

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

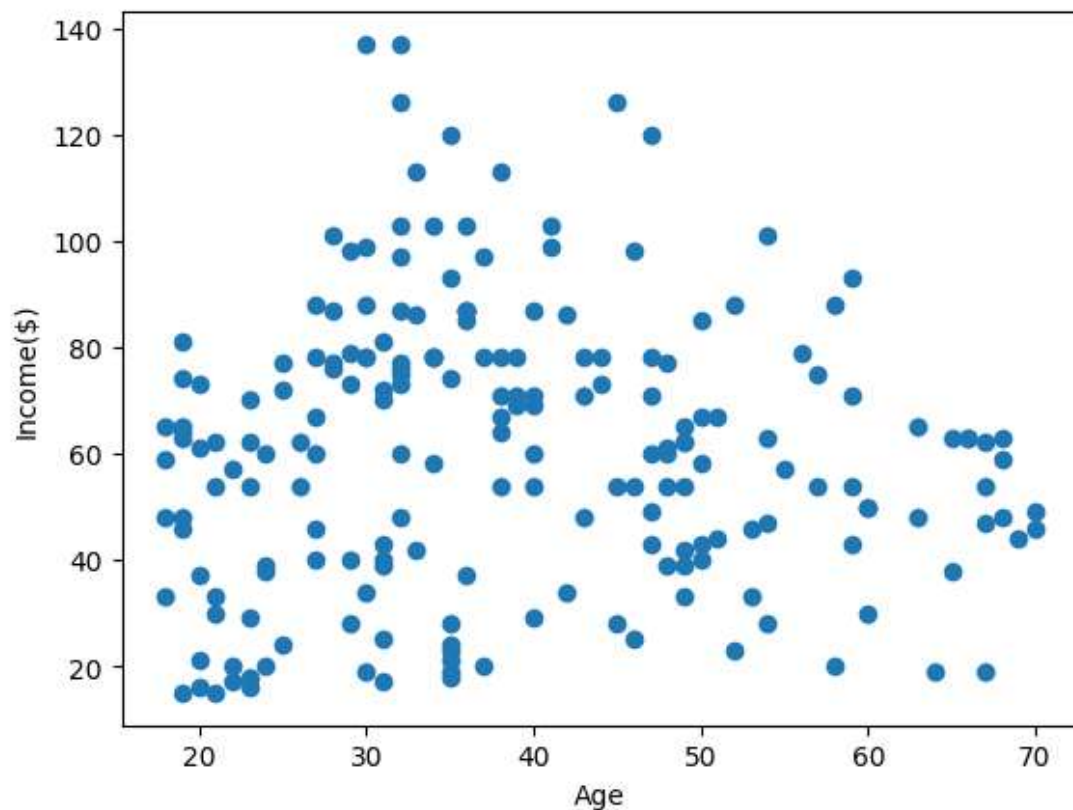
```
In [2]: df=pd.read_csv(r"C:\Users\venka\Downloads\Income.csv")
df.head()
```

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

```
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\venka\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([6, 6, 6, 6, 6, 6, 6, 6, 0, 6, 0, 6, 0, 6, 6, 6, 6, 6, 0, 6, 6, 6,  
              0, 6, 0, 6, 0, 6, 0, 6, 0, 6, 0, 3, 0, 3, 0, 3, 3, 0, 3, 0, 3,  
              0, 3, 0, 3, 3, 3, 0, 3, 3, 0, 0, 0, 0, 4, 3, 0, 4, 3, 4, 4, 3,  
              0, 4, 3, 3, 4, 4, 4, 4, 4, 7, 4, 7, 7, 4, 4, 7, 4, 4, 7, 4, 4, 7,  
              7, 4, 4, 7, 4, 7, 7, 7, 4, 7, 4, 7, 7, 4, 4, 7, 4, 7, 4, 4, 4, 4,  
              4, 7, 7, 7, 7, 7, 4, 4, 4, 4, 7, 1, 1, 1, 7, 1, 1, 1, 4, 1, 1, 1,  
              7, 1, 7, 1, 1, 1, 7, 1, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,  
              1, 1, 1, 1, 1, 1, 4, 1, 1, 1, 5, 1, 5, 1, 1, 1, 5, 1, 1, 1, 5, 1,  
              5, 1, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 2, 2, 2, 2, 2, 2,  
              2, 2])
```

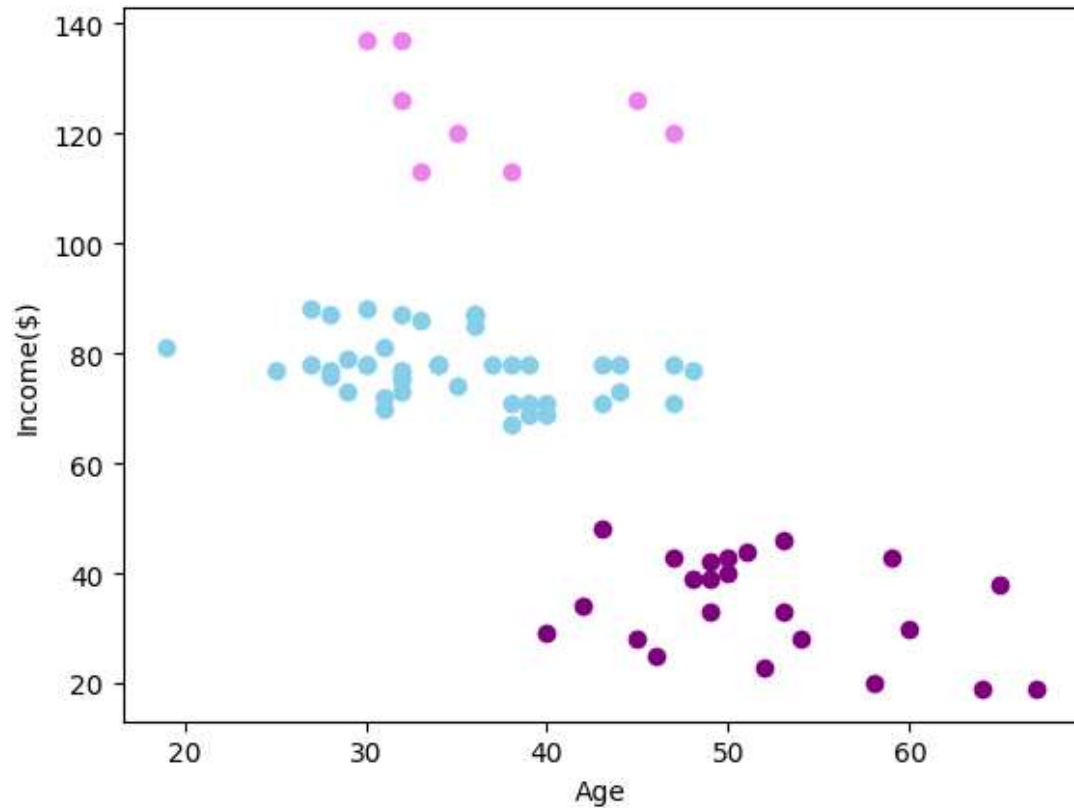
```
In [7]: df["cluster"]=y_predicted  
df.head()
```

```
Out[7]:
```

	Gender	Age	Income(\$)	cluster
0	Male	19	15	6
1	Male	21	15	6
2	Female	20	16	6
3	Female	23	16	6
4	Female	31	17	6

```
In [8]: df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="purple")
plt.scatter(df2["Age"],df2["Income($)"],color="skyblue")
plt.scatter(df3["Age"],df3["Income($)"],color="violet")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

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	6
1	Male	21	0.000000	6
2	Female	20	0.008197	6
3	Female	23	0.008197	6
4	Female	31	0.016393	6

```
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	6
1	Male	0.057692	0.000000	6
2	Female	0.038462	0.008197	6
3	Female	0.096154	0.008197	6
4	Female	0.250000	0.016393	6

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

Out[13]:

▼ KMeans
KMeans()

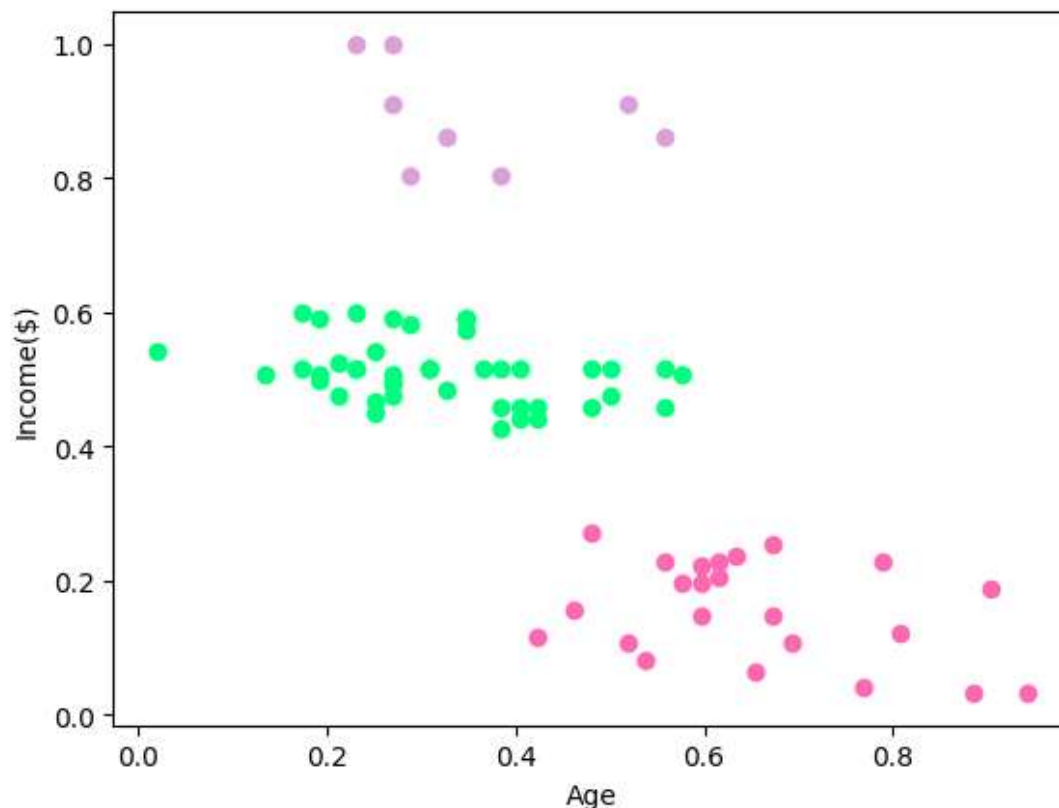
```
In [14]: y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

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

```
In [15]: df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="hotpink")
plt.scatter(df2["Age"],df2["Income($)"],color="SpringGreen")
plt.scatter(df3["Age"],df3["Income($)"],color="plum")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

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

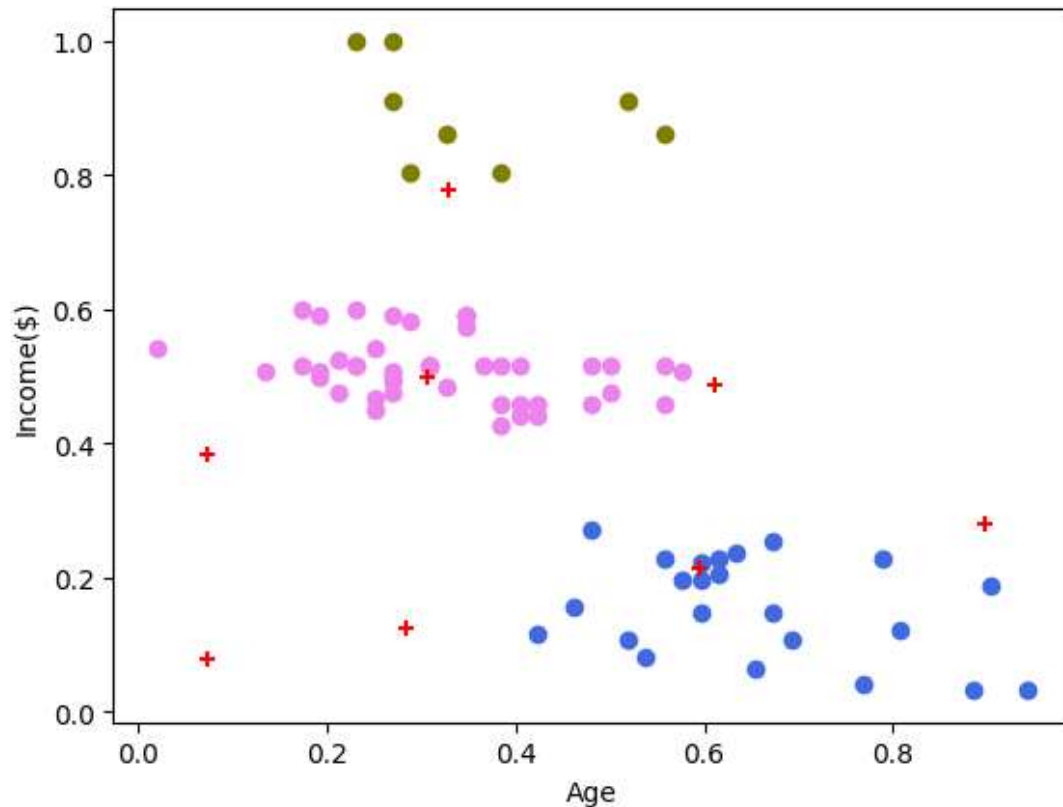


```
In [17]: km.cluster_centers_
```

```
Out[17]: array([[0.32894737, 0.77782571],
 [0.61111111, 0.48876746],
 [0.07322485, 0.38272383],
 [0.89799331, 0.28011404],
 [0.59461538, 0.21508197],
 [0.07239819, 0.08003857],
 [0.30540293, 0.49921936],
 [0.28388278, 0.1245121 ]])
```

```
In [19]: df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="royalblue")
plt.scatter(df2["Age"],df2["Income($)"],color="violet")
plt.scatter(df3["Age"],df3["Income($)"],color="olive")
plt.scatter(km.cluster_centers_[0],km.cluster_centers_[1],color="red",marker="+")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

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



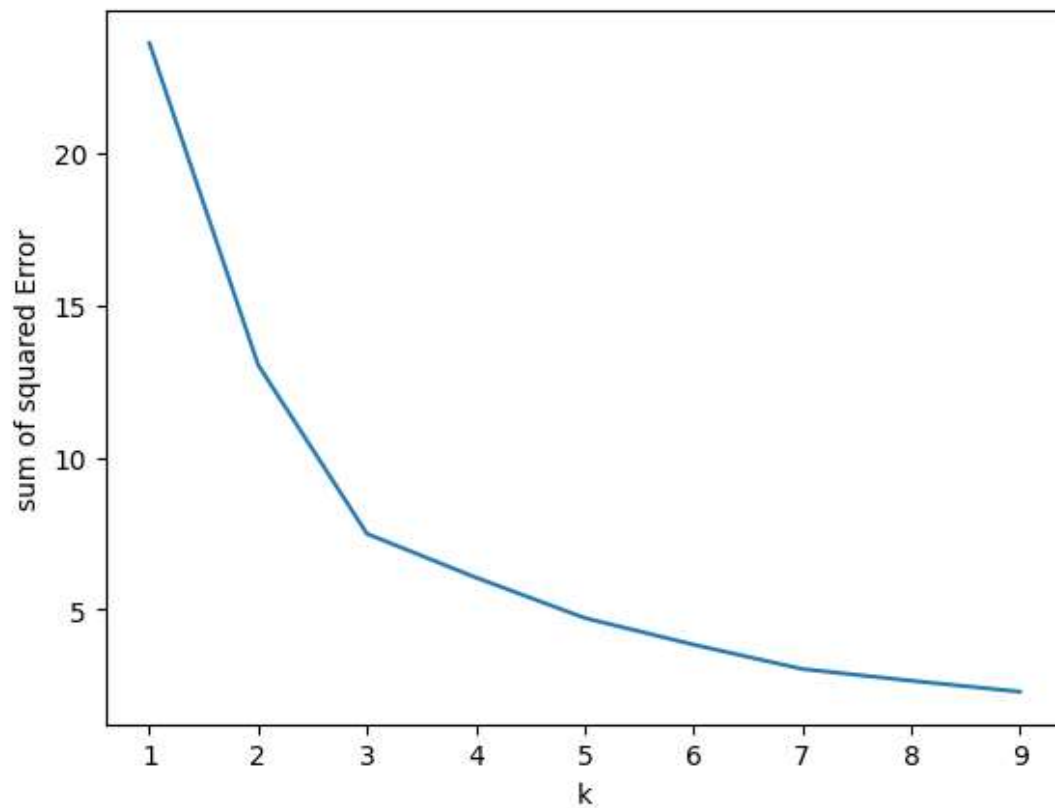
```
In [20]: k_rng=range(1,10)
          sse=[]
          for k in k_rng:
              km=KMeans(n_clusters=k)
              km.fit(df[["Age", "Income($)"]])
              sse.append(km.inertia_)
          sse
```

```
C:\Users\venka\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\venka\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\venka\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\venka\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\venka\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\venka\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\venka\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\venka\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\venka\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[20]: [23.583906150363603,
13.02893842801829,
7.492113413237459,
6.058372453353155,
4.7320582035076235,
3.8578918221646457,
3.058061107078988,
2.6738190874812107,
2.3135720353543285]
```

```
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 []: