

Capstone Project – The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

Analysis and clustering of Chinese restaurants in the neighborhoods of Delhi, India

Rohan Karthikeyan

August, 25, 2020

1. Introduction: Business Problem

Delhi, officially known as the National Capital Territory of Delhi (NCT) is a city and union territory of India containing New Delhi, the capital of India. According to the 2011 Indian census, the population of just the city of Delhi was over 11 million, the second-highest in India after Mumbai. The city is of great historical significance as an important commercial, transport, and cultural hub, as well as the political center of India.

As a resident of Delhi for some time now, it would be an understatement to say that the people of Delhi love street food and enjoy other cuisines. The dearth of food habits among the city's residents have created a unique style of cooking which has become popular world over, such as Kebab, biryani and tandoori. But, talking about international cuisines, there may be no doubt that Chinese cuisine is the favorite among so many Delhiites.

There is also no overlooking the fact that Delhi is home to so many educational institutions of repute such as the Indian Institute of Technology, Delhi University, among many others. This attracts a lot of students from Indian cities as well as from other countries, including students from China. Consequently, Chinese restaurants are of huge demand especially in areas which has a larger concentration of Chinese populations.

We will try to analyze and select the best locations in Delhi to open a new Chinese restaurant. Using data science methodology and tools such as data analysis and visualization, we will try to provide the answer to these business questions:

- What is / are the best locations for opening a Chinese restaurant in Delhi?
- In what neighborhood should we open our restaurant to have better chances of success?

2. Data acquisition and cleaning

Based on the definition of our problem, we will need the following data:

- All the neighborhoods of Delhi
- Venue data, particularly related to restaurants. This data will be used to perform further analysis of the neighborhoods.
- Latitude and longitude of the neighborhoods

2.1 Data sources

Following data sources were needed to extract/generate the required information:

- Geo coordinates of Delhi are obtained using Geopy Nominatim
- Number of restaurants and their type and location in every neighborhood is obtained using the Foursquare API
- List of all neighborhoods of Delhi is obtained by web scraping the page: [Neighborhoods of Delhi](#)

2.2 Data selection

The Get Venue Explore endpoint of Foursquare API is used to retrieve the list of all venues along with the location and other details. This helps us to identify the areas with more footfall and then explore if there are any Chinese restaurants nearby.

One could also first collect the list of all Chinese restaurants (however, this uses the Get Venue Search endpoint of Foursquare API) in the neighborhoods of Delhi and then analyze the locations, cluster them and find out the areas where there is a lesser concentration of restaurants.