

Clustering Jerárquico

Recomendación de Compras por Agrupación

El **Clustering Jerárquico** es un algoritmo de aprendizaje automático que agrupa datos en una jerarquía de *clústers*. Se utiliza para encontrar patrones y conexiones en los datos. Con este modelo, fue que se realizó la práctica en cuestión, donde era requerido agrupar diferentes clientes de una base de datos que contenían información sobre su comportamiento de compras anteriores. Todo esto con el objetivo de **recomendar productos relacionados** que puedan ser de interés dependiendo del grupo al que pertenezca cada cliente.

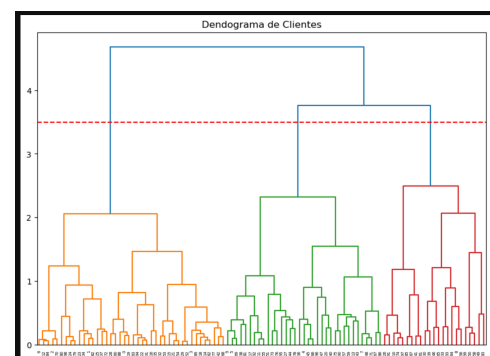
```
df = pd.read_excel("Amazon.xlsx").rename(columns = {"Unnamed: 0" : "Cliente"})
df
```

	Cliente	Velocidad Entrega	Precio	Durabilidad	Imagen Producto	Valor Educativo	Servicio Retorno	Tamaño Paquete	Calidad Producto	Número Estrellas
0	Adam	205	3	345	235	24	23	26	21	17
1	Anna	9	15	315	33	25	4	42	215	28
2	Bernard	17	26	285	3	43	27	41	26	33
3	Edward	135	5	355	295	18	23	39	195	17
4	Emilia	3	45	48	39	34	46	225	34	43
...
95	Teofan	3	8	32	25	7	21	42	17	1
96	Teofil	305	25	46	24	33	28	355	26	45
97	Teofila	1	14	26	25	24	27	42	185	23
98	Teon	155	11	335	34	26	29	42	215	27
99	Teresa	125	9	45	25	22	3	3	22	18

100 rows x 10 columns

Registro de compras de 100 clientes diferentes.

Para poder realizar la agrupación con el algoritmo antes mencionado, es requerido **normalizar** los datos, para que estos se encuentren en una misma escala y poder obtener mejores resultados. Habiendo realizado la normalización se procedió a generar el *dendograma*, el cual nos muestra tres grupos, como número óptimo.



Dendograma

Ahora, teniendo los datos clasificados por medio de PCA, y uniendo las etiquetas de cada grupo con el DataFrame original, se puede observar la clasificación que tiene cada Cliente.

```
df_etiquetado = pd.concat([nombre_clientes, data, grupos_df], axis = 1, join = "inner")
df_etiquetado = df_etiquetado.set_index("Cliente")
df_etiquetado
```

Cliente	Velocidad Entrega	Precio	Durabilidad	Imagen Producto	Valor Educativo	Servicio Retorno	Tamaño Paquete	Calidad Producto	Numero Estrellas	Grupo
Adam	205	3	345	235	24	23	26	21	17	1
Anna	9	15	315	33	25	4	42	215	28	1
Bernard	17	26	285	3	43	27	41	26	33	1
Edward	135	5	355	295	18	23	39	195	17	1
Emilia	3	45	48	39	34	46	225	34	43	2
...
Teofan	3	8	32	25	7	21	42	17	1	2
Teofil	305	25	46	24	33	28	355	26	45	2
Teofila	1	14	26	25	24	27	42	185	23	0
Teon	155	11	335	34	26	29	42	215	27	1
Teresa	125	9	45	25	22	3	3	22	18	0

100 rows x 10 columns

Clientes etiquetados con su **Grupo** Respectivo

Recomendación de productos

Con el análisis inicial, y el proceso de Machine Learning terminado, podemos enfocarnos en los tres clientes de interés, “Salomé”, “Stephanía” y “Lydia.”

Como se observa en la siguiente imagen, **Salomé** pertenece al **Grupo 2**, **Stephanía** al **Grupo 1** junto con **Lydia**.

Cliente	Velocidad Entrega	Precio	Durabilidad	Imagen Producto	Valor Educativo	Servicio Retorno	Tamaño Paquete	Calidad Producto	Numero Estrellas	Grupo
Salome	17	23	275	41	4	44	315	28	32	2
Stephanía	215	125	465	315	34	4	37	305	45	1
Lydia	19	4	435	145	16	21	28	185	24	1

Clientes de Interés

Con una visualización breve de los grupos en cuestión, podemos concluir que **Salomé** podría estar interesado en productos que suelen ser grandes “*Tamaño de Paquete*”, con un “*Servicio de Retorno*” mayor a **20 y menor que 50**, y un precio

que pudiera llegar hasta los **180 - 200**. Como claro ejemplo, podemos tener a **Fabián, o Emilia** quienes tienen comportamientos similares.

	Velocidad Entrega	Precio	Durabilidad	Imagen Producto	Valor Educativo	Servicio Retorno	Tamaño Paquete	Calidad Producto	Numero Estrellas	Grupo
Cliente										
Emilia	3	45	48	39	34	46	225	34	43	2
Fabian	95	165	395	24	26	19	485	22	3	2
Frank	65	21	31	255	28	22	345	215	29	2
Gabriel	2	175	325	3	37	32	435	27	39	2
Henry	195	11	455	23	3	25	415	25	32	2

Grupo 2

Por ultimo, **Stephanía y Lydia** parecen ser candidatas a obtener productos cuyo “Precio” sea **menor a 100**, “Tamaño de Paquete” **pequeño** y una “Durabilidad” excepcional con un indice **mayor a 200 hasta 480-500**. Como ejemplos podemos observar a **Adam o Edward** quienes presentan intereses similares a la hora de comprar.

	Velocidad Entrega	Precio	Durabilidad	Imagen Producto	Valor Educativo	Servicio Retorno	Tamaño Paquete	Calidad Producto	Numero Estrellas	Grupo
Cliente										
Adam	205	3	345	235	24	23	26	21	17	1
Anna	9	15	315	33	25	4	42	215	28	1
Bernard	17	26	285	3	43	27	41	26	33	1
Edward	135	5	355	295	18	23	39	195	17	1
Philip	23	12	475	33	35	45	38	29	31	1

Grupo 1