

Geospatial

analysis of Melbourne and Sydney, AU

BY: NIKKO DOTE





Both domestic migration is common in Australia, moving between states means taking a chance on what benefits you will be gaining or compromise you will be taking in terms of your lifestyle, work, necessities. A lot of people try to find it in 2 of the world's most liveable cities, **Sydney and Melbourne.**

But for a long time, there has always been a heated debate about which Australian city is better.

40

There has been no attempt to conduct a suburb segmentation of these 2 cities.

Maybe this time, an informed choice is what the big-dreamers need.

Let's turn to Data Science to gain insights



First, Data

Using BeautifulSoup, webscrape the details of suburbs including their post codes from:

Melbourne:

https://en.m.wikipedia.org/wiki/List_of_Melbourne_suburbs

Sydney:

<https://namecensus.com/igapo/australia/postcodes/sydney-numeric.html>

Data Wrangling and Acquisition

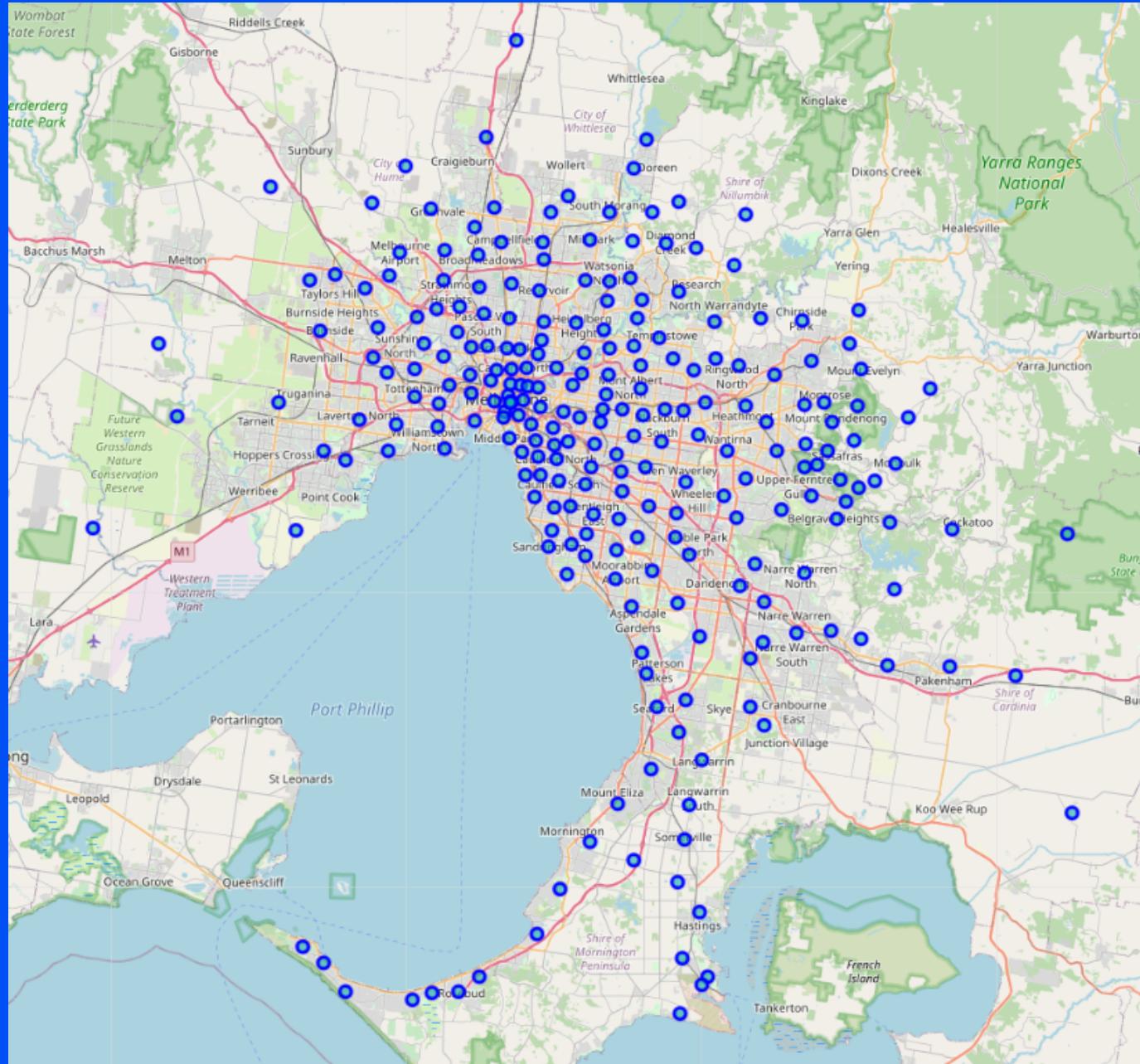
The data is cleaned using various techniques and it boils down to each post code with all its corresponding suburbs.

	Postal Code	Suburb
0	2000	Australia Square Post Office, Circular Quay, ...
1	2006	Sydney University
2	2007	Broadway, Ultimo

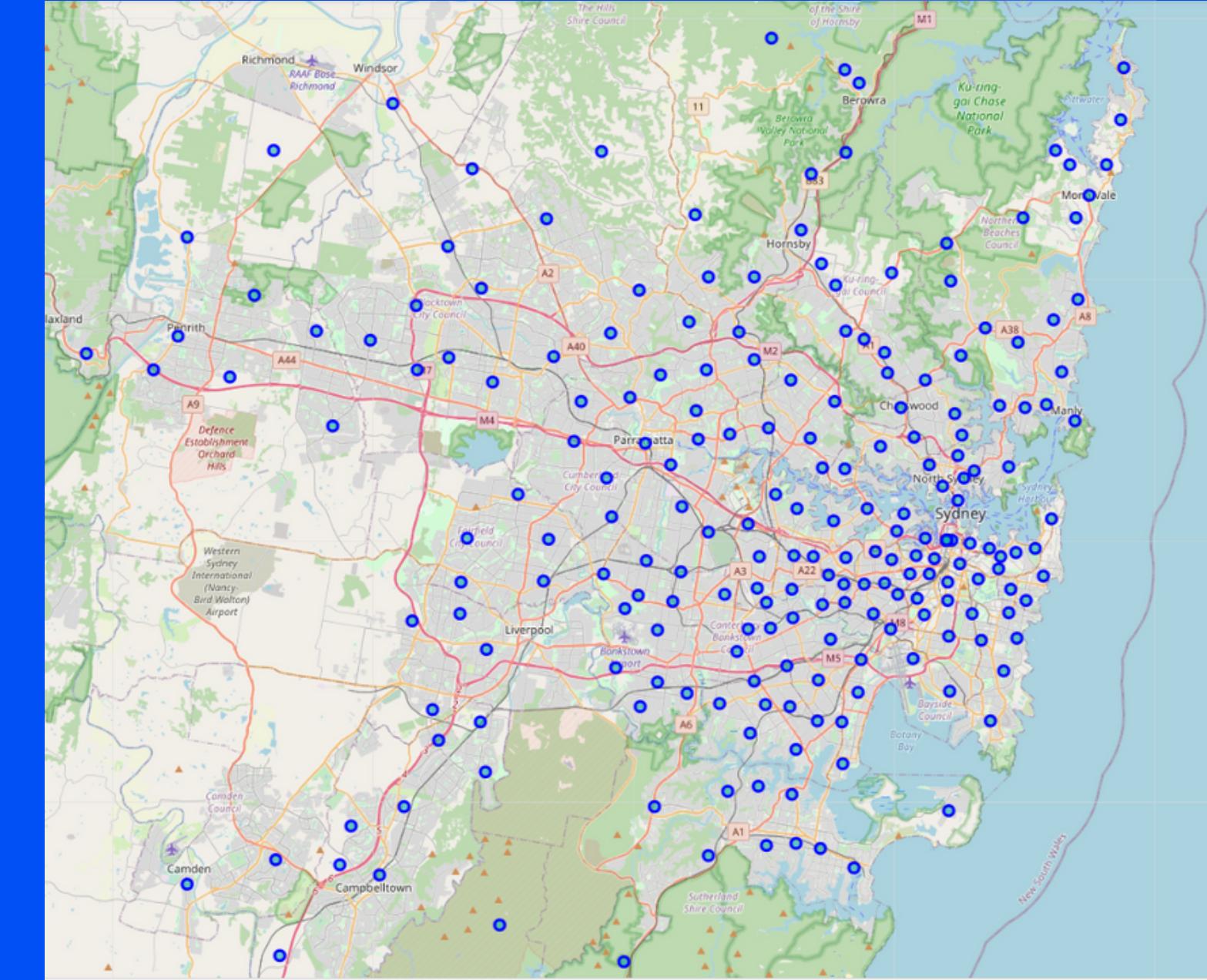
Then the postal codes are used to find the coordinates, **latitude and longitude**, using geocoder arcgis

Postal Code	Suburb	Latitude	Longitude
2000	Australia Square Post Office, Circular Quay, ...	-33.869815	151.209985
2006	Sydney University	-33.869600	151.206910
2007	Broadway, Ultimo	-33.879982	151.198450

Visualize maps with suburbs marked by circles



Map of Melbourne



Map of Sydney

Explore suburbs using **FOURSQUARE API**

Foursquare API can be used to derive the nearby venues once we have the coordinates of a certain place which represents the foot traffic data.

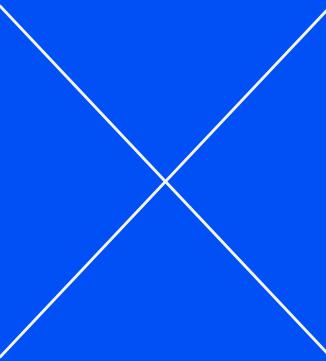
One-hot coding and exploratory data analysis

Since the venues are categorical variables, we can use one-hot coding to represent each category as a number for each suburb. These numbers can then be used to get the mean frequency which will represent how many of a certain venue is in a certain suburb.



Top 5 venues

we can explore the top 5
venues for each suburb to
get a feel for the data.



**Lyndhurst ,Lynbrook ,Lyndhurst **		
	Venues	Frequency
0	Train Station	0.5
1	Golf Course	0.5
2	African Restaurant	0.0
3	Park	0.0
4	Paintball Field	0.0

**Macclesfield ,Emerald ,Avonsleigh ,Clematis **		
	Venues	Frequency
0	Bar	0.3333
1	Train Station	0.3333
2	Café	0.3333
3	Pastry Shop	0.0000
4	Paper / Office Supplies Store	0.0000

**Macleod ,Yallambie ,Macleod **		
	Venues	Frequency
0	Business Service	0.25
1	Motel	0.25
2	Convenience Store	0.25
3	Café	0.25
4	African Restaurant	0.00

**Main Ridge **		
	Venues	Frequency
0	Asian Restaurant	0.1000
1	Ice Cream Shop	0.0667
2	Theater	0.0667
3	Café	0.0667
4	Music Venue	0.0333

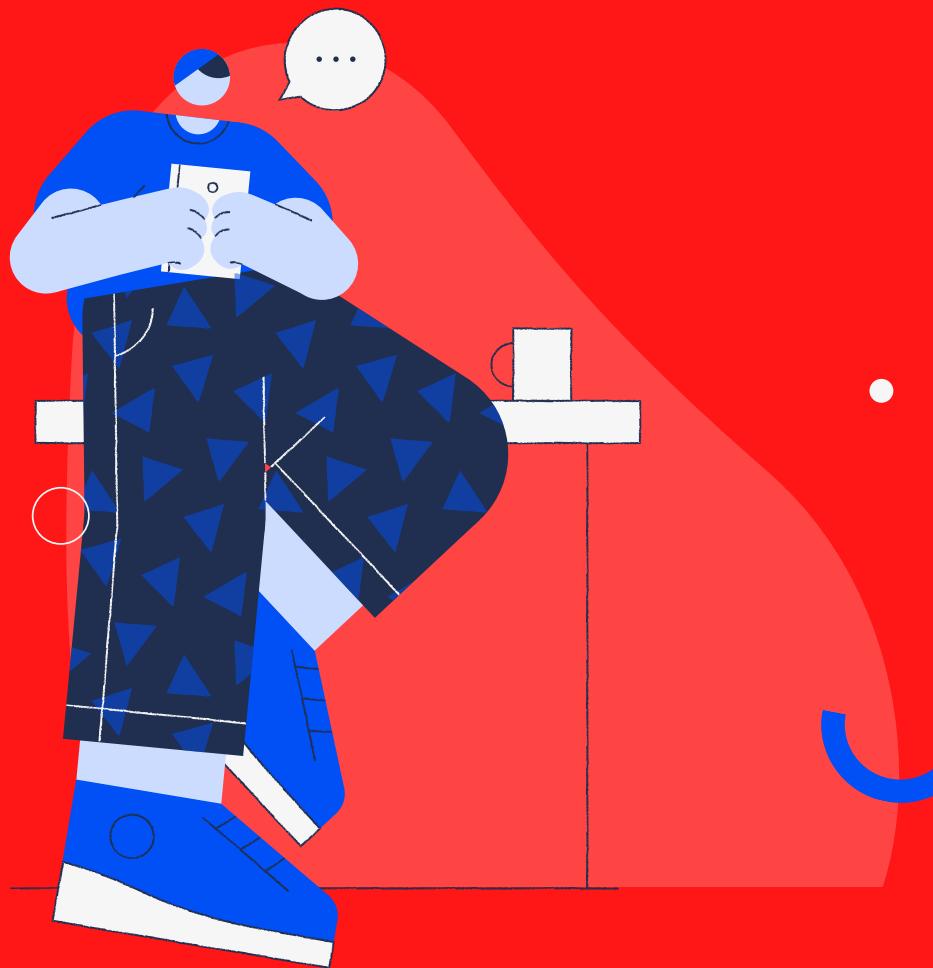
**Maribyrnong ,Ascot Vale ,Travancore **		
	Venues	Frequency
0	Café	0.1000
1	Electronics Store	0.1000
2	Dessert Shop	0.0667
3	Department Store	0.0667
4	Burger Joint	0.0333

**Matlock **		
	Venues	Frequency
0	Asian Restaurant	0.1000
1	Ice Cream Shop	0.0667
2	Theater	0.0667
3	Café	0.0667
4	Music Venue	0.0333

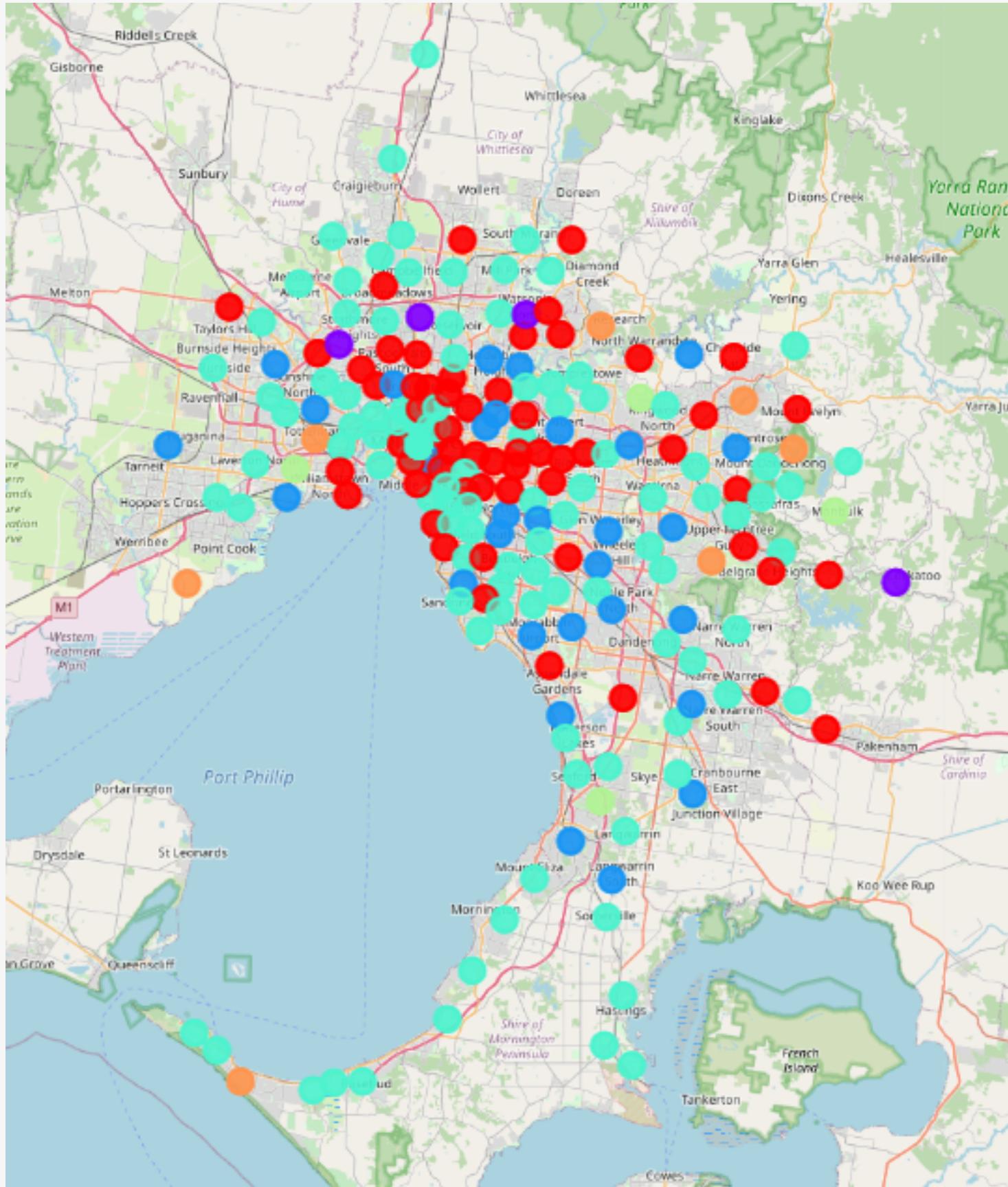
 Top 5 most numerous venues in a suburb and their corresponding frequency

Suburb clustering

- use machine learning algorithm K-means clustering
 - to separate or cluster the datapoints according to the nearest mean. The number of means or centroid is k.
 - In this project, we used 6 clusters.
 -
- in other words, the suburbs are then grouped according to its similarities may be separated by its dissimilarities



Melbourne



Cluster 1 - Cafe
58 post codes

Cluster 2 - Grocery Store
3 post codes

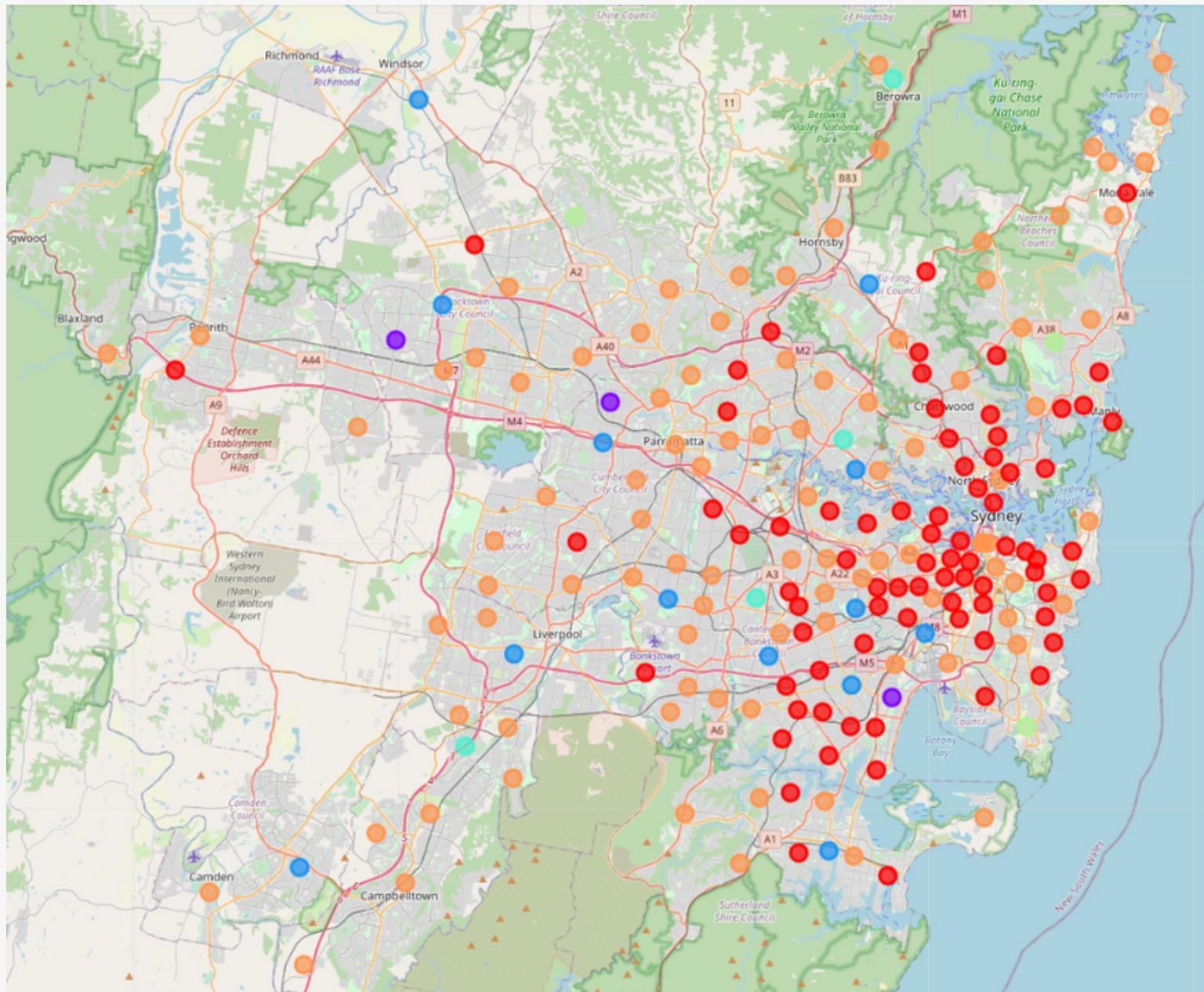
Cluster 3 - Parks
30 post codes

Cluster 4 - Restaurant and Entertainment
138 post codes

Cluster 5 - Business Service and Zoo
4 post codes

Cluster 6 - Home Service, Event Space
4 post codes

Sydney



Cluster 1 - Cafe
78 post codes

Cluster 2 - Gym and Yoga Studio
3 post codes

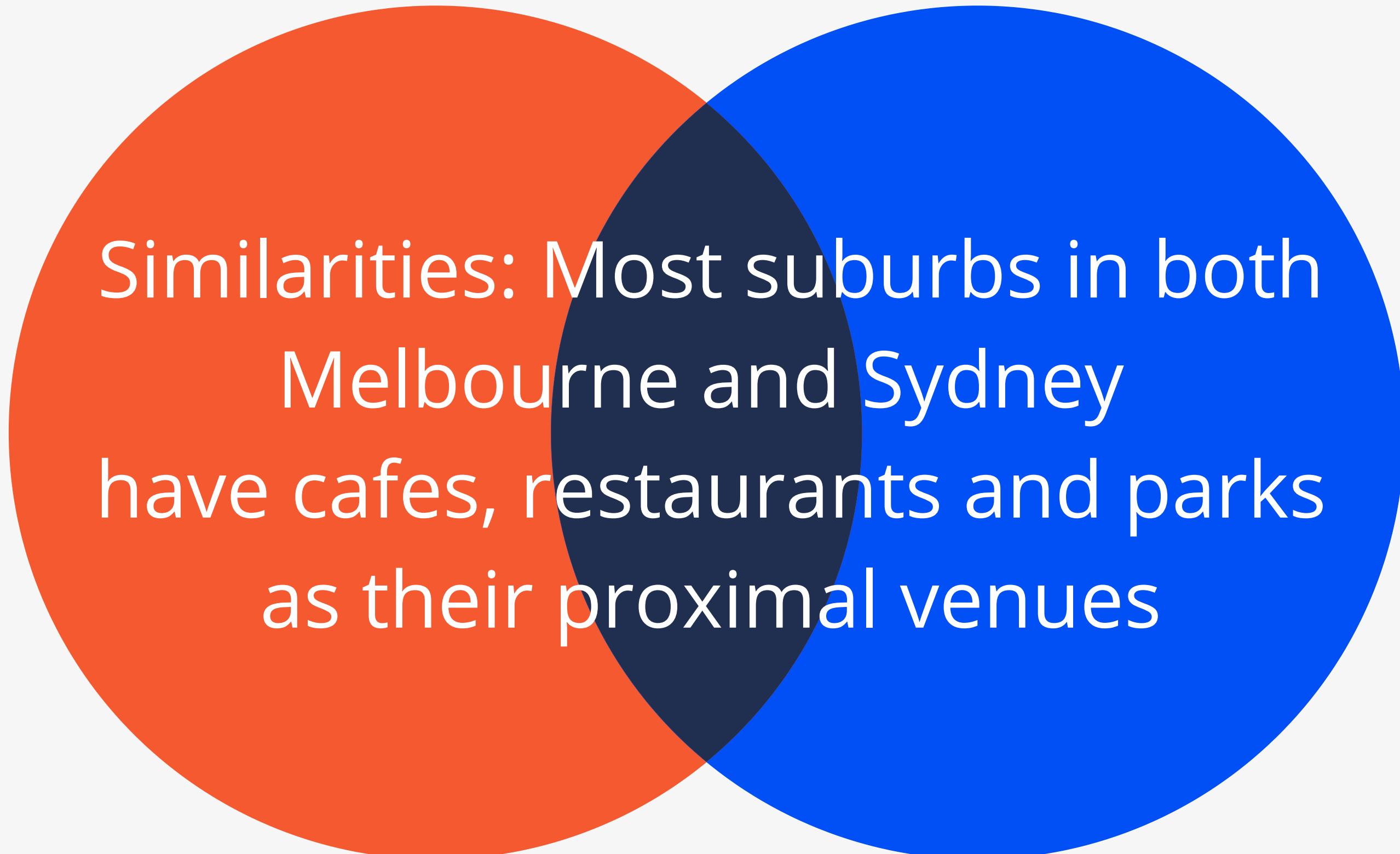
Cluster 3 - Fastfood Restaurants
13 post codes

Cluster 4 - Parks
4 post codes

Cluster 5 - Playground and Yoga
3 post codes

Cluster 6 - Shopping, Bar, Coffee Shop
107 post codes

Data Interpretation



Interesting:

There are quite a few suburbs near the central business district that have numerous park. It may be enticing for who work in the inner part of the cities who want to spend time in the park with their family and pets

contrast

The outer perimeter of the cities have very different landscape in terms of venue. Melbourne with zoos and business/home Service and Sydney with more fitness-related like gym and yoga studios.

Conclusion



- The 2 cities have a lot in common but also have peculiar differences that may affect one's decision making.
- Although the data is limited to foot traffic recorded in foursquare, we can still derive some useful insights.
- This project can be expanded to more specific purposes like analysing the crime rate per suburb or establishing business strategies.