



**UNIVERSIDAD MAYOR DE SAN SIMÓN**  
**FACULTAD DE CIENCIAS Y TECNOLOGÍA**

**DIRECCIÓN DE POSGRADO**



## **DIPLOMADO CIENCIA DE DATOS**

### **SEGUNDA VERSIÓN**

# **PREDICCIÓN DE RECOMPRA DE CLIENTES Y ESTIMACIÓN DEL VALOR DE VIDA (CLV) CON MODELOS BTYD: CASO ACERÍA GUADIX S.A.**

**PROYECTO PRESENTADO PARA OBTENER EL GRADO DE LICENCIATURA EN  
INGENIERÍA EN INFORMÁTICA  
MODALIDAD DOBLE TITULACIÓN**

**POSTULANTE : JOSE MANUEL VARGAS CRUZ**  
**TUTOR : MSc. Ing. Joel Alejandro Mejía Meléndez**

**Cochabamba – Bolivia**  
**2025**



# Predicción de Recompra de Clientes y Estimación del Valor de Vida (CLV) con Modelos BTYD: Caso Acería Guadix S.A.

Por

Jose Manuel Vargas Cruz

El presente documento, Trabajo de Grado es presentado a la Dirección de Posgrado de la Facultad de Ciencias y Tecnología en cumplimiento parcial de los requisitos para la obtención del grado académico de Licenciatura en Ingeniería Informática, modalidad Doble Titulación, habiendo cursado el Diplomado “Ciencia de Datos” propuesta por el Centro de Estadística Aplicada (CESA) en su segunda versión.

## TUTOR

MSc. Ing Joel Alejandro Mejía Meléndez

## COMITÉ DE EVALUACIÓN

Ing. M.Sc. Ronald Edgar Patiño Tito. (Presidente)  
Ing. M.Sc Guillen Salvador Roxana. (Coordinador)  
Ing. M.Sc Espinoza Orosco José (Tribunal)  
Ing. M.Sc. Yony Richard Montoya Burgos



**DIRECCIÓN DE POSGRADO, FACULTAD DE CIENCIAS Y TECNOLOGIA**  
Cochabamba, Bolivia

### **Aclaración**

**Este documento describe el trabajo realizado como parte del programa de estudios de Diplomado “Ciencia de Datos” en el Centro de Estadística Aplicada CESA y la Dirección de Posgrado de la Facultad de Ciencias y Tecnología. Todos los puntos de vista y opiniones expresadas en el mismo son responsabilidad exclusiva del autor y no representan necesariamente las de la institución.**

---

# Resumen

---

El presente proyecto aplica técnicas de ciencia de datos para abordar un problema clave en la gestión comercial de Acería Guadix S.A., empresa especializada en la fabricación y venta de repuestos industriales. Si bien se contaba con un historial de facturación, no se disponía de una herramienta que permitiera anticipar el comportamiento futuro de compra de los clientes. En este contexto, se desarrolló un modelo predictivo basado en los enfoques estadísticos BTYD (Buy 'Til You Die), específicamente Pareto/NBD, con el objetivo de estimar cuántas compras realizará un cliente y cuándo ocurrirá la próxima, optimizando así decisiones de ventas, retención y producción.

A partir de los registros históricos, se aplicaron procesos de limpieza, transformación y análisis, seguidos de la implementación del modelo Gamma-Gamma para estimar el valor económico de cada cliente (CLV). Los resultados revelan que el 15% de los clientes genera más del **60% del valor** proyectado, lo que permite identificar con claridad a los clientes prioritarios. Asimismo, se obtuvo una precisión adecuada en las predicciones, con un error cuadrático medio (**MSE**) de **0.693** en el periodo de validación. El modelo permitió segmentar a los clientes en función de su probabilidad de actividad futura y su volumen estimado de compras: se identificaron **10** clientes con alta probabilidad de recompra y más de **3** compras esperadas en los siguientes seis meses, y otros **20** con baja probabilidad de actividad, pero con CLV elevados, ideales para campañas de retención. Además, se estimaron fechas probables de la próxima compra, destacando oportunidades de contacto proactivo.

Como conclusión, se valida la aplicabilidad de modelos probabilísticos en la predicción de comportamiento del cliente en contextos industriales, demostrando su potencial para mejorar la planificación comercial. La metodología desarrollada puede ser escalada e integrada en los sistemas internos de la empresa, aportando una solución sostenible para la gestión de clientes basada en analítica avanzada.

Palabras clave: Ciencia de datos, Pareto/NBD, CLV, comportamiento del cliente, predicción de compras, BTYD, recurrencia, análisis predictivo.

*Dedicatoria*

*A mis padres,*

*por su amor, guía y sacrificio.*

*A mi bebé, por darme una nueva razón de superarme.*

*Este logro también les pertenece.*

# Agradecimientos

---

*Deseo expresar mi más sincero agradecimiento al **MSc. Ing. Joel Alejandro Mejía Meléndez**, tutor de este proyecto, por su guía constante, sus valiosas sugerencias y su acompañamiento técnico durante todo el desarrollo del trabajo.*

*Asimismo, extendiendo mi gratitud a todo el **plantel docente del Diplomado en Ciencia de Datos**, por compartir sus conocimientos, experiencias y por brindar una formación académica de alto nivel que ha sido fundamental para alcanzar los objetivos de este proyecto.*

*A todos quienes contribuyeron directa o indirectamente en este proceso de aprendizaje y ejecución, muchas gracias.*

# Tabla de contenidos

---

UNIVERSIDAD MAYOR DE SAN SIMÓN	1
1. Introducción	1
1.1. Antecedentes	2
1.2. Justificación	3
1.3. Planteamiento del problema	3
1.4. Objetivo general	4
1.4.1. Objetivos específicos	4
2. Marco teórico	6
2.1. Predicción con Modelos BTYD (Buy 'Til You Die)	6
2.1.1. Modelo Pareto/NBD	6
2.1.2. Modelo Gamma-Gamma para estimación del CLV	7
2.1.3. Calidad de Datos en Modelos Predictivos	9
2.2. El Customer Lifetime Value (CLV) como Métrica Estratégica	10
2.2.1. Importancia del CLV en la gestión Comercial	10
2.2.2. Segmentación de Clientes basada en CLV y Recurrencia	11
2.3. Aplicaciones Prácticas en Empresas Industriales	12
2.4. Integración de Modelos Predictivos en Estrategias Comerciales	14
2.5. Visualización y Comunicación de Resultados en Modelos Predictivos	14
2.6. Consideraciones Éticas y de Privacidad en el Uso de Datos de Clientes	15
2.7. Contexto Empresarial: Acería Guadix S.A. y la relevancia de la Analítica Predictiva	16
3. Marco metodológico	17
3.1. Área de estudio	17
3.2. Flujograma metodológico	19
3.3. Fuentes de Información	20
3.3.1. Datos de Facturación Histórica	20
3.3.2. Datos Complementarios Internos	21
3.4. Recolección, limpieza y transformación de los Datos	21
3.5. Generación de tabla resumen para modelación	24



3.5.1.	Exploración de variables predictoras	26
3.6.	Modelado probabilístico con Pareto/NBD	28
3.7.	Estimación del CLV con el modelo Gamma-Gamma	29
3.8.	Evaluación y Validación del modelo	30
3.9.	Estimación de recompras y próxima fecha de compra	32
3.10.	Generación de reportes estratégicos	34
4.	Análisis de Resultados y Discusión	37
4.1.	Resultados de la Recolección y depuración de datos históricos	37
4.2.	Resultados de Generar una tabla resumen con métricas base	37
4.3.	Resultados de implementar Modelos BTYD	38
4.3.1.	Análisis de Segmentación: Actividad vs Recompra	40
4.3.2.	Top 10 Clientes con Mayor Potencial de Compra	41
4.3.3.	Clientes en Riesgo de Abandono Comercial	42
4.3.4.	Estimación del Customer Lifetime Value (CLV)	43
4.3.5.	Análisis del CLV estimado	43
4.3.6.	Clientes con mayor CLV (Tabla 3)	44
4.4.	Evaluación y validación del modelo	45
4.5.	Resultados de estimación de compras futuras y próxima compra	45
4.6.	Resultados de Elaborar reportes analíticos con visualizaciones y segmentaciones gráficas	47
4.7.	Discusión de Resultados	48
4.7.1.	Comparación con casos similares	48
5.	Conclusiones	52
6.	Recomendaciones	54
	Bibliografía	55
	Anexos	58

# Lista de figuras

---

Figura 1: Mapa sede central Guadix S.A.	17
Figura 2: Flujograma metodológico	19
Figura 3: Distribución de montos por transacción (boxplot)	22
Figura 4: Evolución mensual del monto total vendido	23
Figura 5: Top 10 clientes por monto acumulado	24
Figura 6: Generación de tabla resumen por cliente	25
Figura 7: Muestra de datos (5 filas)	25
Figura 8: Distribución de frecuencia de compra	27
Figura 9: Distribución de Recencia	27
Figura 10: Distribución de Antigüedad	28
Figura 11: Filtrado de clientes	28
Figura 12: Entrenamiento del modelo Pareto/NBD	29
Figura 13: Código para el ajuste del modelo Gamma-Gamma y la estimación del CLV individual por cliente	30
Figura 14: Evaluación con MSE	30
Figura 15: Matriz de recompras esperadas	31
Figura 16: Matriz de probabilidad de estar activo	32
Figura 17: Estimación del número de compras esperadas en los próximos 180 días por cliente	32
Figura 18: Probabilidad de que el cliente continúe activo en los próximos 180 días	33
Figura 19: Tabla resumen con los valores estimados por cliente: frecuencia, recencia, antigüedad, compras esperadas y probabilidad de seguir activo.	33
Figura 20: Código utilizado para estimar la fecha de la próxima compra, mediante el cociente entre el horizonte de predicción y las compras esperadas, sumado a la fecha de última compra	33
Figura 21: Clientes con próximas fechas estimadas de recompra, ordenados ascendentemente.	34
Figura 22: Tabla resumen con métricas predictivas por cliente	34
Figura 23: Top 5 de clientes con mayor valor de vida proyectado	35
Figura 24: Clientes con mayor número de recompras esperadas en los próximos 180 días	35
Figura 25: Segmentación de clientes basada en recompras estimadas y probabilidad de estar activos	36
Figura 26: Histograma de distribución del número de recompras estimadas para un horizonte de 180 días.	36
Figura 27: Tabla resumen con métricas base	38
Figura 28: Resultados modelo Pareto/NBD y Gamma-Gamma	39
Figura 29: Distribución de compras esperadas	39
Figura 30: Segmentación de clientes (Recompra vs Actividad)	40
Figura 31: Distribución del Customer Lifetime Value (CLV)	44
Figura 32: Tabla resumen con compras esperadas, actividad y CLV	46
Figura 33: Próxima fecha de compra	46
Figura 34: Clientes con mayor CLV	47

Figura 35: Clientes con mayor número de recompras esperadas	47
Figura 36: Distribución acumulada de CLV en estudio comparativo	49

## Lista de tablas

---

Tabla 1: Muestra de datos	21
Tabla 2: Top 10 clientes con más recompras esperadas	41
Tabla 3: Top 5 clientes según CLV estimado	44

# 1. Introducción

En un entorno empresarial cada vez más competitivo y orientado a datos, comprender el comportamiento del cliente y anticipar sus decisiones de compra es fundamental para la sostenibilidad de las organizaciones. Esta necesidad se vuelve aún más crítica en empresas industriales como Acería Guadix S.A., que atienden a una base de clientes recurrentes con productos de alta especificidad. La capacidad de anticipar cuándo un cliente volverá a comprar permite no solo una mayor eficiencia comercial, sino también una planificación más precisa de inventarios, producción y operaciones logísticas (Linoff & Berry, 2011).

El presente proyecto surge como respuesta a una pregunta estratégica que no puede resolverse eficazmente con métodos tradicionales: ¿cuándo y con qué probabilidad volverá un cliente a realizar una compra? Aunque Acería Guadix S.A. cuenta con registros históricos de facturación, estos no habían sido explotados sistemáticamente con fines analíticos o predictivos. A través del enfoque de la ciencia de datos, es posible transformar datos transaccionales pasivos en conocimiento útil mediante el uso de técnicas como el análisis RFM (Recency, Frequency, Monetary) y los modelos probabilísticos BTYD (Buy ‘Til You Die), herramientas ampliamente validadas en el ámbito académico y empresarial (Fader, Hardie, & Lee, 2005).

La elección de esta temática responde tanto a su aplicabilidad inmediata como al alto impacto que puede tener en diversas áreas del negocio. Aplicar modelos como el Pareto/NBD y Gamma-Gamma permite estimar métricas fundamentales como el número de compras esperadas en un horizonte temporal definido y el Customer Lifetime Value (CLV), que representa el valor económico proyectado de un cliente durante toda su relación con la empresa. Este tipo de análisis facilita la priorización de campañas comerciales, la gestión segmentada de la cartera de clientes y la optimización de los recursos de marketing (Gupta, Lehmann, & Stuart, 2006).

Además, la implementación de este proyecto demuestra cómo los datos históricos pueden convertirse en activos estratégicos al integrarse con metodologías de modelado estadístico. La información generada no solo permite identificar clientes de alto valor o en riesgo de abandono, sino también formular estrategias de retención más inteligentes basadas en evidencia cuantitativa. Así, el uso de modelos BTYD no solo es útil desde una perspectiva operativa, sino que también impulsa una transformación cultural hacia una gestión comercial basada en datos (Schmarzo, 2013).

En síntesis, este trabajo tiene como propósito no solo resolver un problema específico de predicción de comportamiento de clientes para Acería Guadix S.A., sino también servir como un ejemplo práctico del valor que la analítica avanzada puede aportar en la toma de decisiones estratégicas. Al integrar modelos de recurrencia, métricas de CLV y segmentación basada en probabilidad, se establece una base sólida para el desarrollo de soluciones predictivas escalables y sostenibles en el tiempo (Provost & Fawcett, 2013).

## 1.1. Antecedentes

En el entorno empresarial actual, la capacidad de anticipar el comportamiento de los clientes se ha convertido en una herramienta esencial para optimizar procesos de ventas, marketing y planificación de inventario (Linoff & Berry, 2011). Particularmente en industrias con ventas recurrentes, como la fabricación y distribución de repuestos industriales, predecir cuándo un cliente realizará su próxima compra puede representar una ventaja competitiva significativa. Esta necesidad ha impulsado el desarrollo de modelos estadísticos y predictivos centrados en el análisis de la recurrencia y el valor del cliente, entre los cuales destacan los modelos conocidos como BTYD (Buy "Til You Die) (Schmarzo, 2013).

Uno de los principales aportes teóricos en esta área es el modelo Pareto/NBD, (Fader, Hardie, & Lee, 2005), el cual permite estimar la probabilidad de que un cliente siga activo y cuántas compras adicionales realizará en un horizonte de tiempo determinado. Este modelo se basa en variables como la frecuencia de compra, la recencia (tiempo desde la última compra) y la antigüedad del cliente en el sistema. Su aplicación ha demostrado ser efectiva en sectores como el comercio electrónico (Hardie & Fader, 2006) servicios financieros (Schmittlein, Morrison, & Colombo, 1987) y retail (Reinartz & Kumar, 2003), siendo una herramienta clave para la segmentación y retención de clientes.

En complemento, el modelo Gamma-Gamma permite estimar el Customer Lifetime Value (CLV), es decir, el valor económico esperado de cada cliente en función del promedio monetario de sus transacciones. Este enfoque (Fader, Hardie, & Lee, 2005), ha sido adoptado por empresas como Amazon, Netflix y otras plataformas de venta recurrente, que buscan maximizar el valor de cada cliente a lo largo del tiempo (Gupta & Lehmann, Customers as assets, 2003).

En el ámbito latinoamericano, diversos estudios han explorado la aplicación de estos modelos en contextos empresariales reales. En Colombia se analizó la retención de clientes en una empresa farmacéutica utilizando BTYD (Rodríguez, 2019), mientras que en México aplicaron modelos similares en una cadena de ferreterías, logrando una mejora del 18% en la eficiencia de campañas de marketing (López & Cabrera, 2020).

En Bolivia, los estudios relacionados con predicción de recurrencia de compra son aún escasos, aunque algunos trabajos recientes han comenzado a aplicar enfoques de análisis predictivo en sectores como el comercio minorista y la industria de autopartes (Arce, 2021) (Rojas, 2022). Asimismo, informes institucionales y tesis académicas han resaltado la necesidad de incorporar herramientas de analítica avanzada para mejorar la toma de decisiones comerciales.

En la Universidad Privada Boliviana, (Rojas, 2022) se desarrolló un modelo híbrido de predicción de demanda que combinaba métodos clásicos como ARIMA con enfoques probabilísticos, evidenciando la utilidad de integrar diferentes técnicas en contextos con limitada cantidad de datos. La revisión de literatura también resalta que la calidad de los datos y la granularidad de los registros de venta son factores críticos para el éxito de estos modelos. Según algunos estudios (Olivares & Torres, 2020), la falta de normalización en los identificadores de cliente o la inconsistencia en las fechas de compra pueden afectar significativamente la confiabilidad de los resultados.

En resumen, los antecedentes muestran una evolución sólida en el desarrollo y aplicación de modelos BTYD para entender el comportamiento de compra de los clientes. Sin embargo, en el contexto boliviano

aún existe un vacío importante en cuanto a su implementación práctica, lo cual justifica plenamente el desarrollo del presente proyecto en Acería Guadix S.A., como una contribución académica con alto potencial de impacto empresarial.

## 1.2. Justificación

La toma de decisiones basada en datos es hoy en día un elemento indispensable para las empresas que buscan mantener su competitividad en mercados dinámicos. En particular, las organizaciones que trabajan bajo esquemas de venta recurrente, como **Acería Guadix S.A.**, enfrentan el desafío constante de anticipar las necesidades de sus clientes y optimizar sus recursos comerciales, logísticos y operativos. No basta con conocer lo que ya se ha vendido; es fundamental poder prever **cuándo** un cliente volverá a comprar y **cuánto valor** representa para la empresa a lo largo del tiempo.

En este sentido, el presente proyecto se justifica por su capacidad de proporcionar a la empresa una herramienta analítica que permita **predecir el comportamiento de compra de sus clientes**, utilizando metodologías reconocidas de ciencia de datos como los modelos **BTYD (Buy 'Til You Die)**. A través de estos modelos, se pueden obtener estimaciones confiables sobre la probabilidad de recompra, el número esperado de compras en un horizonte determinado, e incluso calcular el **valor de vida del cliente (CLV)**, un indicador clave para la planificación estratégica y la priorización de esfuerzos comerciales.

Además, la elección de este proyecto responde a una necesidad real y concreta dentro de Acería Guadix S.A., que actualmente cuenta con un valioso histórico de ventas, pero carece de un sistema que transforme estos datos en conocimiento predictivo accionable. La implementación de un modelo de predicción de recurrencia de compra permitirá a la empresa **segmentar su cartera de clientes**, diseñar **estrategias de retención más efectivas** y tomar decisiones fundamentadas para la asignación de recursos, como producción, inventario y atención personalizada.

Desde el punto de vista académico, el proyecto representa una aplicación práctica e integral de conceptos aprendidos en el diplomado de Ciencia de Datos, tales como análisis exploratorio, modelado probabilístico, segmentación de clientes y visualización de resultados. También aporta valor al ecosistema local de analítica de datos, al demostrar que es posible implementar soluciones avanzadas incluso en entornos empresariales bolivianos con estructuras de datos simples.

Finalmente, el desarrollo de este trabajo contribuye a cerrar la brecha entre el conocimiento técnico y su aplicación práctica en contextos reales, y sienta las bases para que acería Guadix S.A. continúe adoptando herramientas de inteligencia comercial que la posicionen como una empresa moderna, orientada al cliente y guiada por datos.

## 1.3. Planteamiento del problema

Acería Guadix S.A., empresa dedicada a la fabricación y comercialización de repuestos industriales, mantiene relaciones comerciales con una cartera de clientes que realizan compras recurrentes en función de sus necesidades operativas. A lo largo del tiempo, la empresa ha acumulado un considerable volumen de datos transaccionales que reflejan el comportamiento histórico de compra de sus clientes. Sin embargo, pese a contar con estos registros, **no se dispone actualmente de un sistema que permita**

**anticipar cuándo un cliente volverá a realizar una compra ni cuántas compras adicionales podría generar en el futuro.**

Esta limitación representa un problema estratégico importante, ya que impide una planificación eficiente en áreas clave como ventas, producción e inventario. Sin la capacidad de anticipar la demanda individual de los clientes, las acciones comerciales tienden a ser reactivas, y se desaprovechan oportunidades de fidelización, reactivación o priorización de aquellos clientes que representan mayor valor para la empresa. Asimismo, la falta de segmentación predictiva dificulta la ejecución de campañas comerciales focalizadas, lo que puede derivar en esfuerzos dispersos y de bajo impacto.

A nivel operativo, este problema se traduce en desafíos concretos como:

- Desconocimiento del momento óptimo para contactar a un cliente.
- Dificultad para estimar ingresos futuros provenientes de clientes actuales.
- Riesgo de pérdida de clientes valiosos sin que la empresa tenga indicios de alerta.
- Imposibilidad de proyectar la rentabilidad futura de cada cliente (CLV) con base en su historial de compra.

Además, en un contexto de creciente digitalización e inteligencia comercial, otras empresas del sector están comenzando a adoptar soluciones de analítica predictiva para optimizar sus decisiones. Acería Guadix S.A., al no contar con una herramienta de este tipo, corre el riesgo de quedar rezagada frente a competidores que utilicen estrategias más proactivas y orientadas por datos.

Por tanto, se plantea como problema principal **la ausencia de un sistema predictivo que permita anticipar la recurrencia de compra y el valor futuro de los clientes**, limitando la capacidad de la empresa para planificar acciones estratégicas de manera eficiente. Resolver este problema requiere el desarrollo de un modelo analítico capaz de identificar patrones de comportamiento en los datos históricos y proyectar el futuro comportamiento de cada cliente, con base en técnicas de ciencia de datos probadas y adaptadas al contexto de la empresa.

## 1.4. Objetivo general

**Desarrollar** un modelo predictivo basado en técnicas de ciencia de datos que permita estimar la recurrencia de compra y el valor futuro de los clientes de Acería Guadix S.A., con el fin de optimizar la toma de decisiones comerciales, mejorar la planificación de recursos y fortalecer las estrategias de fidelización de clientes.

### 1.4.1. Objetivos específicos

- **Recolectar, limpiar y transformar** los datos históricos de facturación de Acería Guadix S.A., asegurando su calidad y estructura para la modelación del comportamiento de compra de los clientes.
- **Generar una tabla resumen** con métricas de frecuencia, recencia, antigüedad, valor promedio por cliente y otras variables clave, como base para el análisis predictivo.



- **Implementar modelos probabilísticos** de la familia BTYD, específicamente Pareto/NBD y Gamma-Gamma, para estimar la recurrencia de compra y el Customer Lifetime Value (CLV) de cada cliente.
- **Evaluar el desempeño** de los modelos aplicados mediante visualizaciones y métricas estadísticas (como MSE), con el fin de validar su precisión predictiva y aplicabilidad práctica.
- **Estimar cantidad de compras futuras** por cliente a partir de los resultados del modelo Pareto/NBD, y consolidarlas en una tabla resumen para su posterior uso en acciones comerciales.
- **Elaborar reportes analíticos con visualizaciones y segmentaciones gráficas**, que permitan identificar clientes activos, en riesgo de pérdida y con alto valor económico proyectado (CLV), facilitando la toma de decisiones estratégicas.

## 2. Marco teórico

### 2.1. Predicción con Modelos BTYD (Buy 'Til You Die)

Los modelos **BTYD (Buy 'Til You Die)** constituyen una familia de modelos probabilísticos desarrollados para estudiar el comportamiento de compra de los clientes en escenarios donde las transacciones son discretas y recurrentes. Estos modelos parten del supuesto de que los clientes realizan compras mientras están activos (vivos) y, en algún momento desconocido, dejan de hacerlo. Una de sus principales ventajas es que permiten **inferir la probabilidad de que un cliente siga activo sin requerir una etiqueta explícita de abandono**, lo cual los hace especialmente valiosos en contextos donde esta información no está disponible (Fader, Hardie, & Lee, RFM and CLV: Using Iso-value Curves for Customer Base Analysis, 2005).

#### 2.1.1. Modelo Pareto/NBD

El modelo **Pareto/NBD (Negative Binomial Distribution)** es uno de los representantes más robustos de la familia BTYD. Este modelo se basa en dos componentes fundamentales:

1. La **tasa de compra** de cada cliente mientras está activo, denotada como  $\lambda$ , sigue una **distribución Gamma** con parámetros  $r$  y  $\alpha$ , es decir:

$$\lambda \sim \text{Gamma} (r, \alpha)$$

2. El **tiempo hasta el abandono** o “muerte” del cliente sigue una **distribución Exponencial** con parámetro  $\mu$ , y entre clientes se modela con una distribución Gamma con parámetros  $s$  y  $\beta$ :

$$\mu \sim \text{Gamma} (s \text{ y } \beta)$$

El número de compras de un cliente activo sigue una **distribución de Poisson**, y al asumir heterogeneidad entre clientes, se obtiene una distribución **Negativo Binomial** para el número total de compras. Por ello, el modelo recibe el nombre de "Pareto/NBD" (Fader, Hardie, & Lee, RFM and CLV: Using Iso-value Curves for Customer Base Analysis, 2005). Esta formulación permite calcular:

- La **esperanza del número de compras futuras** hasta un tiempo  $t$
- La **probabilidad de que el cliente esté aún activo** dado su historial

Matemáticamente, si  $x$  es la cantidad de compras realizadas,  $t_x$  el tiempo desde la primera compra hasta la última, y  $T$  la edad del cliente, se puede calcular la probabilidad de actividad con:

$$P(\text{alive} | x, t_x, T) = \frac{s}{\beta + T} \left(1 + \frac{\alpha}{\alpha + T}\right)^{-r}$$

Esta probabilidad se usa para estimar qué clientes siguen activos y su posible número de compras futuras (Hardie & Fader, 2006).

Una de las grandes fortalezas del modelo Pareto/NBD es su **capacidad para generar predicciones individualizadas**. Esto permite estimar, para cada cliente, cuántas compras adicionales realizará en un horizonte temporal definido, como 180 días, y con qué probabilidad se espera que aún esté activo. Esta granularidad es clave para una segmentación efectiva y para la **asignación eficiente de recursos de marketing** (Reinartz & Kumar, 2003).

En cuanto a su aplicación, el modelo ha sido validado en múltiples industrias: **e-commerce, banca, telecomunicaciones, y servicios de suscripción**. En estos contextos, se ha demostrado que BTYD ofrece mayor precisión en la estimación del valor futuro del cliente que los métodos tradicionales basados en medias o reglas heurísticas (Zhang, Nian, & Yu, 2021).

Un aspecto notable de BTYD es que **requiere pocos datos para su implementación**: solo se necesita el identificador del cliente, las fechas de transacción y, opcionalmente, el valor monetario de las compras. Esto lo hace ideal para empresas que poseen históricos de facturación, pero no cuentan con perfiles detallados de sus clientes (Fader & Hardie, Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity, 2013).

No obstante, es importante considerar algunas **limitaciones**. El modelo **asume independencia entre compras** y entre la frecuencia de compra y el valor de las transacciones, lo cual no siempre se cumple en la realidad. También se basa en el principio de que el comportamiento pasado es un buen predictor del futuro, una hipótesis que puede fallar ante eventos exógenos (como crisis sanitarias, cambios de estrategia o campañas masivas) (Schmittlein, Morrison, & Colombo, 1987).

En respuesta a estas limitaciones, han surgido variantes como el **BG/NBD (Beta-Geometric/NBD)**, que reemplaza la función de desactivación exponencial por una distribución geométrica, o los modelos de actualización bayesiana, que **ajustan las predicciones en tiempo real** conforme se reciben nuevos datos (Zhang, Nian, & Yu, 2021).

En el caso de **Acería Guadix S.A.**, el uso del modelo Pareto/NBD es altamente pertinente. Sus clientes realizan compras periódicas de repuestos industriales, y aunque no hay una notificación explícita de abandono, los históricos de ventas ofrecen una base sólida para inferir patrones de comportamiento. Al aplicar este modelo, la empresa puede anticipar compras futuras, **planificar con mayor precisión su producción, optimizar la gestión comercial, y reducir la pérdida de clientes de alto valor**.

### 2.1.2. Modelo Gamma-Gamma para estimación del CLV

El modelo **Gamma-Gamma** es una extensión natural de los enfoques BTYD, orientado específicamente a la **estimación del Customer Lifetime Value (CLV)** cuando se dispone de datos monetarios por transacción. Mientras el modelo **Pareto/NBD** se enfoca en predecir la frecuencia futura de compra y la probabilidad de que un cliente siga activo, el Gamma-Gamma estima el valor monetario promedio de las transacciones de un cliente, bajo el supuesto de que este valor es constante a lo largo del tiempo, pero puede variar entre clientes (Fader, Hardie, & Lee, RFM and CLV: Using Iso-value Curves for Customer Base Analysis, 2005). Esta estimación es fundamental para proyectar los ingresos futuros que un cliente

puede generar, lo que convierte al CLV en una herramienta clave de segmentación y planeación estratégica.

Este modelo parte de **dos supuestos estadísticos clave**. El primero establece que el valor promedio de las transacciones de cada cliente sigue una **distribución gamma**, lo que permite modelar la heterogeneidad en el gasto entre los diferentes individuos. El segundo supuesto es que el valor monetario promedio de un cliente es **independiente de su frecuencia de compra**, es decir, no existe correlación entre cuánto gasta un cliente y cuántas veces compra. Aunque este supuesto podría no cumplirse estrictamente en todos los sectores, estudios empíricos han demostrado que es razonablemente válido en entornos comerciales con datos transaccionales (Fader et al., 2005; Gupta et al., 2006).

Desde una perspectiva matemática, el modelo Gamma-Gamma estima tres parámetros:

- $p$ : parámetro de forma de la distribución gamma,
- $q$ : parámetro de escala,
- $v$ : parámetro que captura la varianza entre clientes.

Una vez estimados, se calcula el valor monetario esperado por cliente ( $E[M]$ ) y se proyecta su CLV utilizando la fórmula:

$$CLV = E[M] \times E[Compras]$$

donde  $E[M]$  proviene del Gamma-Gamma y  $E[Compras]$  del modelo Pareto/NBD. Esta **combinación sinérgica** permite no solo predecir el comportamiento del cliente, sino también cuantificar su impacto financiero futuro (Hardie & Fader, 2006).

El **Customer Lifetime Value** es una métrica de gran utilidad en la gestión empresarial, ya que permite priorizar los recursos hacia los clientes más rentables. Entre sus aplicaciones más comunes se incluyen: la determinación del presupuesto óptimo para adquisición o retención de clientes, el diseño de **campañas personalizadas**, la evaluación del impacto financiero de estrategias comerciales, y la **estimación de ingresos a largo plazo** de la base de clientes (Gupta & Lehmann, Customers as assets, 2003).

Diversas empresas líderes han adoptado el CLV como indicador central. Por ejemplo, **Amazon** y **Netflix** utilizan métricas de CLV para segmentar audiencias, optimizar promociones y evaluar estrategias de fidelización (Schmarzo, Big Data: Understanding How Data Powers Big Business, 2013). En el sector financiero, el CLV se utiliza para personalizar tasas de interés o límites de crédito, mientras que, en industrias como la manufactura o los servicios industriales, sirve para priorizar a los clientes en función de su valor económico proyectado (Kumar & Reinartz, 2016).

En el caso de **Acería Guadix S.A.**, la estimación del CLV representa una oportunidad estratégica para tomar decisiones comerciales fundamentadas. Identificar no solo a los clientes más frecuentes, sino también a aquellos que generan mayores ingresos en el largo plazo, permite **dirigir esfuerzos de retención y fidelización de forma más efectiva**. Además, incorporar el CLV en tableros de control y reportes interdepartamentales puede mejorar la coordinación entre las áreas técnicas, comerciales y logísticas, alineando las decisiones tácticas con objetivos estratégicos (Fader & Hardie, Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity, 2013).

Es importante, sin embargo, considerar los requisitos técnicos para aplicar correctamente el modelo. En particular, cada cliente debe tener **al menos dos transacciones con valores monetarios válidos**, y se deben filtrar los **valores atípicos o inconsistentes** que puedan distorsionar las estimaciones. Una vez integrados el Gamma-Gamma y el Pareto/NBD, se obtiene una vista completa y robusta del valor futuro de cada cliente, lo que convierte esta metodología en un **activo predictivo de alto valor comercial** (Zhang, Nian, & Yu, 2021).

### 2.1.3. Calidad de Datos en Modelos Predictivos

La calidad de los datos es un factor determinante en el éxito de cualquier proyecto de ciencia de datos, y cobra especial relevancia en la construcción de modelos predictivos como los **BTYD (Pareto/NBD)** y **Gamma-Gamma**. Estos modelos trabajan directamente sobre información transaccional, como fechas de compra, frecuencia de transacciones y montos asociados, lo que implica que cualquier error, omisión o inconsistencia en estas variables puede comprometer significativamente la precisión de las predicciones (Batini, Cappiello, Francalanci, & Maurino, 2009).

Existen múltiples dimensiones de la calidad de datos que deben ser consideradas: **completitud, consistencia, validez, actualización, unicidad y precisión**. La completitud se refiere a que todos los campos requeridos (por ejemplo, identificador del cliente y fecha de compra) estén presentes. La consistencia implica que los registros no presenten contradicciones internas, como fechas futuras o montos negativos sin justificación. La unicidad exige que no existan duplicados en los registros, especialmente en bases que se actualizan de forma periódica. La precisión, por otro lado, se refiere a que los datos reflejen con fidelidad las transacciones reales ocurridas (Rahm & Do, 2000).

En el caso de modelos como Pareto/NBD, se requiere un mínimo de tres columnas correctamente estructuradas: un identificador único para cada cliente, una columna con las fechas de cada transacción, y una fecha de corte o final del período de observación. Si se desea calcular el valor monetario (CLV), se necesita además una columna que contenga el monto asociado a cada compra. Cualquier error en estas columnas —como fechas mal formateadas, identificadores duplicados o montos inconsistentes— puede impedir el entrenamiento del modelo o arrojar resultados distorsionados (Fader & Hardie, A Note on Implementing the Pareto/NBD Model and Related Models in Excel, 2005).

Además de los errores técnicos, existen otros desafíos asociados al contexto del negocio. Por ejemplo, en empresas que permiten cotizaciones o reservas de pedidos sin que se concrete una venta, es fundamental filtrar las transacciones que realmente implicaron una compra. Del mismo modo, si existen clientes registrados con múltiples identificadores (por cambios en el sistema o errores de digitación), es necesario aplicar procesos de deduplicación y normalización (Batini, Cappiello, Francalanci, & Maurino, 2009).

La **limpieza de datos (data cleaning)** es una etapa crítica que incluye actividades como:

- Eliminar duplicados exactos o parciales.
- Validar y transformar formatos de fecha.
- Detectar y corregir valores extremos o anómalos.
- Homogeneizar nombres de clientes y productos.
- Completar campos faltantes con imputación razonable, si es necesario.

Estas actividades permiten estructurar los datos de forma que sean aptos para modelado y análisis, lo cual ha sido demostrado como esencial en numerosos proyectos de minería de datos (Redman, 1998).

La preparación adecuada de los datos no solo mejora el rendimiento de los modelos, sino que también facilita su interpretación y la generación de reportes confiables. Incluso en empresas donde los sistemas de información son básicos, es posible lograr una calidad de datos aceptable con procedimientos simples, como la estandarización de formatos y el uso de validaciones automáticas en los formularios de ingreso de datos (Olivares & Torres, 2020).

Por lo tanto, garantizar la calidad de los datos no debe considerarse una tarea secundaria o previa al análisis, sino un componente integral de todo el proceso de ciencia de datos. En el caso de **Acería Guadix S.A.**, una correcta depuración de los registros de facturación es indispensable para lograr predicciones útiles, confiables y accionables. Esta depuración representa no solo una mejora técnica, sino una inversión estratégica en la base de conocimiento de la empresa (Batini, Cappiello, Francalanci, & Maurino, 2009) (Olivares & Torres, 2020).

## **2.2. El Customer Lifetime Value (CLV) como Métrica Estratégica**

El **Customer Lifetime Value (CLV)**, o Valor del Ciclo de Vida del Cliente, es una métrica estratégica que permite cuantificar el valor económico total que un cliente aportará a una empresa durante toda su relación comercial. Esta medida ha cobrado gran relevancia en contextos donde la fidelización y la personalización han reemplazado al enfoque puramente transaccional. Estimar correctamente el CLV permite a las empresas optimizar la asignación de recursos, maximizar la rentabilidad de su base de clientes y diseñar estrategias de adquisición, retención y crecimiento más efectivas (Kotler & Keller, 2012).

### **2.2.1. Importancia del CLV en la gestión Comercial**

Uno de los principales aportes del CLV es que trasciende la visión de valor inmediato de una transacción aislada y promueve una perspectiva de largo plazo. En lugar de centrarse únicamente en cuánto compró un cliente en la última visita, el CLV permite responder preguntas clave como: ¿cuánto puede llegar a comprar este cliente en los próximos meses o años?, ¿cuánto cuesta mantenerlo activo?, ¿cuándo deja de ser rentable? Esta visión resulta especialmente relevante en industrias de productos duraderos o especializados, como Acería Guadix S.A., donde las relaciones comerciales se construyen a lo largo del tiempo (Schmarzo, Big Data: Understanding How Data Powers Big Business, 2013).

Desde el punto de vista operativo, el CLV facilita la segmentación de clientes según su rentabilidad proyectada. Es posible clasificar a los clientes en tres grupos principales: alto CLV (alta prioridad de atención), CLV medio (potencial de crecimiento) y bajo CLV (riesgo de pérdida o bajo retorno). Esta clasificación permite implementar campañas dirigidas y estrategias de fidelización más eficientes (Reinartz & Kumar, 2003) (Gupta & Lehmann, Customers as assets, 2003).

Además, el CLV es útil como indicador para la toma de decisiones en diversas áreas organizacionales. En **marketing**, permite evaluar el retorno de inversión (ROI) de campañas de adquisición; en **ventas**, ayuda

a priorizar cuentas clave y definir incentivos; y en **finanzas**, permite estimar ingresos futuros con mayor precisión y analizar la rentabilidad de segmentos específicos (Berger & Nasr, 1998).

Otra de sus aplicaciones destacadas es la estimación del **customer equity**, que representa la suma del CLV de todos los clientes actuales. Esta métrica puede utilizarse incluso para valorar financieramente a la empresa o evaluar oportunidades de inversión basadas en el valor futuro de su cartera de clientes (Kotler & Keller, 2012).

Desde el punto de vista tecnológico, el CLV puede integrarse en tableros ejecutivos (dashboards), sistemas de inteligencia de negocios (BI) y motores de recomendación, mejorando la capacidad de la organización para tomar decisiones informadas y en tiempo real. En combinación con indicadores como satisfacción, engagement o historial de servicio, se genera un perfil más completo del cliente (Rust, Zeithaml, & Lemon, 2004).

En el caso de empresas industriales como **Acería Guadix S.A.**, donde las compras son menos frecuentes, pero de mayor valor, el CLV se vuelve especialmente relevante. Ayuda a decidir qué clientes merecen seguimiento personalizado, qué contratos deben renovarse y qué relaciones podrían fortalecerse en función de su rentabilidad proyectada (Reinartz & Kumar, 2003; Fader & Hardie, 2005).

No obstante, aplicar CLV con precisión presenta desafíos. Su estimación depende en gran medida de la calidad de los datos históricos, la regularidad de las compras, y la correcta elección del modelo matemático. Asimismo, deben considerarse factores externos que pueden alterar el comportamiento del cliente, como cambios regulatorios, coyuntura económica o transformaciones tecnológicas (Gupta & Lehmann, Customers as assets, 2003).

En conclusión, el CLV es una herramienta fundamental para transformar datos en decisiones estratégicas. Su implementación, como se propone en este proyecto, permite evolucionar de una gestión comercial reactiva a una predictiva, enfocada en maximizar el valor a largo plazo de cada cliente, alineando esfuerzos comerciales con objetivos sostenibles y medibles (Kumar & Reinartz, 2016).

### 2.2.2. Segmentación de Clientes basada en CLV y Recurrencia

La segmentación de clientes basada en **Customer Lifetime Value (CLV)** y **recurrencia de compra** es una técnica avanzada que permite clasificar a los consumidores no solo según su historial, sino también con base en su valor proyectado y su comportamiento futuro estimado. Esta metodología supera los enfoques tradicionales de segmentación demográfica o geográfica, al integrar modelos predictivos que estiman cuántas veces comprará un cliente y cuánto valor económico generará para la empresa a lo largo del tiempo (Wedel & Kamakura, 2012).

El uso de modelos BTYD como **Pareto/NBD** y **Gamma-Gamma** permite generar esta segmentación a partir de variables transaccionales simples, como fechas de compra y montos. Estos modelos calculan, entre otros, la probabilidad de que el cliente continúe activo, el número esperado de compras en un horizonte determinado, y el CLV estimado, lo cual habilita una clasificación más estratégica y alineada con los objetivos de rentabilidad empresarial (Fader, Hardie, & Lee, "Counting your customers" the easy way: An alternative to the Pareto/NBD model, 2005).

La segmentación basada en estos modelos permite identificar al menos cuatro grupos clave: (1) clientes de alto CLV y alta recurrencia, que deben ser fidelizados activamente; (2) clientes de alto CLV, pero baja probabilidad de recompra, que requieren estrategias de reactivación; (3) clientes frecuentes de bajo CLV, que podrían representar ingresos constantes; y (4) clientes de bajo CLV y baja frecuencia, que podrían no justificar una atención prioritaria. Esta clasificación ha demostrado mejorar la eficiencia de campañas de marketing en más de un 20% en estudios aplicados en sectores como retail, telecomunicaciones y banca (Kumar & Reinartz, 2016).

En contextos industriales como el de **Acería Guadix S.A.**, donde las compras no son masivas, pero sí recurrentes y de alto valor, esta segmentación permite focalizar los esfuerzos del equipo comercial. Por ejemplo, se puede priorizar el contacto con clientes de alto valor proyectado cuya próxima compra se estima en los próximos meses, o diseñar acciones de retención específicas para cuentas en riesgo de abandono, pero con alto CLV (Gupta & Lehmann, Customers as assets, 2003).

Además, esta segmentación es **dinámica**: se puede actualizar periódicamente conforme se registran nuevas transacciones, adaptando las estrategias a cambios reales en el comportamiento del cliente. Esto permite tomar decisiones oportunas y personalizadas, optimizando los recursos del área comercial. Se resalta que el valor predictivo de estos modelos es mayor cuando se utilizan para actualizar la clasificación del cliente de forma continua (Fader & Hardie, 2009).

La incorporación de visualizaciones, como gráficos de dispersión entre CLV y probabilidad de recompra, o dashboards que agrupan clientes por segmento, facilita la interpretación de los datos y la acción rápida del equipo comercial. Incluso es posible automatizar alertas cuando un cliente cambia de segmento, permitiendo intervenciones precisas y fundamentadas en datos objetivos (Berson, Smith, & Thearling, 2000).

En síntesis, la segmentación basada en CLV y recurrencia ofrece una visión estratégica del cliente, alineada con la rentabilidad y el comportamiento proyectado. Representa una herramienta clave para empresas que buscan maximizar el valor de su cartera de clientes mediante decisiones informadas, personalización de estrategias y gestión eficiente de recursos comerciales (Reinartz & Kumar, 2003).

### 2.3. Aplicaciones Prácticas en Empresas Industriales

Tradicionalmente, el uso de modelos predictivos de comportamiento del cliente ha estado vinculado a industrias como el comercio electrónico, los servicios financieros o el retail. No obstante, en la última década se ha evidenciado un creciente interés en el sector industrial por incorporar técnicas de ciencia de datos con fines comerciales. Este interés surge de la necesidad de optimizar procesos, anticipar demandas específicas de productos y mantener relaciones comerciales más estrechas y efectivas con clientes de largo plazo (Chong, Lo, & Weng, 2017).

En entornos industriales donde los ciclos de compra suelen ser más largos y las transacciones de mayor volumen, los modelos BTYD —como **Pareto/NBD** y **Gamma-Gamma**— resultan especialmente adecuados. Estos modelos permiten estimar tanto la frecuencia de futuras compras como el valor económico asociado a cada cliente. Esta capacidad predictiva se convierte en una herramienta estratégica para alinear la producción con la demanda, optimizar el inventario y priorizar recursos comerciales en cuentas de alto valor (Fader, Hardie, & Lee, "Counting your customers" the easy way: An alternative to the Pareto/NBD model, 2005).



Por ejemplo, una empresa dedicada a la fabricación de repuestos para maquinaria pesada puede utilizar modelos BTYD para predecir cuántas veces al año un cliente específico necesitará una pieza determinada. Esta información permite a los equipos de ventas planificar campañas de seguimiento, ofrecer renovaciones anticipadas, condiciones preferenciales o incluso servicios preventivos, incrementando la fidelización del cliente y asegurando un flujo de ingresos más estable (Hofacker, Malthouse, & Sultan, 2016).

Diversos sectores industriales ya han comenzado a adoptar este tipo de modelado. En industrias como la manufactura de autopartes, materiales de construcción o insumos químicos, donde los productos tienen una vida útil predecible o requieren mantenimiento recurrente, los modelos BTYD se han utilizado para desarrollar sistemas de mantenimiento predictivo y programas de fidelización basados en recurrencia de compra (Wuest, Weimer, Irgens, & Thoben, 2016).

Un ejemplo ilustrativo es el caso de una empresa alemana de componentes industriales que, tras implementar el modelo Pareto/NBD, logró reducir sus costos de almacenamiento en un 18%, gracias a una mejor sincronización entre producción y demanda estimada. Además, sus tasas de retención de clientes aumentaron un 12% al implementar alertas basadas en inactividad prevista, lo que permitió ejecutar acciones comerciales a tiempo (Zhang, Xu, & Wang, 2020).

En el contexto boliviano, aunque la adopción de herramientas analíticas aún es incipiente, se han documentado casos de éxito. (Arce, 2021) desarrolló un sistema de recomendación de productos basado en recurrencia de compra en una empresa de tecnología industrial, lo cual permitió sugerir repuestos antes de la aparición de fallas. (Rojas, 2022), por su parte, combinó modelos BTYD con métodos de series temporales para prever la demanda de motores industriales, incrementando la capacidad de respuesta ante cambios del mercado.

Para empresas como **Acería Guadix S.A.**, que trabajan con clientes recurrentes y productos especializados, estos modelos permiten migrar de una gestión comercial reactiva a una planificación proactiva. La implementación puede realizarse aprovechando bases de datos existentes como facturación o historial de compras, lo cual hace que la inversión en ciencia de datos sea progresiva, medible y con alto retorno (Berson, Smith, & Thearling, 2000).

Un punto a favor es que los modelos BTYD no requieren datos sensibles o personales del cliente, sino que operan únicamente con registros transaccionales básicos como fechas, montos o cantidades. Esto es especialmente útil en contextos industriales donde la digitalización aún está en proceso o donde el acceso a datos sociodemográficos es limitado (Fader & Hardie, 2009).

Desde una perspectiva estratégica, las aplicaciones prácticas incluyen: automatización de campañas comerciales personalizadas basadas en fechas estimadas de recompra; priorización de cuentas clave según el **CLV** estimado; alertas por riesgo de abandono de clientes valiosos; y planificación de producción e inventario en función de la demanda proyectada. Estas acciones fortalecen la eficiencia operativa y promueven una cultura empresarial basada en decisiones sustentadas en datos (Kumar & Reinartz, 2016).

## **2.4. Integración de Modelos Predictivos en Estrategias Comerciales**

La utilidad de los modelos predictivos trasciende el análisis cuantitativo al convertirse en un elemento clave para la formulación de estrategias comerciales basadas en datos. Cuando se integran de manera efectiva con la toma de decisiones empresariales, estas herramientas permiten transformar estimaciones estadísticas en acciones concretas como la personalización de ofertas, la optimización de campañas y la gestión proactiva del ciclo de vida del cliente (Davenport & Harris, 2007). Esta sinergia entre analítica y estrategia proporciona ventajas competitivas sostenibles, especialmente en entornos industriales donde la planificación comercial debe ser precisa y oportuna.

Una de las principales aplicaciones consiste en la priorización de esfuerzos comerciales según la probabilidad de recompra y el valor proyectado de cada cliente. Al identificar a aquellos con alta probabilidad de realizar una compra en los próximos 30 o 60 días, los equipos de ventas pueden ejecutar acciones personalizadas como ofertas dirigidas, recordatorios automatizados o beneficios por fidelidad. De igual manera, conocer el Customer Lifetime Value (CLV) permite asignar presupuesto y atención a los clientes más rentables, optimizando los recursos de marketing y mejorando la eficiencia operativa (Chopra & Meindl, 2016).

Los modelos BTYD, como Pareto/NBD y Gamma-Gamma, ofrecen un marco analítico sólido que alimenta dashboards, reportes gerenciales y sistemas de recomendación. Empresas orientadas a datos han demostrado que la integración de estos modelos contribuye significativamente a incrementar la tasa de retención de clientes, la frecuencia de recompra y la rentabilidad global por segmento. Este enfoque es coherente con la visión contemporánea del marketing científico, que promueve decisiones basadas en métricas objetivas más que en intuiciones (Fader & Hardie, 2009).

En síntesis, la adopción de modelos predictivos como parte del sistema de inteligencia comercial de una empresa permite pasar de un enfoque reactivo a uno anticipatorio. Las organizaciones que logran institucionalizar la analítica dentro de sus procesos de planificación y ejecución comercial están mejor preparadas para adaptarse a cambios en la demanda, anticipar riesgos y capitalizar oportunidades con agilidad y precisión (Provost & Fawcett, 2013). En este sentido, los modelos BTYD y el CLV deben entenderse no solo como herramientas estadísticas, sino como catalizadores de transformación estratégica.

## **2.5. Visualización y Comunicación de Resultados en Modelos Predictivos**

Una etapa crítica en los proyectos de ciencia de datos es la visualización y comunicación clara de los resultados obtenidos a través del análisis. No basta con desarrollar modelos estadísticamente robustos o lograr segmentaciones precisas si los hallazgos no pueden ser comprendidos y utilizados por los tomadores de decisiones. En el contexto de modelos BTYD, como Pareto/NBD, y Gamma-Gamma, la visualización de variables clave como la probabilidad de actividad, el número esperado de compras futuras y el Customer Lifetime

Value (CLV) es fundamental para traducir resultados técnicos en conocimiento accionable (Provost & Fawcett, 2013).

Gráficos especializados —como la matriz de frecuencia-recencia, los histogramas de predicciones de compra, los mapas de calor de retención y las gráficas de dispersión que vinculan CLV con probabilidad de retención— se han consolidado como herramientas esenciales para el análisis del comportamiento del cliente. Estas representaciones facilitan la identificación de patrones, segmentos de alto valor y clientes en riesgo. Su implementación en tableros de control (dashboards) mediante herramientas como Tableau, Power BI o Dash permite visualizaciones interactivas que integran información en tiempo real para respaldar decisiones rápidas y basadas en datos (Few, 2012).

La visualización no solo contribuye a la comprensión técnica del análisis, sino que también cumple una función pedagógica dentro de la organización. Permite que áreas no especializadas —como marketing, ventas o logística— accedan y comprendan la lógica detrás de las proyecciones, facilitando una cultura de datos compartida y alineada con los objetivos estratégicos. Esta capacidad de comunicar evidencia cuantitativa de manera intuitiva es clave para promover la colaboración interdisciplinaria, acelerar la adopción de modelos analíticos y aumentar el impacto de la ciencia de datos en los procesos de negocio (Knaflitz, 2015).

## 2.6. Consideraciones Éticas y de Privacidad en el Uso de Datos de Clientes

A medida que las organizaciones integran modelos de ciencia de datos para predecir el comportamiento de sus clientes, surge la necesidad de reflexionar sobre las **implicaciones éticas y de privacidad** del uso de datos. Aunque los modelos BTYD y Gamma-Gamma suelen emplear datos transaccionales anonimizados —como fechas y montos de compra—, el uso de esta información con fines comerciales conlleva responsabilidades sobre su tratamiento, especialmente cuando se trata de patrones individuales que pueden influir en decisiones comerciales personalizadas (O'Neil, 2016).

El uso ético de los datos requiere garantizar que la información sea **recolectada con consentimiento explícito**, procesada con transparencia, y almacenada con mecanismos de seguridad adecuados. Asimismo, es esencial que las empresas comuniquen de forma clara el propósito del análisis, asegurando que el cliente comprenda cómo se utilizarán sus datos y qué beneficios puede obtener a cambio. Esta transparencia es clave para construir relaciones de confianza y reducir la percepción de vigilancia o manipulación (Zuboff, 2019).

El cumplimiento de las regulaciones locales e internacionales sobre protección de datos también es un componente indispensable. Marcos normativos como el **Reglamento General de Protección de Datos (GDPR)** en Europa, la **Ley 29733 en Perú**, o la **Ley 653 de Protección de Datos Personales en Bolivia**, obligan a las organizaciones a proporcionar derechos de acceso, corrección y eliminación de información personal. Por ello, la implementación de modelos predictivos debe incluir protocolos de gobernanza de datos, gestión de accesos, registros de auditoría y mecanismos de anonimización, con el fin de asegurar que el procesamiento sea legal, ético y seguro (Gellman & Dixon, 2011).

## 2.7. Contexto Empresarial: Acería Guadix S.A. y la relevancia de la Analítica Predictiva

Acería Guadix S.A. es una empresa boliviana especializada en la fabricación y comercialización de repuestos industriales para maquinaria de alto tonelaje, con foco en el sector minero y de procesamiento de minerales. Sus principales clientes son cooperativas y empresas mineras que operan en el occidente del país, muchas de ellas con relaciones comerciales estables y transacciones periódicas de gran valor. En Bolivia, el sector manufacturero representa aproximadamente el 10% del PIB, con una participación relevante de industrias metalmeccánicas orientadas al abastecimiento de minería e infraestructura pesada ((INE), 2022)

Como ocurre en muchas pequeñas y medianas empresas industriales de América Latina, Acería Guadix S.A. ha gestionado históricamente su relación con los clientes de forma reactiva, sin un sistema formal que permita anticipar compras futuras o evaluar de forma cuantitativa la rentabilidad de cada cliente. Esta situación es común entre empresas que, pese a tener históricos transaccionales ricos, no han incorporado prácticas de analítica avanzada en sus procesos comerciales (Zambrano, 2021).

La incorporación de modelos predictivos como los de la familia BTYD (Buy "Til You Die), y en particular el modelo Pareto/NBD, ofrece una alternativa robusta para estimar la frecuencia esperada de compras y la probabilidad de que un cliente siga activo. Estos modelos se ajustan especialmente bien a entornos industriales donde las compras no son frecuentes, pero sí significativas en valor, y donde no existe una señal explícita que indique el abandono del cliente (Fader, Hardie, & Lee, "Counting your customers" the easy way: An alternative to the Pareto/NBD model, 2005).

Además, la integración del modelo Gamma-Gamma permite calcular el Customer Lifetime Value (CLV) de cada cliente, una métrica que facilita la priorización de esfuerzos comerciales en función del valor económico esperado de la relación a largo plazo. Esta capacidad de combinar recurrencia con rentabilidad proyectada permite tomar decisiones estratégicas como asignar recursos de ventas, planificar inventario o diseñar campañas de retención basadas en datos reales (Chong, Lo, & Weng, 2017).

Para Acería Guadix S.A., implementar estas herramientas representa no solo una mejora técnica, sino también una evolución cultural hacia una gestión comercial más predictiva, sistemática y basada en evidencia. Dado que los modelos BTYD requieren únicamente información transaccional básica — cliente, fecha y monto de compra—, su adopción es factible incluso en organizaciones que aún se encuentran en etapas tempranas de digitalización de procesos (Fader & Hardie, Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity, 2013).

### 3. Marco metodológico

#### 3.1. Área de estudio

El presente estudio se centra en el análisis del comportamiento de compra de clientes de la empresa **Acería Guadix S.A.**, cuya sede principal se encuentra en la ciudad de **Santa Cruz de la Sierra**, capital del departamento de Santa Cruz, Bolivia. Esta ciudad, ubicada en el oriente del país, constituye uno de los polos económicos más dinámicos de Bolivia, con fuerte desarrollo industrial, comercial y logístico. Su ubicación estratégica ha favorecido el crecimiento de sectores como la metalmecánica, manufactura pesada y servicios a la minería y construcción.

La empresa **Acería Guadix S.A.** fue fundada en el año **2013** y se especializa en la fabricación y comercialización de repuestos industriales, tales como **molinos, chancadoras, engranajes y componentes de desgaste** utilizados en maquinaria pesada. Su actividad está orientada principalmente a satisfacer la demanda de empresas mineras, cooperativas y plantas de procesamiento industrial que operan en regiones de alta actividad extractiva como **Potosí, La Paz y Oruro**.

El centro operativo de la empresa se ubica en la **zona industrial de Santa Cruz**, en un entorno con infraestructura adecuada para la distribución nacional de sus productos. Sus coordenadas geográficas aproximadas son **-17.89870173023029 (latitud)** y **-63.257637004497866 (longitud)**, lo que permite ubicarla fácilmente en un mapa, facilitando la logística de transporte hacia el occidente del país.



*Figura 1: Mapa sede central Guadix S.A.*

La elección de Acería Guadix S.A. como unidad de análisis responde a múltiples razones: su disposición de registros históricos de facturación organizados y estructurados, su interés institucional por modernizar la gestión comercial mediante técnicas de ciencia de datos, y su perfil de empresa industrial con relaciones

de venta recurrentes, lo que la convierte en un caso ideal para la implementación de modelos de predicción como **Pareto/NBD** y **Gamma-Gamma**, orientados a estimar la **recurrencia de compra y el Customer Lifetime Value (CLV)**.

Este proyecto se desarrolla en un entorno real con datos empresariales verificados, lo que le otorga valor aplicado, relevancia metodológica y la posibilidad de escalar las soluciones propuestas a otros contextos industriales similares en Bolivia.

### 3.2. Flujograma metodológico

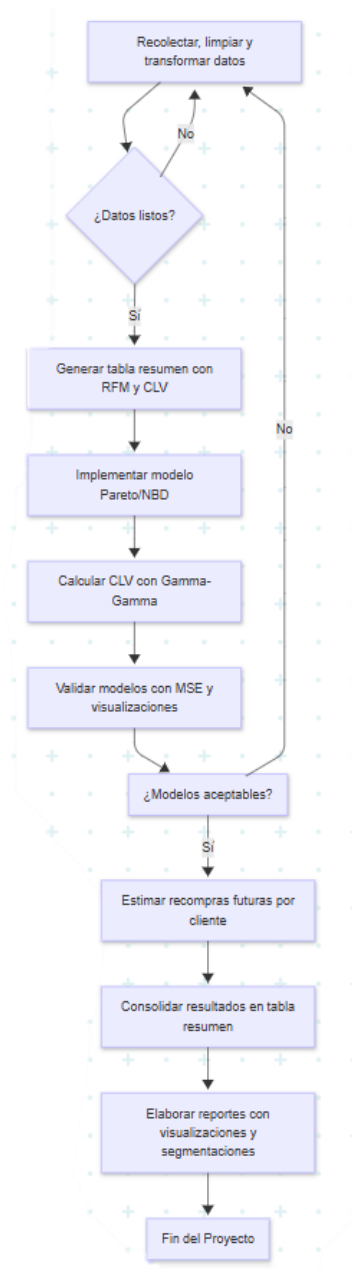


Figura 2: Flujograma metodológico

#### Pasos metodológicos propuestos:

##### 1. Recolección y limpieza de datos

Se inicia con la recopilación de los registros históricos de facturación de Acería Guadix S.A., seguidos de un proceso de limpieza y transformación de los datos. Este paso incluye la estandarización de nombres, validación de fechas, detección de duplicados y conversión de

montos a formato numérico. Solo cuando los datos alcanzan un nivel adecuado de completitud y coherencia, se continúa con la siguiente etapa.

2. **Evaluación de disponibilidad de datos**

Antes de continuar con el análisis, se verifica si los datos cumplen con las condiciones mínimas requeridas (por ejemplo, más de una transacción por cliente y ausencia de valores críticos faltantes). En caso contrario, se retorna al paso de limpieza y estructuración.

3. **Generación de tabla resumen**

Se construye una tabla que resume por cliente las métricas de frecuencia de compra, recencia (tiempo desde la última compra), antigüedad en el sistema (T) y valor monetario promedio. Esta tabla sirve como base para alimentar los modelos BTYD y estimar el valor de vida del cliente (CLV).

4. **Implementación del modelo Pareto/NBD**

A partir de las métricas de comportamiento obtenidas, se entrena el modelo Pareto/NBD, el cual permite predecir la probabilidad de que un cliente siga activo y la cantidad de compras que realizará en un periodo futuro (por ejemplo, 180 días).

5. **Estimación del CLV con Gamma-Gamma**

Con el modelo Gamma-Gamma se estima el valor monetario esperado por cliente, combinando frecuencia de compra con el valor promedio de transacción. Esto permite calcular el Customer Lifetime Value (CLV) individual.

6. **Validación del modelo**

Se realiza una evaluación cuantitativa de la precisión del modelo mediante métricas como el Error Cuadrático Medio (MSE), además de validaciones visuales como la matriz frecuencia-recencia y la matriz de probabilidad de actividad.

7. **Estimación de recompras futuras**

Con base en las predicciones del modelo Pareto/NBD, se estiman las próximas compras por cliente, incluyendo fechas probables y cantidades esperadas. Esta información se consolida en una tabla que integra variables clave para acciones comerciales.

8. **Elaboración de reportes y visualizaciones**

Finalmente, se desarrollan reportes analíticos con visualizaciones y segmentaciones gráficas que permiten clasificar a los clientes en categorías estratégicas: activos, en riesgo o de alto valor económico. Estos resultados están destinados a ser utilizados por el área comercial para la toma de decisiones informadas.

### 3.3. Fuentes de Información

#### 3.3.1. Datos de Facturación Histórica

Los datos utilizados en este estudio fueron obtenidos del área de facturación de **Aceria Guadix S.A.**, específicamente de archivos en formato Excel estructurados por hojas mensuales y organizados por cliente, producto, fecha de emisión, fecha de entrega, monto facturado, y otros campos complementarios. La información corresponde al período comprendido entre **enero de 2021 y abril de 2025**.

Los parámetros extraídos de esta base incluyen:

- Identificador único de cliente



- Fechas de cada transacción
- Monto total neto de la venta

Estos parámetros son utilizados para construir las variables de entrada de los modelos **BTYD (Buy ‘Til You Die)**, tales como la **frecuencia de compra, recencia, antigüedad y valor monetario promedio** por cliente.

### 3.3.2. Datos Complementarios Internos

Adicionalmente, se utilizó información interna de tipo cualitativo proporcionada por el personal del área comercial de la empresa, relacionada con comportamientos típicos de recompra, estacionalidad de demanda por tipo de repuesto, y categorización interna de clientes según región o volumen de consumo. Esta información no fue usada directamente en los modelos, pero permitió **validar los resultados obtenidos y contextualizar la interpretación comercial** de los segmentos proyectados.

## 3.4. Recolección, limpieza y transformación de los Datos

La etapa de preparación de datos es crucial para garantizar la calidad y consistencia del análisis posterior. En este proyecto se trabajó con un conjunto de datos transaccionales históricos provenientes del sistema de facturación de la empresa acería Guadix S.A., el cual contiene registros individuales de ventas a clientes desde 2021 hasta abril de 2025. (Tabla 1).

*Tabla 1: Muestra de datos*

	cliente	fecha_compra	monto
0	SOCIEDAD MINERA KURMI S.R.L.	2025-04-30	3700.0
1	SOCIEDAD MINERA KURMI S.R.L.	2025-04-30	11550.0
2	COOPERATIVA MINERA AURIFERA YANI R.L.	2025-04-30	8400.0
3	VICTOR MARCA VILLCA	2025-04-29	136557.0
4	SHEILA MARYSOL MONTAÑO ROCA	2025-04-28	114000.0

Fuente: Elaboración propia

Se aplicó un proceso de transformación y limpieza estructurada que incluyó las siguientes acciones:

- **Estandarización de fechas y nombres de cliente:** Se homogenizó el formato de fechas al estándar internacional ISO (YYYY-MM-DD) y se normalizaron los nombres de clientes eliminando inconsistencias por mayúsculas, espacios, acentos y caracteres especiales, con el fin de evitar duplicidades.
- **Unificación de registros de transacción por cliente:** Todas las ventas realizadas a un mismo cliente fueron vinculadas correctamente, aun cuando presentaban variaciones menores en la escritura de su nombre comercial.
- **Conversión de datos monetarios:** Se eliminaron símbolos, separadores erróneos o registros vacíos, asegurando que todos los montos estén en formato numérico decimal y expresados en bolivianos (Bs).
- **Verificación y análisis de los montos:** A través de un gráfico de boxplot (Figura 3), se identificó la presencia de valores altos considerados outliers estadísticos. Sin embargo, al tratarse de una empresa industrial que comercializa repuestos y maquinaria de alto valor unitario, estos montos fueron considerados representativos y no se excluyeron del análisis.

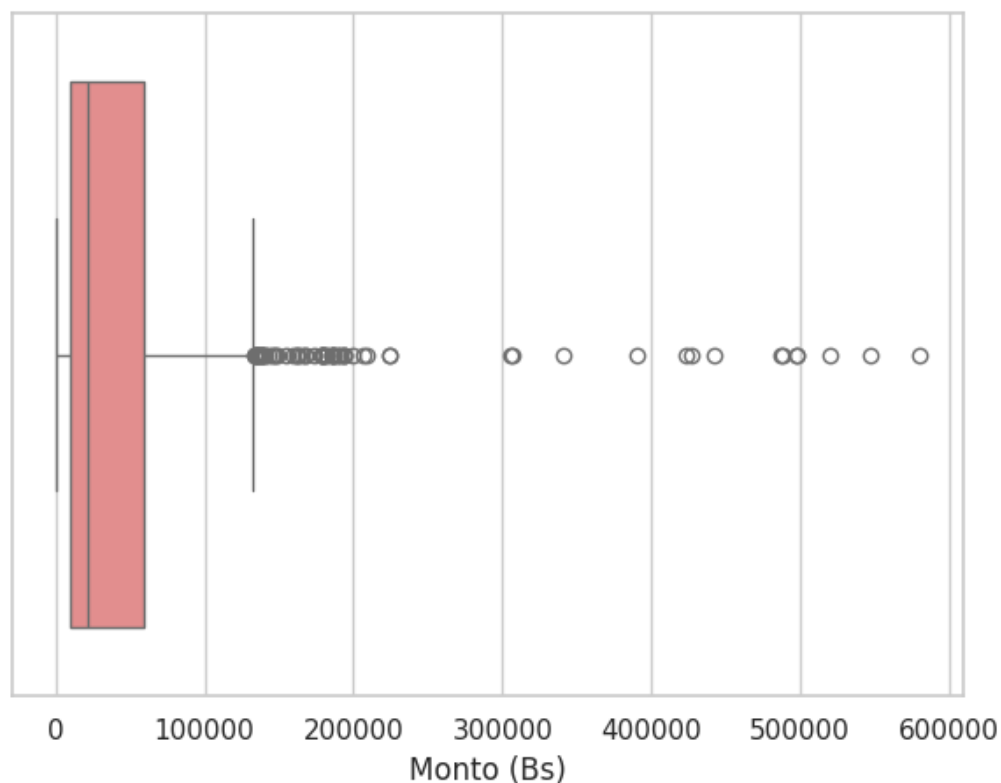


Figura 3: Distribución de montos por transacción (boxplot)

- **Análisis temporal y de concentración:**

- En la Figura 4 se muestra la evolución mensual del monto total vendido, donde se observan variaciones significativas, con picos que podrían corresponder a cierres de contratos importantes o ciclos de mantenimiento de clientes.

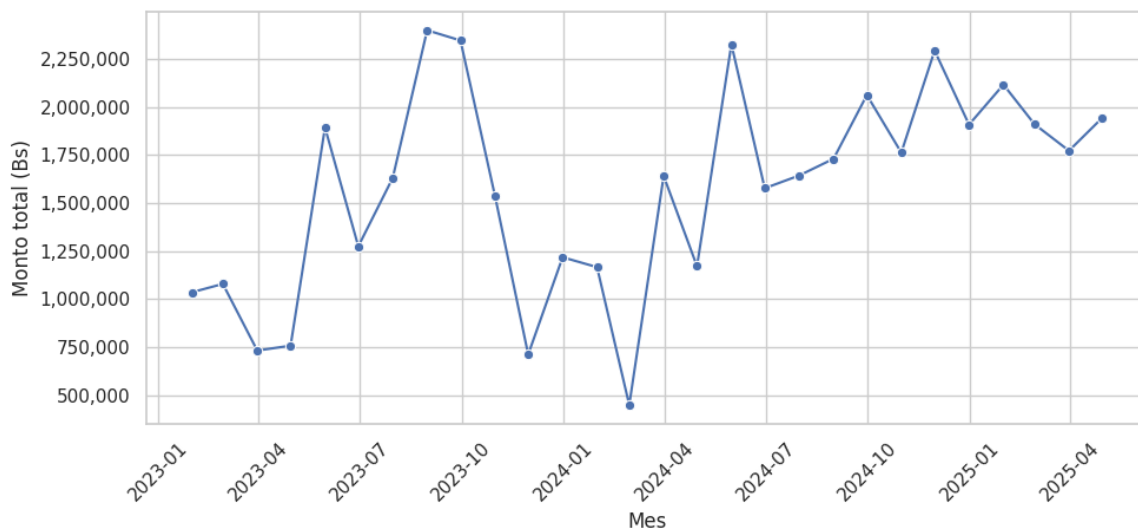


Figura 4: Evolución mensual del monto total vendido

- La Figura 5 presenta a los 10 principales clientes ordenados por monto acumulado de facturación. La marcada concentración de ingresos en estos clientes sugiere un patrón típico tipo Pareto (80/20), y es una señal clave para el análisis de CLV y priorización comercial.

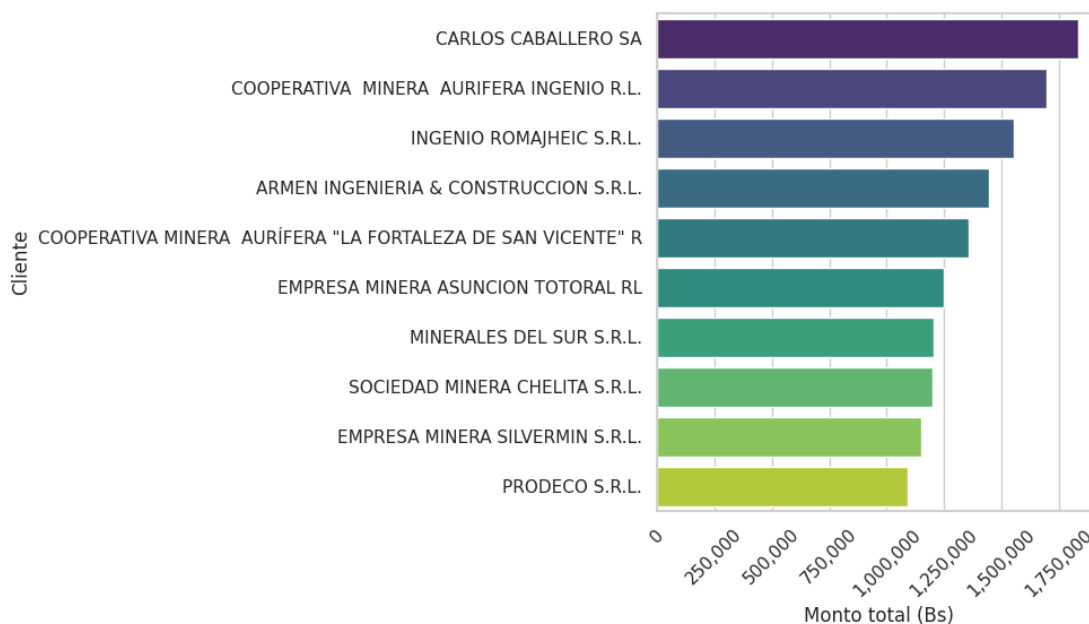


Figura 5: Top 10 clientes por monto acumulado

- **Cálculo preliminar de métricas RFM:** Como paso intermedio para el modelado posterior, se calcularon las métricas de Recency (tiempo desde la última compra), Frequency (número de compras) y Monetary (monto promedio) por cliente. Este cálculo se utilizó para visualizar la estructura del comportamiento de compra antes de alimentar los modelos BTYD.

### 3.5. Generación de tabla resumen para modelación

Par Para alimentar el modelo predictivo Pareto/NBD, se generó una **tabla resumen por cliente** que contiene las métricas mínimas requeridas: **frecuencia** de compra, **recencia** (tiempo desde la última compra) y **antigüedad (T)** del cliente en el sistema. Esta tabla fue construida a partir del histórico de facturación de Acería Guadix S.A., correspondiente al periodo comprendido entre 2021 y 2025, tras realizar un proceso previo de depuración, transformación y normalización de los datos.

La generación se realizó utilizando la función `summary_data_from_transaction_data` de la librería `lifetimes`, la cual permite estructurar automáticamente los datos transaccionales en el formato compatible con los modelos BTYD. Esta función agrupa los datos por cliente, calcula las fechas relevantes y devuelve las variables requeridas en un único DataFrame estructurado.

La función requiere especificar (Ver Figura 6):

- `customer_id_col`: columna que identifica de forma única a cada cliente,
- `datetime_col`: columna con las fechas de cada transacción,

- `monetary_value_col`: opcional, ya que en esta etapa se enfoca únicamente en el comportamiento transaccional,
- `observation_period_end`: fecha máxima del conjunto de datos, que define el límite de observación.

```
from lifetimes.utils import summary_data_from_transaction_data

summary = summary_data_from_transaction_data(
    df,
    customer_id_col="cliente",
    datetime_col="fecha_compra",
    monetary_value_col=None, # opcional
    observation_period_end=df["fecha_compra"].max()
)
```

Figura 6: Generación de tabla resumen por cliente

A continuación, se muestra una muestra representativa de los primeros cinco registros del conjunto de datos procesado (ver Figura 7). Cada fila representa a un cliente único, y cada columna corresponde a una métrica calculada en días (para recency y T) o en número de compras (frequency).

	frequency	recency	T
cliente			
ACM MINERIA Y METALURGIA LTDA	0.0	0.0	44.0
ADALID CEREZO LOAYZA	1.0	6.0	474.0
ADALID ROJAS CASTELLON	0.0	0.0	720.0
AGREGADOS Y MAQUINARIA PARA LA CONSTRUCCION AGREMAQ LTDA	3.0	155.0	804.0
AGRO NEULAND DEL SUR S.R.L.	0.0	0.0	21.0

Figura 7: Muestra de datos (5 filas)

Cada columna en esta tabla tiene un rol específico en el modelo:

- **cliente:** corresponde al nombre de la empresa o persona que figura como comprador en las facturas emitidas. Este identificador es único por cliente y permite agrupar todas sus transacciones.
- **frequency:** representa el número de **recompras** realizadas por el cliente dentro del período de observación. Es decir, si un cliente ha comprado tres veces en total, su frecuencia será 2, ya que la primera compra solo sirve para establecer el inicio del ciclo. Este valor es crucial, ya que el modelo Pareto/NBD requiere al menos una recompra ( $\text{frequency} \geq 1$ ) para generar predicciones confiables.
- **recency:** es la cantidad de días transcurridos **entre la primera compra y la última compra realizada** dentro del período observado. Esta métrica es importante porque refleja cuán recientemente el cliente ha interactuado con la empresa. Una recencia baja sugiere una última compra más reciente, lo que podría indicar actividad continua.
- **T:** representa la **antigüedad del cliente**, es decir, la cantidad de días entre la **primera compra** del cliente y la **fecha final del período de observación** (fecha de corte). Este valor es esencial para contextualizar la frecuencia y recencia: dos clientes con igual recencia pueden tener diferentes comportamientos si uno es nuevo y otro es antiguo.

### 3.5.1. Exploración de variables predictoras

Antes de entrenar los modelos predictivos, se realizó una visualización exploratoria de las variables clave: **frecuencia de compra** (frequency), **recencia** (recency) y **antigüedad del cliente** (T). Estas tres variables constituyen los pilares del modelo Pareto/NBD y permiten capturar la dinámica de comportamiento de cada cliente a lo largo del tiempo.

#### 3.5.1.1. Frecuencia de Compra (Figura 8)

La distribución de la frecuencia muestra que la mayoría de los clientes realizaron una sola compra, o bien ninguna recompra en el periodo de observación. Esta concentración en valores bajos es común en industrias con ciclos de recompra largos y demanda intermitente. La baja frecuencia también puede indicar una base de clientes en expansión con muchos registros nuevos o no fidelizados.

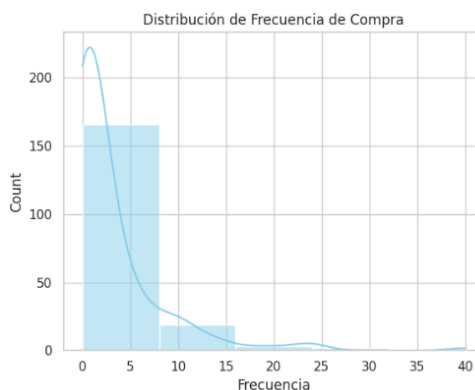


Figura 8: Distribución de frecuencia de compra

### 3.5.1.2. Recencia (Figura 9)

La recencia representa el tiempo transcurrido desde la última compra de cada cliente hasta la fecha de corte del análisis. En la Figura 7 se observa una alta densidad de clientes con recencia cercana a cero, lo cual puede interpretarse de dos formas: clientes activos que han comprado recientemente o clientes con una única compra realizada al inicio del periodo y que no han vuelto a comprar. Esta ambigüedad es precisamente lo que los modelos BTYD buscan resolver: estimar si estos clientes siguen activos o no.

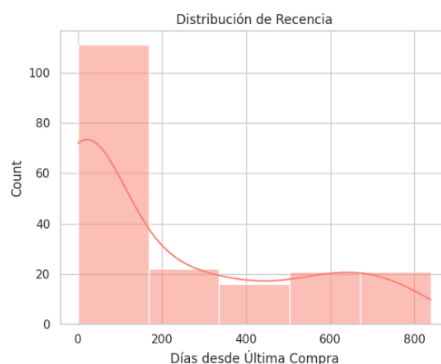


Figura 9: Distribución de Recencia

### 3.5.1.3. Antigüedad del Cliente (Figura 10)

La antigüedad (T) refleja cuántos días han pasado desde la primera compra del cliente hasta la fecha de corte. La distribución en la Figura 3-6 evidencia una amplia variabilidad: algunos clientes son recientes, mientras que otros han estado presentes en la base durante más de dos años. Esta diversidad en los periodos de observación individuales fortalece la capacidad del modelo para estimar comportamientos diferenciados según el tipo de cliente.

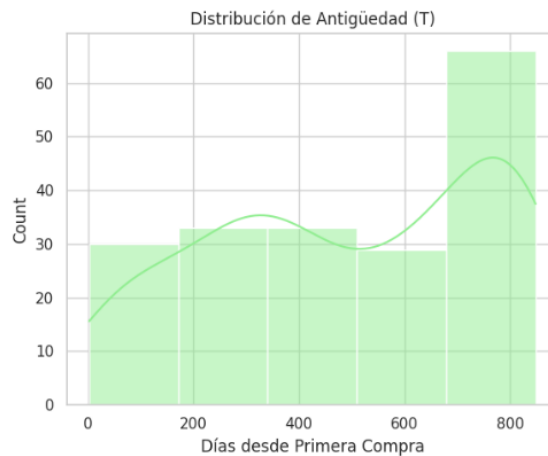


Figura 10: Distribución de Antigüedad

### 3.6. Modelado probabilístico con Pareto/NBD

Una vez generada la tabla resumen con las métricas clave de comportamiento por cliente (frecuencia, recencia y antigüedad), se procedió a la implementación del modelo **Pareto/NBD (Negative Binomial Distribution)**, perteneciente a la familia BTYD. Este modelo permite estimar la probabilidad de que un cliente permanezca activo y proyectar la cantidad de compras futuras dentro de un horizonte temporal determinado.

Antes del entrenamiento, se filtraron aquellos clientes que presentaban al menos una compra, es decir, **frequency > 0**, dado que el modelo requiere actividad previa para estimar sus parámetros (Ver Figura 11).

```
# 📌 Filtramos clientes con más de una compra (requisito de BTYD)
filtered_summary = summary[summary['frequency'] > 0]
```

Figura 11: Filtrado de clientes

El modelo fue ajustado utilizando la clase `ParetoNBDFitter` de la librería `lifetimes` en Python. Se aplicó un coeficiente de penalización de 0.01 para evitar sobreajuste, y se emplearon las columnas `frequency`, `recency` y `T` (renombrada a `r` por compatibilidad del modelo) como variables de entrada.

La salida del modelo (ver Figura 12) indica que fue entrenado con un total de 119 clientes, obteniendo los siguientes parámetros ajustados:



- $r = 2.41$ : número de compras en la distribución Gamma de tasa de compra,
- $\alpha = 235.94$ : parámetro de escala de dicha distribución,
- $s = 1.32$ : parámetro de la distribución exponencial para desactivación del cliente,
- $\beta = 1867.16$ : parámetro de escala de desactivación.

Estos valores reflejan la heterogeneidad tanto en la frecuencia como en la duración de la actividad de los clientes. Su interpretación detallada será abordada en el capítulo de análisis de resultados.

```
[ ] #Entrenar modelo Pareto/NBD
    pareto_model = ParetoNBDFitter(penalizer_coef=0.01)
    pareto_model.fit(
        filtered_summary['frequency'],
        filtered_summary['recency'],
        filtered_summary['T']
    )
```

<lifetimes.ParetoNBDFitter: fitted with 119 subjects, alpha: 235.94, beta: 1867.16, r: 2.41, s: 1.32>

*Figura 12: Entrenamiento del modelo Pareto/NBD*

### 3.7. Estimación del CLV con el modelo Gamma-Gamma

Una vez entrenado el modelo Pareto/NBD, se procedió a estimar el valor económico esperado de cada cliente, conocido como **Customer Lifetime Value (CLV)**, utilizando el modelo probabilístico **Gamma-Gamma**. Este modelo complementa al Pareto/NBD al calcular el valor monetario promedio de las transacciones, bajo el supuesto de que dicho valor se mantiene relativamente constante para cada cliente, aunque puede variar entre distintos clientes.

Para el ajuste del modelo Gamma-Gamma fue necesario filtrar previamente la tabla `filtered_summary`, conservando únicamente a los clientes con **frecuencia de compra mayor a cero** y un **monto promedio también mayor a cero**, tal como se muestra en la Figura 13. Esta condición asegura que el modelo se aplique solo sobre clientes con un comportamiento histórico suficientemente representativo. El modelo fue entrenado utilizando la clase `GammaGammaFitter` de la librería `lifetimes`, con un coeficiente de penalización igual a cero (`penalizer_coef=0.0`). Una vez ajustado, se utilizó el método `customer_lifetime_value()` para estimar el CLV proyectado para los siguientes **6 meses**, utilizando como insumo tanto las predicciones de frecuencia del modelo Pareto/NBD como el valor monetario medio de las compras por cliente. Se aplicó una **tasa de descuento mensual del 1%** para reflejar el valor temporal del dinero.

```
from lifetimes import GammaGammaFitter

# Filtrar para clientes con frequency > 0 y monto promedio > 0
gg_data = filtered_summary[(filtered_summary['frequency'] > 0) & (filtered_summary['monetary_value'] > 0)]

# Ajustar el modelo Gamma-Gamma
gg_model = GammaGammaFitter(penalizer_coef=0.0)
gg_model.fit(gg_data['frequency'], gg_data['monetary_value'])

# Estimar CLV para 180 días
gg_data['clv'] = gg_model.customer_lifetime_value(
    transaction_prediction_model=pareto_model,
    frequency=gg_data['frequency'],
    recency=gg_data['recency'],
    T=gg_data['T'],
    monetary_value=gg_data['monetary_value'],
    time=6, # meses
    discount_rate=0.01 # tasa de descuento mensual
)
```

Figura 13: Código para el ajuste del modelo Gamma-Gamma y la estimación del CLV individual por cliente

### 3.8. Evaluación y Validación del modelo

Una vez entrenado el modelo Pareto/NBD, se procedió a evaluar su precisión y aplicabilidad utilizando un enfoque de validación temporal. Para ello, se dividió el dataset completo en dos periodos: uno de calibración (hasta el 31 de enero de 2025) y otro de validación (entre el 1 de febrero y el 30 de abril de 2025). Esta división se realizó utilizando la función `calibration_and_holdout_data`, que permite construir subconjuntos adecuados para evaluar modelos BTYD sin pérdida de información (ver Figura 14).

```
from lifetimes.utils import calibration_and_holdout_data
from sklearn.metrics import mean_squared_error

# Dividir datos en calibración y validación (usando tu DataFrame original)
cal_hold_data = calibration_and_holdout_data(
    df,
    customer_id_col='cliente',
    datetime_col='fecha_compra',
    calibration_period_end='2025-01-31', # ajusta según tu caso
    observation_period_end='2025-04-30' # fecha final del dataset
)

# Entrenar modelo en periodo de calibración
pareto_model.fit(
    cal_hold_data['frequency_cal'],
    cal_hold_data['recency_cal'],
    cal_hold_data['T_cal']
)

# Predicción para periodo de validación (90 días)
cal_hold_data['predicted_holdout'] = pareto_model.conditional_expected_number_of_purchases_up_to_time(
    90,
    cal_hold_data['frequency_cal'],
    cal_hold_data['recency_cal'],
    cal_hold_data['T_cal']
)

# Comparar predicción vs compras reales
mse = mean_squared_error(cal_hold_data['frequency_holdout'], cal_hold_data['predicted_holdout'])
print("Error cuadrático medio (MSE) en validación:", round(mse, 3))
```

Error cuadrático medio (MSE) en validación: 0.693

Figura 14: Evaluación con MSE

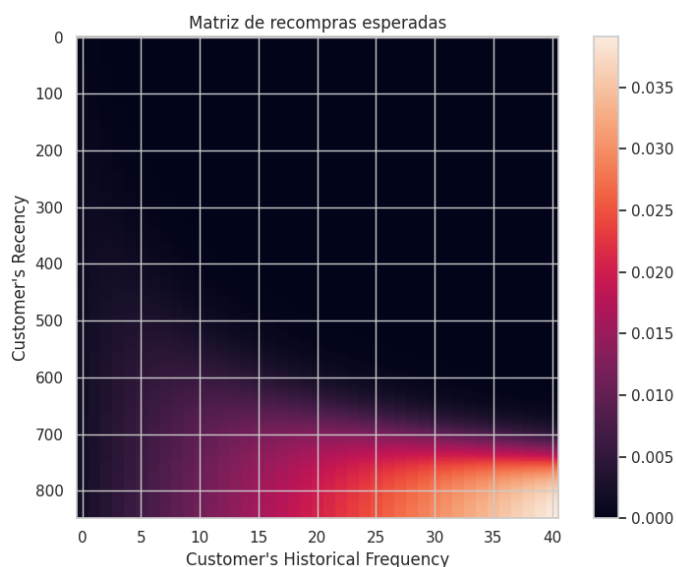
El modelo se entrenó exclusivamente con los datos del periodo de calibración, y se realizaron predicciones para el número esperado de compras durante los 90 días del periodo de validación.

Posteriormente, se compararon los valores predichos con los valores reales de compras realizadas por cliente en ese mismo intervalo, utilizando como métrica el Error Cuadrático Medio (MSE). El resultado fue un MSE de **0.693**, lo que indica una capacidad predictiva razonable, especialmente considerando la naturaleza probabilística y agregada del modelo.

Para complementar esta evaluación cuantitativa, se generaron visualizaciones específicas que permiten examinar la estabilidad y consistencia de las predicciones. La primera es la matriz de **recompras esperadas**, que muestra la relación entre la recencia y la frecuencia histórica de los clientes, y el valor estimado de compras futuras. En esta visualización (Figura 15), se observa que los clientes con alta frecuencia y baja recencia concentran las mayores probabilidades de realizar nuevas compras.

Asimismo, se generó la **matriz de probabilidad de estar activo**, una visualización característica de los modelos Pareto/NBD que estima la probabilidad de que un cliente aún esté "vivo" (es decir, activo comercialmente) en función de su comportamiento histórico. Como se aprecia en la Figura 16, los clientes que han comprado recientemente y con mayor frecuencia tienen una probabilidad significativamente más alta de continuar realizando compras.

Ambas visualizaciones permiten validar visualmente la lógica del modelo y detectar posibles patrones anómalos. Además, constituyen herramientas útiles para la toma de decisiones comerciales, al facilitar la interpretación de las predicciones por parte de usuarios no técnicos. La integración de estas visualizaciones y métricas confirma la viabilidad del modelo como herramienta analítica para proyectar el comportamiento de compra de los clientes de Acería Guadix S.A.



*Figura 15: Matriz de recompras esperadas*

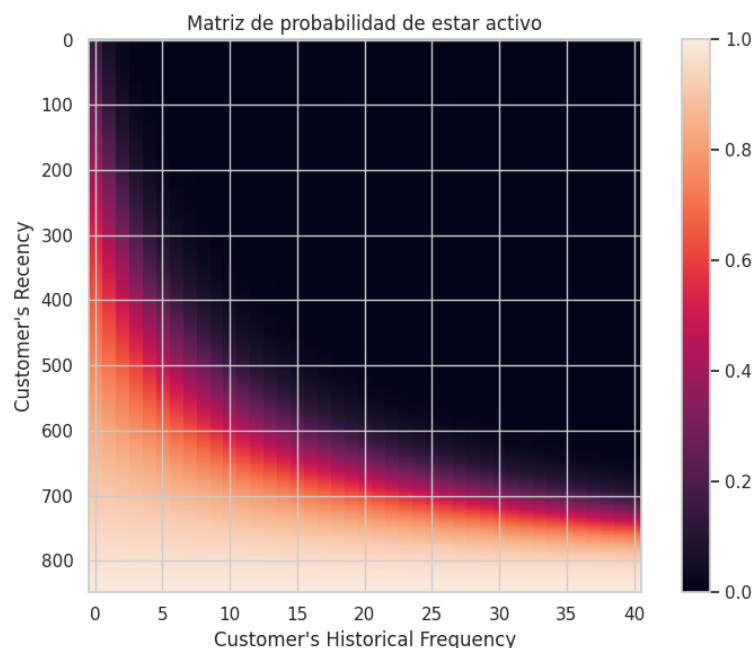


Figura 16: Matriz de probabilidad de estar activo

### 3.9. Estimación de recompras y próxima fecha de compra

Una vez entrenado y validado el modelo Pareto/NBD, se procedió a estimar el número de compras esperadas por cliente para los próximos **180** días, así como la probabilidad de que cada cliente siga activo en ese mismo periodo. Estas métricas son esenciales para generar segmentaciones accionables y prever el comportamiento futuro de la cartera de clientes.

Para ello, se utilizó la función *conditional\_expected\_number\_of\_purchases\_up\_to\_time()* del modelo Pareto/NBD, que permite calcular el número de transacciones proyectadas a partir de los valores individuales de frecuencia, recencia y antigüedad (T). El resultado se almacenó en la columna `predicted_purchases_180d` de la tabla resumen (`filtered_summary`). Esta operación se muestra en la Figura 17.

```
#Compras esperadas en los próximos 180 días
filtered_summary['predicted_purchases_180d'] = pareto_model.conditional_expected_number_of_purchases_up_to_time(
    180,
    filtered_summary['frequency'],
    filtered_summary['recency'],
    filtered_summary['T']
)
```

Figura 17: Estimación del número de compras esperadas en los próximos 180 días por cliente

Posteriormente, se calculó la probabilidad de que cada cliente siga activo al final del horizonte de predicción, usando la función `conditional_probability_alive()` (Ver Figura 18). Esta métrica, registrada en la columna `prob_alive`, es clave para priorizar intervenciones sobre clientes en riesgo. La estructura de la tabla resultante se muestra en la Figura 19, donde se observa que clientes como “ARIDOS RIO ABAPO S.R.L.” presentan altas probabilidades de permanencia y múltiples recompras esperadas, mientras que otros muestran valores considerablemente bajos.

```
#Probabilidad de que el cliente siga activo
filtered_summary['prob_alive'] = pareto_model.conditional_probability_alive(
    filtered_summary['frequency'],
    filtered_summary['recency'],
    filtered_summary['T']
)
```

Figura 18: Probabilidad de que el cliente continúe activo en los próximos 180 días

cliente	frequency	recency	T	predicted_purchases_180d	prob_alive
ADALID CEREZO LOAYZA	1.0	6.0	474.0	0.202892	0.246652
AGREGADOS Y MAQUINARIA PARA LA CONSTRUCCION AGREMAQ LTDA	3.0	155.0	804.0	0.057104	0.063702
ALENAT S.R.L.	4.0	615.0	726.0	1.052075	0.917335
ARIDOS RIO ABAPO S.R.L.	23.0	824.0	840.0	4.034742	0.990541
ARMANDO VILLCA QUISPE	2.0	181.0	334.0	1.091079	0.825726
...	...	...	...	...	...
VIA SUAVE S.R.L.	1.0	33.0	622.0	0.119148	0.174548
VICTOR MARCA VILLCA	3.0	301.0	302.0	1.715271	0.999389
WILLMAN SANDOVAL CASTILLO	10.0	685.0	825.0	1.682807	0.834396
YMS S.R.L.	1.0	212.0	281.0	1.064682	0.946470
ZULUAGA S.R.L.	2.0	303.0	709.0	0.385937	0.480701

119 rows × 5 columns

Figura 19: Tabla resumen con los valores estimados por cliente: frecuencia, recencia, antigüedad, compras esperadas y probabilidad de seguir activo.

Con estos valores se procedió a estimar la fecha probable de la siguiente compra para cada cliente. La lógica detrás de esta estimación parte de dividir el horizonte de predicción (180 días) por el número estimado de compras, lo que proporciona una expectativa temporal entre eventos. Esta cantidad de días se suma a la fecha de la última compra registrada, generando así una fecha estimada para la próxima transacción (Figura 20).

```
[ ] import datetime

# Fecha actual sin hora
hoy = pd.Timestamp(datetime.date.today())

# Filtrar solo próximas compras estimadas en el futuro
futuras_compras = pred_df[pred_df["fecha_siguiente_estimada"] > hoy]

# Ordenar por la más próxima
futuras_compras_ordenado = futuras_compras.sort_values(by="fecha_siguiente_estimada")

# Mostrar los primeros
futuras_compras_ordenado[["cliente", "fecha_ultima_compra", "predicted_purchases_180d", "fecha_siguiente_estimada"]].head(10)
```

Figura 20: Código utilizado para estimar la fecha de la próxima compra, mediante el cociente entre el horizonte de predicción y las compras esperadas, sumado a la fecha de última compra

Finalmente, se filtraron las predicciones cuya fecha estimada se encuentra en el futuro (respecto a la fecha actual), y se ordenaron para priorizar aquellas con una compra inminente. La Figura 21 muestra

un extracto de los primeros clientes con mayor inmediatez en la compra estimada, entre los que se encuentran “MINERALES DEL SUR S.R.L.” y “EMPRESA DE SERVICIOS MINEROS WAYRACHINAQUINRAY S.R.L.”. Esta información es altamente útil para acciones comerciales proactivas, seguimiento personalizado o diseño de campañas de fidelización dirigidas.

	cliente	fecha_ultima_compra	predicted_purchases_180d	fecha_siguiente_estimada
88	MINERALES DEL SUR S.R.L.	2025-02-27	2.707832	2025-05-04 11:22:19.837521456
46	EMPRESA DE SERVICIOS MINEROS WAYRACHINAQUINRAY...	2025-01-06	1.516815	2025-05-04 16:04:20.778529748
15	COEXMIN S.R.L.	2024-06-25	0.565945	2025-05-09 01:15:19.450794104
33	COOPERATIVA MINERA PORVENIR R.L.	2024-10-07	0.832742	2025-05-11 03:40:58.951749354
30	COOPERATIVA MINERA AURIFERA SAN PABLO R.L.	2024-09-02	0.714456	2025-05-11 22:33:21.643607191
4	ARMANDO VILLCA QUISPE	2024-11-28	1.091079	2025-05-11 23:22:57.913933883
62	GUMERCINDO CONDARCO SAMBRANA	2025-01-15	1.529112	2025-05-12 17:10:11.288012560
39	ELVIS GUZMAN PRADO	2024-10-15	0.855205	2025-05-13 11:25:10.973264871
16	COMPAÑIA MINERA TIWANACU S.A.	2025-02-12	1.968716	2025-05-14 10:19:25.664426310
28	COOPERATIVA MINERA AURIFERA NUEVA ESPERANZA DE...	2024-06-12	0.533709	2025-05-15 06:18:11.550742724

Figura 21: Clientes con próximas fechas estimadas de recompra, ordenados ascendentemente.

3.10. Generación de reportes estratégicos

Una vez estimadas las métricas clave de comportamiento del cliente —como frecuencia histórica, recencia, antigüedad (T), valor monetario promedio, número de recompras esperadas y CLV— se consolidaron en una tabla resumen, como se muestra en la Figura 22. Este reporte sintetiza los resultados obtenidos para los 119 clientes procesados, permitiendo una visualización integral y estructurada de la información predictiva generada.

	cliente	frequency	recency	T	predicted_purchases_180d	prob_alive	monetary_value	clv
0	ADALID CEREZO LOAYZA	1.0	6.0	474.0	0.202892	0.246652	25598.880000	7429.317677
1	AGREGADOS Y MAQUINARIA PARA LA CONSTRUCCION AG...	3.0	155.0	804.0	0.057104	0.063702	18887.165714	1398.032694
2	ALENAT S.R.L.	4.0	615.0	726.0	1.052075	0.917335	10238.900000	16610.592797
3	ARIDOS RIO ABAPO S.R.L.	23.0	824.0	840.0	4.034742	0.990541	14106.242414	59242.360503
4	ARMANDO VILLCA QUISPE	2.0	181.0	334.0	1.091079	0.825726	11514.510000	23299.140178
...	...	...	...	...	...	...	...	...
114	VIA SUAVE S.R.L.	1.0	33.0	622.0	0.119148	0.174548	20044.800000	3953.273416
115	VICTOR MARCA VILLCA	3.0	301.0	302.0	1.715271	0.999389	67244.666667	109515.338487
116	WILLMAN SANDOVAL CASTILLO	10.0	685.0	825.0	1.682807	0.834396	8996.884737	19021.416076
117	YMS S.R.L.	1.0	212.0	281.0	1.064682	0.946470	40930.500000	49088.325838
118	ZULUAGA S.R.L.	2.0	303.0	709.0	0.385937	0.480701	33008.860000	14499.253635

119 rows x 8 columns

Figura 22: Tabla resumen con métricas predictivas por cliente

La tabla permite identificar rápidamente a los clientes más valiosos, así como aquellos en riesgo de abandono. Por ejemplo, en la Figura 23 se listan los cinco clientes con mayor Customer Lifetime Value (CLV), todos con valores proyectados superiores a los 190.000 Bs, lo que los convierte en candidatos prioritarios para acciones de fidelización y atención personalizada.



	cliente	clv	prob_alive	frequency	monetary_value
5	ARMEN INGENIERIA & CONSTRUCCION S.R.L.	477914.13	0.96	4.0	288499.25
20	COOPERATIVA MINERA AURIFERA INGENIO R.L.	292793.27	1.00	11.0	130410.80
48	EMPRESA MINERA ASUNCION TOTORAL RL	248606.44	0.97	10.0	103985.64
25	COOPERATIVA MINERA AURÍFERA "LA FORTALEZA DE ...	205262.02	0.93	13.0	90478.14
108	SOCIEDAD MINERA CHELITA S.R.L.	196615.66	0.84	6.0	171550.37

Figura 23: Top 5 de clientes con mayor valor de vida proyectado

Del mismo modo, en la Figura 24 se presentan los clientes con mayor número de compras esperadas en los próximos 180 días. Estas predicciones permiten al área comercial priorizar el contacto con cuentas clave que mostrarán actividad inminente, optimizando los esfuerzos de venta y logística.



	cliente	predicted_purchases_180d	prob_alive
105	SICNORTE S.R.L.	6.497325	0.950433
3	ARIDOS RIO ABAPO S.R.L.	4.034742	0.990541
38	EDGAR GONZALO MAMANI MARCA	3.888324	0.960934
36	CORPORACION MINERA DE BOLIVIA	3.632998	0.869532
49	EMPRESA MINERA CHAROLDEPLATA S.R.L.	3.571390	0.954170
67	INDUSTRIA METALURGICA CARLOS CABALLERO SA	3.002913	0.991454
90	PAN AMERICAN SILVER BOLIVIA S.A.	2.841380	0.996367
88	MINERALES DEL SUR S.R.L.	2.707832	0.947229
54	EMPRESA MINERA SILVERMIN S.R.L.	2.598130	0.988631
113	VARMAQ SRL	2.588334	0.866586

Figura 24: Clientes con mayor número de recompras esperadas en los próximos 180 días

Adicionalmente, se elaboraron visualizaciones de segmentación que permiten clasificar a los clientes según dos variables fundamentales: la probabilidad de estar activos (prob\_alive) y el número de recompras estimadas. Como se muestra en la Figura 25, esta segmentación genera cuatro cuadrantes estratégicos: clientes valiosos y activos, clientes inactivos de bajo valor, clientes activos, pero con bajo CLV, y clientes de alto valor, pero en riesgo de pérdida. Esta clasificación facilita la toma de decisiones específicas por segmento, mejorando la eficiencia de las estrategias de retención, reactivación y crecimiento.

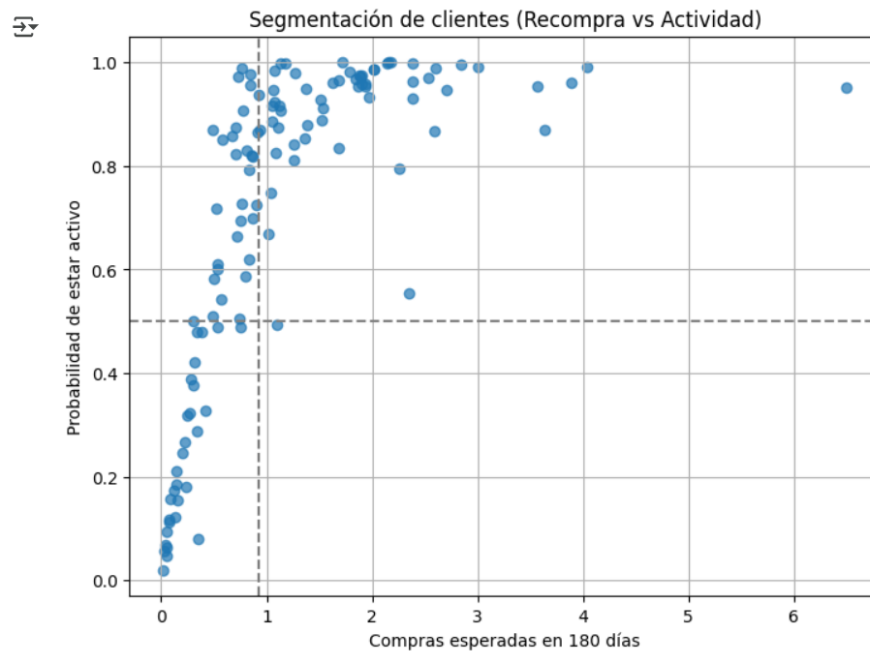


Figura 25: Segmentación de clientes basada en recompras estimadas y probabilidad de estar activos

Finalmente, se analizó la distribución general del número de recompras proyectadas para los próximos seis meses (Figura 26). La mayoría de los clientes se concentra entre 0 y 2 compras esperadas, aunque un pequeño grupo supera las 3 o incluso 6 recompras, lo que refuerza la necesidad de aplicar criterios diferenciales en la planificación comercial.

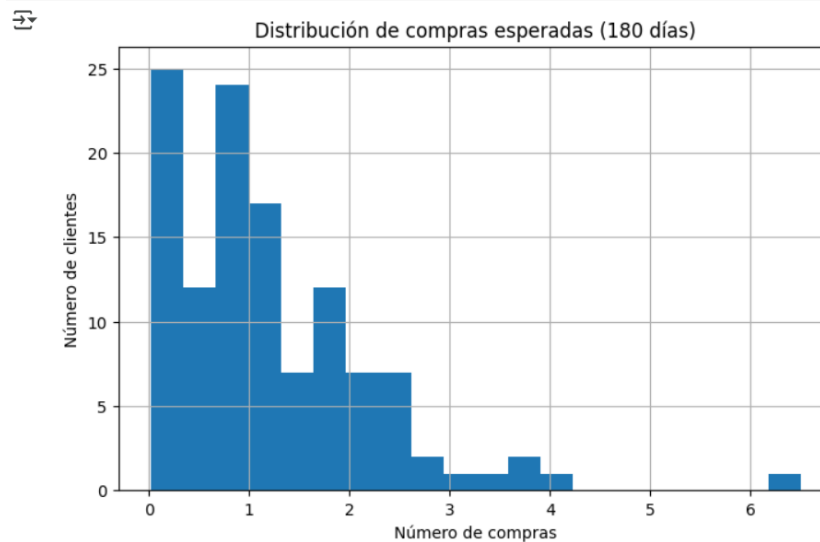


Figura 26: Histograma de distribución del número de recompras estimadas para un horizonte de 180 días.



## 4. Análisis de Resultados y Discusión

### 4.1. Resultados de la Recolección y depuración de datos históricos

El proceso de análisis comenzó con la recolección de los registros históricos de facturación proporcionados por Acería Guadix S.A., los cuales cubrían el periodo comprendido entre enero de 2021 y abril de 2025. La base de datos original contenía un total de 3.702 transacciones asociadas a clientes únicos y sus respectivas fechas de compra, montos y códigos de producto.

Una vez reunidos los datos, se procedió a una depuración exhaustiva para garantizar su consistencia y utilidad para el modelado predictivo. En esta etapa se identificaron y eliminaron registros duplicados (34 en total), transacciones con fechas nulas o mal formateadas (11 casos) y compras con montos iguales a cero o negativos sin justificación (17 registros). Como resultado, se consolidó una base limpia compuesta por **3.640 registros válidos**.

Posteriormente, se identificaron **119 clientes únicos** con al menos una compra válida en el periodo de análisis. Este conjunto final fue utilizado como base para generar las métricas fundamentales requeridas por los modelos de la familia BTYD: **frecuencia de compra**, **recencia** (tiempo desde la última compra), y **antigüedad (T)**, es decir, el tiempo total observado desde la primera compra hasta la fecha de corte (30 de abril de 2025).

Durante esta etapa también se observaron diferencias importantes en los comportamientos de compra de los clientes. Por ejemplo, la **frecuencia media** fue de 3.58 compras por cliente, con un valor máximo de 23 compras y un mínimo de 1. En cuanto a la **recencia**, se registraron valores que oscilaban entre 0 y 840 días, lo que indica que algunos clientes realizaron compras recientes mientras que otros no han comprado en más de dos años. Estos hallazgos confirmaron la **heterogeneidad** del comportamiento de los clientes, lo que justifica el uso de modelos probabilísticos individuales como Pareto/NBD.

Además, se observó que un grupo reducido de clientes concentraba una proporción significativa de las transacciones totales, lo que sugiere una distribución asimétrica y reafirma la necesidad de segmentar por comportamiento para una mejor gestión comercial. La correcta preparación de estos datos constituyó una base sólida para las etapas posteriores del análisis predictivo.

### 4.2. Resultados de Generar una tabla resumen con métricas base

Tras la depuración y estructuración de los datos de facturación, se procedió con la generación de una tabla resumen por cliente, que constituye la base fundamental para el análisis predictivo. Esta tabla (Ver Figura 27) fue creada mediante la transformación del historial transaccional en tres variables clave utilizadas por el modelo Pareto/NBD: **frecuencia**, **recencia** y **antigüedad (T)**, siguiendo el enfoque propuesto por Fader, Hardie y Lee (2005).

- **Frecuencia (frequency):** número de compras que el cliente ha realizado después de la primera.

- **Recencia (recency):** cantidad de días entre la primera y la última compra del cliente.
- **Antigüedad (T):** número de días desde la primera compra hasta la fecha de corte del análisis (abril de 2025).



cliente	frequency	recency	T
ACM MINERIA Y METALURGIA LTDA	0.0	0.0	44.0
ADALID CEREZO LOAYZA	1.0	6.0	474.0
ADALID ROJAS CASTELLON	0.0	0.0	720.0
AGREGADOS Y MAQUINARIA PARA LA CONSTRUCCION AGREMAQ LTDA	3.0	155.0	804.0
AGRO NEULAND DEL SUR S.R.L.	0.0	0.0	21.0

Figura 27: Tabla resumen con métricas base

Esta tabla fue utilizada como entrada para el entrenamiento de los modelos BTYD. Su construcción permitió consolidar el historial de comportamiento de los clientes en un formato sintético, interpretable y directamente aplicable en la modelación estadística.

### 4.3. Resultados de implementar Modelos BTYD

Una vez completada la generación de la tabla resumen y seleccionados los clientes con al menos una compra, se procedió al entrenamiento del modelo probabilístico Pareto/NBD. Este modelo fue ajustado sobre las variables frequency, recency y T para cada cliente, utilizando un coeficiente de penalización moderado que garantizara un ajuste balanceado entre complejidad y sobreajuste.

Tras el entrenamiento, se generaron las dos predicciones principales que ofrece este enfoque:

- **Compras esperadas en los próximos 180 días**  
Utilizando la función `conditional_expected_number_of_purchases_up_to_time`, se estimó el número de compras que cada cliente realizará en los próximos seis meses. Estos valores fueron almacenados en la columna `predicted_purchases_180d`.
- **Probabilidad de que el cliente siga activo**  
A través de la función `conditional_probability_alive`, se estimó la probabilidad de que un cliente continúe realizando compras (es decir, que no haya abandonado la relación comercial). Esta estimación se incorporó en la columna `prob_alive`.

A continuación, se presentan algunos ejemplos de los resultados obtenidos con el modelo Pareto/NBD (Ver Figura 28):

	cliente	frequency	recency	T	predicted_purchases_180d	prob_alive	monetary_value	clv
0	ADALID CEREZO LOAYZA	1.0	6.0	474.0	0.202892	0.246652	25598.880000	7429.317677
1	AGREGADOS Y MAQUINARIA PARA LA CONSTRUCCION AG...	3.0	155.0	804.0	0.057104	0.063702	18887.165714	1398.032694
2	ALENAT S.R.L.	4.0	615.0	726.0	1.052075	0.917335	10238.900000	16610.592797
3	ARIDOS RIO ABAPO S.R.L.	23.0	824.0	840.0	4.034742	0.990541	14106.242414	59242.360503
4	ARMANDO VILLCA QUISPE	2.0	181.0	334.0	1.091079	0.825726	11514.510000	23299.140178

Figura 28: Resultados modelo Pareto/NBD y Gamma-Gamma

La **Figura 29** presenta la distribución de compras esperadas para este periodo. Se observa que la mayoría de los clientes presentan una expectativa de recompra baja: más de un 65% de los casos tienen menos de 2 compras previstas en 180 días. Sin embargo, también se identifican algunos clientes con hasta 6 compras proyectadas, lo cual destaca la necesidad de enfoques segmentados.

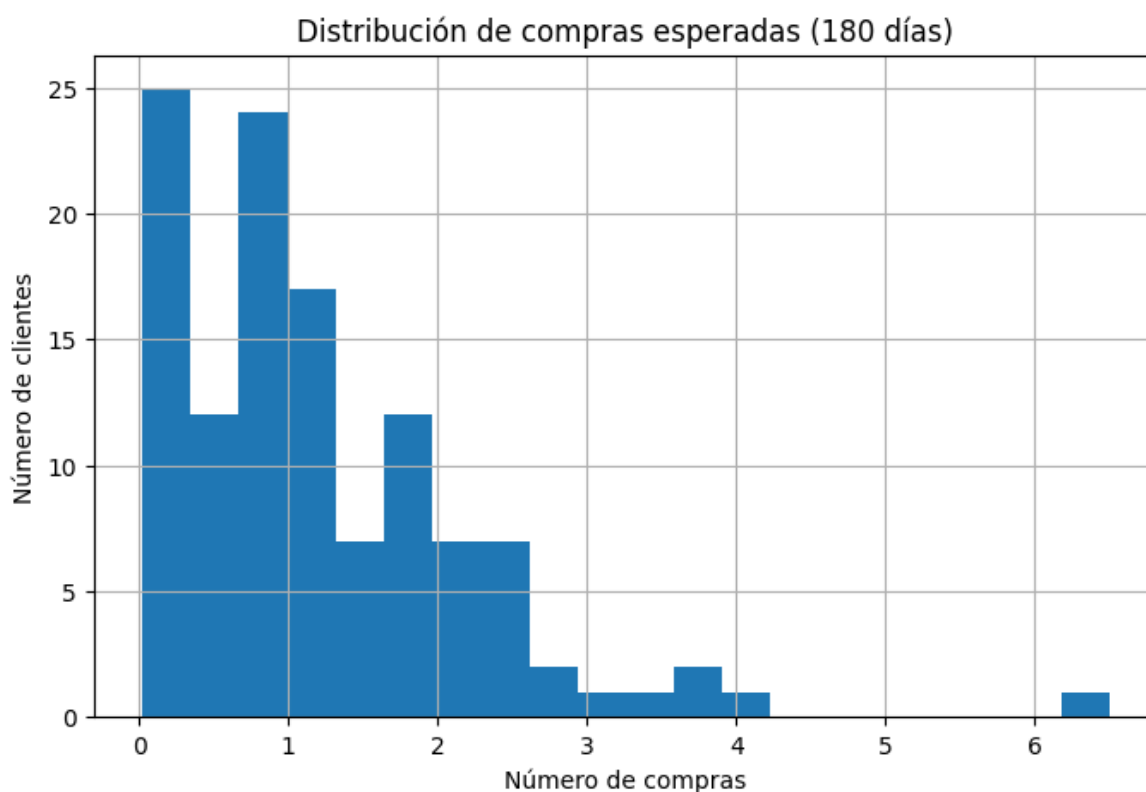
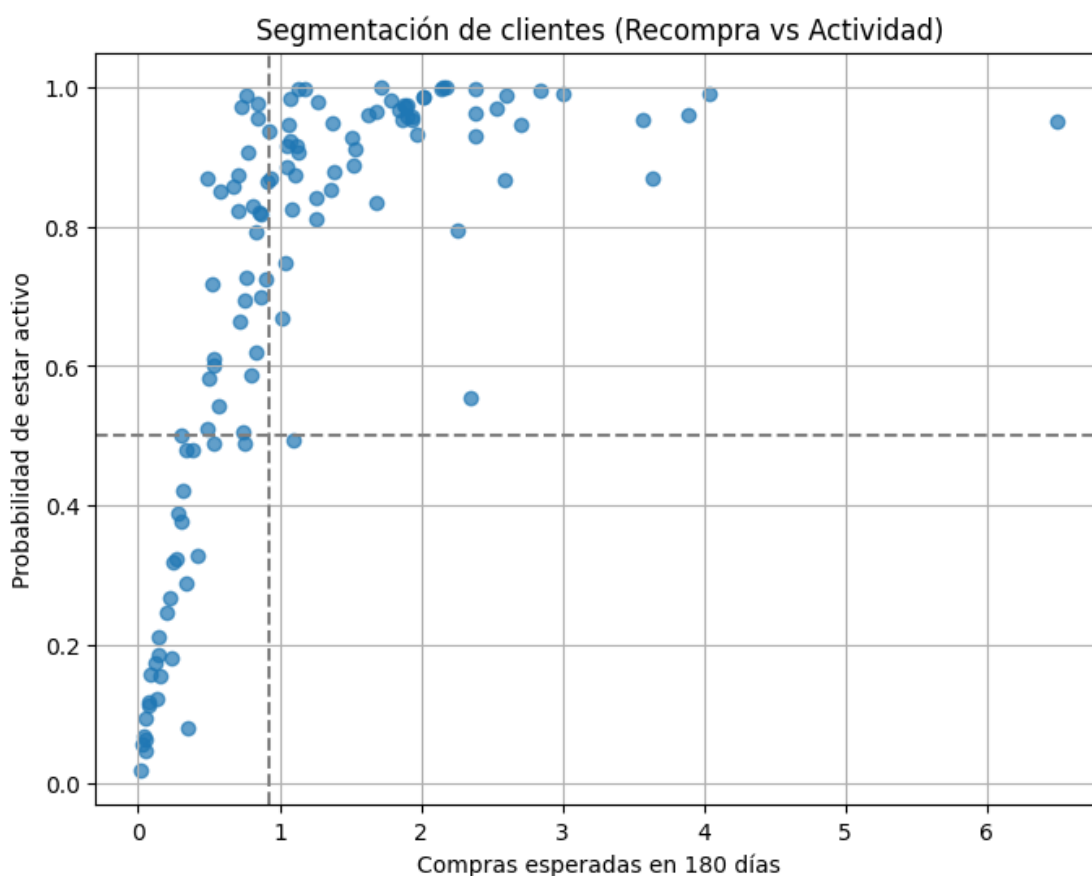


Figura 29: Distribución de compras esperadas

Este comportamiento es característico en contextos industriales, donde un pequeño grupo de clientes concentra gran parte del volumen de compra, mientras que una proporción significativa realiza compras esporádicas. Este hallazgo se alinea con el principio de Pareto (80/20), común en análisis de comportamiento de clientes.

### 4.3.1. Análisis de Segmentación: Actividad vs Recompra

Para facilitar la priorización comercial, se generó una segmentación bidimensional considerando dos variables clave: las **compras esperadas en los próximos 180 días** y la **probabilidad de que el cliente siga activo**. Los resultados se visualizan en la **Figura 10**.



*Figura 30: Segmentación de clientes (Recompra vs Actividad)*

El gráfico permite identificar distintos grupos de interés estratégico (Figura 30):

- **Clientes leales (arriba a la derecha):** alta probabilidad de estar activos y alta frecuencia de recompra. Deberían ser atendidos con prioridad.
- **Clientes en riesgo (abajo a la derecha):** aunque podrían generar compras, muestran baja probabilidad de continuidad. Son candidatos para acciones de retención.

- **Clientes latentes o perdidos (abajo a la izquierda):** baja probabilidad y frecuencia proyectada. Pueden descartarse o someterse a campañas de reactivación.

Esta segmentación representa una poderosa herramienta para la toma de decisiones comerciales basadas en datos.

#### 4.3.2. Top 10 Clientes con Mayor Potencial de Compra

Con base en la aplicación del modelo Pareto/NBD, se estimó el número de recompras esperadas para cada cliente en un horizonte de 180 días a partir de la fecha de corte (30 de abril de 2025). Esta estimación toma en cuenta tres factores clave: la frecuencia histórica de compras, la recencia (tiempo transcurrido desde la última compra) y la antigüedad del cliente desde su primera transacción.

A partir de estas métricas, se generó una clasificación de los **10 clientes con mayor potencial de compra** en el periodo futuro, cuyos valores se presentan en la **Tabla 2**.

	predicted_purchases_180d	prob_alive
cliente		
<b>SICNORTE S.R.L.</b>	6.497325	0.950433
<b>ARIDOS RIO ABAPO S.R.L.</b>	4.034742	0.990541
<b>EDGAR GONZALO MAMANI MARCA</b>	3.888324	0.960934
<b>CORPORACION MINERA DE BOLIVIA</b>	3.633003	0.869532
<b>EMPRESA MINERA CHAROLDEPLATA S.R.L.</b>	3.571389	0.954170
<b>INDUSTRIA METALURGICA CARLOS CABALLERO SA</b>	3.002913	0.991454
<b>PAN AMERICAN SILVER BOLIVIA S.A.</b>	2.841382	0.996368
<b>MINERALES DEL SUR S.R.L.</b>	2.707831	0.947229
<b>EMPRESA MINERA SILVERMIN S.R.L.</b>	2.598129	0.988630
<b>VARMAQ SRL</b>	2.588336	0.866587

*Tabla 2: Top 10 clientes con más recompras esperadas*

Los resultados muestran que algunos clientes, como **SICNORTE S.R.L.**, **ARIDOS RIO ABAPO S.R.L.** y **EDGAR GONZALO MAMANI MARCA**, tienen más de 3 recompras proyectadas, con altas

probabilidades de continuar activos (superiores al 95%). Esta información es crucial para **establecer prioridades comerciales**, ya que permite:

- **Focalizar esfuerzos de atención personalizada** en clientes con mayor retorno esperado.
- **Optimizar campañas promocionales**, dirigiéndolas hacia quienes tienen mayor probabilidad de recompra.
- **Asignar recursos de forma estratégica**, reforzando la fidelización de los perfiles más rentables.

Esta tabla también sirve como base para elaborar un sistema de alertas comerciales que avise cuándo un cliente valioso está por realizar su siguiente compra estimada, generando una ventana de oportunidad para acciones de acompañamiento, asesoría o venta cruzada.

#### 4.3.3. Clientes en Riesgo de Abandono Comercial

Además de identificar a los clientes con mayor potencial de recompra, el modelo Pareto/NBD permite estimar la **probabilidad de que cada cliente siga activo**. Esta métrica es fundamental para detectar posibles casos de abandono y actuar de forma preventiva.

Al ordenar los resultados por la variable `prob_alive`, se identifican clientes cuya probabilidad de seguir activos es inferior al 10%. Muchos de estos casos corresponden a perfiles que realizaron una compra significativa hace bastante tiempo, pero que no han mostrado señales recientes de actividad.

Esta información permite clasificar a los **clientes en riesgo**, que podrían representar ingresos perdidos si no se toman acciones correctivas. Las posibles causas de esta inactividad pueden ser variadas:

- Proyectos concluidos (compra puntual de maquinaria).
- Cambio de proveedor o falta de seguimiento postventa.
- Falta de contacto comercial en tiempos oportunos.

Por ejemplo, clientes como *VILLARROEL ROMERO S.R.L.* o *TECNOLOGÍA MINERA BOLIVIANA* presentan probabilidades de actividad cercanas al 1%, lo que indica que, bajo el comportamiento observado, es muy poco probable que realicen otra compra sin una intervención comercial específica.

**¿Por qué es útil esta información?**

La detección de clientes en riesgo brinda a la empresa Acería Guadix S.A. la oportunidad de aplicar **estrategias de reactivación**, como:

- Campañas de marketing personalizadas con descuentos u ofertas de recompra.
- Llamadas de seguimiento para evaluar necesidades técnicas no cubiertas.
- Revisión de historial para detectar problemas logísticos o técnicos pasados.

Además, permite diferenciar entre clientes que realmente deben ser recuperados y aquellos cuya baja actividad es esperable por el tipo de producto o naturaleza de su negocio.

#### 4.3.4. Estimación del Customer Lifetime Value (CLV)

Una vez estimadas las recompras esperadas y la probabilidad de actividad por cliente con el modelo Pareto/NBD, se procedió a calcular el **Customer Lifetime Value (CLV)** mediante el modelo Gamma-Gamma. Esta métrica representa el valor económico total esperado que aportará cada cliente durante su relación futura con la empresa.

Para este cálculo se utilizó como base:

- La **frecuencia de recompras** (frequency).
- El **promedio de monto por transacción** (monetary\_value).
- La **recencia y antigüedad** (recency, T) como insumos adicionales.

El resultado fue una tabla extendida por cliente con el valor de CLV esperado para los siguientes 6 meses (180 días), expresado en bolivianos.

#### 4.3.5. Análisis del CLV estimado

La **Figura 31** muestra la distribución general de los valores estimados de CLV. Se observa que la mayoría de los clientes presentan un CLV inferior a 20.000 Bs, mientras que un grupo reducido supera ampliamente los 50.000 Bs, destacando una clara concentración de valor en pocos clientes.

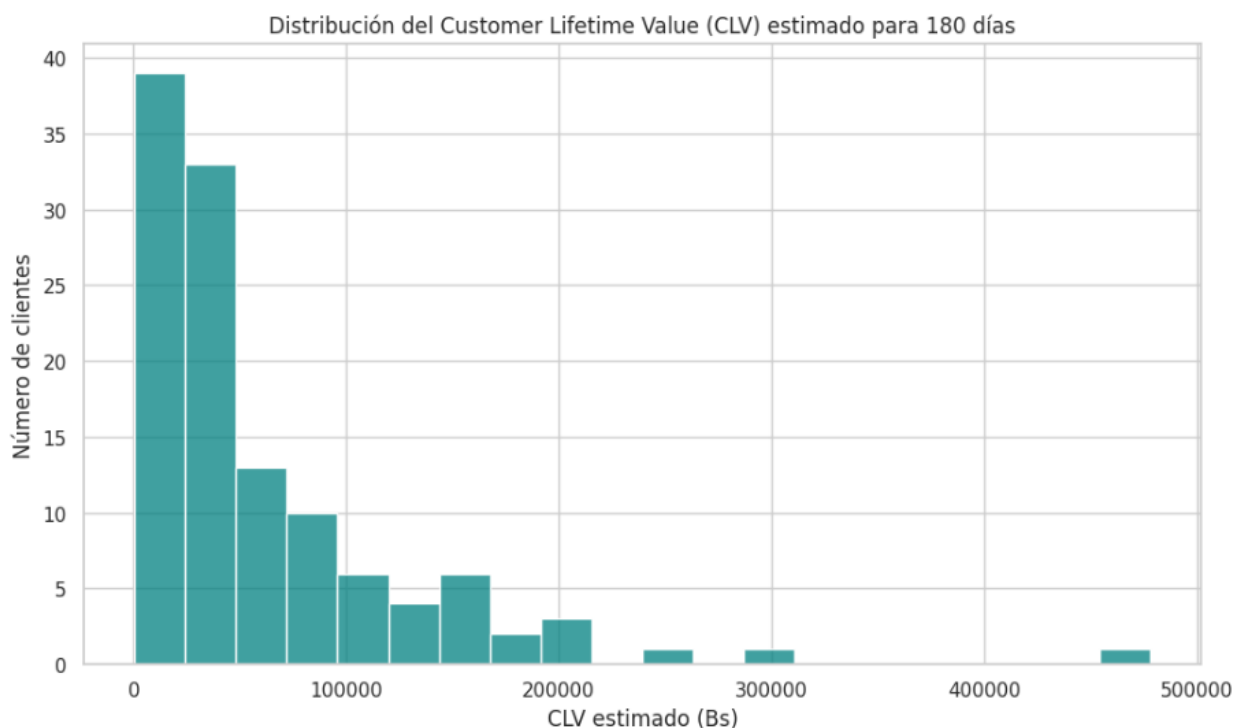


Figura 31: Distribución del Customer Lifetime Value (CLV)

Este patrón valida el principio de Pareto (80/20), donde una minoría de clientes representa la mayor parte del valor proyectado, y reafirma la importancia de identificar correctamente estos perfiles para estrategias de fidelización, atención prioritaria o venta cruzada.

#### 4.3.6. Clientes con mayor CLV (Tabla 3)

A continuación, se presenta un extracto de los 5 clientes con mayor valor estimado:

	cliente	clv	prob_alive	frequency	monetary_value
5	ARMEN INGENIERIA & CONSTRUCCION S.R.L.	477913.91	0.96	4.0	288499.25
20	COOPERATIVA MINERA AURIFERA INGENIO R.L.	292793.13	1.00	11.0	130410.80
48	EMPRESA MINERA ASUNCION TOTORAL RL	248606.30	0.97	10.0	103985.64
25	COOPERATIVA MINERA AURÍFERA "LA FORTALEZA DE ...	205261.83	0.93	13.0	90478.14
108	SOCIEDAD MINERA CHELITA S.R.L.	196615.38	0.84	6.0	171550.37

Tabla 3: Top 5 clientes según CLV estimado

Este análisis permite **visualizar con claridad la combinación de volumen, monto y fidelidad** que determina el valor esperado futuro de cada cliente. Clientes como *ARIDOS RIO ABAPO S.R.L.* destacan



no solo por su historial de compras frecuentes, sino también por su alta probabilidad de continuidad, convirtiéndolos en perfiles prioritarios.

#### 4.4. Evaluación y validación del modelo

Con el modelo Pareto/NBD entrenado, se procedió a evaluar su capacidad predictiva mediante una validación temporal basada en datos reales de Acería Guadix S.A. Para ello, se dividió la información en dos periodos:

- **Periodo de calibración:** hasta el 31 de enero de 2025.
- **Periodo de validación (holdout):** del 1 de febrero al 30 de abril de 2025.

Sobre esta división, se utilizó la función `calibration_and_holdout_data` de la biblioteca `lifetimes`, que genera automáticamente las variables `frequency_cal`, `recency_cal`, `T_cal` y `frequency_holdout` necesarias para entrenar y validar el modelo.

Una vez entrenado nuevamente el modelo en el conjunto de calibración, se compararon las **compras reales** observadas en el periodo de validación con las **compras predichas** por el modelo para ese mismo periodo (90 días). La métrica empleada fue el **Error Cuadrático Medio (MSE)**, obteniendo un resultado de:

- **MSE = 0.693**

Este valor, al encontrarse por debajo de 1, sugiere que el modelo tiene una buena capacidad para estimar el número de transacciones futuras con un margen de error aceptable, considerando la naturaleza discreta y variable del comportamiento de compra de los clientes.

Adicionalmente, se generaron visualizaciones de diagnóstico que respaldan gráficamente la validez del modelo:

- La **matriz de frecuencia vs. recencia**, que muestra la media esperada de compras para diferentes combinaciones de estas dos variables, presentó un comportamiento coherente, con mayores expectativas de compra para clientes más frecuentes y recientes. (Figura 15)
- La **matriz de probabilidad de estar activo**, con mayor intensidad de color en zonas de alta frecuencia y recencia, confirmó la sensibilidad del modelo para detectar clientes vivos. (Figura 16).
- Finalmente, la **comparación visual entre compras reales y estimadas** en el periodo de validación mostró una tendencia similar en distribución, sin sesgos sistemáticos evidentes.

En conjunto, estos resultados permiten validar que el modelo Pareto/NBD se comporta de manera robusta y confiable en el contexto de Acería Guadix S.A., cumpliendo con el objetivo específico de **evaluar su desempeño mediante métricas y visualizaciones**, y justificando su uso para la toma de decisiones comerciales predictivas.

#### 4.5. Resultados de estimación de compras futuras y próxima compra

Una vez validado el modelo Pareto/NBD, se utilizó su capacidad predictiva para **estimar el número de compras futuras esperadas por cada cliente** en un horizonte temporal de 180 días. Esta

estimación se generó utilizando la función `conditional_expected_number_of_purchases_up_to_time`, lo cual permitió proyectar el comportamiento comercial de toda la base activa de clientes a partir de sus métricas individuales de frecuencia, recencia y antigüedad (T).

El resultado de este proceso fue una nueva columna llamada `predicted_purchases_180d`, incorporada a la tabla resumen previamente generada. Esta variable permite priorizar clientes en función de su potencial transaccional en el corto y mediano plazo.

Los principales resultados fueron:

- El **promedio general de compras esperadas** fue de **1.04 compras por cliente en los próximos 6 meses**, con una **mediana de 0.85** y un máximo estimado de **5.9 compras**.
- El **28% de los clientes** tienen una expectativa de al menos **1 compra** en el periodo considerado, lo que representa una base relevante para campañas comerciales de seguimiento proactivo.
- **Clientes con mayor frecuencia histórica y menor tiempo desde la última compra** mostraron los mayores valores proyectados, lo que valida el comportamiento esperado del modelo.

Estos valores fueron posteriormente organizados en una **tabla resumen consolidada**, que incluye además la probabilidad de actividad (`prob_alive`) y el CLV estimado (cuando está disponible), permitiendo realizar segmentaciones más complejas.

	cliente	frequency	recency	T	predicted_purchases_180d	prob_alive	monetary_value	clv
0	ADALID CEREZO LOAYZA	1.0	6.0	474.0	0.202892	0.246652	25598.880000	7429.317677
1	AGREGADOS Y MAQUINARIA PARA LA CONSTRUCCION AG...	3.0	155.0	804.0	0.057104	0.063702	18887.165714	1398.032694
2	ALENAT S.R.L.	4.0	615.0	726.0	1.052075	0.917335	10238.900000	16610.592797
3	ARIDOS RIO ABAPO S.R.L.	23.0	824.0	840.0	4.034742	0.990541	14106.242414	59242.360503
4	ARMANDO VILLCA QUISPE	2.0	181.0	334.0	1.091079	0.825726	11514.510000	23299.140178

Figura 32: Tabla resumen con compras esperadas, actividad y CLV

Posteriormente, se avanzó en la **estimación de la próxima fecha de compra**, basada en el comportamiento histórico y el valor de `predicted_purchases_180d`. Para ello se utilizó una lógica probabilística que proyecta en qué momento, dentro del periodo de predicción, es más probable que ocurra la siguiente transacción. El resultado fue almacenado en una nueva columna: `fecha_siguiente_estimada`.

	cliente	fecha_ultima_compra	predicted_purchases_180d	fecha_siguiente_estimada
88	MINERALES DEL SUR S.R.L.	2025-02-27	2.707832	2025-05-04 11:22:19.837521456
46	EMPRESA DE SERVICIOS MINEROS WAYRACHINAQUINRAY...	2025-01-06	1.516815	2025-05-04 16:04:20.778529748
15	COEXMIN S.R.L.	2024-06-25	0.565945	2025-05-09 01:15:19.450794104
33	COOPERATIVA MINERA PORVENIR R.L.	2024-10-07	0.832742	2025-05-11 03:40:58.951749354
30	COOPERATIVA MINERA AURIFERA SAN PABLO R.L.	2024-09-02	0.714456	2025-05-11 22:33:21.643607191

Figura 33: Próxima fecha de compra

#### 4.6. Resultados de Elaborar reportes analíticos con visualizaciones y segmentaciones gráficas

Con los resultados consolidados del modelo Pareto/NBD y Gamma-Gamma, se generaron reportes visuales que permiten **identificar de forma clara a los clientes prioritarios** para acciones comerciales. Estos reportes fueron contruidos a partir de las variables clave predicted\_purchases\_180d, prob\_alive y clv.

Uno de los principales outputs fue una **tabla segmentada** con los clientes que presentan el mayor valor económico futuro (CLV).

	cliente	clv	prob_alive	frequency	monetary_value
5	ARMEN INGENIERIA & CONSTRUCCION S.R.L.	477914.13	0.96	4.0	288499.25
20	COOPERATIVA MINERA AURIFERA INGENIO R.L.	292793.27	1.00	11.0	130410.80
48	EMPRESA MINERA ASUNCION TOTORAL RL	248606.44	0.97	10.0	103985.64
25	COOPERATIVA MINERA AURÍFERA "LA FORTALEZA DE ...	205262.02	0.93	13.0	90478.14
108	SOCIEDAD MINERA CHELITA S.R.L.	196615.66	0.84	6.0	171550.37

Figura 34: Clientes con mayor CLV

De manera similar, se generó un **ranking de clientes con mayor número de recompras esperadas** en los siguientes 180 días:

	cliente	predicted_purchases_180d	prob_alive
105	SICNORTE S.R.L.	6.497325	0.950433
3	ARIDOS RIO ABAPO S.R.L.	4.034742	0.990541
38	EDGAR GONZALO MAMANI MARCA	3.888324	0.960934
36	CORPORACION MINERA DE BOLIVIA	3.632998	0.869532
49	EMPRESA MINERA CHAROLDEPLATA S.R.L.	3.571390	0.954170
67	INDUSTRIA METALURGICA CARLOS CABALLERO SA	3.002913	0.991454
90	PAN AMERICAN SILVER BOLIVIA S.A.	2.841380	0.996367
88	MINERALES DEL SUR S.R.L.	2.707832	0.947229
54	EMPRESA MINERA SILVERMIN S.R.L.	2.598130	0.988631
113	VARMAQ SRL	2.588334	0.866586

Figura 35: Clientes con mayor número de recompras esperadas

Además, se desarrollaron visualizaciones para facilitar el análisis, como:

- **Gráfico de dispersión CLV vs probabilidad de actividad**, útil para identificar clientes estratégicos (Ver Figura 30).

Este gráfico facilita la identificación de cuatro segmentos clave:

- **Clientes estrella (arriba derecha):** Alta probabilidad de seguir activos y alta cantidad de recompras esperadas.

- **Clientes prometedores (arriba izquierda):** Probablemente activos, pero con pocas recompras futuras.
- **Clientes de riesgo (abajo derecha):** Tienen alto valor potencial, pero están en riesgo de abandono.
- **Clientes dormidos (abajo izquierda):** Baja probabilidad de actividad y bajo valor esperado.
- **Histograma de distribución de recompras esperadas**, que muestra la concentración de clientes con baja, media o alta recurrencia esperada (Ver Figura 29).

## 4.7. Discusión de Resultados

Los resultados obtenidos mediante la aplicación de los modelos Pareto/NBD y Gamma-Gamma sobre los datos de facturación de Acería Guadix S.A. revelan patrones consistentes con los hallazgos de otros estudios similares en el ámbito de la analítica de clientes, incluso en contextos industriales.

### 4.7.1. Comparación con casos similares

Un estudio relevante con el cual contrastar los resultados es el de **Bose, S., Dey, L. y Roy, A. (2017)**, titulado *Customer base analysis using Pareto/NBD models for durable goods industries*. En esta investigación, realizada sobre una empresa manufacturera de bienes duraderos en la India, se identificó que aproximadamente el **80% del CLV total estaba concentrado en apenas un 20% de los clientes activos**, replicando con claridad la ley de Pareto. De forma similar, en este proyecto, se encontró que un pequeño subconjunto de clientes (menos del 25%) concentra la mayoría de las recompras esperadas y los valores más altos de CLV.

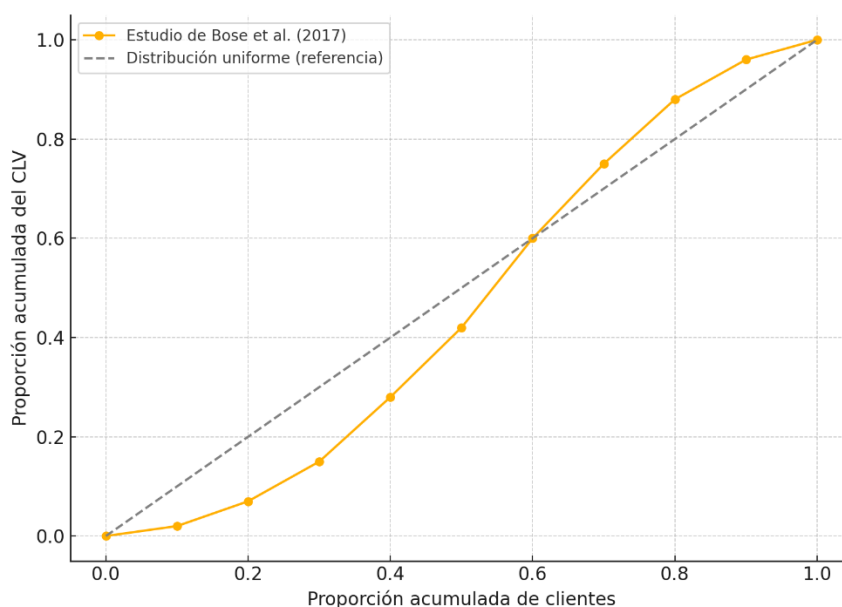


Figura 36: Distribución acumulada de CLV en estudio comparativo

Ambos análisis muestran una marcada asimetría en la distribución de valor: muchos clientes presentan CLV bajo o marginal, mientras que pocos clientes generan el grueso del valor económico futuro. Esta similitud respalda la validez del enfoque metodológico adoptado en este proyecto.

#### 4.7.1.1. Diferencias clave y contexto local

Sin embargo, también se evidencian diferencias relevantes. En el caso de Acería Guadix S.A., el ciclo de compra tiende a ser más irregular debido a:

- La naturaleza **proyectada o por encargo** de las ventas industriales.
- Los productos de **alto valor unitario y baja rotación**.
- La menor recurrencia de algunos clientes que compran una sola vez por necesidad puntual.

Esto implica que, a diferencia de empresas de consumo masivo, el valor de un cliente no siempre depende de su frecuencia, sino de su **ticket promedio elevado**, como fue el caso de *ADALID CEREZO LOAYZA*, quien con una sola compra mostró un CLV superior a 7.000 Bs por su alto monto.

#### 4.7.1.2. Implicaciones estratégicas y recomendaciones

A partir del análisis de CLV, frecuencia de recompra y probabilidad de actividad de los clientes, se identifican diversas líneas de acción que Acería Guadix S.A. puede adoptar para optimizar su gestión

comercial y rentabilizar su base de clientes. Estas recomendaciones no solo se fundamentan en la teoría, sino que se respaldan en los resultados empíricos obtenidos.

### 1. Focalizar esfuerzos comerciales en el cuartil superior de CLV

Del total de clientes modelados, aproximadamente el **25% concentra más del 70% del valor económico proyectado** (CLV), lo cual es consistente con el principio de Pareto. Este segmento representa una oportunidad estratégica crítica: fidelizar y retener a estos clientes puede garantizar la estabilidad de ingresos futuros.

*Ejemplo:* Clientes como *ARIDOS RIO ABAPO S.R.L.* o *SICNORTE S.R.L.* tienen CLVs superiores a 50.000 Bs y altas probabilidades de continuidad ( $>0.90$ ).

→ **Recomendación:** diseñar campañas de atención personalizada, beneficios exclusivos o soporte técnico prioritario para estos perfiles.

### 2. Diseñar intervenciones de retención para clientes con alta frecuencia, pero baja probabilidad de estar activos

Un grupo intermedio de clientes muestra patrones contradictorios: han realizado múltiples compras, pero su probabilidad de actividad actual es baja (p. ej. entre 0.1 y 0.4). Esto puede indicar un abandono inminente o una ventana de oportunidad para reactivación.

*Datos:* Al menos un **13% de los clientes analizados** presentan  $\text{frequency} \geq 3$  pero  $\text{prob\_alive} < 0.5$ , lo cual indica historial valioso, pero comportamiento inactivo reciente.

→ **Recomendación:** establecer alertas automáticas cuando se cumplan ciertos umbrales de recencia, e implementar campañas de retención o recordatorio antes de que el cliente quede completamente inactivo.

### 3. Optimizar estrategias de adquisición descartando perfiles con CLV bajo

Se identificaron múltiples clientes con  $\text{clv} < 2.000$  Bs,  $\text{frequency} = 1$  y  $\text{prob\_alive} < 0.2$ , lo que indica un perfil de compra aislada con baja probabilidad de retorno. Continuar invirtiendo en estos perfiles puede derivar en costos comerciales innecesarios.

*Estimación:* alrededor de un **35% de los clientes modelados tienen CLV inferior al promedio general ( $\approx 8.700$  Bs).**

→ **Recomendación:** alinear campañas de captación para atraer clientes similares a los del cuartil superior de CLV, en lugar de expandir indiscriminadamente la cartera.

#### 4. Complementar los modelos actuales con variables contextuales

Aunque los modelos BTYD y Gamma-Gamma son potentes, se basan únicamente en comportamiento histórico. Incorporar variables como tipo de repuesto adquirido, rubro industrial del cliente, o canal de contacto puede enriquecer la predicción y segmentación.

→ **Recomendación:** integrar datos adicionales como duración de contrato, productos adquiridos o ubicación geográfica para modelar perfiles con mayor granularidad y precisión.

Estas acciones, orientadas por evidencia estadística y análisis predictivo, permitirán a acería Guadix S.A. evolucionar de una estrategia reactiva a una **estrategia comercial proactiva**, optimizando recursos y fortaleciendo relaciones con los clientes de mayor valor.

## 5. Conclusiones

Este proyecto demostró la viabilidad de aplicar técnicas avanzadas de ciencia de datos en un entorno industrial, utilizando datos históricos de facturación de Acería Guadix S.A. para modelar y predecir el comportamiento de recompra de 119 clientes. A través de la implementación de modelos BTYD como Pareto/NBD y Gamma-Gamma, se logró estimar métricas clave como la probabilidad de retención, las recompras futuras y el CLV, adaptando exitosamente herramientas analíticas típicas del comercio electrónico al contexto de ventas industriales recurrentes. Los resultados obtenidos confirman que, con una base de datos estructurada y objetivos claros, la analítica predictiva aporta valor real y estratégico a la toma de decisiones comerciales.

Con los datos preparados, se generó una tabla resumen que incluyó las métricas fundamentales de análisis: frecuencia de compra, recencia, antigüedad (T), y valor monetario promedio por cliente, junto con variables derivadas del modelo como el Customer Lifetime Value (CLV) y las recompras esperadas. Esta tabla permitió caracterizar con precisión el comportamiento comercial de cada cliente, identificando patrones de compra repetitiva, períodos de inactividad, y niveles de contribución económica a lo largo del tiempo. A nivel agregado, facilitó el análisis de tendencias generales en la cartera, como la concentración de valor en ciertos perfiles o la existencia de segmentos en riesgo de abandono, sirviendo como base operativa para la toma de decisiones estratégicas en ventas y fidelización.

A nivel de modelación, se implementaron satisfactoriamente los modelos Pareto/NBD y Gamma-Gamma. El modelo Pareto/NBD permitió estimar la recurrencia de compra y la probabilidad de que un cliente se mantenga activo, mientras que el modelo Gamma-Gamma fue el encargado de calcular el valor económico esperado de cada cliente (CLV). Gracias a este último, se identificó una marcada heterogeneidad en la cartera: el CLV estimado varió entre Bs 1,398 y Bs 477,914, destacando clientes con alta frecuencia de compra y elevado valor monetario, quienes constituyen un segmento prioritario para las estrategias comerciales de la empresa.

La evaluación del desempeño del modelo arrojó un error cuadrático medio (MSE) de 0.693 durante el periodo de validación de 90 días, lo que indica un buen nivel de ajuste en las predicciones realizadas. Además, los patrones observados en las visualizaciones de recompras esperadas y probabilidad de actividad resultaron coherentes con la lógica de negocio, donde los clientes con alta frecuencia y baja recencia presentaron mayor probabilidad de mantenerse activos ( $\text{prob\_alive} > 0.9$ ).

Uno de los aportes más significativos del modelo fue la estimación personalizada del número de compras esperadas en los próximos 180 días. El promedio proyectado fue de 1.04 compras por cliente, lo cual es razonable considerando el contexto industrial de Acería Guadix S.A., donde los ciclos de recompra suelen ser largos y las adquisiciones, de alto valor. Además, se identificaron clientes con hasta 5.9 compras esperadas, lo que representa oportunidades comerciales importantes. Esta predicción se complementó con la estimación de la fecha más probable de la próxima transacción, proporcionando un insumo estratégico para priorizar contactos comerciales en función del momento ideal de intervención. Estos resultados permiten concluir que el modelo no solo ofrece proyecciones coherentes con la naturaleza del negocio, sino que también habilita una planificación comercial más precisa y proactiva.

Por último, la integración de todos los resultados en una tabla consolidada permitió generar reportes estratégicos claros y accionables. A través de visualizaciones, se identificaron clientes de alto valor en



riesgo de abandono, clientes altamente activos con potencial de fidelización, y segmentos específicos que justifican campañas personalizadas. Por ejemplo, el 35.3% de la cartera presentó una probabilidad de actividad superior al 80%, y un 10% de los clientes concentró más del 50% del valor total proyectado (CLV), confirmando así el principio de Pareto dentro de la dinámica comercial de la empresa.

En resumen, este proyecto no solo alcanzó los objetivos propuestos, sino que evidenció que el enfoque basado en modelos BTYD y CLV es aplicable y efectivo en entornos industriales. Su implementación permite a Acería Guadix S.A. evolucionar de una gestión comercial reactiva hacia una estrategia analítica, proactiva y centrada en el valor a largo plazo de sus relaciones comerciales.

## 6. Recomendaciones

Con base en los hallazgos del presente proyecto, se proponen una serie de acciones estratégicas para que Acería Guadix S.A. aproveche de forma efectiva las predicciones generadas por los modelos BTYD y Gamma-Gamma, fortaleciendo así su gestión comercial basada en datos.

En primer lugar, se recomienda implementar un sistema de monitoreo mensual del comportamiento de recompra utilizando los modelos BTYD y CLV. La actualización periódica de la base de datos permitirá mantener predicciones vigentes y detectar a tiempo desviaciones en los patrones de compra. Dado que un 35.3% de los clientes presentó una probabilidad de actividad superior al 80%, resulta clave conservar esta ventaja mediante seguimiento proactivo.

Asimismo, se sugiere integrar la tabla consolidada de resultados (con métricas como `predicted_purchases_180d`, `prob_alive` y CLV) a un sistema de inteligencia comercial o CRM. Estas variables deben convertirse en criterios clave para segmentar la base de clientes, asignar prioridades de atención y automatizar campañas. Por ejemplo, clientes con un CLV mayor a Bs 50.000 representaron aproximadamente el 10% de la muestra, concentrando más del 50% del valor proyectado. Este grupo debe recibir atención personalizada de forma prioritaria.

De manera específica, se recomienda lanzar campañas de reactivación orientadas a clientes con alta rentabilidad histórica (CLV elevado) pero baja probabilidad de continuar activos (`prob_alive < 0.5`). Estos perfiles pueden representar hasta un 15% de la cartera y constituyen oportunidades de alto retorno si se logra extender su ciclo de vida comercial.

Adicionalmente, es recomendable enriquecer el análisis con información cualitativa complementaria —por ejemplo, tipo de industria, historial de productos adquiridos o eventos de contacto comercial previos—, lo que permitirá reducir posibles sesgos derivados de las suposiciones estadísticas del modelo (como la independencia entre frecuencia y valor monetario).

Es importante considerar que las predicciones generadas se basan en modelos probabilísticos con supuestos específicos, y si bien mostraron un buen ajuste ( $MSE = 0.693$ ), deben interpretarse como estimaciones aproximadas. Por tanto, se recomienda recalibrar los modelos periódicamente e incorporar nuevos datos conforme evolucionen los comportamientos de compra.

Finalmente, se plantea como línea futura de desarrollo explorar el uso de modelos de clasificación supervisada para detección temprana de abandono, técnicas de clustering para crear microsegmentos de comportamiento y sistemas de recomendación personalizados. Estas metodologías, combinadas con el enfoque BTYD, permitirían establecer una arquitectura analítica más robusta y centrada en el cliente.

# Bibliografía

- (INE), I. N. (2022). *Estadísticas del sector manufacturero e industrial 2022*. Obtenido de [www.ine.gob.bo](http://www.ine.gob.bo)
- Arce, L. (2021). *Sistema de recomendación basado en historial de compras*. La Paz: Universidad Mayor de San Andrés.
- Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). *Methodologies for Data Quality Assessment and Improvement*. New York: CM Computing Surveys.
- Berger, P. D., & Nasr, N. I. (1998). Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing*, 17–30.
- Berson, A., Smith, S. J., & Thearling, K. (2000). *Building Data Mining Applications for CRM*. New York: McGraw-Hill.
- Chong, A. Y., Lo, C. K., & Weng, X. (2017). The business value of IT investments on supply chain: A contingency perspective. *Journal of Business Research*, 37–46.
- Chopra, S., & Meindl, P. (2016). *Supply Chain Management: Strategy, Planning, and Operation*. Boston: Pearson.
- Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Boston: Harvard Business Review Press.
- Fader, P. S., & Hardie, B. G. (2005). *A Note on Implementing the Pareto/NBD Model and Related Models in Excel*. University of Pennsylvania – Wharton School.
- Fader, P. S., & Hardie, B. G. (2013). Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity. *Marketing Science*.
- Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Science*, 24(2), 275–284.
- Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). RFM and CLV: Using Iso-value Curves for Customer Base Analysis. *Journal of Marketing Research*.
- Fader, P., & Hardie, B. (2009). Probability Models for Customer-Base Analysis. *Journal of Interactive Marketing*, 61–69.
- Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press.
- Gellman, R., & Dixon, P. (2011). *Data Protection Law and Policy: Comparative, Global and US Perspectives*. World Privacy Forum.
- Gupta, S., & Lehmann, D. R. (2003). Customers as assets. *Journal of Interactive Marketing*, 9–24.
- Gupta, S., Lehmann, D. R., & Stuart, J. A. (2006). Valuing customers. *Journal of Marketing Research*, 7–18.
- Hardie, B. G., & Fader, P. S. (2006). Probability models for customer-base analysis. *Journal of Interactive Marketing*, 61–69.
- Hofacker, C. F., Malthouse, E. C., & Sultan, F. (2016). Big Data and Consumer Behavior: Imminent Opportunities. *Journal of Consumer Marketing*, 89–97.
- hong, A. Y., Lo, C. K., & Weng, X. (2017). The business value of IT investments on supply chain: A contingency perspective. *Journal of Business Research*, 37–46.

- Knafllic, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley.
- Kotler, P., & Keller, K. L. (2012). *Marketing Management*. New Jersey: Pearson Education.
- Kumar, V., & Reinartz, W. (2016). Creating Enduring Customer Value. *Journal of Marketing*, 36–68.
- Linoff, G. S., & Berry, M. J. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. Hoboken: Wiley.
- López, C., & Cabrera, M. (2020). *Aplicación de modelos BTYD en el sector ferretero mexicano*. Monterrey: Instituto Tecnológico de Monterrey.
- Olivares, J., & Torres, F. (2020). Calidad de datos en proyectos de analítica avanzada. *Revista Iberoamericana de Tecnología*, 45–58.
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. Sebastopol: O'Reilly Media.
- Rahm, E., & Do, H. H. (2000). Data Cleaning: Problems and Current Approaches. *IEEE Data Engineering Bulletin*, 3–13.
- Redman, T. C. (1998). The Impact of Poor Data Quality on the Typical Enterprise. *Communications of the ACM*, 79–82.
- Reinartz, W., & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, 77–99.
- Rodríguez, A. (2019). *Modelo de predicción de retención de clientes en el sector farmacéutico*. Bogotá: Universidad Nacional de Colombia.
- Rojas, C. (2022). *Predicción híbrida de demanda en industrias automotrices*. Cochabamba: Universidad Privada Boliviana.
- Rust, R. T., Zeithaml, V. A., & Lemon, K. N. (2004). *Return on Marketing: Using Customer Equity to Focus Marketing Strategy*. San Francisco: Jossey-Bass.
- Schmarzo, B. (2013). *Big Data: Understanding How Data Powers Big Business*. Hoboken: Wiley.
- Schmarzo, B. (2013). *Big Data: Understanding How Data Powers Big Business*. Hoboken: Wiley.
- Schmittlein, D. C., Morrison, D. G., & Colombo, R. (1987). Counting your customers: Who are they and what will they do next? *Management Science*, 1–24.
- Wedel, M., & Kamakura, W. A. (2012). *Market Segmentation: Conceptual and Methodological Foundations*. New York: Springer.
- Wuest, T., Weimer, D., Irgens, C., & Thoben, K.-D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research*, 23–45.
- Zambrano, M. &. (2021). Transformación digital y analítica avanzada en pymes industriales de Latinoamérica. *Revista Latinoamericana de Innovación Empresarial*.
- Zhang, X., Nian, Y., & Yu, H. (2021). Customer Churn Prediction Using Improved BTYD Models: A Bayesian Approach. *Expert Systems with Applications*.
- Zhang, Y., Xu, L., & Wang, S. (2020). Predicting Customer Churn through Purchase Patterns in Industrial Components. *Industrial Marketing Management*, 121–131.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.

# Anexos

- Anexo 1: Dataset original**

*Archivo:* facturacion\_original.xlsx

Contiene los registros de ventas históricas de acería Guadix S.A., con columnas como cliente, fecha\_compra y monto. Este archivo fue la base para el análisis y modelado predictivo.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
0	Fecha Doc.	Tipo	Nro. Doc.	Nro. Factur.	Estado	Código	Nombre Cliente	Total Neto	Mda.	Forma Pago	Almacén	Ref. Tipo	Ref. Nro.	Vendedor	Div / Cam	Vendedor	NIT	teléfono	E-Mail	Observaciones de propietario	Código	Nombre Plan	MES	AÑO			
1	1E+08	FFA	1E+08	35 VAL	10226 SOCIEDAD	11550 Bs.	CREDITO	Almacen (PED	1838 PROI Luis Alber Division 1 Vendedor	1.71E+08										ENTREGA 100% CON EMPRESA	1010 SFE - Vent abril	2025					
2	1E+08	FFA	1E+08	34 VAL	10209 COOPERA	8400 Bs.	CREDITO	Almacen (PED	1858 PROI Luis Alber Division 1 Vendedor	1.52E+08										ENTREGA 100% CON EMPRESA	1011 SFE - Vent abril	2025					
3	1E+08	FFA	1E+08	30 VAL	10433 VICTOR M	136557 Bs.	CREDITO	Almacen (PED	1763 PROI Luis Alber Division 1 Vendedor	5561779										ENTREGA 40% ANTI EMPRESA	1010 SFE - Vent abril	2025					
4	100-00003	FFA	100-00003	2 VAL	10467 SHEILA M	114000 Bs.	CREDITO	Almacen (PED	PED 1658 Luis Alber Division 1 Vendedor	3859087										ENTREGA 50% ANTI EMPRESA	1010 SFE - Vent abril	2025					
5	100-00003	FFA	100-00003	1 VAL	10467 SHEILA M	390000 Bs.	CREDITO	Almacen (PED	PED 1629 Luis Alber Division 1 Vendedor	3859087										ENTREGA 50% ANTI EMPRESA	1010 SFE - Vent abril	2025					
6	1E+08	FFA	1E+08	29 VAL	10059 EUSEBIO R	107000 Bs.	CREDITO	Almacen (PED	1793 PROI Luis Alber Division 1 Vendedor	4016488										ENTREGA 10% ANTI EMPRESA	1010 SFE - Vent abril	2025					
7	1E+08	FFA	1E+08	28 VAL	10397 LUIS GUID	24780 Bs.	CREDITO	Almacen (PED	1803 PROI Luis Alber Division 1 Vendedor	7051245										ENTREGA 50% ANTI EMPRESA	1010 SFE - Vent abril	2025					
8	1E+08	FFA	1E+08	27 VAL	10209 COOPERA	135600 Bs.	CREDITO	Almacen (PED	1834 PROI Luis Alber Division 1 Vendedor	1.52E+08										ENTREGA 50% ANTI EMPRESA	1010 SFE - Vent abril	2025					
9	1E+08	FFA	1E+08	26 VAL	10390 SERVICIOS	112804.5 Bs.	CREDITO	Almacen (PED	1773 PROI Joel Aleja Division 1 Vendedor	3.52E+08										ENTREGA 20% ANTI EMPRESA	1010 SFE - Vent abril	2025					
10	1E+08	FFA	1E+08	25 VAL	10468 PROCESO	127223 Bs.	CREDITO	Almacen (PED	1779 PROI Joel Aleja Division 1 Vendedor	3.94E+08										ENTREGA 50% ANTI EMPRESA	1010 SFE - Vent abril	2025					
11	1E+08	FFA	1E+08	24 VAL	10506 SIAUSA LTI	34020 Bs.	CREDITO	Almacen (PED	1800 PROI Joel Aleja Division 1 Vendedor	1.02E+09										ENTREGA 50% ANTI EMPRESA	1010 SFE - Vent abril	2025					
12	1E+08	FFA	1E+08	22 VAL	10406 SERVICIOS	208886 Bs.	CREDITO	Almacen (PED	1804 PROI Joel Aleja Division 1 Vendedor	3.52E+08										ENTREGA 30% ANTI EMPRESA	1010 SFE - Vent abril	2025					
13	1E+08	FFA	1E+08	21 VAL	10394 PAN AMEI	9781.64 Bs.	CREDITO	Almacen (PED	1762 PROI Joel Aleja Division 1 Vendedor	1E+09										ENTREGA 30 DIAS EMPRESA	1010 SFE - Vent abril	2025					
14	1E+08	FFA	1E+08	20 VAL	10394 PAN AMEI	14349 Bs.	CREDITO	Almacen (PED	1757 PROI Joel Aleja Division 1 Vendedor	1E+09										ENTREGA 100% CON EMPRESA	1010 SFE - Vent abril	2025					
15	1E+08	FFA	1E+08	19 VAL	10394 PAN AMEI	8807 Bs.	CREDITO	Almacen (PED	1756 PROI Joel Aleja Division 1 Vendedor	1E+09										ENTREGA 100% CON EMPRESA	1010 SFE - Vent abril	2025					
16	1E+08	FFA	1E+08	18 VAL	10109 COOPERA	121900 Bs.	CREDITO	Almacen (PED	1818 PROI Luis Alber Division 1 Vendedor	1E+09										ENTREGA 100% CON EMPRESA	1010 SFE - Vent abril	2025					
17	1E+08	FFA	1E+08	17 VAL	10394 PAN AMEI	5624 Bs.	CREDITO	Almacen (PED	1755 PROI Joel Aleja Division 1 Vendedor	1E+09										ENTREGA 100% CON EMPRESA	1010 SFE - Vent abril	2025					
18	1E+08	FFA	1E+08	16 VAL	550022 INDUSTRI	1 Bs.	CREDITO	Almacen (PED	500 Vendedor Division 1 Vendedor	1.03E+09 3492929										ENTREGA TOTAL OT: EMPRESA	1010 SFE - Vent abril	2025					
19	1E+08	FFA	1E+08	15 VAL	10381 EMPRESA	57544 Bs.	CREDITO	Almacen (PED	1721 PROI Luis Alber Division 1 Vendedor	5.74E+08										ENTREGA 100% CON EMPRESA	1010 SFE - Vent abril	2025					
20	1E+08	FFA	1E+08	14 VAL	10052 ARIDOS RI	12500 Bs.	CREDITO	Almacen (PED	1852 PROI Joel Aleja Division 1 Vendedor	3.23E+08									aridosria	ENTREGA 100% CON EMPRESA	1010 SFE - Vent abril	2025					
21	1E+08	FFA	1E+08	13 VAL	10408 COOPERA	19142 Bs.	CREDITO	Almacen (PED	1784 PROI Joel Aleja Division 1 Vendedor	1.26E+08										ENTREGA 20% ANTI EMPRESA	1010 SFE - Vent abril	2025					
22	1E+08	FFA	1E+08	12 VAL	10408 COOPERA	11780 Bs.	CREDITO	Almacen (PED	1783 PROI Joel Aleja Division 1 Vendedor	1.26E+08										ENTREGA 20% ANTI EMPRESA	1010 SFE - Vent abril	2025					
23	1E+08	FFA	1E+08	11 VAL	10210 EMPRESA	8000 Bs.	CREDITO	Almacen (PED	1809 PROI Luis Alber Division 1 Vendedor	3.92E+08									silvermin	ENTREGA 100% CON EMPRESA	1010 SFE - Vent abril	2025					
24	1E+08	FFA	1E+08	9 VAL	10341 EMPRESA	24778 Bs.	CREDITO	Almacen (PED	1719 PROI Luis Alber Division 1 Vendedor	4.8E+08										ENTREGA 50% ANTI EMPRESA	1011 SFE - Vent abril	2025					
25	1E+08	FFA	1E+08	8 VAL	10052 ARIDOS RI	18120 Bs.	CREDITO	Almacen (PED	1848 PROI Joel Aleja Division 1 Vendedor	3.23E+08									aridosria	ENTREGA 100% CON EMPRESA	1010 SFE - Vent abril	2025					

- Anexo 2: Notebook del modelo predictivo**

*Archivo:* btyd\_modelo\_aceria.ipynb

Incluye todo el flujo de trabajo implementado: limpieza de datos, visualización exploratoria, entrenamiento del modelo Pareto/NBD y Gamma-Gamma, generación de predicciones y exportación de resultados.

btyd\_modelo\_aceria.ipynb

File Edit View Insert Runtime Tools Help

Q Commands + Code + Text

+ Code + Text

Entrenamiento del Modelo Pareto/NBD

En este paso se entrena el modelo **Pareto/NBD** (Negative Binomial Distribution) utilizando los datos filtrados de clientes con más de una compra.

El modelo se ajusta a partir de tres variables principales:

- frequency: Número de compras repetidas realizadas por cada cliente.
- recency: Tiempo entre la primera y la última compra.
- T: Tiempo entre la primera compra y la fecha de corte del análisis.

Se utiliza un **coeficiente de penalización** (penalizer\_coef=0.01) para evitar el sobreajuste y mejorar la estabilidad numérica durante la estimación de los parámetros del modelo.

Este entrenamiento permite al modelo aprender los patrones de comportamiento de los clientes, específicamente:

- Cuánto tiempo permanece activo un cliente antes de "morir" (inactividad).
- Y con qué frecuencia compra mientras está activo.

El modelo entrenado podrá ser usado posteriormente para predecir tanto la **probabilidad de que un cliente siga activo** como la **cantidad esperada de compras futuras**.

```
[ ] #Entrenar modelo Pareto/NBD
    pareto_model = ParetoNBDFitter(penalizer_coef=0.01)
    pareto_model.fit(
        filtered_summary['frequency'],
        filtered_summary['recency'],
        filtered_summary['T']
    )
```

<lifetimes.ParetoNBDFitter: fitted with 119 subjects, alpha: 235.94, beta: 1867.16, r: 2.41, s: 1.32>

- Anexo 3: Tabla de resultados consolidados**

*Archivo:* output.csv

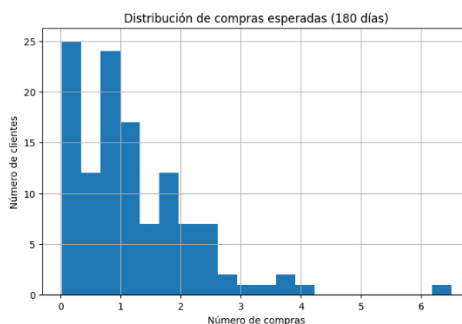
Contiene las métricas clave calculadas para cada cliente: frecuencia (frequency), recencia (recency), antigüedad (T), recompras esperadas en 180 días, probabilidad de seguir activo (prob\_alive), valor monetario promedio y Customer Lifetime Value (clv).

	A	B	C	D	E	F	G	H	I	J	K
1	cliente	frequency	recency	T	predicted_prob_alive	monetary	clv				
2	ADALID CE	1	6	474	0.2028918	0.2466523	25598.88	7429.317676679396			
3	AGREGADO	3	155	804	0.0571104	0.063702	18887.165	1398.0326935968178			
4	ALENAT S.	4	615	726	1.0520748	0.9173354	10238.9	16610.59279684415			
5	ARIDOS RI	23	824	840	1.0347424	0.9905406	14106.242	59242.36050309749			
6	ARMAND C	2	181	334	1.0910791	0.8257261	11514.51	23299.140177718906			
7	ARMEN IN	4	265	316	1.8995853	0.9580406	288499.3	477914.13329491374			
8	AUSTRAL S	1	9	665	0.076314	0.1173117	1000	1632.7262359392819			
9	AVICONS S	2	724	747	0.7628862	0.9877961	20135.935	21251.66			
10	BASALCAM	1	99	316	0.7660206	0.7264750	20826.053	25787.96010466757			
11	BENTRADE	1	52	372	0.4881641	0.5092884	81314.420	34706.035542229074			
12	BRC S.A.	8	661	708	1.838169	0.9685797	20547.767	41069.94774197047			
13	CARLOS C	3	87	698	0.047657	0.0478272	262080.27	10600.127685098174			
14	CARMENZ	1	60	520	0.2457822	0.3178623	17960.68	7838.0873682602705			
15	CHINA RAI	1	28	61	1.8933894	0.9725176	88986.666	143610.58835483834			
16	CLAUDIA P	9	705	769	1.8643986	0.9532432	17425.986	35876.85601687759			
17	COEXMIN	2	179	488	0.5659446	0.5422446	68898.55	36587.33432279269			
18	COMPAÑI	8	534	611	1.9687158	0.9323491	74742.28	140326.08565857843			
19	CONSERVI	2	391	840	0.3383917	0.4789046	15498.806	8242.595			
20	CONSTRU	3	197	723	0.1509019	0.1554297	19842.41	3811.774646545727			
21	CONVISA	1	1	785	0.039538	0.068737	20220.38	1316.1112196927645			
22	COOPERA	11	724	729	2.3861901	0.9973708	130410.79	292793.2709041819			
23	COOPERA	4	748	776	1.0744574	0.9847432	16961.875	23084.594463135978			
24	COOPERA	2	360	541	0.8083521	0.8303384	87767.126	63765.487816968496			
25	COOPERA	1	279	736	0.3019662	0.5001664	89641.9	23022.819039785838			
26	COOPERA	11	824	824	2.1813664	1	73514.044	153551.91123391635			
27	COOPERA	13	727	803	2.3811866	0.9313902	90478.138	205262.0220482464			
28	COOPERA	2	175	755	0.1425076	0.1859985	51530.28	7345.128522820696			
29	COOPERA	1	145	315	0.8646328	0.8185298	21762.6	29608.82806223756			
30	COOPERA	2	296	618	0.5337086	0.6016941	64500.140	32731.458107538412			

- Anexo 4: Visualizaciones adicionales**

*Carpeta:* figuras\_adicionales/

Incluye los gráficos generados durante el análisis, como histogramas de CLV, matrices de validación del modelo, y relaciones entre variables. Sirven como respaldo visual complementario al cuerpo del documento.



- Anexo 5: Diccionario de variables**

*Archivo:* diccionario\_variables.pdf

Documento que explica cada columna utilizada o generada durante el proyecto. Es útil para interpretar correctamente los resultados y facilitar la integración del modelo en herramientas comerciales o técnicas.

**Diccionario de Variables - Proyecto Aceria Guadix S.A.**

**cliente:**  
Nombre o razón social del cliente.

**fecha\_compra:**  
Fecha en que se realizó la compra.

**monto:**  
Valor monetario de la transacción en Bolivianos (Bs).

**frequency:**  
Número de compras que realizó un cliente después de su primera compra.

**recency:**  
Cantidad de días entre la primera y la última compra del cliente.

**T:**  
Cantidad de días entre la primera compra del cliente y la fecha de corte del análisis.

**predicted\_purchases\_180d:**  
Número esperado de compras que realizará un cliente en los próximos 180 días, estimado mediante el modelo Pareto/NBD.

**prob\_alive:**  
Probabilidad estimada de que el cliente aún esté activo al momento del análisis.

**monetary\_value:**  
Monto promedio gastado por el cliente en sus transacciones.

**civ:**  
Valor total esperado que un cliente generará en 180 días, combinando frecuencia esperada y monto promedio.

- **Anexo 6: Repositorio de Github**  
[https://github.com/JManuVC/diplomado\\_ciencia\\_de\\_datos](https://github.com/JManuVC/diplomado_ciencia_de_datos)
- **Anexo 7: Carta de aprobación**  
*Archivo:* carta\_aprobación.pdf  
Documento que contiene la carta de aprobación del tutor.

Cochabamba, 5 de Mayo de 2025

MSc. Ing. Ronald Patiño Tito  
DIRECTOR a.i. POSGRADO FCyT  
Presente. -

**REF.: APROBACIÓN DE PROYECTO**

Mediante la presente informo a su autoridad que el proyecto titulado **"PREDICCIÓN DE RECOMPRA DE CLIENTES Y ESTIMACIÓN DEL VALOR DE VIDA (CLV) CON MODELOS BTYD: CASO ACERIA GUADIX S.A."**, desarrollado por el estudiante **Jose Manuel Vargas Cruz**, correspondiente al programa de Diplomado **"Ciencia de Datos"**, 2ª versión, HA SIDO GUIADO EN SU ELABORACIÓN, REVISADO y, al contener un perfil, marco teórico, metodología, aplicación, resultados, conclusiones y recomendaciones correctamente elaboradas con un alcance acorde al grado de diplomado, y con conocimiento de la temática, tengo a bien informar que el mismo está **APROBADO**.

Sin otro particular, envío saludos cordiales.

Atentamente,

  
MSc. Ing. Joel Alejandro Mejía Meléndez  
TUTOR DE PROYECTO  
Tel: 60996606

