Date - 17/10/2023

Team ID - 4502

Project Title - PUBLIC TRANSPORTATION EFFICIENCY ANALYSIS

```
from google.colab import drive
drive.mount('/content/drive')
```

Import Dependencies

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

Load Dataset

```
dataset = pd.read_csv("/content/drive/MyDrive/20140711 (1).CSV")
```

```
<ipython-input-4-ac3655479c6a>:1: DtypeWarning: Columns (1) have mixed types. Specify dtype option on import or set low_memory=False
  dataset = pd.read_csv("/content/drive/MyDrive/20140711 (1).CSV")
```

Data Eploration

```
dataset
```

	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
...	...	...	...	...	...	...
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

1048575 rows × 6 columns

```
dataset.head()
```

	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
				Zone A Arndale	30-06-2013	

```
dataset.shape
```

```
(1048575, 6)
```

```
dataset.columns
```

```
Index(['TripID', 'RouteID', 'StopID', 'StopName', 'WeekBeginning',
       'NumberOfBoardings'],
      dtype='object')
```

```
dataset.dtypes
```

```
TripID                int64
RouteID              object
StopID                int64
StopName             object
WeekBeginning        object
NumberOfBoardings     int64
dtype: object
```

Data Preprocessing

```
dataset.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 × 6 columns

```
dataset.isnull().sum()
```

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

```
dataset.isnull().sum().sum()
```

```
0
```

```
dataset.describe()
```

	TripID	StopID	NumberOfBoardings
count	1.048575e+06	1.048575e+06	1.048575e+06
mean	2.860299e+04	1.330114e+04	4.132290e+00
std	1.674656e+04	1.119243e+03	6.291338e+00
min	3.017000e+03	1.081700e+04	1.000000e+00
25%	1.162200e+04	1.269800e+04	1.000000e+00
50%	3.423400e+04	1.333500e+04	2.000000e+00
75%	4.512600e+04	1.371600e+04	4.000000e+00
max	6.258500e+04	1.849300e+04	1.930000e+02

```
dataset.describe(include='all')
```

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoarding
count	1.048575e+06	1048575	1.048575e+06	1048575	1048575	1.048575e+(
unique	NaN	36	NaN	583	54	Na
top	NaN	157	NaN	I1 North Tce	08-09-2013 00:00	Na
freq	NaN	95087	NaN	12678	21417	Na
mean	2.860299e+04	NaN	1.330114e+04	NaN	NaN	4.132290e+(
std	1.674656e+04	NaN	1.119243e+03	NaN	NaN	6.291338e+(
min	3.017000e+03	NaN	1.081700e+04	NaN	NaN	1.000000e+(
25%	1.162200e+04	NaN	1.269800e+04	NaN	NaN	1.000000e+(

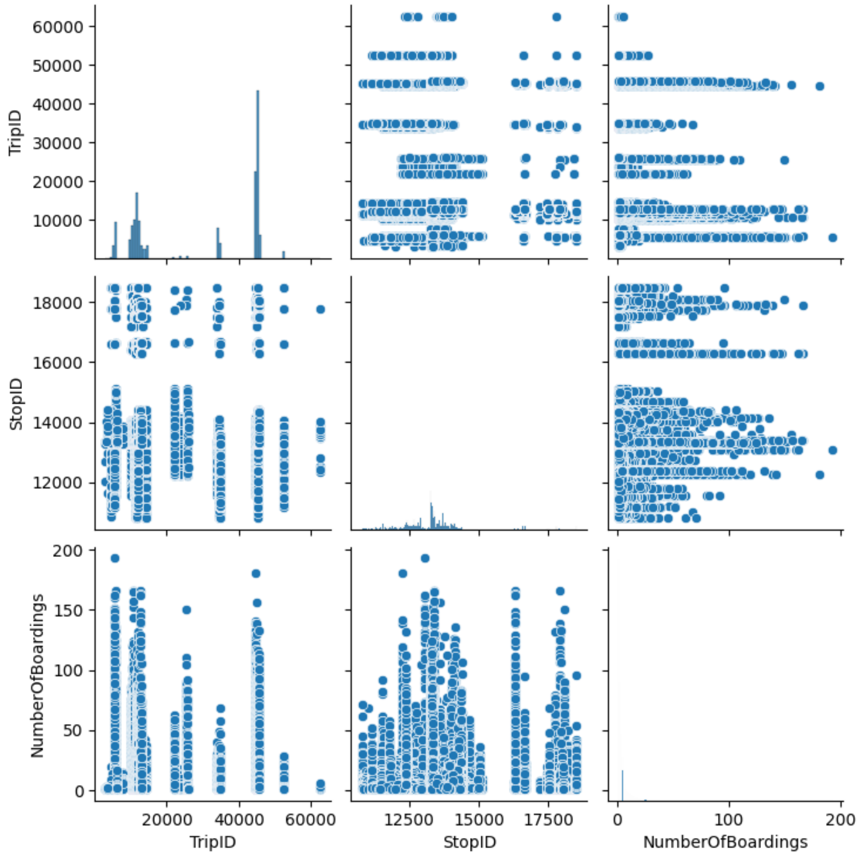
```
dataset.info()
```

```
<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
```

Data Visualization

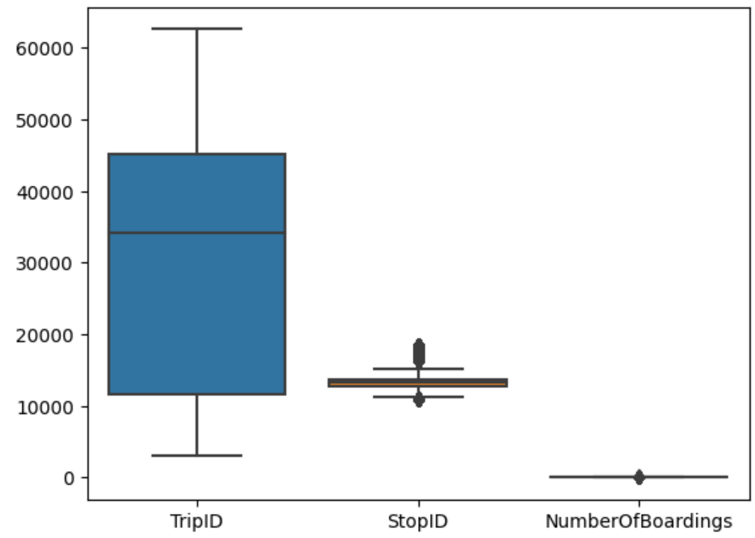
```
plt.figure(figsize=(20,20))
sns.pairplot(dataset)
```

```
<seaborn.axisgrid.PairGrid at 0x7fd0a00f2ef0>
<Figure size 2000x2000 with 0 Axes>
```



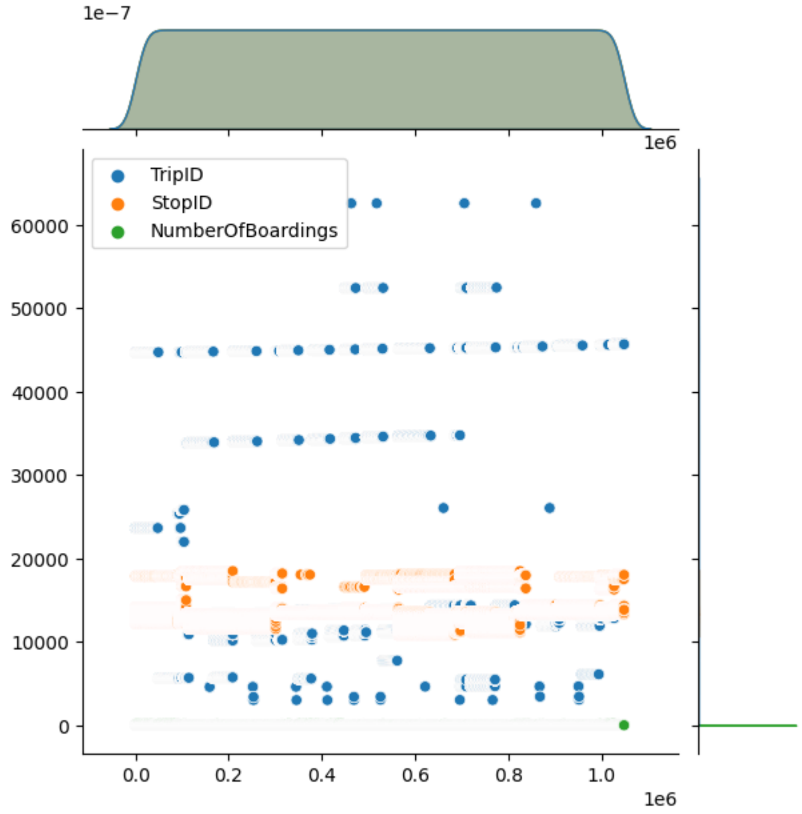
```
sns.boxplot(dataset)
```

```
<Axes: >
```



```
sns.jointplot(dataset)
```

```
<seaborn.axisgrid.JointGrid at 0x7fd09d71fb20>
```



Correlation visualiztion

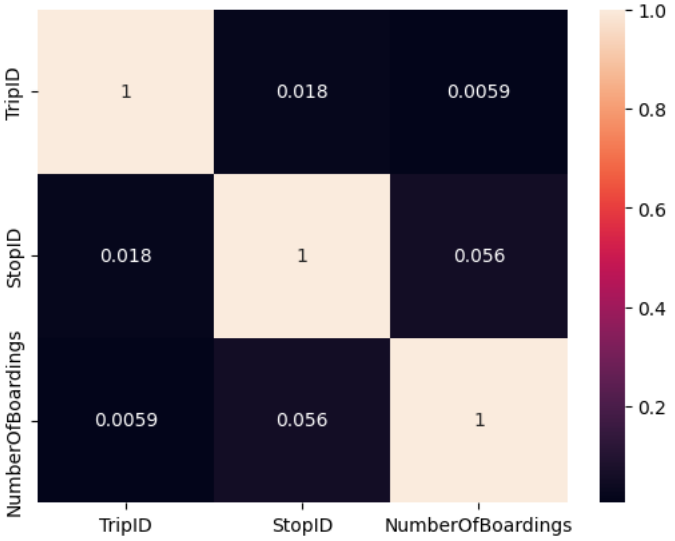
```
dataset.corr()
```

```
<ipython-input-27-c187c74d1e71>:1: FutureWarning: The default value of numeric_only i
  dataset.corr()
```

	TripID	StopID	NumberOfBoardings
TripID	1.000000	0.017946	0.005864
StopID	0.017946	1.000000	0.056094
NumberOfBoardings	0.005864	0.056094	1.000000

```
sns.heatmap(dataset.corr(),annot=True)
```

```
<ipython-input-25-9d3fd451b567>:1: FutureWarning: The default value of numeric_only i
  sns.heatmap(dataset.corr(),annot=True)
<Axes: >
```


TripID
StopID
NumberOfBoardings
TripID
StopID
NumberOfBoardings