

Análisis de Customer Churn – Telco IBM

1. Contexto del negocio

Una empresa de telecomunicaciones enfrenta una alta tasa de cancelación de clientes (*churn*), lo que impacta directamente en los ingresos y el crecimiento. Retener clientes existentes es más rentable que adquirir nuevos, por lo que identificar los factores asociados al abandono es clave para la toma de decisiones.

2. Objetivo del proyecto

Analizar los datos de clientes para:

- Identificar patrones asociados al churn.
 - Detectar segmentos de clientes con mayor riesgo de abandono.
 - Proponer recomendaciones basadas en datos para reducir la tasa de churn.
-

3. Dataset

Fuente: Telco Customer Churn – IBM

Descripción:

- Registros de clientes de una empresa de telecomunicaciones.
- Variables demográficas, tipo de contrato, servicios contratados y cargos.

Variables clave:

- **Contract** : Mensual, Anual, Bienal.
- **MonthlyCharges**
- **Ternure** : Antigüedad en meses
- **Churn** : Variable Objetivo (Si / No)

Variable objetivo:

- **Churn (Sí / No)**
-

4. Herramientas utilizadas

- **Python:** pandas, numpy, matplotlib / seaborn
- **SQL:** consultas exploratorias y agregaciones
- **Power BI:** limpieza en Power Query, creación de KPIs y dashboards
- **Excel:** validación rápida de métricas

5. Metodología

5.1 Limpieza de datos

- Tratamiento de valores nulos
- Conversión de variables categóricas
- Revisión de tipos de datos

5.2 Análisis exploratorio (EDA)

- Distribución del churn
- Análisis por tipo de contrato
- Análisis por cargos mensuales y antigüedad
- Comparación entre clientes que abandonan y los que permanecen

5.3 Definición de KPIs

- Tasa de churn (%)
- Antigüedad promedio del cliente
- Cargo mensual promedio
- Churn por tipo de contrato

5.4 Visualización

- Dashboard interactivo en Power BI
- Segmentación por contrato, servicios y cargos

6. Principales hallazgos

- Los clientes con contrato **mensual** presentan mayor tasa de churn.
 - Cargos mensuales altos están asociados a mayor probabilidad de abandono.
 - Clientes con baja antigüedad tienden a cancelar con mayor frecuencia.
-

6.1 Hallazgos cuantitativos

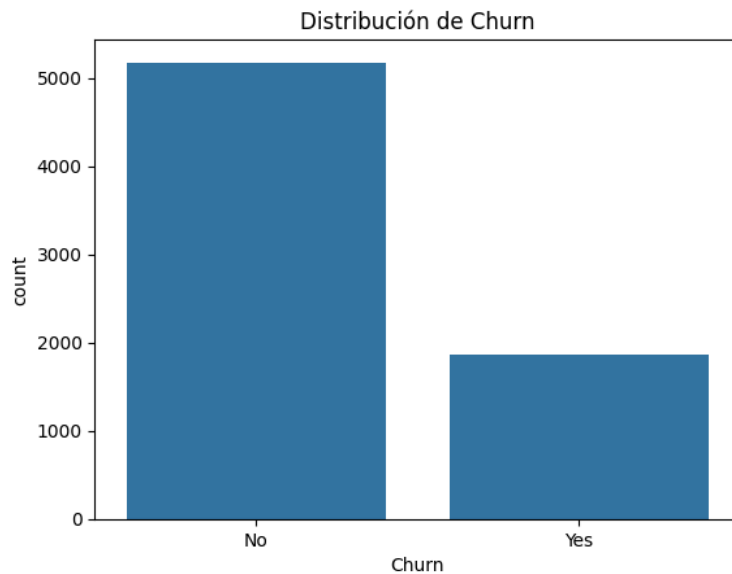


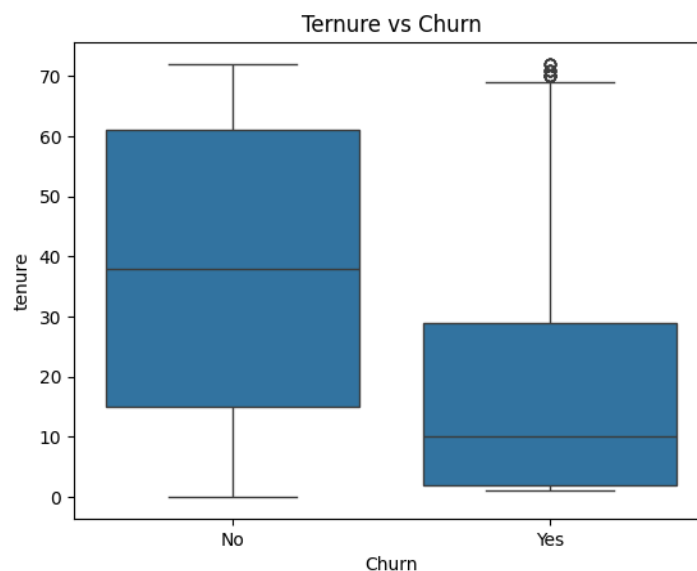
Figura 1 : Distribución de la variable objetivo Churn (Sí vs. No)

La Distribución de Churn muestra un marcado desbalance de clases, donde la gran mayoría de los clientes pertenecen a la categoría 'No'. Esta disparidad justifica el uso de pesos balanceados en el modelo para asegurar que la minoría (clientes que sí abandonan) sea identificada correctamente.

6.2 Factores clave asociados al Churn

El análisis de la **Figura 2** muestra que el abandono se concentra en clientes nuevos, con una mediana de antigüedad significativamente menor a la de los clientes leales. Reforzar la fidelización en los primeros meses es clave, ya que la probabilidad de fuga disminuye drásticamente tras superar los dos años de permanencia.

Figura 2 : Relación entre antigüedad del cliente (tenure en meses) y tasa de churn.



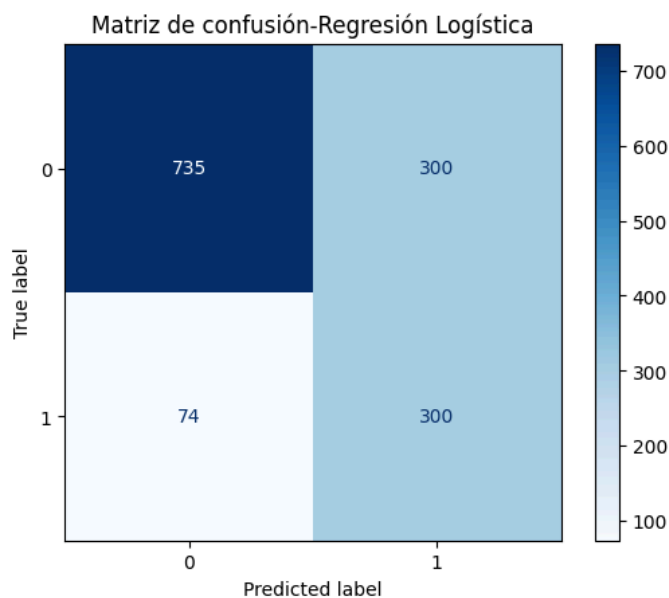
7. Recomendaciones de negocio

- Incentivar contratos de mayor duración.
 - Ofrecer beneficios a clientes nuevos durante los primeros meses.
 - Revisar la estrategia de precios para clientes con cargos elevados.
-

8. Modelo Predictivo

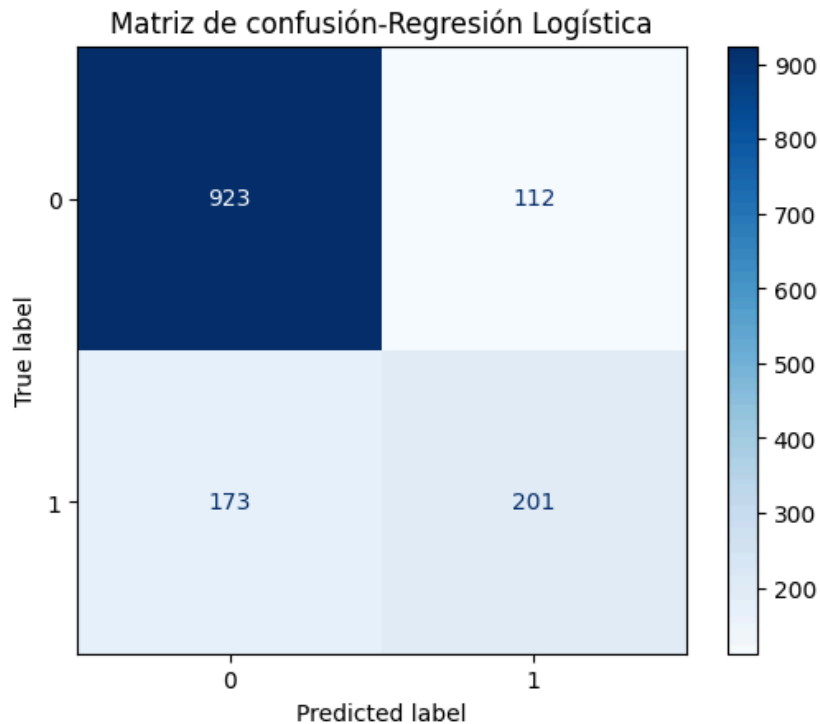
El modelo es capaz de detectar de manera eficiente los casos negativos con un 91% de precisión. Por otro lado logra capturar el 80% de los casos reales de la clase 1, al predecir los falla mucho dado que la mitad de sus predicciones positivas son errores.

Figura 3 : Modelo con Enfoque en Retención



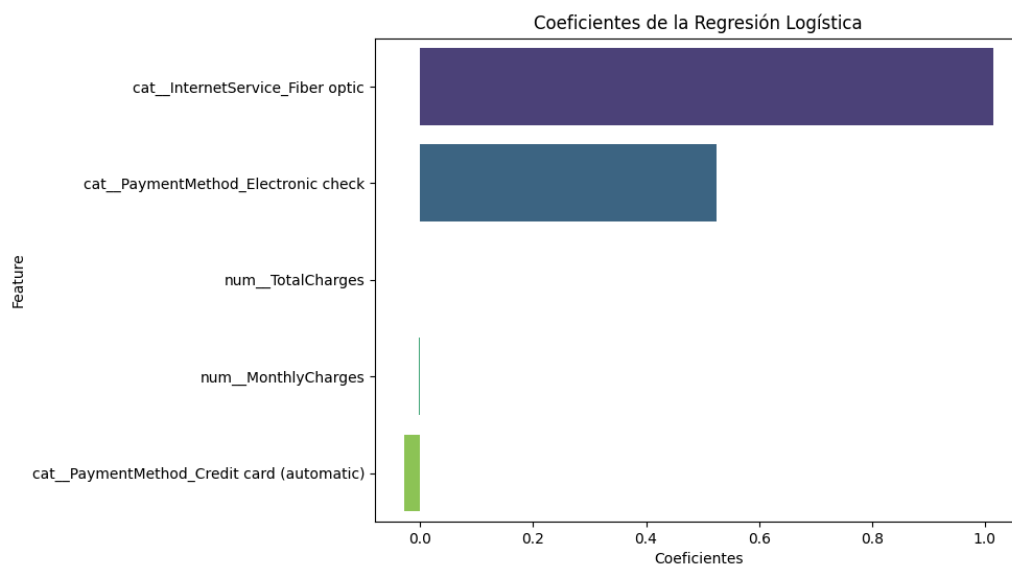
De la Figura 3 concluimos que el modelo arriesga, dado que prefiere marcar casi todo como positivo para no perderse ningún caso real de abandono, aunque los falsos positivos aumentaron drásticamente.

Figura 4: Modelo con Enfoque en Precisión



En esa misma línea la **Figura 4** nos muestra una probabilidad de acierto del 64%, lo cual reduce el desperdicio de dinero en campañas de retención para gente que no pensaba irse. Por otro lado, el recall de la clase positiva registró un 54% indicando que hay fuga de clientes no identificados. En cuanto al Accuracy nos permite observar, que el modelo acierta en mayor medida a los clientes que no abandonarán el servicio.

Figura 5 : Principales Factores Impulsores del Churn: Influencia de Servicios y Métodos de Pago



La **Figura 5** muestra cómo el modelo identifica correctamente a los clientes en riesgo de fuga, destacando la fibra óptica y el pago electrónico como los principales factores que aumentan la probabilidad de abandono.

9. Entregables

- Notebook de análisis exploratorio en Python.
 - Dashboard interactivo en Power BI.
 - Documento de conclusiones y recomendaciones.
 - Modelo de regresión logística
-

10. Próximos pasos

- Construcción de un modelo predictivo sencillo para estimar la probabilidad de churn.
 - Evaluación de impacto potencial de las recomendaciones propuestas.
-