

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

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

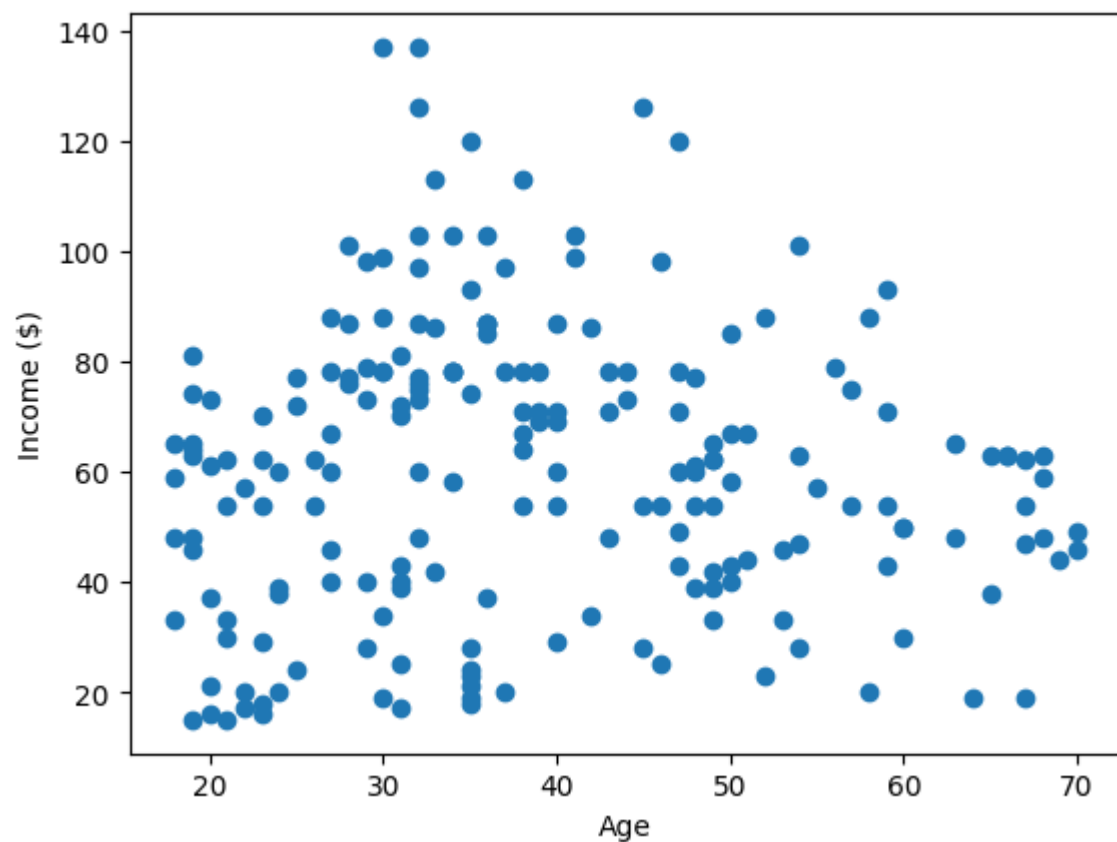
Out[2]:

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

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



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

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

```
Out[5]: 

▼ KMeans

  
KMeans()
```

```
In [6]: y_predicted=km.fit_predict(df[["Age", "Income($)"]])  
y_predicted
```

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

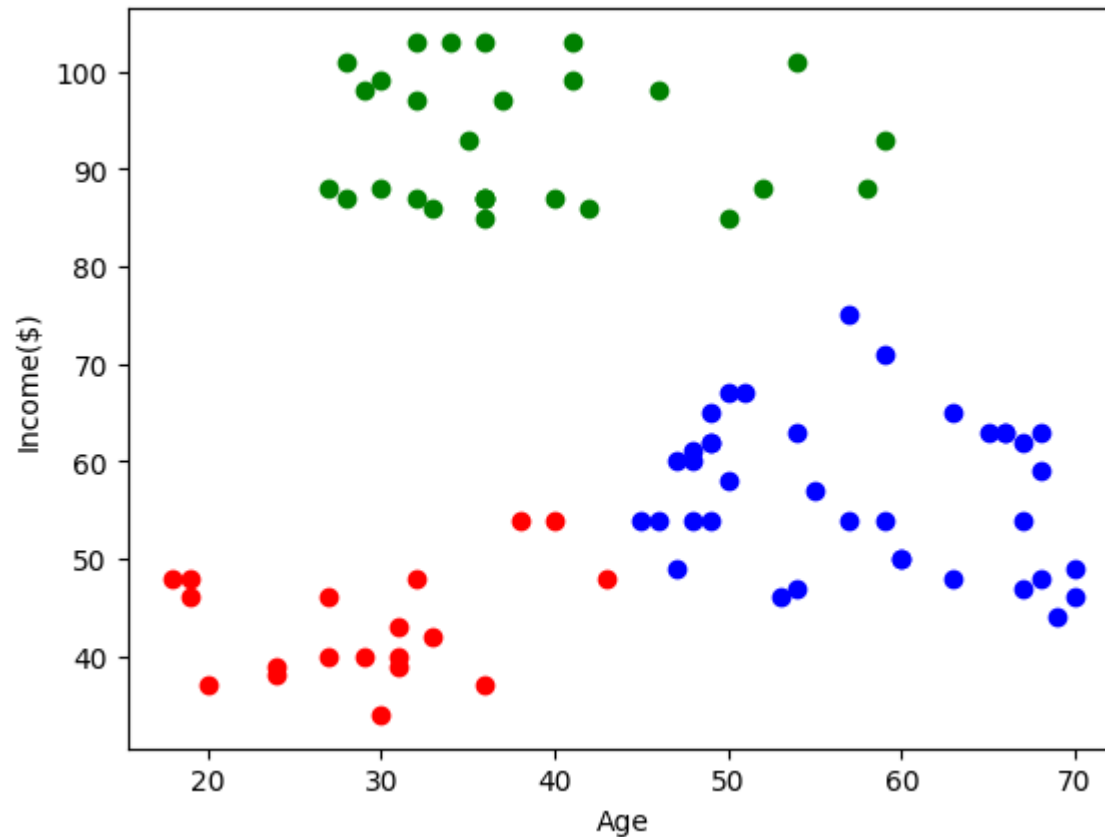
```
In [7]: df["Cluster"]=y_predicted  
df.head()
```

```
Out[7]:
```

	Gender	Age	Income(\$)	Cluster
0	Male	19	15	5
1	Male	21	15	5
2	Female	20	16	5
3	Female	23	16	5
4	Female	31	17	5

```
In [8]: df1=df[df.Cluster==0]
df2=df[df.Cluster==2]
df3=df[df.Cluster==3]
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[8]: Text(0, 0.5, 'Income(\$)')



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

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

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

Out[11]:

	Gender	Age	Income(\$)	Cluster
0	Male	19	0.000000	5
1	Male	21	0.000000	5
2	Female	20	0.008197	5
3	Female	23	0.008197	5
4	Female	31	0.016393	5

```
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	5
1	Male	0.057692	0.000000	5
2	Female	0.038462	0.008197	5
3	Female	0.096154	0.008197	5
4	Female	0.250000	0.016393	5

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

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

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

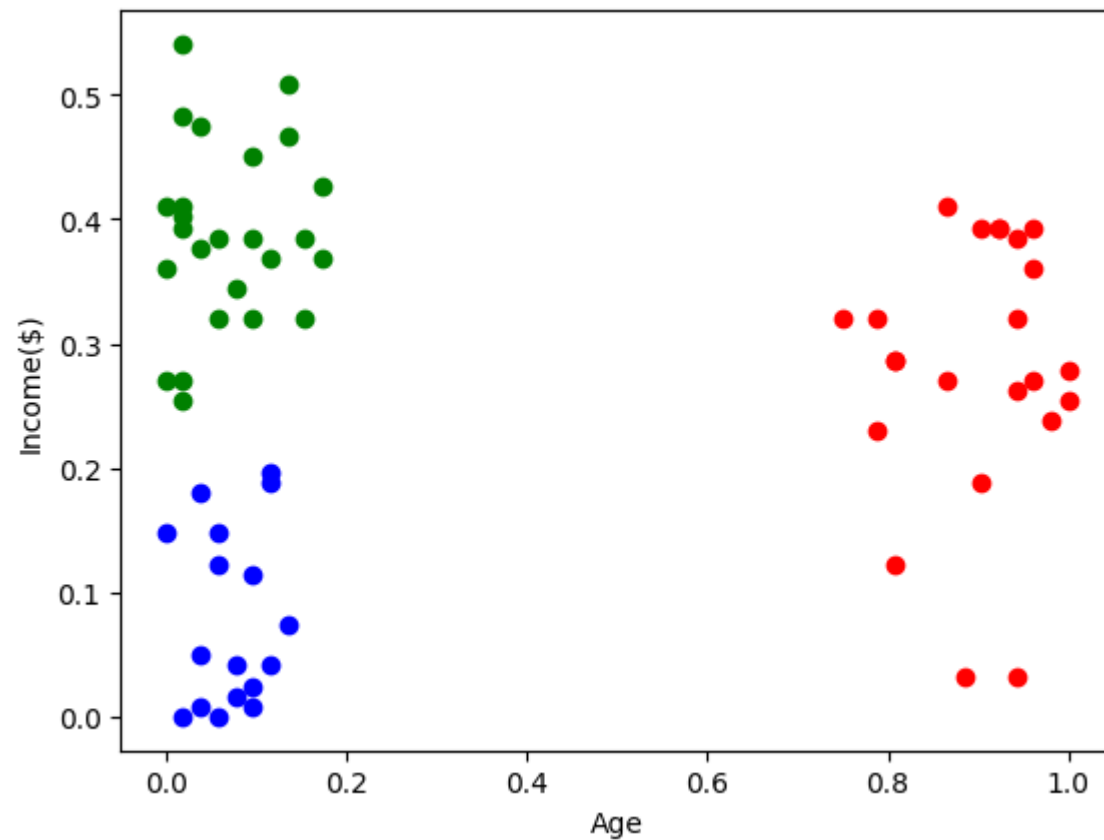
```
In [15]: df["New Cluster"]=y_predicted
df.head()
```

Out[15]:

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

```
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($))')
```



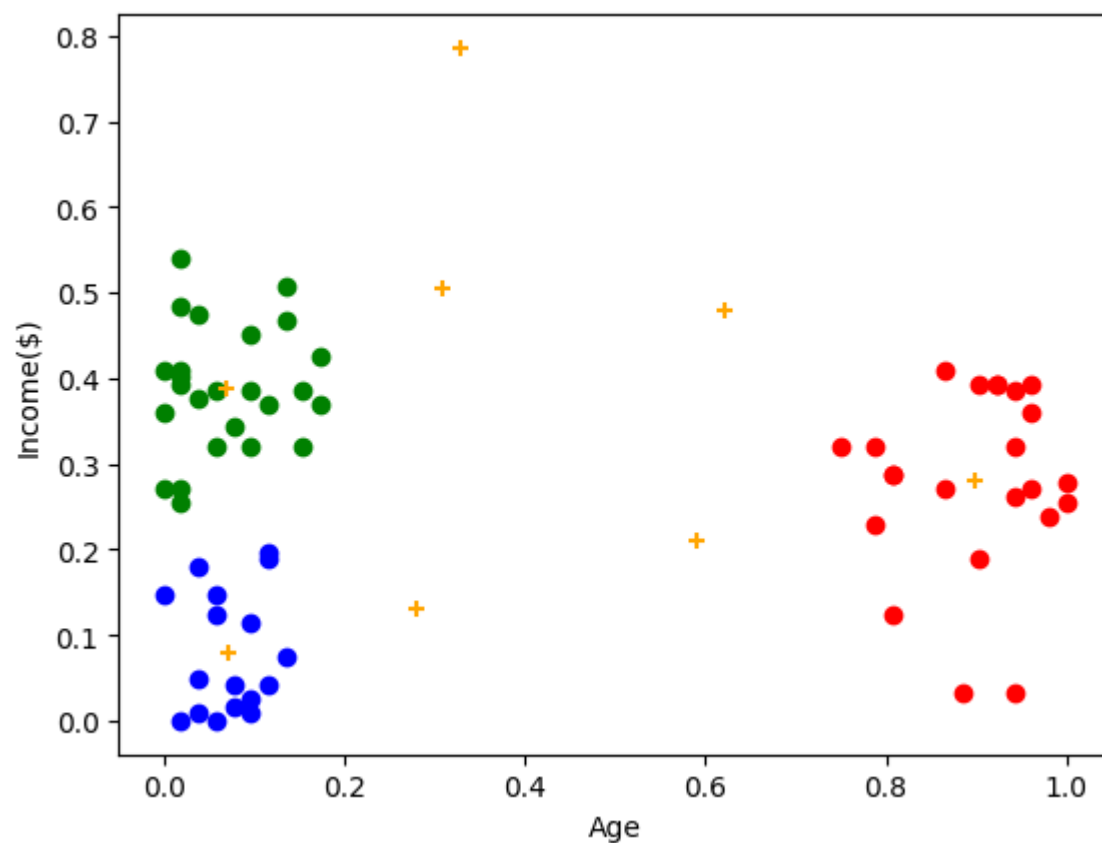
```
In [17]: km.cluster_centers_
```

```
Out[17]: array([[0.89799331, 0.28011404],  
                [0.06923077, 0.38786885],  
                [0.07239819, 0.08003857],  
                [0.62037037, 0.47996357],  
                [0.32905983, 0.78551913],  
                [0.27884615, 0.13040238],  
                [0.58974359, 0.20969945],  
                [0.30944056, 0.50428465]])
```



```
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(\$))')

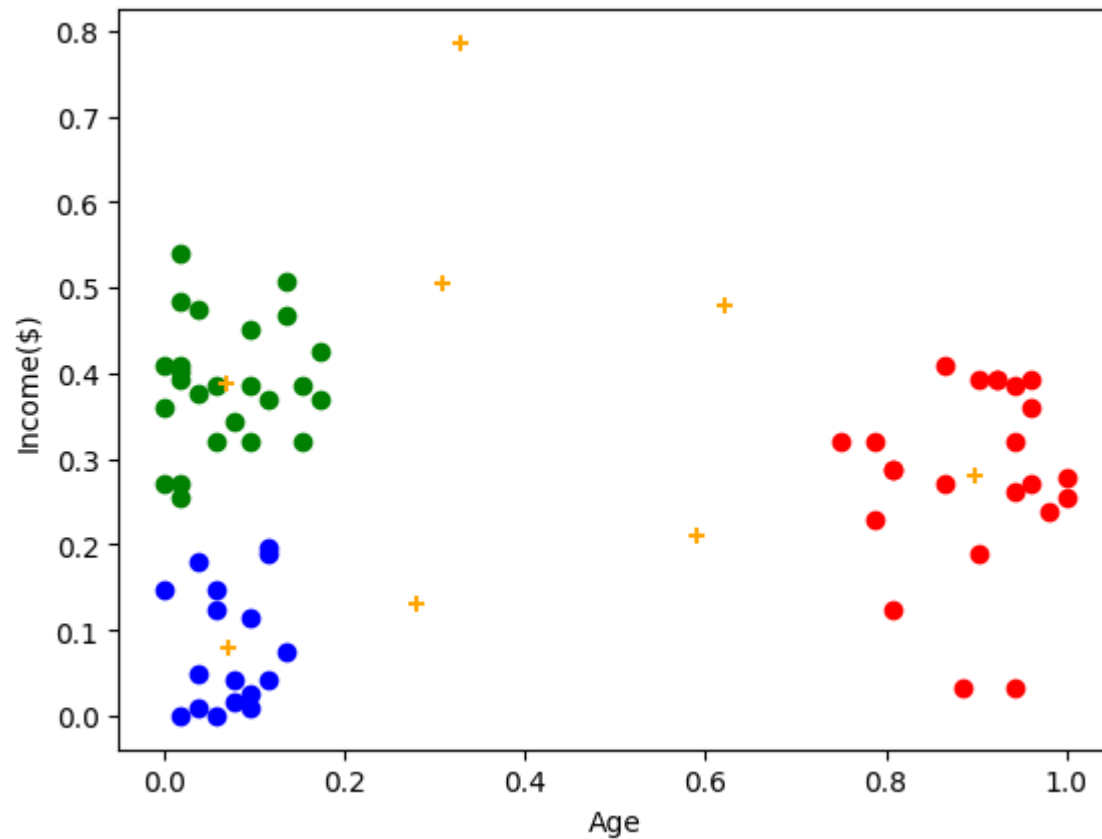


```
In [19]: km.cluster_centers_
```

```
Out[19]: array([[0.89799331, 0.28011404],  
                [0.06923077, 0.38786885],  
                [0.07239819, 0.08003857],  
                [0.62037037, 0.47996357],  
                [0.32905983, 0.78551913],  
                [0.27884615, 0.13040238],  
                [0.58974359, 0.20969945],  
                [0.30944056, 0.50428465]])
```

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

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

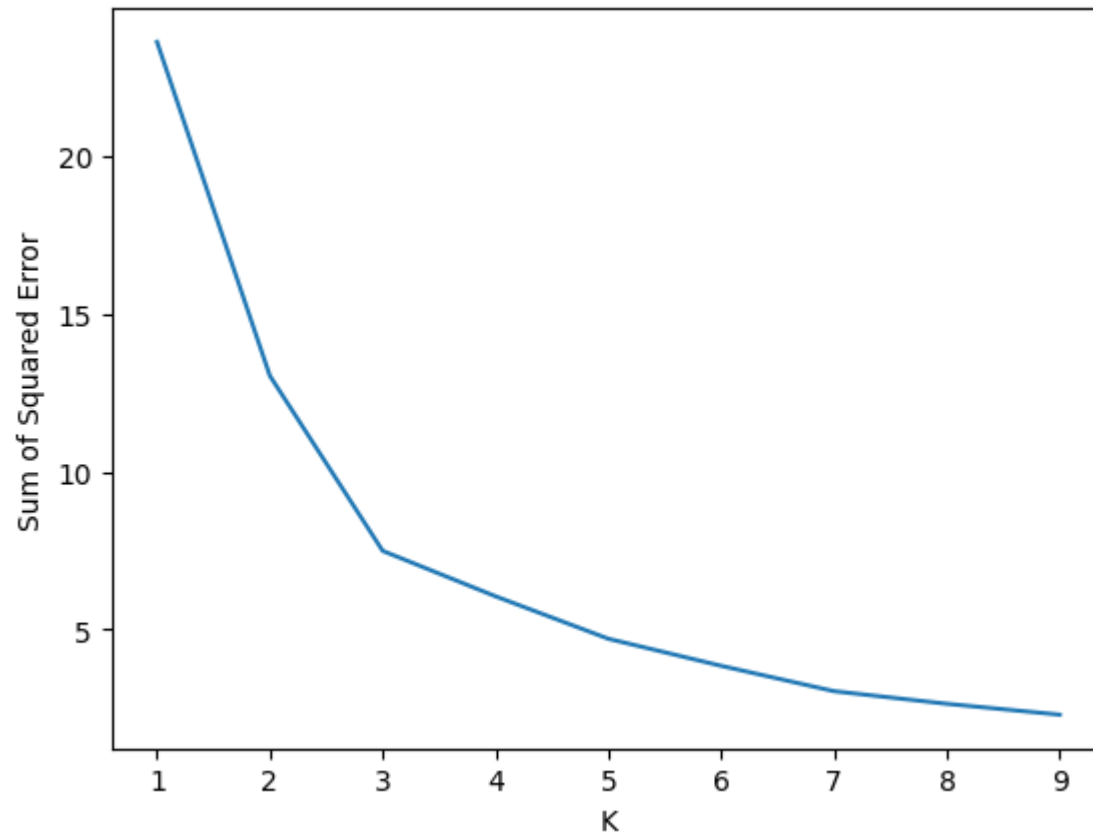


```
In [21]: k_rng=range(1,10)
sse=[]
```

```
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 errorprint(sse)
          plt.plot(k_rng,sse)
          plt.xlabel("K")
          plt.ylabel("Sum of Squared Error")
```

```
C:\Users\DELL\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\DELL\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\DELL\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\DELL\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\DELL\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\DELL\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\DELL\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\DELL\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\DELL\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[22]: Text(0, 0.5, 'Sum of Squared Error')
```



In []: