

## Quiz2



- โปรดดูรายการวิดีโอเพื่อเตรียมตัวก่อนทำข้อสอบ และตรวจคำตอบ
- [https://data.go.th/dataset/item\\_7599910F-03d8-4C75-9d94-112ca38d36d2](https://data.go.th/dataset/item_7599910F-03d8-4C75-9d94-112ca38d36d2)
- ไม่เปิดศึกษา คะแนน ขั้วข้อมูลมีดังนี้ วิทยาศาสตร์ 100 คะแนน คณิต 80 คะแนน

## การแสดงผลการกระจายของข้อมูล Scatter Plot

- `import pandas as pd`
- `from matplotlib import pyplot as plt`
- `df = pd.read_csv("https://raw.githubusercontent.com/pandas-der/pandas/master/pandas/tests/10/data/csv/iris.csv")`
- `df.head()`
- ถ้าใช้ google colab ให้ใช้ `Visualize Data Distribution (UI)`
- สามารถดูข้อมูลแยกเป็น iris

2


ดอกไม้ iris

The image displays three distinct iris flower varieties side-by-side. Each variety is shown in a separate photograph with labels for its petals and sepals. The labels are as follows:

- Iris setosa:** The petals are labeled 'petal' and the sepals are labeled 'sepal'.
- Iris versicolor:** The petals are labeled 'petal' and the sepals are labeled 'sepal'.
- Iris virginica:** The petals are labeled 'petal' and the sepals are labeled 'sepal'.

- ศึกษาข้อมูลก่อนการแปลงค่าเป็น 1 ชุด แบ่งข้อมูลออกเป็น 2 ชุด เพื่อการกระจายค่าข้อมูลระหว่าง 2 normal distribution (ปกติ)
- ใ้ข้อมูลนี้กับ plot.scatter(df['petalLength'], df['sepalLength']) เช่น

```
plt.scatter(df['petalLength'], df['sepalLength'])
```




sepalLength	petalLength	sepalName	petalName
4.9	1.4	set	setosa
5.0	1.5	set	setosa
5.1	1.6	set	setosa
5.2	1.7	set	setosa
5.4	1.8	set	setosa
5.5	1.9	set	setosa
5.6	2.0	set	setosa
5.7	2.1	set	setosa
5.8	2.2	set	setosa
5.9	2.3	set	setosa

4

## Parameter: c กำหนดค่าสีให้ทุกจุด

- สามารถกำหนดสีให้กับทุกจุดใน scatter plot ได้
- `plt.scatter(df['Petal.Length'], df['Sepal.Length'], c="r")`

## Parameter: c กำหนดค่าสีตามกลุ่มข้อมูลที่ต้องการ

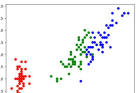


- สามารถกำหนดสีของจุด หรือ เส้นตามสีของแต่ละ class โดยกำหนดให้แต่ละจุดหรือเส้นมีสีตามค่าของกลุ่มตัวแปร หรือสีตามค่าสี class
- ใช้คำสั่ง set(df["Name"]) เพื่อกำหนดค่าให้แต่ละ variable ที่มีค่า
- ["Iris-setosa", "Iris-versicolour", "Iris-virginica"]
- ใช้คำสั่ง replace เปลี่ยนค่าให้ตรงกับสีตามกลุ่มตัวแปร และเปลี่ยนค่าให้ตรงกับสีตามกลุ่มตัวแปร
- df2 = df.replace({"Iris-setosa": 0, "Iris-versicolour": 1, "Iris-virginica":2})
- พล็อต
- df2 = df.replace({"Iris-setosa": "r", "Iris-versicolour": "g", "Iris-virginica": "b"})

### Parameter: c กำหนดค่าสีตามกลุ่มข้อมูลที่ต้องการ

- ใช้กำหนด Parameter: c ตัวนี้ไปเป็นชื่อ Name ของข้อมูลที่เราใช้กับคำสั่งนี้

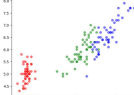
```
plt.scatter(d2["petalLength"], d2["SepalLength"], c=d2["Name"])
```



The scatter plot displays the relationship between 'petalLength' (x-axis, ranging from 1 to 7) and 'SepalLength' (y-axis, ranging from 4.0 to 8.0). The data points are colored based on the 'Name' parameter, showing three distinct groups: 'setosa' (red), 'versicolour' (green), and 'virginica' (blue). The 'setosa' group is clustered at low values for both variables. The 'versicolour' and 'virginica' groups overlap significantly at higher values, but 'virginica' tends to have higher 'SepalLength' values for a given 'petalLength' compared to 'versicolour'.

## alpha กำหนดความโปร่งแสง

- ค่าความโปร่งแสงมีค่า 0 ถึง 1 โดย 1 คือไม่โปร่งแสง เต็มที่โดยจะเห็นฟองซ้อน (Bubble Chart)
- `plt.scatter(df2['PetalLength'],df2['SepalLength'],c=df2['Name'],alpha=0.5)`



8

1