

B4 start running code

```
1 import numpy as np
2 import pandas as pd
3
4 usage
5 def case11():
6     path = 'C:\\Users\\User\\OneDrive\\main\\02-customer_missing.csv'
7     df = pd.read_csv(path)
8     print(df)
9
10 if __name__ == '__main__':
11     case11()
12
13 case11()
14
15 main x
16
17 :
18
19 "C:\\Users\\User\\PycharmProjects\\Final practice2\\.venv\\Scripts\\python.exe" "C:\\Users\\User\\PycharmProjects\\Final practice2\\.venv\\Scripts\\python.exe"
20
21 cust_id    name  gender  national  date_join
22 0      1001    NaN    Male      USA      15/05/2022
23 1      1002  Jane Smith  Female    Canada  16/05/2022
```

Data Cleaning

```
def case11():
    path = 'C:\\Users\\User\\OneDrive\\main\\02-customer_missing.csv'
    df = pd.read_csv(path)
    #print(df)

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

    ## 一下也是data cleaning 但用于(edit "name" column only)
    usage
    def case12():
        path = 'C:\\Users\\User\\OneDrive\\main\\02-customer_missing.csv'
        df = pd.read_csv(path)

        # need to put exactly colum in []
        df['name'] = df['name'].fillna('No data')
        df['name'] = df['name'].replace({
            '-': 'No data',
            '--': 'No data',
            ' ': 'No data'
        })
        print(df)
```

save way
clean 1 by 1 valuable

kidda danger (Must select the right "column name")
clean by "1 column"

same result

```
main x
:
"C:\\Users\\User\\PycharmProjects\\Final practice2\\.venv\\Scripts\\python.exe" "C:\\Users\\User\\PycharmProjects\\Final practice2\\.venv\\Scripts\\python.exe"
cust_id    name  gender  national  date_join
0      1001    No data  Male      USA      15/05/2022
1      1002  Jane Smith  Female    Canada  16/05/2022
```

Data Transformation

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

```
29 usage
30 def case13():
31     print(calculateAge(date(year=1997, month=2, day=3), "years"))
32     text = '2022-05-15' #can't 15/05/2022
33     y = text[0:4]
34     m = text[5:7]
35     d = text[8:10]
36     print(y)
37     print(m)
38     print(d)
39
40 usage
41 def calculateAge(birthDate):
42     days_in_year = 365.2425
43     age = int((date.today() - birthDate).days / days_in_year)
44     return age
45
46 calculateAge()
```

can cut out
after finish "date"
case 14

Can calculate Age / duration

if get code from online sperat it
will ux own

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

```
usage
def case14():
    path = 'C:\\Users\\User\\OneDrive\\เดสก์ท็อป\\02-customer_...'
    df = pd.read_csv(path)

    for x in df.index:
        jdate = df.at[x, 'date_join']
        y = text[0:4]
        m = text[5:7]
        d = text[8:10]

    for x in df.index:
        ...

ain x

ers\\User\\PycharmProjects\\Final practice2\\.venv\\Scripts\\pytho
st_id      name  gender  national  date_join
1001      NaN    Male     USA      15/05/2022
1002    Jane Smith  Female  Canada  16/05/2022
```

→ copy from
case 13

