Tmp to do:

* Really, I should just finish it asap. Just make a few other models, check perfs, blend/stack, make the notebook. Just pray that we are in the top 10.

To do:

* Add a few new models
  + Go back to it and decide
    - ~~SVM~~
    - Light gradient boosting
    - Random forest
* Optimisation
  + What would be the approach if i was really serious about opti???
    - Check best performing models and their pipelines and try to reproduce that
  + Try robust scaling
    - Let’s do this I think
      * New file with my full pipeline
    - On xgboost does not seem to change much (actually, worse on my opti model.
    - Yes, works but need to reoptimize the model
  + Go deeper in understanding best pipeline and opti
* Include my best thing in the the Kaggle notebook

Situation:

Tried:

* Kernel ridge (with various kernels)
* Xgboost
* Lasso
* Elastic net

My impression elastic net useless

What I’d like: check the perfs of these models

The last big question was:

* How to I increase my prediction’s precisions?

My plans was:

* Check ppls pipelines
* At least try robust scaler
* Maybe just fake that I looked for the optimal solution. Like if it’s too hard to fine tune the thing

Optimisation. What I tried:

* Training on full training set (not keeping validation)
  + Did not change a thing.

Open questions for opti:

* Did I overfit the training set?

Maybe try

* From this
  + <https://www.kaggle.com/code/lavanyashukla01/how-i-made-top-0-3-on-a-kaggle-competition#Train-a-model>
  + Add features like have\_bsmt
  + Add a bunch of derived features
  + Could add log and square of features

Other notebooks summary and perfs:

* Reach top 10 with simple model on housing price
  + <https://www.kaggle.com/code/argbile/reach-top-10-with-simple-model-on-housing-prices/edit>
  + I used this for my standard pipeline, except
    - Robust scaler
    - Optimized models myself
    - Used a validation set
  + Models
    - Wighted blending of
      * Lasso
      * Ridge
      * Elasticnet
      * Gradient boosting
  + Perfs
    - Test: 0.12136
* How I made it to top 0.3%
  + <https://www.kaggle.com/code/lavanyashukla01/how-i-made-top-0-3-on-a-kaggle-competition>
  + Pipeline
    - 12 splits cv
  + Models
    - Light gradient boost
    - Xgboost
    - Svm
    - Ridge
    - Random forest
    - Gradient boosting
* House price solution top 1%
  + <https://www.kaggle.com/code/jesucristo/1-house-prices-solution-top-1>
  + Pipeline
    - 10 folds
  + Models
    - Nothing new

Some notes about other notebooks

<https://www.kaggle.com/code/argbile/reach-top-10-with-simple-model-on-housing-prices/edit>

→top ten. Followed this one a lot.

* Lasso
* Ridge
* ElasticNet
* Gradient Boosting

Top 0.3%

<https://www.kaggle.com/code/lavanyashukla01/how-i-made-top-0-3-on-a-kaggle-competition>

lgb: light gradient boost

xgboost

svr: support vector regression

ridge regression (non linear I think)

rf: RandomForestRegressor

Gbr: GradientBoostingRegressor

→ tbh what the diff with xgboost?

<https://www.kaggle.com/code/jesucristo/1-house-prices-solution-top-1>

* Lasso regression model (great perform)
* XGBoost model
* LGBM model
* Dragon model

Also catboost to check

→ <https://www.kaggle.com/code/sarthmirashi07/top-8-house-prices-ensemble-learning>