



SOLUCION
**CASO
PRACTICO**

SANTIAGO MADARIAGA



CONTENIDOS

- 01 Análisis de los datos (EDA)
- 02 Feature Engineering
- 03 Modelos predictivos a usar
- 04 Resultados
- 05 Propuestas de mejora
- 06 Estrategias proactivas

INTRODUCCIÓN

Dataset crudo:

| | COD_CLIENTE | PRODUCTO | NUM_CERTIFICADO | FECHA_INICIO | FECHA_CANCELACION | FLAG_CANCELADO |
|---|---|----------|---|--------------|---------------------|----------------|
| 0 | 30bcb9290daee609d26c5780839c0a48e5faf39f65e1f2... | 1 | c940427a73c497b10f1f223ddb416f7403293120821172... | 2021-03-12 | 2021-03-18 00:00:00 | 1 |
| 1 | bad01b3b590275f7b1729f2012b2eb05a820f1217544d7... | 4 | 5d31273ad02e1f9626faed1ee3a9ac7dab01aef1d239bb... | 2021-07-23 | 2021-10-07 00:00:00 | 1 |
| 2 | e2350ee66445c6e35dcf05a42684e0625b479762e1fcf3... | 1 | c0663d4b54e3901f68d9b70d4e2f2b747831be2a23efc7... | 2021-09-27 | 2022-02-01 00:00:00 | 1 |
| 3 | 8dae800d252d2fb817ca498bb8cbdf4e77f46a02260e7... | 1 | 1c23156ed8e7de34160f513305d84d30ef00a73e044629... | 2022-05-09 | - | 0 |
| 4 | 9a0f0371a3754818a5dc983ab96bc799e25743821d4bdf... | 1 | 2caa655347e3584e3c977f6ee2fd0834e43213272df493... | 2021-09-13 | 2022-12-12 00:00:00 | 1 |

Calidad de los datos:

```
Certificados unicos: 118952 | total non-null values: 120000
Clientes unicos:116188
Clientes duplicados: 3812
```

```
COD_CLIENTE      0
PRODUCTO         0
NUM_CERTIFICADO  0
FECHA_INICIO     0
FECHA_CANCELACION 0
FLAG_CANCELADO   0
dtype: int64
```

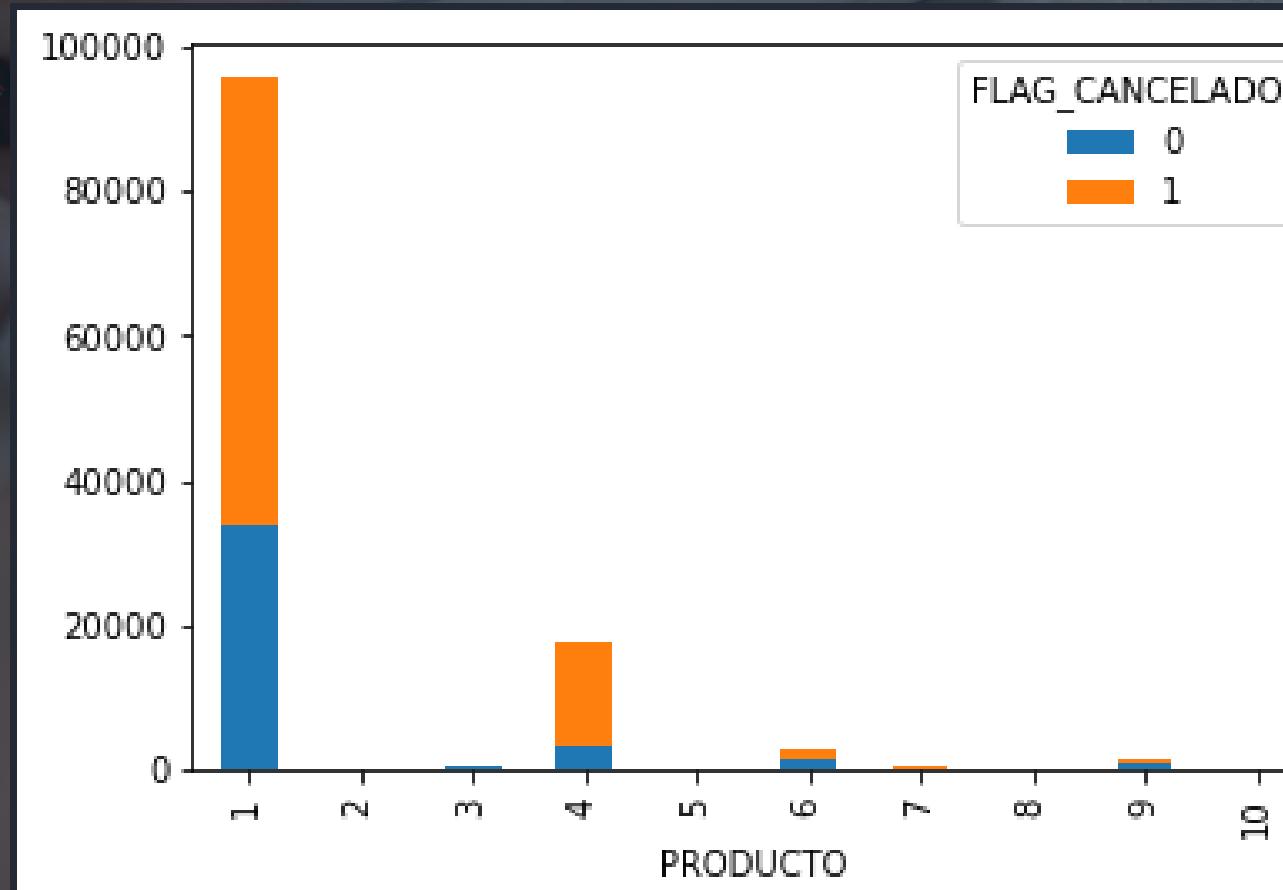
```
count           118952
unique          649
top            2021-04-30 00:00:00
freq            410
first          2021-01-02 00:00:00
last           2022-10-31 00:00:00
Name: FECHA_INICIO, dtype: object
```

No hay valores faltantes (NAs)

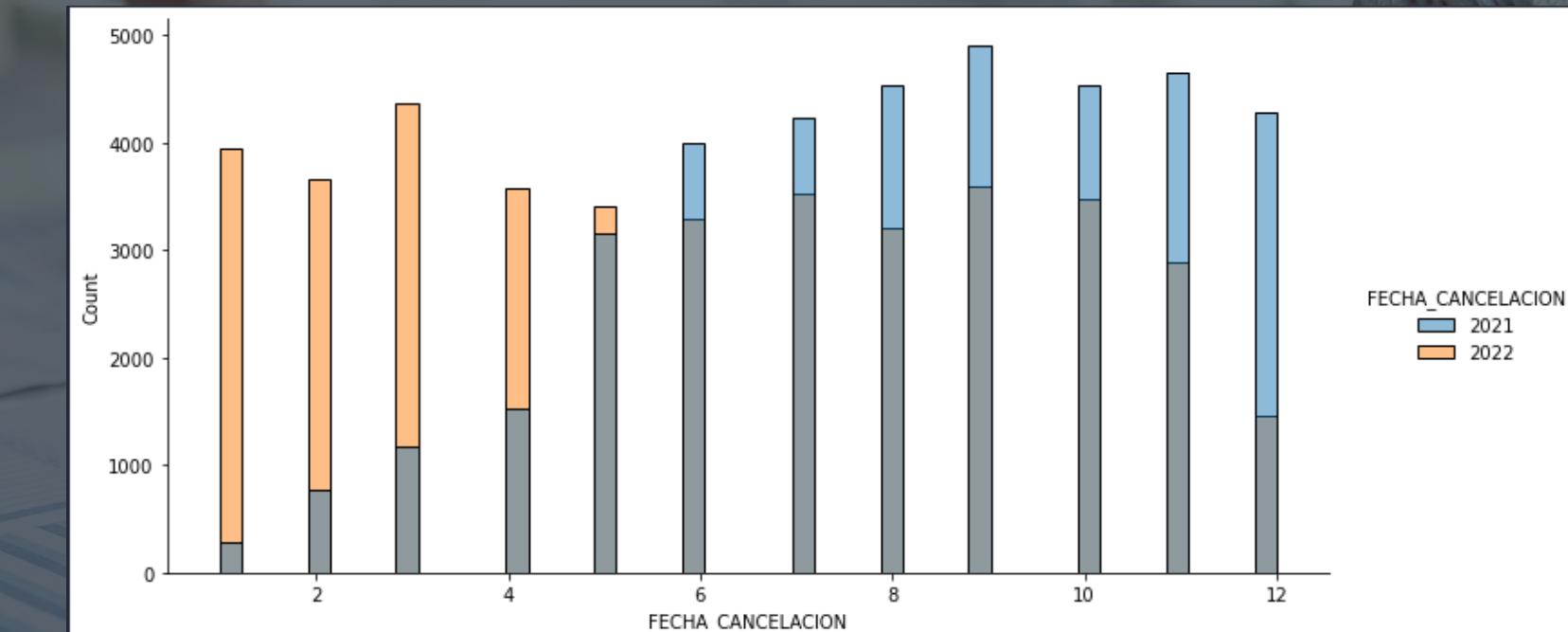
Los datos pertenecen al 2021 y el 2022

EXPLICABILIDAD DE CANCELACIONES

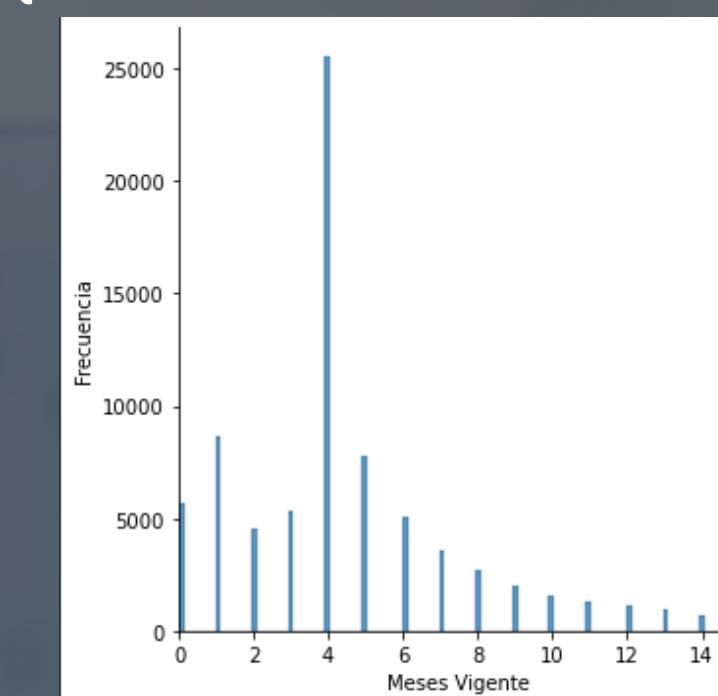
Seguros cancelados por tipo de producto



Seguros cancelados por mes y año



El 63% de los que cancelan lo hacen en los primeros 4 meses

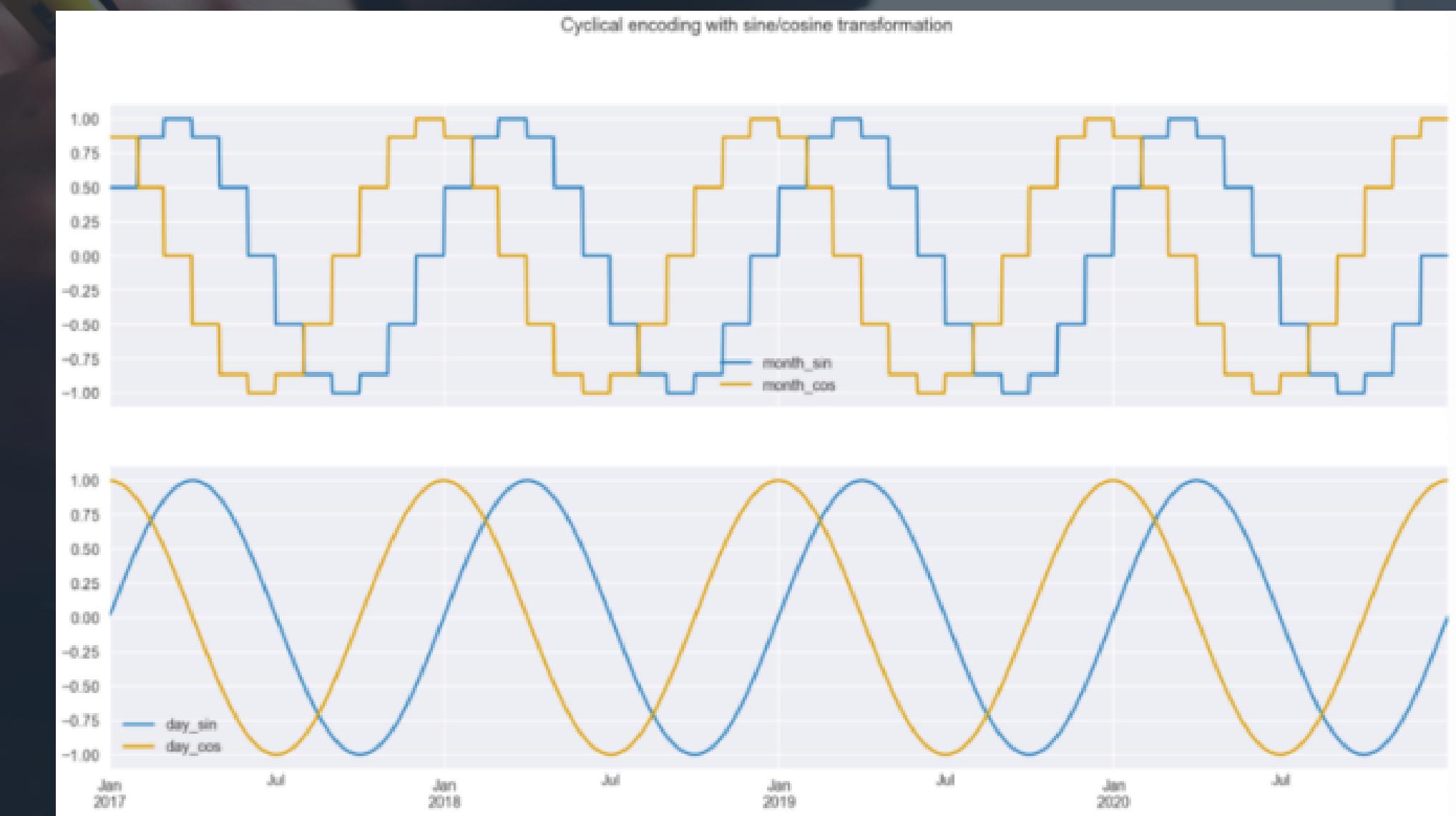


FEATURE ENGINEERING

- Crear una feature que describa si se contrato el servicio en un dia cercano a una festividad pagada (navidad, fiestas patrias).

Encoding de variables categóricas:

El tiempo se tratará como una función periódica de periodo 365 (1 año)



Encoding de variables categóricas:

El tipo de producto no contiene ningun significado numérico. Se trata de una variable categórica ordinal

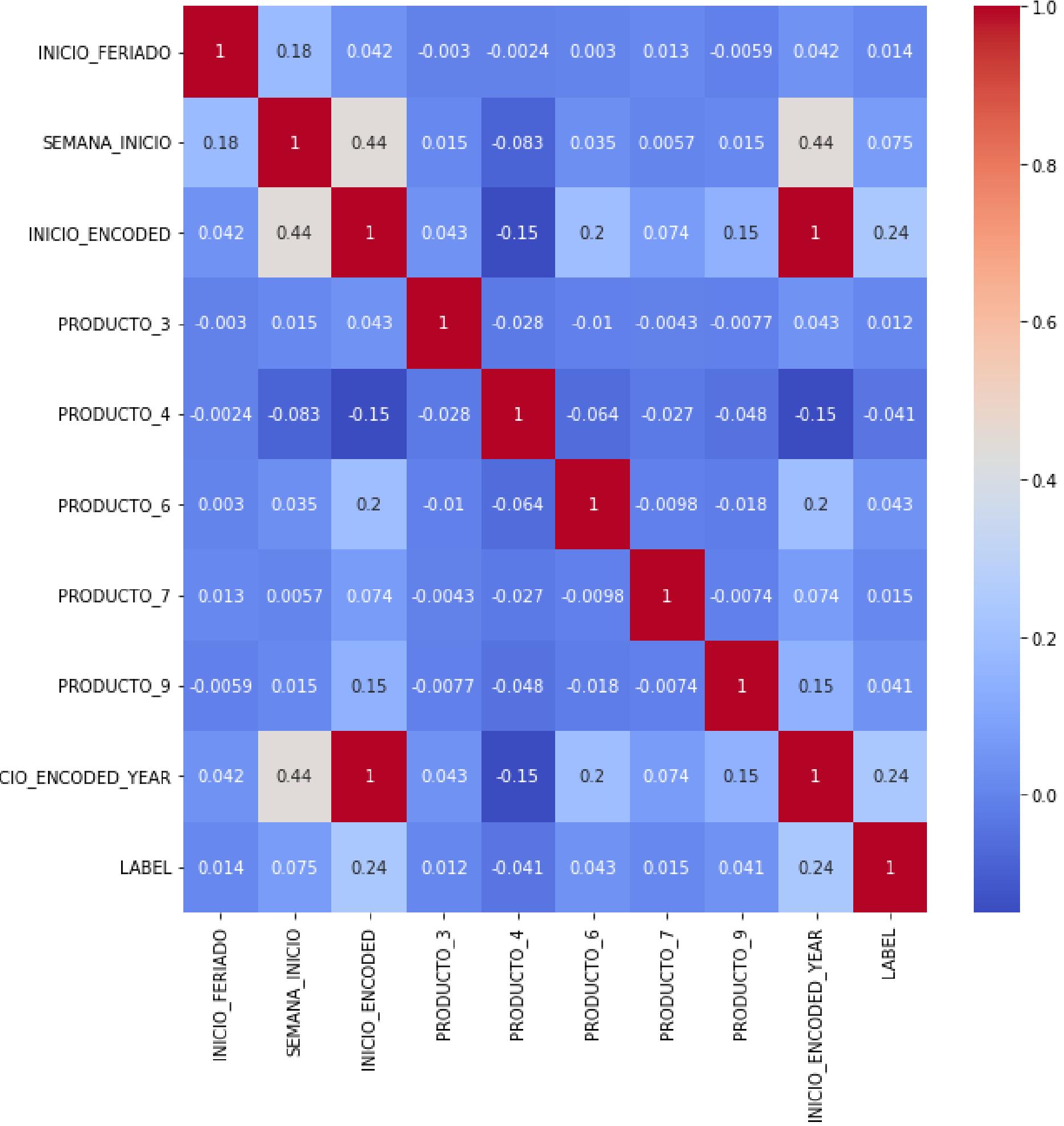
```
# One hot encoding Producto  
df = pd.get_dummies(df, columns=['PRODUCTO'], prefix='PRODUCTO')
```

Creacion del LABEL

Las clases del target se dividen en 2:

- Se cancela la póliza dentro de los primeros 4 meses de ser contratada (**LABEL = 1**)
- Se cancela la póliza fuera de este período o no se cancela (**LABEL = 0**)

AUTOCORRELACION DE LAS FEATURES



SELECCION DEL MODELO

LOGISTIC REGRESSION

- Es fácil de entrenar
- Tiene poca varianza
- No requiere de tuneo de hyperparámetros

DEEP LEARNING

GB DECISION TREES

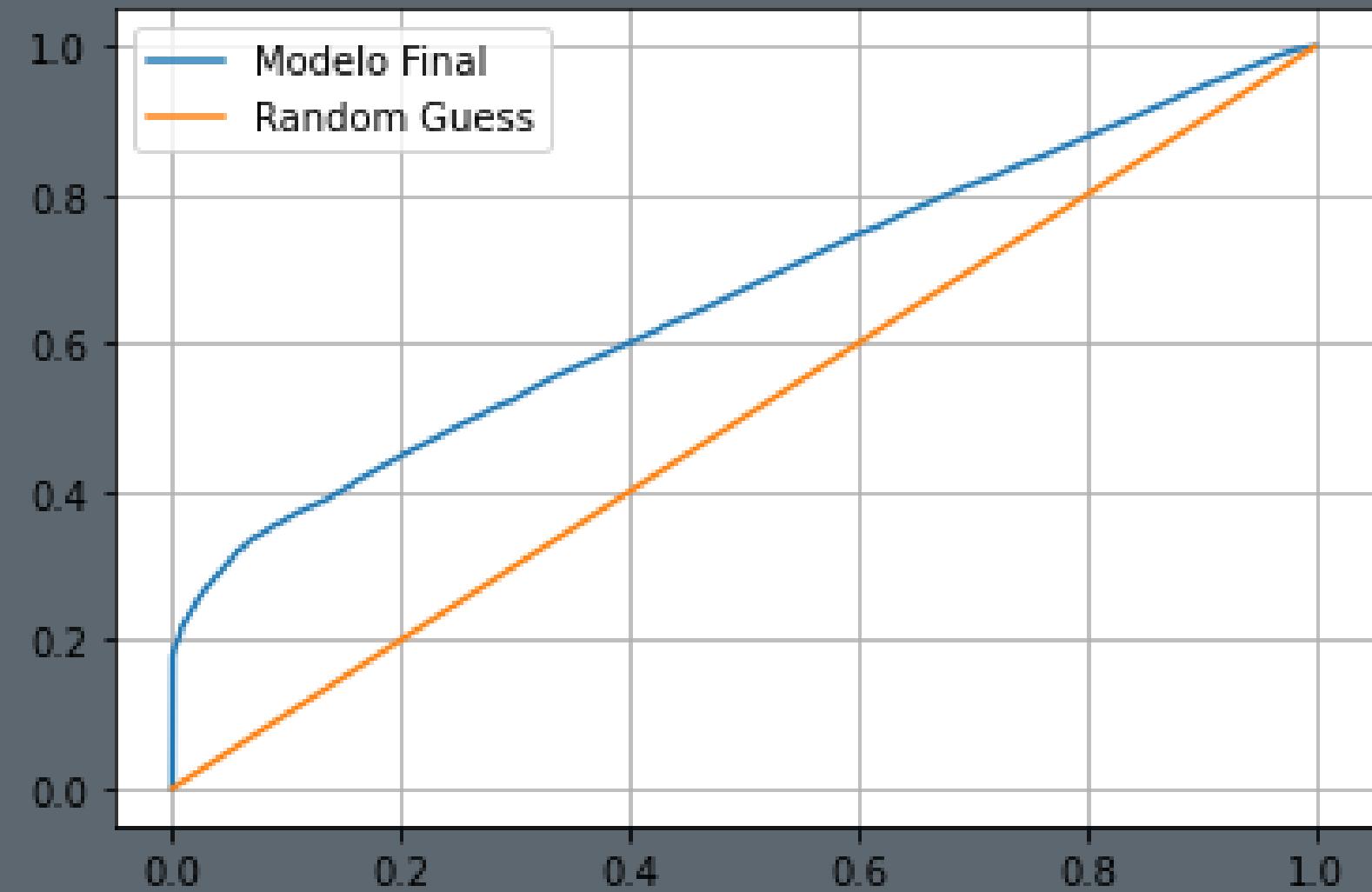
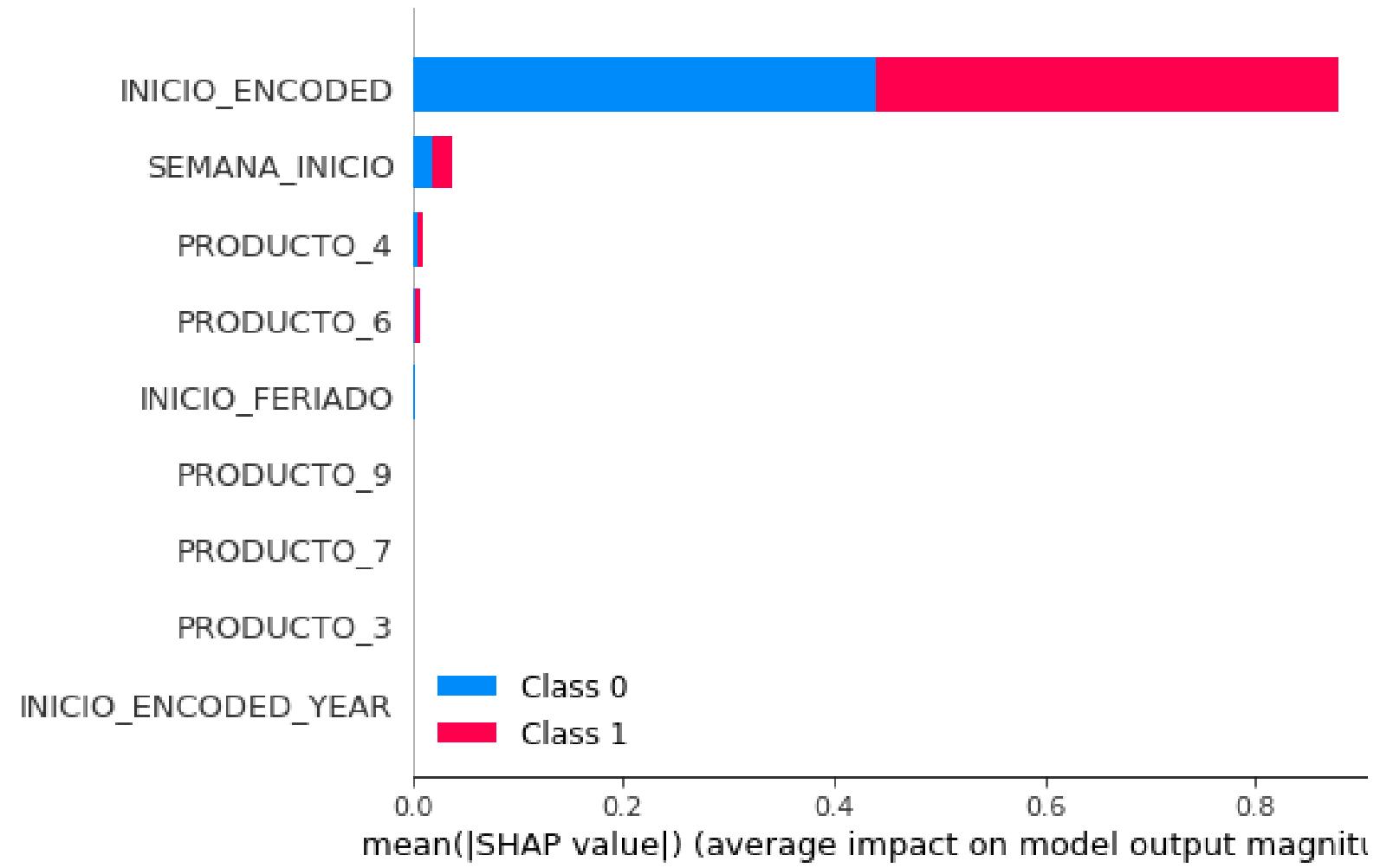
- Ofrecen más flexibilidad por tener más parámetros
- Entregan muy buenos resultados en contraste con su tiempo de entrenamiento
- Son escalables y elásticos

XGB
CATBOOST
LIGHTGBM

RESULTADOS

ROC

SHAP values



04

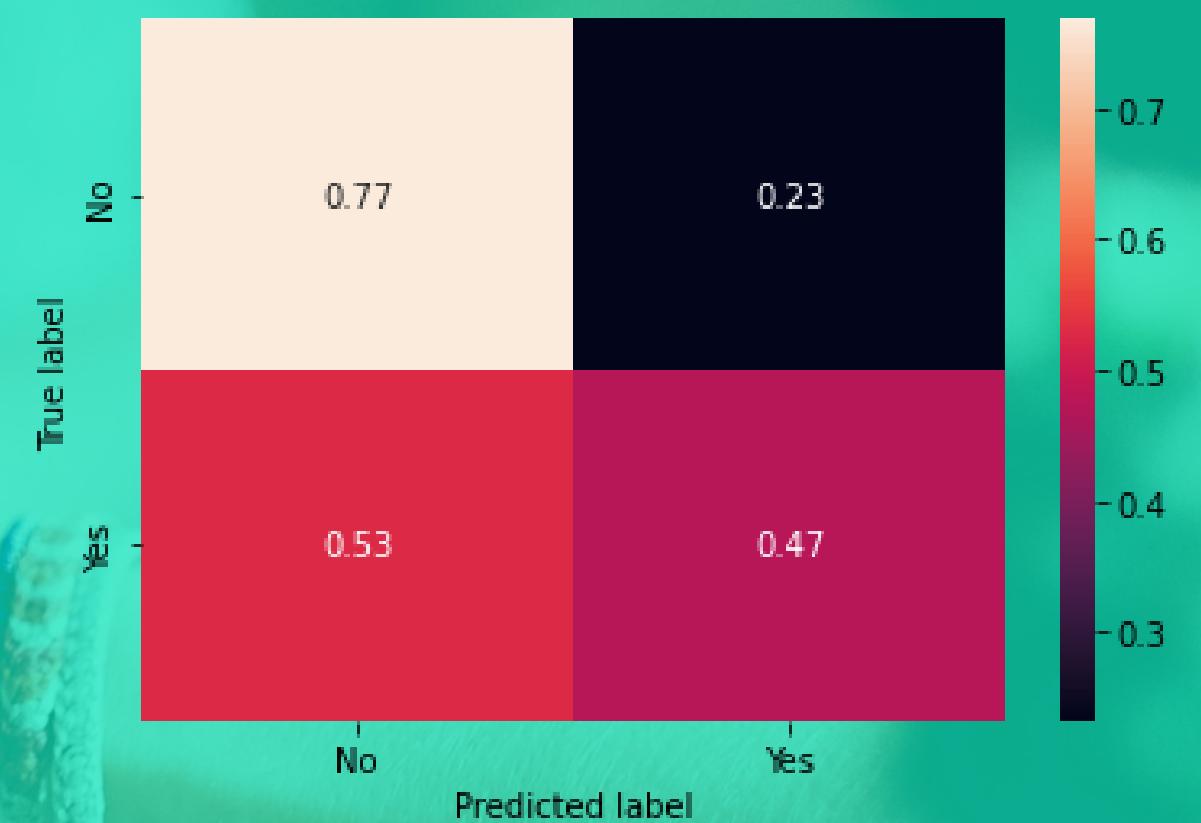
*Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut a
enim nec nisl ullamcorper eleifend.
Praesent risus leo.
Fringilla et nulla at, egestas euismod orci.*

MARGEN DE DECISION

Menor Threshold



Mayor Threshold



Más cobertura de casos positivos

Más precisión

ESTRATEGIA PROACTIVA

Valor de probabilidad entregado por el modelo

**Retroalimentación del cliente
(de ser el caso)**

**Factores exógenos
(riesgo país, recesión, etc.)**



Consultar al cliente si cambiaron sus necesidades

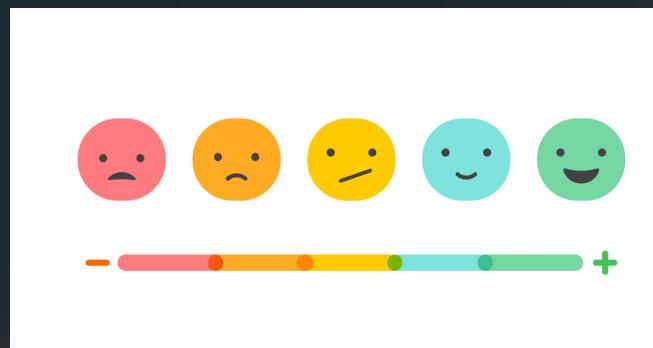
Mencionar la importancia del seguro

Resaltar las ventajas del seguro en contraste con la competencia

PROPUESTAS DE MEJORA

Hacia el cliente

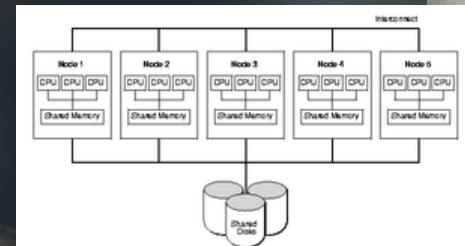
- Encuesta de satisfacción
- Monitoreo continuo



Modelo ensamblado (modelo de segmentación + forecaster + clasificador binario)

Hacia el modelo

- Recopilar más información en el dataset
- Reentrenar el modelo con mayor profundidad en un cluster





ENSIGNA
GRACIAS