2.Data Description

With respect to the above data analysis problem I have taken the help of the following data sources :

- **Foursquare API -** I have extracted data from the Foursquare API, which is one of the most trusted platforms for getting location data from. This entailed in making use of the Developer Tools provided by Foursquare after creating an app instance on their dashboard. Using the Foursquare API we will be able to access the latitude, the longitude, the postal codes of the neighborhoods, various venues that are nearby, their geographical coordinates, and much more.

- **NYU Spatial Data Repository -** The data of the various neighborhoods and the boroughs of the city of new york are present at the repository for free public use. I have downloaded the dataset from a link wherein the same dataset was uploaded. The result of reading the dataset is this. As you can see, our primary focus is to get the names and coordinates of the neighborhoods and the boroughs to which they belong.

```
{'geometry': {'coordinates': [-73.84720052054902, 40.89470517661],
  'type': 'Point'},
 'geometry_name': 'geom',
 'id': 'nyu_2451_34572.1',
 'properties': {'annoangle': 0.0,
  'annoline1': 'Wakefield',
  'annoline2': None,
  'annoline3': None,
  'bbox': [-73.84720052054902,
   40.89470517661,
   -73.84720052054902,
   40.89470517661],
  'borough': 'Bronx',
  'name': 'Wakefield',
  'stacked': 1},
 'type': 'Feature'}
```

- **Kaggle dataset -** There are not too many public datasets available for demographics of the city of new york. Hence, in an effort to provide more accessible analysis of the neighborhoods, I collected The Sales Price of Housing for each neighborhood of New York City from a Kaggle

dataset. We extract specifics in this dataset, namely the neighborhood, borough, total units sold, sales price for these transactions.

I		BOROUGH	NEIGHBORHOOD	BUILDING CLASS CATEGORY	TAX CLASS AT PRESENT	BLOCK	LOT	EASE-MENT	BUILDING CLASS AT PRESENT	ADDRESS	APARTMENT NUMBER	ZIP CODE	RESIDENTIAL UNITS	C(
0	4	1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2A	392	6		C2	153 AVENUE B		10009	5	0
1	5	1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	399	26		C7	234 EAST 4TH STREET		10009	28	3
2	6	1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	399	39		C7	197 EAST 3RD STREET		10009	16	1
3	7	1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2B	402	21		C4	154 EAST 7TH STREET		10009	10	0
4	8	1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2A	404	55		C2	301 EAST 10TH STREET		10009	6	0

Along with the datasets, there were many libraries that were required for this analysis, as it is the first step in quicker and correct code compilation for the analysis such as:

1. Requests

2. Folium

3. Pandas

4. Numpy

5. Geopy.geocoders

6. Json

7. Matplotlib

8. Sklearn

9. Wordcloud

```python
import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analsysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
pd.set_option('float_format', '{:.2f}'.format)

import json # library to handle JSON files

!conda install -c conda-forge geopy --yes
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't completed th
import folium # map rendering library

print('Libraries have been imported.')
```