```
In [4]:  ▶  import numpy
            import matplotlib.pyplot as plt
            import pygad
            import pandas as pd
```

```
In [5]:  ▶  df=pd.read_csv(r"C:\Users\MY HOME\Downloads\Income.csv")
            df
```

Out[5]:

	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 [7]:  ▶  df.shape
```
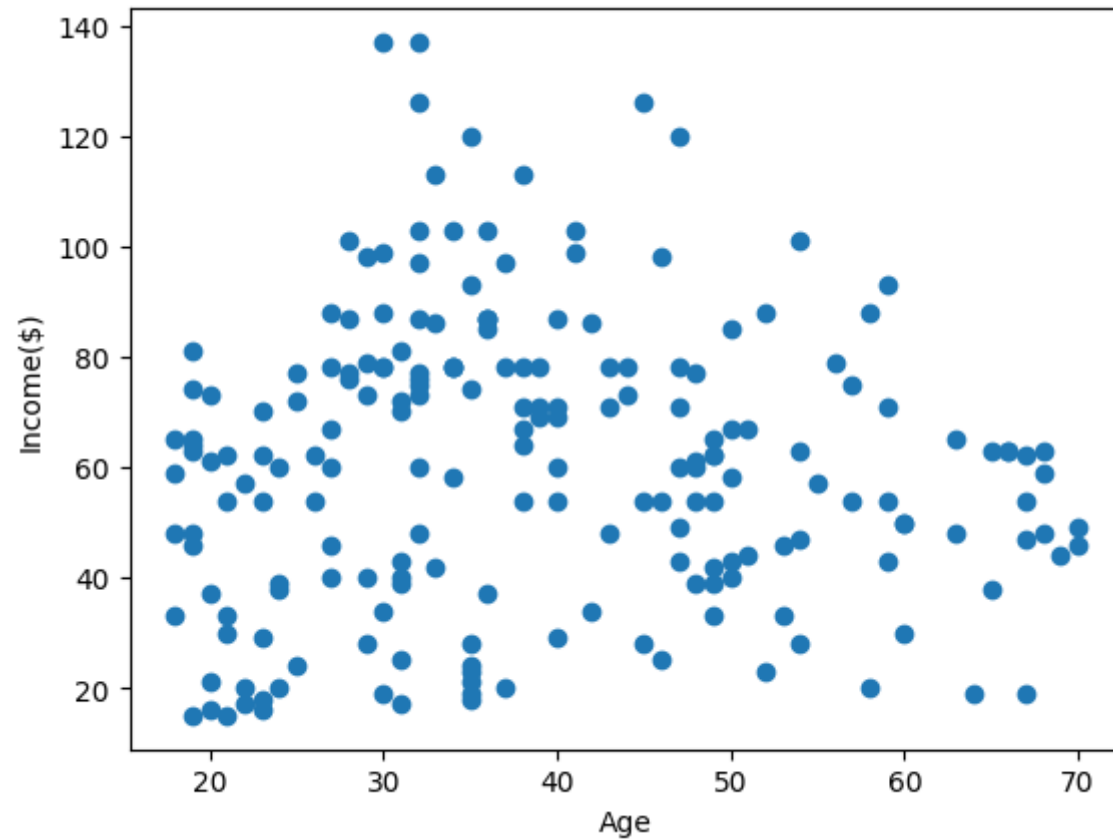
Out[7]: (200, 3)

In [8]:   ▶| df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Gender      200 non-null    object
 1   Age         200 non-null    int64
 2   Income($)   200 non-null    int64
dtypes: int64(2), object(1)
memory usage: 4.8+ KB
```

In [6]:

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

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

In [9]: ▶|
```python
from sklearn.cluster import KMeans
km=KMeans()
km
```

Out[9]:
▼ KMeans

KMeans()

In [10]: ▶|
```python
y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

C:\Users\MY HOME\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster_kmeans.py:870: Futu
reWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(

Out[10]:
```
array([2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 1, 2, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2,
       1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 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, 5, 6, 1, 5, 6, 5, 1, 5, 6,
       1, 5, 6, 6, 5, 1, 5, 5, 5, 6, 7, 7, 6, 7, 5, 7, 5, 7, 6, 7, 5, 3,
       7, 7, 5, 3, 7, 7, 3, 3, 7, 3, 7, 3, 3, 7, 5, 3, 7, 3, 5, 7, 5, 5,
       5, 3, 7, 3, 3, 3, 5, 7, 7, 7, 3, 7, 7, 7, 3, 3, 7, 7, 7, 7, 7, 7,
       3, 3, 3, 3, 7, 3, 3, 3, 7, 3, 3, 3, 3, 3, 7, 3, 3, 3, 7, 7, 7, 3,
       7, 3, 3, 3, 3, 3, 7, 3, 3, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 4, 4, 4, 4, 4, 4,
       4, 4])
```

In [11]: ▶| ```python
df["cluster"]=y_predicted
df.head()
```

Out[11]:

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

In [12]:
```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[12]: Text(0, 0.5, 'Income($)')
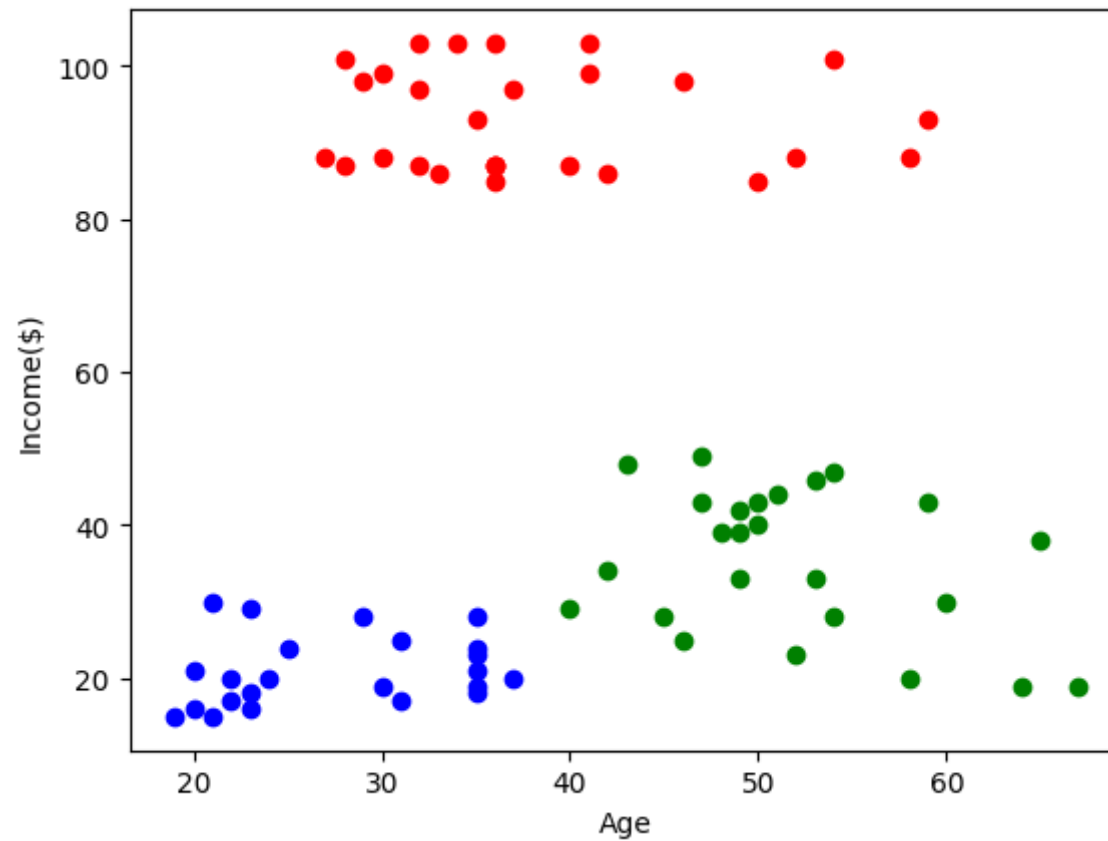
In [13]:
```python
from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
scaler.fit(df[["Income($)"]])
df["Income($)"]=scaler.transform(df[["Income($)"]])
df.head()
```

Out[13]:

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

In [14]:
```python
scaler.fit(df[["Age"]])
df["Age"]=scaler.transform(df[["Age"]])
df.head()
```

Out[14]:

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

In [15]:
```python
km=KMeans()
```

In [16]: ▶|
```python
y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

C:\Users\MY HOME\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster_kmeans.py:870: Futu
reWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(

Out[16]: array([3, 3, 3, 3, 4, 3, 4, 3, 5, 4, 5, 4, 5, 3, 4, 3, 4, 3, 0, 4, 4, 3,
        0, 4, 0, 4, 0, 4, 4, 3, 5, 3, 0, 3, 0, 3, 0, 4, 4, 3, 5, 3, 0, 4,
        0, 3, 0, 4, 4, 4, 0, 4, 4, 5, 0, 0, 0, 5, 1, 0, 5, 1, 5, 0, 5, 1,
        0, 5, 1, 4, 5, 0, 5, 5, 5, 1, 0, 0, 1, 0, 5, 6, 5, 0, 1, 0, 0, 1,
        6, 0, 5, 1, 0, 6, 6, 1, 0, 1, 0, 1, 1, 0, 5, 1, 0, 1, 5, 7, 5, 5,
        5, 1, 6, 1, 1, 1, 5, 7, 7, 7, 1, 6, 6, 6, 1, 6, 7, 6, 7, 6, 7, 6,
        1, 6, 1, 6, 7, 6, 1, 6, 7, 6, 6, 6, 1, 6, 7, 6, 6, 6, 7, 6, 7, 6,
        7, 6, 6, 6, 6, 6, 7, 6, 1, 6, 7, 6, 6, 6, 6, 6, 6, 6, 6, 6, 7, 6,
        7, 6, 7, 6, 2, 2, 7, 2, 2, 2, 7, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
        2, 2])

In [17]: ▶|
```python
df["New Cluster"]=y_predicted
df.head()
```

Out[17]:

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

In [18]:

```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($)"],color="blue")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[18]: Text(0, 0.5, 'Income($)')
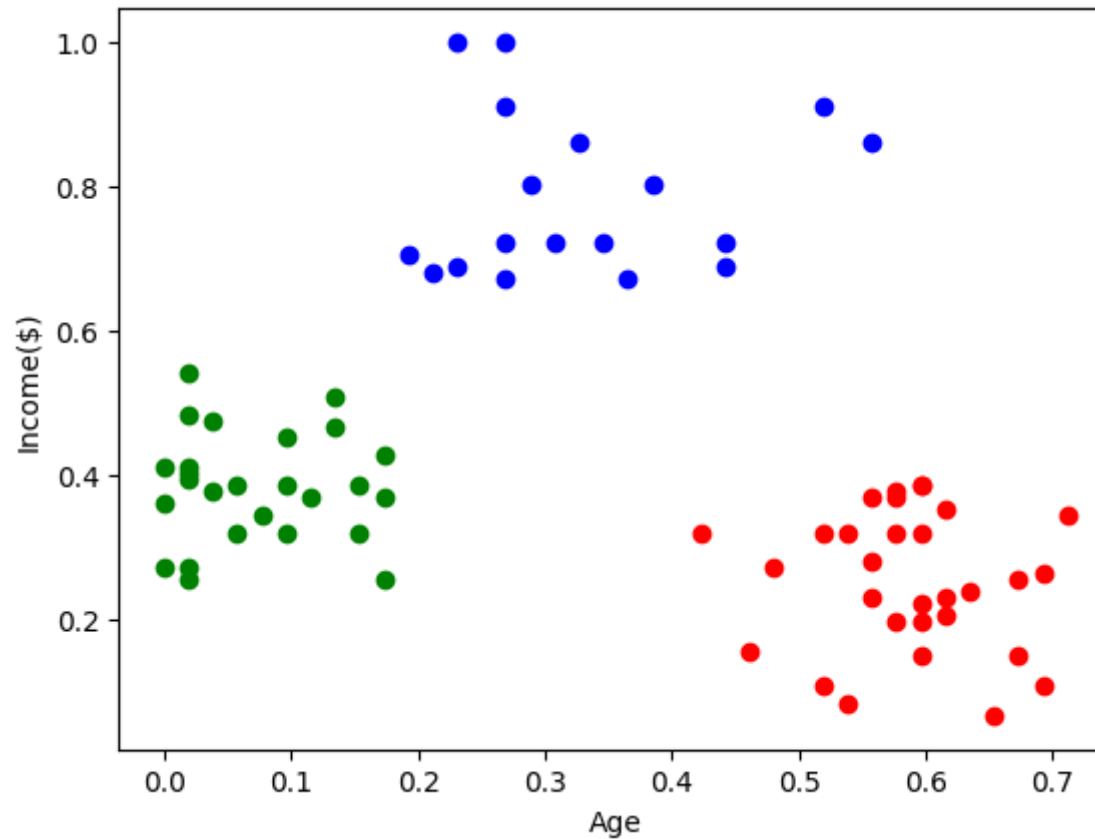
In [19]: ▶| `km.cluster_centers_`

Out[19]: array([[0.58717949, 0.25245902],
                [0.07322485, 0.38272383],
                [0.32905983, 0.78551913],
                [0.07239819, 0.08003857],
                [0.28388278, 0.1245121],
                [0.89262821, 0.27015027],
                [0.30944056, 0.50428465],
                [0.62596154, 0.51885246]])

In [21]:

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

In [22]: ▶|
```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_) #km.inertia_ will give you the value of sum of squa
print(sse)
plt.plot(k_rng,sse)
plt.xlabel("K")
plt.ylabel("Sum of Squared Error")
```

```
C:\Users\MY HOME\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: Futu
reWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\MY HOME\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: Futu
reWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\MY HOME\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: Futu
reWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\MY HOME\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: Futu
reWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\MY HOME\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: Futu
reWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\MY HOME\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: Futu
reWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\MY HOME\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: Futu
reWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\MY HOME\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: Futu
reWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
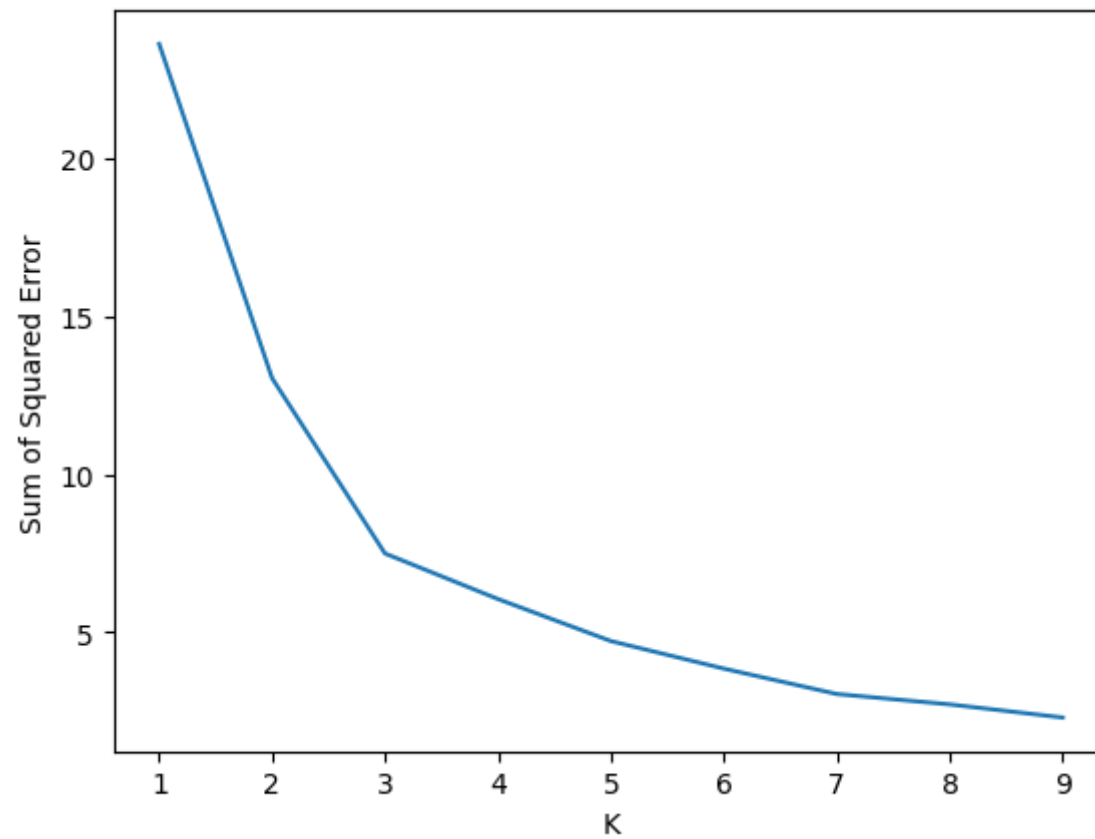itly to suppress the warning
  warnings.warn(
C:\Users\MY HOME\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: Futu
reWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(

[23.583906150363607, 13.028938428018286, 7.49302484330499, 6.055858644812547, 4.727889021361614, 3.857891822164
646, 3.055986211920202, 2.7279821120471457, 2.3135720353543285]
```

Out[22]: Text(0, 0.5, 'Sum of Squared Error')

In [ ]: ▶|