**PUBLIC TRANSPORT EFIICIENCY ANALYSIS**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import datetime

import os

from math import sqrt

data = pd.read\_csv('content/drive/MyDrive/phase4.CSV')

data.shape

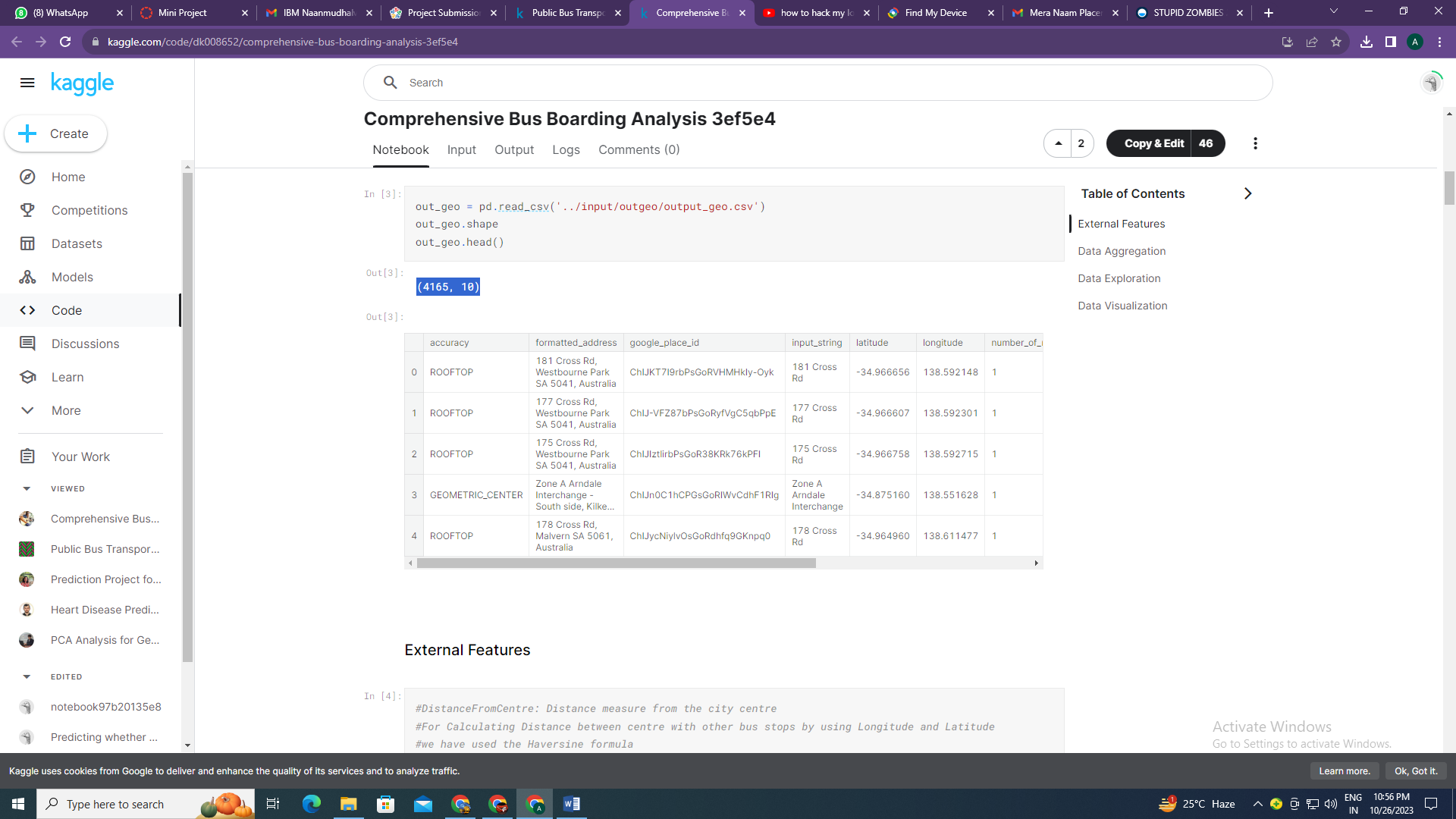
(10857234, 6)

out\_geo = pd.read\_csv('../input/outgeo/output\_geo.csv')

out\_geo.shape

out\_geo.head()

(4165, 10)



from math import sin, cos, sqrt, atan2, radians

def calc\_dist(lat1,lon1):

## approximate radius of earth in km

R = 6373.0

dlon = radians(138.604801) - radians(lon1)

dlat = radians(-34.921247) - radians(lat1)

a = sin(dlat / 2)\*\*2 + cos(radians(lat1)) \* cos(radians(-34.921247)) \* sin(dlon / 2)\*\*2

c = 2 \* atan2(sqrt(a), sqrt(1 - a))

return R \* c

out\_geo['dist\_from\_centre'] = out\_geo[['latitude','longitude']].apply(lambda x: calc\_dist(\*x), axis=1)

out\_geo['type'].fillna('street\_address',inplace=True)

out\_geo['type'] = out\_geo['type'].apply(lambda x: str(x).split(',')[-1])

out\_geo['type'].unique()

array(['street\_address', 'transit\_station', 'premise', 'political',

'school', 'route', 'intersection', 'point\_of\_interest',

'subpremise', 'real\_estate\_agency', 'university', 'travel\_agency',

'restaurant', 'supermarket', 'store', 'post\_office'], dtype=object)

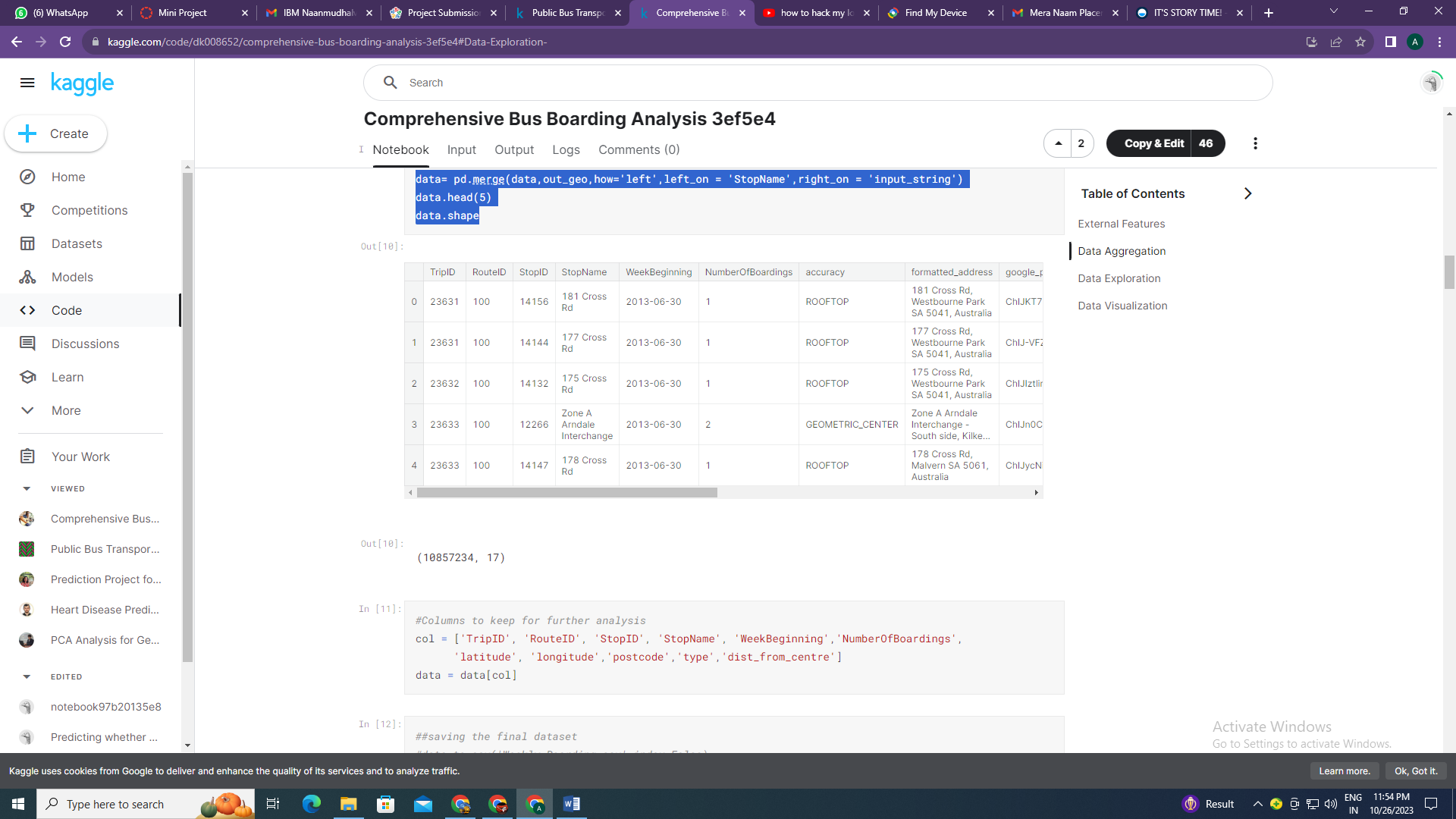
data['WeekBeginning'] = pd.to\_datetime(data['WeekBeginning']).dt.date

data['WeekBeginning'][1]

datetime.date(2013, 6, 30)

data= pd.merge(data,out\_geo,how='left',left\_on = 'StopName',right\_on = 'input\_string')

data.head(5)

data.shape

col = ['TripID', 'RouteID', 'StopID', 'StopName', 'WeekBeginning','NumberOfBoardings',

'latitude', 'longitude','postcode','type','dist\_from\_centre']

data = data[col]

st\_week\_grp = pd.DataFrame(grouped).reset\_index()

st\_week\_grp.shape

st\_week\_grp.head()

(207864, 6)

aa = list(st\_week\_grp1[st\_week\_grp1['WeekBeginning'] == 54]['StopName'])

aa[1:10]

['1 Bartels Rd',

'1 Botanic Rd',

'1 Frome Rd',

'1 Fullarton Rd',

'1 George St',

'1 Glen Osmond Rd',

'1 Goodwood Rd',

'1 Henley Beach Rd',

'1 Kensington Rd']

bb = st\_week\_grp[st\_week\_grp['StopName'].isin(aa)]

bb.head()

bb.shape

type(bb)

(175446, 6)

data.nunique()

TripID 39211

RouteID 616

StopID 5838

StopName 3127

WeekBeginning 54

NumberOfBoardings 359

latitude 2393

longitude 2379

postcode 138

type 8

dist\_from\_centre 2397

dtype: int64

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]

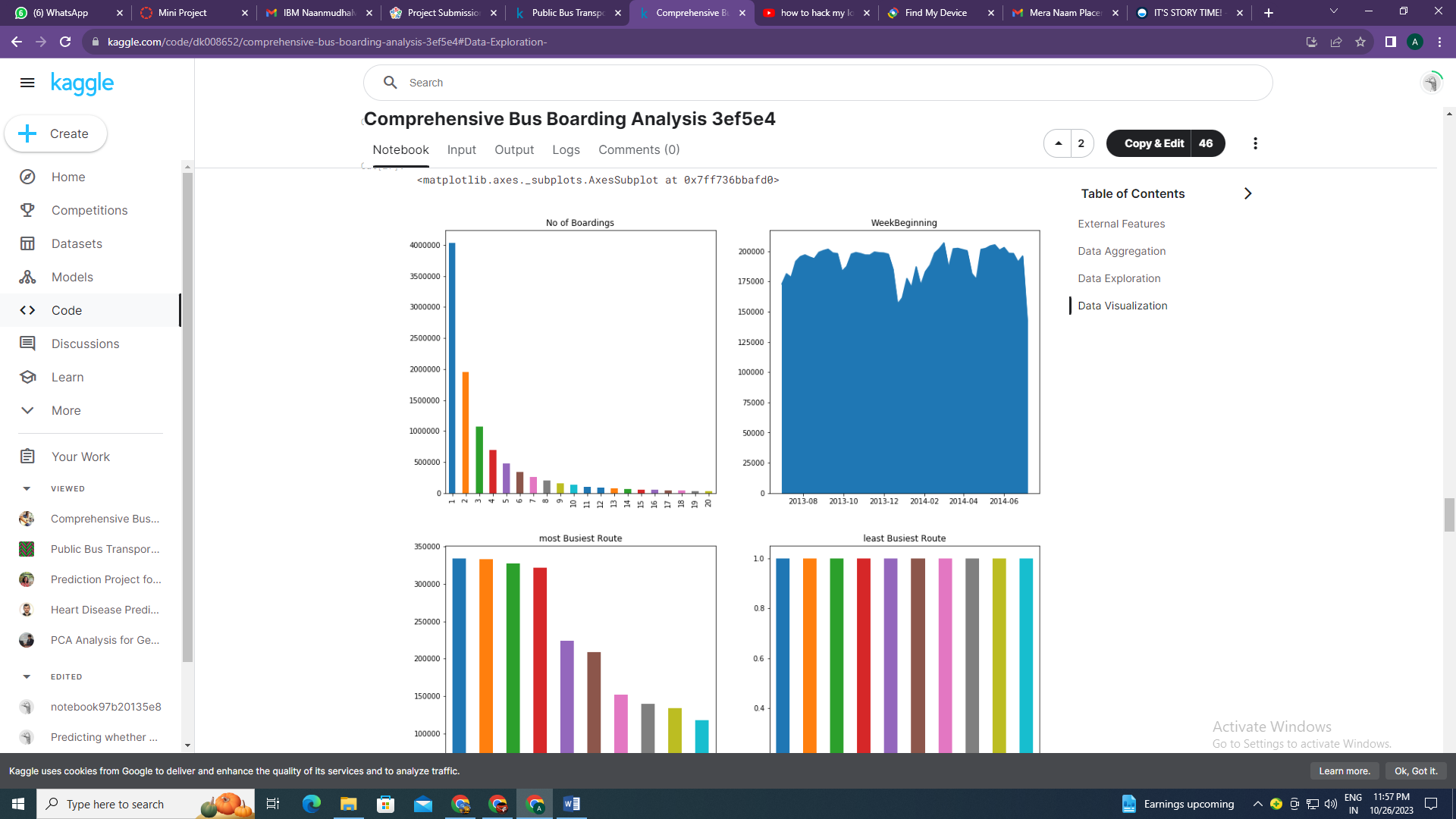
ax.set\_title("most Busiest Route")

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

ax=axrr[1][1]

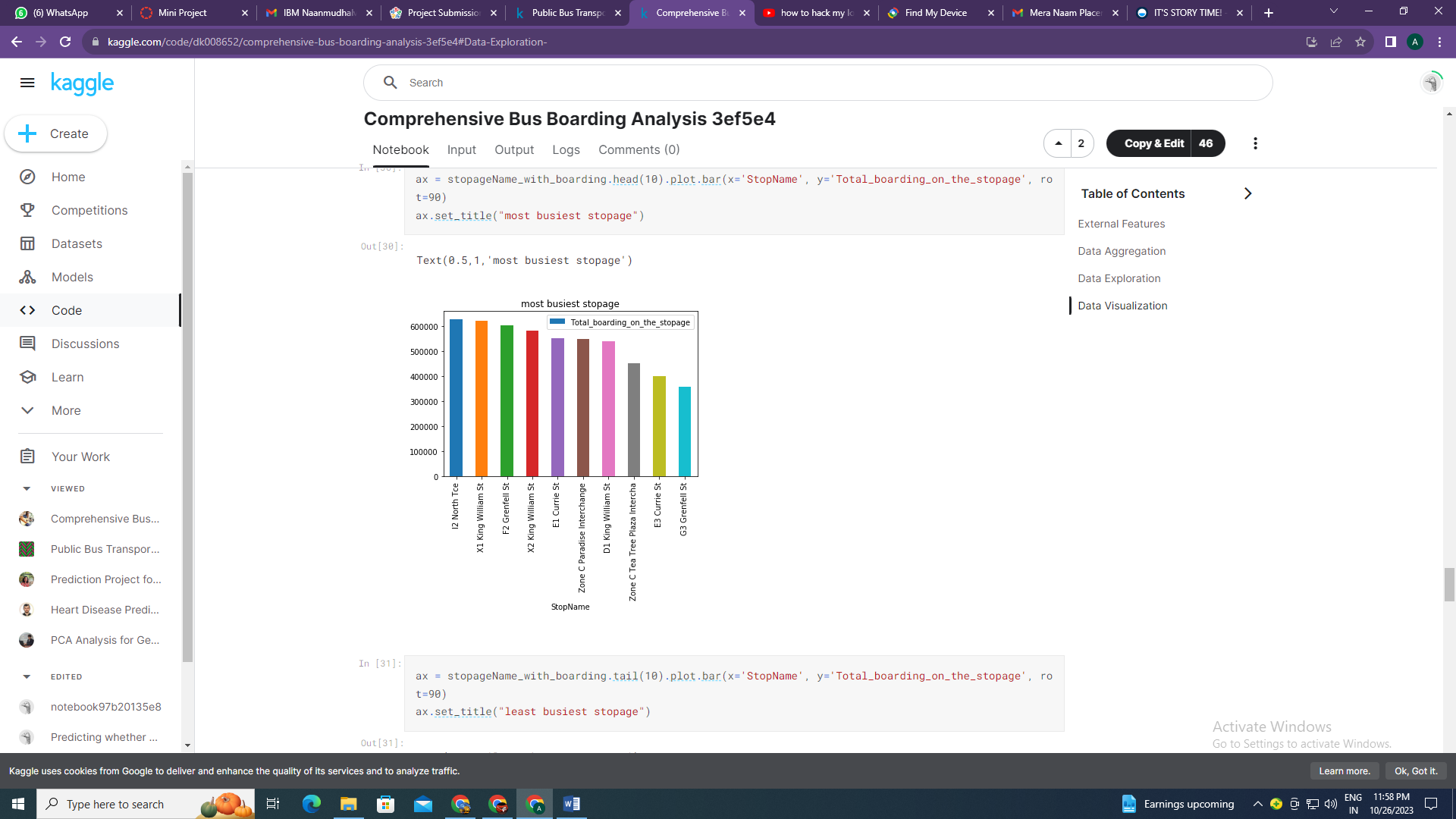
ax.set\_title("least Busiest Route")

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



ax = stopageName\_with\_boarding.head(10).plot.bar(x='StopName', y='Total\_boarding\_on\_the\_stopage', rot=90)

ax.set\_title("most busiest stopage")



data['WeekBeginning'].value\_counts().mean()

191508.66666666666

y = len(distance\_10)+len(distance\_10\_50)+len(distance\_50\_100)

print(total)

print("passangers, boarding the buses in the radious of 10Km from the city center = ", (len(distance\_10)/total)\*100)

print("passanger, boarding the buses from the distance of 10Km to 50Km from the city center = ", (len(distance\_10\_50)/total)\*100)

print("passanger, boarding the buses from the distance of 50Km to 100 from the city center = ", (len(distance\_50\_100)/total)\*100)

10341468

passangers, boarding the buses in the radious of 10Km from the city center = 64.31275521038212

passanger, boarding the buses from the distance of 10Km to 50Km from the city center = 33.16731241638035

passanger, boarding the buses from the distance of 50Km to 100 from the city center = 2.5199323732375327