



Tarea 3: Clustering, K-NN y Regresión

Tema

Sistemas Inteligentes
Clase

Richardson Cárcamo
Estudiante

11711075
#Cuenta

Jueves, 03 de dic de 2020
Fecha



Tabla de contenido

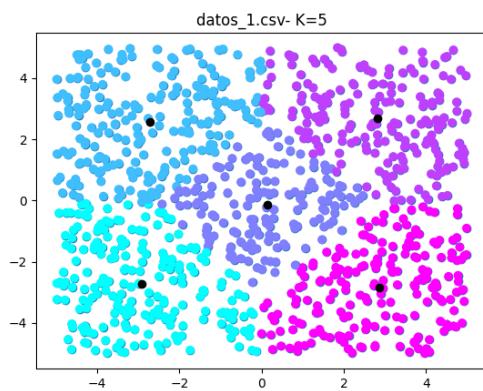
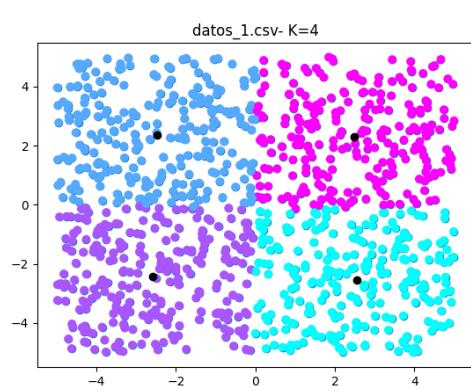
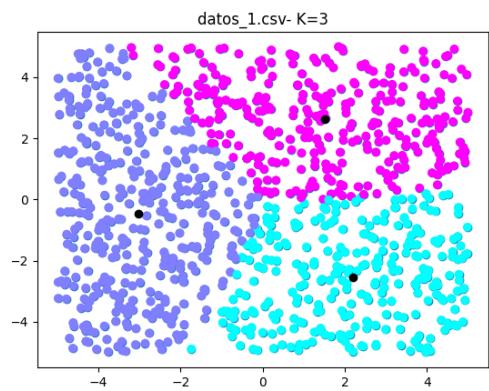
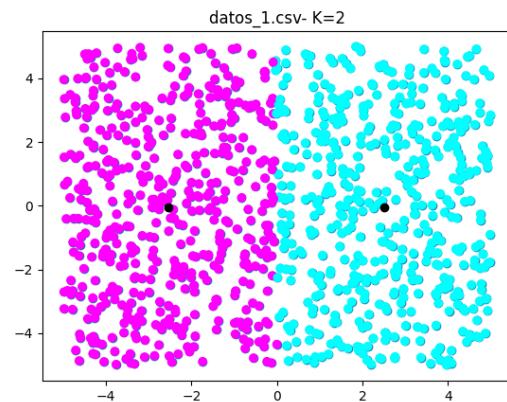
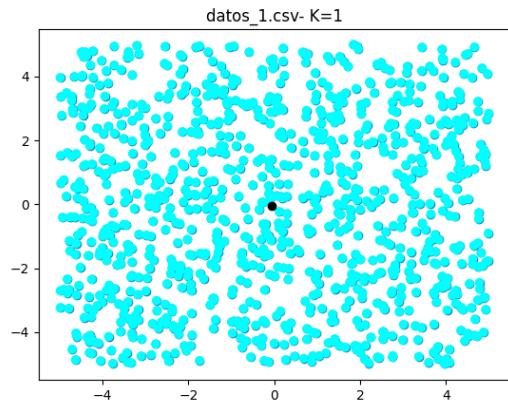
1er Ejercicio.....	3
K-means.....	3
Datos_1.....	3
Datos_2.....	4
Datos_3.....	5
Clusteting Jerárquico.....	6
Datos_1.....	6
Datos_2.....	8
Datos_3.....	10
DBScan.....	12
Datos_1.....	12
Datos_2.....	14
Datos_3.....	16
Mejores resultados según los datasets.....	18
Datos_1.....	18
Datos_2.....	19
Datos_3.....	20

A continuación, se muestran los resultados obtenidos de la Tarea#3 junto con una breve explicación de los resultados considerados más relevantes.

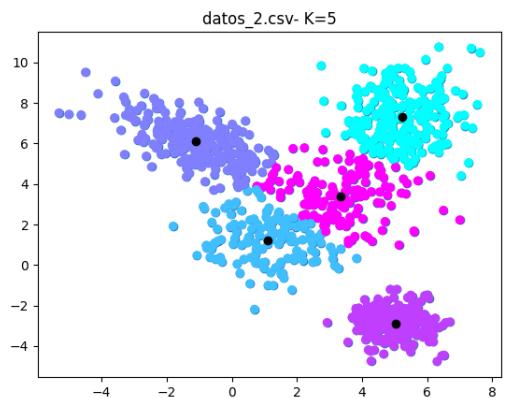
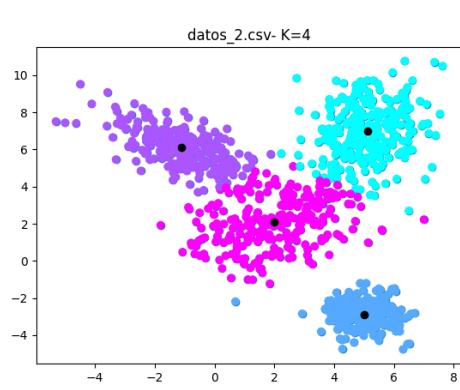
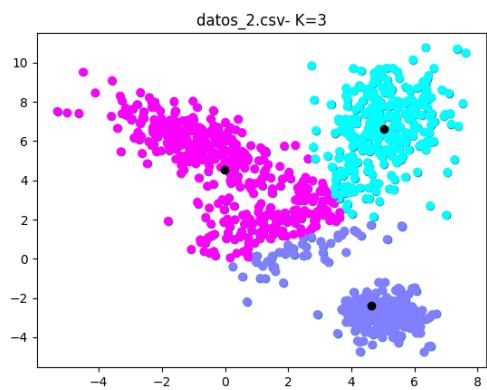
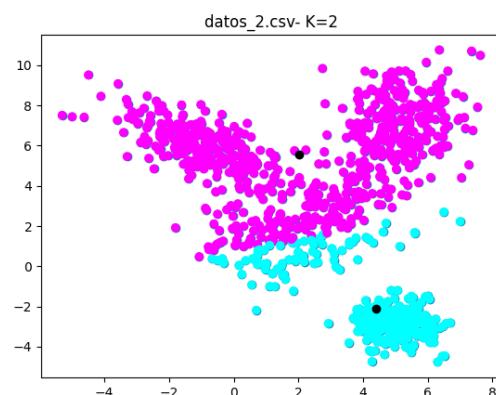
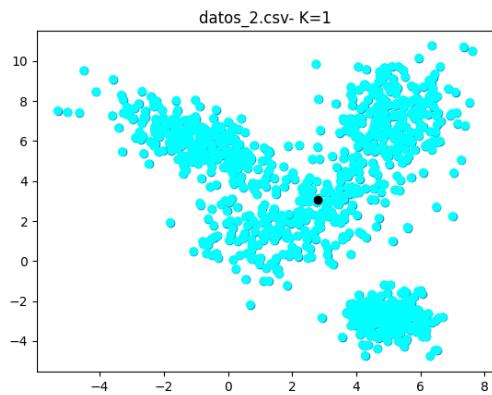
1er Ejercicio

K-means

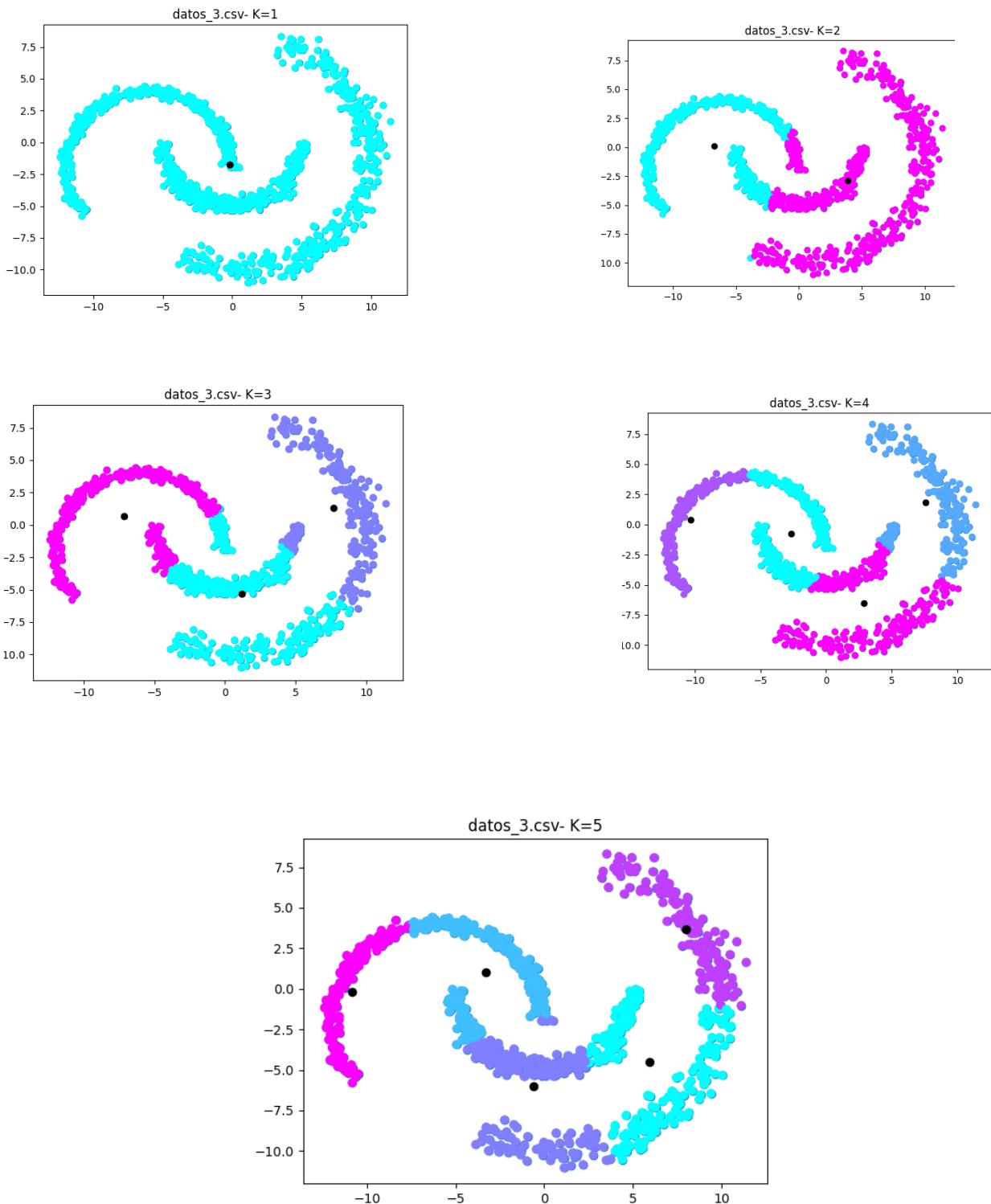
Datos_1



Datos_2



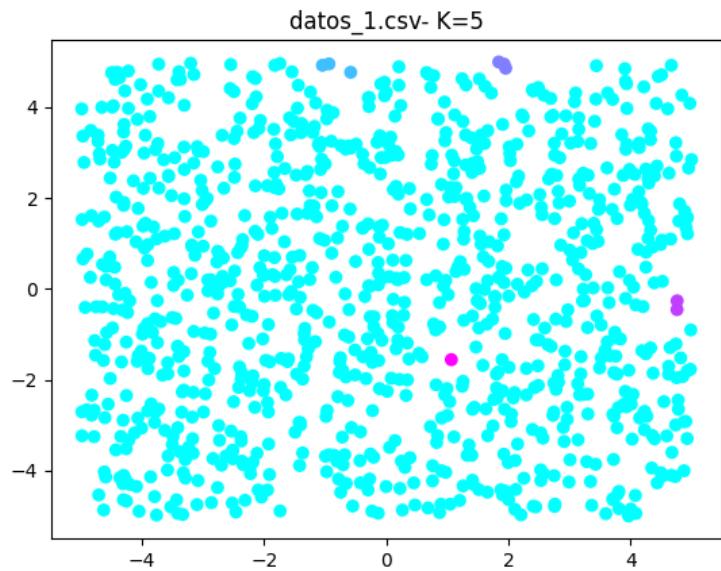
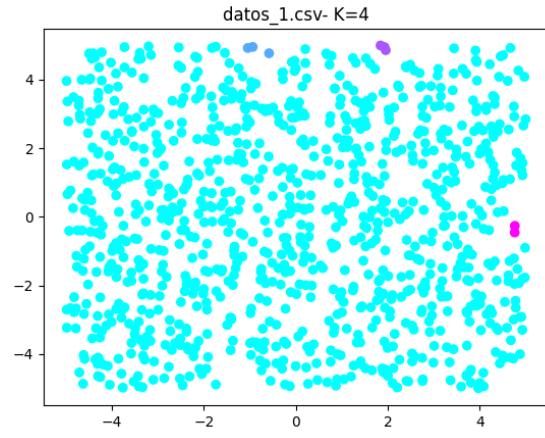
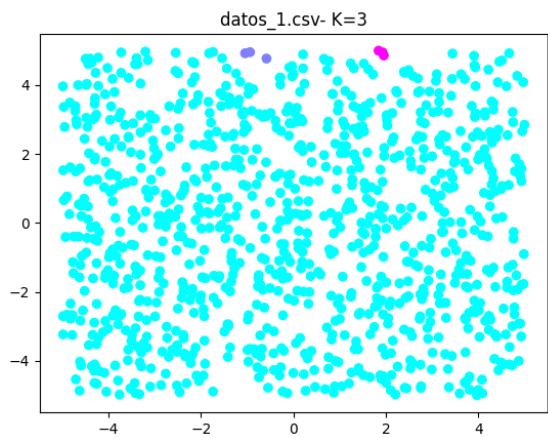
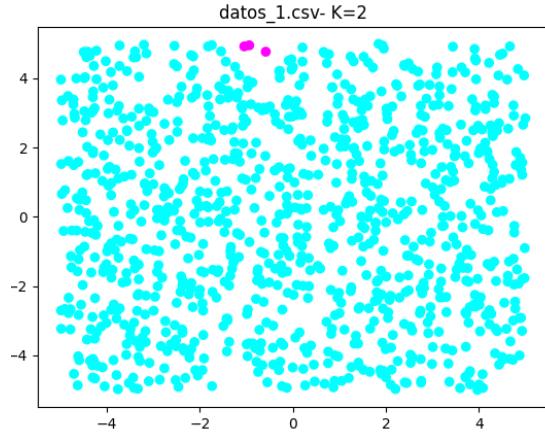
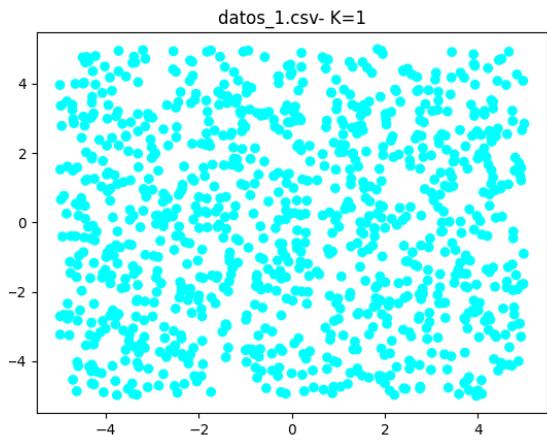
Datos_3



Clusteting Jerárquico

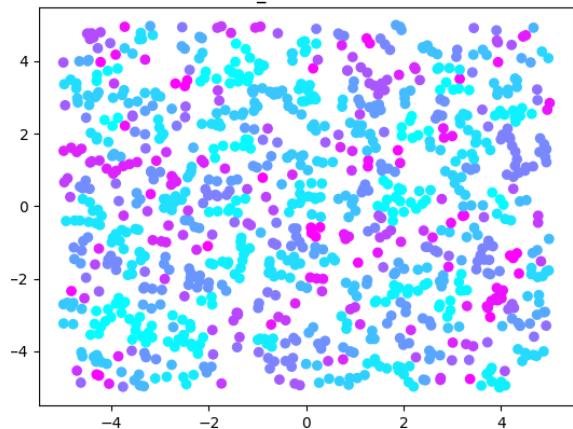
Datos_1

Por Valores K

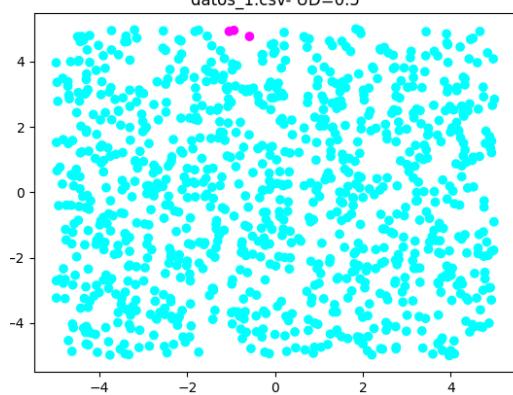


Por Distancia

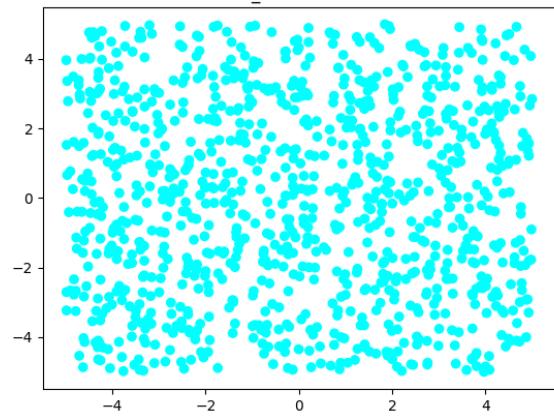
datos_1.csv- UD=0.25



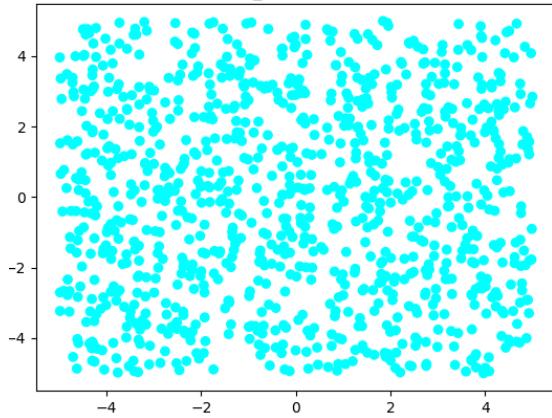
datos_1.csv- UD=0.5



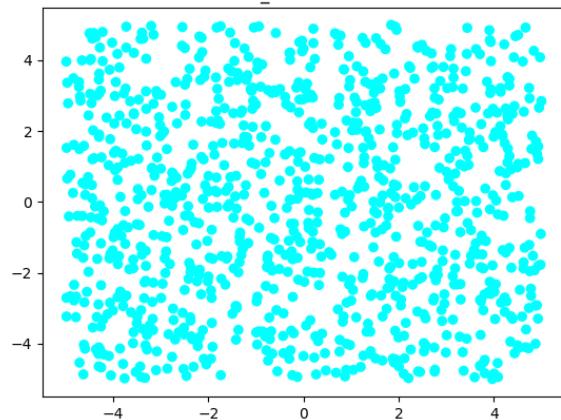
datos_1.csv- UD=0.75



datos_1.csv- UD=1.0

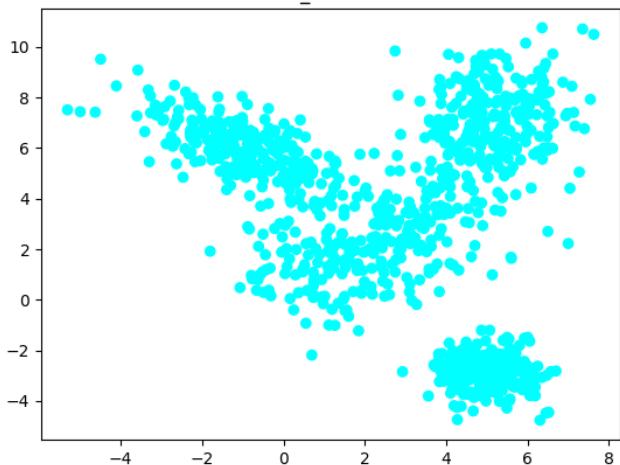


datos_1.csv- UD=1.5

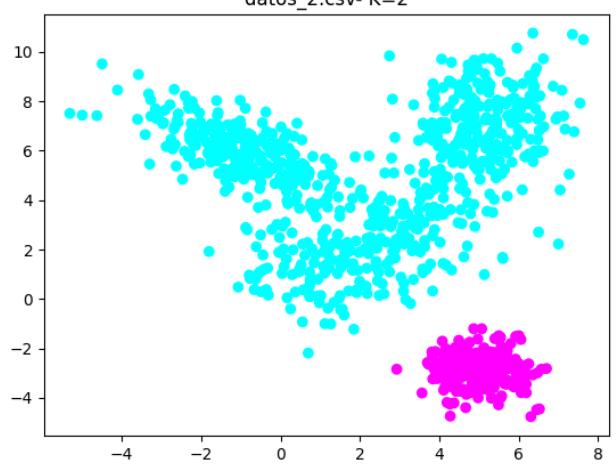


Datos_2
Por Valores K

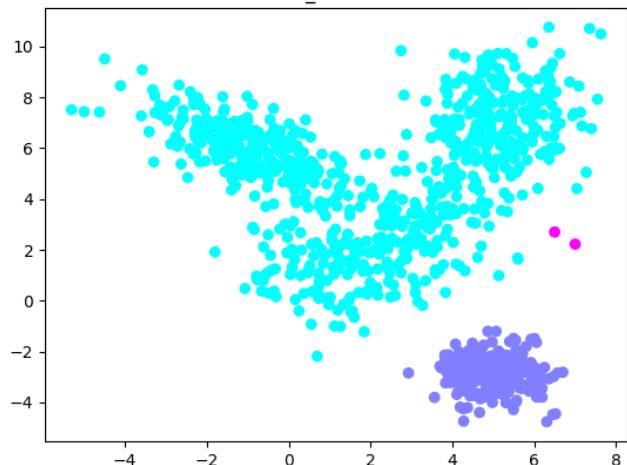
datos_2.csv- K=1



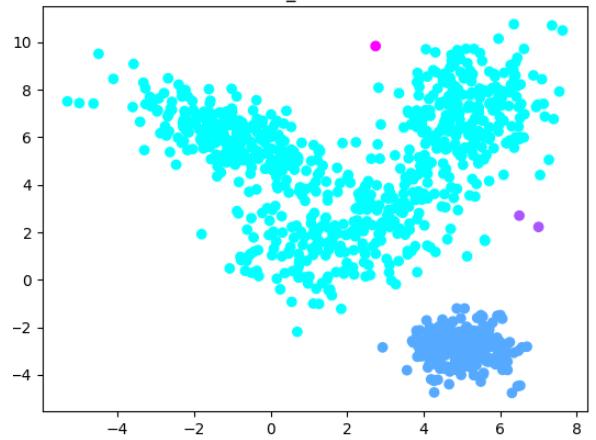
datos_2.csv- K=2



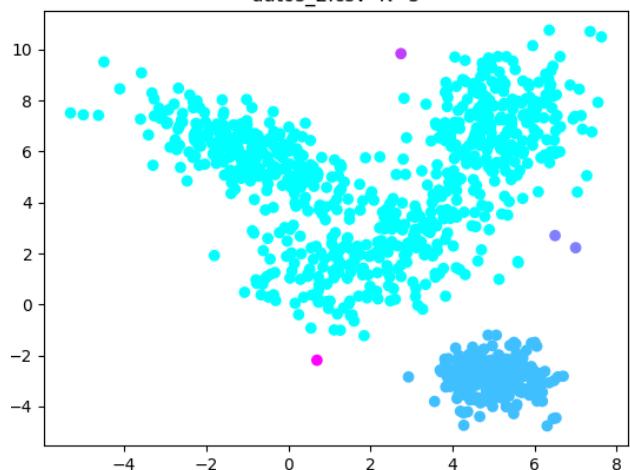
datos_2.csv- K=3



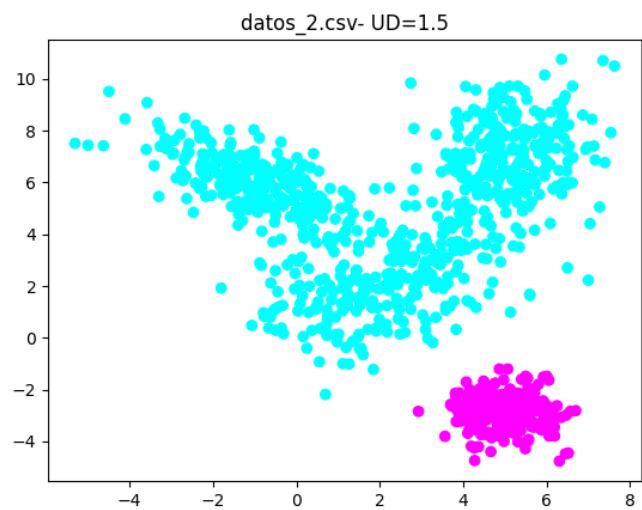
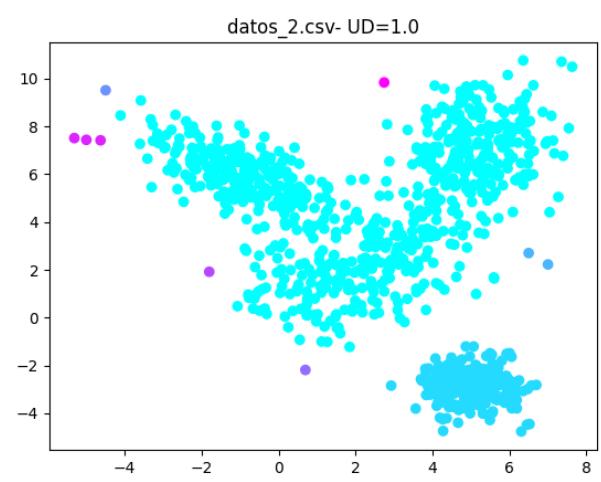
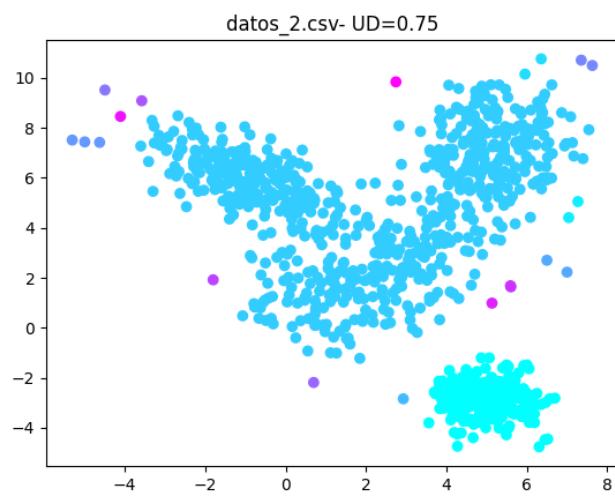
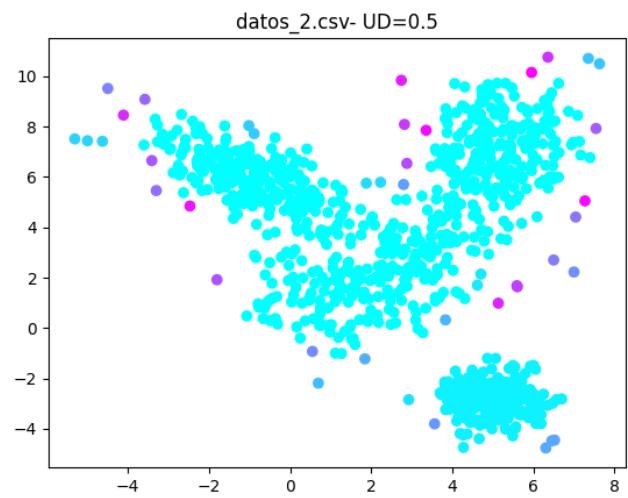
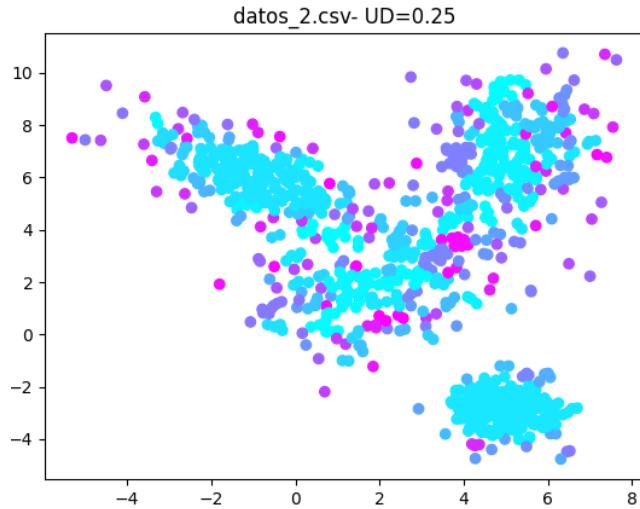
datos_2.csv- K=4



datos_2.csv- K=5



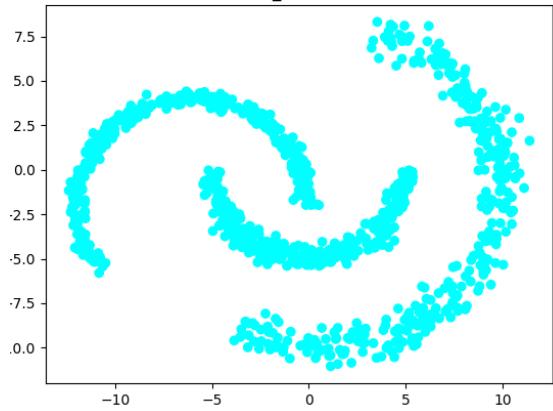
Por Distancia:



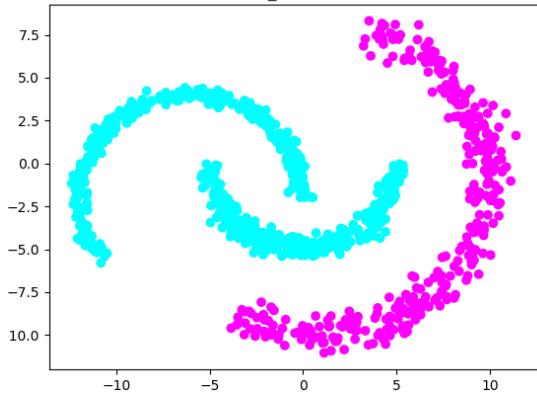
Datos_3

Por valores K

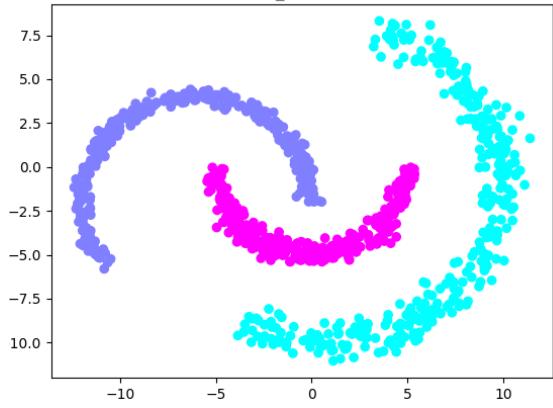
datos_3.csv- K=1



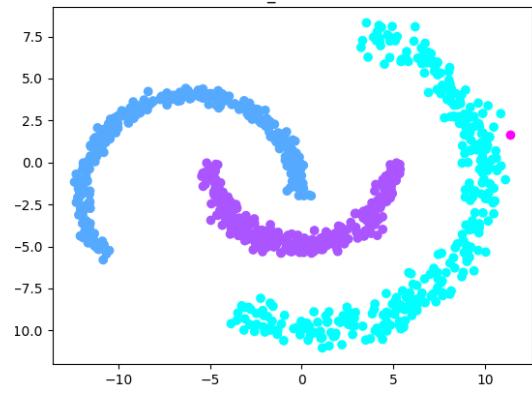
datos_3.csv- K=2



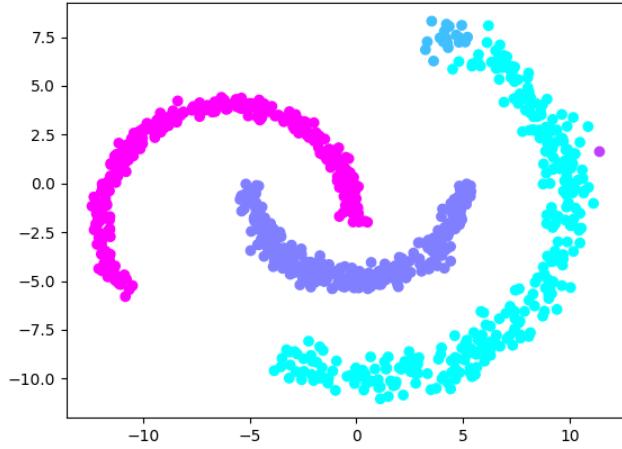
datos_3.csv- K=3



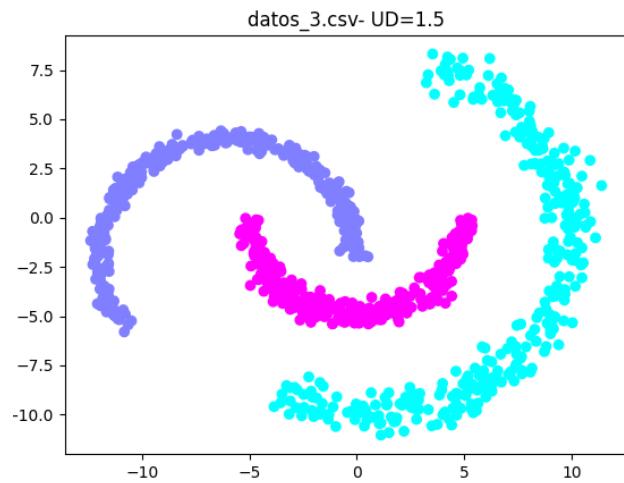
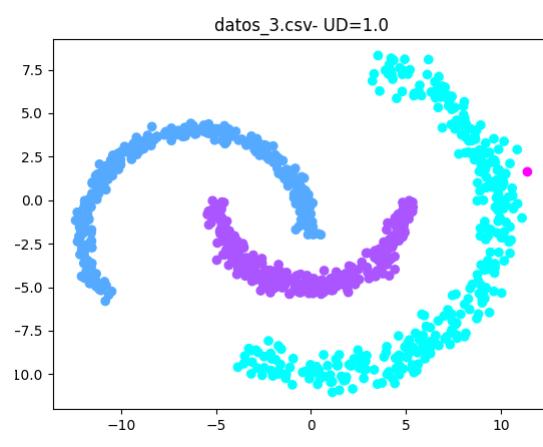
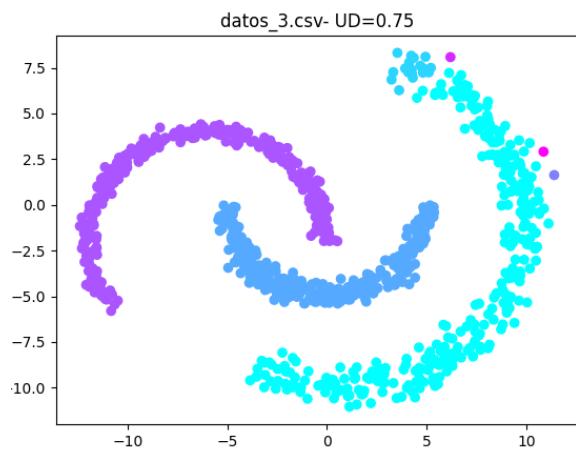
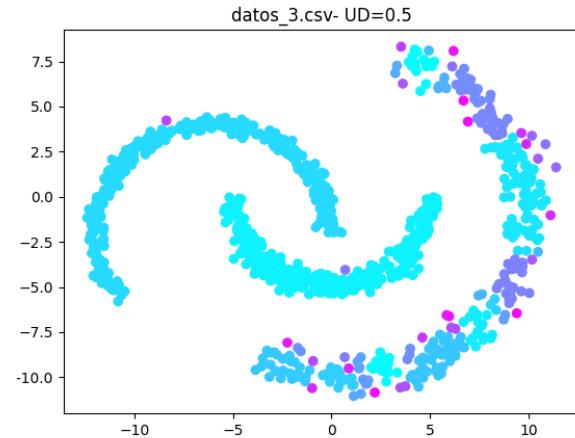
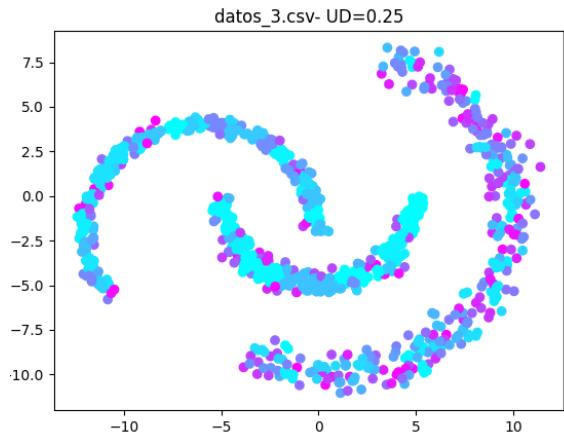
datos_3.csv- K=4



datos_3.csv- K=5



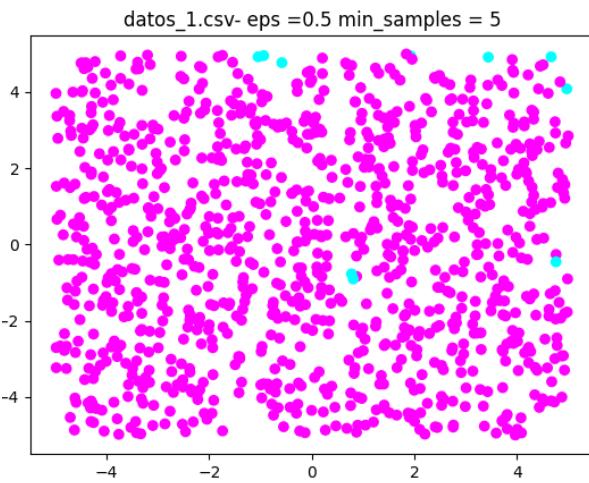
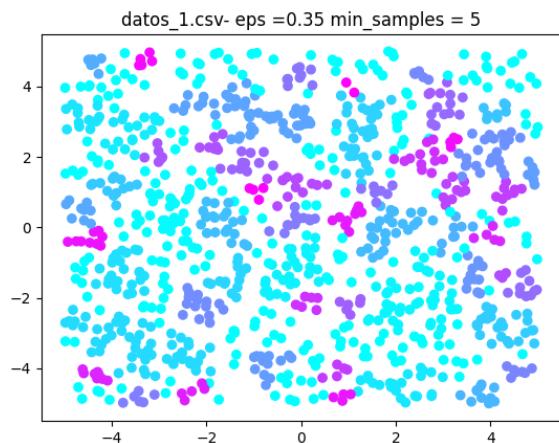
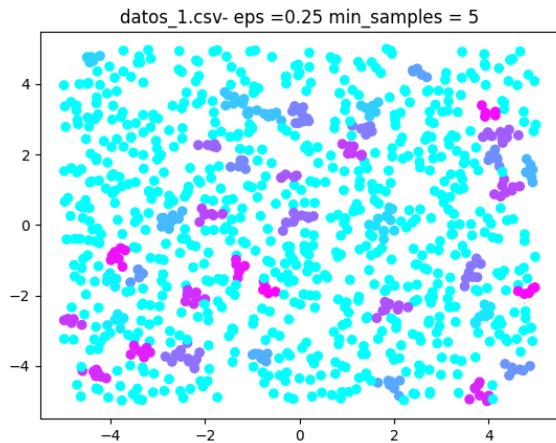
Por distancia



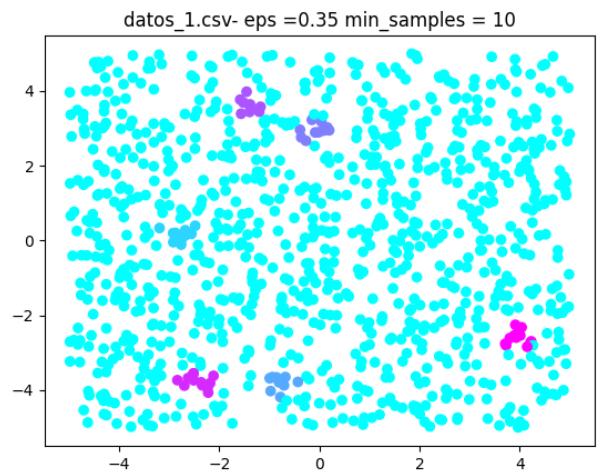
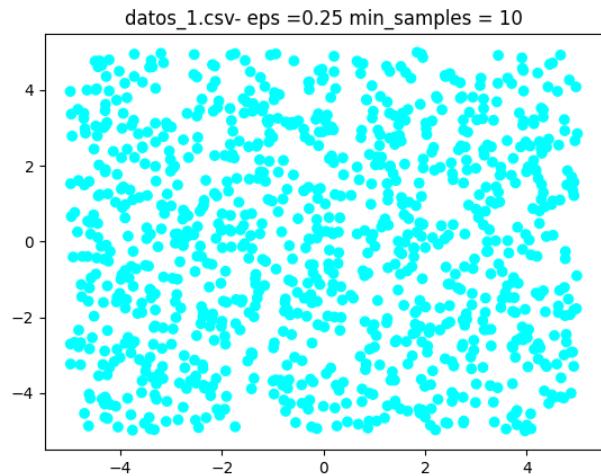
DBScan

Datos_1

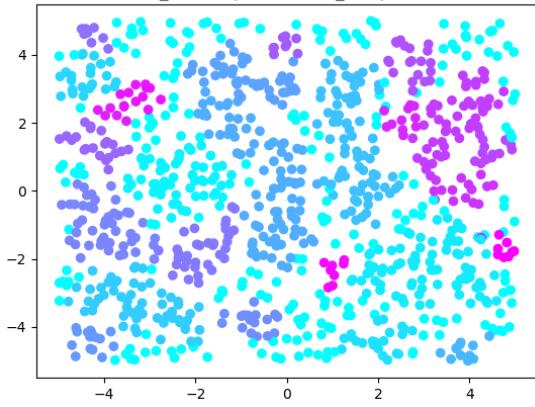
Min_samples = 5



Min_samples = 10

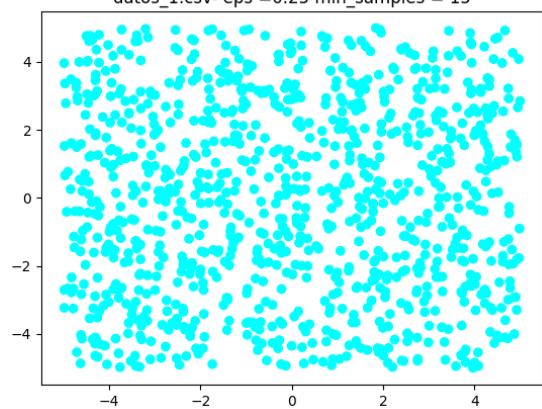


datos_1.csv- eps =0.5 min_samples = 10

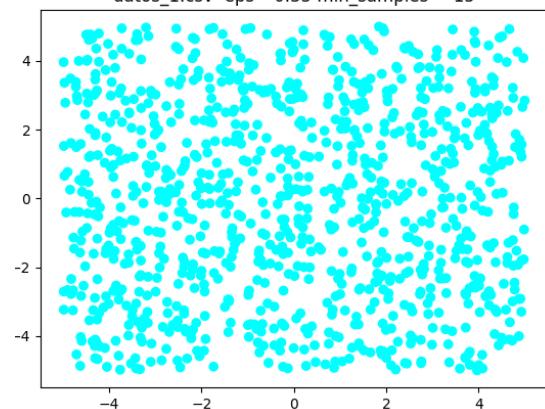


Min_samples = 15

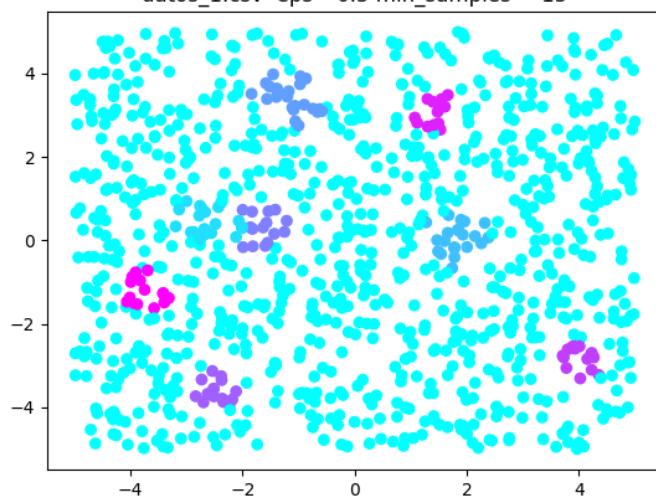
datos_1.csv- eps =0.25 min_samples = 15



datos_1.csv- eps =0.35 min_samples = 15

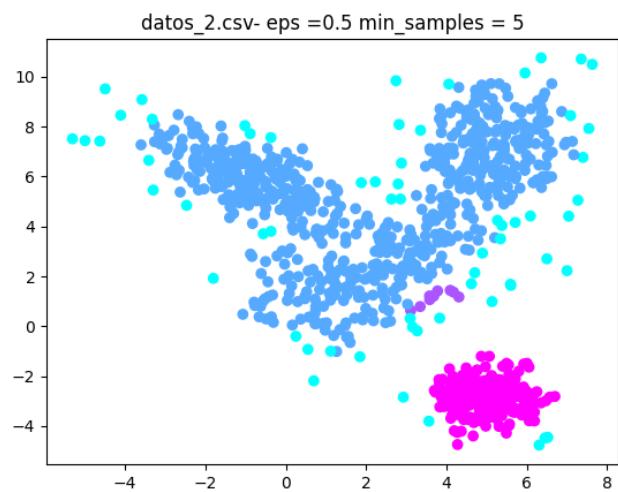
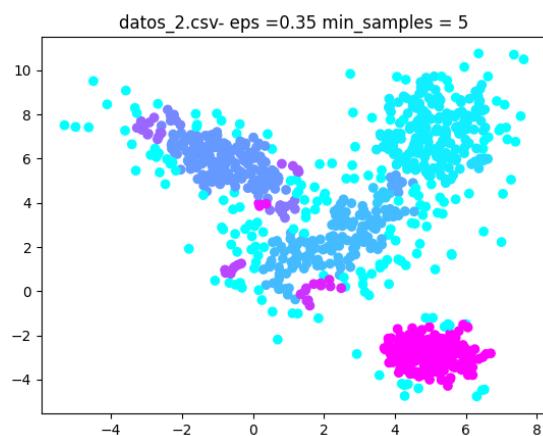
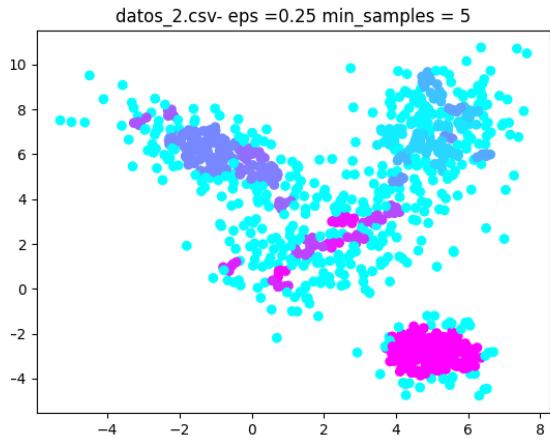


datos_1.csv- eps =0.5 min_samples = 15

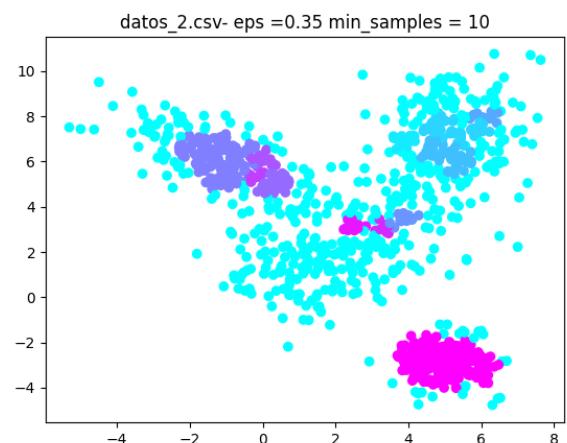
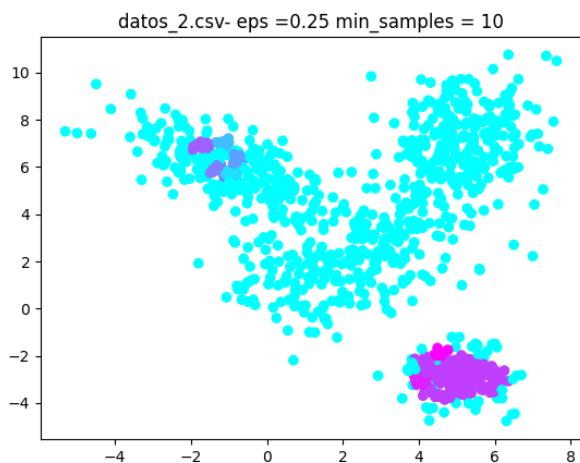


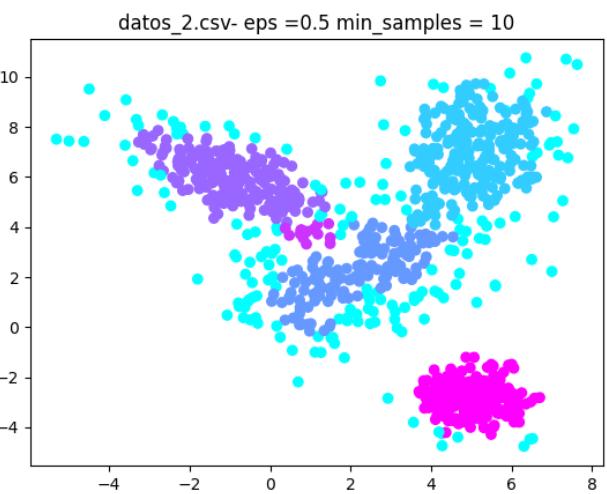
Datos_2

Min_samples = 5

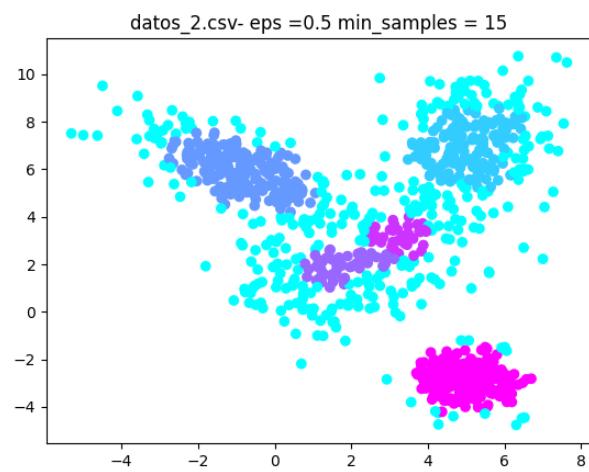
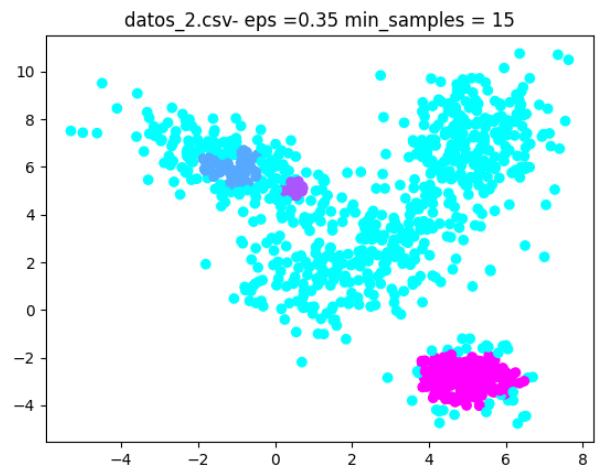
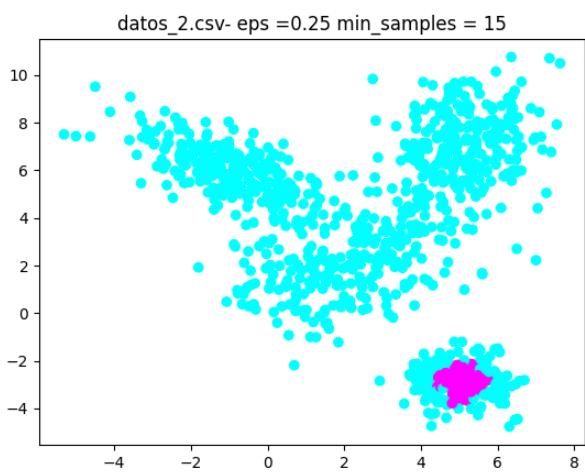


Min_samples = 10



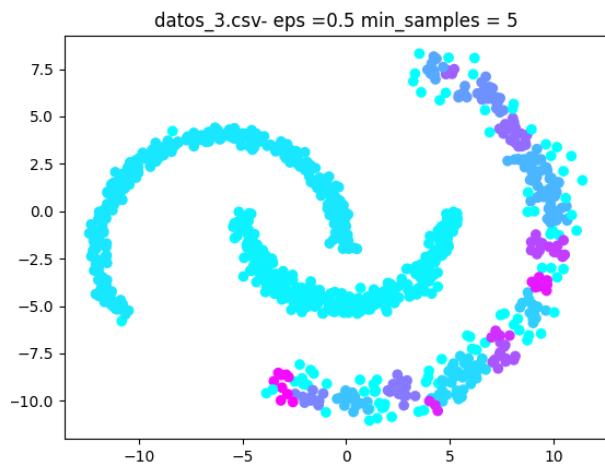
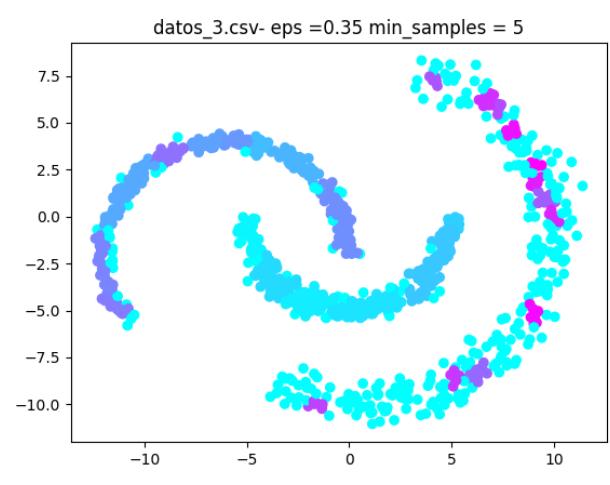
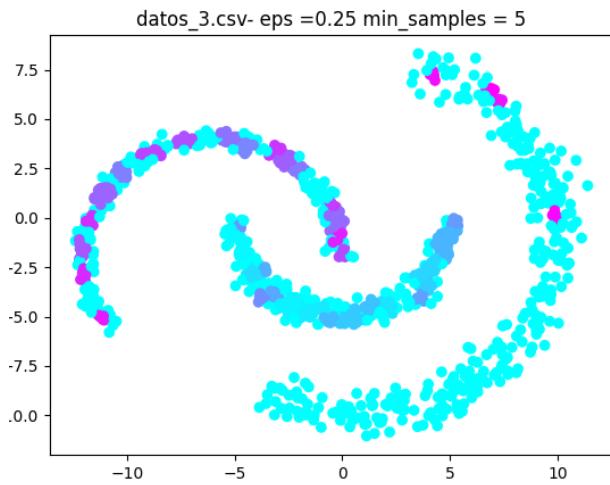


Min_samples = 15

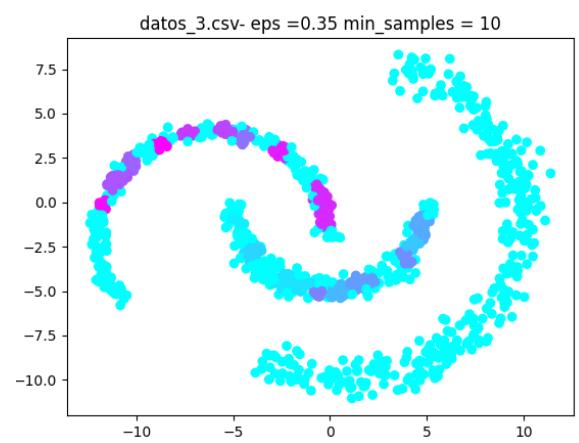
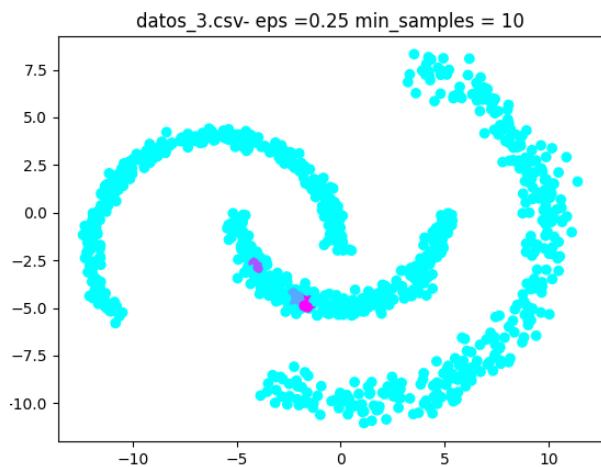


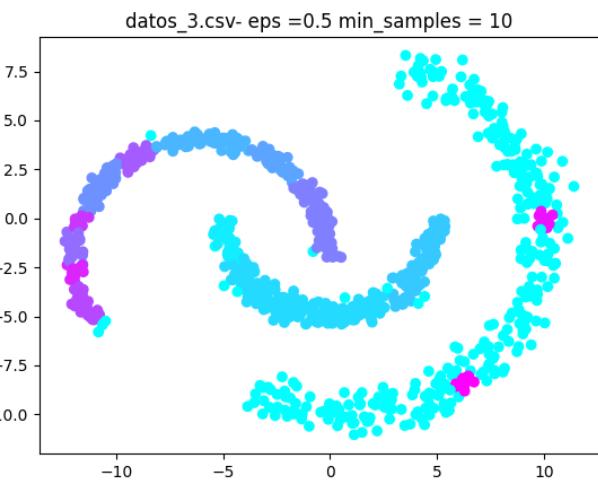
Datos_3

Min_samples = 5

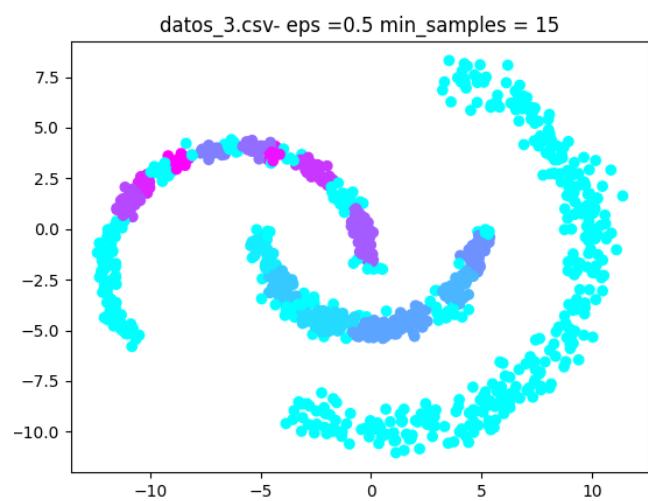
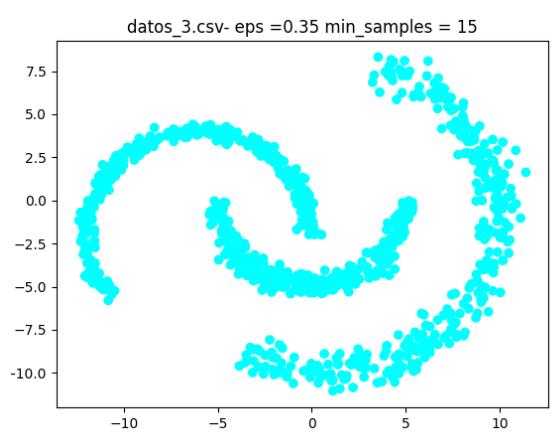
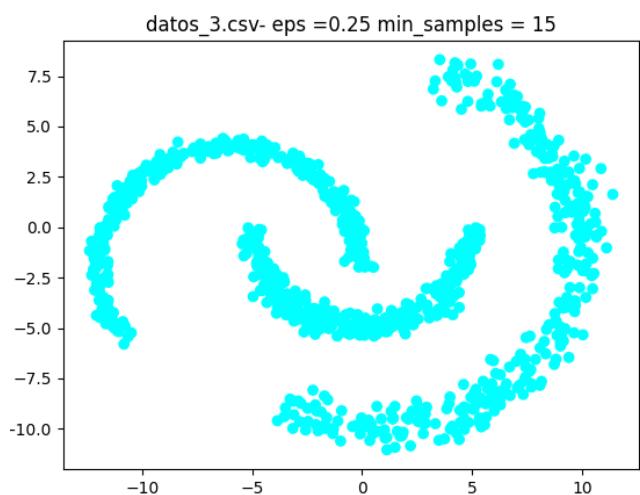


Min_samples = 10





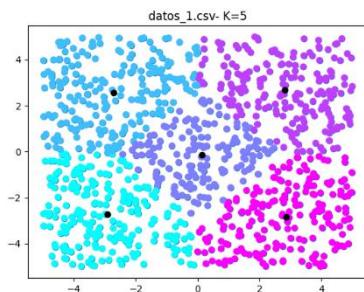
Min_samples = 15



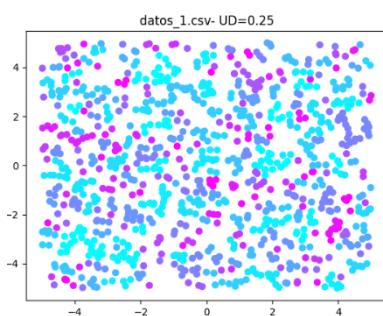
Mejores resultados según los datasets

Datos_1

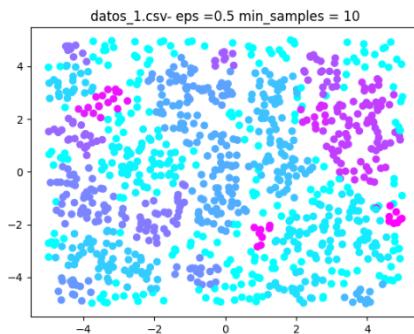
- K-Means



- Clustering jerárquico



- DB Scan

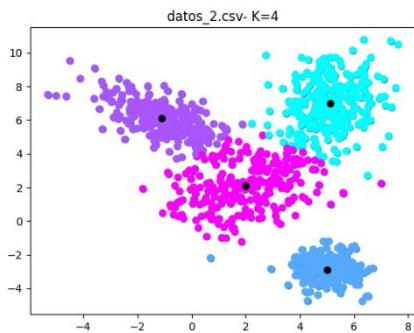


Viendo los resultados en general (no solo más relevantes), se puede notar que cuando tenemos un dataset como este (una gran cantidad de puntos que llenan una región, no tienden a una forma), es mejor el K-Means ya que este busca por el centro de la media y va realizando segmentos uniformes y continuos. Mientras que, los demás forman prácticamente una sola clasificación o realizan una clasificación con segmentos de puntos muy pequeños.

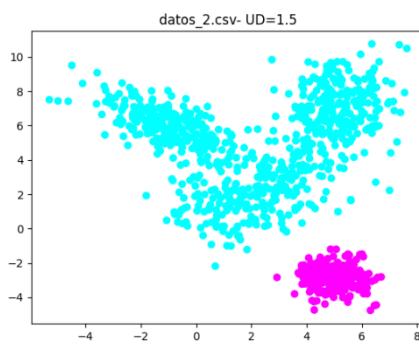
Debido a la cercanía de los puntos, pensaría que fue generado por una sola clase.

Datos_2

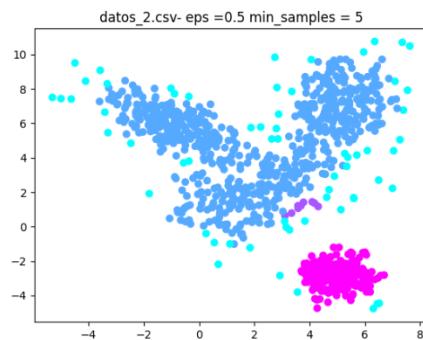
- K-Means



- Clustering jerárquico



- DB Scan

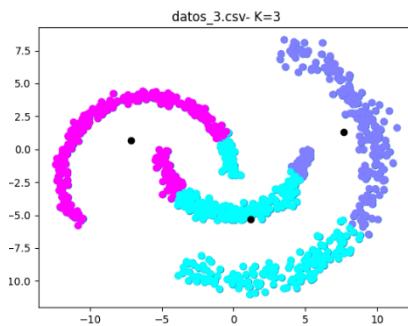


A diferencia del dataset anterior, los grupos en este son más evidente visualmente. Según los gráficos resultantes el que mejor lo clasificó fue Clustering Jerárquico, ya que logró clasificar el conjunto de puntos según la forma de la aglomeración de los mismos. Aunque, me gustaría resaltar un poco el gráfico de DB Scan de este dataset, ya que los puntos de la categoría azul celeste se podrían ver como "ruido", que sería muy provechoso eliminar (cosa que no resalta la clasificación resultante del Clustering Jerárquico).

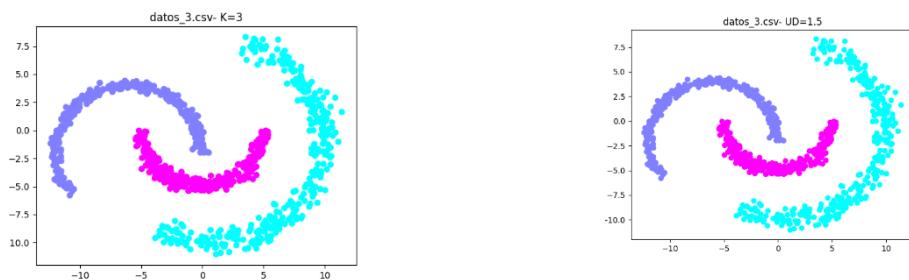
En primera instancia, considero que fue generado por dos clases, aunque, debido a la pequeña división en el centro de la porción superior, podría pensar que fueron tres clases.

Datos_3

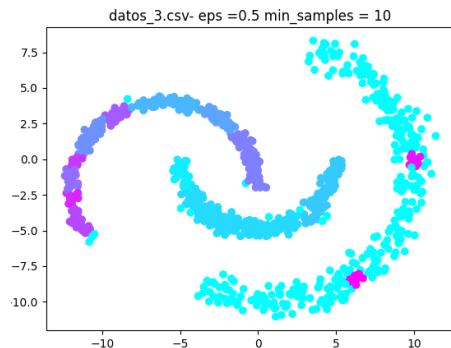
- K-Means



- Clustering jerárquico



- DB Scan



Respecto a este dataset, sin duda alguna el Clustering Jerárquico fue el mejor clasificado, como vemos en la gráfica hay tres grupos evidentes de puntos, este clustering logró identificarlos perfectamente tanto por el umbral de distancia como valores K (este es eficaz ya que tenemos 3 grupos y k=3 busca realizar sí o sí tres clasificaciones).

Debido a su forma, se podría decir que fue construido a partir de tres clases.