



Introducción a Data Mining: Análisis de Churn en Telco

El **churn** o **attrition** indica cuando un cliente abandona el servicio de una compañía en favor de otra. Este análisis se basa en un dataset con 20 variables sobre ~3000 clientes de telecomunicaciones, junto con la indicación de si el cliente abandonó o no la compañía.

Exploraremos las variables, buscaremos correlaciones, analizaremos patrones de abandono y desarrollaremos modelos predictivos para identificar clientes con riesgo de churn.

Variables del Dataset

Datos de Identificación

- State: categórica, 50 estados y Distrito de Columbia
- Account length: entero, tiempo de actividad de la cuenta
- Area code: categórica
- Phone number: identificador del cliente

Planes Contratados

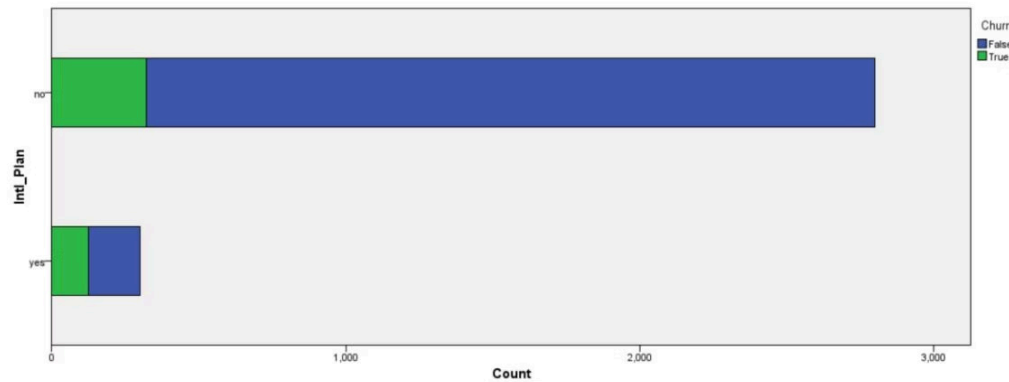
- International Plan: categórica dicotómica (sí/no)
- VoiceMail Plan: categórica dicotómica (sí/no)
- Number of voice mail messages: entero

Uso del Servicio

- Minutos, llamadas y cargos (día, tarde, noche, internacional)
- Number of calls to customer service: entero

La variable respuesta **churn** es dicotómica con valores TRUE y FALSE, indicando si el cliente abandonó o no el servicio.

Análisis de Variables Categóricas



Matrix of Churn by Intl_Plan

Intl_Plan		
Churn	no	yes
False.	2479	173
True.	321	126

Cells contain: cross-tabulation of fields (including missing values)
Chi-square = 205.955, df = 1, probability = 0

Plan Internacional:

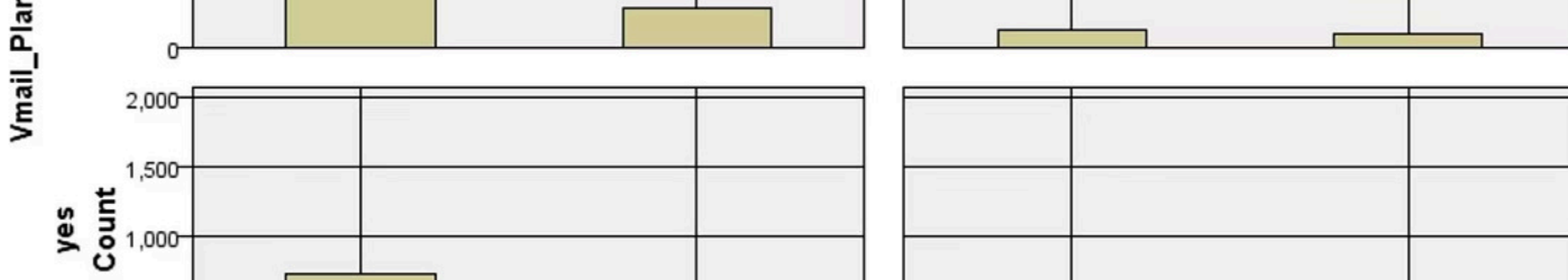
- 9,65% de clientes seleccionaron este plan
- 41% de clientes con Plan Internacional son churners
- Solo 11% de clientes sin Plan Internacional son churners

Conclusión preliminar: Ambos planes influyen significativamente en el churn, pero en direcciones opuestas. El Plan Internacional parece aumentar el abandono, mientras que el Voice Mail parece retener clientes.



Plan Voice Mail:

- 27,59% de clientes tienen este plan
- 16,5% de clientes sin Voice Mail son churners
- Solo 8,9% de clientes con Voice Mail son churners



Interacción entre Planes y Churn

Sin Voice Mail

La proporción de churners es significativamente mayor si tienen Plan Internacional (44%) que si no lo tienen (14%).

Con Voice Mail

La proporción de churners sigue siendo mayor si tienen Plan Internacional (39%) frente a los que no lo tienen (5%).

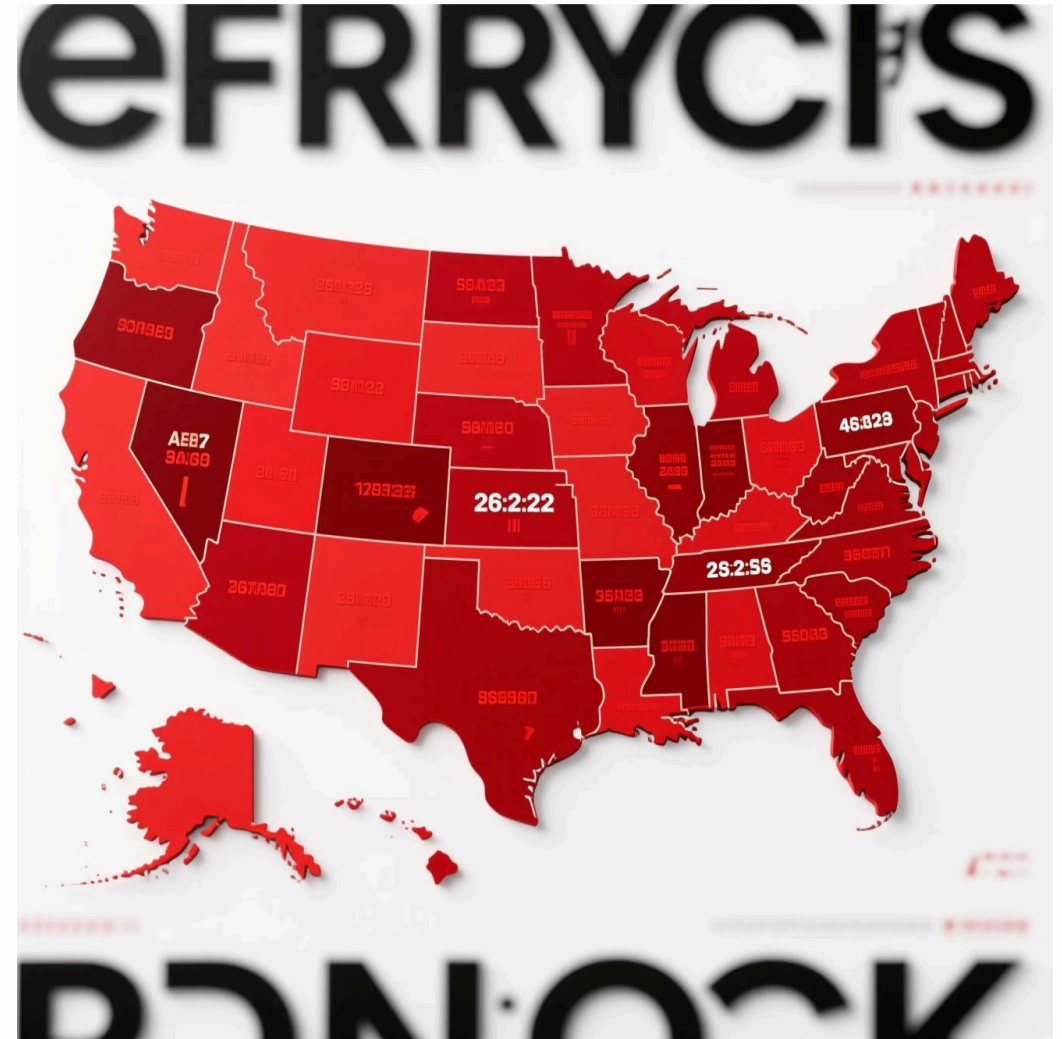
El Plan Internacional parece ser un fuerte predictor de abandono independientemente de si el cliente tiene Voice Mail o no, aunque el Voice Mail reduce el riesgo en ambos casos.

Campos Anómalos: Area Code y State

Distribución de Area Code

- 408: 25,01% (775 clientes)
- 415: 49,79% (1543 clientes)
- 510: 25,2% (781 clientes)

Solo hay 3 códigos de área para todos los estados, lo que resulta anómalo considerando que cada estado debería tener sus propios códigos de área.



Esta anomalía sugiere que los campos **area code** y/o **state** deberían omitirse del análisis hasta obtener clarificación, ya que no representan datos reales.

Exploración de Variables Numéricas

101,18

Account Length

Media de días de antigüedad

Rango: 1-243 días

179,60

Day Minutes

Media de minutos diurnos

Rango: 0-350,8 minutos

Correlaciones importantes:

Se detectó una correlación perfecta ($r=1,0$) entre **day_minutes** y **day_charge**, indicando que el cargo es una función lineal de los minutos (tarifa plana de 17 centavos por minuto).

Fórmula: **day_charge** = $0,000613 + 0,17 \times \text{day_minutes}$

Esta correlación perfecta se repite para los cargos de tarde, noche e internacionales, por lo que estos campos deberían omitirse del análisis.

Llamadas al Servicio al Cliente y Churn

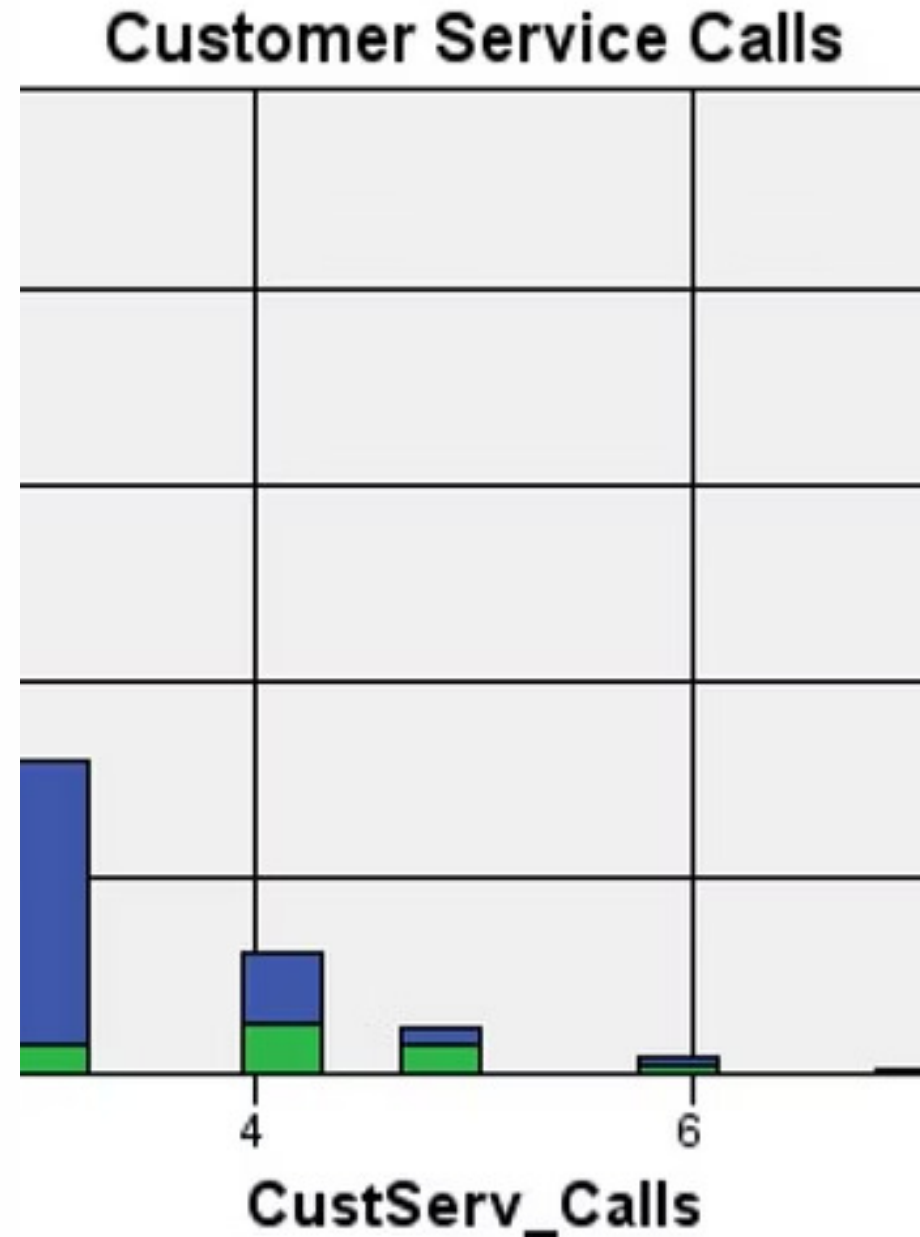
Hallazgos clave:

- Los clientes con 4 o más llamadas al servicio al cliente abandonan con una frecuencia **cuatro veces mayor** que el resto.
- Después de la tercera llamada, es necesario ofrecer incentivos para retener al cliente.

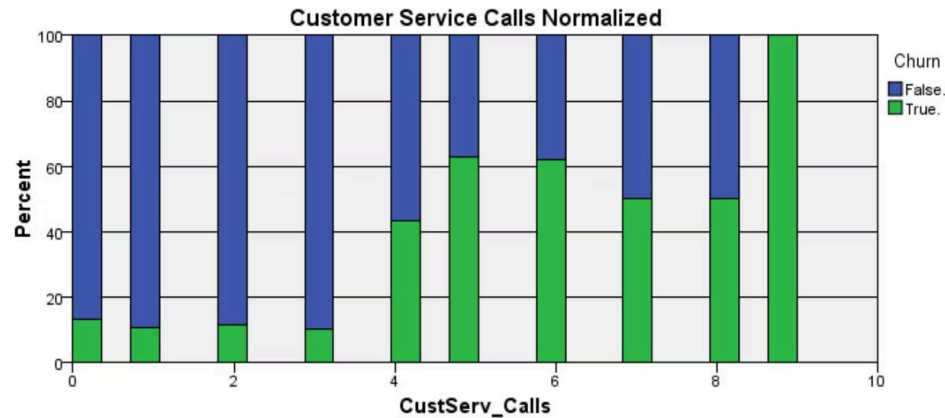
Esta variable será probablemente un predictor importante en cualquier modelo de data mining para predecir el churn.



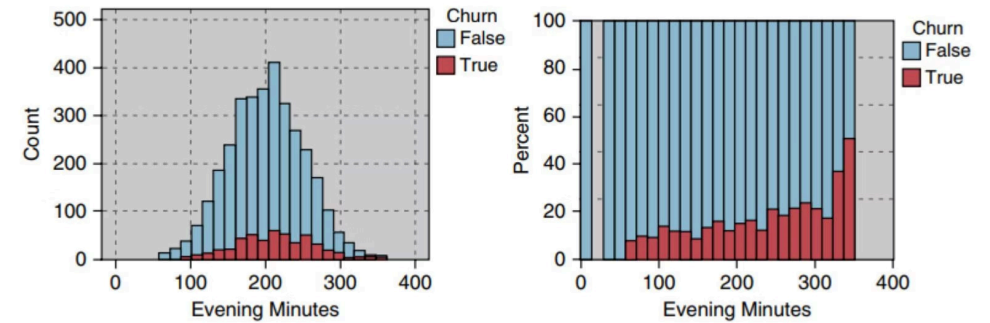
Punto de acción: Implementar un sistema de alerta cuando un cliente realice su tercera llamada al servicio al cliente en un período corto de tiempo.



Minutos de Uso y Churn



Los usuarios con valores altos en **day minutes** tienden a abandonar con tasas más altas.



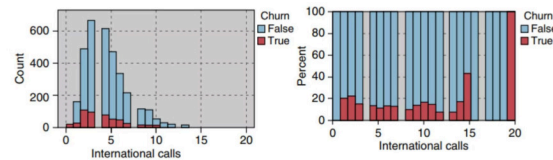
Los usuarios con valores altos en **evening minutes** también muestran tendencia al abandono, aunque menos clara.

Recomendaciones

- Monitorear cuidadosamente los minutos utilizados por el cliente
- Cuando el valor supera 200 minutos diurnos, considerar incentivos especiales
- Investigar por qué los usuarios intensivos están tentados a abandonar



Llamadas Internacionales y Churn



Los churners tienden a realizar **menor cantidad promedio** de llamadas internacionales.

Aunque visualmente la relación no parece fuerte, un test t para la diferencia en la media de llamadas internacionales entre churners y no churners es estadísticamente significativo.

Esto indica que la variable es útil para predecir el churn, a pesar de no ser tan evidente en los gráficos.

Resumen del Análisis Exploratorio

Variables a omitir

- Los cuatro campos **charge** (función lineal de **minutes**)
- Campos **area code** y/o **state** (anomalías)

Predictores de churn

- Clientes con **Plan Internacional** abandonan con mayor frecuencia
- Clientes con **Voice Mail Plan** abandonan con menor frecuencia
- ≥ 4 llamadas al servicio al cliente: tasa de abandono 4× mayor
- Altos **day minutes** y **evening minutes**: tasa de abandono 6× mayor
- Bajo **day minutes** + altas llamadas al servicio: mayor tasa de abandono
- Baja cantidad de llamadas internacionales: mayor tasa de abandono

Estos hallazgos servirán como base para desarrollar modelos predictivos de churn que permitan identificar clientes en riesgo y tomar medidas preventivas.