2. Data Requirement

This prject contians the following data:

1) Zip code of the three three cities with associted latitude and longitude
The Zip code will be used as surrogate of neighborhood of each city. The data is available from: https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/

2) Venues in each zip code area in the three cities:
Venues will be obtained from Foursquare APIs. By using this API, we will get all the venues in each zip code.

2.1 Location Data Preparation

We first import all the needed modules. In this study, we use zip codes as the surrogate for neighborhoods. We find zipcodes of all the cities in the United States along with latitude and longitude from the following website..

```
#https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitu

df = pd.read_csv('us-zip-code-latitude-and-longitude.csv')
df = df.drop(['Timezone','Daylight savings time flag'], axis=1)
```

In the United States, there are many cities in different states with the same name. For example, there is Austin in Texas, Minnesota, Arkansas, and Indiana. Therefore, the dataframe is filtered out by both city name and state. Finally, a dataframe containing all the zip codes in Austin, Seattle, and San Francisco is obtained.

```
city=['Austin','Seattle','San Francisco']
df_1=df.loc[(df['City']=='Austin')&(df['State'] == 'TX')]
df_2=df.loc[(df['City']=='Seattle')&(df['State'] == 'WA')]
df_3=df.loc[(df['City']=='San Francisco')&(df['State'] == 'CA')]
frame=[df_1,df_2,df_3]
df_three=pd.concat(frame)
df_three.reset_index(inplace=True)
df_three.rename(columns={'Zip':'Zip Code'}, inplace=True)
df_three = df_three.drop('index', axis=1)
df_three.head()
```

	Zip Code	City	State	Latitude	Longitude
0	78701	Austin	TX	30.271270	-97.741030
1	78705	Austin	TX	30.292424	-97.738560
2	78727	Austin	TX	30.425652	-97.714190
3	78762	Austin	TX	30.326374	-97.771258
4	78763	Austin	TX	30.335398	-97.559807

2.2 Exploratory Data Preparation

Foursquare Credentials and Version were defined. The Foursquare API will get the top 100 venues in each zip code area within a radius of 5000 meters.

```
CLIENT_ID = 'VZV2XVTGSGOPBFSNTDGZFKNAAOKSSDNORH5S5G4QB244PAVQ' # your Foursquare
CLIENT_SECRET = 'SXK2KXI0BBIXVH0ZSXEAT1X0H5AT5PPSFQPXAZXDKC0ZNJKO' # your Foursqu
VERSION = '20180604'

limit=100
radius=5000
```

The Foursquare API is then used to obtain all the venues for each zip code based on latitude and longitude in Austin, Seattle, and San Francisco.

```
# run function getNearbyVenues on each zip code
three_venues = getNearbyVenues(names=df_three['Zip Code'],
                               latitudes=df_three['Latitude'],
                               longitudes=df_three['Longitude']
                              )
three_venues.head()
```

	Zip Code	Zip Code Latitude	Zip Code Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	78701	30.27127	-97.74103	Chi'lantro BBQ	30.270600	-97.741928	Food Truck
1	78701	30.27127	-97.74103	Caffé Medici	30.270119	-97.742154	Coffee Shop
2	78701	30.27127	-97.74103	Capitol Visitors Center	30.272625	-97.739300	Capitol Building
3	78701	30.27127	-97.74103	Paramount Theatre	30.269457	-97.742077	Movie Theater
4	78701	30.27127	-97.74103	The Townsend	30.269611	-97.742448	Lounge

The venue category at each zip code of each city will be analyzed to explore the similarity between Austin and Seattle or San Francisco.