Coursera Project

## IBM Applied Data Science Capstone

The Battle of Neighborhoods

Introduction:

We are going to find the suitable location to open a new restaurants in Brooklyn, NY. For that we are going to build a data frame of neighborhood in Brooklyn, NY by web scraping the data from internet and will explore the data with the 'Foursquare' API to get the geographical coordinated of the neighborhoods. Once we gets the needed data from our analysis, we will explore and cluster the neighborhoods to get the best cluster to open a new restaurants in the Brooklyn.

Restaurant Industry:

It is no surprise to learn that owning a restaurant comes with its fair share of responsibilities that require to be on own feet. And although it may seem like a dream come true eventually, a restaurant owner initially has to go through several months of constant designing, overlooking and perfecting every little detail that comes hand-in-hand with the project. Recently number of restaurants is growing in New York.
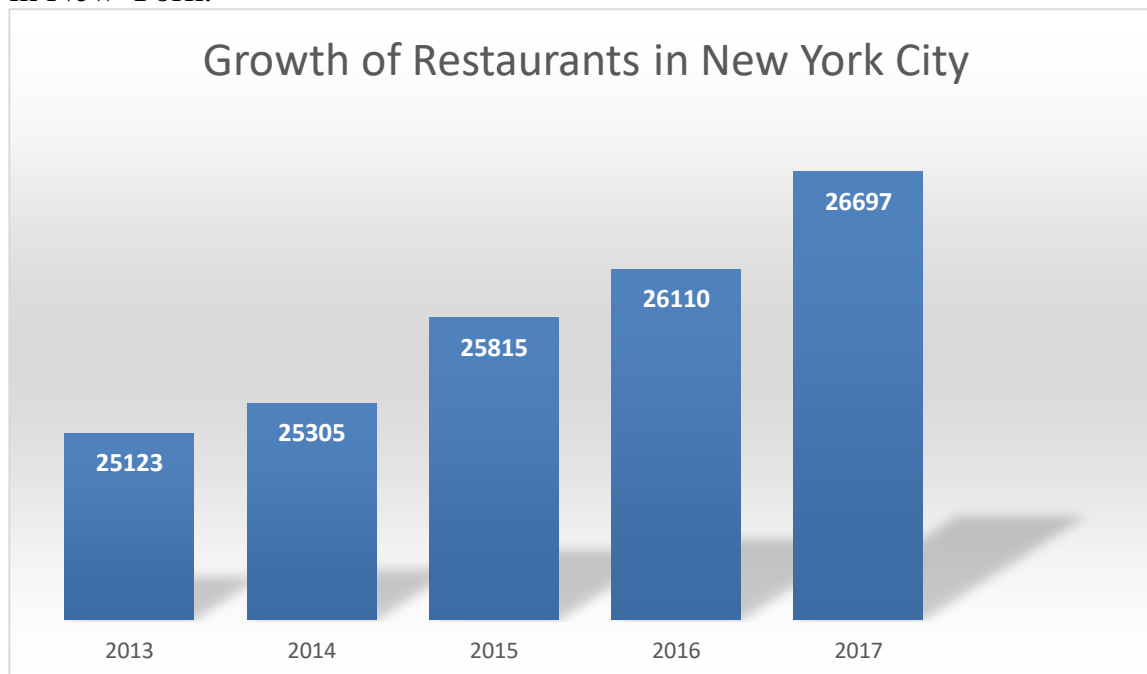
Often there's an issue of time constraint that keeps a restaurant owner preoccupied with intricacies. Whether it is sourcing the right ingredients, obtaining too many licenses, finalizing a space within your budget to interviewing the right staff — restaurateurs have to consider everything beginning from a pin to piano before taking the right foot forward.

Amid the numerous challenges that we've mentioned, here are a few that especially take the forefront for restaurateurs during the teething phase of every project. In this analysis, we are concentrating on choosing the location for restaurants.

- Pre-opening Planning
- Choosing a Location
- Financial
- Planning
- Competition
- Finding the Right Suppliers
- Restaurant Design & Layout
- Restaurant Marketing Tactics
- Staffing
- Guest Management
- Sustainability
- Inventory Management

Business Problem

It is essential to start the restaurants in the best place to have a good growth in the business, not every food-service operation needs to be in a retail location, but for those that do depend on retail traffic, here are some factors to consider when deciding on a location:

How will the location contribute to your sales volume?

- **Accessibility to potential customers.** Consider how easy it will be for customers to get into your business. If you are relying on strong pedestrian traffic, consider whether or not nearby businesses will generate foot traffic for you.

- **The rent-paying capacity of your business.** If you've done a sales-and-profit projection for your first year of operation, you will know approximately how much revenue you can expect to generate, and you can use that information to decide how much rent you can afford to pay.

- **Restrictive ordinances.** You may encounter unusually restrictive ordinances that make an otherwise strong site less than ideal, such as limitations on the hours of the day that trucks can legally load or unload.

- **Traffic density.** With careful examination of food traffic, you can determine the approximate sales potential of each pedestrian passing a given location. Two factors are especially important in this analysis: total pedestrian traffic during business hours and the percentage of it that is likely to patronize your food service business.

- **Customer parking facilities.** The site should provide convenient, adequate parking as well as easy access for customers.

- **Proximity to other businesses.** Neighboring businesses may influence your store's volume, and their presence can work for you or against you.

- **History of the site.** Find out the recent history of each site under consideration before you make a final selection. Who were the previous tenants, and why are they no longer there?

- **Terms of the lease.** Be sure you understand all the details of the lease, because it's possible that an excellent site may have unacceptable leasing terms.

- **Future development.** Check with the local planning board to see if anything is planned for the future that could affect your business, such as additional buildings nearby or road construction.

Who will benefit?

This solution is suitable for organizations looking to establish the restaurant business especially in New York City.

Data Section:

List of neighbourhoods in Brooklyn, New York. This defined the scope of this project.

Neighbourhood has a total of 5 boroughs and 306 neighbourhoods. In order to segment the neighbourhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the latitude and longitude coordinates of each neighbourhood and we are going to use dataset through the url (https://geo.nyu.edu/catalog/nyu_2451_34572)

We will use the Foursquare API to explore neighborhoods in New York City by using the **explore** function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters and will use the *k*-means clustering algorithm to complete this task. Finally, we will be using the Folium library to visualize the neighborhoods in New York City and their emerging clusters.

Methodology:

Firstly, we need to get the list of neighborhoods in the city of Brooklyn. We are using the data from the NYU website https://geo.nyu.edu/catalog/nyu_2451_34572. We will do web scraping using the python and beautiful soup packages to the extract the list of neighborhood data with latitudes and longitude details. After processing the data, we will create a data frame with the relevant information and will be visualize to map using the folium package.

We will use 'FourSquare' API to get the most common venues in the Brooklyn, NY. As we are having the Foursquare account, we will be using the client id and secret key to retrieve the most common venues in the Brooklyn, NY. Based on the venues details, the data will be categorized. We will analyses each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category. After that, we will prepare the data to use in clustering.
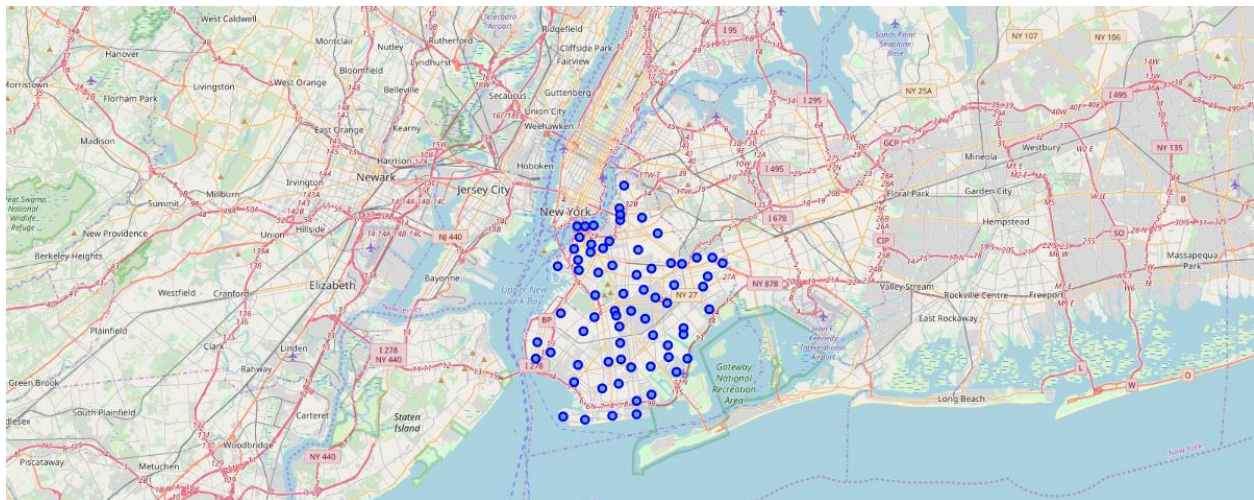
We will perform clustering on the data by using the k-means clustering. k-Means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is the one of the unsupervised machine learning. The result will help us to identify the best place to start the restaurants in the Brooklyn, NY.
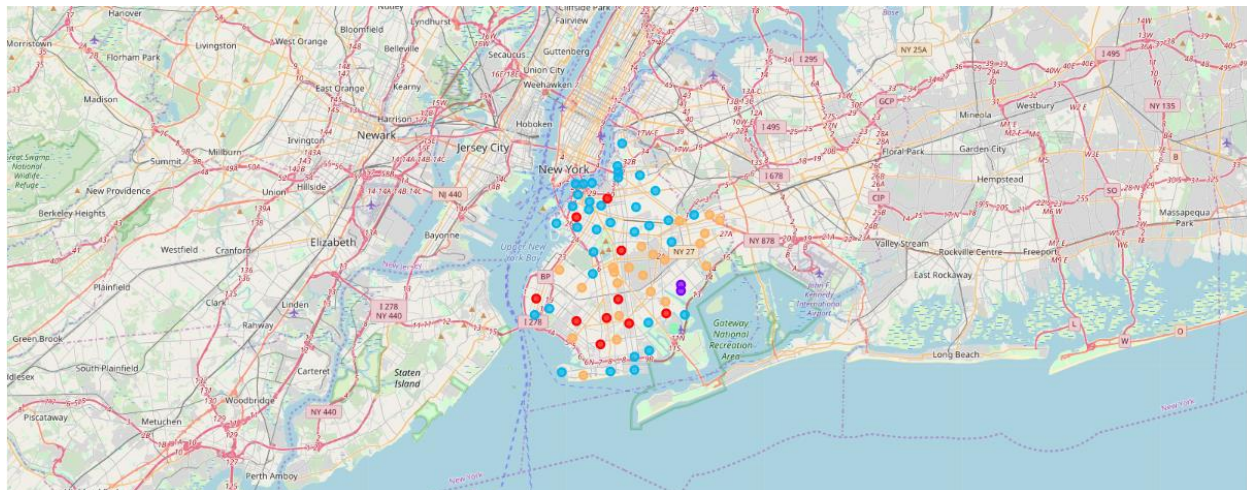
Result:

The result from the k-means clustering show that we can categorize the neighborhoods into 4 cluster based on the frequency of occurrence.

Cluster 0   : Neighborhoods with moderate frequency of operations.

Cluster 1, 2: Neighborhoods with low frequency of operations

Cluster 3, 4: Neighborhood with high frequency of operations

The results were visualized in the map below with color differentiation

Discussion:

Based on the results from the map, the most of the restaurants were in the north side of the Brooklyn, so it is the best place to start the new restaurants. The frequency of occurrence is higher in the 'Brooklyn Heights' and it is the right place to start the restaurant, which mainly concentrate on the organic foods, fruits and salads, as the competency in this place is low and the surroundings were more occupied with Gym, Park and Fitness centers.

Limitations:

In this analysis, we only consider the frequency of occurrence there are other factors like population, income of residents, transportation availability could

influence the decision. Future research can concise the methodology to estimate the better results.

Conclusion:

We have discussed the process of identifying the business problem, data requirement, preparing the data for analysis and performed the machine learning by k-means clustering. The conclusion on the project is the best place to start restaurants in northern Brooklyn and starting the fruit and salad restaurants with less junk foods will be the greatest combination to start new restaurants in the Brooklyn, NY.

References and Appendix:

1. Source of Data: https://geo.nyu.edu/catalog/nyu_2451_34572
2. Foursquare Developers document https://developers.foursquare.com/docs