Coursera Capstone

IBM Applied Data Science Capstone

“The Battle of Neighborhoods”

Finding a location to open an Indian restaurant in Manhattan New York

By: Ankur Mehta
December 2020

1. Introduction:

The United States of America is the country with one of the largest immigrant population in the world. After the IT boom , this immigration in recent times has been driven by IT workers from Asia coming into the states. A large number of them belong to India which due to its high population and educational system is able to provide one of the cheapest most talented manpower for IT firms. Indians flock by the 1000's to America chasing the American Dream. This has led to a growth of new markets being created that caters to the immigrated folk. One of the most lucrative ones of them being the Indian restaurant business. Indian cooking is known across the world for its diverse flavours and spices . Both Indians and American Nationals alike have taken a great fondness to the Indian Cuisine. This has led to a spurt in the Indian restaurant business. These businesses have specifically targeted places with large Indian population earlier but due to market saturation have started moving away to other parts.

In this project we will try to find an optimal location for an India restaurant. Specifically, this report will be targeted to stakeholders interested in opening an Indian restaurant in Manhattan , New York, USA.

Since there are lots of restaurants in New York we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Indian restaurants in vicinity.

We will use our data science powers to generate a few most promising neighbourhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

2. Problem Statement:

The objective of this capstone project will be to find a suitable location to open an Indian Restaurant in Manhattan New York that would have the most chance of being successful by leveraging Data Science and Machine Learning (k - means). The main Business question that will be answered in the Capstone Project will be : “Which neighbourhoods in Manhattan New York are best suitable for an Indian Restaurant ?”

3. Stakeholders/Target Audience

The Capstone Project will be particularly useful for people looking to open a restaurant in a given area. The project can also be modified to go beyond just the restaurant scope as it can be used to scope out other businesses in the area be it gyms , schools etc . The project can be used in an advisory capability by property consultants, realtors etc who can use the project to give their customer an overview of the area and allow them to make a better informed decision. With basic knowledge, customers themselves can use the this project to better understand the option available with them to make informed business decisions thus improving their chances of being successful.

4. Foursquare API

The project largely relies on utilising the Foursquare API, mainly the Places API to gather data related to locations. Foursquare is a location technology platform that allows the user to access its upto date database through an API to provide details of location/ venues the user might be interested in. The details include Name, Category, Location (latitude, longitude) , Ratings , reviews , menu etc as per the customer needs. We will be leveraging the API in this project to find out the venues that are present in the Manhattan Area to look for a suitable place to open our Indian Restaurant.

5. Data

Data used in this project is the New York dataset and was sourced from: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

The coordinates of places if and when required can be sought by using geopy. Geopy is a Python client for several popular geocoding web services. Geopy makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources

Finally Foursquare API will be used for identifying and analysing areas of interests which basically involves using the API to gather the following details –

a. Number of venues in a particular area based on the radius provided by the user based on neighbourhood details.
b. Name of the venue
c. Category of venues (Restaurants, gyms etc)
d. Location of the Venue (Latitude , Longitude)