IBM

Coursera Data Science Specialization

Delhi Neighbourhood and Restaurants Analysis

Capstone Project

By

Simran Verma

May 2020

Introduction

Food is something that has brought communities, cultures, countries and people together for as long as we have existed as a human civilization. Now with the rise in globalization, restaurants from all over the world can be found in all major metropolitan cities of the world. From United States to Japan, Russia to Mexico, you can get all type of cuisines in the **capital of India, New Delhi**, which is the choice of city I have made. I have had the opportunity to spend multiple summers in Delhi with my grandparents throughout my childhood, and I speak from experience, it is one of the most wholesome and beautiful cities I have been to.

Coming back to food, Delivery, Dine-out, Pubs, Bars, Drinks, Buffet, Desserts you name it and Delhi will have a place. Delhi is a great place for foodies and also for businesses around dining and food. The number of restaurants is increasing day by day. Being one of the most densely populated cities in the world, there are customers for all kind of food joints, after all food is a basic necessity. The industry hasn't been saturated yet. Moreover, restaurants in Delhi are not just a sign for food businesses but also give the most prime locations for other businesses in Delhi too.

However, it has become difficult for them to compete with already established restaurants and hangout spots. The key issues that continue to pose a challenge to them include high real estate costs, rising food costs, shortage of quality manpower, fragmented supply chain and over-licensing.

With this project I aim at analysing demography, its impact on food culture and finding the most popular places in Delhi to open up a new consumer centric business around dining. Most importantly it will help new restaurants in deciding the theme, menus, cuisine, cost etc for a particular location, based on the observed trends.

It also aims at finding similarity between neighbourhoods of Delhi on the basis of food. With the analysis the project also will help people in choosing the restaurant based on several other factors. The project will mainly try to answer the question based on restaurants. And what factors should be kept in mind if someone wants to open new restaurant.

• Does the demography of an area matters?

• Does location of a particular type of restaurant also depends on the people living in that area?

• Does the theme of the restaurant matter?

• Are any neighbourhoods similar based on the type of food?

• Is a particular neighbourhood famous for certain kind of food?

• If two neighbourhoods are similar does that mean these are related or particular group of people live in the neighbourhood or these are the places to it?

• What kind of a food is more popular in a locality?

Data Description

The main data used for this project were from two sources:

• A Kaggle repository with neighbourhood and restaurants data of Delhi used for preliminary analysis

https://www.kaggle.com/shaswatd673/delhi-neighborhood-data/version/1

• Explore trending venues in a neighbourhood particularly restaurant (Foursquare API).

Other additional data:

• Coordinates (Geocoder Python)

• Population data for Delhi neighbourhoods to support analysis

https://www.census2011.co.in/census/state/districtlist/delhi.html

Data Collection Process

The Open dataset from Kaggle was downloaded as csv files and cleaned, and for reliability purposes verified against data from Wikipedia.

Foursquare API was used to obtain the nearby venues for each of the locations shortlisted by the preliminary analysis and plotting done on the Kaggle dataset. The location coordinates were obtained for those neighbourhoods using geopy module and then restaurants were filtered from the venues returned by the Foursquare API.

Please note that the preliminary analysis was not sufficient as the data obtained from that was limited and incomplete. However, it was essential to reduce the load on data obtained using Foursquare API and hence the data from the Preliminary stage, speeded up the entire process.