An insight to Housing Prices and Venues in Suburbs of Melbourne

Introduction

Melbourne is the capital and most populous city of the Australian state of Victoria, and the second most populous city in Australia. Recently it is seeing a lot of migration from overseas as well as from onshore [1]. Due to this Melbourne is experiencing high population growth, generating high demand for housing. This housing boom has increased house prices and rents, as well as the availability of all types of housing. Subdivision regularly occurs in the outer areas of Melbourne, with numerous developers offering house and land packages. However, it is always a problem when you want to move far away from the city. The concern would always be if there is availability of shopping mall, bakery, coffee shops, public transport, restaurant and so on which are essential if you are looking to buy a house. The good bargain would be to purchase a house in a suburb where the price is cheaper and nearby availability of all the necessary amenities. Therefore, in the report we first analyze the clusters of different suburbs. Then we compare two suburbs one expensive and the other cheap to see if it is worthwhile to purchase a house in a cheaper suburb.

Data Description

#DATASET FROM KAGGLE.COM MELBOURNE HOUSING MARKET @ AUTHOR TONY PINO [2]

1. The raw data consists of 34857 rows and 21 columns. Out of 21 columns, the price is the target variable and rest are features. The features include both strings and numerical data types. The strings data types are fore example, suburb name, address, type of house (h-house, u-unit, t-townhouse), council and region name etc. The numeric data are for example: number of rooms, bathrooms and car park spaces, distance from the city, land-size and building area, location details (latitude and longitude). The shape of the dataset and the types of data are shown in Fig. 1. The dataset was cleaned to contain relevant data, rooms to be between 1 and 6, Bedroom to be between 1 and 6, bathroom to be between 1 and 3, car parking spaces to be between 1 and 4, and building area and land size between 1 and 1000 m^2. The dataset now became smaller with 10331 rows. Then, an exploratory data analysis was performed to see how price varied with different variables. Further, it was found that there were 307 unique suburbs. A new data-frame was constructed grouped by suburbs, the mean value of price, postcode, latitude and longitude of each suburb.

```
In [6]: df_data.shape
Out[6]: (34857, 21)
```

```
In [10]: df_data.dtypes
Out[10]: Suburb            object
         Address           object
         Rooms              int64
         Type              object
         Price            float64
         Method            object
         SellerG           object
         Date              object
         Distance         float64
         Postcode         float64
         Bedroom2         float64
         Bathroom         float64
         Car              float64
         Landsize         float64
         BuildingArea     float64
         YearBuilt        float64
         CouncilArea       object
         Lattitude        float64
         Longtitude       float64
         Regionname        object
         Propertycount    float64
         dtype: object
```

Fig 1: Shape and the types pf the dataset.

2. Then, I used Foursquare API to get the most common venues of all suburbs present in the data-frame [3].
3. Finally, I collected latest data from Australian Bureau of Statistics webpage to compare the suburbs statistics of number of people, families, details of income and rent/repayments of mortgage for given suburb [4].

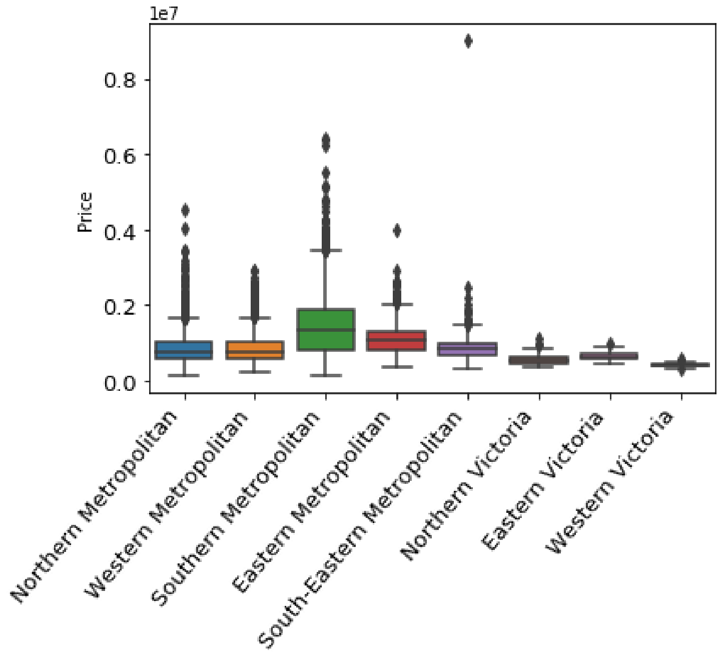
Methodology

The describe function was used to generate a descriptive statistic of the original data as follows:

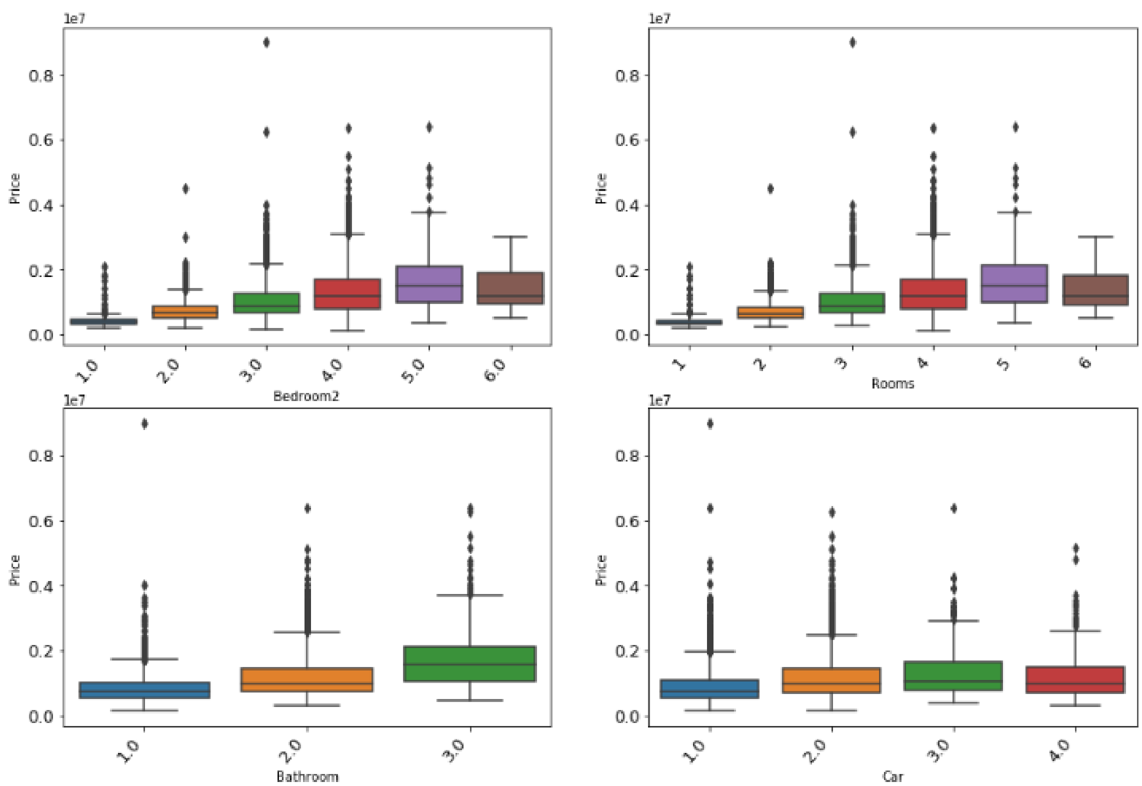
```
In [61]: df_data.describe()
Out[61]:
```

	Rooms	Price	Distance	Postcode	Bedroom2	Bathroom	Car
count	10331.000000	8.058000e+03	10331.000000	10331.000000	10331.000000	10331.000000	10331.000000
mean	3.133869	1.057663e+06	11.352715	3114.394831	3.109283	1.652889	1.748814
std	0.919924	6.163612e+05	6.634150	112.316283	0.917593	0.646478	0.744335
min	1.000000	1.310000e+05	0.000000	3000.000000	1.000000	1.000000	1.000000
25%	3.000000	6.350000e+05	6.900000	3046.000000	3.000000	1.000000	1.000000
50%	3.000000	8.800000e+05	10.500000	3095.000000	3.000000	2.000000	2.000000
75%	4.000000	1.314250e+06	14.000000	3152.000000	4.000000	2.000000	2.000000
max	6.000000	9.000000e+06	48.100000	3977.000000	6.000000	3.000000	4.000000

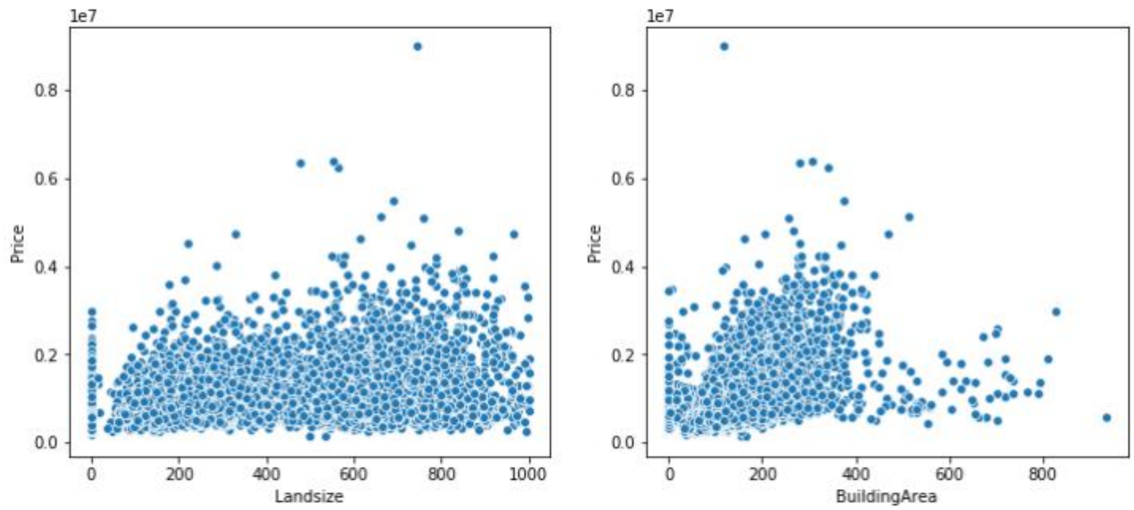
Using seaborn boxplot comparison of housing price was made for different regions in Melbourne. It was found the Southern Metropolitan had the highest median housing price among 8 different regions.



Similarly, box plots were used to compare housing prices for features such as number of bedrooms, room, bathroom and car park space. The median price of the house increased with the increase in the number of rooms, bedrooms, bathroom and car park space indicating a positive linear relationship between price and the features.



Again, a scatterplot was used to observe if any relationship between housing price, the landsize and the building area exists. It can be seen as landsize and building area increases the price increases. However, price for same building area/ landsize could be higher or lower depending upon the suburb location.



Therefore, a new dataset was constructed with Suburb, postcode, price, latitude and longitude to examine top venues for different suburbs as follows:

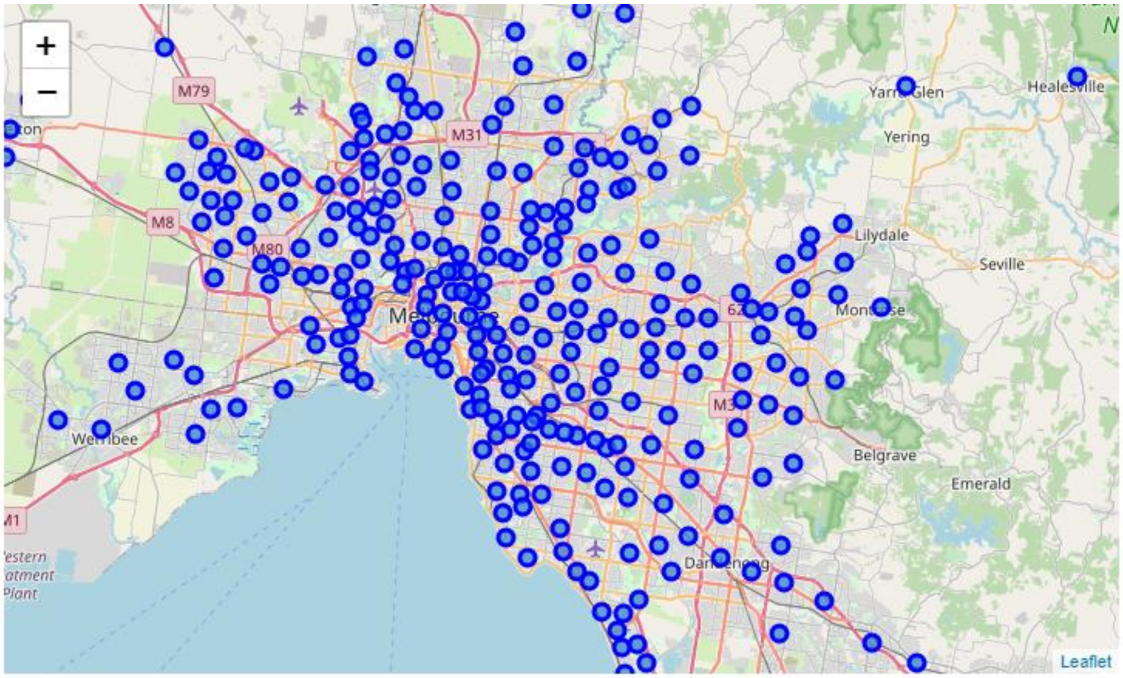
```
In [80]: df_d1.head()
```

Out[80]:

	Suburb	Postcode	Price	Lattitude	Longtitude
0	Abbotsford	3067	1.073117e+06	-37.803990	144.996601
1	Aberfeldie	3040	1.359462e+06	-37.759304	144.897928
2	Airport West	3042	7.402353e+05	-37.724090	144.879632
3	Albanvale	3021	5.403333e+05	-37.744537	144.768773
4	Albert Park	3206	2.257857e+06	-37.844672	144.952423

The geolocator was used to get the geographical coordinates of Melbourne. And, the python folium library was used to generate map of Melbourne and superimpose different suburbs using the latitude and longitude values as below.

Out[83]:



Then, the Foursquare API was utilized to explore the suburbs and segment them. The request to the Foursquare API was set by the limit as 100 venues and the radius of 2000 meter for each suburb from their given latitude and longitude information. The head of the data-frame of Venues name, category, latitude and longitude information from Foursquare API merged with suburb is shown below. There were 14289 venues collected and after investigation 355 unique categories were found.

In [253]:
```
melbourne_venues.groupby('Suburb').count().head()
```

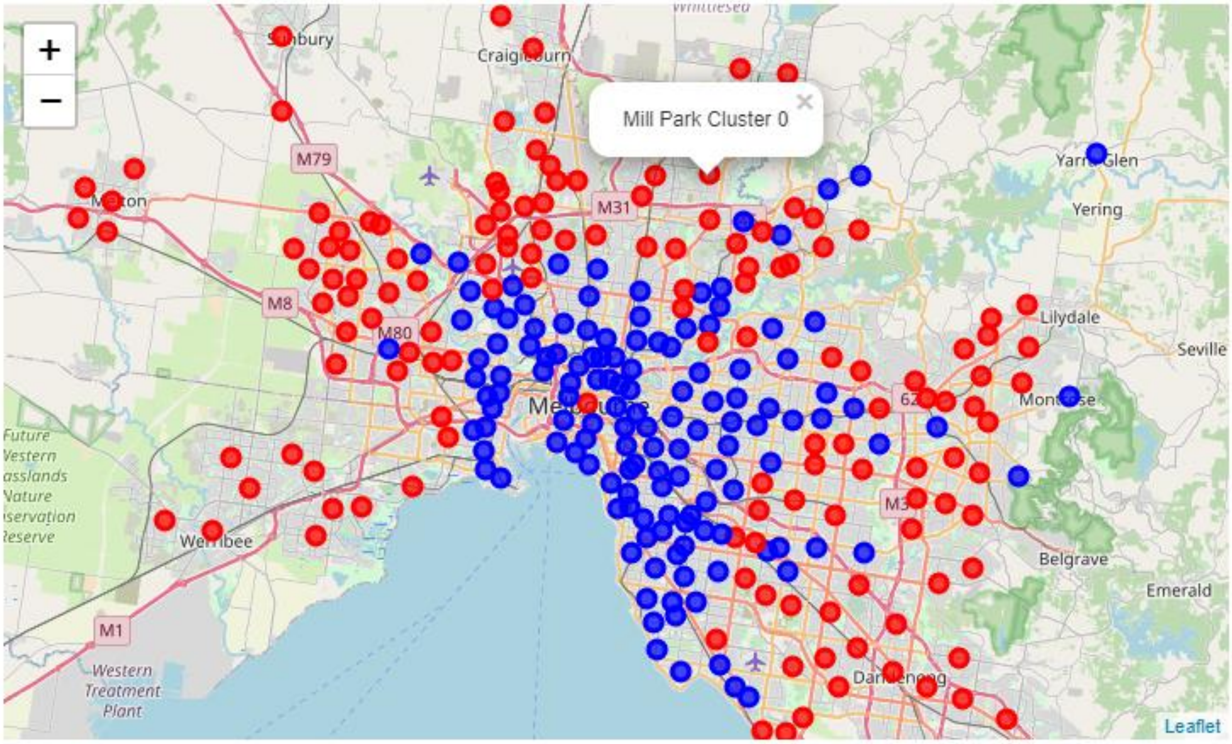
Out[253]:

Suburb	Suburb Latitude	Suburb Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Abbotsford	100	100	100	100	100	100
Aberfeldie	90	90	90	90	90	90
Airport West	53	53	53	53	53	53
Albanvale	12	12	12	12	12	12
Albert Park	100	100	100	100	100	100

In [93]:
```
print('There are {} uniques categories.'.format(len(melbourne_venues['Venue Category'].unique())))
```

```
There are 355 uniques categories.
```

To analyze each suburb one hot encoding was performed for the venue category. And the rows were grouped by suburbs and frequency occurrence. Then, unsupervised learning K-means algorithm was used to cluster the suburbs into two clusters. And, python folium library was used to generate Melbourne map and superimpose the clustered suburbs into the map as follows. Red indicates Cluster 0 and blue indicated cluster 1. In general, the blue clusters represent the suburbs nearby central Melbourne and red represent the outer suburbs.



Then, further comparison for given suburb was made using the census data available in the Australia Bureau of statistics (ABS). After asking the postcode of the suburb, a code was written which would link to the ABS website and collect the information on the suburb. Then a table is generated to compare several Census Stats as follows:

Out[212]:

Census Stats	3168	3174
People	20197	38461
Male	50.4%	50.2%
Female	49.6%	49.8%
Median age	27	35
Families	3824	9905
for families with children	1.5	1.8
for all families	0.4	0.8
All private dwellings	8066	14797
Average people per household	2.7	2.7
Median weekly household income	$1,069	$1,127
Median monthly mortgage repayments	$1,800	$1,500
Median weekly rent	$360	$300
Average motor vehicles per dwelling	1.3	1.6

Results

The average house price of the different clusters was printed as follows. The prices of the house in cluster 0 is significantly cheaper than cluster 1.

```
In [228]: print('The mean price of cluster 0 (red) is AUD {}.'.format(int(cluster1.Price.mean())))
          print('The mean price of cluster 1 (blue) is AUD {}.'.format(int(cluster2.Price.mean())))

          The mean price of cluster 0 (red) is AUD 763919.
          The mean price of cluster 1 (blue) is AUD 1190210.
```

Two postcodes were requested for the purpose of comparison as follows:

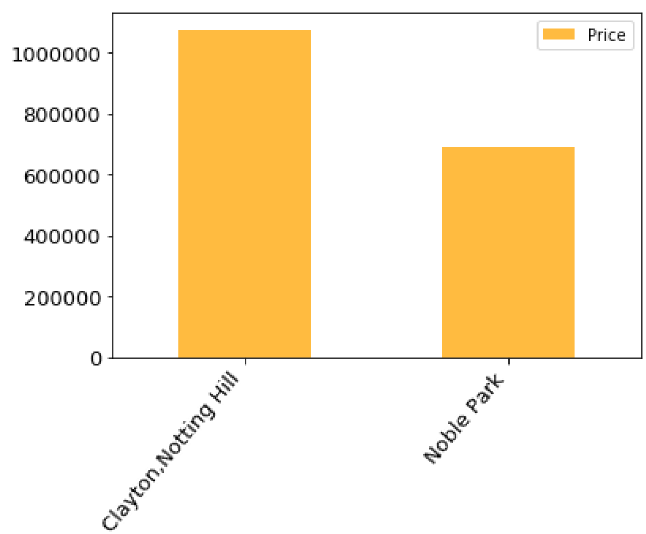
```
In [234]: Nei1=int(input("Enter the Postcode#1: "))

          Enter the Postcode#1: 3168

In [181]: Nei2=int(input("Enter the Postcode#2: "))

          Enter the Postcode#2: 3174
```

After entering the following postcodes, the code would generate a comparison table as follows and a comparison bar plot as shown below:



t[274]:

	3168	3174
Suburb	Clayton,Notting Hill	Noble Park
1st Most Common Venue	Café	Fast Food Restaurant
2nd Most Common Venue	Furniture / Home Store,Sandwich Place	Chinese Restaurant
3rd Most Common Venue	Malay Restaurant,Asian Restaurant	Gym
4th Most Common Venue	Shopping Mall,Indonesian Restaurant	Sandwich Place
5th Most Common Venue	Vietnamese Restaurant,Bakery	Supermarket
6th Most Common Venue	Sandwich Place,Grocery Store	Pizza Place
7th Most Common Venue	Portuguese Restaurant,Fast Food Restaurant	Indian Restaurant
8th Most Common Venue	Supermarket,Korean Restaurant	Vietnamese Restaurant
9th Most Common Venue	Supermarket,Italian Restaurant	Seafood Restaurant
10th Most Common Venue	Electronics Store,Convenience Store	Thrift / Vintage Store
Price	1.07744e+06	688167
Cluster Labels	1	0
People	20197	38461
Male	50.4%	50.2%
Female	49.6%	49.8%
Median age	27	35
Families	3824	9905
for families with children	1.5	1.8
for all families	0.4	0.8
All private dwellings	8066	14797
Average people per household	2.7	2.7
Median weekly household income	$1,069	$1,127
Median monthly mortgage repayments	$1,800	$1,500
Median weekly rent	$360	$300
Average motor vehicles per dwelling	1.3	1.6

Discussion

It can be observed that the given two different post codes indeed belong to two different clusters, therefore they have significantly different housing price. Clayton/ Notting Hill being the expensive one and the Noble Park being the cheaper one. Looking at the top 10 venues it can been both suburbs have nearby amenities ranging from café, grocery store, supermarket, restaurants, etc. Further, details from the census stats Noble park has a greater number of families and families with children living. Also, it has higher number of private dwellings with larger car space. Further, the monthly mortgage repayments and rent is cheaper compared to clayton. Therefore, comparing Noble park and Clayton, buying a house in Noble park seems to be better if you are considering a cheaper price with nearby amenities.

However, the current analysis only factors price and facilities in the suburb. In future, it should be noted considering other factors such as crime rates, school ratings, etc to make an analysis if the suburb is a good purchase place beside price and facilities.

Conclusion

In this report, an analysis was performed to cluster suburbs depending upon the prices and other facilities such as restaurants, café, supermarket. Further, census stats were used to get average statistics people, families, and dwellings in a suburb. Based on the analysis several recommendations such as the suburb is cheap but not enough amenities are present, the suburb is cheap and necessary facilities are nearby, the suburb is expensive but not all facilities are present, etc. could be inferred. For comparison two suburbs were compared and a suggestion was provided.

References

[1] Melbourne- Wikpedia

[2] https://www.kaggle.com/anthonypino/melbourne-housing-market

[3] Foursquare API

[4] https://www.abs.gov.au/