

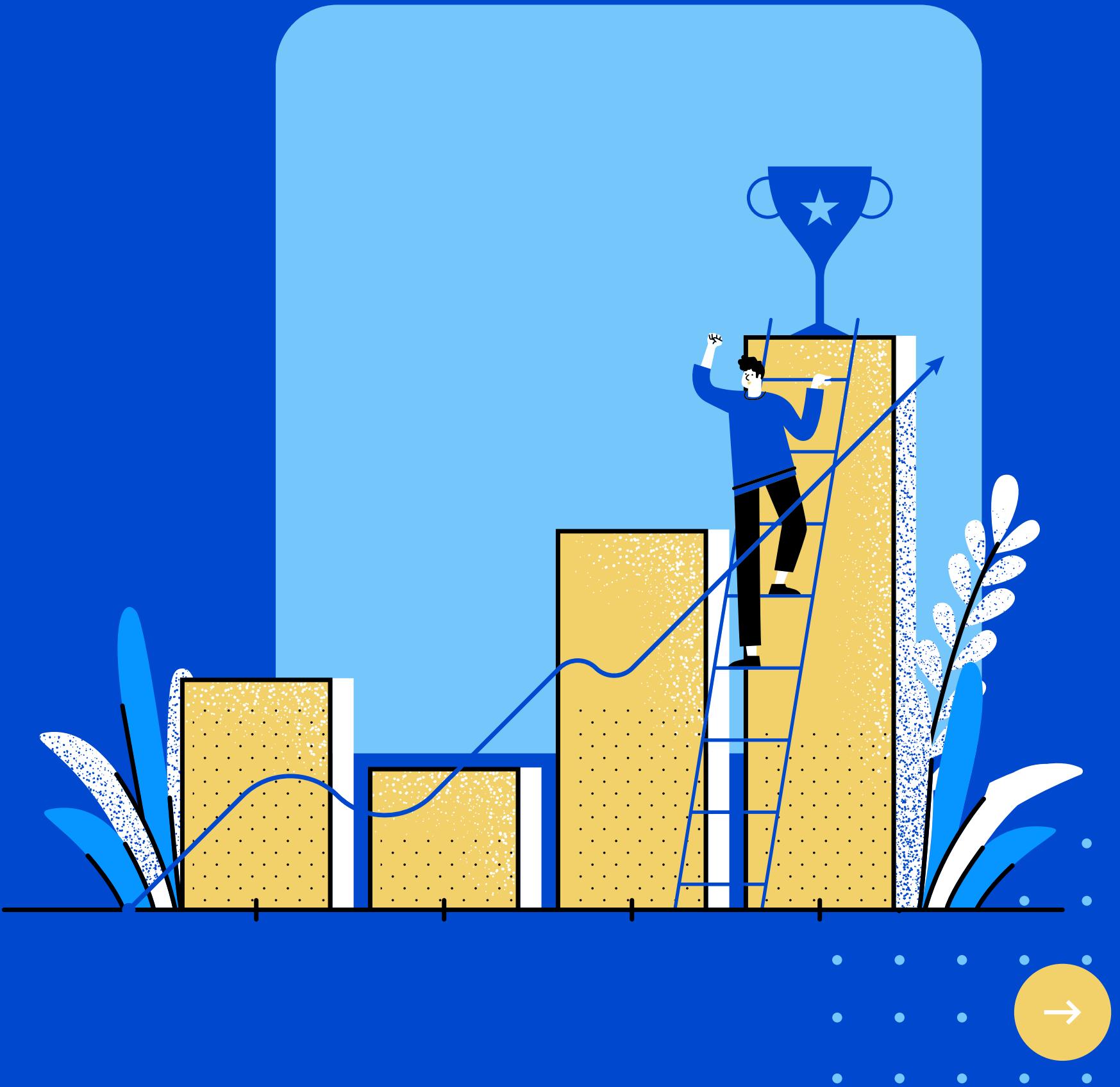
# Proyecto Churn

Caso de uso de prueba por:  
Luis Muroya



# Definición del problema

Canvas de Analytics



## Problema

Disminución del margen operativo por el incremento de 0.5 p.p. en el ratio de fuga (churn) de clientes.

## Objetivo de negocio

Disminuir el ratio de fuga de clientes promedio en 0.5 pp. en una ventana de 3 meses.

## Objetivo analítico

Identificar a los clientes con mayor propensión de fuga para campañas de retención.

## Datos disponibles

- Datos demográficos
- Datos comerciales
- Datos de servicios
- Datos de facturación
- Datos de pago

## Hipótesis

- Disponibilidad de datos continua en el tiempo.
- Comportamiento modelado permanece constante en el tiempo.
- Retención es menos costosa que adquisición de clientes.

## Solución

Modelo analítico para predicción de propensión de fuga de clientes.

## Aplicación

Uso de modelo analítico para perfilamiento de campañas de retención.

## Actores

- Equipo de Campañas
- Equipo de Analytics
- Equipo de Tecnología
- Callcenters

## KPI

- AUC (modelo)
- KS (modelo)
- Ratio de fuga de clientes (campaña)
- Uplift (campaña)

## Impacto y valor

- Disminución de tasa de fuga en grupo gestionado con ofertas de retención.
- Incremento en el margen de ingresos por servicios móviles.
- Incremento en el NPS de los clientes gestionados.

## Riesgos principales

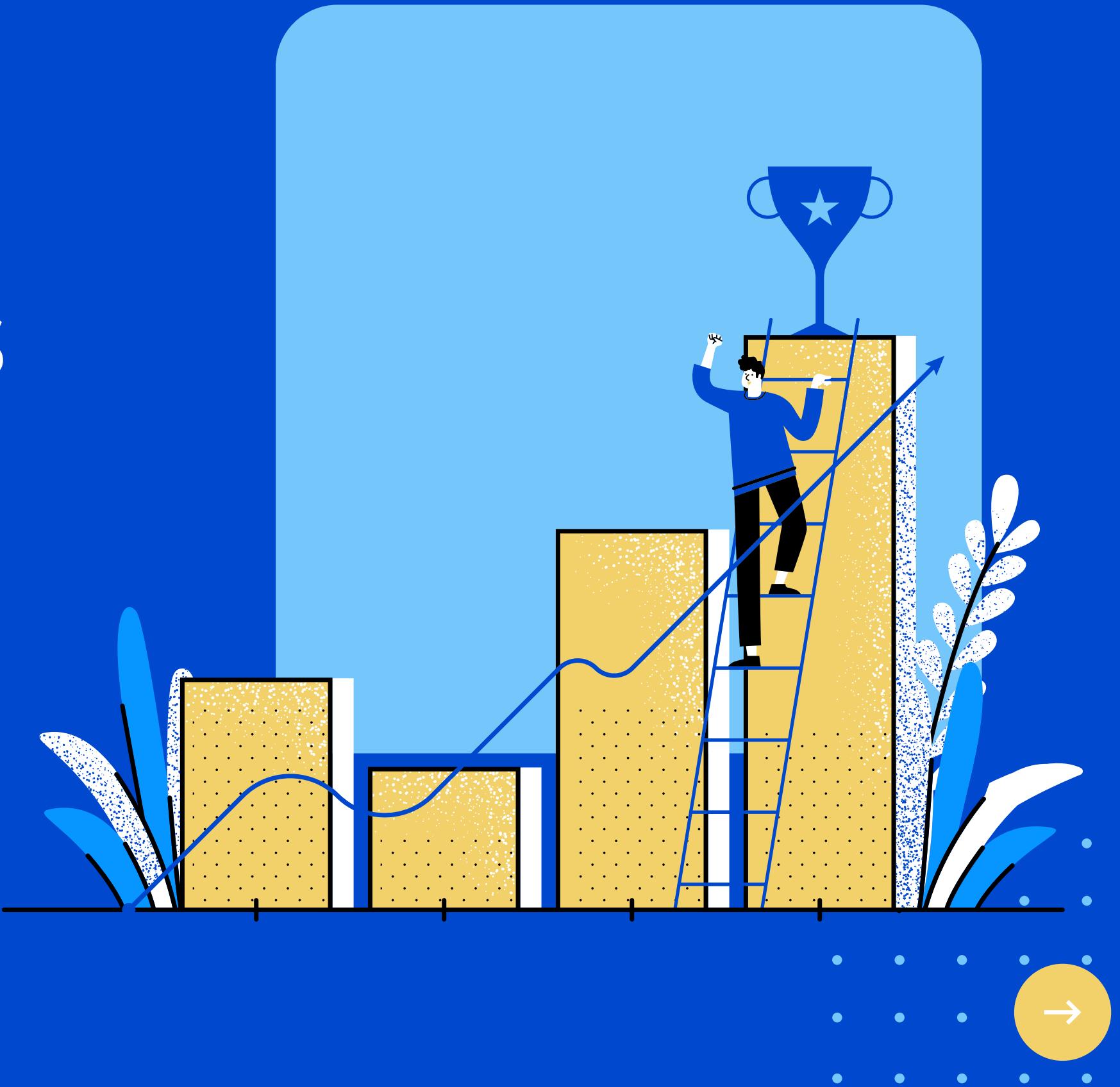
- Factores que afecten la disponibilidad de datos / continuidad del modelo.
- Factores exógenos que modifiquen el comportamiento de la población.
- Herramientas de retención sin potencia suficiente para retener al cliente.
- Factores regulatorios, de privacidad de datos y de contactabilidad al cliente.

## Acciones

- Análisis de datos.
- Desarrollo del modelo.
- Piloto de campañas comerciales de retención (descuentos y ofertas).
- Análisis de resultados de pilotos.
- Puesta en producción de campañas.
- Automatización del modelo.
- Seguimiento del desempeño del modelo.
- Mejora continua del modelo.

# Dominios de datos

Disponibles y deseables



## DOMINIOS DE DATOS DISPONIBLES

### Comerciales

Indicadores del tipo de producto que tiene contratado el cliente: líneas, internet, plazo del contrato, monto de plan.

### Servicios

Indicadores de otros servicios que tiene contratado el cliente: backup, seguridad, tv, streaming, servicio técnico.

### Facturación

Datos sobre la facturación del cliente: modalidad de facturación, cargo mensual, cargo total.

### Pagos

Datos sobre el pago del cliente: forma de pago.

### Demográficos

Datos del cliente como persona: género, estado civil, número de hijos, grupo etario.



## DOMINIOS DE DATOS Deseables

### Consumo

Cantidad de minutos de voz traficados, cantidad de megas traficados, megas destinados al uso de apps, cantidad de llamadas al callcenter, etc. Disponible en información de redes de la empresa.

### Interacciones

Cantidad de reclamos, cantidad de llamadas al centro de atención al cliente, cantidad de visitas a tiendas, cantidad de interacciones con el chatbot, etc. Disponible si se consolida información de áreas comerciales y de experiencia de cliente.

### Comerciales

Indicadores para saber si el cliente ha sido perfilado y gestionado en alguna campaña durante los últimos meses, indicadores de respuesta a estos incentivos, si ha intentado migrar a otra empresa en el pasado, etc. Disponible consolidando información comercial.

### Financieros

Datos del comportamiento financiero del cliente: si tiene deuda activa con bancos, monto de su deuda, monto de deuda vencida, monto de deuda castigada, predictor de ingreso, etc. Disponible por compra de información a una central de riesgos.

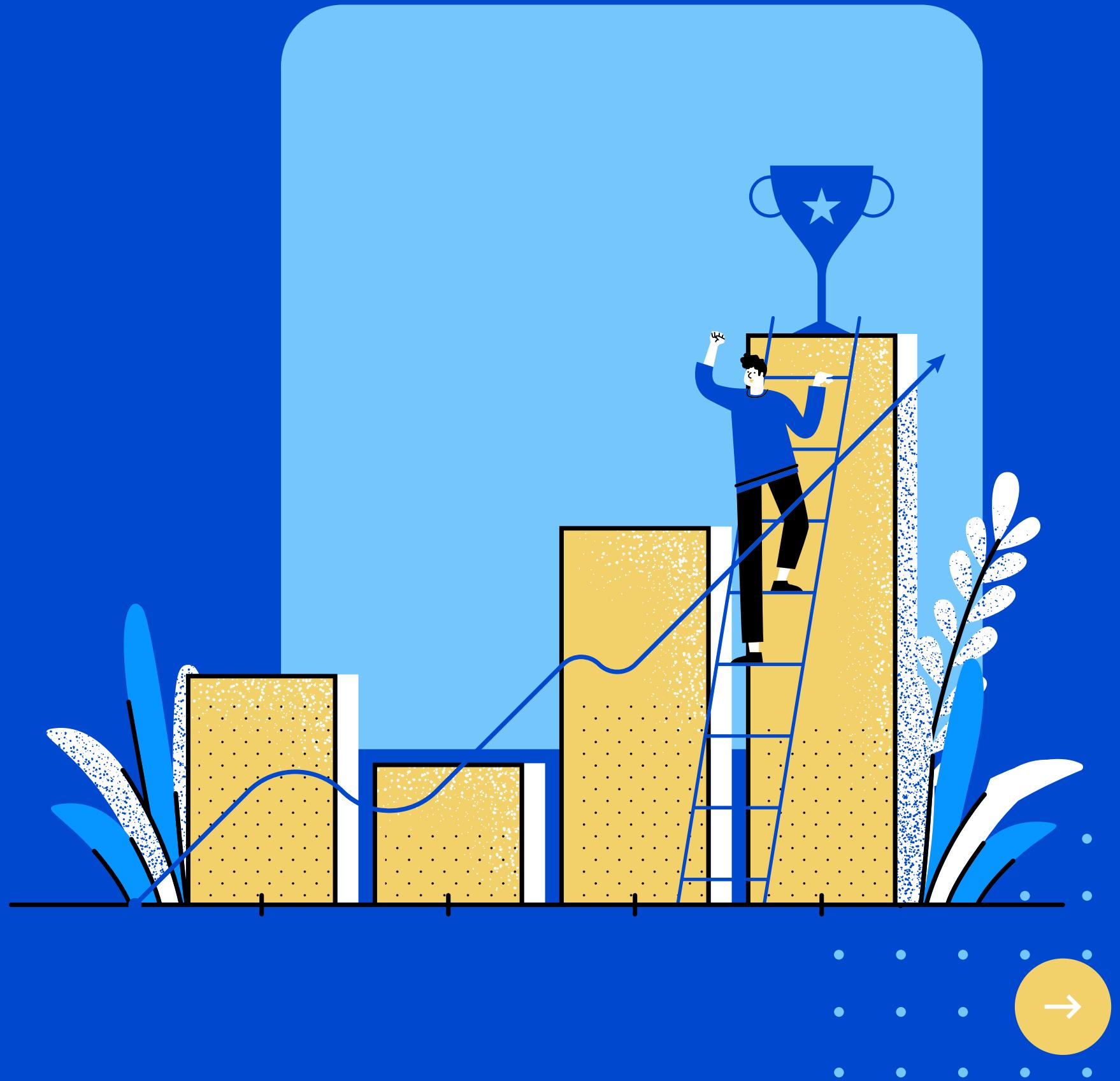
### Pagos

Datos adicionales de los pagos: días de mora, indicador si siempre paga en un mismo rango de días, indicador si hizo pago completo o pago fraccionado de deuda, etc. Disponible trabajando la información transaccional de pagos.



# Analisis Exploratorio

Dashboard de Control





Cálidda

# Tablero de Control Analítico

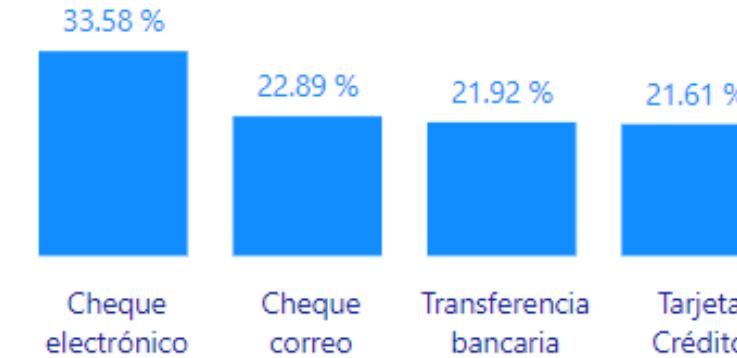
## Característica

- EstadoCivil
- FacturacionElectronica
- Genero
- Hijos
- ModoPago
- Multilinea
- Plazo
- ProteccionTerminal
- SeguridadOnline
- ServicioInternet
- ServicioMovil
- ServicioTecnico
- StreamingMovies
- StreamingTV

## Tasa de Fuga



## Proporción de base



## Resumen por Q

Valor	Cli. Total	Cli. Fugados
Cheque correo	1612	308
Cheque electrónico	2365	1071
Tarjeta Crédito	1522	232
Transferencia bancaria	1544	258

7043

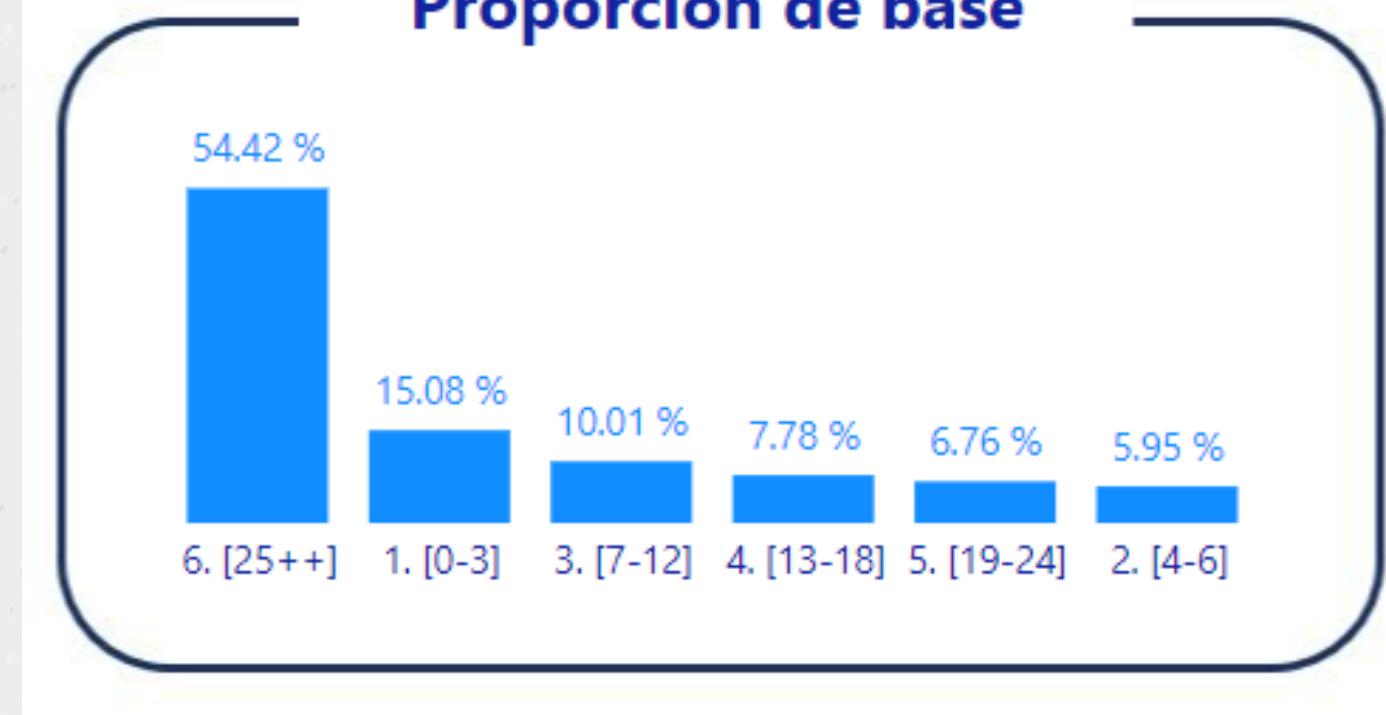
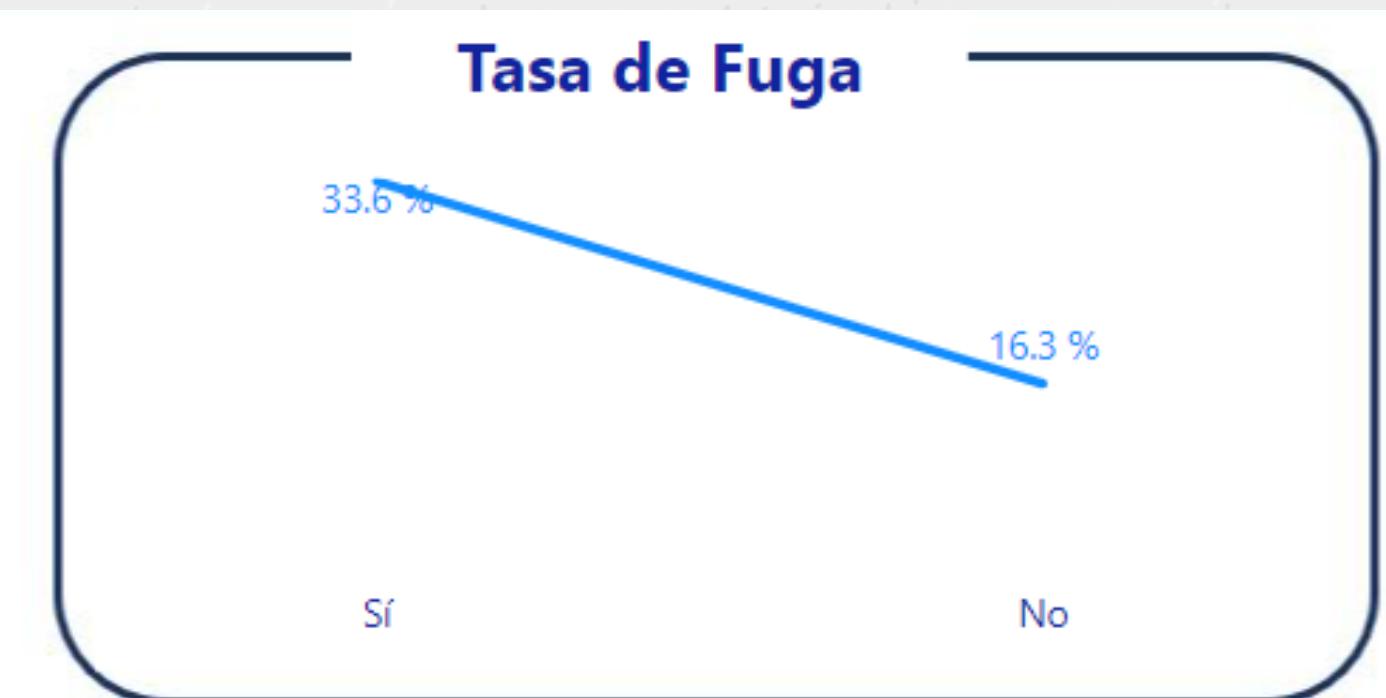
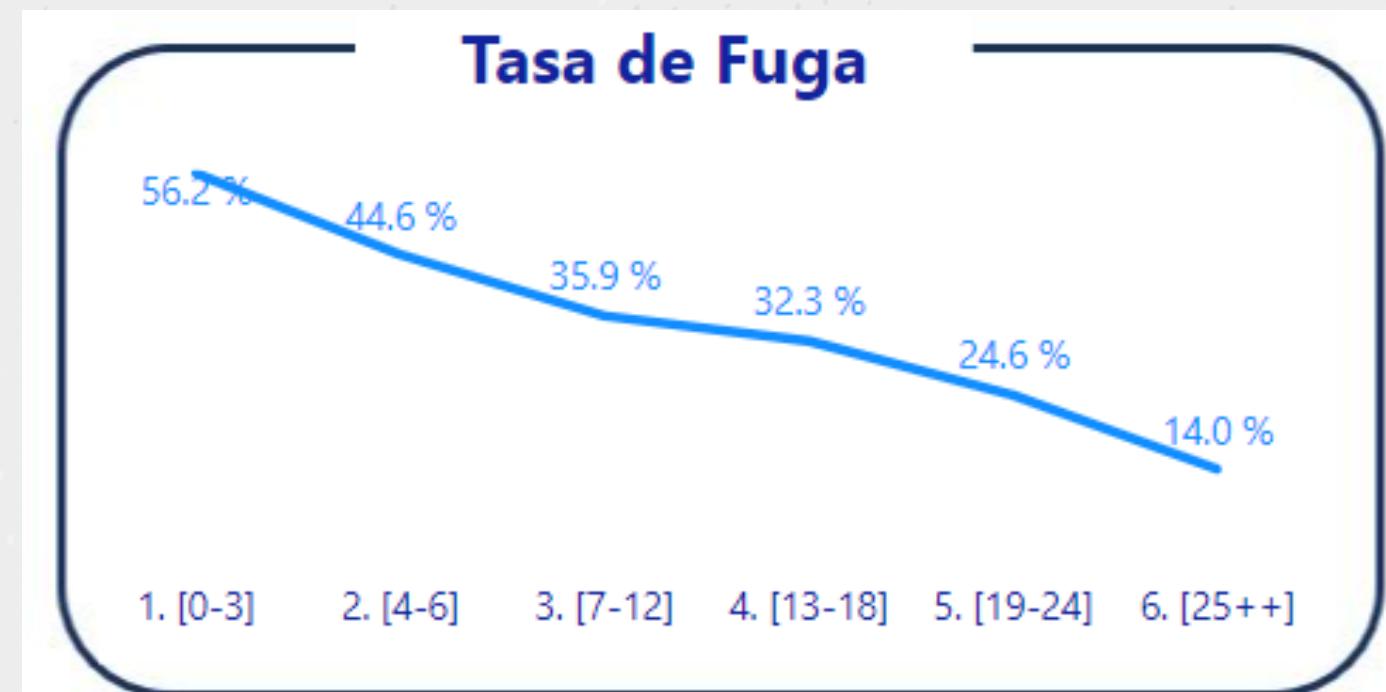
Clientes

26.5 %

Tasa de fuga promedio

## Tablero PowerBI

Se desarrolló Tablero en PowerBI para poder ver aspectos básicos de las variables disponibles. En general, se buscan variables que discriminen en la variable objetivo (fuga) y que tengan cantidades representativas de base (mix).

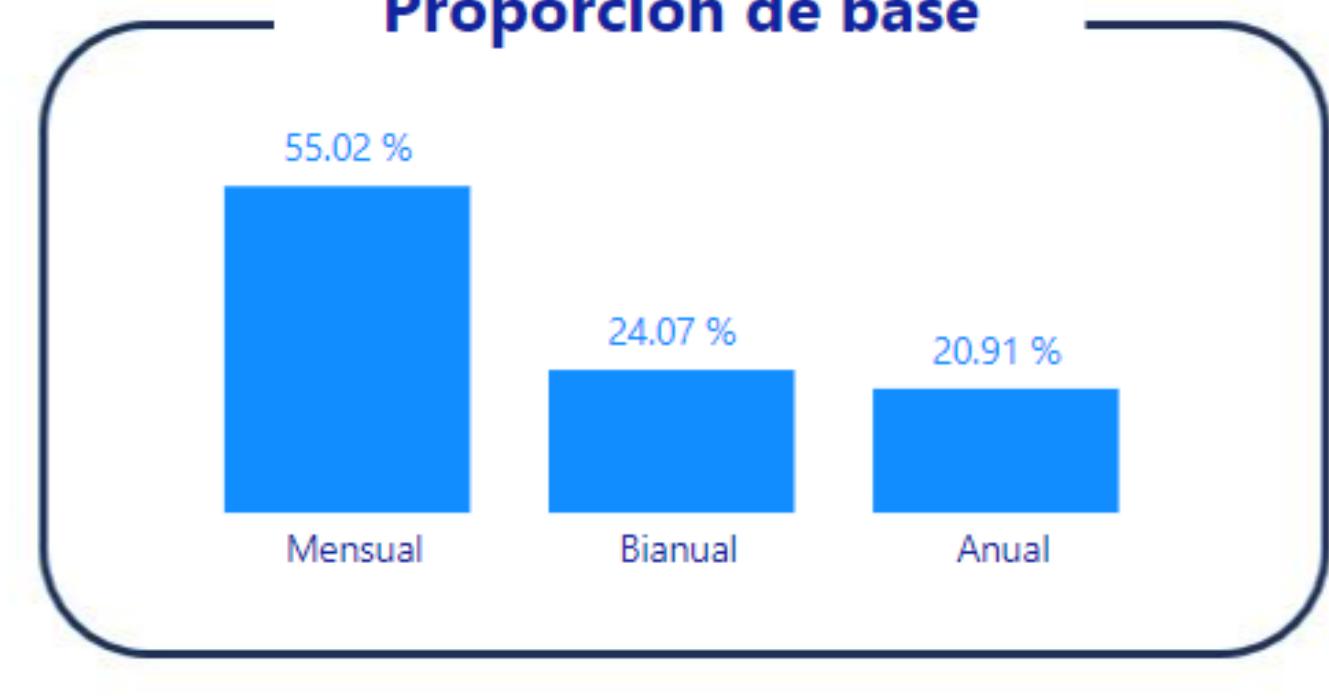
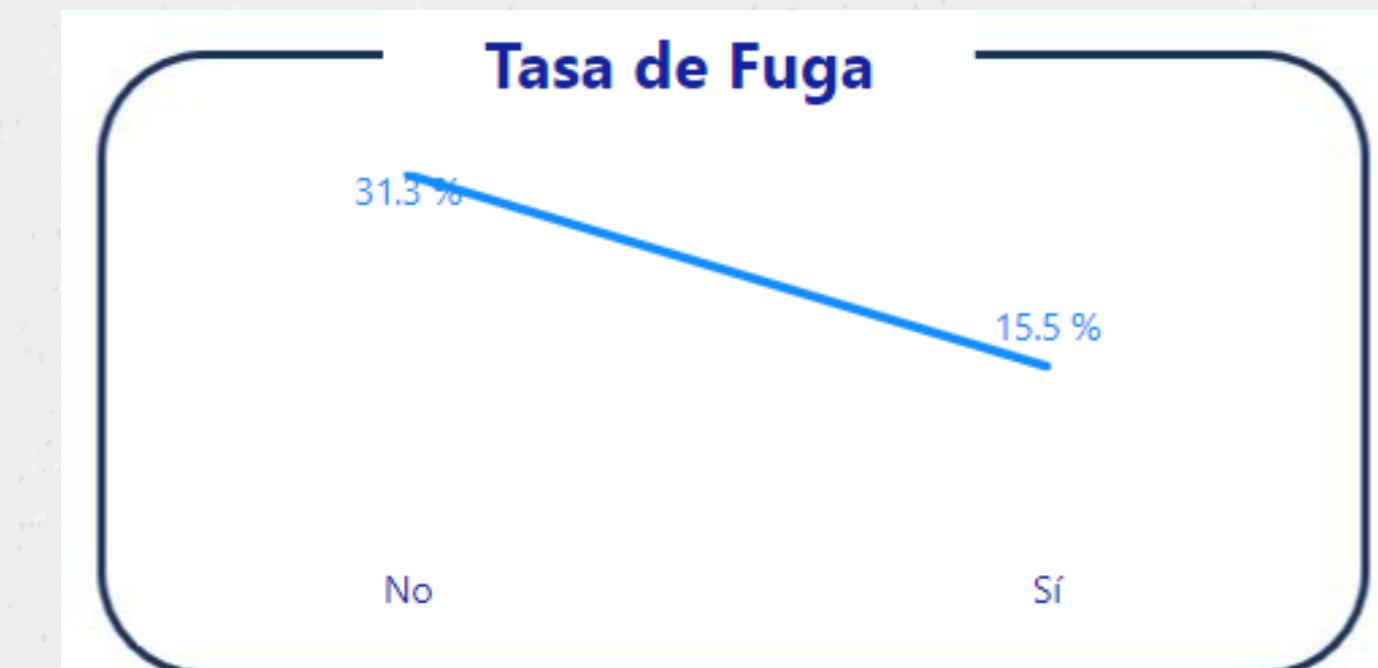
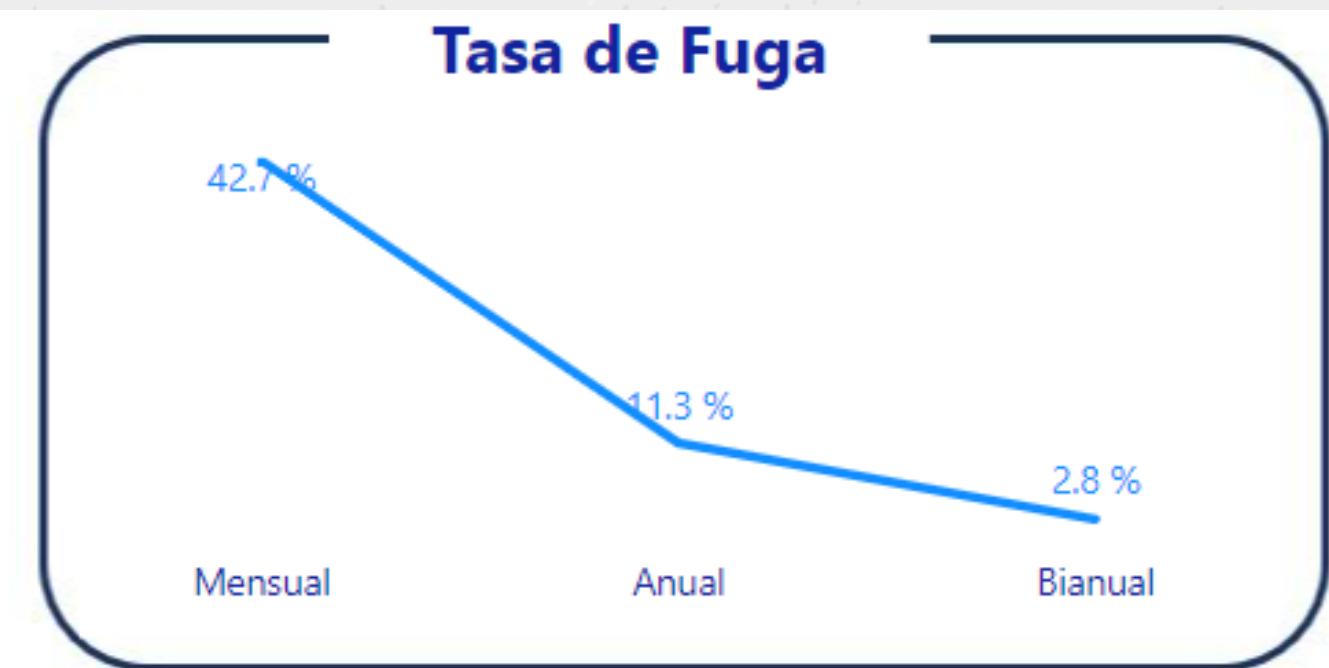


## Antiguedad

Contratos nuevos tienen tasa muy alta de fuga. Se sugiere revisar calidad de las ventas y reclamos por temas de facturación.

## Facturación Electrónica

Si bien hay un 60% de clientes con facturación electrónica, hay un 34% que está cayendo en fuga. Hay que revisar si las personas están recibiendo y revisando su factura, pues pueden terminar desactivando por deuda.



## Contrato

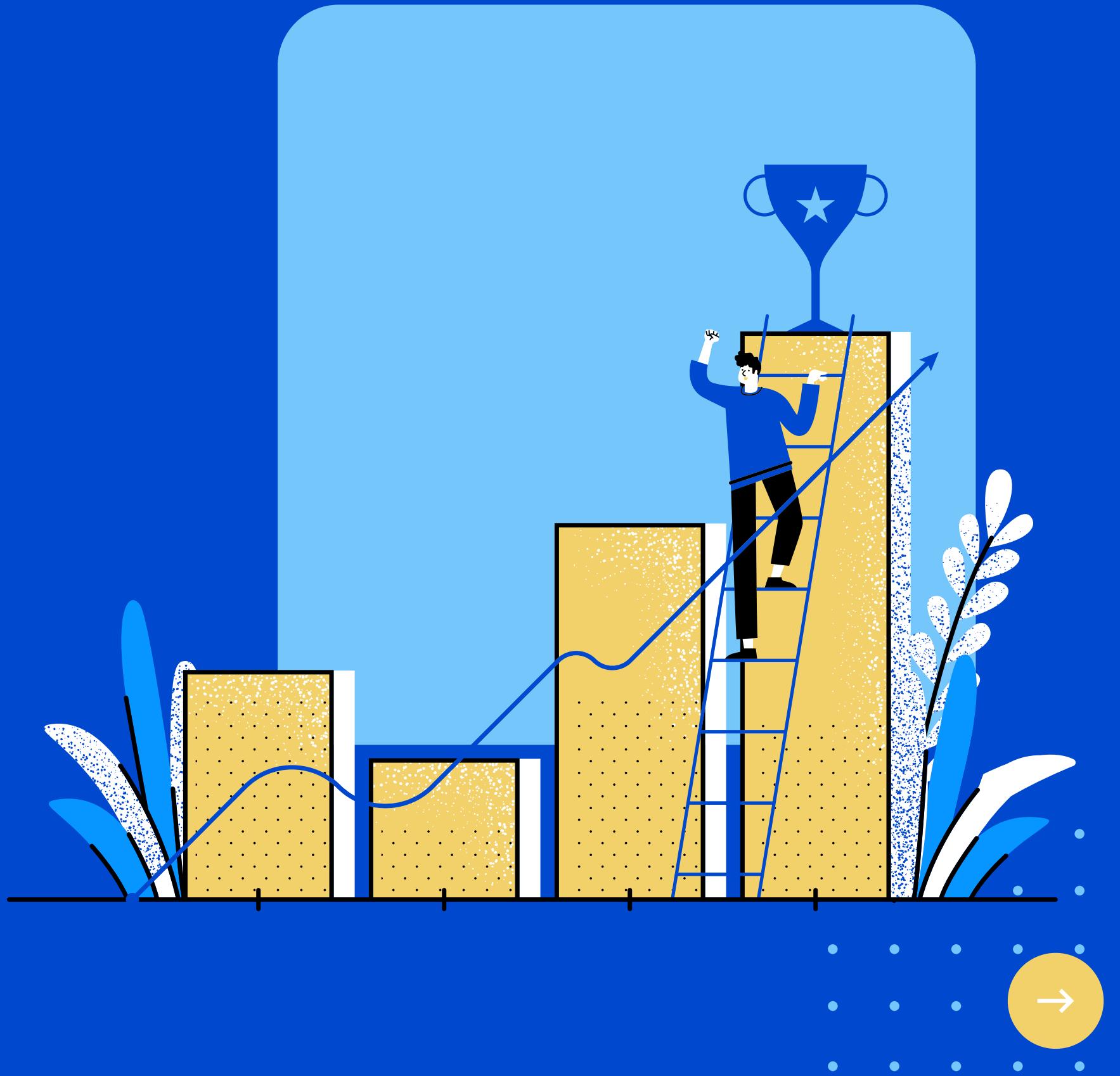
Los clientes con contrato mensual presentan un ratio de fuga significativamente mayor que aquellos que tienen contratos de 1 o 2 años. Podrían promoverse acciones que hagan que los clientes firmen contratos a más largo plazo (atados a equipos o algún servicio). A n

## Tenencia de Hijos

Se ha encontrado que clientes con hijos presentan menor tasa de fuga. Ello, combinado con la variable correspondiente al estado civil y al rango etario parecen señalar que los clientes que fugan suelen ser jóvenes y solteros. Por tanto, las acciones de retención deben estar más enfocadas a este público o tratar de traer más clientes mayores.

# Modelamiento

Modelo e interpretabilidad



# Modelo de churn

## Ficha técnica

Se desarrolló una primera versión de un modelo de predicción de churn para poder identificar a los clientes con mayor propensión de fuga y poder accionar campañas de retención sobre ellos.

### INFORMACIÓN GENERAL

#### Variable objetivo

Indicador si el cliente hizo churn (1) o no (0)

#### Población

Base de clientes del último periodo. Total de 7043 observaciones.  
Se separaron 1407 casos para validación fuera de muestra.

#### Técnica usada

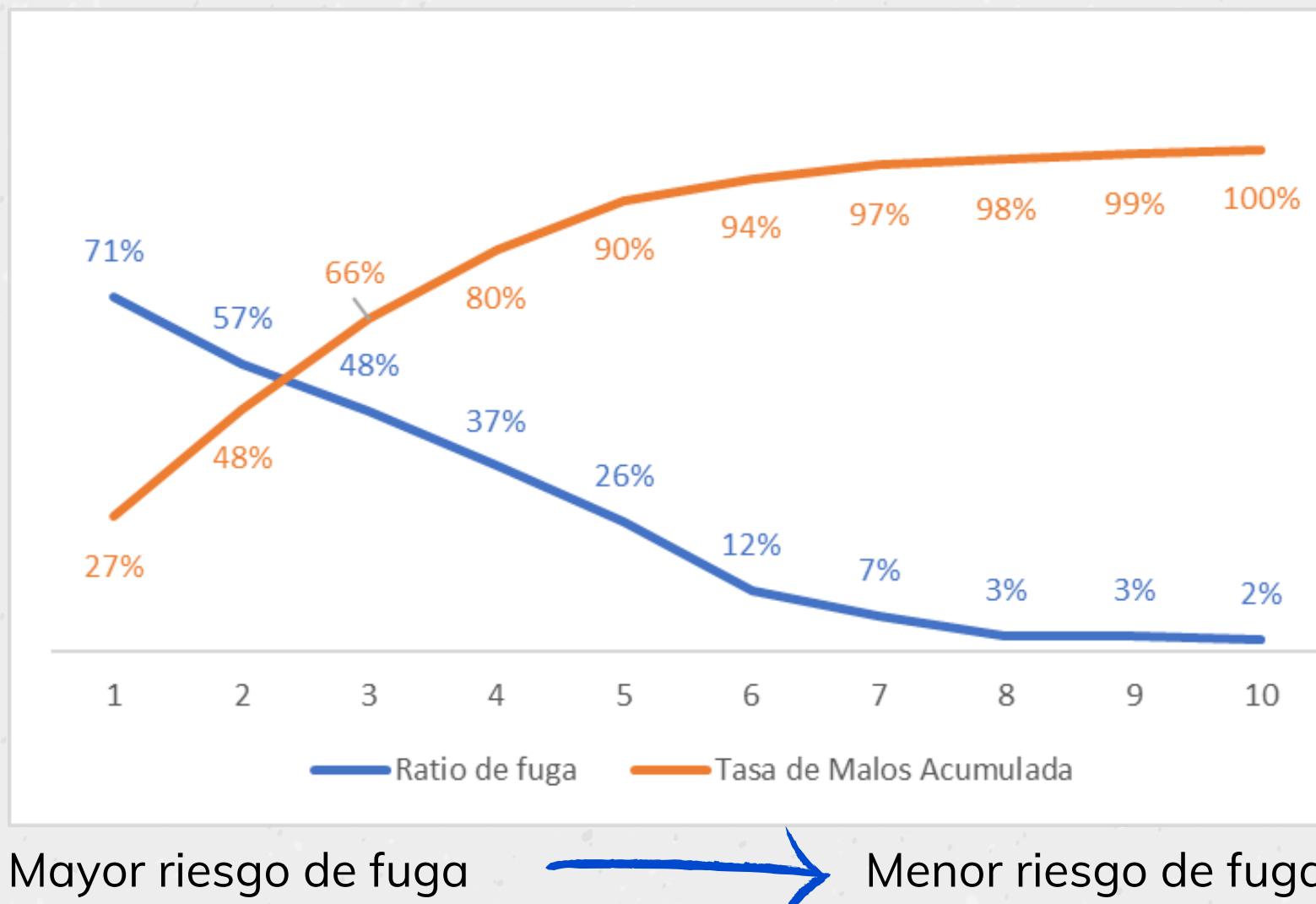
Bosque aleatorio (Random Forest)

#### Datos usados

Dominios de datos: comerciales, servicio, facturación, pago, demográficos.

# Resultados del modelo (set de prueba)

Modelo ordena la tasa de fuga desde el decil 1 (más riesgoso) al 10 (menos riesgoso)



83%

Área bajo curva ROC  
(Referencia: >70% muy bueno)

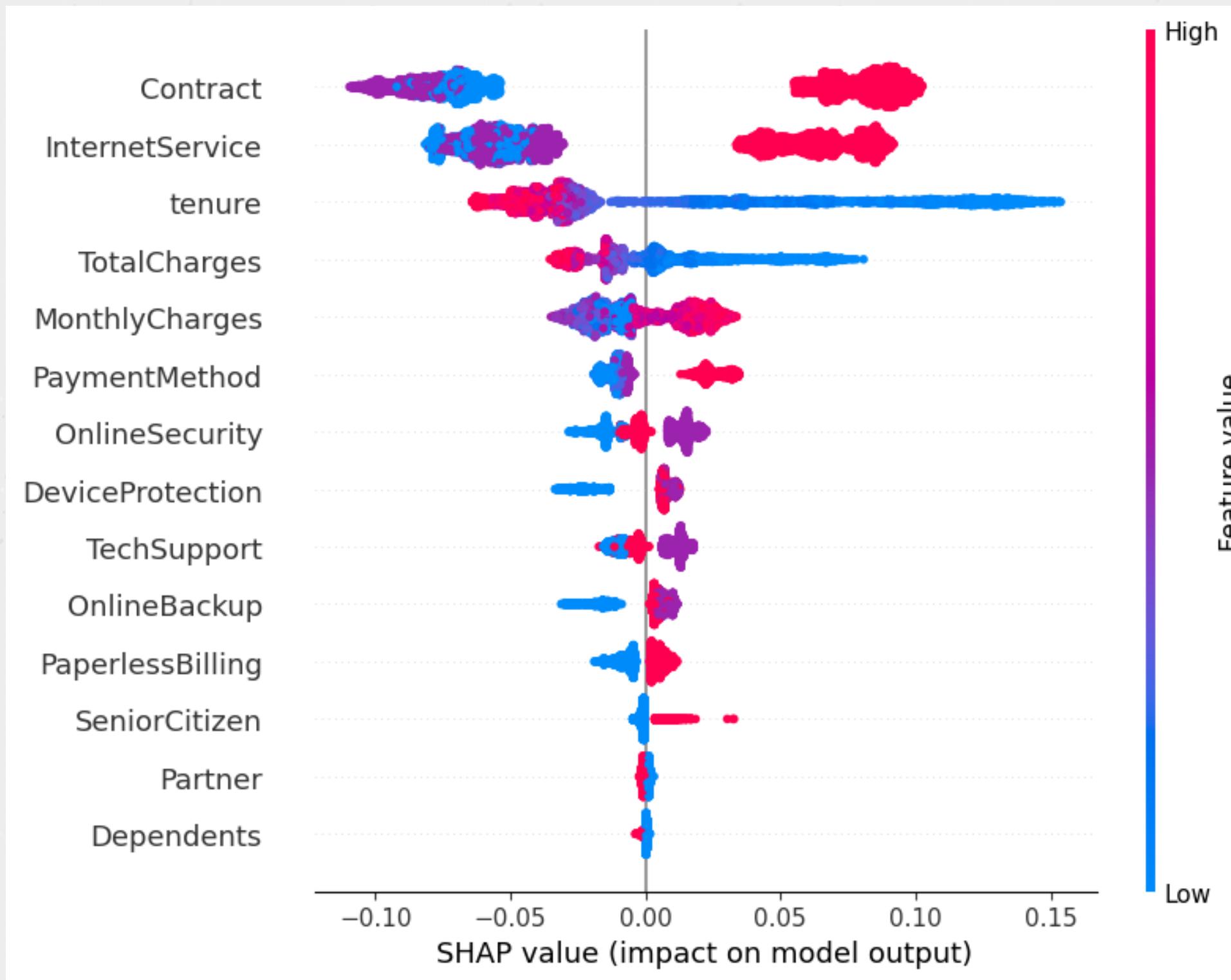
55%

Métrica KS

## Identificación de fuga

Con un 20% de la base de prueba se llegó a identificar el 48% de todos los clientes fugados en la base de prueba.

# Interpretando modelo



## 1. El orden importa

Las variables que están más arriba en la lista son las más relevantes. Conforme se va bajando, su importancia disminuye.

## 2. Lo rojo es alto, lo azul es bajo

En el caso de variables numéricas, cuanto mayor sea su magnitud, más rojo se pone. En el caso de las variables categóricas, se ha puesto los valores afirmativos (Yes) en rojo, No en morado, No aplica en azul.

## 3. A la derecha es peor

Conforme la línea se aleja del cero, hacia la derecha, la probabilidad que el cliente sea propenso a fuga es mayor.

## 4. Ejemplo 1: Tenure

Valores bajos (azules) de tenure están a la derecha del cero en su mayoría. Luego, clientes de poca antigüedad (bajo tenure) son más propensos a fugar.

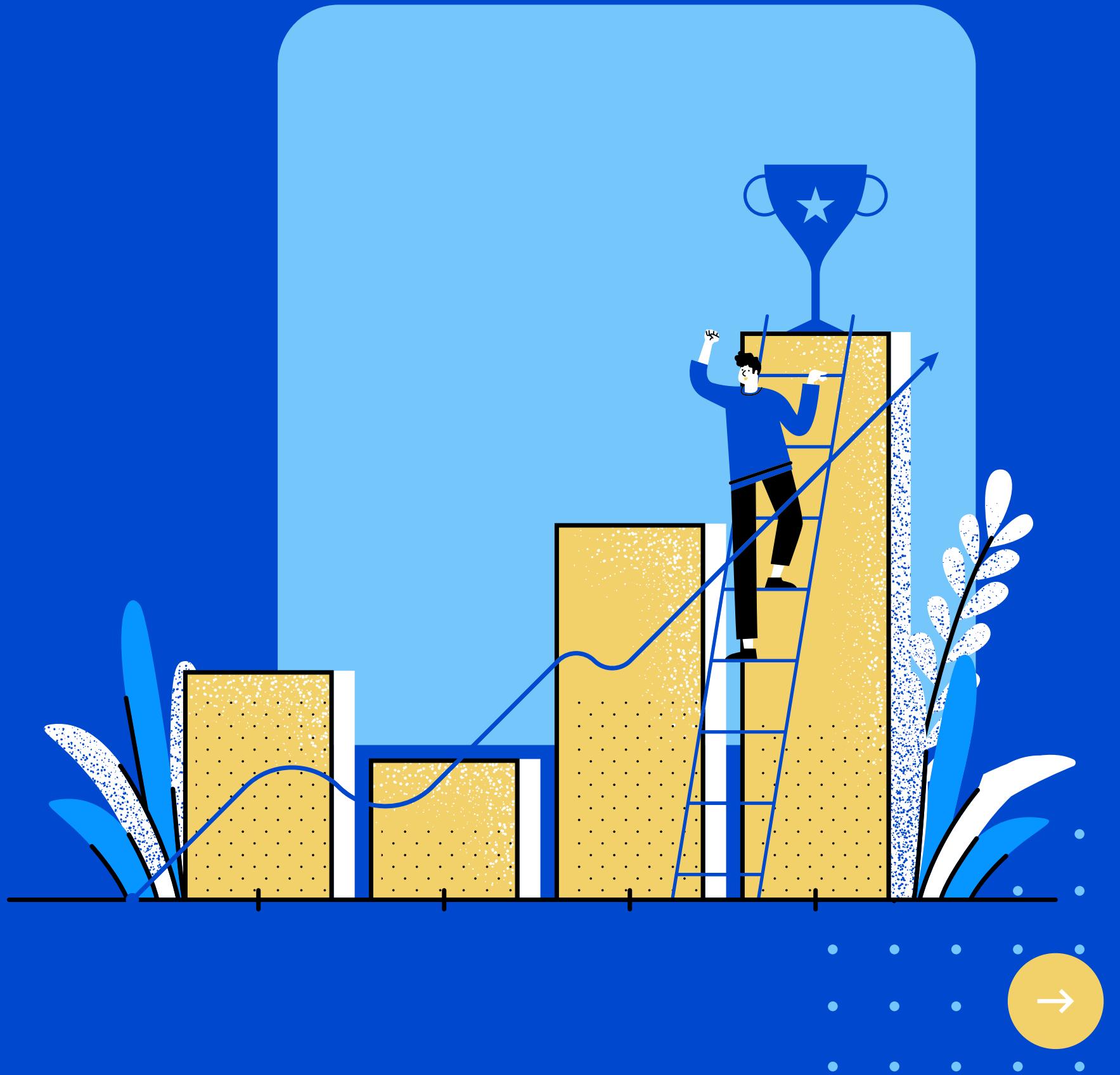
## 5. Ejemplo 2: Internet Service

Clientes con fibra óptica (rojo) son más propensos a fugar que clientes con DSL o sin internet (azul y morado).

## 6. Ejemplo 3: Contract

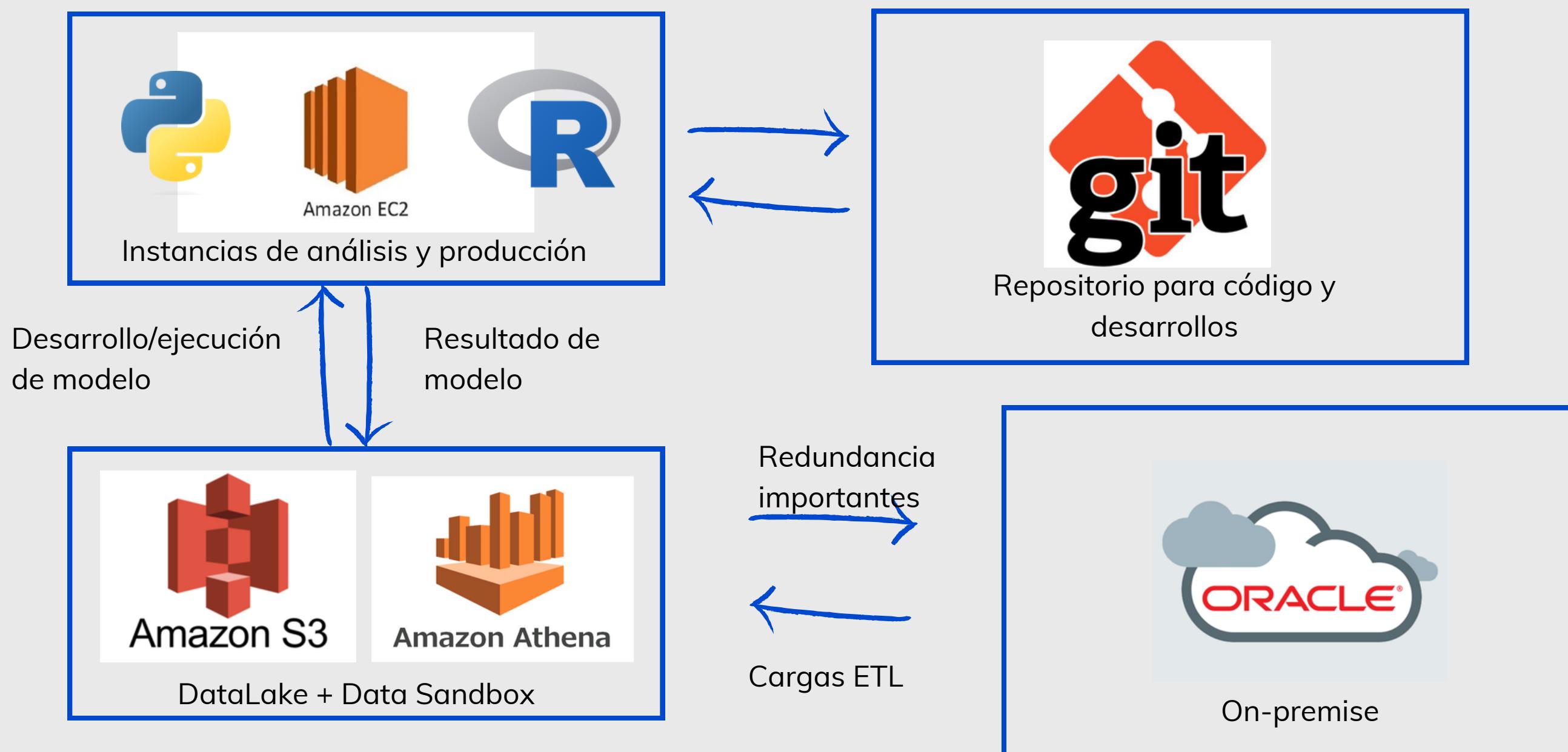
Clientes con contrato de mes (rojo) son más propensos a fugar que clientes con contrato a año o dos años (azul y morado).

# Plan implementación



# Arquitectura de implementación

Propuesta híbrida sobre AWS: disponibilidad + escalabilidad + continuidad.



## Puntos importantes

Arquitectura híbrida busca garantizar disponibilidad de los datos, escalabilidad de procesos y continuidad de operación. Se basa en trabajo en la nube, pero con réplica de lo crítico on-premise.

## Plataforma Web

Se propone el uso de Amazon Web Services (AWS); específicamente, los servicios de Redshift, S3, Athena y EC2. Los primeros, para almacenar los datos. El EC2, para máquinas virtuales para el desarrollo de nuevos modelos y ejecución de los existentes.

## On-Premise

Se propone uso de base de datos Oracle, por su soporte al cliente y garantía que ofrece la marca. El ideal on-premise es servir de espejo para garantizar disponibilidad de datos críticos ante posibles caídas de servidores cloud.

## Inversión

Amazon tiene tarifas on-demand. Evaluar presupuestos y establecer lineamientos para minimizar costos mensuales.

## Actividades

Configuración de las plataformas, integración de plataformas, instalación de Python en EC2, instalación de librerías requeridas, gestionar permisos de firewall con Seguridad.

## Recursos

Tiempo depende de recursos que se asignen al proyecto. Dado que no se puede esperar a que toda la arquitectura esté implementada, pues demorará un tiempo considerable, se plantea fasear la implementación.



## FASE 1 (DURACIÓN ESTIMADA: 2-3 MESES)

Data disponible en fuentes heterogéneas (archivos, bases de datos, etc). Se cargan manualmente las fuentes a una base de datos y se procesan ahí. Se ejecuta scripts de forma local en PC asignada.

## FASE 2 (DURACIÓN ESTIMADA: 3-6 MESES)

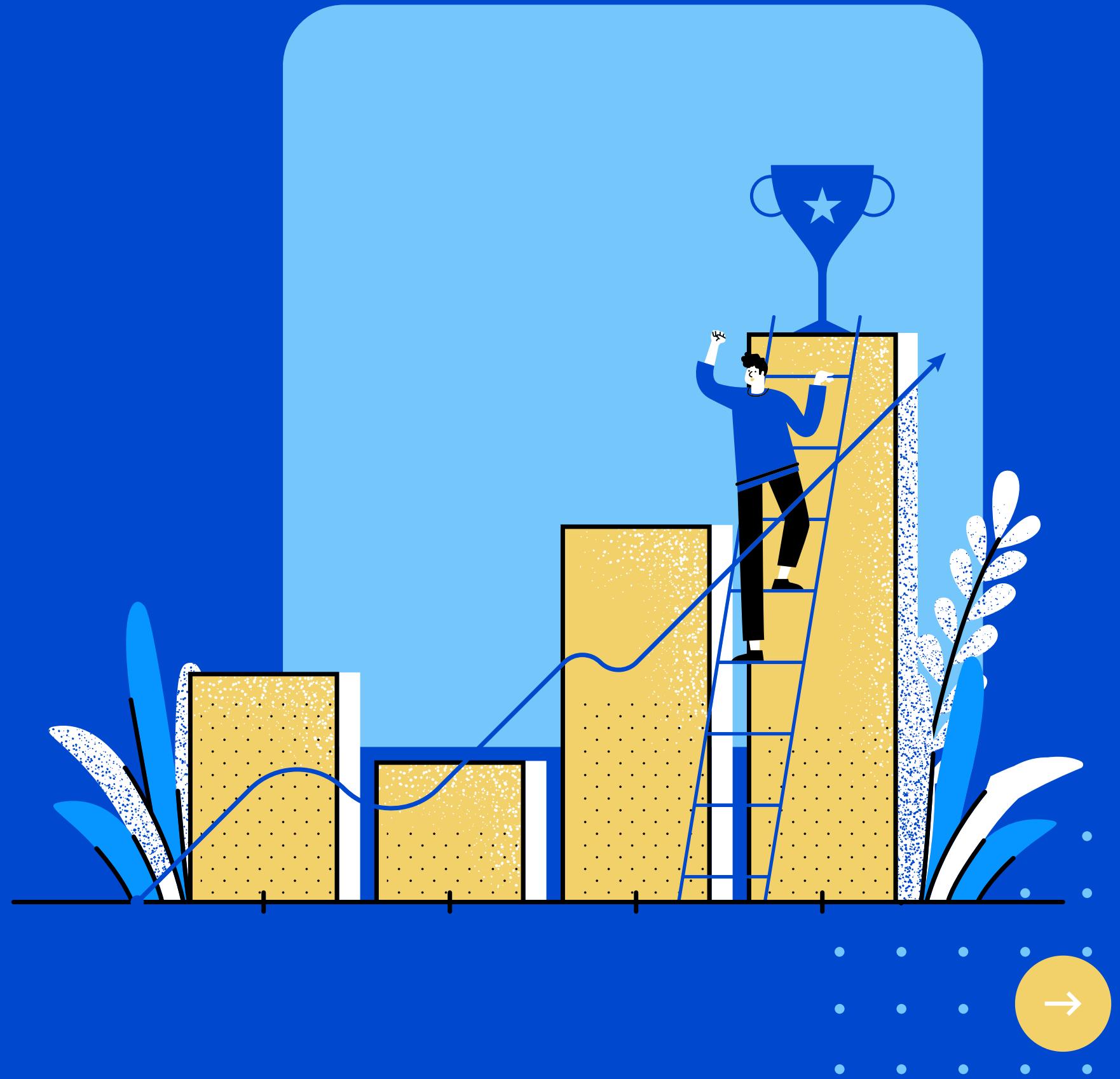
Data disponible en fuentes heterogéneas, pero cargadas automáticamente a la base de datos on-premise seguida por ejecución automática del modelo en servidor local.

## FASE 3

Arquitectura propuesta completamente implementada. Fuentes se cargan en AWS, modelo se ejecuta automáticamente en instancia de producción cloud.



# Next Steps



# Siguientes pasos

Importante no olvidar estas actividades

## Pilotos

Definir con área comercial pilotos de retención para gestionar a los clientes con mayor probabilidad de fuga.

## Calidad del dato

Revisar las distribuciones de las variables y de los deciles luego de cada ejecución del modelo

## Seguimiento

Definir métricas y periodicidad para medir las métricas básicas de los modelos. Definir casos de alerta y de necesidad de calibración.

## Valorización

Definir metodología para valorizar monetariamente el aporte del modelo (con la campaña). Medición de uplift de grupo de gestión vs. grupos de control.

# Muchas gracias

