**The Battle of Neighborhoods (Week 1)**

**Project Objective**

The project objective is to identify the most suitable place in New York or Toronto for providing the recommendations to open a new restaurant by analyzing geographical data, neighborhoods, demography and other relevant key contributors.

Both cities (New York or Toronto) are very diverse in nature and the financial capitals for their respective countries, we will focus on demographics (Source Wikipedia) to select the city for further analysis. By looking at demographics, it is quite evident that the City of New York is more populous and diverse city that gives an edge on Toronto.

**Market Study**

Restaurant business caters the need of good multi-cuisine breakfast, food and drinks to its customers. Restaurant must provide quality and delicious multi-cuisine food. The potential customers would be multi-cultural customers influenced by immigration history of city over the period of time.

The most common multi-cuisine foods given below should be available for some potential immigrants (refer to New York’s food Culture).

Central and Eastern European immigrants, especially Jewish immigrants from those regions, brought bagels, cheesecake, hot dogs, knishes, and delicatessens (or delis) to the city. Italian immigrants brought New York-style pizza and Italian cuisine into the city, while Jewish immigrants and Irish immigrants brought pastrami and corned beef, respectively. Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city. Some 4,000 mobile food vendors licensed by the city, many immigrant-owned, have made Middle Eastern foods such as falafel and kebabs examples of modern New York street food. The city is home to "nearly one thousand of the finest and most diverse haute cuisine restaurants in the world".

**Demographics comparison between New York and Toronto**

**City of Toronto Demographics**

The most prevalent ethnic origins in the City of Toronto are as follows:

|  |  |  |
| --- | --- | --- |
|  | **% of total population (2016)** | **Population (2016)** |
| European | **47.70%** | **1,282,750** |
| South Asian | 12.60% | 338,965 |
| Chinese | 11.10% | 299,460 |
| Black | 8.90% | 239,850 |
| Filipino | 5.70% | 152,715 |
| Latin American | 2.90% | 77,160 |
| West Asian | 2.20% | 60,325 |
| Multiple visible minorities | 1.80% | 47,675 |
| Southeast Asian | 1.50% | 41,645 |
| Korean | 1.50% | 41,640 |
| Visible minority, n.i.e. | 1.40% | 36,975 |
| Arab | 1.30% | 36,030 |
| Japanese | 0.50% | 13,410 |
| First Nations | 0.50% | 14,380 |
| Métis | 0.30% | 7,270 |
| Inuit | 0% | 275 |
| Aboriginal, n.i.e. | 0% | 645 |
| Multiple Aboriginal identities | 0% | 500 |
| **Total population** | **100%** | **2,691,665** |

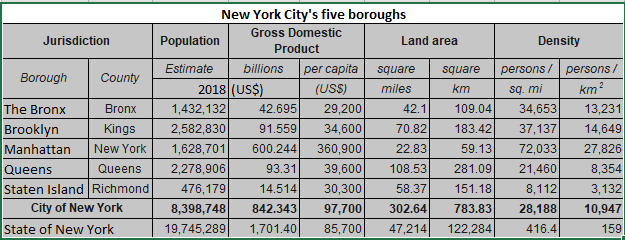
The visible groups per community council is given in [demographics of Neighborhoods of Toronto](https://en.wikipedia.org/wiki/Demographics_of_Toronto#Neighbourhoods).

* **Toronto & East York** (847,045): White: 65.3%, Chinese: 8.9%, South Asian: 6.7%, Black: 5.6%
* **North York** (638,100): White: 47.4%, Chinese: 14.0%, South Asian: 8.5%, Filipino: 8.0%, West Asian: 5.3%, Black: 5.2%
* **Scarborough** (623,135): White: 26.5%, South Asian: 25.4%, Chinese: 19.0%, Black: 10.8%, Filipino: 8.4%
* **Etobicoke York** (583,395): White: 48.9%, Black: 15.7%, South Asian: 11.9%, Latin American: 5.6%

|  |  |
| --- | --- |
| **Religion in Toronto** | |
| **Religion** | **% of total population** |
| Christian | 54.10% |
| Muslim | 8.20% |
| Hindu | 5.60% |
| Jewish | 3.80% |
| Buddhist | 2.70% |
| Sikh | 0.80% |
| Other | 0.60% |
| No religion | 24.20% |

**City of New York Demographics**

|  |  |
| --- | --- |
| **Racial composition** | **% of Population** |
| White | 44.00% |
| —Non-Hispanic | 33.30% |
| —Hispanic | 10.70% |
| Black or African American | 25.50% |
| Hispanic or Latino (of any race) | 28.60% |
| Asian | 12.70% |



**Methodological Approach**

The City of New York, is more populous city than Toronto and is the most populous city in the United States. It is more diverse, multicultural and is financial hub of North America. Being financial hub, it provides more business opportunities and environment. Major industries are banking, insurance, finance, retail, capital market, retail, transport, fashion, entertainment and some more. New York is one of the most costly city in the world due to it’s highly competitiveness for setting up new business. To start a new venture or business expansion needs regressive market study, potential customers and availability of raw material and man power. Based on demographics comparison, **choice of the city is New York.**

It is important to survive in competitive market with strategic plan. Below factors need to be analyzed to make decision to identify the location for new restaurant.

1. New York Population
2. New York City Demographics
3. Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality and cost
4. Nearby venues like Gyms, Entertainment zones, Parks, entertainment and business complexes, so that more customers can be attracted
5. Nearby restaurants and their Cuisine served / Menu
6. Segmentation of the Borough
7. Untapped markets
8. Saturated markets

Even though well-funded XYZ Company Ltd. need to choose the correct location to start its first venture. If this is successful they can replicate the same in other locations. First move is very important, thereby choice of city is very important.

**Choice of City – New York**

**Availability of data related to New York City**

Below datasets will be analyzed for New York City.

1. Neighborhood has a total of 5 boroughs and 306 small neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

<https://geo.nyu.edu/catalog/nyu_2451_34572>

1. **DOHMH Farmers Markets and Food Boxes dataset**

<https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2>

<https://www.grownyc.org/greenmarketco/foodbox>

GrowNYC's Fresh Food Box Program is a food access initiative that enables under-served communities to purchase fresh, healthy, and primarily regionally grown produce well below traditional retail prices.

A farmers' market is often defined as a public site used by two or more local or regional producers for the direct sale of farm products to consumers. In addition to fresh fruits and vegetables, markets may sell dairy products, fish, meat, baked goods, and other minimally processed foods.

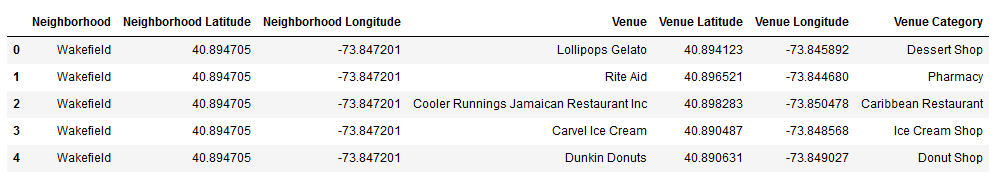
1. **Demographics data from Wikipedia**

* New York Population
* New York City Demographics
* Cuisine of New York city

<https://en.wikipedia.org/wiki/New_York_City>   
<https://en.wikipedia.org/wiki/Economy_of_New_York_City>   
<https://en.wikipedia.org/wiki/Portal:New_York_City>   
<https://en.wikipedia.org/wiki/Cuisine_of_New_York_City>   
https://en.wikipedia.org/wiki/List\_of\_Michelin\_starred\_restaurants\_in\_New\_York\_City

1. **New York Location data**

New York City geographical coordinates data will be utilized as input for the Foursquare API which will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in New York City. The below is image of the Foursquare API data.

[](https://i.imgur.com/MOxxyoY.png)

**Analytic Approach**

Below approaches will be used to identify the location for new restaurant business.

New York City neighborhoods have 5 boroughs and 306 small neighborhoods. Clustering of Manhattan, Brooklyn, Bronx, Queens and Staten Island is needed for analysis.

**Exploratory Data Analysis**

**Part 1 - New York City Geographical Coordinates Data.**

1. We load and explore the data from newyork\_data.json file.

2. Transform the data of nested python dictionaries into a pandas data frame.

3. This data frame contains the geographical coordinates of New York City neighborhoods.

4. This data will be used to get Venues data from Foursquare.

5. We used geopy and folium libraries to create a map of New York City with neighborhoods

Superimposed on top.

**Part 2- DOHMH Farmers Markets and Food Boxes dataset.**

We will be using the data of Farmers Markets. There are total of 144 Farmers Markets in New York City. Majority of Farmers Markets are in Manhattan and Brooklyn and lowest in Queens, Bronx and Staten Island.

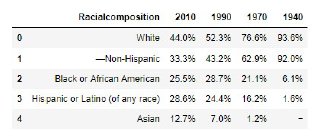
We will use geopy and folium libraries to create a map to visualize farmers markets of New York City.

**Part 3: To analyze New York City Population, Demographics and Cuisine**

To analyze the New York City Population, Demographics and Cuisine, scrapped the data from Wikipedia pages given above in the data section. We will use Beautiful Soup python library. Beautiful Soup is a Python package for parsing HTML and XML documents (including having malformed markup, i.e. non-closed tags, so named after tag soup). It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping.

* **New York City Population** : Insights from the data :
  + Manhattan (New York County) is the geographically smallest and most densely populated borough.
  + Manhattan's (New York County's) population density of 72,033 people per square mile (27,812/km²) in 2015 makes it the highest of any county in the United States and higher than the density of any individual American city.
  + 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.
* **New York City Demographics** : 
  + New York City is the most populous city in the United States with an estimated record high of 8,622,698 residents as of 2017, incorporating more immigration into the city than outmigration since the 2010 United States Census.

The racial composition is as given below. This is the reason New York City has restaurants serving cuisine from many countries such as Indian, African, Japan etc. This also increases the scope for restaurants business in New York City.



* **Cuisine of New York city** :
  + Data is taken from Wikipedia page <https://en.wikipedia.org/wiki/Cuisine_of_New_York_City>
  + Word cloud on this data
* **NEW YORK CITY CUISINE** 
  + Preferred Food in New York City –Italian, Puerto Rican, Mexican, Jewish, Indian, Pakistani & Dominican.
* **BROOKLYN CUISINE**
  + Preferred Food in Brooklyn is –Italian, Puerto Rican & Mexican
* **MANHATTAN CUISINE**
  + Preferred Food in Manhattan is – Italian, American, Puerto Rican and Indian.
* **QUEENS CUISINE**
  + Preferred Food in Queens is – Indian, Irish, Pakistani and Mexican.
* **THE** **BRONX CUISINE**
  + Preferred Food in The Bronx is – Italian, Puerto Rican, Albanian and Dominican.

There is very less data of cuisine relating to Staten Island. So cannot develop word cloud with it.

**Part 4 : New York city geographical coordinates data**

Location data will be utilized as input for the Foursquare API that has been leveraged to provision venues information for each neighborhood. We will use the Foursquare API data to explore neighborhoods in the New York City.

**Brooklyn, Manhattan, Bronx, Queens and Staten Island Neighborhoods Visualization**

Using the geographical coordinates of each neighborhoods foursquare API calls are made to get top 200 venues in a radius of 1000 meters. The venues data is as given below:

**Brooklyn, Manhattan, Bronx, Queens and Staten Island Venues Visualization**

We will generate the Brooklyn and Manhattan Venues Visualization. The "BM\_venues" data frame has 9708 venues and 397 unique venue types.

Bronx, Queens and Staten Island Venues Visualization: The "BQS\_venues" data frame has 10805 venues and 387 unique venue types.

**Outcome**

We will filter and use only restaurant data for Brooklyn, Manhattan, Bronx, Queens and Staten Island clustering using venue data.

Neighborhood K-Means clustering based on mean occurrence of venue category:

We cluster the neighborhoods into two different clusters and will use the K-Means clustering Algorithm. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. It uses iterative refinement approach.

**Brooklyn, Manhattan, Bronx, Queens and Staten Island**

In the Map Visualization, we will see the different types of clusters created by using K-Means for each neighborhoods

Cluster0: The Total and Total Sum of cluster0 has smallest value. It will show that the market is not saturated.

Cluster1: The Total and Total Sum of cluster1 has highest value. It will show that the markets are

Saturated. Number of restaurants are very high.

**Observations**

1. There is scope to increase Farmers markets in Bronx, Queens and Staten Island.

2. There is scope to explore cuisines of various countries in Bronx, Queens and Staten Island.

3. In Manhattan and Brooklyn restaurants of cuisines of many countries are available. So if risk

can be taken with great menu on board. It also shows people love eating cuisines of various

Countries.

This analysis will be performed on limited data which may impact the accuracy of result. It is always better to perform analysis on sizable data to get more results.

The no. of restaurants present shows that there is a lot of demand. Brooklyn and Manhattan have high concentration of restaurant businesses and very competitive market. Bronx, Queens and Staten Island also has good number of restaurants but not as many as required. So there is an option to explore.