Capstone Project Final Report

Finding the Best Neighborhood for You in Brooklyn

Introduction:

The purpose of this project is to assist people in selecting the best neighborhood inside of the Brooklyn to move into. The criteria analyzed for this decision is each neighborhoods level of volume regarding its access to Arts & Entertainment, College & University, Event, Food, Nightlife Spot, Outdoors & Recreation, Professional & Other Places, Residence, Shop & Service, and Travel & Transport. This can be found using the Foursquare API.

With how many people not only move to New York city every year but migrate between the 5 boroughs this information could be incredibly useful for this large and ever-increasing group.

These 10 criteria are often seen as critical features when individuals are deciding where to move to as a new residence, or entrepreneurs looking to build a new business in areas that are booming or needing development. By looking at the current volume of these within different clusters of neighborhoods it will provide potential migrants with awareness of the area and neighborhood before moving to a new city, state, country, or place for their work or to start a new fresh life.

Data Section:

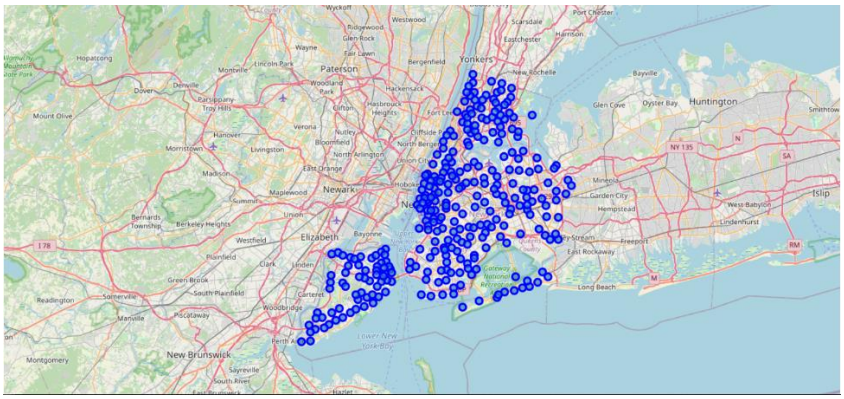
Will use the New York dataset provided in week 3 of the course. Providing boroughs, neighborhoods, latitudes, and longitudes.

Foursquare API Data:

We will need data about different venues in different neighborhoods of that specific borough. To gain access to that information, we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about the top 10 venues inside each neighborhood. For each neighborhood, we have chosen the radius to be 500 meters.

Map of New York



Methodology:

To begin I had to build the data frame using the New York dataset.

```
In [70]: neighborhoods_data = newyork_data['features']

In [71]: # define the dataframe columns
         column_names = ['Borough', 'Neighborhood', 'Latitude', 'Longitude']

         # instantiate the dataframe
         neighborhoods = pd.DataFrame(columns=column_names)

In [72]: for data in neighborhoods_data:
             borough = neighborhood_name = data['properties']['borough']
             neighborhood_name = data['properties']['name']

             neighborhood_latlon = data['geometry']['coordinates']
             neighborhood_lat = neighborhood_latlon[1]
             neighborhood_lon = neighborhood_latlon[0]

             neighborhoods = neighborhoods.append({'Borough': borough,
                                                   'Neighborhood': neighborhood_name,
                                                   'Latitude': neighborhood_lat,
                                                   'Longitude': neighborhood_lon}, ignore_index=True)
```

Using the data set I then structured the database to have 4 columns: Borough, Neighborhood, Latitude, and Longitude.

After constructing the New York database, I zoned in on our desired region of Brooklyn, we had to clean and specify the data set.

```
In [77]: brooklyn_data = neighborhoods[neighborhoods['Borough'] == 'Brooklyn'].reset_index(drop=True)
         brooklyn_data.head()
```

Out[77]:

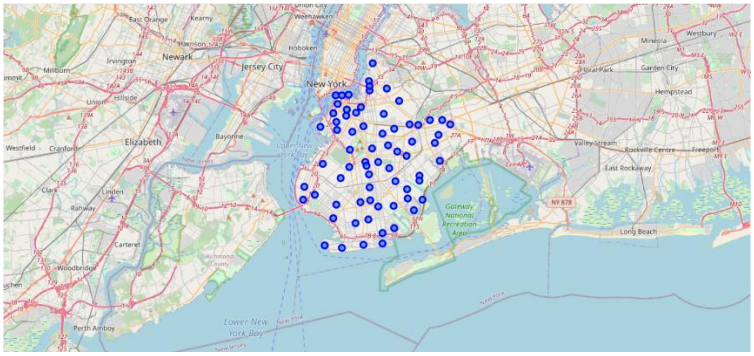
	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471

To have a better visual of the borough I then constructed a folium map of Brooklyn segmented into its neighborhoods.

```
In [78]: address = 'Brooklyn, NY'

         geolocator = Nominatim(user_agent="ny_explorer")
         location = geolocator.geocode(address)
         latitude = location.latitude
         longitude = location.longitude
         print('The geograpical coordinate of Brooklyn are {}, {}.'.format(latitude, longitude))

         The geograpical coordinate of Brooklyn are 40.6501038, -73.9495823.
```

```
In [79]: # create map of Brooklyn using latitude and longitude values
         map_brooklyn = folium.Map(location=[latitude, longitude], zoom_start=11)

         # add markers to map
         for lat, lng, label in zip(brooklyn_data['Latitude'], brooklyn_data['Longitude'], brooklyn_data['Neighborhood']):
             label = folium.Popup(label, parse_html=True)
             folium.CircleMarker(
                 [lat, lng],
                 radius=5,
                 popup=label,
                 color='blue',
                 fill=True,
                 fill_color='#3186cc',
                 fill_opacity=0.7,
                 parse_html=False).add_to(map_brooklyn)

         map_brooklyn
```

This is the resulting map.



From here I used the Foursquare API to find and append the top venue locations in the borough.

```
In [104]: def get_venues_count(lat, long, radius, categoryId):
              explore_url = 'https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&v={}&ll={},{}&categoryId={}'.format(
                          CLIENT_ID,
                          CLIENT_SECRET,
                          VERSION,
                          lat,
                          long,
                          categoryId)

              # make the GET request
              return requests.get(explore_url).json()['response']['totalResults']
```

```
In [105]: brooklyn_venues = brooklyn_data.copy()
          for c in categories_list:
              brooklyn_venues[c[0]] = 0
```

Out[105]:

	Borough	Neighborhood	Latitude	Longitude	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
0	Brooklyn	Bay Ridge	40.625801	-74.030621	0	0	0	0	0	0	0	0	0	0
1	Brooklyn	Bensonhurst	40.611009	-73.995180	0	0	0	0	0	0	0	0	0	0
2	Brooklyn	Sunset Park	40.645103	-74.010316	0	0	0	0	0	0	0	0	0	0
3	Brooklyn	Greenpoint	40.730201	-73.954241	0	0	0	0	0	0	0	0	0	0
4	Brooklyn	Gravesend	40.595260	-73.973471	0	0	0	0	0	0	0	0	0	0

This was then added upon by finding the totals of each within each of the neighborhoods.

```
In [106]: for i, row in brooklyn_venues.iterrows():
              for c in categories_list:
                  brooklyn_venues.loc[i, c[0]] = get_venues_count(brooklyn_venues.iloc[i].Latitude,
                                                                  brooklyn_venues.iloc[i].Longitude,
                                                                  radius=500, categoryId=c[1])

              print('{} ({}, {}) data gathering for is complete'.format(
                  brooklyn_venues.iloc[i].Neighborhood, brooklyn_venues.iloc[i].Latitude,
                  brooklyn_venues.iloc[i].Longitude))
```

```
In [107]: brooklyn_venues.head()
```

Out[107]:

	Borough	Neighborhood	Latitude	Longitude	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
0	Brooklyn	Bay Ridge	40.625801	-74.030621	11	26	7	208	48	75	253	15	201	53
1	Brooklyn	Bensonhurst	40.611009	-73.995180	8	51	8	114	10	43	276	21	201	43
2	Brooklyn	Sunset Park	40.645103	-74.010316	21	38	9	180	15	35	244	12	194	46
3	Brooklyn	Greenpoint	40.730201	-73.954241	55	129	24	121	77	138	288	149	223	65
4	Brooklyn	Gravesend	40.595260	-73.973471	15	93	6	105	20	84	275	35	201	97

To make this easier to work with and plot on a graph I normalized the totals of the venues.

```
In [108]: from sklearn.preprocessing import MinMaxScaler
          X = brooklyn_venues.values[:,4:]
          scaled_dataset = MinMaxScaler().fit_transform(X)
          brooklyn_scaled = pd.DataFrame(scaled_dataset)
          brooklyn_scaled.columns = [c[0] for c in categories_list]
          brooklyn_scaled.head()
```

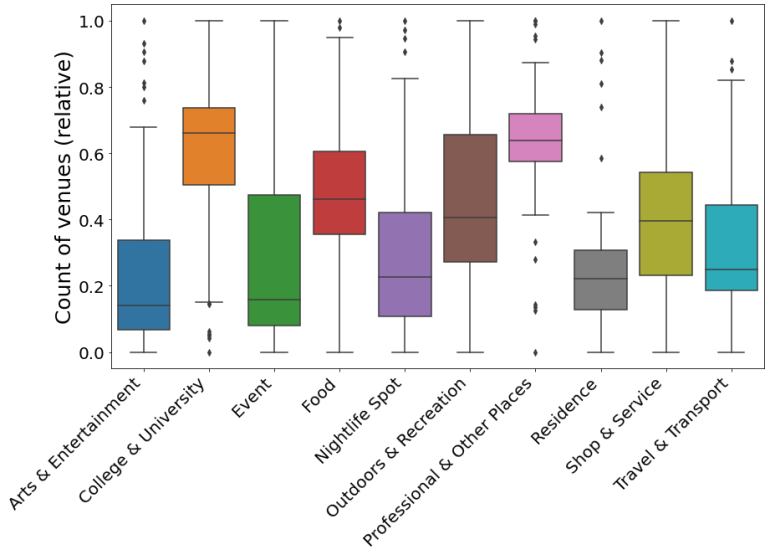
Out[108]:

	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
0	0.066667	0.062069	0.078947	1.000000	0.586667	0.298013	0.414414	0.017751	0.207547	0.089431
1	0.026667	0.234483	0.105263	0.416149	0.080000	0.086093	0.621622	0.053254	0.207547	0.008130
2	0.200000	0.144828	0.131579	0.826087	0.146667	0.033113	0.333333	0.000000	0.075472	0.032520
3	0.653333	0.772414	0.526316	0.459627	0.973333	0.715232	0.729730	0.810651	0.622642	0.186992
4	0.120000	0.524138	0.052632	0.360248	0.213333	0.357616	0.612613	0.136095	0.207547	0.447154

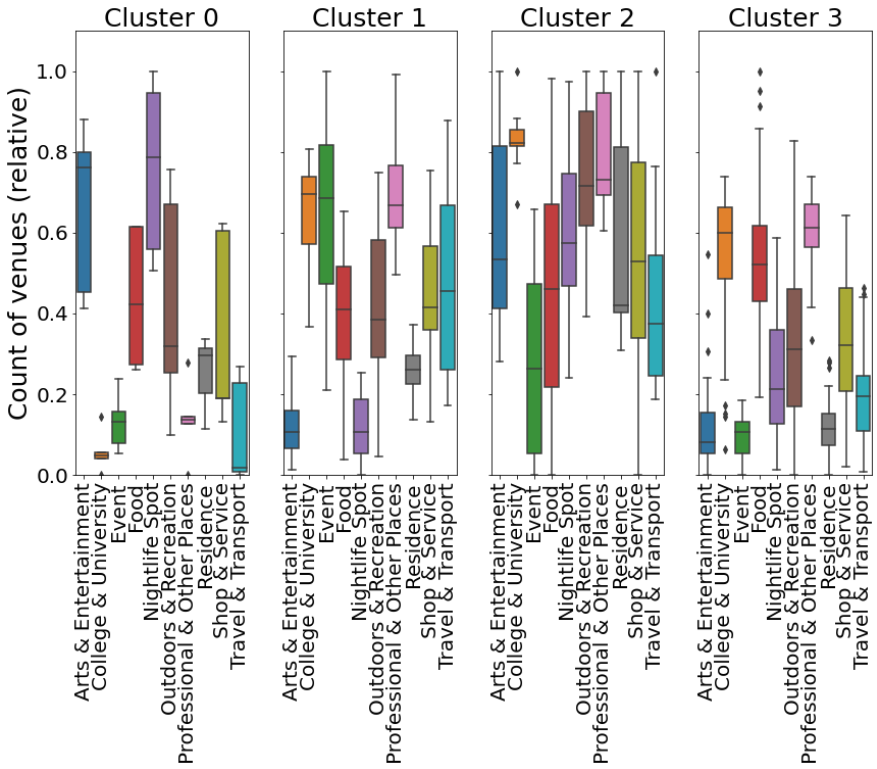
After balancing all that remained was clustering the neighborhoods and then using the clusters to create box plots in order to view the density of venues within each cluster to help visualize which section are booming and which are due for expansion. These will be seen in the results section.

Results Section:

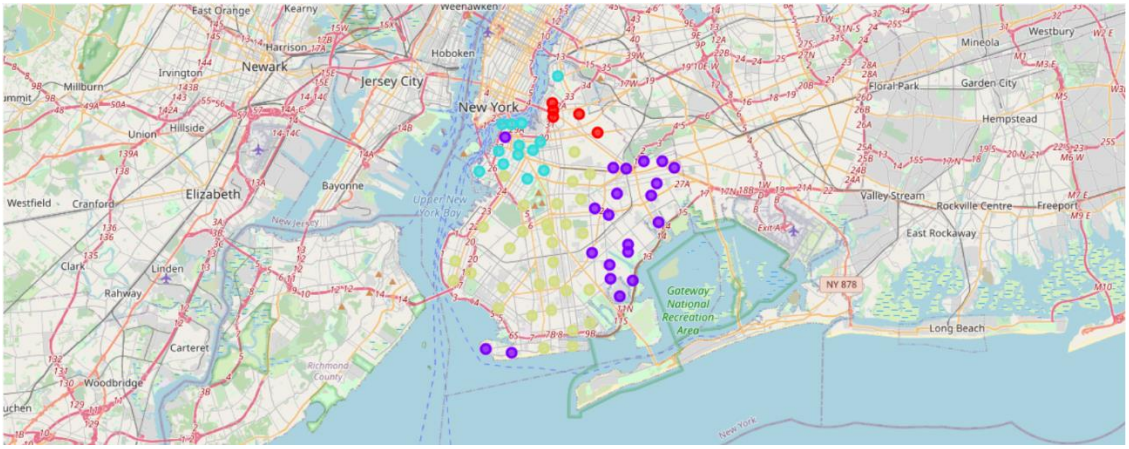
Box plot of venue density in Brooklyn



Box plot of the density of venues within each unique cluster



Map of the clustered neighborhoods in Brooklyn



Red = cluster 0, Purple = cluster 1, Teal = cluster 2, Green/Yellow = cluster 3

After looking at these results we can see that the densest section of Brooklyn regarding venues is cluster 2. This may be due to its location next to main ports and the fact that it is near all of the in traffic from other areas. The least dense section being cluster 3. With its location more inland and away from important trade/travel sections, this makes sense.

Discussion Section:

From looking at the results the 2 main sections of Brooklyn I would recommend new migrants, both entrepreneurs and regular residents, move into would be neighborhoods in either cluster 2 or cluster 3. Cluster 2 because it has the most activities and opportunities. This would allow for increased comfort to new residents and established success for prospective businesses. Cluster 3 because it would most likely be less expensive than the other areas and has untapped potential for new business.

Conclusion Section:

Overall, I believe that this project could provide a large amount of utility to people who are looking to move to Brooklyn. The best part being that this could always be added on to, to provide even more information about the neighborhoods.

This project has shown just how specific and nuanced each individual project can be. It also showed that every problem or question has a solution that can be found and then shared. The power of the Fourquare API combined with the folium mapping and seaborn graph was especially eye opening.