Moving to Melbourne – Australia

By Wael Sarriss

May 23, 2020

1. Introduction

1.1 Background

Immigration has existed during the whole history of humankind and people have moved from their home places for centuries, for all sorts of reasons. While in the old days, the reasons behind immigration were mainly push factors, families running away from their homes for political reasons or climatic disasters, nowadays many of the families and individuals choosing to immigrate are seeking either a healthier life, a change of scenery and culture, or a better job opportunity.

Like many others from my generation, I recently have been granted a permanent residency visa to Australia, and while I am happy to choose to start a new challenge in my life, I feel stuck when it comes to taking the first decision. Where to live in Melbourne?

1.2 The Problem

Melbourne is the Capital of the state of Victoria, and the 2nd most populous city in Australia. With a size of 9993 Km^2, and 31 municipalities, it is an urban agglomeration with many metropolitan areas. The decision to move to Melbourne was an easy one for me, since many of my acquaintances and connections have moved there recently. But unlike them, my decision to choose which area to live in is not going to be based on the location of my job. If the recent pandemic events taught us anything, it is that the neighborhood we choose to live in must be accompanied with all the necessary essentials and that the "Working from Home" scenario might just become the new normal. My decision when choosing the neighborhood to live in shall be based on the following factors:

1. I want the neighborhood I live in to have a nearby hospital with emergency care in case an emergency occurs.
2. I want the neighborhood I live in to have a selection of highly rated primary schools for enrolling my kids.
3. I want the neighborhood I choose to live in not to be densely neither sparsely populated, preferably between 1000 and 3000 people per Square Kilometer.
4. 4. I want the neighborhood I live in to have recreational areas such as parks, pools, playgrounds and so on with a selection of venues such as markets, cafes, etc.

1.3 Interest

Anyone looking to immigrate to a new city/country will go through the same decision process. Some might choose to live nearby a family member or a friend, or even in a location close to their new job or business, but many others might not prefer or find these options, and those individuals or families might find this project study very useful for their decision process.

2. Data acquisition and cleaning

2.1 Data sources

All the data required for this project shall be collected from reliable sources and shall be recently updated to give accurate results.

List of Hospitals in Melbourne

I collected the list of hospitals in Melbourne from the Department of Health of the State of Victoria, Australia. The dataset includes the Hospital Name, address, location, postcode, and whether emergency capable or not. The dataset can be found in a downloadable CSV format here.

List of Schools in Melbourne

The list of schools I used on this project was collected from a private entity called Better education. The online platform of Better Education is the most popular school directory and education information website in Australia. It includes the Overall school Rating, Total student enrollment, School address and postcode. The data can be found here. I read the data using the pandas read HTML function.

List of Australian Cities and Neighborhoods

The Australian Statistical Geography Standard (ASGS) splits Australia into multiple territories starting from the geographical coordinates of the States, going deeper into 5 spatial regions from SA4 to SA1 and ending with the mesh blocks. My list of Australian Neighborhoods or suburbs will be the list of SA2 regions from the ASGS. Statistical Areas Level 2 (SA2) are medium-sized general-purpose areas built up from SA1. Their purpose is to represent a community that interacts together socially and economically. All the required data can be downloaded in CSV format from the Australian Bureau of Statistics here and here.

Population and Median Age Data

The list of Australian Neighborhoods and Cities included the total Square Kilometer (SQKM) size of the neighborhoods. By finding the total population of each neighborhood, we can calculate the Population density of each neighborhood and then filter based on the acceptable range of 1000 to 3000 people per SQKM. The population data was found on the Australian Bureau of Statistics Website – CENSUS Data and was downloaded through the following link by replacing the SA2Code below by the Main SA2 Code from the Neighborhood list:

https://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/{SA2CODE}?opendocument

List of Venues in each Neighborhood

The final piece of data required was the list of neighborhood venues and the category of each venue. This data was downloaded using the Foursquare API. Foursquare is a technology company that built a massive dataset of location data. They crowd-sourced their data and had people use their app to build their dataset and add venues and complete any missing information they had in their dataset. Currently its location data is the most comprehensive out there, and quite accurate that it powers location data for many popular services.

2.2 Data Analysis and Cleaning

Cleaning the data sets and extracting the required information out of each was the most time-consuming step in this project.

Cleaning the Hospitals Dataset

I first started with the Hospital Data Set. After reading the data into a data frame, I dropped any columns with missing data and then filtered the results to match my criteria by finding the Hospitals that were capable of handling emergency cases. In total 40 Hospitals where found across Victoria. Here is how the first 5 rows of the Hospital Data frame look like:

	Hospital ID	Formal Name	Other Name	Emergency Capable	Location Address	Suburb	Postcode	Category	Agency Type
0	5488	Albert Road Clinic	Albert Road Clinic	NO	31-33 Albert Road	South Melbourne	3205	PRIVATE	Private Hospital
1	3485	Albury Wodonga Health	Albury Wodonga Health	YES	69 Vermont Street	Wodonga	3690	PUBLIC	Public Hospital
2	12990	Albury Wodonga Health, Albury Campus	Albury Wodonga Health, Albury Campus	YES	Borella	Albury	2640	PUBLIC	Public Hospital
3	3491	Alexandra District Health	Alexandra District Health	NO	20 Cooper Street	Alexandra	3714	PUBLIC	Public Hospital
4	11519	Alfred Health	Alfred Health	NO	Commercial Road	Melbourne	3004	PUBLIC	Public Hospital

Figure 1 - Hospital Dataframe

Cleaning the Schools Dataset

Next, I read the Schools Data into a Data frame and start by cleaning the data. I first removed the State name and postcode from the School Name, and then dropped any Columns that were unnecessary. I then filtered the list schools by finding the Schools that shared the same Postcode as that of the Emergency Capable Hospitals. Since the Schools Data is of Melbourne only, this step automatically eliminated all postcodes outside of Melbourne. The result was a Data frame with a total of 29 Schools. I further sorted the data by grouping the schools with the same postcode and finding the Average rating of the School per postcode. I then chose the Top 10 School Rated Postcodes as my preference list for the next step. The resulting School Data frame looks as follows:

	School	Postcode	State Overall Score	Total Enrolments
0	Tintern Grammar,Ringwood East	3135	99	812
1	Melbourne Grammar School	3004	98	1795
2	St Margaret's School,Berwick	3806	98	746
3	Heathdale Christian College,Werribee	3030	98	1537
4	Mountain Gate Primary School	3156	98	404

Figure 2 - Schools Dataframe before grouping

Postcode	State Overall Score	Total Enrolments
3156	98.000000	404.000000
3004	97.500000	2470.000000
3030	97.500000	2080.500000
3806	97.000000	836.500000
3199	96.000000	864.000000
3128	95.500000	459.500000
3135	95.333333	649.666667
3084	95.200000	592.800000
3191	94.333333	438.000000
3121	94.000000	327.500000

Figure 3 - Schools Dataframe after grouping

Cleaning the Neighborhoods ad Cities Datasets

After downloading the CSV files from the Australian Bureau of Statistics Webpage, I read the files into Data frames and start the cleaning process. At first, I use the list of Localities and SA3 and SA4 regions to develop a list of the preferred localities with respect to their postcodes. I start by filtering the list into those within the State of Victoria, then into the regions within The Greater Melbourne City. A total of 747 localities were found. I then filter those localities to include the ones that share the same postcodes as that of the Top 10 School rated postcodes. From this list, I generate a list of SA3 Codes that shall be used to filter the smaller community regions within. Figures 4 and 5 below show the Data Frame and the Distribution of Region across Melbourne.

| id	postcode	locality	state	long	lat	sa3	sa3name	sa4	sa4name
4750	3004	MELBOURNE	VIC	144.970161	-37.844246	20605	Port Phillip	206	Melbourne - Inner
4751	3004	ST KILDA ROAD CENTRAL	VIC	144.970161	-37.844246	20605	Port Phillip	206	Melbourne - Inner
4809	3030	CHARTWELL	VIC	144.642090	-37.916240	21305	Wyndham	213	Melbourne - West
4810	3030	COCOROC	VIC	144.642090	-37.916240	21305	Wyndham	213	Melbourne - West
4811	3030	DERRIMUT	VIC	144.642090	-37.916240	21305	Wyndham	213	Melbourne - West

Figure 4 - Preferred Localities across Melbourne
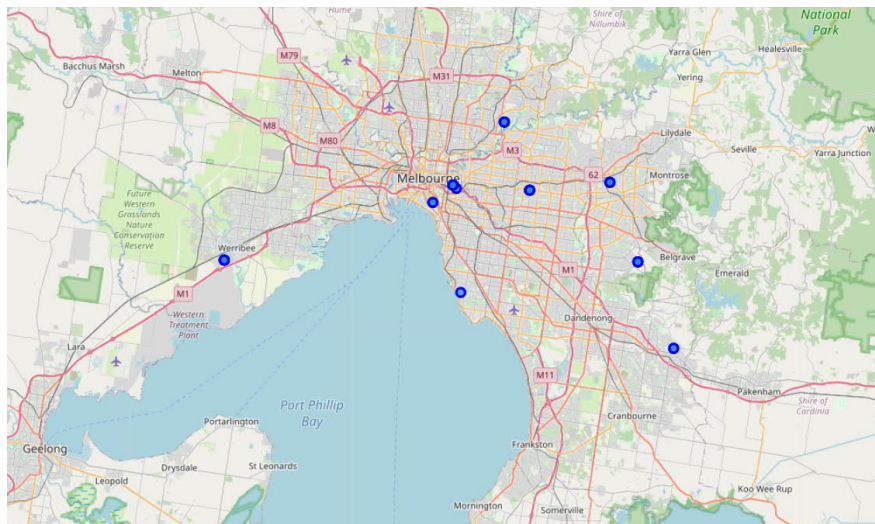
Figure 5 - Preferred Localities in Melbourne

Next, I use the list of SA2 regions, which are essentially our Neighborhood Regions, and I start by filtering the list by the SA2 regions within Victoria State. I then Use the list of SA3 Codes generated earlier to find the SA2 regions within. A total of 75 regions where found.

Adding the Population Data to the Dataset

To the list of Neighborhoods Generated above, I utilize the information available on the Australian Bureau of Statistics to find the Total Populations of each SA2 Region and the Median Age. From that information and the Square Kilometer area already available in our list, I calculate the Population density of each region. I then clean the data by filtering according to the preferred range of population density which is between 1000 and 3000 people per SQKM. A Total of 48 Neighborhoods where found within that range.

I then use that list to add the Geographical Coordinates of each region using the Geopy Library. As Geopy could not find the coordinates of some areas, and found the wrong coordinates of some different areas, those values had to be inputted manually. Figures 6 and 7 below show the first few rows of the resulting Data Frame and the Distribution of Preferred Neighborhoods (SA2 Regions) across Melbourne.

|	Main_ID	ID	Neighborhood	sa3code	Area	Region	State	SQKM	Population	Median_age	Density	Latitude	Longitude
0	207031167.0	21167.0	Surrey Hills (East) - Mont Albert	20703.0	Whitehorse - West	Melbourne - Inner East	Victoria	3.3279	9965.0	40.0	2994.38	-37.824118	145.098621
1	207031163.0	21163.0	Box Hill	20703.0	Whitehorse - West	Melbourne - Inner East	Victoria	7.0120	19828.0	33.0	2827.72	-37.813703	145.123805
2	207031164.0	21164.0	Box Hill North	20703.0	Whitehorse - West	Melbourne - Inner East	Victoria	6.2087	17552.0	38.0	2827.00	-37.805683	145.129575
3	208011170.0	21170.0	Brighton East	20801.0	Bayside	Melbourne - Inner South	Victoria	5.6083	15838.0	43.0	2824.03	-37.917173	145.016366
4	213051362.0	21362.0	Hoppers Crossing - South	21305.0	Wyndham	Melbourne - West	Victoria	7.0114	19477.0	36.0	2777.90	-37.883250	144.701102
5	208011169.0	21169.0	Brighton (Vic.)	20801.0	Bayside	Melbourne - Inner South	Victoria	8.2005	22695.0	45.0	2767.51	-37.908196	144.995799

Figure 6 - Preferred Neighborhood Dataframe
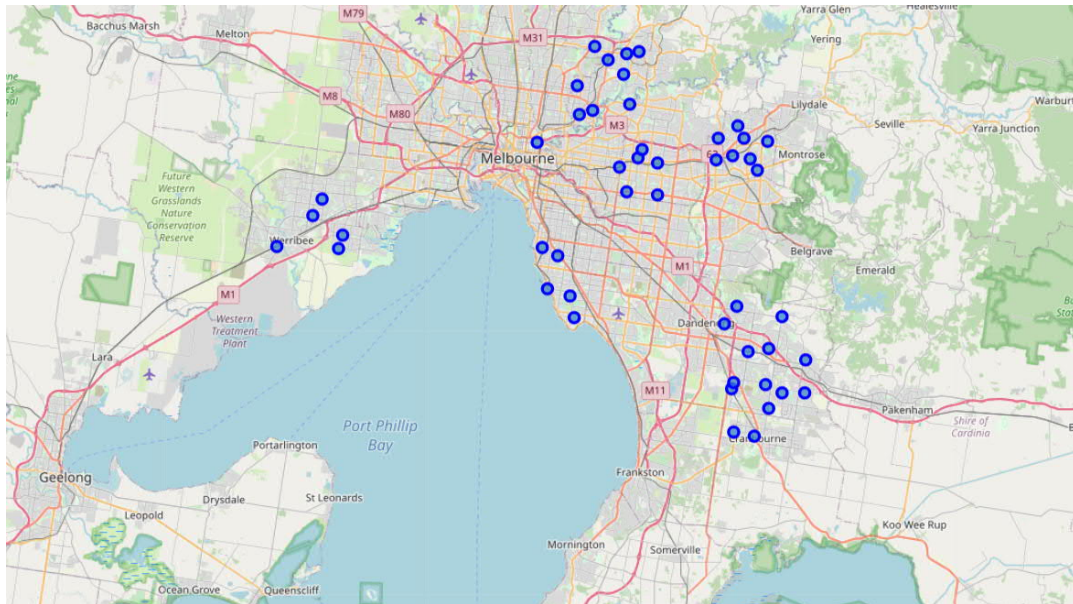
Figure 7 - Preferred Neighborhoods in Melbourne

Exploring the Nearby Venues within each Neighborhood

Finally, with the Data Frame of the Neighborhoods ready, I was able to use the Foursquare API to explore the nearby venues within a radius of 1 Km from each Neighborhood location. The resulting Data frame was of 847 Rows and 8 Columns. Here are the first 5 rows of the Data frame:

	ID	Neighborhood	Neighborhood_Latitude	Neighborhood_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Category
0	21167	Surrey Hills (East) - Mont Albert	-37.824118	145.098621	Burger Burger	-37.823760	145.097880	Burger Joint
1	21167	Surrey Hills (East) - Mont Albert	-37.824118	145.098621	Reunion	-37.823200	145.097570	Café
2	21167	Surrey Hills (East) - Mont Albert	-37.824118	145.098621	Old Kingdom 大中樓	-37.826446	145.102303	Peking Duck Restaurant
3	21167	Surrey Hills (East) - Mont Albert	-37.824118	145.098621	Acorn Nursery and Oaks Café	-37.826085	145.101035	Garden Center
4	21167	Surrey Hills (East) - Mont Albert	-37.824118	145.098621	Zimt	-37.822709	145.098002	Café

Figure 8 - Venues in Melbourne

2.3 Preparing the Data

Now that I have my Data frame ready, it's time to prepare it for the Clustering model.

The first thing I notice is that the venue categories are very specific, such as there are many types of restaurants and recreational areas listed in specific categories. In order to get accurate results with my clustering model, I had to rename all the different recreational areas (my preferred venue according to my initial statement) under one common name:

```
Recreational = ['Garden', 'Beach', 'Playground', 'Basketball Court', 'Recreation Center', 'Lake', 'Field', 'Pool', 'Water Park', 'Skate Park', 'Arcade' , 'Park']
```

Any venue category from the list above was renamed into Recreational Area as they all serve the same purpose to me.

Figure 9 below shows a distribution of the venues across Melbourne.
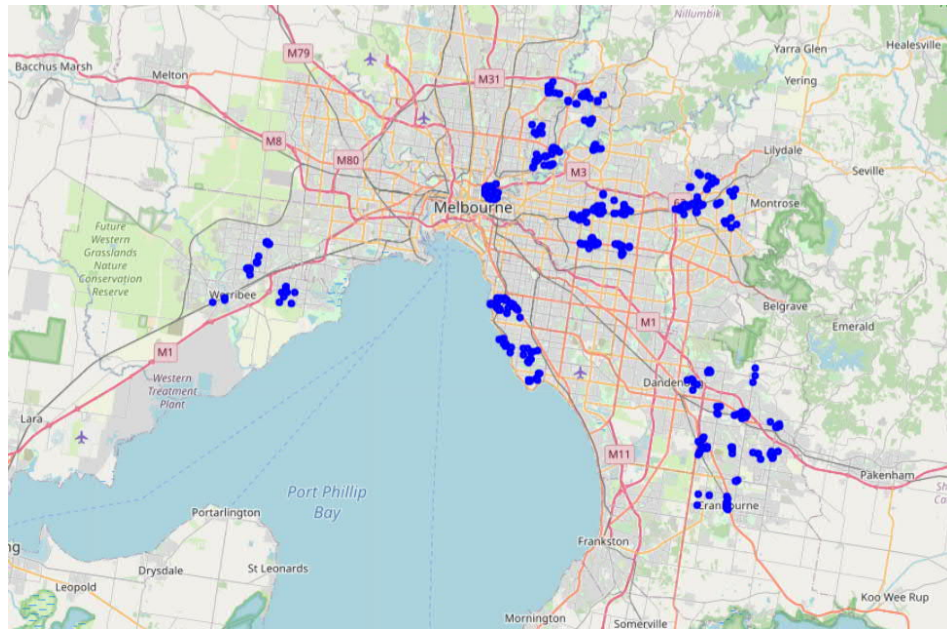

Figure 9 - Venues across Melbourne

Next, I use one-hot encoding to create a data frame of all the neighborhoods with frequency of occurrence of each venue category within that neighborhood. The resulting Data frame was of 48 Rows (# of Neighborhoods) and 157 Columns (# of Venue Categories). Figure 10 below shows a selection of resulting data frame.

	Neighborhoods	Women's Store	African Restaurant	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	Auto Garage	BBQ Joint	Bakery	Bar	Beer Bar	Beer Garden	Big Box Store	Board Shop	Bookstore	Bowling Alley
0	Bayswater North	0.0	0.0	0.0	0.0	0.125	0.00	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	Beaumaris	0.0	0.0	0.0	0.0	0.000	0.00	0.0	0.055556	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	Berwick - North	0.0	0.0	0.0	0.0	0.000	0.00	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	Berwick - South	0.0	0.0	0.0	0.0	0.000	0.08	0.0	0.040000	0.0	0.0	0.040000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	Blackburn	0.0	0.0	0.0	0.0	0.000	0.00	0.0	0.000000	0.0	0.0	0.045455	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Figure 10 - Frequency of Occurrence of venues in each Neighborhood

For the sake of clarity and ease of read and use while identifying the clusters, I create a data frame with each neighborhood along with the top 5 most common venue categories in each. Figure 11 shows the results.

	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bayswater North	Sandwich Place	Sporting Goods Shop	Arts & Crafts Store	Fast Food Restaurant	Recreational Area
1	Beaumaris	Café	Grocery Store	Chinese Restaurant	Sports Club	Recreational Area
2	Berwick - North	Café	Road	Supermarket	Grocery Store	Italian Restaurant
3	Berwick - South	Fast Food Restaurant	Café	Asian Restaurant	Shopping Mall	Bakery
4	Blackburn	Café	Hotel Bar	Sushi Restaurant	Grocery Store	Indian Restaurant

Figure 11 - The top 5 most common Venue Categories in each Neighborhood

3. Clustering the Neighborhoods using K-means

Now that the date is ready, it's time to use Machine learning to cluster the Neighborhoods according to the different type of venues in each. I decide to use K-means clustering approach.

3.1 Finding the best K value

Before attempting to cluster the data, I must first find the best K value to use in order to get reasonable and acceptable results. This can be done in many ways, either by trial and error, were I run the data through different K values until I find the most reasonable results, or I can use the Sum of Squared distances method to find the optimum K value. I decide to go with the mathematical approach. I run the data through a range of K values from 1 to 20 and for each run I calculate the sum of squared distances. Finally, I plot the results to find the optimal K value. Figure 12 below shows the plot.
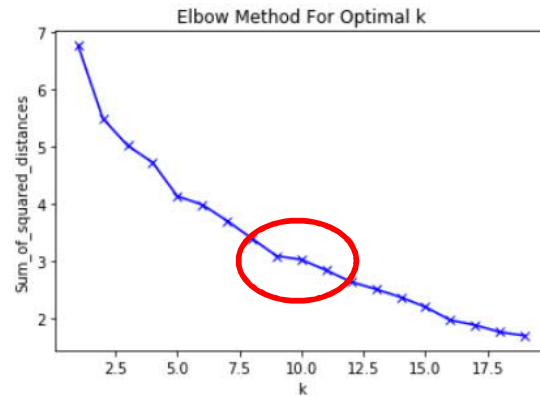
Figure 12 - Finding the optimal K value

From the resulting plot, the optimal K value can be determined from the elbow of the graph. Although the graph doesn't resemble and arm, yet it helps determine the possible optimal K values that can be used to cluster the data. I decide to go with K = 10.

3.2 Using K-means Clustering with K = 10

Using SKlearn library, I cluster my data using K means with K = 10. The resulting data set shows the resulting counts per each cluster.

Clearly, we see some outlier clusters (0, 4, 8, and 9). These clusters will be dismissed from our final analysis. Figure 13 shows the distribution of the clusters on the Map of Melbourne.

Cluster	# of Neighborhoods
0	1
1	2
2	11
3	8
4	1
5	2
6	10
7	11
8	1
9	1

Figure 13 - The resulting Clusters

3.3 Examining the Clusters

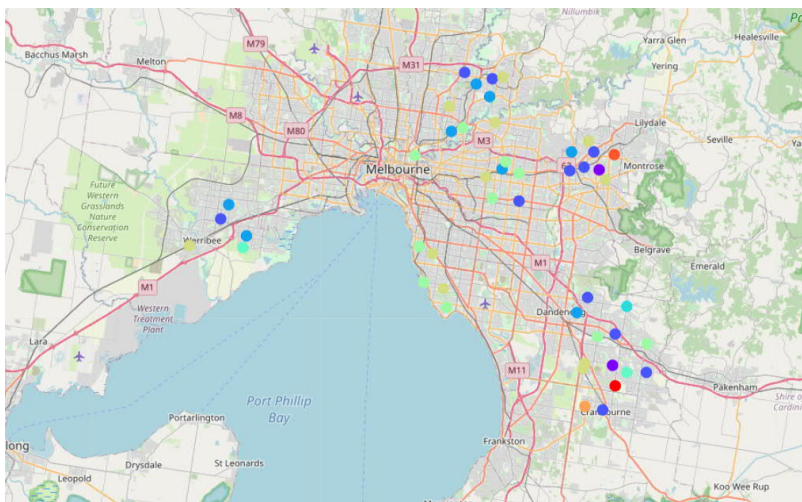Keeping our initial decision terms in mind, I will look for the cluster that is most appealing and satisfies our requirements. Since we determined that Clusters 0, 4, 8, and 9 are outliers, I will not be examining them.

Cluster 1

	ID	Neighborhoods	Area	Region	Median_age	Density	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	21458	Narre Warren South (West)	Casey - South	Melbourne - South East	32	2476.46	1	Supermarket	Video Store	Indian Restaurant	Grocery Store	Fast Food Restaurant
1	21452	Croydon South	Maroondah	Melbourne - Outer East	37	1650.91	1	American Restaurant	Video Store	Grocery Store	Fast Food Restaurant	Shopping Mall

The most common venues in Cluster 1 are mainly Fast-food Restaurants, Video Stores and Grocery Store, none of which are appealing to us.

Cluster 2

	ID	Neighborhoods	Area	Region	Median_age	Density	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	21196	Bundoora - East	Banyule	Melbourne - North East	39	2428.04	2	Pizza Place	Supermarket	Fast Food Restaurant	Indian Restaurant	Light Rail Station
1	21166	Burwood East	Whitehorse - West	Melbourne - Inner East	40	2409.81	2	Fast Food Restaurant	Café	Bakery	Gas Station	Supermarket
2	21454	Endeavour Hills - South	Casey - North	Melbourne - South East	38	2390.60	2	Supermarket	Department Store	Fast Food Restaurant	Food Truck	Discount Store
3	21267	Ringwood East	Maroondah	Melbourne - Outer East	39	2156.95	2	Café	Fast Food Restaurant	Bar	Fish & Chips Shop	Seafood Restaurant
4	21197	Greensborough	Banyule	Melbourne - North East	40	2026.56	2	Café	Mexican Restaurant	Fast Food Restaurant	Supermarket	Electronics Store
5	21451	Croydon - West	Maroondah	Melbourne - Outer East	40	1963.99	2	Fast Food Restaurant	Pizza Place	Australian Restaurant	Supermarket	Beer Garden
6	21361	Hoppers Crossing - North	Wyndham	Melbourne - West	35	1761.47	2	Fast Food Restaurant	Café	Portuguese Restaurant	Pizza Place	Dumpling Restaurant
7	21294	Berwick - South	Casey - North	Melbourne - South East	33	1736.83	2	Fast Food Restaurant	Café	Asian Restaurant	Shopping Mall	Bakery
8	21266	Ringwood	Maroondah	Melbourne - Outer East	37	1730.39	2	Supermarket	Fast Food Restaurant	Coffee Shop	Gastropub	Restaurant
9	21300	Cranbourne	Casey - South	Melbourne - South East	34	1542.49	2	Fast Food Restaurant	Café	Ice Cream Shop	Department Store	Grocery Store
10	21456	Narre Warren - South West	Casey - North	Melbourne - South East	34	1498.70	2	Fast Food Restaurant	Café	Department Store	Grocery Store	Paper / Office Supplies Store

Cluster 2 holds 11 of our Neighborhoods were the most common venue categories are mostly Cafes and Restaurants. This Cluster might be appealing to start a business within these neighborhoods, but not what I am looking for to live in as there aren't any recreational facilities within the most common venues.

Cluster 3

	ID	Neighborhoods	Area	Region	Median_age	Density	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	21163	Box Hill	Whitehorse - West	Melbourne - Inner East	33	2827.72	3	Chinese Restaurant	Asian Restaurant	Korean Restaurant	Café	Dumpling Restaurant
1	21362	Hoppers Crossing - South	Wyndham	Melbourne - West	36	2777.90	3	Pizza Place	Grocery Store	Café	Indian Chinese Restaurant	Doctor's Office
2	21465	Point Cook - North	Wyndham	Melbourne - West	33	2730.08	3	Supermarket	Pharmacy	Japanese Restaurant	Fish & Chips Shop	Shopping Mall
3	21204	Watsonia	Banyule	Melbourne - North East	39	2407.02	3	Grocery Store	Pizza Place	Australian Restaurant	Thai Restaurant	Wine Shop
4	21200	Ivanhoe	Banyule	Melbourne - North East	39	2255.49	3	Japanese Restaurant	Train Station	Pizza Place	Coffee Shop	Café
5	21295	Doveton	Casey - North	Melbourne - South East	34	2176.99	3	Fish & Chips Shop	Recreational Area	Grocery Store	Shopping Mall	Farm
6	21268	Ringwood North	Maroondah	Melbourne - Outer East	41	2013.23	3	Bakery	Asian Restaurant	Grocery Store	Shopping Mall	Coffee Shop
7	21203	Viewbank - Yallambie	Banyule	Melbourne - North East	40	1787.77	3	Fish Market	Supermarket	Pub	Grocery Store	Golf Course

Cluster 3 can be considered as the Cluster of Neighborhoods were the most common venue categories are Markets, mainly Supermarkets, and Grocery Stores or shopping malls. Another interesting cluster of Neighborhoods, but I can barely see any recreational area within the cluster's most common venues.

Cluster 5

	ID	Neighborhoods	Area	Region	Median_age	Density	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	21457	Narre Warren South (East)	Casey - South	Melbourne - South East	34	2109.51	5	Recreational Area	Wine Shop	Electronics Store	Food	Flower Shop
1	21466	Point Cook - South	Wyndham	Melbourne - West	31	1531.81	5	Recreational Area	Brewery	Farm	Food & Drink Shop	Food

Cluster 5 is a cluster of 2 Neighborhoods where the most common venue category in both is Recreational Areas. This fulfills the first requirement of our selection criteria, yet the rest of the common venue categories are not so appealing. Further examination of this cluster might be required.

Cluster 6

	ID	Neighborhoods	Area	Region	Median_age	Density	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	21164	Box Hill North	Whitehorse - West	Melbourne - Inner East	38	2827.00	6	Middle Eastern Restaurant	Italian Restaurant	Dance Studio	Café	Plaza
1	21169	Brighton (Vic.)	Bayside	Melbourne - Inner South	45	2767.51	6	Café	Convenience Store	Movie Theater	Chinese Restaurant	Restaurant
2	21173	Sandringham - Black Rock	Bayside	Melbourne - Inner South	45	2566.99	6	Café	Recreational Area	Pizza Place	Thai Restaurant	Sports Bar
3	21161	Blackburn	Whitehorse - West	Melbourne - Inner East	39	2484.05	6	Café	Hotel Bar	Sushi Restaurant	Grocery Store	Indian Restaurant
4	21165	Burwood	Whitehorse - West	Melbourne - Inner East	29	2430.54	6	Café	Light Rail Station	Gym	Gas Station	College Rec Center
5	21168	Beaumaris	Bayside	Melbourne - Inner South	46	2257.45	6	Café	Grocery Store	Chinese Restaurant	Sports Club	Recreational Area
6	21201	Ivanhoe East - Eaglemont	Banyule	Melbourne - North East	44	1798.27	6	Café	Pizza Place	Grocery Store	Italian Restaurant	Fast Food Restaurant
7	21145	Yarra - North	Yarra	Melbourne - Inner	36	1736.26	6	Café	Bar	Pub	Cocktail Bar	Gastropub
8	21297	Hallam	Casey - North	Melbourne - South East	35	1335.69	6	Convenience Store	Café	Train Station	Music Venue	Wine Shop
9	21293	Berwick - North	Casey - North	Melbourne - South East	42	1159.31	6	Café	Road	Supermarket	Grocery Store	Italian Restaurant

Cluster 6 can be considered as the Cluster of Neighborhoods were the most common venue categories are Cafes and Bars, similar to cluster 2, another interesting cluster of Neighborhoods to start as business in but not what I am looking for to live in.

Cluster 7

	ID	Neighborhoods	Area	Region	Median_age	Density	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	21167	Surrey Hills (East) - Mont Albert	Whitehorse - West	Melbourne - Inner East	40	2994.38	7	Recreational Area	Burger Joint	Train Station	Café	Bakery
1	21170	Brighton East	Bayside	Melbourne - Inner South	43	2824.03	7	Café	Gym / Fitness Center	Recreational Area	Thai Restaurant	Fish & Chips Shop
2	21467	Werribee - East	Wyndham	Melbourne - West	35	2560.43	7	Sports Bar	Football Stadium	Recreational Area	Gas Station	Wine Shop
3	21198	Heidelberg - Rosanna	Banyule	Melbourne - North East	39	2318.85	7	Greek Restaurant	Recreational Area	Miscellaneous Shop	Café	Pub
4	21199	Heidelberg West	Banyule	Melbourne - North East	35	2114.57	7	Recreational Area	Fish & Chips Shop	Convenience Store	Gym	Fried Chicken Joint
5	21306	Lynbrook - Lyndhurst	Casey - South	Melbourne - South East	32	2069.01	7	Recreational Area	Australian Restaurant	Burger Joint	Carpet Store	Shopping Mall
6	21171	Cheltenham - Highett (West)	Bayside	Melbourne - Inner South	40	2056.09	7	Gym	Recreational Area	Dance Studio	Brewery	Italian Restaurant
7	21305	Hampton Park - Lynbrook	Casey - South	Melbourne - South East	33	1913.46	7	Gym / Fitness Center	Bakery	Burger Joint	Chinese Restaurant	Sandwich Place
8	21265	Croydon Hills - Warranwood	Maroondah	Melbourne - Outer East	39	1894.75	7	Tennis Court	Photography Studio	Recreational Area	Theater	Dumpling Restaurant
9	21202	Montmorency - Briar Hill	Banyule	Melbourne - North East	41	1338.23	7	Deli / Bodega	Italian Restaurant	Recreational Area	Restaurant	Bakery
10	21263	Bayswater North	Maroondah	Melbourne - Outer East	37	1188.11	7	Sandwich Place	Sporting Goods Shop	Arts & Crafts Store	Fast Food Restaurant	Recreational Area

Cluster 7 is a Cluster of 11 Neighborhoods were Recreational Areas can be found in within the most 5 common venue categories across all of them, along with a selection of interesting other venues such as Cafes, Gyms, Restaurants, and other stores. This is Definitely an Interesting cluster of Neighborhoods to look for a place to live in. I choose cluster 7 for further examination.

Using the K-means Clustering method, I was able to identify 10 different clusters within our preferred neighborhoods and identify the most favorable cluster to my liking which is Cluster 7.

4. Final Results

Now, I can further explore the preferred cluster (7) and find out the location of each neighborhood and the proximity distance of the neighborhood location and the Schools within and the nearby Hospital.

Below is a summary of each Neighborhood within Cluster 7.

Neighborhood 0
Neighborhood Name: Surrey Hills (East) - Mont Albert
1st Most Common Venue: Recreational Area
2nd Most Common Venue: Burger Joint
3rd Most Common Venue: Train Station
4th Most Common Venue: Café
5th Most Common Venue: Bakery
Population Density: 2994.38
1st School Name: Kingswood College, Box Hill
2nd School Name: St Francis Xavier's Parish Primary School, Box Hill
Hospital Name: Box Hill Hospital

Neighborhood 1
Neighborhood Name: Brighton East
1st Most Common Venue: Café
2nd Most Common Venue: Gym / Fitness Center
3rd Most Common Venue: Recreational Area
4th Most Common Venue: Thai Restaurant
5th Most Common Venue: Fish & Chips Shop
Population Density: 2824.03
School Name: Sandringham East Primary School
School Name: Sacred Heart School, Sandringham
School Name: Sandringham Primary School
Hospital Name: Sandringham Hospital

Neighborhood 2
Neighborhood Name: Werribee - East
1st Most Common Venue: Sports Bar
2nd Most Common Venue: Football Stadium
3rd Most Common Venue: Recreational Area
4th Most Common Venue: Gas Station
5th Most Common Venue: Wine Shop
Population Density: 2560.43
School Name: Heathdale Christian College, Werribee
School Name: Alamanda K-9 College, Point Cook
Hospital Name: Werribee Mercy Hospital

Neighborhood Number: 3
Neighborhood Name: Heidelberg - Rosanna
1st Most Common Venue: Greek Restaurant
2nd Most Common Venue: Recreational Area
3rd Most Common Venue: Miscellaneous Shop
4th Most Common Venue: Café
5th Most Common Venue: Pub
Population Density: 2318.85
School Name: Banyule Primary School, Rosanna
School Name: Heidelberg Primary School
School Name: Viewbank Primary School
School Name: St Martin of Tours School, Rosanna
School Name: Rosanna Golf Links Primary School
Hospital Name: Austin Health - Austin Hospital
Hospital Name: Mercy Hospital for Women

Neighborhood Number: 4

Neighborhood Name:	Heidelberg West
1st Most Common Venue:	Recreational Area
2nd Most Common Venue:	Fish & Chips Shop
3rd Most Common Venue:	Convenience Store
4th Most Common Venue:	Gym
5th Most Common Venue:	Fried Chicken Joint
Population Density:	2114.57
School Name:	Banyule Primary School, Rosanna
School Name:	Heidelberg Primary School
School Name:	Viewbank Primary School
School Name:	St Martin of Tours School, Rosanna
School Name:	Rosanna Golf Links Primary School
Hospital Name:	Austin Health - Austin Hospital
Hospital Name:	Mercy Hospital for Women

Neighborhood Number: 5

Neighborhood Name:	Lynbrook - Lyndhurst
1st Most Common Venue:	Recreational Area
2nd Most Common Venue:	Australian Restaurant
3rd Most Common Venue:	Burger Joint
4th Most Common Venue:	Carpet Store
5th Most Common Venue:	Shopping Mall
Population Density:	2069.01
School Name:	St Margaret's School, Berwick
School Name:	Brentwood Park Primary School
Hospital Name:	Casey Hospital

Neighborhood Number: 6

Neighborhood Name:	Cheltenham - Highett (West)
1st Most Common Venue:	Gym
2nd Most Common Venue:	Recreational Area
3rd Most Common Venue:	Dance Studio
4th Most Common Venue:	Brewery
5th Most Common Venue:	Italian Restaurant
Population Density:	2056.09
School Name:	Sandringham East Primary School
School Name:	Sacred Heart School, Sandringham
School Name:	Sandringham Primary School
Hospital Name:	Sandringham Hospital

Neighborhood Number: 7

Neighborhood Name:	Hampton Park - Lynbrook
1st Most Common Venue:	Gym / Fitness Center
2nd Most Common Venue:	Bakery
3rd Most Common Venue:	Burger Joint
4th Most Common Venue:	Chinese Restaurant
5th Most Common Venue:	Sandwich Place
Population Density:	1913.46
School Name:	St Margaret's School, Berwick
School Name:	Brentwood Park Primary School
Hospital Name:	Casey Hospital

Neighborhood Number: 8

Neighborhood Name:	Croydon Hills - Warranwood
1st Most Common Venue:	Tennis Court
2nd Most Common Venue:	Photography Studio

3rd Most Common Venue: Recreational Area
4th Most Common Venue: Theater
5th Most Common Venue: Dumpling Restaurant
Population Density: 1894.75
School Name: Tintern Grammar, Ringwood East
School Name: Heathmont East Primary School
School Name: Great Ryrie Primary School
Hospital Name: Maroondah Hospital

Neighborhood Number: 9
Neighborhood Name: Montmorency - Briar Hill
1st Most Common Venue: Deli / Bodega
2nd Most Common Venue: Italian Restaurant
3rd Most Common Venue: Recreational Area
4th Most Common Venue: Restaurant
5th Most Common Venue: Bakery
Population Density: 1338.23
School Name: Banyule Primary School, Rosanna
School Name: Heidelberg Primary School
School Name: Viewbank Primary School
School Name: St Martin of Tours School, Rosanna
School Name: Rosanna Golf Links Primary School
Hospital Name: Austin Health - Austin Hospital
Hospital Name: Mercy Hospital for Women

Neighborhood Number: 10
Neighborhood Name: Bayswater North
1st Most Common Venue: Sandwich Place
2nd Most Common Venue: Sporting Goods Shop
3rd Most Common Venue: Arts & Crafts Store
4th Most Common Venue: Fast Food Restaurant
5th Most Common Venue: Recreational Area
Population Density: 1188.11
School Name: Tintern Grammar, Ringwood East
School Name: Heathmont East Primary School
School Name: Great Ryrie Primary School
Hospital Name: Maroondah Hospital

By plotting the resulting Data on the Map of Melbourne, I can get a clear idea of the location of each neighborhood and the Schools and Hospitals within. Figure below illustrates the findings.
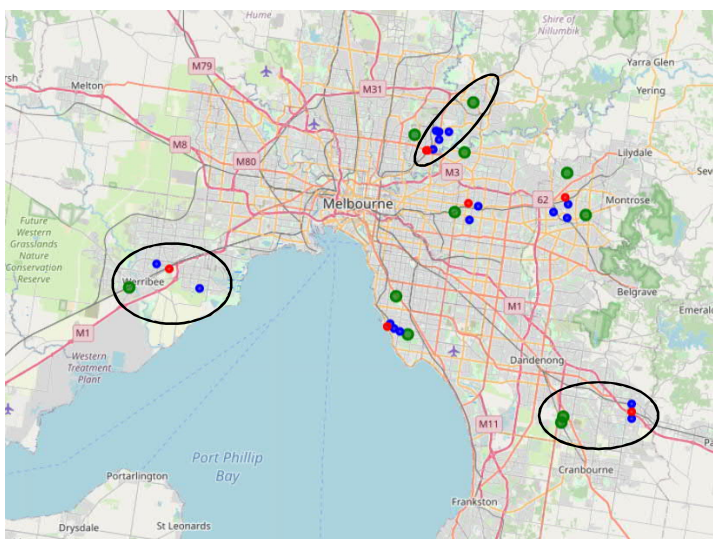
The green dots represent the neighborhoods, whereas the blue dots represent the schools and the red dots represent the hospitals.

Clearly from the plot some neighborhoods appear to be far away from the nearest schools and hospitals (circled in black). These Neighborhoods are numbers (2, 5, 7, and 9) This could be due to several factors, such as:

1. The Neighborhoods and Schools and Hospitals share the same postcode/SA3 are code.
2. The Neighborhood coordinates are towards the edge of the Neighborhood boundaries.

5. Conclusion

The results were satisfying the initial requirements of the project and therefore I consider the results to be helpful in answering the problem question "Where to live in Melbourne?".

From our Results, we conclude the following:

The top locations that fit our initial requirements are the following Neighborhoods:

- Surrey Hills (East)
- Brighton East
- Heidelberg - Rosanna
- Heidelberg West
- Cheltenham - Highett West
- Croydon Hills
- Bayswater North

The remaining 4 Neighborhoods also fit our preferences but are located far away from the nearest Hospital and Primary Schools, yet within the same Postcode or SA3Code.

Finally, it was very satisfying to be able to use analytical thinking and logical elimination, to downsize the search from over 747 localities within Melbourne to finally 11 Neighborhoods that fit our search criteria. This project can be applied to many other similar applications that require location Data such as:

1. Finding the best location to start a business.
2. Finding the best locations stay and visit during a planned vacation.
3. Finding the relation between the average footfall of a venue and the population density of the area.

And many more...

A word cloud showing the most common Venues masked on the Map of Melbourne