

## **Лабораторная работа 1: Анализ данных и построение графиков в Python**

Цель работы: Анализ временных данных и построение графиков в Python.

Ход работы:

Для начала нам понадобится ПО Anaconda, ниже представлена инструкция по его установке и скачиванию.

### **1. Загрузите Anaconda:**

Перейдите на официальный сайт Anaconda по адресу <https://www.anaconda.com/products/distribution> и выберите версию Anaconda, которую вы хотите установить (например, Anaconda Individual Edition). Обычно рекомендуется выбирать последнюю версию для вашей операционной системы (Windows 64-Bit).

### **2. Скачайте установщик:**

Нажмите на ссылку для загрузки установщика Anaconda. Вам предложат сохранить исполняемый файл (обычно с расширением .exe) на ваш компьютер.

### **3. Запустите установщик:**

После завершения загрузки установщика, найдите скачанный файл (например, Anaconda3-2021.05-Windows-x86\_64.exe) и запустите его, дважды кликнув по нему.

### **4. Начните установку:**

После запуска установщика Anaconda откроется окно приветствия. Нажмите "Next" (Далее), чтобы продолжить.

### **5. Прочитайте и принимайте лицензионное соглашение:**

Прочитайте лицензионное соглашение и, если вы согласны с ним, выберите "I Agree" (Я согласен) и нажмите "Next" (Далее).

### **6. Выберите установочное местоположение:**

Выберите местоположение для установки Anaconda. По умолчанию это будет **C:\Users\<ваше\_имя>\Anaconda3** (где <ваше\_имя> - ваше имя

пользователя). Если вы хотите изменить путь установки, вы можете это сделать здесь. Нажмите "Next" (Далее), чтобы продолжить.

#### 7. Выберите опции установки:

При установке Anaconda вы можете выбрать опции, такие как добавление Anaconda к переменной среды PATH и использование Anaconda как системного Python. Рекомендуется оставить все опции по умолчанию, если вы не знаете, что делаете.

#### 8. Установка начнется:

Нажмите "Install" (Установить), чтобы начать установку Anaconda. Процесс установки может занять несколько минут.

#### 9. Завершение установки:

После завершения установки, нажмите "Next" (Далее) и затем "Finish" (Готово).

Теперь надо открыть редактор кода Spyder, его иконка представлена на рисунке 1.

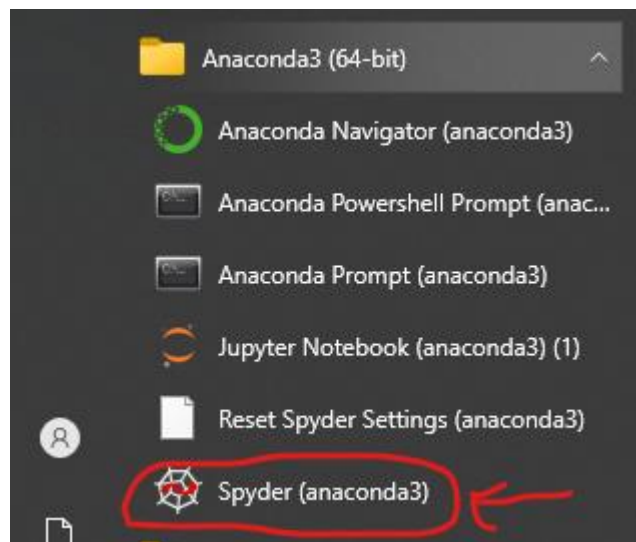


Рисунок 1 — Spyder

Важно!

После запуска spyder нужно будет сохранить файл в ту же папку, где находится ваш будущий файл с данными в формате XLSX. Порядок действий для сохранения файла представлен на рисунке 2.

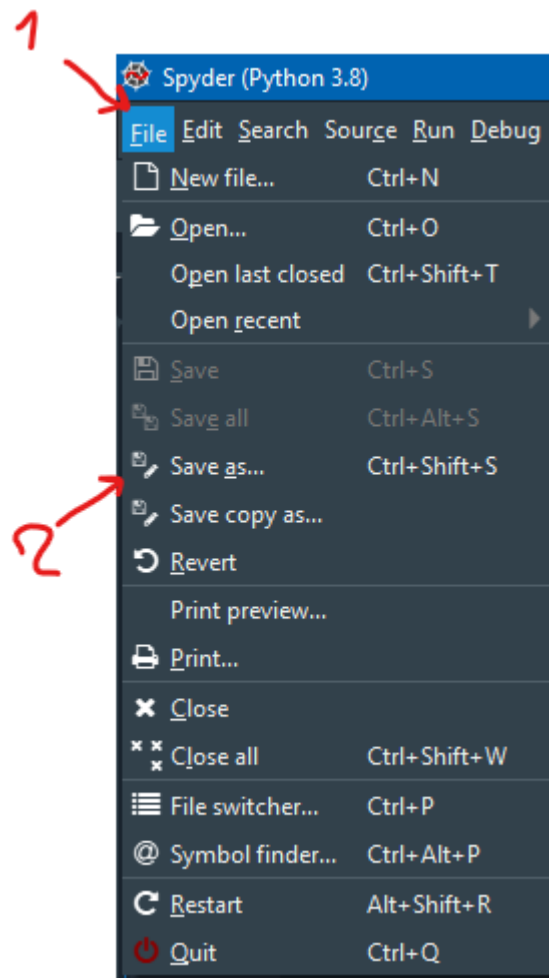


Рисунок 2 — Сохранение

Поздравляю! Теперь можно приступать ко 2 этапу!

Сбор/подготовка данных:

В данной работе в качестве примера будут использоваться временные данные о населении людей в каком нибудь городе. Для начала, нам необходимо собрать данные о населении в этом городе за последние годы. Давайте предположим, что мы найдем такие данные в виде XLSX файла на сайте городской статистики и используем библиотеку Pandas для их сбора. После сбора нужно открыть файл и привести его к виду, как показано на рисунке 3.

The screenshot shows the Microsoft Excel interface with the 'Главная' (Home) tab selected. The ribbon includes options for Font (Шрифт), Paragraph (Выравнивание), Numbers (Число), Styles (Стили), and Editing (Рецензирование). The active cell is H18. The data table is as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Год	Численность															
2	1899	1555															
3	1900	1556															
4	1901	1557															
5	1902	1558															
6	1903	1559															
7	1904	1560															
8	1905	1561															
9	1906	1562															
10	1907	1563															
11	1908	1564															
12	1909	1565															
13	1910	1566															
14	1911	1567															
15	1912	1568															
16																	
17																	

Рисунок 3 — Вид данных

Как можно заметить, в 1 строчке необходимо указывать именно атрибуты (названия ваших столбцов), это важно.

После приведения данных к нормальному виду создайте отдельную папку, где будет находиться ваш XLSX файл, сохраните его туда. Запомните этот путь к папке, он вам еще понадобится.

Теперь, когда данные подготовлены, давайте проведем базовый анализ, например, вычислим среднее значение, медиану и стандартное отклонение населения за это время. Для этого пишем код, как показано на рисунке 4.

The screenshot shows a Python script in a code editor. The script imports pandas and reads data from an Excel file. It then calculates the mean, median, and standard deviation of the population and prints the results.

```

1 import pandas as pd
2
3 # Загрузка данных из файла Excel
4 data = pd.read_excel('data.xlsx')
5
6 # Проведение анализа данных
7 mean_population = data['Численность'].mean()
8 median_population = data['Численность'].median()
9 std_deviation_population = data['Численность'].std()
10
11 print(f'Среднее значение населения: {mean_population}')
12 print(f'Медиана населения: {median_population}')
13 print(f'Стандартное отклонение населения: {std_deviation_population}')
14

```

Рисунок 4 — Начальный код

Методы `mean()`, `median()`, и `std()` применяются к объектам `DataFrame` или `Series` библиотеки `Pandas` и используются для вычисления среднего значения, медианы и стандартного отклонения соответственно. Вот как они работают:

`mean()`:

Метод `mean()` вычисляет среднее арифметическое значение для всех значений в `Series` или столбце `DataFrame`. Среднее значение - это сумма всех значений, деленная на количество значений.

`median()`:

Метод `median()` вычисляет медиану для всех значений в `Series` или столбце `DataFrame`. Медиана - это значение, которое находится в середине отсортированного списка значений. Если у вас есть четное количество значений, медиана будет средним значением двух центральных значений.

`std()`:

Метод `std()` вычисляет стандартное отклонение для всех значений в `Series` или столбце `DataFrame`. Стандартное отклонение измеряет разброс данных относительно их среднего значения. Большое стандартное отклонение указывает на большой разброс, а маленькое стандартное отклонение - на маленький разброс.

Заметьте, что значение “Численность” в 7, 8, 9 строчках на рисунке 4 соответствует моей таблице данных `XLSX` файла, сюда вы должны будете вписать свое название атрибута, которое хотите будете проанализировать.

Запустим код, для этого надо нажать на зеленый треугольник, который показан на рисунке 5.

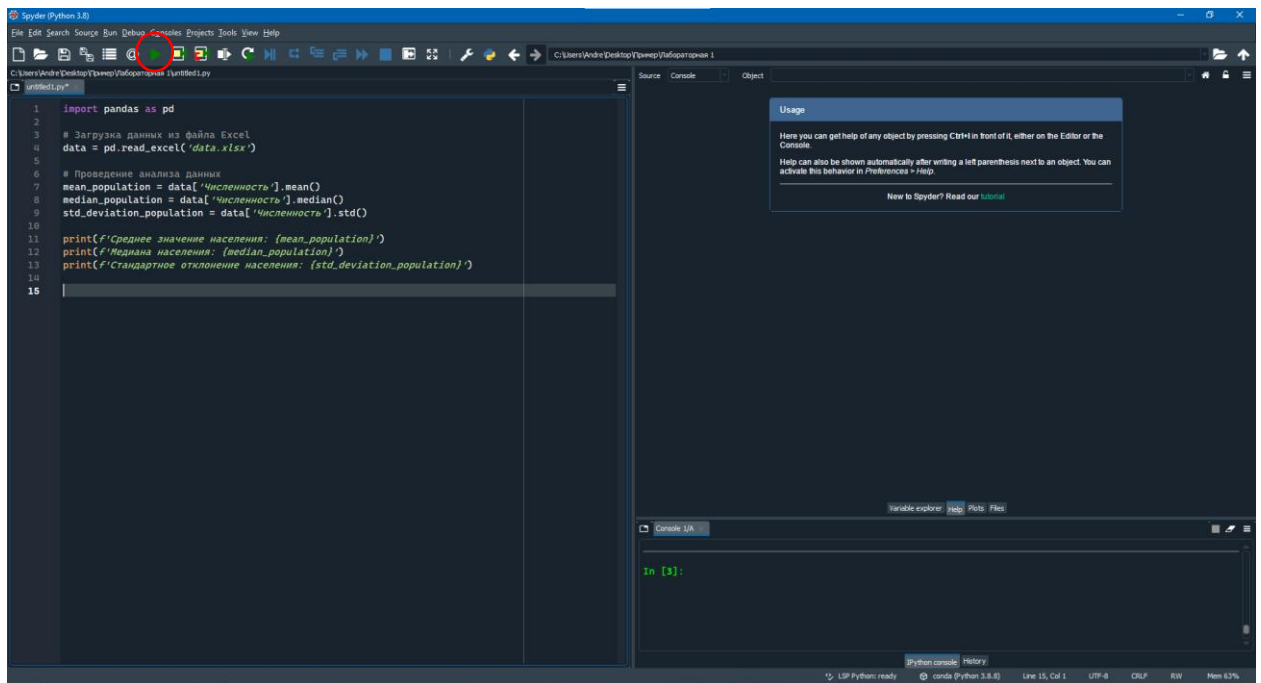


Рисунок 5 — Запуск кода

После запуска в консоли внизу справа будет написано о запуске файла а далее будут выданы данные, которые вы указали в Print, как показано на рисунке 6.

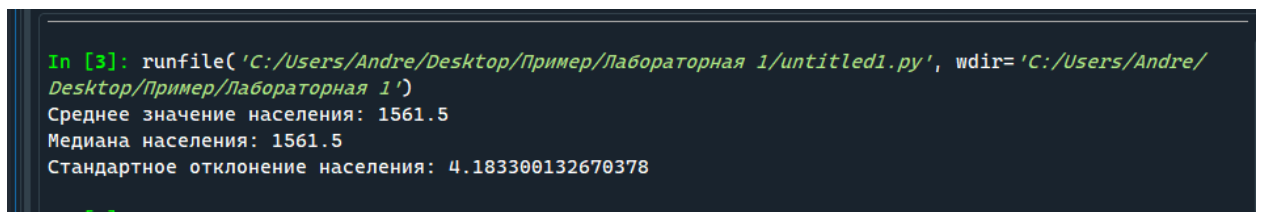


Рисунок 6 — Результат

Теперь давайте построим график. Для этого дополняем код, как показано на рисунке 7, чтобы весь код вышел, как показано на рисунке 8.

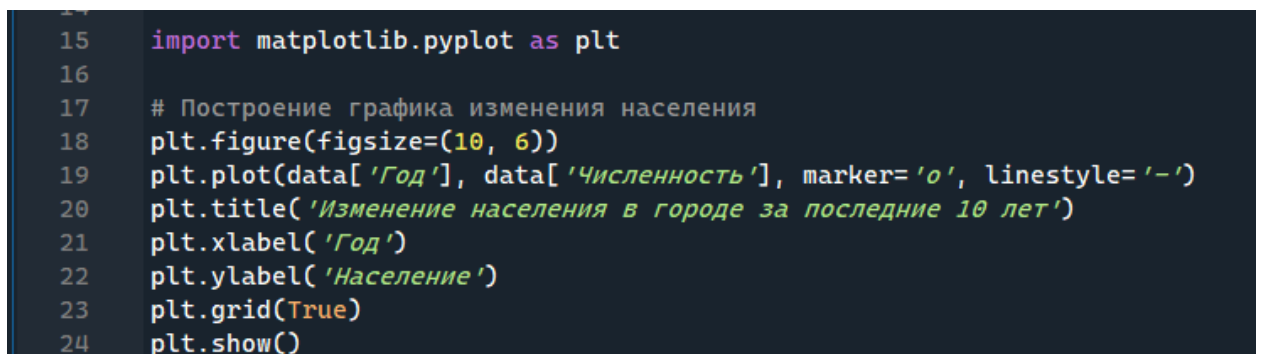


Рисунок 7 — Вторая часть кода

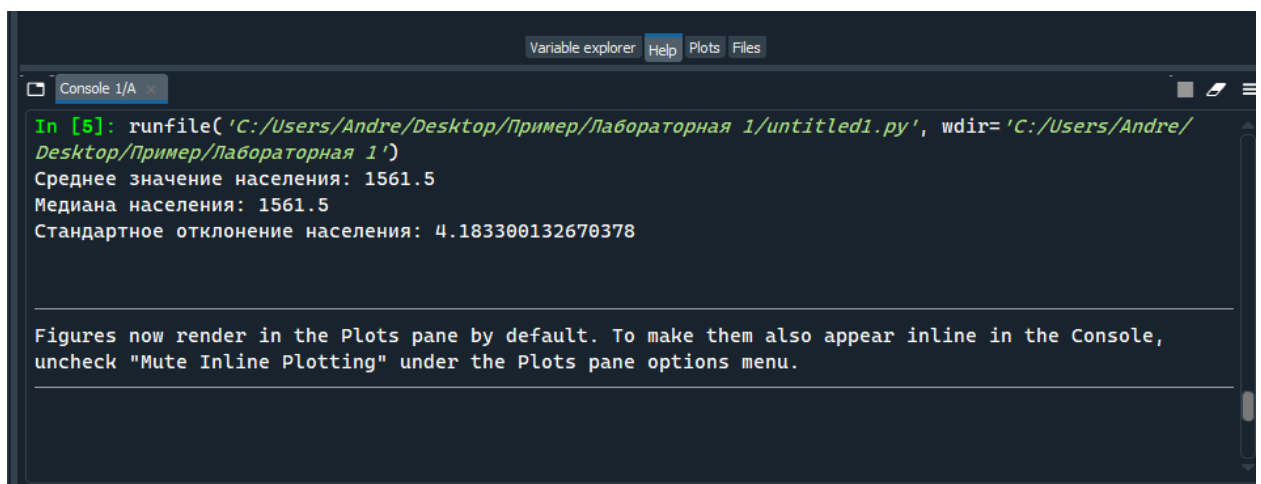
```

1  import pandas as pd
2
3  # Загрузка данных из файла Excel
4  data = pd.read_excel('data.xlsx')
5
6  # Проведение анализа данных
7  mean_population = data['Численность'].mean()
8  median_population = data['Численность'].median()
9  std_deviation_population = data['Численность'].std()
10
11  print(f'Среднее значение населения: {mean_population}')
12  print(f'Медиана населения: {median_population}')
13  print(f'Стандартное отклонение населения: {std_deviation_population}')
14
15  import matplotlib.pyplot as plt
16
17  # Построение графика изменения населения
18  plt.figure(figsize=(10, 6))
19  plt.plot(data['Год'], data['Численность'], marker='o', linestyle='--')
20  plt.title('Изменение населения в городе за последние 10 лет')
21  plt.xlabel('Год')
22  plt.ylabel('Население')
23  plt.grid(True)
24  plt.show()
25

```

Рисунок 8 — Весь код

Вновь запускаем и смотрим в консоль, результат должен быть таким же как показано на рисунке 9.



```

Variable explorer  Help  Plots  Files
Console 1/A
In [5]: runfile('C:/Users/Andre/Desktop/Пример/Лабораторная 1/untitled1.py', wdir='C:/Users/Andre/Desktop/Пример/Лабораторная 1')
Среднее значение населения: 1561.5
Медиана населения: 1561.5
Стандартное отклонение населения: 4.183300132670378

Figures now render in the Plots pane by default. To make them also appear inline in the Console,
uncheck "Mute Inline Plotting" under the Plots pane options menu.

```

Рисунок 9 — Результат

Вы можете увидеть надпись, которая появилась после дополнения нашего кода. Это значит, что график успешно был построен, остается лишь

переключиться с вкладки Help на вкладку Plots и посмотреть на наш график, как показано на рисунке 10.

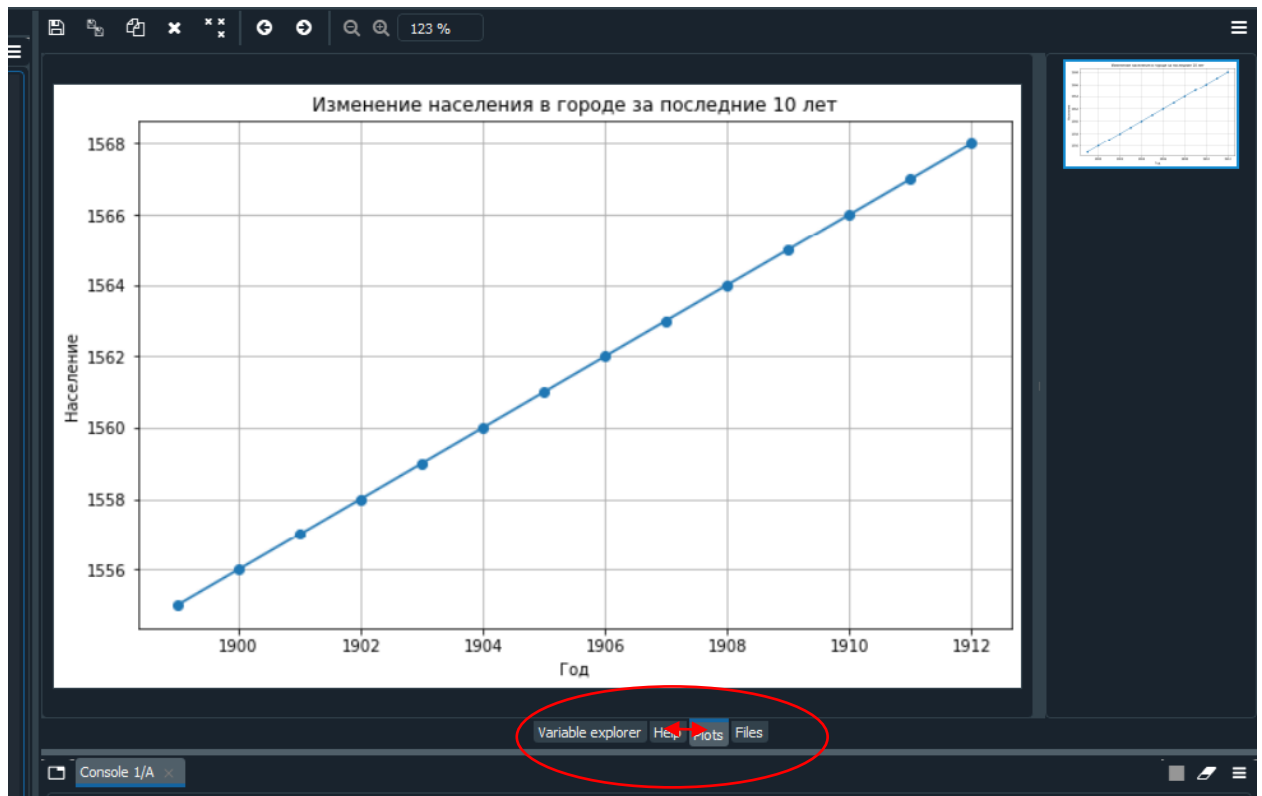


Рисунок 10 — График

После остается лишь проанализировать график и написать о результатах в выводе вашего отчета.

Сделайте отчет по данной работе на основе своих данных или данных, выданными преподавателем.