

# **Capstone Project – The Battle of Neighborhoods**

## **Similar Neighborhoods in two Cities**

Mriganka Medhi

September 8, 2019

### **Table of Contents:-**

1. Introduction
2. Data Acquisition, Wrangling and Utilization
3. Methodology
4. Results and Discussion
5. Conclusion

# **1- Introduction**

## **1.1- Background**

Moving from one city to another can be a very tedious and time-consuming task. On moving, people often tend to find it difficult to adjust to the new surroundings. They definitely prefer to find areas that are similar to their previous areas of residence. This can make it much easier for them to adapt to life in a new city. Also, they would not need to change their habits and lifestyle if they can manage to find similar neighborhoods. Hence, it could be very beneficial for people planning to move between cities to gain valuable information about the neighborhoods in the new city that are somewhat identical to their current city.

## **1.2- Business Problem**

In this project, data related to the neighborhoods of the two cities of Adelaide, Australia and Melbourne, Australia will be used. The aim of this project is to find the similar groups of neighborhoods in the two cities and cluster them based on their location data, i.e. data about the types of venues present in the vicinity of the neighborhood. If the common venues around a neighborhood are similar, then the neighborhoods are likely to be similar to each other.

### 1.3- Interest

The similarity of neighborhoods in different cities is of interest to people who have to move from one city to another (specifically, in this case, from Adelaide to Melbourne or vice-versa). It may also be useful for travel agencies and consultancies that advice people on the suitability of neighborhoods.

## 2- Data Acquisition, Wrangling and Utilization

### 2.1- Data Sources

The list of neighborhoods/suburbs, their postal codes and local government areas for Adelaide and Melbourne were obtained from the Wikipedia pages

[https://en.wikipedia.org/wiki/List\\_of\\_Adelaide\\_suburbs](https://en.wikipedia.org/wiki/List_of_Adelaide_suburbs) and [https://en.wikipedia.org/wiki/List\\_of\\_Melbourne\\_suburbs](https://en.wikipedia.org/wiki/List_of_Melbourne_suburbs)

respectively. The websites were scraped using the **BeautifulSoup** library. The latitude and longitude values were obtained using the python library **geopy**. The location data for the neighborhoods was obtained using the **Foursquare API**.

## 2.2- Data Wrangling

Initially, the two tables containing the neighborhood details were converted to pandas dataframes. Only the columns corresponding to suburb, local government areas and postal codes were kept. The data appears as follows -

	PostalCode	Neighborhood	Local government area
0	5000	Adelaide	City of Adelaide
1	5006	North Adelaide	City of Adelaide
2	5072	Auldana	City of Burnside
3	5066	Beaumont	City of Burnside
4	5067	Beulah Park	City of Burnside

Adelaide data sample after cleaning up

	PostalCode	Neighborhood	Local government area
0	3081	Bellfield	City of Banyule
1	3088	Briar Hill	City of Banyule
2	3083	Bundoora	City of Banyule; City of Darebin; City of Whit...
3	3084	Eaglemont	City of Banyule
4	3095	Eltham	City of Banyule; Shire of Nillumbik

Melbourne data sample after cleaning up

The suburbs with the same postal codes were combined into single rows, such that their names are separated by commas. Two new columns for latitude and longitude of each postal code were added. The two dataframes were concatenated, one after the other so that the details of all neighborhoods in both cities could be obtained in a single dataframe.

## 2.3 –Utilization of Data

The latitude and longitude of each postal code was used to obtain the corresponding location data using Foursquare.

Below is a sample of this data –

	Neighborhood PostalCode	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	3781	-37.677827	145.136993	OZ Tree Services	-37.676120	145.139290	Tree
1	3781	-37.677827	145.136993	Heritage House Garden Centre	-37.678831	145.132984	Garden Center
2	3139	-37.920040	145.016265	The Good Guys	-37.916573	145.018905	Furniture / Home Store
3	3139	-37.920040	145.016265	Bus Stop 15656	-37.922322	145.018311	Bus Stop
4	3139	-37.920040	145.016265	Brighton Golf Course	-37.923769	145.016880	Golf Course

A sample of the location data obtained from Foursquare

Using the location data, the five most common venues for each postal code was obtained as follows -

	Neighborhood PostalCode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	3000	Café	Dessert Shop	Coffee Shop	Bar	Italian Restaurant
1	3002	Hotel	Café	Wine Bar	Park	Convenience Store
2	3003	Flea Market	Asian Restaurant	Fish Market	Flower Shop	Food
3	3004	Café	Cocktail Bar	Dessert Shop	Bar	Italian Restaurant
4	3006	Café	Bar	Hotel	Grocery Store	Bakery

A sample of the dataframe containing the 5 most common venues for each postal code

This data was used to cluster the neighborhoods into several disjoint clusters. The neighborhoods (of both cities) that fall in the same cluster can be considered to be similar to each other. The clusters can be further visualized on a map using Folium

library. Hence, this data can be very useful to gain insights about similarity of neighborhoods among the two cities.

### 3- Methodology

#### 3.1- Calculating important statistics

Here, first we calculated the number of venues returned by Foursquare for each postal code.

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood PostalCode						
3000	30	30	30	30	30	30
3002	30	30	30	30	30	30
3003	2	2	2	2	2	2
3004	30	30	30	30	30	30
3006	30	30	30	30	30	30
3008	30	30	30	30	30	30
3011	4	4	4	4	4	4

A sample of the dataframe showing the no. of venues returned for each postal code

We observed that there were only 362 distinct postal codes in this dataframe whereas there were 395 postal codes which we obtained from the Wikipedia links. Hence there were 33 such neighborhoods for which no nearby venues were returned by Foursquare. These venues were considered to be an additional cluster later on.

Also, we calculated the number of unique venue categories across all the neighborhoods as 295. However, to use this information for clustering, we had to perform one-hot encoding to convert the categories into numerical values (1 if a venue is of a particular category and 0 if it is not of that category). A sample of this dataframe is shown below –

	Neighborhood PostalCode	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Waterfall
0	3781	0	0	0	0	0	0	0	0	0	...	0
1	3781	0	0	0	0	0	0	0	0	0	...	0
2	3139	0	0	0	0	0	0	0	0	0	...	0
3	3139	0	0	0	0	0	0	0	0	0	...	0
4	3139	0	0	0	0	0	0	0	0	0	...	0

A sample of the dataframe containing the category of each venue after one-hot encoding

Using this information, we calculated the mean of the frequency of the occurrence of each category grouped by the postal code.

	Neighborhood PostalCode	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Waterfa
0	3000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.033333	...	0.0
1	3002	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0
2	3003	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0
3	3004	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0
4	3006	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0

A sample of the dataframe showing the mean of the frequency of the occurrence of each category for each postalcode

Using the mean of the frequency of the occurrence of each category, we found out the five most common venues for each neighborhood.

	Neighborhood PostalCode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	3000	Café	Dessert Shop	Coffee Shop	Bar	Italian Restaurant
1	3002	Hotel	Café	Wine Bar	Park	Convenience Store
2	3003	Flea Market	Asian Restaurant	Fish Market	Flower Shop	Food
3	3004	Café	Cocktail Bar	Dessert Shop	Bar	Italian Restaurant
4	3006	Café	Bar	Hotel	Grocery Store	Bakery

A sample of the dataframe showing the 5 most common venues around each neighborhood

### 3.2- Clustering the neighborhoods

We used the k-means clustering algorithm to cluster the neighborhoods into 4 different categories based on the types of venues in their vicinity. We assigned them a cluster label of 0-3 based on which cluster they belonged to. There were also the 33 postal codes for which no venues were returned; hence we assigned all these neighborhoods the cluster 4. We then proceeded to separate the dataframe based on the city, Adelaide or Melbourne.



	PostalCode	Neighborhood	Local government area	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Cor Ven
0	5000	Adelaide	City of Adelaide	Adelaide	-34.9282	138.6	1	Café	Tea Room	Pizza Place	Asia Res
1	5006	North Adelaide	City of Adelaide	Adelaide	-34.9085	138.595	1	Pub	Burger Joint	Italian Restaurant	Caf
2	5007	Bowden, Brompton, Hindmarsh, Welland, West Hin...	City of Charles Sturt	Adelaide	-34.9029	138.58	1	Breakfast Spot	Beer Store	Pizza Place	Gyr
3	5008	Croydon, Devon Park, Renown Park, Ridleyton, W...	City of Charles Sturt	Adelaide	-34.8952	138.566	0	Supermarket	Train Station	Gluten-free Restaurant	Liqui Stoi
4	5009	Allenby Gardens, Beverley, Kilkenny	City of Charles Sturt	Adelaide	-34.9016	138.554	1	Vietnamese Restaurant	Gym	Burger Joint	Hor Ser

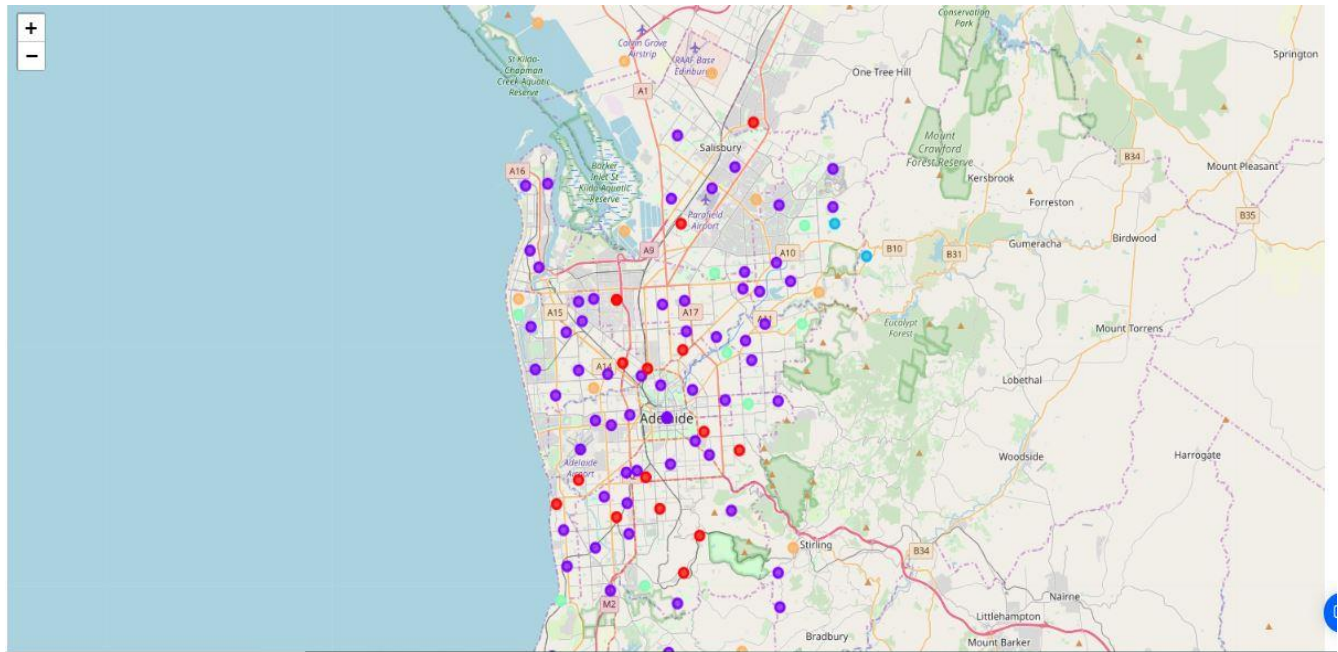
A sample of the final dataframe for Adelaide

	PostalCode	Neighborhood	Local government area	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Com Venu
0	3000	Melbourne CBD	City of Melbourne	Melbourne	-37.8142	144.96	0	Café	Dessert Shop	Coffee Shop	Bar
1	3002	East Melbourne	City of Melbourne	Melbourne	-37.8125	144.986	0	Hotel	Café	Wine Bar	Park
2	3003	West Melbourne	City of Melbourne	Melbourne	-37.8104	144.92	1	Flea Market	Asian Restaurant	Fish Market	Flow Shop
3	3004	Melbourne CBD (St Kilda Road area)	City of Melbourne (east side to High Street, P...	Melbourne	-37.8142	144.963	1	Café	Cocktail Bar	Dessert Shop	Bar
4	3006	Southbank, South Wharf	City of Melbourne; City of Port Phillip	Melbourne	-37.8254	144.964	0	Café	Bar	Hotel	Groc Store

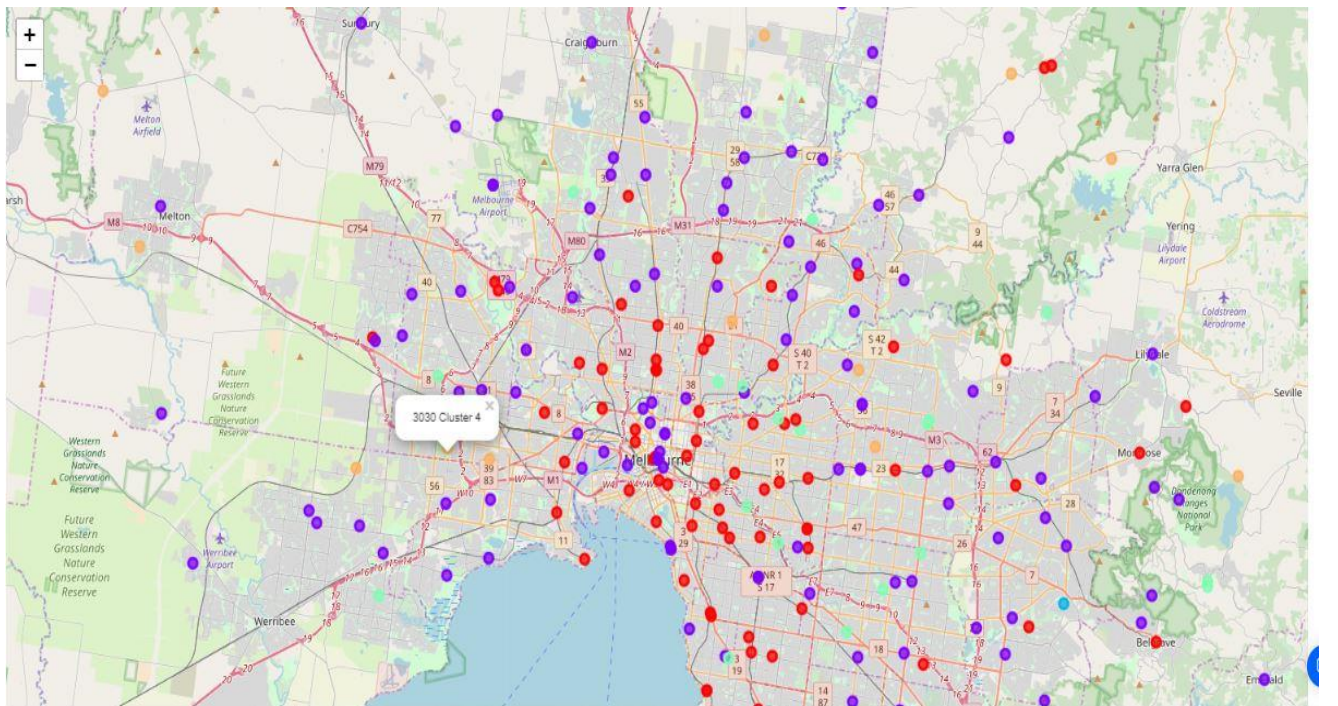
A sample of the final dataframe for Melbourne

### 3.3- Visualizing the results

We used the Folium library to visualize the various clusters of neighborhoods for both Adelaide and Melbourne.



Adelaide map



Melbourne map

## 4- Results and Discussion

### 4.1- Examining the clusters

To get an insight about what separates the clusters of neighborhoods from each other, we examined each of them carefully and looked at the most common venues around them. Using this information, we were also able to predict who might find these venues appealing.

So, here we provide the characteristic venues around the clusters as well as a sample of some of the neighborhoods present in the cluster. To get the table of all neighborhoods in Adelaide and Melbourne that belong to a cluster, visit the link [https://github.com/Mrigangka1998/Coursera\\_Capstone/blob/master/The\\_Battle\\_of\\_Neighborhoods/The%20Battle%20of%20Neighborhoods.ipynb](https://github.com/Mrigangka1998/Coursera_Capstone/blob/master/The_Battle_of_Neighborhoods/The%20Battle%20of%20Neighborhoods.ipynb).

## Cluster 0 - Many cafes, restaurants, bars

### Adelaide

	PostalCode	Neighborhood	Local government area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	5008	Croydon, Devon Park, Renown Park, Ridleyton, W...	City of Charles Sturt	0	Supermarket	Train Station	Gluten-free Restaurant	Liquor Store	Café
1	5010	Angle Park, Ferryden Park, Regency Park	City of Port Adelaide Enfield	0	Restaurant	Racetrack	Athletics & Sports	Park	Sports Bar
2	5034	Clarence Park, Goodwood, Kings Park, Millswood...	City of Unley	0	Train Station	Outdoors & Recreation	Coffee Shop	Café	Zoo Exhibit
3	5040	Novar Gardens	City of West Torrens	0	Grocery Store	Fast Food Restaurant	Bus Station	Café	Football Stadium
4	5041	Colonel Light Gardens, Cumberland Park, Daw Pa...	City of Mitcham	0	Grocery Store	Fish & Chips Shop	Fast Food Restaurant	Café	Football Stadium
5	5043	Ascot Park, Marion, Mitchell Park, Morphettvil...	City of Marion	0	Train Station	Liquor Store	IT Services	Café	Zoo Exhibit

### Melbourne

	PostalCode	Neighborhood	Local government area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	3000	Melbourne CBD	City of Melbourne	0	Café	Dessert Shop	Coffee Shop	Bar	Italian Restaurant
1	3002	East Melbourne	City of Melbourne	0	Hotel	Café	Wine Bar	Park	Convenience Store
2	3006	Southbank, South Wharf	City of Melbourne; City of Port Phillip	0	Café	Bar	Hotel	Grocery Store	Bakery
3	3011	Footscray, Seddon	City of Maribyrnong	0	Café	Skate Park	Campground	Zoo Exhibit	Fish Market
4	3013	Yarraville	City of Maribyrnong	0	Café	Pizza Place	Grocery Store	Burger Joint	Asian Restaurant
5	3015	Newport, Spotswood, South Kingsville	City of Hobsons Bay	0	Café	Pizza Place	Thai Restaurant	Thrift / Vintage Store	Bagel Shop

## Cluster 1 – Many supermarkets, groceries, furniture and departmental stores

### Adelaide

	PostalCode	Neighborhood	Local government area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	5000	Adelaide	City of Adelaide	1	Café	Tea Room	Pizza Place	Asian Restaurant	Wine Bar
1	5006	North Adelaide	City of Adelaide	1	Pub	Burger Joint	Italian Restaurant	Café	Thai Restaurant
2	5007	Bowden, Brompton, Hindmarsh, Welland, West Hin...	City of Charles Sturt	1	Breakfast Spot	Beer Store	Pizza Place	Gym	Café
3	5009	Allenby Gardens, Beverley, Kilkenny	City of Charles Sturt	1	Vietnamese Restaurant	Gym	Burger Joint	Home Service	Beach
4	5011	St Clair, Woodville, Woodville Park, Woodville...	City of Charles Sturt	1	Sushi Restaurant	Grocery Store	Train Station	Mediterranean Restaurant	Shopping Mall

### Melbourne

	PostalCode	Neighborhood	Local government area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	3003	West Melbourne	City of Melbourne	1	Flea Market	Asian Restaurant	Fish Market	Flower Shop	Food
1	3004	Melbourne CBD (St Kilda Road area)	City of Melbourne (east side to High Street, P...	1	Café	Cocktail Bar	Dessert Shop	Bar	Italian Restaurant
2	3008	Docklands	City of Melbourne	1	Middle Eastern Restaurant	Bar	Steakhouse	Coffee Shop	Café
3	3018	Altona, Seaholme	City of Hobsons Bay	1	Diner	Bar	Performing Arts Venue	Pizza Place	Fish & Chips Shop
4	3019	Braybrook	City of Maribymong	1	Pizza Place	Sporting Goods Shop	Pub	Grocery Store	Zoo Exhibit
5	3020	Albion, Sunshine, Sunshine North, Sunshine West	City of Brimbank	1	Furniture / Home Store	Grocery Store	Train Station	Vietnamese Restaurant	Pet Store

## Cluster 2 – Sports/Athletics venues and stadiums

### Adelaide

	PostalCode	Neighborhood	Local government area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	5091	Banksia Park, Tea Tree Gully, Vista	City of Tea Tree Gully	2	Athletics & Sports	Football Stadium	Zoo Exhibit	Fish Market	Flower Shop
1	5131	Houghton, Upper Hermitage	City of Tea Tree Gully	2	Athletics & Sports	Zoo Exhibit	Fish Market	Flower Shop	Food

### Melbourne

	PostalCode	Neighborhood	Local government area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	3335	Bonnie Brook, Grangefields, Plumpton, Rockbank...	City of Melton	2	Athletics & Sports	Zoo Exhibit	Fish Market	Flower Shop	Food

## Cluster 3 – Many parks and playgrounds

### Adelaide

	PostalCode	Neighborhood	Local government area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	5020	West Lakes Shore	City of Charles Sturt	3	Bus Station	Park	Playground	Zoo Exhibit	Flower Shop
1	5049	Kingston Park, Seacliff, Seacliff Park, Marino...	City of Holdfast Bay	3	Park	Harbor / Marina	Train Station	Campground	IT Services
2	5050	Bellevue Heights, Eden Hills	City of Mitcham	3	Park	Soccer Field	Furniture / Home Store	Fruit & Vegetable Store	Frozen Yogurt Shop
3	5068	Kensington Gardens, Kensington Park, Leabrook,...	City of Burnside	3	Pharmacy	Bus Stop	Dog Run	Park	Flower Shop
4	5070	Felixstow, Firle, Glynde, Joslin, Marden, Payn...	City of Norwood Payneham St Peters	3	Park	Sandwich Place	Playground	Bus Station	Zoo Exhibit
5	5076	Athelstone	City of Campbelltown	3	Pizza Place	Grocery Store	Supermarket	Park	Flea Market



## Melbourne

	PostalCode	Neighborhood	Local government area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	3023	Cairnlea, Deer Park, Burnside, Burnside Heigh...	City of Brimbank	3	Park	Liquor Store	Pub	Zoo Exhibit	Flea Market
1	3049	Attwood, Westmeadows	City of Hume	3	Park	Playground	Fish & Chips Shop	Flea Market	Flower Shop
2	3059	Greenvale	City of Hume	3	Park	Fast Food Restaurant	Tennis Court	Convenience Store	Zoo Exhibit
3	3070	Northcote	City of Darebin	3	Food Truck	Playground	Convenience Store	Park	Pool
4	3073	Reservoir	City of Darebin	3	Park	Playground	Dog Run	Zoo Exhibit	Flower Shop
5	3102	Kew East	City of Boroondara	3	Paper / Office Supplies Store	Playground	Café	Park	Flea Market

## Cluster 4 – Sparsely populated areas on the outskirts/countryside

## Adelaide

	PostalCode	Neighborhood	Local government area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	5019	Semaphore Park, Exeter, Semaphore, Semaphore S...	City of Charles Sturt	4	NaN	NaN	NaN	NaN	NaN
1	5025	Flinders Park, Kidman Park	City of Charles Sturt	4	NaN	NaN	NaN	NaN	NaN
2	5089	Highbury	City of Tea Tree Gully	4	NaN	NaN	NaN	NaN	NaN
3	5094	Dry Creek, Gepps Cross, Cavan, Dry Creek	City of Port Adelaide Enfield	4	NaN	NaN	NaN	NaN	NaN
4	5096	Gulfview Heights, Para Hills, Para Hills West,...	City of Salisbury	4	NaN	NaN	NaN	NaN	NaN
5	5110	Waterloo Corner, Bolivar, Burton, Direk, Globe...	City of Playford	4	NaN	NaN	NaN	NaN	NaN

## Melbourne

	PostalCode	Neighborhood	Local government area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	3012	Brooklyn, Kingsville, Maidstone, Tottenham, We...	City of Brimbank; City of Hobsons Bay	4	NaN	NaN	NaN	NaN	NaN
1	3029	Truganina, Hoppers Crossing, Tarneit	City of Melton; City of Wyndham	4	NaN	NaN	NaN	NaN	NaN
2	3030	Derrimut, Point Cook, Werribee, Werribee South...	City of Brimbank	4	NaN	NaN	NaN	NaN	NaN
3	3089	Diamond Creek	Shire of Nillumbik	4	NaN	NaN	NaN	NaN	NaN
4	3105	Bulleen	City of Manningham	4	NaN	NaN	NaN	NaN	NaN
5	3338	Brookfield, Cobblebank, Exford, Eynesbury, Mel...	City of Melton	4	NaN	NaN	NaN	NaN	NaN

## 5- Conclusion

The purpose of this project was to find similar neighborhoods in **Adelaide** and **Melbourne** so that it may be beneficial to people who are planning to move from one of those cities to the other. By using **k-means clustering**, we were able to divide the neighborhoods into 5 different clusters based on the **types of venues around them**, using **location data obtained from Foursquare**. Also, by closely examining the clusters, we were able to find the most common venues around the neighborhoods in each of these clusters which helped us in finding the chief characteristics of the regions. This may also be useful to people as they may be looking to move to neighborhoods of a particular type based on their personal preferences.



Ultimately, this information can be helpful to guide people moving between these two cities and looking for suitable localities. It could help them make further inquiries about these areas and save precious time by narrowing down possible neighborhoods where they might want to stay.