

Проект: Вторая часть

Финальный документ – jupyter notebook с подробными комментариями и кодом + файл с данными.

По любым вопросам, связанным с заданием, можно писать мне в телеграм @anastasiyamaxx

Используя данные собранные в любом проекте из раздела по сбору данных проделайте следующие шаги:

Шаг 2. Предварительная обработка.

- Проанализируйте столбцы. Что в них содержится? Есть ли пропуска и где?
- Обработать пропуски. Если в какой-то колонке пропусков большинство, то ее можно удалить. Сами пропуски обработать: решите удалять их или нет, заполнить медианой/средним или чем-нибудь еще, напишите почему Вы выбрали тот или иной вариант. За это +5 баллов.

Шаг 3. Создание новых признаков.

- Иногда на основании имеющихся признаков можно создать новые. Например, из столбца с датой можно достать: является ли день рабочим, является ли месяц концом квартала, является ли день национальным праздником и т.д. Создавайте признаки, которые логически вписываются в Ваш набор данных. Максимальный балл – до 25 баллов, в зависимости от того сколько признаков и насколько они логически вписываются в исследование.

Шаг 4. Визуализация

- Ваша задача – построить красивые и интересные визуализации, на основании которых можно выдвинуть гипотезы. Максимальный балл – 30, зависит от количества графиков, сложности их построения и интересных идей, которые можно из них заметить. Помимо графиков, построенных на семинаре, можно поискать идеи [в галерее библиотеки matplotlib](#).