**Степан, здравствуйте! Получился отличный проект, который, безусловно, заслуживает высокий итоговый балл. Видно, что все блоки исследовательского анализа были старательно продуманы. Уверен, что на подготовку было потрачено много времени. Мне понравилась данная работа по нескольким причинам. Во-первых, сразу же виден авторский подход и большое количество оригинальных идей. Во-вторых, очень нравится, когда многие блоки сопровождаются подробным описанием. Выглядит работа так, как если бы её готовил специалист с большим опытом работы в данной области. Ещё очень порадовало использование собственных функций. Очень сильный проект, который по праву можно отнести к наиболее удачным решениям. Обсудим достоинства и замечания по работе в данном отзыве. Что сделано хорошо: - Общее впечатление о проекте. Каждый раздел хорошо представлен и раскрыт. Отлично, что в самом начале проекта определено рабочее пространство и подготовлены (объявлены) новые функции. Мне, например, очень нравится, когда для оптимизации решений используются функции. Достаточно полезный подход. Старайтесь использовать его и в будущем. Полнота исследований и результаты не вызывают сомнений. Большая часть выводов и рассуждений правильная. Хорошо, что в работе использованы комментарии. - Создание новых признаков. Можно сказать, что данная часть выполнена лучше остальных. Безусловно, можно выделить анализ дат, тегов, отзывов и других признаков выполнен не хуже. Обычно я рекомендую плюсом использовать дополнительные данные (которых нет в исходном датасете), но и этот пункт вам удалось закрыть. Был выполнен расчёт расстояния от центра до отеля. Если придумать что-то с парсингом данных, то получится отличное дополнение. - Использование ранее неизвестных библиотек. Сразу же отмечаю подробный анализ текстовых полей с отзывами, что для данной задачи можно назвать лучшей стратегией для поиска новых признаков итоговой модели. Выполнен анализ слов, которые позволяют негативно или положительно оценить контекст отзыва. Используется пакет библиотек и программ для символьной и статистической обработки естественного языка (nltk, afinn и textblob), что можно назвать best practice. Удивительно, что для анализа используется несколько библиотек. Можно было ограничиться использование только одной библиотеки. Возможно, в будущем вы будете углубленно изучать NLP. - Кодирование признаков выполнено с пониманием базовых принципов кодирования. Правда, не нашёл в проекте использование бинарного кодировщика. Он очень полезен во многих решениях. Пожалуйста, обратите на это внимание. - Выполнен отбор признаков и анализ мультиколлинеарности (.corr) и значимости признаков (f\_classif). Дополнительно представлены графики, позволяющие интерпретировать результаты анализа. Для оценки распределения использованы статистические тесты. - Полученная модель позволяет достаточно хорошо улучшить значение MAPE для базового шаблона (baseline). Участие в соревновании принято. Не хватает решения на github. Об улучшении метрики можно узнать после просмотра итогового вебинара. Что можно улучшить: - Улучшение MAPE. Для улучшения метрики можно было воспользоваться векторизацией слов. При векторизации, например, каждое слово может быть представлено отдельным числом. Для этого можно использовать библиотеку gensim (Doc2Vec). В итоге получите много новых полезных признаков, которые созданы на основе отзывов. Более подробно, например, можно посмотреть в книге [Б. Бенгфорт, Р. Билбро, Т. Охеда Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. СПб.: Питер, 2019. 368 с.] - Степан, проект нужно сохранить на github в своём аккаунте. Это нужно не только для анализа решений ментора, но и для подготовки будущего портфолио. Обязательно добавьте проект в свой репозиторий и настройте его согласно всем правилам. Есть способ трансляции блокнота kaggle на github:** <https://www.kaggle.com/product-feedback/295170> **Проект завершён и его можно оценить максимальным баллом! Интересный проект с большим содержанием полезных решений. Обработка данных на высоком уровне. Считаю, что проделана очень кропотливая работа. Замечаний нет. Искренне желаю удачи в будущих блоках курса. Будет интересно! Проверку выполнил ментор Сергей Добдин. По любым вопросам, пожалуйста, пишите в общий канал Пачки [DSPR/DST/ЦП\_DST\_м\_03\_project\_3]. Запись итогового вебинара:** <https://us02web.zoom.us/rec/share/qoBLJFQpQHhkoaFS77dl0EkWRRvy226xr1pHO7b7Cz6cwJN8Y148vaJQ-DDCEEz6.aFGotJSRNbS59tUP?startTime=1663866000000> **Код доступа: +5s72pfL**