

INFORME DE CONSULTORÍA:

**Diagnóstico Ejecutivo de Rentabilidad, Riesgo Operativo y Control de Inventario de
la empresa TechLogistics S.A.S**

Equipo consultor:

MIGUEL ROLDAN YEPES

JUAN CAMILO CATAÑO ZULETA

JUAN MANUEL AGUDELO OLARTE

CARLOS JOSE MUÑOZ CABRERA

**Estudiantes consultores de la materia Fundamentos en Ciencia de Datos de la
universidad EAFIT**

Medellín, febrero de 2026

Contenido

1. Resumen.....	4
2. Objetivos de la consultoría	4
2.1. Objetivo General.....	4
2.2. Objetivos Específicos	4
3. Auditoría de Calidad y Transparencia	5
3.1. Dataset Inventario	5
3.2. Dataset Transacciones.....	7
3.3. Dataset Feedback.....	9
4. Respuestas a preguntas de negocio	10
5. Recomendaciones	13

Índice de tablas

Tabla 1. Datos generales	5
Tabla 2. Porcentaje de nulos	6
Tabla 3. Porcentaje de nulos y tipo de dato (dataset original)	7
Tabla 4. Porcentaje de nulos y tipo de dato (dataset final)	8
Tabla 5. Porcentaje de valores nulos y tipo de dato para el dataset Feedback_clientes (dataset original)	9
Tabla 6. Porcentaje de valores nulos y tipo de dato para el dataset Feedback_clientes (dataset limpio)	10

1. Resumen

El siguiente informe es el resultado de la consultoría “Diagnóstico Ejecutivo de Rentabilidad, Riesgo Operativo y Control de Inventario” desarrollado por el equipo de estudiantes de la materia Fundamentos en Ciencia de Datos, y realizada para la empresa TechLogistics S.A.S, con la finalidad de determinar las circunstancias que llevaron a una pérdida de rentabilidad y mala calidad en la experiencia del cliente, haciendo uso de sus sistemas de inventario, logística y retroalimentación.

En primera instancia, se realizó una auditoría de la calidad y transparencia de los datos, en el cual se calculó el porcentaje de los valores nulos, duplicados, outliers y la consistencia en la estructura de los datos, por lo que fue posible identificar problemas de integridad que presentaban dichos sistemas como lo eran los costos extremos, tiempos de entrega irreales, edades invalidas y registros duplicados de feedback por una misma transacción. Luego, se llevó a cabo la etapa de integración y feature engineering, donde se creó una única fuente de datos, consolidando el feedback por Transaccion_ID y uniéndolo con el inventario a través de un LEFT JOIN para identificar explícitamente las ventas sin correspondencia en el catálogo (SKU fantasma).

2. Objetivos de la consultoría

2.1. Objetivo General

Analizar e integrar la información de sus tres sistemas principales de TechLogistics S.A.S. con la finalidad de construir una plataforma que permita identificar las causas en la pérdida de rentabilidad y la mala calidad en la experiencia del cliente

2.2. Objetivos Específicos

- Evaluar la calidad e integridad de los datos mediante una auditoría que identifique nulos, duplicados y valores atípicos, garantizando transparencia y confiabilidad en el análisis.
- Desarrollar un análisis visual que permita evaluar la rentabilidad, el desempeño logístico y el riesgo operativo del negocio, apoyando la toma de decisiones estratégicas.

3. Auditoría de Calidad y Transparencia

Esta fase tiene como finalidad entender y analizar el estado de los datos antes de realizar alguna transformación, asegurando que las decisiones posteriores sean basadas en información cuya calidad sea la más óptima posible. Además, es importante tener en cuenta que, en un proceso de consultoría, los datos no se "limpian" de forma automática, por lo que cada cambio debe ser trazable y justificado desde el aspecto técnico y de negocio.

Para lograr esto, se evaluó cada uno de sus tres sistemas principales (ERP de Inventarios, Logística y Feedback) a través de un conjunto de métricas de calidad, que incluyen el porcentaje de valores nulos por columna, la existencia de registros duplicados, y la presencia de valores atípicos en las variables numéricas, permitiendo identificar las situaciones donde el sistema presenta vacíos de información, posibles errores de captura o comportamientos extremos que podrían distorsionar los análisis financieros y operativos.

Por último, se definieron y registraron aquellas decisiones éticas sobre el tratamiento de la información. En particular, los registros que, aunque estuvieran incompletos se conservarían dado que podrían representar información de negocio relevante para el análisis y que variables podrían ser corregidas o transformadas, como lo es el caso de la imputación, en el cual para las variables numéricas se consideró la mediana u otra métrica en función de cómo este distribuida la información, evitando generar conflicto entre los datos que puedan modificar métricas fundamentales como márgenes, costos o niveles de servicio.

Dataset (Crudo)	Filas	Columnas
Inventario	2500	8
Transacciones	10000	10
Feedback	4500	9

Tabla 1. Datos generales

3.1. Dataset Inventario

En la fase de auditoría y calidad del dataset de inventario, se llevó a cabo una inspección exploratoria inicial para entender como viene los datos crudos provenientes de ERP, en donde inicialmente se examinan el porcentaje de nulos por columna, así como el tipo de dato, tal y como se observa en la tabla de **Porcentaje de nulos**. Esta primera lectura reveló que el inventario no estaba preparado para ser analizado en su estado original, ya que incluía formatos mixtos (como lead times en texto y números), categorías inconsistentes debido a variaciones en la escritura y valores extremos que, si se utilizaban sin tratamiento, podían distorsionar cualquier métrica de margen o planificación logística.

Nombre columna	Porcentaje de nulos [%]	tipo de dato
SKU_ID	0	object
Categoria	12.2	object
Stock_Actual	4.0	float64

Costo_Unitario_USD	0	float64
Punto_Reorden	0	int64
Lead_Time_Dias	16.12	float64
Bodega_Origen	0	object
Ultima_Revision	0	object

Tabla 2. Porcentaje de nulos

Con base en ese diagnóstico, se realizaron funciones de normalización con el propósito de convertir campos "difíciles" en variables que fueran consistentes y comparables, como el caso de la variable Lead_Time_Dias, en la cual, si el valor era "inmediato", se interpretó como un día para reflejar una disponibilidad casi instantánea; si el rango era, por ejemplo, "25-30 días", se empleó la cifra más alta (30) como decisión para tener un margen más alto al momento de realizar una entrega, dado que si es enviado en menor tiempo, mejora la percepción de los usuarios. De igual forma, se normalizó la columna categoría, dado que había productos divididos en varias etiquetas, donde se normalizaron los separadores, se quitaron caracteres raros y se agruparon variantes semánticas (como "smartphone", "smart-phone", etc.)

Luego, se tomó la decisión de que, para la imputación, primero se identifican los registros que presentan más de dos valores nulos en sus columnas; estos registros se eliminan, ya que una imputación tan extensa introduciría un nivel de incertidumbre que podría distorsionar los análisis posteriores. De este modo, solo se imputan aquellos registros con faltantes limitados, donde la reconstrucción de la información es estadística y operativamente confiable. Posterior a ello, con un inventario más consistente se aplicaron imputaciones; en primer lugar, se imputó Lead_Time_Días con la mediana, debido a que es una variable susceptible a valores extremos y se previene que algunos casos aislados lleven a que la mayoría tenga un lead time artificialmente alto o bajo.

Para Stock_Actual se evitó una imputación global y se imputó por mediana dentro de cada categoría, porque el stock tiene escalas muy distintas dependiendo del tipo de producto; imputar un valor único para todo el inventario distorsionaría el comportamiento real de cada categoría. Luego, en aquellos casos donde faltaba la categoría, se imputó dependiendo del perfil en el cual se estaba operando, es decir, se construyó un perfil por Bodega_Origen y Categoría usando la mediana de stock, y a cada registro sin categoría se le asignó la categoría cuyo stock mediano estuviera más cerca del Stock_Actual observado.

Finalmente, se encontraron algunas anomalías que por su comportamiento, no se trataron como simple valores atípicos, sino como posibles errores de captura o codificación explícitamente en la variable Stock_Actual, que presenta casos negativos y se tomó la decisión de no corregirlos ni eliminarlos, porque en muchos ERPs el stock negativo puede representar fenómenos contables reales (reservas, pedidos pendientes, desfases de sincronización, ajustes retroactivos) y, al estar distribuidos en múltiples categorías y bodegas, parecía un patrón sistémico más que un error puntual de digitación; cambiarlo sin evidencia podría ocultar fallas del proceso real.

3.2. Dataset Transacciones

En la fase de auditoría y calidad del dataset de transacciones logísticas, se realizó una inspección exploratoria para entender el estado de los datos crudos provenientes de la operación. Inicialmente se revisó el porcentaje de nulos por columna y el tipo de dato, tal como se muestra en la Tabla 1. Esto evidenció la presencia de valores atípicos y errores de captura (por ejemplo, -5 en cantidad vendida y 999 en tiempo de entrega), fechas futuras, valores vacíos en distintos formatos ("nan", "null", "?"), y variaciones en la escritura de ciudades y canales; elementos que, sin tratamiento, podían distorsionar análisis de desempeño y costos.

El dataset final consolidado contiene 9,815 registros y 12 columnas. El rango de fechas válidas observado es 2024-09-23 a 2026-01-31.

Nombre columna	Porcentaje de nulos [%]	Tipo de dato
Estado_Envio	16.83	object
Costo_Envio	8.34	float64
Transaccion_ID	0.00	object
SKU_ID	0.00	object
Fecha_Venta	0.00	object
Cantidad_Vendida	0.00	int64
Precio_Venta_Final	0.00	float64
Tiempo_Entrega_Real	0.00	int64
Ciudad_Destino	0.00	object
Canal_Venta	0.00	object

Tabla 3. Porcentaje de nulos y tipo de dato (dataset original)

Nota: Para esta auditoría, el cálculo de nulos considera también vacíos disfrazados (p. ej., "?", "--", "null").

Nombre columna	Porcentaje de nulos [%]	Tipo de dato
Fecha_Venta	0.59	object
Transaccion_ID	0.00	object
SKU_ID	0.00	object
Cantidad_Vendida	0.00	float64
Precio_Venta_Final	0.00	float64
Costo_Envio	0.00	float64
Tiempo_Entrega_Real	0.00	float64
Estado_Envio	0.00	object
Ciudad_Destino	0.00	object
Canal_Venta	0.00	object
Ciudad_Destino_norm	0.00	object
Canal_Venta_norm	0.00	object

Tabla 4. Porcentaje de nulos y tipo de dato (dataset final)

Tras la limpieza e imputación, el dataset final contiene 9,815 registros y 12 columnas. El rango de fechas válidas observado es 2024-09-23 a 2026-01-31.

Con base en el diagnóstico, se aplicaron funciones de normalización para estandarizar texto y reducir inconsistencias. Para ciudades y canales se utilizó corrección de encoding, minúsculas, eliminación de acentos y caracteres raros, además de alias comunes (p. ej., "bog" → "bogota", "med" → "medellin") y agrupación por similitud (fuzzy matching) para consolidar variantes semánticas. En ciudades, se pasó de 8 valores únicos a 6 valores canónicos. En canales son la misma cantidad que en el original, pero estandarizados.

Para la imputación, primero se eliminaron los registros con más de dos valores nulos, ya que una reconstrucción tan extensa introduce incertidumbre alta. Posteriormente se imputaron las variables numéricas con un enfoque multivariado (Iterative Imputer) y se controlaron valores extremos por IQR. Finalmente, cuando existían faltantes en Estado_Envio, se imputaron mediante un modelo supervisado (Random Forest) usando variables base de venta, costo y tiempos.

3.3. Dataset Feedback

Tras la inspección exploratoria del Dataset que contiene el feedback de clientes, se identificaron valores atípicos, inconsistencias en categorías y valores faltantes, que podrían distorsionar indicadores de satisfacción. Con base en ese diagnóstico se aplicó la limpieza y estandarización en tres frentes: normalización de las categorías, tratamiento de valores nulos y valores atípicos

En el Dataset crudo se tienen 4500 registros y a partir de ello se comenzó la fase de auditoría y calidad de esta información. Se comenzó por revisar algunos valores importantes de aspectos generales del Dataset como, por ejemplo, el porcentaje de valores nulos en cada columna, el tipo de valores en cada columna y la vista de algunos registros que contiene el dataset. La tabla 5. muestra la relación de valores nulos encontrados en el dataset crudo, se observan que las columnas con valores nulos son Comentario_Texto y Recomendación_Marca, aunque se evaluaron todas las columnas para su verificación.

Nombre columna	Porcentaje de nulos [%]	Tipo de dato
Feedback_ID	0.000000	object
Transaccion_ID	0.000000	object
Rating_Producto	0.000000	int64
Rating_Logistica	0.000000	int64
Comentario_Texto	14.600000	object
Recomienda_Marca	24.866667	object
Ticket_Soporte_Abierto	0.000000	object
Edad_Cliente	0.000000	int64
Satisfaccion_NPS	0.000000	float64

Tabla 5. Porcentaje de valores nulos y tipo de dato para el dataset Feedback_clientes (dataset original)

Inicialmente, se normalizó las categorías. Para la columna de Ticket_Soporte_Abierto, se revisó y se mapearon todas las formas posibles (1, 0, “Sí”, “Si”, “No”, etc.) a una escala binaria consistente (Sí/No). Para la columna de Recomendación_Marca se estandarizaron las respuestas sobre las categorías que se tenían (“Sí”, “No”, “Maybe” y los valores nulos) a categorías donde los valores nulos ahora son “Sin Respuesta”, y “Maybe” a “Tal vez”. Por otro lado, la columna de Comentario_Texto tenía algunos registros donde se encontraba el valor “---” y valores nulos, dichos valores se cambiaron a “Sin comentarios” para poder mantener una mejor clasificación de esos registros. Como se observa, ese proceso de estandarización se realizó a variables categóricas, pero también se debe normalizar las variables numéricas como Satisfacción_NPS la cual toma valores en un rango de -100 a 100. Para el proceso de normalización se realizó una división de cada valor entre 100 y se redondea a 2 decimales, esto hace que se mantenga un rango entre -1 y 1 para el análisis posterior.

En cuanto a la columna de Rating_Producto y Edad_Cliente, se maneja un problema similar de valores fuera del rango establecido, el primero de 1-5 y el segundo presenta edades que no son posibles físicamente. Para la columna relacionada al rating del producto se encontraron 22 filas con un valor de 99 y para la columna que relaciona edad se encontraron 23 filas con un registro de edad

de 195 (Aunque el rango superior se dejó en 90). La decisión de eliminar estos registros se tomó pensando en no modificar o distorsionar los registros de satisfacción.

Por último, se encontraron feedback_id repetidos, pero se decide mantener esos registros porque aún aportan información que puede ser relevante en lo que sigue del proceso de auditoría.

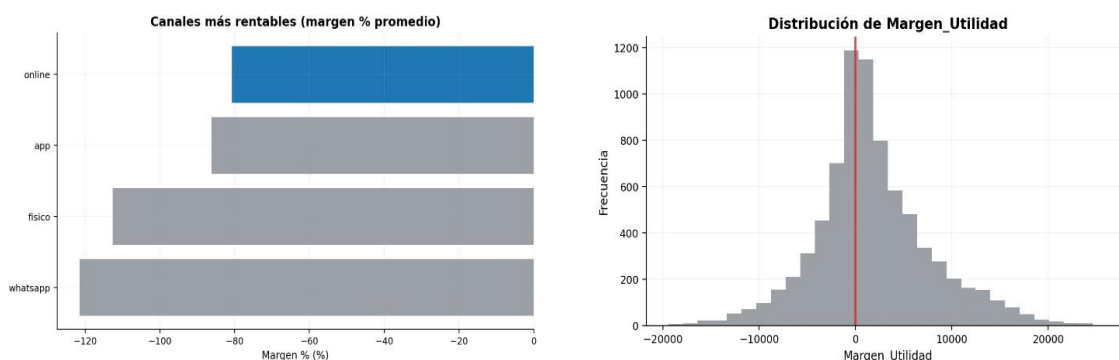
La tabla 6. muestra el resultado después de aplicar todo el proceso descrito anteriormente. El dataset final cuenta con 4455 registros con las columnas necesarias estandarizadas y con los valores nulos tratados según el caso puntual y tomando la decisión que mejor se acopla al problema.

Nombre columna	Porcentaje de nulos [%]	Tipo de dato
Feedback_ID	0.0	object
Transaccion_ID	0.0	object
Rating_Producto	0.0	int64
Rating_Logistica	0.0	int64
Comentario_Texto	0.0	object
Recomienda_Marca	0.0	object
Ticket_Soporte_Abierto	0.0	object
Edad_Cliente	0.0	int64
Satisfaccion_NPS	0.0	float64

Tabla 6. Porcentaje de valores nulos y tipo de dato para el dataset Feedback_clientes (dataset limpio)

4. Respuestas a preguntas de negocio

4.1. Fuga de Capital y Rentabilidad

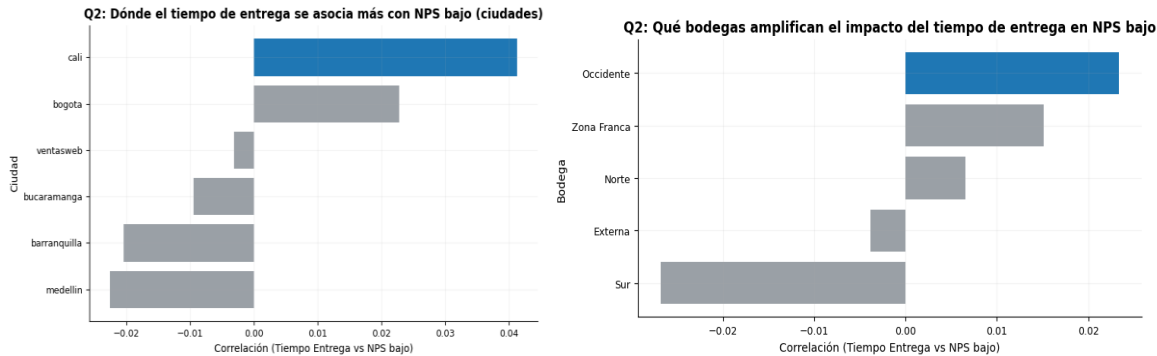


Al preguntarnos si es una falla del canal online, los datos muestran que no lo es, la fuga de capital es generalizada. De hecho, el canal Online tiene la menor incidencia de ventas negativas proporcionalmente, aunque la diferencia es mínima y no altera el diagnostico.

El verdadero problema no es el canal, sino el modelo económico de estos SKUs. En todas las categorías, no funcionan como productos que atraigan clientes, sino que están causando pérdidas de margen por una mala estructura de costos y precios que se repite en cada canal. En la práctica, cada

unidad que se vende aumenta la pérdida. Se recomienda suspender la venta de estos productos o corregir sus costos/precios (cambio de proveedores, mejora en el manejo del producto, optimización de materiales, etc.) inmediatamente en el sistema central, ya que cada venta contribuye en pérdidas significativas.

4.2. Crisis Logística y Cuellos de Botella



Basándonos en las correlaciones observadas, la zona que requiere un cambio inmediato de operadore es el corredor Cali – Bodega Occidente, esto debido a que, entre las ciudades analizadas, está presente una mayor correlación positiva entre el tiempo entrega y la probabilidad de NPS bajo (0.0413), lo que sugiere un aumento de la insatisfacción aumente a medida que las entregas se retrasan. Asimismo, se observa que la Bodega Occidente tiene una correlación más alta, de 0.0233, lo que indica que dichos retrasos desde este origen tienen un impacto tangible en la experiencia del cliente.

A diferencia, ciudades como Medellín o Barranquilla y bodegas como Sur o Externa tienen correlaciones negativas o muy bajas. Esto indica que su insatisfacción no está influenciada por la logística. Por lo tanto, la evidencia indica un problema en el cumplimiento en la Bodega Occidente que afecta especialmente a Cali.

4.3. Análisis de la Venta Invisible

El análisis de la venta Invisible revela una brecha crítica en el control financiero y de inventarios:

Ingreso total 73,340,569.09 <small>suma de Ingreso_Total</small>	Ingreso controlados 58,721,767.93 <small>SKUs presentes en inventario</small>	Venta invisible 19.9% <small>del ingreso está fuera del inventario</small>
--	---	--

Impacto Financiero Cuantificado

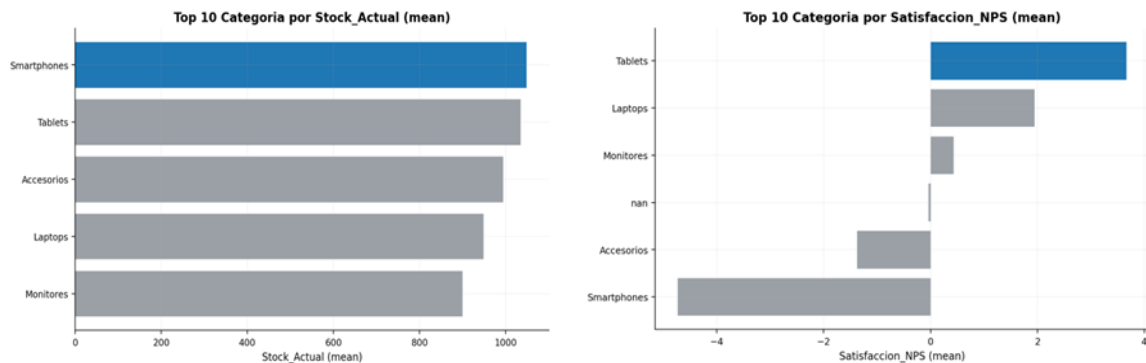
- Ventas en Riesgo (Venta Invisible): \$14,618,801.16 USD
- Porcentaje del Ingreso Total: 19.93%

Casi el 20% de los ingresos de la compañía proviene de productos que no existen en el maestro de inventarios. Esta falta de visibilidad de la operación es un gran problema para la empresa porque les dificulta identificar ganancias reales de posibles falsos positivos, incluso les dificulta analizar el margen de sus productos ya que no sabemos si estos son reales o no. En total son 1,978 transacciones

afectadas y 565 SKUs únicos que se están moviendo fuera del radar del sistema de gestión de inventarios.

Por lo tanto, no es un error marginal; es un gran error de operación que compromete la veracidad del Estado de Resultados de la empresa y afecta no solo sus ingresos sino la confianza de los inversionistas. Se recomienda revisar urgentemente estos 565 SKUs para darlos de alta en caso de ser errores o agregarlos al maestro de inventarios.

4.4. Diagnóstico de Fidelidad



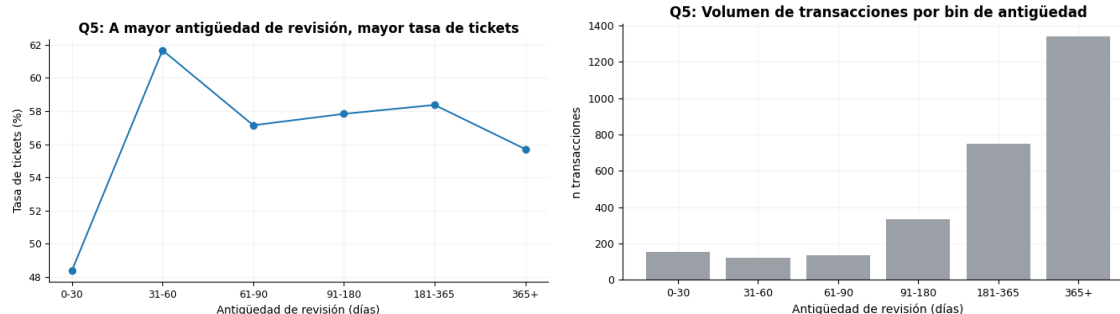
La paradoja se presenta en dos categorías: smartphones y accesorios.

La razón de la alta insatisfacción apunta a un problema de calidad del producto y no a un simple sobreprecio por las siguientes razones:

1. El Margen es Negativo: Como descubrimos en el análisis anterior, estas categorías se están vendiendo con pérdidas, la empresa no está cobrando un sobreprecio; de hecho, está subsidiando el precio de venta. A pesar de comprar barato hay insatisfacción
2. Acumulación de Stock: El alto nivel de stock (más de 1,000 unidades promedio) sugiere que el mercado está rechazando estos productos y la empresa los sigue conservando y peor, comprando.

Lo que sugiere que, la empresa está almacenando grandes cantidades de inventario de tecnología (Smartphones/Accesorios) que es intrínsecamente deficiente. Se está pagando un costo alto por productos de mala calidad que el cliente rechaza incluso cuando se venden con descuento (pérdida).

4.5. Storytelling de Riesgo Operativo



Podemos observar que, aunque la correlación lineal entre la antigüedad de la última revisión y los tickets es cercana a cero, con un valor de -0.0152 , los tramos operativos revelan un gran problema, dado que, cuando una bodega pasa de 0 – 30 días sin revisión a 31 – 60 días la tasa de tickets salta de 48.4% a 61.7%, el punto más alto de toda la serie y luego de este punto, se mantiene en un rango aproximado entre 57% a 58% , incluso cuando la antigüedad sigue aumento, indicando que, una vez se pierde la visibilidad del inventario, el daño en la experiencia del cliente no se revierte. Este riesgo es muy importante porque la mayoría de las transacciones se agrupan en los periodos de 181–365 días y más de 365 días (747 y 1,341 transacciones, respectivamente), lo que sugiere que, las operaciones a gran escala se están realizando con inventarios "a ciegas", es decir, las bodegas que no revisan su stock por más de 30 días están operando con alta probabilidad de errores de disponibilidad, picking y promesas incumplidas, lo que se traduce en más tickets, más fricción y menor satisfacción final.

5. Recomendaciones

En base a los resultados, se debe parar de forma inmediata la pérdida del valor causada por la fuga de capital, dado que, los SKUs con margen negativo no están cumpliendo una función estratégica de atracción de clientes ni de venta cruzada, sino que están erosionando directamente el resultado financiero en cada transacción. Por lo tanto, estos productos deben ser suspendidos o despublicados de todos los canales hasta que se recalcule su estructura real de costos y se establezca un precio mínimo rentable. Al mismo tiempo, se debe comenzar a renegociar con proveedores, buscar reemplazos para insumos y mejorar la logística, porque continuar vendiendo bajo las condiciones actuales significa asumir pérdidas activamente.

Desde el punto de vista operativo, el análisis logístico indica que el problema no es generalizado, sino concentrado en el corredor Cali – Bodega Occidente, lo que exige una intervención quirúrgica y no una reestructuración masiva. Esta bodega debe ser auditada de manera prioritaria en sus procesos de picking, despacho y última milla, ya que es allí donde los retrasos están impactando de forma directa la satisfacción del cliente. Mientras se corrige la causa raíz, parte del volumen de Cali debe ser redirigido hacia bodegas con mejor desempeño, como Sur o Externa, de modo que el deterioro del NPS no continúe amplificándose por una falla localizada.

En el frente financiero y de control, la existencia de casi un 20 % de los ingresos provenientes de la llamada “Venta Invisible” constituye un riesgo crítico para la compañía. No es viable operar con casi dos mil transacciones y más de quinientos SKUs fuera del maestro de inventarios, porque esto

invalida el análisis de márgenes, distorsiona el Estado de Resultados y mina la confianza de los inversionistas. Se requiere un saneamiento urgente que implique congelar temporalmente la venta de estos productos, depurar los registros y dar de alta en el sistema únicamente aquellos SKUs que existan y sean trazables, integrando de forma estricta inventarios, ventas y contabilidad.