Introduction

For many people, who is moving to Australia or planning to relocate with in Australia and looking for better life style, it is always a difficult task to choose the one from two of the best cities, not only in Australia but also in the world, ***Sydney and Melbourne***. These two cities finds place in the ***List of Top 10 most livable cities in the World*** almost every year.

For the year ending 30 June 2019:

- There were over 7.5 million migrants living in Australia.
- 29.7% of Australia's population was born overseas.
- Australia's population increased by 239,600 people due to net overseas migration.
- 404,000 people moved interstate, an increase of 2.9% from the previous year.

Business problem

The objective of this project is to analyze and select the best city to live in Australia between *Sydney and Melbourne.*

This project is mainly focused on geospatial analysis of the cities *Sydney and Melbourne* to understand which would be the best place to migrate to/stay. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: Between the cities *Sydney and Melbourne*, if someone looking to migrate, where would you recommend that they stay?

Data

To solve the problem, we will need the following data:
• List of neighborhoods/Suburbs in *Sydney and Melbourne*. This defines the scope of this project which is confined to the cities Sydney *and Melbourne*, two of the best cities in Australia.
• Latitude and longitude coordinates of those neighborhoods/Suburbs. This is required in order to plot the map and also to get the venue data.
• Venue data, particularly data related to top 10 venues in each neighborhood/Suburb. We will use this data to perform clustering on the neighborhoods.

Sources of Data and methods to extract the Data

The below links from Australia Post website are used as data sources:

1. Australia Post - Sydney
2. Australia Post - Melbourne

Used the web scraping techniques to extract the data from the Australia Post page, with the help of Python requests and Beautiful soup packages. The latitude and longitude details of the neighborhoods using Geocoder package. After that Foursquare API is used to get the venue data for those neighborhoods.

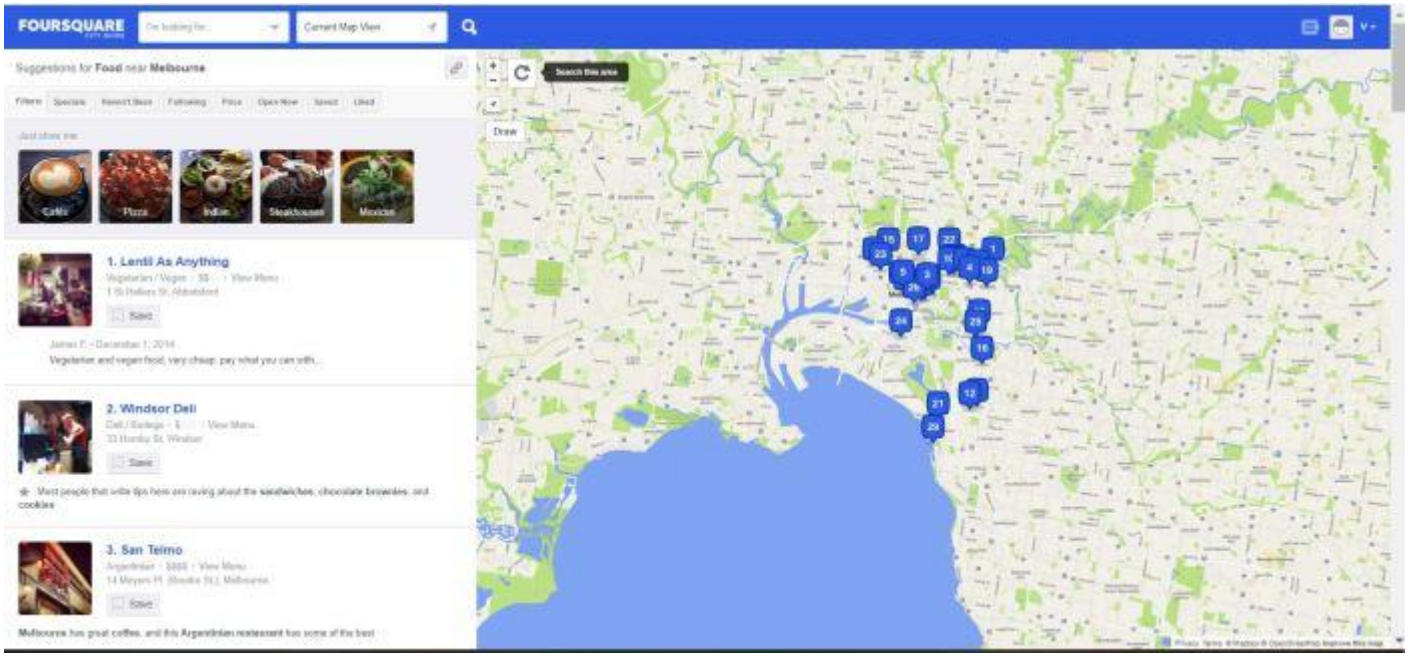
Foursquare API will provide many categories of the venue data, and we are particularly interested in the top 10 venues in order to help us solve the business problem. This is a project that will make use of many data science skills, from web scraping (Australia Post website), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

Methodology

The Foursquare API allows application developers to interact with the Foursquare platform. The API itself is a RESTful set of addresses to which you can send requests, so there's really nothing to download onto your server.

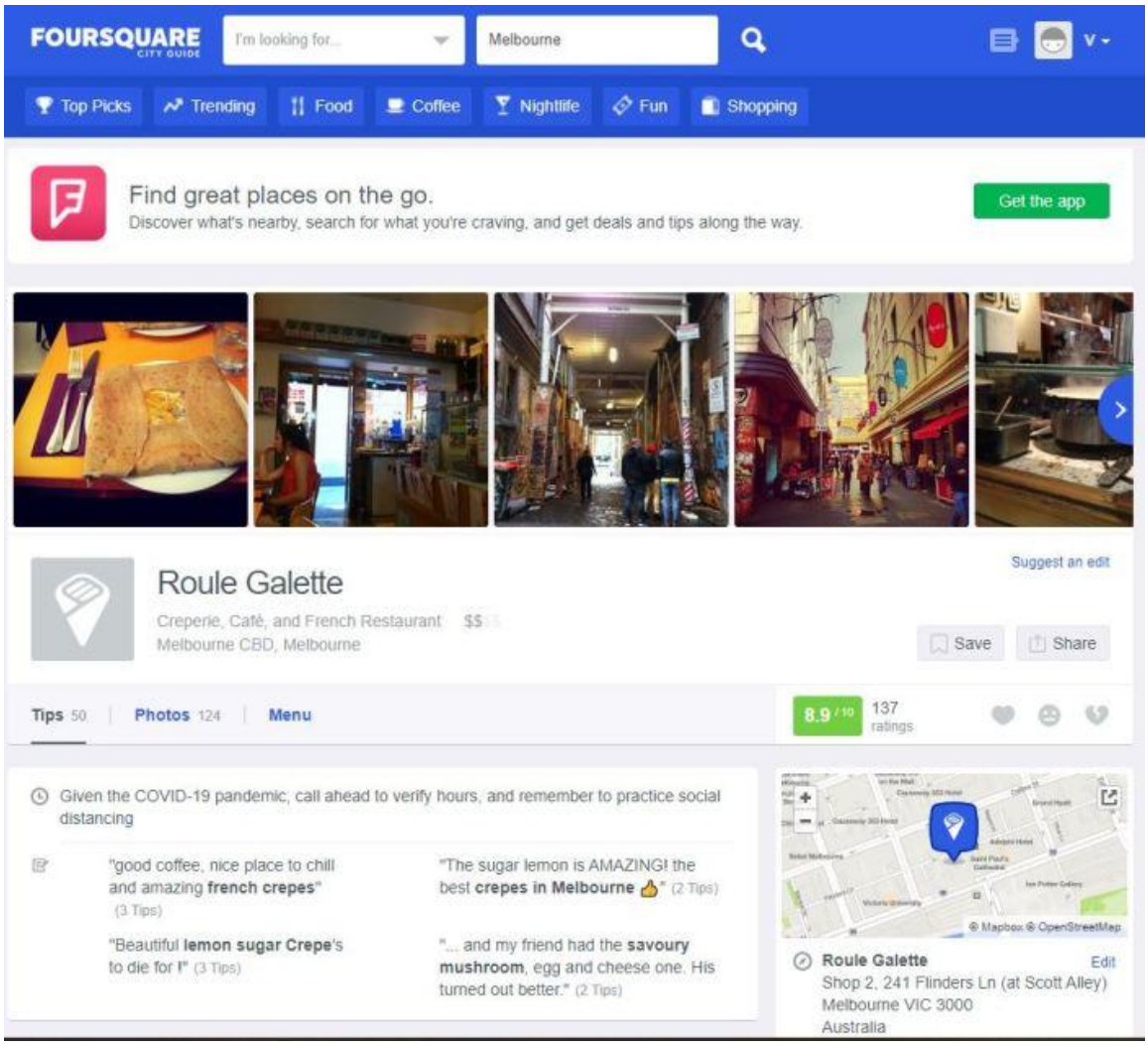
Food in Melbourne

Food returned by Foursquare



On the left, you see all the restaurants and their name, category, address of venues in Melbourne. On the right, you see the map of the venues on the left.

Exploring the results



If you click on any of the result, then you are redirected to the page where you see all the information in the Foursquare dataset about that place (here *Roule Galette*). This includes its name, full address, working hours, tips and images that users have posted about the cafe. So similarly you can explore the other venues of interest in S*ydney and Melbourne cities.*

To explore the Foursquare click here.

Download/Import required Libraries

First we need to install or import all the required libraries.

```
In [1]: import numpy as np # library to handle data in a vectorized manner

        import pandas as pd # library for data analysis

        import json # library to handle JSON files

        # !conda install -c conda-forge geopy --yes # uncomment this line if you haven't completed the Foursquare API lab
        from geopy.geocoders import Nominatim # convert an address into latitude and longitude values


        import requests # library to handle requests
        from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

        # Matplotlib and associated plotting modules
        import matplotlib.cm as cm
        import matplotlib.colors as colors

        # !conda install -c conda-forge scikit-learn
        from sklearn.cluster import KMeans


        # !pip install folium
        # !conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't completed the Foursquare API
        import folium # map rendering library

        from bs4 import BeautifulSoup
        import requests
        import csv
        import matplotlib.cm as cm
        import matplotlib.colors as colors

        print('Libraries imported.')

        Libraries imported.
```

Web Scraping

Perform scraping using Python requests and beautiful soup packages to extract the list of neighborhoods data.

- Send the GET request data.
- Parse the data from html into a beautiful soup object soup.
- Create a list to store Suburbs and Postal Codes.
- Append the data into the lists.
- Create a new Data Frames from the lists.
- Remove any records with duplicate postal codes.

	Postal Code	SubUrb
0	1231	SYDNEY SOUTH, NSW
1	1235	SYDNEY SOUTH, NSW
2	1466	UNSW SYDNEY, NSW
3	2000	SYDNEY, NSW
4	2001	SYDNEY, NSW
5	2006	THE UNIVERSITY OF SYDNEY, NSW
6	2020	SYDNEY DOMESTIC AIRPORT, NSW
7	2055	NORTH SYDNEY, NSW
8	2059	NORTH SYDNEY, NSW
9	2060	NORTH SYDNEY, NSW
10	2127	SYDNEY OLYMPIC PARK, NSW
11	2129	SYDNEY MARKETS, NSW

Sydney DF[/caption]

	Postal Code	SubUrb
0	3002	EAST MELBOURNE, VIC
1	8002	EAST MELBOURNE, VIC
2	3000	MELBOURNE, VIC
3	3001	MELBOURNE, VIC
4	3004	MELBOURNE, VIC
5	8001	MELBOURNE, VIC
6	3045	MELBOURNE AIRPORT, VIC
7	3052	MELBOURNE UNIVERSITY, VIC
8	5006	NORTH ADELAIDE MELBOURNE ST, SA
9	3051	NORTH MELBOURNE, VIC
10	3207	PORT MELBOURNE, VIC
11	3050	ROYAL MELBOURNE HOSPITAL, VIC
12	3205	SOUTH MELBOURNE, VIC
13	3004	ST KILDA ROAD MELBOURNE, VIC
14	3010	UNIVERSITY OF MELBOURNE, VIC
15	3003	WEST MELBOURNE, VIC

Melbourne DF[/caption]

The Data Frames are created after scraping the data. We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the Geocoder package that will allow us to convert the address into geographical coordinates in the form of latitude and longitude.

- Defining a function to get coordinates.
- Initialize empty Data Frame.
- Loop until you get the coordinates.
- Print the Data Frame.

	Postal Code	SubUrb	latitude	longitude
0	1231	SYDNEY SOUTH, NSW	-33.854816	151.216454
1	1235	SYDNEY SOUTH, NSW	-33.854816	151.216454
2	1466	UNSW SYDNEY, NSW	-33.917015	151.225182
3	2000	SYDNEY, NSW	-33.869844	151.207206
4	2001	SYDNEY, NSW	-36.427889	148.614567
5	2006	THE UNIVERSITY OF SYDNEY, NSW	-33.888907	151.189434
6	2020	SYDNEY DOMESTIC AIRPORT, NSW	-33.933728	151.181984
7	2055	NORTH SYDNEY, NSW	-33.841190	151.207469
8	2059	NORTH SYDNEY, NSW	-33.841190	151.207469
9	2060	NORTH SYDNEY, NSW	-33.840468	151.209086
10	2127	SYDNEY OLYMPIC PARK, NSW	-33.834583	151.071612
11	2129	SYDNEY MARKETS, NSW	-33.879847	151.203435

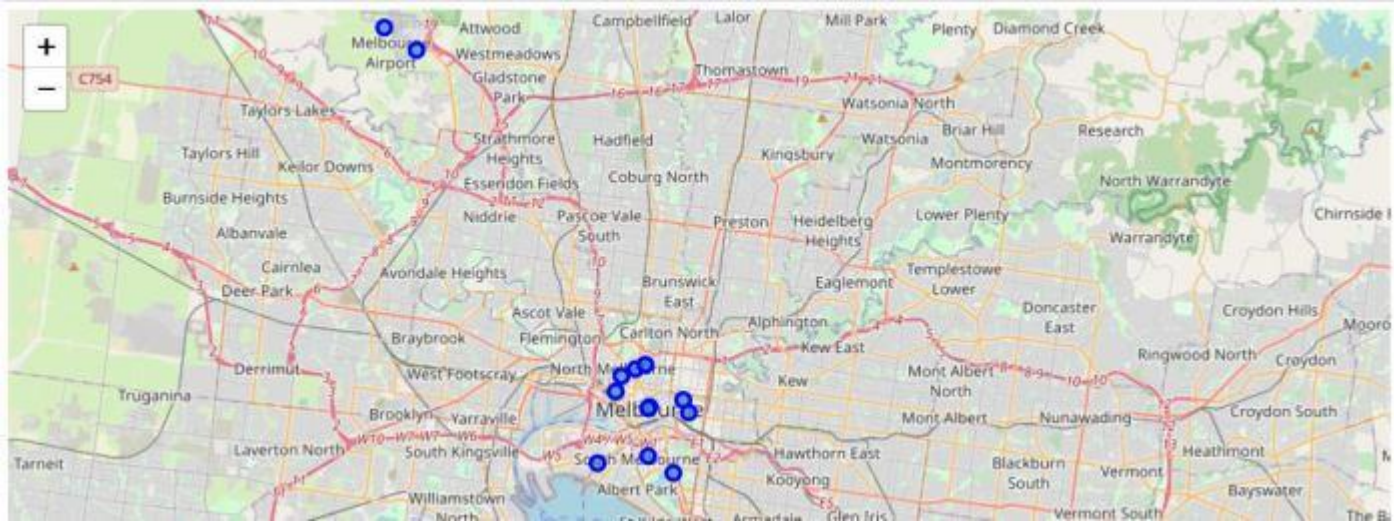
Sydney Lat Long DF[/caption]

	Postal Code	SubUrb	latitude	longitude
0	3000	MELBOURNE, VIC	-37.813562	144.963424
1	3001	MELBOURNE, VIC	-37.814327	144.962604
2	3002	EAST MELBOURNE, VIC	-37.810965	144.979667
3	3003	WEST MELBOURNE, VIC	-37.807744	144.946565
4	3004	MELBOURNE, VIC	-37.838951	144.975018
5	3010	UNIVERSITY OF MELBOURNE, VIC	-37.797080	144.961301
6	3045	MELBOURNE AIRPORT, VIC	-37.675556	144.849071
7	3050	ROYAL MELBOURNE HOSPITAL, VIC	-37.798669	144.956412
8	3051	NORTH MELBOURNE, VIC	-37.801557	144.949367
9	3052	MELBOURNE UNIVERSITY, VIC	-37.797080	144.961301
10	3205	SOUTH MELBOURNE, VIC	-37.832620	144.962225
11	3207	PORT MELBOURNE, VIC	-37.835499	144.937586
12	5006	NORTH ADELAIDE MELBOURNE ST, SA	-34.909838	138.602270
13	8001	MELBOURNE, VIC	-37.667111	144.833481
14	8002	EAST MELBOURNE, VIC	-37.815588	144.982555

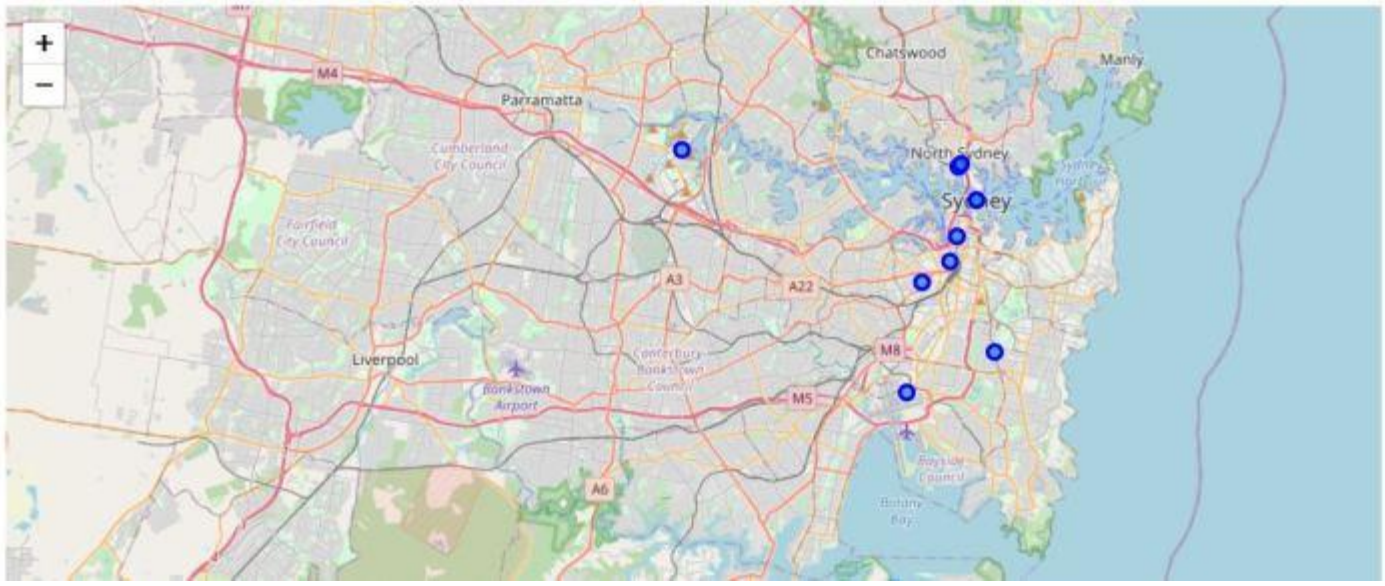
Melbourne Lat Long DF[/caption]

We have obtained the latitude and longitude coordinates for all the places.

After gathering the data, we have to visualize the neighborhoods in a map using Folium package.



Melbourne Map[/caption]



Sydney Map[/caption]

Use the foursquare API to explore the neighborhoods

- Provide Client ID and Client Secret details.
- Get the results in Json file using the dynamic URLs for each Suburb.
- Convert the results into Data Frame.
- Using the onehot coding and group by identify the count of each venue for all the locations.
- Create Data frame with top 10 venue details for each Suburb.

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	MELBOURNE, VIC	-37.813562	144.963424	Tipo 00	-37.813527	144.961978	Italian Restaurant
1	MELBOURNE, VIC	-37.813562	144.963424	Kirk's Wine Bar	-37.813661	144.961351	Wine Bar
2	MELBOURNE, VIC	-37.813562	144.963424	Brother Baba Budan	-37.813445	144.962137	Coffee Shop
3	MELBOURNE, VIC	-37.813562	144.963424	Little Rogue Coffee	-37.810986	144.964059	Coffee Shop
4	MELBOURNE, VIC	-37.813562	144.963424	B'cos Brazil	-37.815486	144.963085	Brazilian Restaurant

Melbourne venue[/caption]

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	SYDNEY SOUTH, NSW	-33.854816	151.216454	Sydney Opera House - Concert Hall	-33.856595	151.215058	Concert Hall
1	SYDNEY SOUTH, NSW	-33.854816	151.216454	Sydney Opera House - Studio	-33.856991	151.214482	Theater
2	SYDNEY SOUTH, NSW	-33.854816	151.216454	Drama Theatre	-33.856863	151.214759	Theater
3	SYDNEY SOUTH, NSW	-33.854816	151.216454	Sydney Opera House	-33.857260	151.215040	Opera House
4	SYDNEY SOUTH, NSW	-33.854816	151.216454	Sydney Opera House - Playhouse	-33.857518	151.214559	Theater

Sydney venue[/caption]

	SubUrb	Airport	Airport Food Court	Airport Lounge	Airport Service	American Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Asian Restaurant	...	Track Stadium	Trail	Tree	Turkish Restaurant	Vegetarian / Vegan Restaurant
0	NORTH SYDNEY, NSW	0	0	0	0	0	0	0	0	3	...	0	6	0	3	0
1	SYDNEY DOMESTIC AIRPORT, NSW	1	1	8	2	0	1	0	0	1	...	0	0	0	0	0
2	SYDNEY MARKETS, NSW	0	0	0	0	0	0	1	0	0	...	0	0	0	0	1
3	SYDNEY OLYMPIC PARK, NSW	0	0	0	0	1	0	0	0	1	...	1	1	0	1	0
4	SYDNEY SOUTH, NSW	0	0	0	0	0	0	2	2	0	...	0	2	2	0	0
5	SYDNEY, NSW	0	0	0	0	0	0	1	0	0	...	0	0	0	0	1
6	THE UNIVERSITY OF SYDNEY, NSW	0	0	0	0	0	0	1	0	0	...	0	0	0	0	1
7	UNSW SYDNEY, NSW	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

8 rows × 156 columns

Sydney grouped[/caption]

	SubUrb	African Restaurant	Airport	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Aquarium	Argentinian Restaurant	Art Gallery	...	Theme Park Ride / Attraction	Track	Turkish Restaurant	Vegetar / Veg Restaur
0	EAST MELBOURNE, VIC	0	0	0	0	0	0	0	2	2	...	0	1	0	
1	MELBOURNE AIRPORT, VIC	0	2	9	2	1	0	0	0	0	...	0	0	0	
2	MELBOURNE UNIVERSITY, VIC	1	0	0	0	0	0	0	0	1	...	0	1	0	
3	MELBOURNE, VIC	0	1	9	2	1	0	0	3	1	...	0	2	0	
4	NORTH ADELAIDE MELBOURNE ST, SA	0	0	0	0	0	1	0	0	0	...	0	0	0	
5	NORTH MELBOURNE, VIC	1	0	0	0	0	0	1	0	1	...	0	0	0	
6	PORT MELBOURNE, VIC	0	0	0	0	0	1	0	0	0	...	0	0	1	
7	ROYAL MELBOURNE HOSPITAL, VIC	1	0	0	0	0	0	0	0	1	...	0	1	0	
8	SOUTH MELBOURNE, VIC	0	0	0	0	0	0	0	0	1	...	0	0	1	
9	UNIVERSITY OF MELBOURNE, VIC	1	0	0	0	0	0	0	0	1	...	0	1	0	
10	WEST MELBOURNE, VIC	1	0	0	0	0	0	0	1	1	...	1	0	0	

11 rows × 169 columns

Melbourne grouped[/caption]

	SubUrb	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	EAST MELBOURNE, VIC	Café	Cocktail Bar	Japanese Restaurant	Park	Bar	Asian Restaurant	Coffee Shop	Australian Restaurant	Wine Bar	Italian Restaurant
1	MELBOURNE AIRPORT, VIC	Airport Lounge	Café	Coffee Shop	Hotel	Noodle House	Rental Car Location	Fast Food Restaurant	Juice Bar	Donut Shop	Scenic Lookout
2	MELBOURNE UNIVERSITY, VIC	Café	Coffee Shop	Bar	Bakery	Korean Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Cocktail Bar	Bookstore	Japanese Restaurant
3	MELBOURNE, VIC	Café	Bar	Coffee Shop	Cocktail Bar	Airport Lounge	Italian Restaurant	Asian Restaurant	Park	Japanese Restaurant	Pizza Place
4	NORTH ADELAIDE MELBOURNE ST, SA	Café	Pub	Coffee Shop	Thai Restaurant	Hotel	Pizza Place	Burger Joint	French Restaurant	Bar	Italian Restaurant

Melbourne Top 10[/caption]

	SubUrb	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	NORTH SYDNEY, NSW	Café	Park	Scenic Lookout	Japanese Restaurant	Bar	Theater	Australian Restaurant	Italian Restaurant	Coffee Shop	Trail
1	SYDNEY DOMESTIC AIRPORT, NSW	Hotel	Airport Lounge	Café	Bakery	Coffee Shop	Fast Food Restaurant	Juice Bar	Thai Restaurant	Donut Shop	Sports Bar
2	SYDNEY MARKETS, NSW	Coffee Shop	Thai Restaurant	Café	Bar	Hotel	Theater	Cocktail Bar	Pub	Japanese Restaurant	Park
3	SYDNEY OLYMPIC PARK, NSW	Café	Park	Japanese Restaurant	Hotel	Fast Food Restaurant	Athletics & Sports	Supermarket	Vietnamese Restaurant	Australian Restaurant	Seafood Restaurant
4	SYDNEY SOUTH, NSW	Café	Scenic Lookout	Australian Restaurant	Hotel	Park	Pub	Theater	Hotel Bar	Speakeasy	Pool

Sydney Top 10[/caption]

Clustering the neighborhoods

Now we need to cluster all the neighborhoods into different clusters. The results will allow us to identify which city is surrounded by diverse group of clusters based on which our conclusion is drawn. If one city has less diverse clusters surrounding than the other help us answer the question that is which city is better place for living.

We set the number of clusters to 4 and run the algorithm. After applying the K-Means clustering algorithm, all the neighborhoods/suburbs get segregated and form different clusters based on the top 10 venues in that area.

	Postal Code	SubUrb	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	3000	MELBOURNE, VIC	-37.813562	144.963424	0	Café	Bar	Coffee Shop	Cocktail Bar	Airport Lounge	Italian Restaurant	Asian Restaurant	Park	Japanese Restaurant
1	3001	MELBOURNE, VIC	-37.814327	144.962604	0	Café	Bar	Coffee Shop	Cocktail Bar	Airport Lounge	Italian Restaurant	Asian Restaurant	Park	Japanese Restaurant
2	3002	EAST MELBOURNE, VIC	-37.810965	144.979667	2	Café	Cocktail Bar	Japanese Restaurant	Park	Bar	Asian Restaurant	Coffee Shop	Australian Restaurant	Wine Bar
3	3003	WEST MELBOURNE, VIC	-37.807744	144.946565	1	Café	Coffee Shop	Korean Restaurant	Japanese Restaurant	Bar	Cocktail Bar	Chinese Restaurant	Italian Restaurant	Korean BBQ Restaurant
4	3004	MELBOURNE, VIC	-37.838951	144.975018	0	Café	Bar	Coffee Shop	Cocktail Bar	Airport Lounge	Italian Restaurant	Asian Restaurant	Park	Japanese Restaurant

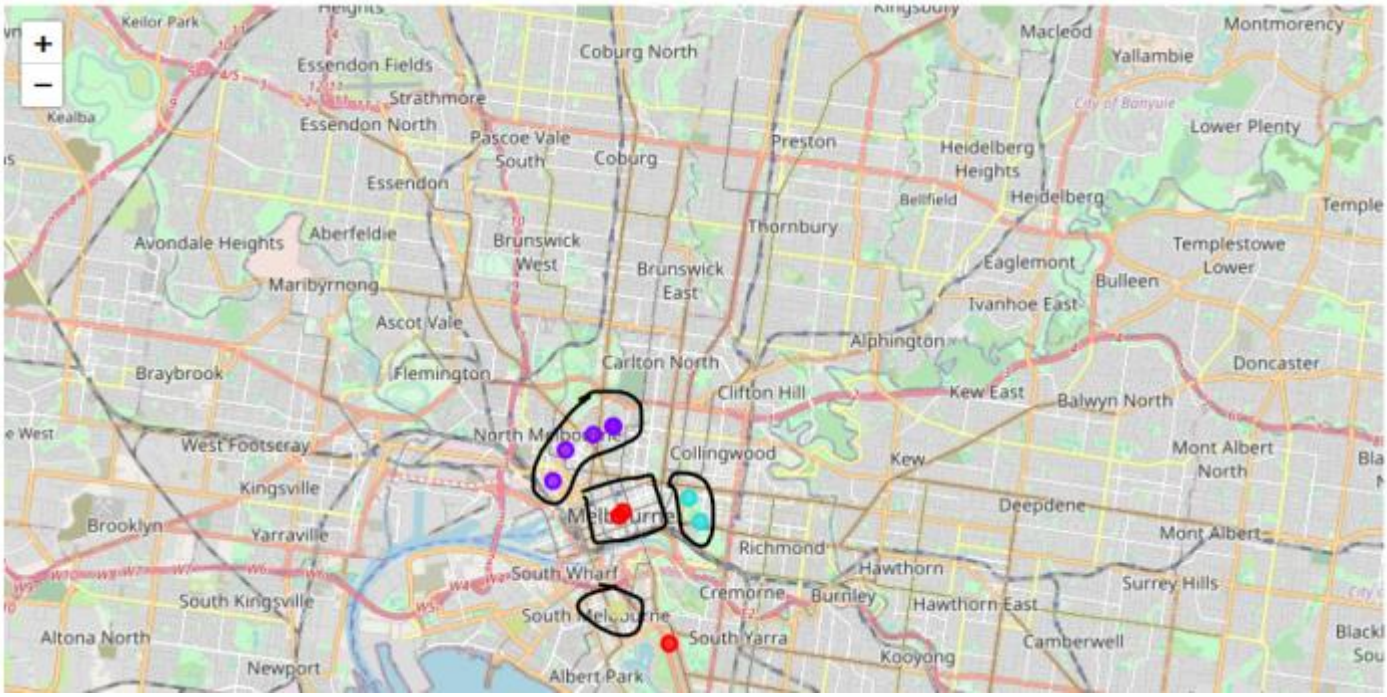
Melbourne Cluster[/caption]

	Postal Code	SubUrb	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	1 C
0	1231	SYDNEY SOUTH, NSW	-33.854816	151.216454	2	Café	Scenic Lookout	Australian Restaurant	Hotel	Park	Pub	Theater	Hotel Bar	Speakeasy	
1	1235	SYDNEY SOUTH, NSW	-33.854816	151.216454	2	Café	Scenic Lookout	Australian Restaurant	Hotel	Park	Pub	Theater	Hotel Bar	Speakeasy	
2	1466	UNSW SYDNEY, NSW	-33.917015	151.225182	3	Café	Thai Restaurant	Indonesian Restaurant	Chinese Restaurant	Sandwich Place	Convenience Store	Bar	Coffee Shop	Malay Restaurant	
3	2000	SYDNEY, NSW	-33.869844	151.207206	0	Café	Coffee Shop	Hotel	Speakeasy	Cocktail Bar	Shopping Mall	Bar	Bakery	Hotel Bar	Re
4	2001	SYDNEY, NSW	-36.427889	148.614567	0	Café	Coffee Shop	Hotel	Speakeasy	Cocktail Bar	Shopping Mall	Bar	Bakery	Hotel Bar	Re

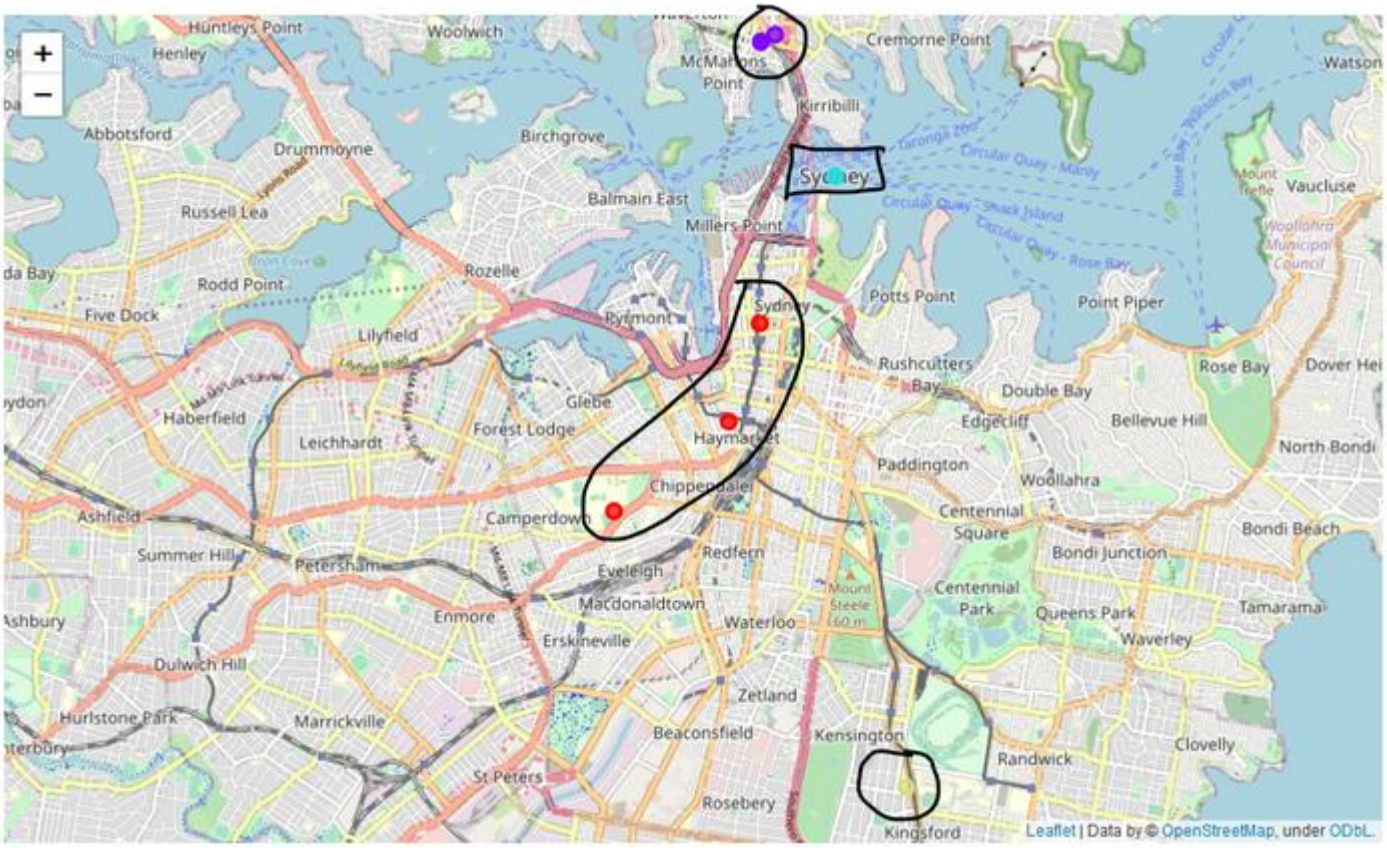
Sydney Cluster [/caption]

Visualizing the resulting clusters

`Both the cities marked in boxes.`



Melbourne Final Map[/caption]



Sydney Final Map[/caption]

Conclusions

Up on examining the results (final maps with clusters), it is clear that the city of Melbourne is surrounded by diverse group of clusters in close proximity where as for city of Sydney, the clusters are spread out far. Also the clusters represented with Yellow color are very far from the city of Sydney.

As these clusters intern represent group of venues, we are making an assumption that Melbourne has access to variety of venues compared to Sydney. This will impact the quality of living.

So with this analysis we are placing Melbourne on top of the Sydney in terms of place for quality of living.

By factoring the other aspects such as population growth, health facilities, education facilities and other venues etc this analysis can be further verified and finding can be very different from what we have here. Given all these are currently not in scope, and with the tools and understanding we have, we hope this analysis and findings are justifies well enough.

Hope you found the article helpful.

Thank you so much for reading!