Project Development Phase – Sprint 4

Team ID	PNT2022TMID18963
Team Lead	Kavya S
Team Members	Poorvasindu Karthick M P Lavakumar
Project Name	A new hint to transportation – Analysis of the NYC bike share system.
Maximum Marks	20 Marks

```python
path = "/content/cleaned_dataset.csv"
edadf = pd.read_csv(path)
print(edadf)
```

```
       tripduration           starttime            stoptime  \
0               695 2013-06-01 00:00:01 2013-06-01 00:11:36
1               693 2013-06-01 00:00:08 2013-06-01 00:11:41
2              2059 2013-06-01 00:00:44 2013-06-01 00:35:03
3               123 2013-06-01 00:01:04 2013-06-01 00:03:07
4              1521 2013-06-01 00:01:22 2013-06-01 00:26:43
...             ...                 ...                 ...
84086          1506 2013-06-08 14:35:22 2013-06-08 15:00:28
84087          1478 2013-06-08 14:35:22 2013-06-08 15:00:00
84088           873 2013-06-08 14:35:31 2013-06-08 14:50:04
84089          2054 2013-06-08 14:34:51 2013-06-08 15:09:05
84090          1179 2013-06-08 14:35:28 2013-06-08 14:55:07

       start station id        start station name  start station latitude  \
0                 444.0           Broadway & W 24 St              40.742354
1                 444.0           Broadway & W 24 St              40.742354
2                 406.0      Hicks St & Montague St              40.695128
3                 475.0          E 15 St & Irving Pl              40.735243
4                2008.0       Little West St & 1 Pl              40.705693
...                 ...                       ...                    ...
84086             422.0            W 59 St & 10 Ave              40.770513
84087             337.0         Old Slip & Front St              40.703799
84088             447.0             8 Ave & W 52 St              40.763707
84089             229.0              Great Jones St              40.727434
84090             259.0     South St & Whitehall St              40.701221
```

```
        start station longitude    end station id              end station name  \
0                     -73.989151            434.0               9 Ave & W 18 St
1                     -73.989151            434.0               9 Ave & W 18 St
2                     -73.995951            406.0          Hicks St & Montague St
3                     -73.987586            262.0                 Washington Park
4                     -74.016777            310.0               State St & Smith St
...                          ...              ...                           ...
84086                 -73.988038            212.0          W 16 St & The High Line
84087                 -74.008387            342.0       Columbia St & Rivington St
84088                 -73.985162            404.0               9 Ave & W 14 St
84089                 -73.993790           2012.0                 E 27 St & 1 Ave
84090                 -74.012342            383.0       Greenwich Ave & Charles St

        end station latitude    end station longitude      bikeid       usertype  \
0                  40.743174               -74.003664     19678.0     Subscriber
1                  40.743174               -74.003664     16649.0     Subscriber
2                  40.695128               -73.995951     19599.0       Customer
3                  40.691782               -73.973730     16352.0     Subscriber
4                  40.689269               -73.989129     15567.0     Subscriber
...                      ...                      ...         ...            ...
84086              40.743349               -74.006818     19225.0       Customer
84087              40.717400               -73.980166     19730.0       Customer
84088              40.740583               -74.005509     15606.0       Customer
84089              40.739445               -73.976806     18597.0     Subscriber
84090              40.735238               -74.000271     14665.0     Subscriber

        birth year   gender
0           1983.0      1.0
1           1984.0      1.0
2              NaN      0.0
3           1960.0      1.0
4           1983.0      1.0
...            ...      ...
84086          NaN      0.0
84087          NaN      0.0
84088          NaN      0.0
84089       1977.0      1.0
84090       1968.0      2.0

[84091 rows x 15 columns]
```

temp = edadf

temp.head()

	tripduration	starttime	stoptime	start station id	start station name	start station latitude	start station longitude	end station id	end station name	end station latitude	end station longitude	bikeid	usertype	birth year	gender
0	695	2013-06-01 00:00:01	2013-06-01 00:11:36	444.0	Broadway & W 24 St	40.742354	-73.989151	434.0	9 Ave & W 18 St	40.743174	-74.003664	19678.0	Subscriber	1983.0	1.0
1	693	2013-06-01 00:00:08	2013-06-01 00:11:41	444.0	Broadway & W 24 St	40.742354	-73.989151	434.0	9 Ave & W 18 St	40.743174	-74.003664	16649.0	Subscriber	1984.0	1.0
2	2059	2013-06-01 00:00:44	2013-06-01 00:35:03	406.0	Hicks St & Montague St	40.695128	-73.995951	406.0	Hicks St & Montague St	40.695128	-73.995951	19599.0	Customer	NaN	0.0
3	123	2013-06-01 00:01:04	2013-06-01 00:03:07	475.0	E 15 St & Irving Pl	40.735243	-73.987586	262.0	Washington Park	40.691782	-73.973730	16352.0	Subscriber	1960.0	1.0
4	1521	2013-06-01 00:01:22	2013-06-01 00:26:43	2008.0	Little West St & 1 Pl	40.705693	-74.016777	310.0	State St & Smith St	40.689269	-73.989129	15567.0	Subscriber	1983.0	1.0

```python
temp.describe()
```

	tripduration	start station id	start station latitude	start station longitude	end station id	end station latitude	end station longitude	bikeid	birth year	gender
count	8.409100e+04	84091.000000	84091.000000	84091.000000	81248.000000	81248.000000	81248.000000	84091.000000	43168.000000	84091.000000
mean	2.240389e+03	449.589813	40.731815	-73.990975	451.956479	40.731570	-73.990982	17604.454698	1973.738857	0.622124
std	1.948179e+04	359.983426	0.021619	0.013123	368.024074	0.021642	0.013242	1661.417010	10.937208	0.673987
min	6.100000e+01	72.000000	40.646607	-74.017134	72.000000	40.646607	-74.017134	14529.000000	1899.000000	0.000000
25%	5.540000e+02	312.000000	40.716059	-74.001547	310.000000	40.716021	-74.001547	16178.000000	1966.000000	0.000000
50%	9.580000e+02	404.000000	40.734011	-73.990697	404.000000	40.733812	-73.990741	17620.000000	1976.000000	1.000000
75%	1.509000e+03	483.000000	40.747804	-73.981923	482.000000	40.747659	-73.981923	19075.000000	1982.000000	1.000000
max	3.876479e+06	3020.000000	40.771522	-73.950048	3020.000000	40.771522	-73.950048	20564.000000	1997.000000	2.000000

```python
temp.info()
```

```
RangeIndex: 84091 entries, 0 to 84090
Data columns (total 15 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   tripduration           84091 non-null  int64
 1   starttime              84091 non-null  object
 2   stoptime               84091 non-null  object
 3   start station id       84091 non-null  float64
 4   start station name     84091 non-null  object
 5   start station latitude 84091 non-null  float64
 6   start station longitude 84091 non-null  float64
 7   end station id         81248 non-null  float64
 8   end station name       81248 non-null  object
 9   end station latitude   81248 non-null  float64
 10  end station longitude  81248 non-null  float64
 11  bikeid                 84091 non-null  float64
 12  usertype               84091 non-null  object
 13  birth year             43168 non-null  float64
 14  gender                 84091 non-null  float64
dtypes: float64(9), int64(1), object(5)
memory usage: 9.6+ MB
```

```python
temp["starttime"] = pd.to_datetime(temp["starttime"])
temp["stoptime"] = pd.to_datetime(temp["stoptime"])
temp.info()
temp["Hour"] = temp["stoptime"].dt.hour - temp["starttime"].dt.hour
temp.head()
```

```python
temp.shape
```

```
(84091, 16)
```

```python
temp['Age'] = 2022 - temp['birth year']
temp.head()
```

	tripduration	starttime	stoptime	start station id	start station name	start station latitude	start station longitude	end station id	end station name	end station latitude	end station longitude	bikeid	usertype	birth year	gender	Hour	Age
0	695	2013-06-01 00:00:01	2013-06-01 00:11:36	444.0	Broadway & W 24 St	40.742354	-73.989151	434.0	9 Ave & W 18 St	40.743174	-74.003664	19678.0	Subscriber	1983.0	1.0	0	39.0
1	693	2013-06-01 00:00:08	2013-06-01 00:11:41	444.0	Broadway & W 24 St	40.742354	-73.989151	434.0	9 Ave & W 18 St	40.743174	-74.003664	16649.0	Subscriber	1984.0	1.0	0	38.0
2	2059	2013-06-01 00:00:44	2013-06-01 00:35:03	406.0	Hicks St & Montague St	40.695128	-73.995951	406.0	Hicks St & Montague St	40.695128	-73.995951	19599.0	Customer	NaN	0.0	0	NaN
3	123	2013-06-01 00:01:04	2013-06-01 00:03:07	475.0	E 15 St & Irving Pl	40.735243	-73.987586	262.0	Washington Park	40.691782	-73.973730	16352.0	Subscriber	1960.0	1.0	0	62.0
4	1521	2013-06-01 00:01:22	2013-06-01 00:26:43	2008.0	Little West St & 1 Pl	40.705693	-74.016777	310.0	State St & Smith St	40.689269	-73.989129	15567.0	Subscriber	1983.0	1.0	0	39.0

```python
Age_Groups = ["<20", "20-29", "30-39", "40-49", "50-59", "60+"]
Age_Groups_Limits = [0, 20, 30, 40, 50, 60, np.inf]
Age_Min = 0
Age_Max = 100
temp["Age_group"] = pd.cut(temp["Age"], Age_Groups_Limits, labels=Age_Groups)
temp.head()
```

	tripduration	starttime	stoptime	start station id	start station name	start station latitude	start station longitude	end station id	end station name	end station latitude	end station longitude	bikeid	usertype	birth year	gender	Hour	Age
0	695	2013-06-01 00:00:01	2013-06-01 00:11:36	444.0	Broadway & W 24 St	40.742354	-73.989151	434.0	9 Ave & W 18 St	40.743174	-74.003664	19678.0	Subscriber	1983.0	1.0	0	39.0
1	693	2013-06-01 00:00:08	2013-06-01 00:11:41	444.0	Broadway & W 24 St	40.742354	-73.989151	434.0	9 Ave & W 18 St	40.743174	-74.003664	16649.0	Subscriber	1984.0	1.0	0	38.0
2	2059	2013-06-01 00:00:44	2013-06-01 00:35:03	406.0	Hicks St & Montague St	40.695128	-73.995951	406.0	Hicks St & Montague St	40.695128	-73.995951	19599.0	Customer	NaN	0.0	0	NaN
3	123	2013-06-01 00:01:04	2013-06-01 00:03:07	475.0	E 15 St & Irving Pl	40.735243	-73.987586	262.0	Washington Park	40.691782	-73.973730	16352.0	Subscriber	1960.0	1.0	0	62.0
4	1521	2013-06-01 00:01:22	2013-06-01 00:26:43	2008.0	Little West St & 1 Pl	40.705693	-74.016777	310.0	State St & Smith St	40.689269	-73.989129	15567.0	Subscriber	1983.0	1.0	0	39.0

```
trips_df = pd.DataFrame()
trips_df = temp.groupby(['start station name','end station name']).size().reset_index(name = 'Number of Trips')
trips_df = trips_df.sort_values('Number of Trips',ascending = False)
trips_df["start station name"] = trips_df["start station name"].astype(str)
trips_df["end station name"] = trips_df["end station name"].astype(str)
trips_df["Routes"] = trips_df["start station name"] + " to " + trips_df["end station name"]
trips_df = trips_df[:50]
trips_df = trips_df.reset_index()
trips_df
```

	index	start station name	end station name	Number of Trips	Routes
0	8964	Central Park S & 6 Ave	Central Park S & 6 Ave	153	Central Park S & 6 Ave to Central Park S & 6 Ave
1	7521	Broadway & W 58 St	Broadway & W 58 St	114	Broadway & W 58 St to Broadway & W 58 St
2	37933	West Thames St	West Thames St	64	West Thames St to West Thames St
3	3990	Atlantic Ave & Furman St	Old Fulton St	58	Atlantic Ave & Furman St to Old Fulton St
4	36141	W 56 St & 6 Ave	W 56 St & 6 Ave	57	W 56 St & 6 Ave to W 56 St & 6 Ave
5	7341	Broadway & W 56 St	Broadway & W 56 St	55	Broadway & W 56 St to Broadway & W 56 St
6	9101	Centre St & Chambers St	Centre St & Chambers St	48	Centre St & Chambers St to Centre St & Chamber...
7	25989	Old Fulton St	Old Fulton St	48	Old Fulton St to Old Fulton St
8	10855	Dean St & 4 Ave	Dean St & 4 Ave	46	Dean St & 4 Ave to Dean St & 4 Ave
9	3945	Atlantic Ave & Furman St	Atlantic Ave & Furman St	46	Atlantic Ave & Furman St to Atlantic Ave & Fur...
10	5161	Broadway & Battery Pl	Broadway & Battery Pl	43	Broadway & Battery Pl to Broadway & Battery Pl
11	20859	Greenwich St & North Moore St	Greenwich St & North Moore St	41	Greenwich St & North Moore St to Greenwich St ...
12	7804	Bus Slip & State St	Bus Slip & State St	39	Bus Slip & State St to Bus Slip & State St
13	25155	Metropolitan Ave & Bedford Ave	Metropolitan Ave & Bedford Ave	39	Metropolitan Ave & Bedford Ave to Metropolitan...
14	25928	Old Fulton St	Atlantic Ave & Furman St	39	Old Fulton St to Atlantic Ave & Furman St
15	38258	Wythe Ave & Metropolitan Ave	Wythe Ave & Metropolitan Ave	36	Wythe Ave & Metropolitan Ave to Wythe Ave & Me...
16	13086	E 17 St & Broadway	E 17 St & Broadway	35	E 17 St & Broadway to E 17 St & Broadway
17	4275	Bank St & Washington St	Bank St & Washington St	34	Bank St & Washington St to Bank St & Washingto...

18	28075	South St & Gouverneur Ln	South St & Gouverneur Ln	33	South St & Gouverneur Ln to South St & Gouvern...
19	4576	Barrow St & Hudson St	Barrow St & Hudson St	32	Barrow St & Hudson St to Barrow St & Hudson St
20	9521	Christopher St & Greenwich St	Christopher St & Greenwich St	31	Christopher St & Greenwich St to Christopher S...
21	7518	Broadway & W 58 St	Broadway & W 51 St	31	Broadway & W 58 St to Broadway & W 51 St
22	37730	West St & Chambers St	West St & Chambers St	31	West St & Chambers St to West St & Chambers St
23	15754	E 39 St & 2 Ave	2 Ave & E 31 St	30	E 39 St & 2 Ave to 2 Ave & E 31 St
24	30225	W 14 St & The High Line	W 14 St & The High Line	29	W 14 St & The High Line to W 14 St & The High ...
25	38065	William St & Pine St	William St & Pine St	29	William St & Pine St to William St & Pine St
26	1005	12 Ave & W 40 St	12 Ave & W 40 St	29	12 Ave & W 40 St to 12 Ave & W 40 St
27	8959	Central Park S & 6 Ave	Broadway & W 58 St	29	Central Park S & 6 Ave to Broadway & W 58 St
28	19612	Front St & Washington St	Front St & Washington St	28	Front St & Washington St to Front St & Washing...
29	5885	Broadway & W 24 St	E 17 St & Broadway	28	Broadway & W 24 St to E 17 St & Broadway
30	31107	W 20 St & 11 Ave	W 20 St & 11 Ave	28	W 20 St & 11 Ave to W 20 St & 11 Ave
31	4115	Bank St & Hudson St	Bank St & Hudson St	27	Bank St & Hudson St to Bank St & Hudson St
32	21265	Greenwich St & Warren St	West Thames St	27	Greenwich St & Warren St to West Thames St
33	5305	Broadway & Berry St	Broadway & Berry St	26	Broadway & Berry St to Broadway & Berry St
34	5398	Broadway & E 14 St	Broadway & E 14 St	26	Broadway & E 14 St to Broadway & E 14 St
35	18105	E 58 St & 3 Ave	E 58 St & 3 Ave	26	E 58 St & 3 Ave to E 58 St & 3 Ave
36	7342	Broadway & W 56 St	Broadway & W 58 St	25	Broadway & W 56 St to Broadway & W 58 St
37	7013	Broadway & W 51 St	Broadway & W 51 St	25	Broadway & W 51 St to Broadway & W 51 St
38	753	11 Ave & W 41 St	11 Ave & W 41 St	25	11 Ave & W 41 St to 11 Ave & W 41 St
39	22335	John St & William St	John St & William St	25	John St & William St to John St & William St
40	10427	Clinton St & Joralemon St	Clinton St & Joralemon St	25	Clinton St & Joralemon St to Clinton St & Jora...

41	37589	West St & Chambers St	Bus Slip & State St	24	West St & Chambers St to Bus Slip & State St
42	868	11 Ave & W 59 St	11 Ave & W 59 St	24	11 Ave & W 59 St to 11 Ave & W 59 St
43	29584	Vesey Pl & River Terrace	Vesey Pl & River Terrace	24	Vesey Pl & River Terrace to Vesey Pl & River T...
44	5841	Broadway & W 24 St	Broadway & W 24 St	24	Broadway & W 24 St to Broadway & W 24 St
45	34286	W 43 St & 10 Ave	W 43 St & 10 Ave	23	W 43 St & 10 Ave to W 43 St & 10 Ave
46	2527	9 Ave & W 14 St	9 Ave & W 14 St	23	9 Ave & W 14 St to 9 Ave & W 14 St
47	1138	12 Ave & W 40 St	West St & Chambers St	23	12 Ave & W 40 St to West St & Chambers St
48	1146	2 Ave & E 31 St	2 Ave & E 31 St	23	2 Ave & E 31 St to 2 Ave & E 31 St
49	25513	Murray St & West St	Murray St & West St	23	Murray St & West St to Murray St & West St

```python
px.pie(values = temp['gender'].value_counts(),
       names =temp['gender'].value_counts().index,
       title ="Gender Variation")
```

```python
px.bar(x=temp["start station name"].value_counts().index,
       y=temp["start station name"].value_counts().values,
       labels={'x':'Start Station Name',"y":"Count"})
```

```python
px.bar(x=temp["end station name"].value_counts().index,
       y=temp["end station name"].value_counts().values,
       labels={'x':'End Station Name',"y":"Count"})
```

```python
px.bar(x=temp["Hour"].value_counts().index,
       y=temp["Hour"].value_counts().values,
       title = "Hour usage of Citi Bikes",
       labels={'x':'Time',"y":"Number of people using bike"})
```