Fit and Food:
Where to go when food is 80% of your progress?

Wouter de Bot

1. INTRODUCTION

During recent years it has become ever more clear that obesity is a public health issue. In the United States the age adjusted obesity rate for adults over the age of 20 is as high as 39.8 %. Shockingly over 71.6 % of the American adults are overweight. "FastStats - Overweight Prevalence", 2020What many people that try to get into shape don't know is that food consists of 80 % of the progress you make towards getting into shape. Because of this a GymCorp has come up with an idea to combine a Gym with a healthy restaurant where you can go for your pre- or post workout meal. In order to determine the optimal location for this gym-restaurant a data analysis will be done for New York city by using publicly available data as well as the FourSquare API.

2. DATA

2.1. **Data Acquisition and example.** The data that is going to be used in order to determine the optimal location for a gym and restaurant combination will consist of several datasources. These datasources are either open source or obtained via a free account on FourSquare.

- FourSquare location Data
- 2014 New York City Neighborhood Names

From the FourSquare API the data that will be used are the locations of Gyms and Restaurants in the city of New York. This data will be combined with the New York City Neighborhood Names dataset. Combined this data will likely yield the optimal neighborhood for the Gym-Restaurant after thorough analysis.

A snapshot of the New York City neighborhoods used are shown in Figure 1.
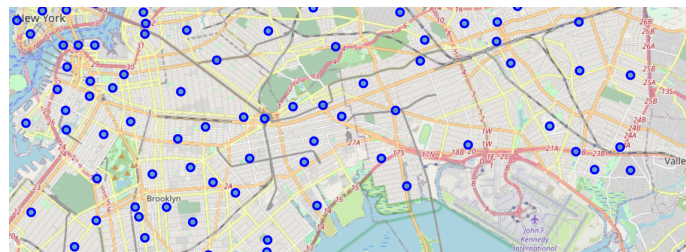
Figure 1. Snapshot of neighborhoods in NYC

An example of the data that is obtained from FourSquare is shown in the table below.

	name	categories	lat	lng
0	The Class by Taryn Toomey	Gym / Fitness Center	40.712753	-74.008734
1	CrossFit 212 TriBeCa	Gym	40.714537	-74.005999
2	Exceed Physical Culture	Gym / Fitness Center	40.715629	-74.007992
3	Equinox Tribeca	Gym / Fitness Center	40.714099	-74.009686
4	The Helena Gym	Gym	40.714276	-74.005967

3. METHODOLOGY

As a quick reminder: The data we have at this moment is: NYC neighborhood data, gyms in an area and the restaurants in an area.

The first part of the analysis will consists of an exploratory data analysis. Here the amount of restaurants and gyms in a neighborhoods will be evaluated. It is expected that insights will arise from this data about general neighborhoods that are crowded with restaurants/gyms or not.

From the information that is obtained in the first part of the analysis additional features will be constructed. An example of this could be the ratio of Restaurants to gyms in a region or total gyms in a neighborhood. Another feature that could be used is the amount of fast food restaurants in a neighborhood which could be an indication of a non-healthy environment.

4. ANALYSIS

The analysis will be done using clustering of neighborhoods. What we want to achieve with this clustering is defining neighborhoods dat are both healthy as well as not being to crowded with restaurants or gyms. In order to achieve this clustering the features that will be selected or created will be health and restaurant/fitness area-density related.

4.1. **Feature selection.** Group by neighborhood For each neighborhood the following features are calculated:

- Ratio of gyms to restaurants: High ratio indicates a healthy region
- Total amount of gyms: A lot of gyms also mean healthier regions however we should strive for a healthy region with the lowest amount of gyms
- Amount of restaurants: We want to minimize the amount of restaurants

4.2. **Modeling. Clustering** For the clustering the K-Means algorithm will be used. This is done because it is possible to suply the number of clusters based on k. The goal of this clustering is to divide the neighborhoods in "Healthy" and "Un-healthy". Because of this the number of clusters is chosen to be 2.

5. RESULTS AND DISCUSSION

The results as obtained from the clustering will be shown in a map of New York City. On this map the different neighborhoods will be shown with their corresponding cluster label in the form of a color. From these neighborhoods a neighborhood is selected in the conclusions as a possible neighborhood in which to start Fit and Food. The map above shows all the neighborhoods of NYC with their corresponding cluster. Here the purple color corresponds with the "Healthy"
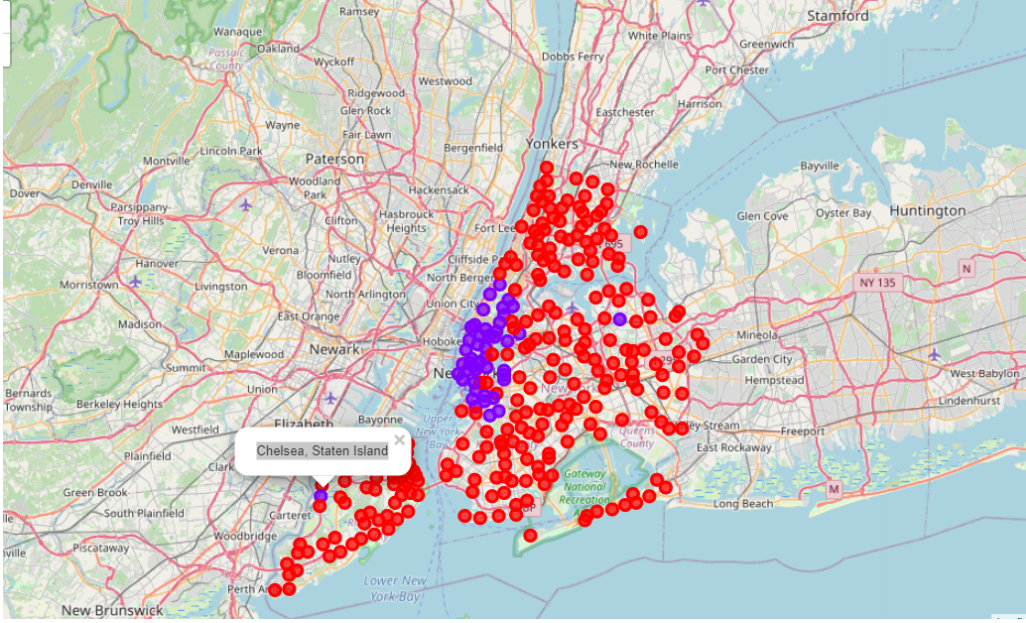
Figure 2. All the neighborhoods of NYC and their corresponding cluster.

neighborhoods and the red color with the "Less-Healthy" neighborhoods. As a reminder, healthier neighborhoods contain more gyms and/or have a higher Gym to Restaurant ratio. In order to increase the accuracy of this clustering several recommendations are given in the recommendation section.

From the results shown below we can see that indeed the neighborhoods with high G/R ratios are clustered together. Another reason why a high G/R ratio is beneficial for Fit and Food is the fact that there are less Restaurants which could make up for the potential surplus of Gyms.

	Cluster Labels	neighborhood	gr_ratio	gym_count	restaurant_count	Borough	Neighborhood	Latitude	Longitude
269	0	University Heights	0.033333	2.0	60.0	Bronx	University Heights	40.855727	-73.910416
270	1	Upper East Side	1.000000	100.0	100.0	Manhattan	Upper East Side	40.775639	-73.960508
271	1	Upper West Side	0.760000	76.0	100.0	Manhattan	Upper West Side	40.787658	-73.977059
272	0	Utopia	0.102041	5.0	49.0	Queens	Utopia	40.733500	-73.796717
273	0	Van Nest	0.046154	6.0	130.0	Bronx	Van Nest	40.843608	-73.866299
274	0	Vinegar Hill	0.893617	42.0	47.0	Brooklyn	Vinegar Hill	40.703321	-73.981116
275	0	Wakefield	0.043478	2.0	46.0	Bronx	Wakefield	40.894705	-73.847201
276	0	Washington Heights	0.160000	16.0	100.0	Manhattan	Washington Heights	40.851903	-73.936900
277	0	Weeksville	0.127907	11.0	86.0	Brooklyn	Weeksville	40.675040	-73.930531
278	0	West Brighton	0.080000	4.0	50.0	Staten Island	West Brighton	40.631879	-74.107182
279	0	West Farms	0.056604	3.0	53.0	Bronx	West Farms	40.839475	-73.877745
280	1	West Village	1.000000	100.0	100.0	Manhattan	West Village	40.734434	-74.006180

Figure 3. Data from which the map is constructed

6. Conclusions

When looking at the results as obtained form the clustering it becomes evident that there are regions in NYC that are more likely to be healthy. However just picking one of the purple neighborhoods to start Fit and Food would be foolish.The reason for this is that rent prices will vary greatly

among the different neighborhoods. For example the rent prices in Manhatten will be a lot higher than in the Bronx. Because of this Murray Hill,Queens and Chelsea, Staten Island are picked as the neighborhoods where Fit and Food is most likely to succeed. However please note that a multitude of other factors come into play in finalizing the possible location. Thoughts on how to improve this analysis and thus mitigate the aforementioned problems are shown in the recommendation section

7. Recommendations

In order to improve the analysis done in this project several recommendations can be made. When comparing the amount of gyms in an area one could look at the population density in that neighborhood. This could be a possible explanation for a decreased number of gyms. Another improvement that could be made is , when possible to obtain, using public health data to cluster the neighborhoods. The reason for this is that a healthier neighborhood will more likely make use of Fit and Food.

Would it be the case that health data is unavailable another way of determining the health of a neighborhood would be to obtain the percentage of fastfood restaurants in a neighborhood compared to the total number of restaurants.

Another way of improving the model is by adding the cost of rent as well as the population density of each neighborhood. Population density can be used to normalize the amount of gyms or restaurants in a neighborhood.

Furthermore during the development of the Fit and Food concept and franchise it is important to continiously re-evaluate the assessments done in this analysis based on changing circumstances or business findings.

References

Faststats - overweight prevalence. (2020). https://www.cdc.gov/nchs/fastats/obesity-overweight. htm