# CS-559-WS Group 12 Report

## Training Data Process

TODO: feature selection

Features were dropped if they met one of several conditions. If a feature had a high percentage of values that were nan, then it was dropped. If a feature was highly correlated with another feature, then one of those two features would be dropped. Categorical variables where most samples fall within one category are not good predictors, so they will be dropped. Any features that lacked relevance were dropped, such as sample id numbers.

TODO: nan replacement

With the remaining features, suitable nan replacements were found. For numerical features, nans will be replaced with the average value. For categorical features, nans will be replaced with the most common category. With these replacements ready, any nans were replaced by these values.

TODO: encode categorical features

Categorical features were encoded as integers.

TODO: standardize the data

All features and the prices are standardized to retain numerical stability and prevent exploding/vanishing gradients.

TODO: use PCA to reduce remaining feature count to 35

The remaining feature count was reduced to 35 using PCA.

## Evaluate Trained Models

TODO: does he want training metrics here? Maybe a discussion of the tuning process?

## Evaluate Test Predictions

TODO: discuss model results I guess…