# Tableau Capstone

31 May 2024

# Table of Contents


- Introduction
- Methodology
- Results
- Extra Materials
- Presentation Improvement
- Innovation and Value Addition
- Appendix


# Introduction

# Introduction

A leading **retail chain** that operates globally, renowned for its diverse product offerings and commitment to customer satisfaction. In order to maintain its competitive edge and optimize business performance, the company is keen on leveraging data-driven insights to enhance sales strategies and operational efficiency.

Conduct a **comprehensive analysis** of the sales performance of the retail chain across different regions and time periods.

# Methodology

# Data Details

1. Data files used: *sales.csv*, *product_hierarchy.csv*, *store_cities.csv*, *store_names.csv*, *city_names.csv* , and *product_names.csv*
2. Data cleanup all files, checking format. Types, filling in missing data
3. Perform basic charts using Tableau
4. Advanced Visualization Using Tableau – Dashboard
5. Creating Engaging Stories in Tableau

# Module 1: Preparing Data and Creating Basic Visualizations

| | product_id | store_id | date | sales | revenue | stock | price | promo_type_1 | promo_bin_1 | promo_type_2 | promo_bin_2 | promo_ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | P0005 | S0001 | 2017-02-01 | 0.00 | 0.00 | 7.00 | 33.90 | PR14 | NaN | PR03 | NaN | |
| **1** | P0011 | S0001 | 2017-02-01 | 0.00 | 0.00 | 10.00 | 49.90 | PR14 | NaN | PR03 | NaN | |
| **2** | P0015 | S0001 | 2017-02-01 | 1.00 | 2.41 | 20.00 | 2.60 | PR14 | NaN | PR03 | NaN | |
| **3** | P0017 | S0001 | 2017-02-01 | 0.00 | 0.00 | 13.00 | 1.49 | PR14 | NaN | PR03 | NaN | |
| **4** | P0018 | S0001 | 2017-02-01 | 0.00 | 0.00 | 49.00 | 1.95 | PR14 | NaN | PR03 | NaN | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **29994** | P0514 | S0008 | 2017-01-18 | 0.00 | 0.00 | 2.00 | 8.90 | PR14 | NaN | PR03 | NaN | |
| **29995** | P0527 | S0008 | 2017-01-18 | 0.00 | 0.00 | 5.00 | 1.95 | PR14 | NaN | PR03 | NaN | |
| **29996** | P0536 | S0008 | 2017-01-18 | 0.00 | 0.00 | 20.00 | 2.95 | PR14 | NaN | PR03 | NaN | |
| **29997** | P0543 | S0008 | 2017-01-18 | 0.00 | 0.00 | 28.00 | 2.50 | PR14 | NaN | PR03 | NaN | |
| **29998** | P0551 | S0008 | 2017-01-18 | 0.00 | 0.00 | 24.00 | 1.85 | PR14 | NaN | PR03 | NaN | |

29999 rows × 13 columns

# Module 1: Preparing Data and Creating Basic Visualizations

```
[13]: df.isnull().sum()
```

```
[13]: product_id                0
      store_id                  0
      date                      0
      sales                     0
      revenue                   0
      stock                     0
      price                     0
      promo_type_1              0
      promo_bin_1           26644
      promo_type_2              0
      promo_bin_2           29999
      promo_discount_2      29999
      promo_discount_type_2 29999
      dtype: int64
```

```
[15]: df.isnull().sum()
```

```
[15]: product_id                0
      store_id                  0
      date                      0
      sales                     0
      revenue                   0
      stock                     0
      price                     0
      promo_type_1              0
      promo_bin_1               0
      promo_type_2              0
      promo_bin_2               0
      promo_discount_2          0
      promo_discount_type_2     0
      dtype: int64
```

# Module 1: Preparing Data and Creating Basic Visualizations

| | product_id | store_id | date | sales | revenue | stock | price | promo_type_1 | promo_bin_1 | promo_type_2 | promo_bin_2 | promo_discount |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | P0005 | S0001 | 2017-02-01 | 0.00 | 0.00 | 7.00 | 33.90 | PR14 | 0 | PR03 | 0.00 | 0. |
| 1 | P0011 | S0001 | 2017-02-01 | 0.00 | 0.00 | 10.00 | 49.90 | PR14 | 0 | PR03 | 0.00 | 0. |
| 2 | P0015 | S0001 | 2017-02-01 | 1.00 | 2.41 | 20.00 | 2.60 | PR14 | 0 | PR03 | 0.00 | 0. |
| 3 | P0017 | S0001 | 2017-02-01 | 0.00 | 0.00 | 13.00 | 1.49 | PR14 | 0 | PR03 | 0.00 | 0. |
| 4 | P0018 | S0001 | 2017-02-01 | 0.00 | 0.00 | 49.00 | 1.95 | PR14 | 0 | PR03 | 0.00 | 0. |

| | product_id | store_id | date | sales | revenue | stock | price | promo_type_1 | promo_bin_1 | promo_type_2 | promo_bin_2 | promo_discount |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | P0005 | S0001 | 2017-02-01 | 0.00 | 0.00 | 7.00 | 33.90 | pr14 | 0 | pr03 | 0.00 | 0. |
| 1 | P0011 | S0001 | 2017-02-01 | 0.00 | 0.00 | 10.00 | 49.90 | pr14 | 0 | pr03 | 0.00 | 0. |
| 2 | P0015 | S0001 | 2017-02-01 | 1.00 | 2.41 | 20.00 | 2.60 | pr14 | 0 | pr03 | 0.00 | 0. |
| 3 | P0017 | S0001 | 2017-02-01 | 0.00 | 0.00 | 13.00 | 1.49 | pr14 | 0 | pr03 | 0.00 | 0. |
| 4 | P0018 | S0001 | 2017-02-01 | 0.00 | 0.00 | 49.00 | 1.95 | pr14 | 0 | pr03 | 0.00 | 0. |

# Module 1: Preparing Data and Creating Basic Visualizations

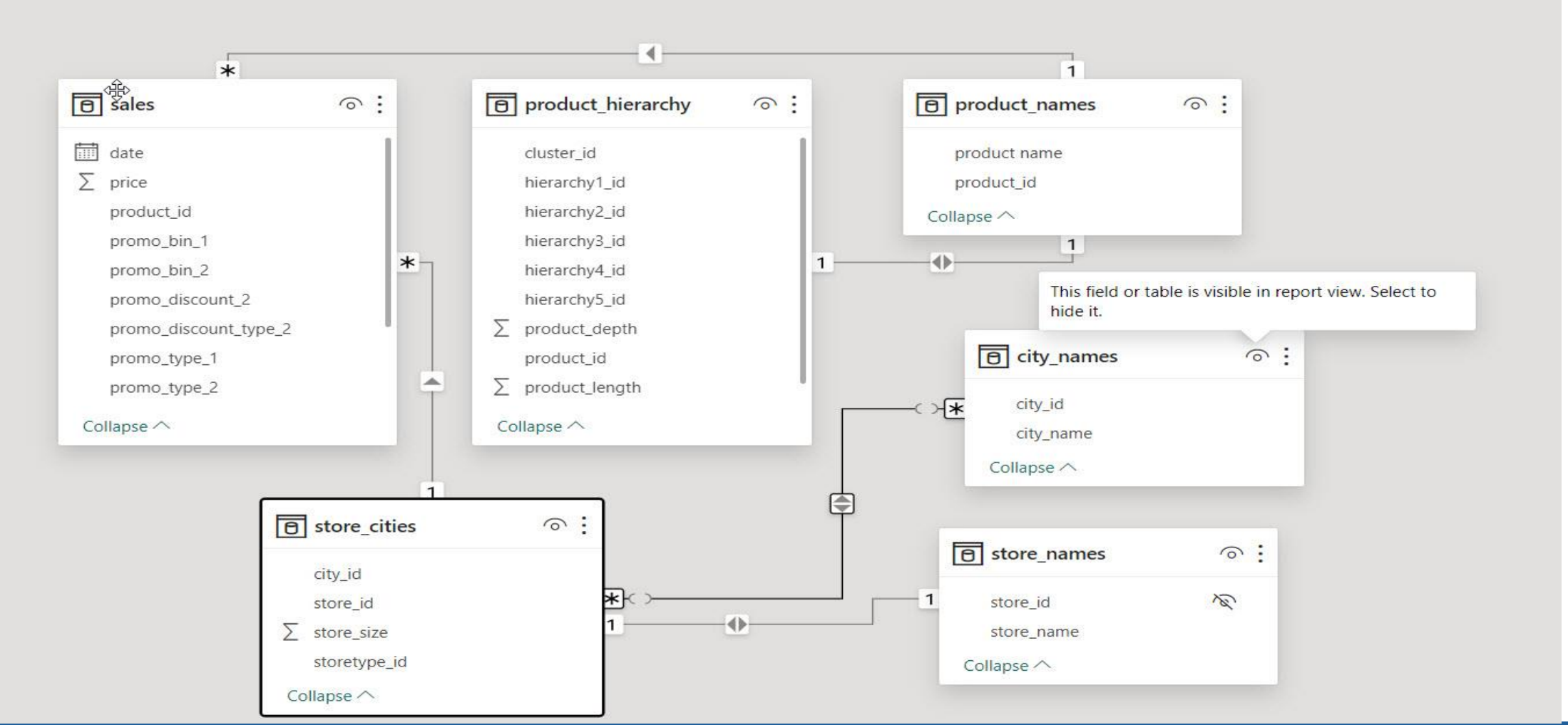
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 699 entries, 0 to 698
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   product_id      699 non-null    object
 1   product_length  699 non-null    float64
 2   product_depth   699 non-null    float64
 3   product_width   699 non-null    float64
 4   cluster_id      699 non-null    object
 5   hierarchy1_id   699 non-null    object
 6   hierarchy2_id   699 non-null    object
 7   hierarchy3_id   699 non-null    object
 8   hierarchy4_id   699 non-null    object
 9   hierarchy5_id   699 non-null    object
dtypes: float64(3), object(7)
memory usage: 54.7+ KB
```

```
df2.dtypes
```

```
product_id          object
product_length     float64
product_depth      float64
product_width      float64
cluster_id          object
hierarchy1_id       object
hierarchy2_id       object
hierarchy3_id       object
hierarchy4_id       object
```

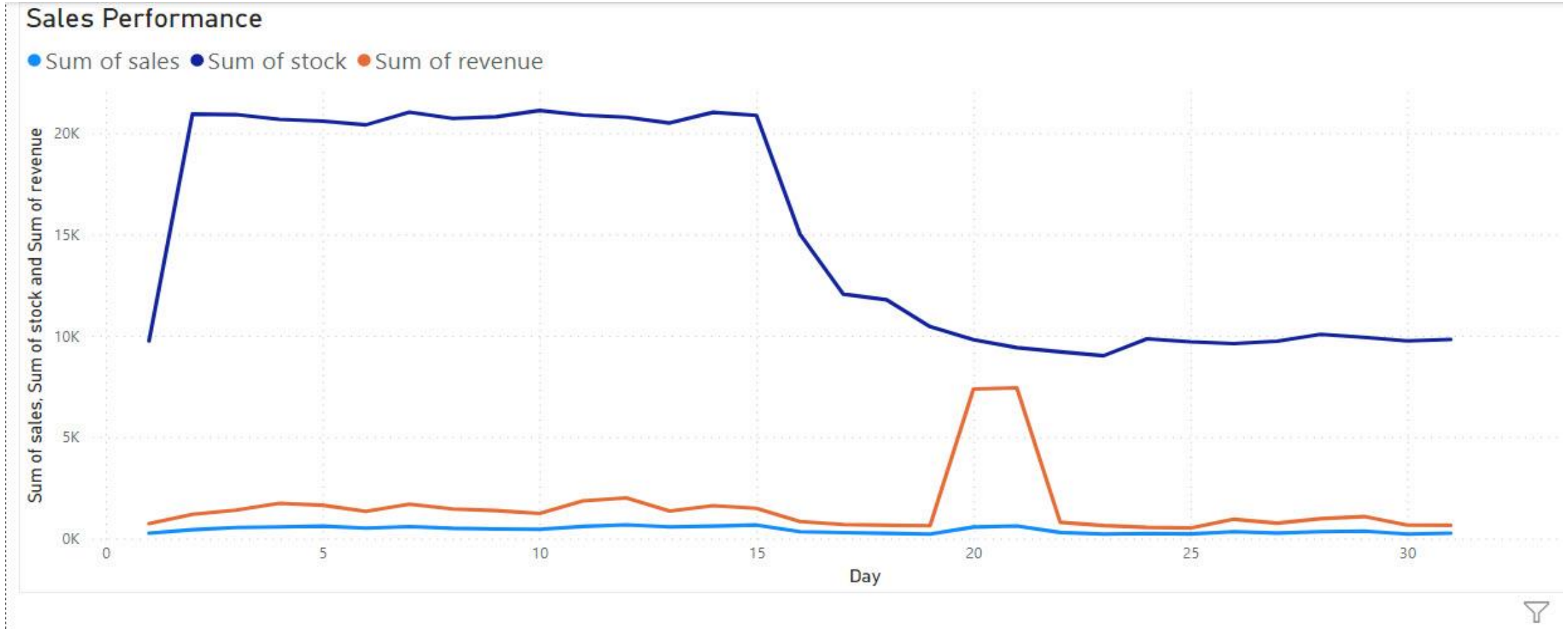
| cluster_id | hierarchy1_id | hierarchy2_id | hierarchy3_id | hierarchy4_id | hierarchy5_id |
|---|---|---|---|---|---|
| 0 | H00 | H0004 | H000401 | H00040105 | H0004010534 |
| cluster_5 | H01 | H0105 | H010501 | H01050100 | H0105010006 |
| cluster_0 | H03 | H0315 | H031508 | H03150800 | H0315080028 |
| cluster_3 | H03 | H0314 | H031405 | H03140500 | H0314050003 |
| cluster_9 | H03 | H0312 | H031211 | H03121109 | H0312110917 |

# Module 1: Preparing Data and Creating Basic Visualizations

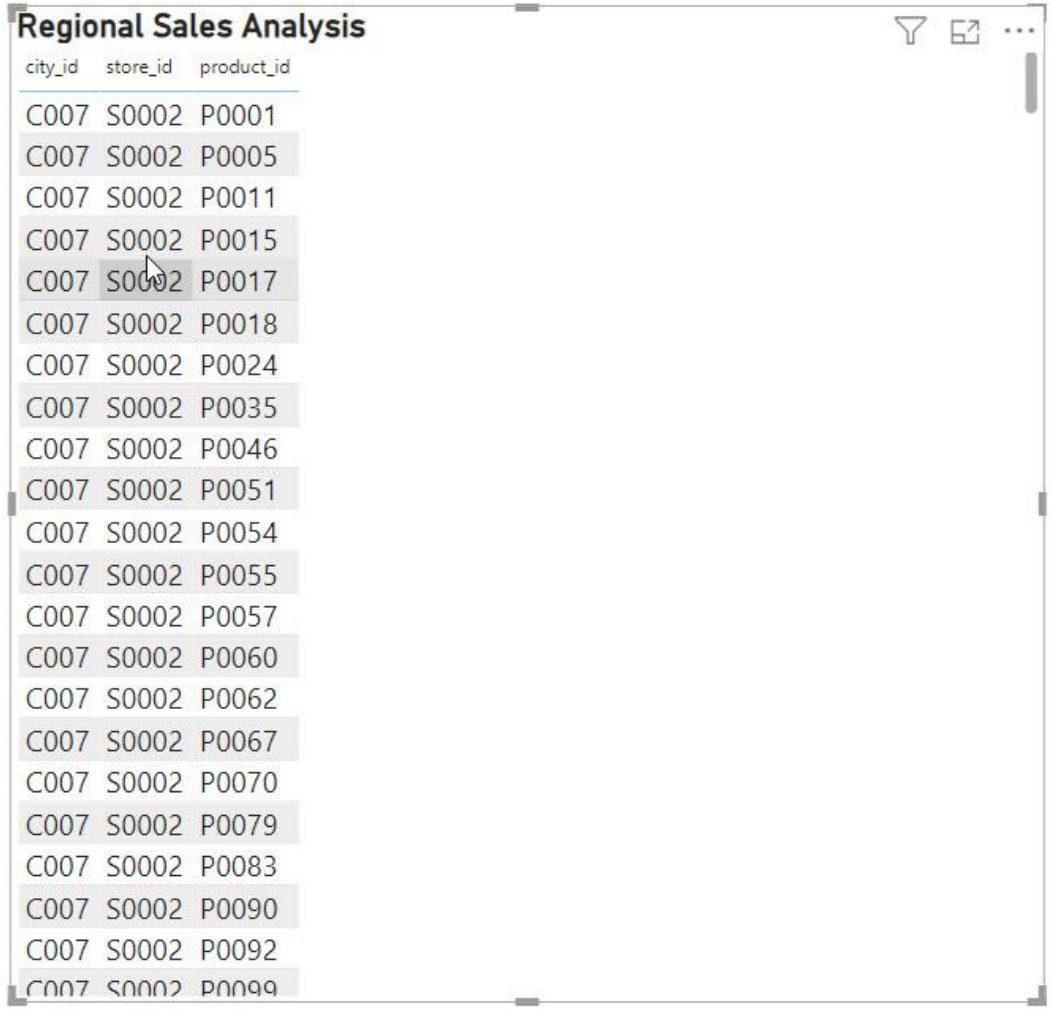


# Results

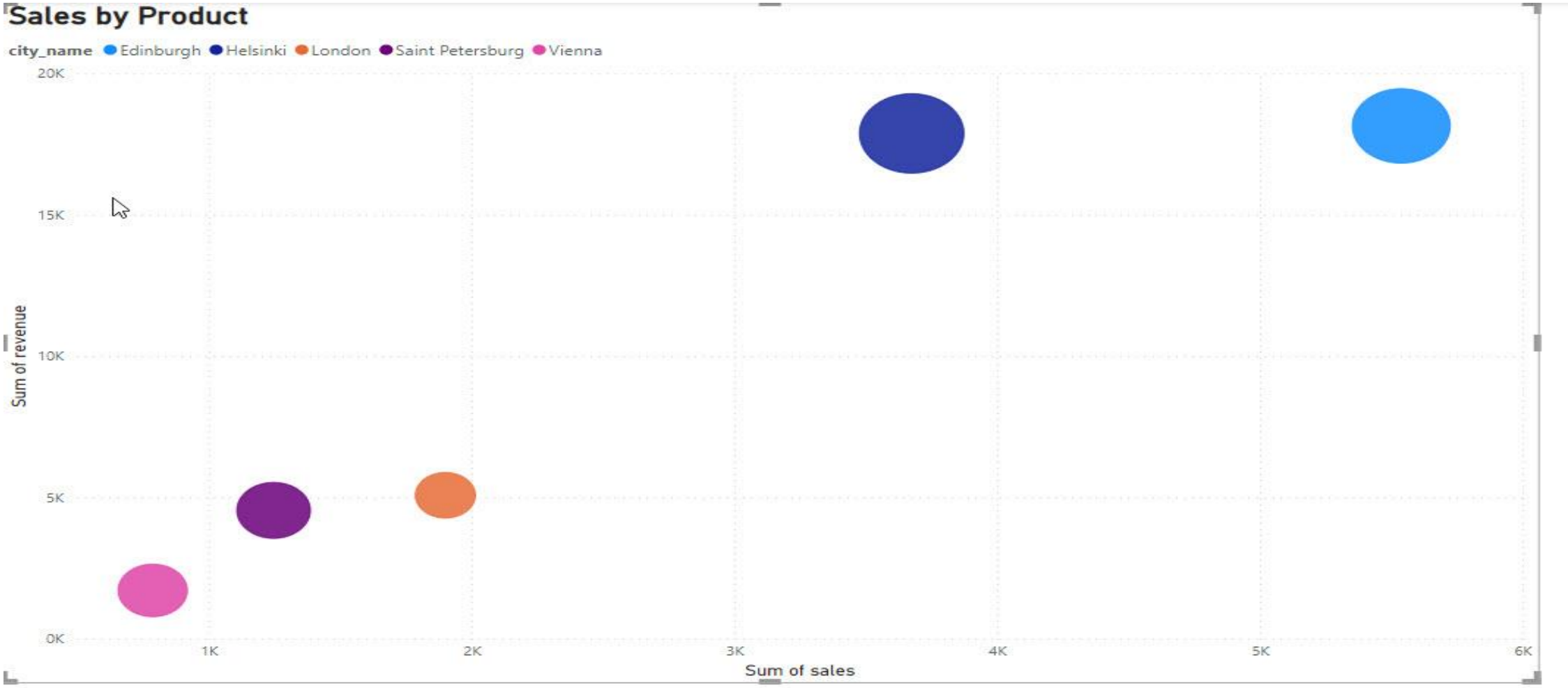
# Module 2: Advanced Visualizations, Dashboards, and Stories with Tableau



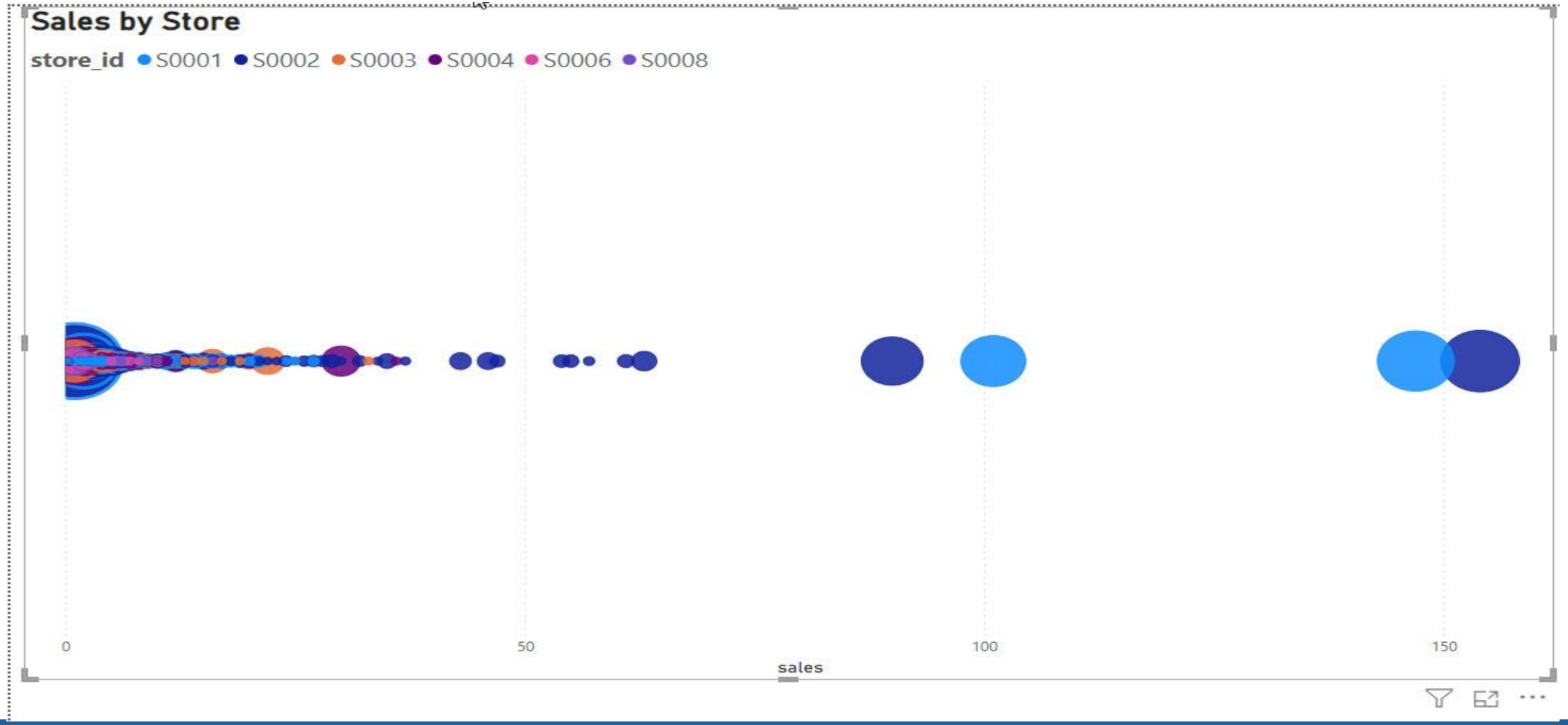
# Module 2: Advanced Visualizations, Dashboards, and Stories with Tableau



# Module 2: Advanced Visualizations, Dashboards, and Stories with Tableau



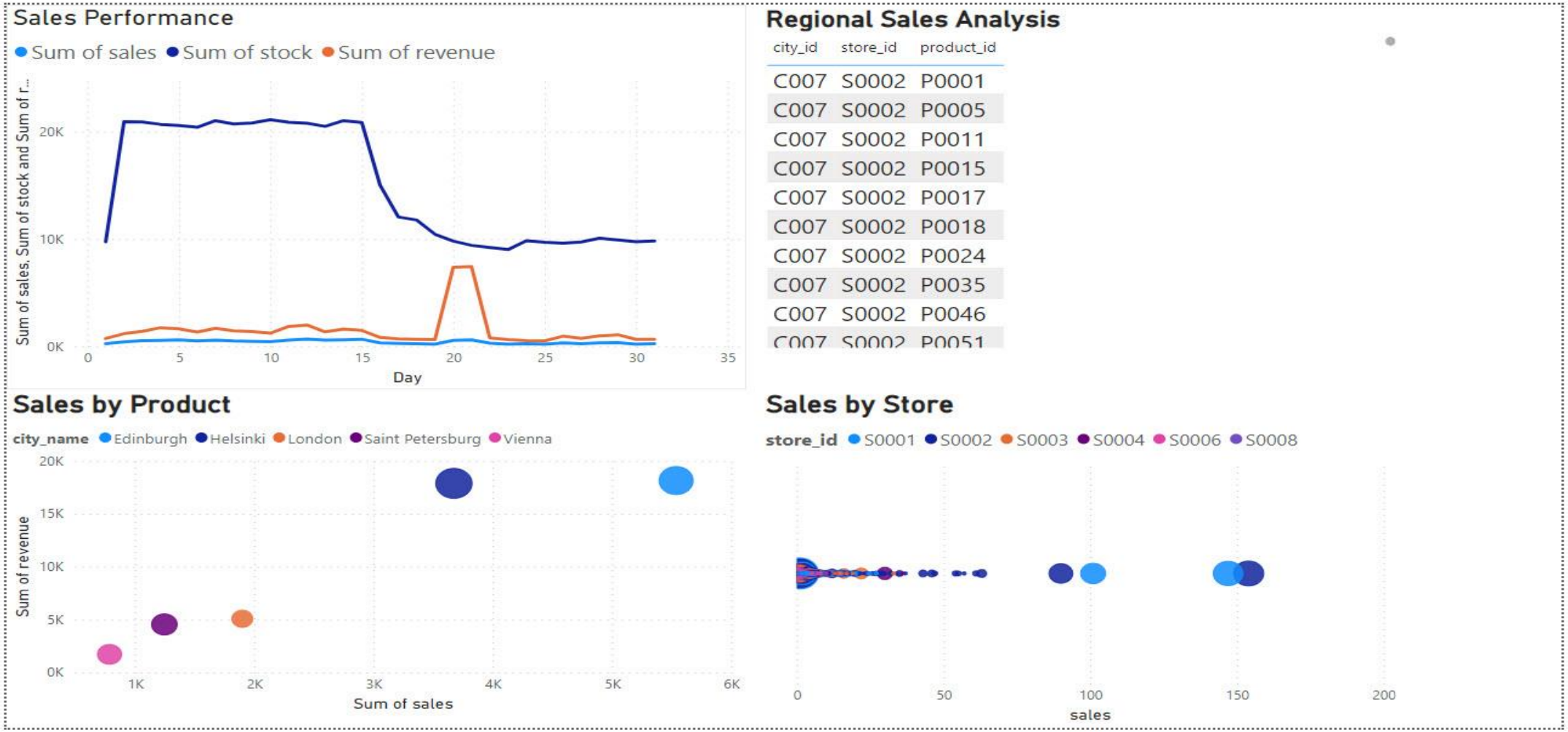
# Module 2: Advanced Visualizations, Dashboards, and Stories with Tableau



# Module 2: Advanced Visualizations, Dashboards, and Stories with Tableau

- A **line chart** is effective for showing trends over time or across ordered categories
- A **bar chart** is ideal for comparing discrete categories or showing distribution.
- A **scatter plot** displays individual data points as dots on a two-dimensional plane.
- A **packed bubbles chart** represents hierarchical data using circles of varying sizes.

# Module 2: Advanced Visualizations, Dashboards, and Stories with Tableau



# Module 3: Tableau Capstone Project Submission and Evaluation

# Extra materials

# Any extra chart or worksheets

- Create a chart or worksheet that segments customers based on their purchasing behavior, such as frequency of purchases, average transaction value, or product preferences. This can help identify different customer segments and tailor marketing strategies accordingly.
- Develop visualizations that depict seasonal sales patterns, including peaks and troughs in sales across different regions or product categories. Understanding seasonal trends can aid in optimizing inventory management and promotional campaigns.
- Incorporate data on competitor sales or market share to compare against your retail chain's performance. Visualizing this data can highlight areas where your company is excelling or where there may be opportunities for improvement.
- Evaluate the effectiveness of past promotional campaigns by visualizing sales data before, during, and after promotions. This can help assess the impact of various promotional strategies on sales performance and inform future promotional planning.

# Presentation Improvement

- Prepare a concise executive summary highlighting the key findings, actionable insights, and recommendations derived from your analysis. This summary should provide a high-level overview of the sales performance analysis and its implications for the business.
- Design infographics to visually represent complex data and concepts in a clear and engaging manner. Infographics can help simplify information, making it easier for stakeholders to grasp important insights at a glance.
- Include case studies or success stories that highlight real-world examples of how data-driven insights have been leveraged to drive business growth or improve operational efficiency within the retail chain. These stories can add credibility to your analysis and inspire confidence in the proposed recommendations.

# Innovation and Value Addition

- Facilitate a crowdsourced data analysis session during the presentation, where audience members can contribute their observations, hypotheses, or interpretations of the sales data in real time. This collaborative approach fosters active participation and encourages diverse perspectives.
- Incorporate interactive physical props or prototypes, such as miniature retail store models or product displays, to complement the data visualizations and provide a tangible representation of key insights. This multisensory approach engages multiple senses and enhances comprehension.
- Integrate live data feeds into the presentation to demonstrate real-time sales performance and trends. This dynamic approach keeps the audience engaged and allows them to witness the immediate impact of data-driven insights on business operations.

# Conclusion

# Summary