

```
In [39]: import pandas as pd

#데이터 스케일링
from sklearn.preprocessing import StandardScaler

#클러스터링
from sklearn.cluster import KMeans

#차원 축소
from sklearn.decomposition import PCA

#데이터 시각화
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [17]: #데이터셋 불러오기
df = pd.read_csv("IRIS.csv")
df.head(4)
```

```
Out[17]:
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa

```
In [40]: #산점도 그리고
```

```
In [41]: #df2 변수를 만들어서 species 지우기
```

```
In [42]: # 데이터 스케일링
```

```
In [43]: #df2 컬럼을 불러서 데이터 프레임 만들기
```

## 클러스터링

```
In [44]: # 그룹 지정할때 선 그래프 그리기
# inertia : 각 데이터 포인트와 그 포인트가 속한 클러스터 중심점 사이 거리의 제곱합
```

```
In [45]: #KMeans로 학습하기
```

```
In [46]: #산점도 그리기
```

## 차원 축소

```
In [47]: #차원 축소 지정
#데이터 변환하기
#DataFrame을 만들고, columns를 pc1,pc2 만들기
```

```
In [48]: #예전 df['species'] 불러와서 새 데이터 컬럼 'species' 붙이기
```

```
In [ ]:
```

```
In [49]: #차원 축소 산점도 그리기
```

```
In [ ]:
```