

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное
образовательное учреждение высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»**

Кафедра инфокоммуникаций

**Отчет по лабораторной работе №2
Введение в Pandas. Создание объекта Series в
Pandas
по дисциплине «Анализ данных»**

Выполнил студент группы ИВТ-б-о-21-1

Лысенко И.А. « » _____ 20__ г.

Подпись студента _____

Работа защищена « » _____ 20__ г.

Проверил Воронкин Р.А. _____

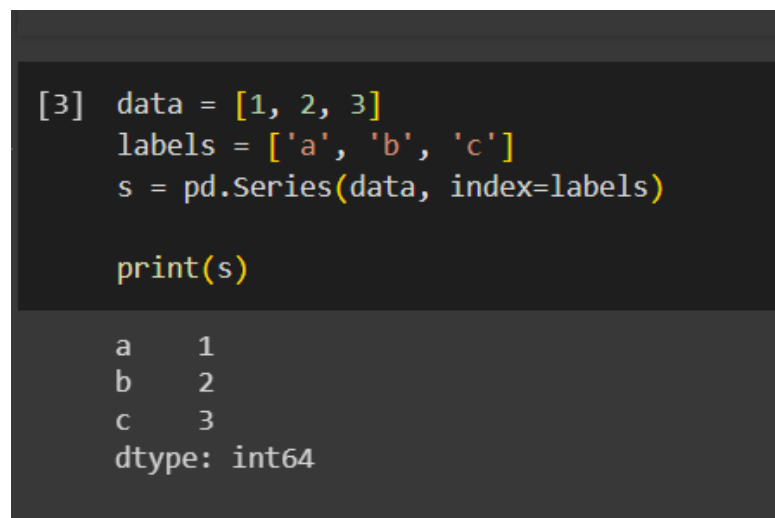
(подпись)

Ставрополь 2023

Цель работы: приобретение навыков по работе с классами и объектами при написании программ с помощью языка программирования Python версии 3. Исследование назначения и способов установки пакета Pandas, исследование способов создания основного объекта Series пакета Pandas..

Ход работы:

1. Создал репозиторий на GitHub: https://github.com/IsSveshuD/AD_lab_2.git . Ссылка на ноутбук: https://colab.research.google.com/drive/1dvWVaa-RrQDtc_Wx9M6G1qDRjp9x7ppm?usp=sharing .
2. Создал объект Series из списка чисел [1, 2, 3] с метками [a, b, c] и вывел его на печать:



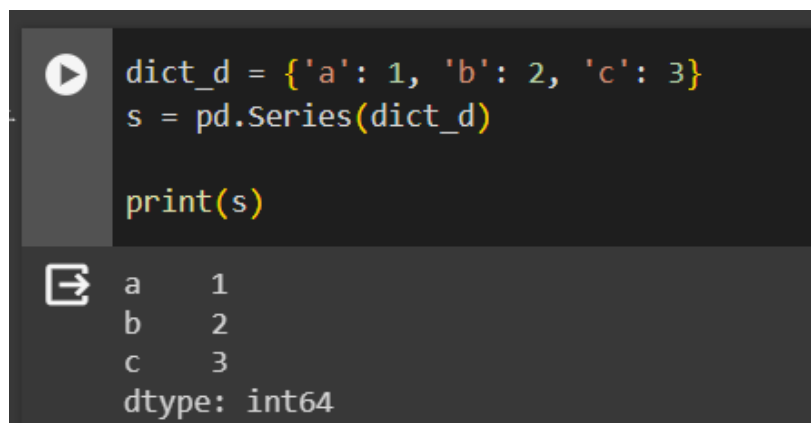
```
[3] data = [1, 2, 3]
     labels = ['a', 'b', 'c']
     s = pd.Series(data, index=labels)

     print(s)
```

a 1
b 2
c 3
dtype: int64

Рисунок 1 – Задание 1

3. Создал объект Series из словаря {'a': 1, 'b': 2, 'c': 3} и вывел его на печать:



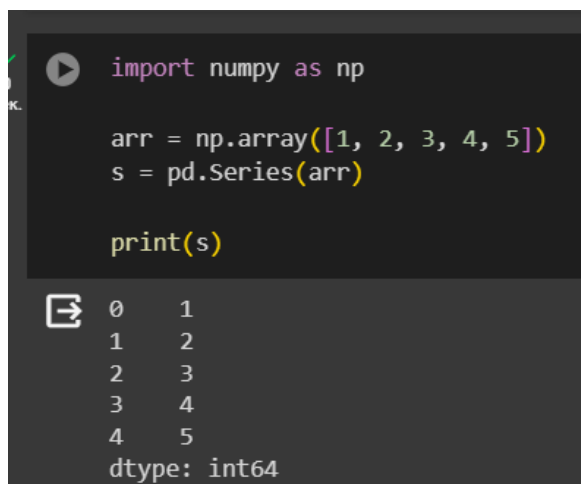
```
dict_d = {'a': 1, 'b': 2, 'c': 3}
s = pd.Series(dict_d)

print(s)
```

a 1
b 2
c 3
dtype: int64

Рисунок 2 – Задание 2

4. Создал объект массива NumPy `arr = [1, 2, 3, 4, 5]` , затем из него создал объект Series и вывел его на печать:



```
import numpy as np

arr = np.array([1, 2, 3, 4, 5])
s = pd.Series(arr)

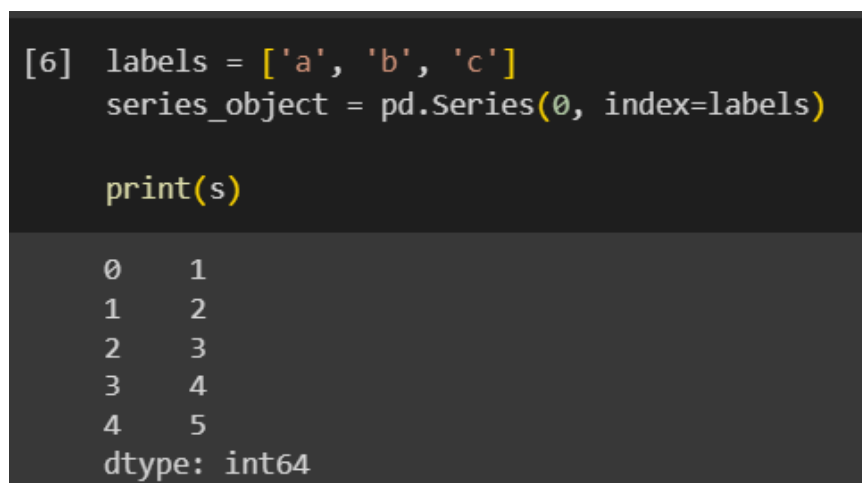
print(s)
```

0	1
1	2
2	3
3	4
4	5

dtype: int64

Рисунок 3 – Задание 3

5. Создал объект Series с заполнением по умолчанию значением 0 и метками [a, b, c] и вывел его на печать:



```
[6] labels = ['a', 'b', 'c']
series_object = pd.Series(0, index=labels)

print(s)
```

0	1
1	2
2	3
3	4
4	5

dtype: int64

Рисунок 4 – Задание 4

6. В первой строке ввода с клавиатуры поступает строка с символами разлепленными пробелом, во второй строка с целыми числами. На их основе создал список кортежей и из него объект Series, вывел его на печать:

```
in_chars = input("Введите символы через пробел: ")
chars_list = in_chars.split()

in_numb = input("Введите целые числа через пробел: ")
numbers_list = [int(num) for num in in_numb.split()]

#Создаём список кортежей
tuples = list(zip(chars_list, numbers_list))

#Создаём объект Series и выводим на печать
s = pd.Series(index=[t[0] for t in tuples], data=[t[1] for t in tuples])
print(s)
```

Введите символы через пробел: й ц у к
Введите целые числа через пробел: 1 2 3 4
й 1
ц 2
у 3
к 4
dtype: int64

Рисунок 5 – Задание 5

7. На вход поступает строка слов, разделенных пробелом. Создал на ее основе объект Series с метками 0, 1, 2,... Вывел его на печать:

```
[11] in_str = input("Введите слова через пробел: ")
words = in_str.split()
s = pd.Series(words)
print(s)
```

Введите слова через пробел: кот мышь собака енот
0 кот
1 мышь
2 собака
3 енот
dtype: object

Рисунок 6 – Задание 6

8. На вход поступает последовательность из чисел 0 и 1, разделенных пробелом. Создал на ее основе объект Series, содержащий булевы значения True (1) , False (0) с метками [a, b, c, d,...]:

```
in_str = input("Введите последовательность чисел 0 и 1 через пробел: ")
numbers_list = [int(num) for num in in_str.split()]
bool_s = pd.Series(numbers_list, dtype=bool)
bool_s.index = [chr(ord('a') + i) for i in range(len(bool_s))]
print(bool_s)
```

Введите последовательность чисел 0 и 1, разделенных пробелом: 0 1 1 0

```
a    False
b     True
c     True
d    False
dtype: bool
```

Рисунок 7 – Задание 7

9. Создал объект Series из списка дат ['2022-01-01', '2022-01-02', '2022-01-03'], вывел его на печать:

```
[13] from datetime import datetime

date_s = ['2022-01-01', '2022-01-02', '2022-01-03']

# Преобразуем в объекты datetime
date_obj = [datetime.strptime(date, '%Y-%m-%d') for date in date_s]

# Создаем объект Series
date_series = pd.Series(date_obj)
print(date_series)
```

```
0    2022-01-01
1    2022-01-02
2    2022-01-03
dtype: datetime64[ns]
```

Рисунок 8 – Задание 8

10. На вход подаются два целых числа n и m через пробел. На их основе создал объект Series с числами от n до m (включительно) с шагом 10 и метками [1, 3, 5, 7,...]. Вывел его на печать:

```
n, m = map(int, input("Введите два целых числа n и m через пробел: ").split())

# Создаем объект Series с числами от n до m (включительно) с шагом 10 и метками [1, 3, 5, 7, ...]
series_object = pd.Series(range(n, m + 1, 10), index=range(1, ((m - n) // 10) * 2 + 2, 2))

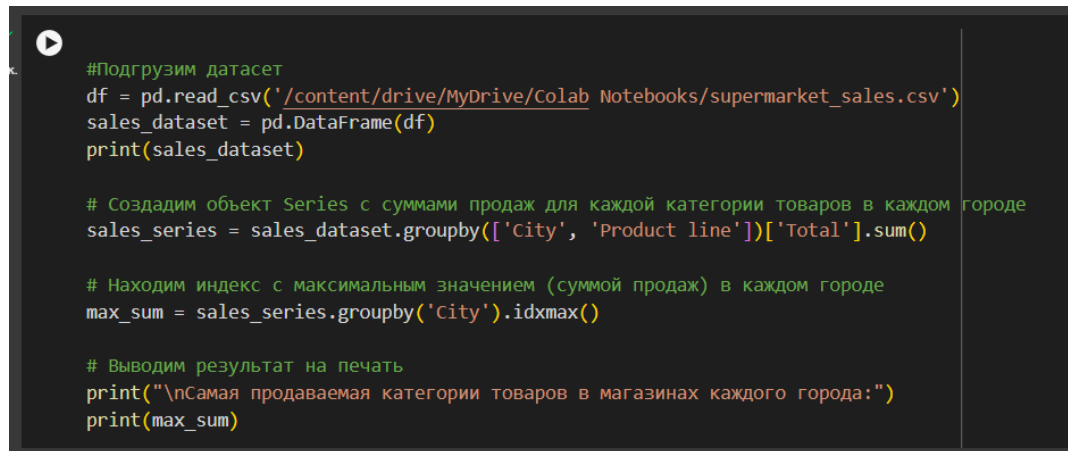
# Выводим объект Series на печать
print(series_object)
```

Введите два целых числа n и m через пробел: 5 30

```
1      5
3     15
5     25
dtype: int64
```

Рисунок 9 – Задание 9

11. Индивидуальное задание. Есть датасет с информацией о продажах продуктов различных категорий в магазинах нескольких городов. Задача - выяснить, какая категория продуктов является самой продаваемой в каждом городе:



```
#Подгрузим датасет
df = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/supermarket_sales.csv')
sales_dataset = pd.DataFrame(df)
print(sales_dataset)

# Создадим объект Series с суммами продаж для каждой категории товаров в каждом городе
sales_series = sales_dataset.groupby(['City', 'Product line'])['Total'].sum()

# Находим индекс с максимальным значением (суммой продаж) в каждом городе
max_sum = sales_series.groupby('City').idxmax()

# Выводим результат на печать
print("\nСамая продаваемая категории товаров в магазинах каждого города:")
print(max_sum)
```

Рисунок 10 – Индивидуальное задание

Ответы на вопросы:

1. Какое назначение библиотеки Pandas?

Pandas – это библиотека для обработки и анализа данных в языке программирования Python. Она предоставляет структуры данных, такие как DataFrame и Series, которые удобны для манипуляции и анализа табличных данных.

2. Каким образом можно загрузить данные в Pandas DataFrame?

Данные можно загрузить в Pandas DataFrame из различных источников, таких как CSV-файлы, Excel-файлы, базы данных SQL, HTML-таблицы и многие другие. Например, функция `pd.read_csv("filename.csv")` используется для загрузки данных из CSV-файла.

3. Какие типы данных могут содержать столбцы в Pandas DataFrame?

В столбцах Pandas DataFrame могут содержаться различные типы данных, включая числовые типы (`int`, `float`), строки (`object`), булевы значения (`bool`) и другие.

4. Какие недостатки могут быть у библиотеки Pandas?

Некоторые из недостатков Pandas включают в себя использование памяти, особенно при работе с большими наборами данных, и относительно медленную производительность при выполнении определенных операций.

5. Какая команда используется для установки библиотеки Pandas в Python?

Библиотеку Pandas можно установить с помощью команды: `pip install pandas`

6. Как проверить, установлена ли библиотека Pandas на компьютере?

Можно использовать команду `import pandas` в среде Python; если она не вызывает ошибку, библиотека установлена.

7. Какие операционные системы поддерживаются библиотекой Pandas?

Pandas поддерживает множество операционных систем, включая Windows, macOS и различные дистрибутивы Linux.

8. Как создать объект Series из списка значений?

```
import pandas as pd
my_list = [1, 2, 3, 4, 5]
my_series = pd.Series(my_list)
```

9. Как создать объект Series с явным заданием индексов?

```
my_series = pd.Series([1, 2, 3], index=['a', 'b', 'c'])
```

10. Как создать объект Series из словаря?

```
my_dict = {'a': 1, 'b': 2, 'c': 3}
my_series = pd.Series(my_dict)
```

11. Как создать объект Series с заполнением по умолчанию?

```
my_series = pd.Series(5, index=['a', 'b', 'c'])
```

12. Как создать объект Series из массива NumPy?

```
import numpy as np
my_array = np.array([1, 2, 3])
my_series = pd.Series(my_array)
```

13. Как создать объект Series из списка кортежей?

```
my_tuples = [('a', 1), ('b', 2), ('c', 3)]
```

```
my_series = pd.Series(dict(my_tuples))
```

Вывод: в результате выполнения лабораторной работы были исследованы назначения и способы установки пакета Pandas, исследованы способы создания основного объекта Series пакета Pandas.