**Task 1: Understanding the Data**

What can you say about the various data features?

The variable house price is affected by the other 4 variables, for example, the location or the amount of bedrooms in the house will have an effect on the final house price. So will the agency which is selling the house since different agencies sell at different rates.

What statistical analysis would you apply over the data to understand it better?

I would use descriptive statistical analysis since the csv file contains a large amount of data and therefore visualisation is possible by making use of graphs, charts, tables, and so forth. Therefore it is easier to summarize and understand the data in a meaningful way.

All of the data seems to be skewed to the right. This can be seen when using the data analysis add-in on the excel sheet, or through histograms.

It doesn’t make sense to use data distribution over the whole dataset, due to the large number of non-integer values.