



---

## *Capstone project*

---

*Auteur :*  
Warren LATA

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Business Problem</b>	<b>2</b>
<b>3</b>	<b>Audience</b>	<b>2</b>
<b>4</b>	<b>Data</b>	<b>2</b>
4.1	To solve the problem, we will need the following data . . . . .	2
4.2	Sources of data and methods to extract them . . . . .	3

## 1 Introduction

For many people, having some Asian restaurants near them is a great way to hang out and discover other specialties. They some diversity on neighborhoods and for the owners of these restaurants it provide consistent rental income. As a result, there are many shopping malls in the city of Denver and many more are being built. Of course, as with any business decision, opening a new restaurant requires serious consideration and is a lot more complicated than it seems. Particularly, the location is one of the most important decisions that will determine whether the mall will be a success or a failure.

## 2 Business Problem

The objective of this capstone project is to analyze and select the best locations in the city of Denver, Malaysia to open a new Asian Restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question : In the city of Denver, Colorado, if a property developer is looking to open a new restaurant, where would you recommend that they open it ?

## 3 Audience

This project is particularly useful to property developers and investors looking to open or invest in new Asian Restaurant in the city of Denver.

## 4 Data

### 4.1 To solve the problem, we will need the following data

To solve the problem, we will need the following data :

- List of neighbourhoods in Denver. This defines the scope of this project which is confined to the city of Denver, Colorado.
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to asian restaurant . We will use this data to perform clustering on the neighbourhoods.

## 4.2 Sources of data and methods to extract them

This Wikipedia page ([https://en.wikipedia.org/wiki/Category:Neighborhoods\\_in\\_Denver](https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Denver)) contains a list of neighbourhoods in Denver, with a total of 70 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods. After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the asian restaurant category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.