Prediction for New Opportunities in the Bakery Business in Pune

Divya Vyas

Introduction

Background

- Baker's hut, a popular bakery in Mumbai

- It is a premium bakery with a legacy of over 50 years of business in Mumbai with over 10 branches currently operating.

- Stakeholders are looking to explore opportunities in the city of Pune, India

- Pune as a city has developed drastically in the past 20 years or so

Problem

❖ This project aims to find out which could be the potential areas for starting the bakery

❖ 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

Data Acquisition And Cleaning

Data Sources

- Names and coordinates with the Postal Codes of all the different areas of Pune District were taken from geographic.org

- Some missing data was taken from Google searches

- Venue data to understand food and lifestyle preferences of audiences were taken using Foursquare API

Data Cleaning

- Many locations of Pune fall in the small town or rural categories

- 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.

Data Division

* All the areas wise Venues details (across all categories)

* Area wise Bakery details

* Zone wise data (North, South, East, West & Central)

* Zone wise Bakery distribution

Exploratory Data Analysis

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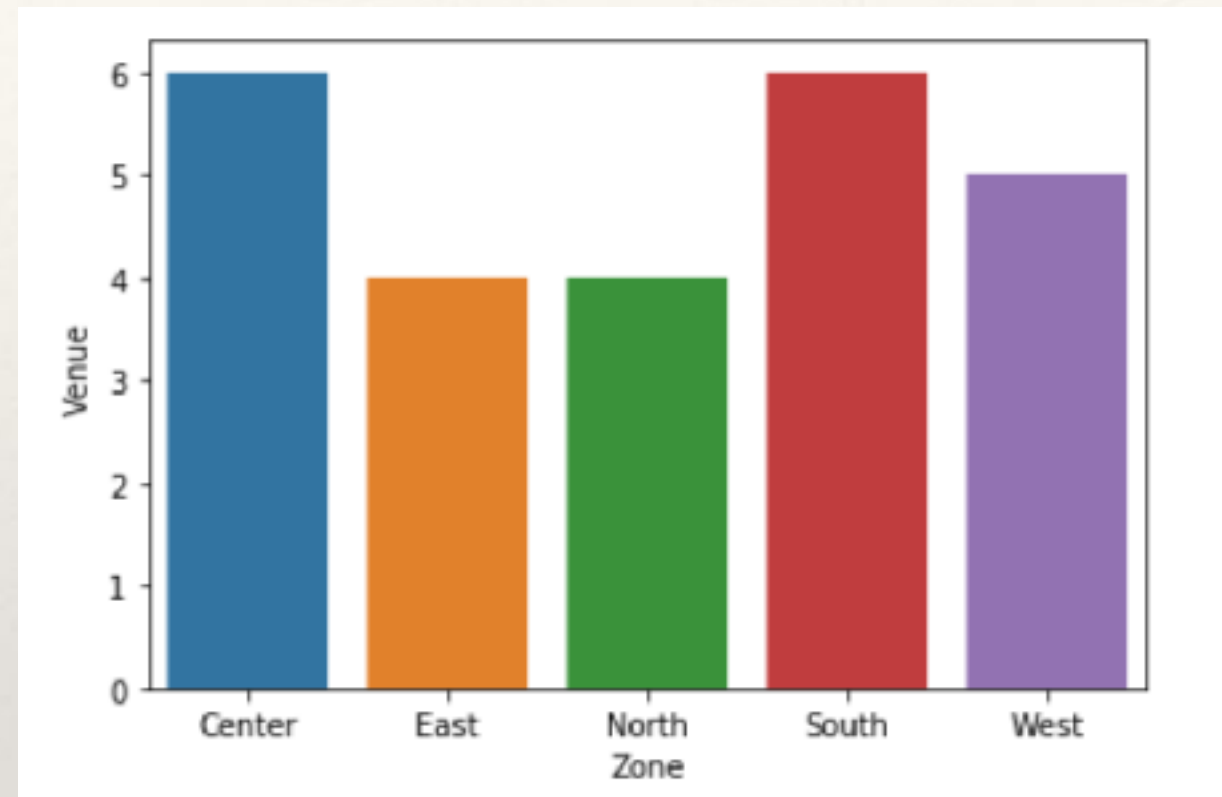
Upon exploring data, according the zones for the bakery category. It was found that the bakeries in Pune were largely evenly distributed. However there seems to be more opportunities in the East and North zone compared to the others. This is evident in the graph below.

Neighborhood and Venue-wise

❖ To understand the category preferences of all the different neighbourhoods

❖ A count to check how many venues were distributed across different neighborhoods. A total of 143 unique categories were found

❖ Top 5 venues in each of the neighborhoods were taken This would help understand the neighborhoods better and also form a strong basis for input to a machine learning algorithm that would follow post data exploration

Heatmap for Key Categories and Neighbourhoods

Heatmap Inferences

❖ 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.
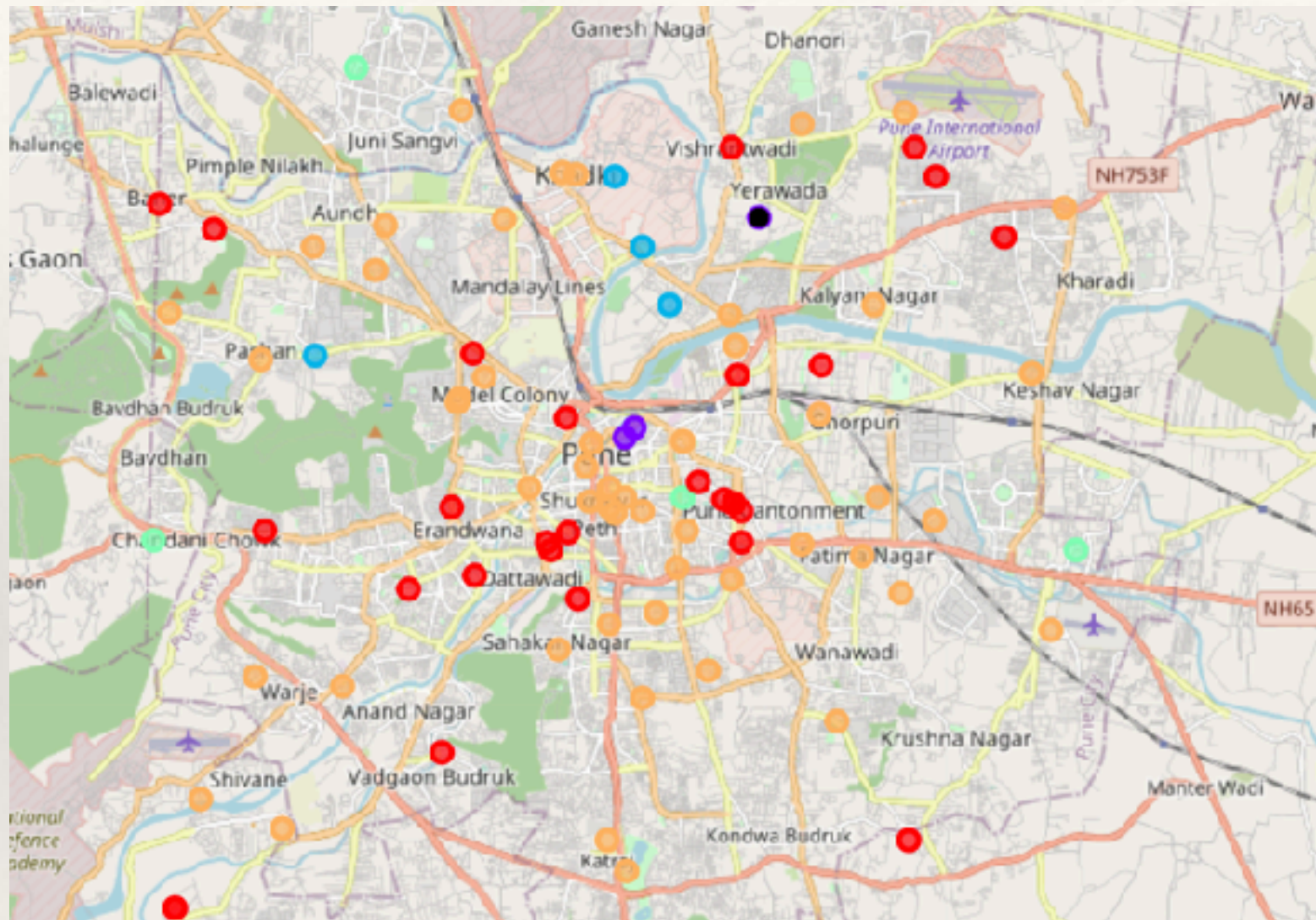
Machine Learning

Best Algorithm

- Since we are dealing with unsupervised data out of the three algorithm that we learned in machine learning either k-means or DBSCAN

- After trying the 2 algorithms, K-means was found to be more suitable for my project

Why K-means?

- DBSCAN does not work well when there is a lot of difference in the densities of the clusters

- The parameters of our data sets consisted of a wide range and variety

Cluster map

The orange clusters seemed to be having the highest scope for food joints, restaurants and bakeries as well

Results

❖ The Machine Learning Algorithm categorised the data into five clusters(0,1,2,3,4)

❖ After analysing and examining the cluster it was 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 viscinity

Recommendations

❖ 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, inspite 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

Recommendations (Contd..)

❖ 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 result shows areas which share similar characteristics to the areas which have good number of bakeries. However, I 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

"The distinction between the past, present and future is only a stubbornly persistent illusion."

–*Albert Einstein*