**IBM-Naan Mudhalvan Data Analytics with Cognous**

**Phase 5**

**Development Part-1**

**Student Name : P. Ajeeth kumar**

**Register Number :** 620821104006

**Branch :** B.E CSE

**Year :** 3rd Year

**Topic :** Data Analytics with Cognos

**Title :** Public Transportation Analysis

**College :** Gnanamani College of Technology

**Introduction:**

Public transportation analytics is a field of study and practice that involves collecting, analyzing, and interpreting data related to public transportation systems. It plays a crucial role in optimizing and improving the efficiency, safety, and accessibility of public transportation services in urban and rural areas. Public transportation analytics leverages various data sources, including real-time information, historical data, and demographic data, to make informed decisions and enhance the overall passenger experience. Here is an introduction to some key aspects of public transportation analytics

**Abstract:**

Public transportation systems play a critical role in urban mobility, but they face numerous challenges, including efficiency, accessibility, and sustainability. This abstract outlines a comprehensive approach to address these issues through a detailed analysis, problem definition, and the application of design thinking principles. The objective is to improve public transportation systems by harnessing data-driven insights, advanced visualization strategies, and seamless code integration.

**Problem Definition:**

Public transportation systems often struggle with problems such as overcrowding, delays, and insufficient accessibility. Additionally, environmental concerns and changing travel patterns demand innovative solutions. The problem definition phase seeks to identify these challenges and prioritize them to guide the analysis process effectively.

**Design Thinking:**

Design thinking principles will be applied to tackle public transportation issues. This human-centered approach involves empathizing with commuters, defining the problem, ideating creative solutions, prototyping, and testing. This iterative process aims to ensure that the final solution meets the needs of both commuters and the environment.

**Analysis Objective:**

The primary objective of the analysis is to enhance the efficiency, accessibility, and sustainability of public transportation systems. Specific goals include:

**1. Optimizing Routes:** Identifying routes with the highest demand and optimizing them to reduce overcrowding and delays.

**2. Accessibility Improvement:** Analyzing current accessibility levels, particularly for individuals with disabilities, and proposing enhancements.

**3. Sustainability Enhancement:** Evaluating the environmental impact of public transportation and suggesting measures to reduce carbon emissions.

**4. Customer Experience:** Gathering and analyzing passenger feedback to improve overall satisfaction and convenience.

**Data Collection:**

Data will be collected from various sources including:

**1. Traffic and Commuter Data:** Utilizing real-time traffic data, passenger counts, and journey histories to understand travel patterns and congestion points.

**2. Accessibility Data:** Examining information on infrastructure, such as ramps, elevators, and accessible vehicles.

**3. Environmental Data:** Gathering data on emissions, fuel consumption, and energy usage of public transportation.

**4. Passenger Surveys:** Conducting surveys and interviews to gain insights into passenger experiences and preferences.

**Visualization Strategy:**

Data will be visualized using advanced technique including:

**1. Interactive Maps:** Visualizing optimal routes, congestion, and accessibility features on interactive maps for easy understanding.

**2. Real-time Dashboards:**Creating dashboards that display live data on delays, vehicle availability, and environmental impact.

**3. Heatmaps:** Using heatmaps to highlight areas with high passenger density and areas with accessibility issues.

**4. Passenger Journey Visualizations:** Visualizing passenger journeys to identify common travel patterns and bottlenecks.

**Code Integration:**

The analysis will involve integrating various data sources and analytics tools, including machine learning algorithms for predictive modeling, geospatial analysis libraries for route optimization, and web development frameworks for creating interactive dashboards. The integration will ensure a seamless and scalable solution that can be continuously updated and improved.

This comprehensive approach, combining problem definition, design thinking, data analysis, advanced visualization, and code integration, aims to transform public transportation systems into more efficient, accessible, and sustainable modes of urban mobility.

Phase – 2

**Predicting Service Disruptions:**

**Data Collection**:

To predict service disruptions, we need to collect historical data from various sources, including:

1. **Transit Logs**: These logs contain information about routes, schedules, and real-time updates.

2. **Weather Data**: Weather conditions can significantly impact public transport. Integrating weather data into the analysis can help predict delays caused by adverse weather.

3. **Maintenance Records**: Data on maintenance schedules, repairs, and vehicle condition can be vital in predicting breakdowns.

4. **Social Media:** Monitor social media platforms for mentions of service disruptions and customer complaints.

**Machine Learning Algorithms**:

1. **Time Series Forecasting**: Utilize time series models like ARIMA or Prophet to predict future delays based on historical data.

2. **Regression Analysis**: Analyze the impact of variables like weather, maintenance, and special events on service disruptions using regression models.

3. **Natural Language Processing**( NLP): Extract insights from passenger feedback using sentiment analysis to identify potential issues leading to disruptions.

**Implementation with Cognos:**

IBM Cognos can be integrated to create visually appealing dashboards and reports for service disruption predictions. The following steps can be followed:

1. **Data Integration**: Import and transform data from various sources into a format suitable for analysis.

2. **Model Training**: Train machine learning models using historical data.

3. **Real-time Updates:** Incorporate real-time data streams to continually update predictions.

4. **Visualization:** Create interactive dashboards and reports in Cognos to monitor and analyze predictions.

5. **Alerting:** Configure alerts in Cognos to notify relevant personnel when disruptions are predicted.

Analyzing Passenger Sentiment:

**Data Collection:**

To analyze passenger sentiment, collect data from multiple sources, including:

1. **Customer Feedback Forms**: Analyze comments, ratings, and feedback from passengers.

2. **Social Media Monitoring:** Monitor social media platforms for passenger sentiment and complaints.

3. **Survey Data:** Conduct regular passenger satisfaction surveys.

**Machine Learning Algorithms:**

1. **Sentiment Analysis**: Utilize NLP techniques to classify passenger sentiment as positive, negative, or neutral.

2. **Topic Modeling:** Identify common topics or issues mentioned in passenger feedback.

3. **Text Analytics**: Extract meaningful insights from unstructured text data.

**Implementation with Cognos**

IBM Cognos can be used to create insightful reports and dashboards for passenger sentiment analysis:

1. Data Integration: Import and preprocess passenger feedback data.

2. Sentiment Analysis: Apply sentiment analysis models to classify feedback.

3. Visualization: Create visualizations in Cognos to display sentiment trends, common issues, and overall passenger satisfaction.

4. Feedback Loop: Use the analysis to inform improvements and monitor changes in passenger sentiment over time.

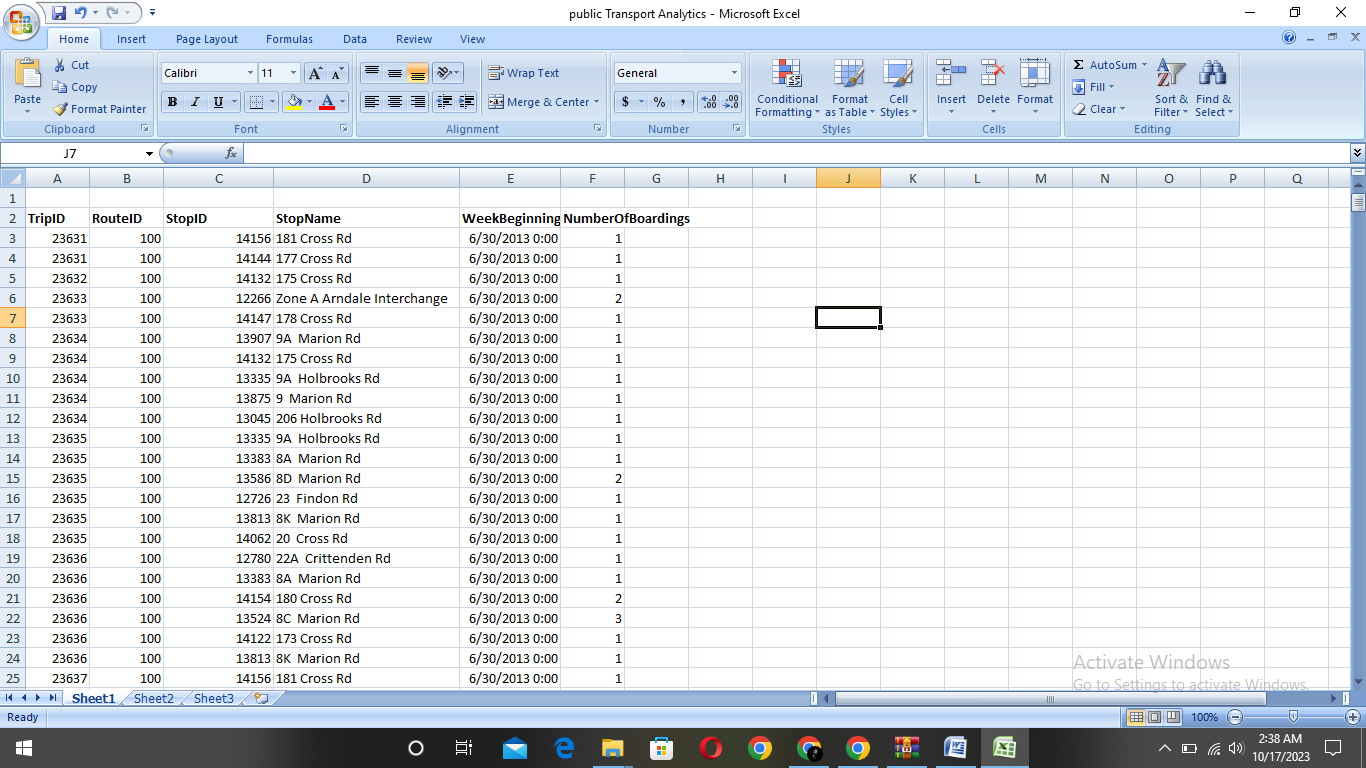
**Phase 3**

**Objectives**

In this phase defines start to building the Project by loading and preprocessing the dataset and perform different analysis and visualization using IBM Cognos.

**Data source**

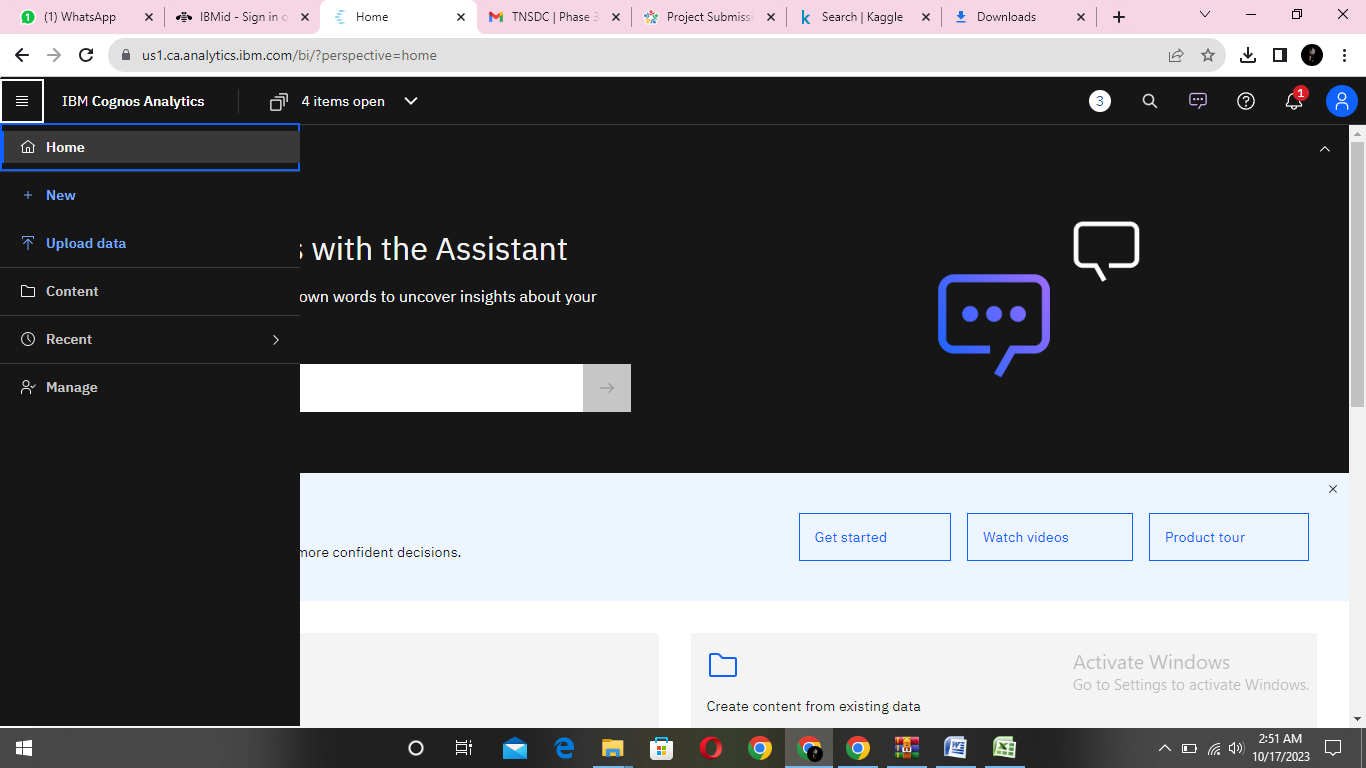
Dataset is collected from the kaggle.com named “daily-website-visitors.csv” which has a data about the Days, Day of week, Date, page Loads, Unique visits, First-time visits, Returning Visits.

Dataset link: <https://www.kaggle.com/datasets/rednivrug/unisys?select=20140711.CSV> 

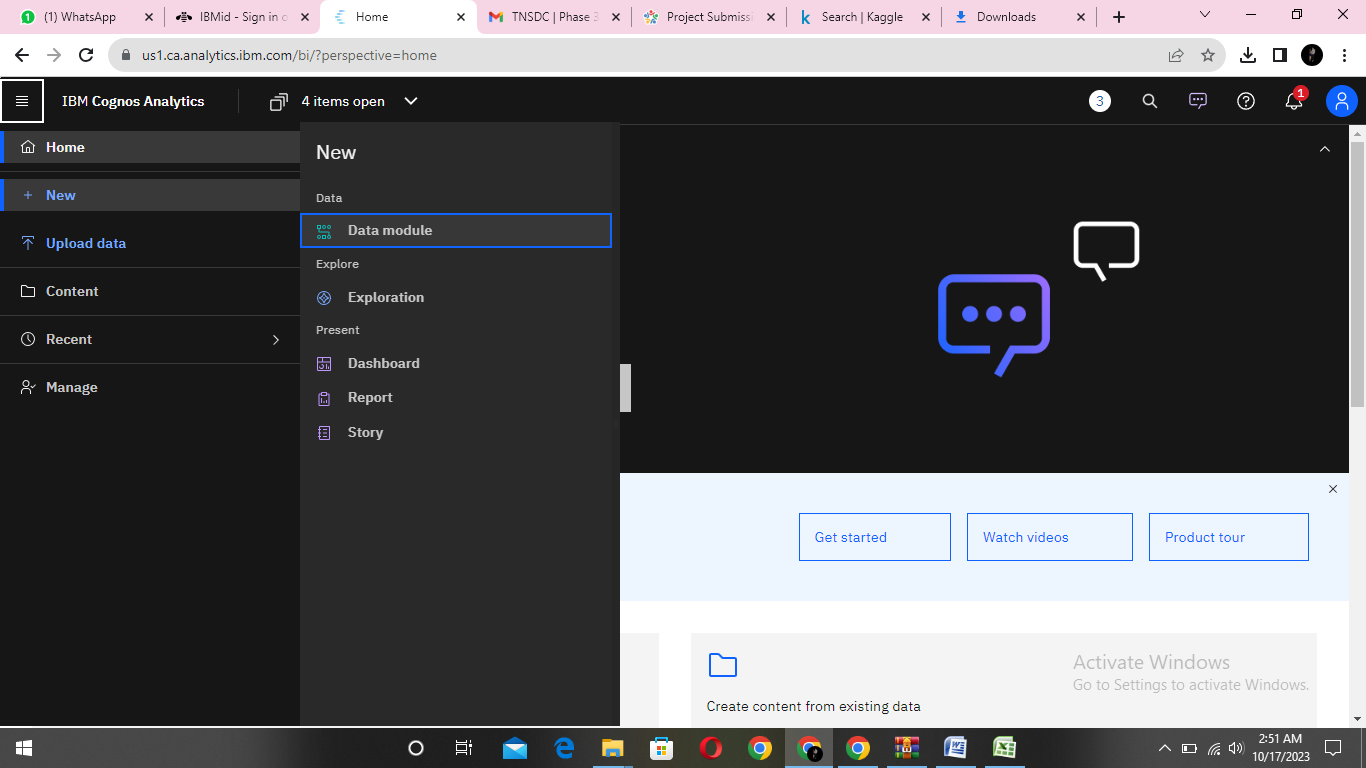
**Data Loading**

Steps Involved in data loading on IBM cognos.

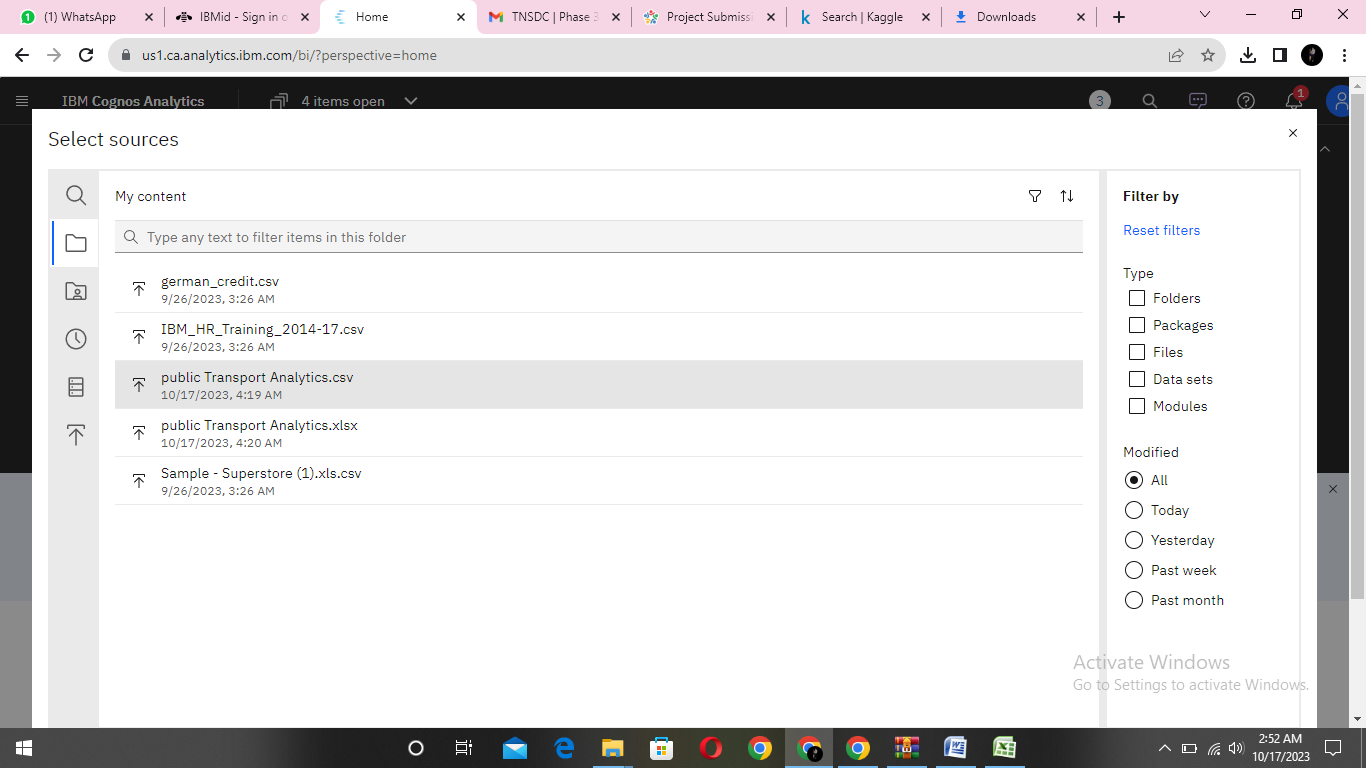
1. Login to your IBM cognos
2. Click more menu from the left side
3. Select new tab



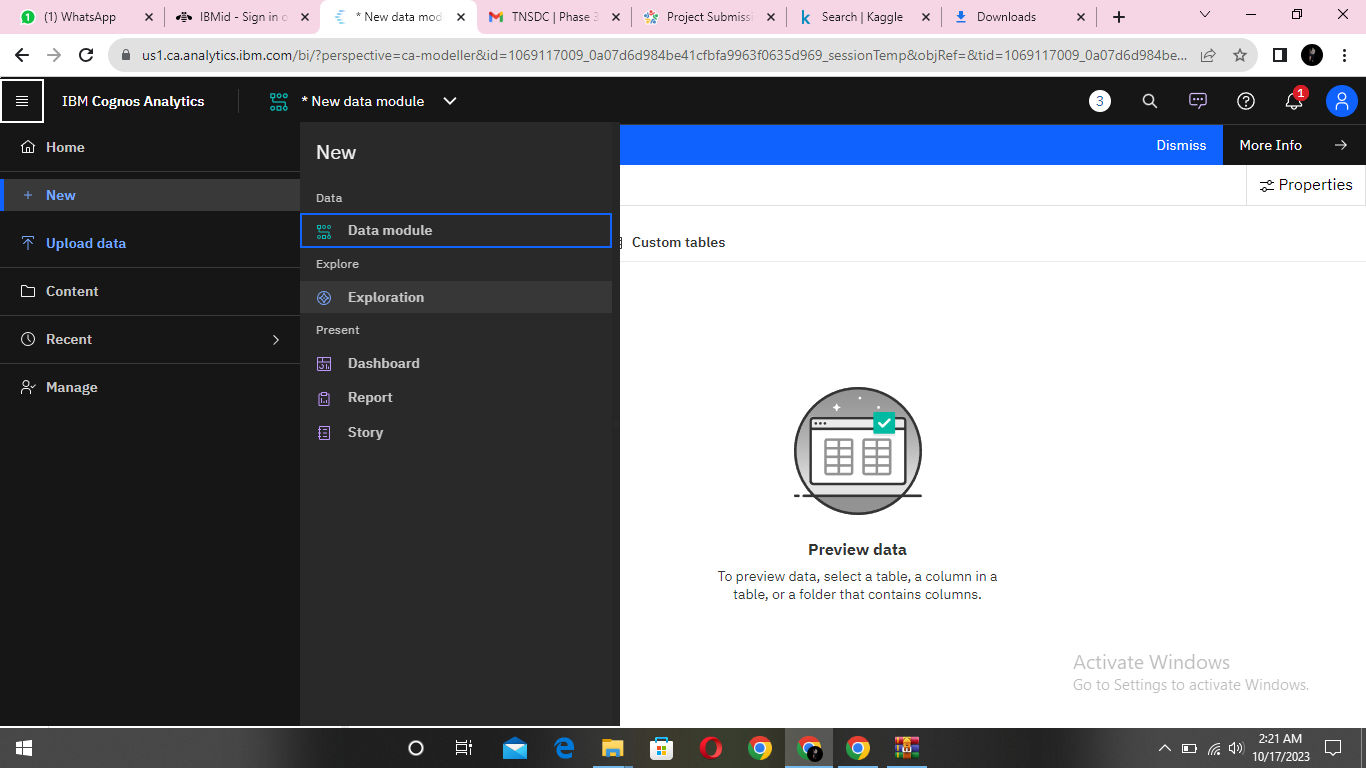
4. Click Data module tab

****

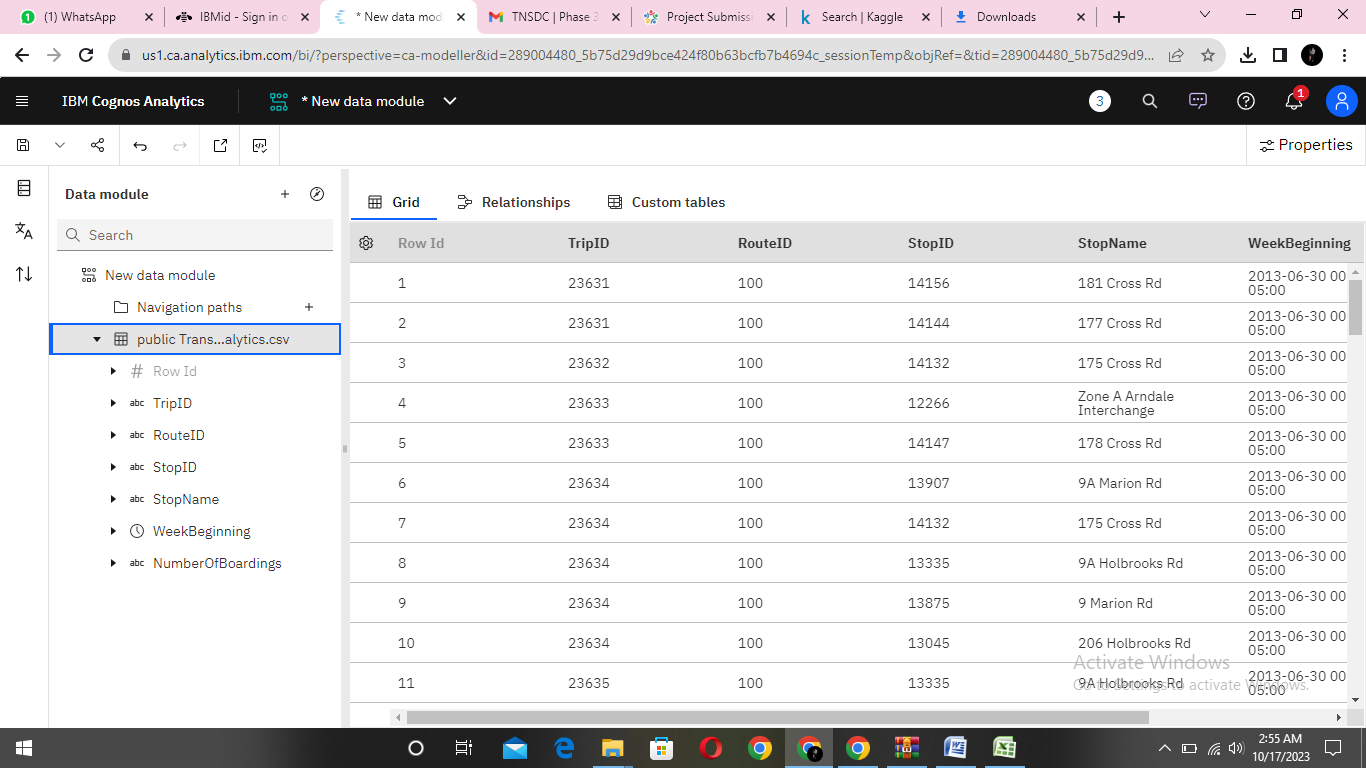
5.Upload the dataset for your project and select the Corresponding file



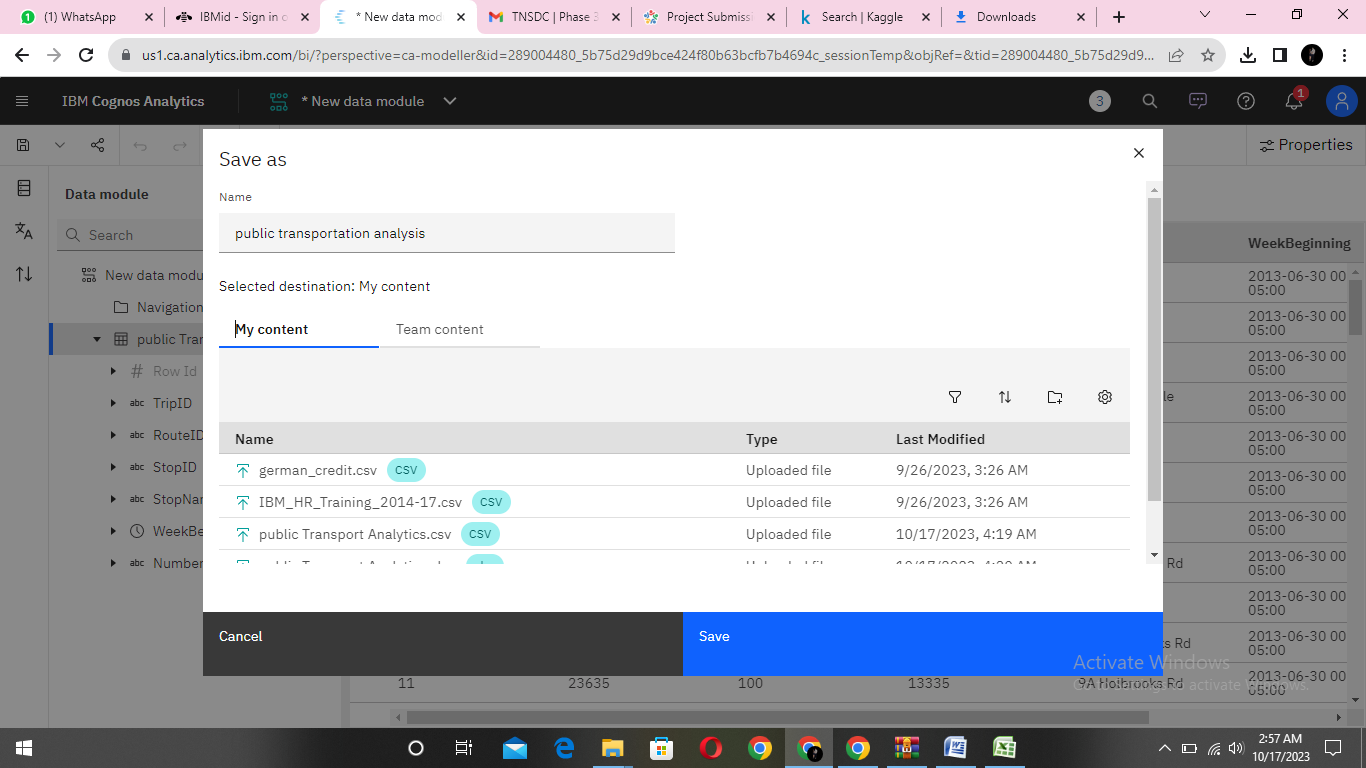
6. preview the data



7.Explore the data



8. save the data module



**Data Preprocessing and Cleaning**

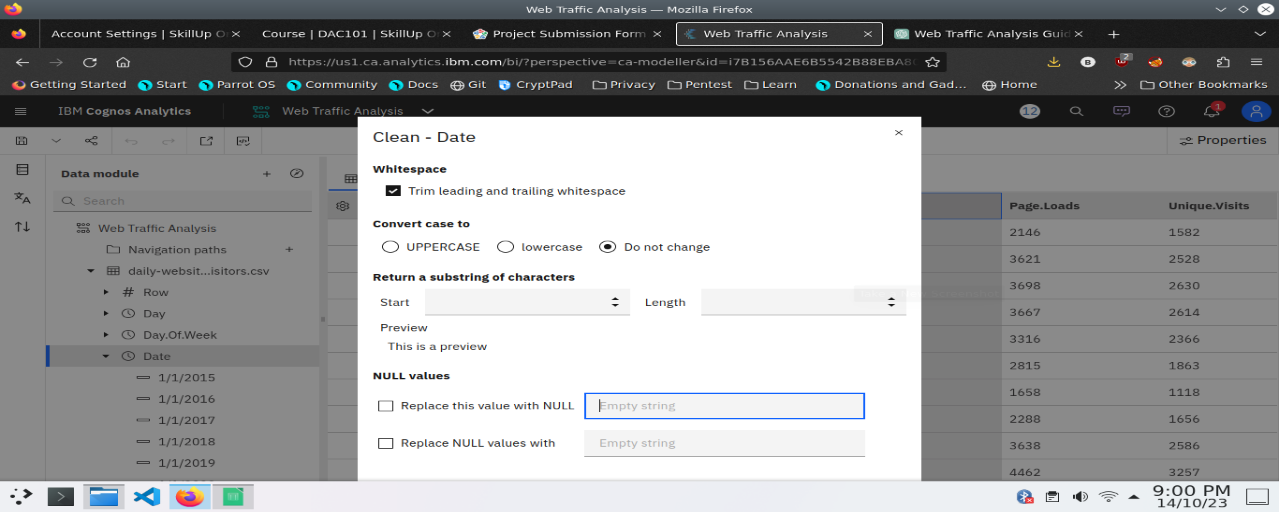
In this phase the following steps will taken

* 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



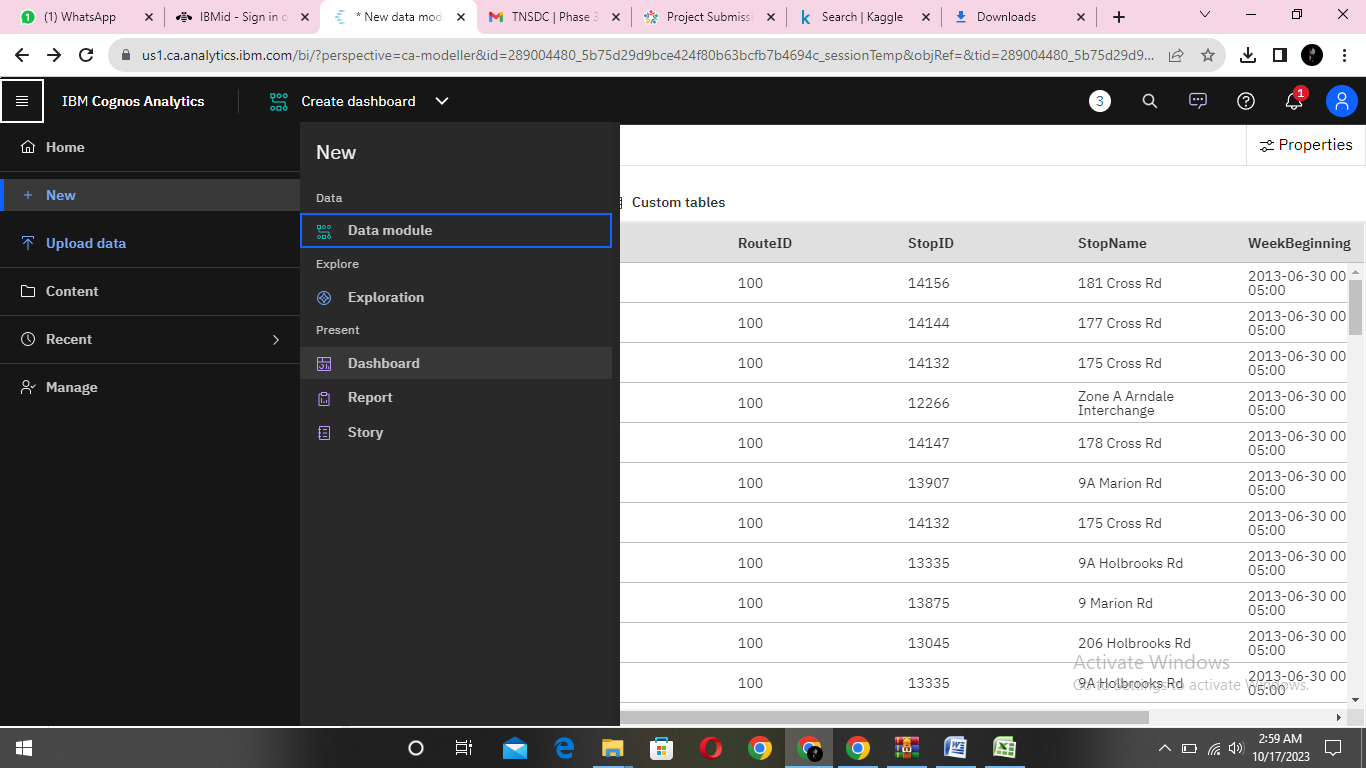
datamodule will be updated by doing the above process

after the completion ofprocess start creating the dashboard for Visualization

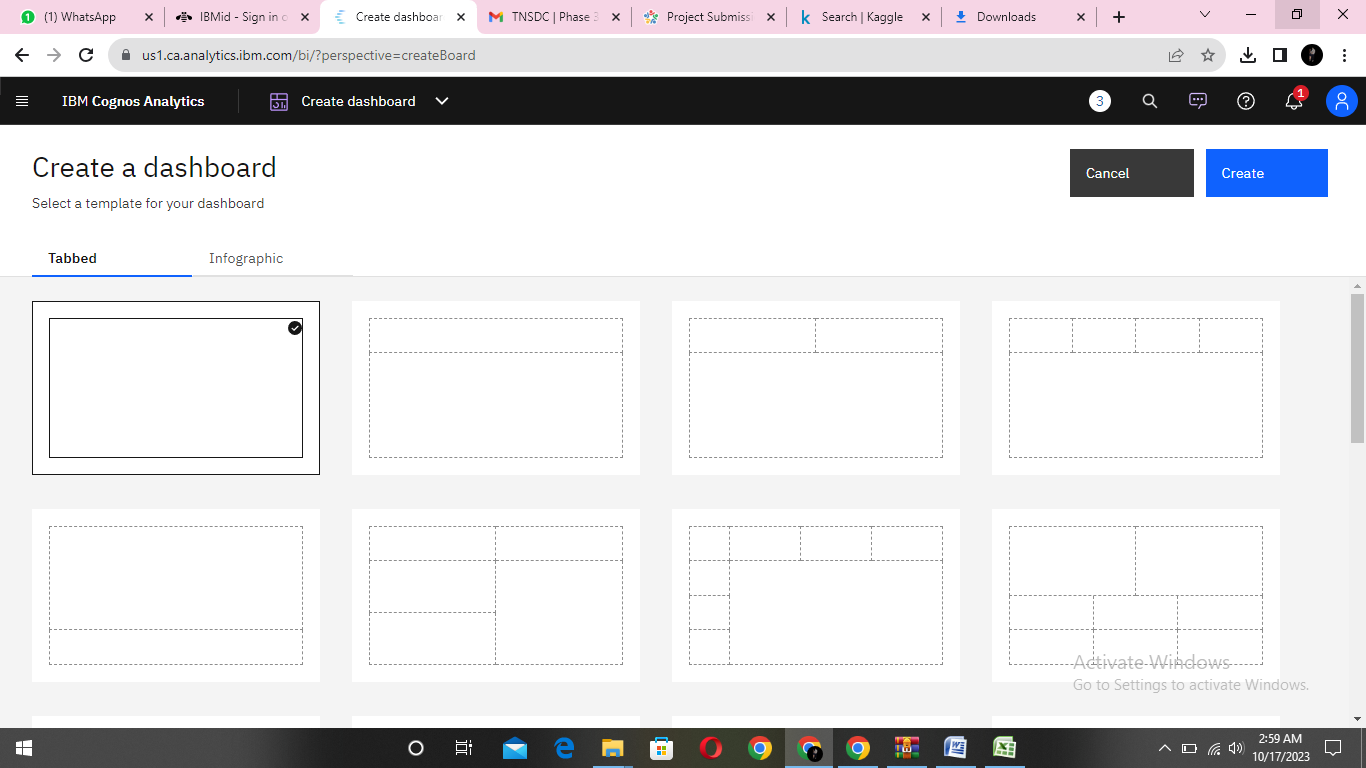
**Dashboard Creation**

Dashboard creation are helpful to visualizing the data

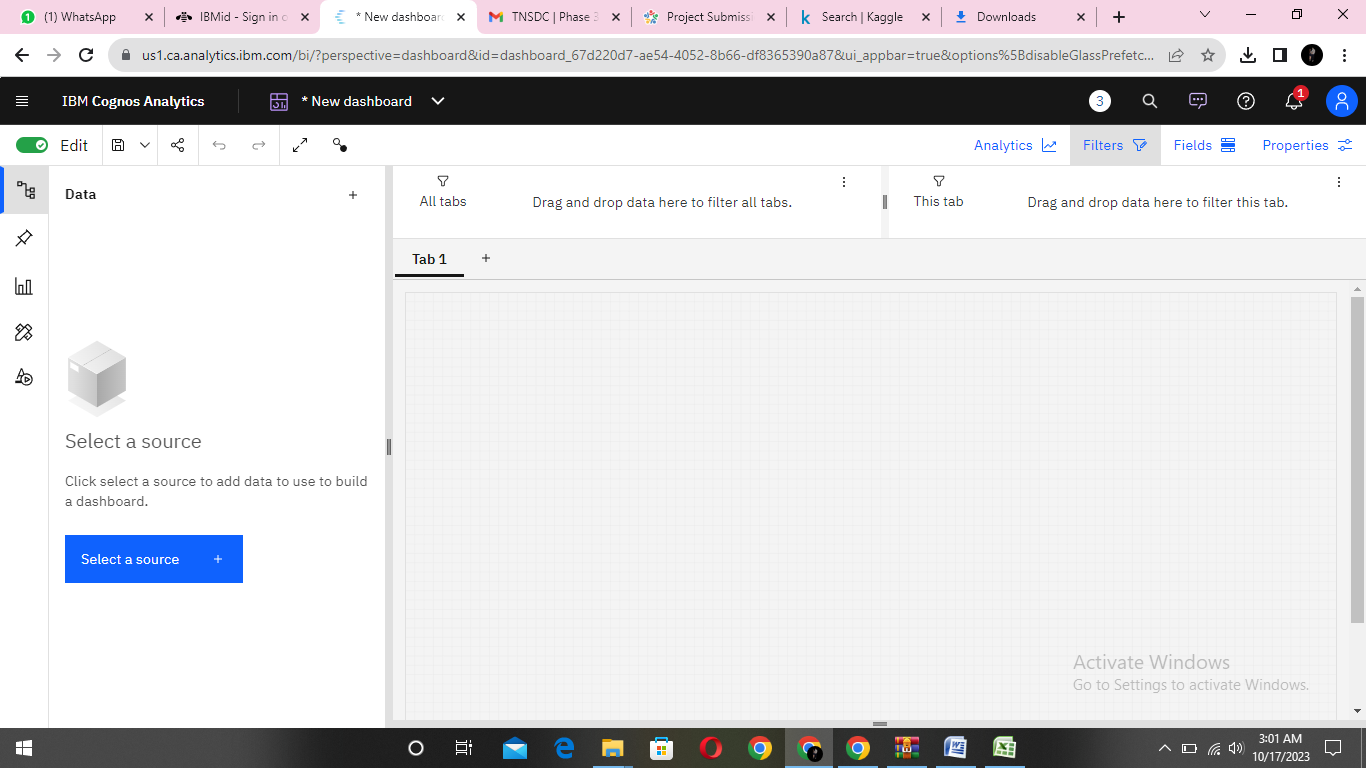
1. Goto Home menu
2. Select the new tab
3. Click dashboard



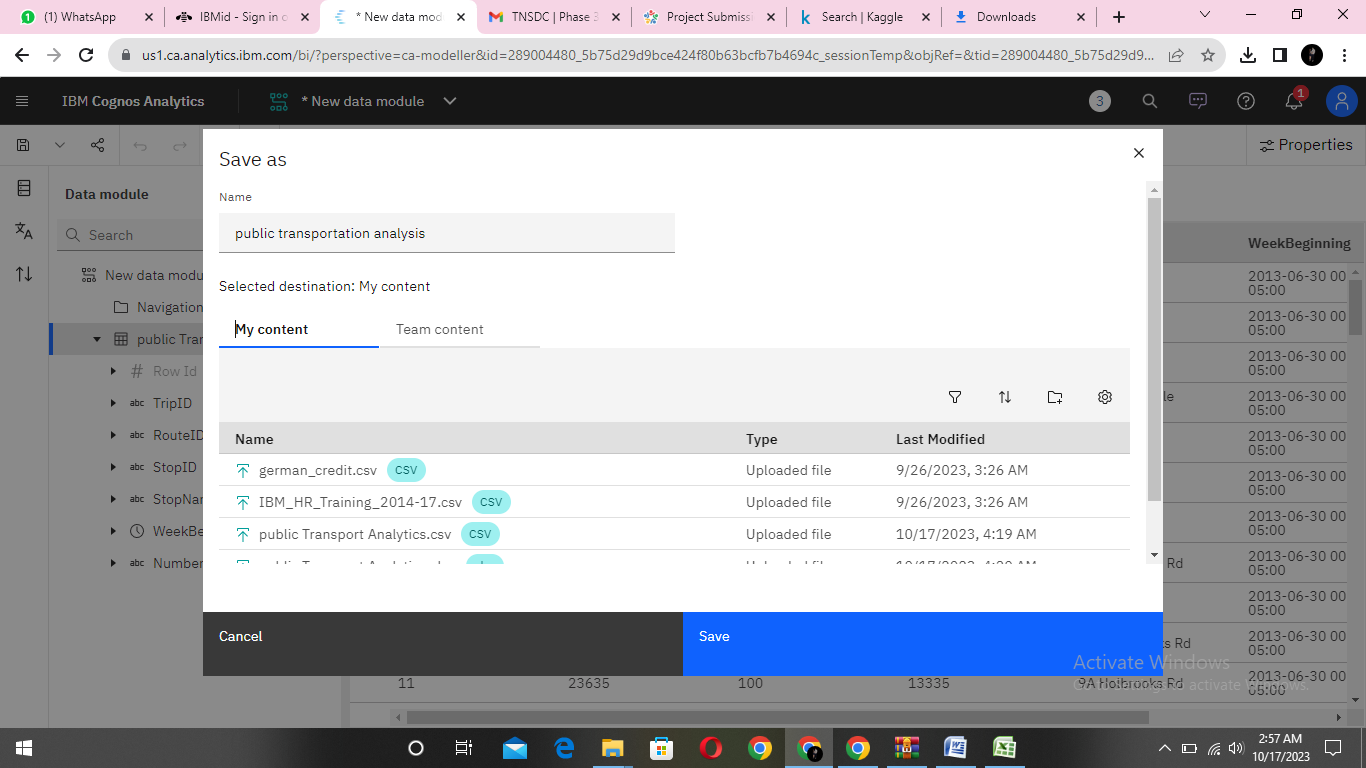
4. Choose the template for your project and click



5.Now Dashboard is created



6. Select the data source



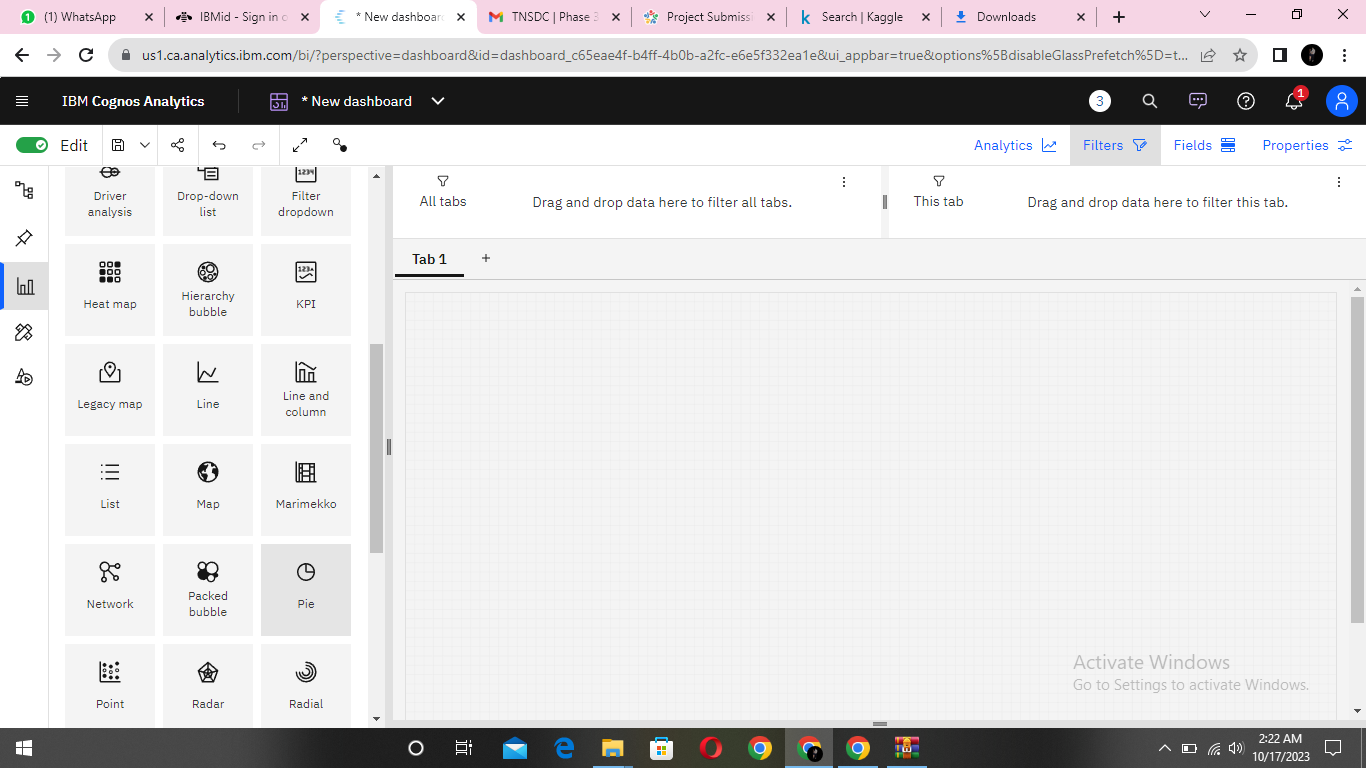
**Visualization**

After creating the dashboard, the next step is to visualize the data

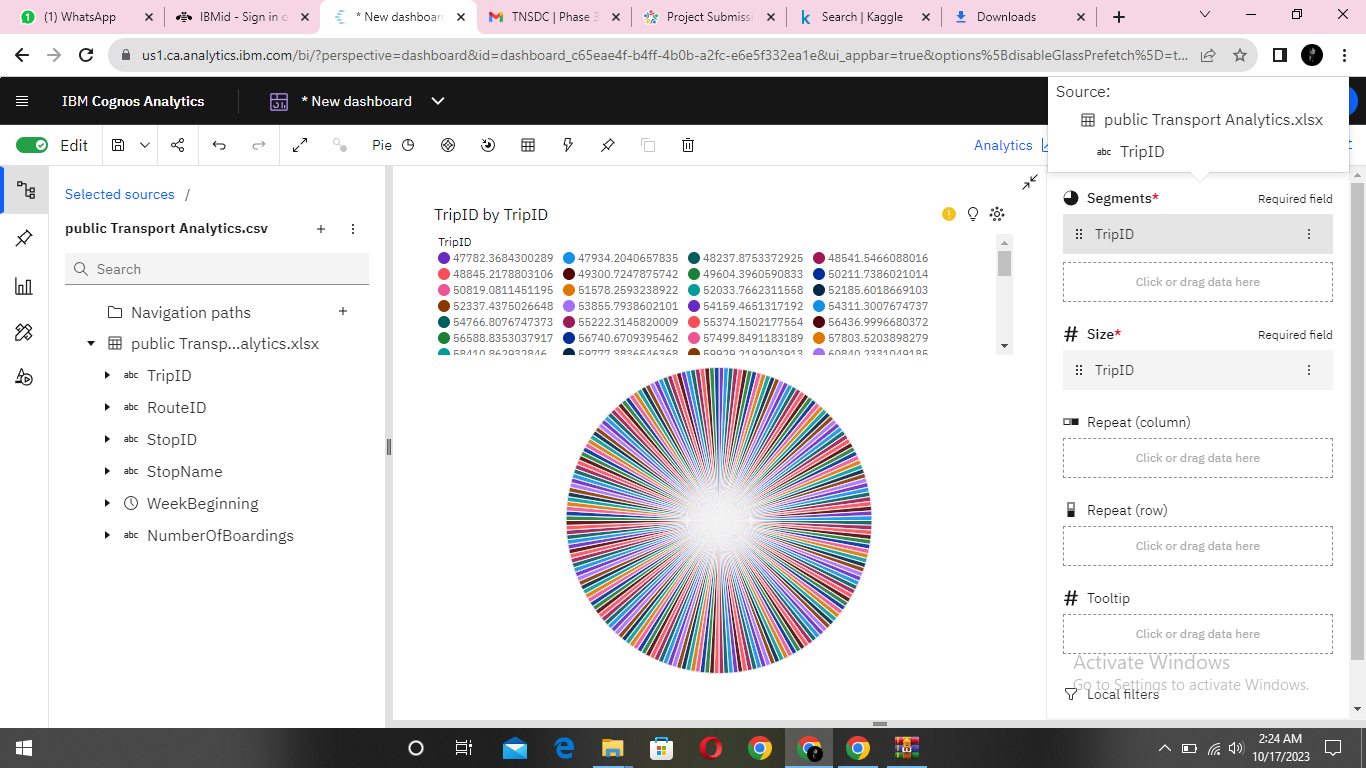
In IBM Cognos

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 Line graph and model compares the “Page.loads” and “Unique.visits” from the time period of 2020

X-axis =Dates

Y-axis = Page.Loads, Unique.visits.

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

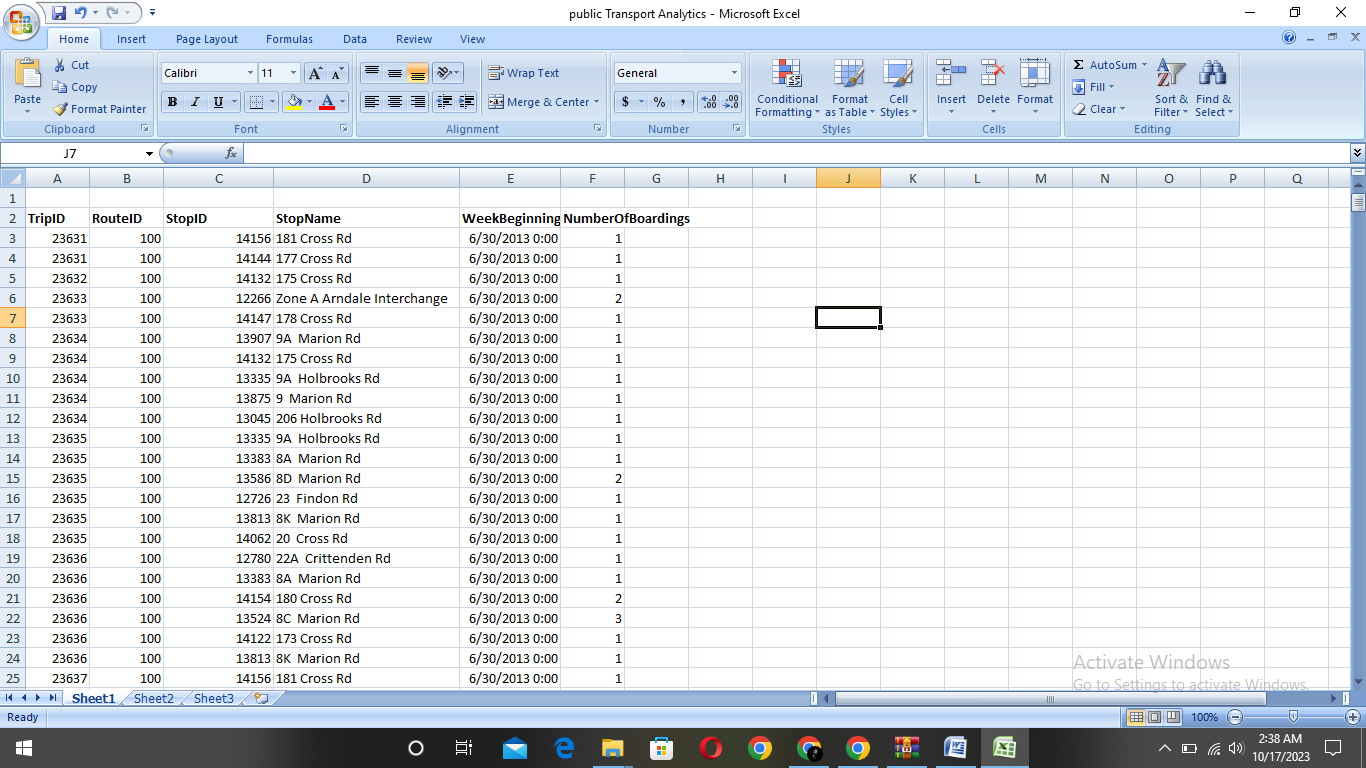
**Phase 4**

**Objectives**

In this phase defines start to building the Project by loading and preprocessing the dataset and perform different analysis and visualization using IBM Cognos.

**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/rednivrug/unisys?select=20140711.CSV> 

**1. Data Preparation:**

Prepare your data for analysis. Assuming you have a CSV file with the necessary data:

**python code**

import pandas as pd

# Load your data into a DataFrame

data = pd.read\_csv('your\_data.csv')

**2. Data Analysis with Python:**

Perform advanced data analysis in Python. For example, calculate the on-time performance rate:

python

Copy code

# Calculate on-time performance rate

on\_time\_count = len(data[data['arrival\_status'] == 'on\_time'])

total\_flights = len(data)

on\_time\_rate = on\_time\_count / total\_flights \* 100

print(f"On-time performance rate: {on\_time\_rate:.2f}%")

Perform sentiment analysis on passenger feedback:

python

Copy code

from textblob import TextBlob

data['sentiment'] = data['feedback'].apply(lambda x: TextBlob(x).sentiment.polarity)

**3. Data Visualization in Cognos**:

Use IBM Cognos to design dashboards and reports. You'll need to follow Cognos-specific procedures to create visualizations. This can involve uploading your cleaned and analyzed data into Cognos and then creating reports and dashboards.

**4. Integration of Python and Cognos:**

To integrate Python analysis results into Cognos reports, you can export the Python analysis results to a format Cognos can consume. For example, you could save your Python analysis results to a CSV file:

python

Copy code

# Save the results to a CSV file

analysis\_results.to\_csv('analysis\_results.csv', index=False)

Then, import this CSV file into Cognos for use in your reports and dashboards.

**5. Documentation:**

You can create a Jupyter Notebook to document your Python data analysis process. In the notebook, include code, explanations, visualizations, and comments to make your analysis clear and reproducible. You can export the Jupyter Notebook to various formats, including HTML or PDF for sharing.

**6. Automation (Optional):**

If this analysis needs to be conducted regularly, you can automate the entire process by scheduling Python scripts to run at specific intervals and update your Cognos reports automatically.

**Conclusion :**

public transportation analytics is a crucial and dynamic field that harnesses data-driven insights to enhance the performance and effectiveness of public transportation systems. It empowers transit authorities, urban planners, and policymakers to make informed decisions, improve passenger experiences, and optimize the overall efficiency of transportation services. The key takeaways regarding public transportation analytics include