CAPSTONE PROJECT – THE BATTLE OF NEIGHBORHOODS

A. Introduction

This project aims to answer the question of where to open a new McDonald's in New York city. This is accomplished through analyzing available New York city's geospatial data and other relevant statistics. The final recommendation is made basing on the number of already existing McDonald's in the neighborhood, the size of the neighborhood, and the population of the neighborhood.

Starting a franchise is a popular investment option for many American entrepreneurs. Operating a franchise offers the independence of small business ownership while still benefiting from the support of a big business network. Franchises like McDonald's has a well-established reputation, proven management and work practices, and constant national-scale advertising campaigns. Combining this with the fact that New York is a populous city with an extremely dynamic economy, it does appear that starting a new McDonald's franchise is attractive to small business owners.

However, there is definitely an abundance of McDonald's in New York City. In fact, it is the city with the most McDonald's in the entire country, with more than 250 locations. However, it is also the biggest city in the United States by population, doubling that of the second largest one. Assuming that after doing thorough market research, you readers have come to the conclusion that there is still untapped potential in The City that Never Sleeps, now you need to consider where should you open your brand-new McDonald's. This report serves as a preliminary spatial analysis to readers with the optimal location for opening their McDonald's to best avoid competition with other McDonald's in the vicinity.

B. Data

This project requires location data on New York City's boroughs, neighborhoods, and all existing McDonald's. Additionally, data on New York City's population by region is also needed. Location data on New York City's boroughs and neighborhoods was acquired from the New York University's Spatial Data Repository. Location data on all existing McDonald's in New York City was queried from FourSquare's database through a series of API calls. Finally, New York City's population statistics by regions were sourced from New York State's Department of Health. All data, as well as the analysis Python script, were uploaded to a GitHub repository.

The datasets are filtered, modified or transformed to be compatible with the programmed Python script. New York City's neighborhoods and boroughs data include 4 columns: Borough, Neighborhood, Latitude, and Longitude. Existing McDonald's data include 4 columns: (Restaurant) Name, Zip Code, Latitude, Longitude. New York City's population data include 4 columns: Neighborhoods within Region, Males, Females, and Total.

There was inconsistency between the way NYU's Spatial Data Repository and NYS's Department of Heath segmented the city into neighborhoods. This project follows the city segmentation standard set by NYS's Department of Health. The Python script includes sections to handle the standardization between the two datasets. Also, since the regions, as set by the NYS's Department of Health, were not named, they are assigned numeric values (index values in the pop_data dataframe) to distinguished from one another.

The existing McDonald's data was queried from FourSquare as followed. For each neighborhood in consideration (there are 122 in total), one API call was sent to FourSquare to search for all McDonald's within a 5000 m radius of the neighborhood's location. Every result was saved into the same dataframe. Afterward, duplicates were removed. Furthermore, the Restaurant Name column was checked and non-McDonald's were filtered out. The Zip Code column was used to make sure that all the recorded McDonald's were indeed New York City's.

C. Methodology

The latitude and longitude values of the neighborhoods and the existing McDonalds were used to put them all onto the New York City map (Figure 1). The blue circles represent the neighborhoods and the red circles represent the existing McDonald's.

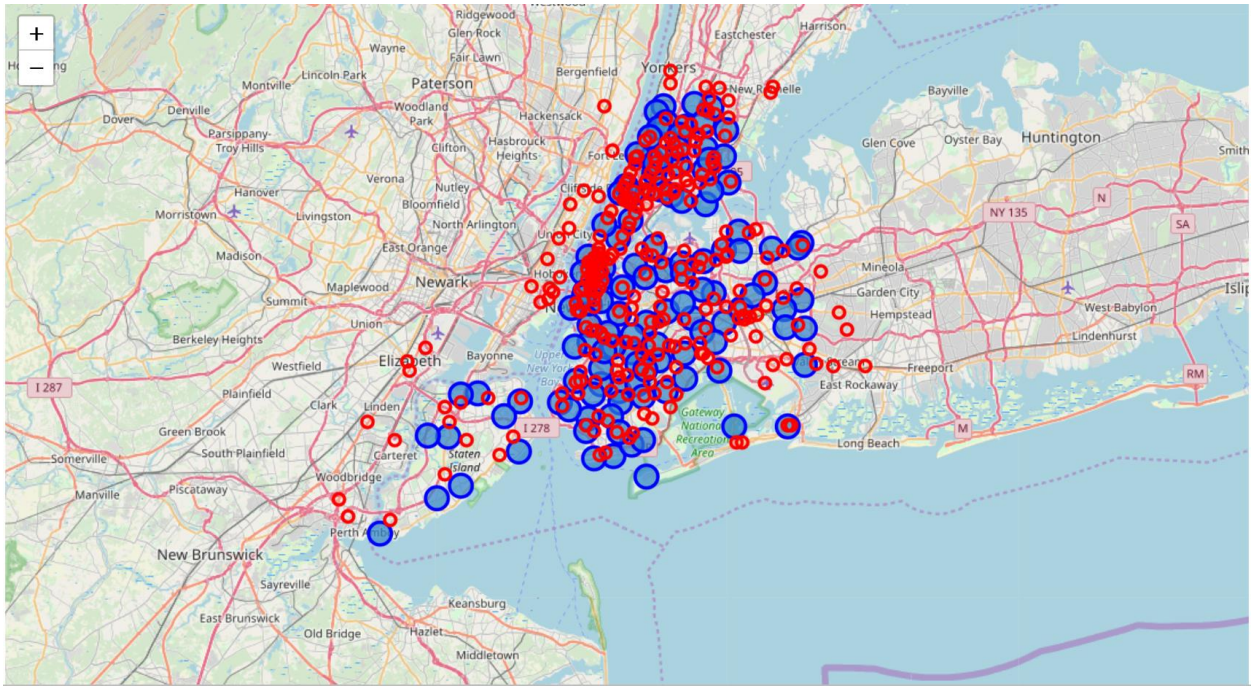

Figure 1: Neighborhoods and Existing McDonald's Locations

The latitude and longitude of the neighborhoods and existing McDonald's were loaded into 2 matrix, X and y. The Scipy package's distance matrix function was used to get a distance matrix of all the neighborhoods versus all the existing McDonald's. Using this distance matrix, each existing McDonald's was assigned to a neighborhood closest to it. From here, it is possible to count the number of McDonald's in each Neighborhood, and subsequently in each defined Region.

Finally, linear regression and quadratic regression models were fitted to find a correlation between the Region population and the number of McDonald's in it. From here, it is possible to identify the region(s) with high population but low number of existing McDonald's. These region(s) would be the ideal candidates to look for a venue of our new McDonald's.

D. Results and Discussions

Table 1 reflects the number of existing McDonald's and population in each region. On average, there are 5 McDonald's in each region. Interestingly, the regions with more than 10 count of McDonald's mostly belong to the Manhattan or Bronx borough.

Table 1: McDonald's Count by Regions

	Region	Males (000s)	Females (000s)	Total (000s)	McD Counts
0	Riverdale, Fieldston, Kingsbridge	51.598	61.481	113.079	3
1	Wakefield, Williamsbridge, Woodlawn	65.216	78.387	143.604	9
2	Co-op City, Pelham Bay, Schuylerville	55.037	65.204	120.241	5
3	Pelham Parkway, Morris Park, Laconia	60.974	67.982	128.956	5
4	Belmont, Crotona Park East, East Tremont	77.119	89.293	166.411	8
5	Bedford Park, Fordham North, Norwood	63.169	68.921	132.090	4
6	Morris Heights, Fordham South, Mount Hope	65.682	72.967	138.648	4
7	Concourse, Highbridge, Mount Eden	68.152	75.978	144.129	5
8	Castle Hill, Clason Point, Parkchester	88.007	100.193	188.201	3
9	Hunts Point, Longwood, Melrose	81.666	79.759	161.425	9
10	Greenpoint, Williamsburg	76.748	77.779	154.527	4
11	Bushwick	66.695	67.378	134.073	2
12	Bedford Stuyvesant	64.655	74.558	139.213	4
13	Brooklyn Heights, Fort Greene	58.271	66.161	124.432	5
14	Park Slope, Carroll Gardens, Red Hook	53.112	58.760	111.871	2
15	Crown Heights North, Prospect Heights	57.936	70.196	128.131	2
16	Brownsville, Ocean Hill	54.311	68.986	123.297	3
17	East New York, Starrett City	73.196	86.466	159.662	3
18	Canarsie, Flatlands	90.847	110.065	200.912	7
19	East Flatbush, Farragut, Rugby	62.365	78.661	141.026	4
20	Crown Heights, Prospect Lefferts, Wingate	50.493	61.025	111.518	2
21	Sunset Park, Windsor Terrace	78.533	75.131	153.664	5
22	Bay Ridge, Dyker Heights	62.055	65.267	127.323	2

23	Borough Park, Kensington, Ocean Parkway	85.578	84.056	169.634	2
24	Flatbush, Midwood	77.849	86.969	164.818	3
25	Sheepshead Bay, Gerritsen Beach, Homecrest	73.985	79.944	153.929	3
26	Bensonhurst, Bath Beach	92.931	96.344	189.275	3
27	Brighton Beach, Coney Island	55.293	64.256	119.550	2
28	Washington Heights, Inwood, Marble Hill	97.142	106.275	203.417	7
29	Hamilton Heights, Manhattanville, West Harlem	61.481	68.085	129.566	8
30	Central Harlem	56.270	65.431	121.701	8
31	East Harlem	56.312	64.124	120.435	4
32	Upper East Side	102.121	127.056	229.177	3
33	Upper West Side, West Side	93.032	108.808	201.840	5
34	Chelsea, Clinton, Midtown Business District	77.568	71.985	149.553	22
35	Murray Hill, Gramercy, tuyvesant Town	71.357	84.491	155.848	6
36	Chinatown, Lower East Side	81.995	87.276	169.271	3
37	Battery Park City, Greenwich Village, Soho	75.851	78.330	154.181	12
38	Astoria, Long Island City	88.021	92.150	180.171	4
39	Jackson Heights, North Corona	98.331	90.820	189.150	5
40	Flushing, Murray Hill, Whitestone	123.550	134.764	258.315	16
41	Bayside, Douglaston, Little Neck	58.155	63.468	121.623	4
42	Queens Village, Cambria Heights, Rosedale	90.730	104.801	195.531	9
43	Briarwood, Fresh Meadows, Hillcrest	76.454	83.163	159.617	10
44	Elmhurst, South Corona	76.755	74.131	150.886	1
45	Forest Hills, Rego Park	54.531	62.495	117.027	3
46	Sunnyside, Woodside	70.244	68.429	138.673	2

Figure 2 shows the linear regression plot of Region's population versus Number of McDonald's. As expected, the higher the more population the region, the higher the count of existing McDonald's. Our ideal region with high potential for opening a new McDonald's should be below the regression line, and to the right side of the plot. These are the regions with low number of existing McDonald's, but high population.

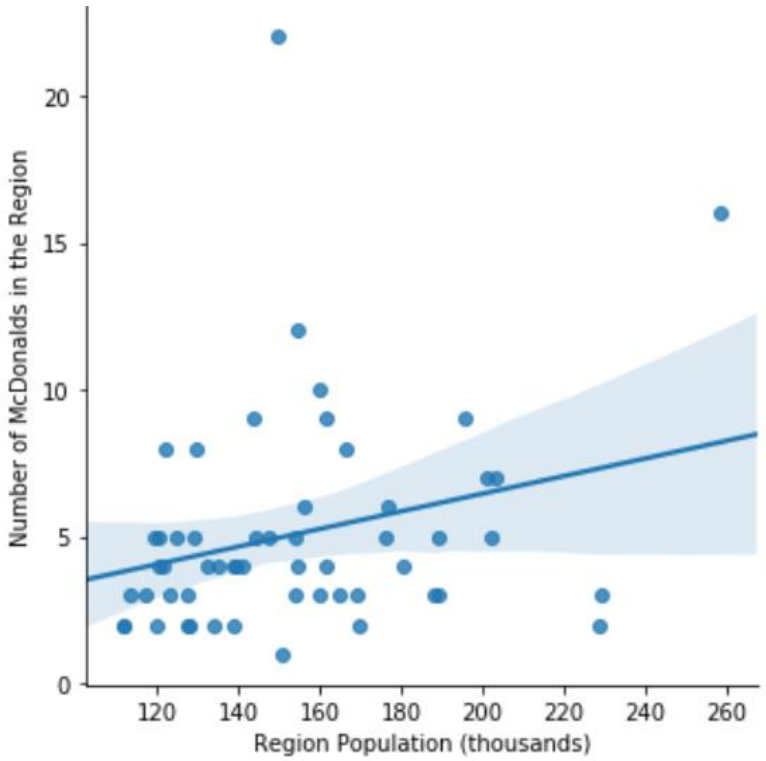

Figure 2: Existing McDonald's Counts as a Linear Function of the Region's Population

Figure 3 shows the quadratic regression plot of the same relationship mentioned above. The regression line seems to fit the data slightly better, but ultimately does not change the results of our analysis. The potential regions to open our new McDonald's are summarized in table 2. Their respective neighborhoods were marked with green circles on Figure 4.

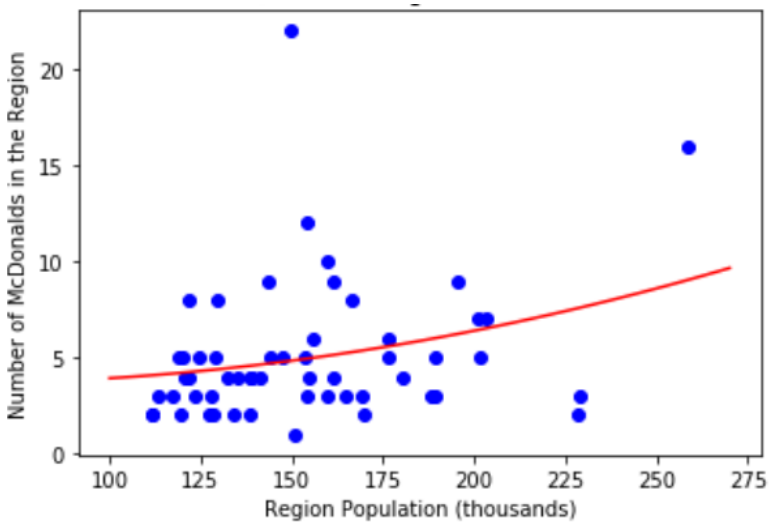

Figure 3: Existing McDonald's Counts as a Quadratic Function of the Region's Population

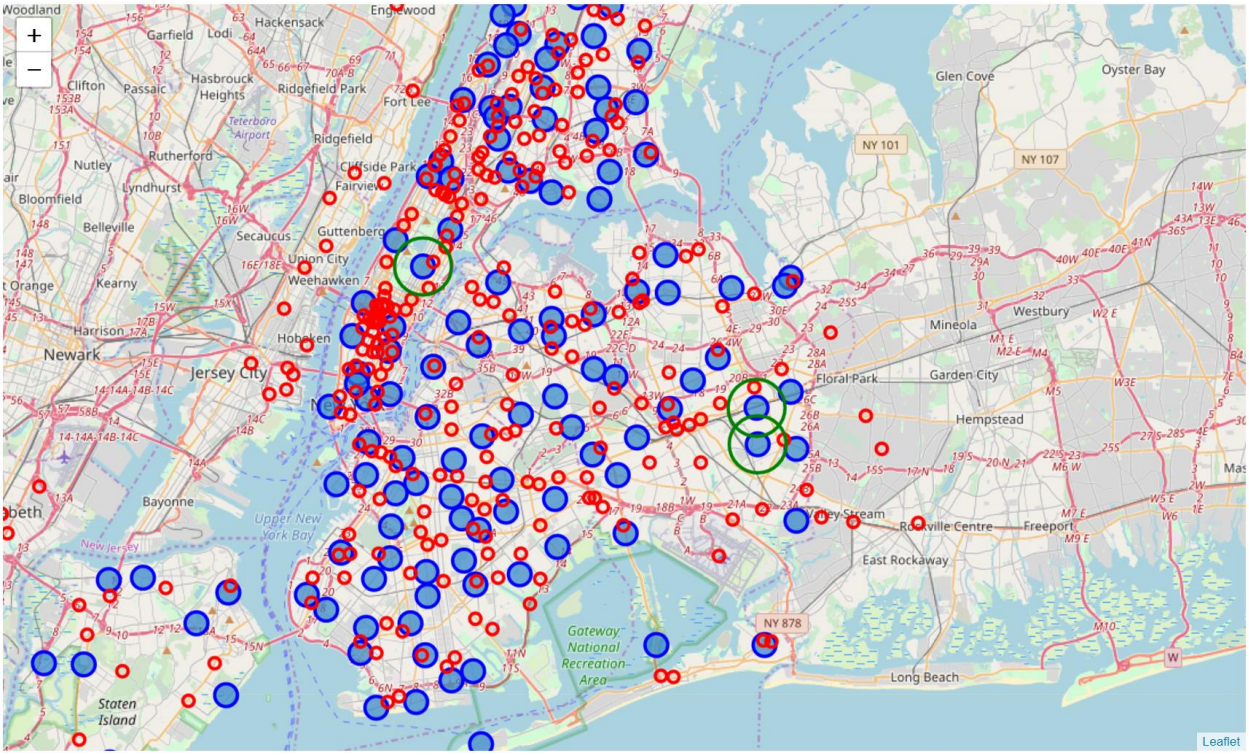

Figure 4: Neighborhoods with High Potential for a New McDonald's

In these regions, the populations are over 225,000, but the count of McDonald's was only 2 or 3. This netted a McDonald per capita of just 0.9 – 1.3 McDonald's per 100,000 population. This is approximately 4 times lower than the New York's average. According to McDonald's own statistics, average McDonald per capita in New York is 3.9 per 100,000 population.

E. Conclusions

Through the all the data sourcing, filtering, and analyzing, we reach the conclusion that there are 3 neighborhoods in New York City with great potential to be our new McDonald's location. These are Upper East Side in Manhattan, Hollis and St. Albans in Queens. These neighborhoods average only about 1 McDonald's per 100,000 population, 4 times lower than that of the state's average. These locations definitely deserve our attention and warrant further studies and analyses.