```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

data = pd.read_csv('Downloads/Universities.csv')
data
```

|	Univ	SAT	Top10	Accept	SFRatio	Expenses	GradRate
0	Brown	1310	89	22	13	22704	94
1	CalTech	1415	100	25	6	63575	81
2	CMU	1260	62	59	9	25026	72
3	Columbia	1310	76	24	12	31510	88
4	Cornell	1280	83	33	13	21864	90
5	Dartmouth	1340	89	23	10	32162	95
6	Duke	1315	90	30	12	31585	95
7	Georgetown	1255	74	24	12	20126	92
8	Harvard	1400	91	14	11	39525	97
9	JohnsHopkins	1305	75	44	7	58691	87
10	MIT	1380	94	30	10	34870	91
11	Northwestern	1260	85	39	11	28052	89
12	NotreDame	1255	81	42	13	15122	94
13	PennState	1081	38	54	18	10185	80
14	Princeton	1375	91	14	8	30220	95
15	Purdue	1005	28	90	19	9066	69

```python
scaler = StandardScaler()
scaled = scaler.fit_transform(data.iloc[:,1:])
scaled
```

```
array([[ 0.41028362,  0.6575195 , -0.88986682,  0.07026045, -0.33141256,
         0.82030265],
       [ 1.39925928,  1.23521235, -0.73465749, -1.68625071,  2.56038138,
        -0.64452351],
       [-0.06065717, -0.76045386,  1.02438157, -0.93346022, -0.16712136,
        -1.65863393],
       [ 0.41028362, -0.02520842, -0.78639393, -0.18066972,  0.29164871,
         0.14422904],
       [ 0.12771914,  0.34241431, -0.32076595,  0.07026045, -0.39084607,
         0.36958691],
       [ 0.69284809,  0.6575195 , -0.83813038, -0.68253005,  0.33778044,
         0.93298158],
       [ 0.4573777 ,  0.71003703, -0.47597528, -0.18066972,  0.29695528,
         0.93298158],
       [-0.10775125, -0.13024348, -0.78639393, -0.18066972, -0.51381683,
         0.59494478],
       [ 1.25797704,  0.76255456, -1.30375836, -0.43159988,  0.85874344,
         1.15833946],
       [ 0.36318954, -0.07772595,  0.24833493, -1.43532055,  2.21481798,
         0.0315501 ],
```

```
wss=[]
for i in range (1,11):
    kmeans=KMeans(n_clusters=i)
    kmeans.fit(scaled)
    wss.append(kmeans.inertia_)

plt.plot(range(1,11),wss)
plt.title('Elbow Curve',fontsize=16,fontweight='bold')
fig=plt.figure(figsize=(16,8))
```



```
<Figure size 1152x576 with 0 Axes>
```
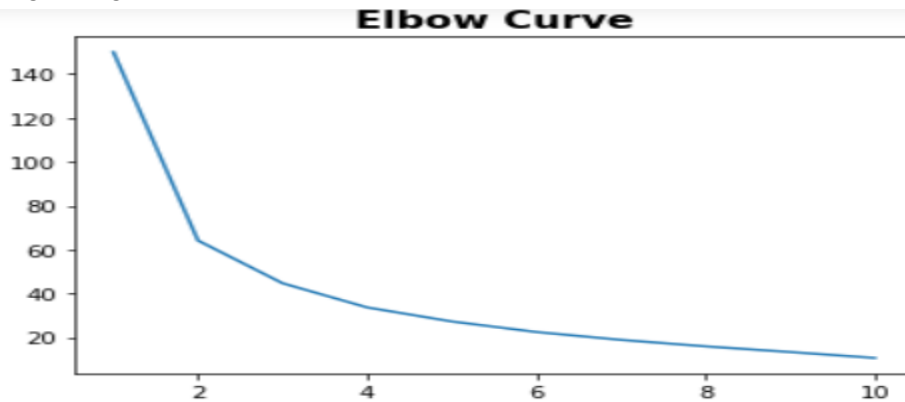
```
clusters = KMeans(n_clusters=4)
clusters.fit(scaled)
```

```
KMeans(n_clusters=4)
```

```
clusters.labels_
```

```
array([0, 3, 2, 0, 2, 0, 0, 2, 0, 3, 0, 2, 2, 1, 0, 1, 0, 1, 2, 2, 2, 2,
       2, 1, 0])
```

```
data['new_clusters'] = clusters.labels_
```

```
clusters.cluster_centers_
```

```
array([[ 0.80273428,  0.68086062, -0.90136381, -0.43159988,  0.44062556,
         0.79526289],
       [-1.93029211, -1.98148647,  1.59348244,  1.63857398, -1.23359906,
        -1.68680366],
       [-0.12658888,  0.06407139,  0.2224667 ,  0.04516743, -0.38064332,
         0.02028221],
       [ 0.88122441,  0.5787432 , -0.24316128, -1.56078563,  2.38759968,
        -0.3064867 ]])
```

```python
data.groupby('new_clusters').agg(['mean']).reset_index(drop=True)
```

| | SAT | Top10 | Accept | SFRatio | Expenses | GradRate |
	mean	mean	mean	mean	mean	mean
0	1351.666667	89.444444	21.777778	11.00	33615.555556	93.777778
1	1061.500000	38.750000	70.000000	19.25	9953.000000	71.750000
2	1253.000000	77.700000	43.500000	12.90	22008.200000	86.900000
3	1360.000000	87.500000	34.500000	6.50	61133.000000	84.000000

```python
data
```

	Univ	SAT	Top10	Accept	SFRatio	Expenses	GradRate	new_clusters
0	Brown	1310	89	22	13	22704	94	0
1	CalTech	1415	100	25	6	63575	81	3
2	CMU	1260	62	59	9	25026	72	2
3	Columbia	1310	76	24	12	31510	88	0
4	Cornell	1280	83	33	13	21864	90	2
5	Dartmouth	1340	89	23	10	32162	95	0
6	Duke	1315	90	30	12	31585	95	0
7	Georgetown	1255	74	24	12	20126	92	2
8	Harvard	1400	91	14	11	39525	97	0
9	JohnsHopkins	1305	75	44	7	58691	87	3
10	MIT	1380	94	30	10	34870	91	0
11	Northwestern	1260	85	39	11	28052	89	2
12	NotreDame	1255	81	42	13	15122	94	2
13	PennState	1081	38	54	18	10185	80	1
14	Princeton	1375	91	14	8	30220	95	0

```python
sns.lmplot('SAT','Top10',data=data,hue='new_clusters',fit_reg=False,size=6)
plt.title('KMeans plot',figsize=16,fontweight='bold')
fig=plt.figure(figsize=(16,8))
```