

PROJECT :PUBLIC TRANSPORTATION EFFICIENCY ANALYSIS

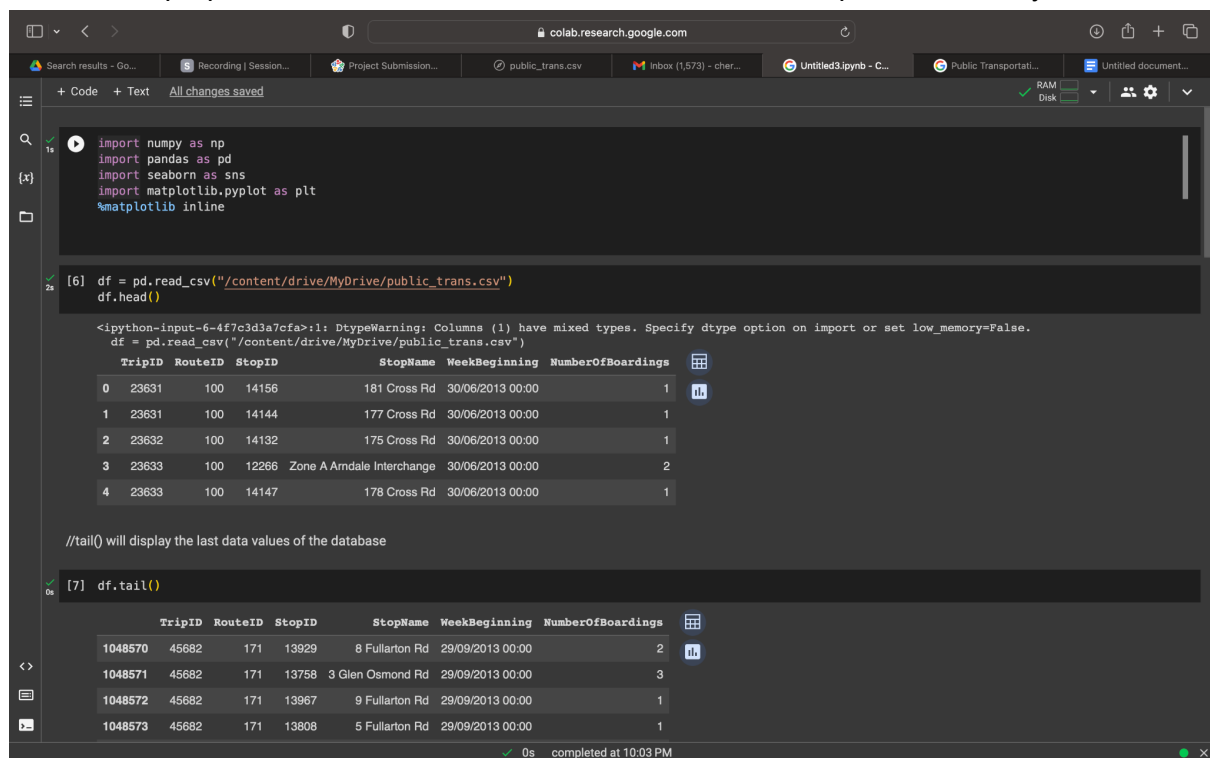
DEVELOPMENT PART- 1

****Step 1: Define Objectives and Scope****

- Clearly define the objectives of your public transportation analysis, such as improving service efficiency, reducing congestion, or increasing accessibility.
- Define the scope of the project, including the geographic area, the modes of transportation to be considered (e.g., buses, subways, trams), and the time frame.

****Step 2: Data Collection and Preprocessing****

- Gather relevant data, including demographic information, traffic data, existing transportation infrastructure, and passenger counts.
- Clean and preprocess the data to ensure it's accurate and compatible for analysis.



The screenshot shows a Google Colab notebook with the following content:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

[6] `df = pd.read_csv("/content/drive/MyDrive/public_trans.csv")`
`df.head()`

<ipython-input-6-4f7c3d3a7cfa>:1: DtypeWarning: Columns (1) have mixed types. Specify dtype option on import or set low_memory=False.

```
df = pd.read_csv("/content/drive/MyDrive/public_trans.csv")
```

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
0	23631	100	14156	181 Cross Rd	30/06/2013 00:00	1
1	23631	100	14144	177 Cross Rd	30/06/2013 00:00	1
2	23632	100	14132	175 Cross Rd	30/06/2013 00:00	1
3	23633	100	12266	Zone A Arndale Interchange	30/06/2013 00:00	2
4	23633	100	14147	178 Cross Rd	30/06/2013 00:00	1

//tail() will display the last data values of the database

[7] `df.tail()`

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
1048570	45682	171	13929	8 Fullarton Rd	29/09/2013 00:00	2
1048571	45682	171	13758	3 Glen Osmond Rd	29/09/2013 00:00	3
1048572	45682	171	13967	9 Fullarton Rd	29/09/2013 00:00	1
1048573	45682	171	13808	5 Fullarton Rd	29/09/2013 00:00	1

0s completed at 10:03 PM

Execute the commands to clean and process data one by one.

colab.research.google.com

Search results - Go... Recording | Session... Project Submission... public_trans.csv Inbox (1,573) - cher... Untitled3.ipynb - C... Public Transportati... Untitled document...

+ Code + Text All changes saved

RAM Disk

```
[6] df = pd.read_csv("/content/drive/MyDrive/public_trans.csv")
df.head()
```

<ipython-input-6-4f7c3d3a7cfa>:1: DtypeWarning: Columns (1) have mixed types. Specify dtype option on import or set low_memory=False.

```
df = pd.read_csv("/content/drive/MyDrive/public_trans.csv")
```

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
0	23631	100	14156	181 Cross Rd	30/06/2013 00:00	1
1	23631	100	14144	177 Cross Rd	30/06/2013 00:00	1
2	23632	100	14132	175 Cross Rd	30/06/2013 00:00	1
3	23633	100	12266	Zone A Armdale Interchange	30/06/2013 00:00	2
4	23633	100	14147	178 Cross Rd	30/06/2013 00:00	1

//tail() will display the last data values of the database

```
df.tail()
```

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
1048570	45682	171	13929	8 Fullarton Rd	29/09/2013 00:00	2
1048571	45682	171	13758	3 Glen Osmond Rd	29/09/2013 00:00	3
1048572	45682	171	13967	9 Fullarton Rd	29/09/2013 00:00	1
1048573	45682	171	13808	5 Fullarton Rd	29/09/2013 00:00	1
1048574	45682	171	13845	6 Fullarton Rd	29/09/2013 00:00	3

//isnull() will display whether there is any null data value of the dataset//

```
[8] df.isnull()
```

0s completed at 10:03 PM

colab.research.google.com

Search results - Go... Recording | Session... Project Submission... public_trans.csv Inbox (1,573) - cher... Untitled3.ipynb - C... Public Transportati... Untitled document...

+ Code + Text All changes saved

RAM Disk

//tail() will display the last data values of the database

```
[7] df.tail()
```

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
1048570	45682	171	13929	8 Fullarton Rd	29/09/2013 00:00	2
1048571	45682	171	13758	3 Glen Osmond Rd	29/09/2013 00:00	3
1048572	45682	171	13967	9 Fullarton Rd	29/09/2013 00:00	1
1048573	45682	171	13808	5 Fullarton Rd	29/09/2013 00:00	1
1048574	45682	171	13845	6 Fullarton Rd	29/09/2013 00:00	3

//isnull() will display whether there is any null data value of the dataset//

```
df.isnull()
```

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...
1048570	False	False	False	False	False	False
1048571	False	False	False	False	False	False
1048572	False	False	False	False	False	False

0s completed at 10:03 PM

```
colab.research.google.com

Search results - Go... Recording | Session... Project Submission... public_trans.csv Inbox (1,573) - cher... Untitled3.ipynb - C... Public Transportati... Untitled document...

+ Code + Text All changes saved

//isnull() will display whether there is any null data value of the dataset//

df.isnull()

TripID RouteID StopID StopName WeekBeginning NumberOfBoardings
0 False False False False False False
1 False False False False False False
2 False False False False False False
3 False False False False False False
4 False False False False False False
... ..
1048570 False False False False False False
1048571 False False False False False False
1048572 False False False False False False
1048573 False False False False False False
1048574 False False False False False False
1048575 rows x 6 columns

//fillna() will fill the null values with 0//

df.isnull().sum()

TripID      0
RouteID     0
StopID      0
StopName    0
WeekBeginning 0
NumberOfBoardings 0

0s completed at 10:03 PM
```

```
colab.research.google.com

Search results - Go... Recording | Session... Project Submission... public_trans.csv Inbox (1,573) - cher... Untitled3.ipynb - C... Public Transportati... Untitled document...

+ Code + Text All changes saved

//fillna() will fill the null values with 0//

df.isnull().sum()

TripID      0
RouteID     0
StopID      0
StopName    0
WeekBeginning 0
NumberOfBoardings 0
dtype: int64

[10] df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1048575 entries, 0 to 1048574
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   TripID                1048575 non-null  int64
1   RouteID               1048575 non-null  object
2   StopID                1048575 non-null  int64
3   StopName              1048575 non-null  object
4   WeekBeginning         1048575 non-null  object
5   NumberOfBoardings     1048575 non-null  int64
dtypes: int64(3), object(3)
memory usage: 48.0+ MB

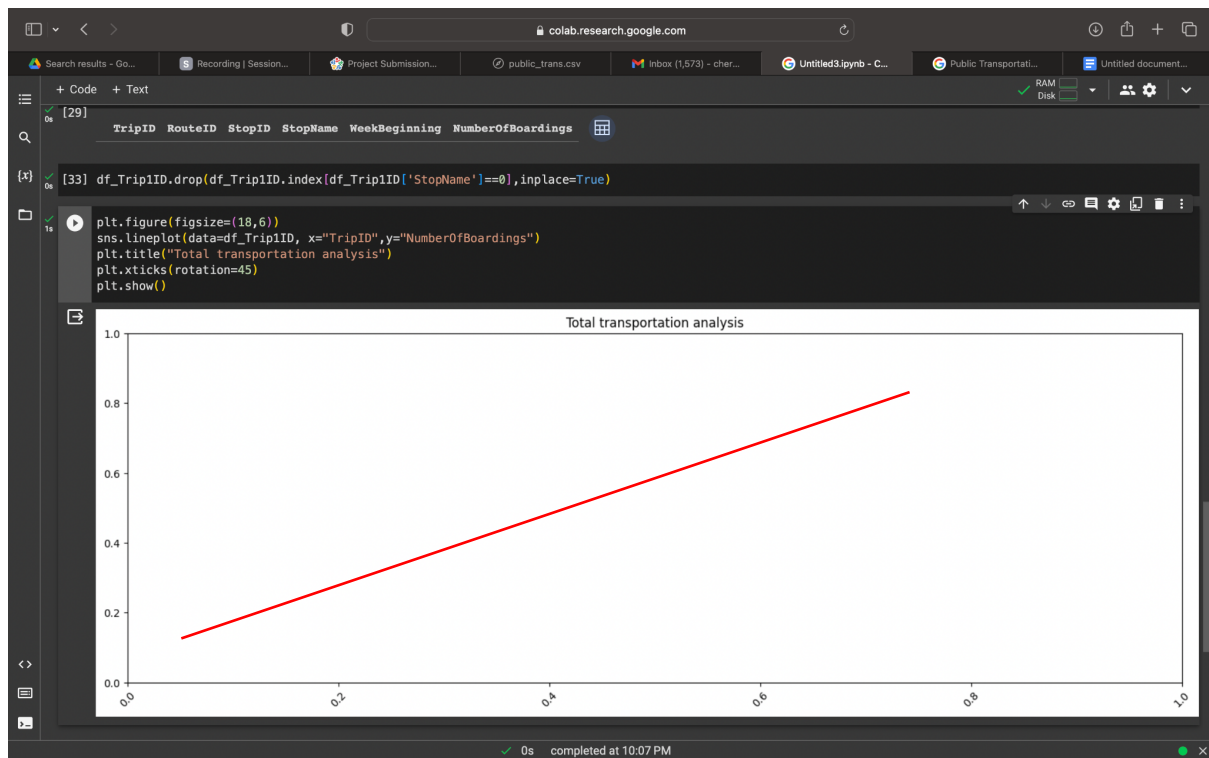
//the data given in the dataset is converted into a specific format//

df_Trip1ID=df[df["TripID"]=="Trip1ID"].copy()
df_Trip1ID

TripID RouteID StopID StopName WeekBeginning NumberOfBoardings

[33] df_Trip1ID.drop(df_Trip1ID.index[df_Trip1ID['StopName']==0], inplace=True)

0s completed at 10:03 PM
```



Step 3: Performance Measurement and Evaluation

- After implementation, continuously monitor the performance of the transportation system.
- Use key performance indicators (KPIs) to evaluate the effectiveness of the project.