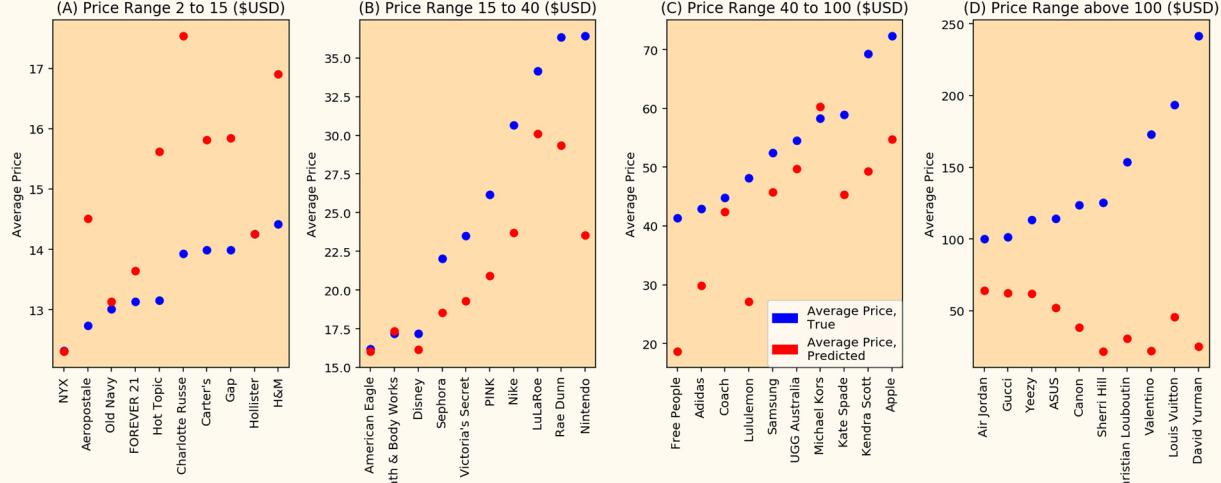


Mercari Price Analysis - Prediction Analysis - Optimal model in predicting prices of products in brands, and predicting products in different main categories

- With our final model, the aim is to evaluate the model on datasets not used in designing the model. The idea is measure how well the model performs in practice. For a starter, we consider predictions on products in brands - to see how well the model performs when different brands of products is considered

Predicted and True Average Price of Products in Brands; Ten Brands in Various Price Ranges;
Predictions on a Validation Set (Independent of Training Set)



(A) The predictions in the low range performs remarkably well, where all predictions are just a few dollars away from the true values.
 - (A) Indicates that our model is in general good for predicting product prices for low-price range brands like H&M, Old Navy and Gap
 - (A) A possible reason might be that a lot of products exists in the low-priced range, which makes our model really efficient in predicting such products

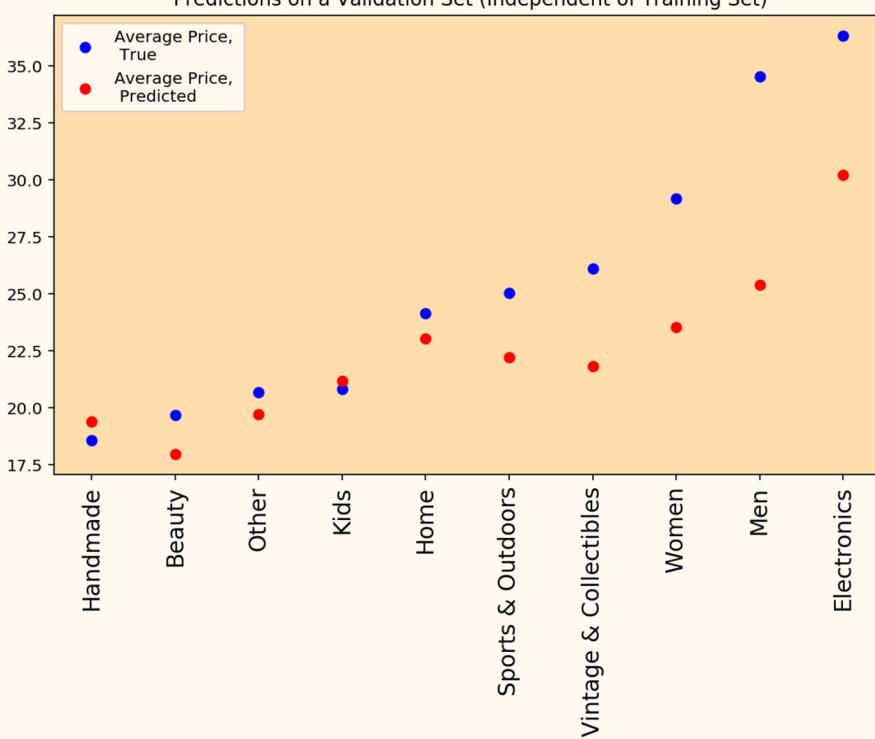
(B) Similarly as in the low-price range, the prediction of price in moderate price range is really well too - again with just a few dollars of margin
 - (B) Hence, our model should perform well predicting prices of products like Nike, Disney and American Eagle
 - (B) The well performance might come, as in (A), from the ubiquity of moderate priced products

(C) The performance in predicting high priced products differs among brands, as can be seen
 - (C) While a few brands can be predicted well, like Coach, Samsung, Michael Kors, the model has a difficulty predicting e.g. Free People, Lululemon
 - (C) However, in general, the predictions are not too bad. One could argue that a margin of 10-15 (\$USD) among high-priced products isn't that much, especially for brands like Apple

(D) Contrary to the other price ranges, the predictions for very-high priced products isn't generally well, as can be seen from the increased margin as the true price increases
 - (D) Most likely, it is due to two factors. Firstly, in our model a portion of training data points corresponding to high cook's distance points were removed, which in the process have decreased our model's ability to predict prices for very-high priced products. Secondly, there is most likely not as many high-priced products in our training data, as opposed to low-, moderate-priced products, which most likely have made our model biased towards predicting the correct price of low- and moderate-priced products

- Further, there is an interest in evaluating our optimal model on products from different main categories.

Predicted and True Average Prices of Products in Main Category;
Predictions on a Validation Set (Independent of Training Set)



The predicted values seem to be quite good, in general, for most of the main categories
 - However, there seems to be some deviations when predicting products in the Women, Men and Electronics main categories. Possibly because products in these categories range from low-priced to very-high priced brands. A major portion of very-high priced brands will have a negative effect on the predictions, as discussed in the predictions of products in brands