A RECOMMENDER SYSTEM FOR ESTABLISHING A FITNESS PRODUCTS STORE

Capstone Project - The Battle of the Neighbourhoods (Week 2)

Applied Data Science Capstone by IBM/Coursera

Table of contents

Introduction: Business Problem

A commercial fitenss products company wants to open their stores across United States. However, it wants to open it's first store in New York city. The company wants to find a perfect location in New York to open it's first store. To start with positive wibes in expanding their business, the company wants to find a good location to open that store. If the company can open a store near most of the top-rated fitness centers, they can have good business. For that, the company wants to perform some analysis about the fitness centers located in New York and wants to make a decision based on the conclusions derived from the analysis.

In this project I will try to find an optimal location to open a store for the company. Specifically, this report will be targeted to stakeholders interested in opening a commercial fitness products store in New York, NY.

First, I will try to find the locations where most the fitness centers are located in New York, NY. Later, I will try to filter those locations based on the number of likes given by the customers. If both crieteia are matched, that location is considered as optimal location for open a store.

I will use my data science capabilities to generate a few most promissing neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

Following data sources will be needed to extract/generate the required information:

- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**

For this Capstone, the Foursquare API is utilized to pull the following location data on Fitness Centers:

- Venue Name
- Venue Location
- Venue Category
- Count of Likes

Data Acquisition Approach

To acquire the data mentioned above, steps involved in are:

Step-1: Get geolocator latitude and longitude coordinates of New York city

Step-2: Use Foursquare API to get a list of all venues in New York city

	name	id	categories	lat	lng	formattedAddress
0	The Bar Room at Temple Court	57f0689d498e7d49d9189369	Hotel Bar	40.711448	-74.006802	[123 Nassau St, New York, NY 10038, United Sta...
1	The Beekman - A Thompson Hotel	56d8c0f8498edb854f926e6a	Hotel	40.711173	-74.006702	[123 Nassau St (at Beekman St), New York, NY 1...
2	City Hall Park	3fd66200f964a520d8f11ee3	Park	40.712241	-74.006977	[17 Park Row (btwn Broadway & Centre St), New ...
3	Alba Dry Cleaner & Tailor	4c606c3e1e5cd13ad1a1a1ed	Laundry Service	40.711434	-74.006272	[140 Nassau St (Beekman and Spruce st), New Yo...
4	The Wooly Daily	56093809498e5344ab8835a6	Coffee Shop	40.712137	-74.008395	[11 Barclay St (Broadway), New York, NY 10007,...
5	Gibney Dance Center Downtown	53373f26498e940581c90985	Dance Studio	40.713923	-74.005661	[280 Broadway (Chambers St), New York, NY 1000...
6	Augustine	58191674ded8f8626ed70af0	French Restaurant	40.711310	-74.006660	[5 Beekman St, New York, NY 10038, United States]
7	The Class by Taryn Toomey	58a253c830ecc66c9e5b40a0	Gym / Fitness Center	40.712753	-74.008734	[22 Park Pl, New York, NY 10007, United States]
8	Takahachi Bakery	4c154c9a77cea593c401d260	Bakery	40.713653	-74.008804	[25 Murray St (at Church St), New York, NY 100...
9	Aahar Indian Cuisine	575dea4c498e2739e43a27e2	Indian Restaurant	40.713307	-74.007994	[10 Murray St (Broadway), New York, NY 10007, ...

Step-3: Get venue name, venue ID, location, category, and likes of all venues

	name	id	categories	lat	lng	formattedAddress	total likes
7	The Class by Taryn Toomey	58a253c830ecc66c9e5b40a0	Gym / Fitness Center	40.712753	-74.008734	[22 Park Pl, New York, NY 10007, United States]	17
11	CrossFit 212 TriBeCa	52001eed498e9ac16ca5e20b	Gym	40.714537	-74.005999	[281 Broadway, New York, NY 10007, United States]	14
15	Kula Yoga Project	4ad79413f964a520610c21e3	Yoga Studio	40.714342	-74.008094	[28 Warren St (Church St), New York, NY 10007,...	48
20	Church Street Boxing Gym	4b8dbba4f964a5201e0b33e3	Boxing Gym	40.713354	-74.009067	[25 Park Pl (btwn Church St. & Broadway), New ...	67
32	Equinox Tribeca	4a6e331af964a52031d41fe3	Gym	40.714099	-74.009686	[54 Murray Street (at W Broadway), New York, N...	246
35	Bikram Yoga Downtown	4b1c5a2ff964a520720624e3	Yoga Studio	40.710021	-74.007224	[121 Fulton St Fl 3 (Dutch), New York, NY 1003...	22
39	Heyday	57ad129c498e05b086594d72	Spa	40.715598	-74.007882	[92 Reade St, New York, NY 10013, United States]	28
44	Exceed Physical Culture	53910ac3498e57a5dc0eb160	Gym / Fitness Center	40.715629	-74.007992	[97 Reade St (bet W Broadway & Church St), New...	91
64	European Wax Center	5400b87a498e87c5ae6e5143	Health & Beauty Service	40.714710	-74.007448	[95 Chambers St, New York, NY 10007, United St...	7
73	Wall Street Bath & Spa 88	49ba7becf964a5206e531fe3	Spa	40.709278	-74.005690	[88 Fulton St, New York, NY 10038, United States]	70
98	New York by Gehry Gym	4e9b03e59a52edbd658ca490	Gym	40.710655	-74.005709	[8 Spruce St., New York, NY 10038, United States]	13

Methodology

Step-1: Get venue name, venue ID, location, category, and likes of all venues in New York city

Step-2: Group the venues into 2-3 clusters based on number of likes

Step-3: Draw the conclusions based on the characteristics of clustered data

Step-4: Make suggestions based on the conclusions drawn

Analysis

Step-1: Find the unique venues and extract the venues related to fitness

Step-2: Find the nuber of likes given to each fitness center and add that info to original data frame

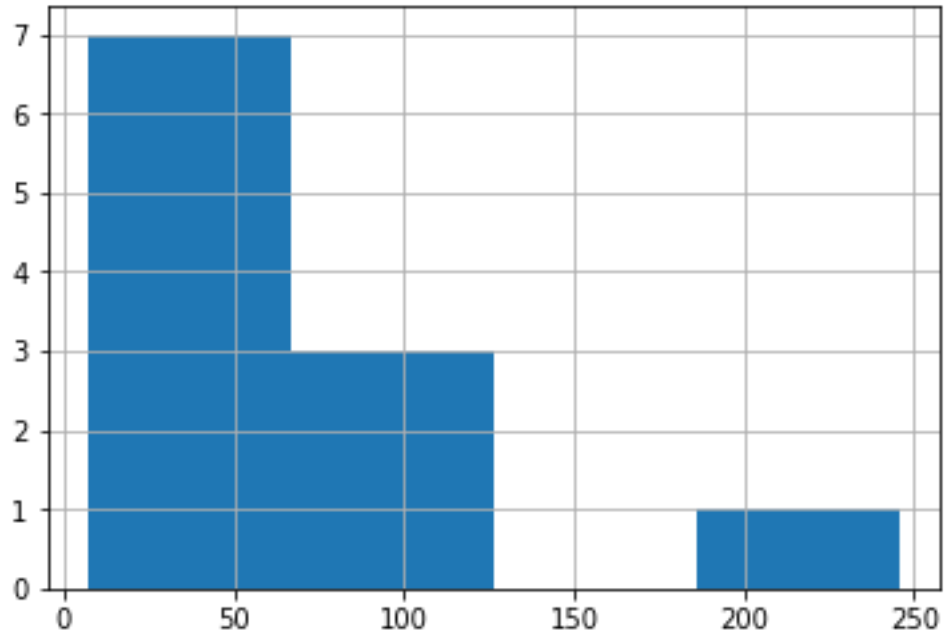
Step-3: Group the data into discrete groups based on number of likes

Maximum number of likes: 246

Minimum number of likes: 7

Median of number of likes: 28.0

Mean of number of likes: 56.63636363636363



25th percentile of the number of likes: 15.5

50th percentile of the number of likes: 28.0

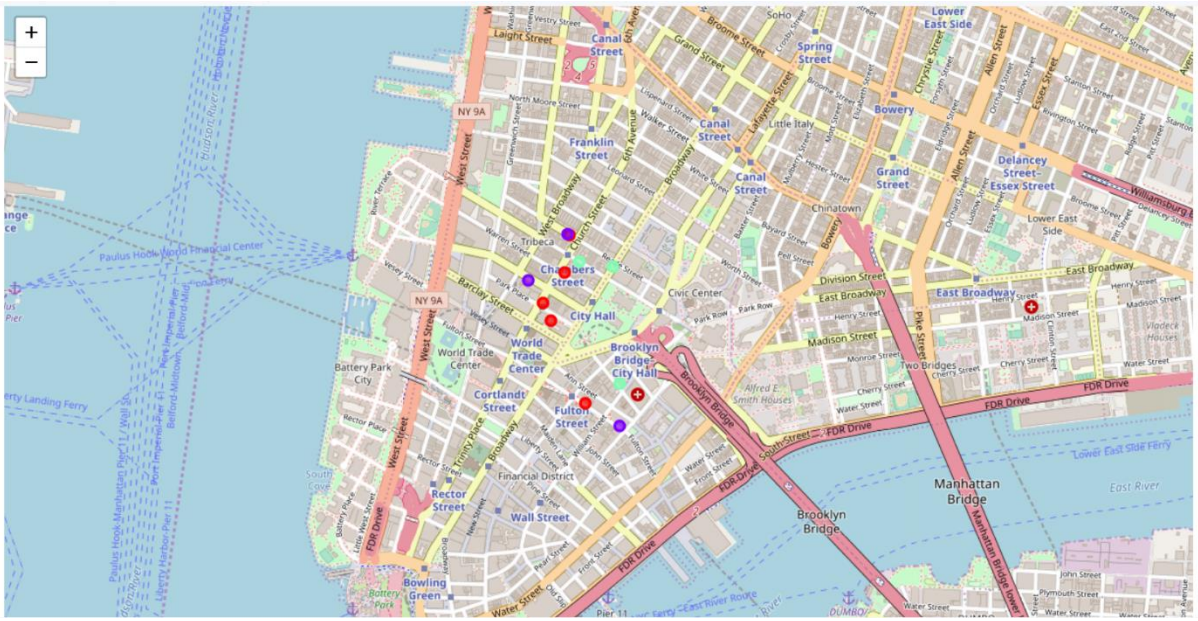
75th percentile of the number of likes: 68.5

Step-4: Create dummy variables for total likes and categories

Step-5: Perform Clustering

	name	id	categories	lat	lng	formattedAddress	total likes	total likes_cat	label
7	The Class by Taryn Toomey	58a253c830ecc66c9e5b40a0	Gym / Fitness Center	40.712753	-74.008734	[22 Park Pl, New York, NY 10007, United States]	17	below avg	0
11	CrossFit 212 TriBeCa	52001eed498e9ac16ca5e20b	Gym	40.714537	-74.005999	[281 Broadway, New York, NY 10007, United States]	14	poor	2
15	Kula Yoga Project	4ad79413f964a520610c21e3	Yoga Studio	40.714342	-74.008094	[28 Warren St (Church St), New York, NY 10007,...	48	above avg	0
20	Church Street Boxing Gym	4b8dbba4f964a5201e0b33e3	Boxing Gym	40.713354	-74.009067	[25 Park Pl (btwn Church St. & Broadway), New ...	67	above avg	0
32	Equinox Tribeca	4a6e331af964a52031d41fe3	Gym	40.714099	-74.009686	[54 Murray Street (at W Broadway), New York, N...	246	great	1
35	Bikram Yoga Downtown	4b1c5a2ff964a520720624e3	Yoga Studio	40.710021	-74.007224	[121 Fulton St Fl 3 (Dutch), New York, NY 1003...	22	below avg	0
39	Heyday	57ad129c498e05b086594d72	Spa	40.715598	-74.007882	[92 Reade St, New York, NY 10013, United States]	28	below avg	0
44	Exceed Physical Culture	53910ac3498e57a5dc0eb160	Gym / Fitness Center	40.715629	-74.007992	[97 Reade St (bet W Broadway & Church St), New...	91	great	1
64	European Wax Center	5400b87a498e87c5ae6e5143	Health & Beauty Service	40.714710	-74.007448	[95 Chambers St, New York, NY 10007, United St...	7	poor	2
73	Wall Street Bath & Spa 88	49ba7becf964a5206e531fe3	Spa	40.709278	-74.005690	[88 Fulton St, New York, NY 10038, United States]	70	great	1
98	New York by Gehry Gym	4e9b03e59a52edbd658ca490	Gym	40.710655	-74.005709	[8 Spruce St , New York, NY 10038, United States]	13	poor	2

Step-6: Visualize the Clustering results



Results and Discussions

Using the k-means algorithm, given data is divided into three clusters as follows.

Cluster 1

```
fitness_centers.loc[fitness_centers['label']==0]
```

2]:

	name	id	categories	lat	lng	formattedAddress	total likes	total likes_cat	label
7	The Class by Taryn Toomey	58a253c830ecc66c9e5b40a0	Gym / Fitness Center	40.712753	-74.008734	[22 Park Pl, New York, NY 10007, United States]	17	below avg	0
15	Kula Yoga Project	4ad79413f964a520610c21e3	Yoga Studio	40.714342	-74.008094	[28 Warren St (Church St), New York, NY 10007,...	48	above avg	0
20	Church Street Boxing Gym	4b8dbba4f964a5201e0b33e3	Boxing Gym	40.713354	-74.009067	[25 Park Pl (btwn Church St. & Broadway), New ...	67	above avg	0
35	Bikram Yoga Downtown	4b1c5a2ff964a520720624e3	Yoga Studio	40.710021	-74.007224	[121 Fulton St Fl 3 (Dutch), New York, NY 1003...	22	below avg	0
39	Heyday	57ad129c498e05b086594d72	Spa	40.715598	-74.007882	[92 Reade St, New York, NY 10013, United States]	28	below avg	0

Cluster 2

```
fitness_centers.loc[fitness_centers['label']==1]
```

3]:

	name	id	categories	lat	lng	formattedAddress	total likes	total likes_cat	label
32	Equinox Tribeca	4a6e331af964a52031d41fe3	Gym	40.714099	-74.009686	[54 Murray Street (at W Broadway), New York, N...	246	great	1
44	Exceed Physical Culture	53910ac3498e57a5dc0eb160	Gym / Fitness Center	40.715629	-74.007992	[97 Reade St (bet W Broadway & Church St), New...	91	great	1
73	Wall Street Bath & Spa 88	49ba7becf964a5206e531fe3	Spa	40.709278	-74.005690	[88 Fulton St, New York, NY 10038, United States]	70	great	1

Cluster 3

```
[154]: fitness_centers.loc[fitness_centers['label']==2]
```

Out[154]:

	name	id	categories	lat	lng	formattedAddress	total likes	total likes_cat	label
11	CrossFit 212 TriBeCa	52001eed498e9ac16ca5e20b	Gym	40.714537	-74.005999	[281 Broadway, New York, NY 10007, United States]	14	poor	2
64	European Wax Center	5400b87a498e87c5ae6e5143	Health & Beauty Service	40.714710	-74.007448	[95 Chambers St, New York, NY 10007, United St...	7	poor	2
98	New York by Gehry Gym	4e9b03e59a52edbd658ca490	Gym	40.710655	-74.005709	[8 Spruce St., New York, NY 10038, United States]	13	poor	2

The following observations have made from the results obtained from the cluster analysis on the data pulled from Foursquare.

Observations made from data in Cluster - 1

- Most of the Fitness centers are located near Park Place
- All the fitness centers are rated as below and above average clusters

Observations made from data in Cluster - 2

- Most of the Fitness centers located near W Broadway
- All the fitness centers are rated as great

Observations made from data in Cluster - 3

- All the fitness centers are rated as poor

After the careful analysis of the geographic locations and number of likes given by the users using clustering approach, some interesting observations have been made from the results. There are more chances to those fitness centers having good rating to attract new customers as the reviews given by the existing customers are good. The new customers also need some fitness accessories and equipment eventually.

Conclusion

This project is aimed to find a optimal location to establish a commercial fitness products store in New York, NY. Data about the fitness centers present in New York city is gathered from Foursquare website using Foursquare API. Then the data is processed to drop the unwanted venues such as malls, restaurants, Pizza shops, Ice Cream shops etc. All the type of venues related to fitness like gyms, fitness centers, spa etc. are rendered and analysed. Then, data is grouped into four bins based on the number of likes. Later, Clustering of that data is performed to divide the data into groups. After clustering we have identified the common traits in the clusters and observed cluster specific information. Based on the observations I have made some conclusions. The company will make a decision where to open a store.