

```
'''
```

1.2.1

Создать массив 5x2. Создать массив 5x2. Вывести Все значения массива, значение элемента с индексом (3,1) и Второй столбец. Индексация начинается с нуля.

```
'''
```

```
import numpy as np
x = np.array([[1,2],[3,4],[5,6],[7,8],[9,10]])
print(x)
print(x[3][1])
print(x[1])
```

Program output:

```
[[ 1  2]
 [ 3  4]
 [ 5  6]
 [ 7  8]
 [ 9 10]]
```

```
8
```

```
[3 4]
```

```
'''
```

1.2.2

Пример. Выполнить следующее:

1. Создать Вектор (одномерный массив) размера 10, заполненный нулями.
2. Создать Вектор размера 10, заполненный единицами.
3. Создать Вектор размера 10, заполненный заданным числом.
4. Создать Вектор со значениями от 10 до 19.

```
'''
```

```
import numpy as np
a = np.zeros(10)
b = np.ones(10)
c = np.full(10, 5)
d = np.arange(10,20)
print(f"{a}\n{b}\n{c}\n{d}")
```

Program output:

```
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
[1. 1. 1. 1. 1. 1. 1. 1. 1. 1.]
[5 5 5 5 5 5 5 5 5 5]
[10 11 12 13 14 15 16 17 18 19]
```

```
'''  
1.2.3  
Создать массив 10x10 со случайными значениями, найти минимум, максимум и среднее значение  
'''
```

```
import numpy as np  
Z = np.random.random((10,10))  
Zmin, Zmax, Zmean = Z.min(), Z.max(), Z.mean()  
print(Zmin, Zmax, Zmean)
```

Program output:

0.0031112348186829175 0.9978660232770116 0.5076605818245621

```
'''
```

1.2.4

Загать матрицу размерности 5 на 5 и поменять 2 строки в матрице местами.

```
'''
```

```
import numpy as np
A = np.arange(25).reshape(5,5)
A[[0,1]] = A[[1,0]]
print(A)
```

Program output:

```
[[ 5  6  7  8  9]
 [ 0  1  2  3  4]
 [10 11 12 13 14]
 [15 16 17 18 19]
 [20 21 22 23 24]]
```

```
'''
```

1.2.5

Выяснить результат следующих Выражений:

```
0 * np.nan
```

```
np.nan == np.nan
```

```
np.inf > np.nan
```

```
np.nan - np.nan
```

```
0.3 == 3 * 0.1
```

```
'''
```

```
import numpy as np
```

```
print(0 * np.nan)
```

```
print(np.nan == np.nan)
```

```
print(np.inf > np.nan)
```

```
print(np.nan - np.nan)
```

```
print(0.3 == 3 * 0.1)
```

Program output:

```
nan
```

```
False
```

```
False
```

```
nan
```

```
False
```

```
'''  
1.2.6  
Отсортировать массив  
'''  
  
import numpy as np  
arr = np.array([2,1,5,3,7,4,6,8])  
print(np.sort(arr))
```

Program output:

```
[1 2 3 4 5 6 7 8]
```



```
'''
```

1.3.2

Создать 5x5 матрицу со значениями в строках от 0 до 4. Для создания необходимо использовать функцию arrange

```
'''
```

```
import numpy as np
arr = []
for i in range(5):
    arr.extend(np.arange(5))
a = np.array(arr).reshape(5,5)
print(a)
```

Program output:

```
[[0 1 2 3 4]
 [0 1 2 3 4]
 [0 1 2 3 4]
 [0 1 2 3 4]
 [0 1 2 3 4]]
```



```
'''
```

1.3.3

Создать массив 3x3x3 со случайными значениями

```
'''
```

```
import numpy as np
a = np.random.random((3, 3, 3))
print(a)
```

Program output:

```
[[[0.0935241  0.61387584 0.31079953]
  [0.70789001 0.3283358  0.02305403]
  [0.98073818 0.43524386 0.88774421]]
```

```
[[[0.87349707 0.0770195  0.55481833]
  [0.71875492 0.33734188 0.2233107 ]
  [0.27069668 0.74050233 0.01807307]]
```

```
[[[0.56343439 0.1937467  0.97907577]
  [0.39676611 0.27437924 0.40200579]
  [0.75989152 0.91960497 0.59174162]]]
```



```
'''
```

1.3.5

Создайте массив и отсортируйте его по убыванию.

```
'''
```

```
import numpy as np
a = np.random.rand(10)
print(np.sort(a)[::-1])
```

Program output:

```
[0.89119962 0.82330437 0.6995577  0.69162395 0.67162794 0.57780624
 0.17627805 0.17386131 0.13332624 0.0761592 ]
```

```
'''
```

1.3.6

Создайте матрицу, Выведите ее форму, размер и размерность

```
'''
```

```
import numpy as np
a = np.random.random((5,5))
print(a)
print(f"Форма: {a.shape}")
print(f"Размер: {a.size}")
print(f"Размерность: {a.ndim}")
```

Program output:

```
[[0.87148507 0.42024261 0.46545946 0.7766243  0.48433206]
 [0.63333089 0.73446627 0.89910479 0.51692507 0.92888673]
 [0.12618987 0.2082944  0.20772871 0.95834626 0.01490062]
 [0.29552041 0.75737081 0.87901716 0.47049728 0.29401004]
 [0.61675206 0.20514933 0.50603416 0.98379765 0.90377257]]
```

Форма: (5, 5)

Размер: 25

Размерность: 2

```
'''
```

2.2.1

Создать Series из списка Python, словаря Python, и массива Numpy (установить буквенные метки для последнего).

```
'''
```

```
import pandas as pd
import numpy as np
lst = [1,2,3,4,5]
d = {'a':1, 'b':2, 'c':3}
ndarr = np.array([1,2,3,4,5])
s1 = pd.Series(lst)
s2 = pd.Series(d)
s3 = pd.Series(ndarr, ['a', 'b', 'c', 'd', 'e'])
print(s1)
print(s2)
print(s3)
```

Program output:

```
0    1
1    2
2    3
3    4
4    5
dtype: int64
a    1
b    2
c    3
dtype: int64
a    1
b    2
c    3
d    4
e    5
dtype: int64
```

```
'''
```

2.2.2

Дано два Series. Напечатать их первые элементы и Все элементы после третьего (Во втором фрейме).

```
'''
```

```
import numpy as np
import pandas as pd
s1 = pd.Series([1, 2, 3, 4, 5], ['a', 'b', 'c', 'd', 'e'])
s2 = pd.Series([5, 4, 3, 2, 1])
print(s1['a'])
print(s2[0])
print(s2[3:])
```

Program output:

```
1
5
3    2
4    1
dtype: int64
```

```
'''
```

2.2.3

Создайте новый фрейм данных.

```
'''
```

```
import pandas as pd
import numpy as np
dataframe = pd.DataFrame ( )
dataframe['Имя'] = ['Джеки Джексон', 'Стувен стувенсон']
dataframe['Возраст'] = [38, 25]
dataframe['Водитель'] = [True, False]
print(dataframe)
```

Program output:

	Имя	Возраст	Водитель
0	Джеки Джексон	38	True
1	Стувен стувенсон	25	False

```
'''
```

2.2.4

Загрузите фрейм данных по ссылке:

https://web.archive.org/web/20201120015841id_/https://raw.githubusercontent.com/chrisalbon/simulated_datasets/master/titanic.csv

```
'''
```

```
import numpy as np
```

```
import pandas as pd
```

```
url =
```

```
"https://web.archive.org/web/20201120015841id_/https://raw.githubusercontent.com/chrisalbon/simulated_datasets/master/titanic.csv"
```

```
dataframe = pd.read_csv(url)
```

```
print(dataframe.head(5))
```

Program output:

	Name	PClass	...	Survived	SexCode
0	Allen, Miss Elisabeth Walton	1st	...	1	1
1	Allison, Miss Helen Loraine	1st	...	0	1
2	Allison, Mr Hudson Joshua Creighton	1st	...	0	0
3	Allison, Mrs Hudson JC (Bessie Waldo Daniels)	1st	...	0	1
4	Allison, Master Hudson Trevor	1st	...	1	0

[5 rows x 6 columns]


```
...
```

2.2.5

Проанализировать характеристики фрейма данных

```
...
```

```
import numpy as np
import pandas as pd
url =
"https://web.archive.org/web/20201120015841id_/https://raw.githubusercontent.com/chrisalbon/simulated_datasets/master/titanic.csv"

dataframe = pd.read_csv(url)
print(dataframe.head(2))
print(dataframe.tail(3))
print(dataframe.shape)
print(dataframe.describe())
```

Program output:

	Name	PClass	Age	Sex	Survived	SexCode
0	Allen, Miss Elisabeth Walton	1st	29.0	female	1	1
1	Allison, Miss Helen Loraine	1st	2.0	female	0	1

	Name	PClass	Age	Sex	Survived	SexCode
1310	Zenni, Mr Philip	3rd	22.0	male	0	0
1311	Lievens, Mr Rene	3rd	24.0	male	0	0
1312	Zimmerman, Leo	3rd	29.0	male	0	0

(1313, 6)

	Age	Survived	SexCode
count	756.000000	1313.000000	1313.000000
mean	30.397989	0.342727	0.351866
std	14.259049	0.474802	0.477734
min	0.170000	0.000000	0.000000
25%	21.000000	0.000000	0.000000
50%	28.000000	0.000000	0.000000
75%	39.000000	1.000000	1.000000
max	71.000000	1.000000	1.000000

```
'''
```

2.2.6

Выберите индивидуальные данные или срезы фрейма данных

```
'''
```

```
import numpy as np
```

```
import pandas as pd
```

```
url =
```

```
"https://web.archive.org/web/20201120015841id_/https://raw.githubusercontent.com/chrisalbon/simulated_datasets/master/titanic.csv"
```

```
"
```

```
dataframe = pd.read_csv(url)
```

```
print(dataframe.iloc[1:4])
```

Program output:

	Name	PClass	...	Survived	SexCode
1	Allison, Miss Helen Loraine	1st	...	0	1
2	Allison, Mr Hudson Joshua Creighton	1st	...	0	0
3	Allison, Mrs Hudson JC (Bessie Waldo Daniels)	1st	...	0	1

```
[3 rows x 6 columns]
```

```
'''
```

2.2.7

Требуется отобрать строки фрейма данных на основе некоторого условия. Необходимо сформировать новый фрейм данных из пассажиров первого класса.

```
'''
```

```
import numpy as np
import pandas as pd
url =
"https://web.archive.org/web/20201120015841id_/https://raw.githubusercontent.com/chrisalbon/simulated_datasets/master/titanic.csv"
dataframe = pd.read_csv(url)
print(dataframe[dataframe['PClass'] == '1st'].head(2))
```

Program output:

	Name	PClass	Age	Sex	Survived	SexCode
0	Allen, Miss Elisabeth Walton	1st	29.0	female	1	1
1	Allison, Miss Helen Loraine	1st	2.0	female	0	1

```
'''
```

2.3.1

Найти евклидово расстояние между двумя Series (точками) a и b, не используя Встроенную формулу.

```
'''
```

```
import pandas as pd
import numpy as np
s1 = pd.Series([1, 2, 3, 4, 5])
s2 = pd.Series([5, 4, 3, 2, 1])

print(sum((s1 - s2)**2)**.5)
```

Program output:

6.324555320336759

```
'''
```

2.3.2

Найдите В Интернете ссылку на любой csv файл и сформируйте из него фрейм данных (например, коллекцию фреймов данных можно найти здесь: <https://github.com/akmand/datasets>).

```
'''
```

```
import numpy as np
import pandas as pd
url = "https://raw.githubusercontent.com/akmand/datasets/refs/heads/main/airlines.csv"
dataframe = pd.read_csv(url)
print(dataframe.head(5))
```

Program output:

	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay
0	CO	269	SFO	IAH	3	15	205	1
1	US	1558	PHX	CLT	3	15	222	1
2	AA	2400	LAX	DFW	3	20	165	1
3	AA	2466	SFO	DFW	3	20	195	1
4	AS	108	ANC	SEA	3	30	202	0

```
'''
```

2.3.3

Проделайте с получившемся из предыдущего задания фреймом данных те же действия, что и в примерах 2.2.5–2.2.7.

```
'''
```

```
import numpy as np
import pandas as pd
url = "https://raw.githubusercontent.com/akmand/datasets/refs/heads/main/airlines.csv"
dataframe = pd.read_csv(url)
print(dataframe.head(2))
print(dataframe.tail(3))
print(dataframe.shape)
print(dataframe.describe())
print(dataframe.iloc[1:4])
print(dataframe[dataframe['Airline'] == 'AA'].head(2))
```

Program output:

```
  Airline  Flight  AirportFrom  AirportTo  DayOfWeek  Time  Length  Delay
0      CO     269           SFO         IAH          3    15     205      1
1      US    1558           PHX         CLT          3    15     222      1
      Airline  Flight  AirportFrom  AirportTo  DayOfWeek  Time  Length  Delay
539380      FL     609           SFO         MKE          5  1439     255      0
539381      UA      78           HNL         SFO          5  1439     313      1
539382      US   1442           LAX         PHL          5  1439     301      1
(539383, 8)
```

	Flight	DayOfWeek	...	Length	Delay
count	539383.000000	539383.000000	...	539383.000000	539383.000000
mean	2427.928630	3.929668	...	132.202007	0.445442
std	2067.429837	1.914664	...	70.117016	0.497015
min	1.000000	1.000000	...	0.000000	0.000000
25%	712.000000	2.000000	...	81.000000	0.000000
50%	1809.000000	4.000000	...	115.000000	0.000000
75%	3745.000000	5.000000	...	162.000000	1.000000
max	7814.000000	7.000000	...	655.000000	1.000000

```
[8 rows x 5 columns]
```

```
  Airline  Flight  AirportFrom  AirportTo  DayOfWeek  Time  Length  Delay
```

1	US	1558	PHX	CLT	3	15	222	1
2	AA	2400	LAX	DFW	3	20	165	1
3	AA	2466	SFO	DFW	3	20	195	1
	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay
2	AA	2400	LAX	DFW	3	20	165	1
3	AA	2466	SFO	DFW	3	20	195	1

```
'''  
3.2.1  
Прошкалируйте числовой признак В диапазон между двумя значениями.  
'''
```

```
import numpy as np  
from sklearn import preprocessing  
feature = np.array([[ -500.5], [-100.1], [0], [100.1], [900.9]])  
minmax_scale = preprocessing.MinMaxScaler(feature_range = (0, 1))  
scaled_feature = minmax_scale.fit_transform(feature)  
print(scaled_feature)
```

Program output:

```
[[0.      ]  
 [0.28571429]  
 [0.35714286]  
 [0.42857143]  
 [1.      ]]
```



```
'''
```

3.2.2

Преобразуйте признак, чтобы он имел среднее значение 0 и стандартное отклонение 1.

```
'''
```

```
import numpy as np
from sklearn import preprocessing
x = np.array([[1000.1], [-200.21], [500.5], [600.6], [9000.9]])
scaler = preprocessing.StandardScaler()
standardized = scaler.fit_transform(x)
print(standardized)
print(f"Среднее: {round(standardized.mean())}")
print(f"Стандартное отклонение: {standardized.std()}")
```

Program output:

```
[[-0.76058191]
 [-0.54177399]
 [-0.35009651]
 [-0.3227144 ]
 [ 1.97516681]]
```

Среднее: 0

Стандартное отклонение: 1.0

```
'''
3.2.3
Дан фрейм данных
dfTest = pd.DataFrame({'A':[14.00,90.20,90.95,96.27,91.21],
                        'B':[103.02,107.26,110.35,114.23,114.68],
                        'C':['big','small','big','small','small']})
```

Необходимо масштабировать его числовые столбцы

```
'''
from sklearn import preprocessing
import pandas as pd
scaler = preprocessing.MinMaxScaler()
dfTest = pd.DataFrame({'A':[14.00,90.20,90.95,96.27,91.21],
                        'B':[103.02,107.26,110.35,114.23,114.68],
                        'C':['big','small','big','small','small']})
dfTest[['A', 'B']] = scaler.fit_transform(dfTest[['A', 'B']])
print(dfTest)
```

Program output:

	A	B	C
0	0.000000	0.000000	big
1	0.926219	0.363636	small
2	0.935335	0.628645	big
3	1.000000	0.961407	small
4	0.938495	1.000000	small

```
'''
```

3.3.2

Загрузить фрейм данных по ссылке: <https://raw.githubusercontent.com/akmand/datasets/master/iris.csv>. Необходимо Выполнить нормализацию первого числового признака (sepal_length_cm) с использованием минимаксного преобразования, а Второго (sepal_width_cm) с задействованием z-масштабирования.

```
'''
```

```
import pandas as pd
from sklearn import preprocessing
url = "https://raw.githubusercontent.com/akmand/datasets/master/iris.csv"
dataframe = pd.read_csv(url)

minmax = preprocessing.MinMaxScaler()
dataframe[['sepal_length_cm']] = minmax.fit_transform(dataframe[['sepal_length_cm']])
std = preprocessing.StandardScaler()
dataframe[['sepal_width_cm']] = std.fit_transform(dataframe[['sepal_width_cm']])
print(dataframe.head(5))
```

Program output:

	sepal_length_cm	sepal_width_cm	petal_length_cm	petal_width_cm	species
0	0.222222	1.032057	1.4	0.2	setosa
1	0.166667	-0.124958	1.4	0.2	setosa
2	0.111111	0.337848	1.3	0.2	setosa
3	0.083333	0.106445	1.5	0.2	setosa
4	0.194444	1.263460	1.4	0.2	setosa