

Лабораторная работа №3

Пакет matplotlib

Рассмотрим датасет, в котором приведены данные о пассажирах с «Титаника» (**titanic.csv**)

0. Провести первичный анализ данных. Предоставить следующую информацию о наборе данных:

- описание набора данных, пояснения, позволяющие лучше понять природу данных.
- описание каждого признака и его тип.
- форма набора данных: количество элементов набора, количество признаков, количество пропущенных значений, среднее значение отдельных признаков, максимальные и минимальные значения отдельных признаков и прочие показатели.
- наличие/отсутствие пустых значений; при наличии пустых значений, необходимо от них избавиться

1. Найти количество наблюдений для каждого значения целевой переменной Survived и применить к полученным данным метод plot, чтобы построить столбчатую диаграмму.

2. Построить такую же диаграмму, используя логарифмический масштаб.

3. Построить две гистограммы по значениям признака Pclass - одну для выживших (Survived равен 1) и другую - нет (Survived равен 0).

4. Подобрать значение аргумента density так, чтобы по вертикали графика было расположено не число наблюдений, а плотность распределения. Число бинов должно равняться 20 для обеих гистограмм, а коэффициент alpha равным 0.5, чтобы гистограммы были полупрозрачными и не загромождали друг друга. Создать легенду с двумя значениями: “Survived 0” и “Survived 1”. Гистограмма выживших должна быть серого цвета, а невыживших - красного. Название горизонтальной оси – “Survived”.

5. Аналогично изобразить гистограмму выживаемости относительно пола пассажиров

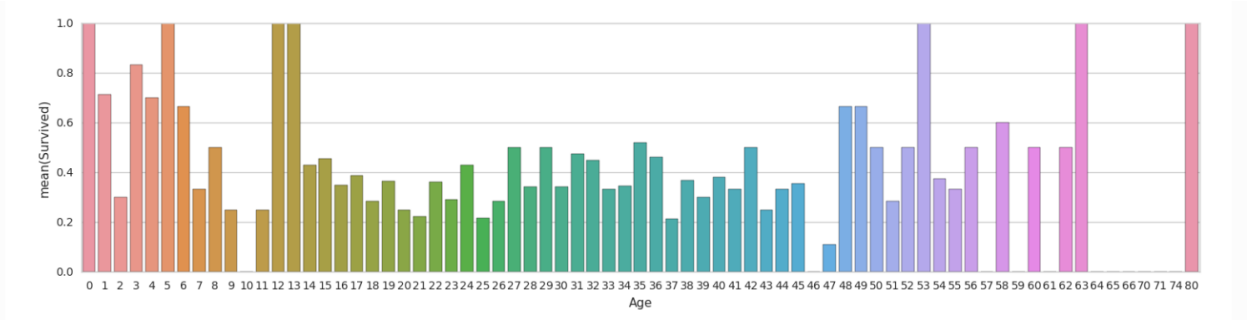
6. Исследовать зависимость выживаемости пассажиров от количества родственников на борту (признак Parch). Построить график, показывающий процент выживших в зависимости от количества родственников

7. Исследовать распределение стоимости билетов для пассажиров разных возрастных групп. Построить ящик с усами, где по горизонтальной оси будут возрастные группы, а по вертикальной - стоимость билетов.

8. Определить процент выживших пассажиров по возрастным группам (например, 0-9 лет, 10-19 лет и т.д.) и отобразить результаты на круговой диаграмме.

9. Получить следующие диаграммы

а)



б)

