

2023

Reto Titanic

**Inteligencia Artificial Avanzada para la
Ciencia de Datos I**

Equipo 5

María Lynne Camacho Padilla

A01423135

Emiliano Ferreira Guadarrama

A01654418

Humberto Alejandro Rosas Téllez

A01659823

Victoria Estefanía Vázquez Morales

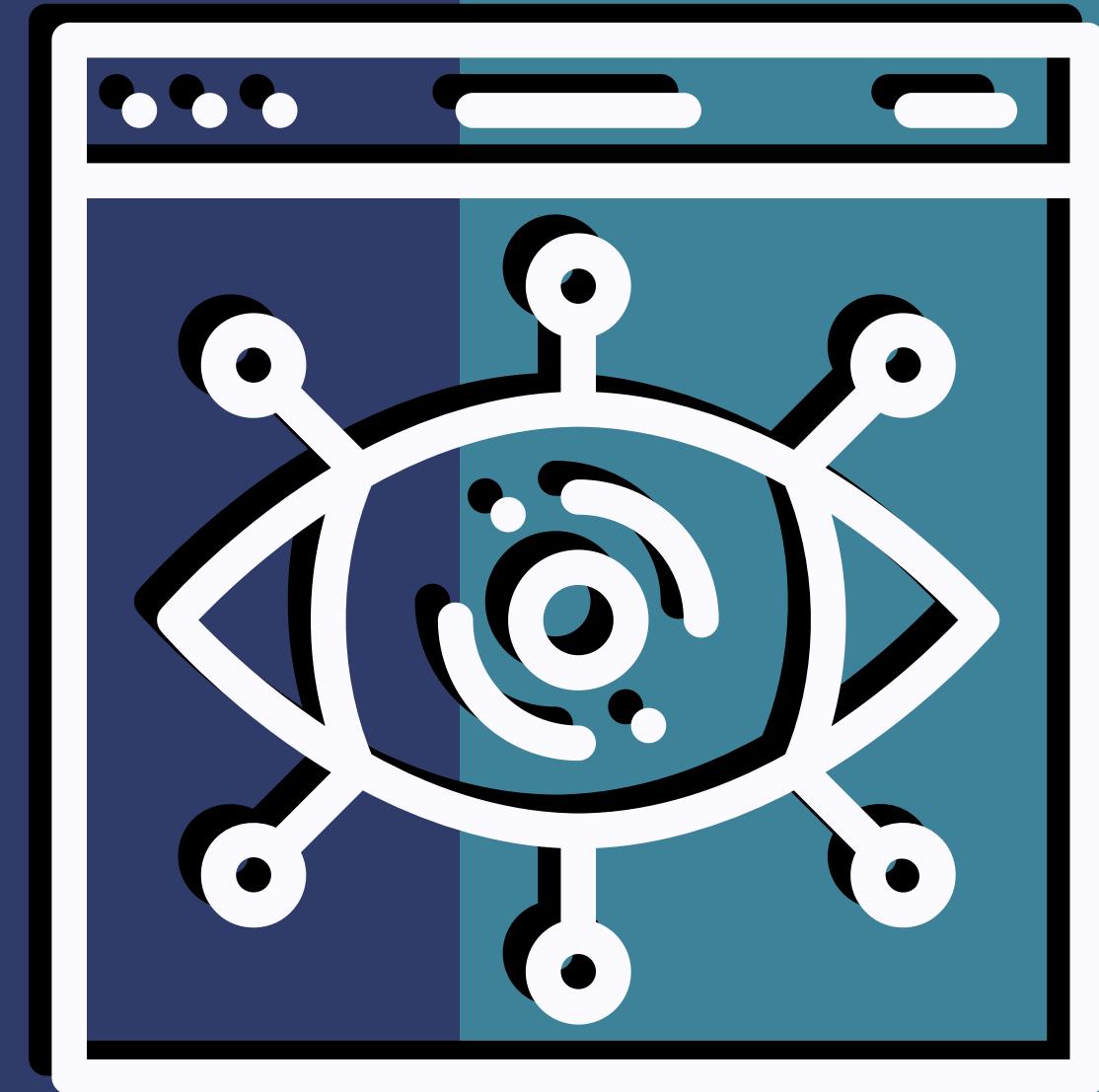
A01654095

Índice

- 03** EXPLORACIÓN
- 08** PREPROCESAMIENTO
- 11** SELECCIÓN DE MODELOS
- 15** RESULTADOS

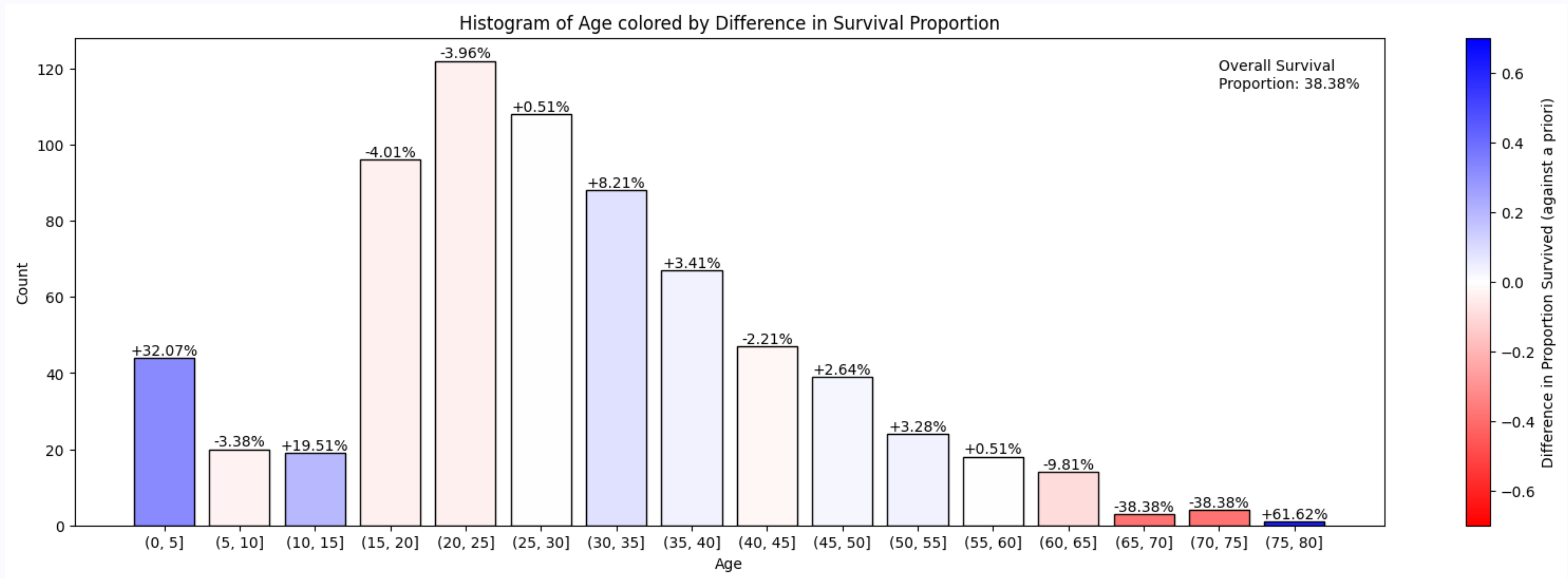
Exploración

- **Relaciones entre variables**
- **Distribuciones de datos numéricos**
- **Datos Faltantes**



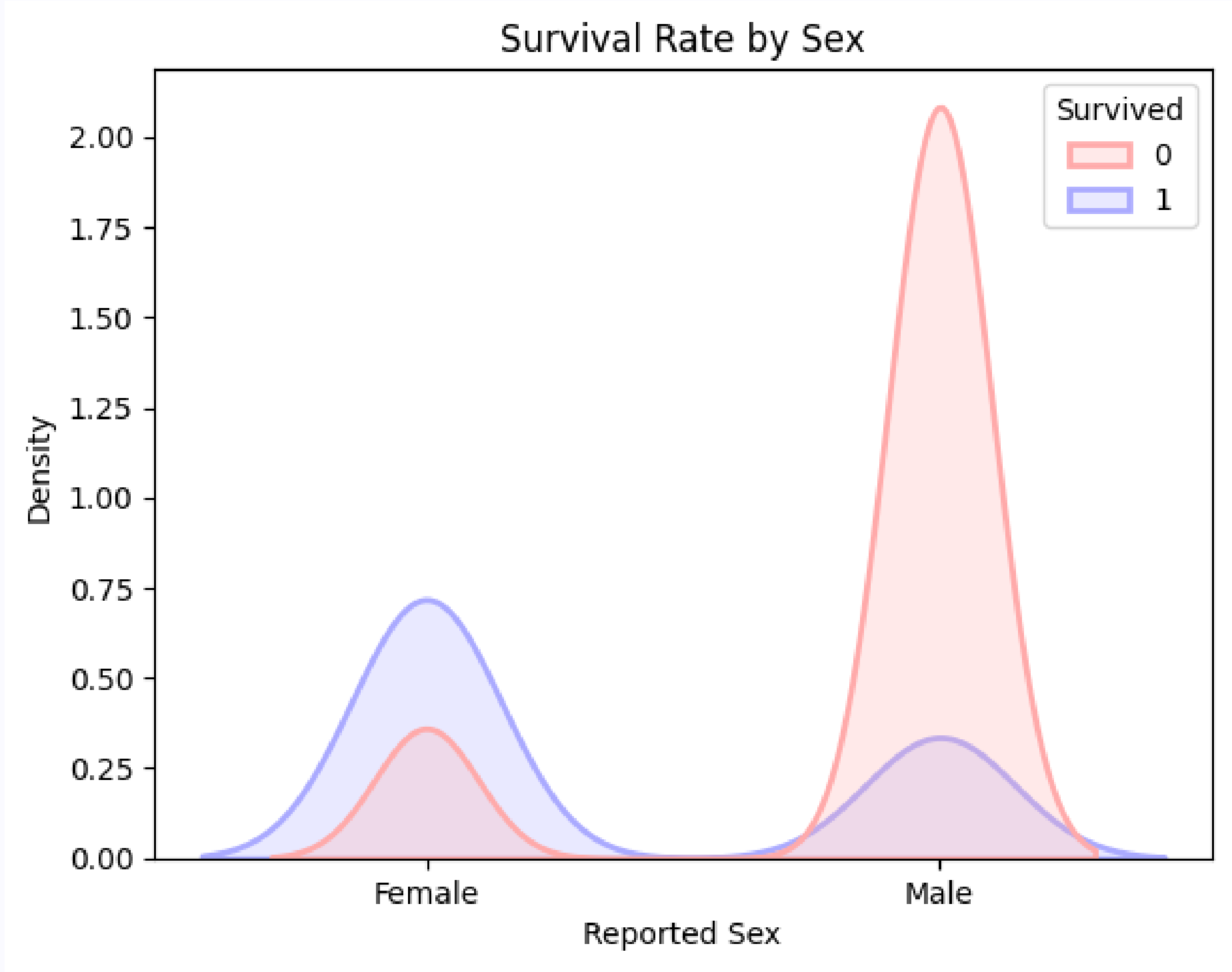
Relaciones Encontradas

AGE VS SURVIVAL

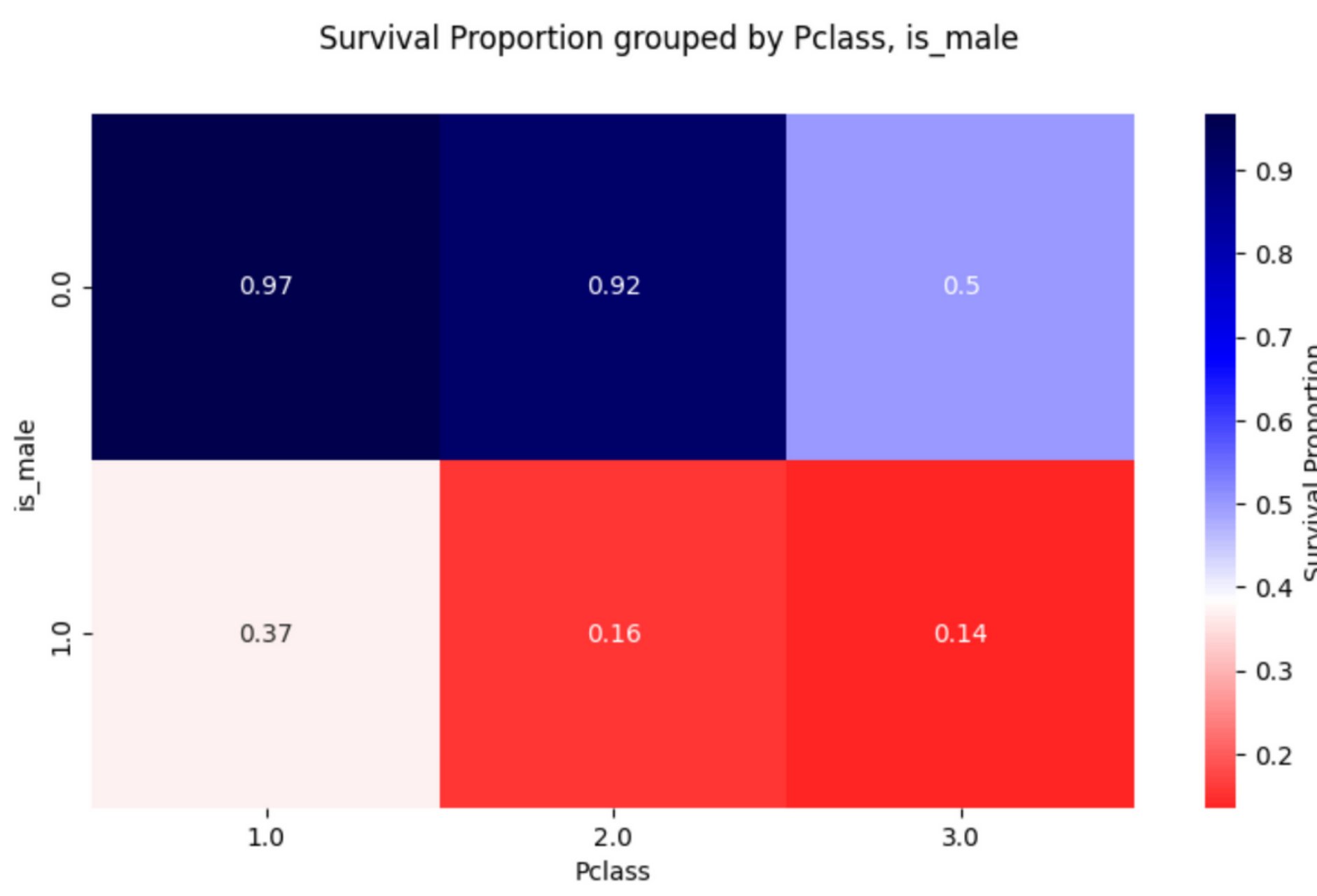




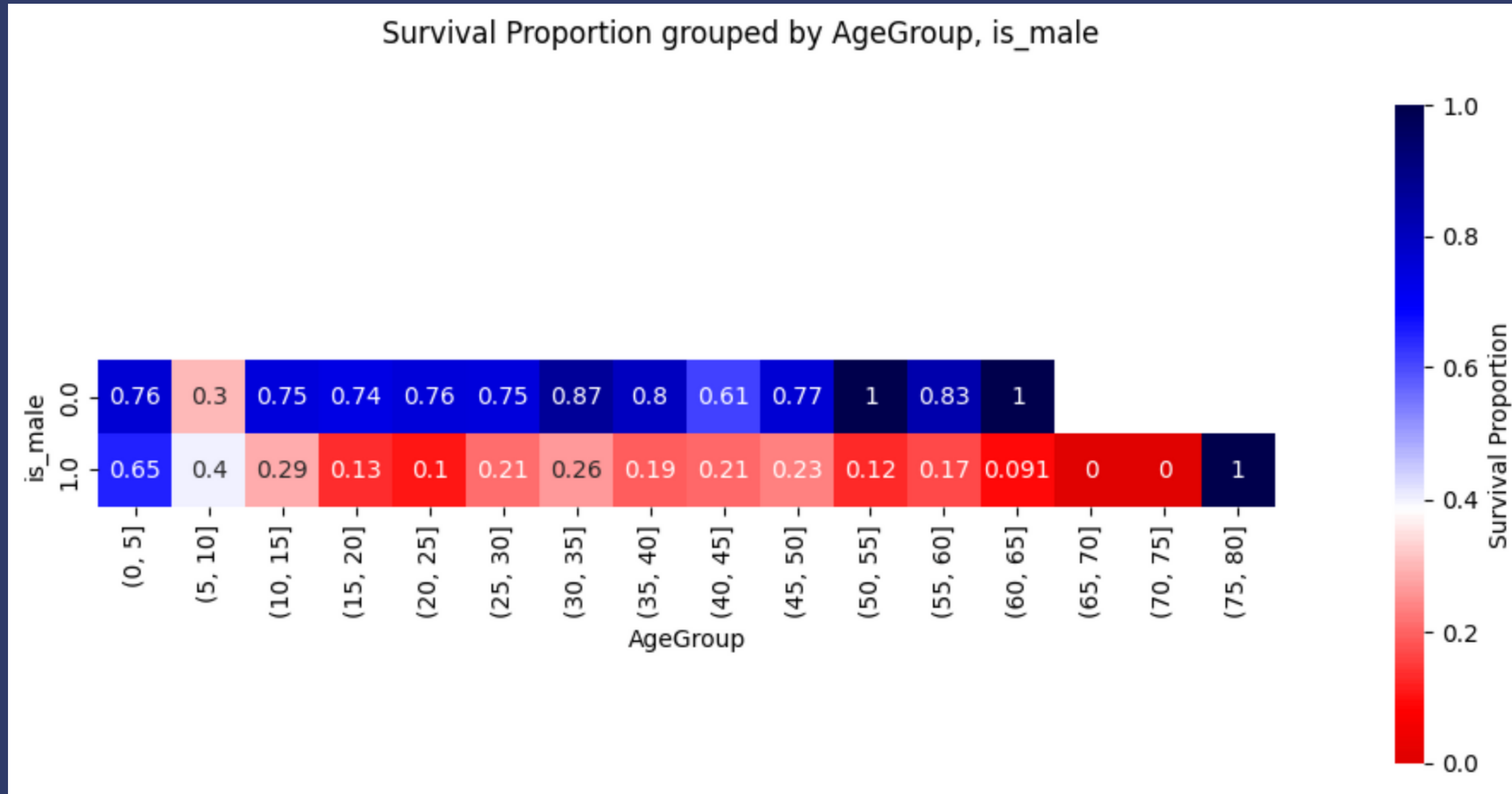
SEX VS SURVIVAL



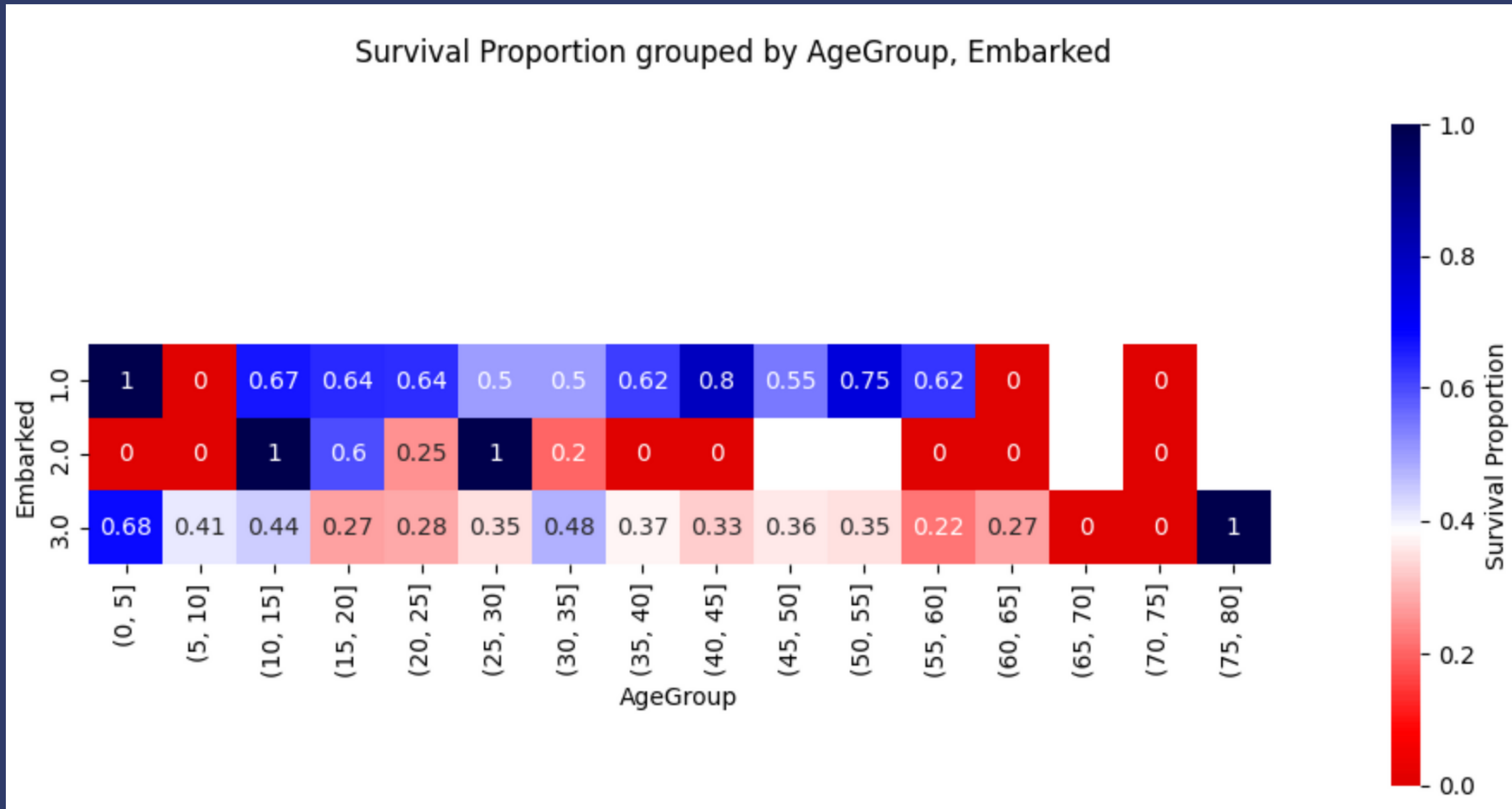
PCLASS AND SEX VS SURVIVAL



AgeGroup vs is_male



AgeGroup vs Embarked



Preprocessing

- **Columnas Eliminadas**
- **Columnas Transformadas**
- **Datos Imputados**
- **Escalamiento**



Preprocessing Pipeline

09

1

~~PassengerID
Name
Cabin
Ticket
Sexo
Embarked~~

2

Discretización
Sex

3

Embarked
One Hot
Encode

4

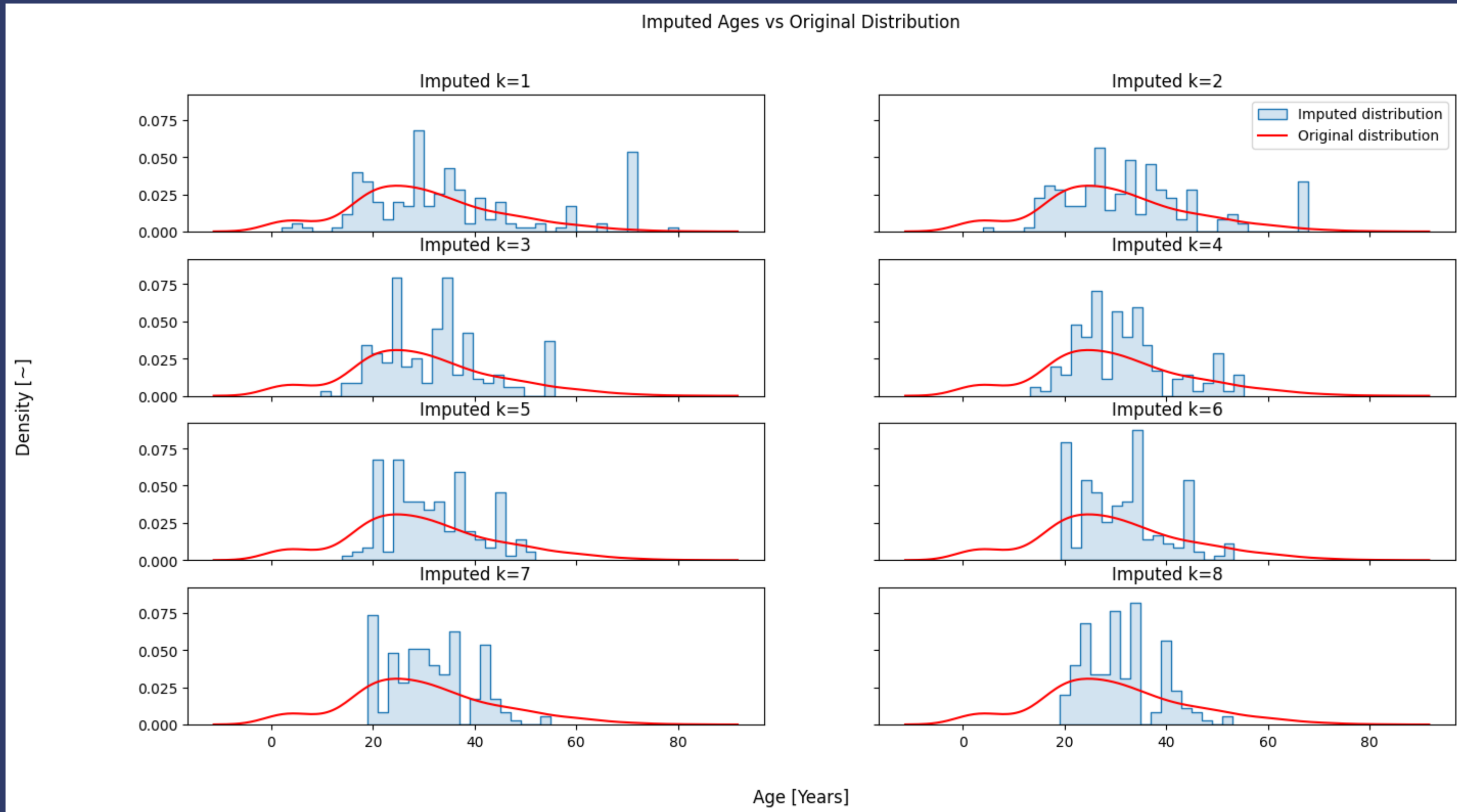
KNN=4
Imputación

5

Estandarización

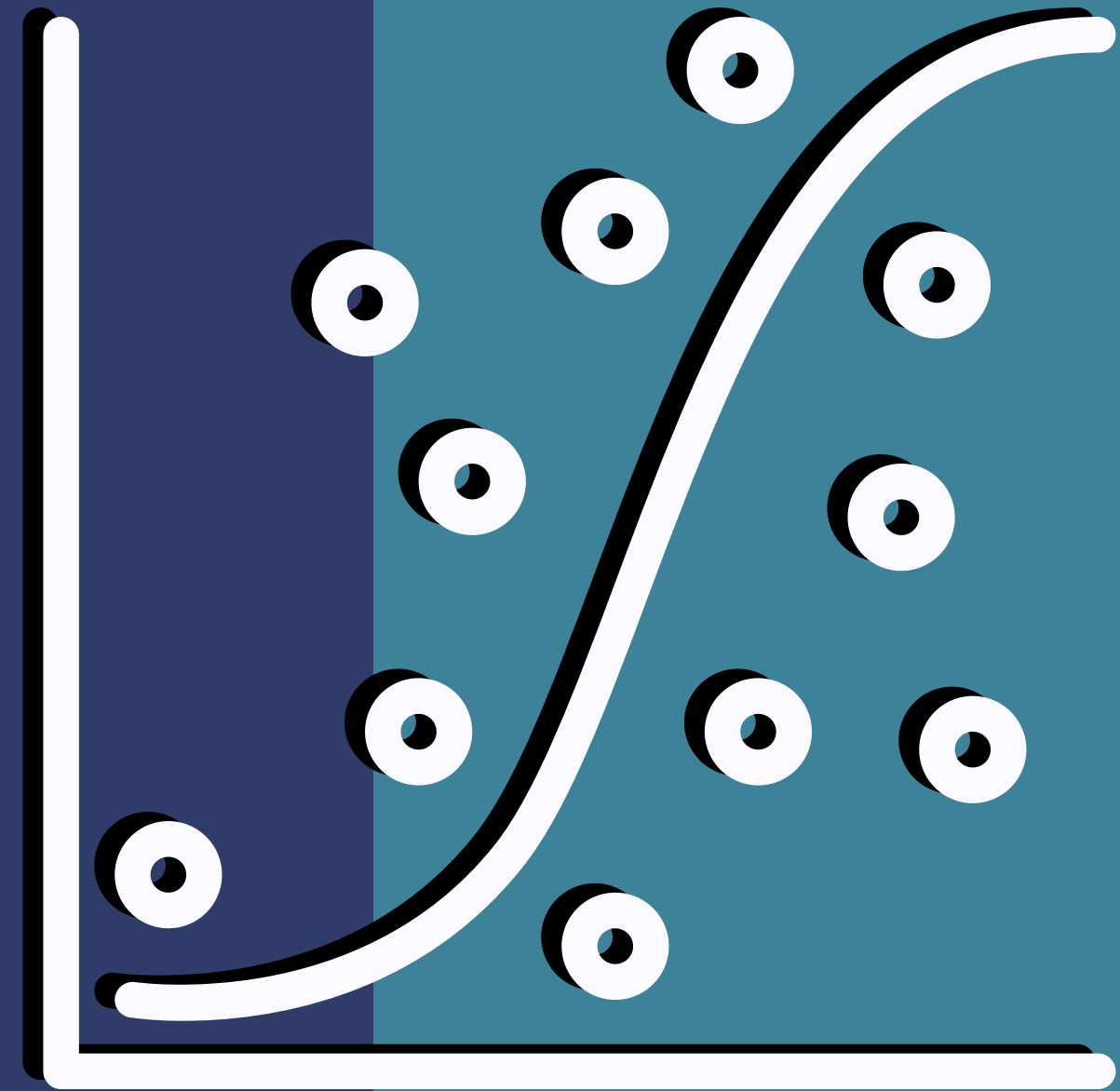
KNN Imputer

10



Selección del Modelo

- **Hyper-Parameter Grid Search**
- **K-Fold Cross Validation**
- **Metrics**



Justificación de uso de k-cross validation

K=10

Maximizar datos
disponibles

Tamaño del Dataset

Justificación de uso de modelos

RANDOM FOREST	SVM KERNEL RBF	KNN
Clasificaciones binarias	Modelos no lineales	Fácil de implementar/ usar
Resistencia a overfitting	Poca Memoria	No requiere de hiperparámetros
Trabajar con datos categóricos y faltantes	Robusta frente a Outliers	Se adapta fácilmente

Parámetros utilizados

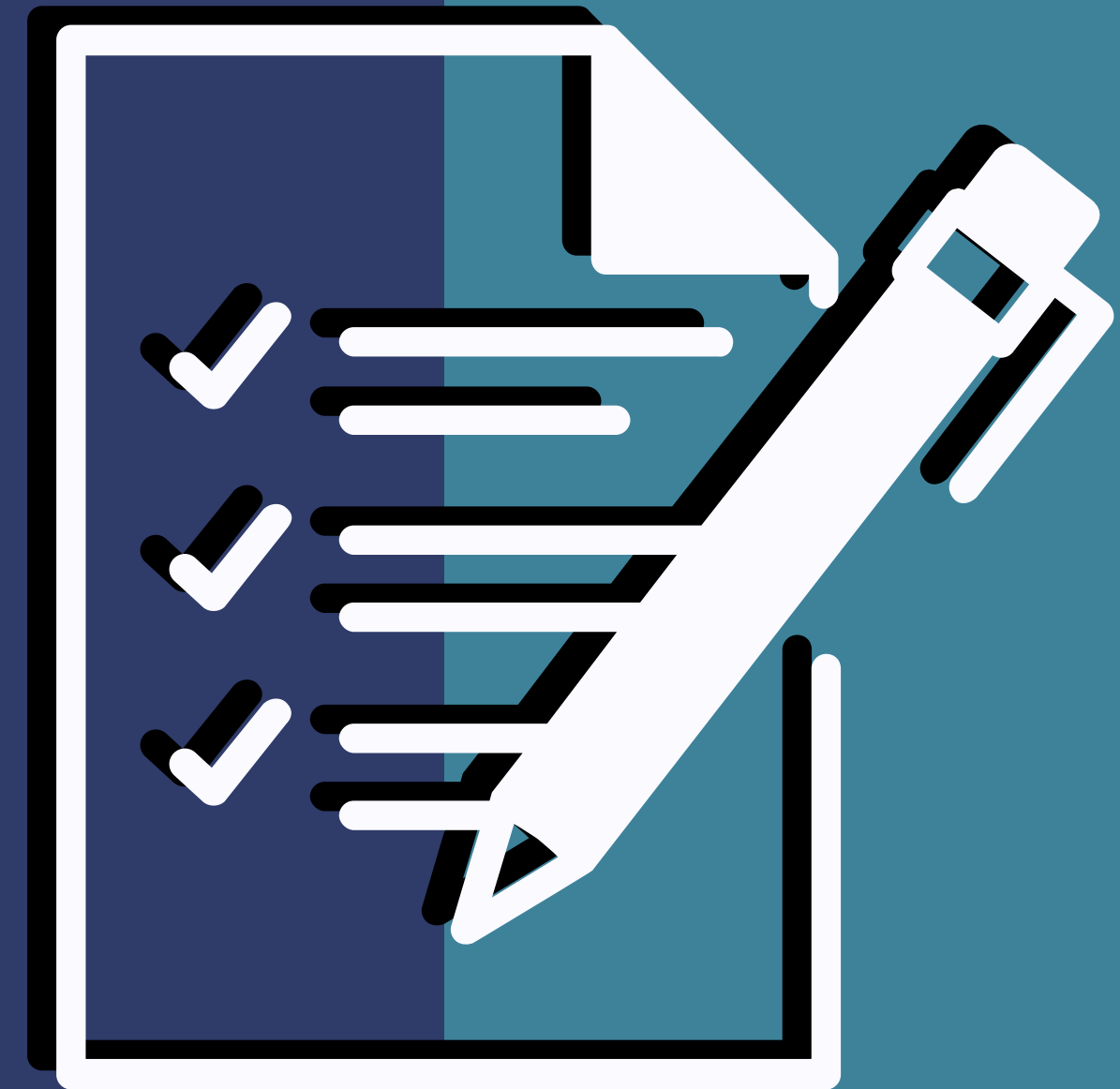
	Random Forest		Kernel SVM		KNN
n_estimators	100	degree	3	n_neighbors	10
criterion	gini	gamma	auto	p	1
min_samples_split	2	kernel	rbf	weights	uniform
min_samples_leaf	2	probability	True		

Reporte de métricas

Train				Validation			
Métrica	Random Forest	SVM Kernel	KNN	Métrica	Random Forest	SVM Kernel	KNN
Exactitud	0.9242	0.8383	0.8385	Exactitud	0.8302	0.8076	0.8118
Precisión	0.9432	0.8345	0.8637	Precisión	0.8221	0.7995	0.8306
AUC (ROC)	0.9840	0.8834	0.9001	AUC (ROC)	0.8624	0.8439	0.8445
F1 Weight	0.9194	0.8359	0.8344	F1 Weight	0.8321	0.8036	0.8063

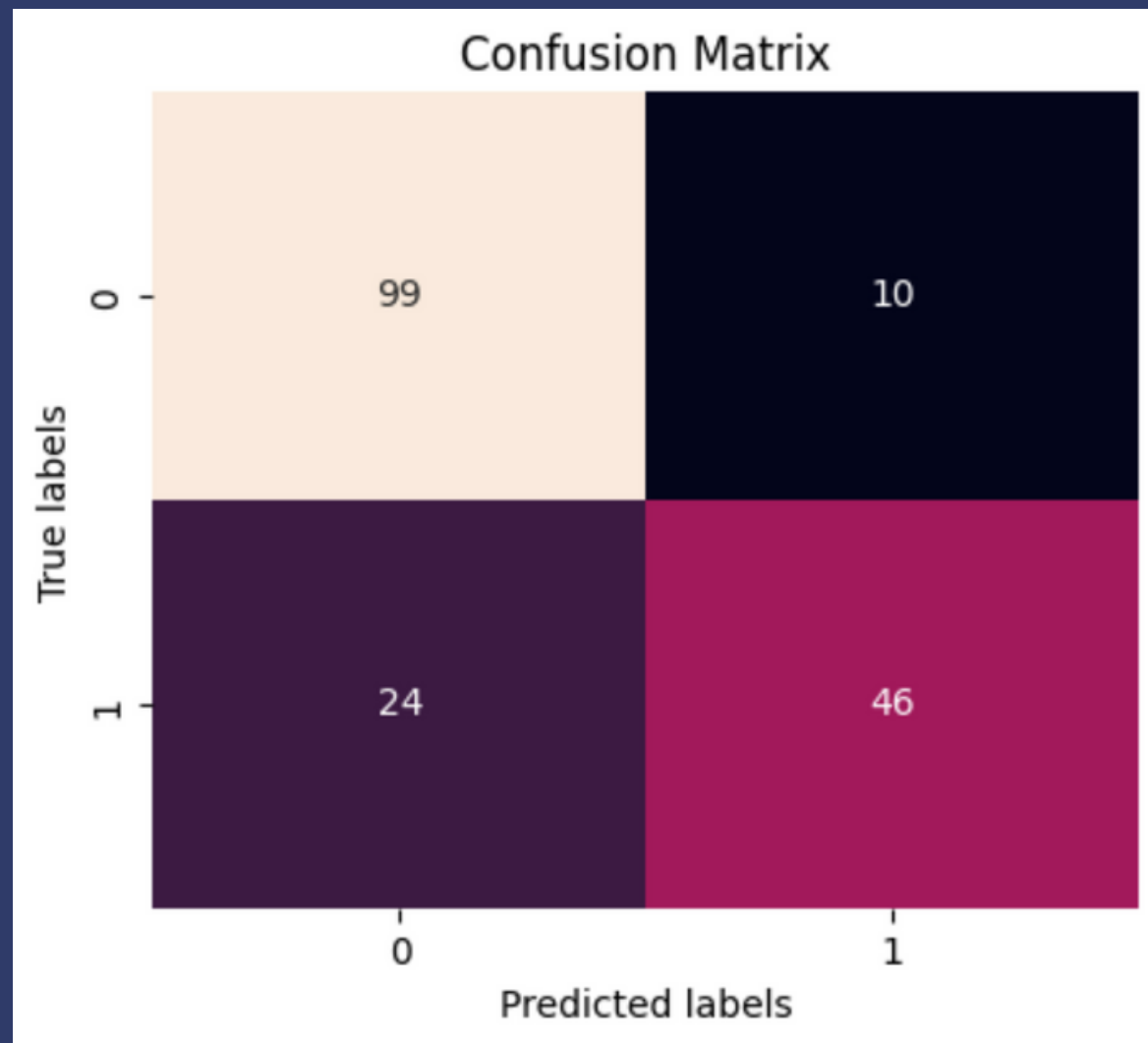
Resultados

- **Confusion Matrices**
- **ROC AUC Curves**
- **Kaggle Test**

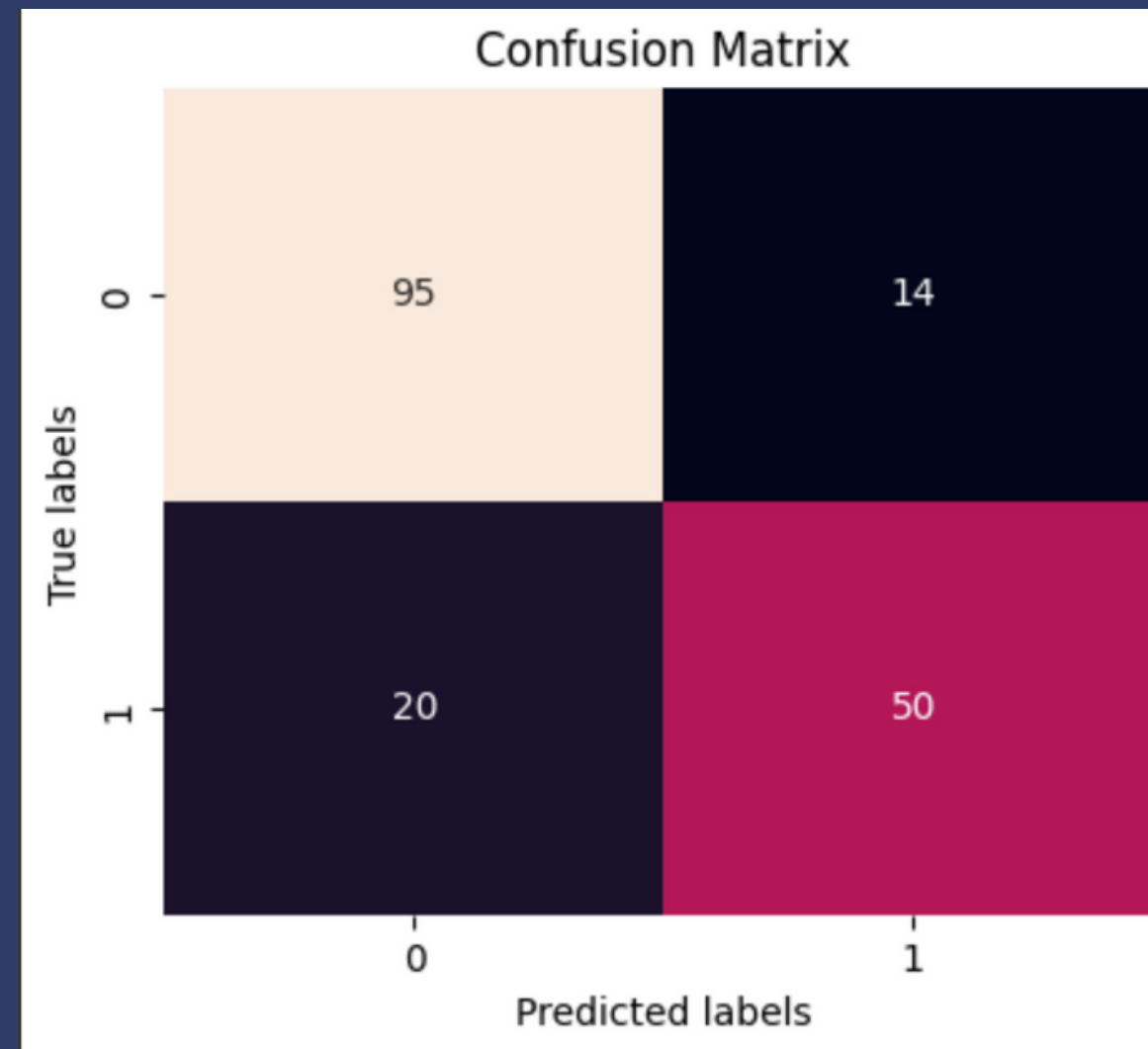


Matrices de confusión

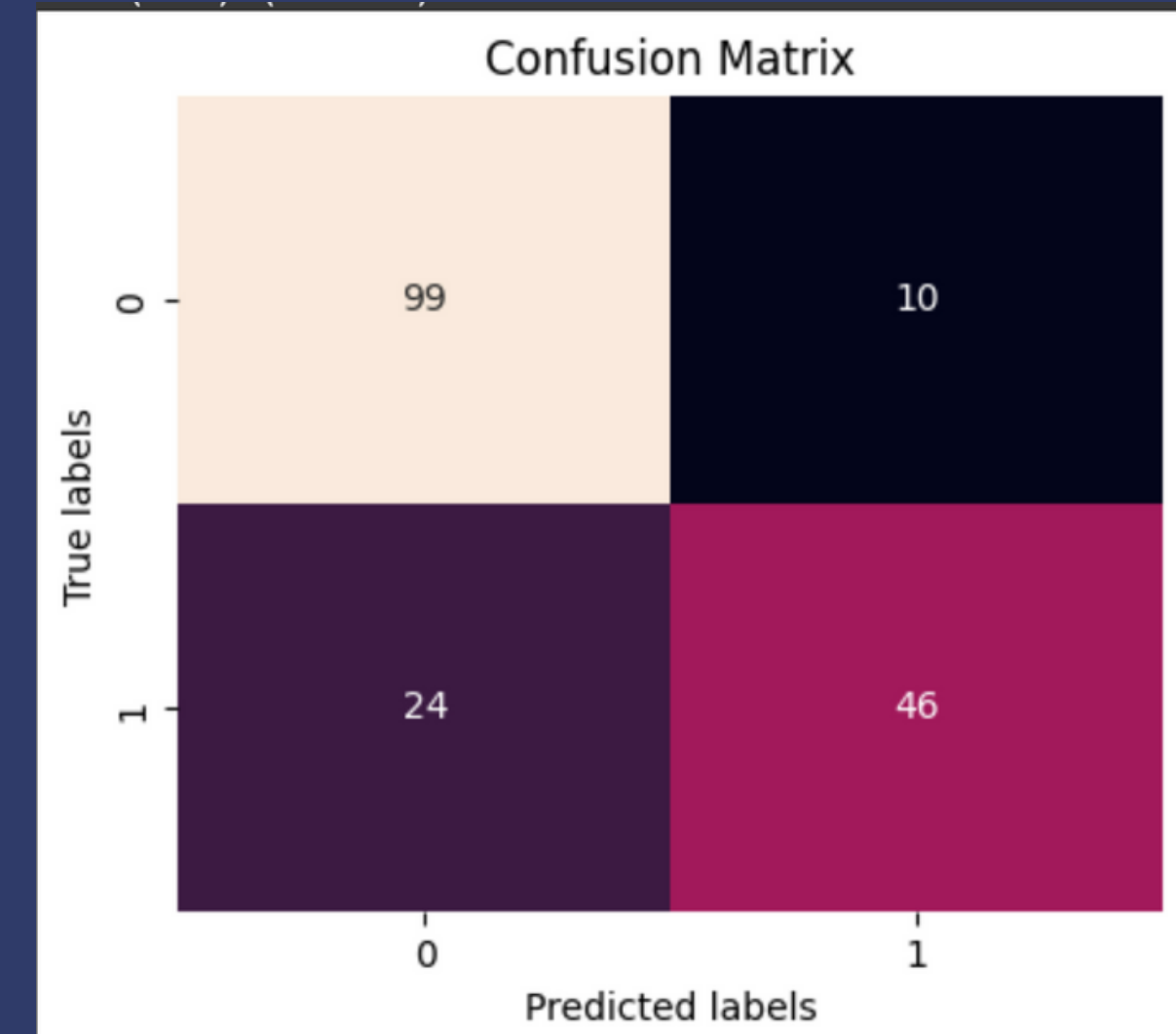
Random Forest



SVM Kernel

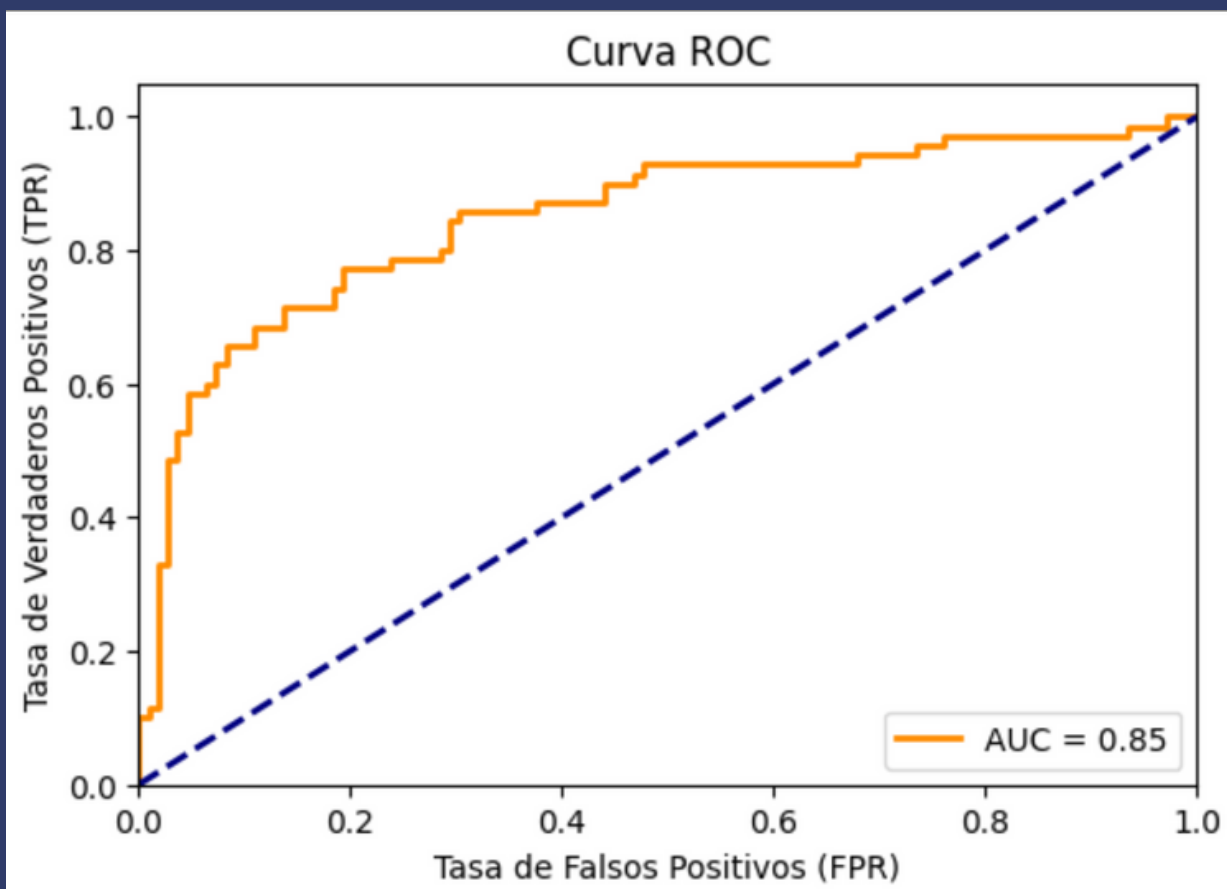


KNN

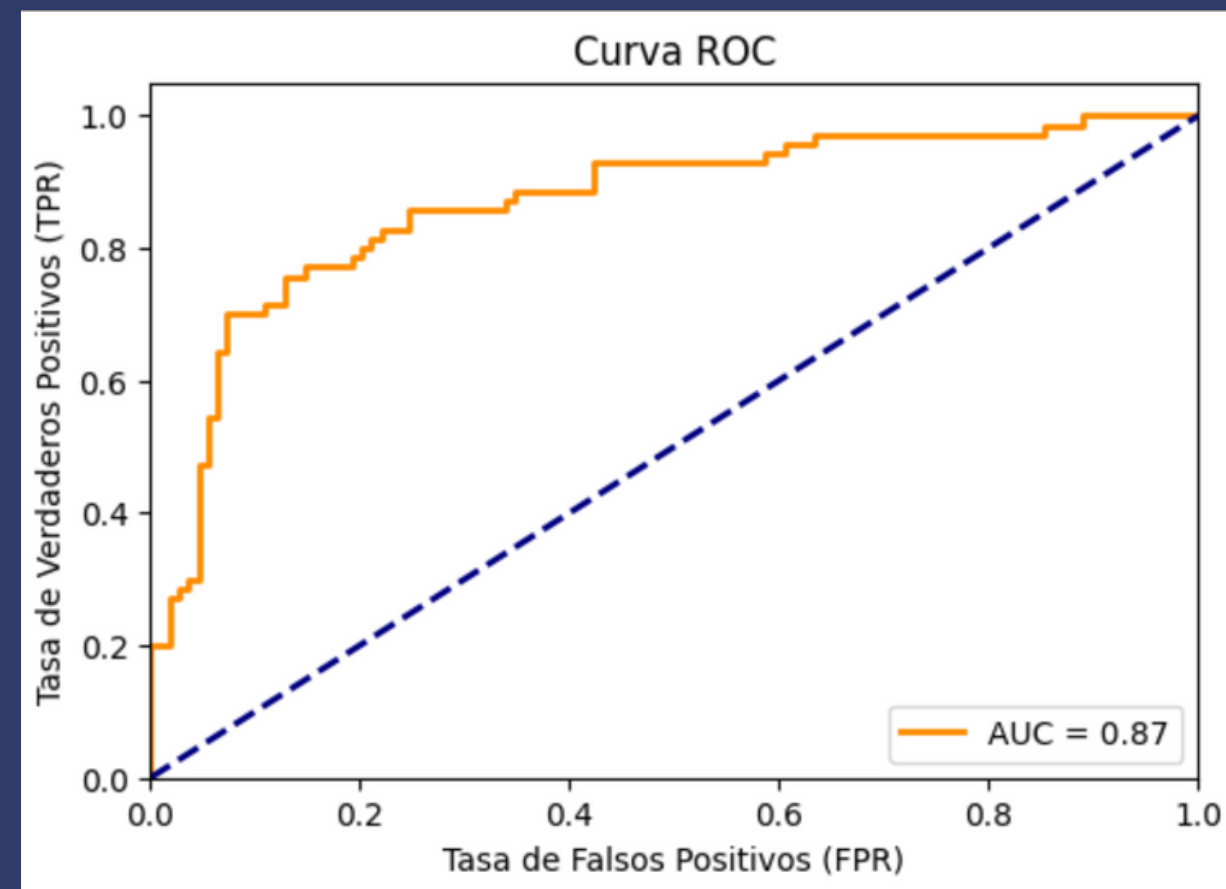


Curvas ROC

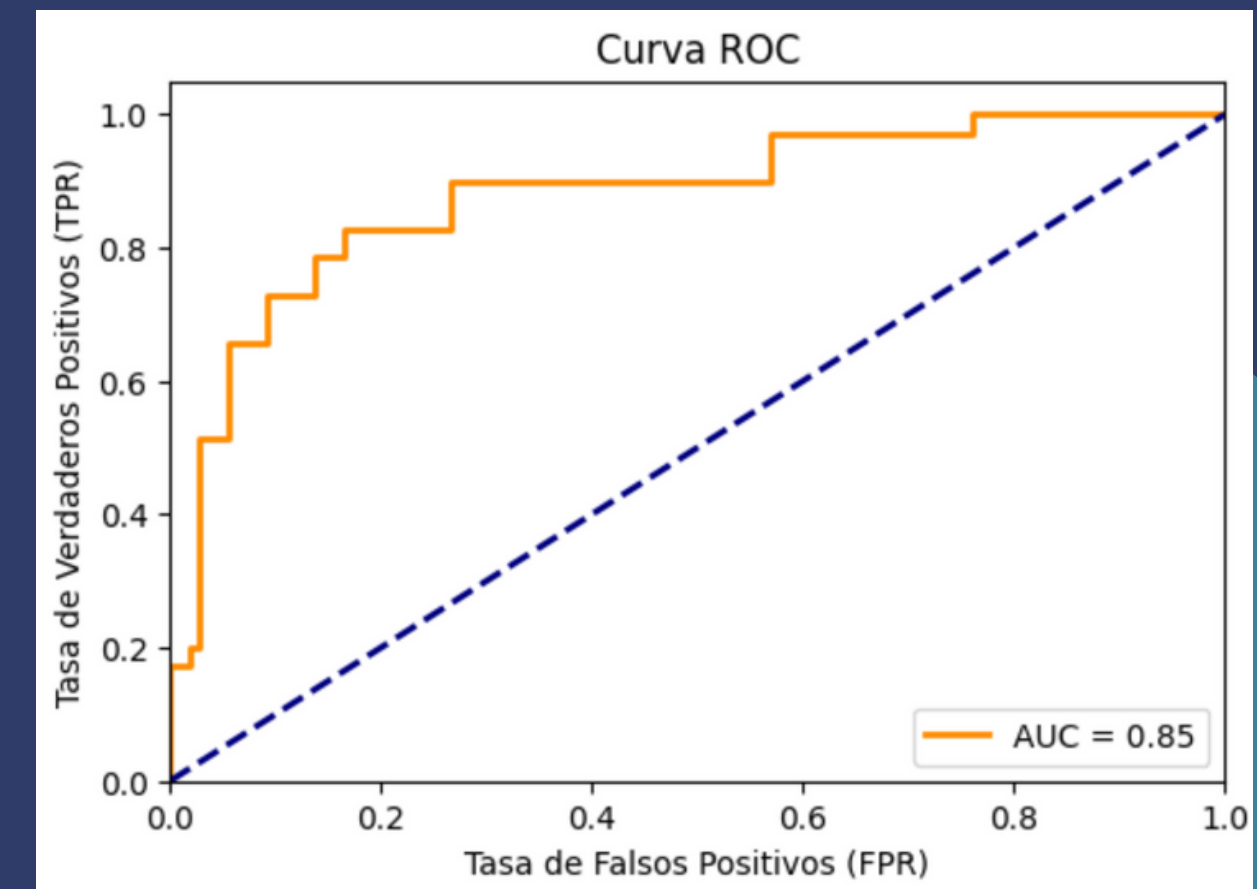
Random Forest



SVM Kernel



KNN



Test Kaggle

Random Forest

0.7799

SVM Kernel

0.7751

KNN

0.7703

**Por su atención,
gracias**

