



Business Intelligence Analysis

by Nataliia Rudnikova

08/09/2025

Table of Contents

- Introduction
- Methodology
- Results
- Discussion
- Conclusion
- Appendix





Introduction

Introduction

The data from a USA-based retail business was analyzed, cleaned, and visualized. An interactive dashboard was created using Microsoft Excel and Tableau to inform better business decisions.





Methodology

Data Details

The data consists of 6 .csv files, the main one regarding sales and transactions and the other 5 being auxiliary regarding details about products, stores and locations.

Module 1: Data Cleaning and Preparation Using Excel

Data cleaning in Excel

- Spell check
- Empty rows: Select all -> Filter -> check only Blanks -> Delete
- Duplicate rows: Select a column which should not have any duplicates -> Conditional formatting -> Highlight duplicate rows
- Whitespace: Select all -> Replace -> *double space* to *single space*; use =TRIM() as with casing for spaces at the beginning and end of cells
- Remove duplicates

Module 2: Data Querying and Analysis Using PostgreSQL

- **Data Querying:**
 - Utilize **SQL queries** to extract and manipulate data from the PostgreSQL database.
 - Perform checks to ensure data is populated in the tables.
- **Data Analysis:**
 - Create **data cubes** using the ROLLUP function to summarize data along hierarchies.
 - Analyze sales trends over time and across various regions to identify patterns.
- **Results Interpretation:**
 - Interpret the results of the SQL queries to derive insights that inform business strategies.

Module 3: Data Visualization and Statistical Analysis

- **Data Visualization:**
 - Use **Microsoft Excel** to create various types of charts, including:
 - Line charts
 - Bar charts
 - Sunburst charts
 - Scatter plots
 - These visualizations help in presenting data trends and comparisons clearly.
- **Statistical Analysis:**
 - Conduct **regression analysis** to understand relationships between variables.
 - Interpret key statistical metrics such as **p-values** and **R-squared values** to assess the strength and significance of the relationships.
- **Insights Communication:**
 - Use the visualizations and statistical results to communicate insights effectively, aiding in decision-making processes.

Module 4: Data Visualization and Dashboards Using Tableau

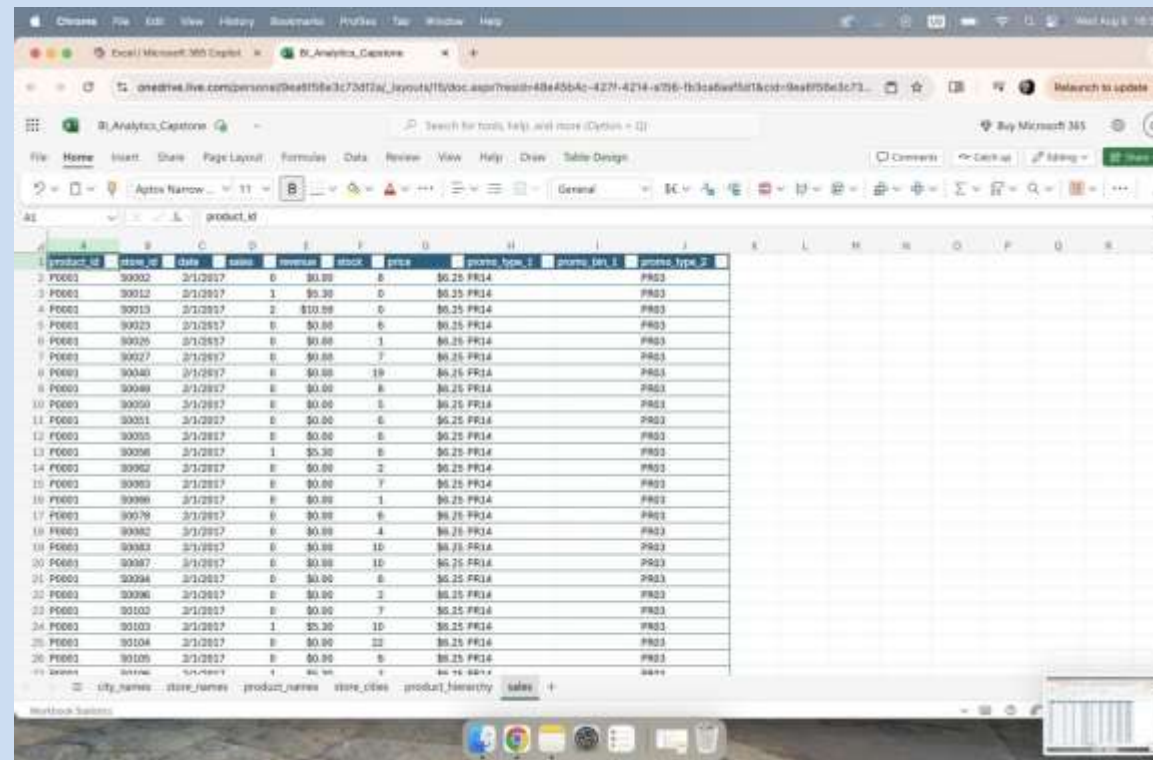
- **Basic Visualizations:**
 - Create fundamental visualizations such as:
 - Line charts
 - Bar charts
 - Scatter plots
 - Packed bubbles charts
 - These visualizations help in understanding data patterns and trends.
- **Dashboard Creation:**
 - Develop a **dashboard** by integrating multiple worksheets into a single view.
 - Arrange the visualizations effectively to provide a comprehensive overview of the data.
- **Interactivity:**
 - Implement **dashboard actions** to add interactivity, allowing users to engage with the data dynamically.
 - This includes setting up filters and actions that enhance user experience and data exploration.
- **Insights Presentation:**
 - Use the dashboard to present insights clearly, supporting decision-making processes for stakeholders.



Results

Module 1, Lesson 1: Data Cleaning and Preparation

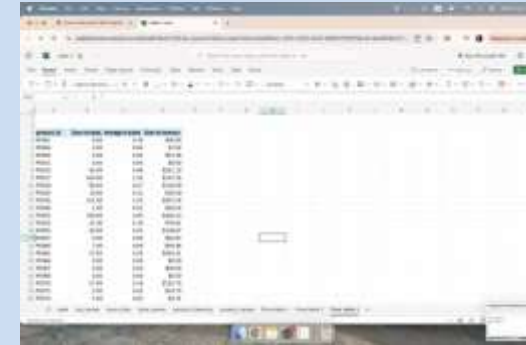
Data Cleaning and Preparation is crucial because it ensures that the dataset is accurate, consistent, and free from errors.



| product_id | store_id | date | sales | revenue | stock | price | promo_type_1 | promo_amt_1 | promo_type_2 |
|------------|----------|----------|-------|---------|-------|--------|--------------|-------------|--------------|
| P0001 | 30001 | 2/1/2017 | 0 | \$0.00 | 8 | \$0.25 | PR04 | | PR03 |
| P0001 | 30012 | 2/1/2017 | 1 | \$0.25 | 0 | \$0.25 | PR04 | | PR03 |
| P0001 | 30013 | 2/1/2017 | 2 | \$0.50 | 0 | \$0.25 | PR04 | | PR03 |
| P0001 | 30023 | 2/1/2017 | 0 | \$0.00 | 6 | \$0.25 | PR04 | | PR03 |
| P0001 | 30026 | 2/1/2017 | 0 | \$0.00 | 1 | \$0.25 | PR04 | | PR03 |
| P0001 | 30027 | 2/1/2017 | 0 | \$0.00 | 7 | \$0.25 | PR04 | | PR03 |
| P0001 | 30040 | 2/1/2017 | 0 | \$0.00 | 19 | \$0.25 | PR04 | | PR03 |
| P0001 | 30049 | 2/1/2017 | 0 | \$0.00 | 8 | \$0.25 | PR04 | | PR03 |
| P0001 | 30050 | 2/1/2017 | 0 | \$0.00 | 5 | \$0.25 | PR04 | | PR03 |
| P0001 | 30051 | 2/1/2017 | 0 | \$0.00 | 6 | \$0.25 | PR04 | | PR03 |
| P0001 | 30055 | 2/1/2017 | 0 | \$0.00 | 6 | \$0.25 | PR04 | | PR03 |
| P0001 | 30056 | 2/1/2017 | 1 | \$0.25 | 0 | \$0.25 | PR04 | | PR03 |
| P0001 | 30062 | 2/1/2017 | 0 | \$0.00 | 2 | \$0.25 | PR04 | | PR03 |
| P0001 | 30063 | 2/1/2017 | 0 | \$0.00 | 7 | \$0.25 | PR04 | | PR03 |
| P0001 | 30066 | 2/1/2017 | 0 | \$0.00 | 1 | \$0.25 | PR04 | | PR03 |
| P0001 | 30079 | 2/1/2017 | 0 | \$0.00 | 6 | \$0.25 | PR04 | | PR03 |
| P0001 | 30082 | 2/1/2017 | 0 | \$0.00 | 4 | \$0.25 | PR04 | | PR03 |
| P0001 | 30083 | 2/1/2017 | 0 | \$0.00 | 10 | \$0.25 | PR04 | | PR03 |
| P0001 | 30087 | 2/1/2017 | 0 | \$0.00 | 10 | \$0.25 | PR04 | | PR03 |
| P0001 | 30094 | 2/1/2017 | 0 | \$0.00 | 6 | \$0.25 | PR04 | | PR03 |
| P0001 | 30096 | 2/1/2017 | 0 | \$0.00 | 2 | \$0.25 | PR04 | | PR03 |
| P0001 | 30102 | 2/1/2017 | 0 | \$0.00 | 7 | \$0.25 | PR04 | | PR03 |
| P0001 | 30103 | 2/1/2017 | 1 | \$0.25 | 10 | \$0.25 | PR04 | | PR03 |
| P0001 | 30104 | 2/1/2017 | 0 | \$0.00 | 22 | \$0.25 | PR04 | | PR03 |
| P0001 | 30105 | 2/1/2017 | 0 | \$0.00 | 6 | \$0.25 | PR04 | | PR03 |

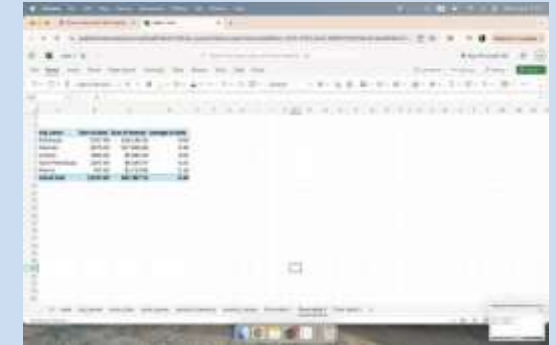
Module 1, Lesson 2: Data Analysis Using Pivot Tables

Basic Analysis Using Pivot Tables allows for quick summarization and aggregation of large datasets, easy identification of trends, patterns, and comparisons across different dimensions and enhanced data exploration, enabling users to derive insights without complex formulas.



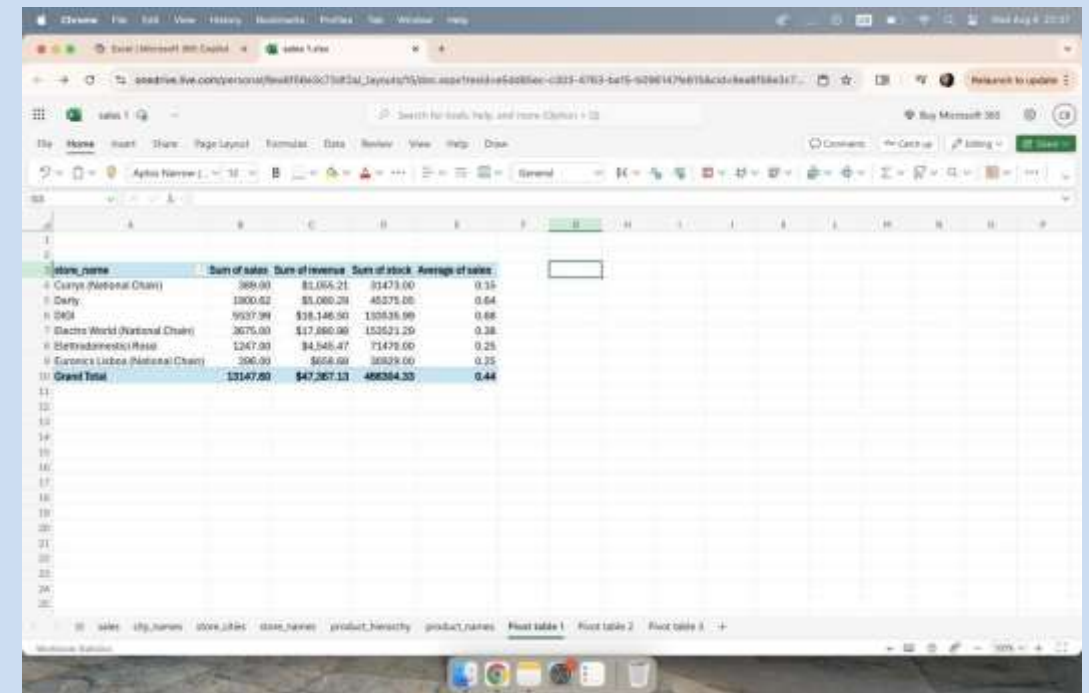
This screenshot shows a PivotTable in Microsoft Excel. The PivotTable is structured with 'Store' in the Rows area and 'Product Category' in the Columns area. The data is summarized by 'Sum of Sales' and 'Sum of Revenue'. The table lists various stores and their corresponding sales and revenue for different product categories.

| Store | Product Category | Sum of Sales | Sum of Revenue |
|---------|------------------|--------------|----------------|
| Store 1 | Electronics | 1000 | 10000 |
| Store 1 | Books | 500 | 5000 |
| Store 1 | Home Goods | 200 | 2000 |
| Store 2 | Electronics | 1500 | 15000 |
| Store 2 | Books | 750 | 7500 |
| Store 2 | Home Goods | 300 | 3000 |
| Store 3 | Electronics | 2000 | 20000 |
| Store 3 | Books | 1000 | 10000 |
| Store 3 | Home Goods | 400 | 4000 |



This screenshot shows another PivotTable in Microsoft Excel, similar to the one above. It displays a summary of sales data by store and product category, with columns for 'Store', 'Product Category', 'Sum of Sales', and 'Sum of Revenue'. The data is organized to show the performance of different stores across various product lines.

| Store | Product Category | Sum of Sales | Sum of Revenue |
|---------|------------------|--------------|----------------|
| Store 1 | Electronics | 1000 | 10000 |
| Store 1 | Books | 500 | 5000 |
| Store 1 | Home Goods | 200 | 2000 |
| Store 2 | Electronics | 1500 | 15000 |
| Store 2 | Books | 750 | 7500 |
| Store 2 | Home Goods | 300 | 3000 |
| Store 3 | Electronics | 2000 | 20000 |
| Store 3 | Books | 1000 | 10000 |
| Store 3 | Home Goods | 400 | 4000 |

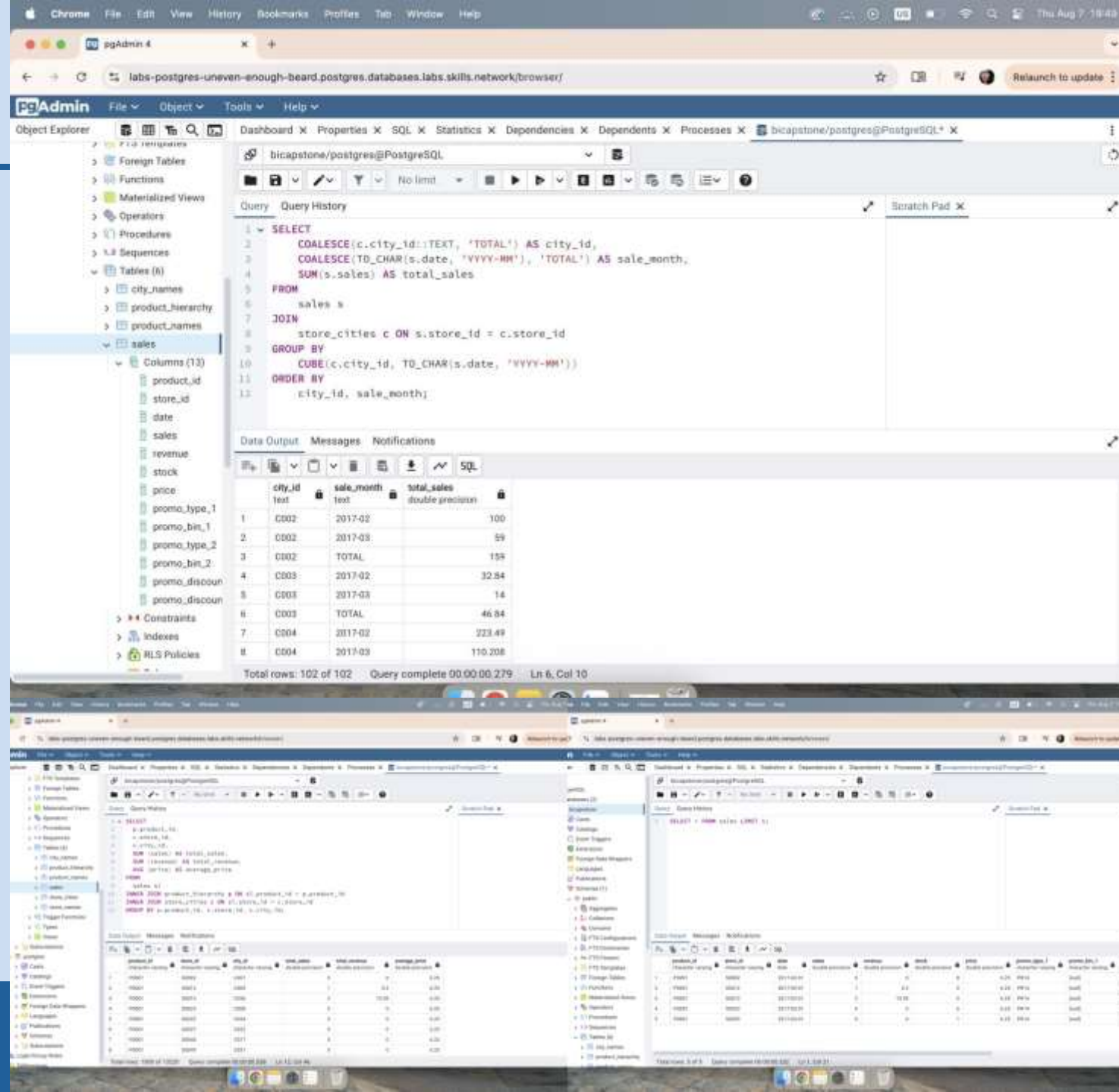


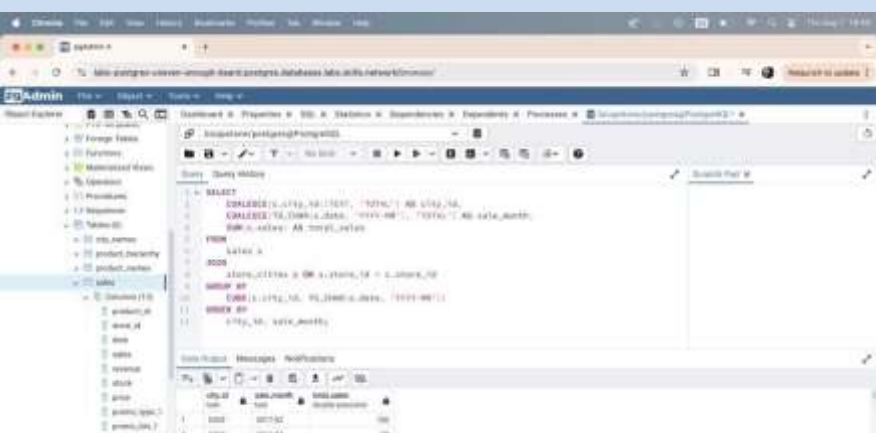
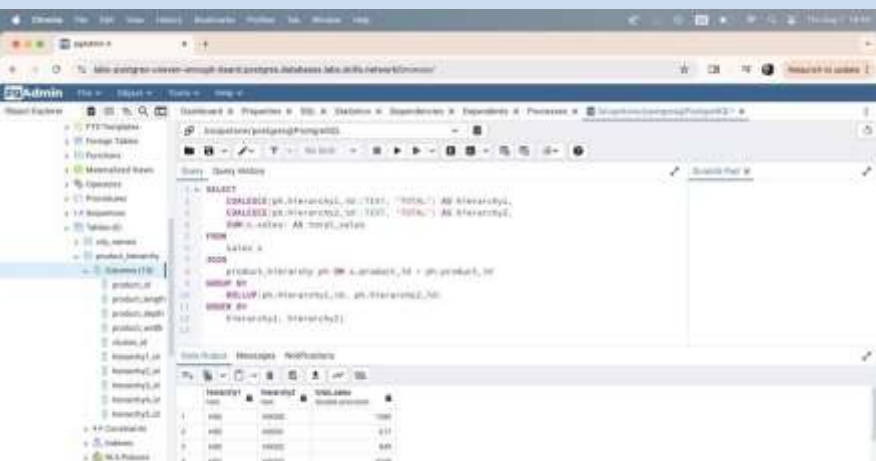
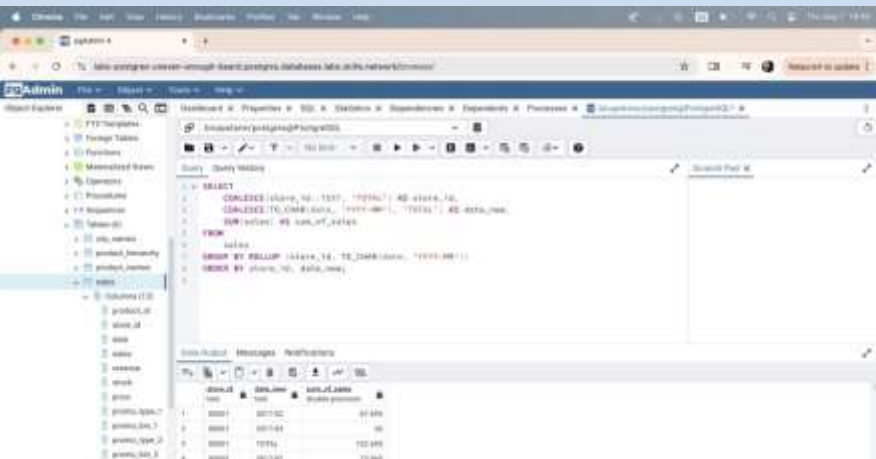
This screenshot shows a PivotTable in Microsoft Excel, displaying a summary of sales data by store and product category. The table includes columns for 'Store Name', 'Sum of Sales', 'Sum of Revenue', 'Sum of Stock', and 'Average of Sales'. The data is organized to show the performance of different stores across various product lines.

| Store Name | Sum of Sales | Sum of Revenue | Sum of Stock | Average of Sales |
|-------------------------------------|-----------------|--------------------|------------------|------------------|
| Curry's (National Chain) | 368.00 | \$1,065.21 | 31473.00 | 0.35 |
| Darty | 1800.62 | \$5,080.28 | 45375.00 | 0.64 |
| Dea | 5937.39 | \$18,146.50 | 139536.99 | 0.68 |
| Electro World (National Chain) | 3675.00 | \$17,880.00 | 153521.20 | 0.38 |
| Electrodomestic House | 1247.00 | \$4,545.47 | 71478.00 | 0.25 |
| Electronics Lisboa (National Chain) | 306.00 | \$658.69 | 35829.00 | 0.25 |
| Grand Total | 13147.60 | \$47,367.13 | 468364.33 | 0.44 |

Module 2, Lesson 1: Data Querying Using PostgreSQL

Setting up the database properly is essential as it organizes data into structured tables, making it easier to query and analyze, facilitates efficient data retrieval and manipulation using SQL and ensures data integrity and relationships between tables.





Module 2, Lesson 2: Data Analysis Using PostgreSQL

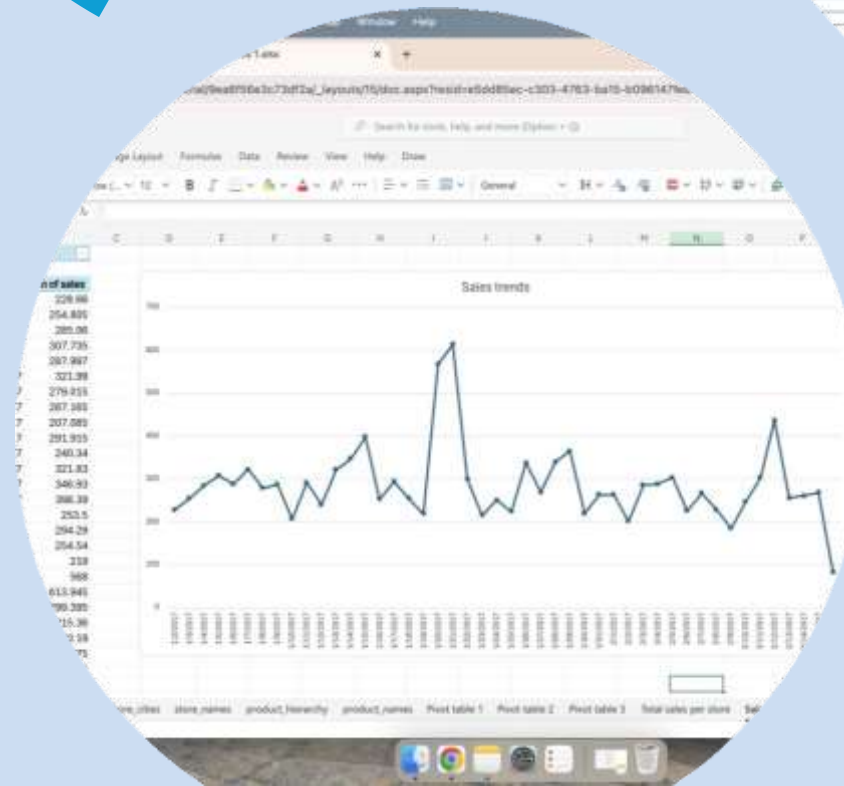
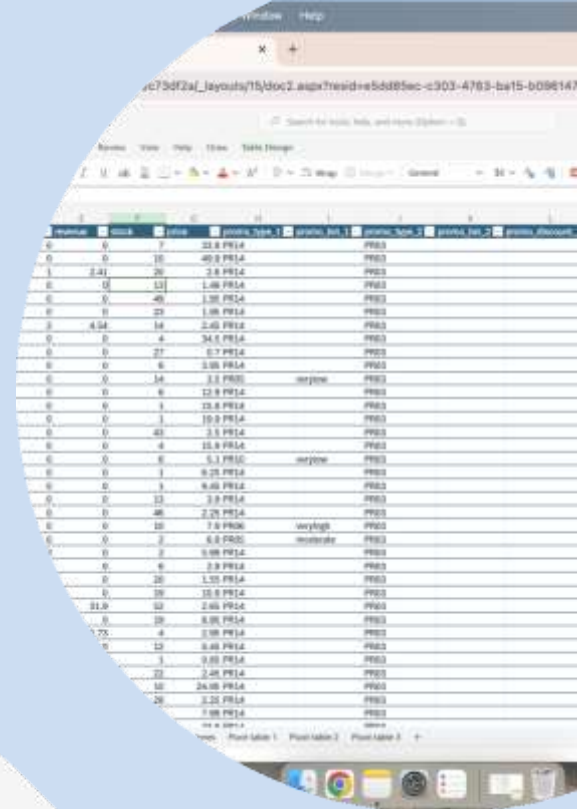
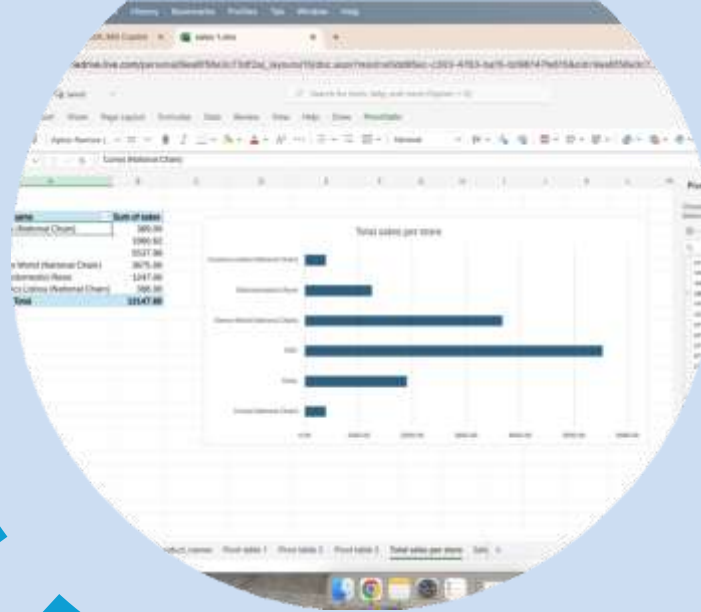
Creation of Data Cubes with ROLLUP: This helps in identifying which products perform best in specific regions or time periods, allowing businesses to tailor their sales strategies and inventory levels accordingly.

Summarizing Data Along Hierarchies: This enables businesses to focus on high-performing categories while identifying underperforming areas that may need attention or strategic changes, thus optimizing marketing efforts and resource allocation.

Identifying Sales Trends Over Time and Across Various Regions: This information is crucial for forecasting demand, planning promotions, and managing stock levels effectively to avoid overstocking or stockouts, ultimately leading to improved customer satisfaction and profitability.

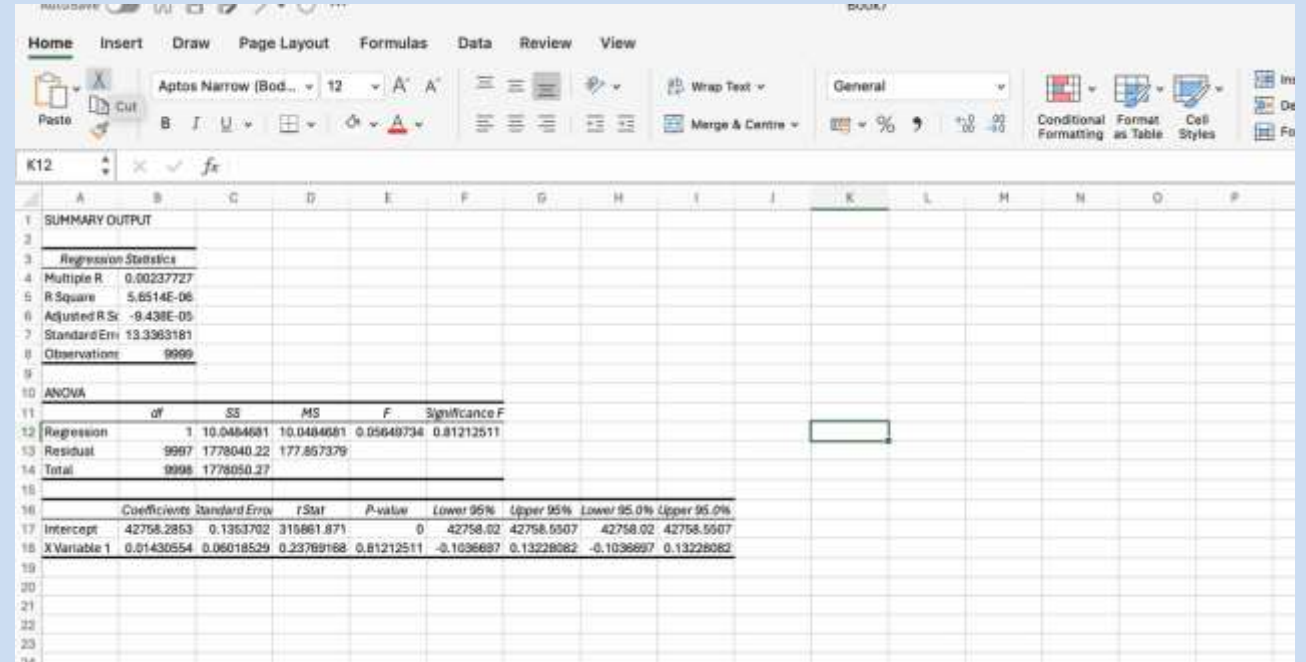
Module 3, Lesson 1: Data Visualization Using Excel

Charts and visualizations play a vital role in enhancing the understanding of data in several ways. First of all, they simplify complex data and effectively illustrate relationships between variables. They allow for quick comparisons across different categories or time period and they tell a compelling story with data, guiding the audience through the analysis and emphasizing key insights.



Module 3, Lesson 2: Statistical Analysis

Since this p-value is less than the common significance level of 0.05, you can conclude that there is a statistically significant relationship between the variables.



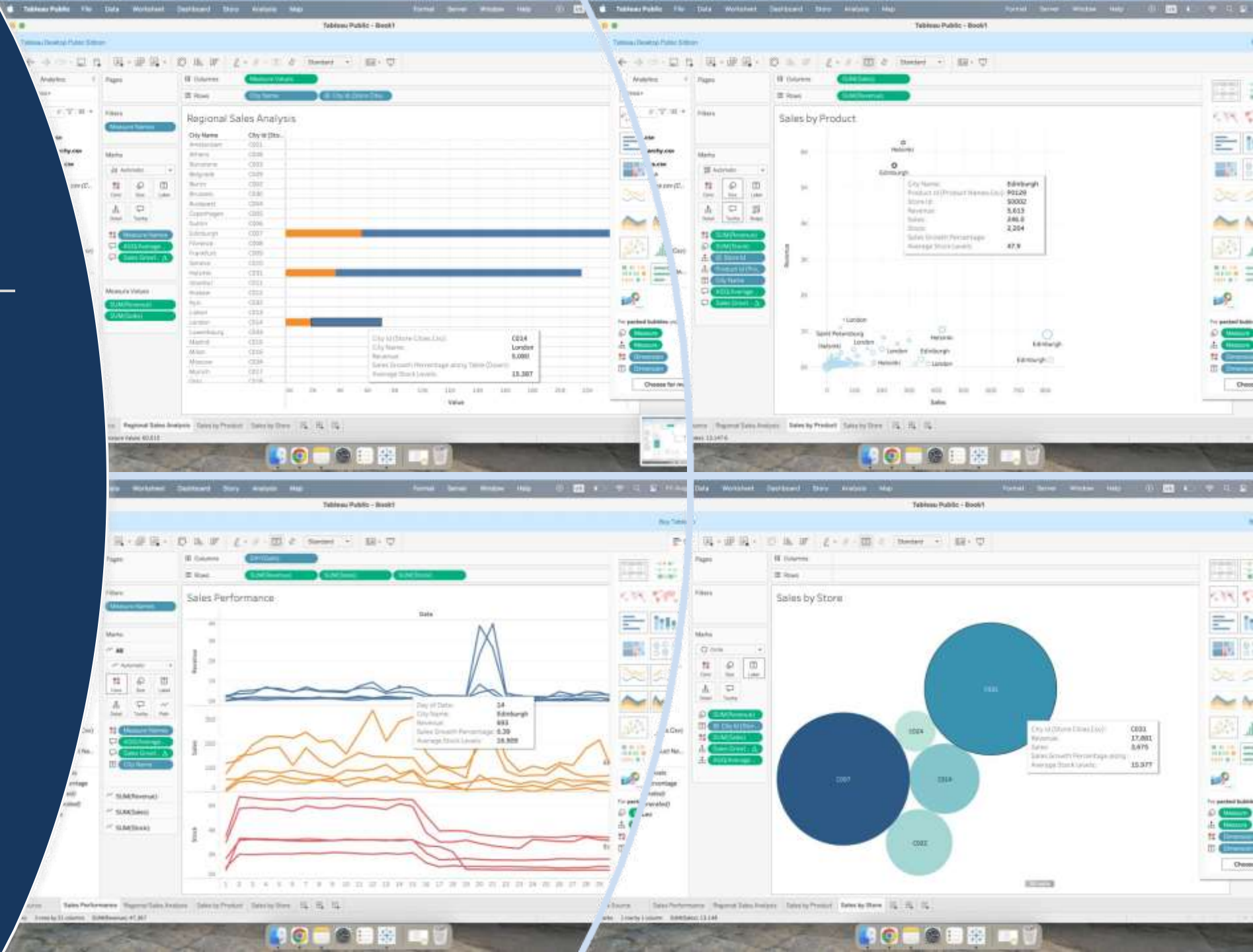
The screenshot shows an Excel spreadsheet with a regression analysis summary. The 'SUMMARY OUTPUT' table is located in the range A1:K14. The 'ANOVA' table is in A11:K14. The 'Coefficients' table is in A16:K18. The p-value for X Variable 1 is 0.81212511, which is greater than the common significance level of 0.05, indicating that there is no statistically significant relationship between the variables.

| SUMMARY OUTPUT | | | | | | | | |
|-----------------------|----------------|------------|------------|------------|----------------|--|--|--|
| Regression Statistics | | | | | | | | |
| | Multiple R | 0.00237727 | | | | | | |
| | R Square | 5.6514E-06 | | | | | | |
| | Adjusted R Sq | -9.438E-05 | | | | | | |
| | Standard Error | 13.3363181 | | | | | | |
| | Observations | 9999 | | | | | | |
| ANOVA | | | | | | | | |
| | df | SS | MS | F | Significance F | | | |
| Regression | 1 | 10.0484681 | 10.0484681 | 0.05649734 | 0.81212511 | | | |
| Residual | 9997 | 1778040.22 | 177.857379 | | | | | |
| Total | 9998 | 1778050.27 | | | | | | |

| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
|--------------|--------------|----------------|------------|------------|------------|------------|-------------|-------------|
| Intercept | 42758.2853 | 0.1353702 | 315861.871 | 0 | 42758.02 | 42758.5507 | 42758.02 | 42758.5507 |
| X Variable 1 | 0.01430554 | 0.06018529 | 0.23769168 | 0.81212511 | -0.1036687 | 0.13228082 | -0.1036687 | 0.13228082 |

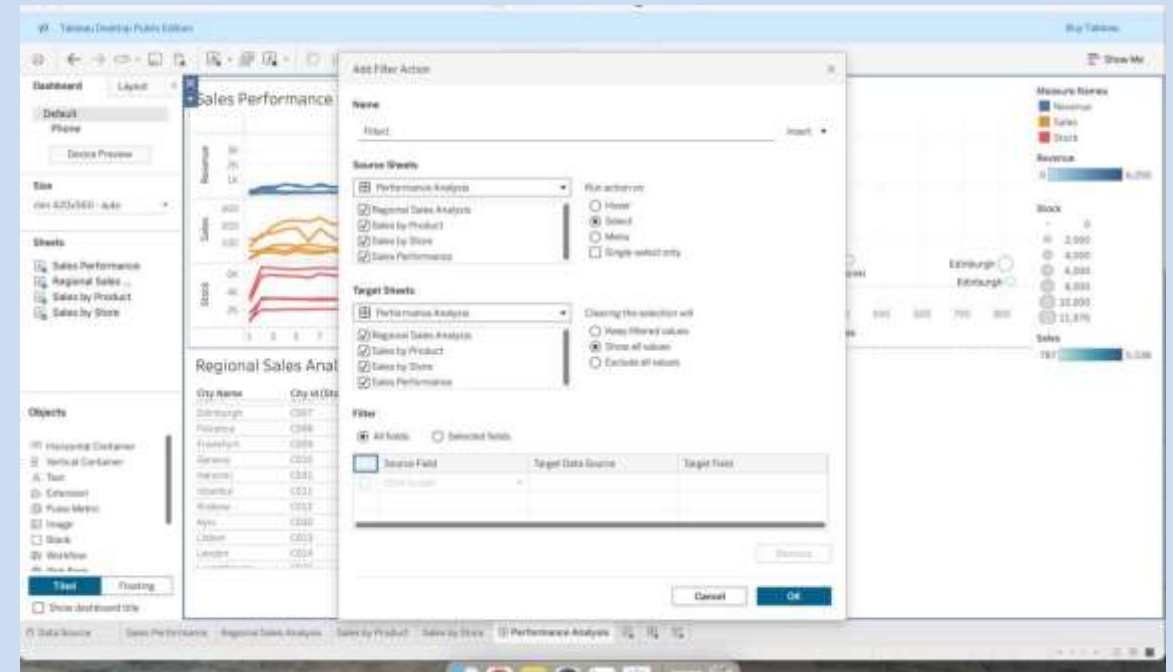
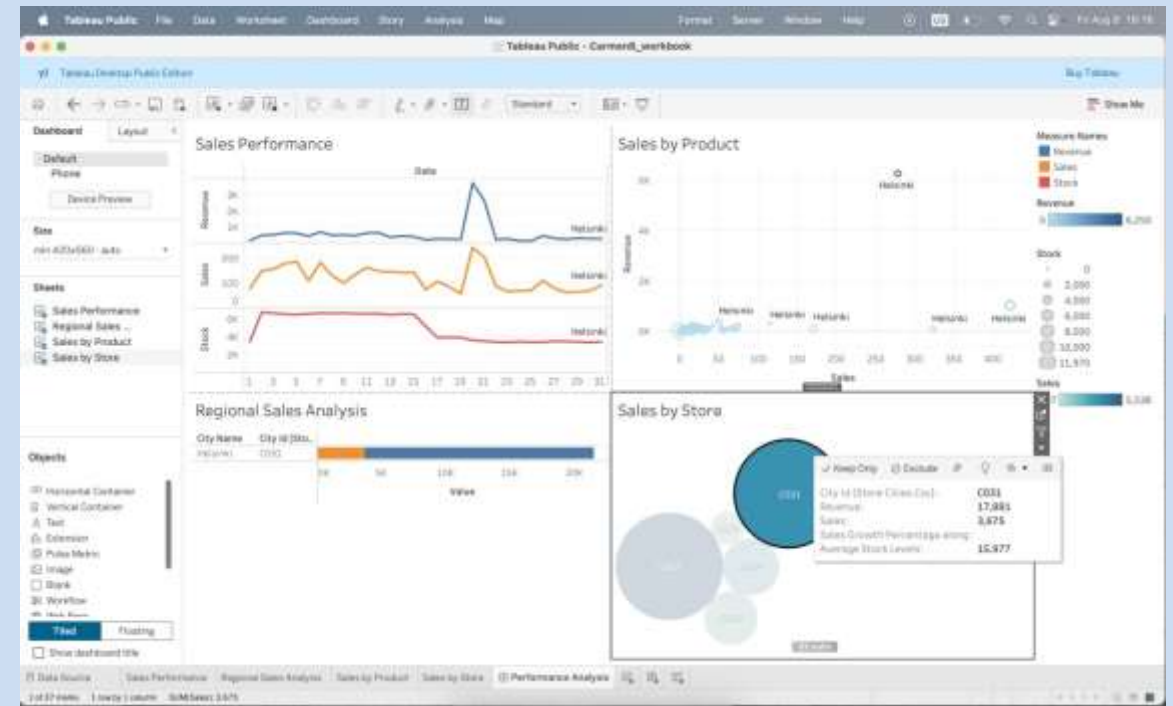
Module 4, Lesson 1: Basic Tableau Visualizations

Charts and visualizations play a vital role in enhancing the understanding of data in several ways. First of all, they simplify complex data and effectively illustrate relationships between variables. They allow for quick comparisons across different categories or time period and they tell a compelling story with data, guiding the audience through the analysis and emphasizing key insights.



Module 4, Lesson 2: Advanced Visualizations Using Tableau

A dashboard that includes visualizations supports decision-making processes by providing real-time insights and enhancing clarity. Dashboards enable users to compare different metrics side by side, aiding in identifying patterns, anomalies, and areas that require attention. By presenting data visually, dashboards encourage decisions based on evidence rather than intuition, leading to more informed and effective outcomes.





Discussion

Insights and Recommendations

- I would recommend focusing on products that sell better and consider eliminating the products that do not account for a large portion of the revenue. I also recommend carefully examining the stocks for each store taking in consideration the products that sell best.



Conclusion

Summary

Comprehensive data analysis for a U.S.-based retail company was executed. By cleaning the data and developing interactive dashboards in Microsoft Excel and Tableau, key insights to improve business strategy were identified. My findings suggest the business should prioritize its best-selling products, potentially phasing out items with low revenue contribution, and optimize inventory management at each store by stocking more of the most popular items.