

# **Análisis de Supervivencia Aplicado en la Gestión del Tiempo de Devolución de Envases**

"Aplicación de conceptos y técnicas de Administración en una situación laboral de revista o ambiente real"

2024

Matias Luciano Gomez

matiaslucianogomez@hotmail.com



<b>Índice</b>	
<b>Resumen</b> .....	4
<b>Introducción</b> .....	5
<b>Situación Problemática</b> .....	6
<b>Preguntas de Investigación</b> .....	6
<b>Objetivo General</b> .....	7
<b>Objetivos Específicos</b> .....	7
<b>Marco Metodológico</b> .....	7
<b>Marco Teórico</b> .....	8
Economía Circular .....	8
Análisis de Supervivencia .....	9
Teoría de la Penalización de Clientes .....	10
Distribuciones de Probabilidad Empíricas y Teóricas .....	11
Empresas Impulsadas por Datos .....	11
Data Mining .....	11
Machine Learning No Supervisado .....	12
<b>APLICACION</b> .....	12
Ajuste del Análisis de Supervivencia a Ciclos Repetitivos .....	12
Implicaciones en el Análisis: .....	12
Solución Propuesta: .....	12
Eventos de interés: .....	13
<b>PROCESAMIENTO DE DATOS</b> .....	13
Limpieza de Base de Datos .....	18
<b>ANÁLISIS INICIAL DE LA DISTRIBUCION DE PROBABILIDAD</b> .....	18
Interpretación del Ajuste a la Distribución Lognormal .....	20
<b>Implementación del Análisis de supervivencia</b> .....	22
Paso 1: Instalación de las Bibliotecas Necesarias .....	22
Paso 2: Importación de Bibliotecas y Carga de Datos .....	22
Paso 3: Creación del Modelo de Kaplan-Meier .....	23
Paso 4: Ajuste del Modelo a los Datos .....	23
Paso 5: Visualización de la Curva de Supervivencia .....	23
Paso 6: Análisis por Grupo .....	24



Paso 7: Interpretación de las curvas .....	24
<b>Recomendaciones</b> .....	25
<b>Conclusiones</b> .....	26
<b>Referencias</b> .....	28
<b>Apéndice</b> .....	29



## Resumen

Circulo S.A., fundada en 2020, opera bajo un modelo de economía circular en la industria de productos de limpieza, promoviendo la reutilización de envases plásticos. Este modelo no solo busca minimizar la generación de residuos, sino también fomentar la sostenibilidad ambiental. Sin embargo, garantizar la devolución oportuna de los envases ha representado un desafío recurrente, generando costos adicionales y riesgos ambientales para la empresa.

El objetivo de este estudio fue modelar el tiempo de devolución de los envases mediante análisis de supervivencia, identificando factores influyentes en la retención o demora, y proponiendo estrategias para optimizar la gestión de los envases. El marco teórico se fundamentó en conceptos de economía circular, análisis de supervivencia y penalización de clientes, complementados con herramientas estadísticas como las curvas de Kaplan-Meier y distribuciones de probabilidad.

Metodológicamente, se adoptó un enfoque cuantitativo con diseño longitudinal, utilizando datos históricos de movimientos de envases procesados en el sistema TRAC. Se llevaron a cabo análisis de distribuciones empíricas y teóricas, seleccionando la distribución Lognormal como la más adecuada para modelar ciclos completos de devolución. Además, las curvas de Kaplan-Meier permitieron incorporar datos censurados, reflejando con mayor precisión la probabilidad de devolución a lo largo del tiempo.

Entre los hallazgos clave, se destacó que los bidones de 5 litros presentaron patrones de devolución más consistentes en comparación con las botellas de 1 litro y los botes de bicarbonato, que mostraron mayores tasas de retención. La probabilidad de no devolución al día 125 varió significativamente entre los tipos de envase, destacando la necesidad de estrategias diferenciadas según el material y uso del envase.

Este estudio demostró que la combinación de análisis de supervivencia y modelos estadísticos avanzados puede generar información estratégica para tomar decisiones basadas en datos. Las recomendaciones incluyen ajustes en las políticas de penalización, incentivos económicos y campañas educativas, orientadas a mejorar la tasa de devolución y fortalecer la sostenibilidad del modelo de negocio.

**Palabras Clave:** análisis de supervivencia, economía circular, penalización de clientes, sustentabilidad, devolución de envases.



## Introducción

La contaminación ambiental es uno de los mayores desafíos que enfrenta la humanidad en el siglo XXI. Los modelos de producción y consumo lineal, caracterizados por la fabricación, uso y eliminación de productos, han contribuido al aumento significativo de residuos sólidos, especialmente plásticos. Estos materiales, al ser desechados de manera irresponsable, terminan contaminando océanos, ríos y suelos, generando un impacto negativo en los ecosistemas y en la salud humana. Según un informe del Programa de las Naciones Unidas para el Medio Ambiente (PNUMA), se estima que cada año se producen más de 300 millones de toneladas de desechos plásticos, de las cuales una gran proporción termina en el medio ambiente, afectando gravemente la biodiversidad y contribuyendo al cambio climático.

En este contexto, el modelo de economía circular ha surgido como una alternativa necesaria para mitigar estos efectos. La economía circular se basa en principios de diseño sostenible que promueven la reutilización, reciclaje y recuperación de materiales, buscando cerrar el ciclo de vida de los productos. Este enfoque no solo reduce la necesidad de extraer nuevas materias primas, sino que también disminuye la cantidad de residuos que se generan, reduciendo la huella de carbono asociada con la fabricación y disposición de productos. La implementación de la economía circular en diversas industrias podría reducir las emisiones globales de gases de efecto invernadero hasta en un 45% para el año 2050, según el Ellen MacArthur Foundation.

En este marco, Circulo S.A. inició sus operaciones en el año 2020, impulsada por la necesidad de adoptar un sistema de economía circular en la industria de productos de limpieza. Este enfoque busca promover la reutilización de envases, reduciendo así la generación de residuos plásticos descartables y la huella de carbono asociada a la producción de estos.

El modelo de negocio de Circulo S.A. se basa en la venta de productos de limpieza, no en los envases que los contienen. Actualmente la empresa posee una cartera de 14 tipos de productos. En la primera venta a un nuevo cliente, la empresa cobra un cargo adicional que corresponde al costo del envase, ya que se entiende que es necesario asignarle uno junto con el producto. A partir de la segunda compra, los clientes sólo pagan por la recarga del producto, sin incurrir en el costo completo del envase.

Es importante destacar que el costo de los bidones reforzados utilizados por la empresa no se transfiere en su totalidad a los clientes, ya que esto podría representar una barrera significativa para la captación de nuevos usuarios. El cargo adicional cubre aproximadamente el 50% del costo del envase, lo que implica que Circulo S.A. asume una parte del costo en cada primera venta. Por lo tanto, desde una perspectiva financiera, es menos beneficioso para la empresa atraer constantemente a nuevos clientes en comparación con fomentar la fidelización de los existentes, quienes ya poseen los envases y solo requieren recargas.



Desde el punto de vista ambiental, la reutilización de los envases es fundamental para la reducción de la huella de carbono. Cada envase reutilizado representa una disminución en la cantidad de plástico desechado y una reducción en las emisiones de gases de efecto invernadero asociadas con la producción de nuevos envases. Además, este modelo contribuye a la sostenibilidad al alentar a los consumidores a participar activamente en la economía circular, retornando los envases vacíos para ser recargados y reutilizados.

## Situación Problemática

En el marco del modelo de economía circular implementado por Circulo S.A., la reutilización de envases es un pilar fundamental tanto para la sostenibilidad ambiental como para la viabilidad económica de la empresa. Sin embargo, garantizar que los clientes devuelvan los envases vacíos en un plazo adecuado ha sido un desafío persistente.

En algunos sistemas de economía circular, especialmente en empresas europeas del sector alimenticio, se ha implementado una política de devolución de envases en un plazo de 15 días. Esta medida busca incentivar la rápida rotación de envases y evitar el acopio innecesario en manos de los consumidores. Sin embargo, este plazo puede no ser adecuado en el contexto de Circulo S.A., donde se comercializan productos de limpieza en grandes volúmenes, como bidones de 5 litros, cuyo tiempo de uso es significativamente mayor al de productos perecederos. Aplicar un plazo de 15 días para la devolución de estos envases podría resultar en penalizaciones percibidas como injustas por parte de los clientes, afectando su satisfacción y, potencialmente, su fidelidad hacia la empresa.

Por otro lado, la demora en la devolución de los envases genera costos adicionales para la empresa. Cada envase no retornado dentro de un período razonable representa una pérdida potencial, ya que la empresa debe incurrir en la adquisición de nuevos envases para continuar operando, afectando su margen de ganancia y sostenibilidad económica. Este escenario se agrava cuando los clientes deciden no devolver los envases en absoluto, forzando a la empresa a asumir el costo total de los mismos.

La situación actual demanda una solución que permita a Circulo S.A. determinar con precisión el momento óptimo para aplicar una penalización o cobro adicional por la no devolución de los envases. Esta solución debe ser lo suficientemente flexible como para adaptarse a los diferentes patrones de consumo de los productos, evitando al mismo tiempo costos innecesarios y preservando la relación con los clientes.

## Preguntas de Investigación

A partir de la situación problemática, se plantean las siguientes preguntas de investigación:

- ¿Cuál es la distribución de probabilidad más adecuada para modelar el tiempo de devolución de los envases por parte de los clientes?
- ¿Cuál es el tiempo promedio que tardan los clientes en devolver los envases de los productos comercializados por Circulo S.A.?



- ¿Cuáles son los factores que más influyen en la demora o en la no devolución de los envases?

## Objetivo General

El objetivo general del trabajo es:

Modelizar el tiempo de devolución de los distintos envases en Circulo S.A., utilizando análisis de supervivencia.

## Objetivos Específicos

Los objetivos específicos que se plantean para el trabajo son:

1. Determinar la distribución de probabilidad que mejor se ajusta al tiempo de devolución de los envases.
2. Identificar los factores que más influyen en la demora o en la no devolución de los envases.
3. Estimar el tiempo promedio que tardan los clientes en devolver los envases de los productos comercializados por Circulo S.A.

## Marco Metodológico

Este estudio se realizó bajo un enfoque cuantitativo y emplea un diseño longitudinal de panel, lo cual permite analizar los ciclos de devolución de envases en Circulo S.A. y observar la evolución temporal del comportamiento de los clientes en cada ciclo de uso. Este diseño resulta particularmente apropiado para estudiar cómo y cuándo los envases completan un ciclo de uso, desde su asignación inicial a un cliente hasta su eventual retorno al almacén. La base de datos, estructurada a partir de los registros históricos del sistema de movimientos de Circulo S.A., permite el seguimiento de cada envase mediante un identificador único. A través de este identificador y de otras variables, como las fechas de asignación y devolución y el estado del envase (en cliente o en almacén), es posible trazar un ciclo de uso completo o, en su defecto, registrar los datos como censurados en caso de no retorno del envase. Esta estructura de datos es compatible con un análisis longitudinal, donde cada ciclo de uso se convierte en una observación independiente dentro del estudio. La construcción de la base de datos incluyó procesos de limpieza y tratamiento de datos para resolver problemas como registros incompletos, valores atípicos, y casos donde el envase fue reasignado a un segundo cliente sin haber pasado por el almacén. En este último caso, se consideró que el envase retornó al almacén el mismo día en que se asignó nuevamente, cerrando el ciclo como un evento completo para mantener la coherencia en el análisis.

El análisis estadístico incluyó la aplicación de técnicas de análisis de supervivencia, que resultan particularmente útiles en estudios donde se analiza el tiempo hasta que ocurre un evento de interés. En este caso, el evento es la devolución del envase. Para modelar adecuadamente el tiempo de devolución, se utilizaron las curvas de supervivencia de



Kaplan-Meier, que permitieron estimar y visualizar la probabilidad de que un envase permanezca sin devolver en distintos momentos a lo largo del tiempo. Estas curvas ofrecen una representación clara y no paramétrica de la supervivencia de los envases en manos de los clientes, y permiten comparar grupos según diferentes características, como el tipo de producto o la modalidad de entrega. Además, el análisis de Kaplan-Meier facilita la identificación de los puntos de tiempo en los que se observan descensos significativos en la probabilidad de supervivencia, lo cual puede informar sobre el momento óptimo para implementar políticas de penalización.

En el proceso de análisis también se emplearon modelos de regresión de Cox, que permitieron evaluar el impacto de varias variables (tipo de producto, frecuencia de devolución, modalidad de entrega) sobre el tiempo hasta la devolución del envase. Este modelo de riesgos proporcionales resulta adecuado para identificar los factores más influyentes en la probabilidad de devolución, ajustando así el análisis a los distintos patrones observados en el comportamiento de los clientes.

Para la ejecución de estos análisis, se utilizó Python, que ofreció herramientas estadísticas avanzadas y facilitó la construcción de visualizaciones, como las curvas de supervivencia y los gráficos de regresión, necesarios para interpretar los resultados del estudio. El análisis en Python incluyó procedimientos de validación cruzada y análisis de residuos para asegurar la estabilidad y precisión del modelo de regresión de Cox.

## Marco Teórico

### Economía Circular

La economía circular según Geissdoerfer es un concepto económico que se propone como alternativa al modelo lineal tradicional de "tomar, hacer, desechar". Este enfoque busca mantener los recursos en uso durante el mayor tiempo posible, extraer el máximo valor de ellos mientras están en uso, y luego recuperar y regenerar productos y materiales al final de su vida útil. La economía circular se basa en varios principios clave, incluyendo la reducción de residuos, la reutilización de productos y materiales, y la regeneración de los sistemas naturales.

El concepto de economía circular ha ganado relevancia a nivel global debido a la creciente preocupación por la sostenibilidad y el impacto ambiental de las prácticas industriales y de consumo actuales. En particular, la industria del plástico ha sido un foco importante debido a la cantidad significativa de residuos plásticos que terminan en vertederos o en el medio ambiente. Según la Fundación Ellen MacArthur (2019), cada año se producen más de 300 millones de toneladas de plástico, de las cuales una gran parte no se recicla y contribuye a la contaminación ambiental.

El modelo de economía circular aplicado a la industria del plástico implica la implementación de sistemas de reciclaje y reutilización de envases, con el objetivo de minimizar la cantidad de desechos plásticos generados. Las empresas que operan bajo este modelo buscan diseñar productos y envases que puedan ser fácilmente reciclados o reutilizados, reduciendo así la necesidad de materias primas vírgenes y disminuyendo la huella de carbono asociada con la producción de plástico.

Una de las estrategias clave en la economía circular es el sistema de devolución y reutilización de envases. Este sistema no solo reduce la cantidad de residuos plásticos, sino





que también puede generar ahorros económicos para las empresas al reducir los costos asociados con la producción de nuevos envases. Sin embargo, para que estos sistemas sean eficaces, es esencial que los clientes devuelvan los envases en un plazo razonable. Aquí es donde surge la necesidad de establecer un tiempo óptimo antes de aplicar penalizaciones a aquellos clientes que no cumplen con este requisito.

### Análisis de Supervivencia

El análisis de supervivencia (Hosmer, Lemeshow, & May, 2008) es una rama de la estadística que se ocupa del análisis del tiempo hasta que ocurre un evento de interés. En muchos casos, este evento puede ser la muerte de un paciente, la falla de un dispositivo mecánico o, en el contexto de este estudio, el tiempo hasta la devolución de un envase por parte de un cliente. El análisis de supervivencia es particularmente útil cuando se estudian eventos que ocurren a lo largo del tiempo y cuando se necesita manejar datos censurados, es decir, datos en los que el evento de interés no ha ocurrido para todos los sujetos al final del período de observación.

Una de las herramientas más utilizadas en el análisis de supervivencia es la curva de Kaplan-Meier, que estima la función de supervivencia, es decir, la probabilidad de que un sujeto (en este caso, un envase) aún no haya experimentado el evento de interés (devolución) en un momento dado. La curva de Kaplan-Meier es útil para comparar diferentes grupos de sujetos, como los que reciben diferentes modos de entrega de pedidos, y observar cómo estas diferencias afectan el tiempo hasta la devolución de los envases.

El análisis de supervivencia se fundamenta en la estimación y evaluación del tiempo hasta que ocurre un evento de interés. En este contexto, la función de supervivencia desempeña un papel central, definiéndose como la probabilidad de que un individuo, o en este caso un envase, no experimente el evento de interés (devolución) hasta un momento específico  $t$ . Matemáticamente, la función de supervivencia se expresa como:

$$S(t) = P(T > t)$$

donde  $T$  es la variable aleatoria que representa el tiempo de supervivencia. Esta función tiene ciertas propiedades fundamentales: inicia con un valor  $S(0) = 1$ , ya que todos los sujetos están "vivos" al comienzo del análisis, y decrece a medida que transcurre el tiempo, indicando la ocurrencia acumulativa de eventos.

Uno de los objetivos principales en el análisis de supervivencia es estimar esta función de supervivencia a partir de los datos observados, especialmente cuando existen observaciones censuradas. Para ello, se emplea comúnmente el estimador no paramétrico de Kaplan-Meier, que proporciona una aproximación de  $S(t)$  en los momentos en los que ocurre el evento. Este método se basa en probabilidades condicionales acumulativas, donde la supervivencia en un instante dado depende de haber sobrevivido en todos los instantes anteriores. La fórmula del estimador de Kaplan-Meier es:

$$S(t_j) = S^{t_{j-1}} * \frac{n_j - d_j}{n_j}$$

donde:



$n_j$  : número de individuos en riesgo en el tiempo  $t_j$ ,

$d_j$  : número de eventos observados en el tiempo  $t_j$  .

Alternativamente, también se puede expresar como el producto acumulativo de las probabilidades de supervivencia:

$$S^{\wedge}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

Este método permite estimar la probabilidad de supervivencia incluso en presencia de datos censurados, lo que lo hace especialmente útil en estudios longitudinales como el presente.

Además de la función de supervivencia, se consideran medidas descriptivas del tiempo de supervivencia, como la mediana y los percentiles. La mediana se define como el tiempo en el cual la probabilidad de supervivencia es del 50%:

A diferencia de la media, la mediana puede calcularse incluso cuando existen observaciones censuradas, proporcionando una medida robusta del tiempo central de supervivencia. Los cuartiles y otros percentiles también son útiles para describir la distribución del tiempo hasta el evento, especialmente en análisis comparativos entre grupos o condiciones diferentes.

### Teoría de la Penalización de Clientes

La penalización de clientes según Thaler & Sunstein (2008) es una estrategia utilizada por las empresas para influir en el comportamiento del consumidor y asegurar el cumplimiento de ciertas reglas o políticas. En el contexto de los sistemas de devolución de envases, la penalización generalmente implica un cargo adicional si los clientes no devuelven los envases dentro de un período especificado. Este enfoque se basa en la teoría económica del comportamiento, que sugiere que los individuos responden a incentivos y desincentivos financieros en su toma de decisiones.

Según la teoría de los incentivos, las penalizaciones pueden ser una herramienta efectiva para motivar a los clientes a devolver los envases, ya que el costo de no devolver el envase se percibe como una pérdida financiera. Sin embargo, para que las penalizaciones sean efectivas, deben ser percibidas como justas y razonables por los clientes. Si la penalización se percibe como excesiva o injustificada, podría generar insatisfacción y afectar negativamente la relación del cliente con la empresa.

En este sentido, es fundamental que las penalizaciones estén alineadas con los patrones de consumo reales de los clientes. Por ejemplo, si un cliente compra un detergente de 5 litros, es poco probable que lo consuma en 15 días, por lo que aplicar una penalización en este plazo podría considerarse injusto. Por lo tanto, la determinación del tiempo óptimo antes de aplicar la penalización debe basarse en un análisis riguroso de los datos de devolución y en una comprensión clara de los patrones de uso de los productos por parte de los clientes. La teoría de la penalización también sugiere que las penalizaciones deben ser lo suficientemente significativas como para motivar el comportamiento deseado, pero no tan altas como para disuadir a los clientes de participar en el programa de devolución. Un equilibrio adecuado es crucial para mantener la eficiencia del sistema de economía circular y asegurar la sostenibilidad a largo plazo.



## Distribuciones de Probabilidad Empíricas y Teóricas

En estadística, las distribuciones de probabilidad son herramientas esenciales para comprender cómo se comportan los datos. Estas pueden ser de dos tipos principales: **empíricas** y **teóricas**.

### 1. Distribuciones Empíricas.

Representan el comportamiento real de un conjunto de datos observado. Estas distribuciones se construyen directamente a partir de los datos recolectados y reflejan cómo se distribuyen las observaciones en una muestra específica.

**Ejemplo:** En Circulo S.A., la distribución empírica del tiempo de devolución de envases muestra cómo se comportan los ciclos completos en la práctica, permitiendo identificar patrones y tendencias específicas.

### 2. Distribuciones Teóricas.

Estas son modelos matemáticos que describen el comportamiento esperado de los datos bajo ciertas condiciones. Ejemplos comunes incluyen la distribución normal, exponencial, y Weibull.

**Ejemplo:** Ajustar la distribución empírica del tiempo de devolución de envases a una distribución teórica permite realizar inferencias sobre los tiempos esperados de devolución y evaluar qué modelo matemático se aproxima mejor al comportamiento real.

## Empresas Impulsadas por Datos

Las empresas impulsadas por datos son aquellas que toman decisiones estratégicas y operativas basadas en la recolección, análisis y uso de datos en lugar de depender de la intuición o la experiencia únicamente. Estas empresas emplean tecnologías y metodologías para captar datos continuamente y transformarlos en información útil que guíe la toma de decisiones. En este contexto, la analítica avanzada, los sistemas de gestión de datos y los modelos predictivos son fundamentales para mejorar la eficiencia, la rentabilidad y la experiencia del cliente. Circulo S.A. al buscar entender y predecir los tiempos de devolución de los clientes a través de datos recolectados adopta una posición de empresa impulsada por datos.

## Data Mining

El Data Mining, o minería de datos, es explicado como el proceso de descubrir patrones significativos, relaciones y tendencias dentro de grandes volúmenes de datos mediante el uso de técnicas estadísticas, de inteligencia artificial y de aprendizaje automático según Han, Kamber, Pei (2011). En investigaciones como la presente, donde existe un gran volumen de datos históricos sobre compras, devoluciones y comportamiento de los clientes, el Data Mining resulta una herramienta crucial para obtener información valiosa que guíe la toma de decisiones.

El objetivo principal del Data Mining es transformar datos crudos en información significativa y útil mediante la identificación de patrones ocultos o relaciones entre múltiples variables. Los algoritmos de clasificación, predicción y agrupamiento, como los



árboles de decisión, redes neuronales y algoritmos de clustering, permiten segmentar clientes, predecir comportamientos futuros y optimizar procesos de negocio.

En el contexto de Circulo S.A., Python fue utilizado para analizar datos históricos de devolución y predecir el comportamiento de los clientes. Técnicas de visualización, como gráficos de dispersión, y algoritmos de clustering permitieron identificar patrones en el tiempo de devolución de envases, facilitando la toma de decisiones sobre el momento óptimo para aplicar penalizaciones.

### Machine Learning No Supervisado

El **Machine Learning No Supervisado** analiza datos sin etiquetas predefinidas, identificando patrones y estructuras subyacentes. Sus principales técnicas incluyen:

1. **Clustering (Agrupamiento):** Agrupa datos en subconjuntos similares. En Circulo S.A., puede segmentar clientes según patrones de devolución, como frecuencia o tipo de envase, para adaptar estrategias específicas.
2. **Análisis de Componentes Principales (PCA):** Reduce la cantidad de variables manteniendo la esencia de los datos. Es útil para identificar los factores más influyentes en el tiempo de devolución, optimizando los análisis.
3. **Análisis de Asociaciones:** Identifica relaciones entre variables. Por ejemplo, puede revelar cómo el tipo de producto y la modalidad de entrega afectan las tasas de devolución.

## APLICACION

### Ajuste del Análisis de Supervivencia a Ciclos Repetitivos

En el análisis de supervivencia convencional, cada unidad de observación experimenta un único evento final (por ejemplo, la muerte o el fallo de un sistema), y el tiempo hasta ese evento es lo que se modela. Sin embargo, en este caso, el evento de interés es la devolución de un envase, pero el mismo envase puede ser asignado a un cliente, devuelto al almacén y luego reasignado a otro cliente. Esto genera un ciclo repetitivo de asignación y devolución para el mismo envase, lo que no es común en el análisis de supervivencia tradicional.

### Implicaciones en el Análisis:

Este patrón de asignación-devolución repetida crea una situación donde un mismo envase puede generar múltiples observaciones, lo que complicaría el análisis si no se maneja adecuadamente. Si no se separan los eventos de asignación y devolución como observaciones independientes, el análisis podría sobrestimar la duración promedio de los eventos o generar inconsistencias en los resultados, dado que algunos envases habrán sido devueltos varias veces.

### Solución Propuesta:

Para evitar este problema, es necesario estructurar el análisis de forma que cada ciclo de asignación-devolución se considere como una observación independiente. Para ello, se ha diseñado un procedimiento que:



1. Identificar cada movimiento de asignación de cada envase a un cliente.
2. Empareja esa asignación con la devolución correspondiente al almacén.
3. Trata las asignaciones sin devolución como observaciones censuradas, asegurando que no se dupliquen ciclos posteriores en el análisis para un mismo envase.

Esta aproximación permite realizar un análisis de supervivencia centrado en cada ciclo individual de uso de los envases, lo que asegura que los resultados reflejen correctamente el tiempo hasta la devolución en cada asignación.

#### Eventos de interés:

El evento de interés en un análisis de supervivencia es aquel suceso cuyo tiempo hasta que ocurre estamos tratando de medir y analizar. En este caso, el evento de interés es la devolución del envase al almacén o warehouse. Se clasifica como el momento en que un envase, previamente asignado a un cliente, es devuelto al sistema logístico, completando así el ciclo de reutilización de la economía circular en la empresa.

La definición del evento de interés es crucial porque determina cómo se estructuran los datos y cómo se interpreta el análisis. En este contexto, podemos definirlo así:

**Evento de interés:** La devolución efectiva del envase desde el cliente al depósito o warehouse, que marca la culminación del ciclo de uso y devolución dentro del sistema de la economía circular. Para los análisis estadísticos, el evento se codifica con un valor de "1".

**Censura:** Si un envase no se ha devuelto dentro del tiempo de observación, el evento no se ha producido. En este caso, los datos estarían censurados, lo que significa que no se sabe cuándo, o si, el envase será devuelto en el futuro. Estos se codifican como "0" en el análisis de supervivencia. La censura puede ser a la derecha, es decir, cuando el tiempo de observación termina antes de que el evento ocurra.

En el análisis de los registros capturados por el sistema, se identifica una situación extraordinaria.

Debido a un error humano por parte de los operarios, algunos envases han sido asignados a un cliente y, de manera consecutiva, a otro cliente sin haber pasado previamente por el almacén. Para abordar este inconveniente, se considerará que el envase retornó al depósito el mismo día en que fue asignado al siguiente cliente. De esta manera, se cerrará el ciclo del envase, registrando el tiempo transcurrido como un evento de interés e iniciando un nuevo ciclo.

## PROCESAMIENTO DE DATOS

### Recolección y Estructura de Datos

Circulo S.A. ha desarrollado un sistema de recolección de datos automatizado a través del software **TRAC**, diseñado para mantener un seguimiento detallado de cada envase mediante un código QR único. Este sistema permite digitalizar el ciclo completo de cada envase, desde su asignación al cliente hasta su devolución, proporcionando información clave para el análisis de supervivencia.

El flujo de datos en cada venta, representado en el **Gráfico N°1**, sigue los siguientes pasos:



1. **Recepción del Pedido (CRM):** El proceso inicia cuando el cliente realiza un pedido, que se registra en el sistema CRM de la empresa. Esto incluye la información del cliente, los productos solicitados y otros detalles relevantes de la transacción.
2. **Carga del Pedido (ERP):** Una vez registrado el pedido, se transfiere al sistema ERP para gestionar el inventario y los aspectos logísticos. Aquí se asignan los productos y se asegura la disponibilidad de estos para su preparación.
3. **Preparación y Asignación de Pedido (TRAC):** En el sistema TRAC, cada envase es asignado al cliente específico y se registra mediante el escaneo del código QR. Esto permite saber en todo momento qué cliente tiene asignado cada envase, facilitando el seguimiento y evitando pérdidas.
4. **Entrega:** El pedido se entrega al cliente, momento en el cual se confirma la asignación del envase mediante el sistema TRAC, garantizando que la información en el sistema se mantenga actualizada.
5. **Carga de Envases Recibidos (TRAC):** Al devolver el envase, este se escanea nuevamente para registrar su retorno al depósito. Esto permite cerrar el ciclo y actualizar el estado del envase, marcándolo como disponible para su próxima asignación.

Este sistema de recolección de datos garantiza un seguimiento preciso de cada envase y permite crear una base de datos detallada de movimientos. Estos datos son cruciales para el análisis de supervivencia, ya que se pueden identificar patrones de devolución y evaluar factores que afectan el tiempo hasta la devolución.

A continuación, se presenta el flujograma de recolección de datos en cada venta.

**Gráfico N°1: Flujograma de Datos recolectados en las ventas**



Fuente: Elaboración propia

## 1. Descripción del Sistema TRAC y los Registros de Movimientos



- **Sistema TRAC:** El sistema TRAC, desarrollado por Circulo S.A., permite la digitalización y seguimiento de cada envase mediante un código único (ID). A través de una aplicación de uso interno, se registra cada asignación de envase a un cliente y su devolución al depósito. Esto crea un historial de movimientos clave para realizar el análisis de supervivencia.
- **Tipos de Datos Recolectados:** Los datos capturados incluyen la fecha de cada movimiento, el ID único del envase, el tipo de envase (1 litro, 5 litros, potes), el lugar del movimiento (cliente o almacén), y el cliente receptor, lo cual permite rastrear el ciclo de cada envase.

Dentro de esta clasificación se encuentran 4 tipos de envases

1. Envase 2 y Envase 5: Son bidones de 5 Litros, pero de diferentes modelos, siendo el envase 2 envases más nuevos, pero más económicos.
2. Envase 3: Botellas de 1 Litro
3. Envase 4: Botes de 455 gramos utilizados para bicarbonato de Sodio

- **Gráfico N°2: Diagrama de Tablas Relacionales de Base de Datos Trac**



Fuente: Elaboración propia

## 2. Integración de Datos de Otros Sistemas

- **Sistema ERP:** Además de TRAC, se incorporan datos del ERP de la empresa, que detallan las ventas por cliente, incluyendo la fecha de compra, productos vendidos y cantidades. Estos datos ayudan a complementar el análisis, permitiendo un enfoque más amplio al estudiar el comportamiento del cliente.
- **Registros Manuales de Repartidores:** Los repartidores de Circulo S.A. registran información sobre quién recibió los pedidos y los envases devueltos. Estos datos manuales ofrecen una perspectiva adicional y ayudan a contextualizar los movimientos de cada envase.

## 3. Estructura de la Tabla de Movimientos



La tabla de movimientos de TRAC originalmente almacena cada movimiento en filas independientes. Para cada registro se incluye:

- **ID\_Movimiento:** Número único que identifica cada movimiento registrado en el sistema.
- **Estado:** Indica si el envase está asignado a un cliente (client) o en el almacén (warehouse).
- **ID\_Envase:** Código único que representa cada envase.
- **ID\_Cliente:** Identificador del cliente al que se asigna el envase. En caso de que el estado sea "warehouse," se asigna un valor de 0.
- **Fecha de Creación:** Fecha y hora exactas de registro del movimiento en el sistema.
- **Cliente:** Nombre del cliente al que fue asignado el envase.

**Tabla N°1: Formato de base de datos generada por el sistema Trac.**

ID_Movimiento	Estado	ID_Envase	ID_Cliente	Fecha de creación	Cliente
1237	client	3831	248	2024-04-17 15:19:22.052027 +00:00	CECILIA MONTERO
1238	client	3223	248	2024-04-17 15:19:22.059782 +00:00	CECILIA MONTERO
1239	client	3105	248	2024-04-17 15:19:22.069924 +00:00	CECILIA MONTERO
1240	warehouse	3249	0	2024-04-17 15:19:54.165666 +00:00	
1244	client	3889	528	2024-04-17 15:20:33.990778 +00:00	MARCELO GOMEZ
1245	client	3424	528	2024-04-17 15:20:34.000351 +00:00	MARCELO GOMEZ

**Fuente: Base de datos sistema Trac.**

Esta tabla permite identificar los momentos de asignación y devolución de cada envase. Para cada envase, la secuencia de movimientos puede determinar si el envase fue devuelto al almacén y luego reasignado, o si permaneció en posesión del cliente. Además, los movimientos con el estado "warehouse" sirven para marcar el cierre de un ciclo de uso y reutilización.





#### 4. Transformación de la Tabla para el Análisis de Supervivencia

La transformación de la tabla es necesaria para asegurar que cada observación represente un ciclo completo de asignación y devolución de los envases, o en su defecto, identificar los ciclos que quedaron sin devolver como censurados. Este ajuste permite definir claramente los eventos y los datos censurados, elementos fundamentales para un análisis de supervivencia adecuado.

Además, la transformación facilita el cálculo de la duración de cada ciclo, que es la variable de tiempo esencial para el análisis de supervivencia. Este dato estructurado permite estimar las tasas de devolución de manera precisa y realizar comparaciones entre distintos tipos de envases. En definitiva, la transformación proporciona una base de datos que permite un análisis detallado y confiable, esencial para modelar los patrones de devolución de los envases

Para adaptar la base de datos al análisis de supervivencia, se realizó una transformación de los datos a una estructura de "ciclos" de asignación y devolución.

**Clasificación de Ciclos:** Se estableció un procedimiento en SQL (Apéndice 1.1) para clasificar cada ciclo de la siguiente forma: +

- **Ciclo Completo:** Cuando el envase es asignado a un cliente y posteriormente devuelto al almacén.
- **Ciclo Censurado:** Si el envase fue asignado, pero no devuelto dentro del período de observación.
- **Ciclo Cliente-Cliente:** En casos de errores operativos, donde un envase fue reasignado a un nuevo cliente sin pasar por el almacén, se consideró como un ciclo completo, usando la fecha de la segunda asignación como fecha de devolución.

#### 5. Estructura Final de la Base de Datos para el Análisis

La base de datos final se configuró de manera que cada ciclo representara una observación independiente. Cada ciclo, se corresponde con un registro en la tabla final que incluye:

- **ID del envase:** identificador único para cada envase.
- **Tipo de envase:** clasificación según la capacidad del envase.
- **Fecha de inicio del ciclo:** fecha en que el envase es asignado al cliente.
- **Fecha de fin del ciclo:** fecha de devolución al depósito, o fecha de la segunda asignación en los ciclos cliente-cliente.
- **Estado del ciclo:** 1 si el ciclo es completo y 0 si es censurado.
- **Indicador del tipo de ciclo:** una columna para identificar de que tipo de ciclo se trata.
- **Días de duración del ciclo:** cantidad de días entre la asignación y la devolución. En el caso de los ciclos censurados se utiliza al día del cierre (26/10/2024).



**Tabla N°2: Base de datos estructurada para el 18 análisis.**

asset_id	tipo_envase	fecha_asignacion	fecha_devolucion	ciclo_completo	tipo_ciclo	dias_cic
2902	2	30/9/2024		0	censurado	26
2903	2	3/5/2024	19/7/2024	1	cliente-warehouse	76
2903	2	19/7/2024	22/8/2024	1	cliente-cliente	34
2903	2	22/8/2024		0	censurado	65
2904	2	2/5/2024	25/6/2024	1	cliente-cliente	53
2904	2	25/6/2024		0	censurado	123
2905	2	2/5/2024	23/7/2024	1	cliente-warehouse	81
2905	2	23/7/2024		0	censurado	95
2906	2	30/4/2024	1/8/2024	1	cliente-cliente	92

Fuente: Elaboración propia.

### Limpieza de Base de Datos

Una vez estructurada la base de datos, se llevaron a cabo procedimientos de limpieza y organización para garantizar la integridad y consistencia de los datos. Este proceso incluyó:

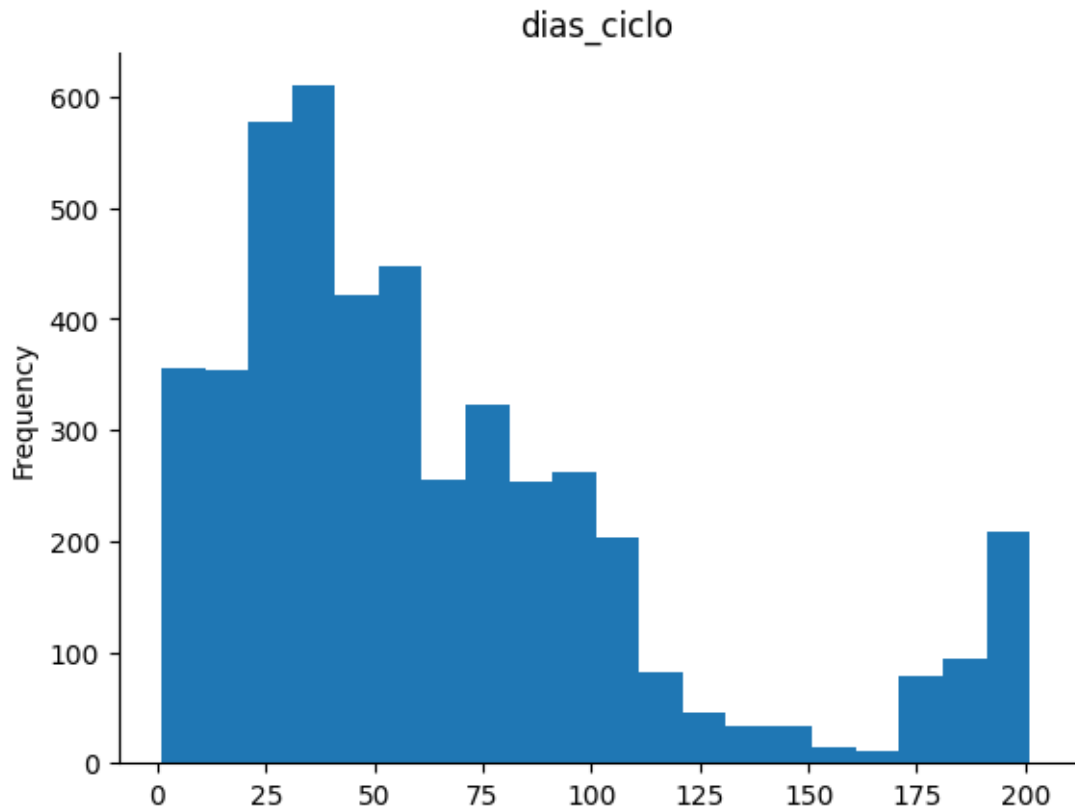
- **Eliminación de valores nulos:** Se revisaron las columnas clave para identificar y eliminar registros con valores nulos, especialmente en la columna de días de duración del ciclo, ya que estos valores afectaban el cálculo de tiempos de supervivencia.
- **Filtrado de registros duplicados:** Se verificaron posibles duplicados en los registros de movimientos para evitar contar múltiples veces un mismo ciclo.
- **Verificación temporal:** Se comprobó la secuencia cronológica de los registros para asegurar que los movimientos tuvieran sentido temporal, por ejemplo, que un envase no apareciera asignado a un cliente antes de haber sido devuelto al depósito.

### ANÁLISIS INICIAL DE LA DISTRIBUCION DE PROBABILIDAD

Para comprender la distribución de los tiempos de devolución, primero se analizó la curva de frecuencia de todos los datos, incluyendo tanto los ciclos completos como los censurados. Este gráfico de distribución inicial mostró una curva peculiar con un aumento al final, lo que sugería la necesidad de separar los ciclos completos de los censurados para una interpretación más precisa.

Luego, se construyó un segundo gráfico de frecuencia excluyendo los datos censurados, lo que permitió observar la distribución de los ciclos completos de forma más clara y sin sesgo. Este análisis preliminar proporcionó una base útil para el análisis de supervivencia, permitiendo entender mejor los patrones de tiempo de devolución.

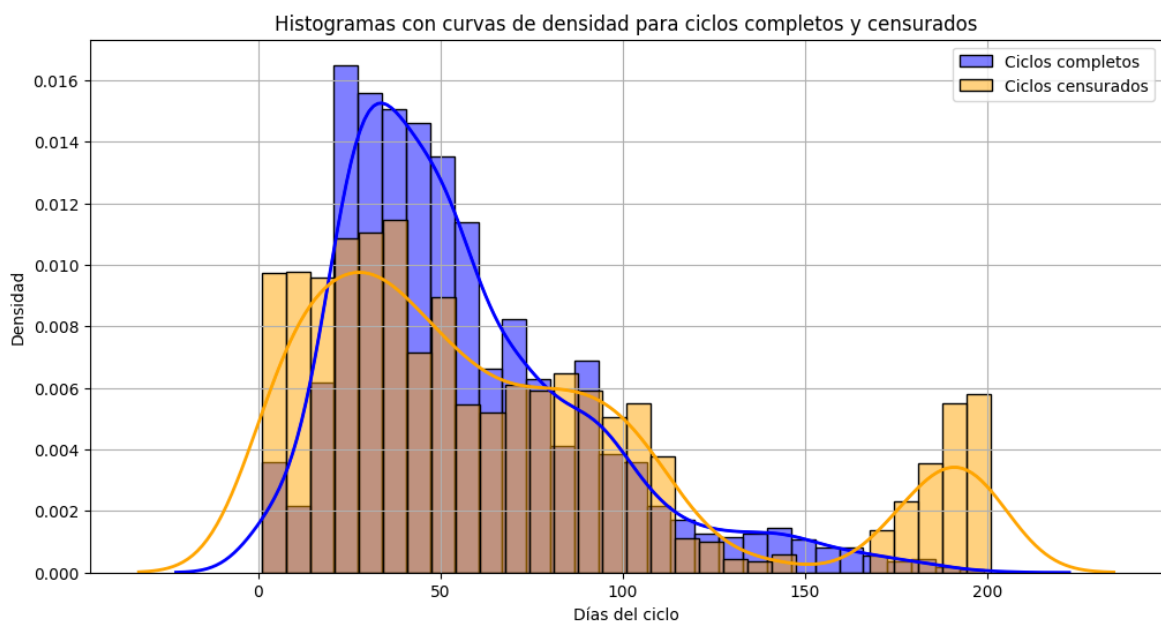
### Gráfico N°3: Distribución de probabilidad empírica de todos los registros



Fuente: Elaboración Propia

Por el crecimiento del final, se realiza un histograma con curvas de densidad para ciclos completos y censurados (Código en Apéndice 2.2)

Gráfico N°4: Histograma con curvas de densidad por tipos de ciclos.



Fuente: Elaboración propia



Teniendo las distribuciones empíricas de los ciclos completos y censurados, se procede a ajustar las mismas a distribuciones teóricas (Apéndice 2.3) obteniendo los siguientes resultados:

### Resultados del ajuste:

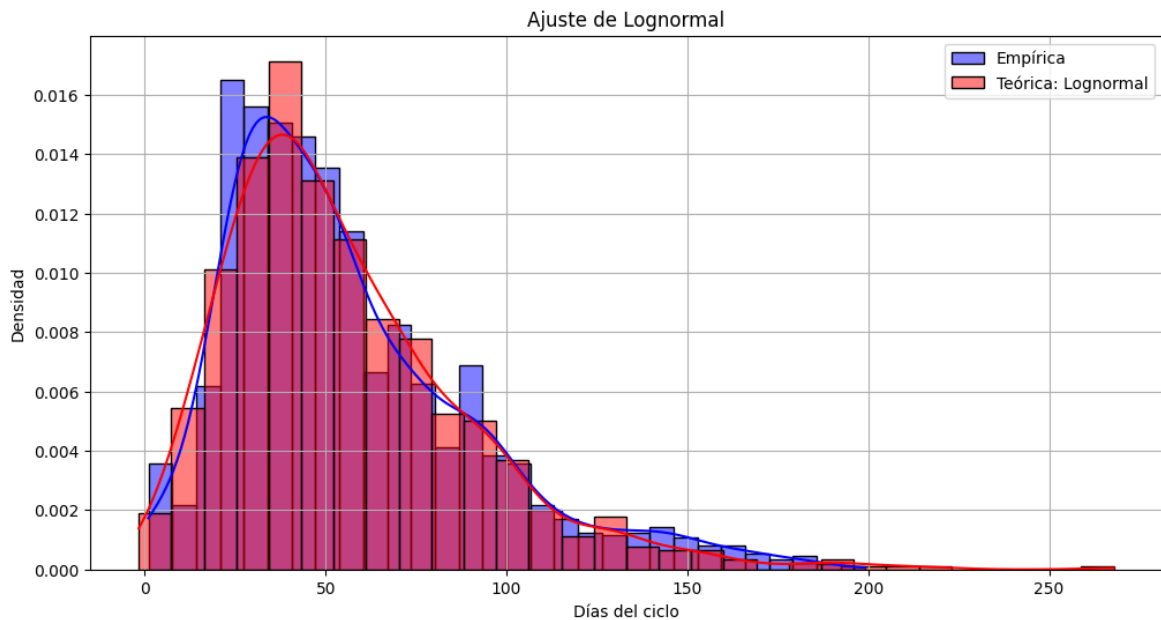
Weibull: KS-Statistic = 0.0581, p-value = 0.0000

Exponencial: KS-Statistic = 0.2287, p-value = 0.0000

Lognormal: KS-Statistic = 0.0355, p-value = 0.0277

La **Lognormal** es la distribución que mejor representa los datos de ciclos completos.

### Gráfico N°5: Ajuste de distribución de ciclos completos a distribución Lognormal.



Fuente: Elaboración propia.

### Interpretación del Ajuste a la Distribución Lognormal

#### 1. Forma de la Distribución:

La **distribución Lognormal** es asimétrica y presenta una "cola" hacia la derecha. Este tipo de distribución es común en situaciones donde los valores no pueden ser negativos y donde hay una tendencia natural a tener una mayor concentración de eventos en valores más bajos, con algunos valores que se extienden mucho más.

En el contexto de los **días de ciclos completos**, la Lognormal se ajusta bien porque la mayoría de los envases se devuelven en un plazo más corto, pero algunos clientes toman mucho más tiempo, generando una distribución sesgada hacia la derecha.

#### 2. Parámetros de la Distribución:



La Lognormal tiene dos parámetros clave: **media y desviación estándar** en el espacio logarítmico. Esto significa que el análisis no solo captura el tiempo promedio de devolución de los envases, sino también la variabilidad en estos tiempos. (Apéndice 2.4)

**Mu (media en log): 3.8242220022699343**

**Sigma (desviación estándar en log): 0.7279570380154001**

**Mediana: 45.8 días**

- En una distribución Lognormal, la mediana es un indicador central más adecuado que la media (en escala no logarítmica) debido a la asimetría de la distribución. Indica que el 50% de los envases son devueltos en un tiempo igual o menor a 45.8 días.

**Percentil 90: 116.4 días**

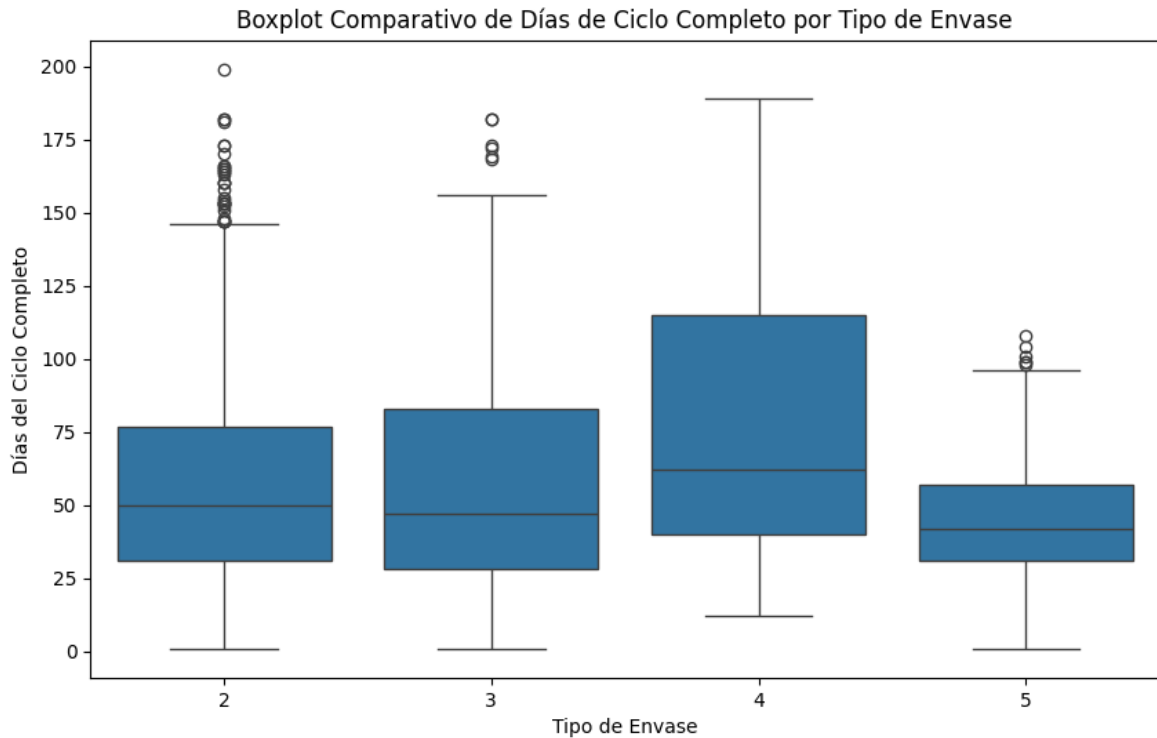
- Este valor indica que el 90% de los envases son devueltos en un tiempo igual o menor a 116.4 días. El percentil 90 es útil para identificar los tiempos más largos en los que se espera que la mayoría de los envases sean devueltos, siendo relevante para decisiones sobre penalizaciones o seguimiento de devoluciones.

### 3. Tiempo de Devolución:

Al ajustarse a una distribución Lognormal, podemos interpretar que el **tiempo de devolución** sigue un patrón que varía considerablemente. Muchos clientes devuelven los envases en tiempos relativamente cortos, pero una proporción más pequeña los retiene mucho más tiempo.

El análisis de *boxplot* es conveniente en este contexto porque permite visualizar de manera clara la dispersión y distribución de los días de ciclo completo según el tipo de envase. Esto es útil para observar la variabilidad en los tiempos de devolución entre los distintos tipos de envase y detectar posibles diferencias que pueden ser importantes para la toma de decisiones. (Apéndice 2.5)

**Gráfico N°6: Grafico de Boxplot de ciclos completos por tipo de envase.**



**Fuente: Elaboración propia**

Este análisis sugiere que el tipo de envase influye en el tiempo que los clientes mantienen los envases antes de devolverlos. Es probable que el tipo 4 (Botes de bicarbonato de sodio) requiera medidas específicas para incentivar una devolución más rápida, mientras que el tipo 5 (Bidones de 5 Litros modelo antiguo) tiene un ciclo de devolución rápido y consistente.

Luego de tener una base de la distribución de los datos se procedió a realizar el análisis de supervivencia.

## Implementación del Análisis de supervivencia

Para llevar a cabo el análisis de supervivencia, se utilizó el lenguaje de programación Python, específicamente con las bibliotecas lifelines y matplotlib. Este análisis se estructuró en varias etapas, comenzando con la revisión de la distribución de probabilidad de los datos y continuando con la aplicación del modelo de Kaplan-Meier. A continuación, se presentan los pasos seguidos y los resultados obtenidos en cada fase del análisis:

### Paso 1: Instalación de las Bibliotecas Necesarias

El primer paso consistió en instalar las bibliotecas lifelines y matplotlib, esenciales para realizar el análisis de supervivencia y para graficar los resultados. (Ver apéndice)

### Paso 2: Importación de Bibliotecas y Carga de Datos

Luego, se importaron las bibliotecas necesarias y se cargó el conjunto de datos a analizar. La base de datos se cargó en un DataFrame de Pandas para facilitar su manipulación y análisis.



### Paso 3: Creación del Modelo de Kaplan-Meier

Se procedió a crear el modelo de Kaplan-Meier, que es adecuado para estimar la función de supervivencia en datos censurados. Para ello, se utilizó el modelo KaplanMeierFitter de la biblioteca lifelines.

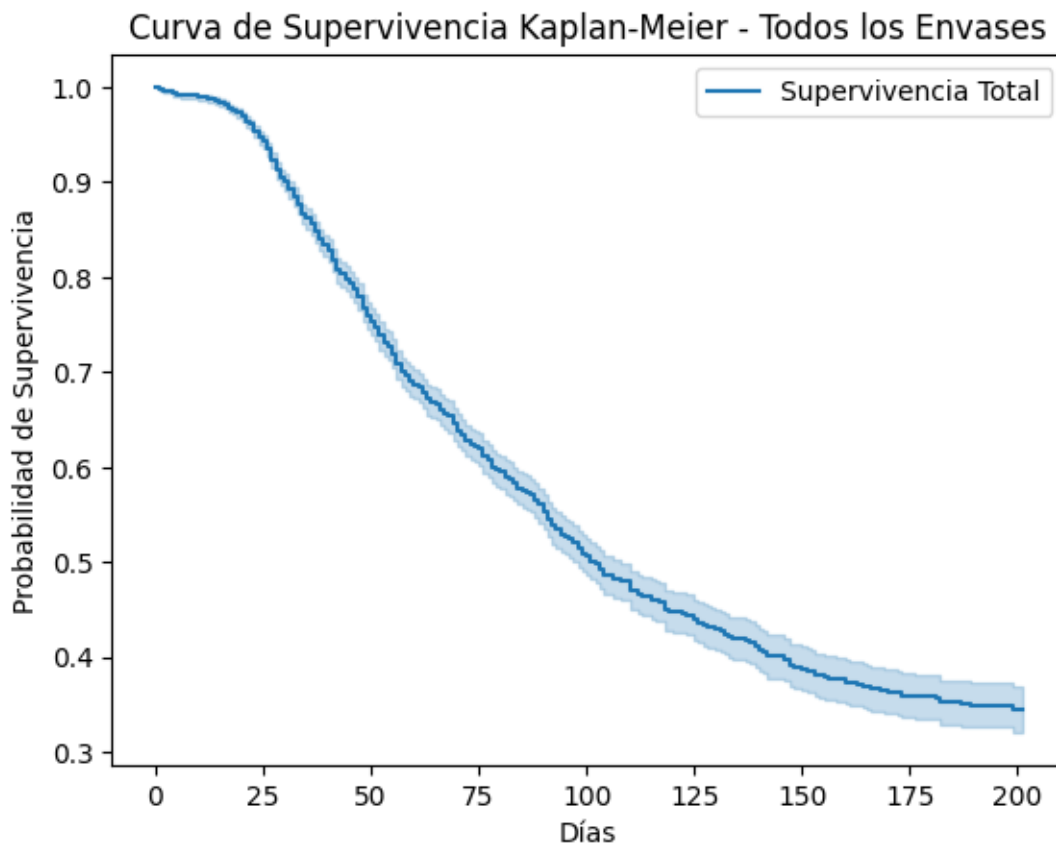
### Paso 4: Ajuste del Modelo a los Datos

El modelo fue ajustado a los datos. La columna que contenía la duración del ciclo (dias\_ciclo) se utilizó para indicar el tiempo transcurrido, mientras que la columna ciclo\_completo especificaba el estado del evento (1 si el evento ocurrió y 0 si el dato fue censurado).

### Paso 5: Visualización de la Curva de Supervivencia

A continuación, se graficó la curva de supervivencia utilizando matplotlib para visualizar la probabilidad de que un envase continúe sin ser devuelto a lo largo del tiempo. (Apéndice 2.6)

**Gráfico N°7: Curva de supervivencia de Kaplan-Meier.**



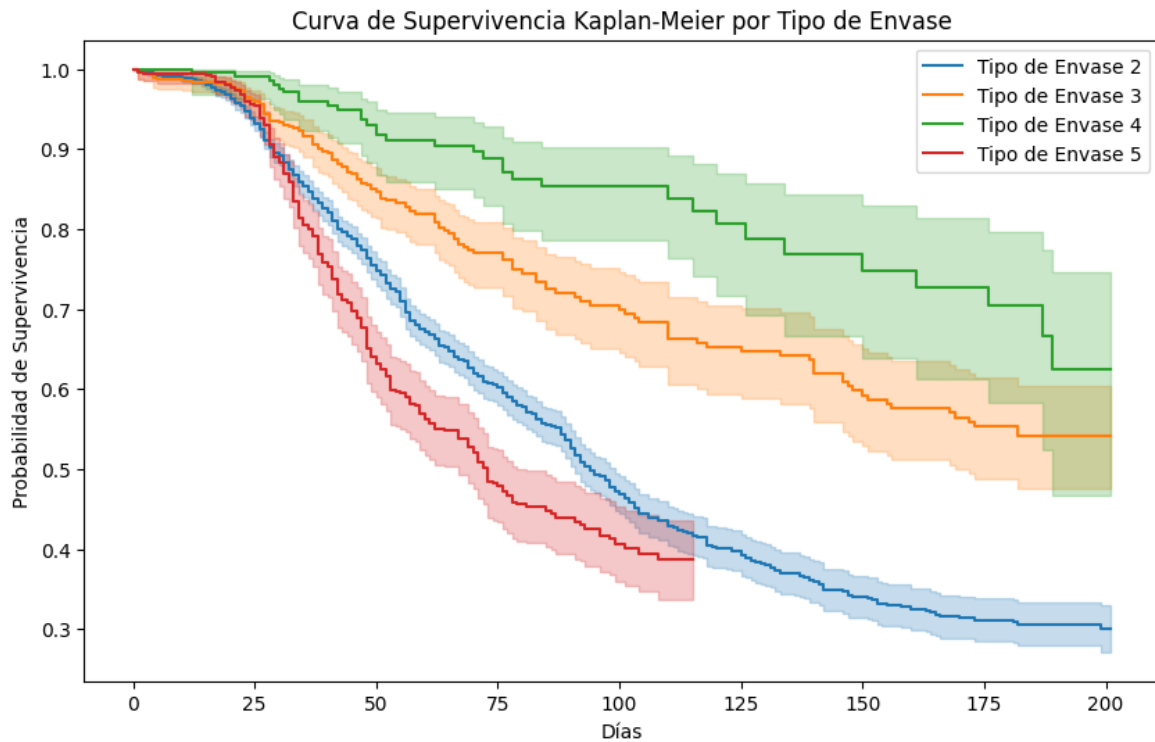
Fuente: Elaboración propia.



### Paso 6: Análisis por Grupo

Para analizar la supervivencia por tipo de envase, se ajustó el modelo de Kaplan-Meier para cada grupo de envases. Se utilizó un bucle para aplicar el modelo a cada valor único de tipo\_envase, comparando así las curvas de supervivencia de cada grupo. (Apéndice 2.7)

**Gráfico N°8: Curva de supervivencia de Kaplan-Meier por tipo de envase.**



**Fuente: Elaboración propia.**

### Paso 7: Interpretación de las curvas

#### Curva General de Todos los Envases

El gráfico que representa la curva de Kaplan-Meier para todos los envases combinados muestra un descenso casi lineal en la probabilidad de supervivencia, lo que indica que la tasa de devolución de envases es relativamente constante con el tiempo. Por ejemplo, en el día 100, la probabilidad de que un envase no haya sido devuelto es del 50%, y esta probabilidad disminuye a aproximadamente el 45% en el día 125. Este comportamiento refleja que, en promedio, los envases mantienen patrones similares en el tiempo de retorno.

#### Curvas por Tipo de Envase

En el análisis desagregado por tipo de envase, se destacan diferencias significativas:

Bidones de 5 litros (Tipos de envase 2 y 5):





Las curvas correspondientes a los bidones de 5 litros muestran patrones similares, aunque los bidones antiguos y de mejor calidad (tipo 5) presentan tasas de devolución ligeramente más rápidas que los bidones económicos (tipo 2).

Probabilidad de no devolución: Entre el 40% y 50% en el día 100, disminuyendo a un 40% en el día 125.

#### Botellas de 1 litro (Tipo de envase 3):

Estas botellas tienen un comportamiento diferenciado, con una menor probabilidad de devolución en comparación con los bidones de 5 litros. Esto podría deberse a que los hogares reutilizan estas botellas para otros fines, extendiendo su vida útil fuera de la economía circular de Circulo S.A.

Probabilidad de no devolución: 70% en el día 100, reduciéndose a 65% en el día 125.

#### Botes de bicarbonato (Tipo de envase 4):

Los botes de bicarbonato tienen la menor tasa de devolución entre todos los tipos de envases analizados. Esto resalta la necesidad de implementar estrategias específicas, como penalizaciones, para fomentar su retorno.

Probabilidad de no devolución: 85% en el día 100, disminuyendo a 80% en el día 125.

#### Comparación entre Curvas

Al comparar la curva general con las curvas segmentadas por tipo de envase, se observa que la curva global está dominada por los patrones de los bidones de 5 litros, que representan la mayor parte de los movimientos registrados. Sin embargo, los envases de menor capacidad (botellas de 1 litro y botes de bicarbonato) presentan desafíos específicos en términos de retorno, ya sea por su uso prolongado en los hogares o por un bajo incentivo para ser devueltos.

### **Recomendaciones**

A partir de la formalización del análisis de supervivencia con las curvas de Kaplan-Meier se pueden proponer las siguientes recomendaciones:

#### **Implementación de penalizaciones específicas por tipo de envase:**

Para los envases de bicarbonato (tipo de envase 4), que presentan la menor probabilidad de devolución, se recomienda establecer penalizaciones más estrictas y claras para incentivar su retorno. Este tipo de envase podría tener mayor probabilidad de ser descartado por los clientes debido a su tamaño y material, lo que refuerza la necesidad de una estrategia de penalización diferenciada.

#### **Revisión de la política de penalización para envases de 5 litros (tipos 2 y 5):**

Dado que los envases económicos (tipo 2) tienen una tasa de devolución ligeramente inferior en comparación con los envases de mayor calidad (tipo 5), se podrían evaluar incentivos económicos para fomentar la devolución de los envases de menor costo.



Además, se sugiere ajustar los plazos de penalización para alinearse con los tiempos promedio de retorno observados en el análisis.

**Estrategias educativas para las botellas de 1 litro (tipo 3):**

Estas botellas tienen una alta probabilidad de reutilización fuera del sistema de economía circular. Se recomienda implementar campañas educativas para concientizar a los clientes sobre la importancia de devolverlas, junto con incentivos para fomentar su retorno, como descuentos en futuras compras al devolver varias botellas.

**Focalización en los puntos de recolección y logística inversa:**

Con base en los patrones observados en el análisis de supervivencia, se podrían optimizar los puntos de recolección para los envases menos devueltos. Por ejemplo, incluir recolecciones programadas o incentivos en lugares estratégicos.

**Monitoreo continuo y segmentación de estrategias por cliente:**

A partir de los datos analizados, es clave segmentar a los clientes según sus patrones de devolución y establecer estrategias personalizadas. Los clientes con tasas de retorno bajas pueden recibir recordatorios o incentivos adicionales.

**Conciencia de una cultura basada en datos:**

Se recomienda fomentar una cultura organizacional que valore la recopilación, análisis y utilización estratégica de los datos. Esto implica capacitar a los equipos en la interpretación de métricas y análisis, promoviendo decisiones fundamentadas en evidencia. La gestión eficiente de los envases puede ser optimizada si se utiliza de manera continua el análisis de los datos recolectados, lo que permitirá identificar patrones de comportamiento y anticiparse a posibles problemas operativos.

**Diseño estratégico de la base de datos:**

Es fundamental diseñar y estructurar la base de datos de manera estratégica, asegurando que recoja todos los datos relevantes para el negocio, desde los movimientos de los envases hasta las características de los clientes. Una base de datos bien diseñada no solo mejora la calidad del análisis de supervivencia y otros estudios estadísticos, sino que también facilita la generación de conocimiento estratégico. Considerar de antemano las variables clave y estructurar los datos con miras a futuros análisis permitirá responder con mayor agilidad a las necesidades del negocio y a los cambios en el entorno competitivo.

## Conclusiones

El presente análisis, centrado en el comportamiento de devolución de envases en el marco del modelo de economía circular de Circulo S.A., ha permitido responder a las preguntas de investigación planteadas, así como alcanzar los objetivos generales y específicos definidos. A partir del análisis de supervivencia y las distribuciones de probabilidad, se obtuvieron hallazgos clave que contribuyen tanto al entendimiento del problema como a la toma de decisiones estratégicas.



La distribución de probabilidad empírica ajustada a una distribución Lognormal permitió observar que el percentil 90 del tiempo de devolución se encuentra en el día 116, es decir, el 90% de los envases son devueltos antes de esa fecha, considerando únicamente los ciclos completos. Sin embargo, al realizar un análisis de supervivencia, que incorpora tanto los datos de ciclos completos como censurados (es decir, envases que no han sido devueltos durante el período de observación), se observó que la probabilidad de que un envase no sea devuelto en el día 125 es del 45%. Esta diferencia refleja la importancia de considerar datos censurados en un análisis más robusto, ya que dichos registros castigan las probabilidades al incluir aquellos casos en los que los envases han permanecido largos periodos sin devolución. Esto resalta la utilidad del análisis de supervivencia para modelar de manera más realista las dinámicas de devolución.

En términos de comparación entre los tipos de envase, las curvas de supervivencia Kaplan-Meier revelaron diferencias significativas. Los bidones de 5 litros (tipos 2 y 5) mostraron patrones similares, aunque los envases de menor costo (tipo 2) presentan una ligera menor probabilidad de devolución, posiblemente asociada a su calidad percibida. Por otro lado, las botellas de 1 litro (tipo 3) y los botes de bicarbonato (tipo 4) presentan tasas de devolución más bajas y tiempos más largos, lo que sugiere la necesidad de estrategias específicas para estos envases. En particular, los botes de bicarbonato tienen la menor probabilidad de retorno, con una probabilidad de alrededor del 80% de no ser devueltos incluso después de 125 días, lo que refuerza la importancia de implementar penalizaciones más estrictas para este tipo de envase.

El análisis permitió identificar que la distribución Lognormal es la más adecuada para modelar el tiempo de devolución de envases completos. Asimismo, el análisis de supervivencia ofreció una estimación del tiempo promedio de devolución, así como las diferencias significativas entre los tipos de envase, donde los factores como el tamaño, la calidad percibida y la reutilización doméstica parecen influir considerablemente en los patrones de devolución.

En conclusión, la incorporación de herramientas estadísticas avanzadas como el análisis de supervivencia y las distribuciones de probabilidad permitió obtener una comprensión integral de los tiempos de devolución de los envases y los factores asociados a las demoras. Los resultados obtenidos no solo reflejan la importancia de considerar datos censurados en los análisis, sino que también ofrecen una base sólida para el diseño de estrategias diferenciadas, adaptadas a las características de cada tipo de envase, que optimicen la sostenibilidad económica y ambiental del modelo de economía circular de Circulo S.A.



## Referencias

- Arana C. (2015). *Análisis de datos en investigación social* (1ª ed.). Editorial Gedisa.
- Barrios J. M., & Torres F. (2016). *Estadística multivariante: Técnicas y aplicaciones* (1ª ed.). Editorial Universitaria Ramón Areces.
- Cox J., Giorgi S., Sharp V., Strange K., Wilson D. C., & Blakey N. (2010). Household waste prevention — a review of evidence.
- Geissdoefer M. (2017). The Circular Economy – A new sustainability paradigm? *Journal of Cleaner Production*.
- Gómez R., & Fernández M. (2008). *Métodos de investigación en ciencias sociales* (2ª ed.). Editorial Síntesis.
- Han J., Kamber M., & Pei J. (2011). *Minería de datos: Conceptos y técnicas* (3ª ed.). Morgan Kaufmann.
- Hair J. F., Black W. C., Babin B. J., & Anderson R. E. (2014). *Análisis multivariante de datos* (6ª ed.). Pearson.
- Hosmer D. W., Lemeshow S., May S. (2008). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. John Wiley & Sons.
- Hernández Sampieri R., Fernández Collado C., & Baptista Lucio M. D. (2014). *Metodología de la investigación* (6ª ed.). McGraw-Hill Interamericana.
- James G., Witten D., Hastie T., & Tibshirani R. (2021). *An Introduction to Statistical Learning with Applications in R* (Second Edition). Springer.
- López A., & Sánchez J. (2019). Diseño de investigación en ciencias sociales. *Revista de Métodos Cuantitativos*, 30(1), 45-67.
- Martínez C. (2010). *Fundamentos de la investigación cuantitativa* (1ª ed.). Ediciones Universitarias.
- Rivas M. (2017). Métodos estadísticos para el análisis de supervivencia: una revisión. *Revista de Estadística Aplicada*, 45(2), 112-130.
- Sharp V., Giorgi S., & Wilson D. C. (2010). Methodologies for monitoring and evaluating household waste prevention.
- Tan P. N., Steinbach M., & Kumar V. (2018). *Introducción a la minería de datos* (2ª ed.). Pearson.
- Thaler R. H., Sunstein C. R. (2008). *Un Pequeño Empujón*. Yale University Press.
- Witten I. H., Frank E., & Hall M. A. (2016). *Minería de datos: Herramientas y técnicas prácticas* (4ª ed.). Morgan Kaufmann.
- Zorpas A. A., & Lasaridi K. (2013). Measuring waste prevention: a challenge for policy makers.



## Apéndice

### 1.1

Script de SQL utilizado para la estructuración de la base de datos

```
WITH movimientos_clientes AS (  
    SELECT  
        m.asset_id,  
        a.type_id AS tipo_envase, -- Tipo de envase  
        m.updated_at AS fecha_asignacion,  
        LEAD(m.updated_at) OVER (PARTITION BY m.asset_id ORDER BY  
m.updated_at) AS fecha_siguiente_movimiento,  
        m.place,  
        LEAD(m.place) OVER (PARTITION BY m.asset_id ORDER BY m.updated_at) AS  
siguiente_lugar  
    FROM movements m  
    JOIN assets a ON m.asset_id = a.id -- Unimos para obtener el tipo de envase  
)  
),  
ciclos_clasificados AS (  
    SELECT  
        asset_id,  
        tipo_envase, -- Tipo de envase en cada ciclo  
        fecha_asignacion,  
        fecha_siguiente_movimiento AS fecha_devolucion,  
        CASE  
            WHEN siguiente_lugar = 'warehouse' THEN 1 -- Ciclo completo cliente-  
warehouse  
            WHEN siguiente_lugar = 'client' THEN 1 -- Ciclo cliente-cliente, tratado como  
completo  
            ELSE 0 -- Censurado, sin devolución o asignación adicional  
        END AS ciclo_completo,  
        CASE  
            WHEN siguiente_lugar = 'client' THEN 'cliente-cliente'  
            WHEN siguiente_lugar = 'warehouse' THEN 'cliente-warehouse'  
            ELSE 'censurado'  
        END AS tipo_ciclo,  
        CASE  
            WHEN siguiente_lugar = 'warehouse' THEN EXTRACT(DAY FROM  
(fecha_siguiente_movimiento - fecha_asignacion)) -- Días para ciclo completo  
            WHEN siguiente_lugar = 'client' THEN EXTRACT(DAY FROM  
(fecha_siguiente_movimiento - fecha_asignacion)) -- Ciclo cliente-cliente  
            WHEN siguiente_lugar IS NULL THEN EXTRACT(DAY FROM (NOW() -  
fecha_asignacion)) -- Censurado, sin devolución registrada  
        END AS dias_ciclo  
    FROM movimientos_clientes
```



```
WHERE place = 'client' -- Solo ciclos que inician con cliente
)
SELECT *
FROM ciclos_clasificados
WHERE ciclo_completo IS NOT NULL;
```

### **Codigos de Python**

#### **2.2**

```
import seaborn as sns
import matplotlib.pyplot as plt

# Configurar el gráfico
plt.figure(figsize=(12, 6))

# Histograma y curva de densidad para ciclos completos
sns.histplot(ciclos_completos, kde=True, color='blue', label='Ciclos completos',
bins=30, stat="density")
sns.kdeplot(ciclos_completos, color='blue', linewidth=2)

# Histograma y curva de densidad para ciclos censurados
sns.histplot(ciclos_censurados, kde=True, color='orange', label='Ciclos censurados',
bins=30, stat="density")
sns.kdeplot(ciclos_censurados, color='orange', linewidth=2)

# Personalización del gráfico
plt.title("Histogramas con curvas de densidad para ciclos completos y censurados")
plt.xlabel("Días del ciclo")
plt.ylabel("Densidad")
plt.legend()
plt.grid()

# Mostrar el gráfico
plt.show()
```

#### **2.3**

```
import scipy.stats as stats
import numpy as np

# Convertir los datos de ciclos completos a un array de numpy
datos_ciclos_completos = ciclos_completos.dropna().values

# Lista de distribuciones a probar
```



```
distribuciones = {
    "Weibull": stats.weibull_min,
    "Exponencial": stats.expon,
    "Lognormal": stats.lognorm
}

# Ajustar las distribuciones y guardar los resultados
resultados_ajuste = {}
for nombre, distribucion in distribuciones.items():
    # Ajustar la distribución a los datos
    parametros = distribucion.fit(datos_ciclos_completos)
    resultados_ajuste[nombre] = parametros

    # Generar datos teóricos basados en los parámetros ajustados
    datos_teoricos = distribucion.rvs(*parametros[:-2], loc=parametros[-2],
scale=parametros[-1], size=1000)

    # Graficar comparación
    plt.figure(figsize=(12, 6))
    sns.histplot(datos_ciclos_completos, kde=True, color='blue', label='Empírica',
bins=30, stat="density")
    sns.histplot(datos_teoricos, kde=True, color='red', label=f'Teórica: {nombre}',
bins=30, stat="density")
    plt.title(f"Ajuste de {nombre}")
    plt.xlabel("Días del ciclo")
    plt.ylabel("Densidad")
    plt.legend()
    plt.grid()
    plt.show()

# Evaluar el ajuste con Kolmogorov-Smirnov
print("Resultados del ajuste:")
for nombre, distribucion in distribuciones.items():
    parametros = resultados_ajuste[nombre]
    ks_stat, ks_p = stats.kstest(datos_ciclos_completos, lambda x: distribucion.cdf(x,
*parametros))
    print(f"{nombre}: KS-Statistic = {ks_stat:.4f}, p-value = {ks_p:.4f}")
```

## 2.4

```
import numpy as np
import scipy.stats as stats
import pandas as pd
```



```
# Cargar la base de datos
df = pd.read_excel('BASE DE DATOS DEFINITIVA.xlsx')

# Filtrar los datos por ciclos completos
df_ciclos_completos = df[df['ciclo_completo'] == 1]

# Asegurarse de que la columna de días esté correctamente seleccionada
dias_ciclo = df_ciclos_completos['dias_ciclo']

# Ajustar la distribución lognormal
shape, loc, scale = stats.lognorm.fit(dias_ciclo, floc=0)

# Calcular mu y sigma
mu = np.log(scale)
sigma = shape

print(f"Mu (media en log): {mu}")
print(f"Sigma (desviación estándar en log): {sigma}")

# Mediana
mediana = np.exp(mu)
print(f"Mediana: {mediana}")

# Percentil 90
percentil_90 = np.exp(mu + stats.norm.ppf(0.90) * sigma)
print(f"Percentil 90: {percentil_90}")
```

## 2.5

```
# Boxplot para comparar los días de ciclo por tipo de envase
plt.figure(figsize=(10, 6))
sns.boxplot(data=df_ciclos_completos, x='tipo_envase', y='dias_ciclo')
plt.title('Boxplot Comparativo de Días de Ciclo Completo por Tipo de Envase')
plt.xlabel('Tipo de Envase')
plt.ylabel('Días del Ciclo Completo')
plt.show()
```

## 2.6

```
# Ajustar el modelo de Kaplan-Meier a los datos
```





```
kmf.fit(durations=data['dias_ciclo'], event_observed=data['ciclo_completo'],  
label="Supervivencia Total")
```

```
# Graficar la curva de supervivencia  
kmf.plot_survival_function()  
plt.xlabel("Días")  
plt.ylabel("Probabilidad de Supervivencia")  
plt.title("Curva de Supervivencia Kaplan-Meier - Todos los Envases")  
plt.show()
```

## 2.7

```
# Obtener los tipos de envase únicos  
tipos_envase = data['tipo_envase'].unique()
```

```
# Graficar Kaplan-Meier para cada tipo de envase  
plt.figure(figsize=(10, 6))  
for tipo in tipos_envase:  
    kmf.fit(durations=data[data['tipo_envase'] == tipo]['dias_ciclo'],  
            event_observed=data[data['tipo_envase'] == tipo]['ciclo_completo'],  
            label=f"Tipo de Envase {tipo}")  
    kmf.plot_survival_function()
```

```
plt.xlabel("Días")  
plt.ylabel("Probabilidad de Supervivencia")  
plt.title("Curva de Supervivencia Kaplan-Meier por Tipo de Envase")  
plt.legend()  
plt.show()
```