

In [1]: #10

```
from google.colab import files
uploaded = files.upload()

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

df = pd.read_csv('Mall_Customers.csv')
print(df.info())
print(df.head())

sns.pairplot(df)
plt.show()

features = df.iloc[:, [3, 4]].values
from sklearn.cluster import KMeans
model = KMeans(n_clusters=5, random_state=42)
model.fit(features)

Final = df.iloc[:, [3, 4]].copy()
Final['label'] = model.predict(features)
print(Final)

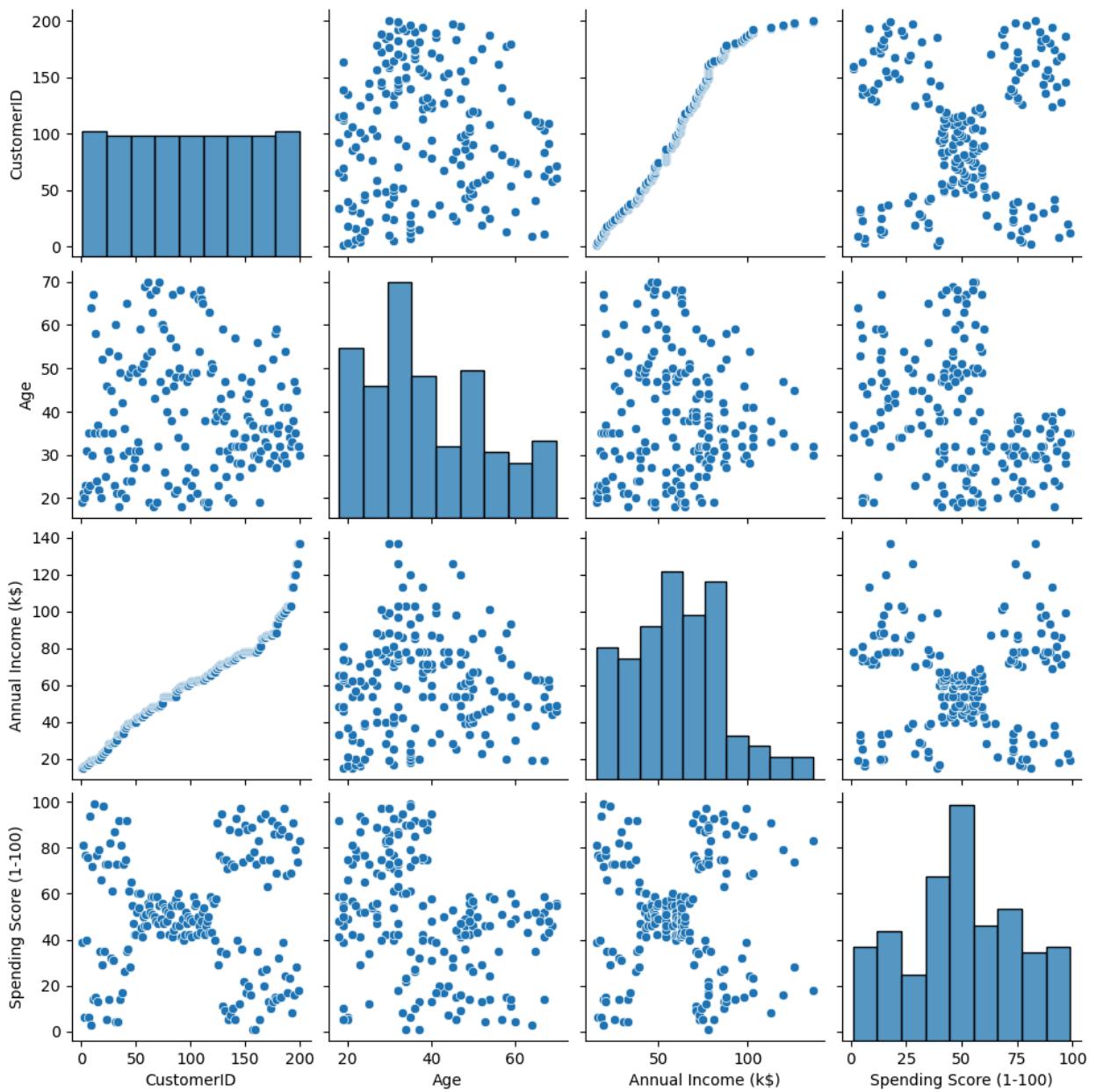
sns.set_style("whitegrid")
plt.figure(figsize=(8,6))
sns.scatterplot(data=Final, x="Annual Income (k$)", y="Spending Score (1-100)")
plt.title("K-Means Clustering of Customers")
plt.show()

features_el = df.iloc[:, [2, 3, 4]].values
wcss = []
for i in range(1, 10):
    model = KMeans(n_clusters=i, random_state=42)
    model.fit(features_el)
    wcss.append(model.inertia_)

plt.figure(figsize=(8,5))
plt.plot(range(1, 10), wcss, marker='o', color='blue')
plt.title("Elbow Method for Optimal K")
plt.xlabel("Number of Clusters")
plt.ylabel("WCSS")
plt.show()
```

Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

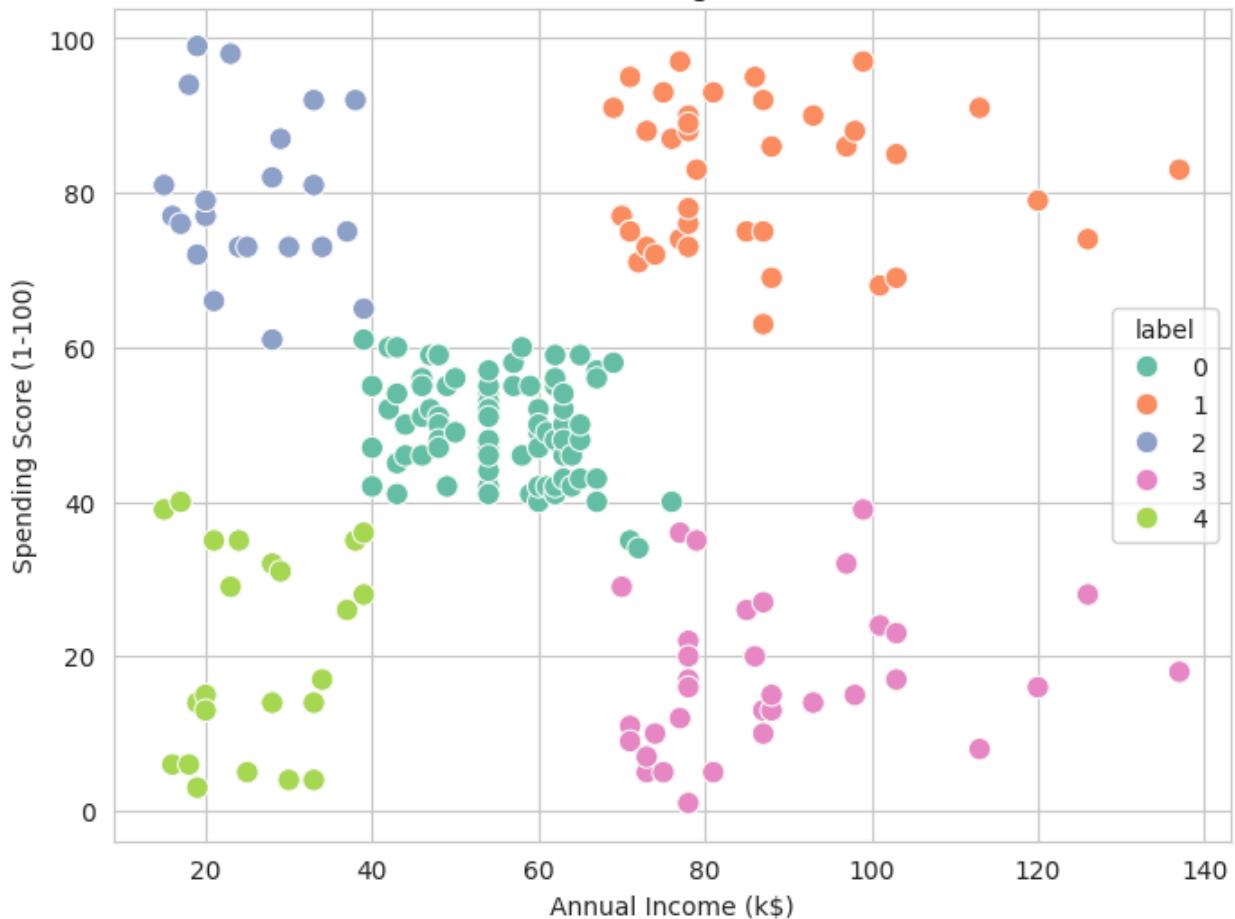
```
Saving Mall_Customers.csv to Mall_Customers.csv
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   CustomerID        200 non-null    int64  
 1   Gender             200 non-null    object  
 2   Age                200 non-null    int64  
 3   Annual Income (k$) 200 non-null    int64  
 4   Spending Score (1-100) 200 non-null  int64  
dtypes: int64(4), object(1)
memory usage: 7.9+ KB
None
   CustomerID  Gender  Age  Annual Income (k$)  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
```



	Annual Income (k\$)	Spending Score (1-100)	label
0	15	15	39 4
1	15	81	2
2	16	6	4
3	16	77	2
4	17	40	4
..	...	...	...
195	120	79	1
196	126	28	3
197	126	74	1
198	137	18	3
199	137	83	1

[200 rows x 3 columns]

### K-Means Clustering of Customers



### Elbow Method for Optimal K

