Name: Abhay Sharma

Class: BE - C

Moodle Id: 20102065

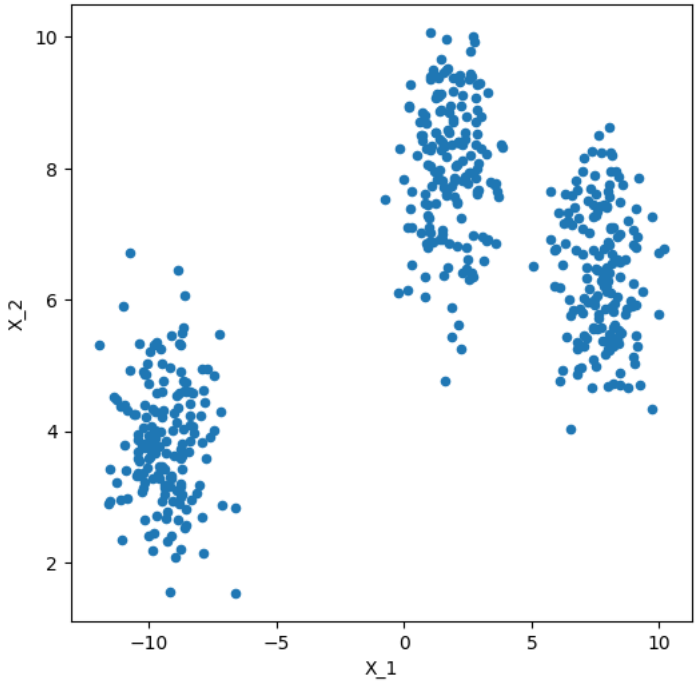
Roll No.: 28

Sub.: Applied Data Science

Experiment No.: 07

```
import numpy as np
from sklearn.datasets import make_blobs
from matplotlib import pyplot as plt
from matplotlib.pyplot import figure
from pandas import DataFrame
X,_ = make_blobs(n_samples=500, centers=3, n_features=2, random_state=20)
```

```
df=DataFrame (dict(x=X[:,0],y=X[:,1]))
fig, ax=plt.subplots (figsize=(6,6))
df.plot(ax=ax, kind='scatter',x='x',y='y')
plt.xlabel('X_1')
plt.ylabel('X_2')
plt.show()
```




```
from sklearn.cluster import DBSCAN
clustering=DBSCAN(eps=1, min_samples=5).fit(X)
cluster=clustering.labels_
len (set (cluster))
```

```
4
```

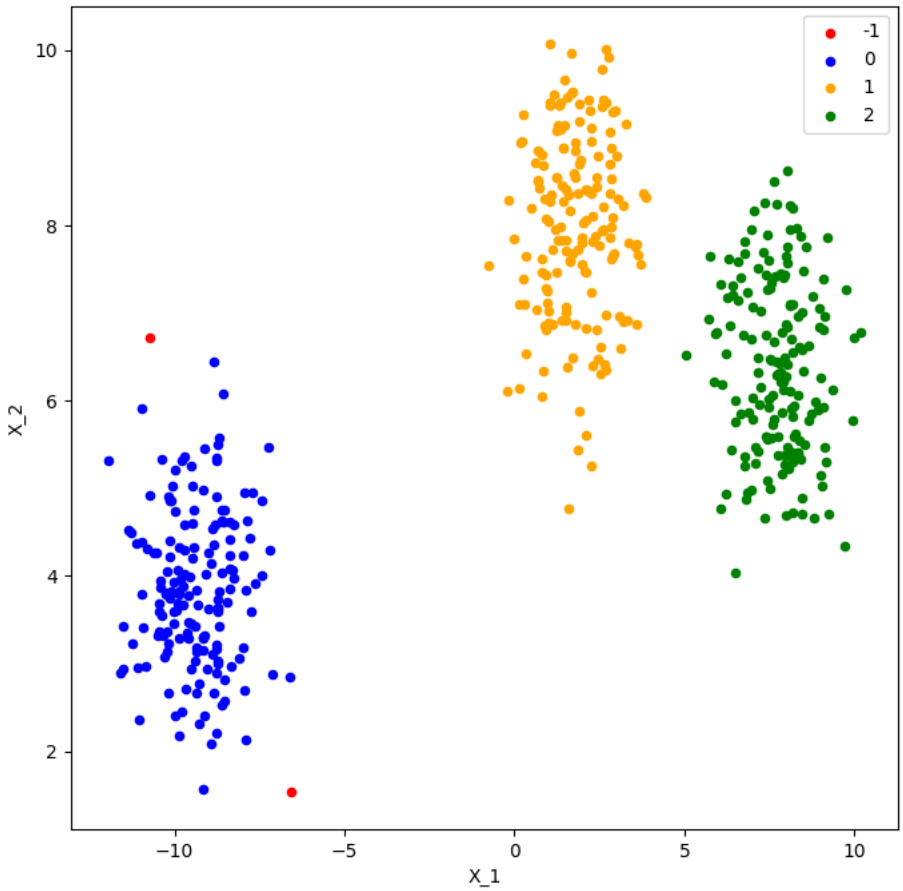
```
clustering.labels_
```

```
array([ 0,  1,  0,  0,  1,  1,  2,  2,  0,  0,  0,  1,  2,  2,  2,  0,  1,
        0,  2,  2,  0,  0,  1,  2,  1,  1,  0,  1,  0,  2,  2,  2,  1,  2,
        0,  1,  0,  2,  1,  1,  2,  1,  2,  2,  2,  0,  0,  0,  2,  2,
        2,  1,  1,  1,  0,  2,  1,  2,  2,  0,  2,  2,  1,  2,  0,  0,  2,
        0,  2,  2,  1,  2,  2,  1,  0,  0,  0,  2,  0,  1,  1,  0,  0,  0,
        2,  2,  2,  0,  2,  0,  2,  2,  0,  0,  2,  1,  0,  2,  0,  1,  2,
        1,  0,  2,  2,  2,  1,  1,  0,  0, -1,  0,  0,  0,  1,  1,  0,  1,
        2,  1,  2,  2,  0,  0,  2,  1,  2,  1,  2,  0,  2,  1,  0,  1,  1,
        2,  0,  2,  1,  1,  2,  2,  1,  1,  1,  2,  0,  2,  1,  0,  0,  0,
        0,  0,  1,  2,  0,  1,  0,  2,  0,  0,  2,  0,  0,  0,  0,  0,  0,
        2,  0,  1,  1,  2,  2,  2,  2,  0,  1,  1,  1,  1,  2,  0,  2,  1,
        2,  1,  2,  1,  1,  0,  1,  1,  1,  2,  2,  1,  1,  0,  2,  2,  1,
        0,  2,  2,  1,  2,  1,  0,  0,  1,  1,  0,  2,  1,  1,  0,  2,  0,
```

```
2,  0,  2,  0,  2,  1,  0,  1,  0,  1,  0,  2,  0,  0,  0,  2,  1,
0,  1,  1,  2,  2,  1,  2,  2,  2,  0,  1,  2,  2,  1,  0,  0,  2,
2,  2,  1,  1,  0,  2,  1,  2,  2,  2,  0,  0,  2,  1,  0,  1,  1,
2,  1,  0,  1,  0,  0,  1,  1,  1,  2,  1,  0,  2,  2,  2,  0,  2,
0,  2,  0,  1,  1,  0,  2,  2,  0,  1,  2,  1,  0,  1,  0,  0,  0,
1,  1,  0,  0,  0,  1,  1,  1,  2,  1,  1,  2,  2,  1,  2,  1,  1,
2,  0,  1,  2,  2,  1,  2,  1,  2,  1,  1,  0,  0,  0,  2,  1,  1,
0,  0,  2,  0,  0,  1,  2,  1, -1,  1,  0,  2,  1,  1,  0,  1,  0,
1,  1,  0,  1,  2,  1,  1,  0,  0,  0,  2,  0,  2,  1,  1,  1,  0,
0,  2,  2,  0,  2,  2,  1,  1,  2,  1,  1,  1,  1,  1,  2,  1,  2,
2,  0,  1,  2,  0,  1,  0,  0,  2,  0,  1,  1,  2,  1,  2,  0,  1,
2,  2,  0,  0,  0,  1,  2,  0,  0,  1,  1,  2,  1,  2,  1,  2,  2,
2,  2,  0,  0,  0,  1,  1,  0,  0,  1,  1,  0,  2,  1,  2,  1,  1,
1,  1,  2,  0,  1,  2,  1,  1,  1,  0,  0,  2,  0,  2,  0,  1,  1,
0,  0,  0,  0,  0,  2,  2,  1,  2,  0,  2,  0,  1,  1,  0,  1,  0,
0,  2,  2,  1,  2,  2,  0,  2,  0,  2,  2,  0,  0,  0,  0,  1,  1,
2,  1,  0,  2,  2,  2,  1])
```

```
def show_clusters(X, cluster):
  df=DataFrame(dict(x=X[:,0],y=X[:,1],label=cluster))
  colors={-1:'red', 0: 'blue', 1:'orange',2: 'green',3: 'pink'}
  fig, ax=plt.subplots(figsize=(8,8))
  grouped=df.groupby('label')
  for key,group in grouped:
    group.plot(ax=ax, kind='scatter',x='x',y='y', label=key, color=colors[key])
  plt.xlabel('X_1')
  plt.ylabel('X_2')
  plt.show()
```

```
show_clusters(X, cluster)
```



Double-click (or enter) to edit

Conclusion:

Python library sklearn is used to implement DBSCAN for density based outlier detection in the experiment and the matplotlib.pyplot library is used for visualizing the clusters and outliers.