GROUP 5

PROJECT 1

# INFLUNCES ON PERTH HOUSING PRICES

Kavish Naran, Jesslyn Lengkong, Dilip Pallapothu

## Introduction

This project utilises a Perth housing dataset to analyse influences Perth housing prices from 1990 to 2020.

Below is a list of questions which were asked during the analysis:

House sales in Perth – Background information

- Number of houses sold per year.
- Best month and year of sales.
- Number of houses sold per suburb.
- Average house prices by suburbs

Influences on house prices - Analyses

- Is there a price trend between houses in the south vs north of the Swan River?
- Does the ranking of schools in the area affect house prices?
- Does the type of school (government vs. non-government) affect average house price?
- Does the distance from the city affect house prices?
- Does the proximity of schools and train stations affect house prices?
- Percentages of bedrooms and bathrooms?
- Does land area influence pricing?

## Data Set

The data set consists of data from 322 Perth suburbs from Kaggle. The data set contains the following parameters:

Address, suburb, price, bedrooms, bathrooms, garage, land area, floor area, build year, CBD distance, nearest station, nearest station distance, date sold, postcode, latitude, longitude, nearest school, nearest school distance and nearest school ranking.

The longitude and latitude were obtained from data.gov.au and school rankings were obtained from better education.

## Data Cleaning

During the data cleaning process, python was used to clean the data set and ensure it was ready for the data analysis process.

The following was applied to the dataset to modify it for analysis:

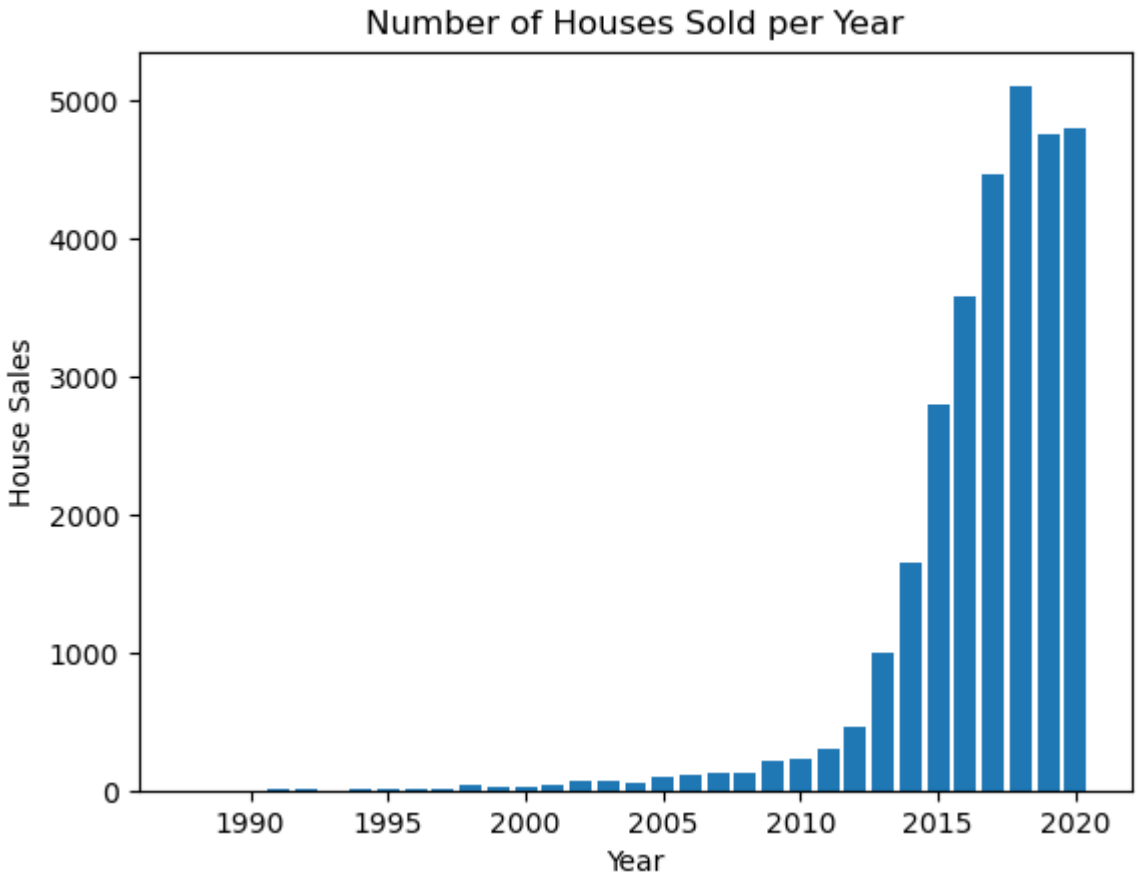
- Removal of null values
- Removal of columns not required in the analysis process, the following was removed from the dataset: address, garage, floor area and school ranking. School ranking was extracted at a later stage on its own.
- Distances in the data set were in kilometres and meters, therefore all distance related data was converted to meter for consistency. This would also align with the APIs (application progressing interface) distance of measurement. The data column was converted from a string variable to datetime to allow for graphing and analysis. Postcodes which were outside of Perth were removed, this involved removing postcodes greater than 6200.

- **Data Analysis and visualization**

House sales in Perth – Background information

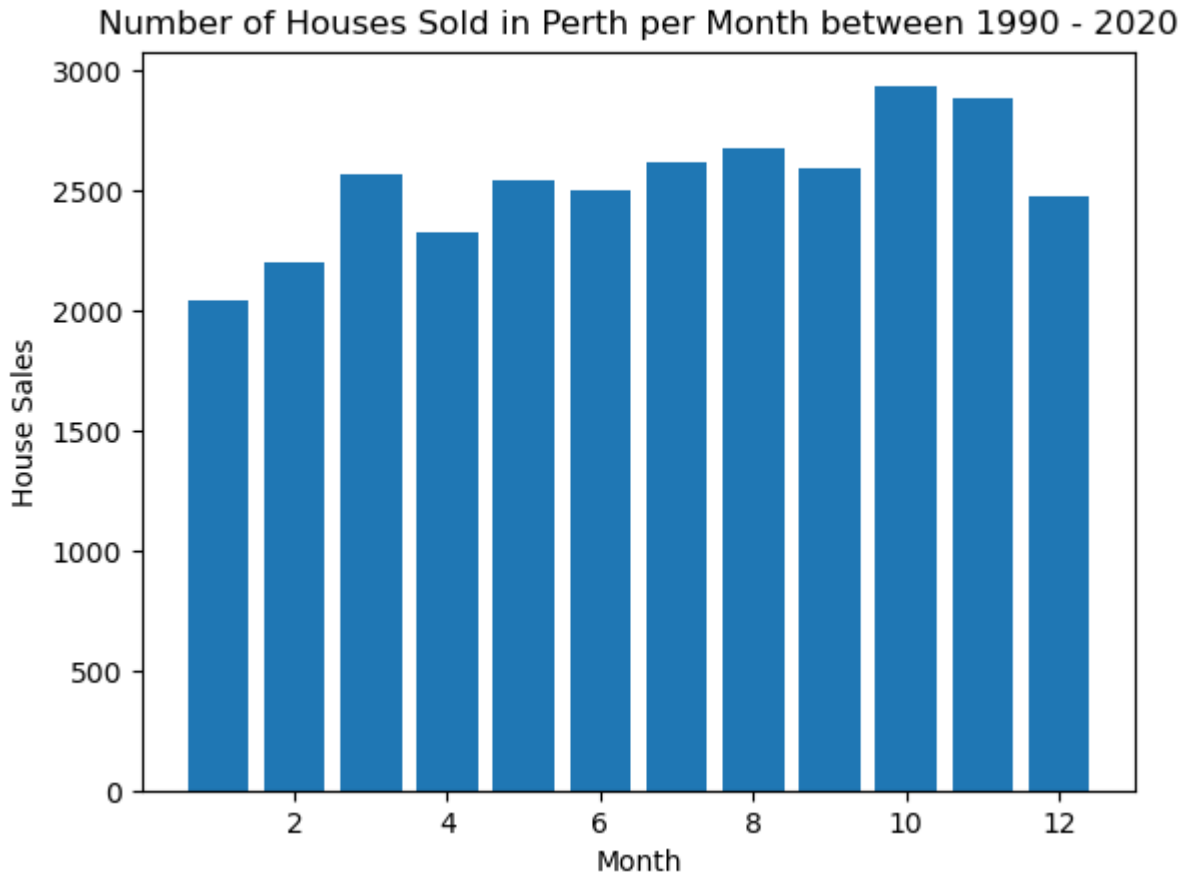
**Number of houses sold per year.**

Annual house sales in the Greater Perth metropolitan area were plotted in a bar graph to find any trends. From the dataset we used, there seems to be exponential growth in sales until 2018 and it slowed. This could be a limitation of the dataset. One of the standout points is that sales seem to be strong despite the covid lockdowns of 2020.



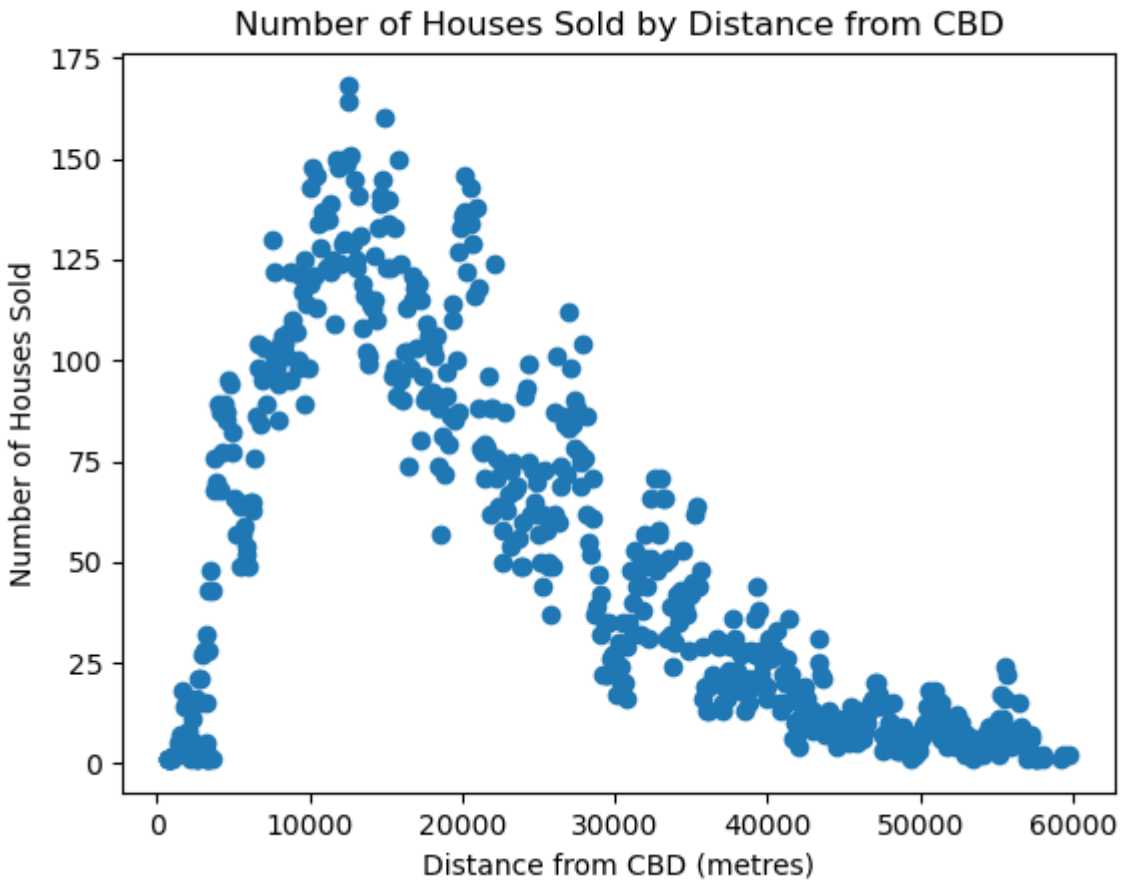
**Best month for sales.**

Sales from across the dataset were plotted by month to observe any seasonality. There seem to be more sales in the months of October and November. This could be because people might be closing in the sales before heading into the holiday season. This also explains the dip in sales during the beginning of the year.



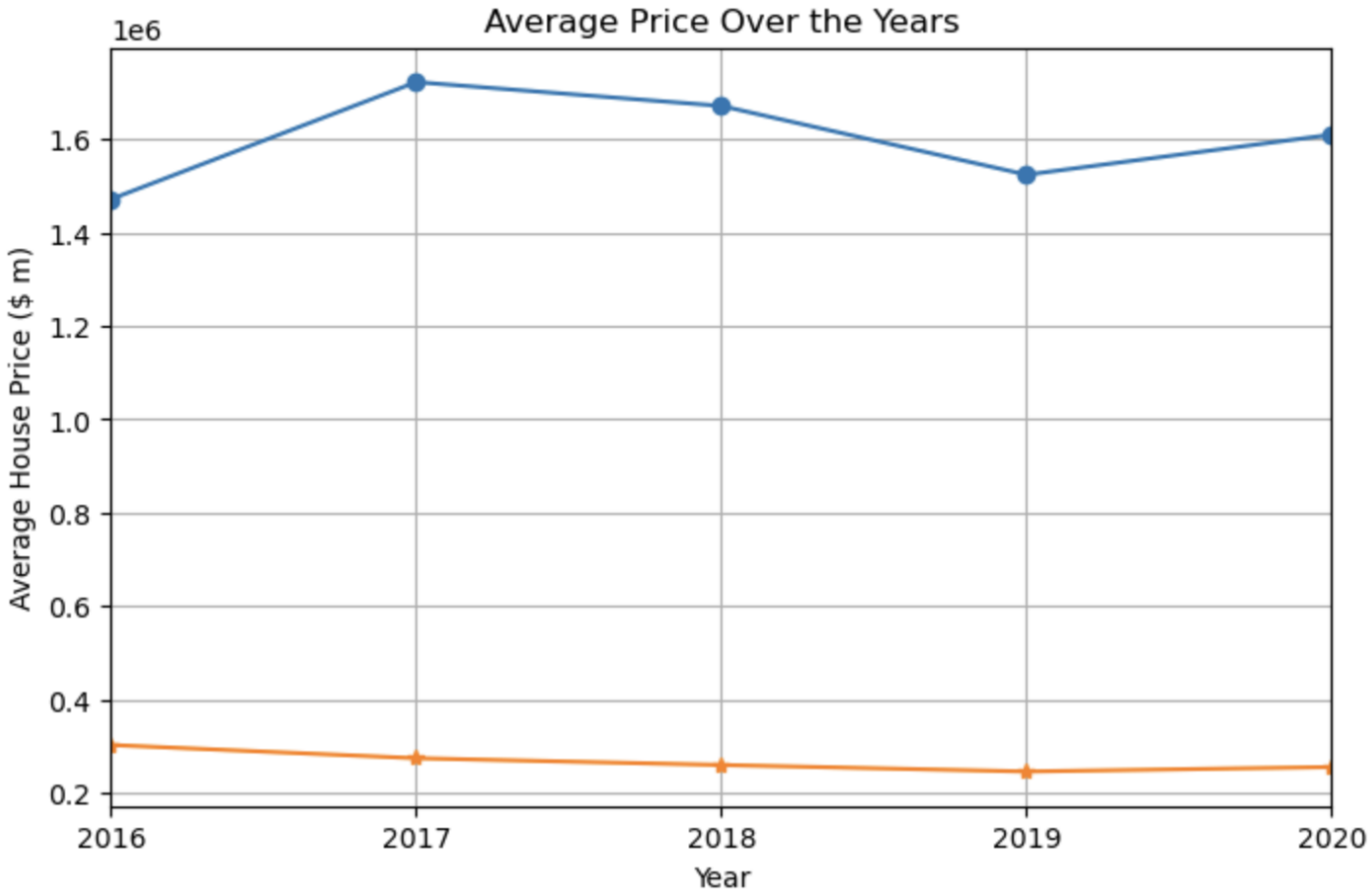
**House Sales by Distance from CBD.**

To observe any trends in sales by distance from the CBD, a scatterplot was generated from the entire data set. Majority of the sales seem to be occurring in the 10-20 km radius from the CBD. This seems to be the sweet spot because the suburbs are not as expensive as the inner-city suburbs, with relatively quick access to the city and they have access to the well established schools. Suburbs within 10 km are more saturated and do not have many sales in comparison to the other suburbs. Future growth will come from the suburbs in the 20-30 km radius as that is where most of the new developments are happening.




**Average house price across top and bottom end of the spectrum.**

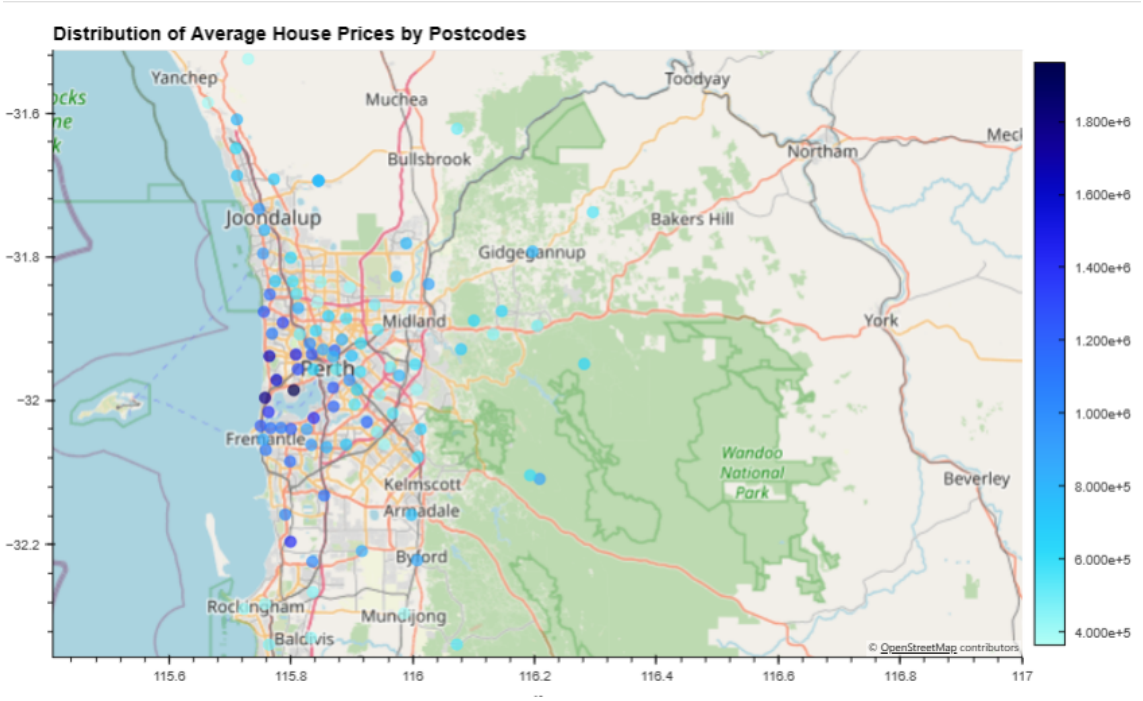
For the years 2016 to 2020, the average house prices of the 10 most expensive suburbs and the 10 least expensive suburbs were plotted to observe any trends. At the pointy end, changes in house prices were volatile. This could be attributed to very few data points. Counter-intuitively, the average house prices in the 10 least expensive suburbs has been declining. Which means getting into the property market for a new buyer seems to be getting easier.




**Average House prices by Postcode/Suburbs:**

A general overview of the average house prices in Perth was completed using a hvplot to visualise the range of average house prices around Perth. Data was cleaned as there were noted duplicate postcodes to simplify the data. The Geoapify API was used to retrieve the longitude and latitude of each postcode and a radius was placed from the city of Perth to include only Perth postcodes.

From this visual map it can be seen that higher average houses tended toward the Swan River and along the coastline, both in the North and South. There seems to also be lower average house price further away towards coastline and further away from the CBD.




**<u>Influences on house prices - Analyses</u>**

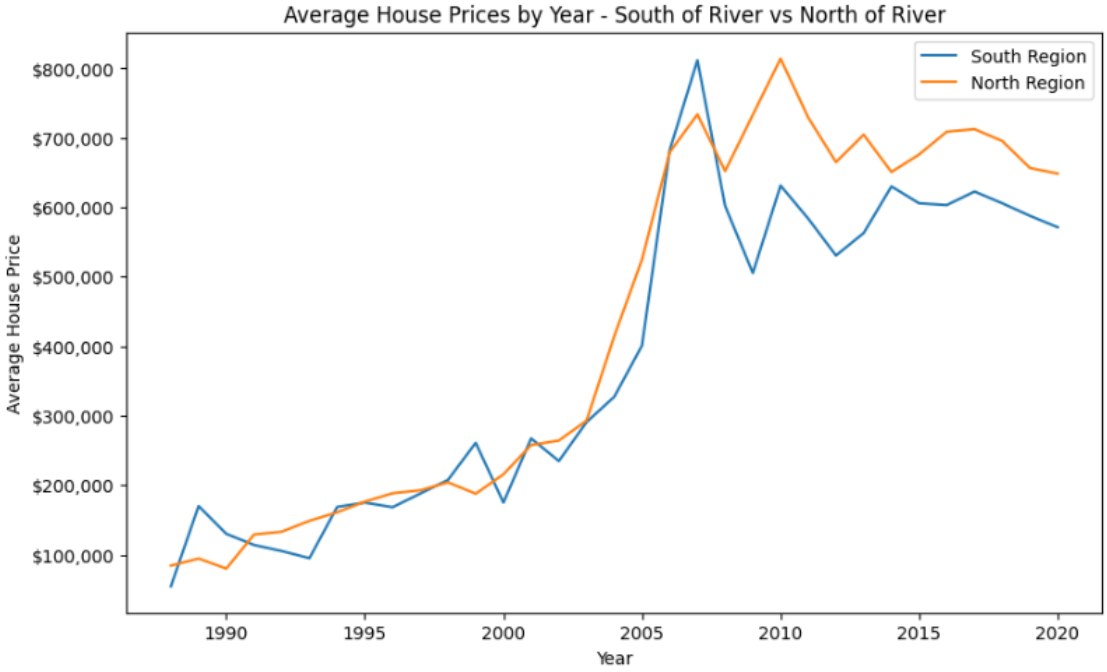
**Is there a price trend between houses in the south vs north of the Swan River?**

The analysis of the houses between North and South of Swan River the dataset was grouped by their postcodes. The definition of South and North were as follows:

- **South of River**: Postcodes equal to and greater than 6100
- **North of Swan River:** Postcodes including 6000 up to 6099
- All suburbs with a postcode of 6200 were excluded from the analysis.

The average price per year was calculated and charted on a line graph for postcodes north and south of the river. There seemed to be an exponential growth of average house prices for both the south and north regions There seemed to be a variation of sale between the years from 2005 to 2010 where the average prices in the north overtook house prices in the south. This trend seemed to continue, which might imply that north house price average will continue to be higher than those in the south.

It is important to note that this graph may also be influenced by total number of data obtained each year where there are possibilities that minimal or less data was accumulated in the first few years of the dataset. Therefore, the average of the house prices may not be accurate.



**Highest and Lowest House Prices by Suburb**

Next, the Top 10 suburbs with the highest average house prices were found by finding the averages of each suburb and sorting them from highest to lowest.

The analysis noted that 8 of the Top 10 suburbs were North of the Swan River, while only 2 of the Top 10 suburbs by average house price were from South of the Swan River. Interestingly further analysis noted that 9 out of the 10 of lowest average house prices were in the South. From here we can conclude there are more houses with higher average house prices in the Top 10 performing suburbs compared to South, and there are more house in the Bottom 10 average house prices in the South than North.

| | SUBURB | POSTCODE | AVERAGE_PRICE |
|---|---|---|---|
| 0 | Dalkeith | 6009 | 1961922.0 |
| 1 | Peppermint Grove | 6011 | 1771982.0 |
| 2 | Cottesloe | 6011 | 1646283.0 |
| 3 | City Beach | 6015 | 1644756.0 |
| 4 | Nedlands | 6009 | 1639099.0 |
| 5 | Swanbourne | 6010 | 1590743.0 |
| 6 | Floreat | 6014 | 1474613.0 |
| 7 | Naval Base | 6165 | 1408000.0 |
| 8 | Applecross | 6153 | 1396130.0 |
| 9 | Mosman Park | 6012 | 1383912.0 |

```
# Check how many postcodes are SOR in the top 10 average house prices
south_count_top_10 = (average_prices.head(10)['POSTCODE'] >= 6100).value_counts()
print(south_count_top_10)
```

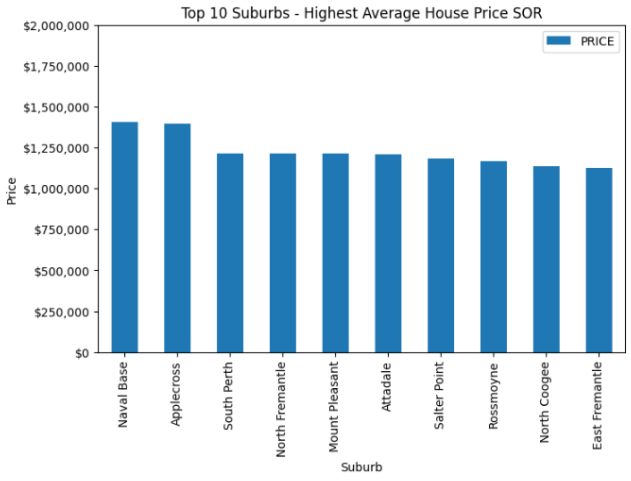
```
POSTCODE
False    8
True     2
Name: count, dtype: int64
```

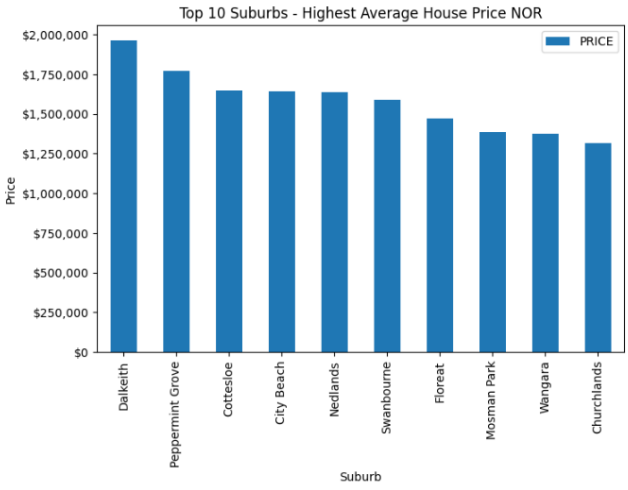
```
# Check how many postcodes are SOR in the top 10 average house prices
south_count_bottom_10 = (average_prices.tail(10)['POSTCODE'] >= 6100).value_counts()
print(south_count_bottom_10)
```

```
POSTCODE
True     9
False    1
Name: count, dtype: int64
```

**Further analysis between North and South average house Prices**

The main data was divided into north and south as per definition of north and south postcodes and a comparison of average house prices between North and South suburbs were also completed with a bar graph. This showed a significant difference in the top 10 suburb average houses between North of Swan River vs. South of Swan River, where average prices of North of Swan River houses at the top 10 were significantly higher than those in the southern suburbs. This would indicate more expensive houses in the North of Swan River in comparison to the South.

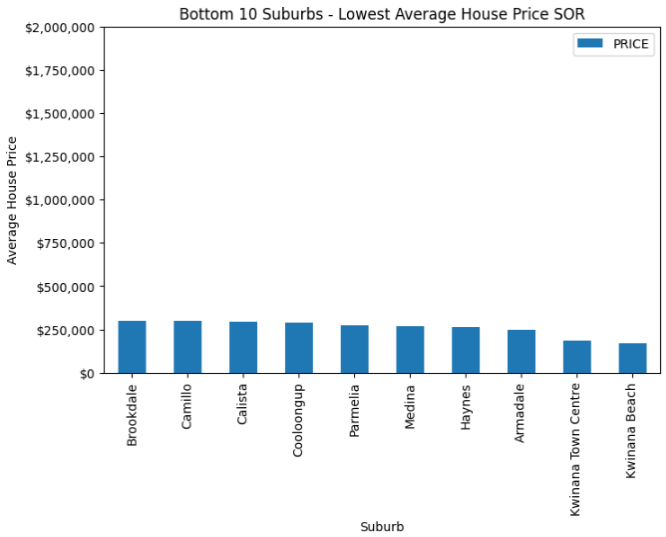


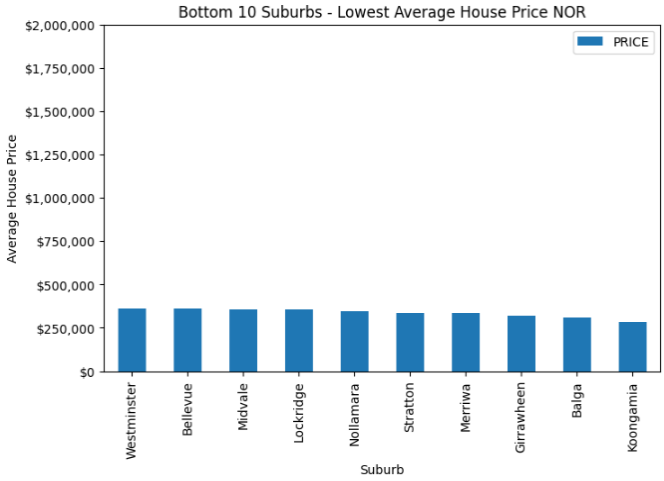




| PRICE_RANGE | SUBURB | AVERAGE_PRICE |
|---|---|---|
| 1m–1.5m | Naval Base | $1,408,000 |
| 1m–1.5m | Applecross | $1,396,130 |
| 1m–1.5m | South Perth | $1,215,011 |
| 1m–1.5m | North Fremantle | $1,213,310 |
| 1m–1.5m | Mount Pleasant | $1,211,816 |

| PRICE_RANGE | SUBURB | AVERAGE_PRICE |
|---|---|---|
| 1.5m–2m | Dalkeith | $1,961,922 |
| 1.5m–2m | Peppermint Grove | $1,771,982 |
| 1.5m–2m | Cottesloe | $1,646,283 |
| 1.5m–2m | City Beach | $1,644,756 |
| 1.5m–2m | Nedlands | $1,639,099 |

Furthermore, the bottom 10 suburbs average house prices were similar, despite the two lowest house averages in Kwinana Town Centre and Kwinana Beach having relatively lower house prices than those of the bottom 10 in North of Swan River. This dataset was further support by placing the averages into appropriate bins where data showed more suburbs with an average house price from 1.5M to 2M inNorth of Swan River compared to South of the Swan River.




| PRICE_RANGE | SUBURB | AVERAGE_PRICE |
|---|---|---|
| 0 - $300k | Medina | $270,912 |
| 0 - $300k | Haynes | $265,126 |
| 0 - $300k | Armadale | $249,044 |
| 0 - $300k | Kwinana Town Centre | $185,598 |
| 0 - $300k | Kwinana Beach | $170,000 |

| PRICE_RANGE | SUBURB | AVERAGE_PRICE |
|---|---|---|
| 300k—500k | Stratton | $337,497 |
| 300k—500k | Merriwa | $333,589 |
| 300k—500k | Girrawheen | $318,077 |
| 300k—500k | Balga | $310,759 |
| 0 - $300k | Koongamia | $282,938 |

**SCHOOL ANALYSIS:**

**Are house prices influenced by the nearest school type (Government vs. Non-Government)?**

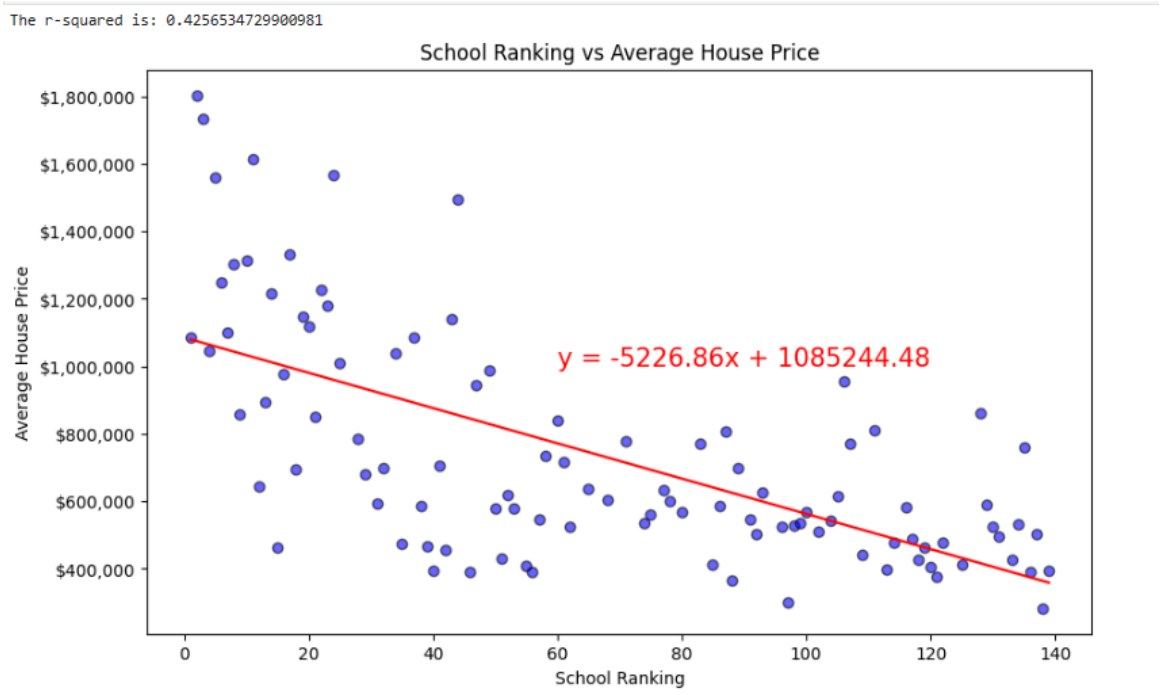
A dataset obtained from the Western Australian Schools List provided by the government was merged into the dataset available to determine whether the high schools provided were classified as government or non-government. The average house prices was then determined by the classification of the nearest schools to the house prices available in the dataset. With the dataset available from 1990 to 2020, the average of houses with non-government high schools nearest to them were above those that only had a government high school nearby.

| School Type | Average House Price |
|---|---|
| GOVERNMENT | $650,877 |
| NON-GOVERNMENT | $757,513 |

**Does the ranking of schools in the area affect house prices?**

To determine whether school ranking had a correlation or effect on house prices, the mean house prices were calculated by grouping then by the nearest school rank. A scatter plot was created and a regression line completed to assess the correlation between school ranking and average house prices.

Below the figure shows a scatterplot with a general relationship between the nearest school ranking and average price of houses. The data indicated a proportion of the variance in the average house price can be explained by the school ranking. This is further supported with the r-value calculated.

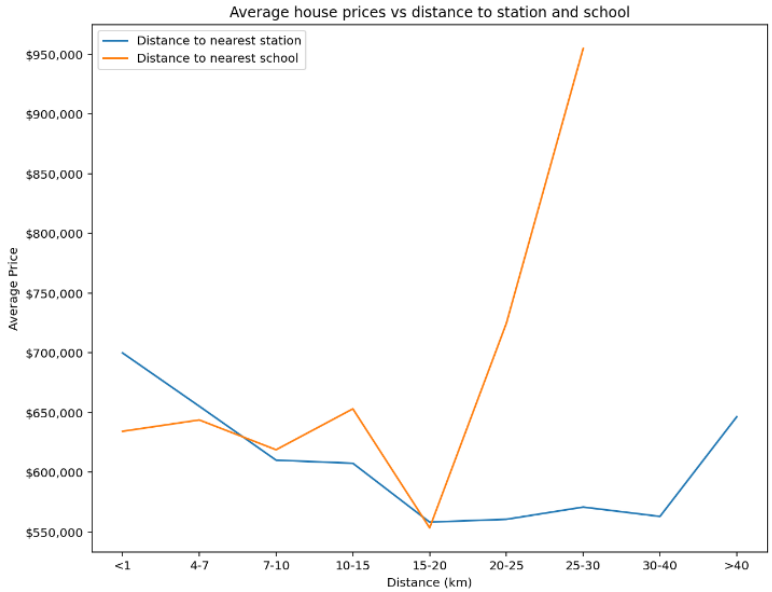


**Does the distance from the city affect house prices**?

To analyse if proximity to the Perth CBD influenced the house prices, distances to the CBD were binned in intervals and plotted against the average property price per bin. The results show a trend from houses between 3-7 km having the highest price and houses greater than 40 km from the city having the lowest price. Houses within 1 km of the CBD partially align with the conclusion that the proximity of the properties influence the price, however the data set does not specify the type of housing such as a unit or apartment. Upon closer look at the data set for housing within 2 km of the CBD, some properties are apartments, units and townhouses. Since the CBD and outer CBD mainly consist of apartments, this may be a reason why the average prices within 2 km of the city are lower than the prices 3-7 km from the CBD. Another consideration is that the suburbs may also contribute to the trend, as prestige suburbs may have high value properties.

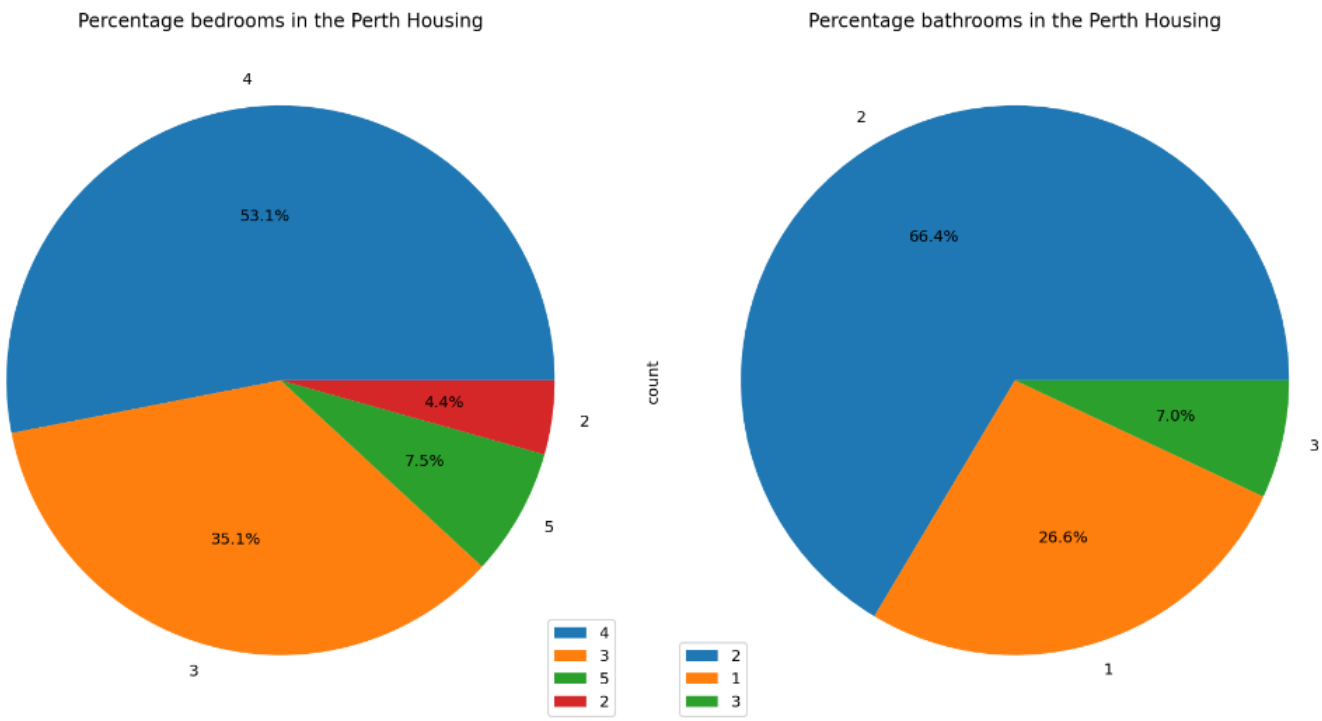

- **Does the proximity of schools and train stations affect house prices?**

To assess if the proximity of schools and train stations had an influence on housing prices, the values were plotted on a line graph. The figure shows a trend between 1 km and 15 kms, where the housing prices were higher when the station or school was in proximity. However overall there is no trench between distance from stations or schools. This could be because some suburbs have bus routes which were not considered in this analysis or the ranking of a school has more of an influence than distance.



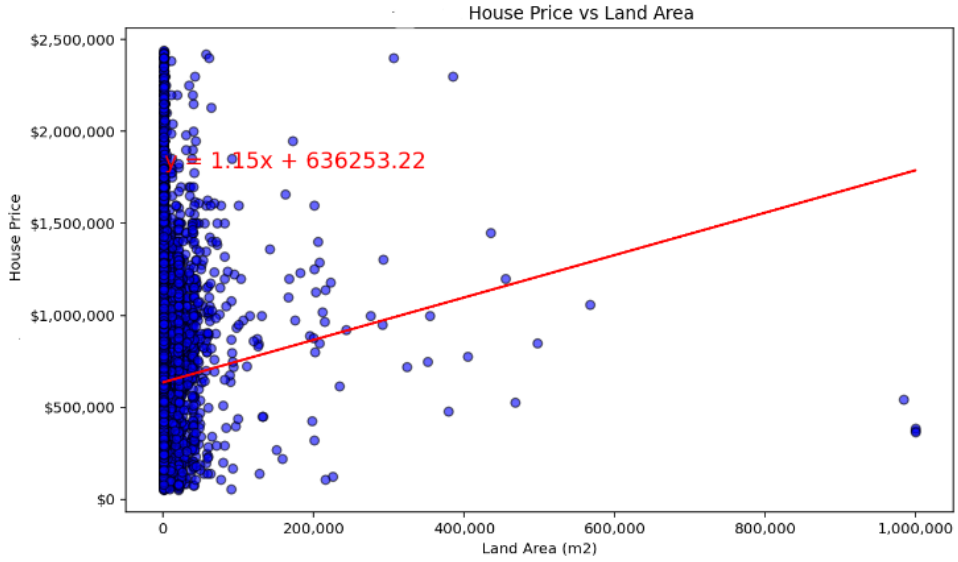

- **Percentages of bedrooms and bathrooms in the Perth housing market?**

The data for the percentage of bedrooms and bathrooms. For an individual or family looking to purchase a house within the Perth housing market, people are most likely to come across houses with four bedrooms and two bathrooms. There are limitations to the dataset provided, the data set did not contain the type of *house,* such as an apartment or Villa. Houses with more than five bedrooms and three bathrooms were excluded from the dataset as they made up less than 1% of the total dataset.

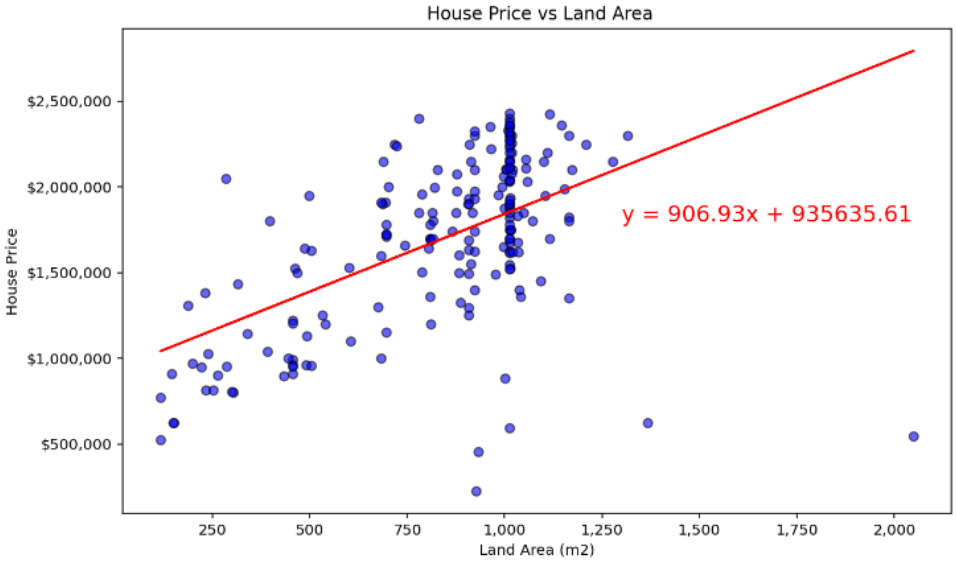



- **Does land area influence pricing?**

The house price and land area were plotted on a scatter plot to identify if there were any trends or correlations (refer to Figure 3 and Figure 4. The correlation coefficient (r) value for the plot below is r = 0.0296 and when the r value is zero (0) it indicates the variables are independent or uncorrelated. There are many factors that could influence these results such as the suburb or distance to Perth CBD. If we choose a single suburb, Dalkeith which is the most expensive suburb north of Swan River. The plot (Figure 4) indicates there is a r-value of 0.286 which is a weak positive correlation. Comparing the houses in a suburb to land area results in a stronger correlation, this is because each suburb's average house value and land size can vary based on factors such as proximity to the CBD.




```
The r-squared is: 0.002961291531073862
```




```
The r-squared is: 0.28625681882207415
```

### Limitations

The following limitations were identified in the dataset:

- The dataset does not contain all the housing information for Perth, the dataset contains 33,000 properties. According to Rentwest (Rentwest, 2023) there are over 818,000 properties in Perth.
- The dataset does not list the type of property such as an apartment or townhouse.
- The dataset only contains information up until 2020.
- The dataset only includes high schools and no primary schools.
- There is also the possibility that not all data were provided including those from the earlier times in the dataset e.g. 1990 to 2010.

### Conclusions

- There are various factors that influences the housing market

- The location, proximity to the city, suburb and school type have an influenced

## 1. References


Rentwest. (2022). *What Can The 2021 Census Tell Us About Perth And The Property Market?* Retrieved from Rentwest: https://www.rentwest.com.au/local-news/what-can-the-2021-census-tell-us-about-perth-and-the-property-market/#:~:text=There%20are%20far%20more%20homes,net%20gain%20of%2064%2C293%20properties.

Zainal, M. (2020). *Perth House Prices.* Retrieved from https://www.kaggle.com/datasets/syuzai/perth-house-prices/code?resource=download

Data WA, (2023), Alphabetical List of Western Australia Schools. Retrieved from. https://apps.det.wa.edu.au/publicreports/SchoolsListExcel0880.xlsx