**PRODUCT SALES ANALYSIS PROJECT-PHASE 5**

**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

Table of Contents

1.Introduction

* Overview of the Project
* Purpose and Objective
* Objective
  + Project Objective
  + Objective of Each Phase

2.Problem Definition and Design Thinking Process

* Problem Definition
* Design Thinking Process
  + - Empathize
    - Define
    - Ideate
    - Prototype
    - Implement

3.Development Phases

* Phase 1: Problem Definition and Design Thinking
* Phase 2: Innovation (Incorporating Machine Learning Techniques)
* Phase 3: Development Part 1 (Data Preprocessing)
* Phase 4: Development Part 2 (Deriving Actionable Insights)

4.Analysis Objective

5.Data Collection

* Data Source
* Dataset Link
* Data Description
* Columns Being Used

6.Innovation-Machine Learning Approach

* Importing Libraries and Loading the Data
* Column Cleanup
* Data Aggregation
* Exploratory Data Analysis (EDA)
* Splitting Training and Testing Data
* Feature Engineering
* Model Training
* Predicting Sales and Revenue
* Visualisation of Prediction
* Project Link

7.Development Part-1 Preprocessing Dataset and Loading into IBM Cognos Analytics

* Data Preprocessing
* Importing Libraries
* Libraries Used
* Loading the Data
* Custom Date Parsing Function
* Applying the Custom Date Parsing Function
* Extracting Year, Month, and Day
* Deriving Weekday
* Calculating Total Sales and Total Revenue
* Renaming Columns
* Reordering the Columns
* Saving Preprocessed Data
* Printing the Updated DataFrame
* Replication Instructions
* IBM Cognos Analytics
* Introduction to IBM Cognos Analytics
* Loading the Data into IBM Cognos
* Building Visualizations

8.Development Part-II Building Visualisations using IBM Cognos and Deriving Actionable Insights

* Visualisation Strategies
* Top Selling Product
* Column Chart Showing Yearly Sales Trends
* Column Chart Showing Yearly Revenue Trends
* Actionable Insights
* Sales Trends and Peak Sales Period
* Line Chart Showing Monthly Sales Trends
* Column Chart Showing Sales Trends Over Years
* Area Chart Showing Sales Trends in Weekdays
* Observation and Actionable Insights
* Dashboard
* Customer Preferences
* Seasonal Trend and Frequency
* Product Specific Preferences
* Actionable Insights based on Customer Preferences
* Estimating Sales on December 31st of Each Year
* Methods for Estimation
* Recommendation
* Cross Tab Chart
* Actionable Insights
* Predicting Future Sales and Revenue for 2024
* Observations for Total Sales Trends
* Observations for Total Revenue Trends
* Actionable Insights for Total Sales Trends

9.Advantages

10.Disadvantages

* Challenges and Limitations

11.Applications

* Practical Use Cases

12.Future Scope

* Potential Improvements and Expansions

13.Conclusion

14.Appendix

* Data Analysis Code

**PRODUCT SALES ANALYSIS**

**Phase 5: Project Documentation**

**Introduction**

This document serves as the comprehensive documentation for the Product Sales Analysis project, encompassing all the phases from initial problem definition to the final insights and recommendations. The project aims to analyze a comprehensive sales and revenue dataset, identify top-selling products, understand sales trends, and uncover customer preferences. Insights derived from the analysis are used to guide inventory management and marketing strategies.

**Overview**

The project is divided into several phases that collectively address the business problem and offer actionable solutions. These phases include:

**Phase 1**:

**Problem Definition and Design Thinking**: Identifying the project's objectives, problem definition, and design thinking process.

**Phase 2:**

**Innovation (Incorporating Machine Learning Techniques)**: Developing machine learning models to enhance the analysis and predict future sales.

**Phase 3:**

**Development Part 1 (Data Cleaning and Preprocessing**):

Preparing the dataset by cleaning and preprocessing it for analysis.

**Phase 4:**

**Development Part 2 (Data Visualization using IBM Cognos and Deriving Actionable Insights):** Creating interactive dashboards and reports using IBM Cognos to visualize the data and derive insights.

**Phase 5:**

**Documentation (This Phase):** Documenting the project's objectives, design thinking process, and development phases. Explaining how insights guide inventory management and marketing strategies.

**Purpose**

The purpose of this documentation is to provide a comprehensive record of the project's objectives, analysis, and actionable insights. It serves as a reference for stakeholders and a guide for replication, further analysis, and decision-making.

**Objective**

The project's main objective is to analyze a sales and revenue dataset to:

* Identify top-selling products among four products (P1, P2, P3, P4).
* Understand sales trends, including monthly and yearly variations.
* Uncover customer preferences, seasonal patterns, and product-specific trends.
* Derive actionable insights that can guide inventory management and marketing strategies.

The specific objectives for each phase are as follows:

* **Phase 2: Innovation:** Develop machine learning models to predict future sales and revenue.
* **Phase 3: Development Part 1**: Prepare the dataset for analysis by cleaning and preprocessing.
* **Phase 4: Development Part 2**: Create interactive dashboards and reports to visualize sales trends and customer preferences. Derive actionable insights for strategic decision-making.

**I.Problem Definition and Design Thinking**

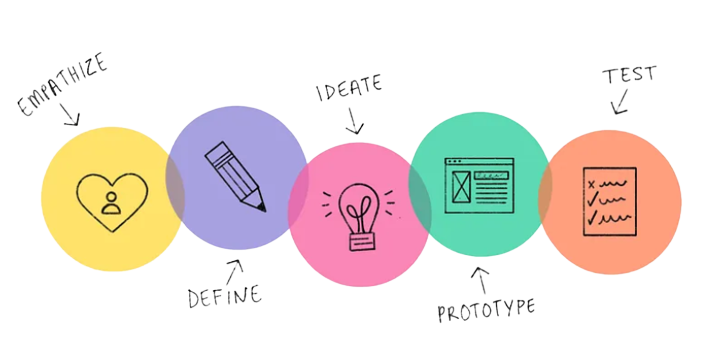
**Problem Definition**

A small-scaled business venture in India that has been selling four products for over ten years. They have collected sales data from their retail centers and organized it into a CSV file. They need help analyzing this data to answer the following questions:

1. Is there any trend in the sales of all four products during certain months?
2. Out of all four products, which product has seen the highest sales in all the given years?
3. The company has all its retail centers closed on the 31st of December every year. Mr: Hariharan, the CEO, would love to get an estimate on the number of units of each product that could be sold on the 31st of December, every year, if all their retail centers were kept open.
4. The CEO is considering an idea to drop the production of any one of the products. He wants you to analyze the data and suggest whether his idea would result in a massive setback for the company.
5. The CEO would also like to predict the sales and revenues for the year 2024. He wants you to give a yearly estimate with the best possible accuracy.

**Design Thinking Process**

In the context of solving companys's business problems, we will utilize a comprehensive design thinking approach that encompasses five key stages, from empathizing with company to implementing the most effective solution:

****

**1.Empathize:** We will first empathize with REC corp and their customers to understand their needs and challenges. This will involve reviewing the sales data and conducting other research. Understand the business problem and the need for data-driven decision-making.

**2. Define:** Once we have a good understanding of the problem, we will define the specific insights we want to extract from the data**.** Clearly define the project objectives and problem statement.

**3. Ideate:** Next, we will brainstorm potential solutions to the problem. This will involve generating as many ideas as possible, without worrying about feasibility. Explore innovative solutions, including machine learning techniques.

**4. Prototype:** Once we have a list of potential solutions, we will prototype and test them to see which ones are most effective. This may involve creating mockups, simulations, or other prototypes. Develop data preprocessing and visualization techniques to test potential solutions.

**5.Implement** :Finally, we will implement the best solution and monitor its effectiveness. This may involve working with REC corp to develop and deploy new processes, tools, or technologies

This design thinking process integrates the five essential stages of empathizing, defining, ideating, prototyping, and implementing into a cohesive approach for tackling REC Corp's business challenges. It emphasizes understanding the problem, defining objectives, fostering creativity, testing potential solutions, and ultimately implementing the most effective strategy to drive REC Corp towards data-driven success.

**Development Phases**

Phase 2: Innovation (Incorporating Machine Learning Techniques)

In this phase, machine learning models are developed to enhance the analysis by providing predictive insights.

Phase 3: Development Part 1 (Data Cleaning and Preprocessing)

This phase focuses on preparing the dataset for analysis. It includes data cleaning, handling missing values, and data preprocessing to ensure data quality.

Phase 4: Development Part 2 (Data Visualization using IBM Cognos and Deriving Actionable Insights)

This phase utilizes IBM Cognos to create interactive dashboards and reports. The primary objectives are to visualize sales trends, product preferences, and derive actionable insights.

**Analysis Objectives**

The specific analysis objectives for this project are to:

* Identify trends in sales of all four products during certain months
* Identify the top selling product out of all four products
* Estimate the number of units of each product that could be sold on the 31st of December, every year, if all retail centers were kept open
* Suggest whether dropping the production of any one of the products would result in a massive setback for the company
* Predict the sales and revenues for the year 2024

**Data Collection:**

**Data Source:**

The data set is collected from Kaggle which is a leading collaborative data analytics 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

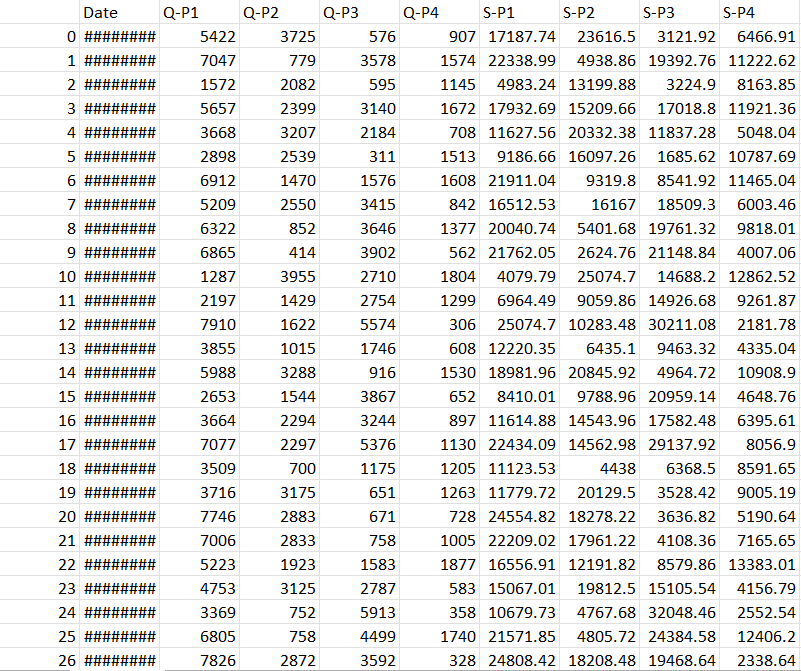
**Data Description:**

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:**

**II.Innovation**

**Machine Learning Approach**

**Machine Learning Approach:**

In our product sales analysis project, we harnessed the SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous variables) model, a recognized technique for time series forecasting to predict sales in 2024.SARIMAX's strength lies in its ability to capture temporal dependencies and seasonality, making it an excellent fit for Company's sales and revenue data analysis. This model excels in handling seasonality and trends through autoregressive and moving average components, as well as exogenous variables. By employing SARIMAX, we empower Company to make informed decisions, allocate resources effectively, and plan strategically, leveraging machine learning for competitive advantage and adaptability in dynamic markets.

**Importing Libraries and Loading the Data**:We loaded the raw data from the CSV file into a pandas DataFrame. To facilitate date-based analysis, we converted the 'Date' column to a datetime format, which enables us to work with dates seamlessly.

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

df['Date'] = pd.to\_datetime(df['Date'], format='%d-%m-%Y', errors="coerce")

df

**Column Cleanup**: To simplify the DataFrame and enhance readability, we removed the 'Unnamed: 0' column as it didn't offer meaningful information. The remaining columns were renamed to better represent their content.

df.drop("Unnamed: 0", axis=1, inplace=True)

**Data Aggregation**: We aggregated the data from daily samples to yearly samples, offering a more comprehensive view of the annual sales and revenue figures. Aggregating data in this manner simplifies time series analysis and visualization.

df\_yearly = df.resample('Y').sum()

df\_yearly

**Splitting Training and Testing Data**

It's essential to split our dataset into two main parts: the training dataset and the testing dataset. This division helps us to assess the performance of your model and ensure that it can make accurate predictions on unseen data

train = df\_yearly.loc['2010-12-31':'2022-12-31', :]

test = df\_yearly.loc['2023-12-31':'2023-12-31', :]

**Feature Engineering:**

Feature engineering is a key step in time series forecasting, as it involves creating relevant features that can enhance the predictive model's accuracy. In our current analysis, we used a basic SARIMAX model to predict product P2's sales for the year 2024. Feature engineering could be extended to include factors such as seasonality, lags, or external variables that may influence product sales.

product\_column = 'Q2' ## This will the column we are predicting

**Model Training:**

Model training is a pivotal step in forecasting. For this project, we trained a SARIMAX **(Seasonal AutoRegressive Integrated Moving Average with eXogenous factors)** model. The model parameters, including the order and seasonal\_order, were selected based on a rigorous analysis of the data.

model = SARIMAX(train[product\_column], order=order,seasonal\_order=seasonal\_order, freq="Y")

results = model.fit()

**Predicting Sales and Revenue:**

Using the trained SARIMAX model, we made predictions for product P2's sales in the year 2024. The forecasted sales value was approximately 722,323 units for product Q2. Such predictions are invaluable for that company as they provide insights into production planning, inventory management, and revenue forecasting.

forecast = results.get\_forecast(steps=3, dynamic=False)

forecast\_values = forecast.predicted\_mean

print(f"PredictedValuesfor{forecast\_values.index.year[1]}:{forecast\_values.values[1]}")

**Output:**

Predicted Values for 2024 : 722323.3350284399

**Visualisation of Prediction:**

vis\_data = df\_yearly.loc[:"2022", ["Q2"]]

vis\_data = vis\_data.reset\_index()

vis\_data

new\_df = pd.DataFrame({"Date": forecast\_values.index.values[0:2], "Q2": forecast\_values.values[0:2]})

new\_df

sns.lineplot(data=final\_df.loc[:"2022", :], x=final\_df.loc[:"2022",:].index.values, y="Q2", label="2010 - 2022")

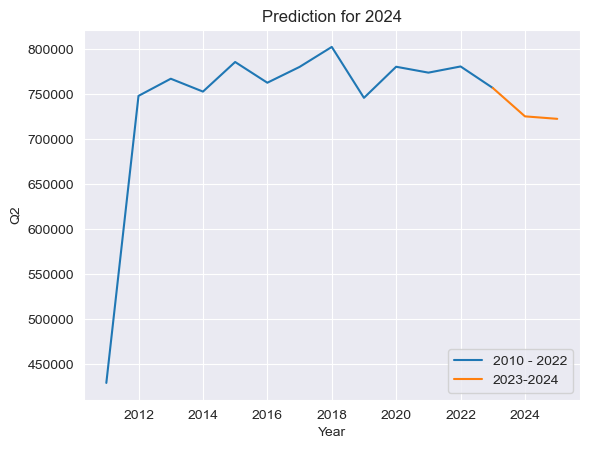
sns.lineplot(data=final\_df.loc["2022":, :], x=final\_df.loc["2022":,:].index.values, y="Q2", label="2023-2024")

plt.title("Prediction for 2024")

plt.xlabel("Year")

plt.ylabel("Q2")

plt.show()

****

**Replication Instructions**

* To replicate the analysis and generate visualizations , follow these steps:
* Clone the GitHub repository: git clone https://github.com/REEMLATHA/ProductSalesAnalysis-Project
* Access the Jupyter Notebook file: DAC\_Phase2\_Code.ipynb
* Execute the cells in the notebook to replicate the analysis.

**III.Development Part-I**

**Preprocessing Dataset and Loading in IBM Cognos for Analysis**

**Data Preprocessing**

In Phase 3, data cleaning and preprocessing techniques were applied to ensure data quality and consistency. This phase included addressing missing values.

**Importing Libraries:**

We begin by importing the necessary Python libraries, primarily `pandas` for data manipulation and `dateutil` for date parsing.

**import pandas as pd**

**from dateutil import parser**

**Libraries Used:**

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

- **dateutil**:A library for parsing dates in various formats, which is used to parse the 'Date' column.

**Loading the Data:**

The raw data is loaded from a CSV file named 'statsfinal.csv' into a pandas DataFrame.

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

**Custom Date Parsing Function:**

A custom date parsing function is defined to handle the 'Date' column, which may contain dates in various formats. This function attempts to parse the date, and if unsuccessful, it returns `None`.

**def parse\_date(date\_str):**

**try:**

**return parser.parse(date\_str)**

**except:**

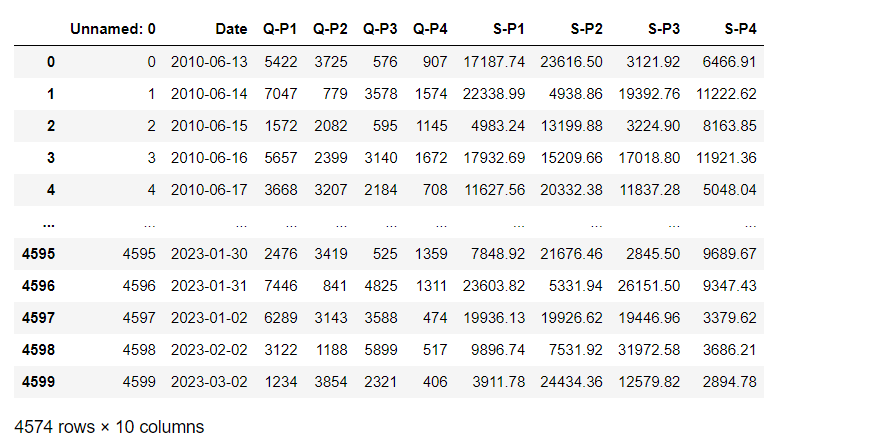
**return None**

**Applying the Custom Date Parsing Function:**

The custom date parsing function is applied to the 'Date' column, converting the dates into a consistent datetime format. Rows with invalid dates are also handled by dropping them from the DataFrame.

**df['Date'] = df['Date'].apply(parse\_date)**

**df = df.dropna(subset=['Date'])**

****

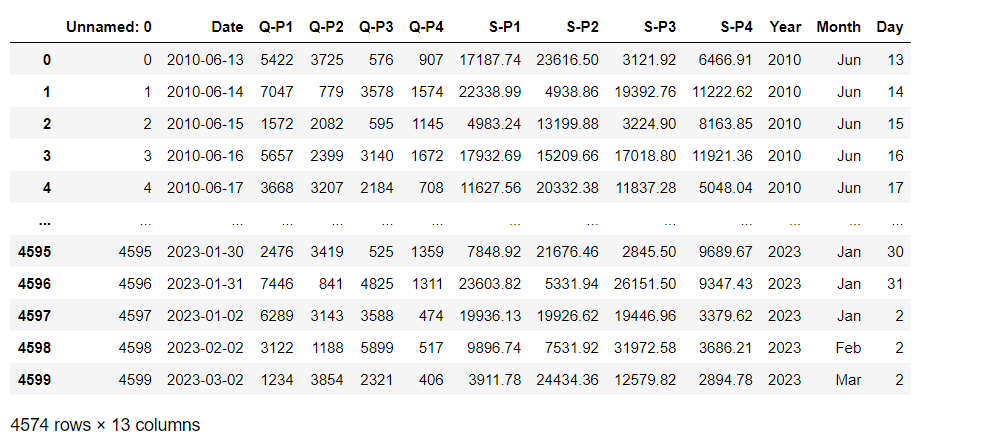
**Extracting Year, Month, and Day:**

Year, month, and day are extracted from the parsed dates to provide more detailed temporal information for analysis. The 'Year' and 'Day' columns are directly extracted, while the 'Month' column is generated with abbreviated month names (e.g., JAN, FEB).

**df['Year'] = df['Date'].dt.year**

**df['Month'] = df['Date'].dt.strftime('%b') # Abbreviated month names**

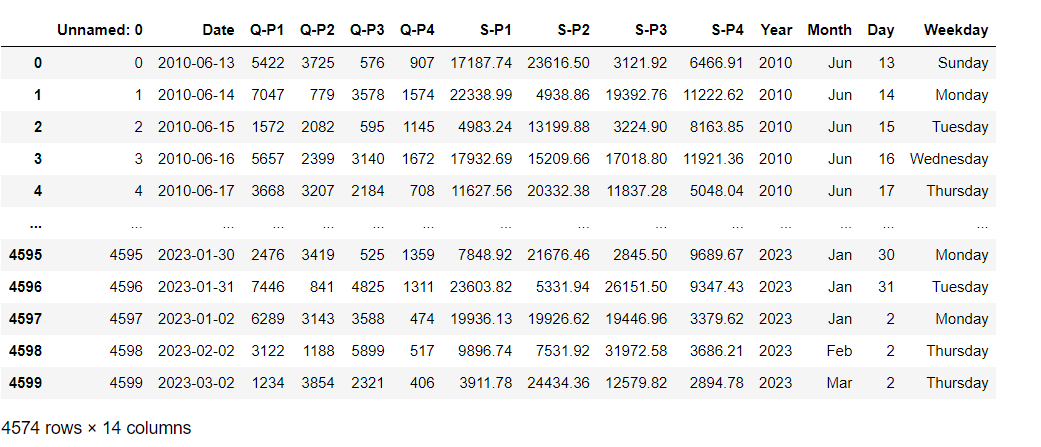
**df['Day'] = df['Date'].dt.day**

****

**Deriving Weekday:**

The weekday information is derived from the 'Date' column, and it is added as a new column 'Weekday.' This provides information about the day of the week for each date**.**

**df['Weekday'] = df['Date'].dt.day\_name()**

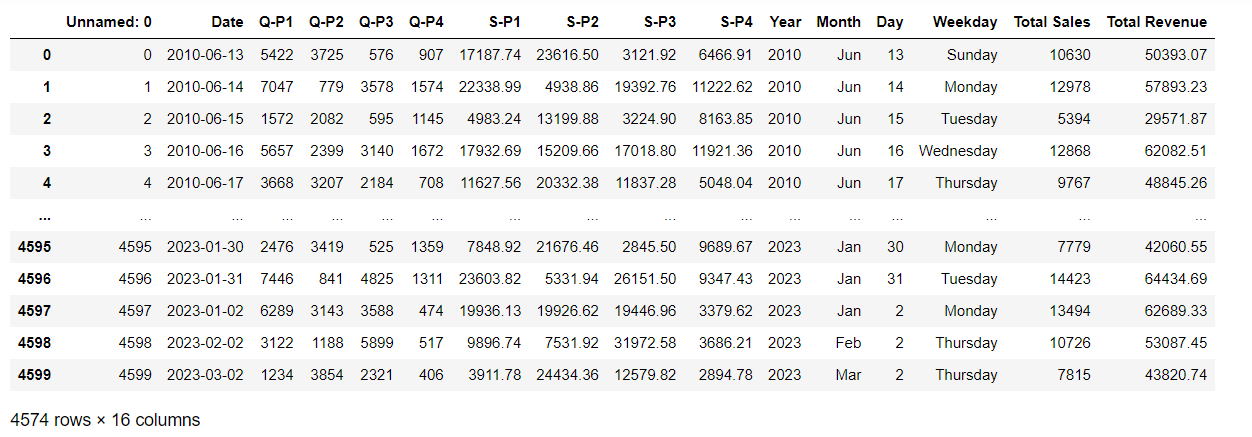
****

**Calculating Total Sales and Total Revenue:**

Total sales and total revenue columns are calculated by summing the corresponding columns for each quarter. This provides a consolidated view of sales and revenue for each entry in the dataset**.**

**df['Total Sales'] = df['Q-P1'] + df['Q-P2'] + df['Q-P3'] + df['Q-P4']**

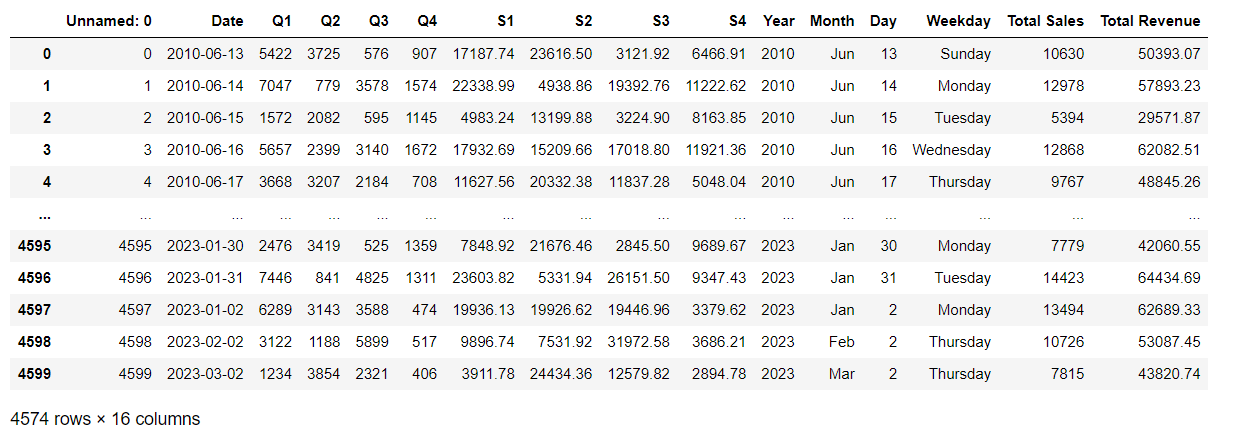
**df['Total Revenue'] = df['S-P1'] + df['S-P2'] + df['S-P3'] + df['S-P4']**

****

**Renaming Columns:**

Column names are renamed to provide concise and consistent names. The 'Q-P1' to 'Q-P4' columns are renamed as 'Q1' to 'Q4,' and the 'S-P1' to 'S-P4' columns are renamed as 'S1' to 'S4.'

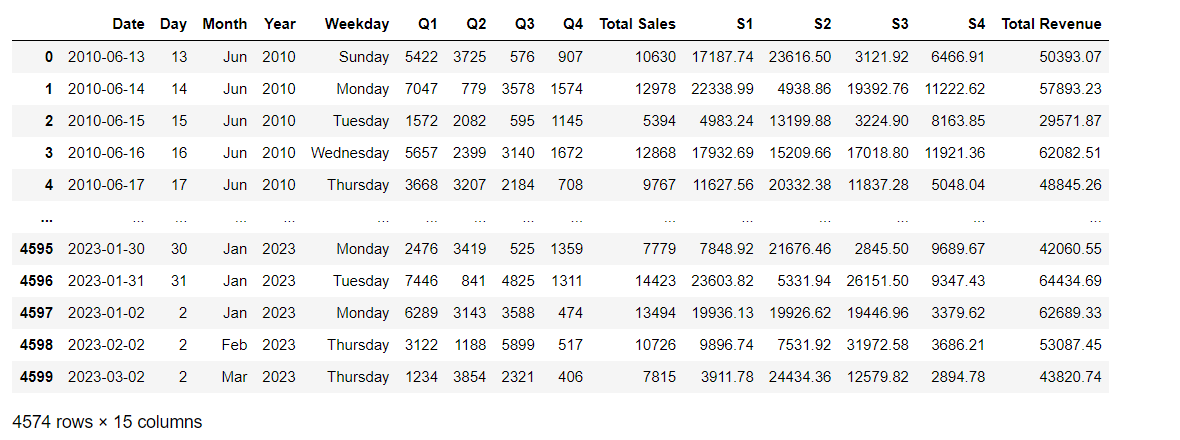
**df = df.rename(columns={ 'Q-P1': 'Q1', 'Q-P2': 'Q2', 'Q-P3': 'Q3', 'Q-P4': 'Q4','S-P1': 'S1','S-P2': 'S2','S-P3': 'S3','S-P4': 'S4'})**

****

**Reordering the Columns:**

The columns in the DataFrame are reordered to follow a specific order: 'Date,' 'Day,' 'Month,' 'Year,' 'Weekday,' 'Q1' to 'Q4,' 'Total Sales,' 'S1' to 'S4,' and 'Total Revenue.'

**df = df[['Date', 'Day', 'Month', 'Year', 'Weekday', 'Q1', 'Q2', 'Q3', 'Q4', 'Total Sales', 'S1', 'S2', 'S3', 'S4', 'Total Revenue']]**

****

**Saving Preprocessed Data:**

The preprocessed data is saved to a new CSV file named 'Preprocessed\_Data.csv' without including the index**.**

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

**Printing the Updated DataFrame:**

The code concludes by printing the updated DataFrame, which now includes the parsed date, derived date components, total sales, total revenue, and renamed columns.

**print(df)**

This code demonstrates a comprehensive approach to loading, cleaning, and preprocessing a dataset, preparing it for further analysis or modeling, while also enhancing its readability and interpretability**.**

**Replication Instructions**

To replicate the analysis and generate visualizations using IBM Cognos, follow these steps:

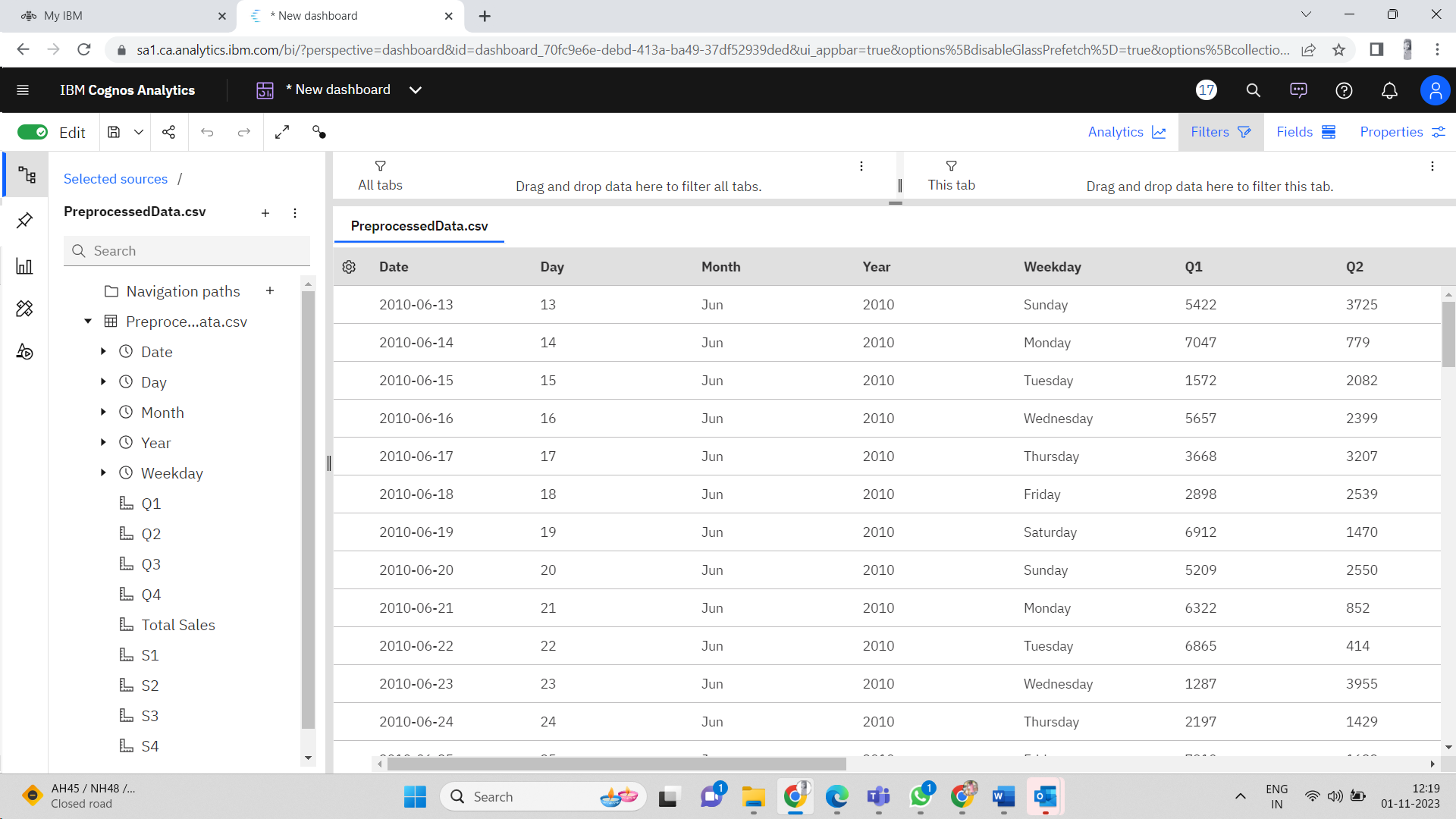
* Clone the GitHub repository: git clone <https://github.com/REEMLATHA/ProductSalesAnalysis-Project>
* Access the Jupyter Notebook file: DAC\_Phase3\_PreprocessingCode.ipynb
* Execute the cells in the notebook to replicate the analysis.
* To generate visualizations using IBM Cognos, refer to the detailed instructions provided in Phase 4 of the documentation**.**

**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

**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 **Preprocessed\_Data.csv** into IBM Cognos using **Upload data and Start Creating Content**  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.

**IV. Development Part-II**

**Data Visualization using IBM Cognos**

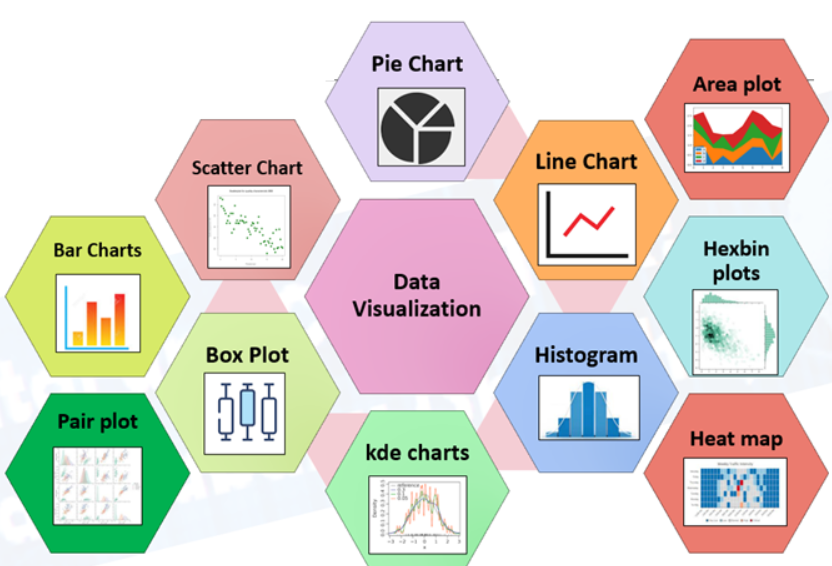
**Data Visualization using IBM Cognos**

Phase 4 focused on creating interactive dashboards and reports using IBM Cognos. Visualizations were designed to identify sales trends, peak sales periods, and customer preferences. Line charts, column charts, and area charts were used to present the data effectively.

**Derived Actionable Insights**

Based on the analysis conducted in Phase 4, actionable insights were derived. These insights include product focus, revenue optimization, pricing strategies, profit margins, and strategies for low-performing products.

**Visualization Strategies:**

****

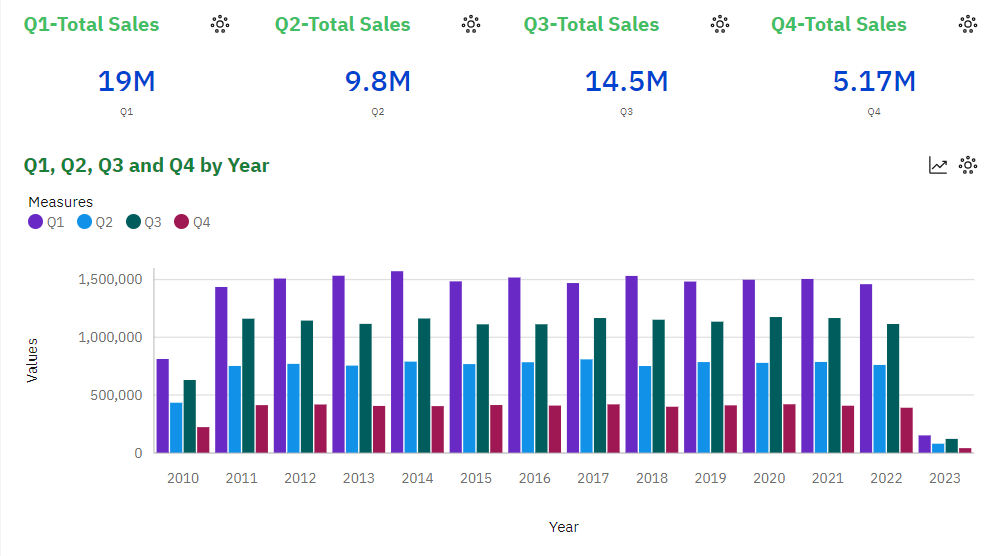
Cognos supports a variety of chart types, including bar charts, line charts, pie charts, scatter plots, heat maps, and more. We can select the chart type that best suits their data and objectives.We will use IBM Cognos to create interactive dashboards to visualize the insights extracted from the data.

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

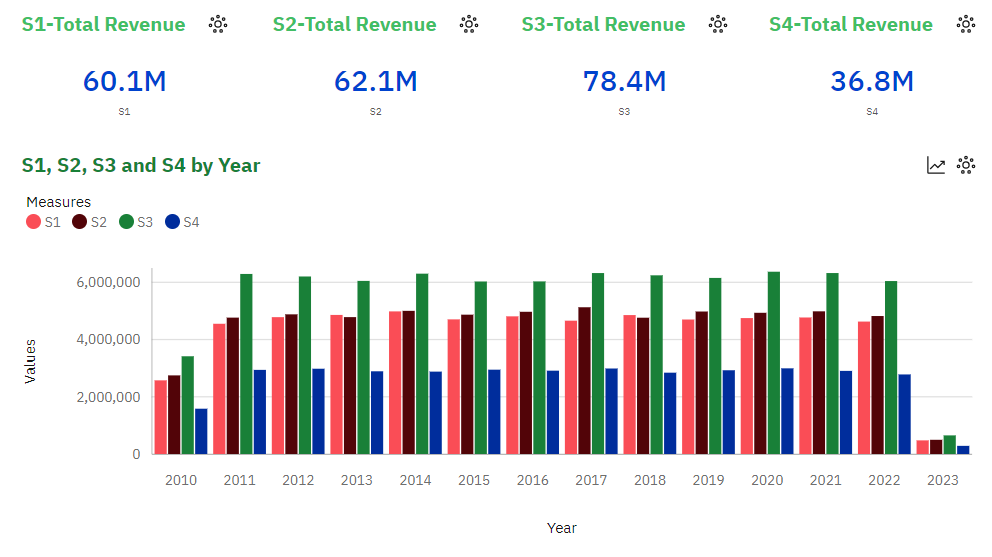
* Line charts showing sales trends over time
* Coulumn charts showing the top selling products

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

We used summary charts to display total unit sales and revenue for products P1, P2, P3, and P4. Additionally, column charts were created to visualize the yearly trends in total sales and revenue. This approach provides a clear and concise overview of product performance and annual financial trends .

**Observation:**

* + 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.

****

**Observation:**

* 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.

**Actionable Insights:**

Based on the observation, here are some actionable insights that can be derived from the data:

**1. Product Focus:**

- Given that P1 consistently has the highest unit sales, it's crucial to focus on maximizing the potential revenue from this product. This might involve pricing strategies, bundling with other products, or marketing campaigns to maintain and potentially increase its revenue contribution.

**2. Revenue Optimization for P3:**

- While P3 generates the most revenue, it's essential to understand why. If it's due to higher prices, consider whether there's room to adjust pricing without sacrificing sales volume. Additionally, explore opportunities to further increase unit sales for P3.

**3. P2 Revenue Enhancement:**

- P2 brings in slightly more revenue than P1. Investigate why P2, with lower unit sales, outperforms P1 in revenue. Consider whether this could be replicated for other products or if there are strategies that can boost P2's revenue further.

**4. P4's Performance:**

- P4 consistently has the lowest unit sales and revenue. Evaluate the viability of this product in your product portfolio. It may be worth considering whether to discontinue or redesign it, or explore marketing strategies to increase its appeal to customers.

**5. Pricing Strategy:**

- Assess the pricing strategy for each product. If P3's high revenue is due to higher prices, evaluate whether similar pricing strategies can be applied to other products. Ensure that the pricing strategy aligns with customer demand and market competition.

**6. Profit Margins:**

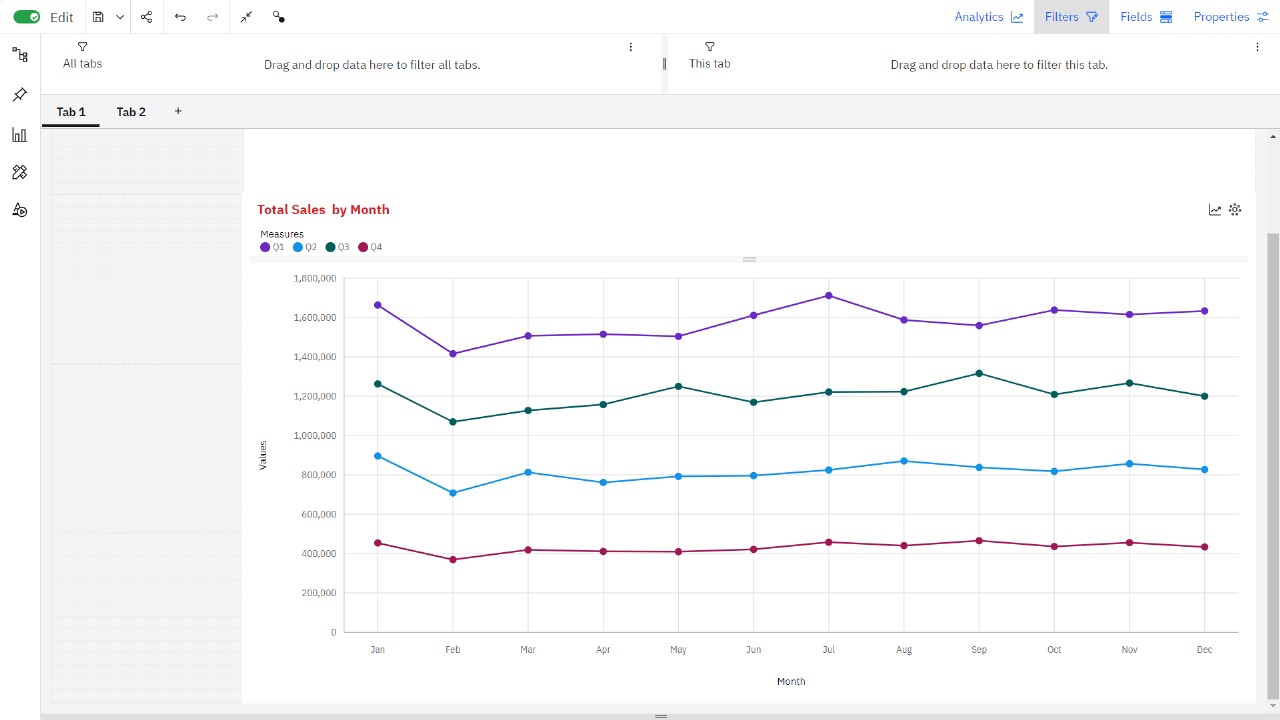
- Calculate and analyze the profit margins for each product (revenue generated minus cost of goods sold). Products with lower unit sales but higher profit margins may represent areas for growth.

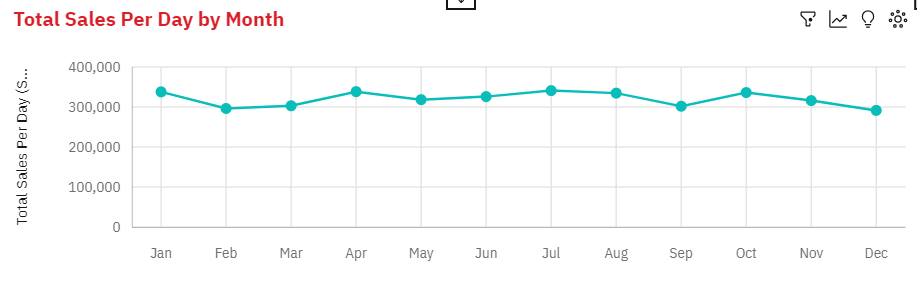
These actionable insights provide a foundation for making informed decisions to enhance product performance, revenue generation, and overall business strategy. It's important to continually monitor and adjust strategies based on evolving market conditions and customer preferences.

**ii.To identify sales trends and peak sales period**

We employed line charts to track sales trends over months and discern peak sales periods. Additionally, column charts illustrated yearly sales for individual products (P1, P2, P3, P4). To analyze weekday patterns, an area chart showcased sales distribution throughout the week. These visualizations offer valuable insights into sales dynamics .

**Line Chart Showing Monthly Sales Trends**

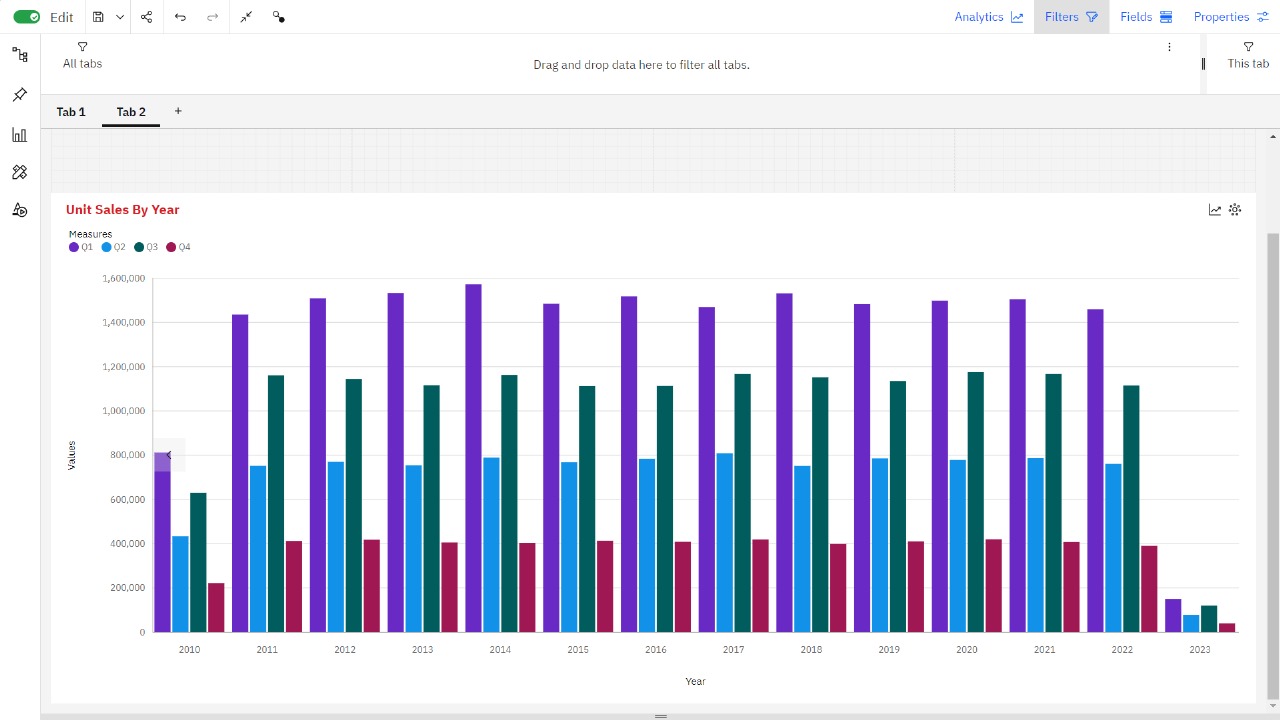


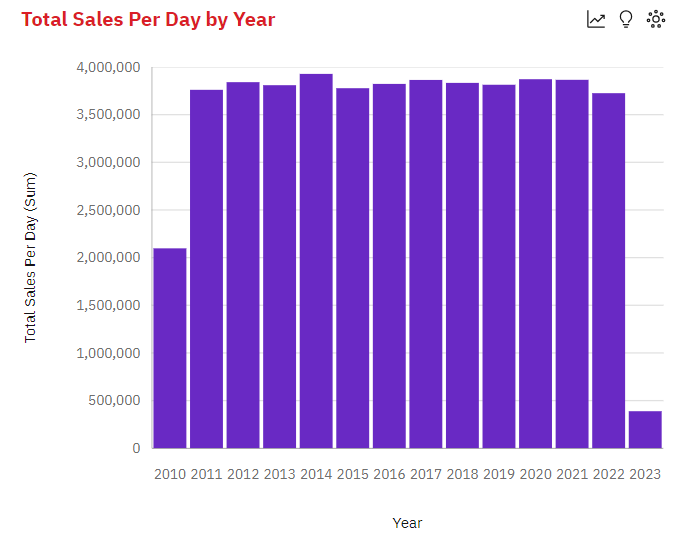
****

**Observation:**

* Q1 has a strong upward trend.
* Q1 ranges from over 1.4 million, in Feb, to over 1.7 million, in Jul.
* Q2 ranges from nearly 708 thousand, in Feb, to almost 896 thousand, in Jan.
* Q3 ranges from almost 1.1 million, in Feb, to over 1.3 million, in Sep.
* Q4 ranges from nearly 369 thousand, in Feb, to over 465 thousand, in Sep.
* Total Sales Per Day ranges from nearly 3.6 million, in Feb, to almost 4.3 million, in Jan.
* All Products drop in February
* All Products Rise in January

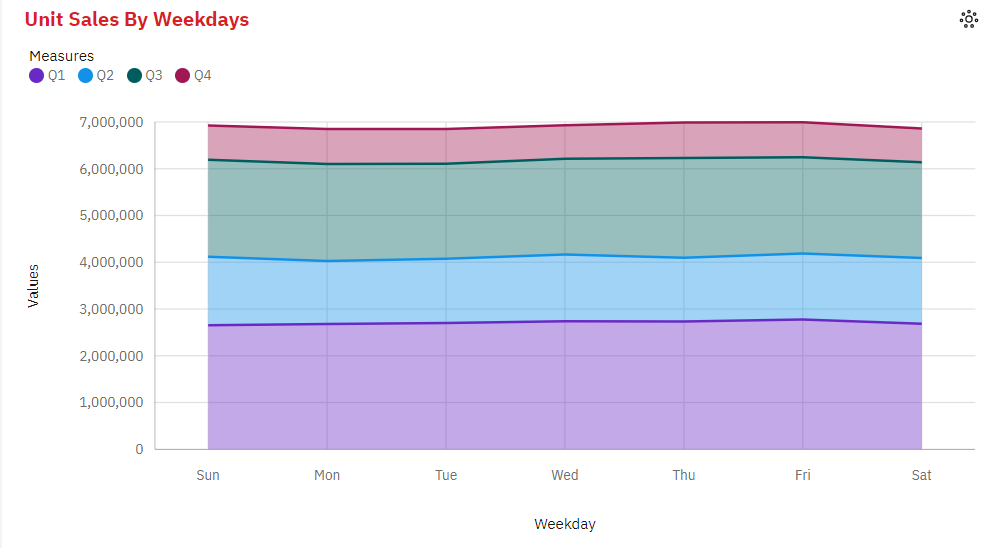
**Column Chart Showing Sales Trends Over Years**

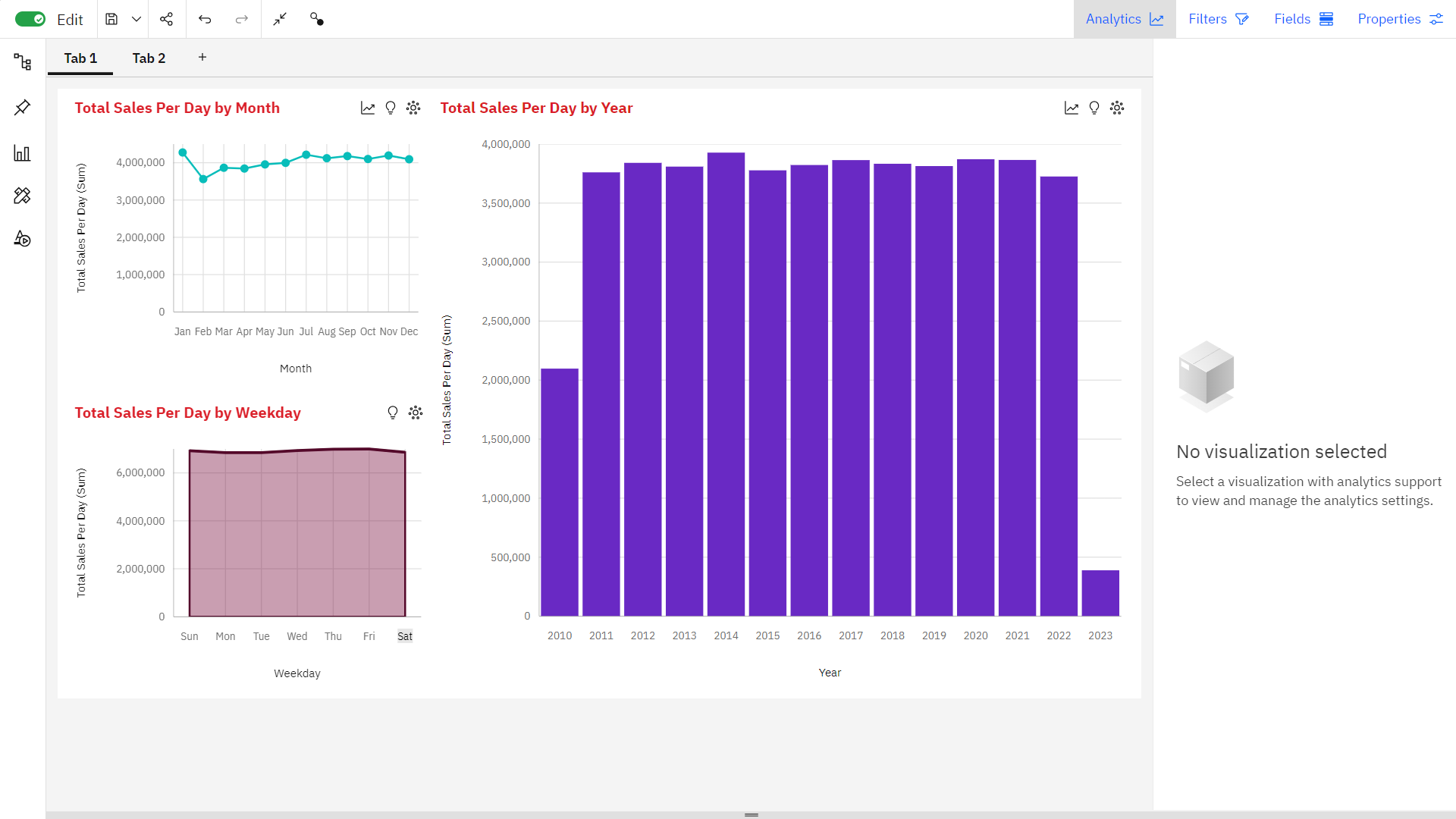


**Observation:**

* Q1 ranges from over 150 thousand, in 2023, to nearly 1.6 million, in 2014.
* Q2 ranges from over 78 thousand, in 2023, to nearly 809 thousand, in 2017.
* Q3 ranges from over 120 thousand, in 2023, to nearly 1.2 million, in 2020.
* Q4 ranges from nearly 40 thousand, in 2023, to almost 420 thousand, in 2020.
* Total Sales Per Day has a weak upward trend.
* Total Sales Per Day is unusually low in 2023 and 2010.
* From 2022 to 2023, Total Sales Per Day dropped by 90%.
* Across all years, the sum of Total Sales Per Day is over 48 million.
* Total Sales Per Day ranges from nearly 389 thousand, in 2023, to over 3.9 million, in 2014.
* For Total Sales Per Day, the most significant values of Year are 2014, 2020, 2021, 2017, and 2012, whose respective Total Sales Per Day values add up to over nineteen million, or 40 % of the total.

**Area Chart Showing Sales Trends In Weekdays**

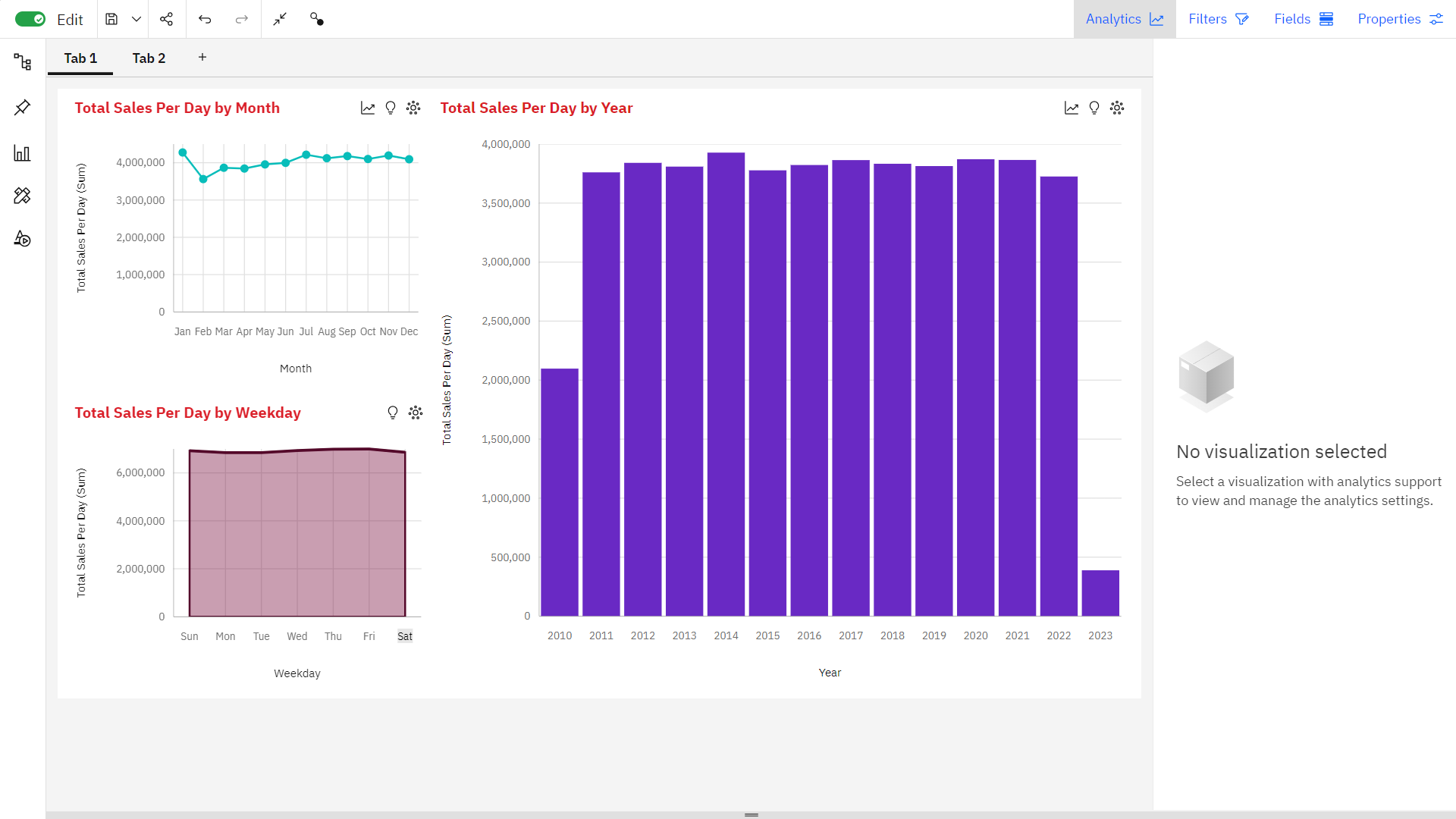
****



**Observation:**

* Q1 ranges from nearly 2.7 million, when Weekday is Sun, to nearly 2.8 million, when Weekday is Fri.
* Q2 ranges from over 1.3 million, when Weekday is Mon, to nearly 1.5 million, when Weekday is Sun.
* Q3 ranges from over 2.0 million, when Weekday is Tue, to over 2.1 million, when Weekday is Thu.
* Q4 ranges from almost 720 thousand, when Weekday is Wed, to over 756 thousand, when Weekday is Thu.
* Across all weekdays, the sum of Total Sales Per Day is over 48 million.
* Total Sales Per Day ranges from over 6.8 million, when Weekday is Mon, to nearly 7.0 million, when Weekday is Fri.
* For Total Sales Per Day, the most significant values of Weekday are Fri, Thu, Wed, and Sun, whose respective Total Sales Per Day values add up to nearly 28 million, or 57.5 % of the total.

**DASHBOARD:**



**Actionable Insights:**

Based on our observations of monthly sales trends, sales trends over the years, and sales trends in weekdays, we can derive some actionable insights for peak sales periods and overall sales strategy:

**Monthly Sales Trends:**

**1. Q1 Peak:**Q1 consistently shows a strong upward trend. The period from February to July appears to be the peak within Q1. Focus your marketing efforts and inventory management to take advantage of this peak.

**2. February Slump**:All products experience a drop in February. Plan for this dip in sales by perhaps running promotions or clearance sales to maintain revenue.

**3. January Surge:** The sales for all products rise in January. Capitalize on this surge by launching new products or special promotions at the beginning of the year.

**Sales Trends over Years:**

**1. Year 2014**:This year had the highest sales in various quarters. Consider studying what worked in 2014 and attempt to replicate successful strategies.

**2. Total Sales Per Day Trends:**While there's a weak upward trend over the years, be cautious of unusually low years, such as 2010 and 2023. Investigate what caused these low sales and work on strategies to prevent such downturns.

**3. 2022-2023 Drop**:The drastic 90% drop in Total Sales Per Day between 2022 and 2023 is alarming. Investigate the reasons behind this decline, whether it's market-related or internal factors, and develop strategies to reverse the trend.

**Sales Trends in Weekdays:**

1**. Weekday Focus**: The weekdays Fri, Thu, Wed, and Sun account for a significant portion of total sales. Concentrate marketing efforts and special promotions on these days to maximize revenue.

**2. Weekday Differences:**The sales figures for each weekday vary significantly. Make sure to have sufficient staff and resources available on days with the highest sales, such as Monday for Q2, and Sunday for Q1.

**Overall Strategy**:

**1. Peak Season Planning**: Focus your inventory buildup and marketing campaigns for Q1, especially during the months of February to July. Be prepared to handle the increased demand and ensure product availability during these months.

**2. February Strategy:** As February sees a slump, plan for clearance sales, promotions, or other incentives to maintain revenue during this period.

**3. January Kickoff:** Take advantage of the January surge by launching new products, running special promotions, or advertising heavily.

**4. Learn from Successful Years:**Study the successful years, like 2014, and try to identify what strategies and products contributed to high sales. Replicate those strategies in other years.

**5. Address Yearly Declines:** Investigate the reasons behind years with unusually low sales, like 2020 and 2023. Take corrective actions to prevent such declines.

**6. Weekday Targeting**:Concentrate your marketing efforts and resources on weekdays with the highest sales, particularly Fri, Thu, Wed, and Sun.

**7. Maintain Quality and Service:** Ensure that your products and customer service remain consistent across all periods and weekdays to maintain customer satisfaction and loyalty.

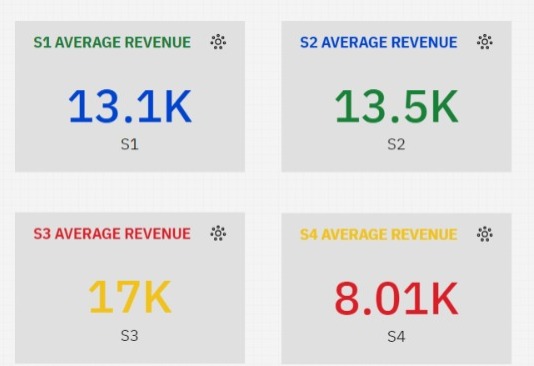
By incorporating these insights into your sales and marketing strategy, you can optimize your business for peak sales periods and ensure steady growth over the years.

**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.

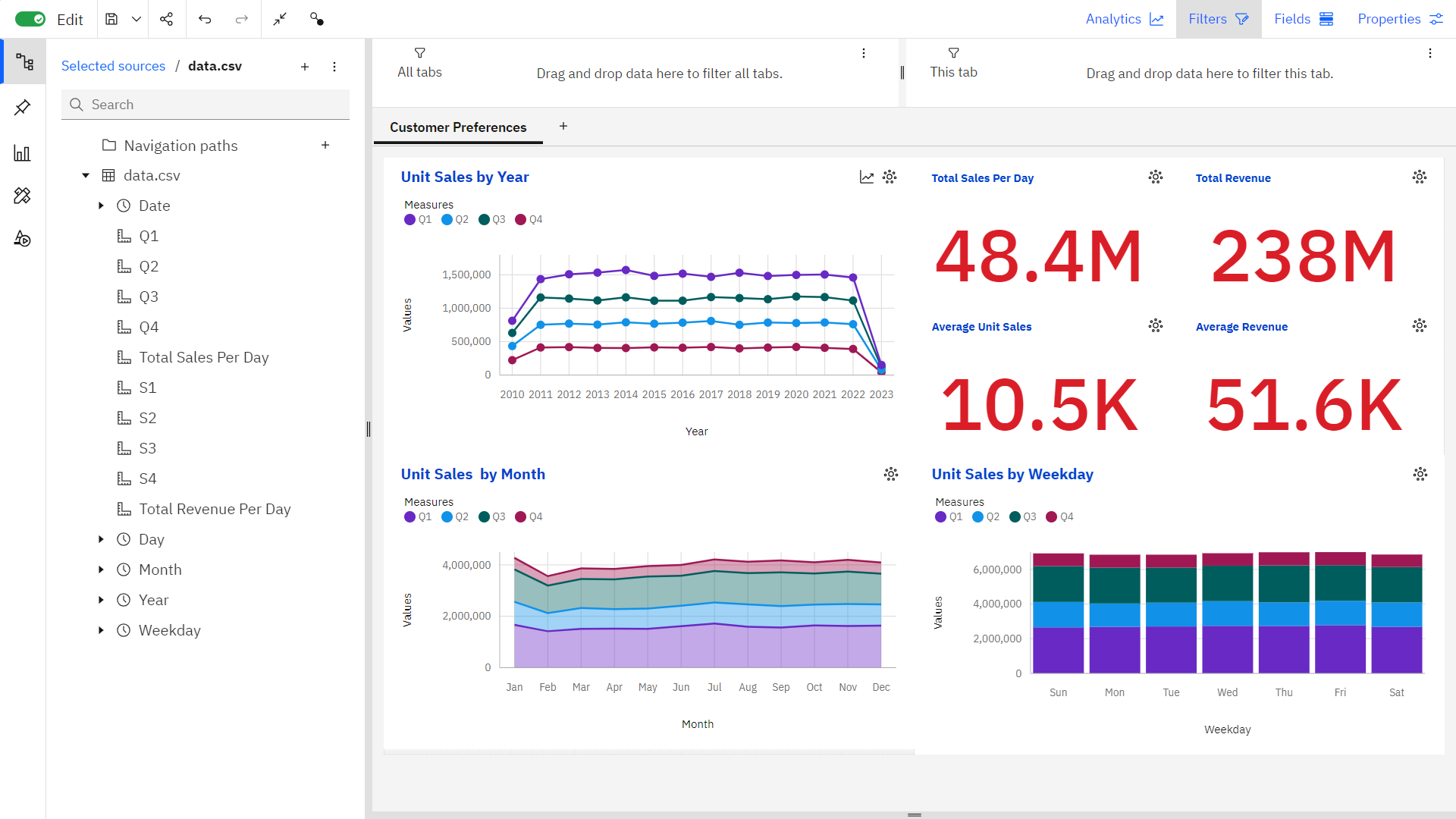
5. **Market Research**:Conduct market research to understand the preferences and needs of your target audience. This can help in product development and marketing strategies.

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

**iv.To identify customer preferences**

To discern customer preferences, we utilized visualizations in IBM Cognos:

* Line charts depicted yearly unit sales trends, aiding in understanding annual customer inclinations.
* Area charts showcased sales patterns by month, unveiling monthly customer preferences.
* Stacked column charts illustrated sales distribution by weekday, revealing customer buying habits across the week.
* Summary charts provided insights into total and average sales and revenue, facilitating a comprehensive overview of customer engagement and preferences**.**



**Observation:**

* July, November, January, September, and October are the most prevalent months across all quarters, representing 43.5% of the total sales counts for each quarter.
* The average sales per month across all quarters are over 4,000 (Q1), over 2,000 (Q2), over 3,000 (Q3), and over 1,000 (Q4).
* The total results for each quarter are over 4,500, emphasizing consistency in dataset entries.
* January consistently displays the highest total sales per day, reaching almost 4.3 million, while February records the lowest figures across both total sales and total revenue per day.
* April has the highest average total sales per day at over 10,600, while May shows the highest average total revenue per day, surpassing 52,300.
* Both total sales and revenue per day are projected to increase by the next month, with estimations exceeding 10,000 for total sales and 51,000 for total revenue per day.

These insights outline seasonal patterns, prevalent months, averages, and trends in sales and revenue, providing a comprehensive understanding of the dataset's dynamics and potential future projections**.**

**Customer Preferences:**

**1. Seasonal Trend and Frequency:**

- Customers exhibit a strong preference for specific months (Jul, Nov, Jan, Sep, Oct) across all quarters (Q1, Q2, Q3, Q4). Focusing on these periods may align with customer buying habits.

**2. Product Specific Preferences:**

- Observing Q1, Q2, Q3, and Q4 trends reveals variations in customer preferences for different products across various timeframes. Analyzing these patterns can guide marketing strategies.

**Actionable Insights based on Customer Preferences:**

**1. Seasonal Marketing Strategy:**

- Tailor marketing campaigns according to the seasonal trends observed in certain months. Leverage this insight to target and engage customers during these periods.

**2. Product Optimization:**

- Develop product-specific marketing and inventory strategies based on the quarterly trends. Addressing product popularity in specific periods could improve sales.

**3. Customer-Centric Approach:**

- Analyze customer preferences across months and quarters to anticipate demands. This approach can guide product development and service enhancements.

**4. Forecast-Driven Strategies:**

- Utilize forecasting insights to prepare for future customer demands. This proactive approach ensures adequate stock and tailored marketing strategies.

**5. Average Sales and Revenue Analysis:**

- Understanding average sales and revenue provides a benchmark to measure performance. Deviations from these averages can signal shifts in customer preferences.

These actionable insights derived from the data provide a roadmap for businesses to better understand customer preferences and align their strategies to meet customer needs effectively. The focus on seasonal trends, product preferences, and proactive planning based on forecasts and averages can significantly impact customer satisfaction and business success.

**v.to estimate the number of units of each product that could be sold on December 31st of each year if all retail centers were kept open**

The company closes all its retail centers on December 31st each year.Estimating sales on December 31st is crucial for planning and decision-making.

**Methods:**

**Method 1: Use Historical Sales Data from Nearby Dates:**

* Analyze sales data from days immediately before and after December 31st.
* Extrapolate estimates based on observed sales patterns during those adjacent days.
* Provides a rough estimate using nearby historical data, but may not account for unique December 31st factors.

**Method 2: Use Average Monthly Sales Patterns:**

* Calculate the average sales for each product over entire months, excluding December.
* Apply this average to estimate sales on December 31st, assuming consistent sales patterns.
* A more systematic approach that considers overall sales trends throughout the year, accounting for variations.

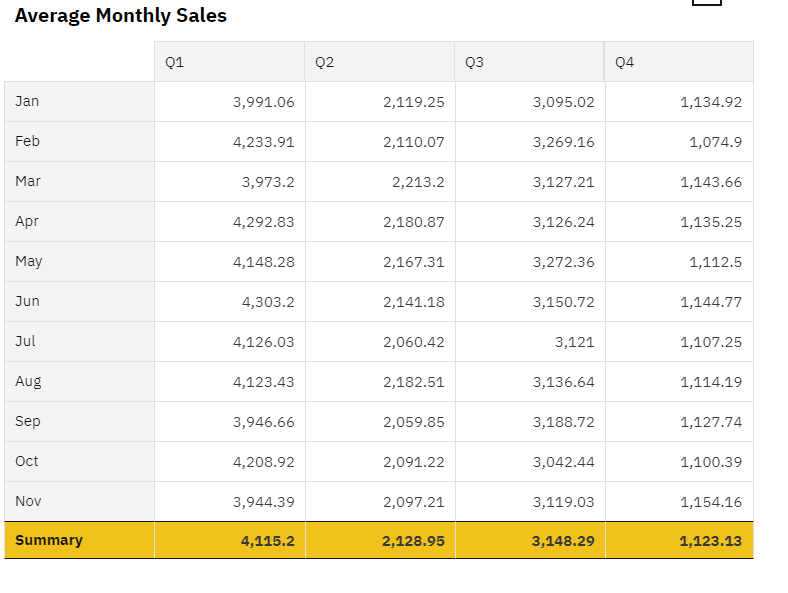
**Method 3: Find Average of Sales on the 31st of Every Month:**

* Determine the average sales on the 31st day of each month.
* Use this average as an estimate for December 31st sales.
* Assumes uniform sales patterns on all 31st days, which may not accurately represent the unique characteristics of December 31st.

**Recommendation:**

* Method 2, using average monthly sales patterns, is recommended for estimating December 31st sales.
* Acknowledge the limitations of all methods and communicate them to stakeholders.
* Provide regular updates and refined estimates as more data becomes availale
* Accurately estimating sales on December 31st is challenging when retail centers are closed.
* The selected method offers a reasonable approximation for planning and decision-making.
* Continuous monitoring and adjustments to estimates are advised**.**

**Cross Tab Chart**

****

Based on the provided average monthly sales data for each product (P1, P2, P3, and P4), we can estimate the sales for December (excluding December) by simply using the respective average sales for each product.

Assuming that we have average monthly sales data for 11 months (excluding December), we can estimate December sales for each product as follows:

* + **P1: 4115 units**
  + **P2: 2128 units**
  + **P3: 3148 units**
  + **P4: 1123 units**

These estimates are based on the average monthly sales data you provided, assuming that sales patterns are relatively consistent from month to month. Please keep in mind that these are approximate estimates and may not account for any unique factors or seasonal variations that might affect December sales differently.

**Actionable Insights:**

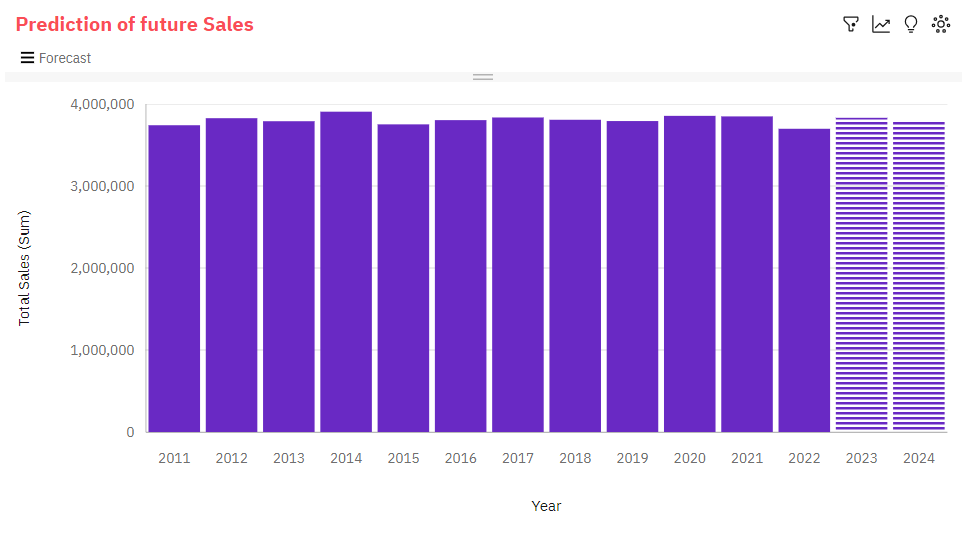
The estimated December sales for each product, based on the provided average monthly sales data, can provide some actionable insights for REC Corp:

* Stocking Inventory: REC Corp can use these estimated figures to ensure that there are adequate product quantities available to meet the expected demand for December. This helps prevent stockouts and ensures a positive customer experience.
* Promotional Planning: While these are approximate estimates, REC Corp can plan promotional or marketing activities for December based on these figures. For example, they could create product-specific promotions or incentives to boost sales further.
* Sales Target Setting: These estimates can serve as a benchmark for setting sales targets for the December month. The sales team can be given specific targets to achieve based on these estimates.
* Production Planning: If REC Corp manufactures these products, the estimated December sales can influence production planning. It helps in allocating resources and workforce accordingly.

**vi.to predict future sales and revenue in 2024**

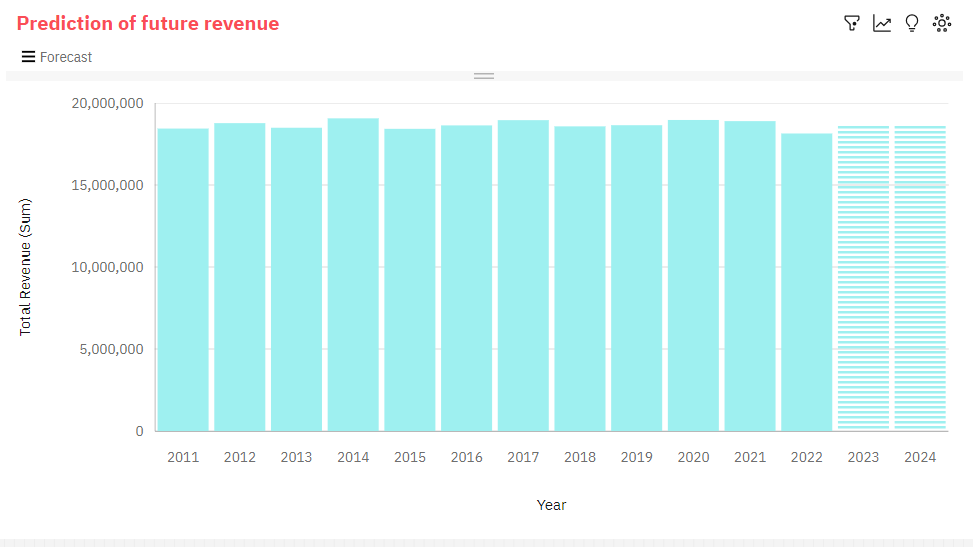
To predict the future sales ,create a column chart that displays historical sales data up to the present.

* After creating the column chart, navigate to the chart options or properties.Look for the "Forecasting" option.
* Access the forecasting parameters where you can specify the target year, which is 2024 in this case.
* Set the desired forecasting method
* Once you've configured the parameters, trigger the forecast generation.
* IBM Cognos will use the selected method to extrapolate future sales data for 2024 based on the historical data provided in the chart.
* Examine the forecasted values displayed in the chart, which will extend into 2024.
* Interpret the forecast to understand the predicted sales trends for the upcoming year.
* Use the Forecast for Decision-Making
* The generated forecast can help with planning, resource allocation, and making informed decisions for the year 2024 based on predicted sales trends.
* Do the same for predicting the revenue



**Observation:**

* Total Sales has a moderate upward trend.
* Based on the current forecasting, Total Sales may reach nearly 3.8 million by Year 2024
* From 2014 to 2015, Total Sales dropped by 4%.
* Total Sales ranges from almost 3.7 million, in 2022, to over 3.9 million, in 2014.

****

**Observation:**

* Total Revenue has a weak upward trend.
* Based on the current forecasting, Total Revenue may reach nearly nineteen million by Year 2024.
* From 2021 to 2022, Total Revenue dropped by 4%.
* Across all years, the sum of Total Revenue is over 224 million.
* Total Revenue ranges from over eighteen million, in 2022, to over nineteen million, in 2014.
* Total Revenue ranges from over eighteen million, in 2022, to over nineteen million, in 2014.

**Actionable Insights:**

**Total Sales Trends:**

* Moderate Upward Trend: The upward trend in Total Sales suggests a positive growth trajectory. REC Corp should focus on sustaining this growth and potentially enhancing it.
* Forecasted Growth: The forecast indicates that Total Sales may reach nearly 3.8 million by Year 2024. REC Corp can use this information for setting ambitious yet achievable sales targets for the upcoming year.

**Total Sales Drops in 2014-2015:**

Identify Causative Factors: REC Corp should investigate the factors that led to a 4% drop in Total Sales from 2014 to 2015. Understanding the causes of such declines can help prevent or mitigate similar drops in the future.

**Total Sales Range:**

Fluctuation in Sales: The range of Total Sales varies from almost 3.7 million (in 2022) to over 3.9 million (in 2014). REC Corp can use this information to understand the seasonality and cyclicality in their sales and plan their inventory accordingly.

**Total Revenue Trends:**

Weak Upward Trend: While Total Revenue has an upward trend, it is described as weak. REC Corp may consider exploring strategies to boost revenue growth more significantly.

**Forecasted Revenue Growth:**

Revenue Growth Potential: The forecast predicts that Total Revenue may reach nearly nineteen million by Year 2024. REC Corp can set revenue targets aligned with this projection and allocate resources accordingly.

**How Insights Guide Inventory Management and Marketing Strategies**

The insights obtained from this project guide inventory management and marketing strategies in the following ways:

**i.Inventory Management:**

* Inventory planning is optimized to align with peak sales periods.
* Strategies are developed to handle February slumps effectively.
* Continuous monitoring helps in addressing yearly declines.

**ii.Marketing Strategies:**

* Marketing efforts are concentrated during peak months and days.
* Product-specific marketing strategies are implemented based on customer preferences.
* Predictive insights help in strategic planning for the year 2024.

**Replication Instruction:**

To replicate the analysis and generate visualizations using IBM Cognos:

* Obtain access to IBM Cognos Analytics.
* Import the preprocessed dataset used in the project.
* Create reports and dashboards using IBM Cognos.
* Follow the visualizations and insights provided in the project to analyze your data.
* Example outputs and further details can be found in the GitHub repository as DAC\_Phase5\_Dashboard and the earlier phases of the project documentation.It is included in this document also

**Advantages:**

* Informed Decision-Making: The project empowers REC Corp with data-driven insights, enabling informed decision-making for inventory management and marketing strategies.
* Forecasting Capability: By incorporating machine learning techniques such as SARIMAX, REC Corp gains the ability to make accurate sales and revenue forecasts, helping in resource allocation and budget planning.
* Seasonal Analysis: The project provides a comprehensive understanding of seasonal sales and revenue trends, which can guide inventory planning and marketing campaigns during peak seasons.
* Historical Insights: Historical data analysis allows to identify trends and patterns, helping to optimize sales and marketing strategies based on past performance.
* Data Visualization: The use of tools like IBM Cognos for data visualization enhances data interpretation, making insights more accessible to various stakeholders.

**Disadvantages:**

* Data Quality Challenges: Inaccurate or incomplete data can lead to flawed analysis and forecasts. Data cleaning and preprocessing are essential but time-consuming processes.
* Complex Models: Implementing machine learning models like SARIMAX may require specialized knowledge, which can be a challenge for companies without data science expertise.
* External Factors: The analysis may not consider all external factors influencing sales, such as market trends, economic changes, or competitive actions.
* Over-Reliance on Past Data: Overreliance on historical data might hinder adaptability to sudden market changes or disruptive events.
* Resource Intensive: Running machine learning models and processing large datasets can be computationally intensive, requiring appropriate hardware and computational resources.

**Applications:**

* Inventory Management: The project's insights can guide efficient inventory management by helping REC Corp adjust stock levels based on sales trends.
* Marketing Strategies: Data-driven insights can optimize marketing strategies, enabling REC Corp to focus efforts on products and periods with the highest sales potential.
* Revenue Forecasting: Accurate revenue forecasts support financial planning, budgeting, and investment decisions.
* Seasonal Campaigns: Understanding seasonal sales trends can assist REC Corp in planning and executing targeted marketing campaigns during peak sales periods.

**Future Scope:**

* Advanced Machine Learning: The project can expand to incorporate more advanced machine learning and predictive modeling techniques to enhance forecasting accuracy.
* Real-Time Analytics: Future iterations could aim to provide real-time analytics, enabling REC Corp to respond swiftly to market changes.
* Integration with IoT: Integration with Internet of Things (IoT) devices and sensors can provide real-time sales data, allowing for even more dynamic decision-making.
* Predictive Maintenance: Beyond sales, predictive analytics can be extended to predict maintenance requirements for manufacturing equipment or delivery schedules for optimizing the supply chain.
* Data Security: Enhancing data security measures is crucial, especially if the project involves sensitive customer or financial data.
* Machine Learning Automation: Automation tools can be explored to streamline model building and deployment processes, reducing the reliance on specialized data science expertise.
* Global Expansion: If REC Corp operates globally, the project can be expanded to analyze sales and revenue data from various regions, offering insights into market-specific trends and strategies.

**Conclusion:**

The Product Sales Analysis project provides valuable insights into sales trends, top-selling products, and customer preferences. These insights offer actionable recommendations for inventory management and marketing strategies. The design thinking process, development phases, and documentation ensure the project's value as a reference and guide for data-driven decision-making and future improvements.

**Github Link:**

For the complete project code and files, please refer the notebook DAC\_Phase5\_AnalysisCode the Github Repository

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

**Appendix:**

**Data Analysis Code:**

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

import numpy as np # library used for working with arrays

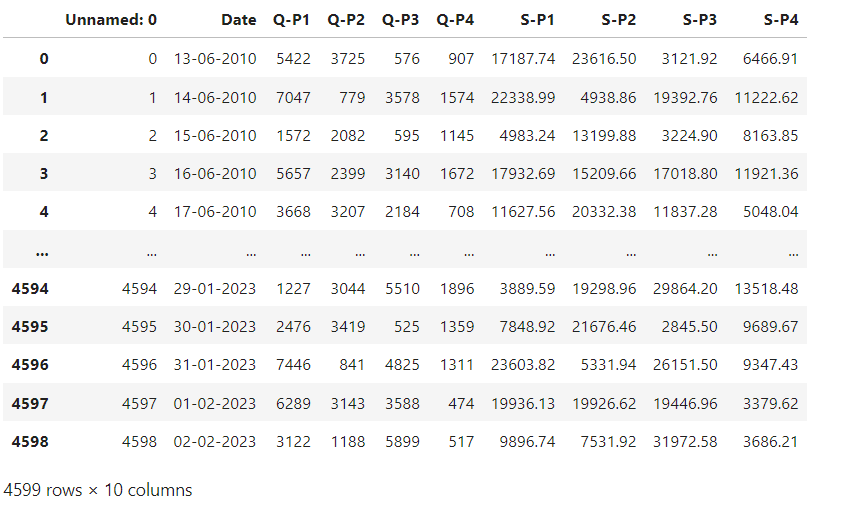
import matplotlib.pyplot as plt # library for plots and visualizations

import seaborn as sns # library for visualizations

%matplotlib inline

*# Checking the first 5 and last 5 rows of the dataset*

data**.**head(**-**1)



*# drop the first column*

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

*# Extract year from the 'Day' 'Month' 'year' from the 'Date' column using a lambda function*

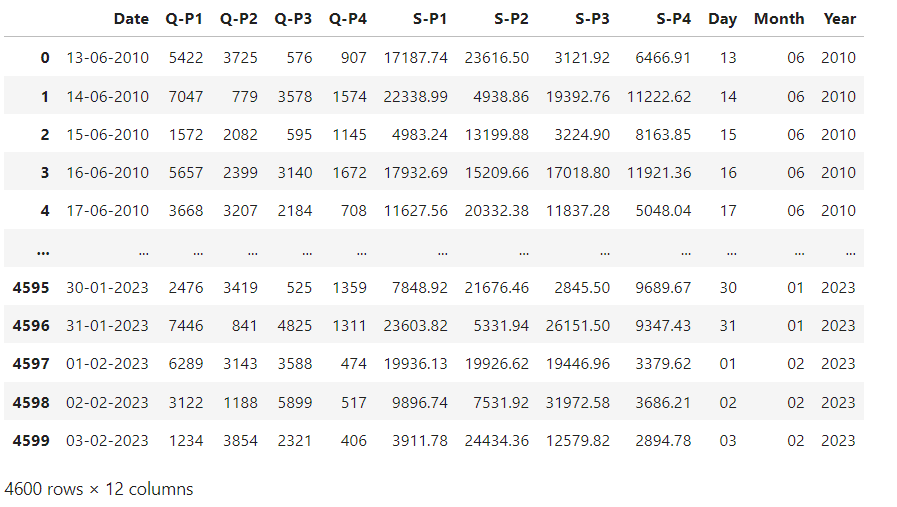
*# We need to get the year from the data to analyse sales year to year*

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

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

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

data



*#Create a function that allows us to plot a bar chart for the 4 products*

**def** plot\_bar\_chart(df, columns, stri, str1, val):

*# Aggregate sales for each product by year, by sum or mean*

**if** val **==** 'sum':

sales\_by\_year **=** df**.**groupby('Year')[columns]**.**sum()**.**reset\_index()

**elif** val **==** 'mean':

sales\_by\_year **=** df**.**groupby('Year')[columns]**.**mean()**.**reset\_index()

*# Melt the data to make it easier to plot*

sales\_by\_year\_melted **=** pd**.**melt(sales\_by\_year, id\_vars**=**'Year', value\_vars**=**columns, var\_name**=**'Product', value\_name**=**'Sales')

*# Create a bar chart*

plt**.**figure(figsize**=**(20,4))

sns**.**barplot(data**=**sales\_by\_year\_melted, x**=**'Year', y**=**'Sales', hue**=**'Product') *#,palette="cividis")*

plt**.**xlabel('Year')

plt**.**ylabel(stri)

plt**.**title(f'{stri} by {str1}')

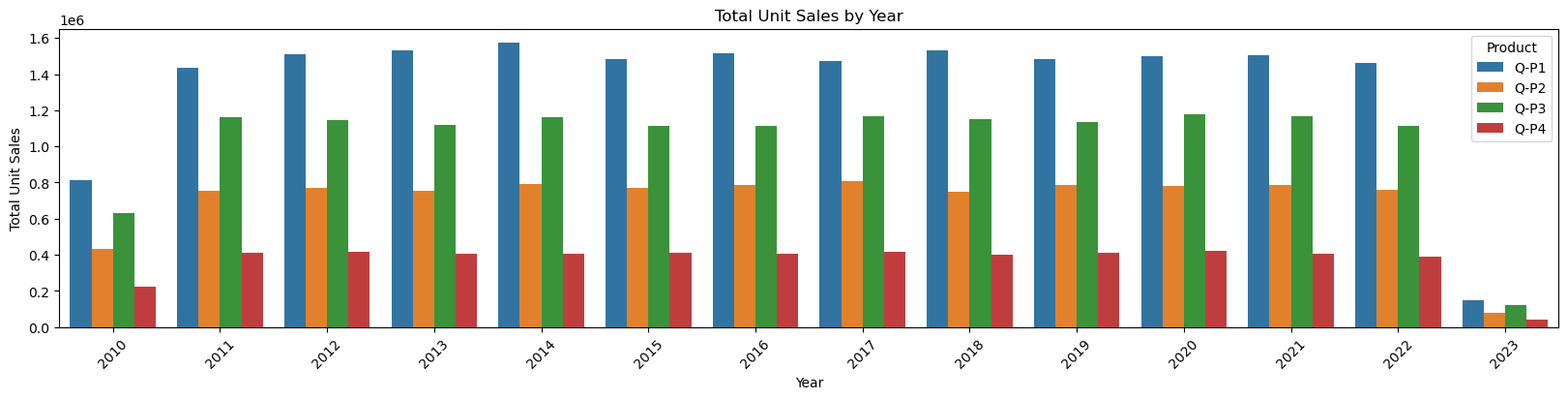
plt**.**xticks(rotation**=**45)

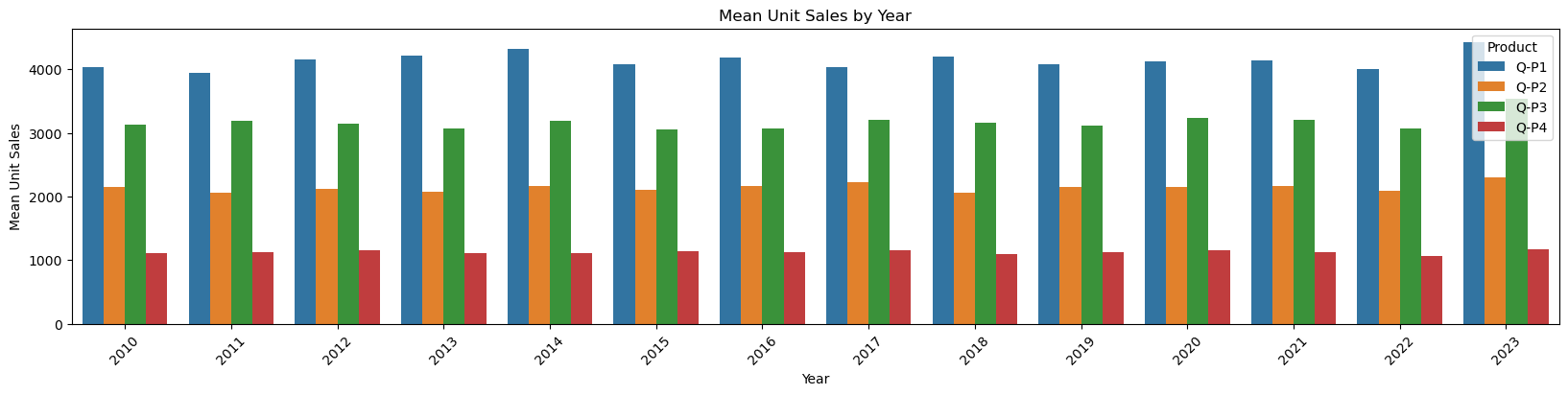
plt**.**show()

*#use the plot\_bar\_chart function, enter the Unit Sales Columns and the Unit Sales string*

plot\_bar\_chart(data, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4'],'Total Unit Sales', 'Year', 'sum')

plot\_bar\_chart(data, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4'],'Mean Unit Sales', 'Year', 'mean')



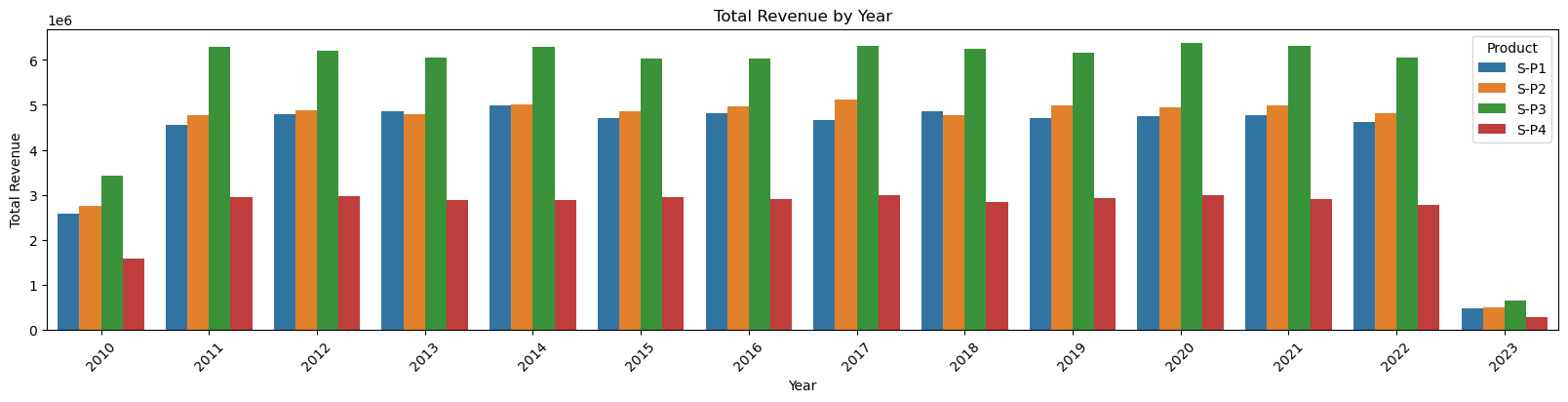


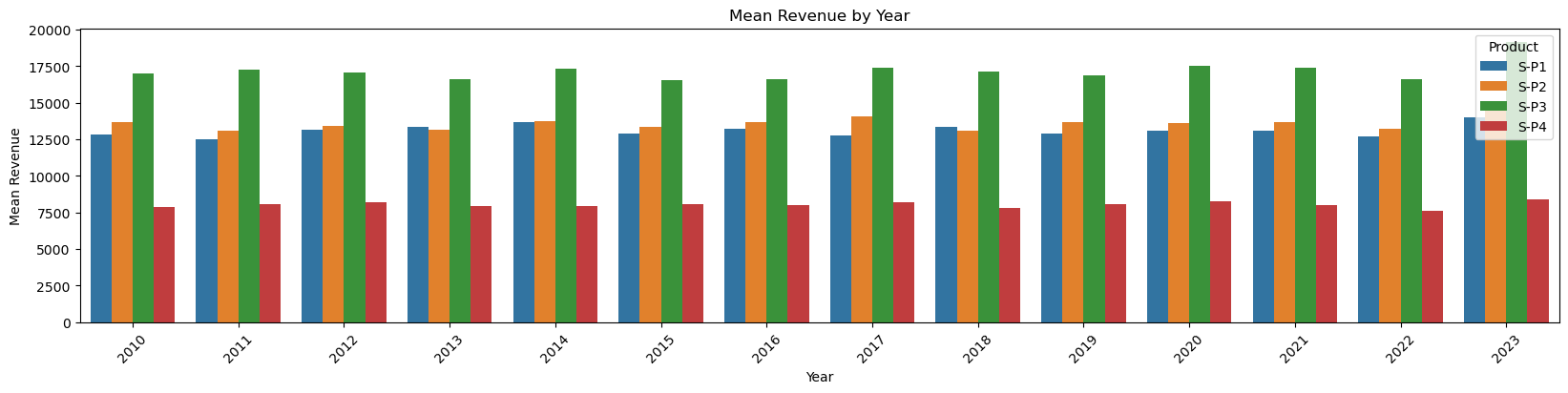
In [11]:

*#use the plot\_bar\_chart function, enter the Revenue Columns and the Revenue string*

plot\_bar\_chart(data, ['S-P1', 'S-P2', 'S-P3', 'S-P4'], 'Total Revenue', 'Year', 'sum')

plot\_bar\_chart(data, ['S-P1', 'S-P2', 'S-P3', 'S-P4'], 'Mean Revenue', 'Year', 'mean')





*# Create a figure and axis*

**def** month\_plot():

fig, ax **=** plt**.**subplots()

*# Plot the sales data for each product by month*

data**.**groupby('Month')[['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4']]**.**sum()**.**plot(ax**=**ax)

*# Set the x-axis limits to only show up to December*

ax**.**set\_xlim(left**=**0, right**=**13)

*# Set the axis labels and title*

ax**.**set\_xlabel('Month')

ax**.**set\_ylabel('Total unit sales')

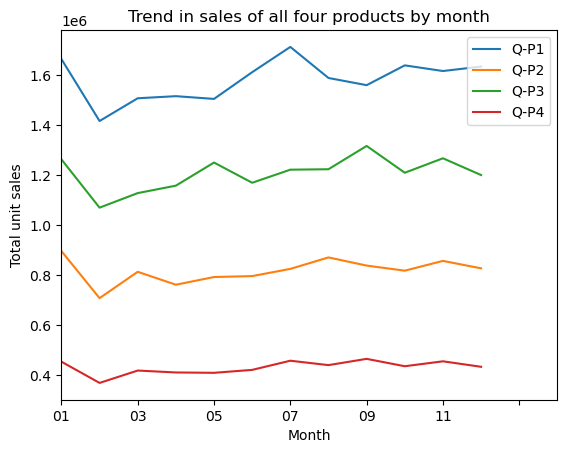
ax**.**set\_title('Trend in sales of all four products by month')

data['Month'] **=** data['Month']**.**replace('9', '09')

*# Show the plot*

plt**.**show()

month\_plot()



*#get the 31st day for each month in each year. Note: not every month has 31 days*

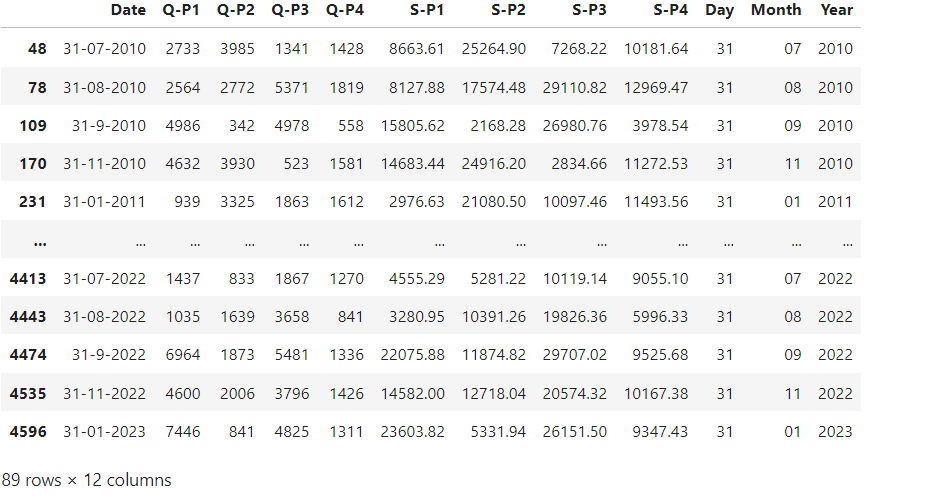
**def** month\_31\_data(df, months):

m31\_data **=** df[df['Month']**.**isin(months) **&** (df['Day'] **==** '31')]

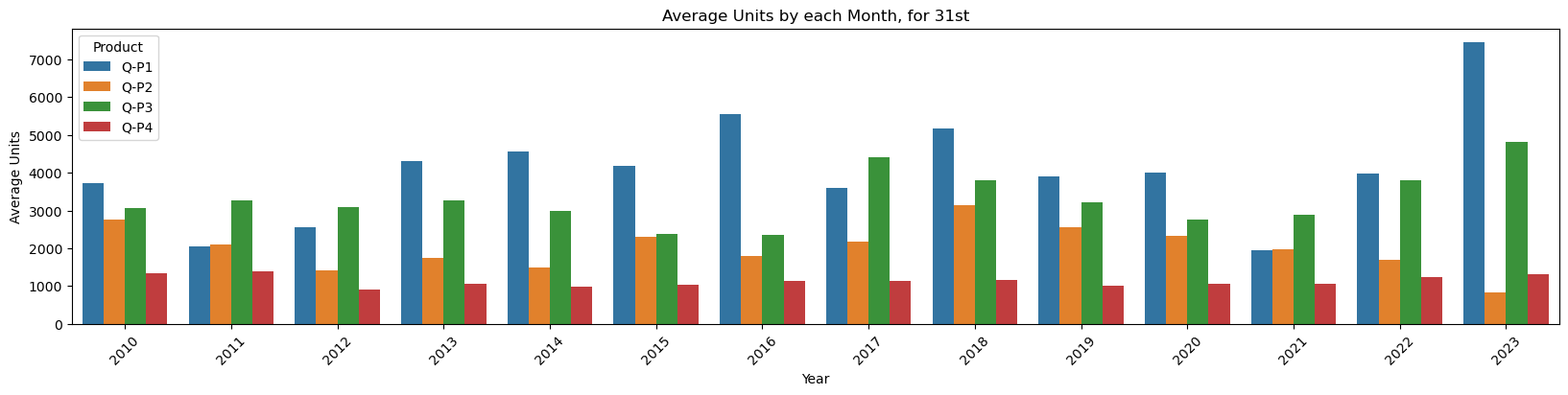
**return** m31\_data

\_31\_months **=** month\_31\_data(data, ['01', '02', '03', '04', '05', '06', '07', '08', '09', '10', '11', '12'])

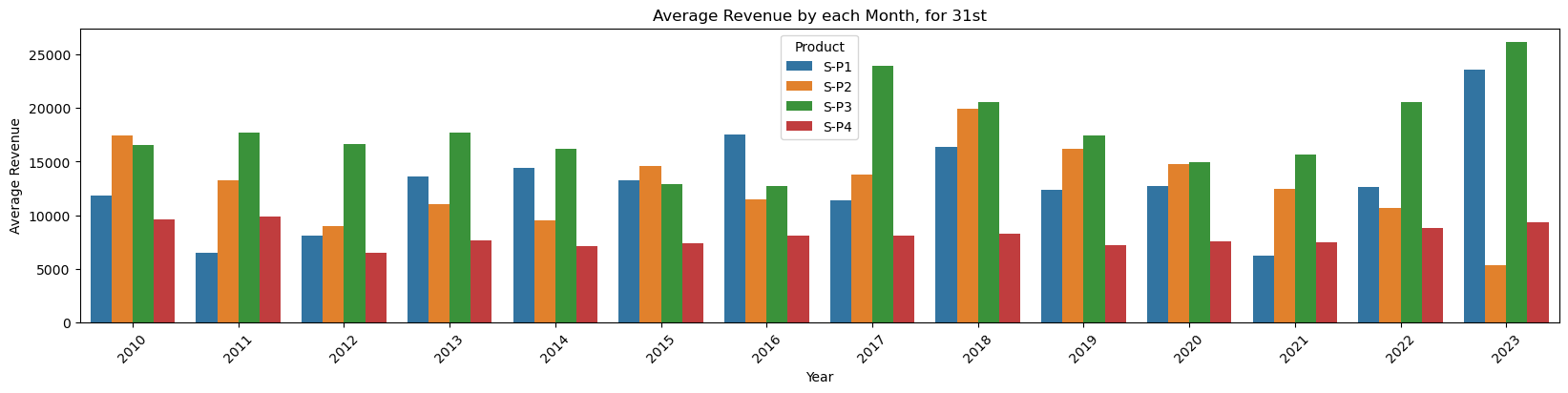
\_31\_months



plot\_bar\_chart(\_31\_months, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4'], 'Average Units', 'each Month, for 31st', 'mean')



plot\_bar\_chart(\_31\_months, ['S-P1', 'S-P2', 'S-P3', 'S-P4'], 'Average Revenue', 'each Month, for 31st', 'mean')



*# gives us the average for all the 31st days across all years for each product*

**def** avg\_on\_31st(df, product):

df\_31 **=** df[df['Day'] **==** '31']

avg\_sales **=** df\_31[product]**.**mean()

**return** avg\_sales

*# Average for Unit Sales*

avg\_on\_31st(data, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4'])**.**round(2)

**Q-P1 3850.73**

**Q-P2 2076.51**

**Q-P3 3196.45**

**Q-P4 1112.13**

**dtype: float64**

*# Average for Revenue*

avg\_on\_31st(data, ['S-P1', 'S-P2', 'S-P3', 'S-P4'])**.**round(2)

**S-P1 12206.82**

**S-P2 13165.05**

**S-P3 17324.76**

**S-P4 7929.52**

**dtype: float64**