KMeans Cluster

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

```
In [2]:   1  df=pd.read_csv(r"C:\Users\Sushma sree\Downloads\Income.csv")
          2  df
```
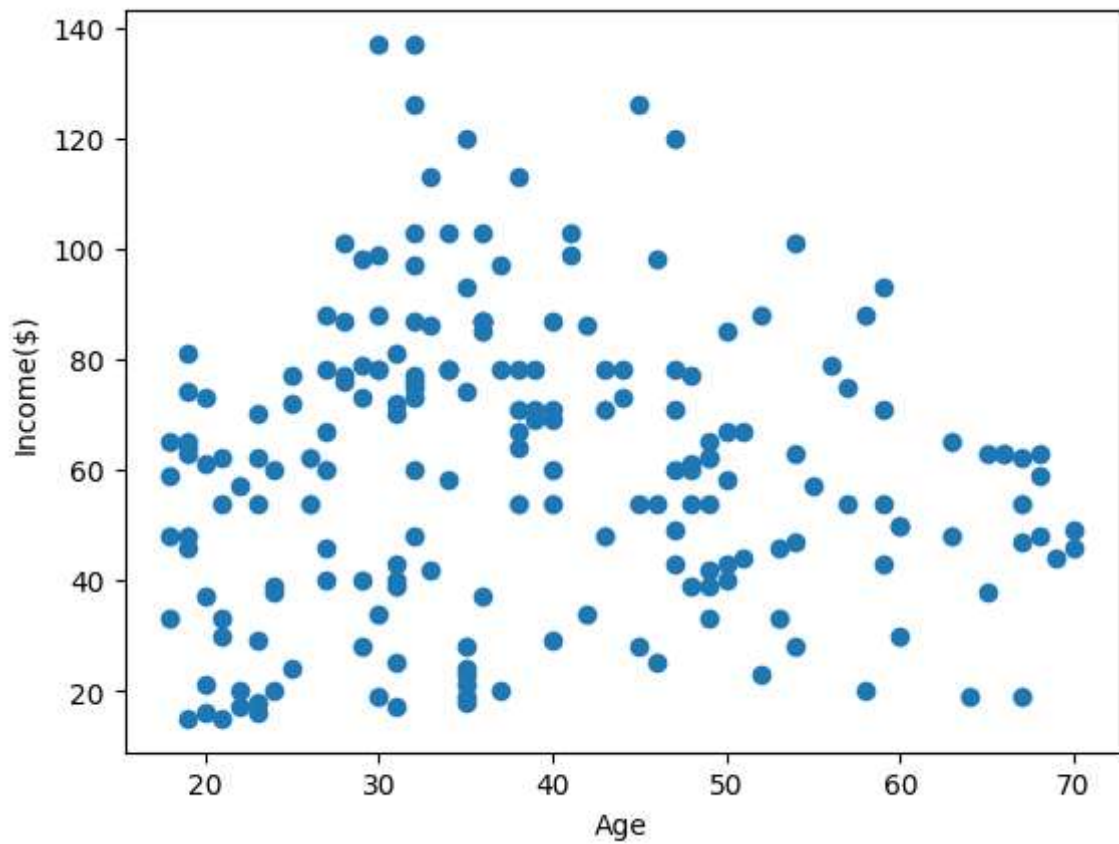
Out[2]:

	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]:   1  plt.scatter(df['Age'],df['Income($)'])
          2  plt.xlabel('Age')
          3  plt.ylabel('Income($)')
```

Out[3]: Text(0, 0.5, 'Income($)')


```
In [4]:   1  from sklearn.cluster import KMeans
```

```
In [5]:   1  km=KMeans()
```

```
In [6]:   1  y_pred=km.fit_predict(df[['Age','Income($)']])
          2  y_pred
```

C:\Users\Sushma sree\AppData\Local\Programs\Python\Python310\lib\site-package
s\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[6]: array([7, 7, 7, 7, 7, 7, 7, 7, 0, 7, 0, 7, 0, 7, 7, 7, 7, 7, 0, 7, 7, 7,
               0, 7, 0, 7, 0, 7, 0, 7, 0, 7, 0, 2, 0, 2, 0, 2, 2, 2, 0, 2, 0, 2,
               0, 2, 0, 2, 2, 2, 0, 2, 2, 0, 0, 0, 0, 5, 2, 0, 5, 2, 5, 0, 5, 2,
               0, 5, 2, 2, 5, 0, 5, 5, 5, 2, 6, 6, 2, 6, 5, 6, 5, 6, 2, 6, 5, 4,
               6, 6, 5, 4, 6, 6, 4, 4, 6, 4, 6, 4, 4, 6, 5, 4, 6, 4, 5, 6, 5, 5,
               5, 4, 6, 4, 4, 4, 5, 6, 6, 6, 4, 6, 6, 6, 4, 4, 6, 6, 6, 6, 6, 6,
               4, 4, 4, 4, 6, 4, 4, 4, 6, 4, 4, 4, 4, 4, 6, 4, 4, 4, 6, 6, 6, 4,
               6, 4, 4, 4, 4, 4, 6, 4, 4, 4, 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, 3, 3, 3, 3, 3, 3,
               3, 3])
```
In [7]:   1  df['cluster']=y_pred
          2  df.head()
```
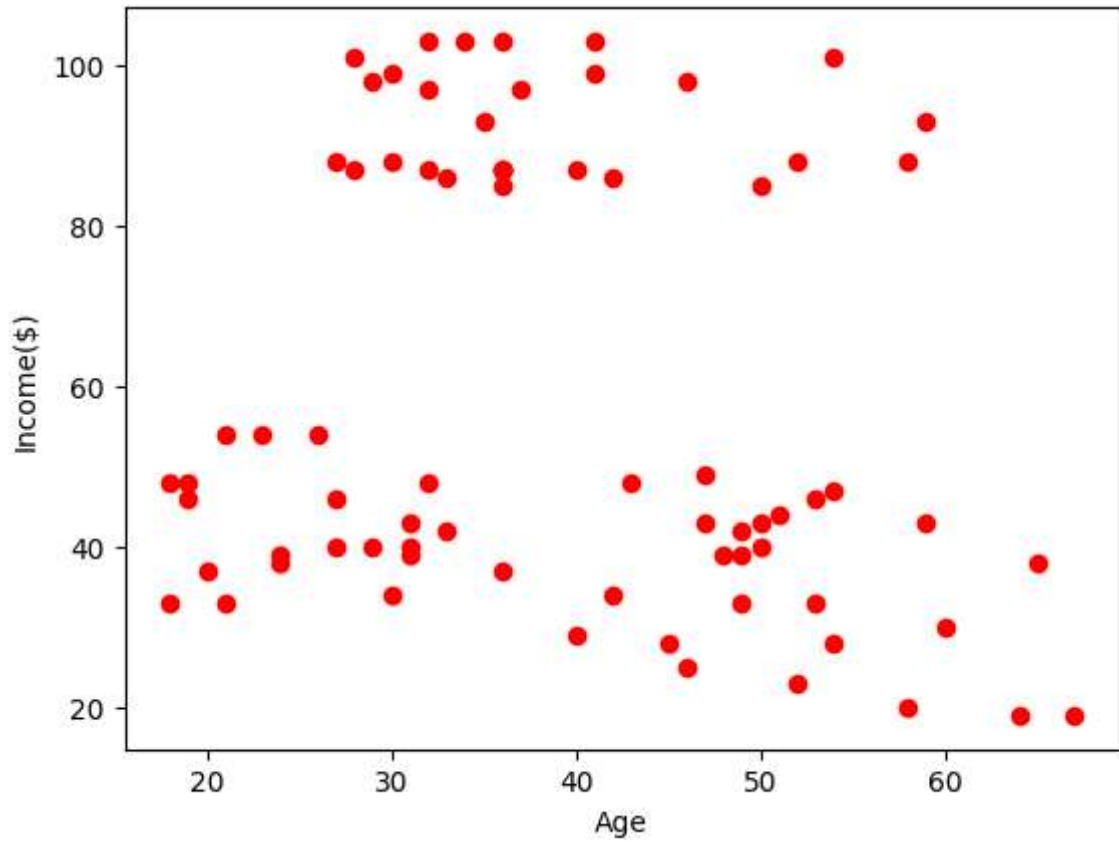
Out[7]:

	Gender	Age	Income($)	cluster
0	Male	19	15	7
1	Male	21	15	7
2	Female	20	16	7
3	Female	23	16	7
4	Female	31	17	7

```
In [8]:  1  df1=df[df.cluster==0]
         2  df2=df[df.cluster==1]
         3  df3=df[df.cluster==2]
         4  plt.scatter(df1['Age'],df1['Income($)'],color='red')
         5  plt.scatter(df2['Age'],df2['Income($)'],color='red')
         6  plt.scatter(df3['Age'],df3['Income($)'],color='red')
         7  plt.xlabel('Age')
         8  plt.ylabel('Income($)')
```

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


```
In [9]:   1  from sklearn.preprocessing import MinMaxScaler
```

```
In [10]:  1  scaler=MinMaxScaler()
```

```
In [11]:   1  scaler.fit(df[["Income($)"]])
           2  df['Income($)']=scaler.transform(df[['Income($)']])
           3  df.head()
```

Out[11]:

	Gender	Age	Income($)	cluster
0	Male	19	0.000000	7
1	Male	21	0.000000	7
2	Female	20	0.008197	7
3	Female	23	0.008197	7
4	Female	31	0.016393	7

```
In [12]:   1  scaler.fit(df[['Age']])
           2  df['Age']=scaler.transform(df[['Age']])
           3  df.head()
```

Out[12]:

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

```
In [13]:   1  km=KMeans()
```

```
In [14]:   1  y_predicted=km.fit_predict(df[['Age','Income($)']])
           2  y_predicted
```

C:\Users\Sushma sree\AppData\Local\Programs\Python\Python310\lib\site-package
s\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([2, 2, 2, 2, 6, 2, 6, 2, 4, 6, 4, 6, 1, 2, 6, 2, 6, 2, 1, 6, 6, 2,
         1, 6, 1, 6, 1, 6, 6, 2, 4, 2, 1, 2, 1, 2, 1, 6, 6, 2, 4, 2, 1, 6,
         1, 2, 1, 6, 6, 6, 1, 6, 6, 4, 1, 1, 1, 4, 7, 1, 4, 7, 4, 1, 4, 7,
         1, 4, 7, 6, 4, 1, 4, 4, 4, 7, 1, 1, 7, 1, 4, 0, 4, 1, 7, 1, 1, 7,
         0, 1, 4, 7, 1, 0, 0, 7, 1, 7, 3, 7, 7, 3, 4, 7, 3, 7, 4, 3, 4, 4,
         4, 7, 0, 7, 7, 7, 4, 3, 3, 3, 7, 0, 0, 0, 7, 0, 3, 0, 3, 0, 3, 0,
         7, 0, 7, 0, 3, 0, 7, 0, 3, 0, 0, 0, 7, 0, 3, 0, 0, 0, 3, 0, 3, 0,
         3, 0, 0, 0, 0, 0, 3, 0, 7, 0, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0,
         3, 0, 3, 5, 5, 5, 3, 5, 5, 5, 3, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5,
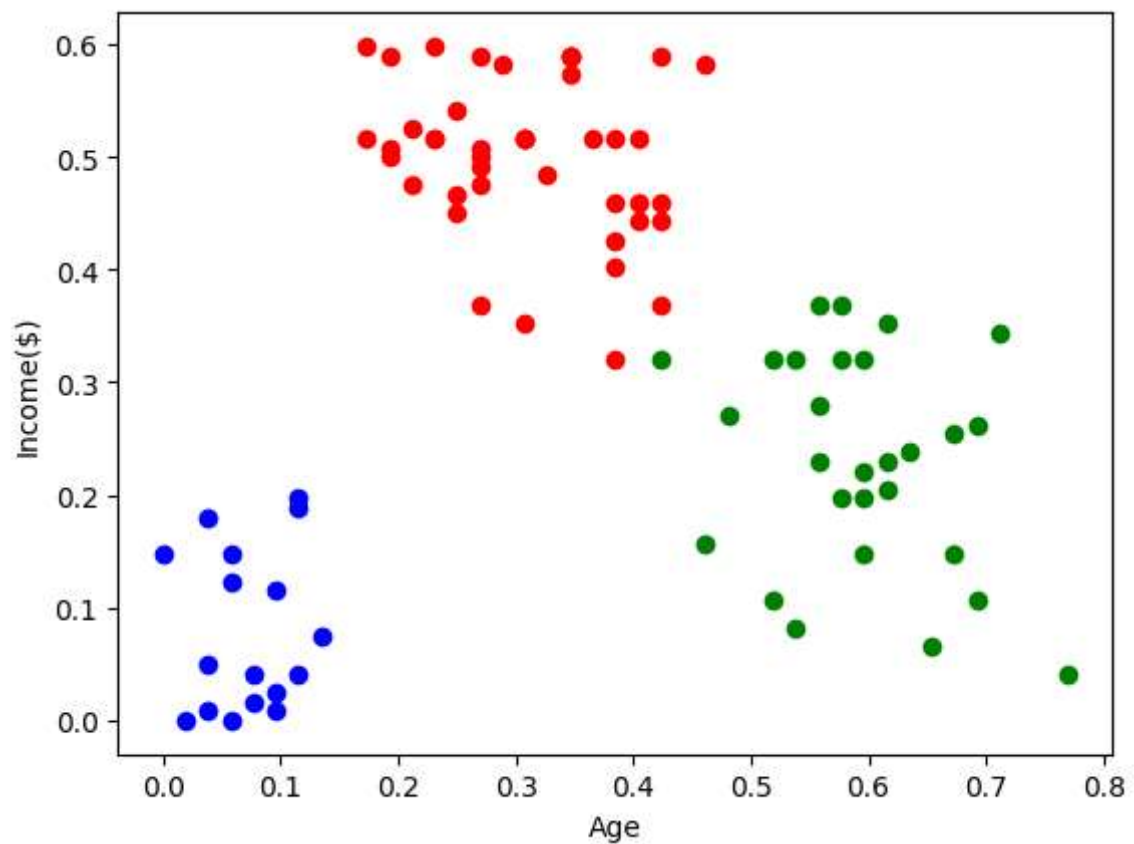   5, 5])
```

```
In [15]:   1  df['New cluster']=y_predicted
           2  df.head()
```

Out[15]:

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

```
In [16]:   1  df1=df[df['New cluster']==0]
           2  df2=df[df['New cluster']==1]
           3  df3=df[df['New cluster']==2]
           4  plt.scatter(df1['Age'],df1['Income($)'],color='red')
           5  plt.scatter(df2['Age'],df2['Income($)'],color='green')
           6  plt.scatter(df3['Age'],df3['Income($)'],color='blue')
           7  plt.xlabel('Age')
           8  plt.ylabel('Income($)')
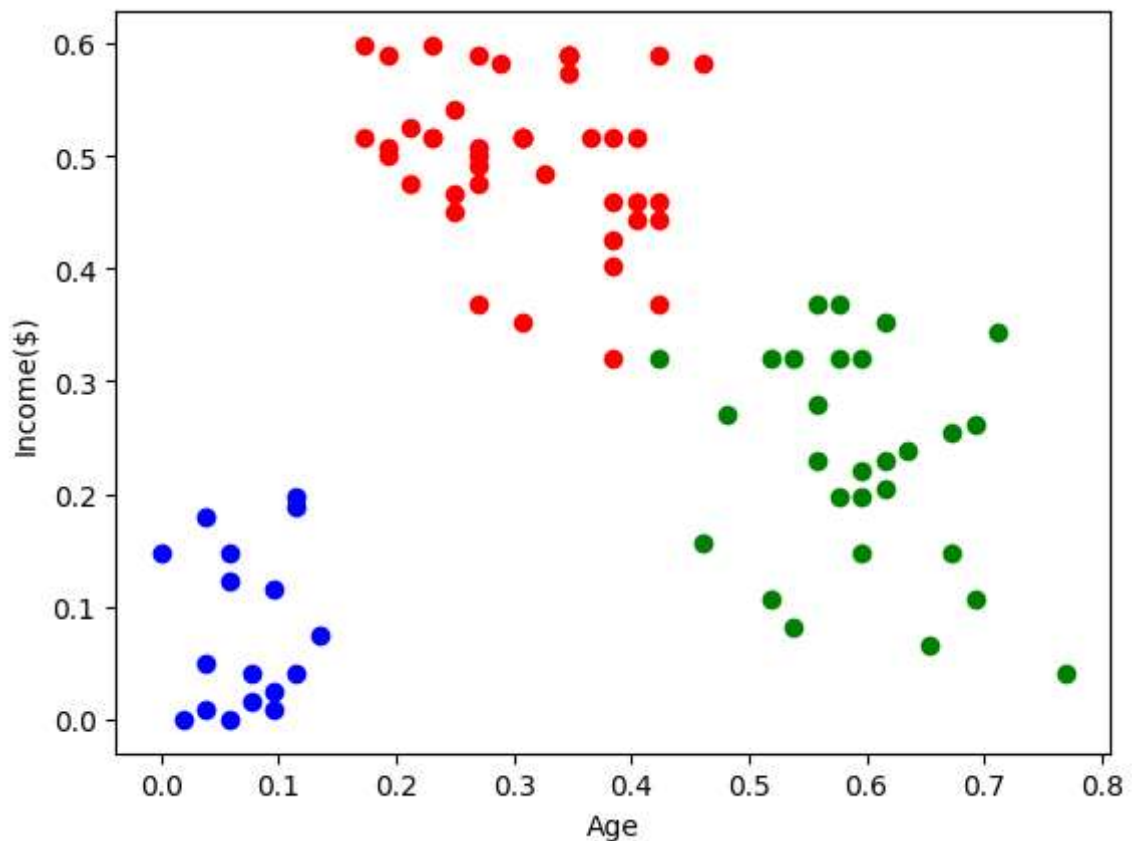```

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

```
In [17]:   1  km.cluster_centers_
```

Out[17]: array([[0.30903399, 0.50114373],
                 [0.59340659, 0.2309719],
                 [0.07239819, 0.08003857],
                 [0.62123746, 0.50106914],
                 [0.89799331, 0.28011404],
                 [0.32894737, 0.77782571],
                 [0.28388278, 0.1245121],
                 [0.07322485, 0.38272383]])

```
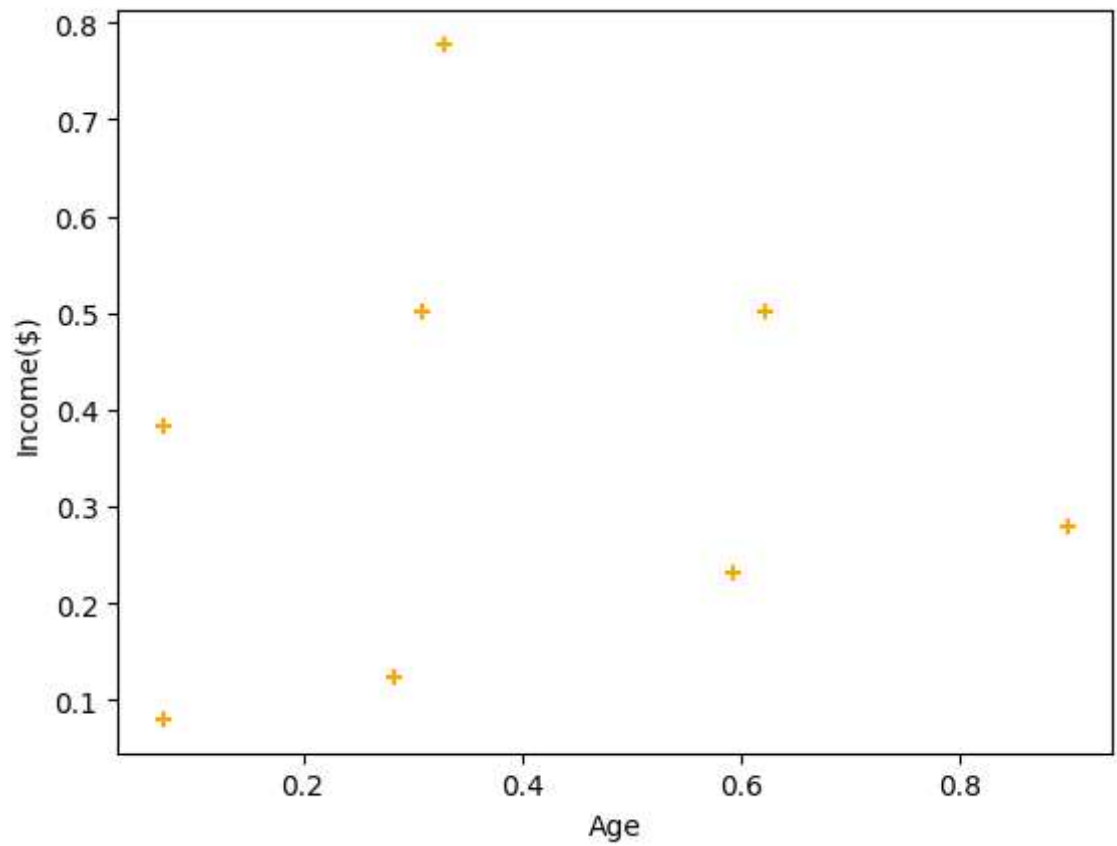In [18]:   1  df1=df[df['New cluster']==0]
           2  df2=df[df['New cluster']==1]
           3  df3=df[df['New cluster']==2]
           4  plt.scatter(df1['Age'],df1['Income($)'],color='red')
           5  plt.scatter(df2['Age'],df2['Income($)'],color='green')
           6  plt.scatter(df3['Age'],df3['Income($)'],color='blue')
           7  plt.xlabel('Age')
           8  plt.ylabel('Income($)')
```

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

```
1  plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color="oran
2  plt.xlabel('Age')
3  plt.ylabel('Income($)')
```

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

In [20]:
```python
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_)
sse
```

C:\Users\Sushma sree\AppData\Local\Programs\Python\Python310\lib\site-package
s\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\Sushma sree\AppData\Local\Programs\Python\Python310\lib\site-package
s\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\Sushma sree\AppData\Local\Programs\Python\Python310\lib\site-package
s\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\Sushma sree\AppData\Local\Programs\Python\Python310\lib\site-package
s\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\Sushma sree\AppData\Local\Programs\Python\Python310\lib\site-package
s\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\Sushma sree\AppData\Local\Programs\Python\Python310\lib\site-package
s\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\Sushma sree\AppData\Local\Programs\Python\Python310\lib\site-package
s\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\Sushma sree\AppData\Local\Programs\Python\Python310\lib\site-package
s\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\Sushma sree\AppData\Local\Programs\Python\Python310\lib\site-package
s\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\Sushma sree\AppData\Local\Programs\Python\Python310\lib\site-package
s\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[20]: [23.583906150363603,
          13.02893842801829,
          7.49302484330499,
          6.055858644812547,
          4.733776701093291,
          3.8584217212358616,
          3.054717436369359,
          2.643962722472653,
          2.3135720353543285]

In [22]:
```
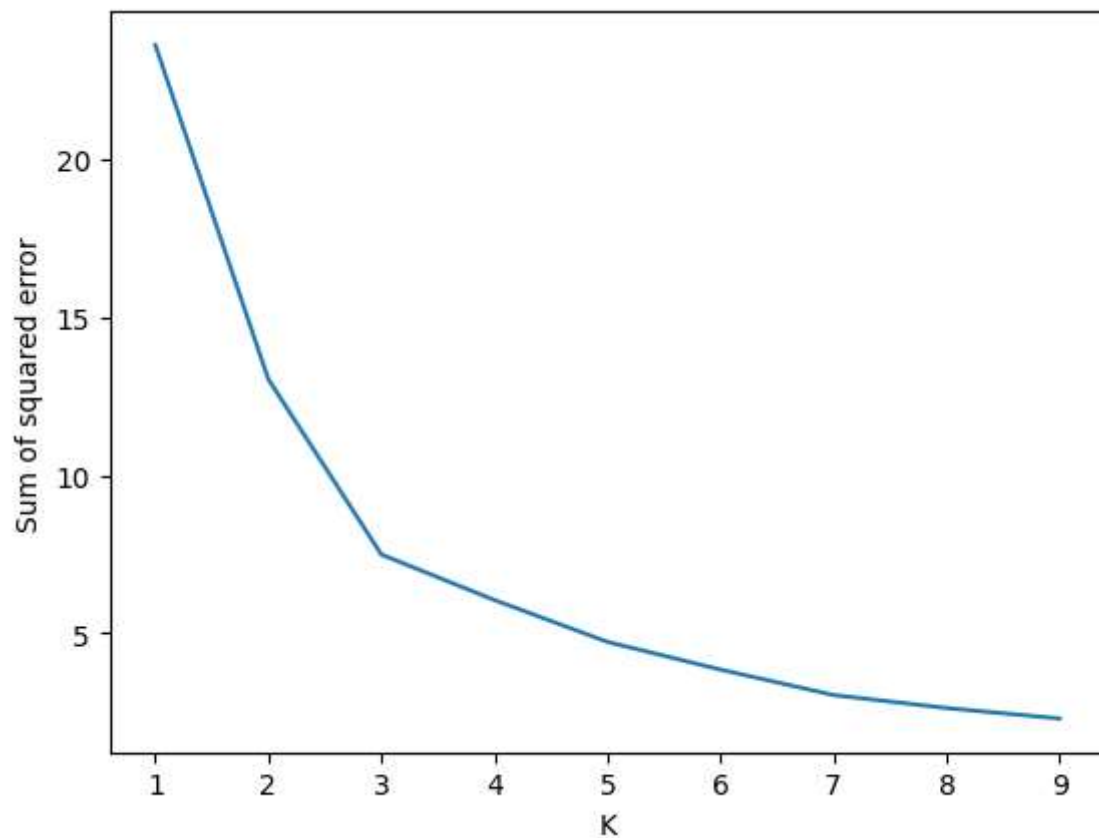1  plt.plot(k_rng,sse)
2  plt.xlabel('K')
3  plt.ylabel("Sum of squared error")
```

Out[22]: Text(0, 0.5, 'Sum of squared error')

In [ ]:
```
1
```