New York City
Battle of Neighborhood

Introduction

- **Background**

  The focus of this analysis is New York City – the financial capital and the most famous and iconic city in USA

- **Problem Statement**

  Identify neighborhoods in New York City that offer the best opportunity to open a new pre-school/child care and after-school care facility

Data

- **Geographical Data**

  New York City is divided into 5 boroughs comprising of 306 neighborhoods

  We have used a free dataset that contains the latitude and longitude for each neighborhood in each of the 5 borough

  The source of the data is: https://geo.nyu.edu/catalog/nyu_2451_34572

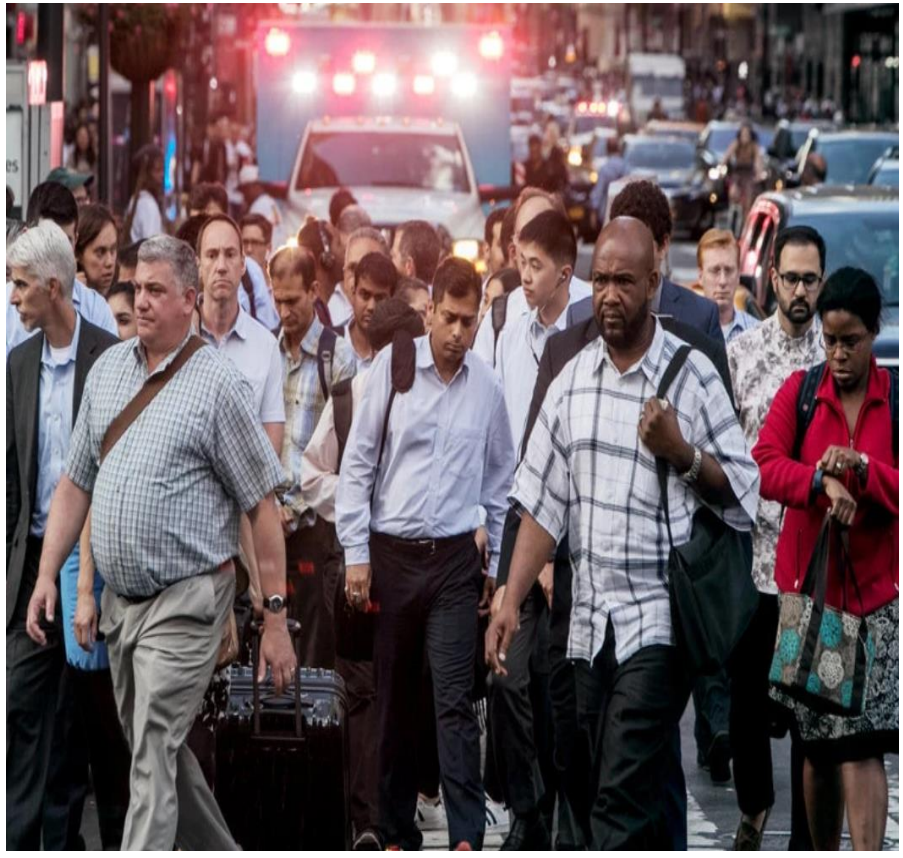
Data

- **Demographic Data**

  To understand the population growth, we have used the population projection data by age group

  The dataset has population projection for each borough thru 2040

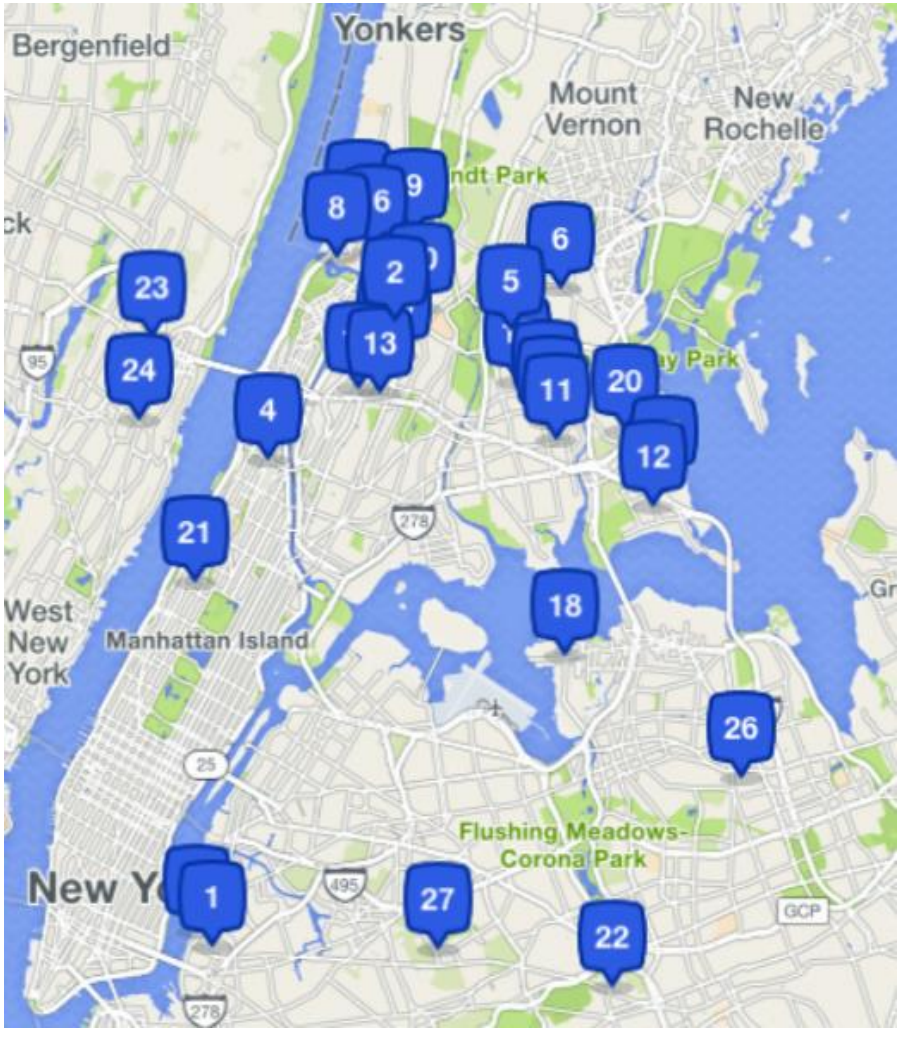
  The source of the data is: https://data.cityofnewyork.us/City-Government/Projected-Population-2010-2040-Total-By-Age-Groups/97pn-acdf



Data

- **Foursquare API**

  We have leveraged the Foursquare API to query list of preschools and parks in all neighborhoods of New York City



Methodology

- **Geo Data**

  New York City geo-data was downloaded as a json file. The nested dictionary data was then transformed into a pandas dataframe

  This data was subsequently used to query Foursquare API and plot maps using the '*Folium*' library

	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

Methodology

- **Demographic Data**

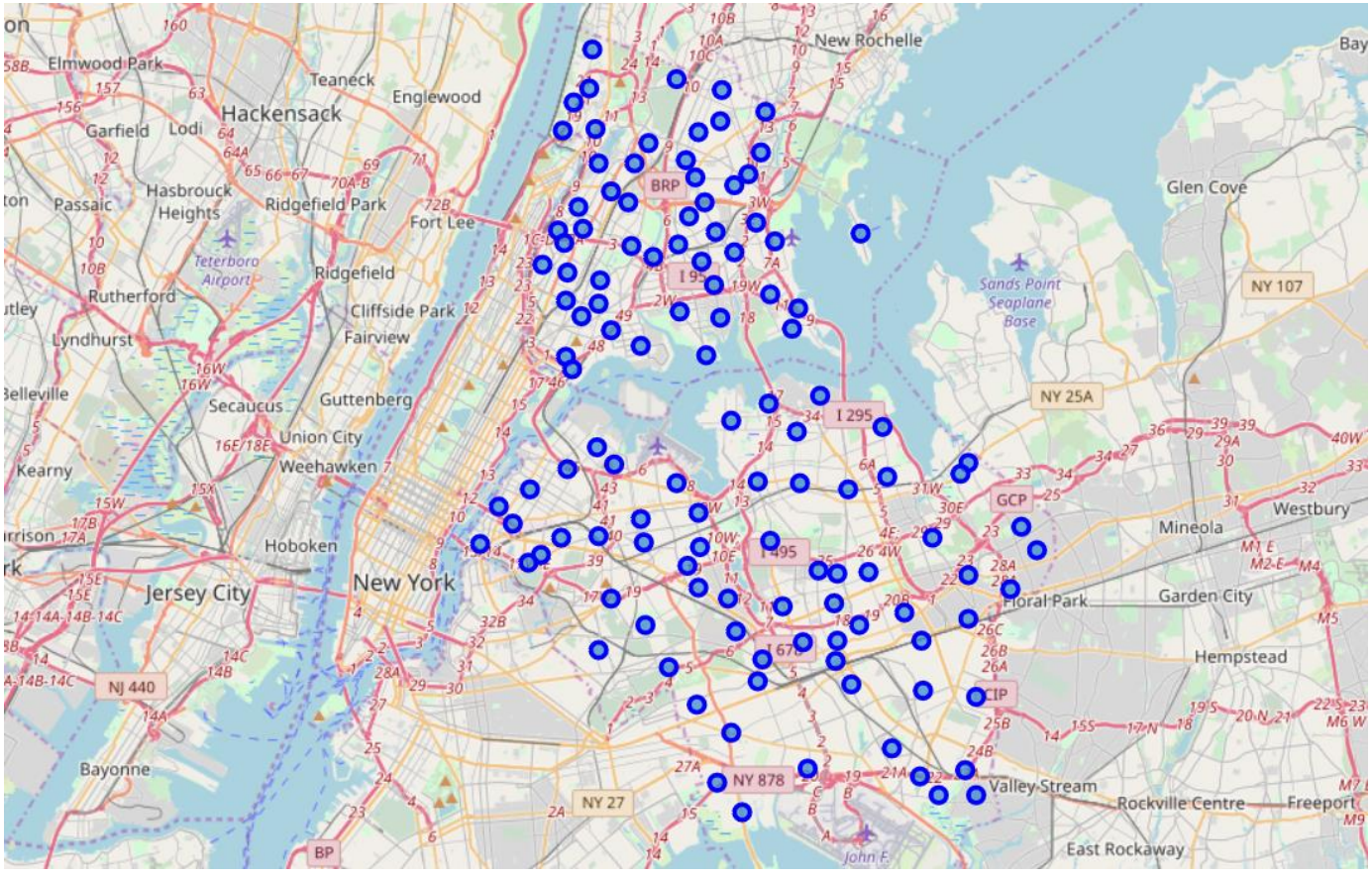
  Population projection data for target demographic (0-9 years) was plotted to understand trend

  Based on the trend, it is clear that the target demographic is set to grow in Bronx and Queens borough

Methodology

- **Bronx and Queens**

  With the demographic growth projected in Bronx and Queens, the subsequent analysis focused on the 2 boroughs



Methodology

- **Pre-School and Parks**

  Utilized the Foursquare API to get a list of Pre-schools and Parks in each of the Bronx and Queens neighborhoods

	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

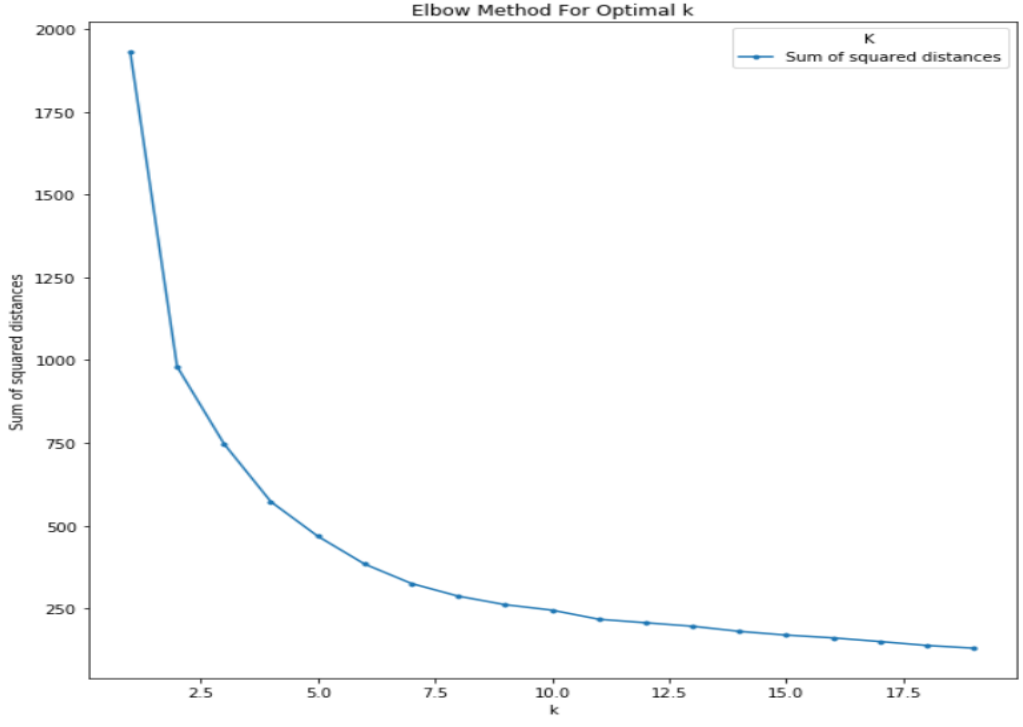
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	P.S 87 community park	40.896	-73.8475	Playground
1	Wakefield	40.894705	-73.847201	my neighbor park	40.8959	-73.8445	Playground
2	Wakefield	40.894705	-73.847201	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

Methodology

- **K-Means Clustering**

  The goal was to cluster the neighborhoods and find clusters with low pre-school and high park presence

  Used the ‘Elbow-Method’ to find the optimal ‘K'



Result

- **K-Means Clustering**

  The neighborhoods were divided into 5 clusters

  Cluster 3 was the best fit with low Pre-School total and high Park Total

  Cluster 4 was second best fit

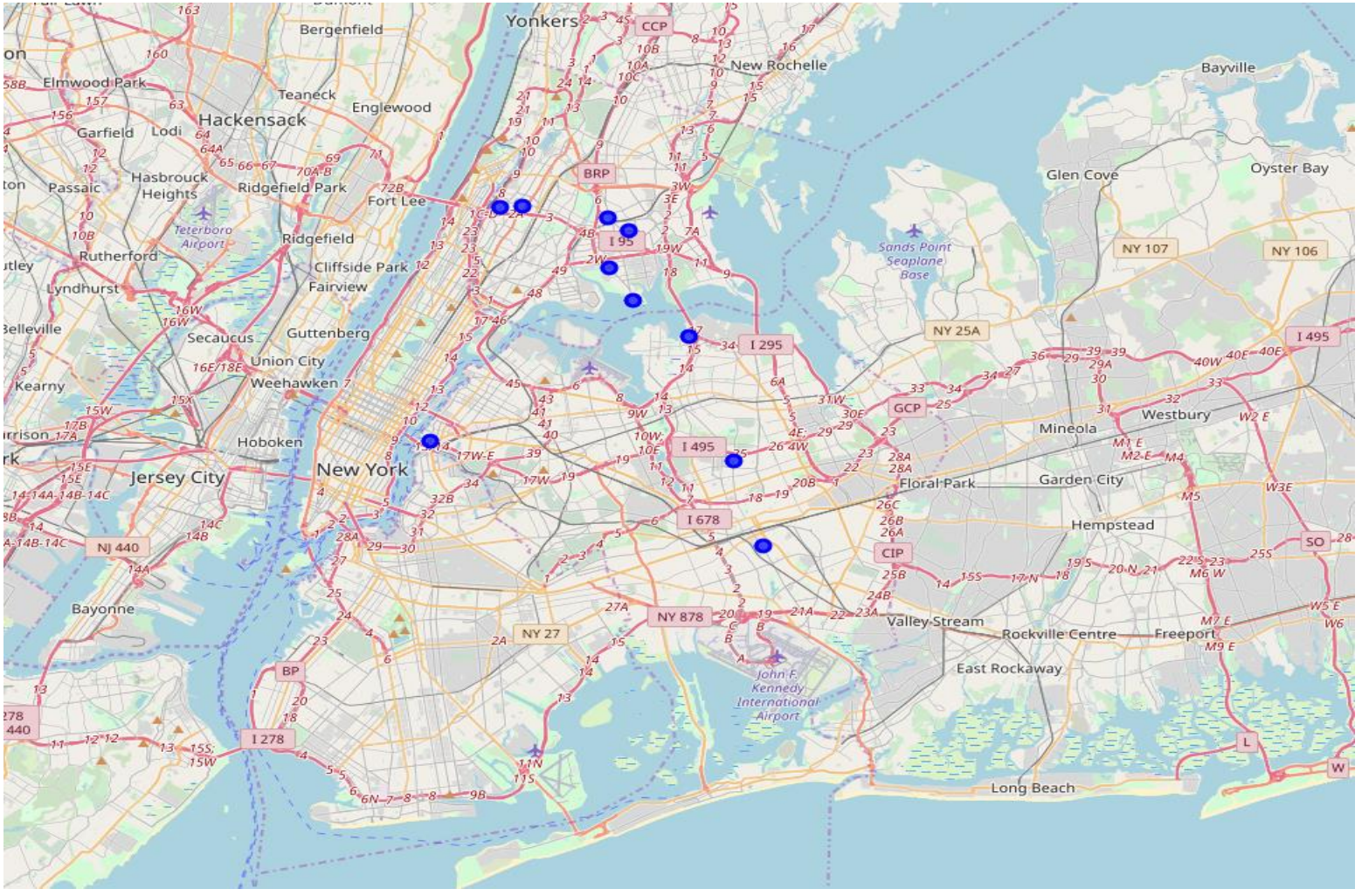
	Daycare	Nursery School	Preschool	Child Care Service	Elementary School	Park	State / Provincial Park	Playground	PreSchool Total	Park Total
0	1.500000	1.500000	2.250000	0.000000	0.250000	2.000000	0.000000e+00	2.750000	5.500000	4.750000
1	0.135593	0.118644	0.050847	0.016949	0.000000	0.084746	3.122502e-17	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

Result

- **Cluster 3**

Result

- **Cluster 4**



Discussion

Population growth for New York City was analyzed to identify growth trends. Growth in next 20 years is projected in Bronx and Queens boroughs

Location of current pre-school and other child care services in the two boroughs was captured

Using the machine learning algorithm, we divided the neighborhoods 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

Further refinement by considering additional datasets that could be obtained from public or 3rd party sources is possible

Crime data can be leveraged to eliminate neighborhoods with higher occurrence of criminal activity

Income data can be used to determine types of services offered and pricing.


THANK YOU