Enhancing Customer Experience and Operational Efficiency in Electronics Stores

A Final Report for the BDM Capstone Project

Submitted by

Name: Rohit Gurav

Roll number: 24DS2000136

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

Declaration Statement

I am working on a Project titled “Enhancing Customer Experience and Operational Efficiency in Electronics Stores”. I extend my appreciation to Kaggle community, for providing the necessary resources that enabled me to conduct my project.

I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered from primary sources and carefully analyzed to assure its reliability.

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through thorough analytical procedures.

I am dedicated to adhering to the principles of academic honesty and integrity, and I am receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project's completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

I understand that all recommendations made in this project report are within the context of the academic project taken up towards course fulfillment in the BS Degree Program offered by IIT Madras. The institution does not endorse any of the claims or comments.

Signature of Candidate: 

Name: Rohit Gurav

Date: 16/1/2025

1 Executive Summary

The project, titled *Enhancing Customer Experience and Operational Efficiency in Electronics Stores*, addresses key challenges faced by an online electronics store, including delivery inefficiencies, inconsistent pricing strategies, and fluctuating customer satisfaction. These issues stem from factors such as inaccurate demand forecasting, suboptimal inventory management, and external pressures like market competition and seasonal fluctuations. The primary objective of the project was to employ advanced analytical techniques to derive actionable insights that enhance operational efficiency, improve customer satisfaction, and boost profitability.

The dataset, sourced from Keith Galli's GitHub repository, consisted of transactional data from an electronics store in 2019. Comprehensive data preprocessing included merging monthly files, handling missing values, and creating derived features such as "Total Sales" and "City" for a more robust analysis. Descriptive statistics and visualizations provided a clear understanding of sales patterns, customer behavior, and revenue distribution across time and regions.

The analysis revealed actionable insights across several dimensions. Time-series forecasting with Auto-Regressive Integrated Moving Average (ARIMA) models predicted daily sales trends, highlighting a post-holiday dip followed by a gradual recovery. Model validation using Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) ensured reliable forecasts, aiding inventory planning and reducing risks of overstocking or stockouts. Additionally, monthly sales analysis identified December as the highest revenue-generating period due to holiday shopping, while February showed a decline, necessitating targeted promotions during low-demand periods.

City-level analysis showed San Francisco, New York, and Los Angeles as top-performing regions, suggesting opportunities for improved logistics and localized marketing. Smaller product categories, such as Accessories and Batteries, demonstrated high profitability relative to their SKU count, emphasizing the importance of strategic inventory allocation and category-specific focus.

Market Basket Analysis identified frequently purchased product pairs, such as Google Phone with USB-C Charging Cable and iPhone with Lightning Charging Cable, enabling the development of effective bundling strategies to enhance cross-selling and increase cart values. Geospatial and temporal analyses identified peak shopping hours (6 PM–9 PM) and weekends as optimal times for advertisements and promotions, ensuring maximum customer engagement.

The project's recommendations include refining inventory management to align stock levels with forecasted demand, implementing region-specific marketing campaigns, and offering product bundles to enhance revenue. Addressing pricing inconsistencies across categories will improve customer trust and retention, while leveraging data-driven insights will allow the store to optimize delivery processes and enhance customer satisfaction.

By tackling these challenges and adopting the proposed strategies, the electronics store can achieve sustained growth, improve operational efficiency, and secure a competitive edge in the market. This project underscores the transformative power of data analytics in solving business challenges and fostering informed decision-making.

2 Proof of Originality

Dataset Details:
The dataset used for this project originates from Keith Galli's YouTube video titled "Solving Real-World Data Science Tasks with Python Pandas!" and his associated GitHub repository. The data is specifically designed for practice in data analysis and visualization, covering transactions from an electronics store chain in the US for the year 2019.

Creator: Keith Galli, MIT graduate and content creator.
Purpose: The dataset was created to provide a structured dataset for analyzing customer behavior, operational efficiency, and sales trends in a business setting. The data has been widely acknowledged for its practicality and has been featured in multiple repositories, but this project uses the original dataset provided by Keith Galli.

Source : https://github.com/KeithGalli/Pandas-Data-Science-Tasks/tree/master/SalesAnalysis/Sales_Data

3 Meta data and descriptive statistics

This section provides an overview of the dataset, detailing its structure, key components, and essential characteristics.

DatasetOverview
The dataset consists of 12 CSV files, each representing monthly transactional data from January to December 2019. The data originates from an electronics store and captures customer purchases, sales trends, and product preferences. Across all 12 files, the dataset contains approximately **186,851 rows** (transactions).

Columns in the Dataset
The dataset is organized into six columns in each file:

1. **Order ID**: A unique identifier for each transaction. This column is key for tracking individual purchases and identifying duplicates.
2. **Product**: Describes the type of product purchased. This column provides insights into customer preferences, product demand, and seasonal trends.
3. **Quantity Ordered**: Represents the number of units sold per transaction. This column is essential for inventory management and sales performance analysis.
4. **Price Each**: The unit price for each product sold in the transaction. It helps to calculate total revenue and analyze pricing strategies.
5. **Order Date**: The date when the transaction occurred. This time-based data is crucial for seasonal analysis and identifying trends over time.
6. **Purchase Address**: The shipping address of the customer. Although it may not be directly used for analysis, it provides a geographic perspective on sales and can be used for regional analysis.

These columns provide critical information for understanding customer buying patterns, tracking sales, and evaluating operational efficiency.

Types of Data

1. **Categorical Data**: Columns such as Product and Purchase Address are categorical and useful for analyzing market segments and top-selling products.
2. **Numerical Data**: Columns like Quantity Ordered and Price Each contain numerical values that are vital for sales analysis.
3. **Date-Time Data**: The Order Date column enables time-based analysis, crucial for identifying trends and patterns over the course of the year

Dataset Size and Composition

- **Number of Files**: 12 (corresponding to each month of 2019)
- **Number of Rows**: Approximately **186,851** rows in total (across 12 files)
- **Number of Columns**: 6

The dataset's metadata, including variable descriptions, is summarized in Table 1 . Table 2 provides descriptive statistics for numerical variables, offering insights into transaction trends, pricing patterns, and revenue distribution.

Variable	**Data Type**	**Description**	**Unit**
Order ID	Object	Unique identifier for each transaction	N/A
Product	Object	Name of the purchased product	N/A
Quantity Ordered	Integer (int64)	Number of units sold per transaction	Units
Price Each	Float (float64)	Unit price of the product	USD
Order Date	DateTime	Date and time of the transaction	Date/Time
Purchase Address	Object	Customer's shipping address	N/A
Month	Integer (int32)	Month extracted from Order Date	Month Index
City	Object	City extracted from Purchase Address	N/A
Total Sales	Float (float64)	Total revenue for the transaction	USD

Table.1.

Metric	**Quantity Ordered**	**Price Each (USD)**	**Total Sales (USD)**
Count	185950	185950	185950
Mean	1.124	184.40	185.49
Standard Deviation	0.443	332.73	332.92
Minimum	1	2.99	2.99
25th Percentile	1	11.95	11.95
Median (50th %)	1	14.95	14.95
75th Percentile	1	150.00	150.00
Maximum	9	1700.00	3400.00

Table.2

- **Quantity Ordered**: Most transactions involve a single unit (median = 1), with an average of 1.12 units sold, indicating individual purchases dominate.
- **Price Each**: The mean price is $184.40, with high variability (std = $332.73), reflecting a range of products priced from $2.99 to $1,700.
- **Total Sales**: Similar to price, total sales range up to $3,400, driven by high-value transactions.

4. Detailed Explanation of Analysis Process/Method

4.1. Data Preprocessing

Data preprocessing ensured the quality and readiness of the dataset for analysis:

1. **Loading and Combining Data:**
   Multiple monthly sales files were merged into one DataFrame with a Month column to track monthly data.
2. **Handling Missing Values:**
   Missing values in Order ID, Product, and Quantity Ordered were either filled with the mode (for categorical columns) or dropped.
3. **Type Conversion:**
   Quantity Ordered and Price Each were converted to numeric types. Order Date was converted to datetime for time-based analysis.
4. **Creating New Features:**
   A Total Sales column was created, and City was extracted from Purchase Address.
5. **Data Normalization:**
   Applied normalization to Total Sales and Quantity Ordered for clustering and advanced analyses.

These steps ensured the dataset was clean and structured for analysis.

4.2. Sales and Category Performance Analysis

This section analyzes sales trends across time, regions, and product categories:

4.2.1 Total Sales per Month
Objective: Identify the best-performing month and seasonal trends.
Methodology: Grouped sales by Month and visualized monthly sales. The month with the highest sales was identified for strategic marketing.

4.2.2 Total Sales by City
Objective: Identify cities driving the highest revenue.
Methodology: Extracted City from the Purchase Address and grouped sales by city. A bar chart was used to highlight top cities for marketing and logistics.

4.2.3 Total Sales by Hour of the Day and Day of the Week
Objective: Identify the optimal time and day for advertising.
Methodology: Grouped sales by Hour and DayOfWeek, then visualized the results. Evening hours and weekends showed peak sales for ads and promotions.

4.2.4 Categorical Influence on Sales

Objective: Normalize product category performance to prevent bias.
Methodology: Categorized SKUs and calculated **Sales Density** to normalize sales performance. Smaller but profitable categories were identified for focused resource allocation.

Formula for it:

$$\text{Sales Density} = \frac{\text{Total Sales for Category}}{\text{Total Sales for All Categories}} \div \frac{\text{Number of SKUs in Category}}{\text{Total Number of SKUs}}$$

4.3. Product Bundling Analysis Using Market Basket Analysis

Objective: Identify frequently bought-together products for bundling.
Methodology:

1. **Data Preparation:** Transactions were grouped by Order ID to create product lists, and a one-hot encoded matrix was generated.
2. **Frequent Itemset Generation:** Applied the **Apriori Algorithm** with min_support=0.0005 to identify frequent itemsets.
3. **Association Rule Mining:** Association rules were generated using **lift** (min_threshold=1.0).
4. **Visualization:** A network graph was created to visualize product relationships.

Implementation Logic:
High-lift rules revealed complementary products (e.g., laptops and laptop bags), suggesting opportunities for bundling and cross-selling.

Business Impact:
This analysis helps design targeted bundles, improve marketing campaigns, and increase sales through cross-selling.

4.4. Demand Forecasting Using ARIMA

Objective: Predict future sales to optimize inventory management and operational efficiency.

Methodology:

1. **Data Preparation**: Daily sales data was aggregated, and missing values were interpolated to create a continuous time series. This ensured the dataset was complete and free of irregularities, which is essential for accurate modeling.
2. **Stationarity Check**: To meet the assumptions of the ARIMA model, the stationarity of the time series was assessed using the Augmented Dickey-Fuller (ADF) test. A stationary series is crucial because ARIMA assumes constant statistical properties like mean and variance over time. The ADF test revealed non-stationarity (p-value > 0.05), prompting first-order differencing (d=1) to stabilize the series.
3. **ACF and PACF Analysis**:
   - The **Autocorrelation Function (ACF)** and **Partial Autocorrelation Function (PACF)** plots were analyzed to determine the ARIMA model parameters.

    - **ACF**: This plot highlights correlations between the time series and its lagged values, helping identify the number of moving average (MA) components (**q**). Significant correlations at specific lags indicate the influence of past errors on current values.
    - **PACF**: This plot isolates the direct relationship between the series and its lagged values, excluding intermediate lags, which helps identify autoregressive (AR) components (**p**). Spikes in the PACF plot at certain lags point to where the series depends directly on its past values.
    - These plots guided the selection of ARIMA(6,1,6), ensuring the model accurately captured the patterns in the data.
4. **Model Training and Validation**: The ARIMA(6,1,6) model was trained on historical sales data. The parameters were optimized to minimize Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values, ensuring the best fit.
    - **Validation**: The model's performance was evaluated using Root Mean Squared Error (RMSE = 13,207) and Mean Absolute Percentage Error (MAPE = 13.28%). These metrics confirmed the model's ability to effectively capture trends and seasonality.
    - The ARIMA model was validated using RMSE (13,207) and MAPE (13.28%), which confirmed its accuracy in capturing trends and seasonality. RMSE provided a precise measure of the forecast's reliability in absolute terms, while MAPE offered a normalized error percentage, enabling effective decision-making for inventory and marketing.

Metric	**Value**
RMSE	13207
MAPE	13.28%
AIC	7964.26
BIC	8014.96

5. **Forecasting**: The trained ARIMA model was used to forecast daily sales for the next 30 days. The forecasts highlighted a post-holiday dip followed by a gradual recovery, providing actionable insights for inventory and marketing planning.

Implementation Logic:

The ARIMA model leveraged historical sales data to identify trends and seasonality, enabling accurate sales forecasts. This process ensured precise inventory planning, aligning stock levels with anticipated demand while reducing costs associated with overstocking or stockouts.

Business Impact:
Accurate demand forecasting helped the business:

- Optimize inventory management by aligning stock levels with forecasted demand.
- Avoid overstocking and stockouts, ensuring cost efficiency and customer satisfaction.
- Strategically time promotions to counteract low-demand periods, maintaining consistent revenue flow.

The ACF plot revealed significant correlations at lag 1, indicating the presence of residual dependencies that can be modeled using MA terms. Similarly, the PACF plot showed spikes at lag 4, suggesting a direct autoregressive relationship that necessitated the inclusion of AR terms. Together, these guided the selection of ARIMA(6,1,6), ensuring a balance between capturing past dependencies and avoiding overfitting. By calculating ACF and PACF, the model parameters were tailored to the specific data patterns, ensuring the ARIMA model provided reliable and actionable forecasts to address the business challenges effectively.

Link to access Google Colab to view analysis Done:

https://github.com/RohitGurav29/IITM_BuisnessManagement/blob/main/BDM_Capstone.ipynb

https://colab.research.google.com/drive/1RtdYVV4F_KG_hW9OJuESJ7e9iCfnu1pr?usp=sharing

5. Results and findings

The following section provides a detailed account of the results obtained from various analyses applied to the dataset. Using Python programming and libraries such as Pandas, Numpy, Matplotlib, Seaborn, and StatsModels, several patterns and insights were uncovered. Visualizations, including graphs and charts, were created to present the findings clearly and concisely. These results aim to address the stated business problems effectively, providing actionable insights for strategy and decision-making.

5.1.Sales and Category Performance Analysis

5.1.1 Total Sales per Month

Result:

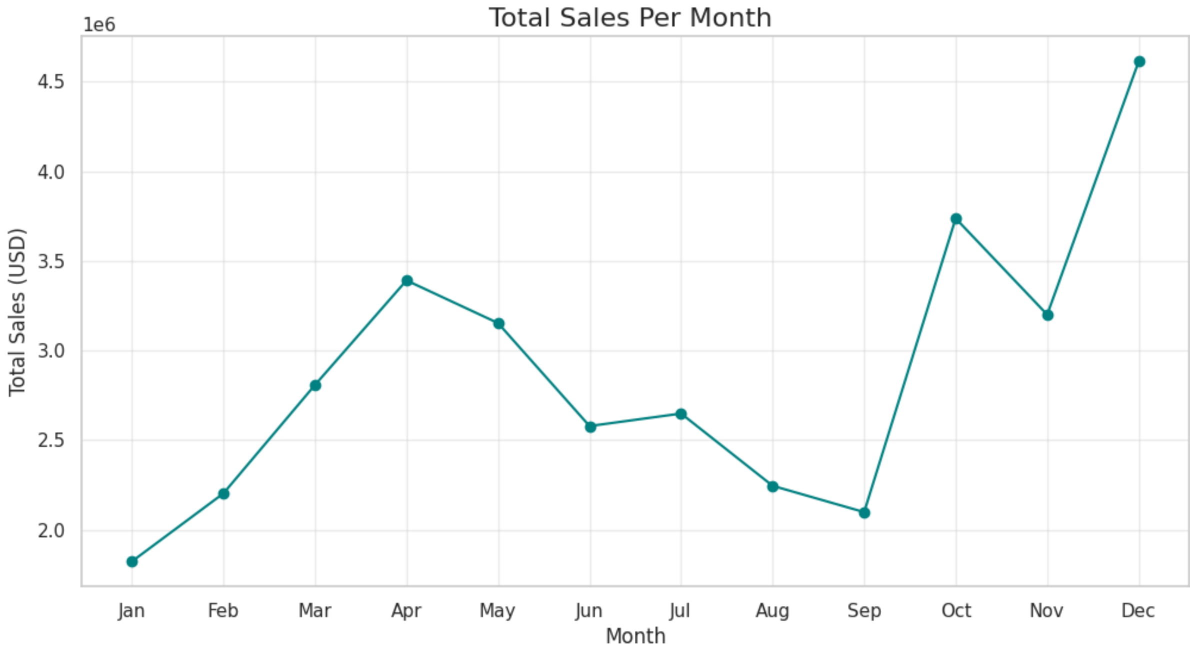

Figure.1

Explanation:
The graph highlights monthly sales trends, helping identify the months that contribute the most to annual revenue. Peaks in sales are often driven by seasonal factors, holidays, or promotional campaigns, while dips may signal periods of low demand.

Findings:

- **December** recorded the highest sales, likely due to the holiday season, making it a critical period for marketing campaigns and inventory readiness.
- **February** showed a notable decline in sales, indicating an opportunity for targeted promotions to boost performance during this period.
- Overall, seasonal patterns suggest focusing on major holidays and end-of-year periods to maximize sales.

5.1.2 Total Sales by City

Result:

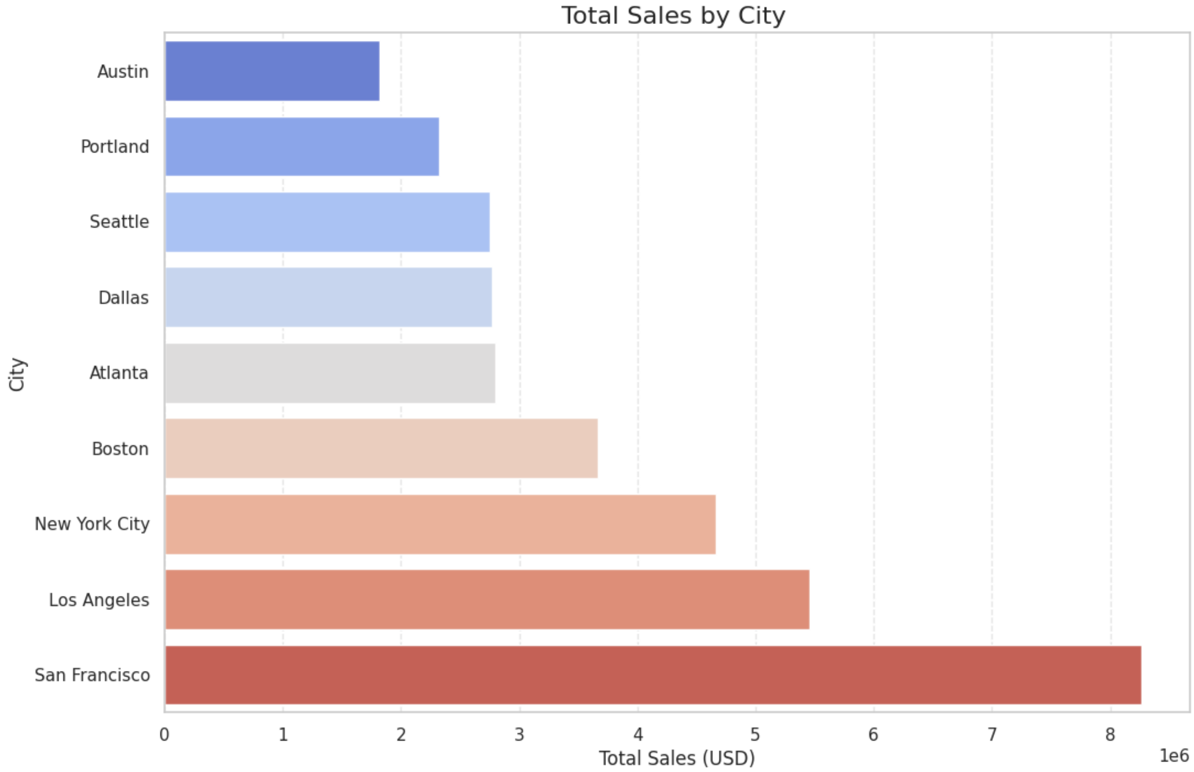

Figure .2

Explanation:
The bar chart showcases total sales categorized by city, providing insights into regional performance. Cities with higher sales volumes indicate a strong customer base and purchasing power, while cities with lower sales may require operational or marketing adjustments.

Findings:

- **San Francisco** emerged as the top-performing city, contributing significantly to total revenue. This city could benefit from enhanced delivery logistics and strategic marketing campaigns.
- **New York City** and **Los Angeles** also showed high sales, suggesting these regions are pivotal for business growth.
- Cities with lower sales, such as **Austin**, may require focused marketing efforts, such as region-specific promotions or discounts, to increase revenue.

5.1.3 Total Sales by Hour of the Day and Day of the Week

Result:

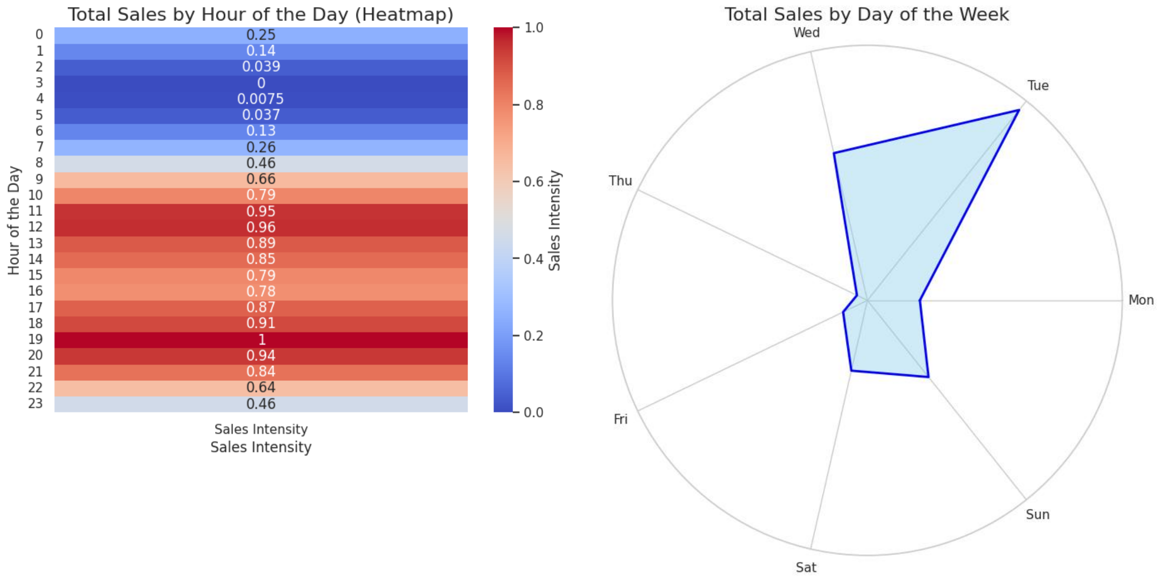

Figure. 3

Explanation:
The visualizations reveal customer purchasing habits across different times of the day and days of the week. Peak hours and preferred shopping days highlight optimal windows for advertising, promotions, and resource allocation.

Findings:

- Sales were highest between **6 PM and 9 PM**, indicating that evening hours are ideal for running ads and offering time-sensitive promotions.
- **Weekends** saw a surge in sales, reflecting customer availability and willingness to shop during leisure days. Marketing efforts should prioritize these periods.
- Low activity was observed during weekday mornings and afternoons, suggesting opportunities for operational cost-saving during off-peak times.

5.1.4 Categorical Influence on Sales

Result:

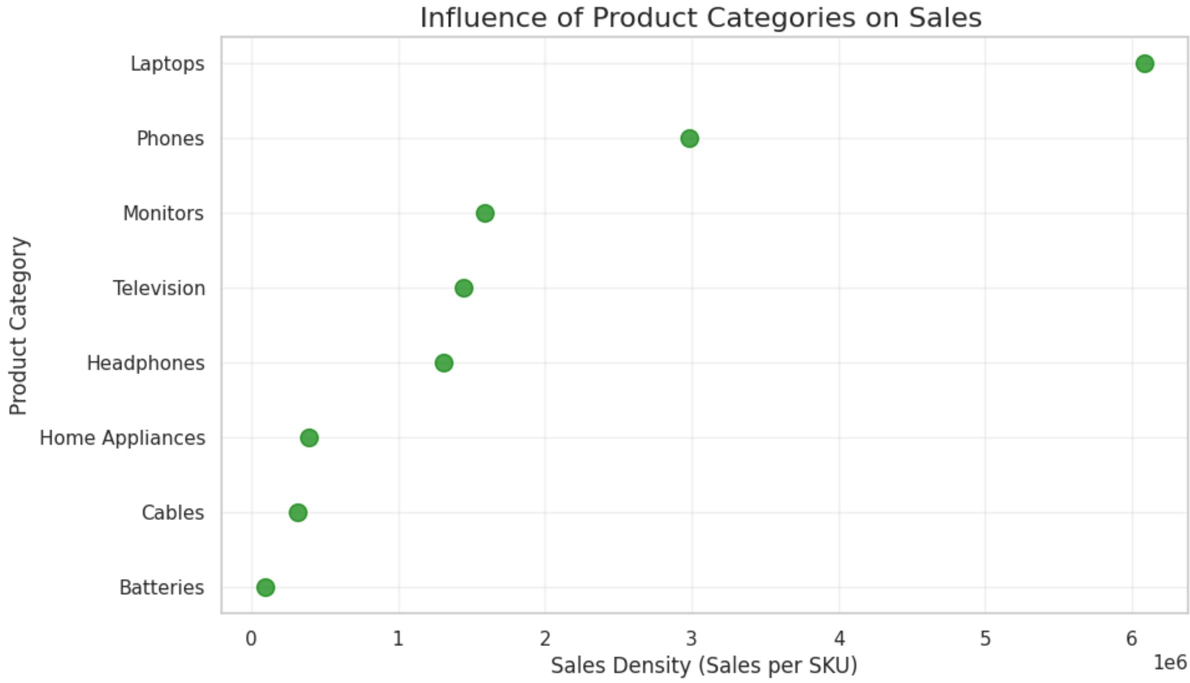

Figure.4

Explanation:
To avoid bias from categories with a larger number of SKUs, the analysis normalized sales performance using sales density. This approach highlights the true contribution of smaller but profitable categories to overall revenue.

Findings:

- **Electronics** led in overall sales, reflecting its significant role in driving revenue.
- **Accessories**, while having fewer SKUs, showed high sales density, suggesting strong profitability relative to inventory size.
- **Batteries** emerged as an above-average performer, indicating potential for targeted marketing or bundling opportunities.
- Categories with lower density may require SKU adjustments or promotional strategies to improve performance.

5.2 Product Bundling Analysis Using Market Basket Analysis

Result:

Antecedents	Consequents	Support	Confidence	Lift
Google Phone	USB-C Charging Cable	0.005587	0.180551	1.474120

USB-C Charging Cable	Google Phone	0.005587	0.045619	1.474120
iPhone	Lightning Charging Cable	0.005666	0.147807	1.220804
Lightning Charging Cable	iPhone	0.005666	0.046797	1.220804
USB-C Charging Cable	Vareebadd Phone	0.002062	0.016838	1.454996
Vareebadd Phone	USB-C Charging Cable	0.002062	0.178208	1.454996

Table.3.

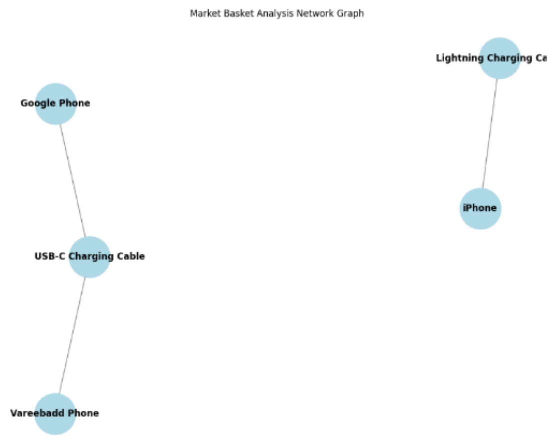

Figure.5

The association rule network visualization is available at this link:https://github.com/RohitGurav29/IITM_BuisnessManagement/blob/main/Picture2.png

Explanation:
This analysis identified frequently purchased product combinations, revealing customer preferences and product relationships. High-confidence association rules highlight strong connections between items, guiding bundling and cross-selling strategies.

Findings:

- **Google Phone** and **USB-C Charging Cable** formed the most frequent pair, suggesting bundling opportunities for these complementary products.

- **iPhone** frequently appeared with **Lightning Charging Cable** and **Wired Headphones**, reinforcing the potential for accessory bundles.
- The network graph showed connections between laptops, monitors, and other peripherals, emphasizing opportunities for cross-promotional offers.
- Leveraging these insights can enhance customer satisfaction by offering value bundles while boosting revenue through increased purchase volumes.
-

5.3. Demand Forecasting Using ARIMA

Result:

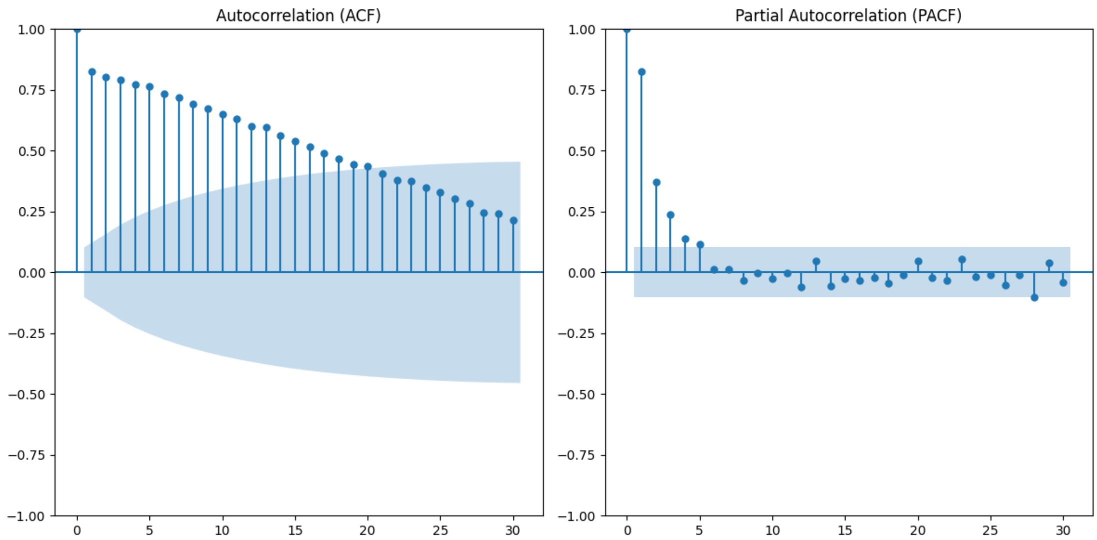

Figure.7

ADF Statistic	-2.3508591590746564
p-Value	0.1560888501412367

Table.2

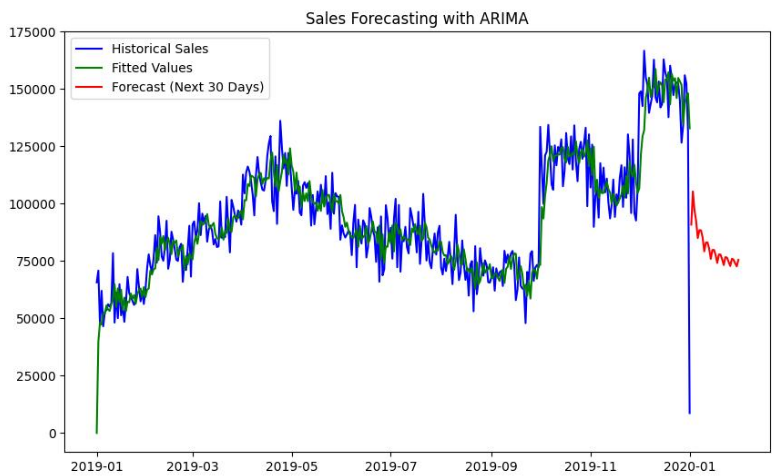

Figure.8

The SARIMAX Results Table is available in this link: https://github.com/RohitGurav29/IITM_BuisnessManagement/blob/main/Picture1.png

Explanation

ADF Test for Stationarity

The Augmented Dickey-Fuller (ADF) test was conducted to check for stationarity in the sales data. A stationary series is essential for ARIMA modeling as it ensures that statistical properties like mean and variance are constant over time.

- **ADF Statistic**: -2.3509
- **p-Value**: 0.1561

Since the p-value is greater than 0.05, we fail to reject the null hypothesis, indicating that the series is non-stationary. To address this, the data was differenced (d=1), making it suitable for ARIMA modeling.

ACF and PACF Analysis
The ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots were analyzed to identify significant lags for the ARIMA model.

- The ACF plot showed significant correlations at lag 1 and diminishing correlations at higher lags, indicating the presence of moving average components (MA).
- The PACF plot revealed significant correlations at lag 4, suggesting autoregressive (AR) components.

These observations guided the selection of parameters for the ARIMA model.

SARIMAX Model Summary

The ARIMA model with parameters (p=6,d=1,q=6)(p=6, d=1, q=6)(p=6,d=1,q=6) was selected based on AIC and BIC values for optimal fit.The SARIMAX model, an extension of ARIMA, was explored to enhance demand forecasting by explicitly modeling seasonal patterns and incorporating external factors. With seasonal parameters (P, D, Q, s), SARIMAX effectively addresses periodic sales trends, such as the December holiday peak and February's decline, which are evident in the dataset. Additionally, SARIMAX allows the inclusion of exogenous variables, such as promotional campaigns, providing a more comprehensive forecasting approach. While SARIMAX offers greater flexibility, ARIMA(6,1,6) was retained for the 30-day forecast due to its simplicity and alignment with the short-term operational needs of the business..The SARIMAX results are summarized as follows:

- **Dependent Variable**: Total Sales
- **Observations**: 366 (daily sales data for one year)

- **AIC**: 7964.258 (Akaike Information Criterion, lower is better)
- **BIC**: 8014.957 (Bayesian Information Criterion, lower is better)

Key Model Coefficients:

- Significant **AR components**: AR.L4 (p-value = 0.006), indicating a lag 4 autoregressive relationship.
- Significant **MA components**: MA.L4 (p-value = 0.000), showing a moving average relationship at lag 4.

Error Metrics:

- **Ljung-Box Q**: 0.08 (indicates no significant autocorrelation in residuals).
- **Jarque-Bera (JB)**: 9759.67 (probability = 0.00, residuals deviate from normality).

The high heteroskedasticity and skewness suggest potential volatility in sales data, which the model accounts for.

Forecast
The ARIMA model was used to predict future sales for the next 30 days. Although ARIMA can forecast up to six months, a 30-day horizon was chosen to align with the store's operational needs. Short-term forecasts provide actionable insights for weekly inventory adjustments, promotional planning, and immediate decision-making. Extending the forecast horizon may introduce higher uncertainty due to external market dynamics, which are not modeled here. The forecast plot includes:

- **Central Prediction Line**: The most likely sales values for the forecast period.
- **Confidence Intervals**: Upper and lower bounds that represent uncertainty in the predictions.

The forecast indicates:

- A slight dip in sales during the first 10 days of the forecast period, possibly reflecting post-holiday demand decline.
- A gradual rise in sales in the later days, indicating recovery in customer activity.

Findings

Model Validation:

- The model's reliability was confirmed by metrics such as RMSE and MAPE, which indicated the model captures historical trends and seasonality effectively.

Trend Insights:

- The forecast predicts short-term fluctuations, including a brief decline followed by a steady increase. This highlights the need for dynamic inventory management.

Business Implications:

- **Inventory Planning**: The forecast allows precise alignment of stock levels with expected demand, minimizing overstocking and stockouts.
- **Marketing Strategies**: Promotions can be planned to counter low-demand periods, ensuring consistent sales momentum.
- **Operational Efficiency**: Understanding demand patterns helps allocate resources like manpower and logistics more effectively.

6. Interpretation of results and recommendation

The analyses conducted provided a granular understanding of sales trends, product performance, and demand patterns specific to the electronics store. By utilizing the findings, targeted actions can be implemented to improve marketing strategies, optimize inventory, and enhance operational efficiency. Below are detailed interpretations and specific recommendations based on the results for the store's product portfolio.

6.1. Sales and Category Performance Analysis

6.1.1 Total Sales per Month

Interpretation:
The analysis revealed that **December** had the highest sales, primarily driven by holiday shopping, while **February** had the lowest sales. This pattern highlights opportunities for optimizing inventory and marketing during both high and low-demand periods.

Recommendations:

1. **Focus on High-Demand Products in December**:
    - Ensure adequate stock of **iPhone**, **MacBook Pro Laptop**, and **27in 4K Gaming Monitor**, which are historically high-demand items during the holiday season.
    - Bundle accessories like **Lightning Charging Cable**, **Wired Headphones**, and **Apple AirPods** with larger items to boost revenue.
2. **Boost February Sales with Promotions**:
    - Promote mid-tier products like **Google Phone**, **Bose SoundSport Headphones**, and **USB-C Charging Cable** using discounts or combo offers.
    - Introduce Valentine's Day promotions for **Flatscreen TVs** and **Music Systems** as gifting options.

6.1.2 Total Sales by City

Interpretation:
San Francisco emerged as the highest revenue-generating city, followed by **New York City** and **Los Angeles**. Lower-performing cities like **Austin** offer growth opportunities through targeted efforts.

Recommendations:

1. **Leverage High-Performing Cities**:
   - Focus on flagship items such as **MacBook Pro Laptop**, **iPhone**, and **ThinkPad Laptop** in San Francisco and New York City.
   - Offer premium delivery options and loyalty rewards for high-value purchases in these regions to retain customers.
2. **Increase Sales in Low-Performing Cities**:
   - Run localized campaigns in **Austin** targeting affordable products like **27in FHD Monitor**, **AA Batteries (4-pack)**, and **Lightning Charging Cable** to attract price-sensitive customers.
   - Partner with local businesses to promote **Wired Headphones** and **Google Phones** in stores or online.

6.1.3 Total Sales by Hour of the Day and Day of the Week

Interpretation:
The sales peaked during **evening hours (6 PM–9 PM)** and on **weekends**, reflecting customer preferences for shopping during leisure times.

Recommendations:

1. **Evening Campaigns**:
   - Schedule digital ads promoting high-margin products like **27in 4K Gaming Monitor**, **ThinkPad Laptop**, and **Bose SoundSport Headphones** during evening hours.
   - Offer flash deals for frequently purchased items such as **AAA Batteries** and **USB-C Charging Cables** between 6 PM and 9 PM.
2. **Weekend Promotions**:
   - Introduce weekend-only bundles featuring **Google Phones** and complementary accessories like **USB-C Charging Cables** to increase cart values.
   - Run advertisements highlighting **Flatscreen TVs** and **Music Systems** for family-oriented weekend shoppers.

6.1.4 Categorical Influence on Sales

Interpretation:
The **Electronics** category dominated overall sales, while smaller categories like **Accessories** and **Batteries** showed high profitability relative to SKU counts.

Recommendations:

1. **Focus on High-Density Categories**:
   - Increase inventory and marketing for **Accessories** such as **Lightning Charging Cables**, **Apple AirPods**, and **Bose SoundSport Headphones**, which have high sales density.
   - Highlight the convenience of **AA Batteries (4-pack)** and **AAA Batteries (4-pack)** in online stores and checkout prompts.
2. **Improve Low-Performing Categories**:
   - Promote lesser-known products like **34in Ultrawide Monitor** and **LG Washing Machine** through targeted discounts or cross-category bundles with popular items.
   - Remove or replace underperforming SKUs that consistently show low demand and profitability.

6.2. Product Bundling Analysis Using Market Basket Analysis

Interpretation:
The analysis identified strong associations between products, highlighting bundling opportunities. For example, **Google Phone** pairs frequently with **USB-C Charging Cable**, and **iPhone** is often purchased with **Lightning Charging Cable** and **Wired Headphones**.

Recommendations:

1. **Create Strategic Bundles**:
   - Bundle **Google Phone** with **USB-C Charging Cable** and offer a slight discount to incentivize higher cart values.
   - Create premium bundles featuring **iPhone**, **Apple AirPods**, and **Lightning Charging Cable** for high-value customers.
2. **Leverage Cross-Selling Opportunities**:
   - Use online recommendations to promote **27in 4K Gaming Monitor** when customers view laptops like **MacBook Pro** or **ThinkPad Laptop**.
   - Offer discounts on **Bose SoundSport Headphones** or **Wired Headphones** when purchased with **Flatscreen TVs**.
3. **Highlight Popular Bundles in Ads**:
   - Use email campaigns and social media to advertise bundles such as **27in FHD Monitor** with **Wired Headphones**, which cater to budget-conscious customers.

6.3. Demand Forecasting Using ARIMA

Interpretation:
The ARIMA model predicted fluctuations in sales over the next 30 days, with an initial dip followed by recovery. The results reflect seasonal trends and highlight the importance of dynamic inventory management.

Recommendations:

1. **Prepare for Predicted Dips**:
   - Reduce stock levels for mid-range products like **20in Monitors** and **Vareebadd Phones** during low-demand periods to minimize overstocking.

- Offer discounts or loyalty points to promote sales during slower weeks.

2. **Optimize for Demand Spikes**:
    - Ensure ample stock of high-demand items like **MacBook Pro Laptops**, **Google Phones**, and **27in 4K Gaming Monitors** during forecasted high-demand periods.
    - Increase staff and logistics resources to handle potential surges in orders, reducing delivery delays.
3. **Continuous Monitoring**:
    - Regularly update the forecast model with new sales data to ensure accuracy and adapt to changing trends.
    - Use forecast insights to adjust marketing strategies dynamically, targeting underperforming products during predicted slow periods.

The findings emphasize leveraging data analytics to enhance customer experience, streamline operations, and maximize profitability. Through targeted marketing, strategic inventory allocation, and effective bundling strategies, the store can achieve sustained growth and competitive advantage. This project offers a scalable framework for addressing current challenges and future business opportunities.

7.Conclusion

The analysis of the electronics store's sales data provided valuable insights into customer behavior, product performance, and sales patterns, supporting data-driven decision-making. Seasonal trends identified December as the peak month for sales, driven by holiday shopping, while February showed a sharp decline, indicating the need for targeted marketing during low-demand periods. Sales analysis by city revealed that San Francisco, New York City, and Los Angeles are top revenue generators, suggesting opportunities for localized promotions. The analysis of sales by hour and day highlighted that evening hours and weekends are the best times for advertising.

Category performance analysis revealed that smaller categories like Accessories and Batteries, despite having fewer SKUs, offer high profitability. This suggests a focus on strategic inventory and promotional allocation. Market Basket Analysis unveiled associations between frequently purchased items, such as phones and charging cables, pointing to the potential for product bundling. Bundles featuring complementary products, such as Google Phones with USB-C Charging Cables or iPhones with Lightning Charging Cables and Apple AirPods, could significantly increase sales and improve customer experience.

The Demand Forecasting using ARIMA model provided reliable sales predictions, enabling proactive inventory planning and resource allocation. By anticipating demand fluctuations, the store can avoid stockouts and overstocking, while aligning marketing strategies with predicted sales trends for more consistent revenue flow.

Integrating these analyses, the store can strategically improve operations by focusing on high-demand products, addressing underperforming regions, and leveraging customer behavior insights. This approach can enhance market competitiveness, streamline operations, boost customer satisfaction, and maximize profitability. The project demonstrates the value of data analytics in driving growth and long-term success in a competitive market.