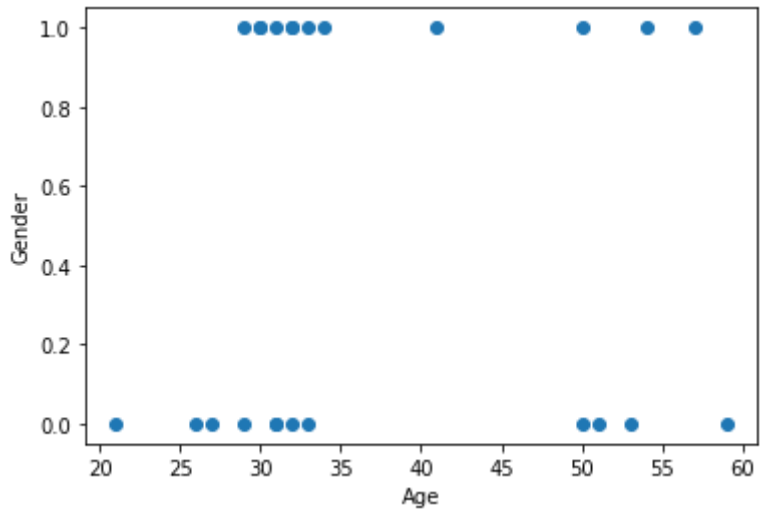
```
from sklearn.cluster import KMeans
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from matplotlib import pyplot as plt
%matplotlib inline
```

```
df = pd.read_csv("diabetes (2).csv")
df.head()
```

	Glucose	Gender	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Out
0	148	1	35	0	33.6	0.627	50	
1	85	0	29	0	26.6	0.351	31	
2	183	1	0	0	23.3	0.672	32	
3	89	0	23	94	28.1	0.167	21	
4	137	0	35	168	43.1	2.288	33	

```
plt.scatter(df.Age,df['Gender'])
plt.xlabel('Age')
plt.ylabel('Gender')
```

```
Text(0, 0.5, 'Gender')
```




```
km = KMeans(n_clusters=3)
y_predicted = km.fit_predict(df[['Age','Gender']])
y_predicted
```

```
array([1, 2, 2, 0, 2, 2, 0, 2, 1, 1, 2, 2, 1, 1, 1, 2, 2, 2, 2, 2, 0, 1,
       2, 2], dtype=int32)
```

```
df['cluster']=y_predicted
df.head()
```

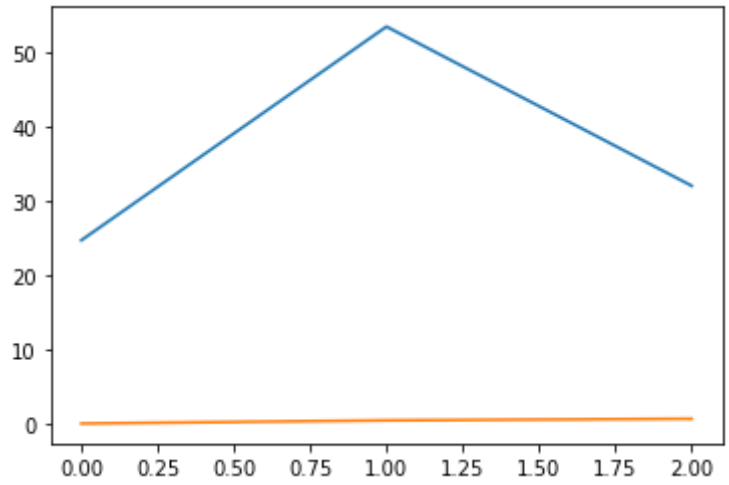
	Glucose	Gender	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Out
0	148	1	35	0	33.6	0.627	50	
1	85	0	29	0	26.6	0.351	31	
2	183	1	0	0	23.3	0.672	32	
3	89	0	23	94	28.1	0.167	21	

```
km.cluster_centers_
```

```
array([[24.66666667,  0.        ],
       [53.42857143,  0.42857143],
       [32.        ,  0.64285714]])
```

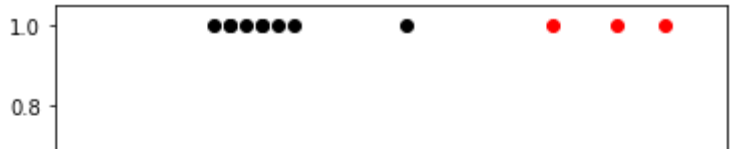
```
plt.plot(km.cluster_centers_)
```

```
[<matplotlib.lines.Line2D at 0x7fb20858e990>,
 <matplotlib.lines.Line2D at 0x7fb20858eb90>]
```




```
df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]
plt.scatter(df1.Age,df1['Gender'],color='green')
plt.scatter(df2.Age,df2['Gender'],color='red')
plt.scatter(df3.Age,df3['Gender'],color='black')
plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color='purple',marker='*',la
plt.xlabel('Age')
plt.ylabel('Gender')
plt.legend()
```

```
<matplotlib.legend.Legend at 0x7fb211cba150>
```



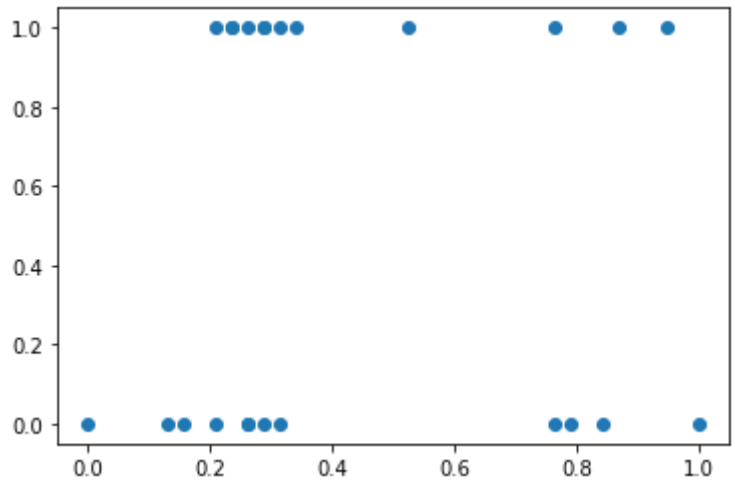

```
scaler = MinMaxScaler()

scaler.fit(df[['Gender']])
df['Gender'] = scaler.transform(df[['Gender']])

scaler.fit(df[['Age']])
df['Age'] = scaler.transform(df[['Age']])
```


```
df.head()
```

	Glucose	Gender	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	148	1.0	35	0	33.6	0.627	0.763158
1	85	0.0	29	0	26.6	0.351	0.263158
2	183	1.0	0	0	23.3	0.672	0.289474
3	89	0.0	23	94	28.1	0.167	0.000000
4	137	0.0	35	168	43.1	2.288	0.315789

```
plt.scatter(df.Age,df['Gender'])
```

```
<matplotlib.collections.PathCollection at 0x7fb208abcf50>
```




```
km = KMeans(n_clusters=3)
y_predicted = km.fit_predict(df[['Age','Gender']])
y_predicted
```

```
array([1, 0, 1, 0, 0, 1, 0, 1, 2, 1, 1, 1, 1, 2, 2, 1, 0, 1, 1, 0, 0, 2,
       1, 0], dtype=int32)
```

```
df['cluster']=y_predicted
df.head()
```

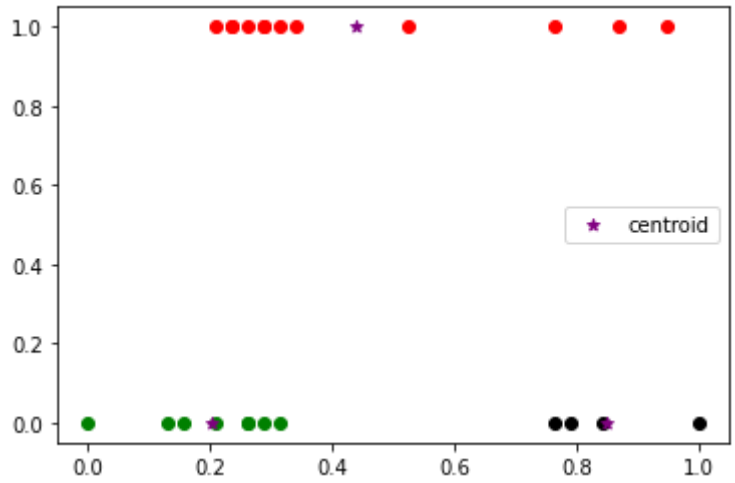
	Glucose	Gender	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	148	1.0	35	0	33.6	0.627	0.763158
1	85	0.0	29	0	26.6	0.351	0.263158
2	183	1.0	0	0	23.3	0.672	0.289474
3	89	0.0	23	94	28.1	0.167	0.000000
4	137	0.0	35	168	43.1	2.288	0.315789

```
km.cluster_centers_
```

```
array([[0.20394737, 0.        ],
       [0.44078947, 1.        ],
       [0.84868421, 0.        ]])
```

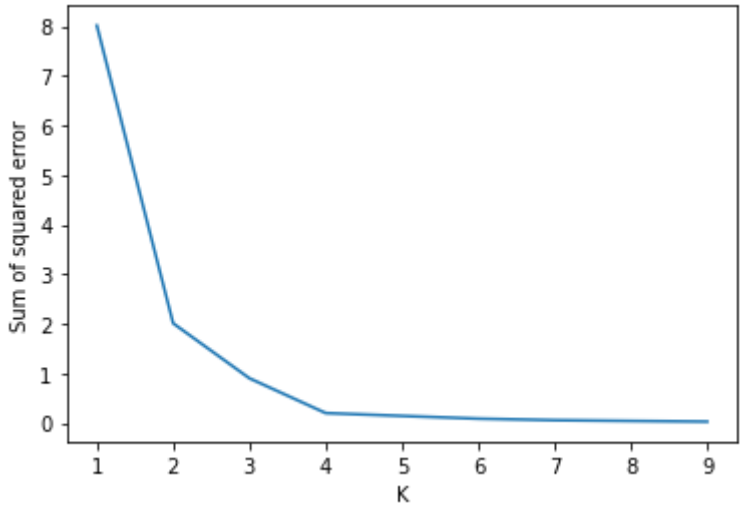
```
df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]
plt.scatter(df1.Age,df1['Gender'],color='green')
plt.scatter(df2.Age,df2['Gender'],color='red')
plt.scatter(df3.Age,df3['Gender'],color='black')
plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color='purple',marker='*',la
plt.legend()
```

```
<matplotlib.legend.Legend at 0x7fb207830790>
```




```
sse = []
k_rng = range(1,10)
for k in k_rng:
    km = KMeans(n_clusters=k)
    km.fit(df[['Age','Gender']])
    sse.append(km.inertia_)
plt.xlabel('K')
plt.ylabel('Sum of squared error')
plt.plot(k_rng,sse)
```

```
[<matplotlib.lines.Line2D at 0x7fb1fefa7e10>]
```



✓ 0s completed at 12:51 PM