Задание на практику Часть 1 Срок сдачи 23.11.

**Первичный анализ**

1. Загрузите данные о выживаемости пассажиров Титаника из файла titanic\_train.csv

2. Представьте данные в виде таблицы. Посмотрите на первые 5 строк.

3. Выведите на экран основную информацию (info) о наборе данных и признаках.

4. Выведите описательные статистики (медиану, среднее, квартили, минимальное, максимальное значения, дисперсию) массива данных в виде таблицы (describe).

6. Избавьтесь от пустых значений, заменив их на медианные значения (для количественных признаков) и на наиболее часто встречающееся (для категориальных). (используйте value\_counts() для определения наиболее часто встречающегося))

7. Выведите на экран сколько мужчин и сколько женщин находилось на борту?

8. Выведите распределение переменной Pclass по всем классам (социально-экономический статус) и это же распределение, только для мужчин / женщин по отдельности. Сколько было мужчин 2-го класса?

9. Каковы медиана и стандартное отклонение платежей (Fare)? Округлите до 2 десятичных знаков.

**Визуализация**

10. Постройте попарные зависимости признаков Age, Fare, Pclass, Sex, SibSp, Parch, Embarked и Survived. (метод scatter\_matrix Pandas или pairplot Seaborn).

11. Постройте гистограммы распределения каждого признака (столбца данных).

12. Как плата за билет (Fare) зависит от класса каюты (Pclass)? Постройте boxplot.

13. Каково соотношение погибших и выживших в зависимости от пола? Отобразите c помощью Seaborn.countplot c аргументом hue.

14. Каково соотношение погибших и выживших в зависимости от класса каюты? Отобразите c помощью Seaborn.countplot c аргументом hue.

15. Как факт выживания зависит от возраста пассажира? Проверьте (графически) предположение, что молодые чаще выживали. Пусть, условно, молодые - младше 30 лет, пожилые – старше 60 лет.

16. Постройте график рассеяния на осях Age и Fare. Cиним отметьте пассажиров, которые не выжили (Survived = 0) и красным — выживших (Survived = 1).

**Machine Learning**

17. Разбить данные на обучающую и проверочную выборки

18. Построить графики целевой переменной Survived

19. Построить корреляционные матрицы переменных (таблица/график)

20. Обучить 3 модели классификаторов и выбрать лучшую по точности (целевая переменная - Survived). 21. Предсказать Survived.