

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

```
In [2]: df=pd.read_csv(r"C:\Users\manasa\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.head()
```

Out[3]:

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

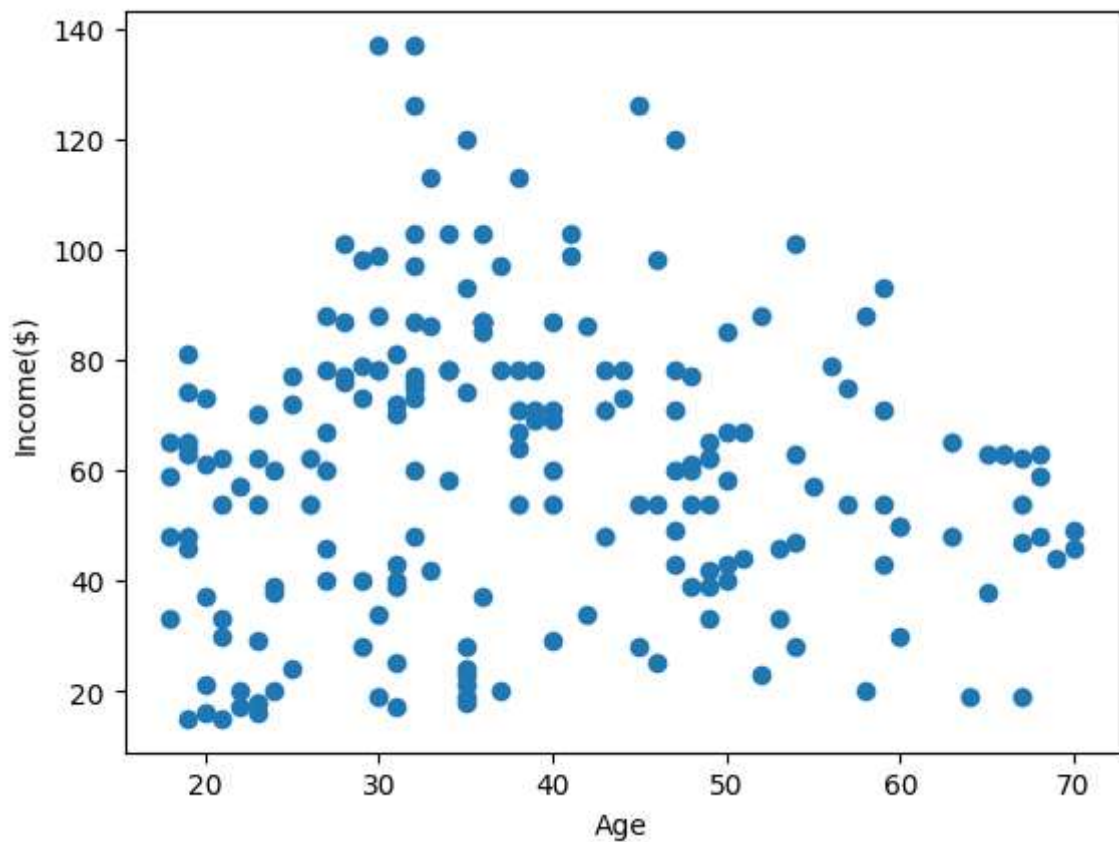
```
In [4]: df.tail()
```

```
Out[4]:
```

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

```
In [5]: plt.scatter(df["Age"],df["Income($)"])
plt.xlabel("Age")
plt.ylabel("Income($)")
```

```
Out[5]: Text(0, 0.5, 'Income($))')
```



```
In [6]: from sklearn.cluster import KMeans
km=KMeans()
km
```

```
Out[6]:
```

```
▼ KMeans
KMeans()
```

```
In [7]: y_predicted=km.fit_predict(df[["Age", "Income($)"]])
y_predicted
```

C:\Users\manasa\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[7]: array([0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 3, 0, 3, 0, 0, 0, 0, 0, 3, 0, 0, 0,
              3, 0, 3, 0, 3, 0, 0, 0, 3, 0, 3, 0, 3, 0, 3, 0, 0, 0, 3, 0, 3, 0,
              3, 0, 3, 0, 0, 0, 3, 0, 6, 3, 3, 3, 3, 4, 6, 3, 4, 6, 4, 3, 4, 6,
              3, 4, 6, 6, 4, 3, 4, 4, 4, 6, 5, 5, 6, 5, 4, 6, 4, 5, 6, 5, 4, 6,
              6, 5, 4, 6, 5, 5, 6, 6, 5, 6, 5, 6, 6, 5, 4, 6, 5, 6, 4, 5, 4, 4,
              4, 6, 5, 6, 6, 6, 4, 5, 5, 5, 6, 5, 5, 5, 2, 2, 5, 5, 5, 5, 5, 5,
              2, 2, 2, 2, 5, 2, 2, 2, 5, 2, 2, 2, 2, 2, 5, 2, 2, 2, 5, 2, 5, 2,
              5, 2, 2, 2, 2, 2, 5, 2, 2, 2, 1, 2, 1, 2, 2, 2, 1, 2, 2, 2, 1, 2,
              1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 7, 7, 7, 7, 7, 7,
              7, 7])
```

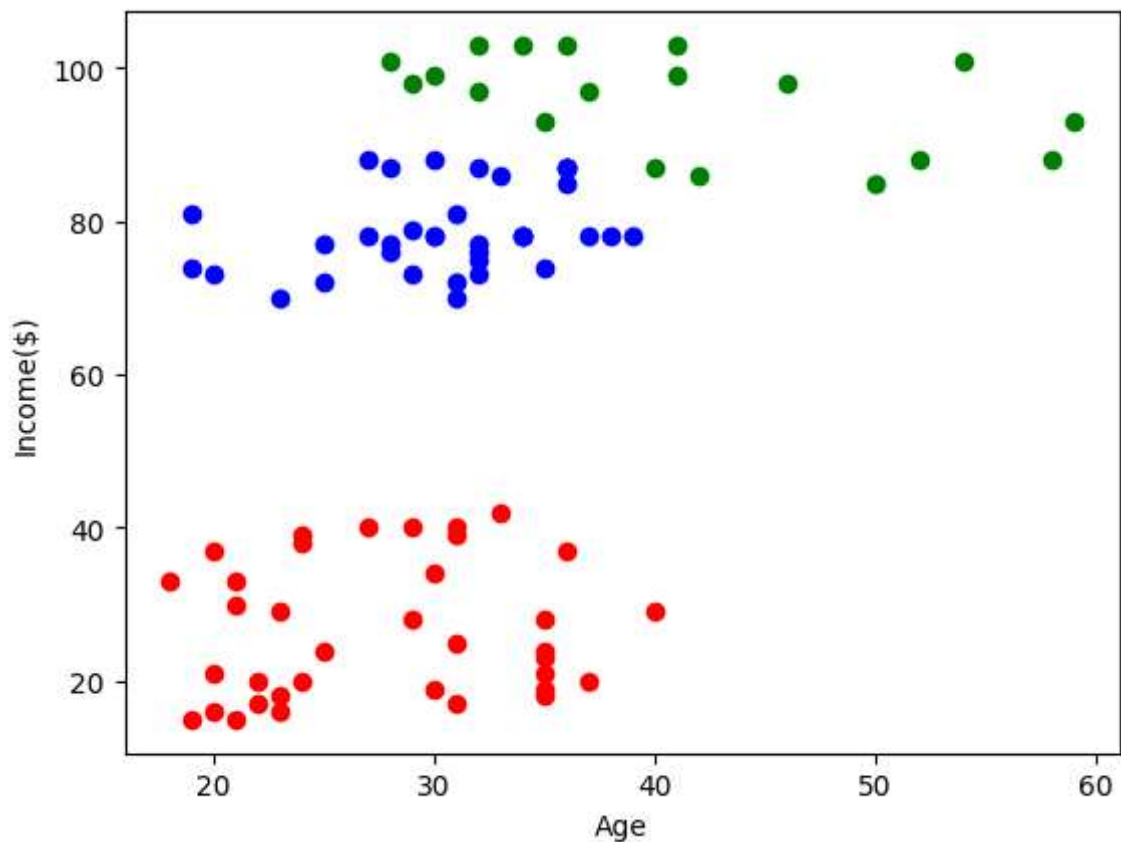
```
In [8]: df["cluster"]=y_predicted
df.head()
```

Out[8]:

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

```
In [10]: 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[10]: Text(0, 0.5, 'Income(\$))')



```
In [11]: from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
scaler.fit(df[["Income($)"]])
df["Income($)"]=scaler.transform(df[["Income($)"]])
df.head()
```

Out[11]:

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

```
In [12]: scaler.fit(df[["Age"]])
df["Age"]=scaler.transform(df[["Age"]])
df.head()
```

Out[12]:

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

```
In [13]: km=KMeans()
```

```
In [14]: y_predicted=km.fit_predict(df[["Age", "Income($)"]])
y_predicted
```

C:\Users\manasa\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[14]: array([7, 7, 7, 7, 2, 7, 2, 7, 1, 2, 1, 2, 6, 7, 2, 7, 2, 7, 6, 2, 2, 7, 6, 2, 6, 2, 2, 7, 1, 7, 6, 7, 6, 7, 6, 2, 2, 7, 1, 7, 6, 2, 6, 7, 6, 2, 2, 2, 6, 2, 2, 1, 6, 6, 6, 1, 5, 6, 1, 5, 1, 6, 1, 5, 6, 1, 5, 2, 1, 6, 1, 1, 1, 5, 6, 6, 5, 6, 1, 0, 1, 6, 5, 6, 3, 5, 0, 3, 1, 5, 3, 0, 0, 5, 3, 5, 3, 5, 5, 3, 1, 5, 3, 5, 1, 3, 1, 1, 1, 5, 0, 5, 5, 5, 1, 3, 3, 3, 5, 0, 0, 0, 5, 0, 3, 0, 3, 0, 3, 0, 5, 0, 5, 0, 3, 0, 5, 0, 3, 0, 0, 0, 5, 0, 3, 0, 0, 0, 3, 0, 3, 0, 3, 0, 0, 0, 0, 0, 3, 0, 5, 0, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 3, 0, 3, 0, 4, 4, 3, 4, 4, 4, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4])

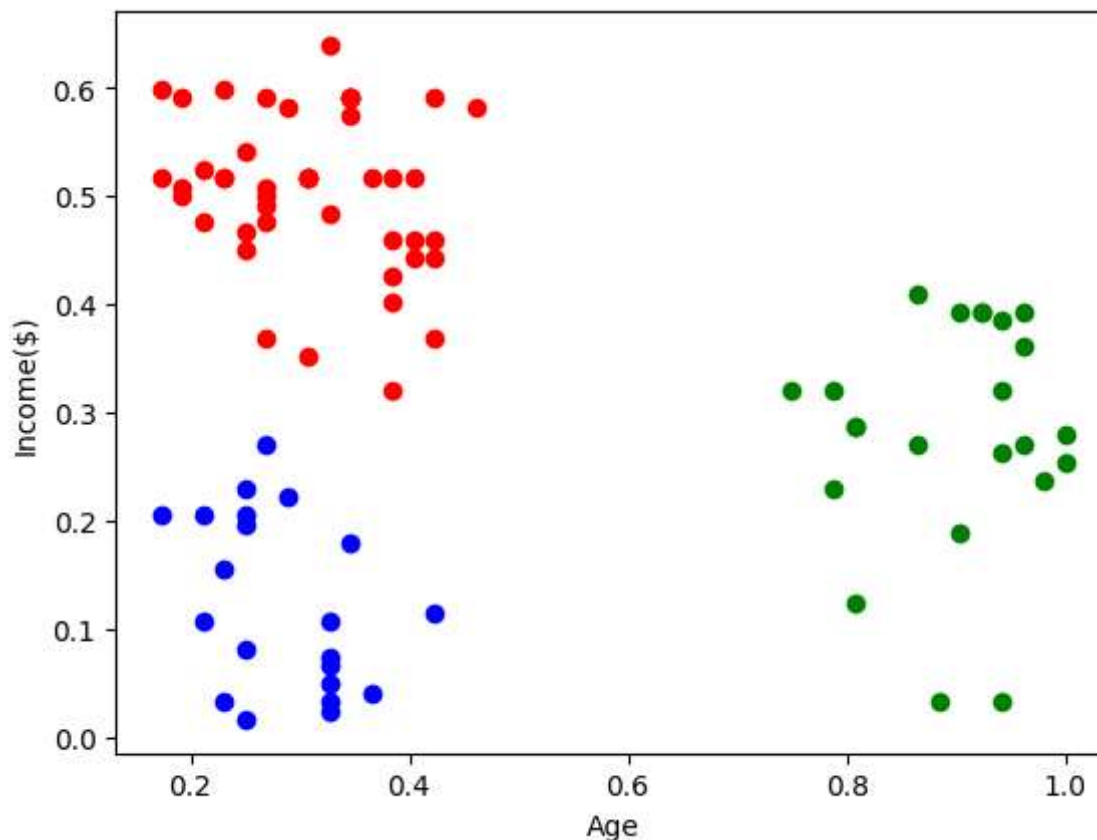
```
In [15]: df["New Cluster"]=y_predicted
df.head()
```

Out[15]:

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

```
In [16]: 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[16]: Text(0, 0.5, 'Income($)')
```

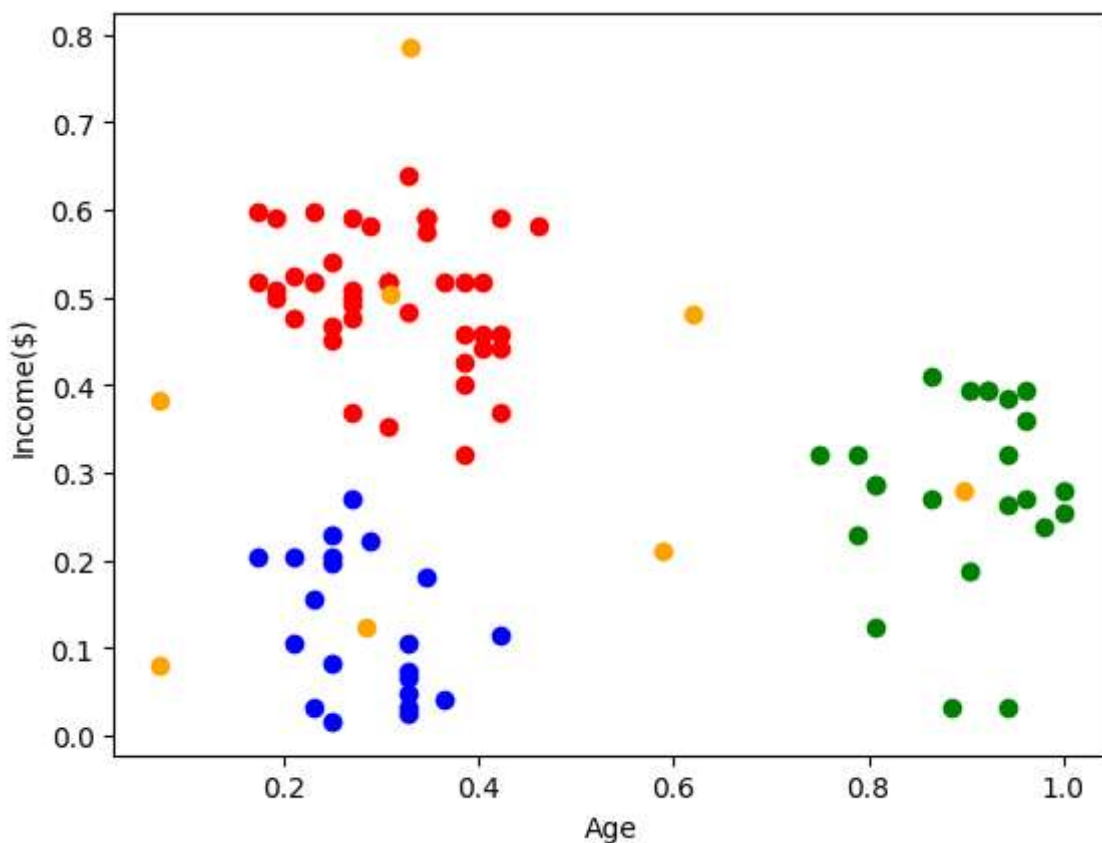


```
In [17]: km.cluster_centers_
```

```
Out[17]: array([[0.30944056, 0.50428465],
 [0.89799331, 0.28011404],
 [0.28388278, 0.1245121 ],
 [0.62037037, 0.47996357],
 [0.32905983, 0.78551913],
 [0.07322485, 0.38272383],
 [0.58974359, 0.20969945],
 [0.07239819, 0.08003857]])
```

```
In [19]: 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($)"],color="blue")
plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color="orange")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[19]: Text(0, 0.5, 'Income(\$))')



```
In [20]: k_rng=range(1,10)
sse=[]
```

```
In [25]: 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 square errorprint(sse)
plt.plot(k_rng, sse)
plt.xlabel("K")
plt.ylabel("Sum of Squared Error")
```

C:\Users\manasa\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\manasa\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\manasa\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\manasa\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\manasa\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\manasa\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\manasa\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(

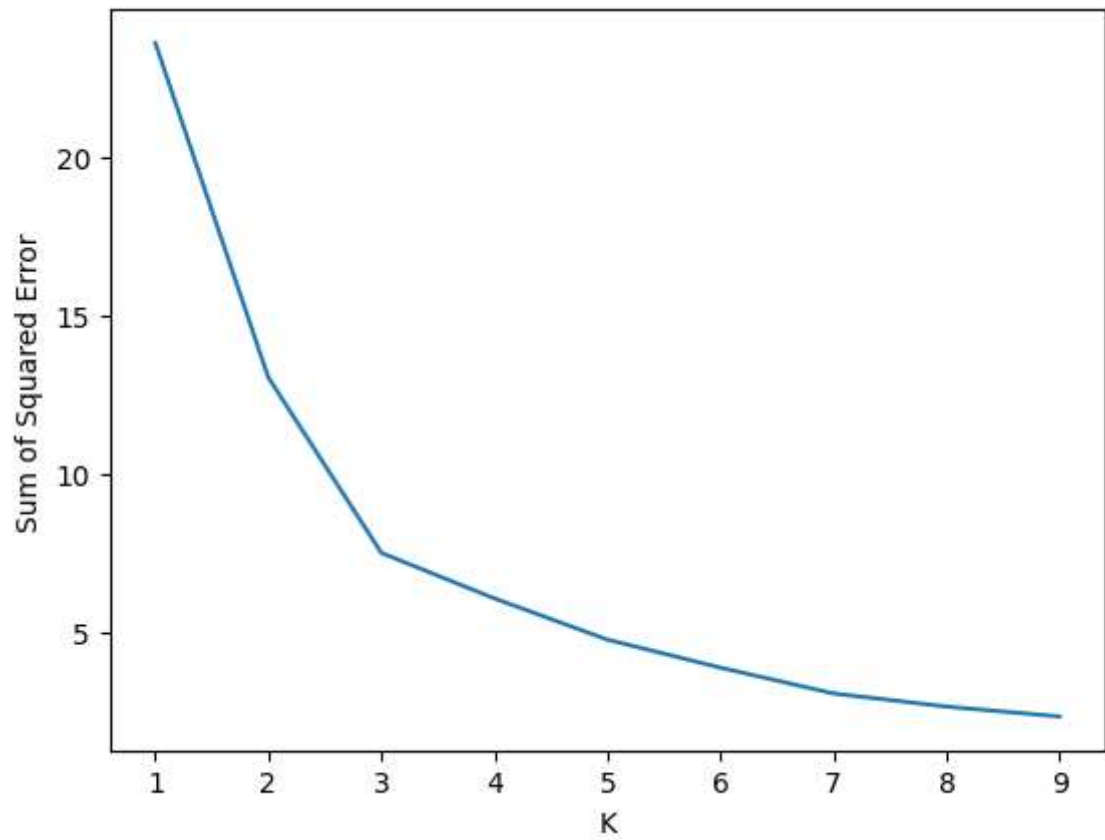
C:\Users\manasa\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning

warnings.warn(

C:\Users\manasa\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[25]: Text(0, 0.5, 'Sum of Squared Error')



In []: