The initial problem presented was to create a model to find a good estimate for the price of a used car to be put to market. To do this, we began with a large sample of used car listings. From a large number of starting listings, very few had to be removed for being incomplete.

Once the data was organized, a decision had to be made about what to do about the excess of categorical data. Each car has not only a make but a specific model. Despite there being only a handful of makes, there are far too many models. Therefore, we replaced the categorical feature with a numerical one by calculating the average price of each model. This left us with 24 features, which was a still large but manageable number.

We then used grid search on multiple models to compare their performance. All training and evaluation used the same 70/30 train/test split of data. All models tended towards large maximum and mean errors. This is likely due to the limited nature of the data. Cars often have things not listed in the common statistics that can drastically affect their value. For example, if there is significant cosmetic damage or a car has had a history that would make it's reliability suspect, it may have a much lower price than expected despite being average in terms of metrics such as mileage. This means that improving the data collected would be far more effective than any potential changes to the model.