Кейс “Телекоммуникационная компания Telco”

Описание кейса: прогнозирование оттока клиентов.

Описание переменных

- `gender`: пол

- `Senior Citizen`: пенсионер

- `partner`: имеет ли супруга / супругу

- `dependents`: имеет ли детей

- `tenure`: продолжительность использования услуг в месяцах

- `PhoneService`: услуги телефонии

- `Internet Service`: услуги интернета

- `Online Security`: безопасность от вирусных сайтов

- `Online Backup`: облачное хранилище файлов для резервного копирования данных

- `Device Protection`: антивирус

- `TechSupport`: техническая поддержка

- `Streaming TV`: стриминговое телевидение

- `Streaming Movies`: каталог фильмов

- `Contract`: формат оплаты услуг: ежемесячно, раз в год

- `Paperless Billing`: электронная оплата

- `Monthly Charges`: размер ежемесячных платежей

- целевой признак: `Churn`

Задания:

(i) построение сводных таблиц

(ii) построение столбиковых диаграмм

(iii) построение коробчатых диаграмм

(iv) построение плотностных гистограмм

(v) построение тепловых диаграмм

(vi) оценка логистической регрессии

Вопросы для обсуждения:

1. как библиотека pandas помогает в загрузке базы данных и определении основной информации выборки?
2. как определить процентную долю признаков (например пол клиентов)?
3. можно ли построить столбиковые диаграммы через библиотеку matplotlib?
4. как создать сводную таблицу?
5. как построить функцию, которая строит сводную таблицу?
6. как построить столбиковую диаграмму через библиотеку seaborn?
7. можно ли вводить дополнительные переменные в цикл построения диаграмм?
8. как построить коробчатые диаграммы?
9. как построить плотностные гистограммы?
10. коробчатая диаграмма через sns?
11. как библиотека pandas создает индикаторные переменные?
12. как построить корреляционную матрицу?
13. как выделить признаки массива данных и целевой вектор?
14. как оценить логистическую регрессию?
15. как проверить точность оценок логистической регрессии?
16. как выделить самые значимые переменные из корреляционной матрицы?