In [1]:
```python
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import MinMaxScaler
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
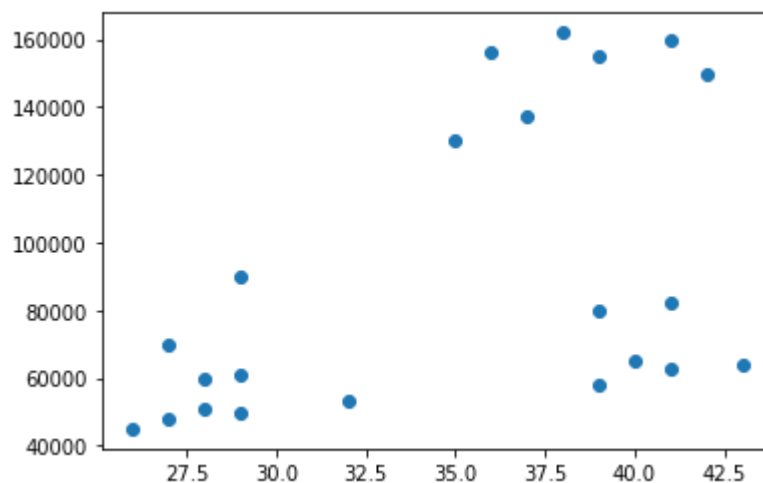```python
df=pd.read_csv('salary.csv')
```

In [3]:
```python
df.head(3)
```

Out[3]:

	Name	Age	Income($)
0	Rob	27	70000
1	Michael	29	90000
2	Mohan	29	61000

In [4]:
```python
plt.scatter(df.Age,df['Income($)'])
```

Out[4]: <matplotlib.collections.PathCollection at 0x2f74860b490>

In [5]:
```python
km=KMeans(n_clusters=3)
y_predict=km.fit_predict(df[['Age','Income($)']])
y_predict
```

Out[5]: array([0, 0, 2, 2, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 0, 0, 2])

In [6]:
```python
df['cluster']=y_predict
```

In [7]:
```python
df.head()
```

Out[7]:

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

In [8]:
```python
df1=df[df.cluster==0]
```

In [9]:
```python
df2=df[df.cluster==1]
df3=df[df.cluster==2]
```

In [10]:
```python
plt.scatter(df1.Age,df1['Income($)'],color='r')
plt.scatter(df2.Age,df2['Income($)'],color='g')
plt.scatter(df3.Age,df3['Income($)'],color='black')
plt.xlabel('Age')
plt.ylabel('Income')
```

Out[10]: Text(0, 0.5, 'Income')
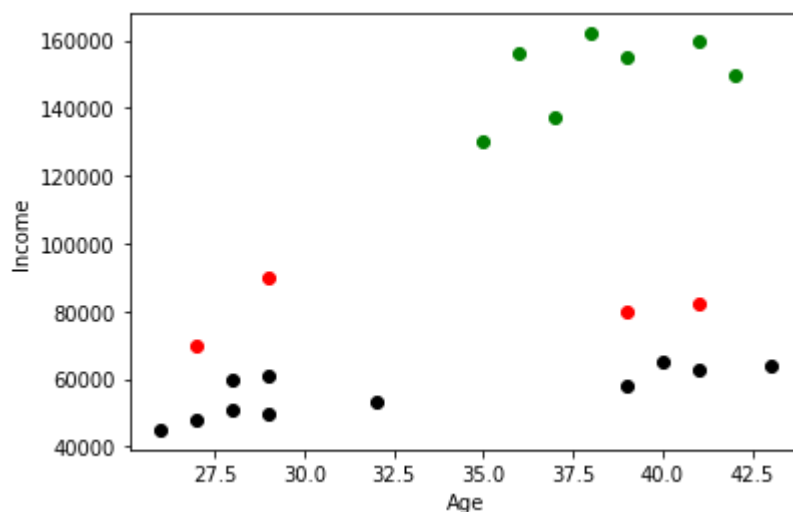
In [12]:
```python
scaler=MinMaxScaler()
scaler.fit(df[['Income($)']])
df['Income($)']=scaler.transform(df[['Income($)']])
```

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

In [15]:
```python
df.head(4)
```

Out[15]:

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

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

In [18]:
```python
df.drop('cluster',axis=1,inplace=True)
```

In [19]:
```python
df.head()
```

Out[19]:

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

In [20]:
```python
df['cluster']=y_predicted
```

In [21]:
```python
df.head()
```

Out[21]:

	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 [22]:
```python
df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
```

In [23]:
```python
plt.scatter(df1.Age,df1['Income($)'],color='r')
plt.scatter(df2.Age,df2['Income($)'],color='g')
plt.scatter(df3.Age,df3['Income($)'],color='black')
```
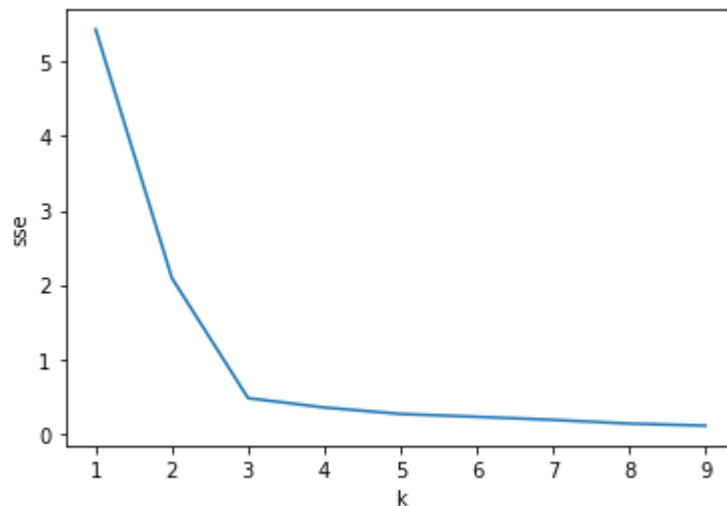
Out[23]: <matplotlib.collections.PathCollection at 0x2f749748d00>
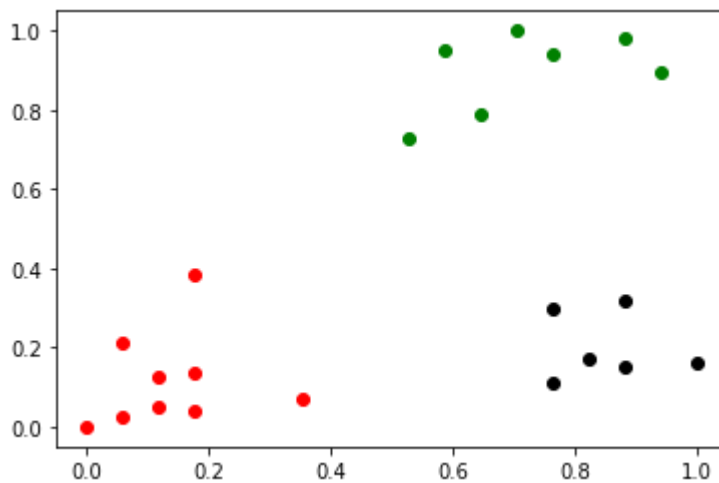
In [27]:
```python
sse=[]
k_rng=range(1,10)

for k in k_rng:
    km=KMeans(n_clusters=k)
    km.fit(df[['Age','Income($)']])
    sse.append(km.inertia_)
```

In [28]:
```python
plt.plot(k_rng,sse)
plt.xlabel("k")
plt.ylabel('sse')
```

Out[28]: Text(0, 0.5, 'sse')

In [ ]: