OPENING A NEW BAKERY IN MUMBAI, INDIA

Ameya Thosar

INTRODUCTION

- Bakeries are an important part of our life as on occasion of Birthdays, Anniversary, a normal day etc we often go to bakery for buying Cakes, cookies, desserts.
- Mumbai has many bakeries but when you want to start your own you need to look for a location where in neighborhood there are minimum number of bakeries.
- Each bakery typically offers breads (bagels, buns, rolls, biscuits and loaf breads), cookies, desserts (cakes, cheesecakes and pies), muffins, pizza, snack cakes, etc.
- So opening a new bakery is a good decision as people love to visit them.

BUSINESS PROBLEM

- The objective of this capstone project is to analyse and select the best location in Mumbai, India to open a new Bakery.
- Using Data Science Methodology and Machine Learning techniques like clustering, this project will provide a solution to the problem.
- In Mumbai, India if anyone wants to open a new Bakery we can suggest him where he can open it.
- Because when you are new in business its better that you have less competitors so that your business can grow at faster rate.

DATA SOURCE

To solve this problem we need following data:

- List of Neighborhood's in Mumbai.
- Latitude and Longitudes of those neighborhood's. This will be required to get plot the map and get venue data.
- Venue data particularly related to Bakery. We will use this data for clustering.
- The data is available on
  https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai
- In this there is only list of neighborhood's to which we will we have to append Latitude and Longitude.

METHODOLOGY

Steps followed In methodology:

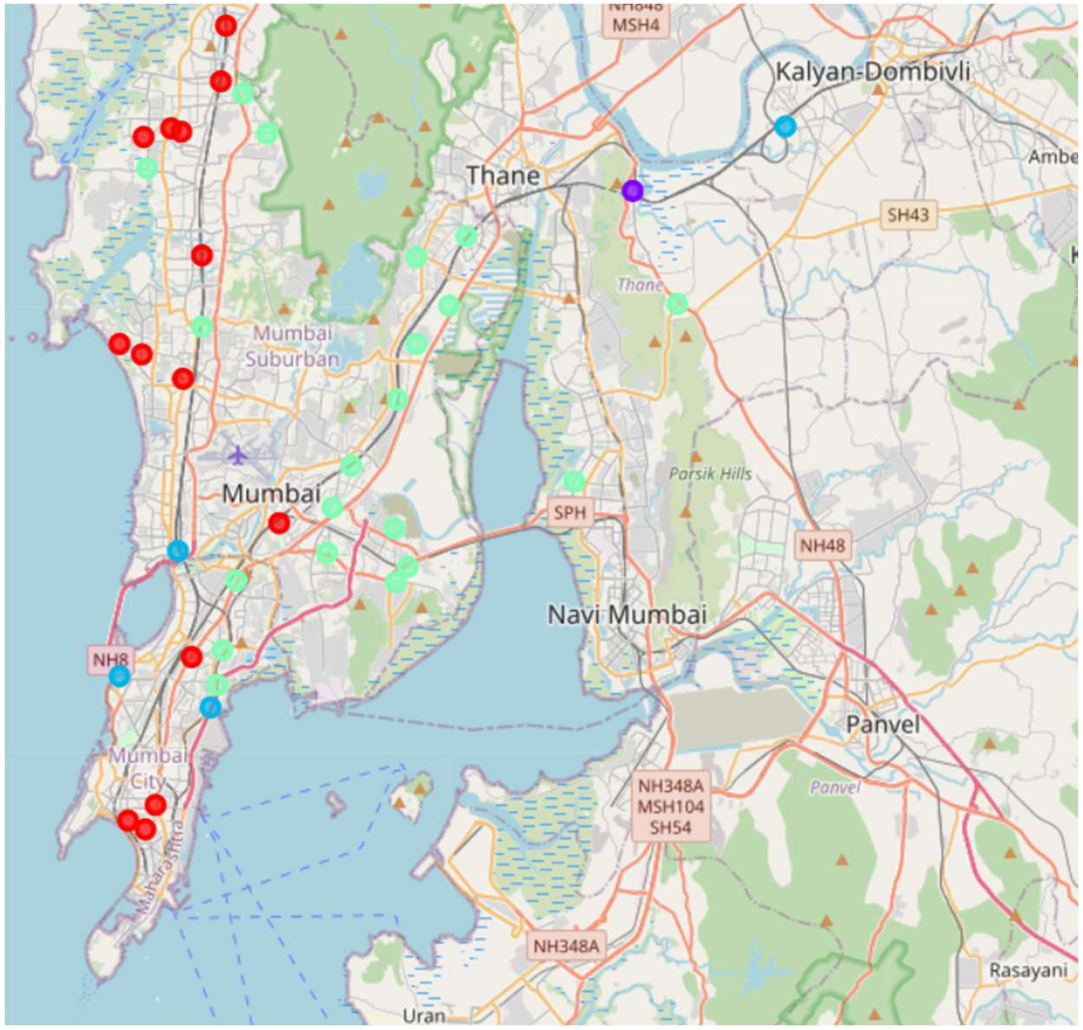
1. Import libraries
2. Scrap data from Wikipedia page into a Data Frame
3. Create a list to store neighbourhood data
4. Get the Geographical Coordinates i.e. Latitude and Longitude
5. Merge the coordinates into data frame
6. Create a map of Mumbai with neighborhoods superimposed on top
7. Use the Foursquare API to explore the neighborhoods
8. Get the top 100 venues that are within a radius of 2000 meters

9. Convert the venues list into the data frame.
10. Find out how many unique categories are there from the returned venues
11. Analyse Each Neighbourhood
12. Group rows by neighborhood and by taking the mean of the frequency of occurrence of each category
13. Create a new Data Frame for Bakery data only
14. Cluster Neighbourhoods
15. Visualize the resulting clusters
16. Examine Clusters

RESULTS

The results from the k-means clustering show that we can categorize the neighbourhoods into 5 clusters based on the frequency of occurrence for “Bakery”:

1. Cluster 0, 3 : Neighbourhoods with high number of shopping malls
2. Cluster 2 : Neighbourhoods with moderate number to no existence of shopping malls
3. Cluster 1,4 : Neighbourhoods with only one shopping malls


MSH4
Kalyan-Dombivli
Thane
SH43
Thane
Mumbai Suburban
Mumbai
Parsik Hills
SPH
NH48
Navi Mumbai
NH8
Panvel
Mumbai City
Panvel
NH348A
MSH104
SH54
NH348A
Rasayani
Uran

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 5 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. property developers and investors regarding the best locations to open a new Bakery.