**МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ**

**Ордена Трудового Красного Знамени**

**Федеральное государственное бюджетное образовательное учреждение высшего образования**

**«Московский технический университет связи и информатики»**

Кафедра «Математическая Кибернетика и Информационные технологии»

Дисциплина «Введение в информационные технологии»

Лабораторная работа №1

«Основы обработки данных с помощью Pandas»

Выполнила:

Студентка группы БВТ2303

Морозова Ольга

**Цель работы:**

Познакомиться с основными возможностями библиотеки Pandas для анализа и обработки данных. Научиться выполнять загрузку данных, их очистку, трансформацию, анализ и визуализацию.

**Ход работы:**

1. **Загрузка и первичный анализ данных.**

Импорт нужных библиотек (pandas, numpy, seaborn) и загрузка данных о пассажирах из csv файла и вывод первых 10 строк с верху таблицы.

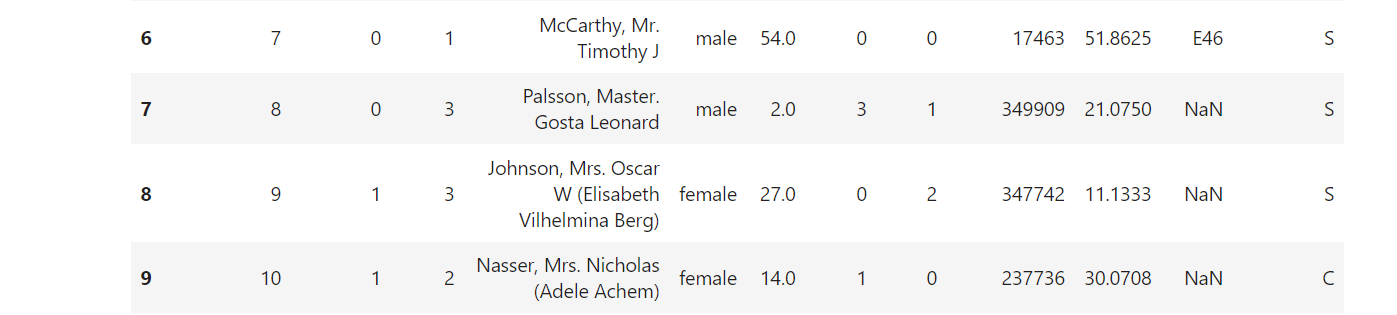
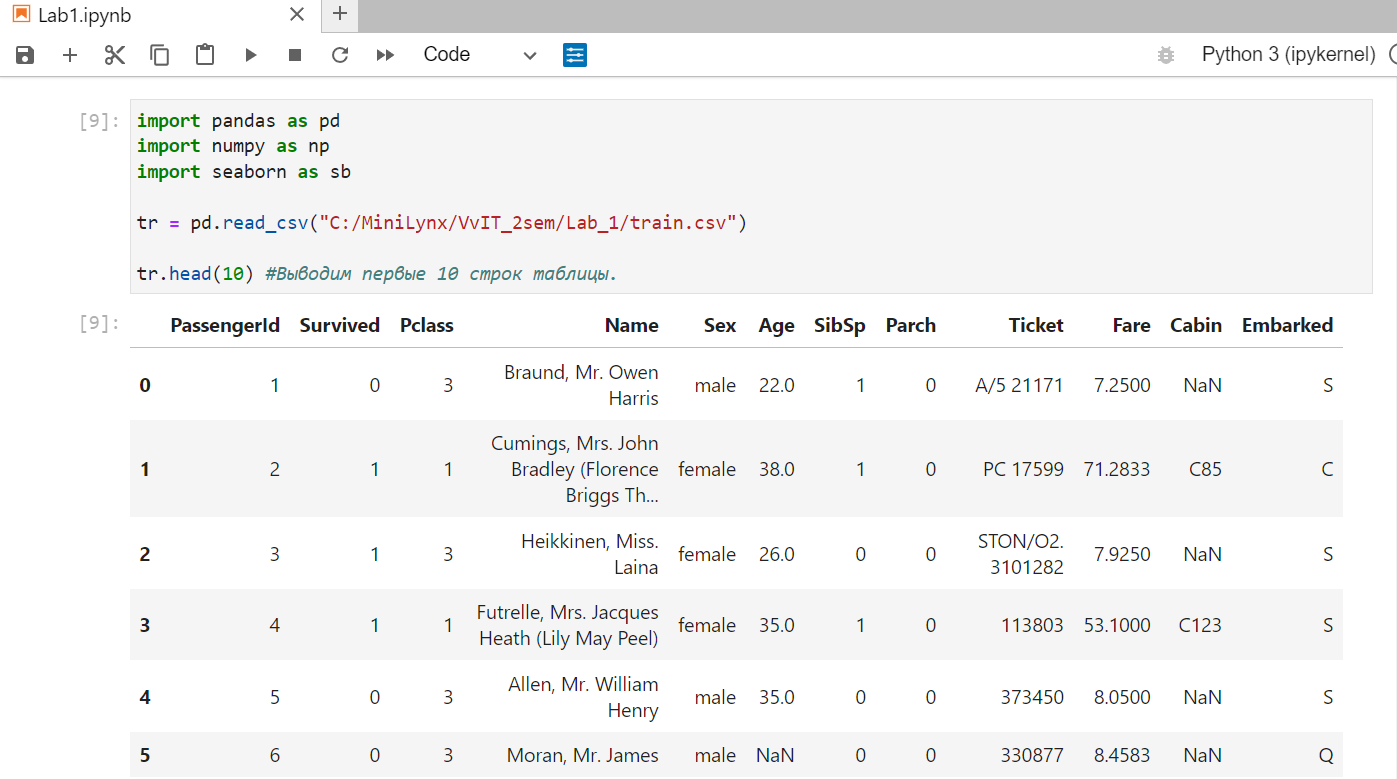
*import pandas as pd*

*import numpy as np*

*import seaborn as sb*

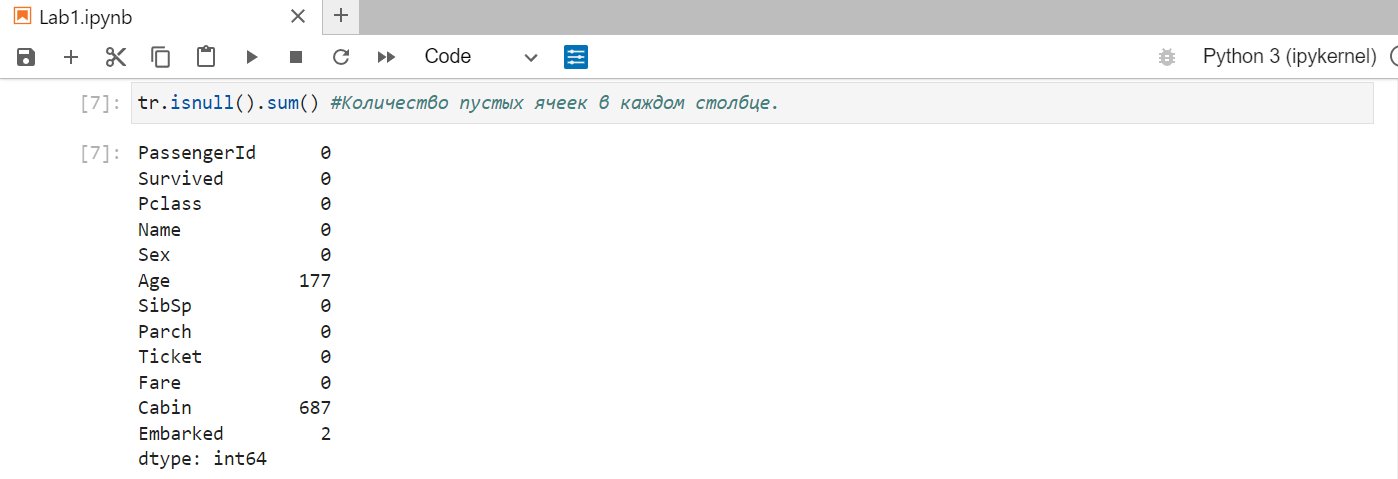
*tr = pd.read\_csv("C:/MiniLynx/VvIT\_2sem/Lab\_1/train.csv")*

*tr.head(10) #Выводим первые 10 строк таблицы.*



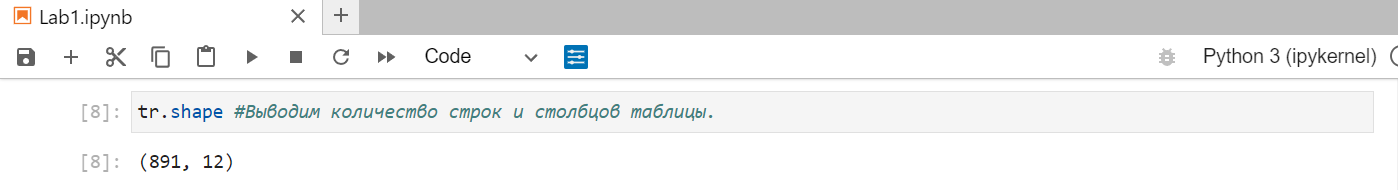
Определение количества пустых ячеек в каждом столбце.

*tr.isnull().sum() #Количество пустых ячеек в каждом столбце.*



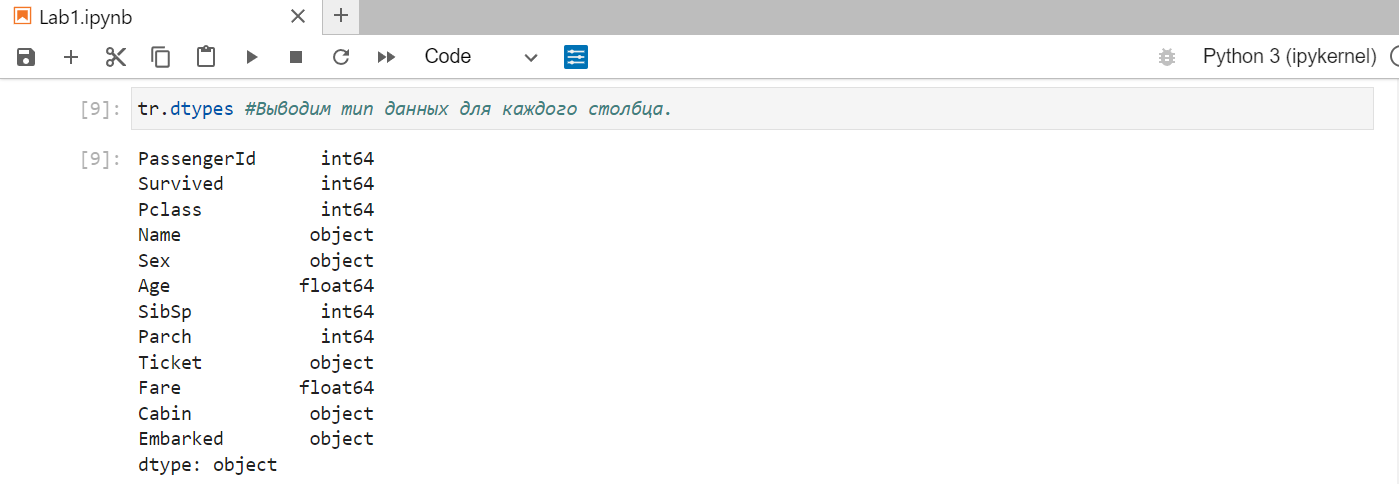
Вывод размерности таблицы (количество строк и столбцов).

*tr.shape #Выводим количество строк и столбцов таблицы.*



Вывод типов данных каждого столбца.

*tr.dtypes #Выводим тип данных для каждого столбца.*



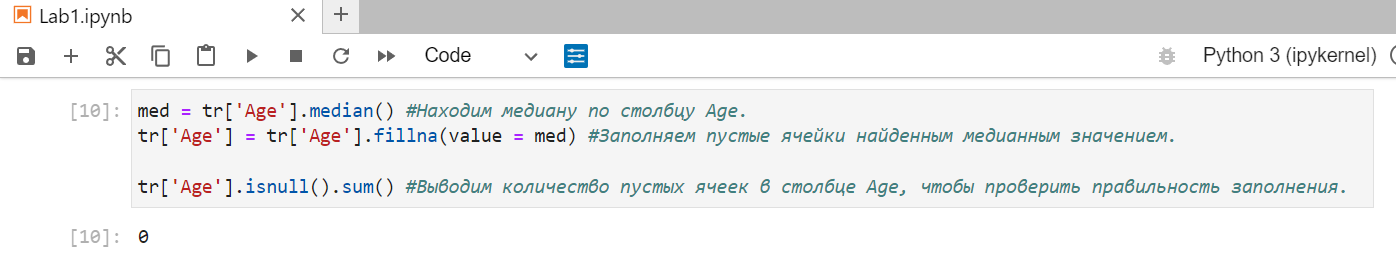
1. **Очистка и подготовка данных.**

Заполняем пустые ячейки в столбце возраста (Age) медианным значением возраста пассажиров.

*med = tr['Age'].median() #Находим медиану по столбцу Age.*

*tr['Age'] = tr['Age'].fillna(value = med) #Заполняем пустые ячейки найденным медианным значением.*

*tr['Age'].isnull().sum() #Выводим количество пустых ячеек в столбце Age, чтобы проверить правильность заполнения.*



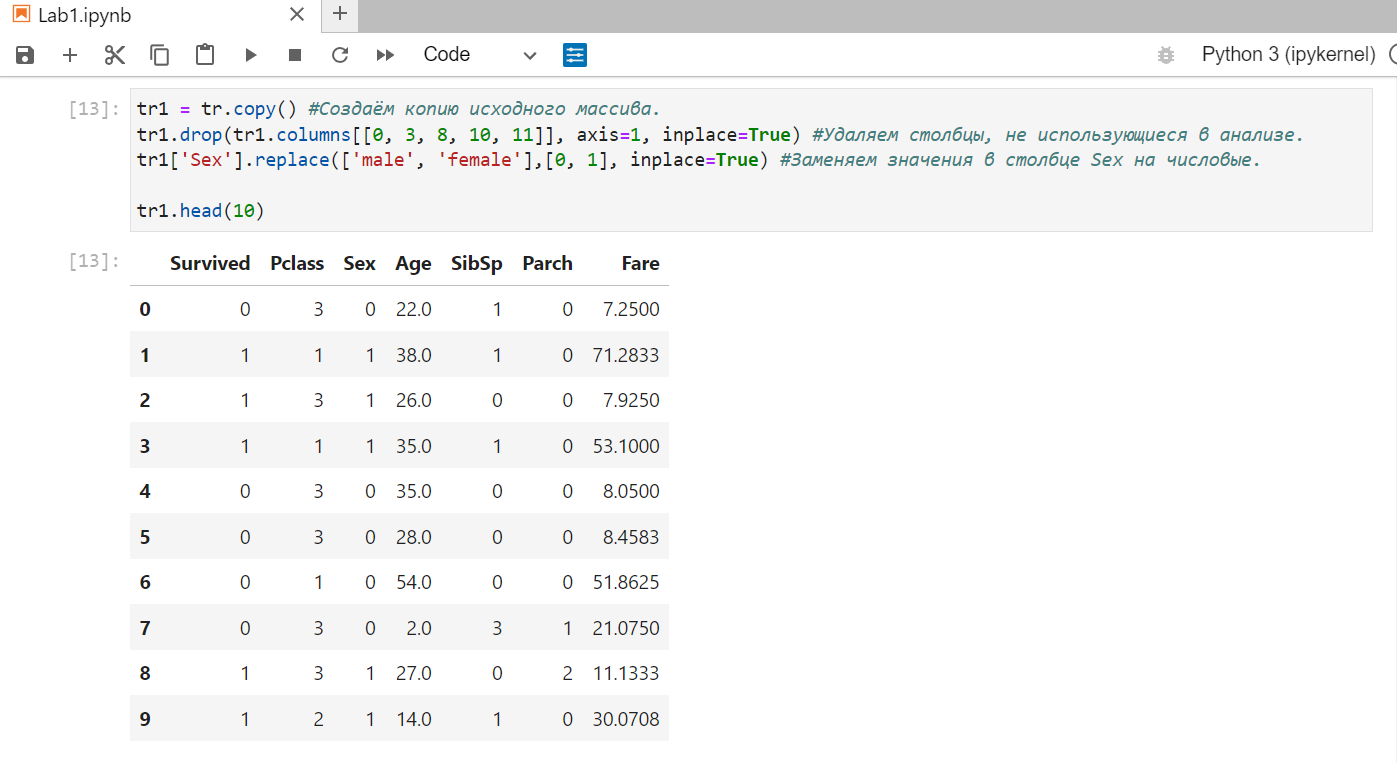
Копирование таблицы и удаление столбцов, которые не будут использоваться в анализе данных, а после вывод первых 10 строк.

*tr1 = tr.copy() #Создаём копию исходного массива.*

*tr1.drop(tr1.columns[[0, 3, 8, 10, 11]], axis=1, inplace=True) #Удаляем столбцы, не использующиеся в анализе.*

*tr1['Sex'].replace(['male', 'female'],[0, 1], inplace=True) #Заменяем значения в столбце Sex на числовые.*

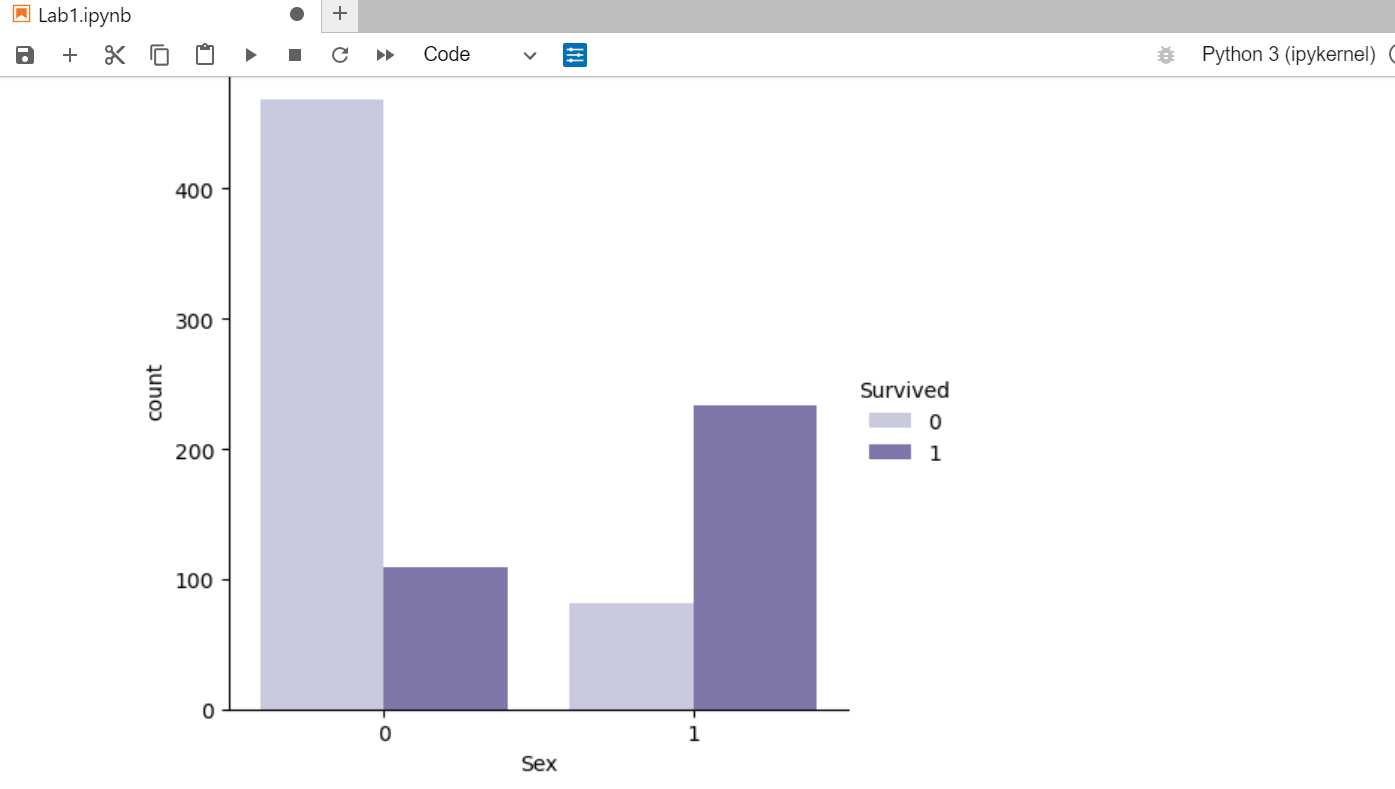
*tr1.head(10)*



1. **Анализ данных с помощью графиков.**

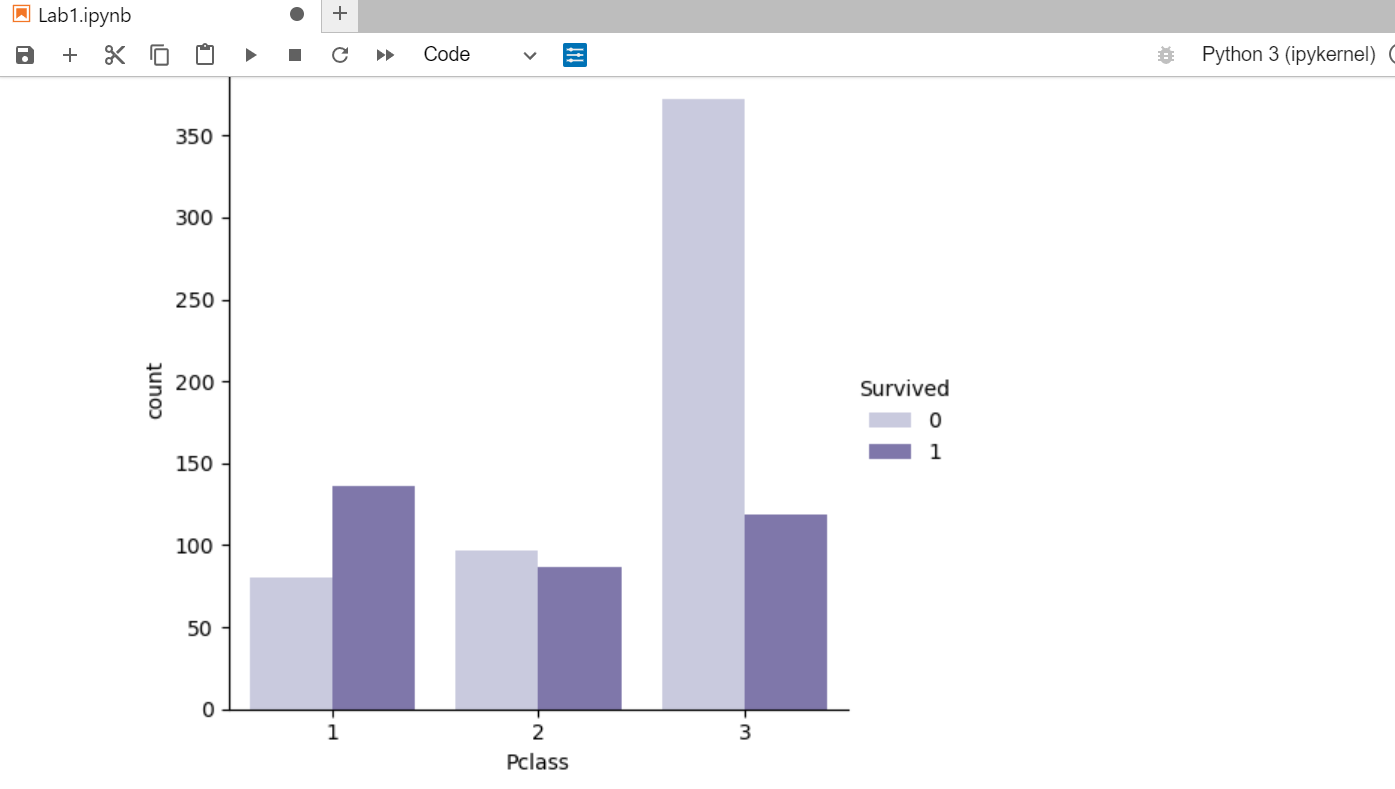
Сравнение выживаемости пассажиров в зависимости от пола (0 – мужской, 1 – женский).

sb.catplot(data=tr1, x="Sex", hue="Survived", kind="count") #Зависимость выживаемости от пола.



Сравнение выживаемости пассажиров в зависимости от класса.

sb.catplot(data=tr1, x="Pclass", hue="Survived", kind="count") #Зависимость выживаемости от класса.



**Вывод:**

Мы научились работать в Pandas, считывать данные из таблицы, работать с ними, изменять и анализировать, выводить с помощью Seaborn диаграммы.