

Identifying New Neighborhoods for launching Gyms / Fitness Centres

Introduction – Problem and Background

With the expanding number of organized fitness chains, companies in the fitness industry would need to identify the most optimal locations for expanding their presence. Since a new gym can involve significant capital expenditure and operating costs, it would be imperative to get the location right. From a customer perspective, location would play a very important role in overall footfalls and repeat customers. Also, customer pay attention to several factors such as parking facilities – which are partly attributable to the location.

In this problem, I have tried to identify the most suitable locations for opening a new fitness centre / gym in New York City. This data could be useful for a fitness chain to plan expansion in the city. It would also be useful for players in relatively recent categories such as Yoga Studios and Cycle Studios to see location preferences of fitness enthusiasts.

Data Description

They key data used in this analysis are as follows:

- New York City neighbourhood data
- Foursquare API location data
- 'Likes' data for gyms in NY using the stats endpoint in Foursquare

I have used the 'likes' data as an indicator of popularity of gyms in the area as well as the location preference of customers. While the checkin data may have been a better indicator – I was unable to retrieve this using the stats API and used the likes data instead. One key assumption is that customers would generally have only one gym subscription at a time.

Methodology

Key steps in the methodology are as follows:

- Using the neighbourhood data json file and the Foursquare API, venue data on all neighborhoods in NY was retrieved, including the Venue ID for each venue
- From the venue data, a subset of rows with the string "Gym" in the Venue Category field was extracted. This subset had 266 venues
 - o The listing of all categories in the venue data indicated that there were multiple fitness venues such as outdoor sports locations, tennis courts, etc., but only gyms have been used for this analysis to limit the number of API calls needed for further analysis
- A loop was run on all the Venue IDs to retrieve additional venue data with the stats endpoint in Foursquare
- Average number of likes for each neighbourhood was calculated and saved in a separate dataframe (num_likes)
- Cluster analysis was performed on the num_likes data frame with k = 5
- The cluster labels were combined with the latitude / longitude data into a new data frame
- A cluster map was plotted using the above analysis

Results

A summary of the clusters in terms of average likes:

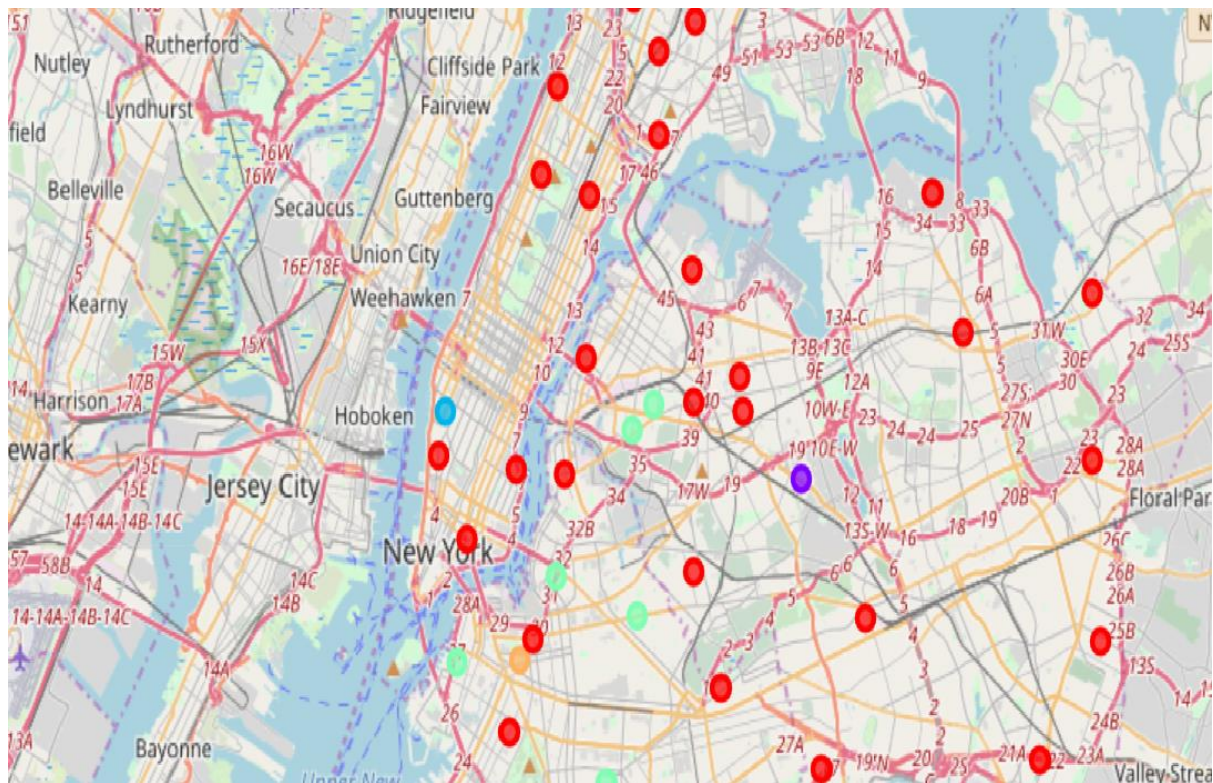
	Likes
Cluster Labels	
0	8.911111
3	40.727273
1	75.000000
4	153.000000
2	321.000000

As can be seen in the above table and the final_df dataframe (please see the notebook provided):

- Most of the Neighborhoods belong to cluster 0, followed by cluster 3

- Neighborhoods in Cluster 3 and Cluster 1 have higher ratings than those in cluster 0
- Clusters 4 and 2 have a limited number of data points and can be treated as outliers

Snapshot of Cluster map:



Discussion

Based on this analysis, some of the best neighborhoods in terms of likes for existing gyms (see final_df in Jupyter notebook provided) would be (top 5 from Cluster 3):

- Hammels
- Bushwick
- Williamsburg
- Cobble Hill
- Bay Ridge

Also, lower data availability for Neighborhoods in Clusters 2 and 4, along with high likes, could indicate potential pockets of high demand in these areas.

While Cluster 0 has a large number of neighborhoods, these generally have zero or significantly lower number of likes – that could also indicate that these are relatively newer gyms, so may be worth exploring for new launches.

Conclusion

While clustering may be a good preliminary approach to identifying locations for new gym launches, there may be a number of other factors involved such as:

- Tiering (high end vs low priced gyms)
- Types of activity / workouts offered
- Availability of other facilities in the area (such as outdoor sports venues, swimming pools, etc.)
- Real estate prices /rentals
- Office locations in the vicinity