Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Российский экономический университет имени Г.В. Плеханова»

**МОСКОВСКИЙ ПРИБОРОСТРОИТЕЛЬНЫЙ ТЕХНИКУМ**

Специальность 09.02.07 Информационные системы и программирование

Квалификация: Программист

ПРАКТИЧЕСКИЕ РАБОТЫ

По дисциплине «Технология разработки и защиты баз данных»

Листов: 48

|  |  |
| --- | --- |
| Выполнил: студент  Студент группы П50-4-21  Игошев Р.В. | Проверил преподаватель  \_\_\_\_\_\_\_\_\_\_\_\_Ю. В. Севастьянов  «\_\_\_\_» \_\_\_\_\_\_\_\_ 2024 года |

Москва 2024

**ОГЛАВЛЕНИЕ**

[ПРАКТИЧЕСКАЯ РАБОТА №1 3](#_Toc160911327)

[«Разведочный анализ» 3](#_Toc160911328)

# ПРАКТИЧЕСКАЯ РАБОТА №1

## «Разведочный анализ»

Цель работы: Провести разведочный анализ данных по датасету.

Начать работу с таблицей следует с импорта пакетов и запись таблицы в переменную.

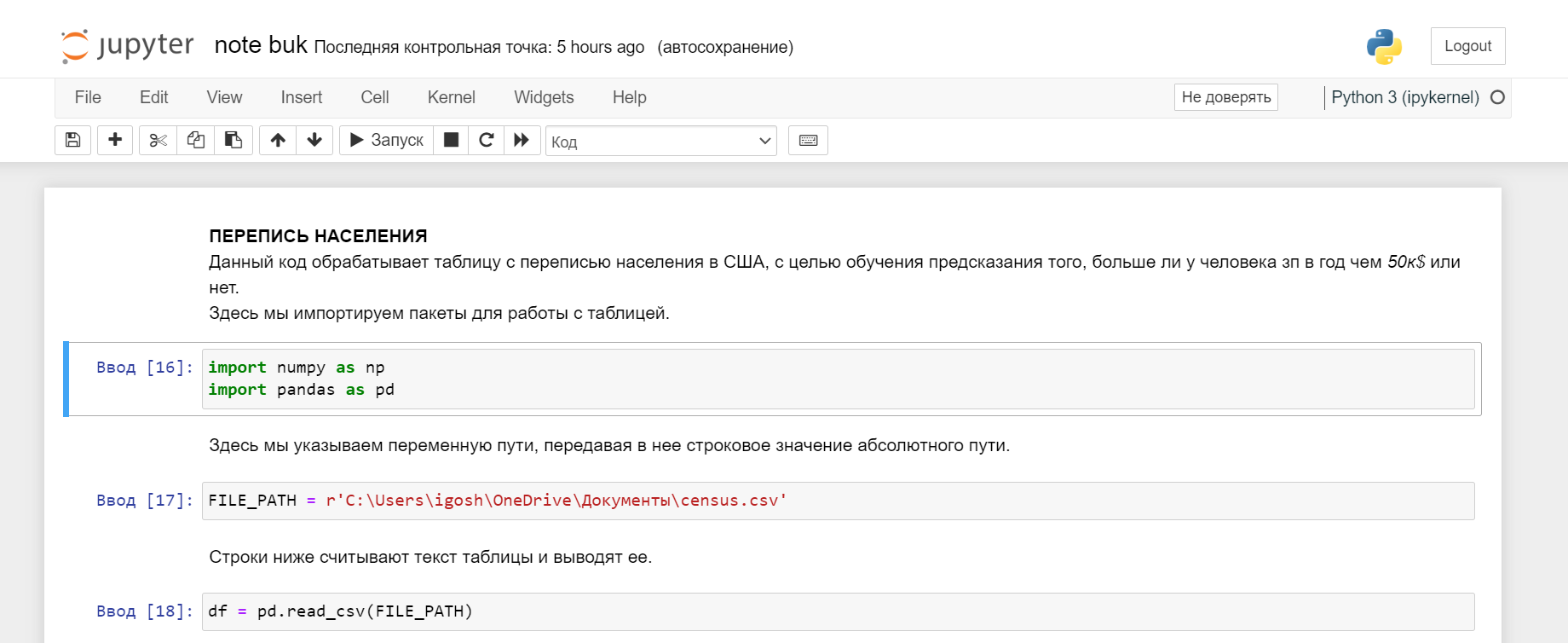


Рисунок 1 – Импорт пакетов и получение таблицы

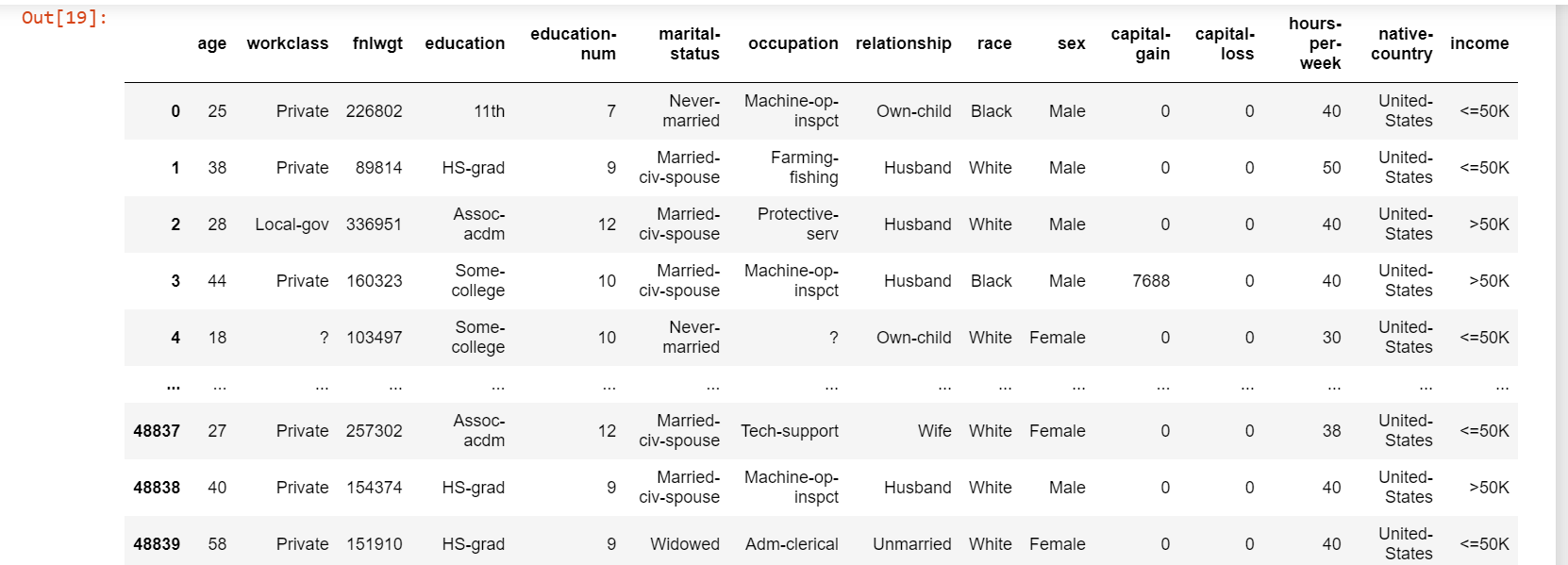


Рисунок 2 – Вывод таблицы

Далее стоит произвести разведочный анализ:



Рисунок 3 – Вывод первых 5-ти строк



Рисунок 4 – Вывод статистических показателей

Кроме того, следует и изменить таблицу, убрав ненужные данные, дупликаты, и отсортировав таблицу.

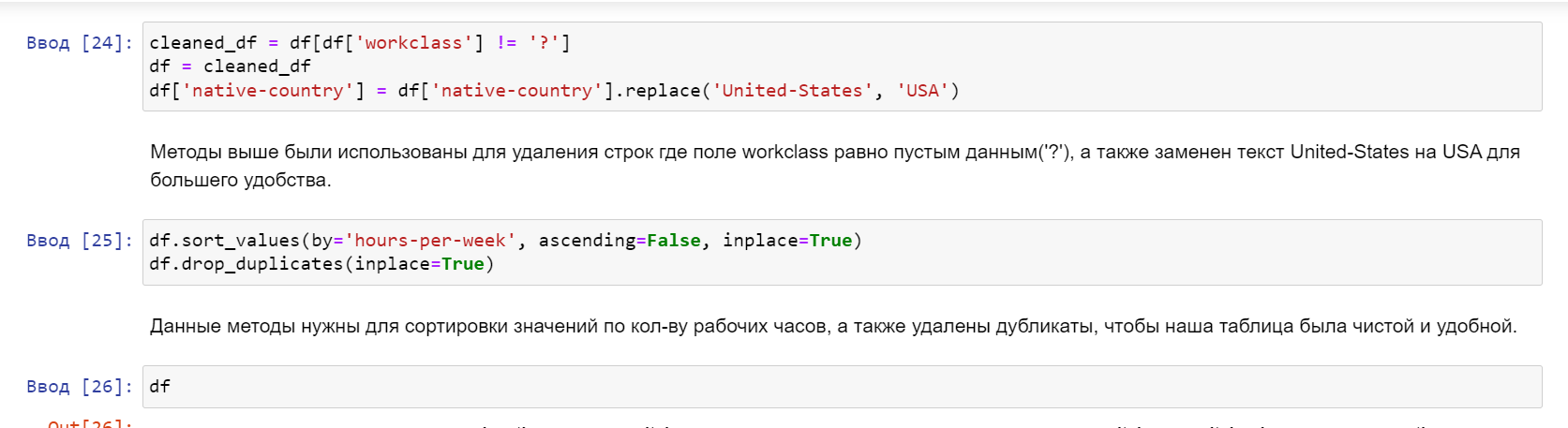


Рисунок 5 – Изменение таблицы

Вывод: выбрав подходящий датасет, был проведен разведочный анализ, включая оптимизацию и документирование.

# ПРАКТИЧЕСКАЯ РАБОТА №2

# «Графики»

Цель работы: Составить графики, проанализировать их.

Практическая работа состоит из двух частей, одна из которых работает с одной таблицей, а уже вторая часть работает с другой таблицей и имеет немного больше графического отображения. В ходе практической необходимо создать графики категориальных признаков, ящики с усами и прочее, что будет видно на изображениях ниже.

Первая часть:

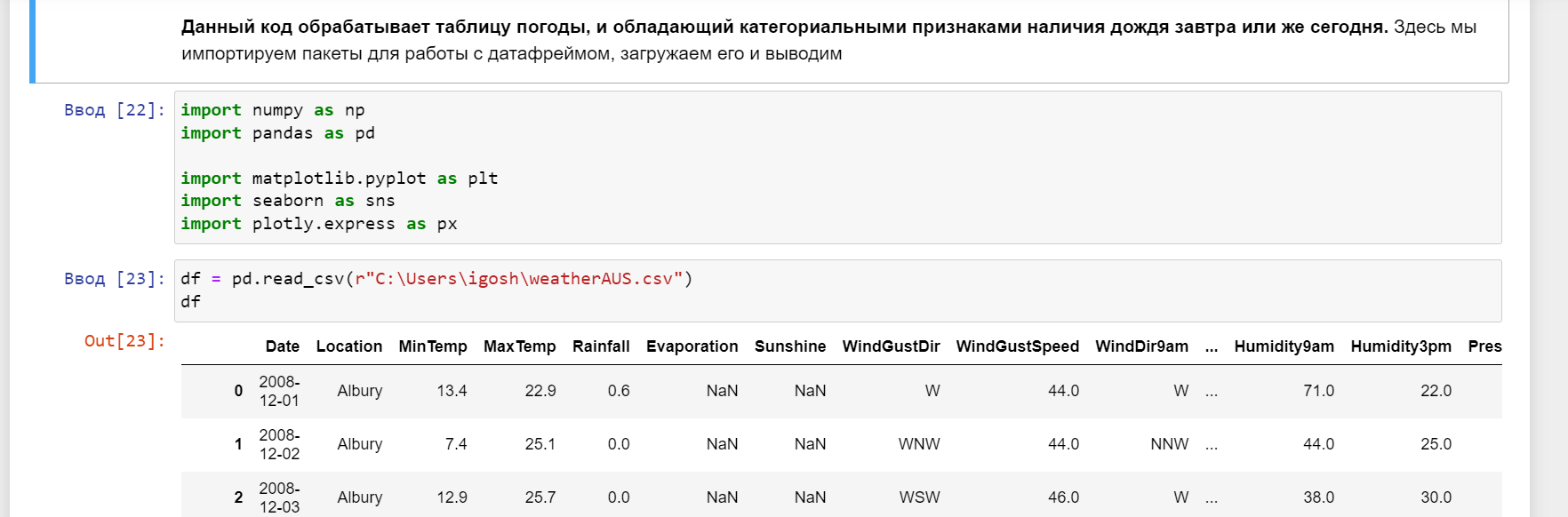


Рисунок 6 – Импорт библиотек и вывод таблицы

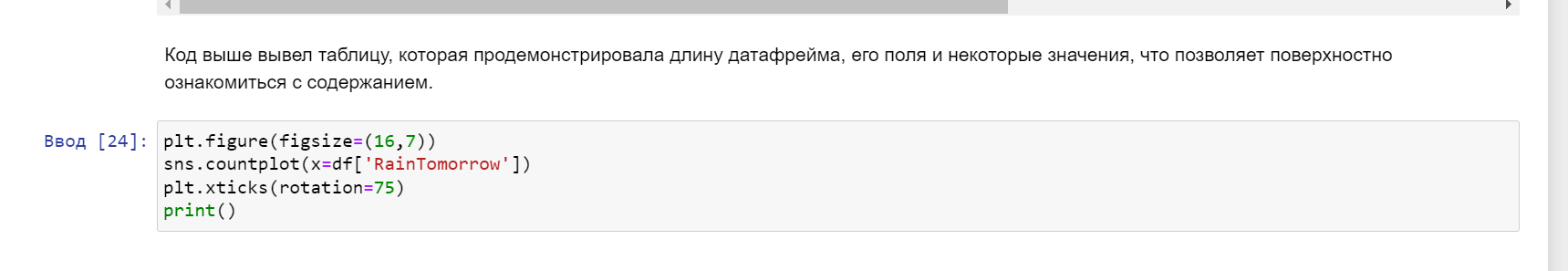


Рисунок 7 – Описание действий

Первый график:

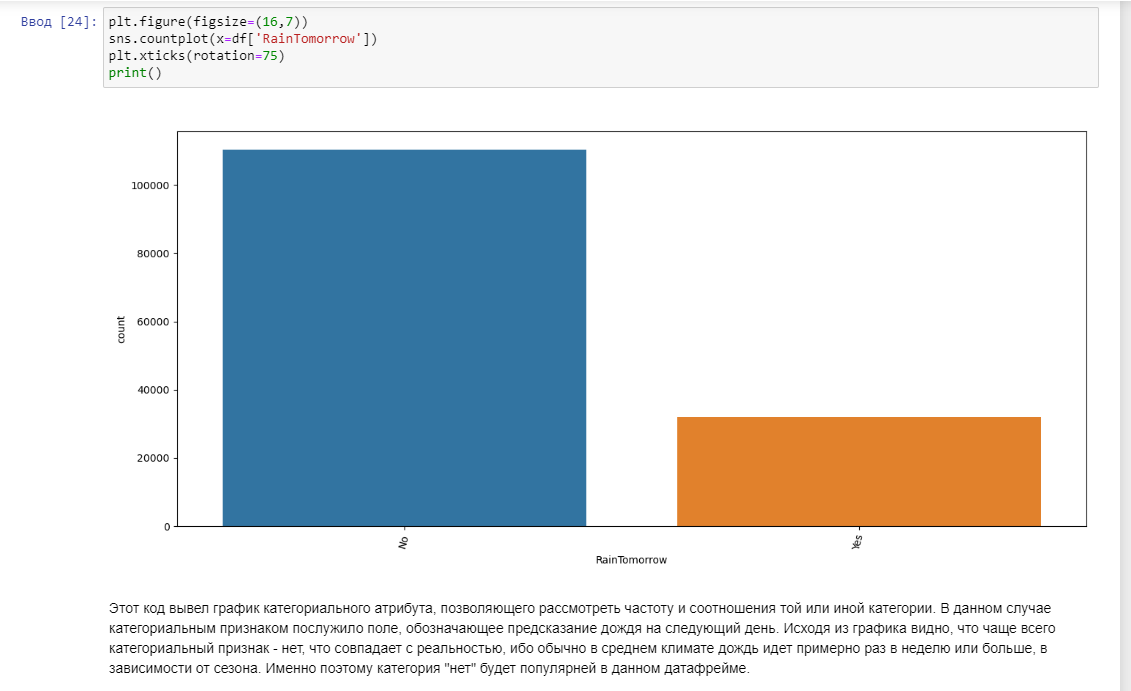


Рисунок 8 – Вывод графика категориального атрибута и описание действий и данных

Второй график:

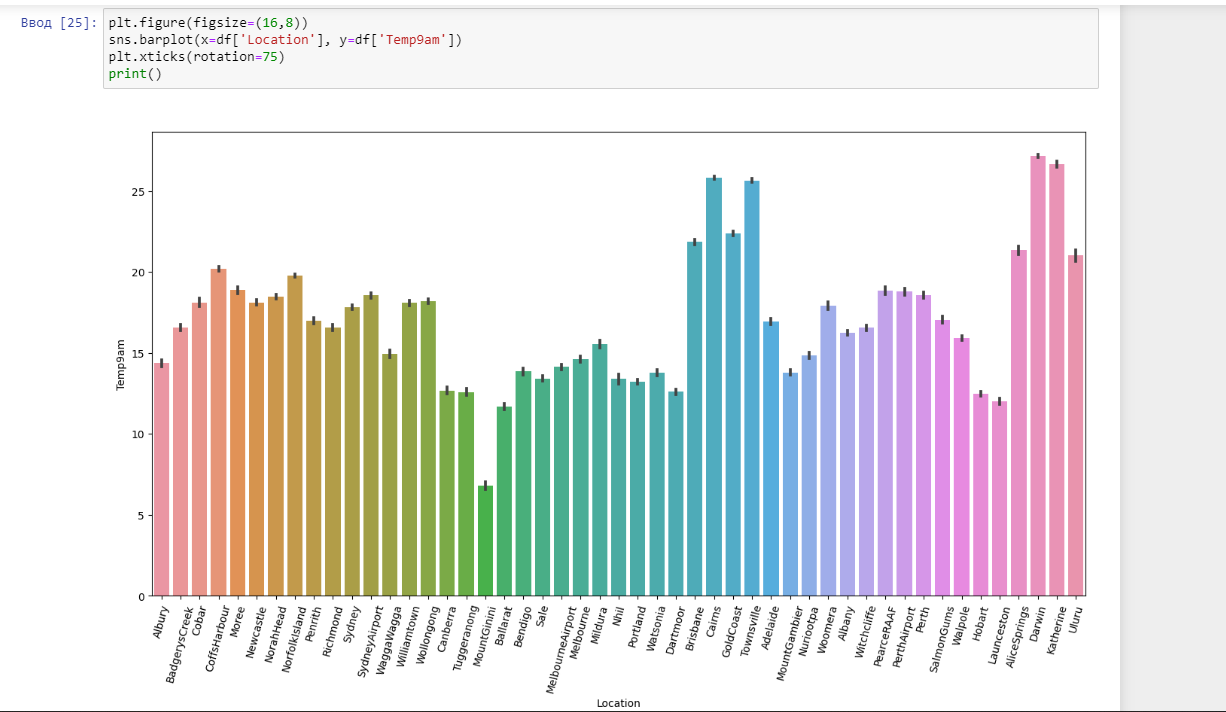


Рисунок 9 – Вывод графика средних температур по городам

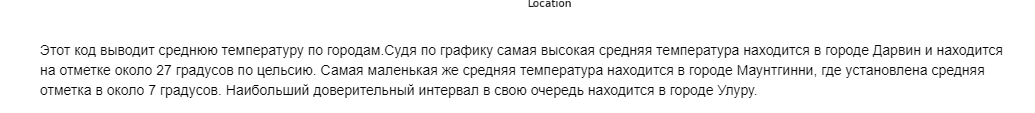


Рисунок 10 – Описание вывода

Третий график:

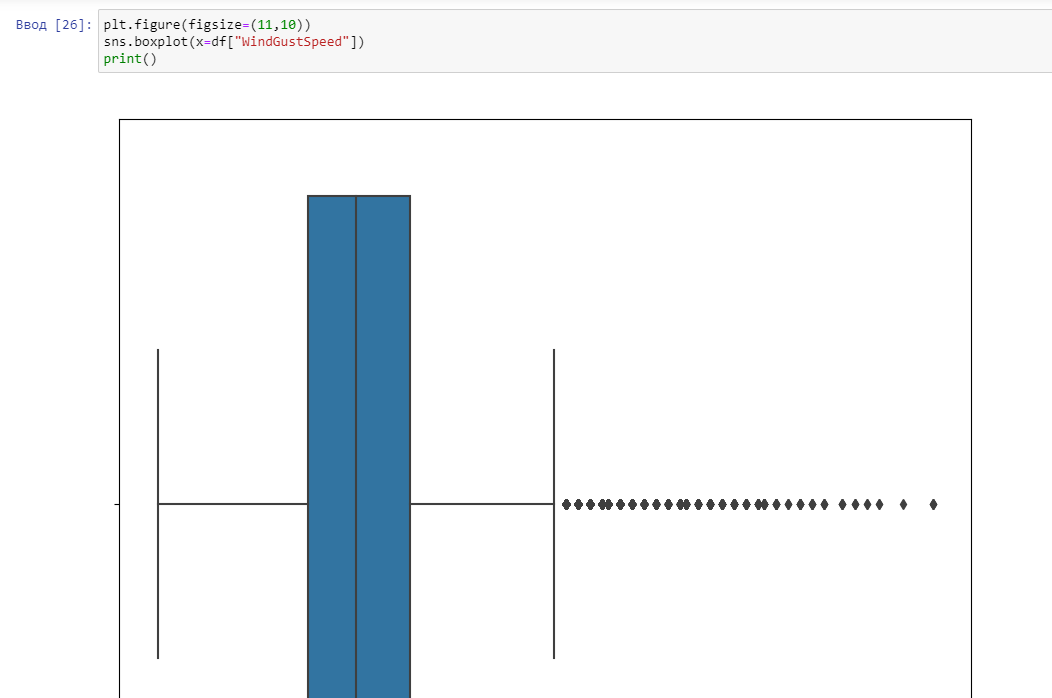


Рисунок 11 – Вывод ящика с усами

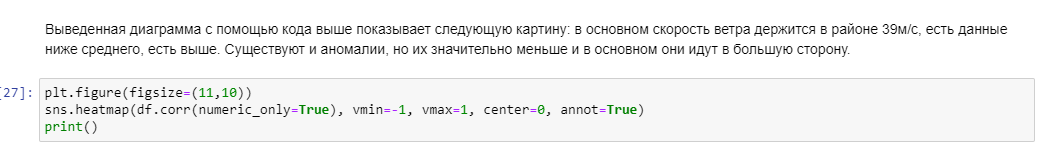


Рисунок 12 – Описание полученного результата и код для вывода тепловой карты

Четвертый график:



Рисунок 13 – Тепловая карта корреляции

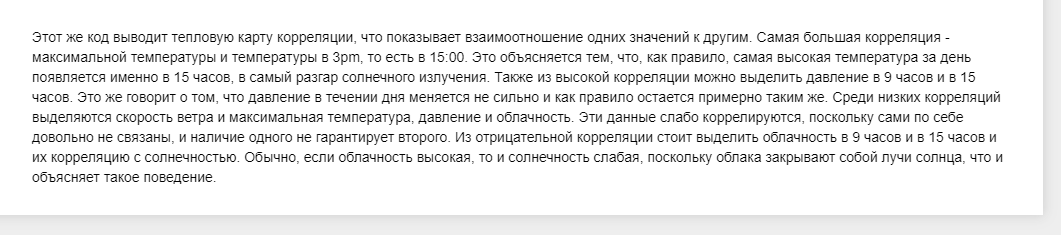


Рисунок 14 – Описание данных тепловой карты корреляции

Вторая часть:

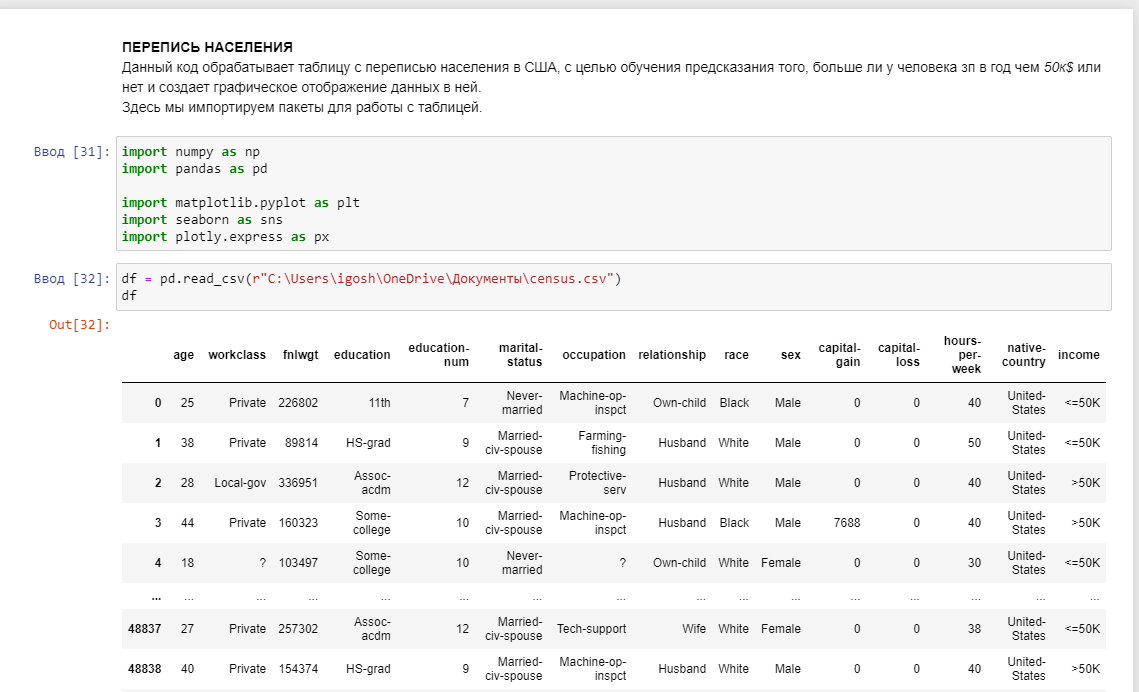


Рисунок 15 – Импорт данных и вывод таблицы

Первый график:

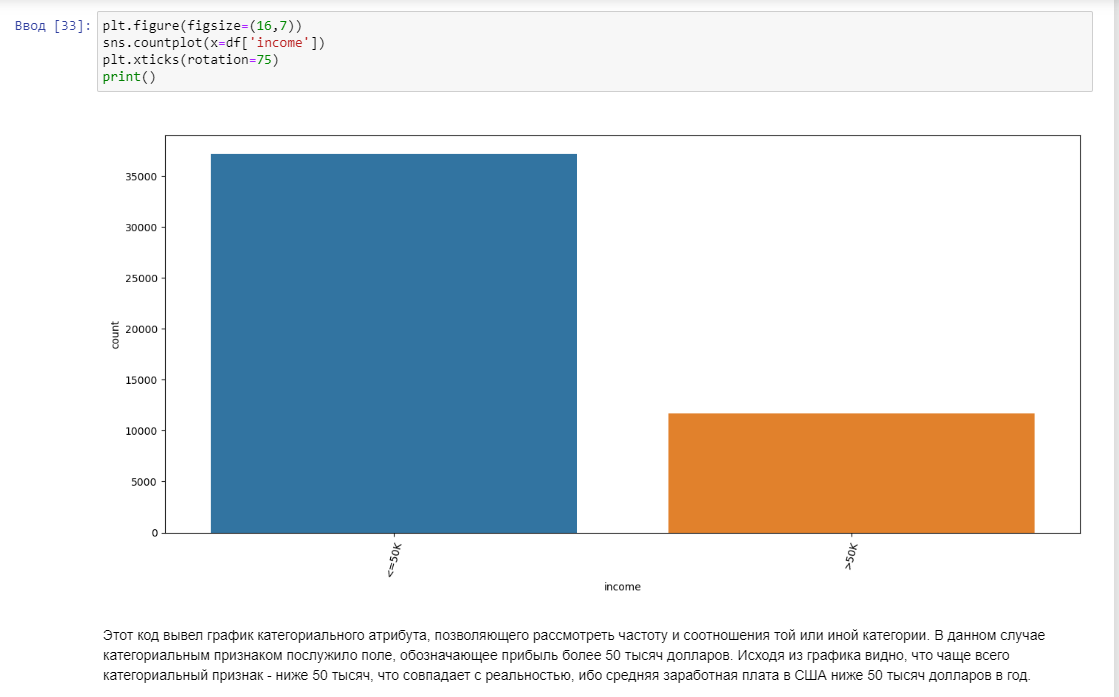


Рисунок 16 – График категориального атрибута и код для него с описанием данных

Второй график:

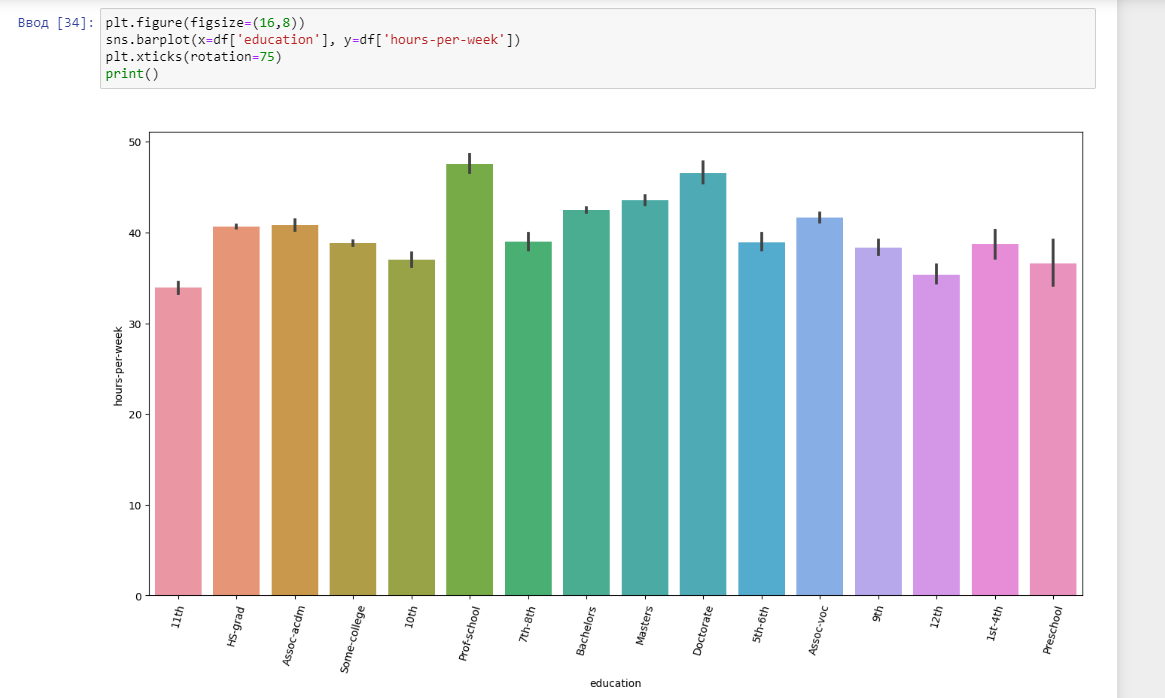


Рисунок 17 – График среднего количества рабочих часов по образованию

Третий график:



Рисунок 18 – Ящик с усами получения капитала

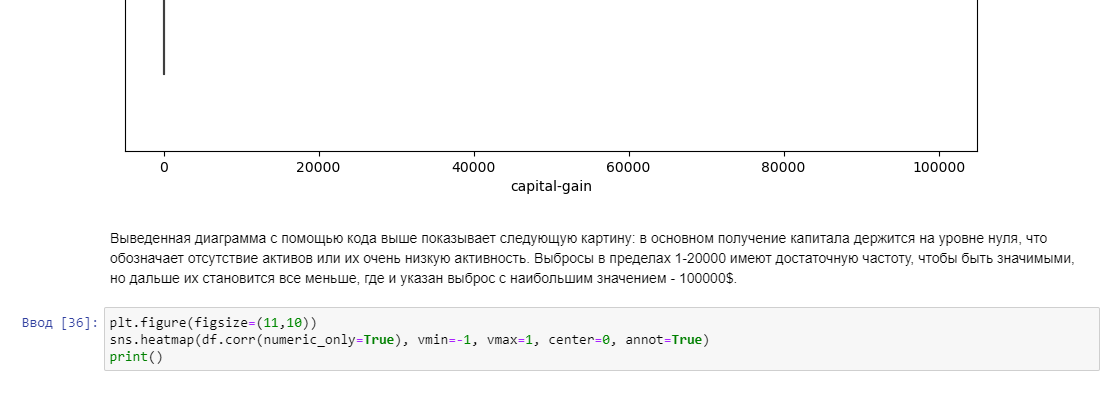


Рисунок 19 – Описание полученного результата и код для тепловой карты корреляции

Четвертый график:

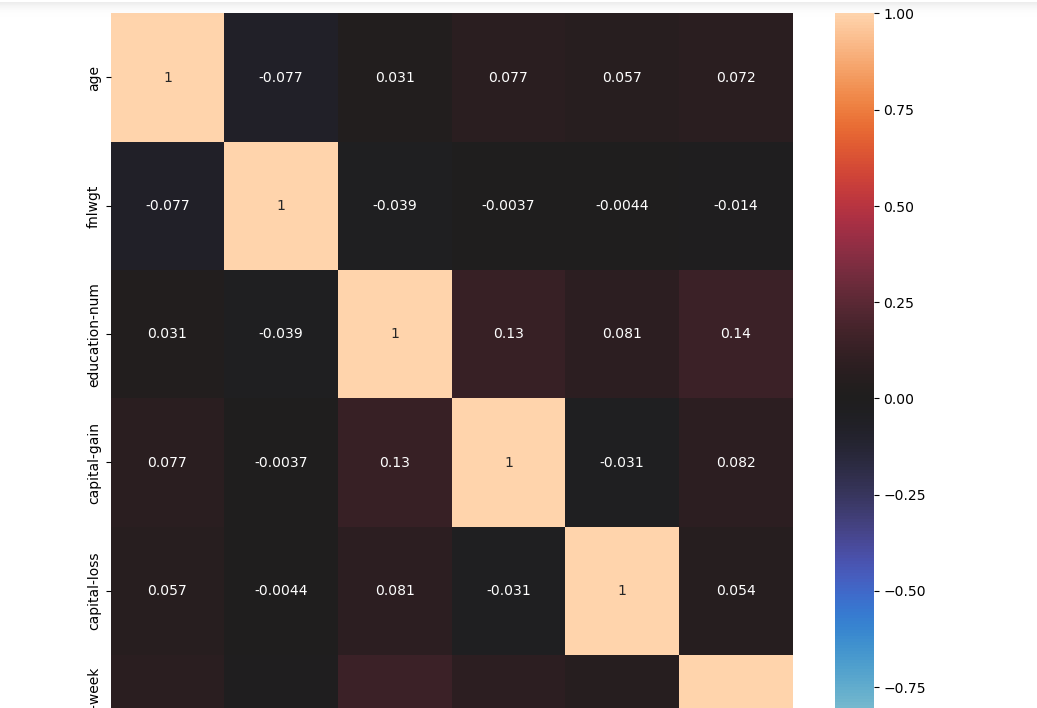


Рисунок 20 – Тепловая корреляция данных

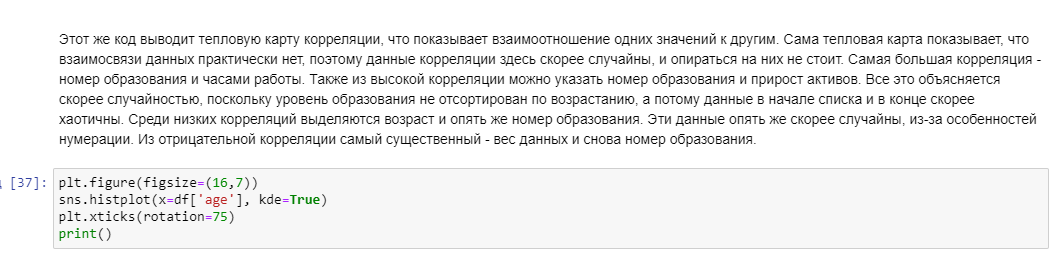


Рисунок 21 – Описание данных корреляции и вывод следующего графика

Пятый график:

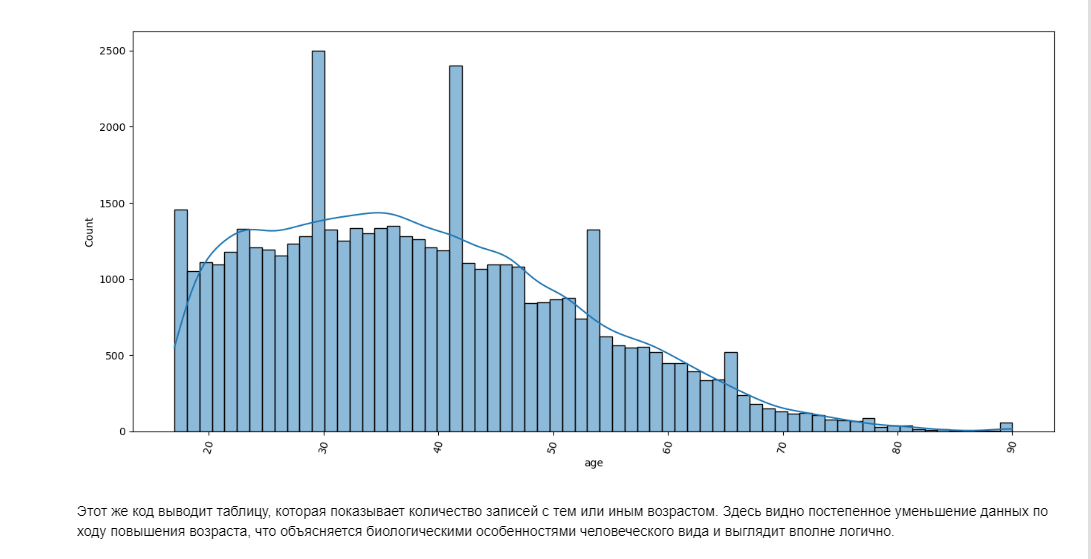


Рисунок 22 – График количества записей определенного возраста

Шестой график:

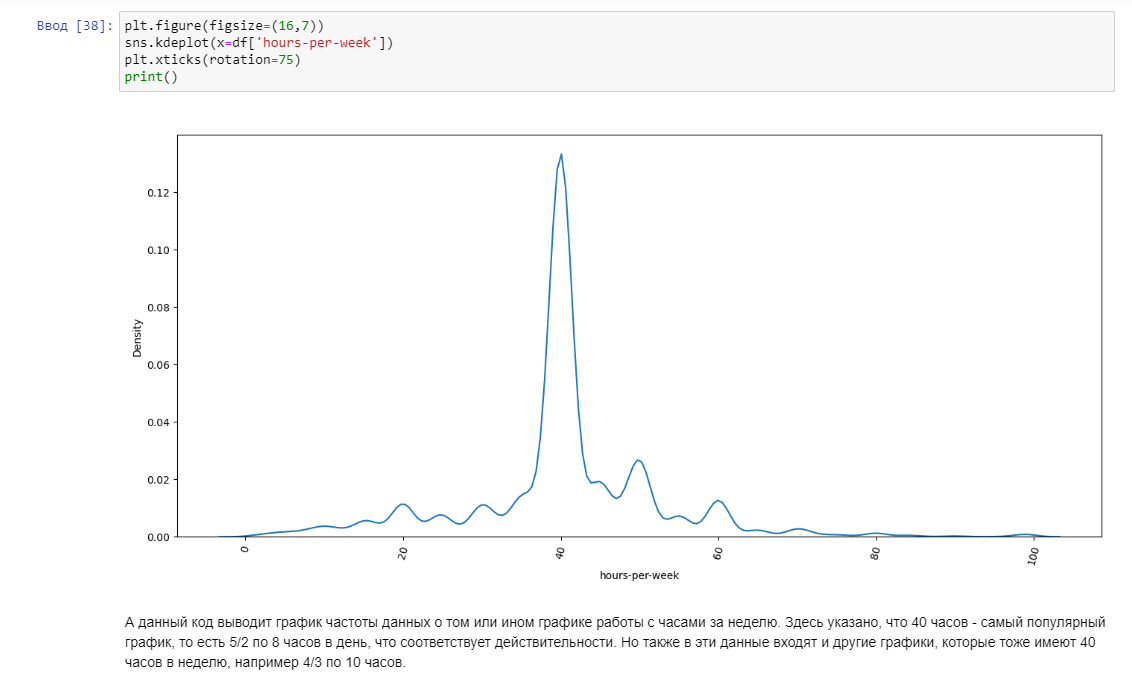


Рисунок 23 – Частота записей с количеством рабочих часов

Вывод: в ходе выполнения работы были составлены графики и проанализированы данные в них, рассмотрены различные виды и получены новые знания.

ПРАКТИЧЕСКАЯ РАБОТА №3

«Преобразование данных»

Цель работы: Закодировать и преобразовать признаки.

Практическая работа состоит из двух частей, одна из которых работает с одной таблицей, а уже вторая часть работает с другой таблицей и имеет другую тему.

Первая часть:

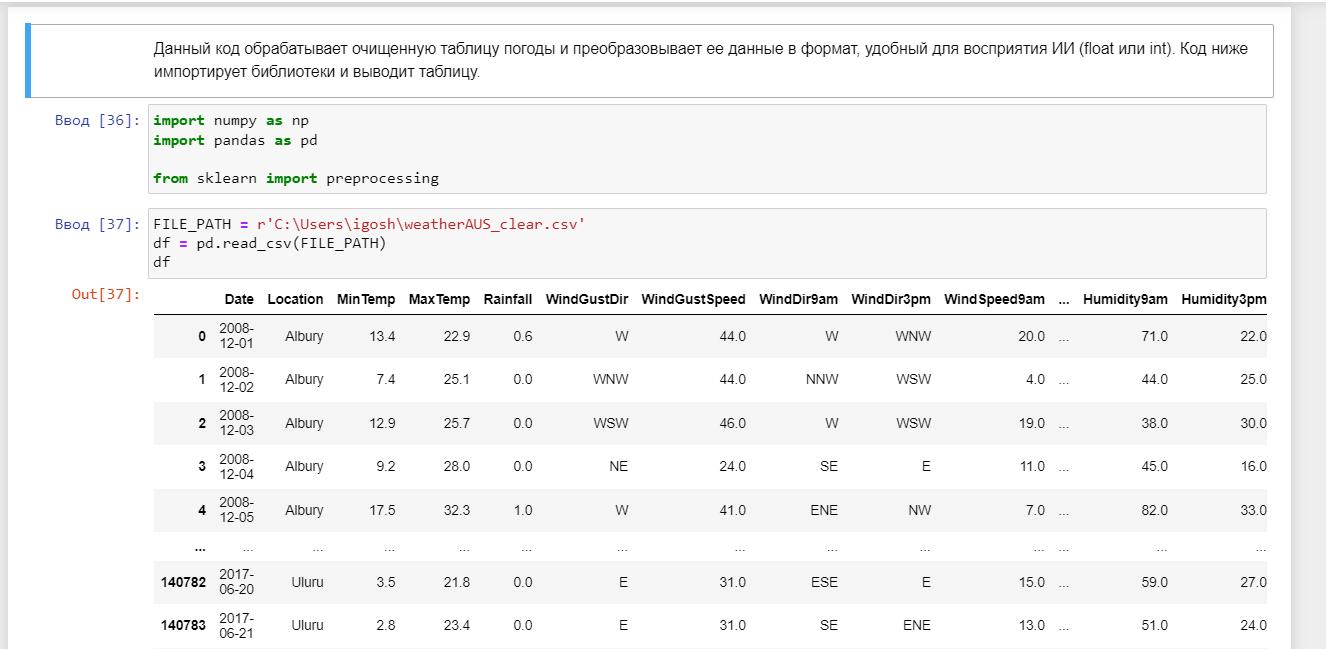


Рисунок 24 – Импорт библиотеки и вывод таблицы

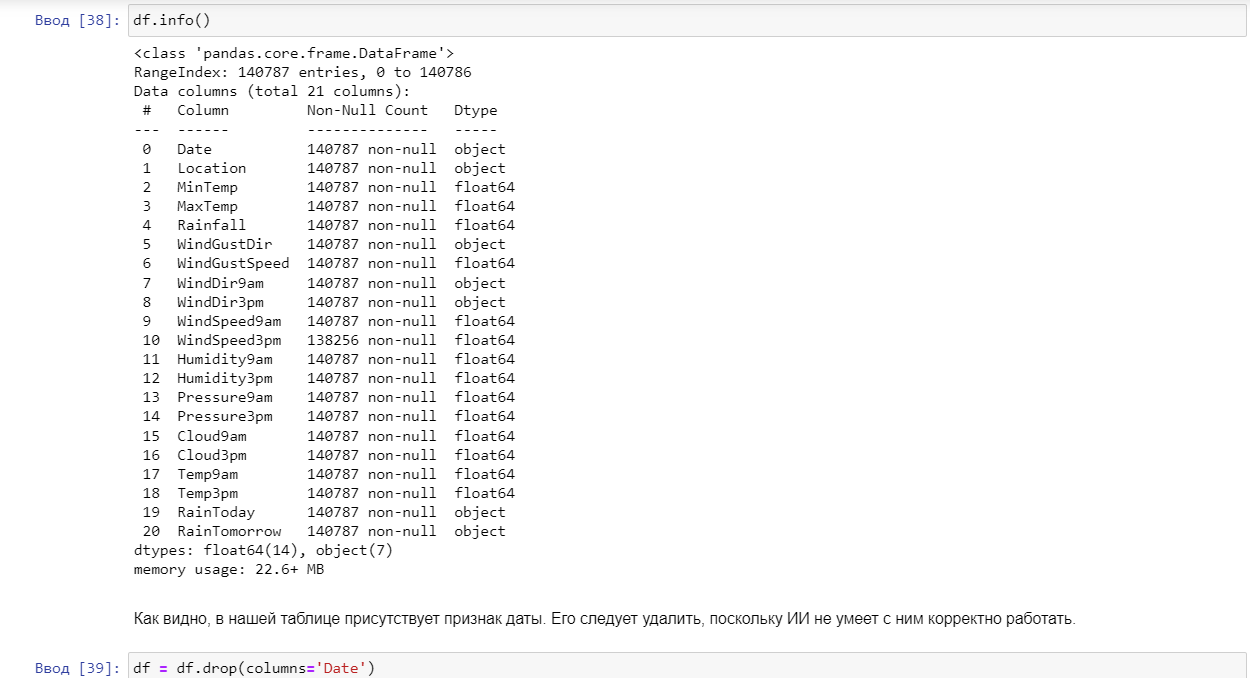


Рисунок 25 – Информация о таблице и удаление даты

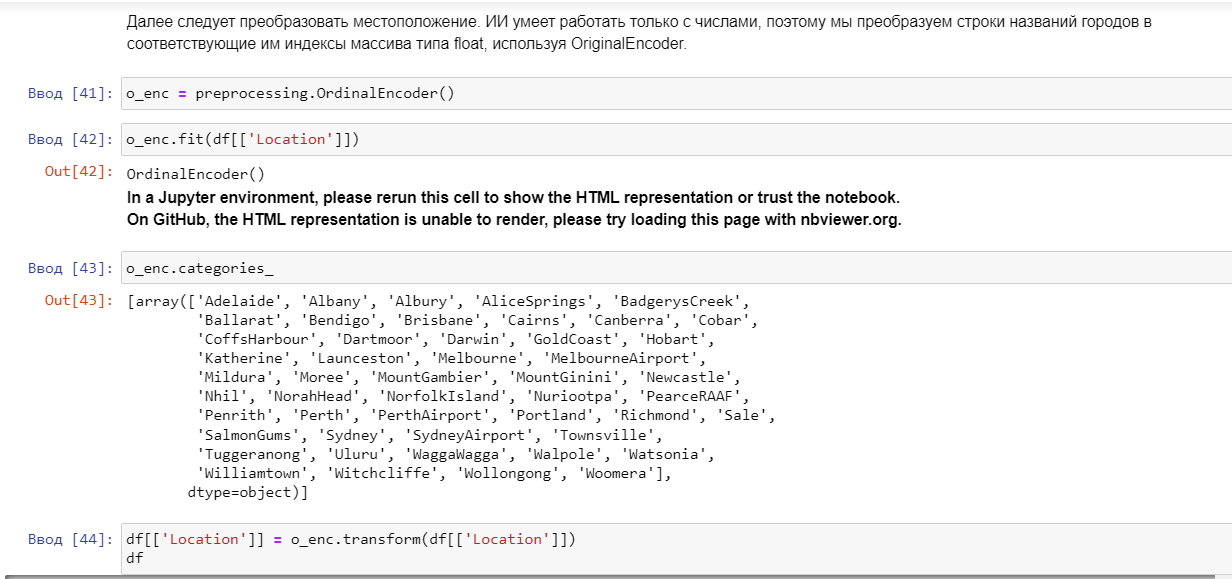


Рисунок 26 – Преображение местоположения

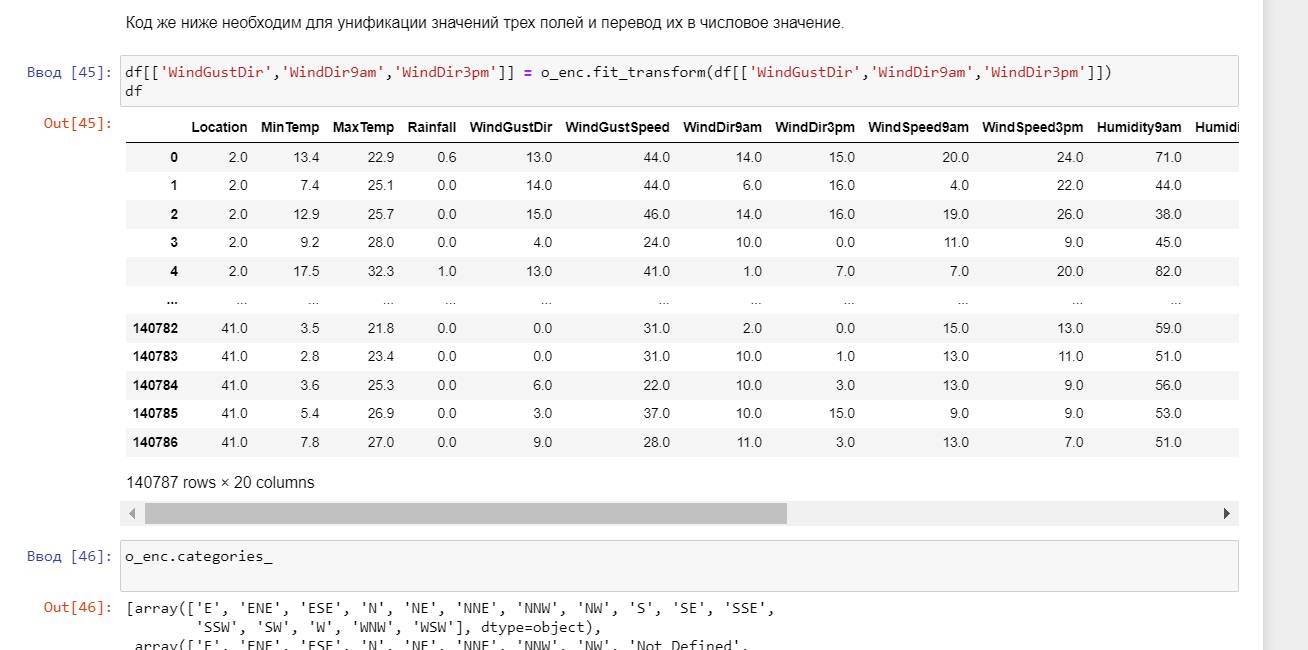


Рисунок 27 – Унификация полей

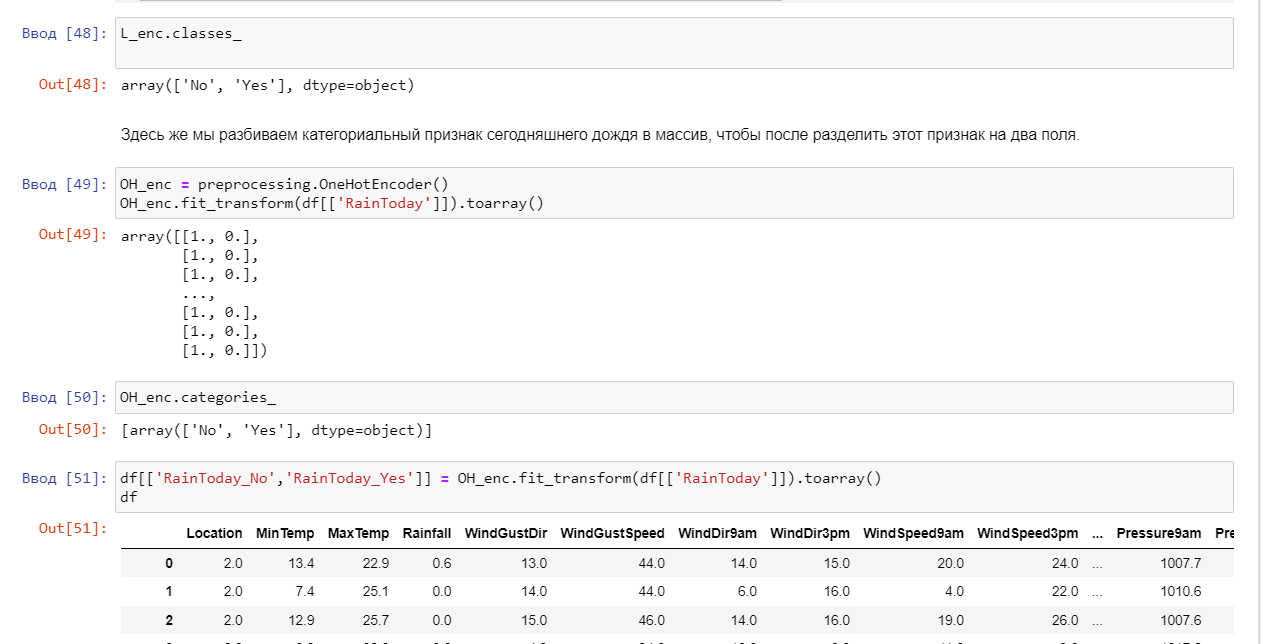


Рисунок 28 – Разбивка категориального признака

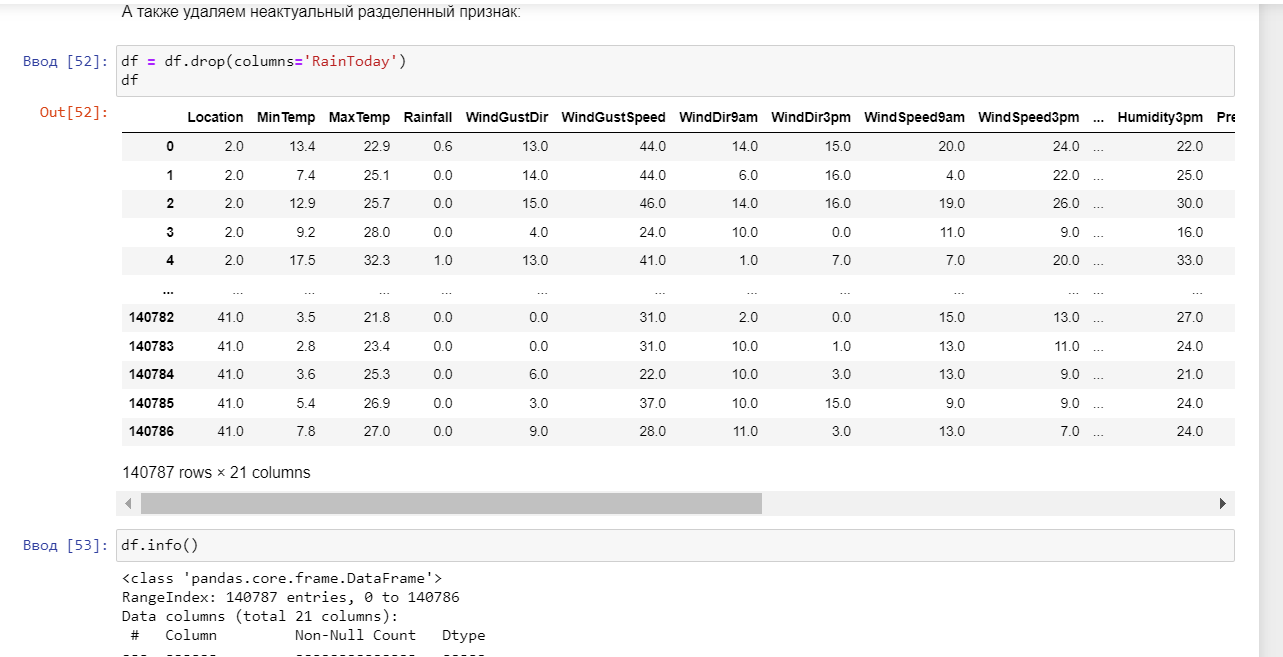


Рисунок 29 – Удаление неактуального признака



Рисунок 30 – Скейл данных

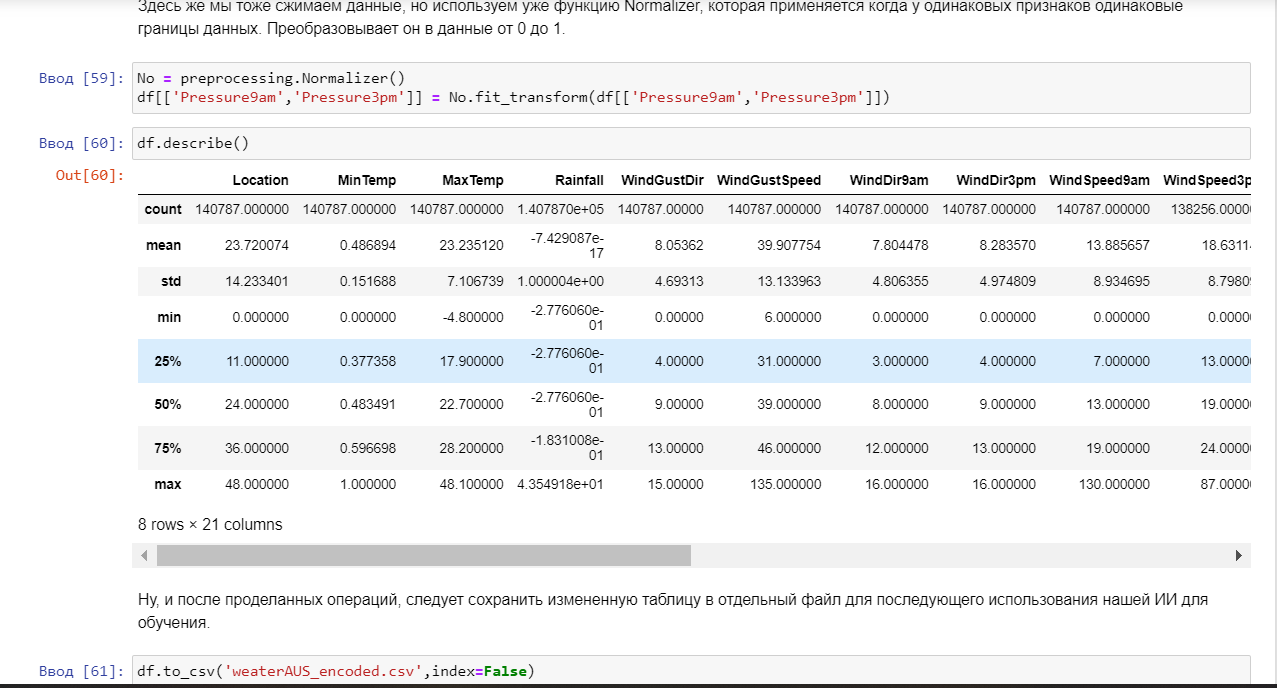


Рисунок 31 – Скейл данных и сохранение в файл

Вторая часть:

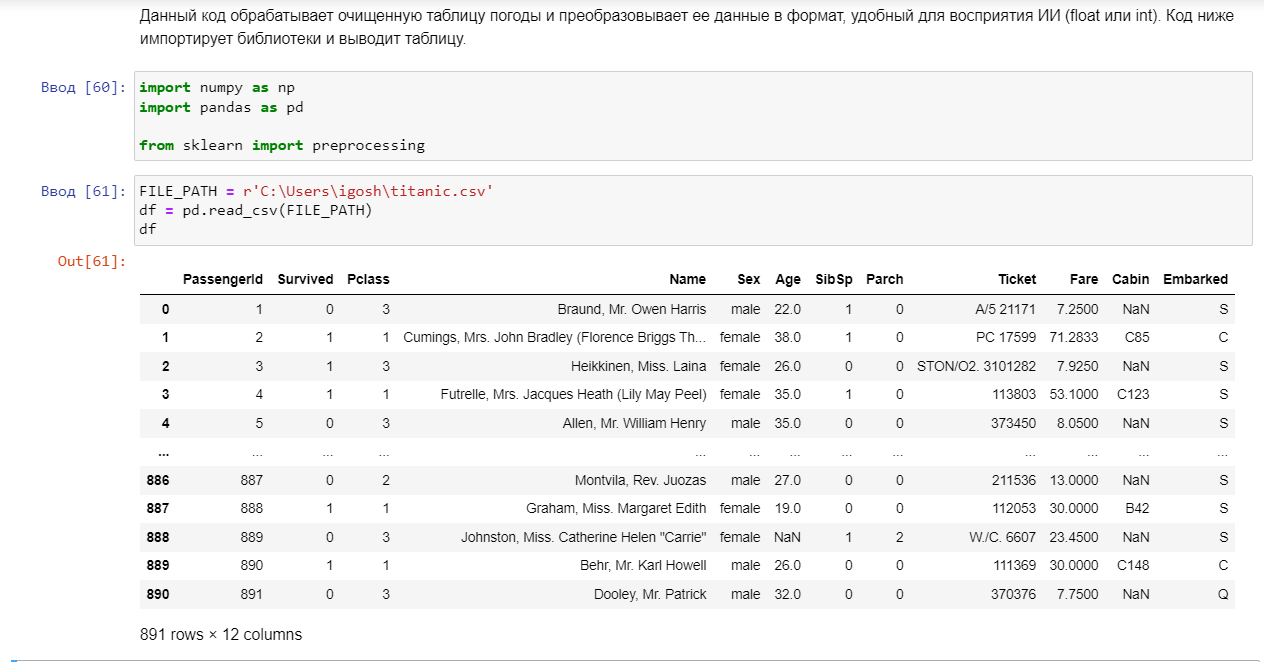


Рисунок 32 – Импорт библиотек и чтение таблицы

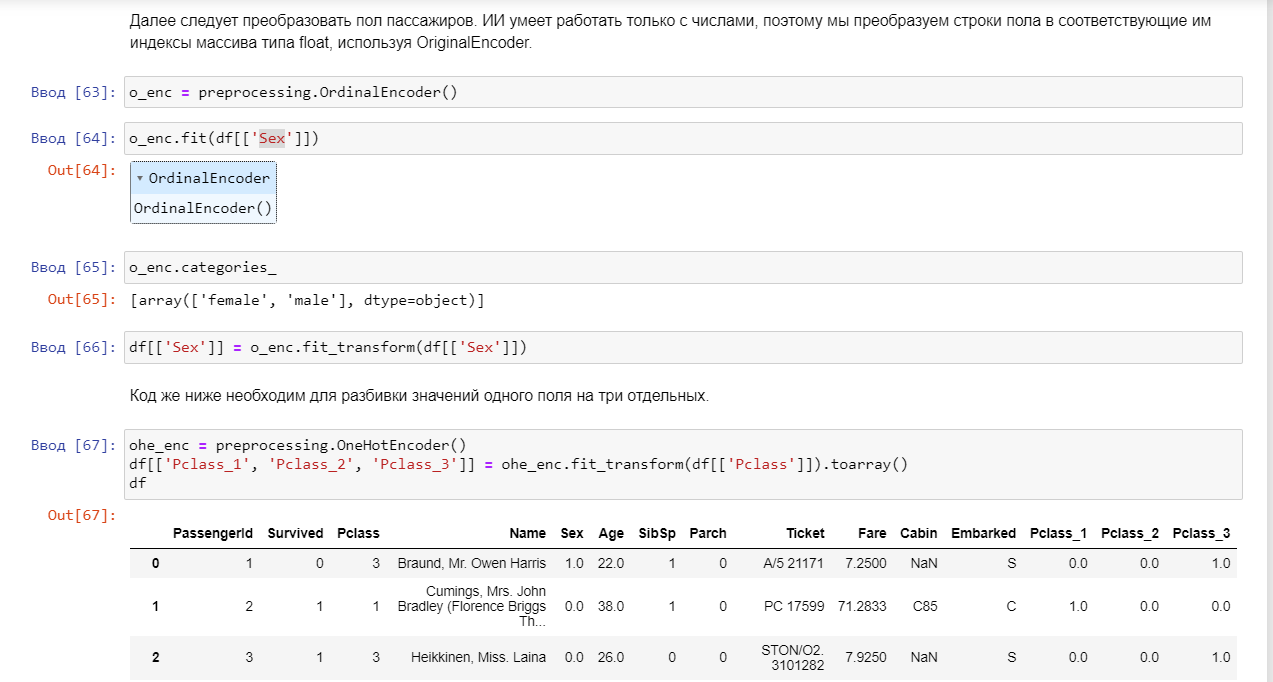


Рисунок 33 – Преобразование пола и разбивка значений

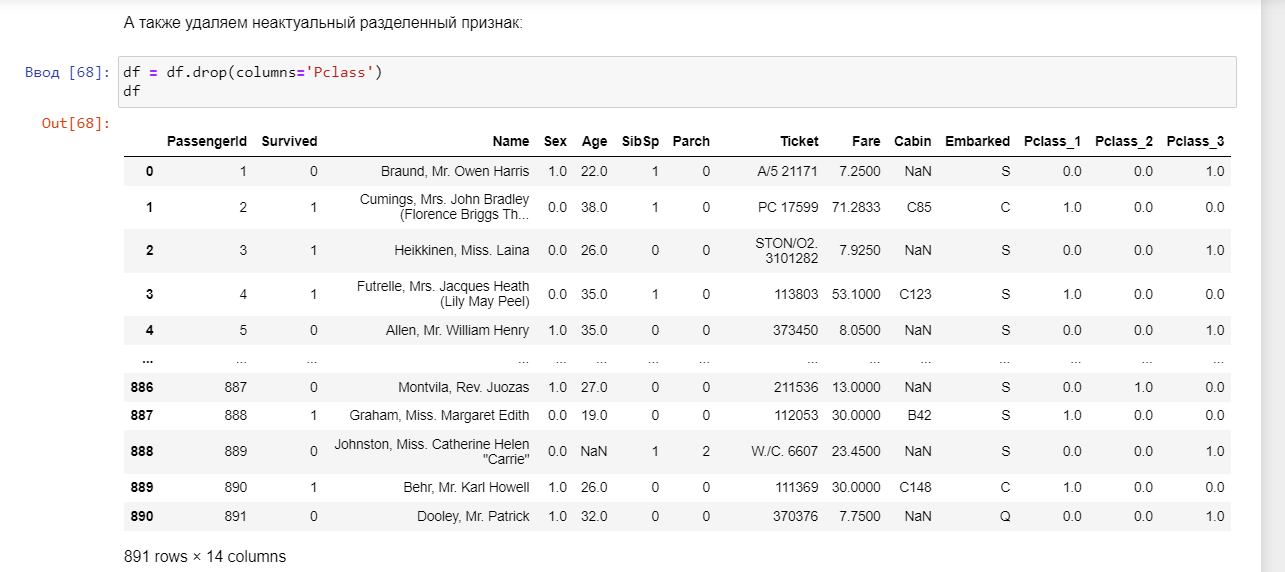


Рисунок 34 – Удаление неактуального признака



Рисунок 35 – Перевод категориального признака в числа

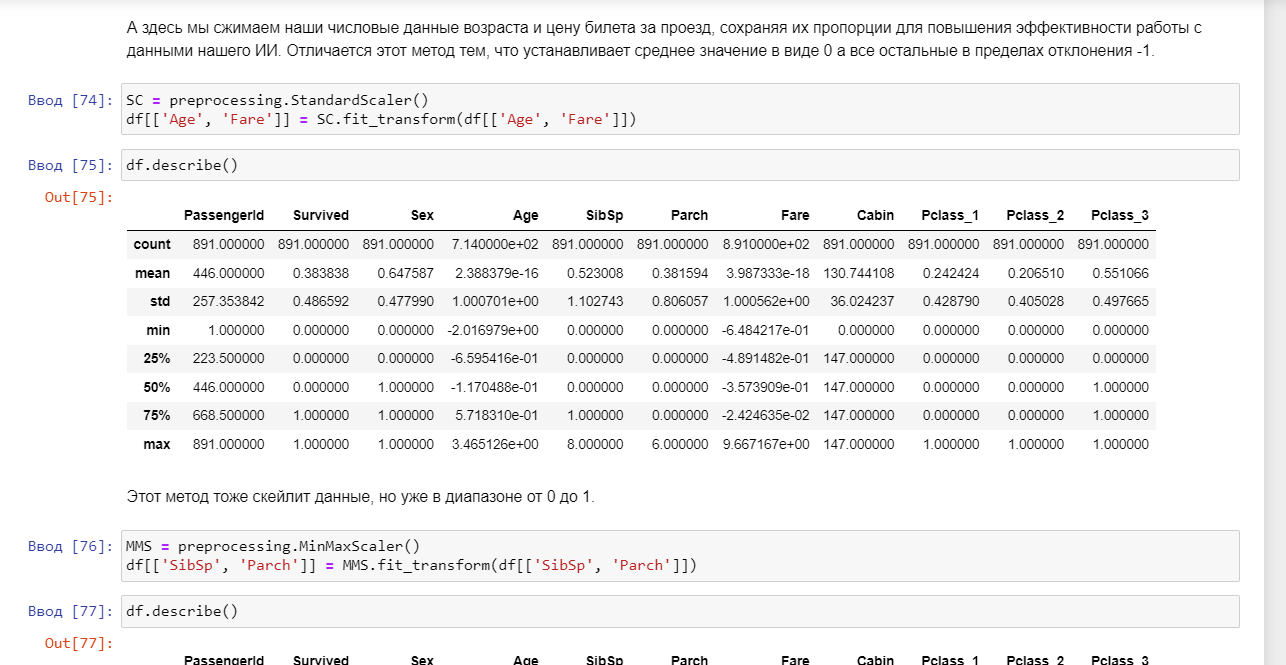


Рисунок 36 – Сжатие данных

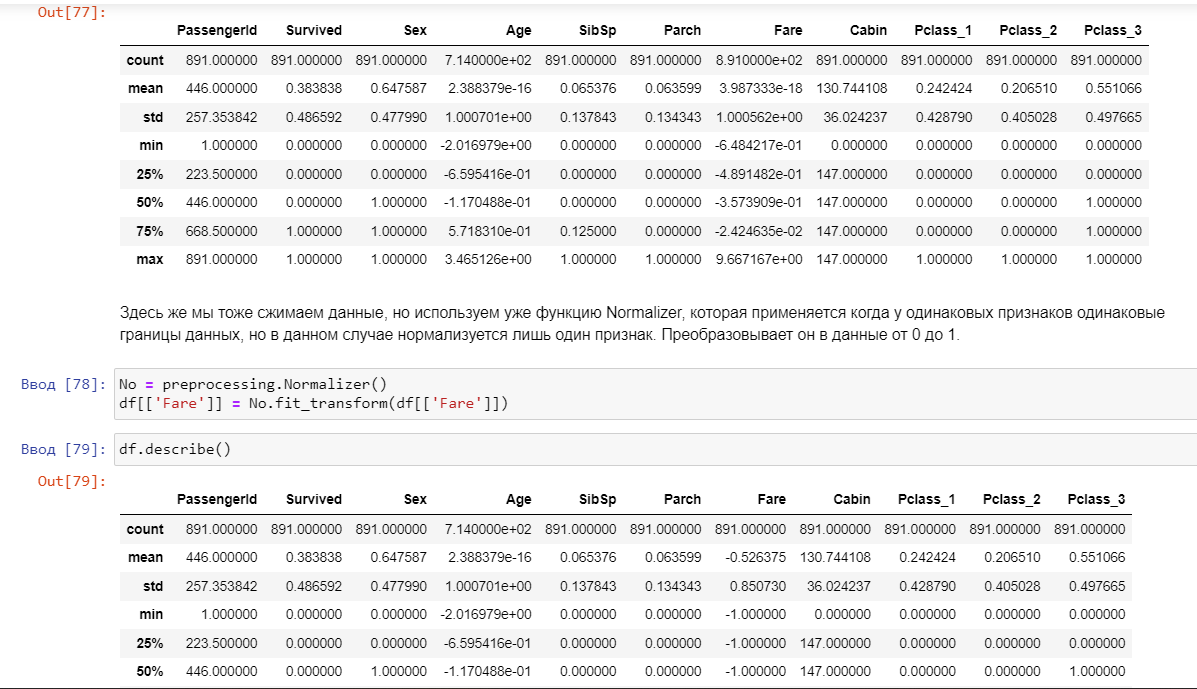


Рисунок 37 – Сжатие данных через нормализацию

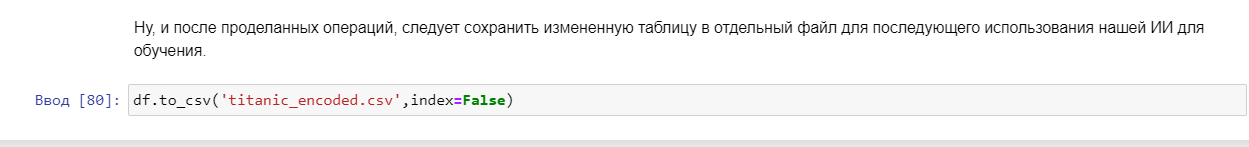


Рисунок 38 – Сохранение таблицы в файл

Вывод: в ходе выполнения практической работы были закодированы и преобразованы признаки двух таблиц с различными темами и несколькими примерами.

# ПРАКТИЧЕСКАЯ РАБОТА №4

«Дерево решений, метрики и классификация»

Цель работы: Обучить модель "Дерево решений" и проверить её точность.

Что необходимо сделать:

1. Разбить данные по KFold;
2. Необходимо обучить модель "Дерево решений" без использования гиперпараметров (настроек);
3. Снова обучить модель "Дерево решений" но уже с использованием гиперпараметров (настроек);
4. Попробовать другие две другие модели классификации.
5. Вывести метрики каждой модели и расписать их.
6. Выбрать наилучшую и написать вывод по модели.

Для начала следует импортировать библиотеки:

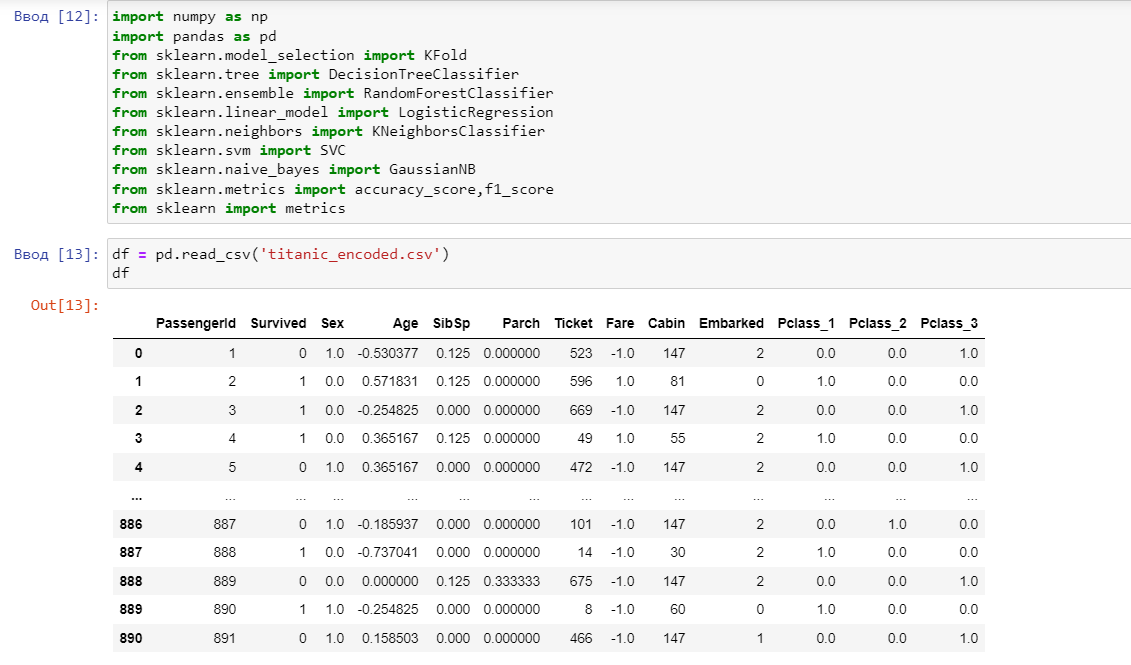


Рисунок 39 – Импорт библиотек и таблицы

После этого необходимо разбить наши данные на некоторые переменные а также разделить данные на 5 частей, для последующего разделения на тестируемые и обучающие данные.

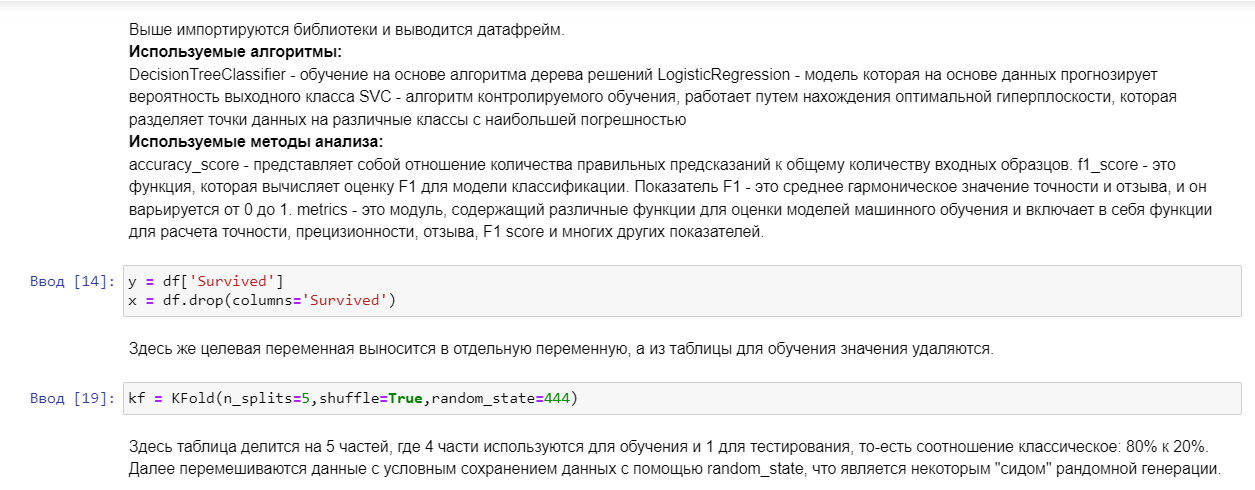


Рисунок 40 – Описание используемых алгоритмов и методов и подготовка таблицы

Далее идет непосредственно само обучение ИИ:

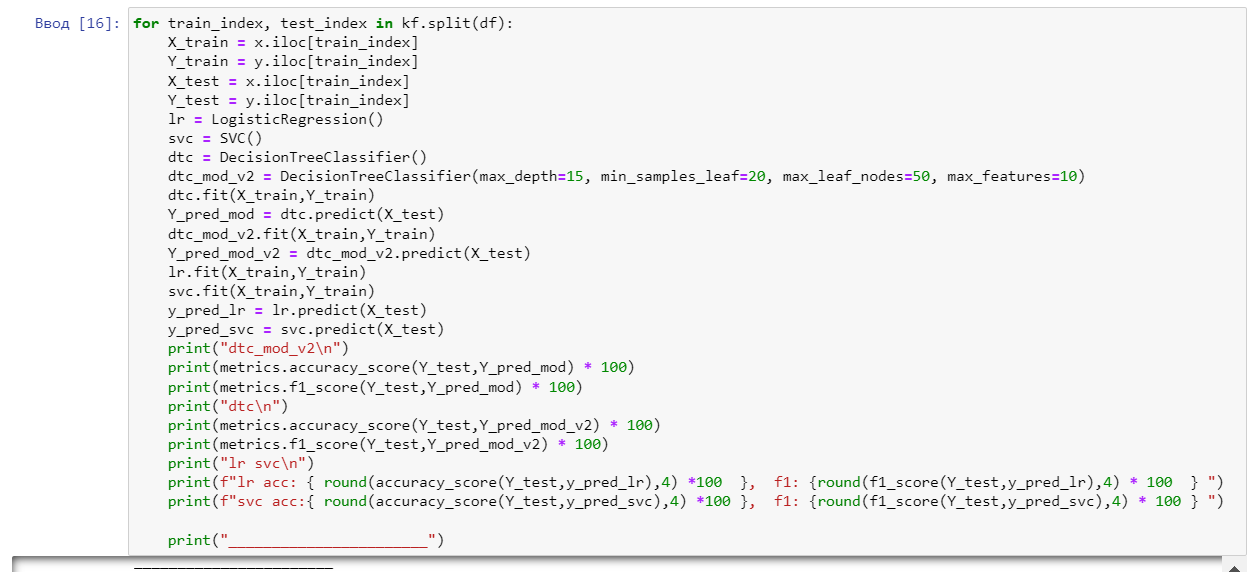


Рисунок 41 – Обучение ИИ

Результаты обучения:

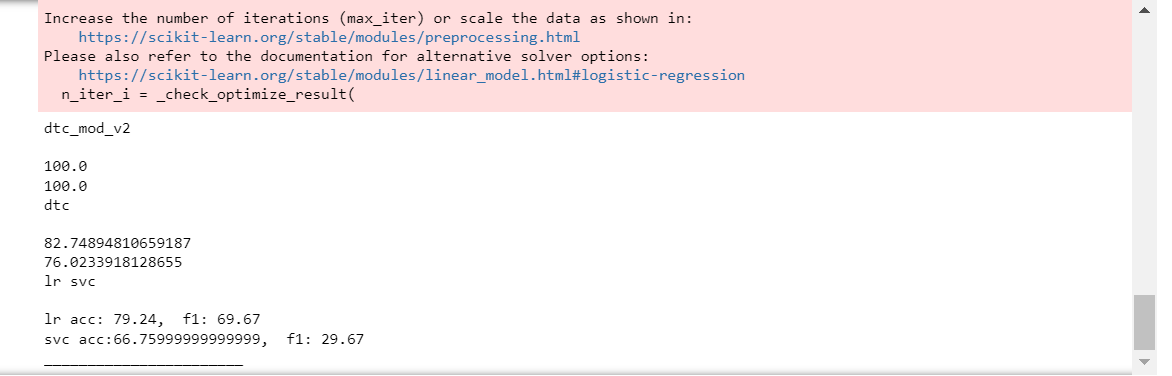


Рисунок 42 – Результаты вычислений ИИ и их предсказаний

Выводы по обучению:

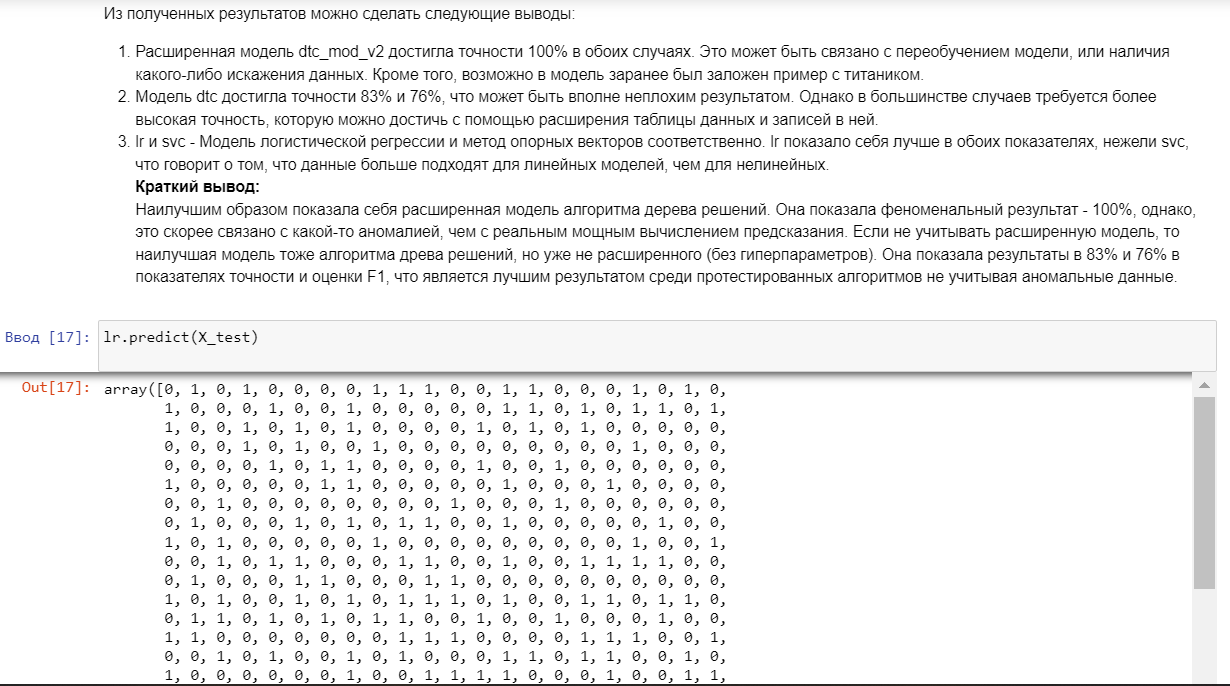


Рисунок 43 – Вывод и список предугадываний

Описание работы кода:

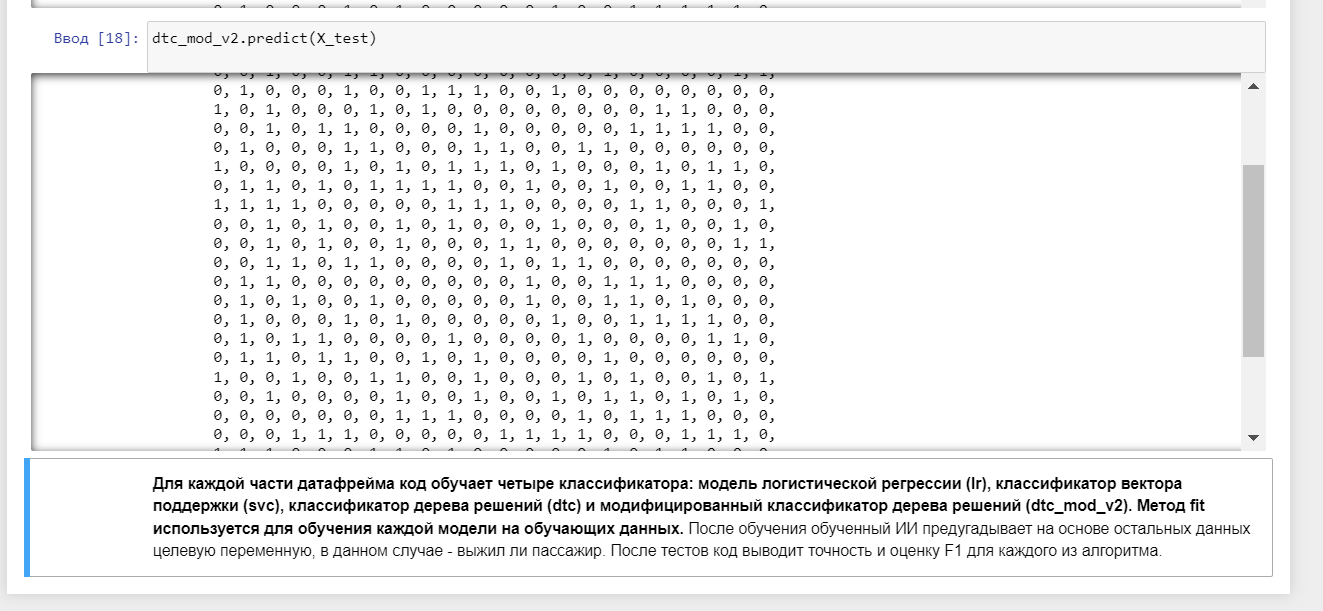


Рисунок 44 – Описание работы кода и вывод предугаданий

Вывод: В ходе выполнения практической работы была обучена модель, используя алгоритм дерева решений.

# ПРАКТИЧЕСКАЯ РАБОТА №5

## «Регрессия и метрики»

Цель работы: Обучить модели на решение регрессионной задачи.

Для этой работы был выбран отдельный датафрейм, найти который можно по ссылке: <https://www.kaggle.com/datasets/mexwell/student-scores>.

Датафрейм содержит в себе информацию об учениках и их оценках по различным предметам. Угадывать же в данном случае ИИ придется оценку по физике. Но, изначальный датафрейм не подходит для обучения, а потому его нужно закодировать. Изначальный датафрейм:

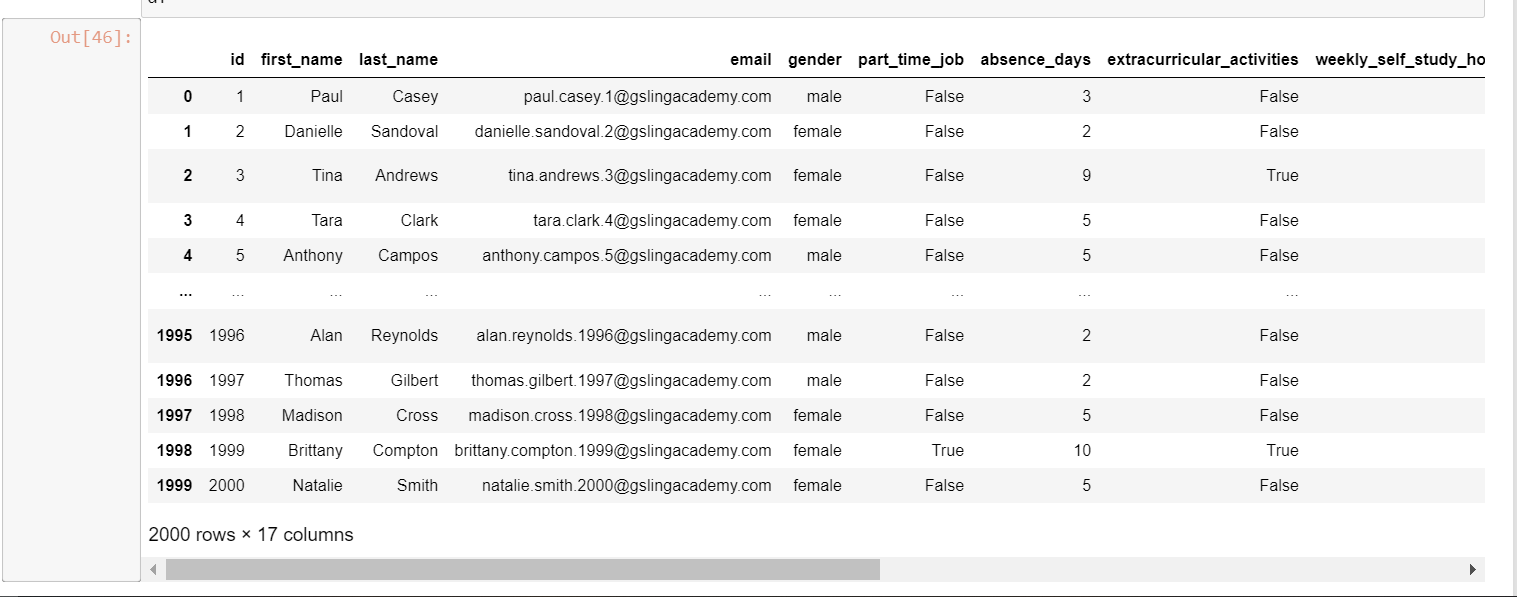


Рисунок 45 – Изначальный датафрейм

Датафрейм после обработки:

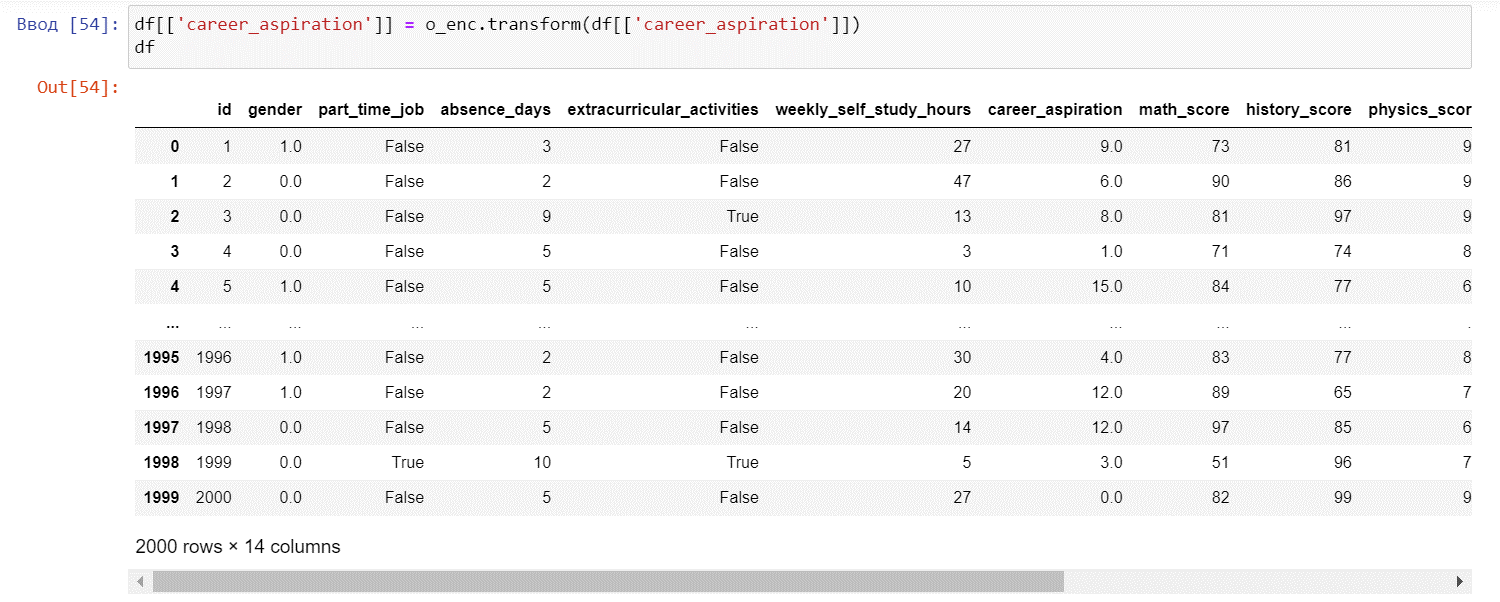


Рисунок 46 – Очищенный датафрейм

Начнем непосредственную работу:

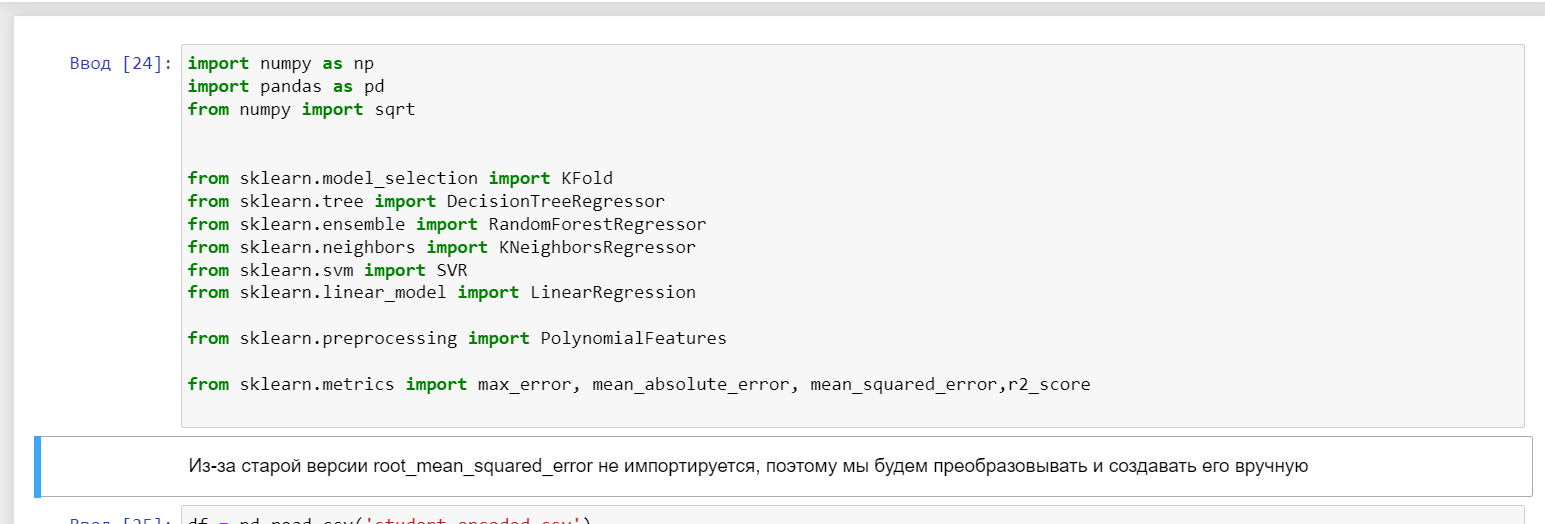


Рисунок 47 – Импорт библиотек

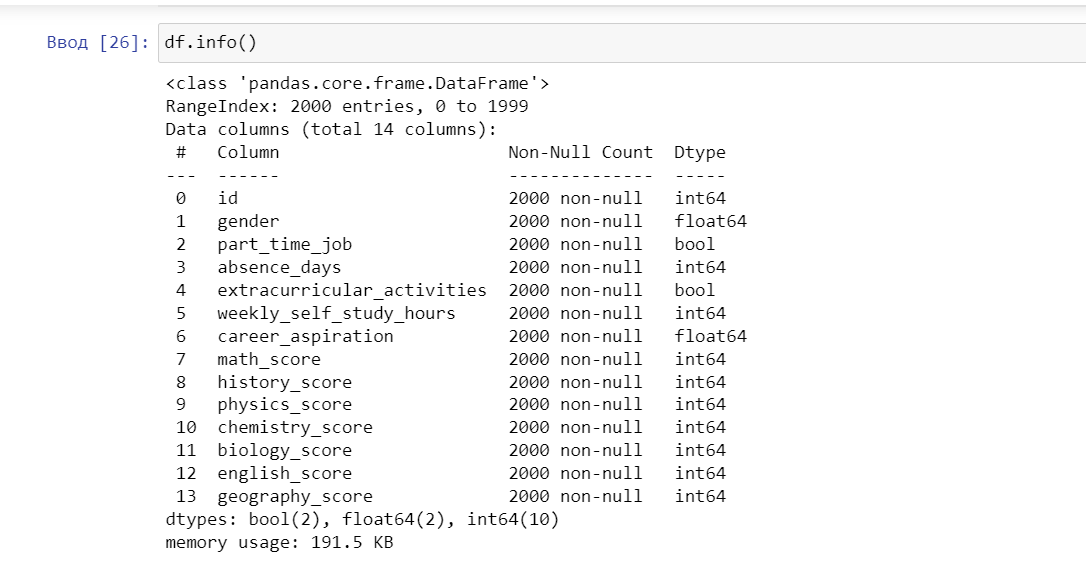


Рисунок 48 – Информация о датафрейме

Разделяем таблицу на части:

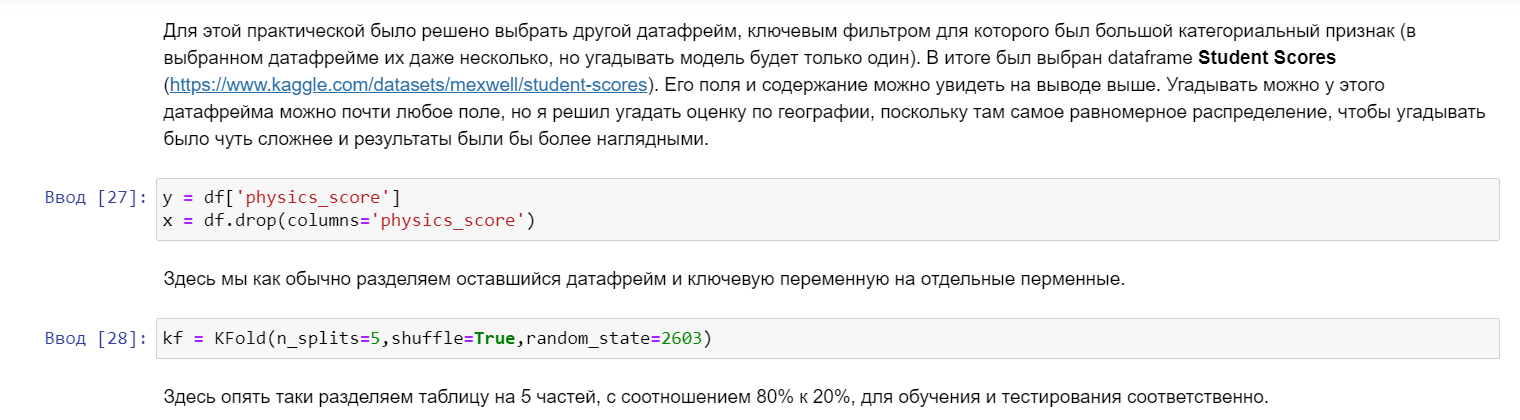


Рисунок 49 – Подготовка таблицы для обучения

А теперь и обучаем сам ИИ.



Рисунок 50 – Обучение ИИ



Рисунок 51 – Метрики обучения

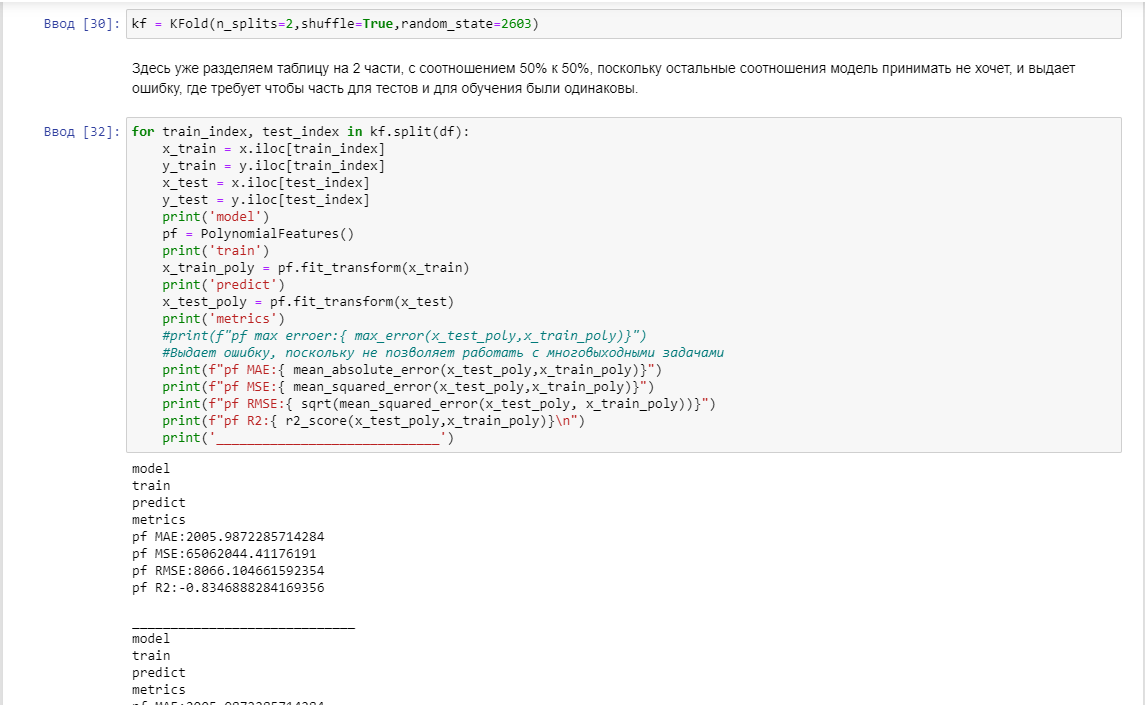


Рисунок 52 – Обучение ИИ и метрики

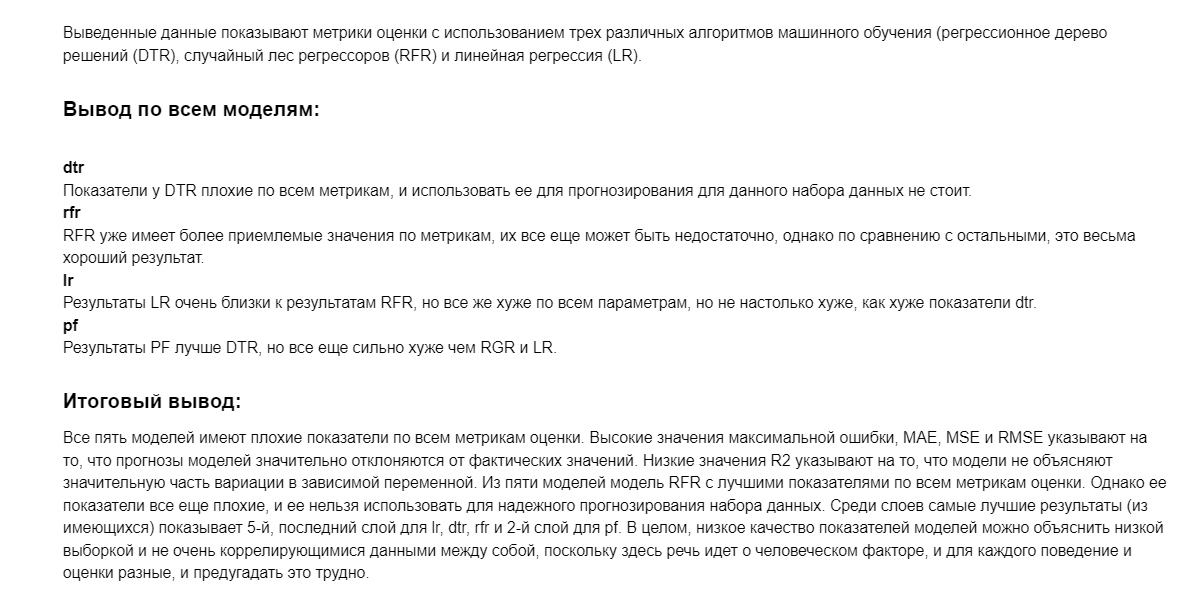


Рисунок 53 – Вывод по обучению

Вывод: в ходе выполнения практической работы, было обучено несколько моделей, используя 4 различных алгоритма, выполнены метрики и сделан анализ и вывод по ним.