

IBM

Data Science Project Week 4th

INTRODUCTION

This is a capstone project for IBM Data Science Professional Certificate. I am developing a hypothetical scenario in this project for a concept that does not have enough Indian Restaurants in the Toronto area. So this might be a perfect opportunity for a Canadian-based entrepreneur. Since Indian food is popular among the Asian community, this entrepreneur may think about opening up his business in areas where the Asian community resides. With the intention in mind, finding the location to open such a restaurant is one of the most important decisions for this entrepreneur and I'm designing this project to help him find the most suitable place.

BUSINESS PROBLEM

The goal of this capstone project is to find the most appropriate place for the entrepreneur to open a new Indian Restaurant in Toronto , Canada. This project aims to provide answers to the business problem by using data science methods and techniques along with machine learning algorithms such as clustering: in Toronto, if an entrepreneur decides to open an Indian Restaurant, where will they consider opening it?

TARGET AUDIENCE

The entrepreneur who wants to find the location to open authentic Indian restaurant.

DATA

We'll need below data to solve this problem:

- List of Toronto neighbourhoods, Canada
- The latitude and longitude of those districts
- Venue data concerning restaurants in India. This will help us find the neighborhoods better suited for opening an Indian Restaurant.

EXTRACTING THE DATA

- Scrapping of Toronto neighborhoods via Wikipedia
- Getting Latitude and Longitude data of these neighborhoods via Geocoder package
- Using Foursquare API to get venue data related to these neighborhoods

In the next section we will discuss the techniques used in data analysis and machine learning.