

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
In [3]: df = pd.read_csv("C:/Users/Swati/Desktop/Unsupervised_ML/Mall_Customers.csv")
```

```
In [4]: df.info()
```

```
<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   Genre                 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
```

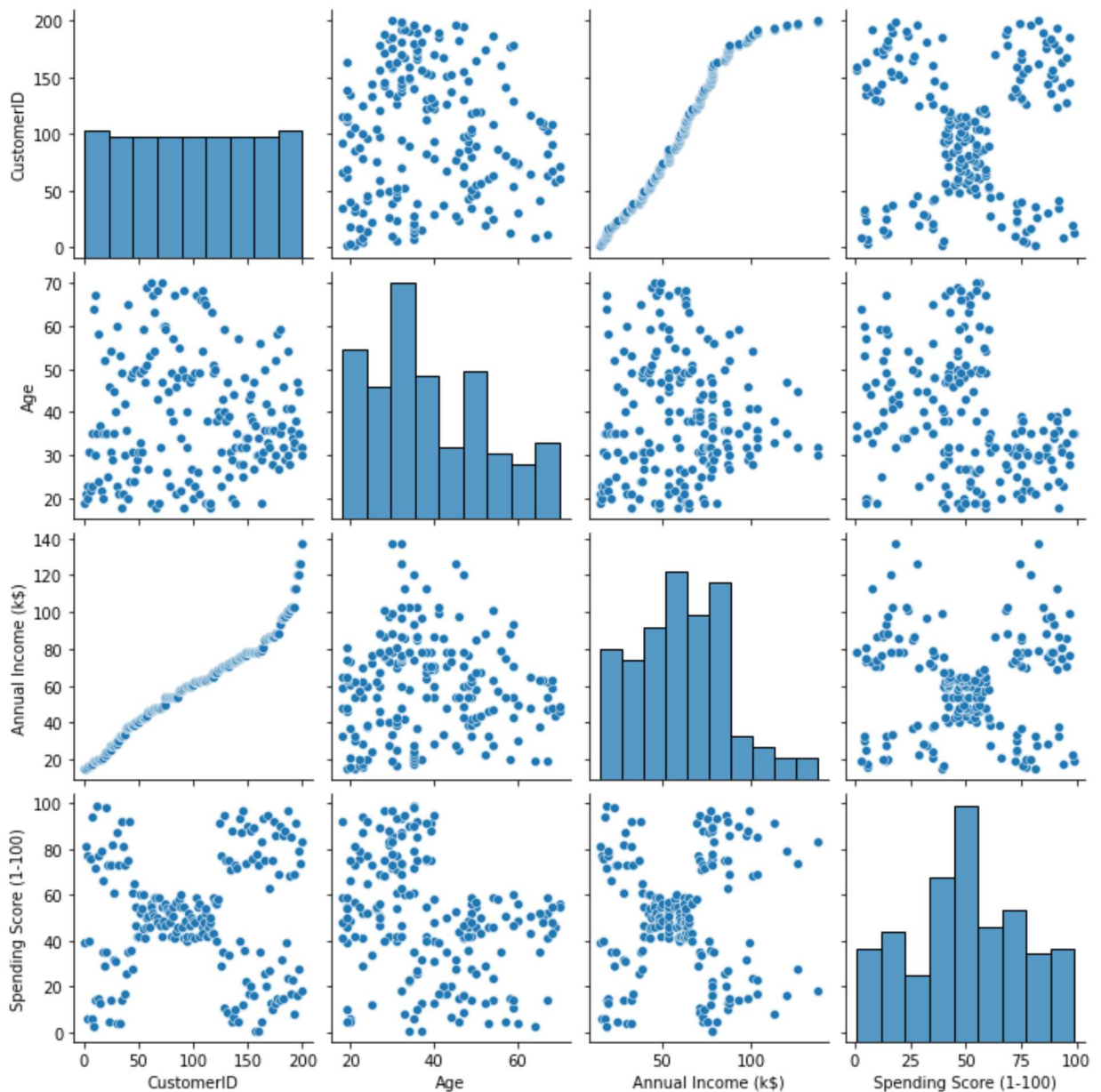
```
In [5]: df.head()
```

```
Out[5]:
```

	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

```
In [6]: sns.pairplot(df)
```

```
Out[6]: <seaborn.axisgrid.PairGrid at 0x26d77ddde20>
```



```
In [7]: df.info()
```

```
<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   Genre                  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
```

```
In [9]: features = df.iloc[:,[3,4]].values
```

```
In [10]: from sklearn.cluster import KMeans
model = KMeans(n_clusters=5)
model.fit(features)
```

```
Out[10]: KMeans(n_clusters=5)
```

```
In [12]: Final = df.iloc[:,[3,4]]
Final['label'] = model.predict(features) #Labels
Final.head()
```

C:\Users\Swati\AppData\Local\Temp\ipykernel\_14148\1120076839.py:2: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
Final['label'] = model.predict(features) #Labels
```

```
Out[12]:
```

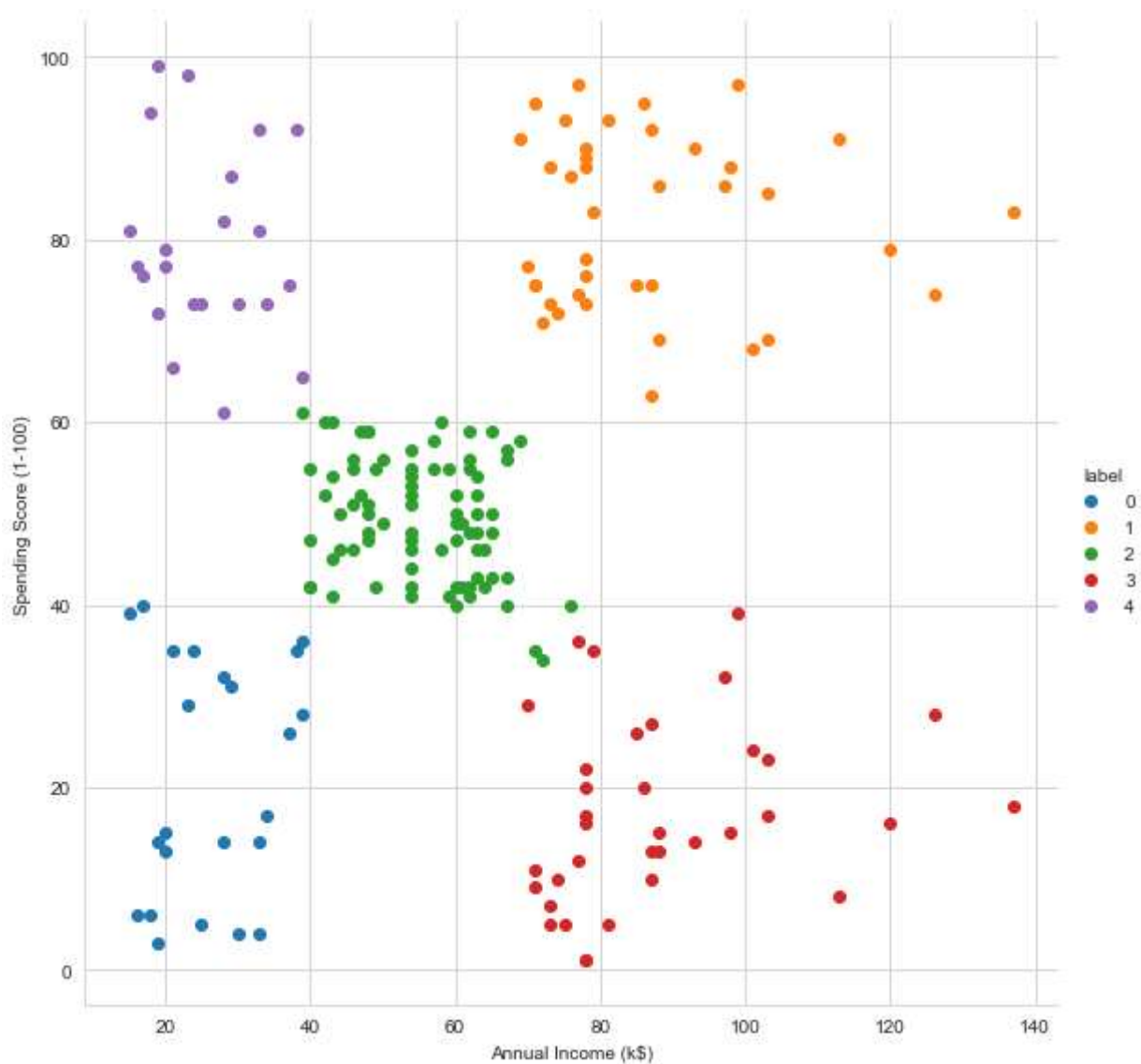
	Annual Income (k\$)	Spending Score (1-100)	label
0	15	39	0
1	15	81	4
2	16	6	0
3	16	77	4
4	17	40	0

```
In [15]: sns.set_style("whitegrid")
sns.FacetGrid(Final, hue="label", size= 8) \
.map(plt.scatter, "Annual Income (k$)", "Spending Score (1-100)") \
.add_legend();

plt.show()
```

C:\Users\Swati\anaconda3\lib\site-packages\seaborn\axisgrid.py:337: UserWarning: The `size` parameter has been renamed to `height`; please update your code.

```
warnings.warn(msg, UserWarning)
```



In [17]: *### Elbow method to check the ideal value of k*

```
features_el = df.iloc[:, [2, 3, 4]].values
from sklearn.cluster import KMeans
wcss=[]
```

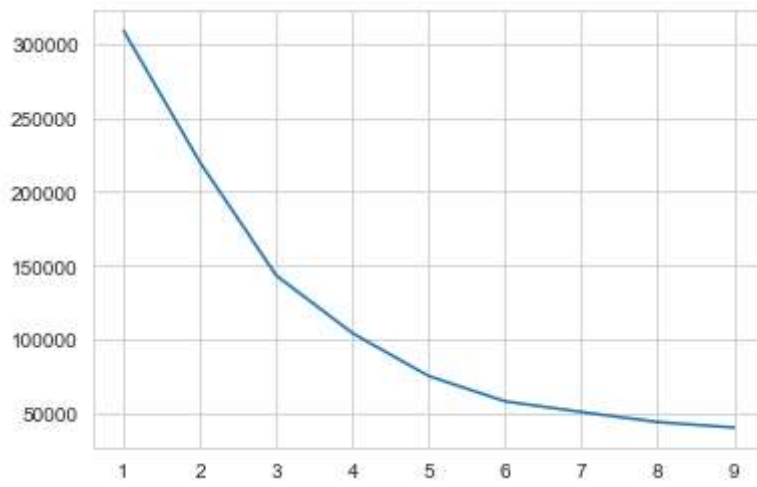
```
for i in range(1,10):
    model = KMeans(n_clusters=i)
    model.fit(features_el)
    wcss.append(model.inertia_)
```

```
plt.plot(range(1,10),wcss)
```

C:\Users\Swati\anaconda3\lib\site-packages\sklearn\cluster\\_kmeans.py:1036: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment variable OMP\_NUM\_THREADS=1.

warnings.warn(

Out[17]: [



In [ ]: