

Coursera Capstone
IBM Applied Data Science Capstone
Opening a New Indian restaurant in San Diego, California

Introduction

For many residents in San Diego, visiting Indian restaurants is a great way to relax and enjoy themselves or even to grab a gift for a friend. They have a wide variety of desserts to choose from. Some Indian restaurants are like a one-stop destination for all types of foodies. For Indian restaurant owners, the central location and the large crowds near Indian Restaurants provide a great distribution channel to market their products and services. Business developers are also taking advantage of this trend and are building more bars to cater to the demand. As a result, there are many Indian restaurants in San Diego and many more are being built. Opening a Indian restaurant allows the business owner to earn a consistent income. Of course, as with any business decision, opening a new location requires serious consideration and is a lot more complicated with a lot of moving parts. Particularly, the location of the Indian restaurant is one of the most important decisions that will determine whether the business will be a success or a failure.

Business Problem

The objective of this capstone project is to analyze and select the best locations in San Diego California to open a new Indian restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In San Diego California if someone is looking to open a new Indian restaurant, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful to business owners and investors looking to open or invest in new Indian restaurant in San Diego California. This project is timely as San Diego is saturated with Indian restaurants. Data from San Diego .gov showed that Indian restaurants are expected to grow by an additional 8 percent.

Data

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

- List of neighborhoods in San Diego. This defines the scope of this project, which is confined to the city of San Diego, which is in California.

- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and to get the venue data.
- Venue data, particularly data related to Indian restaurants. We will use this data to perform clustering on the neighborhoods.

Sources of data and methods to extract them

This wiki page (https://en.wikipedia.org/wiki/Category:Neighborhoods_in_San_Diego) contains a list of neighborhoods in San Diego, with a total of 170 neighborhoods. I will be using web scraping techniques to extract the data from the wiki page, using Python requests and beautifulsoup packages. For the San Diego neighborhoods I will get the geographical coordinates of the neighborhoods using the Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods. After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. The Foursquare API will provide many categories of the venue data, we are particularly interested in the desert bar category which will help us solve the business problem

Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Indian restaurant in San Diego, Californ

stated above. This is a project that will make use of the data science skills covered in this course. Skills such as working with APIs (Foursquare), data cleaning, data wrangling, machine learning (K-means clustering) and visualizing maps with folium. In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning techniques that were used.