

# Capstone Project - Week 4-5

Vishnu Vinayamohanan

July 27, 2021

## 1 Introduction

Australia is a multi-cultural country which has got more than 200 nationalities living in the country. The diversity in population brought numerous cuisines from all over the world to Australia. The people of Australia are very keen to explore different cuisine from across the globe. The aim of this project is to analyze which city in Australia is best suited for starting a new restaurant business.

### 1.1 Problem Statement

Analyzing which cities in Australia is best suited for starting Restaurants.

## 2 Data Collection

Data used in this project are as follows;

### 2.1 Venue Data

The venue data has been extracted using the Foursquare API. This data contains venue recommendations for all cities in Australia and is used to study the popular venues of different neighborhoods.

### 2.2 Australian Cities

Main Australian cities are obtained from the url "<https://www.latlong.net/category/cities-14-15.html>". The data sets consists of the City names, latitude, and longitude information for each neighborhood. The data will be cleaned to remove redundancy and difference in the State Names as explained in the notebook. The data is read into a pandas data frame using the `read_html()` method.

### 2.3 Coordinates

The geographical coordinates for Australia data has been obtained from the GeoPy library in python.

## 3 Methodology

- Details of the Australian Cities would be scrapped from the url mentioned in section 2.2
- The coordinates of each of these Australian Cities will be fetched using the python Geopy package.
- Foursquare API search feature would be enabled to collect the nearby venues of the cities. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 200 and the radius parameter would be set to 1000.
- Folium- Python visualization library would be used to visualize the Australian cities cluster distribution.
- Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters of Australian cities.