

```
In [29]: from sklearn.cluster import KMeans
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from matplotlib import pyplot as plt
%matplotlib inline
```

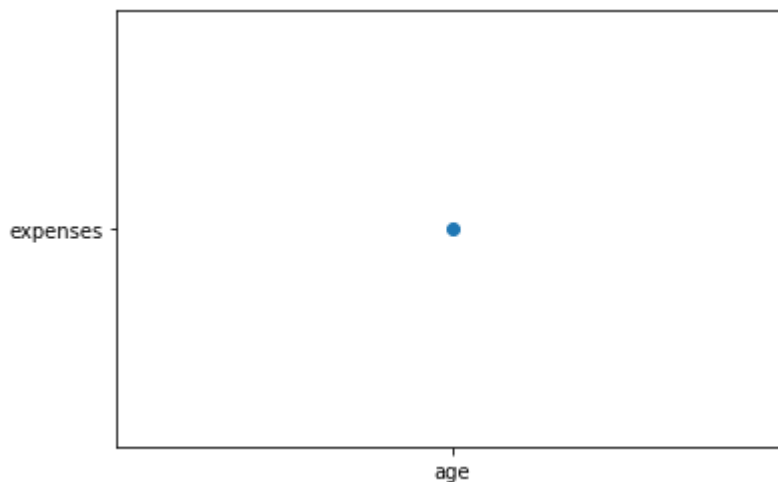
```
In [16]: model= pd.read_csv(r"C:\Users\mayur\insurance.csv")
model.head()
```

```
Out[16]:
```

	age	sex	bmi	children	smoker	region	expenses
0	19	female	27.9	0	yes	southwest	16884.92
1	18	male	33.8	1	no	southeast	1725.55
2	28	male	33.0	3	no	southeast	4449.46
3	33	male	22.7	0	no	northwest	21984.47
4	32	male	28.9	0	no	northwest	3866.86

```
In [17]: plt.scatter(['age'], ['expenses'])
```

```
Out[17]: <matplotlib.collections.PathCollection at 0x2721a2ade80>
```



```
In [18]: model.drop(['sex', 'smoker', 'region'], axis='columns', inplace=True)
```

```
In [19]: km=KMeans(n_clusters=3)
km
```

```
Out[19]: KMeans(n_clusters=3)
```

```
In [25]: km.cluster_centers_
```

```
Out[25]: array([[5.36363636e+01, 3.14987013e+01, 1.09090909e+00, 1.13948815e+04,
                2.85714286e-01],
```

```
[4.50270270e+01, 3.59621622e+01, 1.27027027e+00, 4.27835276e+04,
 1.00000000e+00],
[4.31250000e+01, 2.74718750e+01, 1.18750000e+00, 2.27514238e+04,
 2.00000000e+00],
[3.01755319e+01, 3.01047872e+01, 1.36702128e+00, 4.47392410e+03,
 0.00000000e+00],
[2.77692308e+01, 3.18871795e+01, 8.20512821e-01, 3.46627423e+04,
 1.00000000e+00],
[5.28863636e+01, 2.92090909e+01, 1.11363636e+00, 2.76867427e+04,
 1.84090909e+00],
[4.30000000e+01, 3.71800000e+01, 8.00000000e-01, 6.00182340e+04,
 1.00000000e+00],
[5.44086957e+01, 3.07252174e+01, 9.04347826e-01, 1.38040700e+04,
 2.00000000e+00],
[3.93913043e+01, 3.06987578e+01, 1.58385093e+00, 6.75312783e+03,
 0.00000000e+00],
[2.12857143e+01, 2.96986175e+01, 2.62672811e-01, 2.18701705e+03,
 0.00000000e+00],
[5.55000000e+01, 3.77812500e+01, 1.34375000e+00, 4.76480178e+04,
 1.00000000e+00],
[3.09012346e+01, 2.77370370e+01, 1.27160494e+00, 1.83821283e+04,
 2.00000000e+00],
[4.76329114e+01, 3.06284810e+01, 1.38607595e+00, 9.02481722e+03,
 0.00000000e+00],
[3.37674419e+01, 3.49046512e+01, 1.18604651e+00, 3.86456944e+04,
 1.00000000e+00]])
```

```
In [20]: y=km.fit_predict(model)
model['clusters']=y
model.clusters.unique()
y
```

```
Out[20]: array([2, 0, 0, ..., 0, 0, 2])
```

```
In [21]: model['clusters']=y
model.head()
```

```
Out[21]:
```

	age	bmi	children	expenses	clusters
0	19	27.9	0	16884.92	2
1	18	33.8	1	1725.55	0
2	28	33.0	3	4449.46	0
3	33	22.7	0	21984.47	2
4	32	28.9	0	3866.86	0

```
In [22]: model1 = model[model.clusters==0]
model2 = model[model.clusters==1]
model3 = model[model.clusters==2]
model
```

```
Out[22]:
```

	age	bmi	children	expenses	clusters
0	19	27.9	0	16884.92	2
1	18	33.8	1	1725.55	0

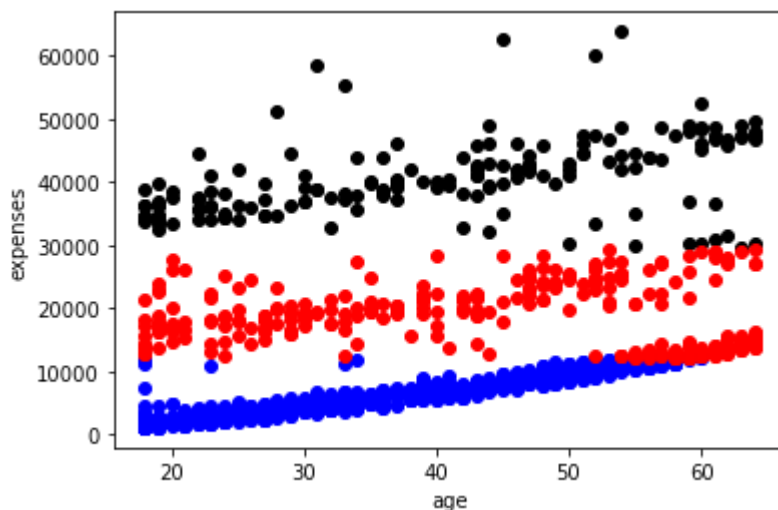
	age	bmi	children	expenses	clusters
2	28	33.0	3	4449.46	0
3	33	22.7	0	21984.47	2
4	32	28.9	0	3866.86	0
...
1333	50	31.0	3	10600.55	0
1334	18	31.9	0	2205.98	0
1335	18	36.9	0	1629.83	0
1336	21	25.8	0	2007.95	0
1337	61	29.1	0	29141.36	2

1338 rows × 5 columns

In [23]:

```
plt.scatter(model1['age'],model1['expenses'],color='blue')
plt.scatter(model2['age'],model2['expenses'],color='black')
plt.scatter(model3['age'],model3['expenses'],color='red')
plt.xlabel('age')
plt.ylabel('expenses')
```

Out[23]: Text(0, 0.5, 'expenses')

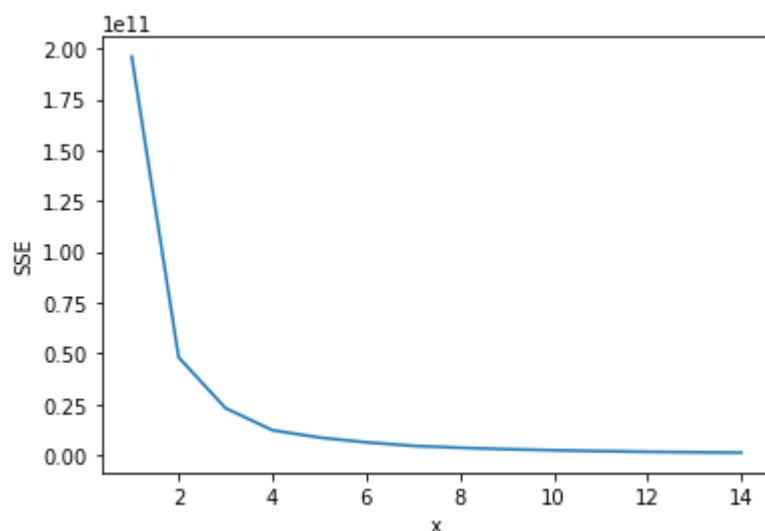


In [24]:

```
rng=range(1,15)
sse=[]
for x in rng:
    km=KMeans(n_clusters=x)
    km.fit(model)
    sse.append(km.inertia_)

plt.xlabel('x')
plt.ylabel('SSE')
plt.plot(rng,sse)
```

Out[24]: [<matplotlib.lines.Line2D at 0x2721a359760>]



```
In [31]: kmm=KMeans(n_clusters=2)
kmm
```

```
Out[31]: KMeans(n_clusters=2)
```

```
In [33]: yy=kmm.fit_predict(model)
model['clusters']=yy
model.clusters.unique()
yy
```

```
Out[33]: array([0, 0, 0, ..., 0, 0, 1])
```

```
In [34]: model11 = model[model.clusters==0]
model12 = model[model.clusters==1]
model
```

```
Out[34]:
```

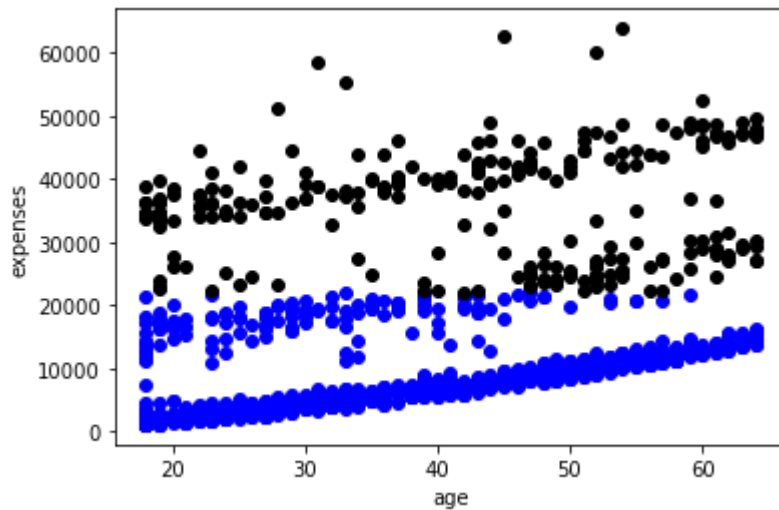
	age	bmi	children	expenses	clusters
--	-----	-----	----------	----------	----------

0	19	27.9	0	16884.92	0
1	18	33.8	1	1725.55	0
2	28	33.0	3	4449.46	0
3	33	22.7	0	21984.47	0
4	32	28.9	0	3866.86	0
...
1333	50	31.0	3	10600.55	0
1334	18	31.9	0	2205.98	0
1335	18	36.9	0	1629.83	0
1336	21	25.8	0	2007.95	0
1337	61	29.1	0	29141.36	1

1338 rows × 5 columns

```
In [36]: plt.scatter(model11['age'],model11['expenses'],color='blue')
plt.scatter(model22['age'],model22['expenses'],color='black')
#plt.scatter(model33['age'],model33['expenses'],color='red')
plt.xlabel('age')
plt.ylabel('expenses')
```

Out[36]: Text(0, 0.5, 'expenses')



In []: