

# Sales Performance Analysis



Prepared by: Sarfaraz Ahmed

# Table of Contents

---

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



**Stock ID: 47857043**



# Introduction

# Introduction

---

- **Business Scenario: Retail Chain Sales Analysis**
- **Business Context:**
  - A retail chain is seeking to optimize its sales performance and identify growth opportunities. The chain operates multiple stores across various locations and offers a diverse range of products.
- **Purpose of the Analysis:**
  - The primary objective of this analysis is to gain a deeper understanding of the factors influencing sales performance and identify areas for improvement. By analyzing sales data, the retail chain aims to:
  - **Identify trends and patterns:** Recognize seasonal variations, regional differences, and product-specific trends.
  - **Optimize inventory management:** Determine optimal stock levels for different products and locations.
  - **Evaluate marketing strategies:** Assess the effectiveness of promotional activities and pricing strategies.
  - **Identify growth opportunities:** Discover new markets or product categories with potential for expansion.
- By conducting a comprehensive analysis of sales data, the retail chain can make data-driven decisions to enhance profitability and customer satisfaction.





# Methodology

# Data Details

---

The initial dataset consisted of six CSV files:

- **sales:** Contains detailed sales transaction data, including transaction ID, date, product ID, price, quantity, store ID, city ID and revenue.
- **product\_hierarchy:** Defines the hierarchical structure of products, including product ID, product name, category, and subcategory.
- **store\_cities:** Maps store IDs to city names.
- **store\_names:** Provides additional information about stores, such as store name and address.
- **city\_names:** Contains information about cities, including city name and region.
- **product\_names:** Provides alternative or descriptive names for products.

# Module 1: Data Cleaning and Preparation Using Excel

---

- **Data Collection:** Gathered the necessary datasets from the provided CSV files: sales.csv, product\_hierarchy.csv, store\_cities.csv, store\_names.csv, city\_names.csv, and product\_names.csv.
- **Data Cleaning:** Inspected the data for inconsistencies, missing values, and outliers. Corrected any errors or inconsistencies found.
- **Data Integration:** Joined the relevant tables (e.g., sales.csv with product\_hierarchy.csv and store\_cities.csv) to create a comprehensive dataset.
- **Data Transformation:** Created calculated fields to derive additional information, such as total sales, revenue, and stock levels.

# Module 2: Data Querying and Analysis Using PostgreSQL

---

## 1. Create Database and Tables:

- Create a new PostgreSQL database and tables corresponding to the CSV files (sales, product\_hierarchy, store\_cities, store\_names, city\_names, and product\_names).
- Define appropriate data types for each column based on the data content (e.g., date, integer, text).

## 2. Import Data:

- Use the COPY command or tools like psql or pgloader to import data from the CSV files into the corresponding tables.
- Ensure data consistency and handle any potential data quality issues during the import process.

## 3. Query Data:

- Use SQL queries to explore the data, examine column values, and identify any inconsistencies or errors.
- For example, you can use `SELECT * FROM sales LIMIT 10` to view the first 10 rows of the sales table.

## 4. Data Cleaning:

- Address data quality issues such as missing values, duplicates, and incorrect data types.
- Use SQL functions like COALESCE, NULLIF, and TRIM to handle missing values and inconsistencies.



# Module 3: Data Visualization and Statistical Analysis

---

## 1. Aggregate Data:

- Use aggregation functions like SUM, AVG, COUNT, MIN, and MAX to calculate summary statistics for sales, revenue, and other metrics.
- For example, `SELECT SUM(sales) FROM sales;` calculates the total sales.

## 2. Join Tables:

- Combine data from multiple tables using JOIN operations (INNER JOIN, LEFT JOIN, RIGHT JOIN, FULL OUTER JOIN) to get a comprehensive view of the data.
- For example, `SELECT * FROM sales INNER JOIN product_hierarchy ON sales.product_id = product_hierarchy.product_id;` joins the sales and product\_hierarchy tables.

## 3. Filter Data:

- Use the WHERE clause to filter data based on specific conditions.
- For example, `SELECT * FROM sales WHERE product_category = 'Electronics';` filters sales for the "Electronics" category.

## 4. Group Data:

- Use the GROUP BY clause to group data by specific columns and apply aggregation functions.
- For example, `SELECT product_category, SUM(sales) FROM sales GROUP BY product_category;` calculates total sales for each product category.

# Module 4: Data Visualization and Dashboards Using Tableau

---

## 1. Connect to Data:

- Connect to the CSV files containing your sales data.

## 2. Data Cleaning and Preparation:

- Inspect the data for any inconsistencies, missing values, or errors.
- Clean the data as needed, using Tableau's data cleaning features or calculated fields.
- Create calculated fields to derive additional information (e.g., total sales, revenue, growth rates).

## 3. Data Visualization:

Create visualizations to explore and analyze the data. Consider using:

- Line charts: To visualize trends over time.
- Bar charts: To compare values across categories.
- Bubble Packed chart: To find out biggest selling product.

## 4. Filtering and Drilling Down:

- Use filters to focus on specific parts of the data (e.g., filter by product category, city, or date range).
- Create hierarchies for dimensions like product or location to drill down into details.

## 5. Creating Dashboards:

- Combine multiple visualizations into a single dashboard for a comprehensive view of the data.
- Use filters, parameters, and actions to make the dashboard interactive.



# Results

# Module 1, Lesson 1: Data Cleaning and Preparation

- Data Quality: The dataset contained some missing values and inconsistencies, which were addressed through data cleaning techniques.
- Data Transformation: Calculated fields were created to derive additional information, such as total sales, revenue, and stock levels.

	A	B	C	D	E	F	G	H	I	J	K	L
1	product_id	store_id	date	sales	revenue	stock	price	promo_type_1	promo_bin_1	promo_type_2	promo_bin_2	promo_discount
2	P0001	S0122	01-04-17	1	5.3	0	6.25	PR14		PR03		
3	P0001	S0049	01-04-17	1	5.3	7	6.25	PR14		PR03		
4	P0001	S0027	01-05-17	1	5.3	6	6.25	PR14		PR03		
5	P0001	S0104	01-03-17	1	5.3	21	6.25	PR14		PR03		
6	P0001	S0106	01-02-17	1	5.3	3	6.25	PR14		PR03		
7	P0001	S0012	01-02-17	1	5.3	0	6.25	PR14		PR03		
8	P0001	S0013	01-02-17	2	10.59	0	6.25	PR14		PR03		
9	P0001	S0056	01-02-17	1	5.3	6	6.25	PR14		PR03		
10	P0001	S0103	01-02-17	1	5.3	10	6.25	PR14		PR03		
11	P0004	S0085	01-03-17	1	3.81	28	4.5	PR14		PR03		
12	P0004	S0006	01-03-17	1	3.81	17	4.5	PR14		PR03		
13	P0004	S0044	01-02-17	1	3.81	7	4.5	PR14		PR03		
14	P0005	S0066	01-03-17	1	28.73	5	33.9	PR14		PR03		
15	P0011	S0115	01-04-17	1	42.29	1	49.9	PR14		PR03		
16	P0011	S0038	01-04-17	1	42.29	8	49.9	PR14		PR03		
17	P0011	S0108	01-05-17	1	42.29	7	49.9	PR14		PR03		
18	P0015	S0014	01-05-17	1	2.41	10	2.6	PR10	low	PR03		
19	P0015	S0024	01-05-17	1	1.85	14	2.6	PR10	low	PR03		
20	P0015	S0026	01-05-17	1	1.85	10	2.6	PR10	low	PR03		
21	P0015	S0012	01-05-17	4	7.96	14	2.6	PR10	low	PR03		
22	P0015	S0066	01-05-17	1	2.41	20	2.6	PR10	low	PR03		
23	P0015	S0002	01-05-17	1	1.85	18	2.6	PR10	low	PR03		
24	P0015	S0009	01-05-17	1	1.85	19	2.6	PR10	low	PR03		
25	P0015	S0097	01-05-17	1	1.85	22	2.6	PR10	low	PR03		
26	P0015	S0085	01-05-17	3	5.56	45	2	PR03	verylow	PR03		
27	P0015	S0082	01-05-17	1	1.85	2	2.6	PR10	low	PR03		
28	P0015	S0135	01-05-17	2	3.7	8	2.6	PR10	low	PR03		
29	P0015	S0130	01-05-17	1	2.41	9	2.6	PR10	low	PR03		

sales

product\_hierarchy

store\_cities

store\_names

product\_names

city\_names

# Module 1, Lesson 2: Data Analysis Using Pivot Tables

## Sales by City Analysis:

Row Labels	Average sales	Total Sales	Total Revenue
Edinburgh	0.675117518	5537.989	18146.5
Helsinki	0.382453429	3674.995	17880.98
London	0.644494744	1900.615	5080.29
Saint Petersburg	0.245859621	1247	4545.47
Vienna	0.188910226	787	1713.89
Grand Total	0.438267909	13147.599	47367.13

## Sales Data Analysis:

Row Labels	Average Sales	Total Sales	Total Stock	Total Revenue
Currys (National Chain)	0.152429467	389	31473	1055.21
Darty	0.644494744	1900.615	45375.05	5080.29
DIGI	0.675117518	5537.989	133535.99	18146.5
Electro World (National Chain)	0.382453429	3674.995	153521.285	17880.98
Elettrodomestici Rossi	0.245859621	1247	71470	4545.47
Euronics Lisboa (National Chain)	0.246592317	398	30929	658.68
Grand Total	0.438267909	13147.599	466304.325	47367.13

## Product performance analysis:

Row Labels	Average sales	Total Sales	Total Revenue
Stackable Washer and Dryer	0.108695652	10	18.09
AutoBlenderhub	0.335526316	51	155.48
AutoBlenderwave	0.029411765	3	42.28
AutoDryerdrive	0.06122449	12	119.4
AutoFridgetron	0	0	0
AutoGrillflow	0.111111111	5	9.25
AutoHeatercast	0.243243243	36	163.26
AutoHeaterdrive	0.176056338	25	54.78
AutoTVdrive	0.071428571	14	105
AutoVacuumcast	0.4	2	99.06
AutoVacuumhub	0.255555556	23	305.21
AutoWasherflow	0.086092715	13	135.7
Baby Monitors with Video and Temperature Monitoring	0.008474576	1	8
Beverage Center	0.043478261	2	7.62
Bottom Freezer Refrigerator with Water Dispenser	0.14953271	16	96.61
Bread Machine with Gluten-Free Setting	0.333333333	14	55.41
Canister Vacuum Cleaner	0.544827586	79	116.84
Canister Vacuums with HEPA Filter and Variable Suction Control	0.181818182	12	17.78
Chest Freezer with Lock and Key	0	0	0
Convection Oven	0.462686567	93	201.1
Counter-depth Refrigerator	0.414634146	102	212.36
Countertop Pizza Oven with Rotisserie Function	0.019230769	2	9.16
Curling Iron with Interchangeable Barrels	0.2	7	19.46
CyberBlenderflow	0.052631579	8	38.3

# Module 1, Lesson 2: Data Analysis Using Pivot Tables

---

## **Significance of Data Cleaning and Preparation:**

Data cleaning and preparation is a crucial step in data analysis for several reasons:

- **Data Quality:** Ensures that the data is accurate, consistent, and reliable, which is essential for drawing valid conclusions.
- **Analysis Accuracy:** Inaccurate or inconsistent data can lead to erroneous results and misleading insights.
- **Efficiency:** Clean and well-prepared data is easier to work with and analyze, saving time and effort.
- **Data Integrity:** Maintains the integrity of the data, preventing errors and ensuring that analysis is based on accurate information.



# Module 2, Lesson 1: Data Querying Using PostgreSQL

Tables stored in BICapstone database:

```
BICapstone=# \dt
               List of relations
 Schema |      Name      | Type  | Owner
-----+-----+-----+-----
 public | city_names     | table | postgres
 public | product_hierarchy | table | postgres
 public | product_names  | table | postgres
 public | sales          | table | postgres
 public | store_cities   | table | postgres
 public | store_names    | table | postgres
(6 rows)
```

```
BICapstone=#
```

Query in BICapstone database:

The screenshot shows a PostgreSQL query interface with a query editor at the top containing the SQL query: `SELECT * FROM SALES LIMIT 5;`. Below the editor, the 'Data Output' tab is active, displaying a table with 5 rows of data. The table has columns: `product_id` (character varying), `store_id` (character varying), `date` (date), `sales` (double precision), `revenue` (double precision), `stock` (double precision), and `price` (double). A green status bar at the bottom indicates the query was successfully run with a total runtime of 144 msec and 5 rows affected.

	product_id character varying	store_id character varying	date date	sales double precision	revenue double precision	stock double precision	price double
1	P0001	S0002	2017-02-01	0	0	8	
2	P0001	S0012	2017-02-01	1	5.3	0	
3	P0001	S0013	2017-02-01	2	10.59	0	
4	P0001	S0023	2017-02-01	0	0	6	
5	P0001	S0025	2017-02-01	0	0	1	

✓ Successfully run. Total query runtime: 144 msec. 5 rows affected. ✕

Total rows: 5 of 5    Query complete 00:00:00.144    1 n 1 Col 20

# Module 2, Lesson 1: Data Querying Using PostgreSQL

- **Significance of Database Setup**
- Setting up the database in PostgreSQL provides a structured and organized environment for storing and managing the data. This facilitates efficient data querying, analysis, and visualization. Key benefits include:
- **Data Integrity:** Ensures data consistency and accuracy through proper data types, constraints, and indexing.
- **Data Security:** Provides mechanisms to protect sensitive data from unauthorized access.
- **Scalability:** Can handle large datasets and support complex queries.
- **Performance Optimization:** Allows for fine-tuning query performance through indexing and other optimization techniques.
- **Integration with Visualization Tools:** Easily integrates with tools like Tableau and Power BI for creating interactive dashboards.

## Sales Performance Analysis:

Query

Query History

Scratch Pad

1

2

3

4

5

6

7

8

9

10

11

12

```
SELECT
  p.product_id,
  c.store_id,
  c.city_id,
  SUM (sl.sales ) AS total_sales,
  SUM (sl.revenue) AS total_revenue,
  AVG (sl.price) AS average_price
FROM
  SALES sl
INNER JOIN product_hierarchy p ON sl.product_id = p.product_id
INNER JOIN store_cities c ON sl.store_id = c.store_id
GROUP BY p.product_id, c.store_id, c.city_id;
```

Data Output

Messages

Notifications

SQL

	product_id character varying	store_id character varying	city_id character varying	total_sales double precision	total_revenue double precision	average_price double precision
1	P0454	S0004	C022	0	0	
2	P0103	S0084	C022	9	22.080000000000002	2
3	P0390	S0138	C022	1	8.35	9
4	P0154	S0060	C029	0	0	1
5	P0436	S0011	C031	1	2.78	
6	P0681	S0049	C031	0	0	

Total rows: 1000 of 13220    Query complete 00:00:00.194    Ln 6, Col 18



## CREATION OF DATA CUBES WITH ROLLUP:

Query
Query History
Scratch Pad X

```

1 SELECT
2     sl.store_id,
3     TO_CHAR(DATE_TRUNC('month', sl.date), 'YYYY-MM') AS sales_mon
4     SUM(sales) AS monthly_sales
5 FROM
6     sales sl
7 GROUP BY
8     ROLLUP (store_id, DATE_TRUNC('month', sl.date))
9 ORDER BY
10    store_id, sales_month;
11
12
13

```

Data Output
Messages
Notifications

+
File
Dropdown
Clipboard
Dropdown
Trash
Database
Download
Refresh
SQL

	store_id character varying	sales_month text	monthly_sales double precision
1	S0001	2017-02	67.695
2	S0001	2017-03	35
3	S0001	[null]	102.695
4	S0002	2017-02	73.965
5	S0002	2017-03	28
6	S0002	[null]	101.965

Total rows: 352 of 352
Query complete 00:00:00.148
Ln 13, Col 1

## SUMMARIZING DATA ALONG HIERARCHIES:

**Query**
Query History
Scratch Pad X

```

1 SELECT
2     COALESCE(ph.hierarchy1_id, 'Total') AS level1_id,
3     COALESCE(ph.hierarchy2_id, 'Total') AS level2_id,
4     SUM(s.sales) AS total_sales
5 FROM
6     sales s
7 JOIN
8     product_hierarchy ph
9 ON
10    s.product_id = ph.product_id
11 GROUP BY
12     ROLLUP(ph.hierarchy1_id, ph.hierarchy2_id)
13 ORDER BY
14     level1_id, level2_id;
```

**Data Output**
Messages
Notifications

	level1_id character varying	level2_id character varying	total_sales double precision
1	H00	H0000	1380
2	H00	H0001	617
3	H00	H0002	849
4	H00	H0003	4268
5	H00	H0004	613.366
6	H00	Total	7727.366000000001

# Module 2, Lesson 2: Data Analysis Using PostgreSQL

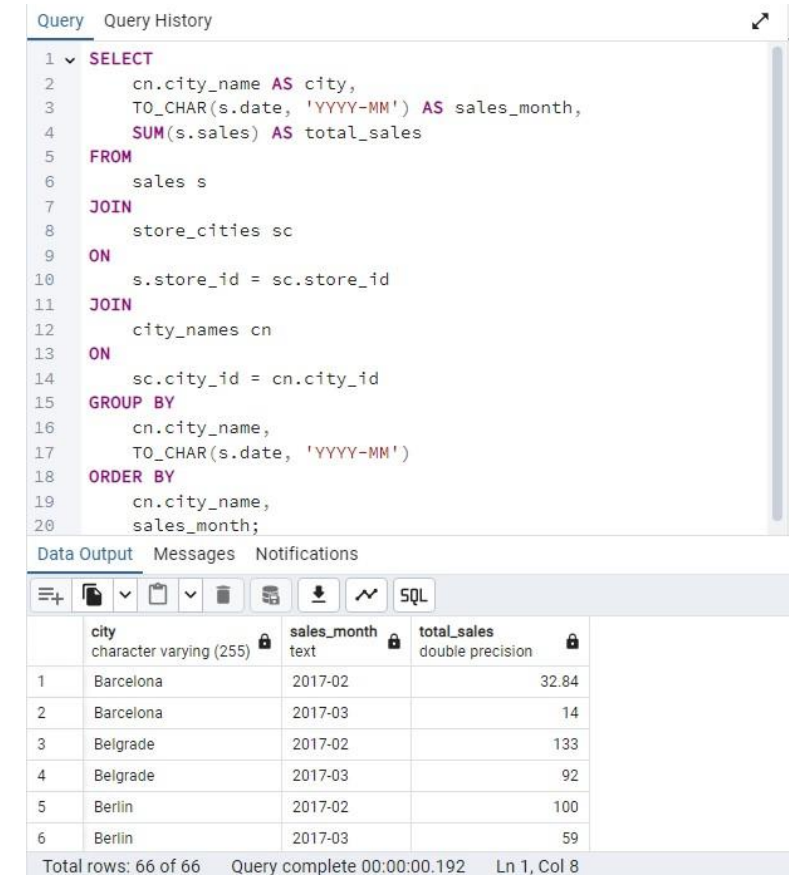
## Optimizing Sales Strategies and Stock Management

The insights gained from these analyses can be used to:

- Identify seasonal trends:** Determine peak sales periods and adjust inventory levels accordingly.
- Optimize product mix:** Identify high-performing and low-performing product categories to focus resources and adjust inventory.
- Improve store operations:** Analyze sales performance by store to identify areas for improvement and optimize store-level operations.
- Enhance marketing efforts:** Target promotional activities based on seasonal trends and customer preferences.

By leveraging data cubes and hierarchical analysis, businesses can gain a deeper understanding of their sales data and make informed decisions to optimize sales strategies and stock management

## SALES TRENDS OVER TIME AND ACROSS VARIOUS REGIONS:



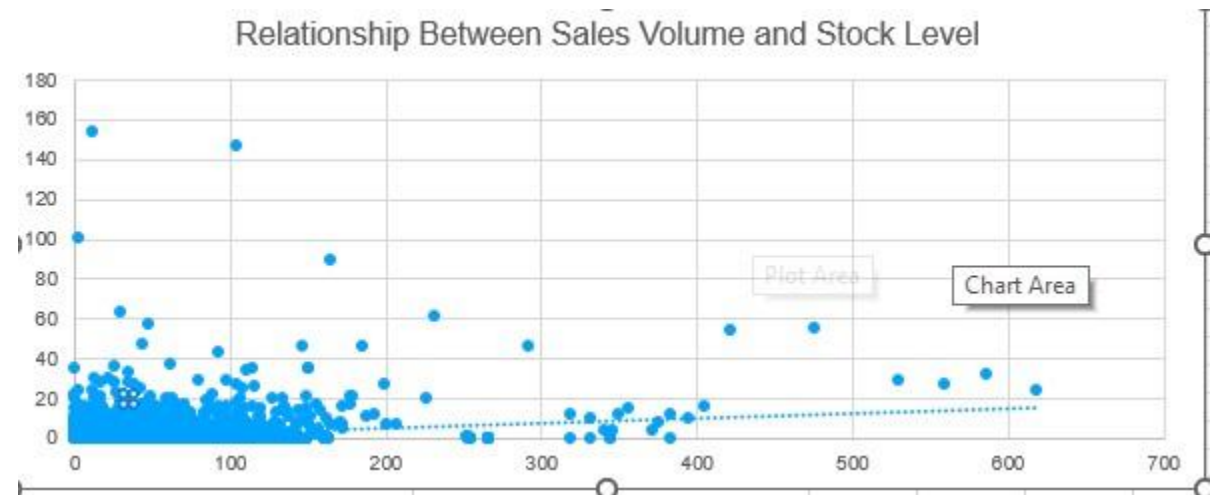
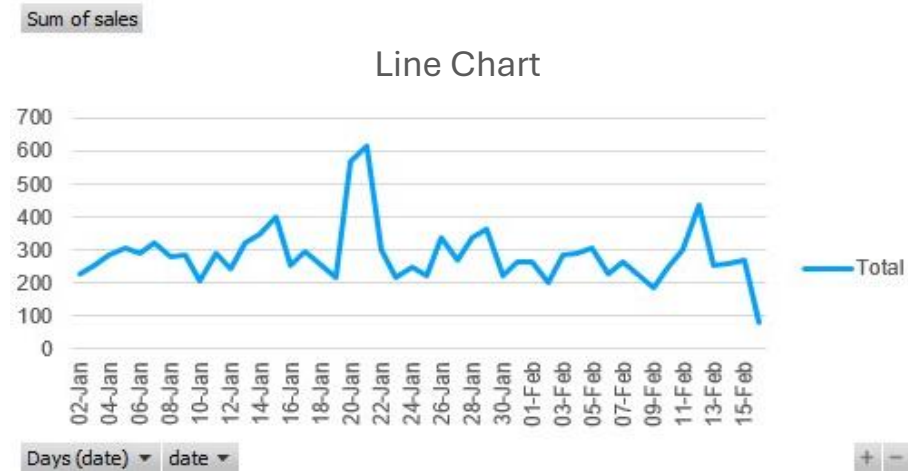
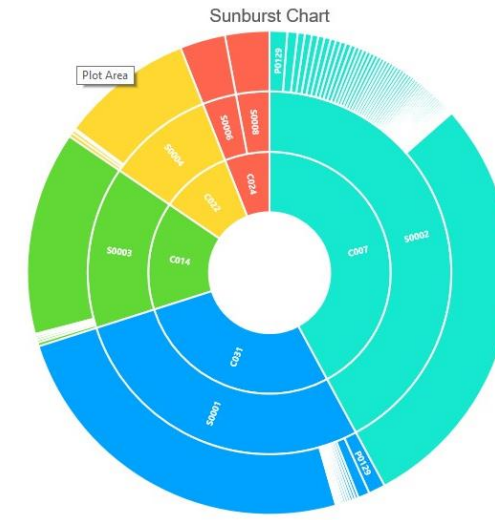
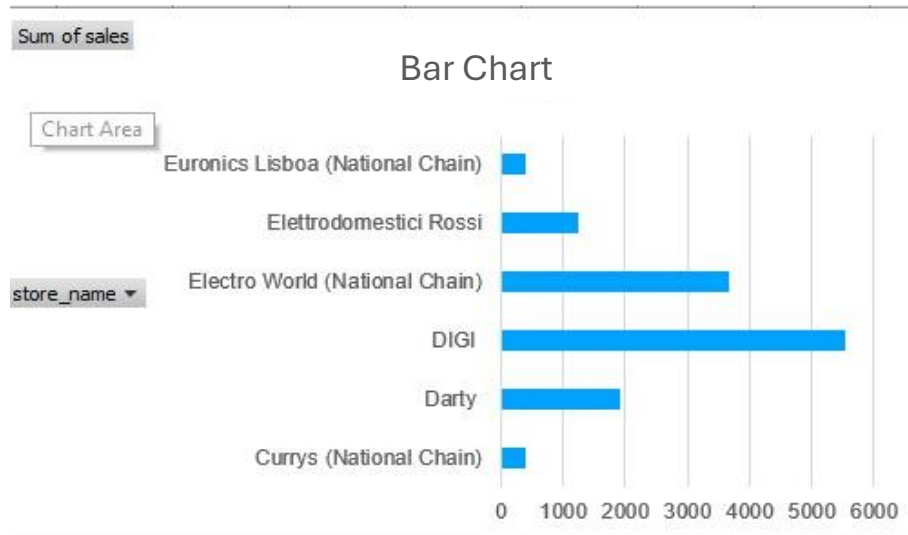
The screenshot displays a PostgreSQL query editor interface. The top section shows a SQL query for selecting sales data by city and month. The bottom section shows the results of the query in a table format.

```
1 SELECT
2   cn.city_name AS city,
3   TO_CHAR(s.date, 'YYYY-MM') AS sales_month,
4   SUM(s.sales) AS total_sales
5 FROM
6   sales s
7 JOIN
8   store_cities sc
9 ON
10  s.store_id = sc.store_id
11 JOIN
12  city_names cn
13 ON
14  sc.city_id = cn.city_id
15 GROUP BY
16   cn.city_name,
17   TO_CHAR(s.date, 'YYYY-MM')
18 ORDER BY
19   cn.city_name,
20   sales_month;
```

	city character varying (255)	sales_month text	total_sales double precision
1	Barcelona	2017-02	32.84
2	Barcelona	2017-03	14
3	Belgrade	2017-02	133
4	Belgrade	2017-03	92
5	Berlin	2017-02	100
6	Berlin	2017-03	59

Total rows: 66 of 66    Query complete 00:00:00.192    Ln 1, Col 8

# Module 3, Lesson 1: Data Visualization Using Excel



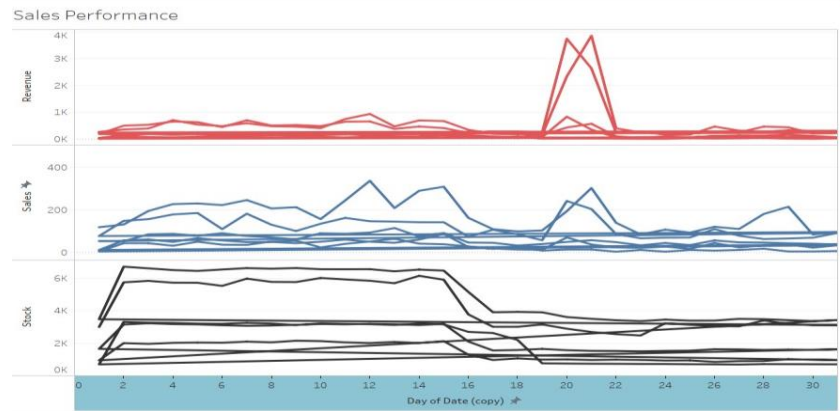
# Module 3, Lesson 1: Data Visualization Using Excel

---

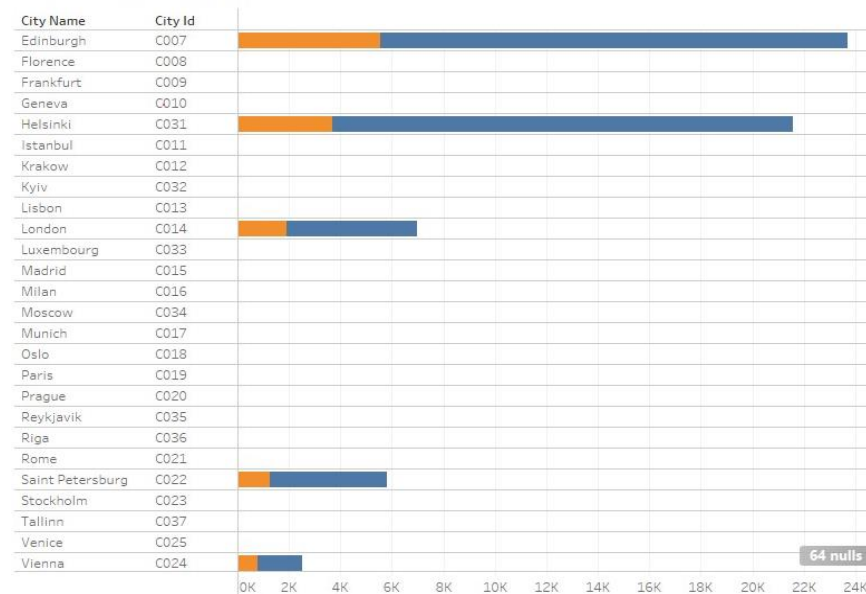
- **Significance of the Charts**
- These charts contribute to a comprehensive understanding of the data by:
- **Visualizing Trends:** Identifying patterns, trends, and outliers that might not be easily apparent in tabular data.
- **Comparing Categories:** Comparing different groups or categories to understand their relative performance.
- **Revealing Relationships:** Identifying relationships between variables that can inform decision-making.
- **Communicating Insights:** Effectively conveying complex information to stakeholders in a visual format.
- By using a combination of these chart types, you can gain valuable insights into your retail sales data and make data-driven decisions.

[illegible]

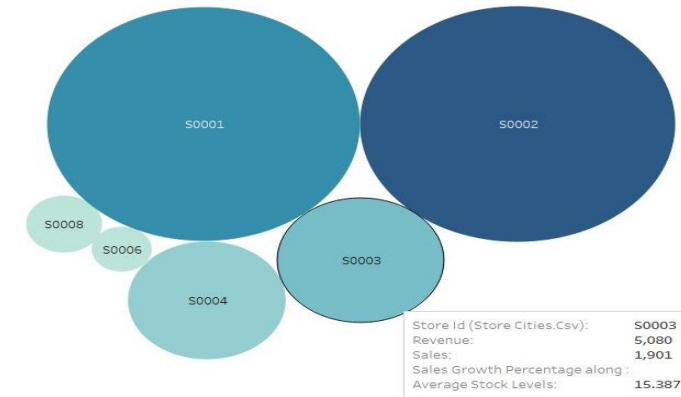




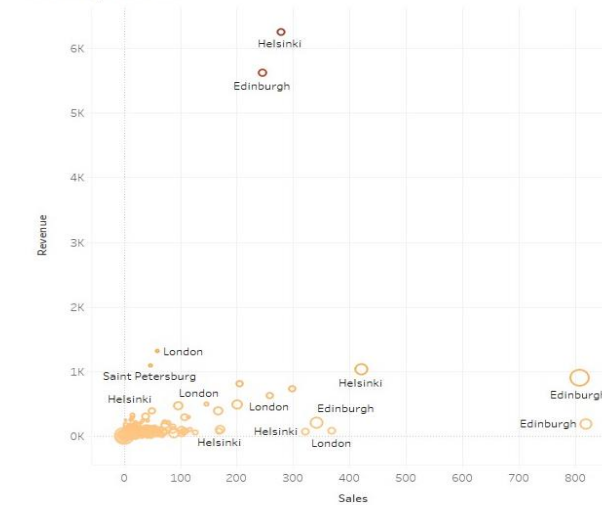
## Regional Sales Analysis



### Sales by Store



### Sales by Product



# Module 4, Lesson 1: Basic Tableau Visualizations

---

- **Significance of Visualizations**
- These visualizations facilitate clear communication of insights by:
- **Visualizing Trends:** Making patterns and trends in the data easier to identify.
- **Comparing Categories:** Providing a visual comparison of different groups or categories.
- **Revealing Relationships:** Identifying relationships between variables that might not be apparent in tabular data.
- **Storytelling:** Creating a compelling narrative through the use of visuals.
- By effectively using these visualizations, you can communicate your findings to stakeholders in a clear, concise, and engaging manner.

# Module 4, Lesson 2: Advanced Visualizations Using Tableau

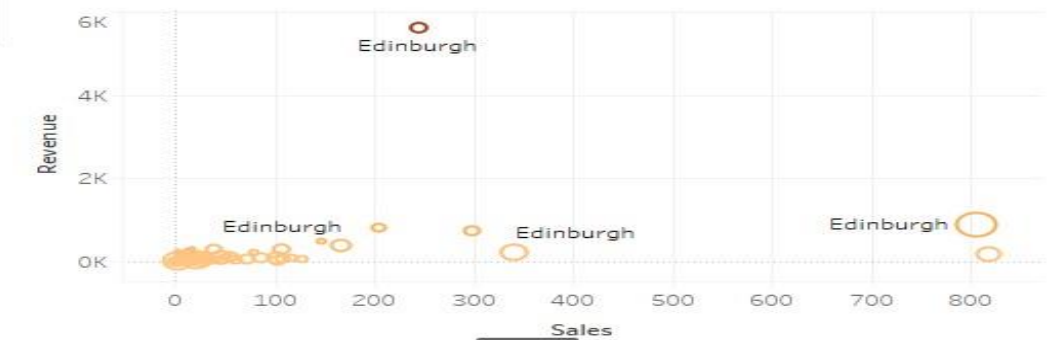
Regional Sales Analysis



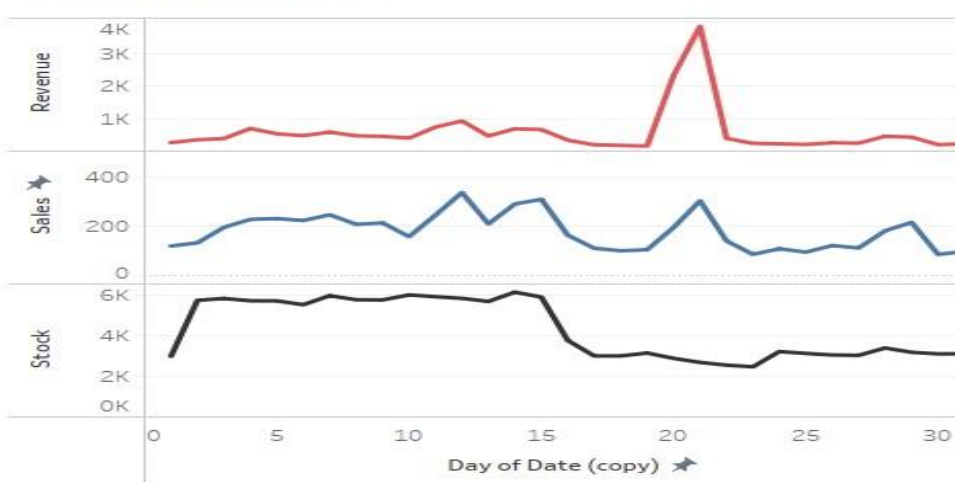
Measure Names



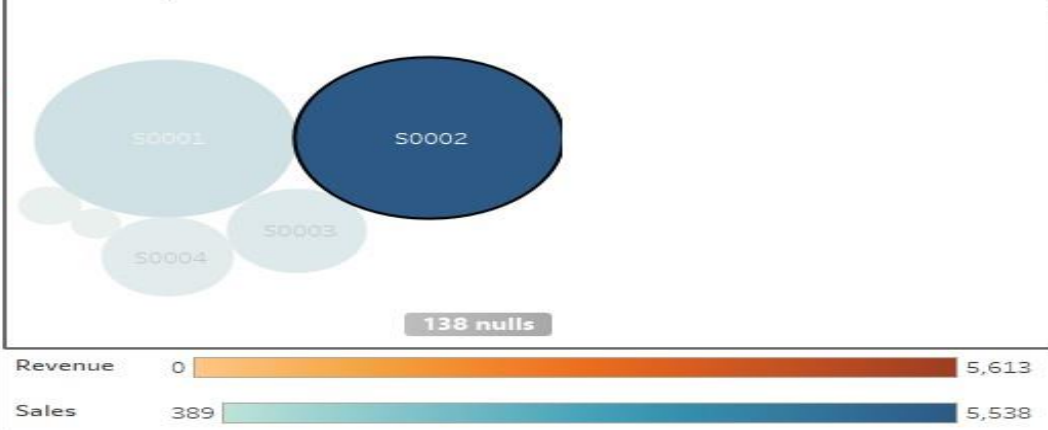
Sales by Product



Sales Performance



Sales by Store





# Module 4, Lesson 2: Advanced Visualizations Using Tableau

---

## Significance of Visualizations and Actions

- **Enhanced Insights:** Dashboards provide a comprehensive view of the data, allowing for deeper insights and analysis.
- **Interactivity:** Actions make the dashboard interactive, enabling users to explore the data dynamically.
- **Decision Support:** Well-designed dashboards can support decision-making by providing actionable insights and visualizations.

By following these steps and leveraging Tableau's features, you can create effective dashboards that communicate your analysis results and facilitate decision-making.



# Discussion

# Insights and Recommendations

---

## Key Insights:

- **Seasonal Trends:** The analysis identified distinct seasonal patterns in sales, with peak periods during holidays and festive seasons. This information can be used to optimize inventory levels and promotional activities.
- **Product Performance:** Certain product categories consistently outperform others, indicating a need to focus on these areas and explore opportunities for growth.
- **Regional Variations:** Sales performance varies across different regions, highlighting the importance of tailoring marketing strategies and inventory management to local preferences.
- **Customer Segmentation:** Customer segmentation revealed distinct groups with different purchasing behaviors, enabling targeted marketing and promotions.

## Communication Effectiveness:

- The presentation effectively communicates findings to stakeholders by:
- **Visualizing Data:** Using clear and informative visualizations to present complex data in a digestible format.
- **Storytelling:** Organizing the presentation in a logical and engaging manner, guiding the audience through the key findings.
- **Supporting Evidence:** Providing concrete examples and evidence to support the claims and recommendations.
- **Addressing Questions:** Being prepared to answer questions and provide additional context.

# Insights and Recommendations

---

## Recommendations:

- **Inventory Optimization:** Adjust inventory levels based on seasonal trends and product performance to minimize stockouts and excess inventory.
- **Targeted Marketing:** Develop tailored marketing campaigns for different customer segments to increase customer engagement and loyalty.
- **Product Portfolio Optimization:** Focus on expanding and promoting high-performing product categories while discontinuing or modifying underperforming ones.
- **Geographical Expansion:** Evaluate opportunities for expanding the retail chain into regions with high growth potential.
- **Continuous Monitoring:** Implement a system for ongoing data monitoring and analysis to identify emerging trends and adjust strategies accordingly.

By implementing these recommendations, the retail chain can optimize its operations, enhance customer satisfaction, and drive sustainable growth.



# Conclusion

# Summary

---

- The retail chain can optimize sales performance by aligning inventory levels with seasonal trends and focusing on high-performing product categories.
- Tailored marketing strategies based on customer segmentation can enhance customer engagement and loyalty.
- Geographical analysis can inform decisions regarding store expansion and resource allocation.
- Continuous data monitoring and analysis are essential for identifying emerging trends and making informed business decisions.



# Appendix

# Appendix #

---

These are the only cities who have sales in their region.

Row Labels	Average sales	Total Sales	Total Revenue
Edinburgh	0.675117518	5537.989	18146.5
Helsinki	0.382453429	3674.995	17880.98
London	0.644494744	1900.615	5080.29
Saint Petersburg	0.245859621	1247	4545.47
Vienna	0.188910226	787	1713.89
<b>Grand Total</b>	<b>0.438267909</b>	<b>13147.599</b>	<b>47367.13</b>