Файл – датасет.csv

Разделитель – запятые.

Заголовки – да.

Количество строк - 3333

Описание данных:

В файле представлена выборка из 3333 абонентов сотовой связи и данные по пользованию услугами связи за один месяц.

|  |  |
| --- | --- |
| Колонка | Описание |
| state | Штат |
| account length | длина учетной записи |
| area code | Код зоны |
| phone number | Номер телефона |
| international plan | Международный тарифный план |
| voice mail plan | Тарифный план голосовой почты |
| number vmail messages | Количество сообщений голосовой почты |
| total day minutes | Общее количество использованных минут в день |
| total day calls | Сделанные дневные звонки |
| total day charge | Общая дневная плата |
| total eve minutes | Общее количество вечерних минут |
| total eve calls | Общее количество вечерних звонков |
| total eve charge | Общая вечерняя плата |
| total night minutes | Общее количество ночных минут |
| total night calls | Общее количество ночных звонков |
| total night charge | Общая плата за ночь |
| total intl minutes | Общее количество использованных международных минут |
| total intl calls | Общее количество международных звонков сделано |
| total intl charge | Общая международная плата |
| customer service calls | Количество сделанных звонков в службу поддержки клиентов |
| churn | Отток (да -1, 0 - нет) |

Задачи -

1. Определить жители какого штата говорят по телефону больше всего? меньше всего?
2. Верно ли, что люди, совершающие много международных звонков, приобретают international plan?
3. Как связаны количество обращений в колл-центр и Оттока?
4. Какие из признаков имеют наибольшую степень корреляции?
5. Построить несколько моделей предсказания оттока на основе данных. Выбрать из них лучшую. Разделение на тренировочную и тестовую выборки 60:40.
6. Показать значимость признаков в модели.

Результирующий файл – ipynb.

Deadline – 23:59 12.05.2022