

Spaceship Titanic

Proyecto final

Alexei Arenas
Andrés Angel
María José Núñez
Tomás Rojas

Universidad Técnica Federico Santa María

Jueves 12 de Diciembre 2912



Contexto



Figura: Anomalía espacial.

Tabla de Contenidos

- 1 Introducción
 - Contexto
- 2 **Visualización**
 - **Distribución**
- 3 Selección de modelo
- 4 Métricas y resultados
- 5 Conclusiones



Datos Parte I

Los datos de train.csv son los siguientes

PassengerId	HomePlanet	CryoSleep	Cabin	Destination
0001_01	Europa	False	B/0/P	TRAPPIST-1e
0002_01	Earth	False	F/0/S	TRAPPIST-1e
0003_01	Europa	False	A/0/S	TRAPPIST-1e
0003_02	Europa	False	A/0/S	TRAPPIST-1e
0004_01	Earth	False	F/1/S	TRAPPIST-1e



Datos Parte II

Los datos de train.csv son los siguientes

Age	VIP	RoomService	FoodCourt	ShoppingMall
39.0	False	0.0	0.0	0.0
24.0	False	109.0	9.0	25.0
58.0	True	43.0	3576.0	0.0
33.0	False	0.0	1283.0	371.0
16.0	False	303.0	70.0	151.0



Datos Parte III

Los datos de train.csv son los siguientes

Spa	VRDeck	Name	Transported
0.0	0.0	Maham Ofracculy	False
549.0	44.0	Juanna Vines	True
6715.0	49.0	Altark Susent	False
3329.0	193.0	Solam Susent	False
565.0	2.0	Willy Santantines	True



Datos Parte IV

Los datos de test.csv son los siguientes

PassengerId	HomePlanet	CryoSleep	Cabin
0013_01	Earth	True	G/3/S
0018_01	Earth	False	F/4/S
...
9273_01	Europa	False	D/297/P
9277_01	Earth	True	G/1498/S

Esta tabla tiene 4277 filas.



Datos Parte V

Los datos de test.csv son los siguientes

Destination	Age	VIP	RoomService	FoodCourt
TRAPPIST-1e	27.0	False	0.0	0.0
TRAPPIST-1e	19.0	False	0.0	9.0
...
NaN	NaN	False	0.0	2680.0
PSO J318.5-22	NaN	False	0.0	0.0

Esta tabla tiene 4277 filas.



Datos Parte VI

Los datos de test.csv son los siguientes

ShoppingMall	Spa	VRDeck	Name
0.0	0.0	0.0	Nelly Carsoning
0.0	2823.0	0.0	Lerome Peckers
...
0.0	0.0	523.0	Kitakan Conale
0.0	0.0	0.0	Lilace Leonzaley

Esta tabla tiene 4277 filas.



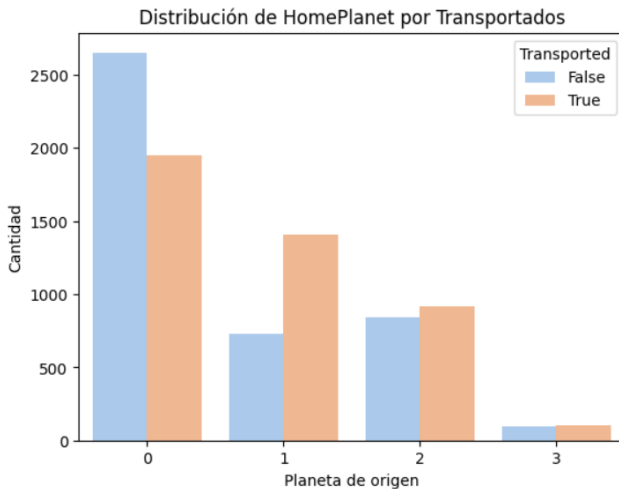


Figura: Distribución respecto al planeta de origen.



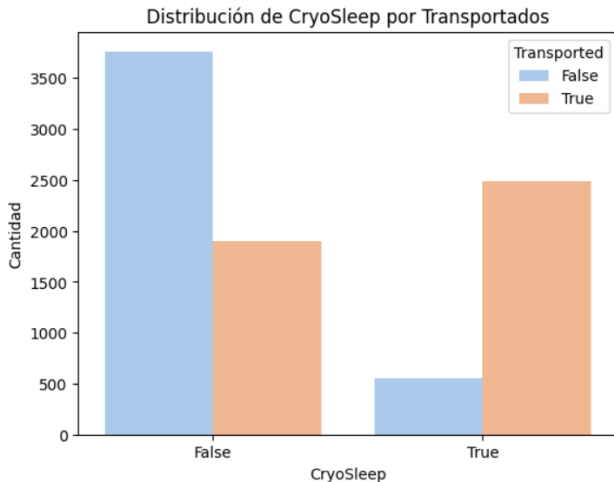


Figura: Distribución respecto si estado de criogenización.



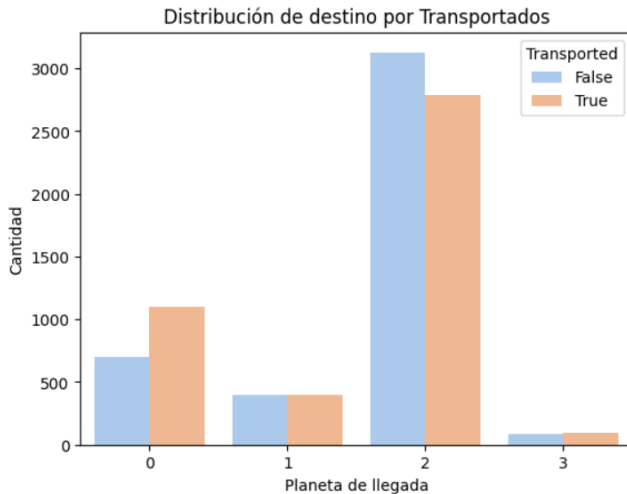


Figura: Distribución respecto al destino



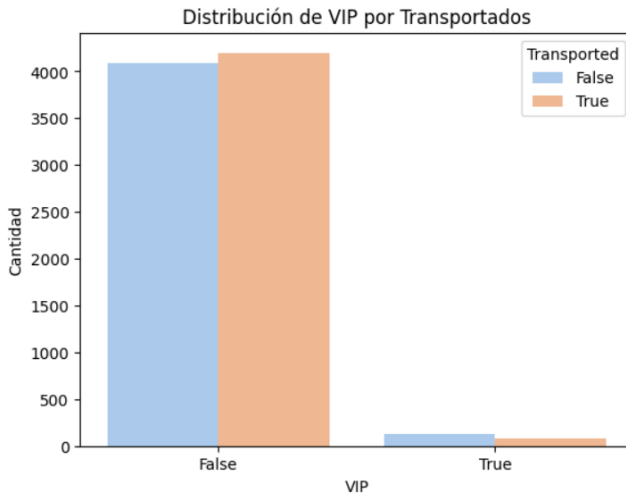


Figura: Distribución respecto VIP



Tabla de Contenidos

1 Introducción

- Contexto

2 Visualización

- Distribución

3 Selección de modelo

4 Métricas y resultados

5 Conclusiones



Los modelos escogidos fueron **Random Forest**, **LightGBM**, **Logistic Regression** y **XGBoost**.



Tabla de Contenidos

1 Introducción

- Contexto

2 Visualización

- Distribución

3 Selección de modelo

4 Métricas y resultados

5 Conclusiones



	Accuracy	F1-Score	ROC-AUC	Precision	Recall
Random Forest	0.799885	0.802945	0.799814	0.796629	0.809361
LightGBM	0.815986	0.817560	0.815967	0.816629	0.818493
Logistic Regression	0.789534	0.794151	0.789411	0.782705	0.805936
XGBoost	0.811386	0.808858	0.811530	0.826190	0.792237



Tabla de Contenidos

1 Introducción

- Contexto

2 Visualización

- Distribución

3 Selección de modelo

4 Métricas y resultados

5 Conclusiones



- 1 Se observó una correlación entre CryoSleep y la probabilidad de ser transportado.
- 2 Los pasajeros VIP tienen una probabilidad mayor de sobrevivir o evitar ser transportados.
- 3 Los modelos de clasificación, como Random Forest y LightGBM lograron resultados notables, destacando el LightGBM.
- 4 Se logró construir un modelo predictivo efectivo para identificar qué pasajeros fueron transportados por la anomalía.
- 5 Se espera investigar otras características que puedan influir en el transporte y considerar el uso de modelos más avanzados para mejorar aún más las predicciones.



Muchas gracias.

