**PRODUCT SALES ANALYSIS PROJECT-PHASE 3**

**STUDENT NAME**: Reemlatha V -210621104040

**COLLEGE NAME:** JEPPIAAR INSTITUTE OF TECHNOLOGY

**COLLEGE CODE:** 2106

**DEPARTMENT**: B.E.COMPUTER SCIENCE AND ENGINEERING

**SEMESTER:**5

**COURSE NAME :** DATA ANALYTICS WITH COGNOS – GROUP 1 (IBM:DAC101)

**TEAM MEMBERS NM ID**

1.A.SHERLIN BRITTO au210621104048

2.V.REEMLATHA au210621104040

3.B.SHAMILA au210621104045

4.P.MOHANAPRIYA au210621104032

5.P.KEERTHIGA au210621104028

**PROJECT TITLE:** PRODUCT SALES ANALYSIS

**Phase 3: Development Part -1**

In this part you will begin building your project by loading and preprocessing the dataset. Start building the product sales analysis using IBM Cognos for visualization. Define the analysis objectives and collect sales data from source shared. Process and clean the collected data to ensure its accuracy and reliability.

**Project Objective:**

The project involves using IBM Cognos to analyze sales data from a company that has been selling four products for over ten years. The objective is to help the company improve inventory management and marketing strategies by understanding sales trends and customer behavior. This project includes defining analysis objectives, collecting sales data preprocessing and cleaning the provided dataset for ensuring the data accuracy and reliability, designing relevant visualizations in IBM Cognos, and deriving actionable insights.



**Analysis Objectives:**

The specific analysis objectives for this project are to:

* To identify trends in sales of all four products during certain months
* To identify the top selling product out of all four products
* To suggest whether dropping the production of any one of the products would result in a massive setback for the company

**Data Collection and Preparation:**

**Data Source:**

The data set is collected from Kaggle which is a leading collaborative data analytics and data science platform.

**Dataset Link :**

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

All the data are stored in the statsfinal.csv (Comma Separated Values) format, which is used to store the data efficiently.Using pandas, we can use the .csv format for Data Processing and Manipulation

**Data Description:**

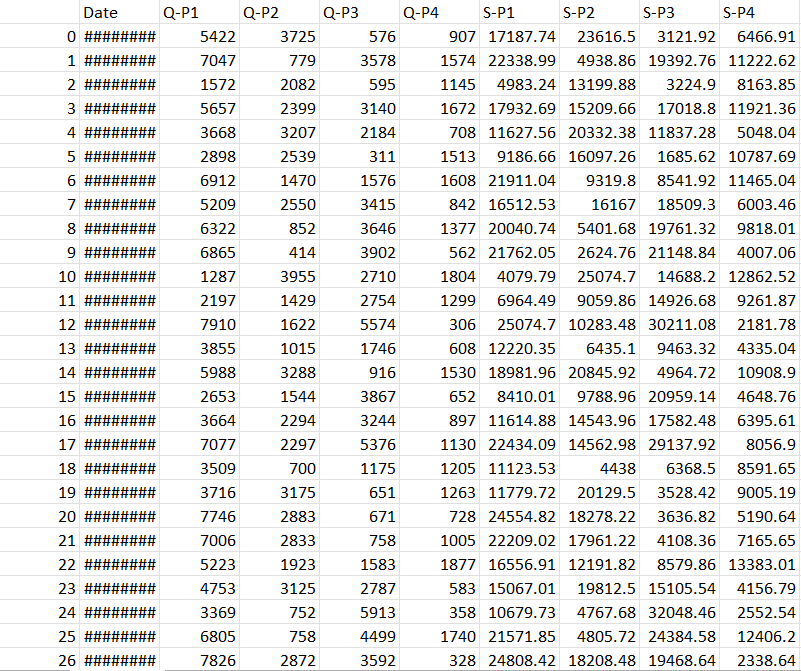
The dataset is comprised of various columns, each offering unique insights:

**Date**: This column marks the date of each data entry, allowing us to track the progression of sales and revenue over time.{ 13-06-2013 to 03-02-2023}

**P-Q1, P-Q2, P-Q3, P-Q4**: These columns denote the total unit sales for products P1, P2, P3, and P4, respectively. These figures provide an understanding of the sales volumes for each product.

**P- S1,P- S2, P-S3, P-S4**: These columns represent the total revenue generated from products P1, P2, P3, and P4, respectively. These monetary values offer a glimpse into the financial performance of each product.

**Columns Being Used:**

****

**Loading and Preprocessing Dataset:**

Data preprocessing is a critical step to ensure data quality and consistency:

1.**Importing Libraries**: We began by importing essential Python libraries, such as pandas, numpy, matplotlib, seaborn. These libraries provide powerful tools for data analysis and visualization.

**PROGRAM:**

import pandas as pd # library used for data manipulation and analysis

import numpy as np # library used for working with arrays

import matplotlib.pyplot as plt # library for plots and visualizations

import seaborn as sns # library for visualizations

**Libraries Used:**

**1.pandas**: A fundamental data manipulation library that allowed us to work with structured data efficiently.

**2.numpy**:Used for numerical operations and array handling.

**3.matplotlib**: A versatile library for creating static, animated, and interactive visualizations in Python.

**4.seaborn**: Built on top of matplotlib, seaborn provides an enhanced interface for drawing informative and attractive statistical graphics

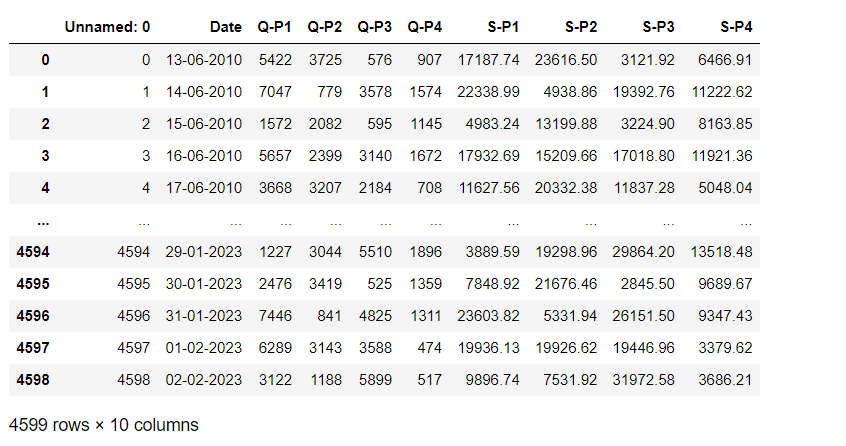
**2.Loading the Data**:We loaded the raw data from the CSV file into a pandas DataFrame

df= pd.read\_csv('statsfinal.csv')

**3.Checking the first 5 and last 5 rows**

df.head(-1)

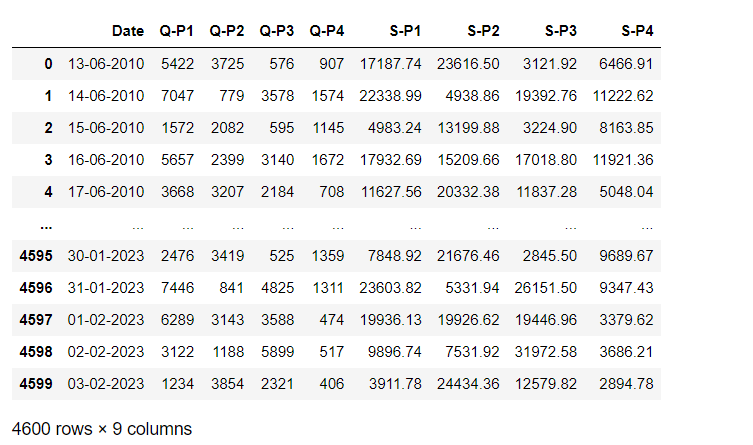
**OUTPUT:**

****

* There are 4600 rows and 10 columns.We can observe the data, starts at 13-06-2010. ends at 02-02-2023

**4.Column Cleanup**: To simplify the DataFrame and enhance readability, we removed the 'Unnamed: 0' column as it didn't offer meaningful information and it is a repeat of our ID.

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



**5. Extracting Year, Month, and Day:**

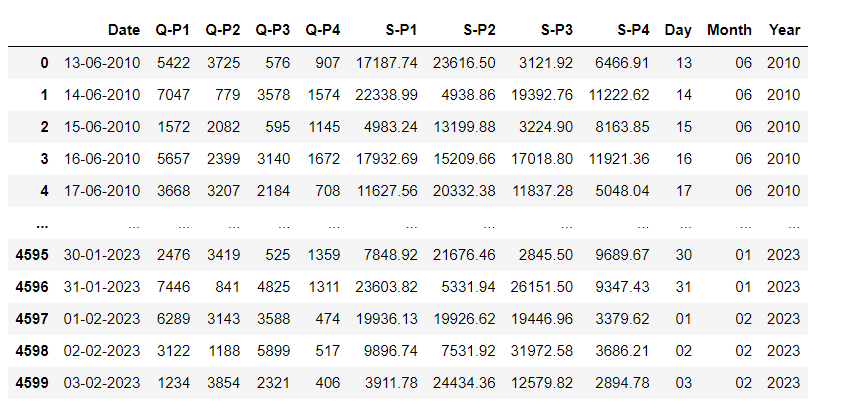
To facilitate year-to-year sales analysis, we extract the 'Year,' 'Month,' and 'Day' from the 'Date' column. This is done using a lambda function and the apply method. The 'Date' column is split into its components, creating new columns for 'Year,' 'Month,' and 'Day' in the DataFrame.

df['Day'] = df['Date'].apply(lambda x: x.split('-')[0])

df['Month'] = df['Date'].apply(lambda x: x.split('-')[1])

df['Year'] = df['Date'].apply(lambda x: x.split('-')[2])

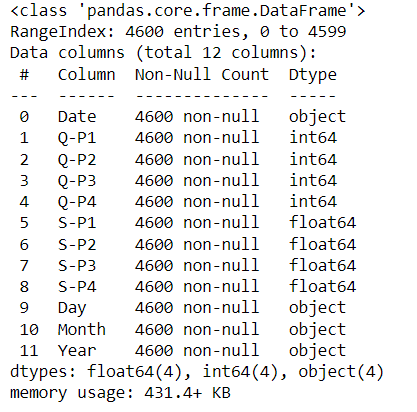
df

****

**6. Dataset Information**

We use df.info() to obtain information about the dataset. This includes data types, non-null values, and memory usage. It's crucial for understanding the structure and integrity of the dataset.

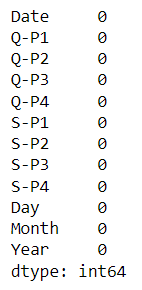
df.info()

****

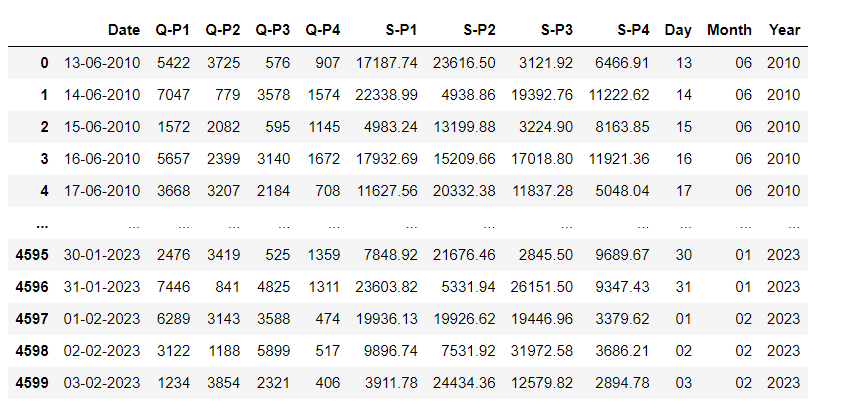
**7.Handling Missing Data**: Handling missing data is a crucial consideration in more comprehensive projects.

df.isnull().sum()

OUTPUT:



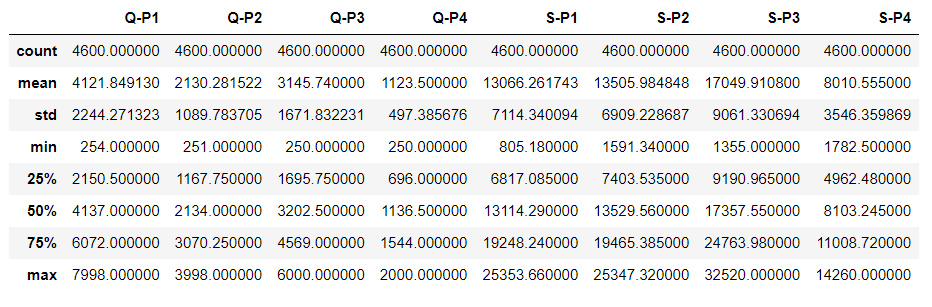
We can observe that there is no missing values

****

**8. Dataset Description**

We generate descriptive statistics for the dataset using df.describe(). This provides an overview of the distribution, central tendencies, and spread of numerical data.

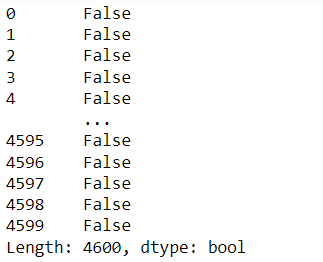
df.describe()



**9. Identifying Duplicate Rows**

Using df.duplicated(), we check for duplicated rows in the dataset. Duplicate entries, if present, can skew the analysis, and identifying and handling them is an important data cleaning step.

df.duplicated()

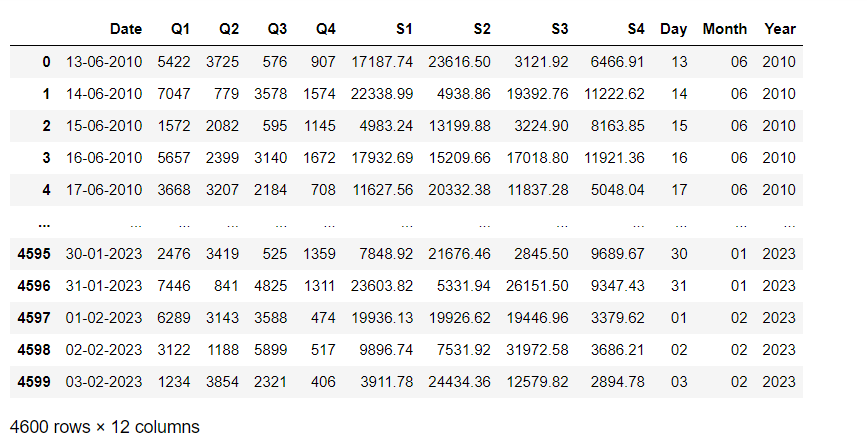


There is no duplicate values.

**10. Renaming Columns**

Finally, we update the column names of the DataFrame for clarity and consistency. The columns are renamed to reflect the information they hold.

df.columns = ['Date', 'Q1', 'Q2', 'Q3', 'Q4', 'S1', 'S2', 'S3','S4','Day','Month','Year']



**11.Saving Preprocessed Data**

This part of the code saves the preprocessed data to a new CSV file named 'Preprocessed\_Data.csv' without including the index.

df.to\_csv('Preprocessed\_Data.csv', index=False)

Code Link: The above code is provided as DAC\_Phase3\_Code.ipynb in this github link

<https://github.com/REEMLATHA/ProductSalesAnalysis-Project>

**IBM Cognos Analytics:**

IBM Cognos Analytics is a business intelligence and analytics platform developed by IBM. It empowers organizations to transform data into actionable insights by providing tools for data visualization, reporting, dashboards, and advanced analytics. With its user-friendly interface and support for data integration from various sources, Cognos Analytics enables users to create interactive visualizations, generate custom reports, and design dynamic dashboards. Its incorporation of AI and machine learning features enhances data analysis, while the platform's security, governance, and collaboration capabilities make it a comprehensive solution for data-driven decision-making across industries, fostering efficiency and informed strategic choices.

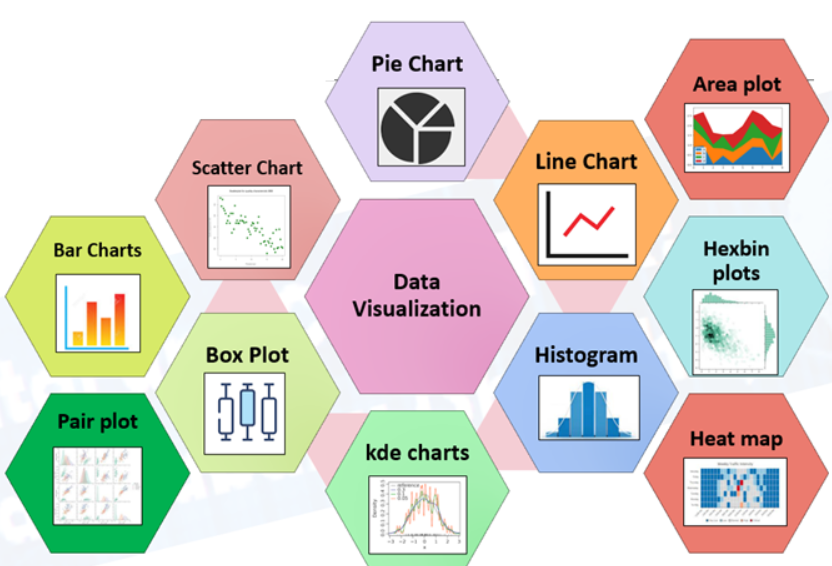
**Loading the Data into IBM Cognos**

Once our data is clean and processed, launch IBM cognos analytics on cloud trial version and load our dataset into IBM Cognos using **Upload data** option.

**Building Visualizations**

Once the data is loaded into Cognos, we can start building visualizations to help analyze our sales data. Cognos offers a variety of visualization tools, such as charts, graphs, and maps.

**Visualization Strategies:**

****

Cognos supports a variety of chart types, including bar charts, line charts, pie charts, scatter plots, heat maps, and more. We can select the chart type that best suits their data and objectives.We will use IBM Cognos to create interactive dashboards to visualize the insights extracted from the data.

The following are some examples of data visualizations that we may create:

* Line charts showing sales trends over time
* Coulumn charts showing the top selling products

**i.Identify trends in sales of all four products during certain month**

To identify trends in the sales of all four products during certain months, Line charts or area charts are commonly used for this type of analysis.

**1.Create a New Report:**

Log in to IBM Cognos Analytics.Create a new report to visualize the sales trends.

**2.Select Data Source:**

In the report, select the data source that contains your sales data.

**3. Choose Visualization Type:**

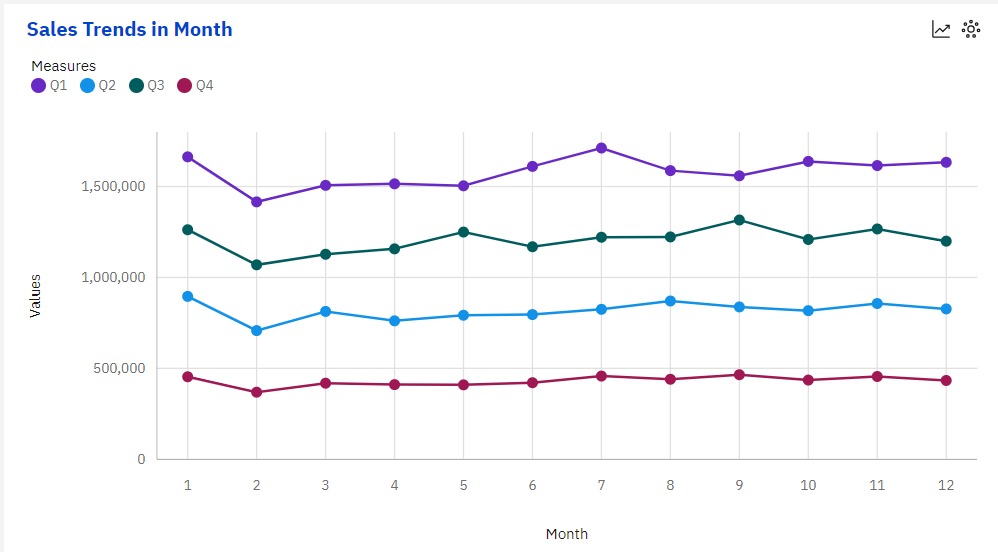
Select the appropriate visualization type to represent sales trends. Line charts or area charts are commonly used for this type of analysis.

**4. Add Data to the Chart:**

Drag and drop the "Month" column to the x-axis .Drag and drop the columns for the sales of each product (e.g., "Q1," "Q2," "Q3," and "Q4") to the y-axis .

**5.Analyze the Trends:**

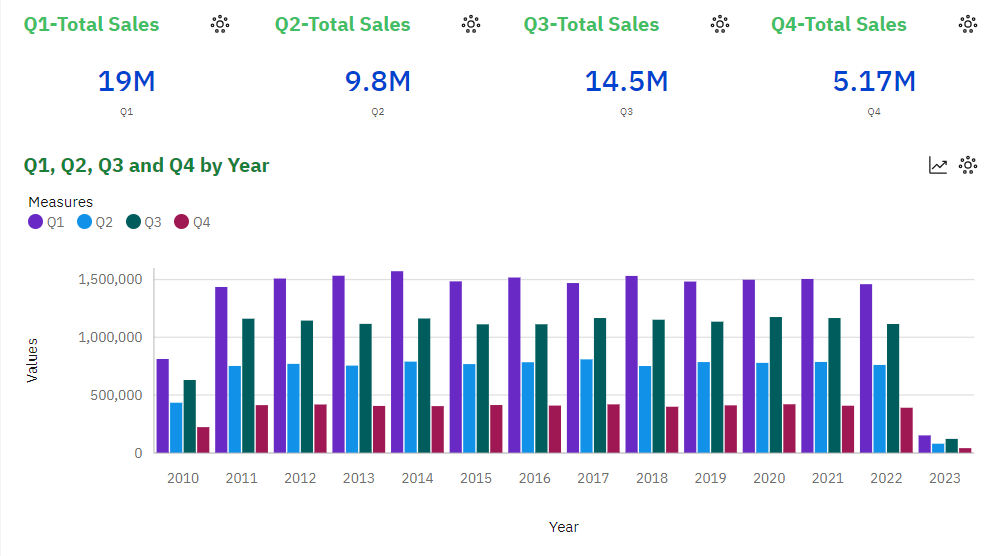
Review the chart to identify trends. Look for patterns, spikes, or dips in sales for each product during specific months.



* We can observe that all products drop in month 2 (Feb)
* We can observe that all products rise in month 1(Jan)

**ii. To identify the top selling product out of all four products**

We can use column graph to find this**.**

****

Q1-UNIT SALES OF P1

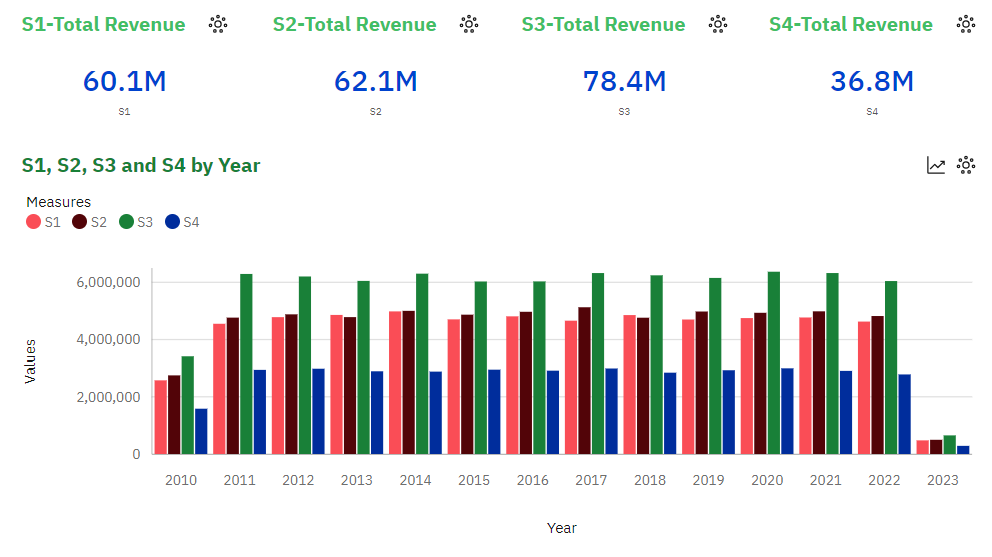
Q2-UNIT SALES OF P2

Q3-UNIT SALES OF P3

Q4-UNIT SALES OF P4

**Insights:**

* + We can observe that P1 has the highest unit sales for each year. And it's highest is in year 2014.
  + We can observe that P4 has the lowest unit sales of all the products.

****

S1-REVENUE GENERATED BY P1

S2-REVENUE GENERATED BY P2

S3-REVENUE GENERATED BY P3

S4-REVENUE GENERATED BY P4

**Insights:**

* We can observe that P3 brought in the most revenue. This could be as a result of multiple things:

-P3 was sold for higher than the rest, as it had the second highest unit sales for each year.

* We can observe than P1 and P2 brought in similar revenues for each year. With P2 bringing in slightly more.

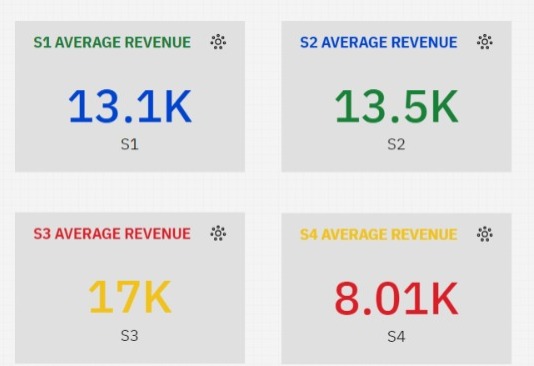
-P1 despite having the most unit sold, brought in the second lowest revenue each year.

**iii.Suggest whether dropping the production of any one of the products would result in a massive setback for the company**

We can use summary chart or KPI chart to find the above



**P4 has the least unit sales**



**P4 has generated the least revenue.**

**Suggestion:**

Dropping the production of any product, especially when it's one of the existing product lines, is a significant business decision that should be carefully considered. In our analysis, we've identified that Product P4 has the least unit sales and has generated the least revenue. Here's an elaboration on this decision and how to mitigate the potential setback:

**Analysis of Dropping Product P4:**

1. **Impact on Revenue**:Dropping Product P4 may indeed result in a decrease in overall company revenue. This is because it's contributing the least to the revenue. However, it's essential to understand the extent of the impact. A financial analysis should be conducted to quantify the revenue loss.

2. **Customer Segmentation**:Analyze the customer segments that purchase Product P4. Are there any loyal customers who rely on this product? Dropping it may result in customer churn if there is a customer base that specifically prefers this product.

3. **Market Demand**:Assess whether there's a continuing market demand for a product similar to P4. If there's no demand, dropping it may not be as significant. However, if there is demand, consider the potential loss of market share.

4. **Profit Margin**: Evaluate the profit margin for Product P4. If it's low or negative, the impact on profitability may not be as severe. On the other hand, if it's a high-margin product, dropping it may hurt profitability.

**Mitigation Strategies:**

1. **Increase Production of Highest Selling Product (P1):** To compensate for the potential setback, you can increase the production and marketing efforts for the highest-selling product, which is presumably generating the most revenue. This can help offset the revenue loss from dropping Product P4.

2. **Marketing Campaign**:Conduct a targeted marketing campaign to promote the highest-revenue-generating product (e.g., P1). This campaign should focus on retaining existing customers who were purchasing Product P4 and attracting new customers.

3.**Diversification**: Consider diversifying your product portfolio by introducing new products that align with market demand. This can help offset the revenue loss and reduce the dependence on any single product.

4. **Customer Retention**: Implement customer retention strategies to retain existing customers. Offer incentives, discounts, or promotions to encourage them to continue buying from the company even if Product P4 is discontinued.

5. **Market Research**:Conduct market research to understand the preferences and needs of your target audience. This can help in product development and marketing strategies.

6. **Cost Reduction**:Evaluate cost reduction strategies to minimize the impact of dropping Product P4 on the company's cost structure.

**Conclusion:**

In conclusion, the analysis of product sales data, performed using IBM Cognos Analytics, has provided valuable insights. The identification of top-selling products, sales trend of products over months,impact assessment of dropping a product, and the strategic recommendation to mitigate setbacks by promoting the highest-revenue-generating product demonstrate informed decision-making. The company's commitment to customer retention, market diversification, and cost management will play a crucial role in ensuring continued growth and financial stability. This analysis equips the company with actionable strategies to navigate potential challenges and foster sustained success in its competitive landscape.In phase 4,we will predict the future sales and revenue in 2024 and to get an estimate on no of units of each product that could be sold on 31st of Dec , every year , if all their retail centers were kept open.