Project Report

Enhancing Business Insights through Data Analysis

Created by: Suma Anumolu

Table of Contents

1. Introduction

In today's competitive business landscape, data-driven decision-making is paramount for achieving growth and operational efficiency. This project undertakes a comprehensive analysis of sales data from the United States (US), sourced from Kaggle, aiming to extract actionable insights that drive strategic business decisions.

2. Objectives and Methodology

Objective

The primary goal of this project is to leverage data analytics to uncover key trends and patterns within the sales dataset, enabling informed business strategies.

Approach

1. **Data Import and Cleaning:**
   - **Data Source:** Kaggle dataset imported into Jupyter Notebook using pandas.
   - **Cleaning:** Addressed missing values, transformed data types, and standardized representations for data integrity.
   - **Data Enhancement:** Calculated metrics such as discount price, sale price, and profit to facilitate deeper analysis.
2. **Data Visualization:**
   - **Time Series Analysis:** Analyzed month-over-month sales growth (2022-2023) to identify seasonal trends and optimize resource allocation.
   - **Regional Performance:**
   - **Category Performance:** Determined the highest sales month for each product category to target promotional efforts effectively.
   - **Profit Analysis:** Explored category and subcategory profit margins to optimize pricing and product mix.
   - **Discount Impact:** Investigated the relationship between discounts and profitability to refine pricing strategies.
   - **Shipping Mode Efficiency:** Evaluated the effectiveness of different shipping modes in meeting customer expectations and managing costs.
3. **Correlation Matrix Analysis:**
   - Utilized seaborn to visualize correlations between discount, quantity, sale price, and profit, providing deeper insights into factors influencing business performance and profitability.
4. **Database Integration:**
   - **SQL Database:** Imported cleaned and enhanced data into SQLite3 for scalable querying and further analysis.
   - **MySQL Server Integration:** Detailed steps on connecting to MySQL server and exporting data from Python to MySQL Workbench, enabling robust querying capabilities and collaborative data analysis.
5. **Queries and Reports:**
   - Executed queries to derive insights such as:

- Top 10 Highest Revenue Generating Products
- Top 5 Highest Selling Products in Each Region
- Month-over-Month Growth Comparison (2022 vs. 2023)
- Highest Sales Month for Each Category
- Subcategory with Highest Profit Growth (2023 vs. 2022)

3. Data Import and Cleaning

Data Source

The dataset used for this project, named "Retail-orders.csv," was sourced from Kaggle. The dataset was downloaded using the Kaggle API command and then extracted for analysis.

Data Cleaning

1. **Importing the Dataset:**
   - The dataset was imported into a Jupyter Notebook using pandas.
2. **Unzipping the File:**
   - The downloaded dataset was in a zipped format. It was unzipped and the extracted file was loaded into a pandas DataFrame.
3. **Initial Data Exploration:**
   - Conducted an initial exploration of the data using the head() function to display the first few rows of the dataset.
   - Checked for null values in the dataset to identify any missing data that needed to be addressed.
4. **Handling Missing Values:**
   - Found that there were missing values in the dataset.
   - Specified na_values parameter to handle specific missing value indicators like 'Not Available' and 'unknown'.
5. **Normalizing Column Names:**
   - Standardized the column names to lower case and replaced spaces with underscores to ensure consistency and ease of use in subsequent analyses.
6. **Exploring Data Types:**
   - Explored the data types of each column to ensure they were appropriate for the analyses to be performed.
   - Converted the 'order_date' column to datetime format for accurate time series analysis.

Data Enhancement:

1. **Calculated Metrics:**
   - To facilitate deeper analysis, additional metrics were calculated:
     - **Discount Price:** Calculated as a percentage of the list price.
     - **Sale Price:** Determined by subtracting the discount from the list price.
     - **Profit:** Computed as the difference between the sale price and the cost price.
2. **Creating New Columns:**

- o Created new columns for the calculated metrics:
       - ▪ discount
       - ▪ sale_price
       - ▪ profit
3. **Dropping Unnecessary Columns:**
   - o Dropped columns that were deemed least useful for the analysis to streamline the dataset and focus on the most relevant data:
       - ▪ list_price
       - ▪ cost_price
       - ▪ discount_percent

4. Data Visualization

Overall Sales Performance and Growth: Across the years 2022 and 2023, we observed an overall increase in revenue, indicating positive growth in sales performance. This growth trajectory suggests effective operational strategies and market responsiveness.

Figure 1:
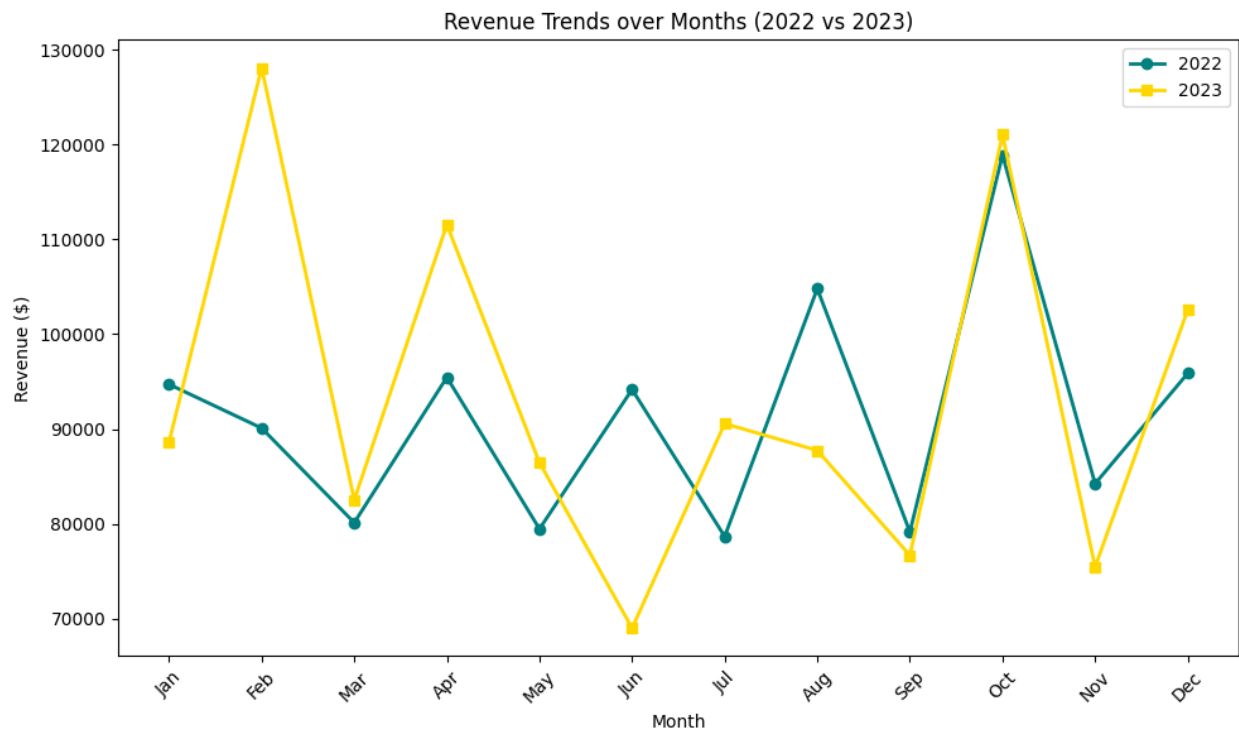
Regional Performance Insights: Regional analysis revealed significant variations in performance:

- **East and South Regions:** Both regions experienced substantial sales growth in 2023 compared to 2022, highlighting promising markets for expansion and resource allocation.

- **Central Region:** Maintained stable sales performance across both years, indicating a mature market with consistent demand.
- **West Region:** Witnessed a decline in profit in 2023, signaling a need for deeper investigation into potential market challenges or competitive pressures.

Figure 2:

Profit Metrics by Region and Year
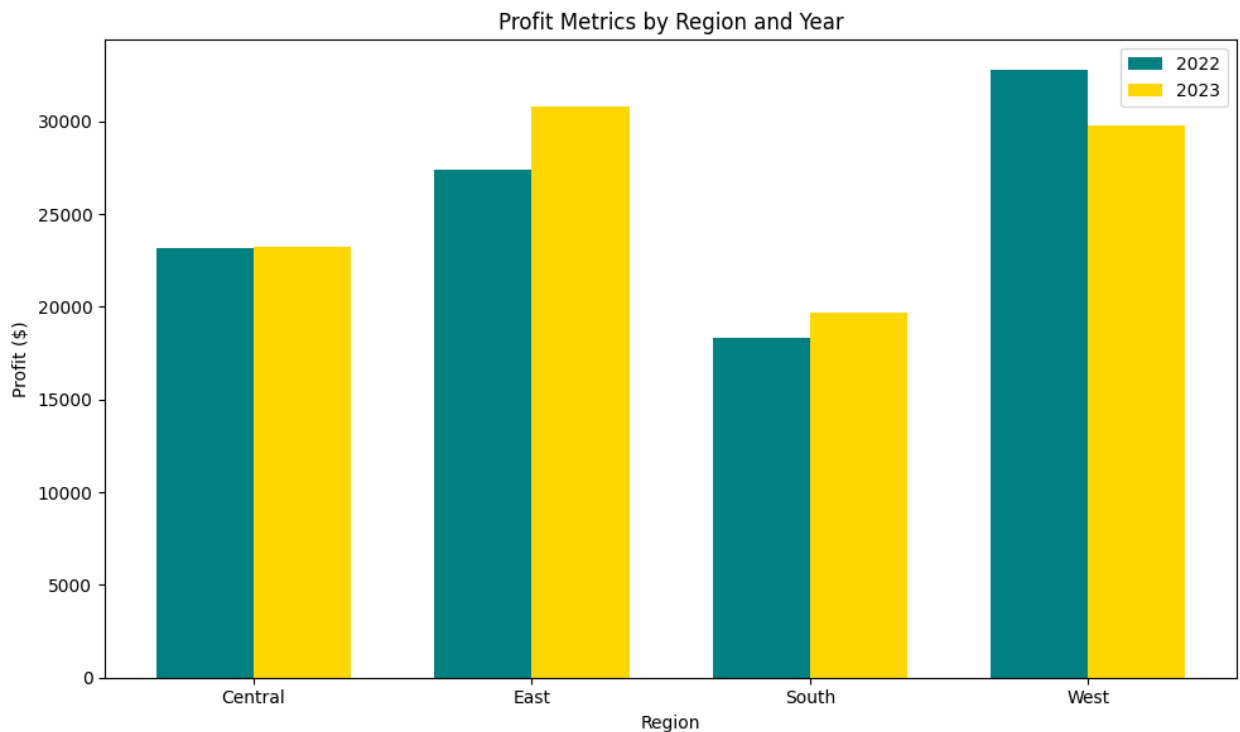
Category Performance:

- **Top profit Generating Categories:** Technology and office supplies has high profits. This insight underscores the importance of focusing resources and promotional efforts in these categories to maximize revenue growth (Figure 3).
- **Top Selling Categories:** Office supplies dominated sales across all regions, with technology products leading in the West. The absence of furniture in top-selling lists suggests opportunities for targeted marketing or product enhancement strategies within this category (Figure 4).
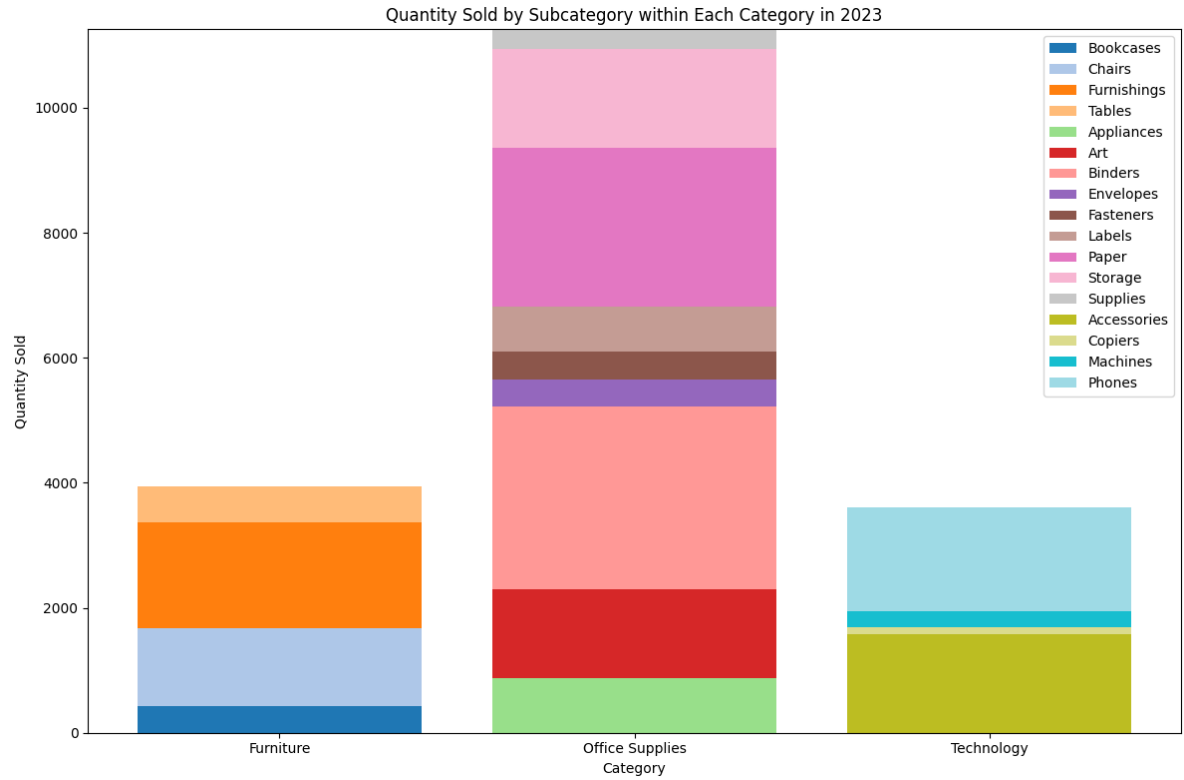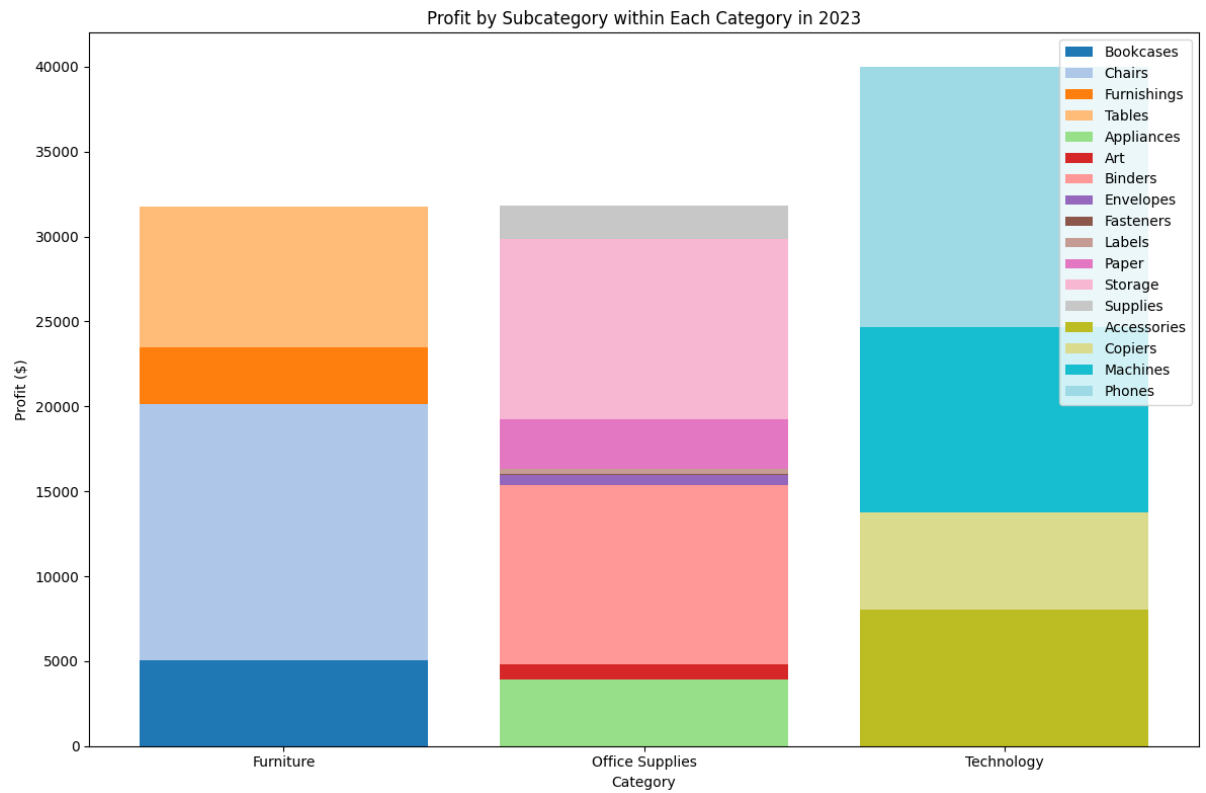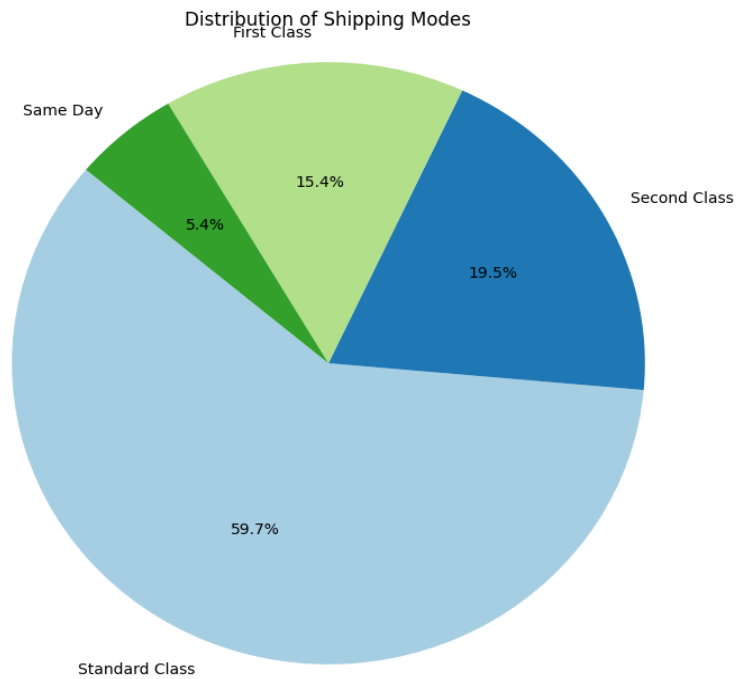
Figure 3:

Figure 4:

Shipping Efficiency and Discount Impact:

- **Shipping Modes:** Standard Class shipping demonstrated the highest efficiency in meeting customer expectations and managing costs, underscoring its reliability and cost-effectiveness.
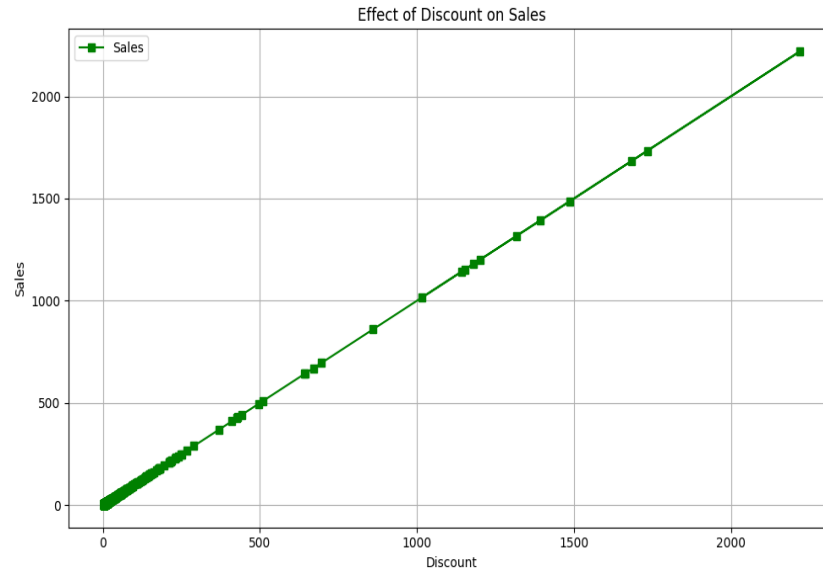
  Figure 5:

Distribution of Shipping Modes

- **Discount Impact:** A positive correlation between discounts and sales quantity was observed, indicating that strategic discounting can effectively boost sales volumes. However, careful management is crucial to mitigate potential impacts on profit margins.

  Figure 6:

Effect of Discount on Sales

Correlation Matrix Insights:

- The correlation matrix emphasized relationships between key variables:
  - **Quantity and Discount:** Higher discounts correlated with increased sales volume.
  - **Discount and Sale Price:** Discounts were closely tied to lower sale prices.
  - **Sale Price and Profit:** Higher sale prices generally led to increased profitability.
  - **Profit and Discount:** While discounts stimulated sales, they also influenced profit margins, highlighting the need for balanced pricing strategies.

Figure 7:

Correlation Heatmap
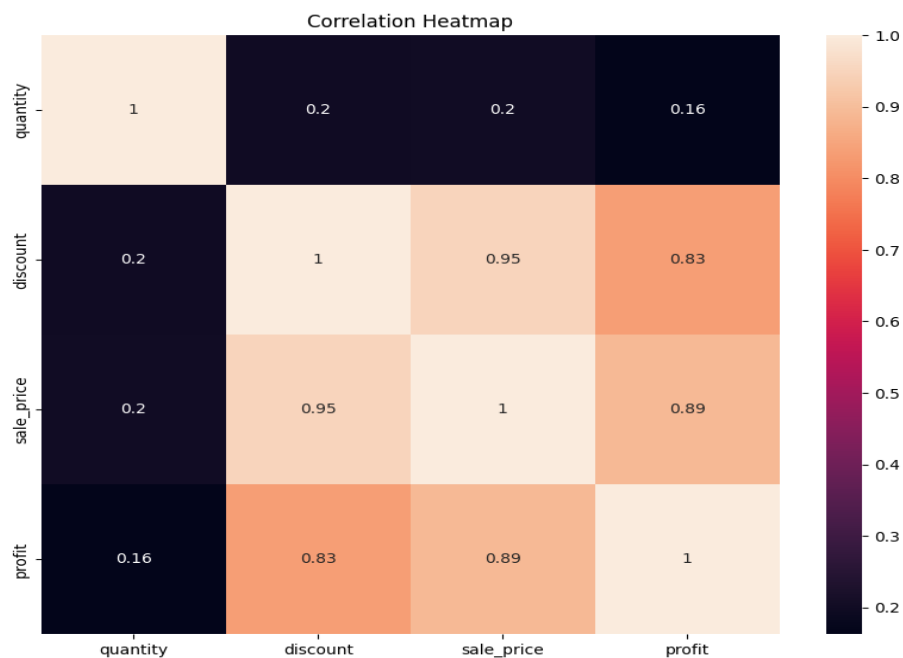
5. Database Integration

SQL Database

Imported the cleaned and enhanced data into SQLite3 for scalable querying and further analysis.

MySQL Server Integration

Detailed the steps on connecting to a MySQL server and exporting data from Python to MySQL Workbench. This integration facilitates robust querying capabilities and collaborative data analysis.

6. Advanced Data Analytics using SQL

Top 10 Highest Revenue Generating Products

Insights from this analysis highlight that technology and office supplies were the top revenue-generating product categories. This information guides strategic decisions on where to invest resources and promote products for maximizing revenue growth.

Top 5 Highest Selling Products in Each Region

Analysis revealed regional preferences, with office supplies dominating sales across all regions. In the West region, technology products emerged as top sellers followed by office supplies.

The absence of furniture in the top-selling products suggests potential areas for focused marketing or product improvement strategies within that category.

Month-over-Month Growth Comparison (2022 vs. 2023)

Comparative analysis between 2022 and 2023 indicated varying sales trends over months.

Highest Sales Month for Each Category

This analysis highlights peak sales months for each category: Furniture led in October 2022, Office Supplies peaked in February 2023, and Technology saw its highest sales in October 2023.

Subcategory with Highest Profit Growth (2023 vs. 2022)

This analysis identifies the subcategories that experienced the highest profit growth from 2022 to 2023. Machines led with a substantial growth of 50.2%, followed by Supplies, Binders, Storage,

and Phones.

7. Strategic Recommendations:

Based on these insights, strategic recommendations for maximizing growth and profitability include:

- **Investing in High-Performing Categories**: Focus resources on technology and office supplies to capitalize on their revenue-generating potential.
- **Targeting Regional Growth Opportunities**: Expand operations in the East and South regions to leverage their strong sales growth.
- **Improving Product Performance**: Enhance offerings in underperforming categories like furniture through targeted marketing and product innovation.
- **Optimizing Pricing and Discounts**: Implement balanced discounting strategies to stimulate sales while safeguarding profit margins.
- **Enhancing Operational Efficiency**: Continue to prioritize Standard Class shipping for its efficiency in meeting customer expectations and reducing costs.

8. Conclusion:

In conclusion, this analysis provides a clear picture of the company's sales dynamics, highlighting areas of strength and opportunities for improvement. By leveraging these insights, stakeholders can make informed decisions to drive sustainable growth, optimize resource allocation, and maintain competitive advantage in the market.

This report serves as a foundation for strategic planning and operational refinement, emphasizing the importance of data-driven decision-making in achieving long-term business success.

9. Skills highlighted:

- **ETL (Extract, Transform, Load):** Demonstrated proficiency in importing, cleaning, and enhancing data for analysis.
- **Data Analysis:** Expertise in using Python (pandas, matplotlib, seaborn) for comprehensive data exploration and visualization.
- **Business Acumen:** Ability to translate complex data findings into actionable insights for strategic decision-making.
- **SQL Proficiency:** Competence in utilizing SQL for database management, querying, and generating analytical reports.