



**UNIVERSIDAD DE ANTIOQUIA**

1 8 0 3

**Automatización y reestructuración de los informes de  
efectividad en la venta de productos bancarios y  
asignación de oportunidades comerciales en  
Bancolombia**

**Juan Gabriel Colorado Restrepo**

Universidad de Antioquia  
Facultad de Ciencias Exactas y Naturales  
Instituto de Matemáticas  
Medellín  
2022

# **Automatización y reestructuración de los informes de efectividad en la venta de productos bancarios y asignación de oportunidades comerciales en Bancolombia**

**Juan Gabriel Colorado Restrepo**

Trabajo de grado presentado como requisito parcial para optar al título de:  
Estadístico

## **Orientador Interno**

Raúl Alejandro Morán Vásquez  
Instituto de Matemáticas

## **Orientador Externo**

Sebastian Betancur Téllez  
Grupo Bancolombia

Universidad de Antioquia  
Facultad de Ciencias Exactas y Naturales  
Instituto de Matemáticas  
Medellín, Colombia  
2022

Dedicado a mi familia, amigos y docentes, todos han hecho a su manera su mejor aporte a mi vida y educación.

## **Agradecimientos**

Deseo expresar mi enorme gratitud a mis colegas durante mi paso por el Grupo Bancolombia, David Alejandro Hernández Pérez y Alex Santiago Contreras Henández, quienes dedicaron tiempo y esfuerzo en mi instrucción y constante aprendizaje durante el período de prácticas. A mi tutor externo de grado, jefe de sección del Área Analítica Personas del Grupo Bancolombia, Sebastian Betancur Téllez, quién confió en mi trabajo y manifestó un total apoyo en mis labores diarias. Finalmente deseo agradecer al profesor Raúl Alejandro Morán Vásquez, por su dedicación, constancia y consejos para llevar a buen término mi proceso académico.

## Resumen

Durante el proceso de prácticas académicas surgió una cuestión elemental respecto a los datos adquiridos en el área de Analítica Personas de Bancolombia para la elaboración de los diferentes informes que allí se tramitan, relacionada con la funcionalidad de la información obtenida, ya que, en algunos casos, los valores que se presentaban no correspondían adecuadamente con lo estimado por otros departamentos que usaban los informes como base de verificación de la efectividad en las ventas de los productos financieros, cuestión fundamental para los analistas estratégicos quienes se encargan de la planeación y distribución de los productos bancarios por los diferentes canales de atención y sucursales, y la importancia de la revisión de las cifras se enmarca en buscar donde está la dificultad para que los clientes adquieran un producto. Es por eso que la preparación de los datos es una herramienta crucial para la gestión comercial y contribuye en la evaluación de los indicadores financieros del banco.

**Palabras clave:** automatización, clasificación, bases de datos, algoritmos, tableros de negocios.

## Abstract

During my internship, an elemental topic raised from the data gathered by “Analítica Personas de Bancolombia” area, with the objective to elaborate different informs related to the functionality of the obtained information, since in some cases, the presented results didn’t match with the ones that other departments estimated based on the efficiency of the financial product sales, main feature for the strategic analysts who were in charge of the banking products planning and distribution through the different service channels and branches, and the importance of data review is to address the problem when customers are trying to acquire a product. That’s why data preparation is really important for the commercial management and it contributes to the evaluation of the financial indicators of the bank.

**Keywords:** automation, clustering, databases, algorithms, dashboards

# Automatización y reestructuración de los informes de efectividad en la venta de productos bancarios y asignación de oportunidades comerciales en Bancolombia

Juan Gabriel Colorado Restrepo \*

26 de marzo de 2022

---

\*E-mail: [jgabriel.colorado@udea.edu.co](mailto:jgabriel.colorado@udea.edu.co), Instituto de Matemáticas, Universidad de Antioquia, Medellín, Colombia.

## Resumen

El proceso de consolidar los datos que se necesitan para obtener con la mayor precisión y brevedad posible acerca de los productos adquiridos por los clientes, tanto de la categoría de preaprobados como de potenciales, y a su vez tener con relativa certeza a cuales clientes ofrecer las correctas oportunidades comerciales de acuerdo a su capacidad de endeudamiento e historial crediticio, es uno de los principales servicios que ofrece el área de Analítica Personas dentro de Bancolombia. Para ello, se requiere una reestructuración y un proceso de automatización que genere una base de datos sólida y confiable que minimice en cuanto sea posible los riesgos económicos y que sirva como insumo para los modelos estadísticos que se proponen desde las diferentes áreas que solicitan la información.

*Palabras clave: automatización, clasificación, bases de datos, algoritmos, tableros de negocios*

## Contenido

<b>Agradecimientos</b>	<b>4</b>
<b>Resumen</b>	<b>5</b>
<b>1. Introducción</b>	<b>8</b>
<b>2. Marco Teórico</b>	<b>9</b>
2.1. Declaraciones y conceptos usados en programación de bases de datos . . . . .	9
2.2. Herramientas usadas para la elaboración de los informes	10
<b>3. Metodología</b>	<b>11</b>
<b>4. Informes</b>	<b>13</b>
4.1. Resultados . . . . .	13
4.2. Discusión . . . . .	20
<b>5. Conclusiones y Recomendaciones</b>	<b>21</b>
<b>6. Anexos</b>	<b>22</b>

## 1. Introducción

Desde el área de Analítica Personas de Bancolombia, perteneciente a la Vicepresidencia de Negocios, Personas y Pymes, se ejecuta manualmente de manera periódica el informe de asignación de oportunidades comerciales y también el informe de efectividad sobre las ventas de los productos bancarios adquiridos por los usuarios. En el primer caso, implica aquellos productos que se encuentran dentro del portafolio de servicios para personas naturales, asignados a los clientes de forma proactiva, es decir, a través de unos rigurosos filtros de verificación y observación del comportamiento crediticio del cliente; así el banco puede ofrecer un producto particular antes de que el cliente lo solicite, de forma que se agilice el proceso para otorgar dicho producto. El segundo informe anteriormente mencionado, verifica la estadística descriptiva sobre el total de ventas y los porcentajes de efectividad de acuerdo al canal de promoción del producto y la sucursal que frecuenta el cliente. Estos resultados son la base para que las diferentes secciones encargadas de gestionar los productos bancarios puedan tomar las decisiones correctas al momento de ofrecerlos, así como también aumentar o disminuir la capacidad de oferta y asignar el cupo adecuado para cada servicio.

Los esquemas actuales de producción de conocimiento acerca de los clientes usando la información suministrada por ellos y la gestionada por la organización desde el ingreso del cliente, permite construir posibles perfiles que permiten visualizar con cierto grado de exactitud la vida económica de los clientes. De esta forma es factible admitir que los modelos predictivos evolucionan a medida que se ingresa más información que ayude a complementar lo que ya se conoce sobre el cliente, y así establecer un modelo que entienda mejor las necesidades, preferencias y proyectos de cada persona. Por lo anterior, es necesario consolidar esa información de forma óptima consultando las fuentes proporcionadas por la organización, construir la lógica adecuada de manera que la información llegue con la menor pérdida posible y finalmente estructurarla de manera que se facilite su lectura y sea accesible para su tratamiento desde las diferentes áreas que la solicitan, todo esto bajo el marco de las buenas prácticas definidas por la empresa.

Dentro de la dinámica que está implementando la sección Analítica Personas, mediante metodologías eficientes de producción, es fundamental la automatización de procesos, ya que disminuye el tiempo que los analistas invierten en la ejecución manual de algunas actividades para enfocarse en resolver problemas de importancia capital o crear nuevos procedimientos para consolidar la oferta de servicios, apuntando a un conocimiento más profundo de los clientes a través del análisis de datos. De esta manera, urge reestructurar las bases de datos que contienen la información de los clientes, haciendo énfasis en los requerimientos sobre los registros, incluyendo atributos de los clientes que estaban excluidos, definiendo nuevos parámetros que especifiquen el posible acceso u omisión de clientes a cierto tipo de productos. Finalmente, es necesario consolidar la ejecución de los informes de manera automática y periódica, como también diseñar su presentación en tableros de Power BI (software para carga y visualización de datos mediante tableros interactivos), con el fin de entregar la información de manera concisa, organizada y accesible para su uso y distribución interna.

## 2. Marco Teórico

### 2.1. Declaraciones y conceptos usados en programación de bases de datos

El lenguaje de programación para la construcción de bases de datos es muy amplio y se emplean varios programas para su diseño, principalmente SQL (lenguaje de programación usado para administrar bases de datos relacionales) y Python (lenguaje de programación multiplataforma de código abierto). A continuación se mencionan solo las declaraciones de lenguaje de definición de datos (DDL) que con mayor frecuencia se usaron para construir los scripts que realizan las consultas en la LZ (infraestructura de almacenamiento y análisis de grandes cantidades de información utilizada por el Grupo Bancolombia).

- CREATE: utilizado para crear tablas y definir columnas.
- ALTER: utilizado para alterar tablas que incluyen agregar y quitar columnas y modificar sus tipos de datos.
- TRUNCATE: se usa para eliminar datos en una tabla pero no para borrar la tabla en si.
- DROP: se usa para eliminar tablas.
- INSERT: se utiliza para insertar una fila o varias filas de datos en una tabla.
- SELECT: lee o selecciona fila o filas de una tabla.
- UPDATE: se usa para editar una o varias filas en una tabla.
- DELETE: elimina una o varias filas de datos en una tabla.
- COUNT: recupera el número de filas que coinciden con los criterios de consulta.
- DISTINCT: se usa para eliminar valores duplicados de un conjunto de resultados.
- LIMIT: se usa para restringir el número de filas recuperadas de la base de datos.
- GROUP BY: esta declaración agrupa filas que tienen los mismos valores en otras filas con los resúmenes de los datos, es a menudo usada con funciones de agregación para agrupar el conjunto de resultados en una o más columnas.
- ORDER BY: se utiliza para ordenar el conjunto de resultados en orden ascendente o descendente.
- EXISTS: utilizado para comprobar la existencia de cualquier registro en una subconsulta.
- CASE: es una declaración que evalúa las condiciones propuestas y devuelve un valor cuando se cumple la primera condición, como una declaración if-then-else.
- PRIMARY KEY: esta declaración identifica de forma única cada registro en una tabla.
- FOREIGN KEY: es un atributo (o varios atributos) en una tabla, que indica que son PRIMARY KEY en otra tabla. Se utiliza para evitar procesos que destruyan los enlaces entre tablas.
- WHERE: es una declaración utilizada para filtrar los registros.
- BETWEEN: es un operador que selecciona valores dentro de un rango especificado. Los valores pueden ser numéricos, texto o fechas.
- JOINS: esta declaración se usa para combinar filas de dos o más tablas, en función de una columna relacionada entre ellas.

## 2.2. Herramientas usadas para la elaboración de los informes

Las herramientas que se usaron durante el proyecto son:

- **Python:** es un lenguaje de programación orientado a objetos, interactivo e interpretado. Es un lenguaje de programación muy versátil, que se puede utilizar para escribir programas de una baja complejidad como también para crear soluciones a problemas complejos de escala empresarial o científica. Los usos de este lenguaje de programación se pueden encontrar en el desarrollo de aplicaciones de escritorio y videojuegos, análisis matemático y científico de datos, y desarrollo web e internet.
- **Impala:** es un motor de consultas SQL open source de Cloudera para el procesamiento masivo en paralelo de los datos almacenados en un clúster de computadoras corriendo Apache Hadoop.
- **Spyder:** es un entorno open source escrito en Python, posee un enfoque científico y fue diseñado por y para científicos, ingenieros y analistas de datos. Presenta diversas funcionalidades como: edición, análisis, depuración y creación de perfiles de una herramienta de desarrollo con la exploración de datos, ejecución interactiva, inspección profunda y capacidades sofisticadas de visualización de un paquete científico.
- **Visual Studio Code:** es un editor de código fuente ágil, potente y versátil. Viene con soporte incorporado para Javascript, Typescript y Node.js, y contiene un amplio sistema de extensiones para otros lenguajes de programación como C++, C#, Java, Python, PHP, Go, entre otros.
- **Power BI:** software que permite la carga y visualización de los datos mediante tableros interactivos con enfoque en inteligencia de negocios, con el fin de obtener conocimiento sobre los datos.

### 3. Metodología

Para abordar la situación presente se debe tener en cuenta el contexto general bajo el cual se iba a proceder, que consistía principalmente en la preparación de los datos; proceso que llevaba un gran avance en su construcción, pero que no lograba traer toda la información requerida. Por lo demás, debían realizarse las modificaciones pertinentes para mejorar su implementación y disminuir las interrupciones por errores durante el proceso de ejecución ajenos al código (por ejemplo, problemas en la LZ). Cuatro bases de datos eran el objetivo, tres de ellas debían ser intervenidas y solo una se iba a ver afectada por las modificaciones que se le hacían a las demás, ya que casi toda la información se extraía de ellas y se concatenaba con una información adicional de una tabla maestra. Esta última no es de interés fundamental en el proyecto.

Las tres bases de datos que debían organizarse fueron: inicialmente la relacionada con el informe de efectividad que almacenaba la información de todos los productos financieros adquiridos por personas naturales, tanto de clientes potenciales como de clientes preaprobados. En segundo lugar, la base de datos relacionada con el informe de asignación que contiene a todos los posible clientes, tanto potenciales como también preaprobados, que pueden adquirir productos financieros debido a la capacidad de oferta del banco y la capacidad de cada cliente de contraer la obligación. Finalmente, la tercera base de datos consiste en un historial que contiene la información de los clientes únicamente catalogados como preaprobados que habían adquirido algún servicio dentro del banco. Las dos primeras bases de datos tenían una periodicidad mensual, es decir, mensualmente se borran los registros anteriores para incluir los del mes en cuestión. La última tabla creada debía contener la información de los clientes preaprobados desde hace aproximadamente dos años, es decir 2020 y 2021.

La fase inicial se enfocó en automatizar el proceso para la generación de cada tabla, con el propósito de disminuir la intervención manual al momento de ejecutar los scripts. La que tenía los procesos más simplificados era la creación de la tabla de asignación de oportunidades comerciales, a la cual se le modificaron líneas de código que evitaron el ingreso periódico de valores que posteriormente toma durante la ejecución: asociados a fechas, extracción de datos específicos de tablas almacenadas en la LZ, declaraciones para verificar compatibilidad entre registros y extraer lo necesario, entre otras acciones que permitieron que la tabla básicamente no necesitara ninguna modificación posterior. En los anexos se puede observar un ejemplo del tipo de script que trae la nueva información (ver anexo 1).

Posteriormente, se continuó con la generación de la tabla que contenía el histórico de clientes preaprobados, la cual tenía una mayor complejidad debido a que se usan varios archivos durante la implementación del algoritmo. Uno de ellos era un diccionario que contenía los parámetros que se usarían en los códigos, cuyos valores de las claves se cambiaban mensualmente en cada ejecución reemplazando los valores a los que hacen referencia los parámetros en los scripts (ver anexo 2).

A continuación siguieron las modificaciones pertinentes al archivo denominado “orquestador.py”, el cual contiene un algoritmo que se encarga de realizar paso a paso el proceso de creación de la base de datos, desde la verificación de las credenciales de la ODBC para validar los accesos a las tablas fuentes, como también la conexión con Impala, la búsqueda de los archivos alojados localmente, la ejecución de los scripts de SQL que tienen las consultas para construir las tablas intermedias y que son necesarias para consolidar la tabla principal, y finalmente, almacenarla en la zona de resultados de la LZ (ver anexo 3).

Luego se reestructuraron los archivos en Python que realizan la extracción de información a partir de

los insumos locales que se proporcionaban desde el área de Servicio de Clientes. Uno de ellos contenía la información mensual de los clientes preaprobados y el otro contenía la información de los clientes del segmento preferencial. Ambos archivos necesitaban ajustes en los parámetros requeridos.

Finalmente, el archivo que debía ser ampliamente modificado era el de las consultas de SQL que construían la tabla, esto a partir de una plantilla anterior creada para extraer la información de los clientes en el año 2019 y que debía acoplarse a los requisitos del área para contemplar la información de los años 2020 y 2021 (ver anexo 4).

Más adelante los esfuerzos se centraban en la preparación de los datos para la tabla con la que se generaba el informe de efectividad, que iba a contener los productos financieros adquiridos por los clientes, tanto potenciales como preaprobados; con una implementación análoga a la construcción de la tabla del histórico de preaprobados. Inicialmente se ejecutaba un archivo .json que contiene un diccionario con los parámetros permanentes y los que debían cambiarse mensualmente antes de ejecutarse el proceso (ver anexo 2).

La construcción de esta base de datos se componía de dos grandes fases, la extracción de la información de los clientes potenciales por una parte, toda a partir de las tablas maestras y tablas diccionarios alojadas en la LZ, las cuales van adquiriendo la información mediante las consultas en SQL. Por otra parte, la información restante se obtenía en los archivos sobre las ventas de los clientes preaprobados y clientes preferenciales entregados por el área de Servicio de Clientes, alojados localmente en el equipo personal. Estos scripts que contienen todas las consultas procesan, en general, las siguientes fases: inicialmente se traen de tablas maestras los nombres de los productos, claves o códigos de identificación de productos, de sucursales, de asesores comerciales, entre otro tipo de información principalmente relacionada con la estructura bancaria. Posteriormente se generan las métricas de los indicadores por producto con respecto a variables como las anteriormente mencionadas y otras que se van adquiriendo en el proceso, como por ejemplo: las regiones y zonas delimitadas por el banco, segmentos de clientes, agrupación de sectores económicos definidas también por el banco, canales de venta, tipo de gestión que realizó el banco, entre otras. Finalmente, el último script, se encarga de organizar la información concatenando las consultas anteriores con las nuevas declaraciones que traerán la información de los clientes y así consolidar la información sobre el volumen de los productos adquiridos, volumen de ventas en dinero, cantidad de productos ofrecidos y su respectivo volumen en dinero, construyendo así la tabla de efectividad mensual. Este proceso anterior es análogo para la obtención de las ventas de los clientes preaprobados (ver anexo 5).

Inmediatamente se construían estas tablas, se organizaba la información de tal manera que se conseguían las mismas variables para ambas tablas, considerando que provienen de distintos procesos, luego el último script “unir\_bases\_y\_cargas.sql” une totalmente ambas tablas consolidando así la base de datos de efectividad de productos bancarios, la cual es almacenada en la zona de resultados de la LZ.

## 4. Informes

### 4.1. Resultados

En esta etapa del proceso se llevaría a cabo la migración de los informes hacia Power BI; el objetivo era construir tableros de negocios interactivos que presentarán de forma gráfica la estadística requerida por las diferentes áreas del banco sobre ciertas medidas fundamentales para ellos, como por ejemplo las oportunidades asignadas y el volumen de dinero que el banco dispone para ofrecerlas, así como también la cantidad de productos que se entregaron a los clientes y su respectivo volumen en dinero. Lo anterior filtrado a través de canales de venta, sucursales, regiones o zonas del país, tipo de cliente, segmento al que pertenece, categoría del producto, entre otros factores de interés para quienes requieren el informe.

Un bosquejo de la base de datos obtenida contiene los siguientes atributos o variables:

- **Oportunidades comerciales asignadas:** hace referencia a la cantidad de productos, diferenciados por código de acción, ofrecidos por el banco a los posible clientes que lo podrían tomar.
- **Volumen asignado:** es la cantidad de dinero que el banco dispone para ofrecer las oportunidades comerciales a los clientes.
- **Efectivos:** son las ventas que el banco logró concretar, es decir, la cantidad de productos que adquieren los clientes del banco.
- **Volumen efectivo:** se refiere al dinero desembolsado por el banco a los clientes que adquieren sus productos.
- **Segmento:** es una clasificación propia del banco para catalogar los clientes y permite una mayor personalización de los servicios ofrecidos. Los valores que puede tomar son: 4 que indica el segmento Personas, 6 el segmento Preferencial, M el segmento Plus, S el segmento Social y el 9 el segmento Independientes.
- **Clúster persona:** es una clasificación propia del banco para catalogar el sector productivo al que pertenece cada cliente; entre algunos de los que allí se encuentran están: agroindustria, inmobiliario y construcción, salud, educación, comercio, inversionistas, etc.
- **Tipo canal:** son los diferentes canales de atención para los clientes de diferentes segmentos. Algunos de ellos son: Sucursal, Emergia, Calm, Vivienda, Sufi, Carterización, Preferencial, entre otros.
- **Código de acción:** se considera como el número de identificación de cada producto o servicio que ofrece el banco.
- **Mes:** es el número correspondiente al mes en que se analizan las ventas y las oportunidades asignadas.
- **Zona:** indica cual es la localidad, zona o región donde fue ofrecido y adquirido el producto por el cliente, esta también es una segmentación propia del banco.
- **Detallado:** hace referencia al nombre detallado del producto que se maneja en el Área de Analítica Personas. Algunos son: Cuentas, Consumo, Captaciones, Tarjetas de Crédito, Hipotecario, entre otros.

- **Producto:** indica cual es el nombre general de determinado producto o servicio, por ejemplo, tarjeta de crédito es un nombre de producto, mientras que los productos detallados podrían ser: tarjeta de crédito AMEX oro, tarjeta de crédito Master Platinum o tarjeta de crédito Puntos Colombia.
- **Foco:** es un mecanismo para agrupar los beneficios y productos que se le van a ofrecer a un grupo específico de clientes, donde el foco IVC indica ciertos productos nuevos para los clientes que intentan mejorar su experiencia en el banco mientras que el foco USO indica productos que no le otorgan mejores prestaciones al cliente sino otro tipo de beneficios.
- **Categoría:** es una clasificación que indica cuales clientes tienen productos disponibles para su uso inmediato en el banco y cuales podrían adquirir los productos del banco, los primeros son los Preaprobados y a diferencia de los segundos, que son los Potenciales, tienen un cupo especificado para el producto que el banco les ofrece, que en algunos casos puede ser menor, mientras que los Potenciales no tienen un cupo disponible debido a que no se le ha aplicado un estudio crediticio a ese cliente específico.
- **Tipo de gestión:** indica si el banco fue el sujeto activo en el proceso de ofrecer el producto, es decir, usa alguno de los canales de atención para contactar al cliente. Por otra parte, existe la posibilidad de que si el cliente visita una sucursal bancaria tramitando alguna solicitud o requiriendo cierto servicio, el banco decida ofrecerle algún producto que el cliente puede tomar inmediatamente; el primer proceso se le conoce como Proactivo y al segundo como Reactivo.

A continuación se observan los diseños de los tableros de negocios, todos deben contener la tabla como requisito del área para analizar el comportamiento de los valores según los diferentes filtros aplicados a lo largo de la presentación del informe. El primero de ellos permite visualizar las cifras de las oportunidades asignadas según el clúster de personas.

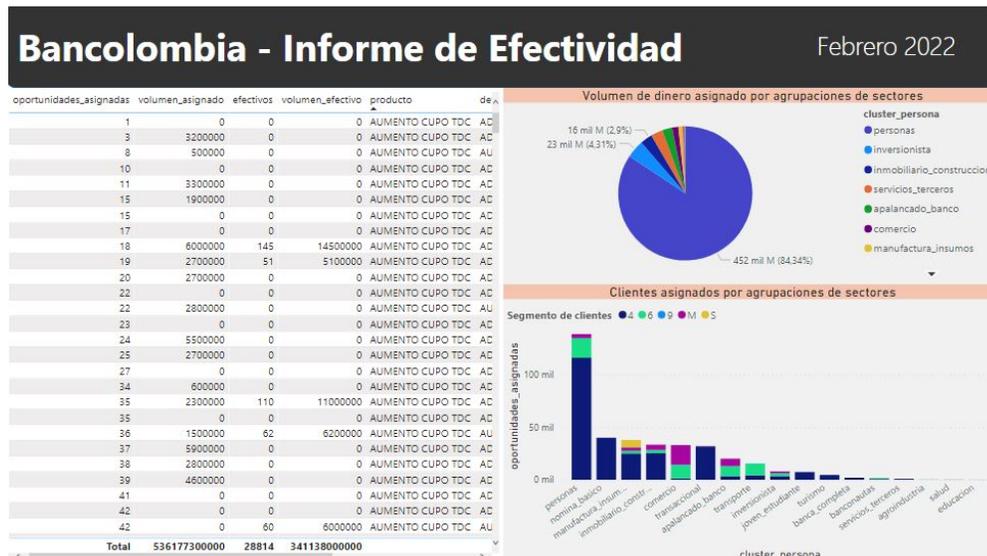


Figura 1: Oportunidades comerciales según el clúster de personas

Considerando la base de datos simulada, en la Figura 1 correspondiente al primer dashboard, se

puede observar que la mayor cantidad de oportunidades comerciales ofrecidas por el banco se dirigen hacia este sector económico Personas, con un 84.34% del volumen total de dinero, lo que concuerda con el número de oportunidades asignadas, la cual también fue en su mayoría para este mismo sector, siguiéndole nómina básico y manufactura de insumos, mientras que en cuanto a volumen los que siguen son inversionistas e inmobiliario y construcción con valores menores al 5%.

En la Figura 2 se detalla las cifras de los clientes efectivos según también el clúster de personas.

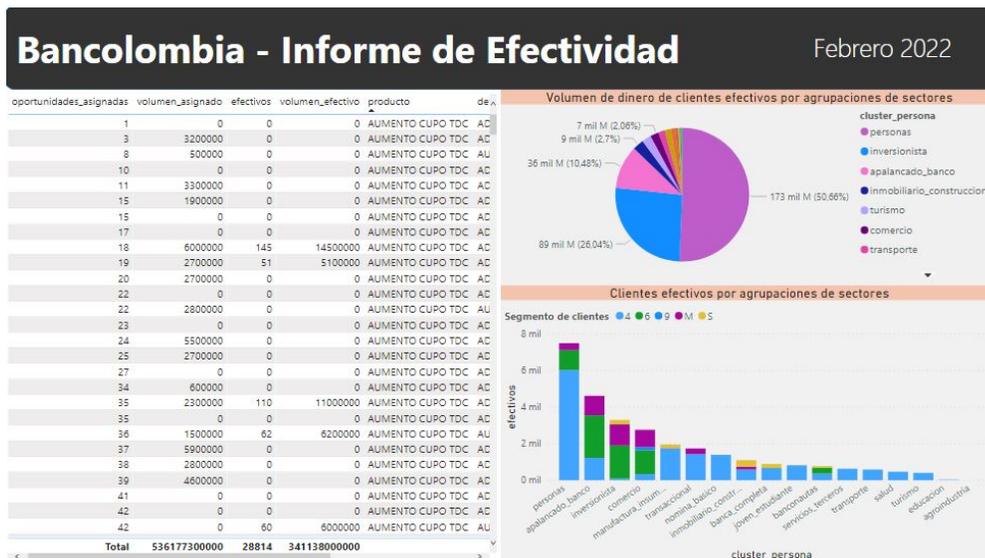


Figura 2: Efectividad según el clúster de personas

En este dashboard se observa por ejemplo que el clúster de personas que mayor volumen de dinero obtuvo a través de los diferentes productos bancarios fue el de Personas, con el 50.66% del total negociado, representando alrededor de 173 mil millones de pesos; luego sigue el clúster de Inversionistas con el 26.04% que representan 89 mil millones de pesos, y a continuación está el grupo de Apalancados Banco con el 10.48% que equivalen a 36 mil millones de pesos. La gráfica inferior nos muestra que el clúster de Personas fue el que más solicitudes de servicios efectivos generó, mientras que le sigue Apalancados Banco y luego el clúster de Inversionistas, que a pesar de haber generado menos productos efectivos, consiguió un mayor volumen de dinero que el segundo.

En la Figura 3 se observa el dashboard que explica las cifras de la cantidad de oportunidades comerciales y el volumen de dinero que se puede distribuir a través de estas.

En este tablero se evidencia que casi la tercera parte del volumen de dinero para las oportunidades comerciales estuvo gestionado por el canal de ventas Emergia, mientras que la otra tercera parte del volumen de dinero era manejada por los canales Calm y Brm, casi en la misma proporción entre ellos; el equivalente en dinero de estos tres canales equivale a aproximadamente 382 mil millones de pesos aproximadamente. Aún con estos resultados, no fueron los canales que mas gestionaron

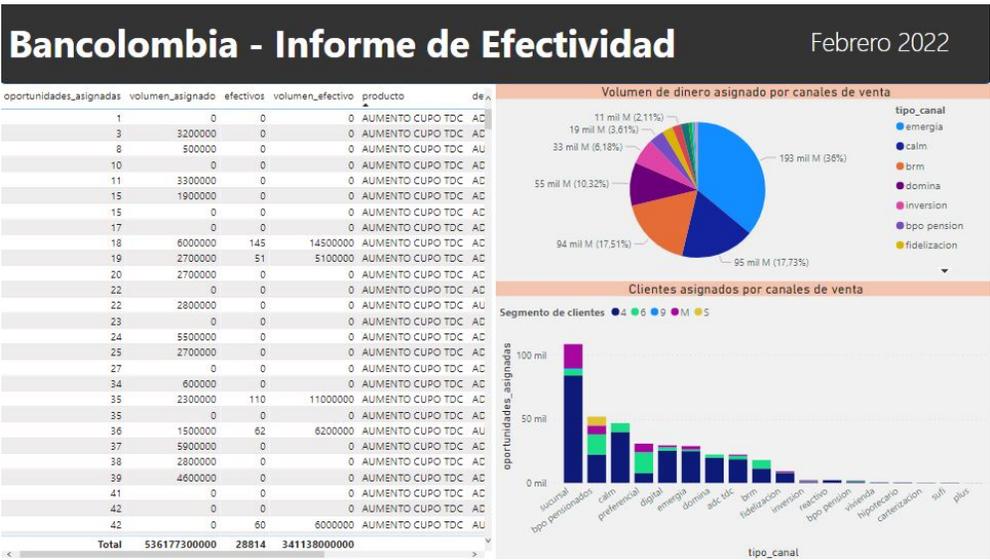


Figura 3: Oportunidades comerciales según el canal de atención al cliente

oportunidades comerciales, solo Calm aparece en tercer lugar, en el segundo Bpo pensionados y en primer lugar Sucursales; esto podría dar un indicio de la posible rentabilidad de los productos ofrecidos por los tres primeros canales de venta analizados, dado el enorme volumen de dinero que manejan para la asignación de productos bancarios.

En la Figura 4 se encuentra el dashboard que muestra las cifras de efectivos y el volumen efectivo por los diferentes canales de atención al cliente.

El canal Emergia, en este caso, ocupando el segundo lugar, hizo efectivo el 20.26% del volumen total de dinero en los productos bancarios, cuando fue el que más volumen en dinero tenía asignado; mientras que el primer lugar lo ocupa el canal de Inversión, con el 28.33% del volumen total de dinero en productos efectivos. Esto es muy importante dado que, este canal de atención ocupaba el quinto lugar en el volumen de dinero de oportunidades comerciales. Lo que muestra que los productos que ofrecen son de alto valor dado que este canal ocupa el séptimo lugar en la cantidad de clientes que tomaron sus productos, superado por los canales Sucursales, Preferencial y ADC TDC, que son respectivamente los tres primeros canales de atención por donde los clientes tomaron más productos bancarios.

En la Figura 5 se presenta el comportamiento de los segmentos en cuanto a oportunidades comerciales y productos efectivos.

Todos los gráficos apuntan a una misma conclusión, el segmento 4 correspondiente a Personas es el que mayor volumen de dinero asigna a las oportunidades y a su vez el mayor volumen efectivo genera. Lo mismo sucede con la cantidad de oportunidades y de productos efectivos. Un análisis análogo puede aplicarse al segmento 6 que corresponde a clientes preferenciales. Entre ambos se llevan el alrededor del 72% del volumen efectivo, equivalente a a 246 mil millones de pesos.



Figura 4: Efectividades según el canal de atención al cliente

La Figura 6 muestra el volumen del dinero de la distribución de oportunidades comerciales y el volumen de efectividad por zonas.

Los gráficos ponen en evidencia cuatro zonas claves: Bogotá y Preferencial Bogotá, como también Antioquia y Preferencial Antioquia. Lo relevante es que estas cuatro regiones se hicieron con aproximadamente 250 mil millones de pesos en volumen efectivo del total que es 341 mil millones de pesos, es decir aproximadamente el 72%.

La Figura 7 nos muestra el tablero que indica cuales son los productos más solicitados por los clientes que se hicieron efectivos y cual fue el volumen de dinero generado por estos productos, donde claramente se observa que la tarjeta de crédito fue el producto que mas volumen de dinero generó, aproximadamente 168 mil millones de pesos, siguiendo inversión y en tercer lugar vivienda; mientras que tarjetas de crédito fue el segundo en la cantidad de clientes que hicieron efectivos los productos ofrecidos, indicando el alto valor de dichos productos; el primero en esta lista fue Cuentas y el tercero Captaciones, ninguno de estos dos aparece dentro de los cinco primeros productos que generó mayor volumen de dinero.

Finalmente, el último dashboard, correspondiente a la Figura 8, muestra los valores totales de las oportunidades comerciales asignadas y del volumen de dinero que el banco dispuso para ofrecer estos servicios, como también la cantidad de productos ofrecidos que se hicieron efectivos y su volumen de dinero correspondiente a dichas ventas.

# Bancolombia - Informe de Efectividad

Febrero 2022

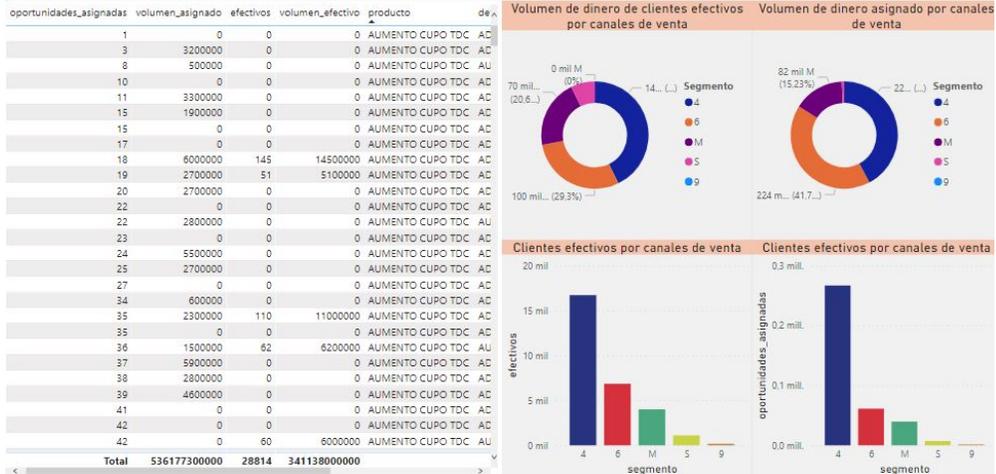


Figura 5: Análisis financiero por segmentos de clientes



Figura 6: Análisis financiero por zonas o regiones

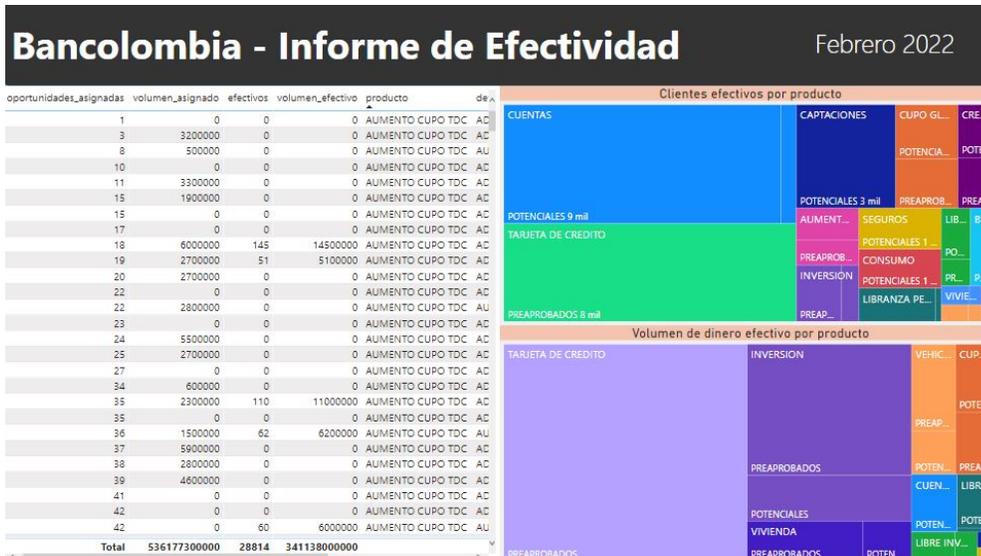


Figura 7: Efectividad de los productos o servicios financieros



Figura 8: Valores totales de las oportunidades comerciales asignadas y su respectiva efectividad

## 4.2. Discusión

La información generada y reestructurada a partir de los cambios implementados durante el tiempo de práctica académica es susceptible de modificación, mejoras o actualizaciones que puedan adaptarse a las necesidades del área encargada de gestionarla o también de los requerimientos de otras secciones del banco. La reestructuración de las bases de datos es un objetivo a largo plazo, ya que constantemente se modifican los códigos para excluir, retirar o agregar variables que ayuden a obtener información de interés sobre los clientes.

Aunque no es un tópico que se consideró en los temas de la práctica, vale la pena mencionar el rumbo que está tomando el área con respecto a la extracción de la información en sus diversas fuentes, ya que se planean usar modelos de análisis de comportamiento en redes sociales, solo por tomar un ejemplo. Esto puede indicar el trayecto que planea tomar el área de Analítica Personas con respecto a la comprensión de los clientes desde los datos, aprovechando que hay fuentes con información prácticamente ilimitada que las personas (potenciales clientes) constantemente proveen con sus preferencias, deseos, obligaciones, opiniones, entre otro tipo de manifestaciones que, si se emplean los algoritmos correctos para organizar dicha información podría convertirse en una herramienta muy sofisticada para continuar mejorando el portafolio de servicios y ofrecerlo a las personas en el momento adecuado.

Esto último no indica que habrá un desplazamiento a corto plazo de las prácticas estadísticas actuales y del manejo tradicional de las bases de datos, pero si inicia una disyuntiva donde lo mejor de ambos marcos de referencia pueden converger para complementarse mutuamente y generar modelos robustos que mejoren la experiencia de los clientes lo mejor posible al adquirir productos financieros.

## 5. Conclusiones y Recomendaciones

Las bases de datos que se consolidaron durante el tiempo de práctica académica lograron satisfacer los requerimientos de las diferentes áreas en cuanto a la información que logra extraer de las tablas maestras alojadas en la infraestructura de almacenamiento usada por el banco (LZ), y de otros insumos enviados desde otras áreas que se alojaban localmente en el equipo. La idea al momento de automatizar estos procesos consiste también en que esos insumos que enviaban al área se hicieran de forma automática a través de la LZ, de manera que se omitieran todos los permisos de seguridad con que se debía respaldar cada archivo, ya que esto presentaba inconvenientes cuando se trataba de acceder a la información usando las librerías de Python. Por ello, en su debido momento, se planeaba una migración a la LZ para que el acceso a la información se ejecutara sin inconveniente únicamente otorgando la clave de acceso a la ODBC.

Finalmente, como se mencionó con anterioridad, la estructura de las bases de datos entregadas son susceptibles de cambios en la medida en que crece y se conoce con mayor profundidad el negocio, haciendo de estas que sean más limpias, es decir, con registros más exactos, también con una menor cantidad de registros nulos, atributos más dicentes sobre la información del cliente, parámetros de filtración que identifique personas con ciertas características de manera más eficiente, entre otros cambios que pretenden proporcionar mejoras continuas que permitan comprender el comportamiento financiero del cliente.

Los reportes que migraron a Power BI son una herramienta más completa, dinámica y de fácil acceso entre quienes toman las decisiones respecto a la asignación de productos bancarios por canal, región, sucursal, entre otras características; ya que el software permite usar herramientas para filtrar respecto a las atributos de mayor interés y así identificar la efectividad de los productos financieros, logrando visualizar, desde la estadística descriptiva, donde hay retenciones y oportunidades de mejora en los aspectos técnicos, como también donde enfocar los esfuerzos desde el ámbito estratégico y comercial.

## 6. Anexos

Anexo 1.

Ejemplos de consultas (SQL) asociadas a la creación de la tabla de asignación de oportunidades comerciales a los clientes.

```
DROP TABLE IF EXISTS proceso.informe_asignacion_client PURGE;

CREATE TABLE proceso.informe_asignacion_client STORED AS
PARQUET AS WITH clientes_parrilla AS (
SELECT
    DISTINCT id ,
    tipo_id
FROM
    proceso_clientes_*****_$. parrilla_actual
WHERE
    tipo_id IS NOT NULL
UNION ALL
SELECT
    DISTINCT num_id AS id ,
    tipo_id
FROM
    proceso_clientes_*****_%$preaprobados_herramientas
WHERE
    vigencia >= CAST(REPLACE(SUBSTR(CAST(DATE_TRUNC(
        "MONTH", NOW()) AS STRING), 1, 10),
        "-", "")) AS INT)
AND tipo_id IS NOT NULL
UNION ALL
SELECT
    DISTINCT id ,
    tipo_id
FROM
    proceso_clientes_*****_$$$. priorizador_principal
WHERE
    tipo_id IS NOT NULL
),
dist_preap_etiqueta AS (
SELECT
    id ,
    t1.tipo_id ,
    asesor ,
    CASE
    WHEN tipo_canal = "BPO_GEST" THEN "colectivos"
    WHEN tipo_canal = "PREFE" THEN "preferencial"
    WHEN tipo_canal = "ANALIT_PERSON" THEN "reactivo"
    WHEN TRIM(tipo_canal) = "plus" THEN "MR_plus"
    WHEN TRIM(tipo_canal) IN ("", "_")
    OR tipo_canal IS NULL THEN "reactivo"
```

```

ELSE tipo_canal
END AS tipo_canal
FROM
    resultados_*****_$$$ .nptb_distribuidor
),
ingestiones AS (
SELECT
    ingestion_year ,
    ingestion_month ,
    ingestion_day
FROM
    resultados_%%&&&&&&$.catalogo_sucursales
GROUP BY
    year ,
    ingestion_month ,
    ingestion_day
ORDER BY
    year DESC,
    ingestion_month DESC,
    ingestion_day DESC
LIMIT
    1;

```

Anexo 2.

Diccionario de “parámetros” (python) que se cambian de forma manual mensualmente antes de iniciar la construcción del Informe de Efectividad y el histórico de clientes preaprobados.

```

{
    "token": true ,
    "mes": "mes_anterior" ,
    "path_insumos": "C:/Users/user_name/Folder" ,
    "user": "*****" ,
    "code": "*****" ,
    "zona": "zona_de_la_LZ_donde_se_almacenan_las_tablas" ,
    "preferencial": {
        "preferencial": "con_o_sin_preferencial" ,
        "zona": "donde_se_almacenan_las_tablas_(LZ)"
    },
    "SQLParams": {
        "cdts": {
            "Param_year": "año_actual" ,
            "Param_month": "mes_anterior" ,
            "zona": "donde_se_almacenan_las_tablas_(LZ)"
        },
        "medicion": {
            "Param_year": "año_actual" ,
            "Param_month": "mes_anterior" ,
            "zona": "donde_se_almacenan_las_tablas_(LZ)"
        },
        "query_potencial": {

```

```

        "Param_year": "año_actual",
        "Param_month": "mes_anterior",
        "Para_name_month": "nombre_mes_anterior",
        "campo_year": "ultimos_digitos_mes_anterior",
        "zona": "donde_se_almacenan_las_tablas_(LZ)"
    },
    "preaprobados": {
        "Param_year": "año_actual",
        "Param_month": "mes_anterior",
        "Para_name_month": "nombre_mes_anterior",
        "campo_year": "ultimos_digitos_mes_anterior",
        "zona": "donde_se_almacenan_las_tablas_(LZ)"
    },
    "union": {
        "Para_name_month": "nombre_mes_anterior",
        "zona": "donde_se_almacenan_las_tablas_(LZ)"
    }
}
}

```

### Anexo 3.

Conexión con los servidores del banco y resumen de las funciones del orquestador (python).

```

## CONEXION CON IMPALA
import sys
import os
import os.path
sys.path.append("C/Herramientas_comunidad/aljibe_helper")
from helper.helper import Helper
cache = {
    "connStr": "DSN=impala-prod",
    "verbose": True
}
hp = helper(cache)

## PAQUETES UTILIZADOS
import io
import json
from os import get_ipython
from collection import OrderedDict
from glob import glob
import pandas as pd
import sys
from pyfiglet import figlet_format

"""
load_config: funcion que abre el archivo de configuraciones
(parametros)

Args:

```

```

path: ruta donde esta el archivo

Return:
Parametros en un Dict
"""

"""
exe_queries: funcion que ejecuta los archivos SQL

Args:
path: ruta de ejecucion de la rutina
SQLParams: Dict de parametros a formatear y a ejecutar
sqls: lista de archivos a ejecutar
ih: instancia del impala helper para ejecutar las consultas

Return:
no retorna ninguna variable
"""

"""
save_file_exe: funcion que formatea las variables de los
SQL's y guarda los archivos ejecutables

Args:
path: ruta de la rutina
file: string de archivo a formatear y guardar
SQLParams: Dict de parametros a formatear

Return:
file_exe: archivo formateado con su ruta de ejecucion
"""

"""
get_files: funcion que busca en la carpeta de scripts , todos
los archivos terminados en .sql

Args:
path: ruta donde se alojan los scripts sql

Return:
files: lista de archivos encontrados
"""

"""
se declaran las rutas de los insumos a utilizar para esta
parte del informe , que inicia ejecutando los queries en sql
"""

"""

```

```

el orden de ejecucion de los queries es:
-cdts.sql
-medicion.sql
-query_potencial.sql
"""

"""
main: funcion principal que ejecuta las demas
"""

```

#### Anexo 4.

Ejemplos de declaraciones que consultan (SQL) la información correspondiente a la construcción del histórico de clientes preaprobados.

```

DROP TABLE IF EXISTS {zona }.distribuidor_acum PURGE;

CREATE TABLE {zona }.distribuidor_acum STORED AS PARQUET
AS WITH fecha AS (
SELECT
    year AS year_max
FROM
    resultados_*****_clientes.distribuidor_acum
WHERE
    year = {Param_year}
GROUP BY
    year
ORDER BY
    year DESC
LIMIT
    1
),
dis_ivc_1 AS (
SELECT
    t1.num_doc ,
    t1.tipo_doc ,
    t1.nivel_1 ,
    t1.nivel_2 ,
    t1.segm ,
    t1.cluster_persona ,
    t1.tipo_can ,
    t1.cod_accion_ivc_1 AS cod_accion ,
    t1.month AS mes
FROM
    result_**_clientes.distribuidor_acum AS t1
INNER JOIN fecha ON t1.year = year_max
WHERE
    t1.cod_accion_ivc_1 IS NOT NULL
),
dis_ivc_2 AS (
SELECT

```

```

        t1.num_doc ,
        t1.tipo_doc ,
        t1.nivel_1 ,
        t1.nivel_2 ,
        t1.segm ,
        t1.cluster_persona ,
        t1.tipo_can ,
        t1.cod_accion_ivc_2 AS cod_accion ,
        t1.month AS mes
FROM
    resultados_*****_clientes.nptb_acum AS t1
WHERE
    t1.cod_accion_ivc_2 IS NOT NULL
),
final AS (
    SELECT
        *
    FROM
        dis_ivc_1
    UNION ALL
    SELECT
        *
    FROM
        dis_ivc_2
)
SELECT
    *
FROM
    final
WHERE
    (
        (
            CAST(cod_accion AS INT) NOT BETWEEN ## AND ##
        ) AND
        (
            CAST(cod_accion AS INT) NOT BETWEEN ## AND ##
        )
    );

```

Anexo 5.

Ejemplos de consultas (SQL) usadas para traer la información de los clientes potenciales y los productos bancarios desde las tablas maestras alojadas en la LZ.

```

DROP TABLE IF EXISTS {zona }.informe_###_inversiones PURGE;

CREATE TABLE {zona }.informe_fecha_###_inversiones STORED AS
PARQUET AS WITH dates AS (
SELECT

```

```

        year AS year_max ,
        ingestion_month AS month_max ,
        ingestion_day AS day_max
FROM
    resultados_*****_depositos.####
WHERE
    (
        year = extract(NOW(), "year")
    AND
        month = extract(NOW(), "month")
    ) OR (
        year = extract(NOW() - INTERVAL 1 MONTH, "year")
    AND
        month = extract(NOW() - INTERVAL 1 MONTH, "month")
    ) OR (
        day = extract(NOW(), "day")
    AND
        day = extract(NOW() - INTERVAL 1 DAY, "day")
    )
GROUP BY
    year ,
    ingestion_month ,
    ingestion_day
ORDER BY
    year DESC,
    ingestion_month DESC,
    ingestion_day DESC
LIMIT
    1;

DROP TABLE IF EXISTS {zona}.*****_inversiones_depost PURGE;

CREATE TABLE {zona}.*****_inversiones_depositos STORED AS
PARQUET AS
SELECT
    numid ,
    bank_suc ,
    cod_suc ,
    amount_cd ,
    t1.ingestion_year ,
    t1.ingestion_month ,
    t1.ingestion_day
FROM
    tabla_maestra.**** $$$ijij AS t1
INNER JOIN {zona}.informe_fecha_###_inversiones AS t2
WHERE
    t1.year = t2.ingestion_year
    AND t1.ingestion_month = t2.ingestion_month
    AND t1.ingestion_day = t2.ingestion_day

```

## Bibliografía

- Real Python (2020). Core Python Language Tutorials.  
<https://realpython.com/tutorials/python/>
- Wikipedia, la enciclopedia libre (2021). Cloudera Impala.  
[https://es.wikipedia.org/wiki/Cloudera\\_Impala](https://es.wikipedia.org/wiki/Cloudera_Impala)
- Spyder (2021). The Scientific Python Development Environment.  
<https://www.spyder-ide.org/>
- Visual Studio Code (2021). Overview.  
<https://code.visualstudio.com/docs>
- Microsoft (2021). ¿Qué es Power BI?.  
<https://powerbi.microsoft.com/es-es/what-is-power-bi/>
- W<sup>3</sup> schools (2022). SQL Tutorial.  
<https://www.w3schools.com/sql/default.asp>
- Khan Academy (2022). Introducción a SQL: consulta y gestión de los datos.  
<https://es.khanacademy.org/computing/computer-programming/sql>
- DataCamp (2021). Introduction to SQL.  
<https://app.datacamp.com/learn/courses/introduction-to-sql>
- Coursera (2022). Introduction to Structured Query Language (SQL).  
<https://www.coursera.org/learn/intro-sql>
- Kaggle (2022). Pandas.  
<https://www.kaggle.com/learn/pandas>