

## БДЗ

### «Основы обработки данных на Python»

#### Задания

Рассмотрим пример популярной задачи предсказания выживших пассажиров Титаника. Ознакомьтесь с файлом *train.csv*

Столбец	Описание	Значения
PassengerId	ID пассажира	
Survived	Выжил ли пассажир	0 = Нет, 1 = Да
Pclass	Класс билета	1 = 1st, 2 = 2nd, 3 = 3rd
Name	Имя пассажира	
Sex	Пол	male / female
Age	Возраст (в годах)	
SibSp	Братья / сестры на борту	
Parch	Родители / дети на борту	
Ticket	Номер билета	
Fare	Тариф	
Cabin	Номер каюты	
Embarked	Порт посадки	C = Cherbourg, Q = Queenstown, S = Southampton

Файл *test.csv* содержит аналогичные данные, кроме столбца *Survived*.



### Задание №1

С помощью модуля *pandas* выведите статистику погибших/выживших отдельно для мужчин и женщин в каждом классе (*Pclass*).

### Задание №2

С помощью модуля *pandas* выведите статистику по всем числовым полям, отдельно для мужчин и женщин.

### Задание №3

Определите, влияет ли порт посадки на выживаемость.

### Задание №4

4.1. Выведите топ 10 популярных имён.

4.2. Выведите топ 10 популярных фамилий.

### Задание №5

Заполните все отсутствующие в *train.csv* значения медианой (по столбцу).

### Задание №6

На основе статистики попытайтесь предсказать выживаемость для пассажиров из файла *test.csv*.

### Задание №7

С помощью библиотеки *matplotlib* отобразите гистограмму зависимости возраста (в годах) от выживаемости.