Module 02

Pandas: Operations

Data Science Developer

Purwadhika
Startup and Coding School

Outline

- Columns Operation using Function
- Columns Operation using Lambda
- Math Operation
- Pivot Table

Purwadhika
Startup and Coding School

Create DataFrame

In [1]:
```
import pandas as pd
df = pd.DataFrame({'col1':[1,2,3,4],'col2':[444,555,666,444],'col3':['abc','def','ghi','xyz']})
df.head()
```

Out[1]:

	col1	col2	col3
0	1	444	abc
1	2	555	def
2	3	666	ghi
3	4	444	xyz

Purwadhika
Startup and Coding School

Columns Operation using Function

Applying Functions

```
In [7]: def times2(x):
            return x*2
```

```
In [8]: df['col1'].apply(times2)
```

```
Out[8]: 0    2
        1    4
        2    6
        3    8
        Name: col1, dtype: int64
```

```
In [9]: df['col3'].apply(len)
```

```
Out[9]: 0    3
        1    3
        2    3
        3    3
        Name: col3, dtype: int64
```

```
In [4]: len(df['col3'])
```

```
Out[4]: 4
```

Purwadhika
Startup and Coding School

Columns Operation using Lambda

Applying Lambda Functions

```
In [11]: df['col1'].apply(lambda x: x*2)

Out[11]: 0    2
         1    4
         2    6
         3    8
         Name: col1, dtype: int64

In [12]: df['col3'].apply(lambda x: x[1])

Out[12]: 0    b
         1    e
         2    h
         3    y
         Name: col3, dtype: object

In [13]: df['col3'].apply(lambda x: len(x))

Out[13]: 0    3
         1    3
         2    3
         3    3
         Name: col3, dtype: int64
```

Purwadhika
Startup and Coding School

Math Operation

Purwadhika
Startup and Coding School

Math Operation

```
df['col4'] = df['col1'] + df['col2']
```

```
df
```

	col1	col2	col3	col4
0	1	444	abc	445
1	2	555	def	557
2	3	666	ghi	669
3	4	444	xyz	448

```
df['col5'] = df['col1'] - df['col2']
```

```
df
```

	col1	col2	col3	col4	col5
0	1	444	abc	445	-443
1	2	555	def	557	-553
2	3	666	ghi	669	-663
3	4	444	xyz	448	-440

```
df['col6'] = df['col2'] / df['col1']
```

```
df
```

	col1	col2	col3	col4	col5	col6
0	1	444	abc	445	-443	444.0
1	2	555	def	557	-553	277.5
2	3	666	ghi	669	-663	222.0
3	4	444	xyz	448	-440	111.0

```
df['col7'] = df['col1'] * df['col1']
```

```
df
```

	col1	col2	col3	col4	col5	col6	col7
0	1	444	abc	445	-443	444.0	1
1	2	555	def	557	-553	277.5	4
2	3	666	ghi	669	-663	222.0	9
3	4	444	xyz	448	-440	111.0	16

Purwadhika
Startup and Coding School

Pivot Table

Purwadhika
Startup and Coding School

Pivot Table

In [22]:
```
data = {'A':['foo','foo','foo','bar','bar','bar'],
     'B':['one','one','two','two','one','one'],
       'C':['x','y','x','y','x','y'],
       'D':[1,3,2,5,4,1]}

df = pd.DataFrame(data)
```

In [23]:
```
df
```

Out[23]:

	A	B	C	D
0	foo	one	x	1
1	foo	one	y	3
2	foo	two	x	2
3	bar	two	y	5
4	bar	one	x	4
5	bar	one	y	1

In [24]:
```
df.pivot_table(values='D',index=['A', 'B'],columns=['C'])
```

Out[24]:

	C	x	y
A	**B**		
bar	**one**	4.0	1.0
	two	NaN	5.0
foo	**one**	1.0	3.0
	two	2.0	NaN

Purwadhika
Startup and Coding School

Pivot Table

```
In [1]: df
Out[1]:
         date variable     value
0  2000-01-03        A  0.469112
1  2000-01-04        A -0.282863
2  2000-01-05        A -1.509059
3  2000-01-03        B -1.135632
4  2000-01-04        B  1.212112
5  2000-01-05        B -0.173215
6  2000-01-03        C  0.119209
7  2000-01-04        C -1.044236
8  2000-01-05        C -0.861849
9  2000-01-03        D -2.104569
10 2000-01-04        D -0.494929
11 2000-01-05        D  1.071804
```

```
In [3]: df.pivot(index='date', columns='variable', values='value')
Out[3]:
variable           A         B         C         D
date
2000-01-03  0.469112 -1.135632  0.119209 -2.104569
2000-01-04 -0.282863  1.212112 -1.044236 -0.494929
2000-01-05 -1.509059 -0.173215 -0.861849  1.071804
```

Purwadhika
Startup and Coding School

Reference

- When Should I Ever Want to Use Pandas Apply in my Code. https://stackoverflow.com/questions/54432583/when-should-i-ever-want-to-use-pandas-apply-in-my-code
- Reshaping and Pivot Table. https://pandas.pydata.org/pandas-docs/stable/user_guide/reshaping.html

Purwadhika
Startup and Coding School