



Proyecto Final Data Science

Comision 46270

CoderHouse

Alexander La Cruz

Tabla de Contenidos

01

Introducción

02

Objetivos y Audiencia

03

Dataset Utilizado

04

Hipotesis

05

Respuestas y
Visualizaciones

06

Resumen





01

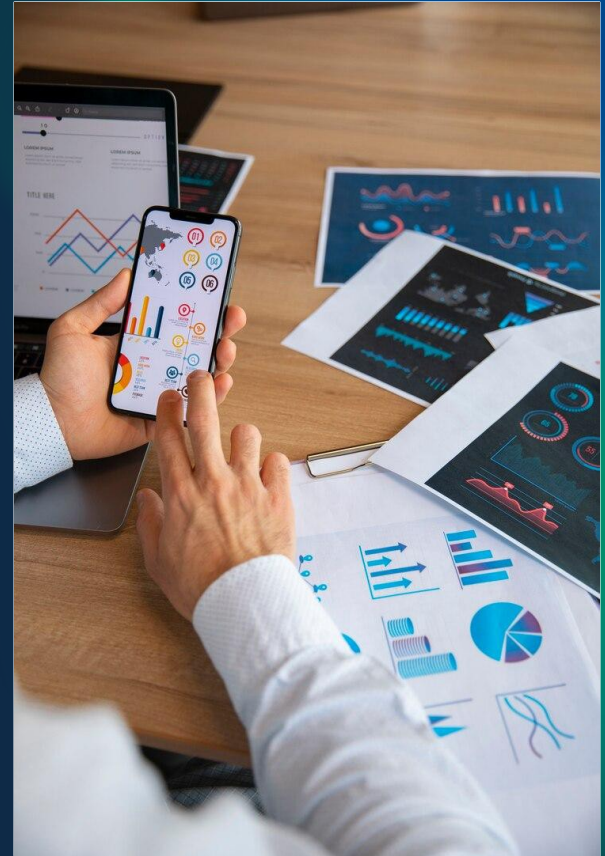
Introducción

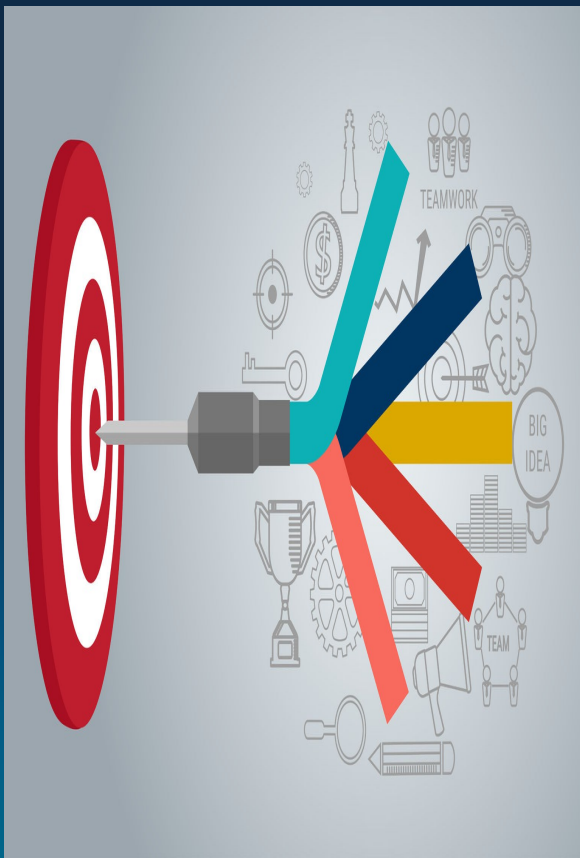
Analisis de Datos; Fidelizando Clientes



Predecir la pérdida de clientes es fundamental para las empresas puedan retener clientes de forma eficaz. Es más costoso adquirir nuevos que retener a los existentes. Por este motivo, las grandes corporaciones de telecomunicaciones por ejemplo, buscan desarrollar modelos para predecir cuales de sus usuarios, tienen más probabilidades de terminar sus contratos y tomar medidas en consecuencia.

En este proyecto, construimos un modelo para predecir la probabilidad de que un cliente abandone analizando datos de usuarios tales como: (1) información demográfica, (2) información de su cuenta, (3) información de los servicios contratados. El objetivo es obtener una solución basada en datos que permita reducir las tasas de abandono (churn rate en ingles) buscando, aumentar la satisfacción del cliente y los ingresos de la empresa.





02

Objetivos y Audiencia

Objetivos



Identificar cuales son los factores que influyen en la tasa de abandono.



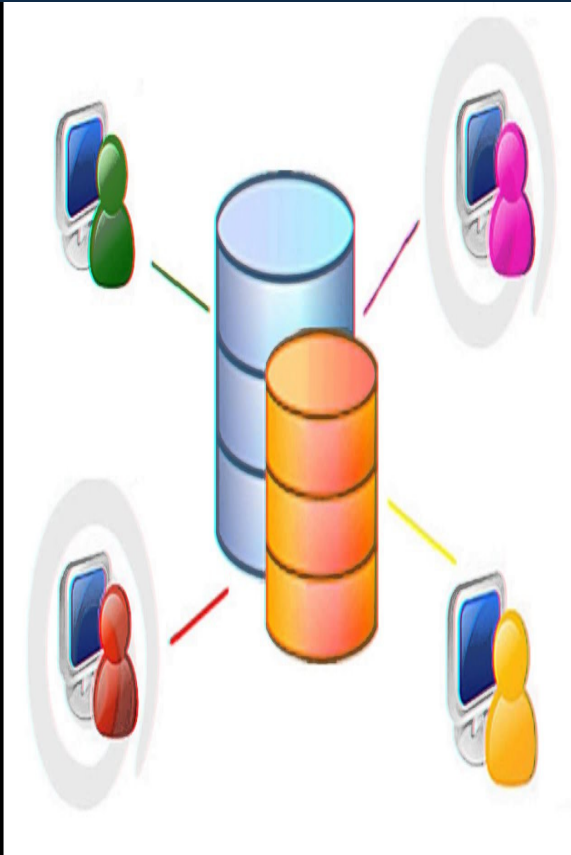
Desarrollar un modelo predictivo usando machine learning que permita predecir la probabilidad de que un cliente abandone su contrato con la empresa

Audiencia



Este proyecto está dirigido a cualquier persona interesada en aprender técnicas de análisis de datos y modelado predictivo. También podría ser útil para gerentes de empresas de servicios interesados en identificar los factores que influyen en la tasa de abandono de los clientes.





03

Dataset Utilizado



Conjunto de Datos Utilizados

El conjunto de datos utilizado para este proyecto está disponible en Kaggle; consta de 21 columnas (variables independientes) y 7043 registros que indican las características de los clientes de una ficticia corporación de telecomunicaciones.

La columna de Churn (variable objetivo) indica si el cliente termino su contrato no. La clase NO, incluye a los clientes que no abandonaron la empresa, mientras que la clase SI, contiene a los clientes que decidieron terminar sus relaciones con la empresa. El objetivo del análisis es obtener la relación entre las características del cliente y el abandono.



Descripción de las variables

Información Demográfica

- gender: Sexo del cliente.
- SeniorCitizen: indica si es un adulto mayor o no.
- Partner: indica si tiene una pareja o no.
- Dependents: indica tiene dependientes o no.

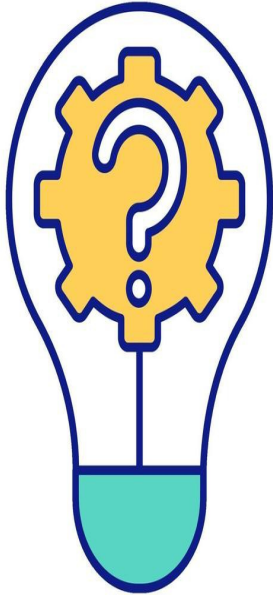
Información de Cuenta

- tenure: indica la cantidad de meses que ha sido cliente de la empresa.
- Contract: indica el tipo de contrato que tiene.
- PaperlessBilling: indica si recibe facturas electrónicas o no.
- PaymentMethod: indica el método de pago que utiliza.
- MonthlyCharges: indica la cantidad que paga cada mes.
- TotalCharges: indica la cantidad total que ha pagado hasta la fecha.

Información de Servicios

- PhoneService
- MultipleLines
- InternetService
- OnlineSecurity
- OnlineBackup
- DeviceProtection
- TechSupport
- StreamingTV
- StreamingMovies





04

Hipótesis

Preguntas/Problema que buscamos resolver



El objetivo principal de este proyecto es desarrollar un modelo predictivo para identificar a los clientes que tienen más probabilidades de abandonar la empresa de telecomunicaciones, basandonos en un conjunto de características. Algunas preguntas específicas que buscamos responder son:

1. ¿La edad, el sexo del cliente influye en la tasa de abandono?
2. ¿Hay alguna relación entre la duración del contrato y/o el tipo de pago utilizado con la tasa de abandono?
3. ¿Cómo afecta a la tasa de abandono, los servicios adicionales que el cliente tenga contratados?
4. ¿Por tanto, cuáles son los factores que más influyen en la tasa de abandono de los clientes?

Para responder a estas preguntas, se realizará un análisis exploratorio de datos, se creará un modelo predictivo utilizando técnicas de aprendizaje automático y se evaluará su precisión.





05

Repuestas y Visualizaciones

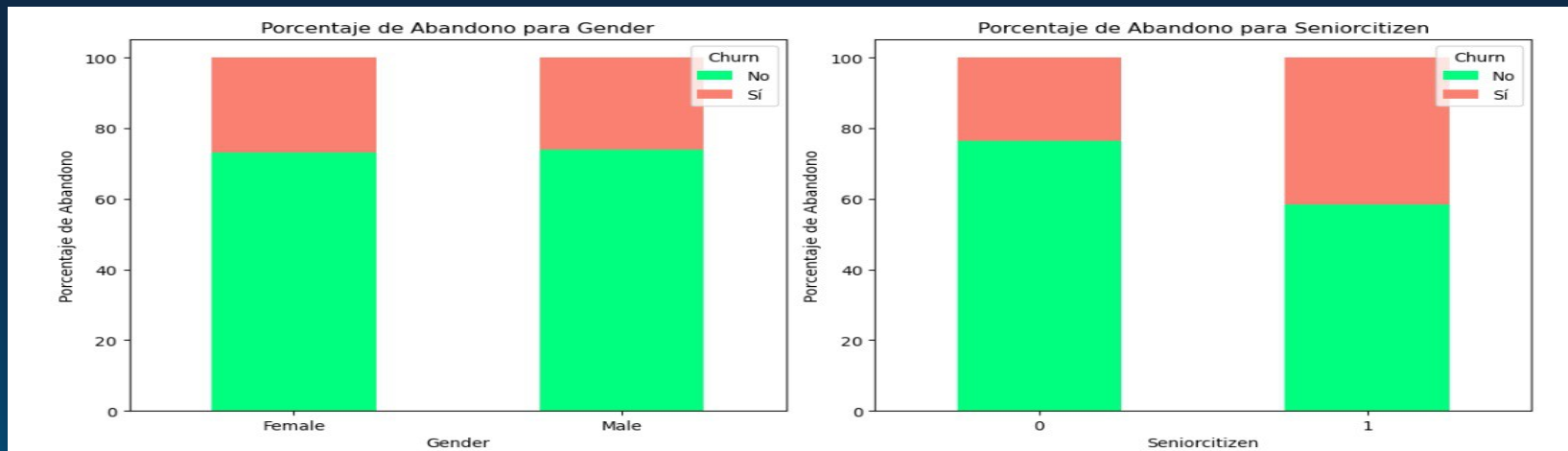
Preguntas/Problema que buscamos resolver



1. ¿La edad, el sexo del cliente influye en la tasa de abandono?

Podemos extraer las siguientes conclusiones analizando los atributos demográficos:

- La tasa de abandono de las personas mayores es casi el doble que la de los jóvenes.
- No esperamos que el género tenga un poder predictivo significativo. Un porcentaje similar de abandono se muestra tanto cuando un cliente es hombre como si es mujer.



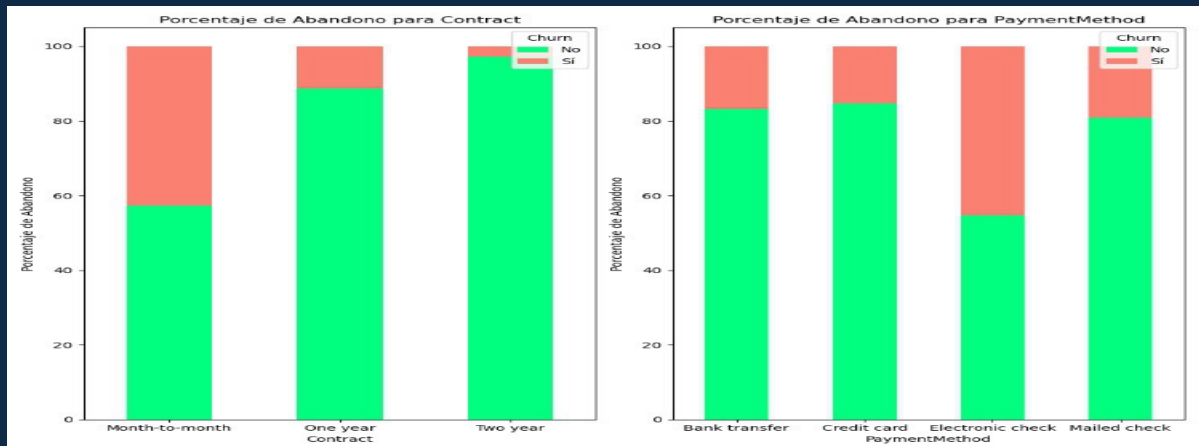
Preguntas/Problema que buscamos resolver



2. ¿Hay alguna relación entre la duración del contrato y/o el tipo de pago utilizado con la tasa de abandono?

Analizando los atributos de cuenta cliente podemos deducir que:

- Los clientes con contratos de mes a mes tienen tasas de abandono más altas en comparación con los clientes con contratos anuales.
- Los clientes que optaron por un cheque electrónico como método de pago tienen más probabilidades de abandonar su contrato.



Preguntas/Problema que buscamos resolver

3. ¿Cómo afecta a la tasa de abandono, los servicios adicionales que el cliente tenga contratados?

Podemos extraer las siguientes conclusiones evaluando los atributos relacionados con los servicios contratados:

- No esperamos que los atributos PhoneService y MultipleLine tengan un poder predictivo significativo. El porcentaje de abandono de todas las clases en ambas variables es casi el mismo.
- Los clientes con OnlineSecurity abandonan menos que aquellos que no la tienen.
- Los clientes sin TechSupport tienden a abandonar con más frecuencia que aquellos con soporte técnico.



Preguntas/Problema que buscamos resolver



4. ¿Por tanto, cuáles son los factores que más influyen en la tasa de abandono de los clientes?

Luego de realizar un análisis exhaustivo de los datos, se pueden identificar varios factores que influyen en la tasa de abandono de los clientes.

- En primer lugar, la duración del contrato parece ser un factor clave, ya que los clientes con contratos más cortos tienen una tasa de abandono significativamente mayor que los clientes con contratos más largos.
- Otro factor importante es el tipo de servicio contratado, donde los clientes que contratan servicios de Internet de fibra óptica tienen una tasa de abandono más alta en comparación con los clientes que contratan servicios de DSL o sin conexión a Internet.
- También se puede observar que los clientes que tienen servicios de seguridad en línea como "OnlineSecurity" o "DeviceProtection" tienen una tasa de abandono más baja que los clientes que no tienen estos servicios.
- Los clientes que utilizan servicios de pago electrónicos como "Electronic check" tienen una tasa de abandono más alta que los clientes que utilizan otros métodos de pago.



06

Resumen



Resumen

Como resumen final; en este proyecto se trabajó con un conjunto de datos que contiene información de clientes de una compañía de telecomunicaciones, con el objetivo de predecir la tasa de abandono. A través de un análisis exploratorio de datos, se identificaron varios factores que influyen en la tasa de abandono de los clientes, como la duración del contrato, el cargo mensual y el tipo de servicio contratado. Comenzamos limpiando los datos y analizándolos con visualizaciones. Luego, para poder construir un modelo de aprendizaje automático, transformamos los datos categóricos en variables numéricas (ingeniería de características) y creamos nuevas columnas. Después de transformar los datos, probamos 6 algoritmos de aprendizaje automático diferentes utilizando parámetros predeterminados. Finalmente, combinamos los dos mejores modelos resultantes y ajustamos los hiperparámetros de ambos para la optimización del modelo combinado, obteniendo una precisión de casi el 80% (cerca de un 6% más que la línea de base).

