

Capstone Project - The Battle of Neighborhoods Report

INTRODUCTION

1.1 Background

ABC Company located in Manhattan, New York is planning to host a company lunch for its employees to celebrate its success in landing a huge business contract. For this purpose, it wants to choose a restaurant offering Italian cuisine in Manhattan, New York that is known for excellent service and serving great food.

1.2 Problem

Find an Italian restaurant in Manhattan, New York with highest number of likes by customers so that its employees will really have an excellent dining experience and awesome food.

1.3 Interest

Anyone interested in finding any popular category of venue in any geographic location may be interested in this project by modifying search criteria.

DATA

2.1 Data sources

Dataset containing geographical coordinates of New York neighborhoods is retrieved from scrapping web https://geo.nyu.edu/catalog/nyu_2451_34572. Transformed this data into Pandas dataframe and sliced it to contain only Manhattan borough of New York.

Used Foursquare API to get the venues using the search criteria 'Italian' for the Manhattan latitude and longitude and transformed it into Pandas dataframe. As the data returned by API also contained venues other than Italian restaurant, cleaned the dataset removing those venues.

Afterwards, used again Foursquare API to retrieve venue details relating to 'likes' 'count' for each venue and added this data point as 'likes count' column to venues dataset. The 'likes count' used for determining the top 3 restaurant choices and offering preferred choice to solve the problem

Foursquare API was used again to retrieve venue details relating to 'rating', though rating would not be available for all venues to verify the venues ranked high based on likes count would also be rated high

Recommendation of top three venues choices would be based on twin criteria: high likes count and high rating to solve the business problem

METHODOLOGY

Used pandas library to hold data, clean, explore and analyze dataset
Used numpy library to handle data in a vectorized manner

Used geopy, folium libraries to create maps of New York and Manhattan with neighborhoods and restaurant venues superimposed on top

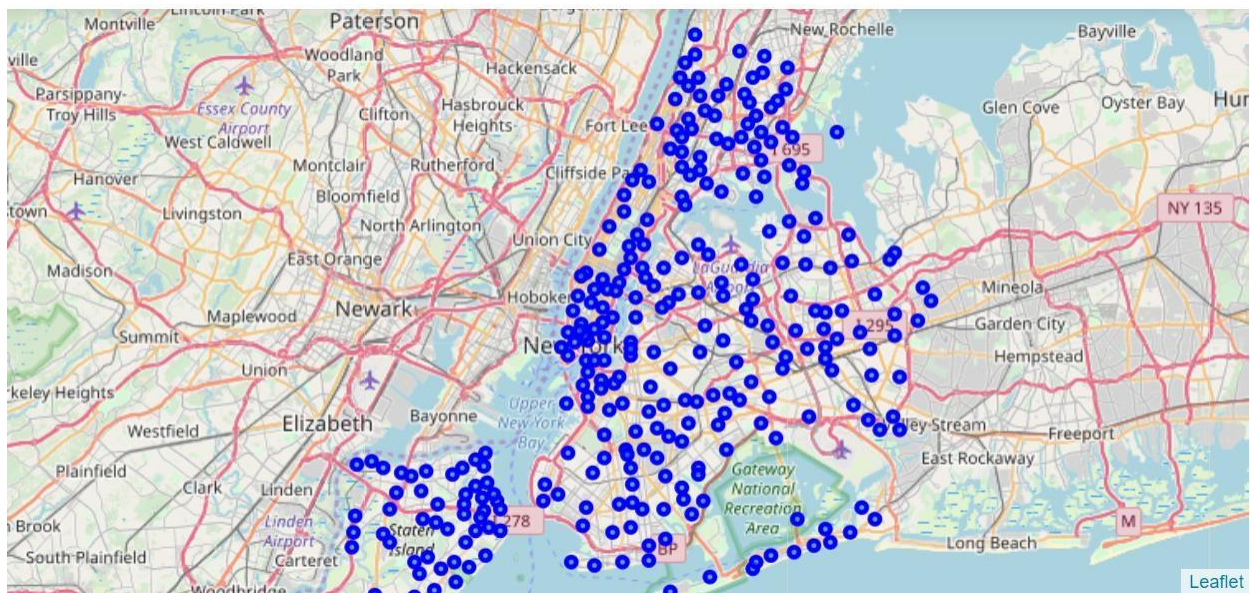
Used Foursquare API to search for venues categorized as Italian Restaurant and explore venue details relating to 'likes' count

Run unsupervised learning K-means algorithm to cluster the venues

Used Foursquare API to get venue rating, though rating is not available for all the venues

Recommendation of top three venues choices would be based on twin criteria: high likes count and high rating.

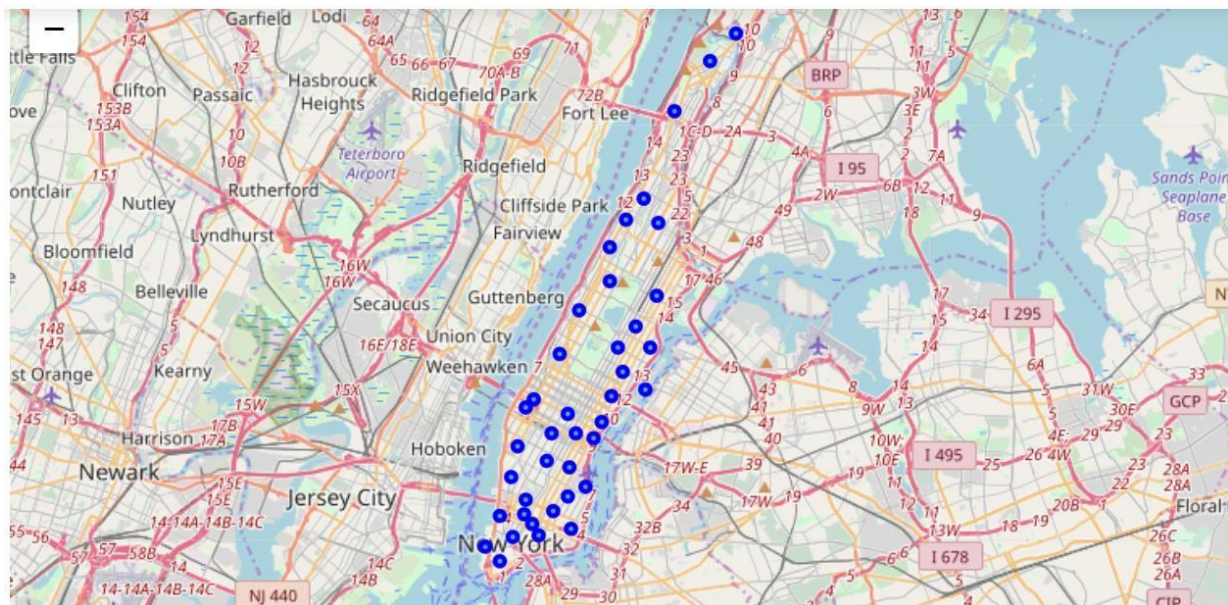
New York map with neighborhoods superimposed on top. New York has 5 boroughs and 306 neighborhoods.



Created Manhattan dataframe slicing New York dataset. First 10 rows of Manhattan neighborhoods dataset

| | Borough | Neighborhood | 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 |
| 5 | Manhattan | Manhattanville | 40.816934 | -73.957385 |
| 6 | Manhattan | Central Harlem | 40.815976 | -73.943211 |
| 7 | Manhattan | East Harlem | 40.792249 | -73.944182 |
| 8 | Manhattan | Upper East Side | 40.775639 | -73.960508 |
| 9 | Manhattan | Yorkville | 40.775930 | -73.947118 |

Manhattan has 40 neighborhoods. Created map of Manhattan with neighborhoods superimposed on top



Venues data set retrieved using search criteria 'Italian' for Manhattan geographic coordinates. 50 venues are returned by Foursquare API

Here're the first 10 rows of Venues dataset

| | id | name | categories | lat | lng |
|---|--------------------------|--|------------------------------|-----------|------------|
| 0 | 4a7778a1f964a5209be41fe3 | Carmine's Italian Restaurant | Italian Restaurant | 40.791096 | -73.973991 |
| 1 | 4a90dbbef964a520a11920e3 | Italian Village Pizzeria | Pizza Place | 40.772669 | -73.952319 |
| 2 | 4ba00318f964a520285237e3 | The Italian Academy (Casa Italiana) | General College & University | 40.807645 | -73.960396 |
| 3 | 51e7310c498e639ed27062b1 | Quality Italian | Italian Restaurant | 40.764476 | -73.977032 |
| 4 | 3fd66200f964a5209ee81ee3 | Carmine's Italian Restaurant | Italian Restaurant | 40.757497 | -73.986779 |
| 5 | 4b786cecf964a52052cd2ee3 | Bellini Italian Restaurant & Brick Oven Pizza | Italian Restaurant | 40.784656 | -73.973522 |
| 6 | 4f32883519836c91c7e0fe4f | Italian Eatery | Italian Restaurant | 40.801706 | -73.961285 |
| 7 | 4e7a3a40775b9ed4d3844232 | Italian Cousin | Pizza Place | 40.803357 | -73.963440 |
| 8 | 556f1642498e3a292961c735 | Italian lessons with Bretta Bracali Lundell | Language School | 40.785728 | -73.978029 |
| 9 | 4d7024805b5ca1cdaadd7644 | Italian Academy Theater, Columbia University 1161 Amsterdam Ave. | College Theater | 40.807664 | -73.959910 |

As we notice the venues dataset created from the results returned by Foursquare API contained venues other than restaurants, dataset is cleaned. Now, the venues dataset has 20 rows

First 10 rows of cleaned venues dataset

| | id | name | categories | lat | lng |
|---|--------------------------|---|--------------------|-----------|------------|
| 0 | 4a7778a1f964a5209be41fe3 | Carmine's Italian Restaurant | Italian Restaurant | 40.791096 | -73.973991 |
| 1 | 51e7310c498e639ed27062b1 | Quality Italian | Italian Restaurant | 40.764476 | -73.977032 |
| 2 | 3fd66200f964a5209ee81ee3 | Carmine's Italian Restaurant | Italian Restaurant | 40.757497 | -73.986779 |
| 3 | 4b786cecf964a52052cd2ee3 | Bellini Italian Restaurant & Brick Oven Pizza | Italian Restaurant | 40.784656 | -73.973522 |
| 4 | 4f32883519836c91c7e0fe4f | Italian Eatery | Italian Restaurant | 40.801706 | -73.961285 |
| 5 | 4e4c4d06bd413c4cc66856b9 | Italian Caterers | Italian Restaurant | 40.805768 | -73.966000 |
| 6 | 4f44a2bd19836ed0019571be | Mama Mia 44sw Italian Cuisine | Italian Restaurant | 40.759716 | -73.991829 |
| 7 | 4c8174d42f1c236a66213043 | Italian Tomato | Italian Restaurant | 40.819351 | -73.960352 |
| 8 | 4c5742f5d12a20a1545766bd | Italian Tomato | Italian Restaurant | 40.816234 | -73.979814 |
| 9 | 5a457b8d3b83076b58991a87 | The Italian Brothers | Italian Restaurant | 40.759424 | -73.969038 |

Afterwards, used Foursquare API again to retrieve venue details relating to 'likes' 'count' for each venue and added this data point as 'likes count' column to venues dataset

First 5 rows of the venues dataset after adding the 'likes count'

| | id | name | categories | lat | lng | likes count |
|---|--------------------------|---|--------------------|-----------|------------|-------------|
| 0 | 4a7778a1f964a5209be41fe3 | Carmine's Italian Restaurant | Italian Restaurant | 40.791096 | -73.973991 | 385 |
| 1 | 51e7310c498e639ed27062b1 | Quality Italian | Italian Restaurant | 40.764476 | -73.977032 | 678 |
| 2 | 3fd66200f964a5209ee81ee3 | Carmine's Italian Restaurant | Italian Restaurant | 40.757497 | -73.986779 | 1569 |
| 3 | 4b786cecf964a52052cd2ee3 | Bellini Italian Restaurant & Brick Oven Pizza | Italian Restaurant | 40.784656 | -73.973522 | 32 |
| 4 | 4f32883519836c91c7e0fe4f | Italian Eatery | Italian Restaurant | 40.801706 | -73.961285 | 0 |

Categorized venues based on 'likes count' by adding computed 'grade' column to venues dataset.

Likes Count Grade

| | |
|-----------|---------|
| <=50 | poor |
| 51 - 100 | average |
| 101 – 200 | good |
| 201 – 300 | great |
| >300 | awesome |

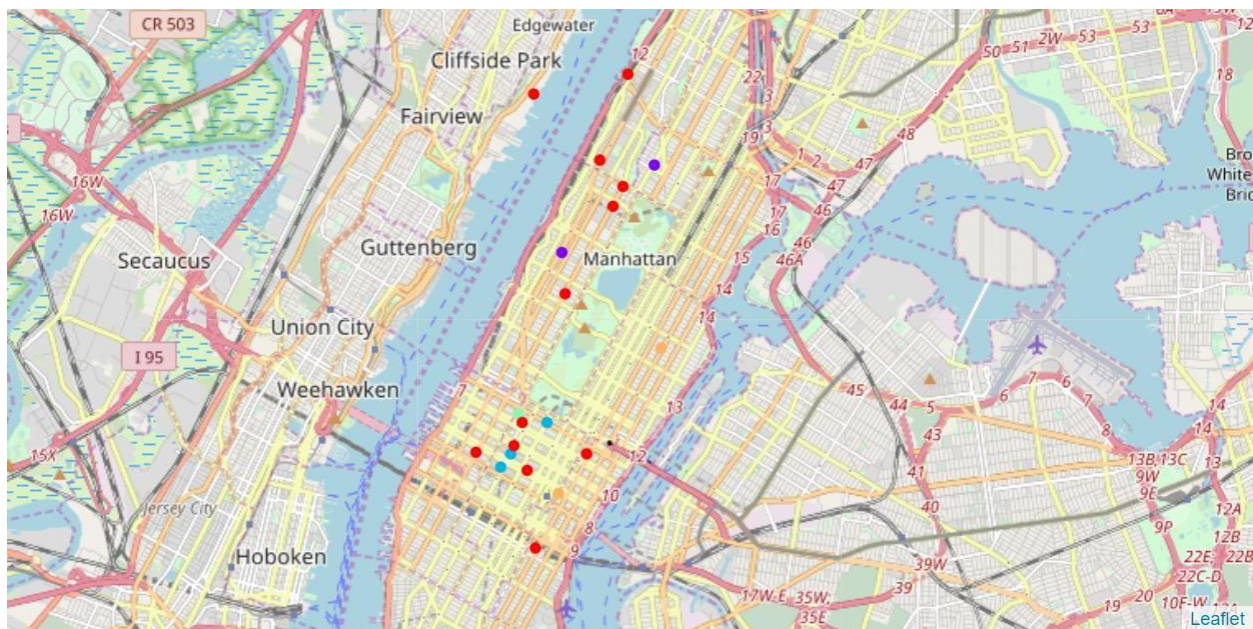
First 8 rows of venues dataframe after adding the grade column.

| | id | name | categories | lat | lng | likes count | cluster | grade |
|---|--------------------------|---|--------------------|-----------|------------|-------------|---------|---------|
| 0 | 4a7778a1f964a5209be41fe3 | Carmine's Italian Restaurant | Italian Restaurant | 40.791096 | -73.973991 | 385 | 3 | great |
| 1 | 51e7310c498e639ed27062b1 | Quality Italian | Italian Restaurant | 40.764476 | -73.977032 | 678 | 2 | awesome |
| 2 | 3fd66200f964a5209ee81ee3 | Carmine's Italian Restaurant | Italian Restaurant | 40.757497 | -73.986779 | 1569 | 2 | awesome |
| 3 | 4b786cecf964a52052cd2ee3 | Bellini Italian Restaurant & Brick Oven Pizza | Italian Restaurant | 40.784656 | -73.973522 | 32 | 1 | poor |
| 4 | 4f32883519836c91c7e0fe4f | Italian Eatery | Italian Restaurant | 40.801706 | -73.961285 | 0 | 1 | poor |
| 5 | 4e4c4d06bd413c4cc66856b9 | Italian Caterers | Italian Restaurant | 40.805768 | -73.966000 | 0 | 1 | poor |
| 6 | 4f44a2bd19836ed0019571be | Mama Mia 44sw Italian Cuisine | Italian Restaurant | 40.759716 | -73.991829 | 17 | 1 | poor |
| 7 | 4c8174d42f1c236a66213043 | Italian Tomato | Italian Restaurant | 40.819351 | -73.960352 | 0 | 1 | poor |

Run unsupervised learning K-means algorithm to cluster the venues. Here're the first 8 rows of venues dataset with cluster label

| | id | name | categories | lat | lng | likes count | cluster | grade |
|---|--------------------------|---|--------------------|-----------|------------|-------------|---------|---------|
| 0 | 4a7778a1f964a5209be41fe3 | Carmine's Italian Restaurant | Italian Restaurant | 40.791096 | -73.973991 | 385 | 3 | great |
| 1 | 51e7310c498e639ed27062b1 | Quality Italian | Italian Restaurant | 40.764476 | -73.977032 | 678 | 2 | awesome |
| 2 | 3fd66200f964a5209ee81ee3 | Carmine's Italian Restaurant | Italian Restaurant | 40.757497 | -73.986779 | 1569 | 2 | awesome |
| 3 | 4b786cecf964a52052cd2ee3 | Bellini Italian Restaurant & Brick Oven Pizza | Italian Restaurant | 40.784656 | -73.973522 | 32 | 1 | poor |
| 4 | 4f32883519836c91c7e0fe4f | Italian Eatery | Italian Restaurant | 40.801706 | -73.961285 | 0 | 1 | poor |
| 5 | 4e4c4d06bd413c4cc66856b9 | Italian Caterers | Italian Restaurant | 40.805768 | -73.966000 | 0 | 1 | poor |
| 6 | 4f44a2bd19836ed0019571be | Mama Mia 44sw Italian Cuisine | Italian Restaurant | 40.759716 | -73.991829 | 17 | 1 | poor |
| 7 | 4c8174d42f1c236a66213043 | Italian Tomato | Italian Restaurant | 40.819351 | -73.960352 | 0 | 1 | poor |

Visualize Clustering of venues



Examination of each cluster and find the discriminating likes count and grade that distinguish each cluster.

Cluster 0

| | id | name | categories | lat | lng | likes count | cluster | grade |
|----|--------------------------|----------------------------------|--------------------|-----------|------------|-------------|---------|---------|
| 12 | 3fd66200f964a520a2e91ee3 | Erminia Ristorante | Italian Restaurant | 40.775993 | -73.953590 | 57 | 0 | average |
| 13 | 525d7bde498e3a406f590b1d | Davio's North Italian Steakhouse | Italian Restaurant | 40.753233 | -73.974774 | 57 | 0 | average |

Cluster 1 – First 8 rows

| | id | name | categories | lat | lng | likes count | cluster | grade |
|----|--------------------------|---|--------------------|-----------|------------|-------------|---------|-------|
| 3 | 4b786cecf964a52052cd2ee3 | Bellini Italian Restaurant & Brick Oven Pizza | Italian Restaurant | 40.784656 | -73.973522 | 32 | 1 | poor |
| 4 | 4f32883519836c91c7e0fe4f | Italian Eatery | Italian Restaurant | 40.801706 | -73.961285 | 0 | 1 | poor |
| 5 | 4e4c4d06bd413c4cc66856b9 | Italian Caterers | Italian Restaurant | 40.805768 | -73.966000 | 0 | 1 | poor |
| 6 | 4f44a2bd19836ed0019571be | Mama Mia 44sw Italian Cuisine | Italian Restaurant | 40.759716 | -73.991829 | 17 | 1 | poor |
| 7 | 4c8174d42f1c236a66213043 | Italian Tomato | Italian Restaurant | 40.819351 | -73.960352 | 0 | 1 | poor |
| 8 | 4c5742f5d12a20a1545766bd | Italian Tomato | Italian Restaurant | 40.816234 | -73.979814 | 1 | 1 | poor |
| 9 | 5a457b8d3b83076b58991a87 | The Italian Brothers | Italian Restaurant | 40.759424 | -73.969038 | 0 | 1 | poor |
| 10 | 53d0596d498e713d024c6506 | Italian Bistro - Time Square | Italian Restaurant | 40.760656 | -73.984000 | 0 | 1 | poor |

Cluster 2 – Venues

| | id | name | categories | lat | lng | likes count | cluster | grade |
|----|--------------------------|------------------------------|--------------------|-----------|------------|-------------|---------|---------|
| 1 | 51e7310c498e639ed27062b1 | Quality Italian | Italian Restaurant | 40.764476 | -73.977032 | 678 | 2 | awesome |
| 2 | 3fd66200f964a5209ee81ee3 | Carmine's Italian Restaurant | Italian Restaurant | 40.757497 | -73.986779 | 1569 | 2 | awesome |
| 17 | 42f15a80f964a520bb261fe3 | Olive Garden | Italian Restaurant | 40.759414 | -73.984702 | 867 | 2 | awesome |

Cluster 3 – Venues

| | id | name | categories | lat | lng | likes count | cluster | grade |
|----|--------------------------|------------------------------|--------------------|-----------|------------|-------------|---------|-------|
| 0 | 4a7778a1f964a5209be41fe3 | Carmine's Italian Restaurant | Italian Restaurant | 40.791096 | -73.973991 | 385 | 3 | great |
| 11 | 4d573d8aa747b60cc83b332b | Lido Harlem | Italian Restaurant | 40.804987 | -73.954885 | 337 | 3 | great |

Cluster 4 – Venues

| | id | name | categories | lat | lng | likes count | cluster | grade |
|----|--------------------------|----------------------------|--------------------|-----------|------------|-------------|---------|-------|
| 18 | 4af0e7f3f964a52009e021e3 | Patsy's Italian Restaurant | Italian Restaurant | 40.765763 | -73.982816 | 129 | 4 | good |

**Sorted venues based on 'likes count' and created new dataframe with top 5 venues.
Here's the this dataframe.**

| | id | name | categories | lat | lng | likes count | cluster | grade |
|----|--------------------------|------------------------------|--------------------|-----------|------------|----------------|---------|---------|
| 2 | 3fd66200f964a5209ee81ee3 | Carmine's Italian Restaurant | Italian Restaurant | 40.757497 | -73.986779 | 1569 | 2 | awesome |
| 17 | 42f15a80f964a520bb261fe3 | Olive Garden | Italian Restaurant | 40.759414 | -73.984702 | 867 | 2 | awesome |
| 1 | 51e7310c498e639ed27062b1 | Quality Italian | Italian Restaurant | 40.764476 | -73.977032 | 678 | 2 | awesome |
| 0 | 4a7778a1f964a5209be41fe3 | Carmine's Italian Restaurant | Italian Restaurant | 40.791096 | -73.973991 | 385 | 3 | great |
| 11 | 4d573d8aa747b60cc83b332b | Lido Harlem | Italian Restaurant | 40.804987 | -73.954885 | 337 | 3 | great |

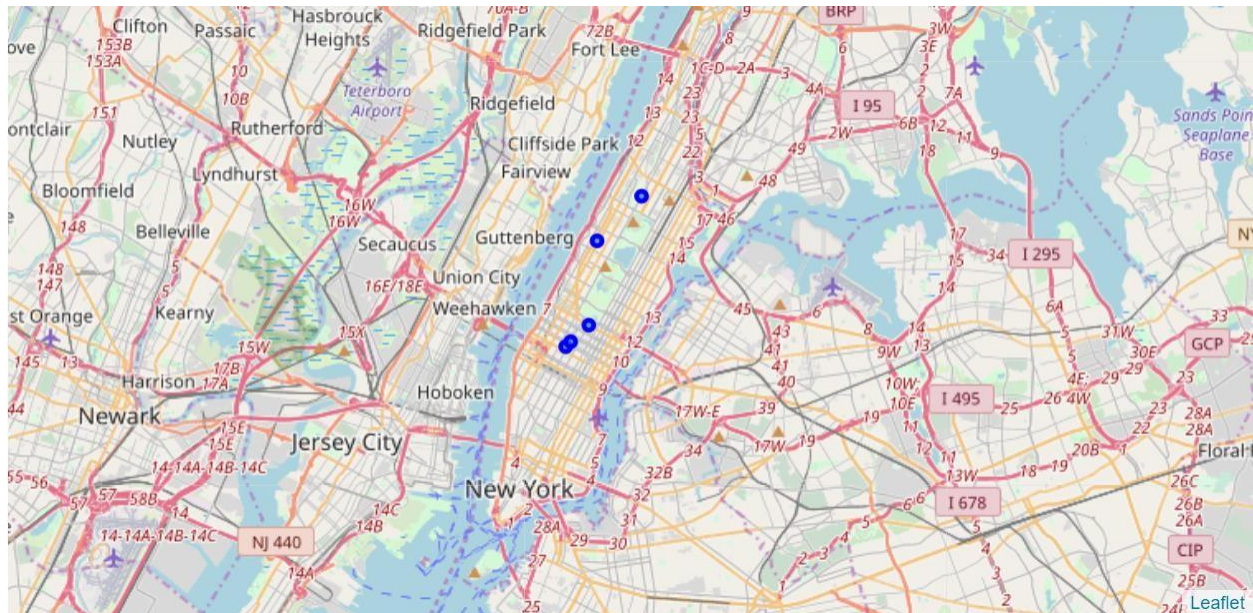
After adding 'rating' retrieved from Foursquare API to the venues dataset, top 5 venues of the dataframe

| | id | name | categories | lat | lng | likes count | cluster | grade | rating |
|----|--------------------------|------------------------------|--------------------|-----------|------------|----------------|---------|---------|--------|
| 2 | 3fd66200f964a5209ee81ee3 | Carmine's Italian Restaurant | Italian Restaurant | 40.757497 | -73.986779 | 1569 | 2 | awesome | 8.8 |
| 17 | 42f15a80f964a520bb261fe3 | Olive Garden | Italian Restaurant | 40.759414 | -73.984702 | 867 | 2 | awesome | 8.2 |
| 1 | 51e7310c498e639ed27062b1 | Quality Italian | Italian Restaurant | 40.764476 | -73.977032 | 678 | 2 | awesome | 8.8 |
| 0 | 4a7778a1f964a5209be41fe3 | Carmine's Italian Restaurant | Italian Restaurant | 40.791096 | -73.973991 | 385 | 3 | great | 8.3 |
| 11 | 4d573d8aa747b60cc83b332b | Lido Harlem | Italian Restaurant | 40.804987 | -73.954885 | 337 | 3 | great | 8.6 |

Created bar chart representing top 5 venues with their like count



Created Manhattan, NY map with top 5 venues superimposed on top



RESULTS

Cleaned dataset representing Italian restaurant venues have 20 rows. Categorizing venues based on 'like count' has resulted in the following break-down: 3 venues with 'awesome', 2 venues with 'great', 1 venue with 'good' and 2 venues with 'average' and the remaining 12 with 'poor' grades.

k-means clustering of venues has resulted in the following 5 clusters and grouping of venues by clustering matches with grade computed based on likes count:

Cluster label venues

- Cluster 0 2 venues with 'average' grade
- Cluster 1 12 venues with 'poor' grade
- Cluster 2 3 venues with 'awesome' grade
- Cluster 3 2 venues with 'great' grade
- Cluster 4 1 venue with 'good' grade

It is noticed that all the top 5 venues ranked by likes count got more than 8 rating and the top ranked venue got the highest rating of 8.8

DISCUSSION

Chosen the criteria of 'likes count' and 'rating' returned by Foursquare API for venues to group and decide which venues to be recommended to ABC Company for selecting the Italian restaurant. It is obvious that there may be other criteria based on other sources of data. This project scope is limited to using the data available through Foursquare API.

This project demonstrated how we can leverage the data freely available on web and Foursquare API(though free version has limits on data points we can retrieve) to search and explore restaurant venues. We can adapt it to choose any other venue by changing the search query. This project is therefore, useful for anyone interested in selecting any other category of venue.

CONCLUSION

Based on two factors, namely 'likes count' and 'rating', Carmine's Italian Restaurant, Olive Garden and Quality Italian are ranked as the top 3 venues and presented as possible choices for consideration of the ABC Company, with the top 1 venue Carmine's Italian Restaurant as the preferred choice. Anyone interested in finding any popular category of venue in any geographic location may also be interested in this project as we can adapt it choose any other venue by changing the search query.