New York - Battle of Neighborhood

Introduction

Background

New York City, is the most famous and most populous city in the United States of America known for its diverse and multicultural population. It is the financial capital of USA and by that virtue provides business opportunities. Every year, the city attracts thousands of new people who dream of making it big in art, theater, fashion, finance and many other industries that prosper in New York

For administration purposes, the city is divided into 5 boroughs: The Bronx, Brooklyn, Manhattan, Queens, and Staten Island.

Manhattan is the most publicly recognized area of New York City, the geographically smallest and most densely populated borough. It is home to most of the city's skyscrapers and prominent landmarks, including Times Square and Central Park

Brooklyn is the most populous borough and is known for its cultural, social, and ethnic diversity. Brooklyn has evolved into a thriving hub of entrepreneurship and high technology startup firms.

Queens is geographically the largest borough, the most ethnically diverse county in the United States has both commercial and residential prominence. It is home to multiple sports arenas and two of the three airports New York metropolitan area.

The Bronx is New York City's northernmost borough and the only New York City borough that is part of the United States mainland. It is home to the largest cooperatively owned housing complex in the United States.

Staten Island is the most suburban of the five boroughs. Staten Island is known for the Staten Island Greenbelt which spans approximately 2,500 acres and one of the last undisturbed forests in the city.

With such diverse areas and population, New York city provides a lucrative opportunity for new businesses. At the same time, it is a competitive market with a high cost of doing business. So, steps must be taken to analyze and identify the right opportunity for a new business to ensure financial success and viability.

Problem Description

As described earlier, thousands of people move to New York City each year in search of opportunities. And after finding a career opportunity of their choice, eventually they find a life partner and start a family. Although the cost of living in the city is astronomically high and a lot of families decide to live outside the city, there are still people who live and raise their families in New York City. For such families, there arises a need of finding a day-care and after-school care facilities where the babies and young children can be looked after while the adults are at their day job.

The problem at hand is to identify neighborhoods that would offer an opportunity to start a new day-care and after-school care facility to serve the families that are looking for such a facility. The various aspects to consider for this would include:

1. Population for each borough and their growth projections
2. Existing Pre-Schools that are competing for the same customer demographic

Target Audience

The target audience of this analysis would be an individual or an investment firm that is exploring the neighborhood to start a new business and is interested in operating a pre-school/child-care facility.

Data

Following datasets are used for the analysis

- **Borough and Neighborhood Data:**

We use a free online dataset that has the latitude and longitude coordinates for all neighborhoods in the 5 New York borough. The source of the dataset is: https://geo.nyu.edu/catalog/nyu_2451_34572

- **Population Projection by Age Group:**

The target demographic for the pre-school and after-school care business is children under 9 years of age. We need to understand what are the growth projections for this target demographic for each borough to focus our analysis in the areas where there is growth. The dataset for the population projection is obtained from the following source:
https://data.cityofnewyork.us/City-Government/Projected-Population-2010-2040-Total-By-Age-Groups/97pn-acdf
Following is a snapshot of the data:

	Borough	Age	2020	2025	2030	2035	2040
0	Bronx	0-4	109972	111127	110880	110982	112571
1	Bronx	5-9	105775	109843	111137	110893	110942
2	Brooklyn	0-4	186886	184273	182152	182974	186465
3	Brooklyn	5-9	180252	182013	179508	177289	178081
4	Manhattan	0-4	82096	83225	80452	77272	76687
5	Manhattan	5-9	68243	72011	72942	69971	66801
6	Queens	0-4	138141	139574	139958	140733	142459
7	Queens	5-9	132342	133933	135413	135764	136512
8	Staten Island	0-4	28683	29137	28984	28562	28244
9	Staten Island	5-9	29394	29890	30397	30253	29816

- **New York City Pre-Schools:**

We need to map the existing pre-schools and child care services in New York City. We will use the Foursquare API query pre-schools in each neighborhood and use the information to create clusters. Following is an example of the data obtained from the Foursquare API:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Anna's Group Family Daycare and Preschool	40.8875	-73.8507	Daycare
1	Wakefield	40.894705	-73.847201	Tender Tots Day Care, Preschool & After School...	40.877	-73.8728	Daycare
2	Wakefield	40.894705	-73.847201	The Lifeskills Preschool	40.8636	-73.8341	Nursery School
3	Co-op City	40.874294	-73.829939	The Lifeskills Preschool	40.8636	-73.8341	Nursery School
4	Co-op City	40.874294	-73.829939	Anna's Group Family Daycare and Preschool	40.8875	-73.8507	Daycare

- **New York City Parks:**

Another aspect to consider for opening pre-schools is access to parks. It proves beneficial for kids to be able to have some outdoor play time. So, we would leverage the Foursquare API to query parks in each neighborhood and add them to our cluster analysis. Following is an example of the data obtained from the Foursquare API:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	P.S 87 community park	40.896	-73.8475	Playground
1	Wakefield	40.894705	-73.847201	my neighbor park	40.8959	-73.8445	Playground
2	Wakefield	40.894705	-73.847201	Seton Falls Park	40.8884	-73.8402	Park
3	Co-op City	40.874294	-73.829939	Haffen Park Pool	40.8745	-73.8375	Pool
4	Co-op City	40.874294	-73.829939	Haffen Park	40.8736	-73.8385	Park

Methodology

The aim of this analysis was to identify neighborhoods in New York City best suitable to open a new pre-school/day care facility for babies, toddlers and young kids. Using the data described in the section about, we performed the following analysis

- **New York City Coordinate Data** – We first analyzed the geographical coordinates of New York City. The data was presented in a *json* file in form of nested dictionaries. The data of nested python dictionaries was then transformed into a pandas dataframe that contains the geographical coordinates of New York city neighborhoods.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

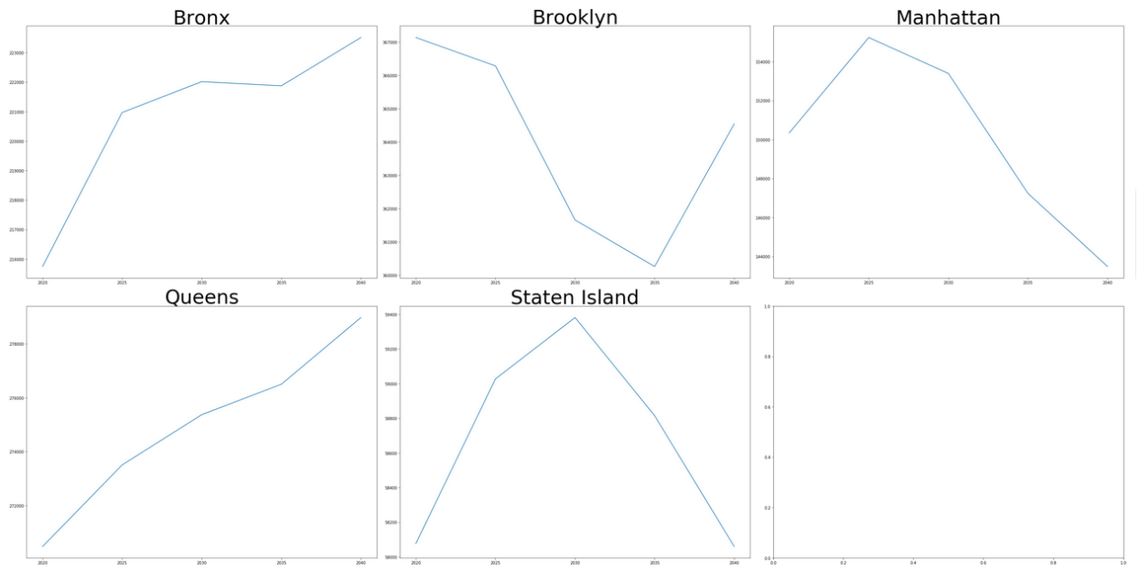
- **New York City Population Projection Data** – Next step was to analyze the New York City Population Projection data. The data had population projection from 2010 thru 2040 for different age ranges across all boroughs.

	Borough	Age	2010	2015	2020	2025	2030	2035	2040
0	NYC Total	0-4	521990	535209	545778	547336	542426	540523	546426
1	NYC Total	15-19	539844	505783	492532	519298	535024	546062	546750
2	NYC Total	20-24	647483	646075	606203	591683	625253	643728	657403
3	NYC Total	25-29	736105	770396	763956	715824	698195	740437	762757
4	NYC Total	30-34	667657	707726	743916	740268	693684	675497	715486

Since our target demographic for pre-school is babies and kids below 9 years of age, we deleted the data for other age groups. Also, we only retained the projections from 2020 thru 2040 resulting in the following dataset.

	Borough	Age	2020	2025	2030	2035	2040
0	Bronx	0-4	109972	111127	110880	110982	112571
1	Bronx	5-9	105775	109843	111137	110893	110942
2	Brooklyn	0-4	186886	184273	182152	182974	186465
3	Brooklyn	5-9	180252	182013	179508	177289	178081
4	Manhattan	0-4	82096	83225	80452	77272	76687
5	Manhattan	5-9	68243	72011	72942	69971	66801
6	Queens	0-4	138141	139574	139958	140733	142459
7	Queens	5-9	132342	133933	135413	135764	136512
8	Staten Island	0-4	28683	29137	28984	28562	28244
9	Staten Island	5-9	29394	29890	30397	30253	29816

Next step was to plot the population growth to identify which borough exhibits the growth in our target demographic and focus the subsequent analysis for selected borough. From the plots, it was evident that only in the borough of Bronx and Queens the population is projected to grow over the next 20 years

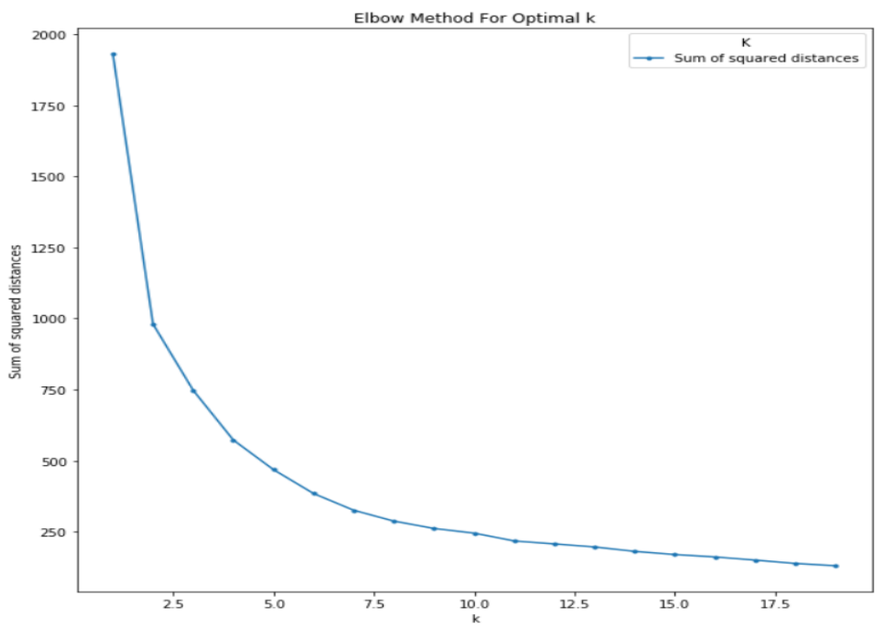


- **Pre-Schools and Park Data** – We utilized the Foursquare API to get the list of Pre-Schools, Day Care Centers and Parks for each neighborhood of Bronx and Queens.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Anna's Group Family Daycare and Preschool	40.8875	-73.8507	Daycare
1	Co-op City	40.874294	-73.829939	The Lifeskills Preschool	40.8636	-73.8341	Nursery School
2	Co-op City	40.874294	-73.829939	Anna's Group Family Daycare and Preschool	40.8875	-73.8507	Daycare
3	Eastchester	40.887556	-73.827806	Anna's Group Family Daycare and Preschool	40.8875	-73.8507	Daycare
4	Fieldston	40.895437	-73.905643	BedRock Preschool	40.8849	-73.9122	School

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	P.S 87 community park	40.896	-73.8475	Playground
1	Wakefield	40.894705	-73.847201	my neighbor park	40.8959	-73.8445	Playground
2	Wakefield	40.894705	-73.847201	Seton Falls Park	40.8884	-73.8402	Park
3	Co-op City	40.874294	-73.829939	Haffen Park	40.8736	-73.8385	Park
4	Co-op City	40.874294	-73.829939	Givens Creek Park	40.8783	-73.8291	Park

We used the Park and Pre-School datasets to neighborhood clusters using the K-Means clustering algorithm on the total Park and Pre-School occurrence in a neighborhood. We utilized the Elbow method to identify the optimal value of 'K' - the number of clusters the data should be divided to get the best result.



Results

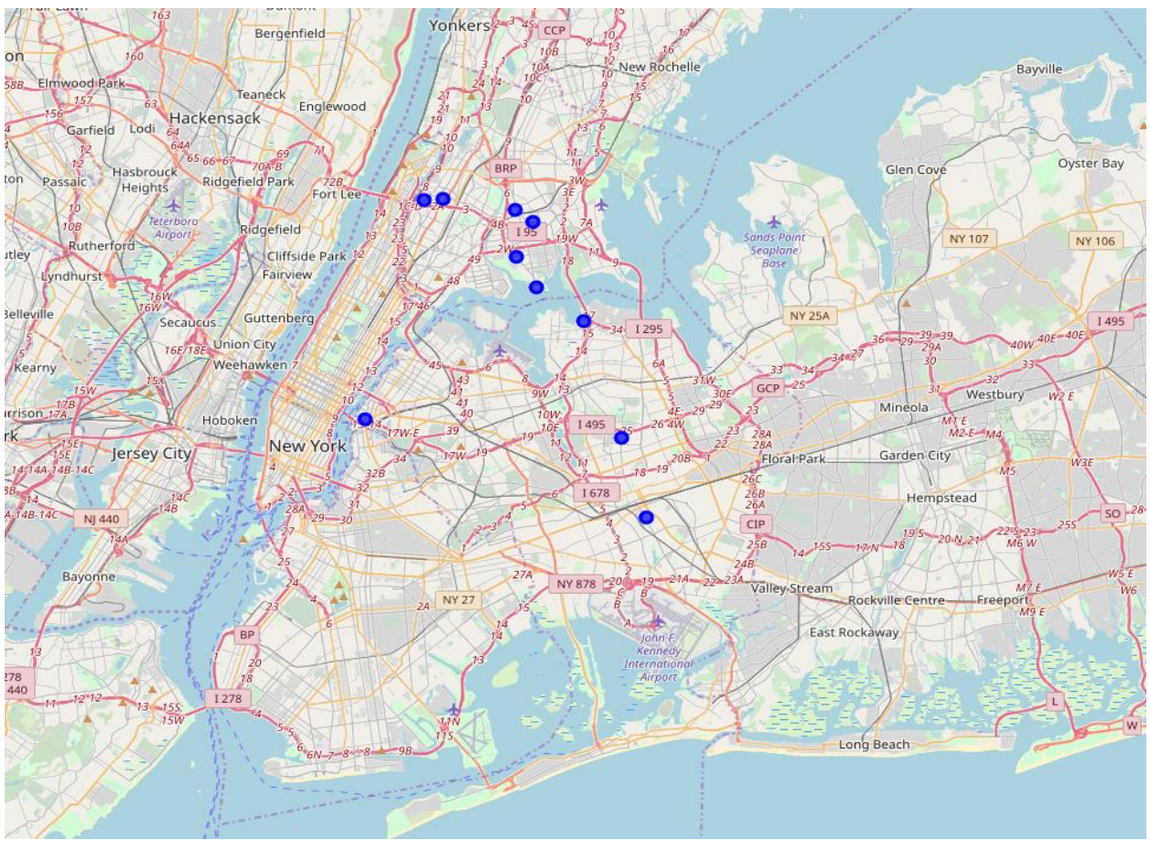
Following the optimal K of 5, the data was divided into 5 clusters using the 'K-Means' clustering algorithm which uses an iterative refinement approach to partition the observations into K clusters with each observation belonging to the cluster with the nearest mean.

	Daycare	Nursery School	Preschool	Child Care Service	Elementary School	Park	State / Provincial Park	Playground	PreSchool Total	Park Total
0	1.500000	1.500000	2.250000	0.000000	0.250000	2.000000	0.000000e+00	2.750000	5.500000	4.750000
1	0.135593	0.118644	0.050847	0.016949	0.000000	0.084746	3.122502e-17	0.016949	0.322034	0.101695
2	1.266667	1.033333	0.833333	0.000000	0.033333	0.033333	1.040834e-17	0.033333	3.166667	0.066667
3	0.300000	0.400000	0.133333	0.000000	0.033333	2.233333	3.333333e-02	1.033333	0.866667	3.300000
4	0.400000	0.400000	0.300000	0.100000	0.100000	5.000000	2.000000e-01	2.900000	1.300000	8.100000

Our goal was to find location suitable for opening a pre-school/daycare facility. So, the best neighborhoods would be the ones that have a low score on Preschool Total. Also, location should provide access to parks and are we were therefore looking for a high score on Park Total
Cluster 3 is the best cluster to meet these criterion and Cluster 4 is second best.

Cluster 3

Cluster 4



Discussion

During this analysis, we have tried to analyze the population growth for New York City and the current presence of pre-school and other child care services to identify the neighborhoods that should be the initial target as potential areas to open and operate a new child care services business. The population projection data provided by the City of New York suggests that the growth in the coming years will be primarily in the Bronx and Queens borough which is where we focused the bulk of our analysis. Using the machine learning algorithm, we could divide the borough into 5 clusters of which 2 clusters were a match for our requirements.

We recommend that the neighborhoods identified in cluster 3 and 4 be considered for further exploring the opportunities for the new business location.

Conclusion

The analysis was limited in its scope and was performed based on publicly available datasets and venues obtained from Foursquare API. There is an opportunity to refine the analysis further by considering additional datasets that could be obtained from public or 3rd party sources. From public sources, we could include crime data to eliminate neighborhoods that have higher percentage of criminal activity. Also, we could look for locations with known sexual predators and ensure that the business is established away areas with their presence.

Another area to explore would be to look at income levels across neighborhoods to determine locations where families with more spending power reside. This can serve as a data point into determining what types of services can be offered and at what cost.

If the scope of the project is changed or access to additional data is provided, we would like to explore the opportunity to enhance the analysis and provide a more refined recommendation for this problem statement.