DATA-MANIPULATION-2

Transforming data to unveil valuable insights

Club Informatique & Télécom
Data Cell

Data Manipulation - 2

- What is data transformation ?
- Applying Functions to Data
- Merging and Joining DataFrames
- Pivot Tables
- Make your own mini assistant

Use function on data

```
data = pd.Series([1, 2, 3, 4, 5])
```

```
>>>
```

0	1
1	2
2	3
4	4
5	5

```python
def square_function(x):
    return x**2
```

```
result = data.apply(square_function)
```

```
>>>
```

0	1
1	4
2	9
4	16
5	25

```
result = data.apply(lambda x : x**2 )
```

Mapping

```python
data = pd.DataFrame({'Category':
                     ['A', 'B', 'A', 'C', 'B','D']
                     })
category_mapping = {'A': 1, 'B': 2, 'C': 3,'D': 4}
```

	category
0	A
1	B
2	C
3	D

Mapping

```python
data['Category_mapped'] = data['Category']\
                        .map(category_mapping)
```

	category	Category_mapped
0	A	1
1	B	2
2	C	3
3	D	4

using applymap

```python
data = pd.DataFrame({'A': [1,2,3,4],
                     'B': [5,6,7,8],
                     'C': [9, 10,11,12]
                    })
```

```python
def add_1(x):
    return x + 1
```

```python
result = data.applymap(add_1)
```

	A	B	C
0	1	5	9
1	2	6	10
2	3	7	11
3	4	8	12

	A	B	C
0	2	6	10
1	3	7	11
2	4	8	12
3	5	9	13

Pivot Table

```python
pivot_table = pd.pivot_table(sales_df,
                             values='Sales',
                             index='Product',
                             columns='Region',
                             aggfunc='sum',
                             fill_value=0)
```

	product	region	sales
0	A	North	100
1	B	South	150
2	A	East	200
3	C	West	120
4	B	North	180

```
                Pivot Table:
Region     East   North   South   West
Product
A          420     100       0      0
B            0     180     150    130
C            0     170      90    120
```

Combining DataFrames using merge

```python
df1 = pd.DataFrame({'ID': [1, 2, 3],
                    'Name': ['Alice', 'Bob', 'Charlie']})
```

	ID	NAME
0	1	ALICE
1	2	BOB
2	3	CHARLIE

```python
df2 = pd.DataFrame({'ID': [2, 3, 4],
                    'Age': [25, 30, 22]})
```

	ID	Age
0	2	25
1	3	30
2	4	22

```python
merged_df = pd.merge(df1, df2, on='ID')
```

	ID	NAME	Age
0	2	ALICE	25
1	3	CHARLIE	30

Combining DataFrames using keys

	ID	NAME
0	1	ALICE
1	2	BOB
2	3	CHARLIE

	ID	Age
0	2	25
1	3	30
2	4	22

```python
joined_df = pd.merge(df1, df2,
                     left_on='ID',
                     right_on='Employee_ID',
                     how='left')
```

	ID	NAME	Employee_id	Age
0	1	ALICE	NaN	NaN
1	2	BoB	2	25
2	3	charlie	3	30

Time to practice !