**Задача:** **Решить соревнование на kaggle** - <https://www.kaggle.com/c/house-prices-advanced-regression-techniques/overview>

**Обзор задания:**

You have some experience with R or Python and machine learning basics. This is a perfect competition for data science students who have completed an online course in machine learning and are looking to expand their skill set before trying a featured competition.

**Competition Description**

Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this playground competition's dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence.

With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, this competition challenges you to predict the final price of each home.

**Practice Skills**

Creative feature engineering

Advanced regression techniques like random forest and gradient boosting

**Acknowledgments**

The [Ames Housing dataset](http://www.amstat.org/publications/jse/v19n3/decock.pdf) was compiled by Dean De Cock for use in data science education. It's an incredible alternative for data scientists looking for a modernized and expanded version of the often cited Boston Housing dataset.

**Решение:** Использовал градиентный бустинг catboost. Основная работа заключалась в выборе и подготовке признаков. Исследовал распределения признаков и их корреляцию с целевым вектором, отобрал важные признаки. Исследовал признаки на выбросы – для работы с выбросами использовал Boxcox нормализацию. Обработал категориальные признаки на предмет пропусков. Пост результат: получил оценку важности каждого признака (библиотека shap) и выгрузил результат в csv.

**Результат:** На тестовой выборке kaggle показал ошибку на уровне 0,14 – это средний результат.

**Stack:** Python, Jupiter, pandas, catboost, sklearn, scipy