Desafío Telecom X

El desafío Telecom X ofrece una oportunidad única para aplicar habilidades esenciales de análisis de datos en un escenario de negocios real.

Aplicación práctica del conocimiento

La limpieza y tratamiento de datos es una habilidad fundamental para cualquier analista de datos. La manipulación de grandes volúmenes de información exige la capacidad de identificar y corregir inconsistencias en los datos, como valores nulos, duplicados y datos fuera de estándar. Garantizar que los datos estén listos para el análisis es un paso esencial para obtener resultados precisos y confiables.

El análisis exploratorio de datos (EDA) es una etapa crucial para comprender en profundidad los datos. La capacidad de aplicar estadísticas descriptivas y generar visualizaciones permite identificar patrones, tendencias y relaciones entre las variables. Esto ayuda a formular hipótesis y generar insights que pueden influir en decisiones estratégicas dentro de la empresa.

Al participar en este desafío, aplicarás conocimientos esenciales para el análisis de grandes volúmenes de datos en un contexto real, donde tus hallazgos podrán impactar directamente en las estrategias de la empresa para mejorar el principal problema que están enfrentando.

Este desafío no solo contribuye a tu crecimiento en el área de Data Science, sino que también ofrece la oportunidad de entender cómo la ciencia de datos puede aplicarse para resolver problemas reales que enfrentan las empresas en el mercado.

Paso 1:  
Telecom X - Análisis de Evasión de Clientes

Has sido contratado como asistente de análisis de datos en Telecom X y formarás parte del proyecto "Churn de Clientes". La empresa enfrenta una alta tasa de cancelaciones y necesita comprender los factores que llevan a la pérdida de clientes.

Tu desafío será recopilar, procesar y analizar los datos, utilizando Python y sus principales bibliotecas para extraer información valiosa. A partir de tu análisis, el equipo de Data Science podrá avanzar en modelos predictivos y desarrollar estrategias para reducir la evasión.

¿Qué vas a practicar?

✅ Importar y manipular datos desde una API de manera eficiente.

✅ Aplicar los conceptos de ETL (Extracción, Transformación y Carga) en la preparación de los datos.

✅ Crear visualizaciones estratégicas para identificar patrones y tendencias.

✅ Realizar un Análisis Exploratorio de Datos (EDA) y generar un informe con insights relevantes.

¡Ahora es tu turno! 🚀 Usa tus conocimientos para transformar datos en información estratégica y ayudar a Telecom X a retener más clientes.

Paso 2:

Descripción

Editar

Para iniciar tu análisis, necesitarás importar los datos de la API de Telecom X. Estos datos están disponibles en formato JSON y contienen información esencial sobre los clientes, incluyendo datos demográficos, tipo de servicio contratado y estado de evasión.

📌 Enlace de la API:

🔗 challenge2-data-science-LATAM/TelecomX\_Data.json at main · ingridcristh/challenge2-data-science-LATAM

🔗GitHub - ingridcristh/challenge2-data-science-LATAM

¿Qué debes hacer?

✅ Cargar los datos directamente desde la API utilizando Python.

✅ Convertir los datos a un DataFrame de Pandas para facilitar su manipulación.

Este es el primer paso para transformar los datos en información valiosa. ¿Listo para programar? 🚀

Paso 3:  
Ahora que has extraído los datos, es fundamental comprender la estructura del dataset y el significado de sus columnas. Esta etapa te ayudará a identificar qué variables son más relevantes para el análisis de evasión de clientes.

📌 Para facilitar este proceso, hemos creado un diccionario de datos con la descripción de cada columna. Aunque no es obligatorio utilizarlo, puede ayudarte a comprender mejor la información disponible.

🔗 Enlace al diccionario y a la API

¿Qué debes hacer?

✅ Explorar las columnas del dataset y verificar sus tipos de datos.

✅ Consultar el diccionario para comprender mejor el significado de las variables.

✅ Identificar las columnas más relevantes para el análisis de evasión.

📌 Tips:

🔗 Documentación de DataFrame.info()

🔗 Documentación de DataFrame.dtypes

Paso 4:

En este paso, verifica si hay problemas en los datos que puedan afectar el análisis. Presta atención a valores ausentes, duplicados, errores de formato e inconsistencias en las categorías. Este proceso es esencial para asegurarte de que los datos estén listos para las siguientes etapas.

📌 Tips:

🔗 Documentación de pandas.unique()

🔗 Documentación de pandas.Series.dt.normalize()

Paso 5:

Ahora que has identificado las inconsistencias, es momento de aplicar las correcciones necesarias. Ajusta los datos para asegurarte de que estén completos y coherentes, preparándolos para las siguientes etapas del análisis.

📌 Tips:

🔗 Manipulación de strings en pandas: lower, replace, startswith y contains | Alura Cursos Online

Paso 6:

Ahora que los datos están limpios, es momento de crear la columna "Cuentas\_Diarias". Utiliza la facturación mensual para calcular el valor diario, proporcionando una visión más detallada del comportamiento de los clientes a lo largo del tiempo.

📌 Esta columna te ayudará a profundizar en el análisis y a obtener información valiosa para las siguientes etapas.

Paso 7:

La estandarización y transformación de datos es una etapa opcional, pero altamente recomendada, ya que busca hacer que la información sea más consistente, comprensible y adecuada para el análisis. Durante esta fase, por ejemplo, puedes convertir valores textuales como "Sí" y "No" en valores binarios (1 y 0), lo que facilita el procesamiento matemático y la aplicación de modelos analíticos.

Además, traducir o renombrar columnas y datos hace que la información sea más accesible y fácil de entender, especialmente cuando se trabaja con fuentes externas o términos técnicos. Aunque no es un paso obligatorio, puede mejorar significativamente la claridad y comunicación de los resultados, facilitando la interpretación y evitando confusiones, especialmente al compartir información con stakeholders no técnicos.

Paso 8:

ara comenzar, realiza un análisis descriptivo de los datos, calculando métricas como media, mediana, desviación estándar y otras medidas que ayuden a comprender mejor la distribución y el comportamiento de los clientes.

📌 Consejos:

🔗 Documentación de DataFrame.describe()

Paso 9:

En este paso, el objetivo es comprender cómo está distribuida la variable "churn" (evasión) entre los clientes. Utiliza gráficos para visualizar la proporción de clientes que permanecieron y los que se dieron de baja.

Paso 10:

Ahora, exploraremos cómo se distribuye la evasión según variables categóricas, como género, tipo de contrato, método de pago, entre otras.

Este análisis puede revelar patrones interesantes, por ejemplo, si los clientes de ciertos perfiles tienen una mayor tendencia a cancelar el servicio, lo que ayudará a orientar acciones estratégicas.

Paso 11:

Ahora, exploraremos cómo se distribuye la evasión según variables categóricas, como género, tipo de contrato, método de pago, entre otras.

Este análisis puede revelar patrones interesantes, por ejemplo, si los clientes de ciertos perfiles tienen una mayor tendencia a cancelar el servicio, lo que ayudará a orientar acciones estratégicas.

Paso 12:

En este paso, explora cómo las variables numéricas, como "total gastado" o "tiempo de contrato", se distribuyen entre los clientes que cancelaron (evasión) y los que no cancelaron.

Este análisis ayuda a entender si ciertos valores numéricos están más asociados con la evasión, proporcionando insights sobre los factores que influyen en el comportamiento de los clientes.

Paso 13:

Finaliza el desafío elaborando un informe dentro del mismo notebook que resuma todo el trabajo realizado. El informe debe incluir:

🔹 Introducción: Explica el objetivo del análisis y el problema de evasión de clientes (Churn).

🔹 Limpieza y Tratamiento de Datos: Describe los pasos realizados para importar, limpiar y procesar los datos.

🔹 Análisis Exploratorio de Datos: Presenta los análisis realizados, incluyendo gráficos y visualizaciones para identificar patrones.

🔹 Conclusiones e Insights: Resume los principales hallazgos y cómo estos datos pueden ayudar a reducir la evasión.

🔹 Recomendaciones: Ofrece sugerencias estratégicas basadas en tu análisis.

Asegúrate de que el informe esté bien estructurado, claro y respaldado por visualizaciones que refuercen tus conclusiones. 🚀

Hecho. Aquí está el informe final del desafío Telecom X, redactado en formato LaTeX para Markdown, listo para ser integrado en tu notebook.

Fragmento de código

# Informe de Análisis de Evasión de Clientes (Churn) - Telecom X

---

## 🔹 Introducción

El presente informe detalla el análisis del problema de \*\*evasión de clientes (Churn)\*\* en la empresa de telecomunicaciones Telecom X. El objetivo principal es identificar y comprender los factores clave que llevan a los clientes a cancelar sus servicios, con el fin de desarrollar estrategias efectivas para la retención. Mediante la aplicación de técnicas de \*\*Análisis Exploratorio de Datos (EDA)\*\* y la manipulación de información de clientes, buscamos obtener \*insights\* valiosos que impacten directamente en las decisiones de negocio.

---

## 🔹 Limpieza y Tratamiento de Datos

El análisis comenzó con la importación de datos desde la API de Telecom X. A continuación, se realizaron los siguientes pasos de limpieza y procesamiento:

1. \*\*Carga de datos:\*\* Se extrajo un conjunto de datos de 7267 registros con 21 columnas en formato JSON y se convirtió a un DataFrame de Pandas.

2. \*\*Manejo de nulos y duplicados:\*\* No se encontraron valores nulos significativos ni registros duplicados, lo que garantizó la calidad inicial de los datos.

3. \*\*Conversión de tipos de datos:\*\* Se identificó que la columna `account.Charges.Total` estaba en formato \*object\*. Se convirtió a un tipo numérico (`float64`) para su correcta manipulación.

4. \*\*Estandarización y traducción:\*\* Se estandarizaron y tradujeron diversas variables para facilitar el análisis. Por ejemplo:

\* Las columnas binarias como `Churn` y `customer.Partner` se convirtieron de 'Yes'/'No' a `1`/`0`.

\* Las variables categóricas `customer.gender`, `account.Contract`, `account.PaymentMethod` e `internet.InternetService` se tradujeron a sus equivalentes en español.

5. \*\*Creación de nuevas variables:\*\* Se creó la columna `Cuentas\_Diarias` para analizar el gasto promedio por día, lo cual ofrece una visión más detallada del comportamiento del cliente.

---

## 🔹 Análisis Exploratorio de Datos

### Distribución de la Evasión (Churn)

El análisis exploratorio reveló una tasa de evasión del \*\*26.54%\*\*. Esto significa que más de una cuarta parte de los clientes de Telecom X abandonan la compañía. Esta cifra subraya la urgencia del problema y la importancia de identificar los factores que lo causan.

### Factores que influyen en la Evasión

Se examinó la relación entre la evasión y variables clave del cliente:

\* \*\*Tipo de Contrato:\*\* Los clientes con contratos \*\*mensuales\*\* tienen una tasa de evasión notablemente superior en comparación con aquellos con contratos de uno o dos años. Esto sugiere que la falta de un compromiso a largo plazo es un indicador de alto riesgo de \*churn\*.

\* \*\*Servicio de Internet:\*\* La evasión es significativamente más alta entre los clientes que utilizan el servicio de \*\*Fibra óptica\*\* que aquellos con DSL. Esto podría indicar problemas de calidad, precio o servicio al cliente relacionados con esta tecnología.

\* \*\*Método de Pago:\*\* Los clientes que utilizan el \*\*Cheque electrónico\*\* como método de pago son los más propensos a la evasión, con una tasa de abandono muy superior a la de otros métodos.

\* \*\*Tiempo de Contrato (Tenure):\*\* Los clientes que cancelan su servicio tienden a tener un \*\*tiempo de contrato mucho más corto\*\* que los clientes que se quedan. El 75% de los clientes que se van tenían menos de 30 meses de servicio, lo que contrasta con la media de 32.3 meses para todos los clientes.

\* \*\*Costo Mensual:\*\* El gasto mensual promedio es mayor para los clientes que evaden. Esto podría indicar que los clientes que pagan más son más sensibles al valor percibido y al servicio recibido.

---

## 🔹 Conclusiones e Insights

El análisis exploratorio reveló que la evasión en Telecom X no es uniforme y está fuertemente relacionada con el comportamiento y perfil del cliente. Los hallazgos clave son:

\* El \*\*tipo de contrato\*\* es el factor más influyente. Los contratos mensuales son un claro indicador de alto riesgo de evasión.

\* El \*\*método de pago\*\* y el \*\*servicio de Internet\*\* también juegan un rol crucial. Los cheques electrónicos y la fibra óptica se asocian con una mayor pérdida de clientes.

\* La evasión ocurre principalmente durante las \*\*primeras etapas\*\* de la relación con el cliente.

---

## 🔹 Recomendaciones

Basado en los insights obtenidos, se sugieren las siguientes acciones estratégicas para reducir la evasión:

1. \*\*Programas de fidelización:\*\* Crear campañas para incentivar a los clientes con contratos mensuales a cambiar a contratos de un año o más, ofreciendo descuentos o beneficios adicionales.

2. \*\*Optimizar el servicio de Fibra Óptica:\*\* Investigar los motivos detrás de la alta tasa de evasión de clientes con fibra óptica. Esto podría implicar revisar la calidad del servicio, la estabilidad de la conexión o la atención al cliente para este segmento.

3. \*\*Análisis de métodos de pago:\*\* Estudiar por qué los clientes que usan cheques electrónicos se van más frecuentemente. Podrían existir fricciones en el proceso de pago que deben ser resueltas.

4. \*\*Enfocarse en nuevos clientes:\*\* Implementar un programa de "bienvenida" robusto para los nuevos clientes, especialmente en los primeros seis meses, para asegurar una experiencia positiva y construir lealtad desde el principio.

Pídele a Gemini que escriba un documento o código que puedes editar

### Informe Final: Análisis de Evasión de Clientes (Churn) en Telecom X

####Introducción

El objetivo de este proyecto es analizar la evasión de clientes (Churn) en Telecom X, una de las principales preocupaciones de la empresa. La evasión se define como el abandono de los clientes del servicio. Al comprender los factores clave que contribuyen a este problema, podemos tomar decisiones estratégicas para retener a los clientes y mejorar la rentabilidad.

####Limpieza y Tratamiento de Datos

La primera fase del proyecto se centró en preparar el conjunto de datos para el análisis. Los pasos clave fueron:

- \*\*Importación\*\*: Se cargaron los datos desde una API pública en formato JSON y se normalizaron en un DataFrame de Pandas, que contenía 7267 registros y 21 columnas.

- Estandarización y Limpieza: Se identificaron inconsistencias como valores nulos en account.Charges.Total que se llenaron con el valor de account.Charges.Monthly para clientes nuevos.

- Traducción y Codificación: Se tradujeron las variables categóricas clave (Tipo\_Contrato, Metodo\_Pago, Servicio\_Internet, Genero) a español y se convirtieron las variables binarias (Yes/No) a formato numérico (1/0) para facilitar el análisis. Se renombraron las columnas para mayor claridad.

Creación de variables: Se creó la columna "Cuentas\_Diarias" a partir del costo mensual, proporcionando una nueva perspectiva sobre los gastos de los clientes.

🔹 Análisis Exploratorio de Datos

El análisis exploratorio reveló información crucial sobre el comportamiento de los clientes y su relación con la evasión.

🔸 Distribución de la Evasión (Churn)

El análisis inicial mostró que el 26.5% de los clientes han cancelado el servicio. Esto confirma que la evasión es un problema significativo que la empresa necesita abordar.

🔸 Evasión por Tipo de Contrato

La evasión es notablemente mayor en clientes con contratos mensuales (Month-to-month).

Los clientes con contratos mensuales tienen una alta tasa de evasión, lo que sugiere que no tienen un compromiso a largo plazo con la compañía.

En contraste, los clientes con contratos de "un año" y "dos años" tienen tasas de evasión mucho más bajas, lo que indica una mayor lealtad.

🔸 Evasión por Método de Pago

El método de pago con la tasa de evasión más alta es el Cheque electrónico.

El 45% de los clientes que usan cheque electrónico se dan de baja, lo que lo convierte en un factor de riesgo importante.

La transferencia bancaria y la tarjeta de crédito (métodos automáticos) tienen tasas de evasión significativamente menores.

🔸 Evasión por variables numéricas

Los clientes que cancelan tienden a tener un menor tiempo de contrato (Tiempo\_Contrato) y un mayor costo mensual (Costo\_Mensual).

Tiempo de contrato: Los clientes que cancelaron están concentrados en los primeros meses de servicio, como se ve en el boxplot.

Costo Mensual: Los clientes con costos mensuales más altos son los que tienen mayor tendencia a la evasión. Esto podría deberse a que perciben que el servicio no justifica el precio, especialmente si no tienen un contrato de largo plazo.

🔹 Conclusiones e Insights

Falta de lealtad: La evasión está fuertemente correlacionada con los contratos mensuales. Los clientes que no están atados a un compromiso a largo plazo son los más propensos a irse.

Alto costo, alta evasión: Los clientes que pagan más por su servicio mensual tienen una mayor probabilidad de cancelar.

Riesgos en el método de pago: El uso de cheques electrónicos está asociado a una alta tasa de evasión, a diferencia de los pagos automáticos que denotan mayor estabilidad del cliente.

Tiempo de permanencia: La evasión es un problema de retención temprana, ya que la mayoría de los clientes que se van lo hacen en los primeros meses.

🔹 Recomendaciones

Basado en este análisis, las siguientes acciones podrían ayudar a reducir la evasión:

Promocionar contratos a largo plazo: Crear ofertas atractivas y descuentos para incentivar a los clientes de contratos mensuales a cambiar a planes de uno o dos años.

Revisar los precios: Analizar los precios de los planes con altos costos mensuales para asegurar que el valor percibido por el cliente justifique el precio.

Incentivar métodos de pago automáticos: Ofrecer pequeñas recompensas (descuentos, puntos) a los clientes que opten por la transferencia bancaria o tarjeta de crédito para sus pagos, en lugar de cheques electrónicos.

Enfocarse en los clientes nuevos: Implementar una estrategia de retención intensiva durante los primeros seis meses de servicio, que es cuando los clientes son más vulnerables a la evasión.