**ME365D - Task 2**

**Dyab Asdi**

Before fitting a regression to check for relationships, I plotted histograms to show the data density of various parameters that we had access to:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

Based on these graphs, we see a lack of data in our middle region for most parameters. The model is skewed as a result but would be most effective for predictions with parameter values near the lower region. We also have quite a few outliers – this data should be filtered out for proper predictions.

Another thing to note is that parameters such as number of bedrooms and baths have significant deviations in their prices for the same value. Although these are important aspects of determining the value of the house, they should not be used to make predictions with a linear regression.

We’ll also want to separate the different house types – “h”, “u”, and “t” for house, apartment, and townhomes, since these are all unique markets, and their prices should be analyzed separately.

Below are linear regressions of various parameters after filtering the data, for some, we can see a relationship that makes sense, for others, there are no proper correlations that can be made.

See attached script called **“housing\_price\_prediction.py”** to generate graphs.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Building Area | Land Size | Year Built |
| Apartment |  |  |  |
| House |  |  |  |
| Townhome |  |  |  |