Laurie Walker, Project 1. Group: The Yellow Elephants

Project: The effect of Covid-19 on Melbourne house prices.

**Overview**

Our project sought to investigate the effect of the ongoing COVID-19 pandemic on median house prices in Melbourne. For simplicity, we confined this to an examination of the Melbourne residential property market (information of commercial real-estate transactions is harder to find), focusing on houses and excluding units and apartments.

We confined our examination to housing sales, rather than changes in the rental market. Whilst there is evidence to suggest the rental market has also been negatively affected by the Covid-19 pandemic, there is also a large informal component to the rental market, and significant variation in rental terms (e.g. renting a single room vs. an entire property, bills included or not, etc.).

We classified sales data obtained from the Domain.com.au API as either pre- and post- Covid-19, taking March 16th (the date a state of emergency was declared in Victoria) as our cut-off period, and then compared median house price data. Initially, we had planned to create a time series with this data, similar to a model dataset of Melbourne house prices found on [Kaggle](https://www.kaggle.com/anthonypino/melbourne-housing-market?select=MELBOURNE_HOUSE_PRICES_LESS.csv). After further exploration, we found this was not possible with the level of access available with a free API access plan. This led us to perform a simple, ‘before and after comparison’ using sales data from January 2020 and April 2020.

To assist our exploration the data, we formulated three distinct research questions, detailed below.

**Question 1:**

**What, if any, effect has Covid-19 had on the Melbourne house sale prices?**

We expected Covid-19 to have a mild, negative impact on house prices in Melbourne, as measured by the change in median house price by suburb in Melbourne before and after Covid-19.

**Question 2:**

**What suburbs would have the largest and smallest drop in median house prices?**

We expected the largest decrease in house prices to occur in suburbs with a higher proportion of young people, where the workforce was more likely to be lower-skilled and employed in highly casualised industries (such as hospitality and retail) and where there were large number of recently arrived migrants and international students. With these assumptions, I expected to see the worst affected suburbs concentrated in the middle to outer northern and western suburbs of Melbourne.

**Question 3:**

**What demographic factors may explain the different impact of Covid-19 between suburbs?**

As foreshadowed in the previous question the demographic factors generally considered focused on the age composition of the suburb, migration and socio-economic factors such as education and employment. Utilising the limited suburb-level demographic data available via the Domain.com.au API we performed a basic linear regression, focusing on the following variables:

-Percentage of suburb aged 20-39

-Percentage of suburb born outside Australia

-Percentage of suburb holding a job requiring a tertiary education

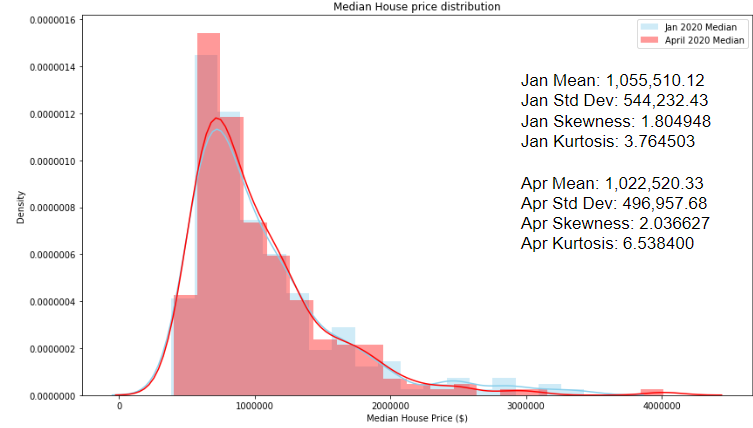
**Findings**

**Question 1**

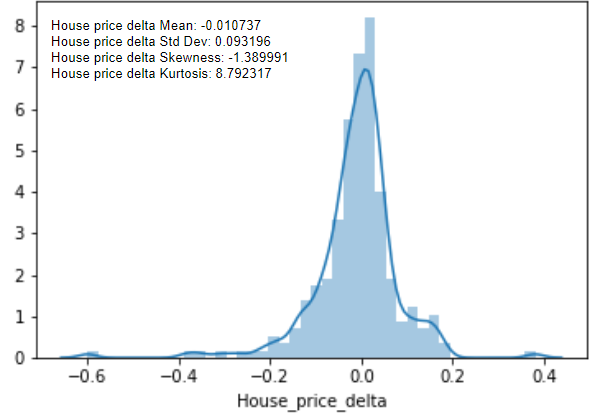
After cleaning the data to remove suburbs that did not have sales in both January and April 2020, we were left with 248 suburbs. Summarising this data, we obtained the following data:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Jan 2020 Median** | **April 2020 Median** | **Jan 2020 Number Sold** | **April 2020 Number Sold** |
| **Total** | $260,711,000 | $251,540,000.00 | 11,000 | 9,669 |
| **Average per suburb** | $1,055,510 | $1,022,520 | 44.53 | 39.15 |
| **Change %** |  | -3.52% |  | -12.08% |

Plotting the distribution of the data using the python module seaborn we see that the main change has been that some of the suburbs with higher median house prices have retreated more towards the mean, as shown by the reduced area to the far right of the x-axis. The decrease in standard deviation between January and April also supports this.



In our dataset we created a new variable for each suburb, named delta, which recorded the percentage change in median house price between January and April 2020. This was plotted, to show the distribution of the variable *delta*, which appears to be broadly normally distributed, albeit with significant excess kurtosis



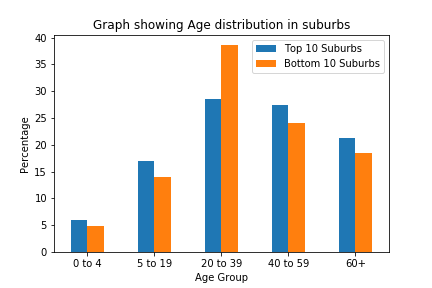
Based on this normal distribution, we were able to do a hypothesis test to see whether the mean for *delta* was statistically significant and less than 0. After testing, we determined it was statistically significant at the 90% confidence level.

**Question 2**

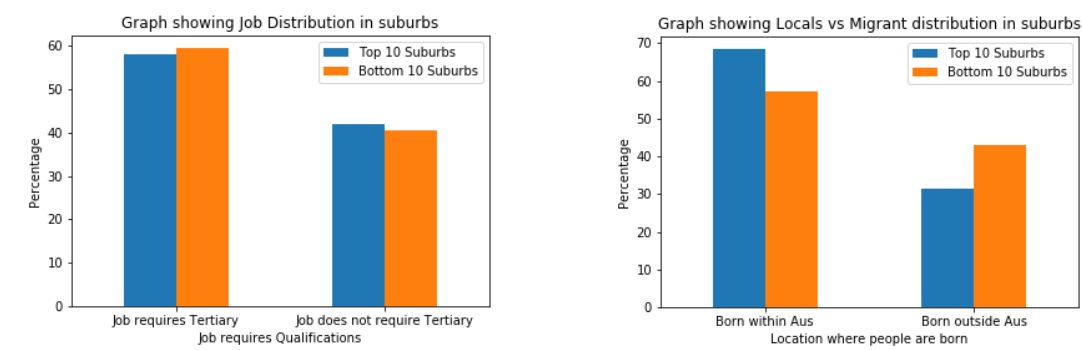
Based on our variable delta, we identified the 10 suburbs with the largest decrease, and 10 suburbs where *delta* was either positive or close to 0. They are shown in the table below:

|  |  |
| --- | --- |
| **Most Affected** | **Least Affected** |
| St Kilda | Lower Plenty |
| Malvern | Carrum |
| Armadale | Fitzroy North |
| Albert Park | Toorak |
| Strathmore | Melton West |
| Fitzroy | Ascot Vale |
| Clayton | Williamstown |
| Warrandyte | Sandringham |
| Healesville | Altona |
| Chelsea | Eaglemont |

Producing histograms for the 10 most affected (bottom 10) and the 10 least affected (top 10) there does seem to be a difference in the proportion of young adults aged 20-39.



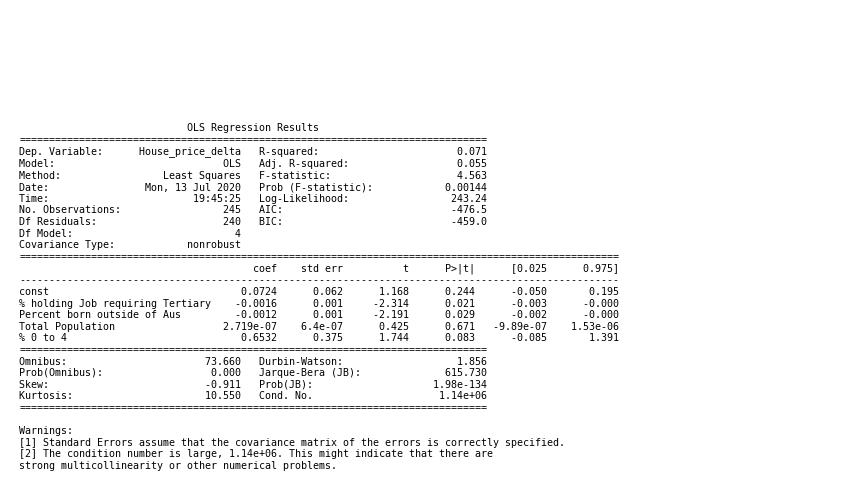
Producing similar charts for migration and jobs requiring a tertiary education, there did seem to be some support for our hypothesis regarding migration, however there did not seem to be much difference between the top 10 and the bottom 10 when it came to the prevalence of people hilding a job requiring a tertiary education.



In general, the most affected suburbs were in unexpected locations.

**Question 3**

After identifying some variables of interest, we performed a linear regression (OLS) using the python module *statsmodel*, with our variable *delta* as the dependant variable. After a couple of regressions with slightly different variables, we obtained the following results:



As can be seen, the overall ‘fit’ of the model, or its explanatory power is quite low, with an R-Squared of just 0.071. On the other hand, it seems that our variables do appear to be at least statistically significant, as indicated by the low p-values associated with the variables *% holding a job requiring tertiary education*, *Percent born outside of Aus* and *% 0 to 4*.

However, all the coefficients are quite small, and in the case of *% holding a job requiring tertiary education*, the co-efficient is in fact negative, which is contrary to our original hypothesis. This is an interesting question, but beyond the scope of our current analysis. Some of this seems to have been reported in an article in [*The Age*](https://www.theage.com.au/national/victoria/data-shows-melbourne-suburbs-worst-hit-by-covid-19-financial-impact-20200608-p550kb.html), based on a far more detailed [analysis of the broader financial impacts of Covid-19](https://taylorfry.com.au/articles/covid-19-financial-impact-index/).

The variable *Percent born outside of Aus* was found to have a negative, and statistically significant relationship with the percentage change in house prices, which is in line with our initial hypothesis. We explain this by the fact that Covid-19 has led to a decline in permanent migration, migrants on working visas and in international students. Demand for housing in suburbs popular with these groups would be expected to be particularly impacted by Covid-19.

In the case of the age variable we found that including a variable for the percentage of suburb between 20-39 was statistically insignificant, whilst the variable *% 0 to 4* has a positive co-efficient, suggesting suburbs with more young families are less likely to be affected by Covid-19. Again, the relationship between demographic factors such as population age and the impact of Covid-19 would benefit from a more detailed analysis than provided here.