**PHASE:3**

**Public Transportation Efficiency Analysis**

**TEAM MEMBERS:**

NAME: Murali V

REG NO:721221104040

NAME: Ragul M

REG NO:721221104044

NAME: Aswin S

REG NO:721221104009

NAME: Arun S

REG NO:72122110400

NAME: Abishek S

REG NO:721221104004

**Introduction:**

Public transportation plays a critical role in urban planning and sustainability. To ensure its optimal functioning, it's essential to analyse and visualize its efficiency. In this project, we will use IBM Cognos for data visualization to gain insights into public transportation efficiency. The project aims to answer questions such as:

**Data Collection:**

* Collect transportation data from the provided source. Ensure that the data includes information about routes, schedules, delays, ridership, and any other relevant metrics.

**DatasetLink:**[**https://www.kaggle.com/datasets/rednivrug/unisys?select=20140711.CSV**](https://www.kaggle.com/datasets/rednivrug/unisys?select=20140711.CSV)

**Data Preprocessing and Cleaning:**

* Clean the collected data to ensure its quality and accuracy.

**#importing data set**

import pandas as pd

import numpy as np

import sklearn

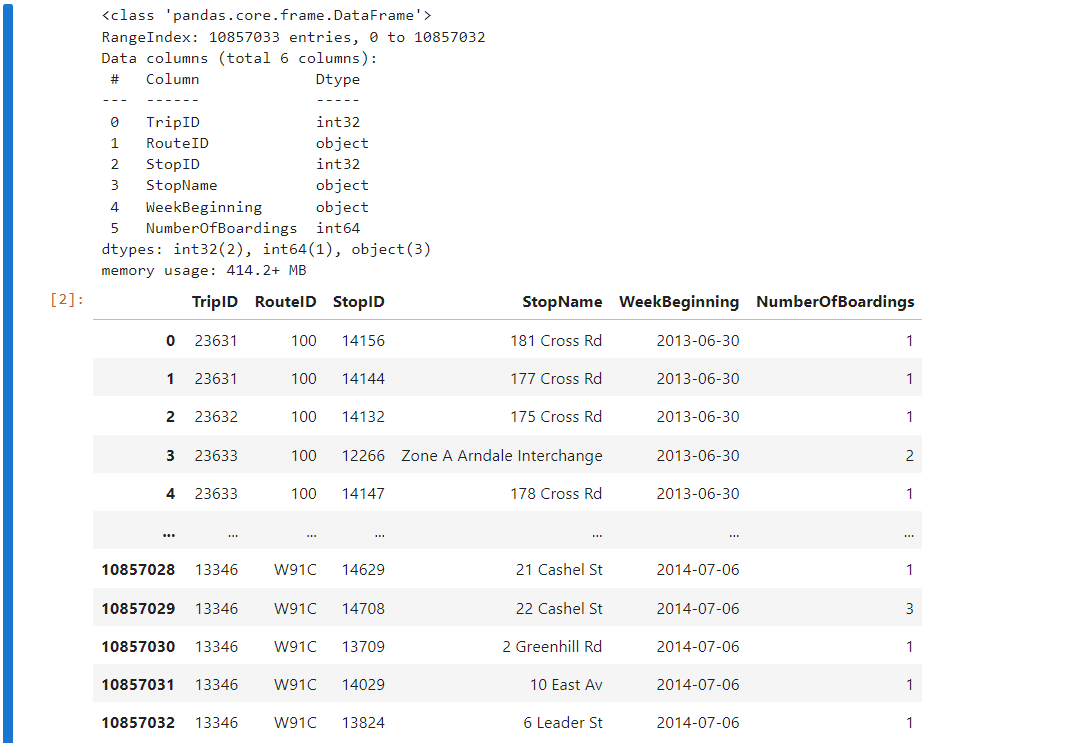
from sklearn.preprocessing import StandardScaler

data = pd.read\_csv("public transport.CSV",dtype={'TripID': int, 'RouteID': str, 'StopID': int, 'WeekBeginning': str})

data.info()

data

**output:**



**#cleansing the data set**

data['StopName'].fillna('Unknown', inplace=True)

data.drop\_duplicates(subset=['TripID', 'StopID', 'WeekBeginning'], keep='first', inplace=True)

data['TripID'] = data['TripID'].astype(int)

data['RouteID'] = data['RouteID'].astype(str)

data['StopID'] = data['StopID'].astype(int)

data['WeekBeginning'] = pd.to\_datetime(data['WeekBeginning'])

#stored the cleaned data into another file

data.to\_csv('cleanddataset.csv', index=False)

**Exploratory Data Analysis (EDA):**

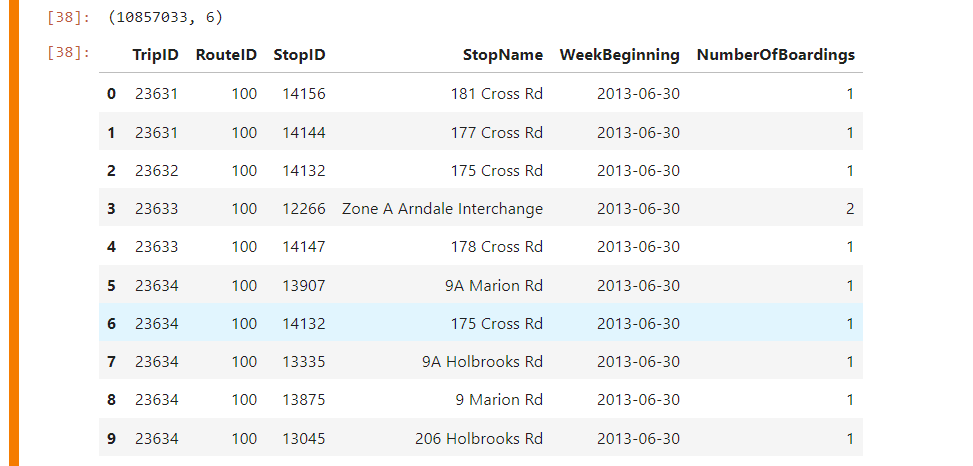
data = pd.read\_csv('cleanddataset.csv')

data.shape

data.head(10)

data.sample(5)

**output:**



data.shape

**output:**

****

data.columns

**output:**



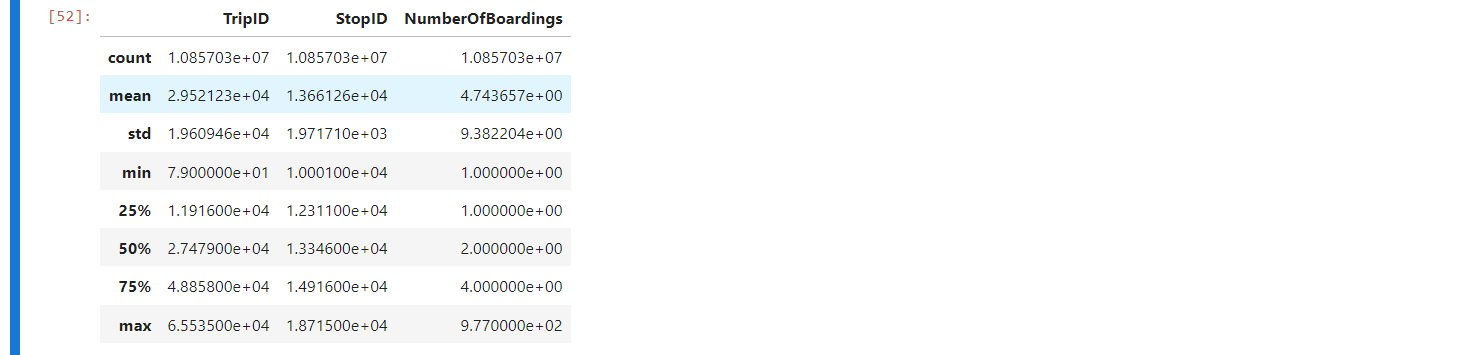
pd.isnull(data).sum()

**output:**



data.describe()

**output:**



data.nunique()

**output:**



**Visualization and Analysis:**

##can assign the each chart to one axes at a time

fig,axrr=plt.subplots(2,2,figsize=(15,15))

ax=axrr[0][0]

ax.set\_title("No of Boardings")

data['NumberOfBoardings'].value\_counts().sort\_index().head(20).plot.bar(ax=axrr[0][0])

ax=axrr[0][1]

ax.set\_title("WeekBeginning")

data['WeekBeginning'].value\_counts().plot.area(ax=axrr[0][1])

ax=axrr[1][0]

ax.set\_title("most Busiest Route")

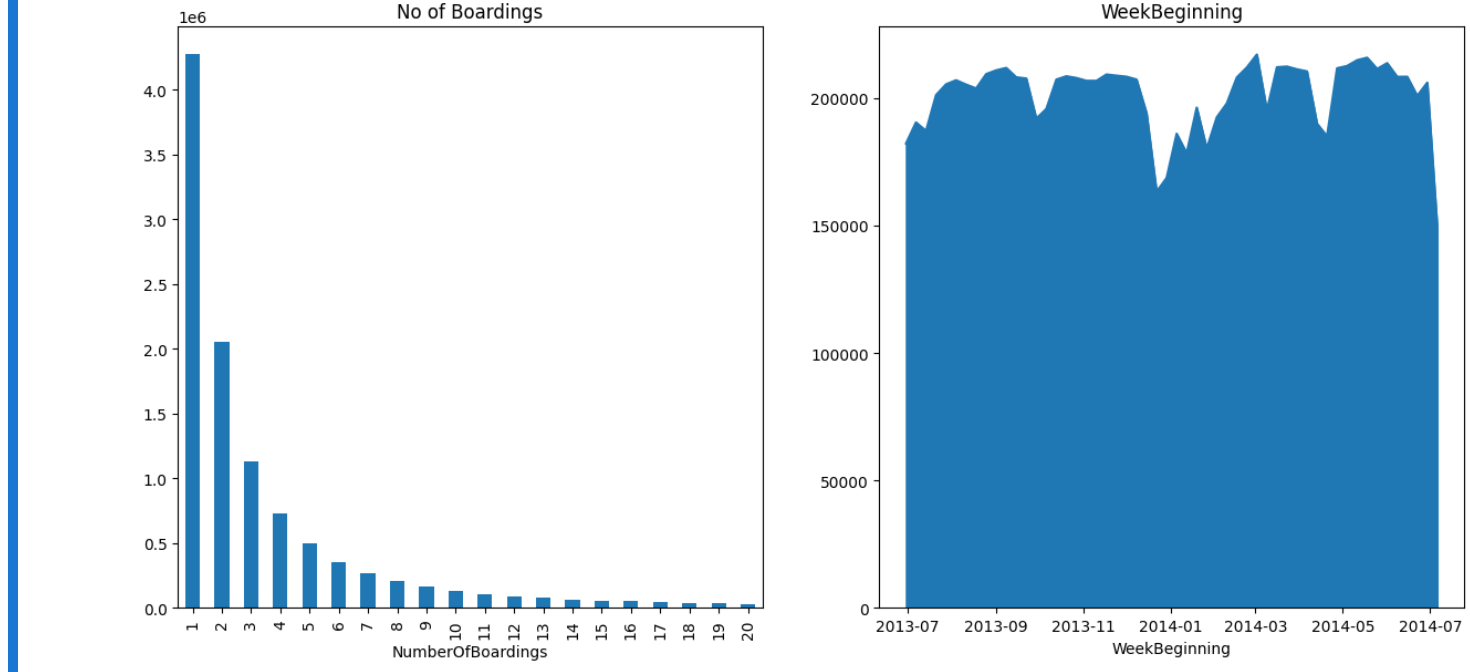
data['RouteID'].value\_counts().head(10).plot.bar(ax=axrr[1][0])

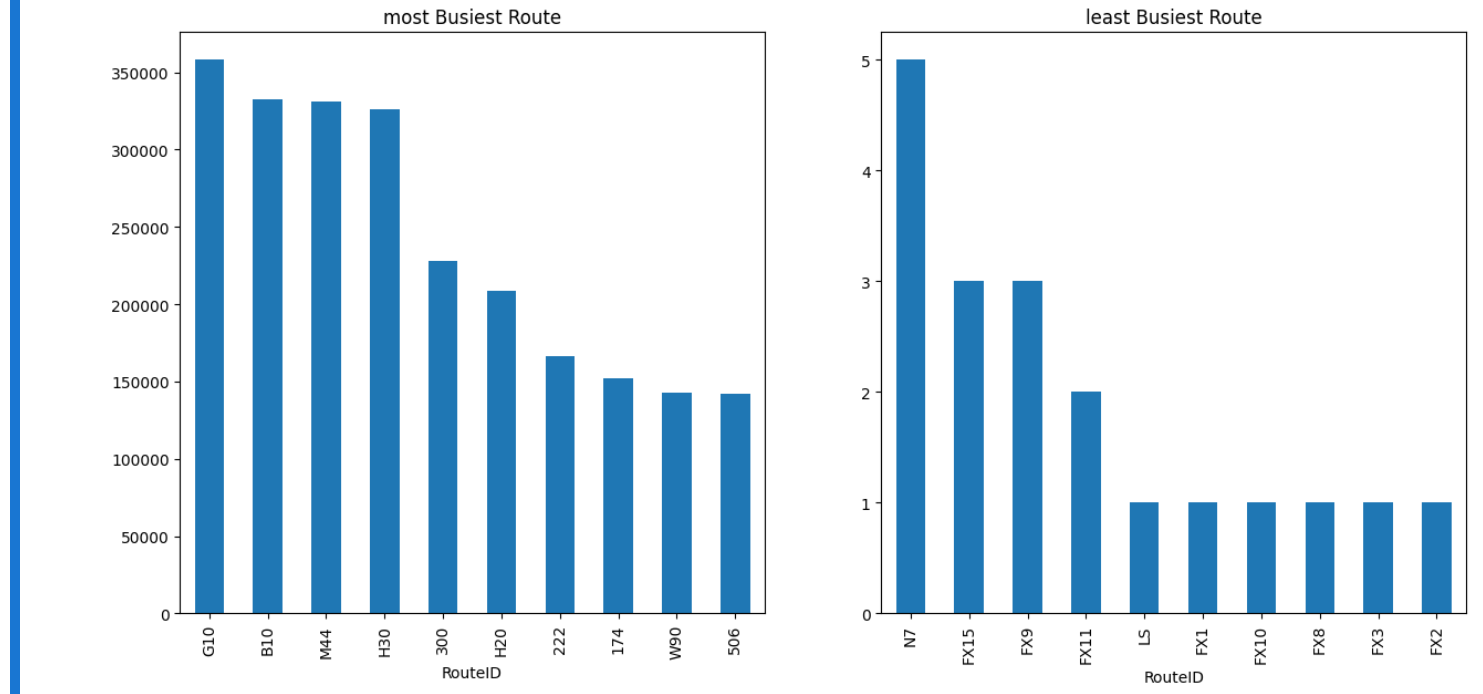
ax=axrr[1][1]

ax.set\_title("least Busiest Route")

data['RouteID'].value\_counts().tail(10).plot.bar(ax=axrr[1][1])

**output:**





**Project Conclusion:**

The public transportation efficiency analysis project aims to provide valuable insights into the performance of public transportation systems. By using IBM Cognos for visualization, we can create interactive and informative dashboards that assist in making informed decisions to enhance public transportation efficiency, ultimately contributing to more sustainable and accessible urban environments.