# Using location data to find profitable locations to setup Healthy Food Centers in New York city

Gourab Ghosh

# Background

- People are becoming more health conscious now a days, which is increasing the demand for healthy foods.
- Business ventures are trying to take up this demand and are coming up with a lot of wellness and healthy food related offerings.
- Location of a healthy food store is vital in the business growth and profitability.
- We can predict the profitable locations in a city where owners can open their stores for selling healthy foods.

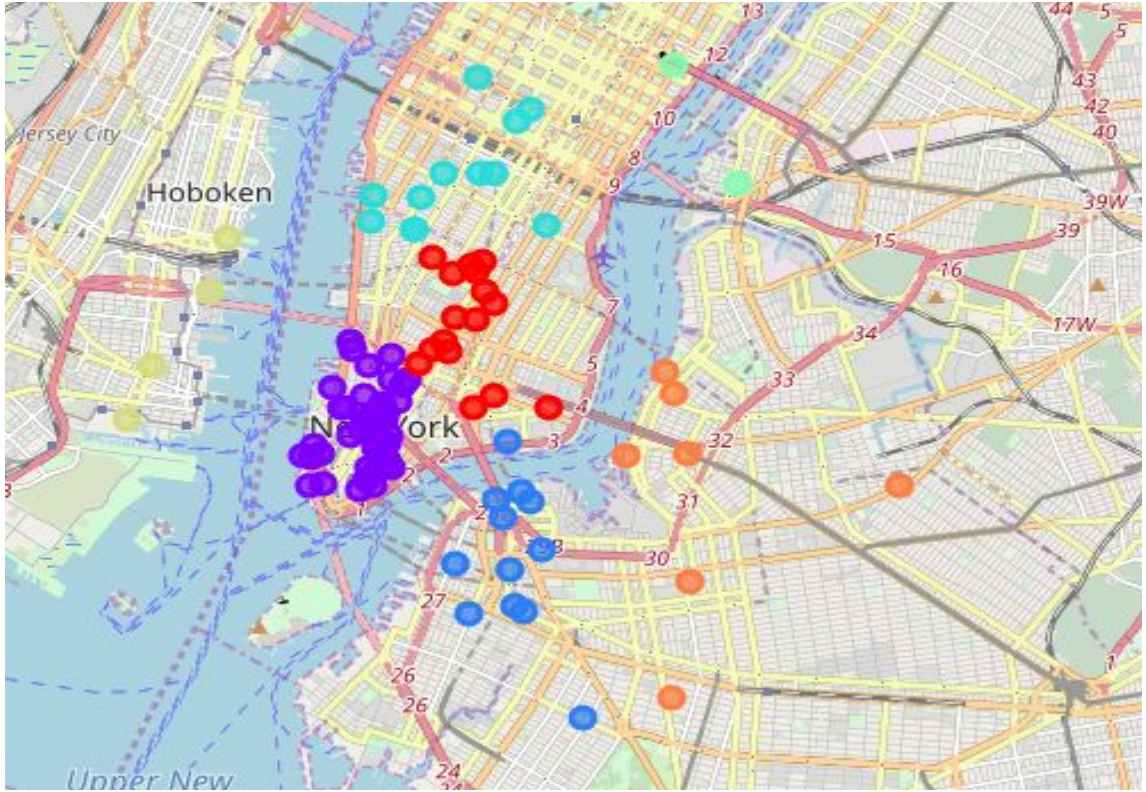
# Data acquisition and cleaning

- We can plan to open the stores near the fitness centers in the city
- Here we take the city of New York as the base location for our prediction.
- Foursquare provides APIs which return venues by search terms - we fetch venues for ‘Gym’ and ‘Yoga’ using Places API by Foursquare for building our base dataset of locations.
- Then we append both the datasets and remove unnecessary columns to arrive at our required set of data containing venue name, address and the latitude and longitude.

# Methodology for prediction

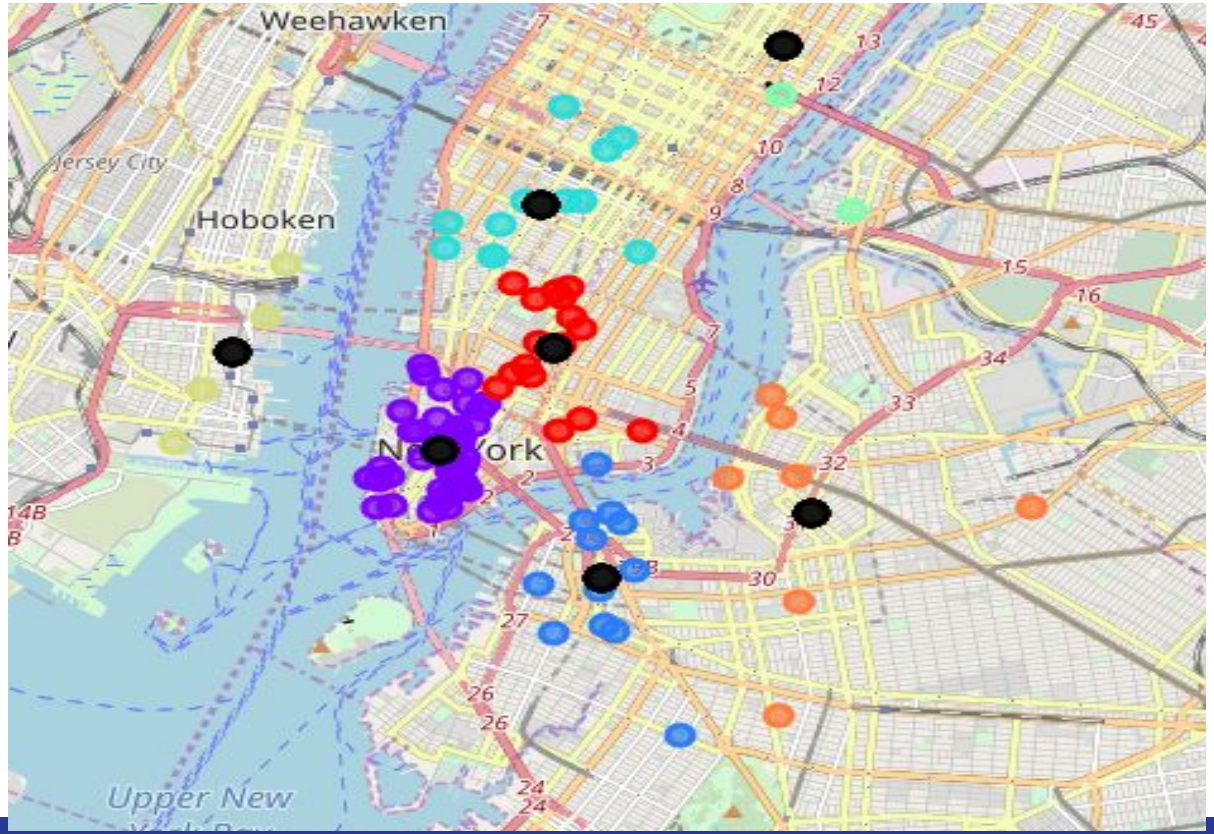
- We plan to open stores in locations where the gym and yoga centers are located in dense numbers. This will lead to more footfall of the targeted customer base.
- For this, we cluster the location dataset into 7 clusters. We use k-means clustering algorithm to do this.
- Then, within every cluster, we fetch the centroid location which can be considered as the predicted location for opening the healthy food store.
- We plot all the location data points as well as the centroids on the map of New York city for a visual depiction.

# Methodology for prediction (continued)



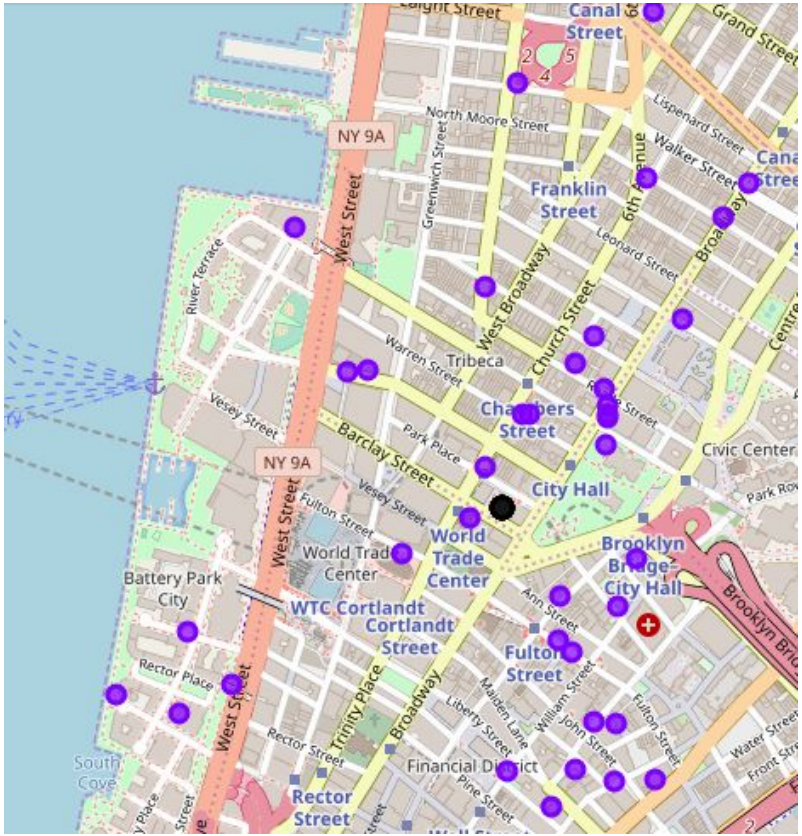

Map of New York with gym & yoga venues colored as per their clusters

# Methodology for prediction (continued)




Map of New York with gym & yoga venues colored as per their clusters along with the centroid locations shown in black

# Discussion




From the map shown above, we can conclude that our clustering mechanism shows lower Manhattan bay area as having the most dense occurrence of gym and yoga centers.

The area is colored in violet. The centroid location centered there with co-ordinates as **(40.71255502453713,-74.00864113580623)** near Barclay Street can be the most profitable venue to open a healthy food center. In real life, the venue of the food center can vary based on budget and other constraints to a certain permissible extent.

# Discussion (continued)


The second most profitable location is the area around Bond Street with co-ordinates as **(40.72619606949153,-73.99339449287002)** marked by deep orange color in this map. It has density lesser than the first one.

# Discussion (continued)


The third most profitable lcoation is the area around Tillary Street with co-ordinates as **(40.6958098909231,-73.98659838836905)** marked by deep blue color in this map. It has density lesser than the first and the second one.

# Conclusion

- Density falls drastically in the other clusters and the venues are located far away from each other. So, it is doubtful whether opening up a healthy food center will really be helpful in those places or not.
- Since the city of New York is very big in size and there are numerous gym and yoga centers, we can improve the efficiency of our method by running the study based on specific boroughs of the city. For example, we can take Manhattan as the base location and by taking its radius, we perform the search for location datasets. This can significantly improve the prediction.
- Also, if some business owner is interested in opening up the store in a preferred borough of his/her, we can take the center of our study around that location and make searches.