APPLIED DATA SCIENCE CAPSTONE PROJECT

INTRODUCTION

This project, I am creating a hypothetical scenario for a concept that there may not be enough Indian Restaurants in New York Area. Therefore it might be a great opportunity for an entrepreneur who is based in USA. As the Indian food is popular among Asian community, so this entrepreneur might think of opening its business in areas where Asian community resides. With the purpose in mind, finding the location to open such a restaurant is one of the most important decisions for this entrepreneur and I am designing this project to help him find the most suitable location.

BUSINESS PROBLEM

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new Indian Restaurant in New York, USA. By using data science methods and tools along with machine learning algorithms such as clustering, this project aims to provide solutions to answer the business question : In New York, if an entrepreneur wants to open an Indian Restaurant, where should they consider opening it?

TARGET AUDIENCE

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

DATA

To solve this problem, we will need below data:

● List of neighbourhoods in New York, USA

● Latitude and Longitude of these neighbourhoods

● Venue data related to Indian restaurants. This will help us find the neighbourhoods that are more suitable to open an Indian Restaurant.

EXTRACTING THE DATA

● Scrapping of New York neighbourhoods via Wikipedia

● Getting Latitude and Longitude data of these neighbourhoods via Geocoder package

● Using Foursquare API to get venue data related to these neighbourhoods