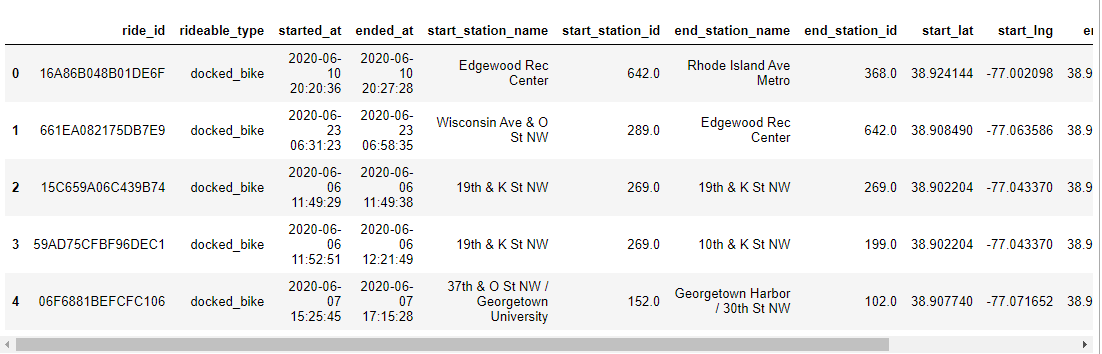
**Scope Of Project:**

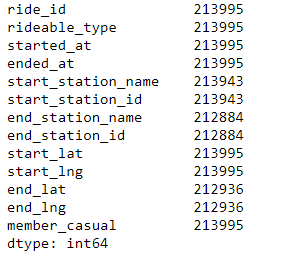
Data preprocessing is important step in while building machine learning model.In this Project we need to perform different Pre-processing Techniques to get data ready for building models.

1. import numpy as np
2. import pandas as pd
3. import seaborn as sns
4. from sklearn.preprocessing import LabelEncoder
5. from sklearn.model\_selection import train\_test\_split
6. from sklearn.preprocessing import StandardScaler

1. dataset=pd.read\_csv('202006-capitalbikeshare-tripdata.csv')
2. dataset.head()



1. dataset.count()



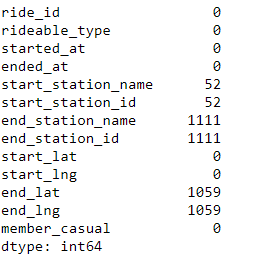
1. dataset['rideable\_type'].unique()



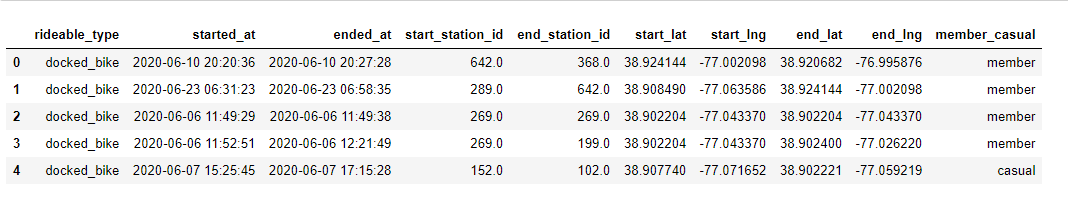
1. dataset['member\_casual'].unique()



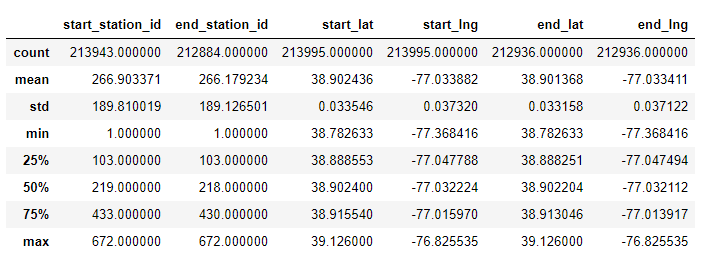
1. dataset.isnull().sum()



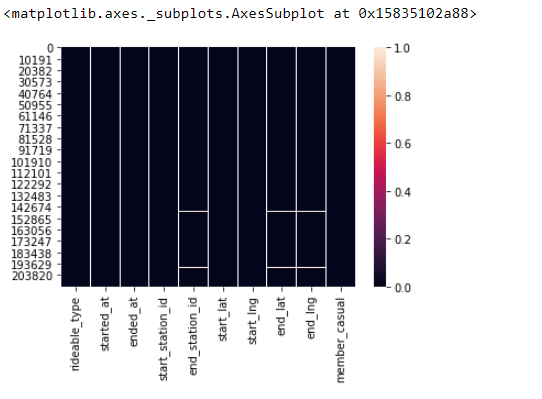
1. dataset=dataset.drop(['start\_station\_name','end\_station\_name','ride\_id'],axis=1)
2. dataset.head()



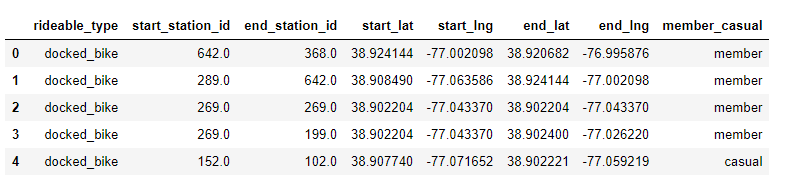
1. dataset.describe()



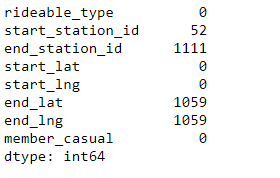
1. sns.heatmap(dataset.isnull())



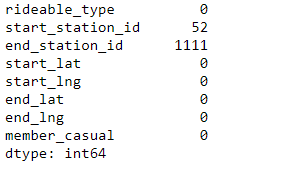
1. dataset=dataset.drop(['started\_at','ended\_at'],axis=1)
2. dataset.head()



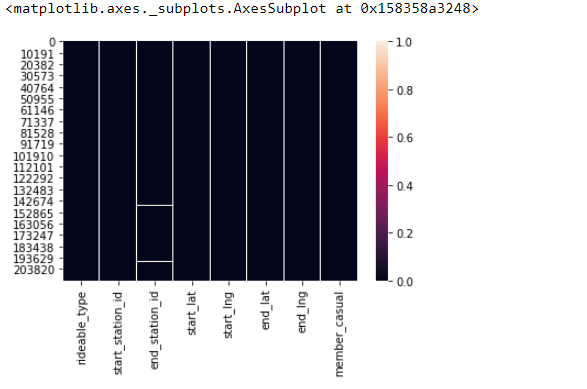
1. dataset.isnull().sum()



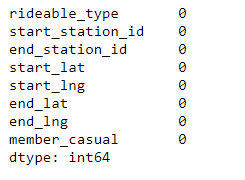
1. dataset['end\_lat']=dataset['end\_lat'].fillna(np.mean(dataset['end\_lat']))
2. dataset['end\_lng']=dataset['end\_lng'].fillna(np.mean(dataset['end\_lng']))
3. dataset.isnull().sum()



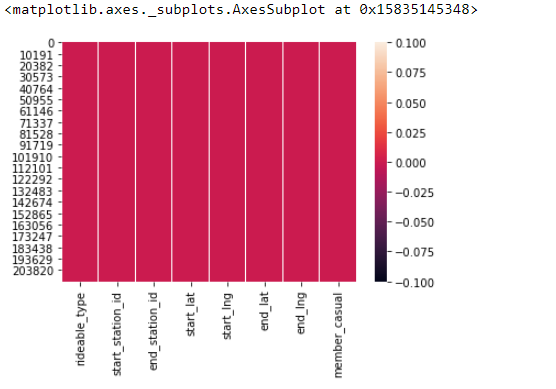
1. sns.heatmap(dataset.isnull())



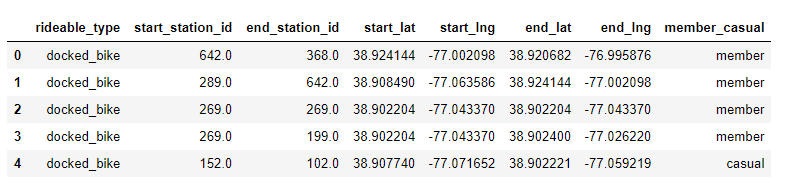
1. dataset['start\_station\_id']=dataset['start\_station\_id'].fillna(dataset['start\_station\_id'].value\_counts().index[0])
2. dataset['end\_station\_id']=dataset['end\_station\_id'].fillna(dataset['end\_station\_id'].value\_counts().index[0])
3. dataset.isnull().sum()



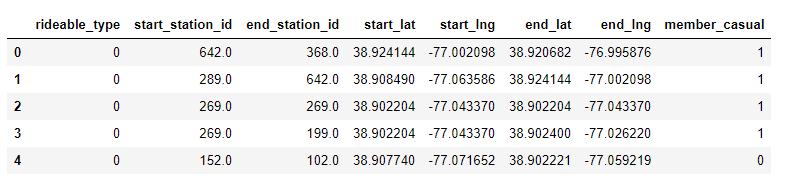
1. sns.heatmap(dataset.isnull())



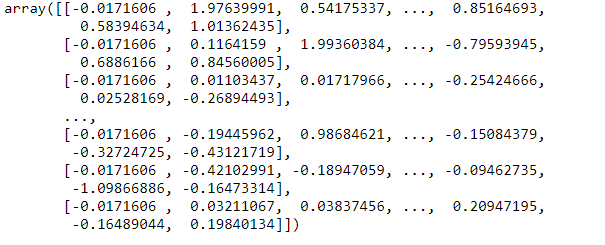
1. dataset.head()



1. enc=LabelEncoder()
2. dataset['rideable\_type']=enc.fit\_transform(dataset['rideable\_type'])
3. dataset['member\_casual']=enc.fit\_transform(dataset['member\_casual'])
4. dataset.head()



1. x\_data=dataset[['rideable\_type','start\_station\_id','end\_station\_id','start\_lat','start\_lng','end\_lat','end\_lng']]
2. y\_data=dataset[['member\_casual']]
3. x\_data=np.array(x\_data)
4. y\_data=np.array(y\_data)
5. ss=StandardScaler()
6. x\_data=ss.fit\_transform(x\_data)
7. x\_data



1. X\_train,X\_test,Y\_train,Y\_test=train\_test\_split(x\_data,y\_data,test\_size=0.2)
2. print(X\_test.shape)
3. print(X\_train.shape)
4. print(Y\_test.shape)
5. print(Y\_train.shape)

