Prediction of New Opportunities for Bakery Businesses in Pune

Divya Vyas

27 May 2020

1. Introduction

1.1 Background

Baker's Hut is a popular Bakery shop in Mumbai, which was established in 1968.It is a leading food & beverage destination in the premium category of bakeries in Mumbai. It has successfully brought exclusive and indulgent brownies, cakes, desserts, pastries, chocolates, breads, savouries, sandwiches, rolls and beverages to the high street and within the reach of many and set a benchmark for delivering high quality food at honest prices. The bakery has been around for more than 50 years and has as many as 10 branches in the city of Mumbai alone.

After its success in Mumbai, the stakeholders are looking to explore opportunities in the nearest biggest city, Pune. Pune as a city has developed drastically in the past 20 years or so and is producing many new businesses opportunities in the food and beverages industry.

1.2 Problem

This project aims to find out which could be the potential areas for starting the bakery branches in the city of Pune, India. We would primarily look at areas which currently do not have sufficient bakeries but share characteristics with other areas that have.

We need to look at data that might contribute to understanding the food and lifestyle preferences of the audiences across the various neighborhoods in the city.

1.3 Interest

Looking at the potential of the city in the bakery products industry, the research could provide the useful information based on statistical analysis of data, Modelling and recommendation which could help Baker's Hut or any similar bakery looking to launch itself in the city of Pune.

2. Data acquisition and cleaning

2.1 Data sources

To get further details about the potential areas first I needed to get names and coordinates with the Postal Codes of all the different areas of Pune District.

1. The same details were taken from the below mention website :-

   https://geographic.org/streetview/india/maharashtra/pune/pune.html

2. . After acquiring the data from above website, it was found that the data of some of the possibly potential areas like "Koregaon Park, Boat Club Road, Fatima Nagar etc" were not available. The data for the same was acquired from google searches.
3. After getting all area locations, we needed to understand food and lifestyle preferences of the city. Hence, we got all venues details corresponding to the respective areas from FourSquare API.

2.2 Data cleaning

Pune district consists of a large geographical area. Many such locations fall in the small town or rural categories. The data set acquired consisted of all such areas as well.

As the category of our bakery does not quite appeal to the audiences of such locations, I needed to focus mainly on areas that fall in the main Pune city area.

The Postal Code of all the main Pune city areas begin with '411'. Hence, I filtered the dataset accordingly.

In order to easily take a glimpse at any particular area I sorted the list in Ascending order. After acquiring venues details from FourSquare I divided data into 2 datasets as follows

1. **All the areas wise Venues details (across all categories)** : This dataset would help us to understand location wise food and lifestyle preferences of the audiences.

2. **Area wise Bakery details**: this would help us to understand the distribution of bakery location in the main city areas so that we could identify locations that do not have sufficient bakeries but fall into the same cluster as the areas which have.

3. **Zone wise data:** A other way to understand data would be to divide the main city areas in to 5 zones. Mainly, “South, North, East, West and central”.

4. **Zone wise Bakery distribution:** The data was then further processed to see the bakery distribution in the above mention’s zones.

3.Exploratory Data Analysis

3.1 Zonal Data

Upon exploring data, according the zones for the bakery category. It was found that the bakeries in Pune were largely evenly distributed. However there seems to be more opportunities in the East and North zone compared to the others. This is evident in the graph below.

Central and South zone seems to be having a higher competition in bakeries.

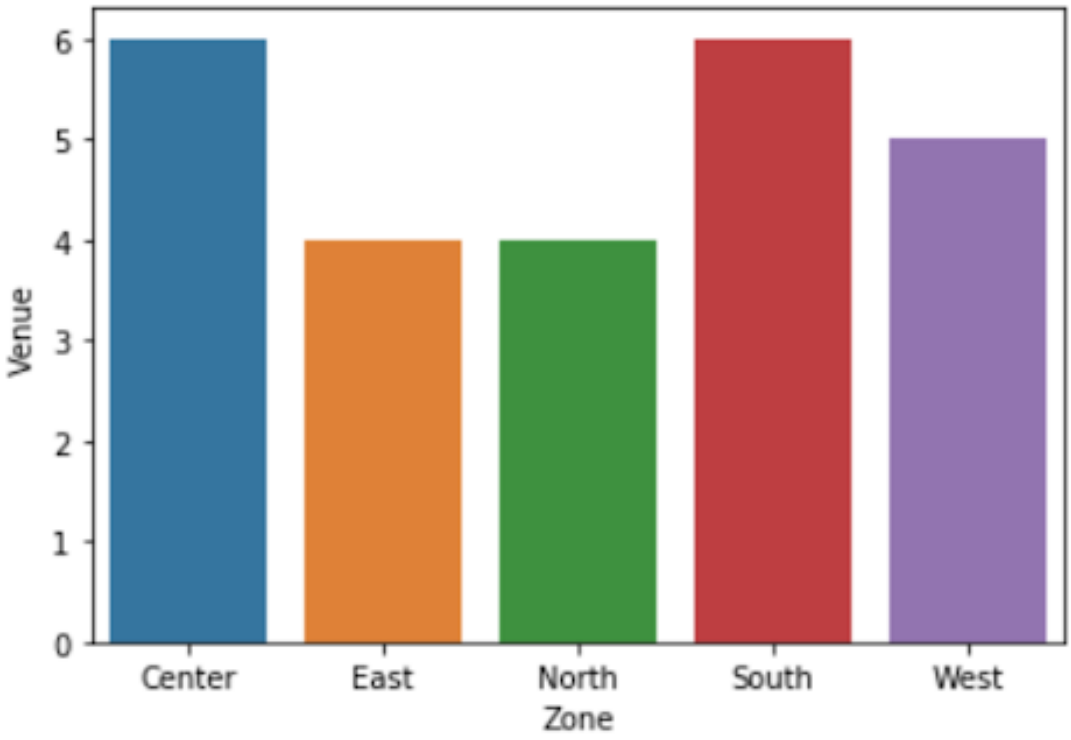

Figure.1 Zone wise venues distribution

3.2 Neighbourhood-wise Venue Distribution

This was to understand the category preferences of all the different neighbourhoods.

I took a count to check how many venues were distributed across different neighborhoods. I got a total of 143 unique categories.

I used “Onehot” encoding after which I took the “groupbymean” to view weightages of categories across neighbourhood.

I found out the top 5 venues in each of the neighborhoods. This would help me understand the neighborhoods better and also form a strong basis for input to a machine learning algorithm that would follow post data exploration

Finally, I put all this data into a data frame. The data now appears in a 2- dimensional format and the columns are arranged in the order of most preferred locations for each of the respective neighborhoods.

3.3 Heatmap for key categories and neighbourhoods

The data set of Pune City consist of large number of data. Further the data acquired for FourSqare also consisted of a large number of venues. In order to better understand the cultural and lifestyle preferences of people the large data set would certainly give us good insights about the audiences.

However, to explore and comprehend the data better, I decided to filter the data into a few key categories which primarily belong food industries. I also filtered only some of the key neighbourhoods in the city of Pune.

Further I decided to plot this on the heat map to draw some key inferences

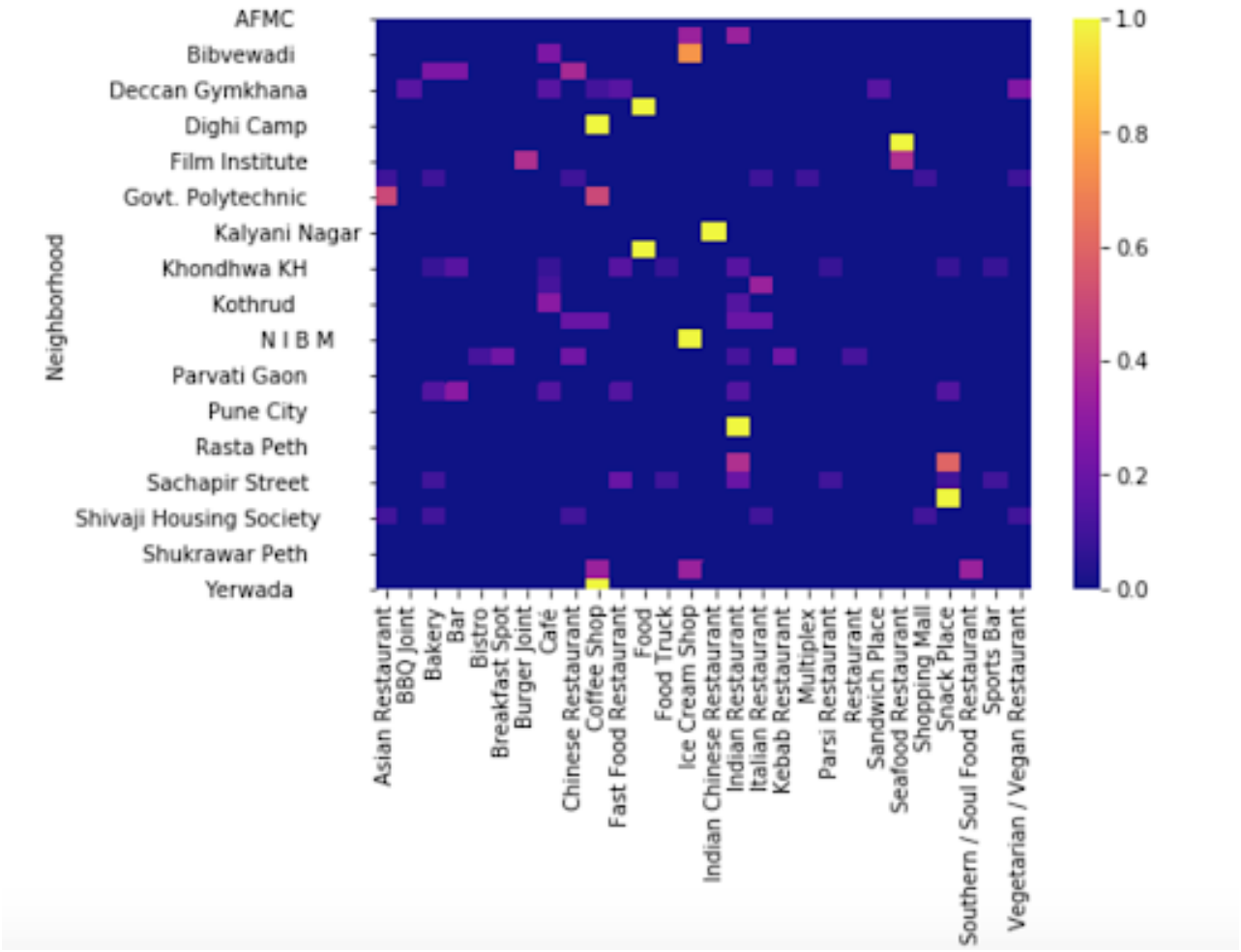

Figure.2 Heatmap for Venue Category and Neighborhoods

Inferences from Heatmap:

- Areas like Kalyani Nagar, Sadashiv Peth, Yerwada and NIBM seem to be bustling with food trucks, restaurants, bakeries, snack joints and cafes.
- Areas like Pune Cent East, Kondhwa and Boat Club Road seem to be having decent number of bakeries around their respective vicinities.

Machine Learning Algorithm:

The research uses k-means clustering algorithm.

Why k-means?

Since we are dealing with unsupervised data out of the three algorithm that we learned in machine learning either k-means or DBSCAN would be the best fit for our research I tried both of the methods and discovered that k-means was more suitable for our particular use case.

This could be because of the following reasons:

1. DBSCAN does not work well when there is a lot of difference in the densities of the clusters.
2. The parameters of our data sets consisted of a wide range and variety

Cluster and Category found as below:

Red circles showing cluster 0 areas

Purple circles showing cluster 1 areas

Sky Blue circles showing cluster 2 areas

Green circles showing cluster 3 areas

Orange circles in the map is showing the cluster 4 areas

Cluster 0 and Cluster 4 seemed to be possible candidates upon further analysis, cluster 4 was found to be strongest candidate.

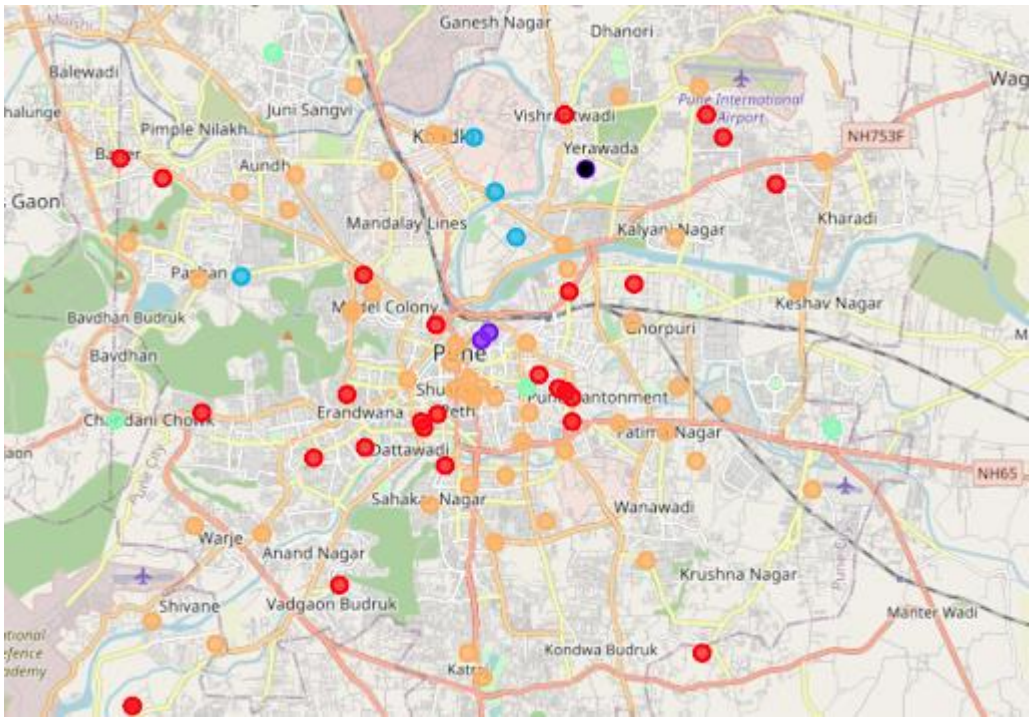

Figure.3 Clustering of all Venues

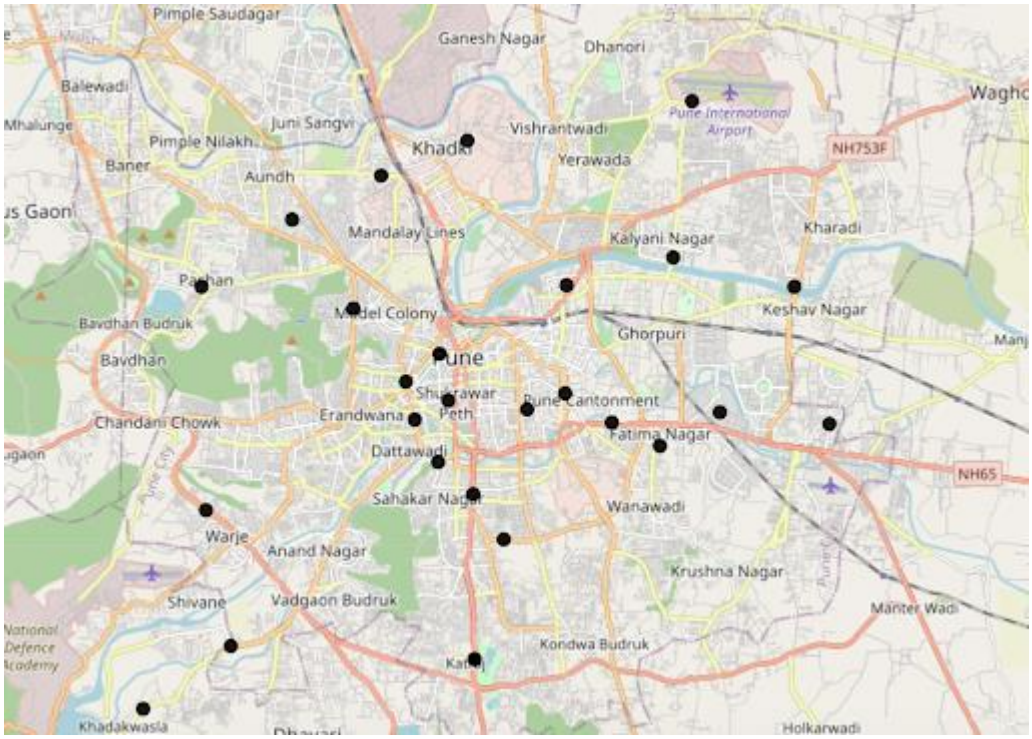

Figure.4 Potential Bakery Location Based on clustering:

Result:

The Machine Learning Algorithm categorises the data into five clusters(0,1,2,3,4) after analysing and examining the cluster, I found that cluster 4 locations were the closest to the type of location that match the audience preferences for bakeries.

Upon extracting the data comprising to Cluster 4 it was found that about 23 such areas in the main Pune city which belonged to Cluster 4 but did not have a single bakery in its vicinity.

Discussion:

Areas like Warje, Katraj and Khadakwasla fall in the main city but are a little on the outskirts of the city. They may not be the best areas to be considered even though they fall in the selected cluster.

Some of the pockets, in spite of being in the central zone of the city might not appeal to an audience looking for premium bakery products. Since our brand falls in the premium category we may not consider these as well.

We can see from the map that areas like Kalyani Nagar, Aundh, Model Colony and Keshav Nagar do not have sufficient bakeries but might be excellent locations for our bakery

Conclusion:

The above map shows areas which share similar characteristics to the areas which have good number of bakeries. However, would like you to note that our data is only based on Bakery category in general and does not involve any sub-categories of bakeries depending on product types, ratings, price, menu, premium or non-premium etc, due to lack of readily available data. Hence these recommendations may not be taken as final recommendations and only be taken as basis for further advanced analysis.