

PROYECTO FINAL – CURSO DE ETL Y CIENCIA DE DATOS



**LUIS CARLOS CORREA ZULUAGA
CODIGO: 22501541**

**UNIVERSIDAD AUTÓNOMA DE OCCIDENTE
FACULTAD DE INGENIERÍA Y CIENCIAS BÁSICAS
MAESTRIA EN INTELIGENCIA ARTIFICIAL Y CIENCIA DE DATOS
2025**

PROYECTO FINAL – CURSO DE ETL Y CIENCIA DE DATOS



**LUIS CARLOS CORREA ZULUAGA
CODIGO: 22501541**

**ETL (EXTRACCION, TRANSFORMACION Y CARGA)
CODIGO: 551464**

**Profesor
FERNANDO BARRAZA ALVARADO**

**UNIVERSIDAD AUTÓNOMA DE OCCIDENTE
FACULTAD DE INGENIERÍA Y CIENCIAS BÁSICAS
MAESTRIA EN INTELIGENCIA ARTIFICIAL Y CIENCIA DE DATOS
CALI - VALLE DEL CAUCA
2025**

CONTENIDO	PÁG.
1. CONTEXTO	6
2. DESCRIPCIÓN DEL PROBLEMA	6
3. JUSTIFICACION DEL PROYECTO	7
4. IDENTIFIACION INICIAL DE FUENTES	8
5. DEFINICIÓN DEL PROBLEMA.....	8
6. EXPLORACION INICIAL DE DATOS.....	9
7. EXTRACCIÓN, TRANSFORMACIÓN Y VALIDACIÓN DE DATOS.....	10

1. CONTEXTO

Actualmente laboro en la empresa Nutriavicola SAS la cual me dio el permiso para utilizar su información comercial, con las siguientes condiciones:

- Se firma una cláusula de confidencialidad
- Se deberá cambiar los nombres de sus clientes por genéricos (Cliente 1, 2...), renombrar zonas de ventas y sus canales de distribución.
- La información que me suministrará será con información no actual para el análisis y el desarrollo del proyecto. Pero con estructura para poderla implementar a futuro.

Información de la empresa:

Nutriavicola SAS es una empresa vallecaucana que se dedica a la producción y distribución de huevos libres de jaula, centrada en el bienestar animal. Sus prácticas incluyen galpones al aire libre, dietas de origen vegetal, sin antibióticos y altos estándares de bienestar para las gallinas.

Nutriavicola SAS comercializa huevos bajo la marca Huevos Oro, y distribuye sus productos a través de diferentes canales de comercialización a nivel nacional.

2. DESCRIPCIÓN DEL PROBLEMA

Actualmente, Nutriavicola SAS enfrenta desafíos en la recopilación, transformación y análisis de los datos generados por su equipo de ventas, lo que dificulta la identificación de tendencias, oportunidades de mejora y la medición del impacto de sus estrategias comerciales. La falta de un sistema centralizado para la gestión de indicadores clave de desempeño (KPI's) limita la capacidad de la gerencia para tomar decisiones oportunas y basadas en información confiable.

La empresa cuenta con más de 100 vendedores a nivel nacional y más de 9,500 clientes activos por mes. Sin una herramienta adecuada para centralizar y visualizar la información, el seguimiento del desempeño comercial se vuelve complejo, lo que dificulta la identificación de tendencias, el monitoreo de ventas y la optimización de estrategias comerciales.

Por esta razón, Nutriavicola SAS requiere el desarrollo de un dashboard en Power BI , la herramienta oficial de la empresa, que permita consolidar los datos de ventas y generar reportes dinámicos con KPI's relevantes. Este sistema proporcionará a la gerencia y dirección información confiable y en tiempo real para mejorar la toma de decisiones, optimizar la gestión de la fuerza de ventas y aumentar la eficiencia operativa.

3. JUSTIFICACION DEL PROYECTO

La implementación de un dashboard en Power BI permitirá:

- Automatizar la recolección y procesamiento de datos provenientes de diferentes fuentes, reduciendo errores manuales y optimizando tiempos de análisis.
- Proporcionar una visualización clara e interactiva de los KPI's que facilitan la toma de decisiones estratégicas en tiempo real.
- Mejorar el seguimiento y desempeño de la fuerza de ventas, identificando oportunidades de mejora y reconociendo el esfuerzo de los vendedores con mejores resultados.
- Fortalecer la planificación comercial mediante el análisis de tendencias en las ventas y comportamiento de los clientes.

Este proyecto no solo permitirá mejorar la eficiencia operativa de la empresa, sino que también impulsará el crecimiento y la rentabilidad al proporcionar herramientas avanzadas de análisis de datos para la optimización de la gestión comercial.

Con la realización del dashboard es el resultado final de extracción, transformación y carga de la información mostrándola de manera amigable al usuario final, fácil de analizar, información organizada y seguimiento a KPI'S.

4. IDENTIFICACION INICIAL DE FUENTES

Para la construcción del dashboard en Power BI, es necesario identificar y analizar las fuentes de datos que contendrán la información relevante para el seguimiento de la fuerza de ventas. A continuación, se presentan diferentes alternativas de fuentes, especificando su tipo, formato y volumen de datos.

Fuente de datos	Tipo de fuente	Formato	Volumen estimado
ERP de la empresa : SAP	Base de datos relacional	Base de datos relacional SAP ASE	+20.000.000 registros históricos
Archivos Excel: Enviados por la fuerza de ventas	Archivos locales	XLSX, CSV	100 - 500 registros por archivo
Aplicaciones de terceros : plataformas de pedidos	API REST o Servicios Web	XML, XLSX	+100,000 registros mensuales

5. DEFINICIÓN DEL PROBLEMA

Nutriavicola SAS requiere optimizar el análisis de sus ventas mensuales de huevo para ajustar su estrategia comercial en función de variables como temporalidad, oferta y demanda en distintas regiones, fluctuación de precios y desempeño de los canales de distribución.

Actualmente, la empresa enfrenta dificultades en la recopilación, transformación y análisis de datos de ventas debido a que en su ERP SAP R/3 no cuenta con reportes diseñados a la necesidad de la empresa y para esto requiere un alto costo de consultoría de SAP para su desarrollo.

A continuación, doy un ejemplo del reporte actual que pueden descargar como una base de datos para construir en Excel u otras herramientas:

Reporte de Ventas y Facturación														
<div><div><div><div><div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><</div></div></div></div></div>														

Ilustración 1: Imagen transacción SAP ZSD004

Estos tipos de reportes impiden una evaluación eficiente de indicadores clave de desempeño (KPI's) y la ausencia de una herramienta visual dinámica también dificulta la identificación de tendencias comerciales, la medición del rendimiento de los vendedores y la optimización de estrategias de distribución.

Para resolver esta problemática, se propone el desarrollo de un dashboard en Power BI, que consolidará la información de ventas, permitiendo a la gerencia acceder a reportes dinámicos y tomar decisiones basadas en datos en tiempo real.

6. EXPLORACION INICIAL DE DATOS

Se realizó una exploración inicial de los datos disponibles relacionados con la fuerza de ventas y su desempeño a medir, encontrando diferentes archivos de Excel no estandarizados donde se llevan los presupuestos de los vendedores y se revisa en SAP las diferentes tablas donde posiblemente se pueda extraer la información para realizar el informe.

Fuentes de Datos Identificadas:

1. Archivos Excel de la Fuerza de Ventas: Archivos donde se encuentran sus presupuestos de ventas, por cliente, producto, canal de distribución y centro logístico.
2. ERP SAP R/3 Se identifican las siguientes tablas:

Tablas con información de clientes.

KNA1: Datos generales del maestro de clientes (nombre, Nit o cedula del cliente, dirección, teléfono).

KNVV: Información específica de ventas del cliente (Tipo de cliente, a que canal de distribución pertenece, zona de ventas, vendedor condiciones de pago).

Tabla con información de los productos.

MAKT: Descripciones de los productos.

Tablas con información de la facturación.

VBRK: Datos de cabecera de las facturas (fechas, cliente, destinatario, canal de distribución, totales).

VBRP: Datos de posición de las facturas (cantidad, producto, descuentos precio).

LIKP: Datos de cabecera de los documentos de entrega (fechas y números de entrega).

7. EXTRACCIÓN, TRANSFORMACIÓN Y VALIDACIÓN DE DATOS

Extracción de Datos

- Para la extracción directa desde SAP utilizaremos la herramienta con la que cuenta la empresa SAP Crystal Reports, permite consultar varias tablas al mismo tiempo de SAP y crear modelos relacionales entre ellas.
- La información generada puede descargarse en formato en XLS, XLSX o CSV.
- La extracción de datos mediante Crystal Reports se realizará mediante una carga incremental diaria, permitiendo la actualización continua sin duplicar registros innecesarios.
- Archivos Excel: Consolidación de los presupuestos para ser comparables con la venta real realizada en el mes.

Transformación de Datos

La transformación de datos se realizará en dos etapas. Primero, en SAP Crystal Reports, se organizarán y relacionarán las tablas extraídas del ERP para estructurar la información de manera coherente. Luego, en Python, se aplicarán procesos de limpieza, estandarización y consolidación para garantizar la calidad de los datos antes de pasarlos a Power BI.

Estos son los procesos que se realizarán:

- **Organización en Crystal Reports:** Se estructurará la relación de las tablas de SAP y generando el siguiente modelo relacional.

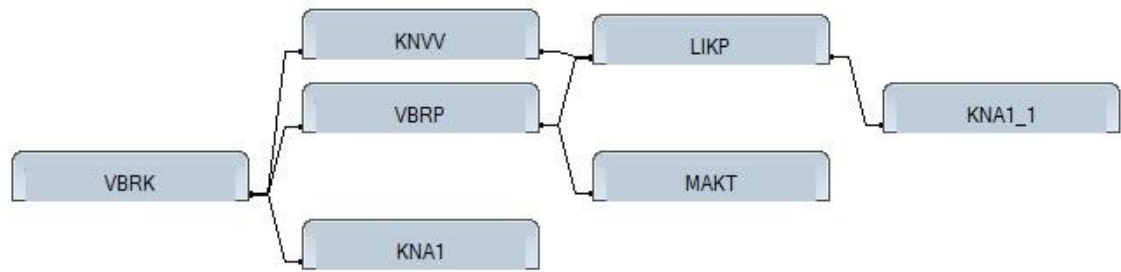


Ilustración 2: Relación de tablas en SAP Crystal Reports para la extracción de datos.

- **Limpieza de datos en Python:** Se eliminarán valores nulos, registros duplicados y datos inconsistentes para mejorar la calidad del análisis.
- **Estandarización de formatos:** Se unificarán las estructuras de datos provenientes de Excel y Crystal Reports, asegurando consistencia en nombres de columnas, formatos de fechas, montos y códigos de identificación.
- **Automatización del proceso:** Se desarrollarán scripts en Python que permitirán programar la transformación de datos, asegurando que la información se procese de manera eficiente y replicable en cada actualización.
- **Protección de la información confidencial:** Para proteger la información confidencial, se desarrollará un script en Python que reemplazará los nombres de clientes con identificadores genéricos (Cliente 1, Cliente 2...), además de renombrar zonas de ventas y canales de distribución para garantizar la privacidad de la información.

Validación de Datos

Para asegurarnos de que los datos sean confiables antes de integrarlos en Power BI, realizaremos varias validaciones clave:

- **Revisión de consistencia:** Compararemos los datos transformados con el reporte de ventas de ese mes para verificar que la información extraída desde SAP y los archivos Excel coincida con indicadores como ventas totales, cantidad de unidades vendidas y cantidad de clientes facturados.
- **Verificación de claves únicas:** Nos aseguraremos de que los identificadores de clientes, productos y vendedores sean correctos y estén correctamente relacionados en todas las tablas.

- **Detección de valores inusuales:** Usaremos Python para identificar posibles errores en los datos, como precios de venta fuera del rango esperado o nombres de productos que no correspondan a la misma categoría.
- **Protección de datos confidenciales:** Usaremos Python para revisar que los nombres de los clientes, zonas de ventas y canales de distribución hayan sido reemplazados correctamente por identificadores genéricos (Cliente 1, Cliente 2...).
- **Automatización de la validación:** Implementaremos scripts en Python para que estos controles se realicen automáticamente en cada actualización de datos.
- **Notificación de errores detectados:** Se generará un reporte automático que documentará cualquier anomalía encontrada en los datos, permitiendo su corrección antes de la carga en Power BI.

Con estas validaciones, nos aseguramos de que la información en Power BI sea precisa y confiable, facilitando el análisis y la toma de decisiones estratégicas.