An insight to Housing Prices and Venues in Suburbs of Melbourne

Siddhartha Shrestha

Background and Business Problem

- 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.
- However, the prices of house in the city is expensive while 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.

Australian Bureau of Statistics

FOURSQUARE

Kaggle

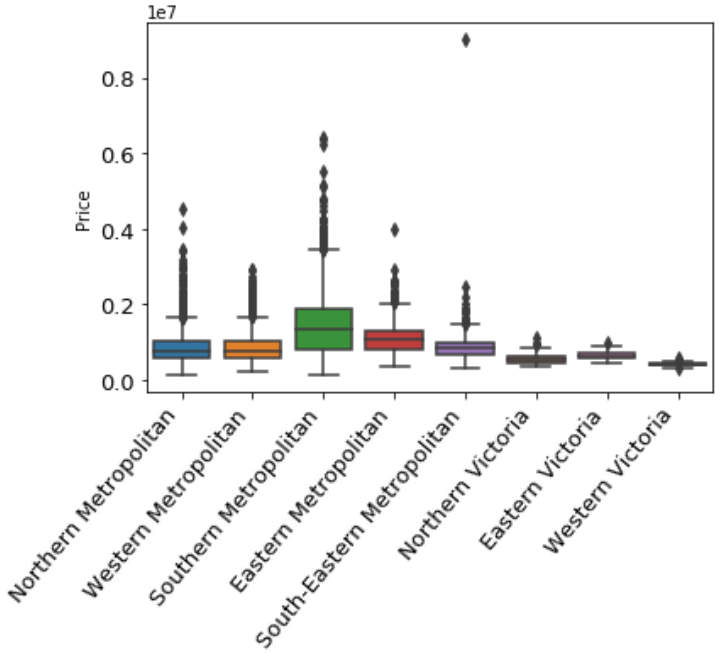

```
In [6]: df_data.shape
Out[6]: (34857, 21)

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
```

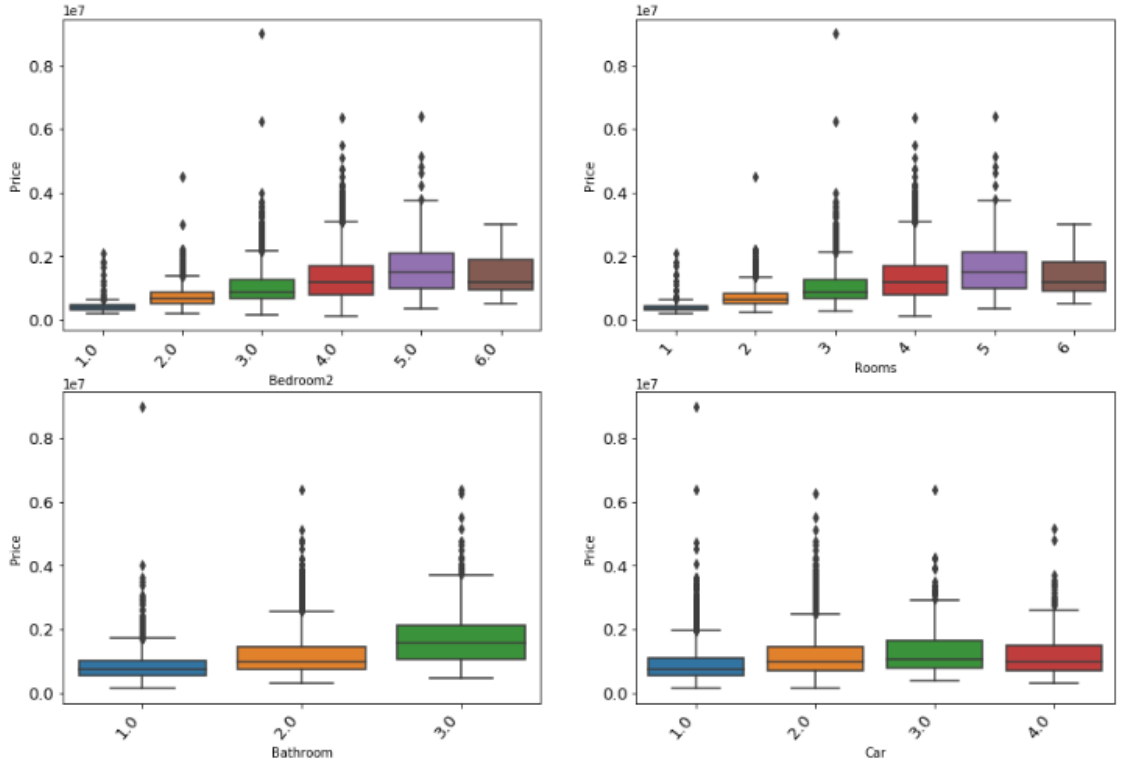
Data Description

- 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).

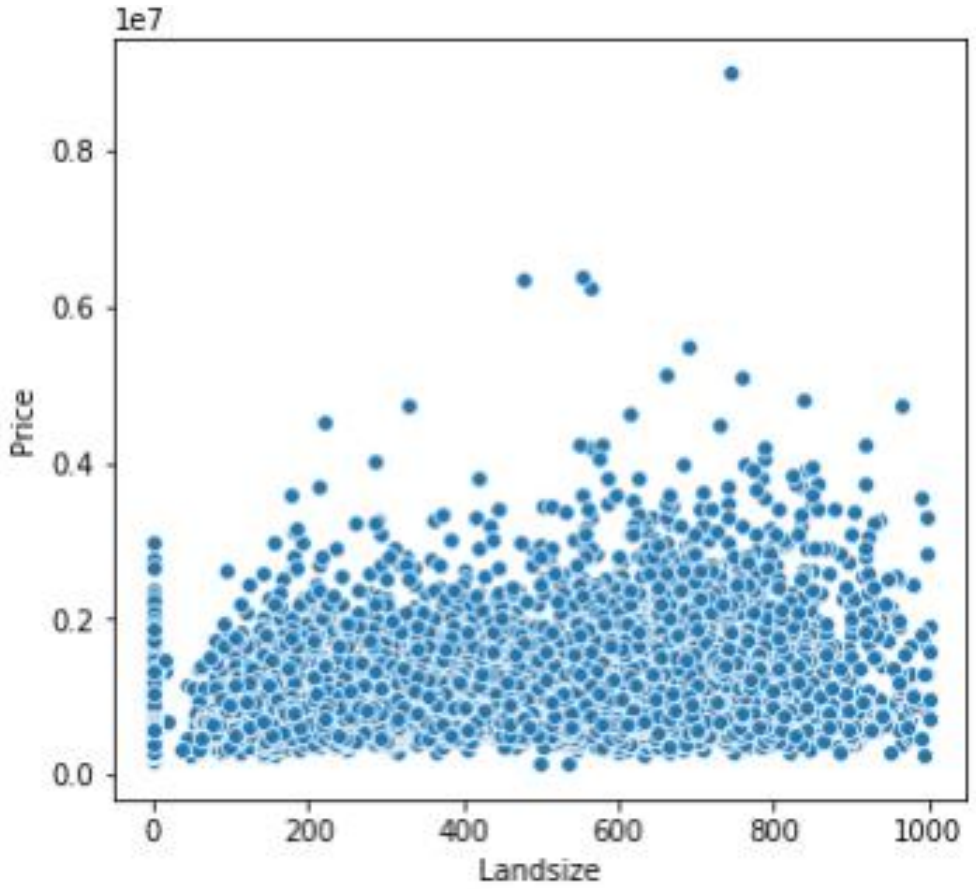
Methodology

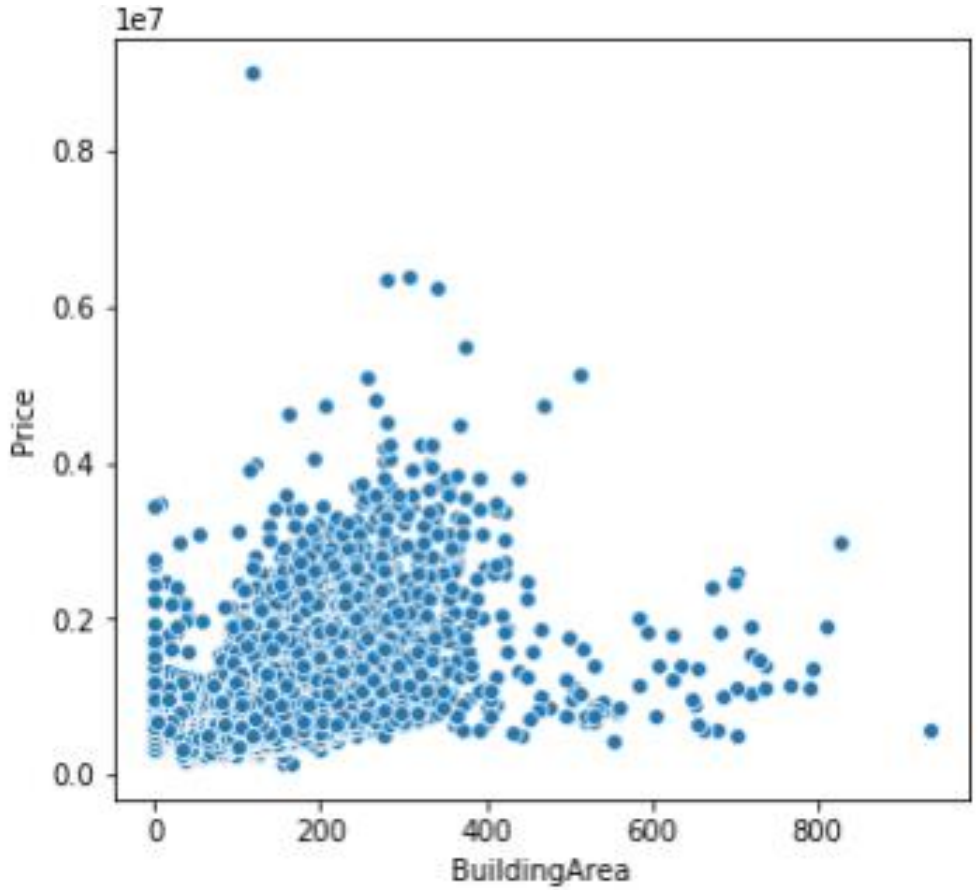
- First it was analyzed if the housing price is affected by regions, number of bedrooms, number of rooms, number of bathroom and number of car park spaces





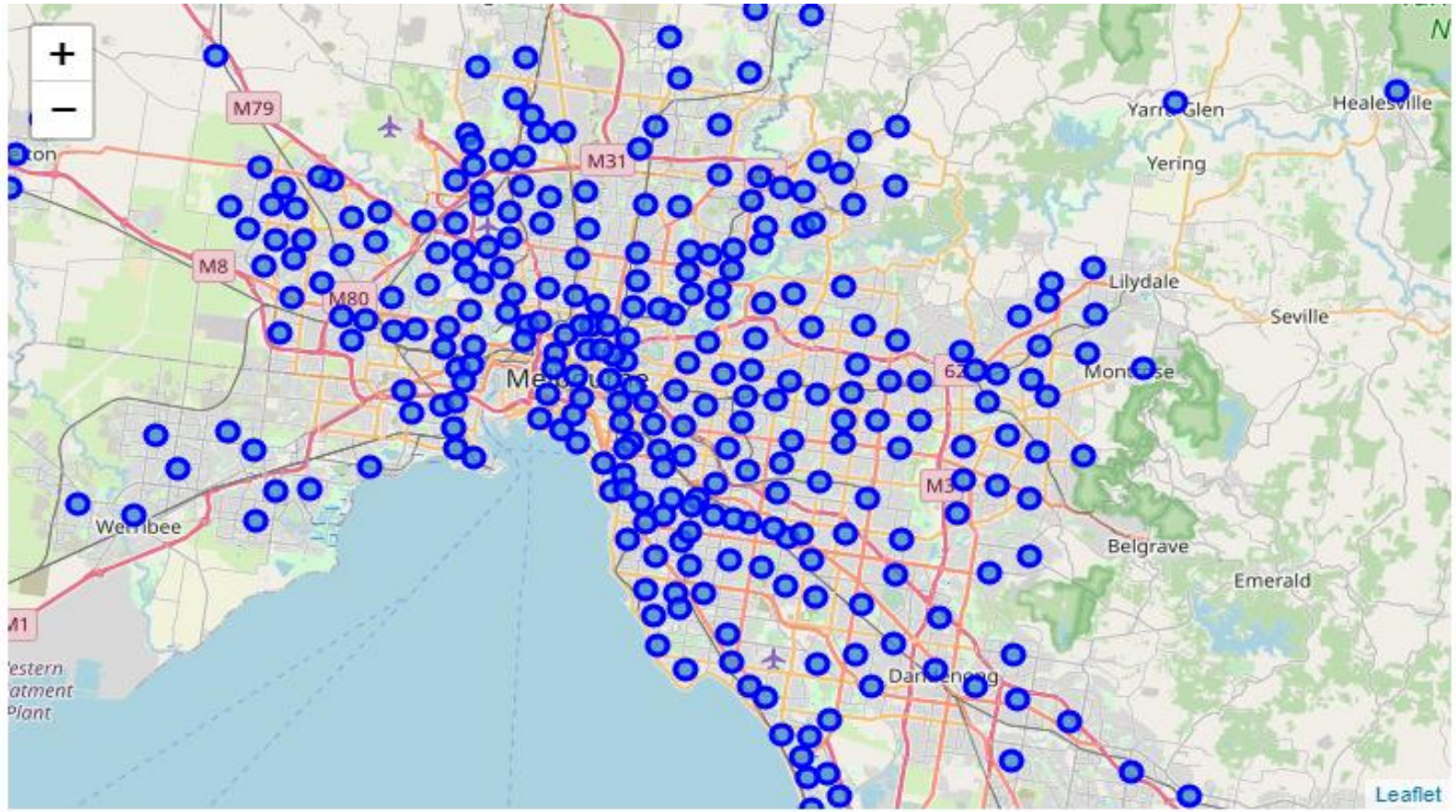
Further it was checked if any correlation of house prices with land size and building area. Both feature tend to increase the house price if it increases however, the data also suggest for the same size of build area and land size the price can be significantly different suggesting it may also depend upon the location



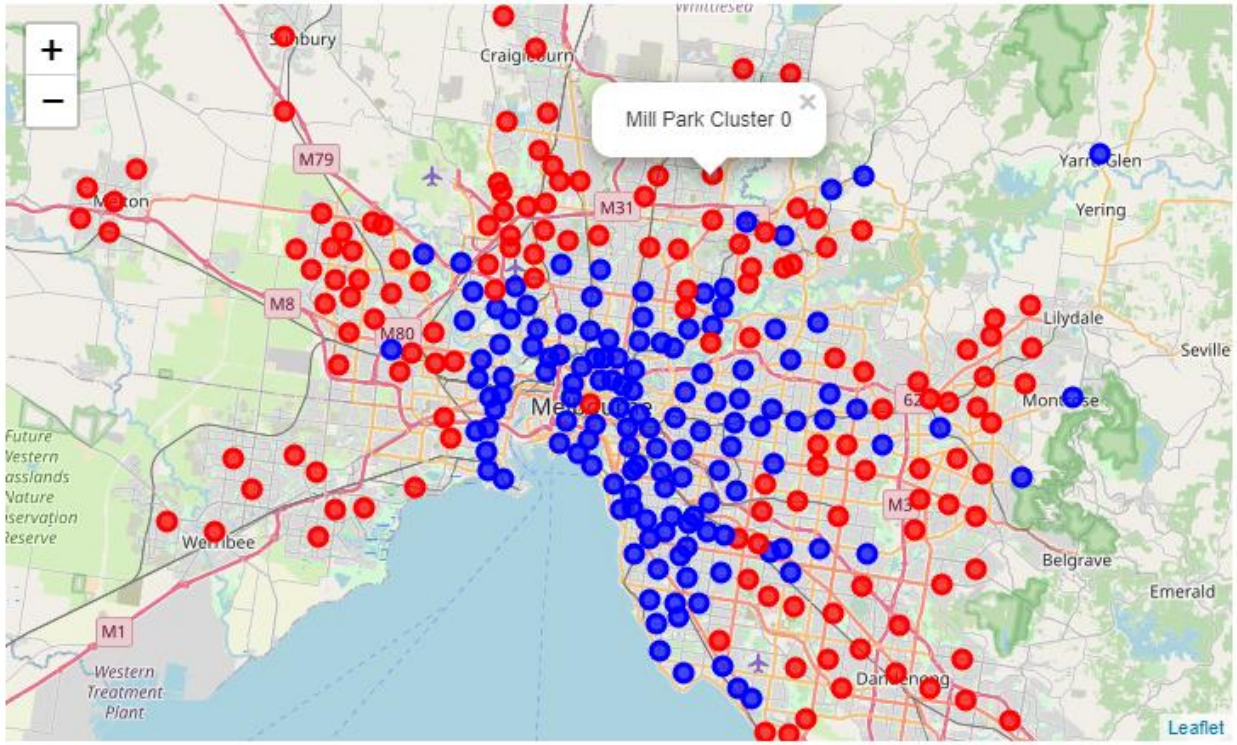


The geolocator was used to get the geographical coordinates of Melbourne. And, the python folium library was used to generate map of Melbourne and superimpose different suburbs using the latitude and longitude values as below.

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.
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.



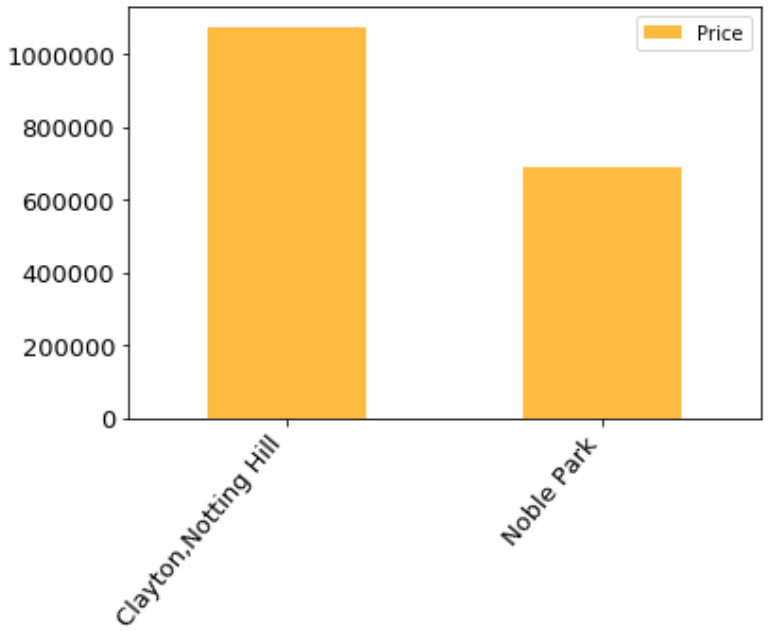
Results

It can be observed that the given two different post codes indeed belong to two different clusters, therefore they have significantly different housing price.

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. a house in Noble park seems to be better if you are considering a cheaper price with nearby amenities.

	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



Conclusion

- 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.