**The Battle of Neighborhoods - New York City**

**Introduction:** New York City is multicultural city with varied of food culture. Explore, analyze and provide suitable location for opening new restaurant to Foodie Inc.

**Business Problem description and background:** Foodie Inc. is an international restaurant chain looking to expand its operations in New York City within one of the following Boroughs – Manhattan, Brooklyn and Bronx. As New York City is the financial capital of the country with multicultural population, Foodie Inc needs help with finding a suitable neighborhood among the 3 Boroughs to open their restaurant. Foodie Inc. requires help with insights on the restaurants present in the neighborhoods, most popular restaurants, frequency of people visiting the restaurants and compares to crime reported in the neighborhoods. We need to provide report of the suitable neighborhood to start the restaurant.

**Data Description:**

**Data 1:** New York City has 5 Boroughs and 306 Neighborhoods. To segment, explore and analyze the neighborhoods we need to use a dataset that contains 5 boroughs and all its neighborhoods with latitude and longitude coordinates. Dataset exists for free in web.

Link to New York City dataset: <https://cocl.us/new_york_dataset>

**Data 2:**  Latitude and Longitude of New York city neighborhoods, from Data 1 will be used to retrieve popular venues in the area using Foursquare API. From the retrieved venues list, categories of type restaurants will be segregated and used for further analysis.

Link to Foursquare API: <https://api.foursquare.com/v2/venues/explore>

**Data 3:** Third data used for the analysis is the crime incidents reported in New York in the last one year. This dataset is available in web.

Link to Crime report: <https://data.cityofnewyork.us/resource/5uac-w243.json>

**Approach of data usage to solve problem:**

The New York City borough and neighborhoods data retrieved from Data 1 along with the latitude and longitude gives the list of neighborhoods to explore. Using the latitude and longitude coordinates as input to Foursquare API (Data 2), retrieve the details on restaurants in the neighborhoods. Now with the crime report available in Data 3 web link find the boroughs with higher crime rate. Compare the crime rate of boroughs, the restaurants clusters and find the neighborhoods with the highest crime rate. Shortlist the neighborhood with the lowest crime rate and the frequency of restaurant is less.

**Methodology:**

The methodology used in the analysis includes:

1. Load each of the data set to dataframe
2. Study the crime reported in all of the Boroughs
3. Analyze the lowest crime Borough and retrieve the neighborhoods
4. Retrieve the restaurants in the New York neighborhoods using Foursquare api
5. Analyze the neighborhood and the frequency of the restaurants
6. Perform k-means statistical analysis on venues by neighborhoods comparing the crime report
7. Determine the neighborhood suitable for restaurant

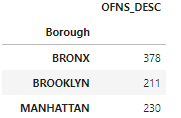
**Analytic Approach:**

New York City has 5 Boroughs and 306 neighborhoods. Out of the 5 Boroughs, we require to analyze 3 Boroughs – Manhattan, Bronx and Brooklyn. First we need to lowest crime reported borough and cluster the neighborhoods for the borough. In this analysis only restaurant data is filtered from Foursquare venues data and utilized. This process is followed due to the below exploratory analysis.

**Exploratory Data Analysis:**

**New York Crime Data:**

Load the required libraries to begin with analysis. The crime dataset is available as json in web link provided in Data 3 is downloaded from Open Data for New York City website. This dataset is loaded to a dataframe for further analysis. The crime data is available for each of the boroughs. Dataset is grouped by boroughs and the count for each borough is calculated. The lowest crime reported data among the 3 selected boroughs is Brooklyn. Manhattan has 230 crime reports, Bronx has 378 crime reports and Brooklyn with 211 crime reports.

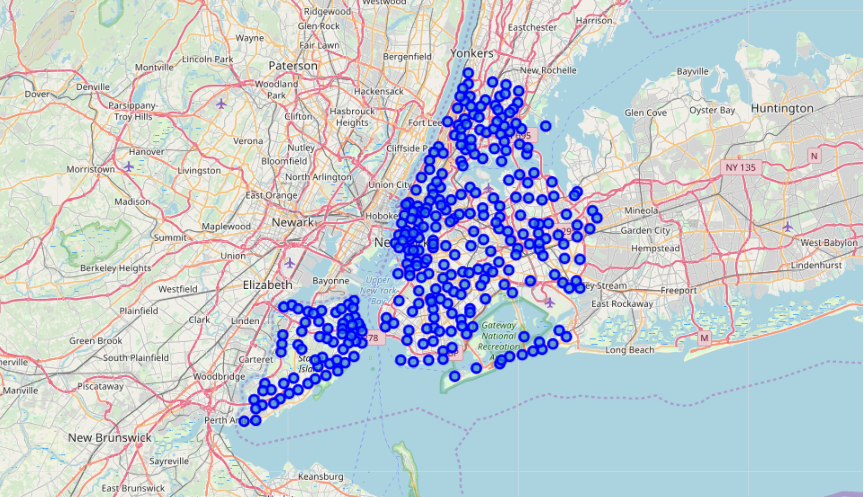


**Neighborhood Data:**

**Analyzing New York Boroughs and Neighborhoods**

New York City neighborhood data available in Data 1 is loaded to dataframe. Neighborhood data has boroughs, neighborhood names, latitude and longitude. With the geographical coordinates of New York City a map is created with neighborhoods superimposed on it as below.

**New York City neighborhoods map**



**Analyzing Brooklyn neighborhoods:**

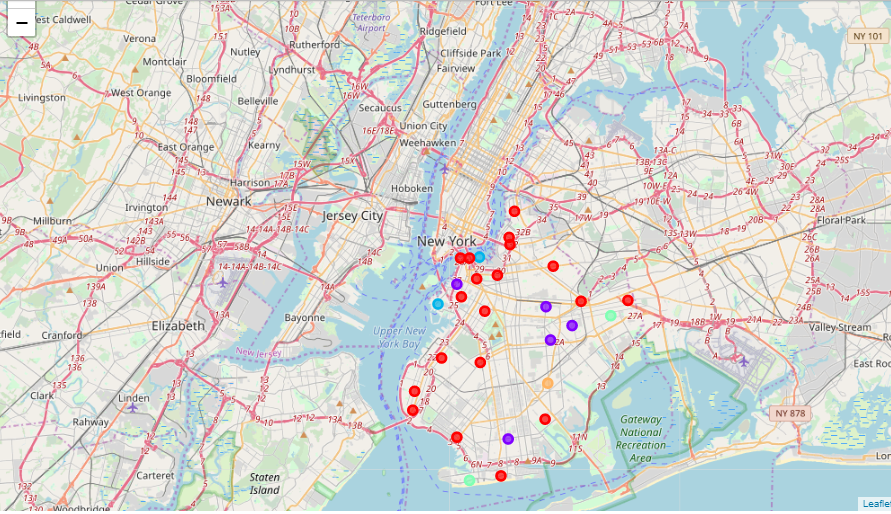
New York City crime report analysis shows Brooklyn is the lowest crime reported borough among Manhattan and Bronx. Neighborhoods dataframe is updated only with Brooklyn boroughs; rest of the boroughs is dropped from the dataframe. Using Foursquare api retrieve the venues in the neighborhoods. Brooklyn has 70 neighborhoods, only the restaurants in the Foursquare response is sorted and analyzed. Neighborhoods in Brooklyn resulted with 34 unique restaurant categories.

Further analyzing the neighborhoods data, grouping by neighborhoods, the mean of the frequency of occurrence of each category of venue categories, we study the top five most frequented venues. Putting this into dataframe we can determine the most common venues by neighborhood and plot it on a map.

**Results:**

The filtered out restaurant data for Brooklyn, is clustered and analyzed further. Neighborhood k-means clustering based on mean occurrence of venue category is performed. In order to cluster the Brooklyn neighborhoods k-means clustering algorithm is used. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. This uses iterative refinement approach.

**Visualizing Clusters in a map:**



**Cluster 1:** First cluster with cluster labels 0 resulted with 19 neighborhoods and its most frequented restaurants locations. This cluster is saturated and has more number of restaurants.

**Cluster 2:** Second cluster with cluster labels 1 resulted with 5 neighborhoods and its most frequented restaurant locations. This cluster has opportunities to open restaurants.

**Cluster 3:** Third cluster with cluster labels 2 resulted with 2 neighborhoods and its most frequented restaurant locations. This cluster has more opportunities to open restaurants.

**Cluster 4:** Fourth cluster with cluster labels 3 resulted with 2 neighborhoods and its most frequented restaurant locations. This cluster has more opportunities to open restaurants.

**Cluster 5:** Fifth cluster with cluster labels 4 resulted with 1 neighborhood and its most frequented restaurant locations. This cluster has the most opportunity to start new restaurants.

**Discussion:**

1. Brooklyn is the borough with lowest crime reported among Manhattan and Bronx
2. Neighborhoods under cluster 1 are saturated and have more number of restaurants. It is not suitable location to begin new restaurant.
3. Neighborhood resulted in cluster 5, Flatlands has the highest opportunity in Brooklyn to open a new restaurant
4. We were able to determine the top 10 most frequented venues by location of interest
5. Statistically, we determined there is more opportunity in Flatlands neighborhood.

**Conclusion:**

Using combination of datasets from New York City Open Data project, New York City crime data and Foursquare venue data we were able to explore, analyze the boroughs and neighborhoods, statistically describe venues by locations of interest. The data available for the analysis is limited, with more descriptive data available for analysis the results could possibly change.