**PHASE 3**

**DEVELOPMENT PART 1**

|  |  |
| --- | --- |
| **DATE** | 26 OCTOBER 2023 |
| **TEAM ID** | 715 |
| **DOMAIN NAME** | DATA ANALYTICS |
| **PROJECT NAME** | PRODUCT SALES ANALYSIS |
| **MAXIMUM MARKS** |  |

**INTRODUCTION:**

The main objective of this phase is to clean and preprocess the data and will be used for predicting the future sales and customer preferences in product sales analysis.

**PROBLEM STATEMENT:**

Predict future sales performance for each product, region depending on top-selling products, peak sales analysis and customer preferences.

**CLEANING AND PREPROCESSING:**

Cleaning and preprocessing of data are done using Jupyter Notebook by importing necessary libraries.

Q-P1: Total unit sales of Product 1

Q-P2: Total unit sales of Product 2

Q-P3: Total unit sales of Product 3

Q-P4: Total unit sales of Product 4

S-P1: Total revenue from Product 1

S-P2: Total revenue from Product 2

S-P3: Total revenue from Product 3

S-P4: Total revenue from Product 4

**Code with Output:**

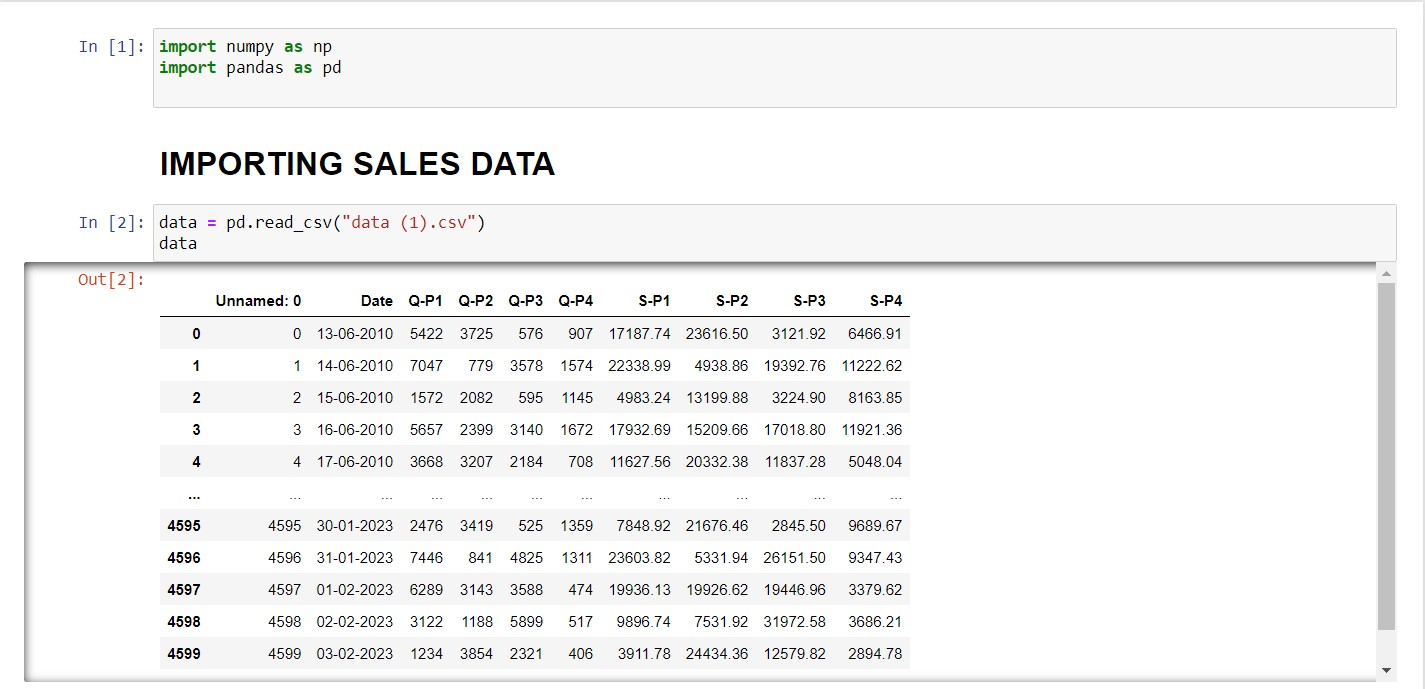
import numpy as np

import pandas as pd

**Import data**

data = pd.read\_csv("data.csv")

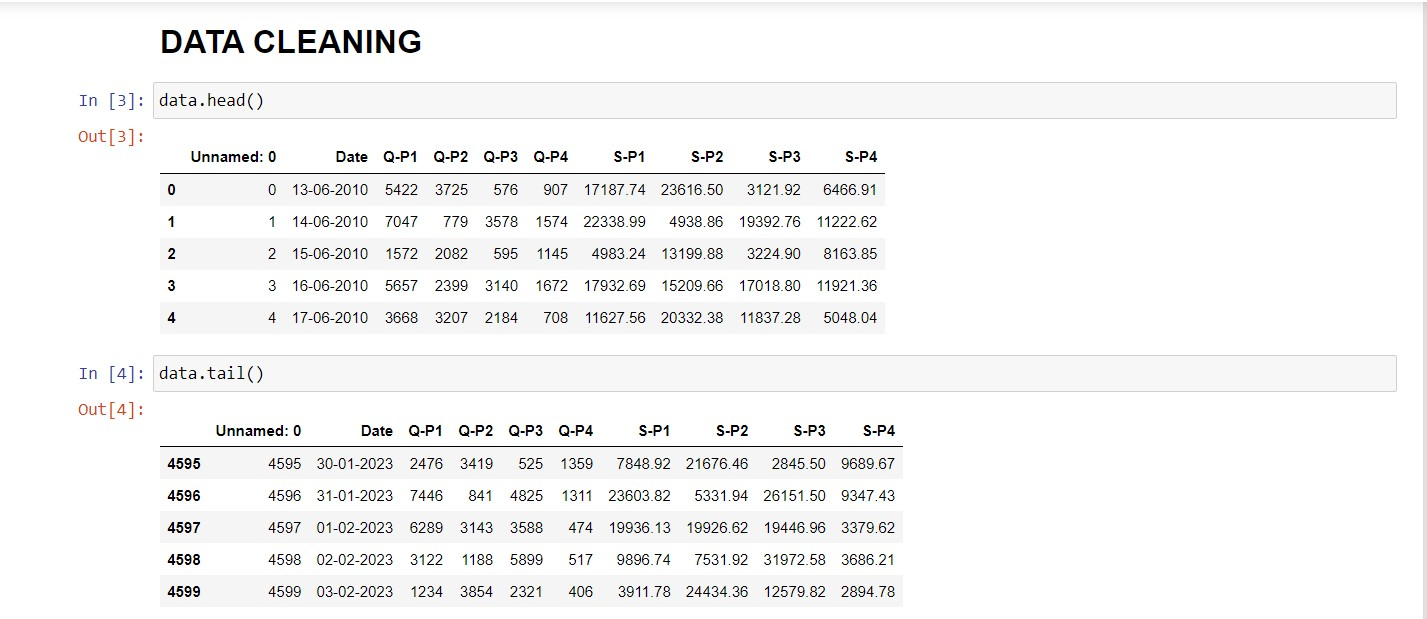
data



**Cleaning the data**

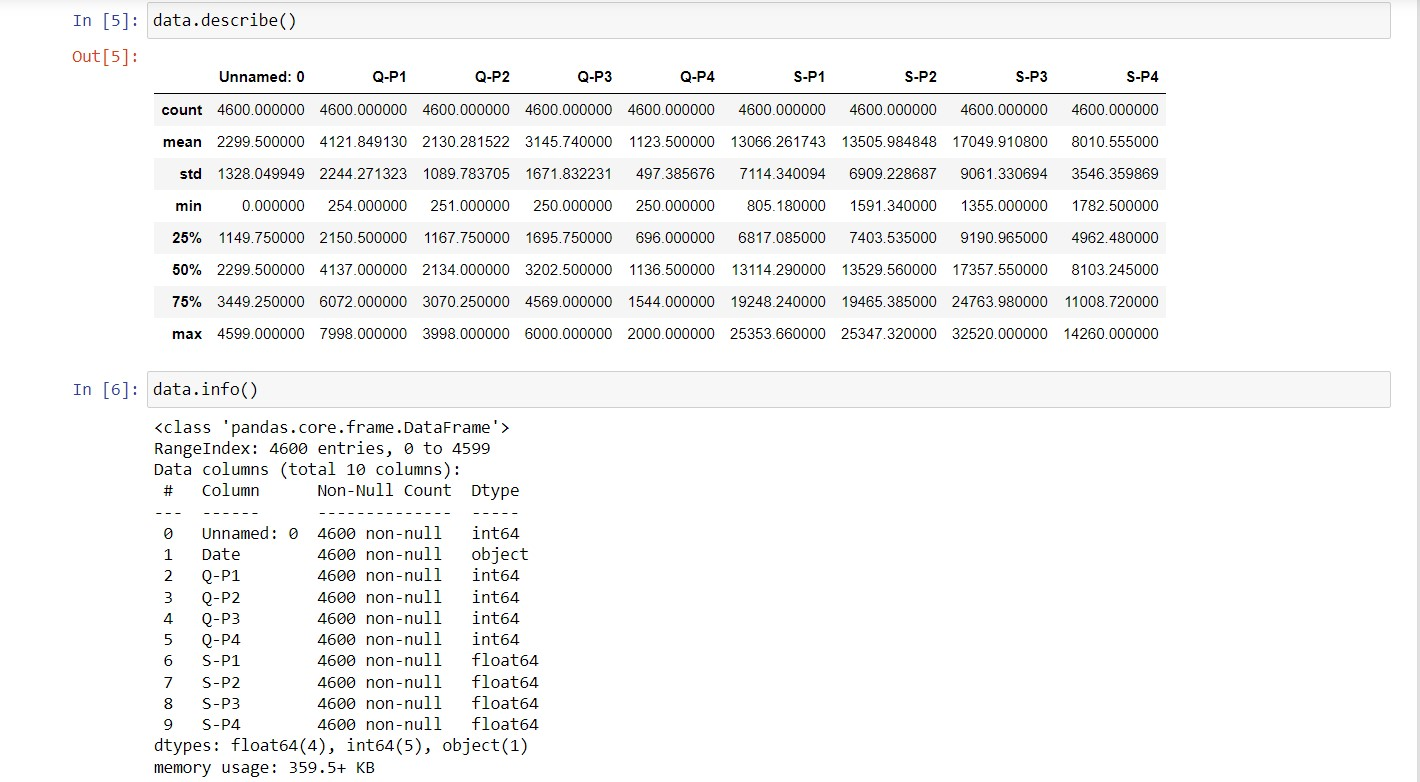
data.head()

data.tail()



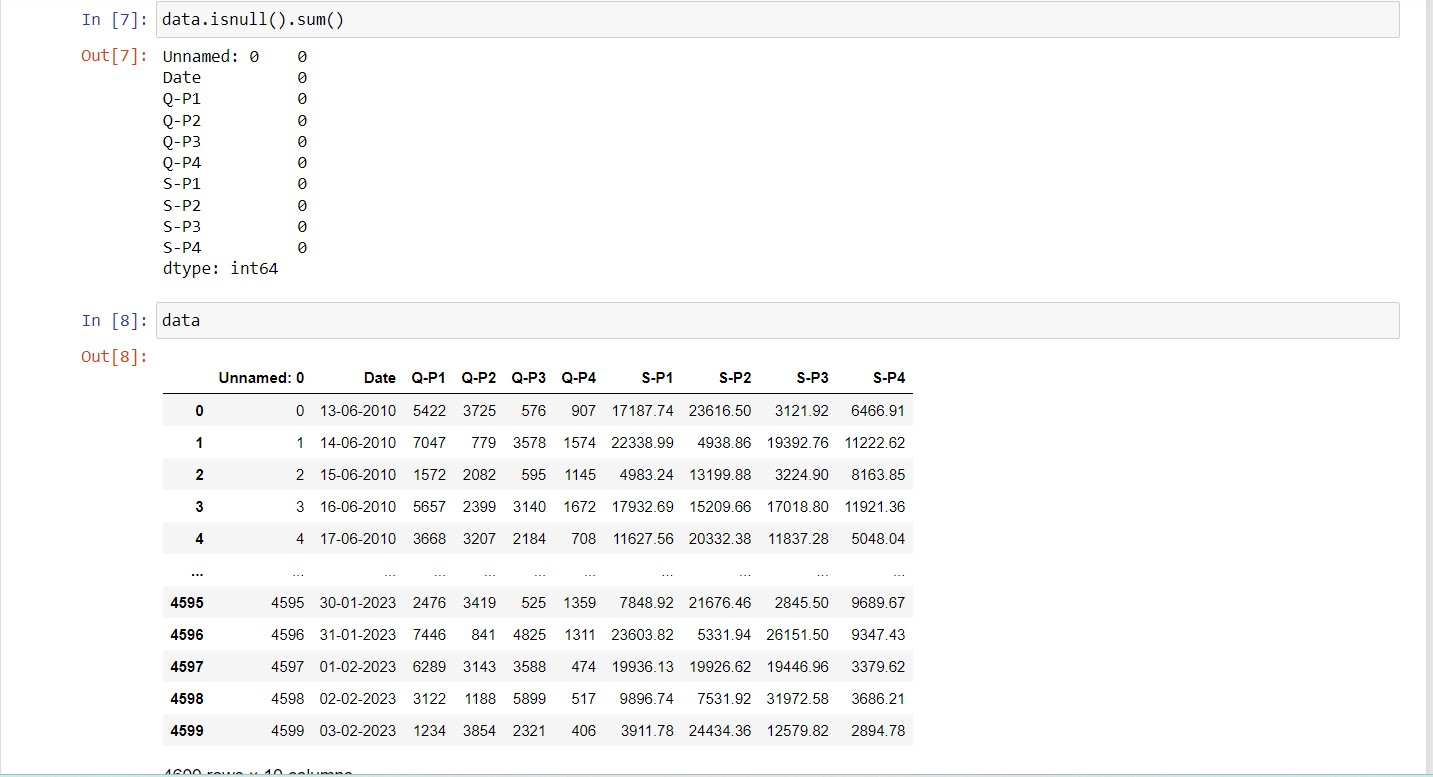
data.describe()

data.info()



data.isnull().sum()

data



data.dropna(how="any").shape

data.dropna(how="all").shape

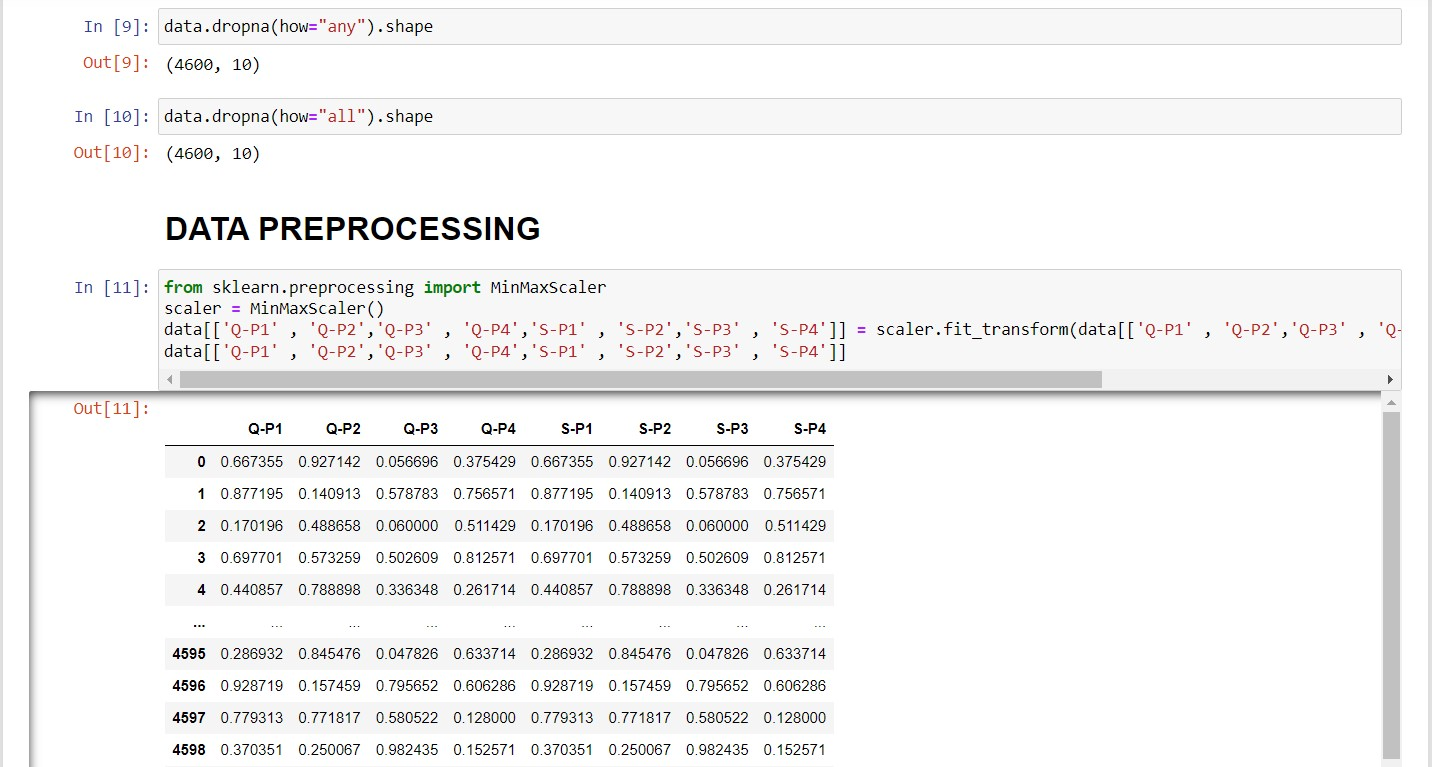
**Preprocessing of data**

from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()

data[['Q-P1' , 'Q-P2','Q-P3' , 'Q-P4','S-P1' , 'S-P2','S-P3' , 'S-P4']] = scaler.fit\_transform(data[['Q-P1' , 'Q-P2','Q-P3' , 'Q-P4','S-P1' , 'S-P2','S-P3' , 'S-P4']])

data[['Q-P1' , 'Q-P2','Q-P3' , 'Q-P4','S-P1' , 'S-P2','S-P3' , 'S-P4']]

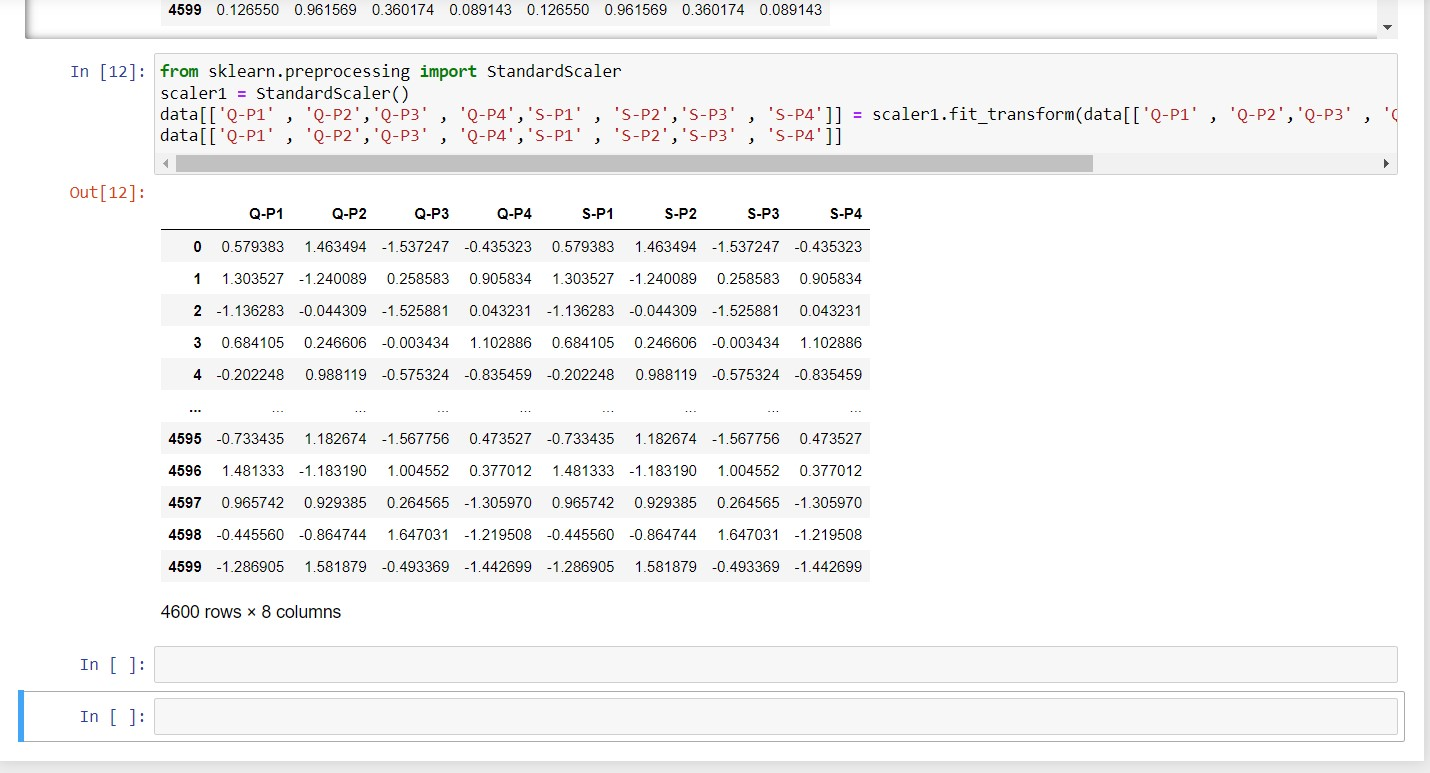


from sklearn.preprocessing import StandardScaler

scaler1 = StandardScaler()

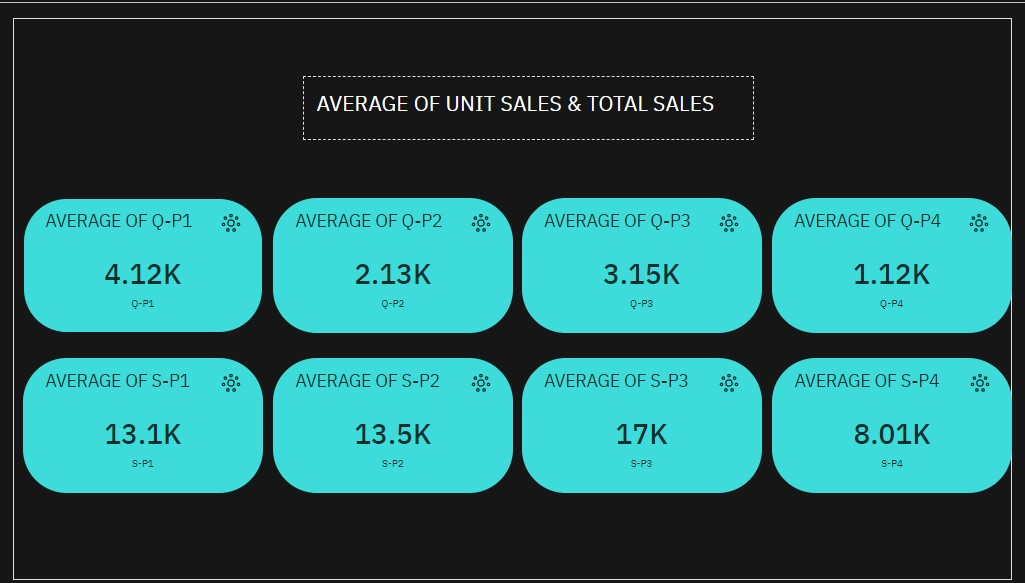
data[['Q-P1' , 'Q-P2','Q-P3' , 'Q-P4','S-P1' , 'S-P2','S-P3' , 'S-P4']] = scaler1.fit\_transform(data[['Q-P1' , 'Q-P2','Q-P3' , 'Q-P4','S-P1' , 'S-P2','S-P3' , 'S-P4']])

data[['Q-P1' , 'Q-P2','Q-P3' , 'Q-P4','S-P1' , 'S-P2','S-P3' , 'S-P4']]

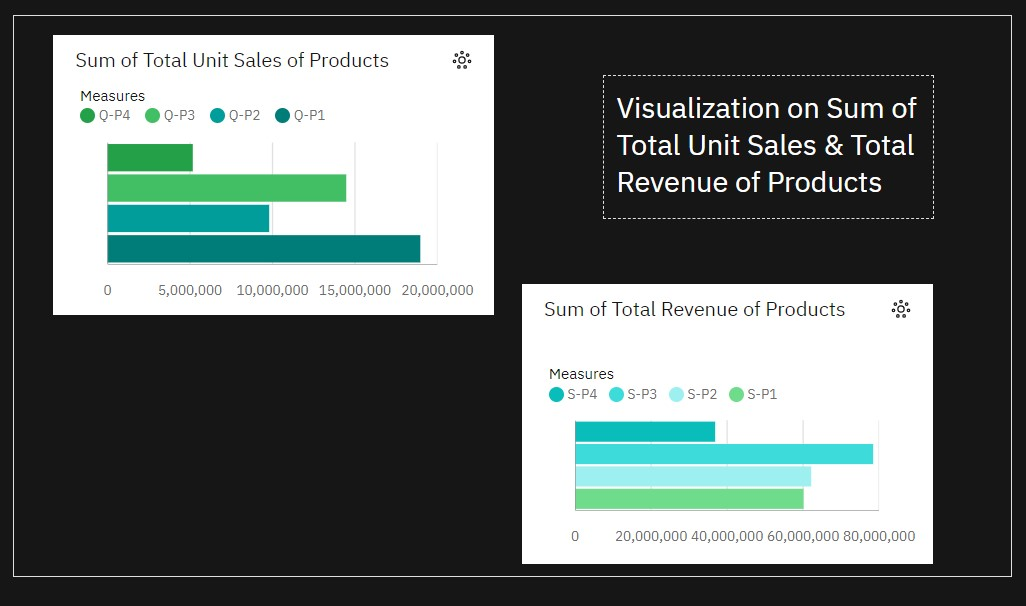


**VISUALIZATION OF DATA USING COGNOS:**

Calculating the average of total unit sales(Q-P1, Q-P2, Q-P3, Q-P4) and total revenue of products(S-P1,S-P2,S-P3,S-P4):



Visualizing the sum of unit sales and total revenue of products:



Comparing each unit sales and its revenue by number:

