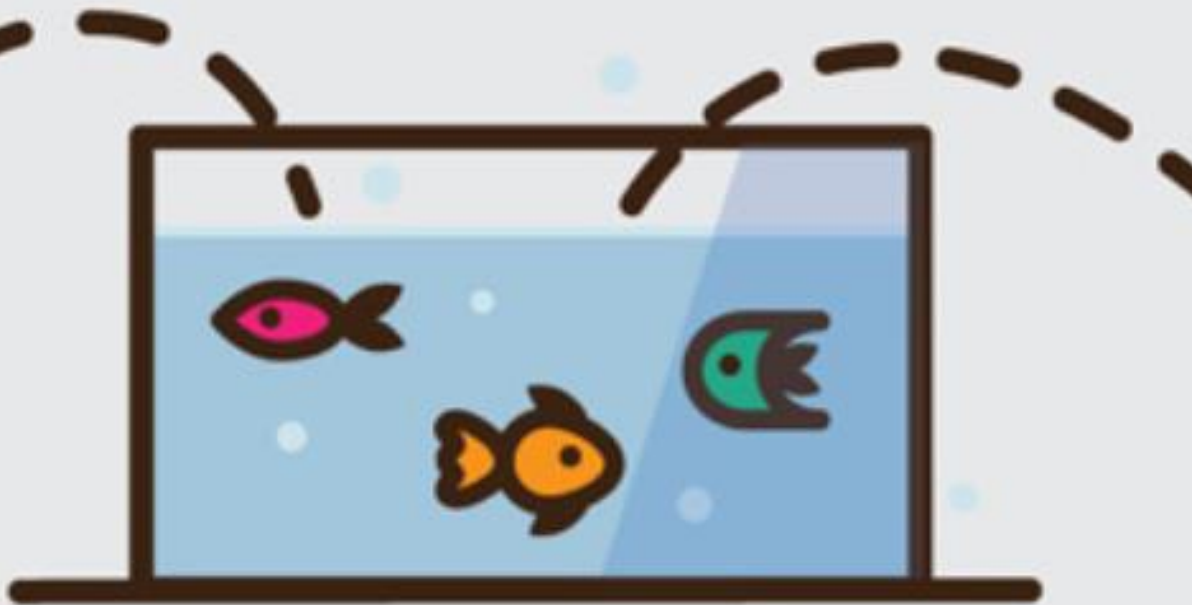


CUSTOMER CHURN




Analisis Predictivo y de Supervivencia a una empresa de Telecomunicaciones (TELECOM)

PRESENTADO POR:

JESÚS ALBERTO LAURA
BALBIN

Índice



1. Descripción del Problema y Objetivo

2. Analisis Descriptivo (Insights)

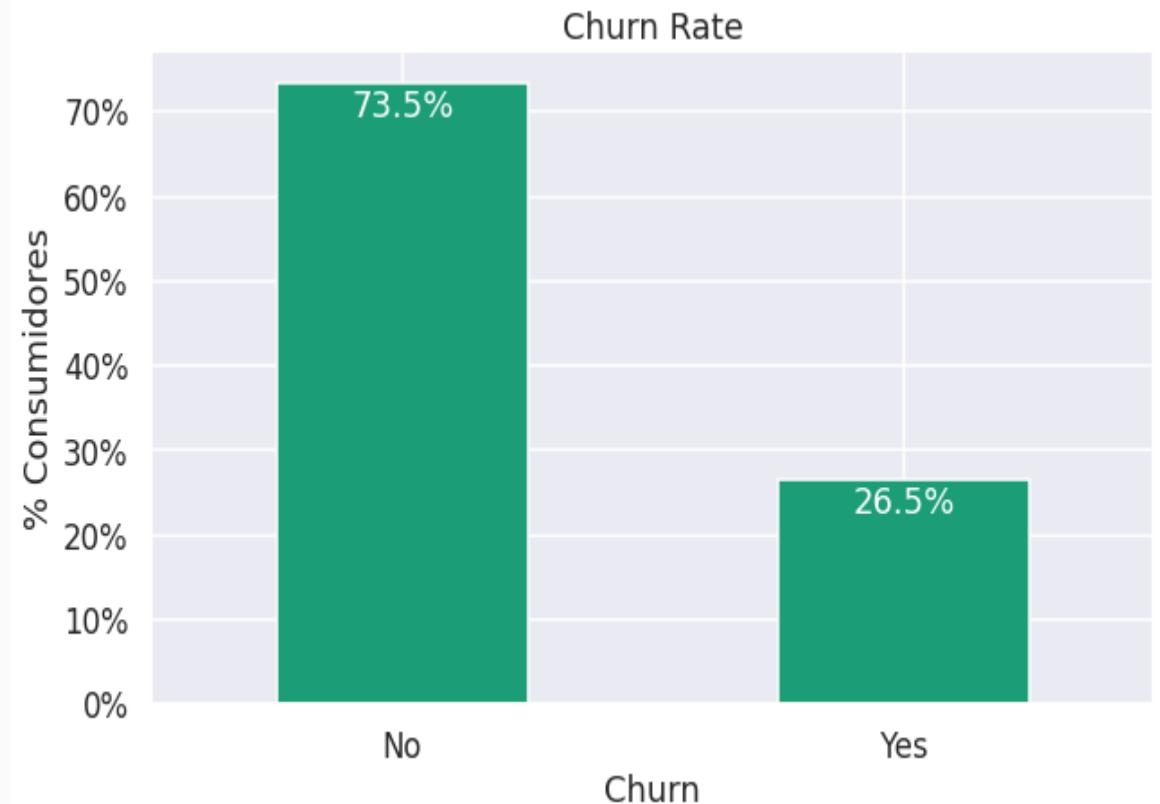
3. Modelo Predictivo

4. Análisis de Supervivencia

Conclusiones

1. Descripción Del Problema y Objetivo

- El desarrollo de una relación con los clientes es un factor crítico para las empresas que buscan permanecer en la industria y aumentar sus ganancias en el segmento de mercado en el que participan (Van den Poel y Larivière, 2004), este tipo de relación puede mejorar entendiendo y prediciendo el churn rate, ya que el objetivo de este indicador es buscar la permanencia de los clientes (fidelización).
- Frente a ello y partiendo de una tasa churn de 26.5% para la empresa TELECOM utilizaremos técnicas de machine learning para predecir el churn y con el análisis de sobrevivencia (survival analysis) podremos identificar que factores aceleran la probabilidad de churn (abandono) de cada grupo de clientes.



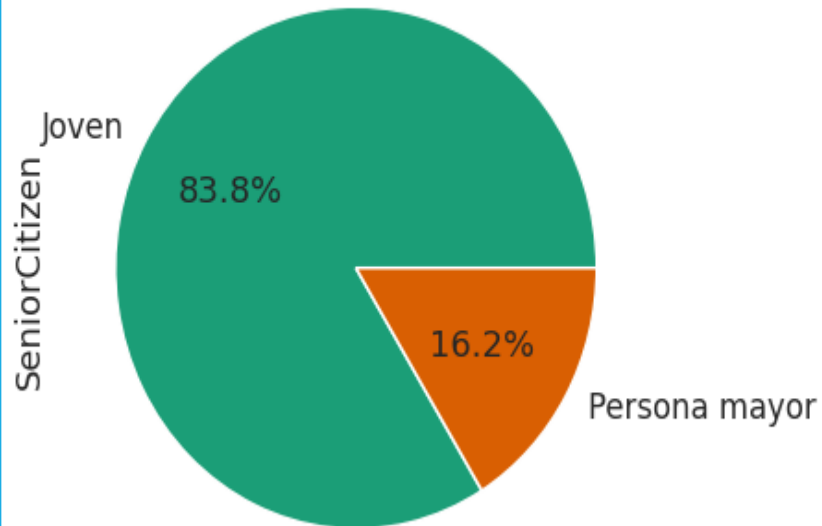
2. Análisis Descriptivo (Insights)

16.2% (1141) son clientes de la tercera edad (senior), y el resto 83.8% (5902) son jóvenes, lo que evidencia que la mayor parte de cliente de la empresa son de este grupo etario

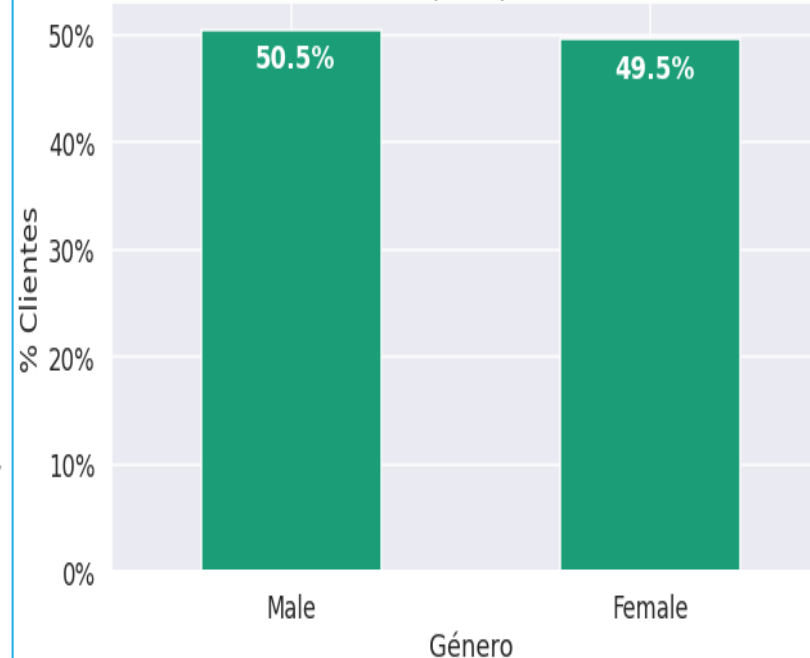
50.5% (3557) son clientes varones, mientras que el otro 49.5% (3486) son mujeres

Muchos clientes han estado con la compañía de telecomunicaciones durante solo un mes, mientras que otro grupo de 550 clientes, permanecen durante aproximadamente 70 meses.

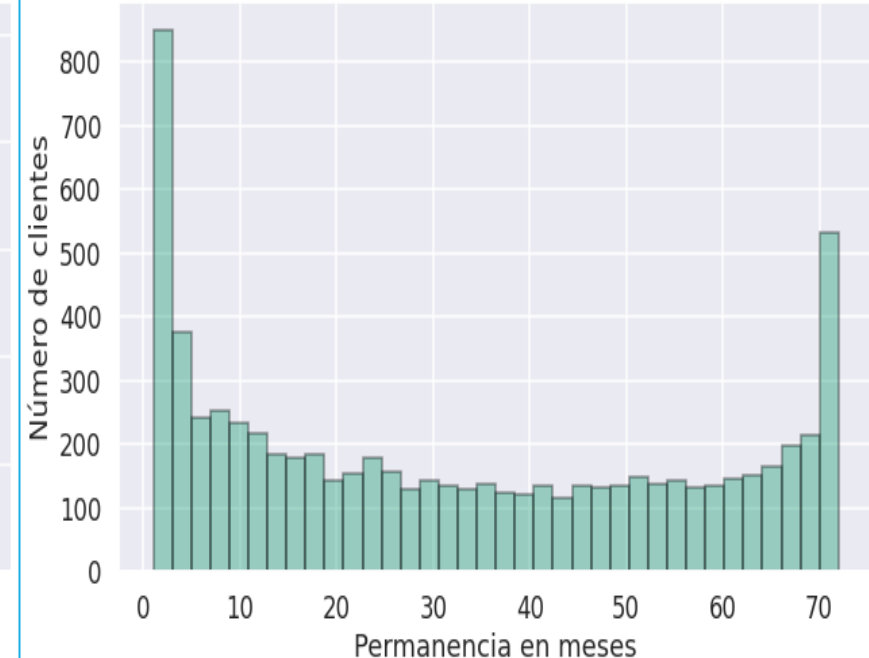
% de Ciudadanos Mayores



Distribución por Tipo de Género

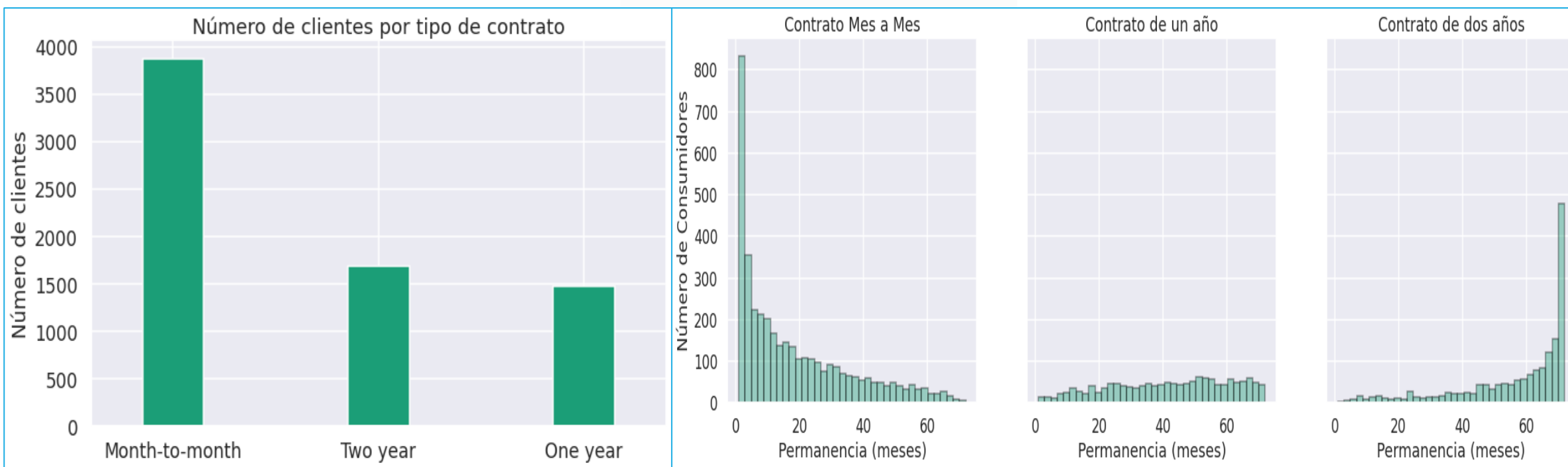


Tiempo de Vida de Clientes (Permanencia)



2. Análisis Descriptivo (Insights)

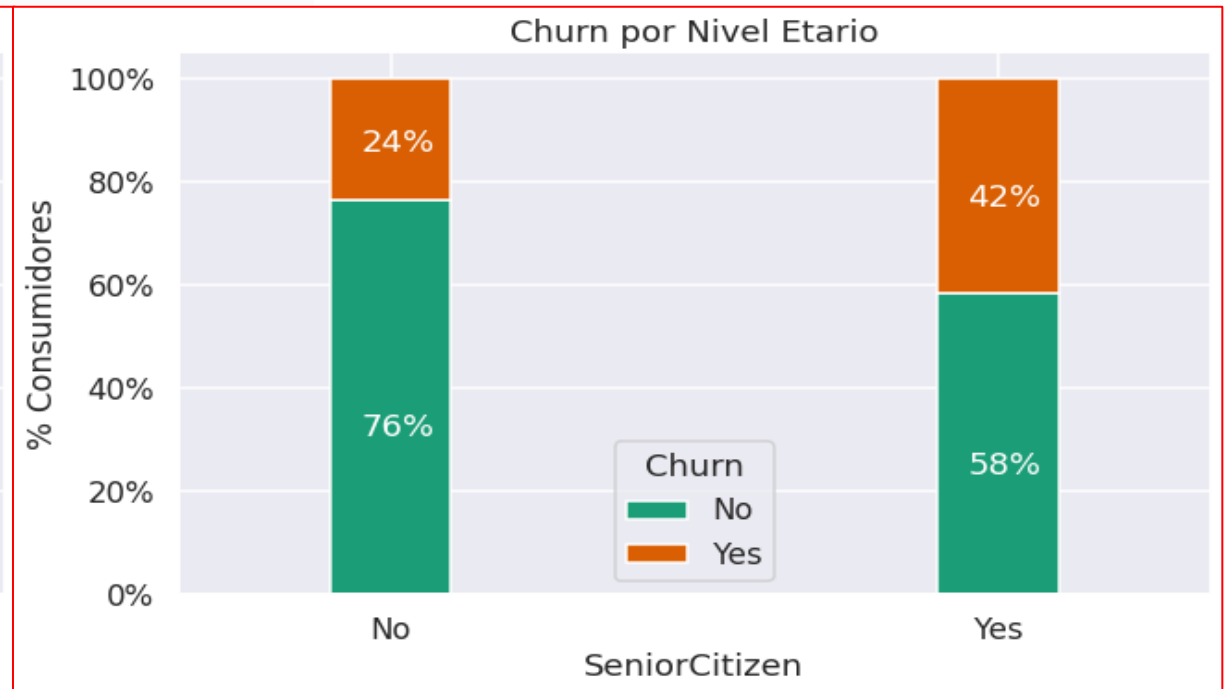
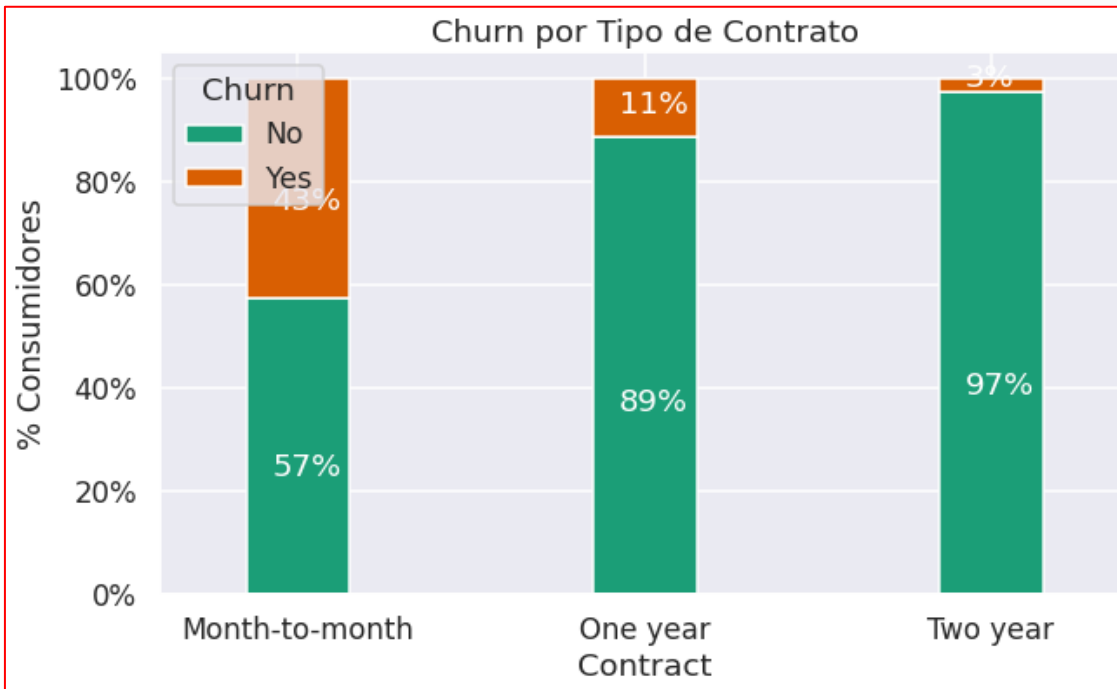
La mayoría de clientes (cerca a 4000) tienen contratos mes a mes (1 a 2 meses) luego de esto pueden seguir en la empresa por mejores promociones o simplemente salir (churn), la otra cara de la moneda es los contratos de 1 año o 2 años con 3000 clientes aproximadamente, esto de cierta forma asegura la fidelidad de los clientes y la retención de los mismos por ese periodo. Hay un gran número de clientes con contratos de 01 meses y con contrato de 02 años, su permanencia parece estar relacionada directamente con el tiempo del contrato (figura derecha).



2. Insights (Relación de Algunas Variables con Churn Rate)

Los clientes que tienen un contrato mensual tienen una tasa de rotación (churn rate) muy alta. Esto proporciona una valiosa aportación al departamento de ventas / marketing

Los clientes mayores (edad) tienen casi el doble de la tasa de rotación que la población más joven, por lo que se podría generar promociones o campañas con alguna temática para este tipo de clientes.



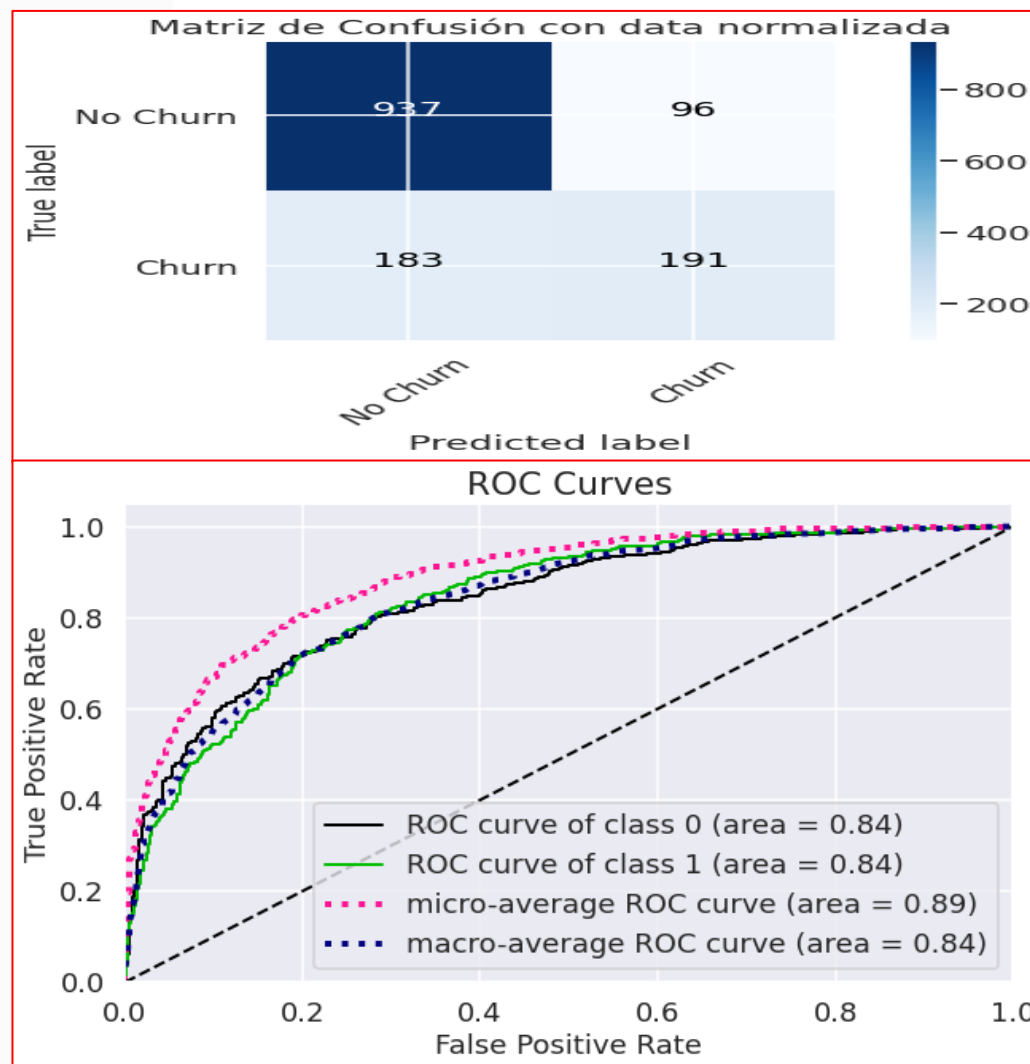
3. Modelo Predictivo (Logistic Regression)



PREDICTOR DESBALANCEADO

MODELO	REGRESION LOGISTICA
VARIABLE DEPENDIENTE	CHURN (SI-NO)
PREDICCCION	80%
PRECISION	66%
CLASIFICACION (RECALL)	51%

- La precisión (accuracy) del modelo es del 80%. Sin embargo, a veces puede no ser la medida correcta, cuando los datos están sesgados (el desbalance del % de churn de los clientes).
- El área bajo la curva ROC (AUROC) representa la probabilidad de que el modelo distinga las observaciones entre dos clases. Se puede señalar que ambas curvas están por encima de 0.5, lo que sugiere que ambos modelos tienen la capacidad de distinguir entre las observaciones de cada clase. AUC es 0.84 lo que significa que hay un 84% de posibilidades de que el modelo pueda distinguir entre Churn y No Churn.

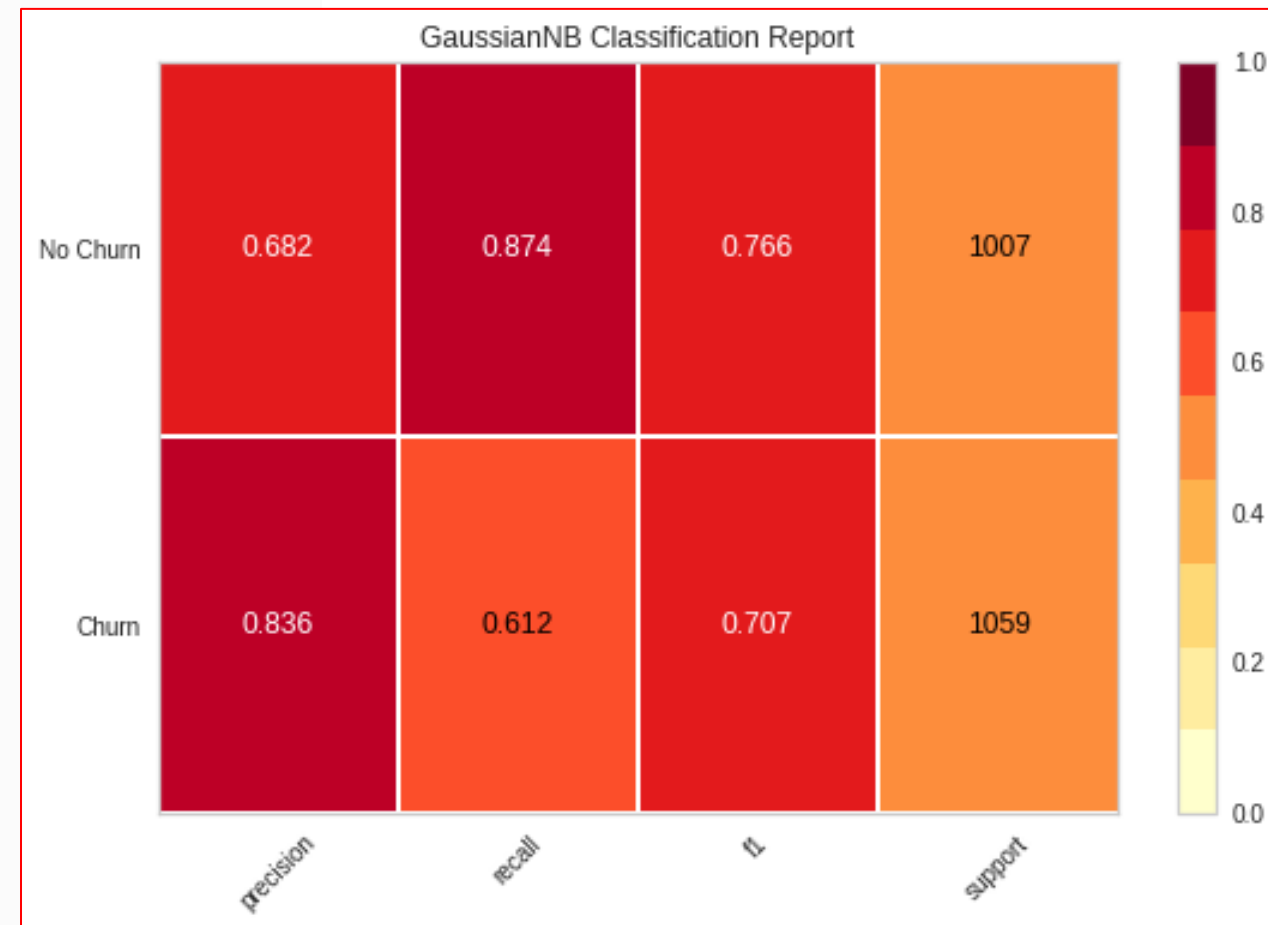


3. Modelo Predictivo (Logistic Regression)

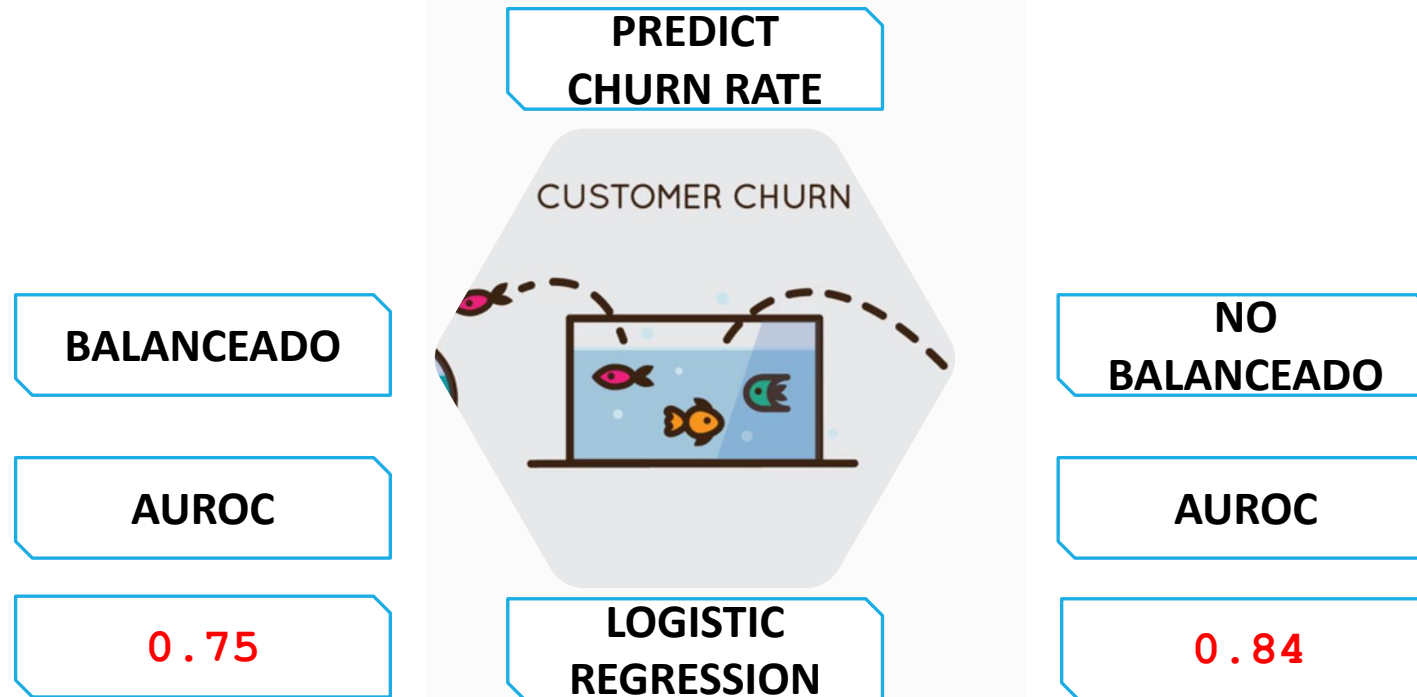


PREDICTOR BALANCEADO

Del resultado anterior se puede ver que la precisión general del modelo ha disminuido, pero la precisión y los puntajes de recuperación para predecir una rotación han mejorado. La precisión del modelo para No Churn es 69% es menor que para Churn 83% pero la precisión (recall) es mayor 86% que Churn 62%. El puntaje F1 para el NoChurn es más alto 76.6%, mientras que para el Churn es del 70%. La precisión general del modelo ha disminuido, pero la precisión y los puntajes de recuperación para predecir un abandono han mejorado en este modelo.



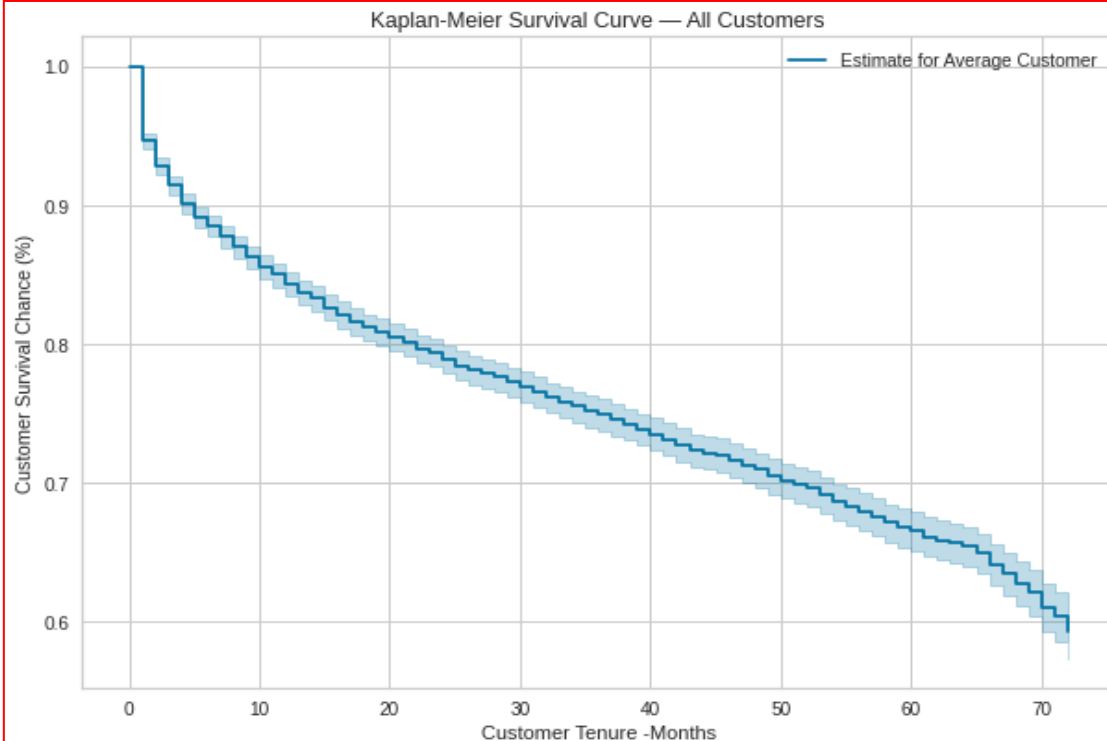
3. Modelo Predictivo (Comparaciones entre Balanceado y No Balanceado)



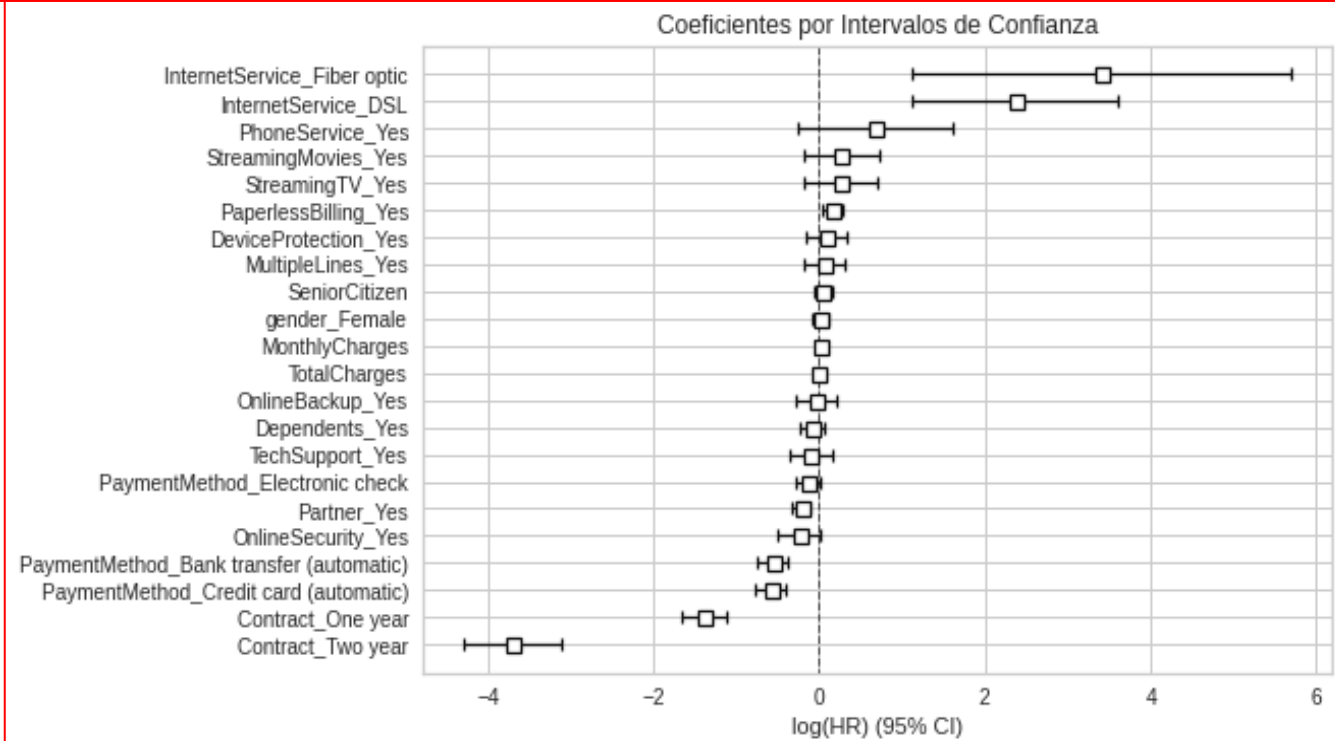
- Los puntajes de AUROC son mayores a 50% en ambos modelos. El modelo no balanceado (desequilibrado) es mayor a 80%. Esto sugiere que ambos modelos tienen la capacidad de distinguir entre observaciones de cada clase.

4. Análisis de Supervivencia (Survival Analysis)

El gráfico, proporciona una intuición básica sobre los clientes. Si no situamos al inicio del gráfico, dentro de 1-2 años, la compañía retuvo casi el 80-90% de sus clientes, pero hay una disminución constante y después de más de 70 meses, la compañía puede retener casi el 60% de sus clientes



Utilizando el modelo de riesgos proporcionales de Cox, podemos identificar las características más significativas que impactan positivamente en la supervivencia, siendo estas: (i) Contrato de 02 años, (ii) Contrato de 01 año, (iii) pago con tarjeta de crédito y (iv) pago con transferencia bancaria.



4. Conclusiones

- El análisis descriptivo proporciono dos insights relevantes: (i) Los clientes que tienen un contrato mensual tienen una tasa de rotación (churn rate) muy alta. Esto proporciona una valiosa aportación al departamento de ventas / marketing y (ii) Los clientes mayores (edad) tienen casi el doble de la tasa de rotación que la población más joven, por lo que se podría generar promociones o campañas con alguna temática para este tipo de clientes.
- La precisión (accuracy) del modelo predictivo logístico es del 80%. Los puntajes de AUROC obtenidos para el modelo predictivo balanceado y no balanceado son mayores a 50% en ambos modelos. El modelo no balanceado (desequilibrado) es mayor a 80%. Esto sugiere que ambos modelos tienen la capacidad de distinguir entre observaciones de cada clase.
- Del análisis de supervivencia se pudo establecer que en un rango 1-2 años, la compañía retuvo casi el 80-90% de sus clientes, pero hay una disminución constante y después de más de 70 meses, la compañía puede retener casi el 60% de sus clientes. Asimismo se identifico los factores de riesgo que afectan la probabilidad de churn siendo estos (i) Contrato de 02 años, (ii) Contrato de 01 año, (iii) pago con tarjeta de crédito y (iv) pago con transferencia bancaria. Las partes interesadas deben comprender cuánto pueden gastar para mantener a los clientes y comparar sus tasas de supervivencia con clientes similares a ellos, que en cambio tienen cada una de estas cuatro características presente

Gracias

EL NOTEBOOK DEL PRESENTE TRABAJO SE ENCUENTRA EN EL
SIGUIENTE ENLACE DE GITHUB [https://github.com/JLaura011/Business-
Analytics/blob/master/Churn%2BSurvival.ipynb](https://github.com/JLaura011/Business-Analytics/blob/master/Churn%2BSurvival.ipynb)