Matplotlib topic

Instructions

- Create new PyCharm project
- Delete unwanted code in main.py
- Go to page 272 / 321
- Go to terminal
  - pip install numpy, pandas, matplotlib
- Go back to main.py
- Import numpy, pandas and pyplot
- Create case06() function in you main.py
- Copy the source code in page 272 to your case06()
- Add plt.show()
- Make sure that main.py call case06()
- Try changing "-10" to "-20"
- Check the content in Series
  - print(x) and print(y)
- Continue 11.1 → Change color of the graph
- About color
  - Goto "color.adobe.com"
  - Try pickup 6-digit color and try it in your code (RGB)
  - #XXXXXX → XX XX XX → x can be 0 1 2 3 ... 9 A B C D E F

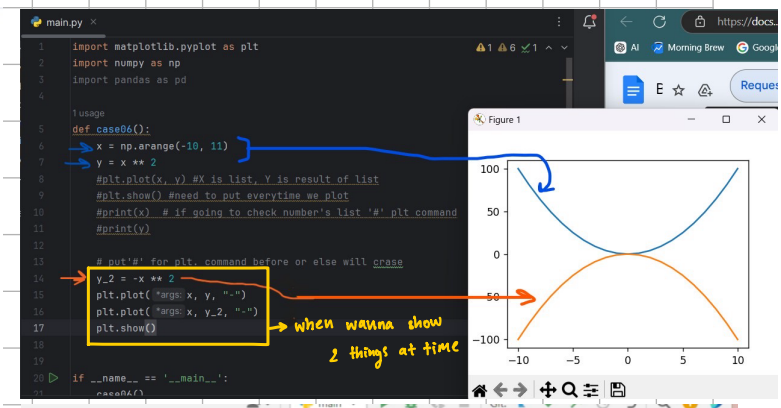
Until 11.1.1.2 (Break 10:15 - 10:30)

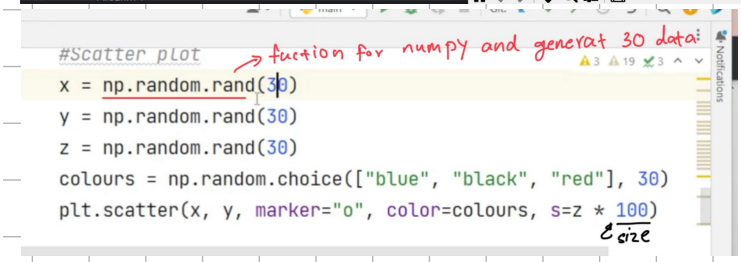
- Go to page 285 / 321
- Create case07()
- Put the source code in case07()
- Create case08()
- Try Histogram → the frequency of data occur in event
- Create case09()
- Try Barchart
- Create case10()
- Try Stacked bar chart
- Go to 11.1.3.3 page (314 / 321)
- Try Legends
- Save file
  - You need to have your dedicated location on your computer.
  - In my case, the position is "C:\data\output"
  - In mac, it is similar to /Users/<user name>/Desktop/<user name>
  - Do not forget to double backslash (Windows only)

meaning

detail / eg

can change





eg

- Data of student score

so Histogram will show which score apear most

- How many ppl walk in to shopping mall

Bins = arrage Grouping numbers that is in same categorise and set a no. that u want to categorize.

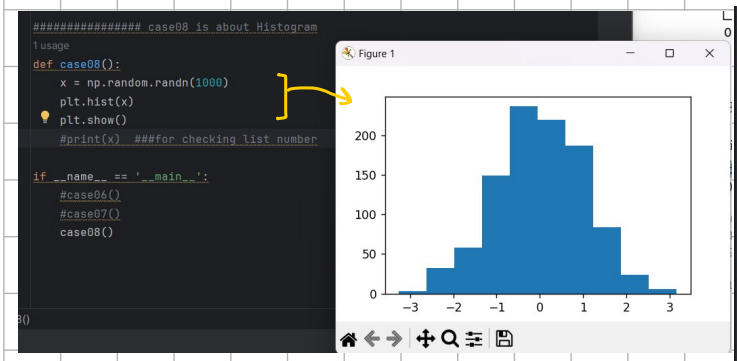
eg • sales in 100 shops Bins in 100,000 = (1)

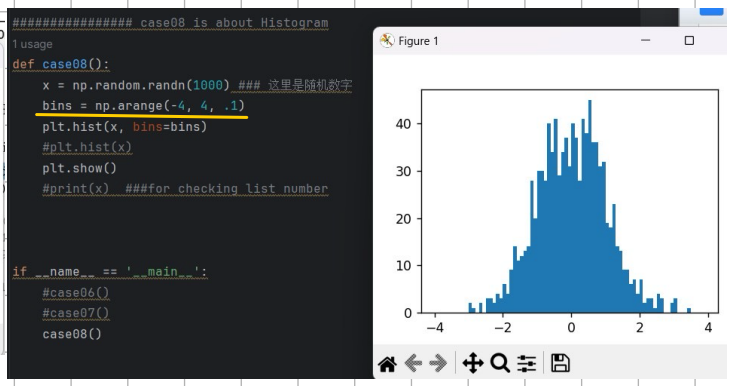
0 – 100,000 , 100,000 – 200,000 etc...

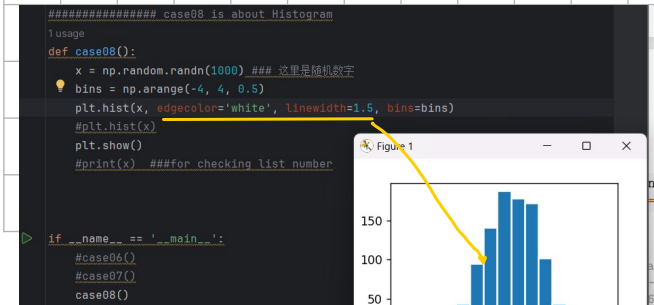
• (-4, 4, 0.1) → Bins

-4 – -3.5 , -3.5 – -3 etc...

categorie 1







n(1000)
= "#00FF00", edgecolor='black', linewidth=1.5)

to make it look better

2., 32., 79., 223., 272., 229., 109.,
array([-3.88089841, -3.18024506, -2.4795917 ,
-1.07828499,
63163, 0.32302172, 1.02367508, 1.72432844,

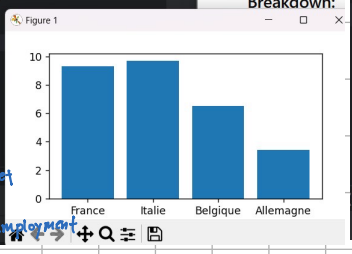
```
################ about Bar chat
1 usage
def case09():
    pays = ["France", "Italie", "Belgique", "Allemagne"]
    unemployment = [9.3, 9.7, 6.5, 3.4]
    plt.bar(pays, unemployment)
    plt.show()


if __name__ == '__main__':
    #case06()
    #case07()
    #case08()
    case09()
```

"Countries" in French

can also use "countries"

can change according to the data set I'm using & the categories I want to represent

eg categories = Pays , Represent = unemployment

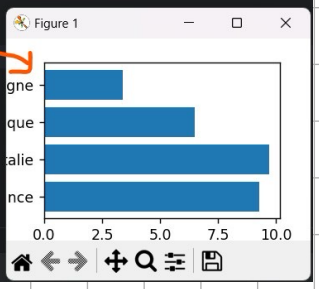


list of country na

"countries."

Breakdown:

```
def case09():
    countries = ["France", "Italie", "Belgique", "Allemagne"]
    unemployment = [9.3, 9.7, 6.5, 3.4]
    #plt.bar(countries, unemployment)
    plt.barh(countries, unemployment)
    plt.show()


if __name__ == '__main__':
    #case06()
    #case07()
    #case08()
```




```
############## Below is for Serveral data/series
    countries = ["France", "Italie", "Belgique", "Allemagne"]
    unemp_f = [9.1, 11.2, 6.6, 2.9]
    unemp_h = [9.5, 9, 6.6, 3.8]
    # Position on the x-axis for each label
    position = np.arange(len(countries))

    # Bar widths
    width = .35
    # Creating the figure and a set of subgraphics
    fig, ax = plt.subplots()
    r1 = ax.bar(position - width / 2, unemp_f, width)
    r2 = ax.bar(position + width / 2, unemp_h, width)
    # Modification of the marks on the x-axis and their labels
    ax.set_xticks(position)
    ax.set_xticklabels(countries)
plt.show()
```

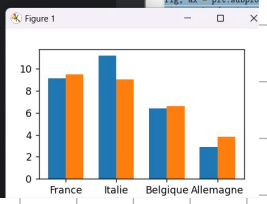
unemployment rate of Males & Females

value of...

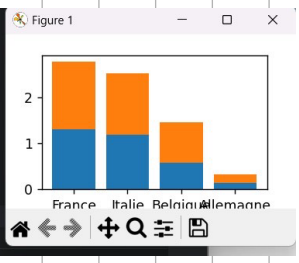
- = Left side

+ = Right side

r1 = Bar for 1st unemployment data set

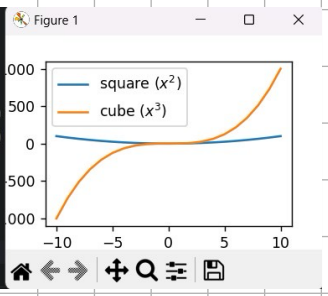



```
### about Stacked bar chart
1 usage
def case10():
    countries = ["France", "Italie", "Belgique", "Allemagne"]
    no_unemp_f = [1.307, 1.185, .577, .148]
    no_unemp_h = [1.46, 1.338, .878, .179]
    plt.bar(countries, no_unemp_f)
    plt.bar(countries, no_unemp_h, bottom=no_unemp_f)

    plt.show()
```

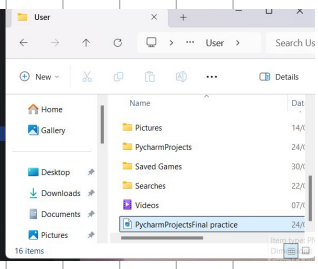


To save pictures of each result

```
### Below is Legends
    x = np.arange(-10, 11)
    y = x ** 2
    y_2 = x ** 3
    plt.plot( *args: x, y, label="square ($x^2$)")
    plt.plot( *args: x, y_2, label="cube ($x^3$)")
    plt.legend()
    plt.legend()

    plt.show()

if __name__ == '__main__':
```




```
### Below is Legends
    x = np.arange(-10, 11)
    y = x ** 2
    y_2 = x ** 3
    plt.plot( *args: x, y, label="square ($x^2$)")
    plt.plot( *args: x, y_2, label="cube ($x^3$)")
    plt.legend()
    plt.legend()
    path = 'C:\\Users\\User\\PycharmProjects' + 'Final practice.png'
    plt.savefig(path)

if __name__ == '__main__':
    #case06()
    #case07()
    #case08()
    #case09()
    case10()
```



x ** 2 = use square = 2

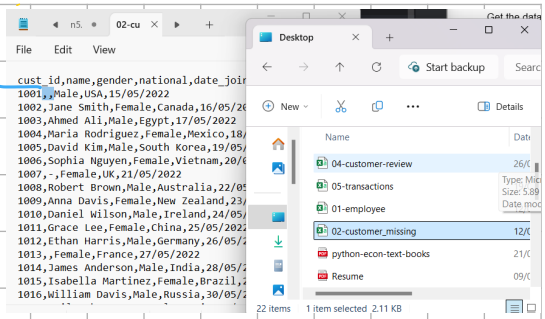
x ** 3 = use cube = 3

Json

```python
import numpy as np
import pandas as pd


1 usage
def case11():
    path = 'C:\\Users\\User\\OneDrive\\เดสก์ท็อป\\02-customer_missing.csv'
    df = pd.read_csv(path)
    print(df)  → checking data (will show after "run")



if __name__ == '__main__':
    case11()
```

or

← Check here →




```
"C:\Users\User\PycharmProjects\Final practice2\.ven
   cust_id          name  gender      national
0     1001  missing data   NaN     Male      USA
            Not A Number
```

Missing data { NAN

Data Cleansing

2 Ways

W1

```python
    ### Data cleancing
        df2 = df.replace( to_replace: '-', value: 'No data')
        print(df2)
```

Replace `-` with word `No data`

But NaN and other types of missing data still there

```
PycharmProjects\Final practice2\.venv\Scripts\python.exe" "
        name  gender      national   date_join
         NaN    Male           USA  15/05/2022
```

```python
### Data cleancing
    df2 = df.replace( to_replace: '-', value: 'No data')
    #print(df2)

    df3 = df2.fillna('No data')
    print(df3)
```

↳ still not a number

First roll is "0"

```
rmProjects\Final practice2\.venv\Scripts\python.exe" "
        name  gender      national   date_join
     No data    Male           USA  15/05/2022
Jane Smith  Female        Canada    16/05/2022
```

```python
### Data cleancing
    df2 = df.replace( to_replace: '-', value: 'No data')
    #print(df2)
    df3 = df2.fillna('No data')
    df4 = df3.replace( to_replace: '--', value: 'No data')
    df5 = df4.replace( to_replace: ' ', value: 'No data')
    print(df5)
```

↳ Continue replace missing data to

```
rmProjects\Final practice2\.venv\Scripts\python.exe" "C:\
     name  gender      national   date_join
  No data    Male           USA  15/05/2022
```

Replace comand is cleaning "all" but can't specifit collum

if want to replace many data → python dicunary { }

W2

```python
### For cleaning "Name" colum ONLY (in specific colum)
1 usage
def case12():
    path = 'C:\\Users\\User\\OneDrive\\เดสก์ท็อป\\02-customer_missing.csv'
    df = pd.read_csv(path)
    df['name'] = df['name'].fillna('No data')
    df['name'] = df['name'].replace({
        '-': 'No data',
        '--': 'No data',
        ' ':'No data'})
    print(df)
```

↳ dicunary

Replace missing data

";" when continue putting data

} Chang for 1 colum

```python
import numpy as np
import pandas as pd
from datetime import date


1 usage
def case13():
    print(calculateAge(date( year: 1997,  month: 2,  day: 3)), "years")
1 usage
def calculateAge(birthDate):
    days_in_year = 365.2425
    age = int((date.today() - birthDate).days / days_in_year)
    return age

case13()
```

give birth date → คิดอายุ ใช้

def ____() :

Get new source code creat new function & paste

Don't put /merger in ur own's

```
"C:\Users\User\PycharmProjects\Final practice2\.venv\Scripts\python.exe" "C:\Users\User\Py
27 years
```

```
def case13():
    print(calculateAge(date( year: 1997,  month: 2,  day: 3)), "years")
    text = '15/05/2022'
```
→ must change to 年,月,日

02-cu

File Edit View

เอาจาก

```
cust_id,name,gender,national,date_join
1001,,Male,USA,15/05/2022
```

For calculate the age in year based on their birthdate.

```
1 usage
def case13():
    print(calculateAge(date( year: 1997,  month: 2,  day: 3)), "years")
    text = '2022/05/15'
    y = text[0:4]
    m = text[5:7]
    d = text[8:10]
    print(y)
    print(m)
    print(d)
case13()
```

get element from 0 to 4 (เริ่มนับจากตัวหน้าไปไว้)

For trying the code only

main

```
"C:\Users\User\PycharmProjects\Final practice2\.venv\Scripts\python.exe" "C
27 years
2022
05
15
```

Finding the work duration of the employee

```
### you want to know the work duration of the employee (one by one)

    path = 'C:\\Users\\User\\OneDrive\\เดสก์ท็อป\\02-customer_missing.csv'
    df = pd.read_csv(path)
    print(df)
    # below is runningg info 1 by 1 (run into 1 by 1)
for x in df.index:
    jdate = df.at[x, 'date_join']
```

join date

Pick up 'date join' to jdate 1 by 1

data transformation

```
    jdate = df.at[x, 'date_join']
    y = int(jdate[0:4])
    m = int(jdate[5:7])
    d = int(jdate[8:10])
    duration_value = calculateAge(date(y, m, d))
    df.at[x, 'duration'] = duration_value
```

(put value back to data from)

```
print(df)
```

joined at year 2

124	Daniel Davis	Male	Mexico	2022-06-08	2.0
125	Grace Johnson	Female	South Korea	2022-06-09	2.0
126	Ethan Garcia	Male	France	2022-06-10	2.0
127	Ava Taylor	Female	Germany	2022-06-11	2.0
128		Male	India	2022-06-12	2.0