**MELBOURNE HOUSING DATA ANALYSIS**

## 1.Business Understanding

### 1.1 Business Overview

Melbourne, the world’s most liveable city as ranked by the Economist, is also one of the most expensive cities to buy a house. There has been a rapid rise in housing prices. By March 2021 house prices increased by $45000, which represented a 7.3 percent annual gain. In the Mornington Peninsula, one of the council areas in Melbourne, property values were up 16.6% year on year. In comparison to houses, units continue to underperform.

Once potential customers realize that housing prices are on the rise, they enter the market, further driving up demand followed by limited supply. The Australian Bureau of Statistics data shows home-ownership rates for people aged under 40 are declining, part of a trend of intergenerational inequality and a growing gap between the haves and have nots.

The rising cost of owning properties can be interpreted as a failure of the housing system to provide the right quantity of housing with the right characteristics and in the right locations over time. With the median price of buying a house now close to 1 million dollars, there is a need for new real estate investment to meet the demand. A comprehensive analysis of the housing market sales data is therefore essential in order to determine the types of properties in high demand which will form a basis for future real estate investments.

### 1.2 Business Objective

This research seeks to aid the housing crisis in Melbourne by analyzing the real estate sales data to determine the property characteristics of the most sold properties in Melbourne and deploying the results of the analysis to guide future housing investments.

Following this approach, the main objectives of this analysis will be:

* To determine the most popular types of properties sold.
* To determine the most popular location among buyers.
* To analyze the relationship between property features and price.

From the objectives above, this analysis will seek to answer the following questions:

1. What were the most popular types of properties sold?
2. Which were the most popular locations among buyers?
3. How do property features, such as number of bedrooms and land size, affect the price of a property?

### 1.3 Business Success Criteria

Provide insights on the most popular types of properties bought based on location, amenities and price that will guide Melbourne’s future real estate investments.

### 1.4 Assessing the Situation

#### A.Resources

i.)Personnel (Technical support, Data mining experts)

ii.)Project Datasets

* Melbourne\_housing\_Full.csv

iii.)Computing Resources

iv.)Software (Colaboratory, Github, JIRA,Google Suite)

#### B. Assumptions

i.) The data sample is an accurate representation of the entire dataset.

#### C .Constraints

There may be bias in our sampled data.

### 1.5 Data Mining Goals

Our data mining goal for this project is to determine the purchase trend of customers and the most popular suburbs.

Potential questions for consideration include:

1. What were the most popular types of properties sold?

* Popularity by Rooms
* Popularity by bedrooms and bathrooms
* Popularity by number of parking slots

1. Which were the most popular locations among buyers?

* Location by Region
* Location by Council Area
* Location by Suburb

1. How do property features, such as number of bedrooms and land size, affect the price of a property?

* What is the correlation between property features and price.

1. Who were the best selling agents
2. Which sale method was most preferred?
3. How were the sales distributed over time?
4. What was the change in price over time?

### 1.6 Project Plan

The **Cross-Industry Standard Process for Data Mining(CRISP-DM)** will be used as a guideline for conducting this research. Below is an overview plan for this study.

|  |  |  |  |
| --- | --- | --- | --- |
| **Phase** | **Time** | **Resources** | **Assignee** |
| Business Understanding | 1hr | Project Datasets/Data Scientist | Asha Deen |
| Data Understanding | 2hrs | Project Datasets/Data Scientist | Joy Machuka  David Oloo |
| Data Preparation | 2hrs | Project Datasets/Data Scientist | Vanessa Ng’eno |
| Data Analysis | 2hrs | Project Datasets/Data Scientist | Maureen Gatu |
| Evaluation | 1hr | Project Datasets/Data Scientist | Joy Machuka |

## 2. **DATA UNDERSTANDING**

### 2.1 Data Understanding Overview.

We have only one dataset that comprises data collected from the sales made in Melbourne by a project management company.

The Dataset below was provided for this research

1. [Melbourne\_housing \_Full](https://drive.google.com/file/d/1fb8hAcYA4p9wHKWjK_pz23FHx1awceJJ/view?usp=sharing)

2.2 Initial data collection

Data collected was sourced from the project management company’s database over one year and five months(May 2016 to March 2018)

2.3 Describing and Exploring Data

The data contained herein include the following columns

1. Suburb: Name of Suburb
2. Address: Address of the property sold
3. Rooms: Number of rooms in a house.
4. Price: Price in Australian dollars
5. Method:
   1. S - property sold;
   2. SP - property sold prior;
   3. PI - property passed in;
   4. PN - sold prior not disclosed;
   5. SN - sold not disclosed;
   6. NB - no bid;
   7. VB - vendor bid;
   8. W - withdrawn prior to auction;
   9. SA - sold after auction;
   10. SS - sold after auction price not disclosed.
   11. N/A - price or highest bid not available.
6. Type:
   1. br - bedroom(s);
   2. h - house,cottage,villa, semi,terrace;
   3. u - unit, duplex;
   4. t - townhouse;
   5. SellerG: Real Estate Agent
7. Date: Date sold
8. Distance: Distance from CBD in Kilometres
9. Regionname: General Region (West, North West, North, North east …etc)
10. Propertycount: Number of properties that exist in the suburb.
11. Bedroom2 : Number of bedrooms on a house
12. Bathroom: Number of Bathrooms
13. Car: Number of carspots
14. Land Size: Land Size in Metres
15. BuildingArea: Building Size in Metres
16. YearBuilt: Year the house was built
17. CouncilArea: Governing council for the area
18. Latitude: Latitudinal coordinates
19. Longitude: Longitudinal coordinates

### 2.4 Data description and exploration

This analysis utilized only one dataset and a full description of the dataset is as follows:

* [**Melbourne\_housing\_Full.csv**](https://drive.google.com/file/d/1fb8hAcYA4p9wHKWjK_pz23FHx1awceJJ/view?usp=sharing)

The dataset contains 21 columns and over 34,000 rows. It contains Melbourne housing prices information. The datasets outline the location of different categories of houses (3 / 4bedrooms etc) located across various suburbs in Melbourne. Further to the description, we have columns containing information about the Residential address, type of unit, price of a unit, Seller details, postal code, and council area.

### . 2.5. Verifying Data quality

Our dataset is complete and does not contain any errors. All of the columns included are necessary for our course. Most of the dataset does not contain missing values except for the year build column that several rows weren’t captured for. With the year column missing, there are minimal conclusions to be made around the column as it is not accurately distributed.

## **3. DATA PREPARATION**

Steps taken during data exploration are as follows:

### 3.1 Selecting Data

The following dataset was used for analysis in this project based on the relevance of our goals and data quality

* [Melbourne\_housing\_Full.csv](https://drive.google.com/file/d/1fb8hAcYA4p9wHKWjK_pz23FHx1awceJJ/view?usp=sharing)

We used data frames to load data from a file, examine basic statistics of the data, change and replace some values and finally output the results.

### 3.2 Cleaning Data

Data cleaning procedures performed during the cleaning process include:

* We dropped columns like ‘BuildingArea’ and ‘YearBuilt’ that seemed irrelevant for analysis and remained with the ones that will give us a good results
* We replaced null values in the ‘Price’ and ‘land size’ columns with the median price. This is the middle price of property in ascending order.
* We replaced all values in the column ‘Type’ to descriptive names: t to townhouse, h to house and u to unit
* We changed Date column data type to datetime
* We changed data type for car,bedroom,bathroom,postcode and property count to int datatype
* We used Describe to look for availability of outliers and used boxplot to show the outliers. Finally, we removed the outliers found.

## **4.ANALYSIS**

During our analysis, the following questions were answered.

1. What is the most popular type of property sold?

house 19548

unit 6713

townhouse 3413

1. What are the most popular suburbs among buyers?

Reservoir 829

Bentleigh East 571

Richmond 498

Preston 473

Brunswick 421

1. What are the least popular suburbs among buyers?

croydon 1

viewbank 1

Bulla 1

Plenty 1

Fawkner Lot 1

1. What are the most popular Council Areas among buyers?

Darebin City Council 2730

Boroondara City Council 2555

Moreland City Council 2028

Glen Eira City Council 1837

Banyule City Council 1710

1. What are the least popular Council Areas among buyers?

Knox City Council 327

Greater Dandenong City Council 306

Melton City Council 128

Nillumbik Shire Council 40

1. What are the most popular Regions among buyers?

Southern Metropolitan 9428

Northern Metropolitan 8892

Western Metropolitan 6245

Eastern Metropolitan 3916

South-Eastern Metropolitan 1156

1. What is the least popular Region among buyers?

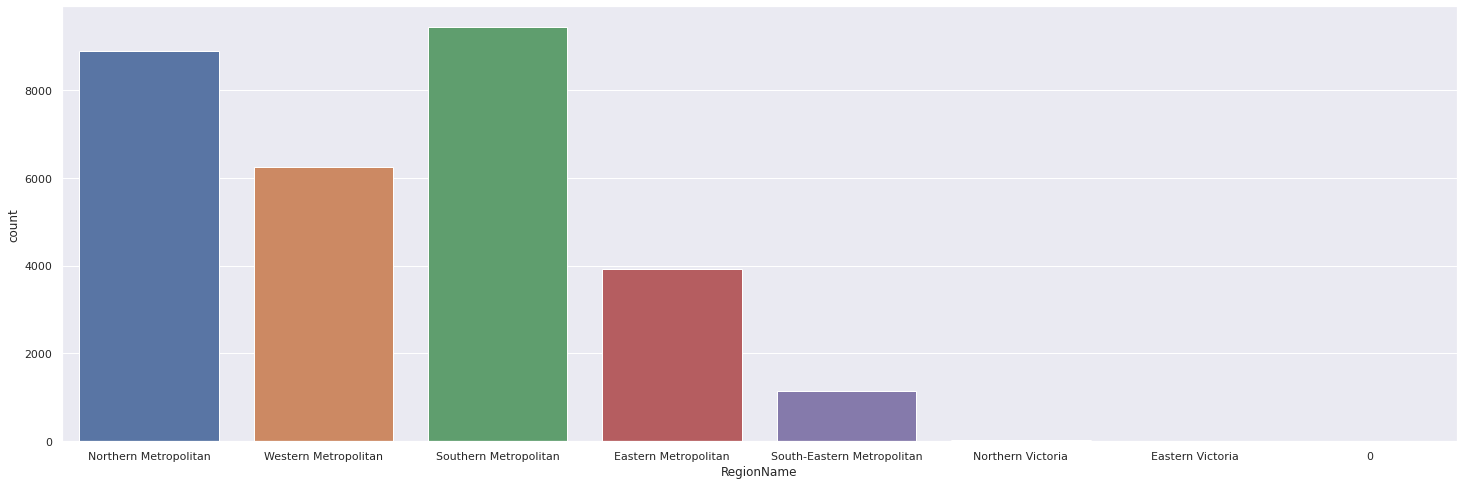
Eastern Metropolitan 3916

South-Eastern Metropolitan 1156

Northern Victoria 32

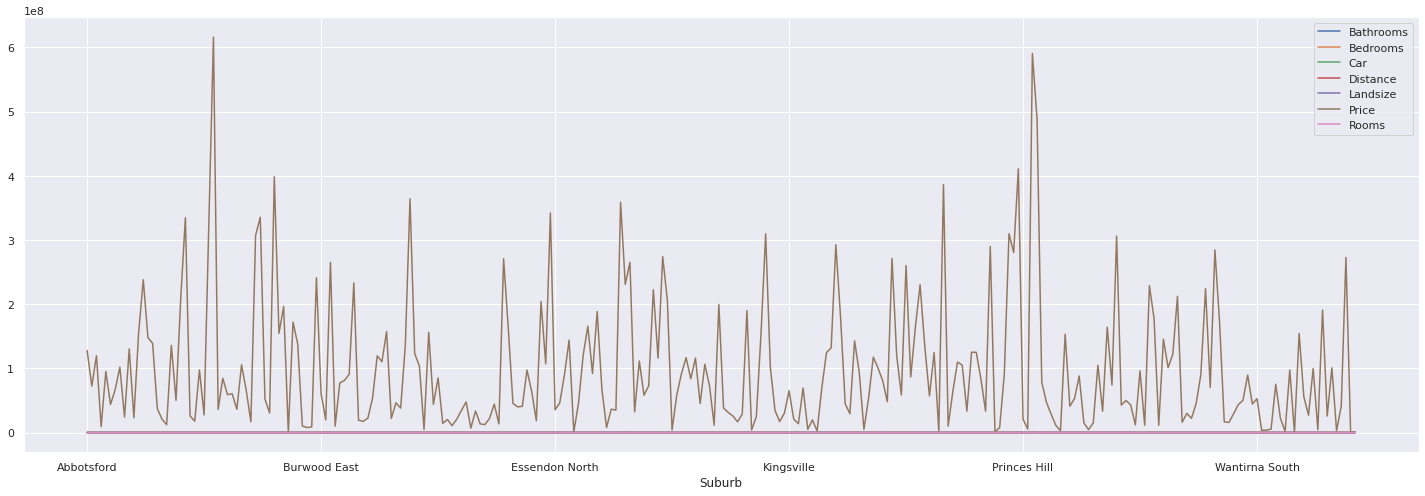
Eastern Victoria 2

The below graph is a representation of the least popular regions among buyers.



1. Is there an influence of distance from CBD on the price of property?

Yes, there is an influence but it is minimal compared to other factors that we have analyzed. This has been done by getting the pivot table and checking correlation between factors.



1. What are the most popular rooms?

3 13251

2 7797

4 6048

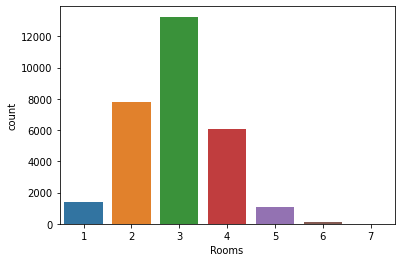
1 1365

5 1097

6 113

7 3

Below is a graphical representation:



1. What is the most popular number of bedrooms?

Three-bedroom houses are the most popular with a total count of 10,268 sold over the two years(2016-2018)

1. What is the least popular number of bedrooms?

Six-bedroom houses with a total sales count of 86

1. What is the most popular number of bathrooms?

Houses with one bathroom are the most popular.

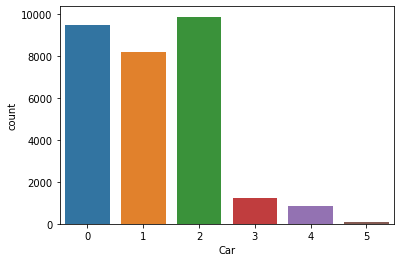
13.What is the most popular number of parking spaces?

2 9870

0 9460

1 8176

Below is a representation of popularity by number of car parking spaces.



14.Who are the top-selling agents?

Nelson 3001

Barry 2976

Jellis 2751

hockingstuart 2218

Ray 1718

15.Who are the worst-selling agents?

Jim 1

Naison 1

Reed 1

Sutherland 1

hockingstuart/Barry 1

16.What is the most popular method of sale?

S 16845

SP 4460

PI 4077

VB 2465

SN 1217

PN 264

SA 184

W 131

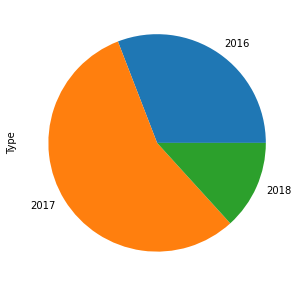
SS 31

17.What is the number of properties sold per year?

2017 16582

2016 9160

2018 3932



18. What is the number of property sold per month?

March 3758

October 3665

September 3237

November 3153

July 2904

August 2886

June 2828

May 2381

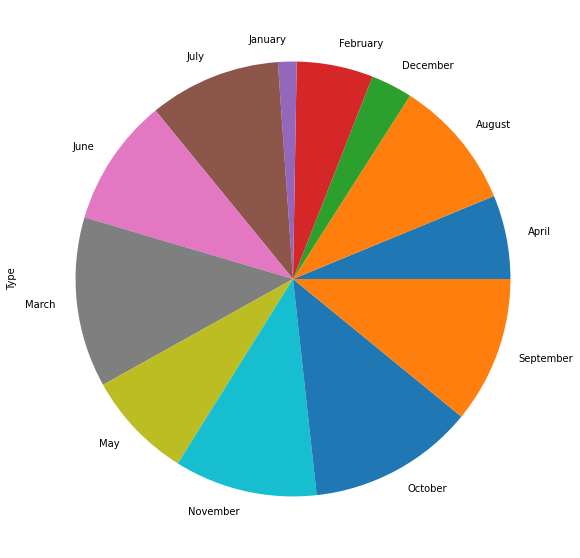
April 1849

February 1687

December 916

January 410

Below, is a representation of the property sold per month.



## 4. RECOMMENDATION

Following our analysis, we provide the following recommendations:

The most popular types of properties sold in the two years of reference are houses followed by units then lastly townhouses. Houses sold were 65.94% of the total sales, Units sold were 22.64% and Townhouses were 11.51%. We recommend investors to invest heavily in houses since they’re 43.3% more preferred than units and 54.43% more popular than Townhouses.

The most popular regions among buyers were Southern Metropolitan, followed closely by Northern Metropolitan. The most popular Council Areas were Darebin City Council and Boroondara City Council while the most popular Suburbs are Reservoir and Bentleigh East. The least popular suburbs are Croydon, Viewbank, Bulla, Plenty, and Fawkner Lot which only sold one piece of property over the two-year period. We therefore recommend investment in the Southern Metropolitan Region, Darebin City Council, and Reservoir compared to other locations because of the high demand in those areas.

The analysis shows that there’s a high positive correlation between property features (such as distance from CBD, number of rooms, bathrooms, bedrooms, parking space, and land size) and the price of the property. The more amenities a property has the higher the price. From the analysis the most popular features in a property were;

* Three rooms
* Three bedrooms
* Two parking spaces
* One bathroom

We, therefore, recommend that future housing projects be of the above characteristics in order to meet the demand in the market.

## 5. Evaluation

From our business success criteria, we have been able to gain important insights to guide future real estate investors in the city of Melbourne. Our approach of identifying the market needs by answering key research questions amongst other questions has enabled gathering relevant information about the market. This information is relevant for future real estate growth to ease out the growth curve of prices in relation to the market need.

**Links to Associated Documents**

* GitHub[[here](https://github.com/MaureenGatu/Melbourne-Housing-Analysis.git)]
* Presentation[[here](https://docs.google.com/presentation/d/1N70rkY8gVBlUbus7PA0Ey_OeUj2Cb_a5XpiRorgdOUM/edit?usp=sharing)]
* Jira [[here](https://maureengatu.atlassian.net/jira/software/projects/MH/boards/5)]

**REFERENCES**

1. <https://www.smh.com.au/politics/federal/sydney-and-melbourne-property-prices-slow-after-record-breaking-boom-20210503-p57oa7.html>
2. <https://www.abc.net.au/news/2021-04-16/the-median-house-price-in-melbourne-has-just-passed-1-million/100073806>
3. <https://www.youtube.com/watch?v=HjaetwKvlQ8>
4. <https://www.invest.vic.gov.au/why-melbourne/a-worlds-livable-city>
5. <https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/rp0809/09rp21>
6. <https://www.abs.gov.au/statistics/people/housing/housing-occupancy-and-costs/2017-18>
7. <https://www.ahuri.edu.au/__data/assets/pdf_file/0021/61239/AHURI-Final-Report-328-Australian-home-ownership-past-reflections-future-directions.pdf>
8. <https://www.theage.com.au/national/victoria/melbourne-the-worlds-sixth-most-expensive-city-20150304-13uilq.html>