

# SEGMENTACIÓN DEL MERCADO COLOMBIANO DE CERVEZA – Grupo 10

Santiago Pulido, Liliana Briceño, Cristhian Barbosa, Diego Peñaloza

Enlace Repositorio en Github: <https://github.com/Cbarbo82/ANS-G10.git>

## 1. RESUMEN

Se realiza un perfilamiento del mercado de cerveza nacional a partir de las características de las marcas que lo componen, estableciendo perfiles cerveceros que agrupen diferentes marcas para identificar posibles oportunidades de negocio.

Para ello, se cuenta con las características químicas, sensoriales y de producción nacional de 52 marcas de cerveza masivas y artesanales del mercado para los años entre 2020 y 2023. Cada conjunto de características se compone de distintas variables que aportan información valiosa sobre las propiedades de cada marca.

Con la información disponible se busca que, empleando técnicas de aprendizaje no supervisado, como el análisis de clústeres, permita segmentar el mercado de cervezas en Colombia a partir de resultados sensoriales, fisicoquímicos y cantidad de ventas. Esto nos permite identificar oportunidades de crecimiento en sectores no explorados, preferencias del consumidor, desarrollo de marcas nuevas o potenciar las ya existentes.

## 2. INTRODUCCIÓN

El mercado de la cerveza en Colombia está experimentando un notable crecimiento, con un consumo anual estimado en 30.2 millones de hectolitros, representando ingresos de aproximadamente 22 billones de pesos en 2023. A pesar de este amplio mercado, la dinamización es limitada debido a la estructura monopolística del mercado. Esto ha restringido la diversificación, limitando la innovación y la introducción de nuevos tipos de cerveza.

No obstante, dentro de las principales oportunidades de negocio se encuentra la inclusión de marcas y expansión de cervezas artesanales, un segmento que ha mostrado un crecimiento gradual en los últimos años [3]. Sin embargo, en comparación con otros países latinoamericanos como Argentina o Chile, donde la cerveza artesanal ha capturado una mayor participación de mercado, Colombia aún tiene un largo camino por recorrer.

Este vacío presenta una oportunidad única para cervecerías que deseen explorar y desarrollar el mercado, ofreciendo productos diferenciados que puedan captar la atención de consumidores que están comenzando a valorar la diversidad y la calidad en su consumo cervecero. Mediante la aplicación de técnicas de aprendizaje no supervisado se busca recomendar los perfiles de cerveza en los que invertir puede resultar en grandes ganancias.

## 3. MATERIALES Y MÉTODOS.

Para el estudio propuesto contamos con una base de datos con 52 marcas de cerveza y 15 variables proporcionadas por Cervecería Colombiana. Algunos de los campos ofrecen información de las propiedades específicas de cada marca de cerveza respecto a sus condiciones fisicoquímicas, tales como el extracto original entendido como la cantidad de azúcares que quedan después de la fermentación, la cantidad de alcohol, el pH como medida de acidez o alcalinidad, la intensidad de color y la cantidad de compuestos amargos.

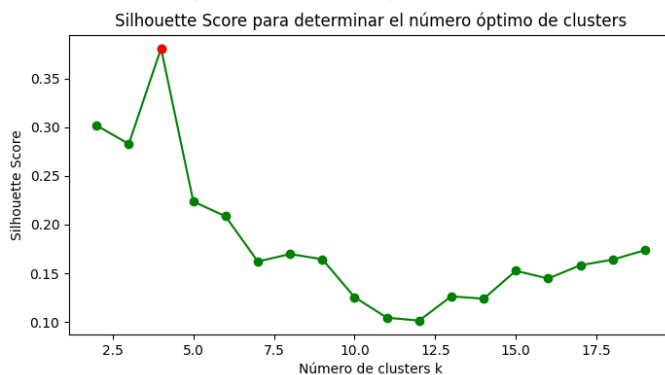
Asimismo, se cuenta con los parámetros sensoriales que ayudan a describir el perfil de cada bebida, tales como, la turbiedad (clara u oscura), el cuerpo (pesada o ligera), el dulzor y la astringencia (sensación de

sequedad). Por último, para el análisis de clustering no se incluyen las ventas anuales de cada marca del 2022 y 2023, sin embargo, estas variables son usadas para calcular el incremento anual de ventas y tamaño de mercado para el análisis de estrategias de la organización. Para más información de las características y unidades de medida de cada atributo, consultar el diccionario de datos dispuesto en el anexo.

Utilizando diferentes algoritmos se realizan análisis de clustering con el objetivo de determinar el número óptimo de clústeres y, a partir de ello, analizar las agrupaciones obtenidas mediante estadísticas descriptivas. El procedimiento realizado para cada metodología es el siguiente:

### 3.1. KMEANS

Primero, se busca identificar el número óptimo de clústeres a partir del **silhouette score**, evaluando esta métrica para valores de  $k$  entre 2 y 19. Cabe aclarar que el **silhouette score** mide la calidad de la separación entre clústeres, donde valores más altos indican una mejor diferenciación entre grupos. Al analizar el comportamiento de esta métrica para los diferentes valores de  $k$ , se determina que el valor máximo corresponde a un  $k$  óptimo de 4.

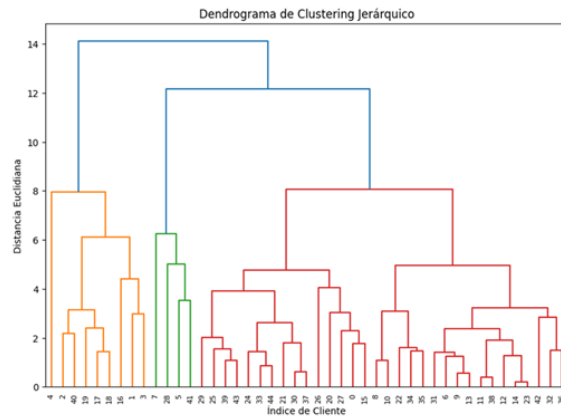


Con el  $k$  óptimo determinado, se entrena el modelo K-Means utilizando los datos previamente procesados, asignando a cada registro el clúster correspondiente. Finalmente, se calculan las estadísticas descriptivas para cada clúster, con el objetivo de analizar las características de las variables evaluadas en cada grupo.

Clúster	#	Extracto Aparente		Alcohol v/v		pH		Color		Amargo				Turbiedad	Cuerpo	Dulzor	Astringencia
		mean	std	mean	std	mean	std	mean	std	mean	std	min	max	moda	moda	moda	moda
0	29	1.77	0.84	4.23	0.55	4.37	0.13	7.28	2.37	13.61	3.72	6.11	18.6	1	2	2	2
1	10	2.66	0.59	4.70	0.39	4.45	0.12	30.30	29.29	26.86	8.60	18.00	44.4	3	3	3	3
2	4	4.55	1.59	2.38	2.00	3.77	0.24	6.64	0.89	10.80	8.35	1.50	20.3	1	1	4	3
3	2	3.92	0.69	5.47	1.05	4.47	0.20	110.65	47.16	17.50	7.78	12.00	23.0	1	3	1	4

### 3.2 CLUSTERING JERARQUICO

El clustering jerárquico aglomerativo permite agrupar recursivamente las marcas de las cervezas con aquellas que le son más similares. En primer lugar, se realiza el dendograma que permite analizar gráficamente el proceso de agrupamiento. Se calcula utilizando la distancia euclidiana por pares de centroides de los clústeres de los datos, cuyo algoritmo agrupará recursivamente los puntos más cercanos. Asimismo, se define el enlace de Ward como el método para combinar los clústeres, el cual busca que al unirse los clústeres haya una reducción de la varianza mejorando la cohesión interna.



Una vez analizado el dendrograma, se establece una distancia de 10 para la generación de 3 clústeres, en concordancia con los resultados obtenidos con el **silhouette score** máximo de 0.35 con un k de 3. Al analizar la composición de cada clúster, se observa:

Cluster	n	Extracto aparente		Alcohol v/v		pH		Color		Amargo				Turbiedad	Cuerpo	Dulzor	Astringencia
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Min	Max	Mode	Mode	Mode	Mode
0	4	1,86	0,87	4,28	0,5	4,38	0,14	10,21	13,1	13,9	3,8	6,11	20	1	2	2	2
1	17	2,92	0,8	4,85	0,65	4,45	0,14	45,42	47,6	28,1	8,22	19,6	44,4	3	3	3	3
2	8	4,55	1,59	2,38	2	3,77	0,24	6,64	0,89	10,8	8,35	1,5	20,3	1	1	4	3

### 3.3 DBSCAN

Para realizar el análisis usando DBSCAN se construye una función que evalúa escenarios con diferentes cantidades de vecinos, encontrando épsilon en el punto de codo para las distancias encontradas con diferentes valores de K. Para cada relación optima de K y épsilon, se entrena el algoritmo y calcula tanto la medida de **silhouette score** para encontrar la mejor definición de clústeres y Davis–Bouldin Score para determinar la menor similitud de agrupamiento. Finalmente, se obtienen los resultados de ambas medidas para cada relación de vecinos y épsilon.

Se destaca, que para diferentes valores de N si tomamos el mejor **silhouette score** (8 vecinos) el modelo solo genera un clúster de 36 datos y 9 outliers, mientras que si se genera un clúster basado en la combinación optima de D-B Score (3 vecinos) se llega a un resultado de 2 clústeres con 29 y 3 marcas agrupadas con 13 outliers.

Cluster	Extracto aparente		Alcohol v/v		pH		Color		Amargo				Turbiedad	Cuerpo	Dulzor	Astringencia
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Min	Max	Mode	Mode	Mode	Mode
-1	3,47	1,27	4,13	1,64	4,2	0,37	37,33	44,75	21,92	11,97	1,5	44,4	1	2	2	3
0	2,07	0,47	4,3	0,55	4,041	0,09	8,93	5,75	14,99	3,52	7,6	20,5	1	2	2	2
1	0,35	0,13	3,88	0,37	4,26	0,14	5,92	0,7	7,3	1,37	6,11	8,8	1	2	2	2

## 4. RESULTADOS Y DISCUSIÓN.

Al realizar la comparación entre los clústeres obtenidos a través de los tres métodos, se observa:

Modelo	Clúster	Descripción
K-Means	0	Cervezas tipo lager que presentan cuerpo, dulzor y astringencia relativamente bajos, con un color dorado y un nivel de amargor moderadamente bajo. Estas cervezas se caracterizan por su baja turbidez, típica de cervezas industriales de alto consumo.

	1	Cervezas artesanales o industriales con cuerpo, dulzor y astringencia moderadamente altos. Este grupo presenta un amplio rango de amargor y un contenido de alcohol relativamente elevado, característico del segmento. Dado que el mercado artesanal ofrece una variedad de estilos, el rango de color es bastante amplio. Asimismo, predomina una alta turbidez, ya que muchas cervezas artesanales no pasan por procesos de filtración.
	2	Cervezas predominantemente dulces con un contenido alcohólico bajo.
	3	Cervezas oscuras, con cuerpo y astringencia moderadamente altos.
<b>Jerárquico</b>	0	Cervezas tipo lager y ale, con un rango de color bastante amplio desde cervezas doradas hasta rojizas. Cuentan con turbiedad clara, cuerpo, dulzor y astringencia moderada. Alcohol moderado con un alto rango de amargor.
	1	Cervezas artesanales o industriales con un cuerpo, astringencia y dulzor alto. Color moderadamente oscuro y rango de amargor relativamente elevado.
	2	Cervezas color dorado, con un bajo nivel de alcohol, dulzor elevado y astringencia moderadamente alta. Segmento enfocado en cervezas de alto consumo.
<b>DBSCAN</b>	-1	Cervezas oscuras y amargas comparadas con las demás. Son cervezas que hacen parte de sector masivo industrial. Por lo tanto, el algoritmo no las agrupa dentro de los clústeres evidenciados si no que las reconoce como outliers independientes.
	0	Corresponde a cervezas con un grado de alcohol medio, relativamente claras y de un amargor medio bajo. Podría describir el sector de consumo masivo más común en el mercado nacional.
	1	Agrupar Cervezas de bajo nivel de alcohol, color más claro y un amargor bajo. Agrupa cervezas tipo light que son ligeras y fáciles para consumo.

El modelo seleccionado es K-means, dado que segmenta las diferentes marcas de cerveza en una agrupación lógica con diferencia significativa en sus propiedades fisicoquímicas y sensoriales, logrando diferenciar segmentos claros de mercado tales como: Cerveza de consumo masivo (lager), artesanal y premium, cervezas dulces y con bajo alcohol, por último, cervezas oscuras y astringentes.

En comparación, el método jerárquico al segmentar por 3 clústeres se incurre en la generalización de las cervezas artesanales en un mismo grupo sin diferenciar parámetros significativamente diferentes tales como cervezas oscuras y astringentes. Por lo tanto, al no tener una segmentación detallada no se considera una agrupación útil para el negocio. Así mismo, la métrica de silhouette score de K-means es más alta que el método Jerárquico.

Por último, para el método de DBSCAN se evidencia una agrupación en 2 clústeres que se enfoca principalmente en los segmentos mayoritarios referentes a las cervezas de consumo masivo, dejando aún lado un segmento potencial como lo son las cervezas artesanales o premium. Para este algoritmo no existen agrupaciones más detalladas en las cervezas de consumo masivo, limitando el análisis de estrategias para posibles expansiones de mercado.

Gráfico de Incremento para Amargo

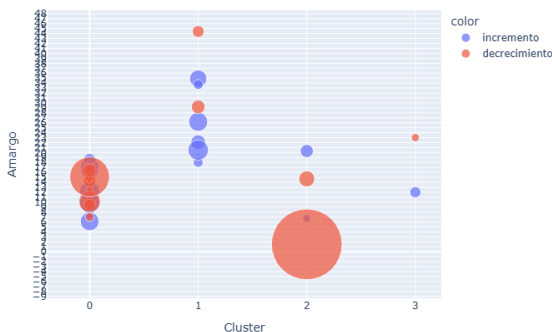
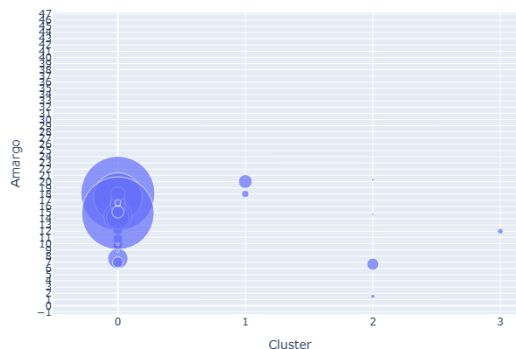


Gráfico de Ventas para Amargo



Con base en los clústeres obtenidos a partir del algoritmo de K-means, se realiza una visualización de burbujas para cada marca por segmento, donde el incremento 2022-2023 y ventas totales 2023, se representan a partir del tamaño de la burbuja. Para el gráfico de incremento el color rojo representa un decrecimiento y azul un incremento respecto al año anterior.

De los resultados podemos analizar:

- A pesar de que el clúster 0 es el segmento con más ventas (97.38%) se puede evidenciar que no es el segmento con más crecimiento. Siendo el grupo con mayor competencia se recomienda invertir solo si se tiene un capital financiero bastante amplio con el fin de impulsar y promocionar nuevas marcas.
- Dado que el clúster 2 es el segmento con el decrecimiento más alto, afectado principalmente por una marca, se recomienda realizar nuevamente el análisis con los datos de ventas 2024 para evaluar el comportamiento de esta marca en específico y concluir si el segmento es viable.
- El clúster 3 es el segmento con menor competencia en el mercado enfocado en cervezas oscuras, con alto cuerpo y astringente, se recomienda que el retorno de inversión en este grupo es de largo plazo dado que su participación en el mercado es la más baja (0.2%) e incremento en ventas (6.52%) relativamente lento.
- El segmento con mayor potencial de crecimiento es el clúster 1 con una participación de mercado del 1.42% y un incremento en ventas para último año del 25.89%. Este el grupo más atractivo para posibles inversiones de corto y mediano plazo, con un capital financiero moderado dada la baja competencia. Por lo anterior, se recomienda el desarrollo de una nueva marca de cerveza que tenga las siguientes características: Bajo amargo, alcohol medio, dorada, cuerpo moderado, dulzor relativamente alto, y leve astringente. Cabe aclarar que la campaña de marketing debe ser enfocada no en el consumo masivo si no en un mercado premium o artesanal.

## 5. CONCLUSIÓN.

El análisis de clustering permitió segmentar el mercado de cervezas en Colombia en función de sus características fisicoquímicas, sensoriales y de ventas. A través de diferentes enfoques tales como K-Means, clustering jerárquico y DBSCAN, se identificaron varios grupos que reflejan las diferentes tipologías de cervezas en el mercado. Estos resultados muestran oportunidades de negocio, destacando perfiles de cerveza artesanal, premium y negra.

De acuerdo con los resultados obtenidos y la métrica de **silhouette score**, el mejor algoritmo para analizar el mercado es K-Means. Este agrupa las cervezas en 4 clústeres que representan perfiles distintos de cerveza: cerveza de consumo rubia masivo, cerveza artesanal o premium, cerveza light con dulzor alto y cervezas negras astringentes. En conjunto con los valores de incremento de ventas y participación en el mercado de las diferentes marcas, se llega las siguientes conclusiones:

- Oportunidad de negocio en el mercado de la cerveza artesanal o premium, con el desarrollo de una nueva marca.
- Para invertir en cervezas de consumo masivo es necesario de un capital bastante alto dado que es un segmento muy competitivo y estancado.
- Existe una oportunidad de negocio en las cervezas negras a largo plazo, dado su crecimiento lento y baja participación en el mercado.

## 6. BIBLIOGRAFÍA

- [1] Mawatari, M., Nagashima, Y., Aoki, T., Hirota, T., & Yamada, M. (1991). Sensory Characterization of Beer Flavor Using Cluster Analysis. *Journal of the American Society of Brewing Chemists*, 49(2), 59–64. <https://doi-org.ezproxy.uniandes.edu.co/10.1094/ASBCJ-49-0059>
- [2] El Tiempo. (2024). *Competir en un mercado que históricamente ha sido un monopolio no ha sido una tarea fácil: Central Cervecera*. <https://www.eltiempo.com/economia/empresas/competir-en-un-mercado-que-historicamente-ha-sido-un-monopolio-no-ha-sido-una-tarea-facil-central-cervecera-3370142>
- [3] La República. (2024). *La industria cervecera artesanal tiene alrededor de 0,5% del mercado total de licores*. <https://www.larepublica.co/consumo/la-industria-cervecera-artesanal-tiene-alrededor-de-0-5-del-mercado-total-de-licores-3444506>
- [4] Castellanos, J. y Sossa, C. (2022). Industria cervecera colombiana: un análisis desde su comercio internacional. *Expresiones, Revista Estudiantil de Investigación*, 9(17), 51-59.
- [5] Mariño, Cynthia, (2023). Análisis de clustering para la segmentación de mercado: un caso de estudio de una aplicación de una bebida alcohólica en las principales ciudades de Colombia. Universidad del Bosque.

## 7. ANEXOS





Variable	Unidad	Definición
<b>Parámetro Físico químicos</b>		
Extracto Original	°P	Concentración de azúcares en el mosto antes de la fermentación. Determina el potencial de alcohol que la cerveza puede alcanzar después de la fermentación.
Extracto Aparente	°P	Concentración de azúcares que quedan después de la fermentación.
Alcohol	% v/v	Cantidad de alcohol en la cerveza
pH		Medida de acidez o alcalinidad de la cerveza. Asegura la estabilidad del sabor, claridad y conservación de la cerveza.
Color	EBC	Se refiere a la intensidad de color de la cerveza. Esto varía dependiendo del tipo de cerveza.
Amargo	IBU	Cantidad de compuestos amargos, sobre todo ácidos alfa de lúpulo en la cerveza. El valor de IBU varía dependiendo del tipo de cerveza.
<b>Medición Sensorial</b>		
Turbiedad	1. Claro / cristalino. 2. Ligeramente turbio. 3. Moderadamente turbio. 4. Turbio.	Se refiere a la clara u opacidad de la cerveza. Influye en la percepción visual de la bebida
Cuerpo	1. Cuerpo bajo. 2. Cuerpo ligero medio. 3. Cuerpo moderado. 4. Cuerpo alto.	Sensación de textura en boca de la cerveza. Se pueden llegar a percibir cervezas más "pesadas" o más "ligeras".
Dulzor	1. Nada dulce. 2. Ligeramente dulce. 3. Moderadamente dulce. 4. Dulce alto.	Percepción de los azúcares residuales en la cerveza.
Astringencia	1. Ninguna. 2. Leve. 3. Moderado. 4. Alto.	Sensación de sequedad o aspereza en la boca.
Ventas 2021	Hl (hectolitros)	Cantidad de hectolitro vendidos por año.
Ventas 2022		
Ventas 2023		