

# Spaceship Titanic

Nombre:  
Esteban Araya Trigo

Profesor:  
Francsico Alfaro



Universidad Técnica Federico Santa María

4 de diciembre de 2023

# Definición del Problema

- PassengerId
- HomePlanet
- CryoSleep
- Cabin
- Destination
- Age
- VIP
- RoomService, FoodCourt, ShoppingMall, Spa, VRDeck
- Name
- TransporterID

# Definición del Problema

- PassengerId
- HomePlanet
- CryoSleep
- Cabin
- Destination
- Age
- VIP
- RoomService, FoodCourt, ShoppingMall, Spa, VRDeck
- Name
- TransporterID

# Definición del Problema

- PassengerId
- HomePlanet
- CryoSleep
- Cabin
- Destination
- Age
- VIP
- RoomService, FoodCourt, ShoppingMall, Spa, VRDeck
- Name
- TransporterID

# Definición del Problema

- PassengerId
- HomePlanet
- CryoSleep
- Cabin
- Destination
- Age
- VIP
- RoomService, FoodCourt, ShoppingMall, Spa, VRDeck
- Name
- TransporterID

# Definición del Problema

- PassengerId
- HomePlanet
- CryoSleep
- Cabin
- Destination
- Age
- VIP
- RoomService, FoodCourt, ShoppingMall, Spa, VRDeck
- Name
- TransporterID

# Definición del Problema

- PassengerId
- HomePlanet
- CryoSleep
- Cabin
- Destination
- Age
- VIP
- RoomService, FoodCourt, ShoppingMall, Spa, VRDeck
- Name
- TransporterID

# Definición del Problema

- PassengerId
- HomePlanet
- CryoSleep
- Cabin
- Destination
- Age
- VIP
- RoomService, FoodCourt, ShoppingMall, Spa, VRDeck
- Name
- TransporterID

# Definición del Problema

- PassengerId
- HomePlanet
- CryoSleep
- Cabin
- Destination
- Age
- VIP
- RoomService, FoodCourt, ShoppingMall, Spa, VRDeck
- Name
- TransporterID

# Definición del Problema

- PassengerId
- HomePlanet
- CryoSleep
- Cabin
- Destination
- Age
- VIP
- RoomService, FoodCourt, ShoppingMall, Spa, VRDeck
- Name
- TransporterID

# Definición del Problema

- PassengerId
- HomePlanet
- CryoSleep
- Cabin
- Destination
- Age
- VIP
- RoomService, FoodCourt, ShoppingMall, Spa, VRDeck
- Name
- Transported

# Estadística descriptiva

	PassengerId	HomePlanet	CryoSleep	Cabin	Destination	Age	VIP	RoomService	FoodCourt	ShoppingMall	Spa	VRDeck	Name	Transported
0	0001_01	Europa	False	B/O/P	TRAPPIST-1e	39.0	False	0.0	0.0	0.0	0.0	0.0	Maham Ofraculy	False
1	0002_01	Earth	False	F/O/S	TRAPPIST-1e	24.0	False	109.0	9.0	25.0	549.0	44.0	Juanna Vines	True
2	0003_01	Europa	False	A/O/S	TRAPPIST-1e	58.0	True	43.0	3576.0	0.0	6715.0	49.0	Altark Susent	False
3	0003_02	Europa	False	A/O/S	TRAPPIST-1e	33.0	False	0.0	1283.0	371.0	3329.0	193.0	Solam Susent	False
4	0004_01	Earth	False	F/1/S	TRAPPIST-1e	16.0	False	303.0	70.0	151.0	565.0	2.0	Willy Santantines	True

Figura 1: Descripción del data frame Train

# Estadística descriptiva

	Age	RoomService	FoodCourt	ShoppingMall	Spa	VRDeck
<b>count</b>	8514.000000	8512.000000	8510.000000	8485.000000	8510.000000	8505.000000
<b>mean</b>	28.827930	224.687617	458.077203	173.729169	311.138778	304.854791
<b>std</b>	14.489021	666.717663	1611.489240	604.696458	1136.705535	1145.717189
<b>min</b>	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
<b>25%</b>	19.000000	0.000000	0.000000	0.000000	0.000000	0.000000
<b>50%</b>	27.000000	0.000000	0.000000	0.000000	0.000000	0.000000
<b>75%</b>	38.000000	47.000000	76.000000	27.000000	59.000000	46.000000
<b>max</b>	79.000000	14327.000000	29813.000000	23492.000000	22408.000000	24133.000000

Figura 2: Metricas del data frame de entrenamiento

# Visualización descriptiva

Distribucion de Transportados

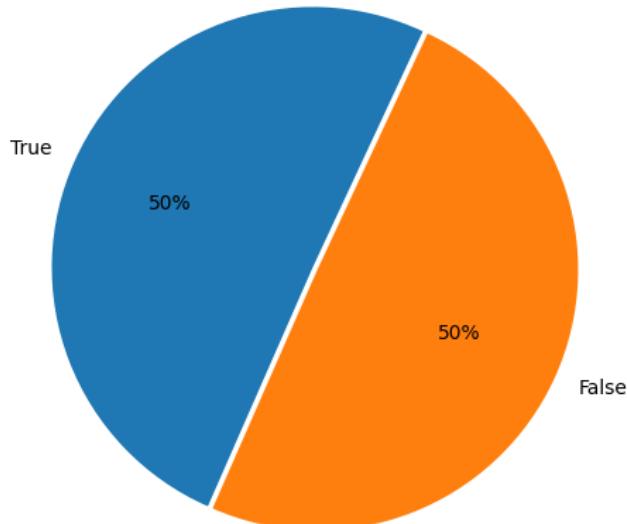
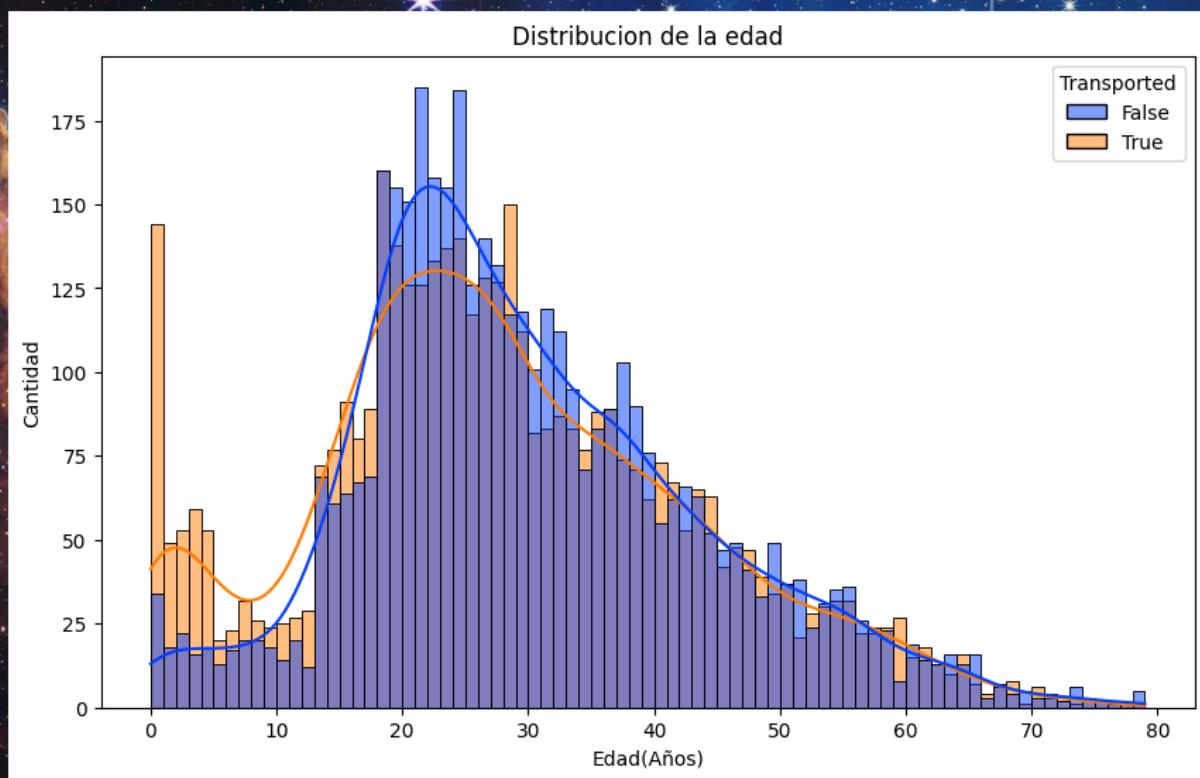


Figura 3: Porcentaje de persona transportadas a otra dimensión.

# Visualización descriptiva



# Visualización descriptiva

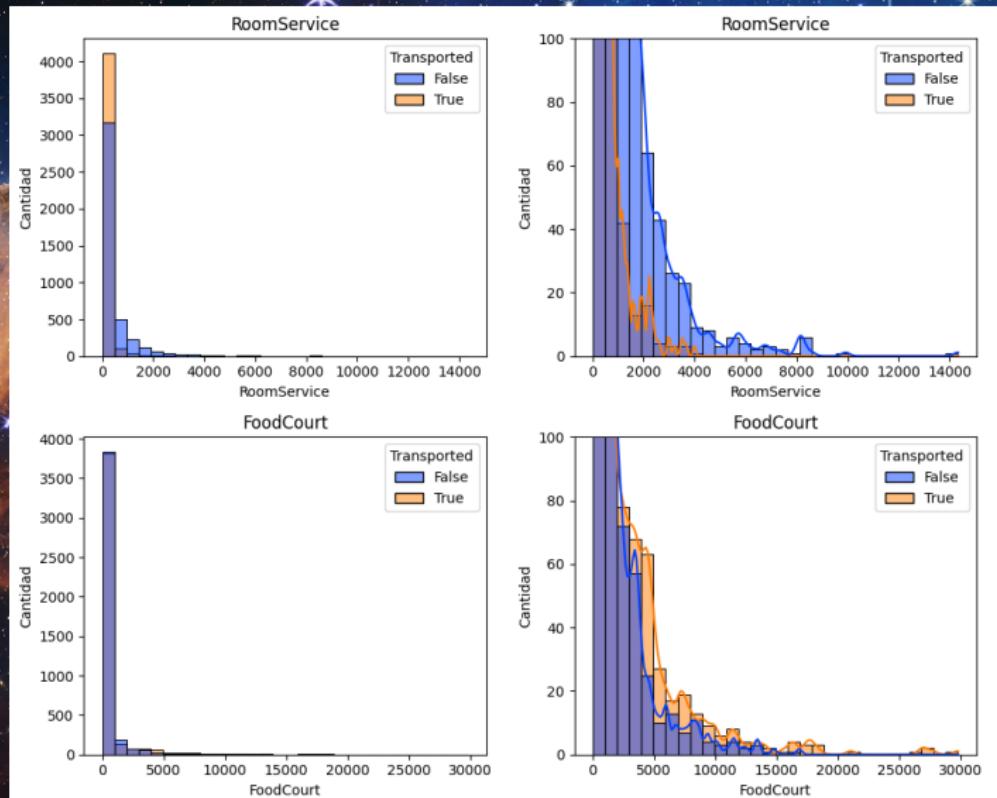


Figura 5: Grafo de RoomService y FoodCourt

# Visualización descriptiva

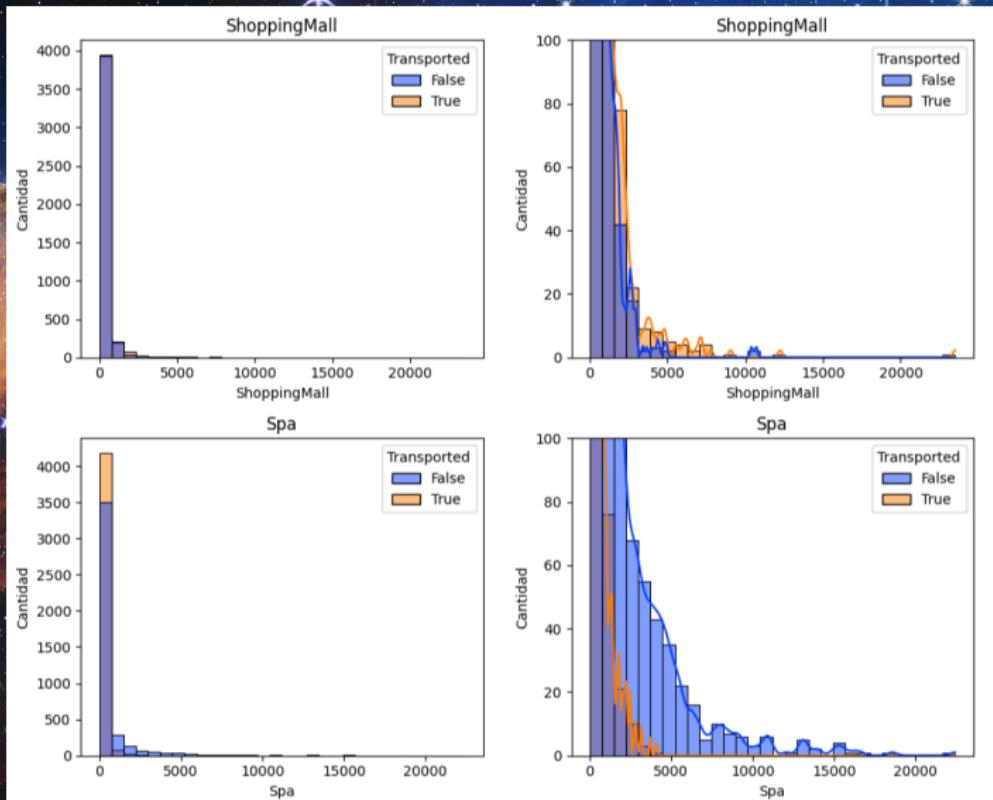


Figura 6: Grafico de ShoppingMall y Spa

# Visualización descriptiva

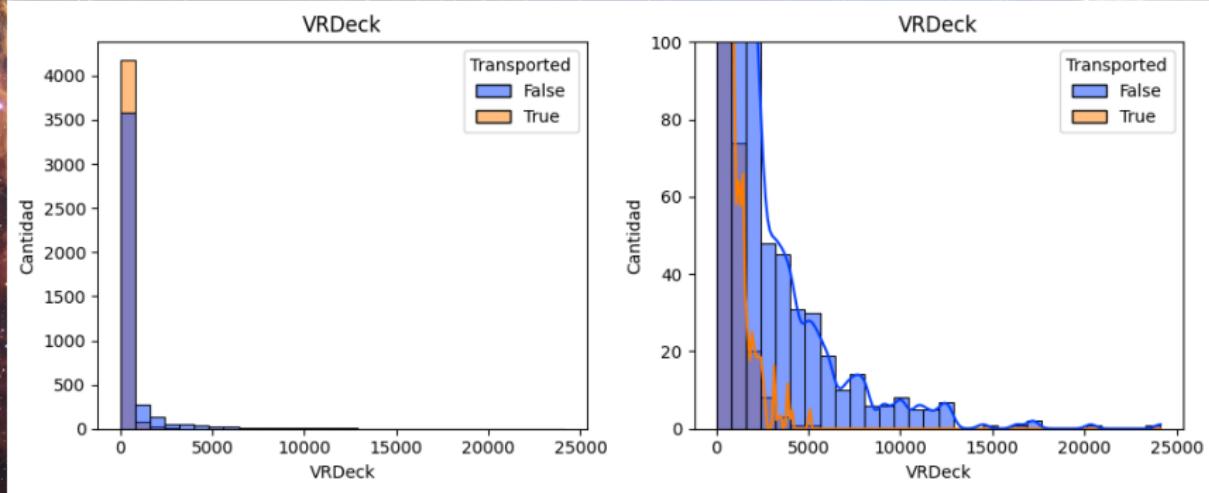


Figura 7: Grafico de VRDeck

# Visualización descriptiva

Distribucion de CryoSleep y VIP

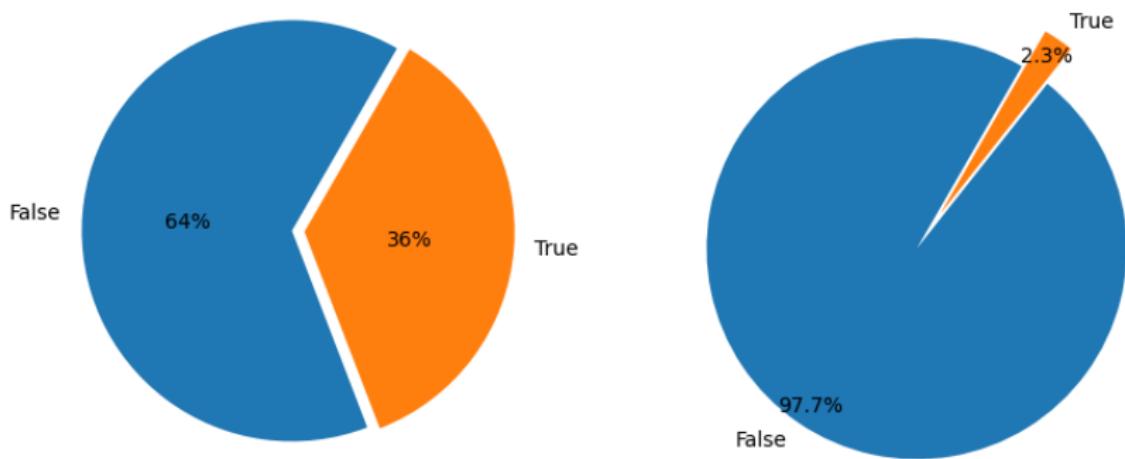


Figura 8: Distribucion de CryoSleep y VIP

# Visualización descriptiva

## Distribucion de HomePlanet y Destination

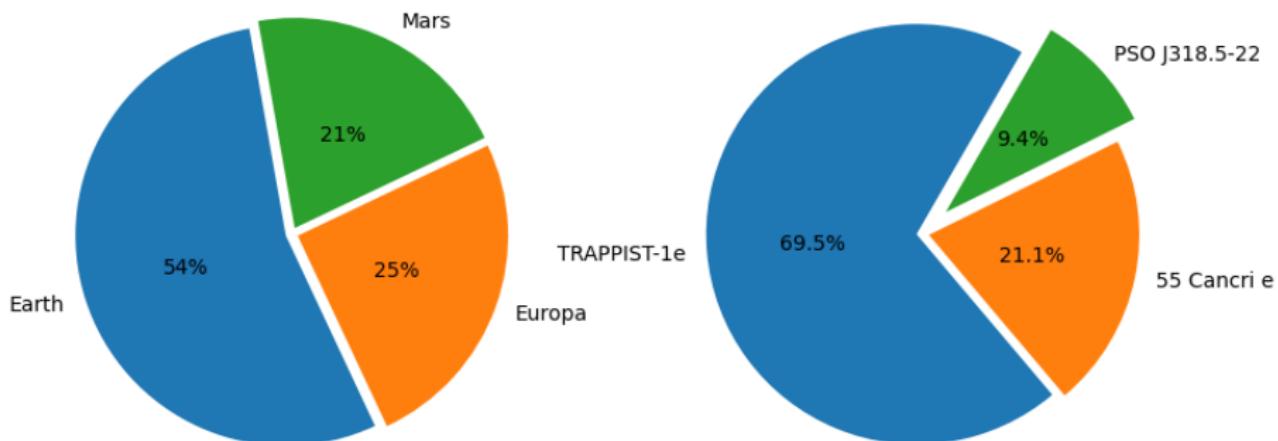


Figura 9: Porcentajes de HomePlanet y Destination

# Visualización descriptiva

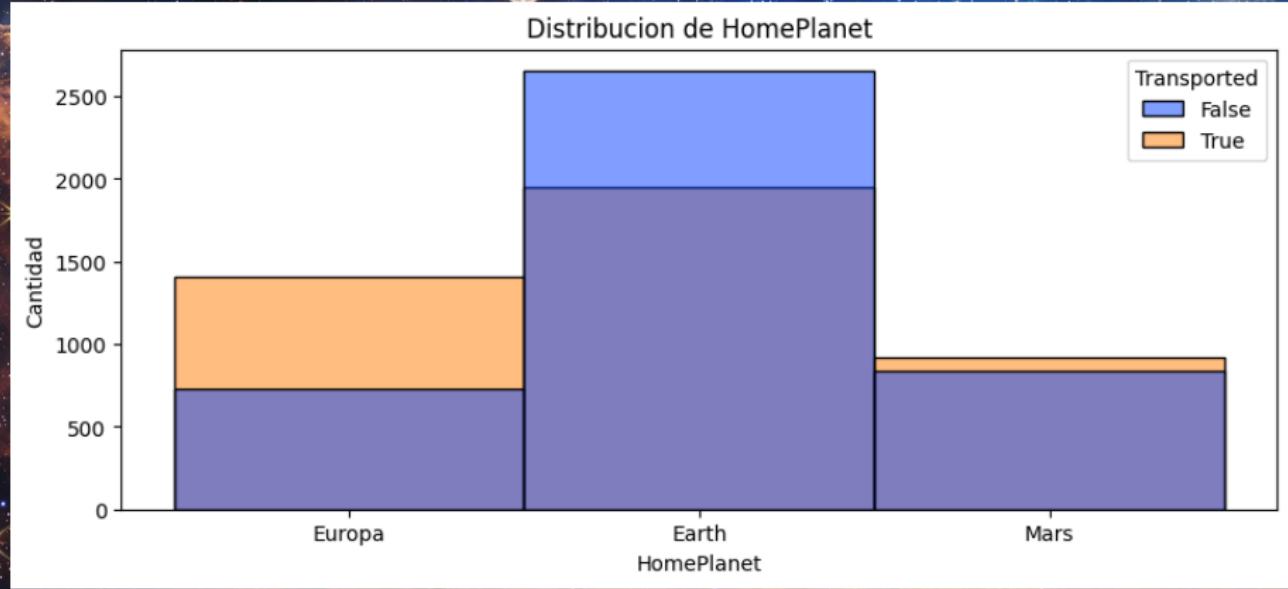


Figura 10: Distribucion de HomePlanet

# Visualización descriptiva

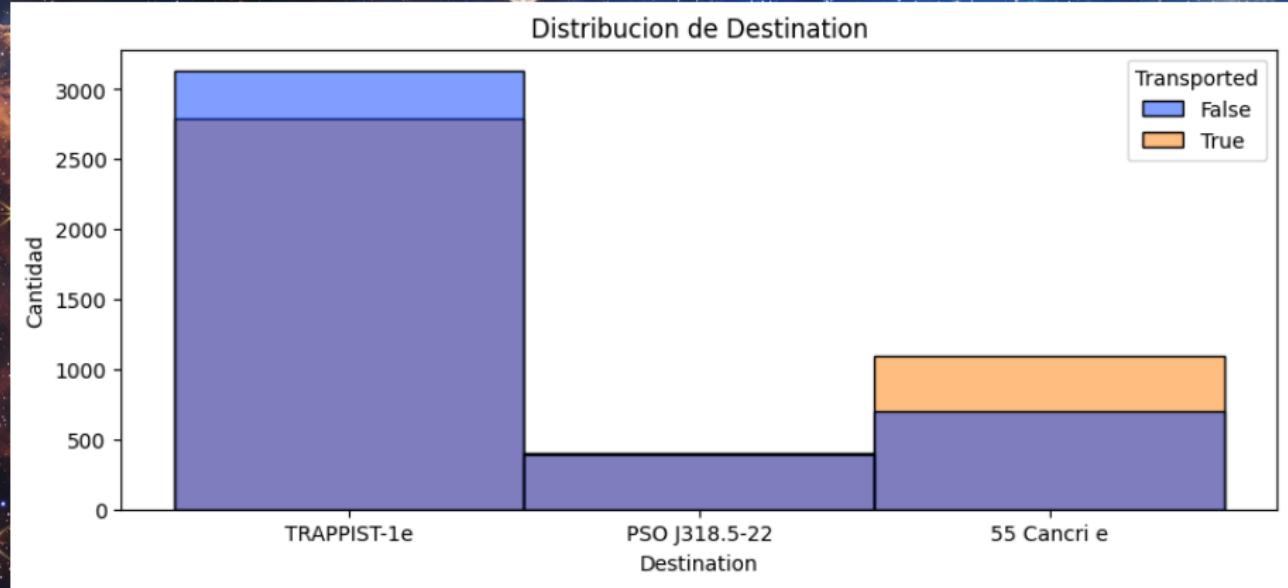


Figura 11: Distribucion de Destination

# Preprocesamiento

Filtramos el dataframe con la informacion relevante para analizar que personas que fueron transportadas y quitamos todos los datos nulos que son aproximadamente 1412 elementos

	HomePlanet	Destination	Age	RoomService	FoodCourt	ShoppingMall	Spa	VRDeck	Transported
0	Europa	TRAPPIST-1e	39.0	0.0	0.0	0.0	0.0	0.0	False
1	Earth	TRAPPIST-1e	24.0	109.0	9.0	25.0	549.0	44.0	True
2	Europa	TRAPPIST-1e	58.0	43.0	3576.0	0.0	6715.0	49.0	False
3	Europa	TRAPPIST-1e	33.0	0.0	1283.0	371.0	3329.0	193.0	False
4	Earth	TRAPPIST-1e	16.0	303.0	70.0	151.0	565.0	2.0	True

Figura 12: Datos procesados

# Modelos

- LogisticRegression
- KNeighbors
- RandomForest
- NaiveBayes

# Modelos

Usamos los siguientes parametros para cada modelos:

Para Logistic Regression:

- penalty: l1, l2
- C: 0.25, 0.5, 0.75, 1, 1.25, 1.5
- max iter: 50, 100, 150

Para KNeighbors:

- neighbors: 3, 5, 7, 9
- p: 1, 2

Para Random Forest:

- estimators: 50, 100, 150, 200, 250, 300
- max depth: 6, 8, 10, 12, 14

Para Naive Bayes:

- var smoothing: 1e-10, 1e-9, 1e-8, 1e-7

Para SVC:

- C: 1.0
- kernel: linear

# Modelos

Mediante testeo observamos cual era el mejor parametro para cada modelo Para Logistic Regression:

- penalty: l2
- C: 0.25
- max iter: 50

Para KNeighbors:

- neighbors: 9
- p: 2

Para Random Forest:

- estimators: 250
- max depth 12

Para Naive Bayes

- var smoothing 1e-10

Para SVC:

- C: 1.0
- kernel: linear

# Métricas y análisis de resultados

Classifier	Validation accuracy	Training time
LogisticRegression	0.790666	0.06
KNN	0.780371	0.09
RandomForest	0.795470	1.93
NaiveBayes	0.702128	0.00
SVC	0.783802	0.01

# Visualizaciones del modelo

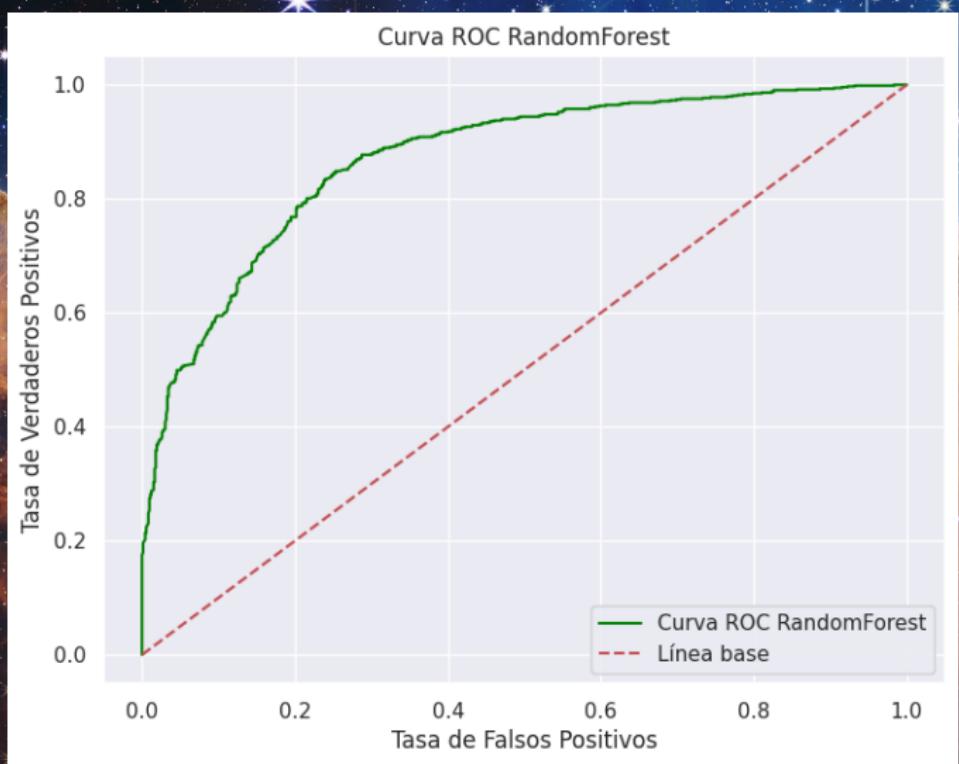


Figura 13: Efectividad del mejor modelo

# Visualizaciones del modelo

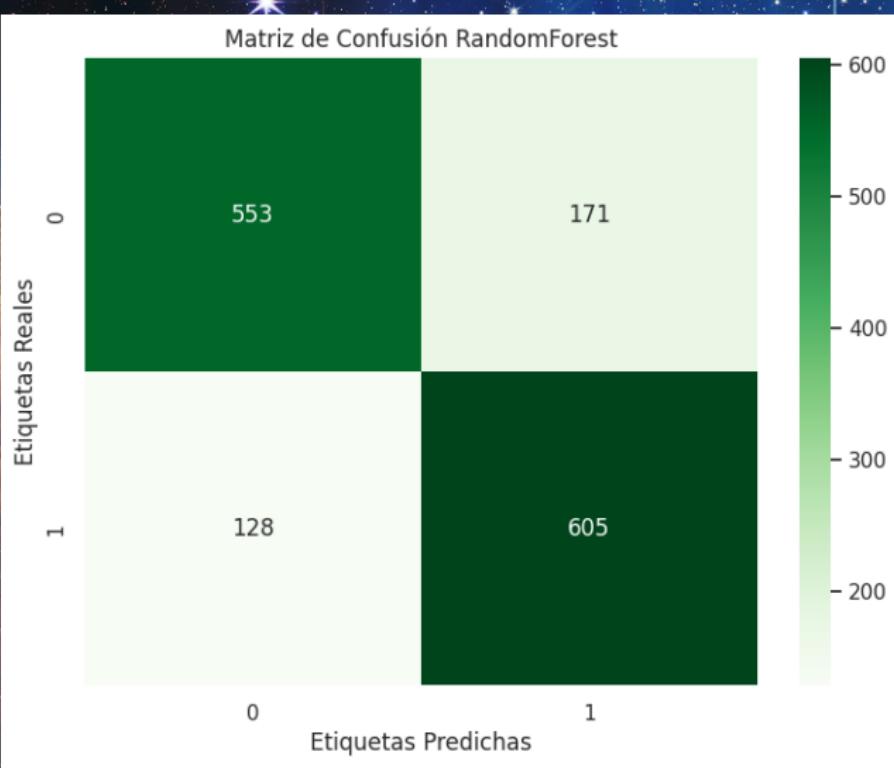


Figura 14: Aciertos del modelo

# Conclusiones

Las principales conclusiones del proyecto

- La elección del modelo fue la mejor
- Aplicamos todo lo aprendido en el curso
- Mejoramiento de la predicción mediante otras técnicas de modelamiento y/o filtros de los datos
- La mayoría de modelos se quedan estables en alguna cantidad de parámetros

The background of the image is a deep space scene featuring a dense cluster of stars of various sizes and colors, primarily white and yellow. Interspersed among the stars are large, wispy clouds of interstellar gas and dust, colored in shades of orange, red, and brown. These nebulae appear to be in the foreground, creating a sense of depth. The overall composition is a high-quality astronomical photograph.

Gracias Por Su Atención