**Capstone Project - The Battle of Neighborhoods**

**Brooklyn Food Truck Location Recommendations**

**I. PURPOSE**

This document provides the details of my peer reviewed assignment for the IBM Data Science Professional Certificate Capstone project.

**II. INTRODUCTION**

New York City is the largest city in U.S.A by population, which also encompasses five divisions called boroughs. Of the five boroughs we will focus on Brooklyn, which is known for its night life and great food.

**III. Objective**

In this project, we will study the areas of Brooklyn in order to determine which neighborhoods have the highest concentration of bars in order to solve our business problem.

**A description of the problem and discussion of the background. (15 marks)**

Our business problem is that we are working with a Food Truck business owner who is trying to determine where he should take his food truck for the greatest chance of success.  
The aim of this project is to segment area's of Brooklyn's neighborhoods based on the concentration of bars using the Foursquare data.  
We therorize that the best places to take a food truck at night is near areas that have high concentrations of bars.  
This theory is based on two main principals.

1. Not all bars have a kitchen and sell food, so targeting these areas fills a need for our customers.
2. People going home from the bars are more likely to buy food to enjoy late at night when many other food options are closed.

Thus the area's with the highest concentration of bars will have the highest concentrations of potential customers.

**A description of the data and how it will be used to solve the problem. (15 marks)**

We will be using the Foursquare Places API for our data source. <https://developer.foursquare.com/docs/api>  
From this API we will use the Venue Categories and Venue Location data to map our data on a Folium map.  
From here we will use K-Means clustering to find the ideal locations to send the Food Trucks to.

**III. Methodology**

In order to determine where is the best place to take our food truck we will take a multi step process.

1. First we will determine which neighborhood in Brooklyn we should focus on.
2. We will then target a specific area of this neighborhood with our Foursquare data in order to determine which area will have the highest number of potential customers for our food truck business.

We will use K-Means clustering to group our neighborhoods by common venues, and from there we will use information from the Foursquare API such as venue rating to determine which bars are the most popular.

# Segmenting and Clustering Neighborhoods in Brooklyn

## 1. Download and Explore Dataset

Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement 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 the latitude and logitude coordinates of each neighborhood.

Luckily, this dataset exists for free on the web. Feel free to try to find this dataset on your own, but here is the link to the dataset: <https://geo.nyu.edu/catalog/nyu_2451_34572>

#### Load and explore the data

Next, let's load the data.

{'type': 'Feature',

'id': 'nyu\_2451\_34572.1',

'geometry': {'type': 'Point',

'coordinates': [-73.84720052054902, 40.89470517661]},

'geometry\_name': 'geom',

'properties': {'name': 'Wakefield',

'stacked': 1,

'annoline1': 'Wakefield',

'annoline2': None,

'annoline3': None,

'annoangle': 0.0,

'borough': 'Bronx',

'bbox': [-73.84720052054902,

40.89470517661,

-73.84720052054902,

40.89470517661]}}

#### Tranform the data into a pandas dataframe

The next task is essentially transforming this data of nested Python dictionaries into a pandas dataframe. So let's start by creating an empty dataframe.

Take a look at the empty dataframe to confirm that the columns are as intended.

| **Borough** | **Neighborhood** | **Latitude** | **Longitude** |
| --- | --- | --- | --- |

Then let's loop through the data and fill the dataframe one row at a time.

Quickly examine the resulting dataframe.

| **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 |

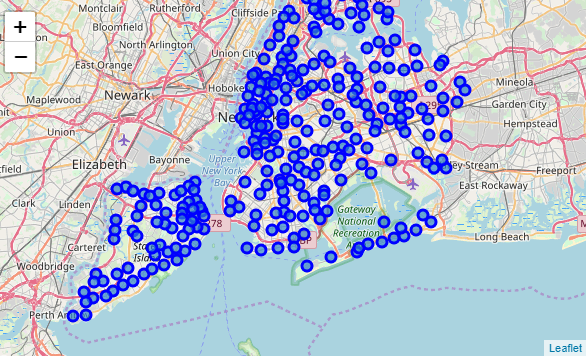
And make sure that the dataset has all 5 boroughs and 306 neighborhoods.

The dataframe has 5 boroughs and 306 neighborhoods.

#### Use geopy library to get the latitude and longitude values of Brooklyn, NY.

The geograpical coordinate of Brooklyn are 40.6501038, -73.9495823.

#### Create a map of New York with neighborhoods superimposed on top.

 **Folium** is a great visualization library. Feel free to zoom into the above map, and click on each circle mark to reveal the name of the neighborhood and its respective borough.

However, for illustration purposes, let's simplify the above map and segment and cluster only the neighborhoods in Brooklyn. So let's slice the original dataframe and create a new dataframe of the Brooklyn data.

## 2. Explore Neighborhoods in Brooklyn

#### Let's create a function to repeat the same process to all the neighborhoods in Brooklyn

#### Now write the code to run the above function on each neighborhood and create a new dataframe called brooklyn\_venues.

In [ ]:

#### Let's check the size of the resulting dataframe

(2845, 7)

|  | **Neighborhood** | **Neighborhood Latitude** | **Neighborhood Longitude** | **Venue** | **Venue Latitude** | **Venue Longitude** | **Venue Category** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | Bay Ridge | 40.625801 | -74.030621 | Pilo Arts Day Spa and Salon | 40.624748 | -74.030591 | Spa |
| 1 | Bay Ridge | 40.625801 | -74.030621 | Bagel Boy | 40.627896 | -74.029335 | Bagel Shop |
| 2 | Bay Ridge | 40.625801 | -74.030621 | Cocoa Grinder | 40.623967 | -74.030863 | Juice Bar |
| 3 | Bay Ridge | 40.625801 | -74.030621 | Pegasus Cafe | 40.623168 | -74.031186 | Breakfast Spot |
| 4 | Bay Ridge | 40.625801 | -74.030621 | Ho' Brah Taco Joint | 40.622960 | -74.031371 | Taco Place |

Let's check how many venues were returned for each neighborhood

|  | **Neighborhood Latitude** | **Neighborhood Longitude** | **Venue** | **Venue Latitude** | **Venue Longitude** | **Venue Category** |
| --- | --- | --- | --- | --- | --- | --- |
| **Neighborhood** |  |  |  |  |  |  |
| Bath Beach | 49 | 49 | 49 | 49 | 49 | 49 |
| Bay Ridge | 86 | 86 | 86 | 86 | 86 | 86 |
| Bedford Stuyvesant | 28 | 28 | 28 | 28 | 28 | 28 |
| Bensonhurst | 31 | 31 | 31 | 31 | 31 | 31 |
| Bergen Beach | 7 | 7 | 7 | 7 | 7 | 7 |

#### Let's find out how many unique categories can be curated from all the returned venues

There are 286 uniques categories.

## 3. Analyze Each Neighborhood

And let's examine the new dataframe size (2845, 286)

#### Lets see which Neighborhoods have Bar as their most common venue. We will use this later to decide which Neighborhoods to focus on.

3 Greenpoint

13 Prospect Heights

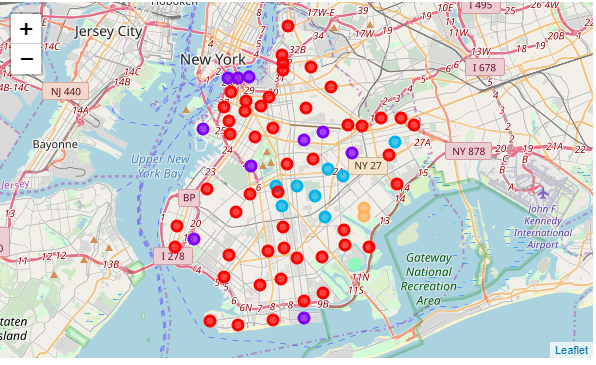
16 Bushwick

49 East Williamsburg

51 South Side

Name: Neighborhood, dtype: object

## Map of all clusters in Brooklyn



#### Lets recenter our map around Williamsburg

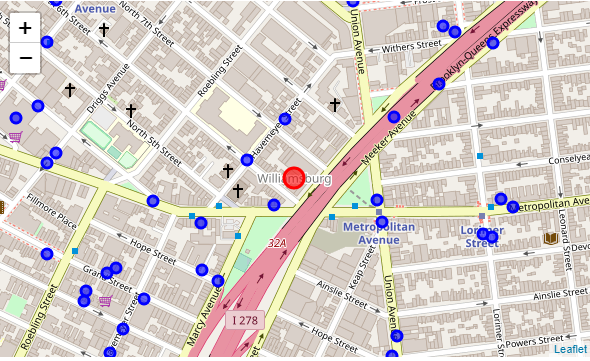
The geograpical coordinate of Williamsburg are 40.714622, -73.95345.

## Map of clusters in Brooklyn where Bars are the most common venue

## According to the data we should focus on venues around Williamsburg because it has the highest density of bars

# Next we will determine which area in Williamsburg is best to send our food truck

#### Let's visualize the Bars that are nearby



## There appears to be a nice cluster of bars around Grand Street and Havemeyer Street.

## Lets investigate Caracas Arepa Bar since it is near that area

9.1

## This is a great rating!

# Explore a location

## Let's explore the popular spots around the bar. In order to explore the area, let's start by getting the latitude and longitude values of Caracas Arepa Barz.

## Let's visualize these venues on the map around our location

## This shows there is a high density of venues and people near this area

# Results Section

#### We followed the methodology that way described above and came to the conclusion that Williamsburg was the best place to send our food truck and of the areas in Williamsburg to focus on, the area near Grand Street and Havemeyer Street would give us the best location to send our food truck

# Discussion Section

#### Observations that I made during this analysis were that Brooklyn has many areas and neighborhoods where bars can be found. For our purposes we wanted to find the area that would have the highest concentration of bars and night life. We came to the conclusion that Williamsburg had the highest concentration of bars within a walking distance of each other.

#### Knowing that people in Brooklyn tend to walk to most places we want to send our food truck to places that will have high foot traffic. We were able to isolate a specific corner in Williamsburg (Grand Street and Havemeyer Street) that would be a good location for our food truck. This is because there are many venues around this location that would have potential customers.

#### Further considerations for improving this model would be the time of day, which has an impact on foot traffic in Brooklyn. Also, we might also want to consider how far away mass transit is from our location. We might be able to get high foot traffic by finding locations that are near subway stops in Williamsburg as well.

# Conclusion Section

### In conclusion, we have shown that k-means clustering combined with Foursquare API data can be used to make business recommendations leveraging machine learning methods. We were able to find a specific area in a specific neighborhood that optimizes our chances for success.[¶](https://nbviewer.jupyter.org/github/JPLAVALLEY/Capstone_Project/blob/master/Capstone.ipynb#In-conclusion,-we-have-shown-that-k-means-clustering-combined-with-Foursquare-API-data-can-be-used-to-make-business-recommendations-leveraging-machine-learning-methods.--We-were-able-to-find-a-specific-area-in-a-specific-neighborhood-that-optimizes-our-chances-for-success.)

In [ ]: