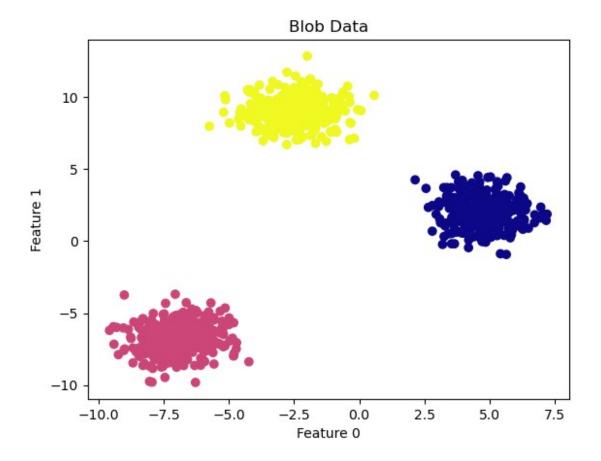
## ML ASSIGNMENT NO.-3

Question NO.1- Using Make\_blob generate data of 1000 data points with three cluster apply kmeans on it with k = 3 and use the metrics and get the accuracy (For Accuracy take reference of DBSCAN evaluation)  $\bigcirc$  Apply DBscan on Cust Segmentation Data

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import make blobs
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette score
from sklearn.metrics.cluster import silhouette score
%matplotlib inline
X, = make blobs(n samples=1000, centers=3, random state=42)
kmeans = KMeans(n clusters=3, random state=77)
y pred = kmeans.fit predict(X)
Accuracy = silhouette score(X, y pred)
print("Silhouette Score:", Accuracy)
plt.scatter(X[:,0],X[:,1],c = y pred,cmap="plasma")
plt.xlabel("Feature 0")
plt.ylabel("Feature 1")
plt.title("Blob Data")
plt.show()
Silhouette Score: 0.8435705873891368
```



**#DBSCAN On Cust Segmentation Data** 

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import DBSCAN

cust = pd.read_csv('Cust_Segmentation.csv')
cust.head()
```

Customer Id	Age	Edu	Years Employed	Income	Card Debt	0ther
0 1	41	2	6	19	0.124	
1.073 1 2	47	1	26	100	4.582	
8.218 2 3	33	2	10	57	6.111	
5.802 3 4	29	2	4	19	0.681	
0.516 4 5 8.908	47	1	31	253	9.308	

Defaulted Address DebtIncomeRatio

```
0
         0.0
              NBA001
                                   6.3
1
         0.0
              NBA021
                                  12.8
2
         1.0
              NBA013
                                  20.9
3
         0.0
              NBA009
                                   6.3
4
         0.0
              NBA008
                                   7.2
cust.columns
Index(['Customer Id', 'Age', 'Edu', 'Years Employed', 'Income', 'Card
Debt',
        Other Debt', 'Defaulted', 'Address', 'DebtIncomeRatio'],
      dtype='object')
cust['Card Debt']=cust['Card Debt'].astype(int)
cust['Other Debt']=cust['Other Debt'].astype(int)
object columns = cust.select dtypes(include = ['object']).columns
cust= cust.drop(object_columns, axis =1)
cust.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 850 entries, 0 to 849
Data columns (total 9 columns):
#
     Column
                      Non-Null Count
                                       Dtype
     -----
     Customer Id
 0
                      850 non-null
                                       int64
 1
     Age
                      850 non-null
                                       int64
 2
     Edu
                      850 non-null
                                       int64
                      850 non-null
 3
     Years Employed
                                       int64
 4
                       850 non-null
     Income
                                       int64
 5
     Card Debt
                      850 non-null
                                       int32
 6
     Other Debt
                      850 non-null
                                       int32
                                       float64
 7
     Defaulted
                      700 non-null
     DebtIncomeRatio 850 non-null
                                       float64
dtypes: float64(2), int32(2), int64(5)
memory usage: 53.2 KB
cust = cust[[ 'Age', 'Edu', 'Years Employed', 'Income', 'Card
Debt','Other Debt']]
cust.head()
   Age
        Edu
            Years Employed
                              Income Card Debt
                                                 Other Debt
0
    41
          2
                                  19
                                              0
                                                           1
                           6
                          26
1
    47
                                 100
                                              4
                                                           8
          1
                                                           5
2
    33
          2
                          10
                                  57
                                              6
3
    29
          2
                                                           0
                          4
                                  19
                                              0
                                              9
    47
          1
                         31
                                 253
                                                           8
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import DBSCAN
from sklearn.metrics import silhouette score
ss = StandardScaler()
```

```
data = ss.fit transform(cust)
clust = DBSCAN(eps = 4, min samples = 20)
clust.fit(data)
print("Accuracy for Cust Segmentation Data :",silhouette score
(cust,clust.labels ))
Accuracy for Cust Segmentation Data: 0.778946043283898
#2-Using dirtydata.csv Demonstrate all the techniques for removing the
null values
#● Replace by MEAN
#● Replace by MEDIAN
#● Replace by MODE
#• Replace by ARBITUARY VALUE
#● Replace by 0
data = pd.read_csv("dirtydata.csv")
data
    Duration
                              Pulse
                                      Maxpulse
                                                 Calories
                        Date
0
               '2020/12/01'
          60
                                 110
                                           130
                                                    409.1
1
           60
               '2020/12/02'
                                 117
                                           145
                                                    479.0
2
           60
               '2020/12/03'
                                 103
                                           135
                                                    340.0
3
           45
               '2020/12/04'
                                 109
                                           175
                                                    282.4
4
           45
               '2020/12/05'
                                 117
                                           148
                                                    406.0
5
           60
               '2020/12/06'
                                 102
                                           127
                                                    300.0
6
          60
               '2020/12/07'
                                 110
                                           136
                                                    374.0
7
         450
               '2020/12/08'
                                 104
                                           134
                                                    253.3
8
           30
               '2020/12/09'
                                 109
                                           133
                                                    195.1
9
           60
               '2020/12/10'
                                  98
                                           124
                                                    269.0
10
           60
                                           147
               '2020/12/11'
                                 103
                                                    329.3
11
                                 100
                                           120
           60
               '2020/12/12'
                                                    250.7
12
          60
               '2020/12/12'
                                 100
                                           120
                                                    250.7
13
           60
               '2020/12/13'
                                 106
                                           128
                                                    345.3
14
           60
               '2020/12/14'
                                 104
                                           132
                                                    379.3
15
           60
               '2020/12/15'
                                  98
                                           123
                                                    275.0
                                                    215.2
16
           60
               '2020/12/16'
                                  98
                                           120
17
           60
                                           120
                                                    300.0
               '2020/12/17'
                                 100
18
           45
               '2020/12/18'
                                  90
                                           112
                                                      NaN
19
           60
               '2020/12/19'
                                 103
                                           123
                                                    323.0
20
           45
                                  97
                                           125
               '2020/12/20'
                                                    243.0
21
           60
               '2020/12/21'
                                 108
                                           131
                                                    364.2
22
                                                    282.0
           45
                         NaN
                                 100
                                           119
23
               '2020/12/23'
           60
                                 130
                                           101
                                                    300.0
24
           45
               '2020/12/24'
                                 105
                                           132
                                                    246.0
25
           60
               '2020/12/25'
                                 102
                                           126
                                                    334.5
26
          60
                   20201226
                                 100
                                           120
                                                    250.0
27
           60
               '2020/12/27'
                                                    241.0
                                  92
                                           118
28
           60
               '2020/12/28'
                                 103
                                           132
                                                      NaN
29
                                           132
           60
               '2020/12/29'
                                100
                                                    280.0
```

```
30
           60
                '2020/12/30'
                                  102
                                             129
                                                      380.3
31
                                   92
                                                      243.0
           60
                '2020/12/31'
                                             115
data.isnull().sum()
             0
Duration
             1
Date
Pulse
             0
Maxpulse
             0
Calories
             2
dtype: int64
#by MEAN
print(data['Calories'].mean())
data['Calories'] = data['Calories'].fillna(data['Calories'].mean())
data
304.68
    Duration
                         Date
                               Pulse
                                       Maxpulse
                                                   Calories
0
           60
                '2020/12/01'
                                  110
                                             130
                                                     409.10
1
                '2020/12/02'
                                  117
                                             145
                                                     479.00
           60
2
                                  103
                                             135
           60
                '2020/12/03'
                                                     340.00
3
           45
                '2020/12/04'
                                  109
                                             175
                                                     282.40
4
           45
                '2020/12/05'
                                  117
                                             148
                                                     406.00
5
           60
                '2020/12/06'
                                             127
                                                     300.00
                                  102
6
           60
                '2020/12/07'
                                  110
                                             136
                                                     374.00
7
          450
                '2020/12/08'
                                  104
                                             134
                                                     253.30
8
           30
                '2020/12/09'
                                             133
                                                     195.10
                                  109
9
           60
                '2020/12/10'
                                   98
                                             124
                                                     269.00
10
           60
                '2020/12/11'
                                  103
                                             147
                                                     329.30
11
           60
                '2020/12/12'
                                  100
                                             120
                                                     250.70
12
           60
                '2020/12/12'
                                  100
                                             120
                                                     250.70
13
                '2020/12/13'
                                                     345.30
           60
                                  106
                                             128
14
           60
                '2020/12/14'
                                  104
                                             132
                                                     379.30
15
           60
                '2020/12/15'
                                   98
                                             123
                                                     275.00
16
           60
                                   98
                                             120
                '2020/12/16'
                                                     215.20
17
           60
                '2020/12/17'
                                  100
                                             120
                                                     300.00
18
           45
                '2020/12/18'
                                   90
                                             112
                                                     304.68
19
           60
                '2020/12/19'
                                  103
                                             123
                                                     323.00
20
           45
                '2020/12/20'
                                                     243.00
                                   97
                                             125
21
           60
                '2020/12/21'
                                  108
                                             131
                                                     364.20
22
           45
                                  100
                                             119
                                                     282,00
                          NaN
23
           60
                '2020/12/23'
                                  130
                                             101
                                                     300.00
24
           45
                '2020/12/24'
                                  105
                                             132
                                                     246.00
25
           60
                '2020/12/25'
                                                     334.50
                                  102
                                             126
26
           60
                    20201226
                                  100
                                             120
                                                     250.00
27
           60
                '2020/12/27'
                                   92
                                             118
                                                     241.00
28
                                                     304.68
           60
                '2020/12/28'
                                  103
                                             132
29
           60
                '2020/12/29'
                                  100
                                             132
                                                     280.00
```

```
30
           60
                '2020/12/30'
                                 102
                                             129
                                                     380.30
31
           60
                '2020/12/31'
                                   92
                                             115
                                                     243.00
# by MEDIAN
data2 = pd.read csv("dirtydata.csv")
print(data1['Calories'].median())
291.2
data2['Calories'] =
data2['Calories'].fillna(data2['Calories'].median())
data2
    Duration
                        Date
                               Pulse
                                       Maxpulse
                                                  Calories
0
           60
                '2020/12/01'
                                 110
                                             130
                                                      409.1
1
           60
                '2020/12/02'
                                 117
                                             145
                                                      479.0
2
           60
                '2020/12/03'
                                 103
                                             135
                                                      340.0
3
           45
                '2020/12/04'
                                 109
                                             175
                                                      282.4
4
           45
                '2020/12/05'
                                 117
                                             148
                                                      406.0
5
           60
                '2020/12/06'
                                 102
                                             127
                                                      300.0
6
           60
                '2020/12/07'
                                  110
                                             136
                                                      374.0
7
          450
                                 104
                                             134
                                                      253.3
                '2020/12/08'
8
           30
                '2020/12/09'
                                             133
                                                      195.1
                                 109
9
           60
                '2020/12/10'
                                   98
                                             124
                                                      269.0
                                                      329.3
10
           60
                '2020/12/11'
                                  103
                                             147
11
                '2020/12/12'
                                             120
           60
                                  100
                                                      250.7
12
           60
                '2020/12/12'
                                 100
                                             120
                                                      250.7
13
           60
                '2020/12/13'
                                  106
                                             128
                                                      345.3
14
           60
                '2020/12/14'
                                             132
                                                      379.3
                                 104
15
                                                      275.0
           60
                '2020/12/15'
                                   98
                                             123
                                                      215.2
16
           60
                '2020/12/16'
                                   98
                                             120
17
           60
                '2020/12/17'
                                 100
                                             120
                                                      300.0
18
           45
                '2020/12/18'
                                             112
                                                      291.2
                                   90
19
           60
                '2020/12/19'
                                 103
                                             123
                                                      323.0
20
           45
                '2020/12/20'
                                                      243.0
                                   97
                                             125
                                                      364.2
21
           60
                '2020/12/21'
                                 108
                                             131
22
                                             119
                                                      282.0
           45
                         NaN
                                 100
                                 130
23
                '2020/12/23'
                                                      300.0
           60
                                             101
24
           45
                '2020/12/24'
                                                      246.0
                                  105
                                             132
25
           60
                '2020/12/25'
                                  102
                                             126
                                                      334.5
26
           60
                    20201226
                                  100
                                             120
                                                      250.0
27
           60
                '2020/12/27'
                                   92
                                             118
                                                      241.0
28
           60
                '2020/12/28'
                                 103
                                             132
                                                      291.2
29
                '2020/12/29'
                                                      280.0
           60
                                 100
                                             132
30
           60
                                 102
                                             129
                                                      380.3
                '2020/12/30'
31
                                                      243.0
           60
                '2020/12/31'
                                   92
                                             115
#by MODE
data2 = pd.read csv("dirtydata.csv")
print(data1['Calories'].mode()[0])
```

```
300.0
print(data1['Calories'].mode()[0])
data1['Calories'] = data1['Calories'].fillna(data1['Calories'].mode()
[0]
data1
300.0
    Duration
                         Date
                               Pulse
                                       Maxpulse
                                                  Calories
0
                '2020/12/01'
                                  110
                                             130
                                                      409.1
           60
                '2020/12/02'
                                  117
                                             145
                                                      479.0
1
           60
2
           60
                '2020/12/03'
                                  103
                                             135
                                                      340.0
3
           45
                '2020/12/04'
                                  109
                                             175
                                                      282.4
4
           45
                '2020/12/05'
                                  117
                                             148
                                                      406.0
5
                '2020/12/06'
                                             127
                                                      300.0
           60
                                  102
                                                      374.0
6
           60
                '2020/12/07'
                                  110
                                             136
7
          450
                '2020/12/08'
                                  104
                                             134
                                                      253.3
8
                '2020/12/09'
                                             133
                                                      195.1
           30
                                  109
9
           60
                '2020/12/10'
                                   98
                                             124
                                                      269.0
10
           60
                '2020/12/11'
                                  103
                                             147
                                                      329.3
11
                                             120
           60
                '2020/12/12'
                                  100
                                                      250.7
12
           60
                '2020/12/12'
                                  100
                                             120
                                                      250.7
13
                                                      345.3
           60
                '2020/12/13'
                                  106
                                             128
14
           60
                '2020/12/14'
                                  104
                                             132
                                                      379.3
15
           60
                '2020/12/15'
                                   98
                                             123
                                                      275.0
16
           60
                '2020/12/16'
                                   98
                                             120
                                                      215.2
17
           60
                '2020/12/17'
                                  100
                                             120
                                                      300.0
18
           45
                                                      300.0
                '2020/12/18'
                                   90
                                             112
19
           60
                '2020/12/19'
                                  103
                                             123
                                                      323.0
20
           45
                '2020/12/20'
                                   97
                                             125
                                                      243.0
21
                                             131
                                                      364.2
           60
                '2020/12/21'
                                  108
22
           45
                                  100
                                             119
                                                      282.0
                          NaN
                                                      300.0
23
           60
                '2020/12/23'
                                  130
                                             101
24
           45
                '2020/12/24'
                                  105
                                             132
                                                      246.0
25
           60
                '2020/12/25'
                                  102
                                             126
                                                      334.5
26
           60
                    20201226
                                  100
                                             120
                                                      250.0
27
           60
                '2020/12/27'
                                   92
                                             118
                                                      241.0
28
                '2020/12/28'
                                  103
                                             132
                                                      300.0
           60
29
           60
                '2020/12/29'
                                  100
                                             132
                                                      280.0
30
                                  102
                                             129
                                                      380.3
           60
                '2020/12/30'
31
           60
                '2020/12/31'
                                   92
                                             115
                                                      243.0
# by ARBITARY VALUE
data4 = pd.read csv("dirtydata.csv")
arbitary value =350
```

data4['Calories'] = data4['Calories'].fillna(arbitary\_value)
data4

0	Duration	Date	Pulse	Maxpulse	Calories
0	60	'2020/12/01'	110	130	409.1
1	60	'2020/12/02'	117	145	479.0
2	60	'2020/12/03'	103	135	340.0
3 4	45 45	'2020/12/04'	109 117	175 148	282.4 406.0
5	60	'2020/12/05' '2020/12/06'	102	146 127	300.0
6	60	'2020/12/07'	102	136	374.0
7	450	'2020/12/08'	104	134	253.3
8	30	'2020/12/09'	104	134	195.1
9	60	'2020/12/10'	98	124	269.0
10	60	'2020/12/11'	103	147	329.3
11	60	'2020/12/11	100	120	250.7
12	60	'2020/12/12'	100	120	250.7
13	60	'2020/12/13'	106	128	345.3
14	60	'2020/12/14'	104	132	379.3
15	60	'2020/12/15'	98	123	275.0
16	60	'2020/12/16'	98	120	215.2
17	60	'2020/12/17'	100	120	300.0
18	45	'2020/12/18'	90	112	350.0
19	60	'2020/12/19'	103	123	323.0
20	45	'2020/12/20'	97	125	243.0
21	60	'2020/12/21'	108	131	364.2
22	45	NaN	100	119	282.0
23	60	'2020/12/23'	130	101	300.0
24	45	'2020/12/24'	105	132	246.0
25	60	'2020/12/25'	102	126	334.5
26	60	20201226	100	120	250.0
27	60	'2020/12/27'	92	118	241.0
28	60	'2020/12/28'	103	132	350.0
29	60	'2020/12/29'	100	132	280.0
30	60	'2020/12/30'	102	129	380.3
31	60	'2020/12/31'	92	115	243.0

## # by ZERO{0}

```
data5 = pd.read_csv("dirtydata.csv")
data5['Calories'] = data5['Calories'].fillna(0)
data5
```

	Duration	Date	Pulse	Maxpulse	Calories
0	60	'2020/12/01'	110	130	409.1
1	60	'2020/12/02'	117	145	479.0
2	60	'2020/12/03'	103	135	340.0
3	45	'2020/12/04'	109	175	282.4
4	45	'2020/12/05'	117	148	406.0
5	60	'2020/12/06'	102	127	300.0
6	60	'2020/12/07'	110	136	374.0

7	450	'2020/12/08'	104	134	253.3
8	30	'2020/12/09'	109	133	195.1
9	60	'2020/12/10'	98	124	269.0
10	60	'2020/12/11'	103	147	329.3
11	60	'2020/12/12'	100	120	250.7
12	60	'2020/12/12'	100	120	250.7
13	60	'2020/12/13'	106	128	345.3
14	60	'2020/12/14'	104	132	379.3
15	60	'2020/12/15'	98	123	275.0
16	60	'2020/12/16'	98	120	215.2
17	60	'2020/12/17'	100	120	300.0
18	45	'2020/12/18'	90	112	0.0
19	60	'2020/12/19'	103	123	323.0
20	45	'2020/12/20'	97	125	243.0
21	60	'2020/12/21'	108	131	364.2
22	45	NaN	100	119	282.0
23	60	'2020/12/23'	130	101	300.0
24	45	'2020/12/24'	105	132	246.0
25	60	'2020/12/25'	102	126	334.5
26	60	20201226	100	120	250.0
27	60	'2020/12/27'	92	118	241.0
28	60	'2020/12/28'	103	132	0.0
29	60	'2020/12/29'	100	132	280.0
30	60	'2020/12/30'	102	129	380.3
31	60	'2020/12/31'	92	115	243.0