**Домашнее задание № 5**

**Тема "Pandas – продвинутый"**

**Автор: Сорокин Максим Евгеньевич**

**Формулировка задания:**

Необходимо загрузить и обработать предложенный датасет (Титаник), провести аналитику данных с помощью фреймворка Python Pandas.

**Описания плана работы:**

1. Создать новый ноутбук в Colab

2. Сохранить датасет в Google диске и загрузить в ноутбук Colab

https://www.kaggle.com/datasets/yasserh/titanic-dataset/code

\*Загрузить датасет по ссылке из интернета (репозитория Kaggle)

3. Перечень и описание столбцов:

Survived - выжил (1-да, 0- нет)

Pclass - Класс круиза

Name - ФИО

Sex - Пол

Age - Возраст

SibSp - число братьев, сестер или супругов на борту у человека

Parch - количество родителей или детей, с которыми путешествовал каждый пассажир

Ticket - Номер билета

Fare - Цена билета

Cabin - Номер каюты

Embarked - Порт посадки

4. Определить количество пустых ячеек

5. \*Заполнить пустые ячейки используя любую логику замещения данных

6. Если пункт 5 не выполнен, то удалить строки имеющие пустые ячейки.

7. По каждому признаку произвести аналитику:

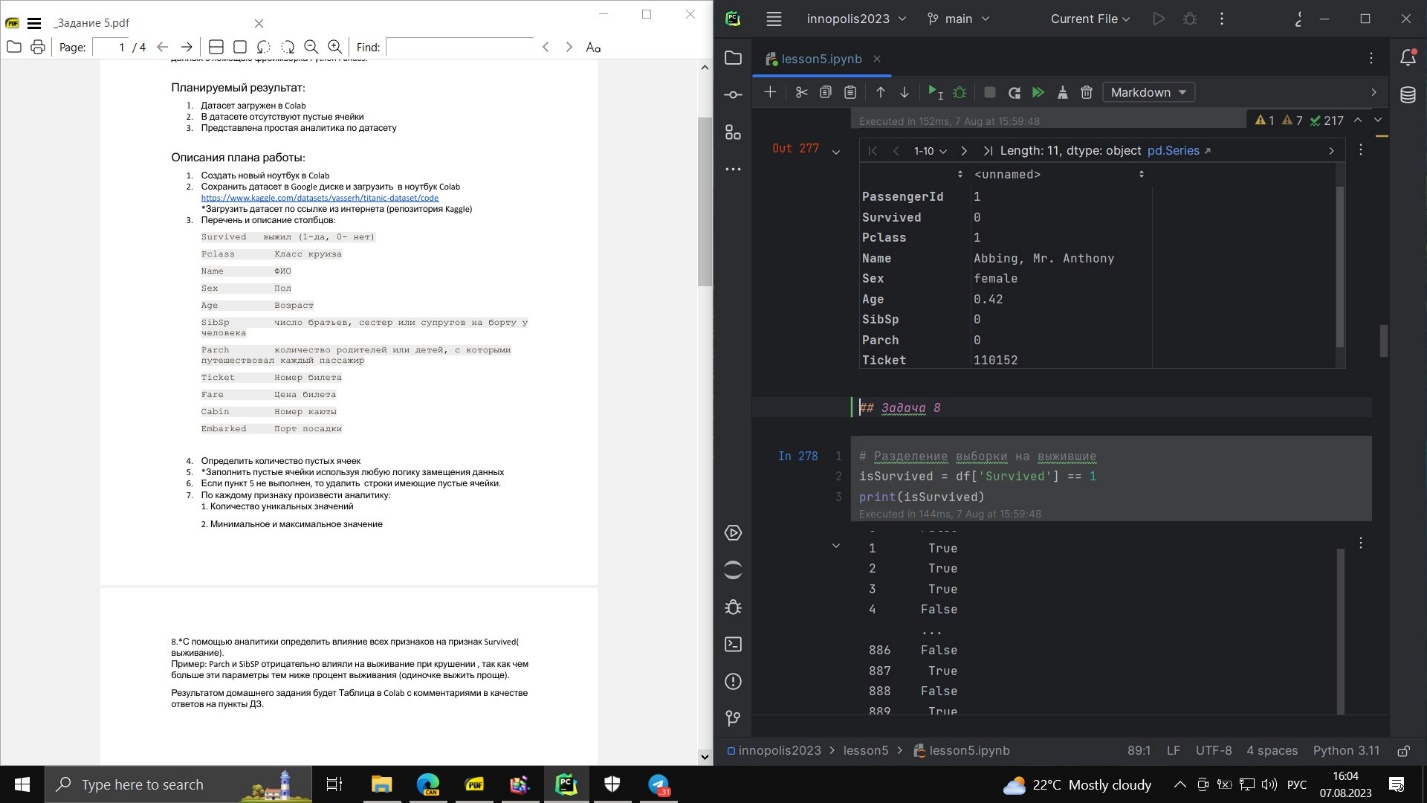
1. Количество уникальных значений

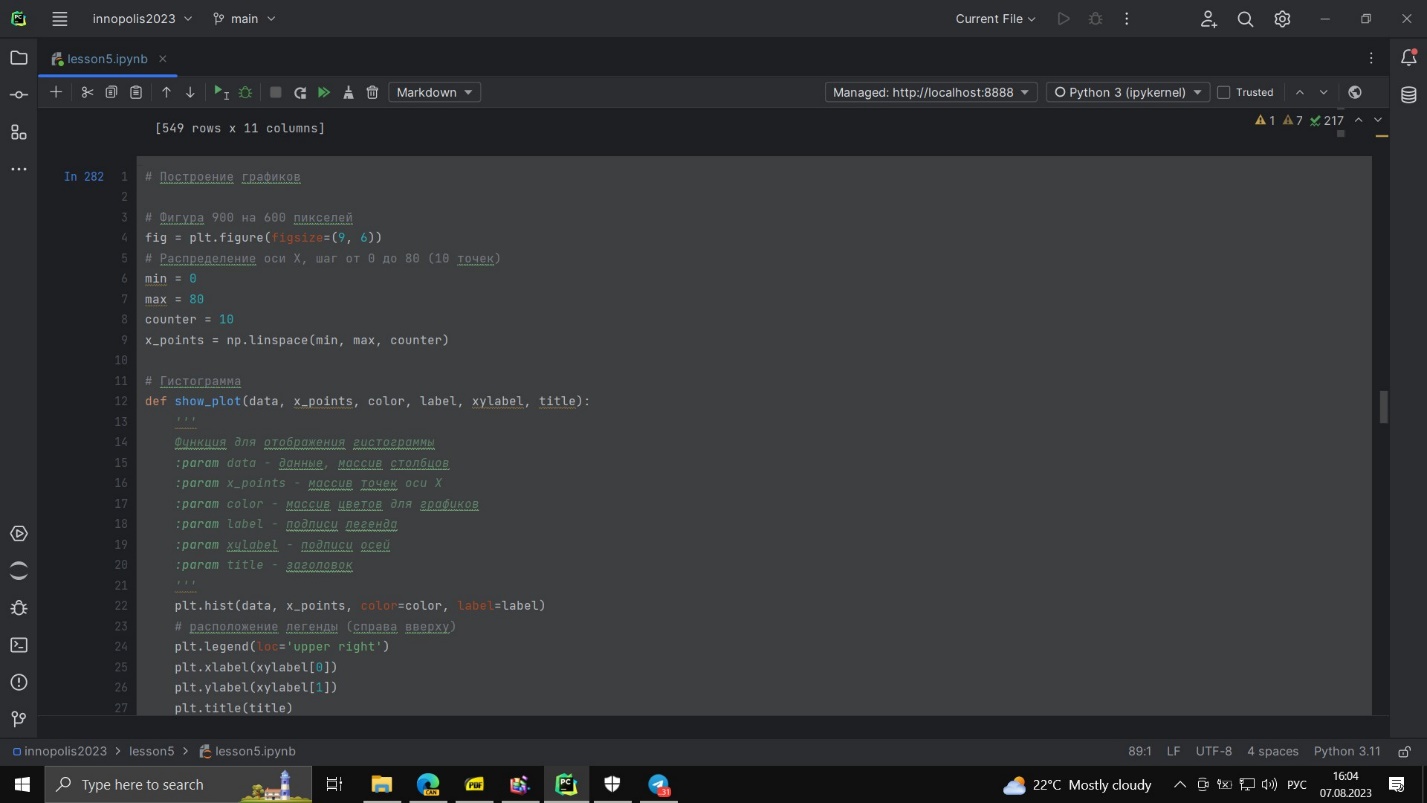
2. Минимальное и максимальное значение

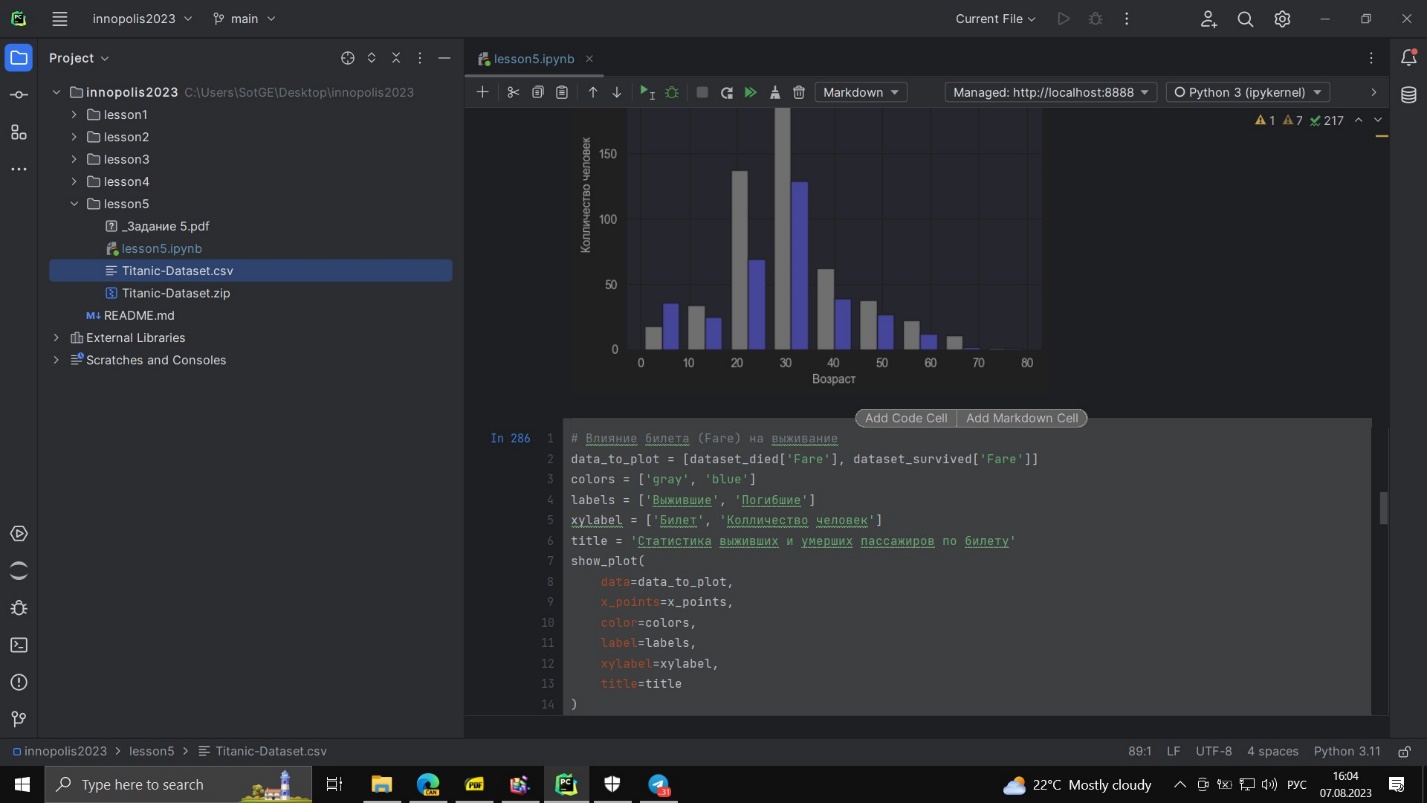
8.\*С помощью аналитики определить влияние всех признаков на признак Survived (выживание).

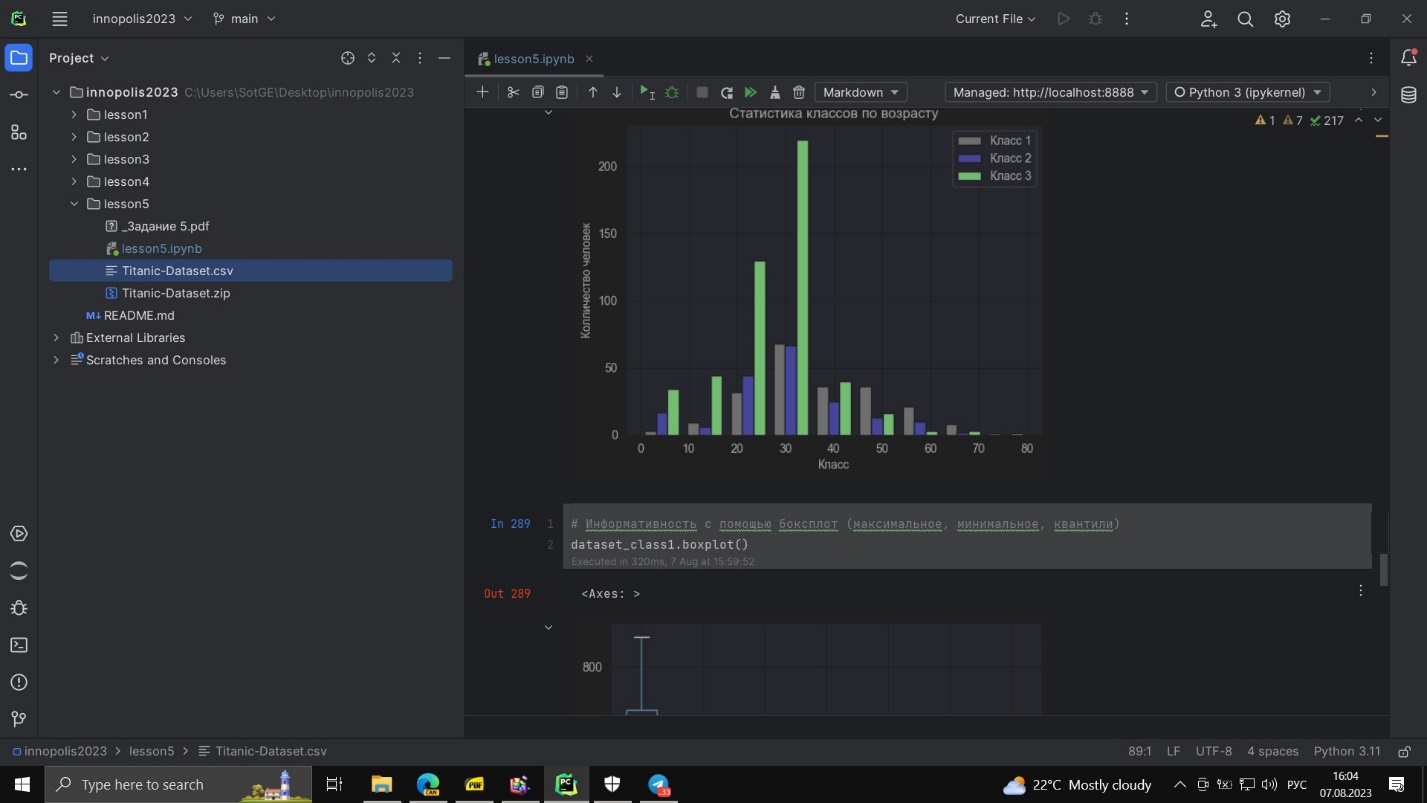
Пример: Parch и SibSP отрицательно влияли на выживание при крушении, так как чем

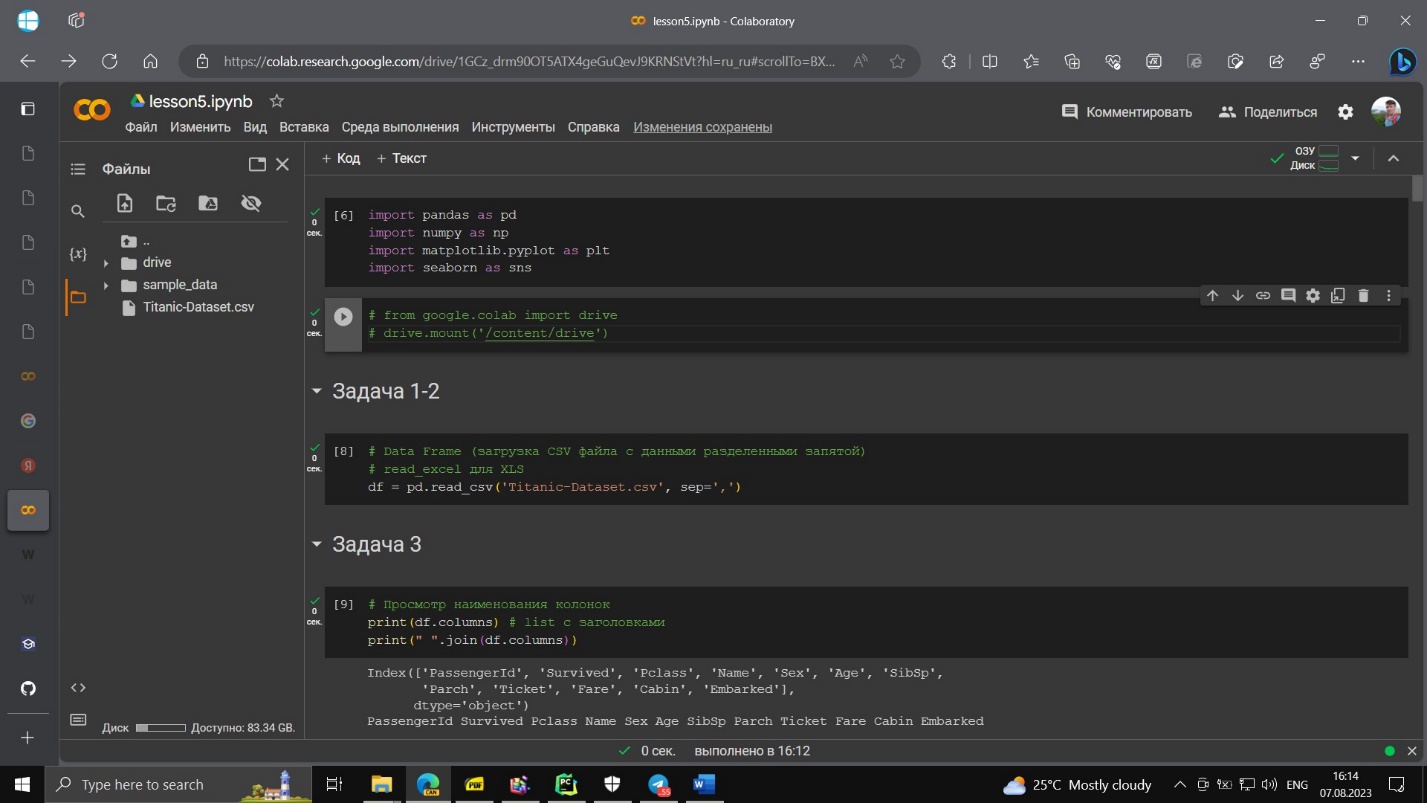
**Решение:**

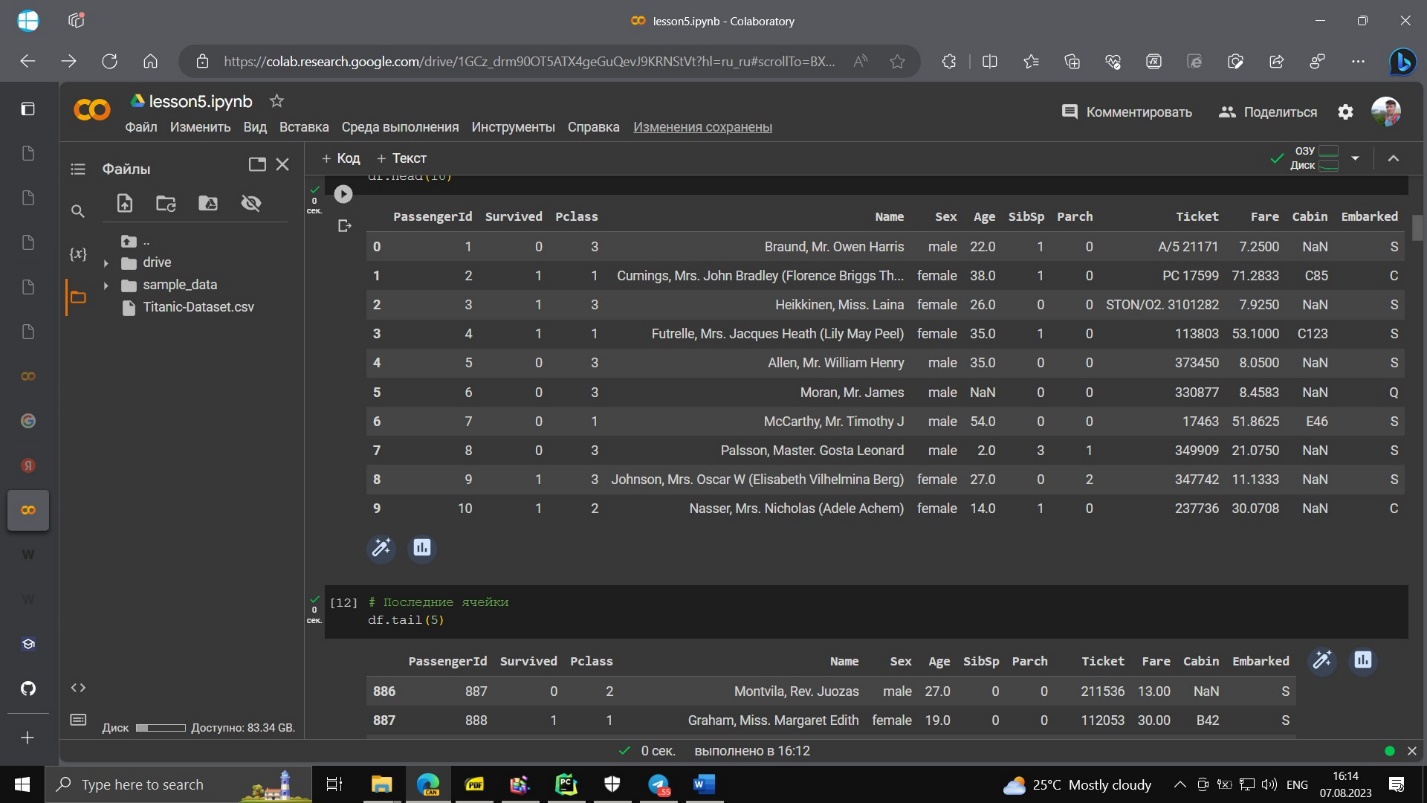


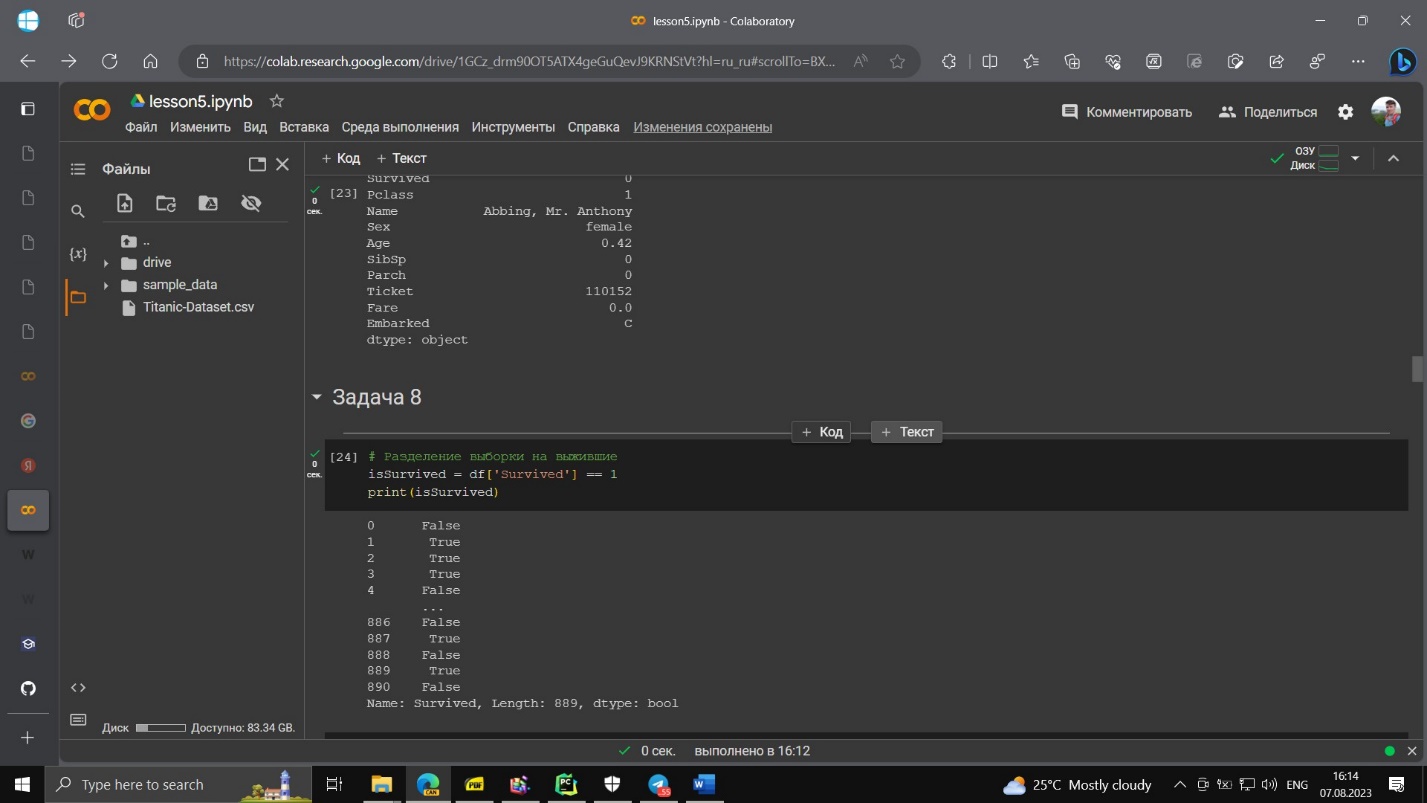


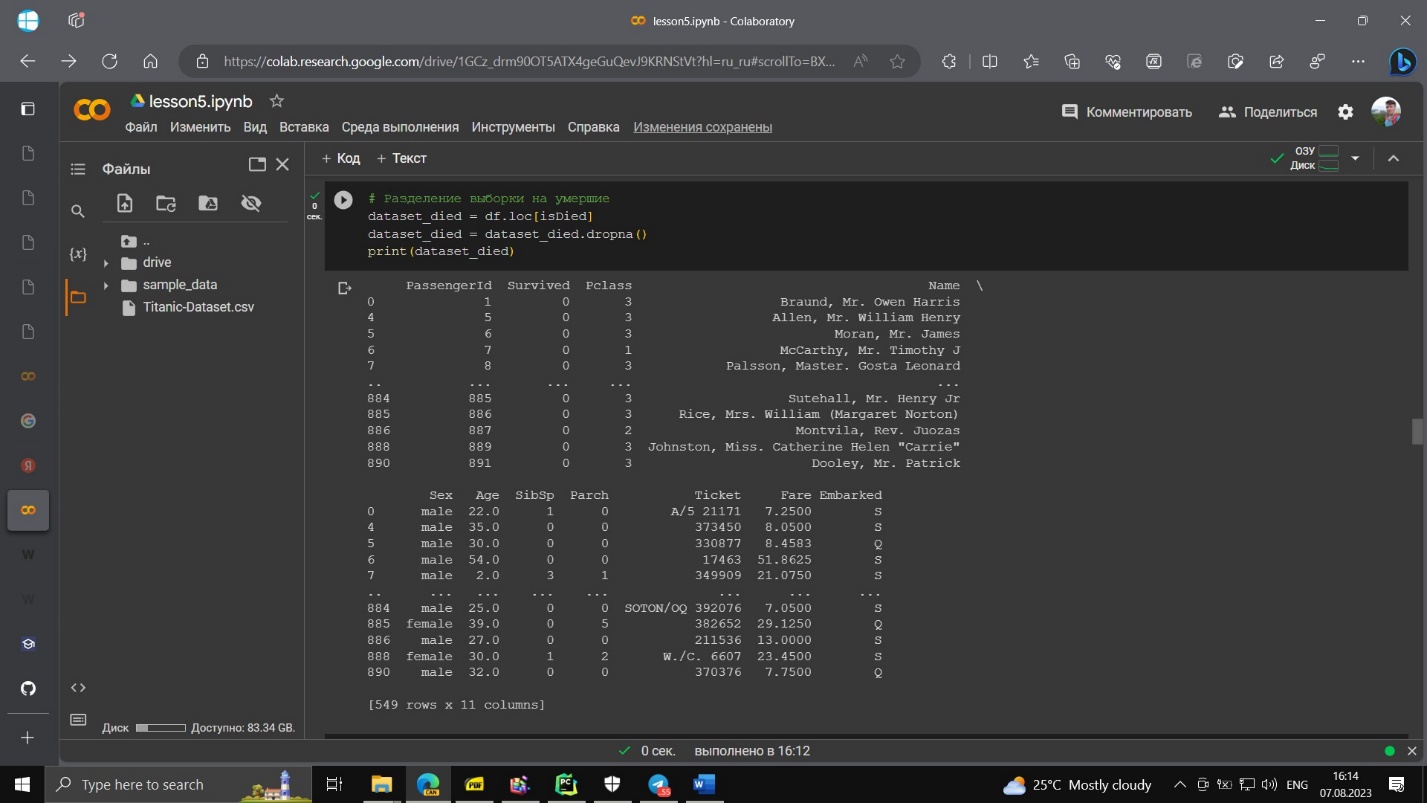


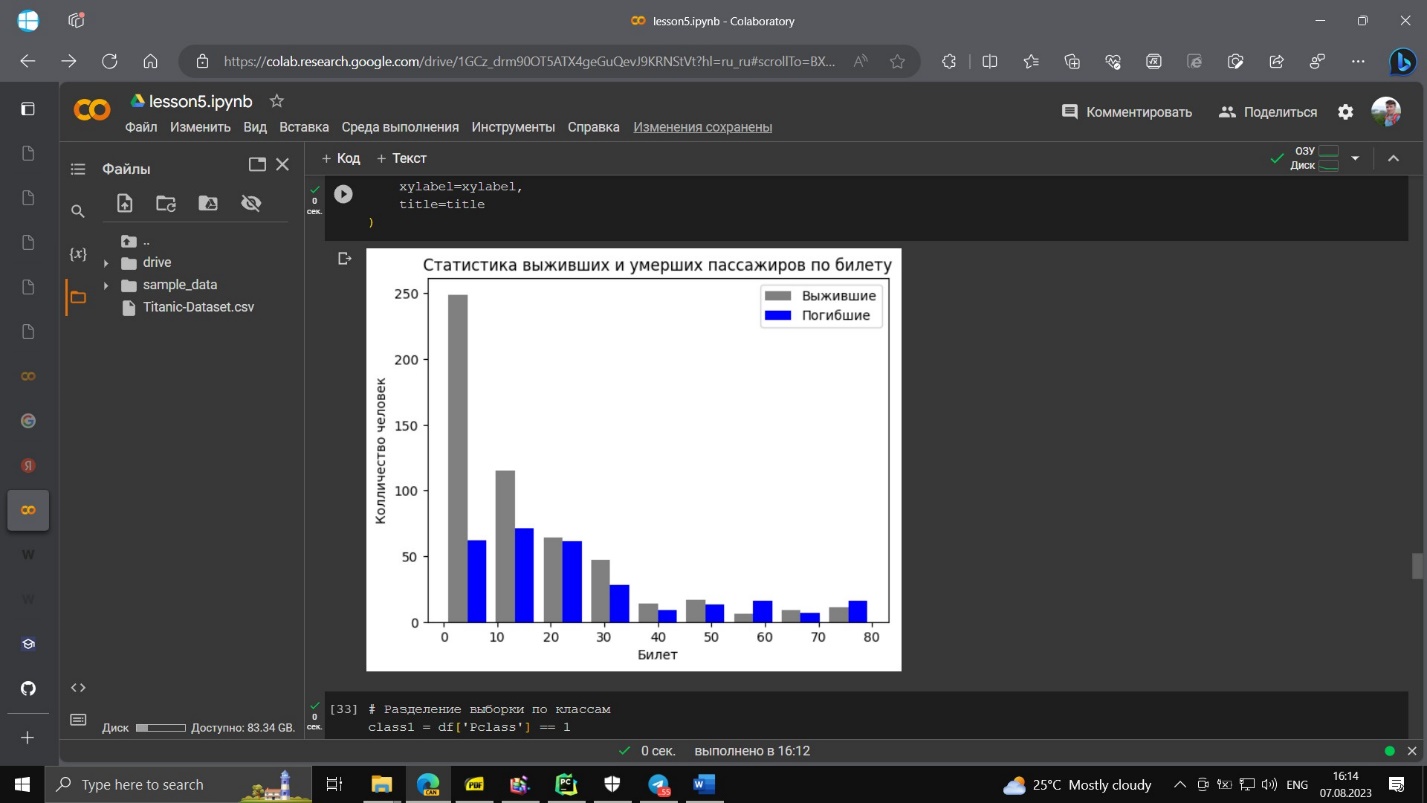


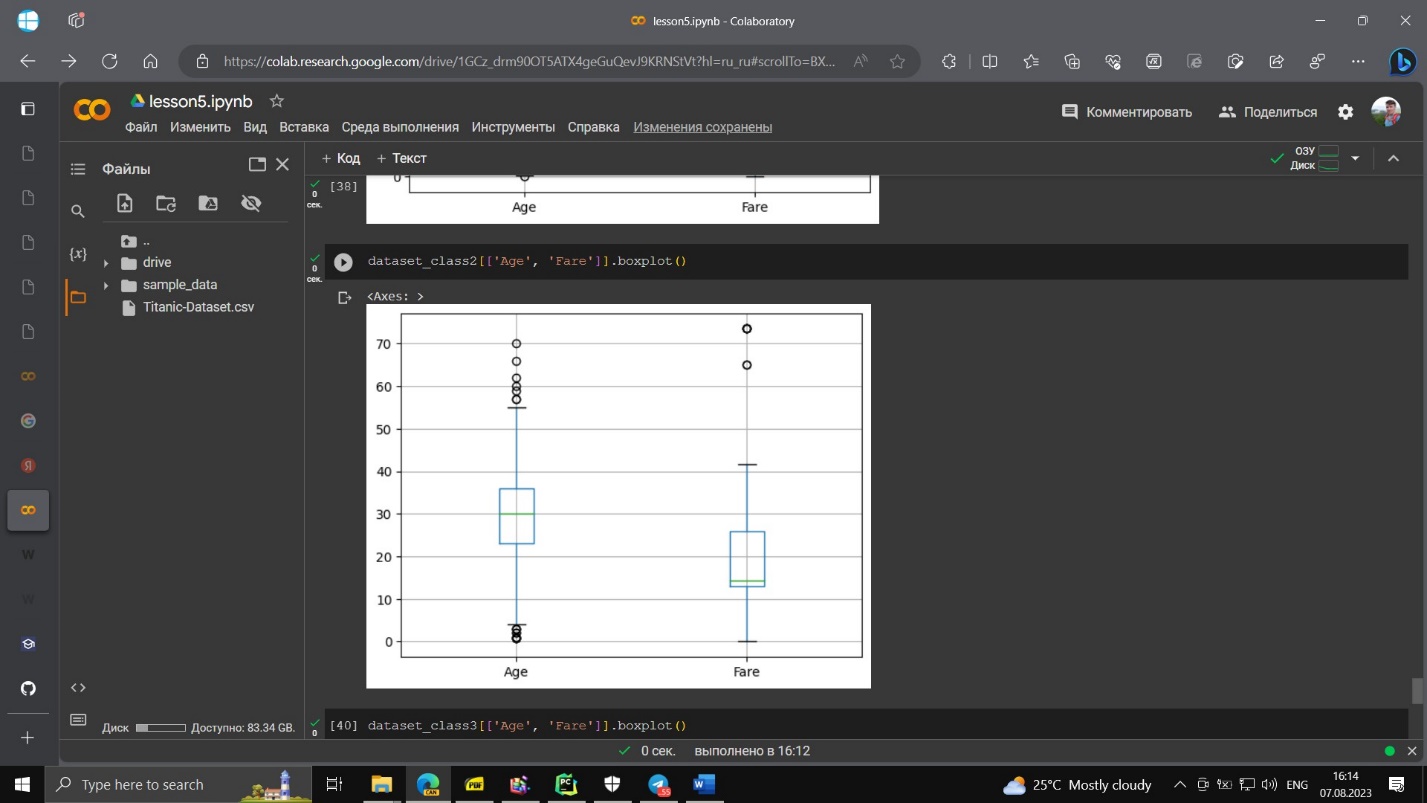












К решению приложены файлы:

«lesson5.ipynb» для Google Colab

«lesson5.py» исходный код для Python

Ссылка на Notebook Colab:

<https://colab.research.google.com/drive/1GCz_drm90OT5ATX4geGuQevJ9KRNStVt?usp=sharing>

Ссылка на GitHub:

<https://github.com/SotGE/innopolis2023/tree/main/lesson5>