Phase 5: Project Documentation & Submission

Project objectives:

Problem Definition and Design Thinking

In this part you will need to understand the problem statement and create a document on what have you understood and how will you proceed ahead with solving the problem. Please think on a design and present in form of a document.

Project Definition: The 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.

Analysis approach:

Assessing the data of sales and revenue of a company to understand the sales of the certain products.

Data Collection: we have collected data from their retail centers and organized by REC corp LTD. is small-scaled business venture established in India. They have been selling FOUR PRODUCTS for OVER TEN YEARS .

Exploratory Data Analysis (EDA): Since our data is an already structured dataset there was no need for data preparation. The data is checked for missing values, duplicates and outliers. In our data there was no such things.

Visualization of Predictions:	
	Create visualizations for the machine learning predictions to make them

easily understandable to business stakeholders. Highlight predicted trends and
behavioural insights in your reports and dashboards.

Program:

```
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
data = pd.read_csv('/content/statsfinal.csv')

# Display the first few rows of the
dataset print(data.head())

# Create histograms for specific
columns plt.figure(figsize=(12, 6))
plt.subplot(2, 2, 1)
plt.hist(data['Q-P1'], bins=30,
color='skyblue', alpha=0.7)
plt.title('Histogram of Q-P1')
plt.xlabel('V
alues')
plt.ylabel('F
requency')

plt.subplot(2, 2, 2)
plt.hist(data['Q-P2'], bins=30,
color='salmon', alpha=0.7)
plt.title('Histogram of Q-P2')
```

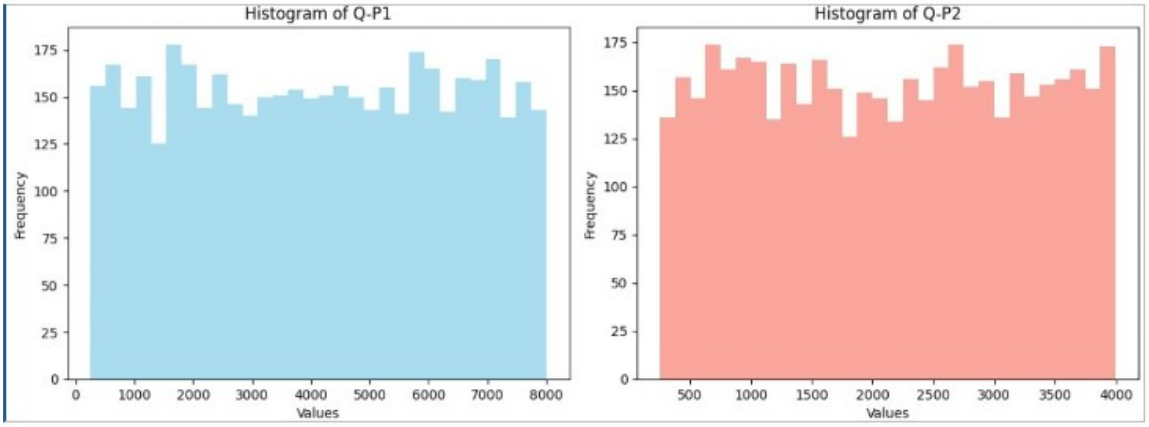
```
plt.xlabel('V
alues')
plt.ylabel('F
requency')
plt.subplot(2, 2, 3)
plt.hist(data['S-P1'], bins=30, color='lightgreen',
alpha=0.7) plt.title('Histogram of S-P1')
plt.xlabel('V
alues')
plt.ylabel('F
requency')

plt.subplot(2, 2, 4)
plt.hist(data['S-P2'], bins=30, color='lightcoral',
alpha=0.7) plt.title('Histogram of S-P2')
plt.xlabel('V
alues')
plt.ylabel('F
requency')

plt.tight_lay
out()
plt.show()
```

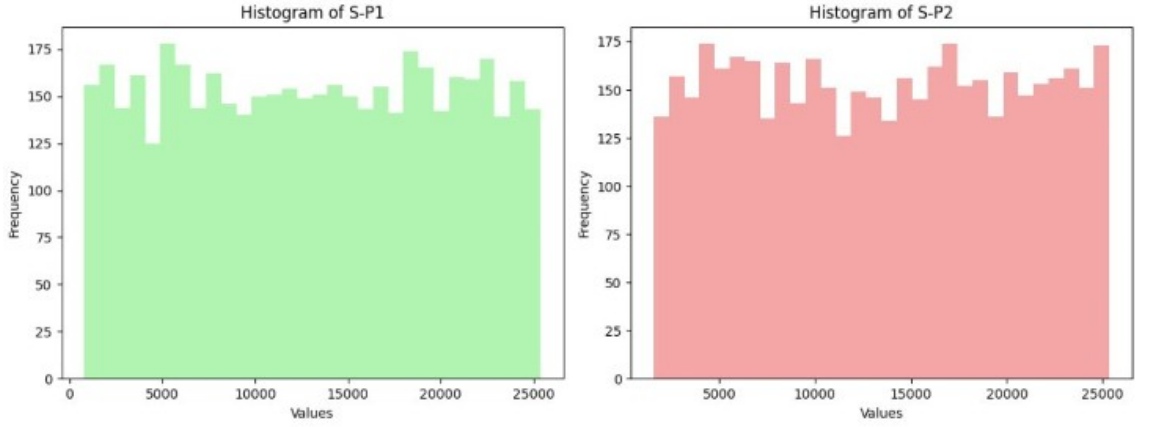
OUTPUT:

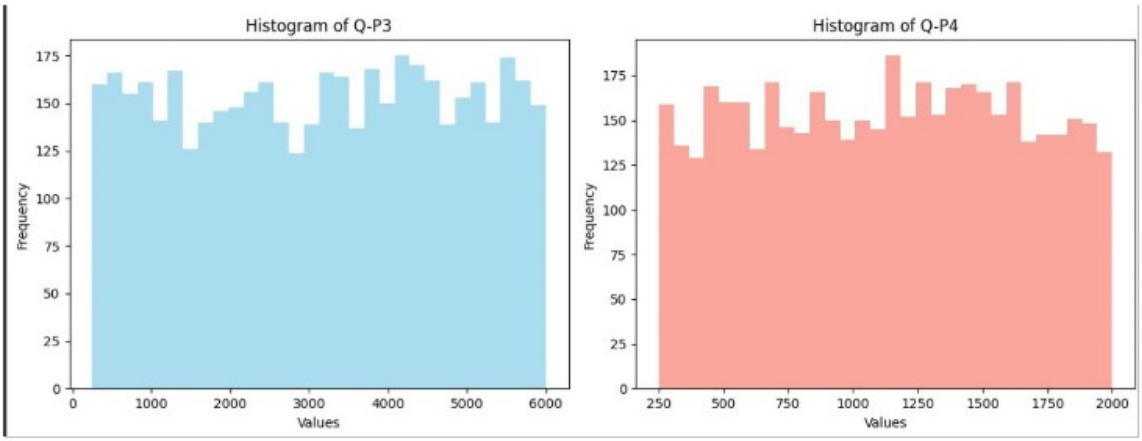
```
  Unnamed: 0  Date  Q-P1  Q-P2  Q-P3  Q-P4     S-P1      S-P2  \

0        0   13-06-   5422 372   576 907  17187.74
                2010            5          23616.50
1        1   14-06-   7047 779 357   1574  22338.99
```

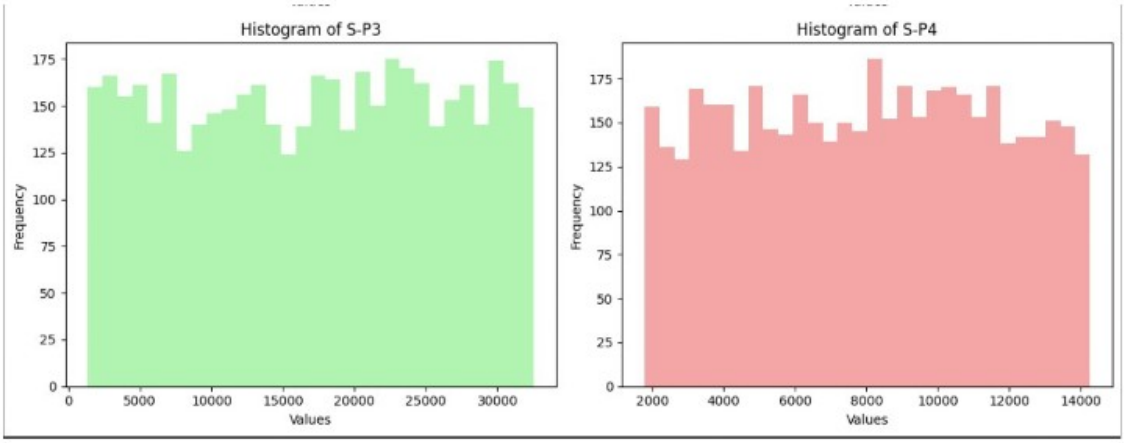
		2010			8	4938.86	
2	2	15-06-2010	1572	208 2	595	1145 13199.88	4983.24
3	3	16-06-2010	5657	239 9	314 0	1672 15209.66	17932.69
4	4	17-06-2010	3668	320 7	218 4	708 20332.38	11627.56

S-P3	S-P4
0 3121.92	6466.9 1
1 19392.76	11222. 62
2 3224.90	8163.8 5
3 17018.80	11921. 36
4 11837.28	5048.0 4







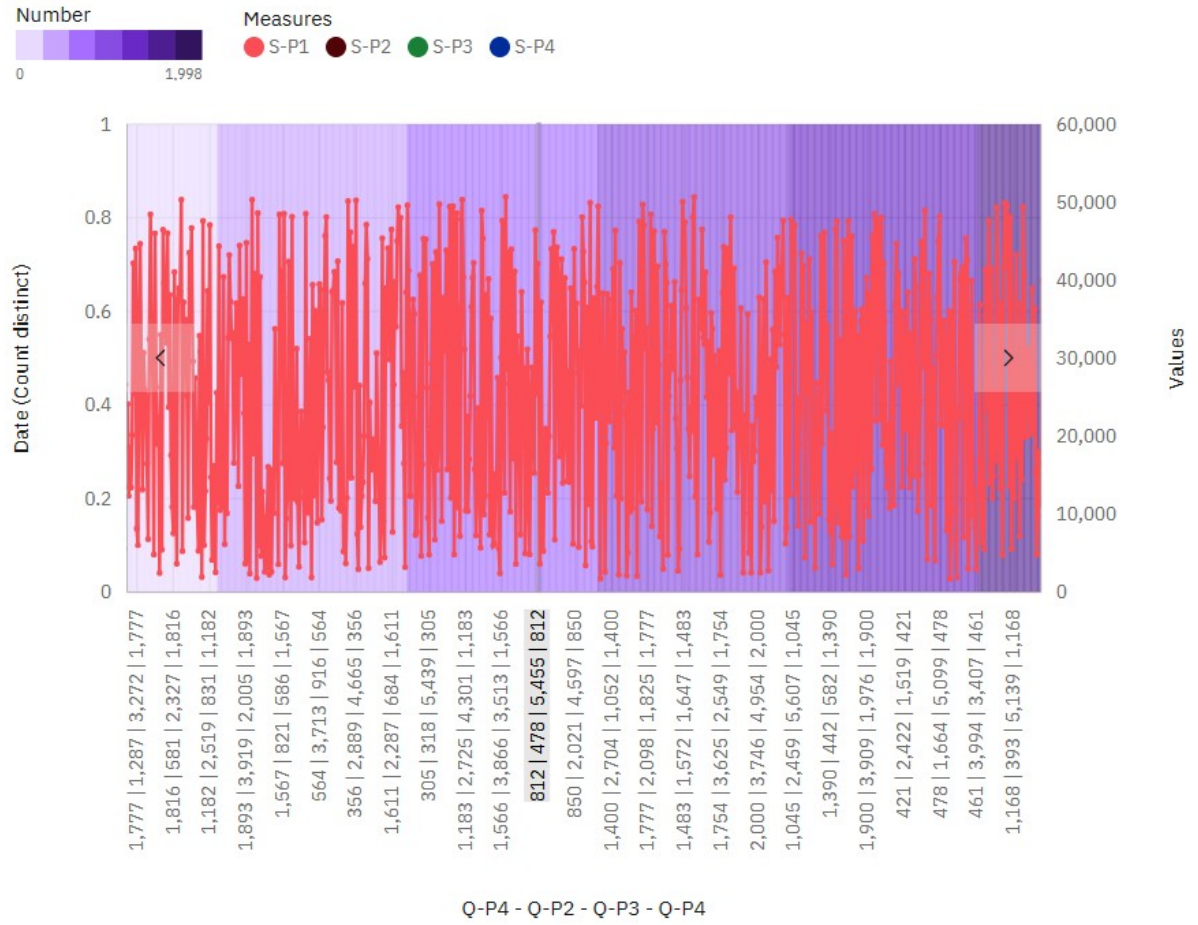


Data Cleaning:

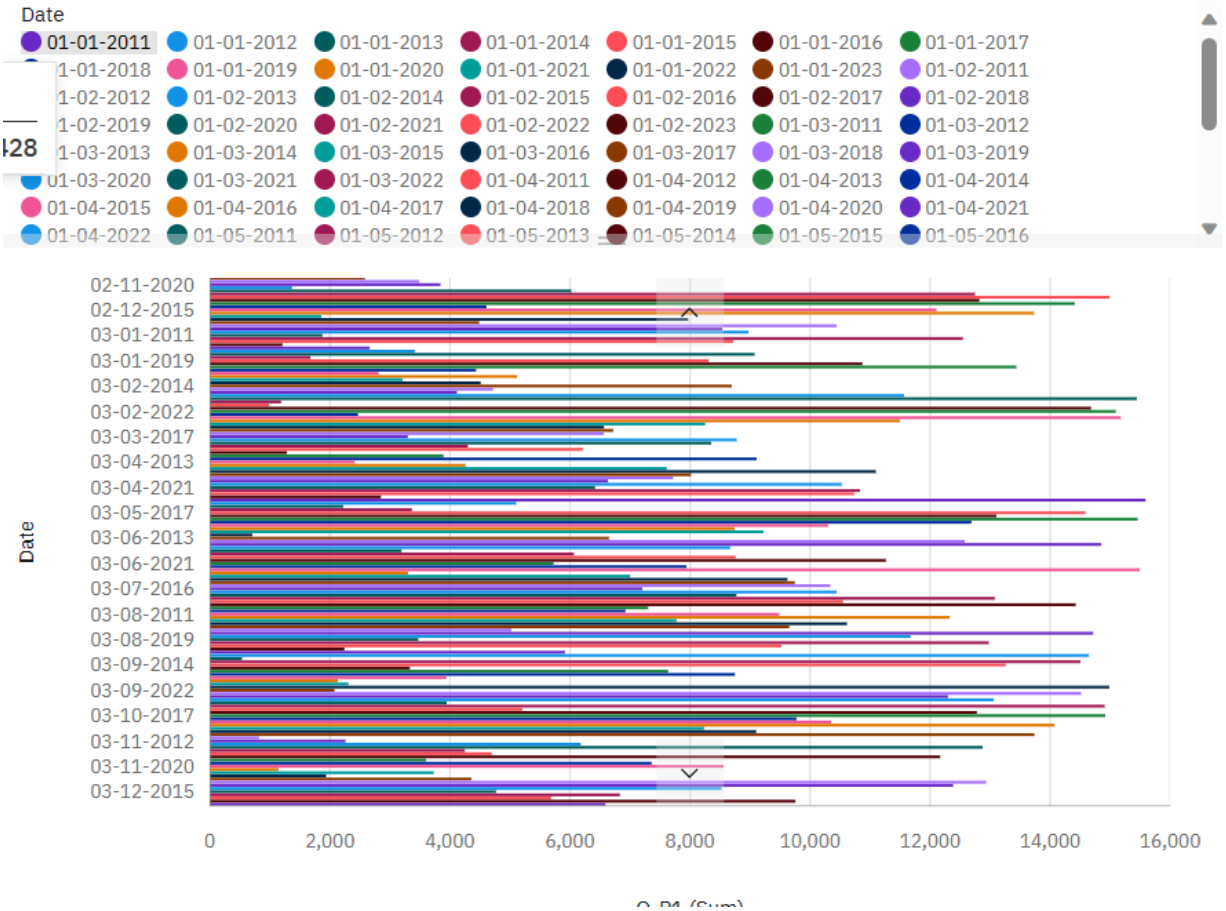
Since there was no missing values or duplicates there wasno need for much data cleaning.

Data Visualization:

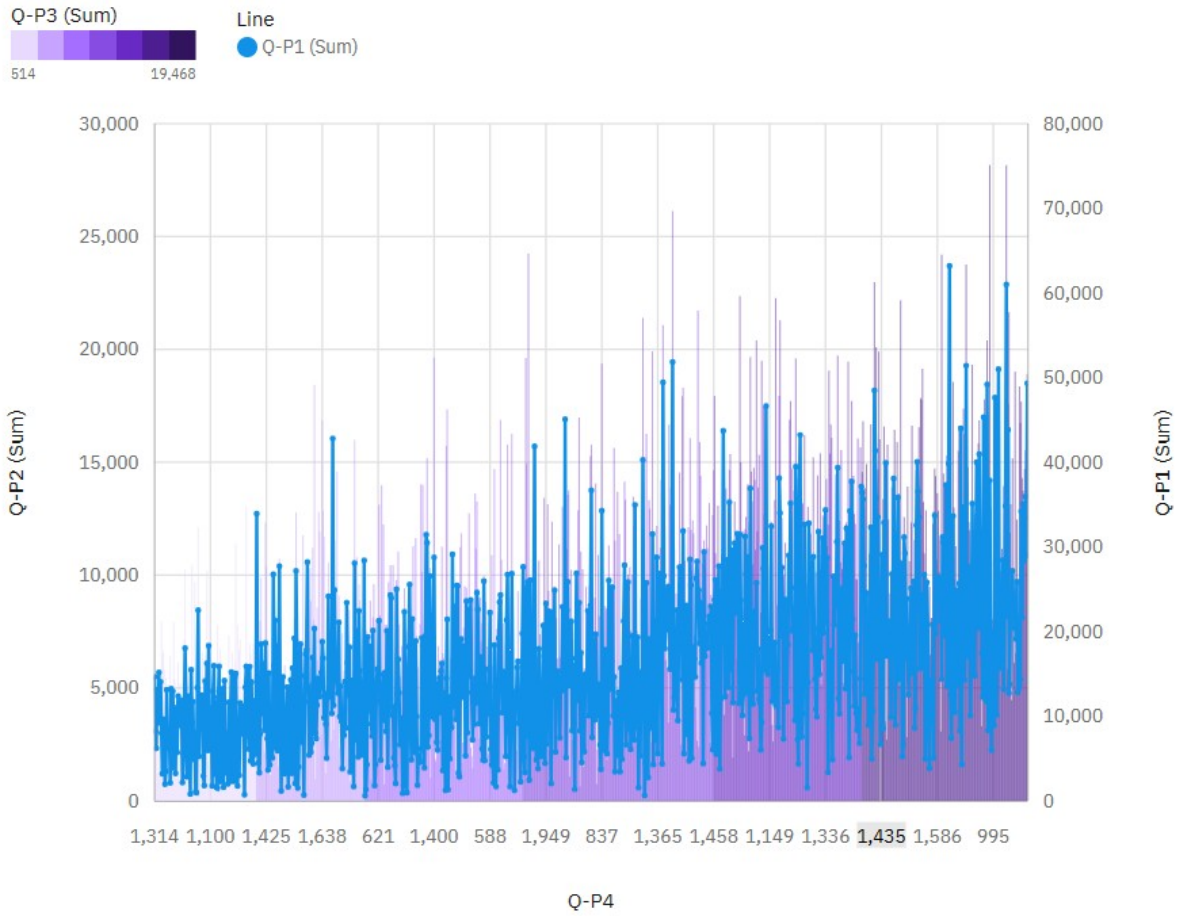
S-P1S-P2S-P3S-P4 and Date for Q-P4, Q-P2, Q-P3 and Q-P4 colored by Number



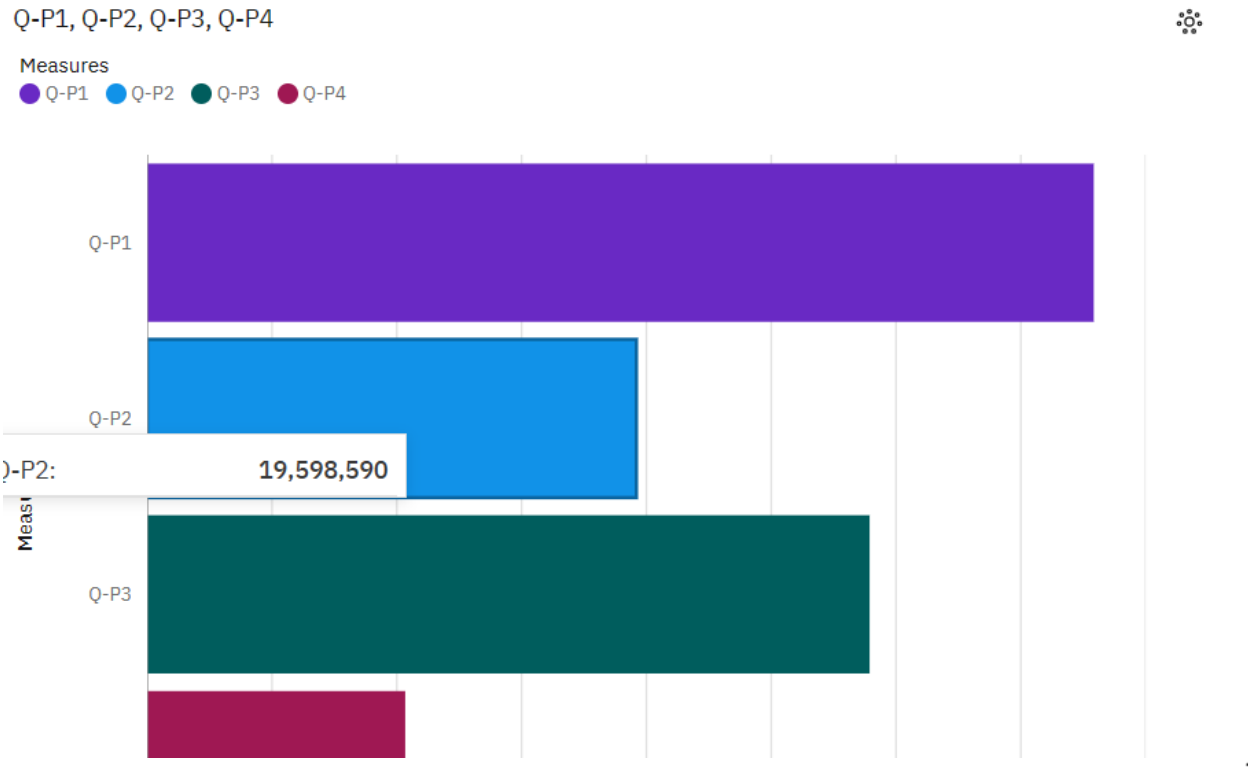
Q-P1 by Date colored by Date

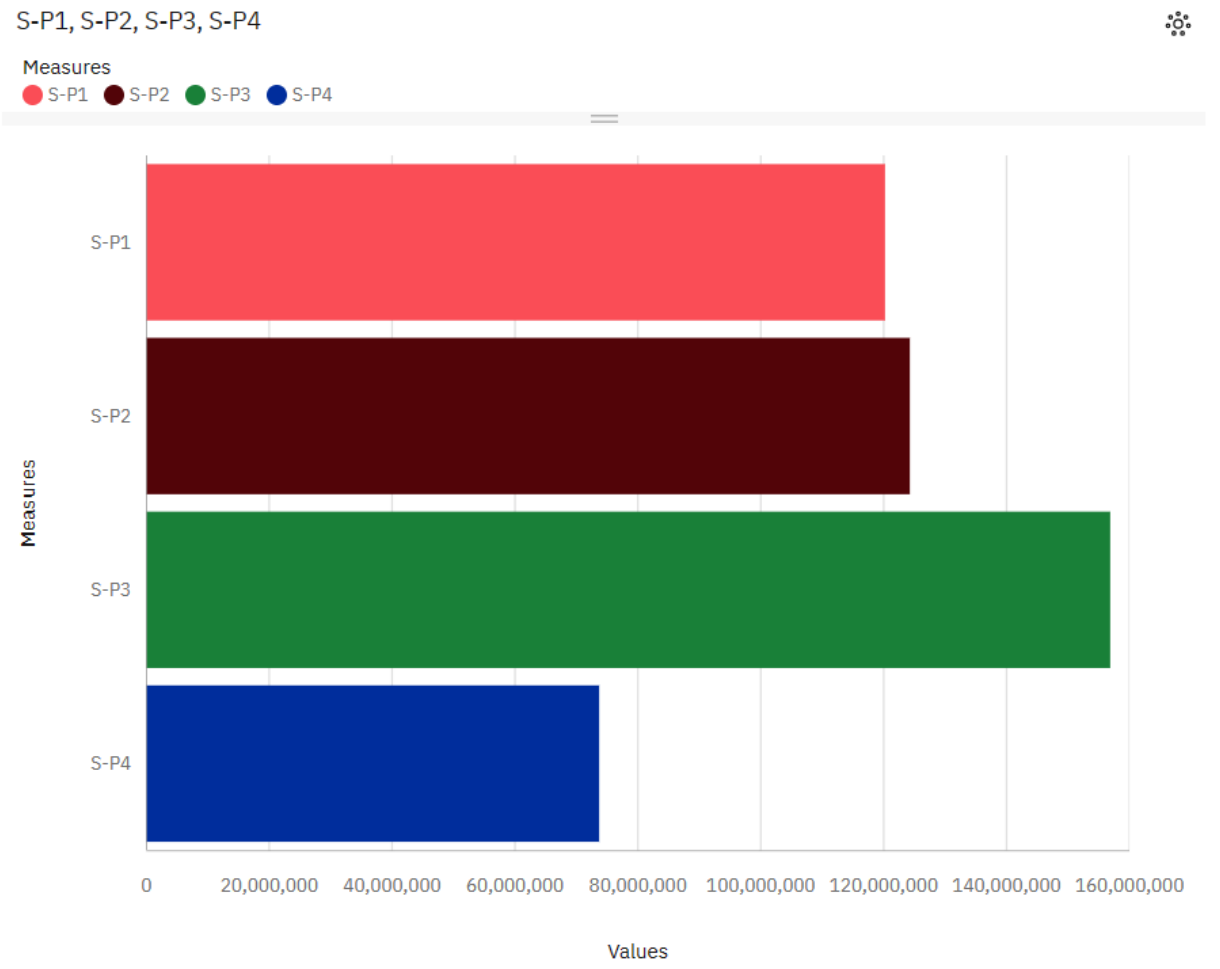


Q-P1 and Q-P2 for Q-P4 colored by Q-P3



Numerous visualizations such as bar charts,column chart were created using IBM cognos analytics and google colab notebook to analyse the relations between the data.





INTERPRETATION:

The chart shows the Total unit sales of product 1,2,3,4

according to this chart we can see that the sales of the product p1 is the highest.

The chart shows the Total revenue of the product 1,2,3,4 according to this chart we can see that the sales of the product p3 is the highest.

By the above charts there we can see that even though the sales of product p1 is high but the revenue generated is very low. Sales of product p3 is moderate but the revenue generated is very high.

Out of all four products , which product has seen the highest sales in all the given years?

In terms of the sales the product p1 is having a highest sales among the given products.

Is there any trend in the sales of all four products during certain months?

Yes we can see a trend of a is high sales of product in the middle of each year.

Conclusion:

From this analysis we can see the relationship between the sales and revenue of the products. This analysis can be used to predict, from which the product has high sales and the amount od sales aquired in a given tim and can be used to predict for the next year also The visualizations included help us better understand the relationship between the variables and make it easier to explain to others.