

# The Battle of the Neighborhoods – Final Report

## Introduction:

New York City is a global hub of business and commerce, as a center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States; while Silicon Alley, metonymous for New York's broad-spectrum high technology sphere, continues to expand. The Port of New York and New Jersey is also a major economic engine, handling record cargo volume. New York City, is the most densely populated city in the United States, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass. It is an environment that provides lot of business opportunities.



# BUSINESS PROBLEM:

- New York is the most densely populated City in the United States with 2019 estimates of over 8.4 Million people.
- It is also a major tourist attraction seeing upwards of 60 Million tourists each year.
- New York is known for its fast paced lifestyle and booming nightlife.
- New York is a competitive markets so a new business venture or expansion needs to be thought out carefully.



# MY OBJECTIVE:

- Find a suitable neighborhood in the city to open a new restaurant based on:
  - Population, demographics, saturated markets and density of nearby venues.
- This information will be made available to ABC company to assist businesses or entrepreneurs looking to open/relocate a restaurant in NYC.



# DATA & METHODOLOGY :

Data from three different areas will be compiled as labeled below to find a prospective area for opening a new restaurant.

- Data 1: I will use a New York City Neighborhood dataset from 2014 that is free from the NYU website at [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572). This data provides information on the five boroughs and neighborhoods contained within for all of NYC and includes latitude and longitude information.
- Data 2: My analysis will include information from the following wikipedia pages to scrape for Population and Cuisine Data.
  1. [https://en.wikipedia.org/wiki/New\\_York\\_City](https://en.wikipedia.org/wiki/New_York_City)
  2. [https://en.wikipedia.org/wiki/Portal:New\\_York\\_City](https://en.wikipedia.org/wiki/Portal:New_York_City)
  3. [https://en.wikipedia.org/wiki/Cuisine\\_of\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Cuisine_of_New_York_City)



# DATA & METHODOLOGY(CONT):

- Data 3:  
I will use Data 1-2 as the reference for my input into the Foursquare API and then leverage that data to find the best location to open a new restaurant in NYC.





# EXPECTED RESULTS/FORECAST:

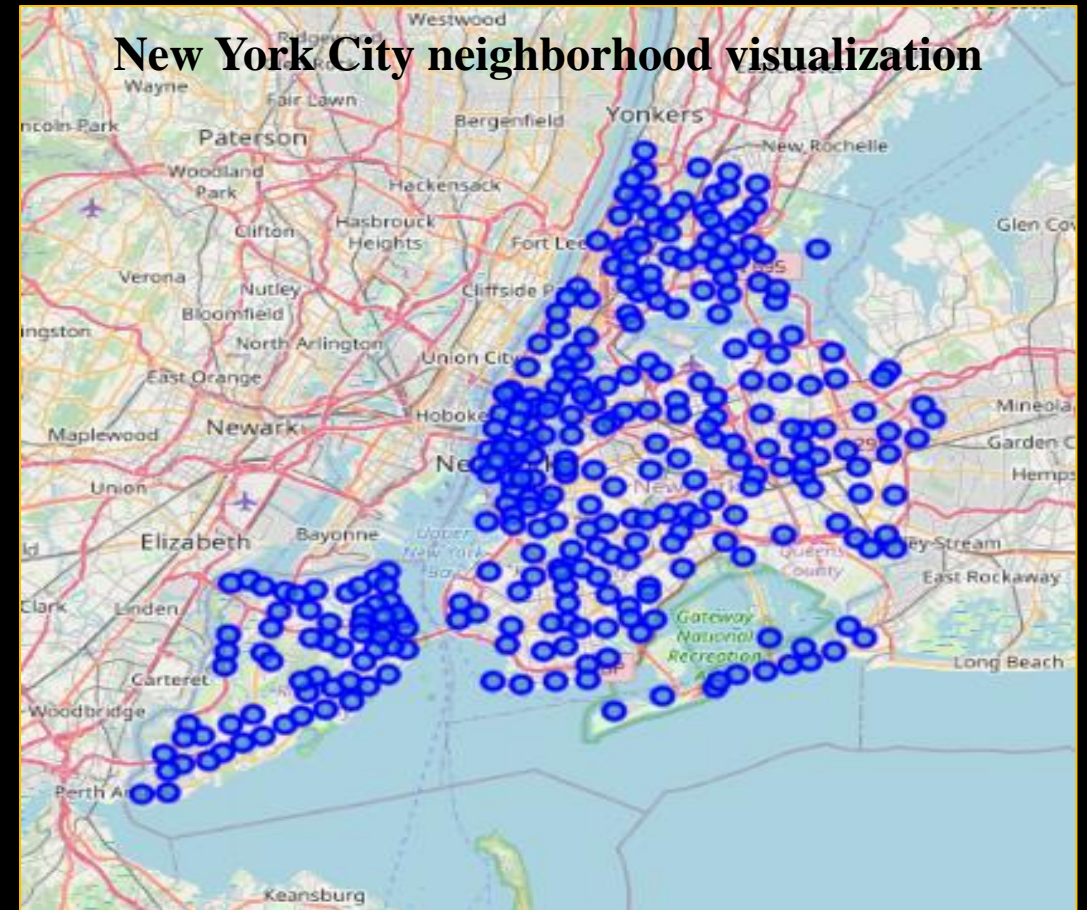
Upon completion of this Data Science analysis we should be able to identify key information to include:

- New York Population by Borough and Neighborhood
- New York City Demographics
- Top Cuisine Served by Borough and Neighborhood
- Segmentation of the Borough
- Untapped Markets
- Saturated Markets



# DATA 1

- Consisted of New York City Geographical Coordinates Data.
  - I loaded the data and explored it from `newyork_data.json` file.
  - I transformed the data of nested python dictionaries into a pandas dataframe containing the geographical coordinates of New York City neighborhoods and tested it by creating a map using Geopy with Folium (see below)



# DATA 2

Consisted of extracting data from NYC Wikipedia pages on population and cuisine. The insights are:

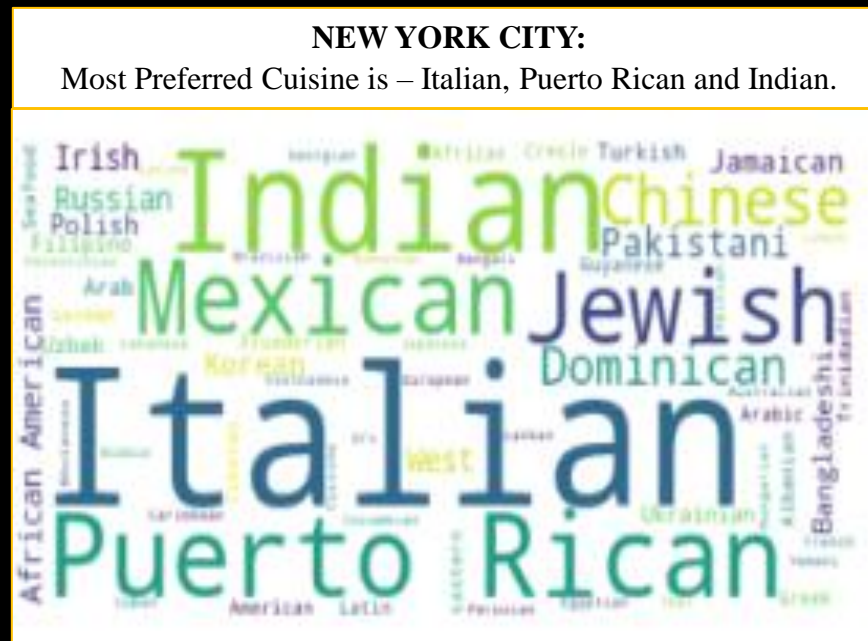
- Manhattan (New York County) is the geographically smallest and most densely populated borough.
- Brooklyn (Kings County), on the western tip of Long Island, is the city's most populous borough.
- Queens (Queens County), on Long Island north and east of Brooklyn, is geographically the largest borough and second most populous.
- Staten Island followed by The Bronx are the least populous boroughs.

	Borough	County	Estimate_2018	GDP	square_miles	square_km	persons_sq_mi
0	The Bronx	Bronx	1,432,132	42.695	29,200	42.10	109.04
1	Brooklyn	Kings	2,582,830	91.559	34,600	70.82	183.42
2	Manhattan	New York	1,628,701	600.244	360,900	22.83	59.13
3	Queens	Queens	2,278,906	93.310	39,600	108.53	281.09
4	Staten Island	Richmond	476,179	14.514	30,300	58.37	151.18
5		City of New York	8,398,748	842.343	97,700	302.64	783.83
6		State of New York	19,745,289	1,701.399	85,700	47,214	122,284



# DATA 2 (CONT)

Cuisine of New York city : Data is taken from Wikipedia page - [https://en.wikipedia.org/wiki/Cuisine\\_of\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Cuisine_of_New_York_City) . This data was using to create word clouds to show the most common cuisine in NYC city first and then broken down by most common cuisine in each borough.



# DATA 2 (CONT)

## MANHATTAN:

Preferred Cuisine is – Italian, Chinese, Puerto Rican and Indian.



**BRONX:**

Preferred Cuisine is – Italian, Puerto Rican, Dominican and Albanian.



**BROOKLYN:**

Preferred Cuisine is – Italian, Puerto Rican and Mexican.



**QUEENS:**

Preferred Cuisine is – Italian, Indian, and Pakistani.

**STATEN ISLAND:**

Preferred Cuisine is – Italian and Indian.





# DATA 3

- New York City data was input into Foursquare API to return venues information and explore neighborhoods in New York City by geographical coordinates.
- FourSquare API calls were made to get the top 200 venues in a radius of 1000 meters and then depicted on this map using folium. →



# RESULTS:

Using the venue data I filtered and used only the FourSquare restaurant data for New York City neighborhoods using K-Means clustering based on mean occurrence of venue category.

- The neighborhoods were put into two clusters using the k-means clustering algorithm.

My Results showed:

- Cluster0 : The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.
- Cluster1 : The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant
cluster0	0.009950	0.019900	0.681592	0.024876	0.004975	0.189055	0.004975
cluster1	0.009901	0.079208	1.584158	0.128713	0.148515	0.603960	0.148515

First few columns

Turkish Restaurant	Udon Restaurant	Ukrainian Restaurant	Vegetarian / Vegan Restaurant	Venezuelan Restaurant	Vietnamese Restaurant	Total	Total Sum
0.009950	1.908196e-17	-8.673617e-18	0.094527	-1.734723e-17	0.049751	9.189055	18.378109
0.277228	2.970297e-02	9.900990e-03	0.435644	1.980198e-02	0.445545	27.940594	55.881188

Last few columns



# RESULTS(CONT):

In New York City there are currently untapped neighborhoods in Staten Island and Bronx.

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Bronx	Clason Point	40.806551	-73.854144	0	0
1	Staten Island	Todt Hill	40.597069	-74.111329	0	0
2	Staten Island	Port Ivory	40.639683	-74.174645	0	0
3	Staten Island	Butler Manor	40.506082	-74.229504	0	0
4	Staten Island	Bloomfield	40.605779	-74.187256	0	0

# RESULTS(CONT):

This map shows the locations of the untapped neighborhoods.





# DISCUSSION:

- There is a wide scope of choices to explore cuisines of various countries and ethnicities in New York City.
- Although the data shows untapped markets in Staten Island and the Bronx. This just shows less competition vs other markets.
- It may not be risky to open a restaurant in Brooklyn where the population is the highest or in Manhattan where tourism data was not really implemented. High risk and high reward can be the outcome of the saturated markets if you have great ambiance and an even greater menu.



# CONCLUSION:

- The analysis was performed on just a fraction of available data.
- Concurrence and validity of the actual data was not confirmed and so may contain errors or bad data.
- Going into more detail on available information could yield better results. Like looking at the hours of operation of nearby businesses, mean income, rents... etc.
- An area saturated with restaurants may just be meeting the demand of that area as you see in Brooklyn and Manhattan. Bronx, Queens and Staten Island have a less robust number of restaurants but also less population, so the demand may or may not be there. It would require more data to be used for a more refined finished product.





# LINKS TO DATA:

- The Notebook for this project was too large to upload to GitHub It can be downloaded/accessed from these areas
- Google Link to Jupyter Notebook file - [Capstone](#)
- Google Link on GitHub Page - [MyGitHub](#)
- IBM Watson Shared Notebook Link - [IBM](#)

