**Prediction of potential locations for a sports shop**

Pratheeksh Eapen

16th June 2020

1. **Introduction**

**1.1 Business Problem**

A person wants to start a sports shop in Manhattan, New York. For this problem we have decided to use the potential customer bases in and around Manhattan to generate locations. Potential customer bases are:

* Colleges and Universities
* Elementary schools
* Middle schools
* High Schools
* Private Schools
* Rec centers
* Stadiums

**2. Data acquisition and cleaning**

**2.1 Data sources**

* Data of **Manhattan, New York** was taken from newyork\_data json file.
* Data of **potential customer bases** including coordinates were obtained using Foursquare API.
* The ID’s of the various customer bases to be searched for using the Foursquare API was obtained from “https://developer.foursquare.com/docs/resources/categories”
* **Addresses of the potential locations** were obtained using  LocationIQ API reverse geocoding

**2.2 Data cleaning**

The data of the potential customer bases once obtained using the Foursquare API was stored in dataframes. These dataframes were then individually inspected and then cleaned as each category had various categories in itself. So to decide which ones to use for the analysis, each dataframe had to be inspected. Figure 1 and Figure 2 below shows how many categories were there in the category “College” and “Rec Center” before it was cleaned:

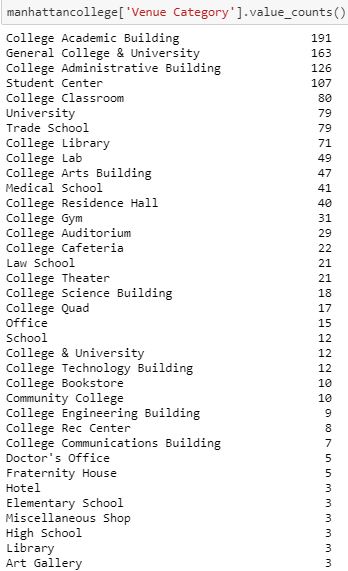
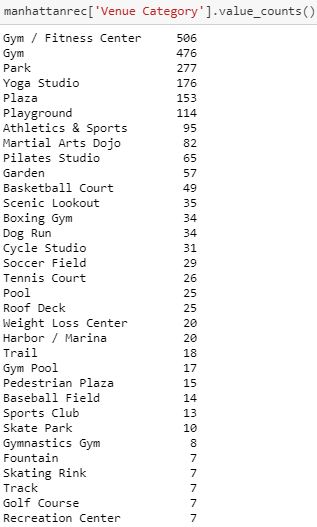
 

Figure 2

Figure 1

After the dataframes of each category was cleaned they were all compiled into one dataframe. A total of 1316 venues were obtained after ignoring several categories as we wanted the cluster centers to be closer to the customer bases which could generate higher revenue.

**3. Analysis**

After the data was cleaned, the coordinates of these customer bases were taken and plotted on a map to get a better idea as shown in Figure 3.

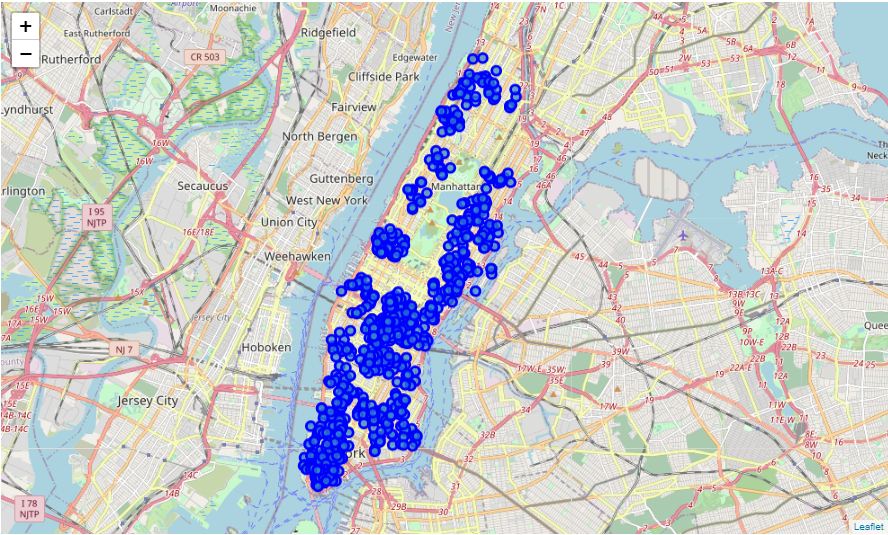


Figure 3

Now that the points have been plotted, these coordinates were used in creating clusters using k-means clustering. The number of locations required can be used as the number of centers i.e. the number of clusters formed. Here we have used 20, so 20 clusters were obtained as shown in Figure 4.

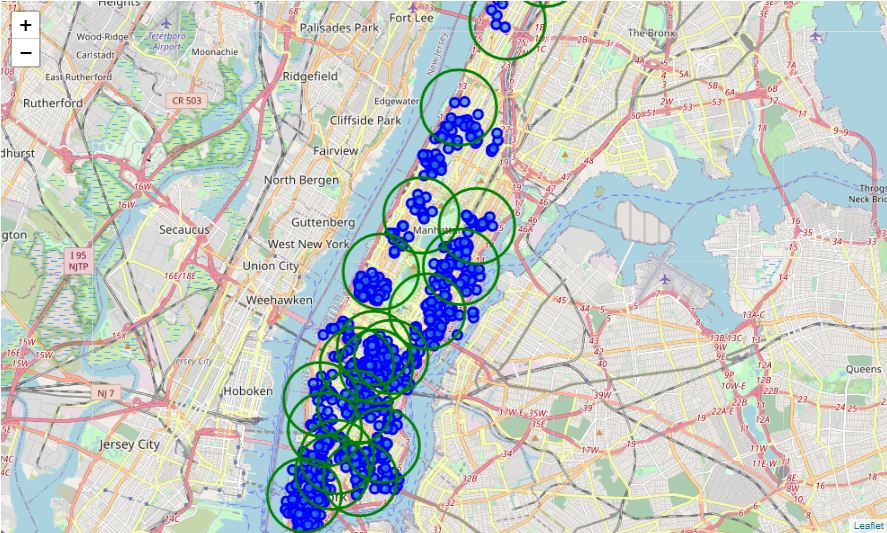


Figure 4

After obtaining these clusters, the centers of these clusters were used as the potential locations to set up the shop. These centers were then stored in a list. To obtain the addresses of these centers, the coordinates will have to be reverse geocoded. In order to do so the LocationIQ API was used. The addresses obtained after reverse geocoding the coordinates is shown in Figure 5.

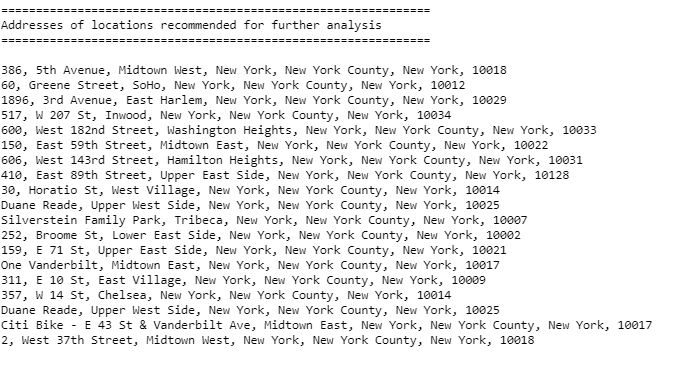


Figure 5

**4. Result**

* The analysis done in this project shows that there is enough and more customer bases in and around Manhattan, New York for a sports shop.
* In this analysis we first got all the potential customers and then we clustered them and used the centers of these clusters as potential locations for the shop/store. Addresses of these locations were obtained using reverse geocoding.
* As a result of this analysis we have generated 20 potential locations. These of course are not the exact optimal locations for the stores. There may be other reasons for which there are no shops in the given location.
* This project only takes into account potential customer bases and doesn’t take into account the close by shops. In this case it was done because a shop like SKECHERS being beside the shop can help boost sales. So a closer manual inspection is required. These conditions can be changed according to the objectives and requirements of the stakeholder.

**5. Conclusion**

* Purpose of this project was to generate potential locations to set up a sports shop depending upon the number of potential customer bases in and around the neighborhood. The potential customer bases and its coordinates were obtained using Foursquare API. The data was inspected and cleaned and then clustered using k-means clustering. The centers of these clusters were taken as potential locations and their addresses were obtained using LocationIQ reverse geocoding API.
* The final decision is up to the stakeholders and may require further manual inspection or on foot inspection as they may have other factors in mind too such as attractiveness of the location, proximity to main roads etc.