

```
In [1]: import pandas as pd
        from matplotlib import pyplot as plt
        %matplotlib inline
```

```
In [2]: df=pd.read_csv(r"C:\Users\LENOVO\Downloads\Income.csv")
        df
```

Out[2]:

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

200 rows × 3 columns

```
In [3]: df.describe
```

```
Out[3]: <bound method NDFrame.describe of      Gender  Age  Income($)
0      Male   19      15
1      Male   21      15
2    Female   20      16
3    Female   23      16
4    Female   31      17
..      ...   ...      ...
195  Female   35     120
196  Female   45     126
197    Male   32     126
198    Male   32     137
199    Male   30     137
```

[200 rows x 3 columns]>

```
In [4]: df.shape
```

Out[4]: (200, 3)

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

```
Out[5]:
```

	Gender	Age	Income(\$)
0	Male	19	15
1	Male	21	15
2	Female	20	16
3	Female	23	16
4	Female	31	17

```
In [6]: df.tail()
```

```
Out[6]:
```

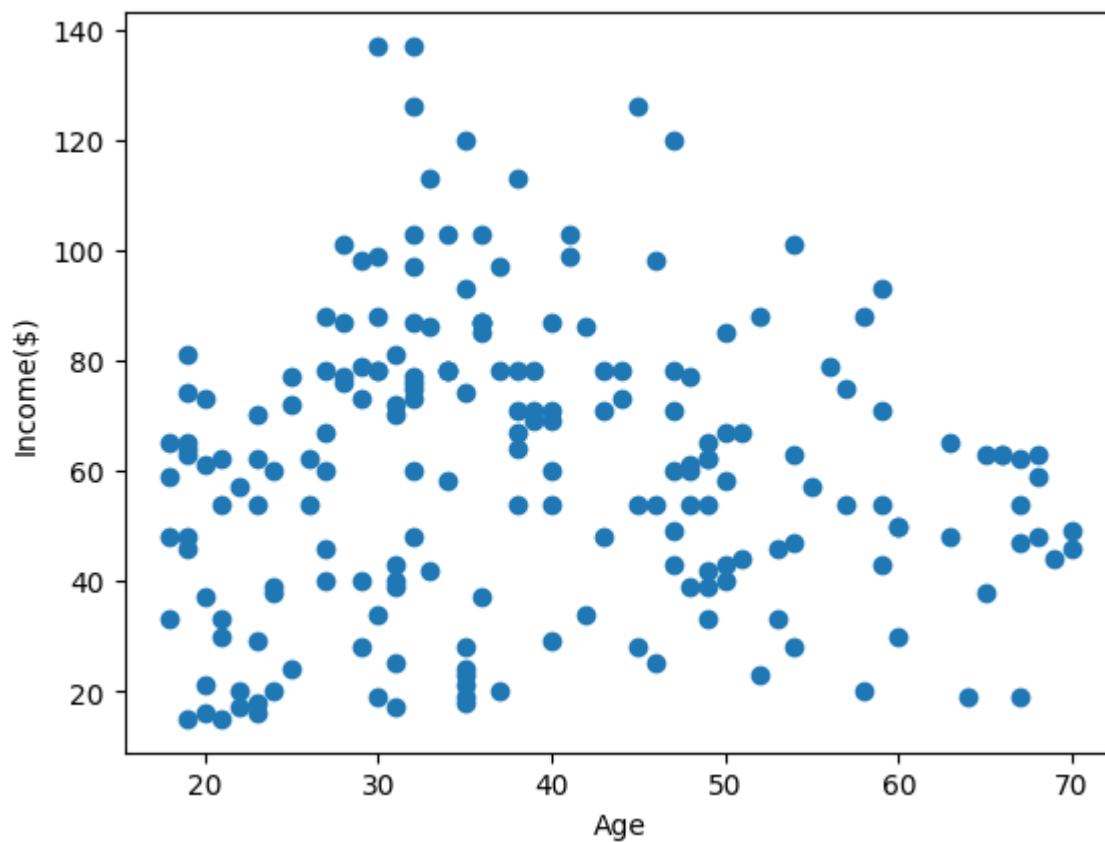
	Gender	Age	Income(\$)
195	Female	35	120
196	Female	45	126
197	Male	32	126
198	Male	32	137
199	Male	30	137

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Gender      200 non-null   object
1   Age         200 non-null   int64
2   Income($)   200 non-null   int64
dtypes: int64(2), object(1)
memory usage: 4.8+ KB
```

```
In [8]: plt.scatter(df["Age"],df["Income($)"])  
plt.xlabel("Age")  
plt.ylabel("Income($)")
```

```
Out[8]: Text(0, 0.5, 'Income($)')
```



```
In [9]: from sklearn.cluster import KMeans  
km=KMeans()  
km
```

```
Out[9]: 

▼ KMeans



KMeans()


```

```
In [10]: y_predicted=km.fit_predict(df[["Age", "Income($)"]])
y_predicted
```

C:\Users\LENOVO\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change from 10 to 'auto' in 1.4. Set the value of `n\_init` explicitly to suppress the warning  
 warnings.warn(

```
Out[10]: array([6, 6, 6, 6, 6, 6, 6, 6, 6, 4, 6, 4, 6, 4, 6, 6, 6, 6, 6, 4, 6, 6, 6,
        4, 6, 4, 6, 4, 6, 4, 6, 4, 6, 4, 2, 4, 2, 4, 2, 2, 2, 4, 2, 4, 2,
        4, 2, 4, 2, 2, 2, 4, 2, 2, 4, 4, 4, 4, 0, 2, 0, 0, 2, 0, 0, 0, 2,
        0, 0, 2, 2, 0, 0, 0, 0, 0, 5, 0, 0, 5, 0, 0, 2, 0, 0, 5, 0, 0, 5,
        5, 0, 0, 5, 0, 3, 5, 5, 0, 5, 0, 5, 5, 0, 0, 5, 0, 5, 0, 0, 0, 0,
        0, 5, 3, 5, 5, 5, 0, 0, 0, 0, 5, 3, 3, 3, 5, 3, 3, 3, 0, 3, 3, 3,
        5, 3, 5, 3, 3, 3, 5, 3, 0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
        3, 3, 3, 3, 3, 3, 3, 3, 5, 3, 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, 7, 7, 7, 7, 7, 7,
        7, 7])
```

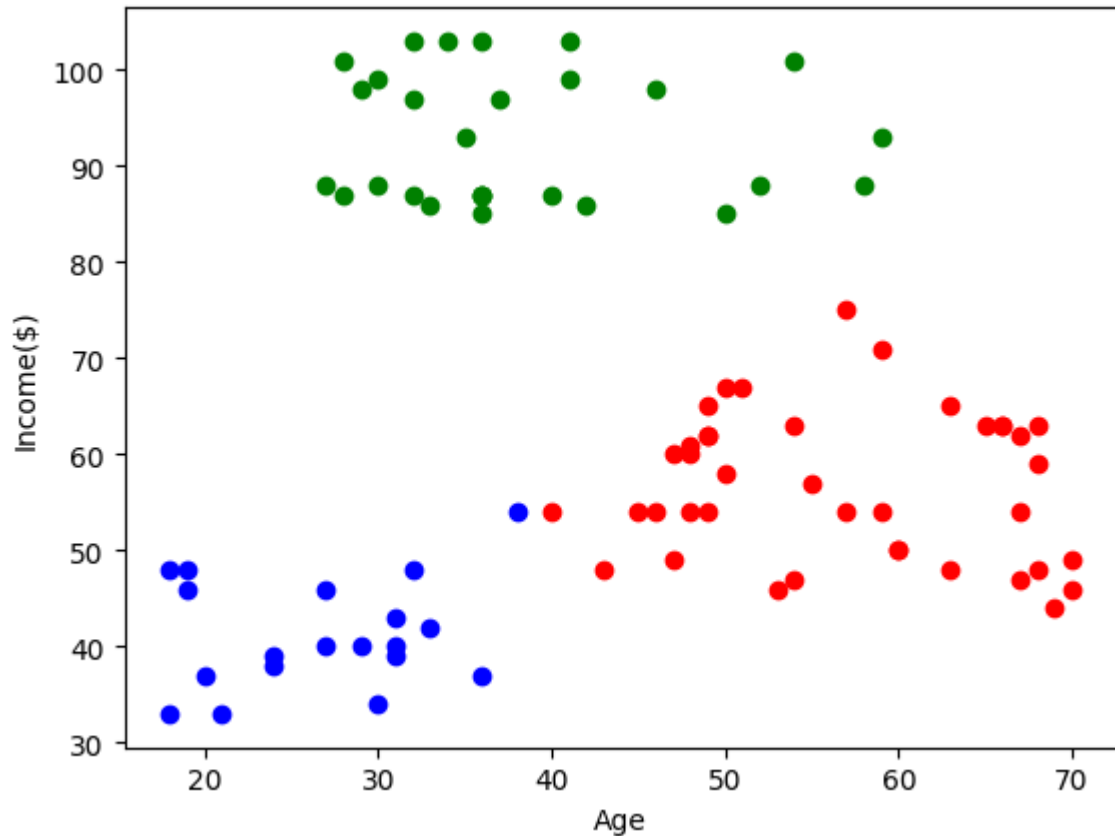
```
In [11]: df["cluster"]=y_predicted
df.head()
```

```
Out[11]:
```

	Gender	Age	Income(\$)	cluster
0	Male	19	15	6
1	Male	21	15	6
2	Female	20	16	6
3	Female	23	16	6
4	Female	31	17	6

```
In [12]: df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="red")
plt.scatter(df2["Age"],df2["Income($)"],color="green")
plt.scatter(df3["Age"],df3["Income($)"],color="blue")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[12]: Text(0, 0.5, 'Income(\$))')



```
In [13]: from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
scaler.fit(df[["Income($)"]])
df["Income($)"]=scaler.transform(df[["Income($)"]])
df.head()
```

Out[13]:

	Gender	Age	Income(\$)	cluster
0	Male	19	0.000000	6
1	Male	21	0.000000	6
2	Female	20	0.008197	6
3	Female	23	0.008197	6
4	Female	31	0.016393	6

```
In [14]: scaler.fit(df[["Age"]])
df["Age"]=scaler.transform(df[["Age"]])
df.head()
```

Out[14]:

	Gender	Age	Income(\$)	cluster
0	Male	0.019231	0.000000	6
1	Male	0.057692	0.000000	6
2	Female	0.038462	0.008197	6
3	Female	0.096154	0.008197	6
4	Female	0.250000	0.016393	6

```
In [15]: km=KMeans()
```

```
In [16]: y_predicted=km.fit_predict(df[["Age", "Income($)"]])
y_predicted
```

C:\Users\LENOVO\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change from 10 to 'auto' in 1.4. Set the value of `n\_init` explicitly to suppress the warning  
 warnings.warn(

Out[16]: array([1, 1, 1, 1, 7, 1, 7, 1, 2, 7, 2, 7, 4, 1, 7, 1, 7, 1, 4, 7, 7, 1, 4, 7, 4, 7, 4, 7, 7, 1, 2, 1, 4, 7, 4, 1, 4, 7, 7, 1, 2, 1, 4, 7, 4, 1, 4, 7, 7, 7, 4, 7, 7, 2, 4, 4, 4, 2, 5, 4, 2, 5, 2, 4, 2, 5, 4, 2, 5, 7, 2, 4, 2, 2, 2, 5, 4, 4, 5, 4, 2, 6, 2, 4, 5, 4, 0, 5, 6, 0, 2, 5, 0, 6, 6, 5, 0, 5, 0, 5, 5, 0, 2, 5, 0, 5, 2, 0, 2, 2, 2, 5, 6, 5, 5, 5, 5, 2, 0, 0, 0, 5, 6, 6, 6, 5, 6, 0, 6, 0, 6, 0, 6, 5, 6, 5, 6, 0, 6, 5, 6, 0, 6, 6, 6, 5, 6, 0, 6, 6, 6, 0, 6, 0, 6, 0, 6, 6, 6, 6, 6, 6, 0, 6, 5, 6, 0, 6, 6, 6, 6, 6, 6, 6, 6, 0, 6, 0, 6, 0, 6, 6, 3, 3, 0, 3, 3, 3, 0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3])

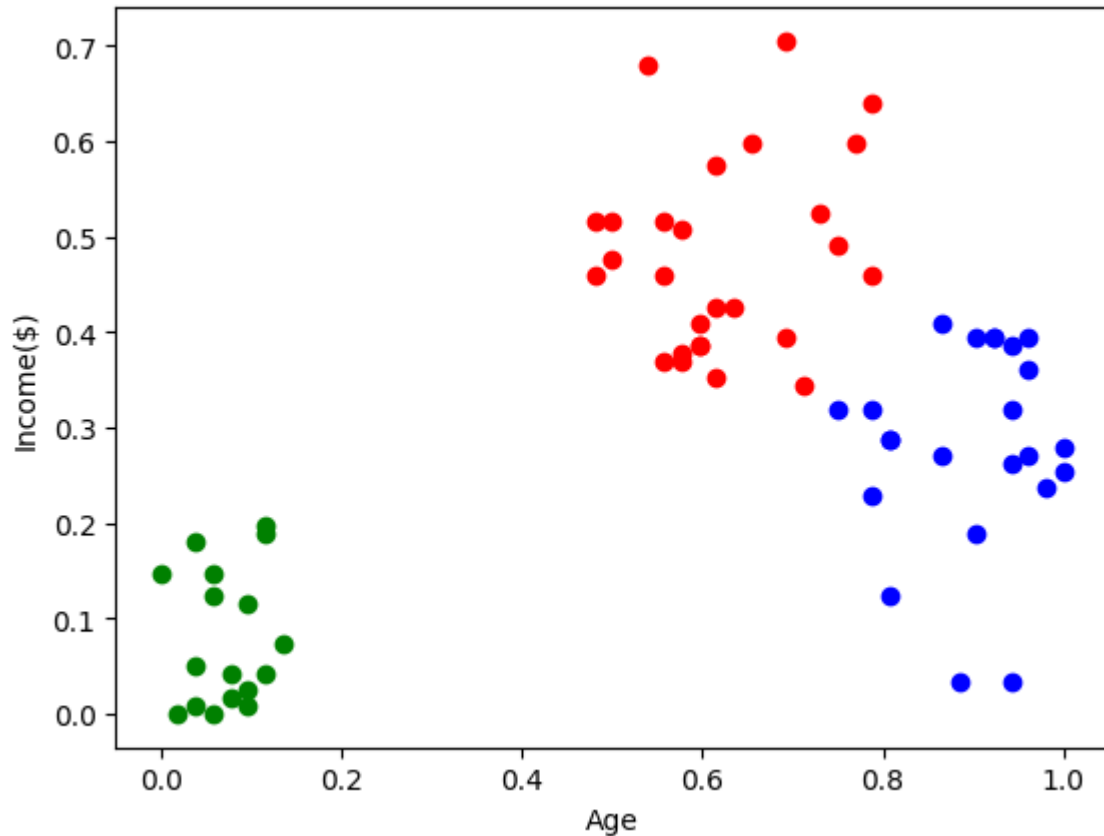
```
In [17]: df["New Cluster"]=y_predicted
df.head()
```

Out[17]:

	Gender	Age	Income(\$)	cluster	New Cluster
0	Male	0.019231	0.000000	6	1
1	Male	0.057692	0.000000	6	1
2	Female	0.038462	0.008197	6	1
3	Female	0.096154	0.008197	6	1
4	Female	0.250000	0.016393	6	7

```
In [18]: df1=df[df["New Cluster"]==0]
df2=df[df["New Cluster"]==1]
df3=df[df["New Cluster"]==2]
plt.scatter(df1["Age"],df1["Income($)"],color="red")
plt.scatter(df2["Age"],df2["Income($)"],color="green")
plt.scatter(df3["Age"],df3["Income($)"],color="blue")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[18]: Text(0, 0.5, 'Income(\$))')

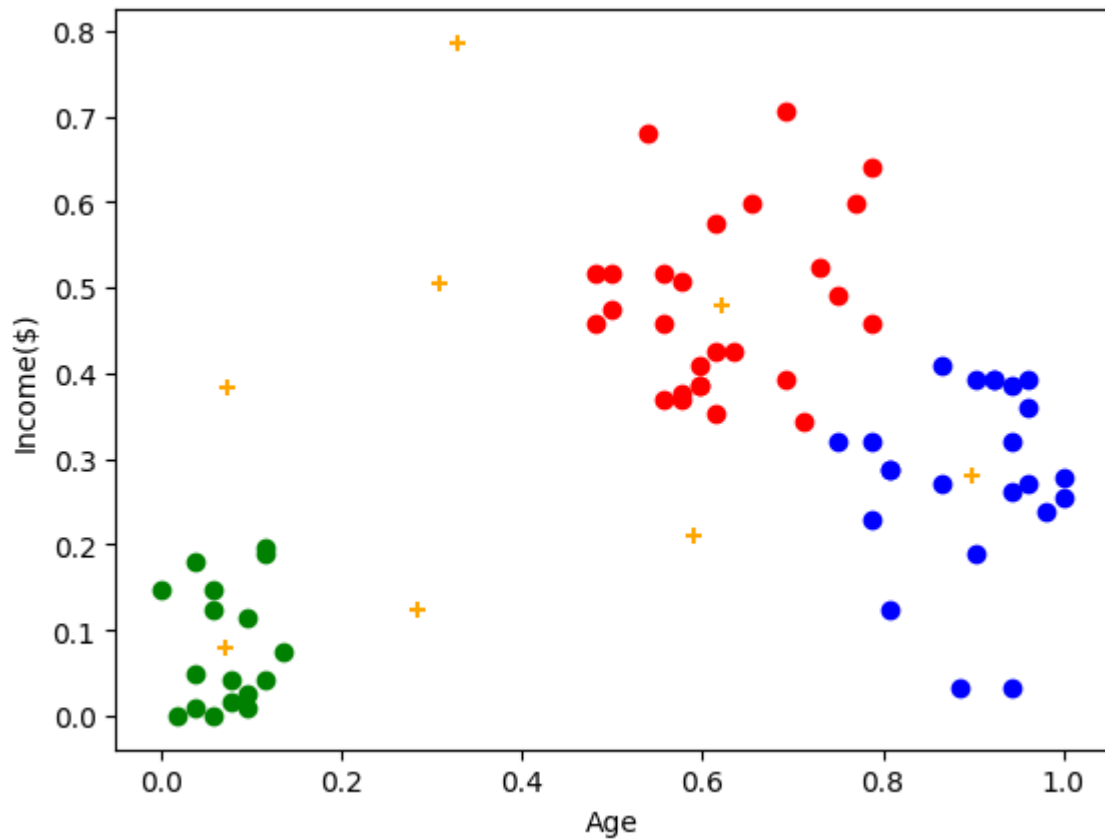


```
In [19]: km.cluster_centers_
```

Out[19]: array([[0.62037037, 0.47996357],  
 [0.07239819, 0.08003857],  
 [0.89799331, 0.28011404],  
 [0.32905983, 0.78551913],  
 [0.58974359, 0.20969945],  
 [0.07322485, 0.38272383],  
 [0.30944056, 0.50428465],  
 [0.28388278, 0.1245121 ]])

```
In [20]: df1=df[df["New Cluster"]==0]
df2=df[df["New Cluster"]==1]
df3=df[df["New Cluster"]==2]
plt.scatter(df1["Age"],df1["Income($)"],color="red")
plt.scatter(df2["Age"],df2["Income($)"],color="green")
plt.scatter(df3["Age"],df3["Income($)"],color="blue")
plt.scatter(km.cluster_centers_[0],km.cluster_centers_[1],color="orange",marker="x")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[20]: Text(0, 0.5, 'Income(\$))')





```
In [22]: k_rng=range(1,10)
sse=[]
for k in k_rng:
    km=KMeans(n_clusters=k)
    km.fit(df[["Age","Income($)"]])
    sse.append(km.inertia_) #km.inertia_ will give you the value of sum of squares
print(sse)
plt.plot(k_rng,sse)
plt.xlabel("K")
plt.ylabel("Sum of Squared Error")
```

```
C:\Users\LENOVO\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
```

```
warnings.warn(
```

```
C:\Users\LENOVO\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
```

```
warnings.warn(
```

```
C:\Users\LENOVO\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
```

```
warnings.warn(
```

```
C:\Users\LENOVO\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
```

```
warnings.warn(
```

```
C:\Users\LENOVO\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
```

```
warnings.warn(
```

```
C:\Users\LENOVO\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
```

```
warnings.warn(
```

```
C:\Users\LENOVO\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
```

```
warnings.warn(
```

```
C:\Users\LENOVO\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
```

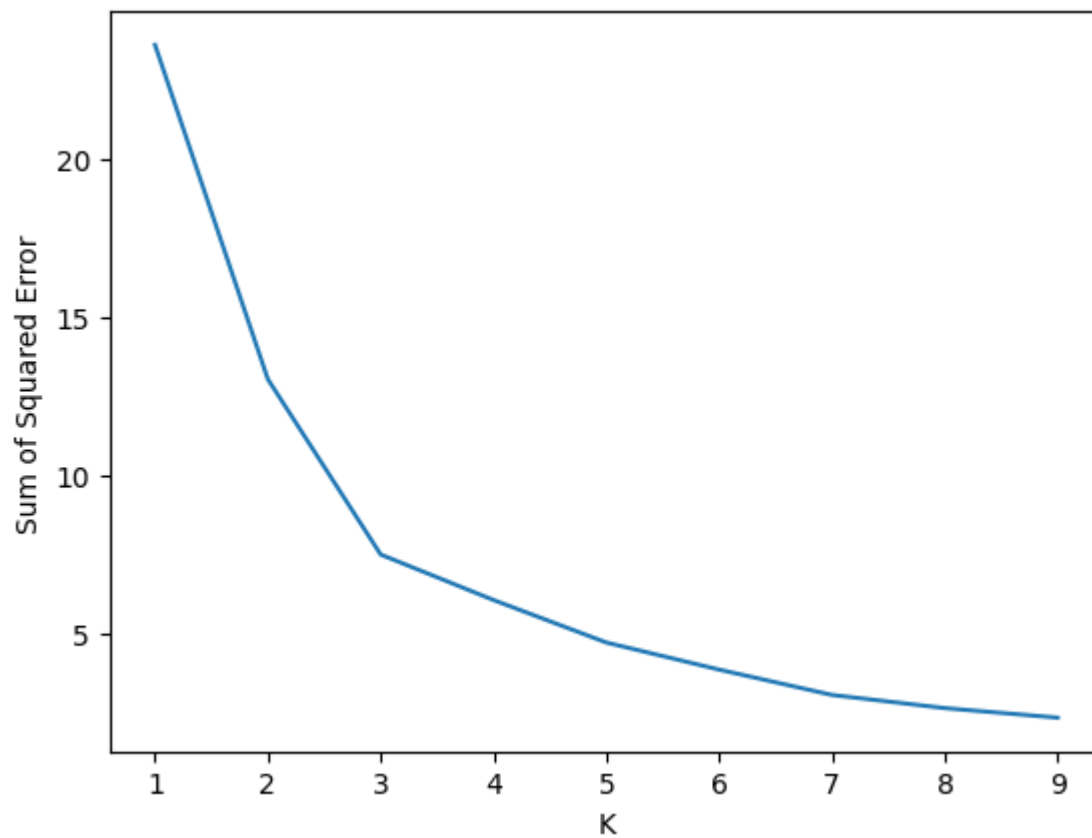
```
warnings.warn(
```

```
C:\Users\LENOVO\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
```

```
warnings.warn(
```

```
[23.583906150363607, 13.028938428018286, 7.49210786858601, 6.055824667599623, 4.713416604872824, 3.857891822164646, 3.0580611070789887, 2.642520343536072, 2.340860343707833]
```

Out[22]: Text(0, 0.5, 'Sum of Squared Error')



In [ ]: