IBM – Coursera

Data Science Specialization

Capstone project - Final report

The Battle of Neighbourhoods – Geolocating ideal venues for Indian Restaurants in San Diego

Divyaprakash Dhurandhar

1. Introduction

Think of India, and one of the first things that come to mind is its diversity. A large country, its population is second only to China. Its languages are numerous, and every state (of which there are 28 and seven Union territories) is unique in its traditions and, very importantly, its food. Food from one region may be alien to a person from another province! The common thread that runs through most Indian food is the use of numerous spices to create flavor and aroma.

Indian restaurants have come a long way from the mid-1960s when the first significant wave of immigrants arrived. While still not entirely as assimilated as Italian and Mexican, Indian food is one of the fastest-growing segments in the culinary scene and is gaining popularity within the American mainstream. And as Indians have spread throughout America, so has their food. Restaurants with cult-like followings are no longer limited to New York City.

2. Business Problem

Concerning the growing demand for Indian food and restaurants across the United States, the need for intelligent business solutions regarding opening Indian restaurants arises. For this project, we assume that a client in the city of San Diego wants to open an Indian restaurant. To maximize revenue and business success, insights into existing Indian restaurants and peer competition are required. Thus, the project's main objective is to find ideal spots in the city where Indian restaurants can be set up.

3. Data

This project's data has been retrieved and processed through multiple sources, giving careful considerations to the accuracy of the methods used.

3.1. Neighborhoods

The data of the neighborhoods in San Diego can be extracted out by web scraping using the BeautifulSoup library for Python. The neighborhood data is scraped from a Wikipedia webpage https://en.wikipedia.org/wiki/Category:Neighborhoods_in_San_Diego.

3.2. Geocoding

The file contents from San-Diego.csv is retrieved into a Pandas DataFrame. The latitude and longitude of the neighborhoods are retrieved using the geocoder API. The geometric location values are then stored in the initial DataFrame.

3.3. Venue Data

From the location data obtained after Web Scraping and Geocoding, the venue data is found out by passing in the required parameters to the FourSquare API, and creating another DataFrame to contain all the venue details along with the respective neighborhoods.