



머신러닝

1주차 과제

명지대학교

경영정보학과

60211356

장유진

이한준 교수님

문제 1

데이터 프레임을 생성하시오.

```
[9]: import pandas as pd
```

```
[5]: df = pd.DataFrame([[2016, 2.8, '1.637M'],  
                        [2017, 3.1, '1.73M'],  
                        [2018, 3.0, '1.83M']], columns = ['year', 'GDP rate', 'GDP'])
```

```
[6]: df
```

```
[6]:
```

	year	GDP rate	GDP
0	2016	2.8	1.637M
1	2017	3.1	1.73M
2	2018	3.0	1.83M

문제 2

2017년부터의 자료만 출력하시오.

```
[7]: df.loc[df['year'] >= 2017]
```

```
[7]:
```

	year	GDP rate	GDP
1	2017	3.1	1.73M
2	2018	3.0	1.83M

GDP rate의 합과 평균을 구할 것

```
[13]: df['GDP rate'].sum()
```

```
[13]: 8.9
```

```
[14]: df['GDP rate'].mean()
```

```
[14]: 2.9666666666666667
```

문제 3

Kaggle 사이트에서 Iris.csv 파일을 다운로드 받아서 데이터프레임으로 가져오시오.

```
[10]: df = pd.read_csv('data/iris.csv')
```

```
[11]: df
```

```
[11]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

Iris(붓꽃)의 종류(Setosa, Versicolor, virginica)를 평균 꽃잎(Petal) 너비(width)와 길이(length), 꽃받침(Sepal)의 너비와 길이를 센티미터 값으로 바꾸시오. (기존 값은 inch라 가정. 1inch = 2.5cm으로 계산)

```
[20]: df.groupby('Species')[['SepalLengthCm', 'SepalWidthCm']].mean() * 2.5
```

```
[20]:
```

	SepalLengthCm	SepalWidthCm
Species		
Iris-setosa	12.515	8.545
Iris-versicolor	14.840	6.925
Iris-virginica	16.470	7.435

```
[21]: df.groupby('Species')[['PetalLengthCm', 'PetalWidthCm']].mean() * 2.5
```

```
[21]:
```

	PetalLengthCm	PetalWidthCm
Species		
Iris-setosa	3.66	0.610
Iris-versicolor	10.65	3.315
Iris-virginica	13.88	5.065

꽃잎과 꽃받침 각각의 너비와 길이에 대한 평균을 품종별로 구하시오.

```
[12]: condition1 = df['Species'] == 'Iris-setosa'
condition2 = df['Species'] == 'Iris-versicolor'
condition3 = df['Species'] == 'Iris-virginica'
```

```
[17]: df.loc[condition1, 'SepalLengthCm'].mean(), df.loc[condition1, 'SepalWidthCm'].mean(), df.loc[condition1, 'PetalLengthCm'].mean(), df.loc[condition1, 'PetalWidthCm'].mean()
< [ ]
```

```
[17]: (5.006, 3.418, 1.464, 0.244)
```

```
[18]: df.loc[condition2, 'SepalLengthCm'].mean(), df.loc[condition2, 'SepalWidthCm'].mean(), df.loc[condition2, 'PetalLengthCm'].mean(), df.loc[condition2, 'PetalWidthCm'].mean()
< [ ]
```

```
[18]: (5.936, 2.7700000000000005, 4.26, 1.3259999999999998)
```

```
[19]: df.loc[condition3, 'SepalLengthCm'].mean(), df.loc[condition3, 'SepalWidthCm'].mean(), df.loc[condition3, 'PetalLengthCm'].mean(), df.loc[condition3, 'PetalWidthCm'].mean()
< [ ]
```

```
[19]: (6.587999999999998, 2.974, 5.5520000000000005, 2.0260000000000002)
```