🏠 Sydney Property 2018-2020
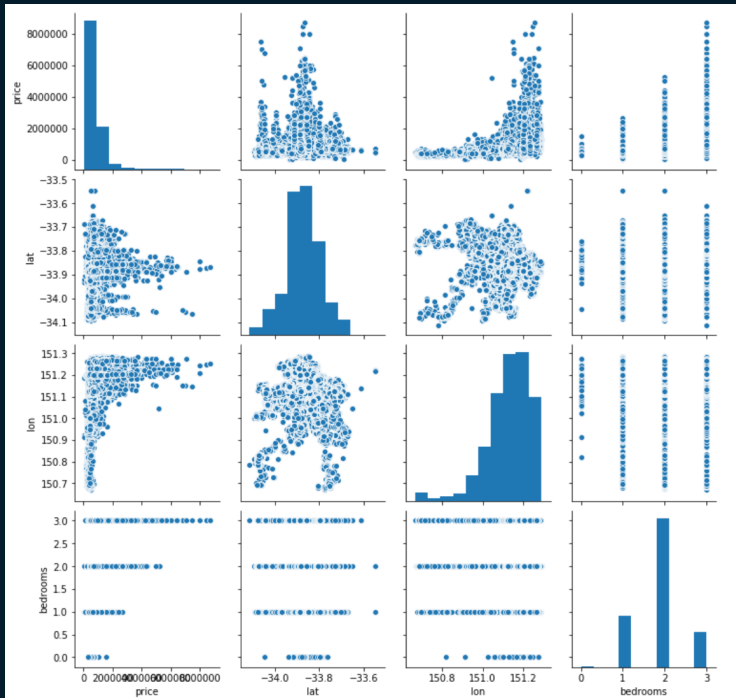
Exploratory Data Analysis (EDA)

## Chart 1: Relationship between all independent quantitative variables

- We observe a positive linear relationship between number of bedrooms and the price.
- Another interesting relationship is between longitude and price, showcasing an exponential increase. This interpretation is prevalent, because it tells us that the more we go east, the more expensive the properties are, whereas the western regions are less expensive.
- On the other hand, latitude does not have an inherent linear relationship with price, as displayed by the scatter plot.
- Furthermore, there is also no relationship between latitude and longitude of the Sydney properties that would be relevant to our data analysis.
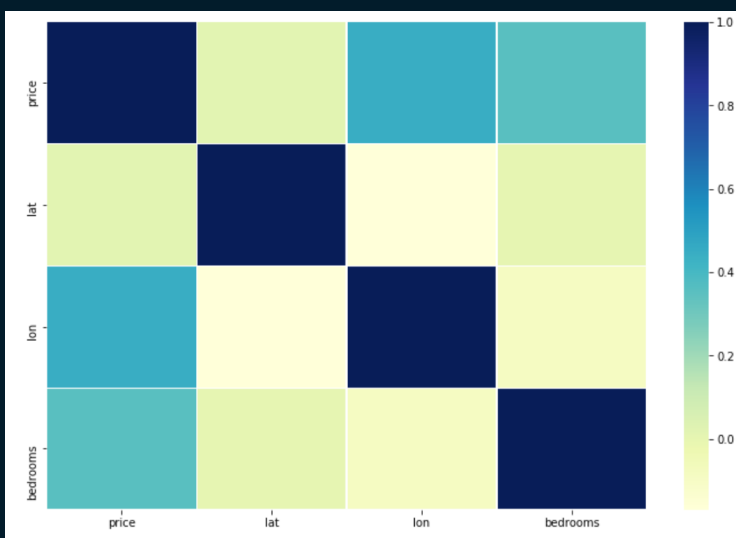
Chart 2: Correlation matrix of independent variables relating to Sydney Properties

- Prior to data analysis, it is imperative to inspect any potential multicollinearity issues.
- Multicollinearity is stemmed from predictors in a model that are linear related to each other, usually indicated from an R-squared value of 0.7 or greater.
- From this chart, the closer the variables are to dark blue, the greater it's correlation coefficients are.
- Overall, all relationships have R-squared value less than 0.7, hence no multicollinearity issues need to be addressed. However, further testing is required.

Chart 3: Missing values in Sydney Properties dataset
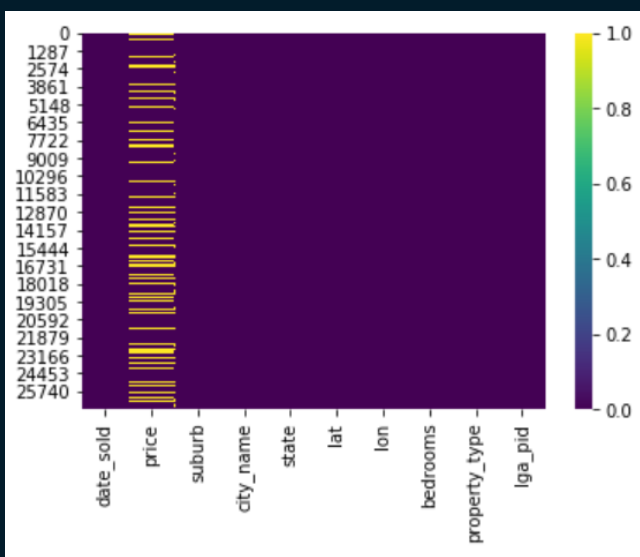
## Chart 4: Missing values in Sydney Properties dataset

The charts above indicate the missing values in the dataset. Chart 3 displays any possible missing values that could be present in the variables. We immediately observe that price has many missing values, however this chart is not accurate, as chart 4 showcases that longitude and latitude has 0.026% missing values and price has 25% missing values. To mitigate this problem, data cleaning is necessary.

Chart 5: Properties in each suburb of Sydney (2018-2020)
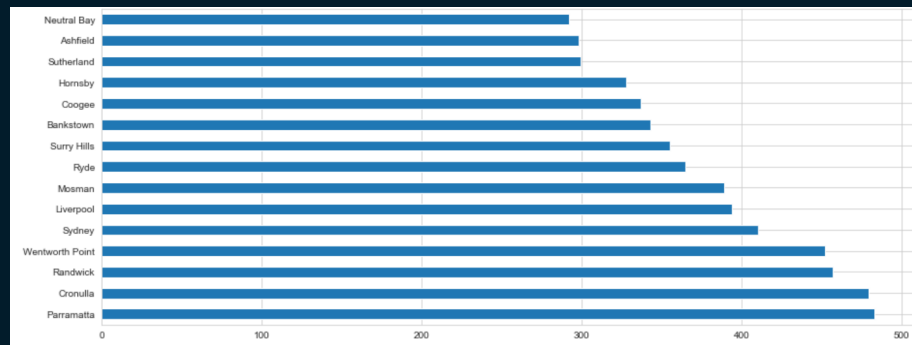
- Chart 5 highlights the number of properties in each suburb.
- It is important to note that there are more suburbs in the dataset, only a subset of 15 were chosen.
- Here we denote that Parramatta comprises of the highest number of properties and Neutral Bay has the lowest amount.

Chart 6: Number of properties in Sydney (Yearly)
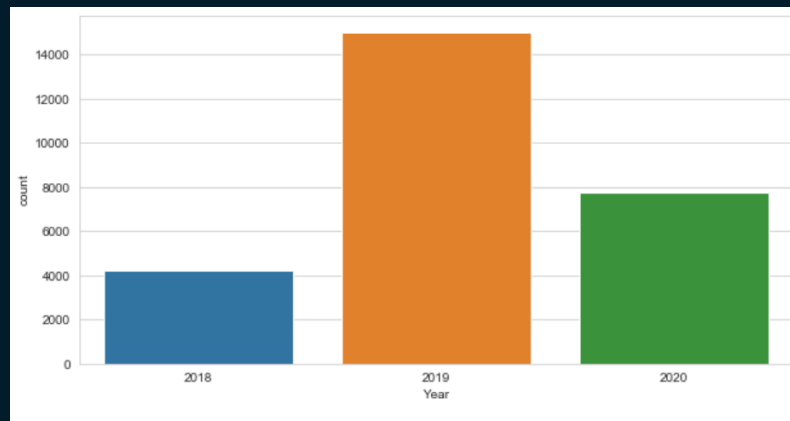
Chart 7: Number of properties in Sydney (Monthly)
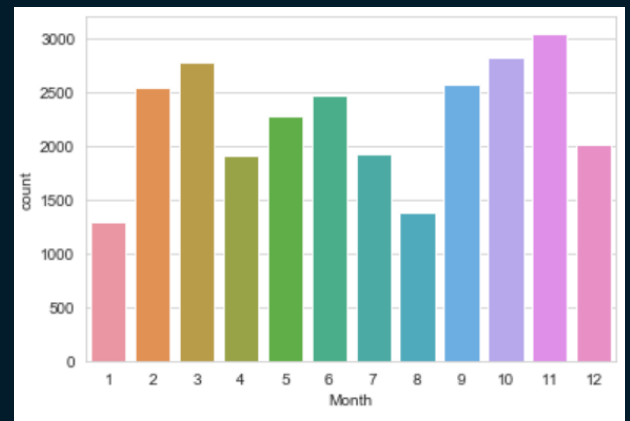
- Chart 6 showcases the number of properties between 2018-2020 on a yearly basis.
- From initial inspection, we see that 2019 saw the highest number of property count, whereas 2018 had the lowest followed by 2020.
- Although we would expect a greater number of properties in 2020 after 2019 saw a major increase, this could be due to the impacts of the COVID-19 pandemic.
- Chart 7 showcases the number of properties on a monthly basis.
- There is no notable value as the count of properties is mostly spread evenly, with January containing the lowest amount of properties and November having the highest number.

Chart 8: Yearly price changes in Sydney Properties (2018-2020)

- Chart 8 is a line graph displaying the price changes between 2018 and 2020.
- It is important to note that the only property type.
- We observe that 2018 had the highest prices ($845,000) and 2019 had the lowest (near $810,000).
- It can be stated that the graph would have a greater spike in 2021 and onwards due to the recent increases in prices for Sydney properties, especially the heavy impacts that COVID-19 has caused.
- Chart 9 on the other hand, displays monthly changes, allows us to analyse the more frequent changes on price.
- February had the greatest price, whereas January had the lowest. The lowest drop in price was between March and April and the highest increase was from January to February.

Chart 9: Monthly price changes in Sydney Properties