



# BDM Capstone Project [Jan 2025]

# Enhancing Customer Experience and Operational Efficiency in Electronics Stores.

Rohit Gurav [24ds2000136]



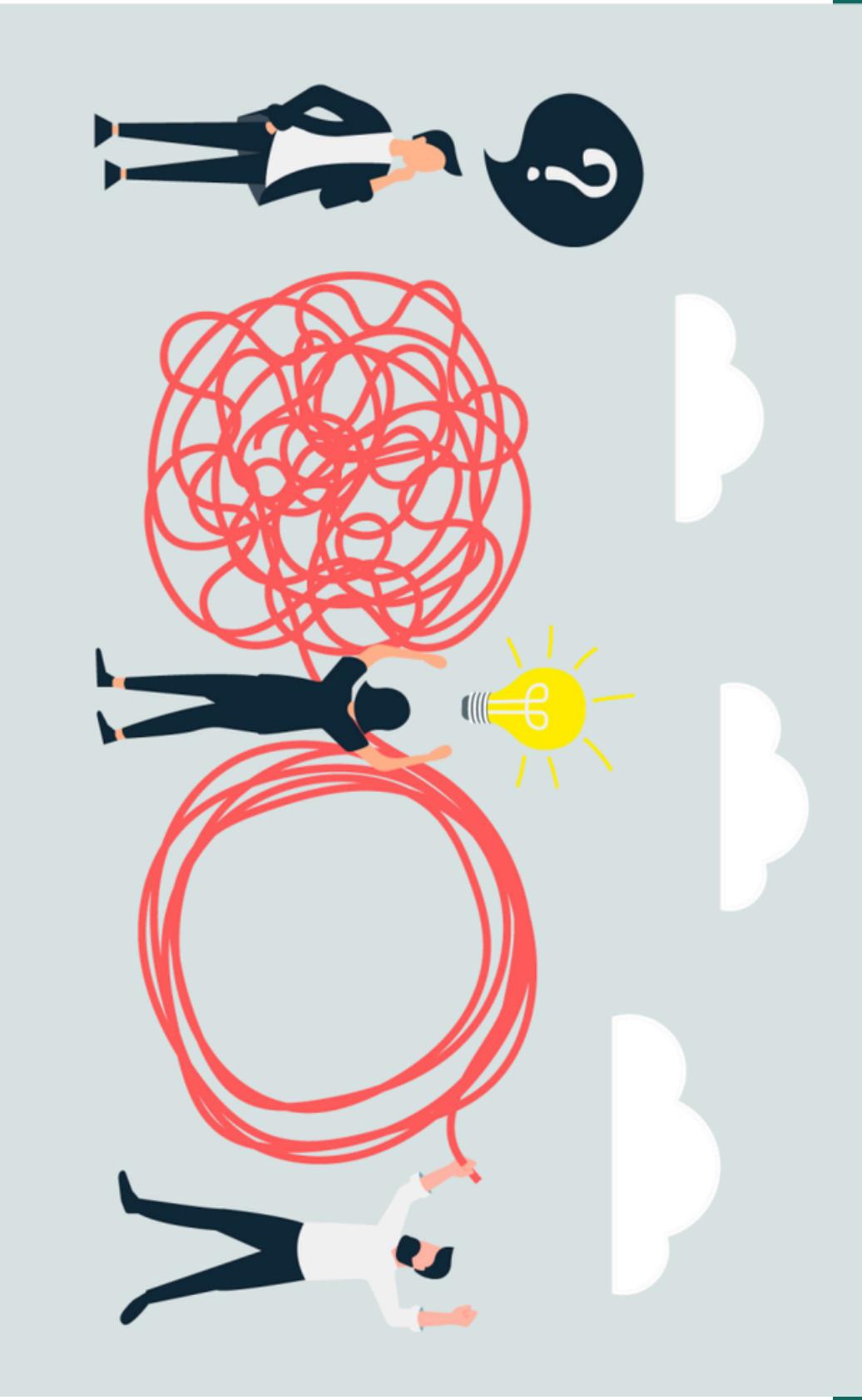
# The Business and The Problems

## About The Business

- **Name :** Keith Galli's Electronics Store Sales Data (2019)
- **Type of Business:** Online retail (B2C)
- **Location:** U.S.-based electronics store
- **Founder:** Keith Gill
- **Background:** The Dataset covers 180k+ transactions over 12 months .

## Identified Challenges

- **Delivery inefficiencies:**  
Inconsistent delivery times causing delays and customer dissatisfaction
- **Inconsistent Pricing Strategies:**  
Variability in pricing across products and regions affecting customer trust.
- **Fluctuating Customer Satisfaction:**  
Influenced by external pressures like seasonal demand and market competition.
- **Ineffective Inventory Planning:**  
Overstocking or stockouts due to inaccurate demand forecasting.
- **Limited Cross-Selling Opportunities:**  
Lack of insights into product relationships and bundling strategies.



# Dataset Overview

## Dataset Details

- Source: Keith Galli's GitHub repository (2019 transactional data).
- Type: Secondary data (online electronics store sales).
- Files: 12 monthly CSV files (January–December 2019).
- Size: 186,851 transactions, 6 key columns.

## Key Features

- Order ID: Unique identifier for each transaction.
- Product: Type of product purchased.
- Quantity Ordered: Units sold per transaction.
- Price Each: Unit price of products.
- Order Date: Transaction date for trend analysis.
- Purchase Address: Customer shipping location, used for regional analysis.

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

# Objective and the Methods

**Objective:** The aim of the project is to analyse the extracted data and draw out actionable insights that may help the business to boost up their sales and build a strong brand image in the market.

**Methods of Analysis:** After a rigorous data cleaning and preprocessing operation, 3 methods of analysis were deployed on the preprocessed dataset to uncover categorical sales influence, product performance, customer behaviour and future sales trends.



**Sales and Performance Analysis**

Analyzed trends by month, city, and product categories.



**Market Basket Analysis**

Identified frequently bought-together items for bundling using the Apriori algorithm.



**ARIMA Sales Forecasting**

Predicted future sales using ARIMA for better inventory planning.

# Data Preprocessing

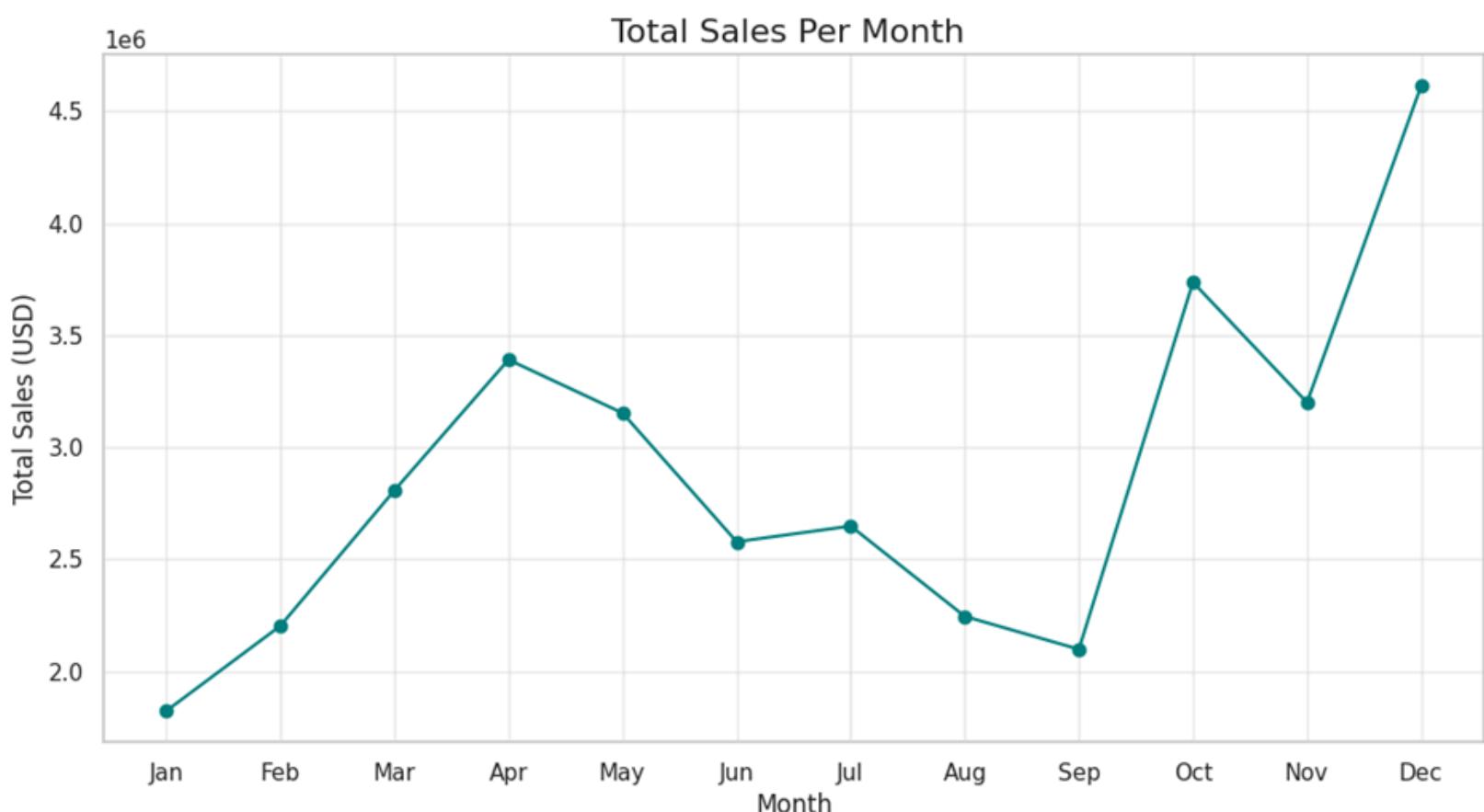
Before proceeding with the analysis, it was essential to clean and preprocess the data. A series of preprocessing steps were performed to prepare the dataset for effective and accurate analysis



# Sales and Performance Analysis

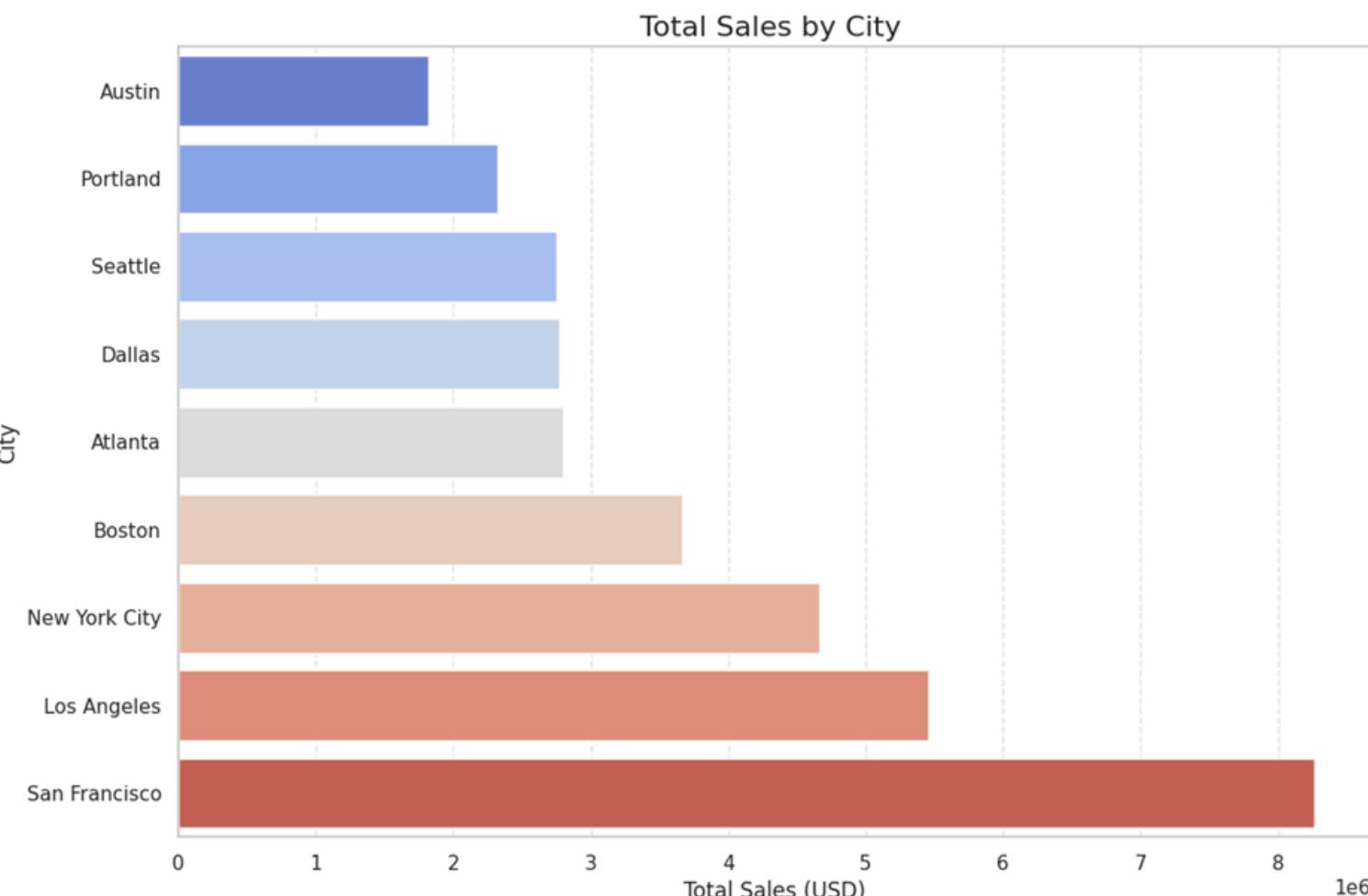
## Monthly Sales Trends

- Explanation:
  - Analyzed total sales for each month to identify seasonal patterns.
- Results and Findings:
  - Peak Month: December (highest sales due to holiday shopping).
  - Low Month: February (target for promotions).



## City-Wise Sales

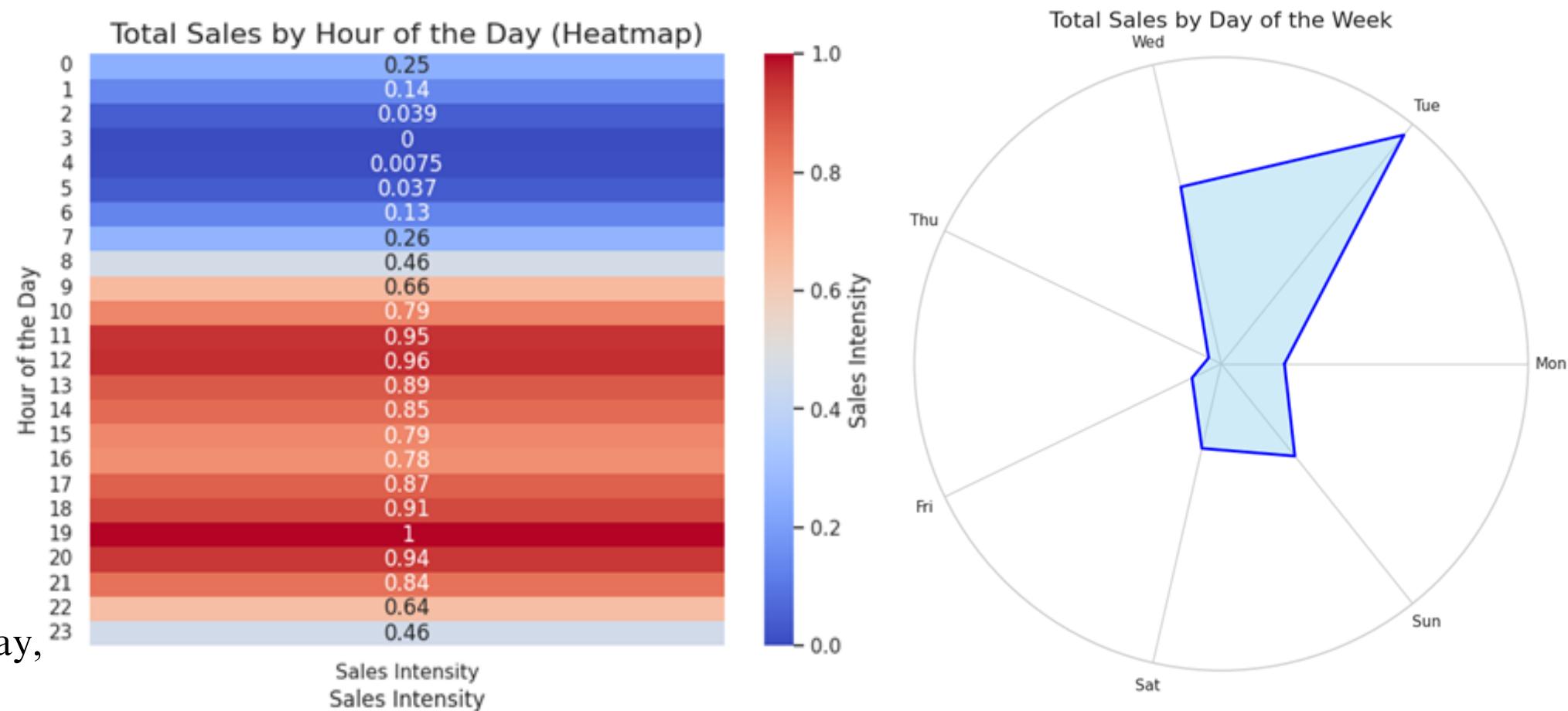
- Explanation:
  - Grouped sales data by city to understand regional performance.
- Results and Findings:
  - Top cities: San Francisco, New York, Los Angeles.
  - Low-performing city: Austin (opportunity for marketing campaigns)



# Sales and Performance Analysis

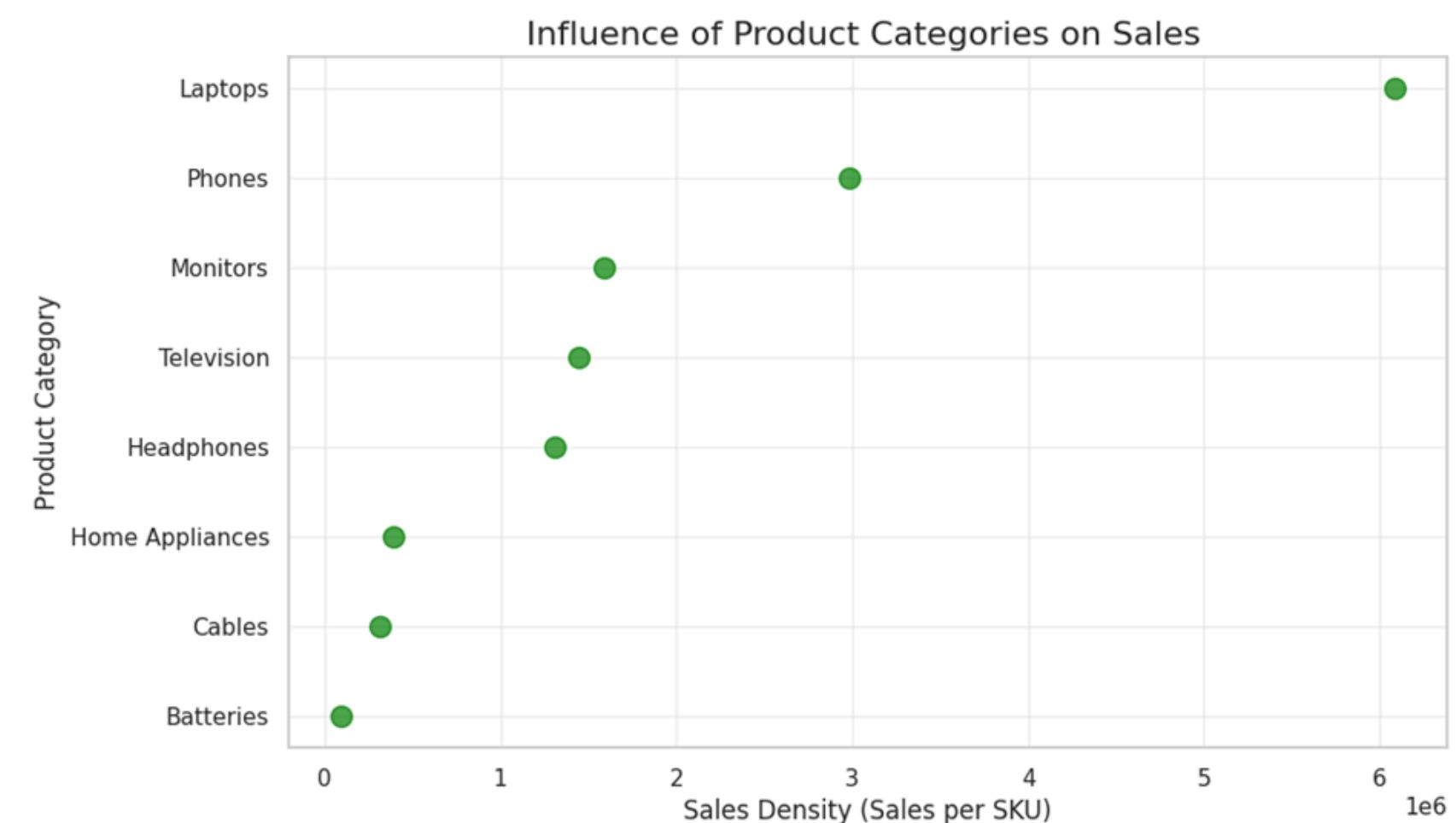
## Hourly and Weekly Trends

- Explanation:
  - Examined sales by hour and day to identify peak shopping times.
- Results and Findings:
  - Peak hours: 6 PM–9 PM.
  - Peak days: Approximately uniform but slightly high on Tuesday, weekends.



## Category Performance

- Explanation:
  - Normalized sales by SKU count to assess profitability of product categories.
- Results and Findings:
  - High-performing categories: Electronics (highest sales), Accessories, and Batteries (high sales density).
  - Opportunity: Improve underperforming categories like some monitors.



# Market Basket Analysis

## Technical Explanation

- Objective: Identify frequently purchased product pairs to design effective bundling strategies.
- Steps:
  - a. Data Preparation: Grouped transactions by Order ID to create product lists and applied one-hot encoding for a binary matrix.
  - b. Frequent Itemset Generation:
    - Utilized the Apriori Algorithm with min\_support = 0.0005 to identify frequent product combinations.
  - c. Association Rule Mining:
    - Generated rules using lift with a threshold of min\_lift = 1.0 to ensure strong relationships between products.
  - d. Visualization: Constructed a network graph to depict associations and highlight the strongest product relationships.

## Results and Findings

- Key Product Pairs:
  - Google Phone ↔ USB-C Charging Cable
    - Support: 0.0056, Confidence: 0.18, Lift: 1.47
  - iPhone ↔ Lightning Charging Cable
    - Support: 0.0057, Confidence: 0.15, Lift: 1.22
  - Vareebadd Phone ↔ USB-C Charging Cable
    - Support: 0.0021, Confidence: 0.18, Lift: 1.45

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

Market Basket Analysis Network Graph



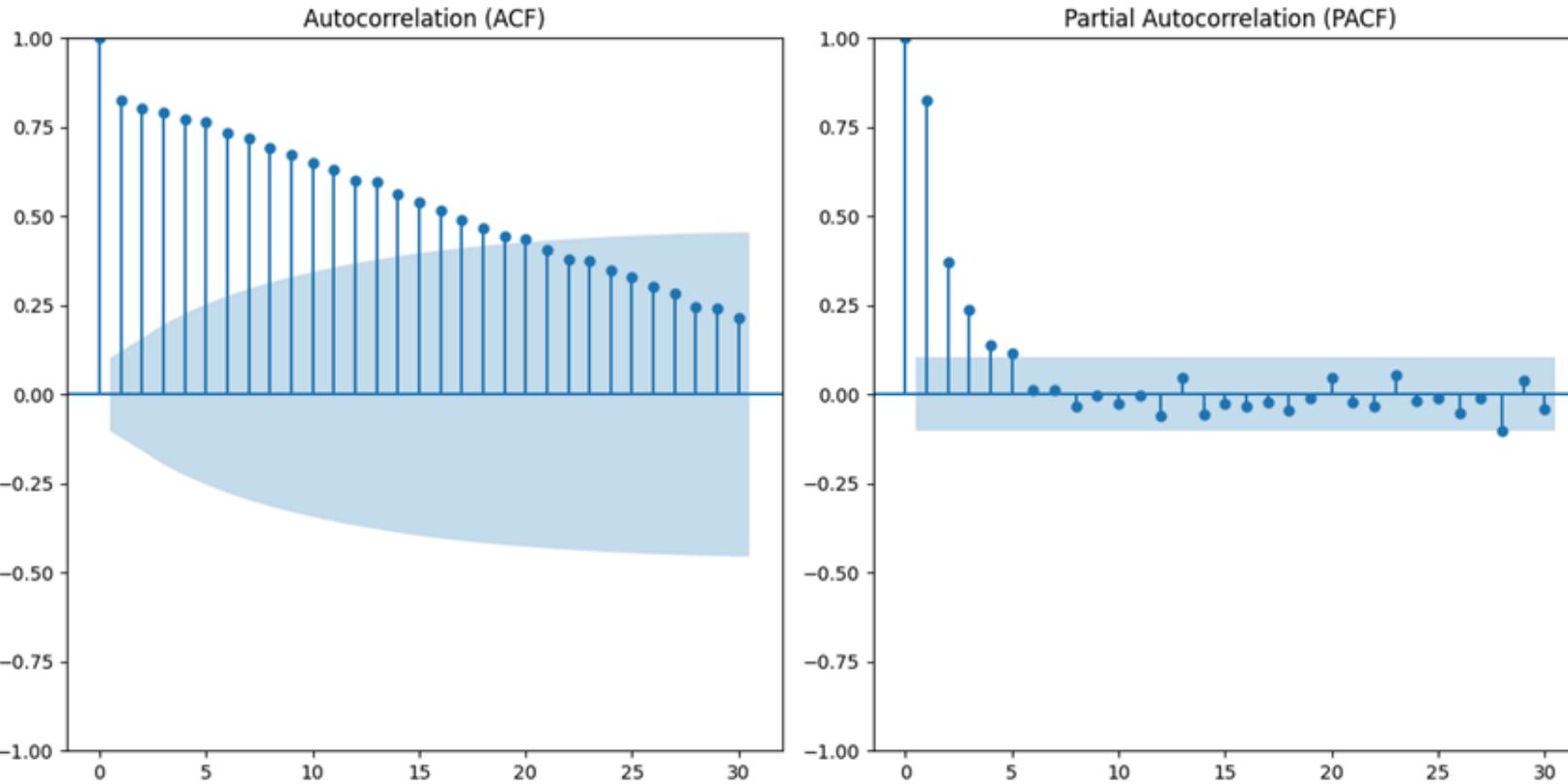
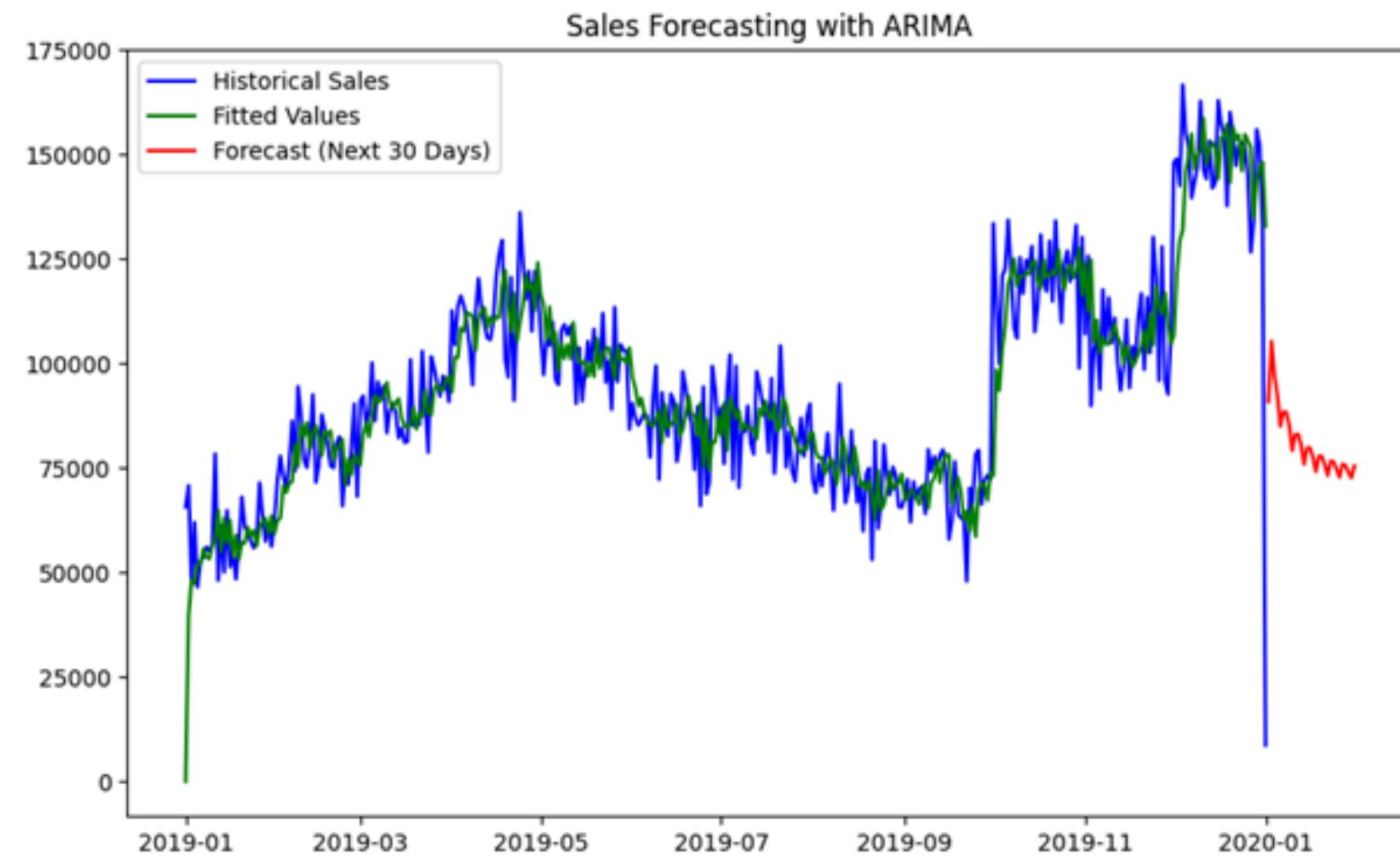
# Demand Forecasting

## Technical Explanation

- Objective: Predict future sales to optimize inventory and operational efficiency.
- Model Used: ARIMA (Auto-Regressive Integrated Moving Average).
  - Stationarity Check:
    - Used the Augmented Dickey-Fuller (ADF) test. Initial data was non-stationary ( $p\text{-value} > 0.05$ ).
    - First-order differencing ( $d=1$ ) applied to stabilize the series.
  - ACF and PACF Analysis:
    - Identified significant lags for AR ( $p$ ) and MA ( $q$ ) components.
    - Model Parameters: ARIMA(6,1,6) selected based on AIC and BIC values for optimal fit.

## Results and Findings

- Validation Metrics:
  - RMSE: 13,207
  - MAPE: 13.28%
  - AIC: 7,964.26
  - BIC: 8,014.96
- Forecast Insights:
  - A short-term dip in sales post-holiday, followed by a gradual recovery.
  - Actionable 30-day forecast for inventory and marketing planning



ADF Statistic	-2.3508591590746564
p-Value	0.1560888501412367

# Recommendations

- **Targeted Marketing Campaigns:**
  - Focus promotions during December to maximize holiday sales.
  - Introduce discounts and bundle offers in February to boost low sales.
  - Leverage peak shopping hours (6 PM–9 PM) and weekends for time-sensitive promotions.
- **Regional Strategies:**
  - Enhance logistics and delivery efficiency in top-performing cities (San Francisco, New York, Los Angeles).
  - Implement localized campaigns in low-performing cities like Austin to increase revenue.
- **Product Bundling:**
  - Bundle complementary products (e.g., Google Phone + USB-C Charging Cable) to increase cart value.
  - Highlight high-lift bundles in advertisements to drive cross-selling opportunities.
- **Inventory Optimization:**
  - Use ARIMA forecasts to align stock levels with anticipated demand, reducing overstocking and stockouts.
  - Prioritize stocking high-demand products (e.g., Accessories, Batteries) during peak periods.
- **Category-Specific Focus:**
  - Increase marketing for high-density categories like Accessories and Batteries.
  - Promote underperforming products with targeted discounts or repositioning.