ASSIGNMENT 2

FOUNDATION OF MACHINE LEARNING

SUBJECT CODE: CS564

Submitted by

Kiran Kumar Dugana

Roll no. 2211MC10

MTech 1st year

1. IMPORTATION OF LIBRARIES:

I have important all libraries required as below

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import metrics
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import scipy.cluster.hierarchy as hcluster
from sklearn.cluster import AgglomerativeClustering
from sklearn.preprocessing import MinMaxScaler
import matplotlib.pyplot as plt
from sklearn.cluster import DBSCAN
from sklearn import metrics
from sklearn import datasets
```

2. DATA CLEANING:

I checked NAN and NULL values presence

```
No of NAN values in given blobs_set_1
No of NAN values in given circles_set_2
No of NAN values in given moons_set_3
0
0
0

No of NULL values in given data-blobs
No of NULL values in given data_circles
No of NULL values in given data_moons
0
0
0
```

I also checked if any duplicated present in all three datasets

```
Printing duplicated rows in blobs data
```

X1	X2

```
Printing duplicated rows in circles data
```

X1	X2

```
Printing duplicated rows in moons data
```

X_1	X_2

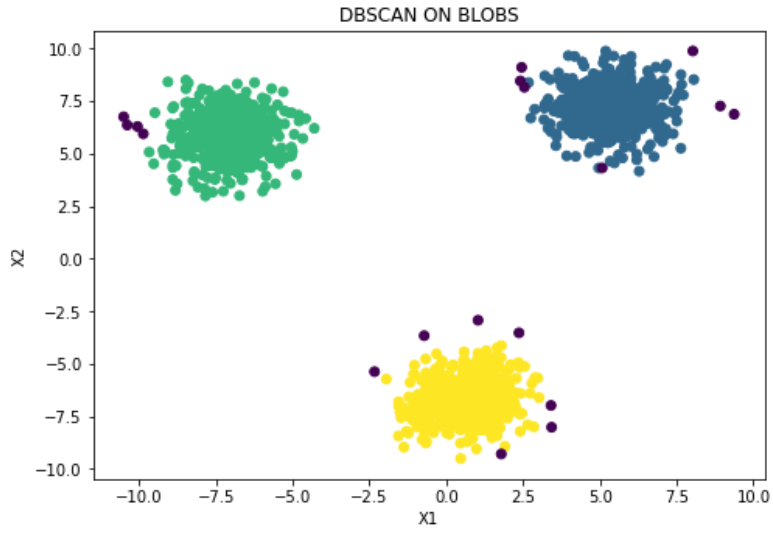
3. DBSCAN AND KMEANS ON BLOBS

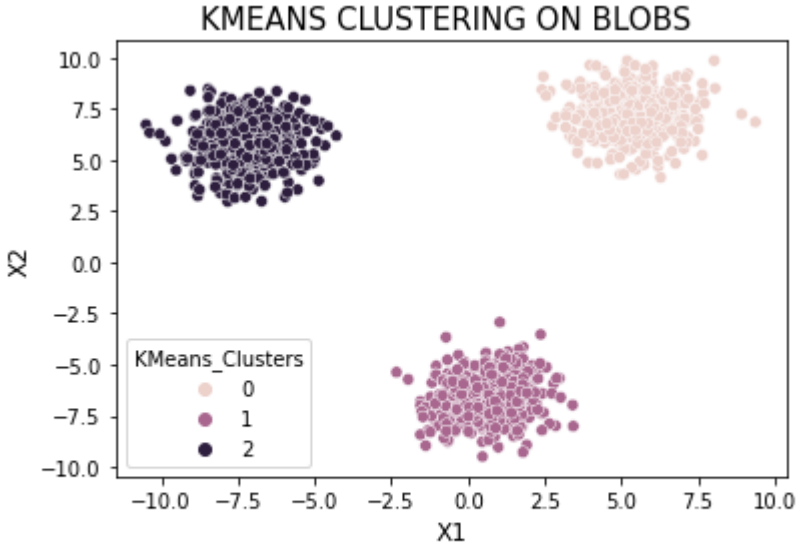
Both the algorithms are working fine on global circular shapes.

We are able to separate out the outliers using DBSCAN but not with KMEANS.

Silhouette score remained same in both the cases because clusters with corresponding points remained same in both cases.

```
Silhouette Coefficient:0.82
```






```
Printing shilhoutte scre using KMeans
0.8638532520315105
```

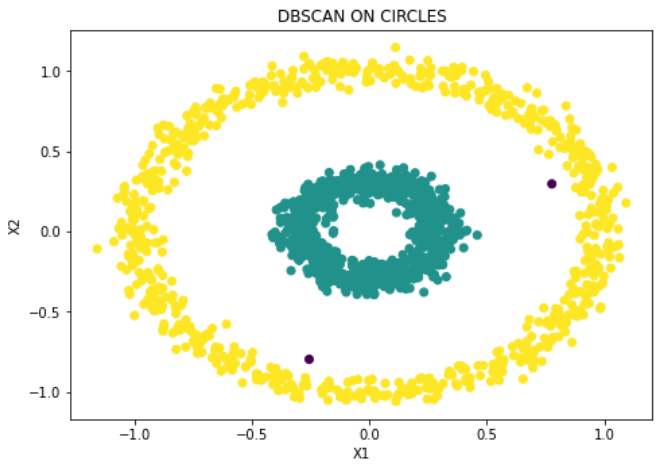
6. DBSCAN AND KMEANS ON CIRCLES

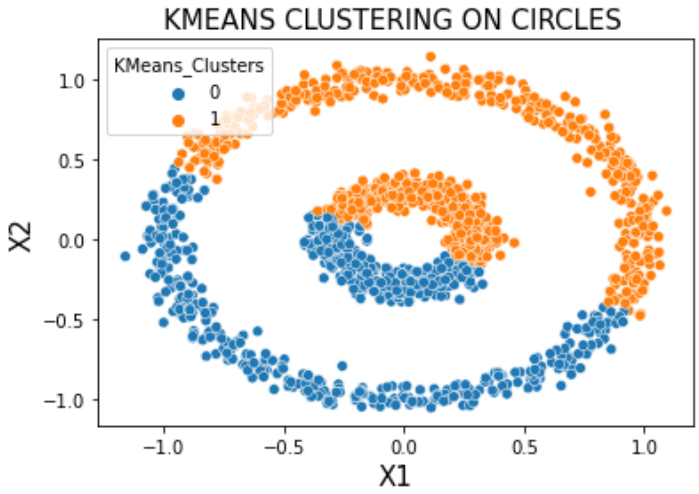
DBSCAN working better on circles compared to KMEANS

We are able to separate out the outliers using DBSCAN but not with KMEANS.

Silhouette score is much lesser in case of DBASCAN compared to KMEANS because of circle shape.

```
Silhouette Coefficient:0.16
```






```
Printing shilhoutte scre using KMeans
0.5068292236193002
```

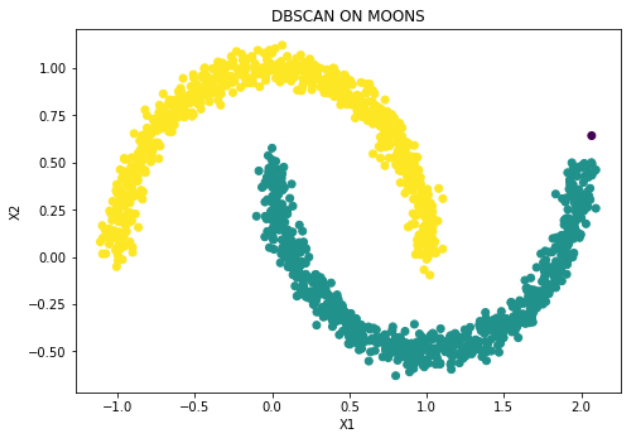
7. DBSCAN AND KMEANS ON MOONS

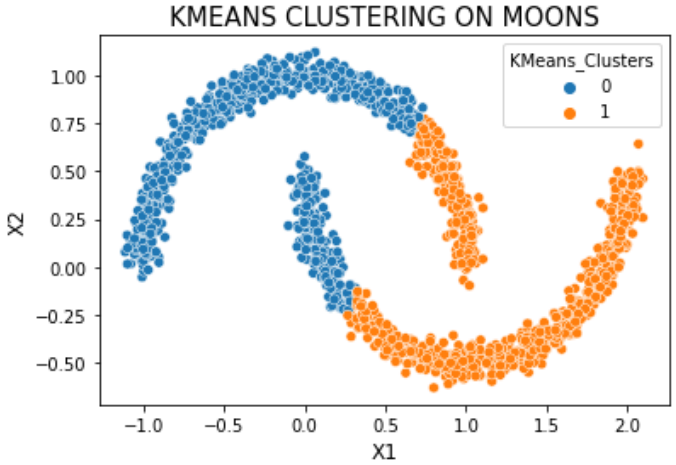
DBSCAN working better on moons compared to KMEANS

We are able to separate out the outliers using DBSCAN but not with KMEANS.

Silhouette score is much lesser in case of DBASCAN compared to KMEANS because of moon shape.

```
Silhouette Coefficient:0.19
```






```
Printing shilhoutte scre using KMeans
0.5856866231023987
```