K Means Clustering

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

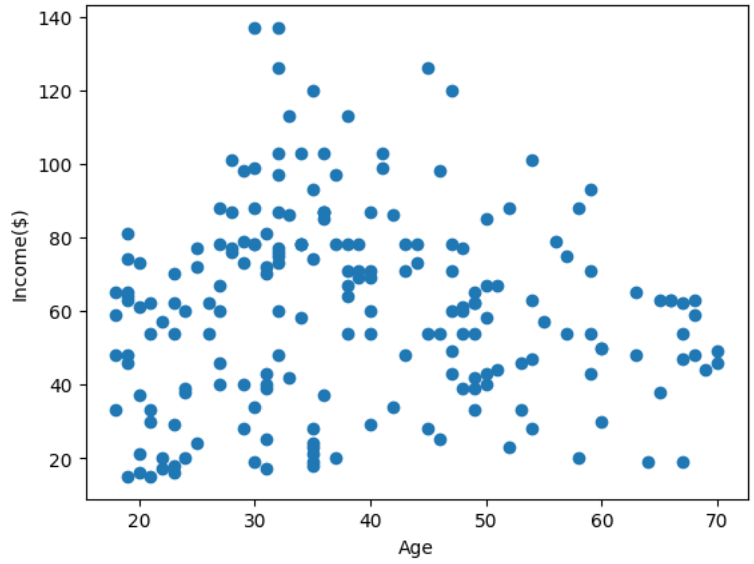
In [2]:
```
df=pd.read_csv(r"C:\Users\anu\Downloads\Income.csv")
df.head()
```

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()
```

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

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

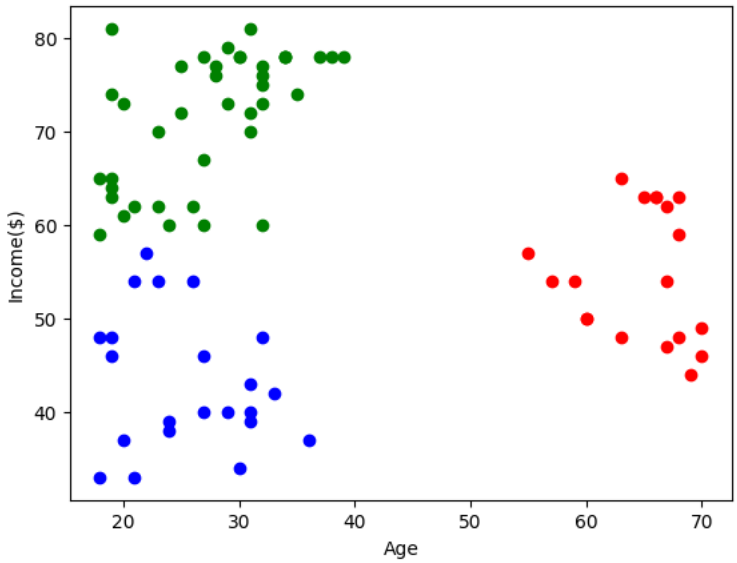
In [7]:
```
df["Cluster"]=y_predicted
df.head()
```

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($)"],color="red")
plt.scatter(df2["Age"],df2["Income($)"],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	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()
```

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

```
C:\Users\anu\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The defa
ult 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([1, 1, 1, 1, 5, 1, 5, 1, 3, 5, 3, 5, 0, 1, 5, 1, 5, 1, 0, 5, 5, 1,
       0, 5, 0, 5, 0, 5, 5, 1, 3, 1, 0, 1, 0, 1, 0, 5, 5, 1, 3, 1, 0, 5,
       0, 1, 0, 5, 5, 5, 0, 5, 5, 3, 0, 0, 0, 3, 5, 0, 3, 4, 3, 0, 3, 4,
       0, 3, 4, 5, 3, 0, 3, 3, 3, 4, 0, 0, 4, 0, 3, 2, 3, 0, 4, 0, 6, 4,
       2, 6, 3, 4, 6, 2, 2, 4, 6, 4, 6, 4, 4, 6, 3, 4, 6, 4, 3, 6, 3, 3,
       3, 4, 2, 4, 4, 4, 3, 6, 6, 6, 4, 2, 2, 2, 4, 2, 6, 2, 6, 2, 6, 2,
       4, 2, 4, 2, 6, 2, 4, 2, 6, 2, 2, 2, 4, 2, 6, 2, 2, 2, 6, 2, 6, 2,
       6, 2, 2, 2, 2, 2, 6, 2, 4, 2, 6, 2, 2, 2, 2, 2, 2, 2, 2, 2, 6, 2,
       6, 2, 6, 2, 7, 7, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7,
       7, 7])
```

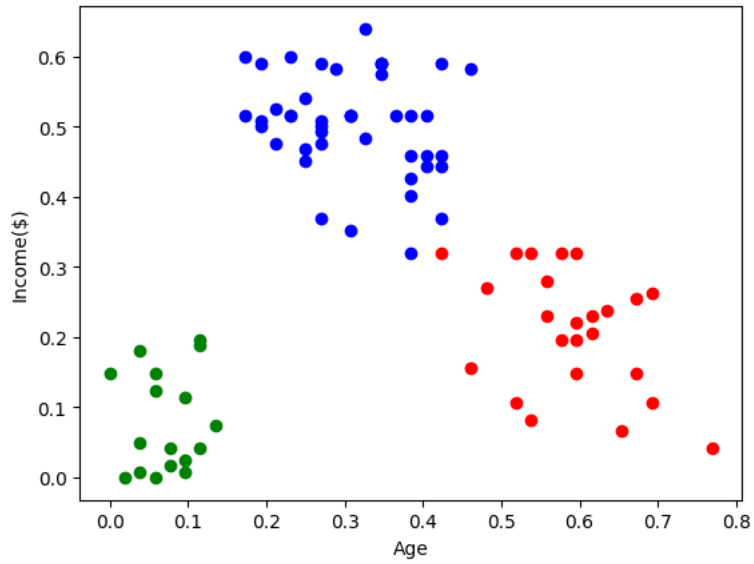
In [15]:
```
df["New Cluster"] = y_predicted
df.head()
```

Out[15]:

	Gender	Age	Income($)	Cluster	New Cluster
0	Male	0.019231	0.000000	6	1
1	Male	0.057692	0.000000	6	1
2	Female	0.038462	0.008197	6	1
3	Female	0.096154	0.008197	6	1
4	Female	0.250000	0.016393	6	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($)')
```

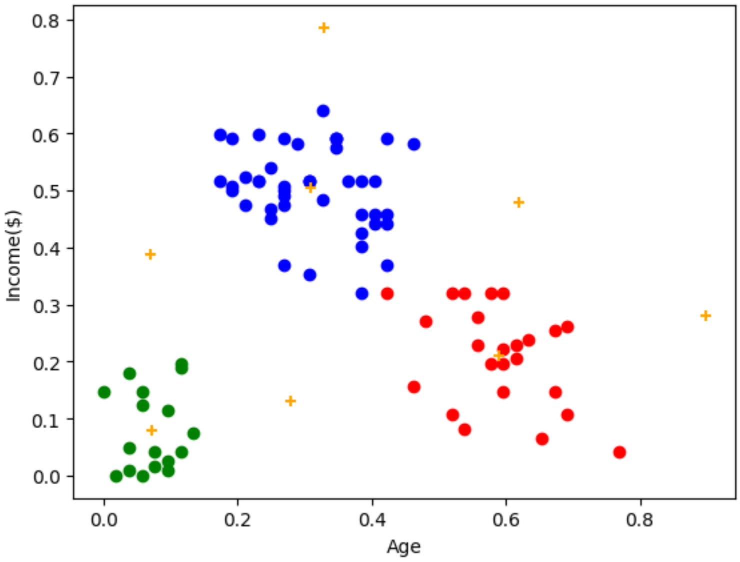


In [17]:
```
km.cluster_centers_
```

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

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($)')
```



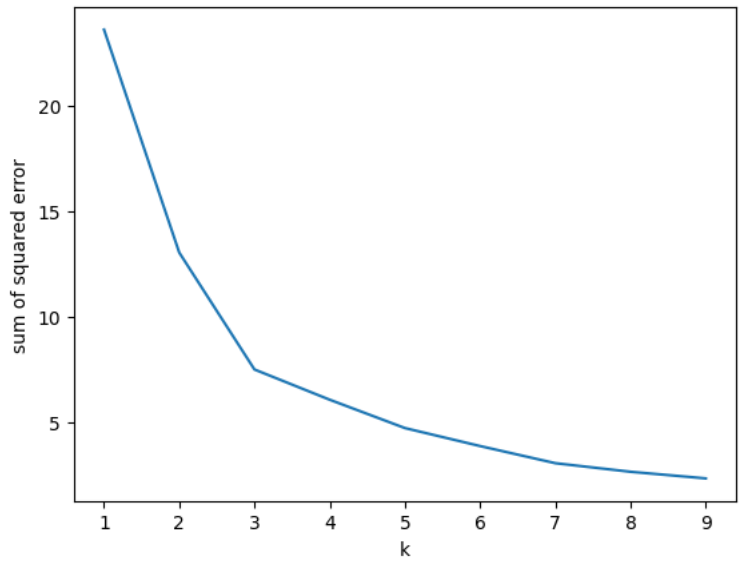
In [19]:
```
k_rng = range(1,10)
sse = []
for k in k_rng:
 km = KMeans(n_clusters=k)
 km.fit(df[["Age","Income($)"]])
 sse.append(km.inertia_)
sse
```

```
C:\Users\anu\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The defa
ult 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\anu\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The defa
ult 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\anu\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The defa
ult 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\anu\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The defa
ult 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\anu\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The defa
ult 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\anu\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The defa
ult 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\anu\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The defa
ult 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\anu\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The defa
ult 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\anu\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The defa
ult 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[19]:
```
[23.583906150363607,
 13.028938428018286,
 7.492113413237458,
 6.060567287104146,
 4.718451903648738,
 3.8659563925296765,
 3.0547174363693586,
 2.6482168727909814,
 2.3343189122246675]
```

In [20]:
```
plt.plot(k_rng,sse)
plt.xlabel("k")
plt.ylabel("sum of squared error")
```

Out[20]:
```
Text(0, 0.5, 'sum of squared error')
```



In [ ]:
```

```