

Monetización de Datos de Proveedores Marketplace

Autor

Jhoon Fredy Ramírez Saa

Director

Elio H. Cables Pérez, Ph.D

Universidad Antonio Nariño
Facultad de Ingeniería de Sistemas
Especialización en Gobierno de Datos
Bogotá D.C

2021

Tabla de Contenido

Listado de Figuras..... 4

Listado de Tablas..... 6

Resumen..... 7

1. Introducción 1

2. Formulación y Descripción del Problema..... 3

 2.1 Objetivo General..... 4

 2.2 Objetivos Específicos..... 4

3. Marco Referencial 6

 3.1 Marco Teórico 6

 3.2 Estado del Arte 11

 3.3 Impacto..... 14

 3.4 Componente de Innovación..... 15

4. Metodología 16

 4.1 Recolección de Información 17

 4.2 Definición de Lineamientos de Gobierno de Datos..... 17

 4.3 Transformación y Almacenamiento de datos..... 18

 4.4 Definición de Planes 19

 4.5 Visualización..... 19

5. Desarrollo de la Propuesta..... 20

 5.1 Recolección de Información 20

 5.2 Definición de Lineamientos de Gobierno de Datos..... 24

 5.2.1 Arquitectura Lógica 24

 5.2.2 Arquitectura Física 25

 5.2.3 Flujo de Datos..... 26

 5.2.4 Lineamientos de Gobierno..... 27

 5.3 Transformación y Almacenamiento de datos..... 28

5.3.1 Desarrollo de los procesos de extracción, transformación y carga	32
5.4 Definición de Planes	40
5.5 Visualización e Implementación.....	42
5.5.1 Plan Free, Dashboard	45
5.5.2 Plan de Pago, Dashboard.....	47
5.5.3 Tablero de Control y Calidad.....	51
6. Conclusiones.....	54
7. Referencias	56

Listado de Figuras

Figura 1. Página Principal Diario El Tiempo.com	12
Figura 2. Acceso a Noticia El Tiempo.com.....	13
Figura 3. Noticia Catalogada Premium.....	13
Figura 4. Información Premium Oculta sin Suscripción.....	14
Figura 5. Metodología – Flujo de trabajo.....	16
Figura 6. Flujo general de un proceso ETL.	18
Figura 7. Arquitectura Lógica de la Solución.....	25
Figura 8. Arquitectura Física de la Solución.....	26
Figura 9. Flujo de Datos.....	27
Figura 10. Proyecto creado en CGP.	33
Figura 11. Creación de Cuenta de Servicio CGP.....	33
Figura 12. Creación del conjunto de Datos.	34
Figura 13 Creación Tabla Ventas / Devoluciones Data Lake.	34
Figura 14. Tablas Creadas en BigQuery GCP.	35
Figura 15. Creación Notebook e Instalación de Librerías.....	35
Figura 16. Lectura Archivos Fuentes en Python.	36
Figura 17. Conexiones a GCP.	36
Figura 18. Carga de datos en Data Set y Tablas BigQuery.	37
Figura 19. Tablas Creadas para Manejo de Errores.	38
Figura 20. Creación del Procedimiento Almacenado.	38
Figura 21. Ejecución del Procedimiento Almacenado.	39
Figura 22. Tabla Ventas del DW (Datawarehouse) cargada después de ejecutar el Procedimiento Almacenado	39
Figura 23. Tabla Errores de Negocio cargada después de ejecutar el SP.....	40
Figura 24. Script de Creación Modelo Asociativo Qlik Sense.	43
Figura 25. Carga de Datos en Qlik Sense.....	43
Figura 26. Modelo Asociativo Qlik Sense.....	44
Figura 27. Archivo de Muestra de Configuración Accesos.....	44
Figura 28. Código de Perfilamiento de Datos en Qlik Sense.	45
Figura 29. Visualización Dashboard Plan Free.	46
Figura 30. Filtros y Barras Unidades x Hora.	46

Figura 31. Gráfico de Tarta y Vista de Datos Detallada.	46
Figura 32. Gráfico de Barras Unidades x Semana e Indicador % Devoluciones.....	47
Figura 33. Plan Premium Panel de Filtros	48
Figura 34. Plan Premium Panel Principal.....	49
Figura 35. Plan Premium Segunda Vista.	49
Figura 36. Plan Premium Vista de Detalle.....	50
Figura 37. Plan Premium Panel de Indicadores.	50
Figura 38. Plan Premium Top 10 Mas Vendidos y más Devueltos.	50
Figura 39. Plan Premium Vetas por Hora, y Participación por Canal.....	51
Figura 40. Plan Premium Bloques por Medio de Pago y Ticket Promedio.....	51
Figura 41. Dashboard Principal Tablero de Control.	52
Figura 42. Graficas Tablero de Control.	52
Figura 43. Detalle Tablero de Control.	53

Listado de Tablas

Tabla 1 Estructura Archivo de Ventas y Devoluciones.....	20
Tabla 2 Estructura Archivo Almacenes	21
Tabla 3 Estructura Archivo Medios de Pago	21
Tabla 4 Estructura Archivo Proveedores.....	21
Tabla 5 Estructura Archivo Productos.....	21
Tabla 6 Estructura Archivo Configuración Proveedor / Plan	22
Tabla 7 Ficha Técnica Levantamiento Set de Datos.....	22
Tabla 8 Estructura Tabla Ventas / Devoluciones Bronce	29
Tabla 9 Estructura Tabla Almacenes Bronce	29
Tabla 10 Estructura Tabla Medios de Pago Bronce.....	30
Tabla 11 Estructura Tabla Proveedores Bronce	30
Tabla 12 Estructura Tabla Productos Bronce.....	30
Tabla 13 Estructura Tabla Configuración Proveedor / Plan Bronce	30
Tabla 14 Estructura Tabla Ventas / Devoluciones Gold.....	31
Tabla 15 Estructura Tabla Errores Técnicos.....	31
Tabla 16 Estructura Tabla Errores de Negocio	32
Tabla 17 Estructura Tabla Configuración Proveedor / Plan Gold.....	32

Resumen

La información es un activo de alto valor y con el desarrollo de este proyecto se busca ofrecer una solución que apoye la toma de decisiones de aquellos proveedores Marketplace y emprendimientos que generan ventas a través de plataformas Marketplace.

Sobre la base de los datos transaccionales entregados por una empresa del sector Retail, se realizó todo un análisis, consolidación e implementación de una solución de datos, que integra varias tecnologías, métodos y buenas prácticas, que al final brinda como resultado visualizaciones presentadas en Dashboard con diferentes perspectivas de análisis, que permiten al usuario final tomar decisiones basadas en datos, para mejorar sus procesos de ventas e inventarios y que al final repercute en un mejor servicio hacia los compradores.

Con la implementación de este producto la empresa dueña de la información y los proveedores que hacen uso de sus canales de ventas, podrán generar ganancias, es decir, que entre más ventas tenga un proveedor, más de su información será analizada y categorizada en insights en el procesamiento de datos y expuesta en Dashboard, para que a través de estos, pueda tomar las mejores decisiones, cada visualización permite, el análisis general y específico si así se quiere de datos relevantes de su operación.

1. Introducción

La gran mayoría de las veces, el éxito comercial de las empresas, refiriéndose en torno a las ventas y utilidades netas, depende de la percepción y reconocimiento de la marca que tengan los clientes hacia ella o alguno de sus productos, esto no es malo, al contrario, que una empresa sea exitosa financieramente es un gran logro, pero con el crecimiento acelerado de datos y herramientas tecnológicas que los explotan, ¿están las empresas pequeñas, grandes y medianas tomando decisiones basadas en datos?, posiblemente sí, pero hasta qué punto o hasta que nivel.

El auge de lo digital, sumado a los efectos de la pandemia del COVID 19 han ayudado a que las ventas por internet, el uso de e-commerce (comercio electrónico) y plataformas, crezcan los datos de forma exponencial, las organizaciones han robustecido sus canales y arquitecturas virtuales y muchos pequeños empresarios han optado por aliarse con las grandes marcas para vender sus productos, lo que les genera más ingresos.

Con el desarrollo de esta solución, dirigida especialmente a empresas del sector Retail con presencia en tiendas físicas y/o digital, se desea complementar esos procesos de ventas para generar muchos más ingresos, ya que la venta no debe terminar con la compra del producto por parte del cliente, al contrario, es el comienzo de una serie de análisis que ayuden a comprender los datos en crudo, poder entender hábitos de compra de los clientes, efectividad en temporadas, estacionalidades del año que ayuden generar mucho más valor y toma de decisiones acertadas. Dicho lo anterior, se propone

desarrollar procesos de extracción, transformación y carga de datos que garanticen datos de alta calidad y que apoyado con el uso de herramientas tecnológicas como BigQuery de GCP para el procesamiento de datos y Qlik Sense para la visualización y análisis de información, proporcionen una visión 360 de lo que sucede con las ventas de los productos. Esta solución está pensada para que todos esos pequeños empresarios y emprendedores que se asocian con las grandes marcas, puedan sacar provecho de sus propios datos de ventas a cambio de un pago por suscripción o plan que al final beneficiara a ambas partes, por lo cual la fuente de datos principal de esta solución y objeto de análisis, son los datos transaccionales de ventas.

2. Formulación y Descripción del Problema

El crecimiento acelerado de tecnologías y aplicaciones, sumado al efecto global que causó la pandemia en el año 2020, ha llevado a muchos proveedores donde la mayoría son Pymes, personas naturales o emprendimientos a no solo ofrecer sus productos y servicios de manera presencial en tiendas o locales físicos, sino a, apoyarse en las grandes plataformas digitales o Ecommerce como Amazon, Mercado Libre, Dafiti, etc. (Ruiz, Villegas, & Hernández, 2012), y también en las grandes cadenas o marcas reconocidas del país para ofrecer y vender sus productos. Esto les representa mayor visibilidad de sus artículos y mayores ingresos, debido a que aprovechan la infraestructura (a nivel tecnológico), la marca (a nivel comercial) y todas las campañas de marketing de esas plataformas para generar confianza y transparencia en sus futuros clientes. A cambio de esto, la compañía dueña de la plataforma o de la infraestructura Marketplace gana una comisión por venta de cada artículo, algunas compañías no solo ofrecen la venta por el Ecommerce (Comercio Electrónico), también pueden ofrecer el despacho, almacenamiento, asesoramiento en diversos temas jurídicos, contables, etc.

Muchos de estos proveedores cuando empiezan a comercializar sus productos por estas herramientas, no tienen la cultura, el inventario o los medios tecnológicos para medir sus ventas y transacciones, cambios y garantías, o lo hacen de forma muy empírica que no les permite tomar decisiones basadas en datos, esto al final, puede convertirse en un problema de no saber, el stock de algún producto, cuándo un producto está generando pérdidas por cambios o defectos, si está entregando a tiempo, así como muchas cosas más que conllevan a un desempeño inadecuado.

Si estos proveedores pudieran acceder a un análisis descriptivo de sus propios datos y/o ventas, quizás podrían mejorar sus estrategias de ventas, y estas plataformas a su vez tendrían un mejor retorno de inversión de sus servicios.

A partir de lo explicado anteriormente se puede identificar el siguiente problema:

¿Cómo aumentar los ingresos de las compañías con plataformas que comercializan a través de Marketplace, a partir de los datos de ventas que generan los proveedores suscritos a este modelo de negocio?

2.1 Objetivo General

Desarrollar una plataforma para las compañías que comercializan a través de Marketplace, tal que a partir de los datos de ventas permita a los Proveedores de Marketplace, tomar de decisiones que los ayude al incremento de sus ventas y que esta a su vez represente ingresos a las compañías.

2.2 Objetivos Específicos

- Caracterizar los modelos existentes para la monetización a partir de los datos.
- Gestionar la información bajo estándares y tecnologías que garanticen la consistencia, confiabilidad y seguridad de la información de ventas de los proveedores de Marketplace.
- Establecer los indicadores o métricas que generen valor a los proveedores.
- Generar los diferentes análisis de datos según los planes de monetización definidos.

- Definir los planes de comercialización del producto de monetización al que van a acceder los proveedores.
- Desarrollar los diferentes Dashboard que permita visualizar la información según el plan contratado por cada proveedor.

3. Marco Referencial

3.1 Marco Teórico

Con el auge de la transformación digital en las organizaciones y el crecimiento de los datos generados por los diferentes medios, resulta importante darle un uso adecuado a la información, aprovechando que las grandes marcas tecnológicas como Google, AWS, Microsoft y Oracle, disponen de gran variedad de herramientas para explotar y monetizar estos datos. Según Patricia, K. (Patricia, 2020) todo lo que abarque contenidos digitales deben proporcionar al usuario final 4 condiciones mínimas que son:

- Conexión o disponibilidad del sitio se debe garantizar en cualquier momento.
- Cualquier persona debe poder acceder al contenido.
- Desde Cualquier lugar.
- Desde Cualquier equipo o dispositivo tecnológico se debe poder acceder.

Implementar un modelo de monetización de datos bien sea en la nube o localmente en la organización, debe garantizarle al usuario final estas 4 características mencionadas anteriormente.

Según Alberto. (Alberto, 2015) cada vez más las organizaciones se basan en datos, el Big Data y el Driven Marketing así lo demuestran, y es por esto que las empresas que recopilan y/o almacenan estos datos deben mejorar las políticas del uso de datos apoyados en la transparencia y la protección de datos personales como parte primordial de sus integraciones y análisis.

Es de conocimiento de cada proveedor y usuario que establecen un contrato comercial con alguna marca o empresa, que, por el solo hecho de tener un contrato

firmado, su información puede ser utilizada para fines comerciales y es responsabilidad de las organizaciones darle el buen uso.

Según Alberto. (Alberto, 2015), Amazon ha sido el pionero en la creación de modelos de monetización de datos, aprovecha las capacidades del Big Data y utilizando datos de comportamiento de sus clientes dentro de sus plataformas, realiza recomendaciones de las mejores ofertas de acuerdo con el hábito de consumo o de búsqueda de cada usuario. En poco tiempo Amazon se convirtió en la más grande cadena de Retail, su idea de negocio apoyada en el manejo de datos y CRM les proporcionaron gran competitividad frente a las otras marcas.

Para Palma, D. (GrupoDot), partners principal de Google en Colombia, el éxito de monetizar los datos en una organización requiere de varios pasos y desafíos (Palma, 2021).

- Organizar un equipo dedicado, cuya misión es analizar, extraer y generar ideas de cómo usar los datos.
- Realizar un inventario de datos, donde se validan fuentes disponibles, internas y externas, y que pueden ser utilizados de acuerdo con las leyes de protección de datos personales.
- Seleccionar el tipo de monetización al que se quiera llegar.
- Buscar en otros mercados y organizaciones casos de éxito que ayuden a comprender los desafíos.
- Desarrollar y probar prototipos, ayuda a hacer juicios de valor para descartar ideas no acordes con el objetivo general.

- ETL, una vez se consideren los datos necesarios y los prototipos adecuados se procesó al manejo, limpieza y almacenamiento de los datos para realizar los respectivos análisis.

Basado en lo anterior, el solo hecho de tener los datos no es suficiente, es de gran importancia conocer el contexto de estos y de las regulaciones legales que estos tienen; una vez se tenga esto, el uso adecuado de los equipos de trabajo (las personas) y tecnologías podrán llevar al éxito o fracaso de los proyectos.

Según Andújar son tres los modelos más utilizados y extendidos para obtener ingresos y beneficios en negocios digitales (Andujar, 2018).

- Modelo de Suscripción: Se concreta a partir de un pago periódico que el usuario o cliente realiza, a cambio de poder disfrutar de un producto o servicio sin la necesidad de tener una membresía o licencia de forma completa, el mayor beneficio de este modelo, es que se puede analizar la evolución del sistema casi que en tiempo real, y de esta manera saber si los usuarios están obteniendo provecho del sistema y están dispuestos a seguir pagando por ello o no.
- Modelo basado en Publicidad: Es la opción más utilizada en estos momentos. Consiste en aprovechar el contenido de blogs y páginas que causan gran interés del público, para vender parte de su espacio y poner banners u otro tipo de contenido y esto representa la principal fuente de ingresos del proyecto. Al final los usuarios acceden a estos blogs, se nutren de su contenido y a su vez están aceptando la visualización de publicidad.

- Modelo Freemium: Este tipo de modelo, se basa en acceder de forma gratuita a un producto o servicio, pero de forma limitada, la idea principal es cautivar al usuario con la propuesta y si este desea mayores funcionalidades, debe cancelar un pago de acuerdo a una tabla de tarifas predeterminadas y al final termina siendo similar a un modelo de suscripción, un gran ejemplo es Dropbox.

Pero definitivamente existe un actor muy importante en el tema de la monetización de datos y contenidos y son las redes sociales, particularmente, en el caso de YouTube, según Patricia (Patricia, 2020) se pueden generar ingresos con los contenidos de 4 formas.

- Ingresos por Publicidad, en anuncios gráficos y videos.
- Membresías del Canal, en donde los suscritos realizan aportes periódicos a cambio de servicios especiales que ofrece el creador del contenido.
- Biblioteca de Productos: Los usuarios pueden revisar y pagar por las promociones oficiales que aparecen en las reproducciones de videos.
- Super Chat y Calcomanías, son pagos que realizan los usuarios para que sus mensajes aparezcan y destaquen en los contenidos.

Son muchas las tecnologías y herramientas en el mercado que pueden garantizar el éxito de un proyecto de este tipo, por ejemplo, Qlik Sense es una plataforma potenciada de Qlik View, está pensada para el análisis y descubrimiento de datos. Tiene un motor de análisis asociativo muy potente, inteligencia artificial sofisticada, su sistema y arquitectura puede residir On Premise o en la nube, su conexión con múltiples orígenes

y tecnologías, velocidad de procesamiento y análisis, brinda a las organizaciones la capacidad de tomar mejores decisiones, creando una empresa verdaderamente basada en datos (Qlik Sense, 2021).

Para Power BI, la gran ventaja de pertenecer a la suite de productos y tecnologías de Microsoft le da muy buenas opciones de hacer parte de proyectos de este tipo, es una solución de análisis empresarial basado en la nube, se pueden integrar múltiples fuentes de información, lo que facilita el análisis a través de informes y paneles (Power BI, 2021). Con Power BI es posible acceder a los datos dentro y fuera de la organización, es web responsive y tiene la ventaja de contar con todo el respaldo y acceso a los datos de Azure.

En el caso de Google proporciona toda una suite de servicios en la nube, por ejemplo, Data Studio (Data Studio, 2021), es una herramienta de visualización para crear informes de datos, analizar y disponibilizar las métricas de una forma sencilla a través de Dashboard dinámicos. El acceso a esta herramienta es gratis y se puede acceder solo teniendo una cuenta de Google. Para los demás servicios de Google Cloud (Google Cloud, 2021) la plataforma ofrece servicios de almacenamiento de datos en Cloud Storage o BigQuery, los Costos BigQuery de almacenamiento son asequibles y tienen un mayor valor de procesamiento, pero puede tener retorno de inversión con la cantidad de usuarios o proveedores que estén dispuestos a pagar por el producto final (Costos BigQuery, 2021).

Para el caso del análisis inicial exploratorio de datos, limpieza y transformación de datos, es muy óptimo trabajar con Python; pues, según Visus (Visus, 2020) Python es ampliamente utilizado por organizaciones de todo el mundo para desarrollar

aplicaciones, especialmente para analizar datos, e industrializar operaciones empresariales, su fiabilidad y escalabilidad ha impulsado a que grandes empresas tecnológicas utilicen este lenguaje de programación y basen su arquitectura tecnológica en él.

3.2 Estado del Arte

Son varias las compañías y/o plataformas que ofrecen diferentes servicios de monetización a través de datos, por ejemplo, en la Revista Compensar (Compensar, 2021), en Julio de 2021 publico un artículo donde menciona y detalla 4 plataformas vigentes para monetizar contenidos, blogs y fans, estas empresas son:

- **Patreon:** Es un sitio web de financiación colectiva, consiste en recaudar dinero a través de una suscripción mensual, en la cual los usuarios, en este caso las personas que realizan el pago reciben contenido exclusivo acerca de temas de su interés.
- **Buy Me a Coffee:** Cualquier usuario que cuente con una comunidad de seguidores amplia, puede sacar bastante beneficio de esta plataforma, ya que permite que sus seguidores realicen donaciones a través de internet utilizando el enlace buymeacoffee.com.
- **Kickstarter:** Esta herramienta está orientada para artistas o personas con bastante creatividad, en este caso el creador diseña contenidos a las necesidades de sus patrocinadores y estos a su vez pagan por esa publicidad.

- Hotmart: Es una herramienta también orientada para crear contenidos, en esta, el creador diseña su contenido en determinado tema (cursos online, e-book, software, podcast, scripts, entre otros) y lo vende. Para ello se debe tener una cuenta en Hotmart, y registrar el producto o contenido.

También es conocido que las grandes plataformas y medios de comunicación aprovechan la era digital para monetizar sus investigaciones y noticias, por ejemplo, el periódico el Tiempo de Bogotá (El Tiempo, 2021), ofrece a los lectores en su página de internet acceso limitado a algunas noticias, pero los grandes titulares y noticias de gran impacto no es posible leerlo o acceder sin antes suscribirse a algún plan de pago, lo mismo ocurre con cada una de sus otras dependencias o medios digitales como la revista Carrusel, Portafolio, Bocas, Revista Alò. A continuación, se validan algunos artículos de este diario informativo y se evidencia la forma en que monetizan y generan ingresos a través de la información que investigan, procesan y almacenan.

Al acceder a la página principal del diario El Tiempo.com, se observa que de forma muy sutil inducen al usuario a suscribirse por un valor muy bajo y atractivo, (ver figura 1).

Figura 1. Página Principal Diario El Tiempo.com



Fuente: Tomado de la Pagina del diario El Tiempo <https://www.eltiempo.com/> (El Tiempo, 2021)

Se puede acceder al contenido de noticias básicas o de poca relevancia, pero el contenido completo de los artículos no es posible verlo e invitan a suscribirse, y realizar un pago para ver la información completa y otra serie de beneficios, (ver figura 2).

Figura 2. Acceso a Noticia El Tiempo.com



Fuente: Tomado de la Pagina del diario El Tiempo <https://www.eltiempo.com/> (El Tiempo, 2021)

Algunas noticias son catalogadas como Premium, la mayoría son noticias de gran impacto y mediáticas que despiertan curiosidad en el usuario final y termina pagando la suscripción, (ver figura 3).

Figura 3. Noticia Catalogada Premium.



Fuente: Tomado de la Pagina del diario El Tiempo <https://www.eltiempo.com/> (El Tiempo, 2021)

Cuando un usuario ingresa al titular sin pagar la suscripción, no puede ver el contenido, (ver figura 4).

Figura 4. Información Premium Oculta sin Suscripción.



Fuente: Tomado de la Pagina del diario El Tiempo <https://www.eltiempo.com/> (El Tiempo, 2021)

De esta forma, se puede evidenciar como las grandes empresas que poseen y almacenan la información, pueden de una u otra forma generar ingresos a partir de esos datos.

3.3 Impacto

El presente trabajo busca contribuir a la buena gestión y análisis de información de ventas que requieren los proveedores de tipo Marketplace, para mejorar la

comercialización de sus productos, disponibilidad de estos, lo que conlleva a diseñar mejores estrategias de ventas, catálogo de productos y mejores ingresos, adicionalmente al aumentar los ingresos de los proveedores por sus ventas, se beneficia e impacta también la empresa dueña de la información que implementa la solución.

3.4 Componente de Innovación

Los datos bien explotados son el futuro de las compañías; Este proyecto busca obtener valor de los datos en crudo de la organización donde se implemente, utilizando técnicas de integración y manejo de datos conocidas en el medio, pero garantizando el buen uso y aplicación de tecnologías aptas para el procesamiento y análisis de datos, generando propuesta de valor al proveedor final que requiere la información y a la empresa que la posee y la explota.

4. Metodología

En el entorno de las TIC (Tecnologías de la Información y Comunicación) existen metodologías para enfrentar diferentes procesos, por ejemplo, en el campo del desarrollo de software existen las metodologías ágiles, pesadas (cascada, de prototipo, en espiral, etc.) que establecen el paso a paso que se debe seguir para obtener el resultado. Sin embargo, en el contexto del manejo y gobierno de datos no se muestran metodologías formalizadas con determinados nombres, pero si existe un framework como el (DAMA International, 2017) Dama-dmbook 2nd Edition, el cual marca las pautas, buenas prácticas y referencias para llevar a cabo un proyecto de datos, por lo cual tomando con referencia algunas de sus 11 áreas de conocimiento, se propone el siguiente flujo de trabajo (ver figura 5).

Figura 5. Metodología – Flujo de trabajo.



Fuente: Elaboración Propia.

A continuación, se describen cada una de las fases:

4.1 Recolección de Información

En esta etapa se procede a definir, capturar y/o recolectar la información que será parte del análisis de datos propuesto en los objetivos específicos del proyecto, partiendo como base fundamental los orígenes de información de ventas y proveedores de la organización dueña de la plataforma, adicionando datos de referencia que ayuden a complementar el modelo.

4.2 Definición de Lineamientos de Gobierno de Datos

El concepto de gobierno de datos es suficientemente amplio, sin embargo, modelos de referencia de gestión de datos como el DAMA, priorizan dentro de un plan de gobierno de datos que este alineado a las estrategias organizacionales cuatro requisitos mínimos a tener en cuenta, estos son.

- Definición de Objetivos.
- Principios y políticas.
- Procesos relacionados con la operación de los datos.
- Reglas de negocio.

Basado en los lineamientos y aprendizaje obtenido durante el periodo académico de esta especialización, y teniendo como referencia el (DAMA International, 2017) Damadmbok 2nd Edition en su capítulo 3 Data Governance, se crearon las políticas y estándares de almacenamiento para garantizar la consistencia y veracidad de los datos, así mismo la arquitectura lógica y física del modelo.

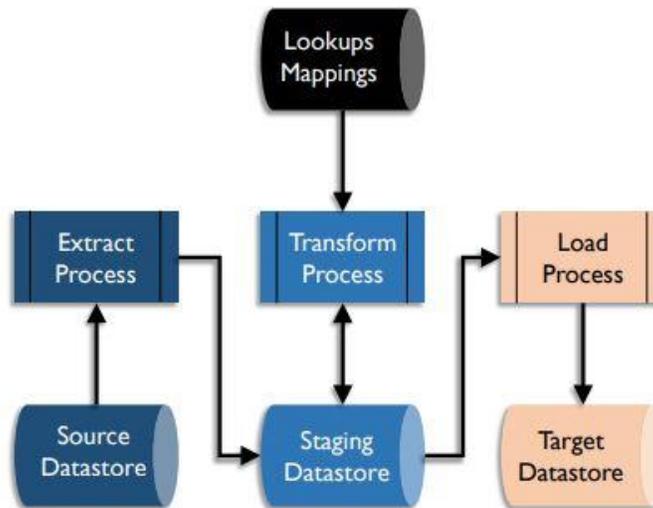
4.3 Transformación y Almacenamiento de datos

Cubiertas las etapas anteriores, se procede a definir las estructuras de almacenamiento que soportaran el modelo de datos y los análisis de información requeridos.

Igualmente se desarrollaron los procesos ETL (del inglés Extract, Transform, Load) que alimentaran las estructuras de almacenamiento, tomando como referencia y según el (DAMA International, 2017) Dama-dmbook 2nd Edition en su capítulo 8 Data Integration and Interoperability, donde se habla de los procesos relacionados con el movimiento y consolidación de datos, buenas prácticas y conceptos esenciales a tener en cuenta en el proceso de integración de datos.

Estos procesos ETL consisten en la integración de las fuentes de información que componen el modelo de datos, aplicando una serie de transformaciones y limpieza que garantizan la calidad final de la información y su posterior carga en las tablas destino, (ver figura 6).

Figura 6. Flujo general de un proceso ETL.



Fuente: (DAMA International, 2017), DAMA-DMBOK 2nd Edition (2 ed.). (Deborah Henderson, CDMP, Ed.) Basking Ridge, NJ 07920 USA: Technics Publications.

Se utilizo como plataforma de almacenamiento GCP – BigQuery debido a su gran capacidad de procesamiento y su escalabilidad en caso de crecer el producto y enfocarse en modelos de Machine Learning, ya que esta tecnología cuenta con modelos preinstalados para dicho fin.

4.4 Definición de Planes

Sobre la base de los resultados obtenidos en la fase de transformación, almacenamiento del modelo y conociendo los datos recolectados, se definen los siguientes planes de acceso a los datos, garantizando que la información a entregar satisfaga a los clientes consumidores del modelo.

- Plan Free.
- Plan Pago.

4.5 Visualización

En esta etapa se desarrollaron los diferentes análisis, gráficos y métricas definidas para cada plan y se implementaron en dashboard dinámicos y asequibles por cada uno de los proveedores, garantizando una buena experiencia de usuario, navegación y entendimiento de los datos mostrados para la toma de las mejores decisiones.

Cada Dashboard permite filtrar la información y mostrar los datos agregados o detallados dependiendo del plan suscrito por cada proveedor, y adicional garantiza la confidencialidad de la información entre proveedores.

5. Desarrollo de la Propuesta

A partir de este capítulo se detalla la forma en que se desarrolló el flujo de trabajo propuesto, describiendo cada uno de los pasos utilizados en cada fase de la metodología.

5.1 Recolección de Información

Se consideró la utilización de datos de ventas, productos, almacenes, medios de pago, proveedores y devoluciones de una empresa del sector Retail, la cual es especialista en productos para el mejoramiento del hogar y la construcción, además, por razones de confidencialidad no se plasma el nombre de la empresa en este documento. Se definieron estructuras generales para cada entidad de datos a recolectar, la información fue entregada en archivos .csv, separados por “;” con encabezados de columnas y con la siguiente estructura (ver tablas 1 -5).

Tabla 1 Estructura Archivo de Ventas y Devoluciones

Columna	Descripción
fecha	Fecha de la transacción
cajapos	Numero de la caja o dispositivo donde se factura la venta
numtransaccion	Numero de la transacción
hora_compra	Hora de la transacción
cod_almacen	Codigo del almacén o tienda
nom_almacen	Nombre del almacén o tienda
cod_medio_pago	Codigo del medio de pago
cod_producto	Codigo del producto comprado
tipo_venta	Venta física o virtual
canal_virtual	Canal virtual por donde ingresa la venta
cliente_tipo	Tipo de Cliente
identificación	Número de Identificación del Cliente
unidades	Cantidad de unidades compradas
valor_unidad	Valor unitario

valor_total	Valor Total del producto
iva	Valor del impuesto de IVA si aplica
venta_neta	Valor neto, puede ser de venta o devolución
tipo_registro	Tipo de registro Venta o Devolución
documento_cliente	Tipo de documento del cliente

Fuente: Elaboración Propia, generada con Excel.

Tabla 2 Estructura Archivo Almacenes

Columna	Descripción
cod_almacen	Codigo del almacén o tienda
nom_almacen	Nombre del almacén o tienda
tipo	Tipo de almacén, físico o virtual
latitud_y	Coordenada de ubicación del local, Latitud
longitud_x	Coordenada de ubicación del local, Longitud

Fuente: Elaboración Propia, generada con Excel.

Tabla 3 Estructura Archivo Medios de Pago

Columna	Descripción
cod_medio_pago	Codigo del medio de pago
nom_medio_pago	Nombre del medio de pago

Fuente: Elaboración Propia, generada con Excel.

Tabla 4 Estructura Archivo Proveedores

Columna	Tipo de Dato	Descripción
prov_cod	numérico	Codigo del proveedor
razon_social	varchar	Nombre o razón social del proveedor
email	varchar	Email del Proveedor
tipo	varchar	Origen, Nacional o Importador
fecha_insercion	datetime	Fecha en que se insertó el registro

Fuente: Elaboración Propia, generada con Excel.

Tabla 5 Estructura Archivo Productos

Columna	Descripción
cod_jerarquia	Codigo de la categoría del producto
nom_jerarquia	Nombre de la categoría del producto
cod_producto	Codigo del producto
nom_producto	Nombre del producto
prov_cod	Codigo del proveedor

De igual forma se consideró un archivo de configuración que contendrá los proveedores y su respectivo plan de acceso a la plataforma (ver tabla 6).

Tabla 6 Estructura Archivo Configuración Proveedor / Plan

Columna	Descripción
prov_cod	Código del proveedor
plan_cod	Código del plan de acceso a la información
estado_plan	Origen, Nacional o Importador
fecha_inicio_vigencia	fecha inicio de la vigencia del plan
fecha_fin_vigencia	fecha inicio de la vigencia del plan
fecha_insercion	Fecha en que se insertó el registro

Fuente: Elaboración Propia, generada con Excel.

En la siguiente tabla se detalla la ficha técnica y metadatos de las fuentes de información del modelo, (ver tabla 7).

Tabla 7 Ficha Técnica Levantamiento Set de Datos

Elementos	Descripciones
Entidad	Ventas
Periodicidad de Carga	Diaria
Volumetría	Entre 1.2 y 1.5 millones registros x mes
Cantidad de columnas	20
Particiones	Mensual por columna Fecha
Base de datos Origen	Sistema Transaccional de Ventas, HCVENTASPOS
Tecnología Base de Datos	Oracle 19
Servidor Origen	Aix
Archivos	Ventas_AAAAMM.csv
Peso archivo	150 MB x Mes
Separador	;
Encabezados de Columna	Si
Entidad	Almacenes
Periodicidad de Carga	Diaria
Volumetría	5 registros
Cantidad de columnas	5
Particiones	NA

Elementos	Descripciones
Base de datos Origen	Sistema Datos Maestros, HCMAESTROS
Tecnología Base de Datos	Oracle 19
Servidor Origen	Aix
Archivos	Almacenes.csv
Peso archivo	1 KB
Separador	;
Encabezados de Columna	Si
Entidad	MediosPagos
Periodicidad de Carga	Diaria
Volumetría	58 registros
Cantidad de columnas	5
Particiones	NA
Base de datos Origen	Sistema Datos Maestros, HCMAESTROS
Tecnología Base de Datos	Oracle 19
Servidor Origen	Aix
Archivos	Almacenes.csv
Peso archivo	2 KB
Separador	;
Encabezados de Columna	Si
Entidad	Proveedores
Periodicidad de Carga	Diaria
Volumetría	6.605 registros
Cantidad de columnas	4
Particiones	NA
Base de datos Origen	Sistema Datos Maestros, HCMAESTROS
Tecnología Base de Datos	Oracle 19
Servidor Origen	Aix
Archivos	Proveedores.csv
Peso archivo	205 KB
Separador	;
Encabezados de Columna	Si
Entidad	Productos
Periodicidad de Carga	Diaria
Volumetría	433.595 registros
Cantidad de columnas	5

Elementos	Descripciones
Particiones	NA
Base de datos Origen	Sistema Datos Maestros, HCMAESTROS
Tecnología Base de Datos	Oracle 19
Servidor Origen	Aix
Archivos	Productos.csv
Peso archivo	25 MB
Separador	;
Encabezados de Columna	Si

Fuente: *Elaboración Propia, generada con Excel.*

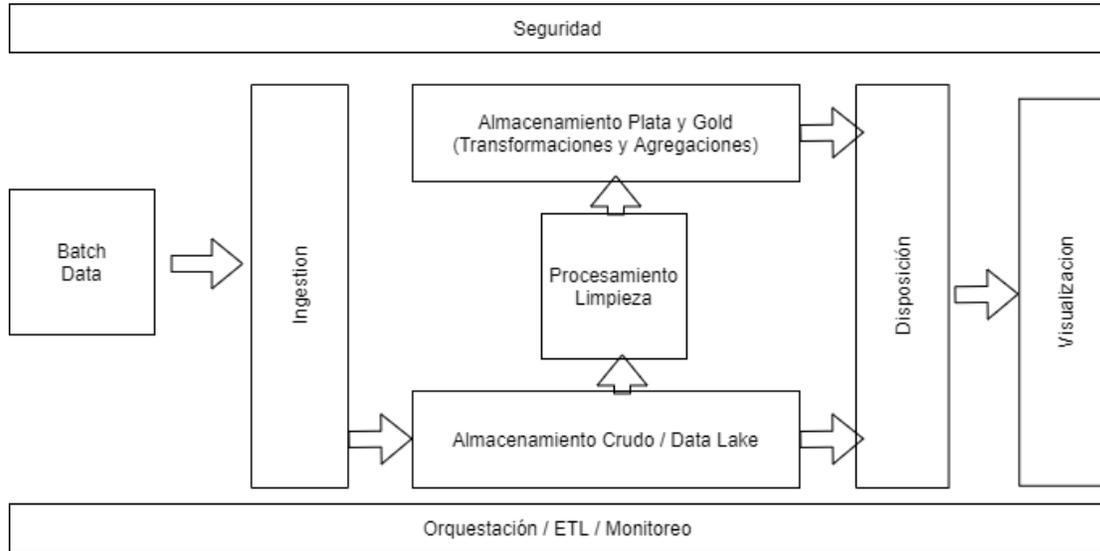
5.2 Definición de Lineamientos de Gobierno de Datos

Se define el flujo de proceso de los datos, las arquitecturas, reglas y políticas de gobierno para garantizar la consistencia de la información y el buen funcionamiento del modelo:

5.2.1 Arquitectura Lógica

A continuación, se presenta el diseño de la arquitectura lógica a utilizada en el proyecto, en el cual se destacan los procesos batch de ingestión de datos a la zona de datos crudos o bronce, proceso de transformación, limpieza y almacenamiento de datos Gold transformados listos para ser consultados en la capa de visualización, todo orquestado y monitoreado por las respectivas consolas de administración y seguridad (ver figura 7).

Figura 7. Arquitectura Lógica de la Solución.

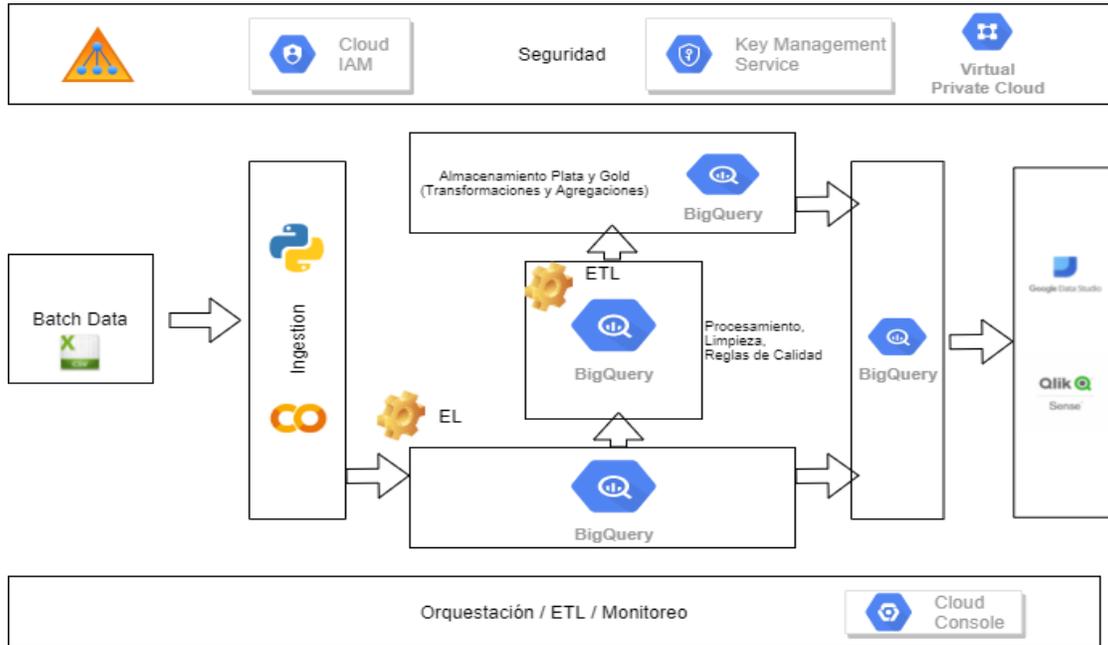


Fuente: Elaboración Propia.

5.2.2 Arquitectura Física

En la arquitectura física se trabajó con tecnologías cloud sobre la plataforma de GCP (Google Cloud Platform), haciendo uso de varios de sus servicios. Existe un proceso inicial de ingestión de datos fuentes al data lake, el cual se realiza con un Notebook de Python, de ahí en adelante los procesos de transformación, aplicación de reglas de calidad, limpieza, procesamiento y almacenamiento están a cargo de BigQuery. Todo está orquestado y monitoreado desde la consola de GCP haciendo uso de la seguridad proporcionada por esta plataforma y el directorio activo de la compañía dueña de la información (ver figura 8).

Figura 8. Arquitectura Física de la Solución.

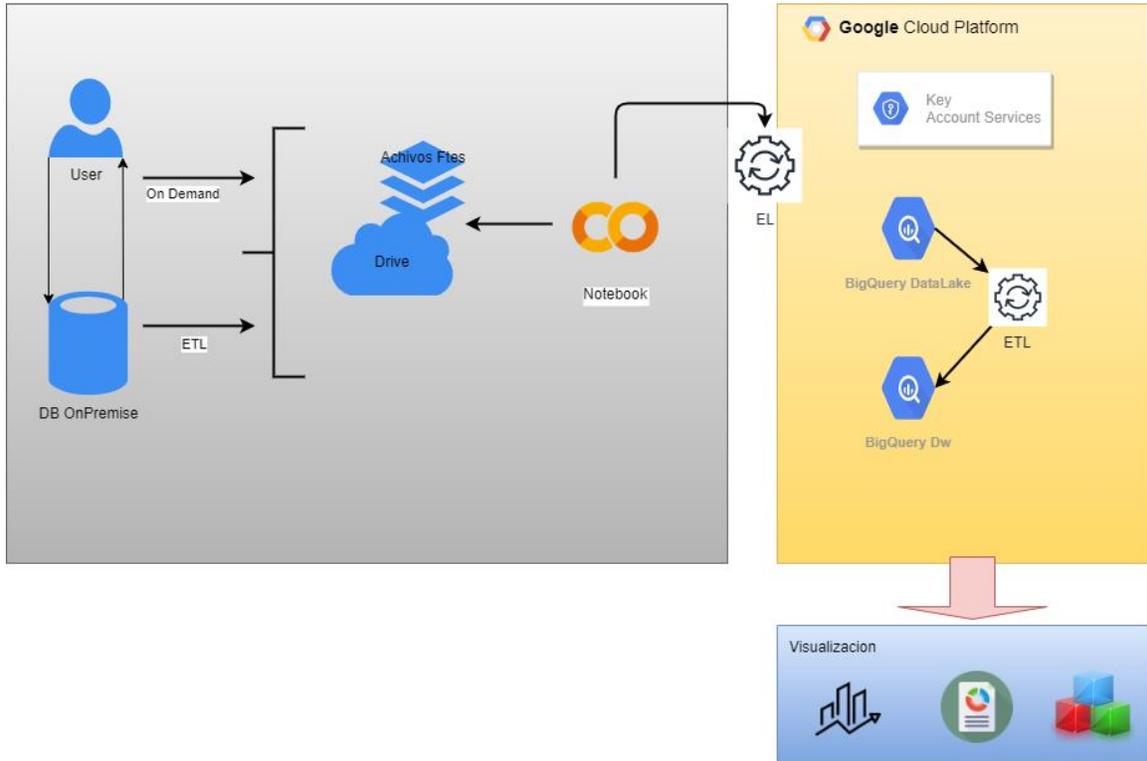


Fuente: Elaboración Propia.

5.2.3 Flujo de Datos

El siguiente flujo de datos representa un acercamiento general a la forma o ciclo de vida del dato dentro del modelo propuesto, donde se inicia con la data transaccional de ventas que tiene la compañía dueña de la plataforma Marketplace, estos archivos por medio de una ETL o a demanda, son ubicados dentro del Drive, para que un notebook con código Python cargue los datos en GCP y a partir de ahí, se realiza todo el proceso de limpieza y transformación de la información hasta dejarla lista para ser analizada por los consumidores de datos, (ver figura 9).

Figura 9. Flujo de Datos.



Fuente: Elaboración Propia.

5.2.4 Lineamientos de Gobierno

Se definen los siguientes lineamientos y directrices de gobierno de dato:

- Se debe entregar la información de las fuentes en archivos .csv, separados por “;” con encabezados de columna, con periodicidad diaria y en horario por establecer.
 - No se deben subir datos sensibles de Proveedores a la Nube, como información de pagos, números de tarjetas, números de cuentas, en caso de ser necesario se debe cifrar el dato.
 - Los Dashboard o visualizaciones no deben mostrar datos de contacto de Proveedores.

- Cada proveedor solo debe tener acceso a su propia información, es decir el modelo de datos debe configurarse de tal forma que cuando se acceda a los Dashboard solo se debe visualizar la información del usuario que ingresa.
- Los datos de la fuente que representen valores y que lleguen vacíos (nulos) se deben transformar a cero (0).
- Los datos de la fuente que representen códigos y que lleguen vacíos (nulos) se deben transformar a menos uno (-1).
- Los datos de la fuente que representen descripciones categóricas y que lleguen vacíos (nulos) se deben transformar en “Sin Identificar”.
- Los datos de la fuente que representen descripciones que no sean descripciones categóricas y que lleguen vacíos (nulos) se deben transformar en “No Aplica”.
- Los datos que representen fechas deben tener formato AAAA-MM-DD.
- Se deben guardar los logs de errores y auditorias de los procesos Bach.

5.3 Transformación y Almacenamiento de datos

El modelo cuenta con 2 capas de datos; Una capa o Data Lake donde se almacenan los datos crudos tal cual como llegan de los sistemas fuentes y una capa de datos Gold o transformados, donde se almacena la información lista para ser consumida por los respectivos análisis. En el medio de estas existe una capa stage o de trabajo donde se realizan las operaciones de limpieza y transformaciones de la información, estas estructuras fueron creadas en la plataforma de GCP y su servicio de base de datos columnar BigQuery.

A continuación, se describen las estructuras de almacenamiento de datos Bronce para cada una de las tablas utilizadas (ver tablas 8 – 13).

Tabla 8 Estructura Tabla Ventas / Devoluciones Bronce

Columna	Tipo de Dato	Descripción
fecha	varchar	Fecha de la transacción
cajapos	varchar	Numero de la caja o dispositivo donde se factura la venta
numtransaccion	varchar	Numero de la transacción
hora_compra	varchar	Hora de la transacción
cod_almacen	varchar	Codigo del almacén o tienda
nom_almacen	varchar	Nombre del almacén o tienda
cod_medio_pago	varchar	Codigo del medio de pago
cod_producto	varchar	Codigo del producto comprado
tipo_venta	varchar	Venta física o virtual
canal_virtual	varchar	Canal virtual por donde ingresa la venta
cliente_tipo	varchar	Tipo de Cliente
identificación	varchar	Número de Identificación del Cliente
unidades	varchar	Cantidad de unidades compradas
valor_unidad	varchar	Valor unitario
valor_total	varchar	Valor Total del producto
Iva	varchar	Valor del impuesto de IVA si aplica
venta_neta	varchar	Valor neto, puede ser de venta o devolución
tipo_registro	varchar	Tipo de registro Venta o Devolución
documento_cliente	varchar	Tipo de documento del cliente
fecha_insercion	datetime	Fecha en que se insertó el registro

Fuente: Elaboración Propia, generada con Excel.

Tabla 9 Estructura Tabla Almacenes Bronce

Columna	Tipo de Dato	Descripción
cod_almacen	varchar	Codigo del almacén o tienda
nom_almacen	varchar	Nombre del almacén o tienda
tipo	varchar	Tipo de almacén, físico o virtual
latitud_y	varchar	Coordenada de ubicación del local, Latitud
longitud_x	varchar	Coordenada de ubicación del local, Longitud
fecha_insercion	datetime	Fecha en que se insertó el registro

Tabla 10 Estructura Tabla Medios de Pago Bronce

Columna	Tipo de Dato	Descripción
cod_medio_pago	varchar	Codigo del medio de pago
nom_medio_pago	varchar	Nombre del medio de pago
fecha_insercion	datetime	Fecha en que se insertó el registro

Fuente: Elaboración Propia, generada con Excel.

Tabla 11 Estructura Tabla Proveedores Bronce

Columna	Tipo de Dato	Descripción
prov_cod	varchar	Codigo del proveedor
razon_social	varchar	Nombre o razón social del proveedor
email	varchar	Email del Proveedor
tipo	varchar	Origen, Nacional o Importador
fecha_insercion	datetime	Fecha en que se insertó el registro

Fuente: Elaboración Propia, generada con Excel.

Tabla 12 Estructura Tabla Productos Bronce

Columna	Tipo de Dato	Descripción
cod_jerarquia	varchar	Codigo de la categoría del producto
nom_jerarquia	varchar	Nombre de la categoría del producto
cod_producto	varchar	Codigo del producto
nom_producto	varchar	Nombre del producto
prov_cod	varchar	Codigo del proveedor
fecha_insercion	datetime	Fecha en que se insertó el registro

Fuente: Elaboración Propia, generada con Excel.

Tabla 13 Estructura Tabla Configuración Proveedor / Plan Bronce

Columna	Tipo de Dato	Descripción
prov_cod	varchar	Codigo del proveedor
plan_cod	varchar	Codigo del plan de acceso a la información
estado_plan	varchar	Estado Activo, Inactivo
fecha_inicio_vigencia	varchar	fecha inicio de la vigencia del plan
fecha_fin_vigencia	varchar	fecha inicio de la vigencia del plan
fecha_insercion	datetime	Fecha en que se insertó el registro

Fuente: Elaboración Propia, generada con Excel.

A continuación, se describen las estructuras de almacenamiento de datos Gold (transformados), (ver tablas 14 – 17).

Tabla 14 Estructura Tabla Ventas / Devoluciones Gold

Columna	Tipo de Dato	Descripción
fecha	date	Fecha de la transacción
hora_compra	integer	Hora de la transacción
cod_almacén	integer	Código del almacén o tienda
nom_almacén	varchar	Nombre del almacén o tienda
cod_medio_pago	integer	Código del medio de pago
nom_medio_pago	varchar	Nombre del medio de pago
cod_producto	integer	Código del producto comprado
nom_producto	varchar	Nombre del producto
cod_jerarquia	integer	Código de la categoría del producto
nom_jerarquia	varchar	Nombre de la categoría del producto
prov_cod	integer	Código del proveedor
razon_social	varchar	Razón social o nombre del proveedor
tipo_proveedor	varchar	Proveedor de tipo nacional o importador
tipo_venta	varchar	Venta física o virtual
canal_virtual	varchar	Canal virtual por donde ingresa la venta
cliente_tipo	varchar	Tipo de Cliente
tipo_registro	varchar	Tipo de registro Venta o Devolución
unidades	numérico	Cantidad de unidades compradas
valor_unidad	bignumeric	Valor unitario
valor_total	bignumeric	Valor Total del producto
iva	bignumeric	Valor del impuesto de IVA si aplica
venta_neta	bignumeric	Valor neto, puede ser de venta o devolución
Cont_trx_prv_hora	bignumeric	Conteo de transacciones distintas
fecha_insercion	datetime	Fecha en que se insertó el registro

Fuente: Elaboración Propia, generada con Excel.

Tabla 15 Estructura Tabla Errores Técnicos

Columna	Tipo de Dato	Descripción
le_fecha	datetime	Fecha y hora de ocurrido el error
le_nom_proceso	varchar	Nombre del proceso donde sucede el error

le_proyecto_id	varchar	Nombre del proyecto
le_job_id	varchar	Nombre del job
le_mensaje	varchar	Mensaje técnico del error
le_query	varchar	Texto o query donde se presenta el error

Fuente: Elaboración Propia, generada con Excel.

Tabla 16 Estructura Tabla Errores de Negocio

Columna	Tipo de Dato	Descripción
fecha	date	Fecha del registro
cod_almacen	integer	Codigo del almacén donde llega el dato errado
nombre_columna	varchar	Nombre de la columna donde se presenta error
valor_columna	varchar	Valor errado
fuentes_de_dato	varchar	Fuente origen del dato
observaciones	varchar	Mensaje
fecha_insercion	datetime	Fecha en que se insertó el registro

Fuente: Elaboración Propia, generada con Excel.

Tabla 17 Estructura Tabla Configuración Proveedor / Plan Gold

Columna	Tipo de Dato	Descripción
prov_cod	numérico	Codigo del proveedor
plan_cod	varchar	Codigo del plan de acceso a la información
estado_plan	varchar	Origen, Nacional o Importador
fecha_inicio_vigencia	date	fecha inicio de la vigencia del plan
fecha_fin_vigencia	date	fecha inicio de la vigencia del plan
fecha_insercion	datetime	Fecha en que se insertó el registro

Fuente: Elaboración Propia, generada con Excel.

5.3.1 Desarrollo de los procesos de extracción, transformación y carga

Para la implementación se trabajó con dos técnicas de carga EL(Extract, Load) y ETL (Extract, Transform, Load), ambas técnicas hacen parte de las buenas prácticas recomendadas por el DAMA (DAMA International, 2017).

5.3.1.1 Creación del proyecto en GCP

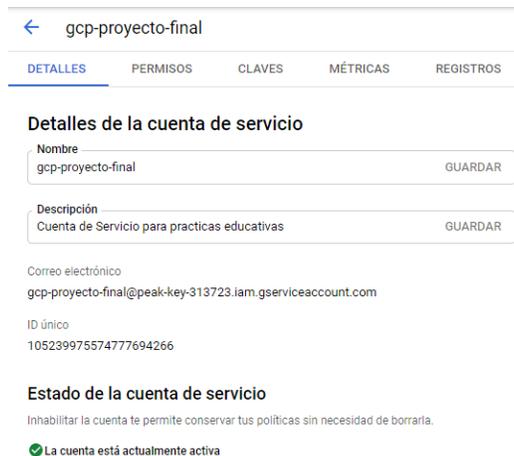
GCP es la plataforma de Google la cual ofrece gran variedad de servicios en la nube para el desarrollo de proyectos, en el caso de proyectos de datos como el que se aborda, es muy eficiente por las ventajas de almacenamiento y procesamiento que ofrece. A continuación, se crea el proyecto de trabajo y la cuenta de servicio, (ver figuras 10 - 11).

Figura 10. Proyecto creado en CGP.



Fuente: Tomado de Google Cloud Platform.

Figura 11. Creación de Cuenta de Servicio CGP.

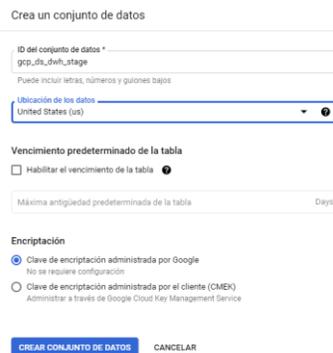


Fuente: Tomado de Google Cloud Platform, IAM.

5.3.1.2 Creación de las estructuras de almacenamiento definidos anteriormente

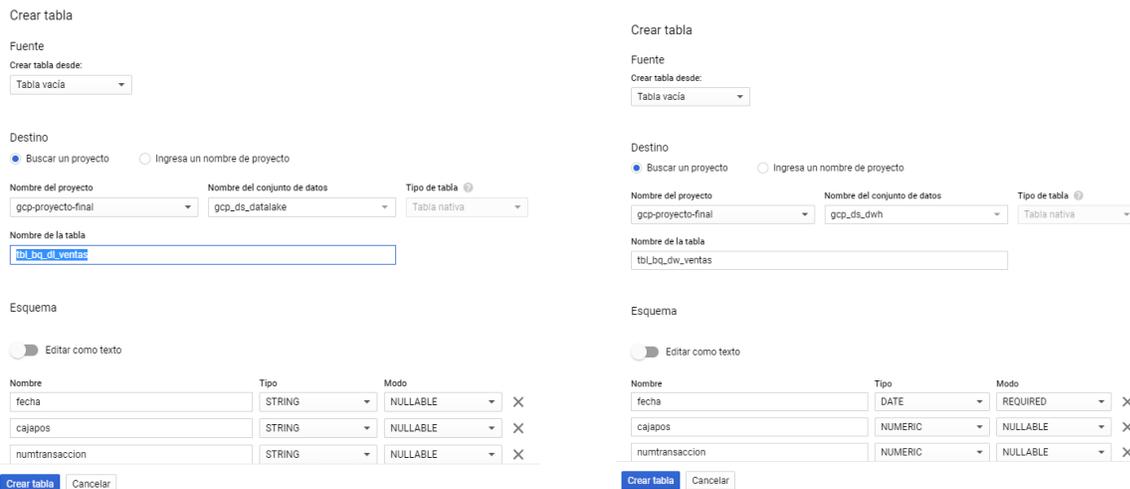
Para el almacenamiento de la información se utilizó el servicio de GCP BigQuery, el cual ofrece gran capacidad de almacenamiento y flexibilidad de trabajar con los datos, ya que se puede hacer uso del lenguaje estándar SQL, a pesar de que por debajo esta sea una base de datos de tipo columnar, además de su gran velocidad de procesamiento. Lo primero que se debe configurar o crear son los conjuntos de datos donde estarán las tablas y objetos necesarios para el almacenamiento (ver figuras 12 – 14).

Figura 12. Creación del conjunto de Datos.



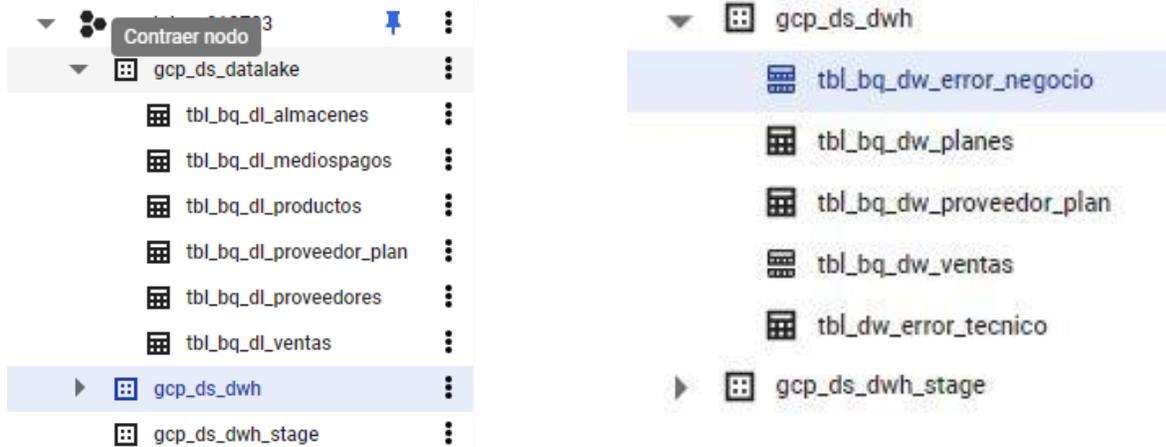
Fuente: Tomado de Google Cloud Platform, BigQuery.

Figura 13 Creación Tabla Ventas / Devoluciones Data Lake.



Fuente: Tomado de Google Cloud Platform, BigQuery.

Figura 14. Tablas Creadas en BigQuery GCP.



Fuente: Tomado de Google Cloud Platform, BigQuery.

5.3.1.3 Creación de Notebook en Google Colab

A continuación, se procede con la creación del notebook en Google Colab, que permite la lectura de los archivos fuentes y su posterior carga en BigQuery. Los archivos se cargarán en el esquema del DataLake sin realizar ninguna transformación, para lograr este objetivo, se deben instalar las librerías necesarias en Python que garanticen la conexión con la plataforma de Google. Los datos de conexión se guardan en un archivo de configuración, garantizando confidencialidad (ver figura 15).

Figura 15. Creación Notebook e Instalación de Librerías.



Fuente: Tomado de Google Colab, Notebook.

5.3.1.4 Desarrollar EL de carga de datos bronce

La lectura de los archivos fuentes se realizó con Python, como se mencionó en el punto anterior y se utiliza la potencia de una librería como pandas para leer los archivos y cargarlos en dataframe y posteriormente subirlos a GCP. Se menciona trabajar con la técnica o concepto EL (Extrae y Carga) ya que no es necesario hacer ninguna transformación sobre Python, la ingesta de datos se realizó en BigQuery y sobre esta misma herramienta se hizo la transformación y limpieza de los datos, (ver figuras 16 – 18).

Figura 16. Lectura Archivos Fuentes en Python.

```

1. LECTURA CSV (Productos, Almacenes, Medios de Pago, Proveedores y Ventas)

[13] df_productos=pd.read_csv('/content/drive/MyDrive/Segundo_Semestre/Proyecto_Aplicacion/Fuentes_Datos/Productos.csv',sep=';',encoding='UTF-8')
print('El Archivo .csv fue leído, ' + str(df_productos.shape[0]) + ' Registros')
El Archivo .csv fue leído, 433594 Registros

[14] df_proveedores=pd.read_csv('/content/drive/MyDrive/Segundo_Semestre/Proyecto_Aplicacion/Fuentes_Datos/Proveedores.csv',sep=';',encoding='UTF-8')
print('El Archivo .csv fue leído, ' + str(df_proveedores.shape[0]) + ' Registros')
El Archivo .csv fue leído, 6605 Registros

[15] df_almacenes=pd.read_csv('/content/drive/MyDrive/Segundo_Semestre/Proyecto_Aplicacion/Fuentes_Datos/Almacenes.csv',sep=';',encoding='UTF-8')
print('El Archivo .csv fue leído, ' + str(df_almacenes.shape[0]) + ' Registros')
El Archivo .csv fue leído, 6 Registros

[16] df_mediospagos=pd.read_csv('/content/drive/MyDrive/Segundo_Semestre/Proyecto_Aplicacion/Fuentes_Datos/MediosPago.csv',sep=';',encoding='UTF-8')
print('El Archivo .csv fue leído, ' + str(df_mediospagos.shape[0]) + ' Registros')
El Archivo .csv fue leído, 58 Registros

df_ventas=pd.read_csv('/content/drive/MyDrive/Segundo_Semestre/Proyecto_Aplicacion/Fuentes_Datos/Ventas.csv',sep=';',encoding='UTF-8')
print('El Archivo .csv fue leído, ' + str(df_ventas.shape[0]) + ' Registros')
  
```

Fuente: Tomado de Google Colab, Notebook.

Figura 17. Conexiones a GCP.

```

2.1 Conexion

#Se obtiene KEY Cuenta de Servicio de GCP
credentials = service_account.Credentials.from_service_account_file('/content/drive/MyDrive/Segundo_Semestre/Proyecto_Aplicacion/Conexiones/cuenta_servicio_gcp.json',

[9] #Se obtienen datos de Conexion GCP
df_conex=pd.read_csv('/content/drive/MyDrive/Segundo_Semestre/Proyecto_Aplicacion/Conexiones/conexionescsv.csv',sep=';',encoding='UTF-8',header=None)
proyecto=df_conex[df_conex.index ==0][0].tolist()
idproyecto=df_conex[df_conex.index ==0][1].tolist()
dataset_datalake=df_conex[df_conex.index ==0][2].tolist()
dataset_dwh=df_conex[df_conex.index ==0][3].tolist()
proyecto=proyecto[1:-1].replace("'",'')
idproyecto=idproyecto[1:-1].replace("'",'')
dataset_datalake=dataset_datalake[1:-1].replace("'",'')
dataset_dwh=dataset_dwh[1:-1].replace("'",'')

tabla_almacenes= dataset_datalake + '.tbl_bq_dl_almacenes'
tabla_productos= dataset_datalake + '.tbl_bq_dl_productos'
tabla_mediospagos= dataset_datalake + '.tbl_bq_dl_mediospagos'
tabla_proveedores= dataset_datalake + '.tbl_bq_dl_proveedores'
tabla_ventas= dataset_datalake + '.tbl_bq_dl_ventas'
  
```

Fuente: Tomado de Google Colab, Notebook.

Figura 18. Carga de datos en Data Set y Tablas BigQuery.

```

2.2 Carga de Datos en GCP - BigQuery

[10] #carga Almacenes
df_almacenes.to_gbq(destination_table=tabla_almacenes,project_id=idproyecto,if_exists='replace',progress_bar=True,credentials=credentials)
print("done, " + "Total Almacenes Cargados en GCP : " + str(df_almacenes.shape[0]))
iit [00:06, 8.17s/it]done, Total Almacenes Cargados en GCP : 6

[11] #carga Producto
df_productos.to_gbq(destination_table=tabla_productos,project_id=idproyecto,if_exists='replace',progress_bar=True,credentials=credentials)
print("done, " + "Total Producto Cargados en GCP : " + str(df_productos.shape[0]))
iit [00:18, 18.04s/it]done, Total Producto Cargados en GCP : 438594

[12] #carga Medios Pagos
df_mediospagos.to_gbq(destination_table=tabla_mediospagos,project_id=idproyecto,if_exists='replace',progress_bar=True,credentials=credentials)
print("done, " + "Total Medios Pagos Cargados en GCP : " + str(df_mediospagos.shape[0]))
iit [00:04, 4.18s/it]done, Total Medios Pagos Cargados en GCP : 58

[13] #carga Proveedores
df_proveedores.to_gbq(destination_table=tabla_proveedores,project_id=idproyecto,if_exists='replace',progress_bar=True,credentials=credentials)
print("done, " + "Total Proveedores Cargados en GCP : " + str(df_proveedores.shape[0]))
iit [00:06, 6.23s/it]done, Total Proveedores Cargados en GCP : 6685

[14] #carga Ventas
df_ventas.to_gbq(destination_table=tabla_ventas,project_id=idproyecto,if_exists='replace',progress_bar=True,credentials=credentials)
print("done, " + "Total Ventas Cargadas en GCP : " + str(df_ventas.shape[0]))
iit [04:35, 275.41s/it]done, Total Ventas Cargadas en GCP : 5028182

```

Fuente: Tomado de Google Colab, Notebook.

5.3.1.4 Desarrollar ETL de Limpieza y Transformación de datos

Una vez los datos están cargados en BigQuery en las estructuras definidas del Data Lake, se procede con la limpieza y transformación de la información, siguiendo las directrices definidas en el capítulo de lineamientos de gobierno. La estrategia definida para lograrlo fue desarrollar un procedimiento almacenado en BigQuery, el cual contiene la lógica adecuada para leer los datos fuentes y cargarlos en la tabla destino.

Adicional a la carga de información, este procedimiento debe ser capaz de identificar aquellos dominios que llegan como parte de los datos transaccionales pero que no se encuentran en los datos maestros, esto para generar alertas a la compañía dueña de la información y realizar la gestión de corrección de fuentes; También debe haber un log de auditoría de errores técnicos que se puedan presentar durante las ejecuciones de carga del modelo. En las siguientes figuras se muestran los artefactos creados y la visualización previa de las tablas finales. (ver figuras 19 - 23).

Figura 19. Tablas Creadas para Manejo de Errores.

Fuente: Tomado de Google Cloud Platform, BigQuery.

Figura 20. Creación del Procedimiento Almacenado.

Información del procedimiento almacenado

ID del procedimiento almacenado	peak-key-313723.gcp_ds_procesos.sp_bp_carga_tbl_bq_dl_ventas
Creado	30 sep. 2021, 21:34:10 UTC-5
Última modificación	30 sep. 2021, 22:24:45 UTC-5
Idioma	SQL
Descripción	Procedimiento que lee las ventas transaccionales y las carga en la tabla destino agregada y que es fuente para la visualización de los dashboard de monetización

```

BEGIN
DECLARE vfechaini String DEFAULT '0';
DECLARE vfechafin String DEFAULT '0';

SET vfechaini = (SELECT min(fecha) FROM `peak-key-313723.gcp_ds_datalake.tbl_bq_dl_ventas`);
SET vfechafin = (SELECT max(fecha) FROM `peak-key-313723.gcp_ds_datalake.tbl_bq_dl_ventas`);

#Delete de las fechas a Reprocesar
DELETE FROM `peak-key-313723.gcp_ds_dwh.tbl_bq_dw_ventas`
WHERE FECHA BETWEEN PARSE_DATE("%Y%m%d", vfechaini) AND PARSE_DATE("%Y%m%d", vfechafin);
    
```

Fuente: Tomado de Google Cloud Platform, BigQuery.

Figura 21. Ejecución del Procedimiento Almacenado.

Trabajo	Etapas finalizadas	Bytes procesados	Acción
✓ 22:34 Procedimiento	2	47.95 MB	VER RESULTADOS
✓ 22:34 Procedimiento	2	47.95 MB	VER RESULTADOS
✓ 22:34 Procedimiento	3	1.38 GB	VER RESULTADOS
✓ 22:35 Procedimiento	18	555.27 MB	VER RESULTADOS
✓ 22:35 Procedimiento	4	251.41 MB	VER RESULTADOS

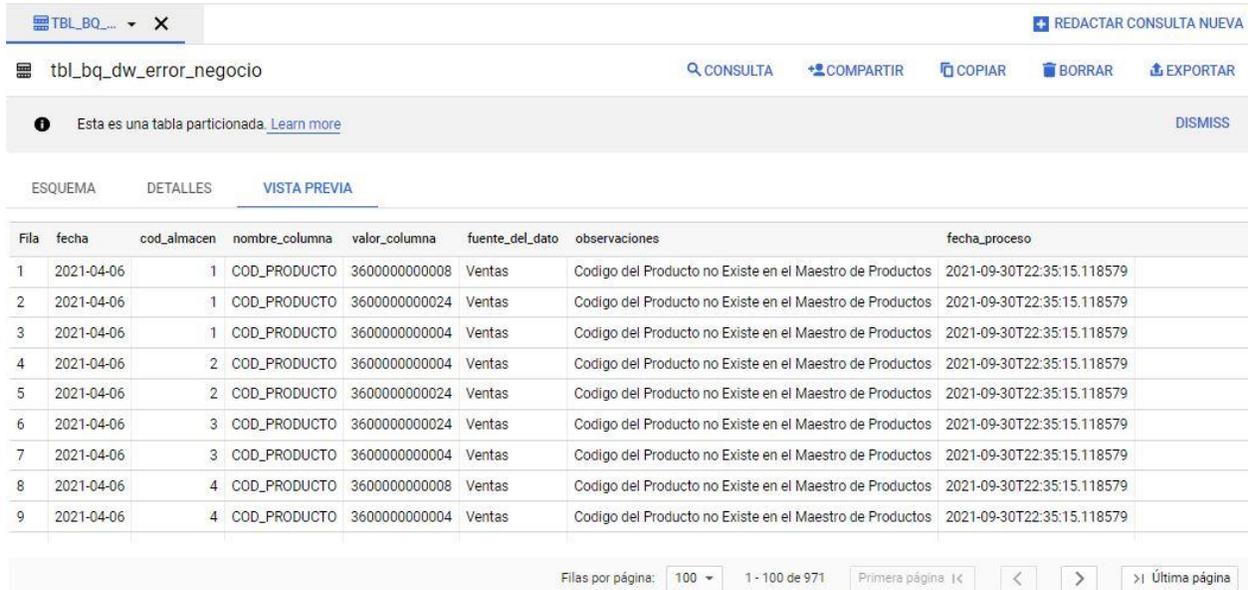
Fuente: Tomado de Google Cloud Platform, BigQuery.

Figura 22. Tabla Ventas del DW (Datawarehouse) cargada después de ejecutar el Procedimiento Almacenado

ESQUEMA	DETALLES	VISTA PREVIA
62	2021-01-28	19 4 ALMACEN 4 46 46 Medio de Pago Tar 81364 PRODUCTO 1256-81364-LIMPIA 1256 PAREDI
63	2021-01-28	17 2 ALMACEN 2 22 22 Medio de Pago Efe 81364 PRODUCTO 1256-81364-LIMPIA 1256 PAREDI
64	2021-01-28	18 4 ALMACEN 4 22 22 Medio de Pago Efe 81364 PRODUCTO 1256-81364-LIMPIA 1256 PAREDI
65	2021-01-28	11 1 ALMACEN 1 124 124 Medio de Pago Tar 81376 PRODUCTO 1256-81376-SELLAD 1256 PAREDI
66	2021-01-28	17 5 ALMACEN 5 22 22 Medio de Pago Efe 81376 PRODUCTO 1256-81376-SELLAD 1256 PAREDI
67	2021-01-28	9 2 ALMACEN 2 22 22 Medio de Pago Efe 211552 PRODUCTO 420-211552-ESTUCO 420 INICIO
68	2021-01-28	9 2 ALMACEN 2 84 84 Medio de Pago Dev 211552 PRODUCTO 420-211552-ESTUCO 420 INICIO
69	2021-01-28	10 2 ALMACEN 2 22 22 Medio de Pago Efe 211552 PRODUCTO 420-211552-ESTUCO 420 INICIO
70	2021-01-28	15 2 ALMACEN 2 22 22 Medio de Pago Efe 211552 PRODUCTO 420-211552-ESTUCO 420 INICIO
71	2021-01-28	-1 6 ALMACEN 6 -2 null 211552 PRODUCTO 420-211552-ESTUCO 420 INICIO

Fuente: Tomado de Google Cloud Platform, BigQuery.

Figura 23. Tabla Errores de Negocio cargada después de ejecutar el SP.



Fila	fecha	cod_almacen	nombre_columna	valor_columna	fFuente_del_datos	observaciones	fecha_proceso
1	2021-04-06	1	COD_PRODUCTO	3600000000008	Ventas	Codigo del Producto no Existe en el Maestro de Productos	2021-09-30T22:35:15.118579
2	2021-04-06	1	COD_PRODUCTO	3600000000024	Ventas	Codigo del Producto no Existe en el Maestro de Productos	2021-09-30T22:35:15.118579
3	2021-04-06	1	COD_PRODUCTO	3600000000004	Ventas	Codigo del Producto no Existe en el Maestro de Productos	2021-09-30T22:35:15.118579
4	2021-04-06	2	COD_PRODUCTO	3600000000004	Ventas	Codigo del Producto no Existe en el Maestro de Productos	2021-09-30T22:35:15.118579
5	2021-04-06	2	COD_PRODUCTO	3600000000024	Ventas	Codigo del Producto no Existe en el Maestro de Productos	2021-09-30T22:35:15.118579
6	2021-04-06	3	COD_PRODUCTO	3600000000024	Ventas	Codigo del Producto no Existe en el Maestro de Productos	2021-09-30T22:35:15.118579
7	2021-04-06	3	COD_PRODUCTO	3600000000004	Ventas	Codigo del Producto no Existe en el Maestro de Productos	2021-09-30T22:35:15.118579
8	2021-04-06	4	COD_PRODUCTO	3600000000008	Ventas	Codigo del Producto no Existe en el Maestro de Productos	2021-09-30T22:35:15.118579
9	2021-04-06	4	COD_PRODUCTO	3600000000004	Ventas	Codigo del Producto no Existe en el Maestro de Productos	2021-09-30T22:35:15.118579

Fuente: Tomado de Google Cloud Platform, BigQuery.

5.4 Definición de Planes

Para la implementación del modelo de monetización y su recaudo, se realizó por el sistema de pago por suscripción, el cual tiene ventajas para el suscriptor y para la compañía dueña de la plataforma. Este método es muy usado en plataformas digitales y consiste en un pago mensual o suscripción que hace el cliente, a cambio del beneficio de un servicio que varía en proporción de acuerdo con el monto pagado, tiene la ventaja para el cliente de poder cancelar la suscripción cuando el servicio no satisfaga sus necesidades y tiene la ventaja para la compañía de ir mejorando y ampliando sus servicios para retener más clientes y suscriptores. Para el inicio del proyecto se definieron 2 planes, uno sin pago de suscripción con opciones limitadas y un plan de pago con múltiples opciones de información, los cuales se describen a continuación:

Plan Free: Plan al que tienen derecho cada uno de los proveedores que comercializan productos en las plataformas Marketplace de la compañía. Este plan incluye:

- Envío una vez al mes por correo electrónico de archivo en formato PDF, con reporte de ventas y devoluciones consolidadas al mes de corte.
- Acceso a la plataforma gratis con uso limitado a la información anteriormente mencionada.

Plan Pago: Suscripción de pago mensual o anual, que incluye:

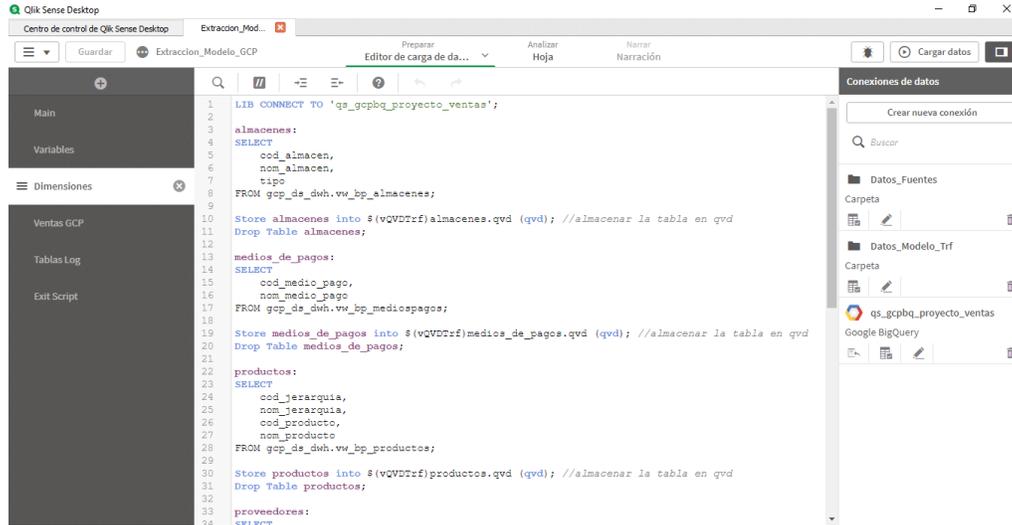
- Acceso a la plataforma con uso ilimitado a la información, podrá ver las diferentes gráficas e indicadores de acuerdo con las condiciones de búsqueda y filtros que desee analizar.
- Indicadores de ventas por año, mes, día.
- Indicadores de crecimientos año/mes actual vs año/mes anterior.
- Indicadores de ventas x almacén, medios de pago, categorías y productos.
- Indicadores de Ticket promedio.
- Indicadores de hora frecuente de compra.
- Indicadores sobre Top de productos más vendidos o devueltos de acuerdo a periodos y filtros de búsqueda.
- Indicadores sobre Tasa de devolución o garantías.
- Envío una vez al mes por correo electrónico de archivo en formato PDF, con análisis consolidados de los indicadores mencionados anteriormente.

5.5 Visualización e Implementación

Una vez cargado el modelo de datos, definidos los planes y con las consultas definidas para cada análisis, se crearon las visualizaciones o Dashboard para cada tipo de suscripción, estos se desarrollaron en la herramienta Qlik Sense (Qlik Sense, 2021) debido a su facilidad de uso, su integración con GCP BigQuery y su potente modelo asociativo que permite generar insights a partir de los datos, Qlik Sense garantiza la confidencialidad de la información ya que se puede segmentar la información por cliente a través de Section Access (código embebido en la aplicación), es decir que un cliente o proveedor solo puede ver su información, el acceso y autenticación se puede integrar con el directorio activo de la compañía, lo que hace necesario que cada cliente o suscriptor sea ingresado de forma especial al directorio activo.

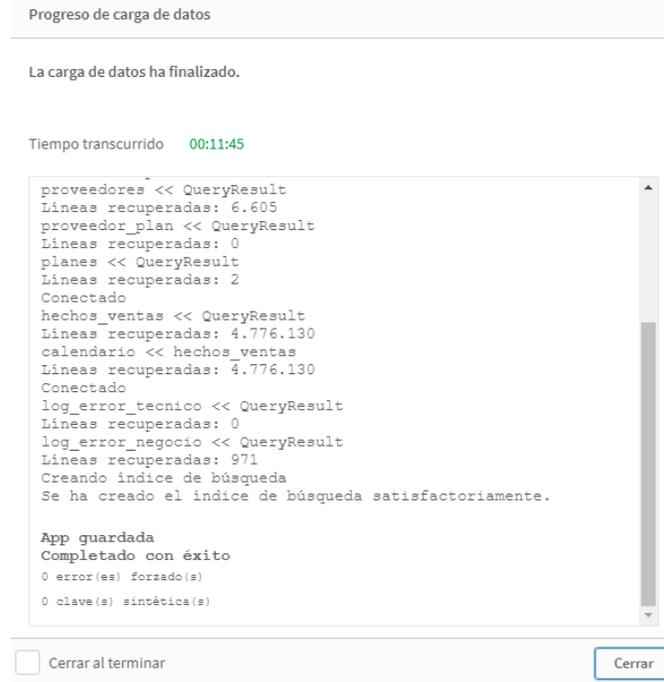
Antes de iniciar con el desarrollo de las visualizaciones, se creó el modelo asociativo o modelo de datos con el que interactuarán los suscriptores en los análisis en la herramienta. Para lograr esto, se hizo en Qlik Sense un desarrollo que lee las fuentes o tablas directamente desde BigQuery y las almacena en archivos de extensión .qvd, estos archivos son nativos de la herramienta y tiene una compresión muy fuerte lo que los deja muy livianos al momento de leerlos, en este proceso se desarrolla la capa semántica del modelo, es decir que los nombres técnicos de columnas y entidades, se pasan a nombres entendibles por el usuario y el negocio, y al final se crea la nube de datos en Qlik Sense simulando un modelo estrella de base de datos dimensional, (ver figuras 24 – 26).

Figura 24. Script de Creación Modelo Asociativo Qlik Sense.



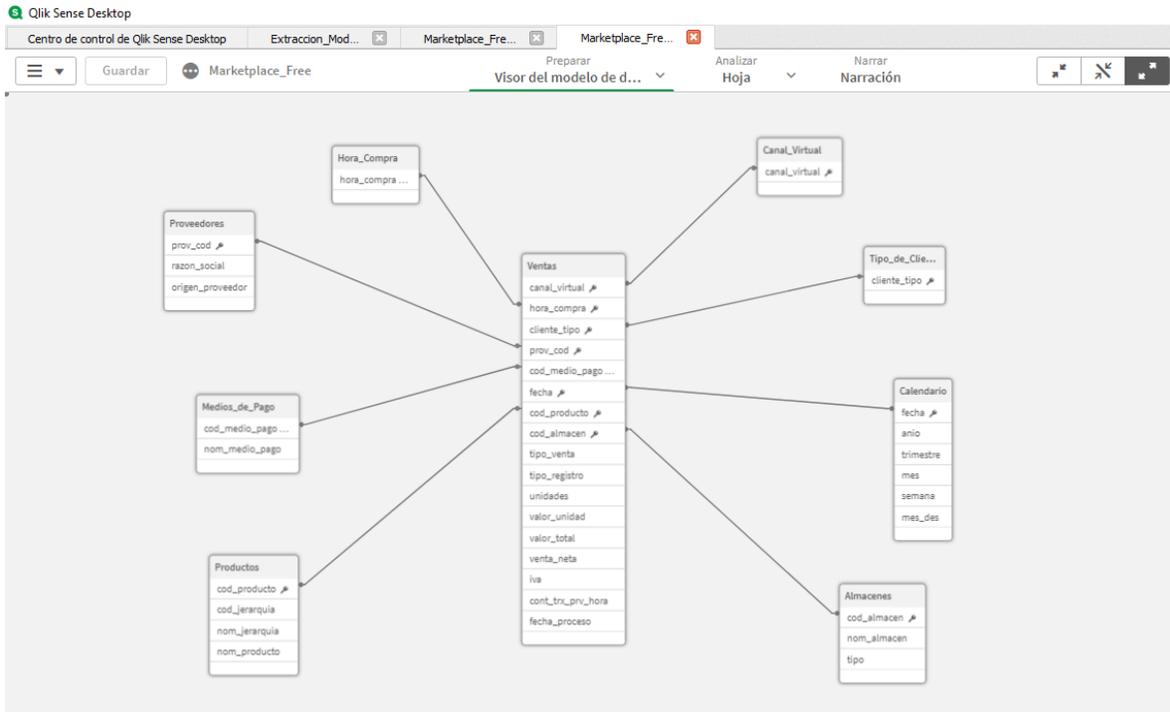
Fuente: Tomado de Qlik Sense Desktop.

Figura 25. Carga de Datos en Qlik Sense.



Fuente: Tomado de Qlik Sense Desktop.

Figura 26. Modelo Asociativo Qlik Sense.



Fuente: Tomado de Qlik Sense Desktop.

Dentro del proyecto es muy importante la seguridad y confidencialidad de la información, es por eso que en este punto es donde se configuran los accesos o perfilamiento de los datos según cada proveedor y plan de acceso, para esto se configuro en un archivo Excel por cada plan, que proveedores están habilitados y se adiciona en el código interno de la aplicación, (ver figuras 27 – 28).

Figura 27. Archivo de Muestra de Configuración Accesos.

	A	B	C
1	ACCESS	USERID	PROV_COD
2	ADMIN	DOMINIO\USER	1130
3			

Fuente: Elaboración Propia, generada de Excel.

Figura 28. Código de Perfilamiento de Datos en Qlik Sense.

The screenshot shows the Qlik Sense 'Editor de carga de da...' interface. On the left, there is a navigation pane with sections: Main, Variables, Section Access (selected), Modelo, and Exit Script. The main area displays a script with line numbers 1 through 21. The script defines a 'Section Access' section and loads data from an Excel file, then concatenates it with a table of user roles.

```

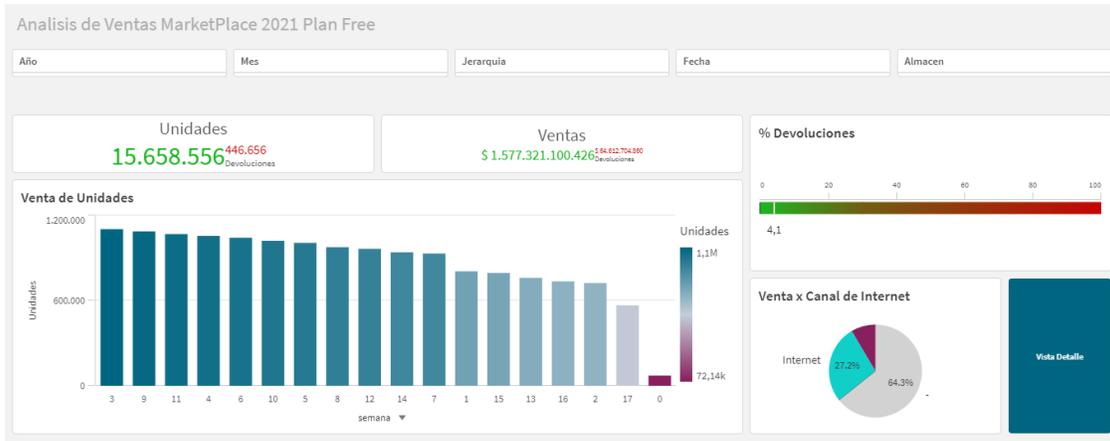
1 Section Access;
2 Section_Access:
3 LOAD
4     "ACCESS",
5     "USERID",
6     PROV_COD
7 FROM [lib://Datos_Modelo_Trf/Section Access.xlsx]
8 (ooxml, embedded labels, table is Hoja1);
9
10
11 Concatenate (Section_Access)
12
13 LOAD
14
15 'ADMIN' as ACCESS,
16 'INTERNAL\SA_SCHEDULER' as USERID,
17 '*' as PROV_COD
18
19 Autogenerate 1;
20
21 Section Application;
  
```

Fuente: Tomado de Qlik Sense Desktop.

5.5.1 Plan Free, Dashboard

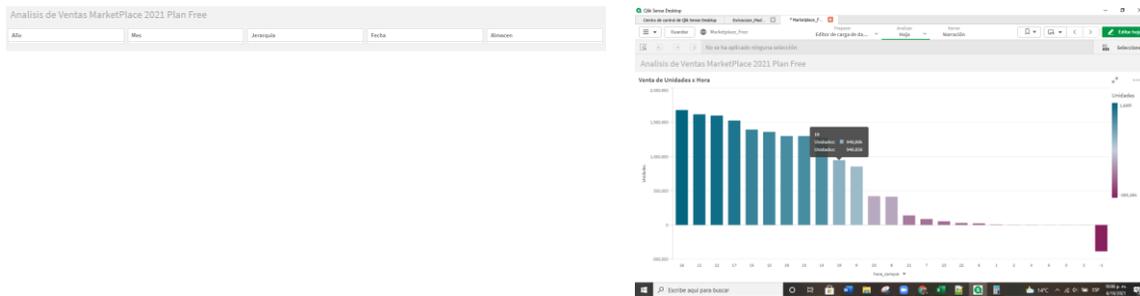
Esta visualización es algo muy general pero a la vez conciso e intuitivo. Está compuesto por una sola hoja o vista principal, donde se destacan los filtros que se pueden realizar en la parte superior, tiene un análisis de ventas de unidades por hora y por semana en gráfico de barras, análisis de ventas en porcentaje por canal de internet representado en gráfico de tarta, una tabla en la parte inferior con el detalle por fecha, producto y categoría de las ventas realizadas y un indicador del porcentaje de devoluciones con respecto a las ventas, el cual tiene un degrade en color rojo que alerta cuando empieza a sobrepasar el 15 %. Estos gráficos son exportables a Excel, se pueden usar a demanda y mensualmente también llegará un correo al suscriptor con el cierre de sus ventas (ver figura 29 - 32).

Figura 29. Visualización Dashboard Plan Free.



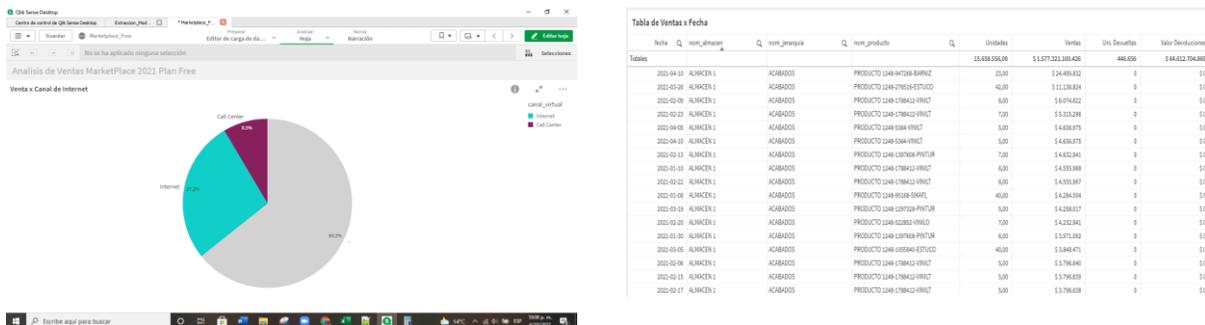
Fuente: Tomado de Qlik Sense Desktop.

Figura 30. Filtros y Barras Unidades x Hora.



Fuente: Tomado de Qlik Sense Desktop.

Figura 31. Gráfico de Tarta y Vista de Datos Detallada.



Fuente: Tomado de Qlik Sense Desktop.

Figura 32. Gráfico de Barras Unidades x Semana e Indicador % Devoluciones.



Fuente: Tomado de Qlik Sense Desktop.

5.5.2 Plan de Pago, Dashboard

Esta visualización es una vista 360 de las ventas de cada proveedor, tiene gran cantidad de filtros (indicar cantidad) y dimensiones por las cuales se pueden hacer análisis, su interfaz y gráficos son muy amigables e intuitivos, lo que facilita la interpretación y el análisis. Está compuesto por 3 hojas o vistas.

La primera ofrece un análisis general de las ventas del proveedor con diversos indicadores muy puntuales, se muestran las ventas y unidades vendidas acumuladas durante el último año, así como las ventas y unidades vendidas durante el mes en curso sino se tiene filtrado, o del mes que el usuario quiera analizar y filtrar, tiene indicador de promedio de ventas del mes en análisis y valor del ticket promedio, tiene un indicador del porcentaje de devoluciones con respecto a las ventas, el cual tiene un degrade en color rojo que alerta cuando empieza a sobrepasar el 15 %, gráfico de barras con las ventas en unidades y valor por cada almacén y un gráfico de distribución que muestra cómo se distribuyeron sus ventas entre los diferentes tipos de clientes y canales virtuales.

La segunda vista muestra dos gráficos de barras insertados dentro de un contenedor, uno evidencia el top 10 de los productos que más venta en dinero

representan y sus unidades y el otro gráfico de barras representa el top 10 de los productos más devueltos en cuestión de dinero. También tiene un gráfico de tarta que representa el porcentaje de ventas por canal físico y virtual, igualmente un gráfico de bloques que representa los medios de pago por los cuales los clientes están pagando sus productos, y finalmente en esta vista se tiene, un gráfico de viñetas que representa las ventas en unidades por cada hora.

La tercera vista es el detalle granular de los datos, que se representa con un gráfico de tabla pivotante donde el usuario puede jugar con las columnas y filas organizándolas según el análisis que quiera realizar. Cada una de estas vistas interactúan de forma dinámica con los filtros principales del modelo, es decir por cada filtro aplicado, las gráficas van cambiando y mostrando los valores que se filtran, y adicional por cada gráfico se pueden seleccionar más filtros, estos gráficos son exportables a Excel y Pdf, se pueden usar a demanda y mensualmente también llegará un correo al suscriptor con el cierre de sus ventas. A continuación se muestra la secuencia de graficas de este plan, (ver figuras 33 – 40).

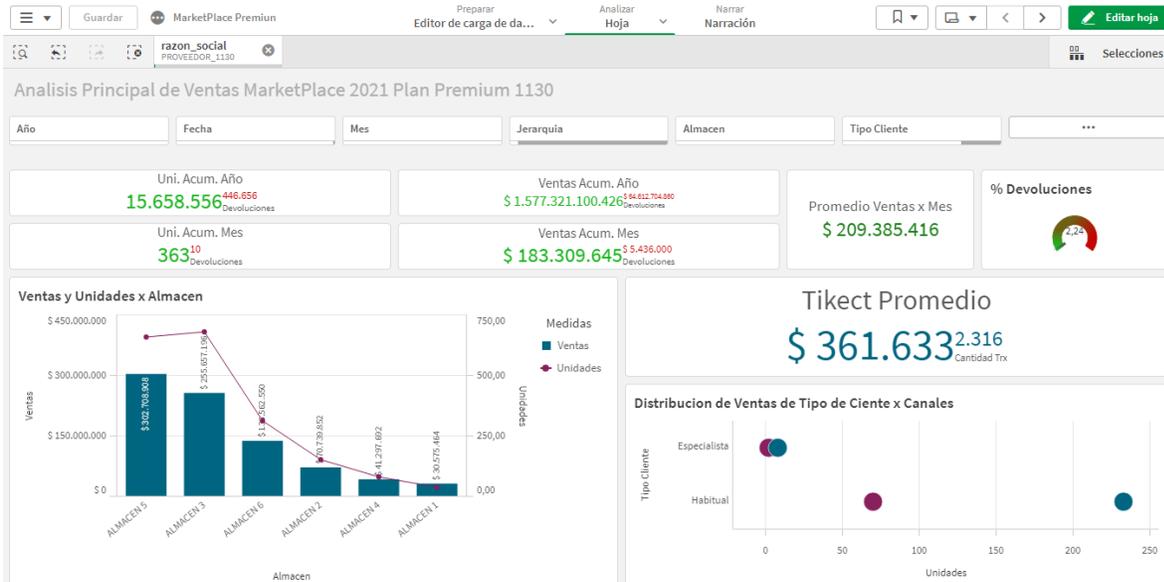
Figura 33. Plan Premium Panel de Filtros .

Analisis Principal de Ventas MarketPlace 2021 Plan Premium

Q, Año	Q, Jerarquia	Q, Almacen	Q, Tipo Cliente	Q, Medio de Pago	Q, Hora	Q, razon_social
2021	ACABADOS	ALMACEN 1	Empresarial	0 Medio de Pago N/A		-1 PROVEEDOR_0
	ACEROS	ALMACEN 2	Especialista	12 Medio de Pago Caj		0 PROVEEDOR_2
Q, Fecha	ASIENTOS	ALMACEN 3	Habitual	22 Medio de Pago Efe		7 PROVEEDOR_4
2021-01-02	CASA ELECTRICA	ALMACEN 4		26 Medio de Pago Cup		8 PROVEEDOR_6
2021-01-03	CASA FERRETERA	ALMACEN 5		28 Medio de Pago Red		9 PROVEEDOR_8
2021-01-04	CASA LIMPIA	ALMACEN 6		30 Medio de Pago Don		10 PROVEEDOR_10
2021-01-05	COCINETAS			32 Medio de Pago Bon		11 PROVEEDOR_12
2021-01-06	ELECTRICOS			34 Medio de Pago Pla		12 PROVEEDOR_14
2021-01-07	ENTRADAS Y BALCONES			36 Medio de Pago Efe		13 PROVEEDOR_16
2021-01-08	GRIFOS Y DEMAS			38 Medio de Pago Con		14 PROVEEDOR_18
Q, Mes	INICIO DE OBRA			42 Medio de Pago Che		15 PROVEEDOR_20
Enero	INOVACIONES			44 Medio de Pago Tar		16 PROVEEDOR_22
	LUCES Y AIRES			46 Medio de Pago Tar		17 PROVEEDOR_24

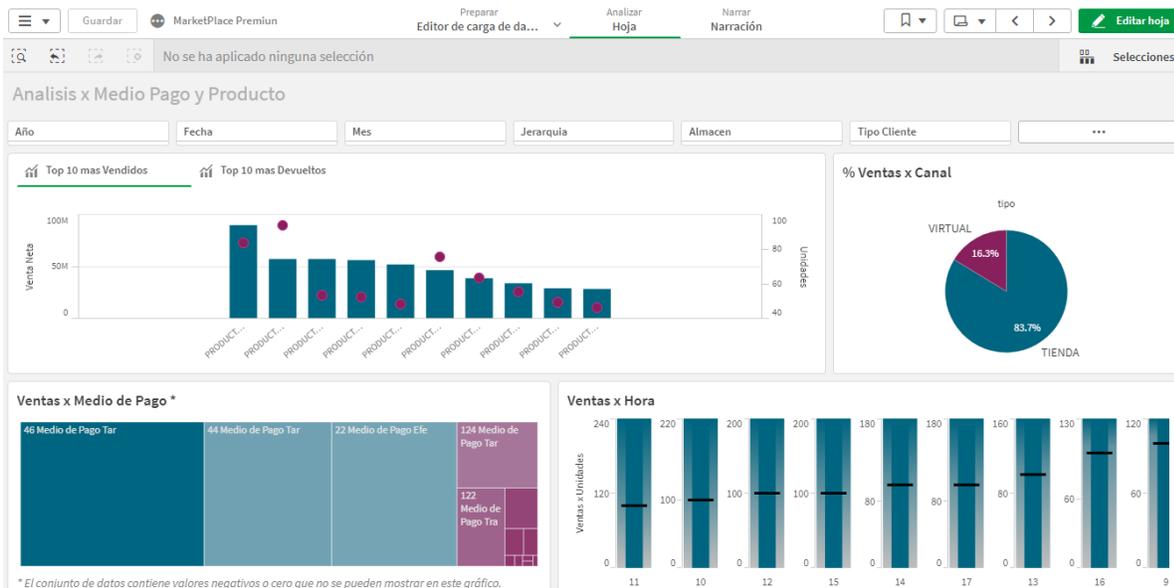
Fuente: Tomado de Qlik Sense Desktop.

Figura 34. Plan Premium Panel Principal.



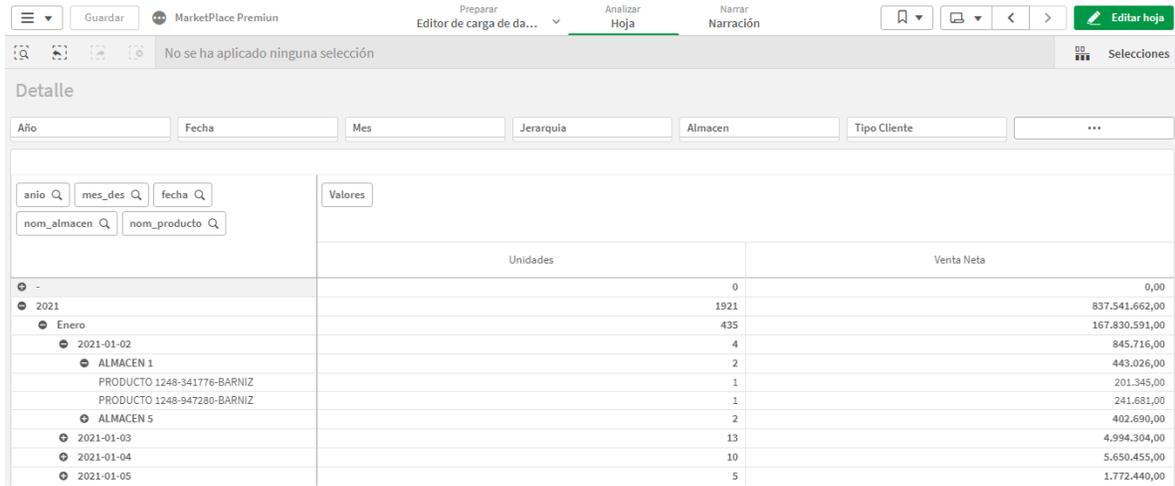
Fuente: Tomado de Qlik Sense Desktop.

Figura 35. Plan Premium Segunda Vista.



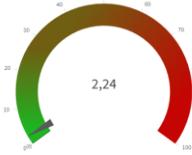
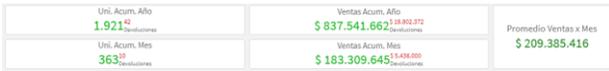
Fuente: Tomado de Qlik Sense Desktop.

Figura 36. Plan Premium Vista de Detalle.



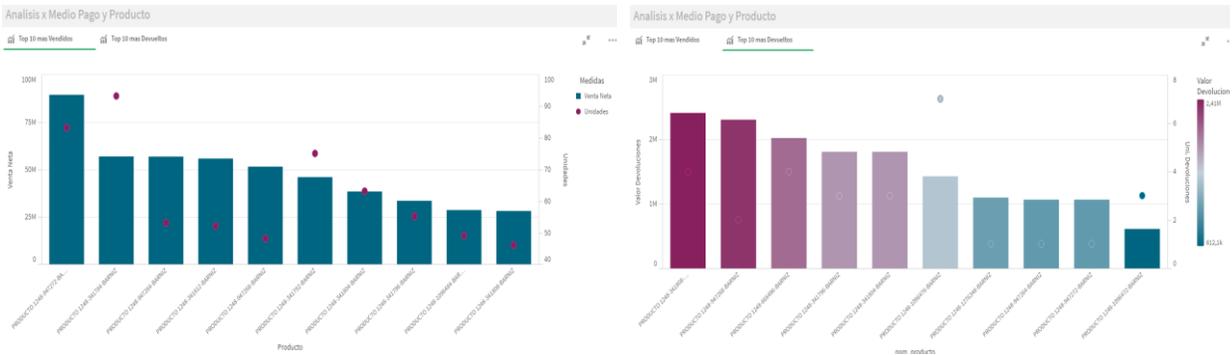
Fuente: Tomado de Qlik Sense Desktop.

Figura 37. Plan Premium Panel de Indicadores.



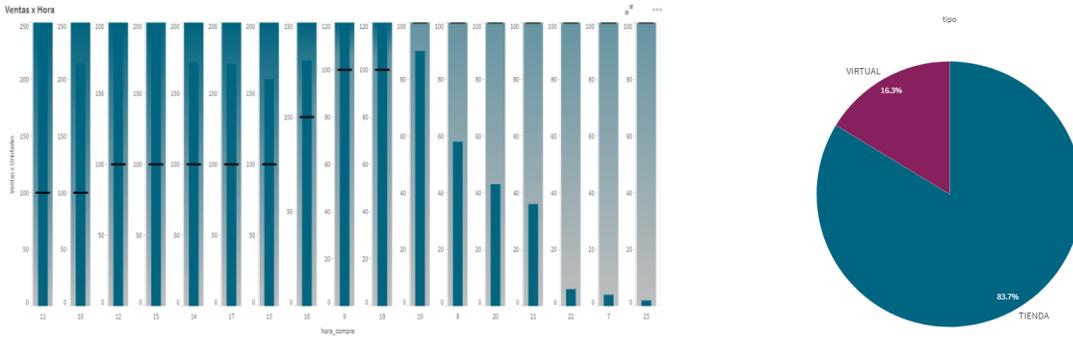
Fuente: Tomado de Qlik Sense Desktop.

Figura 38. Plan Premium Top 10 Mas Vendidos y más Devueltos.



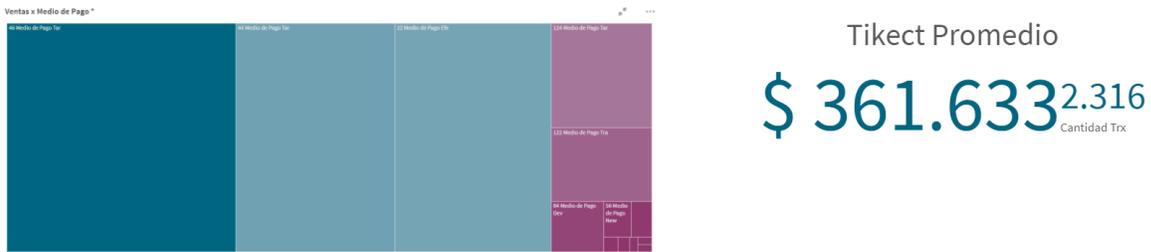
Fuente: Tomado de Qlik Sense Desktop.

Figura 39. Plan Premium Vetas por Hora, y Participación por Canal.



Fuente: Tomado de Qlik Sense Desktop.

Figura 40. Plan Premium Bloques por Medio de Pago y Ticket Promedio.



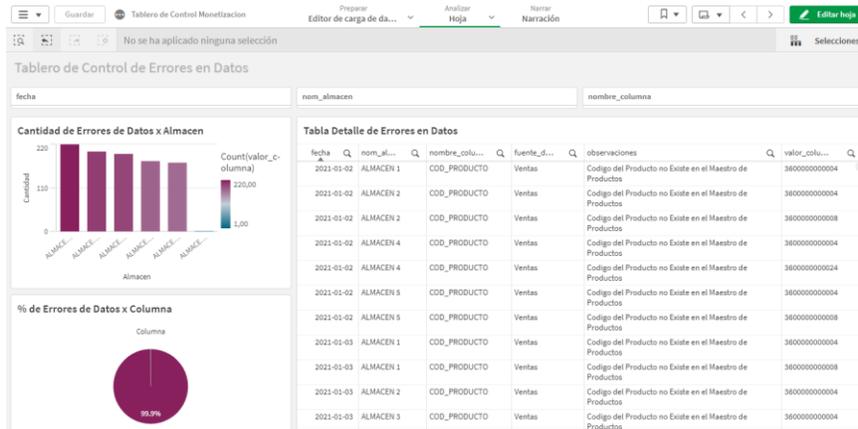
Fuente: Tomado de Qlik Sense Desktop.

5.5.3 Tablero de Control y Calidad

Esta visualización no hace parte de los planes y suscripciones, pero al igual que los otros es muy importantes para el funcionamiento y sobre todo para tener un control de los datos que se procesan, transforman y cargan en el modelo, pues ayuda a generar alertas que deben ser monitoreadas por los dueños de la plataforma y generar controles en los sistemas fuentes. Consta de un Dashboard principal con un gráfico de barras que visualiza la cantidad de errores de datos por almacén, un gráfico de tarta con el porcentaje de errores de datos por columna, es decir del total de errores en datos, que porcentaje corresponde a cada columna, y una tabla de detalle muy clara que informa la

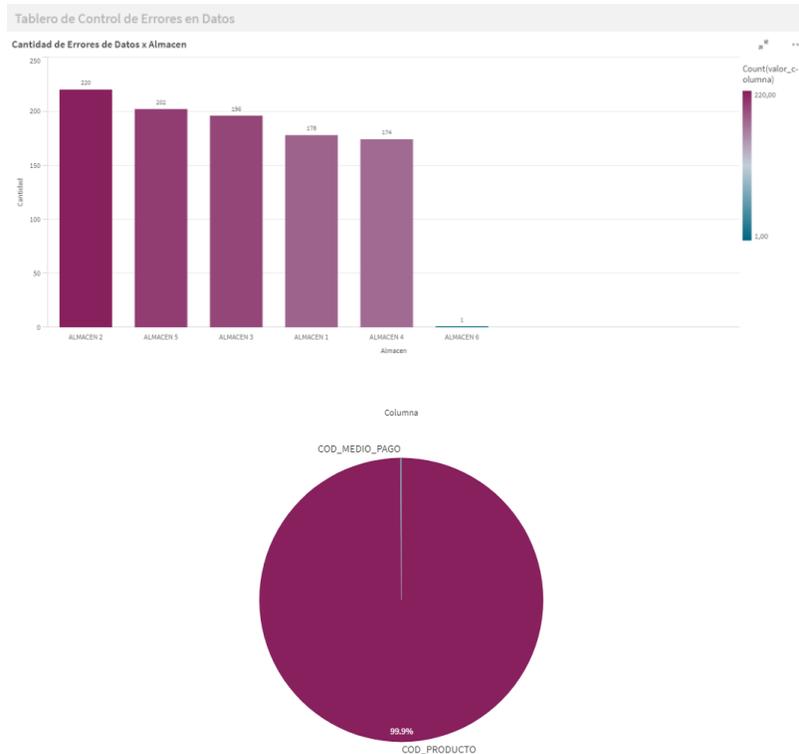
fecha, almacén, columna, observación o nota que dice lo que está mal y el valor que llega en la columna, (ver figuras 41 – 43).

Figura 41. Dashboard Principal Tablero de Control.



Fuente: Tomado de Qlik Sense Desktop.

Figura 42. Graficas Tablero de Control.



Fuente: Tomado de Qlik Sense Desktop.

Figura 43. Detalle Tablero de Control.

Tablero de Control de Errores en Datos

Tabla Detalle de Errores en Datos

fecha	nom_almacen	nombre_columna	fuentes_del_dato	observaciones	valor_columna
2021-01-02	ALMACEN 1	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000004
2021-01-02	ALMACEN 2	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000004
2021-01-02	ALMACEN 2	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000008
2021-01-02	ALMACEN 4	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000004
2021-01-02	ALMACEN 4	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000024
2021-01-02	ALMACEN 5	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000004
2021-01-02	ALMACEN 5	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000008
2021-01-03	ALMACEN 1	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000004
2021-01-03	ALMACEN 1	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000008
2021-01-03	ALMACEN 2	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000004
2021-01-03	ALMACEN 3	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000004
2021-01-03	ALMACEN 3	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000008
2021-01-03	ALMACEN 4	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000004
2021-01-03	ALMACEN 5	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000004
2021-01-04	ALMACEN 1	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000004
2021-01-04	ALMACEN 2	COD_PRODUCTO	Ventas	Codigo del Producto no Existe en el Maestro de Productos	3600000000004

Fuente: Tomado de Qlik Sense Desktop.

6. Conclusiones

A partir de la experiencia adquirida en el desarrollo de este proyecto se puede concluir lo siguiente.

Son varias las plataformas que ya permiten monetizar, algunas lo hacen con videos, contenidos como blogs, redes sociales, etc. Este proyecto lo que busca es desarrollar una solución que aporte valor a las compañías de Retail y sus proveedores de Marketplace.

El mercadeo de productos tecnológicos ha cambiado, hace algunos años las personas o empresas que querían comprar un software, debían hacerlo por un paquete completo así fueran a utilizar una pequeña parte, ahora todo se vende en pequeños servicios y por suscripciones, los clientes pagan lo que usan y quieren ver, es por esto que la solución propuesta, abarca planes y dependiendo del plan adquirido, la información es más específica.

Los datos son un activo de alto valor para las compañías, pero si no gestionan, procesan y almacenan de una forma adecuada que garantice la consistencia, integridad y acceso oportuno de la información, estos no serían aptos para la toma de decisiones.

En esta era tecnológica donde los datos afloran por todas partes, lo importante no es la captura del dato, lo importante no es tenerlo, lo verdaderamente importante es explotarlo, y eso fue lo que se quiso hacer con este producto, generarles valor a los datos de ventas que en otros tiempos y compañías solo sirven para generar informes de balances contables y financieros, con esta aplicación los datos utilizados impulsan más ventas e ingresos.

El uso de Dashboard dinámicos como los que ofrece esta solución, brinda una alternativa muy importante para que los usuarios, en este caso los proveedores de tipo Marketplace puedan tomar decisiones basadas en datos, a partir de los datos obtenidos de sus propias ventas.

7. Referencias

- Alberto. (2015). D'Alessandro. *Conceptualización de Estrategias de Monetización de Datos*. Universidad de San Andres, Buenos Aires, Argentina. Obtenido de <https://repositorio.udes.a.edu.ar/jspui/bitstream/10908/11159/1/%5bP%5d%5bW%5d%20M.%20Mar.%20D'Alessandro,%20Alberto.pdf>
- Andujar, J. A. (01 de 12 de 2018). *Sage.com*. Obtenido de Sage.com: <https://www.sage.com/es-es/blog/los-tres-modelos-de-monetizacion-mas-utilizados-en-los-negocios-digitales/>
- Compensar. (27 de 07 de 2021). Plataformas digitales para ganar dinero extra. *Revista Compensar*, 2. Obtenido de <https://www.revistacompensar.com/comenzando/plataformas-ganar-dinero/>
- Costos BigQuery. (01 de 08 de 2021). *Costos BigQuery*. Obtenido de Costos BigQuery: <https://cloud.google.com/bigquery/pricing?hl=es-419>
- DAMA International. (2017). *DAMA-DMBOK 2nd Edition (2 ed.)*. (Deborah Henderson, CDMP, Ed.) Basking Ridge, NJ 07920 USA: Technics Publications.
- Data Studio. (01 de 07 de 2021). *Soporte Data Studio*. Obtenido de Soporte Data Studio: https://support.google.com/datastudio/answer/6283323?hl=es&ref_topic=6267740
- El Tiempo. (01 de 01 de 2021). El Tiempo Suscripcion. *El Tiempo Suscripcion*. Obtenido de https://seg.eltiempo.com/suscripciones/impreso/eltiempodigital?utm_source=SEM&utm_medium=google_cpc&utm_campaign=PautaAgosto&utm_content=PromoET_900CA&gclid=CjwKCAjwylKJBhBPEiwAu7zllz3ZGEOto84xbJkZDdqSHgEnKIWsRqO_5qG4F1u_yEZm7YU13n4zvRoCKaYQAvD_BwE
- Google Cloud. (01 de 08 de 2021). *Google Cloud*. Obtenido de Google Cloud: <https://cloud.google.com/docs/overview/?hl=es-419>
- Maida, E. G., & Pacienza, J. (01 de Diciembre de 2015). https://scholar.google.es/scholar?hl=es&as_sdt=0%2C5&q=metodolog%C3%ADas+de+desarrollo&btnG=&oq=metodologias+de+desa. Obtenido de <https://repositorio.uca.edu.ar/handle/123456789/522>
- Palma, D. (01 de 06 de 2021). *Grupodot*. (D. Palma, Productor) Obtenido de Grupodot: <https://www.grupodot.com/es/blog-black-box/monetizar-datos-en-tu-empresa/>

Patricia, K. (2020). Rivas Herera. *Marketing de Contenido Para Monetizar Canales de Comunicacion Digital*. Instituto Politecnico Nacional, Ciudad Mexico, Mexico. Obtenido de <https://www.escatop.ipn.mx/especialidadmkt/productividad-academica/tesis/karla-rivas.pdf>

Power BI. (01 de 08 de 2021). *Power BI*. Obtenido de Power BI: <https://powerbi.microsoft.com/es-es/>

Qlik Sense. (01 de 08 de 2021). *Qlik Sense*. Obtenido de Qlik Sense: <https://www.qlik.com/es-es/products/qlik-sense>

Ruiz, A., Villegas, M., & Hernández, H. (2012). https://scholar.google.es/schhp?hl=es&as_sdt=0,5. Recuperado el 2021, de https://d1wqtxts1xzle7.cloudfront.net/37369966/Estudio_de_Usabilidad_en_Aplicaciones_de_Comercio_Electronico_B2C_-_Amazon_-_Dell_-_MercadoLibre_vfinal-with-cover-page-v2.pdf?Expires=1635020025&Signature=JYscJ8N8HbASwj~xa1BQdaWUQCllwdrJva-3wopztdwoE~sdgntp

Visus, A. (01 de 10 de 2020). *www.esic.edu*. Obtenido de *www.esic.edu*: <https://www.esic.edu/rethink/tecnologia/para-que-sirve-python>