The Effect of COVID-19 on Australian Real-Estate

Group 10 PROJECT REPORT

Team Members

Matthew Rea, s4430453

Shuoyuan Zhang, s4607609

Haojia Wang, s4605253

Felix Lui, s4671130

Tom Li, s4658439

May 28, 2021

We give consent for this to be used as a teaching resource.

Executive Summary

COVID-19 has affected the entire world in ways that nobody could ever imagine. People around the world have suffered through lockdowns, loss of work, being isolated from family and friends, and, of course, the virus itself. Along with these changes, there have also been great changes in Australian real-estate, in both the volume of properties being sold and the prices they are being sold at. Brisbane, for instance, in recent months has been experiencing a housing boom, with property prices seeming to soar as people are slowly starting to recover from what was, hopefully, the worst of the pandemic in Australia. Knowing where to buy can be hard as changes in prices vary across the country, as well as the effect of COVID-19 on the market. In such an unpredictable market, knowing whether or not to buy or sell at a time when there are spikes in cases can be very useful in helping one maximise their sell price or minimise their buy price.

Table of Contents

1. Problem Solving with Data

Using design thinking, an analysis was undertaken on this data science problem. The aim of the project was to find factors that impact the volume of properties sold and the prices they sell at, so that three sets of stakeholders could benefit. Namely, buyers could use this insight to make more informed decisions about where they want to purchase or if they want to sell, whereas property developers can build developments that suit the current market, and lastly, the government can introduce policies to ensure a controlled real-estate market. Understanding the impacts of COVID-19 and the unemployment rate can help aid individuals looking at either buying or selling to understand how COVID-19 affects the sale price of different properties across Australia. Further to this, individuals could gain insight into the price differences across the country for varying property types and property characteristics, such as the number of bedrooms the property has. Assisting this analysis exists three datasets: one containing sales data on Australian properties from over a year before the first recorded case of COVID-19 in Australia to the middle of the pandemic in July 2020, one dataset containing the number of recorded daily COVID-19 cases in Australia since the beginning of the pandemic, and one dataset containing the unemployment numbers in Australia since 2005. Together, these datasets were used to gain insights into how COVID-19 affected Australian real-estate prices, as well as the Australian unemployment figures.

2. Getting the Data I Need

2.1 Scope of Data

The first dataset we obtained is from Kaggle, covering 6 major capital cities – Adelaide, Brisbane, Canberra, Melbourne, Perth and Sydney. The Australian property market data was collected by HtAG® via a Web Crawler which systematically browses major real estate portals (HtAG Holdings, 2020), so the source is reliable.

COVID-19 data all over the world each day including confirmed cases, deaths and testing. This dataset is provided by Our World in Data, a scientific online publication that focuses on large global problems (Wikipedia, 2021). It contains location and date columns that we also need. By filtering out the records outside of Australia, we can better query the information we need.

Unemployment rate data is added to analyse specific reason for change of real-state prices, whose source is the Organisation for Economic Co-operation and Development (OECD). Unemployment rate in this dataset is measured in numbers of unemployed people as a percentage of the labour force and it is seasonally adjusted (OECD, 2020).

2.2 Data Privacy

Data privacy is one of the most important processes during analysis, so we promise that these three datasets we use contain no personal information like name, gender, home address. Real-estate dataset does not show the person who had purchase the property. COVID-19 and unemployment rate datasets only include records of Australia as a whole.

3. Is My Data Fit for Use

3.1 Outlier Exclusion

In order to detect and find out outliers, we drew boxplots for necessary columns. Here, we examined real-estate and COVID-19 datasets, for unemployment rate dataset, the records are orderly and complete, we did not need to filter outliers.

3.1.1 Real-Estate Dataset

Price: Delete the record whose price is greater than or equal to 100 million and less than 10 thousand.

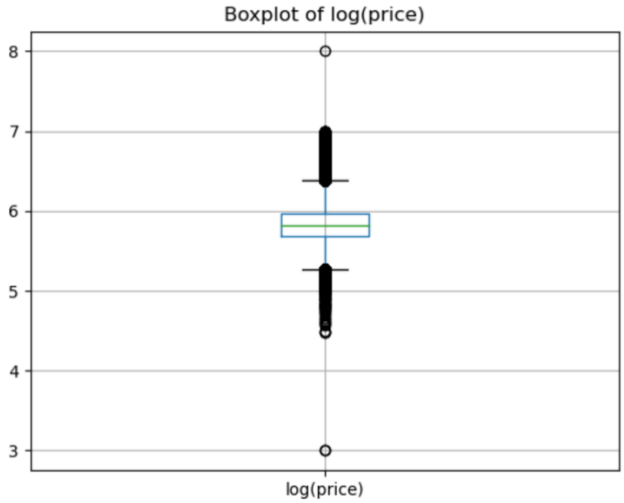

Figure 1 Boxplot of price

Latitude: Delete the value that has positive latitude.

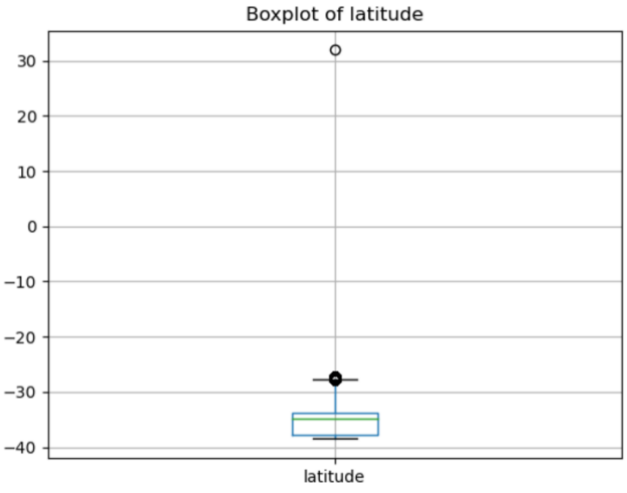

Figure 2 Boxplot of latitude

Longitude: Without obvious outliers.

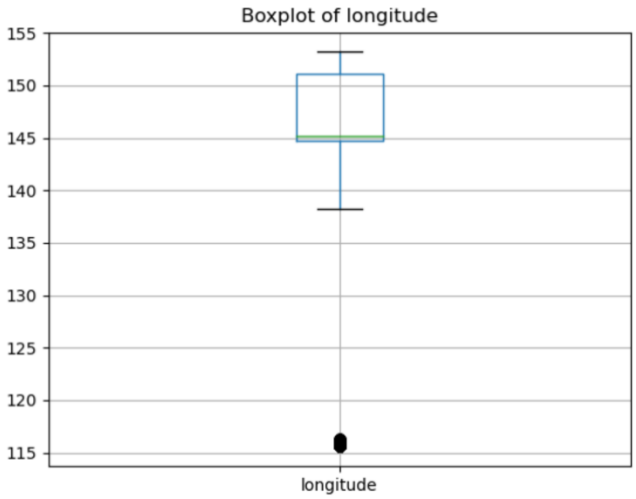

Figure 3 Boxplot of longitude

Bedrooms: Without obvious outliers.

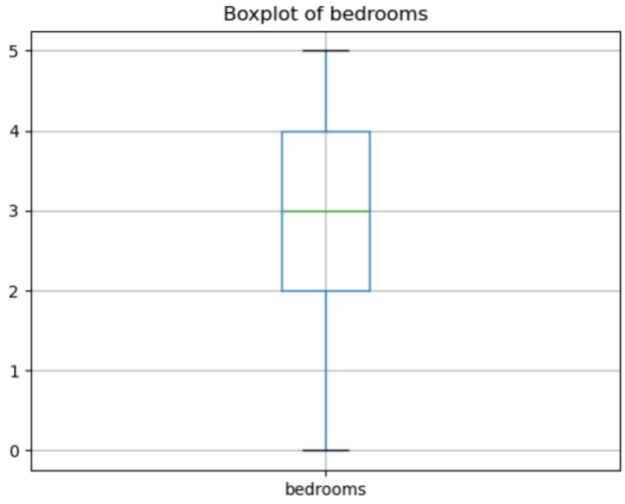

Figure 4 Boxplot of bedrooms

3.1.2 COVID-19 Dataset

Total cases: Without obvious outliers.

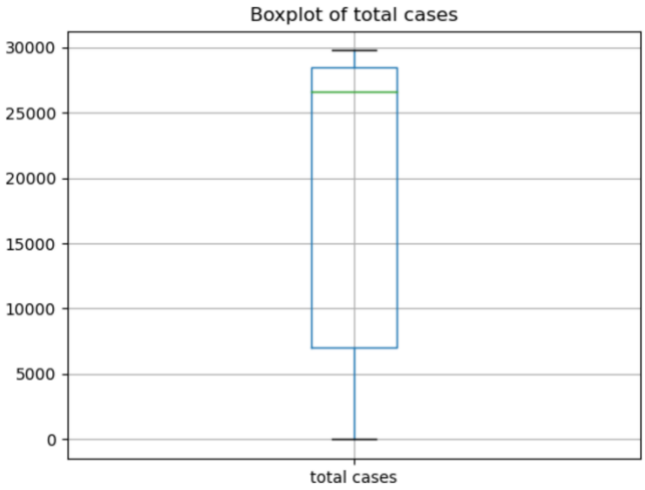

Figure 5 Boxplot of total cases

New cases: Although we can see quite a few outliers on the boxplot, because most of the time, there are not as many new cases, so these "outliers" are actually reasonable.

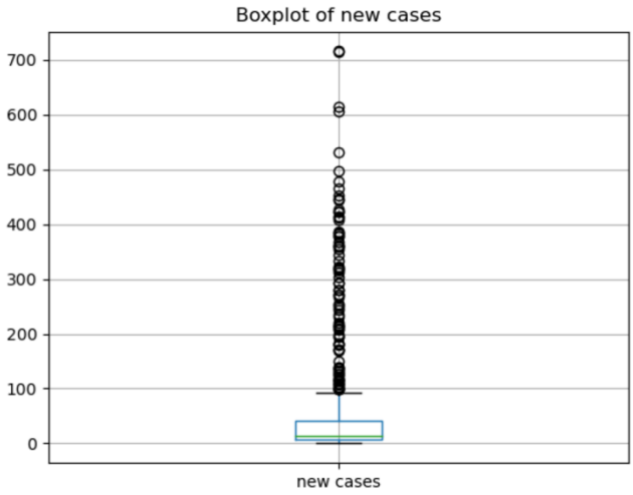

Figure 6 Boxplot of new cases

3.2 Data Imputation

3.2.1 Missing Values

In the real-estate dataset, there are 320334 rows (records) in total, and 62508 missing values in the “price” column and 78 missing values in both “lat” and “lon” columns.

There are a large number of missing values in COVID-19 dataset since at the beginning of the COVID-19 epidemic, many columns have no value, such as “total_ death” and “new_death”. However, we just need columns “total_cases” and “new_cases”, which have no missing values. Therefore, we do not need to deal with missing values of this dataset, and we extract these two columns with the real estate data set and create a new column (more details below).

3.2.2 Dealing with Missing Values

We first dropped all rows with missing values in “lat” and “lon” columns because this kind of missing value is hard to impute and the number of missing values is few. Second, we need to impute the missing values in the “price” column since there are many missing values, which would lead to a severe loss of data features if we just deleted them all.

3.2.3 Imputation Method

KNN imputation (written by python), using “lat”, “lon” (location), “bedrooms”, “property_type”, “date_sold” as predictors and price as response value to create this KNN imputation algorithm.

KNN algorithm is an intuitive algorithm that is easy to understand and does not require too much adjustment to improve performance. The speed of model construction is usually very fast. The last point is that its principle is similar to the distribution between different housing prices, so it is highly explainable.

3.3 Exploratory Data Analysis

3.3.1 Overview of Real-Estate Dataset

We first plot the graph of real-estate sales and sales volume over time and get the result. We can see that the data line has apparent fluctuations, small fluctuations due to the weekend, significant fluctuations due to Christmas (Alex, 2020).

Figure 7 Daily sold vs Average price

Then we plot the graph of weekly average price and focus on the real estate sales from January 2020 to July 2020. We can see that the sales volume rapidly rebounded to the same level as that at the end of 2019 in January and February. Still, the sales volume began to decline for the second time in March and slowly rebounded in May but did not reach the same level as that in the same period of 2019. We can initially speculate that the decline in sales is due to COVID-19, so the next step is to determine our conjecture based on other data sets and policy information.

Figure 8 Weekly sole vs Average price

These six graphs show different region real estate sales and price over time.

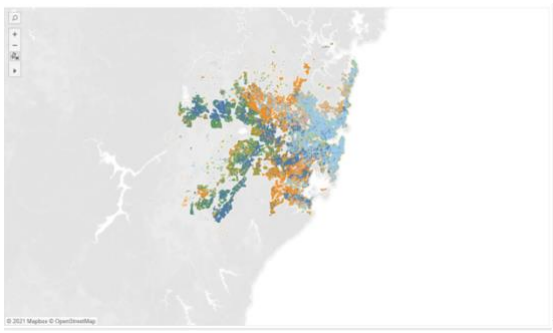

Figure 9 Sydney

Figure 10 Perth

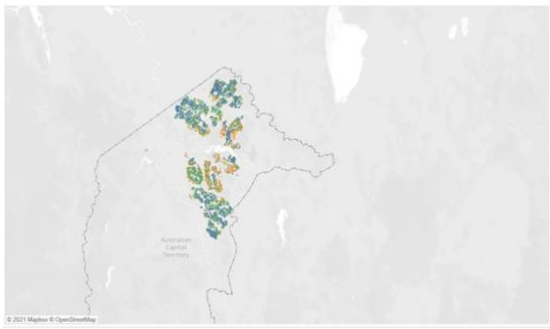

Figure 11 Canberra

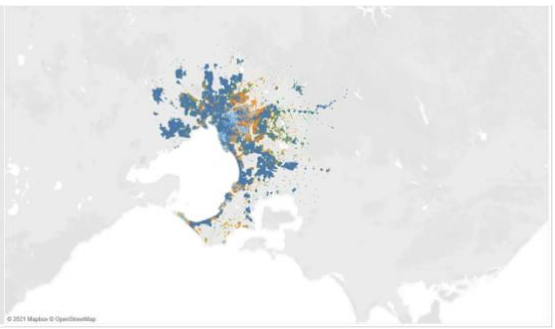

Figure 12 Melbourne

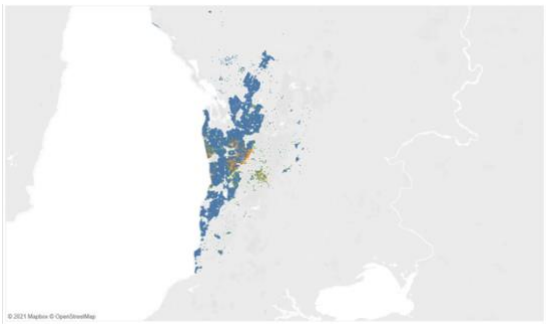

Figure 13 Adelaide

Figure 14 Brisbane

3.3.2 Overview of COVID-19 Dataset

The second dataset is the COVID-19 dataset, including total cases and new cases and so on. Based on the daily newly diagnosed cases data of Australia's COVID-19 epidemic in this dataset, we can find that the number of newly diagnosed cases per day increased sharply from March 13, 2020, to stabilize in mid-April. Therefore, we have reason to believe that people choose to reduce outdoor activities and house viewing activities due to the gradual severity of the epidemic.

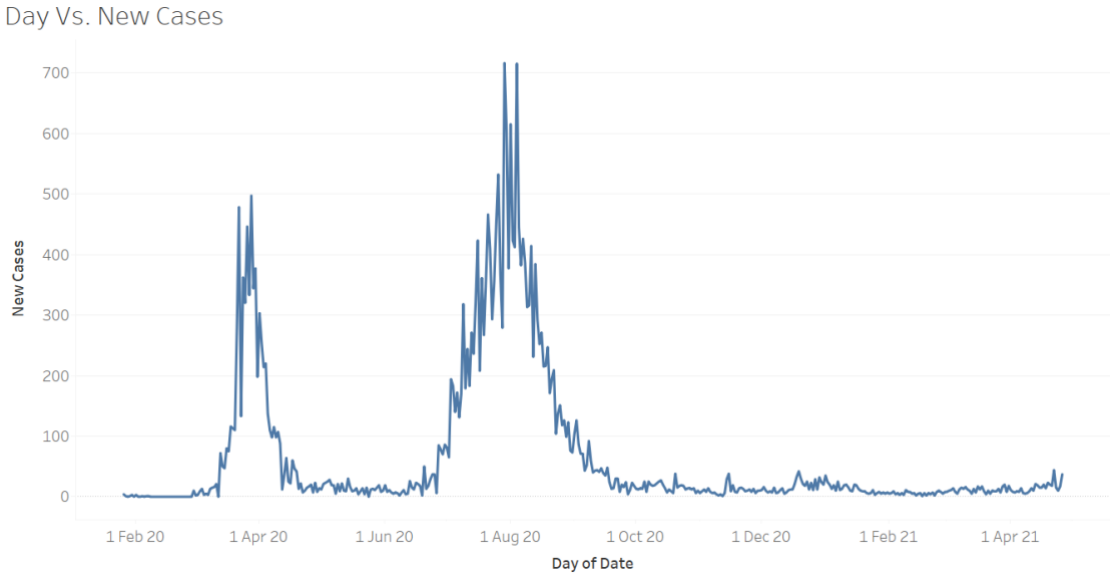

Figure 15 New cases by day

3.3.3 Overview of Unemployment Rate Dataset

Another dataset is Australia unemployment rate dataset. We can also observe from the Australian unemployment rate data set that the unemployment rate began to rise rapidly in April and reached its peak in July, which also explains the negative impact of the COVID-19 on the Australian economy and real estate sales.

Figure 16 Month vs Unemployment rate

3.4 Data Integration

The real-estate dataset was joined with the COVID-19 dataset to get daily and total cases for the date each property was sold. This joined dataset had a new column which indicated the number of new cases since the last date a property was sold, so that models could be trained on a number of cases that were not previously captured if no properties were sold on a given day.

4. Making the Data Confess

4.1 Model Selection

We first construct a prediction model to predict the price based on the AUS real estate dataset. We were using date_sold, suburb, city_name, property_type and bedrooms as predictors and price as the response variable. First, we use 10-fold cross-validation to choose the model with the lowest mean squared error. In other words, it has the best prediction performance. After cross-validation, we choose random forest regression as the final prediction model.

Table 1 Mean squared error of each model

Model Selection	Mean Squared Error
Linear Regression	224,445,384,819
Random Forest Regression	129,751,093,413
K-nearest Neighbours Regression	135,687,918,187
Multi-layer Perceptron Regression	222,690,136,492

4.2 Prediction Using Random Forest Regression

The purpose of this prediction model is to use the data before COVID-19 as training data to predict the real estate price during COVID-19 and then compare the predicted price with the real price to see how COVID-19 influence the real estate price. To gain the most accurate prediction result, we use three different range of data as training data.

Using the data before 2020/01/01 as training data:

Figure 17 First prediction

Using the data before 2020/02/24 as training data:

Figure 18 Second prediction

Using the data before 2020/03/02 as training data:

Figure 19 Third prediction

Although the three results are slightly different, they all show the same trend: the predicted housing price was lower than the actual housing price before May, possibly because the easing of the lending policy mentioned in the previous article stimulated the real estate purchase. The predicted house prices continue to rise and reach the level at the end of 2019 in July, but the real house prices fluctuate at a lower level. This phenomenon may be due to the Australian economic recession caused by the COVID-19 epidemic and the continuous rise of the unemployment rate (from the previous unemployment rate data set, the unemployment rate peaked in July), leading to the decline of purchasing power. Therefore, the real housing price did not return to the same level at the end of 2019.

4.3 Relation between Real-Estate Market and COVID-19

After using the previous prediction model to explain and verify the real estate variation trend, we want to build another prediction model that can help us obtain a more intuitive and direct result on how COVID-19 influence on the real estate price. Therefore, we construct another prediction model based on the joined dataset to find for every new COVID-19 case, the predicted price of a given property rose $92 (this was insignificant)

It was found that there was a large dip in average property prices for each city around Australia on the first day a COVID-19 case was recorded in Australia. Average property prices went up/down after more COVID-19 cases were recorded.

The linear regression model indicated that for every new COVID-19 case, the predicted price of a given property rose $92 (this was insignificant) and the linear regression model showed that the price of a property increased by $190000 for each bedroom it has.

The cheapest city (on average) was Adelaide, whilst the most expensive city was Sydney. On average, units were the cheapest property, followed by townhouses, with houses being the most expensive.

Table 2 Coefficients of different features

Features	**Coefficients**
Adelaide	-117,145
Brisbane	-31,226
Canberra	-19,611
Melbourne	53,471
Perth	-113,466
Sydney	227,978
Bedrooms	195,616
Cases since last sale	92.2

5. Storytelling with Data

We divide property type into unit, house and townhouse for each city. Generally, house is the most expensive and usually unit are the cheapest property type among Australia. Different properties mostly behave in similar pattern over the years. We annotate three key COVID-19 related events date into the graphs. We can see between November 2019 and December 2019, most of the housing price crumble across major cities. However, as previously mentioned, the price drop during Christmas period likely to happen. Therefore, the drop in housing price in this time frame not directly related to the discover of COVID-19 cases. The lockdown and stimulus package have no significant effect on individual property type.

Figure 20 Property prices over time

We merge the property type into a single attribute from the graphs above. Base on the average housing price before COVID-19 as the reference point. We construct the housing price percentage changes after the COVID-19 base on the reference point. The housing price drop by more than 10 percent in cities such as Melbourne, Brisbane and Canberra. These cities housing price then bounce back into much higher level before between the first ever COVID-19 case and first found COVID-19 case in Australia. Meanwhile, cities like Perth and Adelaide keep a rather stable housing price. It partly contributes by the fluctuation in Christmas period. Then, the housing price drop to the usual level and stabilized afterward March 2020. We expected housing price would drop after that because the government impose lockdown, since more people are likely to be unemployed under the unprecedented circumstance. The report there was a global economic recession in that period of time. However, the rollout of first stimulus keeps the housing market steady. Policies such as Home Loan Deposit Scheme, State government first-home buyer incentive and HomeBuilder grants prove working. Comparing the same period in the world, UK housing demands drop by 40% as COVID-19 fears continue to build up (VIRGINIA K. SMITH, 2020). Usually, would create a collapse in housing price as supply were normal. This change over time graph shows Australia's housing price did not affect by the early stages of COVID-19 outbreak.

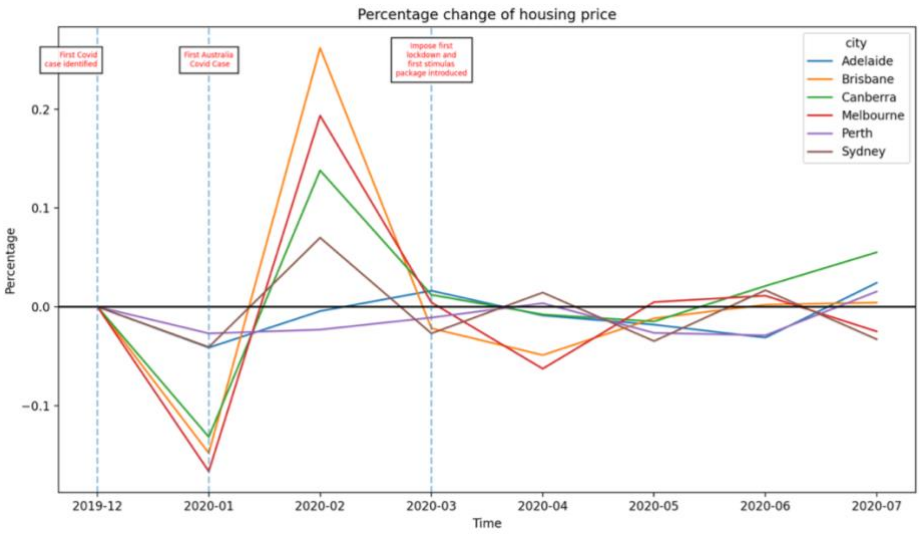

Figure 21 Percent change of housing price

The trend of average price sold for each city before and after COVID-19. From EDA and model prediction, we get a general picture about the average price trend during the COVID-19 pandemic. To support points made previously, we group city column via different city to see whether different cities have different average price trend during the COVID-19 pandemic and the general estate sales in these cities.

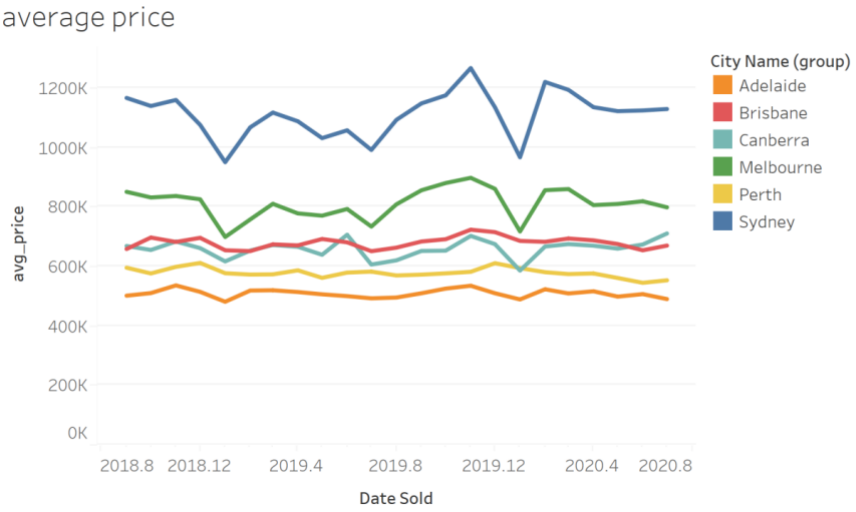

Figure 22 Price vs City

As shown in the figure above, nearly every city especially in Sydney and Melbourne, shows a normal pattern from November to February in the following year, drop dramatically from November to January, then rebounded rapidly till February, part of reason is Christmas vacation. Furthermore, the average price in each city basically maintains at a reasonable fluctuation range, except for Sydney and Melbourne, where real estate price experience not a peaceful trend especially during November to February each year. In the end of 2020.8, average estate price of all the cities approximately maintains the same level as two years ago. As for average price comparison between each city, Sydney's housing price is much higher than other cities, Canberra and Brisbane seems to stand on the same level regarding to average housing price.

Then we want to dig more further to find how COVID-19 had an impact on real estate price in different regions, and thus separate suburb and CBD parts of each city to see whether there is a pattern affecting housing price during COVID-19 pandemic.

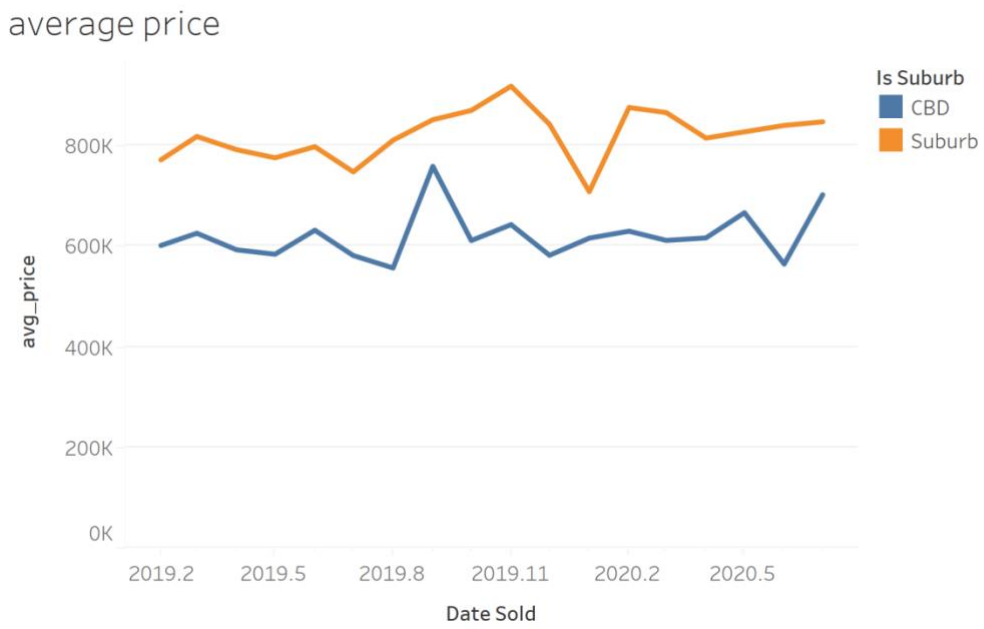

Figure 23 Price vs Region

For the above figure, CBD average housing price is relatively lower than that of the suburb, many influences may take into account, such as difference numbers of each type of house, policies that affect price in specific areas, etc. In addition to the finding that large fluctuation of housing price during Christmas vacation in 2019, like we can see that in the previous chart, we pay more attention on average price changing after April of 2020, the point that COVID-19 has spread into Australia. Compared CBD with suburb in average housing price trend, average housing price in Suburb ticked up a little bit after April, while CBD areas' housing price experience a bigger change.

Some explanation online and our analysis can be made regarding to this phenomenon:

1. Oversupply of vacant rental properties in CBDs led to price falls in these areas (Stephanie McLean, 2021).
2. People prefer properties in suburb, as they realize they can also use teleworking effectively.
3. Property price rebounded in CBD area due to several measures the government introduced to support the economy.

6. Summary and Conclusion

6.1 Feedback Response

On 11th May, we had a meeting with Yoni, and he gave us many useful suggestions which promoted the development of our project. Based on his feedback, we have improved some content in this project.

We fixed up all plots formatting, so they have good labels. All underscores have been deleted, just like “new_cases” was changed to “new cases”, “boxplot of total_cases” was modified to “Boxplot of total cases”. In addition, the abbreviations of some words are written in full. For example, “lat” was written “latitude”, we did not write “lon” anymore, we showed “longitude” instead.

Scaling of graphs was considered. When we drew boxplot of property price, log transform was used to present overview of real-estate market more straightforward. Also, reporting of figures had appropriate precision used.

6.2 Deviation from Project Pitch

We only had one dataset when holding the project pitch presentation, which is about Australian real-estate prices. After analysing data directly in visualising, we saw that there is a precipitous decline in real estate sales from November 2019 to February 2020, including price and turnover. However, we cannot find the specific reason in our dataset, therefore, we sought other datasets to conduct auxiliary analysis.

Finally, two datasets were added, one is about COVID-19 cases in Australia, the other is related to unemployment rate.

7. References

HtAG Holdings. (2020). *AUS Real Estate Sales September 2018 to June 2020.* Retrieved from https://www.kaggle.com/htagholdings/aus-real-estate-sales-march-2019-to-april-2020

Hannah Ritchie. (2021). *Coronavirus Source Data.* Retrieved from https://ourworldindata.org/coronavirus-source-data

OECD. (2020). *Unemployment rate.* Retrieved from https://data.oecd.org/unemp/unemployment-rate.htm

Wikipedia. (2021). *Our World in Data*. Retrieved from https://en.wikipedia.org/wiki/Our_World_in_Data

Alex. (2020). *Visualising the impact of COVID-19 on the Australian property market.* Retrieved from https://www.htag.com.au/covid-19-impact-on-australian-property-market/

VIRGINIA K. SMITH. (2020). *U.K. Housing Demand Plummets as Prices Hold Steady in Response to Coronavirus.* Retrieved from https://www.mansionglobal.com/articles/u-k-housing-demand-plummets-as-prices-hold-steady-in-response-to-coronavirus-213476

Stephanie McLean. (2021). *How 2020 impacted property prices in your suburb.* Retrieved from https://www.realestate.com.au/news/how-2020-impacted-property-prices-in-your-suburb/

Appendix A – Datasets Source

ID	Name	Source	Organisation	Features
1	AUS Real Estate Sales September 2018 to June 2020	https://www.kaggle.com/htagholdings/aus-real-estate-sales-march-2019-to-april-2020	HtAG®	11
2	Coronavirus Source Data	https://ourworldindata.org/coronavirus-source-data	Our World in Data	59
3	Unemployment rate	https://data.oecd.org/unemp/unemployment-rate.htm	OECD	8

Appendix B – Tools and Libraries

1. **Python**

   Modelling for data imputation, analysis and prediction.

   - numpy, pandas, matplotlob, sklearn, copy

2. **R**

   Generate a new dataset with a new column. Analyse data and the COVID-19 effects on housing prices.

   - lubridate, data.table, ggplot2

3. **Tableau**

   Tableau is a good tool to help us see and understand data well. For most visualisations, we used it to generate all kinds of graphs.

4. **Github**

   We created a Git repository and maintained it accordingly, providing an effective integration management approach for the team's overall code version control.

   https://github.com/matthewrea99/DATA7001