NAAN MUDHALVAN

PHASE 3 PROJECT SUBMISSION

**PRODUCT SALES ANALYSIS**

**TEAM MEMBERS:**

PARVATHY A - 2021504533

PAVITHRAN E - 2021504535

SHAMEEM AHAMED S - 2021504542

THUSHIYANTH K - 2021504555

**PROBLEM DEFINITION:**

This project involves using IBM Cognos to analyze sales data and extract insights about top-selling products, peak sales periods, and customer preferences. The objective is to help businesses improve inventory management and marketing strategies by understanding sales trends and customer behavior. This project includes defining analysis objectives, collecting sales data, designing relevant visualizations in IBM Cognos, and deriving actionable insights.

**DATABASE LINK:**

[**https://www.kaggle.com/datasets/ksabishek/product-sales-data**](https://www.kaggle.com/datasets/ksabishek/product-sales-data)

**OBJECTIVES:**

* Analysing data related to product sales.
* Generating valuable insights from the data.
* Based on the insights, recommendations must be formulated to address issues and optimize sales and profitability.

**Cleaning and Preprocessing:**

**Importing the important packages:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

# To ignore warnings

import warnings

warnings.filterwarnings("ignore")

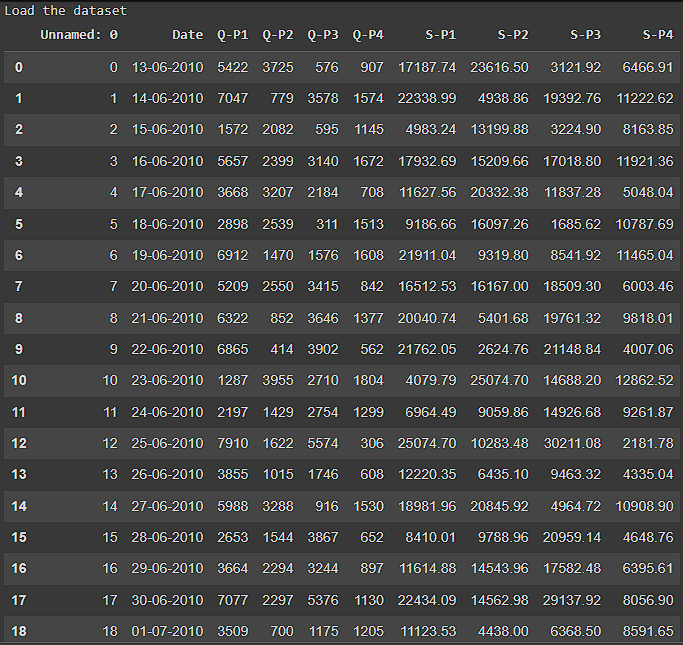
**Loading the dataset:**

print("Load the dataset")

import pandas as pd

data = pd.read\_csv('/statsfinal.csv', low\_memory=False) data.shape

data.head(30)

****

**No. of rows and columns:**

print("Load the dataset")

import pandas as pd

data = pd.read\_csv('/statsfinal.csv', low\_memory=False)

data.shape

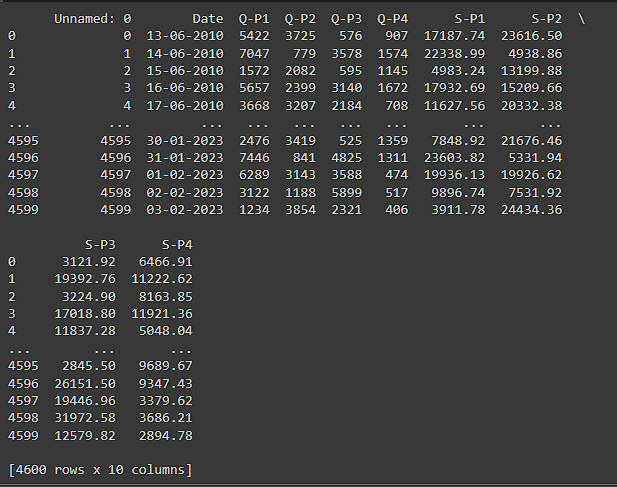
data.head(30)

input\_file = "/statsfinal.csv"

df = pd.read\_csv(input\_file)

print(df)

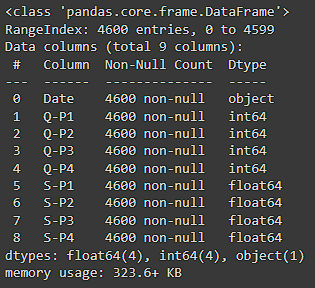
**OUTPUT:**



**DROPPING FIRST COLUMN:**

data = data.drop(columns=['Unnamed: 0'])

data.info()



**Cleaned File:**

output\_file = "cleaned\_file.csv"

df.to\_csv(output\_file, index=False)

print(output\_file)

**Output:**

cleaned\_file.csv

**Loading the cleaned CSV file:**

input\_file = "cleaned\_file.csv"

df = pd.read\_csv(input\_file)

**Histogram visualization:**

plt.figure(figsize=(8, 6))

plt.hist(df['S-P1'], bins=20, color='red')

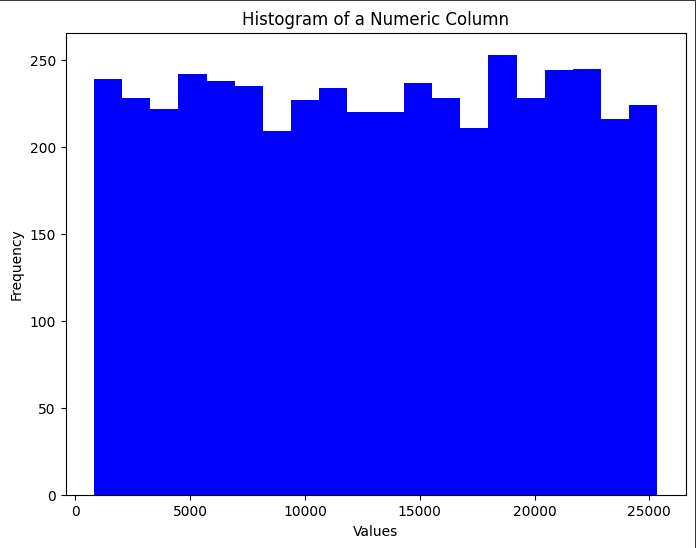
plt.title('Histogram of a Numeric Column')

plt.xlabel('Values')

plt.ylabel('Frequency')

plt.show()

**Output:**

****

**Scatter Plot Visualization:**

plt.figure(figsize=(8, 6))

plt.scatter(df['S-P1'], df['S-P2'], color='blue', alpha=0.5)

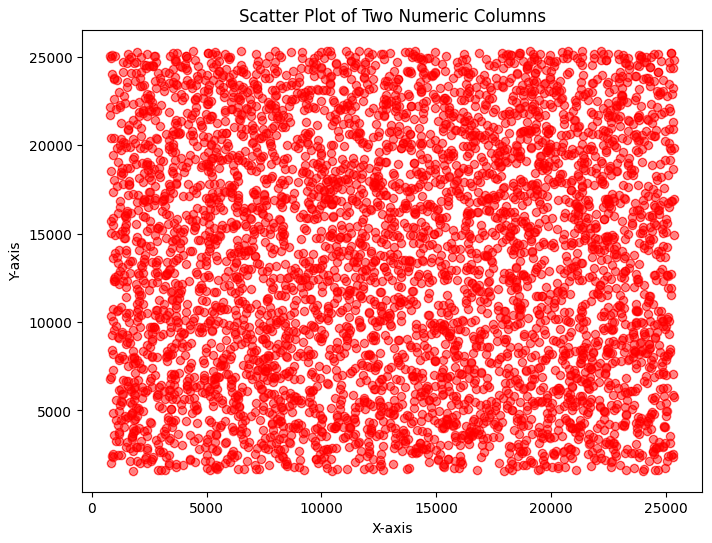
plt.title('Scatter Plot of Two Numeric Columns')

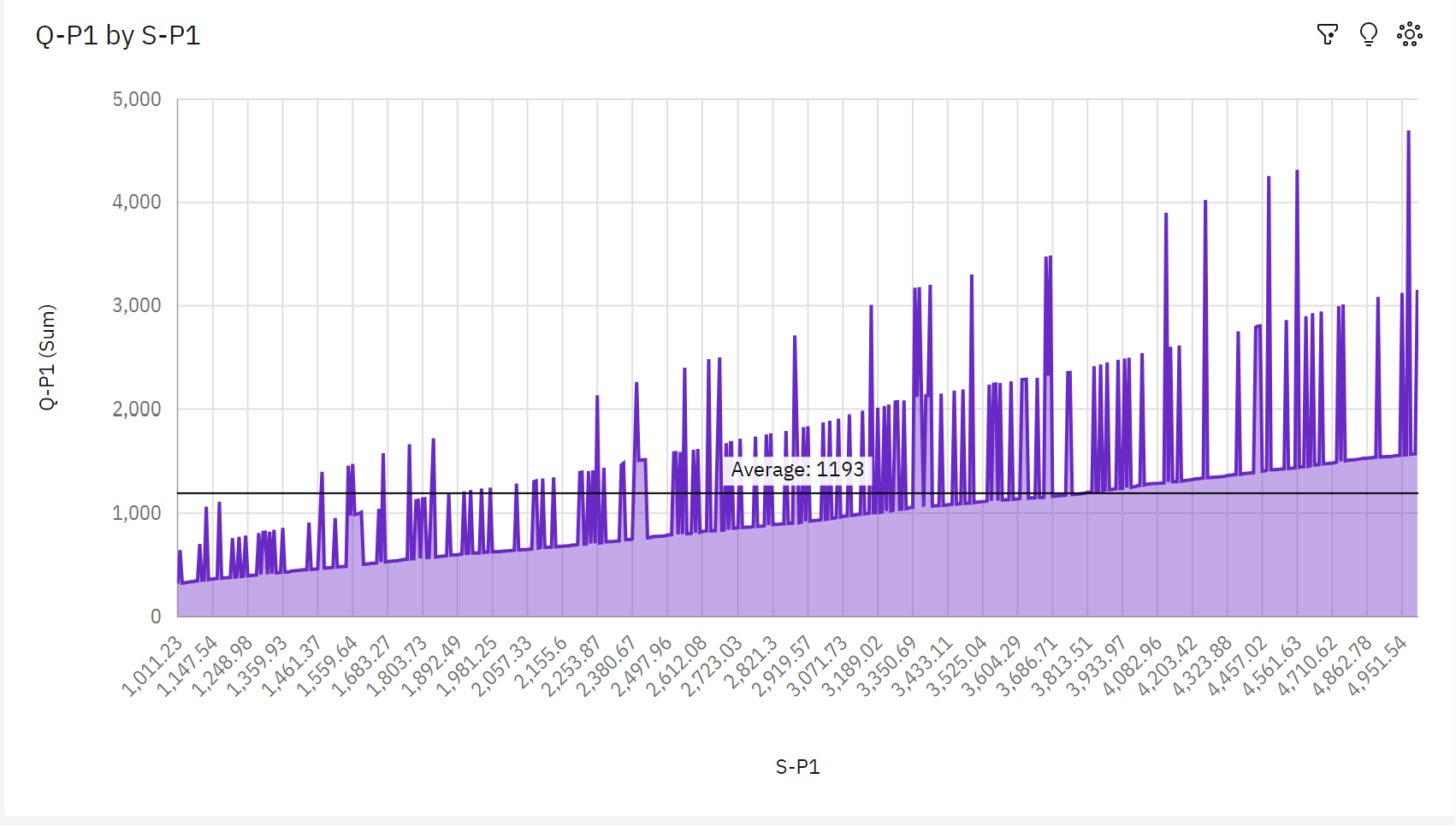
plt.xlabel('X-axis')

plt.ylabel('Y-axis')

plt.show()

**Output:**

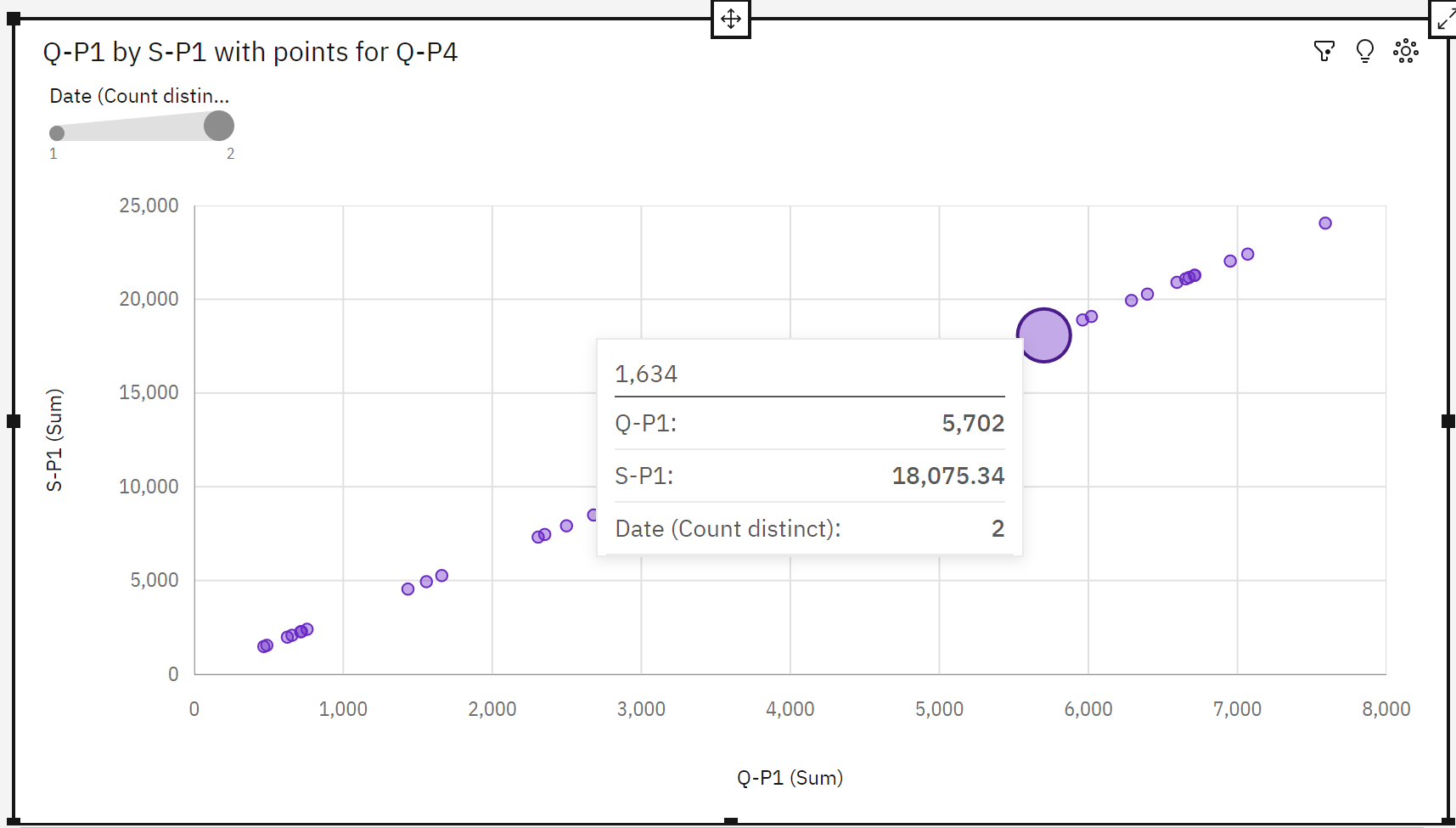
****

**VISUALIZATION 1:**

**Insights:**

* Over all values of S-P1, the sum of Q-P1 is nearly 678 thousand.
* **Q-P1** ranges from **319**, when **S-P1** is **1011.23**, to **over 4500**, when **S-P1** is **4964.22**.
* **S-P1** **3683.54** has the highest **Total Q-P4** but is ranked #**6** in **Total Q-P1**.
* **S-P1** **4964.22** has the highest **Total Q-P1** but is ranked #**238** in **Total Q-P4**.
* **Q-P4** and **Q-P1** diverged the most when **S-P1** is **1122.18**, and when **Q-P4** was **over three thousand** higher than the **Q-P1**.
* **3683.54** **Q-P4** at **over 5 thousand** is **31**% higher than the **Q-P1** of **almost 3500**.

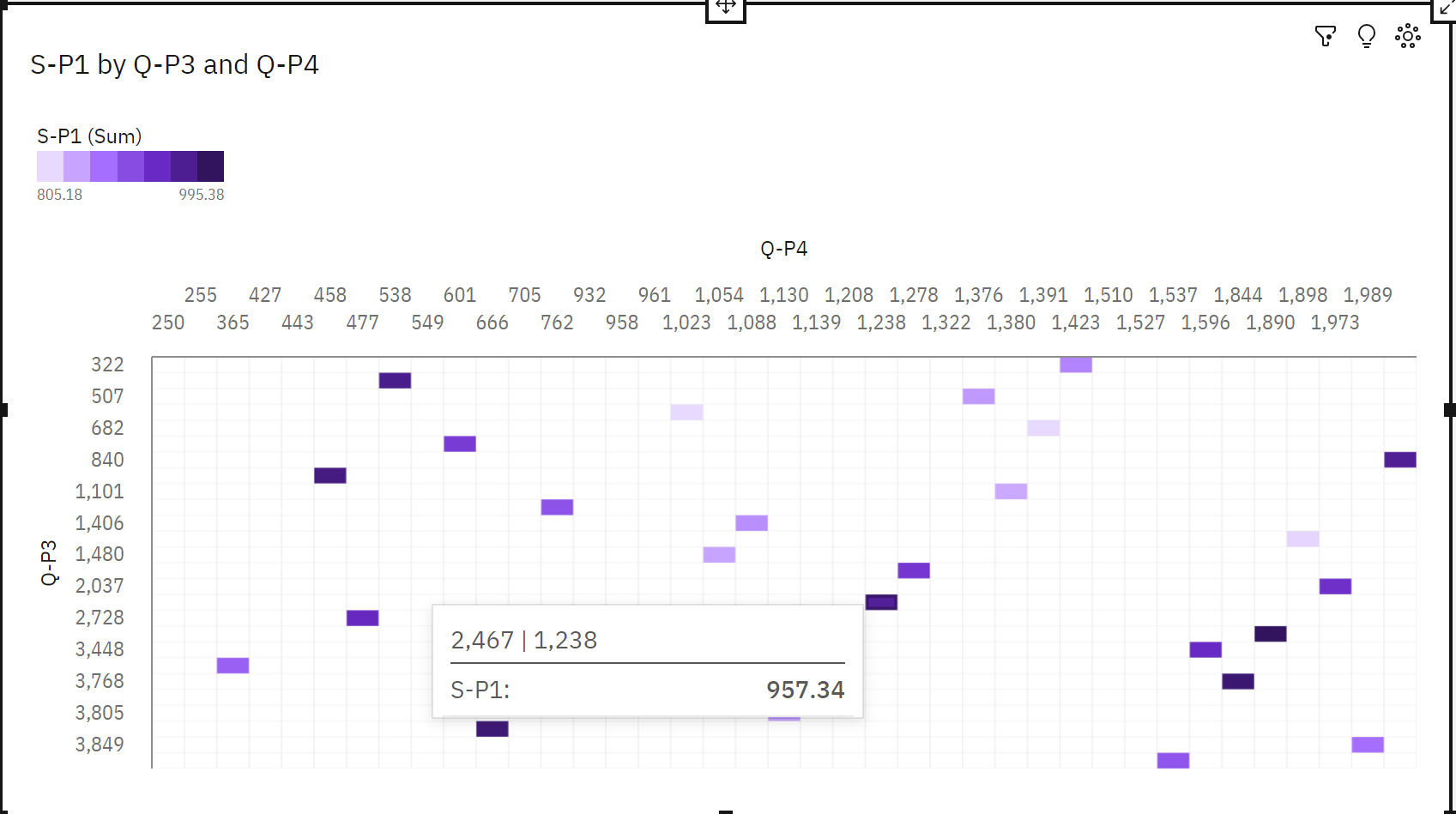
**VISUALIZATION 2:**



**Insights:**

* **S-P4** at **over 23 thousand** is **76**% higher than the **Q-P1** of **over 5500**.
* **S-P4** and **Q-P1** diverged the most when **Q-P4** is **1634**, and when **S-P4** was **nearly 18 thousand** higher than the **Q-P1**.
* **Q-P4** **960** has the highest **Total Q-P1** but is ranked #**30** in **Total S-P4.**
* **Q-P4 1634 has the highest Total S-P4 but is ranked #13 in Total Q-P1.**
* **The total of S-P1 is nearly 575 thousand.**

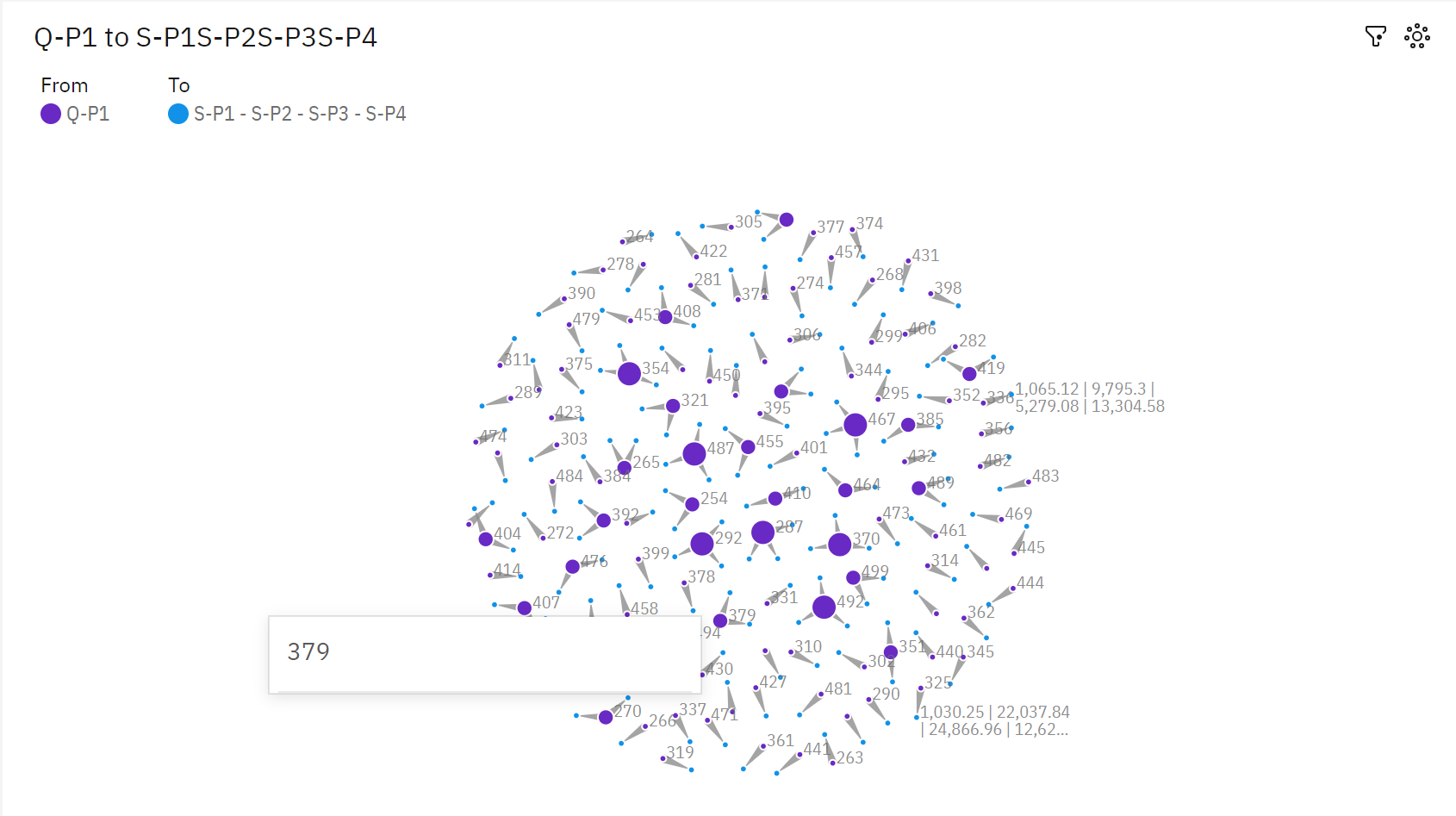
**VISUALIZATION 3:**



**Insights:**

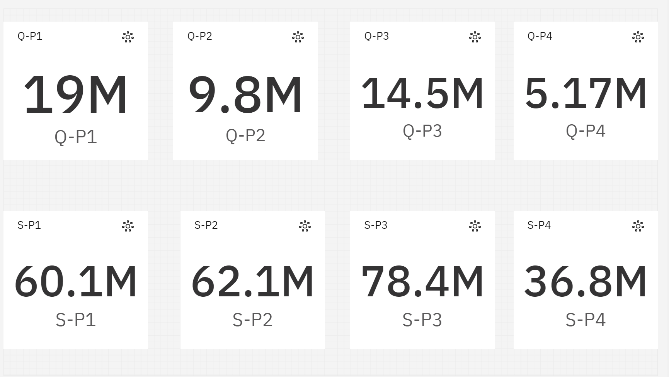
* The summed values of **S-P1** range from **805.2** to **995.4**.
* For **S-P1**, the most significant value of **Q-P4** is **1890**, whose respective **S-P1** values add up to **995.4**, or **2.8** % of the total.
* For **S-P1**, the most significant value of **Q-P3** is **2773**, whose respective **S-P1** values add up to **995.4**, or **2.8** % of the total.
* Across all values of **Q-P3** and **Q-P4**, the sum of **S-P1** is **over 35 thousand**
* **2773** **S-P4** at **13,476** is **93**% higher than the **S-P1** of **995.4**.
* **250** **S-P1** at **985.9** is **68**% higher than the **Q-P1** of **311**.
* **Q-P4** **250** has the highest values of both **S-P1** and **Q-P1**.
* **Q-P3** **2773** has the highest values of both **S-P1** and **S-P4**.
* **Q-P3** **2773** has the highest total **S-P1** due to **Q-P4** **1890**.

**VISUALIZATION 4:**



INSIGHTS:

* **S-P1** **1445.52** has the highest **Unaggregated Q-P1** but is ranked #**99** in **Total Q-P4**
* **S-P1** **1122.18** has the highest **Total Q-P4** but is ranked #**108** in **Unaggregated Q-P1**.

**SUMMARY:**

**NAAN MUDHALVAN - PHASE 1 PROJECT SUBMISSION**

**PROJECT 5 - PRODUCT SALES ANALYSIS**

**TEAM MEMBERS:**

1. Parvathy A - (2021504533)
2. Pavithran E - (2021504535)
3. Shameem Ahamed S - (2021504542)
4. Thushiyanth K - (2021504555)

**PROBLEM DEFINITION:**

This project involves using IBM Cognos to analyse sales data and extract insights about top-selling products, peak sales periods, and customer preferences. The objective is to help businesses improve inventory management and marketing strategies by understanding sales trends and customer behaviour. This project includes defining analysis objectives, collecting sales data, designing relevant visualizations in IBM Cognos, and deriving actionable insights.

**AIM:**

The primary aim of this product sales analysis is to gain a deep understanding of our sales performance, identify key insights, and develop data-driven recommendations to improve sales and profitability.

**DATABASE LINK:**

[**https://www.kaggle.com/datasets/ksabishek/product-sales-data**](https://www.kaggle.com/datasets/ksabishek/product-sales-data)

**PROJECT OBJECTIVES:**

**Sales Performance Evaluation:**

We have to assess and understand the historical sales performance of the organization over a defined period and to identify trends, fluctuations, and patterns in sales data.

**Product Performance Analysis:**

To analyse the sales performance of individual products to determine top sellers and underperforming items. To calculate and compare the contribution of each product to the overall revenue.

**Customer Segmentation:**

To Segment customers based on their purchasing behaviour and preferences. To identify high-value customer segments and understand their needs.

**DECODING THE PROBLEM:**

1. Data Analysis: The core of the problem involves analysing data related to product sales.
2. Insight Generation: The objective is to generate valuable insights from the data.
3. Recommendation Development: Based on the insights, recommendations must be formulated to address issues and optimize sales and profitability.

**DESIGN THINKING AND APPROACH:**

**1. Analysis Objectives:**

**Identify Top-Selling Products:**

Objective: To determine which products have consistently high sales.

Insight: A list of top-selling products and their contribution to overall revenue.

**Analyse Sales Trends:**

Objective: To understand the sales trends over time.

Insight: Visualization of sales trends, including seasonality and growth patterns.

**Customer Preference Analysis:**

Objective: To explore customer preferences for product categories or features.

Insight: Knowledge of which product attributes or categories are most favoured by customers.

**2. Data Collection:**

**Point of Sale (POS) Systems:**

Source: Transaction records generated at physical and online checkout points.

Method: POS systems automatically record sales transactions, including product details, prices, and timestamps.

**E-commerce Platforms:**

Source: Online sales data from e-commerce websites or platforms.

Method: E-commerce platforms often provide analytics dashboards with sales data, including product sales, customer information, and geographic data.

**Sales Reports:**

Source: Internal sales reports generated by our organization.

Method: These reports may be generated regularly and include details on sales revenue, product performance, and customer data.

When collecting sales data, we have to consider data privacy and compliance with relevant regulations, such as data protection laws (e.g., GDPR, CCPA). Additionally, we have to ensure data accuracy and consistency to perform meaningful analysis. Integration between different data sources and the use of data collection tools can streamline the process and improve data quality.

**3. Right Visualization Types:**

Appropriate visualization types can effectively communicate our key messages. Common types include:

* + Bar Charts
  + Line Charts
  + Pie Charts
  + Scatter Plots
  + Heatmaps
  + Tables

**Visualization using IBM Cognos:**

IBM Cognos is a powerful business intelligence and analytics tool that can help to create a wide range of visualizations to communicate insights from our product sales analysis.

**1. Connect to our Data Source:**

* We have to launch IBM Cognos and connect to our data source, which may include databases, spreadsheets, or data warehouses where our sales data is stored.

**2. Create a New Report or Dashboard:**

* Depending on our preferred way of presenting insights, we can create a new report or dashboard. Dashboards are particularly useful for combining multiple visualizations into a single, interactive view.

**3. Select Visualization Types:**

* Within our report or dashboard, the appropriate visualization types based on the insights we want to convey can be selected. IBM Cognos offers a wide variety of chart types, including bar charts, line charts, pie charts, scatter plots, and more.

**4. Drag and Drop Data Fields:**

* The intuitive drag-and-drop interface can be used to add data fields to our chosen visualization. For example, we can drag the "Product Name" field to the X-axis and the "Sales Amount" field to the Y-axis of a bar chart.

**5. Apply Filters and Prompts:**

* Implement filters and prompts can be used to interact with the data. Filters can help users narrow down the data they want to see, such as selecting a specific time period or product category.

**6. Customize Visualizations:**

* We can customize the appearance of our visualizations by adjusting colours, fonts, labels, legends, and other design elements. Visualizations must be visually appealing and align with the organization's branding guidelines.

**7. Add Interactivity:**

* We have to enhance our visualizations with interactive features like drill-through actions, tooltips, and clickable elements that allow users to explore the data in more detail.

**8. Test and Preview:**

* Finally, we have to test our visualizations and dashboard by previewing them within IBM Cognos to ensure they work as intended and convey the insights effectively.

**9. Monitor Usage and Performance:**

* Monitor how users interact with our visualizations to understand which insights are most valuable and where improvements can be made.

IBM Cognos offers a robust set of features and capabilities for creating interactive and informative visualizations. Whether we need simple charts or complex dashboards, it can help us effectively communicate insights from our product sales analysis to drive data-driven decision-making within our organization.

**4. Actionable Insights:**

Actionable insights are valuable pieces of information or findings derived from data analysis that can be directly applied to make informed decisions or take specific actions to achieve business goals. In the context of product sales analysis, actionable insights can guide strategies and tactics to improve sales performance, increase profitability, enhance customer satisfaction, and optimize various aspects of the sales process.

Actionable insights are valuable assets that can drive decision-making, inform strategies, and lead to concrete actions that positively impact product sales and business performance. Regularly reviewing and acting upon these insights is essential for maintaining a competitive edge in the market.

**CONCLUSION:**

The process of product sales analysis is an ongoing one, and businesses should continuously monitor, analyse, and adapt their strategies based on new data and changing market dynamics. By leveraging data effectively, organizations can make informed decisions that drive growth and success in today's competitive business landscape. So, the organization must regularly review and act upon actionable insights to maintain a competitive edge, adapt to market changes, and achieve long-term business success.