

# 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]

	Borough	Neighborhood	Latitude	Longitude	TotalPop	Employed	Income
0	Bronx	Wakefield	40.894705	-73.847201	7703	0	NaN
1	Bronx	Wakefield	40.894705	-73.847201	5403	2308	72034.0
2	Bronx	Wakefield	40.894705	-73.847201	5915	2675	74836.0
3	Bronx	Wakefield	40.894705	-73.847201	5879	2120	32312.0
4	Bronx	Wakefield	40.894705	-73.847201	2591	1083	37936.0

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

	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

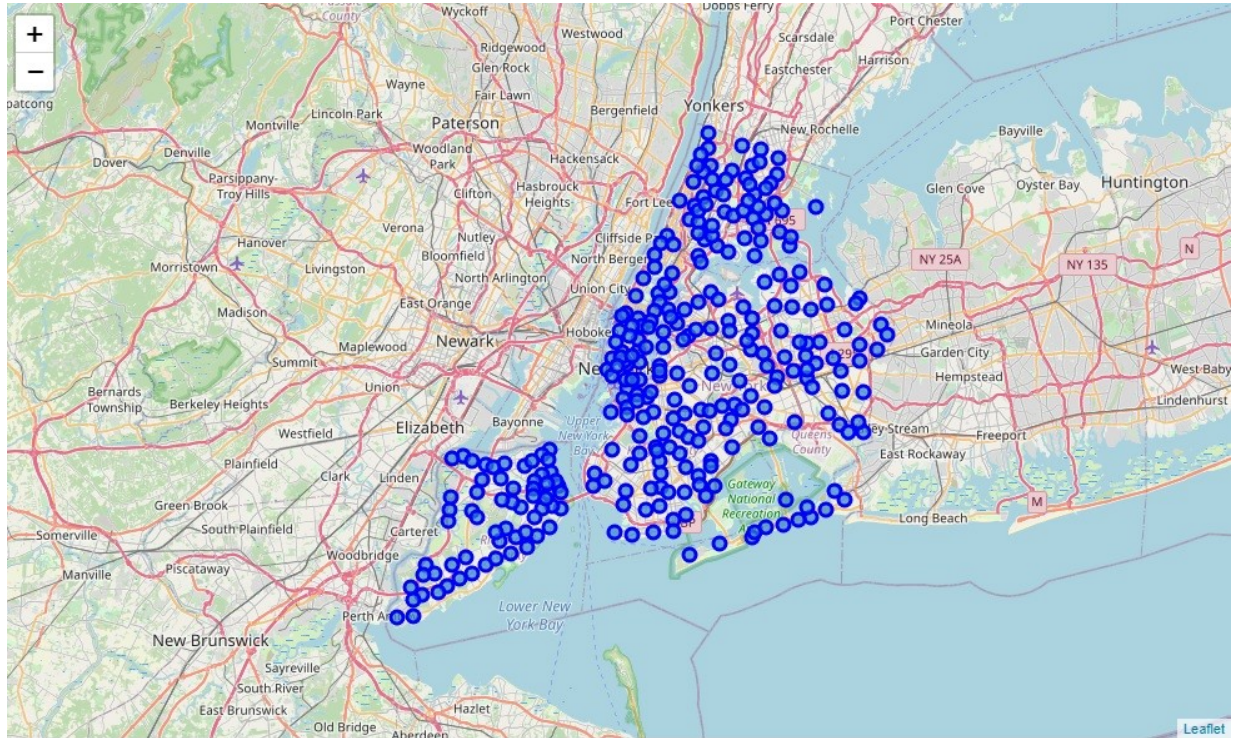
A merge of the two DataFrame gives us our population DataFrame

	Borough	Neighborhood	Latitude	Longitude	TotalPop	Employed	Income	Density
0	Bronx	Wakefield	40.894705	-73.847201	7703	0	NaN	13006
1	Bronx	Wakefield	40.894705	-73.847201	5403	2308	72034.0	13006
2	Bronx	Wakefield	40.894705	-73.847201	5915	2675	74836.0	13006
3	Bronx	Wakefield	40.894705	-73.847201	5879	2120	32312.0	13006
4	Bronx	Wakefield	40.894705	-73.847201	2591	1083	37936.0	13006

## 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

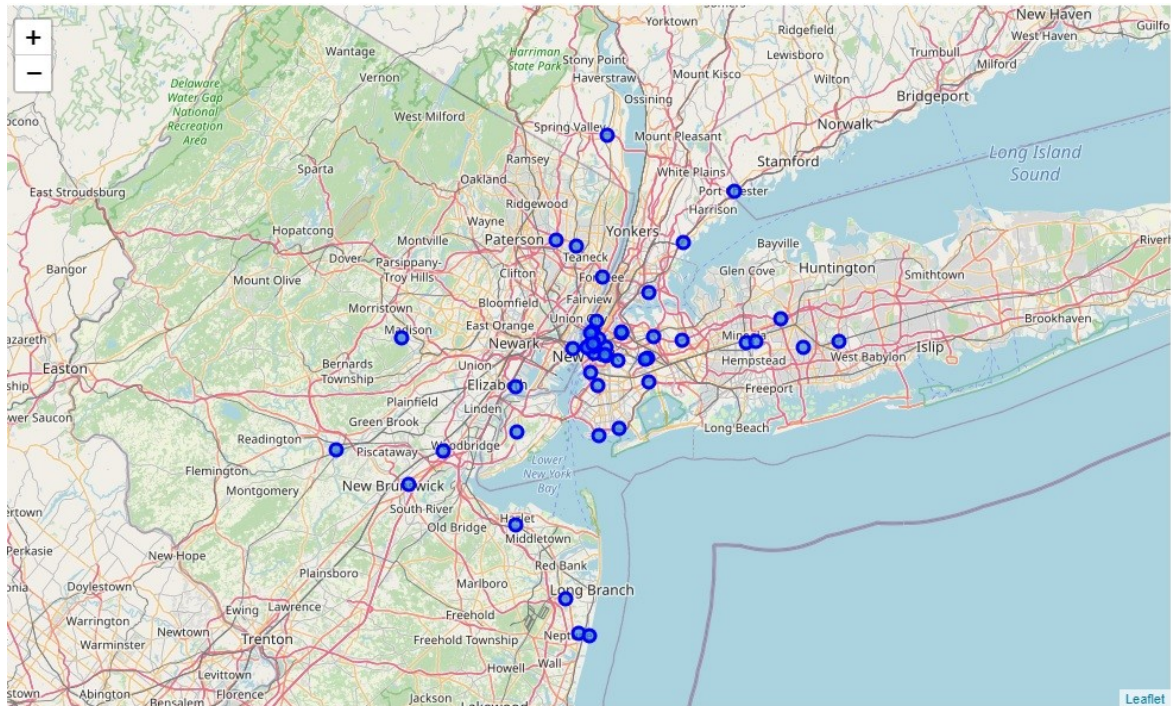
**LIMIT** = 100# A default Foursquare API limit value

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 with 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

```
manhattan_venues.head()
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Sam's Pizza	40.879435	-73.905859	Pizza Place
4	Marble Hill	40.876551	-73.91066	Astral Fitness & Wellness Center	40.876705	-73.906372	Gym

### 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

```
manhattan_dummies.head(3)
```

	Neighborhood	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Animal Shelter	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	...	Video Game Store	Vietnamese Restaurant	Volleyball Court	Waterfror
0	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0	
1	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0	
2	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0	

3 rows x 333 columns

### 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

```
neighborhoods_venues_sorted.head()
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue	12th Most Common Venue	13th Most Common Venue
0	Battery Park City	Park	Coffee Shop	Hotel	Gourmet Shop	Clothing Store	Pizza Place	Memorial Site	Plaza	Playground	Sandwich Place	Shopping Mall	Jewelry Store	
1	Carnegie Hill	Café	Coffee Shop	Bakery	Gym / Fitness Center	Yoga Studio	French Restaurant	Cocktail Bar	Bookstore	Gym	Indian Restaurant	Wine Shop	Pizza Place	Museum
2	Central Harlem	Juice Bar	Seafood Restaurant	Jazz Club	Bar	Park	Grocery Store	French Restaurant	Gym / Fitness Center	Chinese Restaurant	Art Gallery	Public Art	American Restaurant	Restaurant
3	Chelsea	Coffee Shop	Art Gallery	Bakery	American Restaurant	French Restaurant	Ice Cream Shop	Seafood Restaurant	Theater	Park	Bookstore	Café	Thai Restaurant	
4	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Dessert Shop	American Restaurant	Salon / Barbershop	Ice Cream Shop	Optical Shop	Spa	Bar	Mexican Restaurant	Dumpling Restaurant	Fun

## 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.

**Elbow Method**

```
manhattan_grouped.head()
```

	Neighborhood	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Animal Shelter	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	...	Video Game Store	Vietnamese Restaurant	Volleyball Court	Waterfront
0	Battery Park City	0.0	0.0	0.000000	0.010000	0.0	0.0	0.00	0.000000	0.00	...	0.0	0.00	0.0	
1	Carnegie Hill	0.0	0.0	0.000000	0.020000	0.0	0.0	0.01	0.000000	0.02	...	0.0	0.02	0.0	
2	Central Harlem	0.0	0.0	0.035088	0.035088	0.0	0.0	0.00	0.035088	0.00	...	0.0	0.00	0.0	
3	Chelsea	0.0	0.0	0.000000	0.040000	0.0	0.0	0.00	0.060000	0.00	...	0.0	0.00	0.0	
4	Chinatown	0.0	0.0	0.000000	0.040000	0.0	0.0	0.00	0.000000	0.00	...	0.0	0.02	0.0	

5 rows × 333 columns

```
# Features
X = manhattan_grouped.drop('Neighborhood', 1)

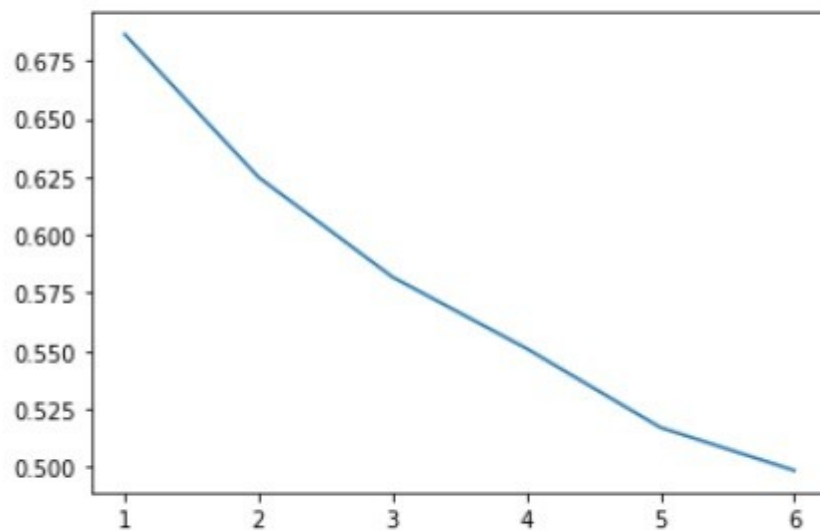
# finding k #
dist_point_from_cluster_center = []

K = range(1,7)
for no_in_clusters in K:
    k_model = KMeans(n_clusters=no_in_clusters)
    k_model.fit(X)
    dist_point_from_cluster_center.append(k_model.inertia_)
```

Plotting the results of the Elbow method

```
# Elbow plot  
plt.plot(K,dist_point_from_cluster_center)
```

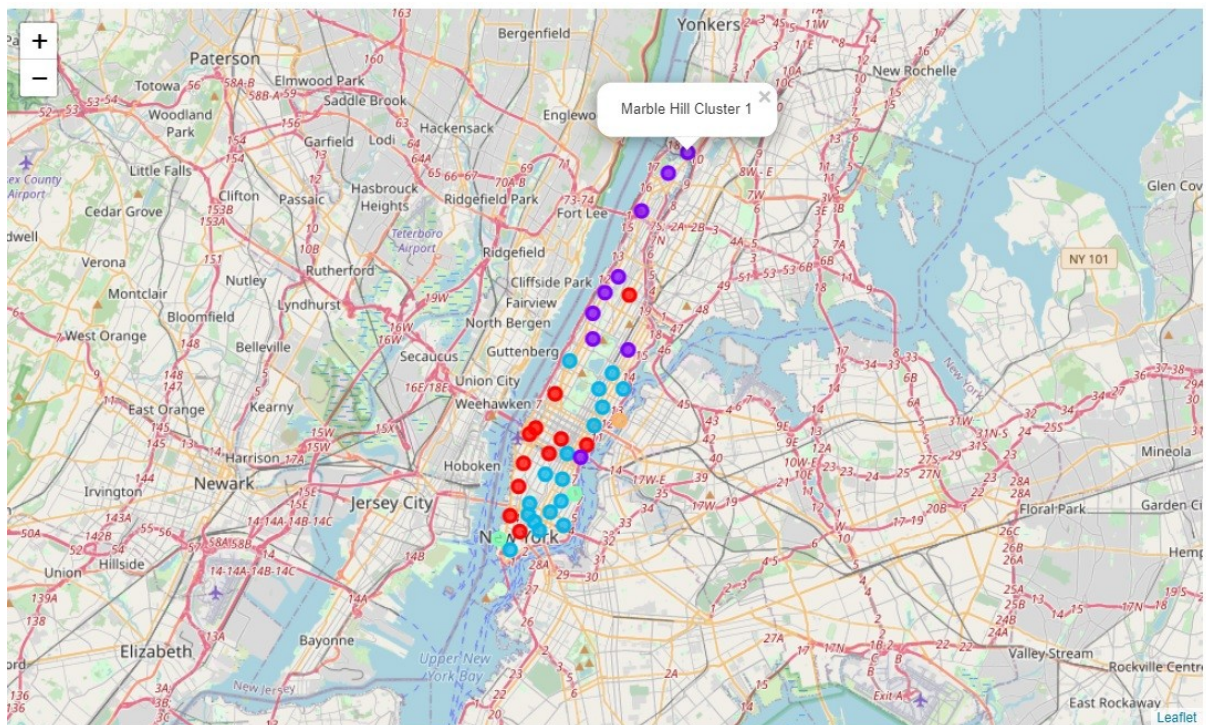
```
[<matplotlib.lines.Line2D at 0x687df1b518>]
```



we will use **k=5**



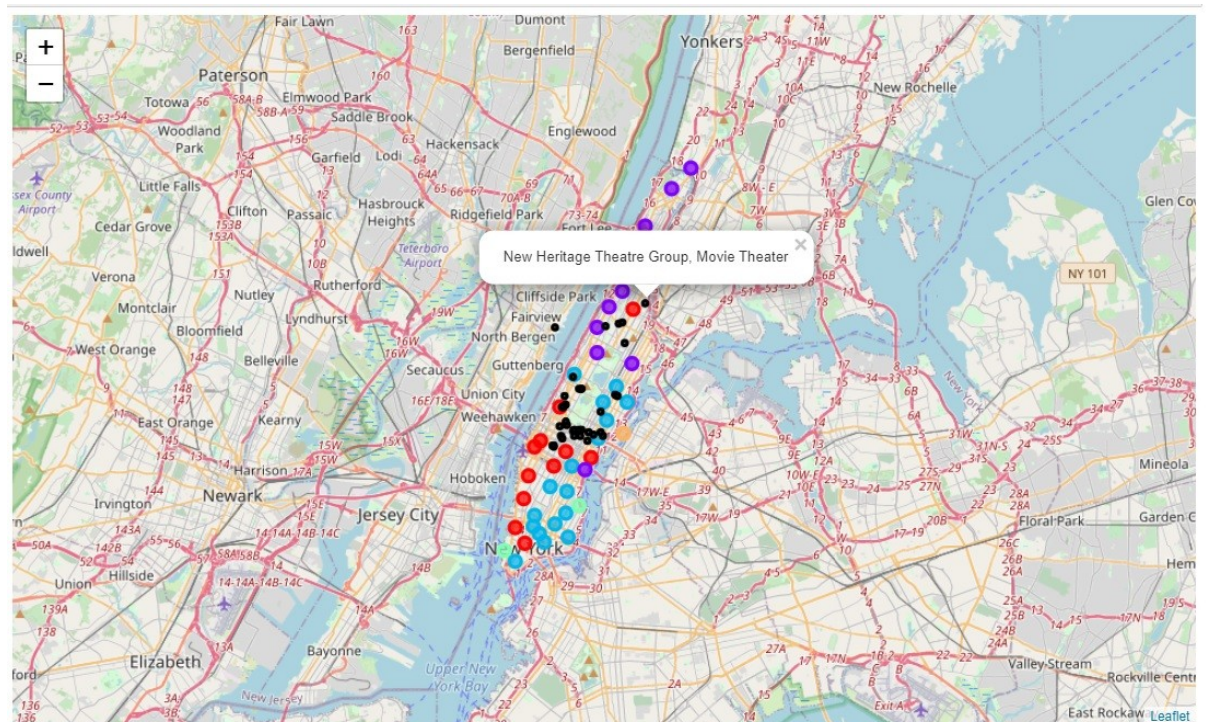
We then run the K Means Clustering Algorithm and we visualize the neighborhoods in each cluster using the Folium map below:



## 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

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue
1	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Dessert Shop	American Restaurant	Salon / Barbershop	Ice Cream Shop	Optical Shop	Spa	Bar	Mexican Restaurant
8	Upper East Side	Exhibit	Italian Restaurant	Coffee Shop	Bakery	Yoga Studio	Hotel	Juice Bar	Cosmetics Shop	Burger Joint	French Restaurant	Grocery Store
9	Yorkville	Italian Restaurant	Coffee Shop	Bar	Deli / Bodega	Mexican Restaurant	Japanese Restaurant	Wine Shop	Ice Cream Shop	Sushi Restaurant	Diner	Gym
10	Lenox Hill	Italian Restaurant	Coffee Shop	Cocktail Bar	Café	Sushi Restaurant	Gym / Fitness Center	Pizza Place	Burger Joint	BBQ Joint	Salad Place	Thai Restaurant
12	Upper West Side	Italian Restaurant	Bakery	Café	Mediterranean Restaurant	Ice Cream Shop	Bar	Coffee Shop	Wine Bar	American Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant
16	Murray Hill	Coffee Shop	Hotel	Sandwich Place	American Restaurant	Gym / Fitness Center	Japanese Restaurant	Taco Place	Gym	Gourmet Shop	Pizza Place	Bar
18	Greenwich Village	Italian Restaurant	Clothing Store	Sushi Restaurant	Dessert Shop	Indian Restaurant	Coffee Shop	Boutique	Cosmetics Shop	Accessories Store	Sandwich Place	Gym
19	East Village	Bar	Pizza Place	Mexican Restaurant	Coffee Shop	Italian Restaurant	Vegetarian / Vegan Restaurant	Wine Bar	Speakeasy	Korean Restaurant	Cocktail Bar	Vietnamese Restaurant
20	Lower East Side	Chinese Restaurant	Bakery	Café	Art Gallery	Tapas Restaurant	Coffee Shop	Tea Room	Bar	Pharmacy	Liquor Store	Park
22	Little Italy	Bakery	Café	Italian Restaurant	Coffee Shop	Pizza Place	Sandwich Place	Ice Cream Shop	Thai Restaurant	Chinese Restaurant	Bubble Tea Shop	Salon / Barbershop
23	Soho	Clothing Store	Boutique	Italian Restaurant	Bakery	Coffee Shop	Mediterranean Restaurant	Salon / Barbershop	Sporting Goods Shop	Shoe Store	Art Gallery	Pizza Place
27	Gramercy	Bagel Shop	Italian Restaurant	Café	Cocktail Bar	Bar	Wine Shop	Pizza Place	Coffee Shop	Indian Restaurant	Ice Cream Shop	American Restaurant
29	Financial District	Coffee Shop	Pizza Place	Cocktail Bar	Hotel	Food Truck	Italian Restaurant	Bar	Gym	Wine Shop	American Restaurant	Jewelry Store
30	Carnegie Hill	Café	Coffee Shop	Bakery	Gym / Fitness Center	Yoga Studio	French Restaurant	Cocktail Bar	Bookstore	Gym	Indian Restaurant	Wine Shop
31	Noho	Cocktail Bar	Italian Restaurant	Mexican Restaurant	Art Gallery	Pizza Place	Bookstore	Coffee Shop	Hotel	Wine Bar	Bakery	French Restaurant
34	Sutton Place	Italian Restaurant	Gym / Fitness Center	Gym	Coffee Shop	Pizza Place	Furniture / Home Store	Salon / Barbershop	Bar	Mediterranean Restaurant	Vegetarian / Vegan Restaurant	Grocery Store
38	Flatiron	Italian Restaurant	New American Restaurant	Japanese Restaurant	Furniture / Home Store	Cosmetics Shop	American Restaurant	Wine Shop	Mexican Restaurant	Coffee Shop	Mediterranean Restaurant	Sporting Goods Shop

## 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:

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue	12th Most Common Venue
Neighborhood												
Lenox Hill	Italian Restaurant	Coffee Shop	Cocktail Bar	Café	Sushi Restaurant	Gym / Fitness Center	Pizza Place	Burger Joint	BBQ Joint	Salad Place	Thai Restaurant	Taco Place
Upper West Side	Italian Restaurant	Bakery	Café	Mediterranean Restaurant	Ice Cream Shop	Bar	Coffee Shop	Wine Bar	American Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Put
Lower East Side	Chinese Restaurant	Bakery	Café	Art Gallery	Tapas Restaurant	Coffee Shop	Tea Room	Bar	Pharmacy	Liquor Store	Park	Cocktail Ba
Little Italy	Bakery	Café	Italian Restaurant	Coffee Shop	Pizza Place	Sandwich Place	Ice Cream Shop	Thai Restaurant	Chinese Restaurant	Bubble Tea Shop	Salon / Barbershop	Mediterranean Restaurant
Gramercy	Bagel Shop	Italian Restaurant	Café	Cocktail Bar	Bar	Wine Shop	Pizza Place	Coffee Shop	Indian Restaurant	Ice Cream Shop	American Restaurant	Restauran
Carnegie Hill	Café	Coffee Shop	Bakery	Gym / Fitness Center	Yoga Studio	French Restaurant	Cocktail Bar	Bookstore	Gym	Indian Restaurant	Wine Shop	Pizza Place
Flatiron	Italian Restaurant	New American Restaurant	Japanese Restaurant	Furniture / Home Store	Cosmetics Shop	American Restaurant	Wine Shop	Mexican Restaurant	Coffee Shop	Mediterranean Restaurant	Sporting Goods Shop	Spa

## 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 there 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](https://en.wikipedia.org/wiki/Demographics_of_New_York_City)

Jupyter notebook

[https://github.com/victorbahlangene/Coursera\\_Capstone/blob/master/Capstone\\_Project.ipynb](https://github.com/victorbahlangene/Coursera_Capstone/blob/master/Capstone_Project.ipynb)