**PROJECT 8:**

**PUBLIC TRANSPORTATION ANALYSIS**

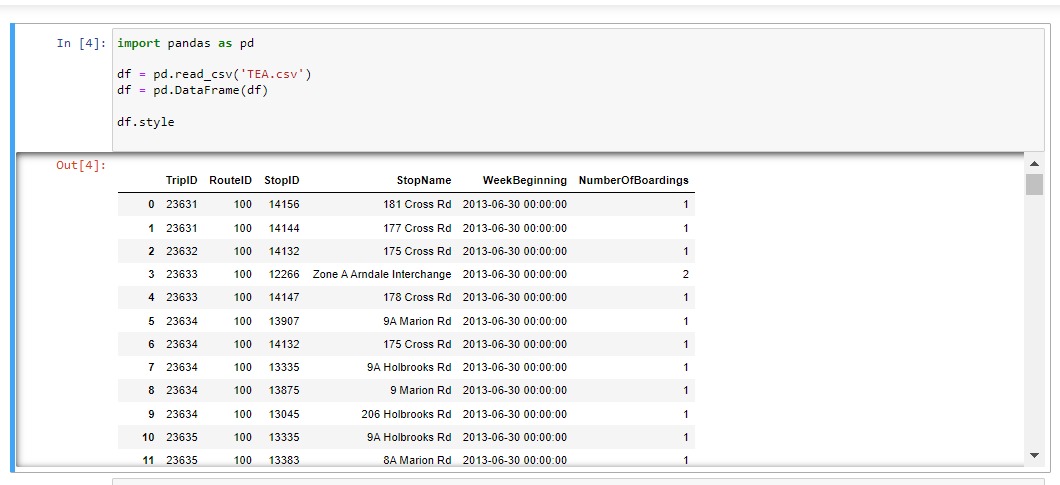
**DEVELOPMENT PART 1**

**LOADING AND PREPROCESSING THE DATASET**

Import pandas as pd

df = pd.read\_csv(‘Transport data’)

df.style

****

Reading of data set for cleaning and preprocessing of data to ensure the accuracy and its quality**.**

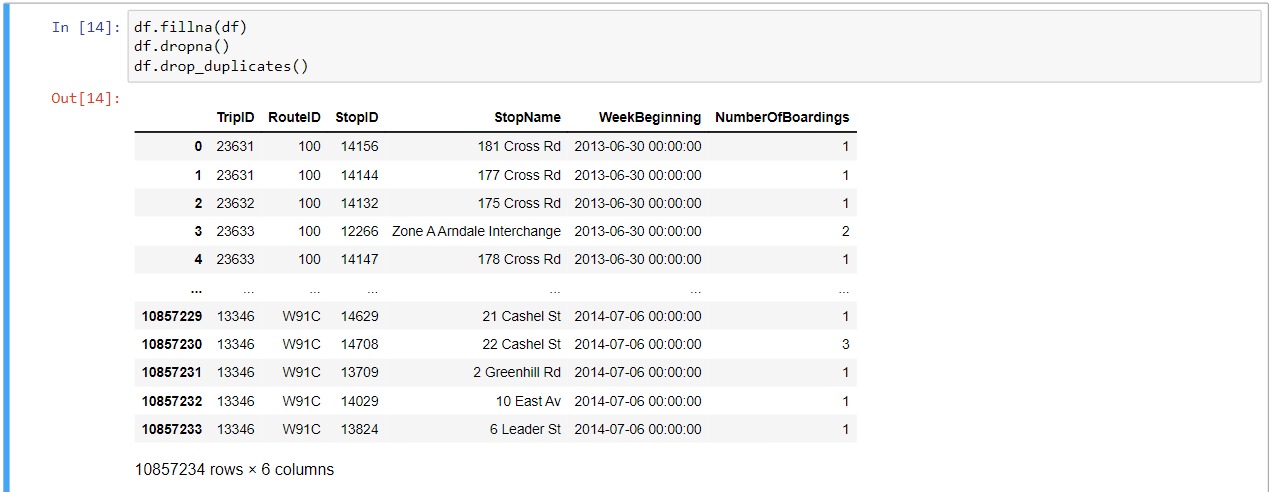
**CLEANING OF DATA AND ENSURING NO REPLICATION’S AND DUPLICATE DATA**

For cleaning of a data set we are going to use the pandas as following:

df.fillna(df)

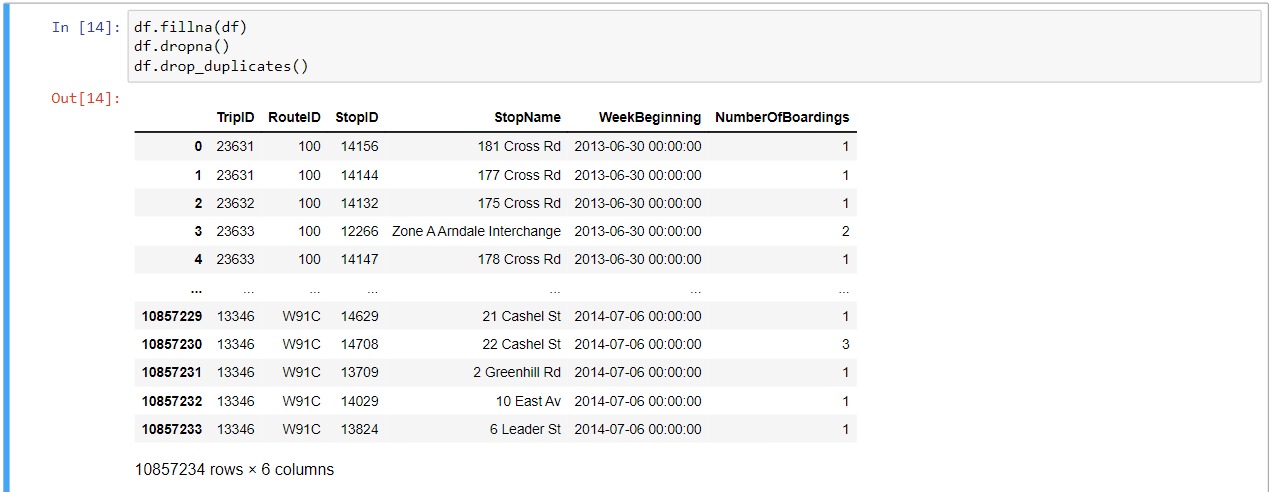
df.dropna()

df.drop\_duplicates()

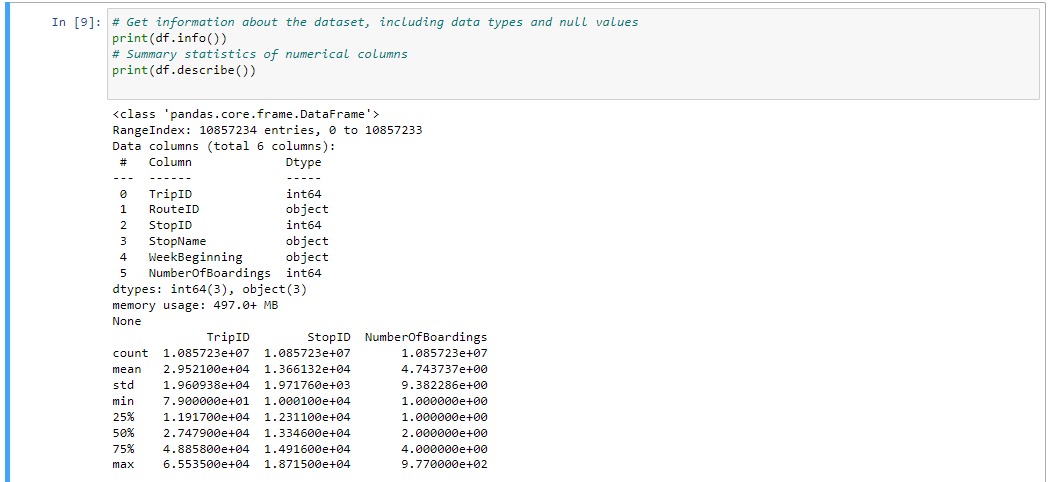


**TO FIND THE DATA TYPES OF THE DOMAINS IN THE DATA SET**

Print(df.dtypes)



**TO GATHER INFORMATIONS ADOUT THE DATA AND DESCRIBING THE DATA SET**

****

**VISUALIZATION OF THE DATA SET USING IBM COGNOS FOR VISUALIZATION**

**1) NO OF BOARDING BY STOP NAME**

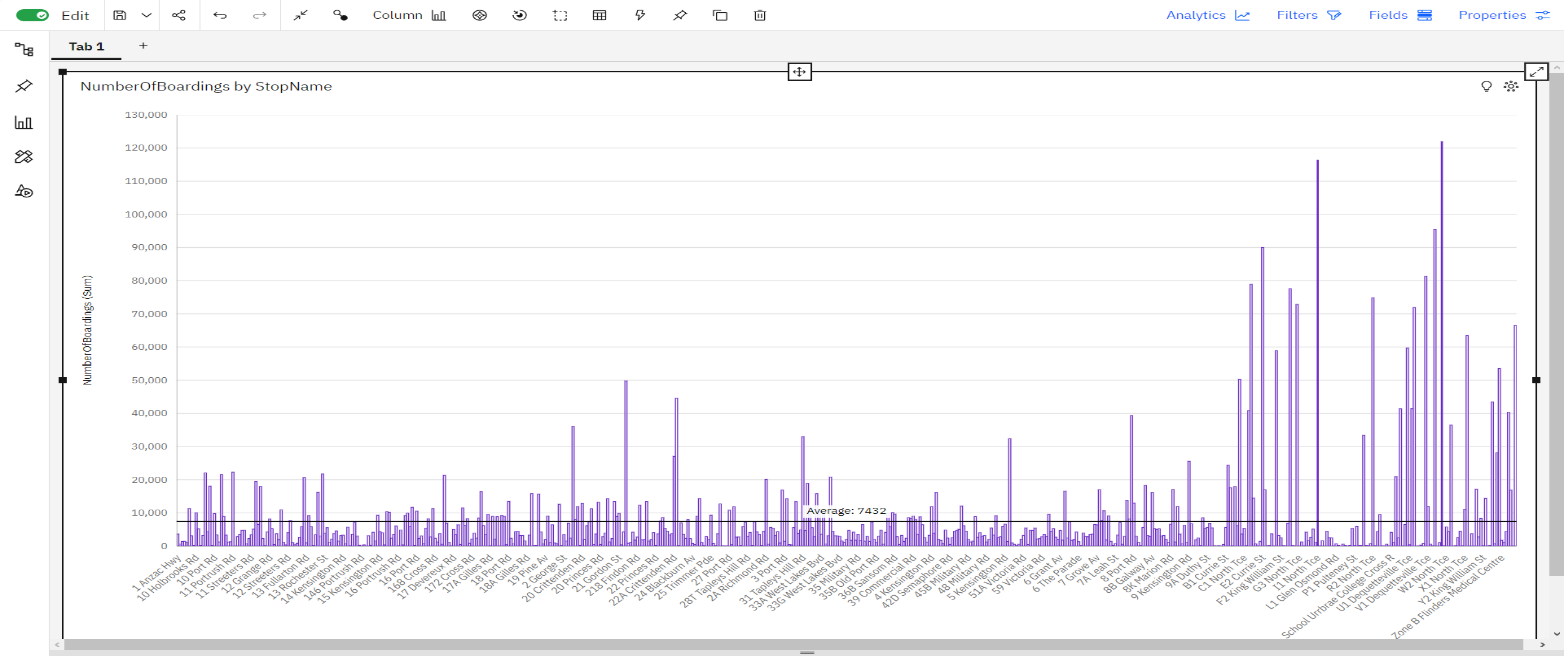
Stop Name W1 North Tce has the highest total Number of Boardings due to Stop ID 13297.

Number Of Boardings is unusually high when Stop Name is W1 North Tce and I1 North Tce.

Across all values of Stop Name, the sum of Number of Boardings is over 4.3 million.

Number Of Boardings ranges from 1, when Stop Name is 11 East Av, to over 122 thousand, when Stop Name is W1 North Tce.

For Number Of Boardings, the most significant values of Stop Name are W1 North Tce and I1 North Tce, whose respective Number Of Boardings values add up to over 238 thousand, or 5.5 % of the total.

****

**2) NO OF BOARDING BY WEEK BEGINNING**

Number Of Boardings is unusually low when Week Beginning is 12/22/2013 0:00 and 12/29/2013 0:00.

Across all values of Week Beginning, the sum of Number Of Boardings is over 4.3 million.

Number Of Boardings ranges from over 37 thousand, when Week Beginning is 12/22/2013 0:00, to almost 99 thousand, when Week Beginning is 3/2/2014 0:00

A screenshot of a graph

Description automatically generated

.

**3)NO OF BOARDINGS BY ROUTE ID**

Route ID 150 has the highest total Number of Boarding due to Stop ID 13297.

Number of Boardings is unusually high when Route ID is 150.

Stop ID 13297 has the highest Number of Boardings at over 122 thousand, out of which Route ID 150 contributed the most at over 45 thousand.

Across all values of Route ID, the sum of Number of Boardings is over 4.3 million.

Number of Boardings ranges from 260, when Route ID is 100S, to nearly 425 thousand, when Route ID is 150

A graph of purple bars

Description automatically generated with medium confidence

**4) NO OF BOARDING BY STOP ID**

Number of Boardings is unusually high when Stop ID is 13297 and 13278.

150 Route ID accounted for 37% of 13297 Number of Boardings compared to 0% for 13278.

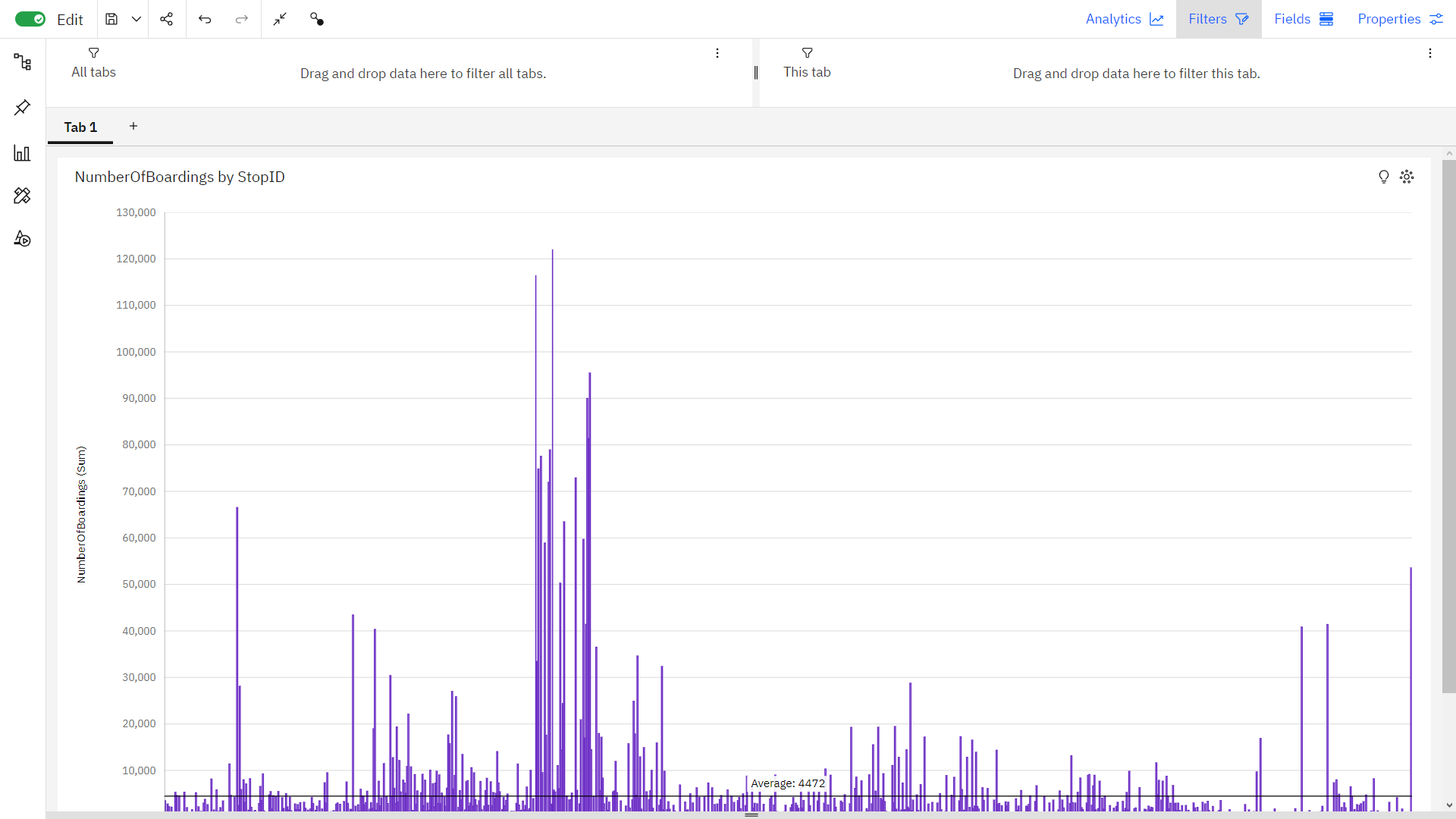
Route ID 150 has the highest Number of Boardings at nearly 425 thousand, out of which Stop ID 13297 contributed the most at over 45 thousand.

Stop ID 13297 has the highest total Number of Boardings due to Route ID 150.

Across all values of Stop ID, the sum of Number of Boardings is over 4.3 million.

Number of Boardings ranges from 1, when Stop ID is 13277, to over 122 thousand, when Stop ID is 13297.

For Number of Boardings, the most significant values of Stop ID are 13297 and 13278, whose respective Number of Boardings values add up to over 238 thousand, or 5.5 % of the total.



**5) NO OF BOARDINGS BY TRIP ID**

A screenshot of a computer

Description automatically generated