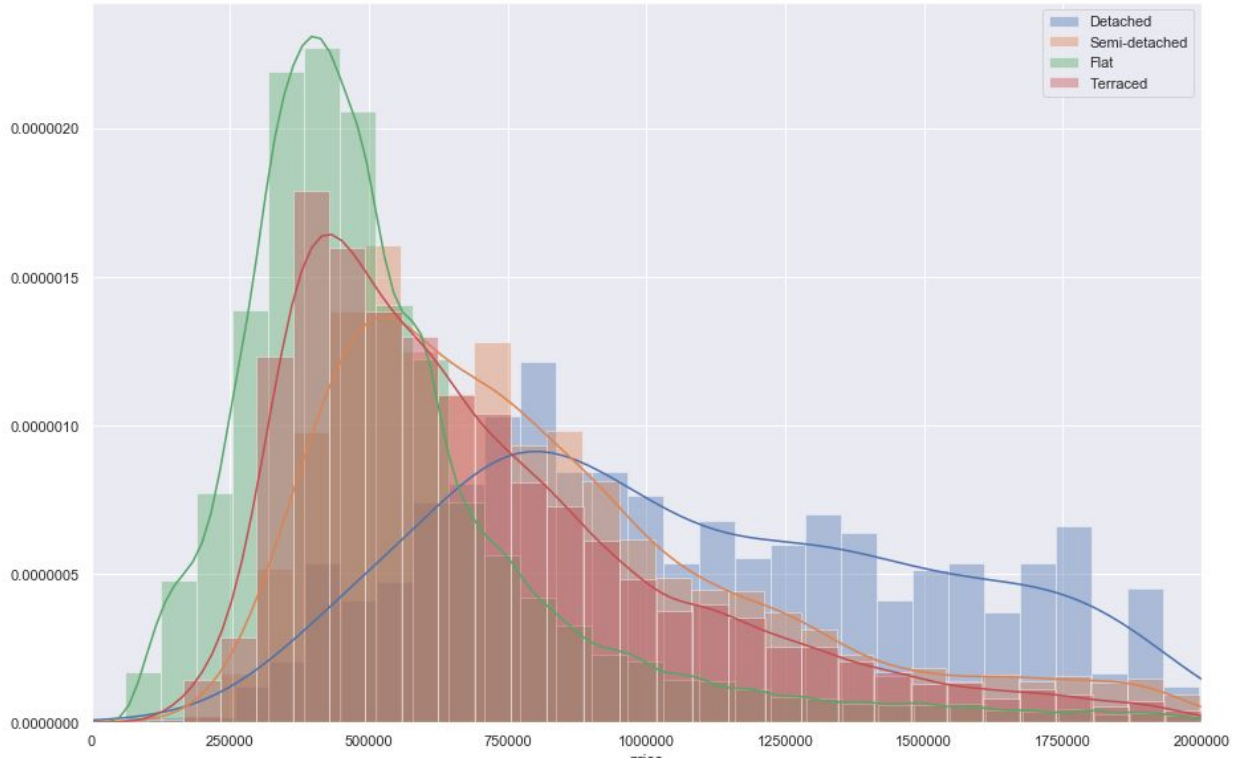
Applying Statistic to Capstone #1

The capstone project examines residential home prices in London, including a focus on the borough of Islington. The main statistical analysis used was to examine the difference between property types - specifically ‘Flats’ and ‘Terraced’ homes. As well, in a similar vein, I used mapping packages to examine the data - although this did not directly apply inferential statistics.

The impetus for examining the difference between ‘Flat’ and ‘Terraced’ properties came after visualising the distribution of prices by property type.

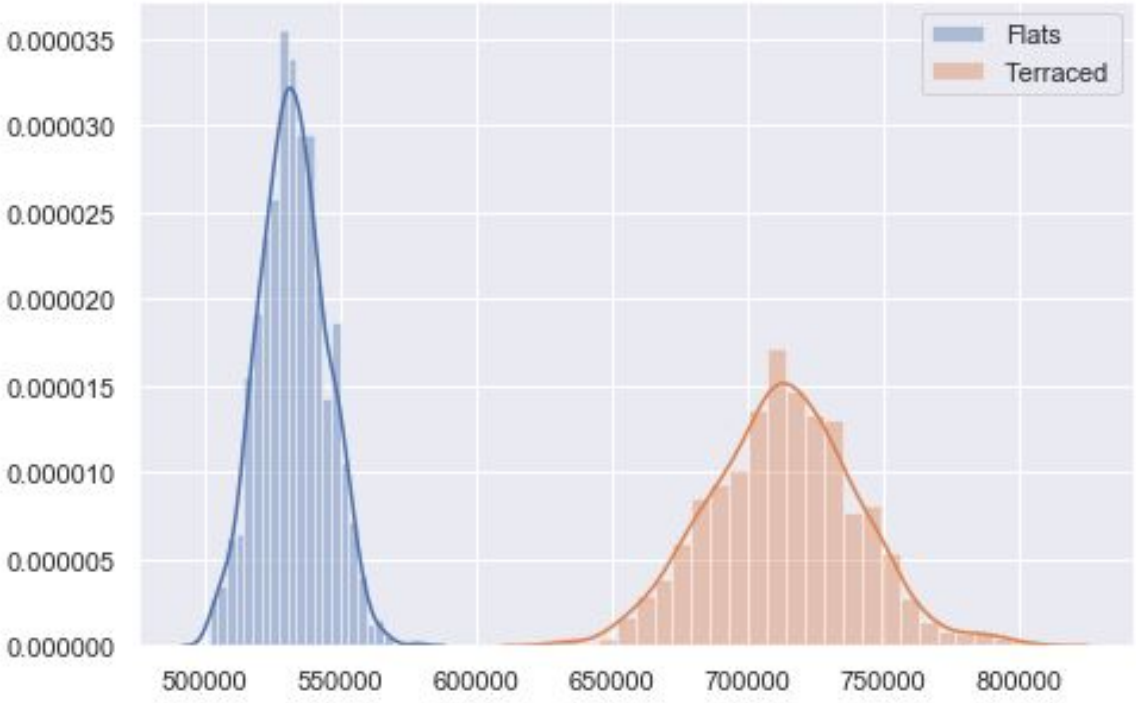


As can be seen, there seems to be a significant overlap between prices for ‘Flat’ and ‘Terraced’ type properties. I am unfamiliar with what ‘Terraced’ means, as this is not a term used to describe homes in Canada. This led to a bootstrapped inference analysis, graphing the means and applying a t-test to the results.

Steps:

1. Determine a bootstrap size (n=1000)
2. Instantiate one empty list to capture means difference
3. Instantiate two empty lists to capture bootstrapped means for both ‘Flat’ and ‘Terraced’

4. Create a for-loop of n size sampled from the subset dataset of Islington (subset of London home prices dataset), using numpy's sample method (with replacement)
    - Take the mean for 'Flat' and 'Terraced' subsets of that bootstrapped sample
    - Calculate the difference between the means of 'Flat' and 'Terraced'
    - Append the three variables to their respective lists
5. Calculate the sum of means differences (results = -181,789,560.05) - which clearly indicates the likelihood of a statistically significant difference between prices of the two properties
6. Graph the bootstrapped results for the two property types



7. Using the stats package, apply the ttest_ind method to the two bootstrapped lists to obtain an independent t-test result with t-statistic and p-value
    - t-statistic = -192.93854369096266
    - p-value = 0.0

These results indicated a statistically significant difference between the prices of 'Flat' and 'Terraced' property types - which tells us that there is a genuine difference.

Future Steps:
In future analysis, it would be interesting to run this analysis across features other than property type, such as geography or demographic factors (schoolage population, average age in area, etc.)