

UNIVERSIDAD PRIVADA FRANZ TAMAYO

FACULTAD DE INGENIERÍA

INGENIERÍA DE SISTEMAS



“ANÁLISIS Y PREDICCIÓN DE CHURN EN TELEFONÍA FIJA DE

COTEL R.L.”

CASO: “ Cotel R.L. ”

AUTOR: Osco Valencia Juan Manuel

DOCENTE: Ing. Enrique Alejandro Laurel Cossio

LA PAZ – BOLIVIA

II – 2025

ÍNDICE

1. Introducción	1
2. Contexto del Proyecto	1
3. Planteamiento del Problema	2
3.1. Problemática de Negocio	2
3.2. Desafíos Técnicos	3
4. Justificación y Motivación	3
5. Historia de los Datos	4
5.1. Origen y Fuentes Primarias	5
5.2. Caracterización del Dataset	5
5.3. Proceso de Adquisición y Preparación	6
5.3.1. Fase 1 Extracción y el Desafío de los Metadatos	6
5.3.2. Fase 2 Integración Temporal y validación	7
5.3.3. Fase 3 Limpieza y Enriquecimiento	7
6. Arquitectura Técnica y Flujo de Trabajo	7
6.1. Pipeline de Procesamiento de Datos	8
6.2. Diagrama de Flujo de Datos	9
6.3. Stack Tecnológico	10
6.4. Estructura del Repositorio y Control de Versiones	11

7. Dashboard y Visualización de Datos	12
7.1. Modelo Dimensional y Diseño	12
7.2. Componentes Visuales y Narrativa	13
7.3. Interactividad y Dinamismo Temporal.....	14
7.4. Justificación de las Visualizaciones.....	14
8. MARCO TEÓRICO	15
8.1. Comunicación	15
8.2. Telecomunicaciones	15
8.3. Telefonía Fija	16
8.4. Gestión de Clientes en COTEL R.L.....	16
8.5. Concepto de Pérdida de Clientes (Churn)	17
8.6. Big Data	18
8.7. Modelos Predictivos de Churn.....	19
8.8. Importancia de la Analítica Predictiva en COTEL.....	20
9. Discusión de Resultados e Insights Clave	20
9.1. El Fenómeno del "Abandono Silencioso" (Riesgo Medio)	20
9.2. La Irreversibilidad del Impago (Histéresis del Churn)	21
8.3. El Uso como Indicador Adelantado de Solvencia	22

10. Impacto y Conclusiones Finales	22
10.1. Conclusiones Generales	22
10.2. Valor Generado para la Organización	23
9.3. Cierre	24

1. Introducción

En la industria moderna de las telecomunicaciones, la retención de clientes se ha convertido en un pilar fundamental para la sostenibilidad financiera, a menudo superando en importancia a la adquisición de nuevos usuarios debido a los altos costos asociados a esta última. El fenómeno conocido como *Churn* (tasa de cancelación o abandono) representa uno de los desafíos más críticos para las operadoras, especialmente en segmentos maduros como la telefonía fija.

El presente proyecto se centra en la Cooperativa de Telecomunicaciones (COTEL R.L.) en Bolivia, abordando la problemática de la deserción de clientes desde una perspectiva analítica basada en datos (*Data Driven*). A través del procesamiento masivo de registros de facturación y tráfico telefónico, este estudio busca no solo cuantificar el abandono, sino entender las dinámicas de comportamiento que lo preceden.

Mediante la aplicación de técnicas de Ingeniería de Datos y Análisis Exploratorio, se ha logrado integrar fuentes de información heterogéneas para construir una visión 360° del cliente. El documento detalla el proceso de transformación de datos crudos provenientes de bases de datos Oracle en información estratégica, permitiendo identificar segmentos de riesgo y patrones de uso que son invisibles bajo los métodos de gestión tradicionales.

2. Contexto del Proyecto

El proyecto se desarrolla dentro del entorno operativo de COTEL R.L., una de las principales cooperativas de telecomunicaciones de Bolivia. El alcance del estudio se circunscribe al

servicio de telefonía fija, un sector que, si bien es tradicional, mantiene una base de usuarios significativa que requiere gestión activa para evitar la obsolescencia y el abandono.

Para el análisis, se ha definido una ventana de observación temporal que garantiza la representatividad de los datos:

- **Datos de Facturación:** Periodo de seis meses, comprendido entre mayo y octubre de 2025.
- **Datos de Tráfico (Llamadas):** Periodo trimestral focalizado en julio, agosto y septiembre de 2025.

La magnitud del conjunto de datos es considerable, abarcando un universo de 138,330 contratos únicos. Esto implica el procesamiento de más de 605,000 registros de facturación y un volumen de tráfico de aproximadamente 3.5 millones de llamadas. La infraestructura de datos original reside en bases de datos Oracle, cuyas exportaciones requirieron un tratamiento exhaustivo para su explotación analítica.

3. Planteamiento del Problema

3.1. Problemática de Negocio

COTEL R.L. enfrenta una incertidumbre respecto al comportamiento real de su base de suscriptores de telefonía fija. La gestión tradicionalmente se ha enfocado en el estado de pago, ignorando la variable de "uso del servicio". Esto ha generado una brecha de información donde no es posible distinguir entre un cliente leal y uno que mantiene el servicio por inercia pero no lo utiliza.

El análisis preliminar revela una tendencia preocupante: la tasa de churn mensual ha mostrado un incremento progresivo, pasando del 2.20% en julio al 4.00% en septiembre de 2025. Esta aceleración en la pérdida de clientes evidencia la necesidad urgente de herramientas que permitan anticipar la baja antes de que ocurra el impago definitivo. Además, se ha detectado que una vez que un cliente interrumpe sus pagos, la probabilidad de recuperación es prácticamente nula, lo que subraya la necesidad de acciones preventivas en lugar de reactivas.

3.2. Desafíos Técnicos

Desde una perspectiva técnica, la organización carecía de un conjunto de datos (dataset) unificado y limpio apto para el análisis avanzado o *Machine Learning*. Los datos se encontraban dispersos en tablas aisladas (CLIENTES, FACTURACION, TRAFICO) y las exportaciones del sistema Oracle presentaban problemas de integridad, específicamente la ausencia de cabeceras (*headers*) en los archivos CSV, lo que impedía su ingestión directa en herramientas de análisis como Python.

La falta de integración entre los datos de facturación (quién paga) y los datos de tráfico (quién usa) impedía la creación de una segmentación precisa, resultando en estrategias comerciales genéricas que no abordaban las necesidades específicas de los clientes en riesgo.

4. Justificación y Motivación

La motivación principal de este proyecto radica en la transformación de datos operativos en activos estratégicos. Resolver el problema del Churn en COTEL R.L. no es solo una cuestión técnica, sino una necesidad financiera imperativa.

Este proyecto se justifica bajo tres pilares:

1. **Valor Económico:** Identificar a los clientes que pagan pero no usan el servicio (segmento de "Riesgo Medio", que representa un 39.2% de la cartera analizada) permite diseñar campañas de fidelización preventivas, asegurando el flujo de caja futuro.
2. **Eficiencia Operativa:** La automatización de la limpieza de datos y la integración de fuentes permiten reducir tiempos de análisis manual, pasando de reportes estáticos a un monitoreo dinámico del comportamiento del cliente.
3. **Innovación Tecnológica:** La implementación de scripts en Python para la corrección de headers y la ingeniería de características (Feature Engineering) sienta las bases para la futura implementación de modelos predictivos de Aprendizaje Automático, modernizando la capacidad analítica de la cooperativa.

5. Historia de los Datos

Para garantizar la robustez del análisis y la fiabilidad de los modelos predictivos posteriores, se llevó a cabo un riguroso proceso de recolección y curación de datos. La información base reside en los sistemas transaccionales de COTEL R.L., gestionados bajo un motor de base de datos Oracle. Esta sección detalla el origen, la naturaleza y el proceso de transformación de dichos activos de información.

5.1. Origen y Fuentes Primarias

Los datos provienen directamente del entorno operativo de la cooperativa. Se identificaron y extrajeron tres fuentes de información primaria, cada una correspondiente a una dimensión crítica del negocio:

- **Dimensión Demográfica y Contractual (Tabla CLIENTES):** Contiene el registro maestro de los suscriptores. Incluye identificadores únicos (contrato, cod_cliente), ubicación geográfica (zona, ciudad) y datos sociodemográficos (fecha_nacimiento, ingresos).
- **Dimensión Financiera (Tabla DESFAC.COBFACTU):** Representa el historial de facturación. Proporciona el estado de solvencia del cliente mes a mes, detallando montos facturados, deudas pendientes y fechas de pago.
- **Dimensión Comportamental (Tabla TRAFICO):** Registra el consumo real del servicio. Almacena el detalle de las llamadas realizadas (CDRs - Call Detail Records), permitiendo medir la intensidad de uso, duración y destinos.

5.2. Caracterización del Dataset

Siguiendo el marco teórico del Big Data, los datos analizados se caracterizan por su Volumen, Variedad y Veracidad:

- **Volumen:** La magnitud de los datos es significativa para un análisis local. El estudio consolida 138,330 contratos de telefonía fija, procesando un historial de 605,946 facturas emitidas durante seis meses y un total de 3,505,487 registros de llamadas en el trimestre analizado.

- **Variedad:** Se integran datos estructurados de naturaleza heterogénea:
 - *Datos Estáticos:* Información del cliente (ej. género, fecha de instalación).
 - *Series Temporales:* Comportamiento de pago mensual.
 - *Datos Transaccionales:* Flujo continuo de llamadas telefónicas.
- **Veracidad y Calidad:** Este fue el desafío principal. Los datos crudos presentaron inicialmente una baja calidad estructural (ausencia de metadatos en las exportaciones) y operativa (fechas nulas, inconsistencias en el registro de edades), lo que requirió una etapa intensiva de limpieza para asegurar que la información reflejara fielmente la realidad del negocio.

5.3. Proceso de Adquisición y Preparación

El "relato" de los datos en este proyecto no fue una extracción directa, sino un proceso de reconstrucción y validación que se desarrolló en tres fases secuenciales:

5.3.1. Fase 1 Extracción y el Desafío de los Metadatos

La extracción inicial desde Oracle generó archivos en formato CSV planos. Se identificó un problema crítico: los archivos carecían de cabeceras (headers), presentando únicamente filas de datos crudos sin identificación de columnas. Esto impedía la ingestión automática en herramientas de análisis.

Solución Implementada: Se desarrolló el script `00_fix_headers.py`, el cual reconstruyó la estructura de los archivos asignando programáticamente los nombres de campo correctos (ej. mapeando la columna 0 a contrato, la 1 a cod_cliente, etc.), basándose en el diccionario de datos institucional.

5.3.2. Fase 2 Integración Temporal y validación

Un reto metodológico fue la disparidad en las ventanas de tiempo. Se disponía de 6 meses de facturación (mayo-octubre) frente a 3 meses de tráfico (julio-septiembre). Para lograr la coherencia, se estableció el trimestre Julio-Septiembre 2025 como el periodo núcleo de análisis (Core Analysis Period), utilizando los meses de mayo y junio como contexto histórico de pagos y octubre como ventana de observación post-evento.

5.3.3. Fase 3 Limpieza y Enriquecimiento

Mediante el script 01_eda.py y 02_feature_engineering.py, se transformaron los datos crudos en variables analíticas:

- **Saneamiento:** Se corrigieron formatos de fecha y se imputaron o descartaron registros con edades inverosímiles (ej. clientes con fecha de nacimiento no registrada).
- **Agregación:** Los 3.5 millones de llamadas se resumieron en indicadores de comportamiento por cliente, tales como minutos_total, cantidad_llamadas y destinos_unicos.

El resultado final de este proceso fue la creación de un Dataset Maestro Unificado, libre de inconsistencias estructurales, que vincula por primera vez en la organización el comportamiento de pago con el patrón de uso del servicio.

6. Arquitectura Técnica y Flujo de Trabajo

Para abordar la complejidad del análisis de churn y la segmentación de clientes, se diseñó una arquitectura de datos modular y escalable. Esta arquitectura sigue el paradigma ETL

(Extract, Transform, Load), permitiendo la trazabilidad completa del dato desde su extracción en los sistemas legados hasta su presentación en tableros de inteligencia de negocios.

6.1. Pipeline de Procesamiento de Datos

El flujo de trabajo (pipeline) se estructuró en cuatro etapas secuenciales, implementadas mediante scripts de Python orquestados para garantizar la reproducibilidad del análisis.

1. Ingesta y Normalización:

- Entrada: Archivos CSV crudos exportados de Oracle (Clientes, Facturación, Tráfico).
- Proceso: Dado que los archivos originales carecían de metadatos estructurales (cabeceras), se implementó una capa de pre-procesamiento (scripts/00_fix_*.py) que inyecta programáticamente los nombres de columnas y tipifica los datos.
- Salida: Archivos CSV normalizados listos para el análisis (_corregido.csv).

2. Ingeniería de Características:

- Proceso: Transformación de datos transaccionales (3.5 millones de registros de llamadas) en variables de comportamiento agregadas por cliente. Se calcularon métricas como minutos_promedio_mes, destinos_unicos y tendencia_llamadas.
- Lógica: Se aplicaron reglas de negocio para derivar la edad del cliente y su antigüedad a partir de fechas crudas, gestionando valores nulos y formatos inconsistentes.

3. Integración y Lógica de Negocio:

- Proceso: Fusión (Merge) de las dimensiones de Facturación y Tráfico.
- Hito Clave: Ejecución del script analisis_churn_mes_a_mes.py, el cual aplica la lógica de segmentación cruzada (Pago vs. Uso) para clasificar a cada cliente en uno de los cuatro segmentos de riesgo (Activos, Riesgo Medio, Riesgo Alto, Churn).

4. Capa Semántica y Visualización:

- Proceso: Modelado de datos para Business Intelligence. Se generaron tablas dimensionales (DIM_CLIENTES, DIM_FECHA) y tablas de hechos (FACT_PAGO, FACT_TRAFICO) para alimentar un modelo en estrella.
- Salida: Ingesta final en Power BI para la generación de dashboards interactivos.

6.2. Diagrama de Flujo de Datos

A continuación, se esquematiza el flujo técnico de la información a través de los distintos módulos desarrollados:

Flujo de Datos - Análisis Churn COTEL

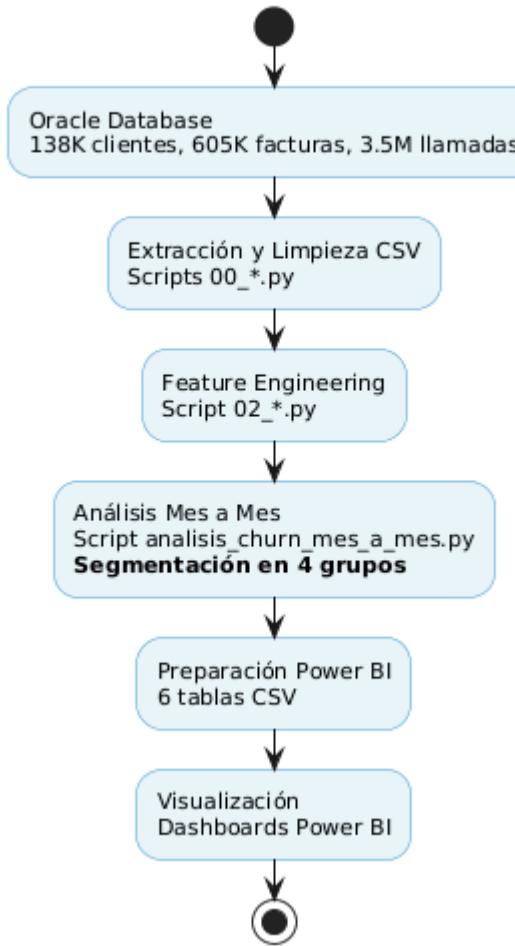


Ilustración 1: Flujo de Datos
Fuente: Elaboración Propia

6.3. Stack Tecnológico

La selección de herramientas respondió a criterios de capacidad de procesamiento, flexibilidad y costo-efectividad:

- **Lenguaje de Programación:** Python 3.x. Seleccionado por su ecosistema robusto para Ciencia de Datos.

- **Pandas:** Utilizado para la manipulación de DataFrames, permitiendo operaciones vectorizadas eficientes sobre millones de registros que serían inviables en hojas de cálculo tradicionales.
 - **NumPy:** Empleado para cálculos numéricos de alto rendimiento.
- **Fuente de Datos:** Oracle Database (Sistema Transaccional).
- **Visualización:** Microsoft Power BI. Se utilizó para la capa de presentación debido a su capacidad de filtrado dinámico (Drill-down) y modelado de relaciones.
- **Entorno de Desarrollo:** Visual Studio Code con extensiones para Python.

6.4. Estructura del Repositorio y Control de Versiones

El desarrollo del proyecto se gestionó bajo un sistema de control de versiones (Git), asegurando la integridad del código y facilitando la experimentación sin riesgo de pérdida de información. La estructura del repositorio refleja la separación lógica de responsabilidades:

- **/data:** Almacenamiento segregado de datos crudos (inmutables) y datos procesados (intermedios y finales).
- **/scripts:** Contiene el código fuente numerado secuencialmente para indicar el orden de ejecución:
 - `00_fix_...`: Scripts de saneamiento de estructura.
 - `01_eda.py`: Análisis Exploratorio.
 - `02_feature_engineering.py`: Creación de variables predictivas.
 - `analysis_churn_mes_a_mes.py`: Script "Core" del negocio.

- preparar_powerbi.py: Script de transformación para capa visual.
- **/output:** Directorio de destino para los datasets finales (dataset_churn_mes_a_mes.csv) y reportes segmentados por categoría de riesgo.

7. Dashboard y Visualización de Datos

La fase final de la arquitectura de datos consistió en la implementación de una capa de visualización interactiva. Se diseñó un tablero de mando (*Dashboard*) en Microsoft Power BI, concebido no meramente como un reporte estático, sino como una herramienta de exploración heurística que permite a los directivos de COTEL R.L. navegar desde los indicadores macroeconómicos hasta el detalle granular de cada contrato.

7.1. Modelo Dimensional y Diseño

Para sustentar las visualizaciones, se transformó el dataset analítico plano en un Modelo en Estrella (Star Schema), optimizado para consultas rápidas y filtrado dinámico. Este modelo consta de:

- **Tablas de Dimensiones:** DIM_CLIENTES (atributos fijos) y DIM_FECHA (jerarquía temporal).
- **Tablas de Hechos:** FACT_PAGO_MENSUAL y FACT_TRAFICO_MENSUAL, que almacenan las métricas cuantitativas.
- **Tabla de Resumen:** SEGMENTACION_COMPLETA, una tabla agregada diseñada específicamente para alimentar los gráficos de alto nivel.

7.2. Componentes Visuales y Narrativa

El tablero se estructura en vistas estratégicas que apoyan la narrativa del "Cliente en Riesgo", permitiendo identificar patrones que los reportes tabulares tradicionales ocultaban.

A. Vista de Contexto y Alcance

El objetivo de esta vista es dimensionar la magnitud del análisis.

- **Visualizaciones Clave:** Tarjetas de KPI (Key Performance Indicators).
- **Métricas:** Total de Contratos (138,330), Periodo de Facturación (Mayo-Octubre) y Volumen de Tráfico (3.5M llamadas).
- **Aporte a la Narrativa:** Establece la confianza en el dato, demostrando que las conclusiones se basan en el censo completo de la base de usuarios y no en una muestra estadística parcial.

B. Vista de Segmentación de Riesgo (El "Core" del Análisis)

Esta es la visualización central del proyecto, donde se responde a la pregunta de investigación sobre la naturaleza del churn.

- **Gráfico de Barras Verticales:** Muestra la distribución absoluta de clientes en los 4 segmentos definidos (Activos, Riesgo Medio, Riesgo Alto, Churn).
- **Gráfico Circular (Donut Chart):** Representa la participación porcentual de cada segmento.
- **Aporte a la Narrativa:** Estas gráficas revelan el hallazgo más contraintuitivo del estudio: el segmento de "Riesgo Medio". Visualmente, se hace evidente que un

bloque masivo de 23,189 clientes (39.2%) mantiene sus pagos al día (color verde/positivo) pero ya no utiliza el servicio (indicador de uso nulo). Esta visualización refuta la hipótesis tradicional de que "mientras paguen, no hay problema", exponiendo un riesgo latente de cancelación masiva.

7.3. Interactividad y Dinamismo Temporal

A diferencia de los reportes estáticos en PDF o Excel, el tablero incorpora Selectores de Tiempo (Slicers). Esto permite al usuario:

1. **Analizar la Evolución:** Cambiar la vista de "Julio" a "Agosto" y "Septiembre" para observar cómo los clientes migran entre segmentos.
2. **Detectar Tendencias:** Visualizar cómo la barra de "Churn" crece mes a mes (del 2.20% al 4.00%), validando visualmente la urgencia del problema planteado en la introducción.

7.4. Justificación de las Visualizaciones

Cada elemento visual fue seleccionado con un propósito cognitivo específico:

- **Uso del Color Semántico:** Se utilizó una paleta de colores lógica (Verde para "Activos", Amarillo para "Riesgo Medio", Rojo para "Riesgo Alto/Churn") que reduce la carga cognitiva y permite interpretar el estado de salud de la cartera en segundos.
- **Jerarquía Visual:** Los totales y agregados se presentan en la parte superior, mientras que los detalles operativos se ubican en la parte inferior, facilitando la lectura ejecutiva.

8. MARCO TEÓRICO

8.1. Comunicación

La comunicación es el proceso mediante el cual dos o más individuos intercambian información, ideas, pensamientos o sentimientos a través de un código común. Según Shannon y Weaver (1949), la comunicación se compone de varios elementos: emisor, mensaje, canal, receptor y retroalimentación. En el ámbito tecnológico, este proceso se traduce en la transmisión de información a través de medios físicos o digitales, lo cual constituye la base del desarrollo de las telecomunicaciones modernas.

La evolución de la comunicación ha permitido que las sociedades amplíen sus formas de interacción, pasando de medios analógicos a digitales, y facilitando la aparición de servicios que permiten conectar a personas y organizaciones a escala global.

8.2. Telecomunicaciones

Las telecomunicaciones son el conjunto de técnicas, dispositivos y sistemas que permiten la transmisión de información a distancia, ya sea en forma de voz, datos o video, mediante señales electromagnéticas. De acuerdo con la Unión Internacional de Telecomunicaciones (UIT, 2019), este sector es esencial para el desarrollo económico y social, al posibilitar la conectividad y el acceso a la información.

En Bolivia, el sector de telecomunicaciones ha experimentado una importante expansión en las últimas décadas, impulsado por la digitalización de los servicios y la competencia entre operadores. Sin embargo, los servicios tradicionales, como la telefonía fija, han sufrido una disminución sostenida debido al crecimiento de la telefonía móvil y el acceso a Internet.

8.3. Telefonía Fija

La telefonía fija es un servicio de telecomunicación que permite la transmisión de voz mediante una red cableada de cobre o fibra óptica, conectando un teléfono residencial o empresarial a una central local. Este servicio fue durante muchos años el principal medio de comunicación domiciliaria y empresarial.

En el contexto de COTEL R.L. (Cooperativa de Telecomunicaciones La Paz), la telefonía fija ha representado históricamente uno de los pilares de su operación y fuente principal de ingresos. No obstante, la migración de usuarios hacia la telefonía móvil y servicios de comunicación basados en Internet (VoIP, mensajería instantánea, etc.) ha generado una disminución constante en la base de clientes activos.

El servicio de telefonía fija en COTEL se basa en un modelo cooperativo, en el cual los usuarios adquieren acciones que les otorgan derecho al uso del servicio. Esto implica un sistema administrativo que maneja altas, bajas, cortes, reconexiones y suspensiones de servicio, además de una estructura de facturación mensual.

En muchos casos, los procesos de morosidad y abandono del pago derivan en la suspensión temporal o definitiva del servicio, lo que constituye una manifestación directa del fenómeno conocido como churn o pérdida de clientes.

8.4. Gestión de Clientes en COTEL R.L.

COTEL administra su cartera de clientes de telefonía fija mediante una estructura basada en contratos y acciones.

Cada socio o cliente cuenta con un número de acción que le otorga derecho a un punto de

conexión (línea telefónica). La relación comercial se rige por un contrato de prestación de servicios, sujeto a obligaciones de pago y mantenimiento.

El ciclo de vida del cliente dentro de la cooperativa incluye etapas como:

- Alta o suscripción: el usuario adquiere una acción y se activa el servicio.
- Facturación periódica: se generan cargos mensuales por uso y mantenimiento.
- Mora o incumplimiento de pago: el cliente acumula deudas y recibe notificaciones.
- Corte temporal o suspensión: el servicio se desactiva por falta de pago.
- Abandono o baja definitiva: el cliente deja de pagar por un periodo prolongado, y la acción se desactiva o pasa a disponibilidad.

Este proceso administrativo refleja la pérdida progresiva de clientes, especialmente en los últimos años, donde factores como la competencia, la digitalización y la migración tecnológica han reducido la rentabilidad del servicio.

8.5. Concepto de Pérdida de Clientes (Churn)

El término churn se utiliza para describir la tasa o cantidad de clientes que dejan de utilizar un servicio durante un periodo determinado. En el ámbito de las telecomunicaciones, se refiere a los usuarios que cancelan, suspenden o abandonan sus servicios (Lemmens & Croux, 2006).

Existen dos tipos de churn:

- Voluntario: cuando el cliente decide cambiar de proveedor o suspender el servicio por decisión propia.

- Involuntario: cuando la empresa corta el servicio por impago o incumplimiento contractual.

Para empresas como COTEL, la identificación de clientes propensos al churn es vital, ya que permite anticipar el abandono, aplicar estrategias de retención y reducir pérdidas económicas.

La predicción del churn mediante técnicas de análisis de datos e inteligencia artificial es actualmente una de las aplicaciones más relevantes del aprendizaje automático en el sector de las telecomunicaciones.

8.6. Big Data

El término Big Data se refiere a la gestión y análisis de grandes volúmenes de datos, caracterizados por las 5V:

- Volumen: enorme cantidad de datos generados.
- Velocidad: rapidez con la que se producen y procesan.
- Variedad: diversidad de fuentes y formatos (textos, registros, sensores, logs, etc.).
- Veracidad: confiabilidad de la información.
- Valor: utilidad de los datos para generar conocimiento.

En COTEL, el concepto de Big Data se aplica en la integración de diversas fuentes:

- Base de datos de contratos y facturación.
- Historial de pagos y cortes de servicio.
- Información de morosidad y reconexiones.
- Registros de uso de servicios y atención al cliente.

Mediante la utilización de herramientas analíticas y modelos predictivos, la cooperativa puede detectar patrones de comportamiento que indiquen riesgo de pérdida de clientes, optimizando la toma de decisiones y la eficiencia operativa.

8.7. Modelos Predictivos de Churn

Los modelos predictivos son algoritmos que permiten estimar la probabilidad de que un cliente abandone el servicio. Entre los más utilizados en el ámbito de telecomunicaciones se encuentran:

- Regresión logística
- Árboles de decisión
- Random Forest
- XGBoost
- Redes neuronales artificiales

Estos modelos procesan variables como:

- Antigüedad del contrato.
- Historial de pagos.
- Frecuencia de cortes y reconexiones.
- Tipo de plan o servicio contratado.
- Ubicación geográfica y segmento del cliente.

La implementación de un modelo predictivo en COTEL permitirá identificar a los clientes con mayor riesgo de abandono, generando alertas tempranas y posibilitando la aplicación de estrategias de retención personalizadas.

8.8. Importancia de la Analítica Predictiva en COTEL

La aplicación de analítica predictiva en COTEL R.L. representa un paso fundamental hacia la transformación digital. Gracias a los modelos de churn basados en Big Data, la cooperativa puede:

- Reducir pérdidas económicas por abandono de clientes.
- Priorizar recursos de atención y fidelización.
- Aumentar la satisfacción del cliente.
- Tomar decisiones basadas en evidencia.

9. Discusión de Resultados e Insights Clave

El análisis cruzado de las dimensiones de facturación y tráfico telefónico ha revelado patrones de comportamiento que permanecían ocultos bajo la gestión tradicional. A continuación, se exponen los tres hallazgos empíricos más relevantes derivados del estudio, su interpretación y sus implicaciones operativas para COTEL R.L.

9.1. El Fenómeno del "Abandono Silencioso" (Riesgo Medio)

El hallazgo más significativo del estudio es la identificación de un segmento masivo de clientes clasificados como "Riesgo Medio".

- **La Evidencia:** El análisis de julio de 2025 reveló que 23,189 clientes (39.2% de la cartera) se encuentran al día en sus pagos pero registran cero minutos de uso.
- **Interpretación:** Tradicionalmente, estos usuarios son considerados "clientes ideales" por el departamento financiero, ya que generan ingresos sin consumir recursos de la red. Sin embargo, desde la perspectiva de retención, representan un riesgo latente extremo. Este comportamiento sugiere que el servicio ha perdido su utilidad práctica para el usuario (posible sustitución por telefonía móvil) y que la suscripción se mantiene probablemente por inercia o descuido.
- **Relevancia:** Este grupo representa ingresos actuales que son altamente volátiles. Sin una estrategia de reactivación de uso o fidelización, es estadísticamente probable que estos clientes migren al segmento de Churn en el corto plazo al darse cuenta del gasto innecesario.

9.2. La Irreversibilidad del Impago (Histéresis del Churn)

Al analizar las matrices de transición de estados entre julio y agosto, se observó un patrón de comportamiento unidireccional respecto a la morosidad.

- **La Evidencia:** Se detectó que 1,189 clientes que eran pagadores activos en julio dejaron de pagar en agosto. En contraparte, la tasa de recuperación (clientes morosos que regularizaron su deuda) fue estadísticamente insignificante (cercana a 0 en la muestra analizada).
- **Interpretación:** Existe un punto de no retorno. Una vez que el cliente rompe el hábito de pago mensual, las estrategias de cobranza actuales de la cooperativa no logran revertir la situación. El churn en COTEL R.L. no es cíclico, es absorbente.

- **Relevancia:** Esto invalida las estrategias reactivas. La inversión de recursos debe focalizarse en la prevención (antes del primer impago) y no en la recuperación (gestión de mora), donde la probabilidad de éxito es nula.

8.3. El Uso como Indicador Adelantado de Solvencia

El estudio confirma una fuerte correlación positiva entre la intensidad de uso del servicio y el cumplimiento de pago.

- **La Evidencia:** Del total de clientes que no pagan sus facturas, el 77.6% (2,359 usuarios) tampoco utilizan el servicio. Por el contrario, solo un 1.2% (681 usuarios) utiliza el servicio activamente sin pagar (fraude o mora técnica).
- **Interpretación:** El abandono del uso precede al abandono del pago. La caída en el tráfico de voz es una "señal temprana" (early warning signal) que aparece meses antes de que se genere la deuda financiera.
- **Relevancia:** Se establece que el monitoreo del tráfico (minutos/llamadas) es un predictor más eficaz y oportuno para el riesgo de baja que el historial de crédito.

10. Impacto y Conclusiones Finales

10.1. Conclusiones Generales

El proyecto "Análisis y Predicción de Churn en Telefonía Fija" ha cumplido con sus objetivos fundamentales, logrando transformar una gran cantidad de datos operativos dispersos en inteligencia de negocio accionable.

1. **Integración Exitosa:** Se superaron barreras técnicas significativas, logrando la unificación de 138,330 contratos y más de 3.5 millones de registros de llamadas,

resolviendo problemas críticos de calidad de datos y ausencia de metadatos en los sistemas de origen.

2. **Redefinición del Problema:** Se demostró que el Churn no es un evento binario (pagó/no pagó), sino un proceso gradual de desvinculación que comienza con la falta de uso.
3. **Cuantificación del Riesgo:** Se determinó que la tasa de cancelación está en una tendencia alcista alarmante, acelerándose del 2.20% al 4.00% en solo un trimestre, lo que exige intervención inmediata.

10.2. Valor Generado para la Organización

La implementación de esta solución analítica aporta valor tangible en tres dimensiones:

- **Estratégico:** Provee a la gerencia de una visión segmentada real. Ya no se gestiona a "los clientes" como un todo, sino que se pueden aplicar estrategias diferenciadas: *retención* para los Activos, *reactivación* para el Riesgo Medio y *cobranza dura* para el Riesgo Alto.
- **Operativo:** El *pipeline* de datos automatizado en Python reduce de días a minutos el tiempo necesario para generar reportes de estado de cartera, eliminando el error humano en la manipulación de Excel.
- **Económico:** Al identificar proactivamente a los 23,000 clientes de "Riesgo Medio", la cooperativa tiene la oportunidad de proteger una porción significativa de sus ingresos recurrentes antes de que estos usuarios decidan cancelar el contrato.

9.3. Cierre

En conclusión, este proyecto sienta las bases para una transformación digital en la gestión de clientes de COTEL R.L. Al pasar de un enfoque reactivo basado en la facturación a un enfoque predictivo basado en el comportamiento, la organización está mejor equipada para enfrentar el desafío de la retención de clientes en un mercado competitivo. La metodología aquí desarrollada es escalable y replicable, constituyendo un activo intelectual duradero para la cooperativa.

Bibliografía

McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (2nd ed.). O'Reilly Media.

VanderPlas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media.

Ferrari, A., & Russo, M. (2016). *The Definitive Guide to DAX: Business intelligence with Microsoft Excel, SQL Server Analysis Services, and Power BI*. Microsoft Press.

ATT (Autoridad de Regulación y Fiscalización de Telecomunicaciones y Transportes). (2024). *Memoria Institucional y Estadísticas del Sector de Telecomunicaciones en Bolivia*. Estado Plurinacional de Bolivia.