

# Proyecto Final

## Churn en Telecomunicaciones



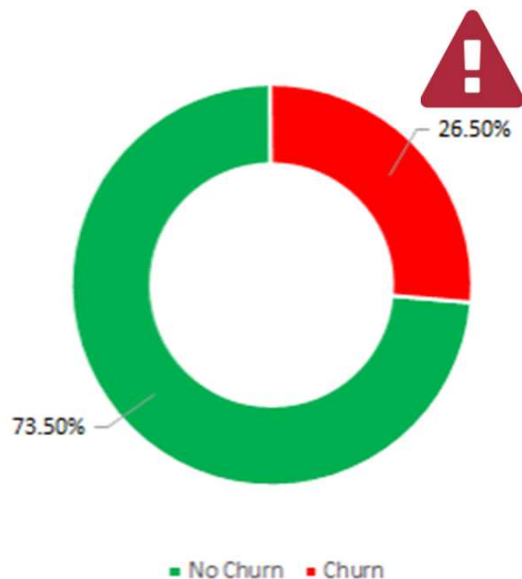
Florencia García Palacio  
Federico Hoffmann  
Gabriel Jourdan  
Victoria Ramondelli

Profesor: Damián Dapuetto  
Tutor: Alejandro Pujol  
Comisión 29730

***CODER HOUSE***

## Contexto

Telco es una empresa estadounidense que brinda servicios de telecomunicaciones a 125M de usuarios en todo el país, capturando el 30% del market share.



## Problema : Alta tasa de abandono

Reducir la tasa de abandono de una empresa es sumamente importante al permitir disminuir los costos de adquisición, ya que es más costoso adquirir nuevos clientes que retener a los existentes.

---

# Objetivo

Predecir qué clientes tienen mayores probabilidades de abandonar la empresa en el siguiente mes mediante el uso de Machine Learning, para poder ejecutar acciones específicas que tiendan a la retención de los usuarios, reduciendo su tasa de abandono actual (churn), en al menos 5pp en el próximo trimestre.

A short horizontal bar with a teal segment on the left and an orange segment on the right.

# Análisis Exploratorio de Datos (EDA)

## Features más relevantes

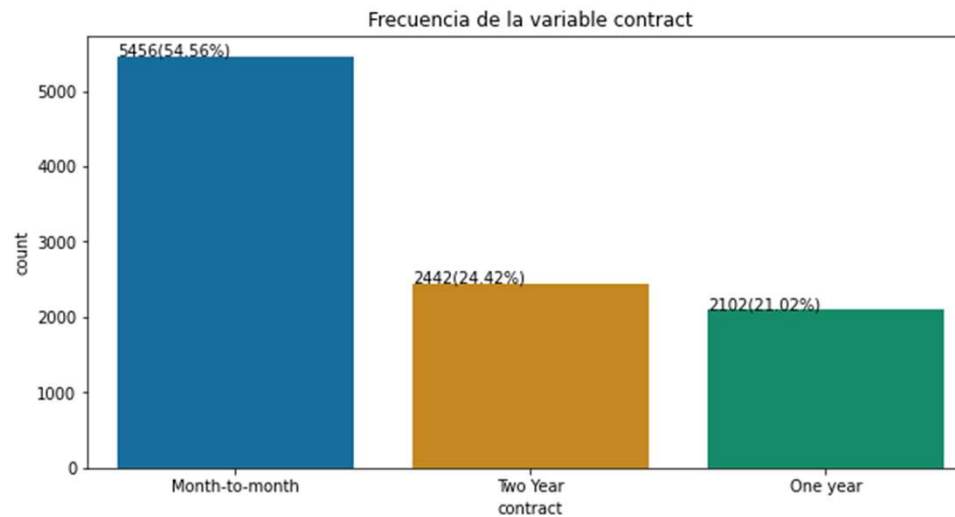
El dataset cuenta con datos de 10.000 clientes (cada fila representa un cliente)

Contamos con información variada de cada cliente: datos demográficos, encuestas y servicios contratados, entre otros.

Se compone por un total de 82 variables. A continuación se resumen las que tienen mayor relación con churn según el análisis EDA:

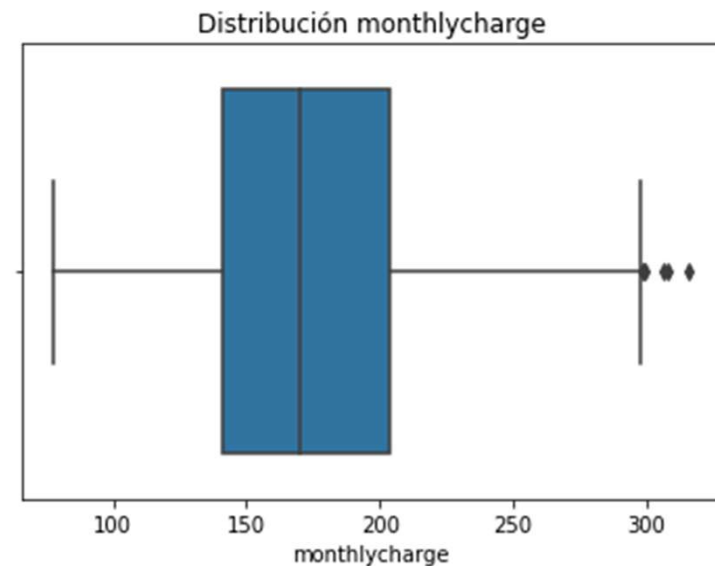
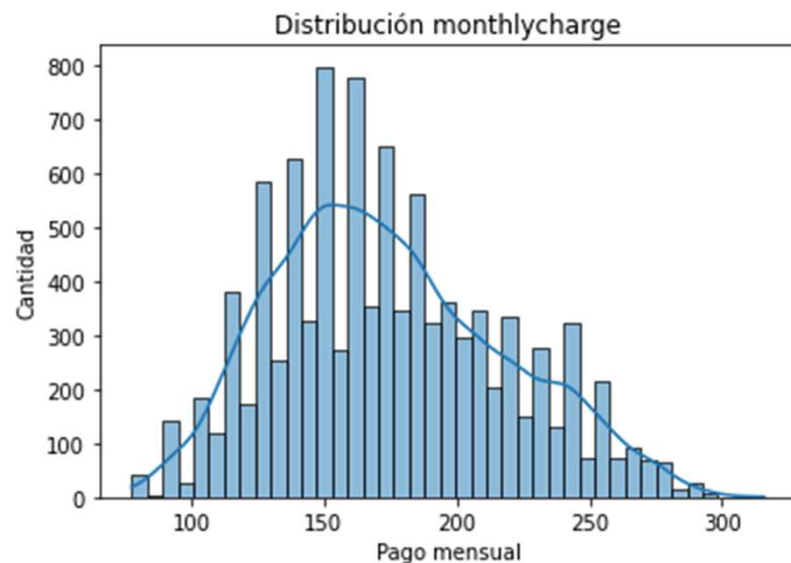
Nombre	Descripción	Relación
Contract	Plazo del contrato del cliente (mes a mes, un año, dos años)	Los contratos mensuales tienen una mayor proporción de churn
MonthlyCharge	Importe cobrado al cliente mensualmente. Este valor refleja un promedio por cliente	Su correlación es levemente positiva (0.36). La mediana de quienes abandonan es mayor
Tenure	Número de meses que el cliente se ha quedado con el proveedor	La mayoría de los churn se concentran en personas con menos de 30 meses de antigüedad
Bandwidth_Gb_Year	Cantidad promedio de datos utilizados (GB), en un año por el cliente	Fuertemente correlacionada con Tenure. Los churn se concentran en valores bajos
q_streaming	Construida en base a servicio de transmisión de TV y de video, cuenta cuántos servicios de este tipo contrata el cliente (0, 1 ó 2)	Son los servicios que presentan mayor porcentaje de churn cuando se los contrata

# Contract



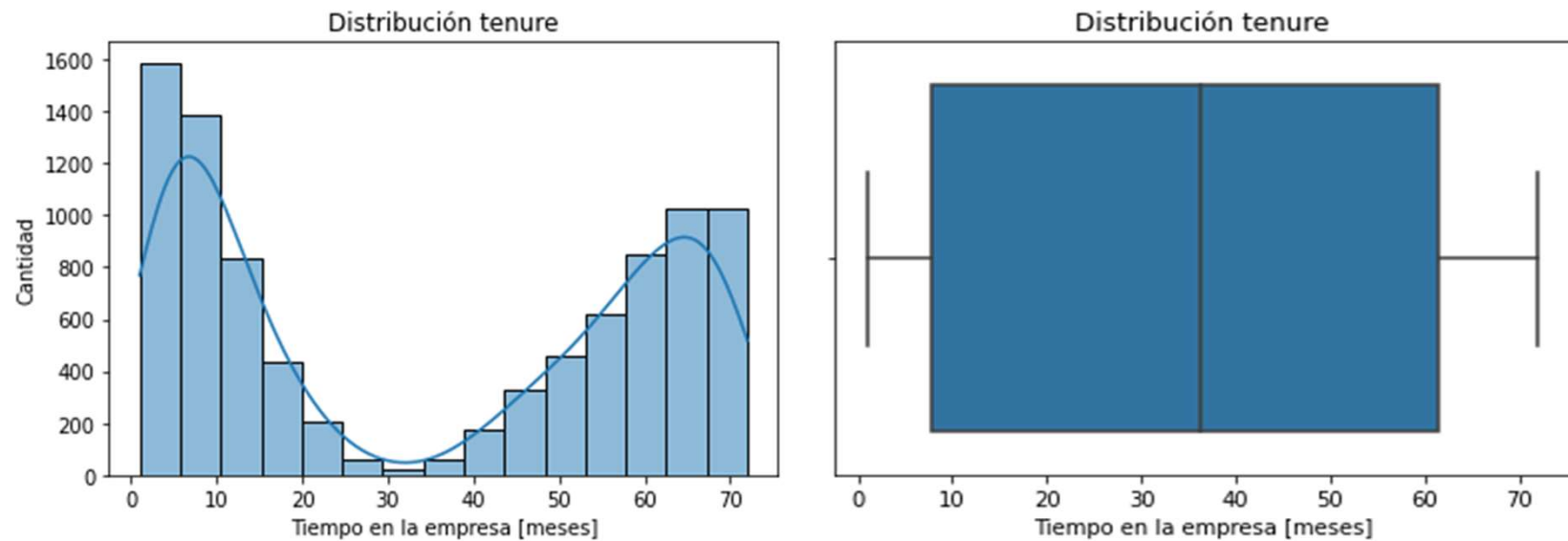
Existe un 55% de clientes que contratan el servicio de forma mensual, otro 21% que prefiere renovarlo anualmente, y el 24% restante lo renueva cada dos años.

## Monthly Charge



Se observa una variable que aparenta estar normalizada. La mayor cantidad de datos se encuentra concentrada en el orden de los USD 150 mensuales, por lo que podríamos afirmar que la distribución es asimétrica con un pequeño sesgo a la derecha.

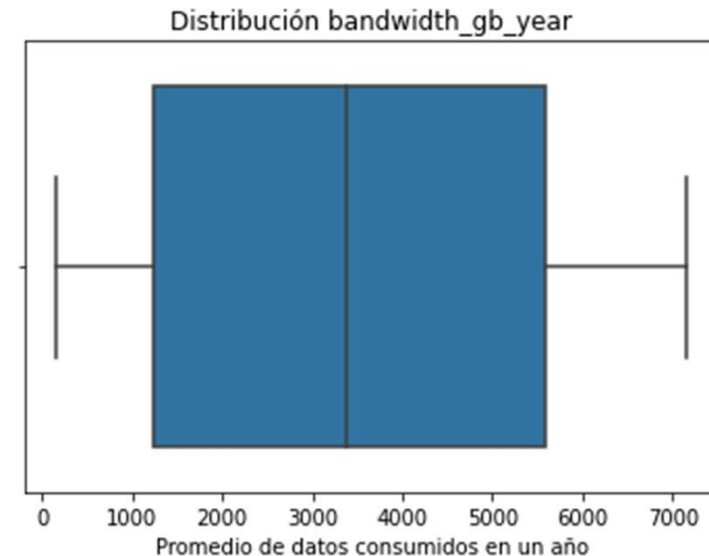
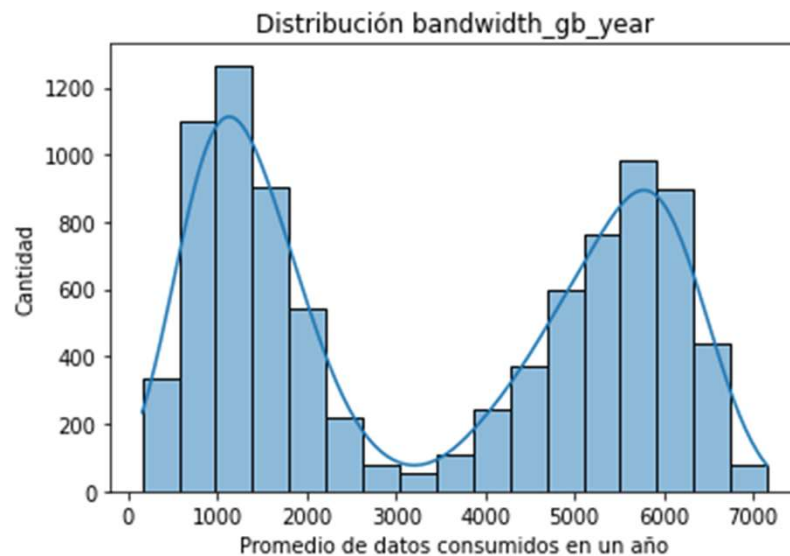
## Tenure



El histograma permite ver dos picos que representan distintas modas, es decir, los dos valores más comunes en nuestro conjunto de datos. Se distinguen dos tipos de clientes principales: quienes están hace más de 40 meses y quienes no superan los 20 meses de antigüedad.



## Bandwidth GB Year

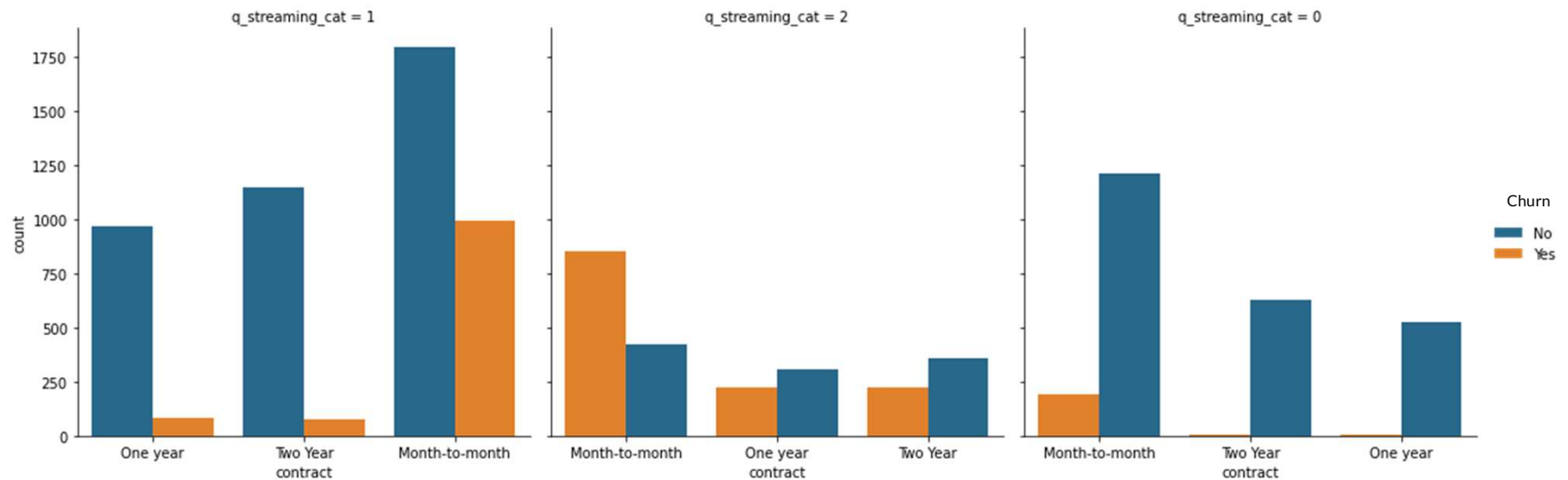


Esta variable también presenta una distribución bimodal, lo que implica que se distinguen dos grupos de clientes principales: quienes consumen anualmente alrededor de 5500 GB, y quienes consumen 1000 GB aprox.

## Principales Conclusiones

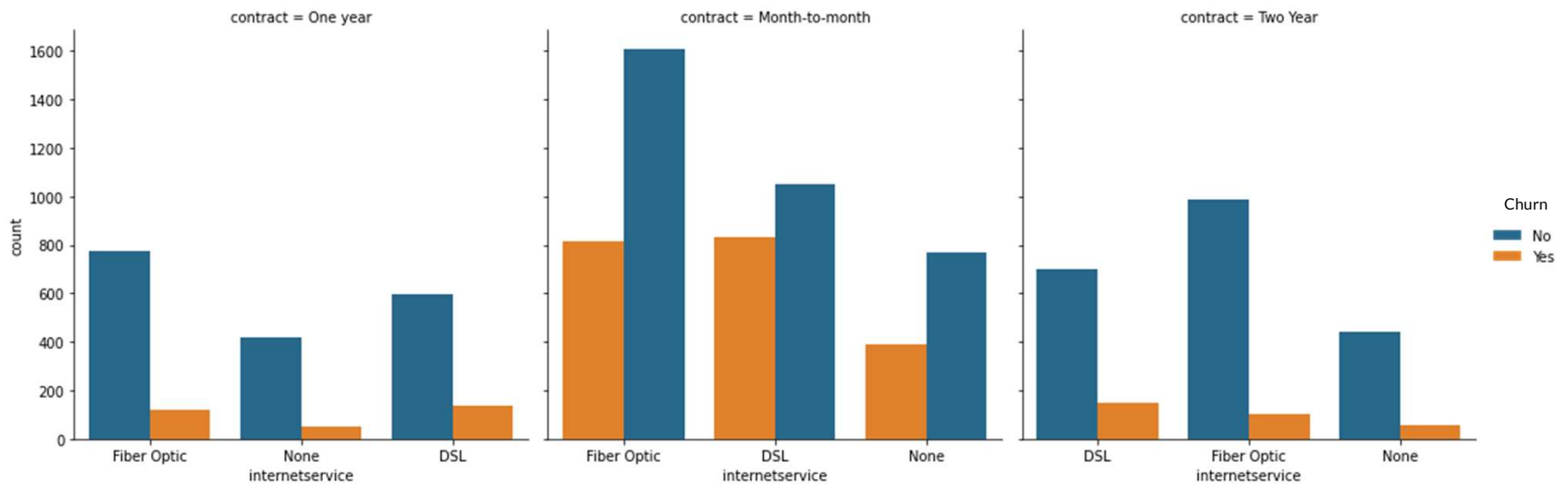
Los clientes que tienen contratos mensuales tienen mayor churn en todas las categorías. Quienes tienen contratos anuales o bi anuales tienen un muy bajo porcentaje de abandono al contratar un servicio de streaming, y casi nulo cuando no contratan.

Hay mayor churn en quienes contratan 2 servicios de streaming en general.



## Principales Conclusiones

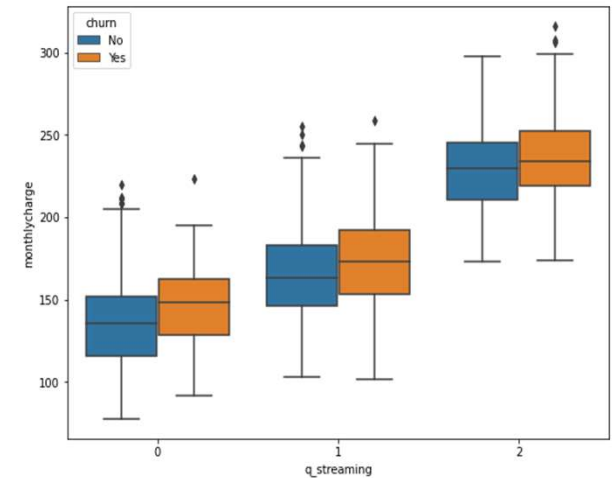
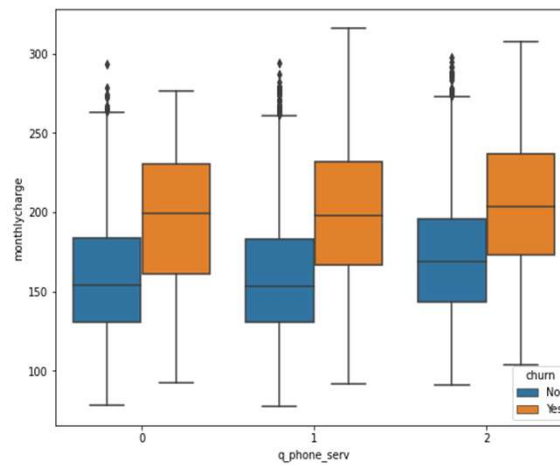
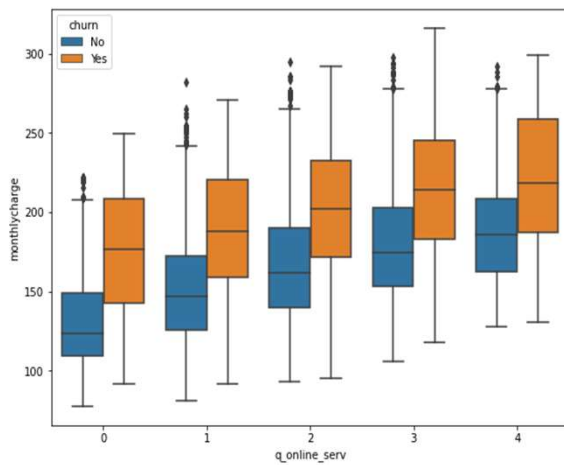
Quienes contratan servicio DSL tienen mayor porcentaje de churn en todos los tipos de contrato.



# Principales Conclusiones

En los servicios de internet y teléfono, aquellos usuarios con abono más caro son quienes tienen mayor propensión al abandono, mientras que en los servicios de streaming, el valor del abono no es un determinante, y la proporción de usuarios que abandonan es similar.

Quienes contratan servicios de streaming pagan abonos más altos.



A small horizontal bar with a teal segment on the left and an orange segment on the right.

# Predicción de Churn

Implementación de modelos de  
Machine Learning

# Modelos implementados

---

Para enfrentar el desafío de descubrir cuáles son las características que más influyen al momento de predecir la condición de Churn / No Churn, implementamos los siguientes modelos de clasificación supervisada con técnicas de balanceo, oversampling, cross validation e hyperparameter tuning:

- **Decision Tree**
- **Logistic Regression**
- **Random Forest**
- **KNN**

## Métricas

A continuación presentamos un resumen de las métricas de cada modelo:

Algoritmo	F1-Score	F-Score-train	Precision	Recall	AUC
Decision Tree base	0.72	1	0.73	0.71	0.80
Decision Tree CV	0.78	0.77	0.77	0.77	0.84
Random Forest base	0.79	1	0.83	0.76	0.85
Random Forest CV	0.87	0.80	0.83	0.78	0.86
Logistic Regression	0.77	0.77	0.80	0.73	0.83
KNN	0.78	0.56	0.66	0.5	0.70

Al probar los algoritmos de Decision Tree y Random Forest con los hiperparámetros predeterminados, los modelos presentaban overfitting, por lo que se aplicó un CrossValidation y búsqueda de hiperparámetros utilizando RandomSearch.

El algoritmo seleccionado es  
**Random Forest** porque presenta  
mejor F1 Score y Recall.

---



## Métricas

El resultado de la **matriz de confusión** resultante luego de probar el modelo, es el siguiente:

R E A L E S	PREDICCIONES	
	No Churn	Churn
	No Churn	Churn
No Churn	1384	86
Churn	117	413

## Métricas

A partir de la matriz de confusión, se obtienen las siguientes métricas:

Balanced Accuracy	Recall	Precision	F1-Score
86.03%	77.92%	82.76%	80.27%

Consideramos que este modelo predice de manera aceptable el churn de la empresa y puede utilizarse para ayudar a detectar y reducir los casos que tengan mayor probabilidad de abandonar en el próximo mes.

A small horizontal bar with a teal segment on the left and an orange segment on the right.

# Próximos pasos

## Próximos pasos

Productivizar el modelo

Construir una aplicación que subiremos a la nube (AWS), para que el equipo de Marketing conecte mediante una API al data lake de usuarios y actualice semanalmente la base y así poder dirigir sus campañas en pos de la reducción del churn.

Como equipo de Data & Analytics, estaremos monitoreando y haciendo revisiones periódicas para ir reajustándolo.



¡GRACIAS!

