**OPENING AN INDIAN RESTAURANT IN TORONTO**

**A DATA SCIENCE APPROACH**

# Introduction

This report is intended to provide an insight on “How Data Science can be used to solve a business problem?” The report sections will discuss a specific business problem and a data science approach to solve the same.

## Business Scenario Background

Canada has been a preferred immigration destination for millions of people over the last decade. As the population has increased exponentially in Toronto, there is a high demand of services catering to the needs of immigrants. One such service is restaurants. The Hotel and restaurant business is on a boom ever since Toronto is flocked with immigrants. Every year, a high number of immigrants settle in the city Toronto who have immigrated from India. Hence there is a high demand for Indian Restaurants in Toronto. However, investing in a restaurant business is make or break life decision for anyone.

With this background, this report will try to address the following question for any entrepreneur who is planning to invest in opening a new India restaurant in Toronto city.

1. How many Indian restaurants are already present in Toronto City?
2. Which neighbourhoods in Toronto city has an Indian restaurant?
3. Which neighbourhoods in Toronto city will be ideal to start a new India Restaurant?
4. What are the descriptive highlights of these ideal neighbourhoods?

With these questions addressed, an entrepreneur will have a fair idea on where to start his business.



## Data Overview

As the scope of the report is around restaurants business in Toronto, the analysis would require data related to the geography of Toronto city. These may include but not limited to postal codes, neighbourhoods of Toronto city with their geo coordinates. Also, as the main business background is related to the restaurant business, data related to the list of restaurants in Toronto city will be required. Below are the list of datasets used in the report and background study.

1. Postal codes of Canada
2. Geographical coordinates of the postal codes.
3. Venues data of Toronto City.

## Data Description and Acquisition

The above mentioned dataset are explained below.

### Postal Codes of Canada – Web Scraping

The study for this report starts with exploring the geography of Toronto. For this, we need to understand the defined Postal codes of Toronto. Each postal code is assigned to a Borough. A Borough is defined as an area within a city which is either made of a single or multiple neighbourhoods. A list of all the postal codes with assigned Boroughs and neighbourhoods is available on a wiki page. Below is the link to the same. <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

Table

Description automatically generated

To acquire this data, we can use two approaches.

1. Using pandas.read\_html(): Pandas library in python offers support to load the data content of any html onto your working environment. Once the data is loaded, we can extract any tables or texts from the same html page.

For the purpose of this study, I have used the same to extract postal codes of Canada table from the wiki page. Once the data was extracted, it was sliced to create a data frame with subset of only postal codes, Borough and Neighbourhoods of Toronto city.

1. Using BeautifulSoup package: This package is also used to scrape the data from web for more complex cases. However, for the purpose of this study I have not used this package.

### Geographical coordinates of the postal codes – Web Scraping

The geographical coordinates data for all the postal codes and Borough is required as this is the main input for acquiring other venue related data. This data can be extracted using Google Maps Geocoding API or Geocoder Python package or using using a direct link to a csv file <http://cocl.us/Geospatial_data> .

For this study, I have used

1. Geocoder Python package and
2. the csv file

to extract geographical coordinates of location.

Table

Description automatically generated Graphical user interface, text, application

Description automatically generated

### Venues data of Toronto City – Foursquare API

As the study requires to investigate restaurant business, we need to acquire the venue data in Toronto. The venue data details all types of venues, their category and geo-coordinates for any desired location. For this study, the venue data was extracted using the Foursquare API.

Graphical user interface, text, application

Description automatically generated