

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

```
In [2]: df=pd.read_csv(r"C:\Users\dinesh reddy\OneDrive\Documents\Income.csv")
```

```
In [3]: df
```

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
...
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 [4]: df.head()
```

Out[4]:

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

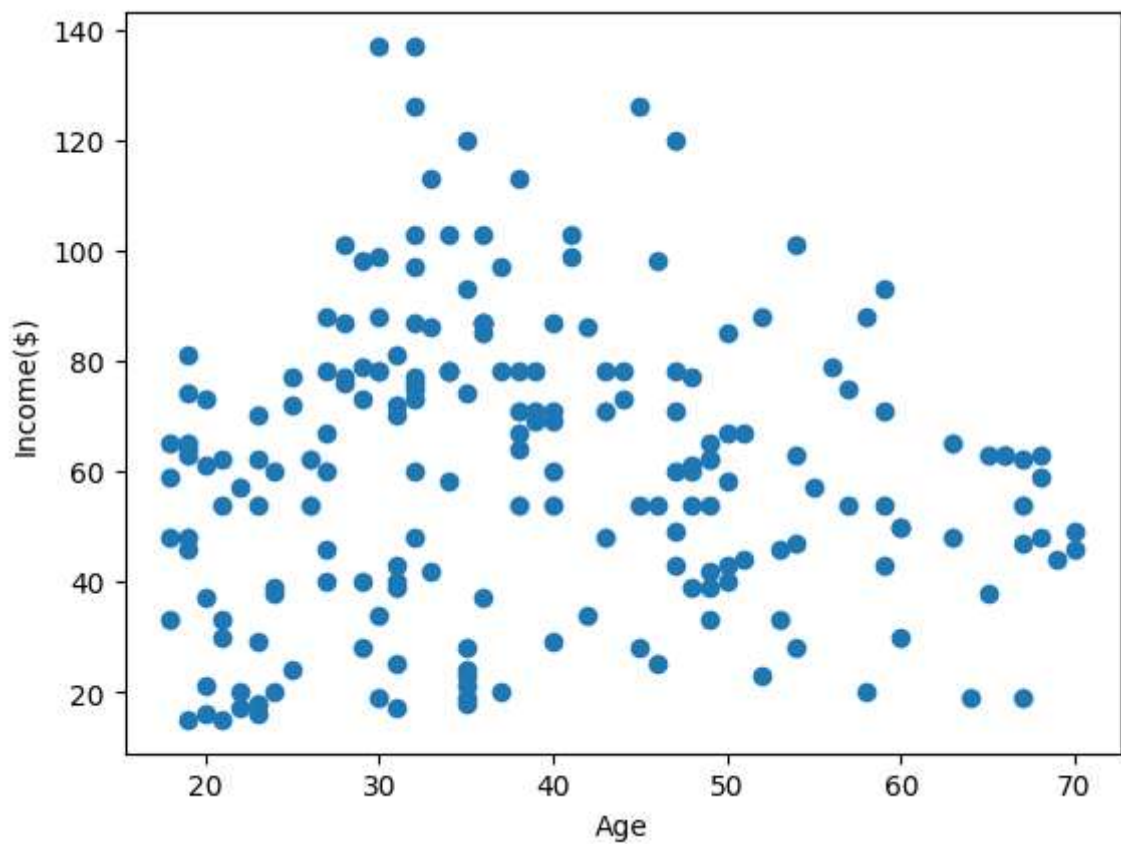
```
In [5]: df.tail()
```

```
Out[5]:
```

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

```
In [6]: plt.scatter(df["Age"],df["Income($)"])  
plt.xlabel("Age")  
plt.ylabel("Income($)")
```

```
Out[6]: Text(0, 0.5, 'Income($)')
```



```
In [7]: from sklearn.cluster import KMeans  
km=KMeans()  
km
```

```
Out[7]:
```

```
▼ KMeans  
KMeans()
```

```
In [8]: y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

```
C:\Users\dinesh reddy\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[8]: array([3, 3, 3, 3, 3, 3, 3, 3, 3, 1, 3, 1, 3, 1, 3, 3, 3, 3, 3, 1, 3, 3, 3,
              1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 6, 1, 6, 1, 6, 6, 6, 1, 6, 1, 6,
              1, 6, 1, 6, 6, 6, 1, 6, 6, 1, 1, 1, 1, 7, 6, 1, 7, 6, 7, 1, 7, 6,
              1, 7, 6, 6, 7, 1, 7, 7, 7, 6, 4, 4, 6, 4, 7, 4, 7, 4, 6, 4, 7, 0,
              4, 4, 7, 0, 4, 4, 0, 0, 4, 0, 4, 0, 0, 4, 7, 0, 4, 0, 7, 4, 7, 7,
              7, 0, 4, 0, 0, 0, 7, 4, 4, 4, 0, 4, 4, 4, 0, 0, 4, 4, 4, 4, 4,
              0, 0, 0, 0, 4, 0, 0, 0, 4, 0, 0, 0, 0, 0, 4, 0, 0, 0, 4, 4, 4, 0,
              4, 0, 0, 0, 0, 0, 4, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
              2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 5, 5, 5, 5, 5, 5,
              5, 5])
```

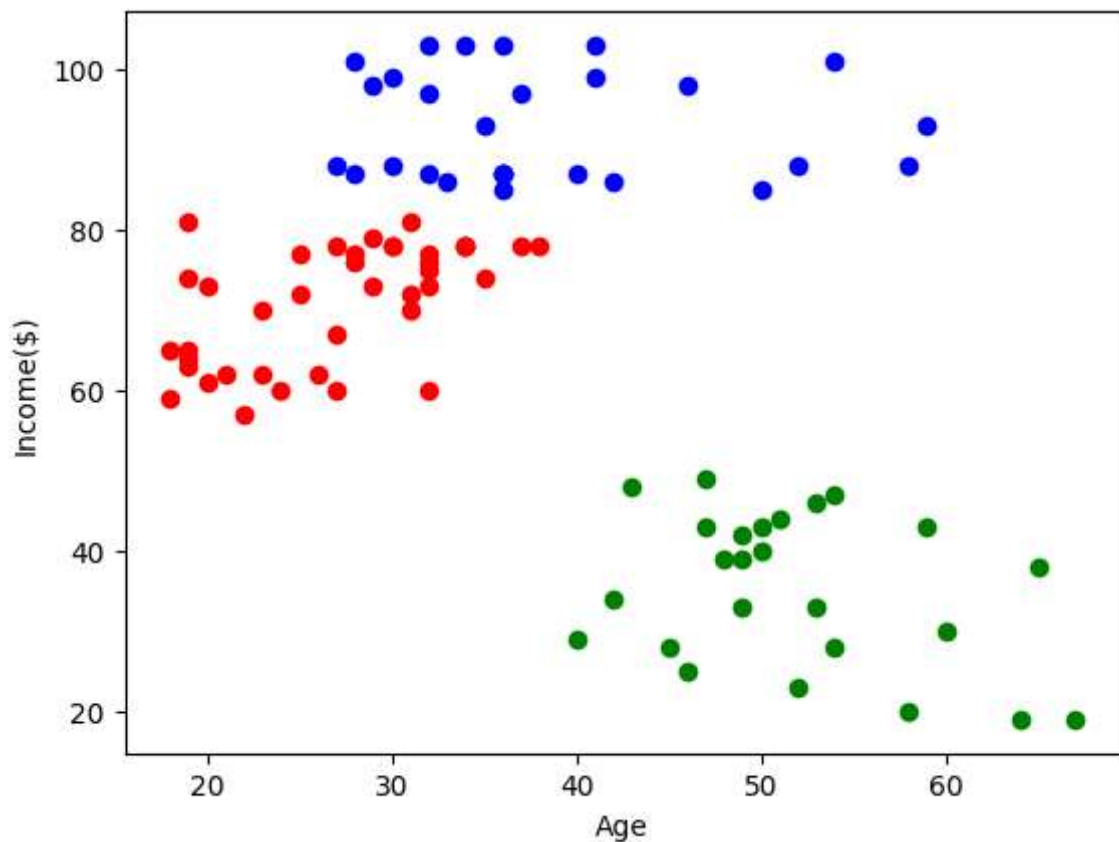
```
In [9]: df["cluster"]=y_predicted
df.head()
```

Out[9]:

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

```
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	3
1	Male	21	0.000000	3
2	Female	20	0.008197	3
3	Female	23	0.008197	3
4	Female	31	0.016393	3

```
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	3
1	Male	0.057692	0.000000	3
2	Female	0.038462	0.008197	3
3	Female	0.096154	0.008197	3
4	Female	0.250000	0.016393	3

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

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

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

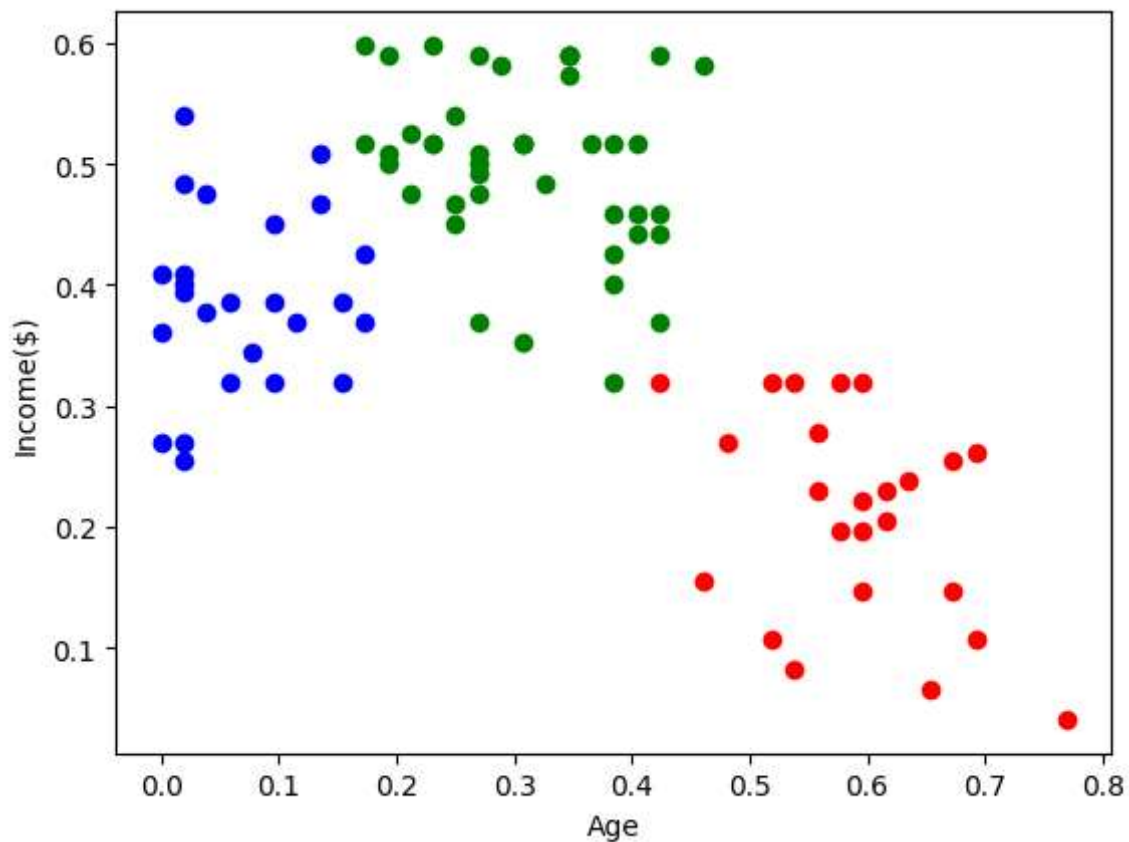
```
In [15]: df["New Cluster"]=y_predicted
df.head()
```

Out[15]:

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

```
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_
```

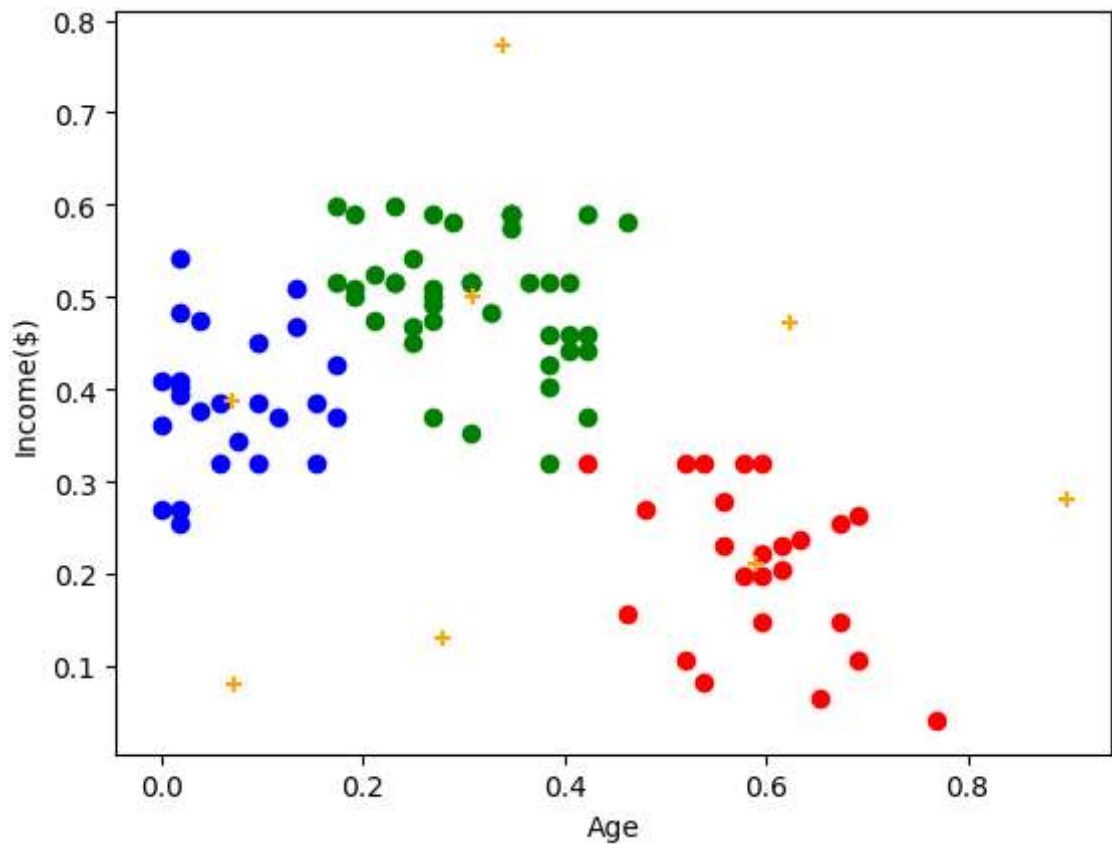
```
Out[17]: array([[0.58974359, 0.20969945],
 [0.30903399, 0.50114373],
 [0.06923077, 0.38786885],
 [0.62352071, 0.47225725],
 [0.07239819, 0.08003857],
 [0.33942308, 0.77295082],
 [0.27884615, 0.13040238],
 [0.89799331, 0.28011404]])
```

```

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

```

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



```

In [19]: k_rng=range(1,10)
sse=[]

```

```
In [20]: for k in k_rng:
          km=KMeans(n_clusters=k)
          km.fit(df[["Age", "Income($)"]])
          sse.append(km.inertia_)
          sse
```

C:\Users\dinesh reddy\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\dinesh reddy\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\dinesh reddy\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\dinesh reddy\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\dinesh reddy\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\dinesh reddy\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\dinesh reddy\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\dinesh reddy\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(
```

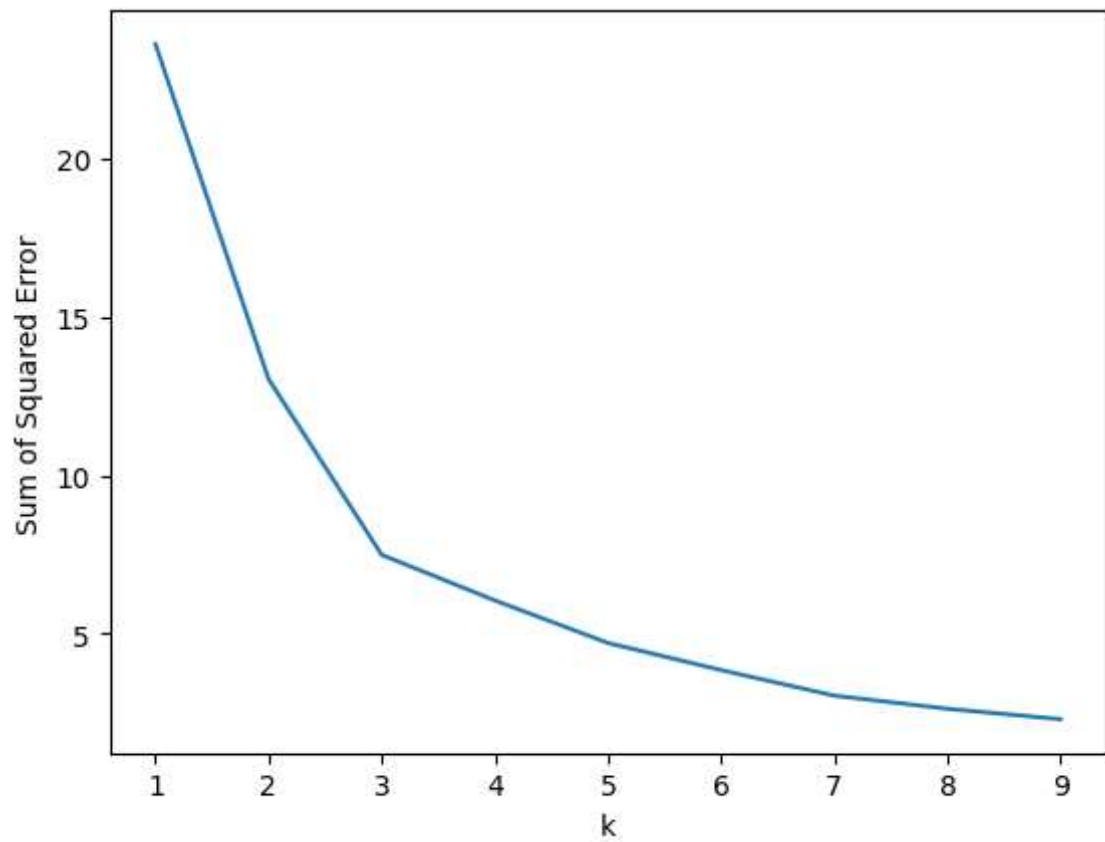
C:\Users\dinesh reddy\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning

```
warnings.warn(
```



```
In [21]: plt.plot(k_rng,sse)
plt.xlabel("k")
plt.ylabel("Sum of Squared Error")
```

Out[21]: Text(0, 0.5, 'Sum of Squared Error')



In []: