**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**

Timeline

Description automatically generated

**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:**

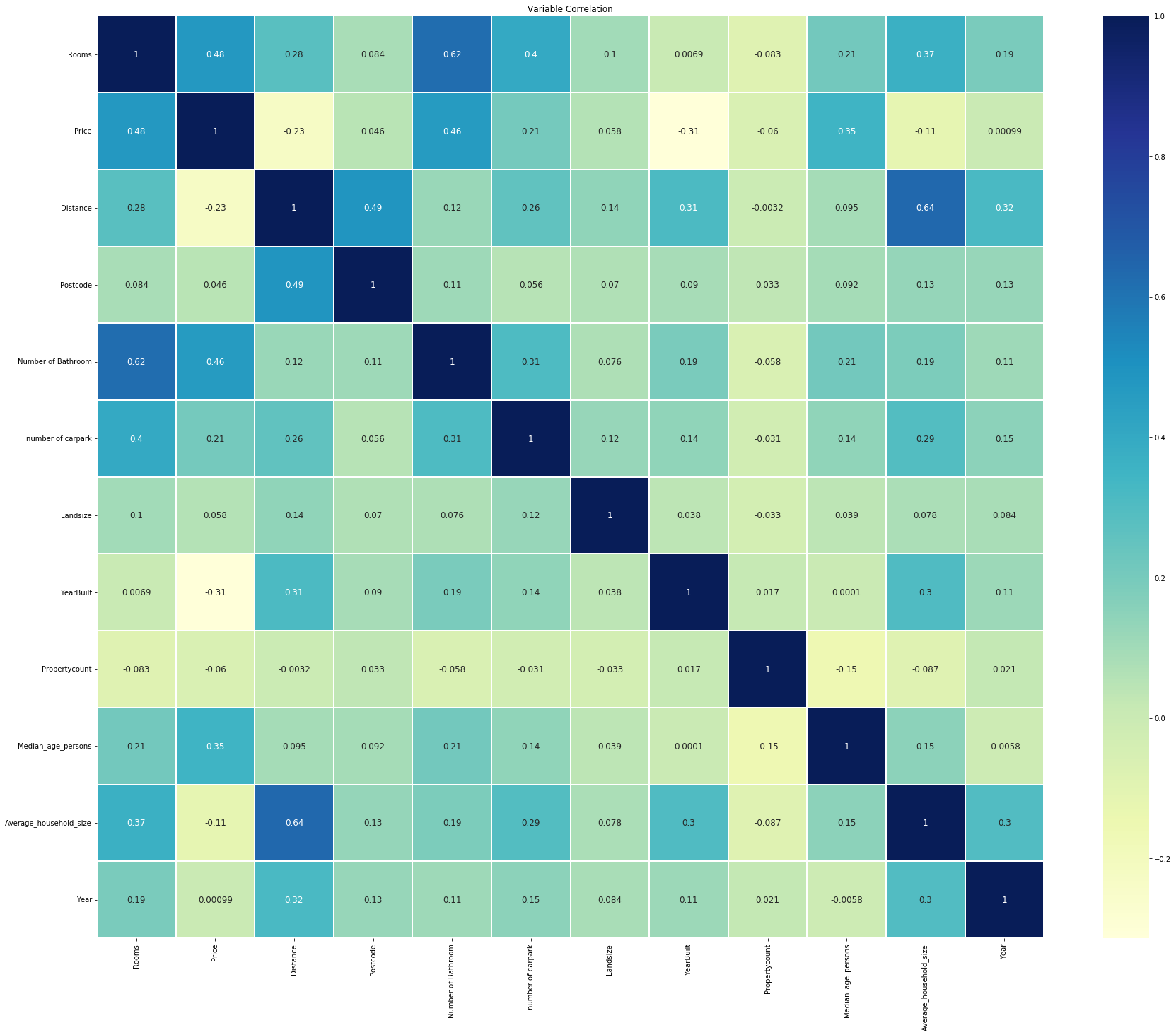
* Census dataset 2021- Australian Bureau of Statistics
* Melbourne Housing Market: Melbourne housing data and Melbourne house prices

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

1. 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.

1. 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.

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

1. 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.

1. 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.

1. 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:

1. 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 ‘geoapify’ API, to search for these locations of interest and their distances from the city within a radius of 15km.

1. Map of the locations of train station.

Chart, map

Description automatically generated

1. Map of the locations of high schools.

Chart, map

Description automatically generated

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

Chart, map

Description automatically generated

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

1. Map of major number of house sales

Chart

Description automatically generated

**Acknowledge limitations:**

**Major Findings and Conclusion:**