

# Predicción de fuga de clientes “The Bridge Bank”

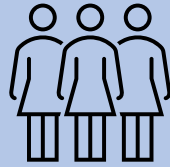
Marzo 2022

Jorge Marchan



# Objetivo del proyecto

El objetivo de este proyecto es hacer un assessment de la fuga de la base de clientes de The Bridge Bank y desarrollar un modelo de clasificación para predecir y reducir la fuga de los clientes



**Muestra de 10.000...**



**...en 3 geografías...**

**...en distintas etapas de su vida....**



Estudiantes  
Universitarios



Jóvenes  
profesionales



Profesionales  
con experiencia



Profesionales  
consolidados



Jubilados y  
pensionistas

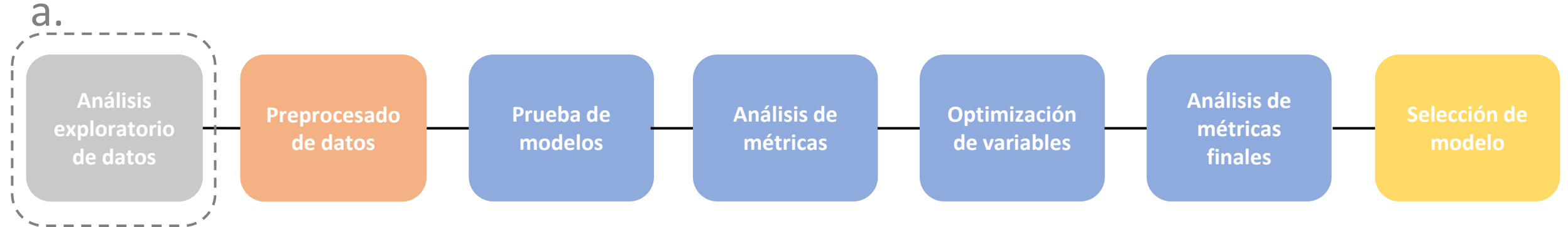
**... con distintos niveles de ingresos ....**

- 0 €
- 1 € - 25.000 €
- 25.000 € - 75.000 €
- 75.000 € - 125.000 €
- 125.000 € - 175.000 €
- 175.000 € <


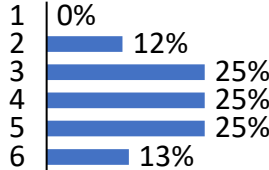
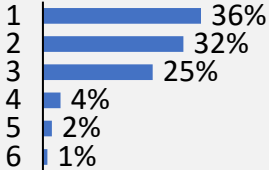

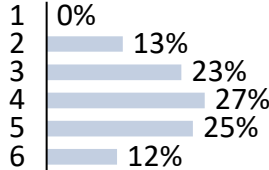
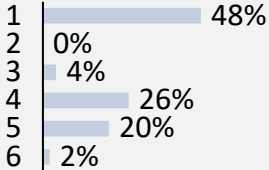

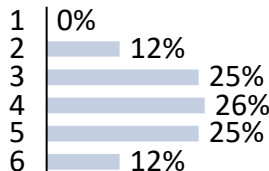
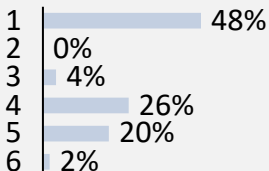

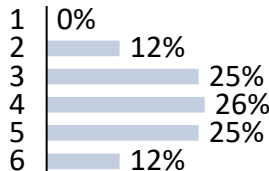
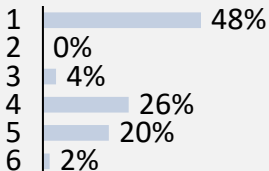
**...y con distintas  
necesidades financieras**

# Pipeline de trabajo

a.

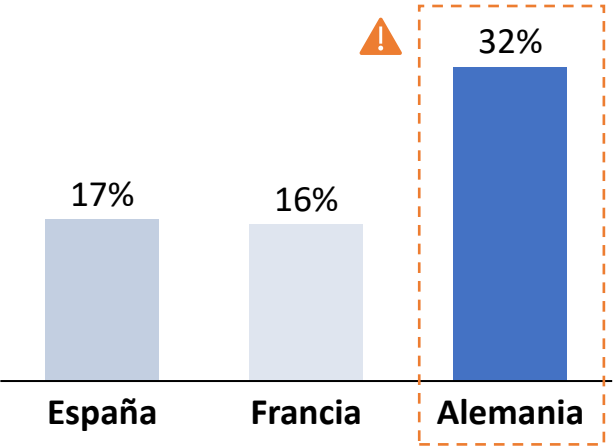


# Descripción de la base de clientes

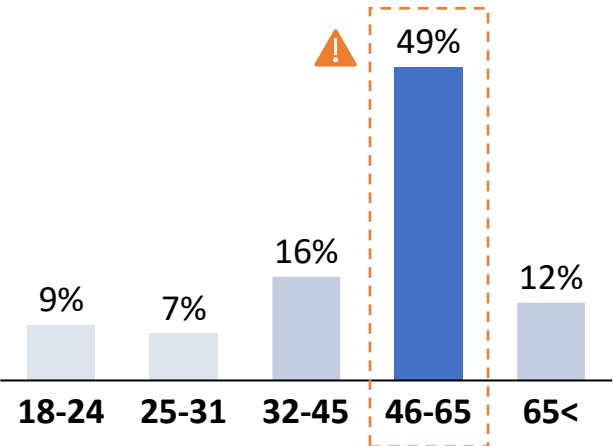
|   | Total Clientes   | Clientes activos | Distribución salario  | Distribución balance  | Score medio | Permanencia | Fuga           |
|---|------------------|------------------|---|---|-------------|-------------|----------------|
|  General    | 10.000           | 5.151<br>(51,5%) |    |    | 650,5       | 5 años      | 2.037<br>(20%) |
|  España     | 2.477<br>(24,8%) | 1.312<br>(53,0%) |    |    | 651,3       | 5 años      | 413<br>(17%)   |
|  Francia    | 5.014<br>(50,1%) | 2.591<br>(51,7%) |    |    | 649,6       | 5 años      | 810<br>(16%)   |
|  Alemania | 2.509<br>(25,1%) | 1.248<br>(49,7%) |  |  | 651,5       | 5 años      | 816<br>(32%)   |

# Fuga por país y edad

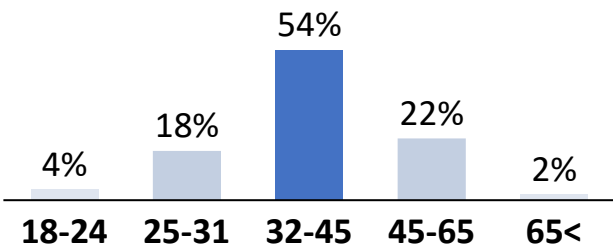
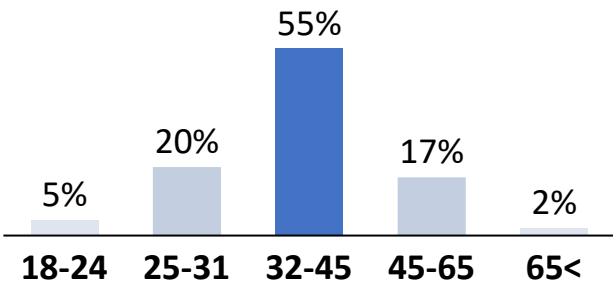
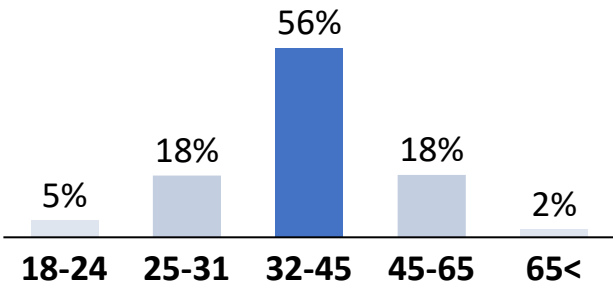
Fuga por país



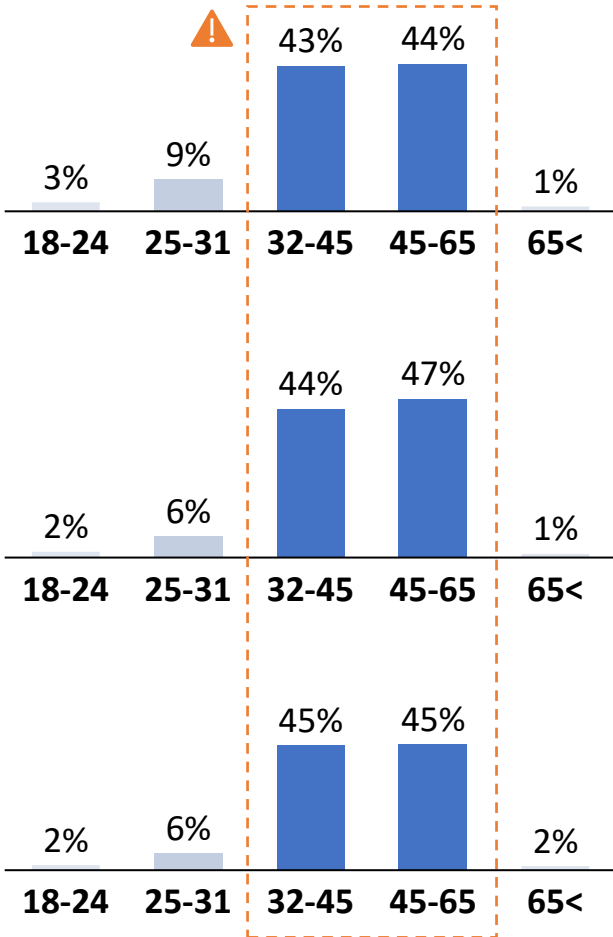
Fuga por rango de edad



Distribución por edad y país

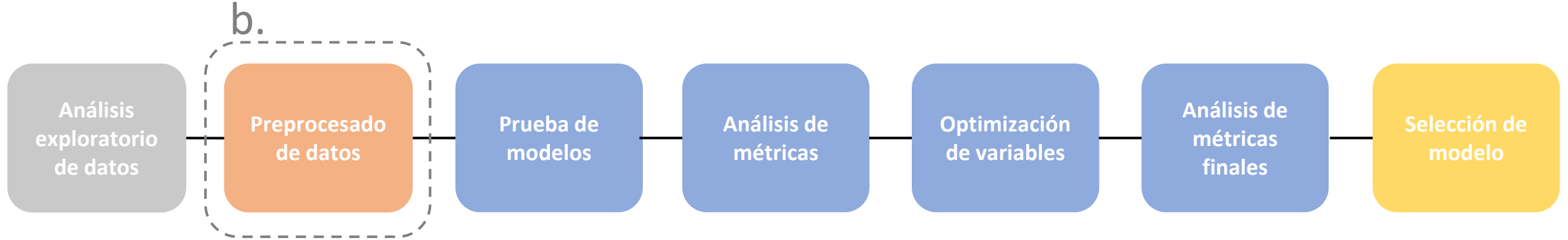


Fuga por edad y país



# Preparamos los datos disponibles...

b.



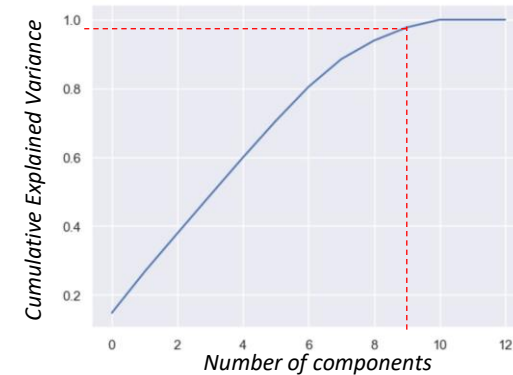
1

| <u>Tipología</u>      | <u>Variables</u>  | <u>Función</u>   |
|-----------------------|---|------------------|
| Variables categóricas | Género y Geografía  | One Hot Encoding |
| Variables numéricas   | Edad, Salario, Balance, Actividad, Productos, Permanencia, Credit score, tarjeta de crédito | Standard Scaler  |
| Variables que remover | Id, Apellido, Rango de edad, Rango de salario y Rango de balance                            | Remove Columns   |

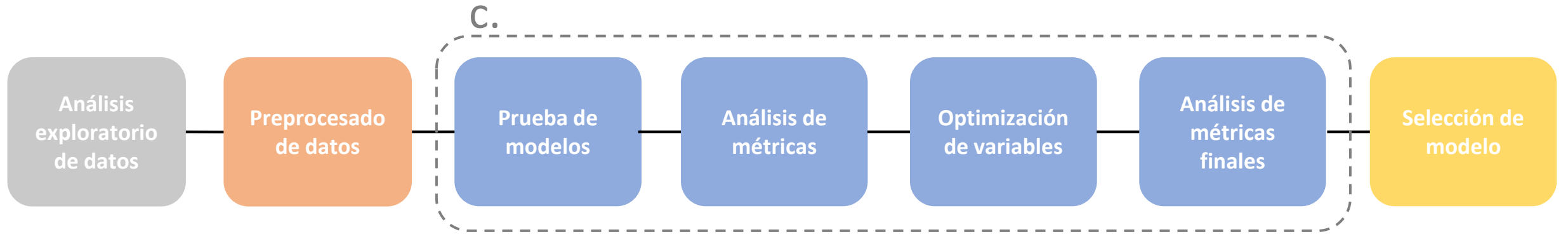
2

## Principal Components Analysis

n\_components = 9



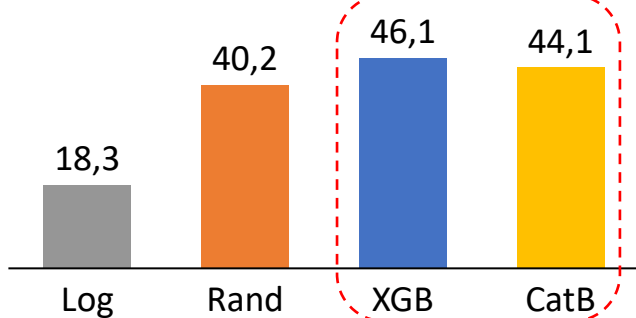
# ...empezamos a probar distintos modelos...



1

## Principal Components Analysis

- *LogisticRegression*
  - *RandomForestClassifier*
  - *XGBClassifier*
  - *CatBoostClassifier*
- CV = 10 Scoring = "recall"



2

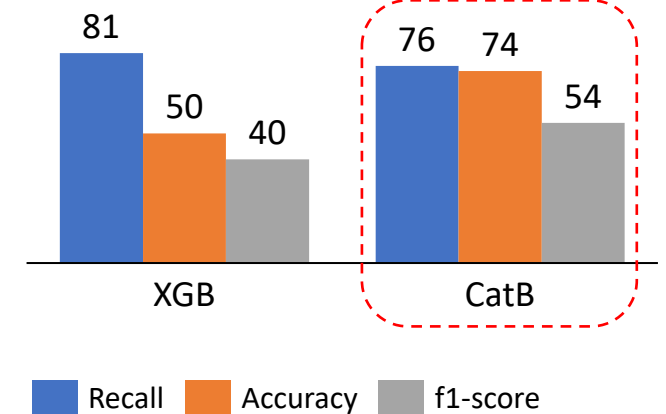
## Grid Search

```
pipe_xgb_param = {  
    'Modelo__n_estimators': [400, 450, 480, 490],  
    'Modelo__scale_pos_weight': [4],  
    'Modelo__max_depth': [5, 6, 7],  
    'Modelo__learning_rate': [2.0, 2.5, 3.0]}  
  
pipe_catb_param = {  
    'Modelo__n_estimators': [330, 345, 360, 365],  
    'Modelo__scale_pos_weight': [8],  
    'Modelo__max_depth': [2, 3, 4],  
    'Modelo__learning_rate': [0.88, 0.9, 0.92]}
```

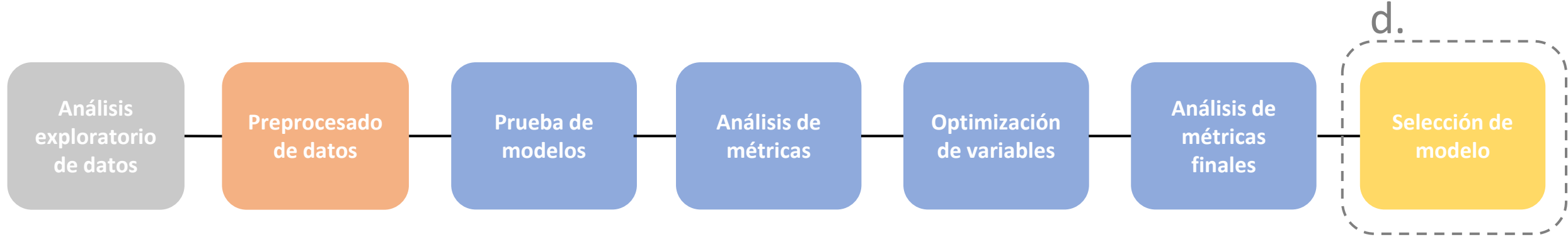
CV = 8 Scoring = "recall"

3

## Métricas finales



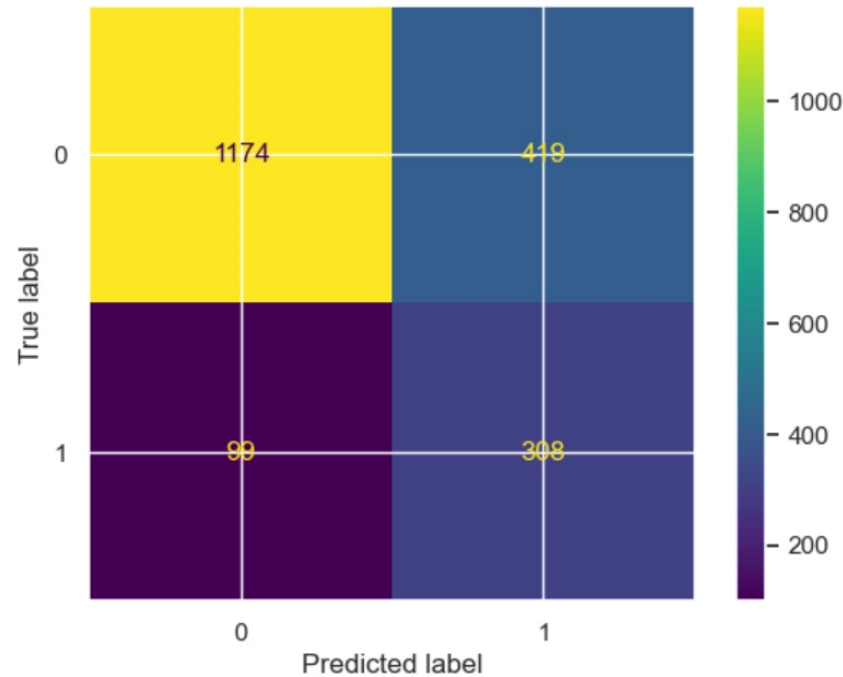
# ...y el ganador es...



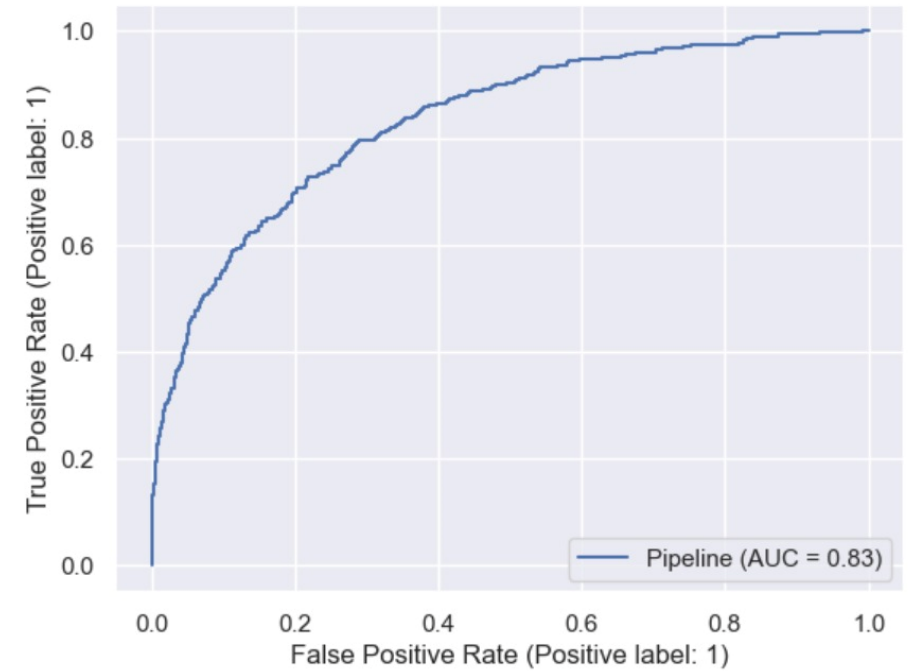
CatBoost Classifier



Matriz de Confusión



ROC Curve





# Puntos de mejora:

Disponer de **datos de negocio**  
como **coste de retención o**  
**LTV** por cliente

1

2

3

4

Hacer un **estudio por geografía** y  
ver si separando las bases de  
clientes por país podemos  
**optimizar más el modelo**

Hacer un business case para  
**optimizar recursos** y en  
función de este análisis  
**calibrar el accuracy y recall**  
del modelo

Probar un **sistema de**  
**votación** con varios  
modelos **para robustecer**  
la **predicción de la fuga**

¡¡Muchas Gracias!!

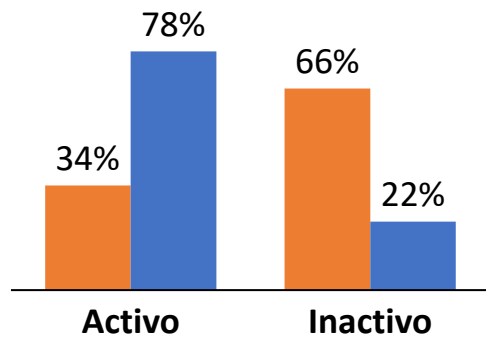
# Anexos



# Alemania: Deep dive de la fuga por edad

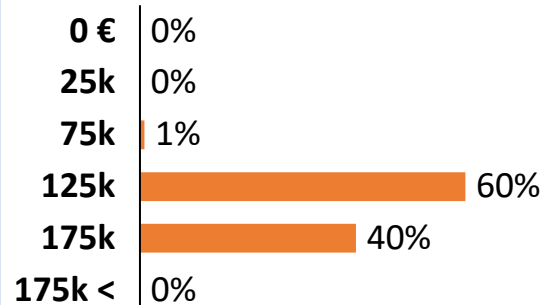
1

## Fuga por actividad



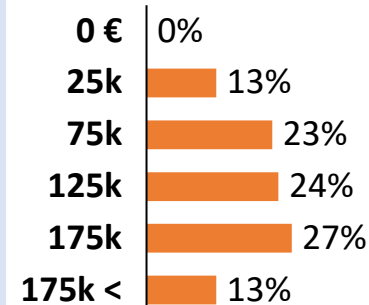
2

## Fuga por balance



3

## Fuga por salario

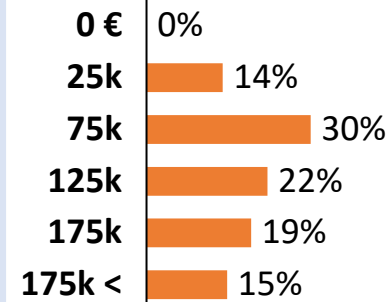
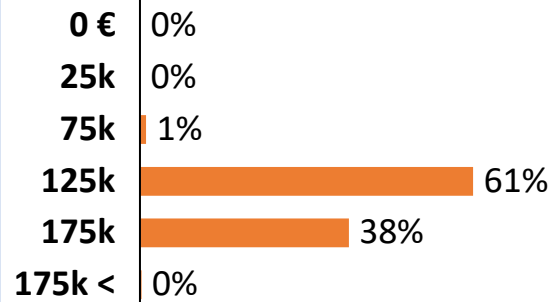
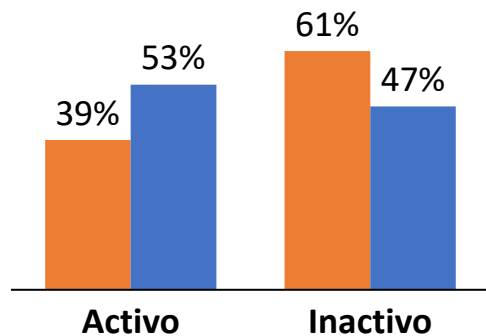


1

46 – 65

2

32 - 45



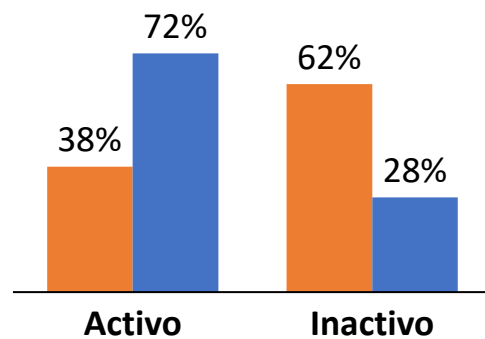
Fuga No fuga



# Francia: Deep dive de la fuga por edad

1

## Fuga por actividad

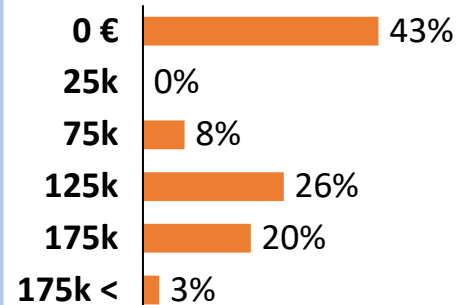


1

46 – 65

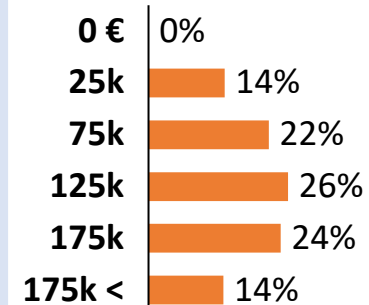
2

## Fuga por balance



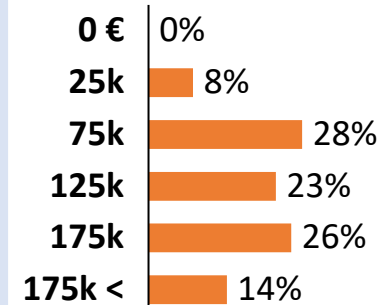
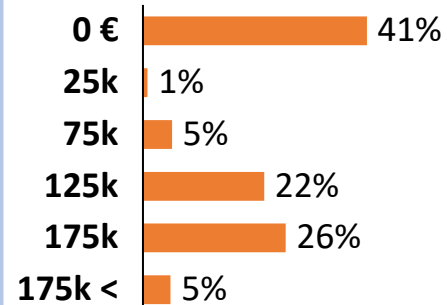
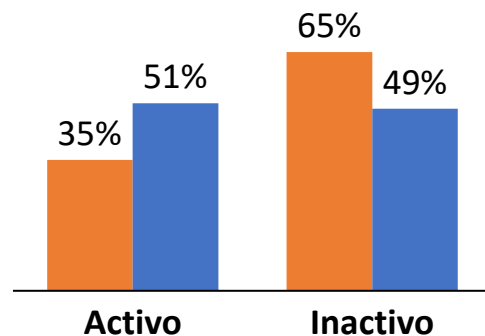
3

## Fuga por salario



2

32 - 45



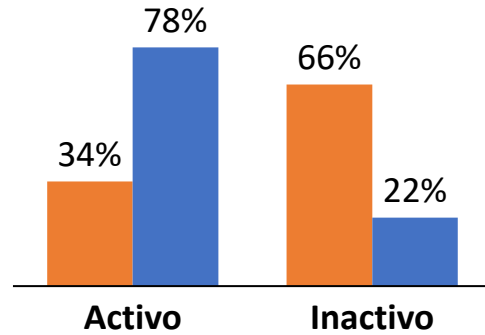
Fuga No fuga



# España: Deep dive de la fuga por edad

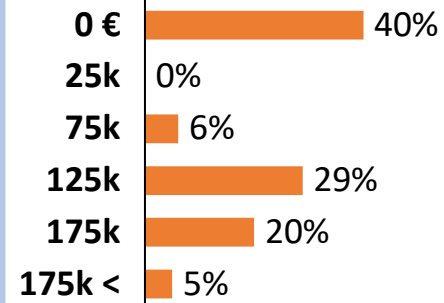
1

## Fuga por actividad



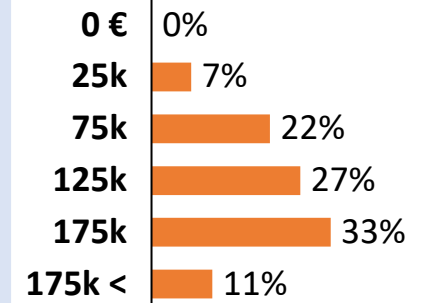
2

## Fuga por balance



3

## Fuga por salario

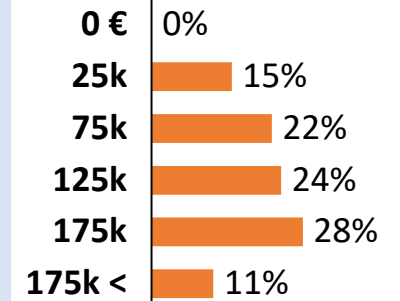
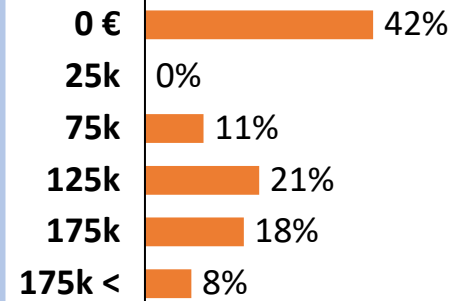
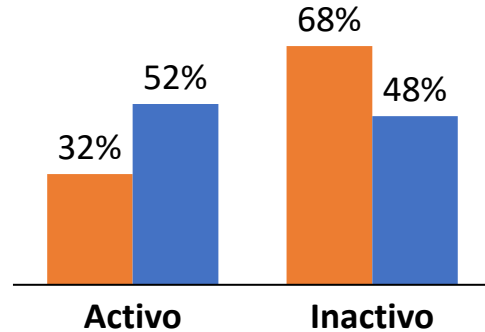


1

46 – 65

2

32 - 45

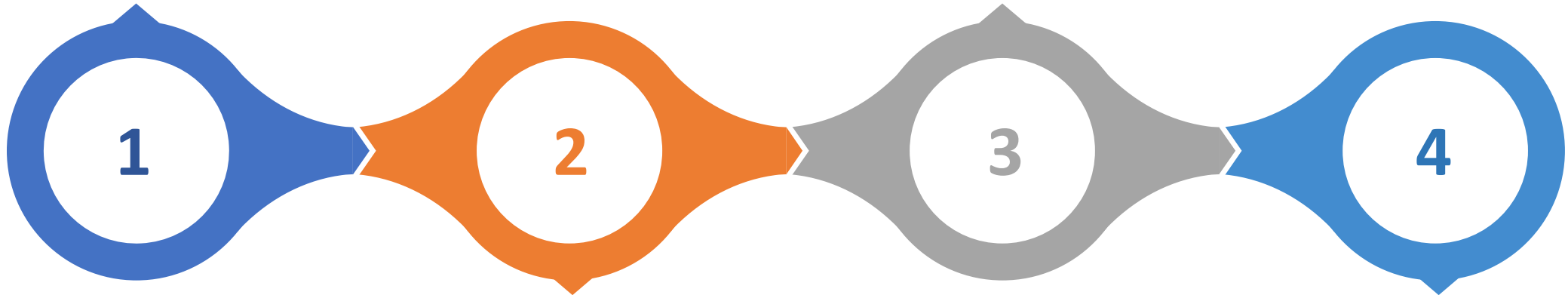


Fuga No fuga

# Takeaways del análisis

La fuga se concentra en clientes entre **32 y 65 años** de edad, especialmente en **Alemania**

En España y Francia más del **40% de lo clientes que salen** tienen un **balance en cuenta de 0€**, mientras que en **Alemania la fuga se centra en más de 125k €**



La **mayor parte** de la fuga, independientemente del país, viene de la **inactividad de los clientes**

Más del **50%** de los **clientes que se fugan** tienen un **salario** estimado de **más de 125k €**