1. Business Problem

In this project we will try to find an optimal location for investing in a rental property in Melbourne. This report is targeted for investors looking to invest in residential rental property in Melbourne at a suitable location which can generate a high rental yield.

Rental yield measures the profit you generated investment property as a percentage of its value. A good rental yield is dependent of two factors a) profit from the property i.e. earnings minus all expenditures and b) purchase price. In other words, rental yield will be high if earnings i.e. rentals are maximised with minimum investment.

A property can be rented through rental agencies for long term rentals and also short-term letting services like Airbnb. Short term rentals have risen in popularity in recent years and with over 20,000 active listings across Metropolitan Melbourne has experienced one of Australia’s most robust growths in short-term rentals over the past two years. Hence, we will be looking at Airbnb listing data as an indicator of rental yield prospect.

Melbourne is a very big city with lot of property investment options. Property prices vary quite a bit in different parts of the city depending on various factors. Airbnb listings also varies greatly across the city along with rental tariffs or rates. Finally, number of venues like cafes, restaurants, pubs, shops and other places of interest, tourist locations also vary across the city. A high count of venues in a neighbourhood would imply popularity of that neighbourhood with renters.

Hence, to achieve a best rental yield, we need to look at areas with -

* relatively high Airbnb rents
* a high number of venues
* with not so high property prices
* less competition from other properties / listings in that area i.e. lower count of listings will indicate less competition.

We will use data science tools and methodology to identify a few most promising investment locations based on the above criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

1. Data Requirements

Based on the business problem, we have looked at below data –

* Current residential property prices in Melbourne metropolitan region from Residential sales [data](#ResSalData)
* Median daily rent of Airbnb [listing](#AirbnbData) in a particular area
* Number of Airbnb [listings](#AirbnbData) in a particular area
* Venues and amenities (restaurants, cafes, pubs) in and around the area from Foursquare [data](#FourSqrData).
  1. Residential sales data -

We have extracted Melbourne residential property prices from Australian real estate sales Sep’18-Jun’20 dataset as uploaded on Kaggle by HtAG Holdings.

<https://www.kaggle.com/htagholdings/aus-real-estate-sales-march-2019-to-april-2020>

Note

The residential sales data is available for suburbs while the Airbnb listing data is available for neighbourhoods or local government areas (LGAs). An LGA will have multiple suburbs. Hence, we will have to extract corresponding neighbourhoods for all suburbs in the residential sales data. This is available on Wikipedia at <https://en.wikipedia.org/wiki/List_of_Melbourne_suburbs>.

* 1. Airbnb listings

We have extracted Airbnb listings data from ‘Inside Airbnb’ dataset

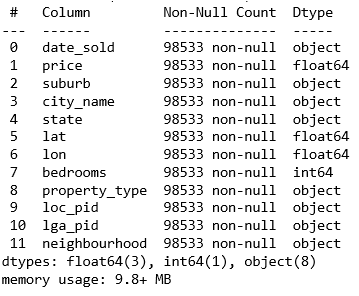
<http://insideairbnb.com/melbourne/?neighbourhood=&filterEntireHomes=false&filterHighlyAvailable=false&filterRecentReviews=false&filterMultiListings=false>

* 1. Foursquare data

Finally, we have extracted venues data from Four square. An area or a location with high number of venues will indicate a popular area which short term renters might prefer.

1. Methodology
   1. Extract and cleanse property sales data

After extracting the residential property sales data, we have filtered Melbourne data from this dataset. Below is a snapshot of fields in the data –



We have also cleansed the data by removing any row where any of the cells are Null or NAN. We have also removed outliers to get a correct perspective of the data. This is the property sales data after removing outliers



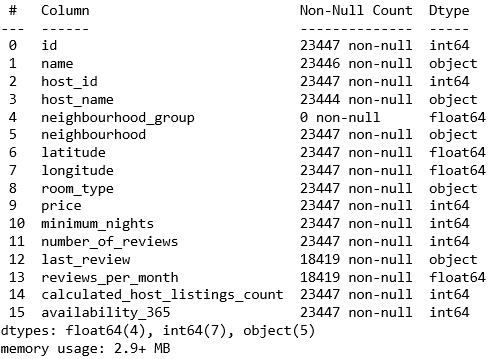
We have used price, latitude (lat), longitude(lon) and neighbourhood in our assessment. The rest of the fields are not required.

We will then find median price of properties for all neighbourhoods in Melbourne metropolitan.

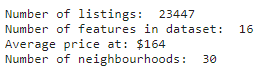


* 1. Extract and cleanse Airbnb listing data

Similar to property prices, we have extracted Airbnb listing data for all suburbs. Below is a snapshot of the Airbnb data –

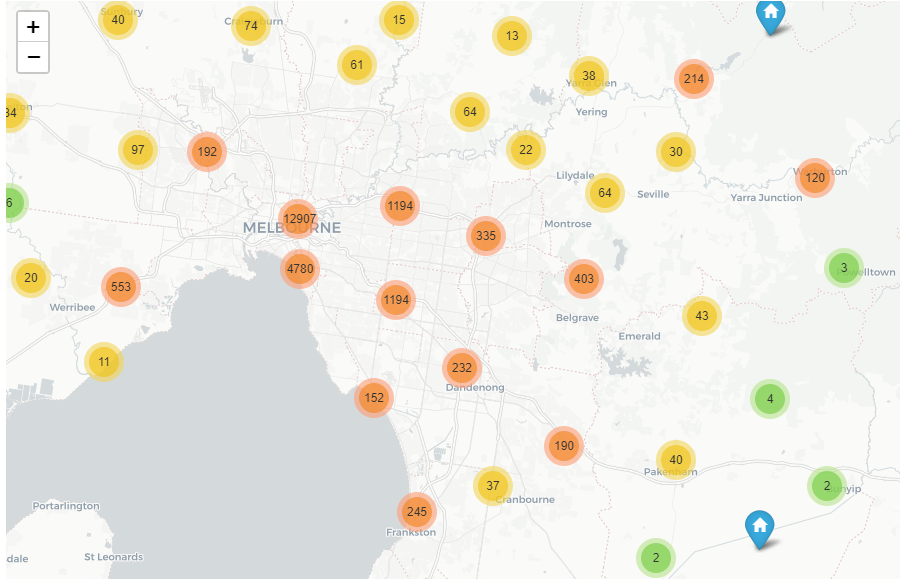


As with residential sales data, we have also cleansed the data by removing any row where any of the cells are Null or NAN. We have removed outliers to get a correct perspective of the data.



We have only considered price and neighbourhood in our assessment. The rest of the fields are not required.

This map shows us spread of listings throughout the Melbourne Metropolitan Region.



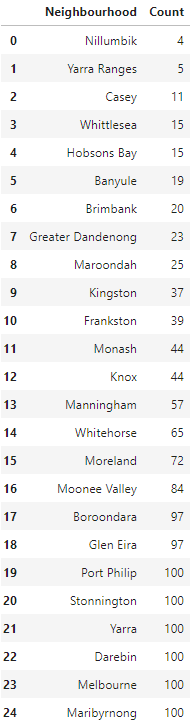
After data cleansing, analyse, average daily rentals across Melbourne neighbourhoods for different accommodation types and find out total listings per neighbourhood.



* 1. Extract venues data from Four Square API



Let’s find out total number of venues in each neighbourhood



* 1. Merge data of all datasets

Let’s merge the airbnb listing data, venues data and property data

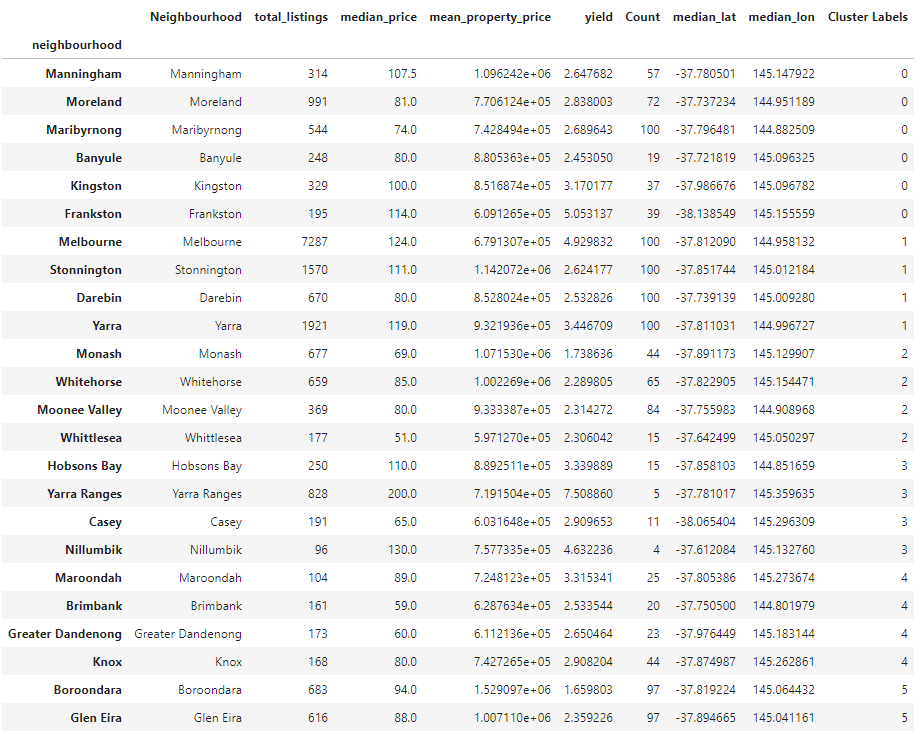


As we are looking at rental yield as one of the important criteria for investment property, lets calculate the gross yield



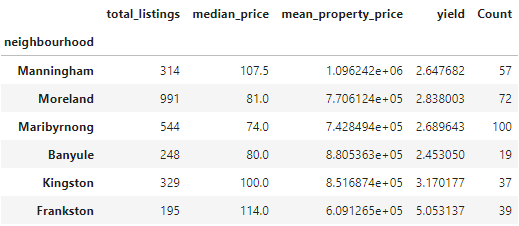
* 1. Cluster data

We will use Kmeans clustering to cluster data. We have considered 6 clusters which is the optimum k of the K-Means what derived from the elbow method

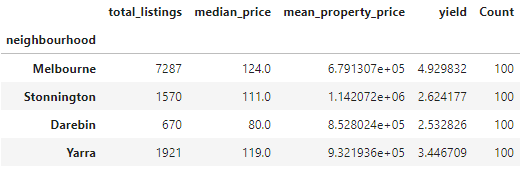


1. Results
   1. Individual cluster data

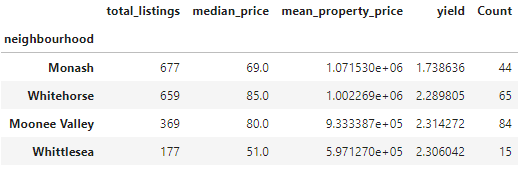
Now let’s take a closer look in the data for neighbourhoods in cluster 0



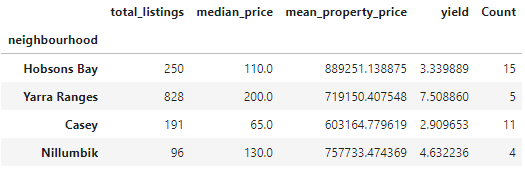
Now let’s take a closer look in the data for neighbourhoods in cluster 1



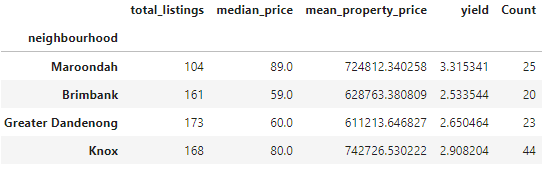
Now let’s take a closer look in the data for neighbourhoods in cluster 2



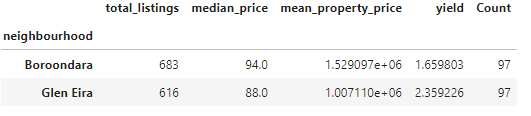
Now let’s take a closer look in the data for neighbourhoods in cluster 3



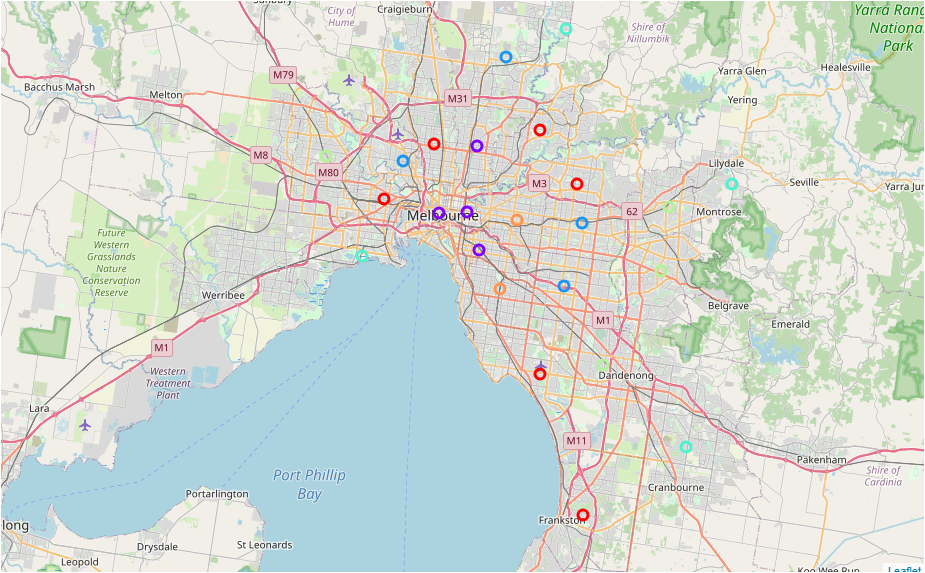
Now let’s take a closer look in the data for neighbourhoods in cluster 4



Now let’s take a closer look in the data for neighbourhoods in cluster 5



* 1. Cluster data map



1. Discussion
   1. Observations

Here is a rundown of the clusters –

Cluster 0: Most of the neighbourhoods in this cluster have medium levels of yield, competition, and mostly medium number of venues. Most of the neighbourhoods have medium property prices.

Cluster 1: All neighbourhoods have high count of venues. Most of the neighbourhoods have high competition. Though half of the neighbourhoods have high yields, (Melbourne and Yarra), property prices are close to high property prices for most of the neighbourhoods.

Cluster 2: Most of the neighbourhoods have high property prices and low yields.

Cluster 3: All neighbourhoods have high yield, medium property prices and medium competition. However, all these neighbourhoods have low count of nearby venues.

Cluster 4: Characteristics of this cluster is similar to Cluster 0 with all parameters except yield lower than cluster 0.

Cluster 5: This cluster has high property prices and low yields. However, there are high number of venues in the vicinity and medium competition.

* 1. Recommendations

As we have stated before, a solid investment property needs to have high yield with minimum competition and should be in a neighbourhood popular with renters. With this objective, Clusters 2 and 5 are not suitable. Though cluster 1 is a very popular area, competition is the highest with relatively high property prices which would make neighbourhoods in this cluster not so ideal for investment. As mentioned above, cluster 0 and 4 are mostly similar and considering property prices, cluster 4 is preferable over cluster 1. Except for the number of venues, Cluster 3 seems to have the best conditions for investment with highest yields with medium investment and medium competition.

Hence neighbourhoods in Cluster 3 and Cluster 4 are ideal investment opportunities. A choice between these two will depend on investors preference. Investors looking for best yields should look at neighbourhoods in Cluster 3 while investors looking for a balance of everything should look at Cluster 4.

1. Conclusion

This analysis will enable potential investors in the residential investment property to shortlist some neighbourhoods based on few key criteria. This is help investors not familiar with the Melbourne property market or the Melbourne Metropolitan Region in general. However, a property investor might have other considerations to make before investing in a property like budget, type of property etc. Hence it would be imperative to know an investor requirements to determine and recommend most appropriate neighbourhoods for an investment property in the Melbourne Metropolitan Region.