

Proyecto Churn

Entidad Bancaria

¿Quiénes somos?

Somos un equipo multidisciplinario que tiene como misión potenciar y transformar negocios, a partir de la búsqueda de soluciones basadas en datos.

Este proyecto será abordado por:

- Caligaris Gabriela - Data scientist y analista especializado de producto
- Churgovich Nahuel – Data scientist y developer
- Jerez Francisco – Data scientist, especialista en Supply chain



Estructura de **nuestro trabajo**

Introducción

Actualmente la reconocida entidad bancaria XX se encuentra transitando una fuerte transformación que impacta en el volumen de clientes.

Esto obliga a contar con indicadores que permitan medir, controlar y tomar acciones preventivas frente a posibles abandonos de parte de los consumidores, considerando que resulta mucho más costoso captar nuevos clientes que retenerlos

La realización de este proyecto le permitirá a la compañía obtener líneas de trabajo para desarrollar planes de acción concretos a fin de lograr retener a sus clientes, a partir de la predicción de estos sucesos



Objetivos del proyecto

Identificar cuáles son los indicadores de alerta para aplicar una anticipada maniobra de retención sobre los clientes que potencialmente se podrían perder.

Nos preguntaremos y responderemos a los siguientes interrogantes:

- ¿Interviene una sola variable en la deserción o múltiples?
- De ser múltiples variables, ¿Cuáles son y cómo afectan sus posibles combinaciones? ¿Qué relación hay entre ellas?
- ¿Podemos identificar cuál es el grupo que más bajas nos representan?



Metodología de Obtención de datos

Para abordar la problemática consideraremos características generales de los clientes que nos ayuden a determinar patrones comunes (género, edad, nacionalidad, score crediticio, antigüedad en la empresa, cantidad de productos obtenidos, salario, saldo, si posee tarjeta de crédito y si aún siguen siendo clientes)

Al tratarse de una empresa joven con una política de datos recientemente implementada, decidimos tomar un set de datos de una fuente externa confiable, con experiencia en datos y del mismo rubro que nuestro cliente.



Esquema de trabajo

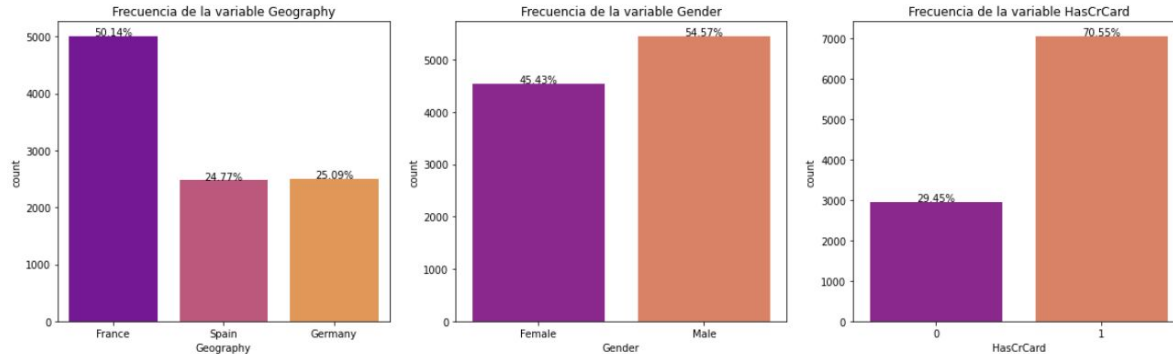
PRIMERA ETAPA



Principales **Hallazgos**

Cientes en General

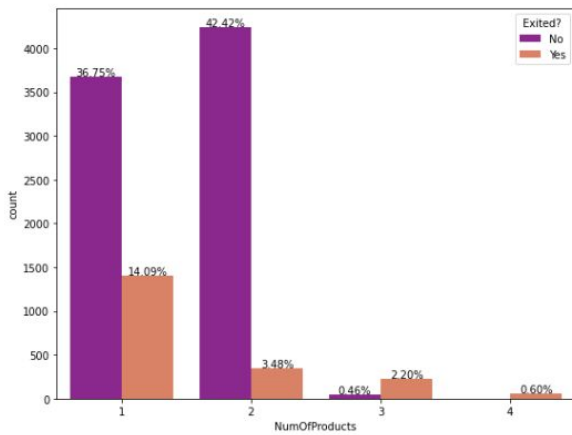
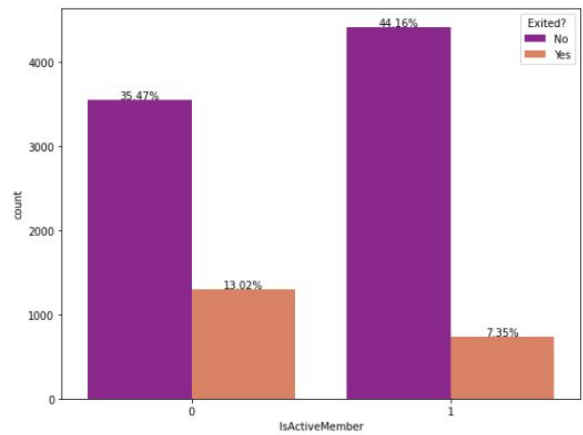
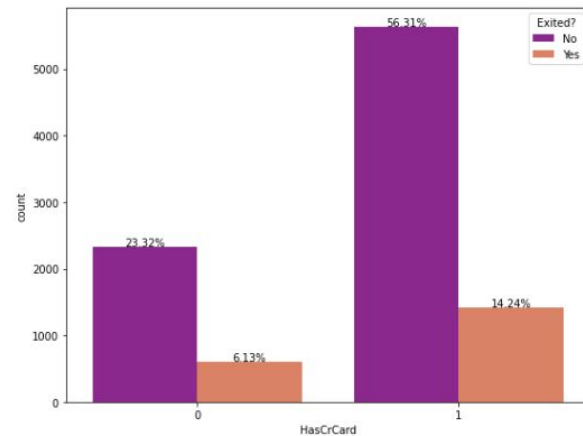
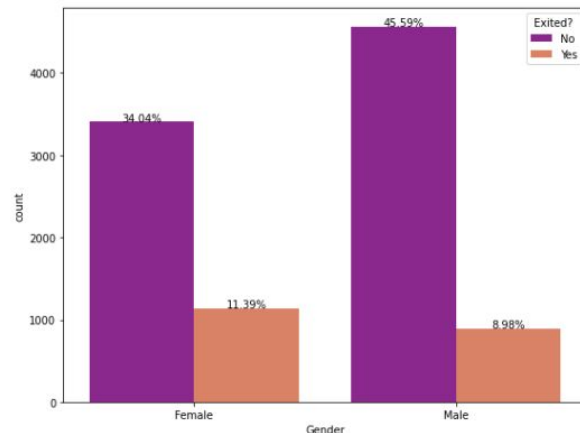
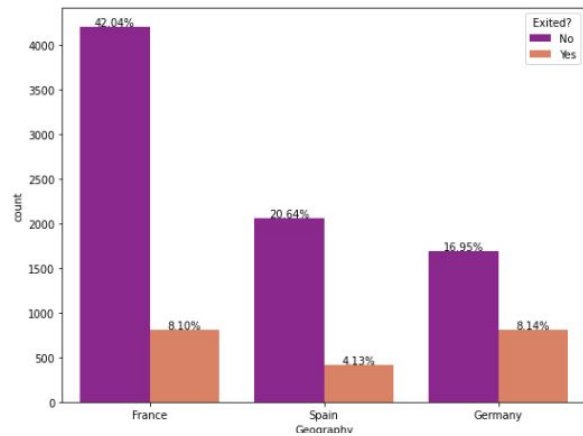
- La distribución de clientes por **género** es **pareja** (55% Masculino / 45% Femenino)
- La **mitad** de los **clientes** se concentran en **Francia**
- Más del **70%** de los clientes poseen **tarjeta de crédito**
- **No hay diferencias** significativas entre la cantidad de **clientes activos y no activos**
- **No hay patrón marcado de permanencia**, los datos se distribuyen en forma similar en los años,
- El 90% de los clientes ha adquirido 2 productos o menos



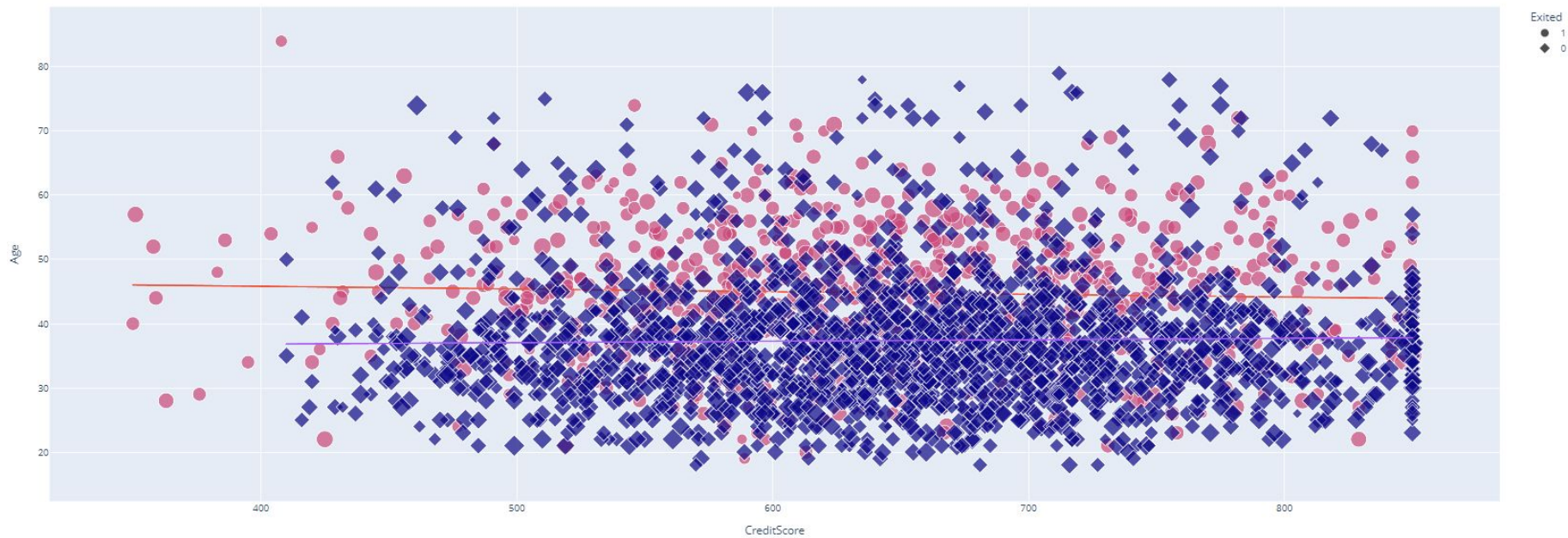
Gráficos resultantes del análisis univariado trabajado en el notebook

Cientes que abandonaron

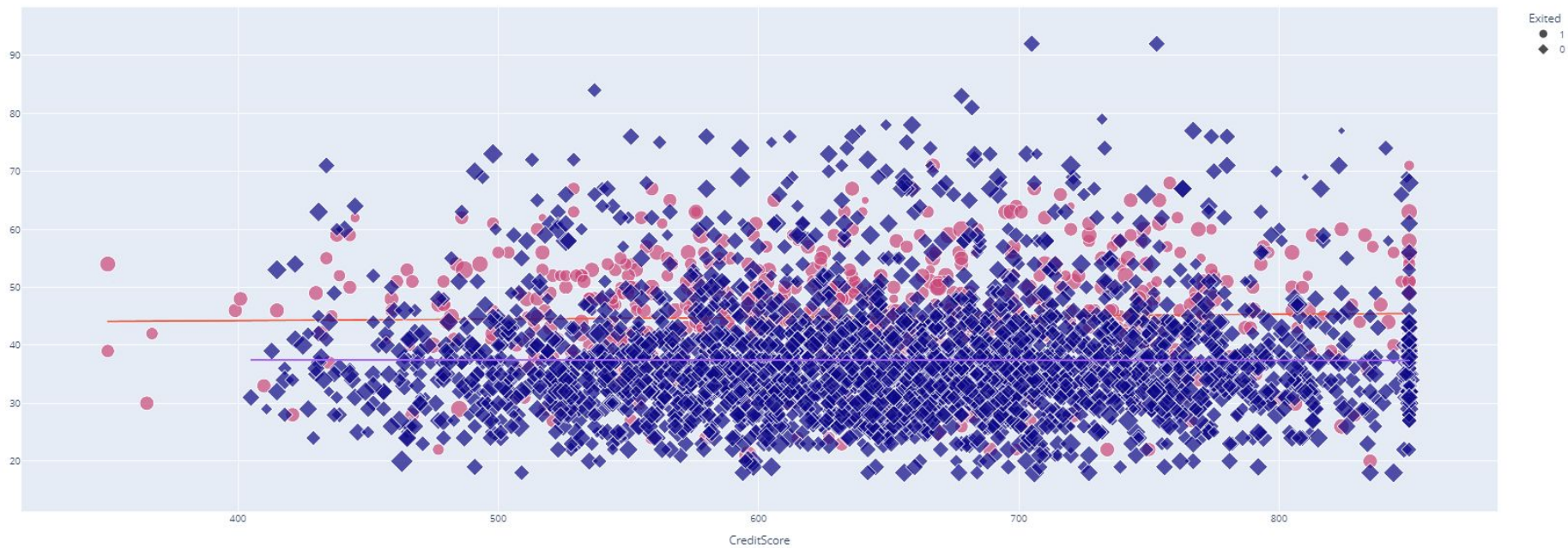
- El 20% de los clientes ha abandonado la compañía
- Casi el 50% de los clientes de Alemania se han ido vs el 20% de los demás países
- Hay mayor tasa de abandono de mujeres en la entidad que respecto a hombres
- El 25% de los que tienen tarjeta de crédito han abandonado la compañía
- El 40% de los clientes que consumieron un producto se han ido, aunque prácticamente el 100% de los que consumieron 3 y 4 también.
- luego de los 50 años se dan muchas salidas de la empresa, sin importar el género



Relacion entre Score y edad en Mujeres



Relacion entre Score y edad en Hombres



Esquema de trabajo

SEGUNDA ETAPA



Modelo de Predicción

Se entrenaron distintos modelos de clasificación y se eligió el **Random Forest** para predecir, ya que es el que obtuvo la mejor performance en base a las métricas analizadas,

Evaluación de Métricas

	precision	recall	f1-score	support
0	0.87	0.95	0.91	1593
1	0.69	0.44	0.54	407
accuracy			0.85	2000
macro avg	0.78	0.70	0.72	2000
weighted avg	0.83	0.85	0.83	2000

Recall: de cada 100 clientes que abandonaron la compañía, el modelo acertó 44

Precision: de 100 veces que el modelo predijo que los clientes abandonaban, se fueron 69

* No se considera el accuracy como métrica por tratarse de un set desbalanceado