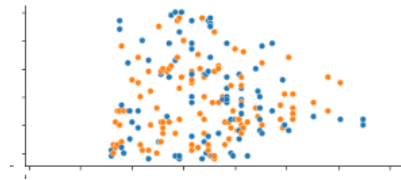
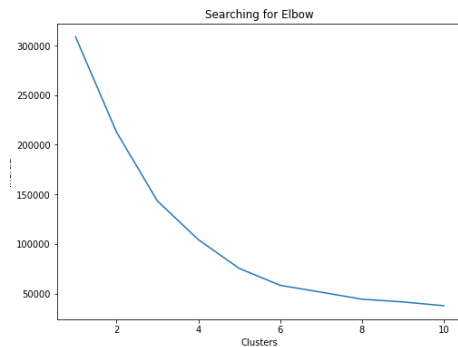
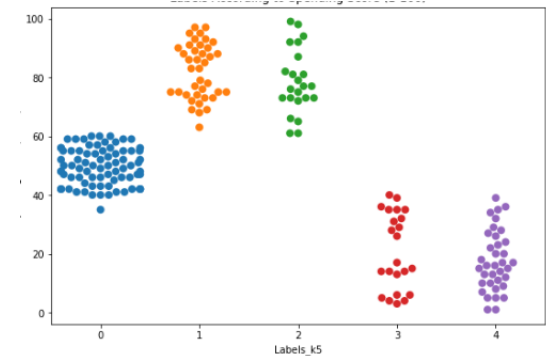


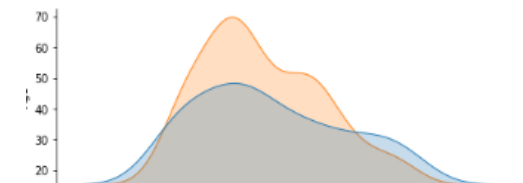
Estado de los conocimientos de los estudiantes sobre el tema de las máquinas eléctricas de corriente continua.



Maestrante Andrea Menco Tovar



Universidad Tecnológica de Bolívar
2022



Base de datos

La data esta compuesta por 6 columnas y 403 filas.

	STG	SCG	STR	LPR	PEG	UNS
0	0.00	0.10	0.50	0.26	0.05	Very Low
1	0.05	0.05	0.55	0.60	0.14	Low
2	0.08	0.18	0.63	0.60	0.85	High
3	0.20	0.20	0.68	0.67	0.85	High
4	0.22	0.22	0.90	0.30	0.90	High

STG (El grado de tiempo de estudio para los materiales del objeto objetivo)

SCG (El grado de número de repetición del usuario para los materiales del objeto objetivo)

STR (El grado de tiempo de estudio del usuario para objetos relacionados con el objeto objetivo)

LPR (El desempeño del usuario en el examen de objetos relacionados con el objeto objetivo)

PEG (El desempeño del usuario en el examen de objetos objetivo)

UNS (El nivel de conocimiento del usuario Muy bajo: 50; Bajo: 129; Medio: 122; Alto 130)

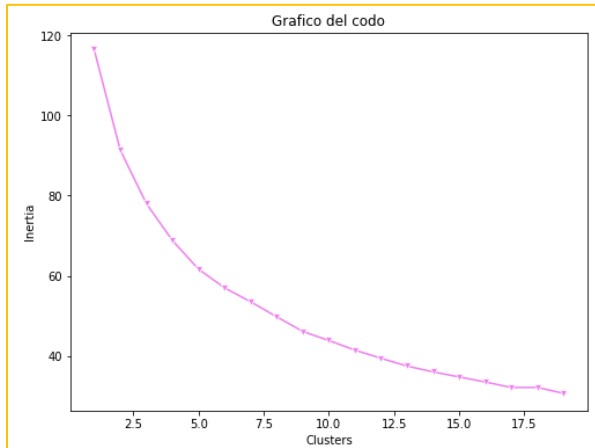
Estadística descriptiva de los datos

	STG	SCG	STR	LPR	PEG	UNS
count	403.000000	403.000000	403.000000	403.000000	403.000000	403.000000
mean	0.353141	0.355940	0.457655	0.431342	0.456360	1.684864
std	0.212018	0.215531	0.246684	0.257545	0.266775	0.986195
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.200000	0.200000	0.265000	0.250000	0.250000	1.000000
50%	0.300000	0.300000	0.440000	0.330000	0.400000	2.000000
75%	0.480000	0.510000	0.680000	0.650000	0.660000	3.000000
max	0.990000	0.900000	0.950000	0.990000	0.990000	3.000000

En esta data se logra evidenciar el cambio realizado a la variable UNS, debido a que los datos cualitativos de los cuales ella esta compuesta se acomodaron con respecto a una categoría numérica.

Algoritmos de agrupación no supervisada

1. Aplicando K-means:



Al observar los cambios en la evolución de la inercia y dado que no se observa tan claro el comportamiento de la misma, se puede afirmar que el número de clústers óptimos estará dentro de un rango de 4 a 6.

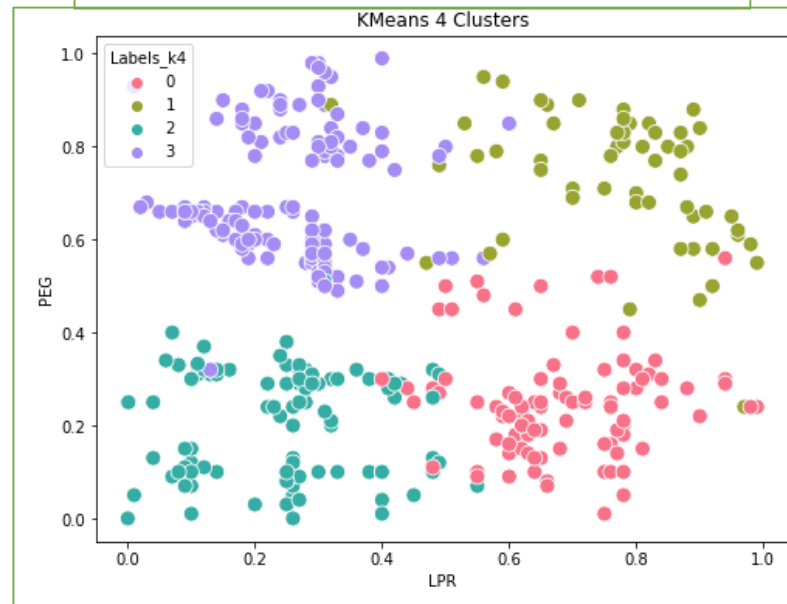
Los 4 clústers parecen ser el mejor agrupamiento de los datos (Hasta el momento), teniendo en cuenta que se esta observando un análisis bivariado de LPR y PEG.

Clúster 0: Alto LPR y PEG Bajo

Clúster 1: Bajo LPR y PEG Alto

Clúster 2: Bajo LPR y PEG Bajo

Clúster 3: Alto LPR y PEG Alto

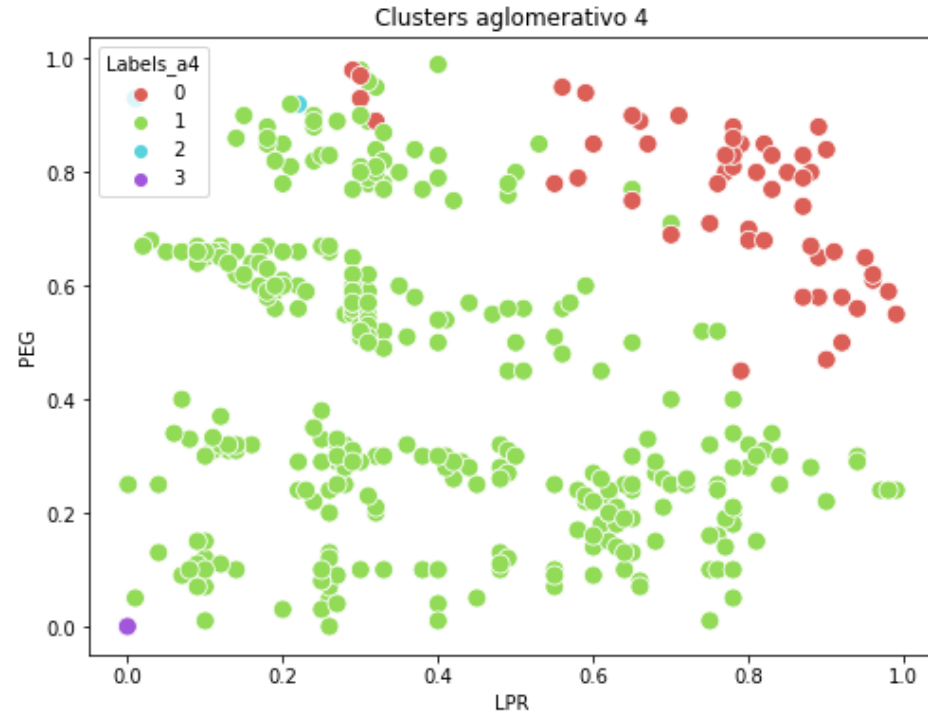


El error de detección 1 (Muy bajo) es = 14.000000000000002
El error de detección 2 (Bajo) es = 0.819672131147541
El error de detección 3 (Medio) es = 12.745098039215685
El error de detección 4 (Alto) es = 5.426356589147287

el error de detección promedio para un K-Mean de 4 cluster es= 8.25

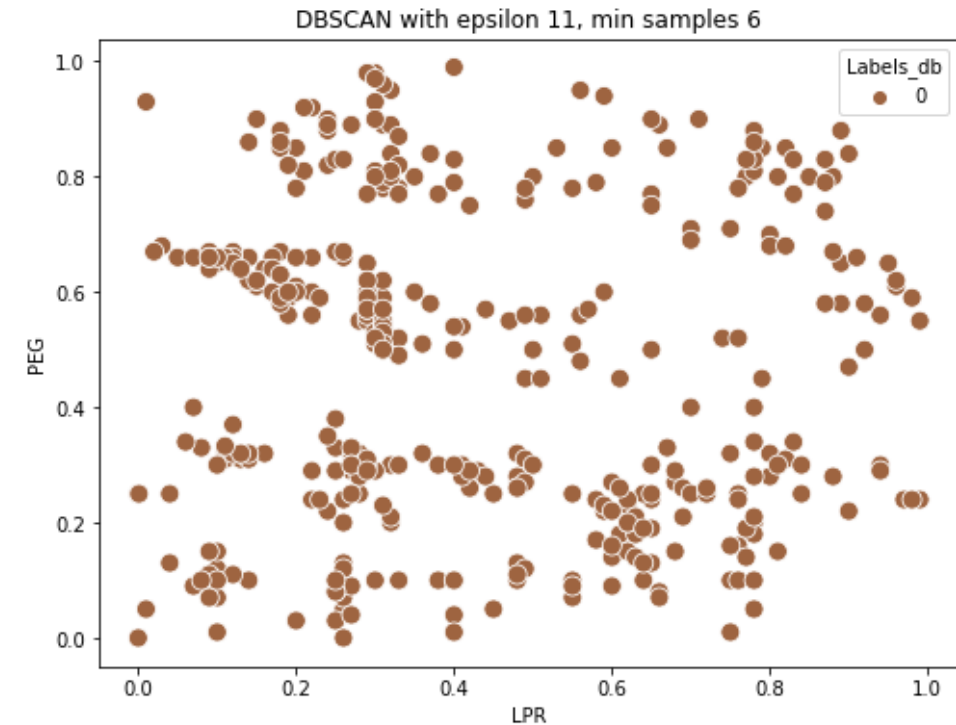
Debido a lo anterior, se cuenta con suficiente evidencia estadística para afirmar que el mejor algoritmo de agrupación no supervisada para la data en estudio es el K-Mean, a juzgar por el grafico y los datos de error de detección.

2. Clustering Jerárquico (Aglomerativo)



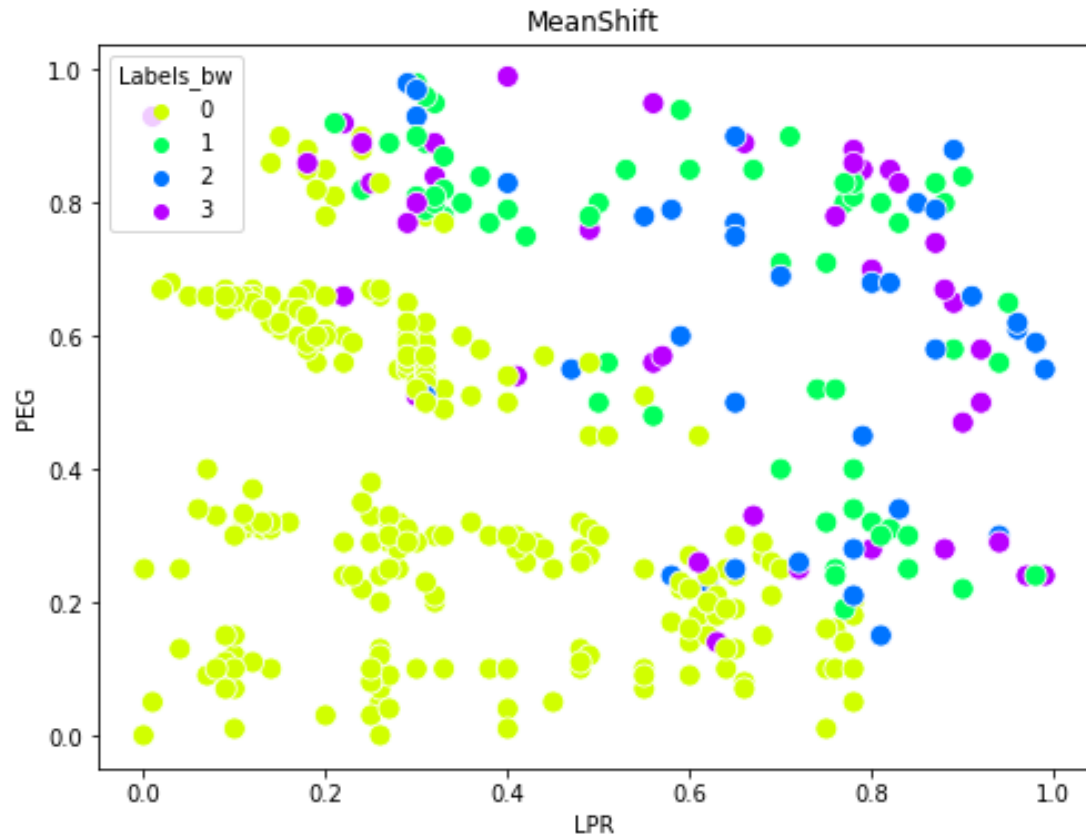
Definitivamente se puede evidenciar que este algoritmo no es el adecuado para los datos en estudio, ya que solo toma dos grupos, así la elección de este clustering no tendría sentido, debido a los grupos que ya están establecidos en el artículo de estudio.

3. Clustering basado en la densidad (DBSCAN)



Se observa que DBSCAN no es adecuado para el análisis de la data, debido a que realiza un solo agrupamiento de los datos.

4. Algoritmo Mean Shift



Podemos evidenciar que los datos no convergen hacia un mismo punto de máxima densidad, lo que provoca que los clústers no encajen en un mismo color, además existe un inconveniente y es que la selección del quantil puede ser no trivial. Así se puede afirmar que este algoritmo es poco aceptado.