SuperStore Sales Analysis

Data Visualization & Dashboard Project

Submitted to:
Mr. Nitin Shelke

Submitted By:
Shubhangi Kundu
E23MCAG0078

Table Of Contents

Project Objective

The project endeavors to contribute significantly to the business's success through the strategic application of advanced data analysis techniques, specifically concentrating on time series analysis. The primary objective is to furnish the organization with valuable insights and accurate sales forecasting. By leveraging meticulous analytical processes and technological proficiency, the project aims to enhance decision-making capabilities, thereby fortifying strategic planning and fostering sustainable business growth and prosperity.

Methodology

Data Pre-Processing in Python

Data Visualization using python

Importing pre-processed Csv file to Power BI Desktop

Analysing the Data in Power BI
(using suitable visualization tools, filters and DAX Queries)

Concluding Insights

Data Preprocessing in Python

The initial step involves preparing and cleaning the data to ensure accurate and meaningful insights. I will delve into the essential aspects of data preprocessing, focusing on feature scaling, handling null values, and the impactful choices made to refine our dataset.

1. **Feature Scaling**

   Feature scaling is a critical preprocessing step to standardize the range of independent variables or features of the dataset.

2. **Null value Handling**

   Null values in a dataset can compromise the integrity of our analysis. We meticulously examined our dataset for any missing values. Detecting null values is pivotal for understanding the data completeness and reliability.

3. **Null Value Imputation**

After identifying null values, we implemented a strategic approach to handle them. Leveraging the powerful “fillna” function, we filled the missing values with appropriate measures.

4. **Removing Duplicate values**

Duplicates in a dataset can adversely impact data integrity and analysis outcomes. Identifying and eliminating duplicates is essential to ensure that statistical measures and insights are based on genuine data, preventing skewed results.

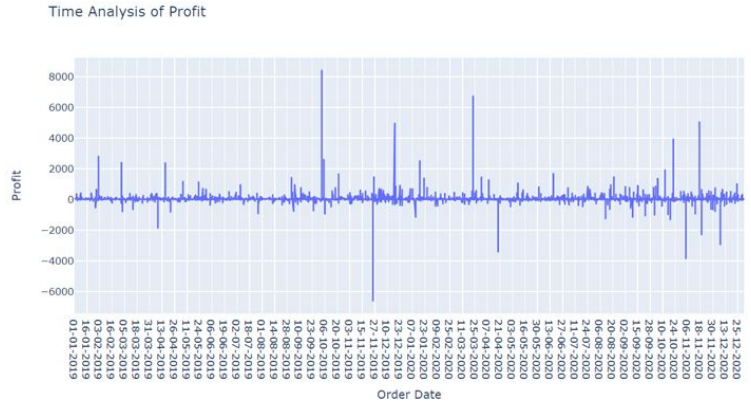
5. **Saving clean data in CSV file**

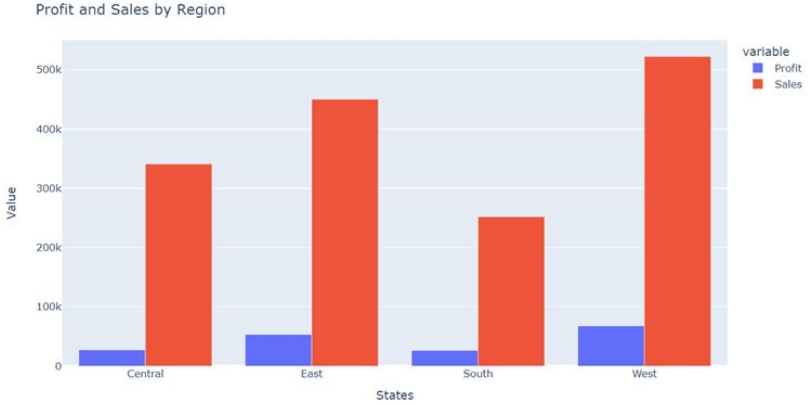
Using the “to_csv” we saved the data in a csv file for using it in Power Bi.

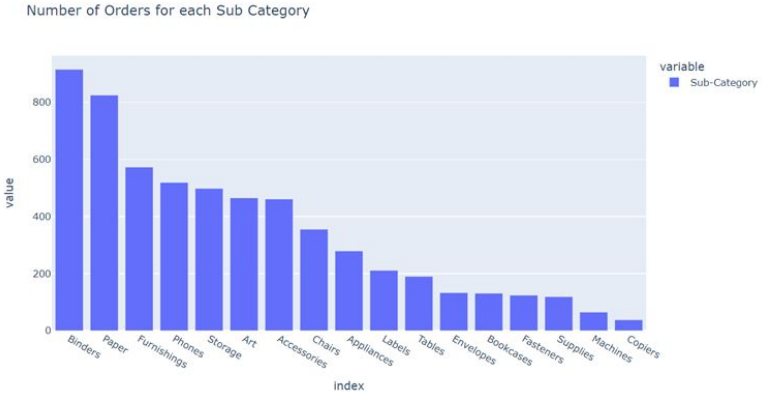
Libraries used in Python

- **Pandas**- Pandas is a Python library for data manipulation and analysis. It offers data structures like DataFrame for efficient data handling, cleaning, and analysis. Pandas is widely used for working with structured data.
- **Plotly Express**- Plotly Express is a high-level data visualization library built on Plotly. It enables the creation of interactive and expressive visualizations with a concise syntax. Plotly Express supports various chart types and is known for its ease of use.
- **Matplotlib.pyplot**- Matplotlib.pyplot is a popular 2D plotting library in Python. It provides a MATLAB-like interface for creating static, animated, and interactive visualizations. Matplotlib is highly customizable and is often used in conjunction with other libraries for data visualization.
- **Seaborn**- Seaborn is a statistical data visualization library based on Matplotlib. It simplifies the process of creating informative and attractive statistical graphics. Seaborn comes with built-in themes and color palettes to enhance the visual appeal of plots and is particularly useful for working with complex datasets.
- **CSV (Comma-Separated Values) Library**- CSV is a file format commonly used for tabular data storage. In Python, the built-in csv module provides functionality for reading from and writing to CSV files. It simplifies the process of handling comma-separated data, allowing easy manipulation and interaction with tabular datasets in Python programs.

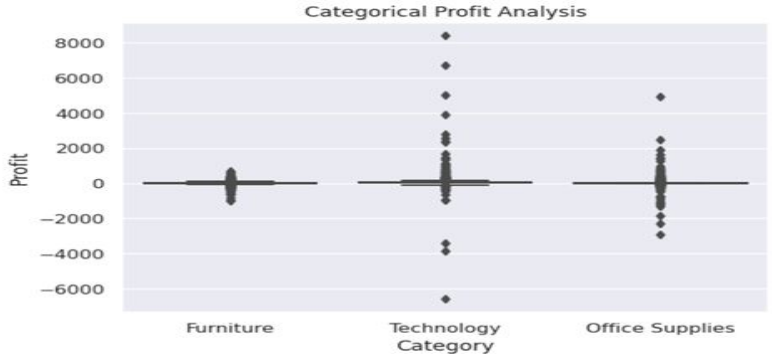
Data Visualization using Python

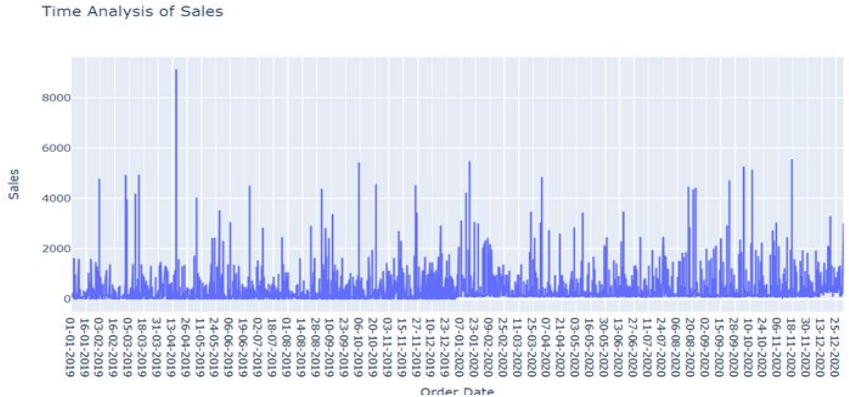


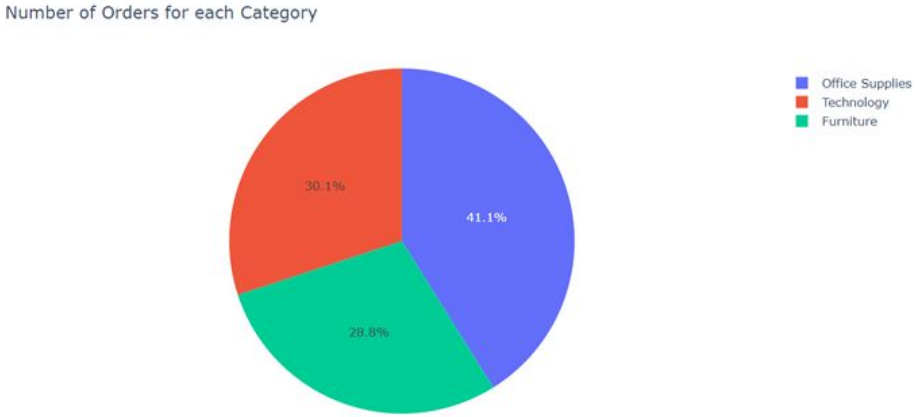


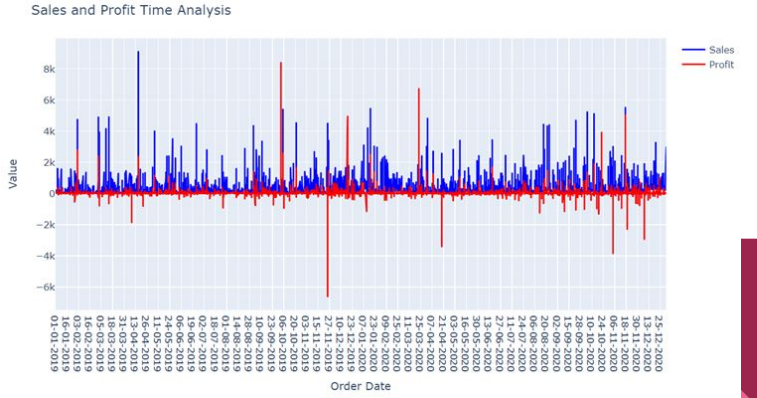


Data Visualization using Python

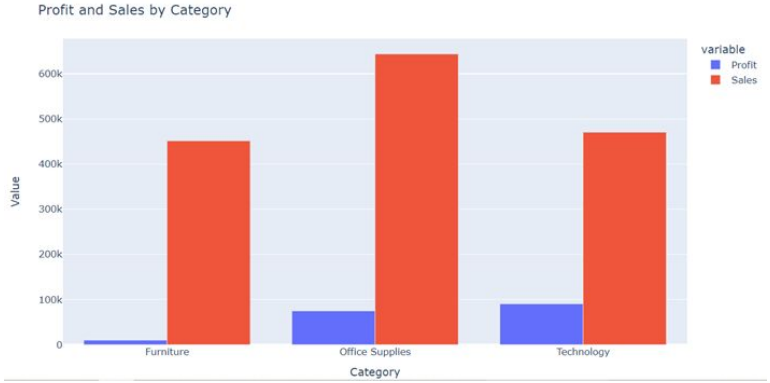


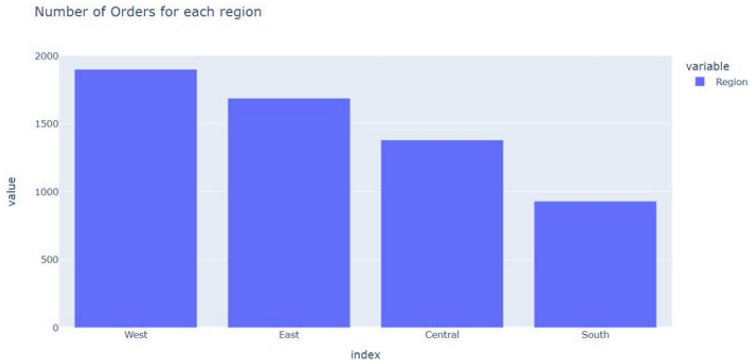


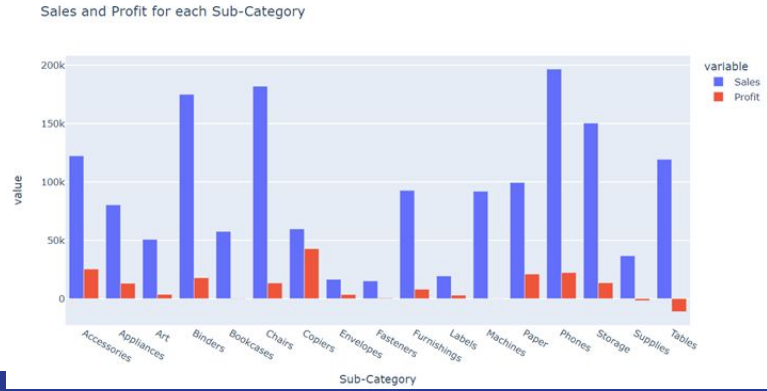




Data Visualization using Python







Dashboard Creation and Visualization in PowerBI

1. **Stacked Area Chart:**

   A stacked area chart is used to represent the cumulative total of multiple series over time. It's effective for showing the composition of a whole over a continuous axis.

2. **Clustered Bar Chart:**

   A clustered bar chart is used to compare values across categories. Each category has its set of bars, and bars within the same category are grouped together.

3. **Donut Chart:**

   A donut chart is a variant of a pie chart with a hole in the center. It's useful for displaying the proportion of parts to a whole.

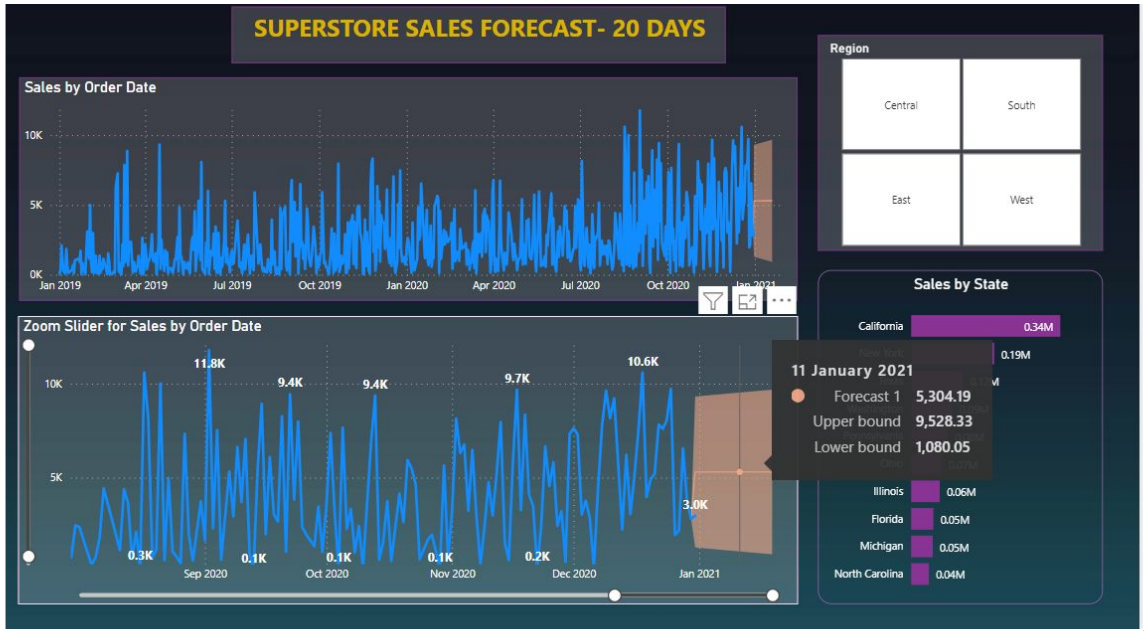
4. **Map:**

   A map visualization represents the data points on a geographical map . It’s ideal for displaying spatial distributions and trends.

Dashboard Creation and Visualization using PowerBI

Forecasting using PowerBI

Forecasting in Power BI involves using predictive analytics to estimate future trends based on historical data. Power BI provides built-in forecasting capabilities that make it easy to generate forecasts for your data.



Insights

From PowerBI:

1. Preferred Payment Method:
    - Cash on Delivery (COD) emerges as the most favored payment method across the country, surpassing other payment options.
2. Consumer Dominance:
    - The consumer segment stands out as the primary contributor to overall sales, demonstrating its significant impact on revenue.

From Python:

1. Product Categories:
    - Office supplies dominated sales across all categories.
    - Technology category led in terms of maximum profit.
2. Regional Impact:
    - The West region emerged as the top contributor for both sales and profits.

Recommendations

1. **Revamp Payment Processes for Better COD Experience:**

   - Identify and address any pain points in the Cash on Delivery (COD) process to enhance customer satisfaction. This could include faster processing, real-time tracking, and clearer communication on payment expectations.

2. **Encourage Digital Payments with Incentives:**

   - To steer customers towards card payments, introduce enticing incentives such as exclusive vouchers or discounts. Communicate the benefits of using cards, emphasizing the added value they bring to the overall shopping experience.

3. **Tailor Marketing Efforts for Diverse Consumer Preferences:**

   - Diversify marketing strategies to cater to different consumer segments. Conduct surveys or customer feedback sessions to understand specific preferences and refine campaigns accordingly. This personalized approach can enhance engagement and drive sales.

Result

To optimize overall performance, businesses should leverage the preference for standard shipping and strategically plan marketing and inventory management around the peak sales and profit months. By understanding these insights and recommendation, businesses can refine their strategies, enhance customer experiences, and capitalize on key market trends for sustained success.

Thank You.