Capstone Project – The Battle of Neighborhoods

Section 1 – Introduction

Location is one of the most important reasons why a new business will succeed or fail. In this example, you are a luxury gym builder and you want to use an analytical approach with location data to find which Manhattan neighborhood would be best to build your next gym. There are dozens of zip codes to choose from, so how would one go about determining a ranking of which zip codes would be best to build in?

The audience here would be anyone that needs to choose a neighborhood to build a project (in this case, a luxury gym builder). The analysis done can be broadly applied to another problem where you want to look at data to determine the best neighborhood to build in (ex: wine bar, upscale Italian restaurant, etc.)

Section 2 – Data

We need to gather data for each zip code in order to determine a ranking. First you will need median income data for each zip code in Manhattan. This data will come from here:

http://zipatlas.com/us/ny/new-york/zip-code-comparison/median-household-income.htm

Then you will feed each zip code into Foursquare's API to find all gyms within a 1500-meter radius of the center of that zip code. This will come from the venue recommendations API call:

https://developer.foursquare.com/docs/api-reference/venues/explore/

Then you will gather data on each gym and determine the max rating, average rating, and number of likes for all the gym returned from the previous section. This will come from Foursquare's venue details call:

https://developer.foursquare.com/docs/api-reference/venues/details/

Section 3 – Methodology

Once we gather the data, we will need to organize a dataframe with the income, max rating, average rating, and number of likes for all the gyms within a 1500-meter radius of each zip code. This data will then be normalized to a range between 0 and 1 by using sklearn's minmaxscaler function. I will then create a metric called "Gym Score" where I multiply the median income by 5 and subtract max rating times 2, average rating, and sum of likes by 2. I

chose these values because income is the most important factor for a luxury gym – people with more dispensable income are willing to pay extra for a better gym. I subtract the max rating times two because the gym with the highest rating in the neighborhood will likely be our direct competition and the higher the rating, the worse the zip code is for the builder. I also subtract average rating in the gym score, but I put less weight on this because not all gyms will be the builder's competition. A higher average rating in a neighborhood means more choices of gyms in that zip code people that people enjoy. Finally, I subtract sum of likes times two from the gym score. This metric gives an overall sense of how satisfied people are with the gyms in their area, and an area with a high number of likes is less appealing to build a new gym in.

Section 4 – Results and Discussion

Here is a table with the top 5 and bottom 5 results:

Zip Code	Income	MaxRating	AvgRating	SumLikes	GymScore
10004	0.887008	0.333333	0.397727	0.045812	3.279022
10162	0.953789	0.523810	0.625000	0.164921	2.766485
10007	1.000000	0.809524	0.795455	0.513089	1.559320
10280	0.955013	0.809524	0.840909	0.500000	1.315109
10025	0.355295	0.285714	0.284091	0.014398	0.892158
10013	0.238733	0.857143	0.909091	0.325916	-2.081545
10011	0.480260	0.904762	0.890152	1.000000	-2.298374
10029	0.074818	0.904762	0.681818	0.193063	-2.503377
10036	0.266249	0.952381	0.984848	0.556937	-2.672239
10038	0.167464	0.809524	0.844156	0.592277	-2.810439

The gym builder would likely want to build in zip codes 10004 and 10162, with the possibility of expanding to 10007 and 10280. Both 10004 and 10162 stand out because they are some of the highest income neighborhoods with relatively low max ratings, average ratings, and number of likes for the gyms in that area. Although zip code 10025 is close to the bottom 3rd of median income zip codes, it gets a relatively high gym score due to the bad ratings and low likes of the gyms in that area. The bottom five neighborhoods in this list also make sense – they have incomes in the lower half of Manhattan zip codes with gyms that have relatively high rankings in the max rating, average rating, and likes categories.

Section 5 – Conclusion

We have used a data driven and analytical approach to determine which are the best zip codes to build a luxury gym in. A similar type of analysis can be used to rank where to build other type of businesses by modifying the calls to Foursqaure's API.