

## **Challenge 02**

### **Documento de hallazgos y soporte analítico**

Curso: Fundamentos en Ciencia de Datos  
Universidad EAFIT

Juan Andrés Montoya  
Julián David Mejía

Periodo 2026-1  
Fecha: 5 de febrero de 2026

TechLogistics S.A.S. presenta señales consistentes de deterioro del margen y de la experiencia del cliente. La hipótesis directiva es que el problema no es un evento aislado, sino una debilidad estructural en la calidad, integración y trazabilidad de los datos operativos. Este trabajo aborda el reto con un enfoque de consultoría: audita la calidad del dato, ejecuta una limpieza transparente y construye un dashboard en Streamlit que convierte datos inconsistentes en indicadores accionables para la toma de decisiones.

El resultado principal es un Sistema de Soporte a la Decisión que permite medir calidad antes/después de la limpieza, cuantificar riesgo operativo (por SKU fantasma, margen negativo, brechas de entrega y tickets), y sintetizar hallazgos con un módulo opcional de IA vía Groq.

### **Datos y alcance**

El análisis se fundamenta en tres archivos CSV cargados por el usuario en la app:

- Inventario Central: maestro de productos con costos atípicos, existencias negativas y variabilidad en campos de fecha.
- Transacciones Logísticas: histórico de ventas con formatos de fecha mixtos, tiempos de entrega atípicos y posibles problemas de integridad referencial.
- Feedback de Clientes: registros duplicados, edades inválidas y escalas de satisfacción mezcladas.

Los errores se interpretan como señales de riesgo (gobernanza, captura y control interno), no solo como fallas técnicas

## Metodología implementada en Streamlit

- **Fase 1. Ingesta, limpieza y healthcheck**

La app realiza una carga controlada de archivos, aplica reglas de limpieza explícitas y ejecuta un healthcheck que resume faltantes, duplicados, tipos de dato y validaciones especializadas. El Health Score funciona como semáforo de preparación del dato para análisis posterior.

- **Cálculo de Health Score Inicial:**

El Health Score se calcula sobre 100 puntos y penaliza datos faltantes, duplicados, y hallazgos críticos según cada dataset. En la implementación actual, la penalización por faltantes corresponde a la suma de porcentajes de valores nulos por columna dividida entre 10, y cada duplicado reduce 0.5 puntos. Adicionalmente, se aplican penalizaciones por reglas críticas: inventario (stock negativo/NaN), feedback (duplicados, edades inválidas, NPS faltante) y transacciones (SKU fantasma, fechas inválidas, outliers extremos de entrega).

- **Limpieza de feedback:**

Se eliminan duplicados completos y duplicados por medio de una llave compuesta para evitar sesgo en agregaciones. Se depuran edades imposibles. En el puntaje de satisfacción, se aplica un valor absoluto para convertir los negativos en un valor válido, y luego se normalizan todos los valores en una escala de 0 a 10. Registros sin NPS o fuera de rango se excluyen.

- **Limpieza de inventario central:**

Se homogenizan los campos de fecha y se convierten al formato adecuado. Detección de outliers con un criterio estadístico (IQR), eliminación de negativos y NaN para mantener coherencia operativa. Depuración de costos menores o iguales a cero. Stock\_Actual se gestiona con lógica de negocio: stock negativo sin Lead\_Time\_Dias se elimina; si el Lead\_Time\_Dias existe, el stock negativo se imputa con la mediana del stock no negativo; remanentes negativos se eliminan.

- **Limpieza de transacciones logísticas:**

Se normalizan fechas de venta con parsing flexible. Se valida SKU\_ID: valores nulos/vacíos se eliminan. Si el inventario fue cargado, se marca SKU\_Fantasma cuando el SKU no existe en el maestro; estos casos se preservan por constituir un porcentaje considerable de los datos. Se eliminan outliers de entrega superiores a 999 días y combinaciones sospechosas como un SKU Fantasma con un tiempo de entrega mayor a 120 días. También se eliminan anomalías como cantidades negativas sin contexto y transacciones con fecha futura.

- **Fase 2. Integración, variables derivadas y riesgo operativo**

Luego de la limpieza, la app integra transacciones e inventario por SKU\_ID (left join). Las transacciones sin correspondencia quedan como FANTASMA. Con el dataset integrado se calculan indicadores de ingresos, costos, margen y brechas de entrega.

- $\text{Ingreso} = \text{Cantidad\_Vendida} \times \text{Precio\_Venta\_Final}$

- $\text{Costo\_Total} = (\text{Cantidad\_Vendida} \times \text{Costo\_Unitario\_USD}) + \text{Costo\_Envio}$

- $\text{Margen\_Utilidad} = \text{Ingreso} - \text{Costo\_Total}$

- $\text{Margen\_Pct} = \text{Margen\_Utilidad} / \text{Ingreso}$  (si Ingreso > 0)

- $\text{Brecha\_Entrega\_Dias} = \text{Tiempo\_Entrega\_Real} - \text{Lead\_Time\_Dias}$

Con estas variables se construye un indicador binario de riesgo operativo por transacción cuando ocurre al menos una condición: SKU fantasma, margen negativo, brecha de entrega mayor a 2 días, o ticket de soporte abierto.

- **Cálculo de Health Score final:**

En Fase 2, el dashboard calcula un Health\_Score transaccional (0–100) que representa riesgo operativo: inicia en 100 y descuenta 40 puntos si el SKU es fantasma, 30 si el margen es negativo, 20 si la brecha de entrega supera 2 días, y 10 si existe ticket de soporte abierto. El valor final se recorta al rango [0, 100].

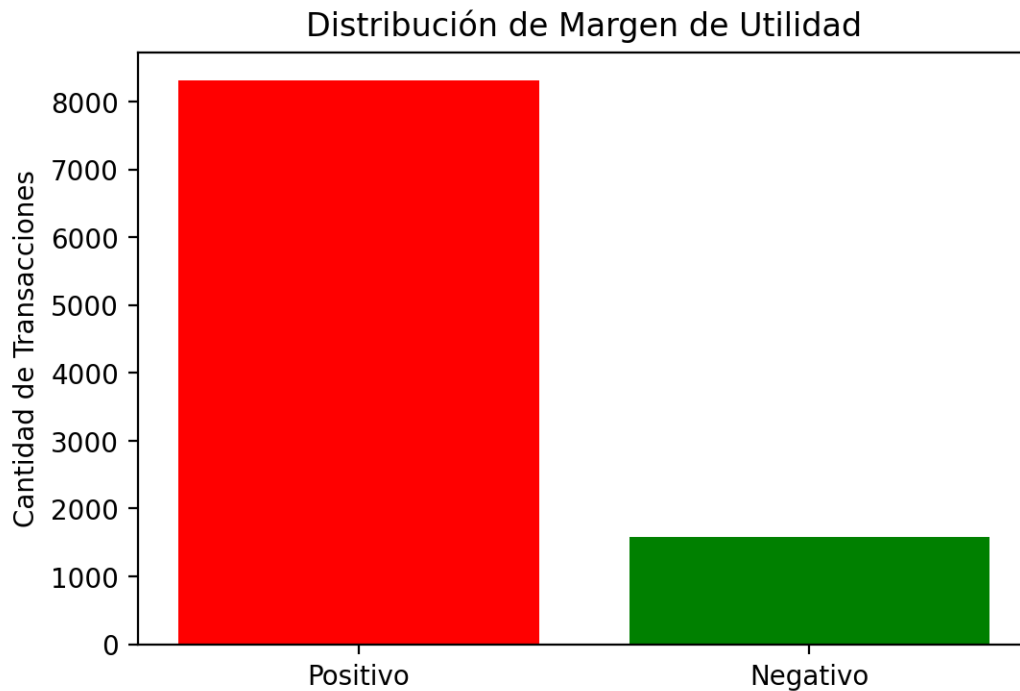
- **Fase 3. Integración de Groq:**

La app incluye un módulo opcional de análisis asistido por IA mediante Groq. El usuario ingresa una API key y selecciona el modelo. El sistema construye un contexto con estadísticas agregadas por bodega y genera recomendaciones operativas. La IA se usa como apoyo de síntesis y comunicación, no como sustituto del análisis.

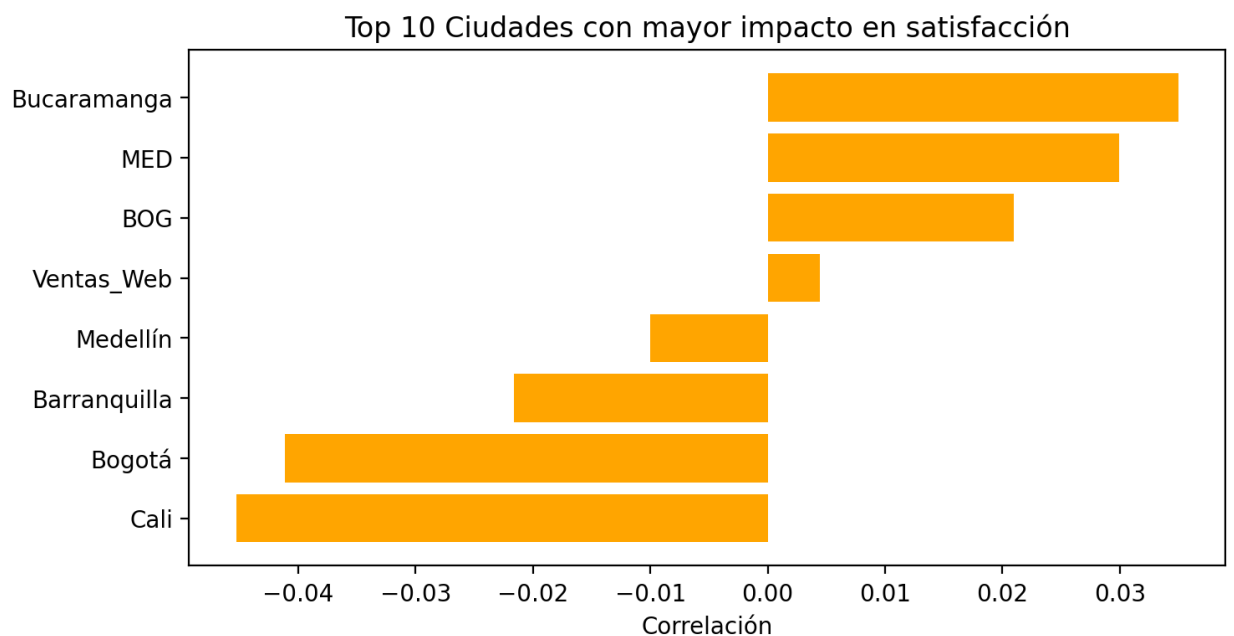
<b>Análisis de Riesgos Operativos por Bodega</b>
<b>Resumen</b>
El análisis de los datos proporcionados muestra que existen variaciones significativas en la antigüedad de revisión, tasa de tickets y satisfacción promedio entre las diferentes bodegas. A continuación, se presentan los resultados del análisis y se identifican los outliers, posibles causas y acciones recomendadas.
<b>Outliers</b>
<div><div>1.</div><div>BOD-EXT-99: Tiene la mayor antigüedad de revisión promedio (367,52 días) y la mayor tasa de tickets (0,1118). Esto sugiere que esta bodega puede tener problemas de eficiencia y calidad en su proceso de revisión.</div></div> <div><div>2.</div><div>Occidente: Tiene la segunda mayor antigüedad de revisión promedio (363,39 días) y una satisfacción promedio alta (2,2267). Esto puede indicar que la bodega está priorizando la calidad sobre la eficiencia.</div></div> <div><div>3.</div><div>Norte: Tiene una antigüedad de revisión promedio baja (342,29 días) y una tasa de tickets moderada (0,1066). Esto sugiere que la bodega puede estar funcionando de manera eficiente, pero puede requerir atención adicional para mejorar la calidad.</div></div>
<b>Posibles Causas</b>
<div><div>1.</div><div>Falta de personal capacitado: La antigüedad de revisión promedio alta en algunas bodegas puede ser causada por la falta de personal capacitado para realizar las revisiones de manera eficiente.</div></div> <div><div>2.</div><div>Procesos ineficientes: La tasa de tickets alta en algunas bodegas puede ser causada por procesos ineficientes o falta de automatización en la revisión y resolución de problemas.</div></div> <div><div>3.</div><div>Priorización de la calidad sobre la eficiencia: La satisfacción promedio alta en algunas bodegas puede ser causada por una priorización de la calidad sobre la eficiencia, lo que puede llevar a tiempos de revisión más largos.</div></div>
<b>Acciones Recomendadas</b>
<div><div>1.</div><div>Capacitación y desarrollo de personal: Proporcionar capacitación y desarrollo de personal para mejorar la eficiencia y la calidad en la revisión y resolución de problemas.</div></div> <div><div>2.</div><div>Análisis y optimización de procesos: Realizar un análisis detallado de los procesos de revisión y resolución de problemas para identificar áreas de mejora y optimizar la eficiencia.</div></div> <div><div>3.</div><div>Implementación de tecnologías de automatización: Implementar tecnologías de automatización para mejorar la eficiencia y reducir la tasa de tickets.</div></div> <div><div>4.</div><div>Revisión y ajuste de prioridades: Revisar y ajustar las prioridades de la bodega para equilibrar la calidad y la eficiencia.</div></div> <div><div>5.</div><div>Monitoreo y seguimiento: Establecer un sistema de monitoreo y seguimiento para evaluar el progreso y hacer ajustes según sea necesario.</div></div>

- **Hallazgos ejecutivos para alta gerencia:**

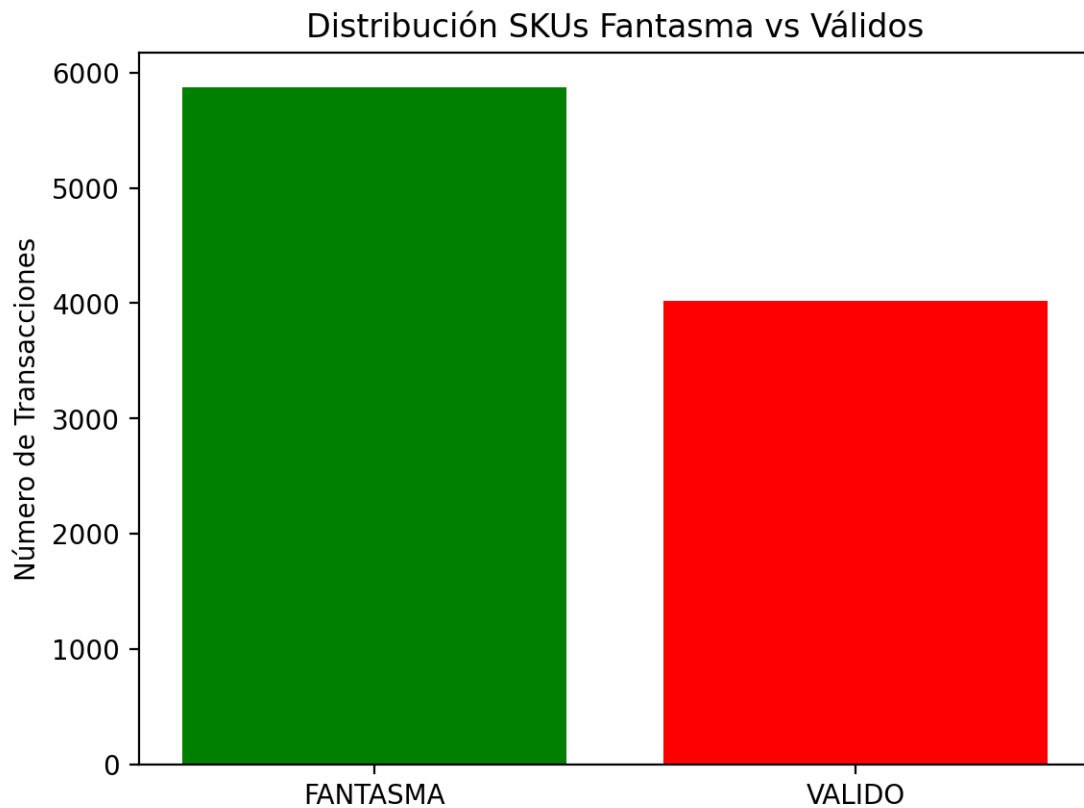
- Fuga de capital: se observan transacciones con margen negativo, lo que sugiere precios finales por debajo del costo total o costos mal registrados.



- Riesgo logístico: retrasos y brechas de entrega se asocian con mayor apertura de tickets y deterioro del NPS promedio, indicando cierto conflicto operacional.



- Venta invisible (SKU fantasma): existe un grupo de transacciones cuyo SKU no está en el inventario; esto obstaculiza auditoría precisa, obtención de datos para un assessment adecuado de la operación, y posibles malos cálculos en gastos para insumos invisibles, o faltantes.



- Fidelidad y percepción: combinaciones de baja satisfacción con condiciones operativas como tiempo de entrega, demuestran una oportunidad de mejora en el área de relacionamiento con el cliente.

- **Recomendaciones:**

- Implementar validaciones automáticas en la captura: rangos para edades, costos, stock y fechas; y obligatoriedad de SKU en transacciones.
- Utilizar el Health Score como indicador de calidad. Alertas cuando un dataset caiga por debajo de un umbral acordado.
- Gestionar SKUs fantasma como incidente de control interno. Investigar orígenes y consecuencias en el resto de la operación.
- Usar el dashboard como tablero de monitoreo continuo, incorporando responsables y cadencia de revisión.