



ANÁLISIS DE RIESGO CREDITICIO

UTILIZANDO TÉCNICAS DE INTELIGENCIA

ARTIFICIAL

JUAN SEBASTIÁN OTERO VEGA - 2220053

FARID CAMILO ROJAS VARGAS - 2220051

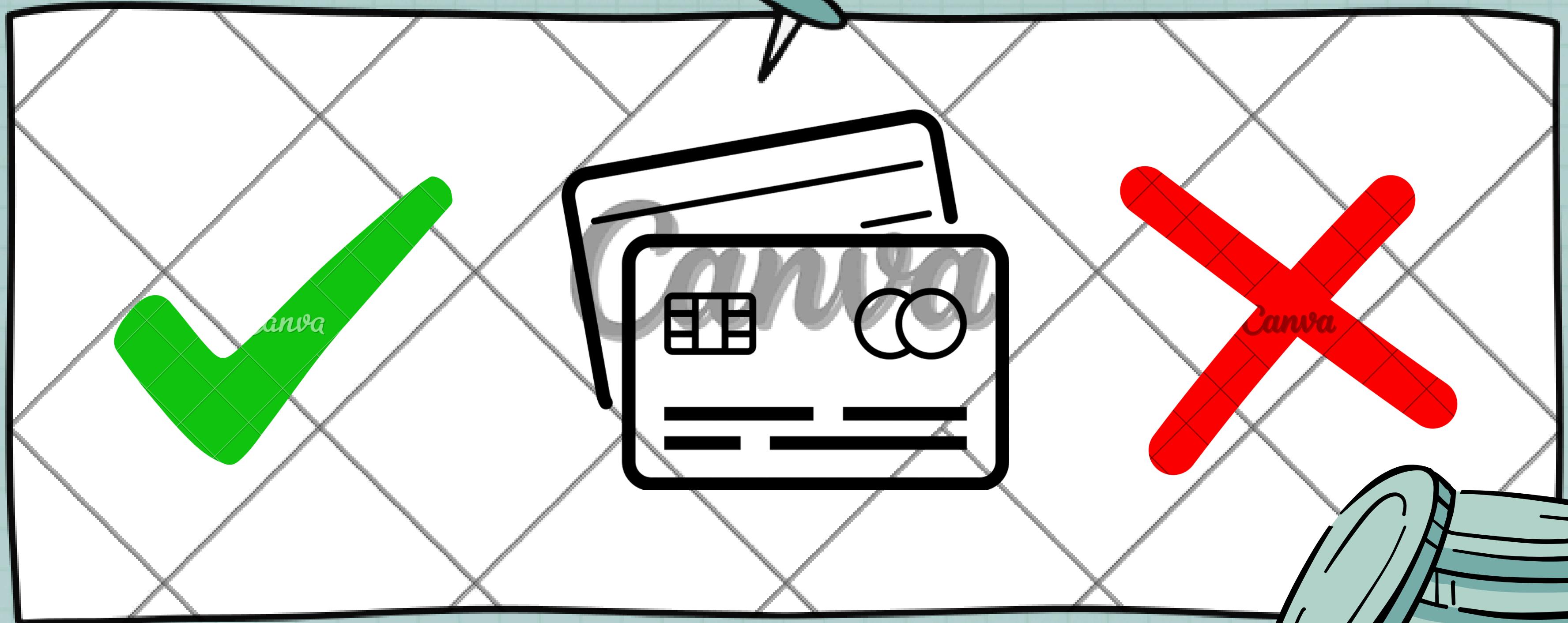
JUAN MANUEL RAMÍREZ SALAMANCA - 2192279

CONTENIDO:

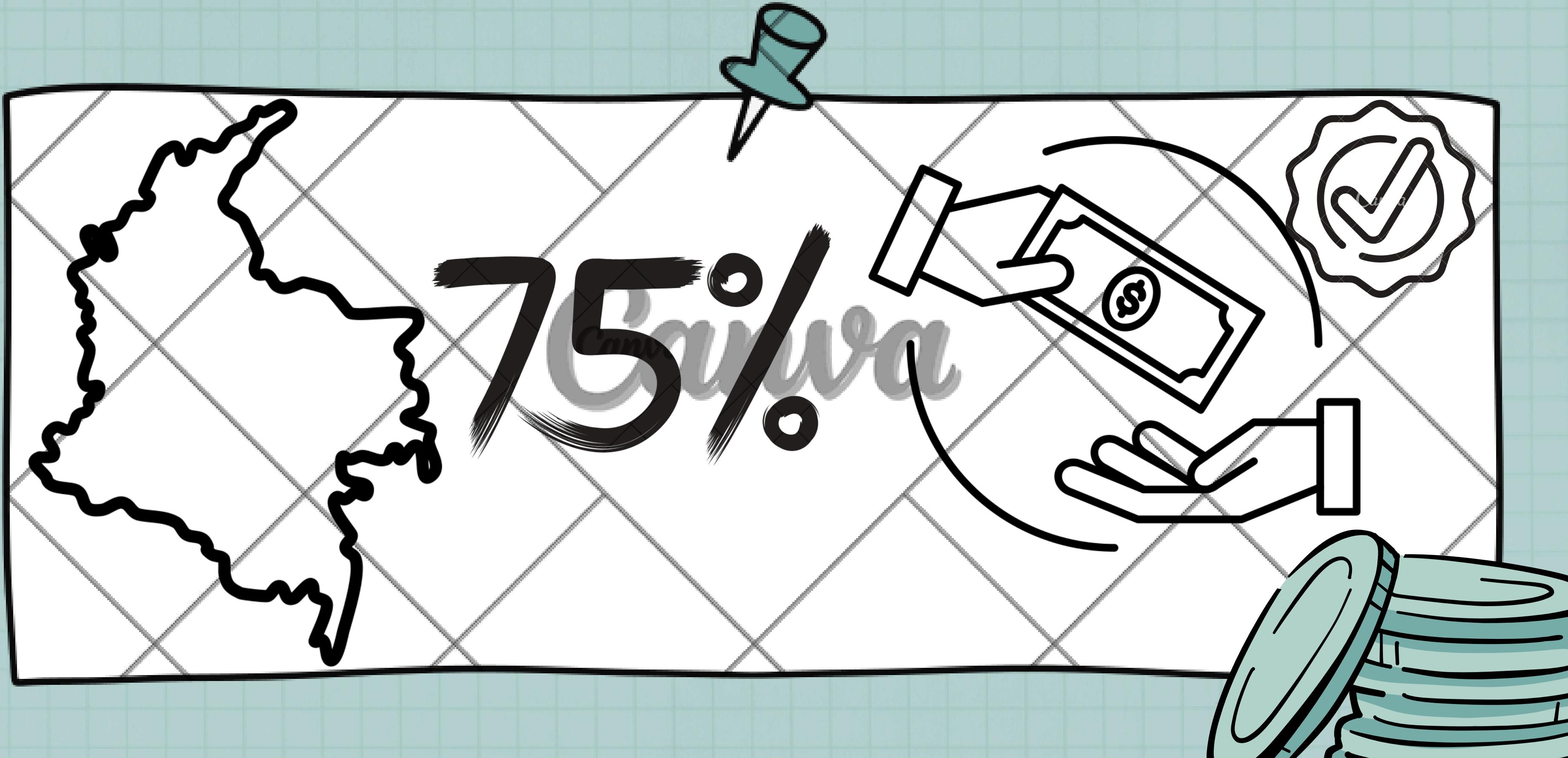
- Presentación del proyecto
- Información del dataset
- Pre-Procesamiento
- Estadística
- Aprendizaje supervisado
- Aprendizaje no supervisado
- Comparación de metricas



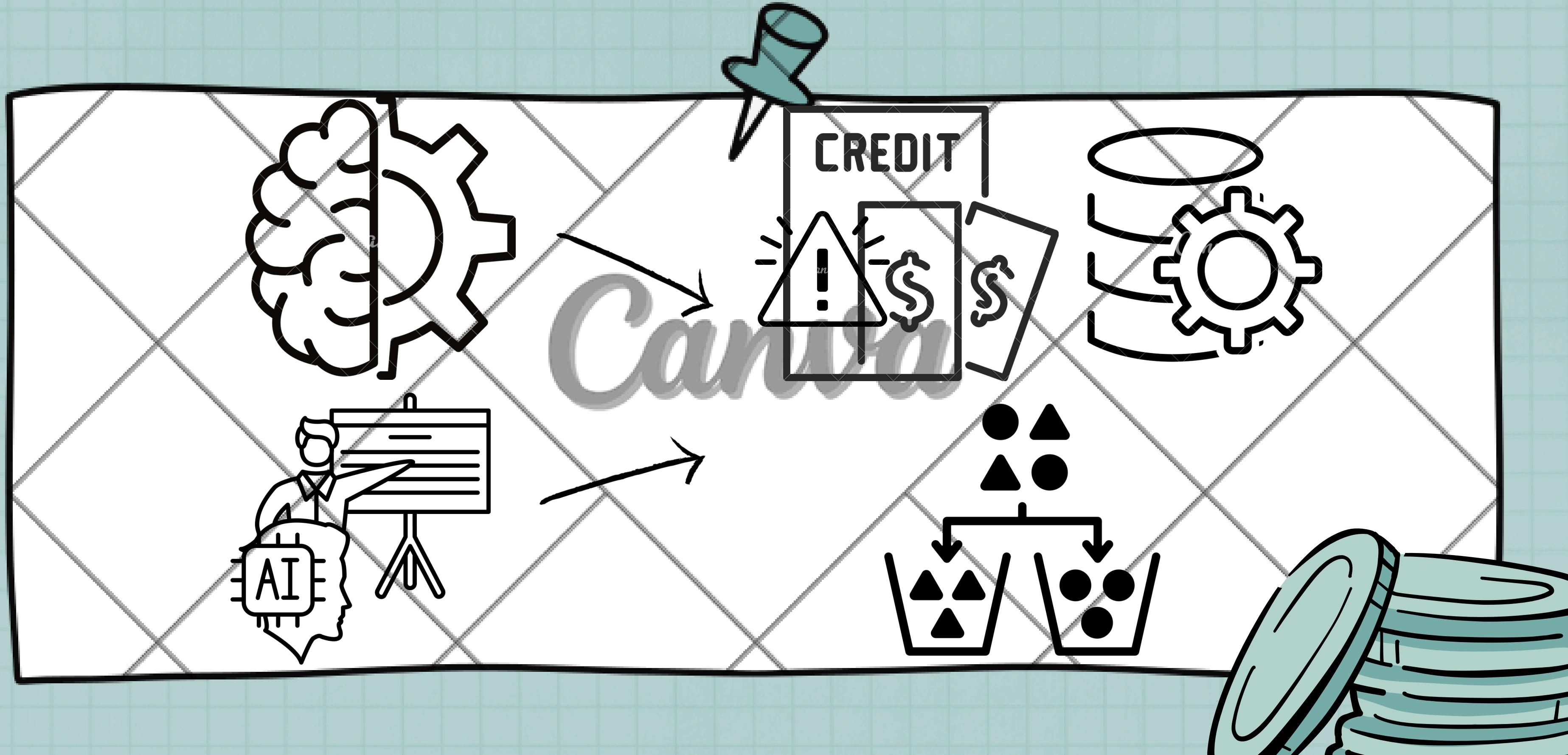
PRESENTACIÓN DEL PROYECTO



MOTIVACIÓN



OBJETIVO



INFORMACIÓN DEL DATASET

Clasifica a personas como de buen o mal riesgo crediticio basado en sus características personales y financieras.

- **Nombre del dataset :** German Credit Data
- **Fuente:** UCI Machine Learning Repository (OpenML ID: 31)
- **Autor:** Dr. Hans Hofmann
- **Año:** 1994
- **DOI:** 10.24432/C5NC77

Hofmann, H. (1994). Statlog (German Credit Data) [Dataset]. UCI Machine Learning Repository. <https://doi.org/10.24432/C5NC77>.

INFORMACIÓN DEL DATASET

MATRIZ DE
COSTO

	Malo (Predicho)	Bueno (Predicho)
Malo (Actual)	0	5
Bueno (Actual)	1	0

Clasificar incorrectamente a un cliente malo como bueno tiene un costo mayor, cuando es malo (costo 5), que clasificar a un cliente como malo cuando es bueno (costo 1).

INFORMACIÓN DEL DATASET

#Características Numéricas

duration: Meses de duración del crédito.

credit_amount: Monto total del crédito solicitado.

installment_commitment: Porcentaje de ingreso dedicado al crédito.

residence_since: Años viviendo en la residencia actual.

age: Edad del cliente.

existing_credits: Número de créditos activos en otras instituciones.

num_dependents: Número de dependientes económicos.

INFORMACIÓN DEL DATASET

#Características Categóricas

checking_status – Estado de cuenta corriente.
credit_history – Historial crediticio.
purpose – Propósito del préstamo.
savings_status – Estado de cuenta de ahorros.
employment – Años de empleo.
personal_status – Estado civil y género.
other_parties – Otros garantes o fiadores.
property_magnitude – Tipo de propiedad o activo.
other_payment_plans – Otros planes de pago existentes.
housing – Tipo de vivienda.
job – Tipo de empleo.
own_telephone – Si tiene teléfono propio.
foreign_worker – Si es trabajador extranjero.
class – (Variable objetivo): bueno / malo.

PRE-PROCESAMIENTO

Análisis inicial: Se revisaron estadísticas, valores nulos y distribución de la variable objetivo (class).

Codificación de la variable objetivo: class fue transformada a valores numéricos: good → 1, bad → 0.

Separación de variables: X: características, y: variable objetivo.

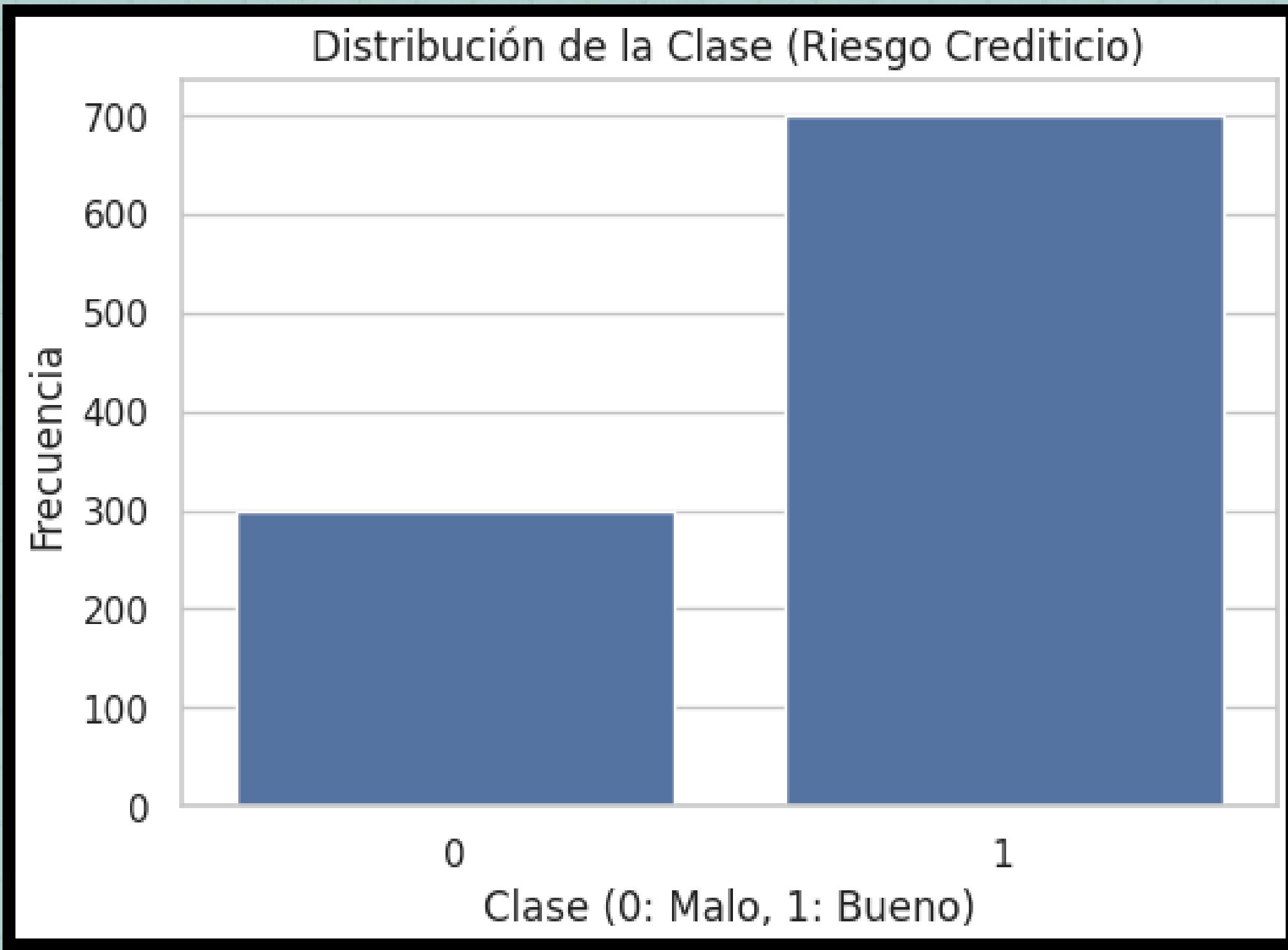
Identificación de tipos de variables: Se separaron en numéricas y categóricas.

Transformaciones: Numéricas → StandardScaler.
Categóricas → OneHotEncoder (con drop='first').

Aplicación del preprocessamiento: Se transformó X y se creó un nuevo DataFrame con las variables procesadas.

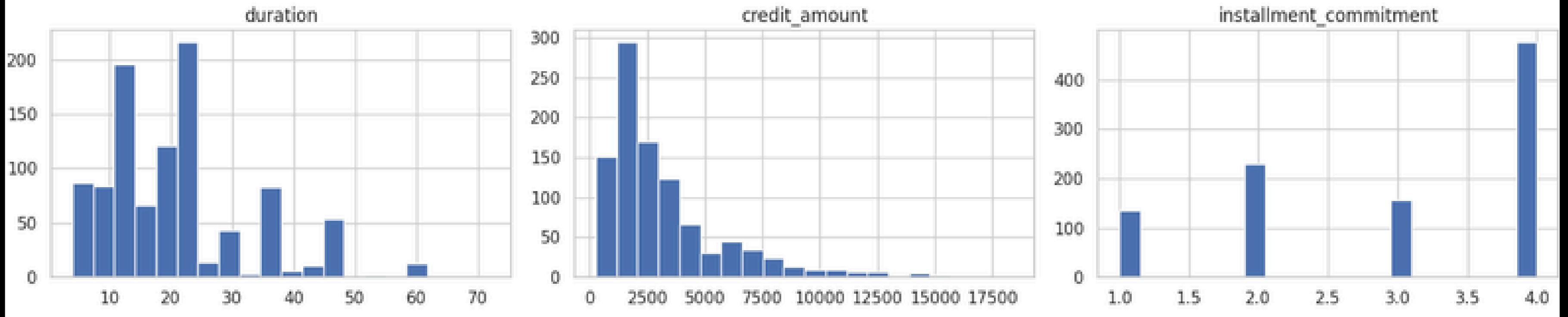
División del dataset: Con train_test_split (80% entrenamiento, 20% prueba, con estratificación).

ESTADÍSTICA

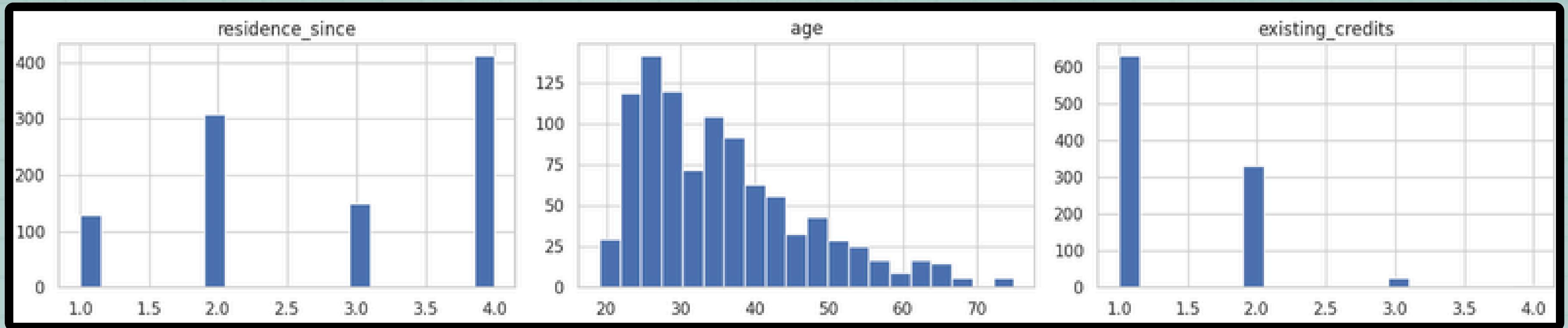


ESTADÍSTICA

Histogramas de Características Numéricas Originales

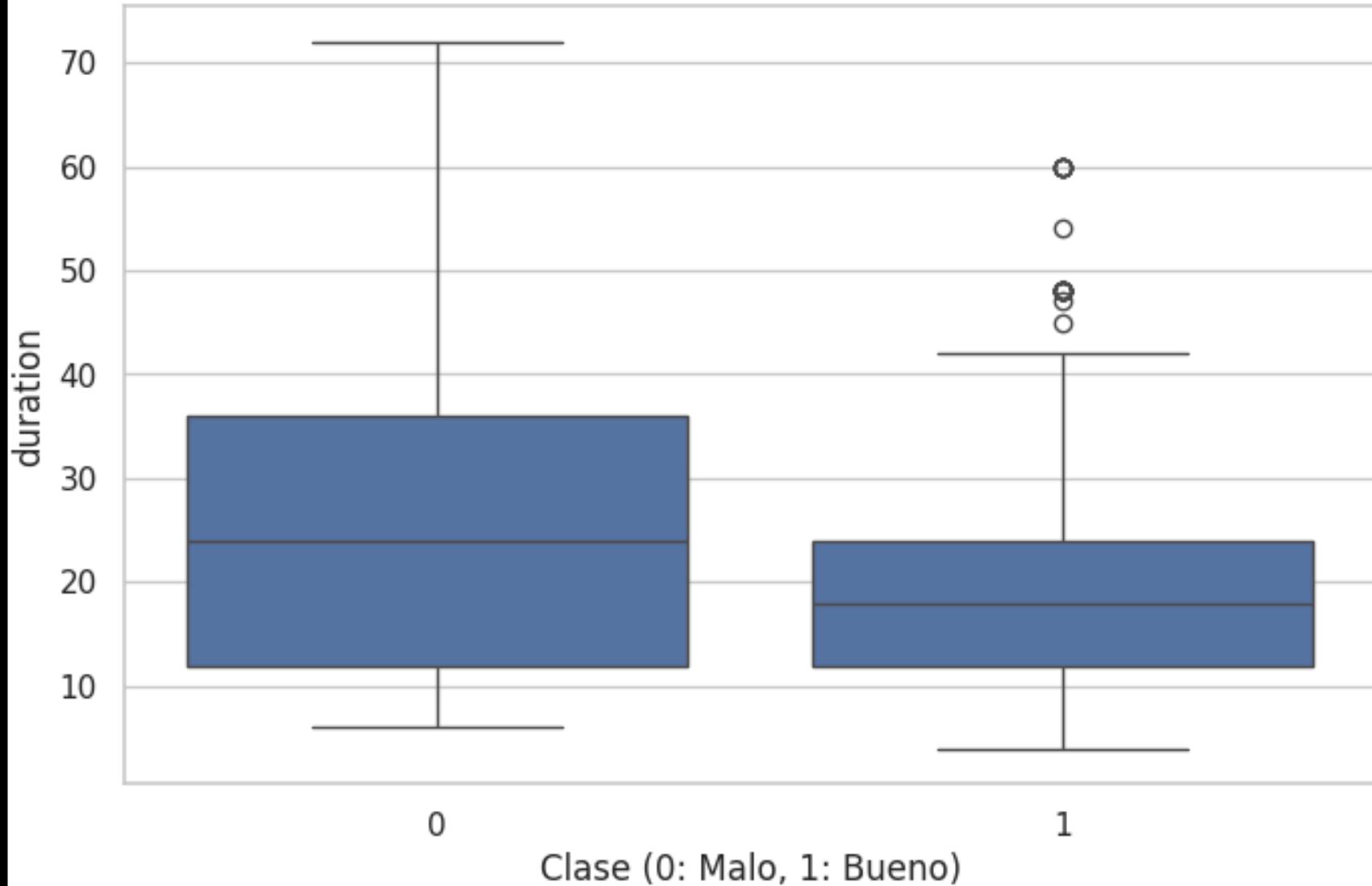


ESTADÍSTICA

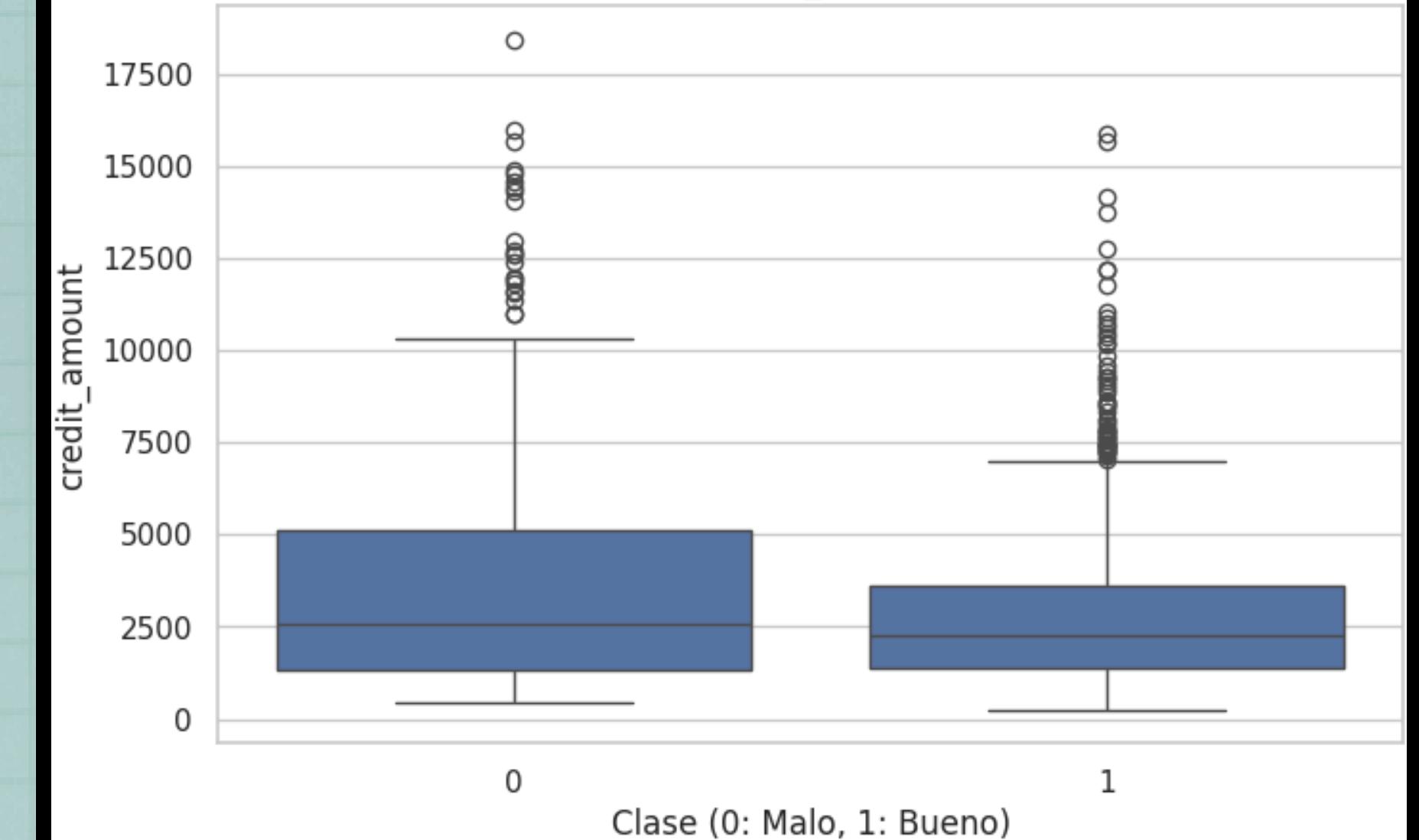


ESTADÍSTICA

Boxplot de duration vs Clase

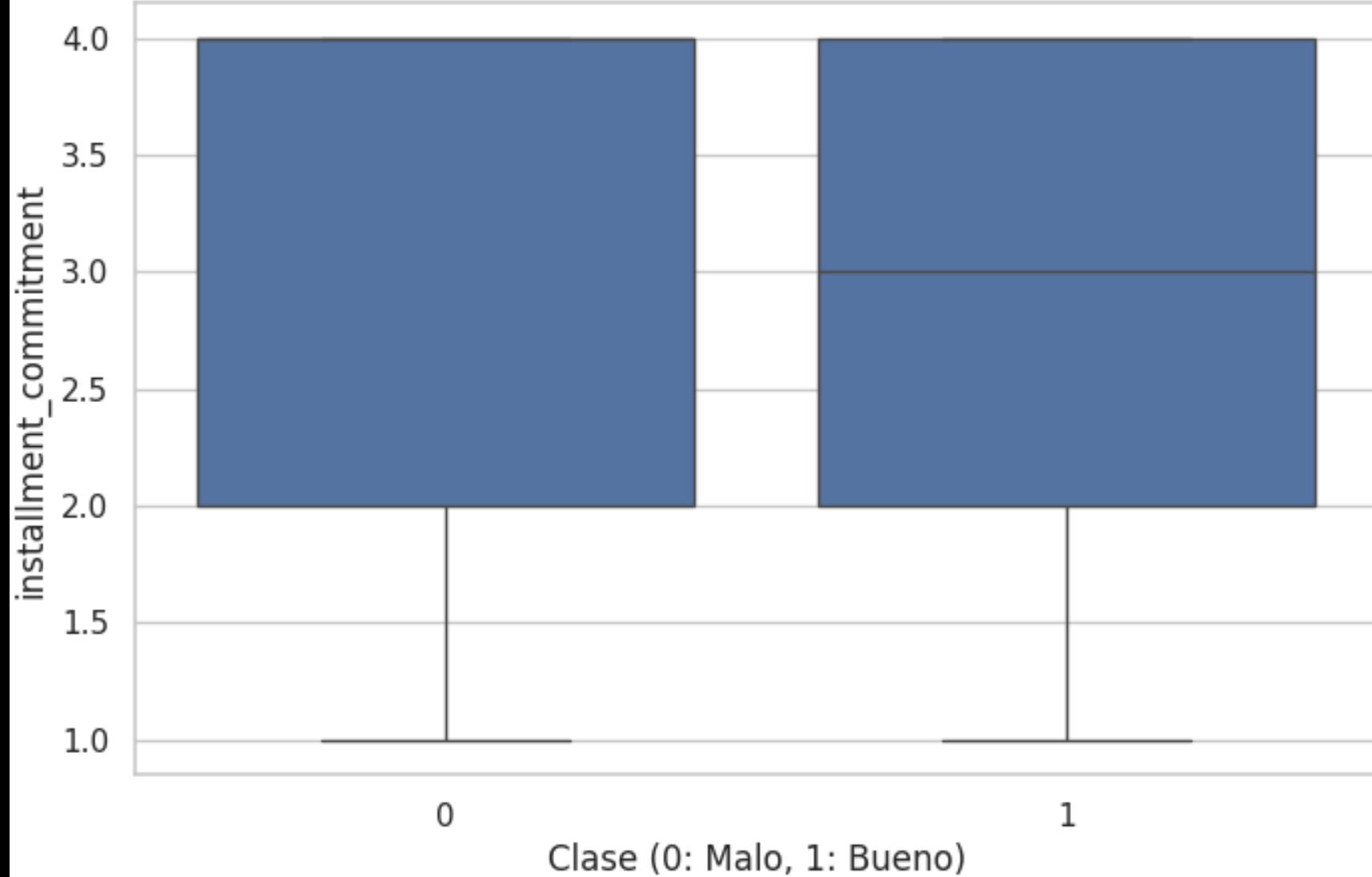


Boxplot de credit_amount vs Clase

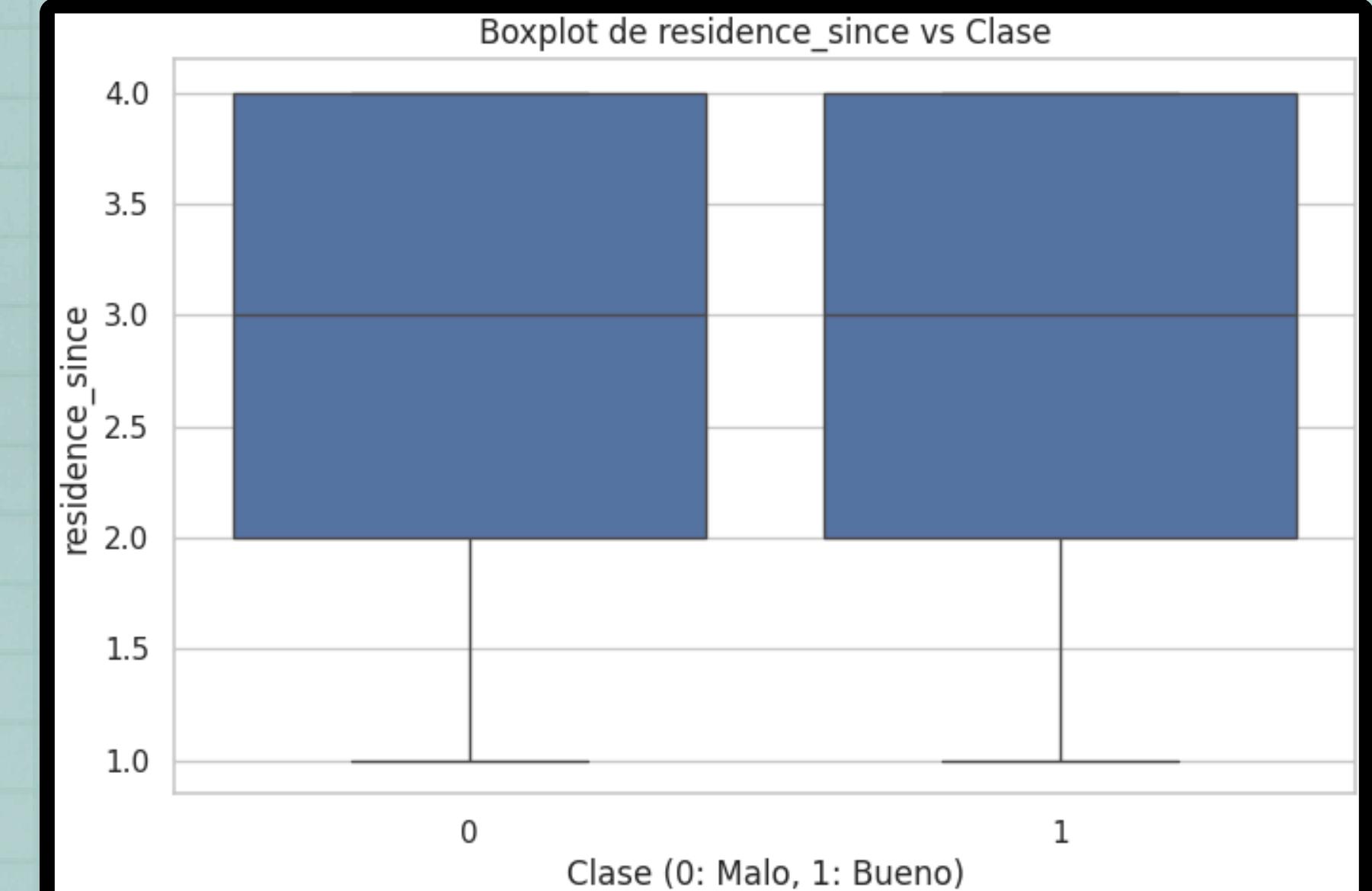


ESTADÍSTICA

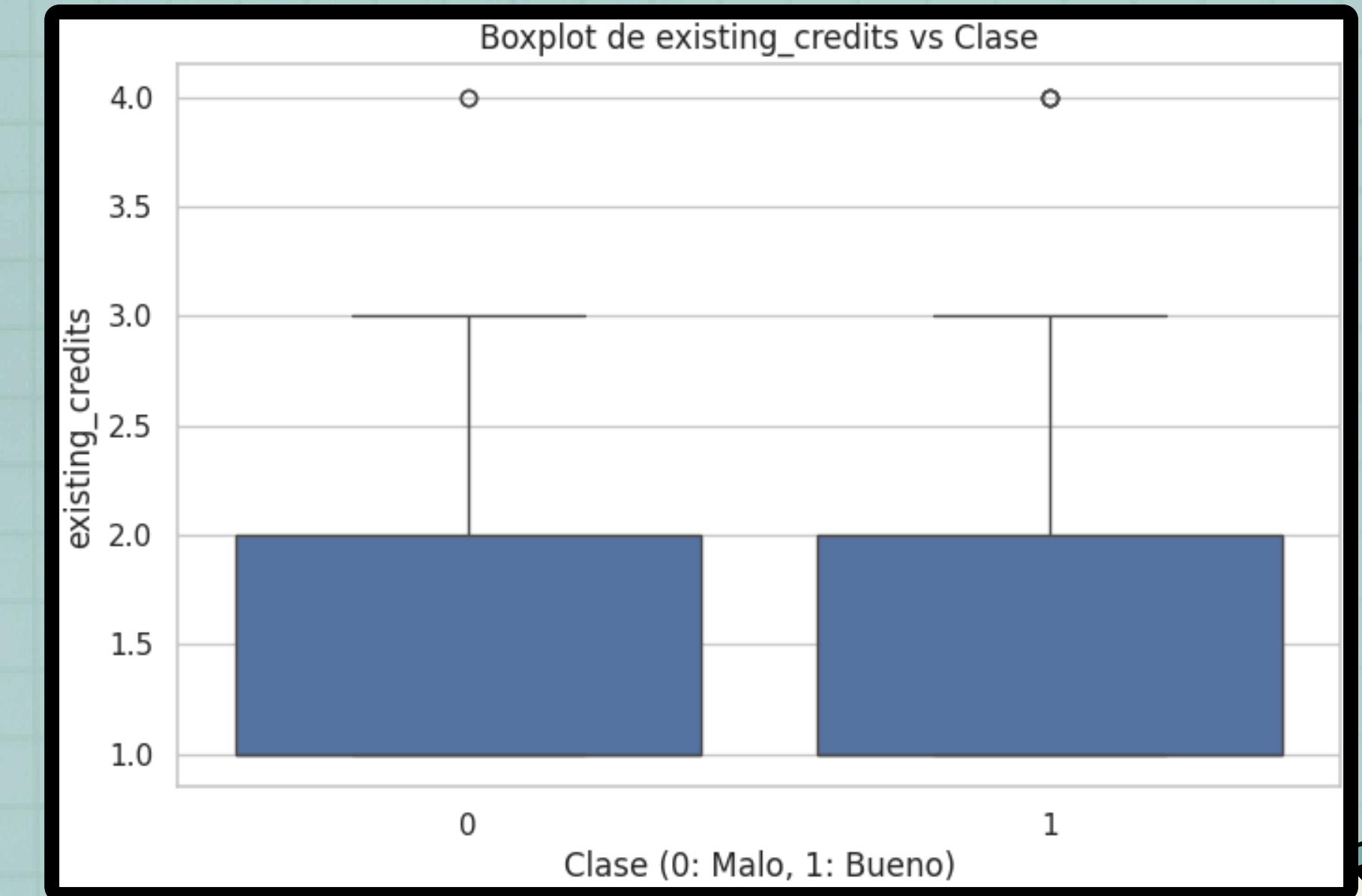
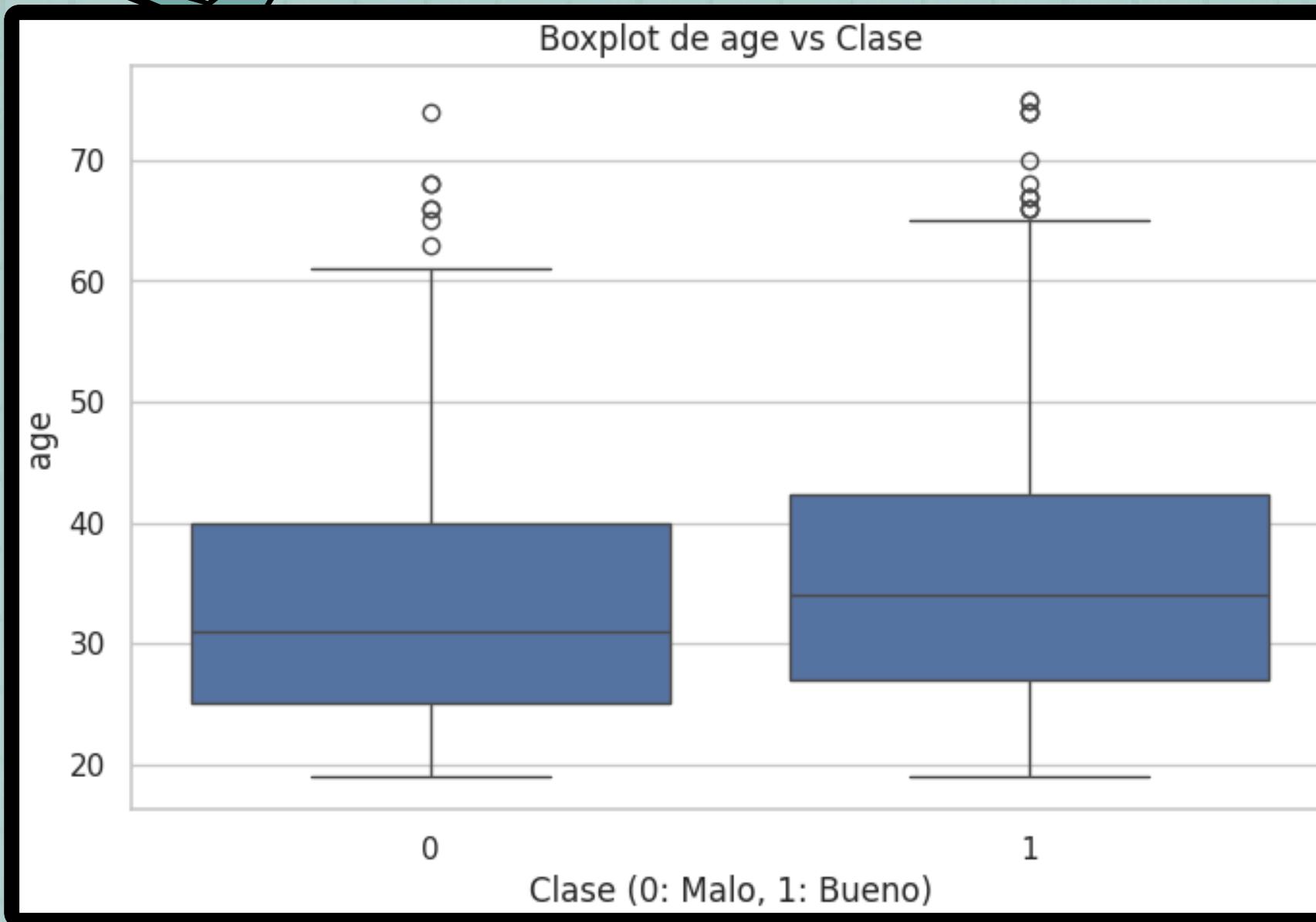
Boxplot de installment_commitment vs Clase



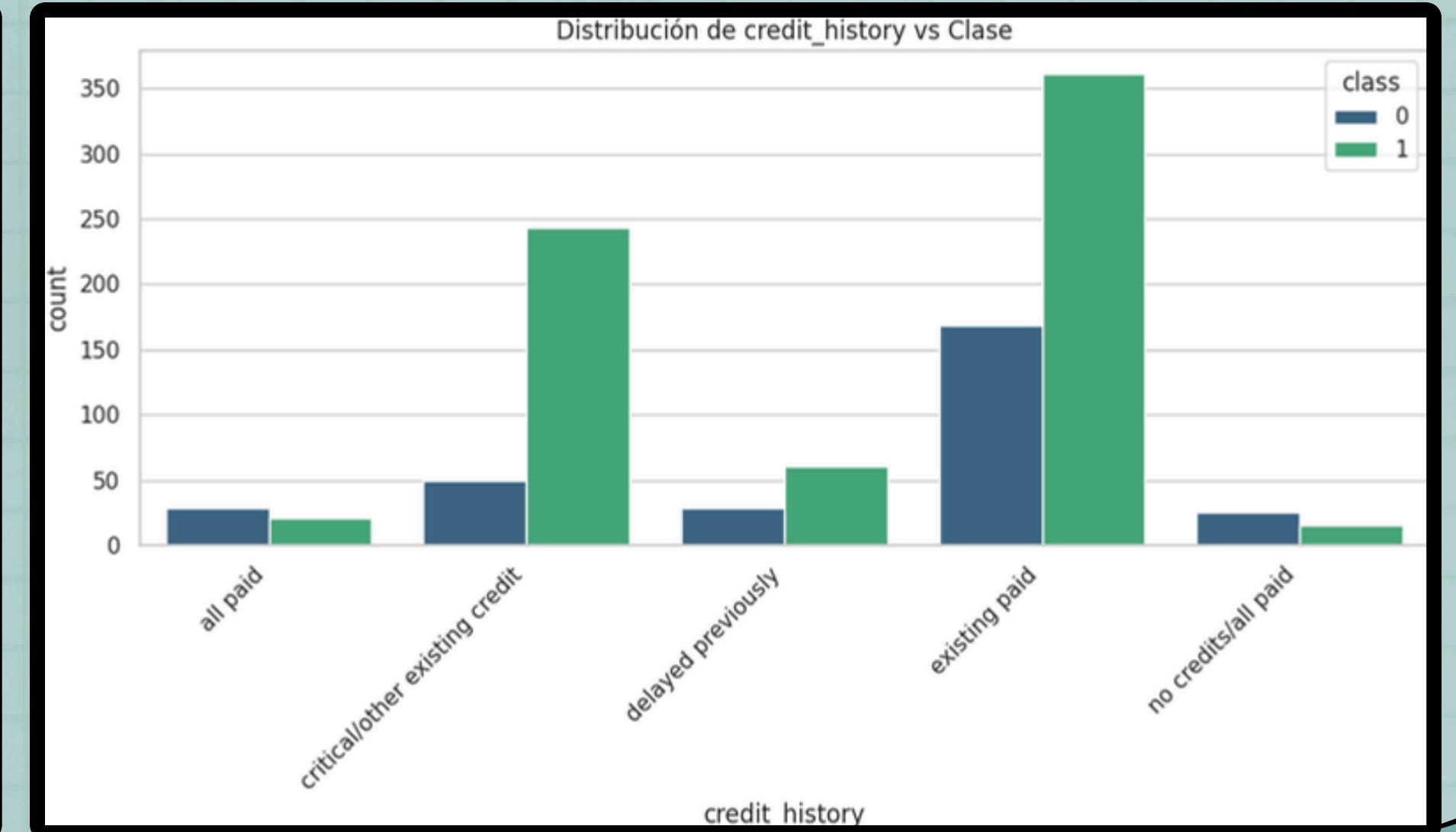
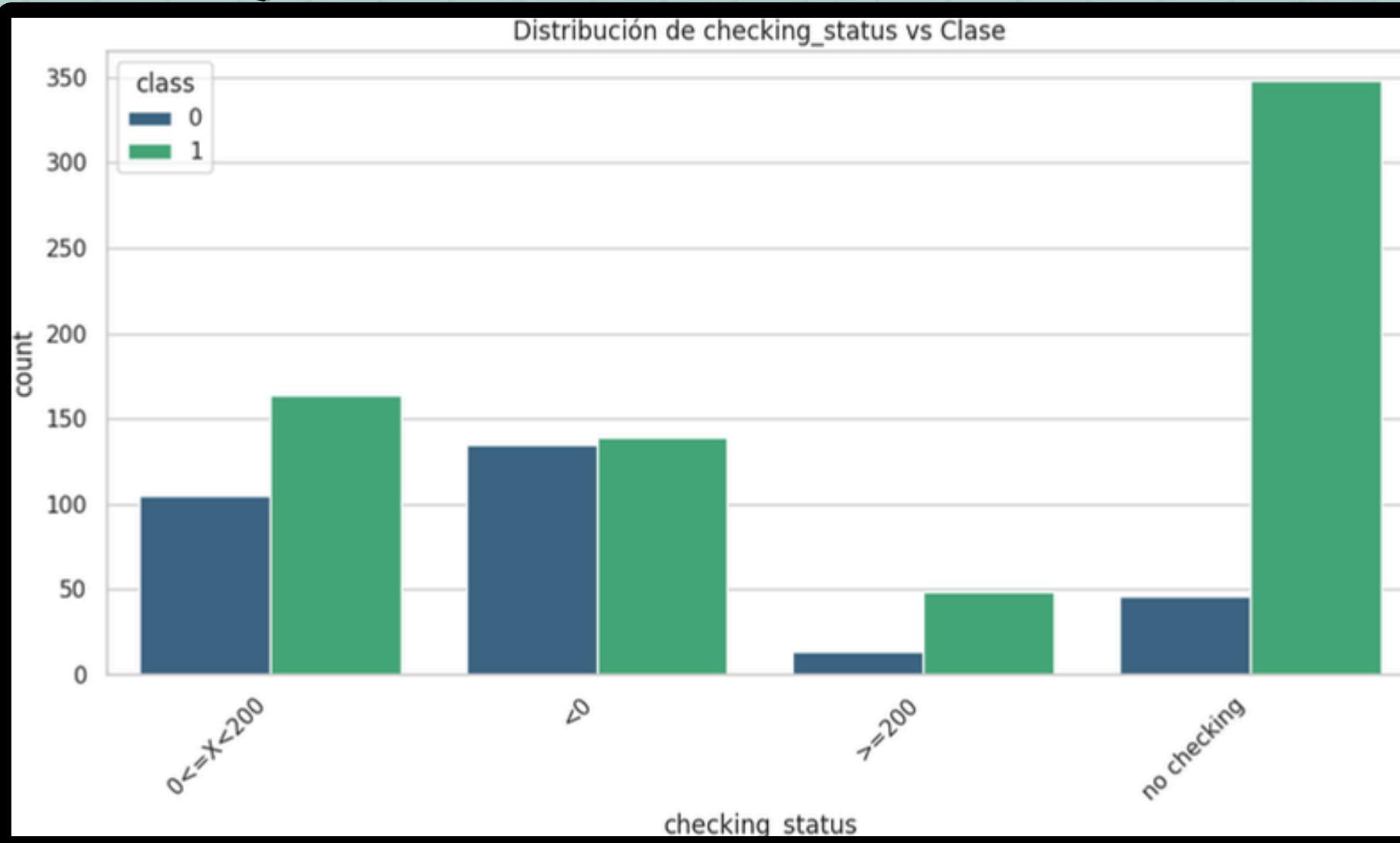
Boxplot de residence_since vs Clase



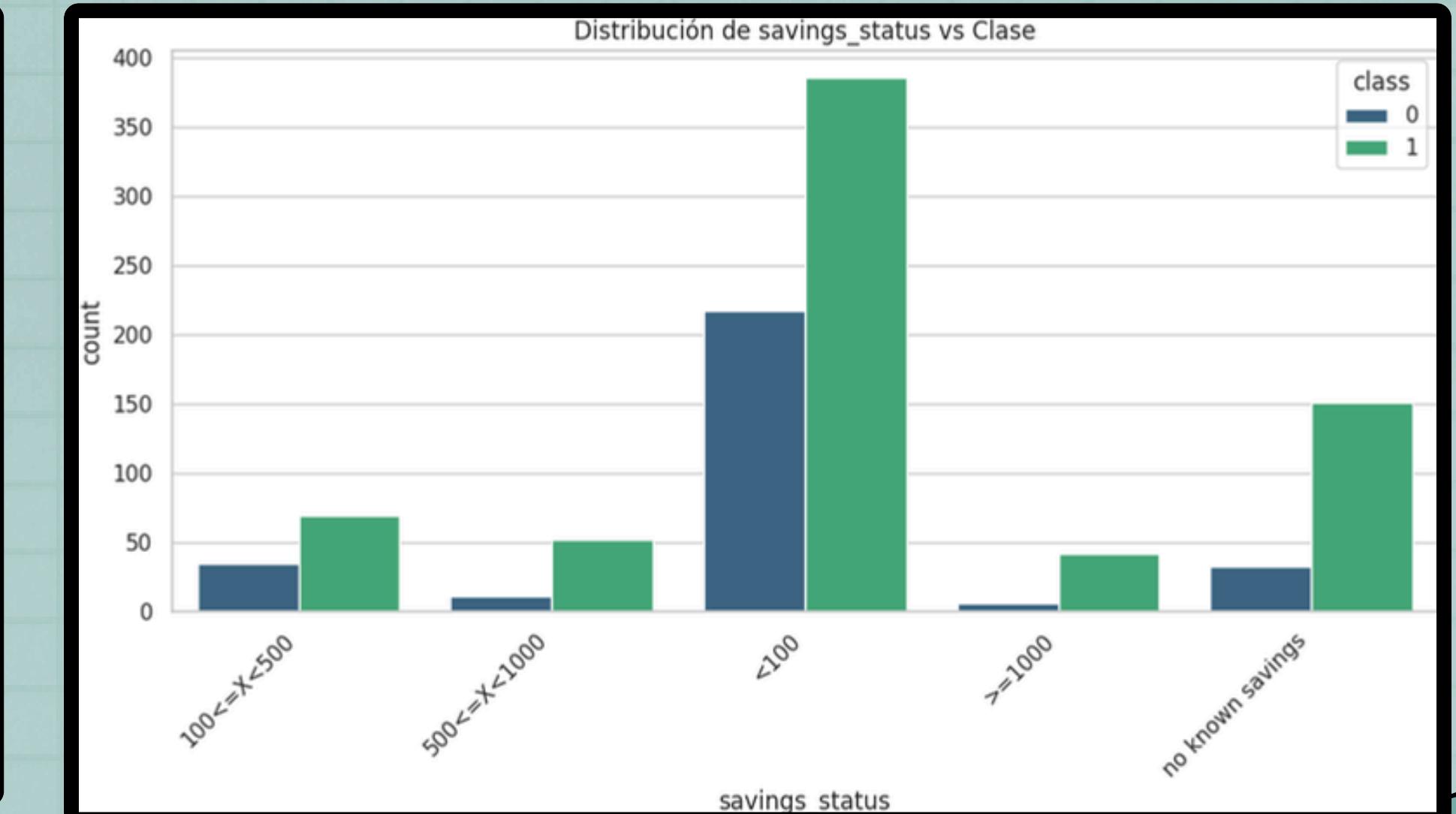
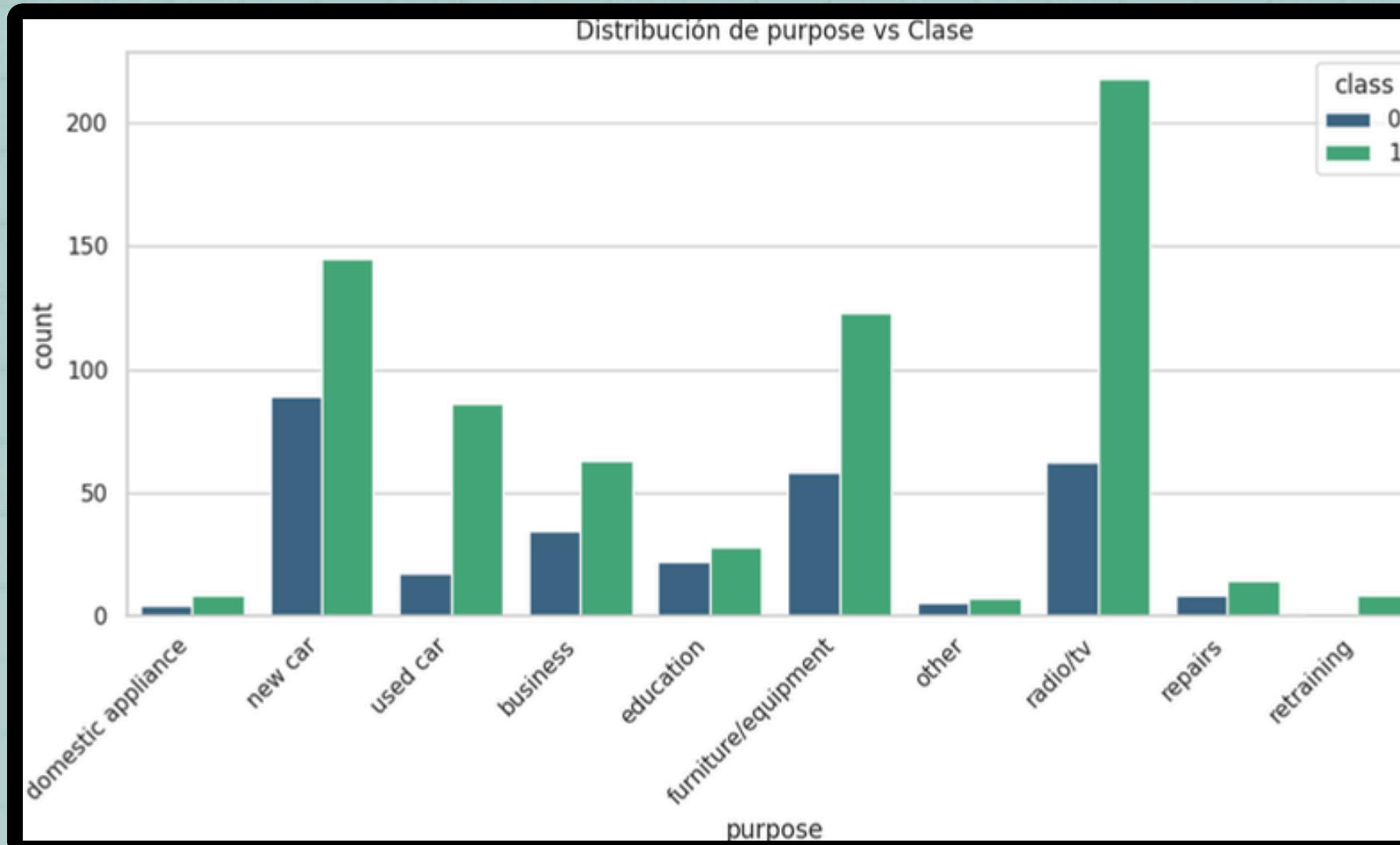
ESTADÍSTICA



ESTADÍSTICA

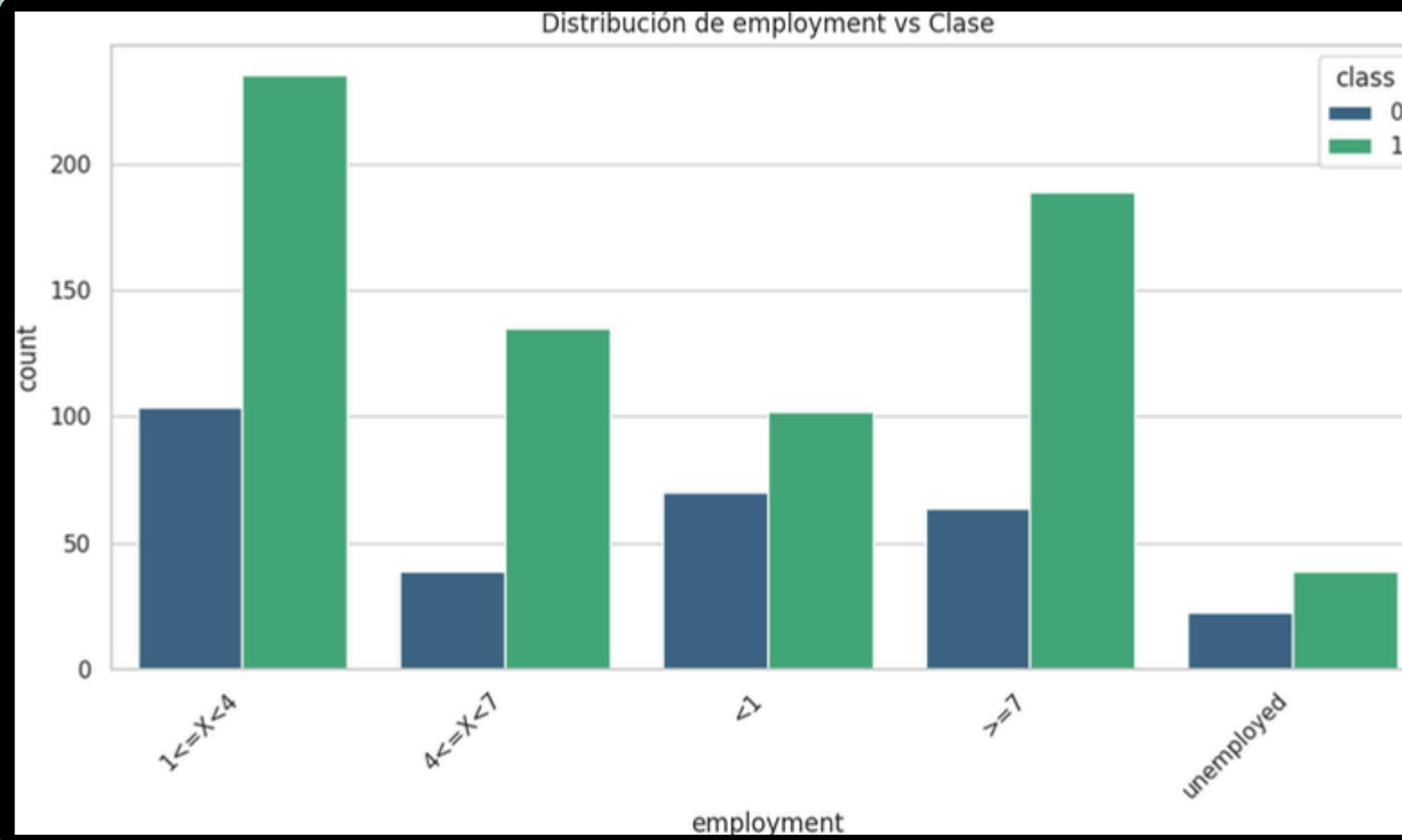


ESTADÍSTICA

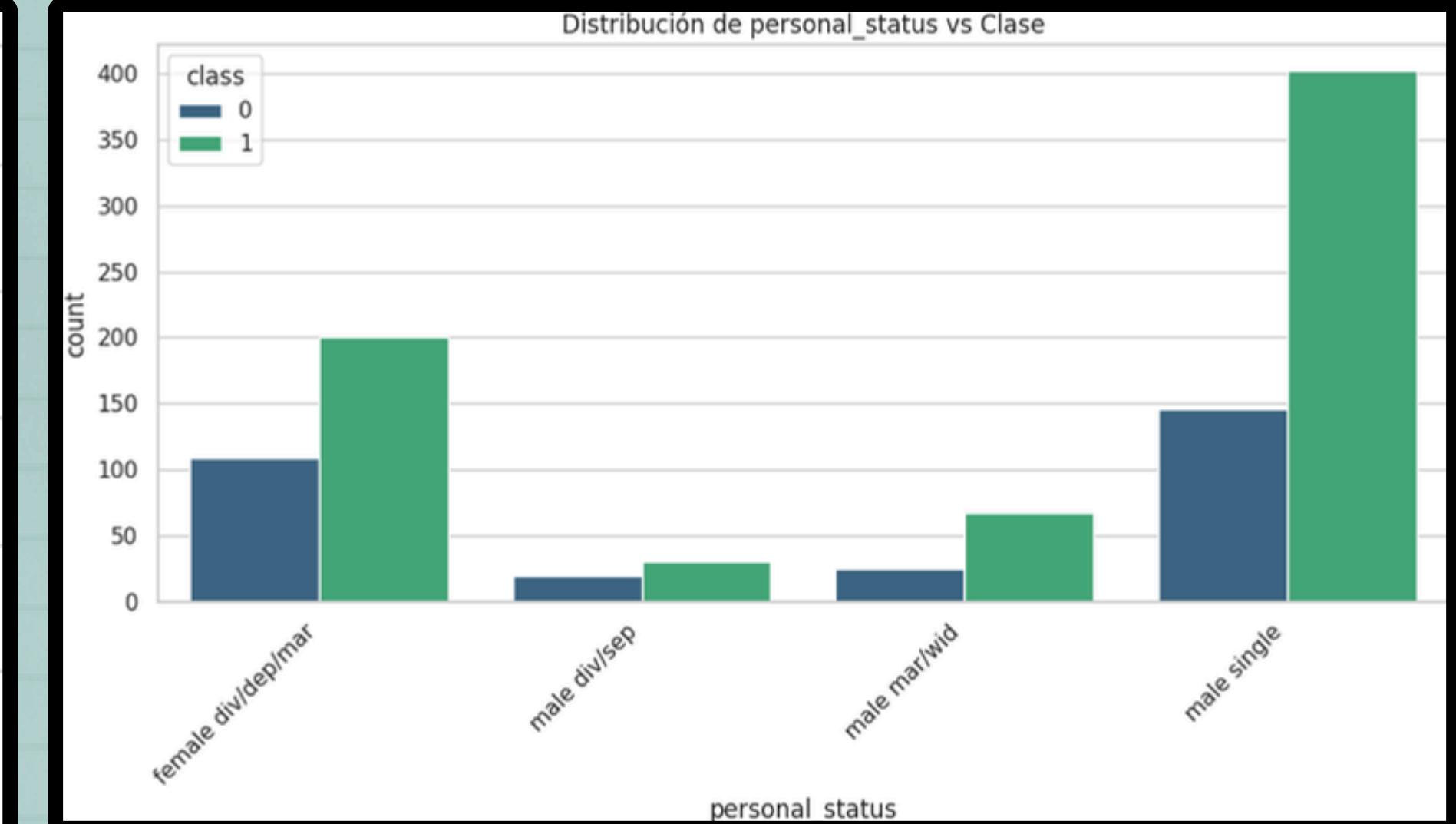


ESTADÍSTICA

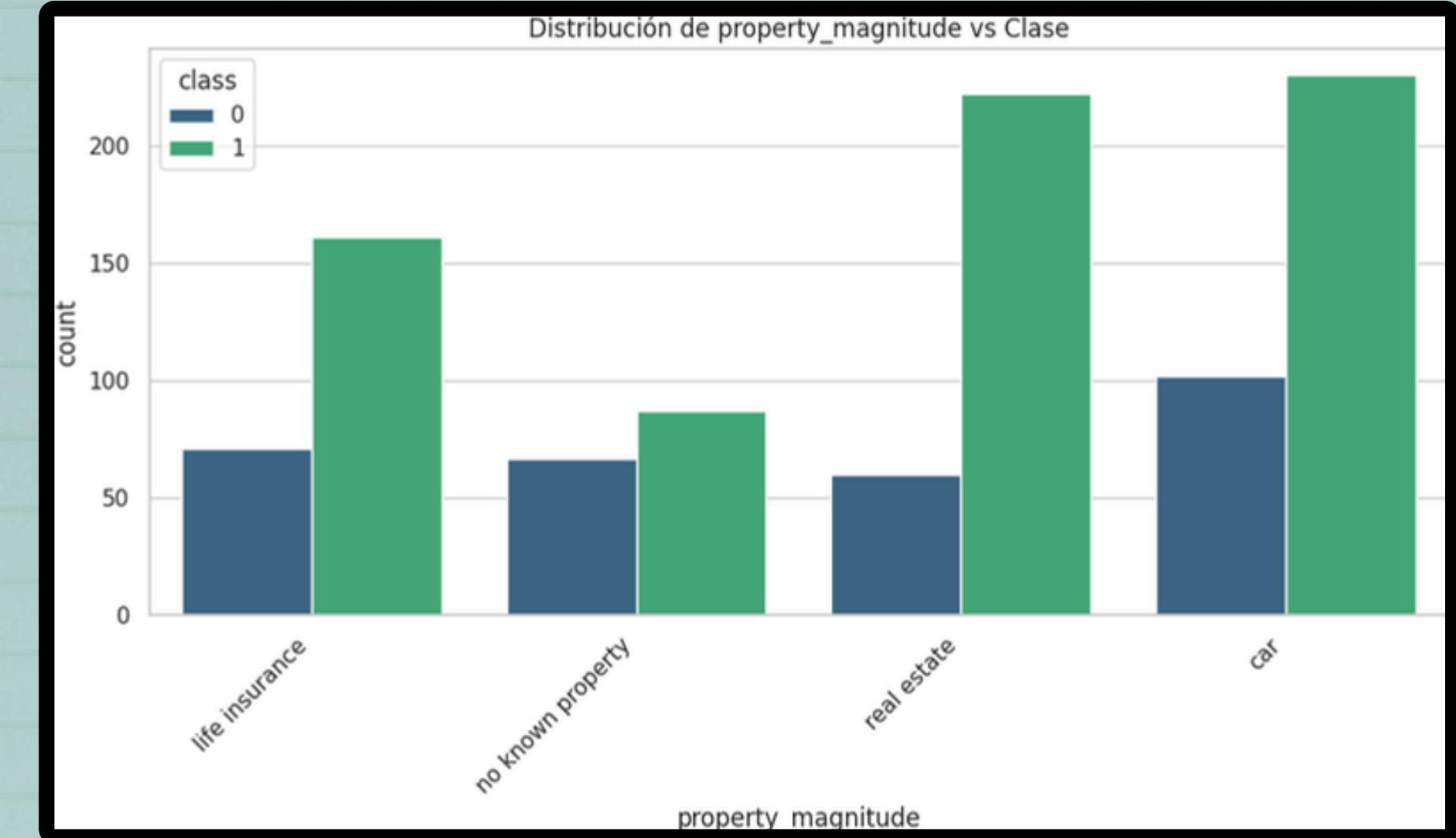
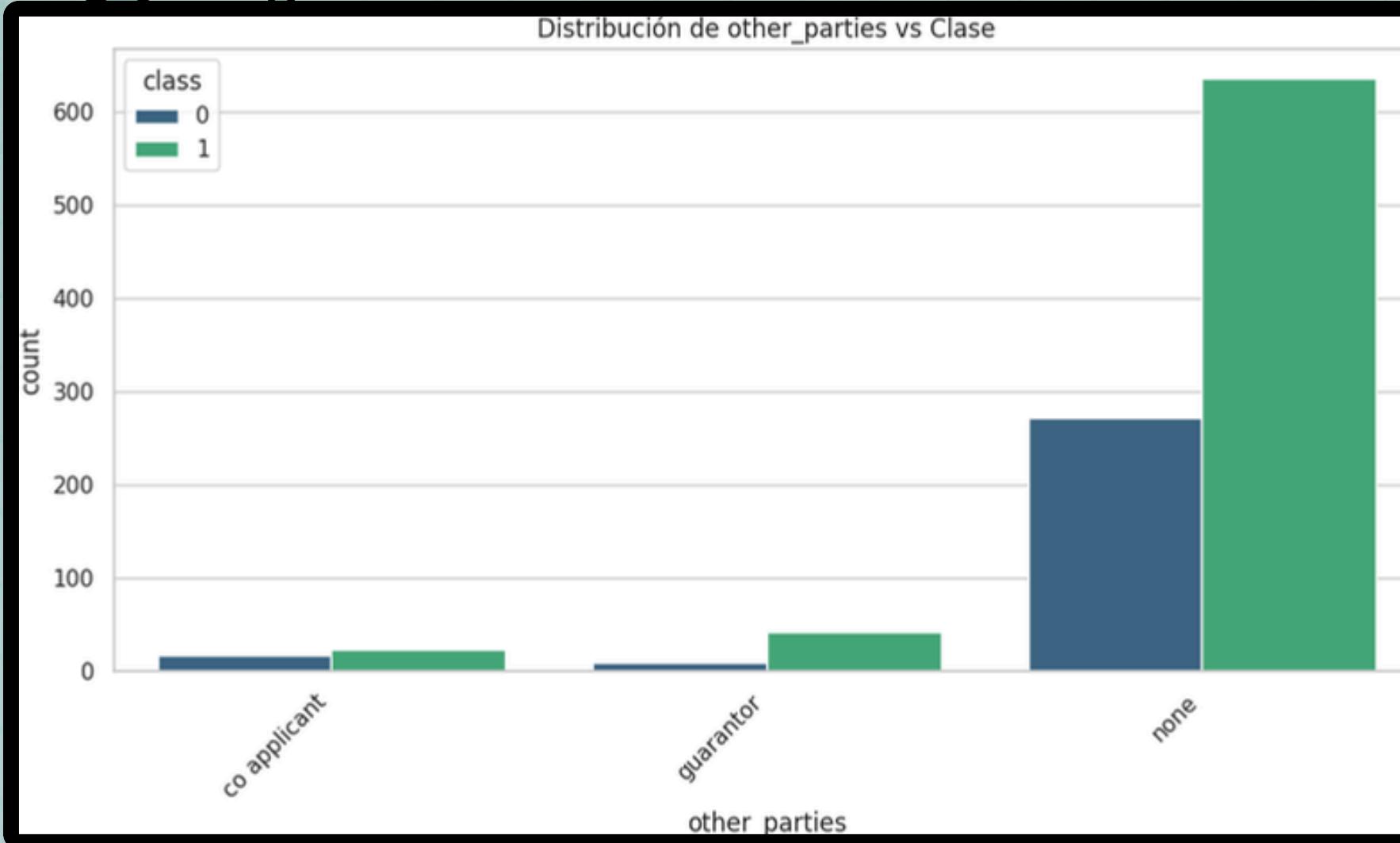
Distribución de employment vs Clase



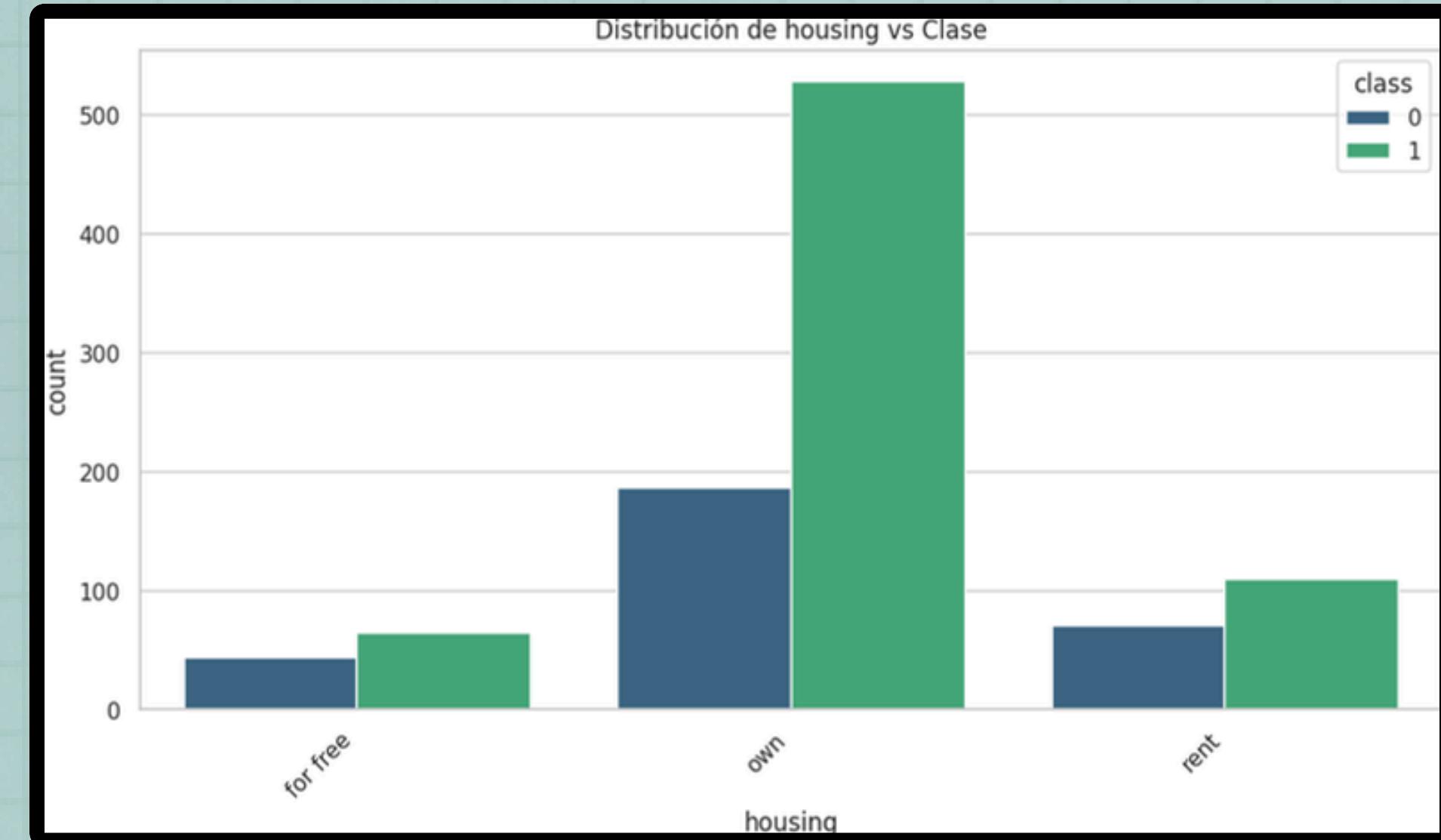
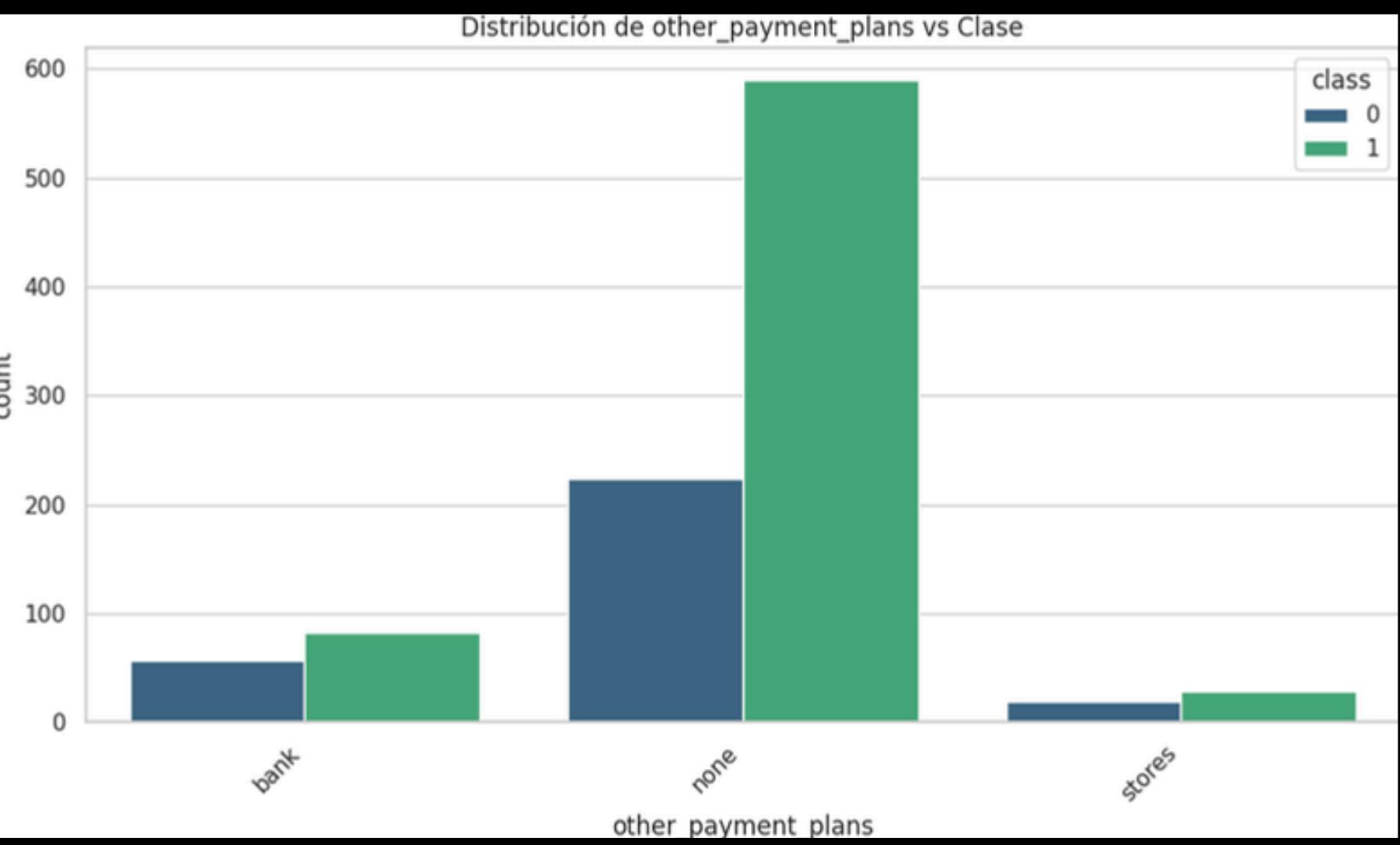
Distribución de personal_status vs Clase



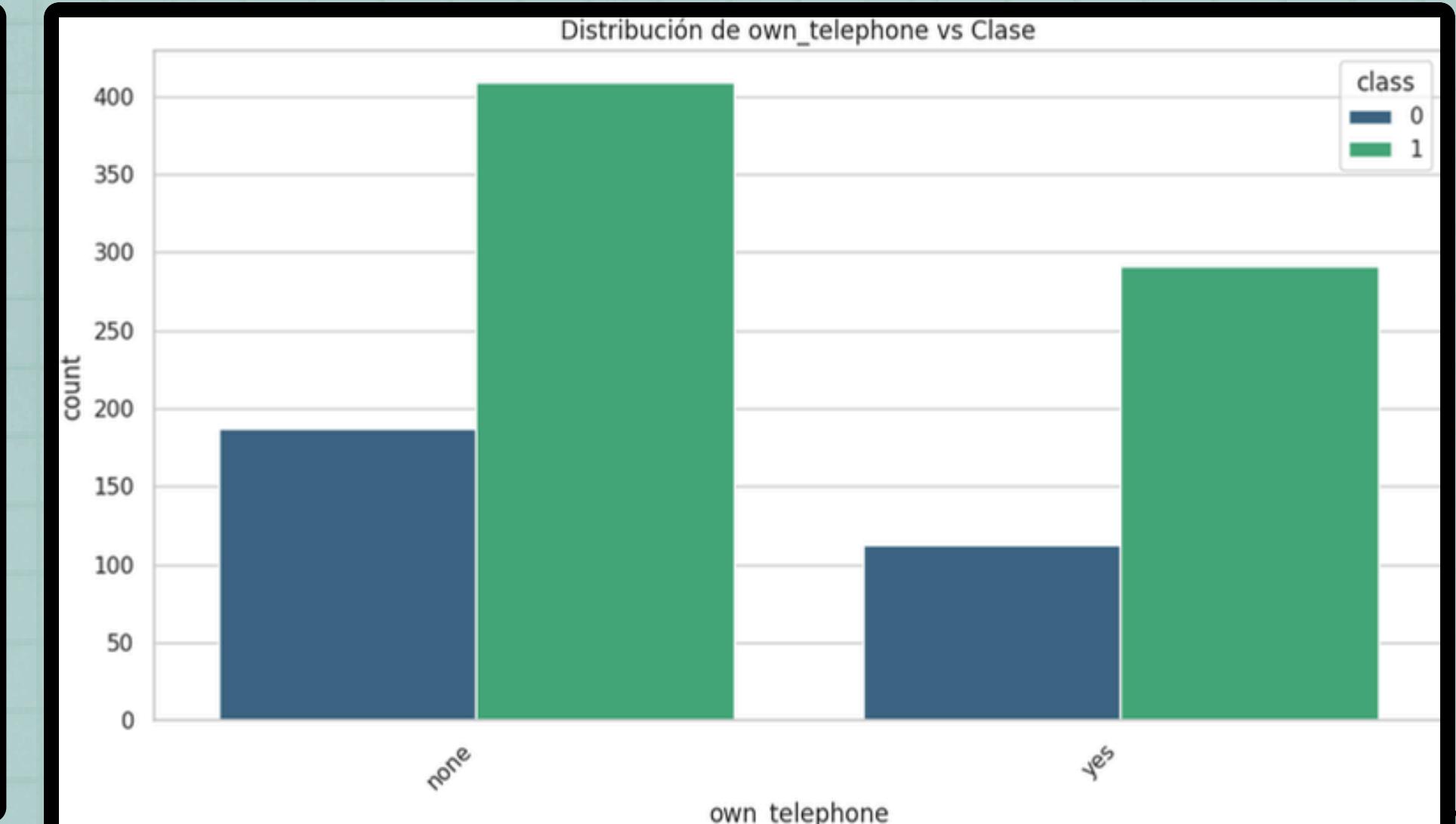
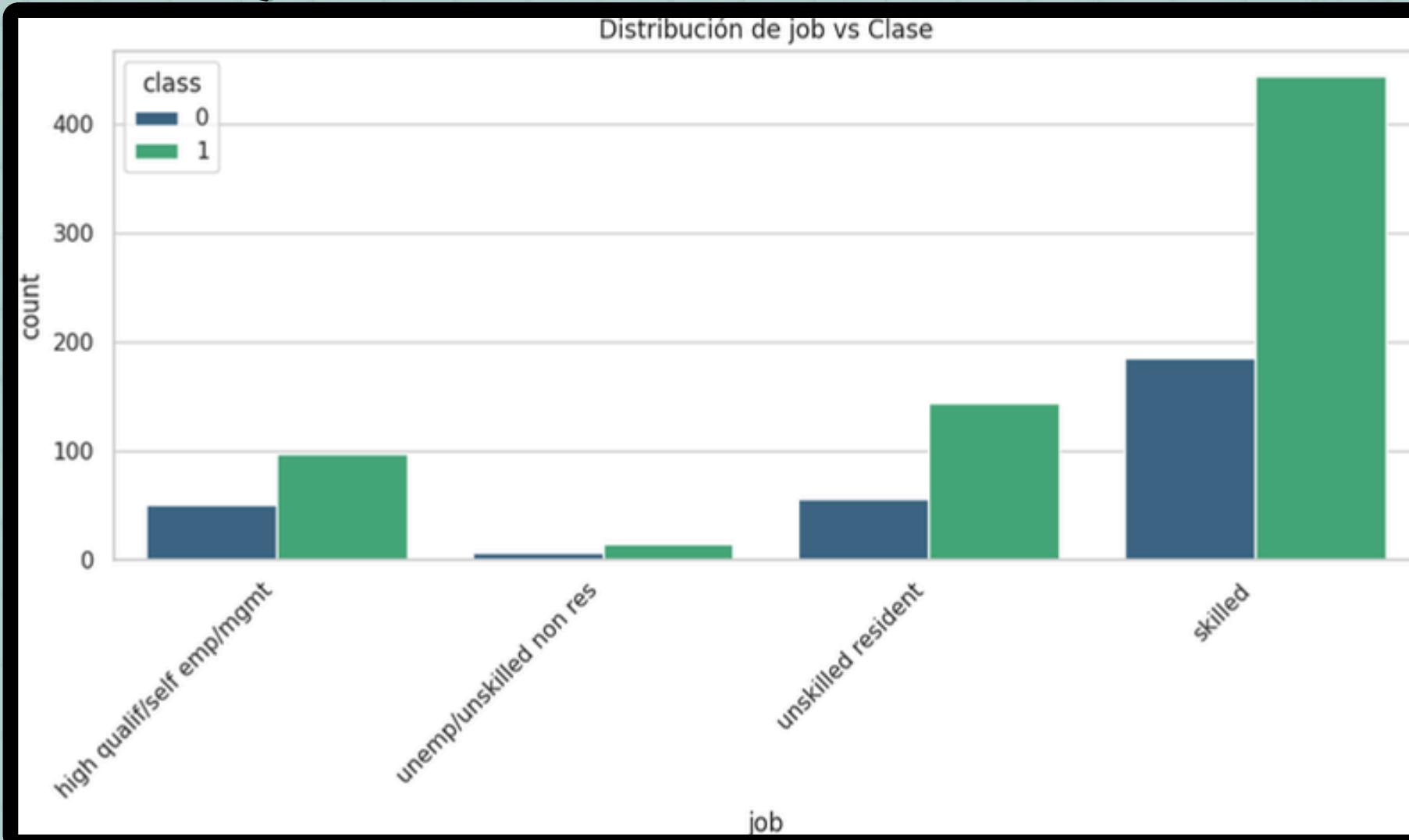
ESTADÍSTICA



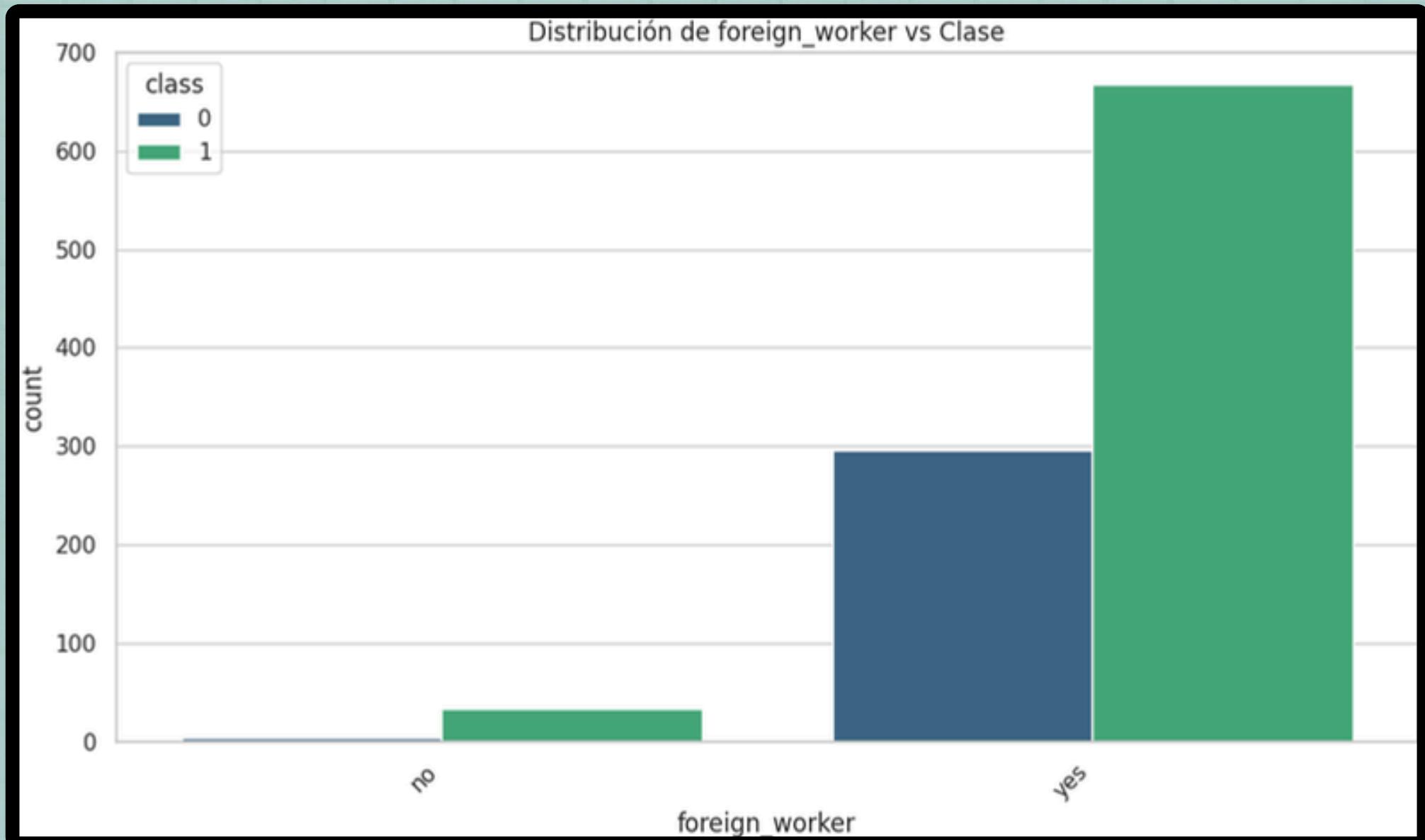
ESTADÍSTICA



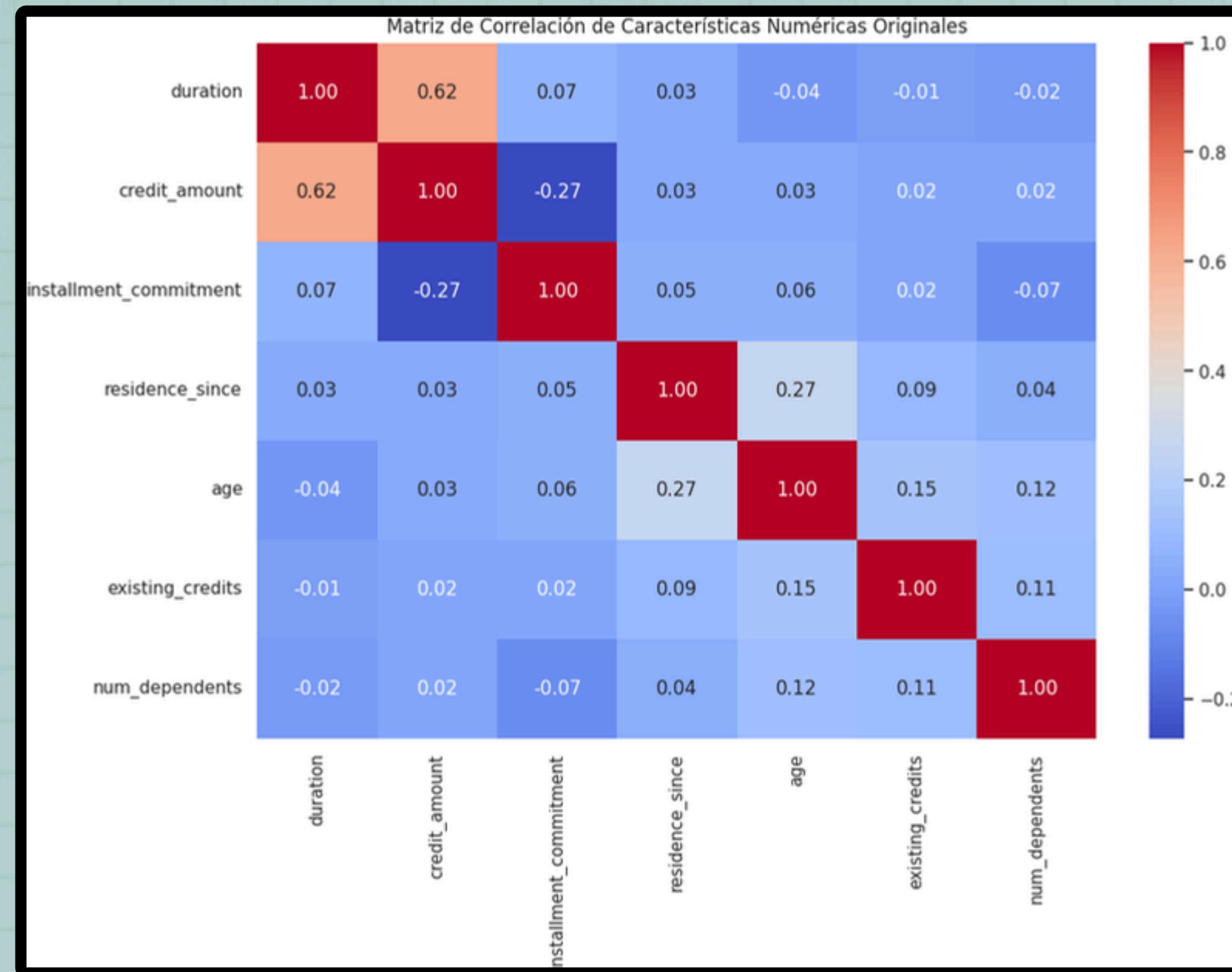
ESTADÍSTICA



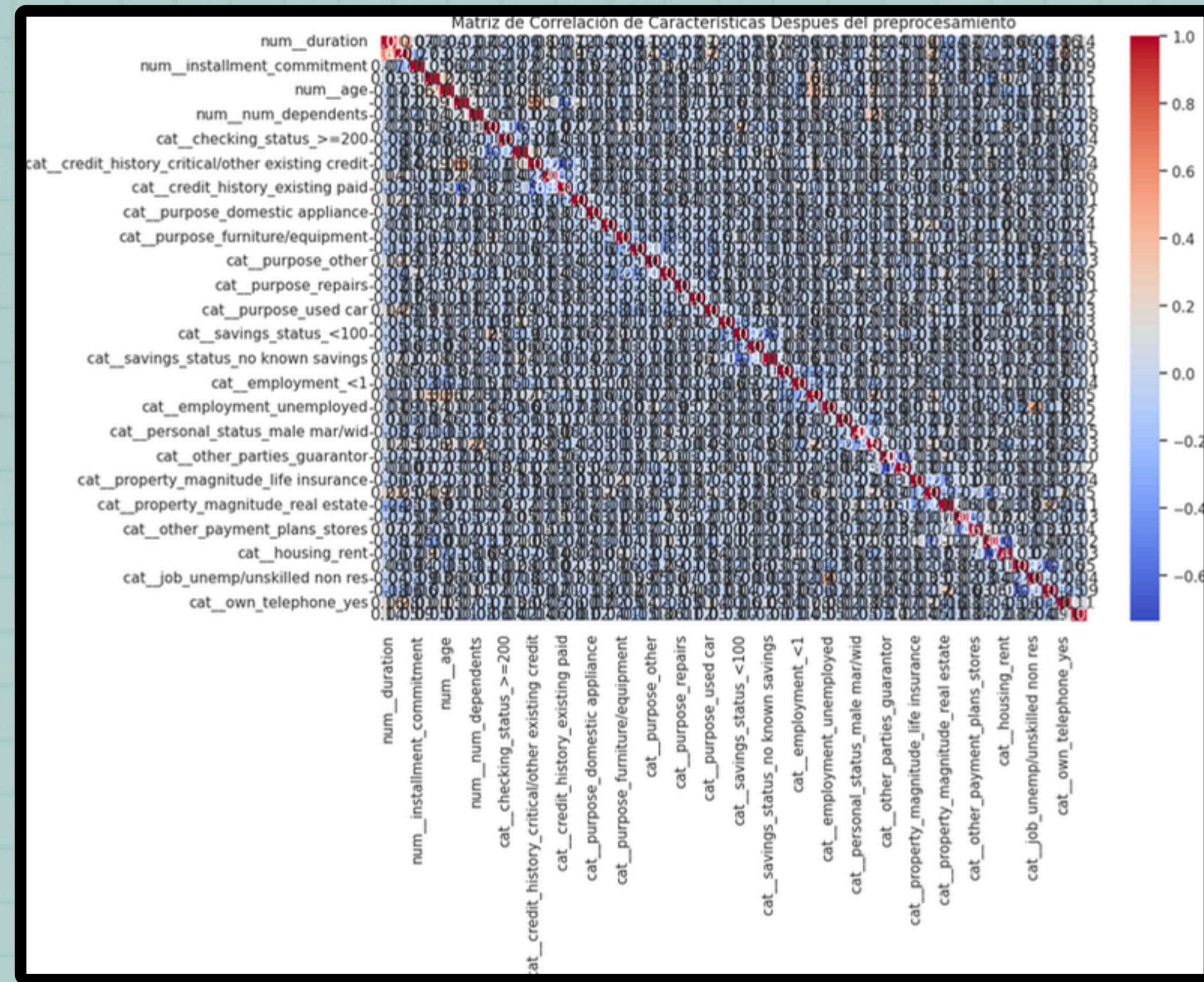
ESTADÍSTICA



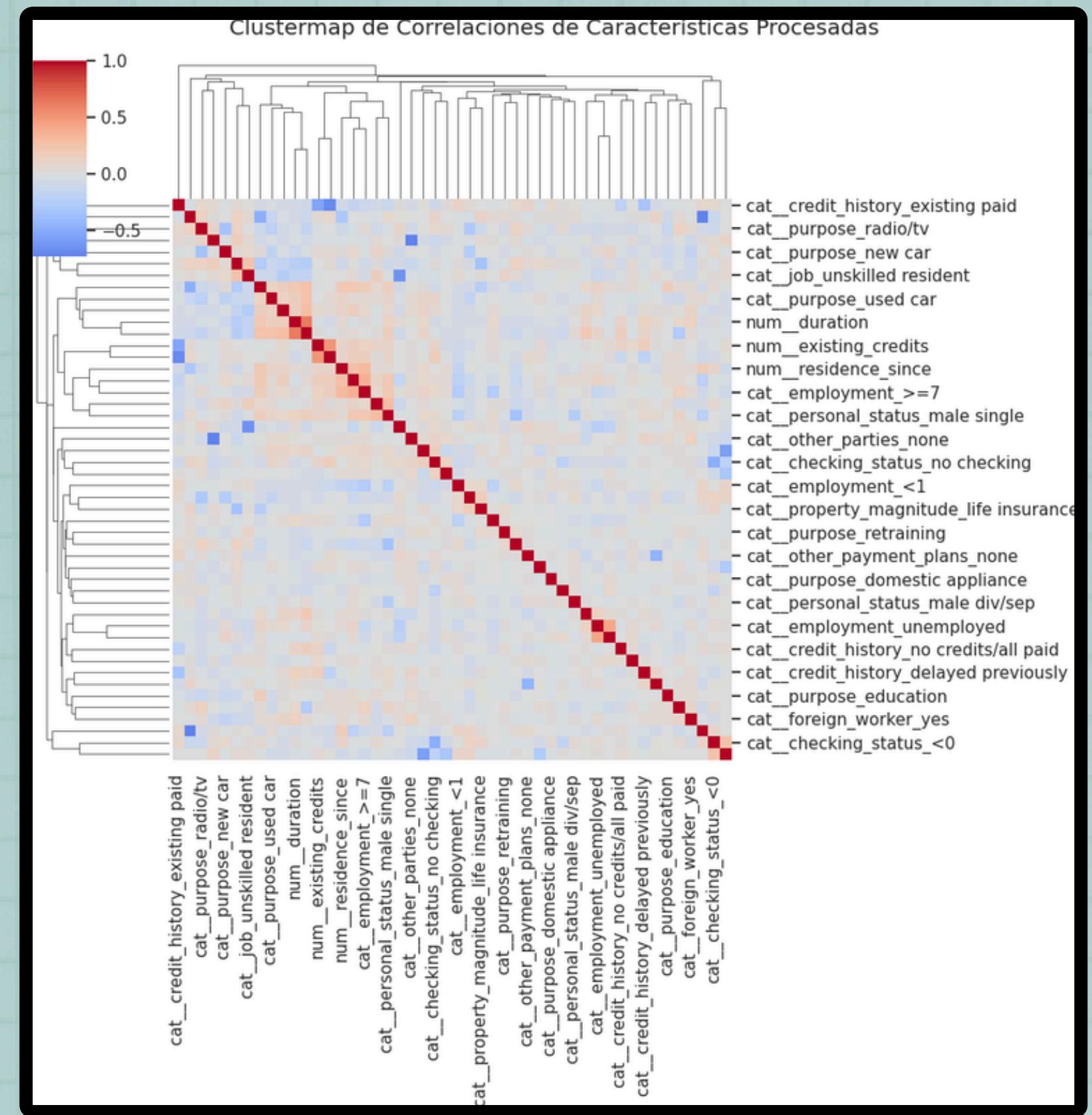
ESTADÍSTICA



ESTADÍSTICA



ESTADÍSTICA



APRENDIZAJE SUPERVISADO

RANDOM FOREST

Métricas para Random Forest (Directo):

Accuracy: 0.7400

Precision: 0.7683

Balanced Recall (Macro Recall): 0.6333 <--- Métrica Principal

Confusion Matrix:

```
[[ 22  38]
 [ 14 126]]
```

DL

Metricas para DNN (Directa):

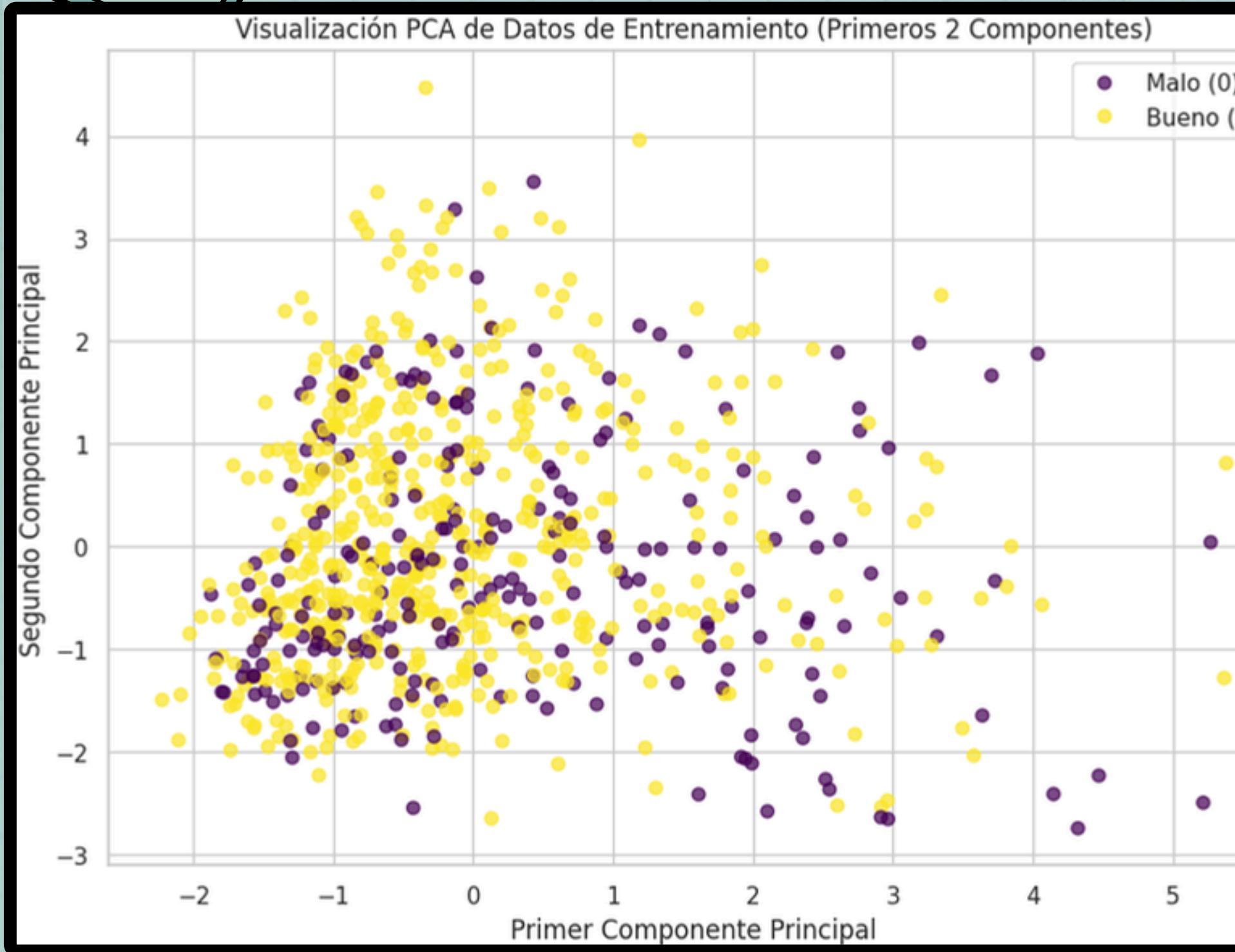
Accuracy: 0.6950

Precision: 0.7724

Balanced Recall (Macro Recall): 0.6250 <--- Métrica Principal

APRENDIZAJE NO SUPERVISADO

PCA



Datos transformados por t-SNE para evaluacion.

Métricas para Random Forest en t-SNE:

Accuracy: 0.6200

Precision: 0.7025

Balanced Recall (Macro Recall): 0.5048 <-- Métrica Principal

Confusion Matrix:

```
[[ 13  47]
 [ 29 111]]
```

Classification Report:

Métricas para SVC en t-SNE:

Accuracy: 0.6000

Precision: 0.7113

Balanced Recall (Macro Recall): 0.5190 <-- Métrica Principal

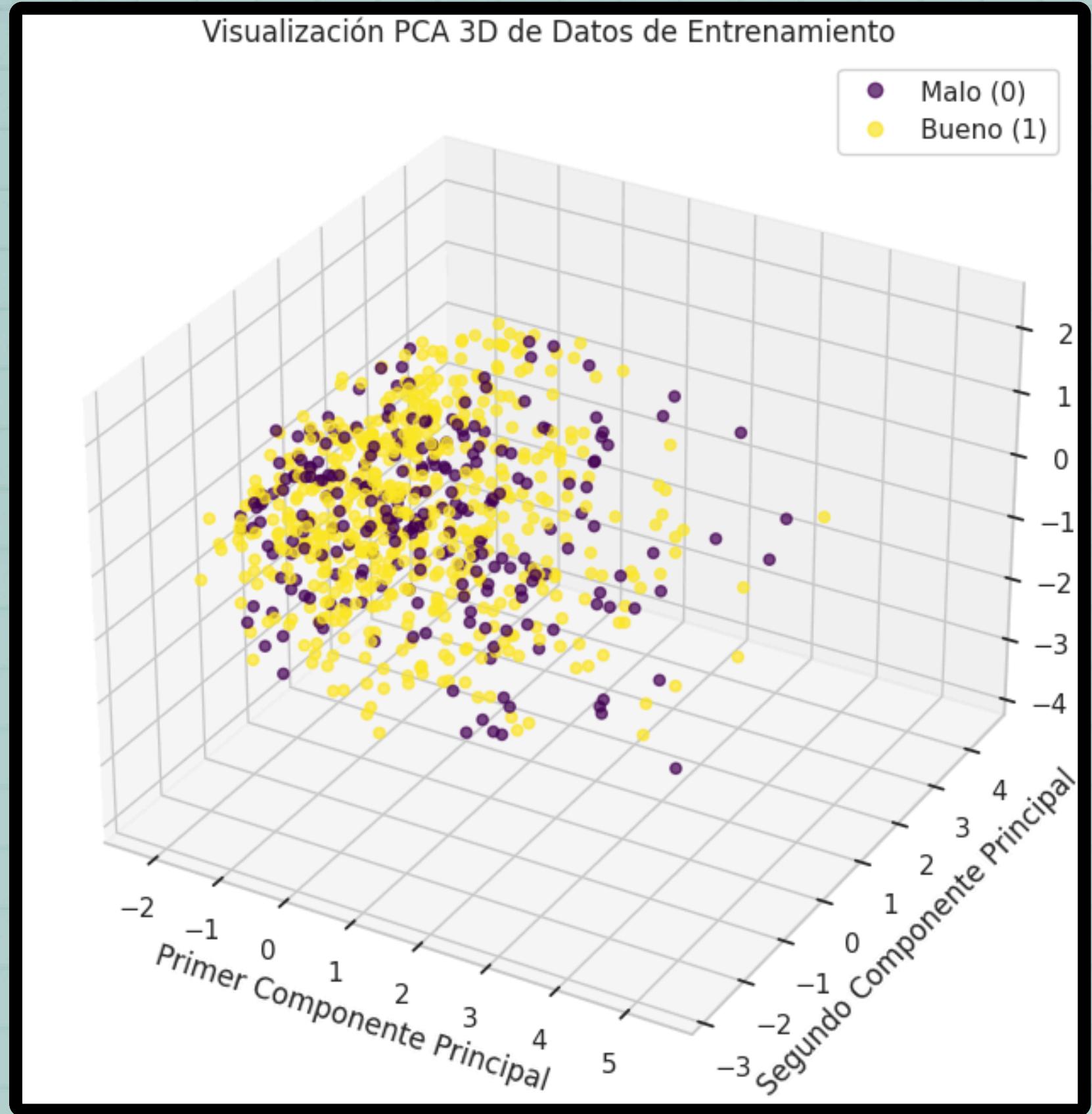
Confusion Matrix:

```
[[ 19  41]
 [ 39 101]]
```

Classification Report:

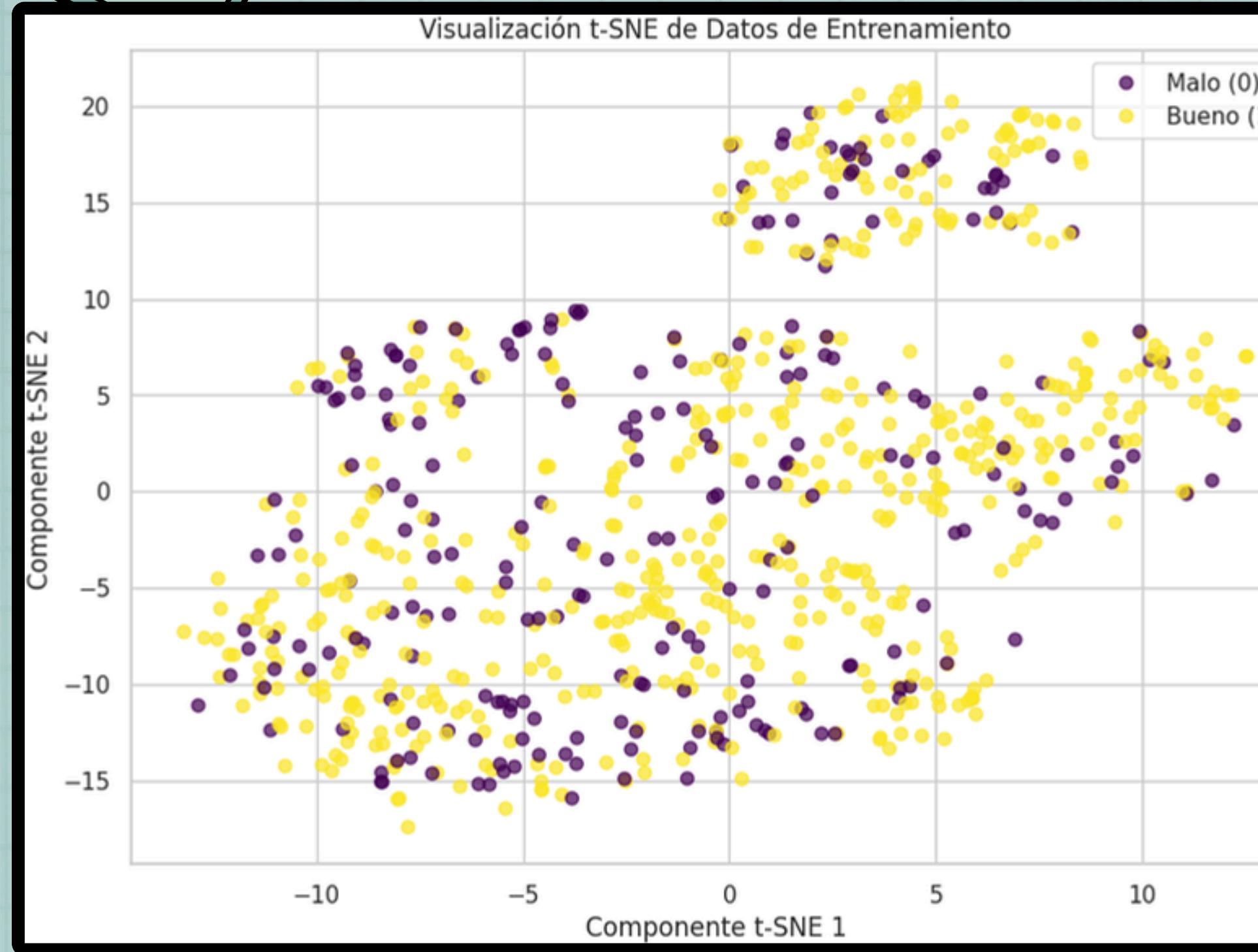
APRENDIZAJE NO SUPERVISADO

PCA



APRENDIZAJE NO SUPERVISADO

TSNE



Datos transformados por t-SNE para evaluacion.

Métricas para Random Forest en t-SNE:

Accuracy: 0.6200

Precision: 0.7025

Balanced Recall (Macro Recall): 0.5048 <-- Métrica Principal

Confusion Matrix:

```
[[ 13  47]
 [ 29 111]]
```

Classification Report:

Métricas para SVC en t-SNE:

Accuracy: 0.6000

Precision: 0.7113

Balanced Recall (Macro Recall): 0.5190 <-- Métrica Principal

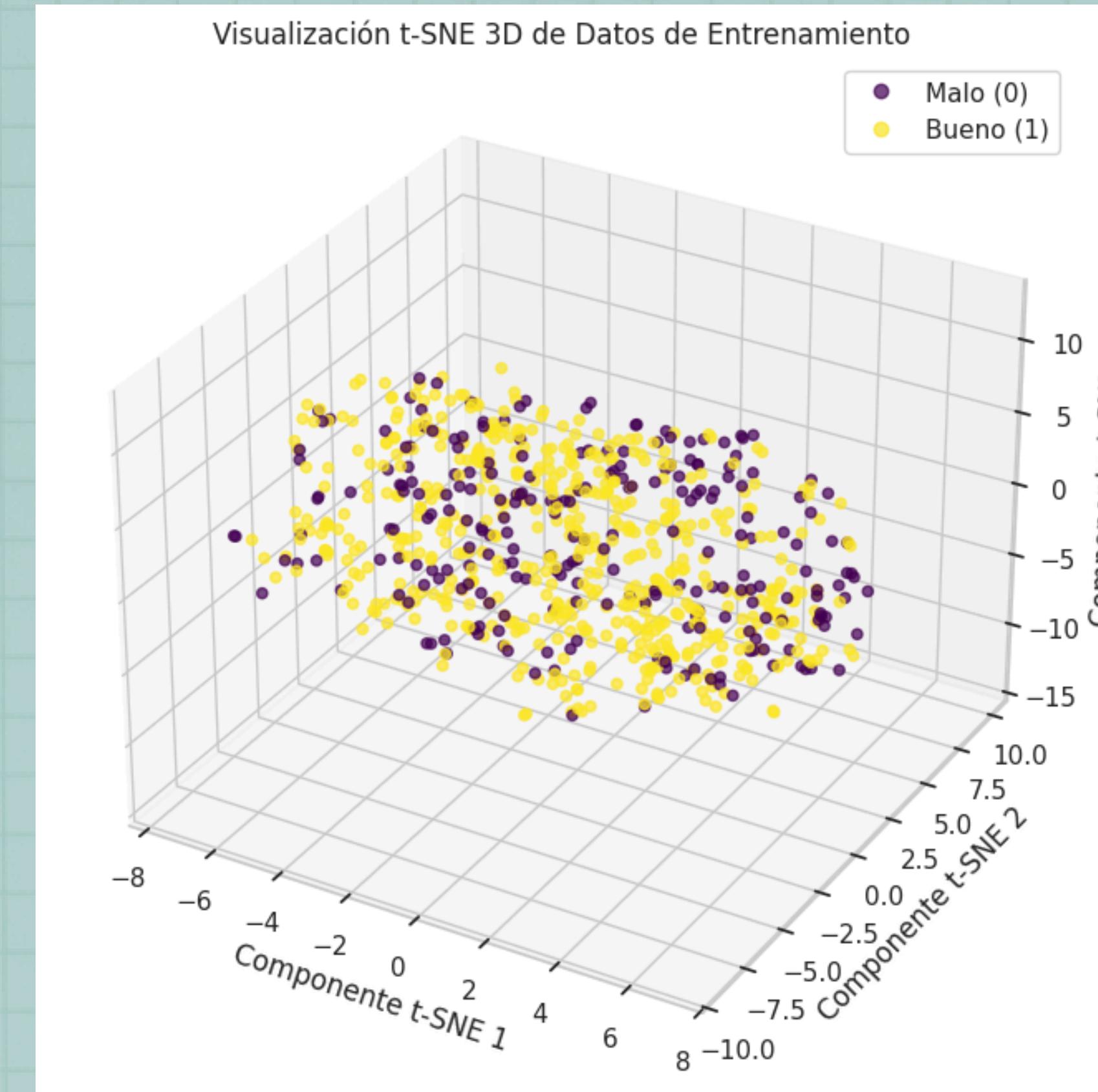
Confusion Matrix:

```
[[ 19  41]
 [ 39 101]]
```

Classification Report:

APRENDIZAJE NO SUPERVISADO

TSNE

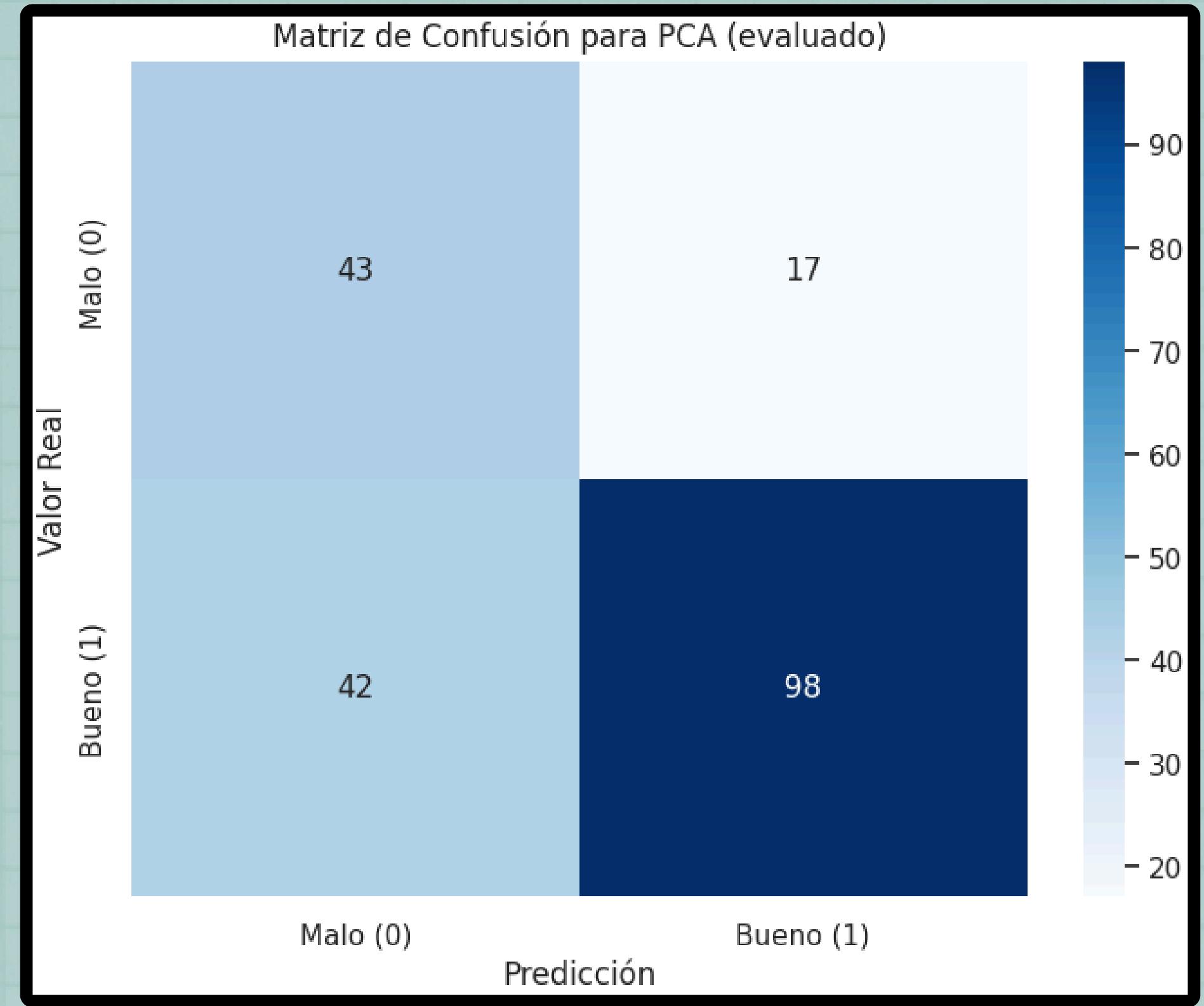


COMPARATIVA MÉTRICAS

Tabla Comparativa General de Modelos:

Modelo	Tipo	Balanced Recall (Macro Recall)	Accuracy	Precision
PCA (evaluado)	No Supervisado (PCA)	0.7083	0.7050	0.8522
SVC	Supervisado	0.7048	0.7000	0.8509
Random Forest	Supervisado	0.6333	0.7400	0.7683
DNN	Supervisado	0.6250	0.6950	0.7724
t-SNE (evaluado)	No Supervisado (t-SNE)	0.5190	0.6000	0.7113

PCA + SVC MEJOR MODELO





¡Gracias!
POR LA ATENCIÓN