

El Reto del Titanic

Kevin Horacio Molina Peñuñuri A01253141

Santiago Juárez Roaro A01705439

David Alejandro Acuña Orozco A00571187

Iván Alejandro López Valenzuela A01284875

Joaquin Sainz Muleiro A01783801

Un Caso de Estudio en Supervivencia

El hundimiento del RMS Titanic en 1912 se ha convertido en un clásico desafío para la ciencia de datos. Nuestro objetivo es analizar a fondo los factores que influyeron en la supervivencia de los pasajeros.

A través de este proyecto, exploraremos cómo variables como la clase de pasaje, la edad y el género desempeñaron un papel crucial en el desenlace de esta histórica tragedia marítima, buscando patrones predictivos en los datos.



Hipótesis Clave

01.

Clase social

02.

**¿El criterio
“mujeres y
niños primero”
fue siempre
aplicado?**

03.

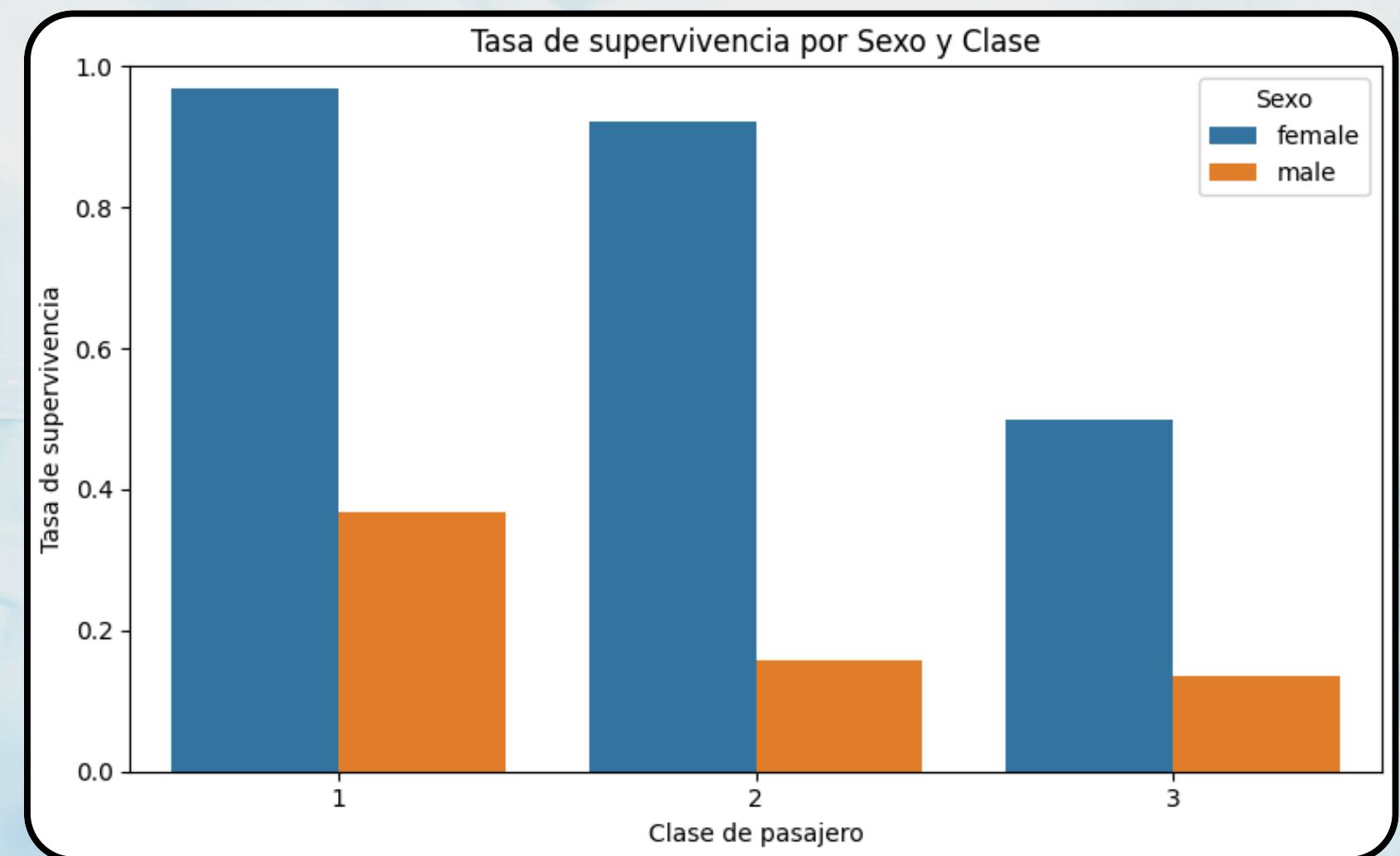
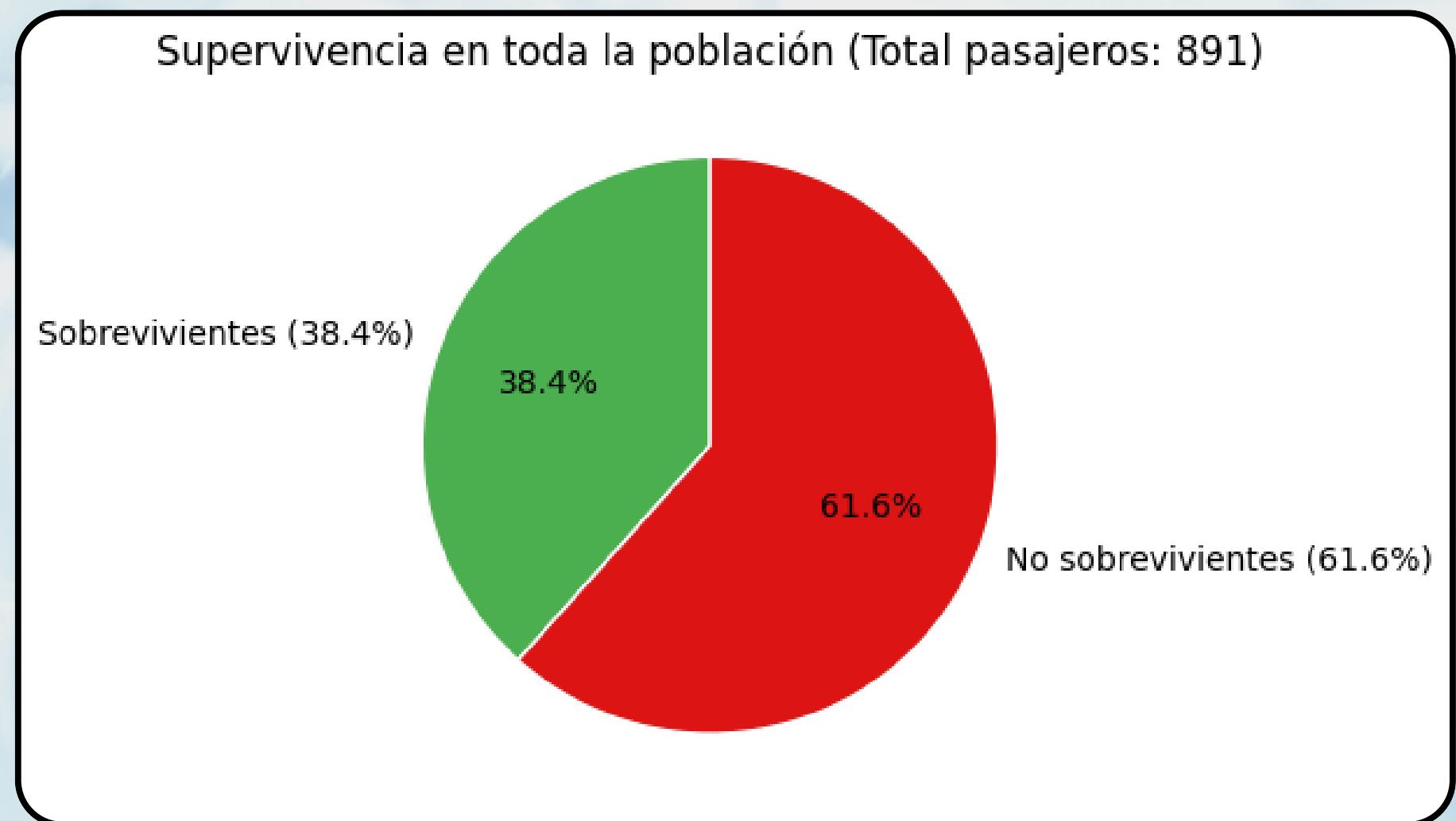
**Hombres con
familia**

04.

**¿Que factores
socioeconómico
s determinaron la
supervivencia de
las personas?**

Dataset

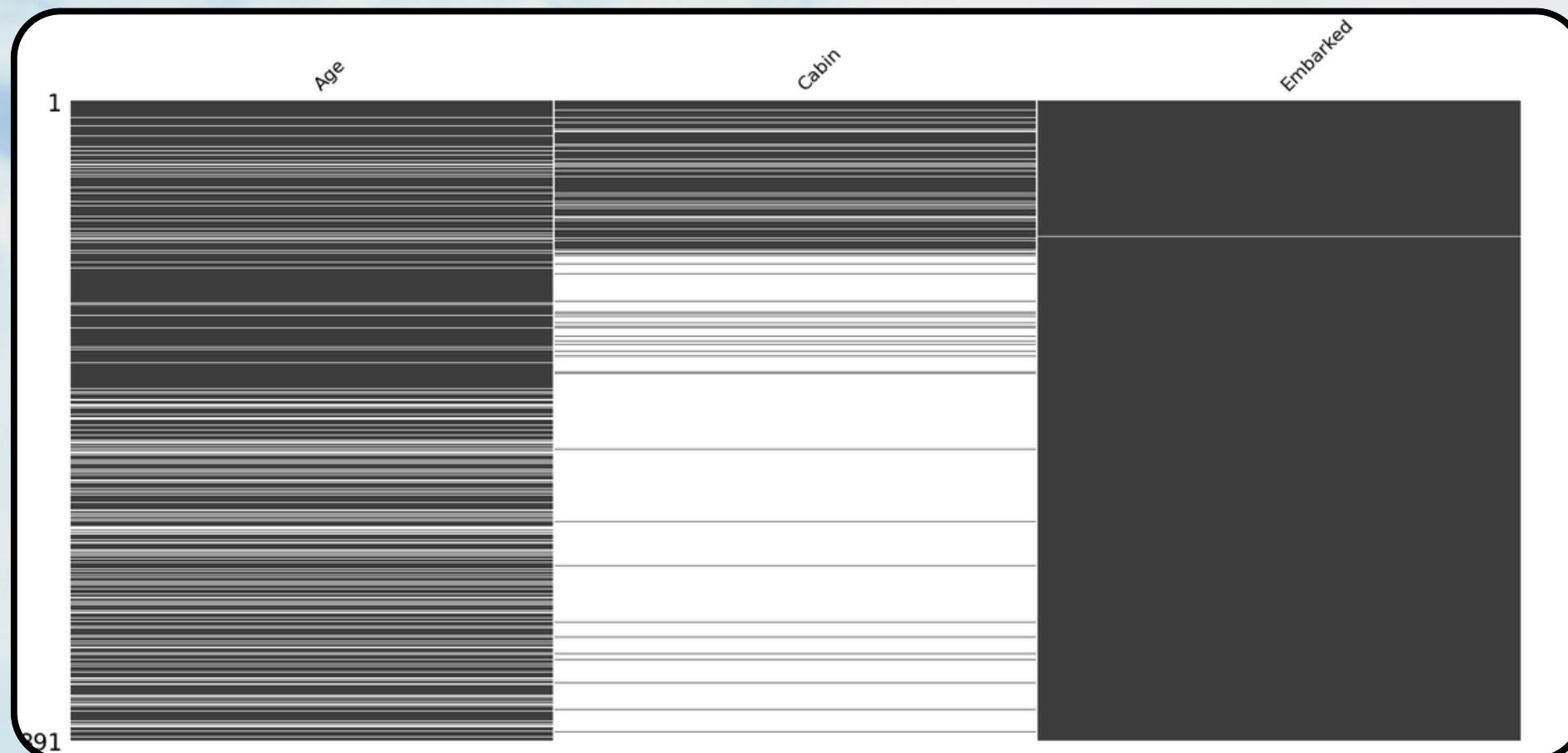
El dataset contiene 891 registros con 12 características, incluyendo datos demográficos y de viaje.



Preprocesamiento

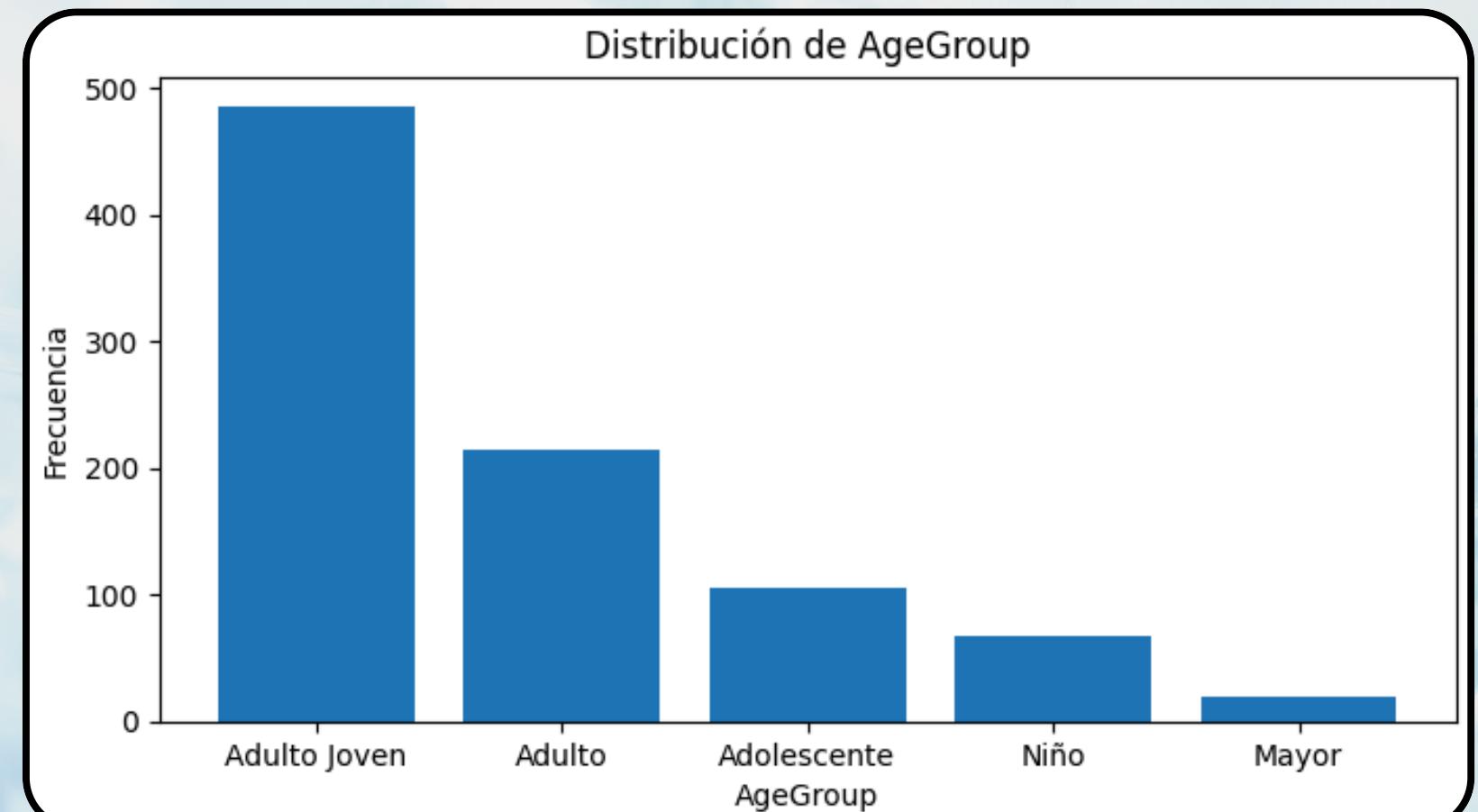
Tratamiento de datos nulos

Se encontraron valores nulos en Cabin (687), Age (177) y Embarked (2).



Feature Engineering

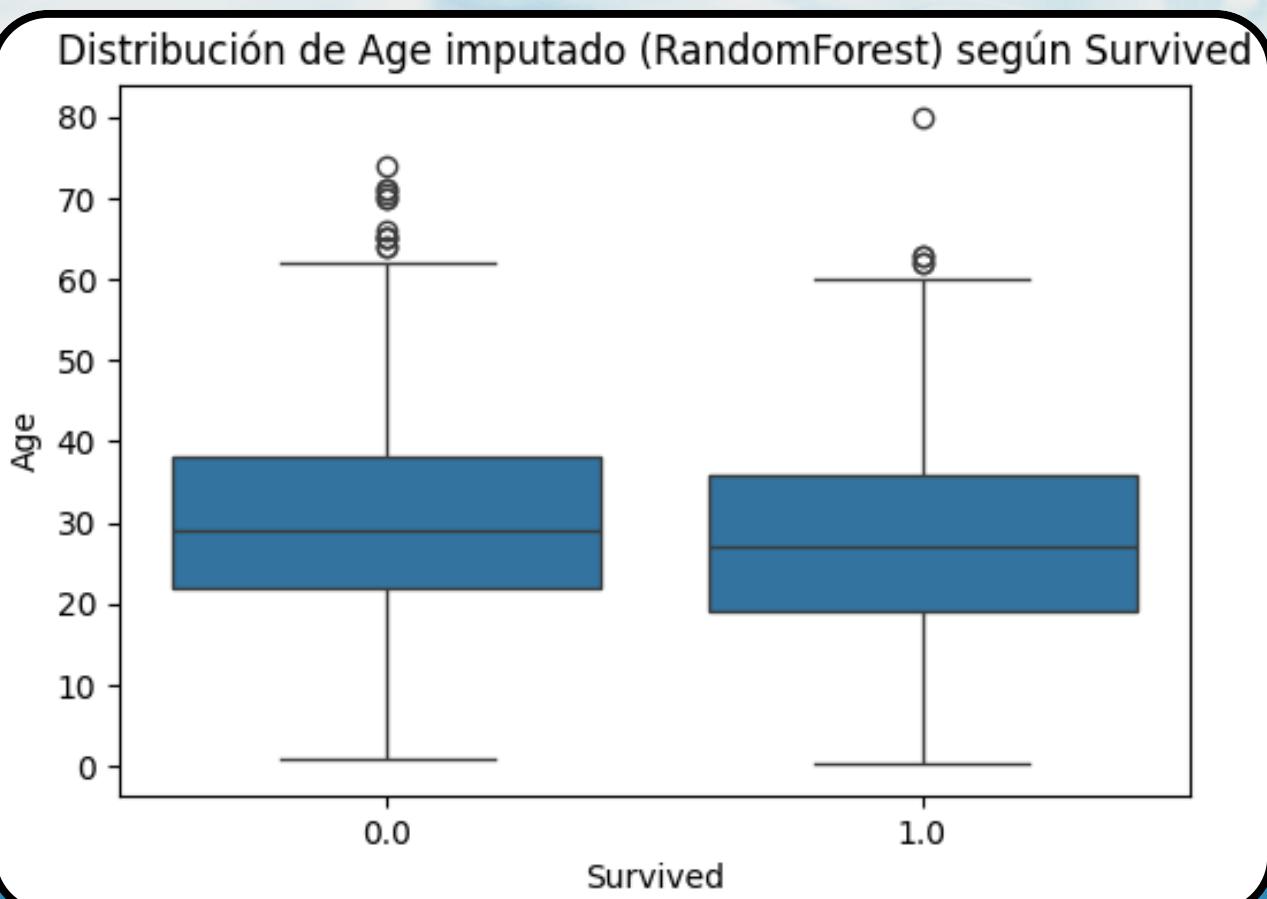
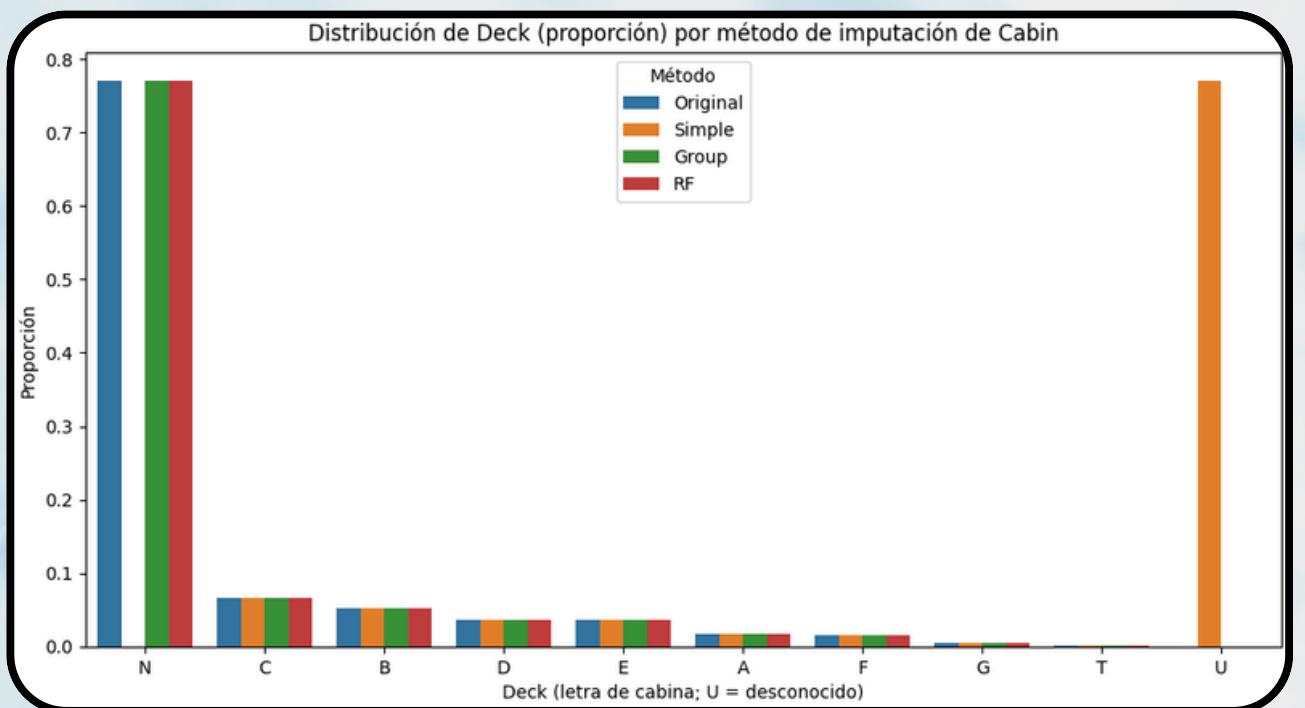
Creación de nuevas variables dependientes de las originales.



Escalado y codificación

Preprocesamiento

Tratamiento de datos faltantes



Feature Engineering

Newly Engineered Features

Title	Deck
FamilySize	CabinKnown
IsAlone	TicketFrequen
AgeGroup	NameLength
FarePerPerson	TitleGroup
FareLevel	

Comparación de Datasets: Original vs. Transformado

Dataset Original

Datos brutos del Titanic tal como se obtuvieron. Shape: (891, 12)

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. C male	male	22	1	0	A/5 21171	7.25	S	
2	1	1	Cumings, Mr. female	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, M female	female	26	0	0	STON/O2. 31	7.925	S	
4	1	1	Futrelle, Mrs. female	female	35	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr. Wil male	male	35	0	0	373450	8.05	S	
6	0	3	Moran, Mr. Ja male	male		0	0	330877	8.4583	Q	
7	0	1	McCarthy, Mi male	male	54	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Mas male	male	2	3	1	349909	21.075	S	
9	1	3	Johnson, Mrs female	female	27	0	2	347742	11.1333	S	
10	1	2	Nasser, Mrs. female	female	14	1	0	237736	30.0708	C	
11	1	3	Sandstrom, N female	female	4	1	1	PP 9549	16.7	G6	S
12	0	1	Bonnell, Miss female	female	58	0	0	113783	26.55	C103	S
13	0	3	Saundercock male	male	20	0	0	A/5. 2151	8.05	S	
14	0	3	Andersson, M male	male	39	1	5	347082	31.275	S	
15	0	3	Vestrom, Mis female	female	14	0	0	350406	7.8542	S	
16	1	2	Hewlett, Mrs. female	female	55	0	0	248706	16	S	
17	0	3	Rice, Master. male	male	2	4	1	382652	29.125	Q	
18	1	2	Williams, Mr. male	male		0	0	244373	-		
19	0	3	Vander Plank female	female		0	0	345763	0.18836581	S	
20	1	3	Masselmanni, female	female		0	0	2649	-0.9329479	C	
21	0	2	Fynney, Mr. J male	male		0	0	239865	0.395659174	S	
22	1	2	Beesley, Mr. male	male		0	0	248706	-0.1031598	S	
23	1	3	McGowan, M female	female		0	0	328652	0.510479533	Q	
24	1	2	Williams, Mr. male	male		0	0	244373	-0.319323	S	
25	0	3	Vander Plank female	female		0	0	345763	0.018836581	S	
26	1	3	Asplund, Mrs female	female		0	0	347077	0.585729852	S	
27	0	3	Emir, Mr. Far male	male		0	0	2631	-0.9329479	C	
28	0	1	Fortune, Mr. male	male		0	0	19950	0.254989826	S	
29	1	3	O'Dwyer, Mis female	female		0	0	330959	-0.8426743	Q	
30	0	3	Todoroff, Mr. male	male		0	0	349216	-0.8404796	S	
31	0	3	Vander Plank female	female	31	1	0	345763	18	S	
32	1	3	Masselmani, female	female		0	0	2649	7.225	C	

Dataset Transformado

Es la versión preprocesada, lista para el modelado. Incluye la imputación de valores faltantes y nuevas características creadas y transformadas. Shape: (891, 63)

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	Title	FamilySize	Isalone	AgeGroup	PclassFirstDeck	CabinUnknown	TicketLength	TitleOrName		
1	0	3	Braund, Mr. C male	male	-0.5348708	1.373636424	0	A/5 21171	-0.9293552	S	Mr	0.836696417	0	Adulto Joven	-1.4014375	UK	0	-0.7758032	-0.2704553	Common
2	1	1	Cumings, Mr. female	female	0.728472495	1.373636424	0	PC 17599	1.385216171	C	Mrs	0.836696417	0	Adulto	1.287873425	C	1	-0.7758032	1.938388173	Common
3	1	3	Heikkinen, M female	female	-0.8736294	-0.67985	0	STON/O2. 31	-0.8366299	S	Miss	-0.7910075	1	Adolescente	-0.4483938	UK	0	-0.7758032	-0.4280325	Common
4	1	1	Futrelle, Mrs. female	female	0.500368464	1.373636424	0	113803	1.103871987	S	Mrs	0.836696417	0	Adulto Joven	0.968346572	C	1	0.94268399	1.606494014	Common
5	0	3	Allen, Mr. Wil male	male	0.500368464	-0.67985	0	373450	-0.8203051	S	Mr	-0.7910075	1	Adulto Joven	-0.4292994	UK	0	-0.7758032	-0.1234285	Common
6	0	0	Moran, Mr. Je male	male	-0.1263903	-0.67985	0	330877	-0.7686572	Q	Mr	-0.7910075	1	Adulto Joven	-0.3689999	UK	0	-0.7758032	-1.684699	Common
7	0	1	McCarthy, M male	male	1.896316898	-0.67985	0	17463	1.081058388	S	Mr	-0.7910075	1	Adulto	1.677557287	E	1	-0.7758032	-0.2704553	Common
8	0	3	Palsson, Mas male	male	-2.2023489	1.718889313	0	349909	0.181282028	S	Master	1.667511506	0	Niño+	-1.2197643	UK	0	1.616112446	0.588938996	Common
9	1	3	Johnson, Mrs female	female	-0.1263903	-0.67985	0	347742	-0.481295	S	Mrs	1.343027306	0	Adulto Joven	-1.3732986	UK	0	1.425479839	1.851473293	Common
10	1	2	Nasser, Mrs. female	female	-1.2259887	1.373636424	0	237736	0.542667217	C	Mrs	0.836696417	0	Adolescente	0.320114405	UK	0	0.94268399	1.027604445	Common
11	1	3	Sandstrom, N female	female	-2.2023489	1.373636424	0	PP 9549	-0.0587476	S	Miss	1.343027306	0	Niño+	-0.8807194	G	1	0.94268399	0.685744184	Common
12	1	1	Bonnell, Miss female	female	0.11210216	-0.67985	0	113783	0.416893540	S	Miss	-0.7910075	1	Adulto Joven	0.968346572	C	1	-0.7758032	-0.1234289	Common
13	0	3	Saundercock male	male	-0.7027518	-0.67985	0	A/5. 2151	-0.8203051	S	Mr	-0.7910075	1	Adulto Joven	-0.4292994	UK	0	-0.7758032	0.588938996	Common
14	0	3	Andersson, M male	male	0.803774038	1.373636424	0	347082	0.582126285	S	Mr	1.768449015	0	Adulto	-1.1490798	UK	0	1.786725032	0.264503503	Common
15	0	3	Vestrom, Mis female	female	-1.2259887	-0.67985	0	350406	-0.845988	S	Miss	-0.7910075	1	Adolescente	-0.4593472	UK	0	-0.7758032	1.103377131	Common
16	1	2	Hewlett, Mrs female	female	1.967015690	-0.67985	0	248706	-0.1031598	S	Mrs	-0.7910075	1	Adulto	0.392830643	UK	0	-0.7758032	0.777647263	Common
17	0	3	Rice, Master male	male	-2.2023489	1.749364363	0	328652	0.510479533	Q	Master	1.729853725	0	Niño+	-1.0481092	UK	0	1.707495022	-0.7801287	Common
18	1	2	Williams, Mr. male	male	0.190631666	-0.67985	0	244373	-0.319323	S	Mr	-0.7910075	1	Adulto Joven	0.148475149	UK	0	-0.7758032	0.378837313	Common
19	0	3	Vander Plank female	female	0.190631666	1.373636424	0	345763	0.0188365											



Diseño Experimental

01. Preprocesamiento

02. División train
test

03. Grid Search /
StratifiedKFold

04. Test

05. Métricas

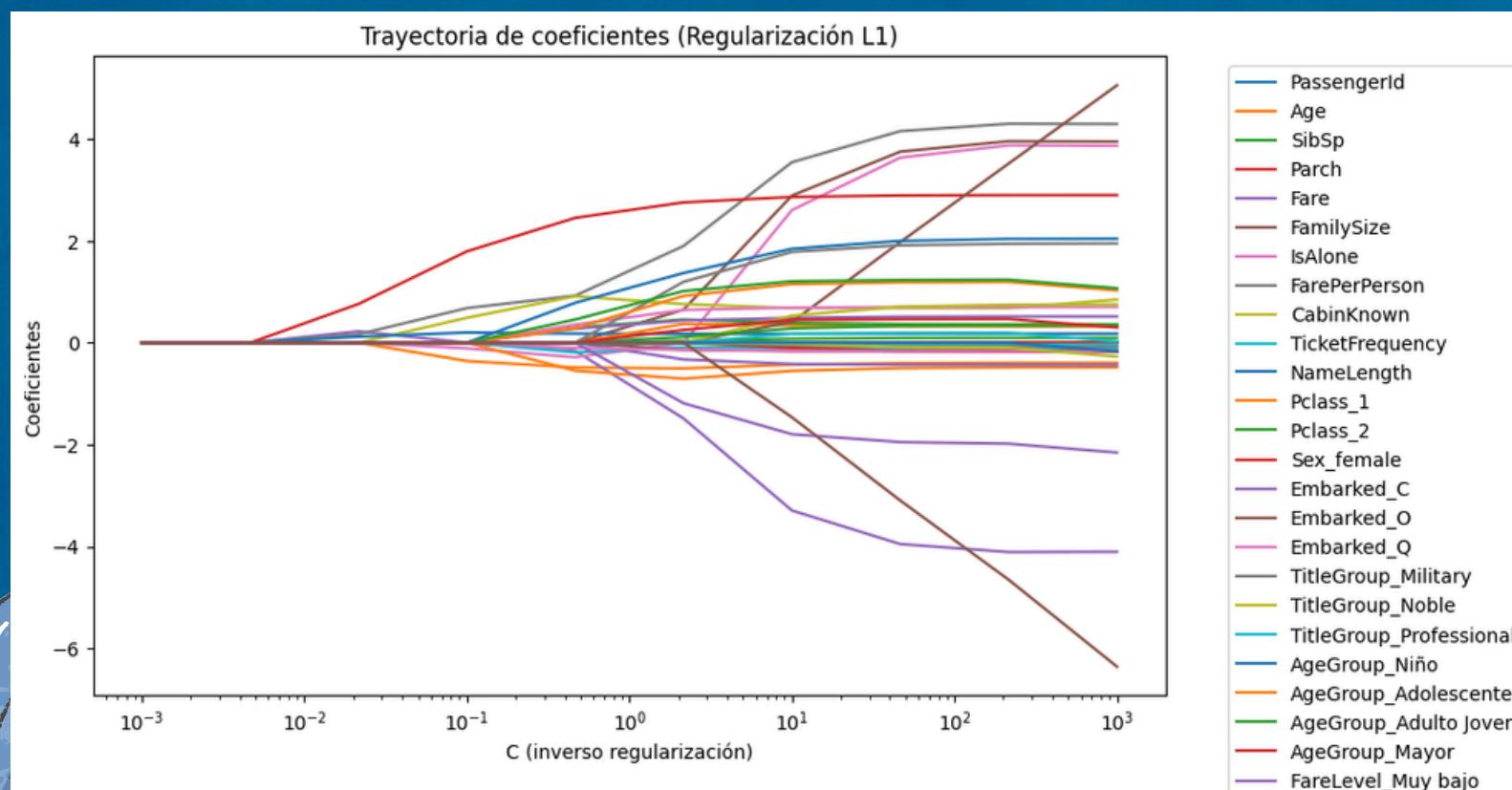


Modelos

Regresión Logística

- F1 : 0.808
- ROC-AUC: 0.864

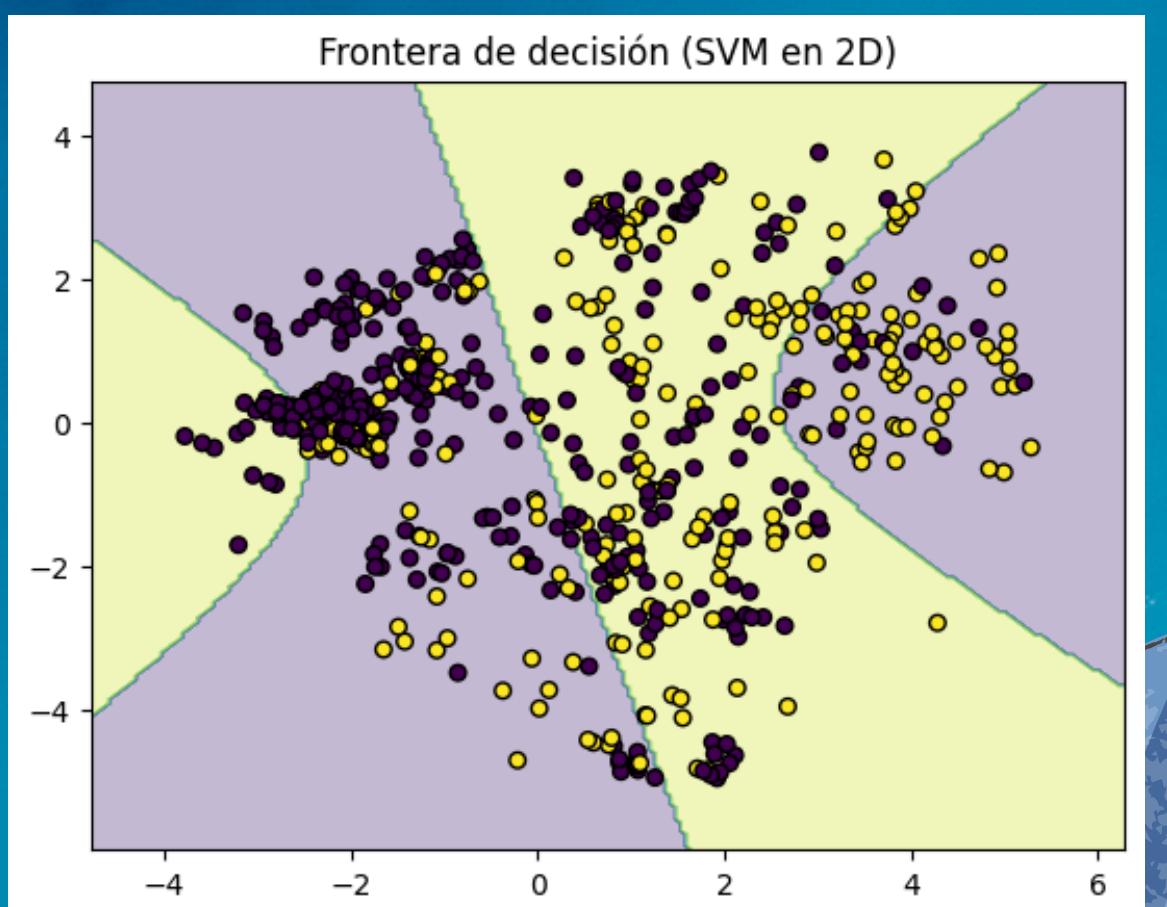
max iter 1000, l1, liblinear



SVM

- F1 : 0.802
- ROC-AUC: 0.746

degree 2, gamma scale, kernel rbf



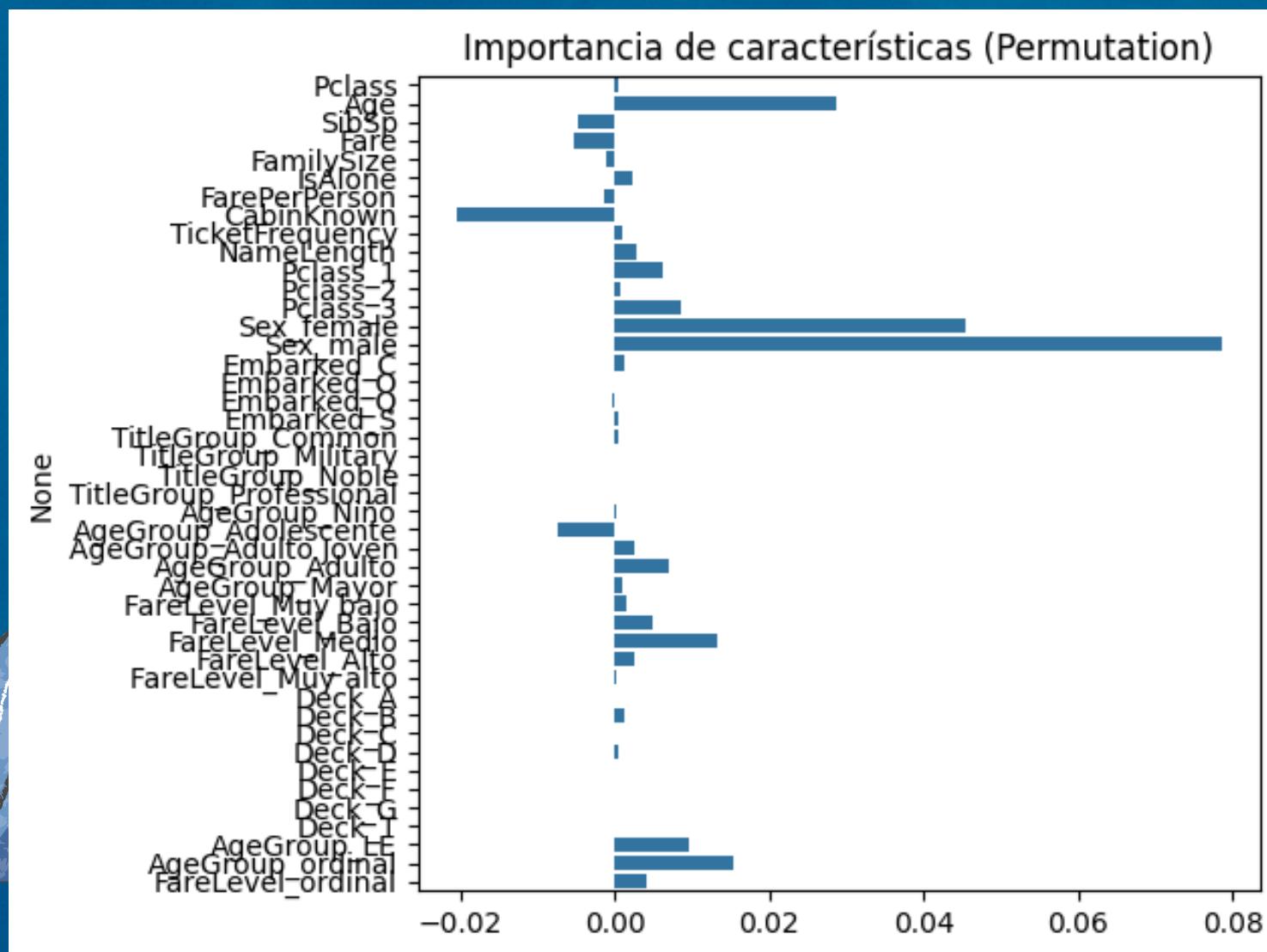


Modelos

Random Forest

- F1 : 0.812
- ROC-AUC: 0.8621

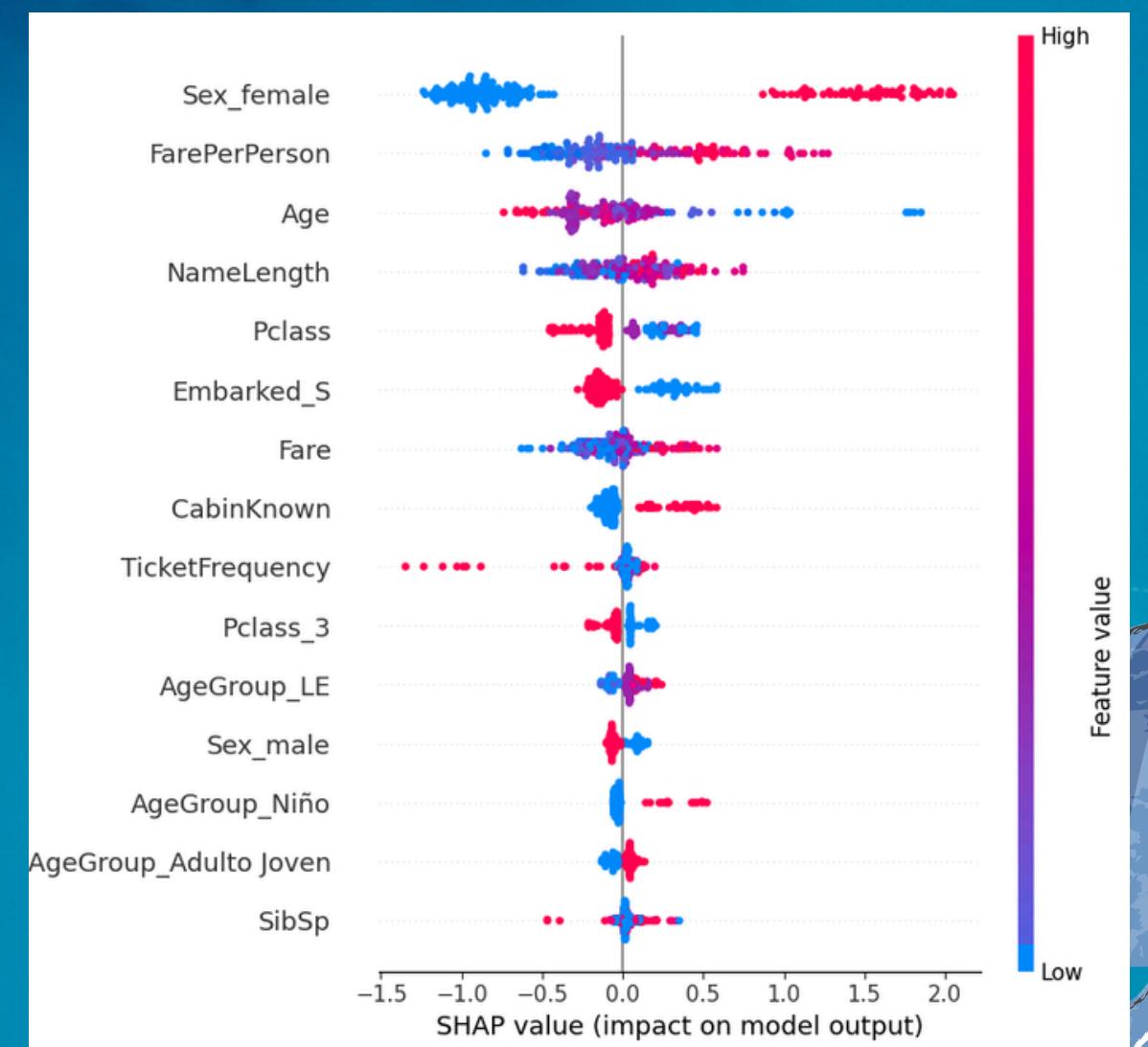
max depth 20, n estimators 200, min samples 10



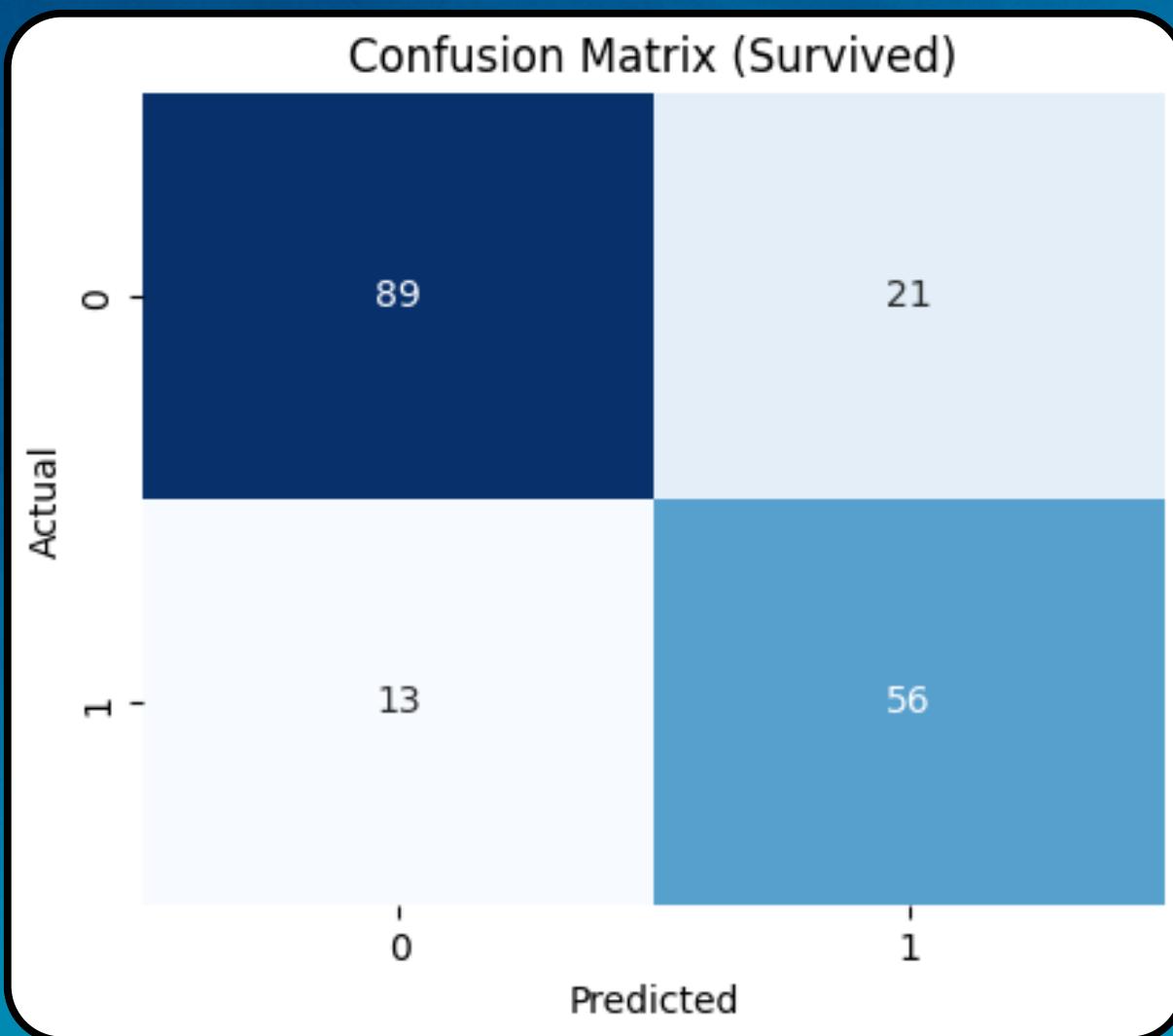
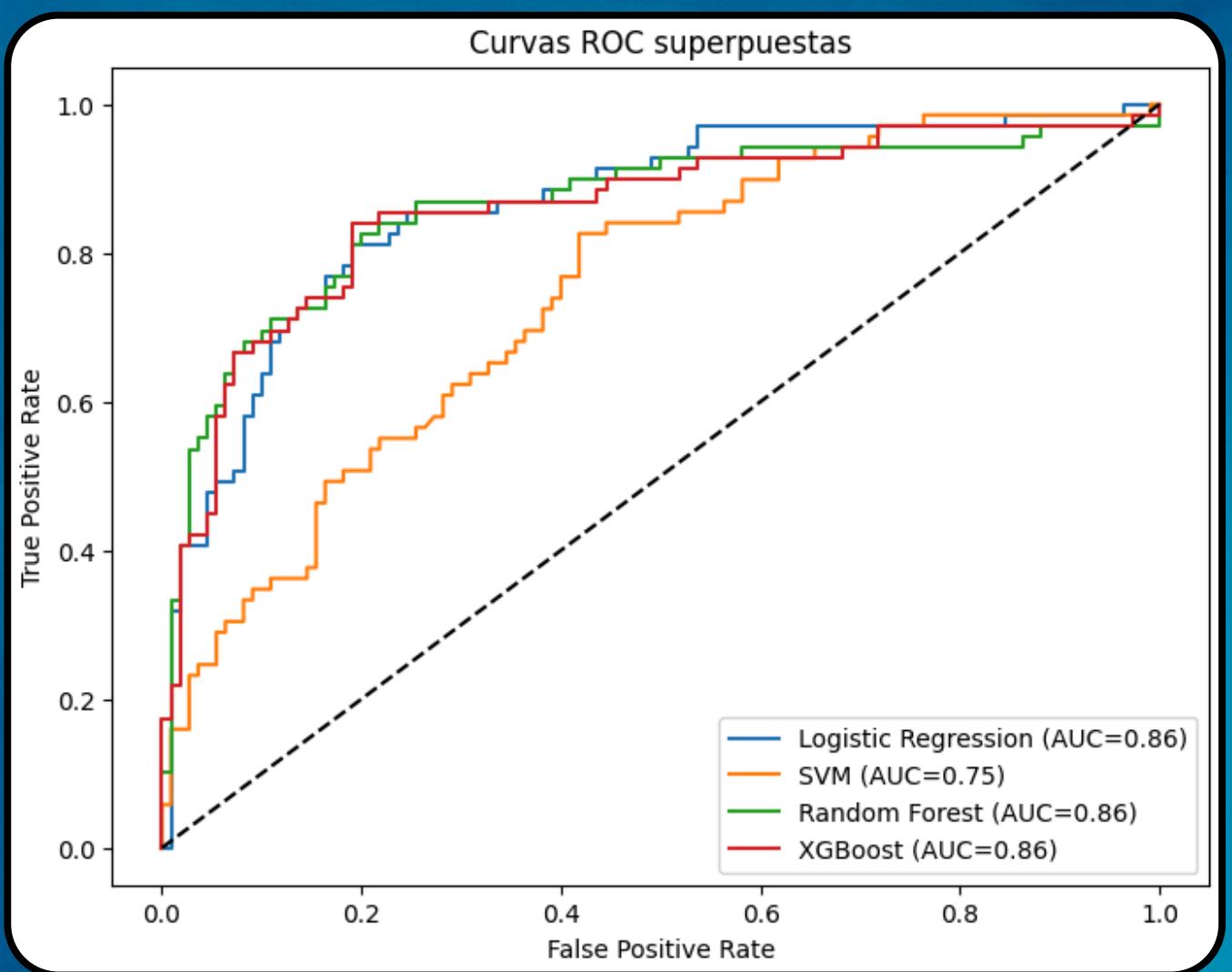
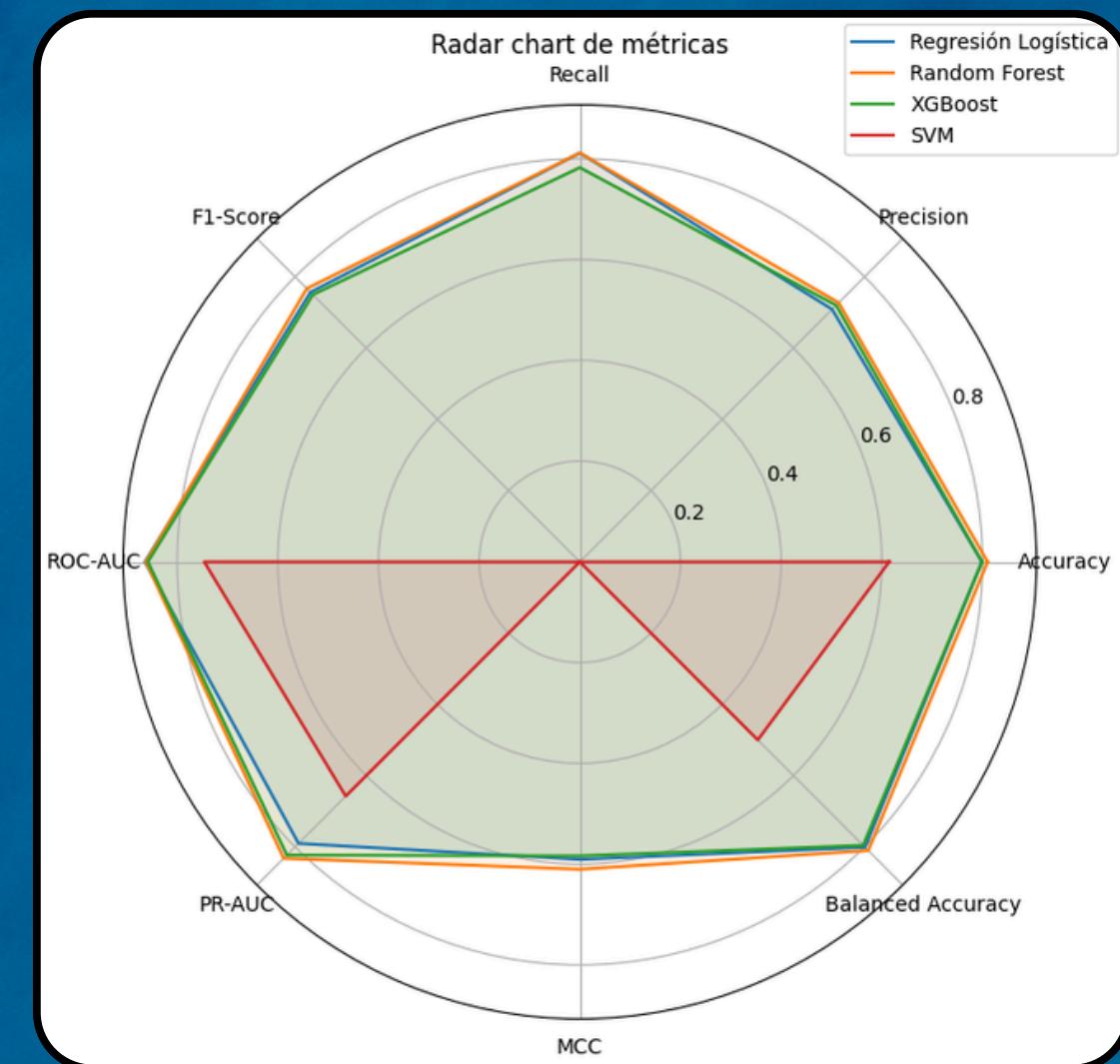
XGBoost

- F1 : 0.800
- ROC-AUC: 0.8534

gamma 0.5, n estimators 200, lr 0.05

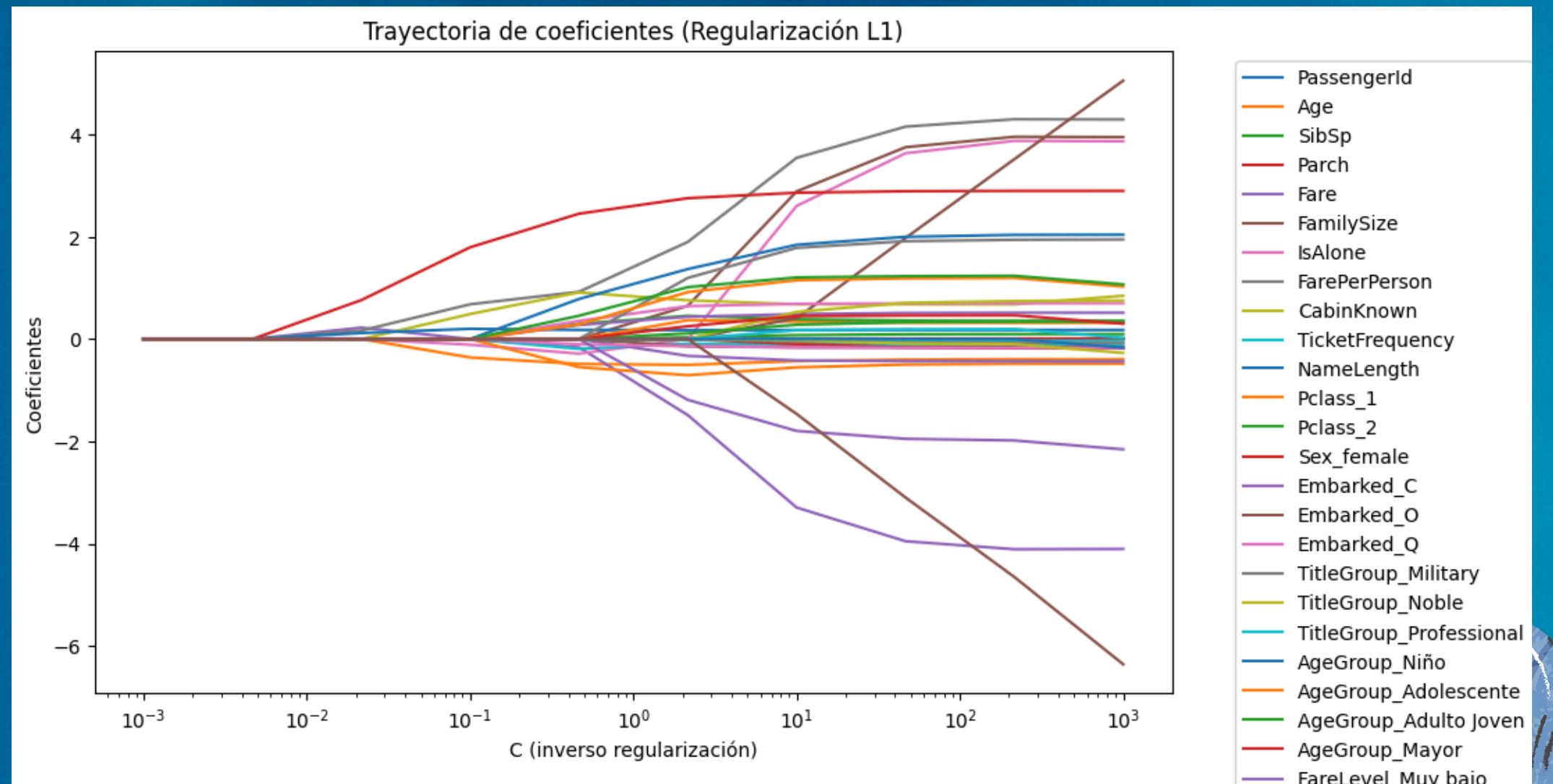
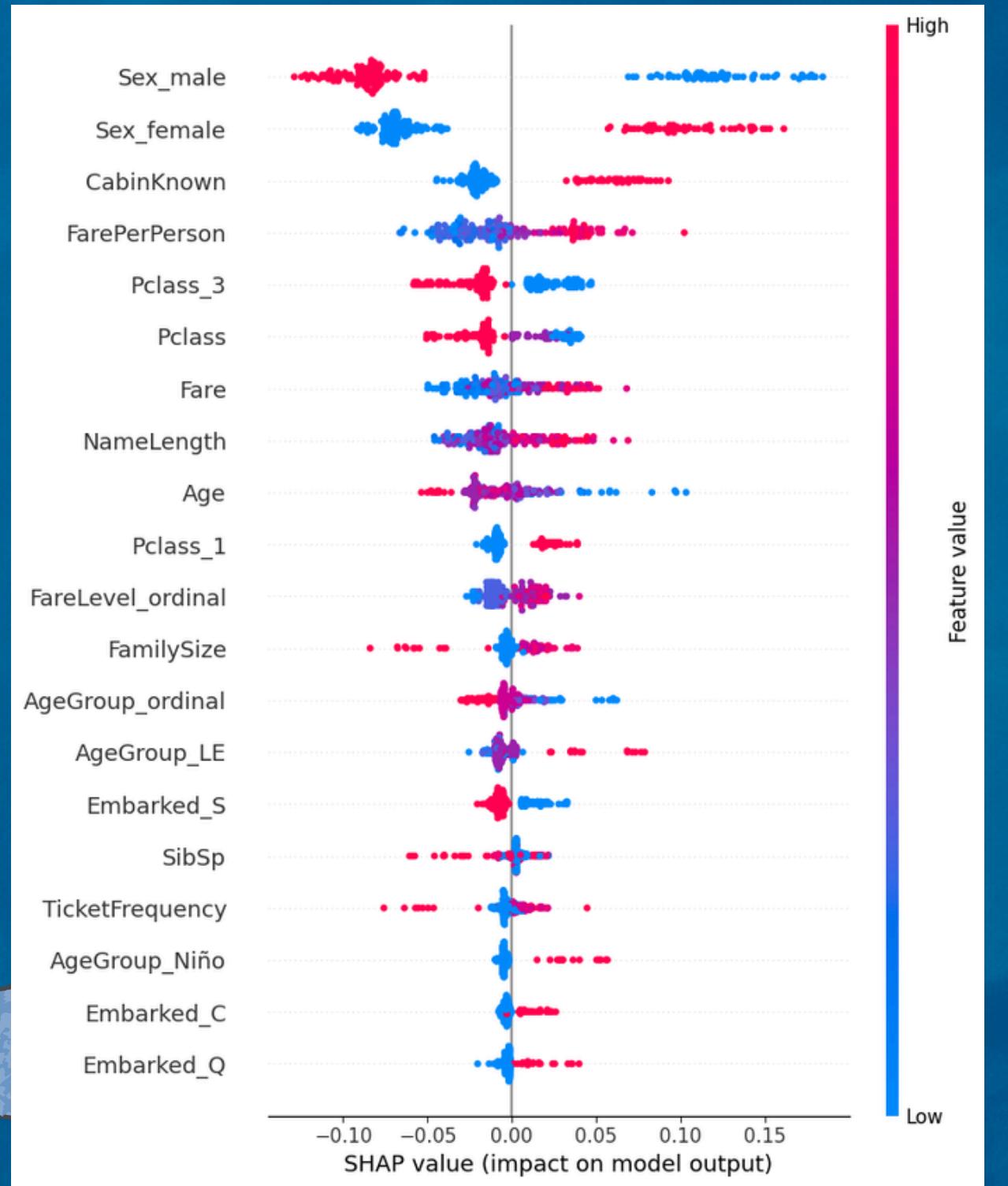


Comparación de modelos



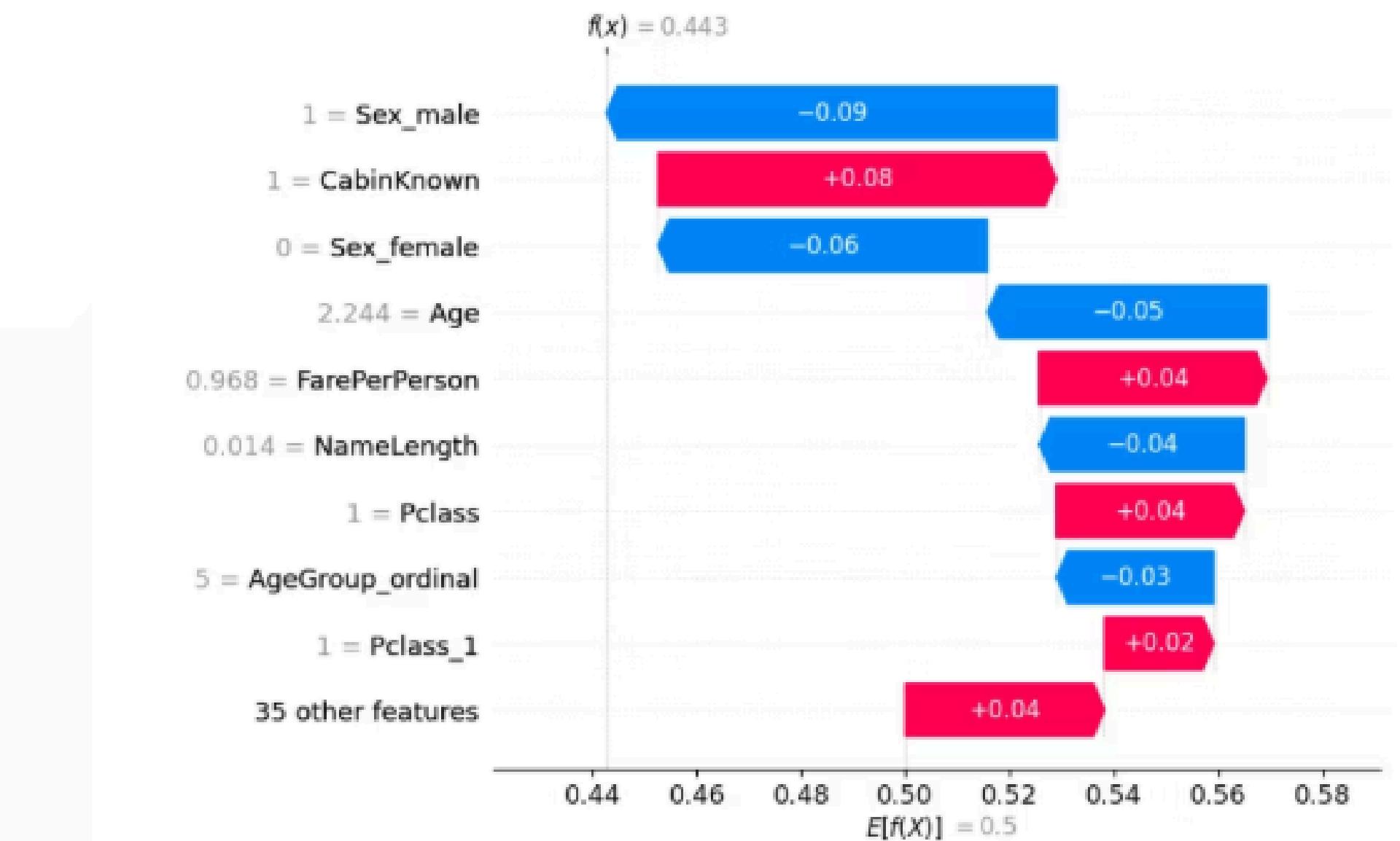


Hallazgos



Interpretabilidad del Modelo

Comprender cómo y por qué un modelo llega a sus predicciones es tan crucial como su precisión.



Hallazgos Principales e Hipótesis

El análisis del caso Titanic revela patrones clave sobre la supervivencia y valida nuestras hipótesis:



Clase Social y Supervivencia

Fuertemente soportada. Mayor supervivencia en 1^a clase (63%) vs. 3^a clase (24%).



"Mujeres y Niños Primero"

Parcialmente soportada. Clase social y ubicación atenuaron la ventaja.



Hombres con Familia

Parcialmente soportada. Hombres con cónyuge o hijos mostraron ligera mayor supervivencia.

Los modelos confirman estos patrones de clase, género y estructura familiar, resaltando el dilema entre datos históricos y equidad moderna.

Ética del Modelo

¿Vida o muerte por un algoritmo?



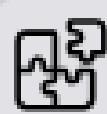
La IA asiste; no decide sobre vidas.



Los algoritmos no son neutrales: heredan sesgos.



Titanic: no predice igual por género.



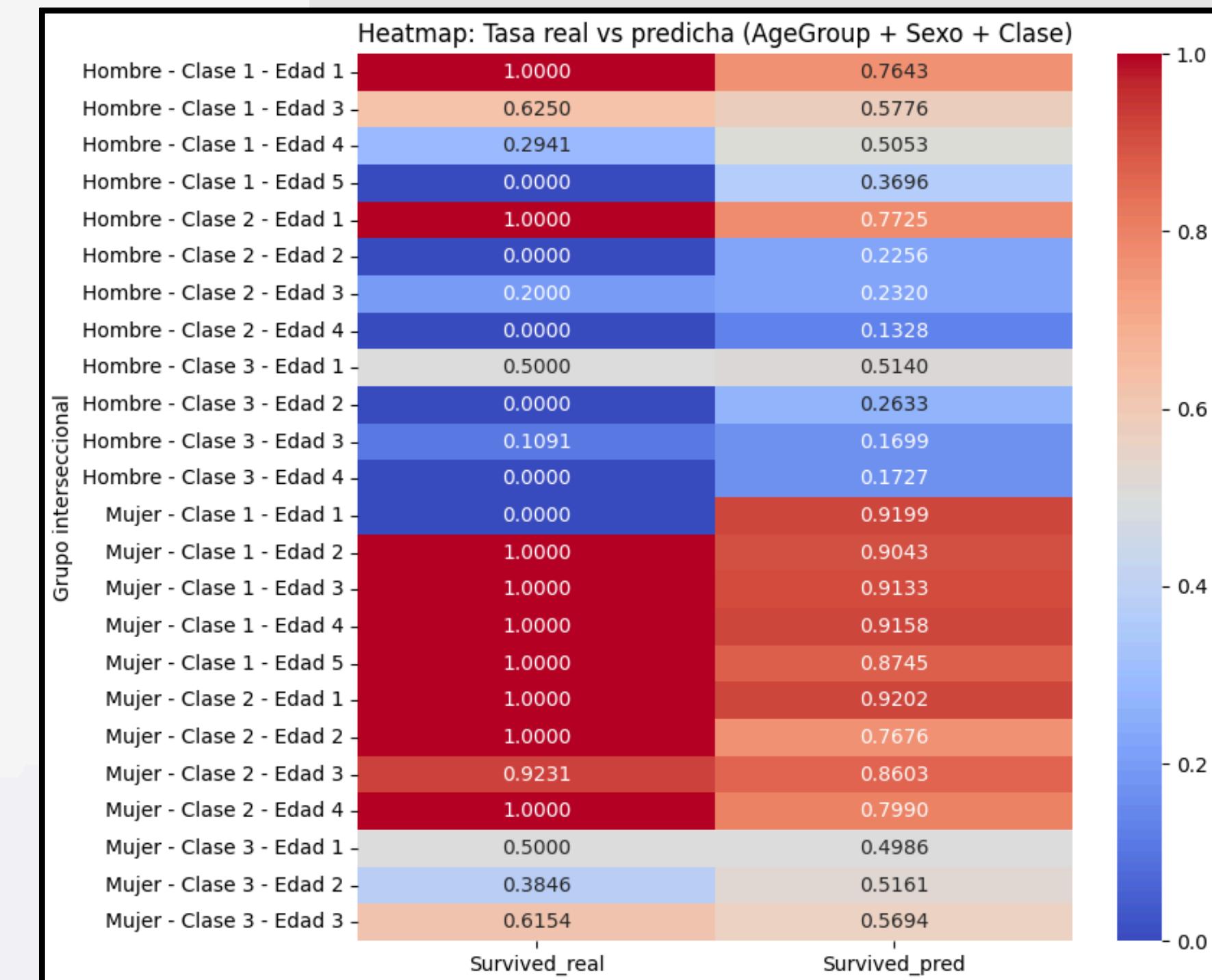
Imposible “igualar todo” sin perder precisión.



Exigir explicabilidad (SHAP/LIME) ≠ causalidad.



Principios: justicia, no discriminación, transparencia, responsabilidad.



Conclusiones



🏆 Contribuciones principales

- XGBoost y Random Forest = mejores métricas.
- Variables clave: Sexo, Pclass, Fare, FamilySize.
- SHAP/LIME: entendemos qué y por qué.

🔮 Trabajo futuro

- Dataset Titanic = caso didáctico, no real.
- No predice igual para todos.
- Aplicar fairness y explicabilidad en contextos reales.

🌟 Mensaje final

- La IA debe asistir, no decidir
- No debe decidir sola
- Usada con responsabilidad, puede construir un futuro más igualitario

Gracias!