

|  |
| --- |
| Capstone Project for Data Science  2020 |
|  |
| September 9  Coursera  Authored by: Anisha Rajpal |

|  |
| --- |
| **Business Problem** ‘Health is wealth’ is quite known and old saying. Eating well and exercising are supplementary to each other. Many doctors recommend that after a good workout, one needs to rest and eat something healthy as body absorbs it rampantly. We have gyms and fitness centers for working out but for keeping the dietary restrictions in place we need healthy food centers. The question is where to open such centers so that they are able to generate profits and people are able to make well use of it. The initial cities that are on the mind are Los Angeles and New York City primarily because of their diversity and population.   **Proposed Solution** Data for gyms/fitness centers/yoga studios/wellness centers would be collected for the two most populous cities of USA and compared with each other in terms of which one has more spread of the fitness centers and type of institutes. This would help to understand the comparison between the cities and prospective areas of opening Healthy Snack Centers. These healthy snack centers can also deal with workout equipment/accessories if required. The reason for comparison is to check which city can be targeted first in regard to dense cluster of centers.  Data Gathering  For collection of data, we need to find health centers in the cities and then plot the map to see how they are distributed within the city. More the dense cluster, more probability of project working out. Also, we need to determine which type of center is more dominant in any city to reckon the scope of equipment sale along with snack bar. For example, people who go to gyms would immediately need protein supplements and those going for Yoga would prefer a fresh juice/fruit salad.  Data Sources  To get the centers, we would be utilizing Foursquare API which give the coordinates of the centers so that we can easily plot them on map. We would be using bar graph to group the type of health centers on basis of their category and mention their number. For example, for LA, this is how a bar diagram looks like:    And then there would be a map using Folium library, which would show the density of health centers within the city and are prospective places to have healthy snack centers/equipment store.  Methodology  The main component of the analysis here has been the plotting of the health centers which API has brought forward. The places/health centers were searched on basis of the Latitude and Longitude of the cities and not specifically neighborhood. The reason for this selection was my target audience would be looking at city as a whole and find the respective health centers packed areas. Foursquare API has brought the list of 100 health centers each as this was the limitation with using Foursquare API. These places have been marked using Folium map and can be zoomed in and zoomed out to see palatial spacing between two or more places of interest.  It is noted which areas have maximum presence of health centers whether be stand-alone or inside a building or a hotel. Wherever, there are health freaks there is a good chance of the market for healthy food and equipment.  For exploratory and comparative analysis , I did an online research to find which of the city is more expensive and as per this (<https://www.bestplaces.net/cost-of-living/los-angeles-ca/new-york-ny/100000>) and many other websites Los Angeles is cheaper than New York City.  In respect of utilizing Machine Learning, I have used the creation of Map using Folium library and use of Matplotlib, Numpy and Pandas.  Results  On basis of the data received from Foursquare API , I have plotted the places for both New York City and Los Angeles.  For Los Angeles (LA), it looks something like this:    It can be inferred that Financial District has more of clustered fitness centers than the other parts. And then there is equal scope in Historic Core district for the same. Also, it is seen that most of the health centers are a bit distant from the main city point. But there is lot of potential in many areas where at least more than 4 centers are present.  For New York City, it looks like this:    It can be seen that although most of the health centers are far away from the main city point, there are clusters of places which can be targeted. There are many areas with at least 4 health centers placed together.  Now checking the type of centers both of these cities have which would also define the focus which storekeeper has to take care of for better sales:  For LA,  This is how distribution of health centers look like. |
| For NYC, it looks like: |
| The query used to get the data from API was ‘gym’ which specifically resulted the data with category as gym or name as gym. If we compare the data in both, it is seen in LA approx. 96% are established centers and 4% places are either in building or residential complex and for NYC, number is almost similar with 95% established centers and 5% building or shared complexes.  Discussion  It is well noted from above maps and graphs that there is not much difference in how the center distribution is there for the cities. Mostly its gym that dominates the market.  For NYC, its approximately 5 prospective areas with 4 or more health centers that are prospective areas. For LA, its approximately 6 prospective areas but health centers in LA are more nearby placed than they are in New York.  Also, as per the research done as mentioned in previous section, people use more of their own vehicle in LA than in NYC which makes them more accessible to the nearby probable healthy snack bar even though they are physically exhausted. |
| Conclusion  On basis of the requirement and analysis done, I am concluding that there are lot many health/fitness centers present in both cities and some of the centers are nearby located ,hence offering the area as potential candidate for opening up a healthy snack bar plus health equipment store.  Suitable areas in LA include Financial District, Historic Core District, Bunker Hill, Hope Place, Orsini, South Main Street.  Suitable areas in NYC include Bleecker Street, Union Square, Lincoln Tunnel, 86th Street, 70th Street  On basis of other studies done on expense and accessibility to store, I think LA offers an edge over NYC. |
|  |