**USING MACHINE LEARNING TO FIND LOCATIONS TO OPEN AN INDIAN RESTAURANT IN MELBOURNE METROPOLITAN AREA**

**(IBM CAPSTONE PROJECT)**



**Jananee Ganesapiragas**

Contents

[Introduction 2](#_Toc12964184)

[Business Problem 3](#_Toc12964185)

[Target Audience 3](#_Toc12964186)

[Data 3](#_Toc12964187)

[Methodology 3](#_Toc12964188)

# Introduction

For the capstone project I am considering a scenario of selecting a suitable site for opening an Indian restaurants in Melbourne metropolitan area.

Melbourne has a vibrant Indian community with Indian-born migrants making up 3 per cent of Melbourne’s total population. Since 2001 the number of Indian-born migrants in Melbourne has more than tripled. There may not be enough Indian restaurants in Melbourne and it might present a great opportunity for an entrepreneur who is wants to open an Indian restaurant in Melbourne.

# 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 Melbourne,Australia.

# Target Audience

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

# Data

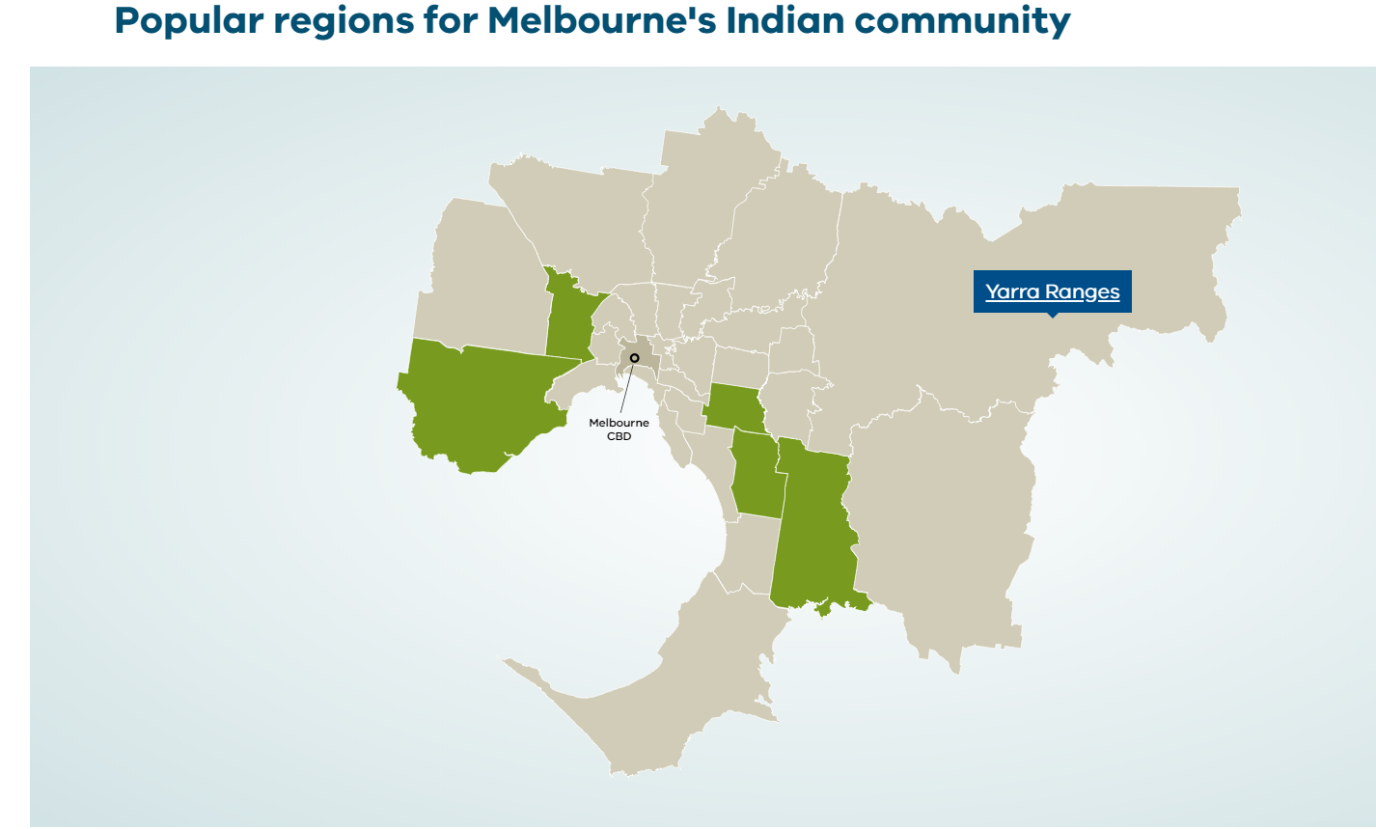
To solve this problem, I will need below data:

* List of suburbs where the Indian population is high.
* Latitude and Longitude of these suburbs.
* Venue data related to Indian restaurents.

# Methodology

As the first step to the project, I need to identify the suburbs where Indian community is high. This information is obtained from the following website <https://liveinmelbourne.vic.gov.au/discover/multicultural-communities/indian>

Cassy, Greater Dandenong, Monash, Brimbank and Wyndham are the major cities where Indian population is high.



The following suburbs of these five cities were considered for the analysis,

Endeavour Hills, Cranbourne, Berwick, Narre Warren, Tooradin, Springvale, Keysborough, Noble Park, Dandenong, Lyndhurst, Chadstone, Oakleigh, Glen Waverley, Mulgrave,Clayton, Point Cook, Werribee, Tarneit, Taylors Lakes, Deer Park, St Albans, Sunshine, and Keilor

The second step is to find the coordinates of these suburbs. Since Geocoder is not working, I prepared a csv file by manually extracting latitude and longitudes of these suburbs and read the file as pandas data frame.

I used folium package to visualize Melbourne and its suburbs.

Next, I use Foursquare API to pull the list of top 100 venues within 500 meters radius. From Foursquare, I am able to pull the names, categories, venues. With this data, I can also check how many unique categories that I can get from these venues. Then, I analyze each city by grouping the rows by city and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later.

Lastly, I performed the clustering method by using k-means clustering. I have clustered the cities into 5 clusters based on their similarity of venues. Then I looked for existent of Indian restaurant to assess the competition in those area. Then find the optimum location.