

PYTHON & POWER BI EXERCISE

GLOBAL SUPERSTORE DATASET

DHANUSHA SRI M J

20MIY0054

Introduction

The "Global Superstore Sales Analysis" project leverages Python to explore and visualize sales data, aiming to uncover significant trends and insights. The analysis includes data cleaning, transformation, and visualization, providing a comprehensive view of sales performance, customer behavior, and regional differences.

Detailed Analysis

1. Data Exploration

- Libraries and Data Importation: Utilized NumPy, pandas, seaborn, and matplotlib for data handling. The dataset includes 24 columns and 51,290 rows.
- Column Overview: Important columns include Order Date, Ship Mode, Customer ID, Segment, Category, Sales, Quantity, and Profit.
- Data Inspection: Used data.info() and data.describe() to examine the dataset structure, detect missing values, and obtain summary statistics.
- Order Priority Distribution: Analyzed the frequency of each order priority level to understand its distribution.

2. Data Cleaning

- Copying and Renaming Columns: Created a copy of the dataset and renamed columns for clarity and consistency.
- Date Conversion: Converted order_date and ship_date from strings to date formats to facilitate time-based analysis.
- Adding Sales Year Column: Added a sales_year column to segment data by year.
- Categorical Data Optimization: Converted categorical columns to the category datatype to improve processing efficiency.

3. Data Analysis

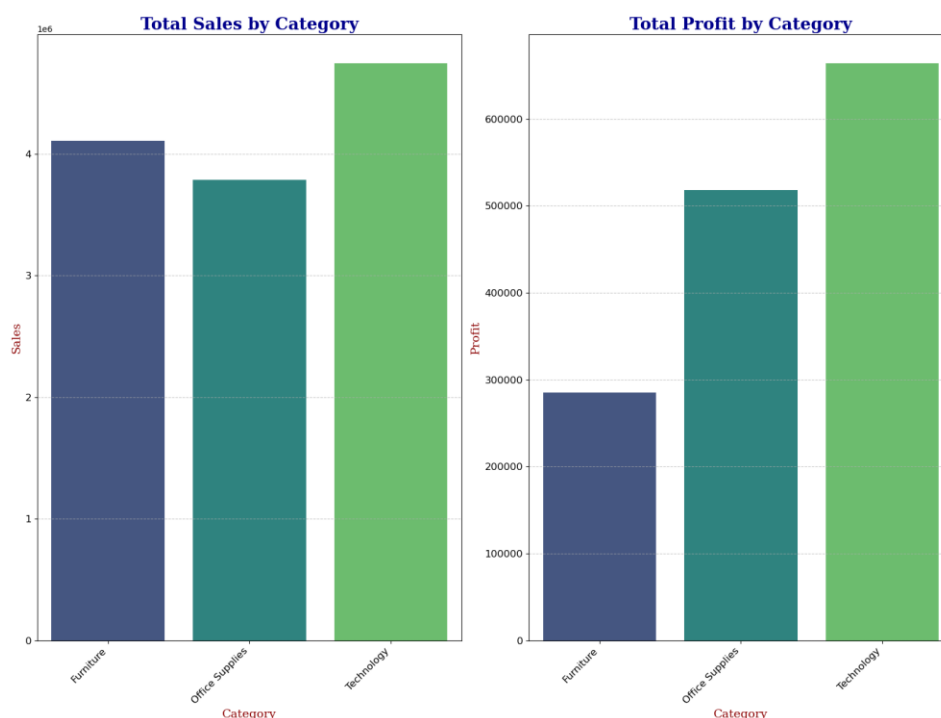
- Correlation Analysis: Calculated the correlation between shipping costs and sales, as well as shipping costs and profits, to understand their relationships.
- Order Priority and Profit Analysis: Aggregated sales and profit by order priority and category to identify high-performing segments.

- Sales Performance by Region: Aggregated sales data by region to analyze regional performance.
- Customer Segmentation: Analyzed sales and profit by customer ID, segmented customers based on sales volume into 'Low', 'Medium', and 'High' categories.
- Top Products: Identified the top 5 products by sales and profit.
- Profit Margin Analysis: Calculated profit margins for product categories and sub-categories, sorting by profitability.
- Segment Contributions: Analyzed sales contributions by segment to understand their impact.
- Monthly Trends: Analyzed sales and profit trends on a monthly basis to observe temporal variations.

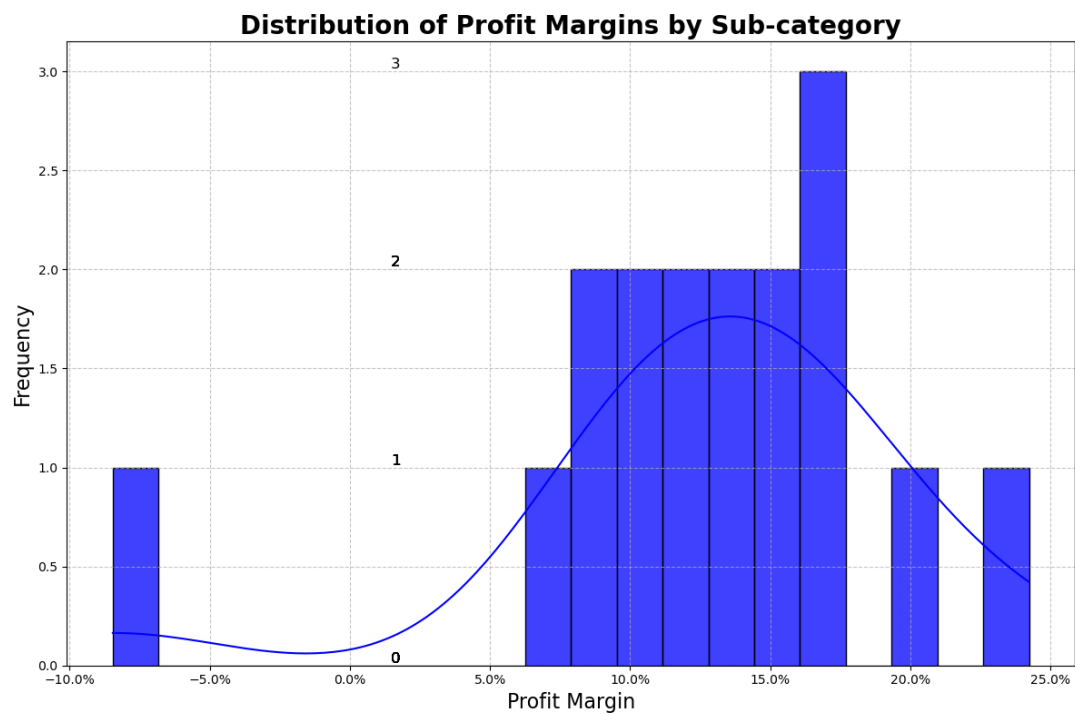
4. Data Visualization

- Total Sales and Profit by Category: Created bar charts to visualize total sales and profit across categories.
Profit Margin Distribution: Plotted a histogram of profit margins by sub-category, showing the frequency distribution.
- Sales by Category per Segment: Used a stacked bar chart to display sales distribution by category across different customer segments.
- Monthly Sales and Profit Trends: Created line charts to visualize monthly trends in sales and profit.
- Sales Contribution by Segment: Visualized sales contributions by segment using a bar chart.

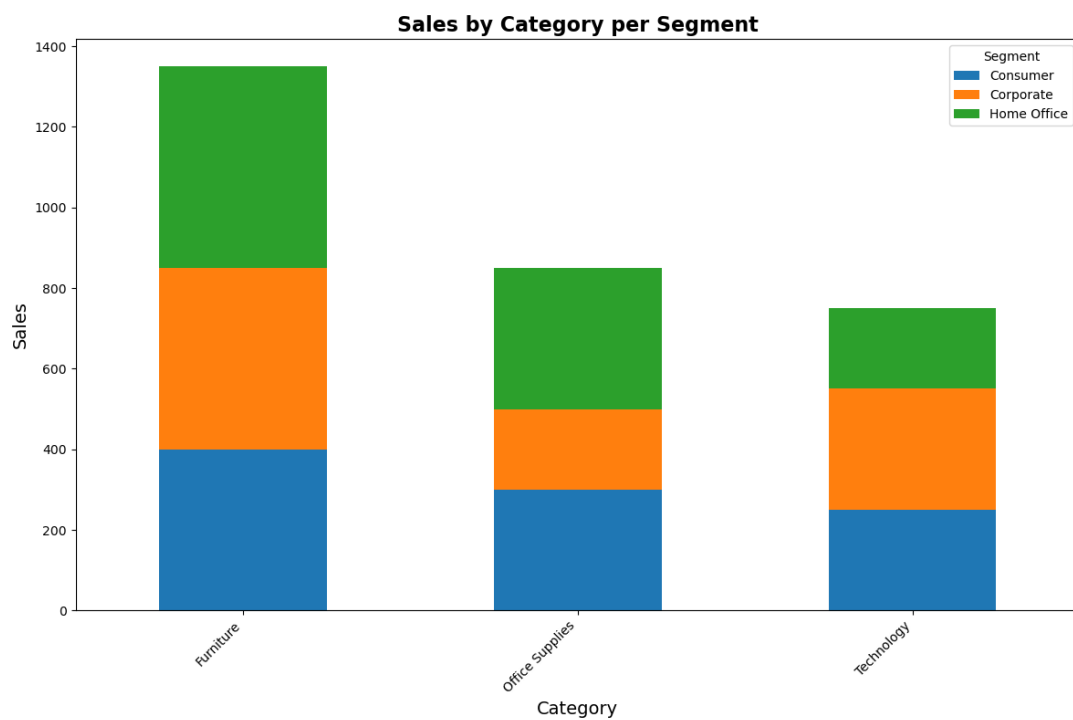
BAR CHART FOR TOTAL SALES AND PROFIT BY CATEGORY



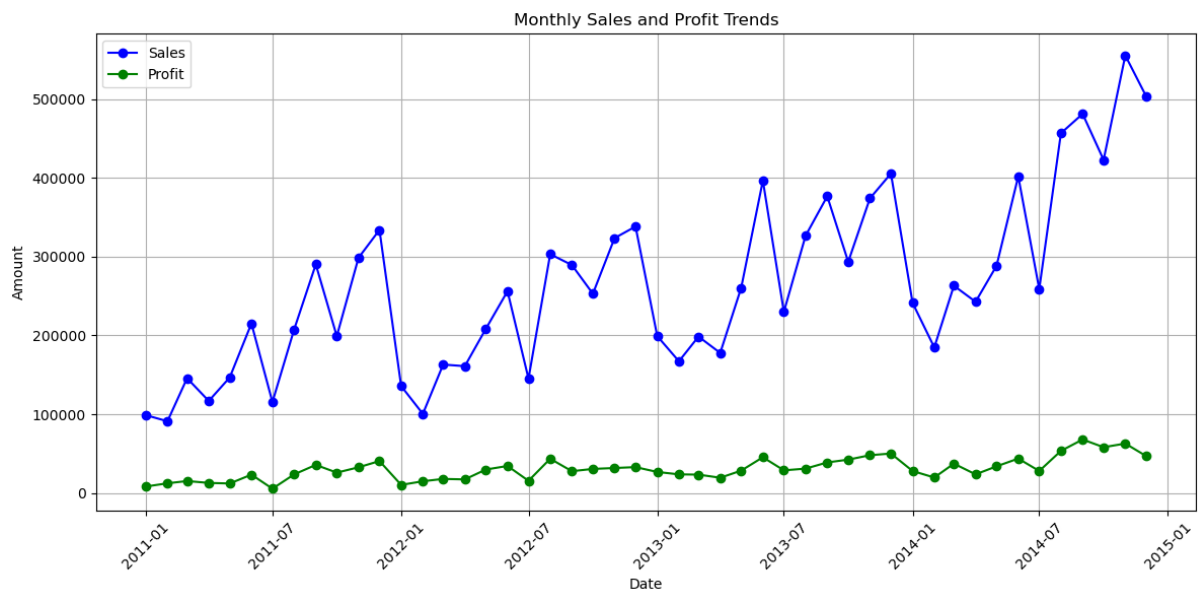
HISTOGRAM FOR DISTRIBUTION OF PROFIT MARGINS BY SAUB CATEGORY



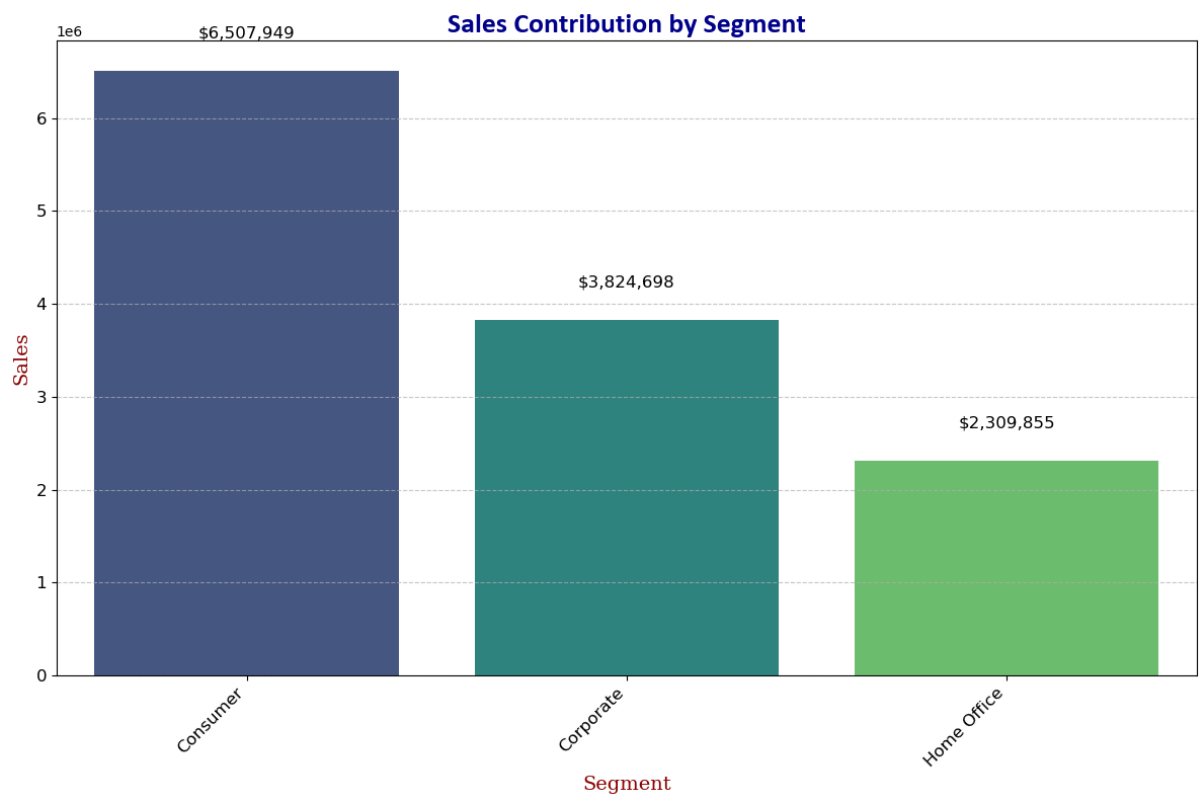
DISTRIBUTION OF SALES BY CATEGORY PER SEGMENT



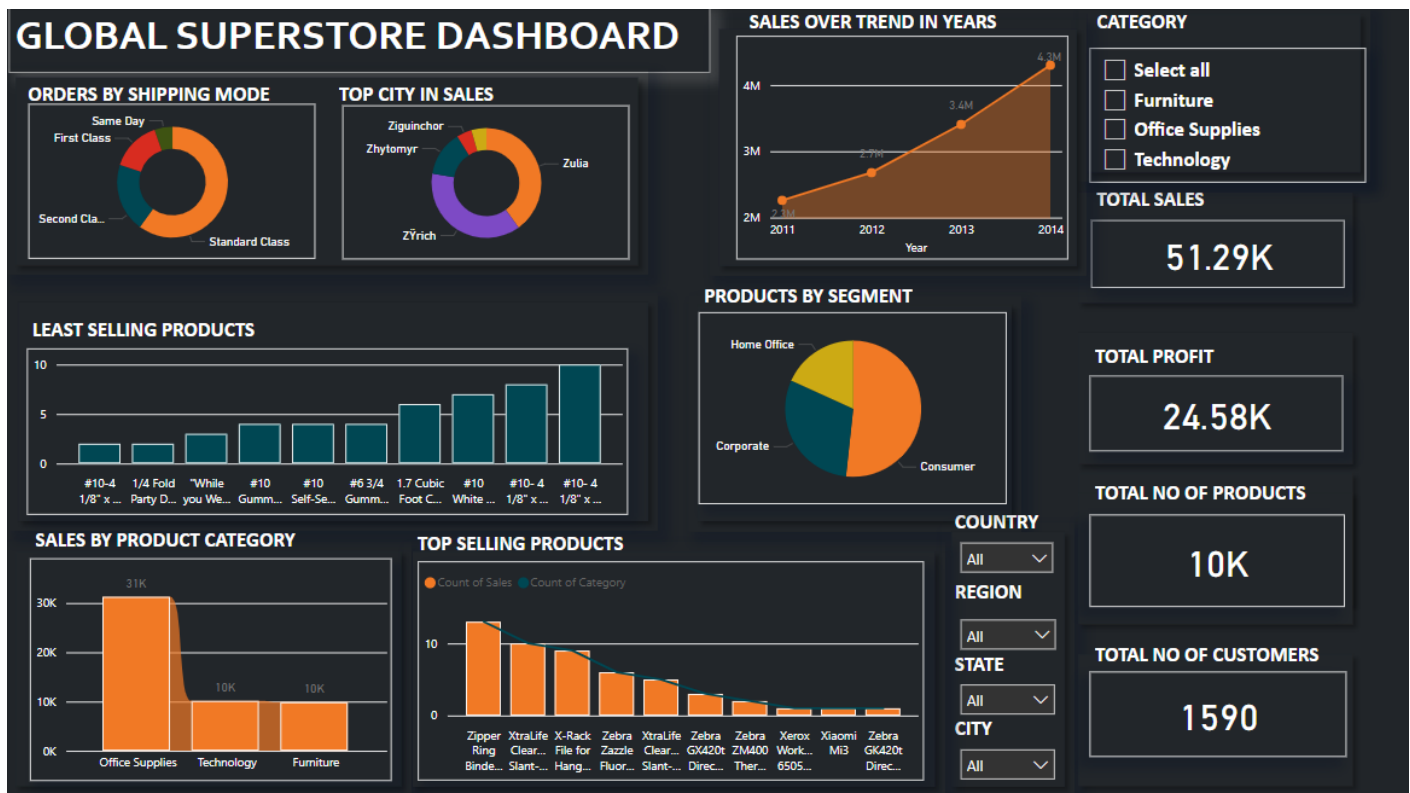
MONTHLY SALES AND PROFIT TRENDS



BAR CHART FOR SEGMENT AND SALES



POWER BI DASHBOARD



5. Conclusion

The **Global Superstore Sales Analysis** project provides valuable insights into sales performance, customer behavior, and regional differences. Key findings include:

- ⇒ Understanding how shipping costs impact sales and profits.
- ⇒ Identifying top-performing categories and products.
- ⇒ Analyzing customer segments and their sales contributions.
- ⇒ Observing trends in sales and profit over time.

The visualizations and analyses support informed decision-making and strategic planning, offering a clear picture of sales dynamics within the global superstore context.

6. Reference

- https://docs.google.com/spreadsheets/d/1KagwoQLy1quKvT_82amuS-x3UnsoIX4J6p02ewbjQNA/edit?usp=sharing
- https://docs.google.com/document/d/1luBvcyB81k5Ucz2QcZjlr8xQRJpdpMhMLSVVYGX_k78/edit?usp=sharing
- <https://www.geeksforgeeks.org/exploratory-data-analysis-in-python/>
- <https://www.analyticsvidhya.com/blog/2015/04/comprehensive-guide-data-exploration-sas-using-python-numpy-scipy-matplotlib-pandas/>

- <https://medium.com/@jscvcds/data-exploration-in-python-with-examples-30a5324472aa>
- <https://openstax.org/books/introduction-python-programming/pages/15-5-data-visualization>