```
In [1]: import pandas as pd
        from matplotlib import pyplot as plt
        %matplotlib inline
```
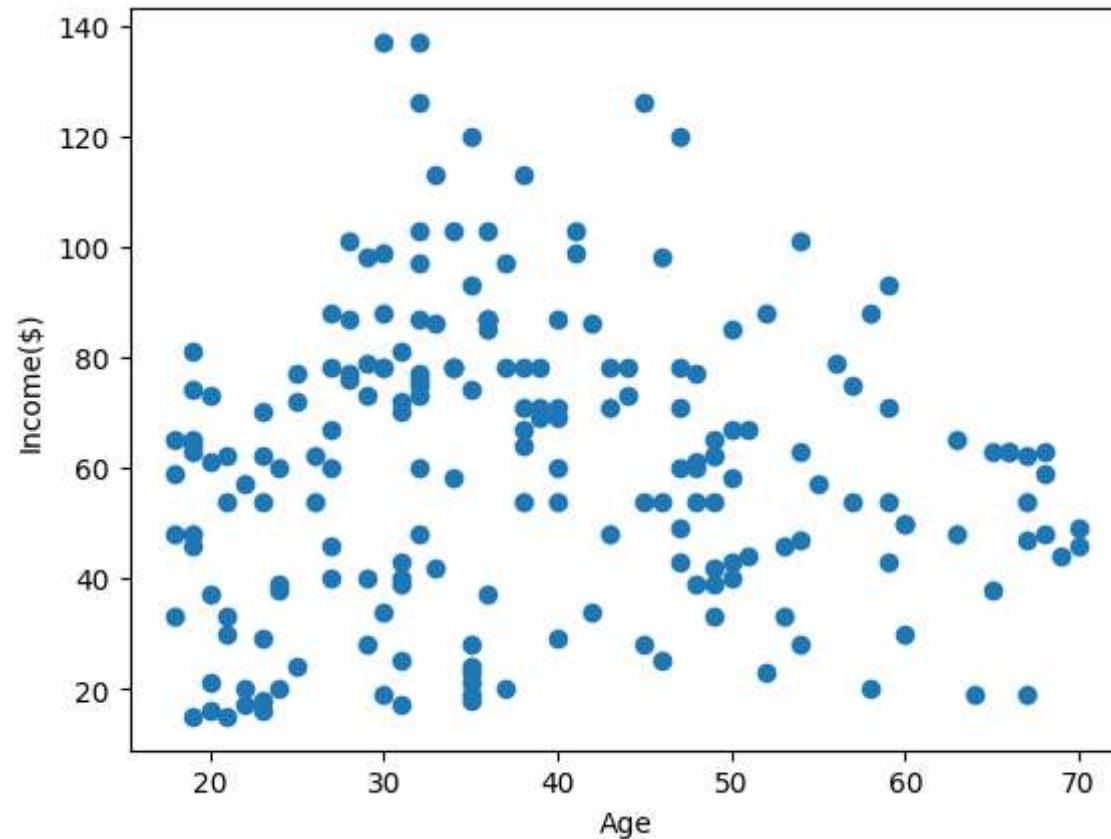
```
In [2]: df=pd.read_csv(r"C:\Users\arshiha\Downloads\Income.csv")
        df.head()
```

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

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

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

In [5]:
```python
Km=KMeans()
Km
```

Out[5]: KMeans()

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [20]:
```python
y_predicted=Km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

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

In [7]:
```python
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]:
```python
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[8]: Text(0, 0.5, 'Income($)')
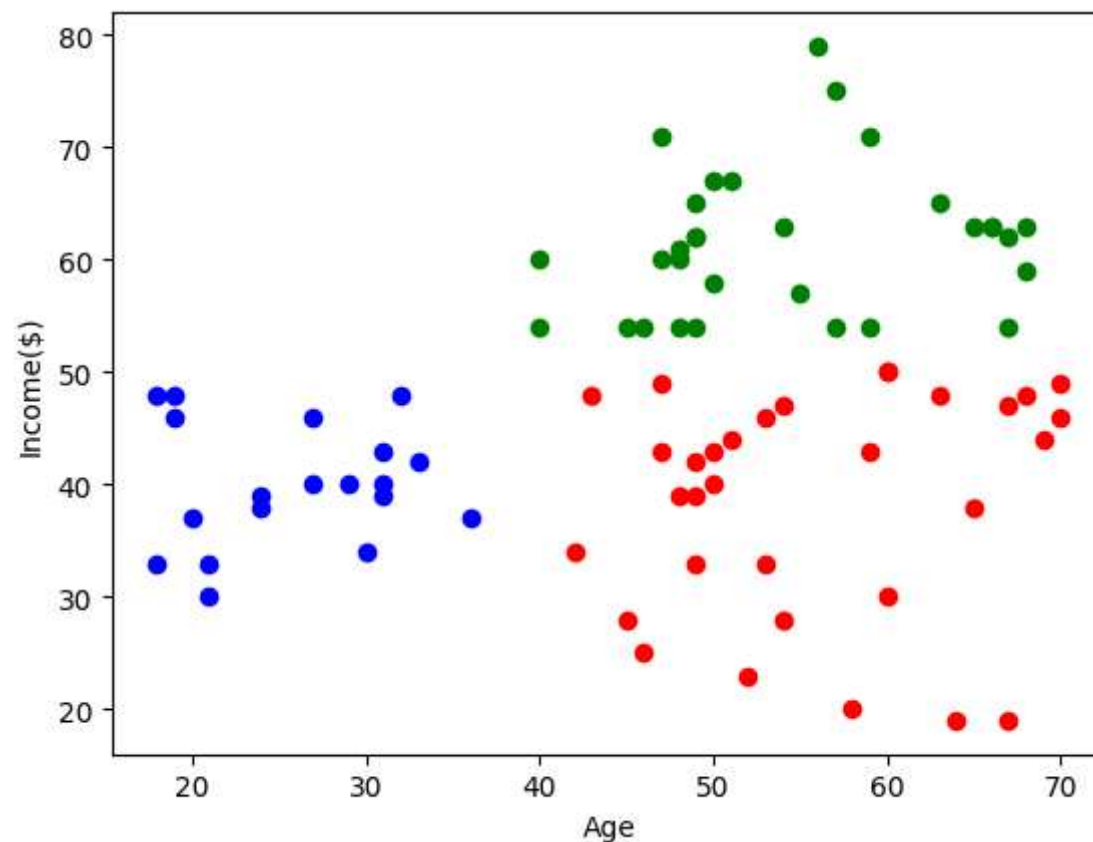
In [9]:
```python
from sklearn.preprocessing import MinMaxScaler
```

```
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]: Km=KMeans()
```

```
In [13]: df["New Cluster"]=y_predicted
         df.head()
```

Out[13]:

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

```
In [14]: 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[14]: Text(0, 0.5, 'Income($)')
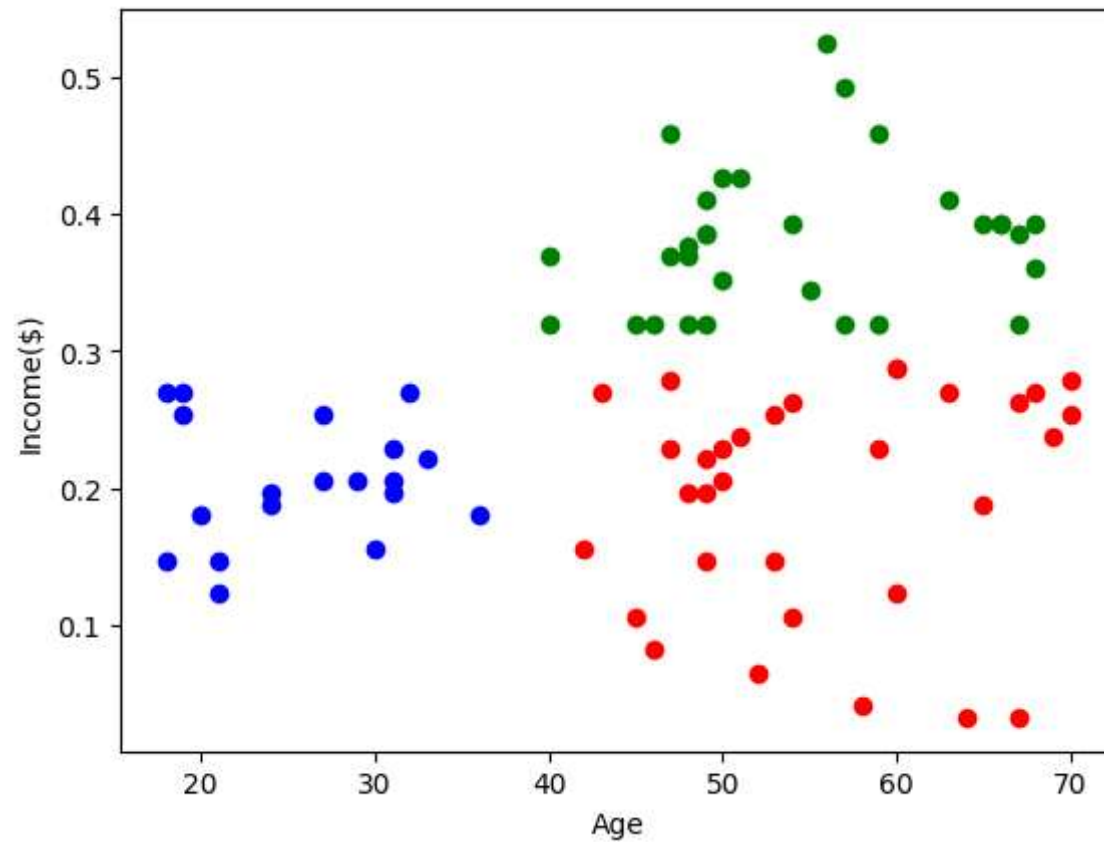
In [22]: Km.cluster_centers_

Out[22]: array([[41.66666667, 0.44925839],
               [19.8   , 0.25639344],
               [66.64705882, 0.28688525],
               [30.68421053, 0.47648835],
               [48.90625   , 0.34836066],
               [57.23529412, 0.36065574],
               [35.86206897, 0.42622951],
               [24.9047619 , 0.29625293]])

In [28]:
```python
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($)"],color="blue")
plt.scatter(Km.cluster_centers_[:,0],Km.cluster_centers_[:,1],color="orange",marker="+")
plt.xlabel("Age")
plt.ylabel("Income($)")
```
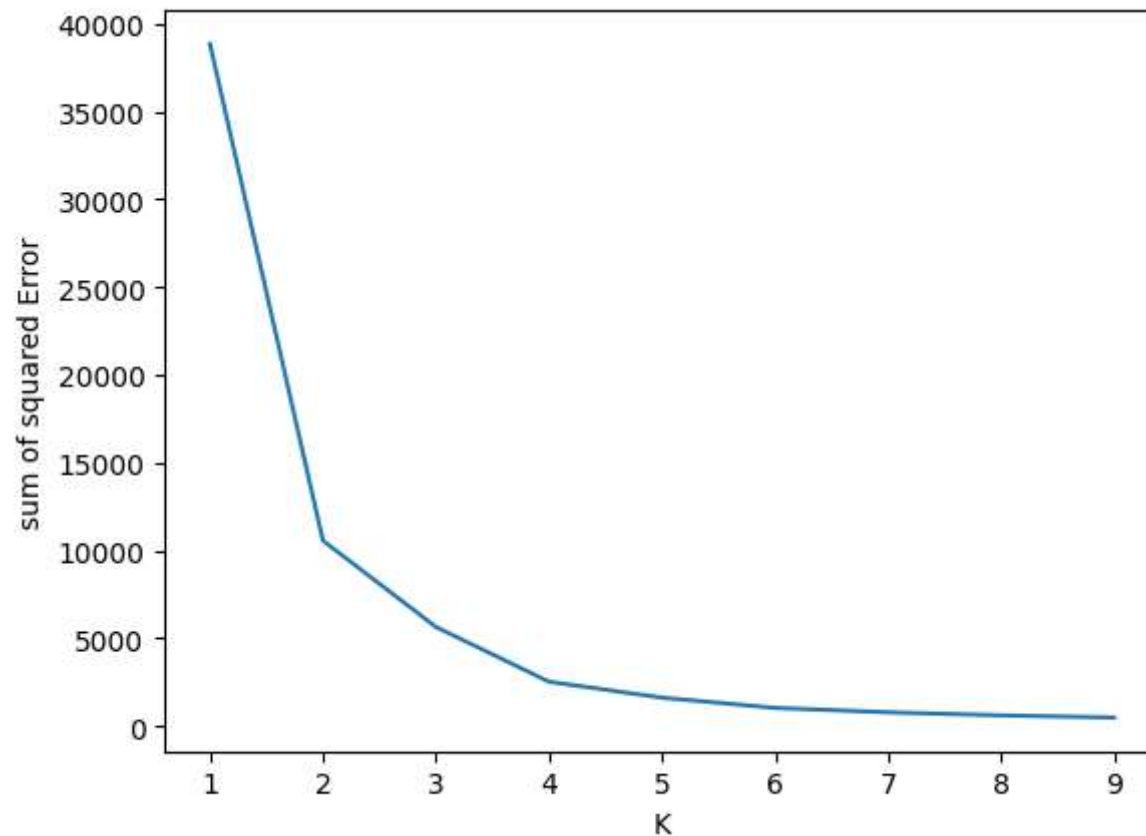
Out[28]: Text(0, 0.5, 'Income($)')
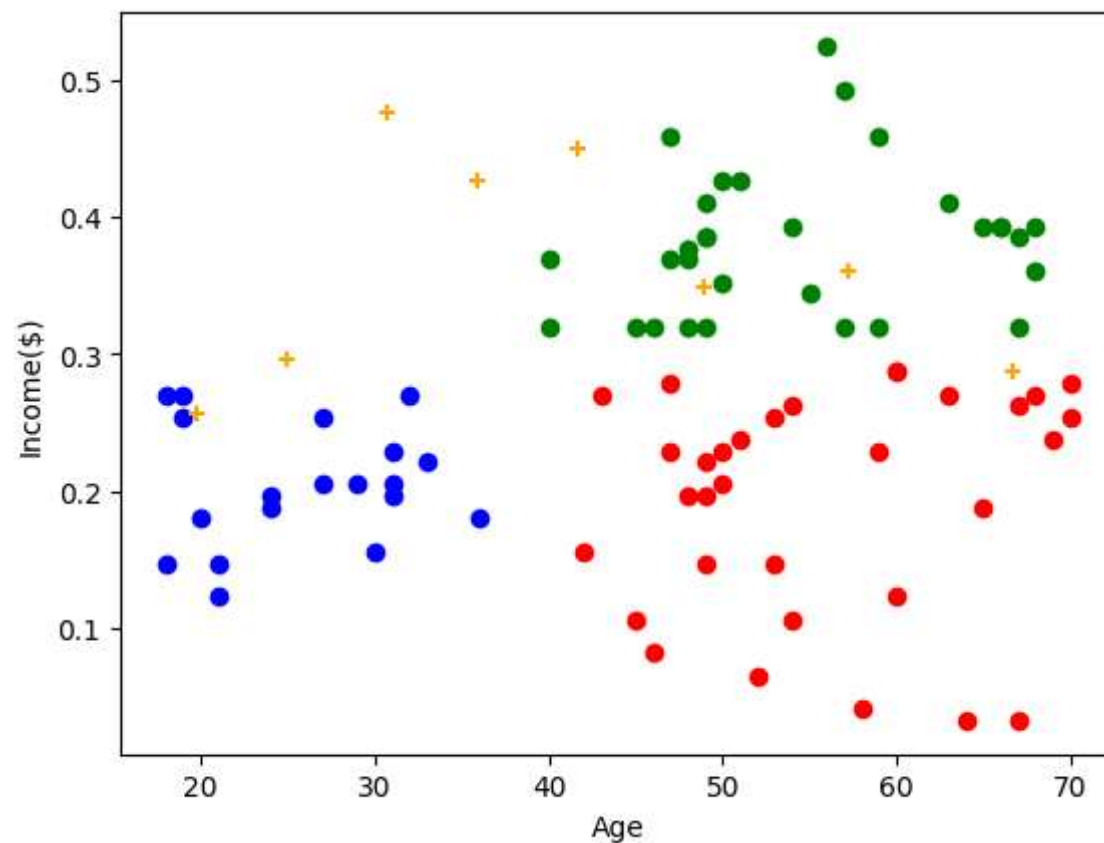
In [29]:
```python
K_rng=range(1,10)
sse=[]
for K in K_rng:
    Km=KMeans(n_clusters=K)
    Km.fit(df[["Age","Income($)"]])
    sse.append(Km.inertia_)
sse
```

```
C:\Users\arshiha\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\arshiha\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\arshiha\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\arshiha\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\arshiha\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\arshiha\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\arshiha\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\arshiha\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\arshiha\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[29]: [38840.723144316034,
         10558.82532963463,
         5641.051276951294,
         2521.859262993656,
         1614.5937582632425,
         1031.7292698285246,
         782.0798354177841,
         604.885525249197,
         484.2535461232586]

In [30]:
```
plt.plot(K_rng,sse)
plt.xlabel("K")
plt.ylabel("sum of squared Error")
```

Out[30]: Text(0, 0.5, 'sum of squared Error')

In [ ]: