|  |
| --- |
| **Retail Store Analytics**  **Proyecto de Inteligencia de Negocios** |



**Autor**:

Bryan Gustavo Guapulema Arellano

**Plataforma:**

Talend Cloud

**Tabla de contenido**

[**1. Introducción 3**](#_Toc212151867)

[**1.1. Descripción del documento 3**](#_Toc212151868)

[**1.2. Criterios técnicos 3**](#_Toc212151869)

[**1.3. Criterios de evaluación 3**](#_Toc212151870)

[**1.4. Definiciones 3**](#_Toc212151871)

[**1.4.1. Retail Stores 3**](#_Toc212151872)

[**1.4.2. Talend Cloud(TC) 4**](#_Toc212151873)

[**2. Infraestructura del proyecto 5**](#_Toc212151874)

[**2.1. Fuentes de datos 5**](#_Toc212151875)

[**2.2. Herramientas y servicios 5**](#_Toc212151876)

[**2.3. Arquitectura 6**](#_Toc212151877)

[**3. Metodología y Desarrollo del Proyecto 7**](#_Toc212151878)

[**3.1. Ingesta de datos: Talend Data Inventory (TDI) 7**](#_Toc212151879)

[**3.2. Limpieza de datos (ETL): Talend Data Preparation 7**](#_Toc212151880)

[**3.3. Verificación de casos (ETL): Talend Data Stewardship 12**](#_Toc212151881)

[**3.4. Volcado de datos (Single Store) 15**](#_Toc212151882)

[**4. Visualización 18**](#_Toc212151883)

[**4.1. Conexión a Power BI 18**](#_Toc212151884)

[**4.2. Modelo de datos 18**](#_Toc212151885)

[**4.3. Medidas DAX 19**](#_Toc212151886)

[**5. Visualización 20**](#_Toc212151887)

[**5.1. Dashboard 20**](#_Toc212151888)

[**5.2. Historia construida a partir del dashboard 20**](#_Toc212151889)

[**6. Recursos 22**](#_Toc212151890)

# Introducción

## Descripción del documento

El presente documento tiene como objetivo describir de forma detallada el desarrollo técnico y analítico del proyecto de Inteligencia de Negocios “Retail Store Analytics”, implementado en la plataforma Talend Cloud.

Este informe recoge todos los procesos ejecutados para la creación de una solución end-to-end, que aborda el ciclo completo de vida de los datos desde la extracción y transformación de datos hasta la visualización final de indicadores comerciales.

El documento está estructurado para reflejar las fases del ciclo de vida del proyecto y los componentes utilizados en cada una de ellas. Además, se detallan las fuentes de información utilizadas, las herramientas de Talend empleadas, el modelo de datos creados, y el dashboard gerencial resultante.

## Criterios técnicos

* **Fuente Datos**: Utilizar una fuente de datos que contenga estructuras de tipo descriptivo y que presente oportunidades de mejora en la calidad de la información, por ejemplo: clientes, productos o proveedores.
* **Reglas**: Aplicar al menos 10 reglas de limpieza y validación de datos, orientadas a garantizar la consistencia, integridad y precisión de la información.
* **Integración Plataforma**: Durante el desarrollo, se deberán emplear todas las siguientes herramientas de la suite Talend: Talend Data Inventory (TDI), Talend Data Preparation (TDP), Talend Data Stewardship (DTS), Talend Management Console (TMC)

## Criterios de evaluación

* Extracción, transformación y carga correctamente ejecutadas en la plataforma elegida.
* Reglas de negocio a fuentes datos
* Integración de toda la suite Talend
* Claridad en la explicación de resultados final.
* Claridad en la explicación de resultados final, importancia de gobierno de datos dentro de las organizaciones.
* Soluciones novedosas, visualizaciones impactantes, enfoque original.

## Definiciones

## Retail Stores

Retail Stores es una empresa ficticia del sector comercial minorista (retail) dedicada a la venta de una amplia gama de productos, que incluye artículos de consumo masivo, tecnología, moda, hogar y entretenimiento. Opera a través de una red de tiendas físicas distribuidas en diferentes regiones, complementadas con un canal de ventas en línea que permite atender a clientes en todo el país.

La compañía basa su estrategia en la eficiencia operativa, la experiencia del cliente y la inteligencia de datos, utilizando la analítica avanzada para optimizar decisiones sobre inventario, precios, promociones y rentabilidad.

En el contexto del proyecto, Retail Stores representa una organización moderna que gestiona grandes volúmenes de datos transaccionales y de clientes

## Talend Cloud(TC)

Talend Cloud es una plataforma integral de integración y gobierno de datos en la nube que permite a las organizaciones conectar, transformar, limpiar, gobernar y compartir datos de forma segura y escalable.

Su arquitectura combina herramientas visuales, conectores listos para usar y capacidades de automatización que facilitan la creación de canales de datos confiables, alineados con principios de calidad, trazabilidad y cumplimiento normativo.

Talend Cloud está compuesto por varios módulos especializados que, trabajando en conjunto, cubren todo el ciclo de vida de los datos: desde su descubrimiento y preparación hasta su validación, monitoreo y explotación analítica.

1. **Talend Data Inventory (TDI)**

* Es el catálogo central de datos dentro de Talend Cloud.
* Permite conectar fuentes de datos (bases de datos, archivos, APIs, etc.) y explorar su contenido.
* Evalúa la calidad de los datos mediante indicadores automáticos (completitud, unicidad, validez, etc.) y permite documentar, clasificar y certificar datasets para promover su reutilización.
* Actúa como punto de partida para proyectos de preparación, limpieza o gobierno de datos.

1. **Talend Data Preparation (TDP)**

* Herramienta orientada al perfilado, limpieza y transformación de datos de manera visual e interactiva.
* Permite aplicar operaciones como normalización, reemplazo de valores, fusiones, cálculos, formatos y filtros sin necesidad de programar.
* Los datasets preparados pueden exportarse a otros entornos (como bases de datos, archivos o Data Stewardship) para ser usados en procesos analíticos o de gobierno.
* Facilita el trabajo de analistas de datos o business users en la mejora continua de la calidad de la información.

1. **Talend Data Stewardship (TDS)**

* Plataforma de gestión colaborativa de la calidad y validación de datos.
* Permite crear campañas de revisión donde los *data stewards* (responsables de datos) corrigen, validan o completan registros según reglas de negocio.
* Se integra con Data Preparation o Data Inventory para recibir los datos “dudosos” o con errores detectados.
* Asegura que los datos finales sean precisos, coherentes y aprobados antes de ser reutilizados.

1. **Talend Management Console (TMC)**

* Es el panel de administración central de Talend Cloud.
* Permite gestionar usuarios, roles, entornos, ejecuciones, planes y permisos.
* Controla el despliegue y la ejecución de tareas (jobs o pipelines), ya sea en la nube o mediante Remote Engines locales.
* Incluye herramientas de monitoreo, auditoría y automatización, esenciales para la gobernanza operativa del ecosistema Talend.

1. **Talend Open Studio (TOS)**

* Es la versión de escritorio (on-premise) de Talend, orientada a la creación de Jobs ETL/ELT (extracción, transformación y carga).
* Ofrece una interfaz gráfica basada en componentes reutilizables para construir flujos de integración complejos.
* Aunque no es parte directa del entorno cloud, se integra con Talend Cloud para subir y ejecutar Jobs diseñados localmente.
* Ideal para desarrolladores técnicos que requieren control detallado sobre los procesos de integración.

# Infraestructura del proyecto

## Fuentes de datos

|  |  |
| --- | --- |
| **Fuente de datos** | **CSV** |
| **Descripción:** | Archivos estructurados con información de clientes, productos y detalles de compra |
| **Conexión** | Disponible en: <https://github.com/BryanGuapulema/Hackaton-3>  Obtenido de: <https://www.kaggle.com/datasets/ahmedmohamed2003/retail-store-sales-dirty-for-data-cleaning> |
| **Contenido** | * mappings.csv * retail\_store\_sales\_details.csv |

## Herramientas y servicios

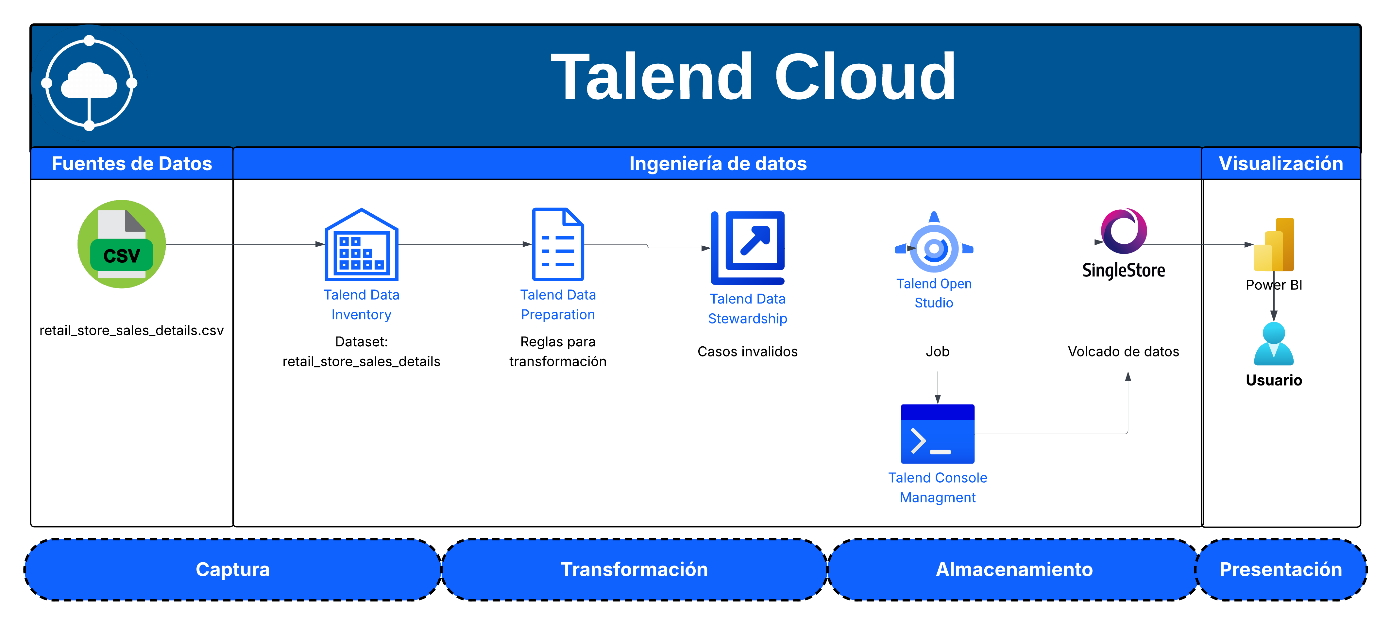
1. **Lenguajes**:

|  |  |
| --- | --- |
| **Lenguaje** | **Uso principal** |
| **SQL** | Consultas y transformaciones en los datasets publicados dentro de Talend Cloud y conexión con fuentes tabulares. |
| **Expresiones Talend / Preparations** | Aplicación de reglas, filtros y transformaciones dentro de Talend Data Preparation y Data Stewardship (condiciones, validaciones, sugerencias, etc.). |
| **DAX (Power BI)** | Cálculos analíticos y de control de calidad de datos en los dashboards (por ejemplo, % de registros válidos, completitud, cumplimiento de reglas). |

1. **Herramientas**

|  |  |  |
| --- | --- | --- |
| **Herramienta / Servicio** | **Función** | **Función principal** |
| **Talend Data Inventory (TDI)** | Descubrimiento / Evaluación | Punto de partida del proyecto. Permite conectar y explorar fuentes de datos, evaluar su calidad automática, clasificar columnas y documentar datasets. Desde aquí se identifican los registros o campos con baja calidad. |
| **Talend Data Preparation (TDP)** | Limpieza / Transformación visual | Espacio de trabajo interactivo y visual donde se preparan y limpian los datos detectados en TDI. Se aplican transformaciones, reemplazos, uniones y reglas de negocio. Los registros con valores sospechosos o sugeridos son enviados a Data Stewardship. |
| **Talend Data Stewardship (TDS)** | Validación colaborativa / Gobierno de datos | Herramienta de resolución colaborativa que permite crear campañas de validación para que los *data stewards* revisen, corrijan o aprueben registros. Permite controlar el estado de cada tarea (To Do, Resolved, Rejected). |
| **Talend Management Console (TMC)** | Administración / Ejecución | Panel de gestión centralizada donde se administran usuarios, permisos y pipelines. Permite ejecutar tareas y flujos de integración desde el Remote Engine Cloud y hacer seguimiento a la ejecución de procesos. |
| **Talend Open Studio (TOS)** | Integración / Desarrollo técnico | Entorno de desarrollo ETL local para diseñar y probar procesos de integración, exportación o carga de datos. Permite conectar con fuentes externas (por ejemplo, CSV o bases de datos) y preparar datasets para subir a Talend Cloud. |
| **Power BI Desktop** | Visualización | Conexión en modo Import a las vistas de Athena. Cálculos DAX para ventas acumuladas, brechas y porcentajes de cumplimiento de presupuesto. |

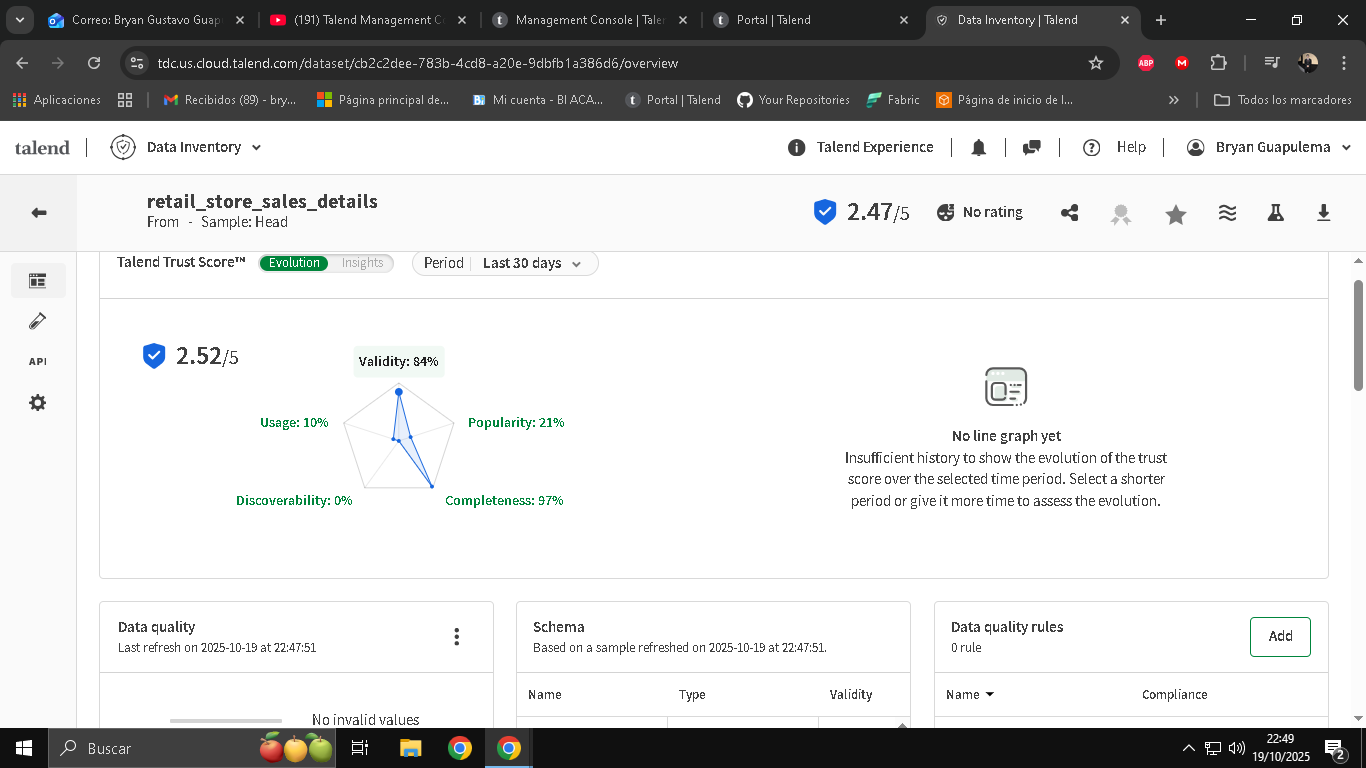
## Arquitectura



# Metodología y Desarrollo del Proyecto

## Ingesta de datos: Talend Data Inventory (TDI)

Tras la búsqueda de la fuente de información con las características indicadas se hizo la carga del archivo mediante un pipeline creado en Talend Pipeline Designer. Así pues, con la fuente de datos disponible se hizo la carga de esta en Talend Data inventory

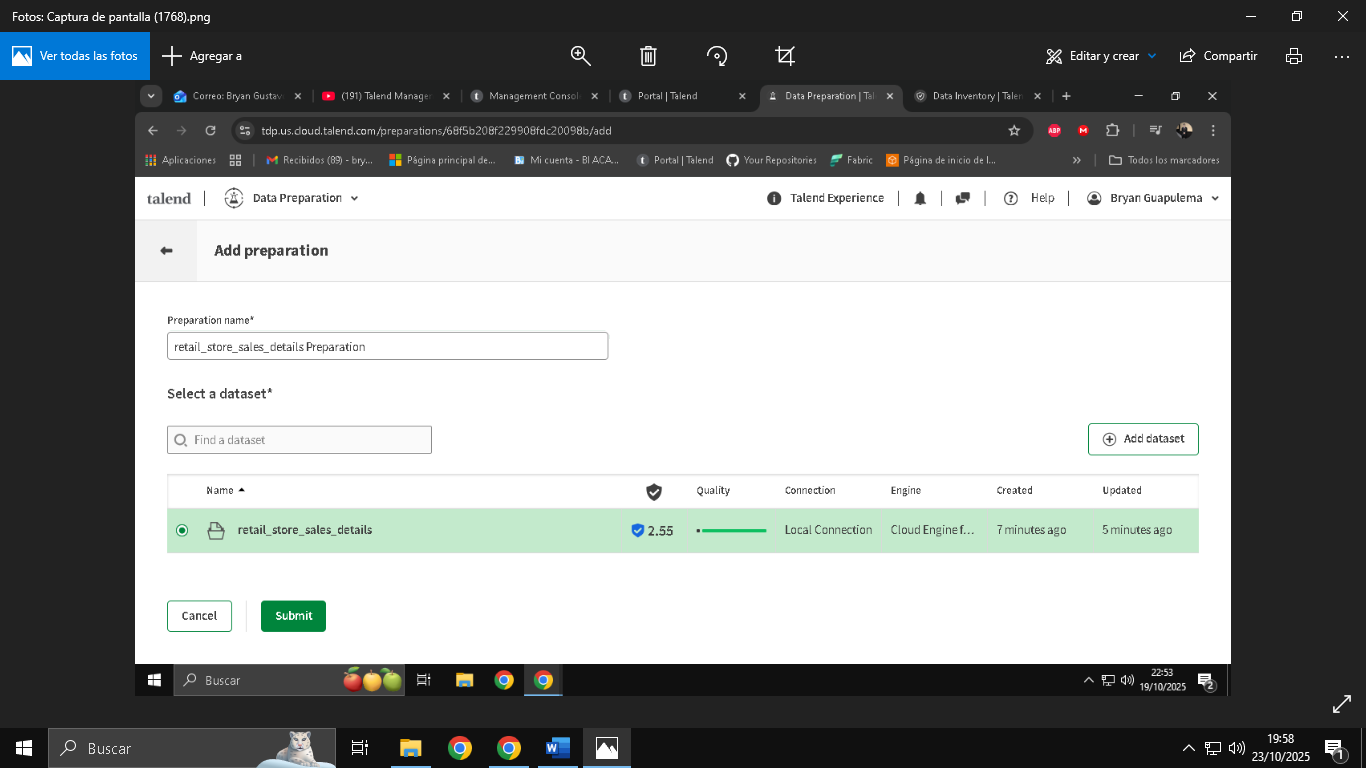


Como resultado de la carga se obtuvo un porcentaje de 2.47 puntos sobre 5 en lo referente a la calidad de los datos con notable presencia de valores inválidos, nulos y pobre documentación de semántica y metadata.

Adicionalmente se cargaron una fuente de mappings que contiene la información de los clientes y los productos por categoría, mismos que se unieron al dataset principal en Talend Data Preparation.

## Limpieza de datos (ETL): Talend Data Preparation

Para el proceso de limpieza se crea una receta en Talend Data Preparation basada en la fuente de datos almacenada en Talend Data Inventory.



Esta receta es la encargada de la limpieza de datos basada en reglas de calidad tanto de negocio como técnicas. Así pues, se definieron las siguientes reglas para la completitud, validación y limpieza del dataset.

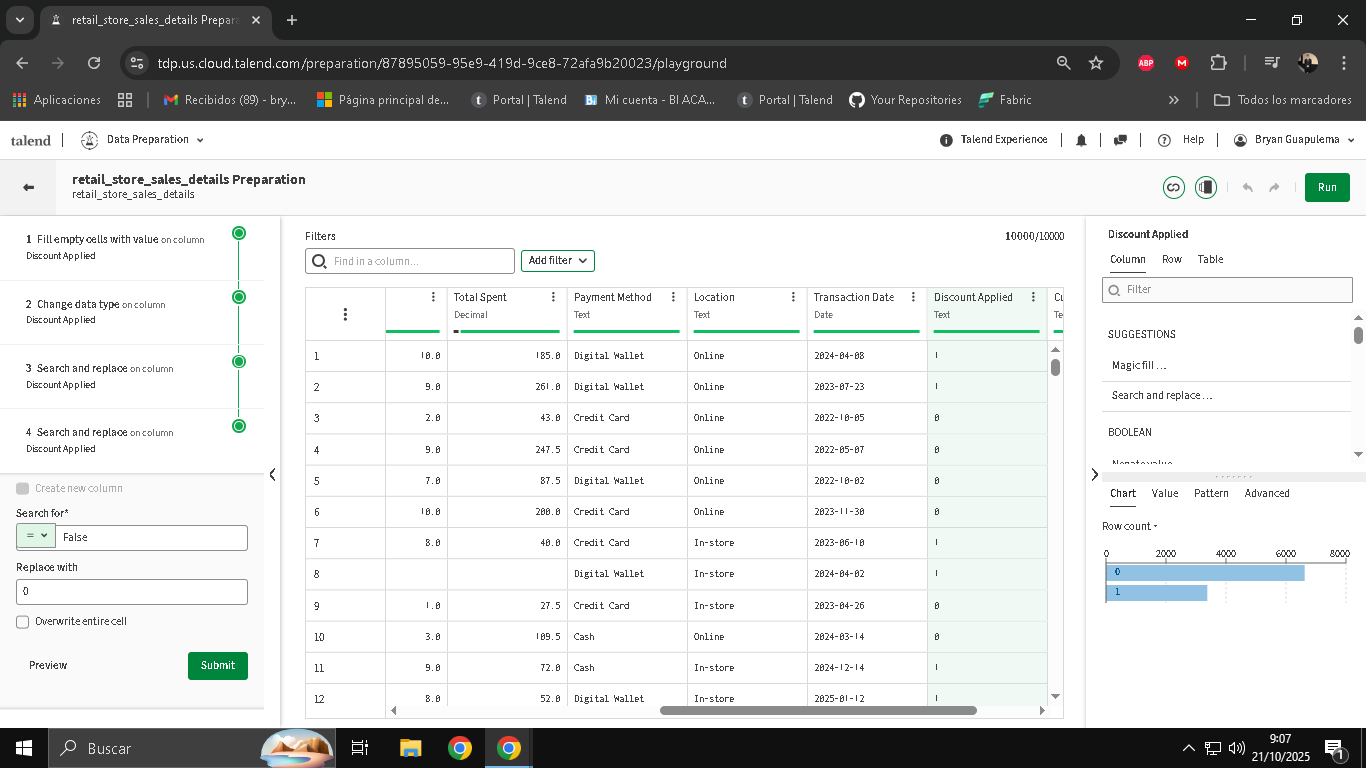
1. **Regla 1: Normalizar** **Discount Applied a 0/1**

Dado a que se encontraron múltiples valores para representar la presencia o ausencia de descuento en la compra como “True”,”False”,”No Discount”,”Yes”,”No” se normalizaron los valores que representan afirmaciones a 1 y las negaciones a 0 de modo que se tenga una bandera estandarizada

1. **Regla 2: Corregir valores nulos en la columna Discount Applied**

Para el tratamiento de valores nulos en la columna se interpretó el valor vacío como la ausencia de descuente, de modo que se imputaron los valores nulos con el valor de 0

Con estas dos reglas se unifica semántica y elimina 33% de nulos; hace medible el descuento en modelos y BI



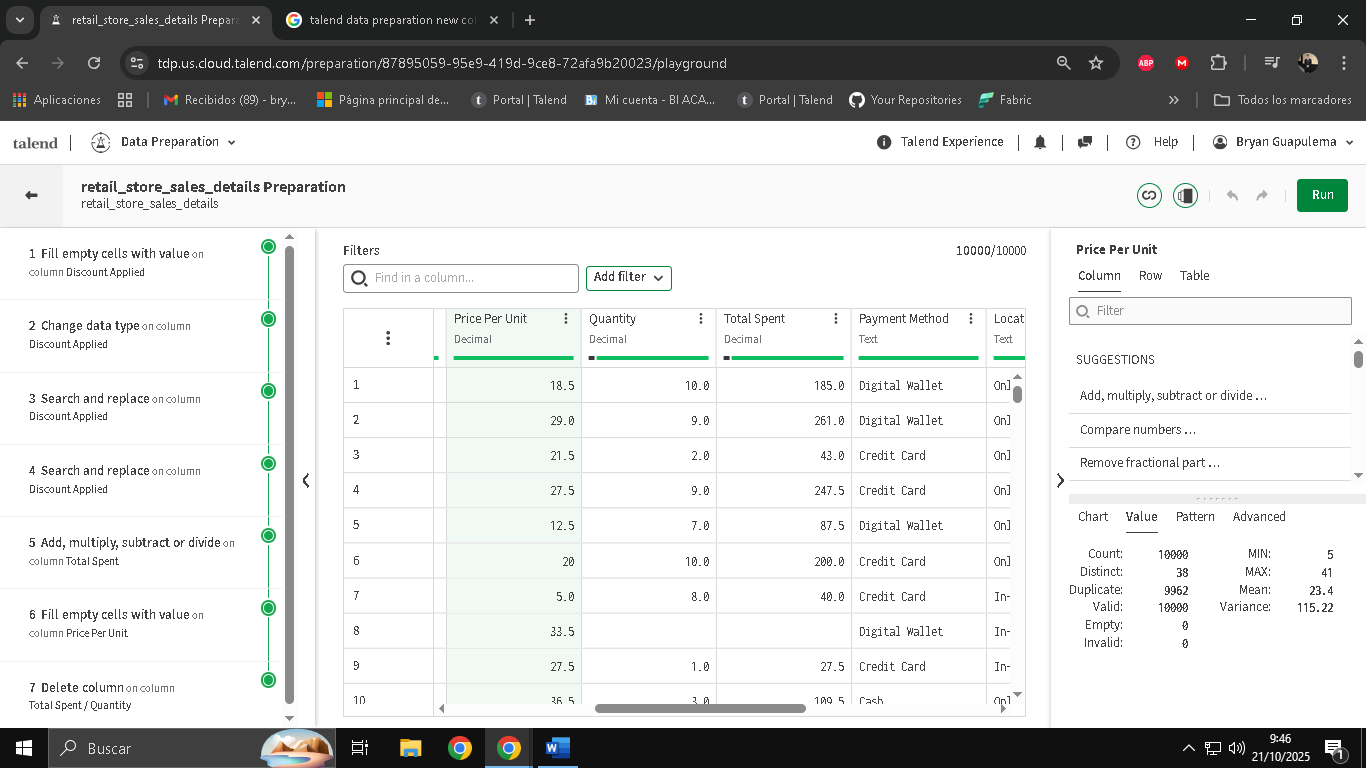
1. **Regla 3: Completar Price Per Unit (PPU) con cálculos**

Para la columna de precio unitario se noto la ausencia de 482 registros. Así pues para asegurar su completitud se calculo el precio unitario como la división entre el total gastado dividido para la cantidad para los casos en que esto sea posible

1. **Regla 4: Completar Price Per Unit (PPU) con mediana por categoría**

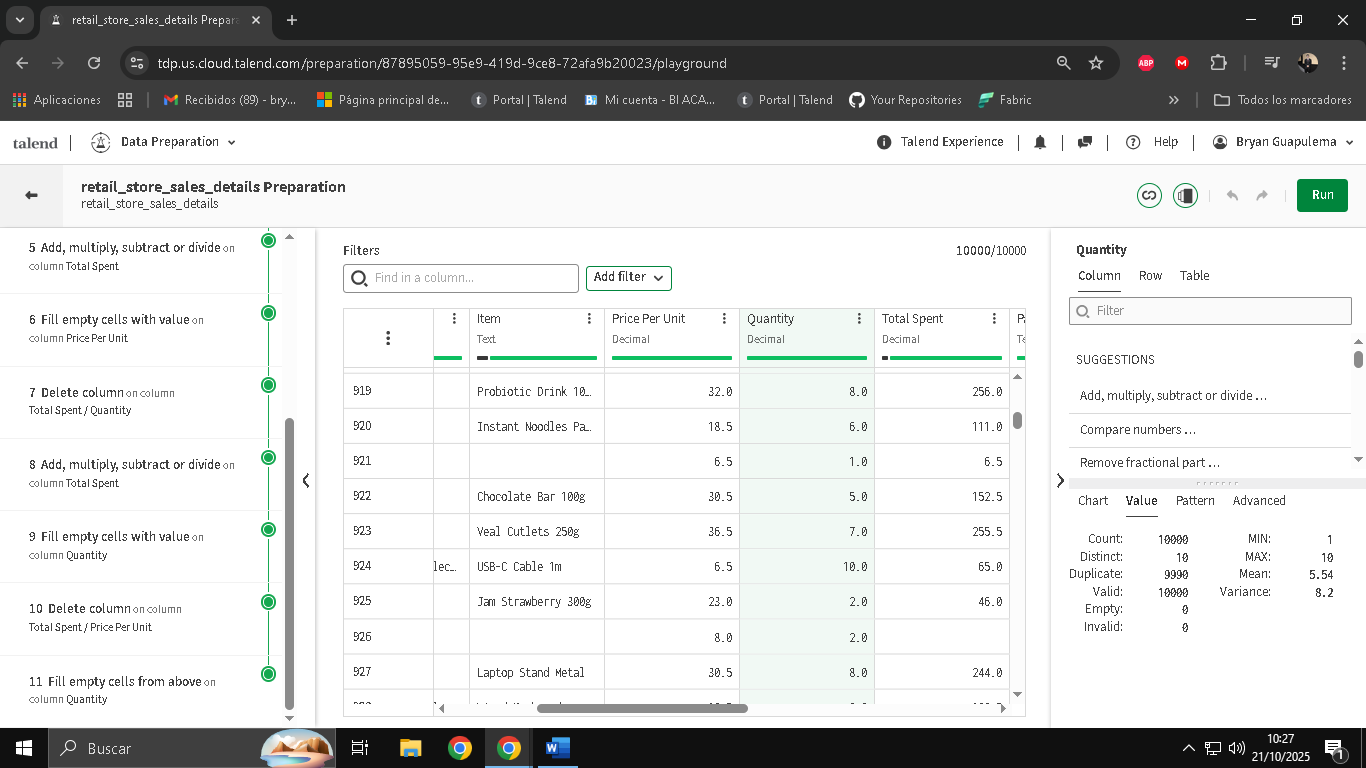
Para los casos de la columna de precio unitario que no se llenaron con la regla anterior se completaron mediante la imputación de precio por mediana de precios del producto basado en la categoría.

Estas reglas permitieron completar los valores faltantes y evitaron eliminar la porción de datos faltantes que podría representar un porcentaje significativo de la realidad de la empresa.



1. **Regla 5: Calcular Quantity si es posible**

Con la columna de precio unitario completa se ejecutó el llenado de la cantidad basado en la división del total gastado sobre la cantidad. En caso de no ser posible el cálculo se podría optar con la cantidad promedio basado en la categoría del producto. El fin es eliminar los valores nulos por completo en esta columna



1. **Regla 6: Total Spent = Quantity \* Price Per Unit**

Con las columnas de cantidad y precio unitario se hizo una validación de la columna del gasto total multiplicando la cantidad y el precio unitario de modo que se detecten irregularidades que puedan ser mandadas a revisar al Data StewardShip.

Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.

1. **Regla 7: Completar nulos en Total Spent**

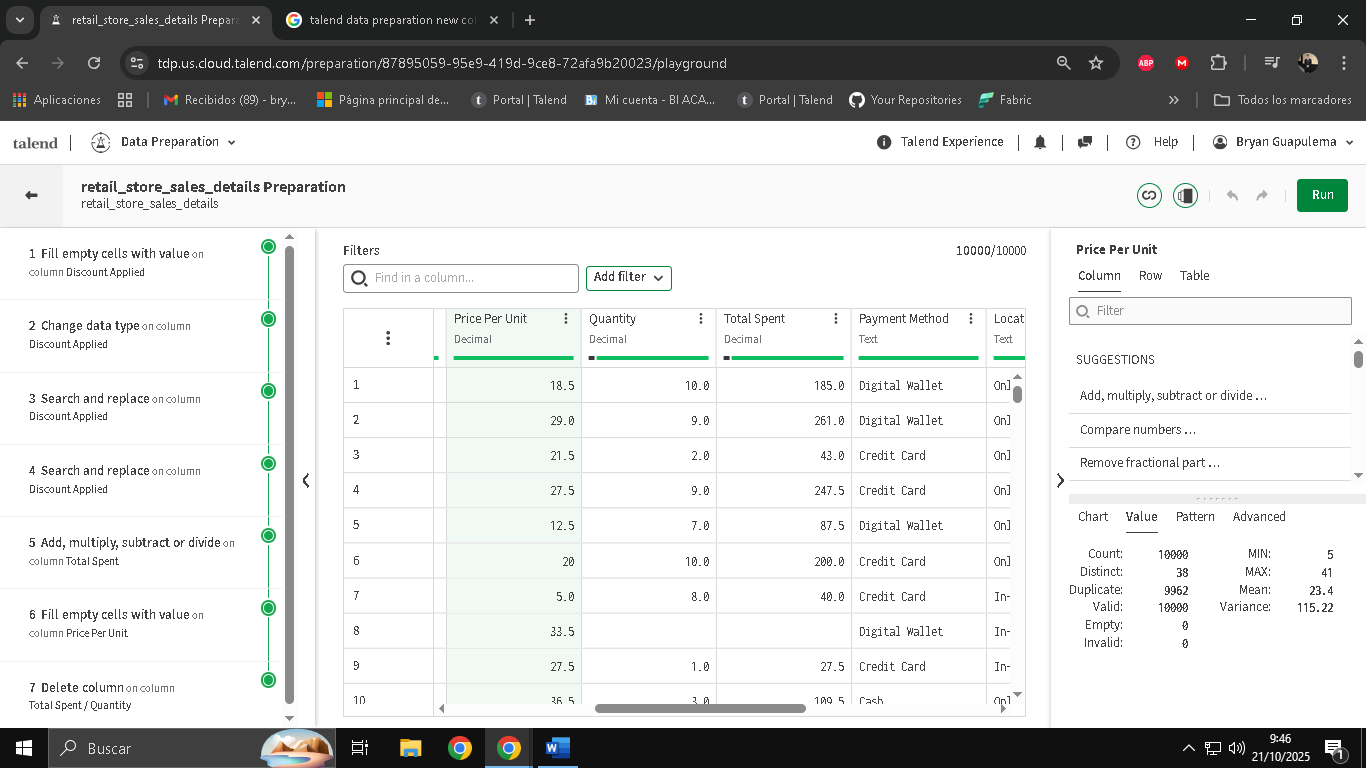
Tras la validación se completaron los valores nulos con el resultado de la multiplicación.

Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.

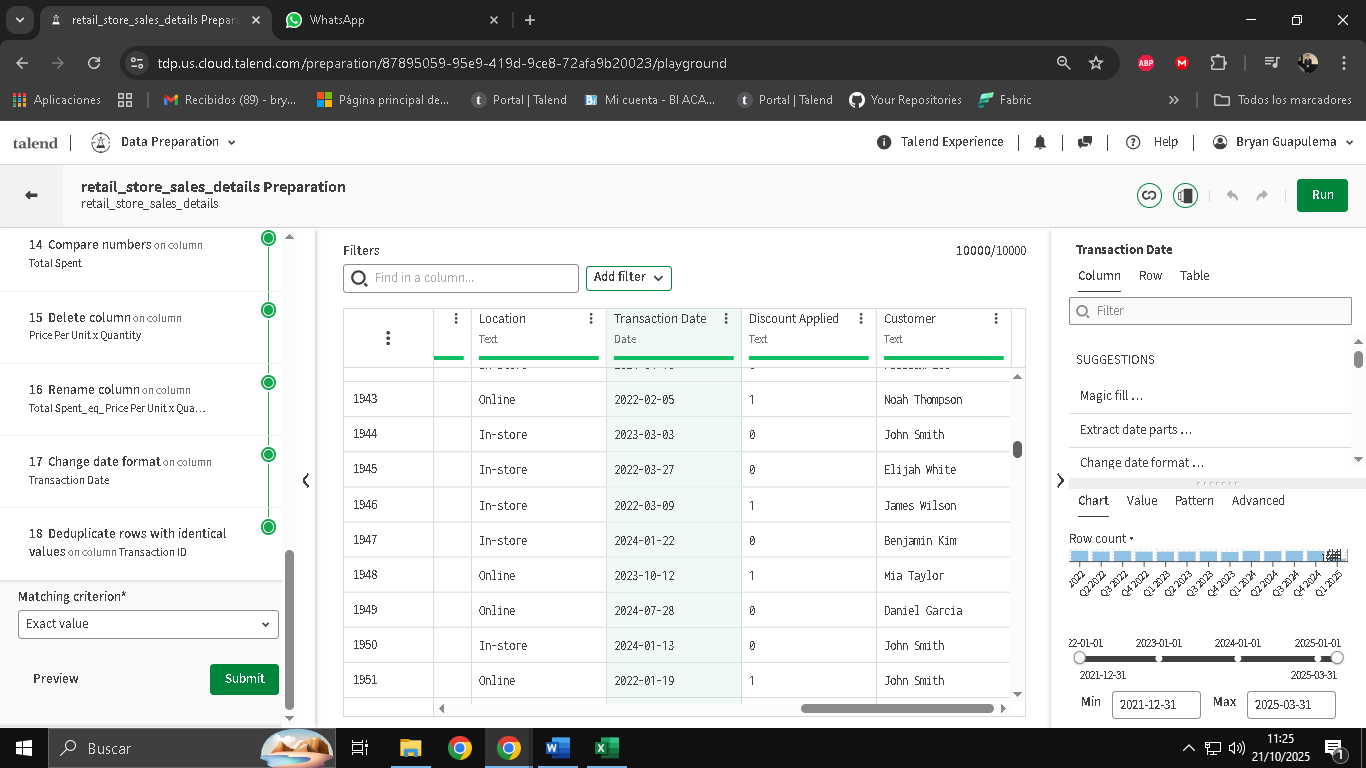
1. **Regla 8: Normalización de Paymeny Method**

Se establecieron todos los valores posibles dentro de 3 grupos “Cash”, “Credit Card” y “Digital Wallet”. Esto incluyó la conversión de todas las variantes de estos valores con mayúscula, minúsculas, entre otros.



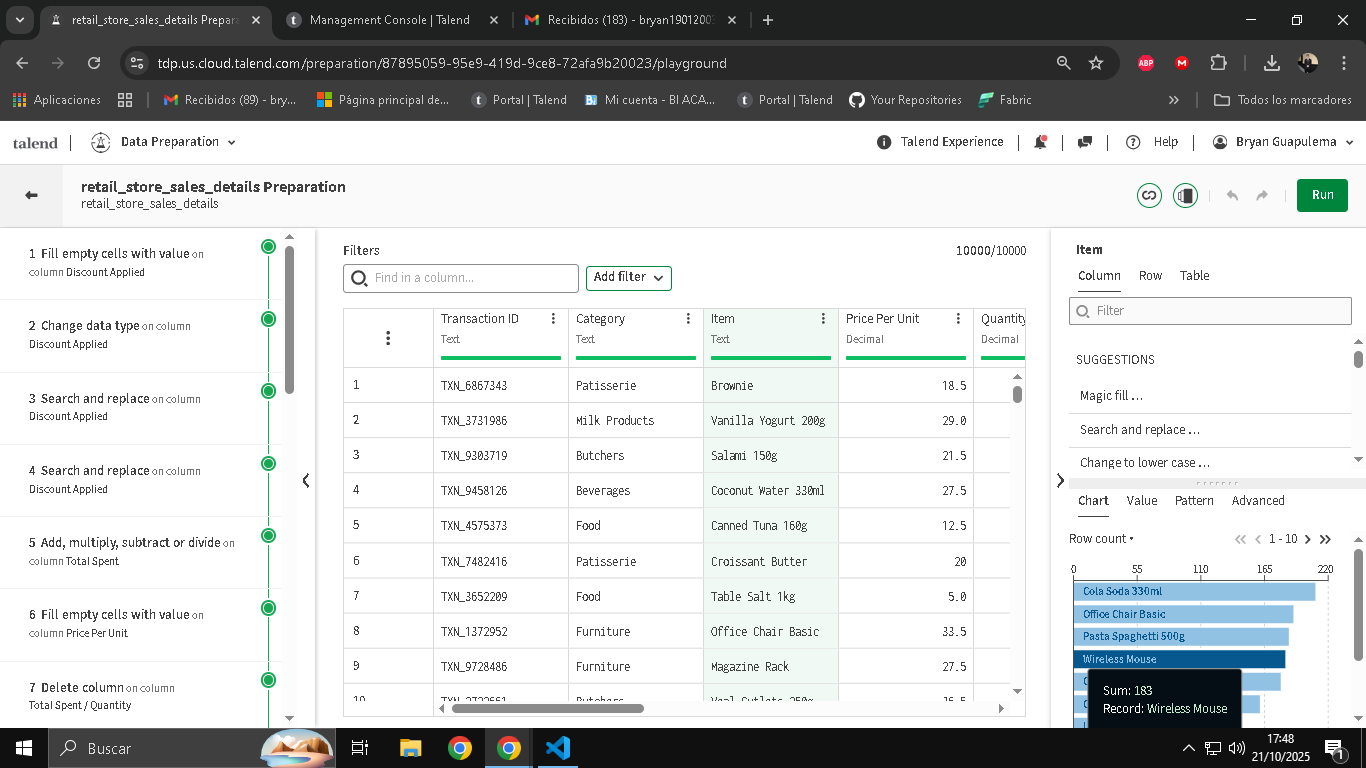
1. **Regla 9: Estandarizar Transaction Date y validar no-futuro**

Se convirtió la fecha al formato de fecha ISO con el formato “”yyyy-MM-dd”. Además se validó que la fecha sea coherente y sea menor que el día de ingesta. En caso de errores se dirige a Stewardship



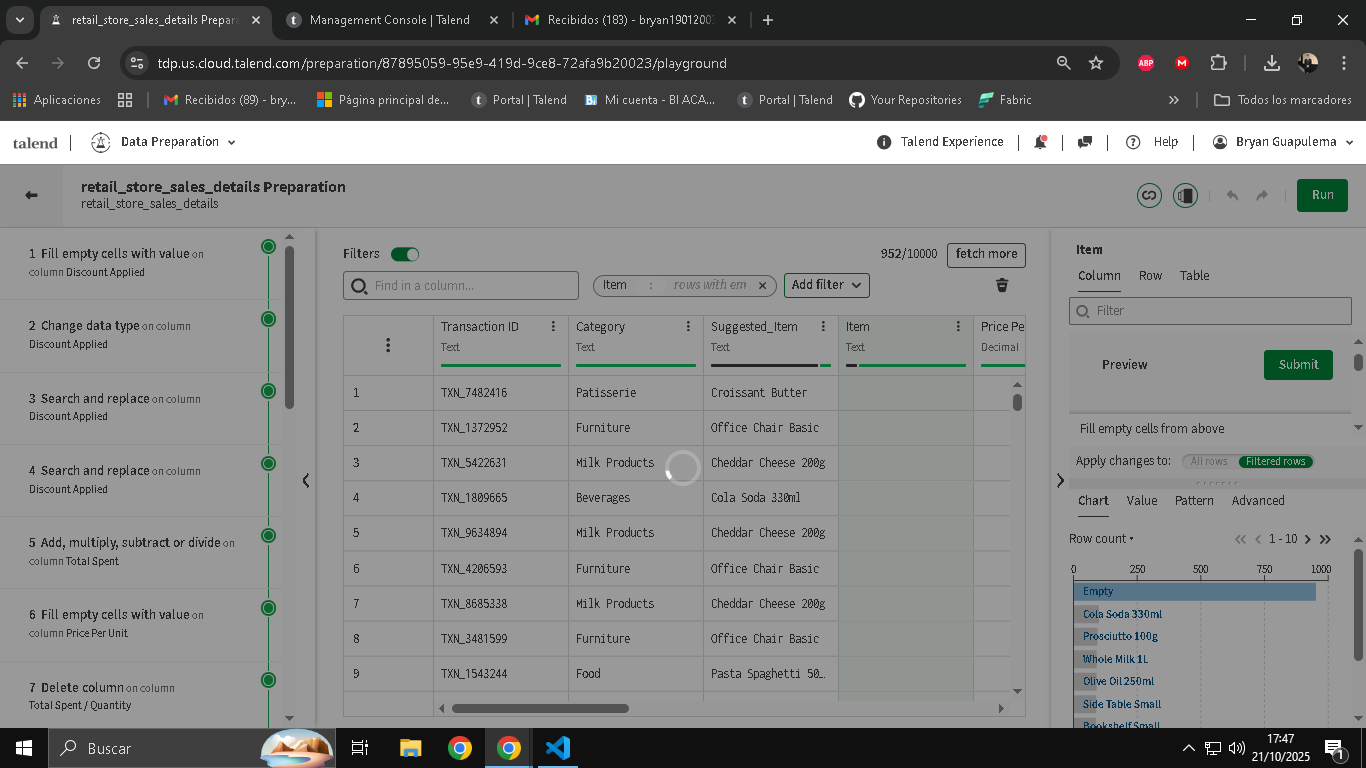
1. **Regla 10. Transaction ID obligatorio y único**

Adicionalmente se aplicó una deduplicación de los datos por Transaction ID a priori de eliminar duplicados exactos de fila.



1. **Regla 11: Ítem no puede estar vacío**

Finalmente se hizo el llenado de la columna Item. Pero esta no fue llenada de manera calculada pues corresponde a una decisión de negocio pues si se agrega al producto más vendido puede crearse outliers y si se agrega al de menos ventas puede crearse información falsa. Así pues, esta decisión el corresponde al custodio de la información: el Data Steward. Así pues, no se lleno directamente sino que se creó una columna Suggested Item con el producto más vendido para que sea validado e insertado o se cambie por el producto que se decida



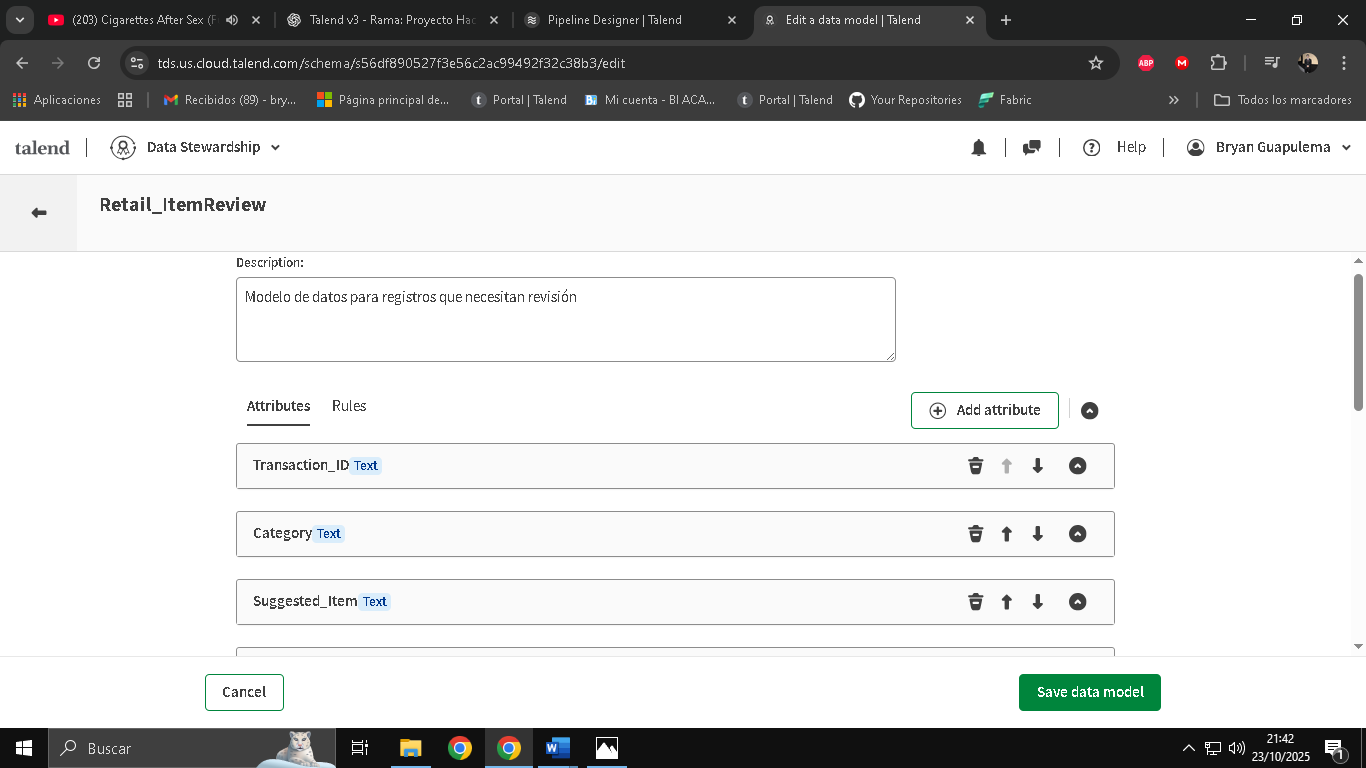
1. **Regla 12: Tipos coherentes**

Finalmente se eliminaron columnas auxiliares y se verificó que cada una de las columnas tenga un tipo de dato coherente con su contenido

**Nota**: se creó una columna bandera que registró las columnas que necesitaban revisión por ser excepciones a la regla

## Verificación de casos (ETL): Talend Data Stewardship

Para la revisión por parte del Data Steward fue necesario primero crear el modelo de datos con que la información que va a revisar cuenta:



Una vez creado el modelo se crea una campaña basada en el modelo que servirá para la conexión y asignación de tareas con los datos por validar el TDP.

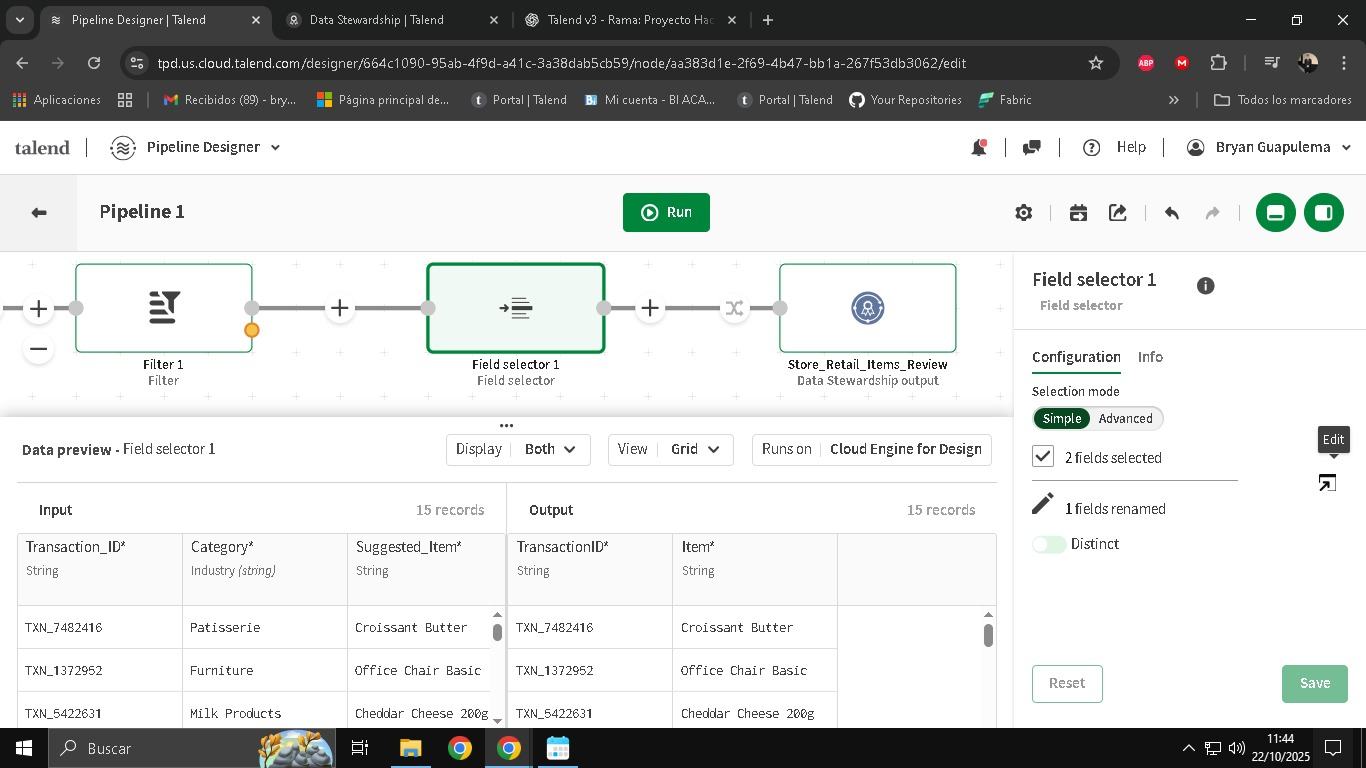
Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

El contenido generado por IA puede ser incorrecto.

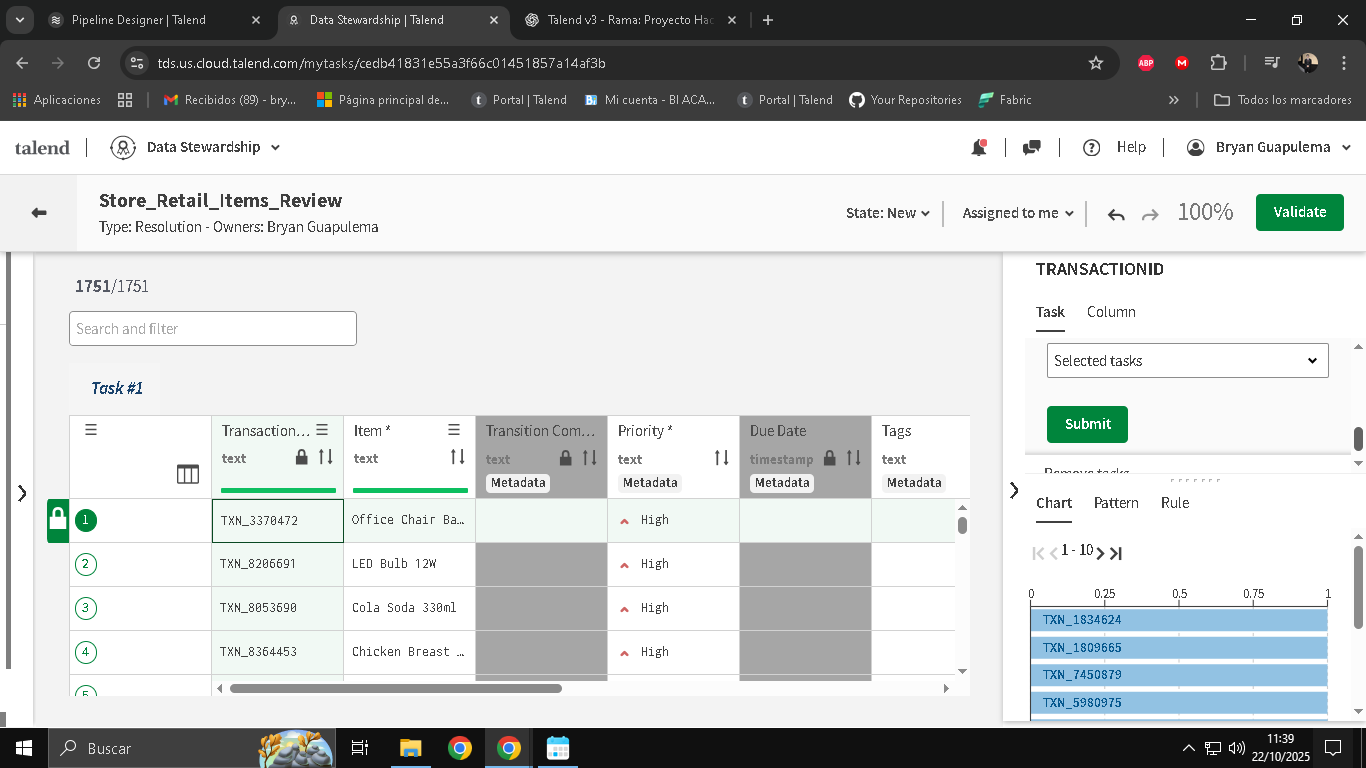
Interfaz de usuario gráfica, Aplicación, Sitio web

El contenido generado por IA puede ser incorrecto.

Para poder enviar los registros de TDP a TDS se usó Talend Pipeline Designer. Una herramienta incluida en la suite de Talend y que se encuentra en el plan gratuito de esta. Así pues, se creó un pipeline que envió los registros que necesitan revisión a la campaña creada a priori



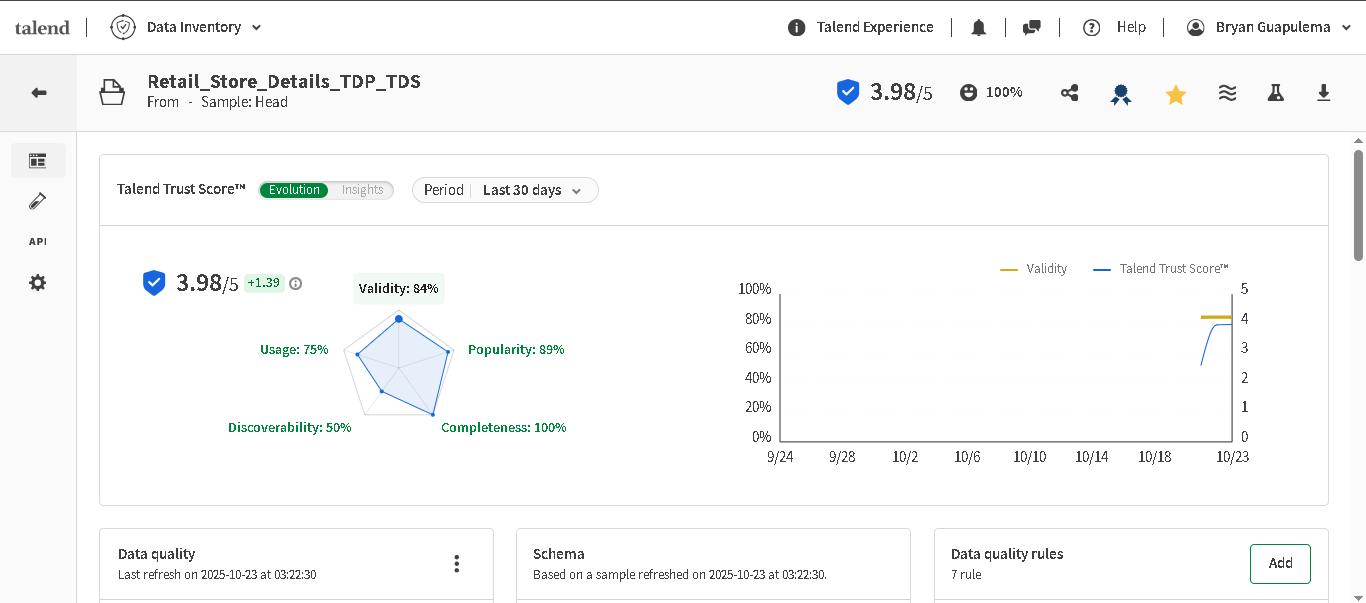
Tras ello se crean tareas con los registros que pueden ser revisados y validados por el usuario asignado. Así pues se revisaron y validaron dichos registros dejandolos con el estado resolved



Tabla

El contenido generado por IA puede ser incorrecto.

Finalmente se unieron los datos validos al inicio con los validados con TDS en un solo dataset y se exportaron a un dataset almacenado en TDI cuyas métricas aumentaron en medida

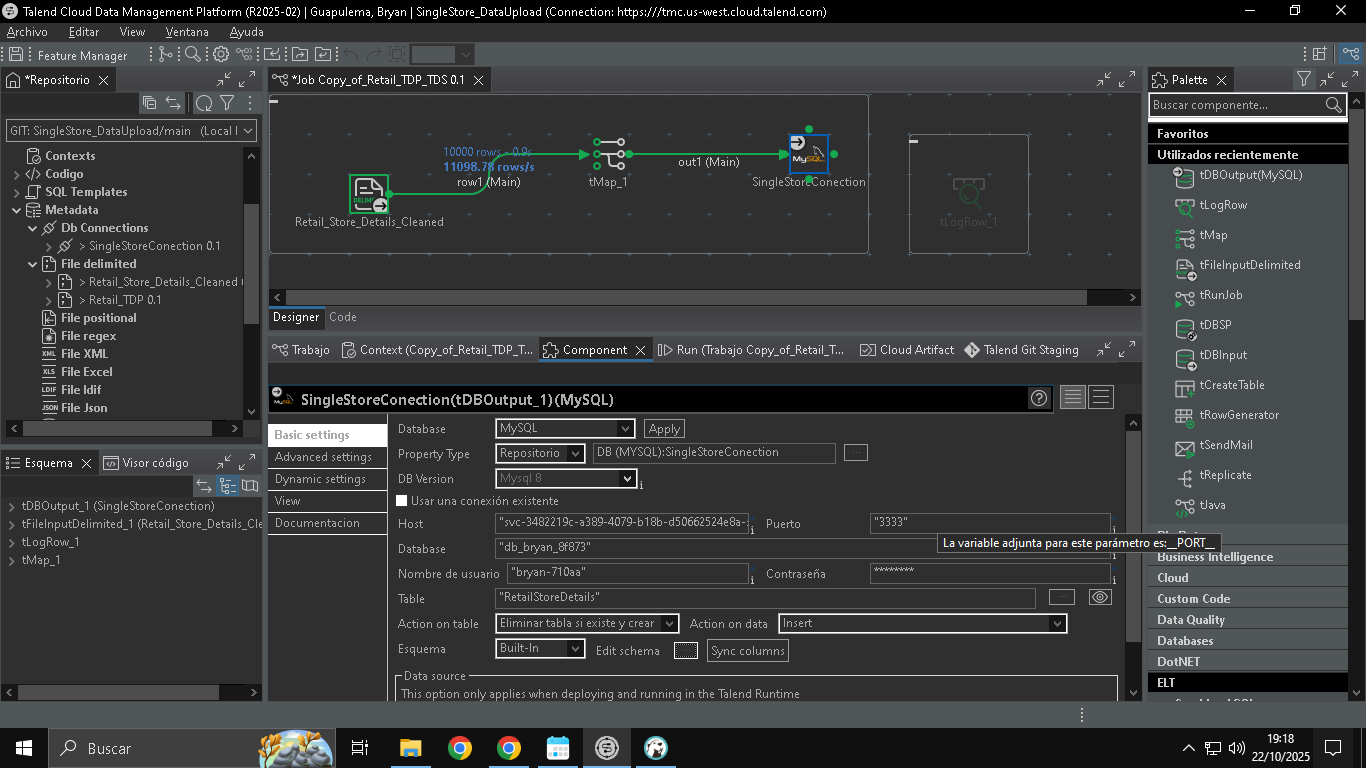


Nótese que el criterio de descubrimiento se ve afectado por el uso que se le da al dataset en la suite. Al ser el dataset final solo se usa para el volcado a single store mediante un job.

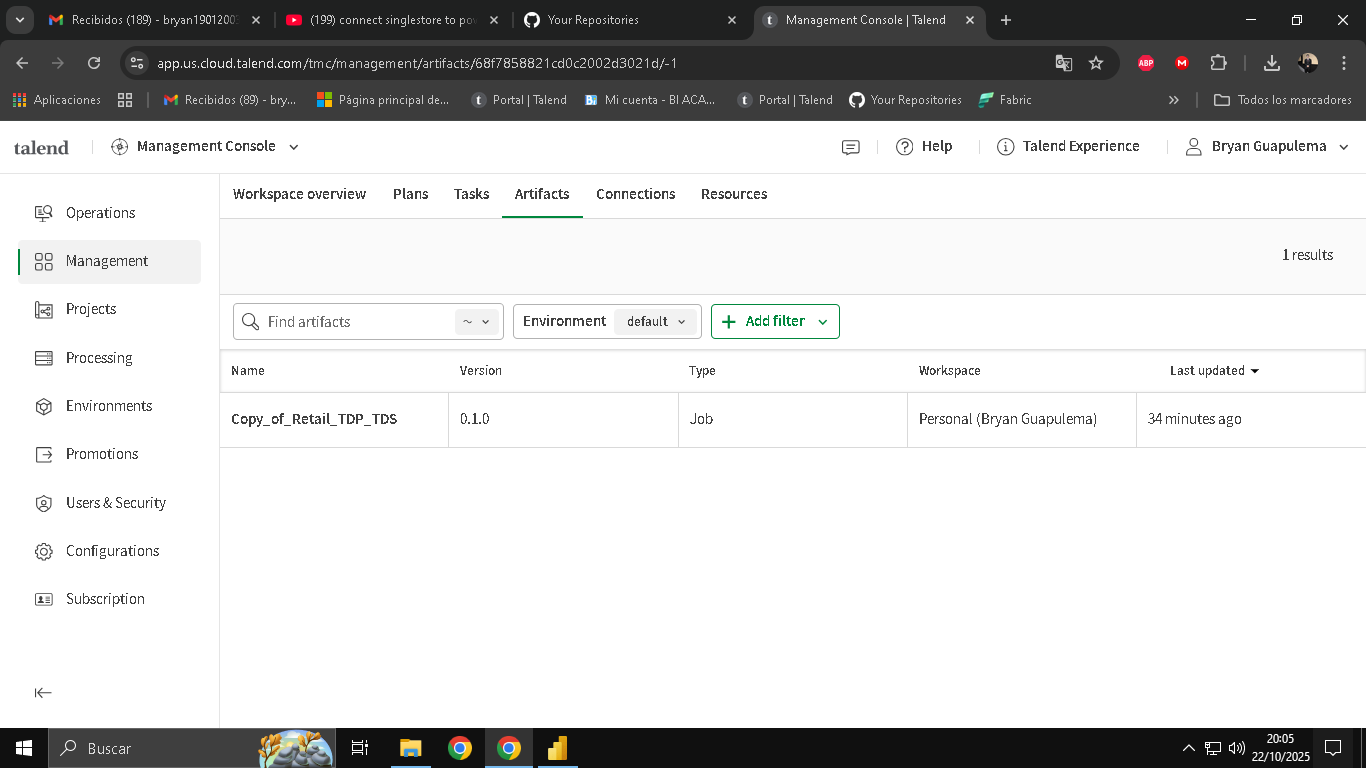
## Volcado de datos (Single Store)

Para el volcado de datos se utilizó la consola de Talend: TMC junto con Talend Open Studio (la última de las herramientas de la suite de Talend. Así pues se utilizó la consola para conectar la nube de Talend con el estudio local mediante la creación de u token en TMC.

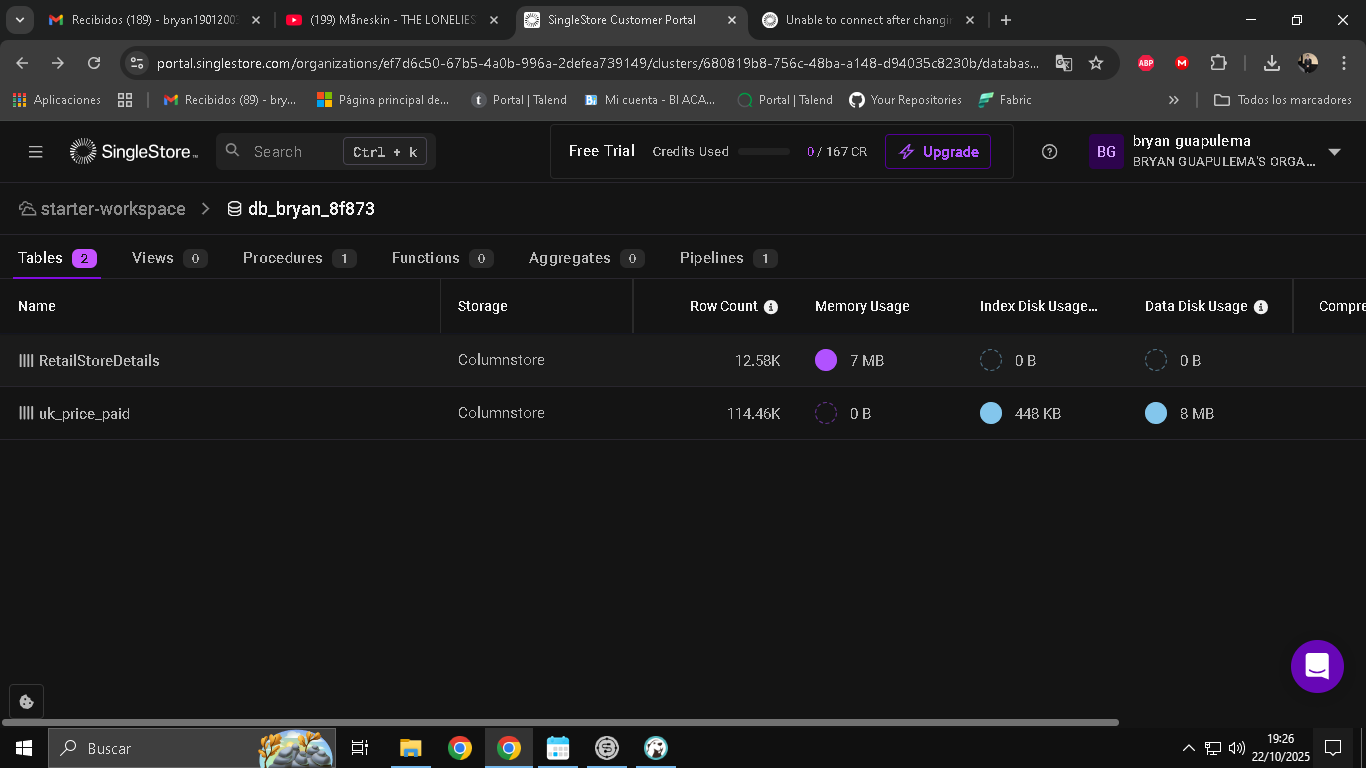
Así pues, se creó un job en Talend que consuma el dataset final generado, lo mapee asegurando el orden, semántica y tipo de datos y se publicó la tabla resultante a una base de datos en SinlgeStore previamente creada.



Finalmente se publicó el job a la nube de cloud como un artefacto que se puede ejecutar de manera automatizada



Finalmente se ejecutó el job desde Talend Managment Console y se verificó el volcado de datos en SingleStore desde DBeaver y en la propia plataforma de SingleStore



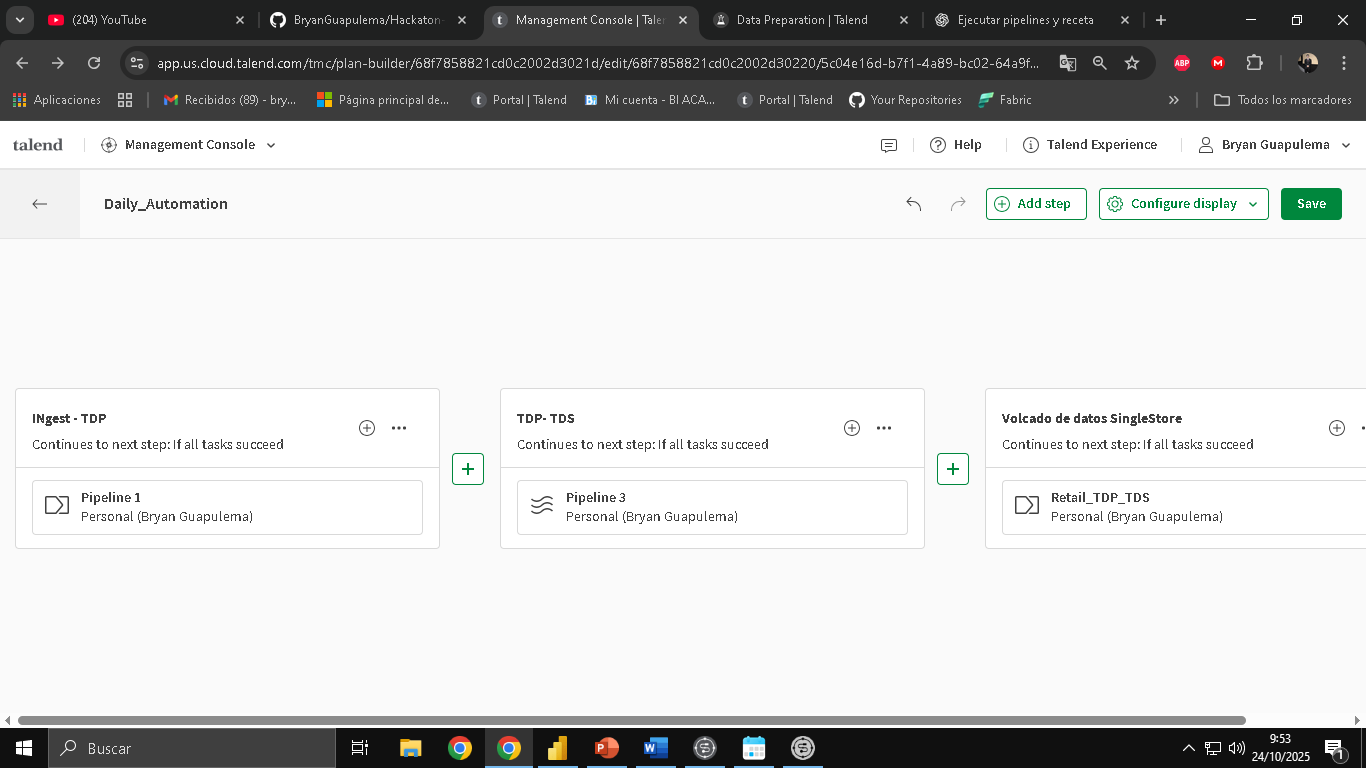
Captura de pantalla de computadora

El contenido generado por IA puede ser incorrecto.

Captura de pantalla de computadora

El contenido generado por IA puede ser incorrecto.

Finalmnete se configuro la carga automática creando un artefacto por cada pipeline y añadiéndolos a un plan en la consola de Talend TMC.



Finalmente se configuró la ejecución diaria mediante una tarea cron a ejecutarse a las 21:00 de cada día:

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

# Visualización

## Conexión a Power BI

Para conectar Single Store con Power BI se usó el driver de la base de datos incluida en el software de visualización. Para iniciar sesión se utilizaron las credenciales proporcionadas en la plataforma y se configuró el certificado SSL.

Se eligió el modo Import, que permite mayor rendimiento y refrescos del modelo cuando se actualiza con los nuevos meses cargados por la lambda incremental.

## Modelo de datos

Tras conectar la tabla de SingleStore Power BI Desktop, se construyó el esquema presentado enseguida (nótese que seestá usando una big table con toda la información de la empresa):

Imagen que contiene Escala de tiempo

El contenido generado por IA puede ser incorrecto.

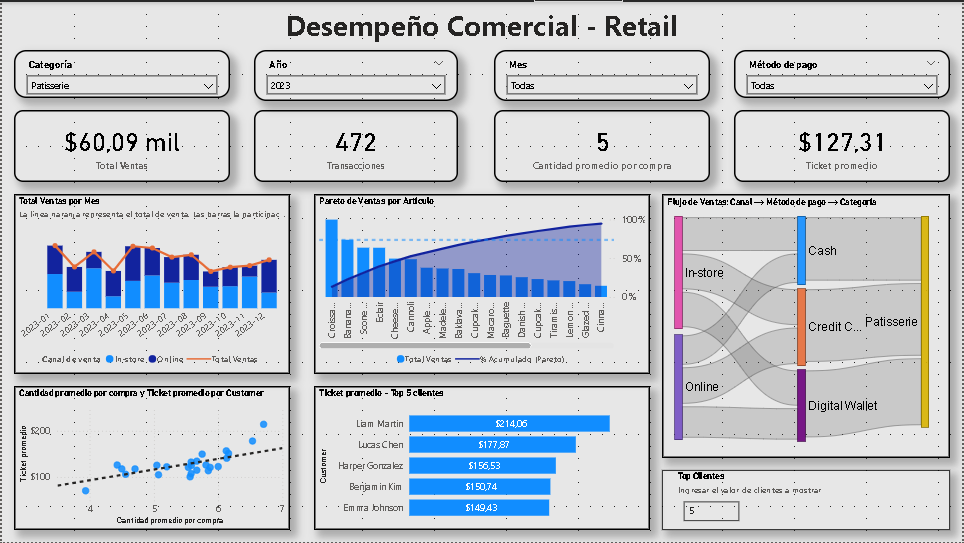
## Medidas DAX

Se definió la siguiente lista de medidas dentro de una tabla contendora de medidas:

|  |  |
| --- | --- |
| **Medida** | **Fórmula** |
| **% Acumulado (Pareto)** | % Acumulado (Pareto) =  DIVIDE([Ventas acumuladas (Pareto)], [Ventas total visible (Pareto)]) |
| **% Participación canal** | % Participación canal =  DIVIDE([Ventas (contexto mes)], [Ventas total mes (todos los canales)]) |
| **Cantidad** | Cantidad = SUM(RetailStoreDetails[Quantity]) |
| **Cantidad promedio por compra** | Cantidad promedio por compra = DIVIDE([Cantidad], [Transacciones]) |
| **Cantidad promedio por compra** | Clientes únicos = DISTINCTCOUNT(RetailStoreDetails[Customer]) |
| **Mostrar cliente (Top N)** | Mostrar cliente (Top N) =  IF( [Ranking cliente (Ticket Promedio)] <= SELECTEDVALUE('TOP'[TOP]), 1, 0 ) |
| **Ticket promedio** | Ticket promedio = DIVIDE([Total Ventas], [Transacciones]) |
| **Título gráfico Top N** | Título gráfico Top N =  "Ticket promedio - Top " & SELECTEDVALUE('TOP'[TOP]) & " clientes" |
| **Total Ventas** | Total Ventas = (SUM(RetailStoreDetails[Total\_Spent])) |
| **Transacciones** | Transacciones = DISTINCTCOUNT(RetailStoreDetails[Transaction\_ID]) |
| **Valor Sankey =** | Valor Sankey =  VAR origen  = SELECTEDVALUE('Enlaces Sankey'[Origen])  VAR destino = SELECTEDVALUE('Enlaces Sankey'[Destino])  VAR isCanal\_Metodo =      origen  IN VALUES(RetailStoreDetails[Channel]) &&      destino IN VALUES(RetailStoreDetails[Payment\_Method])  VAR isMetodo\_Categoria =      origen  IN VALUES(RetailStoreDetails[Payment\_Method]) &&      destino IN VALUES(RetailStoreDetails[Category])  RETURN  SWITCH(      TRUE(),      isCanal\_Metodo,          CALCULATE([Total Ventas],              RetailStoreDetails[Channel] = origen,              RetailStoreDetails[Payment\_Method] = destino          ),      isMetodo\_Categoria,          CALCULATE([Total Ventas],              RetailStoreDetails[Payment\_Method] = origen,              RetailStoreDetails[Category] = destino          ),      BLANK()  ) |
| **Ventas (contexto mes)** | Ventas (contexto mes) = [Total Ventas] |
| **Ventas acumuladas (Pareto)** | Ventas acumuladas (Pareto) =  VAR r = [Ranking Item]  VAR Tabla =      ADDCOLUMNS(          ALLSELECTED(RetailStoreDetails[Item]),          "VentasItem", CALCULATE([Total Ventas])      )  VAR TopR =      TOPN(r, Tabla, [VentasItem], DESC)  RETURN SUMX(TopR, [VentasItem]) |

# Visualización

## Dashboard



## Historia construida a partir del dashboard

Este proyecto contempla el análisis de una tienda de retail especializada en productos de consumo, con información transaccional comprendida entre los años 2022 y 2025. El objetivo principal fue evaluar el desempeño comercial durante este periodo, identificando los factores que impulsan las ventas, los productos más rentables y los clientes de mayor valor para el negocio.

Una visión general del desempeño muestra que en este periodo, la tienda ha registrado ventas totales por 60 mil dólares, distribuidas en 472 transacciones, con un ticket promedio de 127 dólares y una cantidad promedio de cinco unidades por compra, lo que refleja una estructura de ventas saludable y sostenida.

En la parte superior del dashboard se pueden seleccionar los filtros por categoría, año, mes y método de pago, lo que permite analizar de manera dinámica el comportamiento comercial bajo distintos escenarios.

Ahora bien, al observar la evolución de ventas por mes, notamos un comportamiento estacional pero con una tendencia general de crecimiento, especialmente en el canal Online, que ha ganado terreno frente a la tienda física. Este aumento del canal digital demuestra que los clientes están adoptando cada vez más las compras en línea, lo cual abre oportunidades para fortalecer las estrategias de marketing digital, descuentos web y experiencia omnicanal.

Sin embargo, se observa que ciertos meses —como marzo y octubre— presentan leves caídas en el volumen total de ventas, lo que podría relacionarse con ciclos de consumo y con la falta de campañas promocionales específicas durante esos periodos.

Pasando al análisis de productos, el gráfico de Pareto muestra que un grupo reducido de artículos concentra la mayoría de los ingresos. En concreto, se confirma la regla 80/20: el 20% de los productos genera aproximadamente el 80% de las ventas. Esto significa que la rentabilidad del negocio está altamente concentrada en pocos artículos, por lo que las decisiones de reposición, marketing y descuentos deben centrarse en ese núcleo de productos estrella. Al mismo tiempo, los productos del tramo inferior del Pareto deberían ser revisados para determinar si conviene mantenerlos, liquidarlos o reemplazarlos por variantes más rentables.

Si analizamos el flujo de ventas desde el canal hasta la categoría, podemos ver cómo se comporta el cliente en su proceso de compra. En el Sankey diagram se observa que el canal Online representa el flujo más grande, con predominio de pagos con tarjeta de crédito, especialmente en la categoría Patisserie. En cambio, las ventas In-store muestran una mayor proporción de pagos en efectivo. Este análisis es clave porque revela las preferencias de los consumidores según el canal: mientras los clientes online valoran comodidad y medios de pago digitales, los clientes presenciales prefieren compras inmediatas y pagos tradicionales.

En la parte inferior izquierda, el gráfico de dispersión relaciona la cantidad promedio comprada con el ticket promedio por cliente, lo cual permite identificar los distintos tipos de comportamiento.

En la zona superior derecha del gráfico se ubican los clientes VIP, aquellos que compran con alta frecuencia y alto ticket promedio; en la zona inferior derecha, los premium ocasionales, que gastan mucho pero compran con menor frecuencia; y en la parte superior izquierda se ubican los clientes frecuentes de bajo ticket, que podrían ser fidelizados con promociones o descuentos.

Este análisis es fundamental para segmentar a los clientes y establecer estrategias diferenciadas: por ejemplo, programas de fidelización para los VIP, combos para los frecuentes de bajo ticket y campañas de recompra para los premium ocasionales.

En la parte inferior derecha se observa el Top de clientes con mayor ticket promedio, el cual es totalmente dinámico gracias al parámetro Top N. En este caso, los clientes Liam Martin y Lucas Chen encabezan el ranking con tickets promedio superiores a los 170 dólares. Este tipo de visualización permite concentrar los esfuerzos de fidelización y recompra en los clientes que más contribuyen a la rentabilidad general del negocio.

Finalmente, si observamos el conjunto de KPIs y tendencias, podemos concluir que el negocio mantiene una estructura comercial saludable, con crecimiento en el canal online, altas oportunidades de segmentación de clientes, y una cartera de productos concentrada pero altamente rentable.

En términos estratégicos, las recomendaciones principales son:  
primero, fortalecer el canal online mediante campañas dirigidas y beneficios de pago digital;  
segundo, optimizar el catálogo de productos, priorizando los del tramo superior del Pareto y revaluando los de baja contribución; y tercero, desarrollar programas de fidelización segmentados por tipo de cliente para aumentar el valor promedio de cada compra.

# Recursos

**Documentación completa:** <https://github.com/BryanGuapulema/Hackaton-3>