



OPENING A BAKERY SHOP IN MUMBAI- INDIA

IBM APPLIED DATA SCIENCE CAPSTONE PROJECT



JUNE 1, 2020
SREYA SARKAR

Introduction

Bakeries are a popular type of foodservice establishment, and they allow us to express our culinary creativity while also serving a unique market. The business for selling baked products has been there for thousands of years. However, it has also evolved over the years due to the fame and desire that the art of baking has received. While grocery stores and supermarkets are keen on selling pre-sliced breads, cakes, and pastries, many still prefer to get their kind of flour baked from a small artisanal bakery. This may be because of taste or age-old household tradition. The recent move of commercialization of baked goods has increased the quality characteristic in the trade of baking. Therefore, the passion of baking is now catching up into the small and big bakery business.

People in India are fueling their passion for food through their open food business ventures. Along with the startup wave hitting the country, and a major disruption seen by the Restaurant Industry has encouraged a lot of people to start up their restaurant venture, especially women. Professionals and home cooks alike are now fuelling their passion for food and to get into the food business. Baking is one such passion that has been catching up lately, with many big and small bakery businesses cropping up. Opening a bakery presents many unique challenges that are different from other types of businesses.

The profit potential for a bakery is good, but starting a food business is never easy. A lot of things have to be taken care of. Opening a bakery presents many unique challenges that are different from other types of businesses. The total approximate cost of opening a bakery in India can be figured but the cost of equipment and location can lead to considerable variance in the approximate cost.

One of the first steps to how to start a bakery business in India is therefore to decide a good location. The location plays a huge role in its success. Although a great location may not guarantee success, a bad location will almost always guarantee failure.

- Look out for the competitors in the area. **Lesser the competition**, easier is the sales.
- The best location combines visibility, affordability and lease terms you can live with.

Business Problem

The objective of this Capstone project is to analyze and select the best location to opening a Bakery Business in the India's Financial capital, Mumbai. Using the data Science methodology and machine learning techniques like clustering, the aim is to provide answer to the question:

If a Restaurateur is planning to open a new Bakery business in Mumbai, where would you recommend it?

Target Audience

To solve this problem, data scientist team led by myself has been engaged by a Restaurateur. The objective is to locate and recommend to the management which neighborhood will be the best choice to start off their first bakery shop. The management also expects to understand the rationale of the recommendations in the report.

Success Criteria

The success criteria of this project will be a good recommendation of the neighborhood choice in Mumbai to the Restaurateur based on 2 key factors; lack of Bakery Shops available (less competition) and higher number of residences presented (higher demand).

Data

We would need the following data to analyze the opportunity:

- List of neighborhoods in Mumbai, India. This defines the scope of the project which is confined to the city of Mumbai.
- Latitude and Longitude coordinates of those neighborhoods. This will help us to plot the locations on the map and also get the venue data.
- Venue data, particularly Venue equals to Bakery. We will use this data set to perform Clustering.

Data Sources

We will use the Wikipedia page ("https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai") which contains list of neighborhoods in Mumbai which is close to 42. With the help of different Python Packages and using the scraping techniques, we will extract the data from the Wikipedia page. Post that, we will fetch the latitude and longitude coordinates of the neighborhoods using the geospatial data.

We will then use the Foursquare API to get the venue data for those neighborhoods. This API will provide us with many categories of venue data for those neighborhoods. As we are interested in Bakery, we will further filter out Bakery from the venue data and proceed with the analysis. This project will make use of few data science techniques like Web Scraping, Working with Foursquare API, data cleaning, data wrangling, Machine learning algorithm like K-Means clustering and Map visualization.

Going forward we will explain the methodology i.e. the steps taken in this project.

Methodology

Firstly, we need to get the list of neighborhoods in the city of Mumbai. The list is available in the Wikipedia page ("https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai"). We will do web scraping using Python requests and BeautifulSoup packages to extract the list of neighborhoods data. However, this is just a list of names. Next using the Geocoder package we will fetch the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. The Geocoder package allows us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into a pandas DataFrame and then visualize the neighborhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Mumbai.

Next, we will use Foursquare API to get the top 100 venues that are within a radius of 1000 meters. One should have a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key using which API calls are made to Foursquare passing in the geographical coordinates of the neighborhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many

venues were returned for each neighborhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyze each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering.

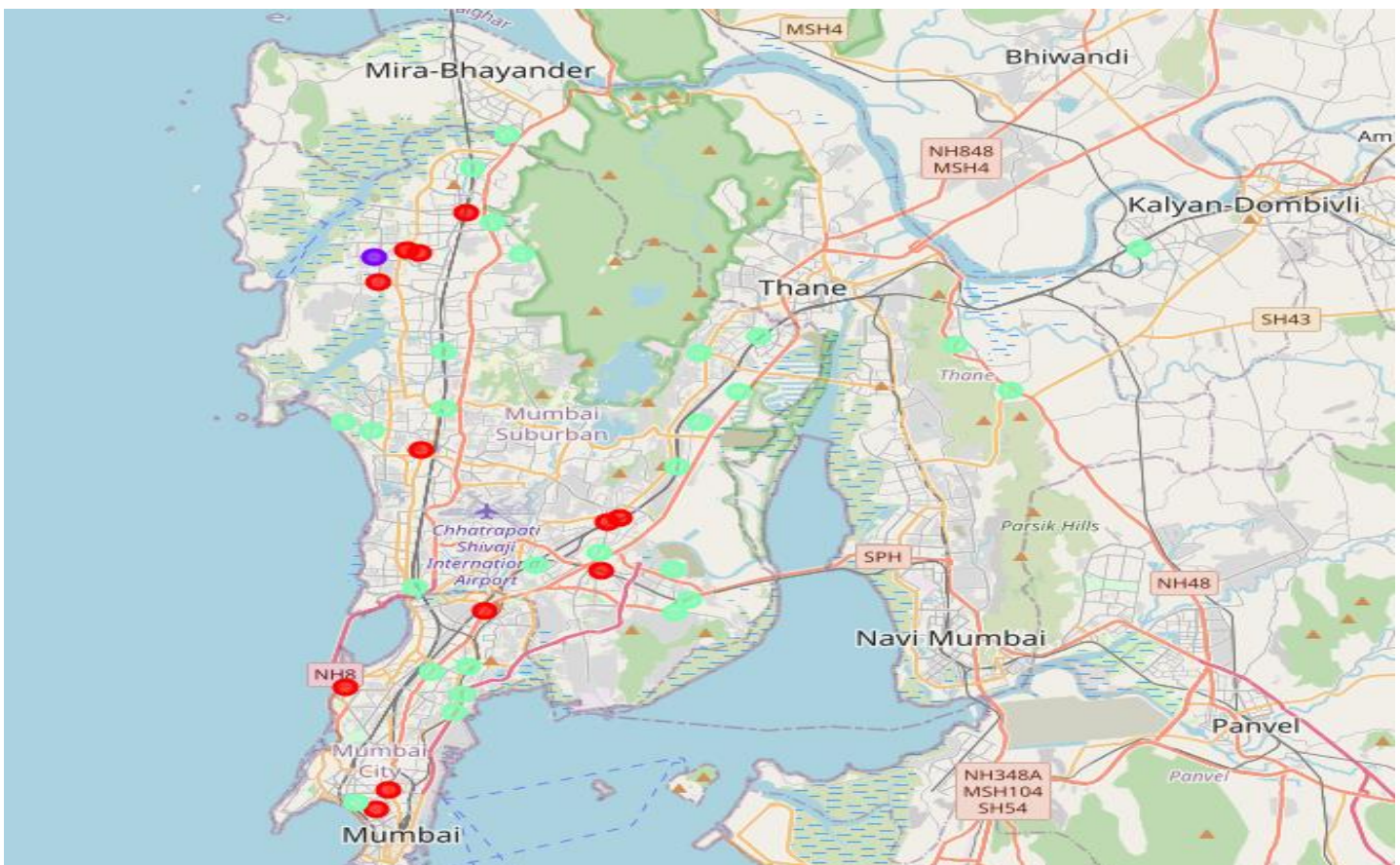
Since we will be analyzing the “Bakery” data, we will filter the “Bakery” as venue category for the neighborhoods.

Finally, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighborhoods into 3 clusters based on their frequency of occurrence for “Bakery”. The results will allow us to identify which neighborhoods have higher concentration of Bakery shops while which neighborhoods have fewer number of Bakery shops. Based on the occurrence of Bakery shops in different neighborhoods, it will help us to answer the question as to which neighborhoods are most suitable to open new Bakery shops.

Results

From the k-means clustering we see that, we can categorize the neighborhoods into 3 clusters based on the frequency of occurrence for “Bakery”:

The results of the clustering are visualized in the map below with cluster 0 in red color, cluster 1 in purple color, and cluster 2 in mint green color.



- Cluster 0: Neighborhoods with moderate number of Bakery shops**

	Neighborhood	Bakery	Cluster Labels	Latitude	Longitude
0	Andheri	0.027027	0	19.118459	72.841763
37	Vashi	0.026316	0	19.084650	72.904810
32	Sion, Mumbai	0.034483	0	19.043410	72.863320
24	Matharpacady, Mumbai	0.035714	0	18.950694	72.827268
22	Mahavir Nagar (Kandivali)	0.068966	0	19.210940	72.841370
40	Western Suburbs (Mumbai)	0.043478	0	19.197010	72.827680
18	Kandivali	0.058824	0	19.211982	72.837573
14	Grant Road	0.026316	0	18.959290	72.831080
12	Ghatkopar	0.027778	0	19.086523	72.909008
41	Worli	0.041667	0	19.007440	72.816880
7	Chembur	0.023810	0	19.062180	72.902410
5	Borivali	0.020408	0	19.229360	72.857510

- Cluster 1: Neighborhoods with low to no concentration of Bakery shops:**

	Neighborhood	Bakery	Cluster Labels	Latitude	Longitude
6	Charkop	0.125	1	19.20866	72.82612
36	Uttan	0.200	1	26.86634	80.93884

- Cluster 2: Neighborhoods with high concentration of Bakery shops:**

	Neighborhood	Bakery	Cluster Labels	Latitude	Longitude
29	Pestom sagar	0.000000	2	19.070640	72.902170
30	Seven Bungalows	0.014706	2	19.131342	72.816342
31	Shil Phata	0.000000	2	19.146580	73.040050
2	Baiganwadi	0.000000	2	19.062940	72.926630
33	Sonapur, Bhandup	0.000000	2	19.163940	72.935440
10	Dombivli	0.000000	2	19.212750	73.083240

	Neighborhood	Bakery	Cluster Labels	Latitude	Longitude
34	Thakur village	0.000000	2	19.210200	72.875410
35	Tilak Nagar (Mumbai)	0.000000	2	18.996160	72.852790
1	Anushakti Nagar	0.000000	2	19.042830	72.927340
38	Vikhroli	0.000000	2	19.111090	72.927810
39	Wadala	0.000000	2	19.017200	72.858170
28	Mumbra	0.000000	2	19.167632	73.021408
27	Mulund	0.000000	2	19.171830	72.955650
25	Mira Road	0.000000	2	19.265674	72.870681
9	Devipada	0.000000	2	19.224690	72.866050
3	Bandra	0.000000	2	19.054370	72.840170
23	Mankhurd	0.000000	2	19.048530	72.932220
4	Bhandup	0.000000	2	19.145560	72.948560
21	Kurla	0.000000	2	19.064980	72.880690
19	Kanjurmarg	0.000000	2	19.131380	72.935680
17	Kalyan	0.016393	2	18.953937	72.820367
16	Juhu	0.000000	2	19.014920	72.845220
15	Jogeshwari	0.000000	2	19.137920	72.849410
13	Goregaon	0.000000	2	19.164550	72.849460
8	Dahisar	0.000000	2	19.250030	72.859070
11	Eastern Suburbs (Mumbai)	0.000000	2	19.004270	72.855792
26	Mogra Village	0.000000	2	24.375900	75.954570
20	Kausa	0.000000	2	19.127580	72.825390

Discussion:

From the results section, most of the Bakery shops are concentrated in the so called Old city area of Mumbai, with the highest number in cluster 2 and moderate number in cluster 0. On the other hand, cluster 1 has very low number to no bakery shops in the neighborhoods. This would not be an ideal location to start a bakery shop as it's the New Mumbai area with comparatively lesser footfall. Cluster 0 represents a great opportunity and high potential areas to open new Baking shops as there is comparatively less competition from existing Bakeries. Meanwhile, Bakery shops in cluster 2 are likely suffering from intense competition due to oversupply and high concentration of bakery shops.

Therefore, this project recommends the Restaurateur to capitalize on these findings to open new Bakery shops in neighborhoods in cluster 0 with comparatively less competition. Restaurateur with unique selling propositions may want to enter the not explored area with new bakery shops in neighborhoods in cluster 1. Lastly, it is advised to avoid neighborhoods in cluster 2 which already have high concentration of bakery shops and suffering from intense competition.

Limitations and Suggestions for Future Research

In this project we have considered only one factor i.e. frequency of occurrence of shopping malls, there are other factors such as population and footfall that could influence the location decision of a new shopping mall. However, to the best knowledge of this researcher such data are not available to the neighbourhood level required by this project. Also Foursquare fee account was used to make the API calls to fetch data which comes with certain limitations.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and lastly providing recommendations to the restaurateurs regarding the best locations to open a new Bakery shop.

Therefore to answer the Business problem: The neighborhoods in cluster 0 are the most preferred locations to open a new bakery shop. The findings of this project will help to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new Bakery Shop.

References

Category:Suburbs in Mumbai. *Wikipedia*. Retrieved from ("https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai")

Foursquare Developers Documentation. *Foursquare*. Retrieved from https://developer.foursquare.com/docs