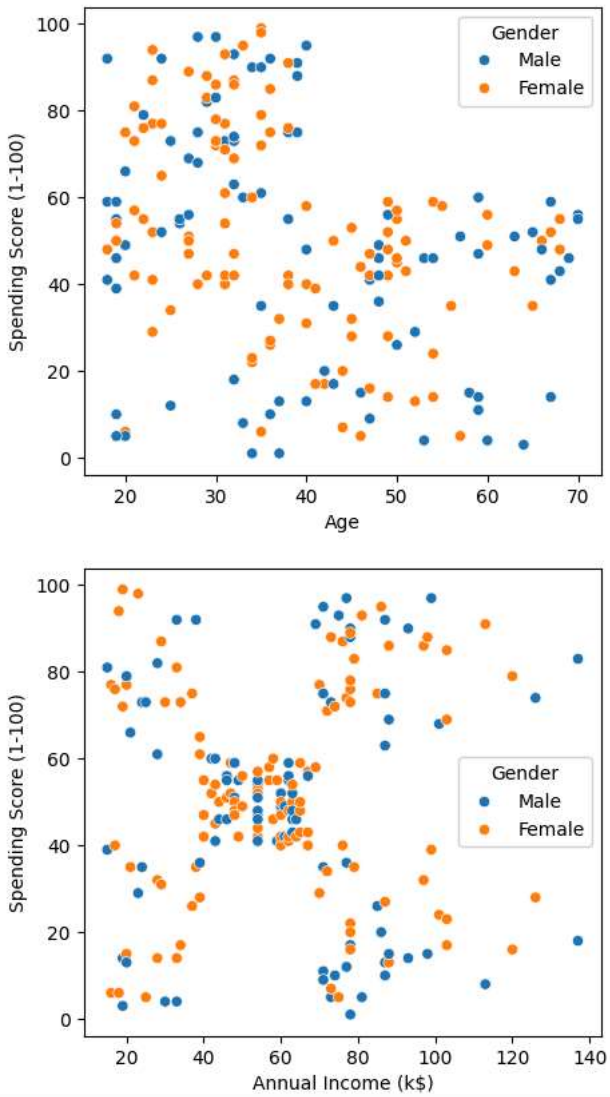
```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns


from sklearn.cluster import KMeans


df=pd.read_csv('/content/Mall_Customers.csv')
df.head()
```

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

+ Code + Text

```
fig,ax=plt.subplots(2,figsize=(5,10))
sns.scatterplot(ax=ax[0],data=df,x='Age',y='Spending Score (1-100)',hue='Gender')
sns.scatterplot(ax=ax[1],data=df,x='Annual Income (k$)',y='Spending Score (1-100)',hue='Gender')
plt.show()
```




```
km=KMeans(n_clusters=5)
df['cluster']=km.fit_predict(df[['Annual Income (k$)','Spending Score (1-100)']])
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
```

```
df.head()
```

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)	cluster
0	1	Male	19	15	39	3
1	2	Male	21	15	81	0
2	3	Female	20	16	6	3
3	4	Female	23	16	77	0
4	5	Female	31	17	40	3

```
km.cluster_centers_
```

```
array([[25.72727273, 79.36363636],
       [88.2       , 17.11428571],
       [55.2962963 , 49.51851852],
       [26.30434783, 20.91304348],
       [86.53846154, 82.12820513]])
```

```
km.cluster_centers_
```

```
array([[25.72727273, 79.36363636],
       [88.2       , 17.11428571],
       [55.2962963 , 49.51851852],
       [26.30434783, 20.91304348],
       [86.53846154, 82.12820513]])
```

```
km.inertia_
```

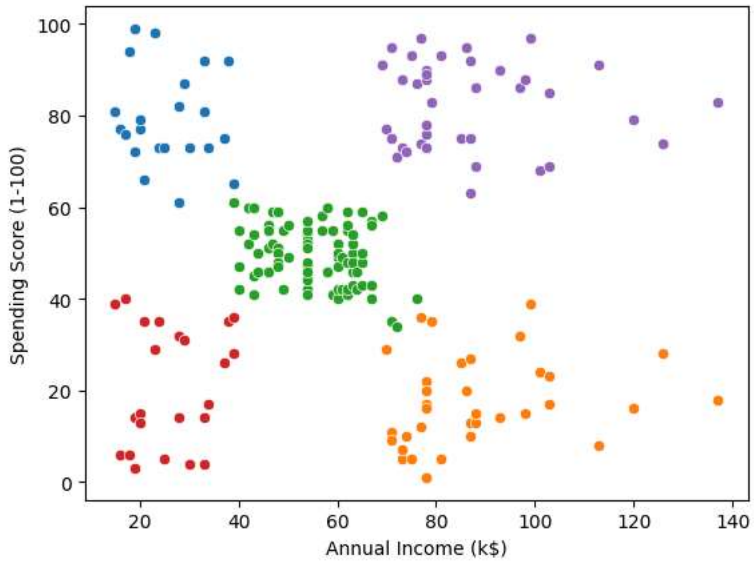
```
44448.45544793369
```

```
df['cluster'].unique()
```

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

```
df0=df[df['cluster']==0][['Annual Income (k$)','Spending Score (1-100)','cluster']]
df1=df[df['cluster']==1][['Annual Income (k$)','Spending Score (1-100)','cluster']]
df2=df[df['cluster']==2][['Annual Income (k$)','Spending Score (1-100)','cluster']]
df3=df[df['cluster']==3][['Annual Income (k$)','Spending Score (1-100)','cluster']]
df4=df[df['cluster']==4][['Annual Income (k$)','Spending Score (1-100)','cluster']]
```

```
sns.scatterplot(data=df0,x='Annual Income (k$)',y='Spending Score (1-100)')
sns.scatterplot(data=df1,x='Annual Income (k$)',y='Spending Score (1-100)')
sns.scatterplot(data=df2,x='Annual Income (k$)',y='Spending Score (1-100)')
sns.scatterplot(data=df3,x='Annual Income (k$)',y='Spending Score (1-100)')
sns.scatterplot(data=df4,x='Annual Income (k$)',y='Spending Score (1-100)')
plt.show()
```




```
sse=[]
for k in range(1,20):
    kmk=KMeans(n_clusters=k)
    kmk.fit(df[['Annual Income (k$)','Spending Score (1-100)']])
    sse.append(kmk.inertia_)
```

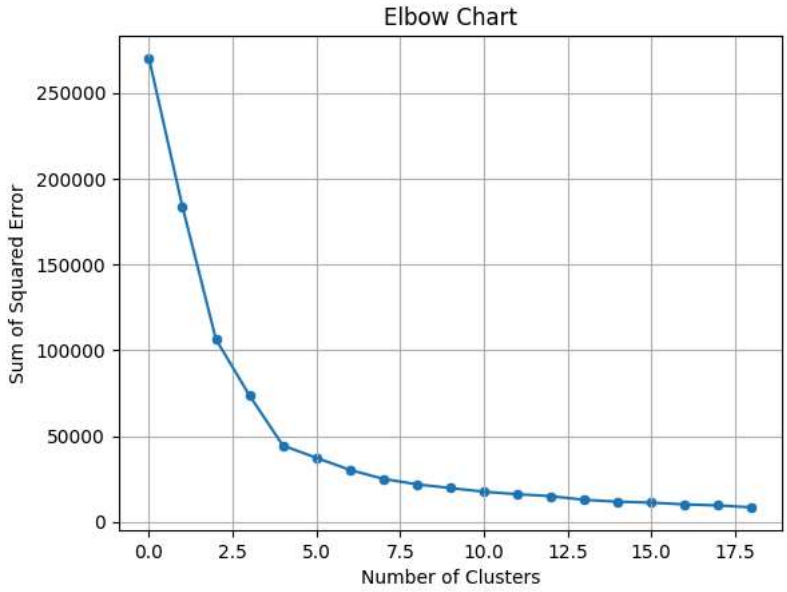
```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 1
  super()._check_params_vs_input(X, default_n_init=10)
```

```
sse
```

```
[269981.28000000014,
 183653.3289473683,
 106348.37306211119,
 73679.78903948837,
 44448.45544793369,
 37265.86520484345,
 30259.657207285458,
 25022.485004530332,
```

```
    21841.978256748636,
    19676.612585602812,
    17546.928000046544,
    16099.92568836392,
    14946.724459143003,
    12814.183489601206,
    11723.801925267711,
    11165.151003881174,
    10157.744683335979,
    9537.97646175663,
    8424.892959671906]
```

```
sns.lineplot(data=sse)
sns.scatterplot(data=sse)
plt.grid(True)
plt.xlabel('Number of Clusters')
plt.ylabel('Sum of Squared Error')
plt.title('Elbow Chart')
plt.show()
```




```
sns.scatterplot(data=df0,x='Annual Income (k$)',y='Spending Score (1-100)')
sns.scatterplot(data=df1,x='Annual Income (k$)',y='Spending Score (1-100)')
sns.scatterplot(data=df2,x='Annual Income (k$)',y='Spending Score (1-100)')
sns.scatterplot(data=df3,x='Annual Income (k$)',y='Spending Score (1-100)')
sns.scatterplot(data=df4,x='Annual Income (k$)',y='Spending Score (1-100)')
plt.show()
```