Coursera Capstone

Opening a new Bakery in Pune

By: Ninad Khare

June 2020

Introduction:

My sister is good at baking and she recently told me that she would like to open her own bakery shop or a franchise. Hence her requirement was to find the best location to start her business in the city of Pune where we reside.

Objective:

The objective of this Capstone project is to find the best location in Pune, Maharashtra where there is little to no competition in order to open a new bakery or expand a franchise. This will help my sister to grow her business in the fastest way possible.

Data:

1.Getting different areas in Pune based on their pin code

2.Location data of these neighbourhoods

3.Venue data, related to Bakeries in order to perform Clustering

Sources:

This website:

http://zipcodepincode.com/India/Maharashtra/Pune/Pune-City/index.html

contains a postal code directory with all the areas in Pune.

We will use web scraping through pandas to get the data directly into a data frame.

Then we can use the Geocoder package in python to get the location data.

Finally, we will use the Foursquare API to get venues near our locations which are related to Bakeries

Project Report:

First and foremost is importing all the necessary libraries.

The next step is to extract data from the website and convert it to a Pandas data frame. This data frame consists of the Post code (Neighbourhoods) and Pin Code of the data.

The next step is to get the coordinates or location data of the areas in the data. Then cleaning it to get accurate results, this can be done by taking only the coordinates which are close to each other since it is the same city and also plotting the points on the map to remove any inaccuracies.

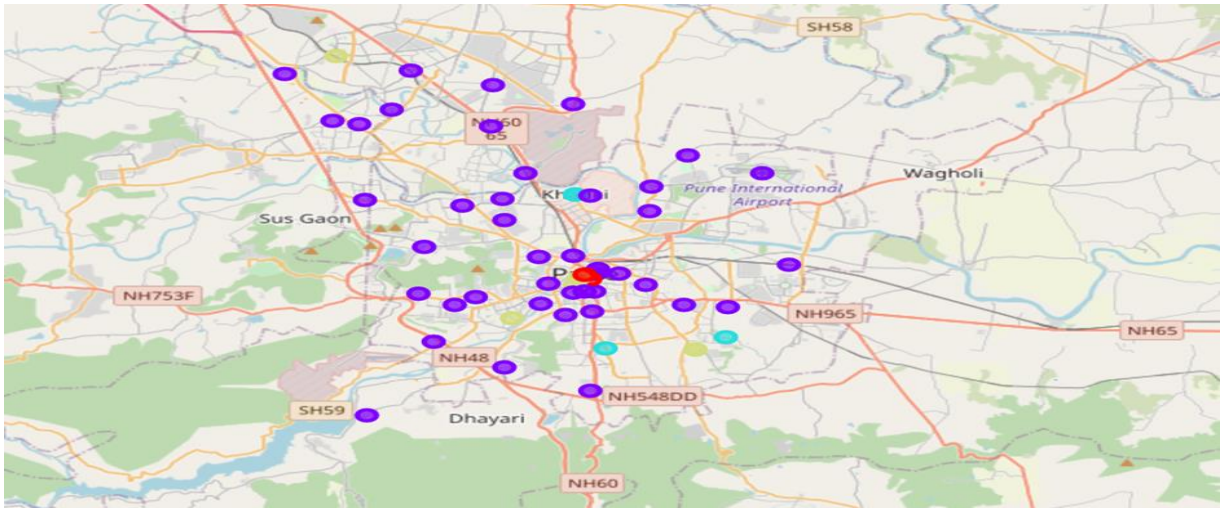
Then we get the Foursquare credentials required to access the API and explore the neighbourhoods using Foursquare.

We consider data relevant to our business problem and procced to form clusters in order to determine the best location.

We will use K means clustering to do this and we form 4 clusters. I took the value as 4 based on a quick look at the data frame which immediately suggests High concentration, Medium concentration, Low concentration and Little to no concentration of bakeries.

Finally, we plot our data to visualize our clusters and form better conclusions.

Results:



Here we can see that clusters containing bakeries are concentrated in the middle and southern part of the city, represented by red, yellow and light green.

Whereas the areas in blue are opportunities because these are areas where there is hardly any competition.

Further analysis:

As this project only determines prospective locations , this cannot be the only reliant information as further analysis is necessary in order to find the best location which can be done by analysing our identified cluster of neighbourhoods on multiple other factors and variables.

References:

Postal codes:

http://zipcodepincode.com/India/Maharashtra/Pune/Pune-City/index.html

Foursquare Developers Doc:

https://developer.foursquare.com/