

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

```
In [2]: df=pd.read_csv(r"C:\Users\Sudheer\AppData\Local\Microsoft\Windows\INetCache\IE
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

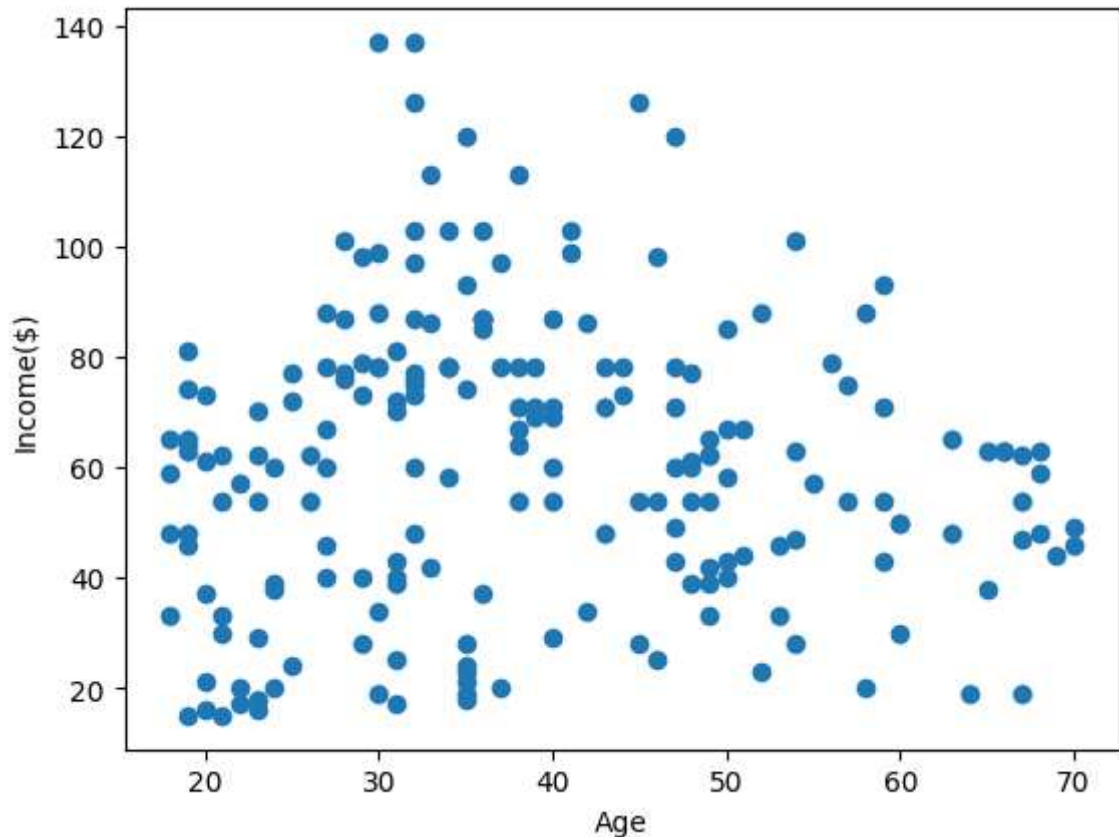
```
In [4]: df.tail()
```

```
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()
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.

On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

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

C:\Users\Sudheer\AppData\Local\Programs\Python\Python310\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([1, 1, 1, 1, 1, 1, 1, 1, 5, 1, 5, 1, 5, 1, 1, 1, 1, 1, 5, 1, 1, 1,
        5, 1, 5, 1, 5, 1, 5, 1, 5, 1, 5, 1, 5, 1, 5, 1, 5, 1, 5, 7, 5, 7,
        5, 7, 5, 7, 7, 7, 5, 7, 7, 5, 5, 5, 5, 4, 7, 5, 4, 7, 4, 5, 4, 7,
        5, 4, 7, 7, 4, 5, 4, 4, 4, 7, 0, 0, 7, 0, 4, 7, 4, 0, 7, 0, 4, 7,
        7, 0, 4, 7, 0, 0, 7, 7, 0, 7, 0, 7, 7, 0, 4, 7, 0, 7, 4, 0, 4, 4,
        4, 7, 0, 7, 7, 7, 4, 0, 0, 0, 3, 0, 0, 0, 3, 3, 0, 0, 0, 0, 0, 0,
        3, 3, 3, 3, 0, 3, 3, 3, 0, 3, 3, 3, 3, 3, 0, 3, 3, 3, 0, 3, 0, 3,
        0, 3, 3, 3, 3, 3, 0, 3, 3, 3, 6, 3, 6, 3, 6, 3, 6, 3, 6, 6, 6, 3,
        6, 3, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 2, 2, 2, 2, 2, 2,
        2, 2])
```

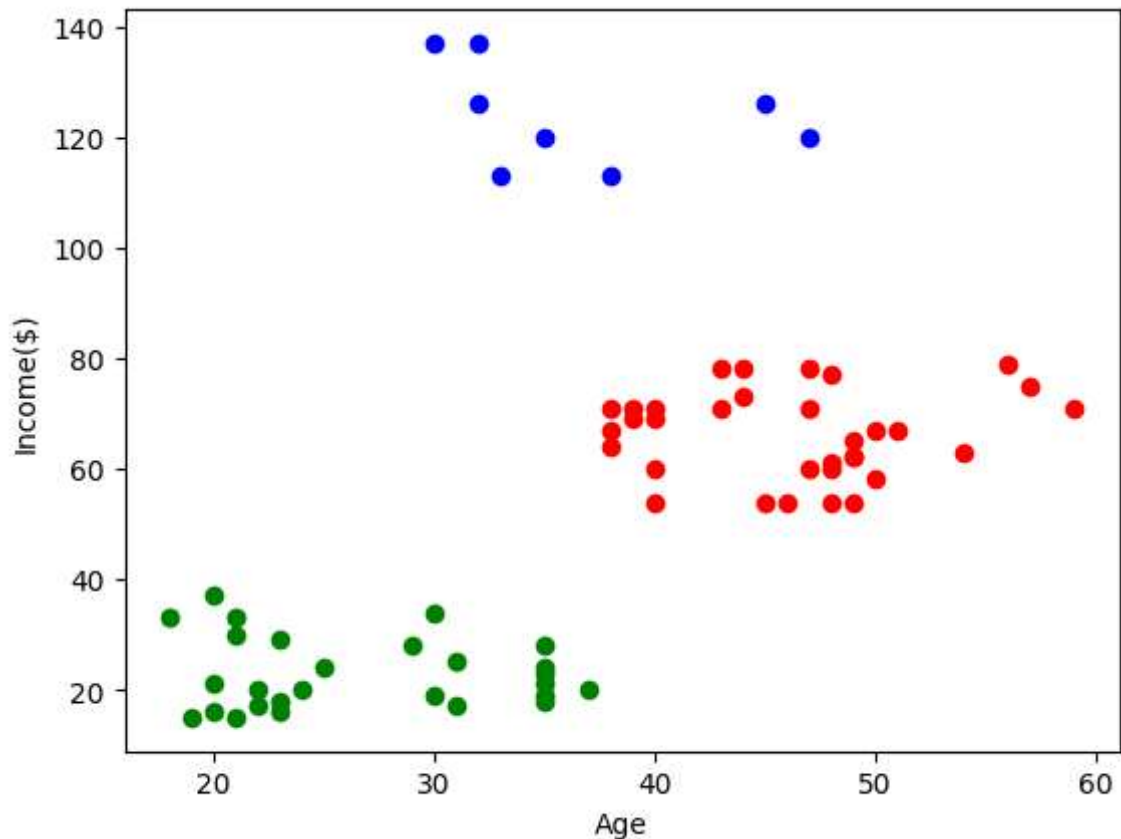
```
In [8]: df["cluster"]=y_predicted
df.head()
```

```
Out[8]:
```

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

```
In [9]: 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[9]: Text(0, 0.5, 'Income(\$))')



```
In [10]: from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
scaler.fit(df[["Income($)"]])
df["Income($)"]=scaler.transform(df[["Income($)"]])
df.head()
```

Out[10]:

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

```
In [11]: scaler.fit(df[["Age"]])
df["Age"]=scaler.transform(df[["Age"]])
df.head()
```

```
Out[11]:
```

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

```
In [12]: km=KMeans()
```

```
In [13]: y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

C:\Users\Sudheer\AppData\Local\Programs\Python\Python310\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[13]: array([5, 5, 5, 5, 2, 5, 2, 5, 4, 2, 4, 2, 0, 5, 2, 5, 2, 5, 0, 2, 2, 5,
                0, 2, 0, 2, 0, 2, 2, 5, 4, 5, 0, 5, 0, 5, 0, 2, 2, 5, 4, 5, 0, 2,
                0, 5, 0, 2, 2, 2, 0, 2, 2, 4, 0, 0, 0, 4, 1, 0, 4, 1, 4, 0, 4, 1,
                0, 4, 1, 2, 4, 0, 4, 4, 4, 1, 0, 0, 1, 0, 4, 7, 4, 0, 1, 0, 3, 1,
                7, 3, 4, 1, 3, 7, 7, 1, 3, 1, 3, 1, 1, 3, 4, 1, 3, 1, 4, 3, 4, 4,
                4, 1, 7, 1, 1, 1, 4, 3, 3, 3, 1, 7, 7, 7, 1, 7, 3, 7, 3, 7, 3, 7,
                1, 7, 1, 7, 3, 7, 1, 7, 3, 7, 7, 7, 1, 7, 3, 7, 7, 7, 3, 7, 3, 7,
                3, 7, 7, 7, 7, 7, 3, 7, 1, 7, 3, 7, 7, 7, 7, 7, 7, 7, 7, 3, 7,
                3, 7, 3, 7, 6, 6, 3, 6, 6, 6, 3, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,
                6, 6])
```

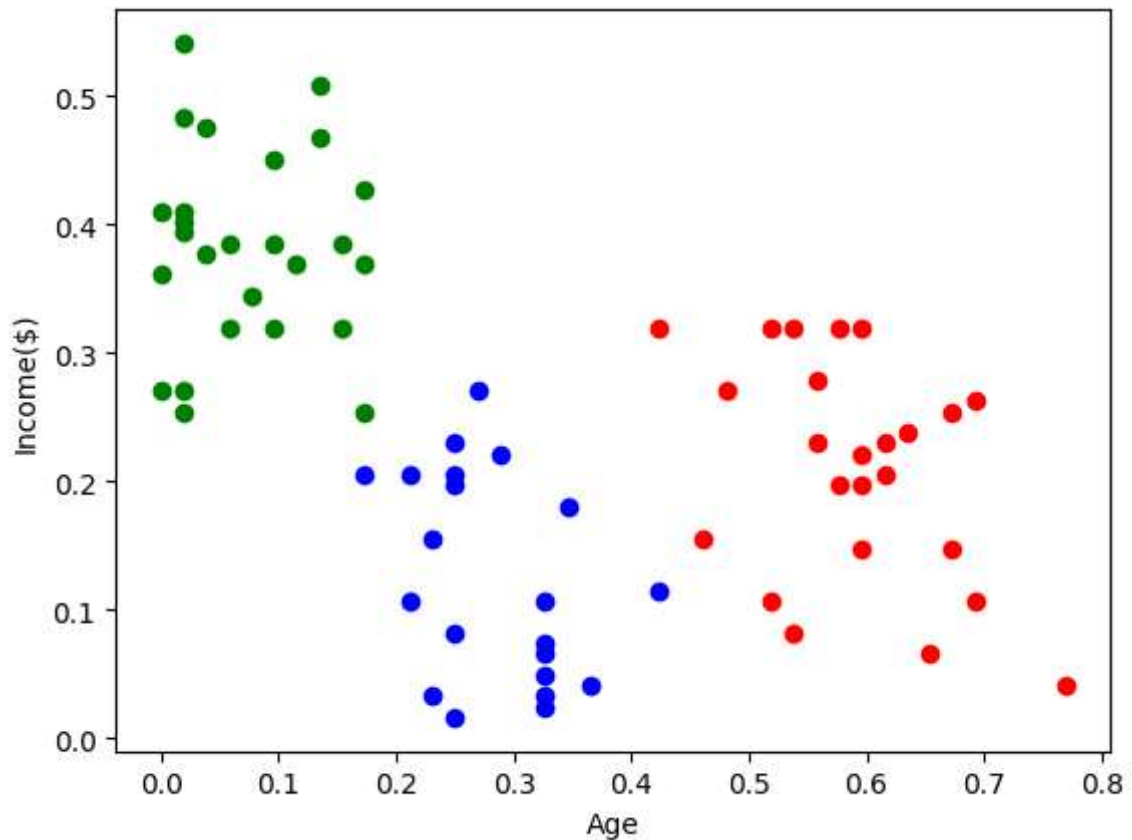
```
In [14]: df["New Cluster"]=y_predicted
df.head()
```

```
Out[14]:
```

	Gender	Age	Income(\$)	cluster	New Cluster
0	Male	0.019231	0.000000	1	5
1	Male	0.057692	0.000000	1	5
2	Female	0.038462	0.008197	1	5
3	Female	0.096154	0.008197	1	5
4	Female	0.250000	0.016393	1	2

```
In [15]: 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[15]: Text(0, 0.5, 'Income($)')
```



```
In [16]: km.cluster_centers_
```

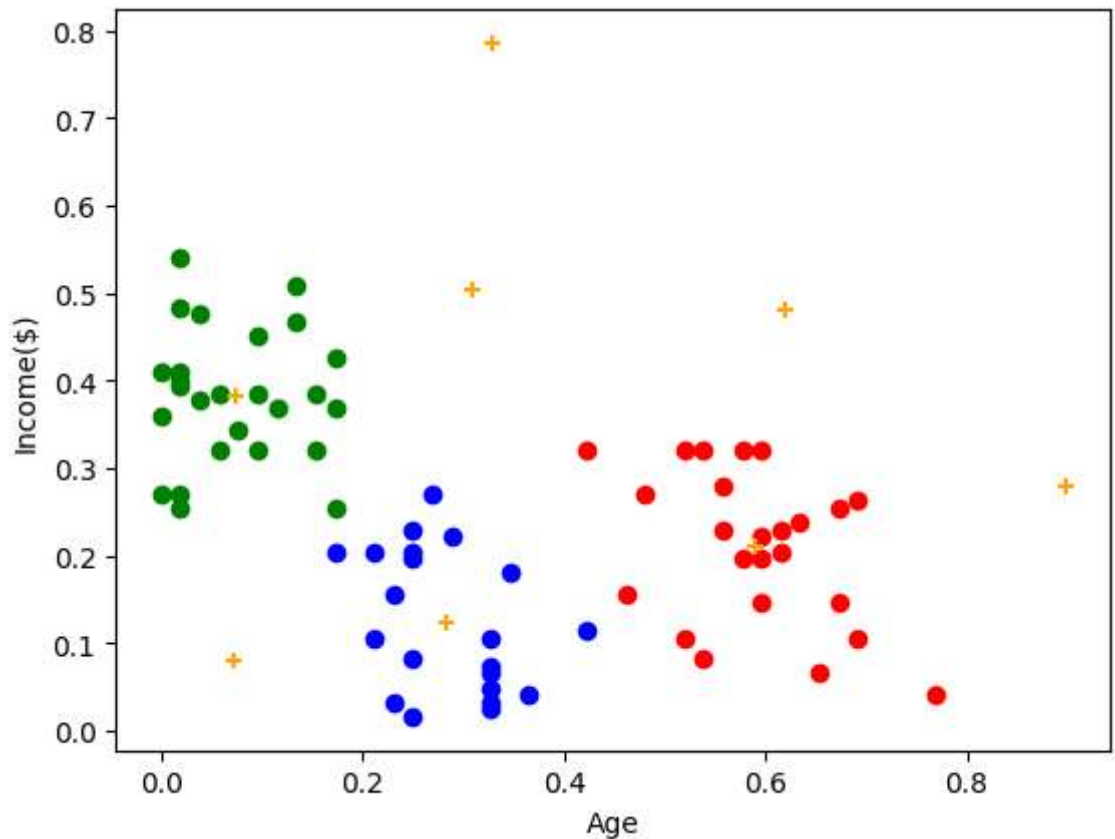
```
Out[16]: array([[0.58974359, 0.20969945],
 [0.07322485, 0.38272383],
 [0.28388278, 0.1245121 ],
 [0.62037037, 0.47996357],
 [0.89799331, 0.28011404],
 [0.07239819, 0.08003857],
 [0.32905983, 0.78551913],
 [0.30944056, 0.50428465]])
```

```

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($)"],color="blue")
plt.scatter(km.cluster_centers_[0],km.cluster_centers_[1],color="orange",marker="+")
plt.xlabel("Age")
plt.ylabel("Income($)")

```

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



```

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

```

```
In [22]: 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 error
    print(sse)
    plt.plot(k_rng, sse)
    plt.xlabel("K")
    plt.ylabel("Sum of Squared Error")
```

C:\Users\Sudheer\AppData\Local\Programs\Python\Python310\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\Sudheer\AppData\Local\Programs\Python\Python310\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\Sudheer\AppData\Local\Programs\Python\Python310\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\Sudheer\AppData\Local\Programs\Python\Python310\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\Sudheer\AppData\Local\Programs\Python\Python310\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\Sudheer\AppData\Local\Programs\Python\Python310\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\Sudheer\AppData\Local\Programs\Python\Python310\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\Sudheer\AppData\Local\Programs\Python\Python310\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

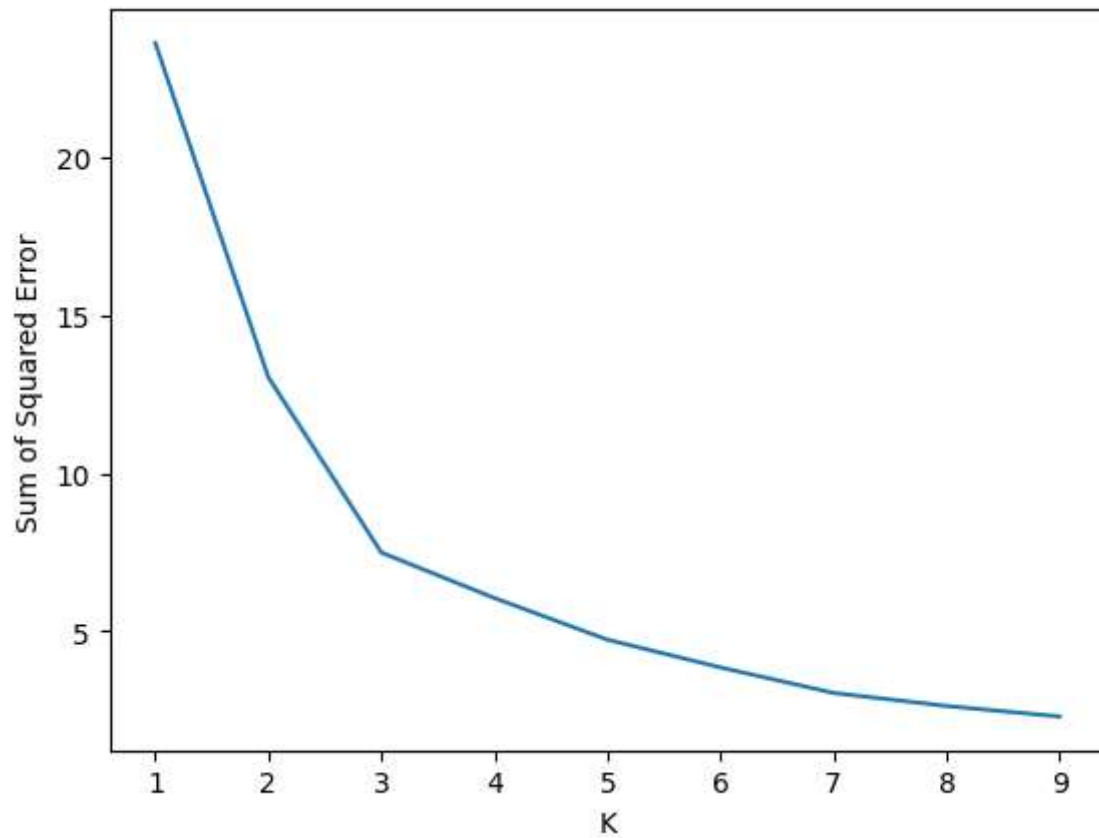
warnings.warn(

C:\Users\Sudheer\AppData\Local\Programs\Python\Python310\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, 13.028938428018291, 7.4921134132374565, 6.055858644812546, 4.7389183247905855, 3.863114332829242, 3.0580611070789887, 2.646037617631439, 2.313572035354328]
```

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



In []: