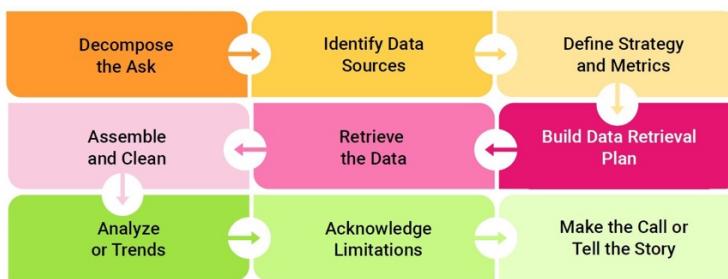


Project Description (The Ask):

A Lottery winner, in her 40s, has won \$1M and would like to invest it in her dream family home. She has 2 teenage children, a daughter and a son, who enjoy their own privacy and would love to have a place big enough to invite friends over. Both herself and her husband work in the Melbourne city and they only own 1 car. Considering that this \$1M will be her only substantial saving enough to make her dream home come true, she has decided to approach our property investment consultancy firm in order to make the most informed decision for herself and her family. The client has also expressed her preference that if she would be able to find the perfect home below \$1M, it would help her to upgrade her car and to buy a second one.

Analytical approach used to address client's inquiry



Decompose the Ask:

- Client has a budget of \$1 million to buy large family home
- Client's preference is to invest portion of her budget to upgrade her car and buy a second one hence ideally her budget is set to start from \$800k - \$1m
- Client's intention is to eventually own 2 cars and the ideal place must have 2 garages.
- Client has a family of 4, husband and 2 children, requiring 4 bedrooms (3 bedrooms and a guest room) and 2 bathrooms, in a free-standing home with a large enough household size, providing for comfort and entertainment.
- Client and her husband both work in the Melbourne city and short distance to CBD is preferred.
- Client currently owns only 1 car and close distance to public transport is preferred.
- Client children are teenagers and schools' locations nearby are of importance to her as well as proximity to local shops and shopping centers.
- Client is using her main and only capital for this property investment and the growth in the value of the ideal property and future return of this investment for her family is of importance. As such, to minimize the potential ongoing cost of maintenance and renovations, the properties built from 2015 onwards are to be considered.
- Client and her husband are in their 40s and her family is a young family in general, with both their children enjoying a balanced social life. Hence, prioritizing the lifestyle of their children, they are in pursuit of a younger suburb.

Identify Data sources:

- Australian Census
- Real state

Define strategy and metrics and build data retrieval plan:

- To find insightful real state associated datasets, large enough to draw trends from.
- To then define trends and correlations based on:
 - Suburbs, postcodes
 - Number of bedrooms, carparks and overall household/land size
 - Real state type and year built
 - Price and growth potential
- To use APIs and geodata and find distances from main areas of interest for the client

Retrieve the Data:

- Census dataset 2021- Australian Bureau of Statistics
- Melbourne Housing Market: Melbourne housing data and Melbourne house prices

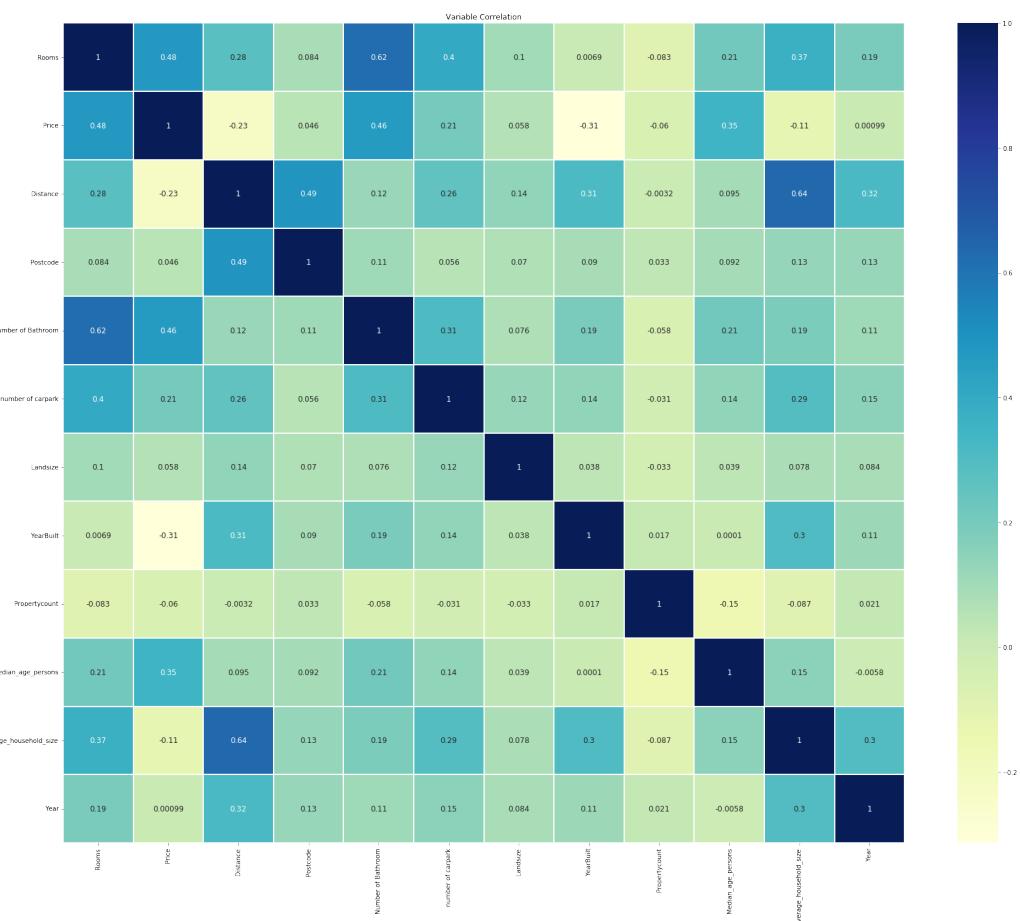
Assemble and clean:

- 'Melbourne housing dataset' is cleaned through:
 - Dropping 'NaN' missing values, which reduced total number of values under each column of the dataset to 8,887, originally varying from 13,742 to 34,857.
 - All 'float' type values were converted to 'integer' except for 'latitude and longitude' and 'postcode', which are remained as 'float' and changed to 'object', respectively.
- 'Census dataset 2021' is cleaned through:
 - In order to merge both datasets on 'postcode', this title was edited to match its identical from 'Melbourne housing dataset'.
- 'Melbourne housing dataset' and Census dataset 2021 are merged on 'postcode':
 - Columns titles are renamed in line with the key data required within the scope of the research.
- The new merged data set is sorted:
 - A new column for 'year', based on the date each property is listed in the original data set, is extracted and added to our new merged dataset.

- 8 main regions are identified with a total of 315 suburbs as: Northern Metropolitan', 'Western Metropolitan', 'Southern Metropolitan', 'Eastern Metropolitan', 'South-Eastern Metropolitan', 'Northern Victoria', 'Eastern Victoria', 'Western Victoria'.
- New data set is sorted based on 'Price' in descending order.

Analyse or Trends:

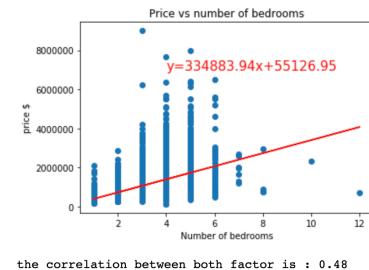
- The sorted new data set is filtered to include only properties from \$800k - \$1m and. Only free-standing homes, reducing the number of suburbs by 103 and to 212.
- For each of the main regions then data is filtered to include properties that:
 - Have 4 bedrooms, 2 carparks and 2 or more bathrooms,
 - Have a Land size of more than 600 square meters,
 - Are within 10km of the Melbourne city,
 - Are within postcodes with median age of residents from 30 to 40 years old,
 - Are built no earlier than 2015.
- The filtered data returned a total of 348 properties listed in 158 suburbs, with each region's breakdown, in descending order, as:
 - 'Northern Metropolitan' =99 properties, 29 suburbs
 - 'Western Metropolitan' =93 properties, 42 suburbs
 - 'Eastern Metropolitan' =68 properties, 35 suburbs
 - 'South-Eastern Metropolitan' =50 properties, 22 suburbs
 - 'Eastern Victoria' =10 properties, 6 suburbs
 - 'Southern Metropolitan' =9 properties, 7 suburbs
 - 'Northern Victoria' =9 properties, 7 suburbs
 - 'Western Victoria' =3 properties, 2 suburbs
- In terms of availability of properties matching our criteria, the 4 top main regions are,
 - 'Northern Metropolitan', 'Western Metropolitan', 'Eastern Metropolitan', 'South-Eastern Metropolitan'.
- A correlation heatmap for a correlation matrix, representing the correlations between different criteria considered, is generated, to provide more insight on relatedness of these filters on the price of the property:



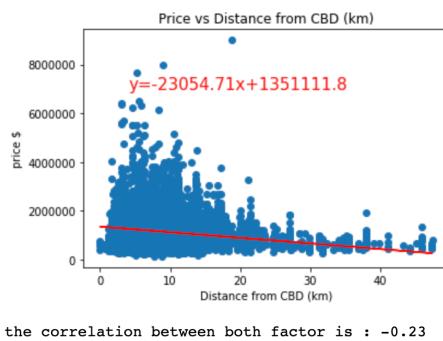
- Main observations from correlation heatmap are:
 - Strongest positive correlations between price and number of bedrooms, number of bathrooms, median age and number of car parks, in descending order.
 - Weakest positive correlations between price and land size.
 - Strongest negative correlations between price and year built, distance and household size, in descending order.
 - Strongest positive correlations between distance and household size, year built, number of rooms and carpark in descending order.

- To further confirm the observations above, below graphs are generated:

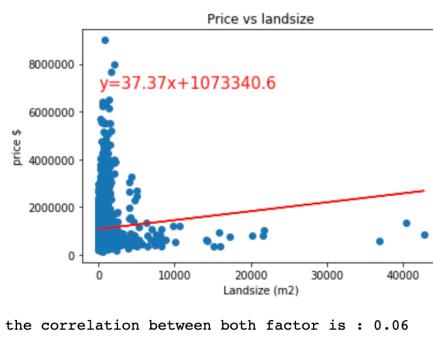
- Price vs. Number of bedrooms



- Price vs. Distance from the city (km)



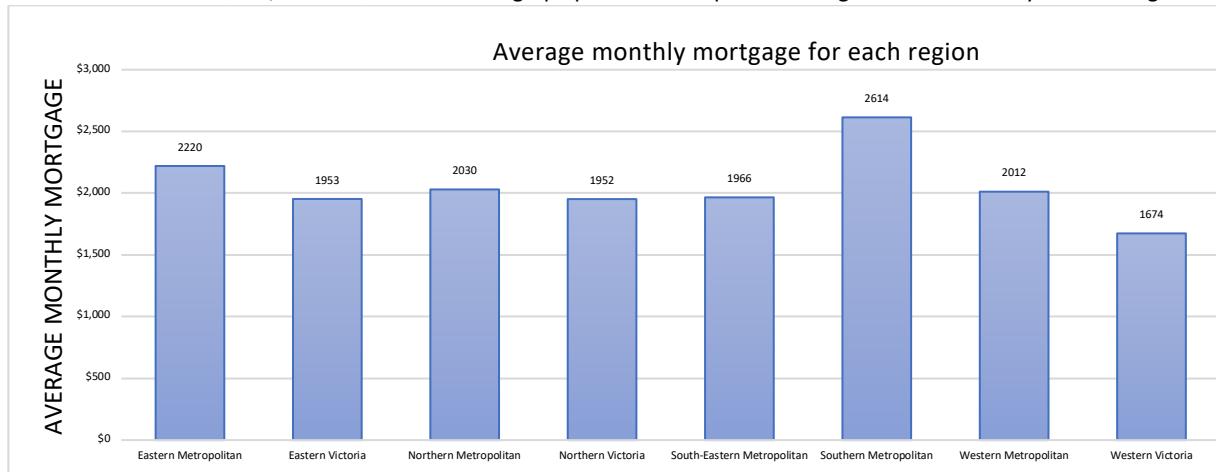
- Price vs. land size (m²)



- To better understand the affordability and spread of properties in different regions, meeting the budget of the client, below graphs are generated:

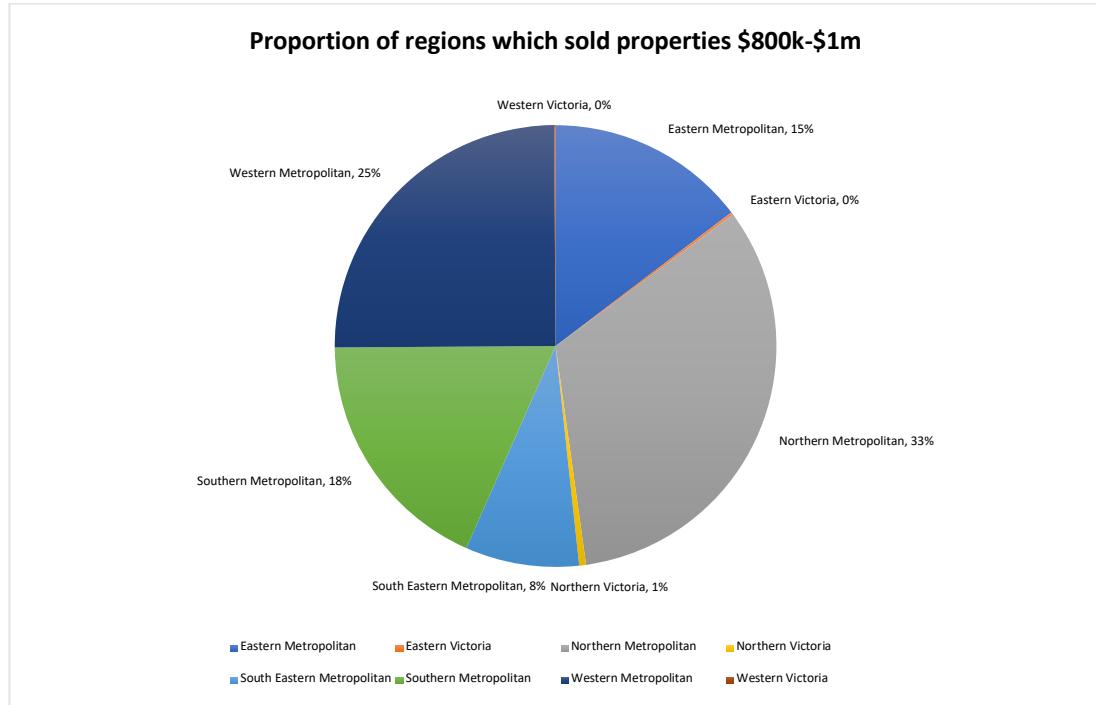
- Bar graph: average monthly mortgage for each Victorian region.

The client is between \$800,000 and \$1 million. This graph provides a snapshot of the general affordability for each region.

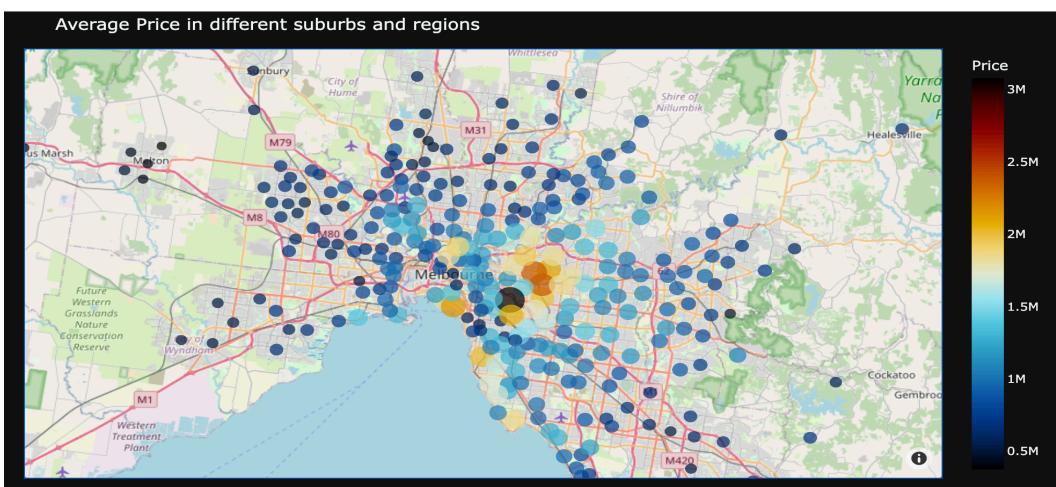


5. Pie chart of properties sold within budget for each Victorian region.

This graph provides a snapshot of which regions are more likely to have properties that fit within client's budget.



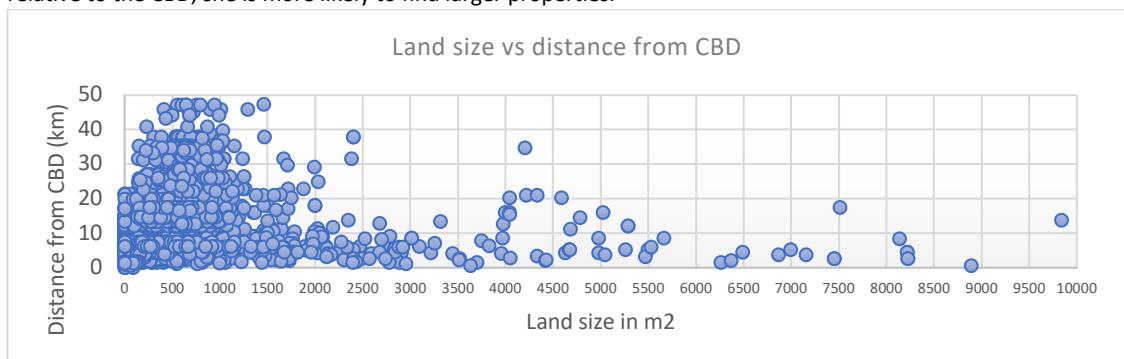
6. Average price in different suburbs and regions:



- In order to further investigate correlations between the distance criteria and the space associated criteria requested by the client, below graphs are generated:

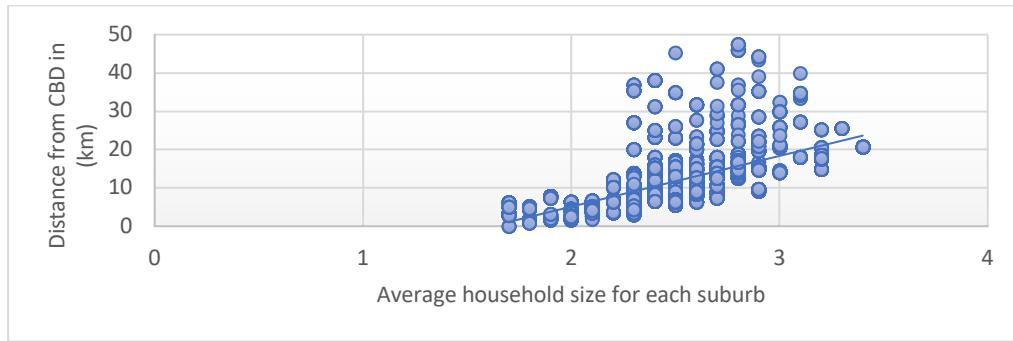
7. Scatterplot: land size vs distance from CBD.

The client wants to find a spacious family home. This graph is relevant for her to understand where, in terms of location relative to the CBD, she is more likely to find larger properties.



8. Scatterplot: Average household size vs distance from CBD.

This is plotted to show whether there is a correlation between how far people chose to live from the CBD, and the size of families.



9. Bar graph: median number of bedrooms for each Victorian region

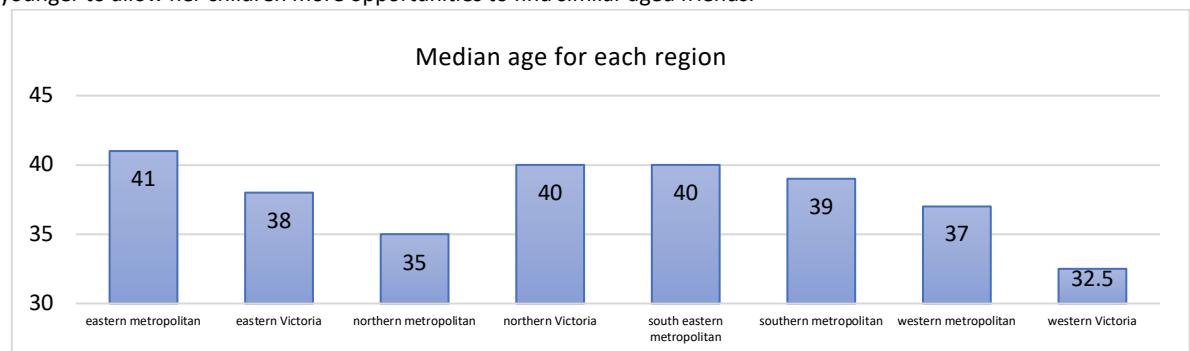
The client is interested in finding a spacious house for her family of four and to entertain their friends. This graph is relevant for her to identify regions where she is more likely to find more spacious houses, with more bedrooms.



- To further investigate the spread of median age of residents in different regions the graph below was generated:

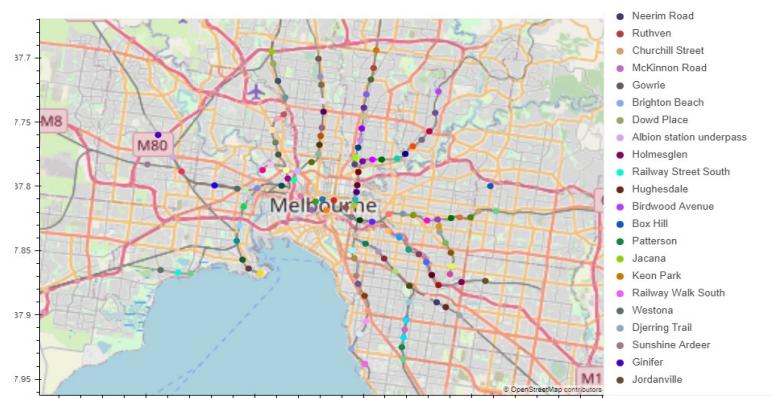
10. Bar graph: Median age for each Victorian region.

This was plotted to show the median age for each region. The client can use this graph to choose to live in regions where she would find more middle-aged individuals like herself, or she may choose to live in regions where the median age is younger to allow her children more opportunities to find similar aged friends.

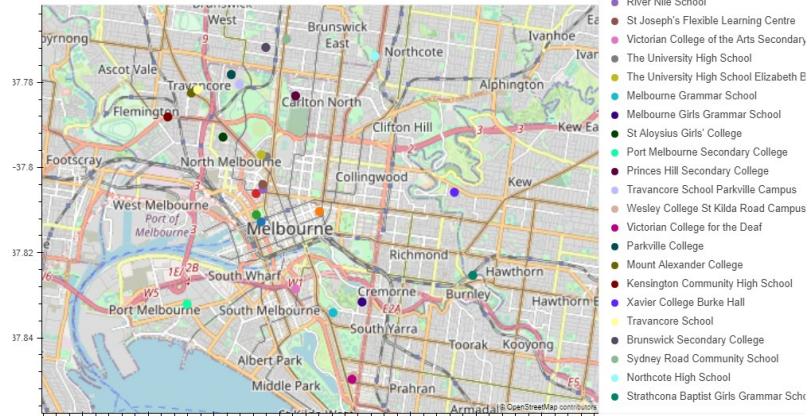


- In order to better understand the proximity of properties of interest to public transports, high schools and shopping centres, below maps are generated. This is done, through use of 'latitude and longitude' values from our data set with 'geopify' API, to search for these locations of interest and their distances from the city within a radius of 15km.

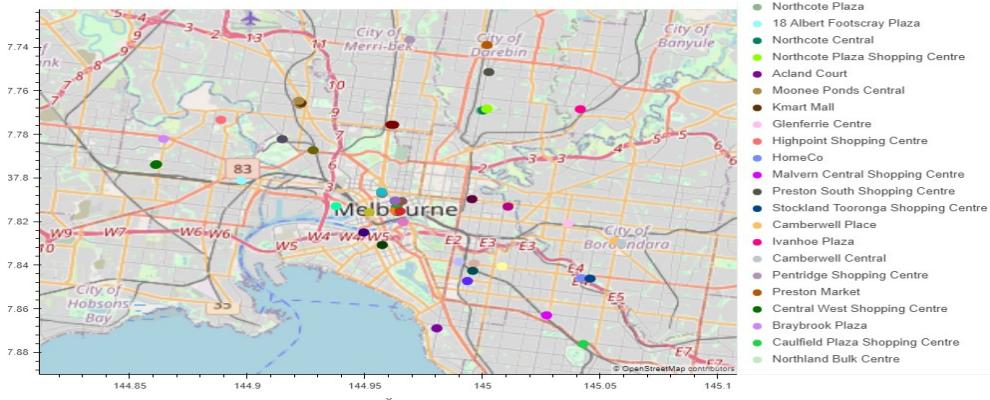
11. Map of the locations of train station.



12. Map of the locations of high schools.

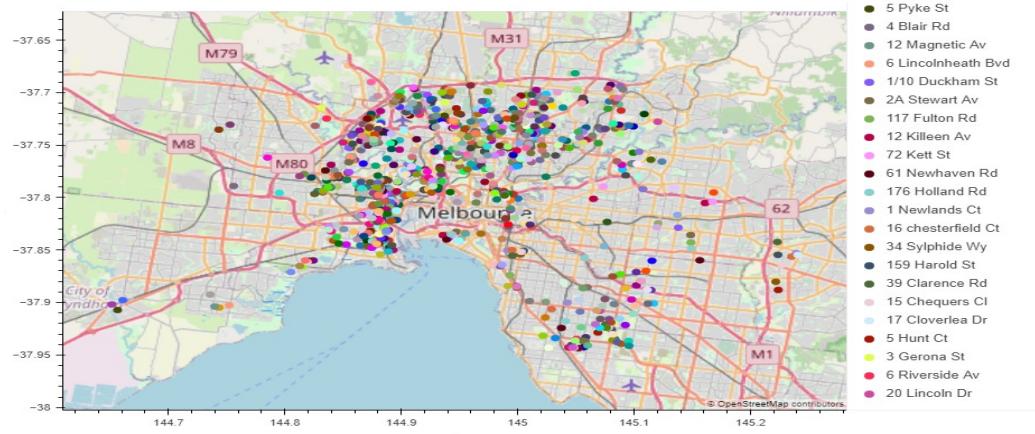


13. Map of the locations of some of major shopping centres



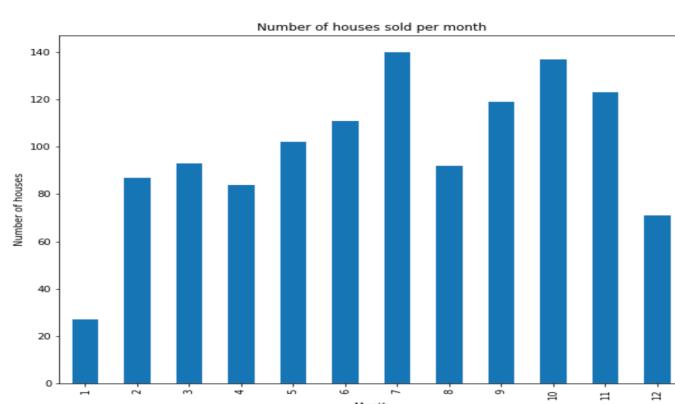
- In order to better understand the growth potential of properties the of interest below map is generated.

14. Map of major number of house sales



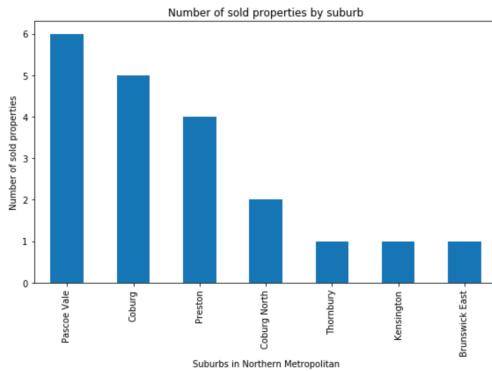
- To further investigate the growth in the value of the ideal property and future return of this investment for her family, the graphs below are generated:

15. Number of houses sold per month, to indicate the right timing to invest in buying a house so maximum saving could be done, given that summertime prices tend to rise it seems like demand is lower, however following June and second half of the year in winter when it is also the end of financial year it seems the selling and buying market is stronger.

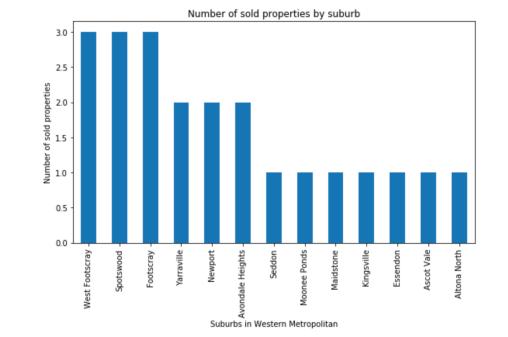


- To investigate the top two regions in terms of sales and understand which suburbs have sold the highest number of properties the two graphs below are generated:

16. Number of properties sold by suburbs in 'Northern Metropolitan'



17. Number of properties sold by suburbs in 'Western Metropolitan'



Acknowledge limitations:

- Datasets are from 2016-2017 for housing and 2021 for census, both of which would not represent the current trends but for sake of our project they were deemed sufficient.
- The growth for value of the property is dependent on many factors that cannot be considered within the scope of this project, hence long-term return cannot be precisely guaranteed.
- The ongoing construction projects within suburbs can increase or decrease the value of properties, which cannot be taken into considerations.
- Public transports considered are only trains.
- Other factors that are not considered for finding an ideal family home, could have been the rate of crimes and accidents to account for liveability of the suburbs considered and safety of the client's family.
- Some better data sources with larger and more up to date information were available, however not used as they were not free.

Major Findings and Conclusion:

In order to assist the client with making an informed decision based on her criteria for her ideal family home, we investigated free-standing homes with 4 bedrooms, 2 car parks, +2 bathrooms and with a land size of over 600 square meters. Moreover, only homes built in and after 2015 were considered, with close proximity of 10km or less to city, schools and shops in regions where the median age of residents is from 30 to 40 years old.

Based on our correlation heat map analyses as expected price showed a positive correlation and increase with number of bedrooms and an inverse relationship with the distance from the city. For our homes of interest, the data for 4-bedroom houses within 10km of the city still is well within the price range of up to \$800k. However, the positive correlation between the land size and the price is so weak and given that most of the homes closest to the city have the highest price and smallest land size, their stronger association with proximity to the city, takes over the effect of land size on their price and hence majority of homes with land sizes of 1000 square meter or smaller fall between \$800k-\$1m, meeting and exceeding our criteria. This is further confirmed using graph 7, showing how the further we get from the city the land sizes increases and how the concentration of smaller land sizes expectedly is higher within 20km distance from the city. The graph 8 further shows how the household size increases as we move away from the city, yet within 10km from the city it still fits our criteria.

Given above considerations and following the application of our criteria to our data set, the top 4 regions of Melbourne with highest number of houses spread over highest number of suburbs, matching our criteria were found to be:

- 'Northern Metropolitan' =99 properties, 29 suburbs
- 'Western Metropolitan' =93 properties, 42 suburbs
- 'Eastern Metropolitan' =68 properties, 35 suburbs
- 'South-Eastern Metropolitan' =50 properties, 22 suburbs

Further to this through investigating the median no of bedrooms for houses sold in each region, graph 9 indicates western and eastern metropolitan to have sold the highest number of homes with 4 bedrooms, making the two regions more plausible choices.

Our graph of average monthly mortgage across the top 4 regions, indicates similar affordability between Northern and western metropolitan and ranks eastern metropolitan the least and south-eastern metropolitan the most affordable regions amongst the 4. Additionally, the graph 6 also confirms eastern suburbs to account for highest price ranges and some anomalies. Given this analysis, western suburbs seem preferable for our budget comparing to eastern suburbs. Moreover, our pie chart of properties sold at each regions and within \$800k-\$1m, shows, northern followed by western metropolitan as the strongest markets accounting for 58% of total sales, which again singles out western suburbs as a more preferred choice, where there's a healthy market for sellers and buyers. This is further confirmed by map 14. We can also see from graph 10 that the median range for western region is 37 as opposed to the eastern metropolitan where it is 41, which puts it well within the age restriction preferred by the client.

From the maps 11 it could be seen that there are independent train lines with over 30 train stations spread around the west side of the city making it easily accessible. Map 12 suggest that northern suburb has better spread go high schools, however the map doesn't cover the western side, being set within 10km radius. Yet the travel distance and public transport between the west and north is like west to the city making it still considerable to reside in west. Map 13 also shows a good spread of major shopping centres around west with expected higher concentration at city.

Considering meeting most criteria set by the client Western metropolitan seems the most promising region to investigate and from our plot 17 the 6 suburbs of Avondale heights, Newport, Yarraville, Footscray, west Footscray and spots wood show highest number of sales making them better investment for when the client may like to sell and also potentially set at better price due to higher availabilities. Lastly from graph 15 it could be said that the right timing to invest in buying a house so maximum saving could be done, given that summertime prices tend to rise it seems like demand is lower, however following June and second half of the year in winter when it is also the end of financial year it seems the selling and buying market is stronger.