**Albert Ripoll Batlló**

Máster Big Data La Salle, Barcelona

Práctica Introducción BI

**Feb 10, 2024**

# Resumen

"TheLook" es una empresa de moda que ha experimentado una transformación significativa en los últimos años, pasando de un enfoque tradicional en tiendas físicas a una estrategia de comercio electrónico desde 2020. Durante este período, la empresa ha logrado vender productos de moda de más de 2000 marcas en 16 países a través de su plataforma de comercio electrónico. Con el objetivo de gestionar eficientemente el gran volumen de datos generado por estas operaciones, la empresa planea implementar un proyecto de Business Intelligence en 2024.

Actualmente, "TheLook" opera tres sitios web, cada uno enfocado en un continente específico, y centraliza los datos en una plataforma de comercio electrónico, utilizando Shopify. El equipo de e-commerce es responsable de la gestión de datos y actualiza los registros consolidados de ventas en la plataforma de BigQuery cada 15 minutos. Posteriormente, el equipo de Business Intelligence se encarga de procesar estos datos a través de un proceso de extracción, carga y transformación (ELT) en BigQuery. Finalmente, generan contenido analítico utilizando la plataforma de dashboarding Holistics.

El objetivo principal de este proyecto de Business Intelligence es permitir una monitorización efectiva de las ventas y facilitar la identificación de nuevas oportunidades de crecimiento para la empresa. Al consolidar y analizar los datos de manera más eficiente, se espera que "TheLook" esté mejor posicionada para tomar decisiones estratégicas informadas y continuar expandiendo su presencia global en el competitivo sector de la moda.

# 

# Implementar vista para cada entidad del diseño estrella

### Vista DT\_PRODUCT

| Create Or Replace View lasalle-albert-ripoll.caso\_practico.DT\_PRODUCT As  Select id as product\_id,  Case  When (name Is Null Or Trim(name)='')  Then 'SENSE-NOM'  Else Trim(name)  End as name,  Case  When (category Is Null Or Trim(category)='')  Then 'SENSE-CATEGORIA'  Else Trim(category)  End as category,  Case  When (brand Is Null Or Trim(brand)='')  Then 'SENSE-MARCA'  Else Trim(brand)  End as brand  From `bigquery-public-data.thelook\_ecommerce.products`; | |
| --- | --- |

### Vista DT\_TIME

| Create Or Replace View lasalle-albert-ripoll.caso\_practico.DT\_TIME AS  Select Distinct created\_at As time\_id,  Cast(Extract(YEAR From created\_at) As String) As year,  LPad(Cast(Extract(MONTH From created\_at) As String), 2, '0') As month,  LPad(Cast(Extract(DAY From created\_at) As String), 2, '0') As day,  FORMAT\_TIMESTAMP('%Y\_%m', created\_at) As yyyy\_mm,  FORMAT\_TIMESTAMP('%Y\_%m\_%d', created\_at) As yyyy\_mm\_dd  from `bigquery-public-data.thelook\_ecommerce.orders`; |
| --- |

### 

### Vista DT\_TICKET\_LINE

| Create Or Replace View lasalle-albert-ripoll.caso\_practico.DT\_TICKET\_LINE As  Select items.id As ticket\_line\_id,  items.order\_id As ticket\_id,  orders.status As order\_status,  items.status As order\_line\_status,  orders.created\_at As order\_created\_at  From `bigquery-public-data.thelook\_ecommerce.order\_items` items  Inner Join `bigquery-public-data.thelook\_ecommerce.orders` orders  on (orders.order\_id=items.order\_id); |
| --- |

### Vista DT\_CUSTOMERS

| Create Or Replace View lasalle-albert-ripoll.caso\_practico.DT\_CUSTOMER As  With users\_ids As (  Select email, max(id) as max\_user\_id, min(id) As min\_user\_id  From `bigquery-public-data.thelook\_ecommerce.users`  Group By email  ),  users\_info\_max As (  Select cust.email,  cust.first\_name,  cust.last\_name,  concat(trim(cust.first\_name),', ',trim(cust.last\_name)) As full\_name,  cust.age,  Case  When (cust.age < 18) Then '(0,18)'  When (cust.age >=18 And cust.age<25) Then '[18,25)'  When (cust.age >=25 And cust.age<35) Then '[25,35)'  When (cust.age >=35 And cust.age<45) Then '[35,45)'  When (cust.age >=45 And cust.age<60) Then '[45,60)'  When (cust.age >=60) Then '[60,-)'  Else '<N/A>'  End As age\_range,  cust.gender,  cust.country,  cust.city,  cust.traffic\_source  From `bigquery-public-data.thelook\_ecommerce.users` cust  Inner Join users\_ids ids  On (cust.id=ids.max\_user\_id)  ),  users\_info\_min AS (  Select cust.email, cust.created\_at,  ROUND(CEILING(TIMESTAMP\_DIFF(CURRENT\_TIMESTAMP(), cust.created\_at, DAY)/365),0)  As creation\_lifespan\_years  From `bigquery-public-data.thelook\_ecommerce.users` cust  Inner Join users\_ids ids  On (cust.id=ids.min\_user\_id)  )  Select max.\*, min.\* EXCEPT(email)  From users\_info\_max max Inner Join users\_info\_min min  On (max.email=min.email); |
| --- |

### Vista Tabla de Hechos FT\_SALES

| Create Or Replace View lasalle-albert-ripoll.caso\_practico.FT\_SALES As  With orders\_ext As (  Select users.email, orders.order\_id As ticket\_id, orders.created\_at As time\_id,  ROUND(CEIL(TIMESTAMP\_DIFF(orders.delivered\_at, orders.created\_at,DAY)),0)  As days\_from\_creation\_to\_delivery,  ROUND(CEIL(TIMESTAMP\_DIFF(orders.shipped\_at, orders.created\_at,DAY)),0)  As days\_from\_creation\_to\_shipment,  ROUND(CEIL(TIMESTAMP\_DIFF(orders.delivered\_at, orders.shipped\_at,DAY)),0)  As days\_from\_shipment\_to\_delivery  From `bigquery-public-data.thelook\_ecommerce.orders` orders  Join `bigquery-public-data.thelook\_ecommerce.users` users  on (orders.user\_id=users.id)  ), items\_ext As (  Select items.order\_id As ticket\_id,  items.id As ticket\_line\_id,  products.id As product\_id,  1 As total\_products,  Case  When items.status in ('Complete','Shipped','Processing')  Then items.sale\_price Else 0  End As net\_sales,  products.retail\_price As gross\_sales,  (products.retail\_price - items.sale\_price) As total\_discount  From `bigquery-public-data.thelook\_ecommerce.order\_items` items  Inner Join `bigquery-public-data.thelook\_ecommerce.products` products  On (items.product\_id=products.id)  )  Select  items.ticket\_line\_id, orders.email, items.product\_id, orders.time\_id,  orders.days\_from\_creation\_to\_delivery, orders.days\_from\_creation\_to\_shipment,  orders.days\_from\_shipment\_to\_delivery, items.total\_products,  items.net\_sales, items.gross\_sales, items.total\_discount  From orders\_ext orders Left Join items\_ext items  On (orders.ticket\_id = items.ticket\_id); |
| --- |

# 

# Capa semántica

Crear una nueva vista (V\_SALES\_SEMANTIC\_LAYER) que exponga la capa semántica a partir de las 5 tablas base en una única vista.

### V\_SALES\_SEMANTIC\_LAYER

| Create Or Replace View lasalle-albert-ripoll.caso\_practico.V\_SALES\_SEMANTIC\_LAYER As  Select FT\_SALES.email, \* EXCEPT(ticket\_id,ticket\_line\_id,product\_id,time\_id,email)  From lasalle-albert-ripoll.caso\_practico.FT\_SALES  Inner Join lasalle-albert-ripoll.caso\_practico.DT\_TICKET\_LINE  On (FT\_SALES.ticket\_line\_id=DT\_TICKET\_LINE.ticket\_line\_id)  Left Join lasalle-albert-ripoll.caso\_practico.DT\_PRODUCT  On (FT\_SALES.product\_id=DT\_PRODUCT.product\_id)  Left Join lasalle-albert-ripoll.caso\_practico.DT\_TIME  On (FT\_SALES.time\_id=DT\_TIME.time\_id)  Left Join lasalle-albert-ripoll.caso\_practico.DT\_CUSTOMER  On (FT\_SALES.email=DT\_CUSTOMER.email); |
| --- |

# 

# Preguntas sobre la capa semántica

## ¿Cuál es el producto más vendido (gross\_sales)? ¿A qué marca (brand) pertenece? ¿A qué categoría? ¿Cuántos ingresos (net sales) ha proporcionado? Asumiremos todos los importes están en EUR.

| -- Si la pregunta realmente es: el producto más vendido (max productos\_vendidos):  Select name, category, brand,  Sum(total\_products) as productos\_vendidos,  Sum(gross\_sales) as sum\_ventas\_brutas,  Sum(net\_sales) as sum\_ventas\_netas  From lasalle-albert-ripoll.caso\_practico.V\_SALES\_SEMANTIC\_LAYER  Group By name, category, brand  Order By Sum(total\_products) Desc  Limit 1;  -- Si la pregunta realmente es: el producto que más ventas brutas generó (max sum\_ventas\_brutas):  Select name, category, brand,  Sum(total\_products) as productos\_vendidos,  Sum(gross\_sales) as sum\_ventas\_brutas,  Sum(net\_sales) as sum\_ventas\_netas  From lasalle-albert-ripoll.caso\_practico.V\_SALES\_SEMANTIC\_LAYER  Group By name, category, brand  Order By Sum(gross\_sales) Desc  Limit 1; |
| --- |

## El producto más veces vendido es el “Jeans” “Wrangler Men’s Premium” ya que ha vendido 53 veces.

## 

## El producto con más ingresos brutos es el “Outwear & Coat” “Canada Goose Men’s The Chat…” ya que ha generado un total de 10595 EUR de ingresos brutos con solo 13 productos vendidos.

## 

## ¿Quién es el mejor cliente? ¿Cuántos ingresos ha generado? Utilizar indicador net\_sales

| Select email, full\_name, sum(net\_sales) as sum\_net\_sales  From lasalle-albert-ripoll.caso\_practico.V\_SALES\_SEMANTIC\_LAYER  Group By email, full\_name  Order By sum(net\_sales) Desc  Limit 1; |
| --- |

## 

## ¿Qué año ha sido el que “theLook” ha ingresado más? Utilizar indicador net\_sales

| Select year, SUM(net\_sales) As sum\_net\_sales  From lasalle-albert-ripoll.caso\_practico.V\_SALES\_SEMANTIC\_LAYER  Group By year  Order By SUM(net\_sales) Desc  Limit 1; |
| --- |

## 

## ¿Cuántos clientes únicos tiene “theLook”?

| Select count(distinct email) as clientes\_unicos  From lasalle-albert-ripoll.caso\_practico.V\_SALES\_SEMANTIC\_LAYER; |
| --- |

## 

## ¿Cuál es el promedio en días entre creación y envío, creación y entrega, y entre envío y entrega de los pedidos? En caso “theLook” quiera optimizar el tiempo total del pedido, ¿en qué proceso de la cadena de envío debería invertir recursos para mejorar?

| Select avg(days\_from\_creation\_to\_shipment) as creacion\_envio  , avg(days\_from\_creation\_to\_delivery) as creacion\_entrega  , avg(days\_from\_shipment\_to\_delivery) as envio\_entrega  From lasalle-albert-ripoll.caso\_practico.V\_SALES\_SEMANTIC\_LAYER; |
| --- |

## De promedio los tiempos de

## • Compra a Creación del Envío son de 1 día: Esta etapa implica el tiempo desde que se realiza la compra hasta que se genera la etiqueta de envío. Aquí se podría evaluar si hay posibles demoras en la confirmación del pedido, en la preparación del producto para el envío o en la generación de la etiqueta de envío. Mejorar la eficiencia en estas áreas podría reducir este tiempo.

• Envío hasta Entrega son de 2 días: Esta etapa implica el tiempo desde que el paquete se envía hasta que se entrega al destinatario. Aquí se podría evaluar la eficiencia de las opciones de transporte y logística, así como la velocidad de procesamiento en los centros de distribución y los tiempos de tránsito. Mejorar la eficiencia en estas áreas podría acortar este período.

# 

En el análisis resulta sorprendente que los tiempos de creación hasta envío más los tiempos de envío hasta la entrega no sean iguales al tiempo total del envío a la entrega (1+2=3,5). Eso es debido a que los resultados son aproximados porque en el análisis los valores han perdido precisión al redondearlos a días enteros. Además puede ser que algunos paquetes se hayan enviado pero no se hayan llegado a entregar por lo que se podría tener un desajuste de valores nulos entre las distintas columnas. Por ejemplo valores nulos en envío\_entrega y creación\_entrega pero no nulos en creación\_envío.

# 

# Implementación de dashboard en preset, que permite generar conocimiento

## ¿El negocio prospera? Mostrar a nivel temporal la evolución de ingresos.

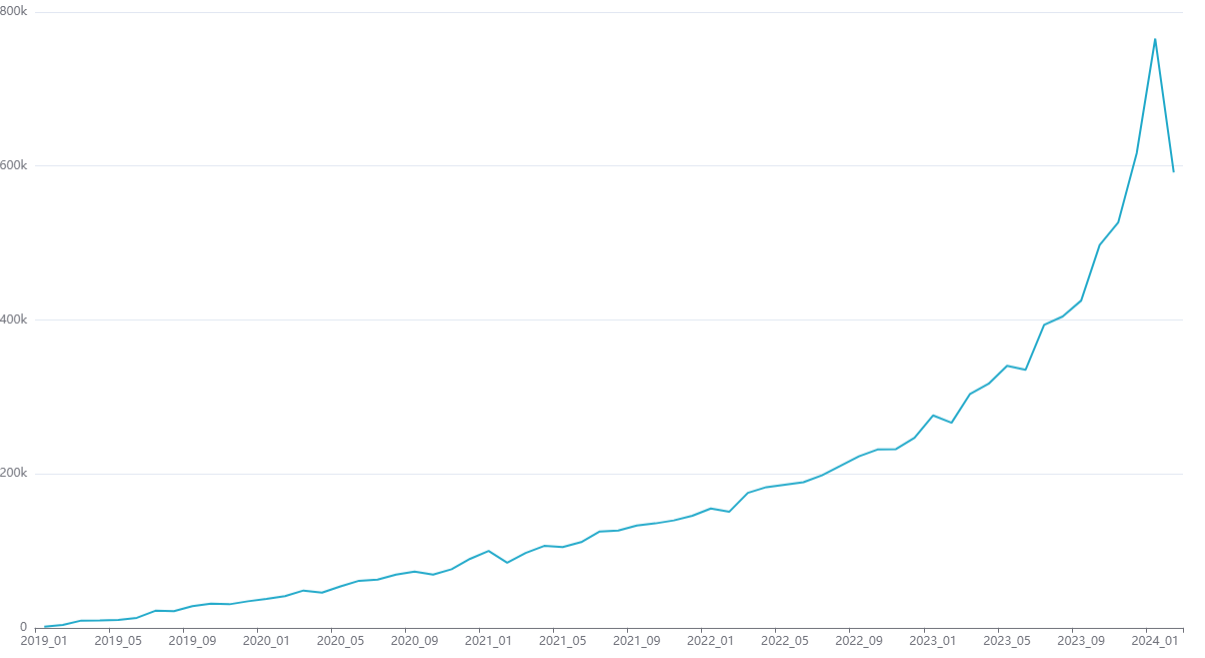
| SELECT  DATE\_TRUNC(time\_id, MONTH) as month, --trunca la columna time\_id al primer día del mes, lo que te permite agrupar por mes.  SUM(gross\_sales) as total\_sales --calcula la suma de las ventas brutas para cada mes.  FROM  `lasalle-albert-ripoll.caso\_practico.FT\_SALES`  GROUP BY -- agrupa los resultados por mes.  month  ORDER BY -- ordena los resultados en orden cronológico.  month; |
| --- |

## A pesar del bajón del último mes, el negocio prospera.

## Ingresos totales por mes vs mes

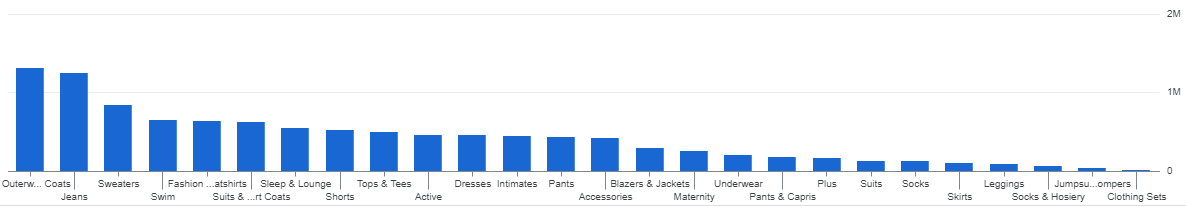
Con preset:

|  |
| --- |



## ¿Qué tipología de productos debemos posicionar más en la web? Mostrar los ingresos por categoría de producto, con un gráfico que permita visualizar la distribución.

| SELECT  P.category,  SUM(S.gross\_sales) as total\_sales --suma las ventas brutas para cada categoría de productos.  FROM  `lasalle-albert-ripoll.caso\_practico.FT\_SALES` S  JOIN  `lasalle-albert-ripoll.caso\_practico.DT\_PRODUCT` P --se utiliza para combinar las tablas FT\_SALES y DT\_PRODUCT mediante la columna product\_id  ON  S.product\_id = P.product\_id  GROUP BY --agrupa los resultados por la categoría de productos.  P.category  ORDER BY  total\_sales DESC;--ordena los resultados en orden descendente según las ventas totales. |
| --- |

Los productos de “Outwear&Coats” y “Jeans” son los productos que generan más de un millón de ingresos a la empresa.Ingresos totales por categoría de producto

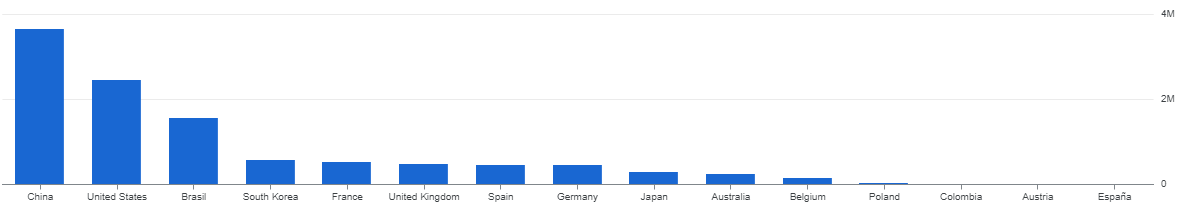
|  |
| --- |

## 

## 

## ¿En qué mercados (países) debemos invertir en marketing? Mostrar los ingresos a nivel de país.

| SELECT  country,  SUM(gross\_sales) as total\_sales  FROM  `lasalle-albert-ripoll.caso\_practico.V\_SALES\_SEMANTIC\_LAYER`  GROUP BY  country  ORDER BY  total\_sales DESC; |
| --- |

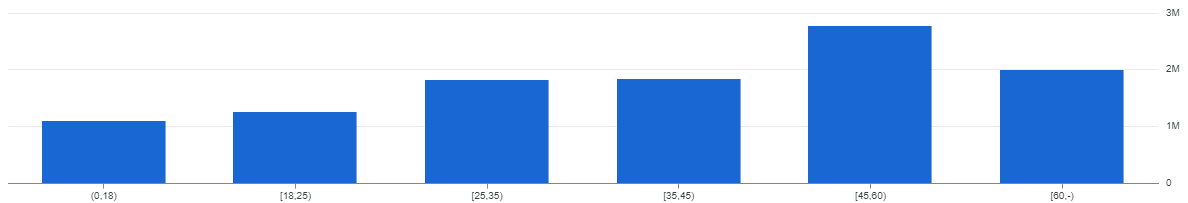


## 

## 

## ¿En qué medio de marketing deberíamos invertir las campañas? Mostrar los ingresos por rango de edad para facilitar al equipo de marketing saber la edad del consumidor que más monetiza (muy jóvenes TikTok, rango medio Facebook, mayores Televisión).

| SELECT  age\_range,  SUM(gross\_sales) as total\_sales  FROM  `lasalle-albert-ripoll.caso\_practico.V\_SALES\_SEMANTIC\_LAYER`  GROUP BY  age\_range  ORDER BY  age\_range; |
| --- |



Ingresos por rango de edad del consumidor

Si consideramos muy jóvenes de 0 a 25 años que usan principalmente TikTok, rango medio de 25 a 45 años que usan Facebook y mayores de 45 que consumen Televisión:

| SELECT  CASE  WHEN age\_range IN ('(0,18)', '[18,25)') THEN 'TikTok'  WHEN age\_range IN ('[25,35)', '[35,45)') THEN 'Facebook'  WHEN age\_range IN ('[45,60)', '[60,-)') THEN 'Televisión'  ELSE 'Otro'  END AS marketing\_channel,  SUM(gross\_sales) as total\_sales  FROM  `lasalle-albert-ripoll.caso\_practico.V\_SALES\_SEMANTIC\_LAYER`  GROUP BY  marketing\_channel  ORDER BY  total\_sales DESC; |
| --- |

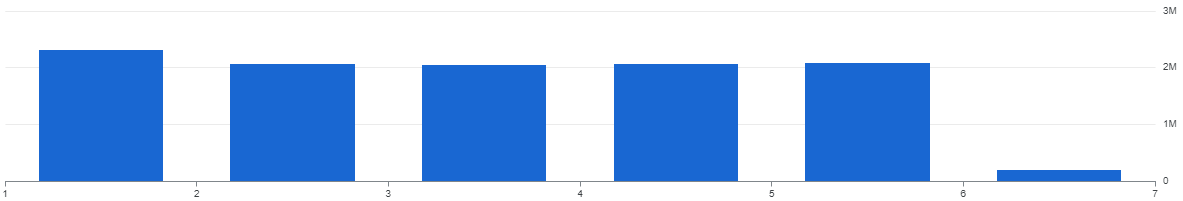


Ingresos por red de consumo principal de los consumidores

Observamos que se debería invertir más en televisión pero sin menospreciar los otros medios.

## ¿Nuestros clientes son fieles? Se propone mostrar los ingresos por el campo “creation\_lifespan\_years”, que mostrará los años que el cliente es fiel a la empresa.

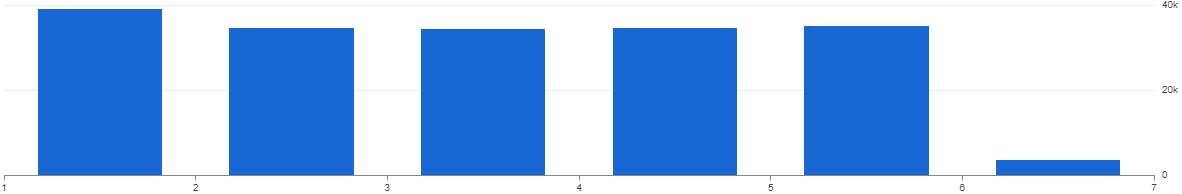
| SELECT  creation\_lifespan\_years,  SUM(gross\_sales) as total\_sales  FROM  `lasalle-albert-ripoll.caso\_practico.V\_SALES\_SEMANTIC\_LAYER`  GROUP BY  creation\_lifespan\_years  ORDER BY  creation\_lifespan\_years; |
| --- |

No observamos mucha diferencia de las ventas entre los clientes que llevan 1, 2, 3, 4, 5 o 6 años con la cuenta creada.

Ingresos por antigüedad de creación de la cuenta

Lo que sí vemos es que los usuarios que hace más de 6 años que tienen la cuenta creada generan muchos pocos ingresos. Pero esta razón radica seguramente en que hay pocos usuarios con cuenta de antiguedad mayor a 6 años. Como podemos comprobar:

| SELECT  CASE  WHEN creation\_lifespan\_years < 1 THEN '(0,1)'  WHEN creation\_lifespan\_years >= 1 AND creation\_lifespan\_years < 2 THEN '[1,2)'  WHEN creation\_lifespan\_years >= 2 AND creation\_lifespan\_years < 3 THEN '[2,3)'  WHEN creation\_lifespan\_years >= 3 AND creation\_lifespan\_years < 4 THEN '[3,4)'  WHEN creation\_lifespan\_years >= 4 AND creation\_lifespan\_years < 5 THEN '[4,5)'  WHEN creation\_lifespan\_years >= 5 AND creation\_lifespan\_years < 6 THEN '[5,6)'  WHEN creation\_lifespan\_years >= 6 THEN '[6,-)'  ELSE 'Otro'  END AS lifespan\_range,  COUNT(\*) as count\_in\_range  FROM  `lasalle-albert-ripoll.caso\_practico.V\_SALES\_SEMANTIC\_LAYER`  GROUP BY  lifespan\_range  ORDER BY  lifespan\_range; |
| --- |



Cantidad de cuentas consumidoras por antigüedad de creación de la cuenta