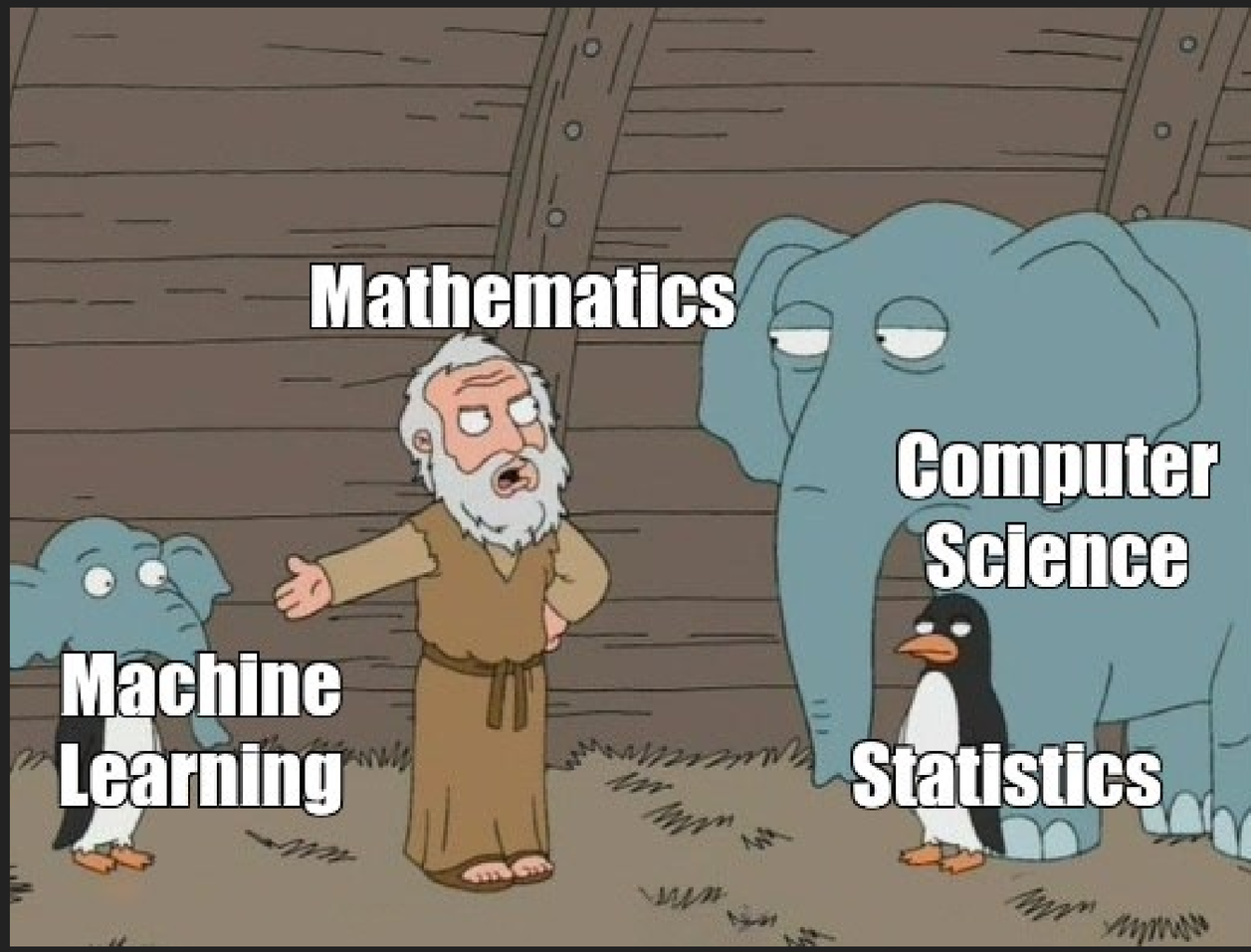
GRUPO 1

01

MODELO DE ATTRITION



Machine learning

02

Contenidos

de la presentación

Problema de negocio
- Retención de clientes
- Industria bancaria

Análisis exploratorio de datos
Preparación de datos
Selección de variables
Modelado
Pruebas del modelo
Predicciones
Estrategia de retención

Cómo encaramos el desafío

04

Comprender la situación actual

Cuál es la situación actual de deserción y cuales son las variables que explican este fenomeno

Estudio de los datos

Cuales son los datos disponibles, qué datos extras podemos generar?
Limpieza, preparación y control de los datos.

Desarrollo del modelo

Prueba de diferentes modelos de ML para detectar aquellos clientes que se van a dar de baja.
Cuál es el mejor camino para la mejor predicción posible?

Estrategia de retención

Qué acciones tomamos con los resultados obtenidos?

03

Detectar clientes próximos a dejar la compañía

El banco busca detectar aquellos clientes próximos a darse de baja (en los siguientes dos meses) con el objetivo de implementar acciones de retención sobre ellos.

Con esto se busca maximizar la ganacia de la compañía

05

Por qué retener clientes?

Es 5 veces más costoso conseguir un cliente nuevo que conservar a un cliente actual.

Se incrementa al menos 25% la utilidad como resultado de mejorar tan solo 5% en la retención de clientes.

Clientes fidelizados son 50% más propensos a probar nuevos productos y gastar un 30% más.

Clientes a largo plazo resultan menos sensibles a actividades publicitarias de la competencia

Clientes fidelizados refieren nuevos clientes por boca a boca y brindan feedback valioso

HTTPS://WWW.CRAZYEGG.COM/BLOG/CUSTOMER-RETENTION/
HTTPS://WWW.DCRSTRATEGIES.COM/CUSTOMER-INCENTIVES/5-REASONS-CUSTOMER-RETENTION-BUSINESS/

Industria Bancaria

06

Expectativas

Qué servicios buscan los clientes.

Fidelización

Por que servicios estarían dispuestos a quedarse

Deserción

Por cuales estarían dispuestos a irse.

Expectativas de Clientes

07

Descuentos y servicios

La principal razón por la cual son leales a su banco son los descuentos. También valoran un servicio personalizado, conveniencia y relevancia .

Experiencias personalizadas

Los programas de lealtad no funcionan.
Para volver estos programas más atractivos es necesario crear experiencias personalizadas a través del análisis de data.
Ofrecer beneficios alcanzables que el cliente valore.

Resultados

Los consumidores, especialmente los millennials, buscan resultados; no buscan una caja de ahorro sino tener más dinero.
Es más útil ofrecer un servicio que los notifique cuando sus gastos se salgan de presupuesto.

Nuevas tecnologías

El 46% de los encuestados por Accenture dijeron que usarían “Robo-advice”, técnicas automatizadas para hacer recomendaciones financieras basadas en cuestionarios y algoritmos.

Fidelización de Clientes

Nuevos bancos

Según Qualtrics, el 40% de usuarios de bancos se pasaría a un banco completamente online.

Costos acorde a los clientes

El 56% de quienes planean irse podrían cambiar de idea, principalmente si el banco baja los costos o mejora el servicio.

Centrado en el cliente

El 70% de los desertores dicen que su decisión fue impulsada por una acumulación de veces en las cuales no se cumplieron del todo sus expectativas.

Ubicación y atención

El 70% de las operaciones bancarias son digitales pero el 75% de los clientes que dejan un banco lo hacen por problemas con la atención personal en los bancos.

Deserción de Clientes

09

Deserción de nuevos clientes

El estándar de deserción es del 11% y sube a 25% entre nuevos clientes, de los cuales la mitad no pasa los 3 meses.
La rentabilidad depende del ciclo de vida del cliente

Deserción parcial

Está el caso de quienes aparentemente siguen activos pero abren cuentas en otros bancos (cerca del 50%). Esto merma las ganancias y es más difícil de detectar

Facilidad de cambio

Ya no es necesario trasladarse físicamente hasta el banco. Aparecen nuevas opciones de banca digital. La relación con el banco es transaccional (emocionalmente fácil de cortar).

Tasas y mal servicio

El motivo principal que los lleva a cambiar de banco son tasas elevadas, en segundo lugar un mal servicio

Etapas del trabajo

Análisis exploratorio

Transformación y generación de variables

Selección de variables

Modelado

Pruebas del modelo

Predicción y estrategia de retención

Análisis exploratorio

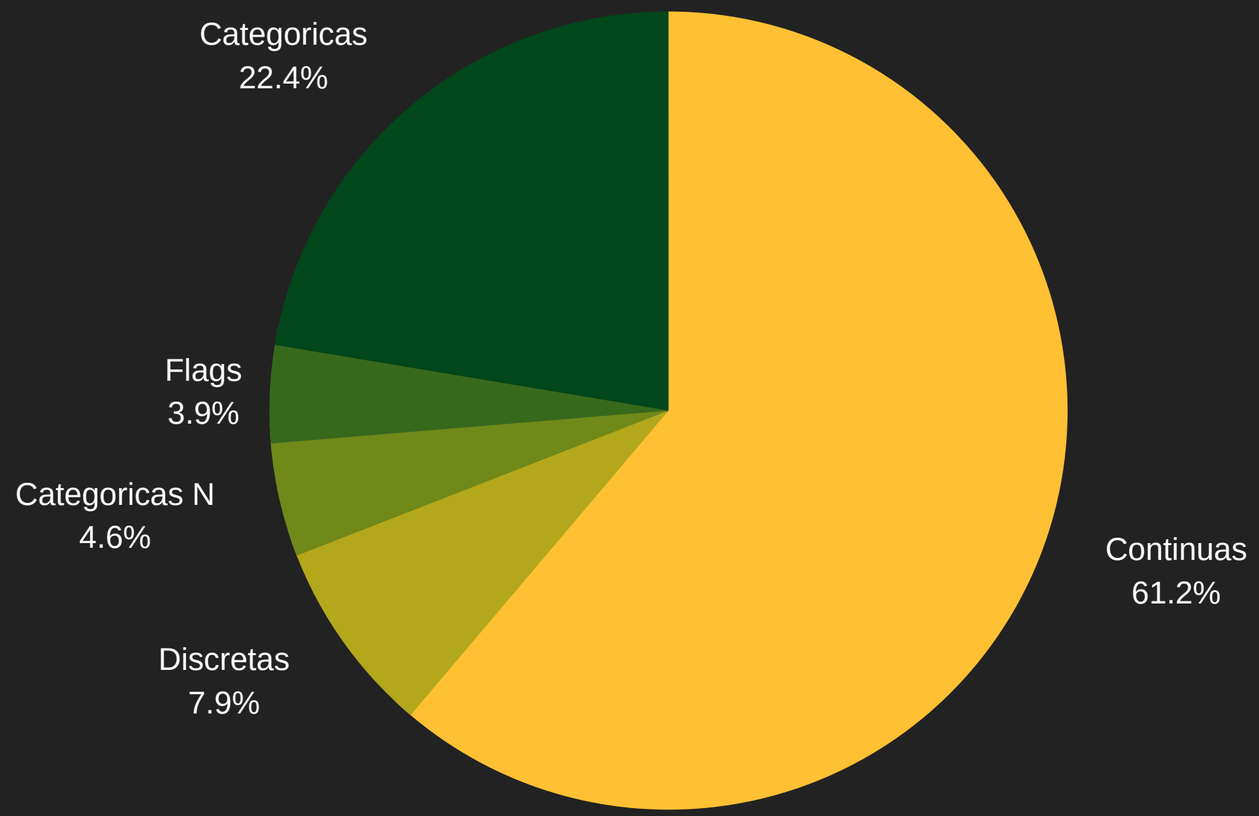
2.694.630

REGISTROS

La base cuenta con 2.694.630 registros de clientes y su información bancaria a lo largo de 2013 y primeros meses del 2014.

Trabajamos con diferentes sets de datos para asegurarnos no sesgar los resultados por los meses tomados.

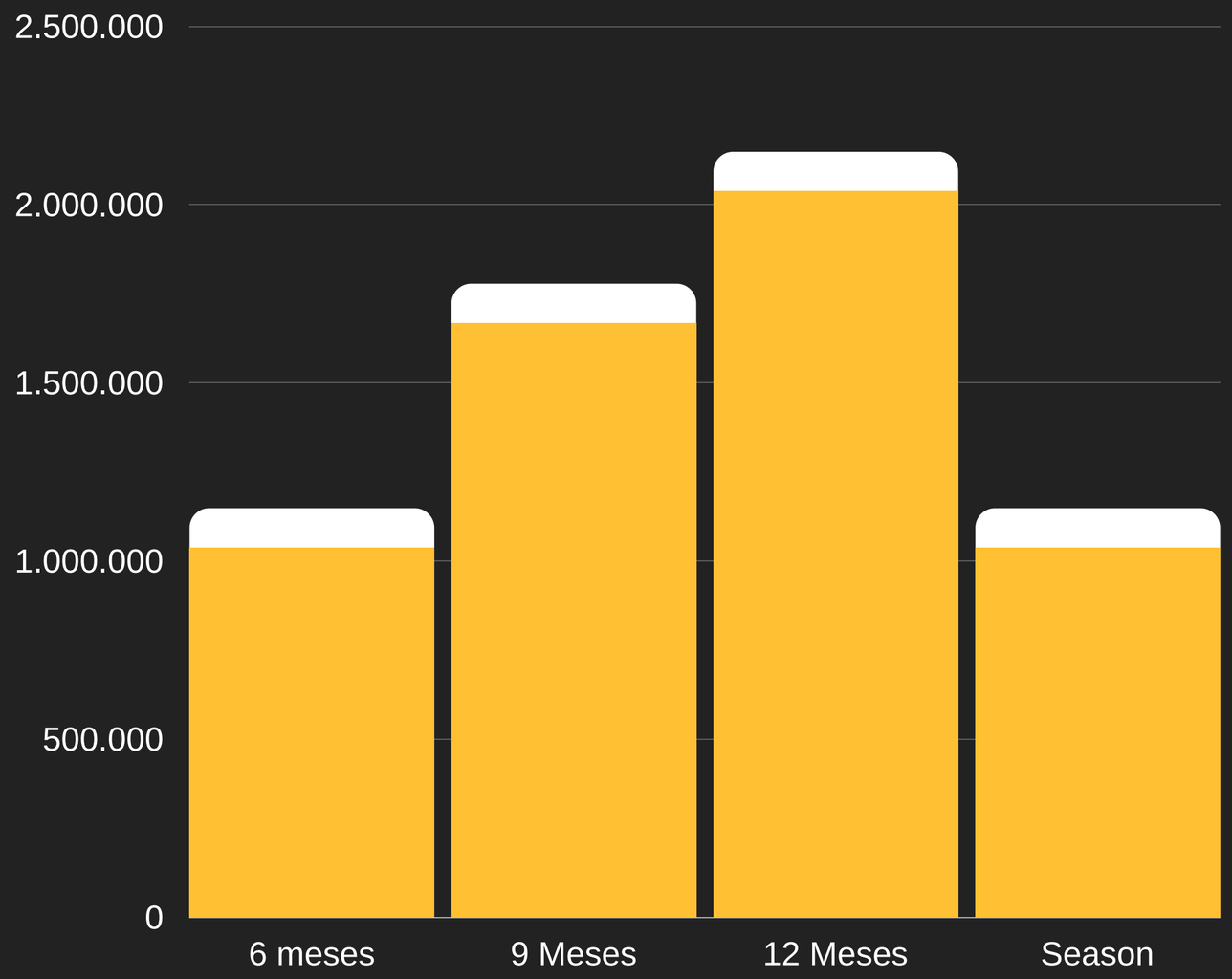
- Últimos 6 meses
- Últimos 9 meses
- Últimos 12 meses
- 4 meses 2013 + 4 meses 2014



Distribución de variables

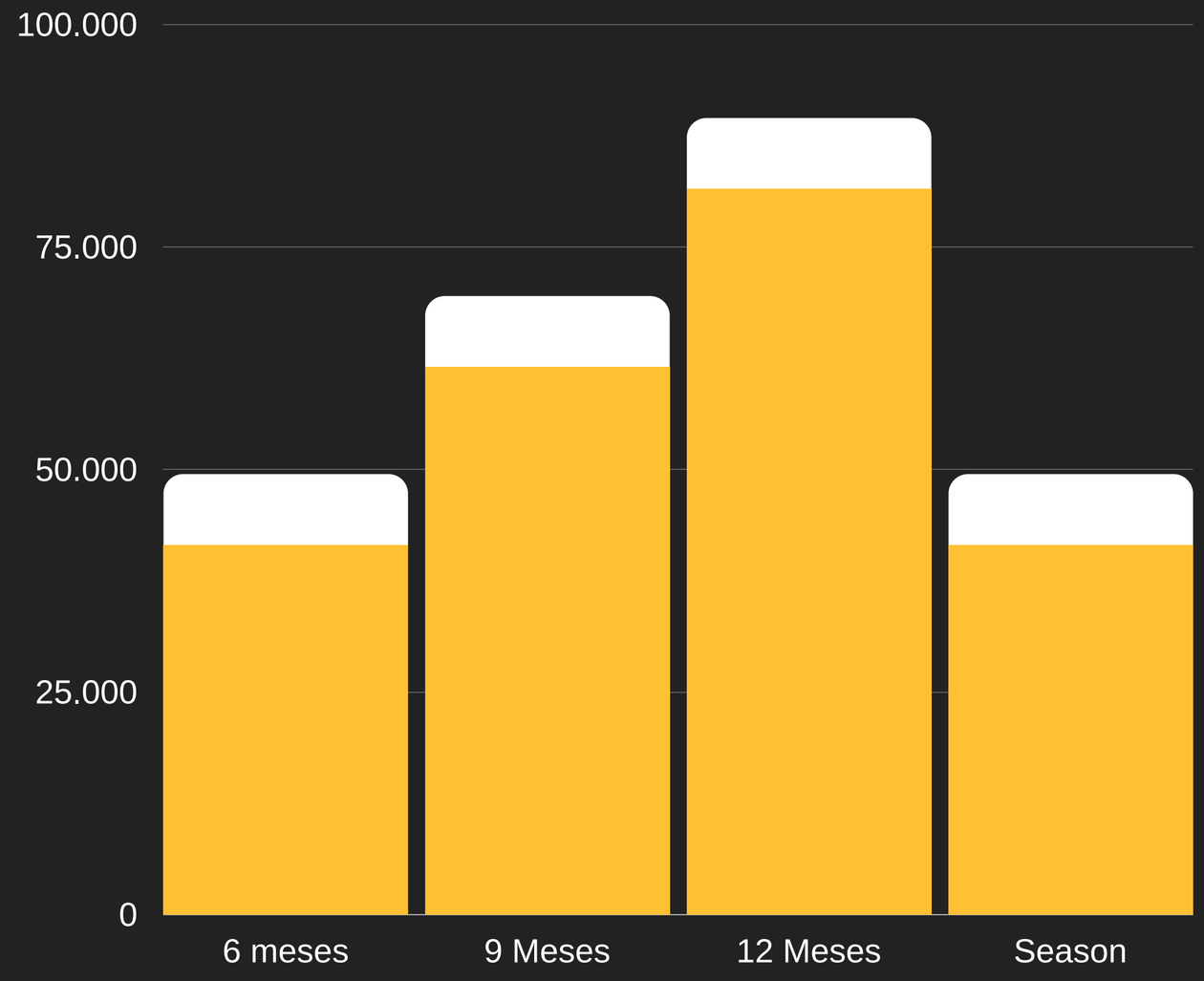
Distribución de las bases

En todas las particiones de base los clientes que se dan de baja en los siguientes dos meses representan menos del 1%.



Balanceo de las bases

Balanceamos las bases eliminando, de forma aleatoria, registros de clientes que seguían en el banco durante todo el período estudiado.



Análisis preliminar

192.852

CLIENTES UNICOS

95% de los clientes activos en 2013, continuaron activos durante todo el periodo estudiado.

10% llevan 19 años como clientes.

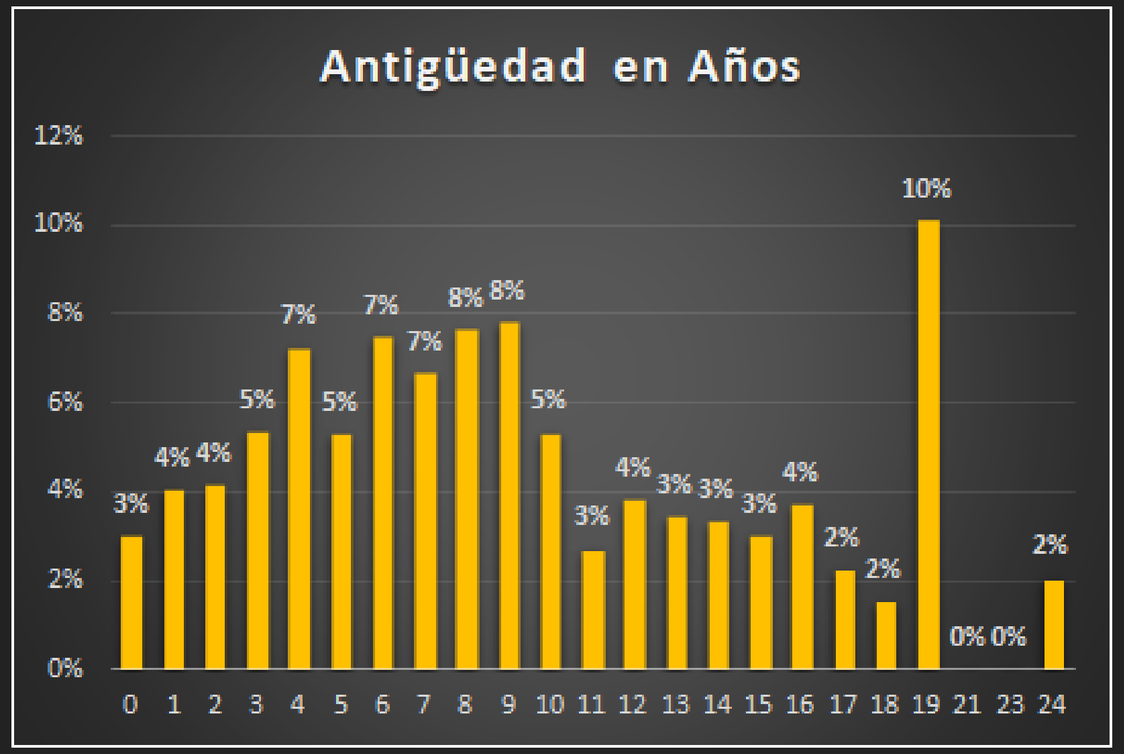
3% llevan menos de 1 año como clientes.

29% llevan 5 o menos años como clientes.

Llama la atención la baja que se evidencia entre 10 y 18 años de antigüedad.

Podría suceder que en esos años ingresaron menos clientes nuevos o que ocurrió algo que los hizo darse de baja masivamente.

Más interesante aún es la cantidad de clientes que llevan exactamente 19 años con este banco



Clientes nuevos

Crecimiento inusual

Se evidencia una cantidad inusitada de nuevos clientes en enero 2014 que fácilmente triplica la de otros meses.

Altas vs Bajas

Evaluando los clientes que se dan de alta y al mes siguiente dejan de estar activos vemos que casi todos los meses se mantiene en el 2%.
Esta proporción muestra una baja a 1% entre quienes ingresaron en enero 2014 y luego sube a 3,8% al mes siguiente.

Red flags

Esto podría indicar que febrero 2014 es un mes atípico para realizar predicciones o que se modifica la tendencia anterior, al menos entre nuevos clientes.

Clientes nuevos vs stock

Continuidad de clientes

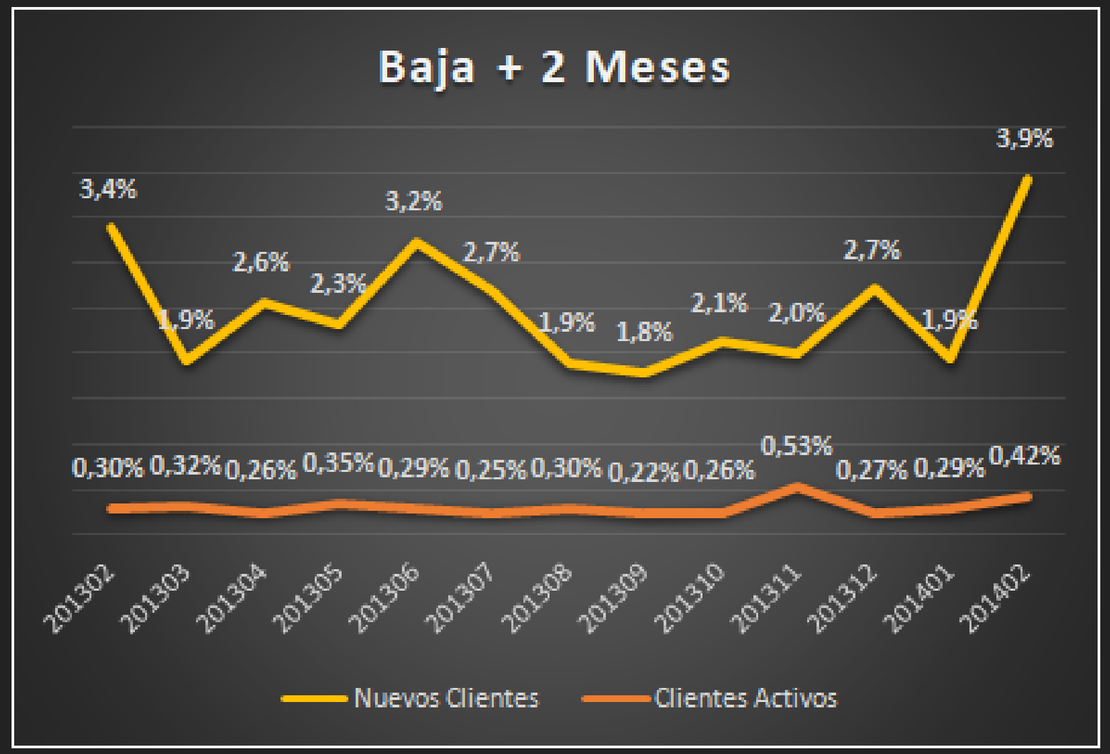
Cruzando los clientes en stock de cada mes con su continuidad dos meses después, en general el 99,7% seguían en el banco.

Propensión a baja

Podemos ver que los nuevos clientes son notoriamente más propensos a darse de baja en 2 meses que los clientes activos de meses anteriores.

Fluctuación mensual

También es llamativa la fluctuación mensual entre el 1,8% que se dieron de alta en septiembre 2013 y 3,9% en febrero 2014.
Mientras que entre los clientes en stock varía entre 0,2% y 0,5%.



Transformación y Generación de Variables

Duplicados

Quitamos registros duplicados (mismo usuarios en el mismo periodo)

Limpieza de campos

Quitamos columnas que no aportan valor

Valores Nulos

Reemplazamos todos los nulls por 0

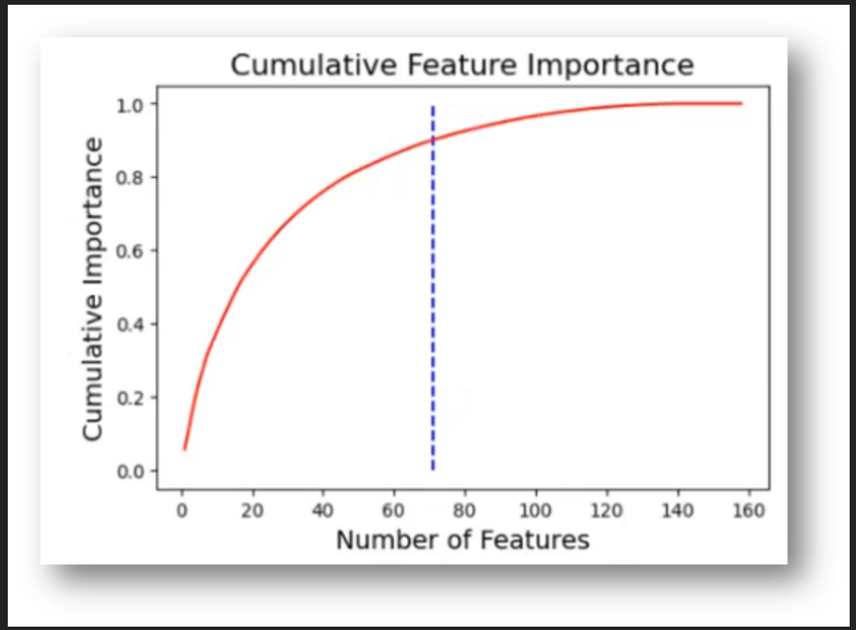
Nuevas variables

Combinamos subtotales de cuentas corrientes, cajas de ahorro y consumos con tarjeta

Transformación

Convertimos variables categóricas S/N en 1/0.

Reducción de dimensionalidad



Balanceo

Eliminamos registros clase "CONTINUA" para balancear la muestra de trabajo, dejando 1/25 clase "CONTINUA"

Importancia de variables

Quitamos las variables clase, número_de_cliente y foto_mes para la selección de variables de importancia. Quitamos 99 columnas de variables con bajo threshold (< 0.9)

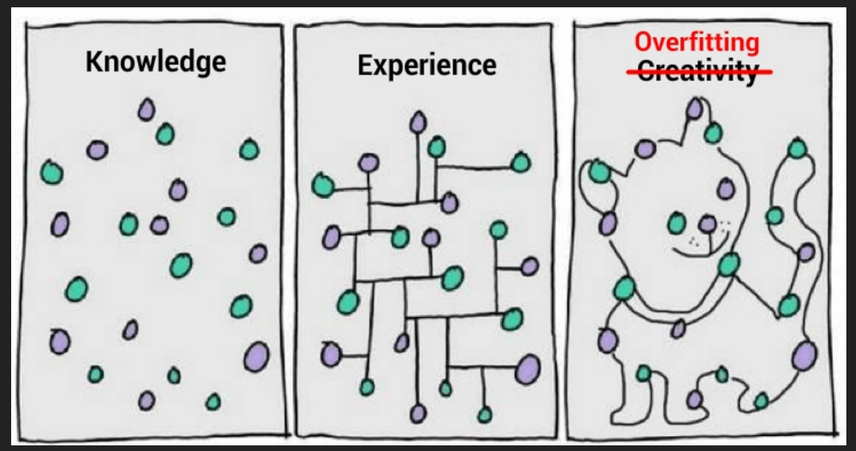
- -19 sin importancia,
- -76 con importancia baja

Selección de variables

72 (de 160) features poseen una importancia mayor al threshold establecido y fueron incluidas dentro del modelo.

Data sets

21 # Training y test



X sin Clase v1

Eliminamos la variable target clase y dropeamos las columnas eliminadas anteriormente quedandonos con 72 variables.

X sin Clase v2

Agregamos las variables calculadas y volvimos a correr los distintos modelos.

Re-escalado

Reescalamos las variables para poder comparar manzanas con manzanas y no manzanas con unicornios.

Modelado

Buscamos el mejor modelo posible alternando los parámetros con Grid Search y Random Search

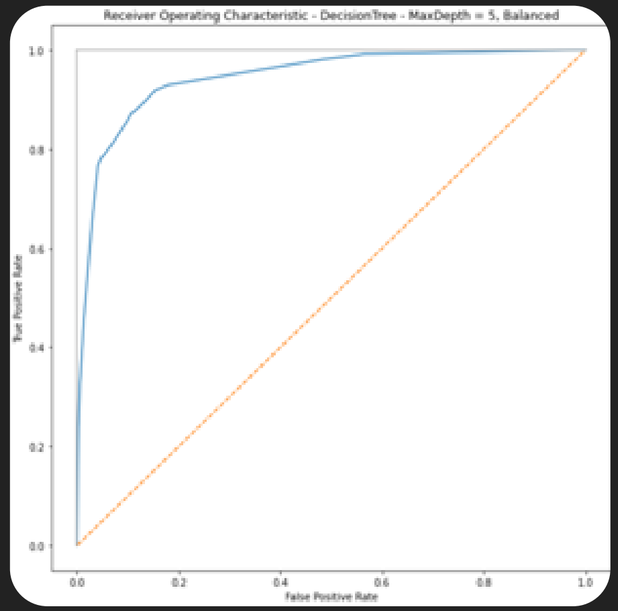
0.94 Decision Tree Classifier
(max_depth=5, class_weight='balanced')

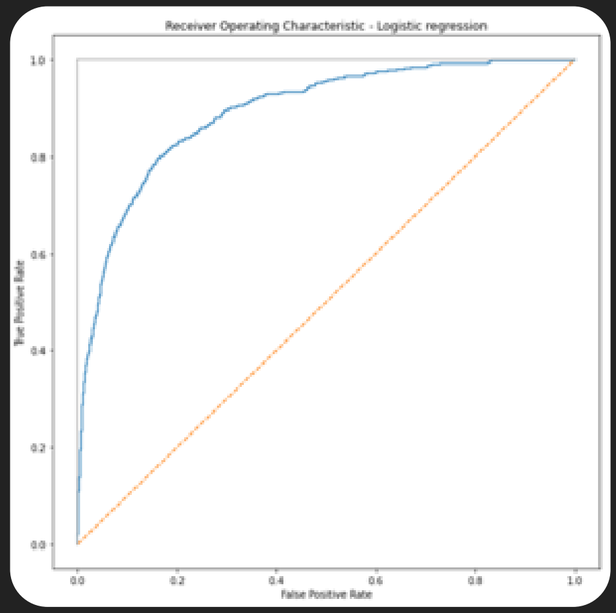
0.99 Gradient Boosting Classifier
(n_estimators=100, max_depth=5).

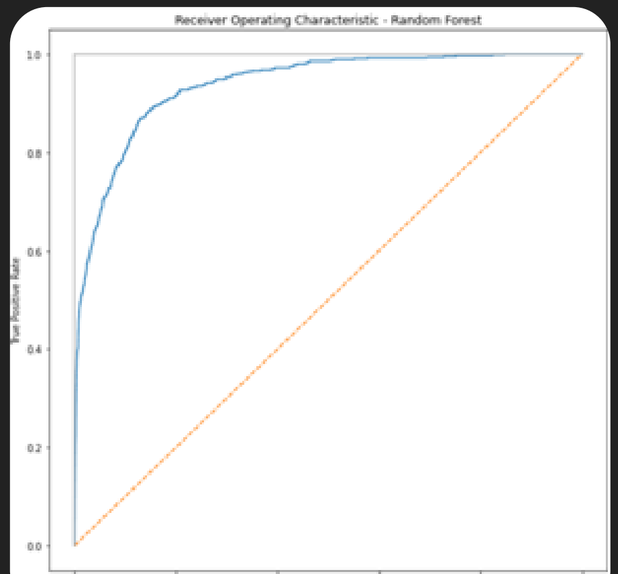
0.94 Random Forest Classifier
(n_estimators=50, max_depth=5, min_samples_split=10).

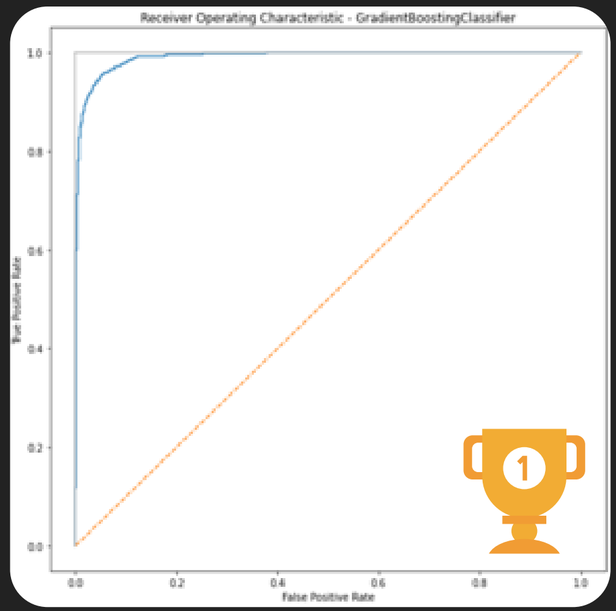
0.89 Logistic Regression

0.86 KNN Classifier









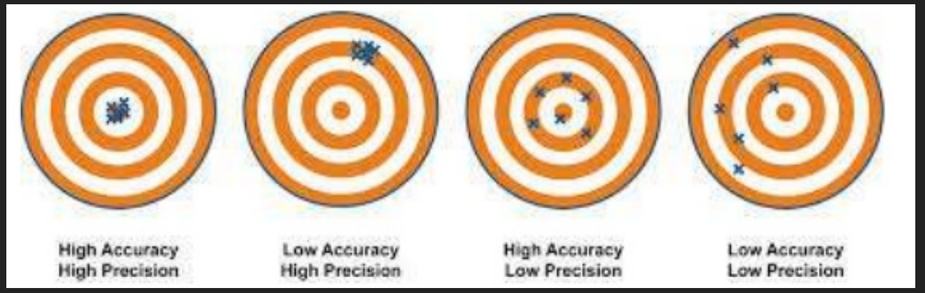
Accuracy, Precision & Recall

Gradient Boosting Classifier

es mejor en todas las evaluaciones, contra todos los modelos, excepto en Recall contra Decision Tree, es decir que se escapan algunos que no logramos detectar.

Accuracy test: 0.97728
Recall test: 0.80893
Precision test: 0.91943

Comparación de modelos




```
DecisionTreeClassifier
Accuracy
Train: 0.8979200997639514
Test: 0.8911155644622578
Recall
Train: 0.9024640657084189
Test: 0.8736517719568567
Precision
Train: 0.455440414507772  ✖
Test: 0.43615384615384617

matriz de confusión en train
 [[18403  2102]
 [  190  1758]]
matriz de confusión en test
 [[6103  733]
 [  82  567]]
```

```
GradientBoostingClassifier
Accuracy
Train: 0.9869505188616221
Test: 0.9772879091516367
Recall
Train: 0.8860369609856262
Test: 0.8089368258859785
Precision
Train: 0.9604897050639956
Test: 0.9194395796847635

matriz de confusión en train
 [[20434    71]
 [  222  1726]]
matriz de confusión en test
 [[6790   46]
 [ 124  525]]
```

```
RandomForestClassifier
Accuracy
Train: 0.9446844519663297
Test: 0.9424181696726787
Recall
Train: 0.4152977412731006 ✖
Test: 0.3882896764252696
Precision
Train: 0.8870614035087719
Test: 0.8811188811188811

matriz de confusión en train
 [[20402   103]
 [ 1139   809]]
matriz de confusión en test
 [[6802   34]
 [ 397  252]]
```

```
LogisticRegression
Accuracy
Train: 0.930922371175344
Test: 0.9286573146292585
Recall
Train: 0.3809034907597536 ✖
Test: 0.3420647149460709
Precision
Train: 0.6826126954921803 ✖
Test: 0.6747720364741642

matriz de confusión en train
 [[20160   345]
 [ 1206   742]]
matriz de confusión en test
 [[6729  107]
 [ 427  222]]
```

Mejora de modelos

```
DecisionTreeClassifier Mejorado
Accuracy
Train: 0.9519440609272704
Test: 0.9480293921175684
Recall
Train: 0.6380903490759754
Test: 0.6163328197226502
Precision
Train: 0.7687074829931972
Test: 0.7407407407407407

matriz de confusión en train
 [[20131   374]
 [  705  1243]]
matriz de confusión en test
 [[6696  140]
 [ 249  400]]
```

```
GradientBoostingClassifier Mejorado
Accuracy
Train: 0.9530574978844698
Test: 0.9500334001336005
Recall
Train: 0.5395277207392197
Test: 0.5115562403697997
Precision
Train: 0.8700331125827815
Test: 0.8534704370179949

matriz de confusión en train
 [[20348   157]
 [  897  1051]]
matriz de confusión en test
 [[6779   57]
 [ 317  332]]
```

```
RandomForestClassifier Mejorado
Accuracy
Train: 0.9997327751302721
Test: 0.9569806279225117
Recall
Train: 0.9974332648870636
Test: 0.5824345146379045
Precision
Train: 0.9994855967078189
Test: 0.8811188811188811

matriz de confusión en train
 [[20504     1]
 [    5  1943]]
matriz de confusión en test
 [[6785   51]
 [ 271  378]]
```

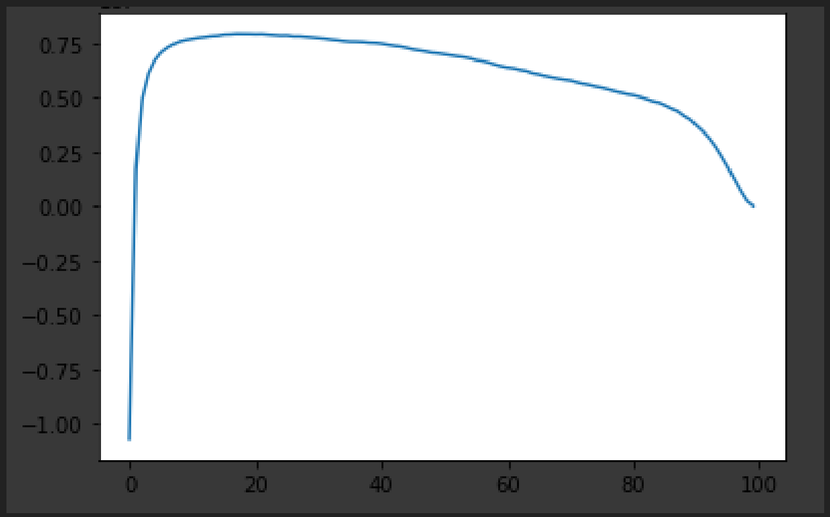
```
KNN Classifier
Accuracy
Train: 0.9402752416158197
Test: 0.931997327989312
Recall
Train: 0.45071868583162217
Test: 0.37442218798151
Precision
Train: 0.7641427328111401
Test: 0.7023121387283237

matriz de confusión en train
 [[20234   271]
 [ 1070   878]]
matriz de confusión en test
 [[6733  103]
 [ 406  243]]
```

Testeo del Modelo

Probamos el modelo entrenado con diferentes set de datos para comprobar su funcionamiento.

Predicciones del Modelo



Curva de ganancia

Calculamos la curva de ganancia correspondiente a la ganancia y perdida generada por los verdaderos positivos y los falsos positivos respectivamente.

Selección de treshold

En el gráfico podemos observar dos puntos de inflexión que marcan donde empezamos a perder plata si nos equivocamos en la predicción.
Para correr el modelo elegimos un treshold de 0.85

Clientes que van a desertar en los próximos 2 meses

```
predicions
0    175209
1      2793
```

De los 178.002 clientes activos en el período 2014-04, 2.793 desertarán la compañía en 2014-06 (baja+2)

Ganancia generada por retener a los clientes

```
Max: 4900
Min: -100
Total: 2552600
Bajas: 649
Predicciones: 574
Max: 2552700
Total: 2552600
```

Con una ganancia estimada de $4900 por cada cliente por darse de baja bien identificado y una perdida de $100 por los mal identificados, podemos estimar que logrando retener a los clientes identificados obtenemos una ganancia de $2.500.000

Estrategias de Retención

A quienes retenemos?

Una variable importante a la hora de predecir los clientes propensos a darse de baja es la **edad del cliente**. Un grupo etario muy probable a darse de baja son los mayores de 80 años. ¿Queremos invertir plata en acciones para este grupo?

Estrategias a implementar

- Campañas de tiempo limitado con bajas tasas de interés
- Ofrecer programas VIP para los clientes de mayor antiguedad/mayor actividad
- Personalizar los programas de fidelización, generando una conexión emocional con el banco

Próximos pasos

DATA DRIVEN INSIGHTS

Customer Journey

Analizar todo el recorrido del cliente en la empresa, especialmente de los desertores, para detectar qué los lleva a dejar el banco.
Integrar con otros datos disponibles (redes sociales, CRM, customer care) para tener una visión holística de la situación.

Deserciones parciales

Realizar un análisis más profundo sobre el volumen de transacciones para detectar clientes, que sin cerrar su cuenta, estén migrando a otros bancos su operatoria mayoritaria.

El equipo

Callejón Alberio, Lucía - 1082770
Romero, Emiliano Patricio - 1073072
Stefanuto, Lucía - 102950
Szulak, Florencia - 1068717

Preguntas?
Dudas?

Aplausos