**PHASE - 3**

**Development Part 1**

|  |  |
| --- | --- |
| **Date** | **25-10-2023** |
| **Team ID** | **8941** |
| **Project Name** | **PRODUCT SALES ANALYSIS** |
| **Team Name** | **Proj\_207143\_Team\_1** |

**ANALYTICS OBJECTIVES**

**Data Preprocessing :**

1.Data Inspection

2.Data Cleaning

3.Data Transformation

4.Data Splitting

5.Data Normalization

6.Data Validation

7.Data Visualization

8.Data Collectio

**1.Loading Data :**

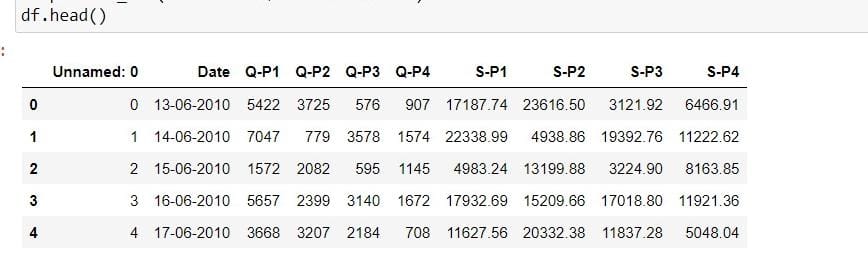
Use pandas.read\_csv() to load data from a CSV file.

Use pandas.read\_excel() for Excel files



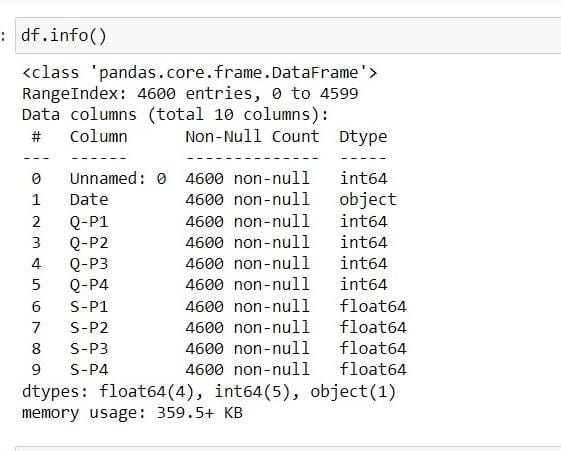
**2. Exploring Data:**

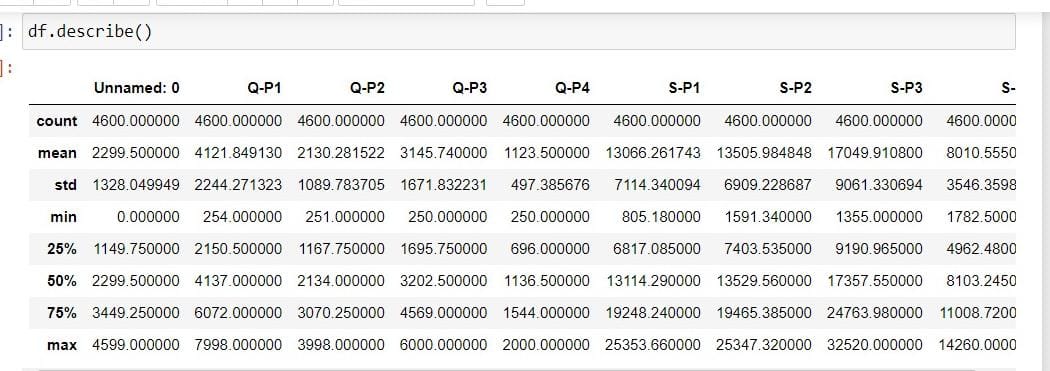
Use df.head() to view the first few rows of the dataset



Use df.info() to get information about data types and missing values.

Use df.describe() for summary statistics.python

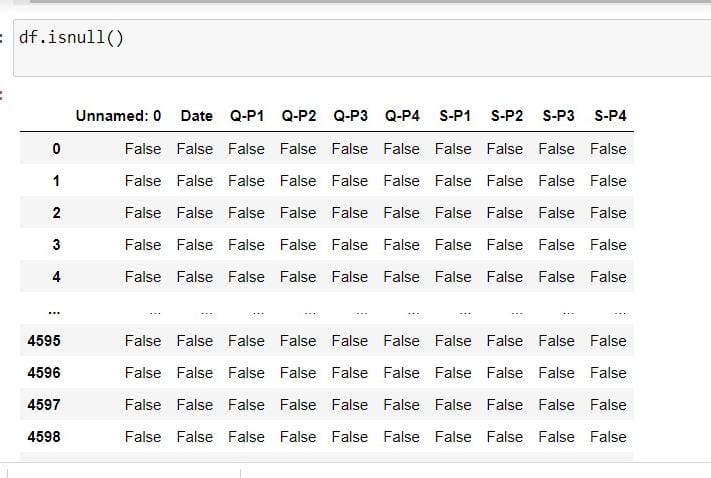




**3.Handling Missing Values:**

Use df.isnull() to identify missing values.

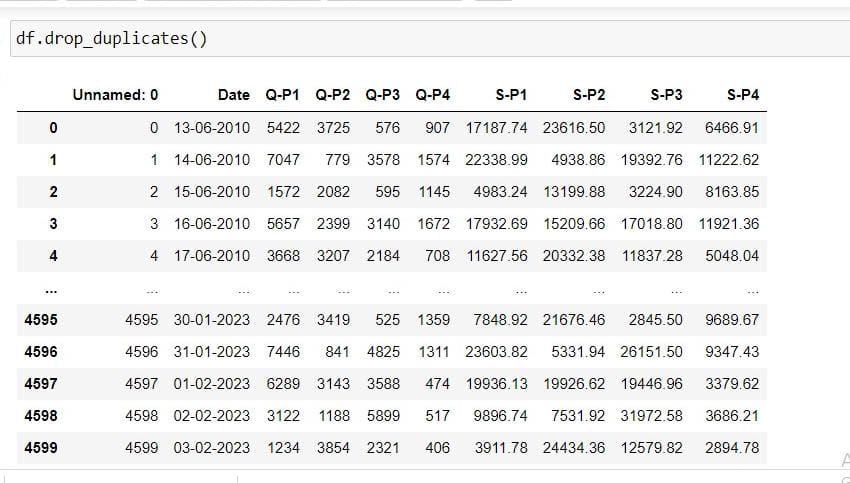
Use df.fillna() or df.dropna() to handle missing values



**4.Data Cleaning:**

Remove duplicate rows with df.drop\_duplicates().

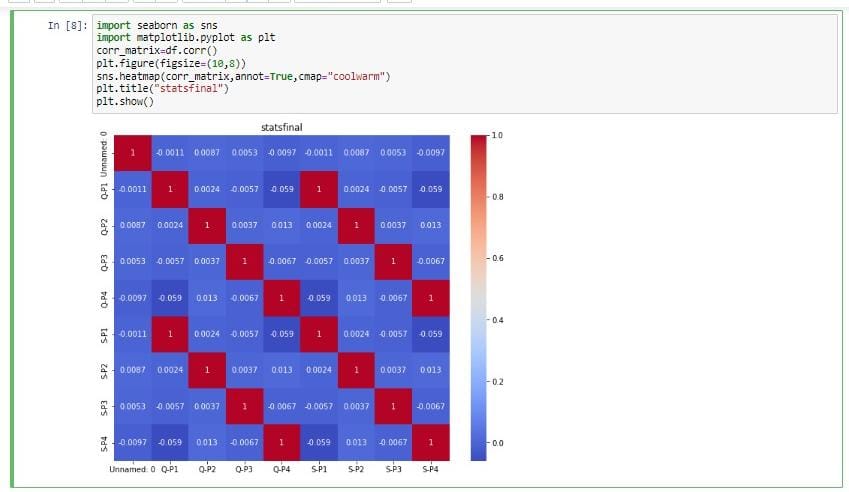
Rename columns using df.rename() if necessary.Convert data types with df.astype()



**5.Handling Outliers:**

Detect and deal with outliers using statistical methods or visualization.

You can use techniques like zscores or IQR (Interquartile Range).



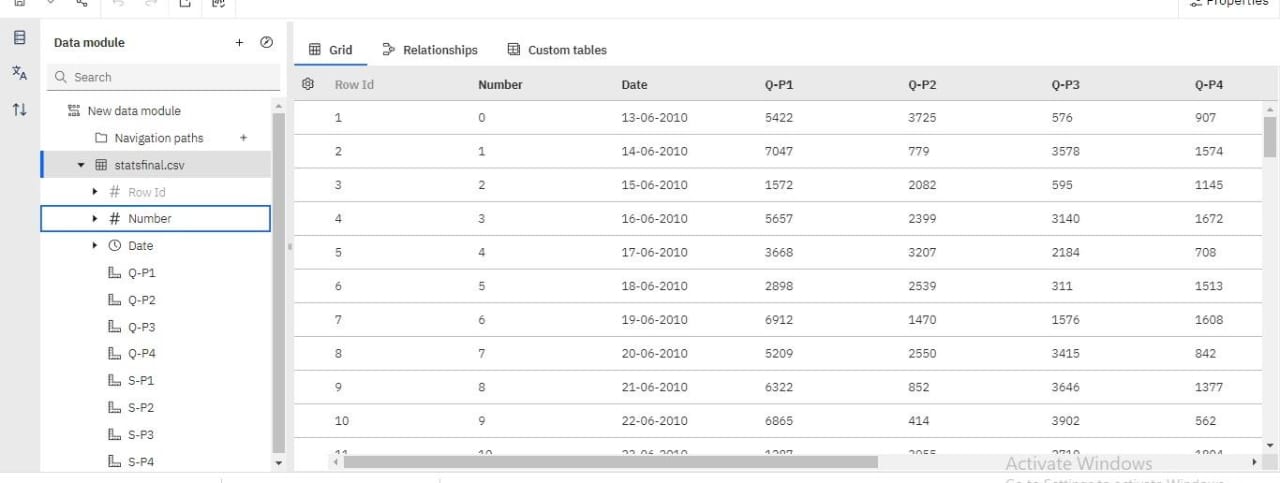
**6.Saving Data:**

Save the preprocessed data back to a file if needed.

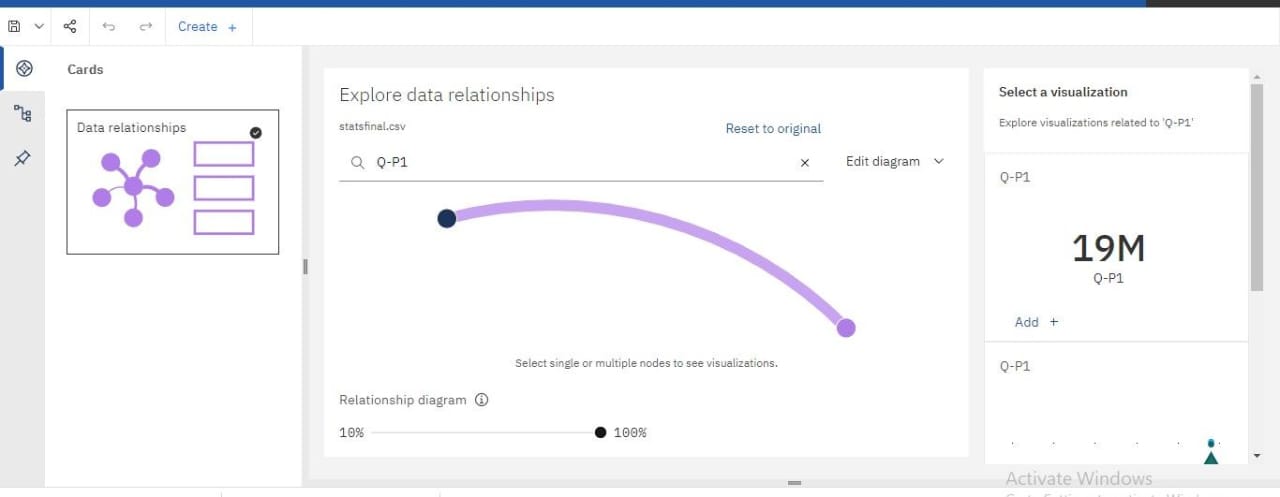


**IBM COGNOS ANALYTICS**

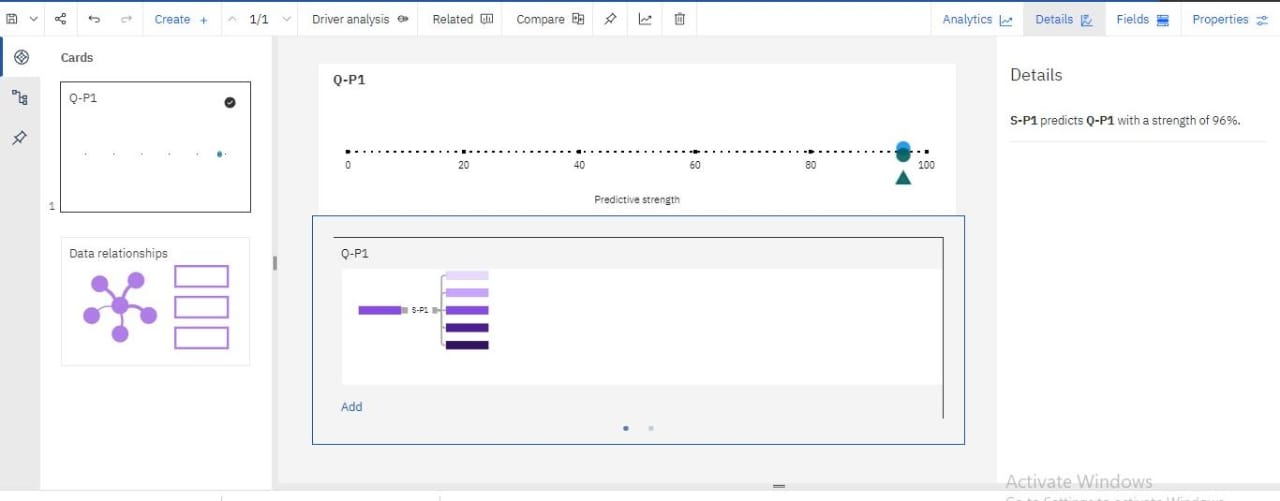
**IN COGNOS - DATA LOADING**



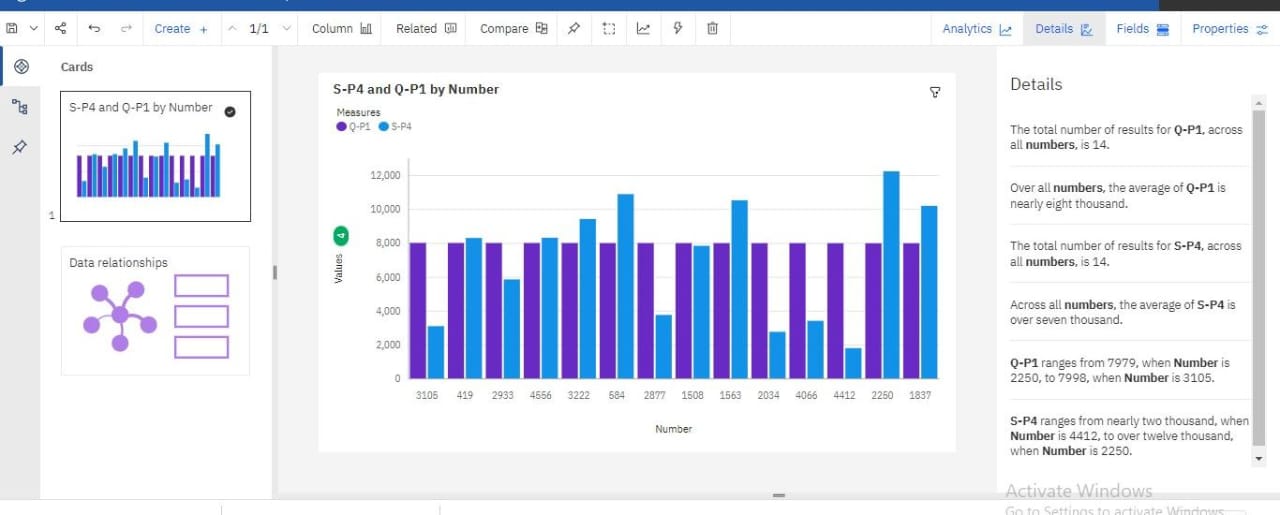
**IN COGNOS - DATA RELATIONSHIP**



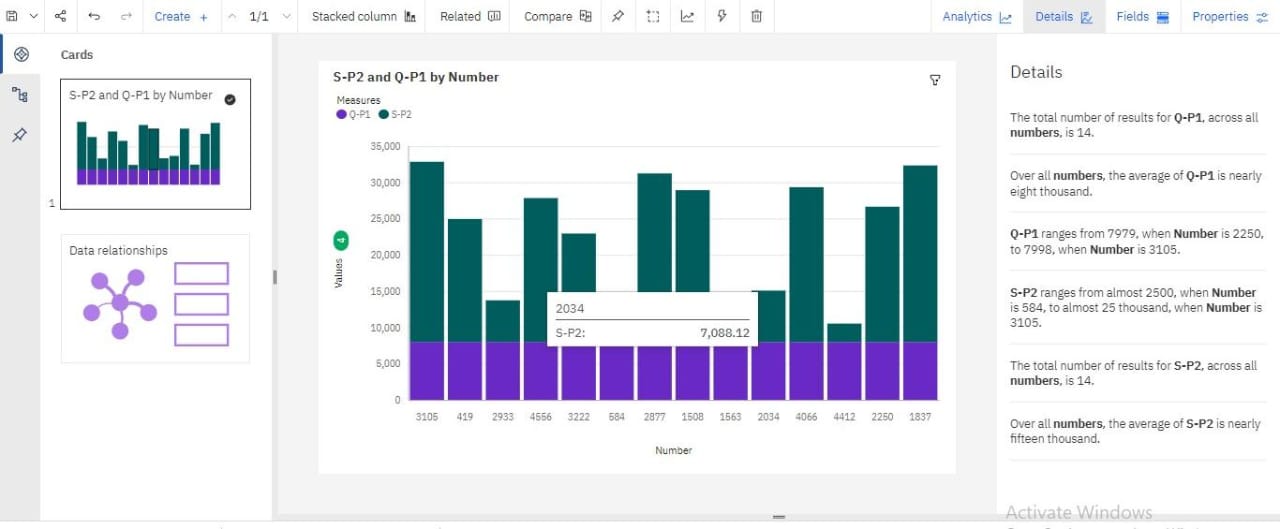
**Q-P1**



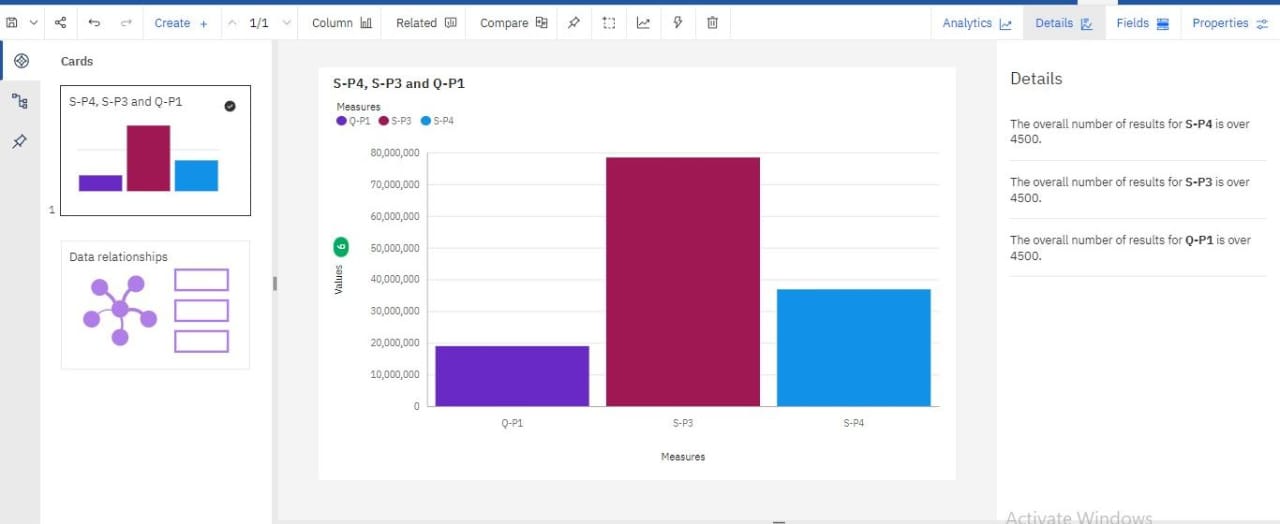
**S-P4 andQ-P1 by Number**



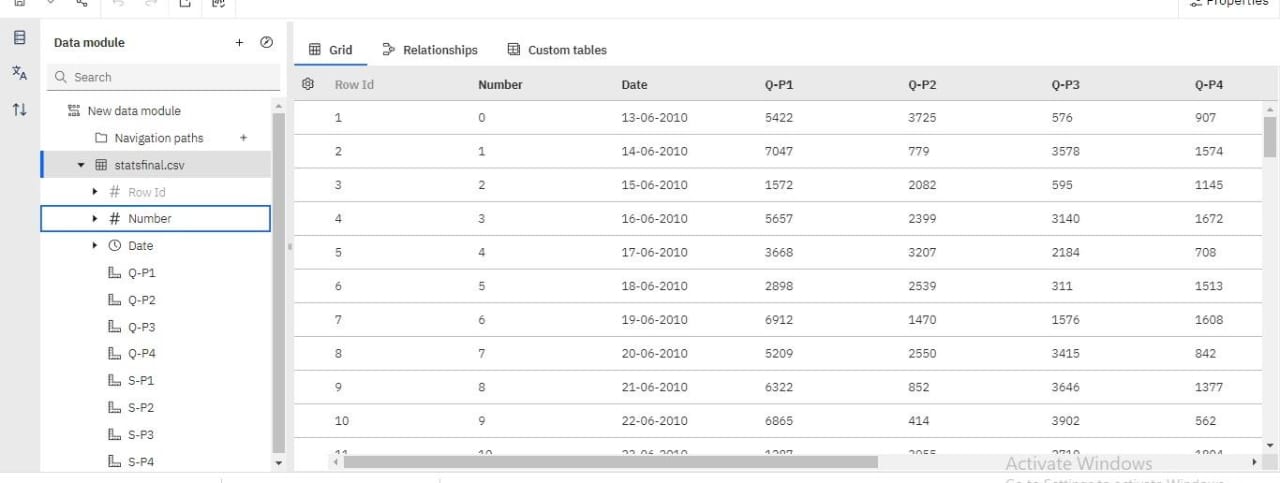
**S-P2 and Q-P1 by Number**



**S-P4,5-P3 and Q-P1**



**VIEW OF TABLE**

****

CONCLUSION

Data preprocessing is a crucial step in preparing data for analysis and machine learning. It involves collecting, inspecting, cleaning, transforming, and organizing data. The main steps include data collection, inspection, cleaning, transformation, splitting, normalization, and validation. It ensures data is ready for analysis and model training, improving the success of data-related projects