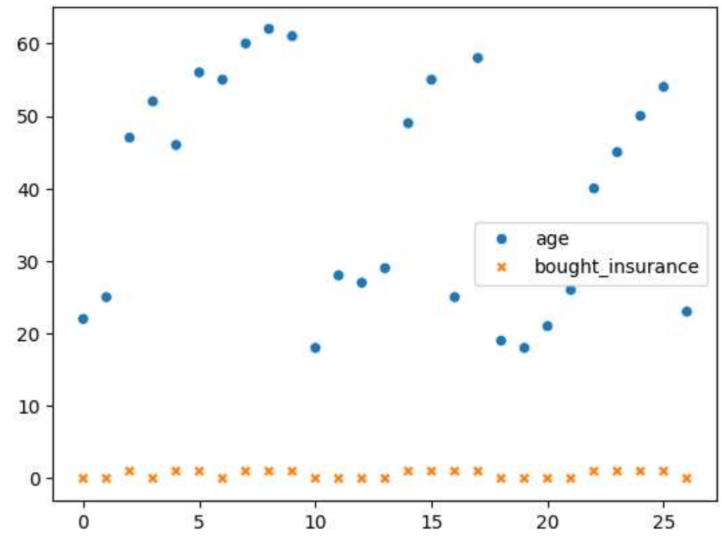
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
from sklearn import metrics
df=pd.read_csv("/insurance_data.csv")
df.head()
```

	age	bought_insurance
0	22	0
1	25	0
2	47	1
3	52	0
4	46	1

```
sns.scatterplot(df[['age','bought_insurance']])
```

```
<Axes: >
```



```
from sklearn.cluster import KMeans
KM_Model = KMeans(n_clusters=3)
cluster_predict_old=KM_Model.fit_predict(df[['age','bought_insurance']])
cluster_predict_old
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change f
  warnings.warn(
array([1, 1, 2, 0, 2, 0, 0, 0, 0, 1, 1, 1, 1, 2, 0, 1, 0, 1, 1, 1, 1,
       2, 2, 2, 0, 1], dtype=int32)
```

```
df['old_cluster'] = cluster_predict_old
df0 = df[df['old_cluster']==0]
df1 = df[df['old_cluster']==1]
df2 = df[df['old_cluster']==2]
```

```
df0
```

	age	bought_insurance	old_cluster
3	52	0	0
5	56	1	0
6	55	0	0
7	60	1	0
8	62	1	0
9	61	1	0
15	55	1	0
17	58	1	0
25	54	1	0

```
df1
```

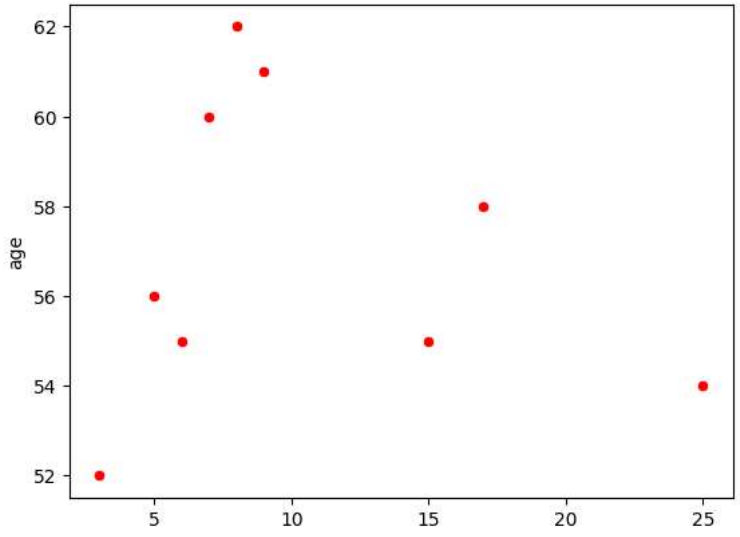
	age	bought_insurance	old_cluster
0	22	0	1
1	25	0	1
10	18	0	1
11	28	0	1
12	27	0	1
13	29	0	1
16	25	1	1
18	19	0	1
19	18	0	1
20	21	0	1
21	26	0	1
26	23	0	1

```
df2
```

	age	bought_insurance	old_cluster
2	47	1	2
4	46	1	2
14	49	1	2
22	40	1	2
23	45	1	2
24	50	1	2

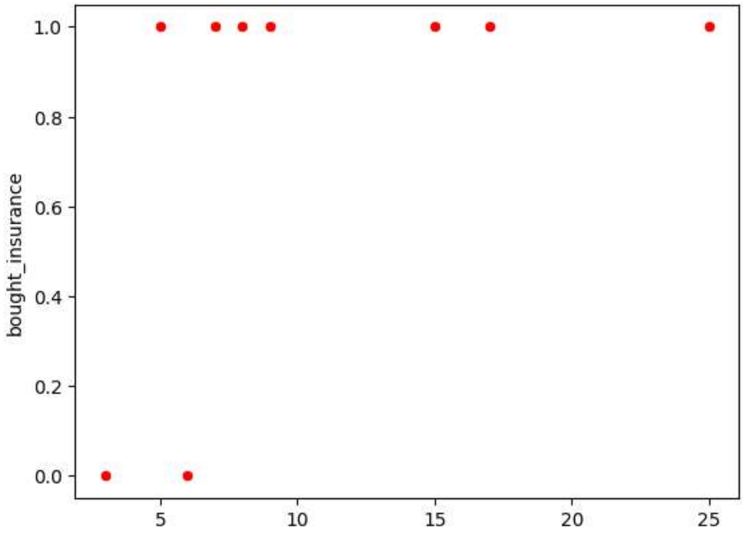
```
sns.scatterplot((df0['age']),color='red')
```

```
<Axes: ylabel='age'>
```



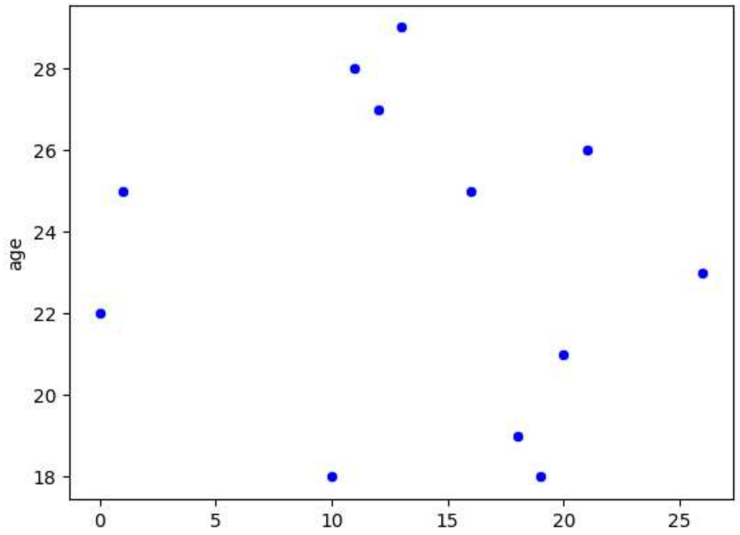
```
sns.scatterplot((df0['bought_insurance']),color='red')
```

```
<Axes: ylabel='bought_insurance'>
```



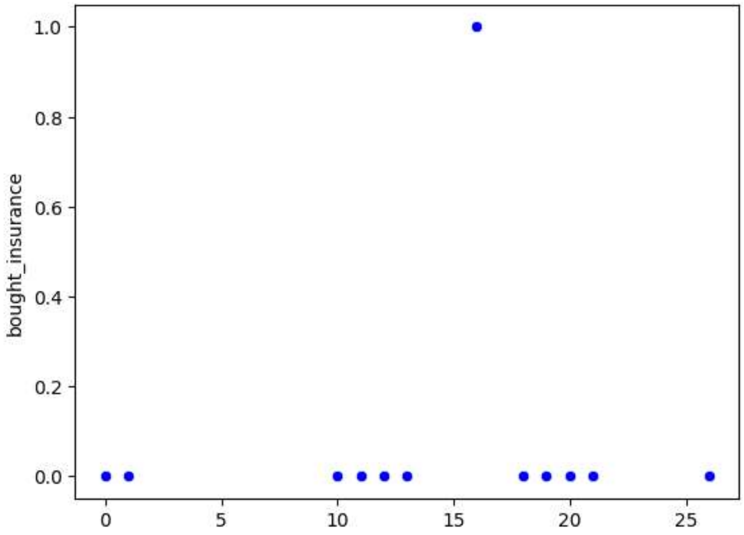
```
sns.scatterplot((df1['age']),color='blue')
```

```
<Axes: ylabel='age'>
```



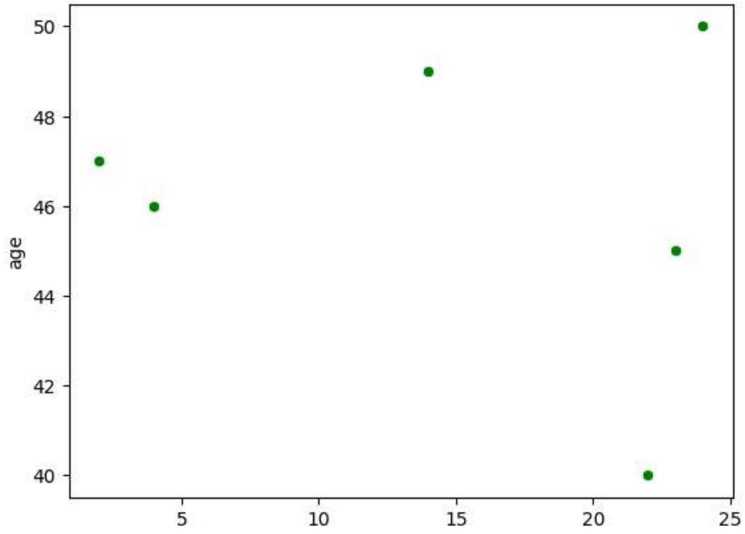
```
sns.scatterplot((df1['bought_insurance']),color='blue')
```

```
<Axes: ylabel='bought_insurance'>
```



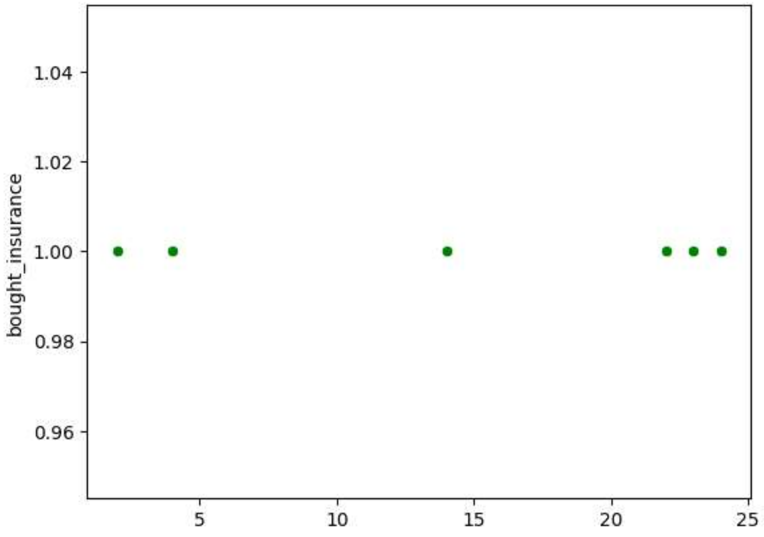
```
sns.scatterplot((df2['age']),color='green')
```

```
<Axes: ylabel='age'>
```



```
sns.scatterplot((df2['bought_insurance']),color='green')
```

```
<Axes: ylabel='bought_insurance'>
```



```
from sklearn.preprocessing import MinMaxScaler
Scaler = MinMaxScaler()
Scaler.fit(df[['age']])
df['age']=Scaler.transform(df[['age']])
Scaler.fit(df[['bought_insurance']])
df['bought_insurance']=Scaler.transform(df[['bought_insurance']])
df
```

	age	bought_insurance	old_cluster
0	0.090909	0.0	1
1	0.159091	0.0	1
2	0.659091	1.0	2
3	0.772727	0.0	0
4	0.636364	1.0	2
5	0.863636	1.0	0
6	0.840909	0.0	0
7	0.954545	1.0	0
8	1.000000	1.0	0
9	0.977273	1.0	0
10	0.000000	0.0	1
11	0.227273	0.0	1
12	0.204545	0.0	1
13	0.250000	0.0	1
14	0.704545	1.0	2
15	0.840909	1.0	0
16	0.159091	1.0	1
17	0.909091	1.0	0
18	0.022727	0.0	1
19	0.000000	0.0	1
20	0.068182	0.0	1
21	0.181818	0.0	1
22	0.500000	1.0	2
23	0.613636	1.0	2
24	0.727273	1.0	2
25	0.818182	1.0	0
26	0.113636	0.0	1

```
from sklearn.cluster import KMeans
KM_Model = KMeans(n_clusters=3)
cluster_predict_new=KM_Model.fit_predict(df[['age','bought_insurance']])
cluster_predict_new
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change f
  warnings.warn(
array([2, 2, 1, 0, 1, 1, 0, 1, 1, 1, 2, 2, 2, 2, 1, 1, 1, 1, 2, 2, 2, 2,
       1, 1, 1, 1, 2], dtype=int32)
```

```
df['new_cluster']=cluster_predict_new
df
```

	age	bought_insurance	old_cluster	new_cluster
0	0.090909	0.0	1	2
1	0.159091	0.0	1	2
2	0.659091	1.0	2	1
3	0.772727	0.0	0	0
4	0.636364	1.0	2	1
5	0.863636	1.0	0	1
6	0.840909	0.0	0	0
7	0.954545	1.0	0	1
8	1.000000	1.0	0	1
9	0.977273	1.0	0	1
10	0.000000	0.0	1	2
11	0.227273	0.0	1	2
12	0.204545	0.0	1	2
13	0.250000	0.0	1	2
14	0.704545	1.0	2	1
15	0.840909	1.0	0	1
16	0.159091	1.0	1	1
17	0.909091	1.0	0	1
18	0.022727	0.0	1	2
19	0.000000	0.0	1	2
20	0.068182	0.0	1	2
21	0.181818	0.0	1	2
22	0.500000	1.0	2	1
23	0.613636	1.0	2	1
24	0.727273	1.0	2	1
25	0.818182	1.0	0	1
26	0.113636	0.0	1	2

Generate   Using ...   10 random numbers using numpy   Close

```
df0 = df[df['new_cluster']==0]
df1 = df[df['new_cluster']==1]
df2 = df[df['new_cluster']==2]
```

```
df0
```

	age	bought_insurance	old_cluster	new_cluster
3	0.772727	0.0	0	0
6	0.840909	0.0	0	0

```
df1
```

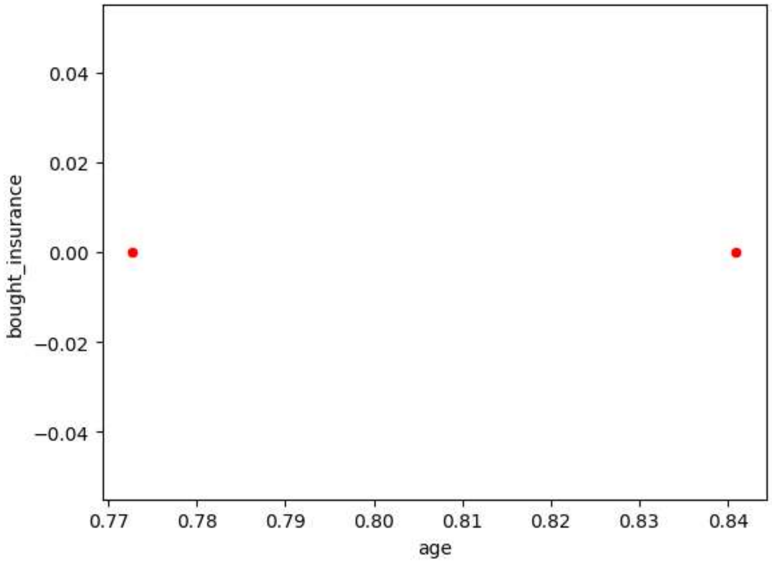
	age	bought_insurance	old_cluster	new_cluster
2	0.659091	1.0	2	1
4	0.636364	1.0	2	1
5	0.863636	1.0	0	1
7	0.954545	1.0	0	1
8	1.000000	1.0	0	1
9	0.977273	1.0	0	1
14	0.704545	1.0	2	1
15	0.840909	1.0	0	1
16	0.159091	1.0	1	1
17	0.909091	1.0	0	1
22	0.500000	1.0	2	1
23	0.613636	1.0	2	1
24	0.727273	1.0	2	1
25	0.818182	1.0	0	1

```
df2
```

	age	bought_insurance	old_cluster	new_cluster
0	0.090909	0.0	1	2
1	0.159091	0.0	1	2
10	0.000000	0.0	1	2
11	0.227273	0.0	1	2
12	0.204545	0.0	1	2
13	0.250000	0.0	1	2
18	0.022727	0.0	1	2
19	0.000000	0.0	1	2
20	0.068182	0.0	1	2
21	0.181818	0.0	1	2
26	0.113636	0.0	1	2

```
sns.scatterplot(data=df0, x='age', y='bought_insurance', color='red')
```

```
<Axes: xlabel='age', ylabel='bought_insurance'>
```



```
sns.scatterplot(data=df1, x='age', y='bought_insurance', color='blue')
```

```
<Axes: xlabel='age', ylabel='bought_insurance'>
```