**README**

**→ основные цели и задачи проекта:**

* очистить довольно сильно загрязнённый датасет, добавить новые признаки и подготовить данные для обучения модели, которая должна будет предсказывать рейтинг ресторана по данным сайта TripAdvisor на основе имеющихся в датасете данных; принять участие в соревновании по машинному обучению на Kaggle и поработать с метрикой МАЕ.

**→ краткая информация о данных:**

* датасет с информацией о 40000 ресторанах состоит из 10 столбцов, из которых только 3 являются числовыми признаками (Ranking, Number\_of\_ Reviews, Rating), остальные 7 необходимо преобразовать,'Reviews' и 'Cuisine\_Style' нужно преобразовать в списки. В 'Cuisine\_Style','Price\_Range','Number\_of\_Reviews' большое количество пропусков, которые нужно заполнить.
* Описание датасета:  
    
  Restaurant\_id — идентификационный номер ресторана / сети ресторанов  
  City — город, в котором находится ресторан  
  Cuisine Style — кухня или кухни, к которым можно отнести блюда, предлагаемые в ресторане  
  Ranking — место, которое занимает данный ресторан среди всех ресторанов своего города Rating — рейтинг ресторана по данным TripAdvisor (именно это значение должна будет предсказывать модель)  
  Price Range — диапазон цен в ресторане  
  Number of Reviews — количество отзывов о ресторане; Reviews — данные о двух отзывах, которые отображаются на сайте ресторана  
  URL\_TA — URL страницы ресторана на TripAdvisor  
  ID\_TA — идентификатор ресторана в базе данных TripAdvisor

→ **этапы работы над проектом:**

* очистить данные
* создать модель, основанную на алгоритмах машинного обучения
* соревновании на *kaggle*

→ **ответы на вопросы саморефлексии:**

**Какой частью своей работы вы остались особенно довольны?**

я создала первую модель! Понравилось, как я заполнила пропуски, что я попыталась создать несколько версий, поиграть с параметрами, добавить признаки, улучшить результат предсказания

**Что не получилось сделать так, как хотелось? Над чем ещё стоит поработать?**

работа с дубликатами, визуализация текста и работа с WordCloud, работа с Review и такими списками, не совсем поняла, как считается MAE, запуталась в версиях

**Что интересного и полезного вы узнали в этом модуле?**

познакомилась с Kaggle, поучаствовала в соревновании, узнала, что такое MAE

**Что является вашим главным результатом при прохождении этого проекта?**

удалось предложить несколько версий модели

**Какие навыки вы уже можете применить в текущей деятельности?**

добавление признаков, dummies

**Планируете ли вы дополнительно изучать материалы по теме проекта?**

хочу попробовать поучаствовать в других соревнованиях