Proximity Measures & Cluster Analysis

Rajesh K. Jat

2021kpad1001@iiitkota.ac.in

Statistics for Data Science, Winter (2021-22)

Dataset Used:

Shape having 8 different datasets named as S1, S2, S3, …… S8.

Each shape has 3 fields, the first two are data fields and the third one is the label. (available at: http://cs.joensuu.fi/sipu/datasets/)

Clustering algorithms applied:

1. K-means Clustering
2. Hierarchical Clustering
3. Dbscan Clustering
4. PAM Clustering
5. Fuzzy Clustering

Cluster evaluation methods used:

1. Plots of Clusters
2. Gap statistics to know the best no of clusters in K-means
3. Silhouette Plot

Final Observations:

1. Plots of different methods on the different datasets are attached below in this document.

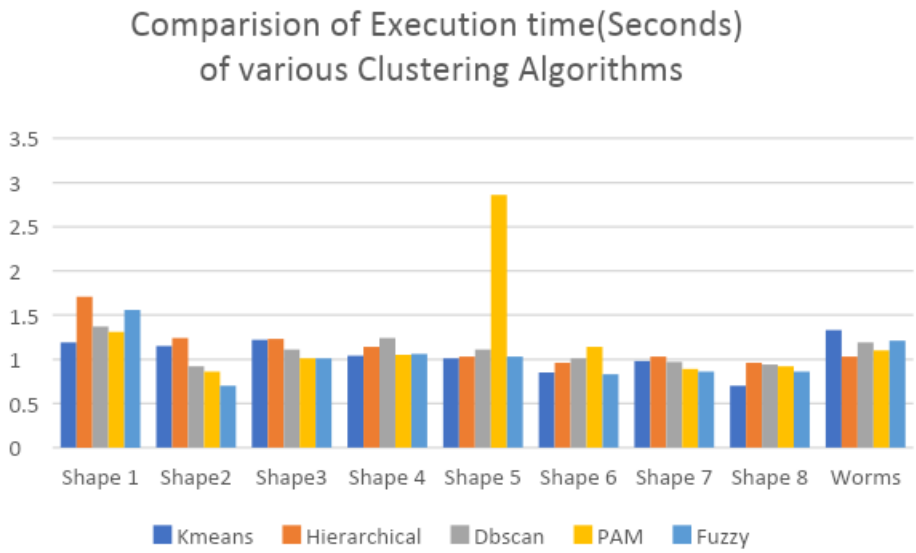
2. Different algorithms work differently because:
    - Kmeans works as per the selection of k value, if it is different it creates different clusters.
    - Hierarchical clustering has some initial clusters, finally they merge into higher clusters.
    - Dbscan is density based, so when the minimum points value is increased, no. of the clusters created is decreased.
    - PAM uses medoids, which are more centrally located values, so gives better results.
    - Fuzzy clustering method works on the basis of the degree of membership so many times a data point may be part of many clusters.
3. Hyperparameters like k in k means and min_pts in Dbscan are changed and plots are given.
4. Execution time of all the algorithms was calculated and then presented using a graph and it is found that on average K-means and Fuzzy clustering methods are giving minimum execution time for all the datasets.
5. Overall Dbscan gave the best performance almost in all the cases because density is a major point when shapes are drawn

Execution time calculation was done through tic() and toc() functions. Here are the results:

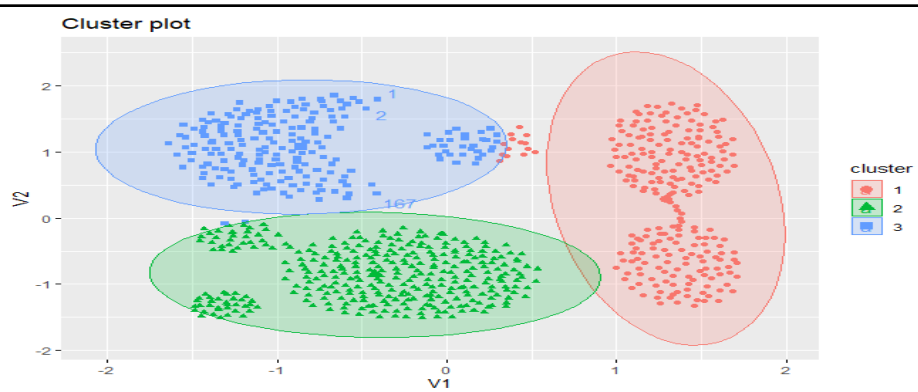
Dataset	**Clustering Model**	**Execution Time(Seconds)**
Shape1	Kmeans	1.19
	Hierarchical	1.71
	Density Based	1.37
	PAM	1.31
	Fuzzy	1.56
Shape 2	Kmeans	1.15

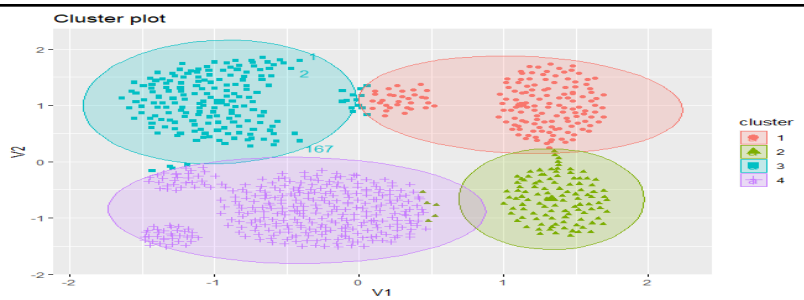
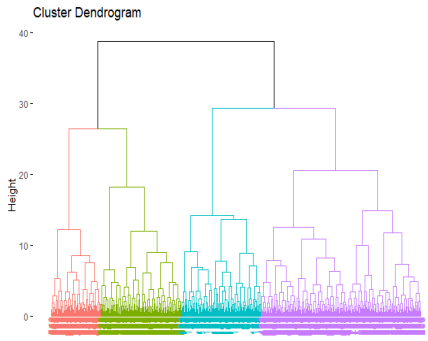
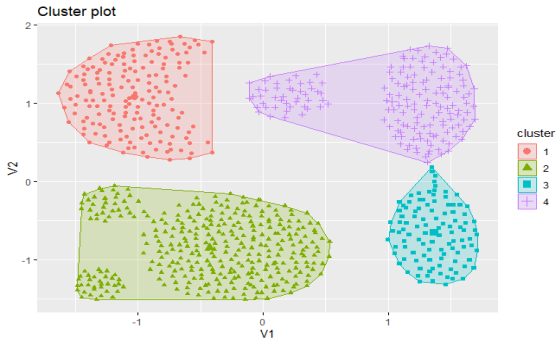
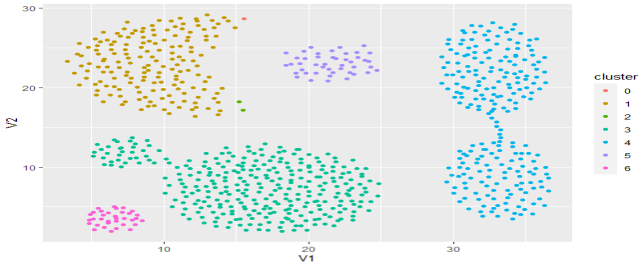
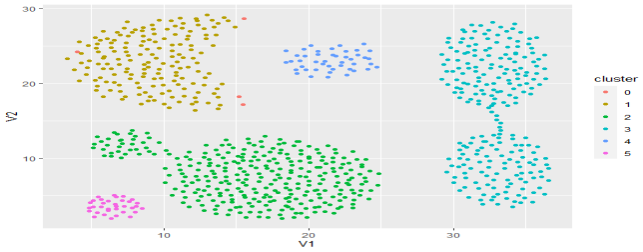
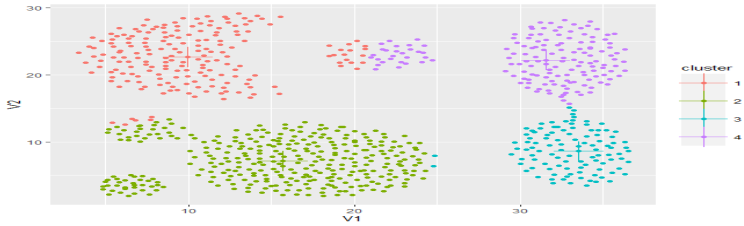
	Hierarchical	1.24
	Density Based	0.92
	PAM	0.86
	Fuzzy	0.70
Shape 3	Kmeans	1.22
	Hierarchical	1.23
	Density Based	1.11
	PAM	1.01
	Fuzzy	1.01
Shape 4	Kmeans	1.04
	Hierarchical	1.14
	Density Based	1.24
	PAM	1.05
	Fuzzy	1.06
Shape 5	Kmeans	1.01
	Hierarchical	1.03
	Density Based	1.11
	PAM	2.86
	Fuzzy	1.03
Shape 6	Kmeans	0.85
	Hierarchical	0.96
	Density Based	1.01
	PAM	1.14
	Fuzzy	0.83
Shape 7	Kmeans	0.98
	Hierarchical	1.03
	Density Based	0.97
	PAM	0.89
	Fuzzy	0.86
Shape 8	Kmeans	0.70
	Hierarchical	0.96
	Density Based	0.94
	PAM	0.92
	Fuzzy	0.86

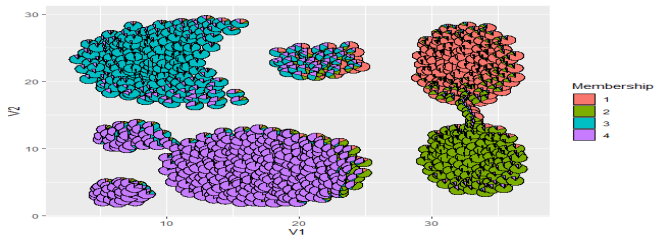
Execution Time :



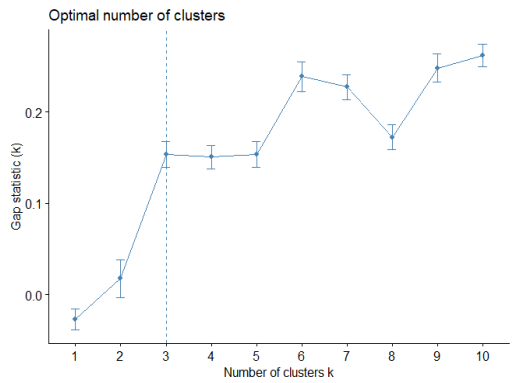
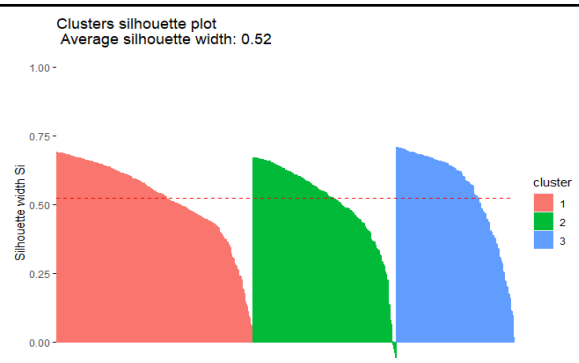
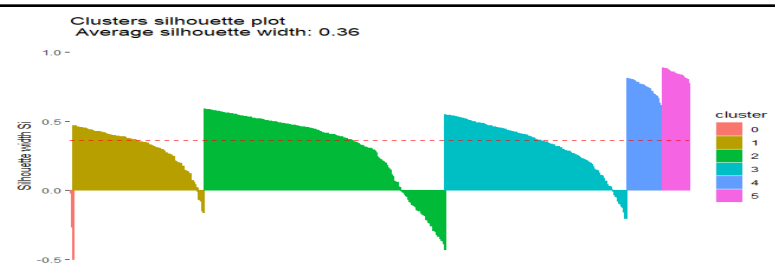
Shape 1 :

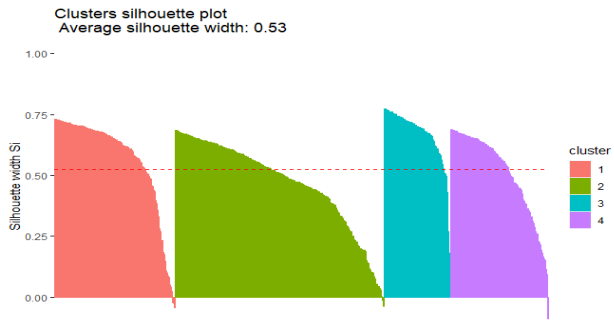
Kmeans(K=3)	

Kmeans(K=4)	
Hierarchical	 
Dbscan Minimum point=2	
Dbscan Minimum points=4	
PAM	

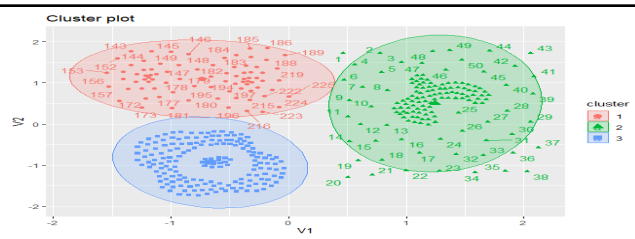
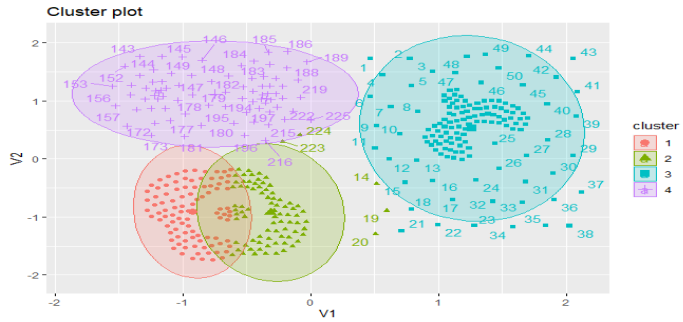
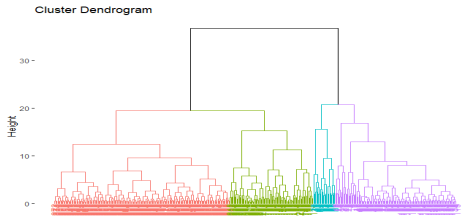
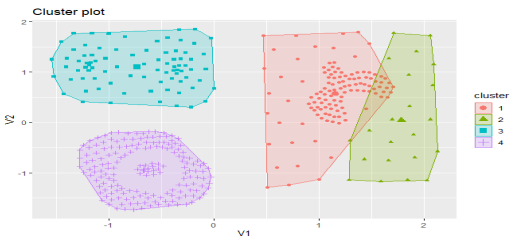
Fuzzy	

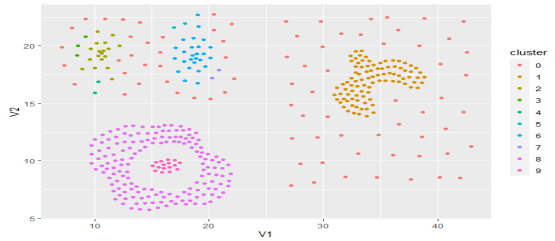
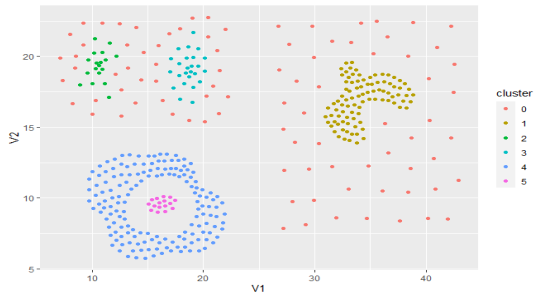
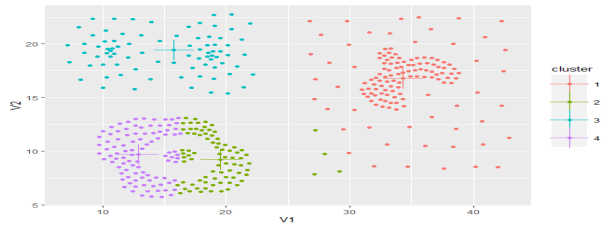
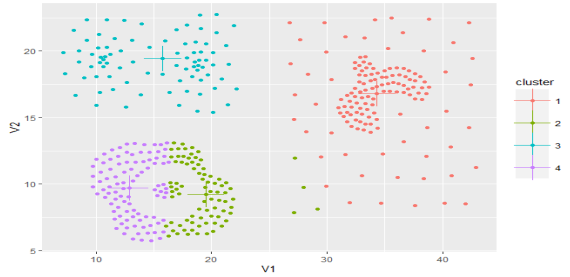
Cluster Evaluation for Shape 1:

Kmeans (optimal no of Cluster)	
Silhouette plot for k-means	
Silhouette plot for Dbscan	

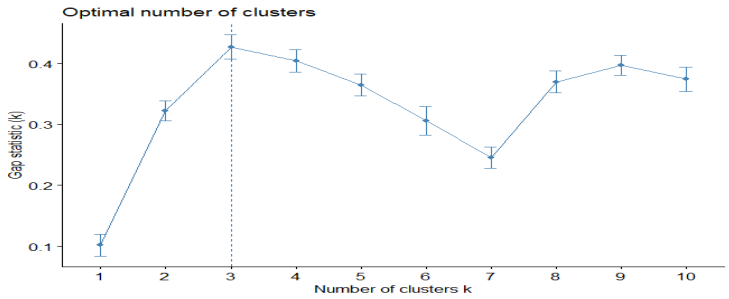
Silhouette plot for PAM	

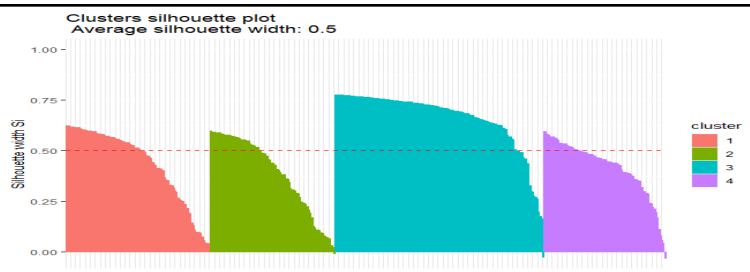
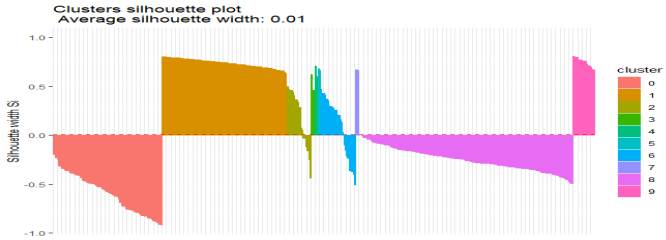
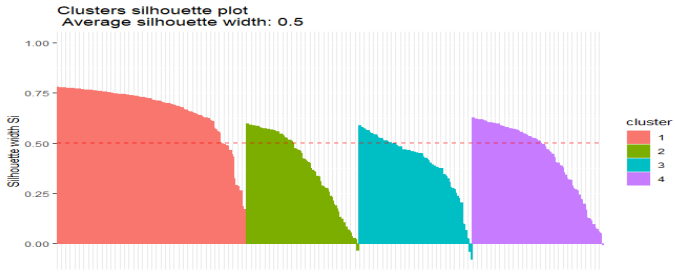
Shape2

Kmeans(K=3)	
Kmeans(K=4)	
Hierarchical	 

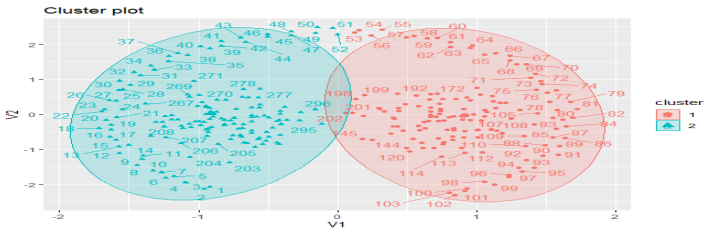
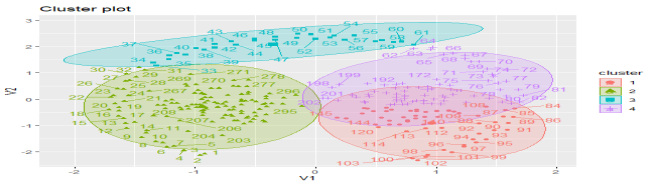
Dbscan Minimum point=2	
Dbscan Minimum points=4	
PAM	
Fuzzy	

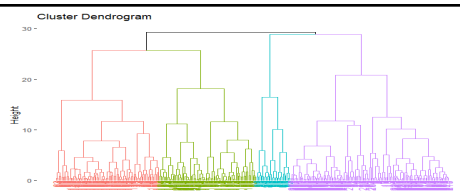
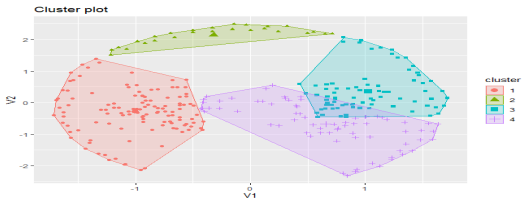
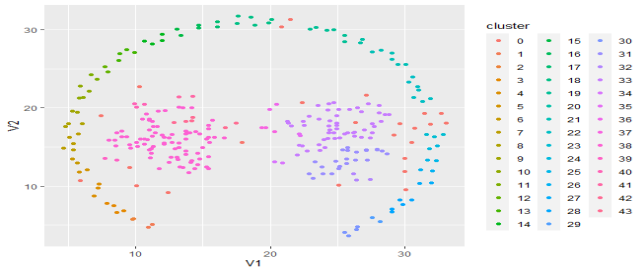
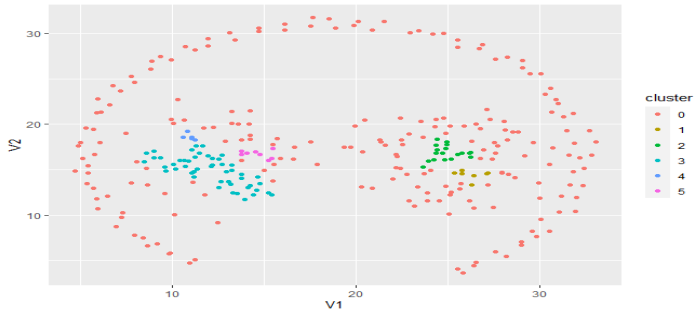
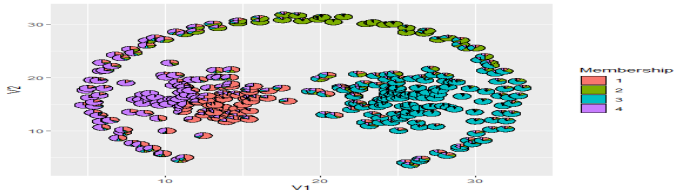
Cluster Evaluation for shape 2:

Kmeans (optimal no of Cluster)	

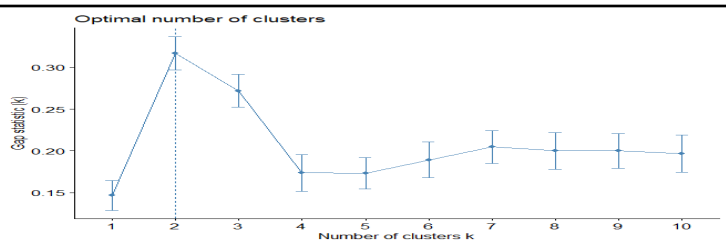
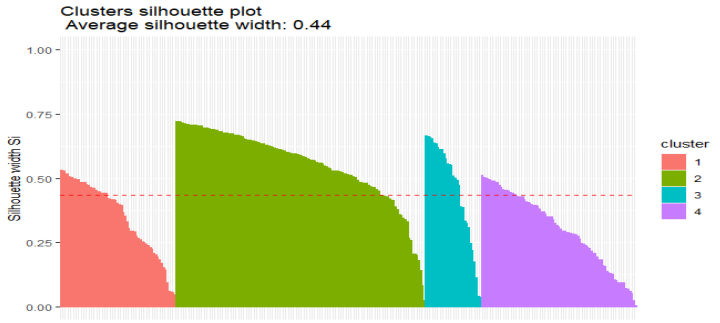
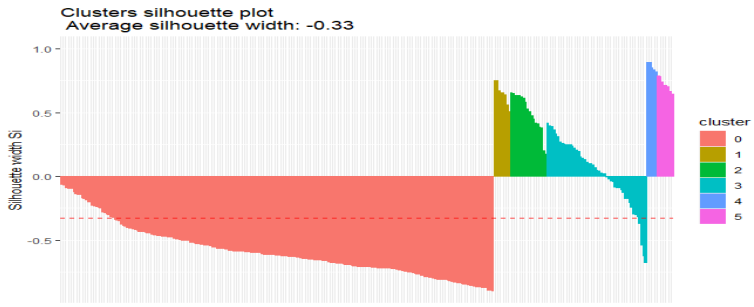
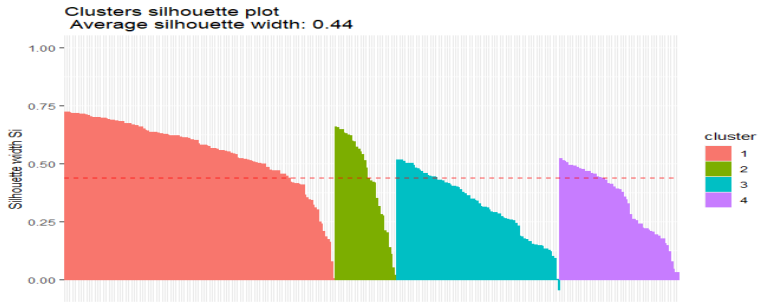
Silhouette plot for k-means	
Silhouette plot for Dbscan	
Silhouette plot for PAM	

Shape 3:

Kmeans(K=2)	
Kmeans(K=4)	

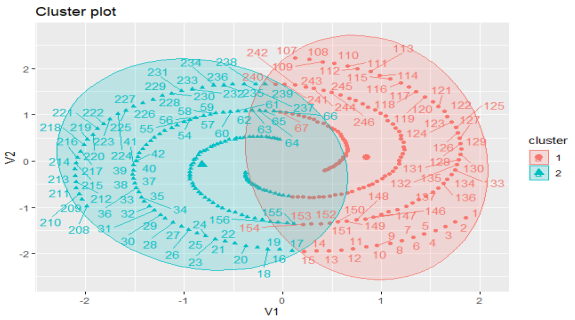
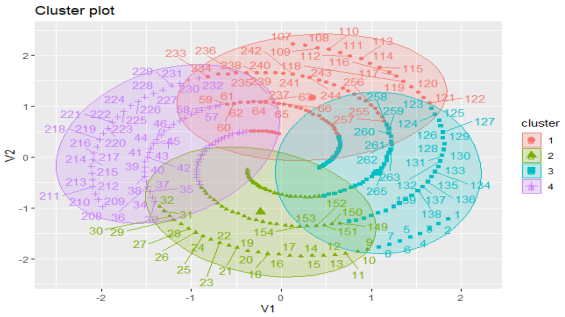
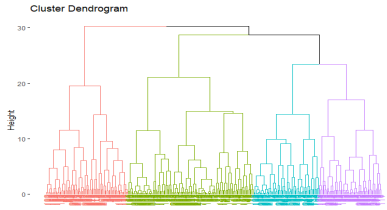
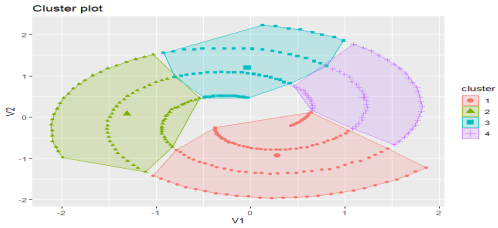
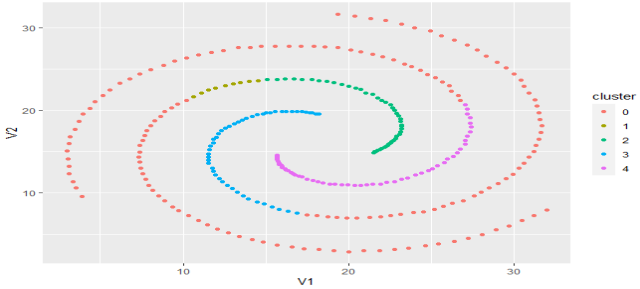
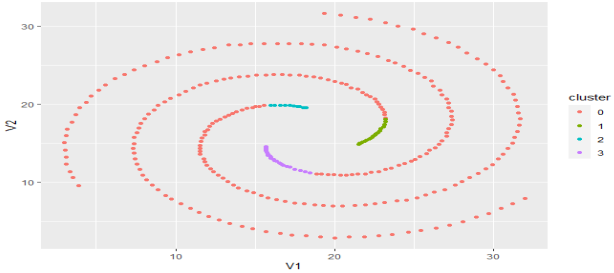
Hierarchical	 
Dbscan Minimum point=2	
Dbscan Minimum points=8	
PAM	
Fuzzy	

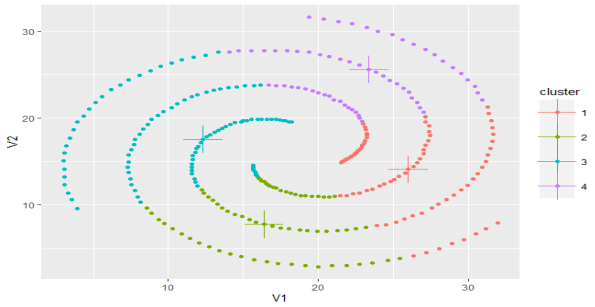
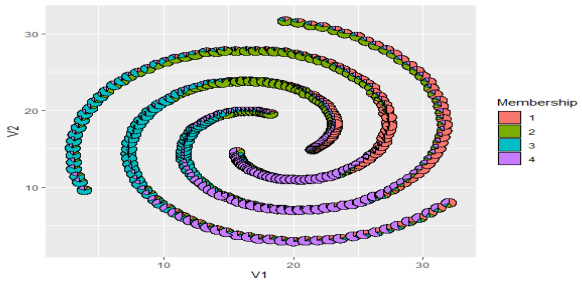
Cluster Evaluation Shape 3:

Kmeans (optimal no of Cluster)	
Silhouette plot for k-means	
Silhouette plot for Dbscan	
Silhouette plot for PAM	

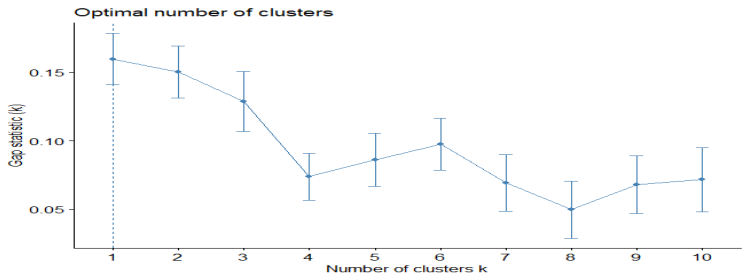
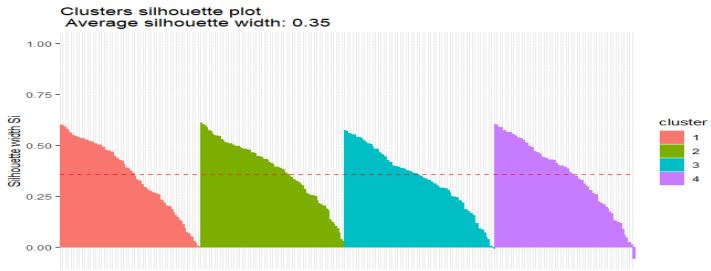
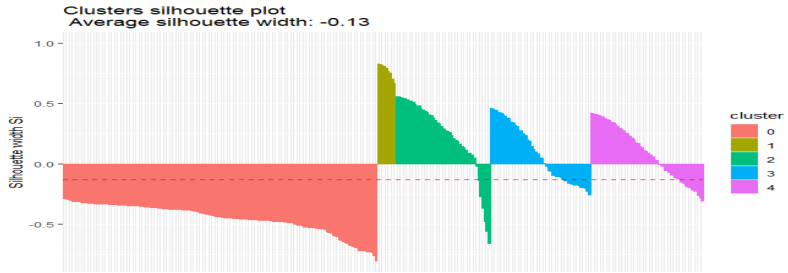
Shape 4 :

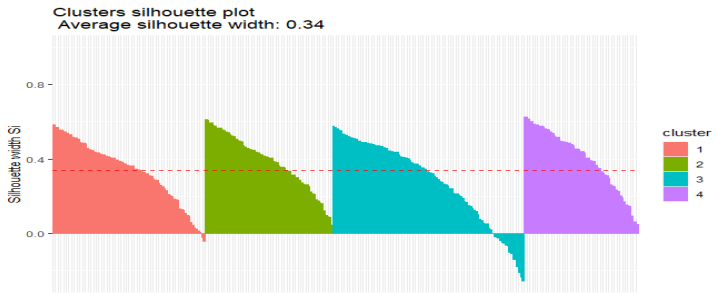
Kmeans(K=2)	

	
Kmeans(K=4)	
Hierarchical	 
Dbscan Minimum point=4	
Dbscan Minimum points=8	

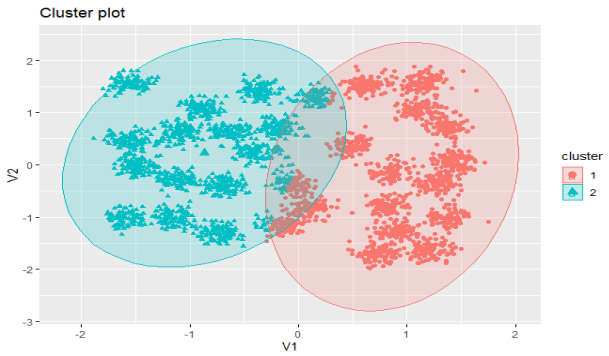
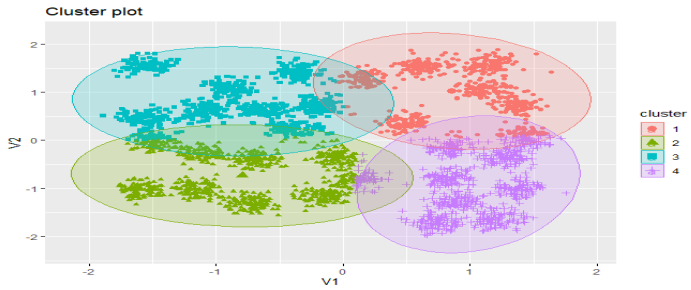
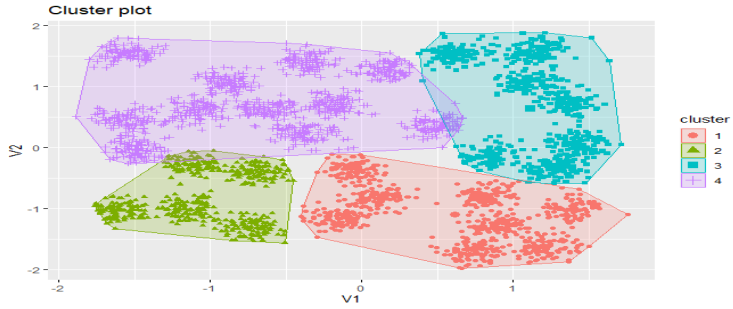
PAM	
Fuzzy	

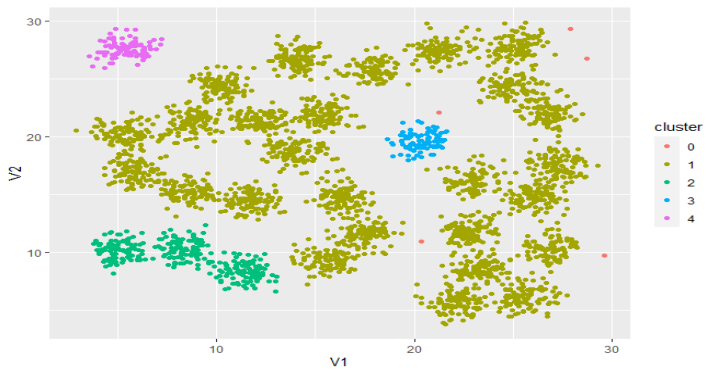
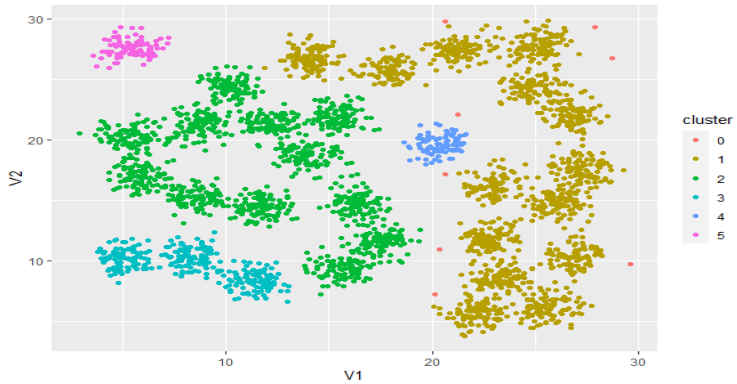
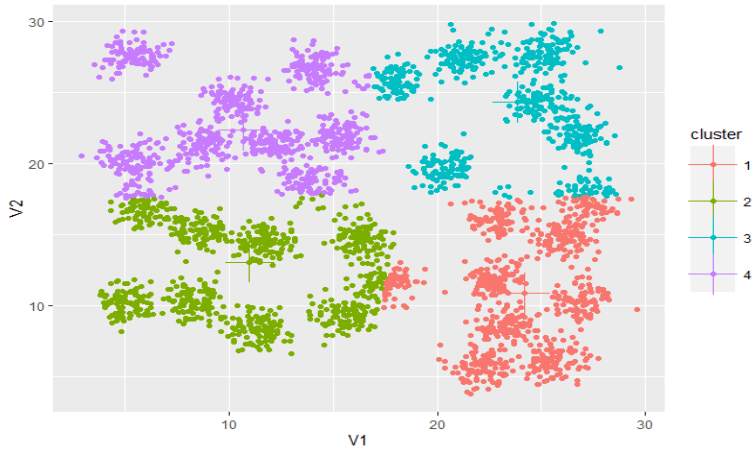
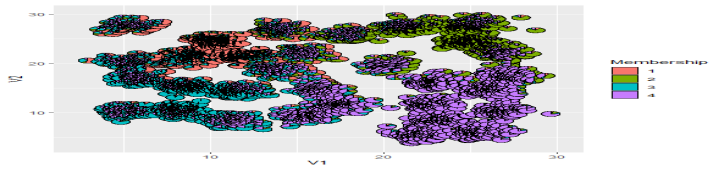
Cluster Evaluation for Shape 4.

Kmeans (optimal no of Cluster)	
Silhouette plot for k-means	
Silhouette plot for Dbscan	

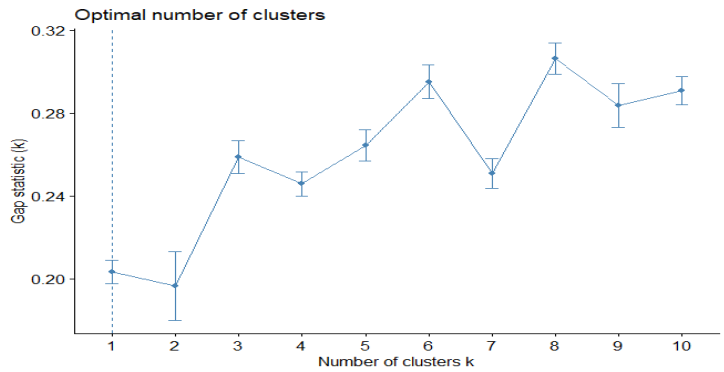
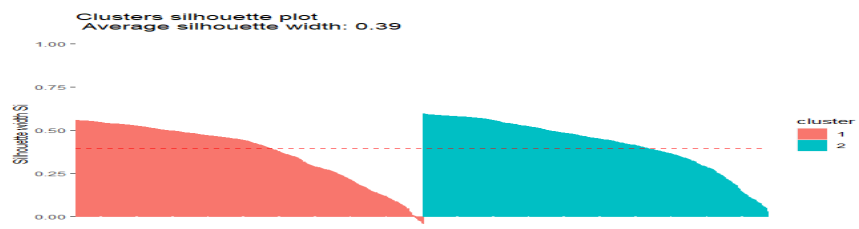
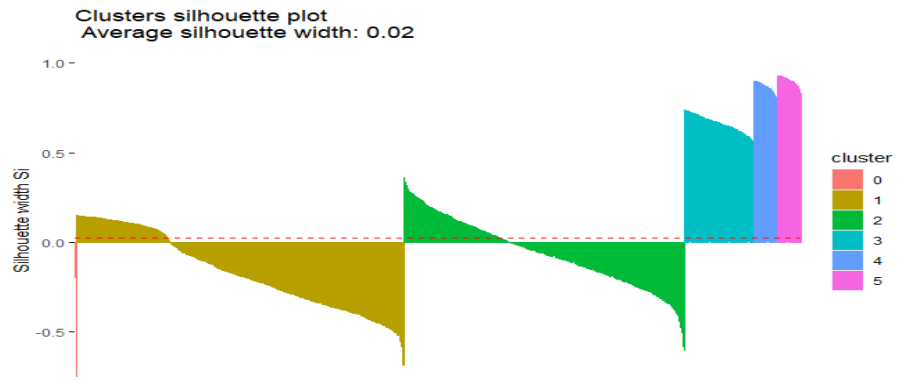
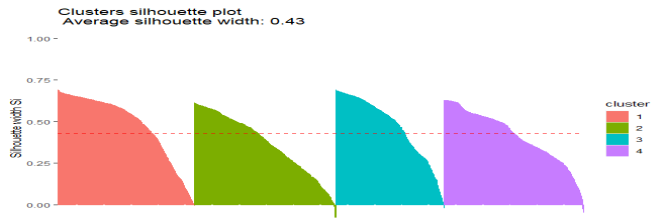
Silhouette plot for PAM	

Shape 5 :

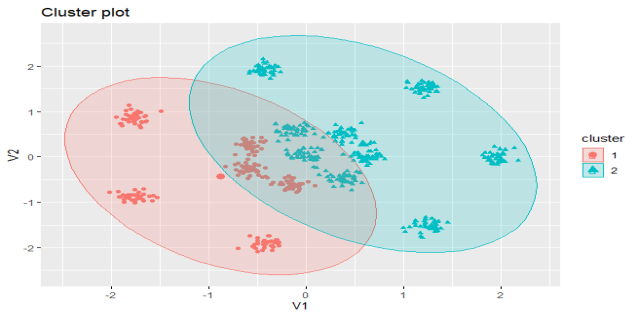
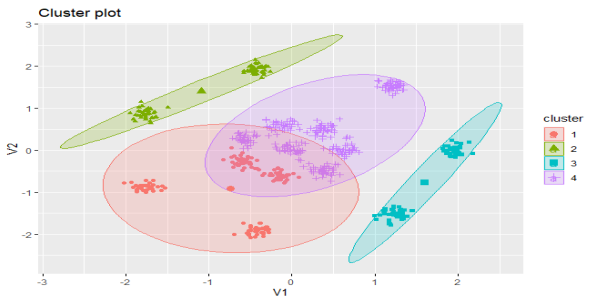
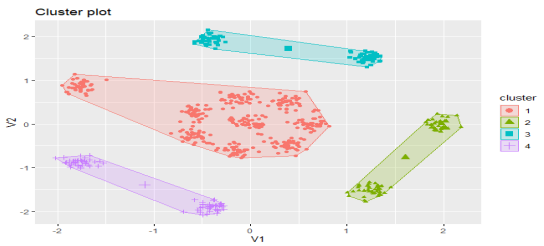
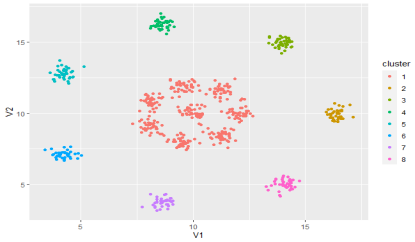
Kmeans(K=2)	
Kmeans(K=4)	
Hierarchic al	

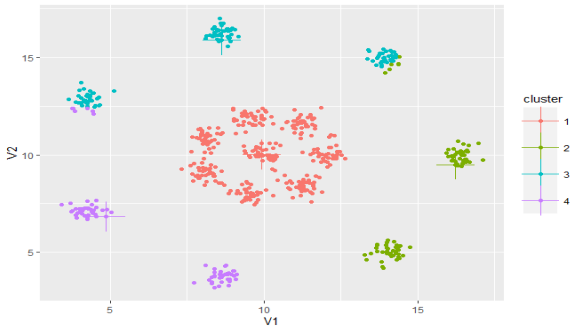
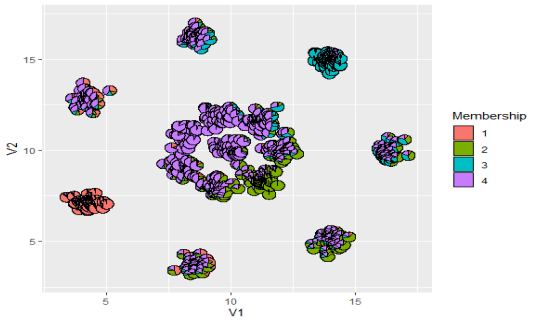
Dbscan Minimum point=4	
Dbscan Minimum points=8	
PAM	
Fuzzy	

Cluster Evaluation Shape5 :

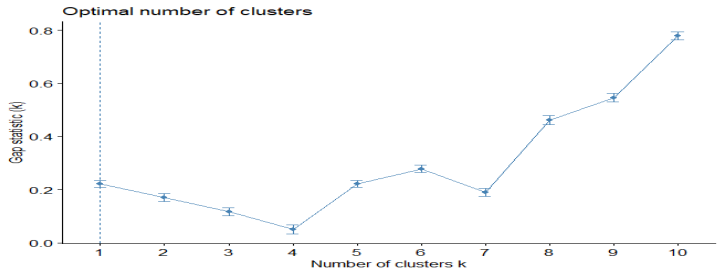
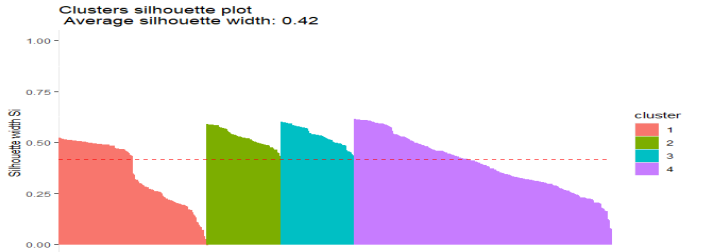
Kmeans (optimal no of Cluster)	
Silhouette plot for k-means	
Silhouette plot for Dbscan	
Silhouette plot for PAM	

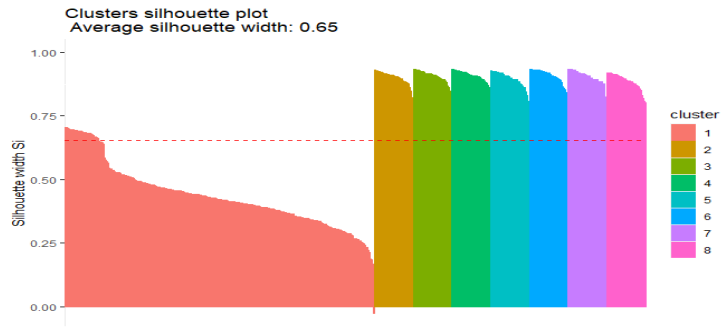
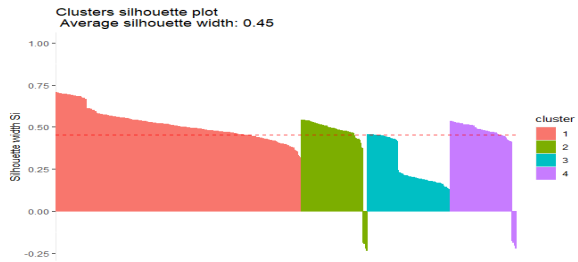
Shape 6:

Kmeans(K=2)	
Kmeans(K=4)	
Hierarchical	
Dbscan Minimum point=4	
Dbscan Minimum points=8	

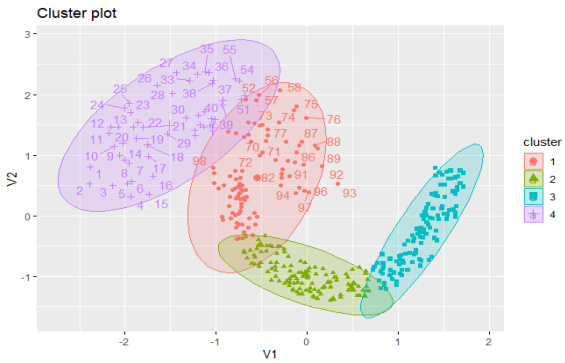
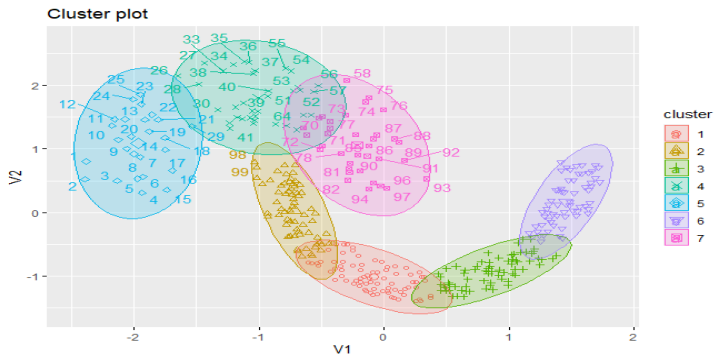
PAM	
Fuzzy	

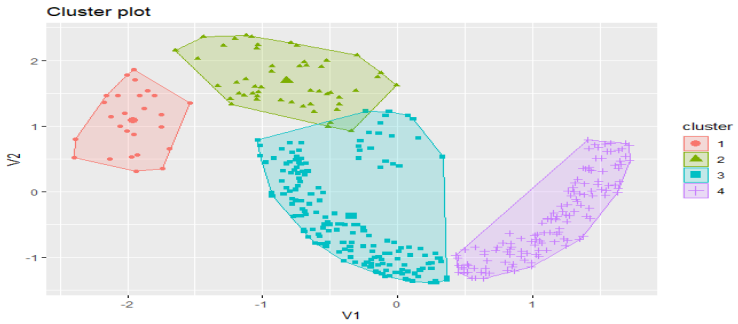
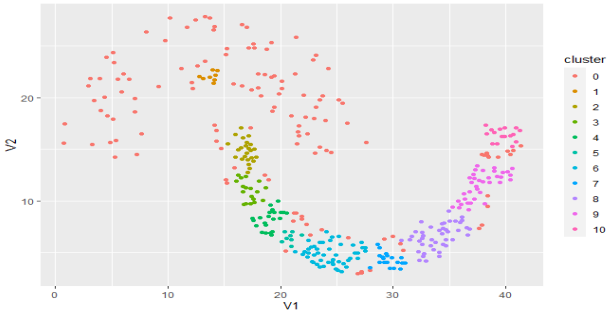
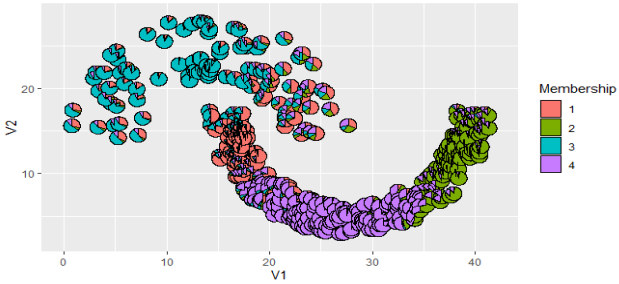
Cluster Evaluation Shape 6 :

Kmeans (optimal no of Cluster)	
Silhouette plot for k-means	

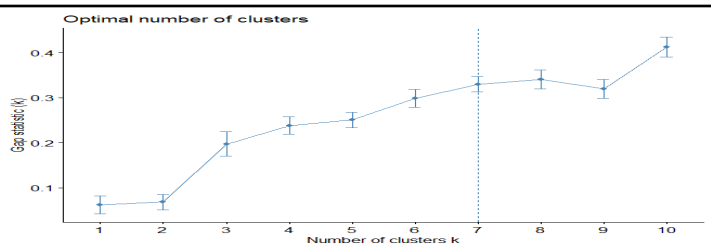
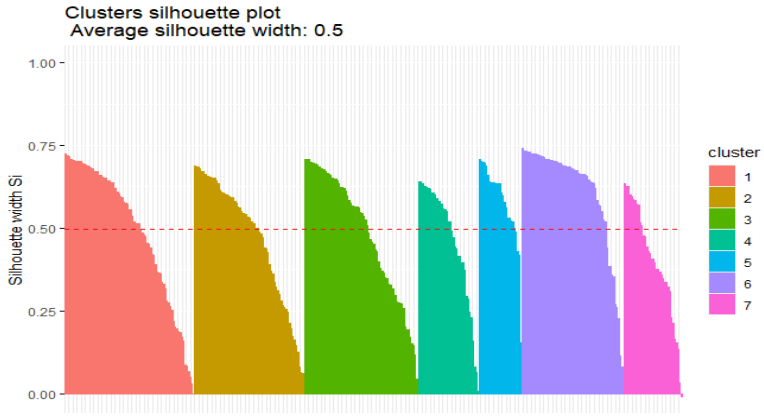
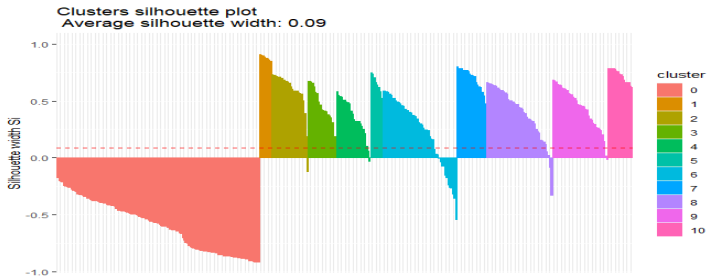
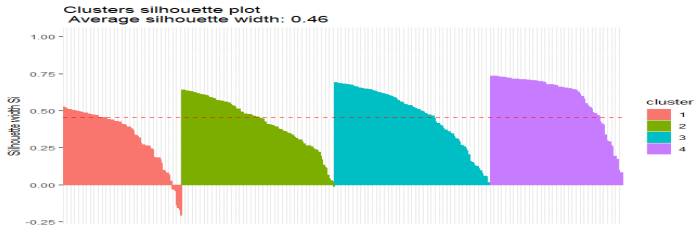
Silhouette plot for Dbscan	
Silhouette plot for PAM	

Shape 7

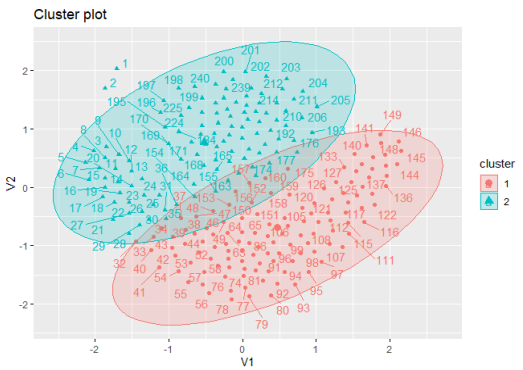
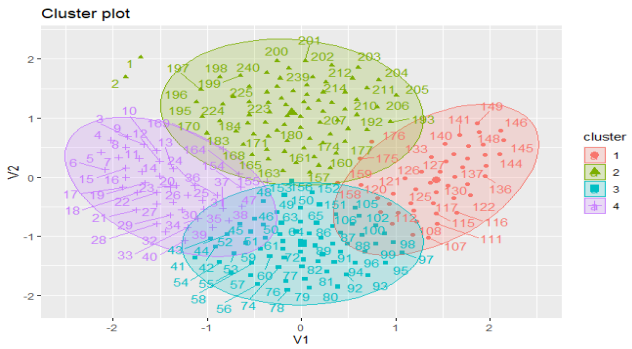
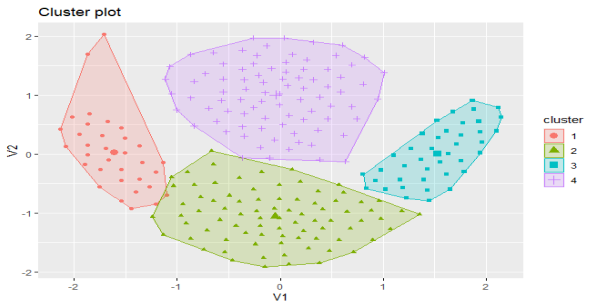
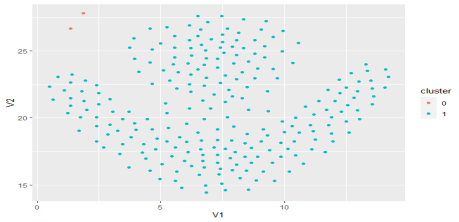
Kmeans(K=4)	
Kmeans(K=7)	
Hierarchic al	

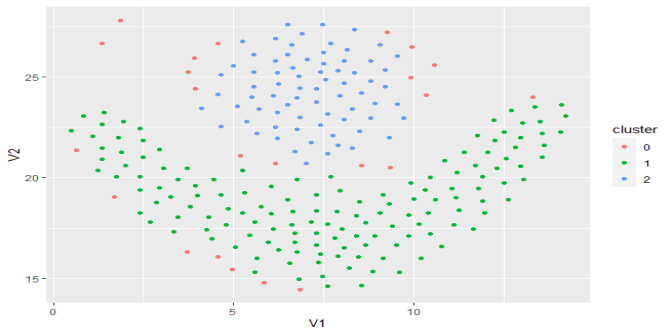
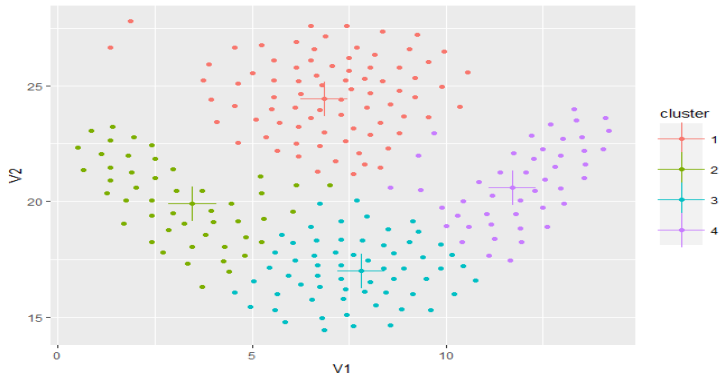
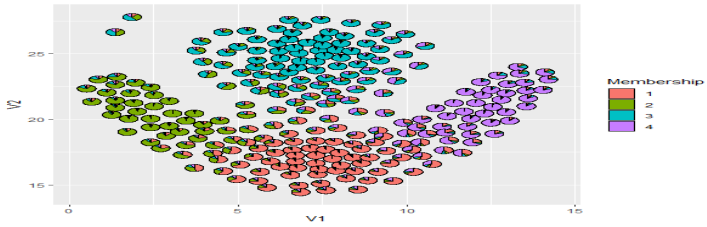
	
Dbscan Minimum point=4	
Dbscan Minimum points=8	
PAM	
Fuzzy	

Cluster Evaluation Shape 7:

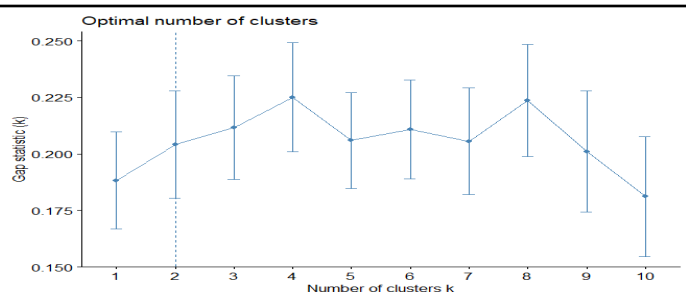
Kmeans (optimal no of Cluster)	
Silhouette plot for k-means	
Silhouette plot for Dbscan	
Silhouette plot for PAM	

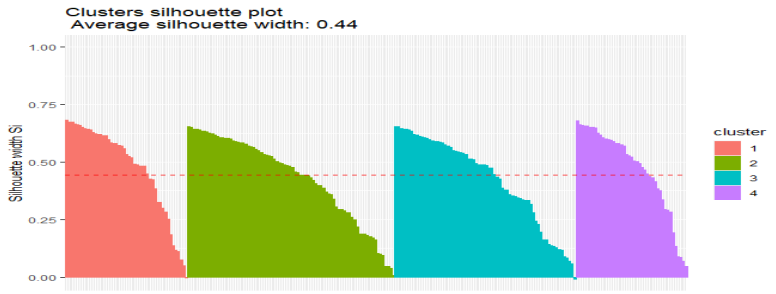
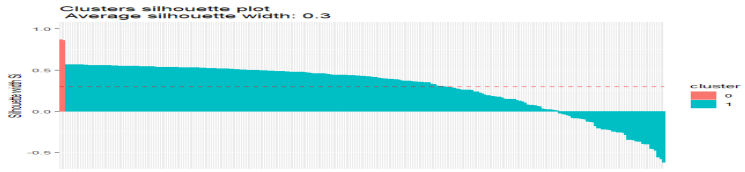
Shape 8:

Kmeans(K=2)	
Kmeans(K=4)	
Hierarchical	
Dbscan Minimum point=4	

Dbscan Minimum points=8	
PAM	
Fuzzy	

Cluster Evaluation Shape 8 :

Kmeans (optimal no of Cluster)	

Silhouette plot for k-means	
Silhouette plot for Dbscan	
Silhouette plot for PAM