

---

# SEGMENTATION AND RECOMMENDATION SYSTEM FOR MANHATTAN BOROUGH

*Battle of the neighbors*  
*ibm professional certificate*

*by*  
***Nelson ogbeide***

## INTRODUCTION

In this project, we have been instructed as data scientists by our stakeholder humtourist.co to come up with a brilliant idea to help make the experience of their client worthwhile during their stay in Manhattan, New York City. We came up with an idea to create a segmentation and recommendation system which recommends the best venues to visit according to collective ratings on each venue and neighborhood to visit based on some characteristics.

---

## DATA SECTION

Data used during this project was collected from different sources.

1. New York database [cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)
2. Foursquare

Data from the New York database was our base data, this data included features like

1. Borough (different borough in New York)
2. Neighborhood (existing neighborhood in each borough)
3. Latitude

---

#### 4. Longitude

	Borough	Neighbourhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Note! The project is specifically based on the manhattan borough.

With the help of the base data and Forusquare, we were able to get the most important data for the project .

For the segmentation task, the dataframe had features like

1. Neighborhood (existing neighborhood in each borough)
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue (Name of venue)
5. id (Unique id of venue)
6. Venue Latitude
7. Venue Longitude

Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	id	Venue Latitude	Venue Longitude	Venue Category
Marble Hill	40.876551	-73.91066	Arturo's	4b4429abf964a52037f225e3	40.874412	-73.910271	Pizza Place
Marble Hill	40.876551	-73.91066	Bikram Yoga	4baf59e8f964a520a6f93be3	40.876844	-73.906204	Yoga Studio
Marble Hill	40.876551	-73.91066	Tibbett Diner	4b79cc46f964a520c5122fe3	40.880404	-73.908937	Diner
Marble Hill	40.876551	-73.91066	Dunkin'	4b5357adf964a520319827e3	40.877136	-73.906666	Donut Shop
Marble Hill	40.876551	-73.91066	Starbucks	55f81cd2498ee903149fcc64	40.877531	-73.905582	Coffee Shop
Marble Hill	40.876551	-73.91066	Astral Fitness & Wellness Center	4cf6ae55d3a8a1cd71a9d243	40.876705	-73.906372	Gym
Marble Hill	40.876551	-73.91066	Blink Fitness	55f751ca498eacc0307d1cfe	40.877271	-73.905595	Gym
Marble Hill	40.876551	-73.91066	T.J. Maxx	546d31ca498e561c698a0320	40.877232	-73.905042	Department Store
Marble Hill	40.876551	-73.91066	Land & Sea Restaurant	4b9c9c6af964a520b27236e3	40.877885	-73.905873	Seafood Restaurant
Marble Hill	40.876551	-73.91066	TCR The Club of Riverdale	4a725fa1f964a520f6da1fe3	40.878628	-73.914568	Tennis Stadium

For the recommendation task, the “Rating” feature was added to the segmentation dataframe.

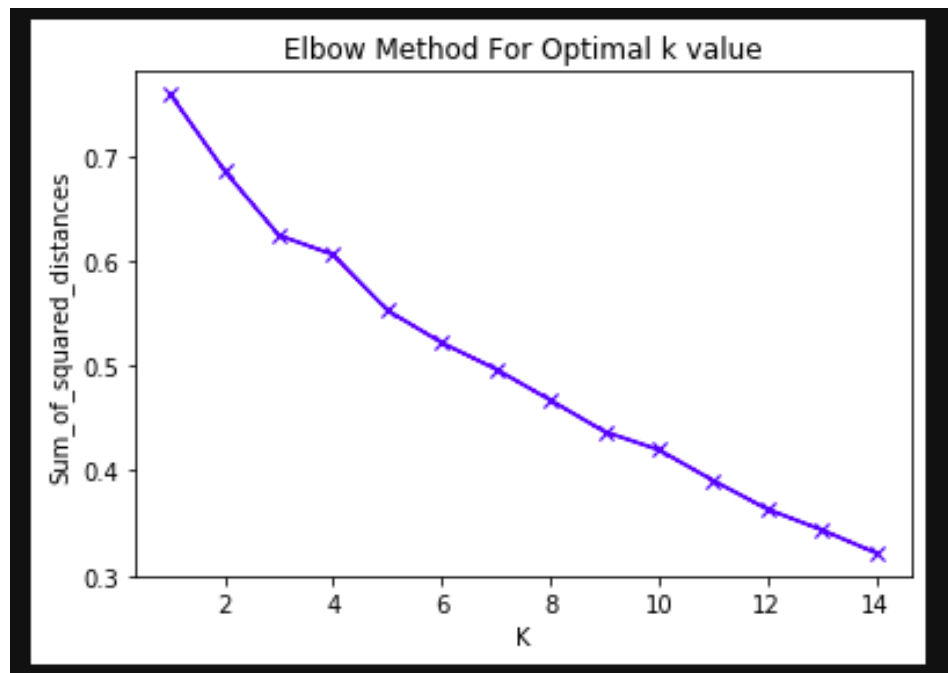
Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	id	Venue Latitude	Venue Longitude	Venue Category	ratings
Civic Center	40.715229	-74.005415	Potbelly Sandwich Shop	4f4d2593e4b00b42c3f1060f	40.714454	-74.005820	Sandwich Place	7.8
Civic Center	40.715229	-74.005415	The Mysterious Bookshop	4b6469d3f964a52037b12ae3	40.714910	-74.009480	Bookstore	8.7
Civic Center	40.715229	-74.005415	Chambers Street Wines	4adcf23cf964a520cc6221e3	40.715773	-74.009718	Wine Shop	9.1
Civic Center	40.715229	-74.005415	Smyth Hotel	49efcc88f964a52006691fe3	40.715144	-74.009183	Hotel	8.4
Civic Center	40.715229	-74.005415	SKY TING YOGA	582dbaabd21e031c1a35423d	40.718124	-74.003051	Yoga Studio	9.3

Note! For the recommendation task, we used less number of data due to the foursquare API restriction on premium API calls. Which is 50 per day. We worked with only 5 popular neighborhoods

---

## METHODOLOGY

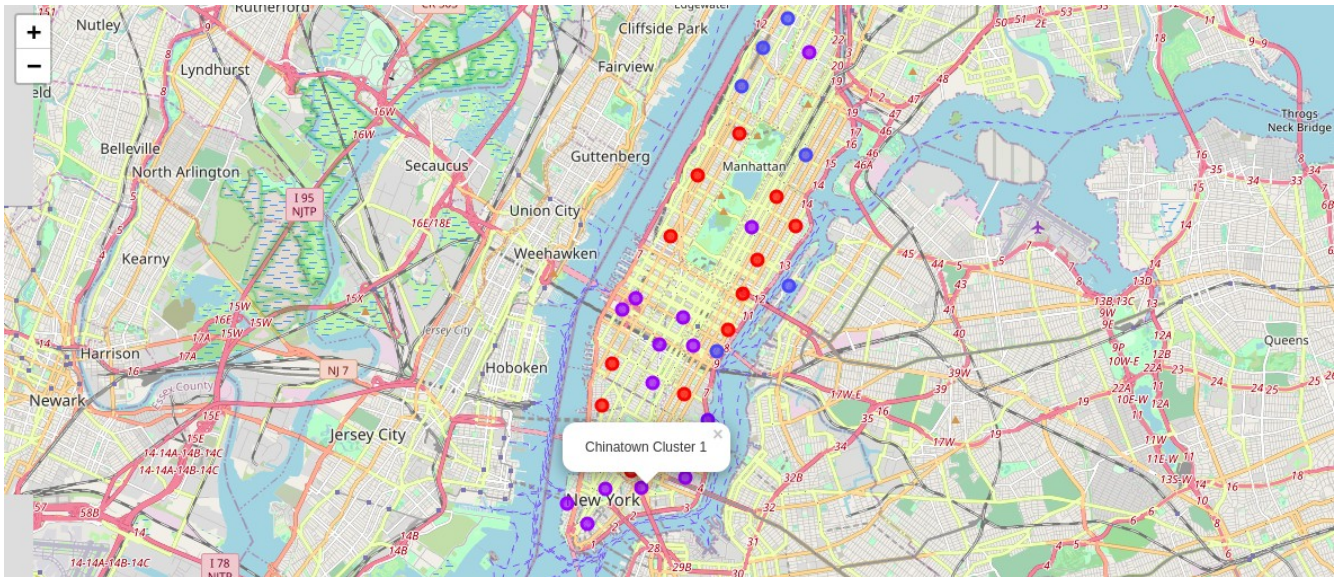
We utilized the Kmeans algorithm for segmentation and Content based filtering for Recommendation. The best K for clustering kmeans was determined using the elbow method.



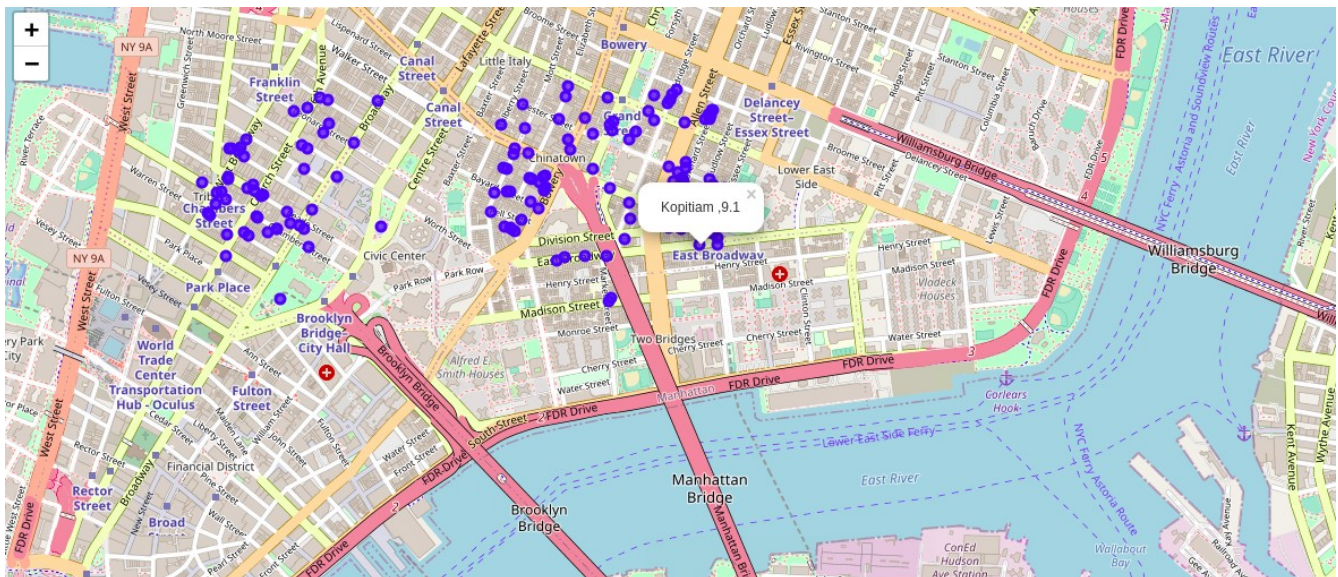
From the elbow method, we concluded on 5 as the best K for our kmeans algorithm, the various clusters were visualized on a map.

---





Ratings from venues for the recommender were also visualized for better understanding. The markers on the map includes the name of the venue and the rating allocated to them.



From our map, we can see that Kopitiam which falls under the Males Restaurant venue category has a 9.1 rating

---

## RESULT

The aim of this project was to identify the best places/venues to visit in manhattan in order to give tourist the best experience while in manhattan without them wasting their time searching and asking for best spots in the city. We achieved this firstly by collecting necessary data and leveraging the foursquare API in searching for venues around manhattan ranging from restaurants to bar, spa, parks, malls etc.

Next we asked ourselves how we can recommend the best venues. We ended up providing both a segmentation and recommendation system . Segmentation system to segment neighborhood with similar properties and recommendation system to recommend venues based on ratings. This produced wonderful results and insights to provide tourist with the best experience.

Our system recommends the best neighborhood and venues to visit when clients of hamtourits are in manhattan also bringing to their notice , the best neighborhood for different activities with the help of our segmentation system. Our analysis shows that majority of the venues in manhattan are restaurants, and majority of venues in neighbohood with large number of venues are also restaurants. Also from our analysis, the coffee shop and Chinese restaurant are the most recommended venue to visit, we also found out that chinatown was highly influential in the overall ratings of the Chinese restaurants and the civic center was also highly influential in the overall ratings of the coffee shop.

For clients visiting manhattan, These are the top recommendations according to venues, "Coffee Shop,Chinese Restaurant, Cocktail, Bar Bakery, American Restaurant, Mexican Restaurant and Sandwich Place. They should be also visit civic center if they are looking to get the best coffee, chinatown for restaurants(not just including Chinese restaurant), Stuyvesant Town if they would love a relaxation and recreational center and marbell hill if they are looking out for a wide variety of activities.