**FINAL SUBMISSION**

**IBM-Naan Muthalvan Data Analytics with Cognos**

**Student Name : Selladurai J**

**Register Number : 620821104103**

**Branch : B.E CSE**

**Year : 3rd year**

**Topic : Data Analytics with Cognos**

**Title : Product Sales Analysis**

**College : Gnanamani College of**

**Technology**

**Introduction:**

In today's highly competitive business landscape, data-driven decision-making has become paramount for organizations aiming to thrive and succeed. Understanding consumer behavior, analyzing sales trends, and identifying top-selling products are critical components of any successful business strategy. This is where Product Sales Analysis, driven by the principles of Design Thinking, comes into play. By adopting a systematic approach, businesses can extract valuable insights from sales data, thereby making informed decisions to enhance their operations and better serve their customers.

**Abstract:**

This document presents a comprehensive framework for Product Sales Analysis using Design Thinking principles. The primary objective of this analysis is to extract specific insights from sales data, including the identification of top-selling products, the analysis of sales trends, and a deep understanding of customer preferences.

The process begins with a focus on data collection, where we delineate the sources and methods for gathering pertinent sales data. This encompasses transaction records, product information, and customer demographics, providing a holistic view of the sales ecosystem.

Subsequently, we delve into the Visualization Strategy, where we outline how to leverage IBM Cognos to create interactive and insightful dashboards and reports. Effective visualization is the bridge that connects data to actionable insights, empowering organizations to make data-driven decisions.

The ultimate goal of this analysis is to derive actionable insights that can guide critical aspects of business operations. These insights will be instrumental in streamlining inventory management, ensuring optimal stock levels, and refining marketing strategies to target the right audience effectively. By employing Design Thinking principles, we foster a human-centered approach to data analysis, ensuring that the end results align with the needs and aspirations of both the business and its customers.

I. **Analysis Objectives**

A. Define specific insights to extract

1. Identifying top-selling products

2. Analyzing sales trends

3. Understanding customer preferences

II. **Data Collection**

A. Sources of data

1. Transaction records

2. Product information

3. Customer demographics

B. Methods for data collection

III. **Visualization Strategy**

A. Utilizing IBM Cognos

B. Creating interactive dashboards and reports

C. Planning effective visualization techniques

IV. **Actionable Insights**

A. Identifying how derived insights can guide

1. Inventory management

2. Marketing strategies

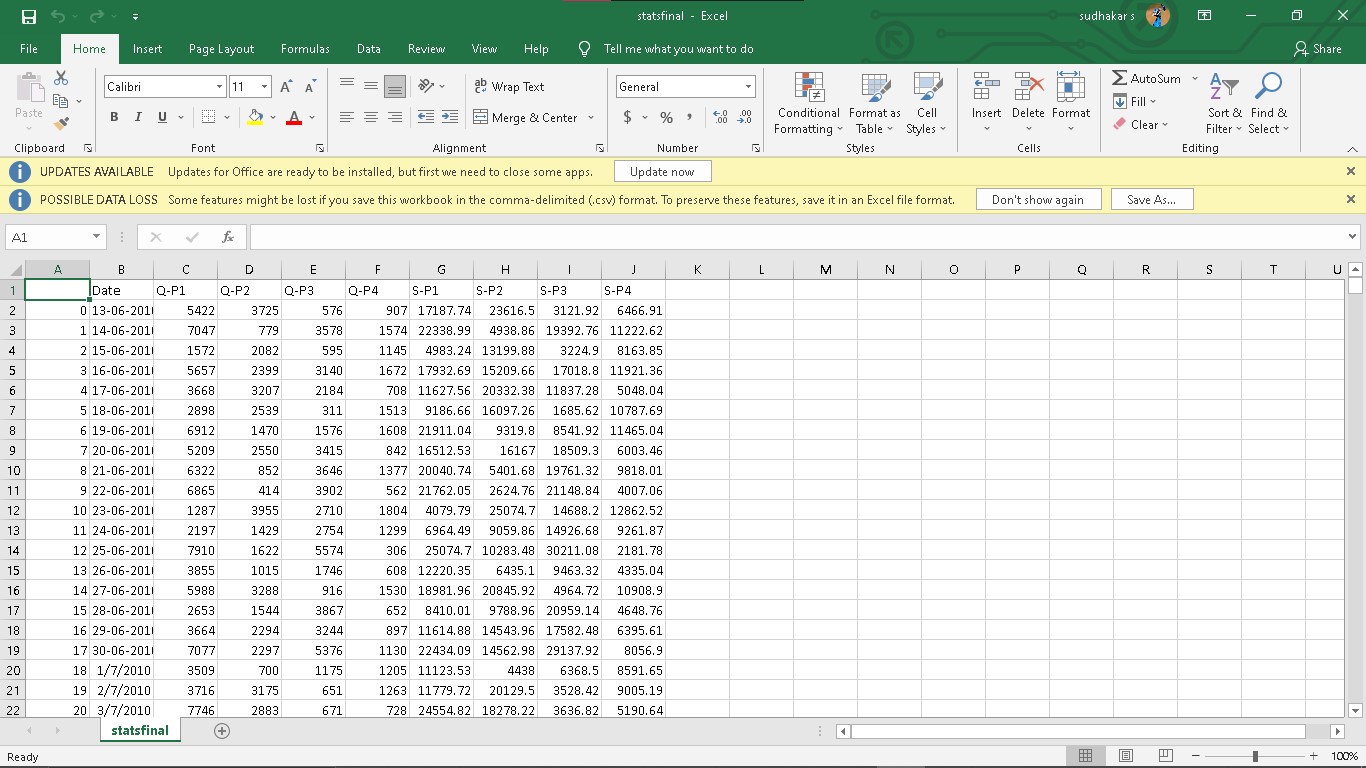
B. Implementing recommendations based on insights

1. **Data Collection and Integration**:
   * Begin by collecting relevant data from various sources, including sales data, customer data, product data, and external sources like market trends, economic indicators, and social media sentiment.
   * Integrate and preprocess the data to ensure it is clean, consistent, and ready for analysis. This may involve data cleaning, transformation, and handling missing values.
2. **Feature Engineering**:
   * Identify and engineer meaningful features from the data that can help improve the predictive model. These features can include historical sales, product attributes, customer demographics, and more.
   * Consider using domain knowledge and creativity to extract valuable insights from the data.
3. **Selection of Machine Learning Algorithms**:
   * Choose appropriate machine learning algorithms for your specific sales and product analysis tasks. Common choices include regression, time series analysis, and classification algorithms.
   * For sales forecasting, time series forecasting methods like ARIMA, LSTM, or Prophet can be particularly effective.
4. **Model Training**:
   * Split your data into training, validation, and test sets to train and evaluate your machine learning models.
   * Experiment with different algorithms, hyperparameters, and feature combinations to find the best-performing model.
5. **Ensemble Methods**:
   * Consider using ensemble methods such as random forests, gradient boosting, or stacking to improve the accuracy of your predictive models.
   * Ensembles combine the predictions of multiple models to reduce overfitting and enhance generalization.
6. **Hyperparameter Tuning**:
   * Optimize hyperparameters of your machine learning models through techniques like grid search, random search, or Bayesian optimization.
   * This step helps fine-tune the models for the best performance.
7. **Regularization Techniques**:
   * Implement regularization techniques such as L1 or L2 regularization to prevent overfitting, especially when dealing with complex models.
8. **Validation and Testing**:
   * Validate your models using cross-validation techniques to ensure they generalize well to unseen data.
   * Evaluate model performance using appropriate metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or accuracy, depending on the specific analysis.
9. **Continuous Learning**:
   * Set up a system for continuous learning and model retraining. As new data becomes available, periodically retrain your models to keep them up to date and accurate.
10. **Visualization and Interpretation**:
    * Visualize the results of your analysis using charts, graphs, and dashboards to make insights more accessible to decision-makers.
    * Interpret the model's predictions and understand the key factors driving sales and product performance.
11. **Deployment**:
    * Deploy your machine learning models in a production environment so that they can be used for real-time predictions or integrated into business processes.
12. **Monitoring and Maintenance**:
    * Continuously monitor the performance of your deployed models and ensure they are performing as expected.
    * Update and retrain models as needed to adapt to changing market conditions and business goals.

**Data source**

Dataset is collected from the kaggle.com named “daily-website-visitors.csv” which has a data about the Days, Day of week, Date, page Loads, Unique visits, First-time visits, Returning Visits.

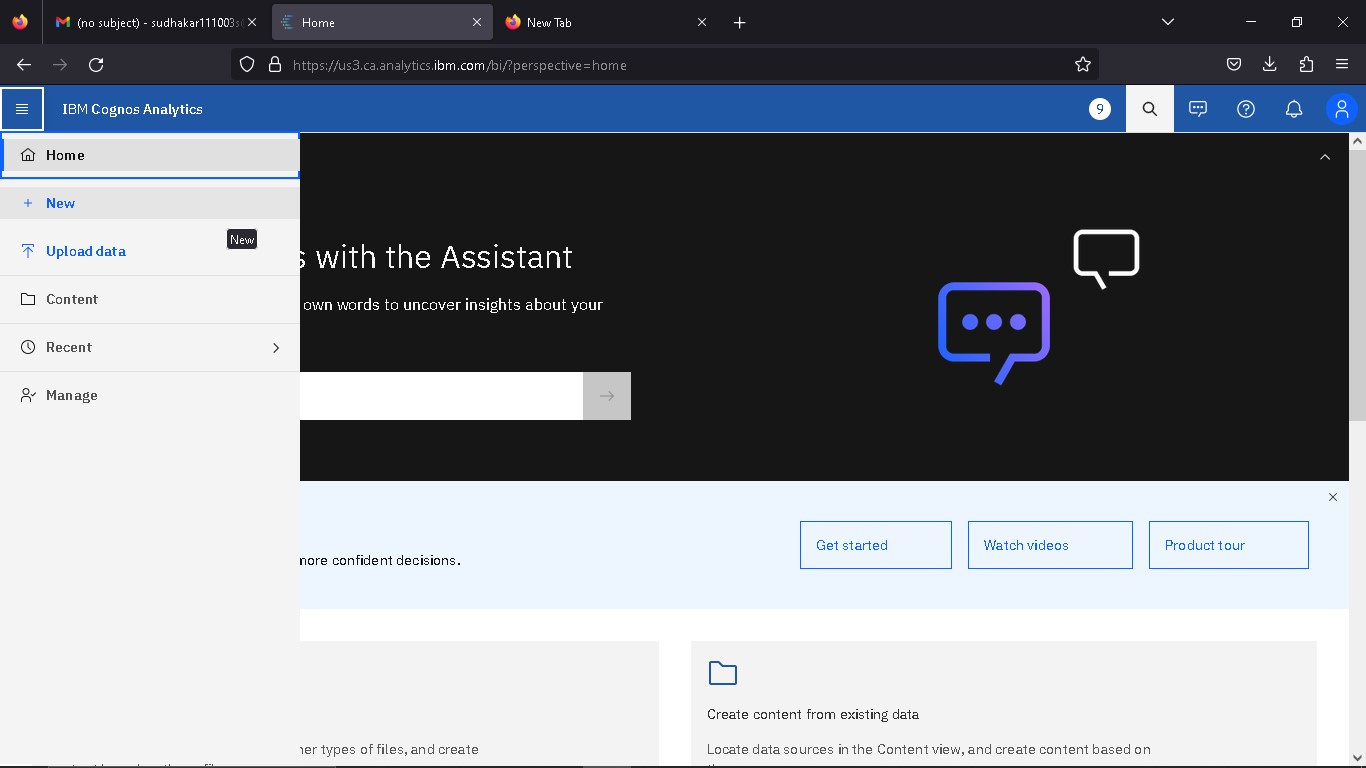
Dataset link : <https://www.kaggle.com/datasets/ksabishek/product-sales-data>

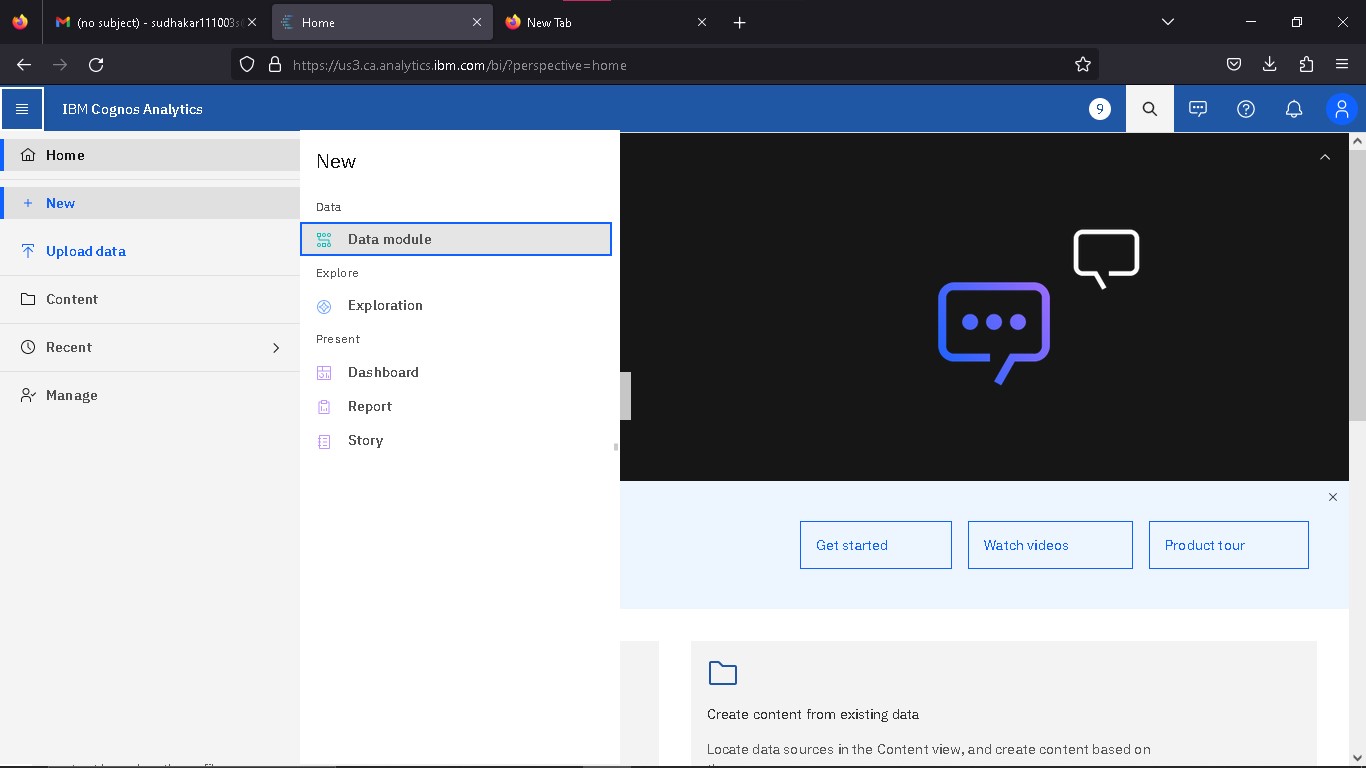


**Data Loading**

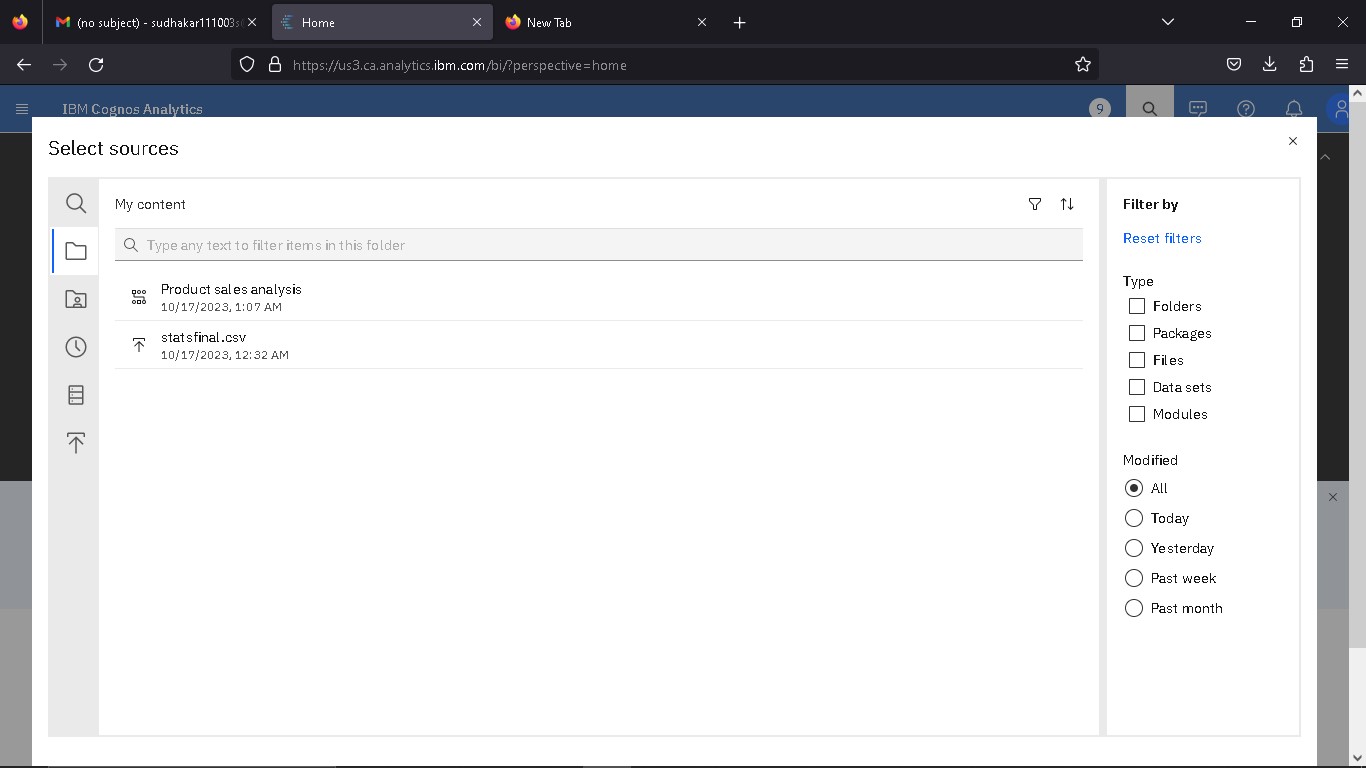
Steps Involved in data loading on IBM cognos.

1. **Login to your IBM cognos**
2. **Click more menu from the left side**
3. **Select new tab**

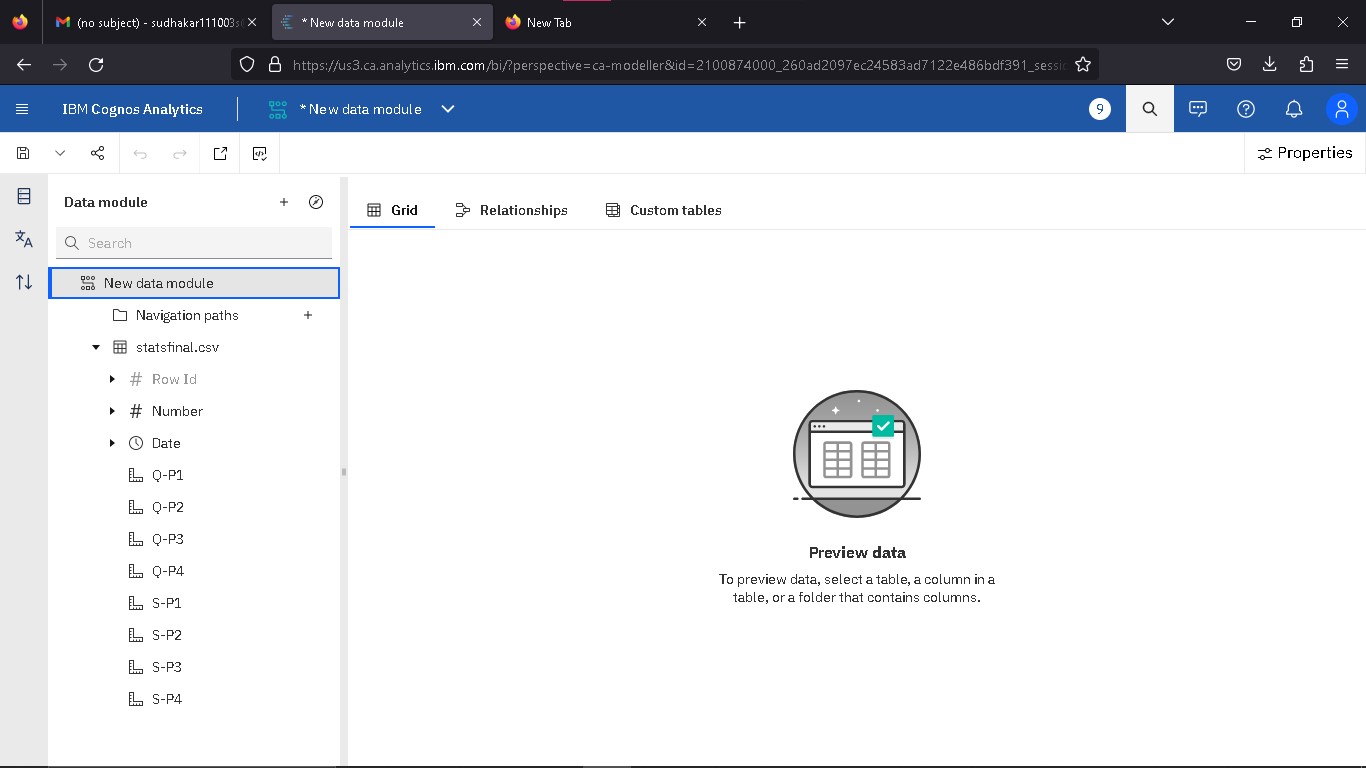


**4. Click Data module tab**

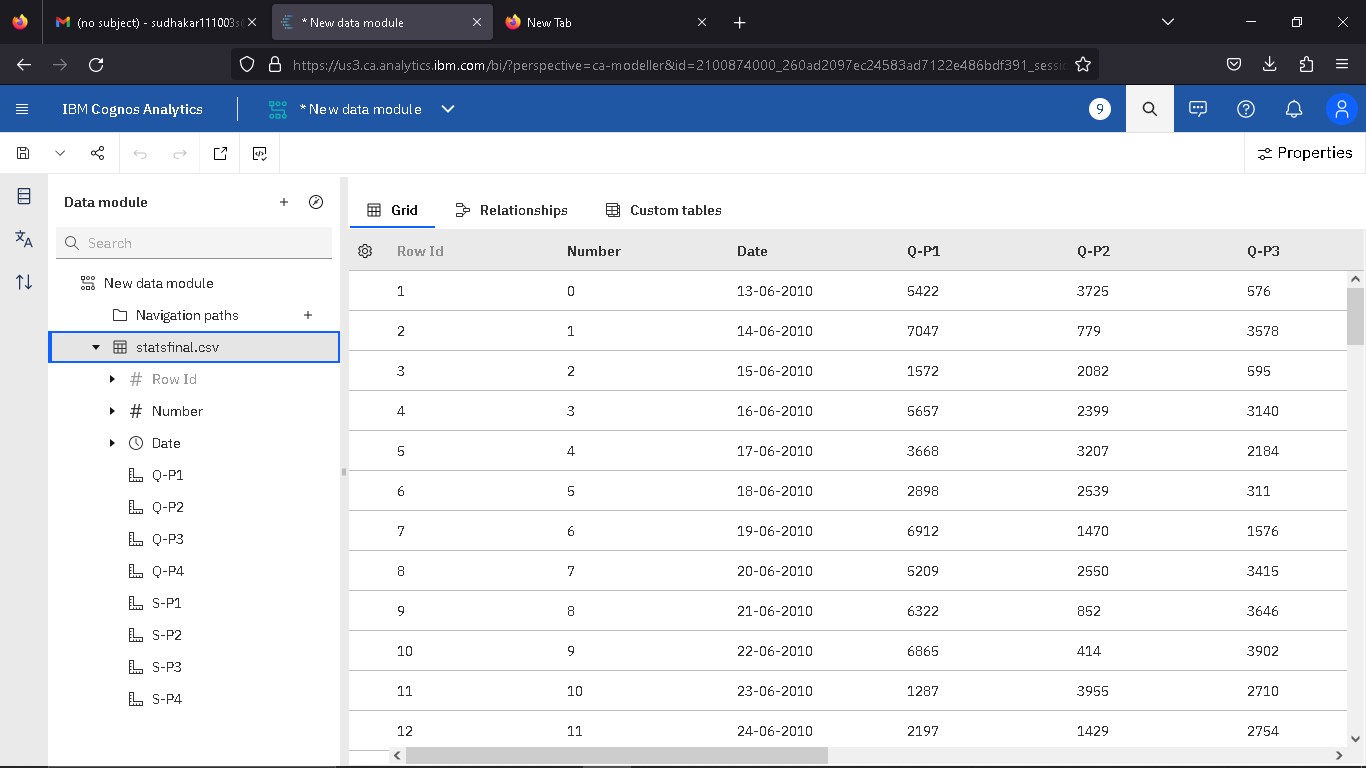
**5.Upload the dataset for your project and select the Corresponding file**



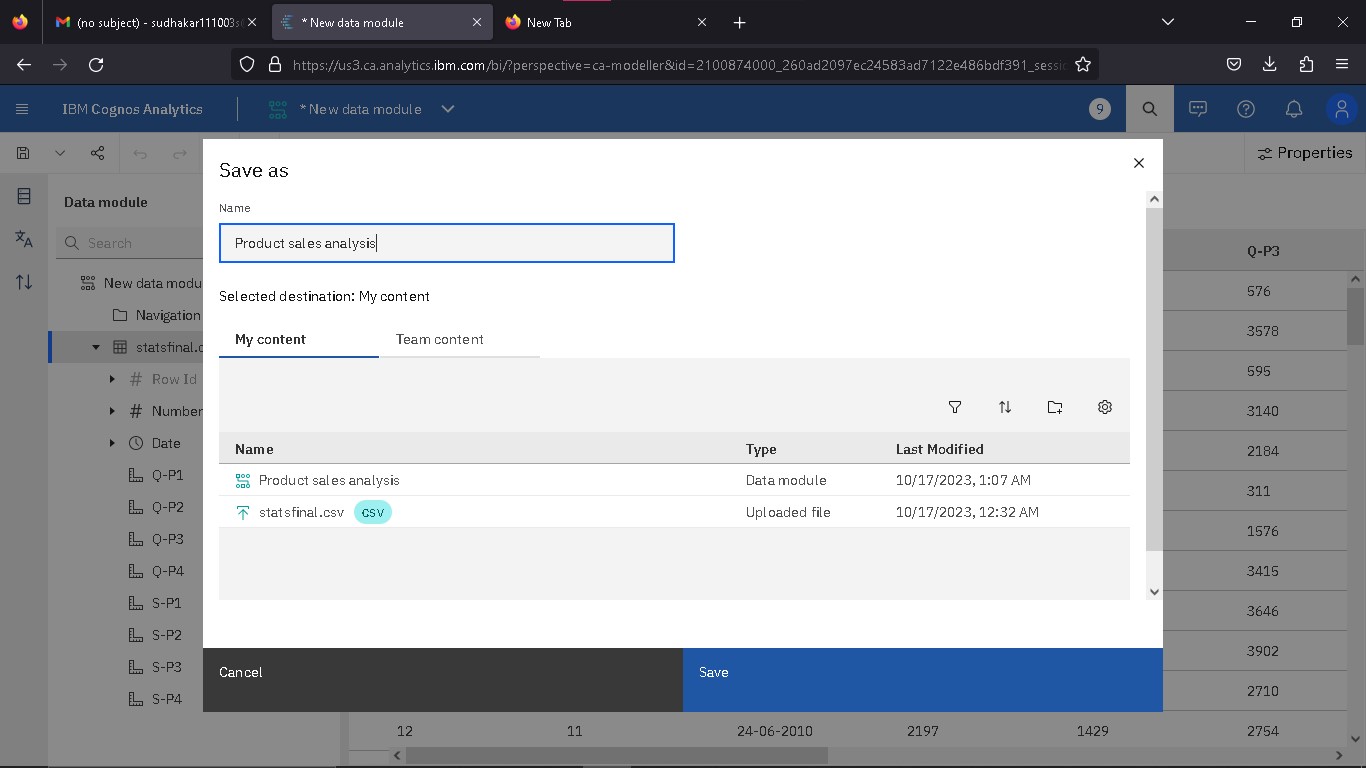
**6. preview the data**



**7.Explore the data**



**8. save the data module**



**Data Preprocessing and Cleaning**

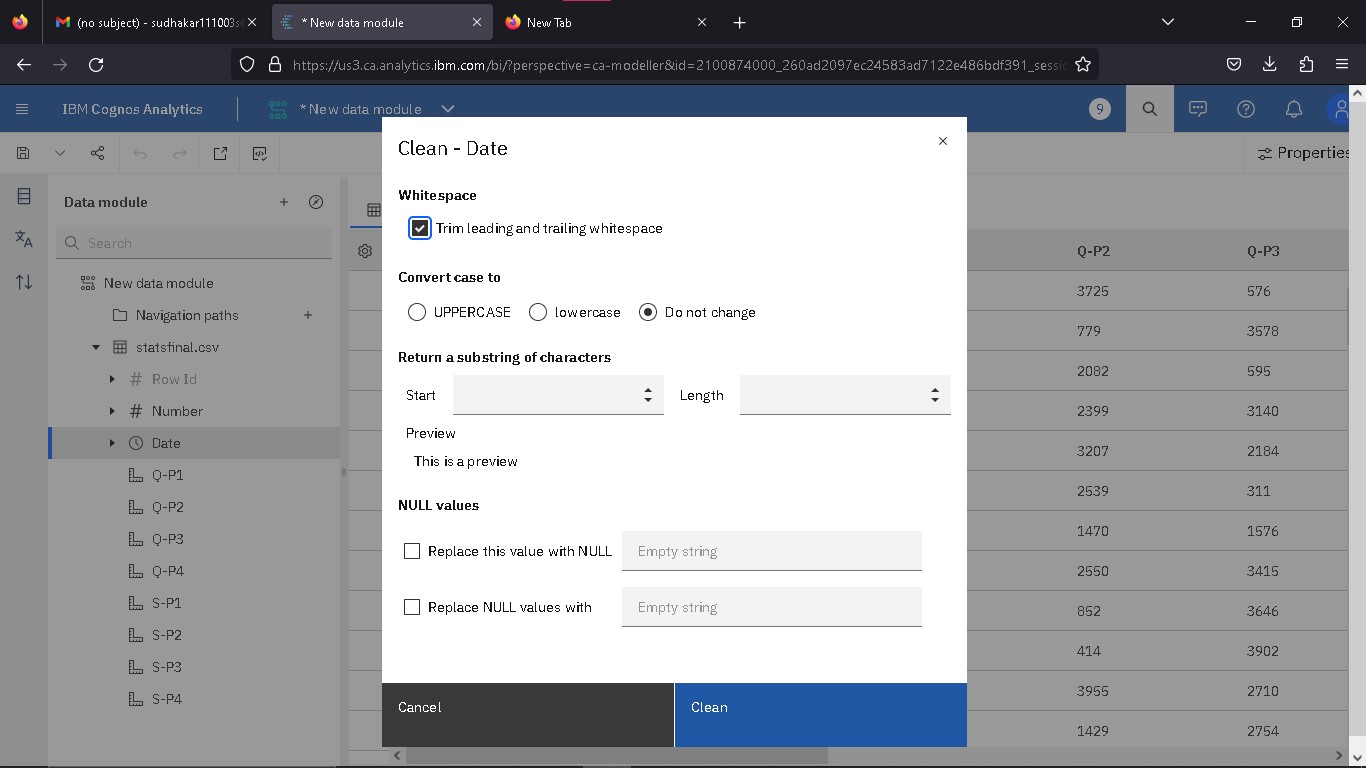
In this phase the following steps will taken

* Handling missing data
* Data Transformation
* Data Type Conversion
* Removing Duplicates
* Dealing Outliers

Once you saved the data module. Click the corresponding dataset on IBM cognos and Preview the mosule

Right Click the row where you want to clean the data

It provides the UI to Clean the data and makes the task easy one, Now Updating and Replacing the Null values are simple



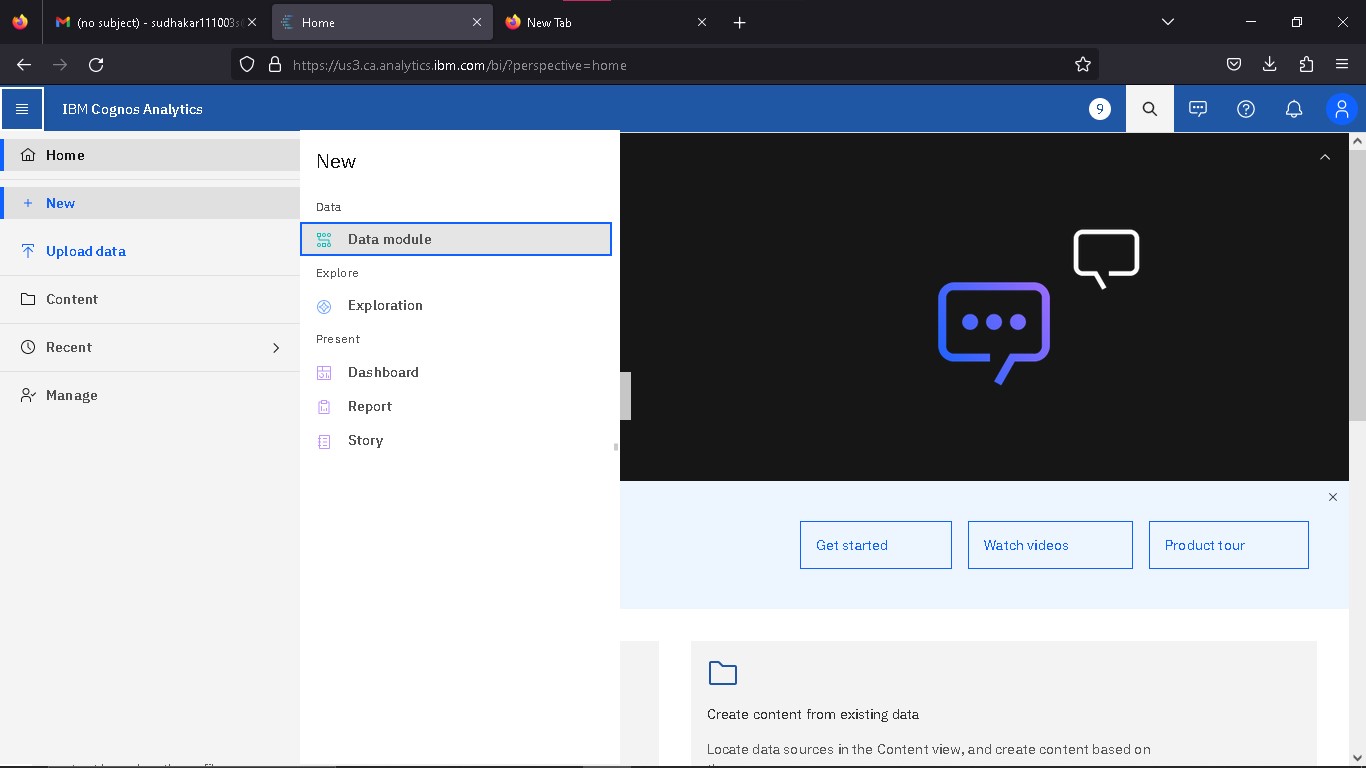
data module will be updated by doing the above process

after the completion of process start creating the dashboard for Visualization

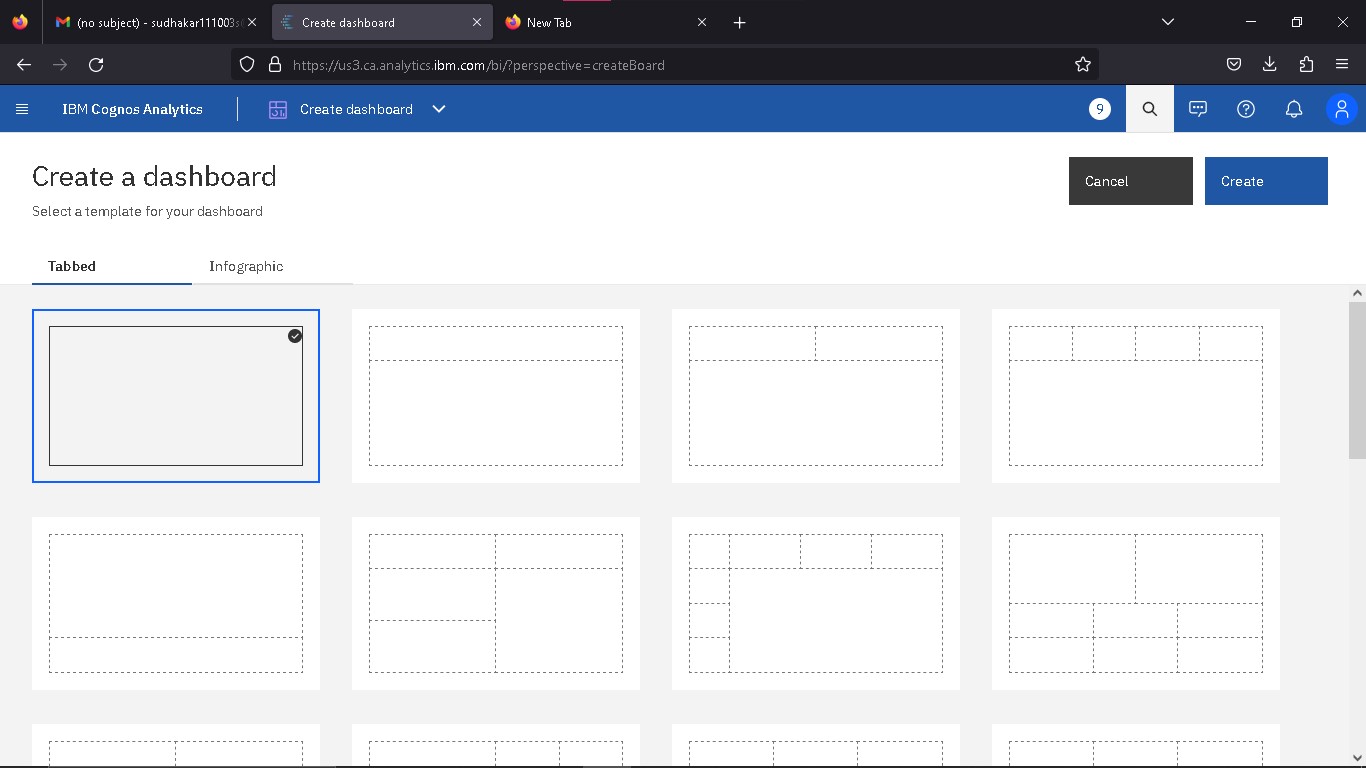
**Dashboard Creation**

Dashboard creation are helpful to visualizing the data

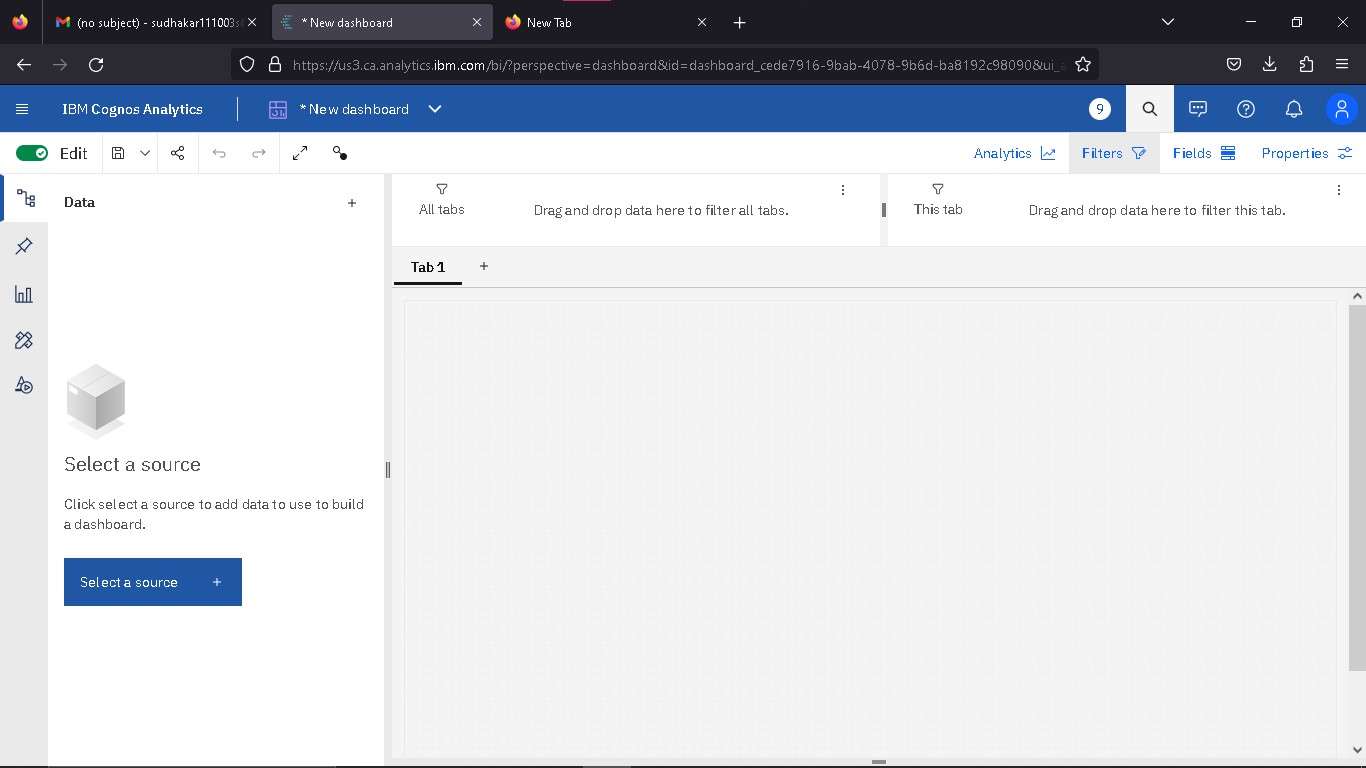
1. Goto Home menu
2. Select the new tab
3. Click dashboard



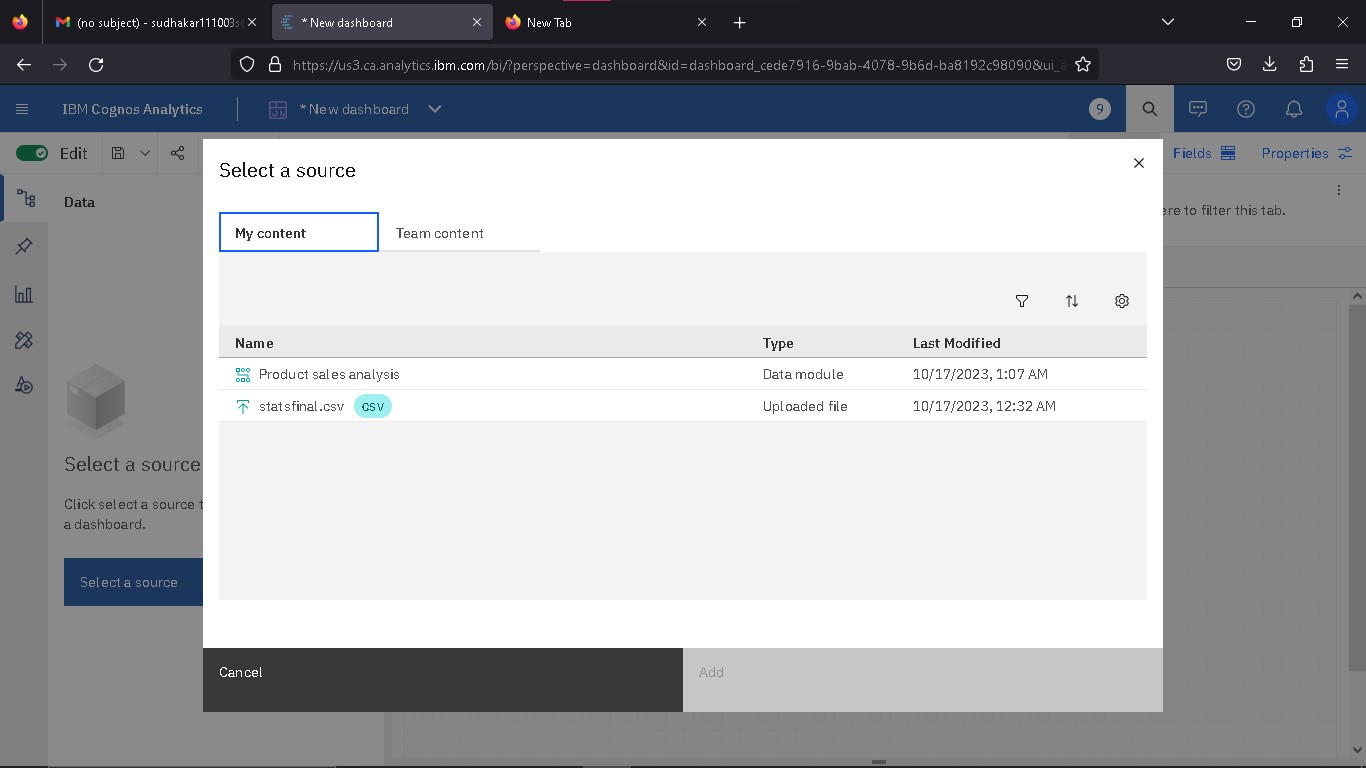
**4. Choose the template for your project and click**



**5.Now Dashboard is created**



**6. Select the data source**

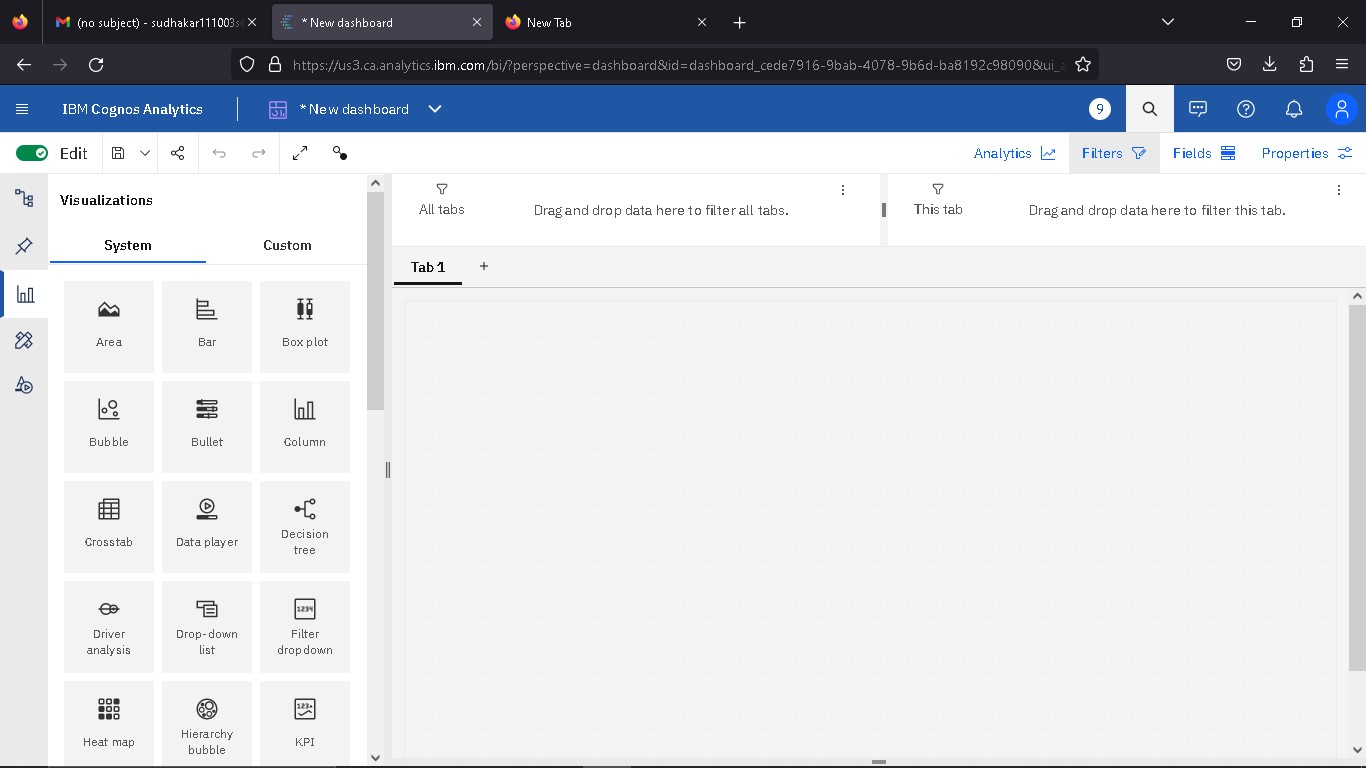


**Visualization :** After creating the dashboard, the next step is to visualize the data

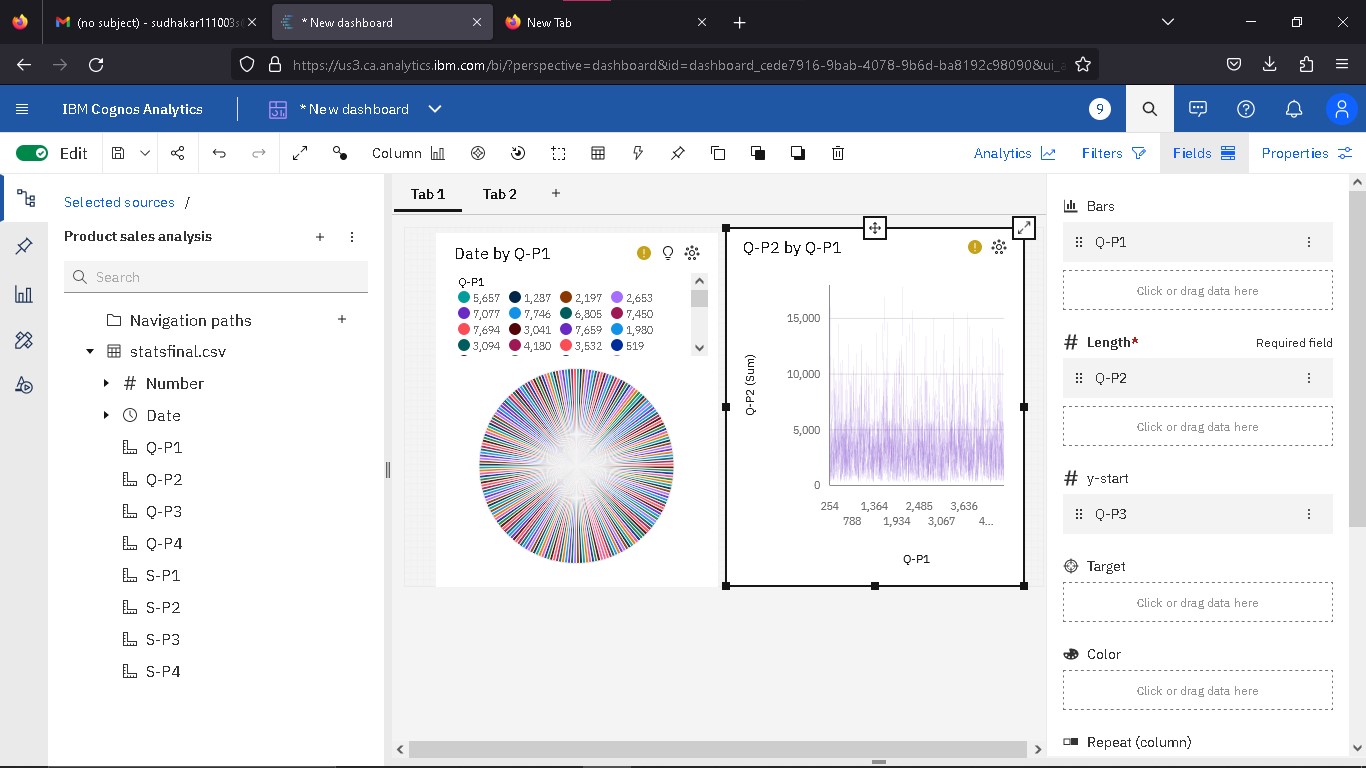
In IBM Cognos

1. Goes to the Corresponding Dashboard

2. select the visualizations tab in the left side of title bar



**3.Choose the system as you want and put the data source for the required columns**



In the above screen shot displays the Pie chart and model compares the “Date” and “Q-P1” .

X-axis =Q-P1

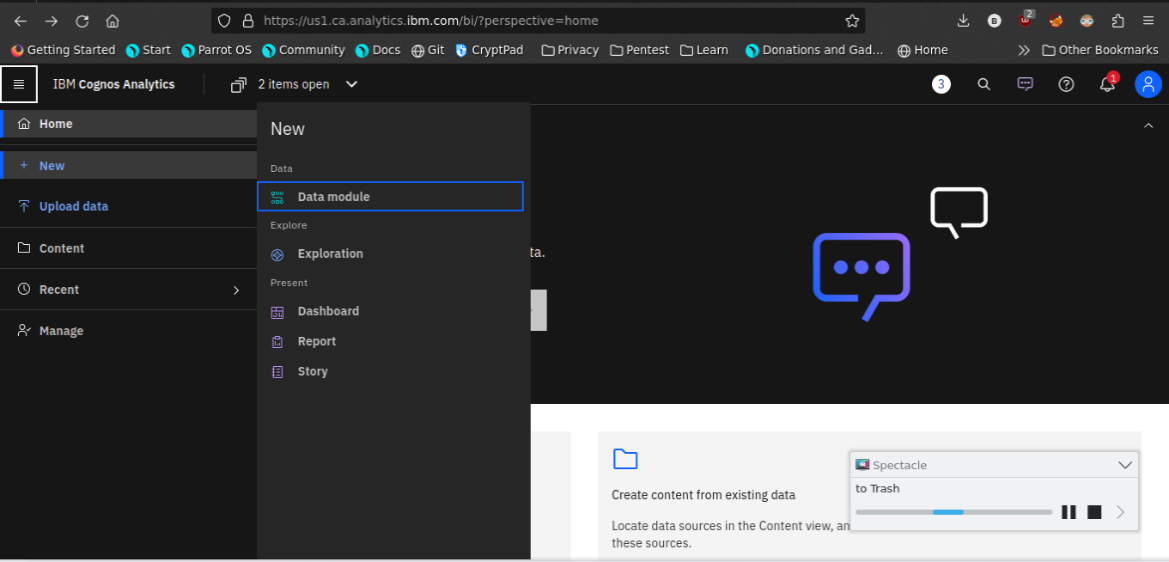
Y-axis = Q-P2

After performing these activities a comprehensive document will be created to demonstrate the ability to Communicate and share finding.

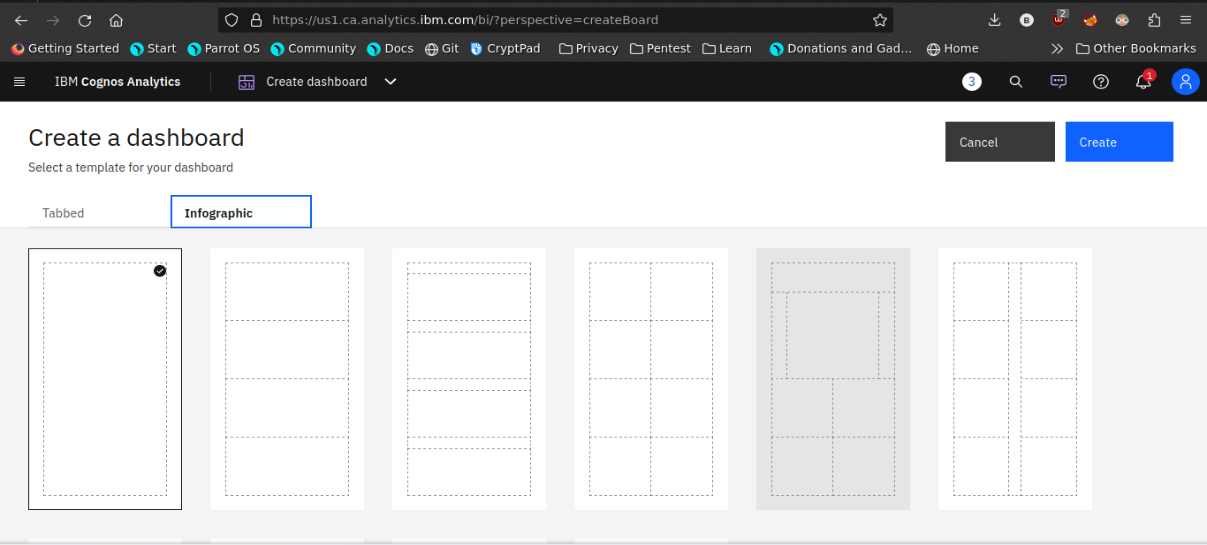
**Visulaization in IBM Cognos**

**Step 1:**

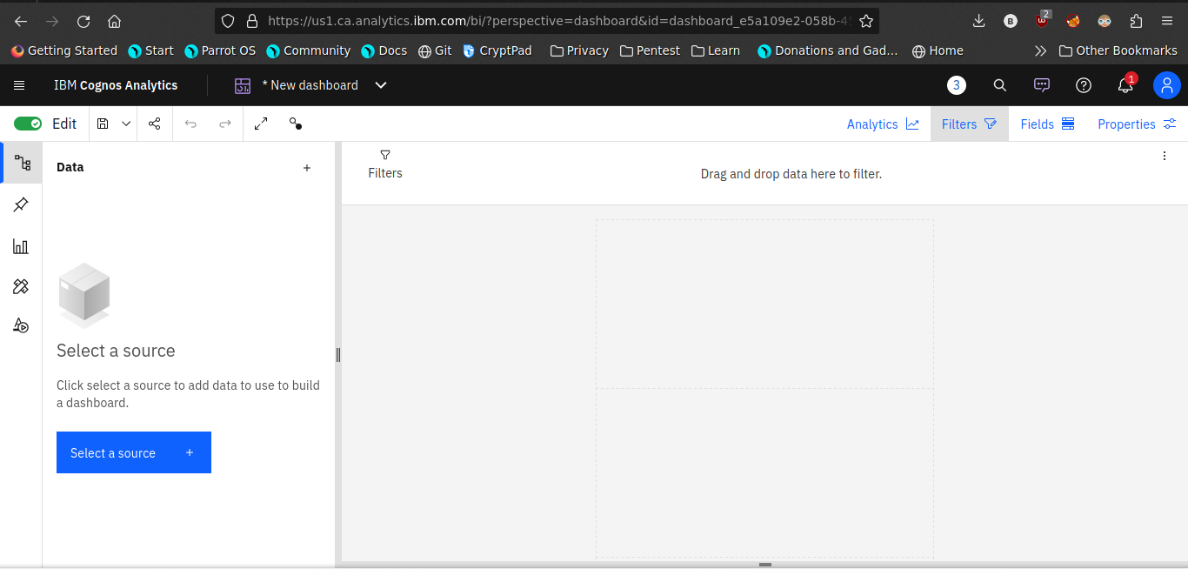
1. Login to your IBM cognos
2. Click more menu from the left side
3. Select new tab
4. Click Dashboard tab



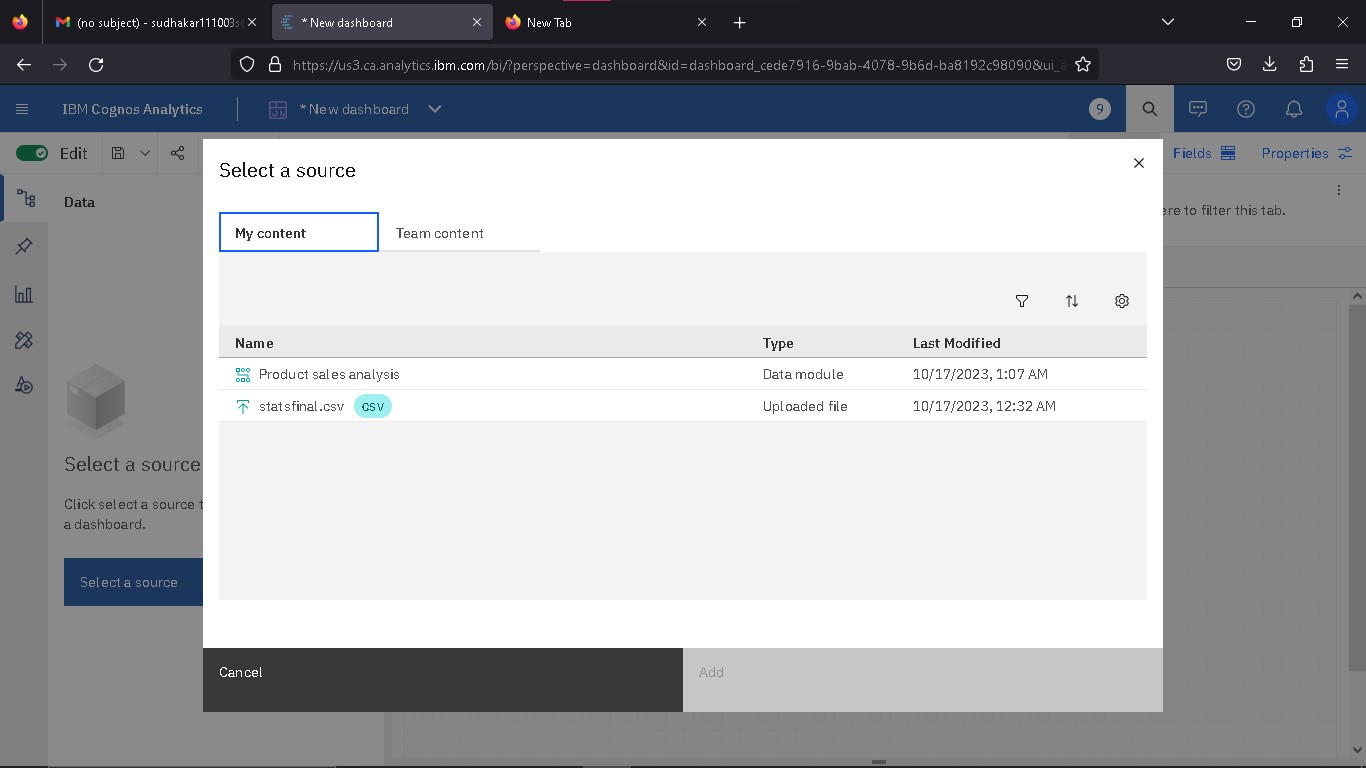
1. Select Template for your dashboard



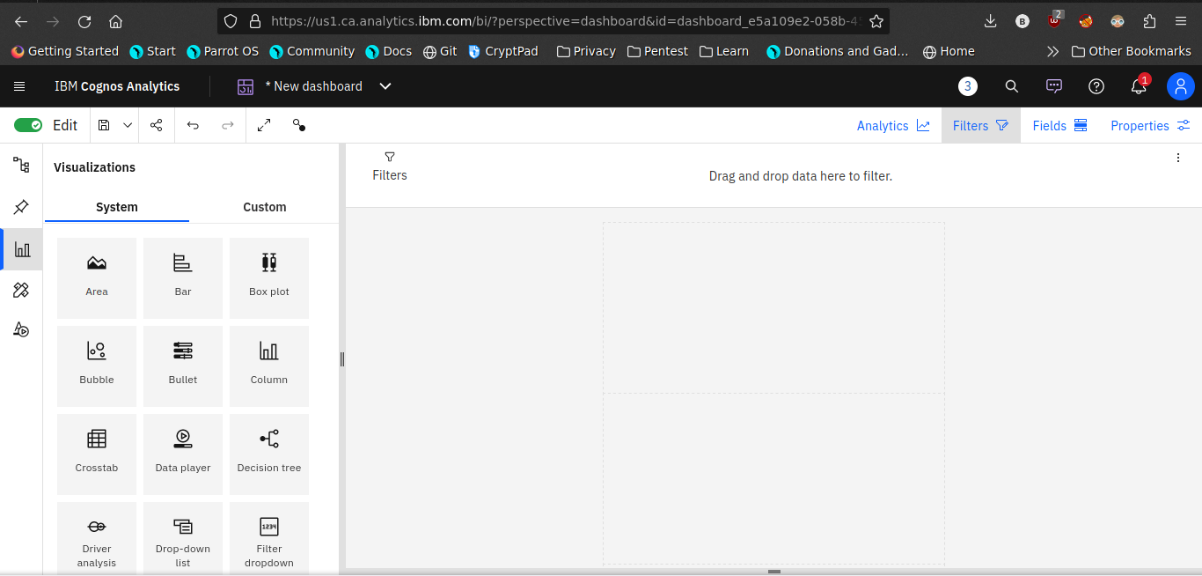
1. Now the dashboard is created and select your data-source.



1. Select your Corresponding dataset

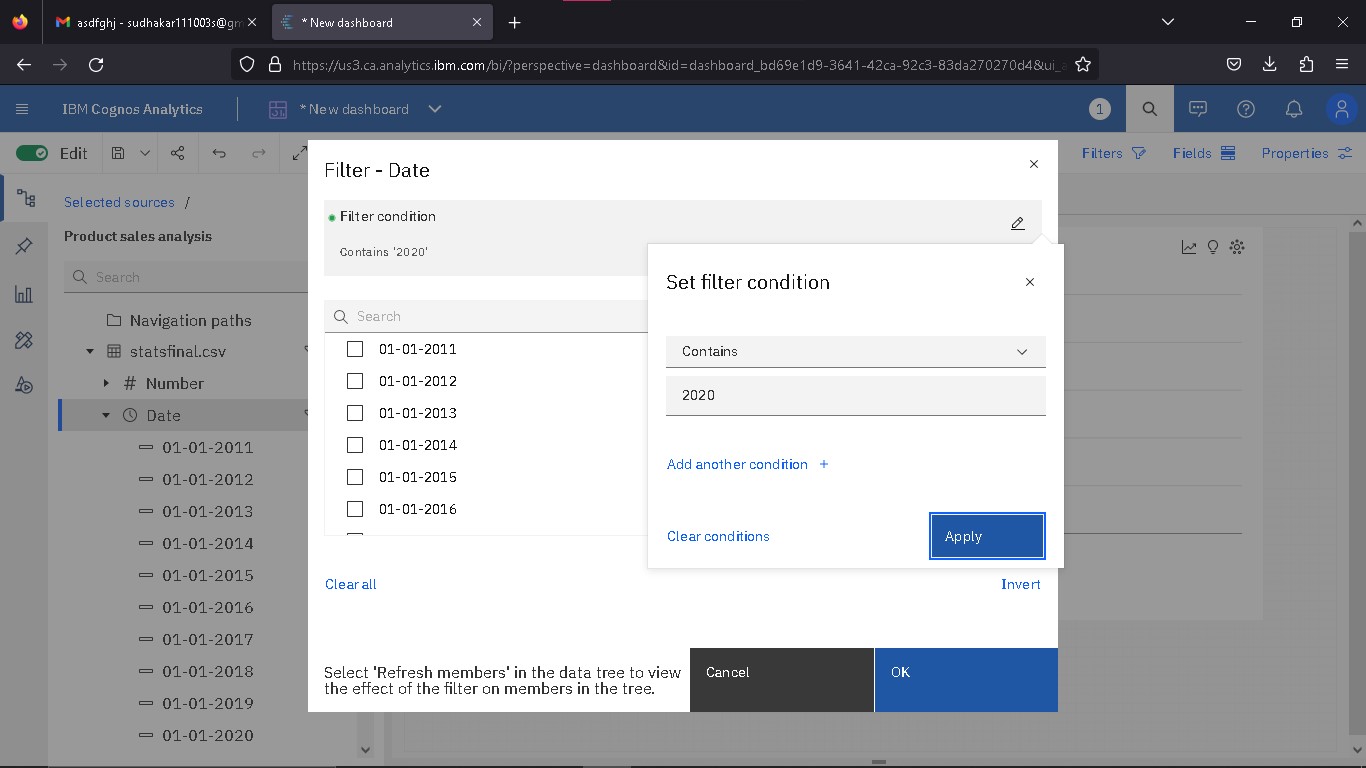


1. Select favourable visualization system



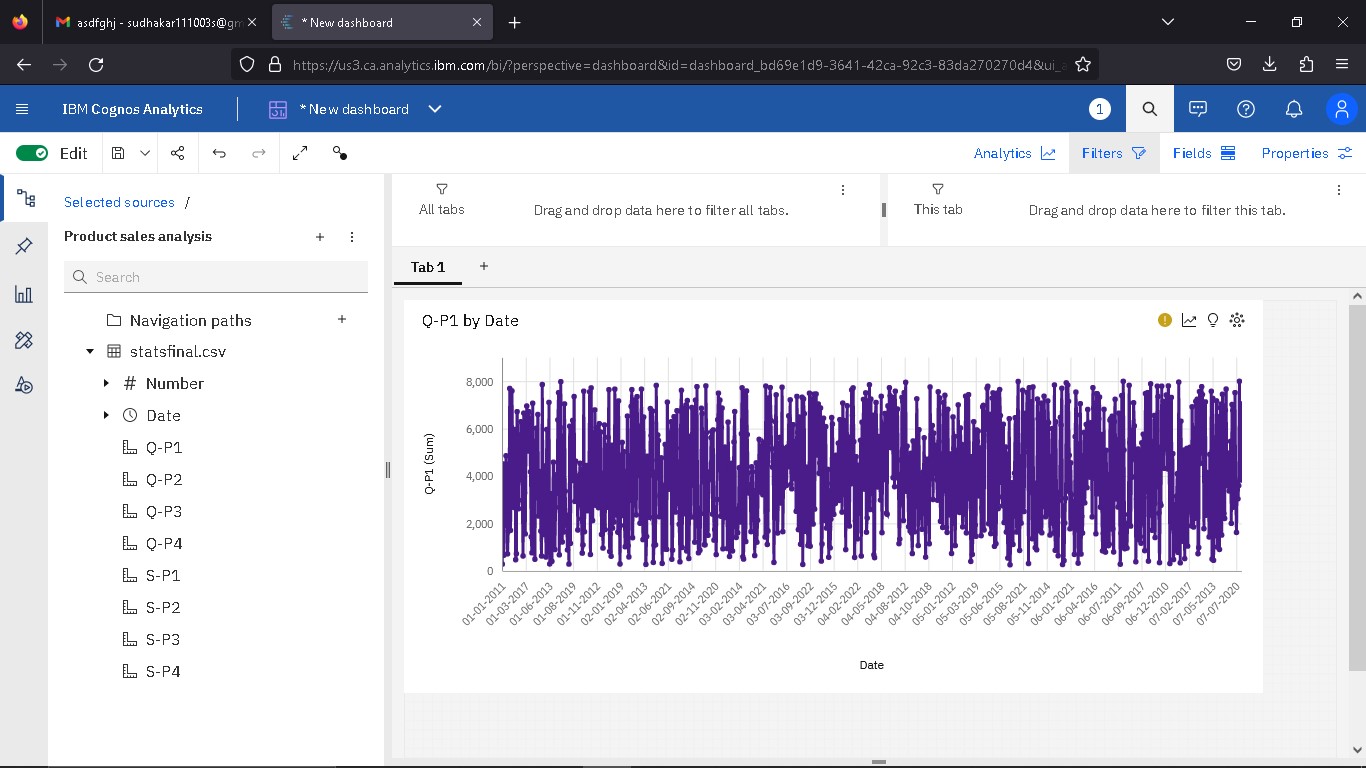
1. If needed Filter the data

Here the datasets are filtered by the date Which contains 2020



1. Line plot

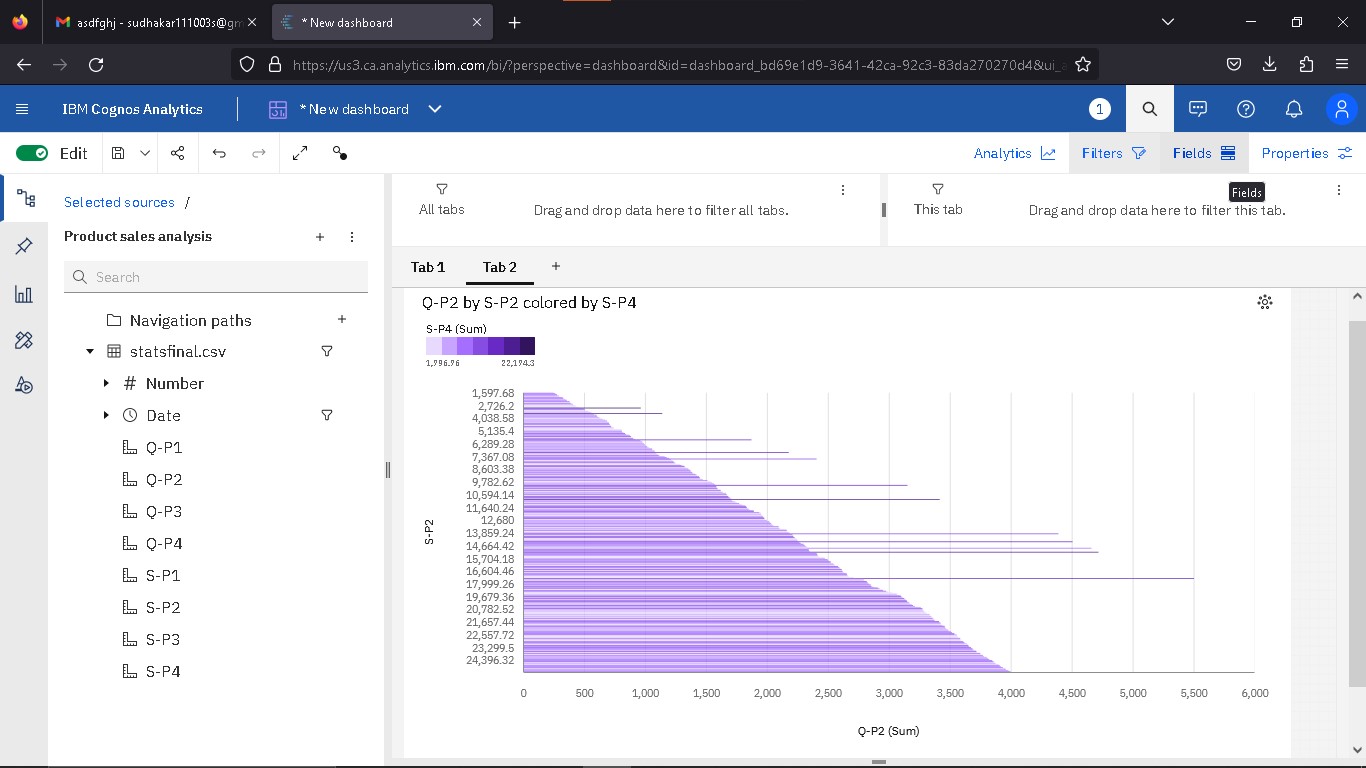
In this line plot X\_axis are dates and Y\_axis are Date and Q-P1



1. **Barchart**

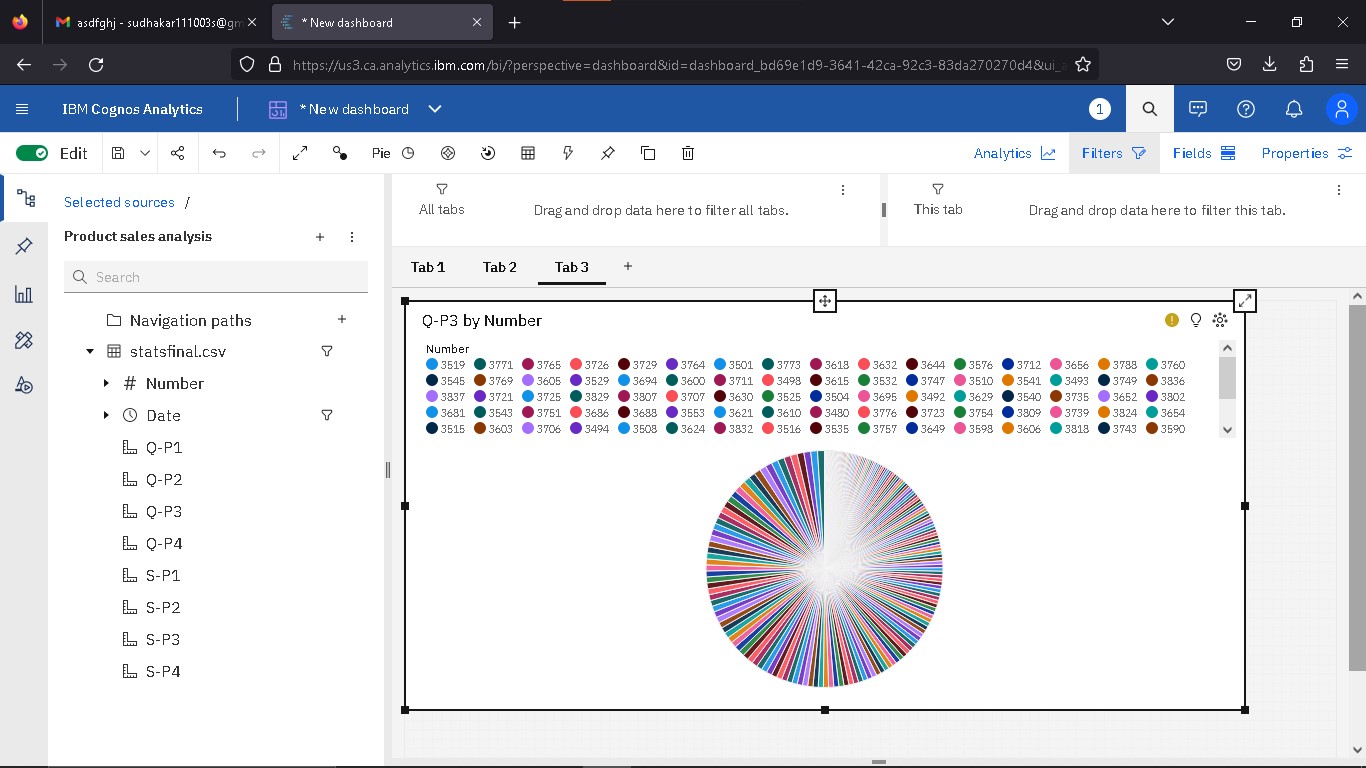
In this Bar chart the bars represent the Q-P2 and length defines S-P4 and colored by S-P2

It helpful to visualize the maximum S-P4 occurs on a day



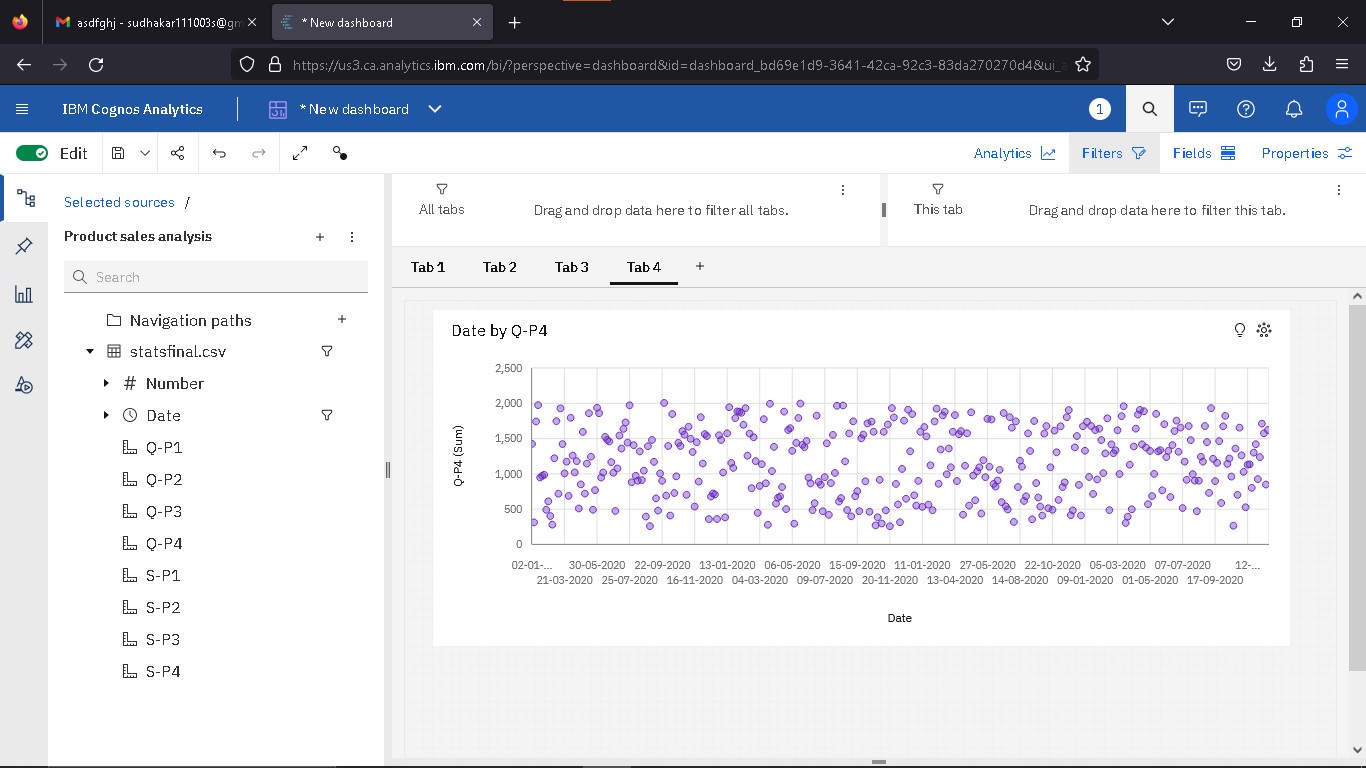
1. **Pie chart,**

This is same as a bar chart. it helpful to analyze the Numbers by Q-P3



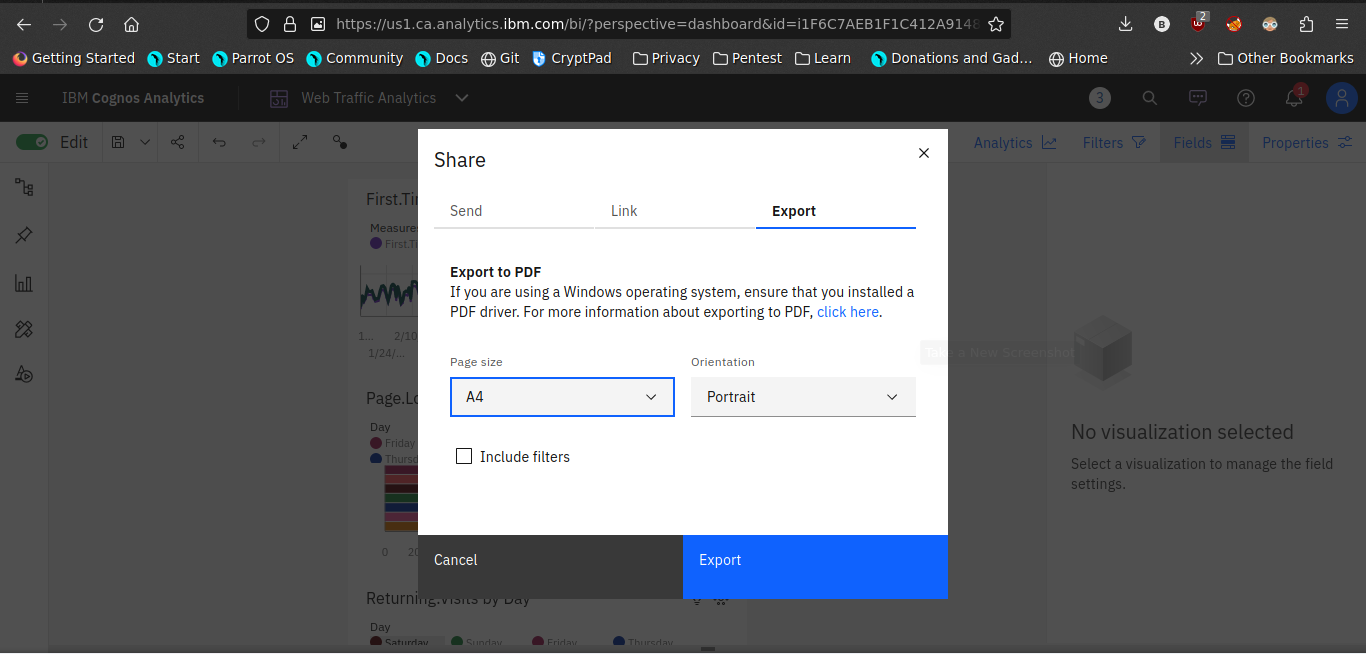
1. **Scatter plot**

It is used to display the relationship between two variables and observe the nature of the relationship. The relationships observed can either be positive or negative, non-linear or linear



1. After completing the dashboard lets export to another format

Like pdf





Now the visualization phase where over. lets start analyze the dataset using Python libraries use machine learning models for predictive analyisis.

**Data Analysis using python**

In this steps are used to analyse the given dataset using python libraries

**Steps:**

1. **Import Necessary packages**

Pandas

seaborn

Machine learning models

Linear regression

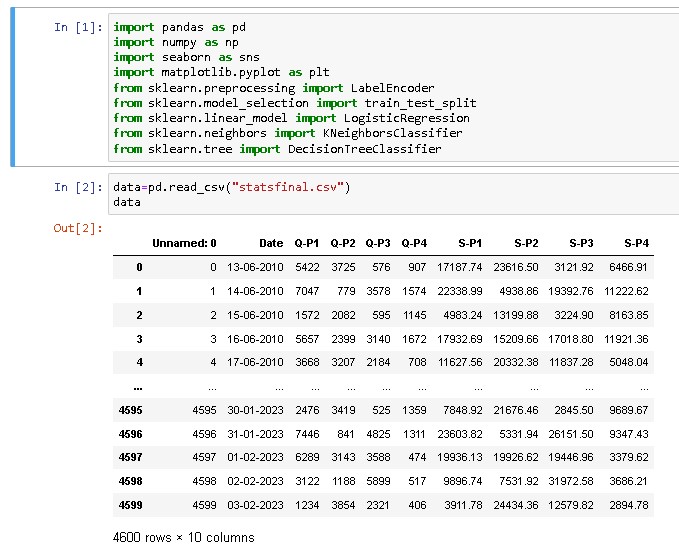
1. **Make a training and test data**

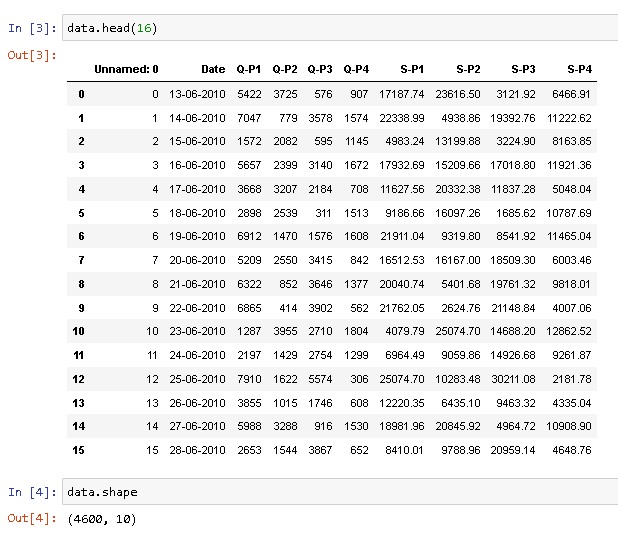
Use the train test split model

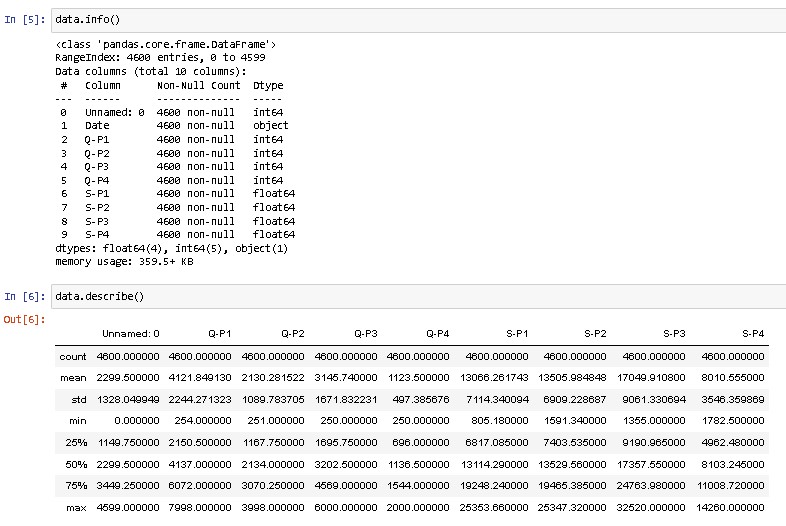
Compare the testing and training data set by visualization library

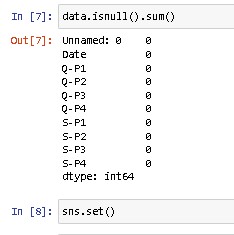
1. **Calculate the accuracy of the model**

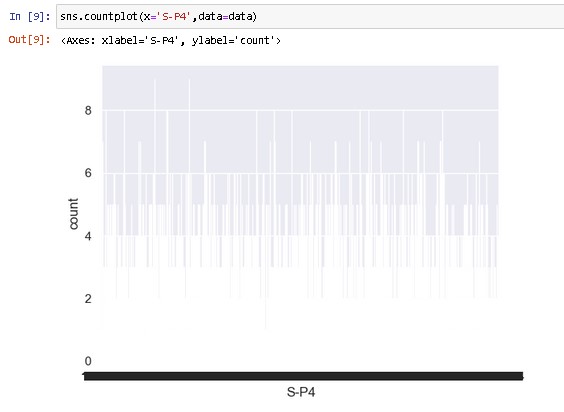
Use r2\_score to measure the accuracy of the model

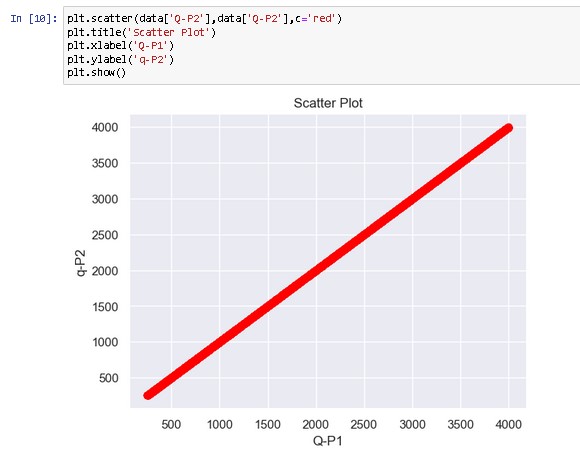


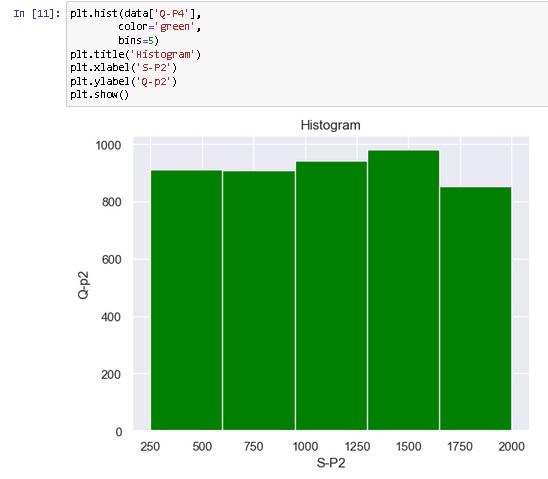


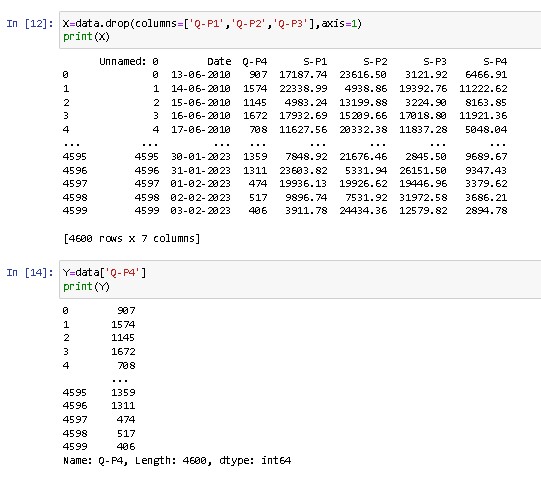


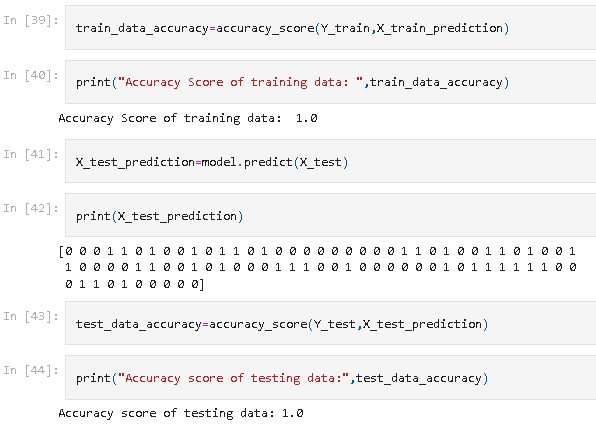












The accuracy for the above machine learning model is

1.0

**Conclusion:**

In the dynamic landscape of modern business, where data reigns supreme, the fusion of Product Sales Analysis and Design Thinking proves to be an invaluable ally. This holistic approach not only allows us to extract critical insights from sales data but also ensures that these insights are directly aligned with the needs and aspirations of both the business and its customers.Top of Form

Bottom of Form