МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования

НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ

УНИВЕРСИТЕТ им. Р.Е.АЛЕКСЕЕВА

Институт радиоэлектроники и информационных технологий

Кафедра информатики и систем управления

Лабораторная работа 1

"Предварительная обработка данных" (1 Вариант)

ОТЧЕТ

по лабораторной работе

по дисциплине

Машинное Обучение

РУКОВОДИТЕЛЬ:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Гай В.Е

(подпись)

СТУДЕНТ:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Агалаков А.С.

(подпись)

15-АС

Работа защищена «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_

С оценкой \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Нижний Новгород, 2018

**Цель лабораторной работы:**

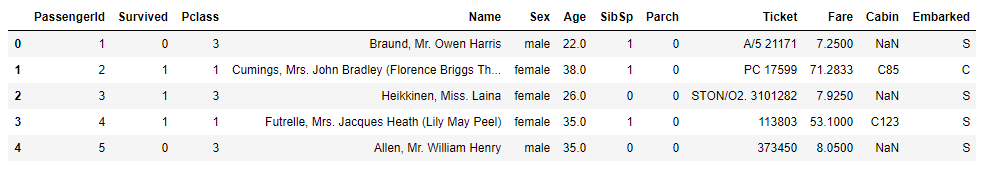
Получение навыков работы с библиотеками анализа данных, предварительной обработки данных, средствами визуализации.

Выполнить следующие пункты:

1. Определите количество мужчин и женщин, которые ехали на корабле.
2. Определите какой части пассажиров удалось выжить. Посчитайте долю выживших пассажиров.
3. Какую долю пассажиры первого класса составляли среди всех пассажиров?
4. Какого возраста были пассажиры? Посчитайте среднее и медиану возраста пассажиров.
5. Коррелируют ли число братьев/сестер с числом родителей/детей? Посчитайте корреляцию Пирсона между признаками SibSp и Parch. Оцените значение p-value. Постройте плотность распределения признаков SibSp и Parch.
6. Какое самое популярное женское имя на корабле? Извлеките изполного имени пассажира (колонка Name) его личное имя (FirstName). Попробуйте вручную разобрать несколько значений столбца Name и выработать правило для извлечения имен, а также разделения их на женские и мужские.
7. Коррелирует ли класс, которым ехал пассажир, с выживаемостью?
8. Визуализируйте гистограммы возраста для выживших и не выживших пассажиров. Сделайте выводы. Отобразите данные на одном и нескольких графиках
9. Визуализируйте гистограммы возраста для выживших и не выживших пассажиров по классам. Сделайте выводы.
10. Постройте столбчатую диаграмму количества людей: мужчины, женщины, дети.

**Ход работы:**

Таблица с исходными данными содержит 12 полей:



**PassengerID** - уникальный номер пассажира;

**Survived** - выживший или нет;

**Pclass** - класс пассажира;

**Name** - имя пассажира;

**Sex** - пол пассажира;

**Age** - возраст пассажира;

**Sib Sp** - братьев / сестер на борту корабля;

**Parch** - родителей / детей на борту корабля;

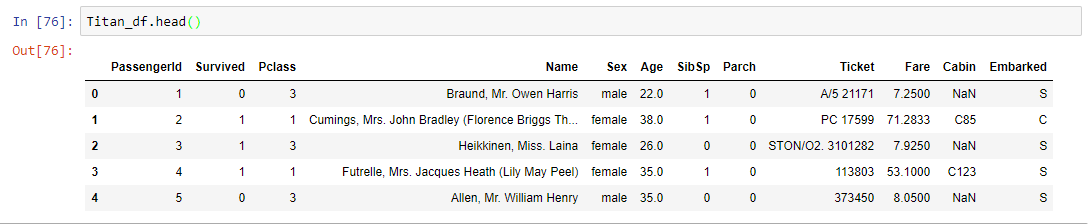
**Ticket** - номер билета;

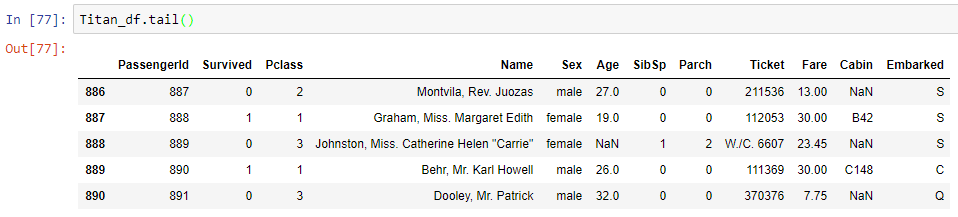
**Fare** - стоимость проезда;

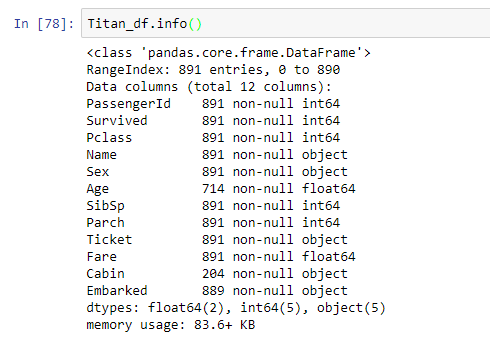
**Cabin** - каюта;

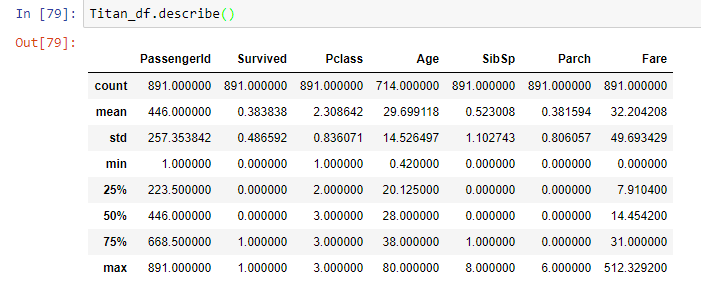
**Embarked** - название порта;

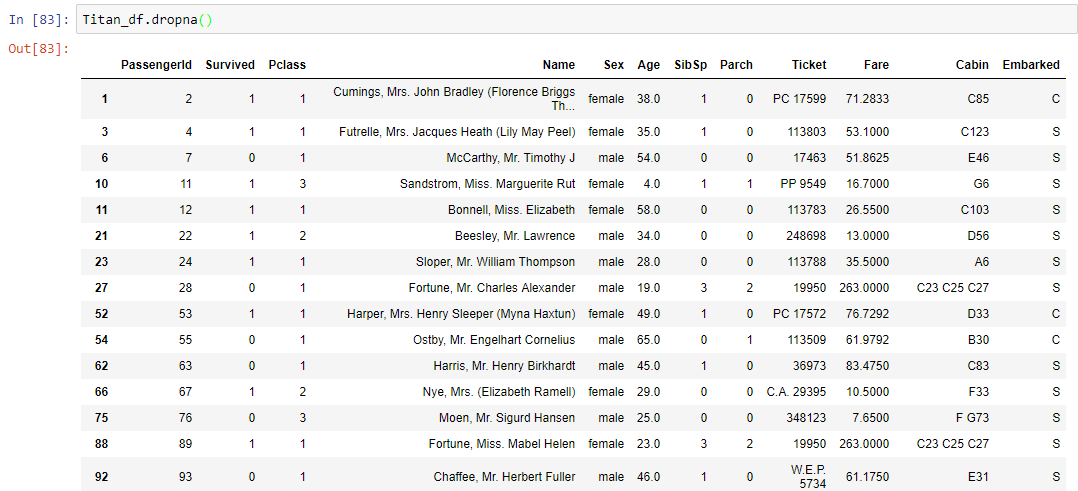
Результат выполнения следующих функций библиотеки Pandas: head, tail, info, describe, dropna, drop\_duplicates, shape:

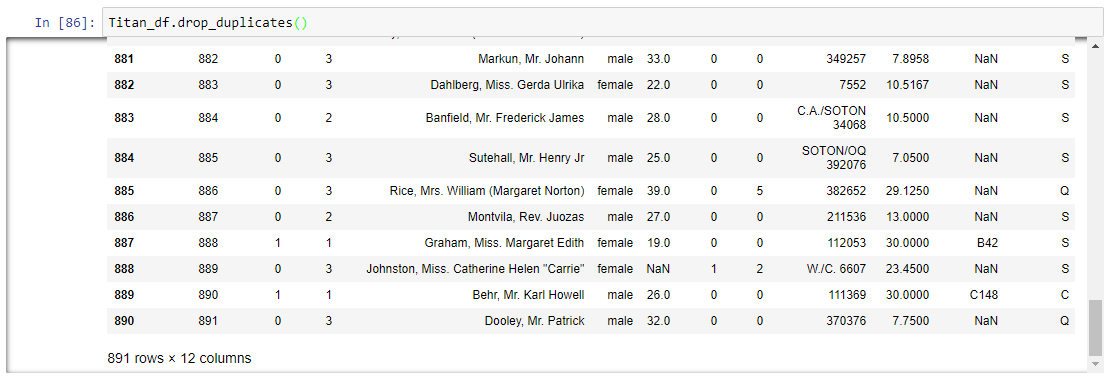


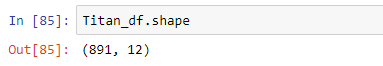












Исходный код приложен в дополнительном файле **Laba1.ipynb**.