Problem description:

First step was looking into data and check their format. Then I split data into a few categories:

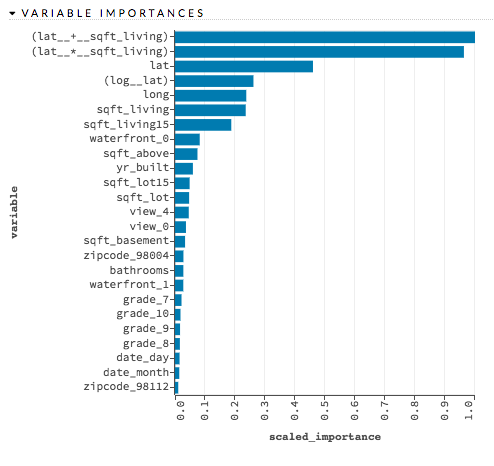
* Reg\_features
* Cat\_featires
* Int\_features

Next step was start to build a pipeline. First step of pipeline was ‘1\_manual\_preparing\_data’. During this step I change all categorical features by one hot encoding. From column with data I take number of month and day. Output of first step is DataFrame with numerical features.

During second step I generated more features by automotive approach. I build XGBoost model. Select most important features and generate extra features. Extra features are sum and multiplication of 2 most important features and log for best features. I do it in a few iterations or until my score (MAE) is improving.

Then my last step is using auto ml generate final model. I used h2o for this.

Below there is a table with variable importance. We can see that the best param is mixed from two other params.



Potential next steps:

* During second step (creating auto features) combine different type of features (cat, reg and int)
* Deeper analysis of data, searching for outliers
* Generate more artificial features