# Borough and Neighborhood for a New Movie Theatre

**Victor bahlangene**

**May 2021**

## **Introduction: Business Problem**

The Goal of this project is to find which Borough and Neighborhood in New York City United States Of America (since it contributes the most to the Global Box office numbers) is best to open up a new theatre.

In early 2020 there was a worldwide pandemic stopped people from going to the theatre. Many entertainment companies rely on the Hollywood box office as a major source of income. Because the movie theatres were closed for over a year many of them were forced to shut down. Beginning of spring 2021 vaccines have made it possible for movie theatres to open, and some at full capacity. There is however an opportunity for media companies to open new theatres to replace the ones that went out of business.

Prior to the pandemic, there was a gradual drop in the number of people going to the movie theatre, so the new theatre cannot necessary be in the location of the closed one, hence the need to find a new location.

We will use data science tools to fetch the raw data, visualize it then generate a few most promising areas. Stakeholders can make the final decision base on the analysis.

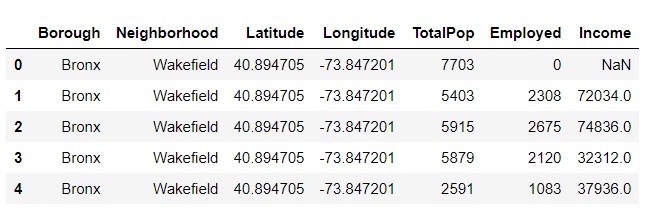
## **Data**

According to the project definition, these are the properties we are looking for in our data:

* **Richest borough**
* **Population data of New York City**
* **Busy entertainment area**
* **Fewest number of theatres in the borough (*ie low competition*)**
* **Of the five Boroughs in New York, which is the richest(*highest income*)**

**Richest borough**

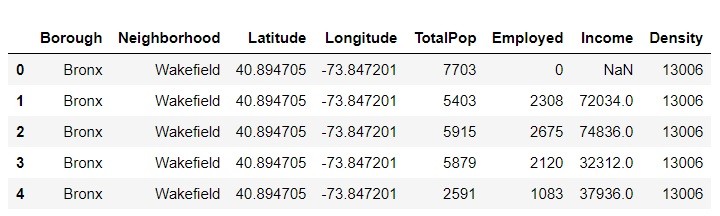
We use the kaggle data set [1]

****

We also use a Json file and convert it to a DataFrame[2]

****

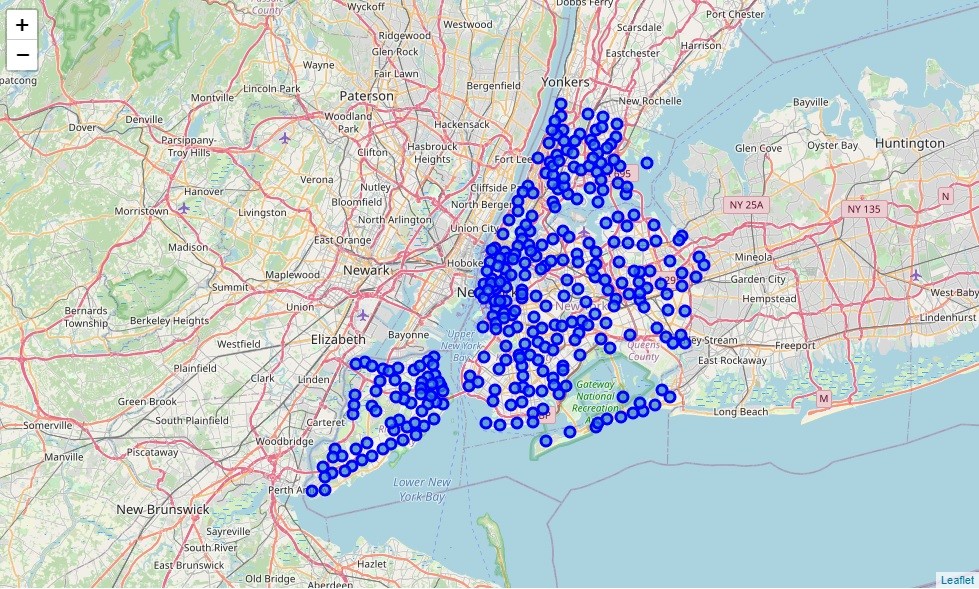
A merge of the two DataFrame gives us our population DataFrame

****

**Population data of New York City**

Folium map: this is a Library use to visualize interactive geographical plots

Map of the Neighborhoods in New York City

****

**Busy entertainment area - we use the Foursquare API**

**API CREDETIALS**

**CLIENT\_ID** = # your Foursquare ID

**CLIENT\_SECRET** = # your Foursquare Secret

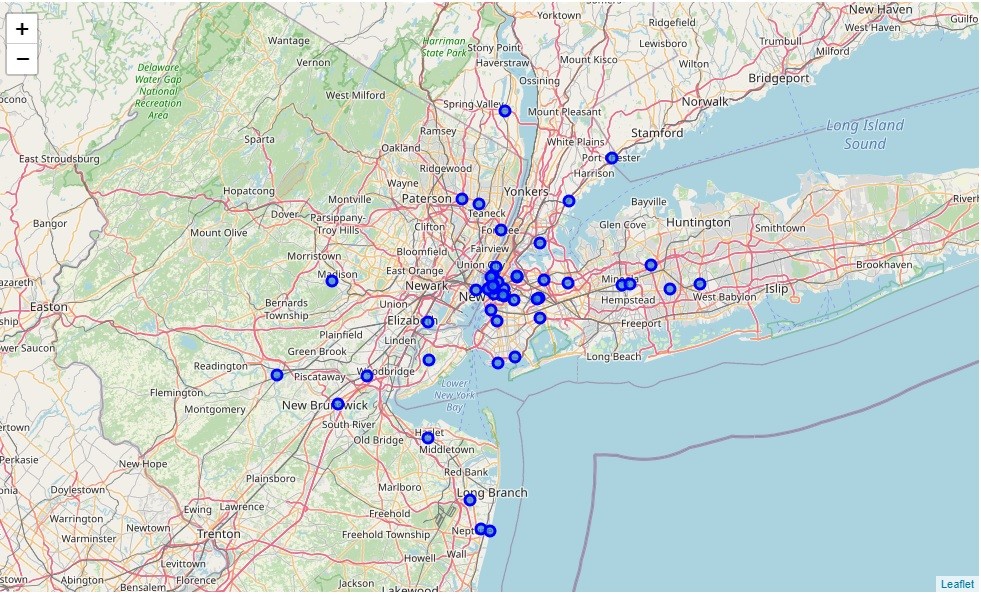
**VERSION** = # Foursquare API version

**LIMIT** = 100# A default Foursquare API limit value

From Foursquare API documentation, we can find the corresponding movie theater category in Venue Categories. The corresponding ID of Movie Theater in Foursquare API is 4bf58dd8d48988d17f941735 which is under Arts & Entertainment main category. It contains several sub-categories:

* Drive-in Theater, id: 56aa371be4b08b9a8d5734de
* Indie Movie Theater, id: 4bf58dd8d48988d17e941735
* Multiplex, id: 4bf58dd8d48988d180941735

From this we can use Folium to view a map of all the Movie Theatres in New York City

****

**Of the five Boroughs in New York, which is the richest(highest income)**

We use our population DataFrame will a bit of data preprocessing, we can determine that Manhattan is the richest Borough(*with respect to our data collected*)

**Movie theatres in Manhattan**

****

## **Methodology**

We conduct our project analysis using the following procedure:

1. Once the Borough is selected (MANHATTAN), for each neighborhood in Manhattan get top 100 venues within a 600 meter radius.
2. Now we have a DataFrame with the venues in each neighborhood, for each neighborhood venue we create a dummies value so as to later put it in the clustering algorithm.
3. We group the rows by the neighborhood so as to find the frequency of each venue type.
4. I need to get the neighborhoods where in the top 15 most common venues, there are Restaurants, shopping malls, things that should be around a movie theatre to make it successful (i.e. finding the busiest neighborhoods).
5. Cluster the neighborhoods and add the cluster label to each neighborhood and visualize the clusters.
6. Then examine each cluster and decide which cluster is the most popular, but has the least number of theatres. That cluster is where the movie theatre will be located. For a potential movie location use the clusters Centre.
7. Within your chosen cluster, decide the best neighborhood for a movie Theatre.

## **Analysis**

**PART 1**

* for each neighborhood in Manhattan get 100 venues within a 600 meter radius.

In the Foursquare API there is an explore tag. This allows use to take a location and use that location as a Centre point so as to get up to 100 different venues within a specified radius. We will use these venues to determine the best location to have a new movie theatre.

Manhattan venues DataFrame

****

**Part 2**

* for each neighborhood venue we create a dummies value so as to later put it in the clustering algorithm

A Dummies DataFrame is a DataFrame where we convert categorical data(*venues*) into binary form.1 represents the presence of the venue,0 represents the absence of a venue

****

**Part 3**

* We group the rows by the neighborhood so as to find the frequency of each venue type

The frequency is how often we see a particular venue point. The frequency helps us determine if a venue category is popular or not.

**Part 4 - The Goal**

* I need to get the neighborhoods where in the top 15 most common venues, there are Restaurants, shopping malls, things that should be around a movie theatre to make it successful (i.e. finding the busiest neighborhoods)

The resulting DataFrame

****

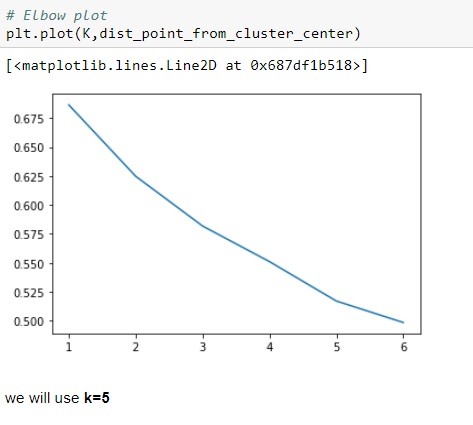
**Part 5**

* Cluster the neighborhoods and add the cluster label to each neighborhood and visualize the clusters.
* Elbow method to find best value for K

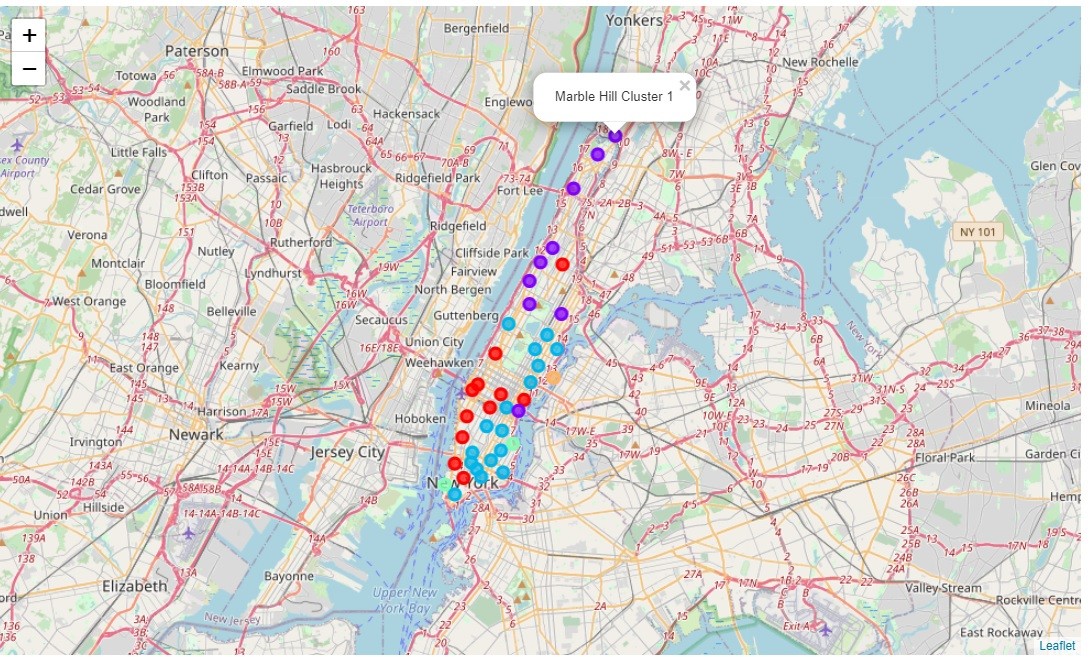
The Elbow method is a method that uses Sum of Squared Errors(SSE) to help us select the value K which represents the number of clusters we will use in our K means algorithm.

****

Plotting the results of the Elbow method

****

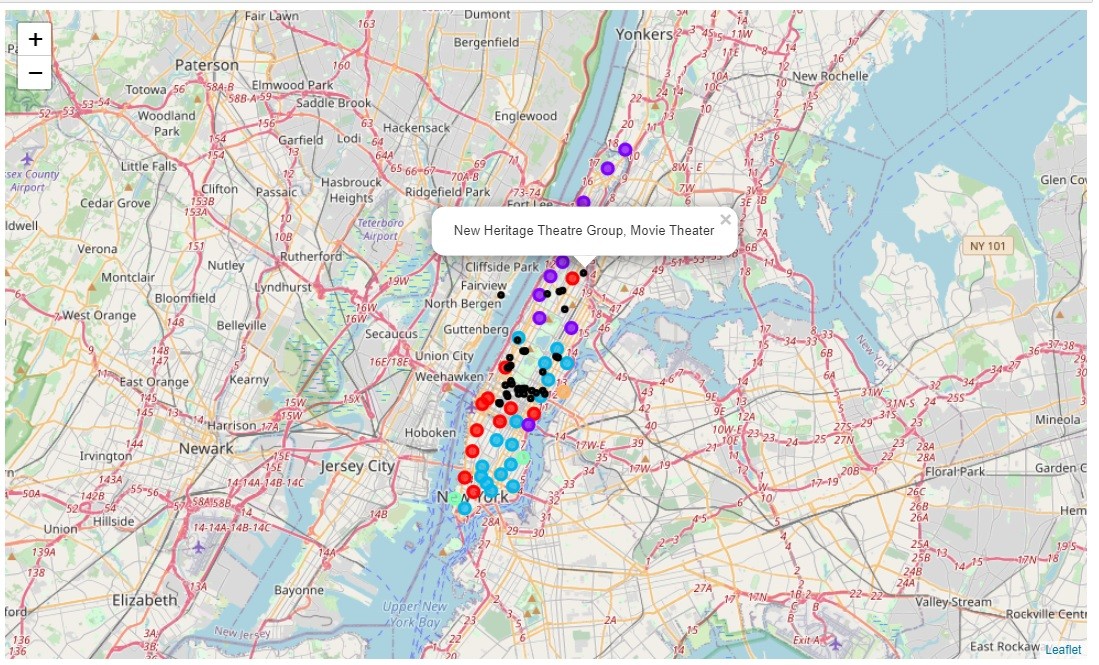
We then run the K Means Clustering Algorithm and we visualize the neighborhoods in each cluster using the Folium map below:

****

**Part 6**

* Display a map with the cluster and theatres on it to view which cluster has the least movie Theatres

The Theatres are represented with the black data point. The colorful points are the clusters

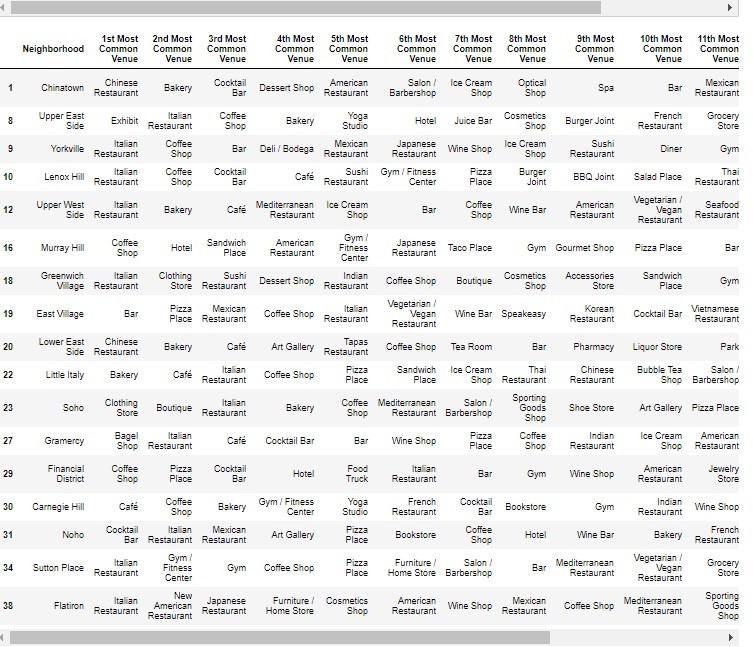
****

**Part 7**

* Within your chosen cluster, decide the best neighborhood for a movie Theatre

From our clustering results we choose the cluster label 2. Further explanation for this decision will be found in the results section.

The neighborhoods in cluster label 2 are as shown below

****

**Filtering neighborhoods by venue names**

Finally we use a Lambda expression(*anonymous function*) to return the neighborhoods with the most entertainment facilities, Yet without a movie theatre in it top 15 venues. The reason being that the ideal neighborhood should have restaurants, cafes, and other entertainment facilities hence making desirable to movie goers.

Below a DataFrame of the most ideal neighborhoods for a new Movie Theatre:

****

## **Results and Discussion**

* most of the movie theatres are located in Cluster label 0. So we eliminate this cluster since the shareholders would not want to open a theatre where there is massive competition.
* Cluster label 2: form pure observation seems the likely candidate because the are a few theatres in the cluster(low competition) but the cluster is very popular(has a lot of restaurants, shopping malls etc.).We determined its popularity by the number of neighborhoods in the cluster
* Cluster label 1 and Cluster label 3 and Cluster label 4: has too few neighborhoods to be an ideal location for a new theatre

**List of each neighborhoods unwanted venues in cluster\_label\_2**

* Lenox Hill - gym, sporting goods shop, Cosmetics
* Upper west side - gym
* Chelsea - art gallery, Theater, Park, Bookstore, cycle studio
* Tribeca - hotel, Park, spa, men's store
* Little Italy - Salon/Barbershop, Hotel, Optical shop
* Gramercy - gym, Art Gallery
* Carnegie - Bookstore, gym, gym/fitness
* Flatiron - Furniture/Home store, Cosmetics Shop, Sports goods shop, Salon/Barbershop, Toy/Game Store

From the cluster of Neighborhoods, we see that the **Upper west side** is the best neighborhood to build a theatre in because it has only one venue out of its 15 most popular venues, that does not have a quality we look for in buildings near a theatre.

## **Conclusion**

The purpose of this project was to determine a good location within the city of New York to open a new Movie Theatre. Stakeholders are interested in this location seeing as a Movie theatre is a prospect that generates a large income if located an ideal location.

We used the foursquare API to narrow it all the way down to the best Neighborhood where we could build a new Movie Theatre. The Final decision on the best Movie Theatre location will be made by stakeholders based on specific characteristics of neighborhoods.

## **References**

[1] Kaggle Data Set :

<https://www.kaggle.com/muonneutrino/new-york-city-census-data>

[2] New York Population Density: <https://en.wikipedia.org/wiki/Demographics_of_New_York_City>

Jupyter notebook

<https://github.com/victorbahlangene/Coursera_Capstone/blob/master/Capstone_Project.ipynb>