

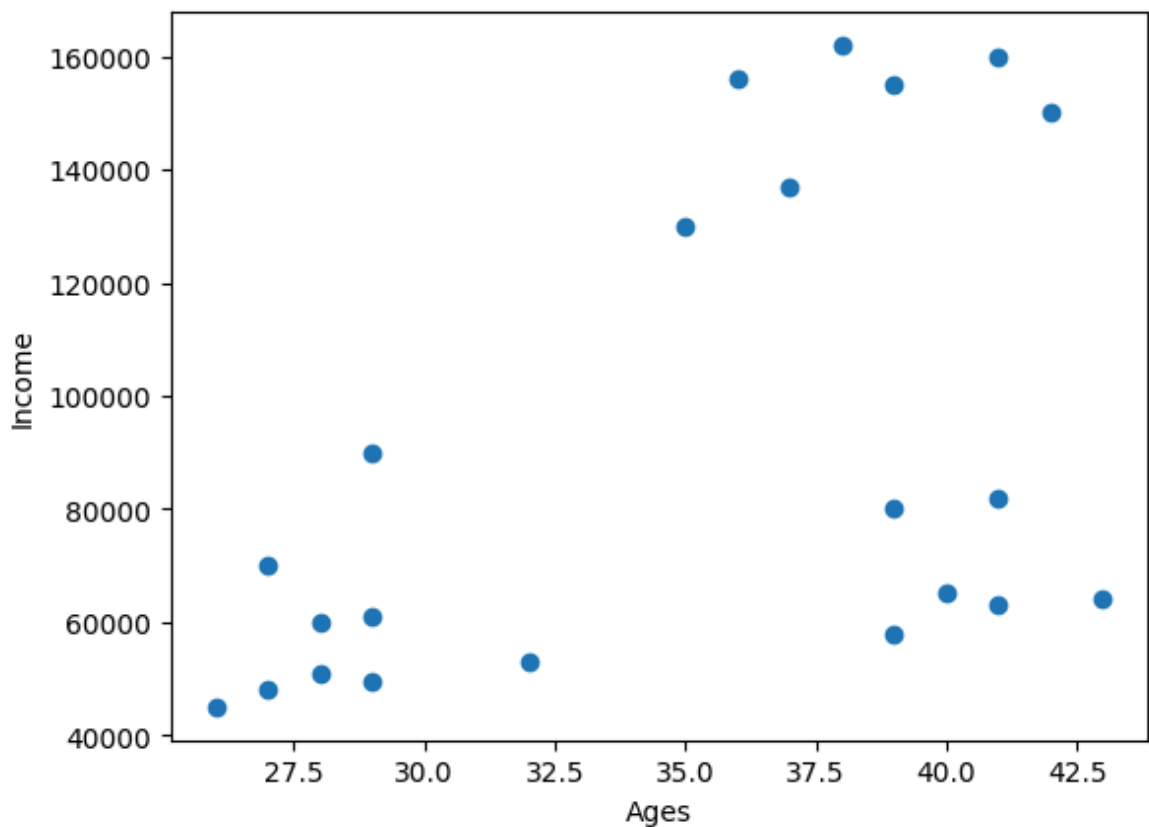
```
In [18]: import pandas as pd
data = pd.read_csv("C:\\Users\\sriha\\OneDrive\\Desktop\\pb excel\\salary for c
df=pd.DataFrame(data)
df.head()
```

```
Out[18]:
```

	Name	Age	Income(\$)
0	Rob	27	70000
1	Michael	29	90000
2	Mohan	29	61000
3	Ismail	28	60000
4	Kory	42	150000

```
In [19]: import matplotlib.pyplot as plt
plt.scatter(df.Age, df['Income($)'])
plt.xlabel('Ages')
plt.ylabel('Income')
```

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



```
In [20]: from sklearn.cluster import KMeans
km = KMeans(n_clusters=3)
y_pred = km.fit_predict(df[['Age', 'Income($)']])
y_pred
```

```
Out[20]: array([2, 2, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 2, 2, 1])
```

```
In [21]: df["clusters"] = y_pred
df.head()
```

```
Out[21]:
```

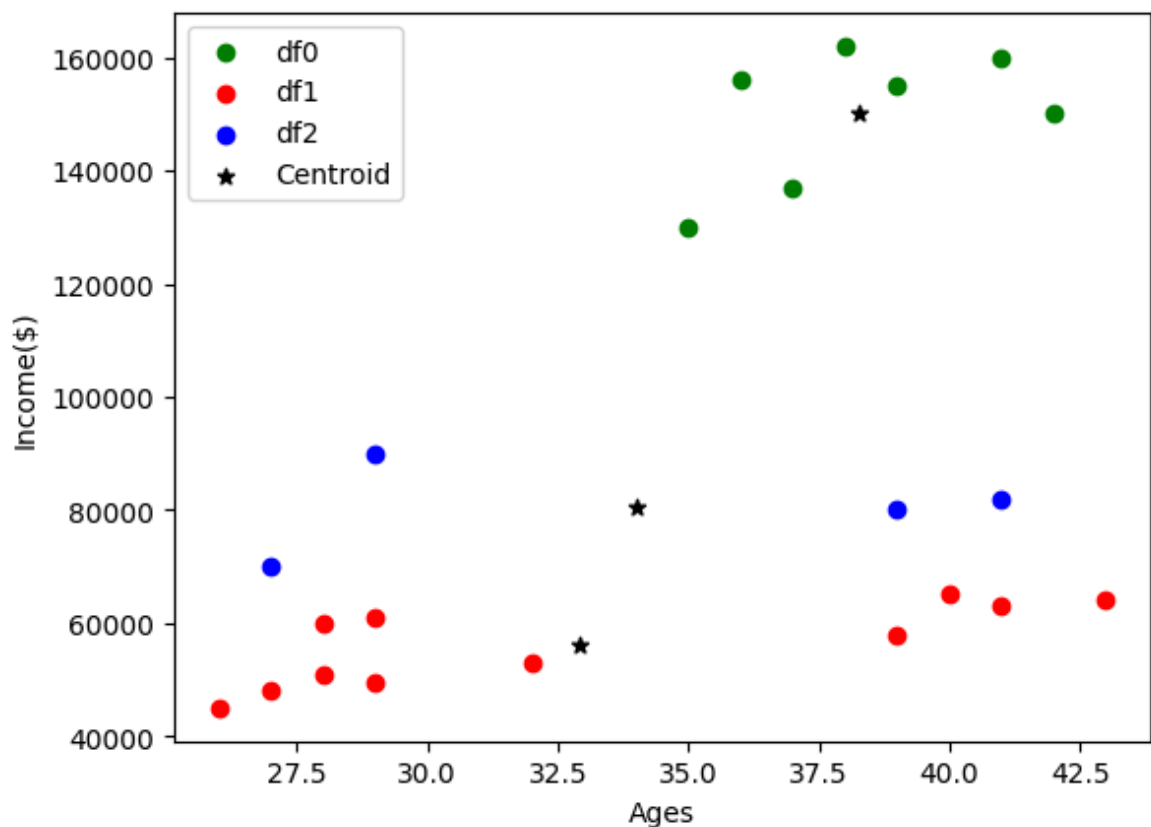
	Name	Age	Income(\$)	clusters
0	Rob	27	70000	2
1	Michael	29	90000	2
2	Mohan	29	61000	1
3	Ismail	28	60000	1
4	Kory	42	150000	0

```
In [22]: km.cluster_centers_
```

```
Out[22]: array([[3.82857143e+01, 1.50000000e+05],
               [3.29090909e+01, 5.61363636e+04],
               [3.40000000e+01, 8.05000000e+04]])
```

```
In [23]: df0 = df[df.clusters==0]
df1 = df[df.clusters==1]
df2 = df[df.clusters==2]
```

```
In [24]: pb.scatter(df0.Age ,df0['Income($)'] , color ="green",label= "df0" )
pb.scatter(df1.Age ,df1['Income($)'] , color ="red" ,label= 'df1')
pb.scatter(df2.Age ,df2['Income($)'] , color ="blue" ,label= 'df2')
pb.scatter (km.cluster_centers_[:,0],km.cluster_centers_[:,1], color ="Black",ma
pb.xlabel('Ages')
pb.ylabel("Income($)")
pb.legend()
pb.show()
```



```
In [25]: from sklearn . preprocessing import MinMaxScaler  
mm= MinMaxScaler()  
  
mm.fit_transform(df[['Age']])
```

```
Out[25]: array([[0.05882353],  
                [0.17647059],  
                [0.17647059],  
                [0.11764706],  
                [0.94117647],  
                [0.76470588],  
                [0.88235294],  
                [0.70588235],  
                [0.58823529],  
                [0.52941176],  
                [0.64705882],  
                [0.        ],  
                [0.05882353],  
                [0.11764706],  
                [0.17647059],  
                [0.35294118],  
                [0.82352941],  
                [0.88235294],  
                [1.        ],  
                [0.76470588],  
                [0.88235294],  
                [0.76470588]])
```

```
In [26]: mm.fit_transform(df[['Income($)']])
```

```
Out[26]: array([[0.21367521],  
                [0.38461538],  
                [0.13675214],  
                [0.12820513],  
                [0.8974359 ],  
                [0.94017094],  
                [0.98290598],  
                [1.        ],  
                [0.94871795],  
                [0.72649573],  
                [0.78632479],  
                [0.        ],  
                [0.02564103],  
                [0.05128205],  
                [0.03846154],  
                [0.06837607],  
                [0.17094017],  
                [0.15384615],  
                [0.16239316],  
                [0.2991453 ],  
                [0.31623932],  
                [0.11111111]])
```

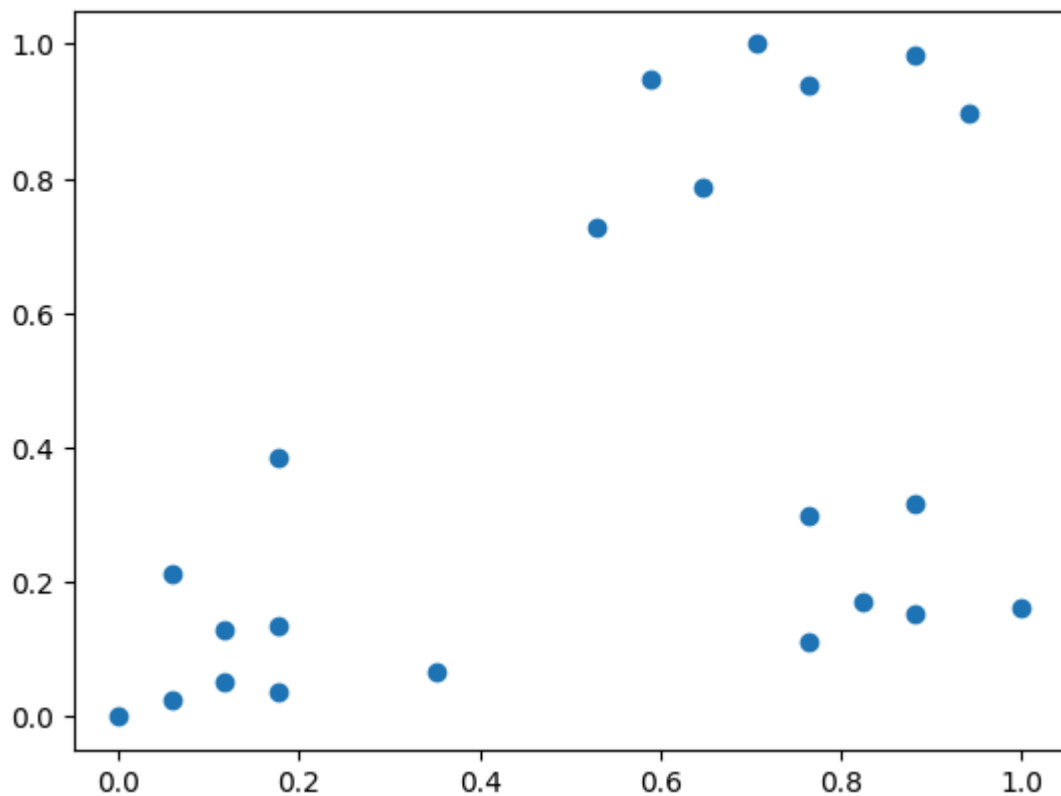
```
In [27]: df['Age'] = mm.fit_transform(df[['Age']])  
df['Income($)'] = mm.fit_transform(df[['Income($)']])  
df.head()
```

Out[27]:

	Name	Age	Income(\$)	clusters
0	Rob	0.058824	0.213675	2
1	Michael	0.176471	0.384615	2
2	Mohan	0.176471	0.136752	1
3	Ismail	0.117647	0.128205	1
4	Kory	0.941176	0.897436	0

In [28]: `pb.scatter(df.Age , df["Income($)"])`

Out[28]: `<matplotlib.collections.PathCollection at 0x2b9736def28>`



In [29]: `km = KMeans(n_clusters=3)`  
`y_pred = km.fit_predict(df[['Age', 'Income($)']])`  
`df['cluster'] = y_pred`

Out[29]: `array([0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2])`

In [39]: `df.head()`

```
Out[39]:
```

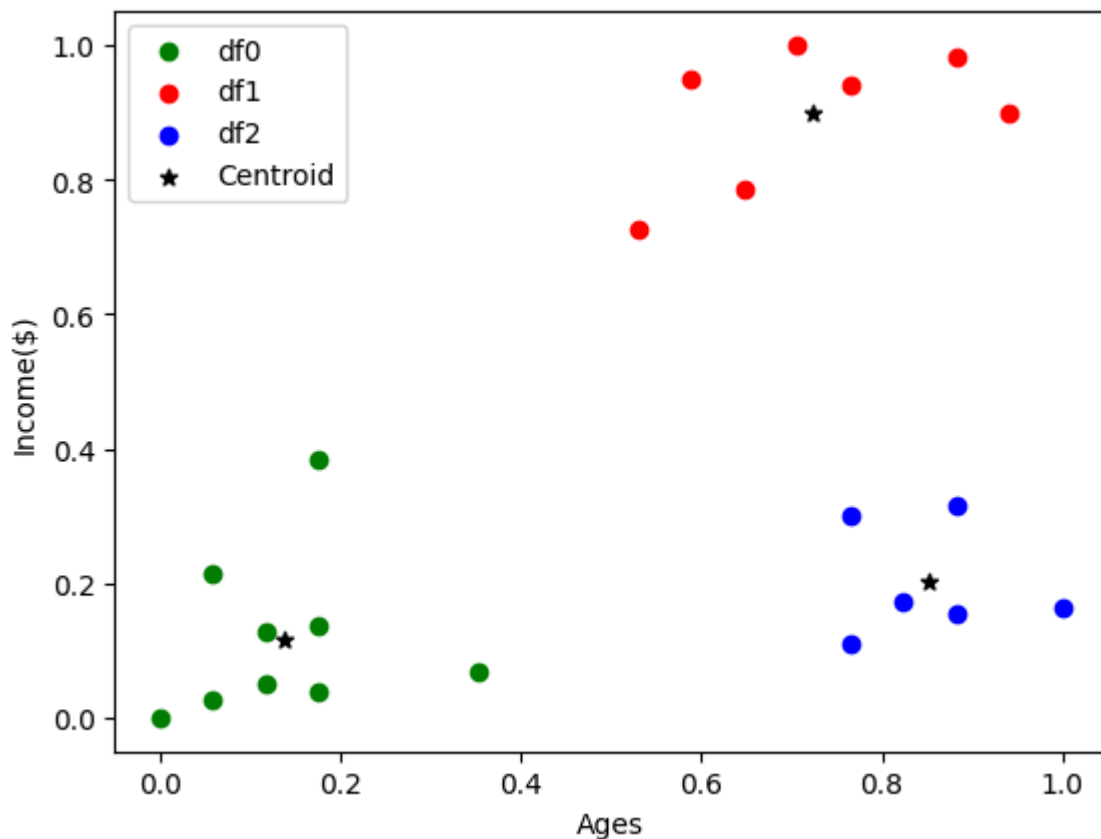
	Name	Age	Income(\$)	cluster
0	Rob	0.058824	0.213675	0
1	Michael	0.176471	0.384615	0
2	Mohan	0.176471	0.136752	0
3	Ismail	0.117647	0.128205	0
4	Kory	0.941176	0.897436	1

```
In [31]: km.cluster_centers_
```

```
Out[31]: array([[0.1372549 , 0.11633428],
                [0.72268908, 0.8974359 ],
                [0.85294118, 0.2022792 ]])
```

```
In [40]: df0 = df[df.cluster==0]
df1 = df[df.cluster==1]
df2 = df[df.cluster==2]
```

```
In [45]: pb.scatter(df0.Age ,df0['Income($)'] , color ="green",label= "df0" )
pb.scatter(df1.Age ,df1['Income($)'] , color ="red" ,label = 'df1')
pb.scatter(df2.Age ,df2['Income($)'] , color ="blue" ,label = 'df2')
pb.scatter (km.cluster_centers_[ :,0],km.cluster_centers_[ :,1], color ="Black",ma
pb.xlabel('Ages')
pb.ylabel("Income($)")
pb.legend()
pb.show()
```

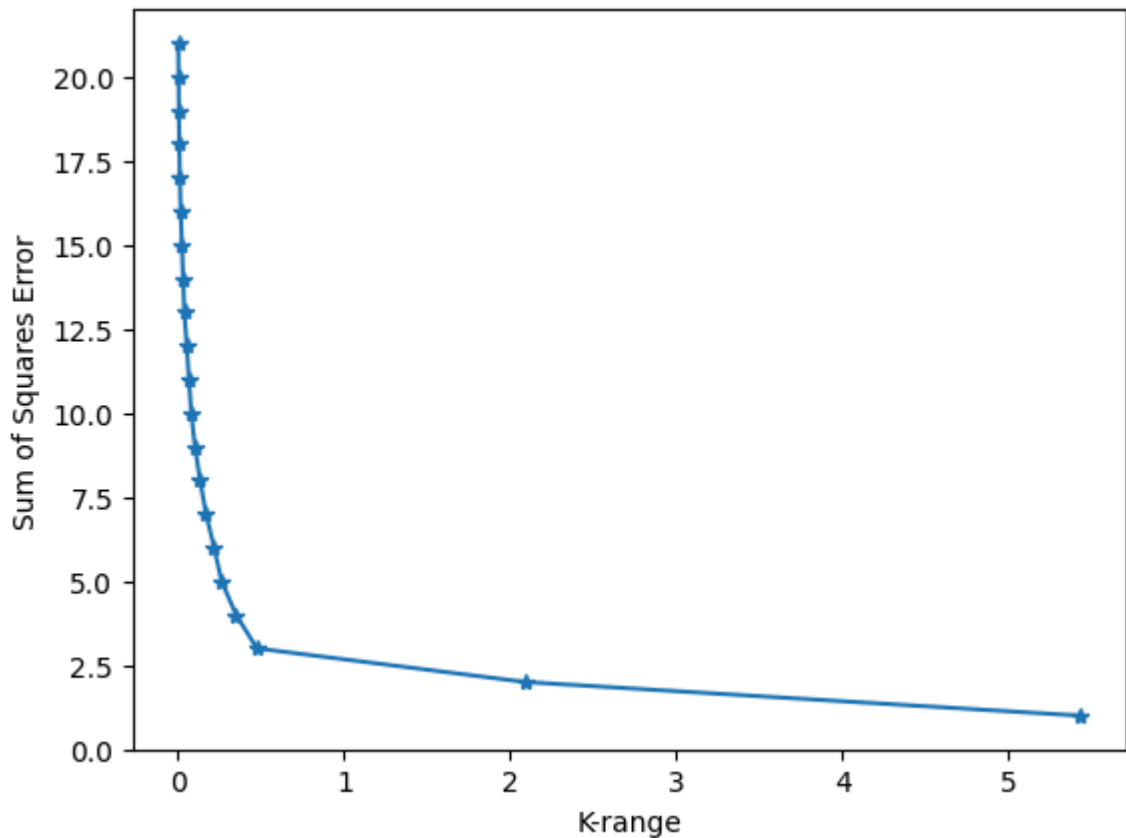


## Elbow Polt

```
In [59]: k_range = range(1,22)
sse = []
for k in k_range:
    km=KMeans(n_clusters=k)
    km.fit(df[['Age', 'Income($)']])
    sse.append(km.inertia_)
```

```
In [60]: pb.xlabel('K-range')
pb.ylabel('Sum of Squares Error')
pb.plot(sse,k_range , marker = "*")
```

Out[60]: [<matplotlib.lines.Line2D at 0x2b973d212e8>]



In [ ]: