

# DM Lab2-Data (Student)

วัตถุประสงค์

1. import pandas [https://en.wikipedia.org/wiki/Pandas\\_\(software\)](https://en.wikipedia.org/wiki/Pandas_(software))  
([https://en.wikipedia.org/wiki/Pandas\\_\(software\)](https://en.wikipedia.org/wiki/Pandas_(software))), <https://pandas.pydata.org/>  
(<https://pandas.pydata.org/>)
2. การสร้าง data frame (df) จากข้อมูลที่เรารวบรวมไว้ iris.csv
3. การวัดแนวโน้มเข้าสู่ศูนย์กลาง Mean Mode Median
4. Data type (continuous, series)
5. Bar chart จาก magic function %matplotlib inline <https://matplotlib.org/users/index.html>  
(<https://matplotlib.org/users/index.html>)

1. จง import pandas

In [13]: `import pandas as pd`

2. จงสร้าง data frame เปิดไฟล์ iris.csv

In [14]: `df=pd.read_csv("iris.csv")`  
`df`

Out[14]:

	Spellength	Sepalwidth	Petallength	Petalwidth	Class
0	51	35	14	2	1
1	49	30	14	2	1
2	47	32	13	2	1
3	46	31	15	2	1
4	50	36	14	2	1
...	...	...	...	...	...
145	67	30	52	23	3
146	63	25	50	19	3
147	65	30	52	20	3
148	62	34	54	23	3
149	59	30	51	18	3

150 rows × 5 columns

3. จงแสดงตัวอย่างข้อมูลแบบสุ่มมา 5 records

In [15]: `df.sample(5)`

Out[15]:

	Spellength	Sepalwidth	Petallength	Petalwidth	Class
148	62	34	54	23	3
139	69	31	54	21	3
110	65	32	51	20	3
37	49	31	15	1	1
50	70	32	47	14	2

4. จงเรียกดูข้อมูล information

In [16]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Spellength      150 non-null    int64
1   Sepalwidth      150 non-null    int64
2   Petallength     150 non-null    int64
3   Petalwidth      150 non-null    int64
4   Class           150 non-null    int64
dtypes: int64(5)
memory usage: 6.0 KB
```

5. จงหาค่า Mean ของฟิลด์ Petallength ได้

In [17]: `df.Petallength.mean()`

Out[17]: 37.586666666666666

6. จงหาค่า Mode ของฟิลด์ Sepalwidth ได้

In [19]: `df.Sepalwidth.mode()`

Out[19]: 0 30  
dtype: int64

7. จงหาค่า Median ของฟิลด์ Petalwidth ได้

In [20]: `df.Petalwidth.median()`

Out[20]: 13.0

8. จงหา สถิติเชิงพรรณนาอย่างง่ายได้ (descriptive statistics)

In [21]: `df.describe()`

Out[21]:

	Spelength	Sepalwidth	Petallength	Petalwidth	Class
<b>count</b>	150.000000	150.000000	150.000000	150.000000	150.000000
<b>mean</b>	58.433333	30.540000	37.586667	11.986667	2.000000
<b>std</b>	8.280661	4.335943	17.644204	7.631607	0.819232
<b>min</b>	43.000000	20.000000	10.000000	1.000000	1.000000
<b>25%</b>	51.000000	28.000000	16.000000	3.000000	1.000000
<b>50%</b>	58.000000	30.000000	43.500000	13.000000	2.000000
<b>75%</b>	64.000000	33.000000	51.000000	18.000000	3.000000
<b>max</b>	79.000000	44.000000	69.000000	25.000000	3.000000

9. จงใช้ Transpose ในการกลับแถวเป็นหลักกับ `.describe()`

In [22]: `df.describe().T`

Out[22]:

	count	mean	std	min	25%	50%	75%	max
<b>Spelength</b>	150.0	58.433333	8.280661	43.0	51.0	58.0	64.0	79.0
<b>Sepalwidth</b>	150.0	30.540000	4.335943	20.0	28.0	30.0	33.0	44.0
<b>Petallength</b>	150.0	37.586667	17.644204	10.0	16.0	43.5	51.0	69.0
<b>Petalwidth</b>	150.0	11.986667	7.631607	1.0	3.0	13.0	18.0	25.0
<b>Class</b>	150.0	2.000000	0.819232	1.0	1.0	2.0	3.0	3.0

10. หาขนาดของ data frame ของ iris data set % ผลลัพธ์ (150\*5)

In [23]: `df.shape`

Out[23]: (150, 5)

11. จงนับจำนวนในแต่ละ class ว่าแต่ละ class (ทั้ง 1,2,3) มีจำนวนเท่าใด

In [25]: `df['Class'].value_counts()`

Out[25]:

```
3    50
2    50
1    50
Name: Class, dtype: int64
```

13. ลองเลือกมาแสดงแค่ 2 column ของ Petal (กลีบดอก)

```
In [26]: df[['Petallength', 'Petalwidth']].head()
```

```
Out[26]:
```

	Petallength	Petalwidth
0	14	2
1	14	2
2	13	2
3	15	2
4	14	2

14. จงสุ่มข้อมูล iris data set มาสัก 10 records

```
In [27]: df[['Petallength', 'Petalwidth']].sample(n=10)
```

```
Out[27]:
```

	Petallength	Petalwidth
127	49	18
114	51	24
145	52	23
61	42	15
84	45	15
6	14	3
53	40	13
44	19	4
96	42	13
80	38	11

15. จงสร้างข้อมูลแบบเป็น series (ตย. ข้อมูลทองคำ ข้อมูลตลาดหุ้น ข้อมูลการขาย) ข้อมูลของนักศึกษาแต่ละคนห้ามซ้ำกัน ถ้าอ.ตรวจสอบแล้วใครซ้ำ คะแนนเต็มหาร 2 (array มีขนาดตั้งแต่ 5 ตัวแปรขึ้นไป)

```
In [28]: data=pd.Series([15,52,66,48,92,107,156])
data
```

```
Out[28]: 0    15
1    52
2    66
3    48
4    92
5   107
6   156
dtype: int64
```

16. เลือกแถวที่ 2 กับ 4 มาแสดง

```
In [29]: data[2]
```

```
Out[29]: 66
```

```
In [30]: data[4]
```

```
Out[30]: 92
```

17. หาผลรวม sum

```
In [31]: data.sum()
```

```
Out[31]: 536
```

16. จงใส่ index ตามตัวอย่าง อาทิเช่น ทองคำ ในเดือน มกราคม กุมภาพันธ์ มีนาคม เป็นต้น

```
In [42]: mydata = pd.Series([15,52,66,48,92,107,156],index=[ 'Jan', 'Feb', 'March', 'Apr'  
mydata
```

```
Out[42]: Jan      15  
Feb      52  
March    66  
Apr      48  
May      92  
Jun     107  
Aug     156  
dtype: int64
```

17. จงเพิ่ม matplotlib (magic function) แล้วสั่ง plot() ตามตัวอย่างที่สอนในชั้นเรียน  
ไปอ่าน user guide เพิ่มเติมนะคะ <https://matplotlib.org/users/index.html>

```
In [36]: import matplotlib.pyplot as plt
```

```
In [37]: %matplotlib inline
```

18. จงสร้าง bar chart ของข้อมูลที่นักศึกษาสร้างขึ้นเอง ตามตัวอย่างที่สอนในชั้นเรียน

```
In [41]: mydata.plot(kind = 'bar')
```

```
Out[41]: <AxesSubplot:>
```

