В целом своим результатом не очень доволен. На Kaggle результат 30.0. При этом это получилось в начале разработки на минимальных изменениях. После этого обработка признаков и создание новых только ухудшало ситуацию. Я понял, что просто надо начинать все с начала, но уже не успевал.

Работали вместе с Сергеем Пинаевым. Мы с ним делали еще предыдущее задание («Кмпьютер говорит нет»). Ну как вместе – обсуждали идеи и подходы. Модели делали отдельно. Должен признать – он предоставил мне спарсеные данные. Я из Киева и у нас весь Yandex.ru забаненый. Парсить через VPN было сложно, долго да и пока не знаю как.

Разделил модель на несколько файлов. Два файла (my\_functions.py mainCode.py) – это подготовка модели. Выделил так, чтобы потом можно было играться разными подходами на одинаковой базе и вносить изменения в одном месте.

Структура модели:

**my\_functions.py** – обработка признаков, создание новых, чистка модели от ненужных признаков, расчет метрик. В начале было намного больше шагов в def putting\_order. Но когда начал играться с разными моделями остались только эти.

**mainCore.py** – основные шаги по подготовке файла. То же, что и выше. Это то, что осталось после работы с разными моделями.

Далее модели:

**Module\_6\_Naive.py** – ключевые базовые модели регресии. Дали плохие результаты. MAPE в районе 83%

**Module\_6\_bagging.py -** бэггинг на базе BaggingRagressor с дефолтным базовым алгоритмом. Подставлял в базовый алгоритм LinearRegression , Lasso, Ridge. Получился чистый ужас. Думаю, что учитывая эти модели в базе (**Module\_6\_Naive.py**) дали плохой результат то бэггинг его только усилил. Отрабатывает долго.

**Module\_6\_boosting.py –** после обсуждения с Сергеем и просмотра кода на Kaggle применил CatBoosting. Тут получил по МАРЕ лучший результат – 17,17%. Но загрузка submission на Kaggle мой рейтинг не улучшило.

**Module\_6\_stacking.py** – после хорошего результат с бустингом от CatBoost решил засунуть его в стакинг. МАРЕ сильно лучше не улучшилась – до 17.00%. Зато R2 выросла до 0.835. Но работает долго.

**Module\_6\_review.ipynb** – визуализация признаков, просматривал распределение данных.