What is The Impact of Crime on Affordability of Housing Within Victorian Communities?

1. Significance of Research question

House prices and crime directly impacts the liveability of communities. House prices affect the community's quality of life and affordability while crime not only influences the community's safety and sense of security but also the neighbourhood stability, urban economic development, and the perceived quality of life. Based on these reasons, an average person would like to live in an area with minimal crime to have minimal risk.

2. Datasets used

Crime Statistics Agency Data Tables – Criminal Incidents (xlsx) provides information on crime incidents/rates of different types within Victorian communities.

Victorian Property Sales Report – Median House by Suburb Time Series (xlsx) provides the median house prices within those same communities.

Median House by Suburb Time Series can be linked with *Crime Incidents* by their shared yearly range (2011-2019) and matched based on the community areas.

GI - Working Age Employment and Income (Suburb) 2011 by AURIN (xlsx) was used to match suburb codes and names.

2016 Census GCP (General Community Profile) State Suburbs for VIC by ABS (Australian Bureau of Statistics) (csv) was used to find population in Victorian suburbs.

Victorian Property Sales Report - Time Series (xlsx) was used to find property data on Victorian LGAs.

Working Age Employment, *2016 Census GCP State Suburbs*, and *Time Series* can be linked by their shared yearly range (2011-2020) and matched based on suburb codes and names.

3. Justification for wrangling and analysis methods applied

Wrangling methodology

The *Crime Incidents* and *Median House by Suburb Time Series* datasets were read into python, cleansed of unneeded data, and merged together. First the datasets were read and converted into panda dataframes. For the *Crime Incidents* dataset, only Table 3 was read in as it contained the required data. The *Crime Incidents* dataframe was grouped by year and suburb and crime incidents were summed up. After dropping the unnecessary data, this resulted in the "total incidents per suburb" for each year – which was subsequently written to the disk. For the *Median House by Suburb Time Series* dataset, the data was read and subsequently merged using an inner join with the "total incidents per suburb" data for each year. String methods and functions were utilised to cleanse the suburb names to allow matching.

As a suburb's population had not been considered, the *2016 Census GCP State Suburbs for VIC* dataset was used for recent suburb populations. Additionally, the *Working Age Employment and Income 2011* dataset was utilised to translate suburb codes (i.e., the 2016 census' identification for suburbs) to names. The datasets were merged using suburb codes (which had been cleansed). Further merging occurred with the crime and price data from earlier using suburb names (mostly cleansed with a regular expression). Crime rates per 1000 people were derived from "total incidents" and "population" to use in analysis. This usable data was written as "Crime_Per_Suburb_Per_Year.xlsx" - showing crime incidents, crime rate, and median house prices for a suburb during 2011-2019.

Due to population data inaccuracies, crime rates and crime incidents for Local Government Areas (LGAs) from 2011-2020 from the *Crime Incidents* dataset were used. Table 1 was read, separated by year, and written back. The *Victorian Property Sales Report – Time Series* dataset was used for LGA data. Due to ill-formatting, the data had to be loaded and searched for LGA data's table positions and parsed for each LGA. Only houses were considered for consistency with the suburb data from earlier. Years 2011-2020 were written back to the disk for linking to *Crime Incidents*. Finally, both the property and crime data were merged with an inner join on LGA names – requiring data cleansing using string slicing and string methods. The merged data was then written back as "Local_Crime_And_Property_Per_Year.xlsx" – a completed dataset showing crime incidents, crime rates, median house prices, and frequency of house sales for a LGA during 2011-2020.

Analysis Methodology

Line graphs are used to conduct time-series analysis. They display the trend of the variables clearly through time and help make predictions about the results of data not yet recorded.

For the spatial analysis, the crime and property variables were plotted on scatter plots. Regressions were run to check for linear relationships between variables.

The limitation posed by the two axes layout prompted the use of bubble plots to simultaneously display three variables.

Box plots were used to identify outliers.

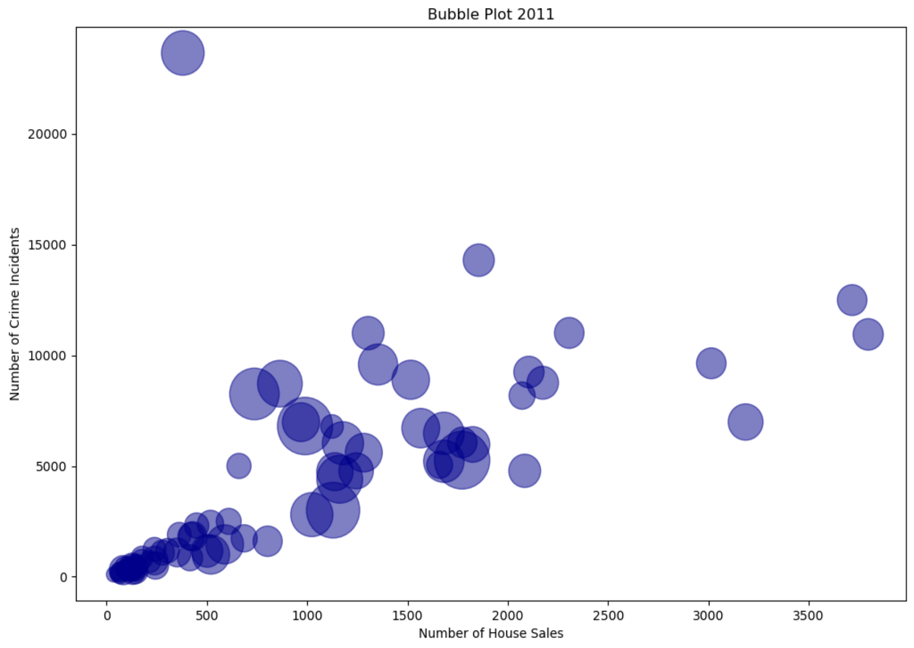
Assumptions:

Crime was generalised as a whole as house-buyers are concerned about all types of crime. Only the median prices for houses and units were available. Without the raw data, an accurate calculation of the median prices for houses and units combined wasn't possible. Thus, houses were chosen over units as they better reflect the market. LGAs were chosen as its data was more reliable than suburban community divisions. Median prices were chosen rather than mean prices as they're not influenced by outliers.

As the variables are discrete and quantitative, Pearson Correlation was used to find the strength of the linear relationship between them. It was not appropriate to use Mutual information as there's no obvious non-linear relationship between the variables. A linear regression is performed on the scatter plots to visualise the extent of the variables' correlation.

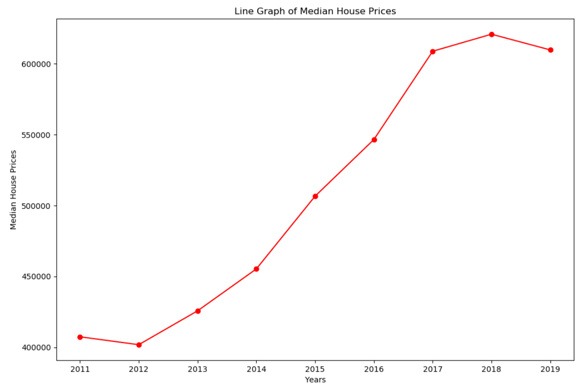
4. **Key Results**

Figure 4.1: Bubble Plot 2011 Showing the relationship between Number of house sales, Number of crime incidents and Median house prices. Bubble size corresponds to Median house prices.

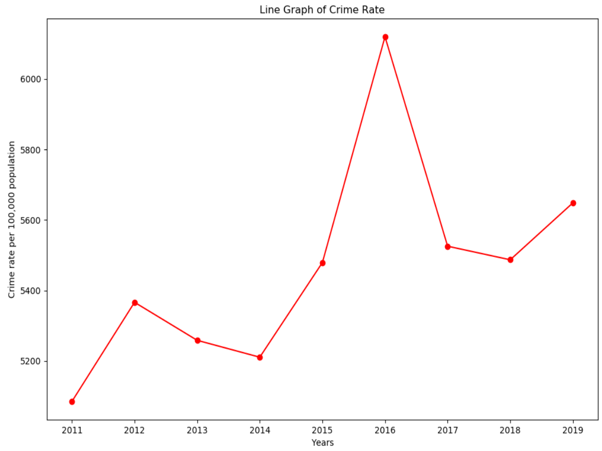


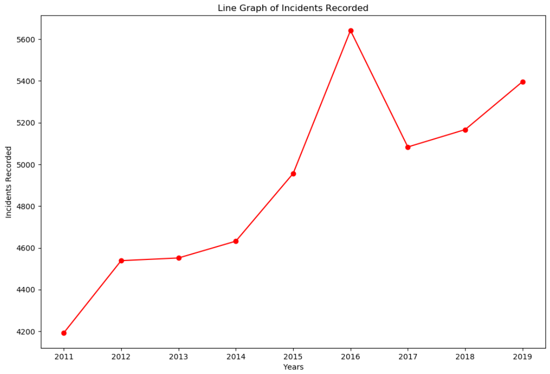
There's a possible correlation between the increase in the number of crime incidents and the increase in house sales. This Is due to the places with high crime having a higher turnover rate. There's a concentration of data points on the bottom left corner of the plot, suggesting that areas with low frequency of sales tend to have low crime rate. The lack of discernible patterns in median house prices suggests a wide variation in prices. Consequently, it can be inferred that there may not be a significant relationship between median house sales and the number of crime incidents versus the number of house sales.

The Big Picture

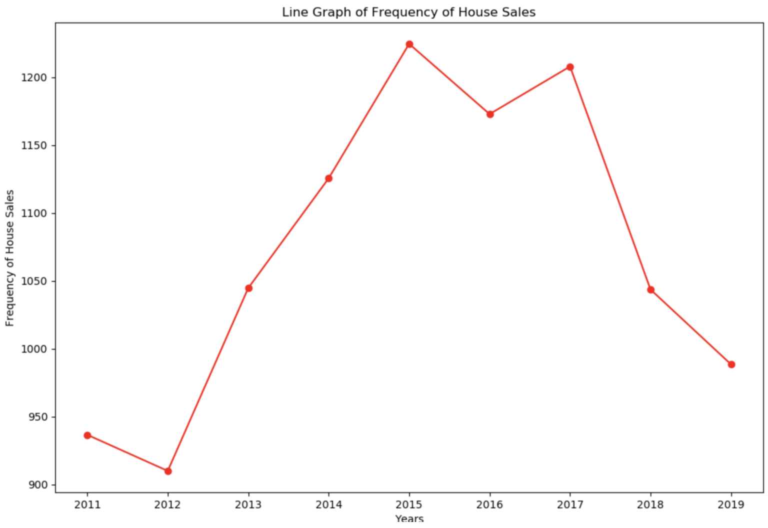


Both the median house price and the total crime incidents in Victoria have climbed steeply over the last decade. To determine whether Victoria's population growth is the driving force behind this rise in crime, a examination of Victoria's crime rates over the years was also conducted. The line graph of the crime rate also rose significantly, corresponding to total crime incidence,





suggesting that the increase in population in Victoria over the last 10 years is not a factor.

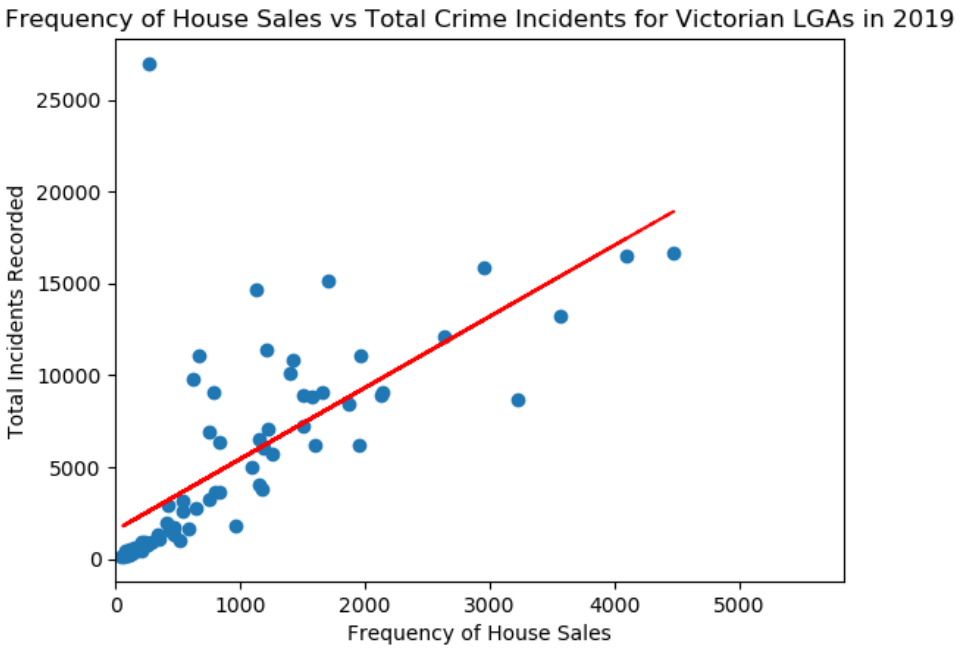


The frequency of house sales has increased from 2012 to 2015 and has been decreasing since then.

One key insight obtained is the strong positive correlation between total crime incidence and the frequency of house sales in the local government area.

Table 1: Pearson correlation coefficient of frequency of house sales and total crime incidents recorded for Victorian LGAs from 2011 to 2019.

Year	Pearson Coefficient
2011	0.670819
2012	0.678088
2013	0.683229
2014	0.699408
2015	0.699231
2016	0.709961
2017	0.677083



2018	0.665628	
2019	0.697132	
The average Pearson correlation is 0.687, which shows a strong positive correlation.		

Table 2: Pearson correlation coefficient of frequency of house sales and crime rate per 100,000 people for Victorian LGAs from 2011 to 2019

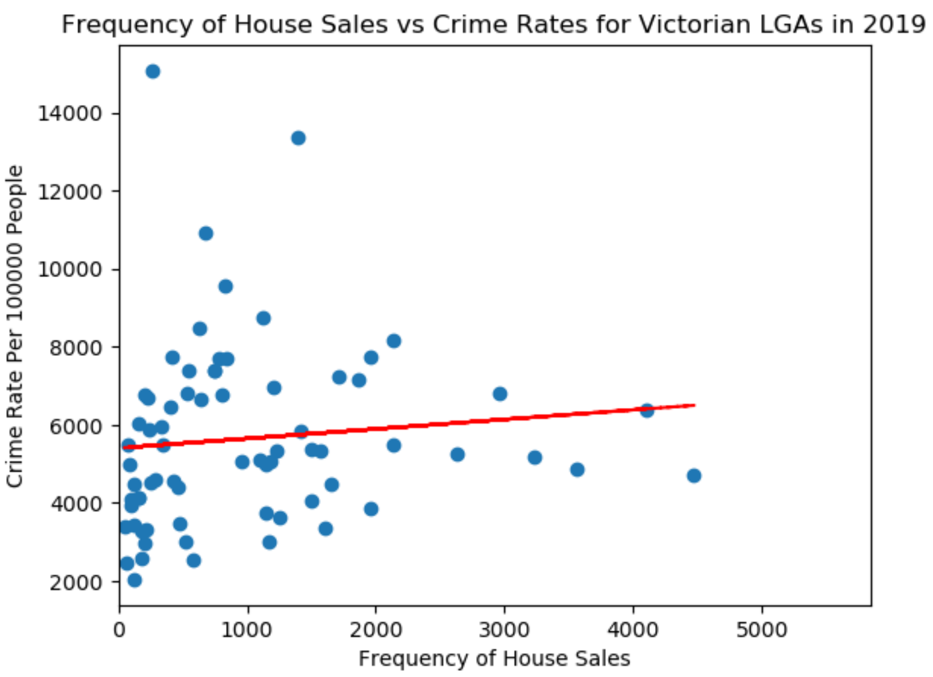
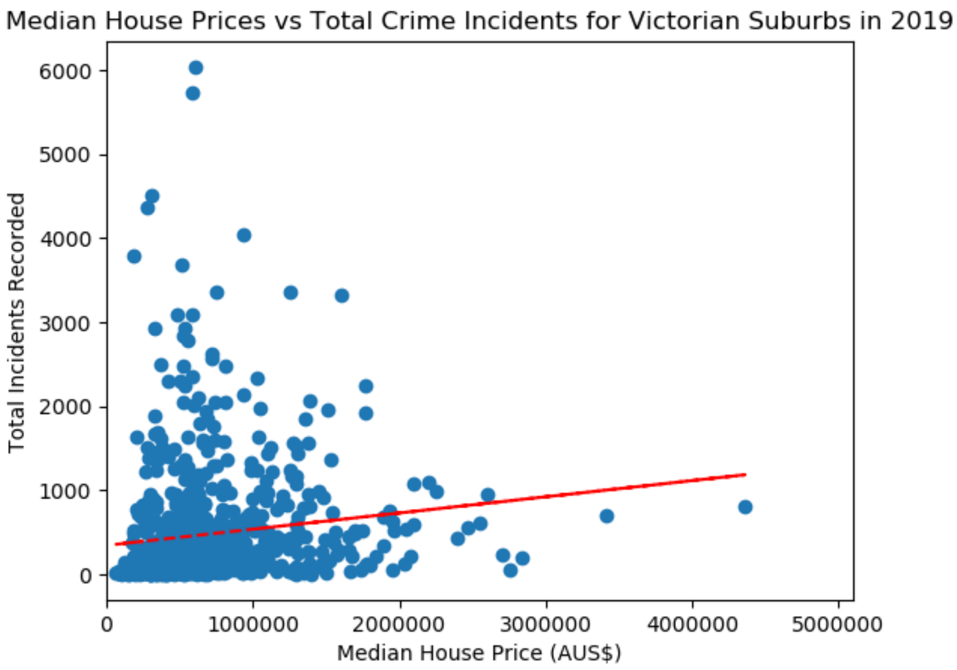
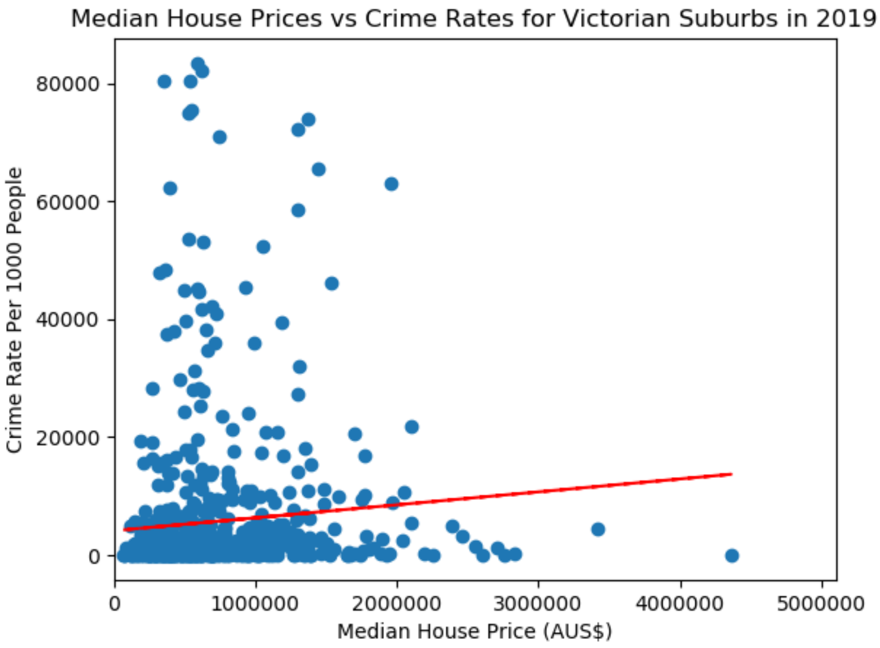
Year	Pearson Coefficient	
2011	0.122407	
2012	0.1396398	
2013	0.1359605	
2014	0.1421561	
2015	0.1271157	
2016	0.138618	
2017	0.0829441	
2018	0.1114659	
2019	0.101970	
The average Pearson correlation is 0.122, suggesting that there’s a low positive linear relationship.		

Table 3: Pearson correlation coefficient of median house price and total crime incidents for suburb population year 2011 to 2019

Year	Pearson Correlation	
2011	0.1375	
2012	0.1408	
2013	0.13681	
2014	0.14032	
2015	0.13492	
2016	0.15026	
2017	0.1593	
2018	0.1496	
2019	0.1278	

The average Pearson correlation is 0.1419, suggesting that there’s a low positive linear relationship.

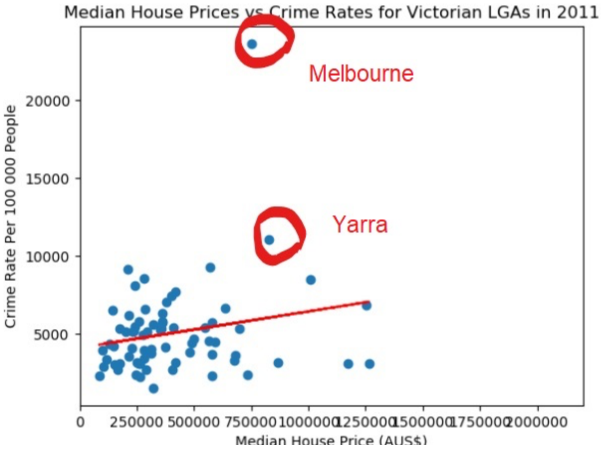
Table 4: Pearson correlation coefficient of median house price and crime rate per 1000 people for suburb population year 2011 to 2019

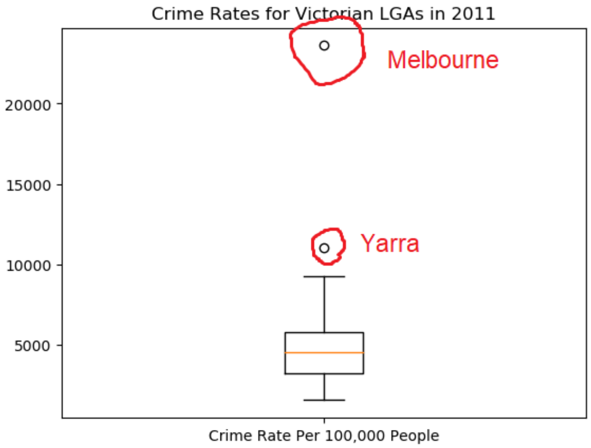
<table>
<tr><td>Year</td><td>Pearson Correlation</td><td rowspan="10">
</td></tr>
<tr><td>2011</td><td>0.077915</td></tr>
<tr><td>2012</td><td>0.097081</td></tr>
<tr><td>2013</td><td>0.089015</td></tr>
<tr><td>2014</td><td>0.061643</td></tr>
<tr><td>2015</td><td>0.058628</td></tr>
<tr><td>2016</td><td>0.107263</td></tr>
<tr><td>2017</td><td>0.10665</td></tr>
<tr><td>2018</td><td>0.094097</td></tr>
<tr><td>2019</td><td>0.082339</td></tr>
<tr><td colspan="3">The average Pearson correlation is 0.0861, suggesting that there’s low positive linear relationship.</td></tr>
</table>

Investigation of Outliers

First outlier: Melbourne LGA
23599 crime incidences/ 100,000 people

Second outlier: Yarra LGA
11049 crime incidences/ 100,000 people



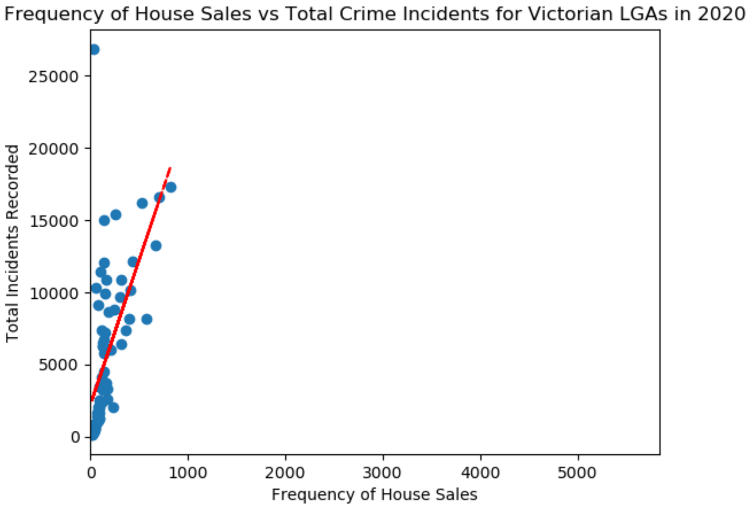


The Melbourne LGA is the area of the city centre. This LGA is the most densely populated LGA so more events and activities provide more opportunities for criminals to conduct crimes. This explains the extraordinarily high levels of crime.

5. Significance of results

Stakeholders need to fully understand how crime impacts house pricing. Conclusions drawn identify places which need improvement; liveability and sustainability could potentially gain extra value in places with low crime. Knowing the crime frequency in an area gives peace of mind to the household's security. This lowers the necessity to invest in security measures compared to other communities. High crime rates imply (disregarding the price) there is less demand in the market, so identifying if the rate of crime has an impact helps stakeholders defer from investing in risky lands. Hard statistics and information derived from information would be much more valuable in general, comparative to implicit information gained from a community's reputation.

6. Limitations of results and improvements



Effect of the Pandemic

Since there are other factors influencing house affordability besides crime, there is statistical noise present in every plot created. For example, the frequency of house sales experienced a significant decline in 2020 due to the COVID-19 pandemic and the subsequent lockdown in Victoria. While the pandemic's unique circumstances are unrelated to crime or property, they nonetheless had a substantial impact on the observed relationships. However, it is noteworthy that despite the pandemic, the correlation between total recorded incidents and the frequency of house sales remained positive.

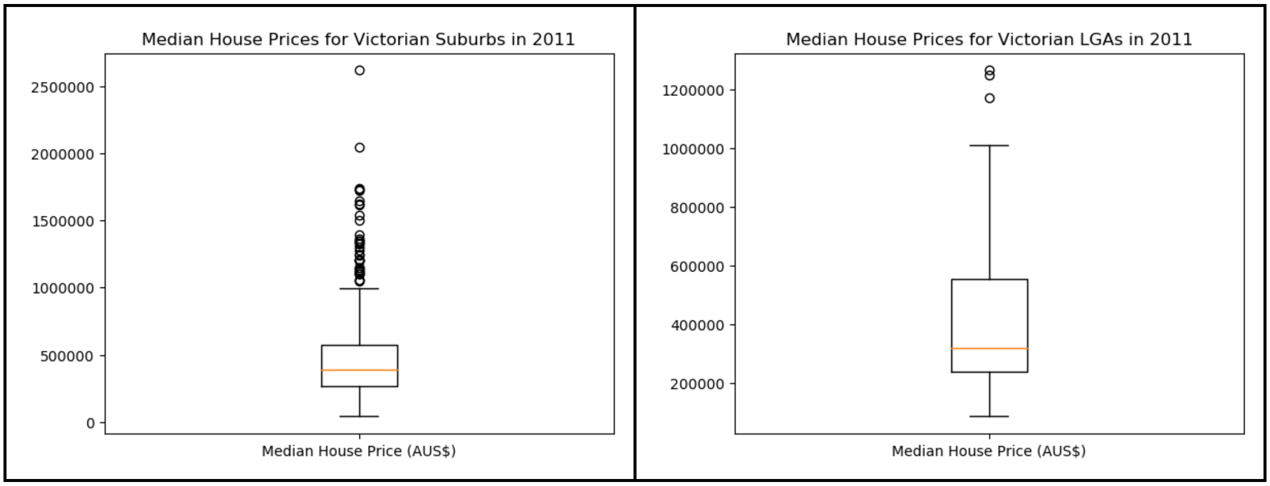
Inaccurate Population Data

The population data on the suburb level is not accurate and this caused the data based on suburbs to be unusable.

The suburb population data retrieved from the census is conducted once every five years, limiting our ability to assess population changes annually. An assumption was made that the population remained constant over the years, using the 2016 suburb population to represent the population for each year.

Moreover, because each suburb spans over such a small area, there were only a few houses sold in each suburb. Consequently, the median house price of the small sample is not an accurate representation of the true value of property, leading to numerous outliers in the box plot (see plots below). Therefore, location was divided into LGAs in the analysis to mitigate statistical noise and inaccuracy.

Location divided into suburbs	Location divided into LGAs



Median House Prices for Victorian Suburbs in 2011

Median House Prices for Victorian LGAs in 2011