

```

1
2 import pandas as pd
3 import numpy as np
4 import matplotlib.pyplot as plt
5 import seaborn as sns

```

```

1 data = pd.read_csv("Mall_Customers.csv")
2 data.head()

```

	CustomerID	Genre	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

```
1 data.describe()
```



	CustomerID	Age	Annual Income (k\$)	Spending Score (1-100)
<b>count</b>	200.000000	200.000000	200.000000	200.000000
<b>mean</b>	100.500000	38.850000	60.560000	50.200000
<b>std</b>	57.879185	13.969007	26.264721	25.823522
<b>min</b>	1.000000	18.000000	15.000000	1.000000
<b>25%</b>	50.750000	28.750000	41.500000	34.750000
<b>50%</b>	100.500000	36.000000	61.500000	50.000000
<b>75%</b>	150.250000	49.000000	78.000000	73.000000
<b>max</b>	200.000000	70.000000	137.000000	99.000000

```
1 data.isnull().sum()
```

```

CustomerID      0
Genre            0
Age             0
Annual Income (k$)  0
Spending Score (1-100)  0
dtype: int64

```

```

1 data['Genre'] = data['Genre'].map({'Male':0, 'Female' : 1})
2 data

```

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	0	19	15	39
1	2	0	21	15	81
2	3	1	20	16	6
3	4	1	23	16	77
4	5	1	31	17	40
...	...	...	...	...	...
195	196	1	35	120	79
196	197	1	45	126	28
197	198	0	32	126	74
198	199	0	32	137	18
199	200	0	30	137	83

200 rows × 5 columns

```
1 df = pd.DataFrame(data['Annual Income (k$)'])
```

```
1 df.head()
```

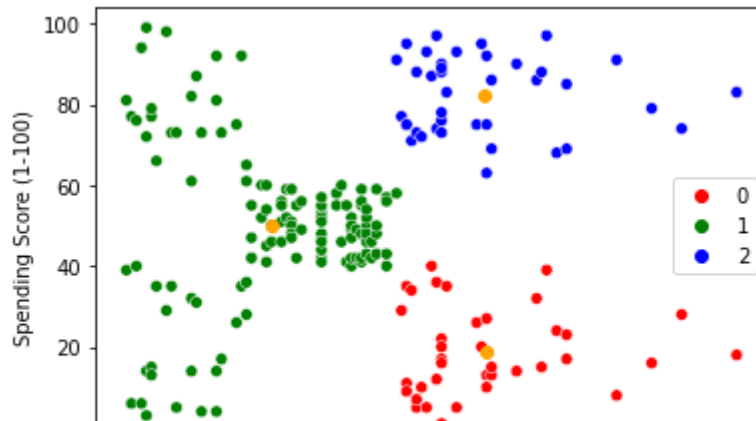
	Annual Income (k\$)
0	15
1	15
2	16
3	16
4	17

```
1 df['Spending Score (1-100)'] = data['Spending Score (1-100)']
2 df.head()
```

	Annual Income (k\$)	Spending Score (1-100)
0	15	39
1	15	81
2	16	6
3	16	77
4	17	40

```
1 plt.scatter(data['Annual Income (k$)'],data['Spending Score (1-100)'])
```





```

1 wcss = []
2 for i in range(1, 11):
3     kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 0)
4     kmeans.fit(df)
5     wcss.append(kmeans.inertia_)
6 wcss

```

```

[269981.2800000001,
 181363.59595959596,
 106348.3730621112,
 73679.78903948834,
 44448.45544793371,
 37265.86520484346,
 30259.65720728547,
 25095.703209997548,
 21830.041978049438,
 20736.679938924128]

```

```

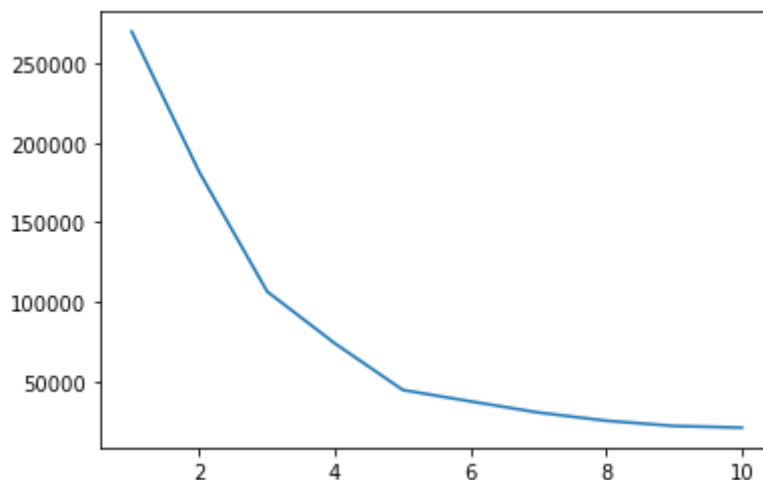
1 li = [1,2,3,4,5,6,7,8,9,10]
2 sns.lineplot(li,wcss)

```

```

/home/pvg/anaconda3/lib/python3.9/site-packages/seaborn/_decorators.py:36: FutureWarning:
  warnings.warn(
<AxesSubplot:>

```



```

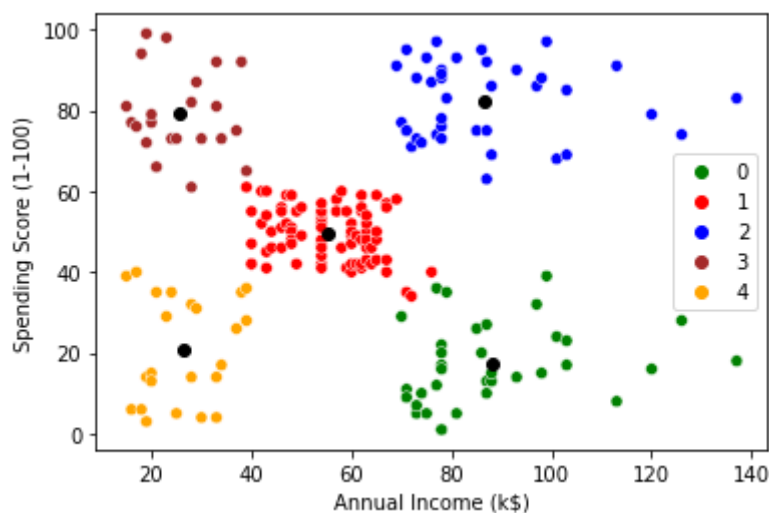
1 kmeans = KMeans(n_clusters = 5, init = "k-means++", random_state = 0)
2 y_kmeans = kmeans.fit_predict(df)

```

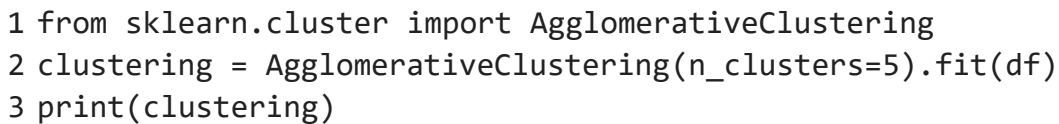
[illegible]

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

```
1 sns.scatterplot(data=df, x = df['Annual Income (k$)'], y = df['Spending Score (1-100)'])
2 plt.scatter(xc,yc,color='black')
3 plt.show()
```



```
1 import scipy.cluster.hierarchy as sch
2 dendrogram = sch.dendrogram(sch.linkage(df,method = 'ward'))
```

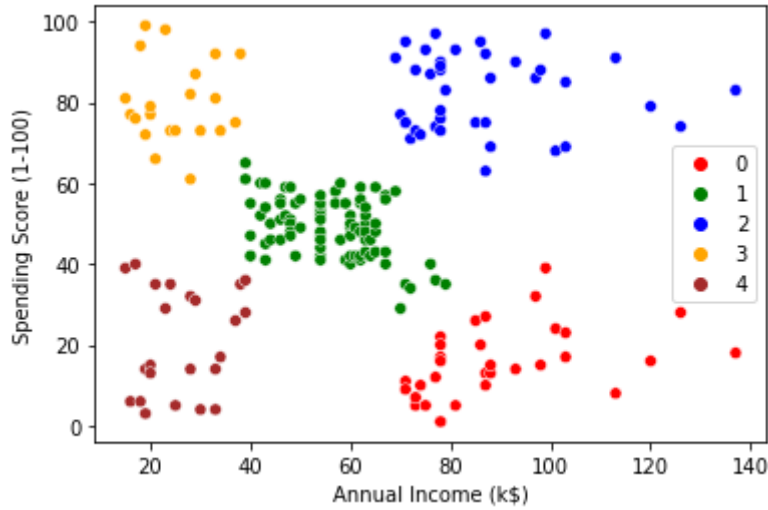


```
1 clustering.labels_
```

```
1 df.info()
```

```
1 sns.scatterplot(data = df, x = df['Annual Income (k$)'], y = df['Spending Score (1-100)'])
```

<AxesSubplot:xlabel='Annual Income (k\$)', ylabel='Spending Score (1-100)'\>



[Colab paid products](#) - [Cancel contracts here](#)

