**Контрольная работа**

**по дисциплине «Обработка и визуализация информации»**

**Тема**: Обработка и визуализация данных на Python с использованием датасета Boston Housing

**Цель работы:**

1. Изучение обработки и визуализации данных в Python.
2. Закрепление навыков:

• Загрузки и предобработки данных в Pandas.

• Построения статистических графиков (Matplotlib, Seaborn).

• Анализа корреляций и группировки данных.

**Номер варианта,** чтобы подписать на титульнике – (последняя цифра пароля +1)

**Для определения варианта задания** в таблицах 1,2,3 использовать последнюю цифру пароля.

**Задание:**

* Провести анализ датасета Boston Housing, включая визуализацию распределений, корреляций и группировку по заданным критериям.
* Сформировать отчет, куда занести результаты по каждому заданию.

Для этого надо выполнить 6 отдельных заданий:

* **Задание 1**: Написать программный код на базе файла «Контрольная.pynb» для подключения набора данных (датасет), который загружается по ссылке

″<http://lib.stat.cmu.edu/datasets/boston>″. Открыв по ссылке датасет, вы сможете ознакомиться с набором данных (наименования полей, типы значений, есть ли пустые значения).

**Программный код** должен содержать:

* Подключение датасета и создание массива *data* с названиями полей датасета и массива *target* со значениями датасета.

**В отчет** вставить описание набора данных, с которым будете работать в следующих заданиях (назначение датасета, какие данные там полей хранятся (не просто как это указано в датасете, а саму суть-назначение) и типы значений.

**Внимание:** Насколько грамотно проанализируете, настолько грамотно сможете подобрать аппарат для статистического анализа и формы вывода результатов анализа.

* **Задание 2:** Написать программный код, в котором выведите несколько строк из датасета в виде таблицы.

**Программный код** должен содержать:

* Cоединие *data* и *target* в DataFrame под названием boston.
* Вывод нескольких строк в виде таблицы

**В отчет** вставить скрин с программным кодом, скрин результата вывода таблицы.

* Задание 3: Написать программный код, в котором датасет boston проанализируете, построив матрицу корреляции (.corr()) в виде тепловой карты (sns.heatmap).

**В отчет** вставить скрин с программным кодом и результат его работы в виде тепловой карты.

* Задание 4: Проанализировать для датасета два заданных значения и их зависимость (таблица 1):
* Определить, что в датасете означает каждое значение.
* Провести анализ заданной корреляции этих значений и интерпретировать его, т.е словесно вывести, что на основании такого значения корреляции следует, что …

**Внимание:** Программный код не пишите, а учитесь анализировать наборы данных, их зависимость и делать выводы.

**В отчет** вносите описание назначения заданных значений, интерпретацию зависимости этих значений при заданном показателе корреляции.

Таблица 1 (брать вариант по последней цифре пароля)

|  |  |
| --- | --- |
| **Последняя**  **цифра пароля** | **Значения и корреляция между ними** |
| 0 | INDUS и LSTAT = 0.6 |
| 1 | DIS и AGE = -0.75 |
| 2 | NOX и CRIM = 0.42 |
| 3 | RM и MEDV = 0.7 |
| 4 | TAX и RAD = 0.91 |
| 5 | PTRATIO и LSTAT = 0.53 |
| 6 | ZN и DIS = -0.45 |
| 7 | B и MEDV = 0.33 |
| 8 | CHAS и CRIM = -0.18 |
| 9 | AGE и TAX = 0.64 |

* **Задание 5:** Написать программный код для построения графика распределения целевой переменной (таблица 2) при анализе данных датасет boston.

**В отчет** вставить скрин с программным кодом и результат его работы в виде графика распределения целевой переменной.

Таблица 2 (брать вариант по последней цифре пароля)

|  |  |
| --- | --- |
| **Последняя**  **цифра пароля** | **Целевая переменная** |
| 0 | MEDV |
| 1 | CRIM |
| 2 | AGE |
| 3 | RM |
| 4 | PTRATIO |
| 5 | LSTAT |
| 6 | DIS |
| 7 | NOX |
| 8 | INDUS |
| 9 | TAX |

* **Задание 6**: Написать программный код, в котором необходимо:
* Сгруппировать дома по критерию приближенности к реке Чарльз.
* Определить по полученному результату, в какой группе выше показатель, заданный в таблице 3.

**В отчет** вставить скрин с программным кодом, результат в виде рассчитанного значения показателя и ваш вывод, что означает это вычисленное значение.

Таблица 3 (брать вариант по последней цифре пароля)

|  |  |
| --- | --- |
| **Последняя**  **цифра пароля** | **Показатель** |
| 0 | MEDV |
| 1 | CRIM |
| 2 | AGE |
| 3 | RM |
| 4 | PTRATIO |
| 5 | LSTAT |
| 6 | DIS |
| 7 | NOX |
| 8 | INDUS |
| 9 | TAX |

**Отчет должен содержать**:

* + Тема, ФИО, группа, вариант, дата.
  + Цель работы
  + Задание на контрольную работу.
  + Описание хода работы.
  + Блок-схема выполнения (текст + код).
  + Графики с подписями осей и заголовками (3,5,6 задание).
  + Выводы по каждому заданию.
  + Заключение (личные выводы, умозаключения).

**Пример формата отчёта и какие требования надо соблюдать при его оформлении приведены в приложении 1.**

**Критерии оценки контрольной:**

• Корректность кода (50%).

• Качество визуализаций (30%).

• Глубина выводов (20%).

**Рекомендации по выполнению:**

Для того чтобы приступить к выполнению контрольной работы, необходимо перейти на сайт Google Colab (<https://colab.google/> ) и **загрузить** файл «Контрольная.pynb». Пример загрузки файла показан на рисунке 1.

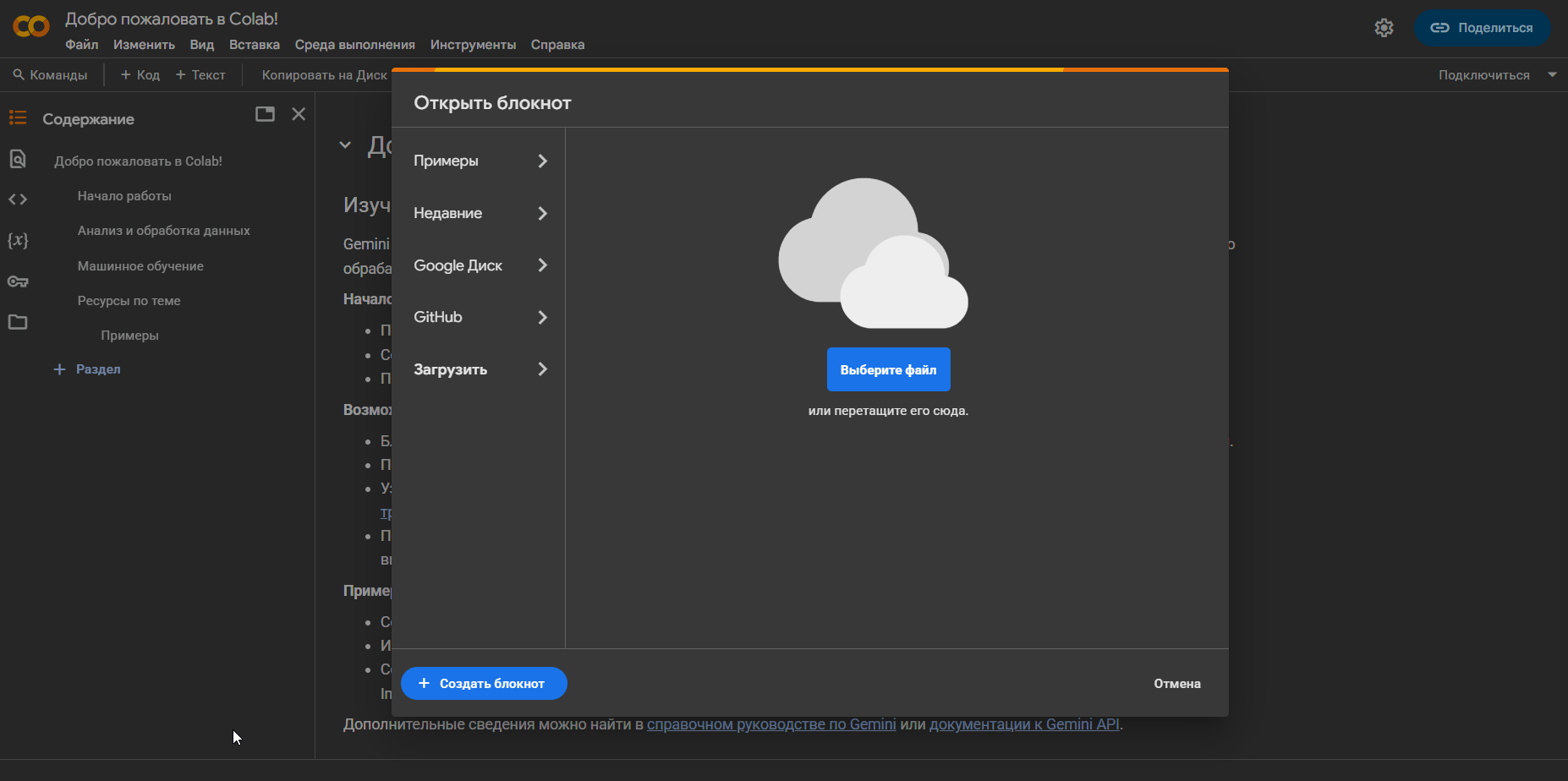


Рисунок 1 – Загрузка исходного файла для начала контрольной работы

**Выполнение заданий возможно после изучения всех лекций-презентаций. Внимательно изучите лекционный материал, там описаны схожие примеры.**

Приложение 1

Отчёт по контрольной работе

«Обработка и визуализация данных на Python с использованием датасета Boston Housing»

Выполнил: Фамилия Имя Отчество, XX-00 (последние 2 цифры пароля), вариант 01

25.25.2025 г.

1. Цель работы

Цель работы следует переписать из методических указаний к контрольной работе.

2. Задание на контрольную работу

Задание на контрольную работу также следует переписать из методических указаний к контрольной работе.

3. Ход работы

Краткое описание задач и подготовка к работе

3.1. Задание 1

Раздел всегда начинается только текстовым блоком (а не рисунком).

Если описание хода работы объемное его логично разбить на части, это можно сделать с помощью выделения первого в абзаце предложения или словосочетания.

Выделение предложения или словосочетания. Такое выделение выполняется в тексте в начале абзаца. Оно позволяет разграничить несколько абзацев и структурировать описание. Также допустимо выделение подчеркиванием в середине текста, если на это нужно обратить особое внимание.

На каждый рисунок в тексте до рисунка должна быть ссылка на него (рисунок 1). Эта ссылка может быть в скобках или без скобок по тексту. У рисунка и подписи рисунка не должно быть отступов красной строки.

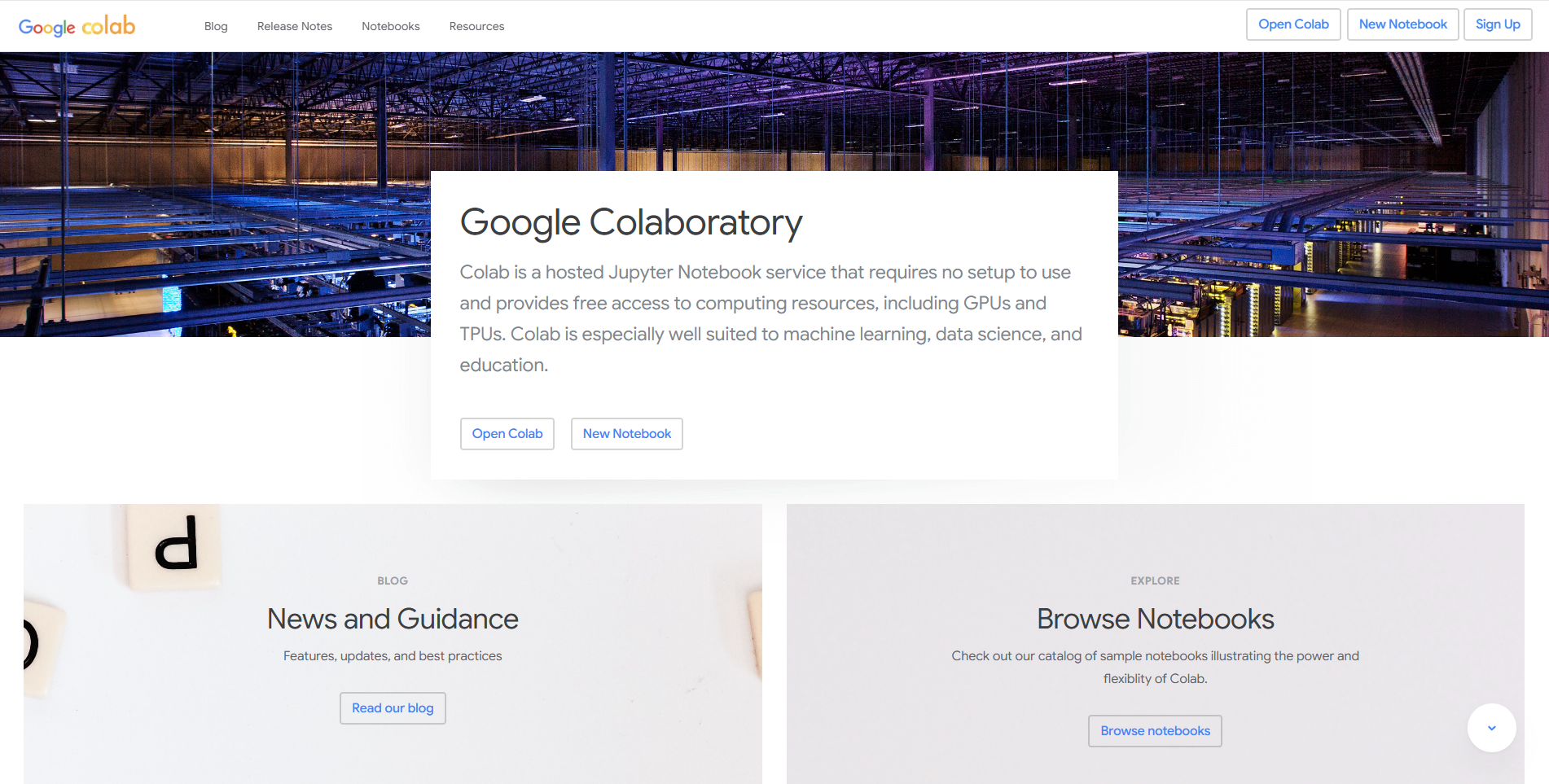
В конце выполнения каждого из заданий должен быть краткий вывод.

3.2. Задание 2

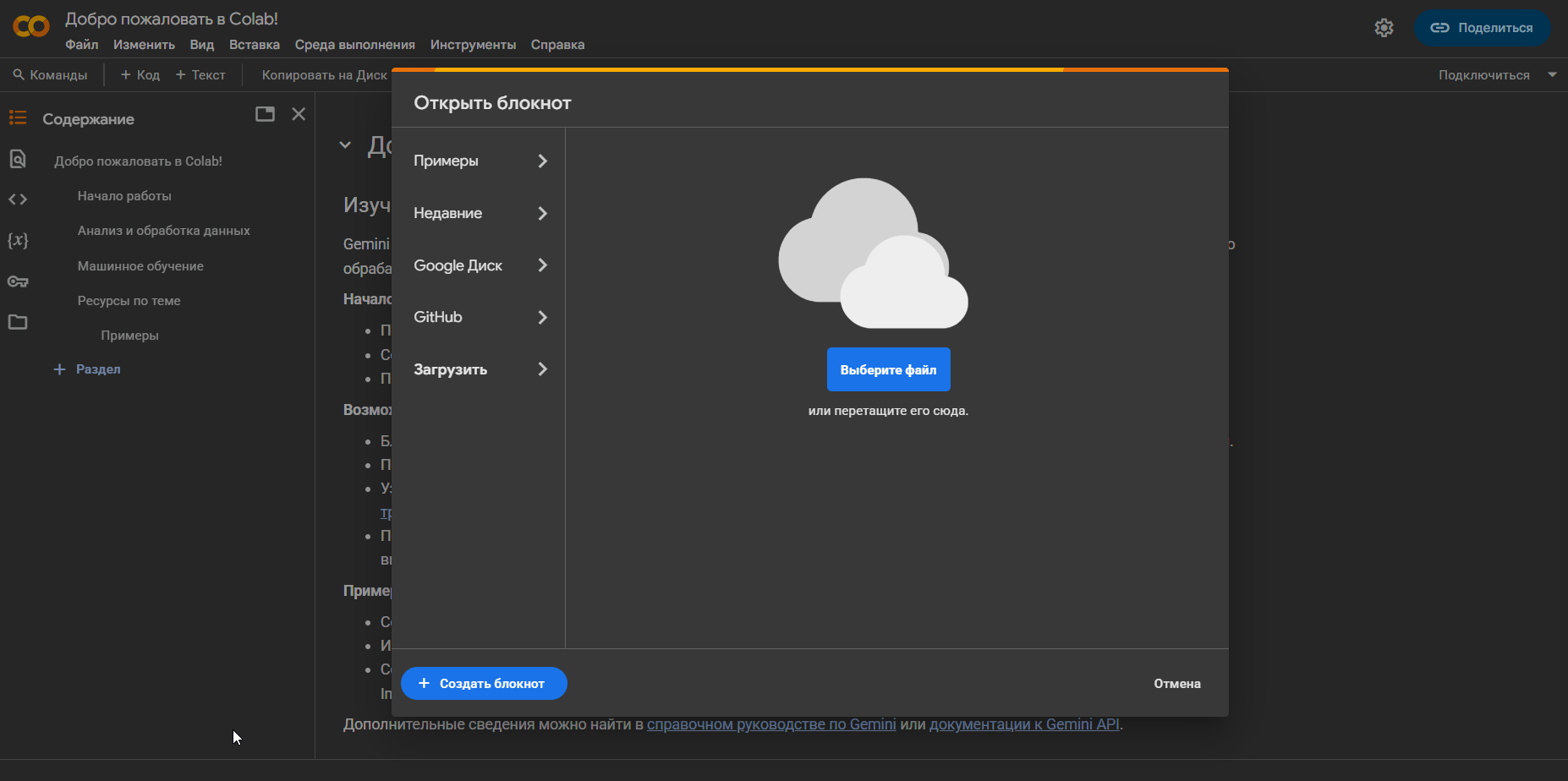
В документе не должно быть заголовков/подзаголовков или подписей рисунков, оторванных от текста на другую страницу. Также нежелательно наличие неиспользуемого пространства документа (например, четверть страницы осталась пустой из-за того, что крупный рисунок расположен на следующей странице).

Размер изображения следует подбирать самостоятельно исходя из содержания изображения. Если в одном поле приведены два или несколько изображений, их следует нумеровать буквами, как на рисунке 1. В тексте обязательно должна быть ссылка на каждый рисунок (если необходимо, можно сделать ссылку на его часть, например, на рисунок 1,б).

Разные изображения не должны сливаться в одно, их следует разграничить пробелом.



а)



б)

Рисунок 1 – Начало работы в Google Colab: а) отображение начальной страницы сайта; б) отображение окна загрузки блокнота

Раздел всегда заканчивается текстовым блоком, и никогда – таблицей или рисунком с его подписью. Т.е. после рисунка должен быть минимум один абзац (хотя бы из пары предложений). Здесь нужно подвести итог тому, что написано в данном разделе.

4. Заключение

Личные выводы.