**IBM Data Science Capstone – Coursera**

**Where to open a new pizza place in Melbourne, Australia?**

**Introduction**

Melbourne is one of the largest cities in Australia. It happens to be a sports and culture capital of the country as well. Melbourne is rated as one of the most livable and student friendly cities in the world. This attracts a lot of foreign investors planning to gain residency in the city by investing a large sum of money in it.

Pizza is a universally loved food item. So if a pizza place is opened and the operation is executed well, it can lead to a great amount of profit create a good return for the investor. So if a foreign investor wants to open a pizza place in Melbourne, in which Suburb should he/she open it? This report will attempt to answer this question by using data science methods.

**Business problem**

The aim of this project is to analyze and recommend the best suburbs/areas Melbourne, Australia to open up a pizza place. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: Where would you recommend a new investor to open a new pizza place in the city of Melbourne?

**This project will be useful to… (target audience)**

This project is useful for any investors who are willing to open a new pizza place in the city of Melbourne.

**Data**

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

• List of suburbs in Melbourne. This defines the scope of this project

• Lat/long coordinates of those suburbs. This is required in order to plot the map and to get the venue data.

• Venue data, particularly data related to pizza places which is required to perform clustering on the suburbs.

Sources of data and methods to extract them

The Wikipedia page (<https://en.wikipedia.org/wiki/List_of_Melbourne_suburbs>) contains a list of suburbs in and around Melbourne, with a total of 539 suburbs. We will use the web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and pandas package. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those suburbs. Foursquare has one of the largest databases 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 Pizza places 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.