

**Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»**

Институт цифрового развития
Кафедра инфокоммуникаций

**ОТЧЕТ
ПО ЛАБОРАТОРНОЙ РАБОТЕ №2
по дисциплине «Анализ данных»**

«Введение в Pandas. Создание объекта Series в Pandas»

Вариант 19

Выполнила:

Уланбекова Айканыш Уланбекова
студентка 3 курса, группы ИВТ-б-о-21-1
направление подготовки Информатика и
вычислительная техника, очная форма
обучения

(подпись)

Проверил:

Воронкин Роман Александрович

(подпись)

Отчет защищен с оценкой _____ Дата защиты _____

Ставрополь, 2023 г

Цель: исследование назначения и способов установки пакета Pandas, исследование способов создания основного объекта Series пакета Pandas.

Порядок выполнения:

1. Создала общедоступный репозиторий на github.
2. Выполнила клонирование репозитория.
3. Дополнила файл gitignore.
4. Организовала репозиторий в соответствии с моделью ветвления git-flow.

```
User@WIN-IV6UOF9FR99 MINGW64 ~/Documents/3,1 курс/Анализ данных/lab2 (master)
$ git checkout -b develop
Switched to a new branch 'develop'

User@WIN-IV6UOF9FR99 MINGW64 ~/Documents/3,1 курс/Анализ данных/lab2 (develop)
$ git flow init
No branches exist yet. Base branches must be created now.
Branch name for production releases: [master] main
Branch name for "next release" development: [develop] develop

How to name your supporting branch prefixes?
Feature branches? [feature/] feature
Bugfix branches? [bugfix/] bugfix
Release branches? [release/] release
Hotfix branches? [hotfix/] hotfix
Support branches? [support/] support
Version tag prefix? [] version
Hooks and filters directory? [C:/Users/User/Documents/3,1 курс/Анализ данных/lab2/.git/hooks] hooks

User@WIN-IV6UOF9FR99 MINGW64 ~/Documents/3,1 курс/Анализ данных/lab2 (develop)
$ |
```

Рисунок 1. Организация репозитория в соответствии с git flow

5. Решила следующие задачи:

Задание 1. Создать объект Series из списка чисел [1, 2, 3] с метками [a, b, c] и вывести его на печать.

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

a    1
b    2
c    3
dtype: int64
```

Рисунок 2. Выполнение задания 1

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

```
pd.Series({ "a": 1, "b": 2, "c": 3 })
```

a	1
b	2
c	3

dtype: int64

Рисунок 3. Выполнение задания 2

Задание 3. Создать объект массива NumPy `arr = [1, 2, 3, 4, 5]` , затем из него создать объект `Series` и вывести его на печать.

```
arr = [1, 2, 3, 4, 5]
pd.Series(arr)
```

0	1
1	2
2	3
3	4
4	5

dtype: int64

Рисунок 4. Выполнение задания 3

Задание 4. Создать объект `Series` с заполнением по умолчанию значением 0 и метками `[a, b, c]` и вывести его на печать.

```
pd.Series(0, index=["a", "b", "c"])
```

a	0
b	0
c	0

dtype: int64

Рисунок 5. Выполнение задания 4

Задание 5. В первой строке ввода с клавиатуры поступает строка с символами разлепленными пробелом, во второй строка с целыми числами. На их основе создать список кортежей и из него объект `Series`, вывести его на печать.

```
indexesInput = input("Введите строку с символами, разделенными пробелами")
valuesInput = input("Введите строку с целыми значениями, разделенными пробелами")

indexes = indexesInput.split(" ")
values = [int(value) for value in valuesInput.split(" ")]

pd.Series(values, index=indexes)
```

one	10
two	20
three	30
four	40
five	50

dtype: int64

Рисунок 6. Выполнение задания 5

Задание 6. На вход поступает строка слов, разделенных пробелом. Создать на ее основе объект Series с метками 0, 1, 2,... Вывести его на печать.

```
wordsInput = input("Введите строку слов, разделенных пробелом")

pd.Series(wordsInput.split(" "))
```

0	Ulanbekova
1	Aikanysh
2	Ulanbekova
3	IVT-21-1

dtype: object

Рисунок 7. Выполнение задания 6

Задание 7. На вход поступает последовательность из чисел 0 и 1, разделенных пробелом. Создать на ее основе объект Series, содержащий булевы значения True (1), False (0) с метками [a, b, c, d,...]

Задание 7. На вход поступает последовательность из чисел 0 и 1, разделенных пробелом. Создать на ее основе объект Series, содержащий булевы значения True (1), False (0) с метками [a, b, c, d,...]

```
letters = ["a", "b", "c", "d", "e", "f", "g", "h", "i", "j", "k", "l", "m", "n", "o", "p", "q", "r", "s", "t", "u", "v", "w", "x", "y", "z"]

valuesInput = input("Введите последовательность чисел 0 и 1, разделенных пробелом")
values = [bool(int(value)) for value in valuesInput.split(" ")]
indexes = letters[0:len(values)]

pd.Series(values, index=indexes)
```

a	False
b	False
c	False
d	True
e	True
f	False

dtype: bool

Рисунок 8. Выполнение задания 7

Задание 8. Создать объект Series из списка дат ['2022-01-01', '2022-01-02', '2022 01-03'], вывести его на печать. Даты в объекте Series должны быть

представлены объектами `datetime` стандартной библиотеки языка программирования Python.

```
pd.Series(data=pd.date_range("2022-01-01", periods=3))

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

Рисунок 9. Выполнение задания 8

Задание 9. На вход подаются два целых числа n и m через пробел. На их основе создать объект `Series` с числами от n до m (включительно) с шагом 10 и метками $[1, 3, 5, 7, \dots]$. Вывести его на печать.

```
numbersInput = input("Введите два целых числа n и m через пробел")
[n, m] = numbersInput.split(" ")

values = np.arange(int(n), int(m), 10)

maxIndex = len(values) * 2 + 1
indexes = np.arange(1, maxIndex, 2)

pd.Series(values, index=indexes)

1      10
3      20
5      30
7      40
9      50
11     60
13     70
15     80
17     90
dtype: int32
```

Рисунок 10. Выполнение задания 9

6. Самостоятельно придумала и решила задачу, где использовался бы объект `Series` библиотеки `Pandas`.

```
Создать объект Series с оценками студентов + Code

names = ["Иван", "Николай", "Оля", "Юля", "Петя", "Вова"]
grades = [4, 5, 4, 5, 3, 5]

data = pd.Series(grades, index=names)

print(data)
```

```
[81]
... Иван      4
   Николай    5
   Оля        4
   Юля        5
   Петя       3
   Вова       5
   dtype: int64
```

Рисунок 11. Выполнение индивидуального задания

Контрольные вопросы:

1. Какое назначение библиотеки Pandas?

Pandas - это библиотека на языке Python для работы с данными, которая предоставляет широкий набор инструментов для анализа и манипуляции табличных данных. Pandas использует основной объект - DataFrame - для хранения и работы с данными, а также предоставляет многочисленные функции для их обработки.

2. Каким образом можно загрузить данные в Pandas DataFrame?

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

Каждый столбец может иметь разные типы данных, например, числа, строки или даты.

4. Какие недостатки могут быть у библиотеки Pandas?

- Потребление памяти. Pandas может быть не очень эффективной в использовании памяти для больших наборов данных. Это может привести к тому, что операции на больших данных могут занимать много времени и памяти.

- Сложность. Pandas имеет множество функций и методов, что делает ее мощной, но возможно, сложной в использовании, особенно для новичков в анализе данных.

- Производительность. Несмотря на то, что Pandas является достаточно быстрой библиотекой, некоторые операции, особенно на больших наборах данных, могут быть довольно медленными.

- Ограниченная поддержка параллелизма. Pandas не имеет полной поддержки параллельных вычислений, что может замедлить операции на больших наборах данных.

- Поддержка типов данных. Pandas может быть не очень гибкой в работе с некоторыми типами данных, например, с данными, которые имеют смешанные типы или с данными, которые требуют высокой точности.

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

```
pip install pandas.
```

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

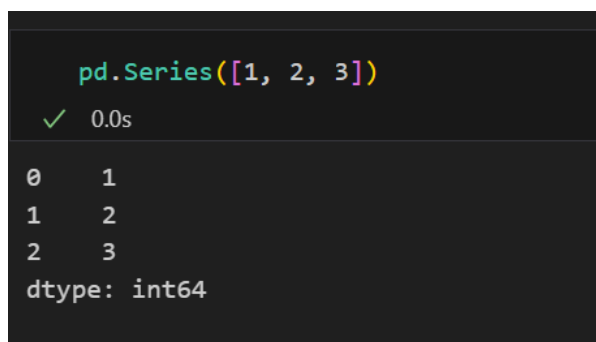
Чтобы убедиться, что Pandas установлен правильно, можно запустить интерпретатор Python и импортировать Pandas:

```
import pandas as pd
```

7. Какие операционные системы поддерживаются библиотекой Pandas?

Mac OS Linux, Windows

8. Как создать объект Series из списка значений?



```
pd.Series([1, 2, 3])
```

✓ 0.0s

0	1
1	2
2	3

dtype: int64

9. Как создать объект Series с явным заданием индексов?

```
pd.Series([1, 2, 3], index=["a", "b", "c"])
✓ 0.0s
a    1
b    2
c    3
dtype: int64
```

10. Как создать объект Series из словаря?

```
pd.Series({ "a": 1, "b": 2, "c": 3 })
a    1
b    2
c    3
dtype: int64
```

11. Как создать объект Series с заполнением по умолчанию?

```
pd.Series(0, index=["a", "b", "c"])
```

12. Как создать объект Series из массива NumPy?

```
pd.Series(np.array([1, 2, 3]))
✓ 0.0s
```

13. Как создать объект Series из списка кортежей?

```
tuples = [("a", 1), ("b", 2), ("c", 3)]
pd.Series(dict(tuples))
```

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