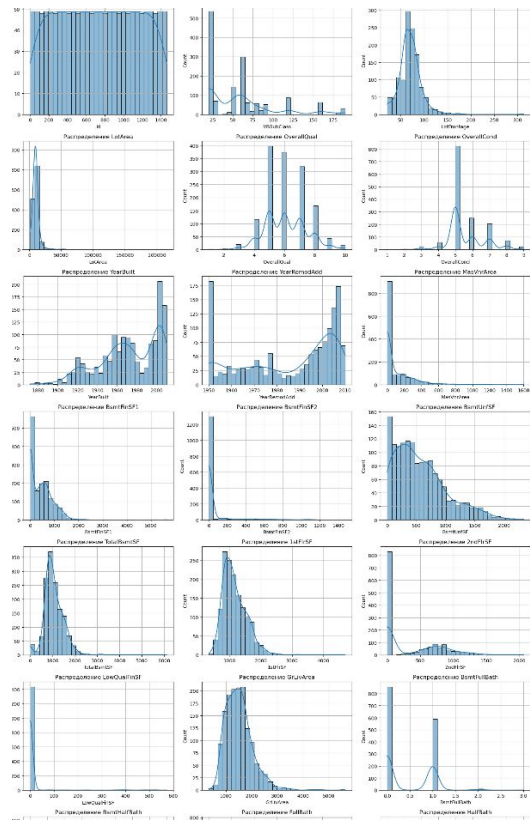


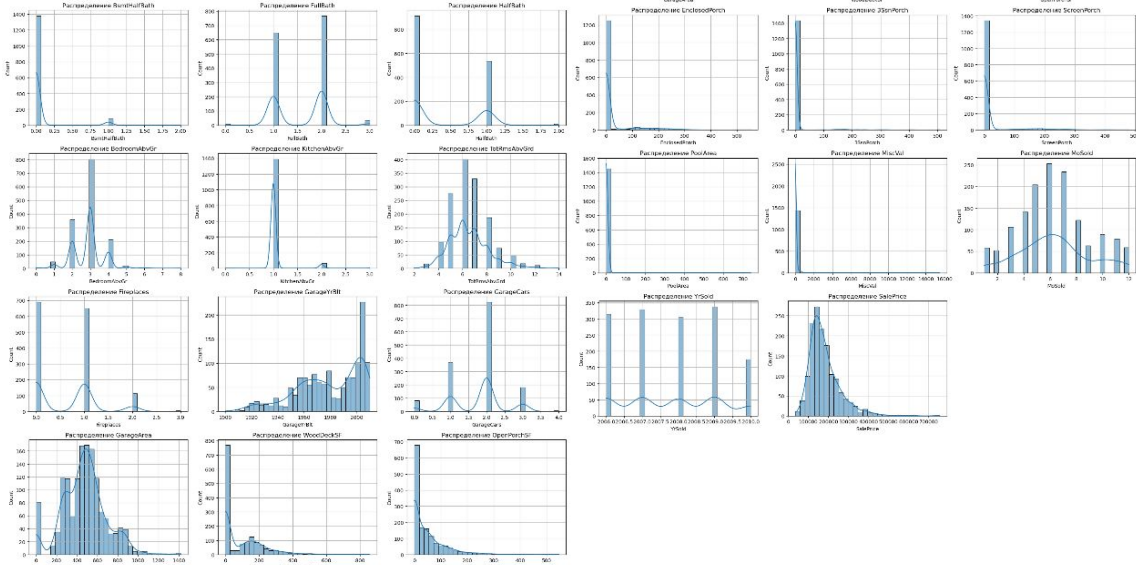
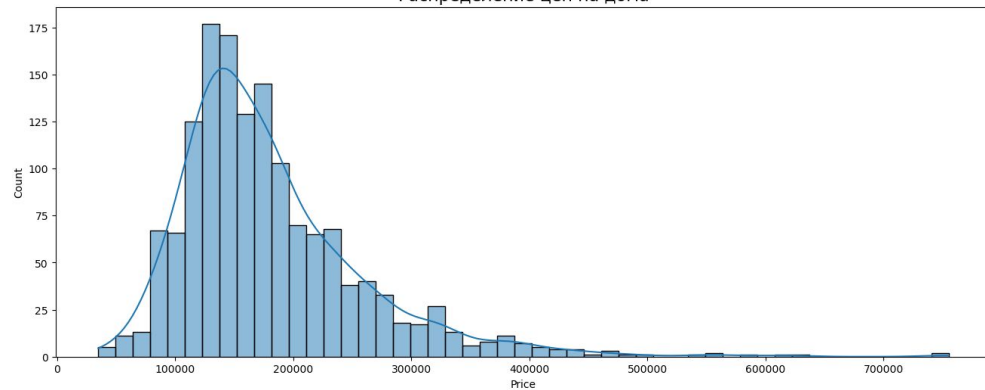
Работа команды

Укротители грифонов

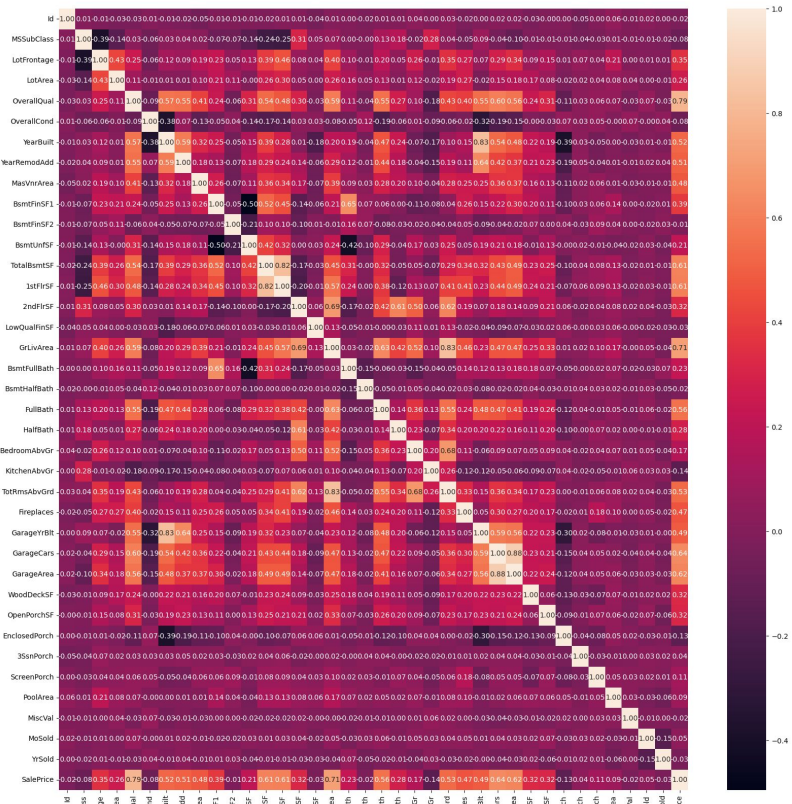
Визуализация данных



Распределение цен на дома



Построение корреляционной матрицы



Изучение корреляции числовых признаков между собой и с таргетом SalePrice

Построение PipeLine: сборка Preprocessor

my_imputer

drop: колонки с наибольшим количеством пропусков, близких к количеству ID

cat_imputer: заполнение модой категориальных признаков

num_imputer_0: заполнение логически ценных NaN у категориальных признаков на «0»

num_imputer_mean: заполнение NaN средним значением у числовых признаков

Построение PipeLine: сборка Preprocessor

scaler_and_encoder

ordinal_encoding: порядковое кодирование колонок, содержащих всего 2 качественно разных значения

one_hot_encoding: one-hot кодирование колонок, содержащих >2 качественно разных значения

standard_scaling: нормализация колонок, содержащих только числовые значения

Построение PipeLine: сборка Preprocessor

Preprocessor

my_imputer

drop: колонки с наибольшим количеством пропусков, близких к количеству ID

cat_imputer: заполнение модой категориальных признаков

num_imputer_0: заполнение логически ценных NaN у категориальных признаков на «0»

num_imputer_mean: заполнение NaN средним значением у числовых признаков



scaler_and_encoder







ordinal_encoding: порядковое кодирование колонок, содержащих всего 2 качественно разных значения

one_hot_encoding: one-hot кодирование колонок, содержащих >2 качественно разных значения

standard_scaling: нормализация колонок, содержащих только числовые значения

Построение PipeLine: выбор Модели

Ключевые метрики по моделям:

	rmsle	r2-score	mae
CatBoost 	0.13285121025883634	0.9109470568833798	15651.608215766824
LightGBM 	0.1358939273870469	0.8910645005327766	16660.285262189802
RandomForestRegressor 	0.15268260654247315	0.8900503769471245	17473.79624785615
VotingRegressor  	0.1393873740559207	0.9083578148584984	16090.453002187041
XGBoost 	0.12457236347382572	0.9210239231788393	15134.387681934932

Загрузка в Kaggle

1170 укротители грифонов



0.13294

1

37s



Your First Entry!
Welcome to the leaderboard!

Выявление мало полезных признаков

С помощью Lasso

Heating_Floor	70.38065036368731
Neighborhood_SawyerW	-50.427623187224924
OpenPorchSF	-39.70202916296212
random	22.6910767711588
GarageFinish_Fin	-17.68154799245815
BsmtFinType2_Rec	7.280567324909044
ExterQual_TA	-3.0557219101408455
BsmtExposure_0	0.47786286138631456
KitchenQual_TA	-0.11142436714658872
GarageQual_Gd	0.0
Neighborhood_SWISU	-0.0
Electrical_Mix	0.0
RoofMatl_Membran	0.0
GarageQual_0	-0.0
Condition2_RRNn	0.0
HouseStyle_SFoyer	-0.0
GarageCond_Fa	0.0
PavedDrive_Y	0.0
LotConfig_Inside	-0.0
RoofMatl_WdShake	-0.0
BsmtCond_TA	-0.0
GarageCond_0	-0.0

Выявление мало полезных признаков

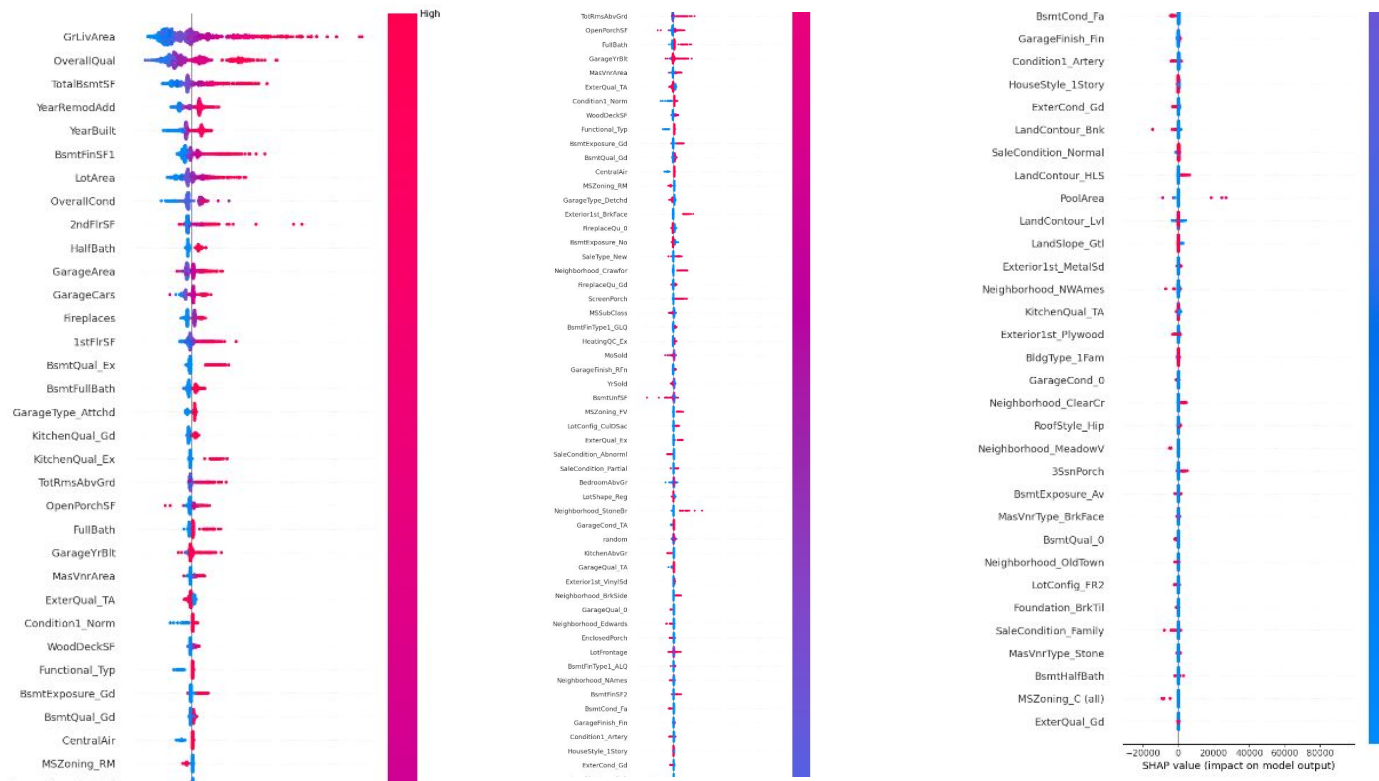
С помощью Feature_importance

Condition2_Feedr	1.2678554695126626e-06
RoofStyle_Mansard	1.2600973332729432e-06
GarageCond_Gd	1.1729645067415973e-06
SaleType_Con	8.975004590011556e-07
RoofMatl_Metal	8.886649961313337e-07
GarageCond_Ex	7.828532755089154e-07
Exterior2nd_Other	6.1597531723344e-07
SaleType_ConLI	5.384196586618439e-07
Neighborhood_NPkVill	5.191635251175855e-07
SaleType_Oth	4.806939058065034e-07
HouseStyle_2.5Fin	4.407973646610428e-07
Foundation_Wood	3.2061026339749854e-07
Street	2.8718273281540237e-07
Heating_Floor	2.556653475131235e-07
Condition1_RRNe	1.9254836483683256e-07
GarageQual_Po	1.8746136752486854e-07

Exterior1st_Stone	1.8731498270175943e-07
ExterCond_Po	1.463706451945681e-07
Condition2_Artery	1.4381935680172687e-07
Exterior1st_CBlock	1.2093292540878965e-07
Condition2_PosN	6.964821439934181e-08
RoofMatl_Roll	5.8117051092225685e-08
Condition2_RRAe	1.8698524845969128e-08
Condition2_RRAn	1.0470839445948454e-08
Condition2_RRNn	6.055519843208459e-09
Utilities	6.033113558853112e-09
BsmtCond_Po	8.463112287769236e-10
Exterior2nd_CBlock	0.0
Condition2_PosA	0.0
Electrical_Mix	0.0
RoofMatl_Membran	0.0

Выявление мало полезных признаков

С помощью Shap



Исправление PipeLine: Preprocessor

Новый Preprocessor

my_imputer

drop: колонки с наибольшим количеством пропусков, близких к количеству ID

cat_imputer: заполнение модой категориальных признаков

num_imputer_0: заполнение логически ценных NaN у категориальных признаков на «0»

num_imputer_mean: заполнение NaN средним значением у числовых признаков

scaler_and_encoder

ordinal_encoding: порядковое кодирование колонок, содержащих всего 2 качественно разных значения

one_hot_encoding: one-hot кодирование колонок, содержащих >2 качественно разных значения

standard_scaling: нормализация колонок, содержащих только числовые значения

after_scale_and_encode_dropper

drop: колонки после кодирования и скейлинга, признанные бесполезными после Lasso, Feature_importance и Shap

Исправление PipeLine: выбор Модели

Ключевые метрики по моделям:

	rmsle	r2-score	mae
CatBoost 🐱	0.13604478192205685	0.9151278827821836	15462.332801804992
LightGBM 👤	0.1351028539811035	0.8920988684533254	16723.571508622597
RandomForestRegressor 🌲 🌲 🌲	0.15196710739875025	0.8888334395615263	17466.302191502808
VotingRegressor ✓ ✗	0.1379342753442541	0.9026725461093361	16322.919371297341
XGBoost 📈	0.12387645214321104	0.9241271951621555	14868.976228060788

Финальная загрузка в Kaggle

1127

укротители грифонов



0.13273

2

7s



Your Best Entry!

Your submission scored 0.13396, which is not an improvement of your previous score. Keep trying!

А теперь портал в Streamlit от Ильи!

