# **INTRODUCTION**

**Business Problem**

Moving across states might be a complicated thing, especially for someone who does not want to take risks living out of comfort zone, living in a completely different neighbourhood. People might want to find similar neighbourhood to live in. In this project, I simulate a person who will move from Melbourne to Adelaide, specifically preferred living within 10km radius from Central Business District (CBD).

**Objective**

This project aims to provide relevant existing data and conduct appropriate analysis to find similar neighbourhood between Melbourne CBD and Adelaide CBD based on their nearby venues.

# **DATA ANALYSIS**

**Data Collection**

We have two data sources to serve the objective of this project:

* Australia post codes, taken from <https://raw.githubusercontent.com/Elkfox/Australian-Postcode-Data/master/au_postcodes.json>

This data takes geographical information: Post Code, Place Name, State Name, and coordinates (longitude and latitude).

* Four Square Data, taken from the provided API

Four square requests can help gather venues information in the target area.

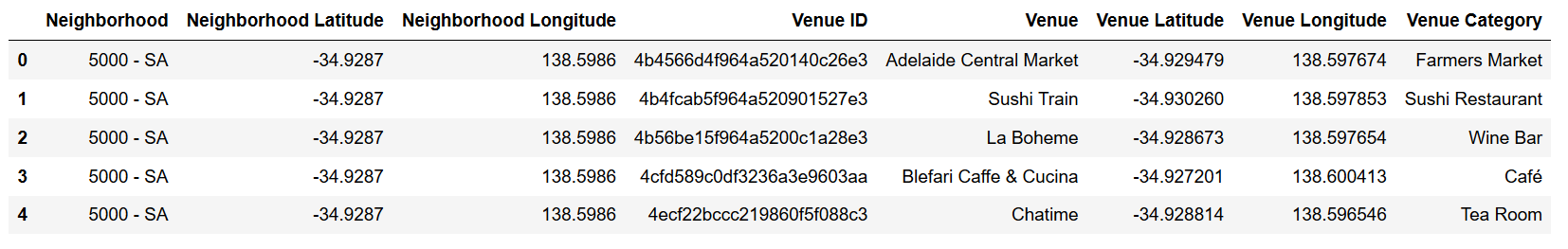
**Methodology**

For data cleaning, >10 km from CBD in each target state is filtered out. Merging two data sets, target Area by post code and Four-Square data by postcode, we obtain neighbourhoods within 10 km from CBD in each target area along with their surroundings venues and venues categories.

The following are the steps taken:

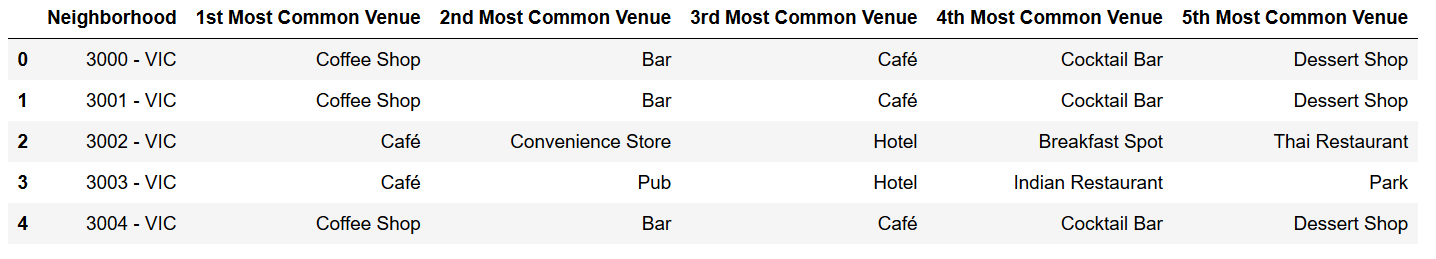
1. Accessing Four Square API databases and limit Four Square to return max 100 venues within 500 metres radius

Parameters fed to access API are: *client id, client secret key,* *coordinates, radius, limited number of venues* and *version.* API is expected to return max 100 venues within 500 metres from each coordinates from list of post codes.



1. Top 5 most common venues in selected areas in Adelaide and Melbourne

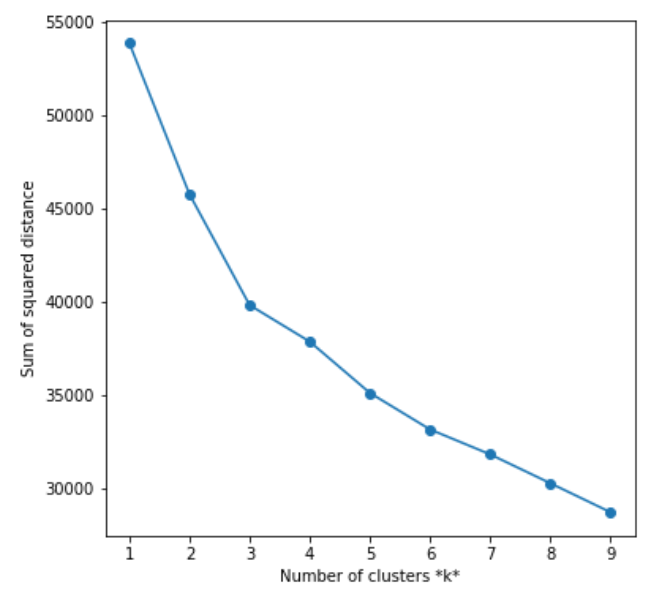
A matrix called one-hot encoding method is applied based on venue category for each post code (neighbourhood), so that we can get visualisation of what type of venues commonly found in a neighbourhood, example below.



Clustering list post codes of Adelaide and Melbourne by venues, it is expected that similar neighbourhoods of both cities are in the same cluster, that can be referred when someone would like to move from Melbourne to Adelaide.

1. Determining number of *k* clusters

Elbow method (depicted below) is applied to get optimal number of clusters as cluster analysis is the unsupervised method. From below picture of fitting data to elbow method, it is acknowledged that *k=3* is the best possible number of clusters, adding more *k* will not improve clustering performance significantly, thus we choose *k=3*  as number of clusters.



1. Applying k-means clustering with k=3

|  |  |  |
| --- | --- | --- |
| Cluster 1: Mixed Venues | Cluster 2: Café, Bar, Coffee Shop | Cluster 3: Coffee Shop |
| Melbourne and Adelaide | Melbourne and Adelaide | Melbourne |
| Mixed Venues : restaurant, stores, market | Café, Bar, Coffee Shop | Coffee Shop |

# **CONCLUSION**

From the three clusters, cluster 3 does not have Adelaide post codes, which leave cluster 1 and cluster 2 to be the options left for someone wants to move from Melbourne to Adelaide. If someone used to live in variety of venues around, list of Adelaide (SA, South Australia) post codes in Cluster 1 can be preferred. If someone used to live only near the local restaurants and more café or coffee shop, cluster 2 can be preferred.