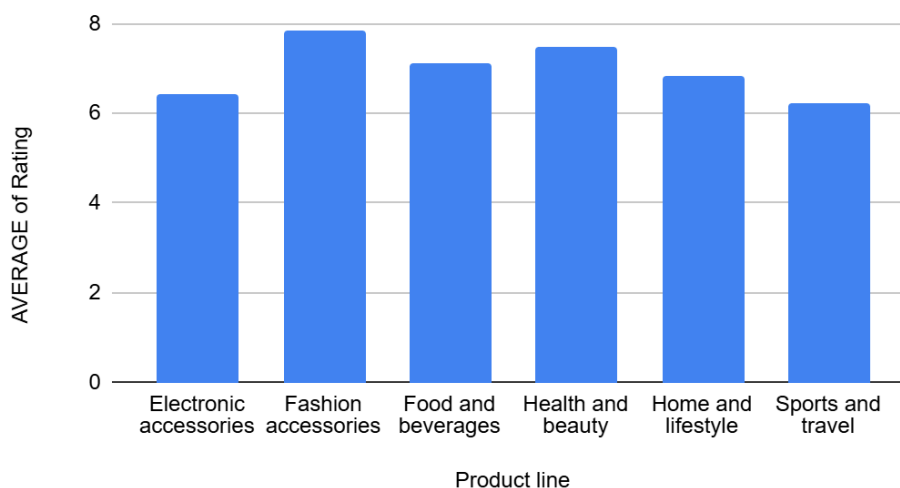


Figure 6

- The graph gives us the average ratings of the Product Categories indicating fashion accessories have the highest rating of 7.8 and the lowest rating of Sports and Travel having a rating of 6.2. Therefore the product quality and pricing must be improvised to increase average rating.

| <i>Product line</i> | SUM of Unit price |
|------------------------|-------------------|
| Electronic accessories | 9103.77 |
| Fashion accessories | 10173.35 |
| Food and beverages | 9745.54 |
| Health and beauty | 8337.88 |
| Home and lifestyle | 8850.71 |
| Sports and travel | 9460.88 |
| Grand Total | 55672.13 |

Table 13

- Fashion Accessories has the highest total unit price at 10,173.35, indicating it might have the highest-priced items or highest sales volume.
- Health and Beauty has the lowest total unit price at 8,337.88, suggesting lower sales or cheaper unit prices.
- Electronic Accessories, Food and Beverages, and Sports and Travel have similar total values, indicating a balanced sales trend among these categories.

| <i>Product line</i> | SUM of Profit |
|---------------------|------------------|
| Electronic acces | 2587.5015 |
| Fashion accessc | 2585.995 |
| Food and beverc | 2673.564 |
| Health and beau | 2342.559 |
| Home and lifesty | 2564.853 |
| Sports and trave | 2624.8965 |
| Grand Total | 15379.369 |

Table 14

- This shows electronic items generate the maximum of the profit, followed by sports and travel items so these items should be promoted more to achieve maximum profit.

| <i>SUM of Total</i> | <i>Gender</i> | | |
|------------------------|-------------------|-------------------|-------------------|
| <i>Product line</i> | Female | Male | Grand Total |
| Electronic accessories | 27102.0225 | 27235.509 | 54337.5315 |
| Fashion accessories | 30437.4 | 23868.495 | 54305.895 |
| Food and beverages | 33170.9175 | 22973.9265 | 56144.844 |
| Health and beauty | 18560.9865 | 30632.7525 | 49193.739 |
| Home and lifestyle | 30036.8775 | 23825.0355 | 53861.913 |
| Sports and travel | 28574.721 | 26548.1055 | 55122.8265 |
| Grand Total | 167882.925 | 155083.824 | 322966.749 |

Table 15

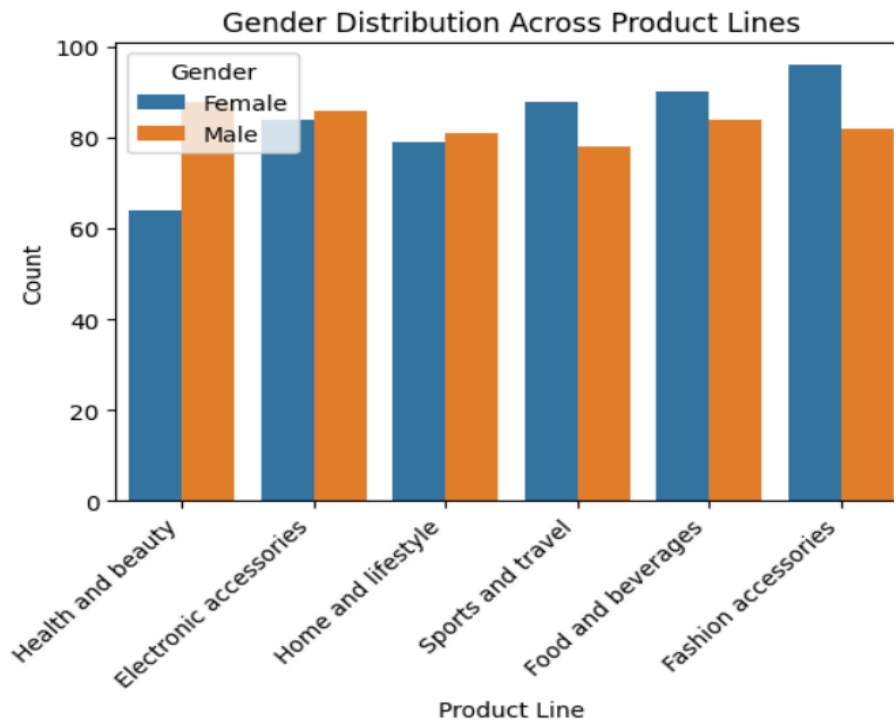


Figure 7

- Total Sales by Gender: Females: 167,882.93 (52%)
- Males: 155,083.82 (48%)
- Observation: Female customers contributed slightly more to total sales.
- Top-Selling Product Line: Food and Beverages (56,144.84) had the highest total sales, with females spending significantly more than males.
- Gender-Based Preferences: Females spent the most on Food and Beverages (33,170.92), suggesting a preference for this category. Males spent the most on Health and Beauty (30,632.75), showing higher interest in this segment.
- Balanced Sales Categories: Electronic Accessories and Home & Lifestyle show relatively balanced spending between both genders

Rationale for Graphs:

1. Gender Distribution Across Product Lines- A grouped bar chart is ideal for comparing two categories (male vs. female) across multiple product lines.
2. Average sales value over time- A line chart effectively captures sales trends over time, highlighting seasonal fluctuations and major spikes in sales. A scatter plot could be used, but it might not convey the trend as smoothly as the line chart did here.
3. Average Weekly Sales per Store- A bar chart is useful for comparing sales across different stores. It provides a clear visualization of which stores perform better in terms of average weekly sales. A box plot could be used to show the distribution of sales per store, but it wouldn't highlight individual store performances as effectively.
4. Average Rating vs Product Line- A bar chart clearly compares the average rating across different product lines. The length of bars makes it easy to see which product lines have higher or lower ratings.

6 Interpretation of Results and Recommendations

After reviewing all the analysis and reports, the following **interpretations** are made to help the business grow and improve performance:

1. Electronic accessories is the most bought product category by the customers, followed by food and beverage items.
2. Fashion Accessories has the highest total unit price, indicating it might have the highest-priced items or highest sales volume, while Health and Beauty has the lowest total unit price.
3. Food and Beverages has the highest total sales, with females spending significantly more than male, therefore the stocks are made available accordingly.
4. Males spent the most on Health and Beauty, showing higher interest in this segment, while females spent the most on Food and Beverages, suggesting a preference for this category. Electronic Accessories and Home & Lifestyle are preferred by both genders.
5. The lowest performing store is store number 38 with the lowest sales value among all the stores, whereas Store number 20 makes the highest sales and revenue generation.
6. Sales tend to increase gradually from July 2011 to October 2012. The highest predicted sales appear in October 2012, which might indicate a peak shopping period.
7. High Sales During Year-End: The data shows significant spikes around year-end, the actual sales show seasonal spikes at regular intervals (around year-end). The fluctuations are generally stable, except for a few sharp peaks.
8. Sales Trends Over Time: The actual weekly sales data show noticeable fluctuations over time, with some extreme spikes around the end of year and dips in the beginning of the year. These fluctuations could be due to seasonal trends or economic factors.

RECOMMENDATIONS

1. Targeted Marketing: Focus promotions based on gender preferences, such as more beauty-related deals for male customers while Food and Beverage for females. Keeping enough stocks of those products to meet the demands of customers.
2. Inventory Management: Stock more of the high-selling categories like Electronic Accessories and Food & Beverages. Fashion accessories should have a good amount of stock present as it has the highest demand by the customers.
3. Pricing Strategies: Analyze why some categories have a gender imbalance and adjust pricing or promotions accordingly, like Health-beauty products, Sports & Lifestyle.
4. Sales peak between July to Oct and in 2012, so ensure adequate stock levels before these periods. Implement early procurement strategies to handle supplier lead times.
5. Run targeted promotions significantly of products which are preferred by customers leading up to peak sales months. Optimize supply chain and continuously monitor key sales drivers and identify the peak sales months to manage the required stock on time.
6. Increase male-focused campaigns for food and beverages. Promote health and beauty products equally to both genders. Stock more sports and travel items favored by females. Maintain equal stock for electronic accessories due to balanced demand.

Lowest performing stores: 38, 12, 28, 33, 44

- Store 38- Introduce aggressive promotional offers. Improve store signage and visibility to attract more foot traffic. Consider bundling products at a discount to encourage larger purchases.
- Store 12- Improve operational efficiency and increase stock availability. Improve customer engagement through social media and email marketing.
- Store 28- Invest in local digital advertising (Google Ads, Facebook promotions). Launch an online ordering system with in-store pickup options.
- Store 33- Expand product selection based on local customer preferences. Provide personalized customer service to build loyalty. Maintain good stocks for electronic accessories due to balanced demand.
- Store 44- Increase male-focused campaigns for food and beverages. Promote health and beauty products equally to both genders. Stock more sports and travel items.