Analyzing New Delhi for Opening a New Business

Abhiroop Santra

July 5, 2020

1. Introduction

1.1 Background

For this project, we assume that our client Acme Corp. is a large conglomerate with successful businesses running across several sectors. They have hotel chains, restaurants, shopping malls and fitness centers. Acme Corp. intends to expand its business in New Delhi, India. As India is one of the most populous countries in the world, hence New Delhi, the capital city of India can be an attractive spot to open a new business. This will give Acme Corp. access to a large customer base. Also, this city can serve as a good foothold from where it can expand to the rest of the country.

1.2 Problem

New Delhi is a vast city with several localities. It is crucial for Acme Corp. to choose the right locality to establish a particular business. If our client opens a certain business where the demand for such services will be low then it is unlikely the business venture will succeed. Hence we need to determine, a good location for our client to open their business. This project aims to answer this question for our client and help them decide which locations are more suitable for a certain type of business.

1.3 Interest

This analysis is essential for our client as opening a business in the ideal location will not only increase profits but will also allow them to expand in the country in the coming years.

2. Data Acquisition and Cleaning

2.1 Data Sources

We will require the data on the various localities of New Delhi. We will need, at least, the names of all the localities that we are planning to analyze. Fortunately, there is a data set available from Kaggle which can be downloaded from here. This data is from 2018, but as the locality details of New Delhi is unlikely to change in this period of time, so we can consider it accurate.

2.2 Data Cleaning

The data that we have obtained contains information of several cities. So first, we need to filter the data only for New Delhi. Once this is done, we will end up with all the listed restaurants in New Delhi from 2018.

Next, we will be selecting the names of all the localities provided in the filtered data. As this data set is a bit old, so the restaurant details are probably not up-to-date. Hence we will ignore the restaurant data, which we will obtain later from Foursquare API. For now, we will simply select the various localities available.

After this step, we need to drop any row that does not have locality data as we will not be able to conclude anything from these rows.

The next step would be to remove any duplicate locality names from the data. This will also reduce the number of rows to work with. Once this is accomplished, we will have a complete list of all localities in New Delhi.

However, we are still missing an important information, that is the coordinates (latitude and longitude) of each locality. We will see how we can obtain this in the next section.

2.3 Fetching the Missing Details

We will be using the Python Geocode library to get the coordinates of each locality. There is a small drawback to this data. As the locality information is not supplied officially by the New Delhi Municipality, it might have some incorrect locality names. In this case, we will be ignoring these incorrect localities as rectifying the locality names is beyond the scope of this project.

We will loop through our data and assign each locality its respective coordinates. Once this is done, we will have the data with which we can start the analysis.