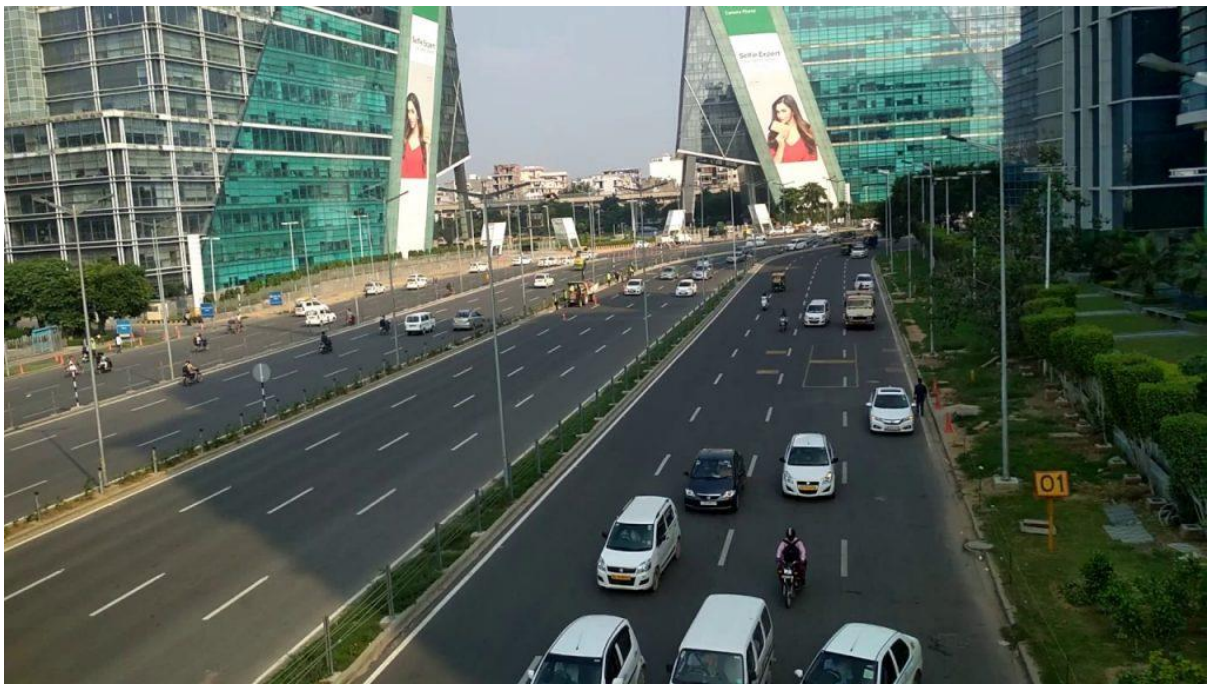
IBM Applied Data Science Capstone

'Exploring the Food Culture and Diversity of Gurgaon, India'

By: Karan Sud

May 2020



Introduction

Gurgaon, officially named Gurugram, is a city located in the Northern Indian state of Haryana with an estimated population of about 10 million as of 2020. It is situated about 30 kilometres southwest of the national capital New Delhi and is one of the leading Metropolitan cities of India, also sometimes referred to as the ‘Job capital of North India’. Many people shift to this city for better job opportunities and hence, it becomes necessary to explore the Food culture and diversity of this city.

Business Problem

Undoubtedly, Food Diversity is an important part of an ethnically diverse metropolis. The idea of this project is to categorically segment the neighbourhoods of Gurgaon city into major clusters and examine their cuisines. A desirable intention is to examine the neighbourhood cluster’s food habits and taste. This project will help to understand the diversity of a neighbourhood by leveraging venue data from Foursquare’s ‘Places API’ and ‘k-means clustering’ unsupervised machine learning algorithm. Exploratory Data Analysis (EDA) will help to discover further about the culture and diversity of the neighbourhood.

Target Audience

This quantifiable analysis can be used to understand the distribution of different food cultures and cuisines over ‘Job capital of North India’— Gurgaon, which can help any individual migrating to the city for work or education. Also, it can be utilized by a new food vendor who is willing to open his or her restaurant or by investors looking to invest in the food and beverage industry in the city.

Data Collection

To solve the problem, we will need the following data:

• List of neighbourhoods in Gurgaon: This defines the scope of this project which is confined to the city of Gurgaon, India.

• Latitude and longitude coordinates of those neighbourhoods: This is required in order to plot the map and also to get the venue data.

• Venue data, particularly data related to restaurants and cuisines: We will use this data to perform clustering on the neighbourhoods. To begin with, we will take a look at Gurgaon on the Map using the folium library.

We will also fetch the data from two different APIs.

- Foursquare API: We will use the Foursquare API to fetch venues in Gurgaon starting from the middle up to 10 Kilometres in each direction.
- Zomato API: The Zomato API provides information about various venues including the food cuisines, user and restaurant ratings, price range and a lot more.

Data Source and Extraction

The page (https://www.mapsofindia.com/pincode/india/haryana/gurgaon/) contains a list of neighbourhoods in Gurgaon, with a total of 166. We will use web scraping techniques to extract the data from the page, with the help of Python requests and beautifulsoup package. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use Foursquare API to get the venue data for those neighbourhoods in order to help us to solve the business problem put forward and the Zomato API to get data on food cuisines and restaurants. Both these APIs are free and easy to use with a limitation on the number of calls. This is a project that will make use of many data science skills, from web scraping, working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).