

HENRY

BOOTCAMP DATA ENGINEER

PROYECTO INTEGRADOR

Alumno:

Angela Sarasti Mera

Cedula de ciudadania:

1.015.474.787

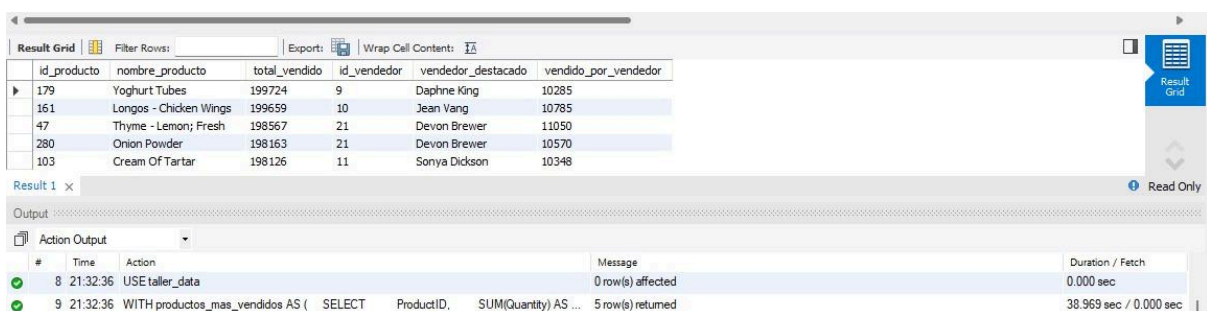
Profesional en Administración de empresas y  
profesional en mercadeo

COLOMBIA

## ÍNDICE:

- Primer avance proyecto integrador

A) ¿Cuáles fueron los 5 productos más vendidos (por cantidad total), y cuál fue el vendedor que más unidades vendió de cada uno? Una vez obtenga los resultados, en el análisis responde: ¿Hay algún vendedor que aparece más de una vez como el que más vendió un producto? ¿Algunos de estos vendedores representan más del 10% de la ventas de este producto?



The screenshot shows a database query result grid with the following data:

id_producto	nombre_producto	total_vendido	id_vendedor	vendedor_destacado	vendido_por_vendedor
179	Yoghurt Tubes	199724	9	Daphne King	10285
161	Longos - Chicken Wings	199659	10	Jean Vang	10785
47	Thyme - Lemon; Fresh	198567	21	Devon Brewer	11050
280	Onion Powder	198163	21	Devon Brewer	10570
103	Cream Of Tartar	198126	11	Sonya Dickson	10348

Below the grid, the 'Output' section shows the execution details of the query:

#	Time	Action	Message	Duration / Fetch
8	21:32:36	USE taller_data	0 row(s) affected	0.000 sec
9	21:32:36	WITH productos_mas_vendidos AS ( SELECT ProductID, SUM(Quantity) AS ...	5 row(s) returned	38.969 sec / 0.000 sec

Tenemos que los 5 productos más vendidos son: yoghurt tubes, longs, chickenwings, thyme-lemon; fresh onion powder y cream of tartar, esto se logró desarrollando un script SQL estructurado en pasos lógicos con la ayuda de las subconsultas temporales como WITH, calculando el total de unidades vendidas con un sumQuantity por cada producto, tambien con la utilización de group by y order by poniendo limit 5 como una condición importante y trayendo información de otra carpeta con un join employés y mostrar su nombre completo como vendedor\_destacado.

En los resultados obtenidos, se observa que Devon Brewer aparece más de una vez como el vendedor con el mayor volumen de ventas en dos productos del top 5.

Y para saber si alguno de los vendedores representan más del 10% de la ventas de los productos se añadió:  $\text{ROUND}((\text{tv.cantidad\_vendida} * 100.0) / \text{pt.cantidad\_total}, 2)$  AS porcentaje\_vendedor

Result Grid						
Filter Rows:						
Export:						
Wrap Cell Contents:						
id_producto	nombre_producto	total_vendido	id_vendedor	vendedor_destacado	vendido_por_vendedor	porcentaje_vendedor
103	Cream Of Tartar	200002	9	Daphne King	10551	5.28
179	Yoghurt Tubes	199724	9	Daphne King	10285	5.15
161	Longos - Chicken Wings	199659	10	Jean Vang	10785	5.40
47	Thyme - Lemon; Fresh	198567	21	Devon Brewer	11050	5.56
280	Onion Powder	198163	21	Devon Brewer	10570	5.33

#	Time	Action	Message	Duration / Fetch
2	15:37:27	USE taller_data	0 row(s) affected	0.000 sec
3	15:37:27	WITH productos_top5 AS ( SELECT ProductID, SUM(Quantity) AS cantidad_t...	5 row(s) returned	15.297 sec / 0.000 sec

Con estos resultados podemos concluir que aunque varios vendedores son constantes en el top, ninguno de ellos representa más del 10% del total vendido del producto.

B) Entre los 5 productos más vendidos, ¿cuántos clientes únicos compraron cada uno y qué proporción representa sobre el total de clientes? Analiza si ese porcentaje sugiere que el producto fue ampliamente adoptado entre los clientes o si, por el contrario, fue comprado por un grupo reducido que generó un volumen alto de ventas. Compara los porcentajes entre productos e identifica si alguno de ellos depende más de un segmento específico de clientes.

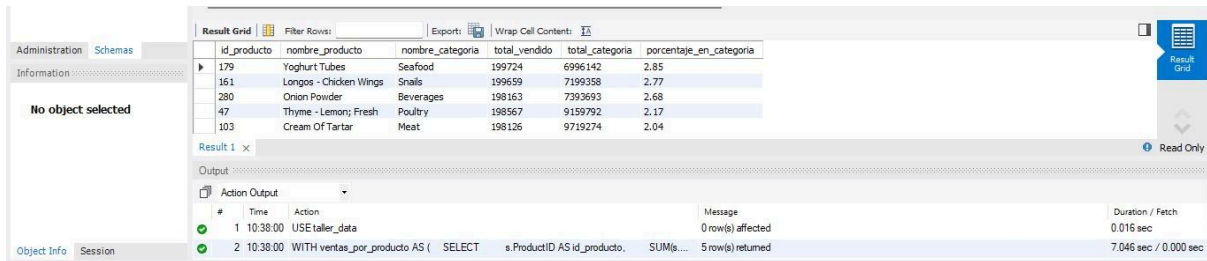
Result Grid						
Filter Rows:						
Export:						
Wrap Cell Contents:						
id_producto	nombre_producto	total_vendido	id_vendedor	vendedor_destacado	vendido_por_vendedor	porcentaje_vendedor
103	Cream Of Tartar	200002	9	Daphne King	10551	5.28
179	Yoghurt Tubes	199724	9	Daphne King	10285	5.15
161	Longos - Chicken Wings	199659	10	Jean Vang	10785	5.40
47	Thyme - Lemon; Fresh	198567	21	Devon Brewer	11050	5.56
280	Onion Powder	198163	21	Devon Brewer	10570	5.33

#	Time	Action	Message	Duration / Fetch
2	15:37:27	USE taller_data	0 row(s) affected	0.000 sec
3	15:37:27	WITH productos_top5 AS ( SELECT ProductID, SUM(Quantity) AS cantidad_t...	5 row(s) returned	15.297 sec / 0.000 sec

En la imagen podemos encontrar que cada uno de los 5 productos más vendidos fueron comprados por aproximadamente el 14% del total de clientes únicos registrados en la base de datos. Por lo cual se concluye que todos los productos analizados presentan una alta adopción entre la base de clientes. No hay productos con una adopción anormalmente baja que sugiera que fueron comprados solo por un pequeño grupo. Y que las ventas de estos productos están bien distribuidas, lo que da a entender que las estrategias de venta y demanda están equilibradas en cuanto a cobertura de mercado.

- C) ¿A qué categorías pertenecen los 5 productos más vendidos y qué proporción representan dentro del total de unidades vendidas de su categoría? Utiliza funciones de ventana para comparar la relevancia de cada producto dentro de su propia categoría.



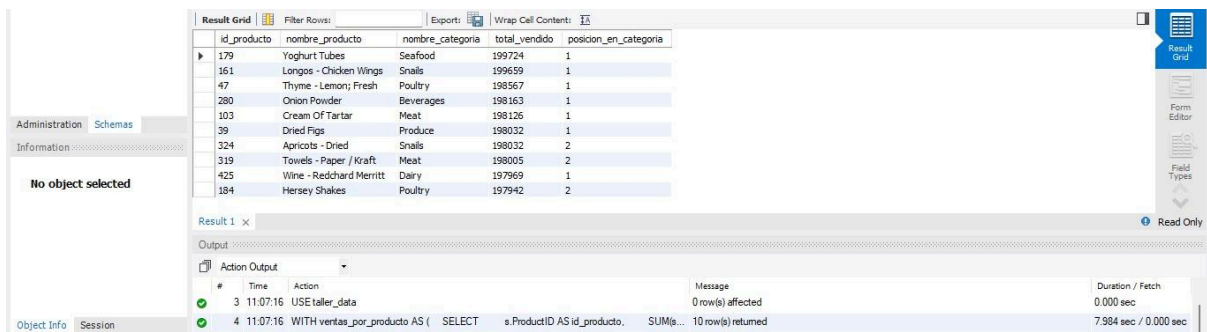
id_producto	nombre_producto	nombre_categoria	total_vendido	total_categoria	porcentaje_en_categoria
179	Yoghurt Tubes	Seafood	199724	6996142	2.85
161	Longos - Chicken Wings	Snails	199659	7199358	2.77
280	Onion Powder	Beverages	198163	7393693	2.68
47	Thyme - Lemon; Fresh	Poultry	198567	9159792	2.17
103	Cream Of Tartar	Meat	198126	9719274	2.04

#	Time	Action	Message	Duration / Fetch
1	10:38:00	USE taller_data	0 row(s) affected	0.016 sec
2	10:38:00	WITH ventas_por_producto AS ( SELECT s.ProductID AS id_producto, SUM(s...	5 row(s) returned	7.046 sec / 0.000 sec

Los productos más vendidos no dominan completamente sus categorías, lo que sugiere un mercado diversificado. Cada uno representa entre el 2% y el 2.85% del total de unidades vendidas dentro de su categoría. A pesar de tener altos volúmenes de ventas, representan solo un pequeño porcentaje del total de su categoría, lo cual puede ser indicio de que comparten espacio con muchos otros productos populares.

- D) ¿Cuáles son los 10 productos con mayor cantidad de unidades vendidas en todo el catálogo y cuál es su posición dentro de su propia categoría? Utiliza funciones de ventana para identificar el ranking de cada producto en su categoría. Luego, analiza si estos productos son también los líderes dentro de sus categorías o si compiten estrechamente con otros productos de alto rendimiento. ¿Qué observas sobre la concentración de ventas dentro de algunas categorías?



id_producto	nombre_producto	nombre_categoria	total_vendido	posicion_en_categoria
179	Yoghurt Tubes	Seafood	199724	1
161	Longos - Chicken Wings	Snails	199659	1
47	Thyme - Lemon; Fresh	Poultry	198567	1
280	Onion Powder	Beverages	198163	1
103	Cream Of Tartar	Meat	198126	1
39	Dried Figs	Produce	198032	1
324	Apricots - Dried	Snails	198032	2
319	Towels - Paper / Kraft	Meat	198005	2
425	Wine - Redchard Merritt	Dairy	197969	1
184	Hersey Shakes	Poultry	197942	2

#	Time	Action	Message	Duration / Fetch
3	11:07:16	USE taller_data	0 row(s) affected	0.000 sec
4	11:07:16	WITH ventas_por_producto AS ( SELECT s.ProductID AS id_producto, SUM(s...	10 row(s) returned	7.984 sec / 0.000 sec

Para este punto se utilizó la función de ventana `RANK() OVER (PARTITION BY CategoryID ORDER BY total_vendido DESC)` para identificar la posición del producto dentro de su categoría, si el producto está en la columna `posicion_en_categoria = 1` con el número 1 significa que es líder en su categoría, pero si por el contrario obtiene un 2 o 3 significa que comparten la categoría con productos que les compiten de cerca o que son fuertes dentro de esa categoría.

- Segunda avance proyecto integrador

A) Crea un trigger que registre en una tabla de monitoreo cada vez que un producto supere las 200.000 unidades vendidas acumuladas.

El trigger debe activarse después de insertar una nueva venta y registrar en la tabla el ID del producto, su nombre, la nueva cantidad total de unidades vendidas, y la fecha en que se superó el umbral.

id_producto	nombre_producto	total_vendido	fecha_superado
-------------	-----------------	---------------	----------------

B) Registra una venta correspondiente al vendedor con ID 9, al cliente con ID 84, del producto con ID 103, por una cantidad de 1.876 unidades y un valor de 1200 unidades.

Consulta la tabla de monitoreo, toma captura de los resultados y realiza un análisis breve de lo ocurrido.

The screenshot displays a database interface with a 'Result Grid' showing the following data:

id_producto	nombre_producto	total_vendido	fecha_superado
103	Cream Of Tartar	200002	2025-06-10 22:39:15

Below the grid, the 'Output' section shows the trigger definition and the execution log. The log indicates that the trigger 'superan\_umbral1' was executed successfully, inserting a row into the 'productos\_superan\_umbral' table.

El trigger se activó después de la inserción de la venta para el producto 103, lo que permitió realizar el cálculo total de unidades ya que el producto Cream of Tartar tenía un total acumulado de 198.126 unidades vendidas sumando las 1.876 nuevas ventas el total serian 200.002 unidades, esto hace que el trigger se active correctamente e inserte el producto en la tabla de monitoreo, registrando la fecha exacta en la que se superó el umbral de 200,000 unidades.

C) Selecciona dos consultas del avance 1 y crea los índices que consideres más adecuados para optimizar su ejecución.

Prueba con índices individuales y compuestos, según la lógica de cada consulta. Luego, vuelve a ejecutar ambas consultas y compara los

tiempos de ejecución antes y después de aplicar los índices.  
Finalmente, describe brevemente el impacto que tuvieron los índices en el rendimiento y en qué tipo de columnas resultan más efectivos para este tipo de operaciones.

### OPTIMIZACIÓN DE LA CONSULTA 1 (A):

Result Grid						
Filter Rows:						
Export:						
Wrap Cell Contents:						
	id_producto	nombre_producto	total_vendido	id_vendedor	vendedor_destacado	vendido_por_vendedor
▶	179	Yoghurt Tubes	199724	9	Daphne King	10285
	161	Longos - Chicken Wings	199659	10	Jean Vang	10785
	47	Thyme - Lemon; Fresh	198567	21	Devon Brewer	11050
	280	Onion Powder	198163	21	Devon Brewer	10570
	103	Cream Of Tartar	198126	11	Sonya Dickson	10348

Result 1 x					Read Only
Output					
Action Output					
#	Time	Action	Message	Duration / Fetch	
8	21:32:36	USE taller_data	0 row(s) affected	0.000 sec	
9	21:32:36	WITH productos_mas_vendidos AS ( SELECT ProductID, SUM(Quantity) AS ...	5 row(s) returned	38.969 sec / 0.000 sec	

Result Grid						
Filter Rows:						
Export:						
Wrap Cell Contents:						
	id_producto	nombre_producto	total_vendido	id_vendedor	vendedor_destacado	vendido_por_vendedor
▶	103	Cream Of Tartar	200002	9	Daphne King	10551
	179	Yoghurt Tubes	199724	9	Daphne King	10285
	161	Longos - Chicken Wings	199659	10	Jean Vang	10785
	47	Thyme - Lemon; Fresh	198567	21	Devon Brewer	11050
	280	Onion Powder	198163	21	Devon Brewer	10570

Result 2 x					Read Only
Output					
Action Output					
#	Time	Action	Message	Duration / Fetch	
25	21:02:24	USE taller_data	0 row(s) affected	0.000 sec	
26	21:02:24	WITH productos_top5 AS ( SELECT ProductID, SUM(Quantity) AS cantidad_t...	5 row(s) returned	7.532 sec / 0.000 sec	

Tiempo 1 consulta 1: 38.969 sec

Tiempo 2 consulta 1: 7.532 sec

Se utilizó el siguiente INDEX para la optimización de los tiempos de ejecución:  
CREATE INDEX idx\_ventas\_producto\_cliente\_qty ON sales (ProductID, CustomerID, Quantity);

### OPTIMIZACIÓN DE LA CONSULTA 1 (B):

Result Grid					
Filter Rows:					
Export:					
Wrap Cell Contents:					
	id_producto	nombre_producto	clientes_unicos	total_clientes	porcentaje_clientes
▶	161	Longos - Chicken Wings	14252	98759	14.4311
	103	Cream Of Tartar	14246	98759	14.4250
	47	Thyme - Lemon; Fresh	14101	98759	14.2782
	179	Yoghurt Tubes	14066	98759	14.2428
	280	Onion Powder	14058	98759	14.2347

Result 1 x					Read Only
Output					
Action Output					
#	Time	Action	Message	Duration / Fetch	
11	21:59:13	USE taller_data	0 row(s) affected	0.000 sec	
12	21:59:13	WITH productos_mas_vendidos AS ( SELECT ProductID, SUM(Quantity) AS t...	5 row(s) returned	15.031 sec / 0.000 sec	

Result Grid					
Filter Rows:		Export:		Wrap Cell Content: <a href="#">T</a>	
id_producto	nombre_producto	clientes_unicos	total_clientes	porcentaje_clientes	
161	Longos - Chicken Wings	14252	98759	14.4311	
103	Cream Of Tartar	14247	98759	14.4260	
47	Thyme - Lemon; Fresh	14101	98759	14.2782	
179	Yoghurt Tubes	14066	98759	14.2428	
280	Onion Powder	14058	98759	14.2347	

Action Output					
#	Time	Action	Message	Duration / Fetch	
22	10:46:36	USE taller_data	0 row(s) affected	0.000 sec	
23	10:46:36	WITH productos_mas_vendidos AS ( SELECT ProductID, SUM(Quantity) AS t...	5 row(s) returned	5.609 sec / 0.000 sec	

Tiempo 1 consulta 2: 15.031 sec

Tiempo 2 consulta 2: 5.609 sec

Se utilizó el siguiente INDEX para la optimización de los tiempos de ejecución:

```
CREATE INDEX idx_ventas_producto_cliente_qty ON sales (ProductID, CustomerID, Quantity);
```

Impacto:

Estos Índices compuestos tuvieron un gran impacto en reducir las lecturas físicas de disco permitiendo usar acceso indexado en vez de hacer un full table scan.