



IBM Applied Data Science Capstone Project

OPENING AN INDIAN RESTAURANT IN WASHINGTON D.C.

Sidhant Mishra | Coursera Capstone Project | 07/22/2021

Introduction

- **Background**

There are tons of factors to be considered when you are planning to open a new restaurant out of which location is one of the major factors. Whether you're opening your first full-service restaurant, your second, or your 50th, it's important to understand what to look out for when choosing a new restaurant location. It is of utmost importance to determine the most strategic location in order to attract customers and maximize profit.

- **Business Problem**

Let us choose a hypothetical situation:

A client having little to no experience in the restaurant and hospitality business seeks to open a restaurant specializing in Indian cuisine in Washington D.C. area. Being from the DC area himself, he kind of understands the demand which Indian food has and certainly believes that given the right strategic location, he could make it a successful one.

Taking into account the price level at which the restaurant will operate, the intent is to find an optimal location in the DC neighborhood, where there is significant demand for other cuisines and which already has a good footfall that can be capitalized.

The objective of this capstone project is to locate the optimal neighborhood for operation using unsupervised machine learning techniques. We would be using the data available on the internet to extract all neighborhoods of Washington D.C and then use Foursquare API to get an idea of different places of interest in each neighborhood.

- **Target Audience**

Entrepreneurs seeking to establish a new restaurant of a certain niche would be able to get an idea of how choosing an optimal location would give them a competitive advantage and help stay ahead in the game.

Data Acquisition, Cleaning and Methodology

- **Data Sources**

To perform this analysis, the following set of data would be required

1. List of Washington D.C. neighborhoods.
2. Geo coordinates of all the neighborhoods.
3. Top venues in each of the neighborhoods.

The list of neighborhoods can be scraped from [here](#). Geographical coordinates for each neighborhood can be obtained using the geocoder tool in the notebook. Data pertaining to top venues would be retrieved using Foursquare API. One has to register for a Foursquare developer account [here](#) to access their API credentials.

- **Data Cleaning and Methodology**

Data downloaded from the web page would be stored in a data frame. We would only be utilizing data that do not contain any junk or unnecessary values that would create problems going forward. In order to do so, we would need to make the data ready for analysis.

1. Only the cells that have an assigned neighborhood would be processed and the rest would be ignored.
2. Sometimes the geocoder package fails to retrieve geo coordinates of a neighborhood. In that case, the neighborhoods that would not have a corresponding geo coordinate would be ignored from any further analysis going forward.

Once the data is available for use, we would be exploring the neighborhoods using Foursquare API and then use machine learning technique (K- means clustering) and map visualization (Folium) to create and visualize different neighborhood clusters.