Introduction to DBSCAN

Let's briefly explore visually the differences between DBSCAN and other clustering techniques, such as K-Means Clustering.

DBSCAN and Clustering Examples

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [24]:
```
blobs = pd.read_csv('data/cluster_blobs.csv')
```

In [25]:
```
blobs.head()
```

Out[25]:

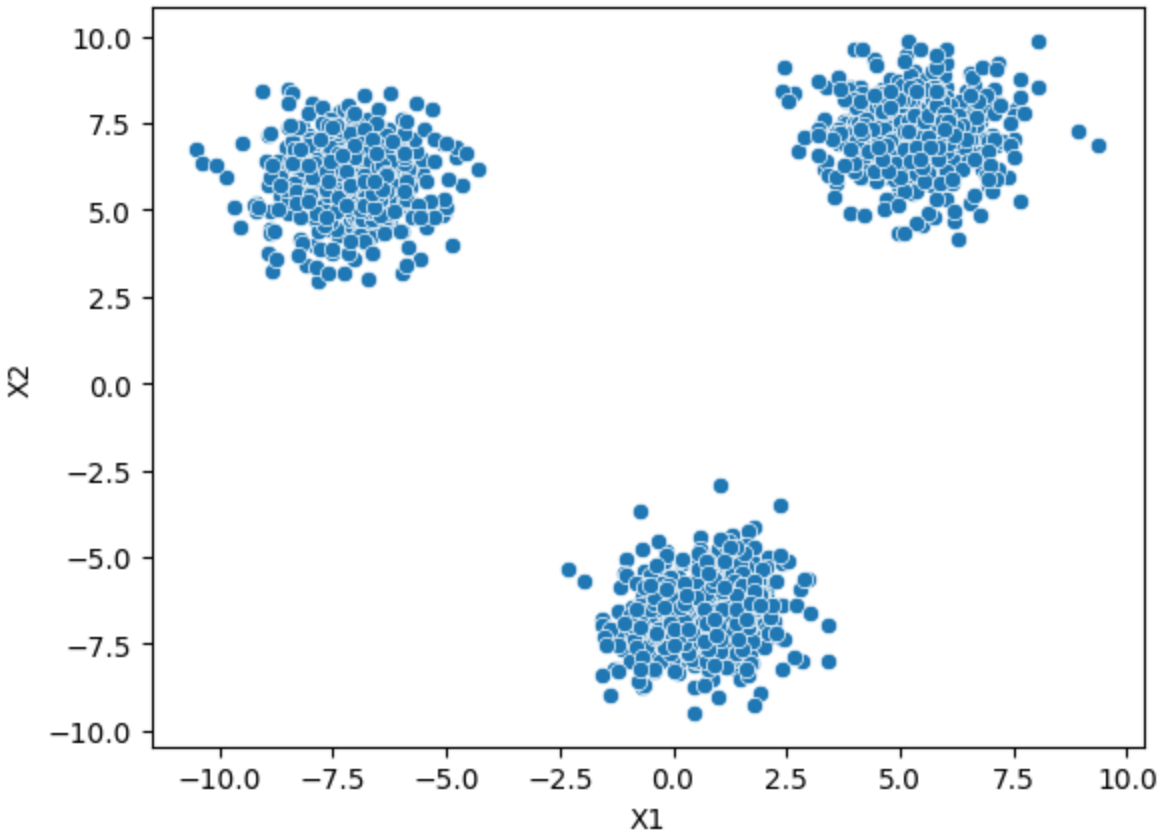
	X1	X2
0	4.645333	6.822294
1	4.784032	6.422883
2	-5.851786	5.774331
3	-7.459592	6.456415
4	4.918911	6.961479

In [22]:
```
blobs.count()
```

Out[22]:
```
X1    1500
X2    1500
dtype: int64
```

In [23]:
```
sns.scatterplot(data=blobs,x='X1',y='X2')
```

Out[23]:
```
<Axes: xlabel='X1', ylabel='X2'>
```



In [5]:
```
moons = pd.read_csv('data/cluster_moons.csv')
```

In [26]:
```
moons.count()
```

Out[26]:
```
X1    1500
X2    1500
dtype: int64
```

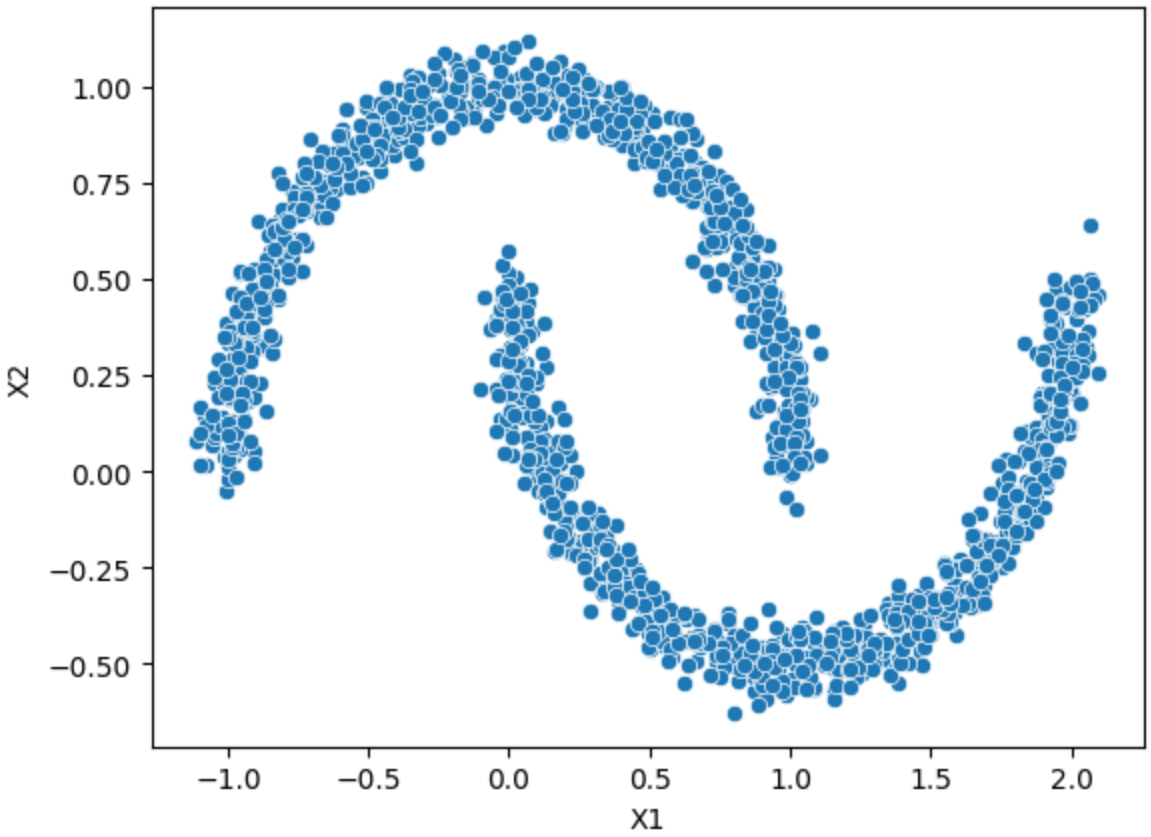
In [6]:
```
moons.head()
```

Out[6]:

	X1	X2
0	0.674362	-0.444625
1	1.547129	-0.239796
2	1.601930	-0.230792
3	0.014563	0.449752
4	1.503476	-0.389164

In [19]:
```
sns.scatterplot(data=moons,x='X1',y='X2')
```

Out[19]:
```
<Axes: xlabel='X1', ylabel='X2'>
```



In [7]:
```
circles = pd.read_csv('data/cluster_circles.csv')
```

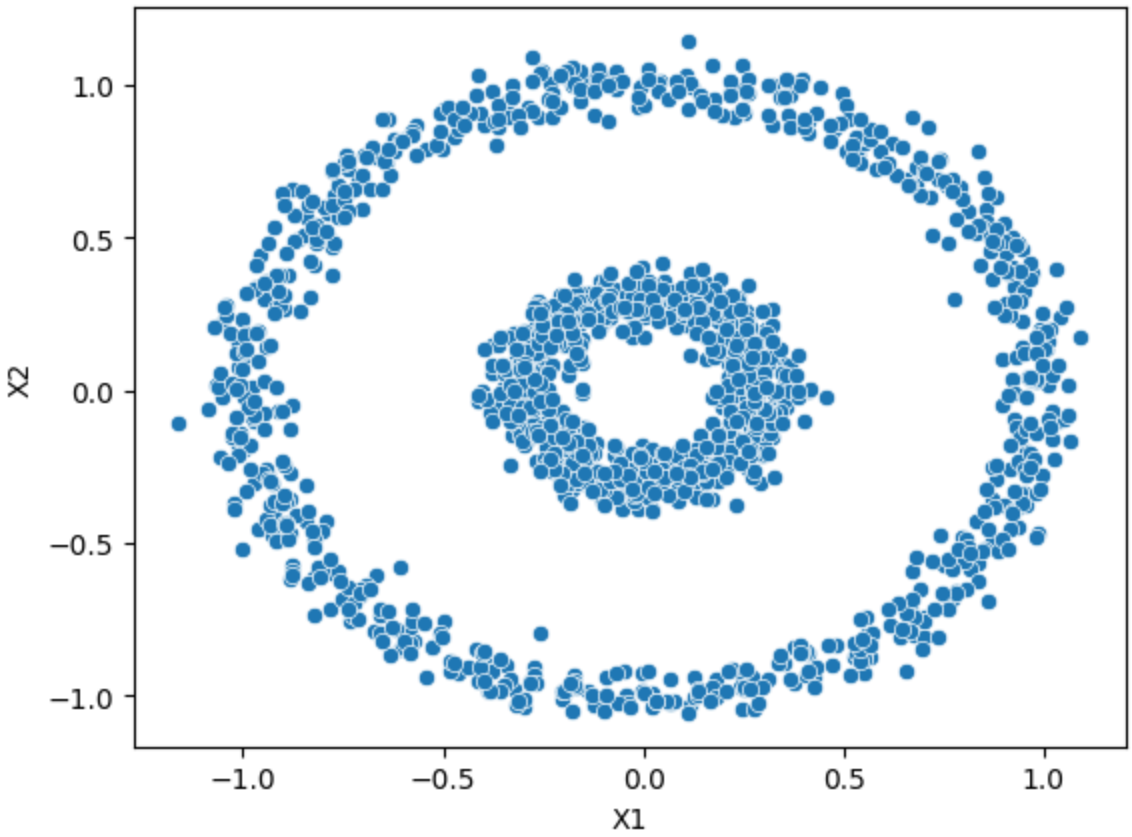
In [8]:
```
circles.head()
```

Out[8]:

	X1	X2
0	-0.348677	0.010157
1	-0.176587	-0.954283
2	0.301703	-0.113045
3	-0.782889	-0.719468
4	-0.733280	-0.757354

In [22]:
```
sns.scatterplot(data=circles,x='X1',y='X2')
```

Out[22]:
```
<Axes: xlabel='X1', ylabel='X2'>
```



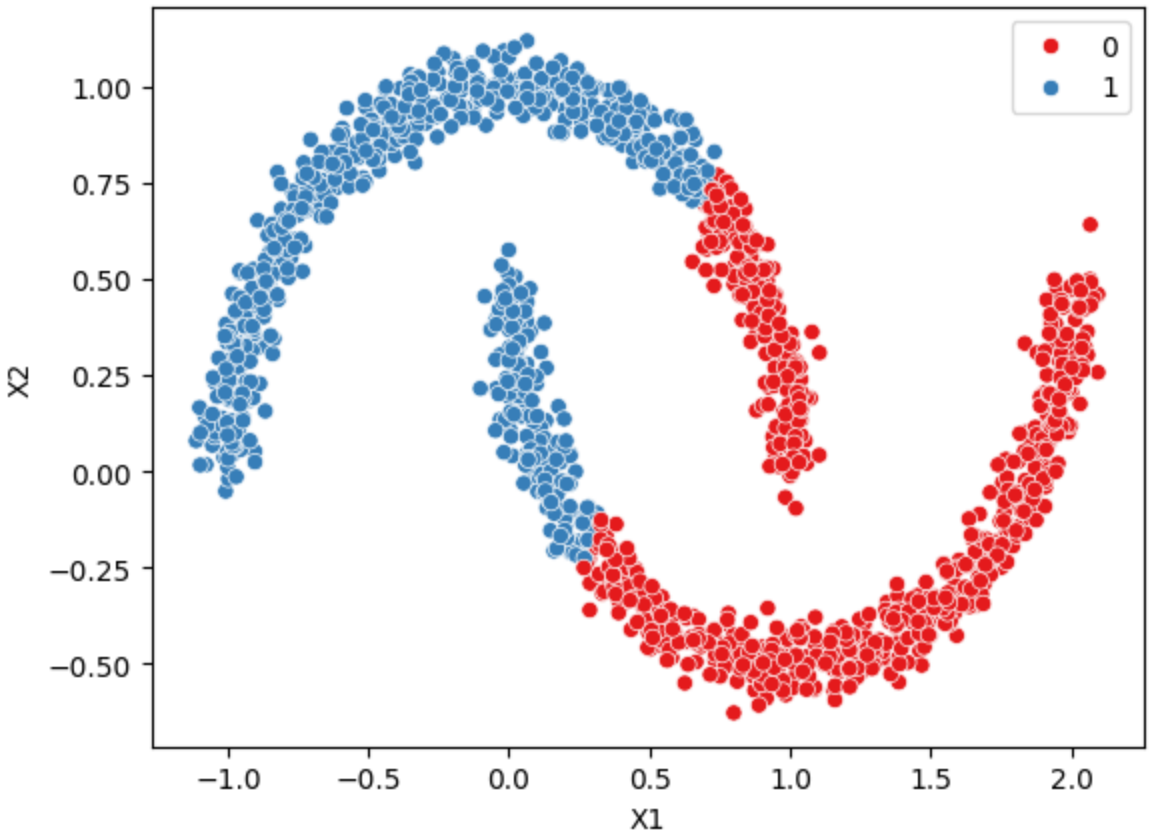
Label Discovery

In [9]:
```
def display_categories(model,data):
    labels = model.fit_predict(data)
    sns.scatterplot(data=data,x='X1',y='X2',hue=labels,palette='Set1')
```

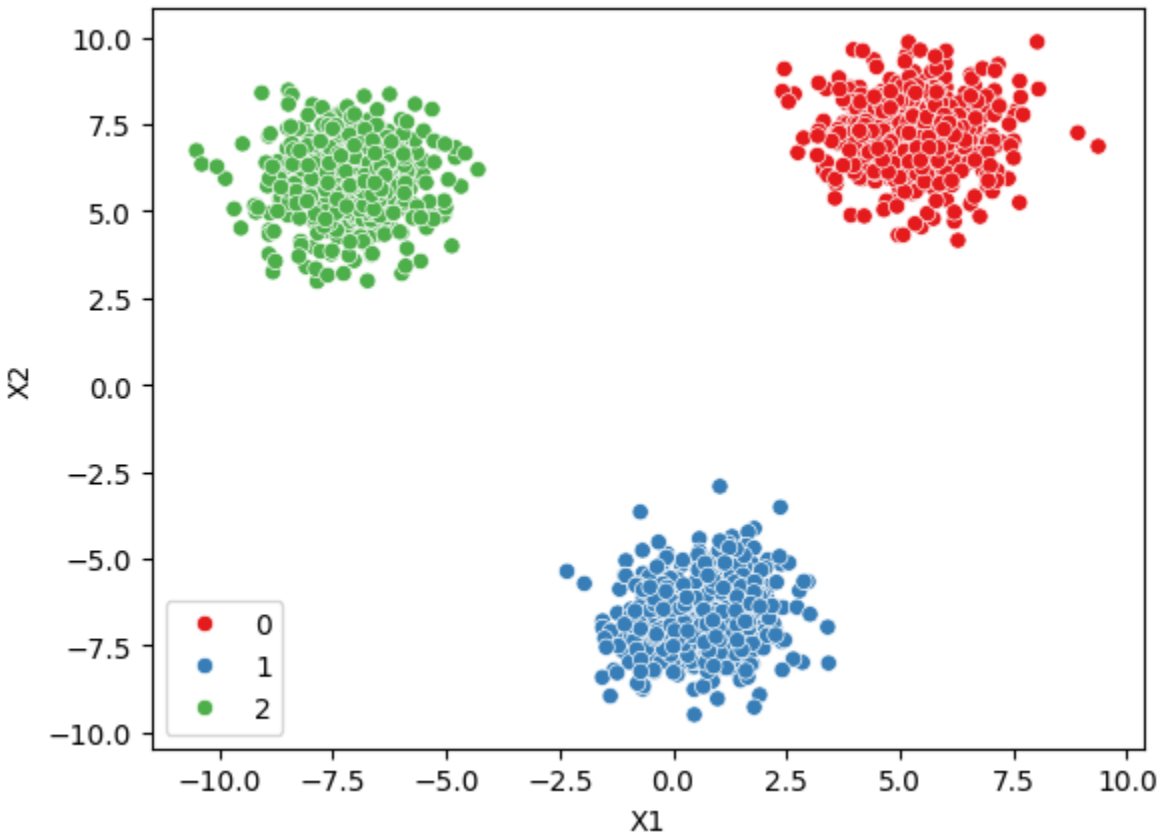
Kmeans Results

In [10]:
```
from sklearn.cluster import KMeans
model = KMeans(n_clusters = 2)
```

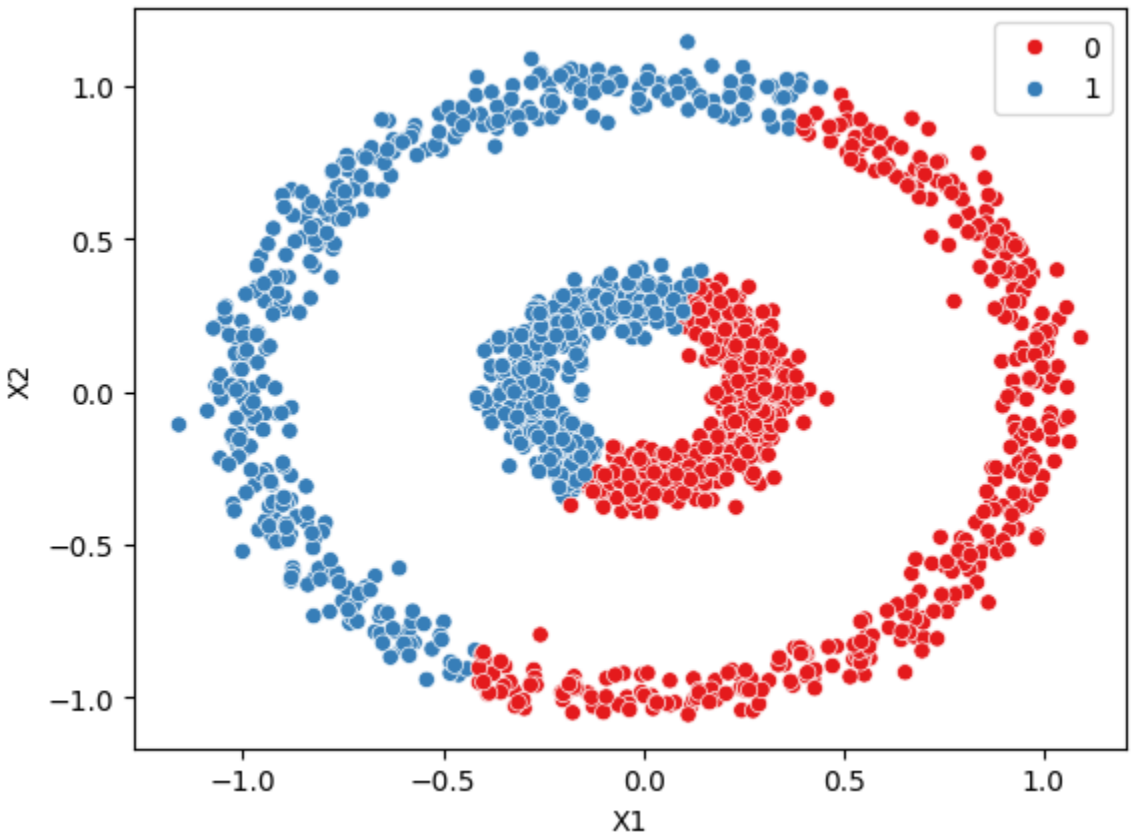
In [11]:
```
display_categories(model,moons)
```



In [12]:
```
model = KMeans(n_clusters = 3)
display_categories(model,blobs)
```



In [13]:
```
model = KMeans(n_clusters = 2)
display_categories(model,circles)
```

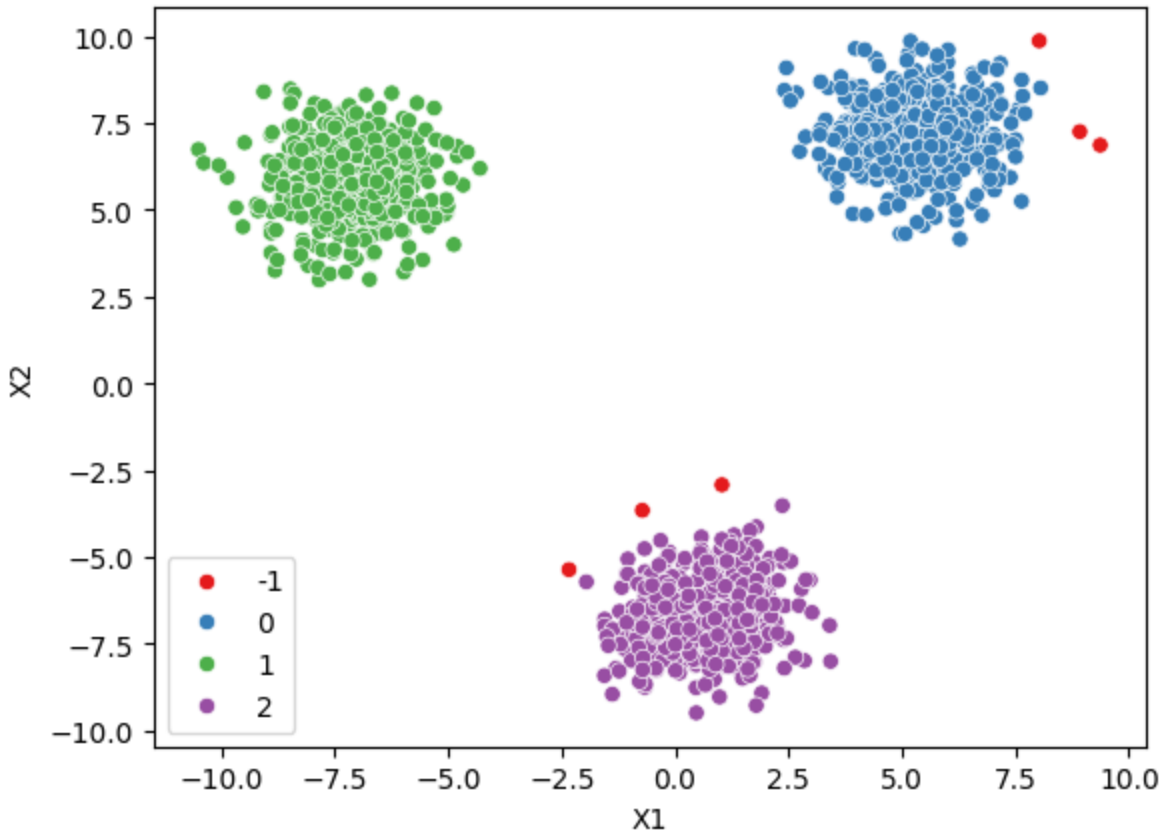


DBSCAN Results

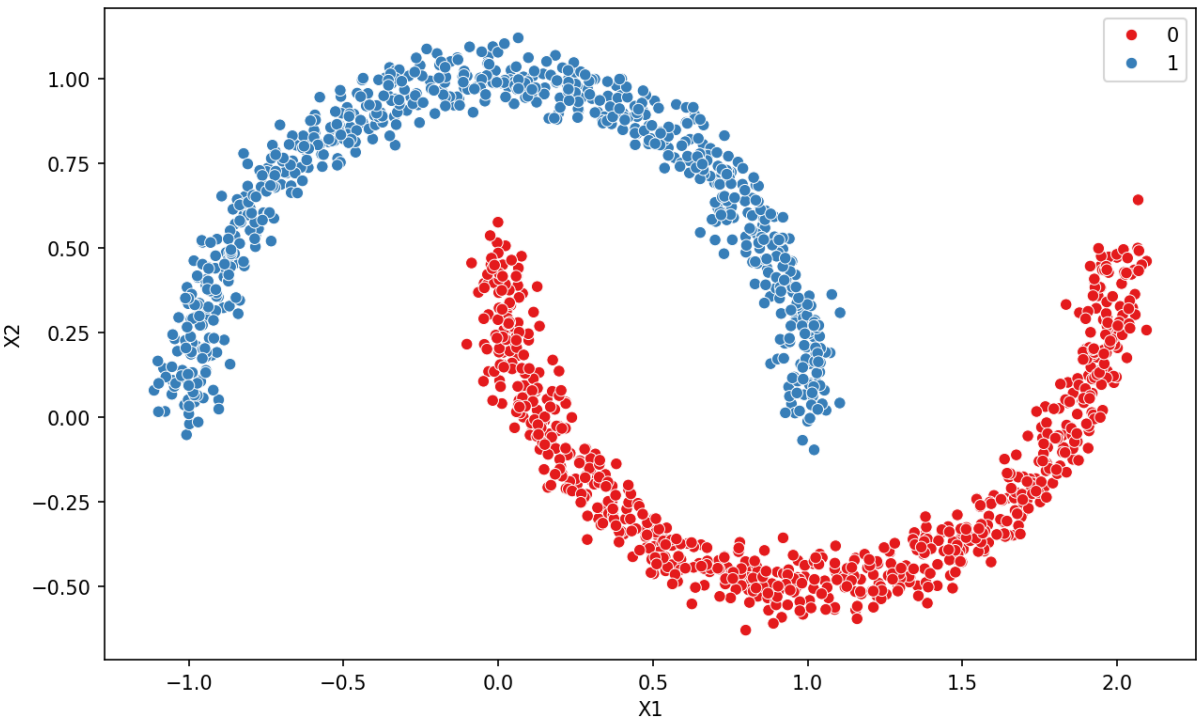
In [15]:
```
from sklearn.cluster import DBSCAN
```

In [16]:
```
model = DBSCAN(eps=0.9)
```

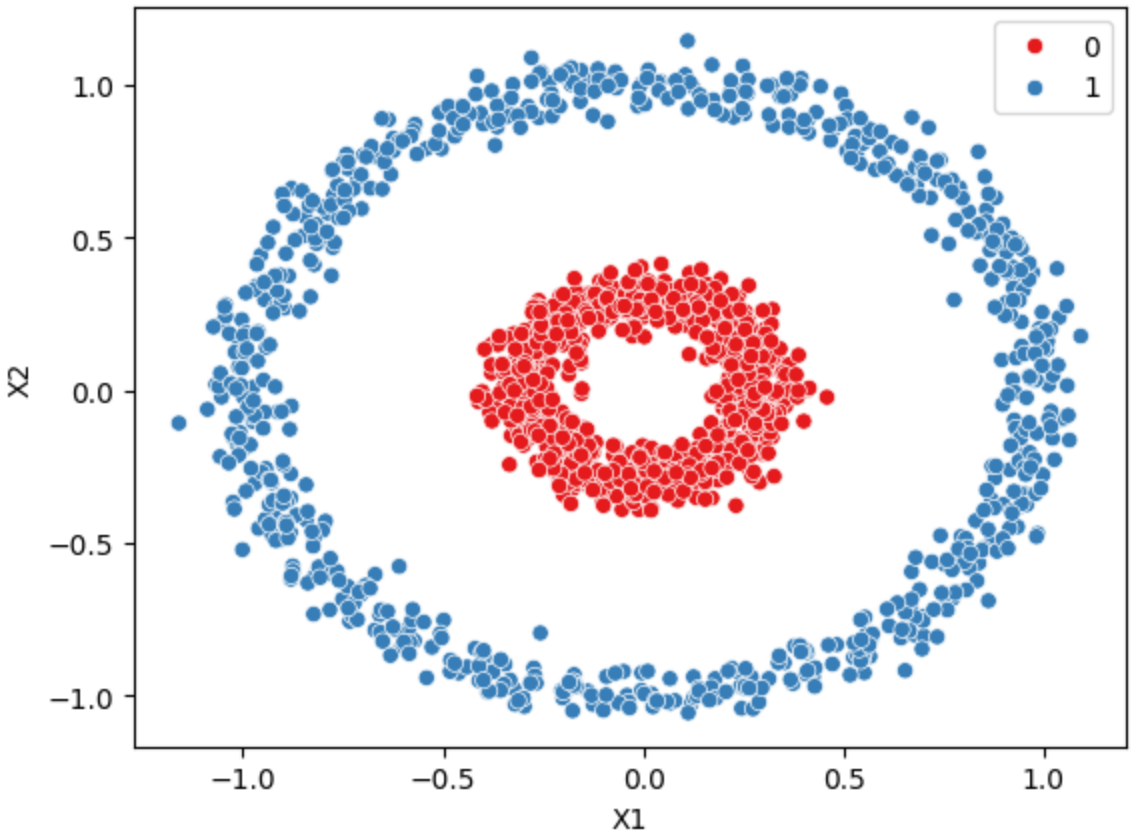
In [17]:
```
display_categories(model,blobs)
```



In [19]:
```
model = DBSCAN(eps=0.15)
plt.figure(figsize=(10,6),dpi=150)
display_categories(model,moons)
```



In [20]:
```
display_categories(model,circles)
```



In [ ]: