IBM APPLIED DATA SCIENCE CAPSTONE PROJECT

PLACES FOR OPENING A NEW VEGAN RESTAURANT COMPLEX AT MUMBAI, ECONOMIC HUB OF INDIA

By: Ardhendu Mahatha

27th June,2020

INTRODUCTION

A restaurant is generally an establishment where the public may obtain meals or refreshments.As with so many ideas that evolve over time, restaurants now serve a larger role in society... they have become places of social contact, of discovering new cultures and tastes from far-away lands, of spending an evening with your loved ones, of clinching business deals over a glass of wine, and so forth. All this, of course, in addition to the basic functions of “restoring” people with the help of good food, service and ambience. When it comes to vegetarian people ,a complete vegan restaurant plays the most vital role. In Mumbai, where a diversity of people stays from all over the world. So, we have to choose such a location that will be beneficial for the person who is trying to open the business. Crowded places where there are very less amount of vegan restaurants will be an idle site for the project.

BUSINESS PROBLEM

The goal of this project is to find the ideal place to launch the vegan restaurant in Mumbai, India. We have to find the best location using the Data science and Machine Learning techniques like clustering and also using data visualisation tools like Folium maps etc..Now we have to spot where Vegetarian people are more in number and Vegan Restaurants are less.

Where would you recommend to open a Vegan Restaurant Complex in Mumbai, India?

TARGET AUDIENCE

My project is solely focussed on the businessman trying to open a new Vegan Restaurant complex in Mumbai. Every year more people are migrating to Mumbai, so restaurants are also growing in more numbers in the city. But Vegan Restaurants are comparatively less in number.

DATA SECTION

The data that we need is :

- The neighbourhood areas list in Mumbai. This will create the extent of the project in the city. More nos of neighborhood included, more places we can explore.
- Next, we will have to get the latitude and longitude of those places and merge it with the neighborhood areas.
- Furthermore, we have to explore these places for vegan restaurants and will use this for clustering.

The neighborhood areas data we will get from wikipedia page: https://en.wikipedia.org/wiki/Category:Suburbs_in_Mumbai. We will use this data to separate the name of the neighborhoods of Mumbai. Next, we will use the geocoder to find the latitude and longitudes of these places and merge it with the previous DataFrame. Next we will use the FourSquare API to explore these neighborhoods using the location details and get the venues data. And finally using the venues to plot on the map using folium maps. Next, we will use ML techniques like K-Means clustering to cluster those points.

METHODOLOGY

We will get the data of the neighborhood we need from wikipedia from the link https://en.wikipedia.org/wiki/Category:Suburbs_in_Mumbai. We will do Web scraping using python requests to fetch the data from wikipedia. After that we will use beautiful soup to extract the data and then use the pandas dataframe to store the data. In this way, I have obtained a list of all the names of the neighborhoods of Mumbai.

Next we need to convert the addresses to their geographical coordinates, we will use geocoder for that, we will use python to put the geocoder in a loop and extract all the places coordinates and put them in a separate data frame. Next we will merge the data frames of the names of the neighborhood with their location data. This allows us to do a sanity check whether the data received are correctly plotted on the Map of Mumbai.

	Neighborhood	Latitude	Longitude
0	Andheri	19.118459	72.841763
1	Anushakti Nagar	19.042830	72.927340
2	Baiganwadi	19.062940	72.926630
3	Bandra	19.054370	72.840170
4	Bhandup	19.145560	72.948560

Next we will use the folium maps to visualise these neighborhoods within the map of Mumbai. After that we will use the FourSquare API to explore the venues nearby in these neighborhoods. We will use it to get 100 venues that are within a radius of 500. We will make an API call to obtain the FourSquare to explore the neighbourhoods and get the json files to get the details of the venue. Now we will use this json file To extract the details of the venues from it like venue category, venue latitude, venue longitude etc. We will then examine the data to find how many unique categories of venues exist and then we will perform one-hot encoding to divide the data in categories. Then we will analyze the data by grouping them as per neighbourhoods. We will filter for Vegan restaurants and sort the data for Vegan Restaurants.

Lastly we will use the K-means clustering to cluster the labels. This algorithm identifies the clusters with k no of centroids. We will cluster the neighbourhoods for k=3 no of clusters depending upon their frequency of presence. This result will allow us to identify the presence of Vegan Restaurants. The sample is shown below:

:

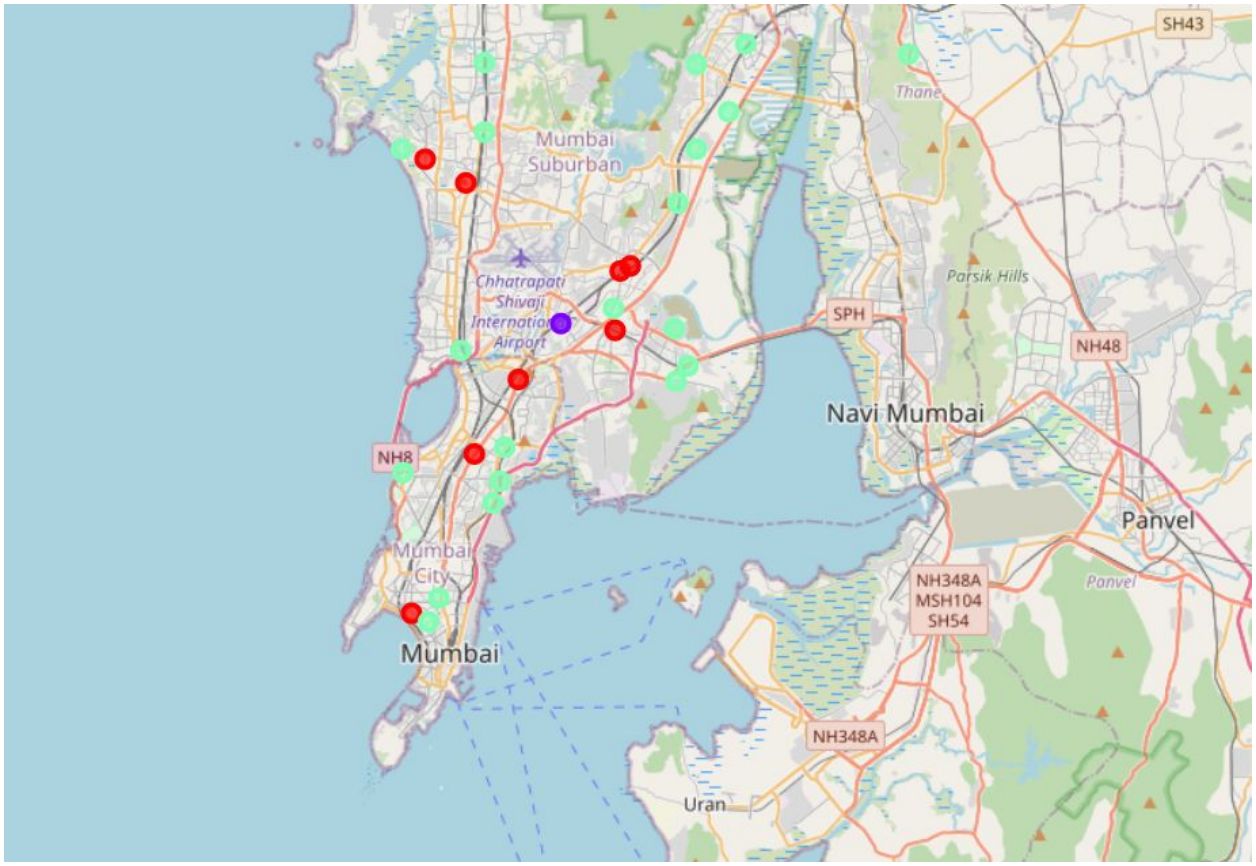
	Neighborhood	Vegetarian / Vegan Restaurant	Cluster Labels
0	Andheri	0.058824	0
1	Anushakti Nagar	0.000000	2
2	Baiganwadi	0.000000	2
3	Bandra	0.000000	2
4	Bhandup	0.000000	2

RESULTS

The results of K-means clustering shows how we can categorize the neighbourhood into 3 clusters based on the amount of occurence of Vegan Restaurants.

1. Cluster0: Moderate no. of Vegan Restaurants.
2. Cluster1: High concentration of Vegan Restaurants.
3. Cluster2: Extremely low no of Vegan Restaurants.

The results are visualised in this map cluster0: Red, cluster1: Blue, cluster2: Green.



Places in

1. Cluster1: Kurla
2. Cluster0: Andheri, Kalyan, Juhu, Vashi, Mira Road, Ghatkopar, Kosa, Chembur, Sion
3. Cluster2: Sonapur, Mumbra, Thakur Village, Mulund, Tilak Nagar etc.

DISCUSSION

Vegan restaurants are present in maximum nos at Kurla area. And it is in moderate amount at areas like Andheri, Juhu, Kalyan, Vashi, Chembur etc. and present in a very low amount at areas like Worli, Borivalli, Bhandup etc. So, businessmen wishing to open Vegan restaurants should not approach the Kurla area,they should focus on the areas of cluster 2 like Worli, Borivalli, Bhandup etc. These areas will have a zero competition and a total monopoly in the market can be achieved. Cluster 1 places can also be used for the business, since there is a little competition and if the food quality of yours is better, then there is a very nice scope. And further it is advised to avoid the Kurla area as there is already a high concentration of Vegan Restaurants existing there.

:

	Neighborhood	Vegetarian / Vegan Restaurant	Cluster Labels	Latitude	Longitude
21	Kurla	0.25	1	19.06498	72.88069

LIMITATIONS:

This project is limited to a certain extent because we haven't included the people's choice in a certain area, where there are more vegetarians. So we should also include these datas. Furthermore we should also include economic data to find in which area the restaurants can get many customers who can afford it. So, for further studies on it we should also include these datas also.

CONCLUSION:

In this project we have gone through identifying the business problem. Then extracting the data, preparing the data and then analyzing the data through several tools like pandas, k-means clustering and then visualising it with folium maps.

Finally we have found that places of cluster2 are best for opening the restaurant and the places of Cluster1 are the second best places for the vegan restaurants. This project is relevant and useful for those who want to open a new Vegan restaurant in Mumbai, India.

REFERENCES:

- WIKIPEDIA: https://en.wikipedia.org/wiki/Category:Suburbs_in_Mumbai.
- MUMBAI CITY: https://mumbaicity.gov.in/document-category/census/ .
- FOUR SQUARE: https://developers/foursquare.com .
- GEOCODERS: https://geocoder.readthedocs.io/ .

THANK YOU