Finding Best Places to open Chinese Restaurants in Pune

Ganisha Bhawsar

April 27, 2020

1. Introduction

1.1. Background

Pune has a mixed food culture. Being one of the major IT cities of India, Pune hosts a huge number of international firms with many employees and student recruits from all over the country. As a result, there's a huge demand of all kinds of outside food. Apart from the traditional Maharashtrian, Gujarati and North Karnataka cuisine, there's a good demand for international delicacies.

This report creates a hypothetical scenario for a concept Chinese restaurateur who wants to explore opening an authentic Chinese restaurant in Pune. The idea behind this project is that there may not be enough Chinese restaurants in Pune and it might present a great opportunity for this entrepreneur who is based in India. With the purpose in mind, finding the location to open such a restaurant is one of the most important decisions for this entrepreneur and this project would help her find the most suitable location.

1.2. Business Problem

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new Chinese restaurant in Pune, India. By using data science methods and machine learning methods such as clustering, this project aims to provide solutions to answer the business question: In Pune, if an entrepreneur wants to open a Chinese restaurant, where should they consider opening it?

1.3. Target Audience

The entrepreneur who wants to find the location to open authentic Chinese restaurant

2. Data acquisition and cleaning

2.1. Data Requirements

To solve this problem, the project uses below data:

- List of neighbourhoods in Pune, India.
- Latitude and Longitude of these neighbourhoods.
- Venue data related to Chinese restaurants. This will help us find the neighbourhoods that are most suitable to open a Chinese restaurant.

2.2. Data Sources and Extraction

- Dataset containing Pune neighbourhoods via geonames.org
- Getting Latitude and Longitude data of these neighbourhoods via Geocoder package
- Using Foursquare API to get venue data related to these neighbourhoods

2.3. Data Cleaning

Data downloaded from multiple sources were combined into one table. There were a lot of missing values, which were discarded.

2.4. Feature Selection

After data cleaning and merging different datasets, there were 917 samples and 10 features in the data.

(917, 10)

	Neighbourhood	Community	Chinese Restaurant	Cluster Labels	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
81	Lonavala	Maval	0.000000	0	18.7528	73.4057	Coopers Fudge	18.749842	73.406306	Dessert Shop
97	N.I.A.	Pune City	0.000000	0	18.5586	73.7794	Platinum Super Store	18.558777	73.777535	Convenience Store
97	N.I.A.	Pune City	0.000000	0	18.5586	73.7794	Manish	18.561910	73.778114	Spa
97	N.I.A.	Pune City	0.000000	0	18.5586	73.7794	Corporate Stay	18.562646	73.779912	Hotel
98	N.W. College	Pune City	0.000000	0	18.5372	73.8551	Shopper's Stop	18.534332	73.851900	Department Store
...	...	...	...	...	...	...	...	...	...	...
45	Film Institute	Pune City	0.066667	2	18.5148	73.8358	Old Canal Jogging Track	18.517328	73.831985	Track
37	Deccan Gymkhana	Pune City	0.066667	2	18.5148	73.8358	Casa Lolo	18.516930	73.833940	Deli / Bodega
1	A.R. Shala	Pune City	0.066667	2	18.5148	73.8358	Old Canal Jogging Track	18.517328	73.831985	Track
45	Film Institute	Pune City	0.066667	2	18.5148	73.8358	Deccan Gymkhana Club	18.515179	73.840528	Gym / Fitness Center
0	9 DRD	Pune City	0.062500	2	18.5685	73.9158	Natural Ice Cream	18.567612	73.917748	Ice Cream Shop

917 rows × 10 columns

3. Methodology

3.1. Exploratory Data Analysis

Based on the data extracted from all the sources, visualize the map of Pune using Folium package.

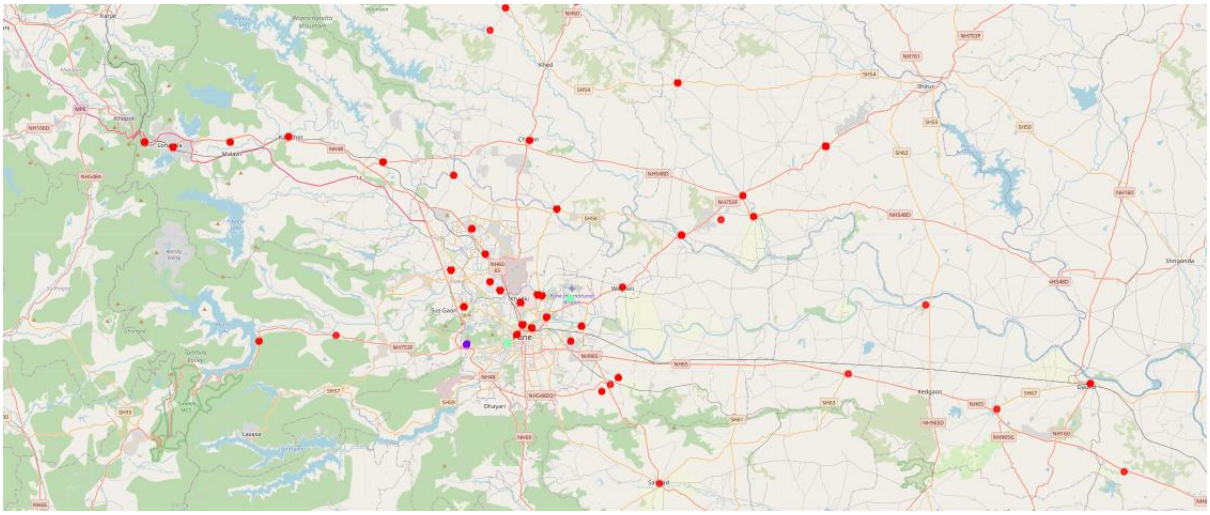
Next, I used Foursquare API to pull the list of top 100 venues within 500 meters radius. I have created a Foursquare developer account in order to obtain account ID and API key to pull the data. From Foursquare, I am able to pull the names, categories, latitude and longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues. Then, I analyze each neighbourhood by grouping the rows by neighbourhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later.

Here, I made a justification to specifically look for “Chinese restaurants”.

Lastly, I performed the clustering method by using k-means clustering. K-means clustering algorithm identifies k number of centriods, 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 it is highly suited for this project as well. I have clustered the neighbourhoods in Pune into 3 clusters based on their frequency of occurrence for “Chinese food”. Based on the results (the concentration of clusters), I will be able to recommend the ideal location to open the restaurant.

4. Results

Clusters:



The results from k-means clustering show that we can categorize Pune neighbourhoods into 3 clusters based on how many Chinese restaurants are in each neighbourhood:

- Cluster 0: Neighbourhoods with little or no Chinese restaurants
- Cluster 1: Neighbourhoods with few Chinese restaurants
- Cluster 2: Neighbourhoods with high number of Chinese restaurants

The results are visualized in the above map with Cluster 0 in red color, Cluster 1 in light green color and Cluster 2 in purple color.

5. 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 the machine learning by utilizing k-means clustering and providing recommendation to the stakeholder.

6. Recommendations

Most of Chinese restaurants are in Cluster 2 which is around Sus, Bavdhan and Armament areas and lowest (close to zero) in Cluster 0 areas which are Lonavala, Narayan Peth, Yerwada areas. Also, there are good opportunities to open near Vadgaon Sheri, Film Institute as the competition seems to be low. Looking at nearby venues, it seems Cluster 1 might be a good location as there are not a lot of Chinese restaurants in these areas. Therefore, this project recommends the entrepreneur to open a Chinese restaurant in these locations with little to no competition.

7. Limitations and Suggestions for Future Research

In this project, I only take into consideration of one factor: the occurrence / existence of Chinese restaurants in each neighbourhood. There are many factors that can be taken into consideration such as population density, income of residents, rent that could influence the decision to open a new restaurant. However, to put all these data into this project is not possible to do within a short time frame for this capstone project. Future research can take into consideration of these factors. In addition, I am relying on the existence of Chinese restaurants only for this project but future research can take into consideration of other variables such as existence of Asian restaurants, Asian population level in each neighbourhood etc.

8. References

1. List of neighbourhoods in Pune: https://download.geonames.org/export/zip/
2. Foursquare Developer Documentation: https://developer.foursquare.com/docs
3. All codes for this project can be found here: https://github.com/GB1111/Coursera_Capstone/blob/master/Pune_Chinese_Restaurants_final.ipynb