In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
PATH_LOAD1 = "us-101.csv"
df1 = pd.read_csv(PATH_LOAD1)
print(df1.shape)
```

```
(4802933, 9)
```

In [3]:
```
PATH_LOAD2 = "lankershim.csv"
df2 = pd.read_csv(PATH_LOAD2)
print(df2.shape)
```

```
(1607319, 9)
```

In [4]:
```
df1.head(1)
```

Out[4]:

	Unnamed: 0	Vehicle_ID	Global_Time	Real_Time	Local_X	Local_Y	v_Vel	Lane_ID	Movement
0	0	5	1118846979700	2005-06-15 10:49:39.7	39.788	39.154	40.0	4	NaN

In [5]:
```
df2.head(1)
```

Out[5]:

	Unnamed: 0	Vehicle_ID	Global_Time	Real_Time	Local_X	Local_Y	v_Vel	Lane_ID	Movement
0	0	63	1118935680200	2005-06-16 11:28:00.2	44.589	67.404	0.0	2	3.0

In [6]:
```
df1.tail(1)
```

Out[6]:

	Unnamed: 0	Vehicle_ID	Global_Time	Real_Time	Local_X	Local_Y	v_Vel	Lane_ID	Movement
4802932	4802932	1948	1118849752200	2005-06-15 11:35:52.2	32.147	2162.848	46.04	3	NaN

In [7]:
```
df2.tail(1)
```

Out[7]:

	Unnamed: 0	Vehicle_ID	Global_Time	Real_Time	Local_X	Local_Y	v_Vel	Lane_ID	Movement
1607318	1607318	1573	1118937747000	2005-06-16 12:02:27.0	99.82	389.216	29.49	2	1.0

In [8]:
```
print(len(df1.groupby(['Vehicle_ID']).size()))
print(len(df2.groupby(['Vehicle_ID']).size()))
```

```
2847
1506
```

In [9]:
```
# Define a fucntion to calculate the distance between two cars
def calDistance(carLoc1, carLoc2):
    x1 = carLoc1[0]
    y1 = carLoc1[1]
    x2 = carLoc2[0]
    y2 = carLoc2[1]
    return round(np.sqrt((x1 - x2)**2 + (y1 - y2)**2))
```

US-101

In [10]:
```
timeList1 = df1.groupby(['Global_Time']).size()
timeList1.index
```

Out[10]:
```
Int64Index([1118846979700, 1118846979800, 1118846979900, 1118846980000,
            1118846980100, 1118846980200, 1118846980300, 1118846980400,
            1118846980500, 1118846980600,
            ...
            1118849751300, 1118849751400, 1118849751500, 1118849751600,
            1118849751700, 1118849751800, 1118849751900, 1118849752000,
            1118849752100, 1118849752200],
           dtype='int64', name='Global_Time', length=27726)
```

In [15]:
```
distanceList1 = []

for time in timeList1.index:
#     print(time)
    dfTime = df1[df1['Global_Time'] == time]
    carLocList = [[x, y] for x, y in zip(dfTime.loc[:, 'Local_X'], dfTime.loc[:, 'Local_Y'])]
    carNum = len(carLocList)
    for i in range(carNum - 1):
        for j in range(i + 1, carNum):
            dis = calDistance(carLocList[i], carLocList[j])
            distanceList1.append(dis)

distanceArray1 = np.array(distanceList1)
print('The maximum vehicle headway is', max(distanceList1), 'feets.')
print('The minimum vehicle headway is', min(distanceList1), 'feets.')
print('The average vehicle headway is', np.mean(distanceArray1), 'feets.')
print('The median vehicle headway is', np.median(distanceArray1), 'feets.')
```

```
The maximum vehicle headway is 2195.0 feets.
The minimum vehicle headway is 0.0 feets.
The average vehicle headway is 688.4350796180776 feets.
The median vehicle headway is 597.0 feets.
```

Lankershim

In [14]:
```
timeList2 = df2.groupby(['Global_Time']).size()
timeList2.index

distanceList2 = []

for time in timeList2.index:
#     print(time)
    dfTime = df2[df2['Global_Time'] == time]
    carLocList = [[x, y] for x, y in zip(dfTime.loc[:, 'Local_X'], dfTime.loc[:, 'Local_Y'])]
    carNum = len(carLocList)
    for i in range(carNum - 1):
        for j in range(i + 1, carNum):
            dis = calDistance(carLocList[i], carLocList[j])
            distanceList2.append(dis)

distanceArray2 = np.array(distanceList2)
print('The maximum vehicle headway is', max(distanceList2), 'feets.')
print('The minimum vehicle headway is', min(distanceList2), 'feets.')
print('The average vehicle headway is', np.mean(distanceArray2), 'feets.')
print('The median vehicle headway is', np.median(distanceArray2), 'feets.')
```

```
The maximum vehicle headway is 1640.0 feets.
The minimum vehicle headway is 0.0 feets.
The average vehicle headway is 512.8853229649293 feets.
The median vehicle headway is 409.0 feets.
```

In [ ]: