Ярослав, здравствуйте!

Прежде всего, хотелось бы отметить отличное использование графических пакетов визуализации. Особенно понравился анализ распределения данных.
Могу сказать, что всё остальное тоже выглядит достойно. Отлично поработали над проектом! Ниже распишу главные моменты.

Что сделано хорошо:
- В работе подготовлено много новых признаков, которые подготовлены на основе существующих данных. Очень подробно всё разобрано.
Для многих признаков реализованы функции для обработки данных. Функции отлично описаны. Многие пункты сопровождаются подробными комментариями.
- Используется кодировка признака на основе TargetEncoder(). Как вариант, можно было немного проанализировать работу OneHotEncoder() и BinaryEncoder()
при создании дополнительных признаков. Отмечу, что понимание принципов кодирования присутствует.
- Отмечаю, что для улучшения итоговой метрики была выбрана верная стратегия. Правда, тему можно развивать и дальше.
Один из лучших вариантов для данной задачи является обработка отзывов. Была использована специальная библиотека,
которая позволяют получить много дополнительной информации. Использование библиотеки nltk и класс SentimentIntensityAnalyzer().
Дополнительно можно было использовать векторизацию слов.
- Выполнен отбор признаков, анализ мультиколлинеарности (.corr) и значимости признаков (f_classif). Выполнен анализ на основе оценки по Пирсону и Спирману.
Результаты анализа для наилучшего восприятия были представлены в виде графиков.
- Для подготовки и обучения модели используется фреймворк lightautoml. Всё-таки нужно было ограничиться работой с RandomForestRegressor из библиотеки sklearn.
- Полученная модель достаточно хорошо позволяет улучшить значение MAPE для базового шаблона (baseline).

На момент проверок работ результат входит в ТОП-10! Считаю, что это достижение.

Что можно улучшить:
- Проект достаточно очень хорошо подготовлен. Всё-таки есть небольшая рекомендация.
Существует такое понятие, как векторизация слов.
Каждое слово может быть представлено отдельным числом. Для этого можно использовать библиотеку gensim. Использование класса Doc2Vec.
В итоге получите множество новых полезных признаков, которые созданы на основе отзывов.

Работу завершена и может быть оценена максимальным баллом! Все пункты очень хорошо представлены. Считаю, что работа получилась образцовой. Удачи в следующих блоках курса!

Проверку выполнил ментор Сергей Добдин. По любым вопросам, пожалуйста, пишите в общий канал slack [#03_project-3].

Запись итогового вебинара:
https://us02web.zoom.us/rec/share/qoBLJFQpQHhkoaFS77dl0EkWRRvy226xr1pHO7b7Cz6cwJN8Y148vaJQ-DDCEEz6.aFGotJSRNbS59tUP?startTime=1663866000000 Код доступа: +5s72pfL