IBM Capstone Project

Investigation into the relationship between surrounding locale on property values of suburbs in Melbourne, Australia

By Cody

This report was created as part of the requirement for completing the Data Science Capstone project course provided by IBM through Coursera

Abstract - It is well known in property investment that location is one of the most important factors when purchasing property. This report will investigate at a high level the relationship between property value for houses and apartment units and it's the locales for selected suburbs in Melbourne, Australia. The datasets used are: house median values, unit median values, 2016 census dataset. The venues were obtained using Foursquare's Places API as per the requirement of this project. Agglomerative clustering was done to cluster the suburbs based on venues and property values. Results showed that there is a relationship between property value and venues such that clusters can be formed from these features.

Report Structure

The report constructed of 6 chapters in total. Beginning with topic introduction in chapter 1, background in chapter 2, project design in chapter 3 and results and discussion in chapter 4 and 5 with conclusion in 6. Diagrams are labeled as "Figures' while tables are listed as "tables". Figures and tables within the body of the report are number coded while figures and tables in appendix begins with an alphabet then numerical value corresponding to their respective sub appendix.

List of Figures and Tables

Tables

Appendix

Chapter 1

Introduction

As the property values continue to fall in Australia's capital cities in what some are calling Australia's property bubble bust, some Australians are seeing it as an opportunity to take advantage of the lower prices and purchase new property despite it being harder to get mortgages. Westpac bank recently reported a 11.8 % increase in respondents who think it's a good time to buy property, led by consumers in New South Wales and Victoria [1]. For those looking to buy property be it first time home owners, investors or business entrepreneurs; location is perhaps the biggest factor in determining purchase price, resale value or in determining if a business will do well. However, what makes a good location or a bad one is a subjective topic that depends on each individual wants and needs.

This study is aimed towards informing home buyers and property investors of select suburbs in the state of Victoria. Buyers will be able to use this information to aid them in decision making on purchasing properties in suburbs that are within their budget and explore popular venue categories in those areas while business owners are able to better decide which areas best suit their business model.

1.1 Objective

The objective of this paper is to investigate the relationship between venues in the form of categories and property values of houses and apartment units in Melbourne, Victoria.

1.2 Study Limitations

This study is limited by the data available as it relies on publicly available datasets. These datasets may not be up to date or may lack specific details. The dataset containing prices of houses and units do not include details such as number of rooms, size, age, building for apartment units or type of houses. Furthermore, a full-scale study of all suburbs in the state of Victoria is not possible due to time constraints and hardware limitations. Therefore, this study will analyse selected suburbs in and around Melbourne city and should be taken as a general overview rather than a definitive indicator. For this project, the terms: housing refers to houses and units refers to apartment units. While terms such as suburbs, neighbourhoods and regions are used interchangeably.

Chapter 2: Background

2.1 The Australian Property Market

The Australian property market has enjoyed national growth of 41.8% in the last 10 years up to January of 2018. This growth has been spearheaded by the country's capital cities that is Sydney and Melbourne. In a recent report by CoreLogic, Sydney has seen a rise of 79.3% in dwelling values over the last 10 years with Melbourne following close behind at 72.4%. The two cities combined with regional Victoria at 42.7% are the only locations to have growth that is over the national value [2]. Together, Sydney and Melbourne comprise of around 55% of the Australian housing market. This is due to the cities attracting many foreign investors as well as experiencing strong population growth with Melbourne alone receiving 35% of all overseas migrants [3].

However, the Australian housing market have recently been experiencing the largest decline since the global financial crisis of 2008, falling -4.8 % nationally throughout 2018. Dwelling values in Sydney fell by -8.9% by the end of 2018 after a peak high in August 2017. Meanwhile, Melbourne fell -7.0% after peaking 3 months later on November of 2017 [2]. But unlike 2008, this decline was not caused by economic shocks but by tighter credit conditions from regulators as a preventive measure to prevent a property market crash in Sydney and Melbourne, if the cities continue to grow at an unstainable rate as they were. The tighter conditions resulted in many not being able to borrow as much as they could a year ago [3].

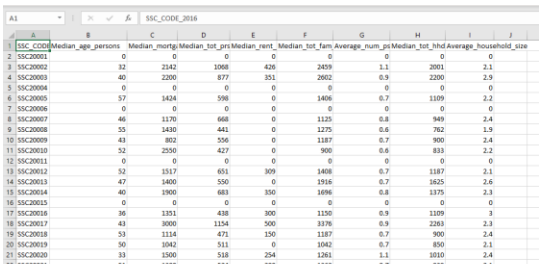
Current numbers for buyers put first time home buyers at 17.4% of all owner-occupied housing finance commitments rising from 12.9% in 2015. Meanwhile, investors makes up the lion share of the housing demand, in 2015, investors reached a historic high, making up 55% of all new housing finance commitments. Due to tightening regulations that number has fallen in 2018 but still holds a large chunk at 42.8% of all mortgage demand [4]. Despite the tighter regulations many Australians see it as an opportunity to take advantage of the falling prices to purchase property. However, even though prices are dropping property prices in Australia's major cities remains high; which means the average household in Sydney would need to dedicate an average of 185.1% of annual household income to raise the 20% deposit required to buy a home while the figure in Melbourne is slightly lower at 159.7% [4]. Making purchasing property a considerably big investment for those living in these capital cities.

Chapter 3: Project Design

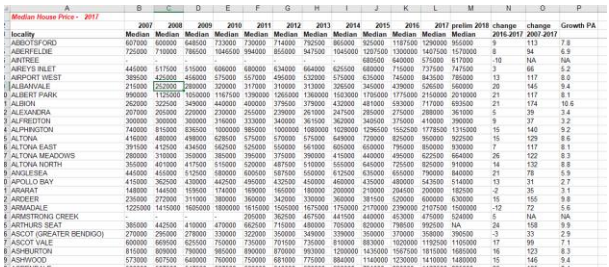
3.1 Data Gathering

The first dataset used for this analysis was obtained from two sources, the first is the suburb profiles collected from the 2016 census of population and housing performed by the Australian Bureau of Statistics (ABS). The dataset is publicly available for download on the ABS website. This is a large dataset containing many features ranging from labour force status, family composition, occupation, qualifications, etc for every suburb in the state of Victoria. For the purpose of this analysis only the dataset label '2016Census_G02_VIC_SSC' and '2016Census_G41_VIC_SSC' will be used. The first one contains the median values for specific features for every suburb including age, mortgage monthly repay, personal income, household income, family weekly income, average household size and number of persons per bedroom. However, emphasis will be placed on weekly household income and age. It should be noted that the data on incomes are from those who are 15 years or older and the survey form list income in various ranges rather than requesting a specific amount. The second dataset contains the total population for each suburb. An example of the data set is shown in figure 1 below (left). Population only counts those who are 1 year or older. It is important to note that because the last census was performed in 2016 and the next one is only scheduled for 2021, the 2016 version will be used for the analysis despite being outdated as this data is the most comprehensive that is publicly available.

The second and third datasets that is used in this analysis are the ones that contain the median house and unit values for every suburb in Victoria. This data set comes from the Victorian State Government of Australia and is available on their website. The 2016 median house and unit prices will be used for consistency with the rest of the data. An example is shown in figure 1 (right).



SSC_CODE	Median_age_persons	Median_mortg	Median_tot_prs	Median_rent_	Median_tot_fam	Average_num_ps	Median_tot_hhd	Average_household_size
SSC20001	0	0	0	0	0	0	0	0
SSC20002	32	2142	1068	426	2459	1.1	2001	2.1
SSC20003	40	2200	877	351	2602	0.9	2200	2.9
SSC20004	0	0	0	0	0	0	0	0
SSC20005	57	1424	598	0	1406	0.7	1109	2.2
SSC20006	0	0	0	0	0	0	0	0
SSC20007	46	1170	668	0	1125	0.8	949	2.4
SSC20008	55	1430	441	0	1275	0.6	762	1.9
SSC20009	43	802	556	0	1187	0.7	900	2.4
SSC20010	52	2550	427	0	900	0.6	833	2.2
SSC20011	0	0	0	0	0	0	0	0
SSC20012	52	1517	651	309	1408	0.7	1187	2.1
SSC20013	47	1400	550	0	1916	0.7	1625	2.6
SSC20014	40	1900	683	350	1696	0.8	1375	2.3
SSC20015	0	0	0	0	0	0	0	0
SSC20016	36	1351	438	300	1150	0.9	1109	3
SSC20017	43	3000	1154	500	3376	0.9	2263	2.3
SSC20018	53	1114	471	150	1187	0.7	900	2.4
SSC20019	50	1042	511	0	1042	0.7	850	2.1
SSC20020	33	1500	518	254	1261	1.1	1010	2.4
SSC20021	51	1300	524	200	1362	0.7	939	2.1
SSC20022	35	1560	680	300	1809	0.8	1527	2.8
SSC20023	60	0	420	0	1074	0.8	474	1.7
SSC20024	45	1343	469	0	1125	0.9	1042	2.5
SSC20025	54	1200	459	0	1125	0.7	1062	2.2
SSC20026	45	1850	676	270	1462	0.8	1341	2.7
SSC20027	39	1470	671	240	1732	0.8	1478	2.7
SSC20028	47	1305	693	0	1416	0.7	1221	2.2
SSC20029	45	1200	481	240	1292	0.7	1199	2.3
SSC20030	44	1365	427	205	900	0.8	814	2.8
SSC20031	30	0	383	0	900	0.9	900	3.1
SSC20032	47	1300	687	0	1541	0.7	1374	1.9
SSC20033	38	2037	915	370	2802	0.9	2123	2.5
SSC20034	42	2000	732	331	1917	0.8	1400	2.3
SSC20035	38	1517	592	310	1586	0.8	1370	2.6
SSC20036	38	2000	503	340	1478	0.9	1240	2.6
SSC20037	49	800	615	138	1708	0.6	1145	2.4



locality	2007 Median	2008 Median	2009 Median	2010 Median	2011 Median	2012 Median	2013 Median	2014 Median	2015 Median	2016 Median	2017 Median	prelim 2018 Median	change 2016-2017	change 2007-2017	Growth PA
ABBOTSFORD	607000	600000	648500	733000	730000	714000	792500	865000	925000	1187500	1290000	950000	9	113	7.8
ABERFELDIE	725000	710000	786000	1045000	994000	855000	947500	1045000	1207500	1300000	1407500	1570000	8	94	6.9
AINTREE										680500	640000	575000	-10	NA	NA
AIREYS INLET	445000	517500	515000	606000	680000	634000	664000	625500	680000	715000	737500	747500	3	68	5.2
AIRPORT WEST	389500	425000	416000	575000	557000	496000	532000	575000	625000	745000	843500	785000	13	117	8.0
ALBANVALE	215000	252000	288000	320000	317000	310000	313000	326500	345000	439000	526500	560000	20	145	9.4
ALBERT PARK	990000	1125000	1050000	1167500	1390000	1265000	1360000	1563000	1705000	1775000	2150000	2010000	21	117	8.1
ALBION	262000	322500	349000	440000	400000	379500	379000	432000	481000	593000	717000	693500	21	174	10.6
ALEXANDRA	207000	205000	220000	230000	255000	239000	261000	247500	285000	275000	288000	361000	5	39	3.4
ALFREDTON	300000	300000	300000	316000	333000	340000	361500	362000	340500	375000	410000	390000	9	37	3.2
ALPHINGTON	740000	815000	836500	1000000	985000	1000000	1080000	1028000	1296500	1552500	1778500	1315000	15	140	9.2
ALTONA	416000	480000	498000	628500	575000	570000	575000	649000	720000	825000	950000	922500	15	129	8.6
ALTONA EAST	391500	412500	434500	562500	525000	550000	561000	605000	660000	796000	850000	930000	7	117	8.1
ALTONA MEADOWS	280000	310000	350000	385000	395000	375000	390000	415000	440000	495000	622500	664000	26	122	8.3
ALTONA NORTH	350000	401000	417500	515000	520000	487500	510000	555000	645000	725000	825000	910000	14	132	8.8
ANGLESEA	445000	465000	512500	580000	600500	587500	560000	612500	635000	655000	790000	840000	21	78	5.9
APOLLO BAY	415000	362500	430000	442500	495000	432500	450000	460000	435000	480000	543500	514000	13	31	2.7
ARARAT	148000	144500	153900	174000	169000	165000	180000	200000	210000	204500	200000	182500	-2	35	3.1
ARDEER	236000	272000	311000	380000	360000	342000	330000	360000	381500	520000	600000	630000	15	155	9.8
ARMADALE	1225000	1415000	1685000	1800000	1615000	1505000	1575000	1750000	2170000	2390000	2107500	1500000	-12	72	5.6
ARMSTRONG CREEK	-	-	-	-	295000	362500	467500	441500	440000	453000	475000	524000	5	NA	NA
ARTHURS SEAT	385000	442500	410000	470000	662500	715000	480000	705000	820000	798500	992500	NA	24	158	9.9
ASCOT (GREATER BENDIGO)	270000	295900	278000	330000	322000	350000	349000	379900	350000	370000	358000	396500	-3	33	2.9
ASCOT VALE	600000	669500	625500	750000	735000	701500	736000	810000	883000	1020000	1192500	1105000	17	99	7.1
ASHBURTON	815000	909000	790000	985000	890000	870000	993000	1200000	1436000	1567500	1815000	1685000	16	123	8.3
ASHWOOD	573000	607500	640000	760000	750000	681000	775000	884000	1140000	1230000	1410000	1480000	15	146	9.4
ASPENDALE	500000	507500	547500	637500	630000	610000	620000	690000	791000	890000	1122500	885000	26	125	8.4
ASPENDALE GARDENS	435000	475000	498500	581000	550000	552000	586500	632000	715000	775000	877500	927500	13	102	7.3
ATTWOOD	365000	440000	485000	576000	592000	500000	540000	515000	629000	660000	740000	795000	12	103	7.3
AVENEL	206000	172500	212500	268500	221500	230500	231000	302900	245000	247000	340000	320000	38	65	5.1
AVOCA	134000	167000	133000	144000	151000	181500	182500	174500	182500	164000	212500	190000	30	59	4.7
AVONDALE HEIGHTS	367500	440000	460000	576000	538000	536000	532500	560000	650000	753000	850000	755000	13	131	8.7
AVONSLEIGH	255000	318000	309000	321500	385500	345000	420000	355000	397500	581000	562000	500000	12	120	8.2
BACCHUS MARSH	248000	255000	279000	305000	330000	325000	335000	348000	355000	376000	415000	435000	11	68	5.3
BADGER CREEK	234500	234500	266500	295000	295000	295000	315000	325000	345000	379000	465000	482500	23	98	7.1
BAIRNSDALE	208000	210000	219000	230000	235000	231000	235000	244000	247500	255000	270000	295000	6	30	2.6
BALACLAVA	705000	782500	750000	916000	793500	717500	866500	942000	1062000	1242500	1390000	1410000	12	97	7.0
BALCOMBE	476500	519000	539000	655000	716500	686000	712500	737000	871000	950000	1100000	1108000	16	131	8.7
BALLAN	203500	220000	250000	260000	310000	305000	352000	330000	362000	363500	414000	439000	14	103	7.4
BALLARAT CENTRAL	235000	240000	251000	280000	320000	313500	322000	327500	351500	385000	430000	465000	12	83	6.2
BALLARAT EAST	190000	195000	205000	225000	240000	255000	265000	265000	275000	295000	309500	335000	5	63	5.0
BALLARAT NORTH	222000	215000	230000	272000	275000	280000	296000	312000	315000	340000	346500	350000	2	56	4.6
BALLARAT SOUTH	332500	385500	388500	480000	472500	501500	525000	544000	595000	670000	815000	811000	22	145	9.4
BALNARRING	636000	709900	710000	787500	630000	660000	750000	700000	887500	994000	896000	1970000	-2	40	3.4
BALNARRING BEACH	636000	709900	710000	787500	630000	660000	750000	700000	887500	994000	896000	1970000	-2	40	3.4
BALWYN	1101000	1170000	1279500	1450000	1363000	1295000	1530000	1780000	2075000	2170000	2360000	2255000	9	114	7.9
BALWYN NORTH	896000	900000	981000	1141000	1100000	997500	1240000	1452000	1850000	1799500	1950000	1960000	8	118	8.1
BANNOCKBURN	276500	287000	305000	342500	378000	351500	370000	405000	389900	382000	430000	477500	13	55	4.5
BARAMBOGIE	292000	327000	313000	363000	349000	340000	344000	330000	333000	367500	360000	383000	-2	30	1.8

Figure 1: Preview of dataset from the 2016 census of population and housing (left) and median housing price (right).

3.2 Methodology

3.2.1 Data Selection and Feature Extraction

For this study the methodology involves sourcing the relevant data, cleaning, exploration and analysis. The selected region is Melbourne located in the state of Victoria, Australia. Melbourne is second largest and populous city in the country behind Sydney. Owing to the large number of suburbs in the Victoria, only the ones located in Melbourne city and its surroundings will be used for this analysis. The suburbs in Victoria are grouped into regions that are under the authority of various local city councils as shown in the figure 2 below.

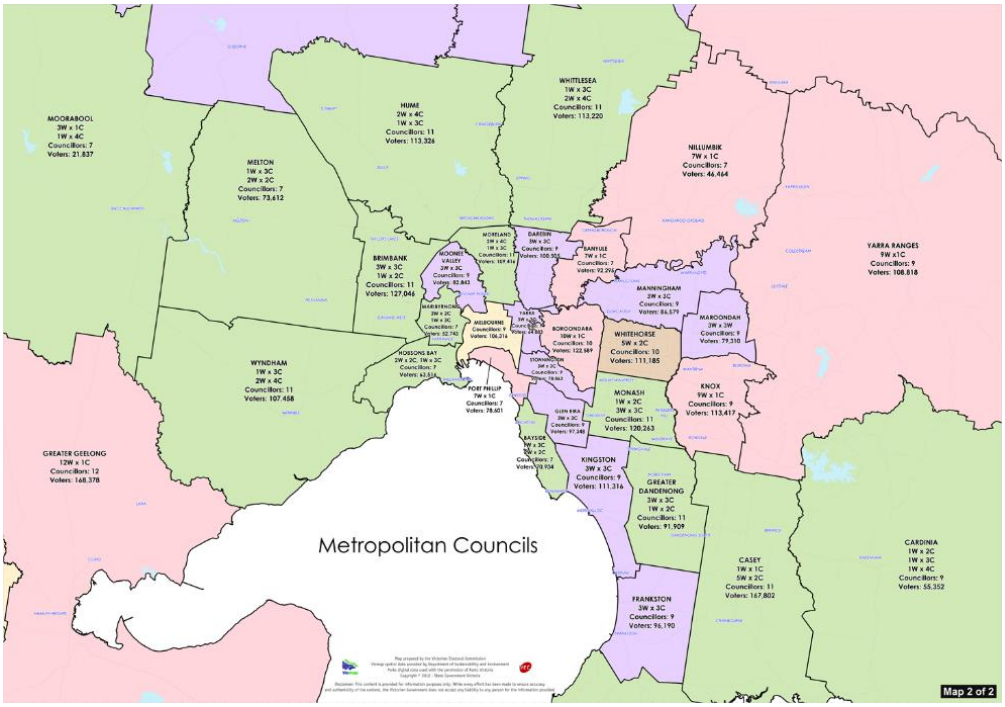


Figure 2: Map showing the city councils in Victoria [5].

The city councils chosen for this study are:

- City of Melbourne
- City of Port Phillip
- City of Yarra
- City of Boroondara
- City of Stonnington
- City of Monash
- City of Whitehorse
- City of Moonee Valley
- City of Maribyrnong

- City of Hobsons Bay
- City of Brimbank
- City of Greater Dandenong
- City of Darebin
- City of Moreland

These councils are specifically chosen as they are in and around the city of Melbourne. Care is taken as to not bloat the analysis area by taking large city councils such as the City of Melton, Hume, Wyndham, etc whilst covering a good portion of the Melbourne region.

The first thing to do in processing the data is to import all the relevant datasets into a data frame. All the datasets need to be merge into a single data frame using the suburbs as a common feature. A text file containing all the suburbs for each city council was used as a starting point to extract data for the desired suburbs from the other datasets. Unfortunately, this file contained some additional location information for some of the suburbs which needed to be removed so data cleaning was required. After cleaning, the file was ready to be merged with the dataset for median housing and unit prices since the suburbs are readily available without further cleaning required. All other columns were dropped except for the median 2016 values. The next datasets to be merge was the ones from the 2016 census; However, unlike the housing and unit datasets they did not directly list the suburbs in names but instead used SSC codes that was assigned by the ABS. To transform this code into plain text, a separate file containing the list of SSC codes and their respective suburb names was obtained. After transforming these codes into names, the file was finally merged with the other datasets to obtain the complete dataset containing all the data of interest as shown in figure 3 below. Separate datasets where created for houses and apartment units since some suburbs have apartment units but not houses such as the Melbourne CBD area. However, another problem arose, some suburbs were duplicated because they are shared between two or more city councils. So, the duplicated data were combined into a single entry while maintaining their respective city councils. An example would be 'BROOKLYN' suburb as seen in figure 3.

	Zipcode	Suburb	City council	SSC code	State	House price $ (2016)	Median age	Median mortgage repay monthly	Median personal inc weekly	Median rent weekly	Median tot family inc weekly	Average person per bedroom	Median houshold inc weekly	Average household size	Total population
0	3002	EAST MELBOURNE	City of Melbourne	20824	Victoria	3655000.0	38.0	2192.0	1341.0	451.0	3120.0	1.0	2285.0	1.9	4899.0
1	3003	WEST MELBOURNE	City of Melbourne	22743	Victoria	1181000.0	30.0	2006.0	852.0	450.0	2246.0	1.1	1766.0	2.2	5477.0
2	3011	FOOTSCRAY	City of Maribyrnong	20929	Victoria	775000.0	32.0	1842.0	623.0	310.0	1660.0	1.1	1314.0	2.2	16131.0
3	3011	SEDDON	City of Maribyrnong	22245	Victoria	934500.0	35.0	2167.0	936.0	380.0	2446.0	1.0	2006.0	2.4	5025.0
4	3012	BROOKLYN	City of Hobsons Bay & City of Brimbank	20350	Victoria	700500.0	33.0	1842.0	730.0	340.0	1684.0	0.9	1460.0	2.3	1818.0

Figure 3: The first 5 entries of data frame for housing.

The Foursquare places API requires the latitude and longitude as input. Unfortunately, a dataset containing this information for suburbs could not be found. So, a function was created to make use of Nominatim by OpenStreetMaps to get the latitude and longitude for each suburb. The returned coordinates were added to the housing and unit data frames respectively, ready to be applied to the Foursquare API. For data exploration purposes, a separate data frame was created containing all suburbs regardless of whether they had both houses and apartment units present. In total, three data frames were created: df_housing, df_unit, all_suburbs.

3.2.2 Data Exploration

The first step in data exploration is to visualise all suburbs on a map. This can be done using the folium package available for Python 3. Figure 4 illustrates the results of this. The blue circle markers represent the suburbs for housing while the green is for units. To prevent them completely overlapping the green markers was slightly off centred.

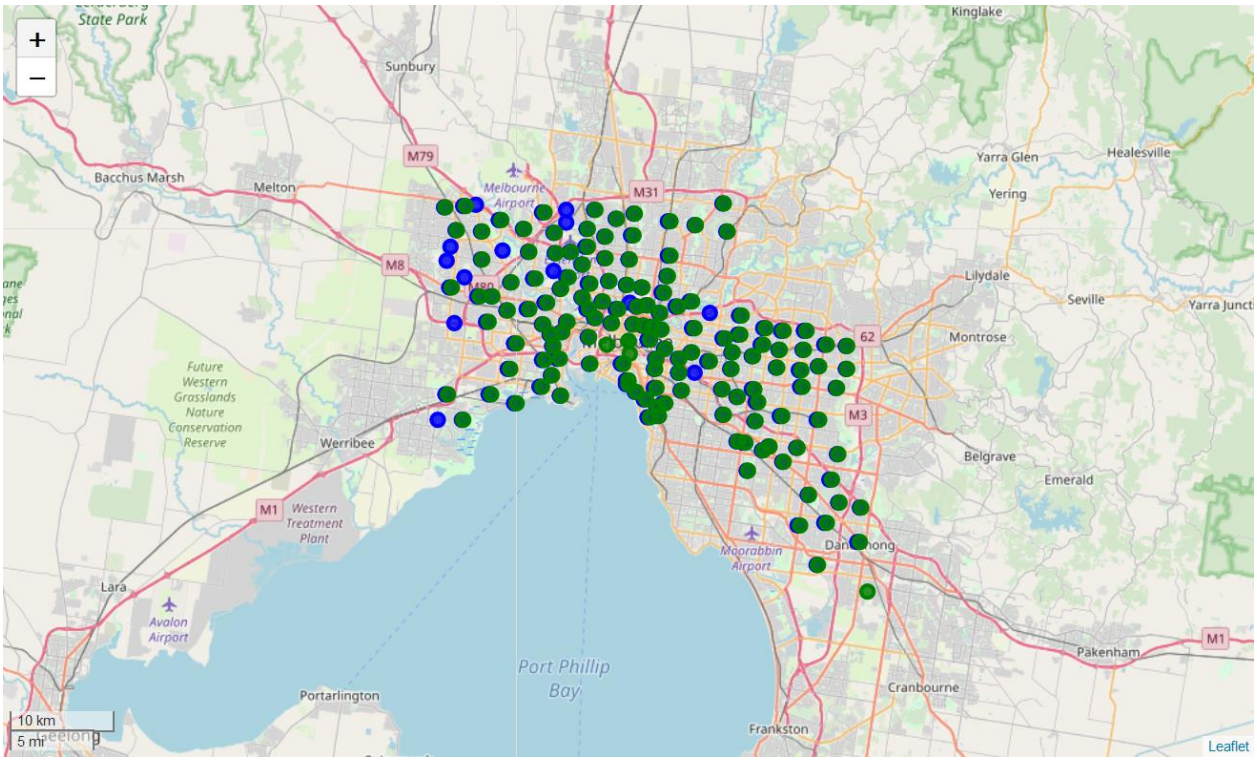


Figure 4: All suburbs plotted using folium.

Folium's choropleth maps can be used to visualise specific features by overlaying it on a map using a json file containing the boundaries for each suburb. This json file can be found on the Victorian government page. The downloaded json file need to be edited to only include the suburbs of interest and saved as a new json file. The choropleth plots were then created using this new boundary file.

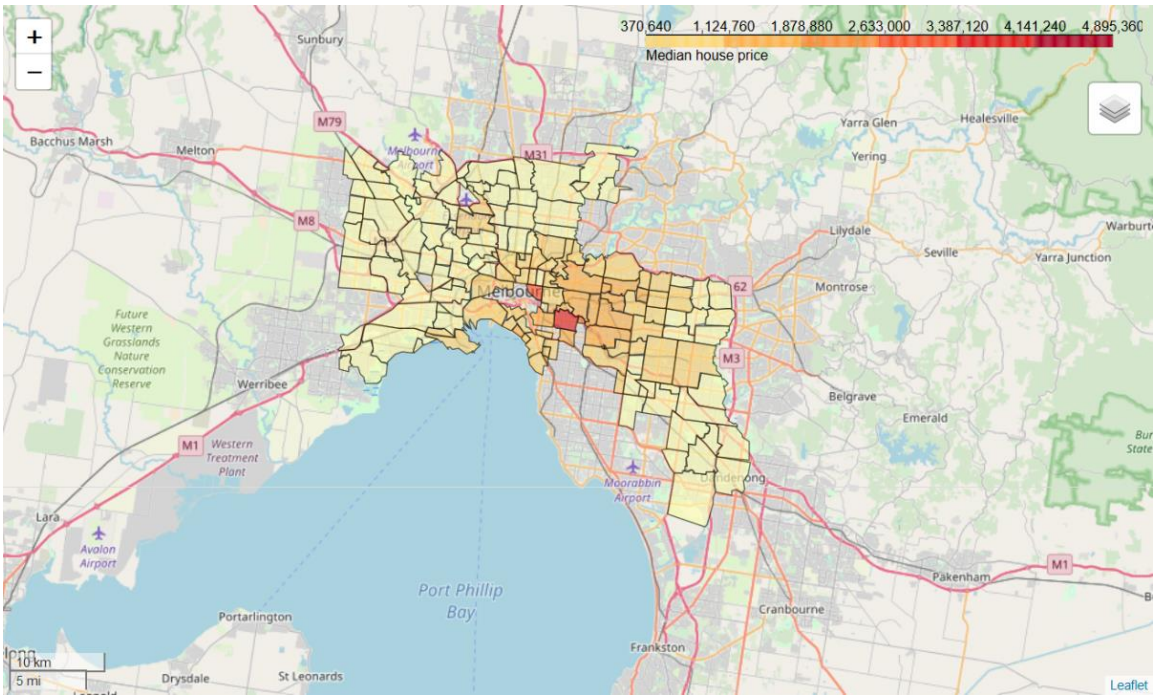


Figure 5: Median housing values per suburb.

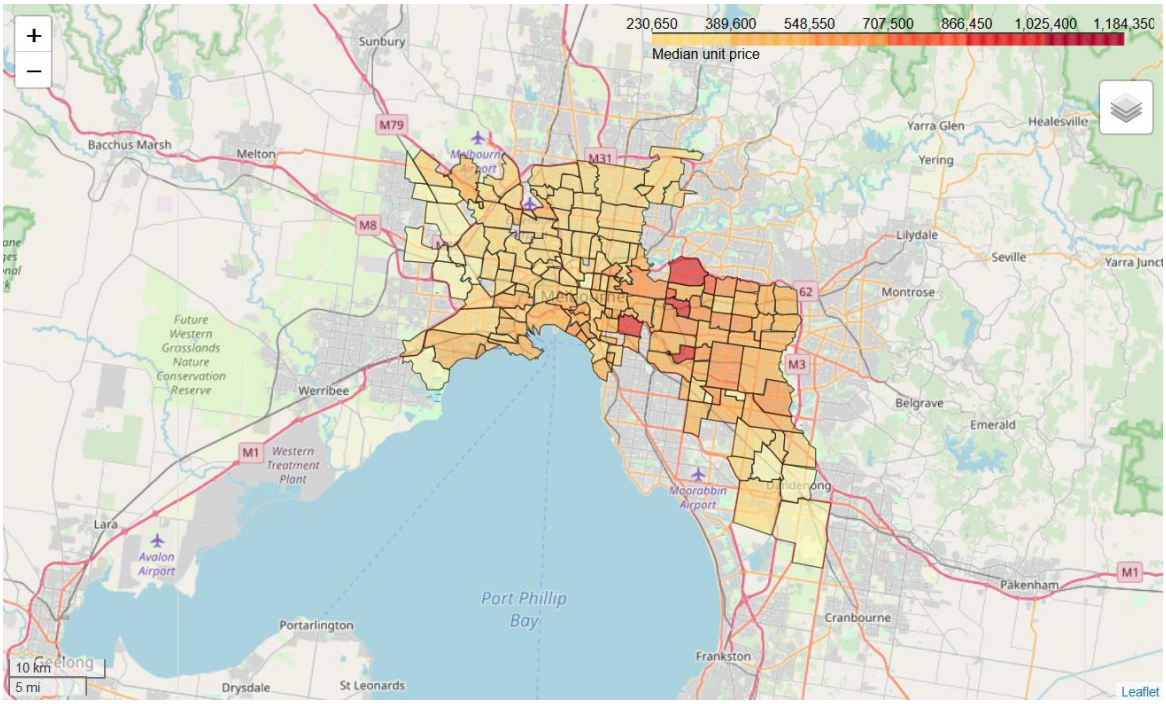


Figure 6: Median unit values per suburb.

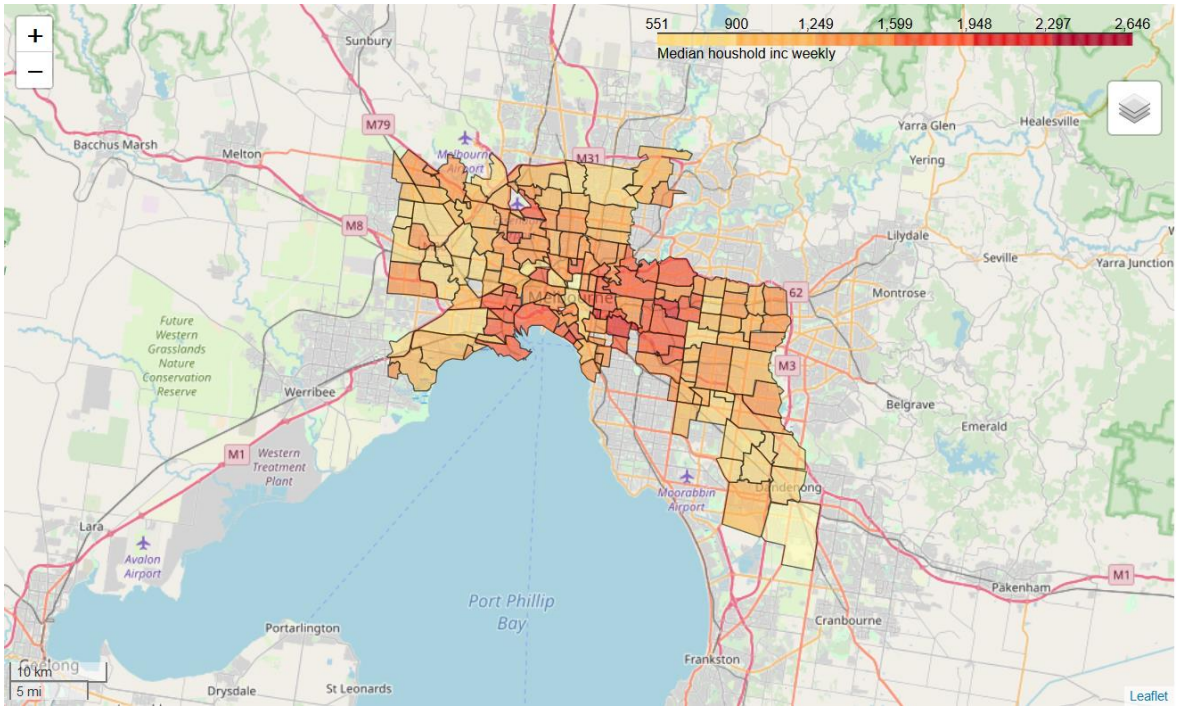


Figure 7: Median household values per suburb.

The choropleth maps for population can be viewed in appendix A. To better understand each suburb, tables and graphs can be plotted. Figures 8 and 9 shows the plots for median housing value and unit value while income and age can be found in appendix A. Figure 10 shows the density plot of the values.

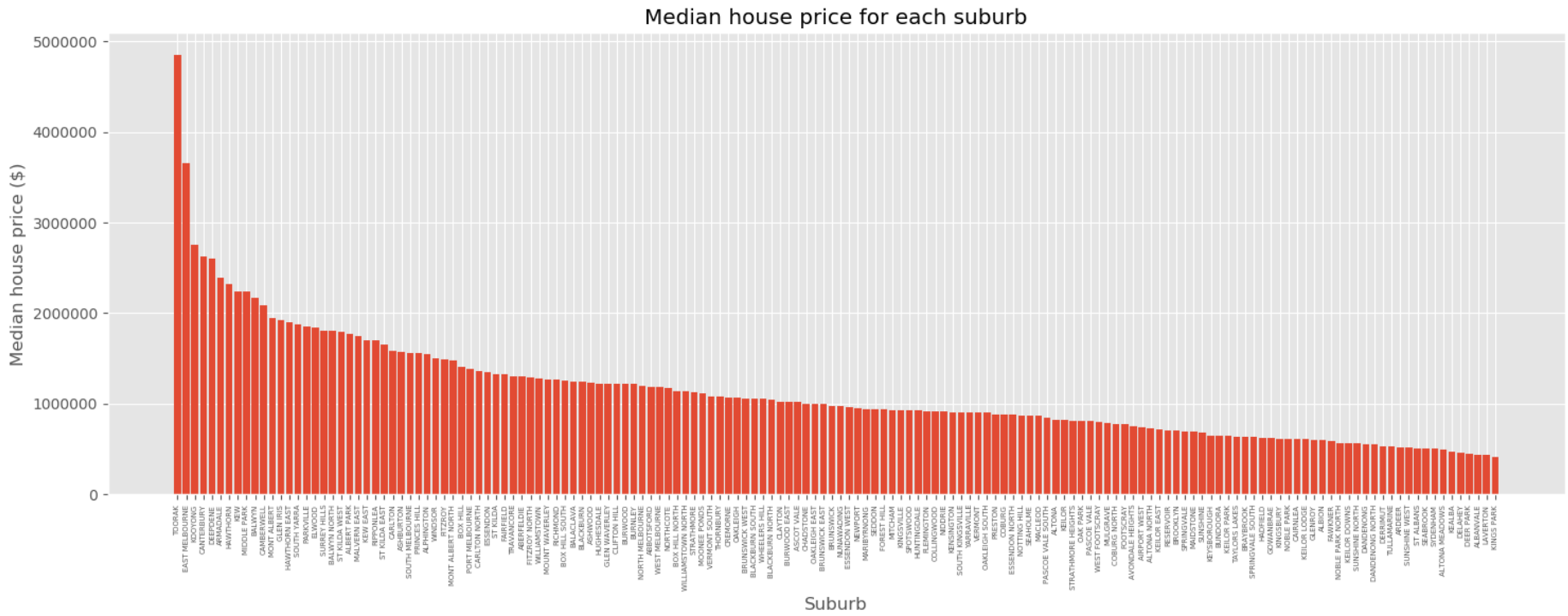


Figure 8: Median housing value per suburb.

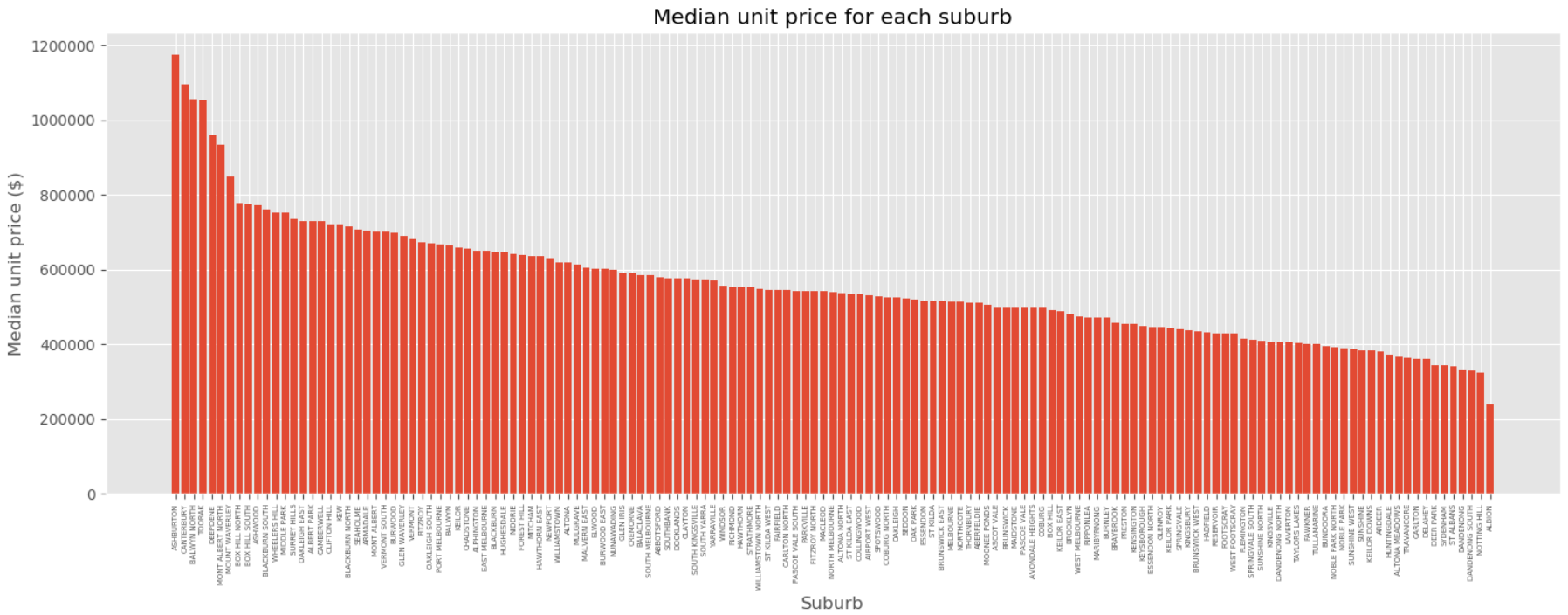


Figure 9: Median unit value per suburb.

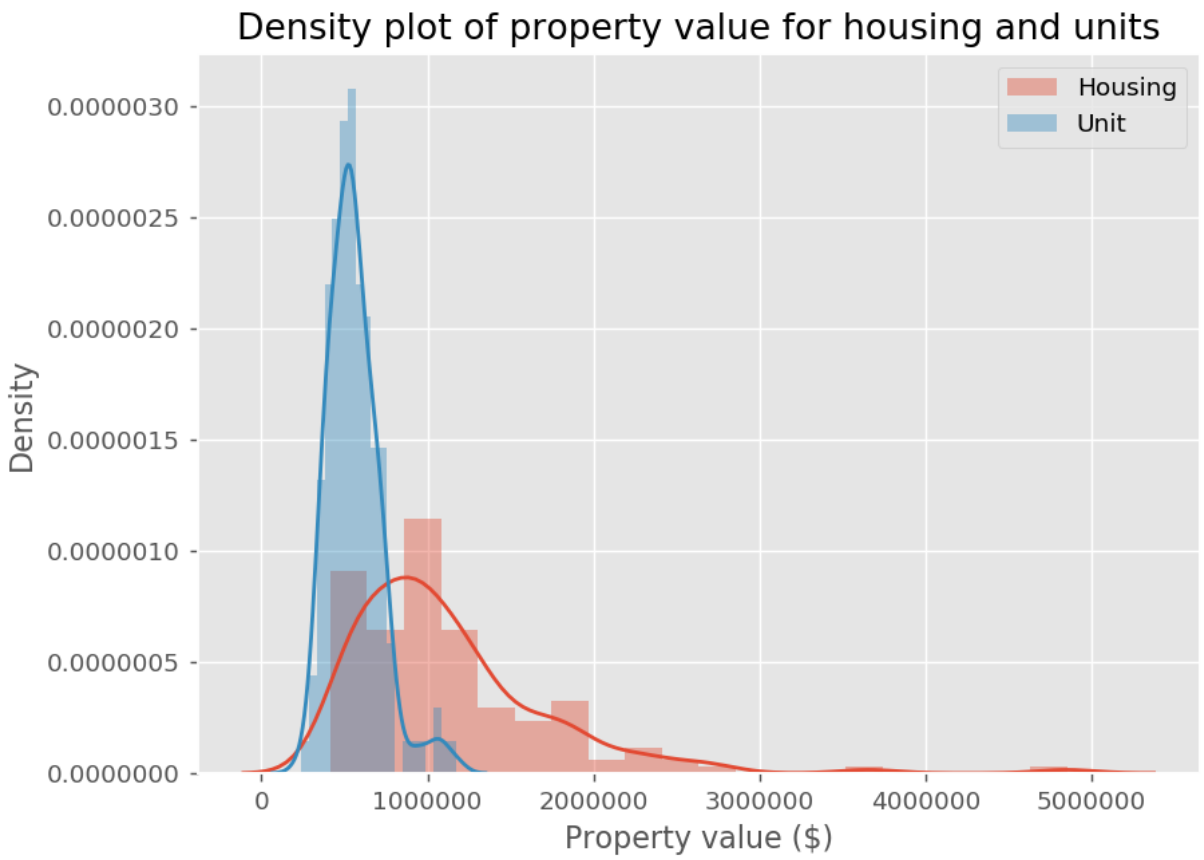


Figure 10: Density of property values.

The visualisations above reveal some important insight to the data. Firstly, Melbourne's most wealthy neighbourhoods in the form of house and unit prices and weekly household income are Canterbury, Kooyong, Toorak, Deepdene, Hawtorn, Middle Park and East Melbourne. With Toorak having the highest house price of all suburbs by a large margin. Meanwhile, the suburb of Deepdene is valuable for both apartment units and housing but interestingly not as high up in median income as the other wealthy suburbs but is tied for the highest median age of all suburbs. A possible reason could be that this suburb might have a larger population of retirees or people who bought these properties back when they were cheaper. On the other hand, there are suburbs like Cremorne with high household and personal income but moderate in house and unit price.

The choropleth maps show that most valuable areas for housing are mostly located in the eastern suburbs of Melbourne while the western suburbs are less valuable excluding areas in the inner Melbourne regions. On the other hand, apartment units' prices are more evenly distributed, but the most expensive areas are still mostly in the eastern parts. Furthermore, the highest value places are all located in the inner eastern suburbs closer to the city and the median age of these areas are also one of the highest which is sensible as older people generally have more money and so are more likely to own valuable property.

Looking at some of the least valuable and lowest income areas, it appears they are mostly located on the outskirts of Melbourne on the far west and south east. Analysing Melbourne city itself, we can note that younger people prefer to live in the city. The median age of Melbourne city at just 27 is among the lowest and the same goes for income. It would seem young professionals and students are the most popular demographic here. An interesting place to note are the suburbs of Carlton and Clayton both of which have very low weekly household incomes (500-1000) and ages (24-26). This is because these suburbs are located adjacent to Melbourne University and Monash University respectively. Thus, it is mostly populated by university students which accounts for these observations.

Finally, the density plot shows that the distribution for house value is skewed to the right due to the high value neighbourhoods such as Toorak. Most houses in Melbourne cost somewhere around 900,000 to 1,100,000. Meanwhile, the distribution for apartment units is more normally distributed with an increase in value to the right tail. This is due to the suburbs 'Ashburton', 'Canterbury', 'Baldwyn North' and 'Toorak' which all have values above $1,000,000 which is in the ball park of the cost of most houses in Melbourne. The distribution plot shows that apartment units cost approximately 500,000 to 600,000.

Another important statistic to look at is the correlation between different features. This can be done using the Pandas correlation matrix method. Figure 11 and 12 shows the correlation of features to housing value and unit value.

```
House price $ (2016)            1.000000
Median tot family inc weekly    0.747847
Median mortgage repay monthly   0.719419
Median personal inc weekly      0.621186
Median houshold inc weekly      0.612368
Median rent weekly              0.594838
Median age                      0.228412
Average household size         -0.408608
Name: House price $ (2016), dtype: float64
```

Figure 11: Correlation matrix for house price.

```
Unit price $ (2016)             1.000000
Median mortgage repay monthly   0.671382
Median houshold inc weekly      0.602231
Median tot family inc weekly    0.599261
Median rent weekly              0.587156
Median age                      0.461252
Median personal inc weekly      0.377649
Average household size          0.064859
Name: Unit price $ (2016), dtype: float64
```

Figure 12: Correlation matrix for unit price.

The correlation ranges from -1 to 1 with 1 being perfectly positively correlated and -1 being perfectly negatively correlated. From the figures, it seems mortgage repay, and income are quite strongly correlated with property values especially for housing prices while average household is negative, meaning more expensive homes have smaller households. For units, the average household size does not seem to be linearly correlated to the value as it is very close to 0. However, for the correlation matrix it is important to note that it assumes a linear relationship and cannot

pick up on non-linear ones. Also, the common phrase “correlation does not imply causation” should be observed, this means that there may be other factors involved that would explain the observation which the case is usually.

3.2.3 Data Exploration with Foursquare API

For this course, IBM requires the use of Foursquare’s Places API as part of the analysis. As such the API’s ‘explore’ attribute is used to obtain the venues for each suburb. The data version request is set to 1st December 2016 to comply with the date of rest of the data. To ensure that enough venues are obtained the limit is set to 200 so up to 200 venues can be retrieved. The radius is automatically selected by the API based on the density of venues in area. The output is reconfigured as a data frame containing the venue’s name, suburb, category, latitude, longitude and address. To prevent having to retrieve the information every time the notebook is run, the file is saved as a excel file. The total number of instances in the output data frame is 14,961 venues. A slice of the data frame can be seen in the appendix.

By counting the number of each category, it can be seen that cafes are by far the most common venue. This is because Melbourne is well known for its ‘café culture’ and is regarded to have one of the best coffee’s in the world. To further explore the data, the venue categories are counted and ranked in descending order for the top 10 most valuable suburbs and top 10 least valuable. The observation of the results are as follows:

- Those in the top 10 most valuable posses a wide selection of restaurants of mainly Asian and Italian. While places such as bars and pubs are also common sight including verities such as breweries, wine bars and beer gardens. Business and services such as parks, recreational and sporting facilities, stores and shops, movie theatres, etc are also considerably prominent in these suburbs.
- In contrast, those in the top 10 least valuable neighbourhoods have a sizeable amount of fast food and Portuguese restaurant, in fact fast food appears to be the most popular. These suburbs generally do not have as many venues or facilities compared to their wealthier counterparts.

The top 10 most common venue for every suburb is extracted from the data frame and saved as a new data frame variable: df_top_venues (Figure 13).

Out[100]:

	Suburb	Most common ranking: 1	Most common ranking: 2	Most common ranking: 3	Most common ranking: 4	Most common ranking: 5	Most common ranking: 6	Most common ranking: 7	Most common ranking: 8	Most common ranking: 9	Most common ranking: 10
0	ABBOTSFORD	Café	Vietnamese Restaurant	Coffee Shop	Pub	Bakery	Pizza Place	Bar	Thai Restaurant	Vegetarian / Vegan Restaurant	Breakfast Spot
1	ABERFELDIE	Café	Electronics Store	Gym	Dessert Shop	Pub	Coffee Shop	Bakery	Japanese Restaurant	Thai Restaurant	Pizza Place
2	AIRPORT WEST	Café	Supermarket	Grocery Store	Fast Food Restaurant	Sandwich Place	Shopping Mall	Convenience Store	Portuguese Restaurant	Electronics Store	Thai Restaurant
3	ALBANVALE	Fast Food Restaurant	Café	Portuguese Restaurant	Grocery Store	Vietnamese Restaurant	Electronics Store	Convenience Store	Shopping Mall	Pizza Place	Burger Joint
4	ALBERT PARK	Café	Coffee Shop	Beach	Breakfast Spot	Bar	Pub	Burger Joint	Mexican Restaurant	Japanese Restaurant	Seafood Restaurant

Figure 13: First 5 entries for the top 10 venues for all suburbs data frame.

3.2.4 Unsupervised Leaning: Clustering

To explore the relationship between the venue categories and the housing and apartment unit worth, unsupervised learning is used as the data is unlabelled. The data will need to be reconfigured since machine learning algorithms expect numerical values. The venue categories were converted into dummy variables and then grouped to obtain the frequency ratio of occurrence for each category. This ratio ranges from 0 to 1. The property values are then added to the new date set but not before normalisation with the MinMaxScaler from Scikit-learn. This operation is done for the housing and unit data set resulting in two new data frames consisting of 154 and 146 instances with both having 306 features each.

However, the high dimensional data might affect the effectiveness of distance-based clustering algorithms, to counter this principle component analysis (PCA) is used to decrease the number of features before instantiating clustering by breaking it down into components. Decreasing the features down to 50 components still retains 98 % of the variance of the original data. t-distributed Stochastic Neighbour Embedding (TSNE) can be used to decrease the features down further to 2 for visualisation purposes. TSNE is a better method for this compared to PCA which if used to reduce the features down to 2 will only retain 52% of the original variance which is very poor. Since TSNE is density based it does not do well in preserving distances, it is not a good technique as an input for clustering, only for visualisation. Scikit-learn documentation recommends using PCA to reduce the features down to at most 50 before using TSNE as it is computationally expensive reduction technique. The figure 14 shows the reduced data for housing cluster and unit cluster data after TSNE.

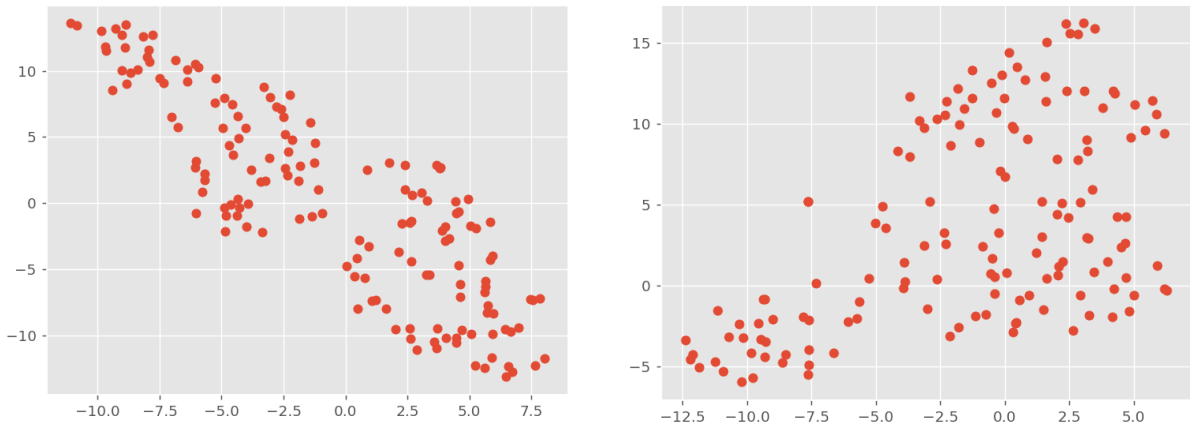


Figure 14: Housing data distribution (left) and Unit data distribution (right) after PCA + TSNE.

Three popular clustering methods were applied to the datasets, these are: K-means, Agglomerative clustering and Gaussian mixture clustering.

3.2.4.1 K-means

K-means clustering is the simplest method for clustering. K-means is an iterative process of assigning data points to cluster centres, recalculating the distances between the data points and the centres and reposition the centres until no change occurs. The algorithm has the advantage of scaling well to large data sets and is very fast in comparison to some other clustering techniques. Although, scalability is not an issue for this study as the number of instances is very small. The disadvantage if K-means is that it is capable of only producing spherical shapes for its clusters, so it is unable to pickup on more complex shapes. Furthermore, there is the issue of selecting the appropriate number of clusters.

There are three methods that can be used to suggest a suitable number of clusters. This includes the elbow method, silhouette score and the gap statistic method. The elbow method was applied to the data sets the results suggests an optimum value of 4 or 5 clusters while the silhouette score suggest 2 as the best while the gap statistic method is more inconsistent depending on the maximum clusters to try higher is better. From the TSNE diagram, it is understandable as to the reason for the silhouette score suggesting 2 clusters. However, after visual confirmation of the results for each number, 5 is chosen as it provides a good balance between being detailed enough and being too detailed.

3.2.4.2 Agglomerative Clustering

In agglomerative clustering each data point is assigned as its own cluster the data points that are closest are merged until some criteria is met in this case it is the number of clusters specified by the user. There are some hyperparameters that can be tuned by the user and one of it is the linkage type which determines how to clusters are formed. The default in Scikit-learn is ‘ward’ and it works by variance reduction, for this purpose the default settings are good enough. Like k-means it requires a user input for the number of clusters. This can be found using a dendrogram as can be seen in appendix A. The 2 red and green colour codes suggest that 2 clusters may be appropriate for this data set. However, 2 would not provide enough definition to the data for analysis so like before by visual confirmation by trial and error, 5 was determine as a good balance. The returned plot is near identical to that produced by k-means except for some data points.

3.2.4.3 Gaussian Mixture Clustering

Gaussian mixture is a case of k-means where in addition to using distances it uses probability to determine data point cluster assignments. Gaussian mixture has many more hyperparameters to tune than the previous two methods. The main ones in Scikit-learns version is n_components and covariance_type. The covariance type is selected to ‘full’ by default which is generally more computationally expensive since each component will have its own covariance matrix. The number of components can be difficult to decide on. Luckily, Scikit-learns’s implementation of this algorithm comes with a method to aid in this task. The method returns AIC and BIC values that can be plotted. Unfortunately, the return AIC and BIC values for this data set is very inconsistent. From the AIC plot the suggested components number is 4, 5 or 6 but BIC does not suggest any clear number. So, the selected chosen value

is 5 together with the default hyperparameters. Overall, the resulting clustering pattern is very similar to k-means and agglomerative but with more differences than between agglomerative and k-means.

Comparing of all three methods, agglomerative and k-means clustering successful picked up on data points that are significantly different and clustered them together which Gaussian mixture failed to do. Moreover, despite being very similar agglomerative produce overall more satisfactory clustering than k-means by clustering data points that are more similar together. Hence, agglomerative would be the better option among all three.

Chapter 4: Results

4.1 Clustering Visualisation

The clusters produced from the agglomerative clustering algorithm is plotted on a map using folium overlaid with the choropleth map of the property values (figure 15, figure 16).

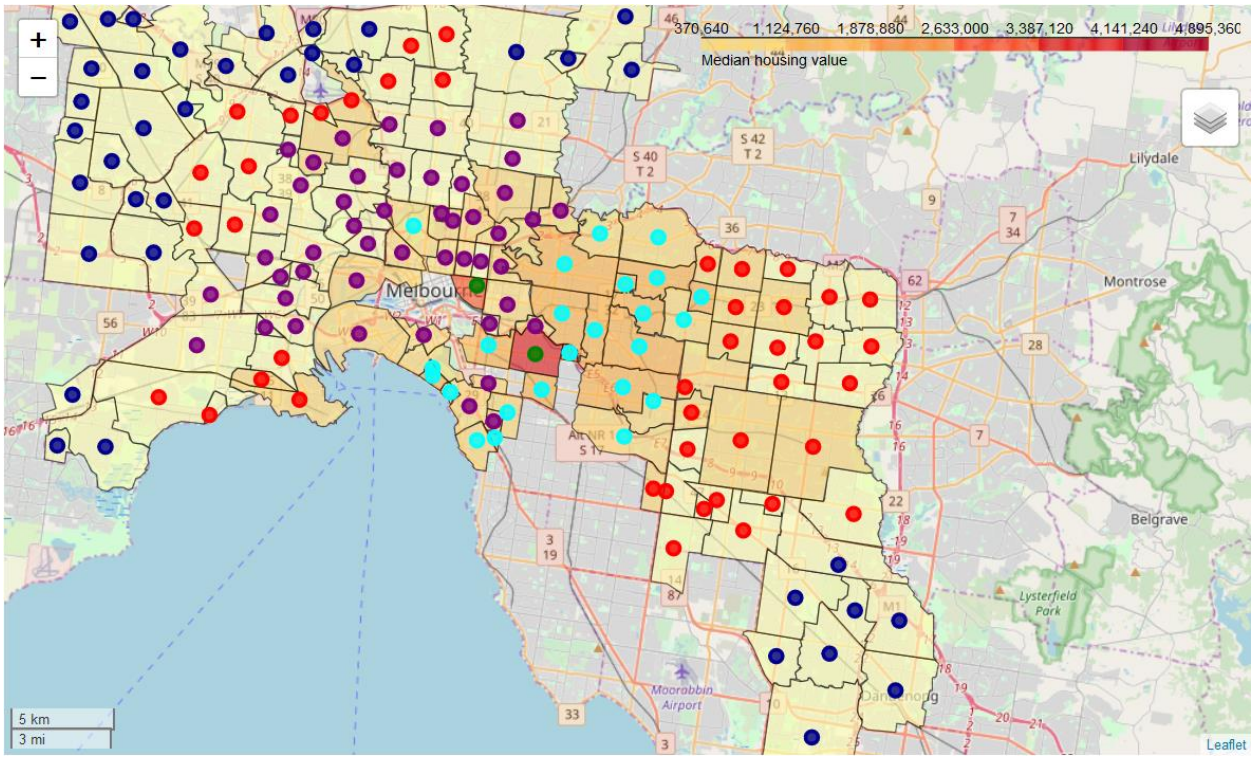


Figure 15: Housing clusters.

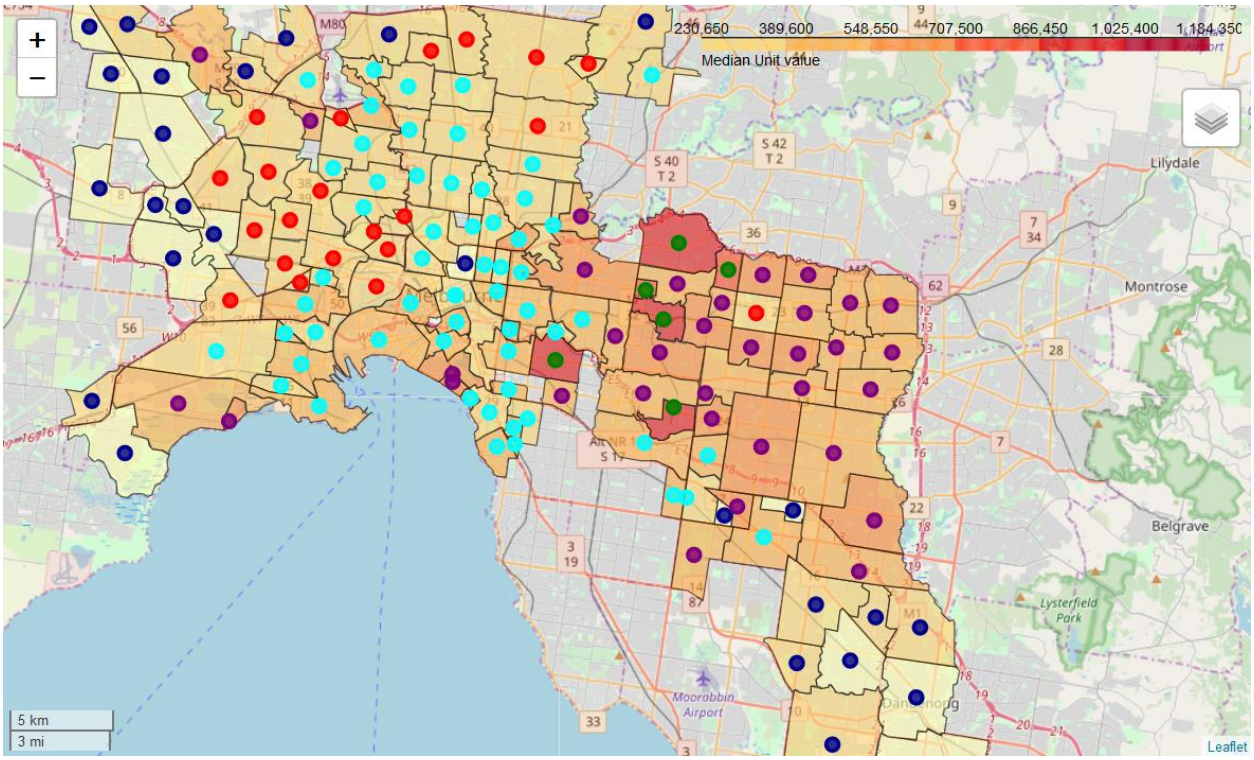


Figure 16: Unit clusters.

Folium enables interaction in the notebook by clicking on the circle markers a popup will show indicating the suburb and cluster label, but since this is impossible to do in this report a legend is given below:

Cluster label	Colour code
0	Purple
1	Cyan
2	Dark blue
3	Green
4	Red

Table 1: Cluster label and colour codes.

The descriptive statistics of for the clusters provides more insights to the clusters.

	House price $ (2016)							
	count	mean	std	min	25%	50%	75%	max
Cluster								
0	47.0	1.114649e+06	249704.628008	694000.0	916500.0	1075000.0	1302500.0	1577000.0
1	24.0	2.013000e+06	332300.381400	1567500.0	1791500.0	1890000.0	2238875.0	2758000.0
2	38.0	5.974474e+05	117387.081422	415000.0	512500.0	596500.0	650000.0	868000.0
3	2.0	4.253000e+06	845699.710299	3655000.0	3954000.0	4253000.0	4552000.0	4851000.0
4	43.0	9.895698e+05	219794.752239	560000.0	872500.0	1000000.0	1137750.0	1480000.0

Figure 17: Basic statistical data for housing values based on clusters.

	Unit price $ (2016)							
	count	mean	std	min	25%	50%	75%	max
Cluster								
0	34.0	6.926029e+05	60742.973083	590000.0	642750.0	698500.0	730000.0	850000.0
1	58.0	5.536983e+05	54029.064600	433500.0	516500.0	542500.0	576500.0	720000.0
2	27.0	3.784444e+05	45152.803247	240000.0	352500.0	382500.0	403500.0	449500.0
3	6.0	1.045333e+06	88751.713598	932500.0	983500.0	1054500.0	1085375.0	1175000.0
4	21.0	4.462143e+05	36192.738025	363000.0	430000.0	445500.0	475000.0	500000.0

Figure 18: Basic statistical data for apartment unit values based on clusters.

4.2 Discussion of Clusters

4.2.1 Housing dataset clusters

The map showing the clusters for housing shows a clear pattern. There are clusters cantered around Melbourne city. eastern Melbourne are populated by mostly 3 clusters while is west is 2 types. From the previous analysis of housing values, we can see that the algorithm clusters the most valuable neighbourhoods together as cluster 1 with a mean value of 2,000,000 which is the second highest. The third highest mean value at 1,100,000 is the nearest to Melbourne city located mostly to towards the north and west of the city. Interestingly, there are two suburbs that are were different enough to be put into a separate cluster containing only the two of them. A closer look reveals that these two suburbs of Toorak and east Melbourne are the highest value areas by a large margin, in fact east Melbourne cost almost 1,000,000 higher than the nearest suburb by worth, while Toorak is 1,000,000 + higher than east Melbourne. The large difference in price is clearly the reason for the separate cluster in addition to the venues being similar.

The least valuable neighbourhoods are clustered on the outer edges of the analysis area. The average cost of property in this cluster is 600,000. Cluster 4 is located mostly on the far eastern suburbs of the city and some on the western suburbs of Melbourne. The average cost of a house for this cluster is approximately 1,000,000. From domain knowledge, this is a reasonably accurate value.

4.2.2 Units dataset clusters

The clustering for apartments units are much more scattered than that of housing. Which is likely due to the nature of apartments purchases which are more ambiguous than houses. However, we can still see a similar pattern to that of houses. Again cluster 1 is situated mostly around Melbourne city but this for units it seems to be more spread out, this cluster is also the largest. The average cost of a unit is 550,000. The algorithm also successfully clustered the most expensive suburbs for apartment units as cluster 3 in green with average value of 1,000,000.

A large cluster can be seen in the far east of Melbourne. This one is on average the second most valuable with a mean of 700,000. Similarly, the suburbs with the lowest cost are grouped together, located mostly on the outer edges, except for Carlton in the city. As mentioned previously, this is because it mostly contains small apartments complexes with studio apartments as well as the place being mostly a commercial area rather than a residential area.

4.3 Cluster Venue Categories

4.3.1 Cluster 0

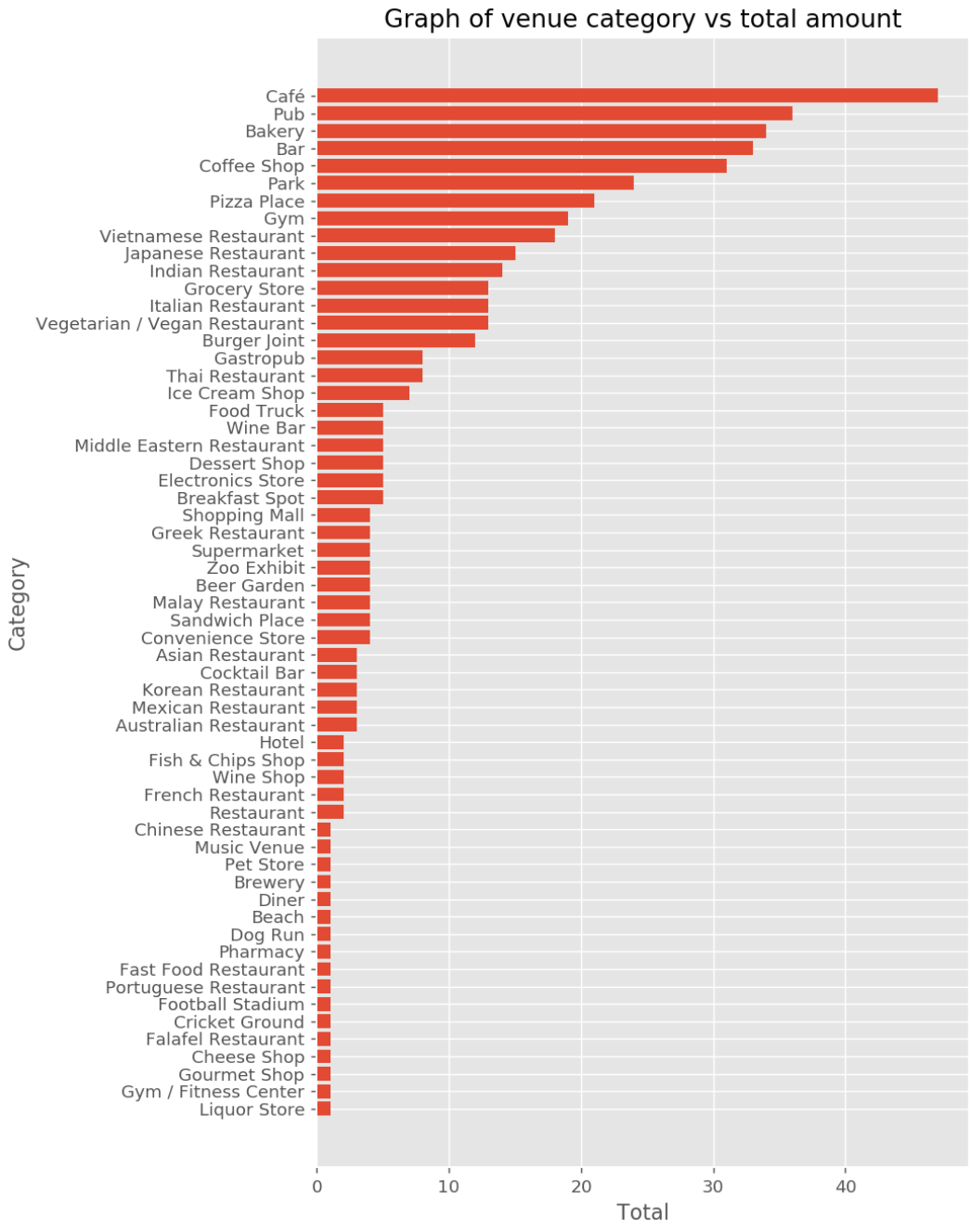


Figure 19: Bar graph of venues categories in cluster 0.

4.3.2 Cluster 1

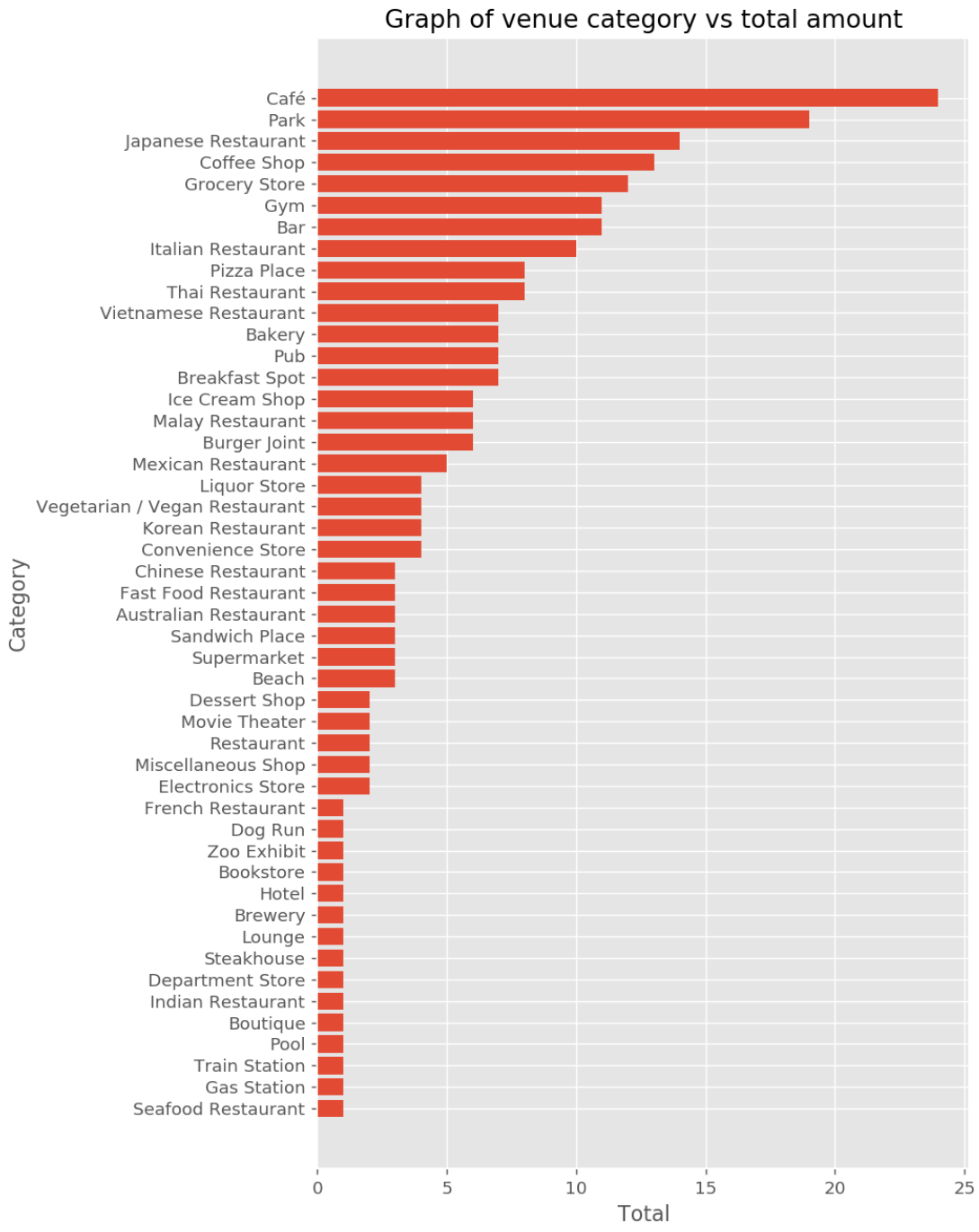


Figure 20: Bar graph of venues categories in cluster 1.

4.3.3 Cluster 2

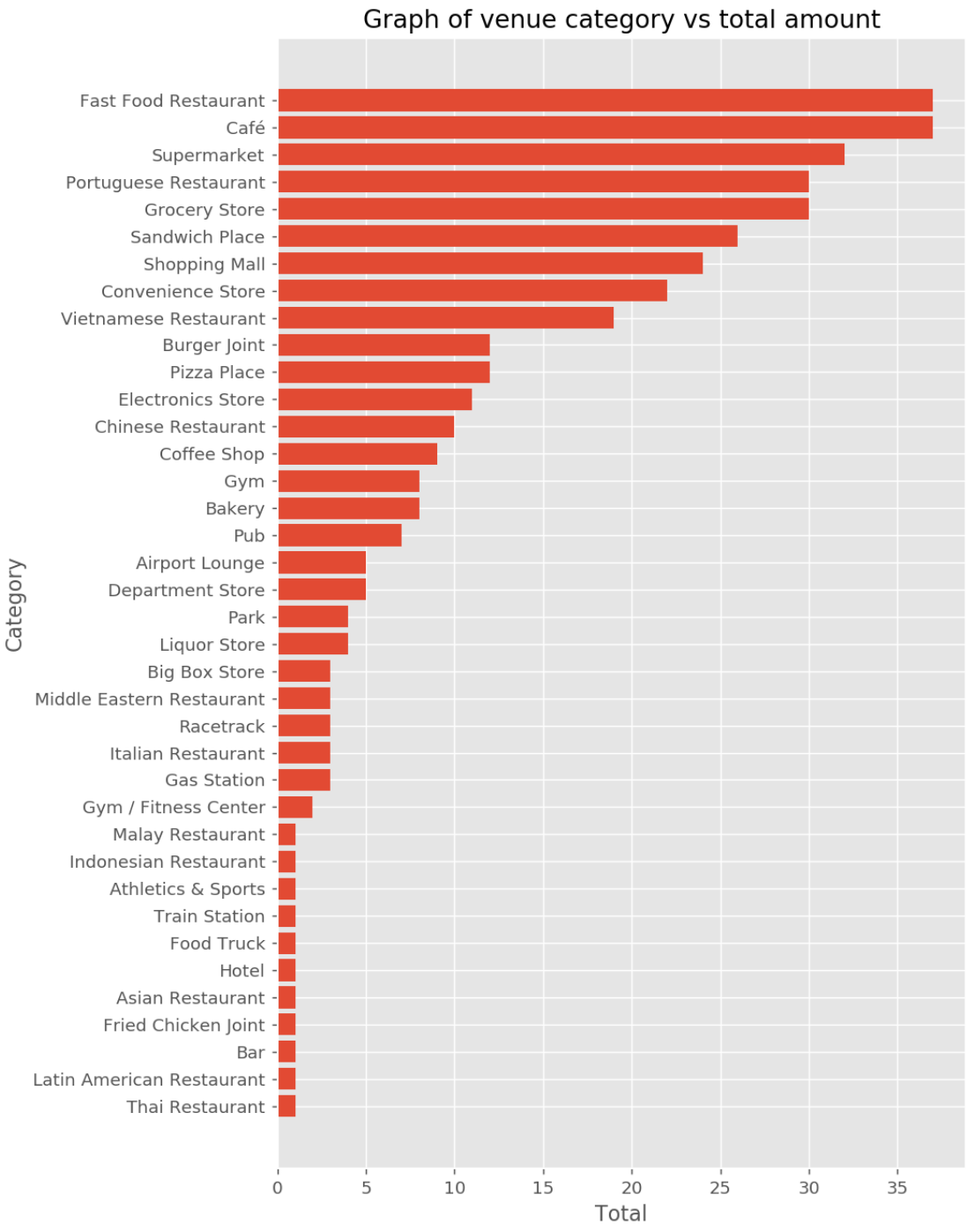


Figure 21: Bar graph of venues categories in cluster 2.

4.3.3 Cluster 3

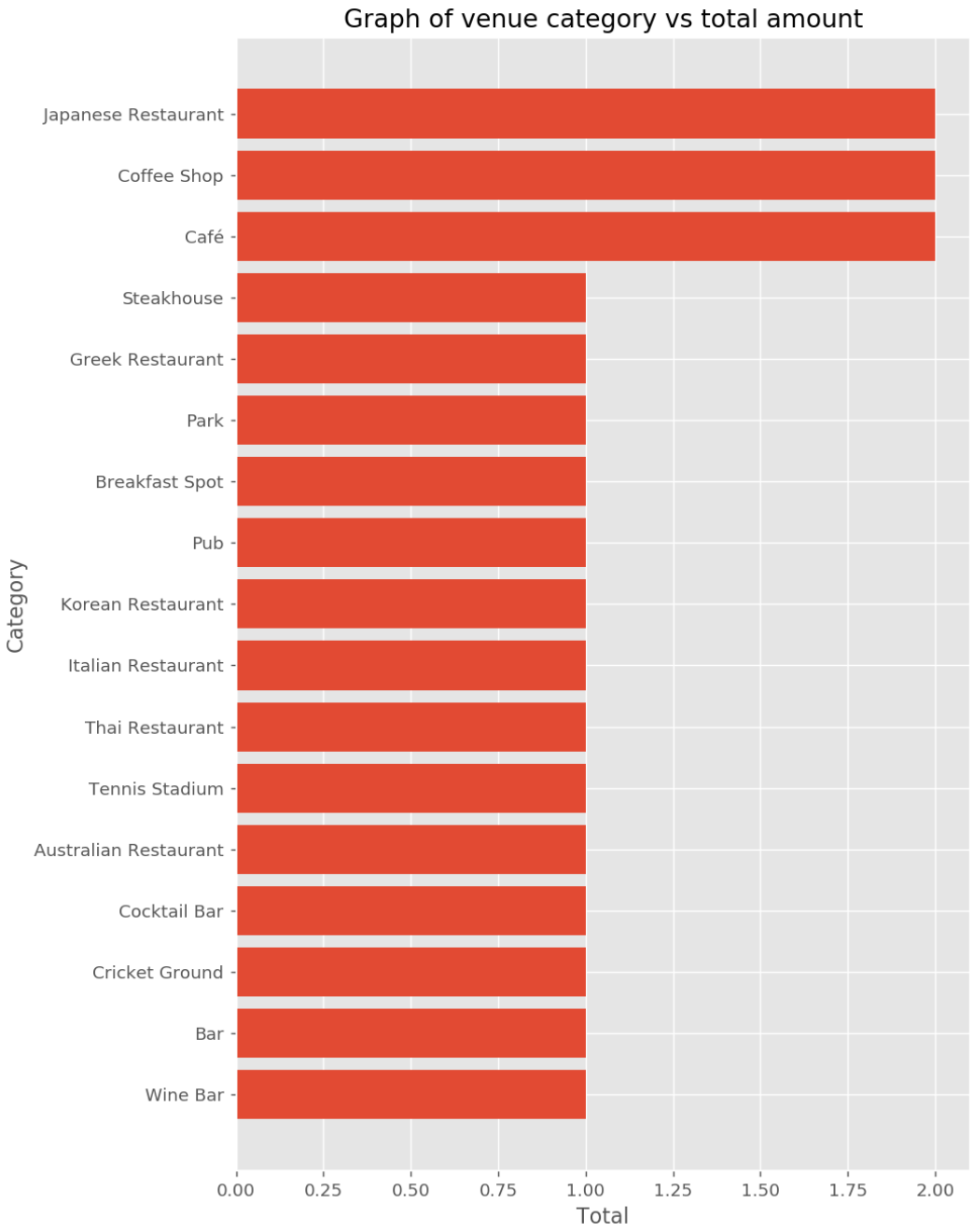


Figure 22: Bar graph of venues categories in cluster 3.

4.3.4 Cluster 4

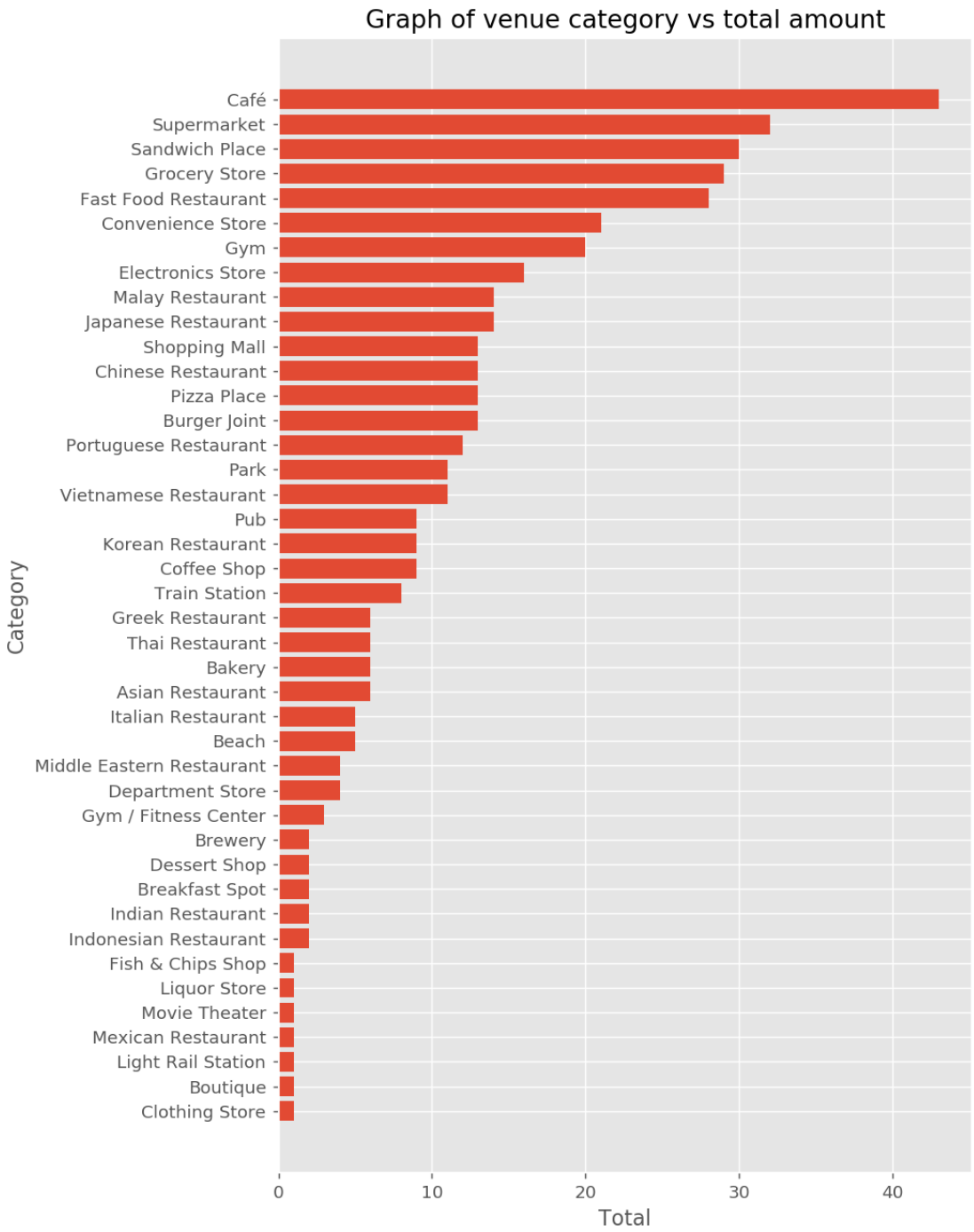


Figure 23: Bar graph of venues categories in cluster 4.

Chapter 5: Discussion

5.1 Housing Clusters and Categories

From the previous analysis of the suburb profiles it was shown that the inner eastern suburbs such as Toorak, Kooyong, Hawthorn, Deepdene and the like, contains the most valuable properties and is populated by higher income earners. Here, houses cost from 1,000,000 up to 5,000,000 while units are priced at 900,000 to 1,100,000. The median household income is also higher from 1600 to 2700 a week. While the lower value neighbourhood are mostly on the outer suburbs of Melbourne.

To begin with, the agglomerative clustering algorithm successfully clustered the high value areas into cluster 1 colour code in cyan. The mean house value here is 1,100,000 with a min of 700,000 and a max of 1,600,000. People living here are generally older with an average age of 38 and most are in their mid-30s with age going up in proportional to house value. As one would expect, the weekly household income is also high with most in the near or significantly above the 2k mark. One of the suburbs in this cluster is Parkville which is mostly made up of Melbourne Zoo with few residential properties. So, it should be taken with a grain of salt as it is an unusual suburb when compared to the rest.

The graph plotting the venue categories for this cluster shows a wide variety of venues especially restaurants. Japanese and Italian restaurants are considerably popular followed by a wide range of ethnic restaurants such as Thai, Vietnamese, Chinese, Korean, Malay, etc. In comparison fast food restaurants are one of the least popular on the list more so than vegan or vegetarian restaurants. This might be because wealthier people are more likely to be health conscious and can afford to buy healthier foods. In addition, other than Japanese and Italian eateries, other premium dining places includes French restaurants, steakhouses and seafood restaurants. Breakfast spots are also rather common. Bars and pubs are highly recommended here with bars being more popular than pubs.

Many shops and services are also available including movie theatres, breweries, hotels, boutiques, dessert shops, bookstores and electronics shops; while department store is the lowest on the list of stores and shops. Furthermore, there are places for sporting and recreational activities for example: pools, beaches and parks. Parks are the second highest common place behind cafes. This would make sense since, since it known that parks generally increase the value of nearby properties. Overall, we can say that this cluster contains premium to casual dining options together with many shops and recreational places.

A point of interest is cluster 3 in green which appears to contain suburbs that are very different such that the algorithm decided to put them into a separate cluster. The cluster has only two members: Toorak and East Melbourne. Their unusually high worth at 3,650,000 and 4,850,000 respectively resulted the suburbs standing out from the rest. Their proximity to the city would likely be a significant factor for their high value. In terms of venues, they are not much different from cluster 1 but in addition to the presence of a wine bar, cocktail bar, cricket ground and tennis stadium.

In contrast, the lowest value areas are put into cluster 2 coloured in blue on the map. The mean value for a house here is only 600,000 going as low as 420,000 and a high of 870,000. The median household income ranges from 1000 to 1800 a week while median ages are from 30s to 40s.

Popular eatery here is fast food restaurants followed by Portuguese restaurants with fast food being the most popular. We also note that there isn't as many restaurants' varieties here, in particular ethnic restaurants are considerably lower. Unlike wealthier suburbs, shopping malls and department stores are much more common. Pubs are bars are also not highly recommended with bars much less than pubs. In addition, the cluster in comparison is lacking in recreational spots with some parks being moderately common, racetracks and an athletic & sports category.

Moving on, cluster 4 in red mostly consist of suburbs in the City of Whitehorse at far eastern Melbourne and some in western Melbourne and south west regions. From the profile analysis of each suburb. These neighbourhoods are also rather valuable for housing with a mean of 1,000,000 a min of 560,000 and a max of 1,500,000.

The types of venues in cluster 4 appears to be a hybrid between cluster 1 and 2. Fast food is the most common restaurant type followed by many ethnic restaurants. Like cluster 1, Japanese restaurants are also popular. The presence of many Asian restaurants may suggest that suburbs here are popular with Asians. Entertainment places like Movie theatres, boutiques, shopping malls and departments store are present. Train stations are recommended here more so than other clusters. This is because a lot of people who work in the city also lives here. Interestingly, bars are not listed but there are several pubs and a couple of breweries. Parks are also a common sight together with a couple of beaches.

Finally, the last cluster is cluster 0 located in right Melbourne city and its surroundings suburbs spreading towards northern Melbourne. As expected, houses here are rather expensive costing on average of 1,100,000 from 700,000 to 1,600,000.

Being in and around the city, it's no surprise that these suburbs has many venues ranging from premium outlets, casual dining, shopping outlets and entertainment. Vietnamese, Japanese, Indian and Italian restaurants are highly rated. Vegetarian/vegan restaurants are also popular. There are ethnic restaurants including Thai, Middle eastern, Greek, Korean, Mexican, Chinese and Malay restaurants. Fast food and Portuguese restaurants are very much less popular than compared to other restaurants. Pubs and bars are the second and fourth highest recommended places behind cafes while other types like cocktail bars, brewery, beer gardens, wine bars, gastropubs are also available. As with any other city, shopping is a popular activity here. There are shopping malls, pet stores, pharmacy, electronics store, wine shops and gourmet shops. Entertainment centres are cricket grounds, football stadium, zoo exhibits, beaches and music venues.

Assigning suitable labels to the clusters for the housing data:

- Cluster 0: ['high value housing', 'shopping', 'premium outlets', 'entertainment spots', 'moderate/high income']
- Cluster 1: ['very high value housing', 'inner suburbs', 'high income', 'premium outlets', 'casual dining', 'recreational spots']
- Cluster 2: ['low value housing', 'outer suburbs', 'fast food', 'casual dining', 'low/moderate income']
- Cluster 3: ['extremely high value housing', 'high/very high income', 'inner city suburb', 'premium outlets', 'recreational spots']
- Cluster 4: ['valuable housing', 'Suburban', 'moderate income', 'ethnic restaurants', 'shopping', 'entertainment spots']

5.2 Units Clusters and Categories

The overall clusters representing the apartment units shares many similarities to that of housing. So, this part of the analysis will not go into too much detail for the venue categories of each suburb.

The first notable cluster is cluster 3. From the suburb profile of top value units, it can be seen that cluster 3 consist of the most valuable units with an average price tag of 1,000,000 a minimum of 900,000 and a maximum of 1,100,000; it is the only cluster to have a mean value of over a million. It also has the lowest number of suburbs of only 6. Like that of housing, these clusters are in the inner eastern suburbs of Melbourne. Thus, the venue categories are also similar to cluster 1 from housing with premium and casual locations present. The median age of the cluster is generally older between 39 to 47 years. As expected, the median household income is also high between 1.8k to 2.4k weekly.

The largest cluster is cluster 1. This cluster contains the city of Melbourne and its surrounding suburbs particularly towards the north. The mean value of units is 550,000. From the age groups it seems that younger people live in these suburbs especially within the city of Melbourne. The median age of Melbourne city itself is only 27 with its closest city suburbs of Docklands and Southbank at 30 years old. Melbourne city has two big universities: Melbourne University and RMIT, which means it has a large population of students while Melbourne city suburbs like Docklands is a financial district and residential area; therefore, is populated mostly by young professionals, hence the low ages. This also explains the lower income at 1.1k weekly for Melbourne while the city suburbs are higher at 1.8k. The overall cluster has a wide range of ages from 25 to early 40s but most people here are in their 30s. The same goes for income groups although this cluster contains some very high-income groups such as those in Cremorne and East Melbourne. Interestingly the unit value of these suburbs is not very high as one would expect being populated by some of the highest income earners. The venues in these areas are similar to that of cluster 0 and 1 from the housing analysis containing many higher end locations typical of cities.

Like housing, the outer suburbs are the lowest value for units as well. These suburbs are grouped into cluster 2, having a mean value of only 380,000 up to 500,000. The ages are between early 30s to late 40s with most being in their mid-30s. The income groups here are not that much different than that of cluster 1. This cluster includes an anomaly that is the suburb of Carlton which is only one located in Melbourne city in the cluster. The reason was explained previously as the suburb is mostly a commercial area as well as being adjacent to Melbourne University, so it is populated by university students thus containing mostly small studio apartments. This is corroborated by the low weekly income of only 572 and low age group of 25. The category venues are similar to cluster 2 from housing with fast food restaurants being the most popular venue and generally less diversity in venues.

Finally, we have cluster 0 and 4 on opposite sides of the city. cluster 0 is located on the far eastern suburbs while cluster 4 is on the western suburbs of Melbourne. Cluster 0 is on average more valuable than 4 with a mean of 700,000 compared to 450,000. It is already known that eastern Melbourne is more valuable than western Melbourne. This is particularly true for units in places like Box hill and Glen Waverly when considering cluster 0 as these suburbs is known to be popular with wealthier people. Glen Waverly has a very high population quite like Melbourne CBD which has 49,000 people, the suburb also has many apartments. The venue categories for these two clusters are like that of cluster 4 in housing.

Overall, we can say that the cluster for units are very similar to that of housing with some exceptions due the differences in unit values.

Assigning suitable labels to the clusters for the apartment units data:

- Cluster 0: ['very high value units', 'Suburban', 'moderate income', 'ethnic restaurants', 'shopping', 'entertainment spots']
- Cluster 1: ['high units', 'city/inner city', 'shopping', 'premium outlets', 'entertainment spots', 'low/moderate/high income']
- Cluster 2: ['low value units', 'outer suburbs', 'fast food', 'casual dining', 'low/moderate income']
- Cluster 3: ['extremely high value units', 'high/very high income', 'inner city suburb', 'premium outlets', 'recreational spots']
- Cluster 4: ['valuable units', 'Suburban', 'moderate income', 'ethnic restaurants', 'shopping', 'entertainment spots']

5.3 Usage and Recommendation Examples

The findings from this study can help people looking to buy property. For example, say a couple is looking to buy a house, their budget is up to $ 1,000,000 and their work place is in the city. The analysis would suggest them houses in cluster 0, 2 or 4 in ordinance with their budget. They are also self-proclaimed ‘foodies’ who enjoy trying new eateries and a drink on Saturday nights. The clusters can then be narrowed down to either cluster 0 or 4. Houses in cluster 0 can be bought for as low as $ 700, 000 while cluster 4 as low as $ 560,000. Between the two, suburbs in cluster 4 is perhaps the best option as houses here generally cost lower than in cluster 0 while still having a large

selection of places including many restaurants and pubs; also train stations are available so they can take the train to the city for work.

In another scenario, let's say a person is planning to open a Teppanyaki restaurant (a type of Japanese grilled cuisine) and he wants to know which area is best for his new restaurant. The analysis would not suggest suburbs in cluster 2 since people there generally do not eat Japanese food based on the types of venues, while all other clusters appear to be suitable for a Japanese restaurant. However, depending on how much he is willing to spend on his new business and how much he's charging for food, we would recommend suburbs in either cluster 0 or 1 as Japanese restaurants appear to be popular with people here and they are wealthier so they may be more willing to spend on good food. Overall, cluster 1 would be the preferred option since it is between cluster 0 and 4, so the restaurant can better attract people living in these two clusters as well.

Chapter 6: Conclusion and Future Work

6.1 Conclusion

The importance of location when dealing with property cannot be understated. This study has successfully obtained and analysed the data. In chapter 1, the scope and the objective of the project was clearly stated and in chapter 3 the selection process for the suburbs was explained. The required data was sourced, cleaned and the relevant features extracted. Data exploration was also carried out and key information was uncovered. The venues for every suburb were extracted using Foursquare's API. Unsupervised learning was applied to the datasets for housing and apartment units to cluster them based on value and venue categories. Three methods of clustering were tested and the agglomerative was chosen as the best.

In chapter 4, the results of the clustering were plotted and discussed and in chapter 5 it was shown that there is a relationship between property value and venues. Properties with higher values will have more of and a wider variety of venues. Premium eateries that are considered more expensive such as Japanese and Italian restaurants are much more common in wealthier neighbourhoods while lower value suburbs have more casual diner areas that are typically cheaper such as fast food. Wealthier neighbourhoods are also closer to the city and contains sporting facilities, entertainment centres and recreational areas. Property value decreases as the distance from the city increases. This means that lower value neighbourhoods are farther away from the city centre and they lack the services and amenities compared to their wealthier counterparts.

6.2 Future Work

Various assumptions were made throughout the course of this study due to time constraints, hardware limitations and the lack of data. It is recommended that the analysis be redone with more data including considering the size of the property, age, number of rooms, type if a house or building if an apartment unit. In addition, further values

should be use other than just the median and it might be worthwhile to view the change of property worth over time. If possible, the analysis area should also be increased by including more suburbs. Finally, using an up to date data set is recommended as this analysis is limited to data that is 2 years old.

Bibliography

1. Chalmers, S., *Australians say it's a good time to buy a house despite thinking prices will fall further: Westpac*. 2018, abc News: abc.com, 14 November 2018, https://www.abc.net.au/news/2018-11-14/australians-think-time-to-buy-house-despite-falling-prices/10496408, Access on: 6 March 2019.
2. Kusher, C., *Little value growth outside of Sydney and Melbroune over the past decade*. 2018, CoreLogic: CoreLogic.com, 19 Feb 2018, https://www.corelogic.com.au/news/little-value-growth-outside-sydney-and-melbourne-over-past-decade, Access on: 6 March 2019.
3. Yardney, M., *End Of 2018: Australian Property Market Report*. 2018, Your investment property magzine, 17 December 2018, https://www.yourinvestmentpropertymag.com.au/market-analysis/end-of-2018-australian-property-market-report-258690.aspx, Access on: 6 March 2019.
4. CoreLogic, *25 years of housing trends*. 2018, CoreLogic: Corelogic.com, https://www.aussie.com.au/home-loans/property-reports/25years.html, Access on: 6 March 2019.
5. Commision, V.E., *Local council maps*. Victorian Electoral Commision, https://enrol.vec.vic.gov.au/ElectoralBoundaries/LocalCouncilMaps.html, Access on: 6 March 2019.

Appendix

Appendix A

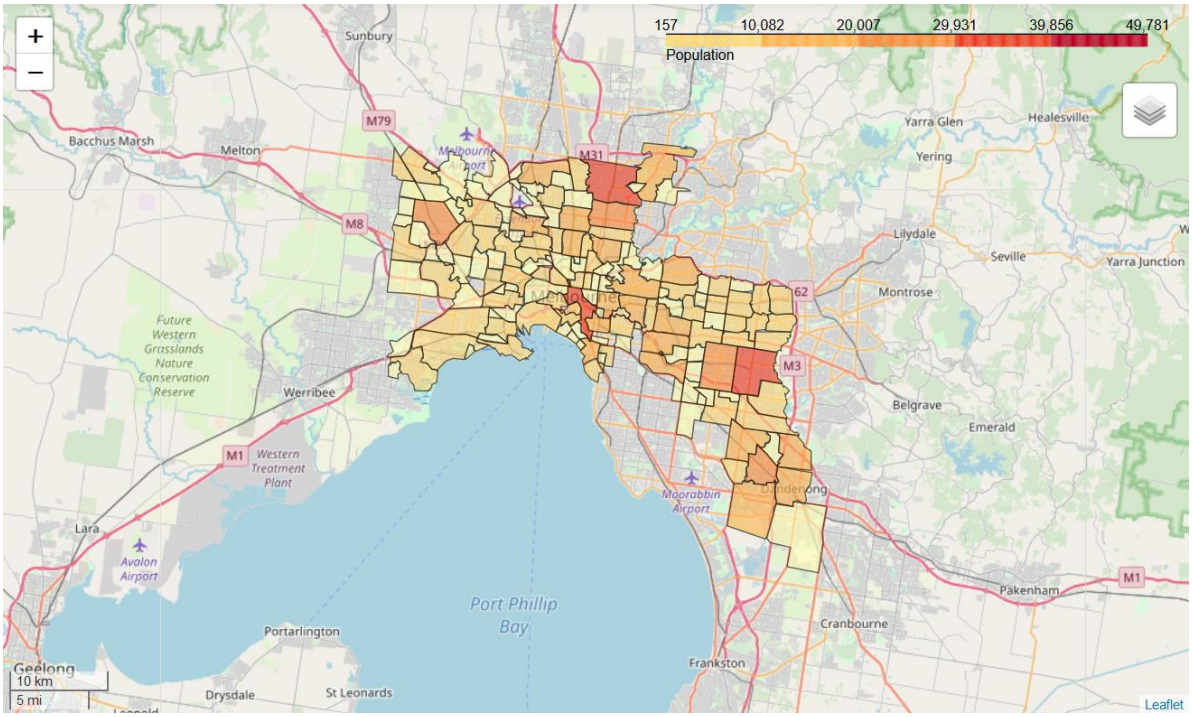


Figure A.1: Total population per suburb.

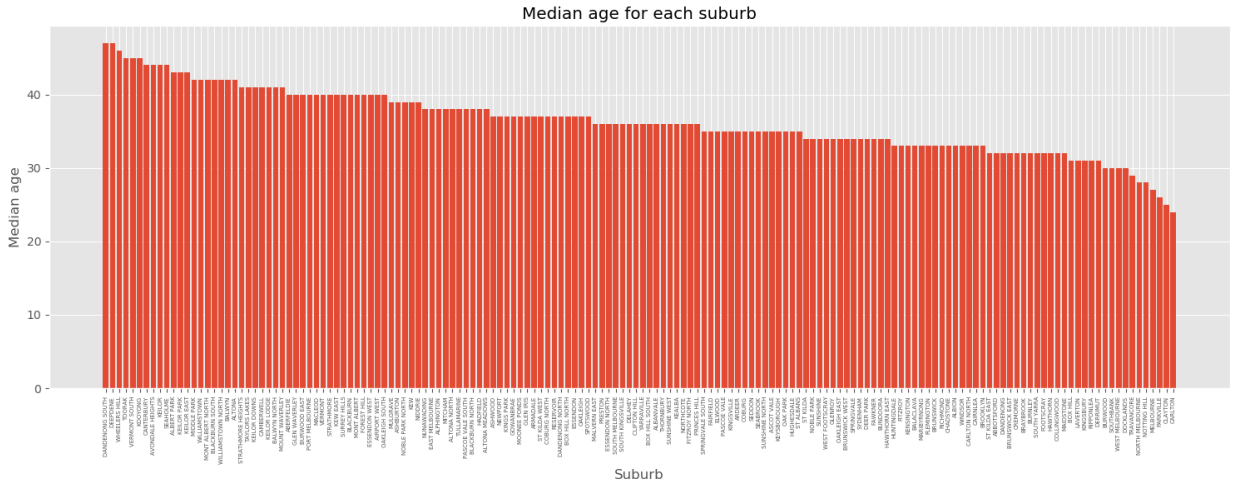


Figure A.3: Graph of median age.

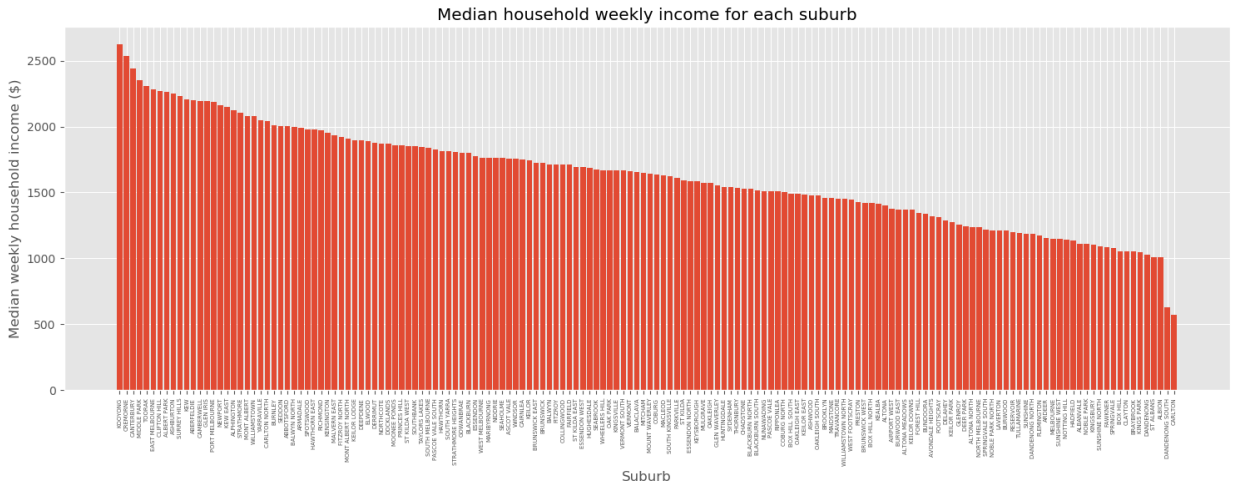


Figure A.4: Graph of median household.

Out[92]:

	Suburb	Category	Suburb lat	Suburb lon	Venue name	Lat	Lon	Venue address
0	BRUNSWICK EAST	Café	-37.76888	144.977682	Milkwood	-37.767576	144.980241	120 Nicholson St., Melbourne VIC 3057, Australia
1	BRUNSWICK EAST	Bar	-37.76888	144.977682	Mr Wilkinson	-37.769401	144.971969	295 Lygon St, Melbourne VIC 3056, Australia
2	BRUNSWICK EAST	Greek Restaurant	-37.76888	144.977682	Hellenic Republic	-37.764496	144.973040	434 Lygon St, Melbourne VIC 3057, Australia
3	BRUNSWICK EAST	Coffee Shop	-37.76888	144.977682	Padre Coffee	-37.764107	144.973280	438-440 Lygon St (Stewart St), Brunswick East ...
4	BRUNSWICK EAST	Café	-37.76888	144.977682	East Elevation	-37.767682	144.972339	351 Lygon St, Melbourne VIC 3057, Australia
5	BRUNSWICK EAST	Sake Bar	-37.76888	144.977682	Kumo Izakaya & Sake Bar	-37.773424	144.971492	152 Lygon St (O'Conner St), Melbourne VIC 3057...
6	BRUNSWICK EAST	Bar	-37.76888	144.977682	The Alderman	-37.773997	144.971299	134 Lygon St, Brunswick East VIC 3057, Australia
7	BRUNSWICK EAST	Chocolate Shop	-37.76888	144.977682	Monsieur Truffe	-37.767579	144.972279	351 Lygon St, Brunswick VIC 3057, Australia
8	BRUNSWICK EAST	Community Center	-37.76888	144.977682	CERES Community Environment Park	-37.765743	144.983281	Roberts St (Stewart St), Brunswick VIC 3057, A...
9	BRUNSWICK EAST	Brewery	-37.76888	144.977682	Temple Brewing Company	-37.776028	144.971710	122 Weston St, Brunswick VIC 3057, Australia

Figure A.5: First 10 values from the data frame created from Foursquare's data.

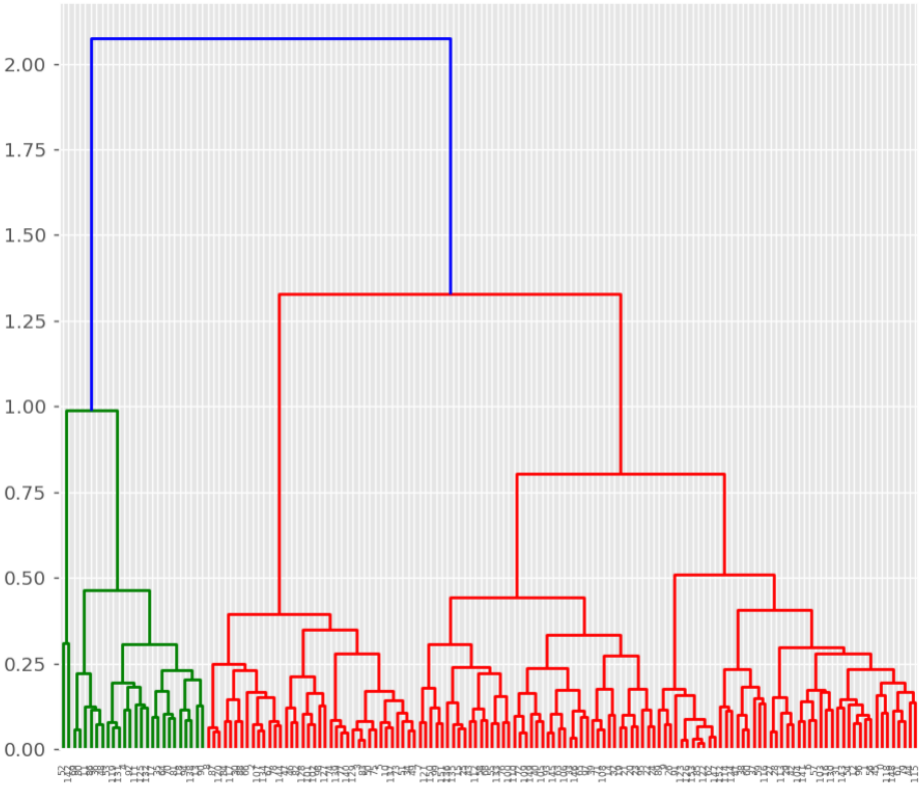


Figure A.6: Agglomerative clustering dendrogram.