

Гребенщиков Данила Александрович

Отчет по работе с данными на хакатоне GPN, трек: Data Science

Сначала я обработал данные, избавился от ненужных значений, а также заменил или вовсе удалил пустые ячейки.

Таблицу `attack_probability` считал, как таблицу с вероятностью нападения той или иной группы на трактирщиков по городу, то есть на их бары. Убрал значения и даты нападения негативных групп, у которых вероятность нападения была меньше 0,05.

После этого я отобрал трактиры, на которые вероятнее всего не напали негативные группы. Сгруппировал их, вывел суммы пользы литров по группе связанными с датой, городом и типом продукта – алкогольным или безалкогольным.

Отобрал по городам, а также времени проведения акций литры выгоды. При этом высчитал те литры, где была выгода в общей сумме по сравнению с тем, что выпили негативные группы. После данного отбора, посчитал суммы для всех акций, то есть их выгоду, учитывая нападение негативной группы (вычел из литров, принесенный по акции, значение выпивки негативной группы). Литры, выпитые негативными группами, считаю убытком.

Итоговый файл `promo_effects.csv` содержит информацию по выгоде от акций для трактиров с учетом нападений на трактиры негативных групп.