```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
import seaborn as sns

df = pd.read_csv('D:\ML\Mall_Customers.csv')
df
```

[79] ✓ 0.0s

...

	CustomerID	Gender	Age	Annual Income (k$)
0	1	Male	19	15
1	2	Male	21	15
2	3	Female	20	16
3	4	Female	23	16
4	5	Female	31	17
...	...	...	...	...
195	196	Female	35	120
196	197	Female	45	126
197	198	Male	32	126
198	199	Male	32	137
199	200	Male	30	137

200 rows × 4 columns

```
df = df.select_dtypes(['int64', 'float64'])
df = df.drop(['CustomerID'], axis=1)
df
```

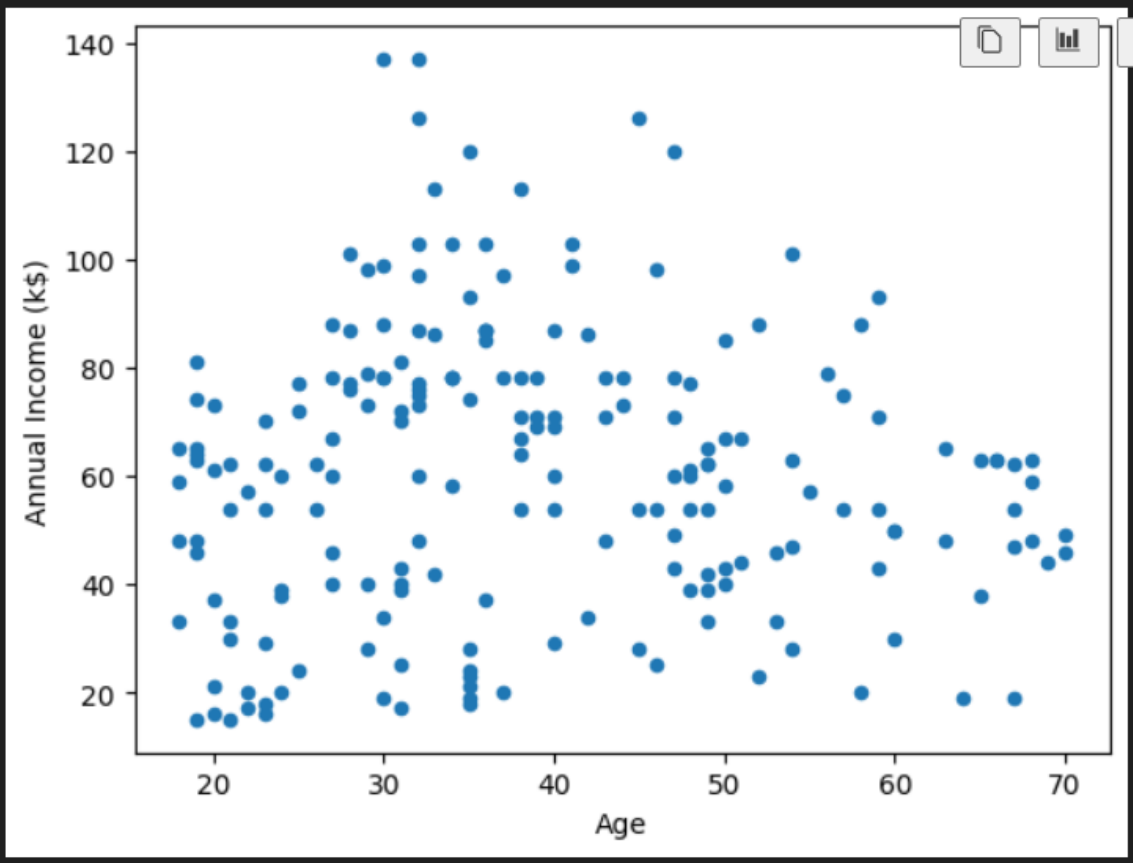
[80] ✓ 0.0s

	Age	Annual Income (k$)
0	19	15
1	21	15
2	20	16
3	23	16
4	31	17
...	...	...
195	35	120
196	45	126
197	32	126
198	32	137
199	30	137

200 rows × 2 columns

```
df.plot(kind ='scatter', x='Age', y='Annual Income (k$)')
plt.show()
```

✓ 0.1s



```
km3 = KMeans(n_clusters = 3)
km3 = km3.fit(df)
km3.labels_
```

[82] ✓ 0.0s

```
array([0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 2, 0, 0, 0, 0, 0, 2, 0, 2, 0, 2, 0, 0, 0, 0, 0, 2, 0, 2, 0,
       2, 0, 2, 0, 0, 0, 2, 0, 0, 2, 2, 2, 2, 2, 0, 2, 2, 0, 2, 2, 2, 0,
       2, 2, 0, 0, 2, 2, 2, 2, 2, 0, 2, 2, 0, 2, 2, 2, 2, 2, 0, 2, 2, 0,
       2, 2, 2, 1, 2, 2, 1, 1, 2, 1, 2, 1, 1, 2, 2, 1, 2, 1, 2, 2, 2, 2,
       2, 1, 1, 1, 1, 1, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1], dtype=int32)
```

```
labels = km3.labels_
labels = pd.DataFrame(labels, columns=['cluster'])
df = pd.concat([df, labels], axis=1)
```

(variable) labels: Any

[83] ✓ 0.0s

```
df.groupby('cluster').size()
```

[84] ✓ 0.0s

```
cluster
0    51
1    91
2    58
dtype: int64
```

```
centroids = km3.cluster_centers_
feature_columns = df.columns[:-1]  # Giả sử cột cuối cùng là 'cluster'
centroids = pd.DataFrame(centroids, columns=feature_columns)
centroids
```

[85] ✓ 0.0s Python

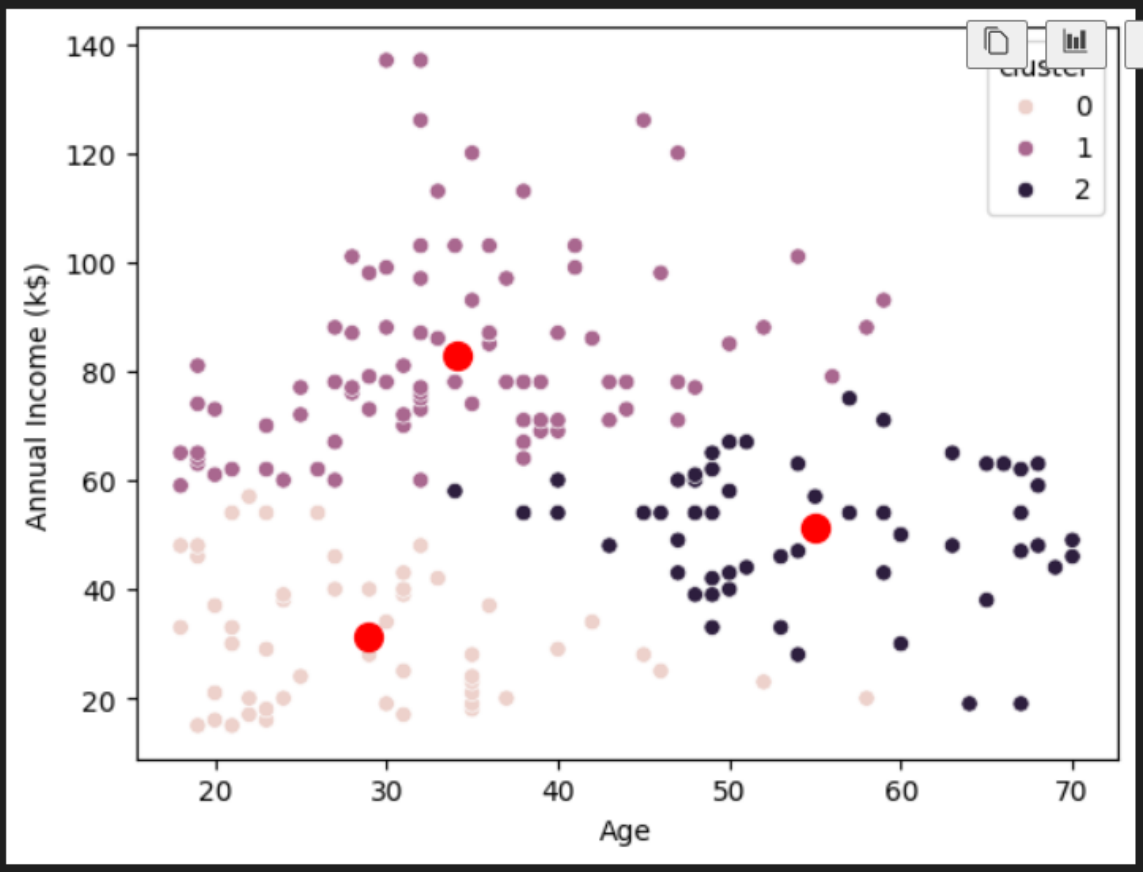
	Age	Annual Income (k$)
0	28.941176	31.215686
1	34.098901	82.912088
2	55.017241	51.293103

+ Code + Markdown

```
s1 = sns.scatterplot(data=df, x='Age', y='Annual Income (k$)', hue='cluster')
centroids.plot(ax=s1, kind ='scatter', x='Age', y='Annual Income (k$)',
color='red', s=100)
```

[87] ✓ 0.1s Python

```
<Axes: xlabel='Age', ylabel='Annual Income (k$)'>
```



```
km5 = KMeans(n_clusters = 5)
km5 = km5.fit(df)
km5.labels_
```

✓ 0.0s

```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 4, 2, 4, 2, 4,
       2, 4, 2, 4, 4, 4, 2, 4, 4, 2, 2, 2, 2, 2, 4, 2, 2, 4, 2, 2, 2, 4,
       2, 2, 4, 4, 2, 2, 2, 2, 2, 4, 2, 4, 4, 2, 2, 4, 2, 2, 4, 2, 2, 4,
       4, 2, 2, 4, 2, 4, 4, 4, 2, 4, 2, 4, 4, 2, 2, 4, 2, 4, 2, 2, 2, 2,
       2, 4, 0, 4, 4, 4, 2, 2, 2, 2, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
       3, 3], dtype=int32)
```

```
labels = km5.labels_
labels = pd.DataFrame(labels, columns=['cluster'])
df = pd.concat([df, labels], axis=1)
```

✓ 0.0s

```
df.groupby('cluster').size()
```

✓ 0.0s

```
cluster
0    60
1    37
2    49
3    20
4    34
dtype: int64
```

```
centroids = km5.cluster_centers_
feature_columns = ['Age', 'Annual Income (k$)', 'Spending Score (1-100)']
centroids = pd.DataFrame(centroids, columns=feature_columns)
centroids
```

✓ 0.0s Python

	Age	Annual Income (k$)	Spending Score (1-100)
0	36.033333	78.050000	1.016667
1	33.864865	23.729730	0.216216
2	56.081633	52.551020	2.000000
3	36.600000	109.700000	1.000000
4	25.735294	52.411765	0.588235

```
s1 = sns.scatterplot(data=df, x='Age', y='Annual Income (k$)', hue='cluster')
centroids.plot(ax=s1, kind ='scatter', x='Age', y='Annual Income (k$)',
color='red', s=100)
```

✓ 0.1s Python

```
<Axes: xlabel='Age', ylabel='Annual Income (k$)'>
```