

<div style="display: flex; flex-direction: column; align-items: center;"> <div style="display: flex; justify-content: space-between; width: 100%;"> AD</div> <div style="display: flex; justify-content: space-between; width: 100%;"> AI</div> </div>	PHÒNG THÍ NGHIỆM THỰC HÀNH ỨNG DỤNG KHOA HỌC DỮ LIỆU VÀ TRÍ TUỆ NHÂN TẠO Applied Data Science and Artificial Intelligence Lab
<div style="display: flex; align-items: center;">  <div style="flex: 1; padding-left: 20px;"> <h2 style="color: yellow;">LẬP TRÌNH PYTHON CHO KHOA HỌC DỮ LIỆU</h2> <h3 style="color: white;">PYTHON FOR DATA SCIENCE</h3> <p style="text-align: right; color: white;">Hà Nội ★ 11.2020</p> </div> </div>	
Nothing is faster than Innovation together	

1

<div style="display: flex; flex-direction: column; align-items: center;"> <div style="display: flex; justify-content: space-between; width: 100%;"> AD</div> <div style="display: flex; justify-content: space-between; width: 100%;"> AI</div> </div>	<h2 style="color: white;">Thông tin giảng viên</h2>	
	TS. TẠ QUANG CHIẾU	<p>Vị trí & Đào tạo</p> <ul style="list-style-type: none"> - Trưởng bộ môn Hệ thống thông tin và Tri thức, Khoa Công nghệ Thông tin - PhD & Post-doc in Polytech Tours, France <p>Lĩnh vực giảng dạy, nghiên cứu</p> <ul style="list-style-type: none"> - Big data, Data science, machine learning, AI - Các phương pháp cho bài toán lập kế hoạch <p>Contact:</p> <p>(E): quangchieu.ta@gmail.com</p> <p>(M): 0913 522 275</p>
Dr. Tạ Quang Chiếu – (E): quangchieu.ta@gmail.com – (M): 0913 522 275		2/91

2

<div style="display: flex; justify-content: space-between;"> A D </div> <div style="display: flex; justify-content: space-between;"> A I </div>	<h2 style="margin: 0;">Objectives</h2>
<ul style="list-style-type: none"> Hiểu được tầm quan trọng của Khoa học dữ liệu Vận dụng được các bước trong quy trình thực hiện một dự án về khoa học dữ liệu. Kiến thức, kỹ năng cần thiết để trở thành một nhà khoa học dữ liệu Áp dụng được Python và các thư viện phổ biến trong giải quyết một số bài toán cơ bản của Khoa học dữ liệu. 	
<div style="display: flex; justify-content: space-between;"> Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275 3/91 </div>	

3

<div style="display: flex; justify-content: space-between;"> A D </div> <div style="display: flex; justify-content: space-between;"> A I </div>	<h2 style="margin: 0;">References</h2>
<ul style="list-style-type: none"> [1]. Introducing Data Science [2]. Python Data Science Handbook [3]. Data Science from Scratch [4]. Python crash course [5]. Bài giảng Khoa học dữ liệu 	
<div style="display: flex; justify-content: space-between;"> Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275 4/91 </div>	

4

A

A

D

I

PHẦN 2: NHẬP MÔN KHOA HỌC DỮ LIỆU

FUNDAMENTALS OF DATA SCIENCE

CONTENT

1

CHAPTER 1: INTRODUCTION

2

CHAPTER 2: PYTHON LIBRARIES FOR DATA SCIENCE

Dr. Tà Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

5/91

5

A

A

D

I

Content of chapter 2

2.1 Python Library for Data Science

2.2 Numpy library

3.3 Pandas library

2.4 Matplotlib library

2.5 Scikit-learn library

Dr. Tà Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

6/91

6

A D
A I

Python Library for Data Science

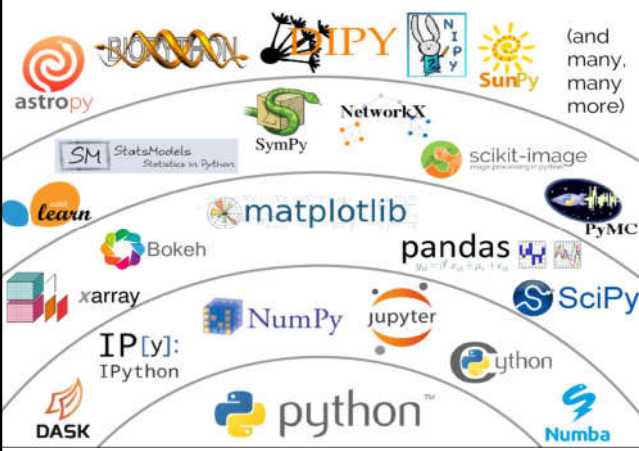
Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275
7/91

7

A D
A I

2.1. Python Library for Data Science

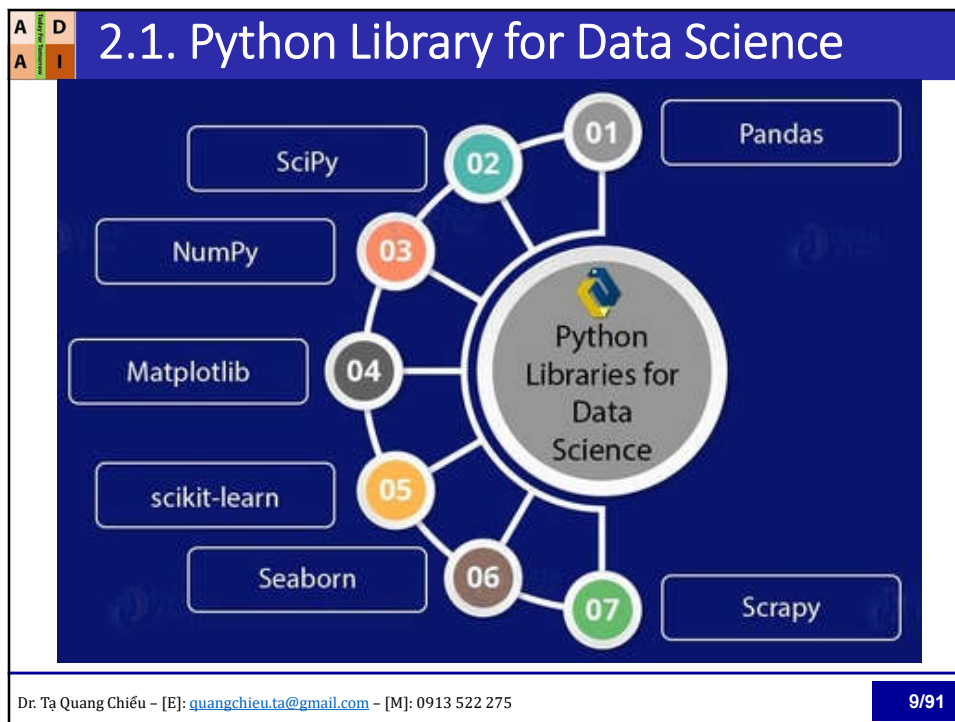
- **Python** có hệ thống thư viện rất phong phú, hỗ trợ nhiều lĩnh vực khác nhau.



- Do đó, tùy thuộc vào lĩnh vực nghiên cứu cụ thể, để lựa chọn và sử dụng các thư viện cho phù hợp.

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275
8/91

8



9

2.1. Python Library for Data Science

- 01 Pandas:** Sử dụng cho quản lý và tương tác với dữ liệu có cấu trúc, được sử dụng rộng rãi trong việc thu thập và tiền xử lý dữ liệu.
- 02 SciPy:** Dựa trên Numpy, cung cấp các công cụ mạnh cho khoa học và kỹ nghệ, như biến đổi fourier rời rạc, đại số tuyến tính, tối ưu hóa và ma trận thưa
- 03 Numpy:** Thư viện chuyên xử lý dữ liệu số (nhiều chiều), bao gồm cả các hàm đại số tuyến tính cơ bản, biến đổi fourier, sinh số ngẫu nhiên nâng cao,...
- 04 Matplotlib:** Thư viện này được sử dụng để trực quan hóa dữ liệu (Data Visualization), chuyên vẽ các biểu đồ, hỗ trợ rất nhiều loại biểu đồ khác nhau...

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

10/91

10

A

D

2.1. Python Library for Data Science

05

Scikit-learn: Thư viện [chuyên về học máy](#); thư viện này có sẵn nhiều công cụ hiệu quả cho học máy và thiết lập các mô hình thống kê như các thuật toán phân lớp, hồi quy, phân cụm và giảm chiều dữ liệu...

06

Seaborn: Thư viện này dựa trên Matplotlib, cung cấp các công cụ hiển thị dữ liệu một cách trực quan, hiệu quả. Mục tiêu của thư viện này là sử dụng việc [trực quan hóa dữ liệu](#) như là trọng tâm của khám phá và hiểu dữ liệu

07

Scrapy: Thư viện này chuyên về việc [thu thập thông tin trên Web](#), rất phù hợp với việc lấy các dữ liệu theo mẫu.

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

11/91

11

A

D

How to Install Python Packages

Một số thư viện Python được cài đặt mặc định, để kiểm tra thư viện đã được cài đặt hay chưa và phiên bản đang sử dụng là bao nhiêu:

A. Sử dụng Jupyter notebook

```

In [1]: #Khởi báo sử dụng thư viện và kiểm tra phiên bản thư viện đang sử dụng
import numpy as np
print("Thu vien Numpy, Version: ",np.__version__)

Thu vien Numpy, Version: 1.15.4

In [2]: #Trong trường hợp thư viện chưa được cài đặt!
import scrapy as sc
print("Thu vien Scrapy, Version: ",sc.__version__)

-----
ModuleNotFoundError                               Traceback (most recent call last)
<ipython-input-2-ef1be0ed66f4> in <module>
      1 #Trong trường hợp thư viện chưa được cài đặt!
----> 2 import scrapy as sc
      3 print("Thu vien Scrapy, Version: ",sc.__version__)

ModuleNotFoundError: No module named 'scrapy'

```

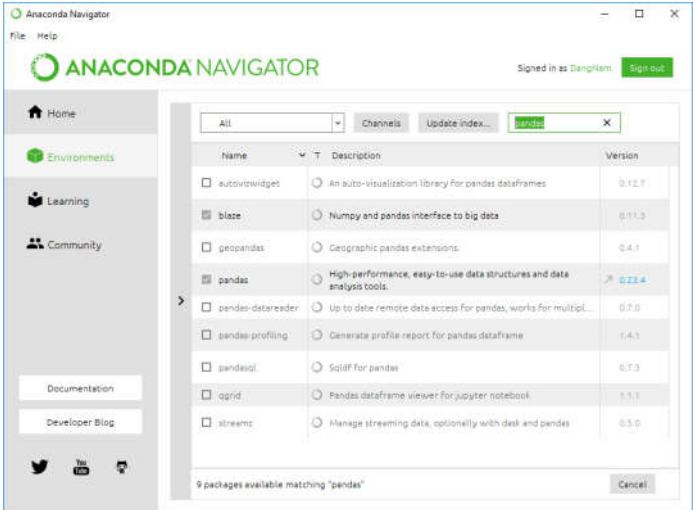
Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

12/91

12

A D A I How to Install Python Packages

B.Sử dụng Anaconda Navigator



The screenshot shows the Anaconda Navigator application window. On the left is a sidebar with navigation options: Home, Environments, Learning, and Community. The main panel displays the 'Channels' tab, which lists various Python packages available for installation. The 'pandas' package is selected, and its details are shown in the right pane. The status bar at the bottom indicates '9 packages available matching "pandas"'. At the bottom of the slide, the contact information for Dr. Tạ Quang Chiếu is provided, along with the slide number 13/91.

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

13/91

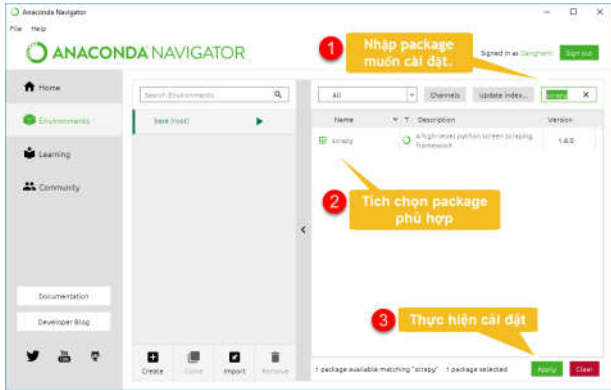
13

A D A I How to Install Python Packages

Trường hợp thư viện chưa được cài đặt, có thể sử dụng lệnh:

!pip install <tên thư viện>

Hoặc sử dụng Anacoda Navigator:



The screenshot shows the Anaconda Navigator application window with three numbered annotations in yellow boxes: 1. 'Nhập package muốn cài đặt' (Enter package to install) pointing to the search bar, 2. 'Tích chọn package phù hợp' (Select appropriate package) pointing to the 'pandas' package in the list, and 3. 'Thực hiện cài đặt' (Perform installation) pointing to the 'Install' button. The status bar at the bottom indicates '1 package available matching "pandas"'. At the bottom of the slide, the contact information for Dr. Tạ Quang Chiếu is provided, along with the slide number 14/91.

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

14/91

14

<div><div>A</div><div>D</div></div> <div><div>A</div><div>I</div></div>	<h1>NumPy Library - phần 01</h1>
Dr. Tà Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275	
15/91	

15

<div><div>A</div><div>D</div></div> <div><div>A</div><div>I</div></div>	<h2>Nội dung phần 01</h2>
<ul style="list-style-type: none">1. Introduction2. NumPy creating array3. Array Indexing and Slicing4. Mathematical statistics functions5. Bài tập	
Dr. Tà Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275	
16/91	

16

<div style="display: flex; justify-content: space-between;"> <div style="background-color: #000080; color: white; padding: 2px;">A</div> <div style="background-color: #000080; color: white; padding: 2px;">D</div> </div> <div style="display: flex; justify-content: space-between;"> <div style="background-color: #000080; color: white; padding: 2px;">A</div> <div style="background-color: #000080; color: white; padding: 2px;">I</div> </div>	<h2 style="margin: 0;">Ref document</h2>
<p>https://www.machinelearningplus.com/numpy-tutorial-part1-array-python-examples</p> <p>https://www.machinelearningplus.com/numpy-tutorial-python-part2</p>	
<div style="display: flex; justify-content: space-between;"> Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275 17/91 </div>	

17

<div style="display: flex; justify-content: space-between;"> <div style="background-color: #000080; color: white; padding: 2px;">A</div> <div style="background-color: #000080; color: white; padding: 2px;">D</div> </div> <div style="display: flex; justify-content: space-between;"> <div style="background-color: #000080; color: white; padding: 2px;">A</div> <div style="background-color: #000080; color: white; padding: 2px;">I</div> </div>	<h2 style="margin: 0;">1. Introduction</h2>
<ul style="list-style-type: none"> Numpy (Numeric Python): là một thư viện toán học phổ biến và mạnh mẽ của Python. Cho phép làm việc hiệu quả với ma trận và mảng, đặc biệt là dữ liệu ma trận và mảng lớn với tốc độ xử lý nhanh hơn nhiều lần khi chỉ sử dụng “core Python” đơn thuần. Ngoài ra, Python cũng hỗ trợ một thư viện khác để mở rộng thêm các tính năng của Numpy là Scipy với ưu thế về các phép hồi quy hay biến đổi Fourier... Tham khảo thêm tại: http://www.numpy.org/ 	
<div style="display: flex; justify-content: space-between;"> Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275 18/91 </div>	

18

A

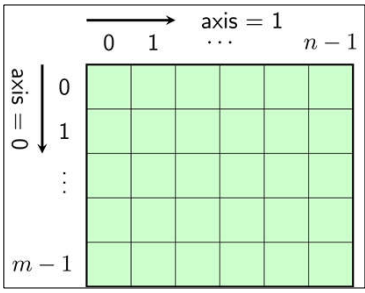
D

I

1. Introduction

• **Đối tượng chính của NumPy** là các mảng đa chiều đồng nhất:

- Kiểu dữ liệu của các phần tử con trong mảng phải giống nhau
- Mảng có thể có 1 chiều hoặc nhiều chiều
- Các chiều được đánh số từ 0 trở đi
- Số chiều được gọi là hạng (**rank**)
- Có đến 24 kiểu số khác nhau.
- Kiểu **ndarray** là lớp chính xử lý dữ liệu mảng nhiều chiều.
- Có rất nhiều hàm và phương thức xử lý ma trận



axis = 1
0 1 ... n - 1
0 1 ... m - 1

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

19/91

19

A

D

I

2. NumPy creating array

1	5	18	23
---	---	----	----

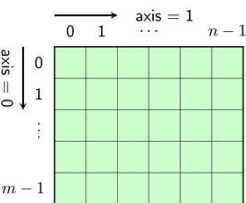
Vector (1D array)
Dimension = 1
(1 index required)

3	12	66
7	9	34
23	45	11

Matrix (2D array)
Dimension = 2
(2 indexes required)

3	12	66
7	9	34
23	45	11

3D array (3rd order Tensor)
Dimension = 3
(3 indexes required)



axis = 1
0 1 ... n - 1
0 1 ... m - 1

ND array
Dimension = N
(N indexes required)

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

20/91

20

A D 2. NumPy creating array A I

- Khởi tạo mảng 1 chiều – 1D (Vector)

```

1 import numpy as np
2
3 a = np.array([1,2,5,7,0,8]) # mảng 1 chiều (1D)
4
5 print(a)
6 print("Loại dữ liệu của biến a:", type(a))
7 print("Kiểu dữ liệu của phần tử trong mảng a:", a.dtype)
8 print("Kích thước của mảng a:", a.shape)
9 print("Số phần tử của mảng a:", a.size)
10 print("Số chiều của mảng a:", a.ndim)

```

[1 2 5 7 0 8]
 Loại dữ liệu của biến a: <class 'numpy.ndarray'>
 Kiểu dữ liệu của phần tử trong mảng a: int32
 Kích thước của mảng a: (6,)
 Số phần tử của mảng a: 6
 Số chiều của mảng a: 1

21/91

21

A D 2. NumPy creating array A I

- Khởi tạo mảng 2 chiều – 2D (Matrix)

```

1 import numpy as np
2
3 b = np.array([[4,5,6],[1,2,3]],dtype=float) # mảng 2 chiều (2D)
4
5 print(b)
6 print("Kiểu dữ liệu của phần tử trong mảng b:", b.dtype)
7 print("Kích thước của mảng b:", b.shape)
8 print("Số phần tử của mảng b:", b.size)
9 print("Số chiều của mảng b:", b.ndim)

```

[[4. 5. 6.]
 [1. 2. 3.]]
 Kiểu dữ liệu của phần tử trong mảng b: float
 Kích thước của mảng b: (2, 3)
 Số phần tử của mảng b: 6
 Số chiều của mảng b: 2

22/91

22

2. NumPy creating array

- Khởi tạo mảng 3 chiều – 3D**

```

1 import numpy as np
2
3 c = np.array([[ (2,4,0,6), (4,7,5,6)],
4               [(0,3,2,1), (9,4,5,6)],
5               [(5,8,6,4), (1,4,6,8)]] #mảng 3 chiều (3D)
6
7 print(c)
8 print("Phần tử đầu tiên của mảng c:",c[0,0,0])
9 print("Kiểu dữ liệu của phần tử trong mảng c:",c.dtype)
10 print("Kích thước của mảng c:",c.shape)
11 print("Số phần tử của mảng c:",c.size)
12 print("Số chiều của mảng c:",c.ndim)

```

[[[2 4 0 6]
 [4 7 5 6]]

 [[0 3 2 1]
 [9 4 5 6]]

 [[5 8 6 4]
 [1 4 6 8]]]
 Phần tử đầu tiên của mảng c: 2
 Kiểu dữ liệu của phần tử trong mảng c: int32
 Kích thước của mảng c: (3, 2, 4)
 Số phần tử của mảng c: 24
 Số chiều của mảng c: 3

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275 23/91

23

2. NumPy creating array

- Khởi tạo mảng với các hàm sẵn có của Numpy**

Initial Placeholders

<pre> >>> np.zeros((3,4)) >>> np.ones((2,3,4),dtype=np.int16) >>> d = np.arange(10,25,5) >>> np.linspace(0,2,9) >>> e = np.full((2,2),7) >>> f = np.eye(2) >>> np.random.random((2,2)) >>> np.empty((3,2)) </pre>	Create an array of zeros Create an array of ones Create an array of evenly spaced values (step value) Create an array of evenly spaced values (number of samples) Create a constant array Create a 2X2 identity matrix Create an array with random values Create an empty array
---	--

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275 24/91

24

A
D

A
I

2. NumPy creating array

• Ví dụ:

```

1 import numpy as np
2 #Tạo ma trận 0 có kích thước 5 hàng x 3 cột
3 a=np.zeros((5,3))
4 print(a)
5 print("Kiểu dữ liệu của các phần tử: ", a.dtype)
6 print("Số chiều của mảng: ", a.ndim)
7 print("Kích thước của ma trận: ", a.shape)
8

```

```

[[0. 0. 0.]
 [0. 0. 0.]
 [0. 0. 0.]
 [0. 0. 0.]
 [0. 0. 0.]]

```

Kiểu dữ liệu của các phần tử: float64
Số chiều của mảng: 2
Kích thước của ma trận: (5, 3)

01

```

1 import numpy as np
2 #Tạo ma trận đơn vị cấp 5
3 array_eye= np.eye(5)
4
5 print(array_eye)
6

```

```

[[1. 0. 0. 0. 0.]
 [0. 1. 0. 0. 0.]
 [0. 0. 1. 0. 0.]
 [0. 0. 0. 1. 0.]
 [0. 0. 0. 0. 1.]]

```

02

```

1 import numpy as np
2 #tạo mảng 10 phần tử bao gồm các số ngẫu nhiên [0,1]
3 b=np.random.random((10))
4
5 print(b)
6 print("Kiểu dữ liệu của các phần tử: ", b.dtype)
7 print("Số chiều của mảng: ", b.ndim)
8 print("Kích thước của ma trận: ", b.shape)
9

```

```

[0.715331 0.83764632 0.13474808 0.26798476 0.68645829 0.35319849
 0.3170353 0.80510223 0.3408744 0.7720831 ]

```

Kiểu dữ liệu của các phần tử: float64
Số chiều của mảng: 1
Kích thước của ma trận: (10,)

03

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

25/91

25

A
D

A
I

2. NumPy creating array

?

Yêu cầu: Tạo một ma trận 10 x 10, bao gồm các phần tử, là những số nguyên ngẫu nhiên trong khoảng (0-100)

```

1 '''
2 Bài tập: Tạo một ma trận 10x10 các số
3 nguyên ngẫu nhiên nằm trong khoảng [0-100]
4 '''
5 import numpy as np
6 b=np.random.random((10,10))*100
7 #Phần tử mặc định tạo ra có kiểu float
8 #Chuyển sang kiểu số nguyên int
9 b = b.astype(np.int64)
10 print(b)
11 print("Kiểu dữ liệu của các phần tử: ", b.dtype)
12 print("Số chiều của mảng: ", b.ndim)
13 print("Kích thước của ma trận: ", b.shape)
14

```

```

[[ 7 87 84 49 96 50 14 87 64 26]
 [35 94 84 48  3 84 37 32 36 39]
 [58 47 23 11 67 58 94 73 97 77]
 [72 26 91 54 78 45 30 46 83 77]
 [56 76 61 36 50 37 24 88 47 98]
 [ 4 31 83 27 63 26 17 54 12 17]
 [ 6 79 27 47 30 64 78 15 60 41]
 [38 65 20 28 22 97  2 63 50 58]
 [77 55 67 76 20 74 29 86 82 24]
 [66 42 44 56 21 42  0 98 88 90]]

```

Kiểu dữ liệu của các phần tử: int64
Số chiều của mảng: 2
Kích thước của ma trận: (10, 10)

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

26/91

26

13

2. NumPy creating array

- Bảng điểm của lớp 2A** (bao gồm 30 học sinh, tương ứng với 30 cột, của 10 môn học, tương ứng với 10 hàng)

Diem_2A - Notepad

Điểm học sinh i
(30 học sinh)

2	4	3	7	5	6	5	6	8	9	3	6	1	9	8	7	3	3	9	5	1	6	5	1
3	5	3	10	9	1	9	8	3	1	6	0	7	10	8	5	2	7	7	1	1	6	1	6
1	10	4	9	6	9	0	2	3	1	8	6	8	4	2	9	2	9	5	0	4	1	7	3
6	3	0	8	3	7	7	2	6	8	7	3	4	1	5	9	1	0	2	10	4	6	8	6
4	3	6	7	4	5	2	6	9	4	3	9	9	4	5	7	2	10	9	4	0	5	3	1
2	3	8	10	4	5	9	5	4	7	10	1	8	4	3	9	6	3	6	7	4	7	3	5
9	9	1	10	9	9	5	9	6	3	9	5	1	10	7	10	2	8	8	1	8	4	5	4
8	8	4	8	0	4	4	8	6	7	1	3	1	6	8	8	4	6	8	4	0	1	8	2
6	7	8	9	10	9	2	2	6	1	10	9	6	3	9	5	9	8	1	1	8	8	8	6
7	8	7	8	6	10	10	6	8	10	8	9	8	8	5	10	8	7	8	7	9	9	8	7

Điểm của môn học j (10 môn học)

- Đọc dữ liệu từ file Diem_2A.txt**

```
diem_2a = np.loadtxt('Diem_2A.txt', delimiter=',')
print('File dữ liệu điểm lớp 2A:\n', diem_2a)
```

File dữ liệu điểm lớp 2A:

```
[[ 2.  4.  3.  7.  5.  6.  5.  6.  8.  9.  3.  6.  1.  9.  8.  7.  3.  3.
  9.  5.  1.  6.  5.  1.  4.  6.  7.  1.  1.  1.]
 [ 3.  5.  3. 10.  9.  1.  9.  8.  3.  1.  6.  0.  7. 10.  8.  5.  2.  7.
  7.  1.  1.  6.  1.  6.  3.  0.  2.  2.  1.  6.]
 [ 1. 10.  4.  9.  6.  9.  0.  2.  3.  1.  8.  6.  8.  4.  2.  9.  2.  9.
  5.  0.  4.  1.  7.  3.  8.  9.  8.  9.  9.  9.]
 [ 6.  3.  0.  8.  3.  7.  7.  2.  6.  8.  7.  3.  4.  1.  5.  9.  1.  0.
  2. 10.  4.  6.  8.  6.  6. 10.  3.  1.  4.  5.]
 [ 4.  3.  6.  7.  4.  5.  2.  6.  9.  4.  3.  9.  9.  4.  5.  7.  2. 10.]
```

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

27/91

27

2. NumPy creating array

```
#a.shape: Cho biết kích thước của mảng a:
print('kích thước của mảng diem_2a:', diem_2a.shape)
```

kích thước của mảng diem_2a: (10, 30)

```
#a.ndim: Cho biết Số chiều của mảng a:
print('Số chiều của mảng diem_2a:', diem_2a.ndim)
```

Số chiều của mảng diem_2a: 2

```
#a.size: Cho biết số phần tử của mảng a:
print('Số phần tử của mảng diem_2a: ', diem_2a.size)
```

Số phần tử của mảng diem_2a: 300

```
#a.dtype: Cho biết kiểu dữ liệu của các phần tử trong mảng a
print('Kiểu dữ liệu của các phần tử trong mảng diem_2a:', diem_2a.dtype)
```

Kiểu dữ liệu của các phần tử trong mảng diem_2a: float64

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

28/91

28

A D
A I

2. NumPy creating array

```

#a.astype(kiểu mới): Chuyển đổi kiểu dữ liệu của mảng a
#Chuyển dữ liệu của mảng diem_2a từ float sang kiểu int

diem_2a_int = diem_2a.astype(np.int16)
print('Dữ liệu mảng diem_2a sau khi chuyển: ', diem_2a_int.dtype)

Dữ liệu mảng diem_2a sau khi chuyển:  int16

#Chuyển dữ liệu sang kiểu string:
diem_2a_str = diem_2a.astype(np.str)
print('Dữ liệu mảng diem_2a sau khi chuyển: ', diem_2a_str.dtype)
print(diem_2a_str)

Dữ liệu mảng diem_2a sau khi chuyển:  <U32
[['2.0' '4.0' '3.0' '7.0' '5.0' '6.0' '5.0' '6.0' '8.0' '9.0' '3.0' '6.0'
  '1.0' '9.0' '8.0' '7.0' '3.0' '3.0' '9.0' '5.0' '1.0' '6.0' '5.0' '1.0']

```

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275
29/91

29

A D
A I

2. NumPy creating array

NumPy dtypes

Basic Type	Available NumPy types	Comments
Boolean	bool	Elements are 1 byte in size
Integer	int8, int16, int32, int64, int128, int	int defaults to the size of int in C for the platform
Unsigned Integer	uint8, uint16, uint32, uint64, uint128, uint	uint defaults to the size of unsigned int in C for the platform
Float	float32, float64, float, longfloat,	Float is always a double precision floating point value (64 bits). longfloat represents large precision floats. Its size is platform dependent.
Complex	complex64, complex128, complex	The real and complex elements of a complex64 are each represented by a single precision (32 bit) value for a total size of 64 bits.
Strings	str, unicode	Unicode is always UTF32 (UCS4)
Object	object	Represent items in array as Python objects.
Records	void	Used for arbitrary data structures in record arrays.

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275
30/91

30

A

D

A

I

2. Array Indexing and Slicing

- Truy cập tới phần tử cụ thể trong mảng sử dụng chỉ số mảng (chỉ số bắt đầu từ 0):
 - `a[i]` : truy cập tới phần tử thứ i của mảng một chiều
 - `a[i,j]` : Truy cập tới phần tử hàng i , cột j của mảng 2 chiều
 - `a[n,i,j]` : Truy cập tới phần tử chiều n , hàng i , cột j của mảng 3 chiều

```
#Truy cập phần tử của mảng:
print('các phần tử của mảng 1 chiều a:\n', a)
print('phần tử thứ 5 của mảng:', a[4]).
```

các phần tử của mảng 1 chiều a:
[3. 5. 3. 10. 9. 1. 9. 8. 3. 1.]

phần tử thứ 5 của mảng: 9.0

```
#Truy cập phần tử của mảng 2 chiều:
print('Điểm môn học thứ 1, của học sinh thứ 3 là:', diem_2a[0,2])
print('Bảng điểm lớp 2A:\n', diem_2a)
```

Điểm môn học thứ 1, của học sinh thứ 3 là: 3.0

Bảng điểm lớp 2A:

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

axis = 1
0 1 ... n-1
0 1 ... m-1

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

31/91

31

A

D

A

I

2. Array Indexing and Slicing

- Truy cập nhiều phần tử trong mảng sử dụng ký hiệu “.”
 - `a[3:8]`: Truy cập tới các phần tử thứ 4 tới phần tử thứ 9 của mảng một chiều `a`
 - `b[:3, :]` : Truy cập tới phần tử từ hàng 0 tới hàng 3, của tất cả các cột trong mảng 2 chiều

```
#Truy cập tới nhiều phần tử
#Lấy điểm tất cả các môn (t)
diem_hs5 = diem_2a[:,4]

print("Điểm các môn của học sinh")
```

Điểm các môn của học sinh

```
diem_hs = diem_2a[:,10:20]
print("Bảng điểm từ học sinh 10 tới học sinh 20:\n", diem_hs)
```

Bảng điểm từ học sinh 10 tới học sinh 20:

```
[ [ 3. 6. 1. 9. 8. 7. 3. 3. 9. 5.]
  [ 6. 0. 7. 10. 8. 5. 2. 7. 7. 1.]
  [ 8. 6. 8. 4. 2. 9. 2. 9. 5. 0.]
  [ 7. 3. 4. 1. 5. 9. 1. 0. 2. 10.]
  [ 3. 9. 9. 4. 5. 7. 2. 10. 9. 4.]
  [10. 1. 8. 4. 3. 9. 6. 3. 6. 7.]
  [ 9. 5. 1. 10. 7. 10. 2. 8. 8. 1.]
  [ 1. 3. 1. 6. 8. 8. 4. 6. 8. 4.]
  [10. 9. 6. 3. 9. 5. 9. 8. 1. 1.]
  [ 8. 9. 8. 8. 5. 10. 8. 7. 8. 7.] ]
```

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

32/91

32

A D
A I

2. Array Indexing and Slicing

```

1 import numpy as np
2 np_matrix=np.random.random((4,6))
3 print(np_matrix)
4 # Hiển thị dòng đầu tiên của matrix
5 print('result1')
6 print(np_matrix[0:1])
7 print(np_matrix[0:1].shape)
8 #Lấy dòng cuối của matrix==>list
9 print(np_matrix[0:1][-1])
10 print(np_matrix[0:1][-1].shape)
11 #Lấy phần tử từ 1 đến 2 của list
12 print(np_matrix[0:1][-1][1:3])

```

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

33/91

33

A D
A I

3.Các phép toán trên mảng (1)

- Cho phép thực hiện +, -, *, /, ...các phần tử trên mảng như các phần tử riêng biệt.

```

1 #Thực hiện phép toán giữa biến array và một giá trị
2 a=np.array((7,5,8,1))
3 print('Mảng a: ', a)
4
5 x = 3
6 # Các phép toán giữa array với một giá trị
7 array_tong = a + x
8 print("Tổng mảng a + x:", array_tong)
9
10 array_hieu = a - x
11 print("Tổng mảng a - x:", array_hieu)
12
13 array_tich=a*x
14 print("Tích mảng a * b:", array_tich)
15
16 array_thuong=a/x
17 print("Tích mảng a / b:", array_thuong)
18
19 array_thuongnguyen=a//x
20 print("Thương nguyên mảng a // b:", array_thuongnguyen)
21
22 array_thuongdu=a%x
23 print("Thương dư của mảng a % b:", array_thuongdu)

```

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

34/91

34

A
D

A
I

Thực hành 1

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275


35/91

35

A
D

A
I

Thực hành 1



Yêu cầu 1: Học viên tạo một ma trận vuông cấp n (n hàng x n cột), bao gồm các phần tử, là những số nguyên ngẫu nhiên trong khoảng (0-100) như minh họa, với n = 12

```

[[61 21 68 72 84 90 81 80 79 80 49 53]
 [92 10 1 43 49 93 76 6 2 6 69 2]
 [12 79 88 10 37 55 37 6 59 75 77 64]
 [97 22 75 32 39 39 93 19 28 64 55 87]
 [88 25 88 58 11 96 58 14 88 16 22 64]
 [ 3 5 60 14 65 50 80 42 8 27 44 52]
 [84 38 54 27 86 13 67 77 77 12 66 40]
 [96 9 94 24 61 19 2 80 95 92 72 32]
 [49 21 78 92 35 92 84 86 85 62 64 29]
 [ 5 53 95 2 43 30 72 66 97 17 8 23]
 [16 27 1 71 19 22 90 81 12 93 14 53]
 [40 50 83 25 37 16 49 73 42 86 11 18]]
        
```

Kiểu dữ liệu của phần tử trong ma trận: int32
 Kích thước của mảng ma trận: (12, 12)
 Số phần tử của mảng ma trận: 144
 Số chiều của mảng ma trận: 2

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

36/91

36

A D
A I

Thực hành 1

?

Yêu cầu 2: Sử dụng ma trận tạo được trong yêu cầu 1, Học viên tạo 2 vector như sau:

- v_chinh:** bao gồm các phần tử nằm trên đường chéo chính của ma trận.
- V_phu:** bao gồm các phần tử nằm trên đường chéo phụ của ma trận

```
[[ 61, 21, 68, 72, 84, 90, 81, 80, 79, 80, 49, 53]
[ 92, 10, 1, 43, 49, 93, 76, 6, 2, 6, 69, 2]
[12, 79, 88, 10, 37, 55, 37, 6, 59, 75, 77, 64]
[ 97, 22, 75, 32, 39, 39, 93, 19, 28, 64, 55, 87]
[ 88, 25, 88, 58, 11, 96, 58, 14, 88, 16, 22, 64]
[ 3, 5, 60, 14, 65, 50, 80, 42, 8, 27, 44, 52]
[ 84, 38, 54, 27, 86, 13, 67, 77, 77, 12, 66, 40]
[ 96, 9, 94, 24, 61, 19, 2, 80, 95, 92, 72, 32]
[ 49, 21, 78, 92, 35, 92, 84, 86, 85, 62, 64, 29]
[ 5, 53, 95, 2, 43, 30, 72, 66, 97, 17, 8, 23]
[16, 27, 1, 71, 19, 22, 90, 81, 12, 93, 14, 53]
[ 40, 50, 83, 25, 37, 16, 49, 73, 42, 86, 11, 18]]
```

Vector các phần tử nằm trên đường chéo chính:
[61, 10, 88, 32, 11, 50, 67, 80, 85, 17, 14, 18.]

Vector các phần tử nằm trên đường chéo phụ:
[40, 27, 95, 92, 61, 13, 80, 14, 28, 75, 69, 53.]

Dr. Tà Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

37/91

37

A D
A I

Thực hành 1

?

Yêu cầu 3: Nhập vào số nguyên x bất kỳ trong khoảng (0-100), đếm xem có bao nhiêu phần tử trong ma trận sinh ra ở yêu cầu 1 có giá trị bằng, lớn hơn và nhỏ hơn giá trị x:

```
[[ 61, 21, 68, 72, 84, 90, 81, 80, 79, 80, 49, 53]
[ 92, 10, 1, 43, 49, 93, 76, 6, 2, 6, 69, 2]
[12, 79, 88, 10, 37, 55, 37, 6, 59, 75, 77, 64]
[ 97, 22, 75, 32, 39, 39, 93, 19, 28, 64, 55, 87]
[ 88, 25, 88, 58, 11, 96, 58, 14, 88, 16, 22, 64]
[ 3, 5, 60, 14, 65, 50, 80, 42, 8, 27, 44, 52]
[ 84, 38, 54, 27, 86, 13, 67, 77, 77, 12, 66, 40]
[ 96, 9, 94, 24, 61, 19, 2, 80, 95, 92, 72, 32]
[ 49, 21, 78, 92, 35, 92, 84, 86, 85, 62, 64, 29]
[ 5, 53, 95, 2, 43, 30, 72, 66, 97, 17, 8, 23]
[16, 27, 1, 71, 19, 22, 90, 81, 12, 93, 14, 53]
[ 40, 50, 83, 25, 37, 16, 49, 73, 42, 86, 11, 18]]
```

Nhập vào giá trị x (0-100):88

- Số phần tử có giá trị bằng x trong ma trận: 4
- Số phần tử nhỏ hơn giá trị x trong ma trận: 124
- Số phần tử lớn hơn giá trị x trong ma trận: 16

Dr. Tà Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

38/91

38

A
D
A
I

4. Tính toán các đặc trưng thống kê với mảng

Dr. Tà Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

39/91

39

A
D
A
I

4 Mathematical statistics functions

The diagram consists of three overlapping circles. The top-left circle is light gray and labeled 'Computer Science/IT' with a laptop icon. The top-right circle is dark gray and labeled 'Math and Statistics' with a line graph icon. The bottom circle is maroon and labeled 'Domains/Business Knowledge' with a smartphone icon. The central intersection of all three circles is labeled 'Data Science'. The intersection of the top two circles is labeled 'Machine Learning'. The intersection of the top-left and bottom circles is labeled 'Software Development'. The intersection of the top-right and bottom circles is labeled 'Traditional Research'.

Toán học và thống kê có một vai trò rất quan trọng trong khoa học dữ liệu!

Dr. Tà Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

40/91

40

A

D


I

4 Statistics functions : Max - Min

- a.max():** Lấy giá trị lớn nhất của mảng a
- b.min():** Lấy giá trị nhỏ nhất của mảng b

```
#Xác định giá trị lớn nhất, nhỏ nhất:
#Liệt kê điểm cao nhất và thấp nhất của mỗi môn học
for i in range(0,diem_2a.shape[0]):
    print('Môn ', i+1,': Điểm Max: ', diem_2a[i,:].max(),
          '-- Điểm Min:',diem_2a[i,:].min())
```

```
Môn 1 : Điểm Max: 9.0 -- Điểm Min: 1.0
Môn 2 : Điểm Max: 10.0 -- Điểm Min: 0.0
Môn 3 : Điểm Max: 10.0 -- Điểm Min: 0.0
Môn 4 : Điểm Max: 10.0 -- Điểm Min: 0.0
Môn 5 : Điểm Max: 10.0 -- Điểm Min: 0.0
Môn 6 : Điểm Max: 10.0 -- Điểm Min: 1.0
Môn 7 : Điểm Max: 10.0 -- Điểm Min: 1.0
Môn 8 : Điểm Max: 9.0 -- Điểm Min: 0.0
Môn 9 : Điểm Max: 10.0 -- Điểm Min: 1.0
Môn 10 : Điểm Max: 10.0 -- Điểm Min: 5.0
```



Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

41/91

41

A

D

I


4 Statistics functions : Sum

- a.sum() – np.sum(a):** Tính tổng tất cả các phần tử của mảng a

```
#Sum:Tính tổng các phần tử
print('Tổng tất các điểm trong của lớp 2A:', np.sum(diem_2a))

#Tính tổng điểm của từng học sinh
for i in range(0,30):
    print('Tổng điểm các môn của học sinh ', i+1, ' : ', diem_2a[:,i].sum())
```

```
Tổng tất các điểm trong của lớp 2A: 1662.0
Tổng điểm các môn của học sinh 1 : 48.0
Tổng điểm các môn của học sinh 2 : 60.0
Tổng điểm các môn của học sinh 3 : 44.0
Tổng điểm các môn của học sinh 4 : 86.0
Tổng điểm các môn của học sinh 5 : 56.0
```



Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

42/91

42

A

D

A

I

4. Statistics functions

Statistics – Mean, Median, Mode and Range
 EZY MATHS

Mean	Median	Mode	Range
Mean = $\frac{\text{Total of all values}}{\text{number of values}}$ 3, 3, 4, 5, 5, 8, 9, 15 $\text{Mean} = \frac{52}{8} = 6.5$	Median = Middle value (Numbers written in order) 3, 3, 4, 5 , 5, 8, 9, 15 $\text{Median} = 5$	Mode = Most common value/item 3 , 3 , 4 , 5 , 5 , 8, 9, 15 Mode = 3 and 5	Range = Largest - Smallest 3, 3, 4, 5, 5, 8, 9, 15 Range = 15 - 3 = 12
Collect it all together and share it out evenly	Finds the middle value	Average usually used for qualitative data	Reveals how close/far apart the values are
Using the mean to find the total amount $\text{Mean} \times \text{Number of values}$ Ezytown FC have scored an average of 3.8 goals per game in their last 15 matches. How many goals have they scored? $3.8 \times 15 = 57 \text{ goals}$	Use of formula to find location of median $\text{Location} = \frac{n+1}{2}$ The median of 45 values would be the 23 rd number when written in order $\frac{45+1}{2} = 23$	Occurrence of no mode If every value appears equally, there is no mode 1, 1, 3, 3, 7, 7 Each value appears twice so there is no mode	Interpreting measures of spread The Smaller the range, the closer and more 'consistent' the values are. The Larger the range, the more varied and more 'inconsistent' the values are.

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

43/91

43

A

D

A

I

4. Statistics functions : Mean

Statistics – Mean, Median, Mode and Range
 EZY MATHS

```
#mean(): Giá trị trung bình
#Tính điểm trung bình của các học sinh trong lớp:
#CÁCH 1:
for i in range(0,30):
    print('Điểm trung bình của học sinh ', i+1, ' : ', diem_2a[:,i].mean())

Điểm trung bình của học sinh 1 : 4.8
Điểm trung bình của học sinh 2 : 6.0
Điểm trung bình của học sinh 3 : 4.4
Điểm trung bình của học sinh 4 : 8.6
Điểm trung bình của học sinh 5 : 5.6

#CÁCH 2:
mean_2a = diem_2a.mean(axis=0)

for i in range(0,mean_2a.size):
    print('Điểm trung bình của học sinh ', i+1, ' : ', mean_2a[i])

Điểm trung bình của học sinh 1 : 4.8
Điểm trung bình của học sinh 2 : 6.0
Điểm trung bình của học sinh 3 : 4.4
Điểm trung bình của học sinh 4 : 8.6
Điểm trung bình của học sinh 5 : 5.6
```

Mean
Mean = $\frac{\text{Total of all values}}{\text{number of values}}$ 3, 3, 4, 5, 5, 8, 9, 15 $\text{Mean} = \frac{52}{8} = 6.5$
Collect it all together and share it out evenly
Using the mean to find the total amount $\text{Mean} \times \text{Number of values}$ Ezytown FC have scored an average of 3.8 goals per game in their last 15 matches. How many goals have they scored? $3.8 \times 15 = 57 \text{ goals}$

Mean

Mean = $\frac{\text{Total of all values}}{\text{number of values}}$
 3, 3, 4, 5, 5, 8, 9, 15
 $\text{Mean} = \frac{52}{8} = 6.5$

Collect it all together and share it out evenly

Using the mean to find the total amount
 $\text{Mean} \times \text{Number of values}$
 Ezytown FC have scored an average of 3.8 goals per game in their last 15 matches. How many goals have they scored?
 $3.8 \times 15 = 57 \text{ goals}$

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

44/91

44

A
D
A
I

4. Statistics functions: Median

```
#median(): Giá trị giữa trong một tập hợp các phần tử.
#Trường hợp số phần tử trong mảng là lẻ
a=diem_2a[:,5]
b=np.array([9,8,6])
a=np.append(a,b) #nối mảng b vào mảng a

print('Mảng a ban đầu: \n', a)
print('Số phần tử trong mảng a: ', a.size)
print('Mảng a đã sắp xếp: \n',np.sort(a,))
print('Giá trị trung bình mean:', np.mean(a))
print('Giá trị trung vị median:', np.median(a))
```

Mảng a ban đầu:
[6. 1. 9. 7. 5. 5. 9. 4. 9. 10. 9. 8. 6.]
Số phần tử trong mảng a: 13
Mảng a đã sắp xếp:
[1. 4. 5. 5. 6. 6. 7. 8. 9. 9. 9. 9. 10.]
Giá trị trung bình mean: 6.769230769230769
Giá trị trung vị median: 7.0

Median

Median = Middle value
(Numbers written in order)

3, 3, 4, 5, 5, 8, 9, 15

Median = 5

Finds the middle value

Use of formula to find location of median

$Location = \frac{n + 1}{2}$

The median of 45 values would be the 23rd number when written in order

$\frac{45 + 1}{2} = 23$

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

45/91

45

A
D
A
I

4. Statistics functions: std

Độ lệch tiêu chuẩn (*standard deviation*) là đại lượng thường được sử dụng để phản ánh mức độ phân tán của một biến số xung quanh số bình quân.

```
#Std: độ lệch chuẩn
#Tính độ lệch chuẩn điểm thi của từng học sinh
for i in range(0,30):
    print('Độ lệch chuẩn của học sinh ', i+1, ' : ', diem_2a[:,i].std())
```

Độ lệch chuẩn của học sinh 1 : 2.6381811916545836
 Độ lệch chuẩn của học sinh 2 : 2.569046515733026
 Độ lệch chuẩn của học sinh 3 : 2.6532998
 Độ lệch chuẩn của học sinh 4 : 1.1135528
 Độ lệch chuẩn của học sinh 5 : 2.9393876
 Độ lệch chuẩn của học sinh 6 : 2.6925824

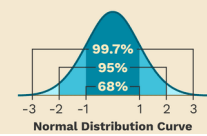
Calculating Standard Deviation

$$S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

n = The number of data points

x_i = Each of the values of the data

\bar{x} = The mean of **x_i**



Normal Distribution Curve

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

46/91

46

A

D

A

I

4. Statistics functions: Mode

#Mode: phần tử xuất hiện nhiều nhất
#Liệt kê điểm xuất hiện nhiều nhất theo môn học
from scipy import stats as sp #sử dụng thư viện scipy để dùng hàm mode

```
for i in range(0,diem_2a.shape[0]):
    a = sp.mode(diem_2a[i,:])
    print('Môn ', i+1,': Điểm xuất hiện nhiều nhất: ', a[0],
          ' số lần: ', a[1])

print(type(a))
```

Môn 1 : Điểm xuất hiện nhiều nhất: [1.]	số lần: [6]
Môn 2 : Điểm xuất hiện nhiều nhất: [1.]	số lần: [6]
Môn 3 : Điểm xuất hiện nhiều nhất: [9.]	số lần: [8]
Môn 4 : Điểm xuất hiện nhiều nhất: [6.]	số lần: [5]
Môn 5 : Điểm xuất hiện nhiều nhất: [4.]	số lần: [6]
Môn 6 : Điểm xuất hiện nhiều nhất: [5.]	số lần: [5]
Môn 7 : Điểm xuất hiện nhiều nhất: [9.]	số lần: [7]
Môn 8 : Điểm xuất hiện nhiều nhất: [8.]	số lần: [9]
Môn 9 : Điểm xuất hiện nhiều nhất: [8.]	số lần: [7]
Môn 10 : Điểm xuất hiện nhiều nhất: [8.]	số lần: [9]

<class 'scipy.stats.stats.ModeResult'>

Mode
 Mode = Most common value/item
3,3,4,5,58,9,15
 Mode = 3 and 5

Average usually used for qualitative data

Occurrence of no mode
 If every value appears equally, there is **no mode**

1,1,3,3,7,7
 Each value appears twice so there is no mode

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

47/91

47

A

D

A

I

4. Statistics functions: Range

Trong thư viện numpy không có hàm tính range, ta có thể xác định giá trị range bằng cách tính thông qua max - min

#Range: là sự khác biệt, khoảng cách giữa phần tử dưới và phần tử trên,
#giữa giá trị nhỏ nhất (Min) với giá trị lớn nhất (Max) trong tập hợp.
#Xác định độ chênh điểm max - min của từng học sinh

```
for i in range(0,30):
    print('Độ chênh điểm của học sinh ', i+1,': ',
          diem_2a[:,i].max()-diem_2a[:,i].min())
```

Độ chênh điểm của học sinh 1 :	8.0
Độ chênh điểm của học sinh 2 :	7.0
Độ chênh điểm của học sinh 3 :	8.0
Độ chênh điểm của học sinh 4 :	3.0
Độ chênh điểm của học sinh 5 :	10.0
Độ chênh điểm của học sinh 6 :	9.0

Range
 Range = Largest - Smallest
 3, 3, 4, 5, 5, 8, 9, 15
 Range = 15 - 3 = 12

Reveals how close/far apart the values are

Interpreting measures of spread
 The Smaller the range, the closer and more 'consistent' the values are.

The Larger the range, the more varied and more 'inconsistent' the values are.

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

48/91

48

A

D

A

I

4. Statistics functions: corrcoef

Hệ số tương quan đo lường mức độ quan hệ tuyến tính giữa hai biến.

- Hệ số tương quan không có đơn vị
- Hệ số tương quan nằm trong khoảng $[-1, 1]$

Strength of correlation

Author: NNB

Correlation Coefficient Formula

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Positive correlation

Negative correlation

No correlation

Author: NNB

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

49/91

49

A

D

A

I

4. Statistics functions: corrcoef

```
#corrcoef: Hệ số tương quan
#Thời gian dành cho học bài
a_giohoc = np.array([4,7,1,2,8,0,3,8,6])
#Điểm thi nhận được:
b_diem = np.array([7,9,3,4,9,0,5,10,8])
co = np.corrcoef(a_giohoc,b_diem)
print(type(co))
print('Hệ số tương quan: \n', co)
```

Ví dụ về mối tương quan giữa **thời gian dành cho việc học bài** với **điểm thi nhận được**!

```
<class 'numpy.ndarray'>
Hệ số tương quan:
[[1.          0.96995403]
 [0.96995403  1.        ]]
```

BIỂU ĐỒ THỂ HIỆN MỐI TƯƠNG QUAN GIỮA GIỜ HỌC BÀI VÀ ĐIỂM THI

Giờ học bài (tuần)

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

50/91

50

Thực hành 2

46

51

A

D

A

I

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

52/91

52

A D
A I

Thực hành 2

• Yêu cầu 1: Sử dụng dữ liệu bảng điểm của lớp 2A. Cho biết:

1. ĐTB của từng học sinh trong lớp.
2. Học sinh có điểm TB cao nhất.
3. Học sinh có điểm trung bình thấp nhất

Điểm TB của từng học sinh trong lớp:
[4.8 6. 4.7 8.6 6.2 6.8 5.6 5.4 5.9 5.1 7.2 5.4 5.9 5.9 6. 7.9 4.2 6.1
6.3 4.4 4.7 5.9 5.6 4.7 6.4 6.2 6.5 4.6 5.8 4.3]

Điểm TB cao nhất: 8.6
Của học sinh thứ: 3
Bảng điểm đầy đủ của học sinh: [7 10 9 8 7 10 10 8 9 8]

Điểm TB thấp nhất: 4.2
Của học sinh thứ: 16
Bảng điểm đầy đủ của học sinh: [3 2 2 1 2 6 2 7 9 8]

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275
53/91

53

A D
A I

Thực hành 2

• Yêu cầu 2: Sử dụng dữ liệu bảng điểm của lớp 2A. Cho biết:

1. ĐTB của từng môn học.
2. Môn học có điểm TB cao nhất.
3. Môn học có điểm trung bình thấp nhất

Điểm TB của từng môn học:
[4.73 4.43 5.5 4.83 4.97 5.6 6.23 7.3 6.13 7.97]

Điểm TB cao nhất của môn học: 7.97
Môn học thứ [9]
Bảng điểm đầy đủ của môn học: [[7 8 7 8 6 10 10 6 8 10 8 9 8 8 5 10 8 7 8 7 9 9 8 7
7 7 10 8 9 7]]

Điểm TB thấp nhất của môn học: 4.43
Môn học thứ [1]
Bảng điểm đầy đủ của môn học: [[3 5 3 10 9 1 9 8 3 1 6 0 7 10 8 5 2 7 7 1 1 6 1 6
3 0 2 2 1 6]]

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275
54/91

54

ADAI

Thực hành 3

Yêu cầu 3: Sử dụng dữ liệu bảng điểm của lớp 2A. Cho biết:

?

1. Sinh viên có điểm đồng đều nhất tất cả các môn. Sinh viên có điểm các môn lệch nhất trong lớp.
2. Môn học có điểm đồng đều nhất. Môn học có điểm chênh lệch nhất.

Học sinh học đồng đều nhất: [3]
 Bảng điểm của học sinh học đồng đều: [[7 10 9 8 7 10 10 8 9 8]]

Học sinh học lệch nhất: [19]
 Bảng điểm của học sinh học lệch: [[5 1 0 10 4 7 1 8 1 7]]

Môn học có điểm đồng đều nhất: [7]
 Bảng điểm của môn học đồng đều: [[8 8 7 8 6 7 7 8 6 7 8 6 7 6 8 8 7 6 8 8 8 7 8 8 8 6 8 7 7 8]]

Môn học có điểm lệch nhất: [2]
 Bảng điểm của môn học lệch: [[1 10 4 9 6 9 0 2 3 1 8 6 8 4 2 9 2 9 5 0 4 1 7 3 8 9 8 9 9 9]]

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

55/91

55

ADAI

5. Hệ số tương quan: corrccoef

Hệ số tương quan đo lường mức độ quan hệ tuyến tính giữa hai biến.

- Hệ số tương quan không có đơn vị
- Hệ số tương quan nằm trong khoảng [-1,1]

Correlation Coefficient Formula

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Strength of correlation

Positive correlation
Negative correlation
No correlation

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

56/91

56

28

A **D**
A **I**

5. Hệ số tương quan: corrcoef

#corrcoef: Hệ số tương quan
 #Thời gian dành cho học bài
 a_giohoc = np.array([4,7,1,2,8,0,3,8,6])
 #Điểm thi nhận được:
 b_diem = np.array([7,9,3,4,9,0,5,10,8])
 co = np.corrcoef(a_giohoc,b_diem)
 print(type(co))
 print('Hệ số tương quan: \n', co)

```
<class 'numpy.ndarray'>
Hệ số tương quan:
[[1.          0.96995403]
 [0.96995403  1.        ]]
```

Ví dụ về mối tương quan giữa **thời gian dành cho việc học bài** với **điểm thi nhận được!**

BIỂU ĐỒ THỂ HIỆN MỐI TƯƠNG QUAN GIỮA GIỜ HỌC BÀI VÀ ĐIỂM THI

Giờ học bài (tuần)	Điểm môn học
0	0
1	3
2	4
3	5
4	7
6	8
7	9
8	9
8	10

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

57/91

57

A **D**
A **I**

Thực hành 3

Thực hành 3

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

58/91

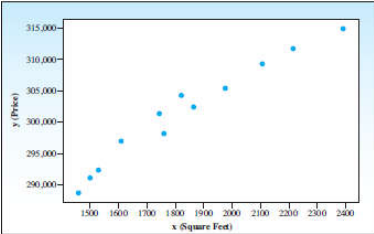
58

Thực hành 3

Yêu cầu: Xác định hệ số tương quan giữa diện tích (1) | Khoảng cách từ trung tâm thành phố (2) và giá bán nhà theo bảng số liệu dưới đây:

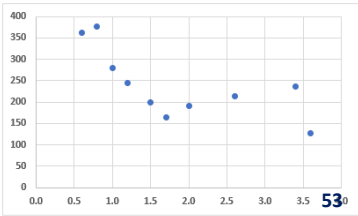
Square Feet, x	Price, y	Square Feet, x	Price, y
1460	\$288,700	1977	\$305,400
2108	309,300	1610	297,000
1743	301,400	1530	292,400
1499	291,100	1759	298,200
1864	302,400	1821	304,300
2391	314,900	2216	311,700

Plot of data for Exercise 12.42



01

Distance (in Miles from the city centre)	Prices (in \$ x 1000)
2.6	214
0.8	376
1.0	280
0.6	362
1.5	200
2.0	190
3.4	236
1.2	244
3.6	128
1.7	165



02

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

59/91

59

6. Kết hợp, chuyển đổi Vector, Matrix

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275


60/91

60

A
D

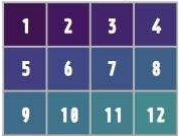
A
I

6.1 Reshape Arrays (3)

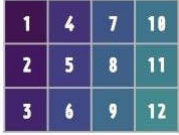
`a1 = np.arange(1, 13)`


Sắp xếp thứ tự các phần tử khi reshape array sử dụng
 thuộc tính **order='C' | 'F'**

→



↓



```

a1.reshape(3, 4) # reshapes or 'fills in' row by row
a1.reshape(3, 4, order='C') # same results as above
a1.reshape(3, 4, order='F') # reshapes column by column
      
```

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275


63/91

63

A
D

A
I


6.2 Flatten|ravel Arrays



↓

`np.ravel()`

↓



```

1 #Chuyển đổi từ Matrix --> Vector
2
3 a1_2d = np.array([(1,2,3,4),(5,6,7,8),(9,10,11,12)])
4 print('Matrix: \n', a1_2d)
5
6 print('-----')
7 print('a) ravel by row (default order=\'C\')')
8 print(a1_2d.ravel())
9
10 print('\n b) ravel by column (order=\'F\')')
11 print(a1_2d.ravel(order='F'))
      
```

Matrix:

```

[[ 1  2  3  4]
 [ 5  6  7  8]
 [ 9 10 11 12]]
      
```

a) ravel by row (default order='C')

```

[ 1  2  3  4  5  6  7  8  9 10 11 12]
      
```

b) ravel by column (order='F')

```

[ 1  5  9  2  6 10  3  7 11  4  8 12]
      
```

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

64/91

64

32

A D A I 6.3 Concatenate Arrays (1)

The function name

The axis along which you will concatenate the arrays

The NumPy arrays you want to concatenate together (required)

```
np.concatenate((arr1, arr2), axis = )
```

Setting `axis=0` concatenates along the row axis

Setting `axis=1` concatenates along the column axis

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

65/91

65

A D A I 6.4 Split Arrays (1)

Split: Tách một vector, ma trận thành các vector, ma trận con

`np.split(2)`

`np.split(x, [2, 6])`

```

1 import numpy as np
2 x = np.arange(0,6)
3 print(x)
4
5 #Tách vector x thành 2 vector
6 #có số phần tử bằng nhau
7 x1, x2 = np.split(x, 2)
8 print(x1, x2)

```

```

1 import numpy as np
2 x = np.arange(1,10)
3 print(x)
4
5 #Tách vector x thành 3 vector
6 #tại các vị trí 2 và 6
7 x1, x2, x3 = np.split(x, [2,6])
8 print(x1, x2, x3)

```

```
[0 1 2 3 4 5]
[0 1 2] [3 4 5]
```

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

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

66/91

66

A D A I 6.4 Split Arrays (2)

Vsplit, hsplit: Tách một ma trận thành các ma trận con theo hàng, cột

np.vsplit(2)

np.hsplit(a, 2)

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

67/91

67

A D A I 6.5 Flip

Ma trận ban đầu:

```
[[ 1  2  3  4  5]
 [ 6  7  8  9 10]
 [11 12 13 14 15]
 [16 17 18 19 20]
 [21 22 23 24 25]]
```

- **np.flip(A,1) ~ np.fliplr(A):** Lật ngược ma trận A theo cột.

```
1 #Lật ma trận theo cột
2 A1 = np.flip(A,1)
3 #Tương đương với
4 A1 = np.fliplr(A)
5 print('Lật ma trận theo cột: \n',A1)
```

Lật ma trận theo cột:

```
[[ 5  4  3  2  1]
 [10  9  8  7  6]
 [15 14 13 12 11]
 [20 19 18 17 16]
 [25 24 23 22 21]]
```

- **np.flip(A,0) ~ np.flipud(A):** Lật ngược ma trận A theo hàng.

```
1 #Lật ma trận theo hàng
2 A2 = np.flip(A,0)
3 #Tương đương với
4 A2 = np.flipud(A)
5 print('Lật ma trận theo hàng: \n',A2)
```

Lật ma trận theo hàng:

```
[[21 22 23 24 25]
 [16 17 18 19 20]
 [11 12 13 14 15]
 [ 6  7  8  9 10]
 [ 1  2  3  4  5]]
```

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

68/91

68

A
D

A
I

Thực hành 4

Dr. Tà Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

69/91

69

A
D

A
I

Thực hành 4

?

Yêu cầu: Tạo một vector gồm 30 phần tử, có giá trị tăng dần từ 1 đến 30.

Vector a:

```
[ 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
25 26 27 28 29 30]
```

Chỉ sử dụng các phương thức reshape, ravel, split...tách Vector a ở trên thành 3 vector con bao gồm:

- a_le:** chứa các phần tử là số lẻ;
- a_chan:** chứa các phần tử là số chẵn.
- a_3:** chứa các phần tử chia hết cho 3

```
Vector a_le : [ 1  3  5  7  9 11 13 15 17 19 21 23 25 27 29]
Vector a_chan: [ 2  4  6  8 10 12 14 16 18 20 22 24 26 28 30]
Vector a_3   : [ 3  6  9 12 15 18 21 24 27 30]
```

Dr. Tà Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

70/91

70

A	D	
A	I	
<h1>BÀI TẬP</h1>		
Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275		71/91

71

A	D	
A	I	
<h2>BÀI TẬP 2-1</h2> <p>Các file dữ liệu cho phần bài tập: Temp.txt Diamonds.txt học viên download trong phần tài nguyên của bài học (Data Exercise).</p>		
Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275		72/91

72

A
D
A
I
Làm việc với numpy - Bài Tập 2-1

File dữ liệu: Temp.txt

Temp - Notepad

File Edit Format View Help

```

25.65 24.79 24.01 25.06 25.48 24.97
25.31 24.21 24.02 24.93 25.16 24.83
25.05 23.73 23.89 24.79 24.80 24.55
24.79 23.36 23.83 24.84 24.74 24.48
24.59 23.05 23.69 24.82 24.80 24.38
24.40 22.80 23.52 24.79 24.87 24.40
24.38 22.79 23.68 25.10 24.71 24.41
26.72 25.61 24.92 26.56 25.03 24.91
28.84 26.93 26.51 26.53 25.75 25.85
30.29 28.72 27.48 26.95 26.64 26.79
31.35 29.97 26.96 27.23 27.68 27.53
32.05 28.93 26.86 27.38 28.43 28.98
31.31 28.94 26.65 27.47 28.29 29.24
30.95 30.25 27.83 27.44 28.00 30.66
30.56 30.62 26.49 27.16 27.67 30.97
31.13 30.58 26.29 26.68 27.29 30.59

```

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275


73/91

73

A
D
A
I
Làm việc với numpy - Bài Tập 2-1

Mô tả file dữ liệu: Temp.txt

- File dữ liệu lưu trữ nhiệt độ (°C) của 6 thành phố lớn dọc theo nước Việt Nam là: Hà Nội, Vinh, Đà Nẵng, Nha trang, Hồ Chính Minh và Cà Mau
- Thời gian từ 0h ngày 15/09/2019 tới 23h ngày 22/09/2019



(https://www.meteoblue.com/en/weather/week/da-nang-vietnam_1583992)

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

74/91

74

A D
A I
Làm việc với numpy - Bài Tập 2-1

Mô tả file dữ liệu: Temp.txt

Hà Nội	Vinh	Đà Nẵng	Nha Trang	HCM	Cà Mau	
25.65	24.79	24.01	25.06	25.48	24.97	Time: 0h 15/09
25.31	24.21	24.02	24.93	25.16	24.83	1h 15/09
25.05	23.73	23.89	24.79	24.80	24.55	
24.79	23.36	23.83	24.84	24.74	24.48	
24.59	23.05	23.69	24.82	24.80	24.38	
24.40	22.80	23.52	24.79	24.87	24.40	
24.38	22.79	23.68	25.10	24.71	24.41	
26.72	25.61	24.92	26.56	25.03	24.91	
28.84	26.93	26.51	26.53	25.75	25.85	
30.29	28.72	27.48	26.95	26.64	26.79	

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275
75/91

75

A D
A I
Làm việc với numpy - Bài Tập 2-1

1) Đọc dữ liệu lưu trữ trong file Temp.txt vào biến mảng **data_temp**, cho biết kích thước, số chiều, kiểu dữ liệu và số phần tử của biến data_temp.

```

[[25.65 24.79 24.01 25.06 25.48 24.97]
 [25.31 24.21 24.02 24.93 25.16 24.83]
 [25.05 23.73 23.89 24.79 24.8 24.55]
 ...
 [24.81 24.47 23.4 25.86 25.05 25.29]
 [23.97 24.22 22.95 25.74 24.92 24.87]
 [22.84 23.99 22.59 25.5 24.77 24.57]]
-----
Kích thước biến: (192, 6)
Số chiều của biến: 2
Kiểu dữ liệu của các phần tử: float64
Số phần tử: 1152

```

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275
76/91

76

A
D
A
I
Làm việc với numpy - Bài Tập 2-1

2) Tìm nhiệt độ cao nhất (Max) – Thấp nhất (Min) – Nhiệt độ trung bình của cả tập dữ liệu.

3) Tìm nhiệt độ cao nhất (Max) – Thấp nhất (Min) – Nhiệt độ trung bình của từng thành phố và hiển thị kết quả.

```

---THÔNG KÊ CHO CẢ 6 THÀNH PHỐ---
Nhiệt độ cao nhất: 33.45
Nhiệt độ thấp nhất: 20.93
Nhiệt độ trung bình: 26.50222222222222
-----
1) Hà Nội
Nhiệt độ cao nhất: 33.45
Nhiệt độ thấp nhất: 21.68
Nhiệt độ trung bình: 27.712291666666667
2) Vinh (Nghệ An)
Nhiệt độ cao nhất: 32.57
Nhiệt độ thấp nhất: 22.6
Nhiệt độ trung bình: 26.719895833333333
3) Đà Nẵng
Nhiệt độ cao nhất: 29.88
Nhiệt độ thấp nhất: 20.93
Nhiệt độ trung bình: 25.522499999999997
4) Nha Trang
Nhiệt độ cao nhất: 28.68
Nhiệt độ thấp nhất: 24.5
Nhiệt độ trung bình: 26.166875000000005
5) TP Hồ Chí Minh
Nhiệt độ cao nhất: 31.06
Nhiệt độ thấp nhất: 23.22
Nhiệt độ trung bình: 26.159218749999997
6) Cà Mau
Nhiệt độ cao nhất: 31.37
Nhiệt độ thấp nhất: 23.99
Nhiệt độ trung bình: 26.732552083333333

```

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275
77/91

77

A
D
A
I
Làm việc với numpy - Bài Tập 2-1

4) Tạo một ma trận **data_thongke** gồm 3 hàng x 7 cột; các hàng lần lượt lưu trữ dữ liệu như sau:

- Hàng 0: Nhiệt độ cao nhất (Max)
- Hàng 1: Nhiệt độ trung bình (Mean), làm tròn đến 2 số sau dấu phẩy
- Hàng 2: Nhiệt độ nhỏ nhất (Min)

Các cột lần lượt theo thứ tự của 6 thành phố và cột cuối cùng là cột thống kê chung cho cả 6 thành phố. Lưu ra file **thongke.txt**

```

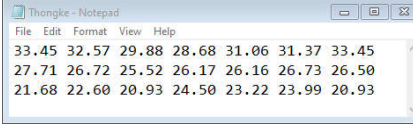
1 print(data_thongke)
2 print(type(data_thongke))
3 print('Kích thước:', data_thongke.shape)

```

```

[[33.45 32.57 29.88 28.68 31.06 31.37 33.45]
 [27.71 26.72 25.52 26.17 26.16 26.73 26.5 ]
 [21.68 22.6  20.93 24.5  23.22 23.99 20.93]]
<class 'numpy.ndarray'>
Kích thước: (3, 7)

```



Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275
78/91

78

A D
A I

BÀI TẬP 2-2

Các file dữ liệu cho phần bài tập: Temp.txt | Diamonds.txt học viên download trong phần tài nguyên của bài học ([Data Exercise](#)).

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275
79/91

79

A D
A I

Làm việc với numpy - Bài Tập 2-2

File dữ liệu: Diamonds.txt

Mô tả file dữ liệu:

- File dữ liệu lưu trữ thông số 50 viên kim cương bao gồm: Trọng lượng (carat) và Giá bán (\$) tương ứng

Dr. Tạ Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275
80/91

80

A D
A I

Làm việc với numpy - Bài Tập 2-2

Học viên thực hiện các yêu cầu sau:

1) Đọc dữ liệu lưu trữ trong file Diamonds.txt vào biến kiểu mảng **data_diamond**, cho biết kích thước, số chiều, kiểu dữ liệu và số phần tử của biến data_diamond

```
[ 3.4000e-01  7.6500e+02]
[ 4.1000e-01  8.2700e+02]
[ 7.5000e-01  3.1200e+03]
[ 1.0700e+00  5.2200e+03]
[ 1.3400e+00  7.4270e+03]
[ 1.7500e+00  9.8900e+03]]
-----
Kích thước biến data_diamond: (50, 2)
Số chiều của biến data_diamond: 2
Kiểu dữ liệu của các phần tử: float64
Số phần tử: 100
```

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

81/91

81

A D
A I

Làm việc với numpy - Bài Tập 2-2

2) Tách mảng data_diamond thành 2 vector: diamond_size và diamond_price lưu trữ trọng lượng và giá bán.

```
Vector diamond_size:
[0.23 0.31 0.2  1.02 1.63 1.14 2.01 1.28 1.7  1.01 0.64 0.97 1.78 3.4
 3.01 1.51 1.37 1.5  0.54 0.72 1.13 2.24 3.01 4.5  0.92 1.05 0.55 0.74
 0.91 1.23 1.52 0.91 0.43 1.24 1.77 1.79 2.05 2.03 2.01 1.  0.9  1.01
 1.14 1.53 0.34 0.41 0.75 1.07 1.34 1.75]

Vector diamond_price:
[ 484.  942.  345.  4459. 14022. 4212. 11925. 9548. 11605. 4642.
 3541. 4504. 13691. 15964. 10453. 11560. 7979. 9533. 1723. 3344.
 6133. 13827. 16538. 18531. 3625. 7879. 1319. 2761. 3620. 6165.
10640. 4138. 1094. 11130. 14561. 10108. 12654. 16280. 13498. 4586.
 3105. 7745. 5047. 10830. 765. 827. 3120. 5220. 7427. 9890.]
```

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

82/91

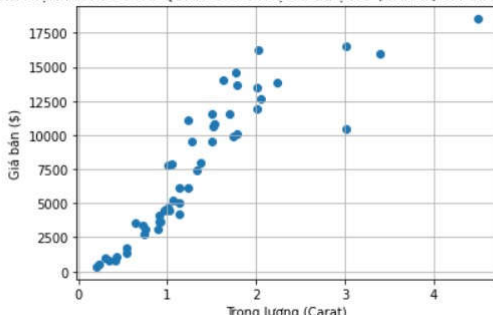
82

A D A I

Làm việc với numpy - Bài Tập 2-2

3) Vẽ đồ thị thể hiện mối quan hệ giữa kích thước và giá bán kim cương. Xác định hệ số tương quan tương ứng giữa 2 thông số này.

BIỂU ĐỒ THỂ HIỆN MỐI TƯƠNG QUAN GIỮA TRỌNG LƯỢNG (CARAT) VÀ GIÁ BÁN KIM CƯƠNG (\$)



Hệ số tương quan giữa trọng lượng và giá bán kim cương: 0.8814849023922127

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

83/91

83

A D A I

Làm việc với numpy - Bài Tập 2-2

4) Cho biết kích thước và giá trung bình của 50 viên kim cương. Hiện thị giá bán của viên kim cương có trọng lượng 3.01 carat.

Trọng lượng trung bình: 1.3448
 Giá trung bình: 7550.78

Viên kim cương trọng lượng 3.01 carat có giá bán:
 Giá bán 1 : 10453.0
 Giá bán 2 : 16538.0

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

84/91

84



Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

85/91