Predicting the best location for a new Wellness-Fitness Center

Anna Ferré

December 9, 2019

1. Introduction

In this project we will try to find an optimal location for a kind of wellness-fitness center: a center where you can go to gym, spa or yoga classes. Specifically, this report will be targeted to stakeholders interested in opening this kind of center in **Manhattan, New York, US**.

Since there are lots of gyms, yoga centers and spas in NY we will try to detect **locations that are mediumly crowded with these services**. We think that if there are a lot of gyms, spas or yoga centers in certain areas, it has to be for a reason. So, lets focus in this areas and also skip the more crowded.

We will use our data science powers to generate a few most promising 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.

2. Data acquisition and cleaning
2.1. Data sources

We will use a dataset that contains the 5 boroughs and the 306 neighborhoods existing in NY city. This dataset exists for free on the web: https://geo.nyu.edu/catalog/nyu_2451_34572.
Since the factor that will inlfuence in our conclusión is the amount of existing gyms, spas and yoga centers in all the differents neighborhoods of Manhattan, we will extract this data from Foursquare API

2.2. Data cleaning and feature selection

Data related to Boroughs and Neighborhoods of New York, will be downloaded from a json file and extracted and set into a dataframe. Then, only the data related to Manhattan will be used and stored in a other dataframe.

Data related to Venues of Manhattan neighborhoods will be cleaned: we will discard all the categories that are not related to Gym, Fitness, Sport, Yoga, Pilates, Spa, Nail Salón, Health & beauty and so on.

3. Exploratory Data Analysis

After cleaning the data, the first steps will be to explore Manhattan looking for Gyms, Spas and Yoga centers. We will use Geopy library in combination with Foursquare API.

Then we will select only the neighborhoods which are médium crowded of this kind of venues. Thus, we chose neighborhoods that have between 80 and 130 venues.

The lasts steps will be group rows by neighborhood and take the mean of the frequency of occurrence of each category's venue. Finally, we will cluster the data in order t osee which neighborhoods are similar in terms of venue's category.

4. Conclusions

Our analysis shows that in Manhattan there are 1662 gyms or fitnes centers, 1834 spas, massage studios and so on and only 20 yoga centers. They are distributed in 39, 40 and 9 neighborhoods, respectively.

The analysis also shows that the distribution of the venues are not uniform among the neighborhoos of Manhattan. Midtown South has the large amount of this kind of venues (180) whereas Morningside Heights, Manhattanville, Stuyvesant Town and Central Harlem only have 12 of them.

By clustering the data, we've seen that the following neighborhoods are similiar among them:

- Lincoln Square, Clinton, Battery Park City and Tudor City
- Yorkville, Lenox Hill, Chelsea, East Village, West Village, Gramercy, Carnegie Hill, Noho and Turtle Bay
- Chinatown and Greenwich Village

Purpose of this project was to identify Manhattan neighborhoods with medium number of gyms, spas and yoga centers in order to aid stakeholders in narrowing down the search for optimal location for a new Wellness-Fitness Center. By analysing the data we decided to not consider the neihborhoods with really low and really high amount of fitness/wellness venues. At this point, we have only 15 neighborhoods to chose.

Clustering of those locations was then performed in order to create major zones of interest.

One important point to consider is that, after we reduced of neighborhoods, only Noho, Lenox Hill and Greenwich Village have yoga centers. It's important if stakeholders want to chose one area with some representation of this sport.

Finally, since the first Noho and Lenox Hill, the ones with yoga centers, are in the same cluster, and also this cluster is the biggest one, chose one of these two locations to start the new Wellness Center would be the best option. Other neighborhood of the same cluster could be a good option since due to the scope of the cluster it will be easier to find variety in additional factors such as attractiveness of each location (proximity to park or water), levels of

noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.