



Instituto Tecnológico  
de Buenos Aires

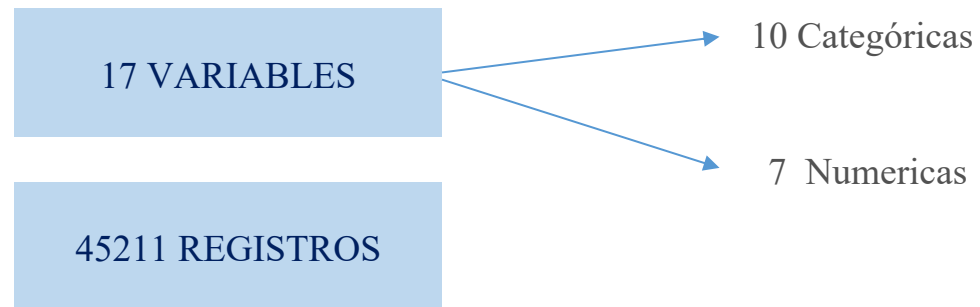
CHALBAUD, MILAGROS

ANÁLISIS PREDICTIVO

—

# AGENDA

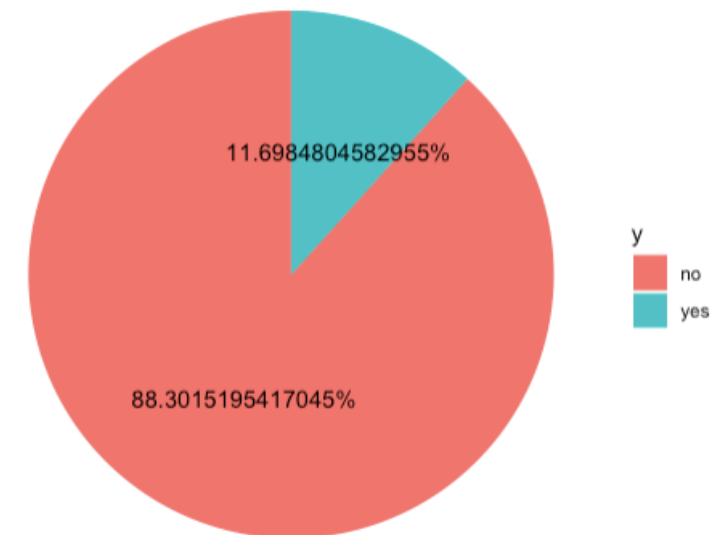
- Objetivo del trabajo
- Descripción de la base
- Análisis Univariado
- Análisis Bivariado
- Modelos
- Conclusiones



### *Variable a predecir : Y*

*El objetivo de este análisis predictivo es desarrollar un modelo capaz de predecir si un cliente se suscribirá a un depósito a plazo fijo. Utilizando datos históricos y características relevantes de los clientes, el modelo buscará identificar patrones y factores determinantes que influyan en la suscripción a este producto financiero.*

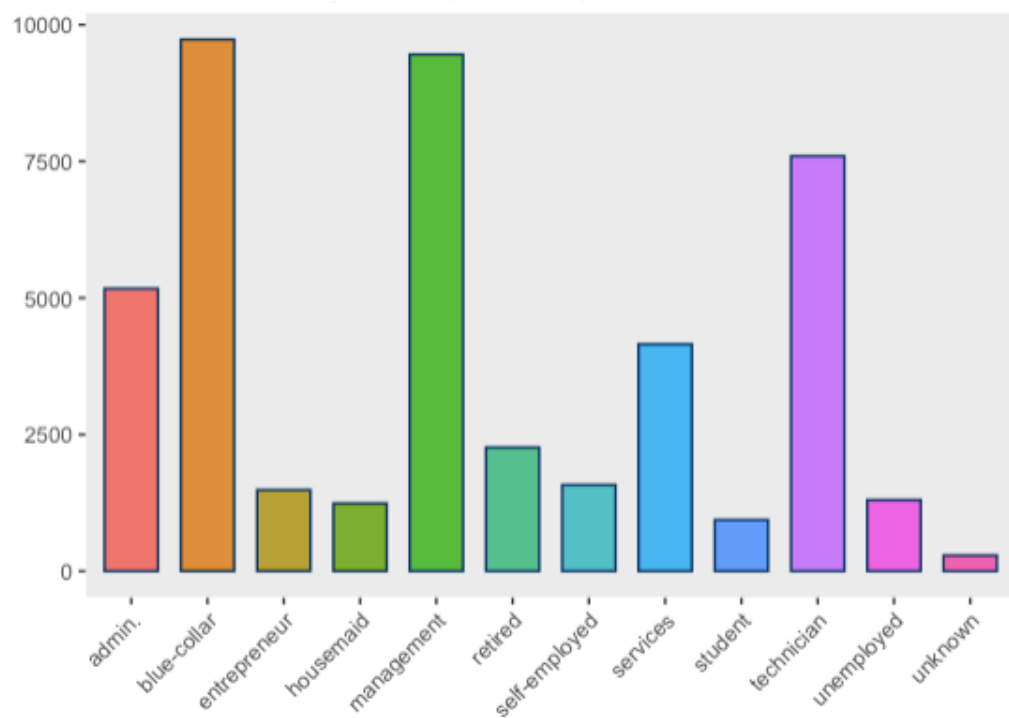
*El resultado esperado es obtener un modelo preciso y confiable que pueda predecir de manera efectiva si un cliente se suscribirá o no a un depósito a plazo fijo, lo que permitirá a la institución financiera tomar decisiones informadas y desarrollar estrategias de marketing más eficientes.*



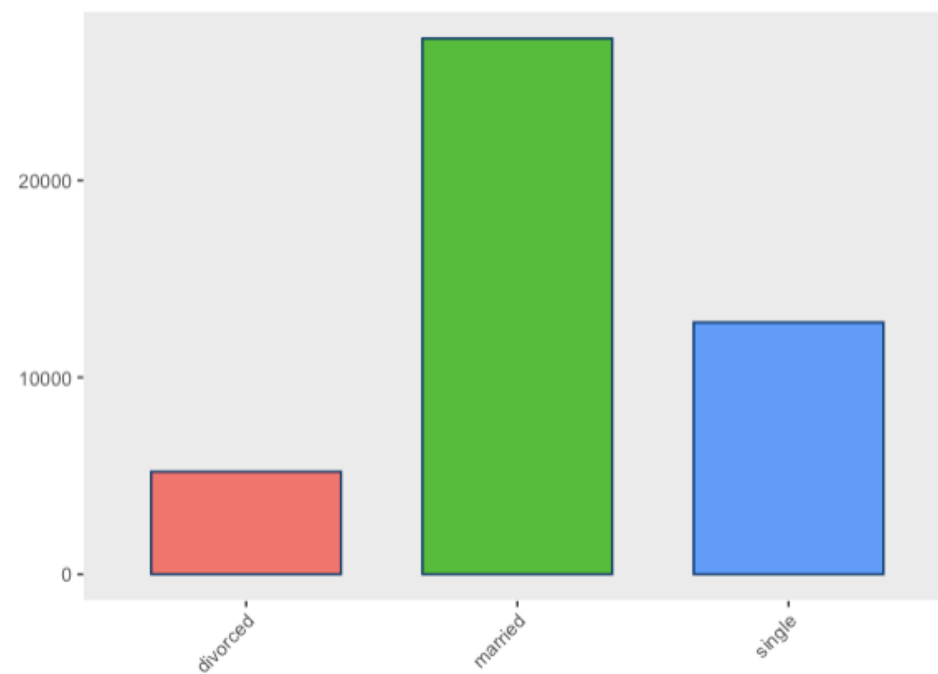
VARIABLE	DESCRIPCION
age	Edad del cliente
job	Tipo de trabajo del cliente
marital	Estado civil del cliente
education	Nivel educativo del cliente
default	Indicador de incumplimiento crediticio
balance	Saldo de la cuenta del cliente
housing	Indicador de si el cliente tiene una hipoteca
loan	Indicador de si el cliente tiene un préstamo personal
contact	Método de contacto utilizado para comunicarse con el cliente
day	Día del mes en el que se realizó el último contacto
month	Mes en el que se realizó el último contacto
duration	Duración del último contacto en segundos
campaign	Número de contactos realizados durante esta campaña
pdays	Número de días que han pasado desde el último contacto anterior
previous	Número de contactos realizados antes de esta campaña
poutcome	Resultado de la campaña de marketing anterior
y	Variable objetivo que indica si el cliente se suscribió o no a un depósito a plazo fijo

# ANÁLISIS UNIVARIADO

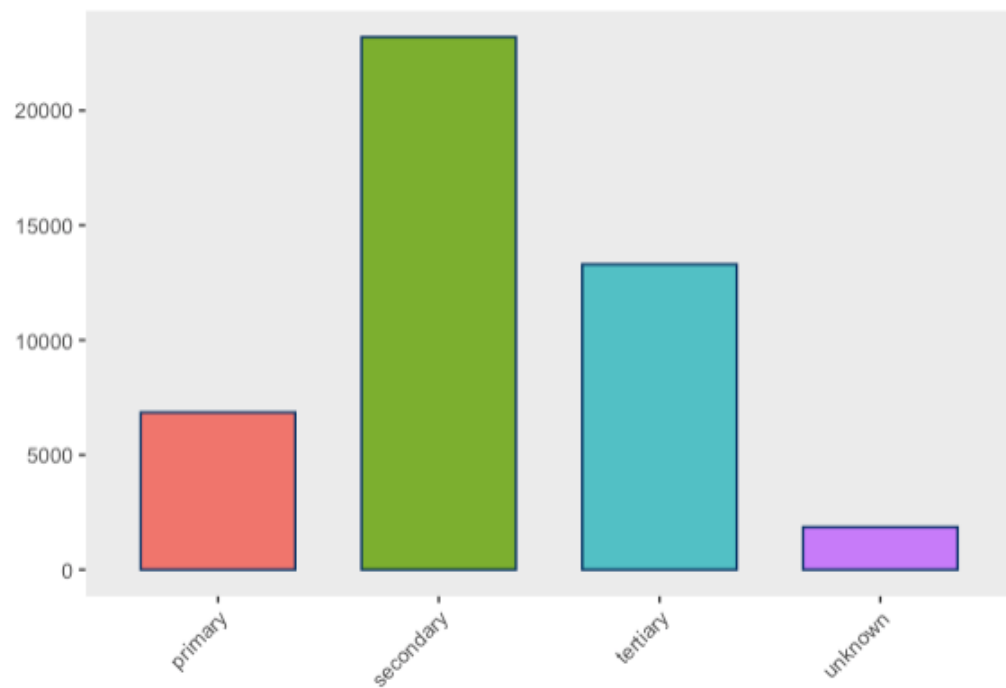
## Job



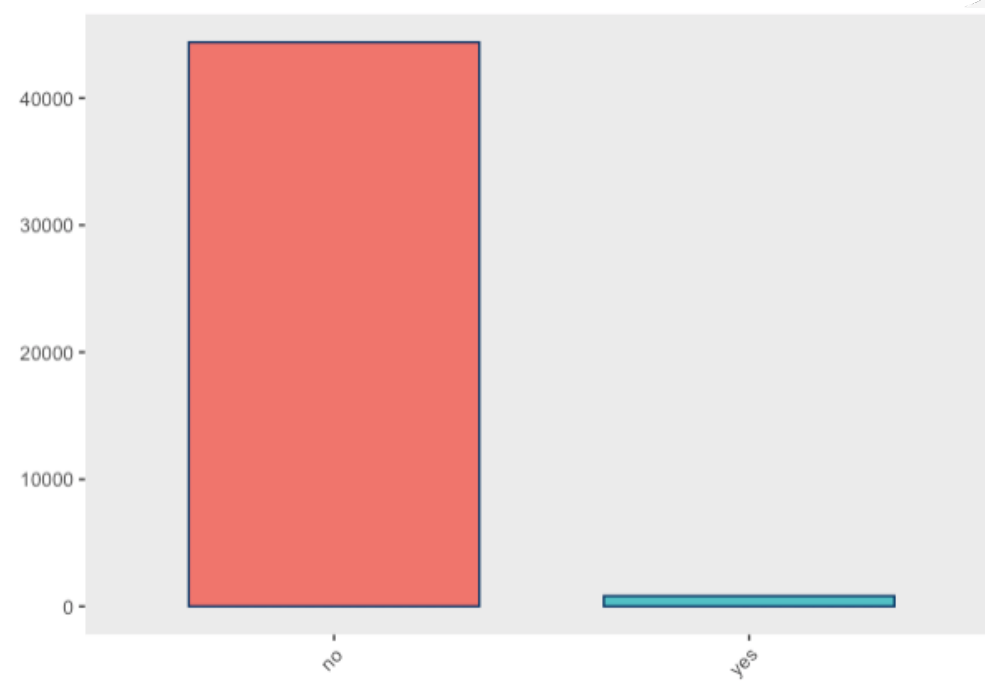
## Marital



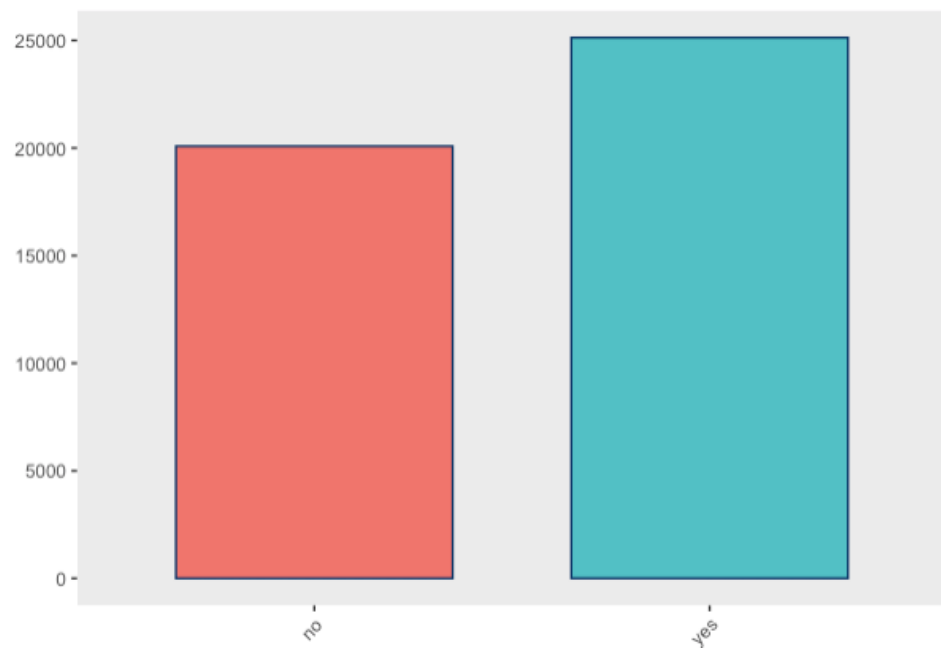
## Education



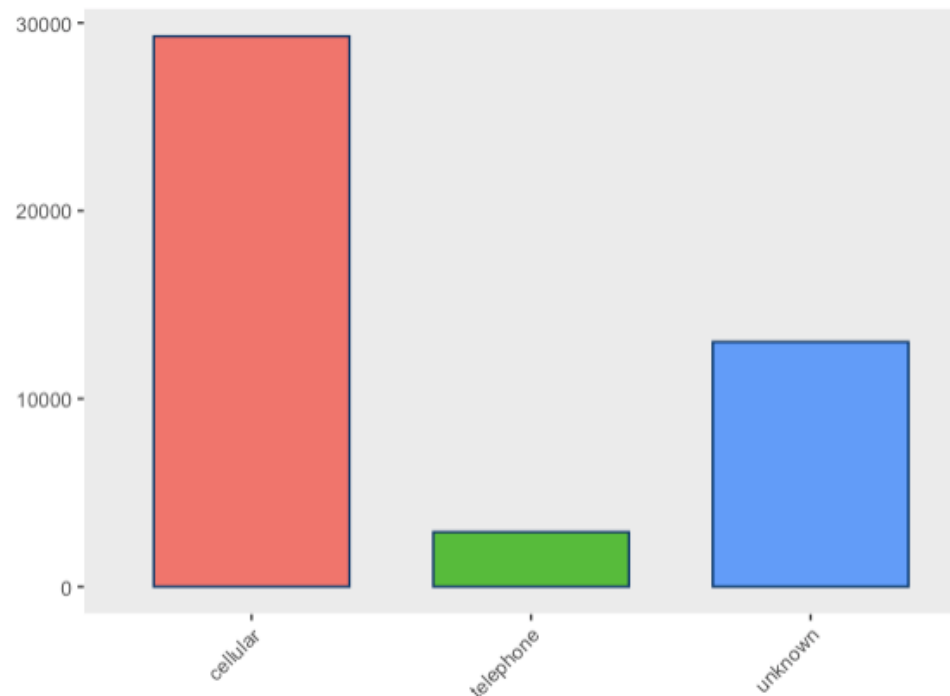
## Default



## Housing

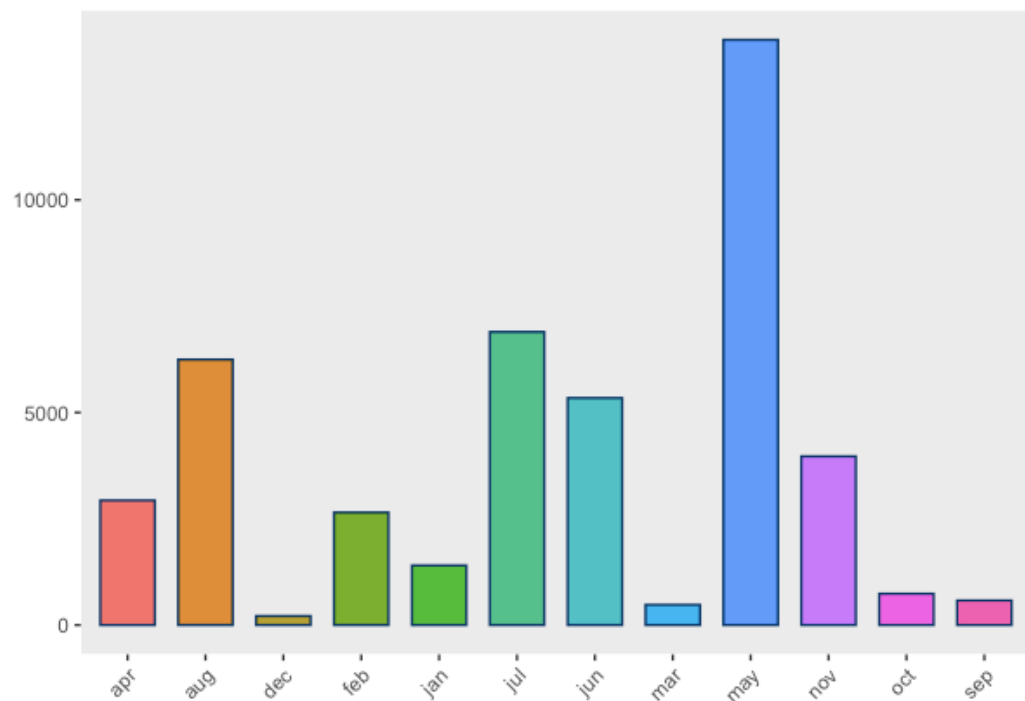


## Contact

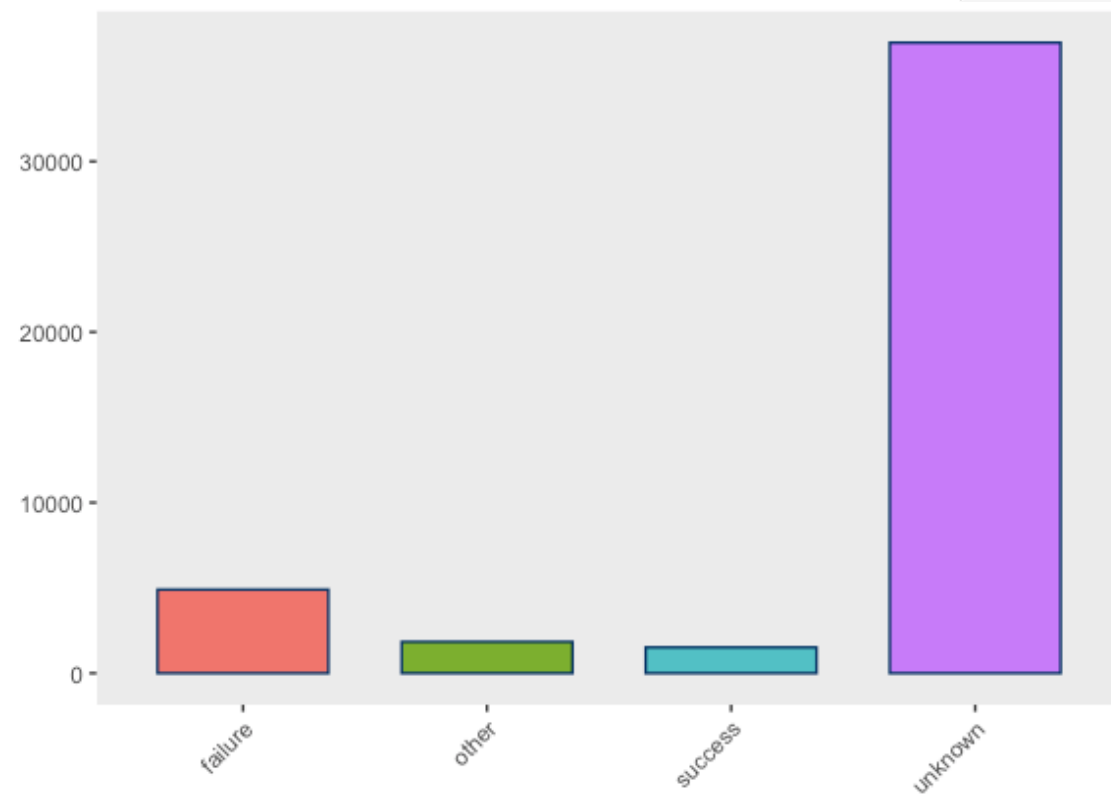




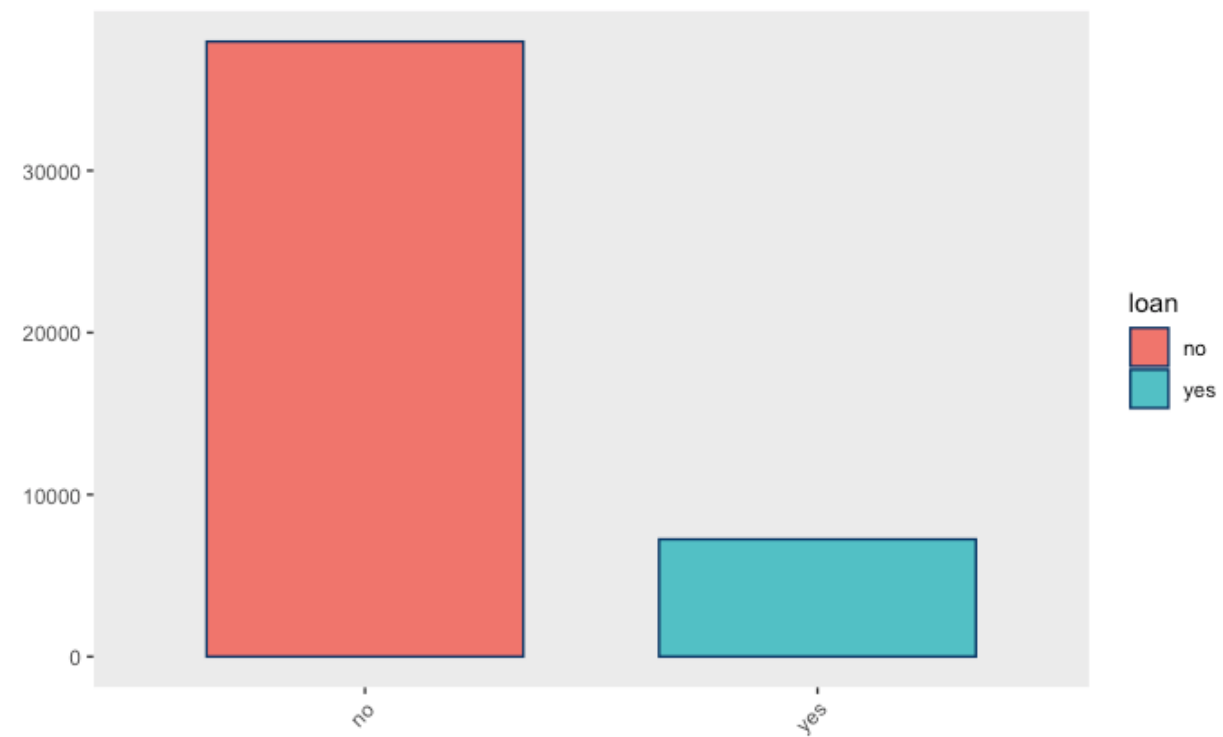
Month



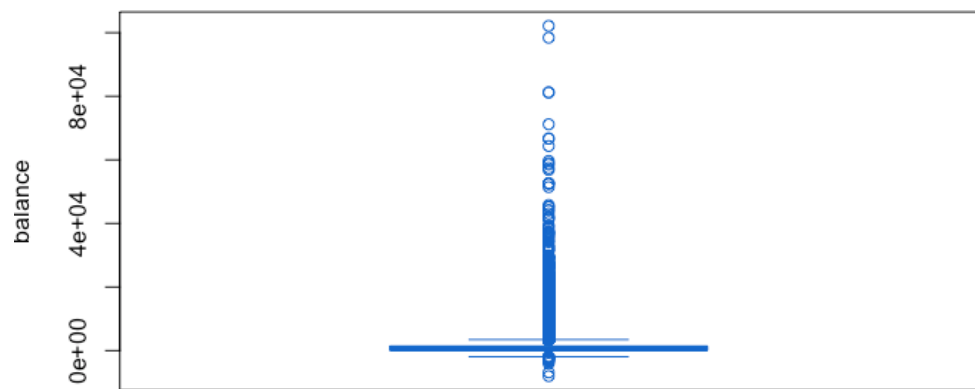
Poutcome



## Loan



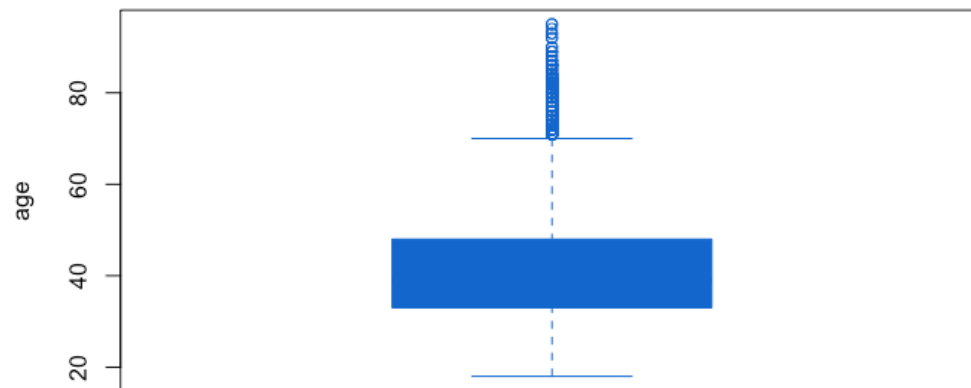
## Balance



Min.	1st Q.	Median	Mean	3rd Q.	Max.
-8019	72	448	1362	1428	102127

Los outliers fueron reemplazados por la media.

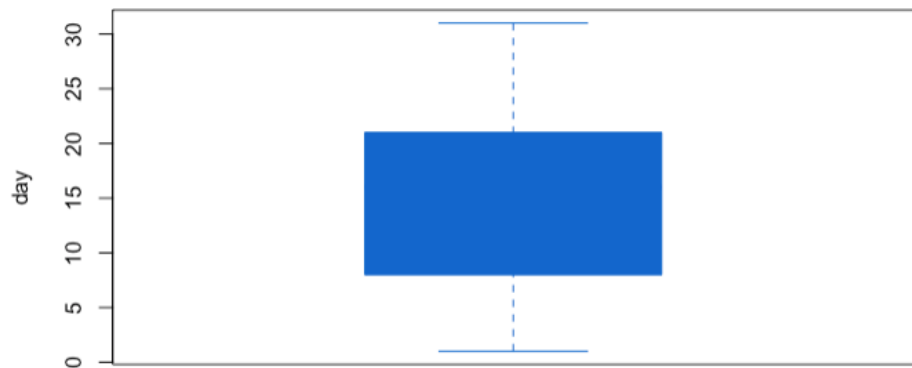
## Age



Min.	1st Q.	Median	Mean	3rd Q.	Max.
18.0	33.0	39.0	40.94	48.0	95.0

Los outliers fueron reemplazados por la media

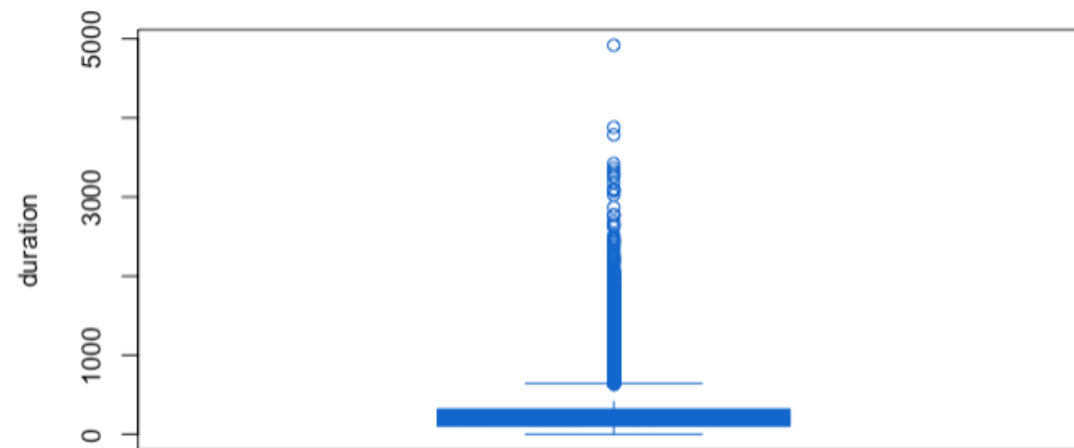
## Day



Min.	1st Q.	Median	Mean	3rd Q.	Max.
1	8	16	15.81	21	31

No presenta outliers.

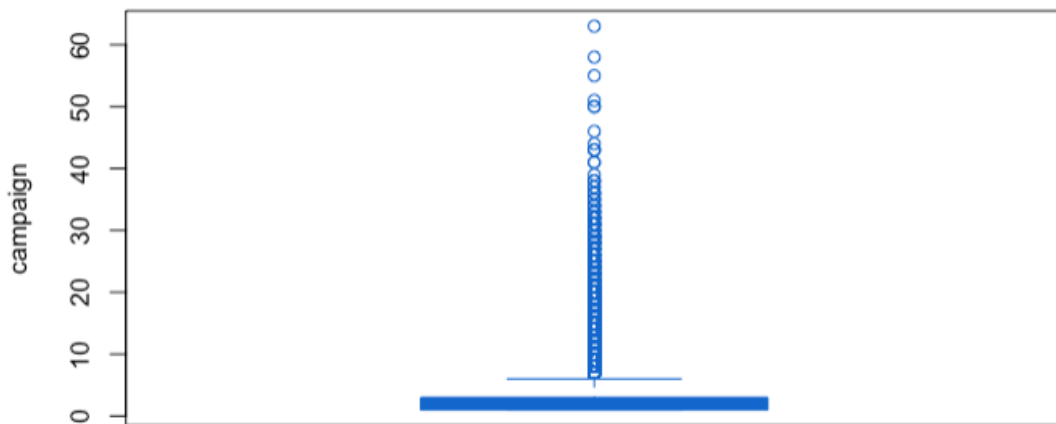
## Duration



Min.	1st Q.	Median	Mean	3rd Q.	Max.
0.0	103.0	180.0	258.2	319.0	4918.0

Los outliers fueron reemplazados por la media.

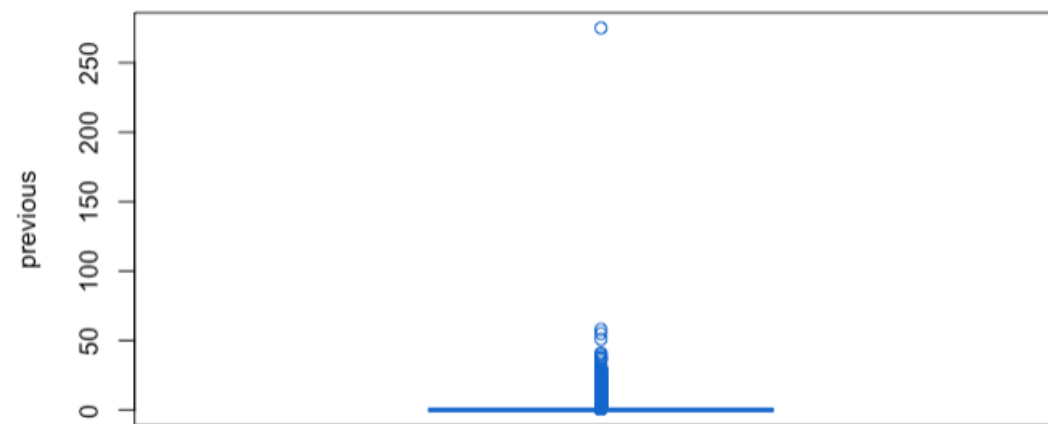
## Campaign



Min.	1st Q.	Median	Mean	3rd Q.	Max.
1	1	2	2.764	3	63

Los outliers fueron reemplazados por la media.

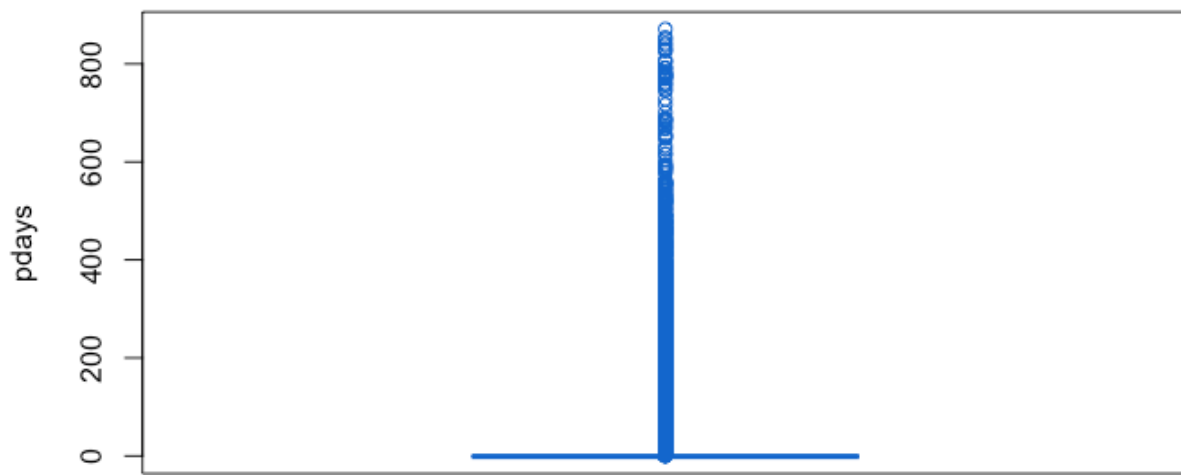
## Previous



Min.	1st Q.	Median	Mean	3rd Q.	Max.
0.0	0.0	0.0	0.5803	0.0	275.0

Los outliers fueron reemplazados por la media.

## Pdays



Min.	1st Q.	Median	Mean	3rd Q.	Max.
-1	-1	-1	40.2	-1	871

Los outliers fueron reemplazados por la media.

# ANÁLISIS BIVARIADO

## CHI-SQUARE TEST

- Job → p-value < 2.2e-16
- Marital → p-value < 2.2e-16
- Education → p-value < 2.2e-16
- Default → p-value = 2.454e-06
- Housing → p-value < 2.2e-16
- Contact → p-value < 2.2e-16
- Month → p-value < 2.2e-16
- Poutcome → p-value < 2.2e-16
- Loan → p-value < 2.2e-16

$H_0$ : las variables son independientes  
 $H_a$ : las variables son dependientes

Para todos los test rechazo  $H_0$ , por lo tanto se puede afirmar con un 95% de confianza que todas las variables son dependientes. Hay relación estadísticamente significativa entre la variable objetivo y las demás.



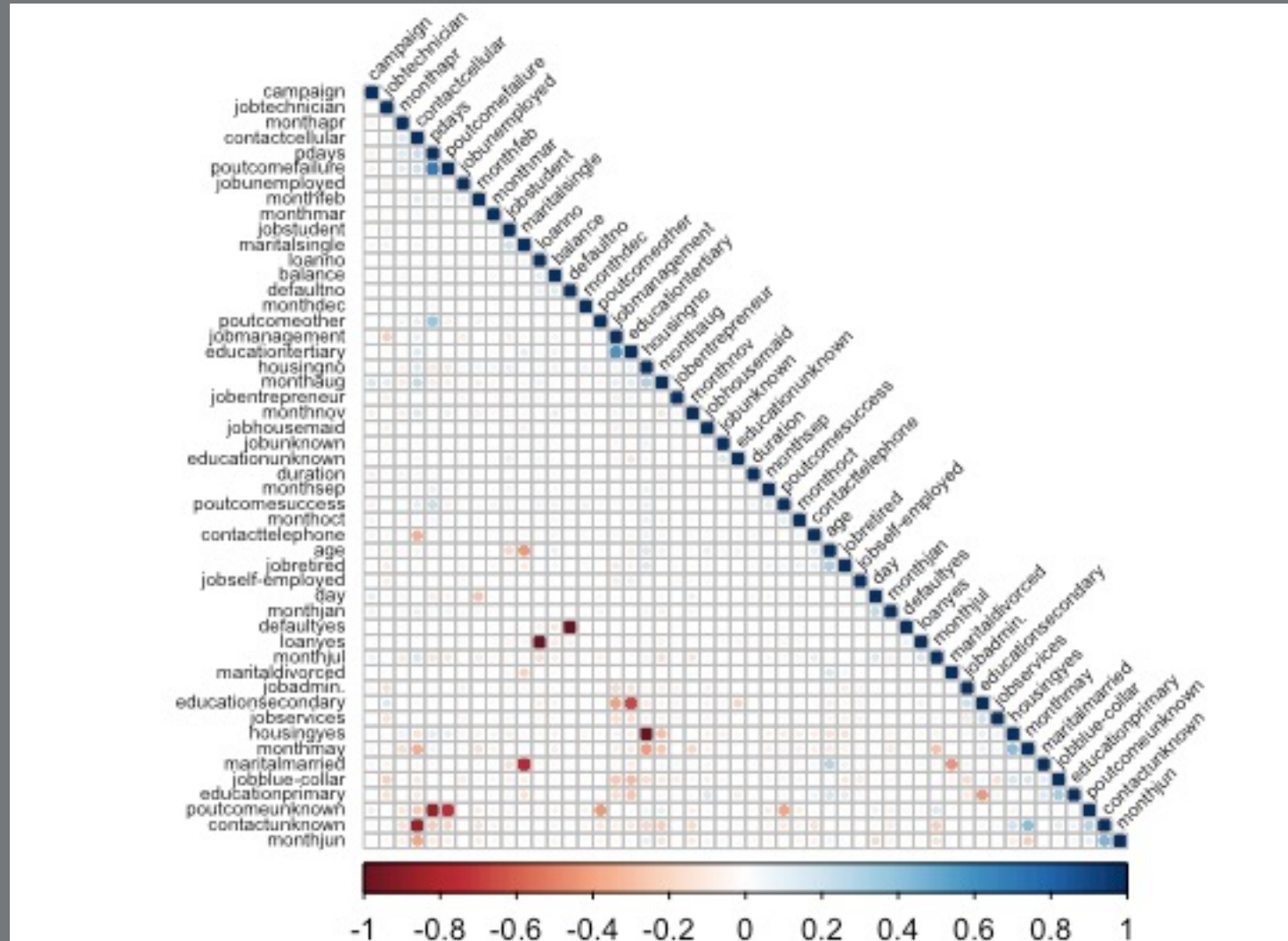
## T STUDENT

- Balance → p-value < 2.2e-16
- Age → p-value = 1.597e-05
- Day → p-value = 3.441e-09
- Duration → p-value < 2.2e-16
- Campaign → p-value < 2.2e-16
- Previous → p-value < 2.2e-16
- Pdays → p-value < 2.2e-16

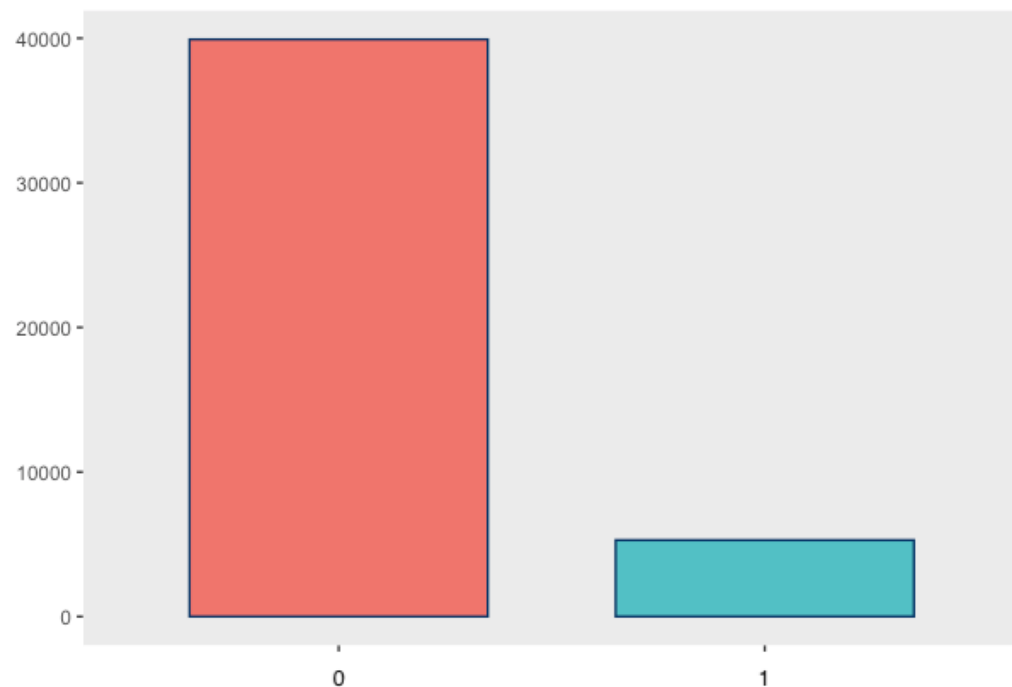
H0 : no hay diferencia entre las medias  
Ha: si hay diferencia entre las medias

Para todos los test rechazo H0, por lo tanto se puede afirmar con un 95% de confianza que todas las variables son dependientes. Hay relación estadísticamente significativa entre la variable objetivo y las demás

# MATRIZ DE CORRELACIÓN



## BALANCEO DE CLASES



**0: 39922**

**1: 5289**

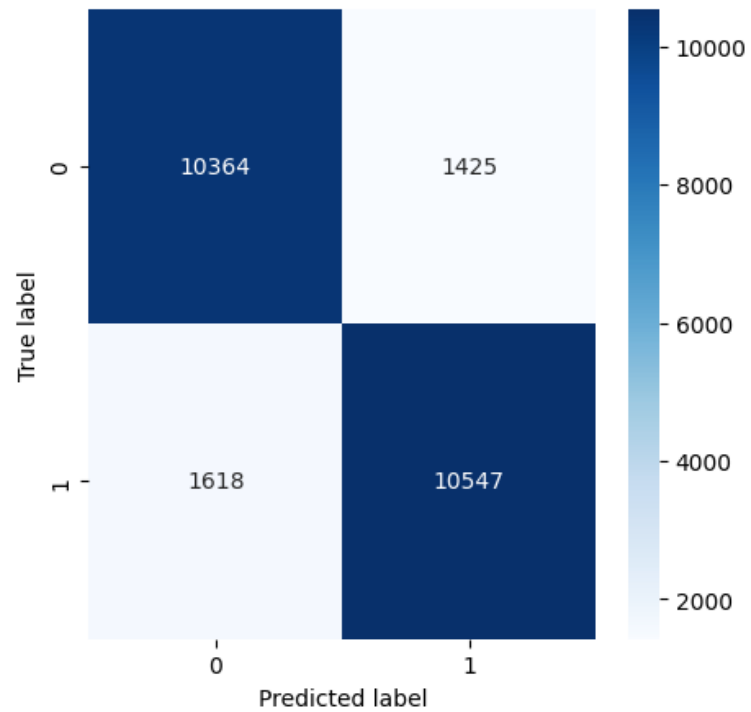
## PARTICIÓN DE LA BASE

Tamaño del conjunto de entrenamiento: 55890 (70%)

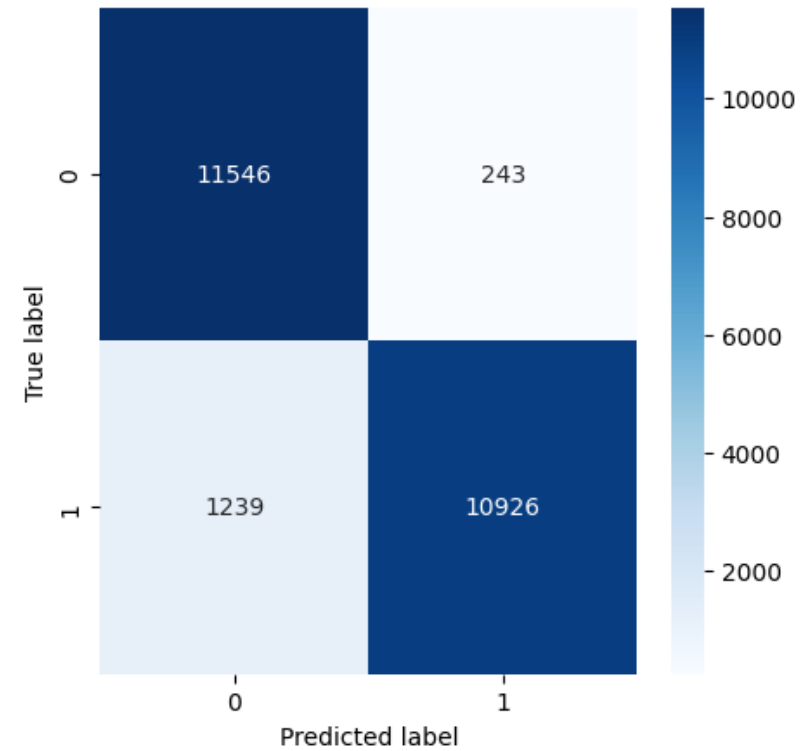
Tamaño del conjunto de testeo: 23954 (30%)

# MODELOS PREDICTIVOS

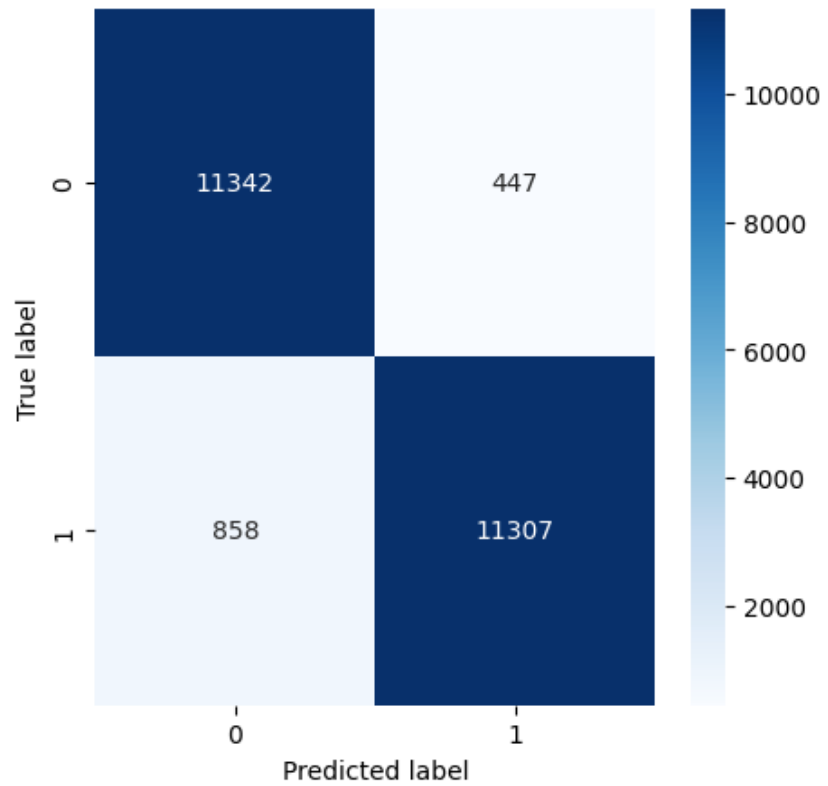
## NAIVE BAYES



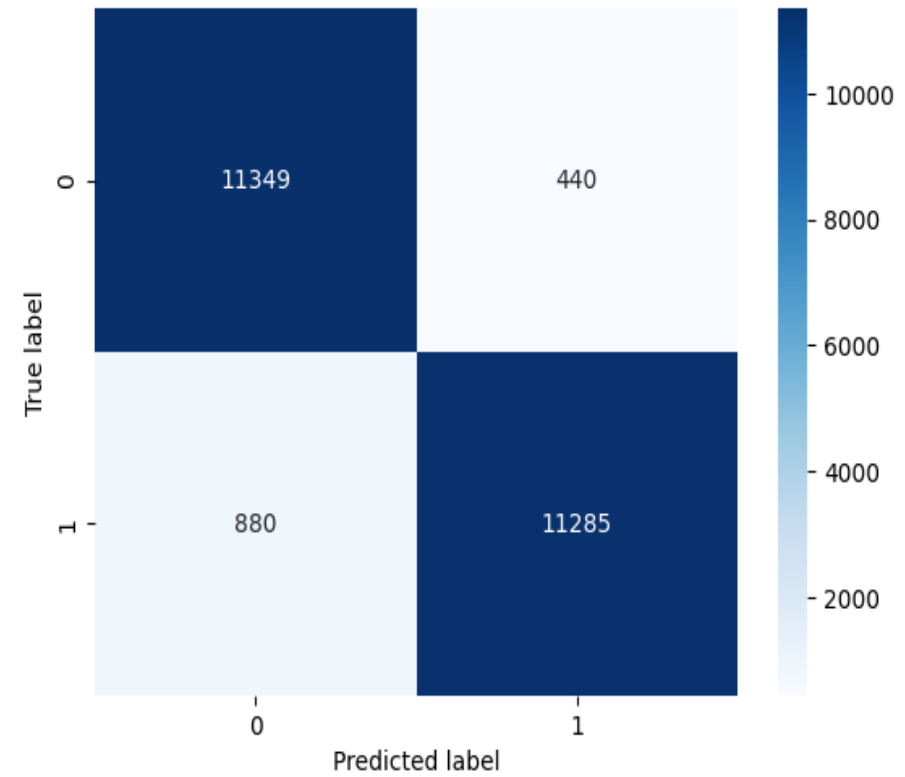
## REGRESIÓN



## XG BOOST



## RANDOM FOREST



## COMPARACIÓN DE MODELOS

	NAIVE BAYES	REGRESIÓN	XG BOOST	RANDOM FOREST
Accuracy	0.8729	0.9381	0.9455	0.9448
Precision	0.8809	0.9782	0.9619	0.9624
Recall	0.8669	0.8981	0.9294	0.9276



Instituto Tecnológico  
de Buenos Aires

**¡MUCHAS**  
**GRACIAS!**

MÁS INFORMACIÓN > [www.itba.edu.ar](http://www.itba.edu.ar)