

22 de marzo 2024

APÉNDICE

DAVID GARCÍA, LLUÍS ITARTE, MARCOS ROCA, DIEGO RODRÍGUEZ

Índice

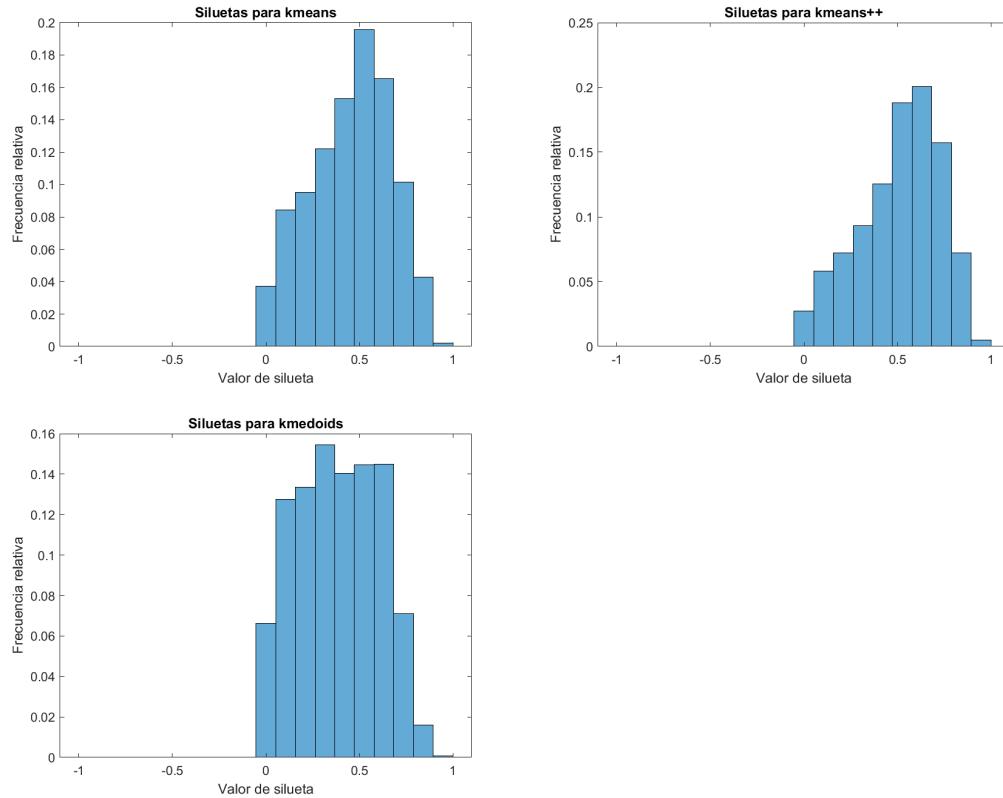
1 Experimentación

1.1 Clustering y medidas internas	2
1.1.1 Dataset 1	3
1.1.2 Dataset 2	8
1.1.3 Dataset 3	13
1.1.5 Dataset 6	18
1.2 Evaluación externa	28
1.2.1 Dataset Propio.	29
1.2.1 Dataset 1	29
1.2.2 Dataset 2	30
1.2.3 Dataset 3	30
1.2.5 Dataset 4	31
1.2.6 Dataset 5	32
1.2.7 Dataset 6	33
1.3 Optimización de k	34
1.3.1 Dataset 1	35
1.3.2 Dataset 2	35
1.3.3 Dataset 3	35
1.3.5 Dataset 4	35
1.3.6 Dataset 5	35
1.3.7 Dataset 6	35

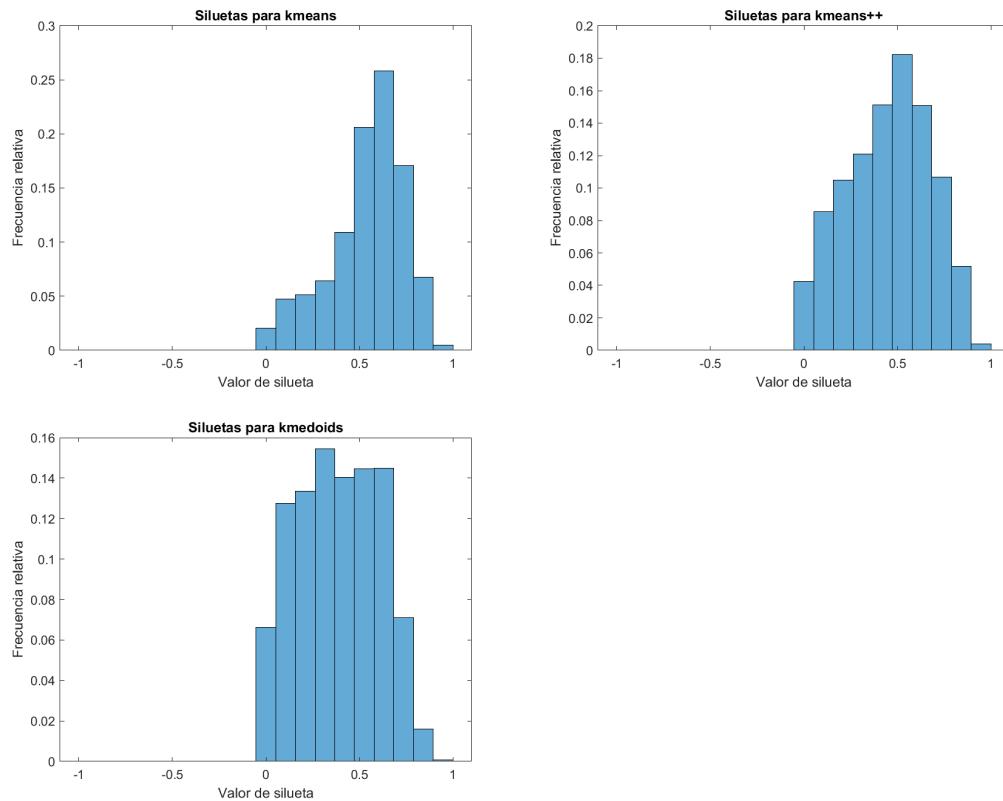
CLUSTERING y MEDIDAS INTERNAS

DataSet 1

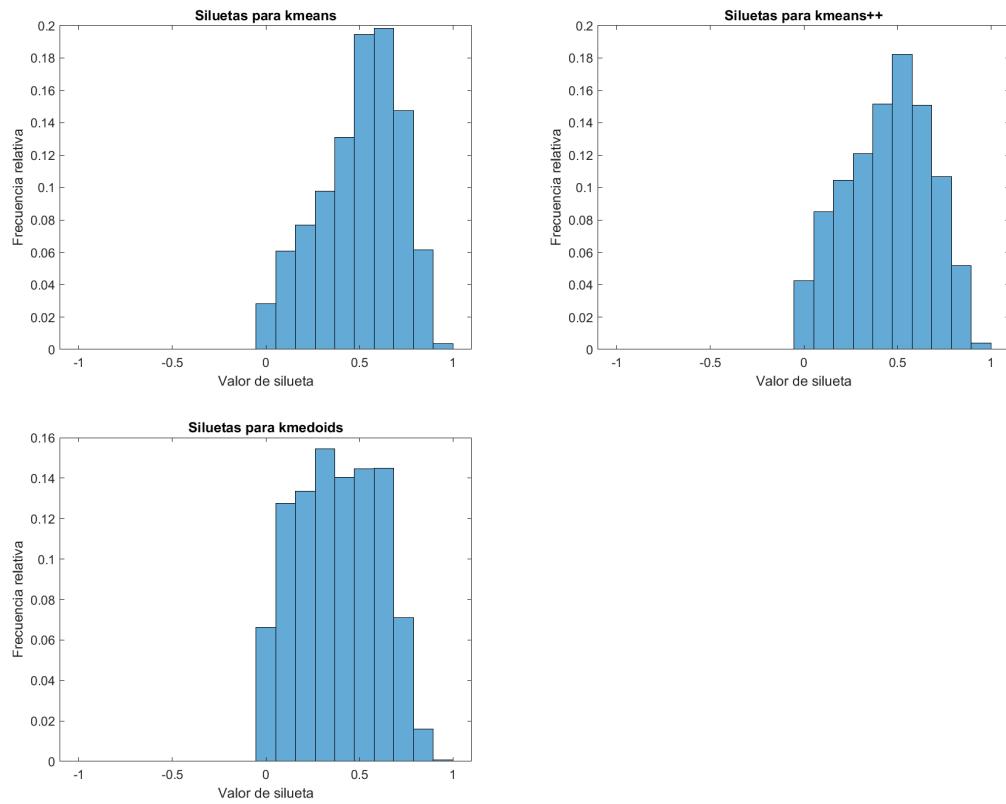
·Experimento 1



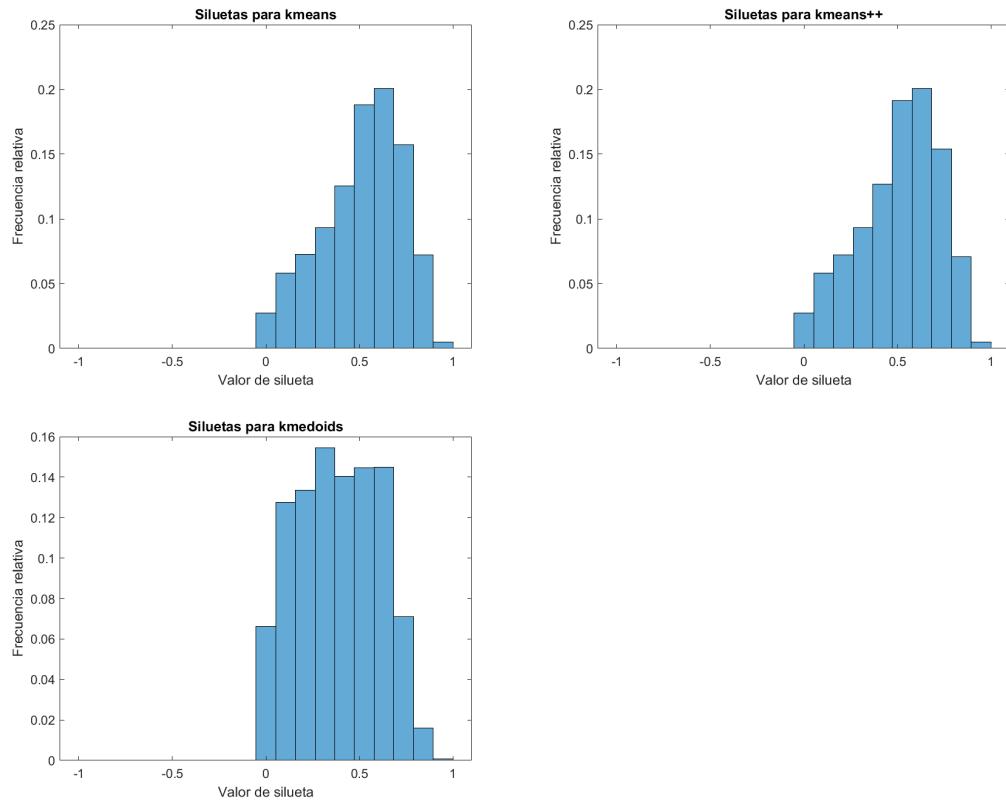
·Experimento 2



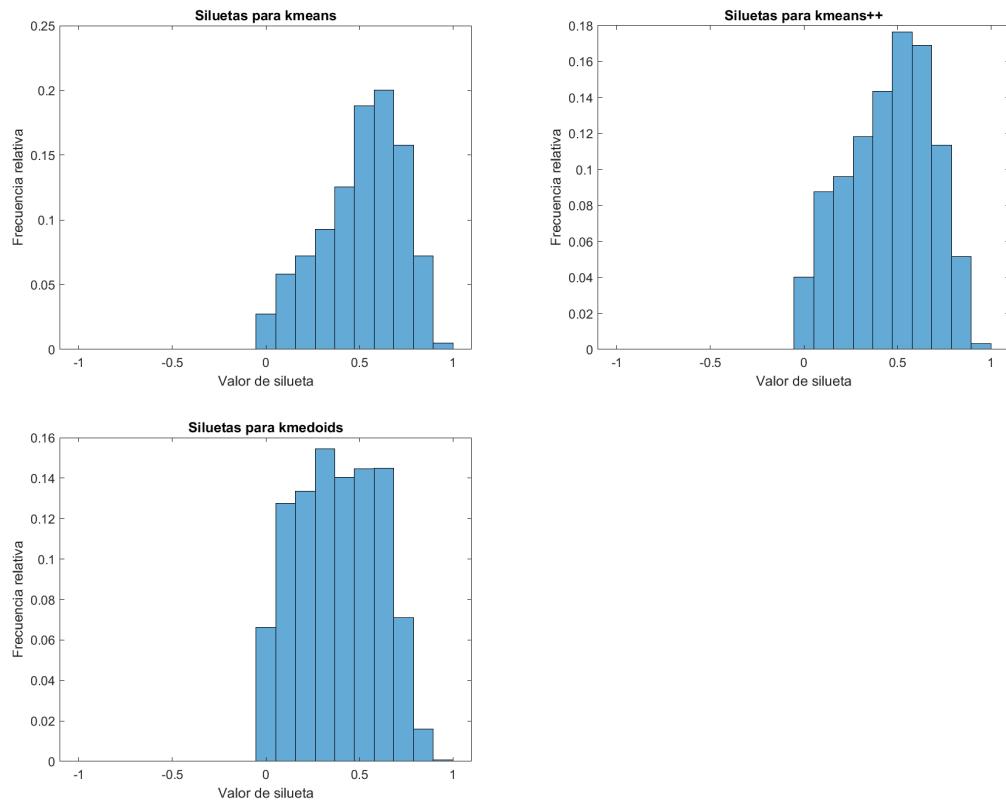
•Experimento 3



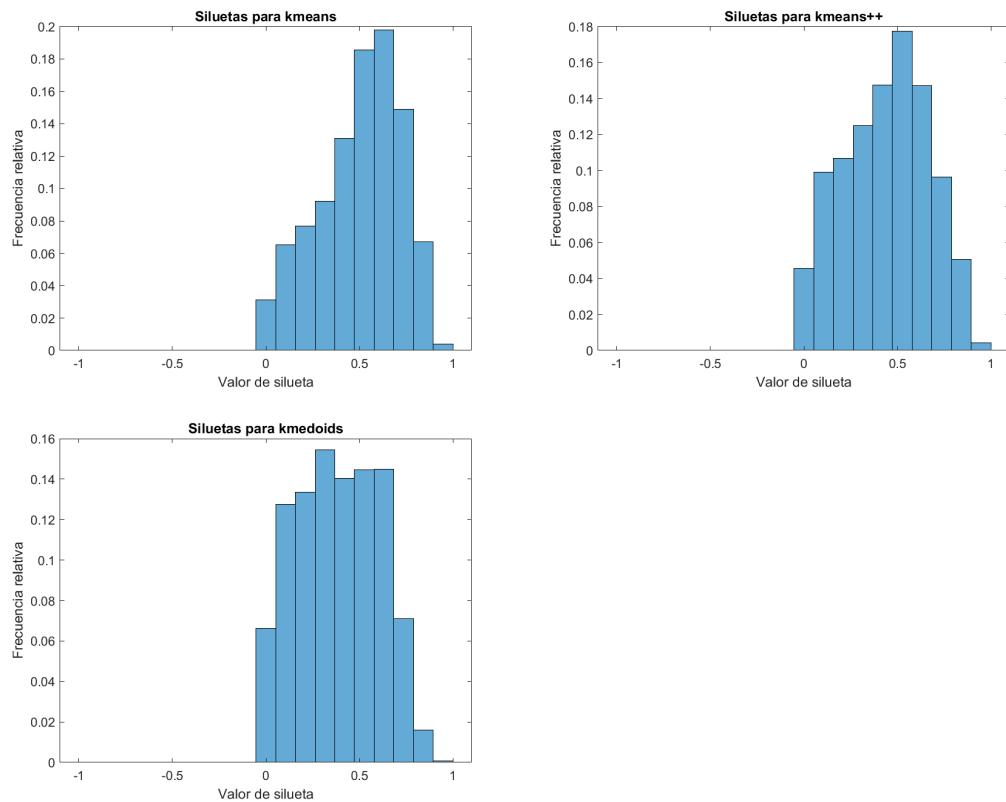
•Experimento 4



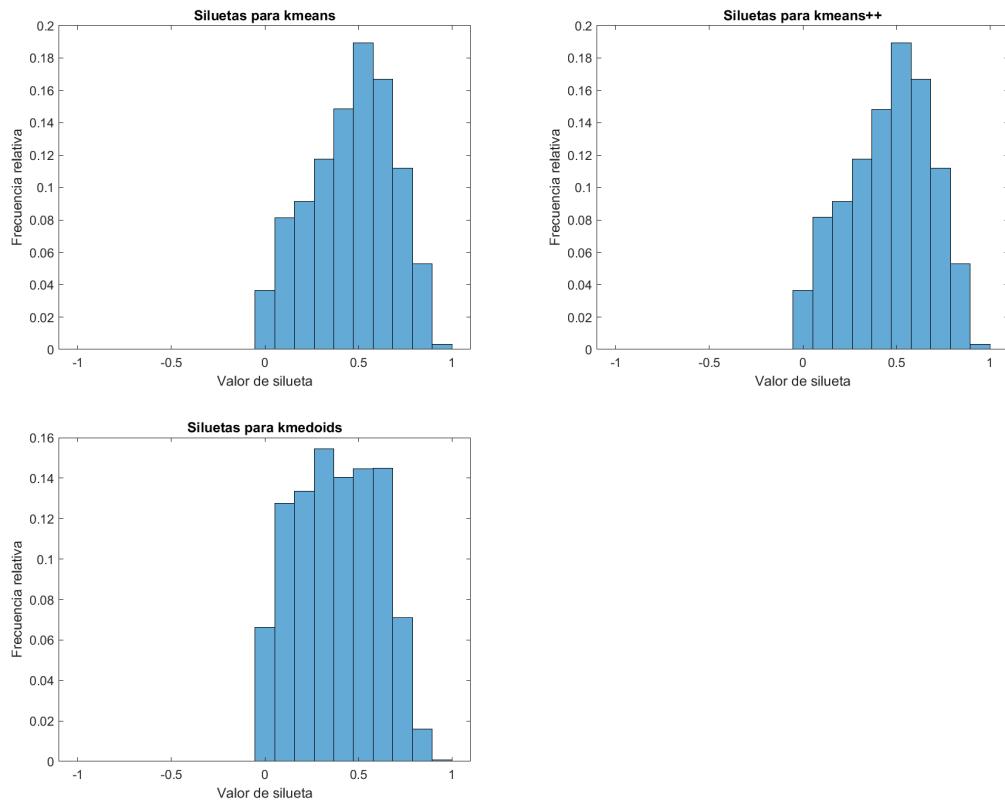
•Experimento 5



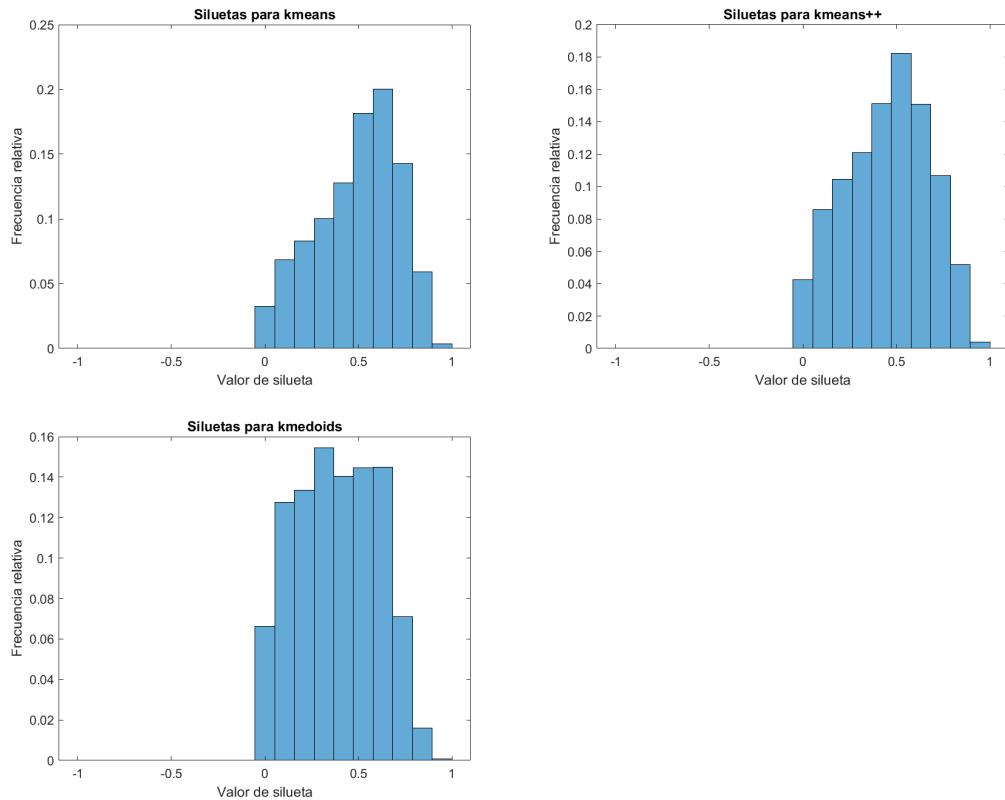
•Experimento 6



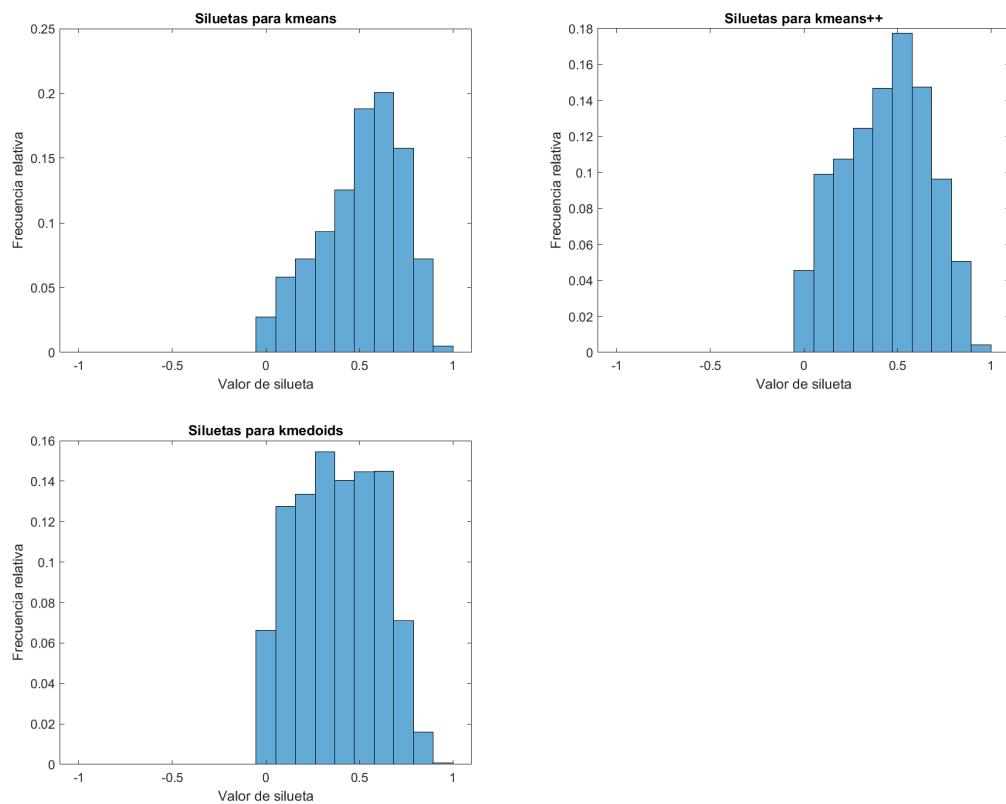
•Experimento 7



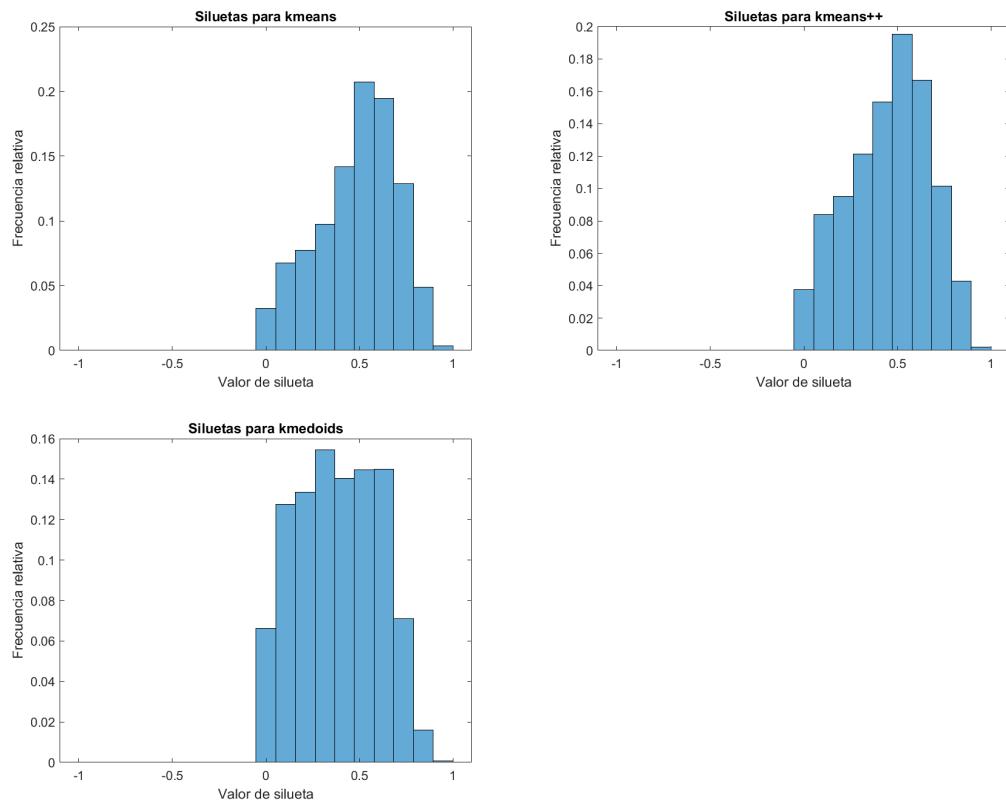
•Experimento 8



•Experimento 9

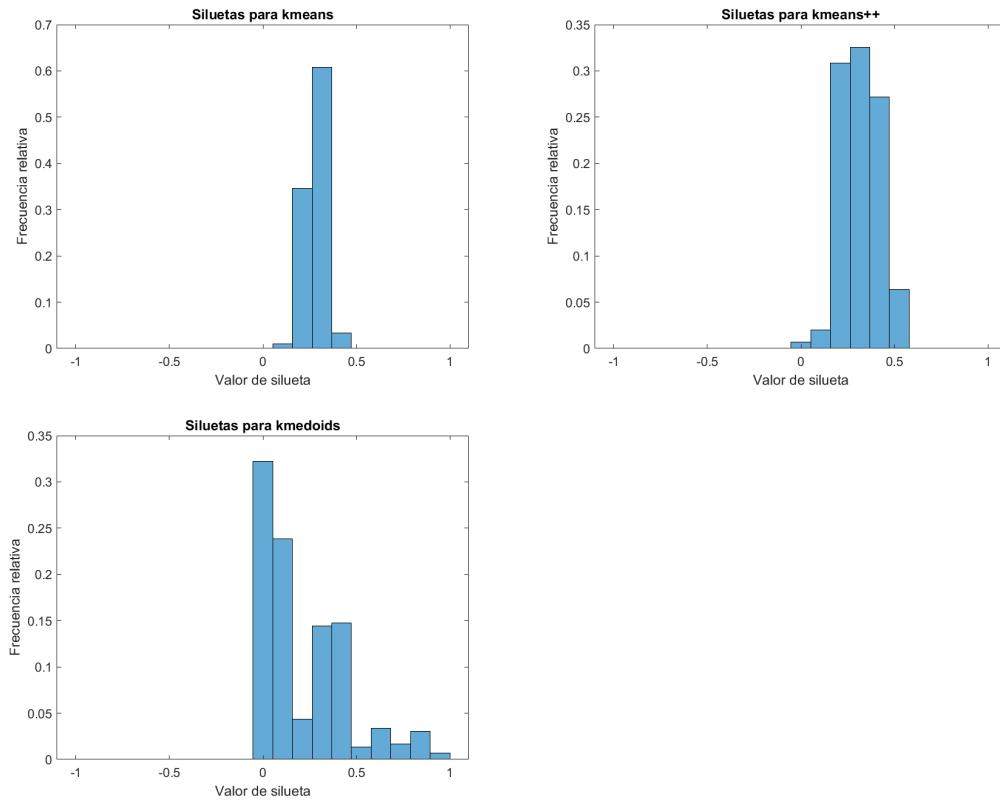


•Experimento 10

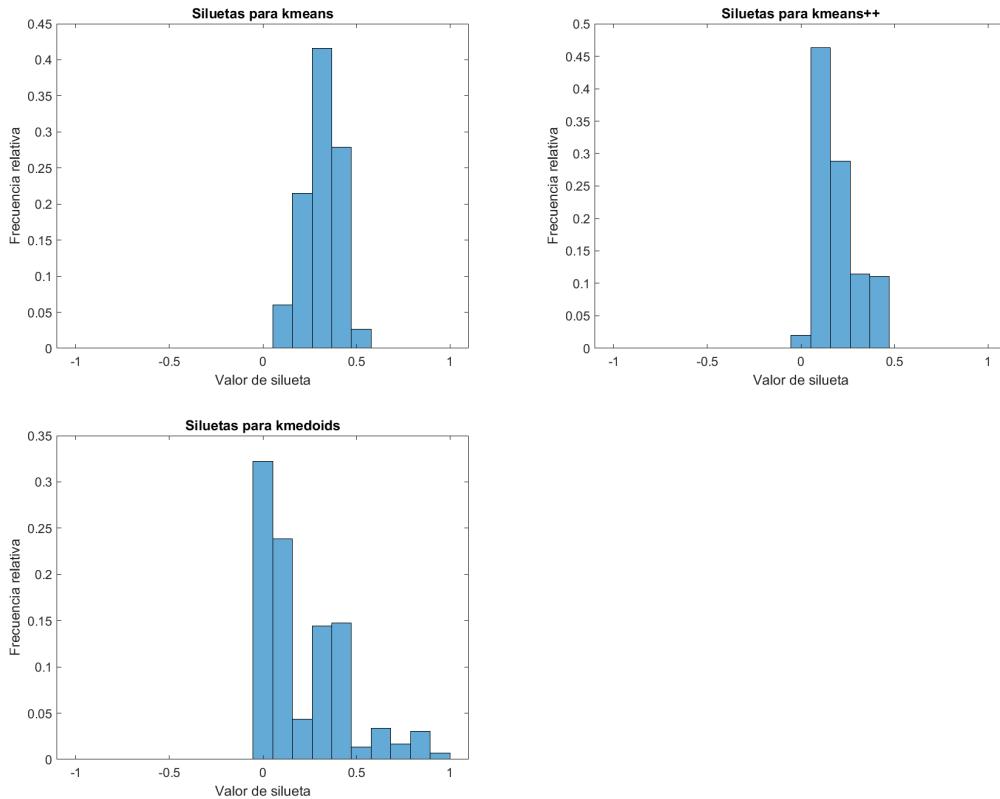


DataSet 2

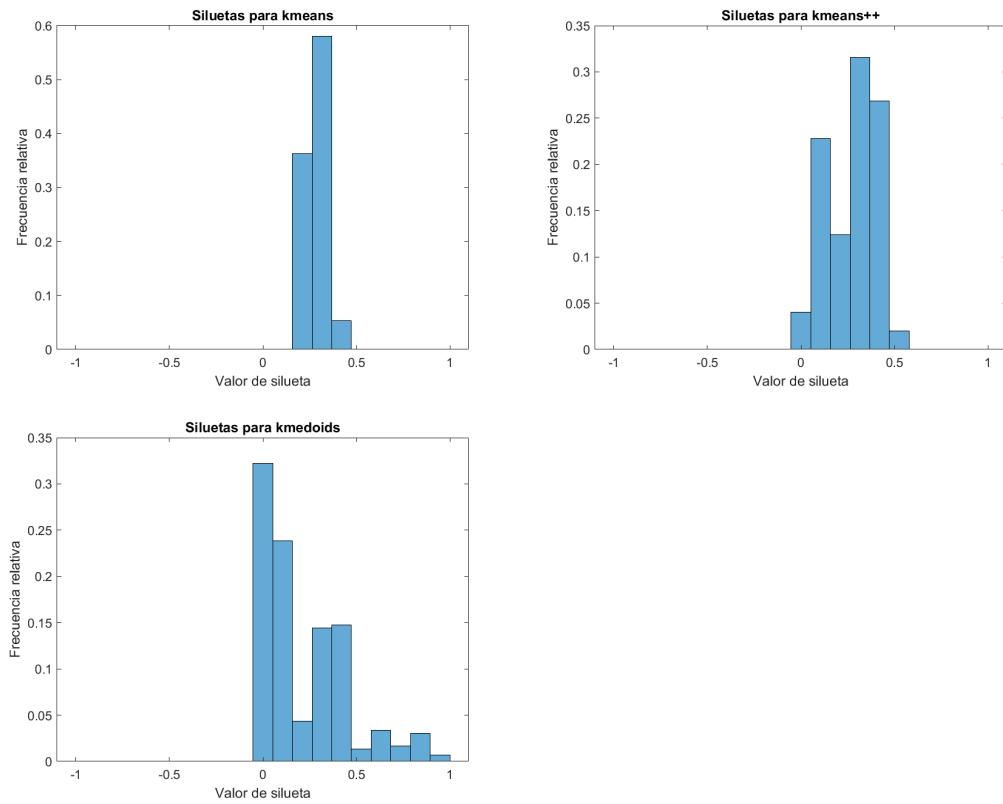
·Experimento 1



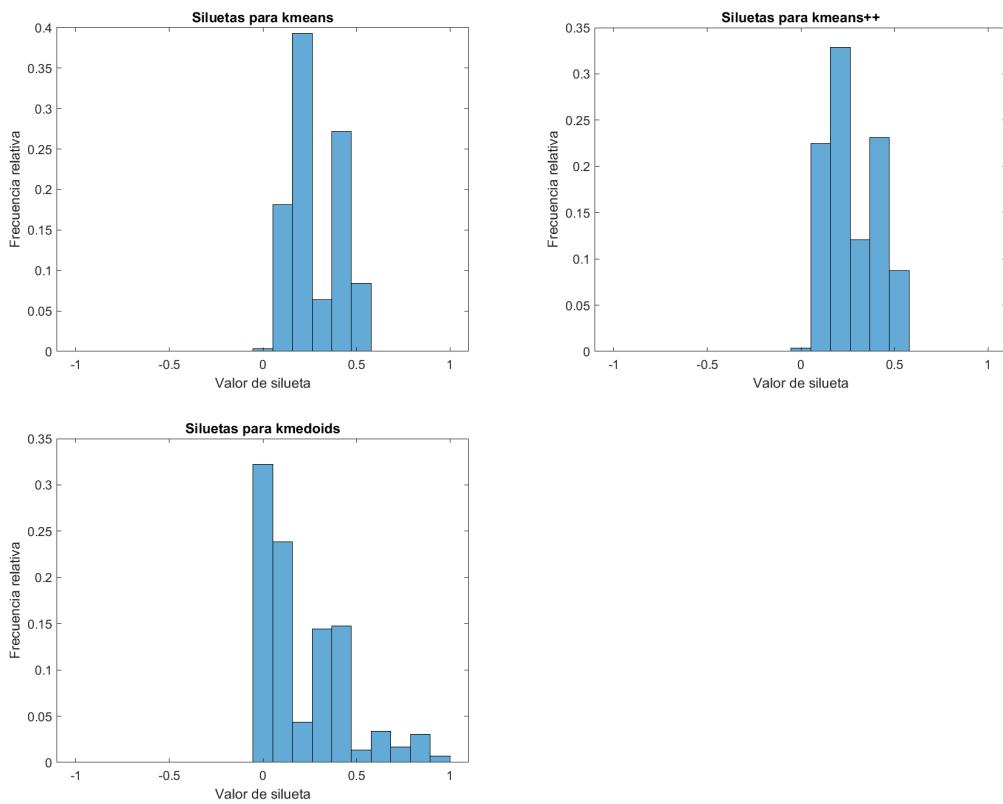
·Experimento 2



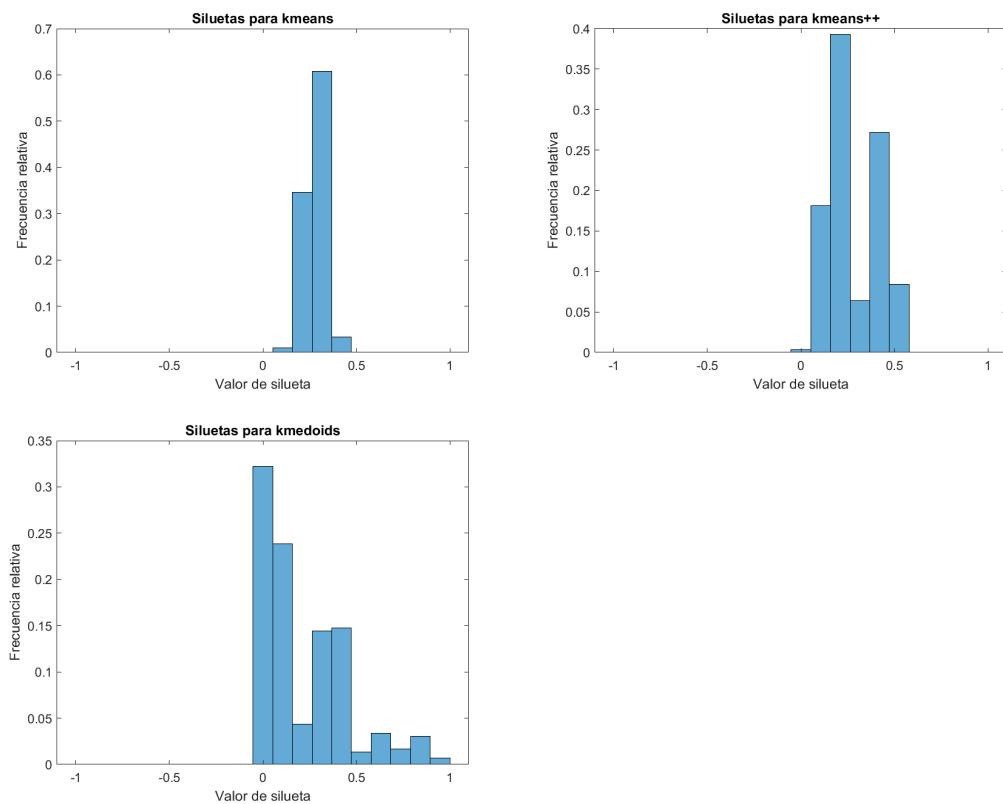
•Experimento 3



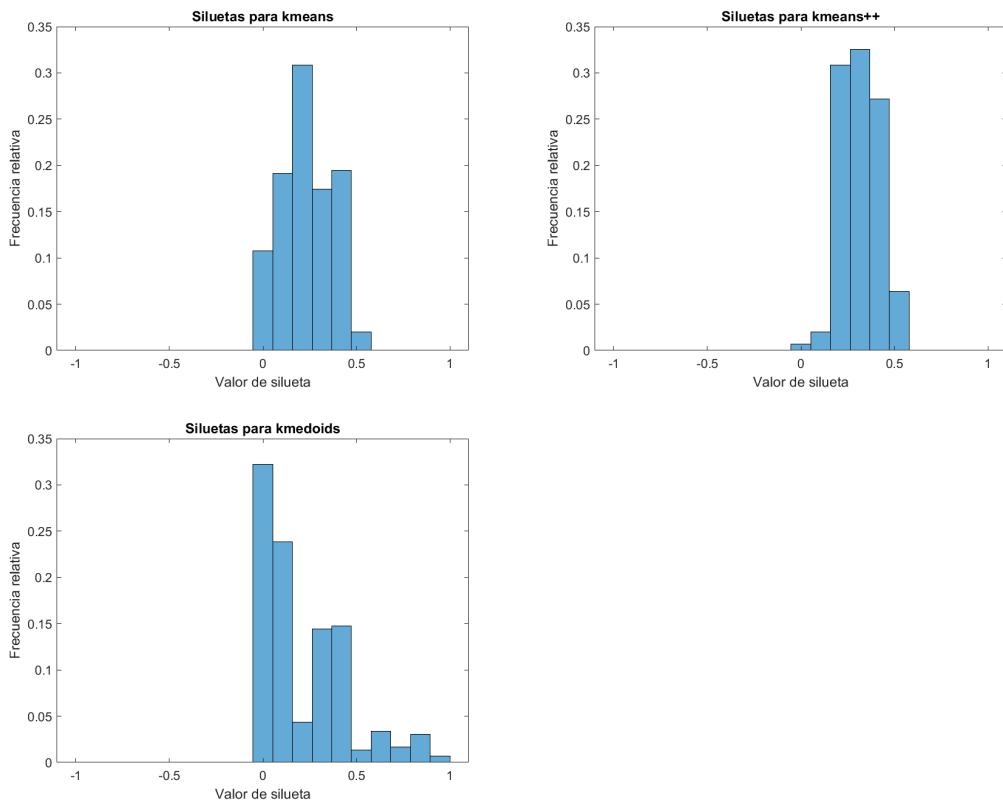
•Experimento 4



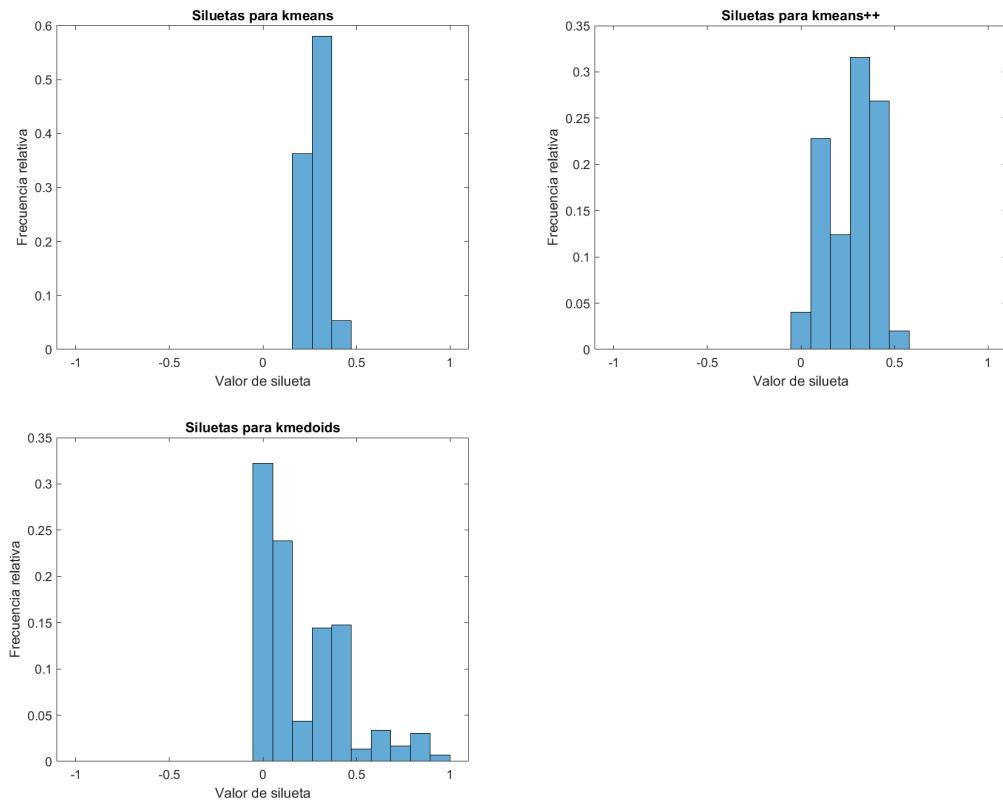
•Experimento 5



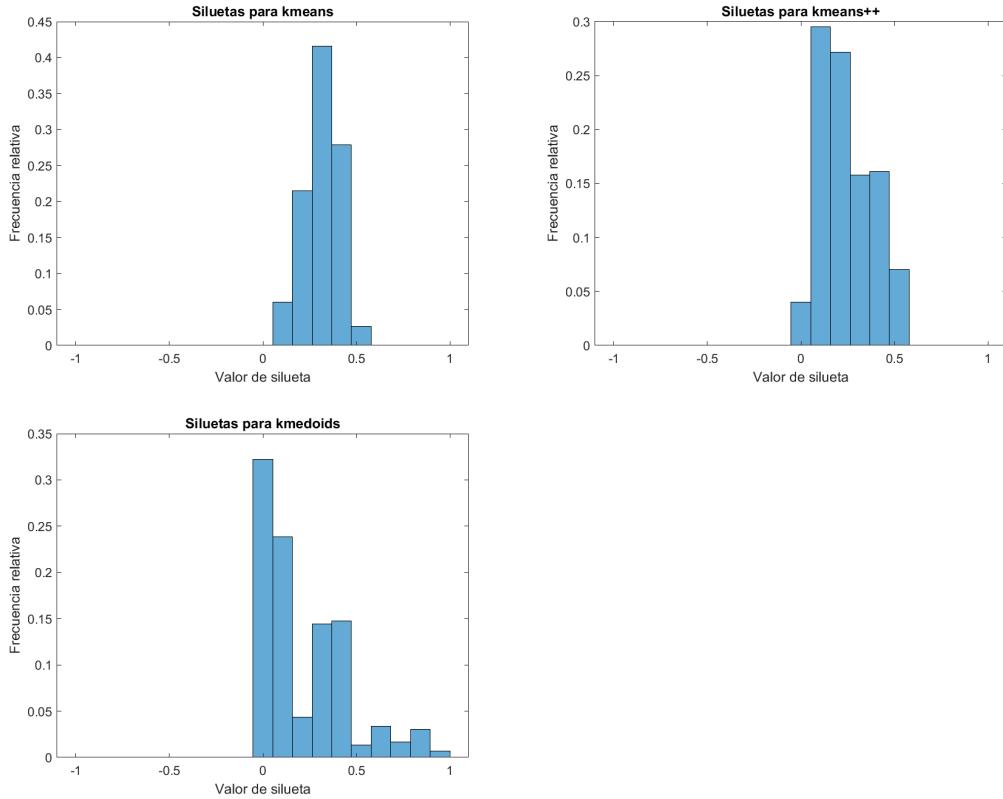
•Experimento 6



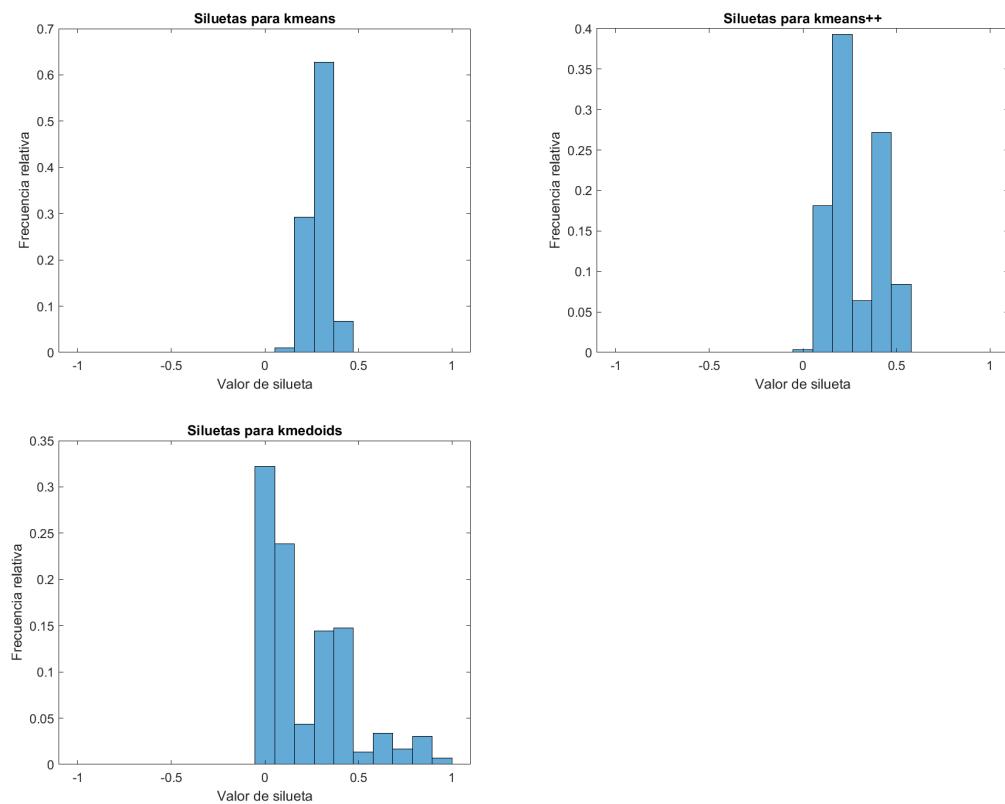
•Experimento 7



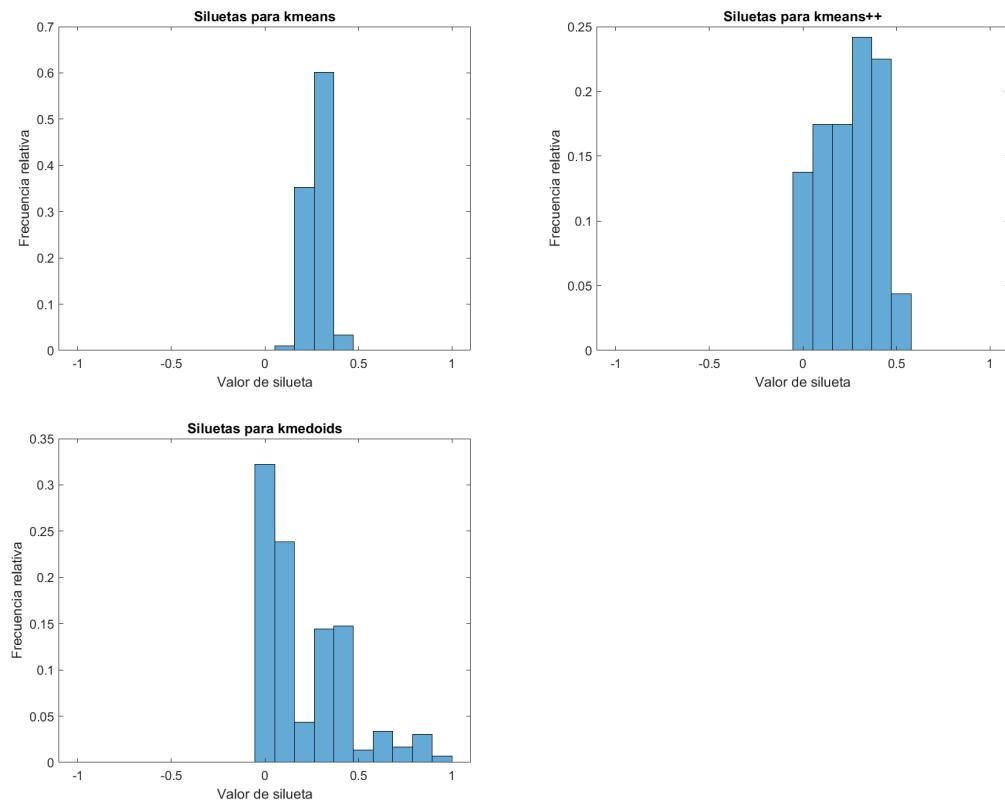
•Experimento 8



•Experimento 9

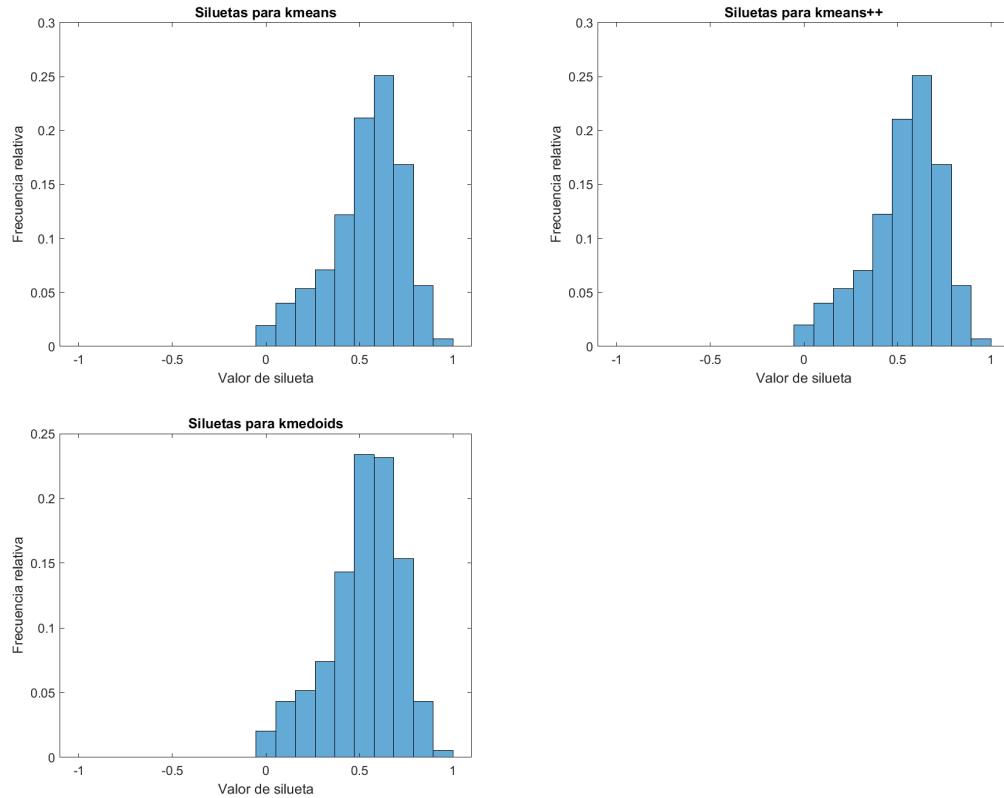


•Experimento 10

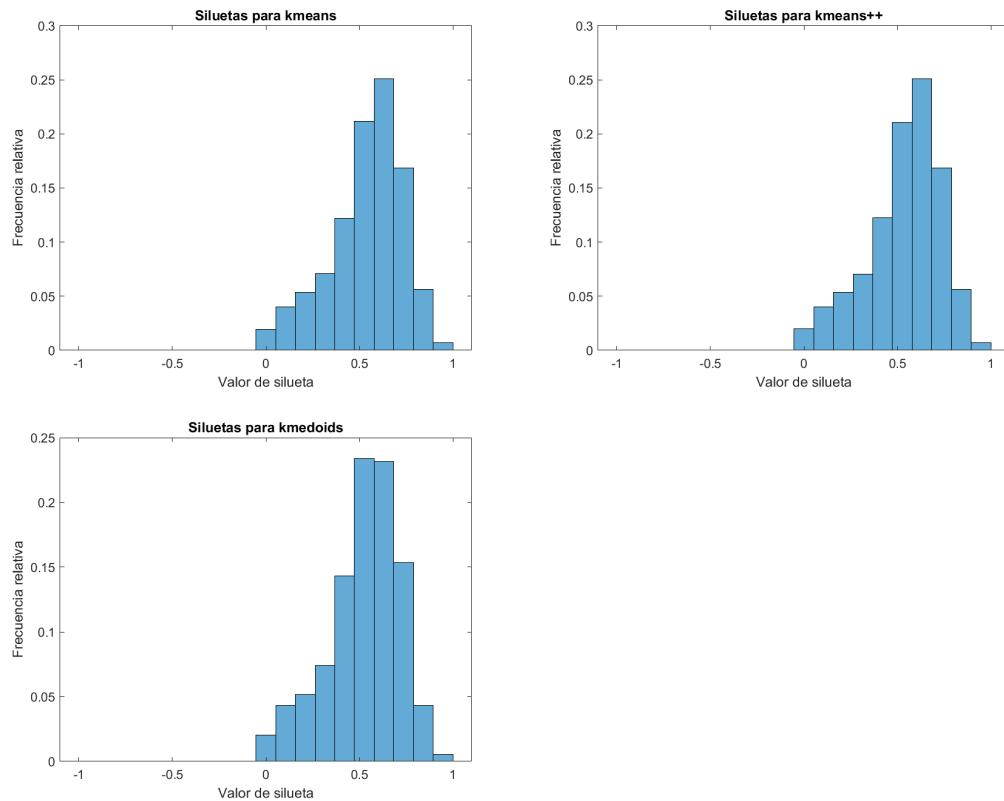


DataSet 3

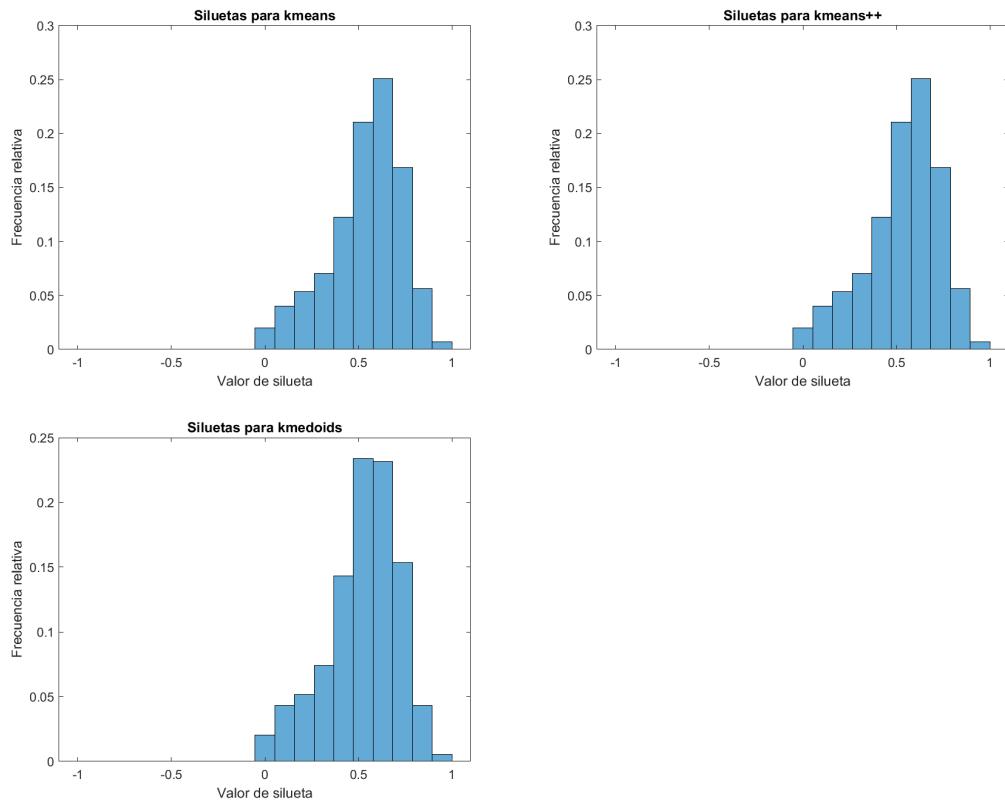
·Experimento 1



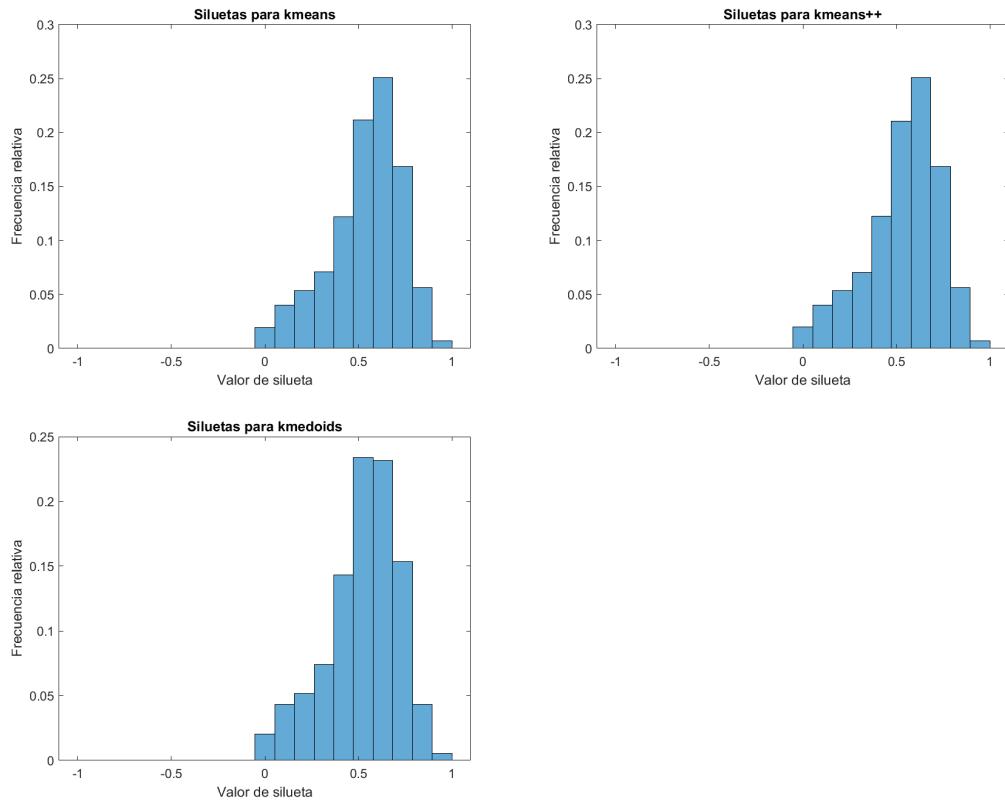
·Experimento 2



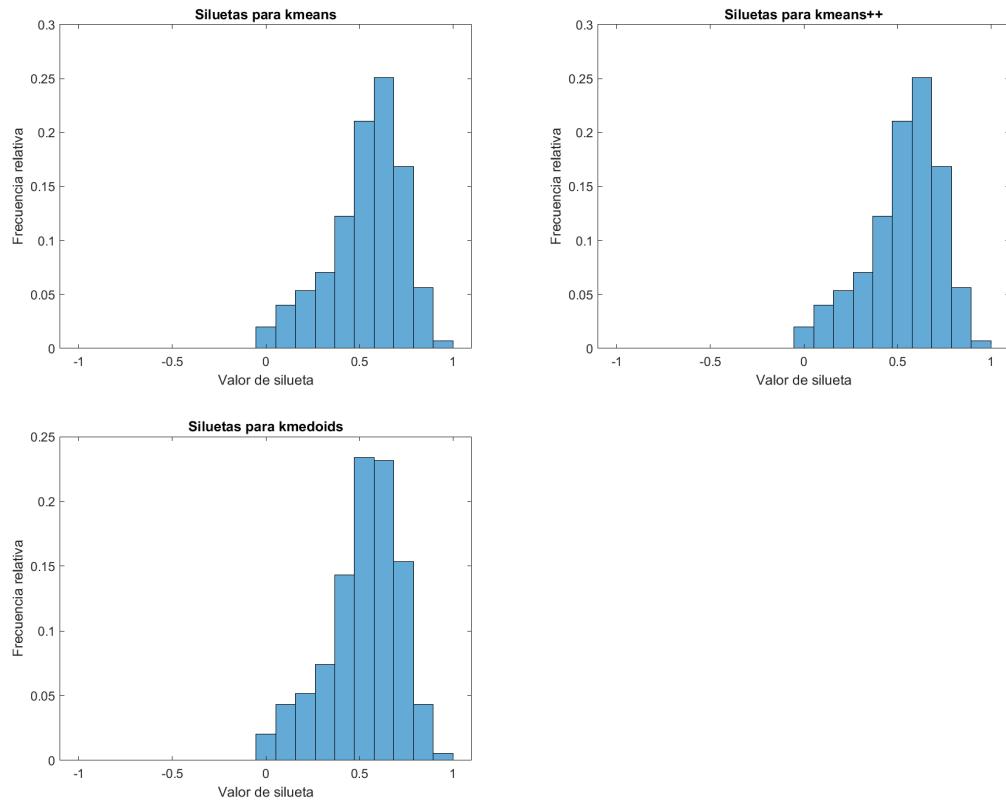
•Experimento 3



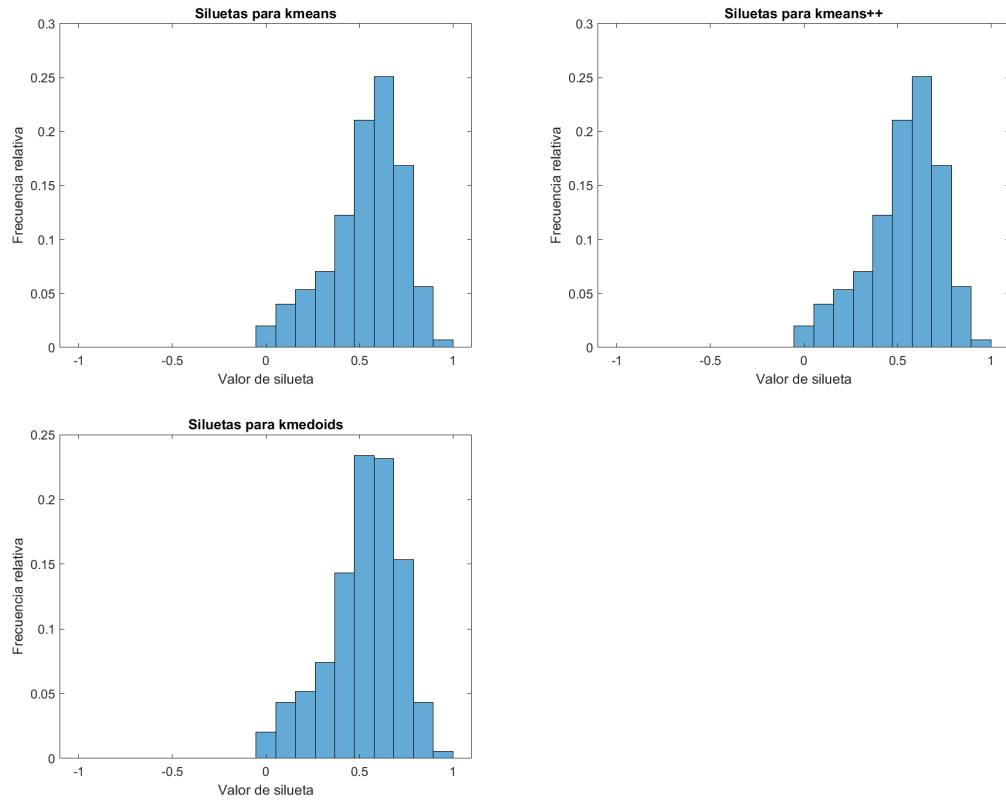
•Experimento 4



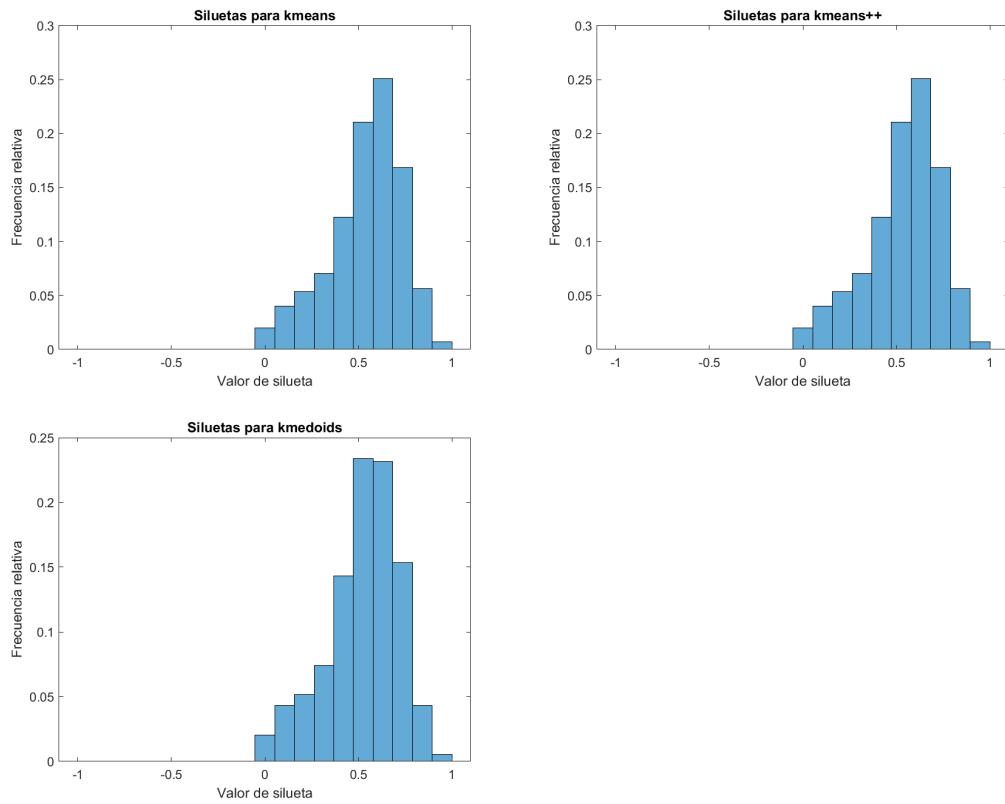
• Experimento 5



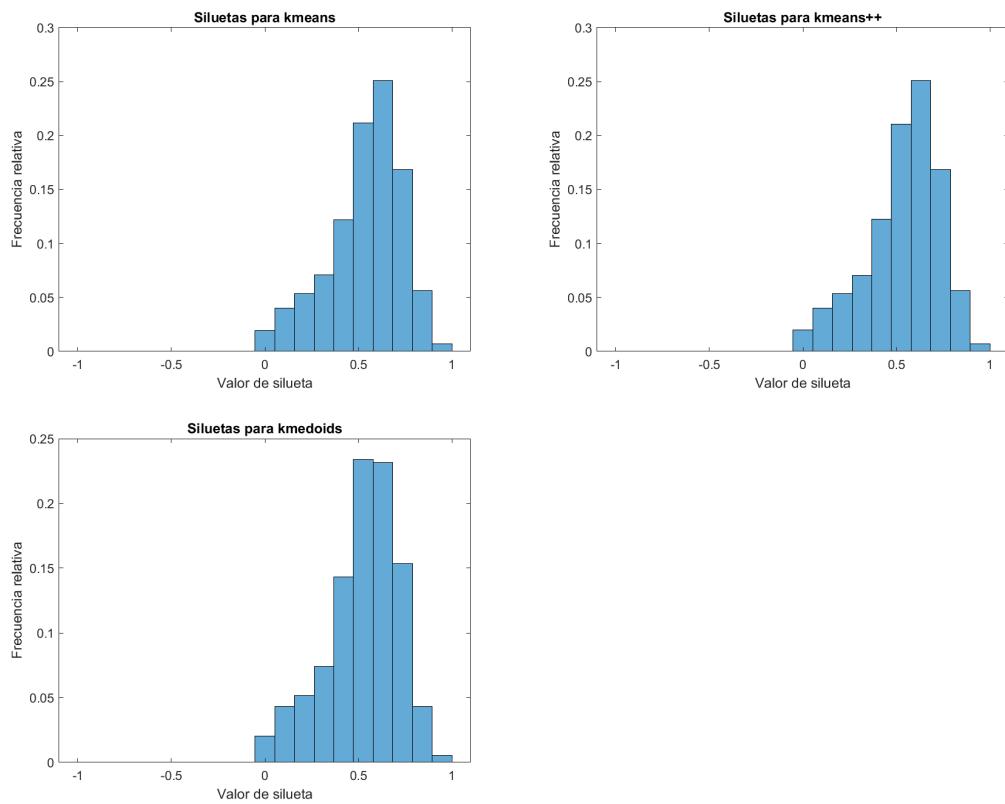
• Experimento 6



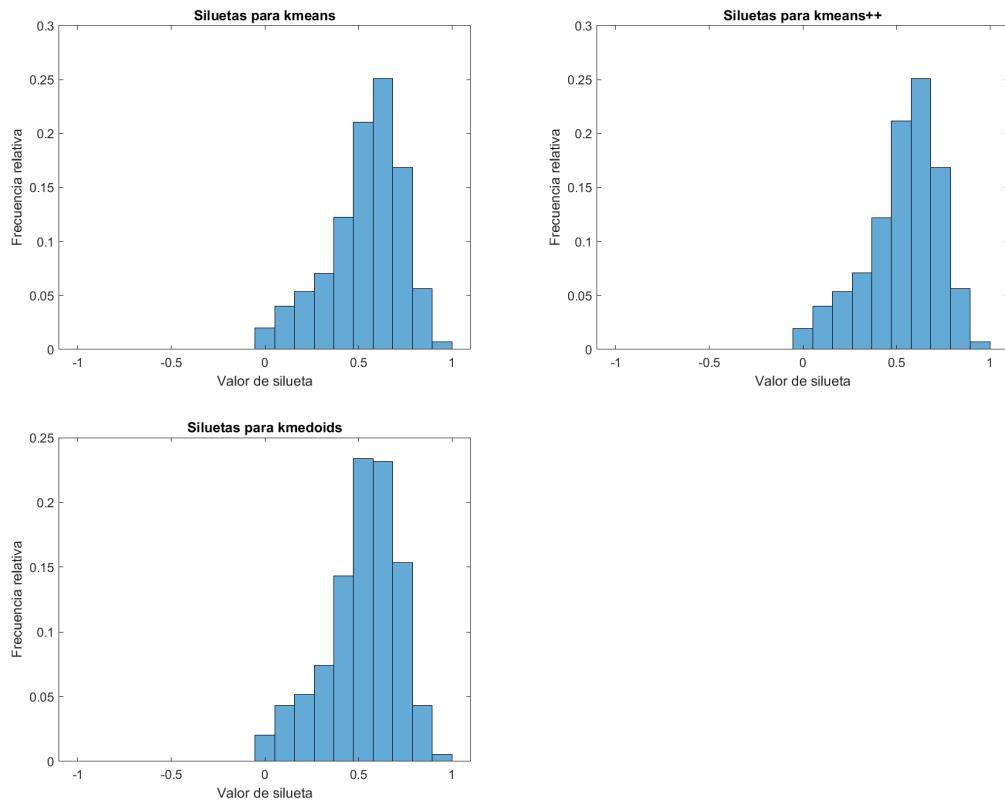
•Experimento 7



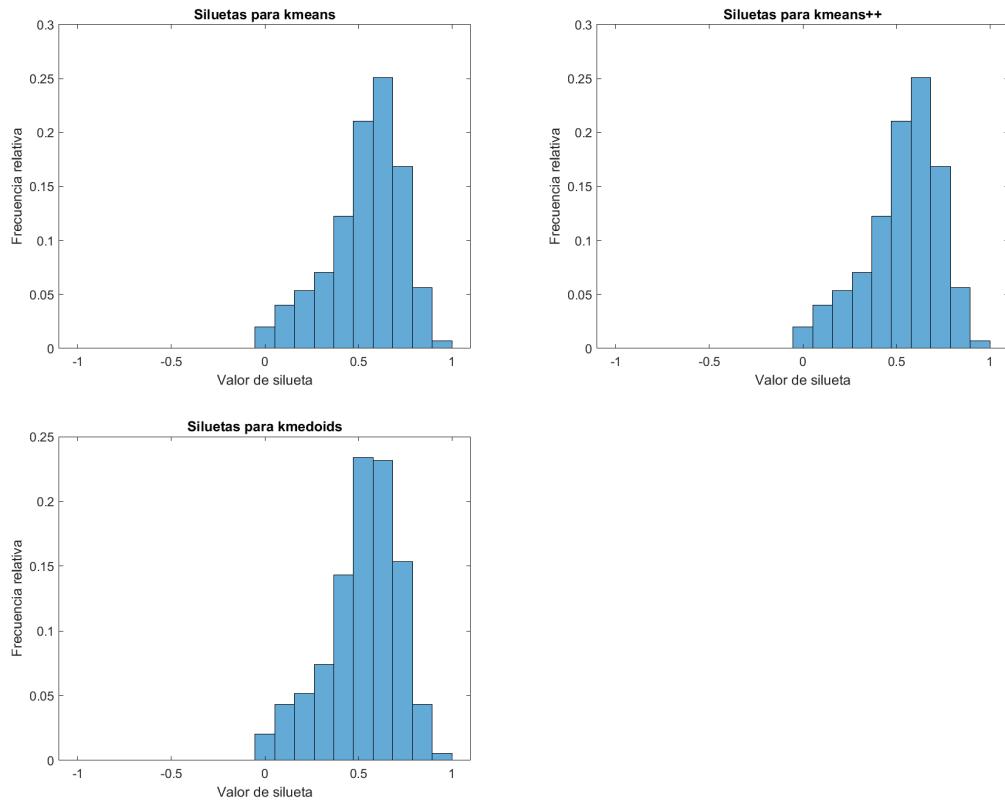
•Experimento 8



•Experimento 9

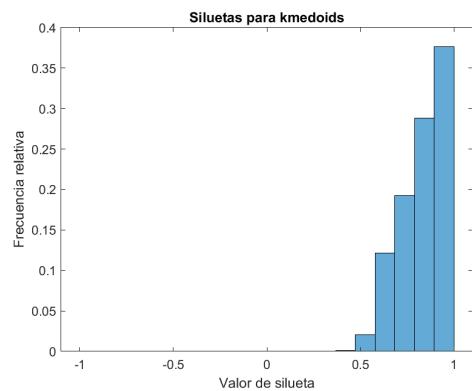
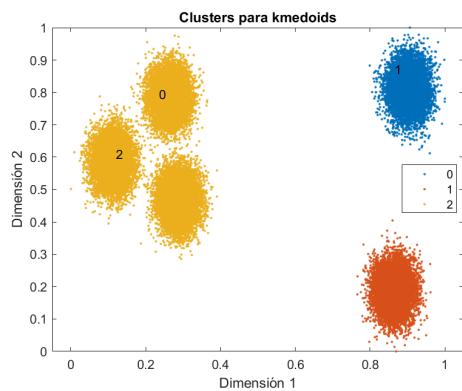
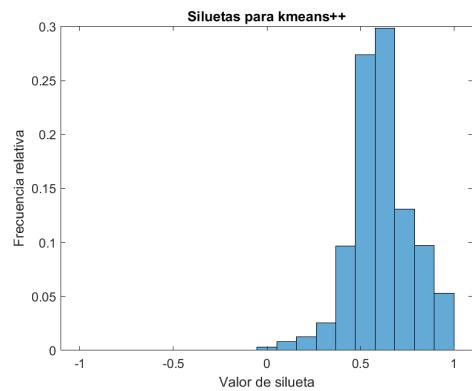
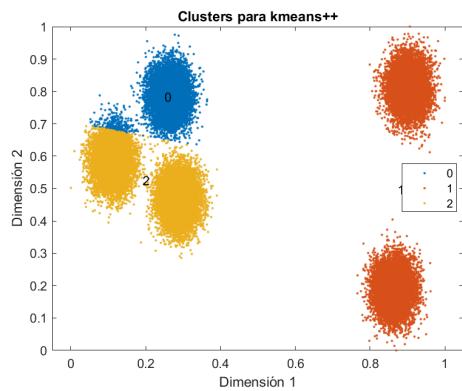
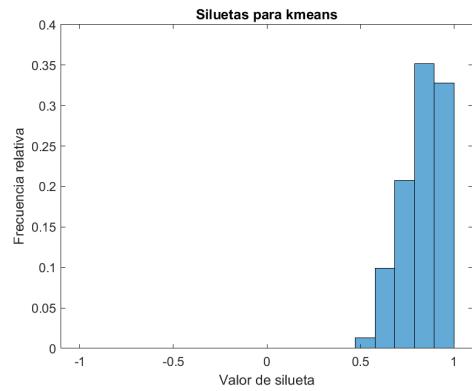
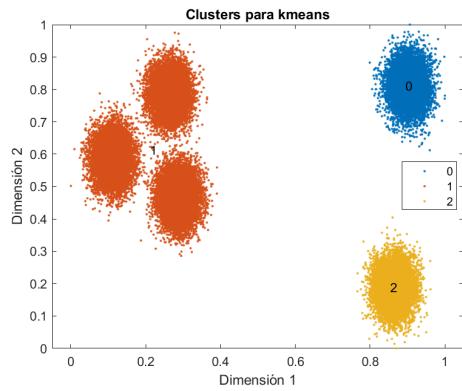


•Experimento 10

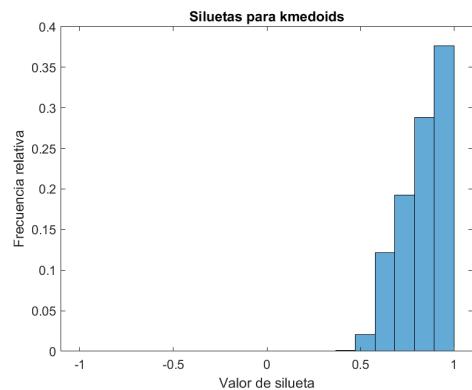
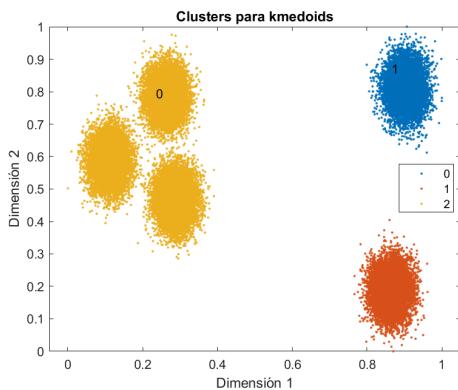
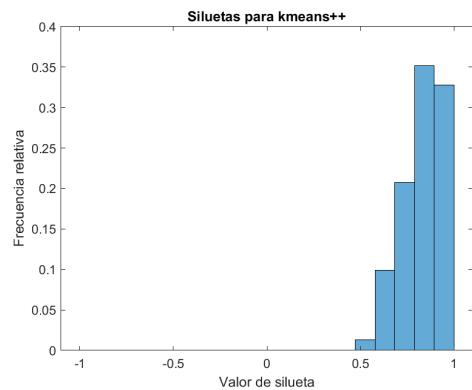
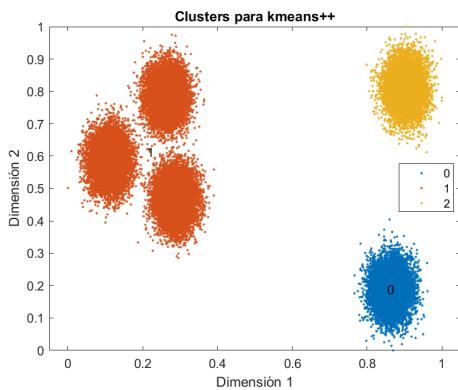
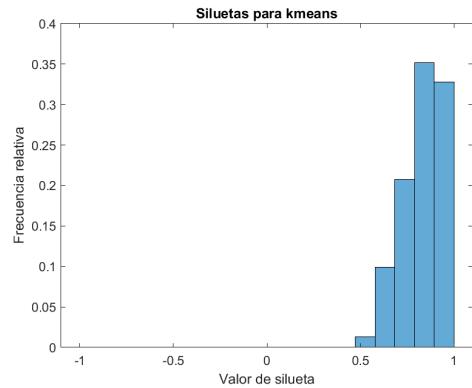
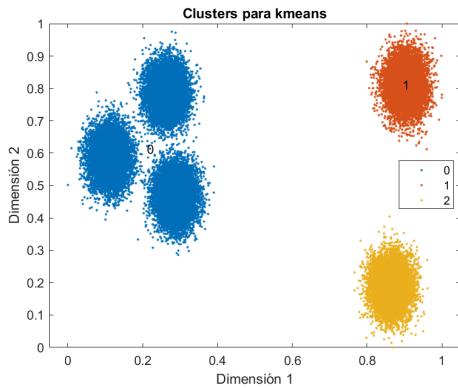


DataSet 6

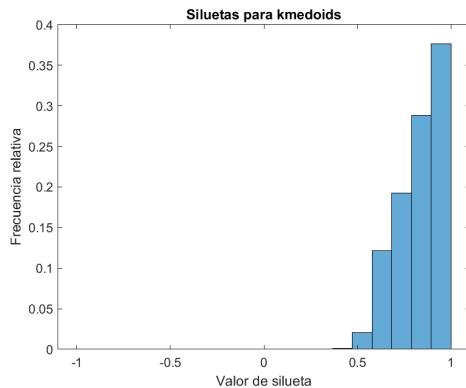
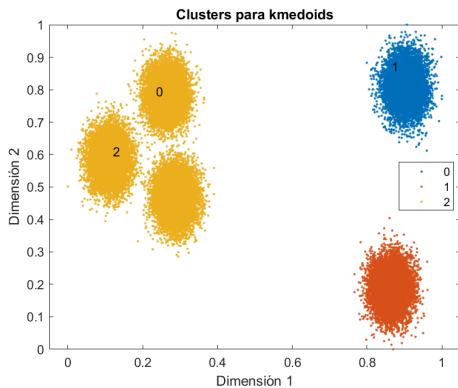
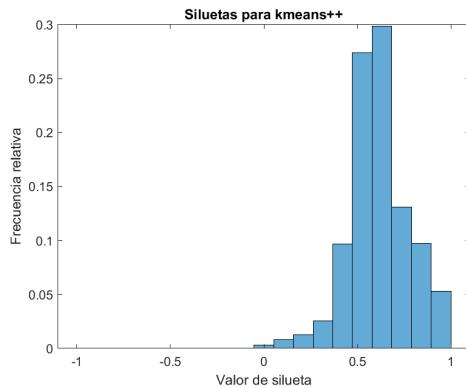
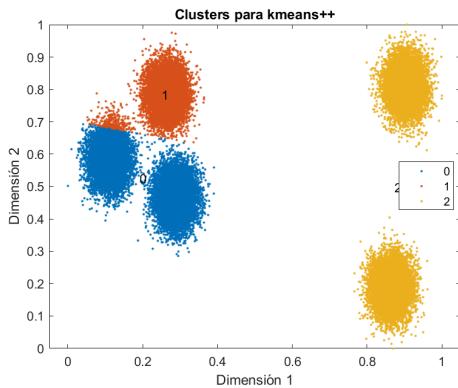
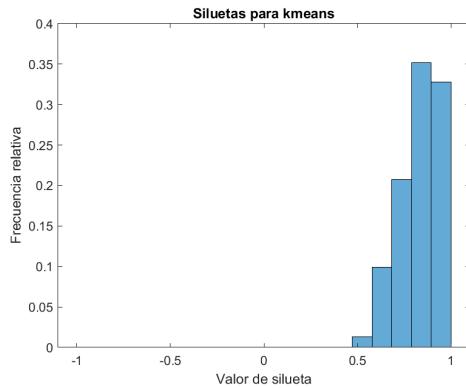
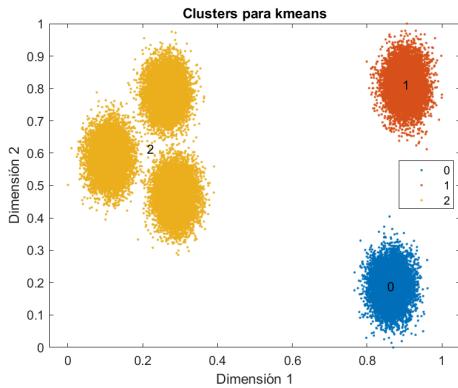
·Experimento 1



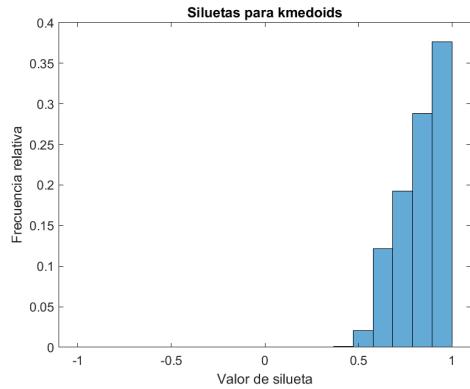
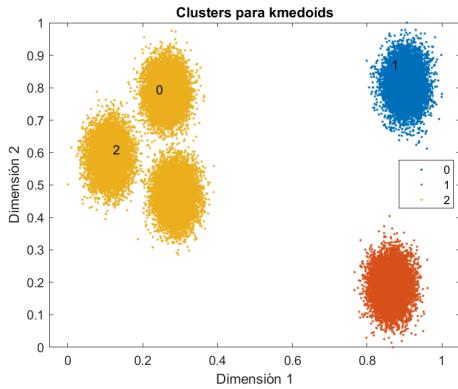
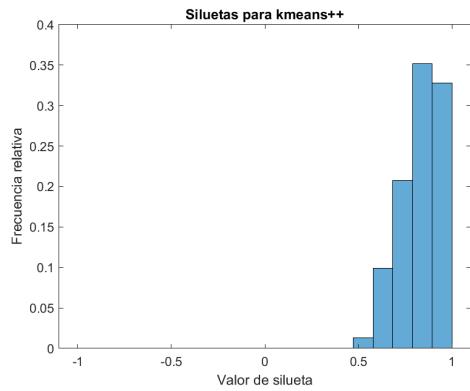
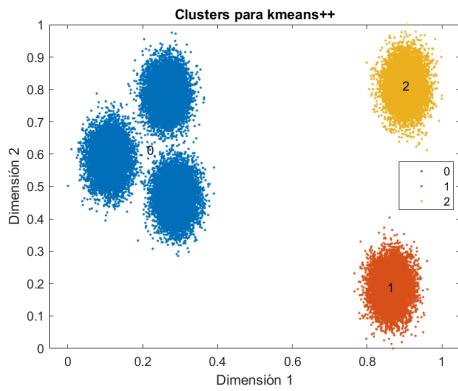
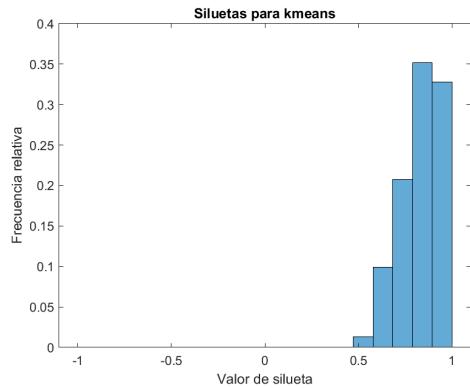
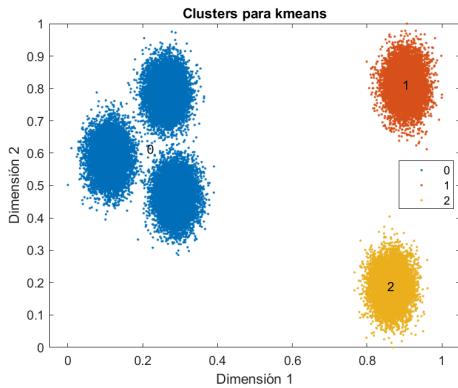
• Experimento 2



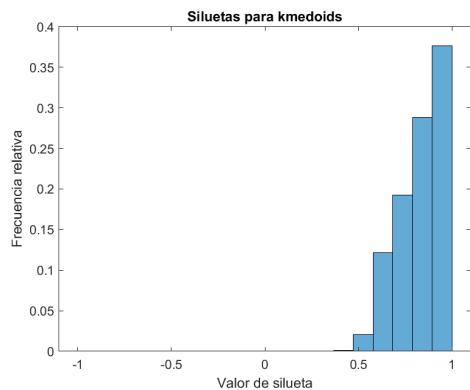
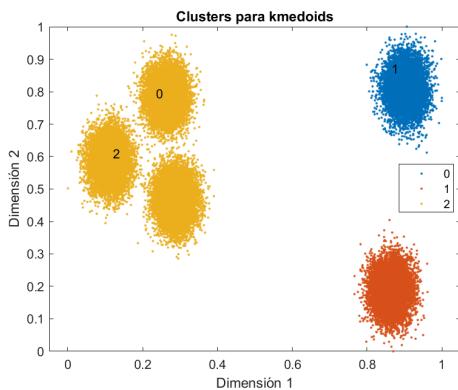
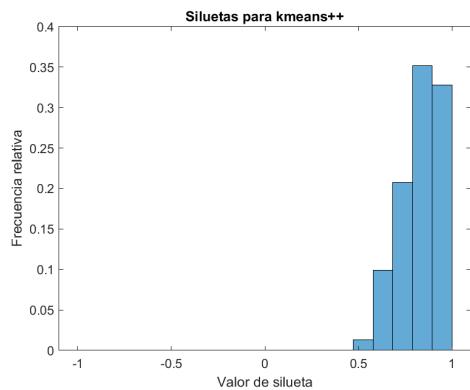
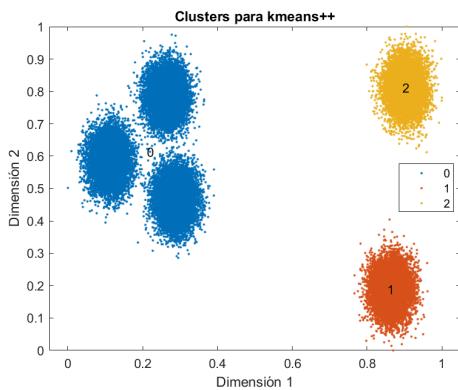
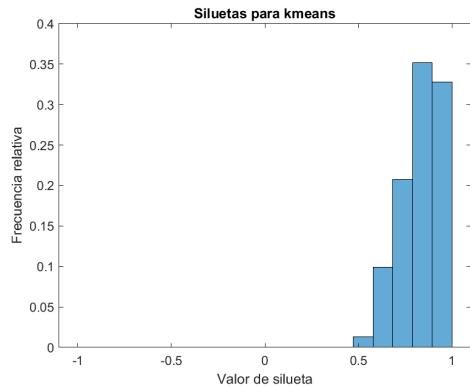
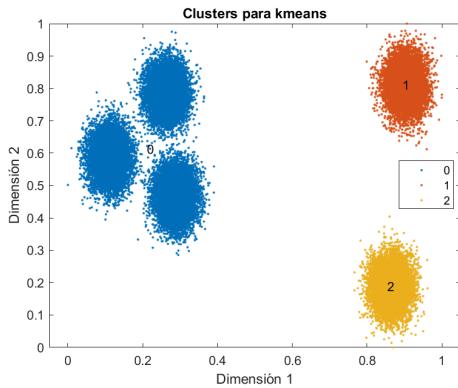
•Experimento 3



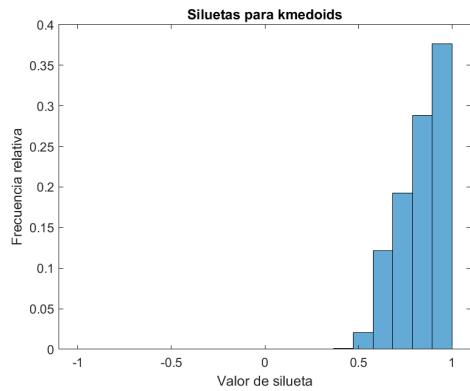
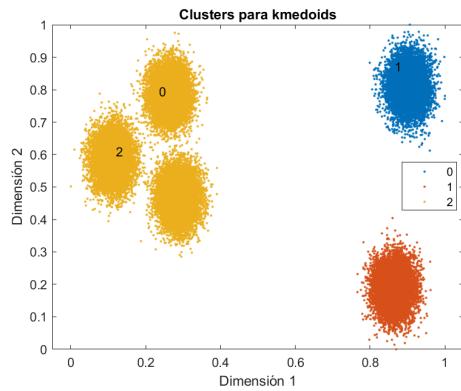
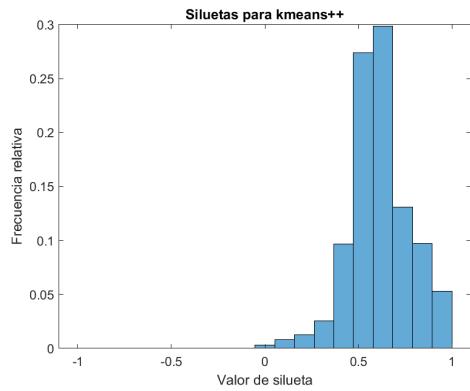
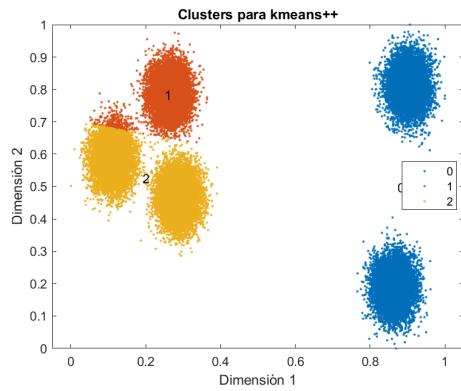
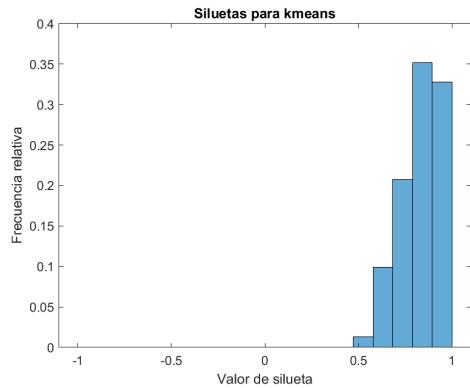
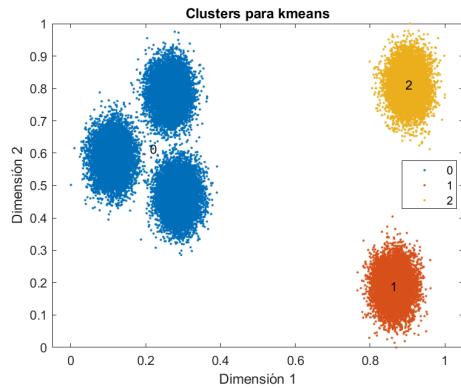
• Experimento 4



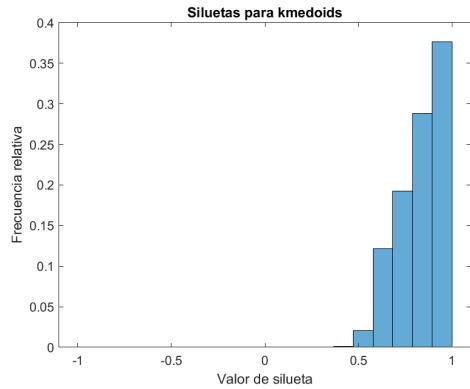
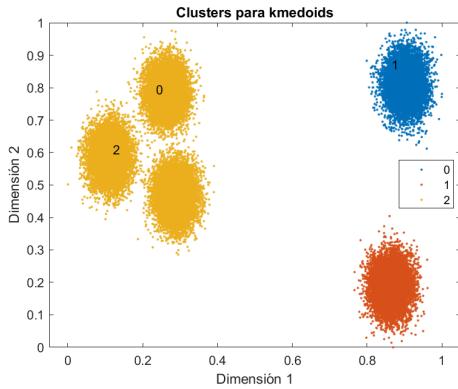
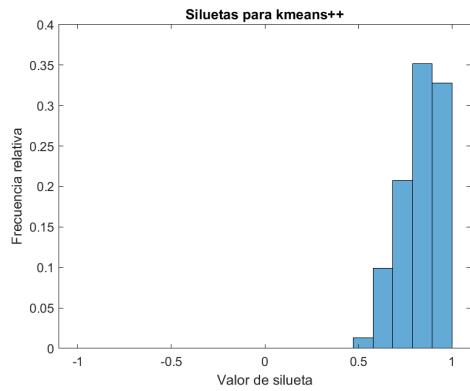
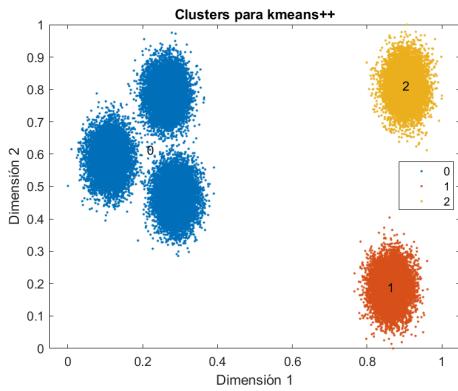
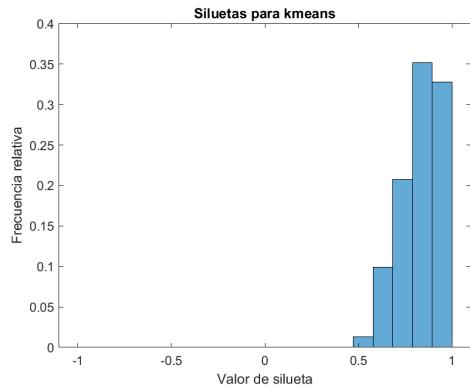
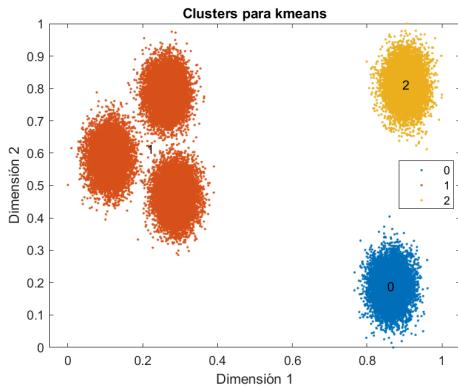
• Experimento 5



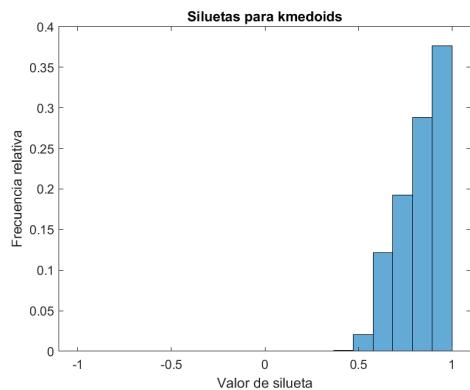
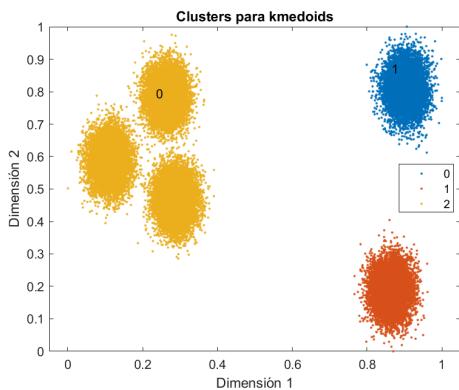
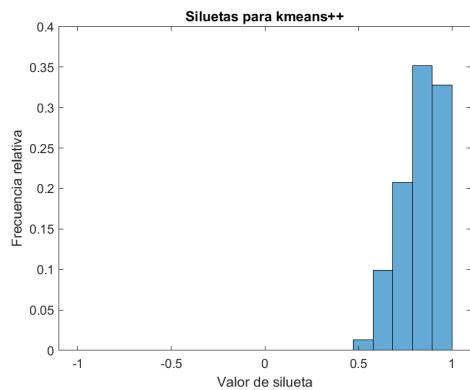
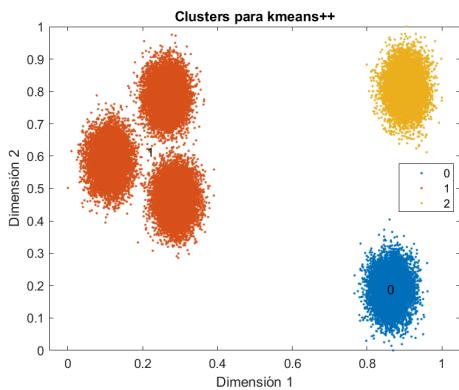
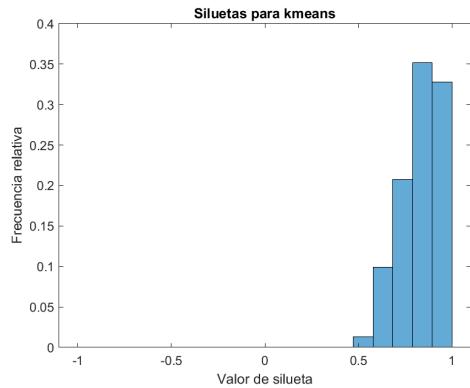
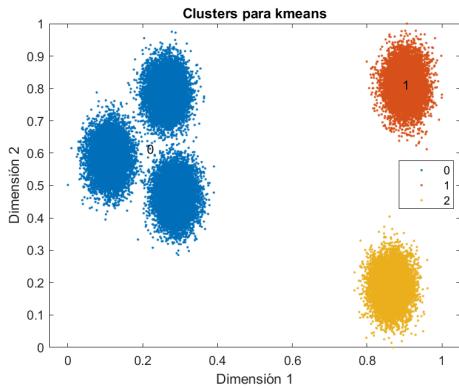
•Experimento 6



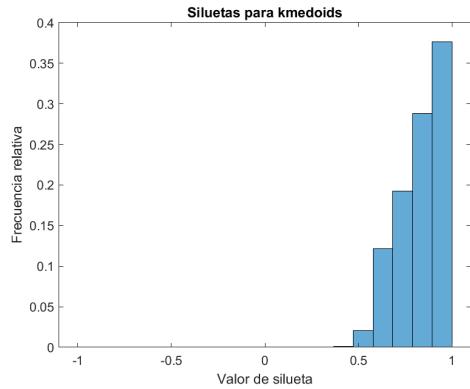
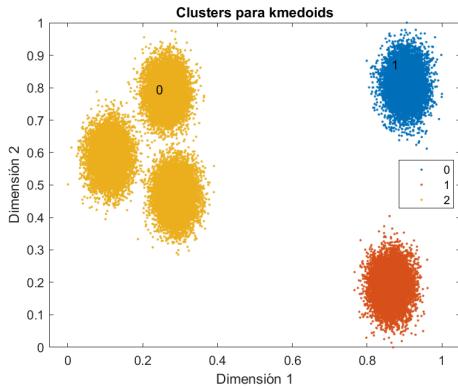
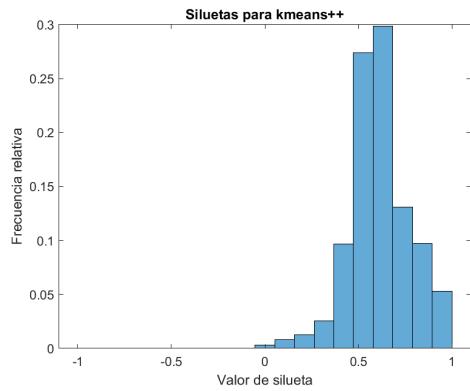
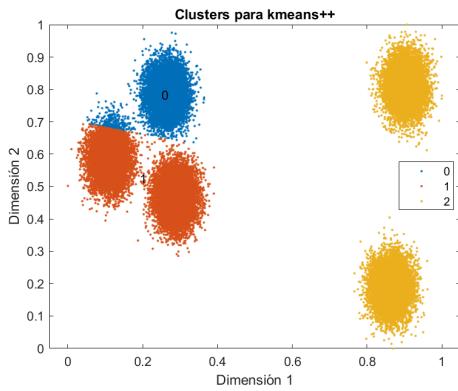
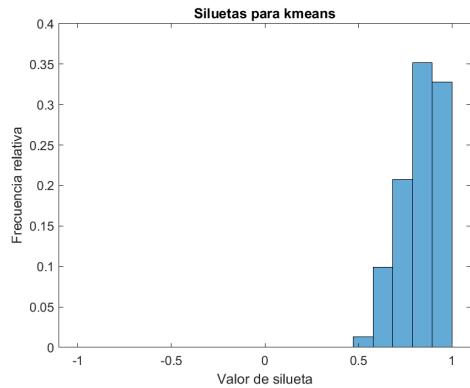
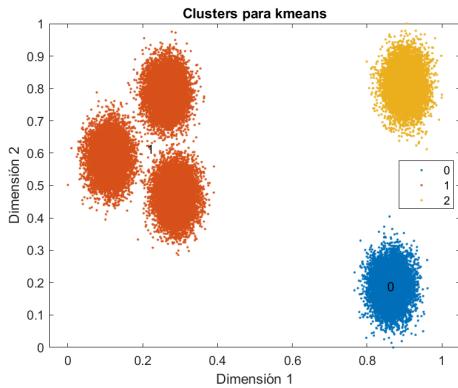
• Experimento 7



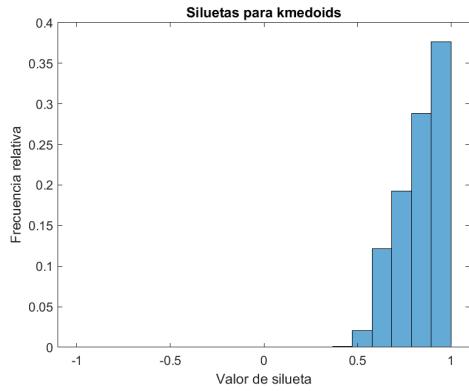
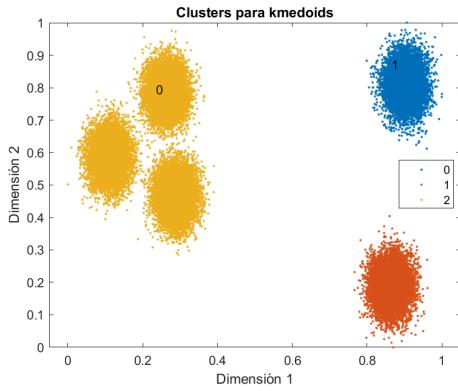
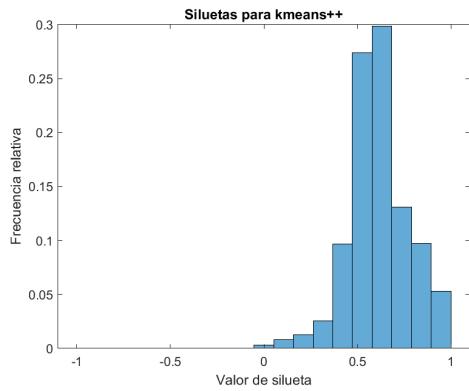
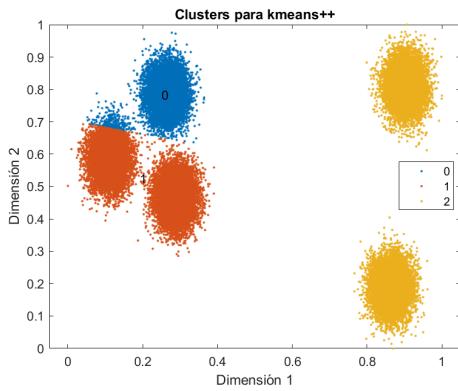
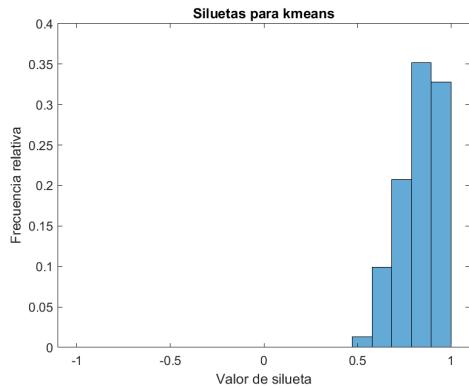
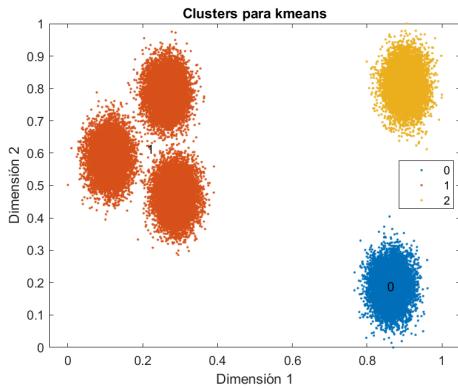
•Experimento 8



•Experimento 9



• Experimento 10



EVALUACIÓN EXTERNA

DataSet Propio (k = 3, d = 8, n = 3000)

Número de Exp.	k-means	k-means++	k-medoids
Experimento 1	0.777704	1	0.723812
Experimento 2	0.777704	0.777704	0.728706
Experimento 3	0.777704	1	1
Experimento 4	1	1	1
Experimento 5	0.333111	0.777704	1
Experimento 6	1	1	1
Experimento 7	0.333111	1	1
Experimento 8	0.333111	1	1
Experimento 9	0.777704	1	1
Experimento 10	0.777704	1	0.723482
Promedio	0,688785	0,955541	0,917600

DataSet 1 (k = 7, d = 16, n = 13611)

Número de Exp.	k-means	k-means++	k-medoids
Experimento 1	0,836955	0,897007	0.854959
Experimento 2	0,907875	0,891388	0.854959
Experimento 3	0,83623	0,897654	0.854959
Experimento 4	0,83623	0,891283	0.854959
Experimento 5	0,836955	0,907808	0.854959
Experimento 6	0,836955	0,897826	0.854959
Experimento 7	0,836946	0,907846	0.854959
Experimento 8	0,836955	0,897847	0.854959
Experimento 9	0,770107	0,836946	0.854959
Experimento 10	0,836955	0,851595	0.854959
Promedio	0,8372163	0,88772	0,854959

Dataset 2 ($k = 2$, $d = 12$, $n = 299$)

Número de Exp.	k-means	k-means++	k-medoids
Experimento 1	0,498597	0,499944	0.514085
Experimento 2	0,506004	0,499944	0.514085
Experimento 3	0,514085	0,49927	0.514085
Experimento 4	0,514085	0,503041	0.514085
Experimento 5	0,504433	0,499001	0.514085
Experimento 6	0,514085	0,504433	0.514085
Experimento 7	0,504433	0,504433	0.514085
Experimento 8	0,504433	0,514085	0.514085
Experimento 9	0,499944	0,504433	0.514085
Experimento 10	0,498328	0,498373	0.514085
Promedio	0,5058427	0,5026957	0,514085

Dataset 3 ($k = 2$, $d = 7$, $n = 3810$)

Número de Exp.	k-means	k-means++	k-medoids
Experimento 1	0.844344	0.844344	0.844344
Experimento 2	0.844344	0.844344	0.844344
Experimento 3	0.844344	0.844344	0.844344
Experimento 4	0.844344	0.844344	0.844344
Experimento 5	0.844344	0.844344	0.844344
Experimento 6	0.844344	0.844344	0.844344
Experimento 7	0.844344	0.844344	0.844344
Experimento 8	0.844344	0.844344	0.844344
Experimento 9	0.844344	0.844344	0.844344
Experimento 10	0.844344	0.844344	0.844344
Promedio	0,844344	0,844344	0,844344

Dataset 4 ($k = ??$, $d = 10$, $n = 980$)

Número de Exp.	k-means	k-means++	k-medoids
k-means	1	0,89	0,82
k-means++	0,89	1	0,84
k-medoids	0,82	0,84	1

Número de Exp.	kmeans vs kmeans++	kmmedoids vs kmeans ++	kmeans vs kmmedoids
Experimento 1	0,851844	0,835786	0,792529
Experimento 2	0,864297	0,83855	0,779217
Experimento 3	0,89477	0,845711	0,837846
Experimento 4	0,85245	0,836124	0,826087
Experimento 5	0,916014	0,846095	0,795683
Experimento 6	0,869692	0,825124	0,841187
Experimento 7	0,887699	0,83715	0,817508
Experimento 8	0,934967	0,850445	0,828109
Experimento 9	0,932786	0,843468	0,8411
Experimento 10	0,855875	0,832789	0,824488
Promedio	0,8860394	0,8391242	0,8183754

Dataset 5 ($k = ??$, $d = 24$, $n = 5456$)

Número de Exp.	k-means	k-means++	k-medoids
k-means	1	0,94	0,88
k-means++	0,94	1	0,88
k-medoids	0,88	0,88	1

Número de Exp.	kmeans vs kmeans++	kmmedoids vs kmeans ++	kmeans vs kmmedoids
Experimento 1	0,917072	0,875296	0,872815
Experimento 2	0,942923	0,865287	0,881408
Experimento 3	0,942303	0,880241	0,881171
Experimento 4	0,939744	0,868533	0,871708
Experimento 5	0,939585	0,877274	0,884037
Experimento 6	0,931074	0,880946	0,879521
Experimento 7	0,966714	0,879787	0,881365
Experimento 8	0,922111	0,883426	0,870944
Experimento 9	0,952189	0,865585	0,870938
Experimento 10	0,921113	0,882113	0,872006
Promedio	0,9374845	0,8758488	0,8765913

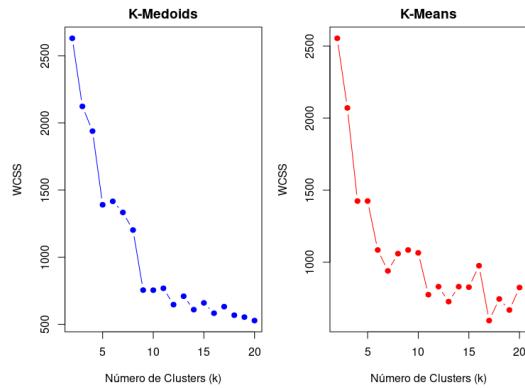
Dataset 6 ($k = 3$, $d = 2$, $n = 50000$)

Número de Exp.	k-means	k-means++	k-medoids
k-means	1	0,83	0,95
k-means++	0,83	1	0,98
k-medoids	0,95	0,98	1

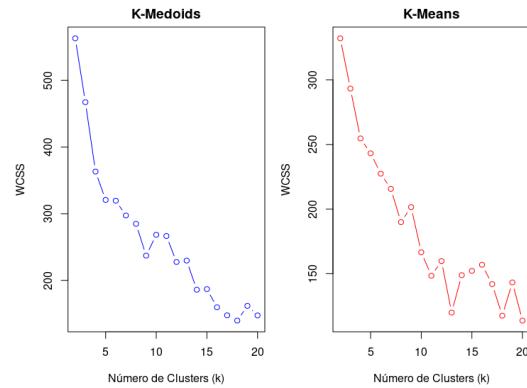
Número de Exp.	kmeans vs kmeans++	kmmedoids vs kmeans ++	kmeans vs kmmedoids
Experimento 1	0,757979	0,757979	1
Experimento 2	0,757995	1	1
Experimento 3	1	1	1
Experimento 4	0,757979	1	1
Experimento 5	1	1	1
Experimento 6	0,757979	1	1
Experimento 7	0,757979	1	1
Experimento 8	0,757979	1	1
Experimento 9	0,757995	1	0,757979
Experimento 10	1	1	0,757995
Promedio	0,8305885	0,9757979	0,9515974

OPTIMIZACIÓN de K

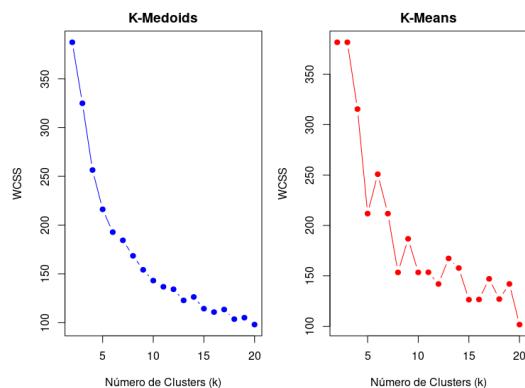
Dataset 1: Comparación de WCSS por Número de Clusters



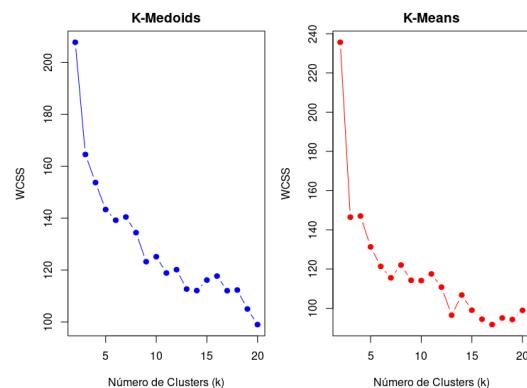
Dataset 2: Comparación de WCSS por Número de Clusters



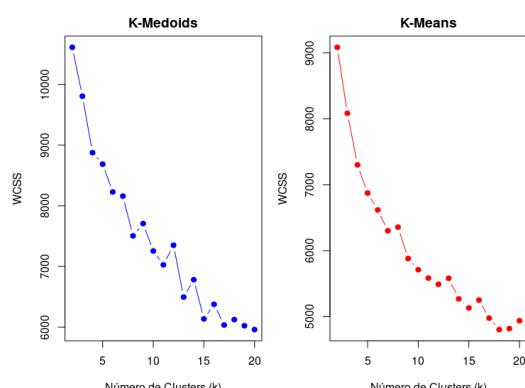
Dataset 3: Comparación de WCSS por Número de Clusters



Dataset 4: Comparación de WCSS por Número de Clusters



Dataset 5: Comparación de WCSS por Número de Clusters



Dataset 6: Comparación de WCSS por Número de Clusters

