

# SPACESHIP TITANIC

CIETE (7)

# INTELIGENCIA ARTIFICIAL AVANZADA PARA LA CIENCIA DE DATOS I

## INTEGRANTES:

Diana Cañibe Valle A01749422

Andrea Viany Díaz Álvarez A01750147

José Benjamín Ruiz García A01750246

Edna Jacqueline Zavala Ortega A01750480

Cristian Aldo Sandoval Suárez A01751137

ESCUELA DE INGENIERÍA Y CIENCIAS  
CAMPUS ESTADO DE MÉXICO

15 de Septiembre del 2022



Tecnológico  
de Monterrey



# AGENDA

- 
- Contexto del reto 01
  - Procesamiento de datos 02
  - Modelado 03
  - Empresa 'Open Space' 04
  - Despliegue del modelo 05
  - Modelo de negocio 06
  - Conclu 07
- 





# — TRAGEDIA ESPACIAL

## INTRODUCCIÓN

En el año 2912 el transatlántico interestelar 'Spaceship Titanic' emprendió su travesía con 13,000 pasajeros a bordo, desde nuestro sistema solar a tres exoplanetas (55 Cancri E,PSO J318.5-22,TRAPPIST-1e).

Lamentablemente mientras rodeaba Alpha Centauri en ruta hacia su primer destino, la nave chocó con una anomalía del espacio-tiempo escondida dentro de una nube de polvo, la cual causó la teletransportación de casi la mitad de sus pasajeros a una dimensión alternativa.





## — MISIÓN

Generar un modelo de predicción utilizando los registros recuperados de la nave espacial para determinar qué pasajeros fueron transportados, y habilitar una plataforma para realizar consultas a el modelo.

- 
- 
- 
- 

## — OBJETIVOS

### ANÁLISIS Y PROCESAMIENTO DE DATOS

-Realizar el estudio inicial de los datos proporcionados, y la limpieza, imputación y cambios considerados adecuados

### MODELADO

-Construcción, pruebas y ajustes de diferentes modelos para la selección del mejor modelo

### DESPLIEGUE DEL MODELO

-Desarrollo de aplicación para demostración práctica del modelo desarrollado, proporcionando los resultados de clasificación



# HERRAMIENTAS UTILIZADAS

Datos y modelado



Despliegue





# PERFIL DE DATOS

## CANTIDAD

- train.csv : registros personales de aproximadamente dos tercios (~ 8700) de los pasajeros
- test.csv : registros personales del tercio restante (~4300) de los pasajeros

## ATRIBUTOS

- 1. PassengerId- Id único para cada pasajero.
- 2. HomePlanet- Planeta del que partió el pasajero.
- 3. CryoSleep- Indica si el pasajero eligió ser puesto en animación suspendida durante la duración del viaje.
- 4. Cabin- El número de cabina donde se hospeda el pasajero. (deck/num/side).
- 5. Destination- Planeta donde desembarcará el pasajero.
- 6. Age- Edad del pasajero.
- 7. VIP- Indica si el pasajero ha pagado por servicio VIP .
- 8. RoomService, FoodCourt, ShoppingMall, Spa, VRDeck- Monto que el pasajero ha facturado en cada uno de los servicios de lujo.
- 9. Name- Nombre completo del pasajero.
- 10. Transported- Si el pasajero fue transportado a otra dimensión.

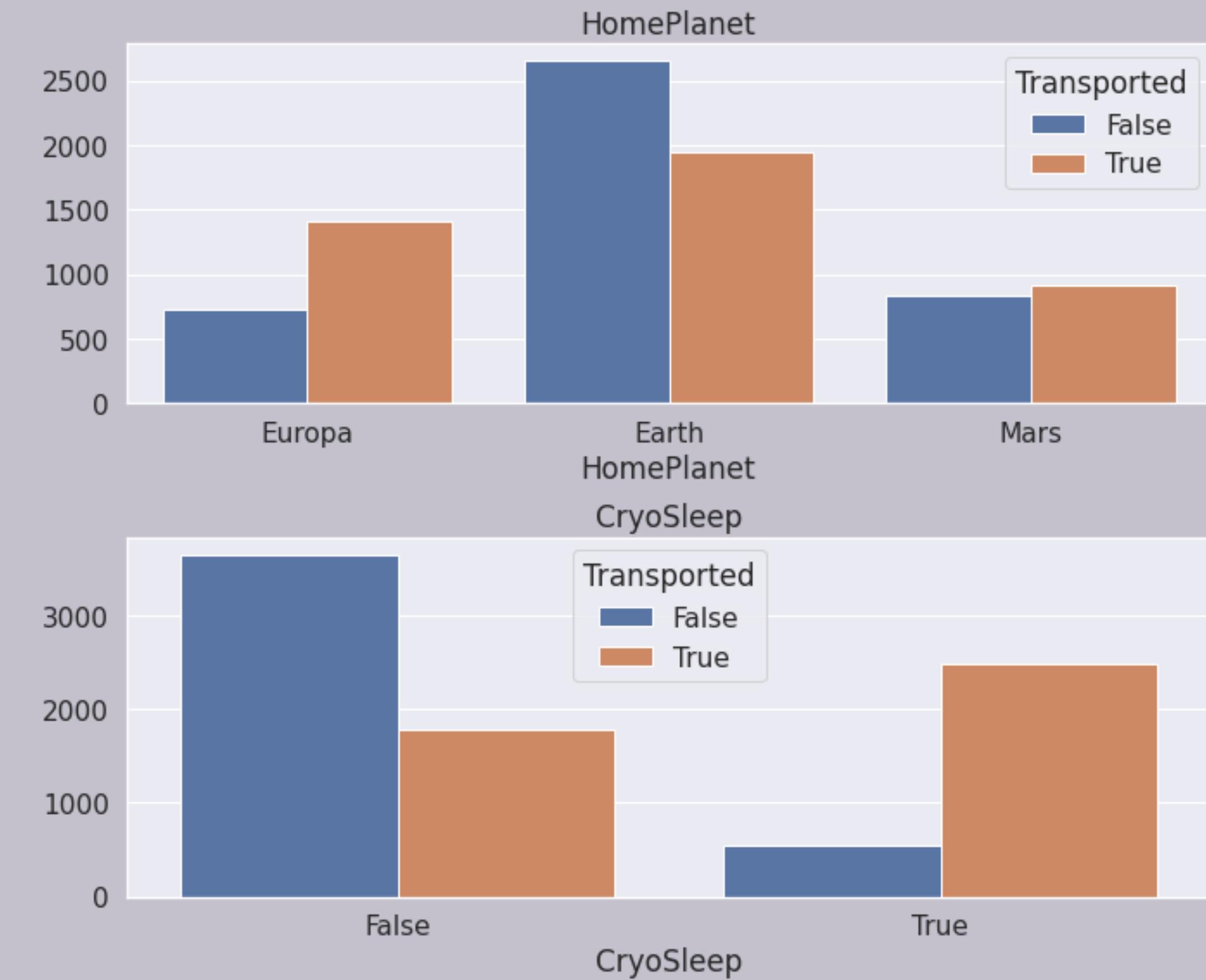
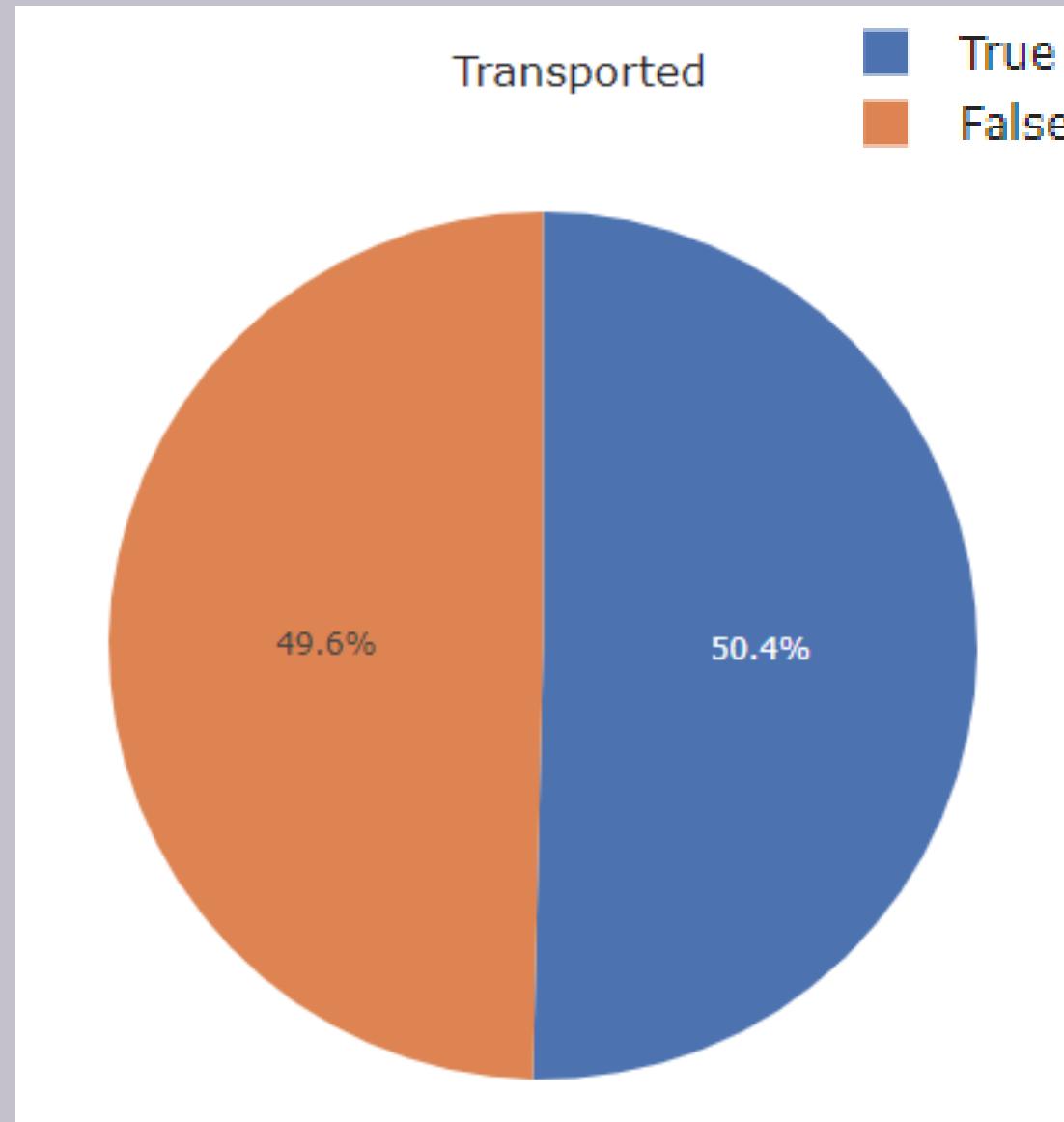
	PassengerId	HomePlanet	CryoSleep	Cabin	Destination	Age	VIP	RoomService	FoodCourt	ShoppingMall	Spa	VRDeck	Name	Transported
0	0001_01	Europa	False	B/O/P	TRAPPIST-1e	39.0	False	0.0	0.0	0.0	0.0	0.0	Maham Ofracculy	False
1	0002_01	Earth	False	F/O/S	TRAPPIST-1e	24.0	False	109.0	9.0	25.0	549.0	44.0	Juanna Vines	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/I/S	TRAPPIST-1e	16.0	False	303.0	70.0	151.0	565.0	2.0	Willy Santantines	True



# PERFIL DE DATOS

TRAIN SET

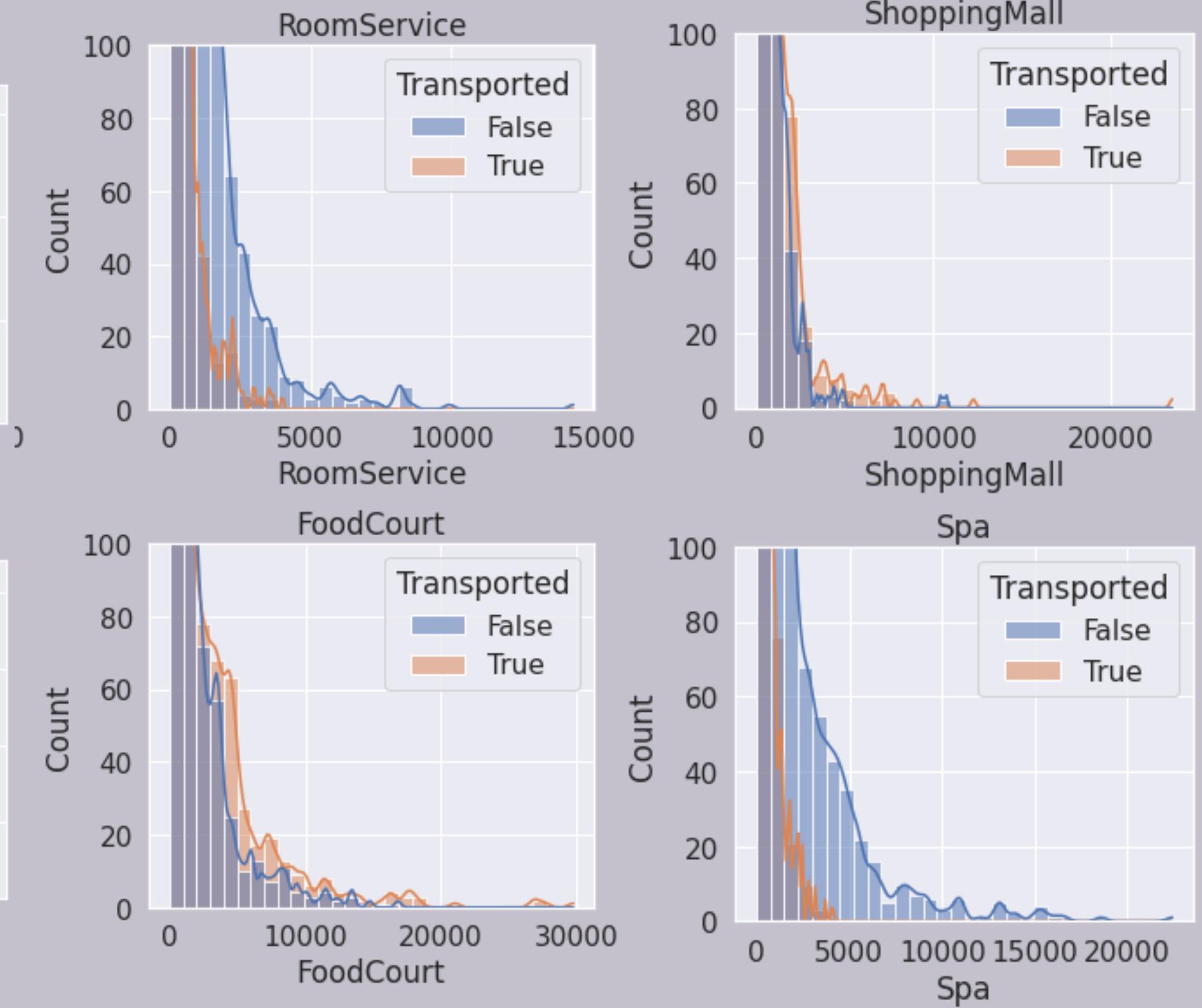
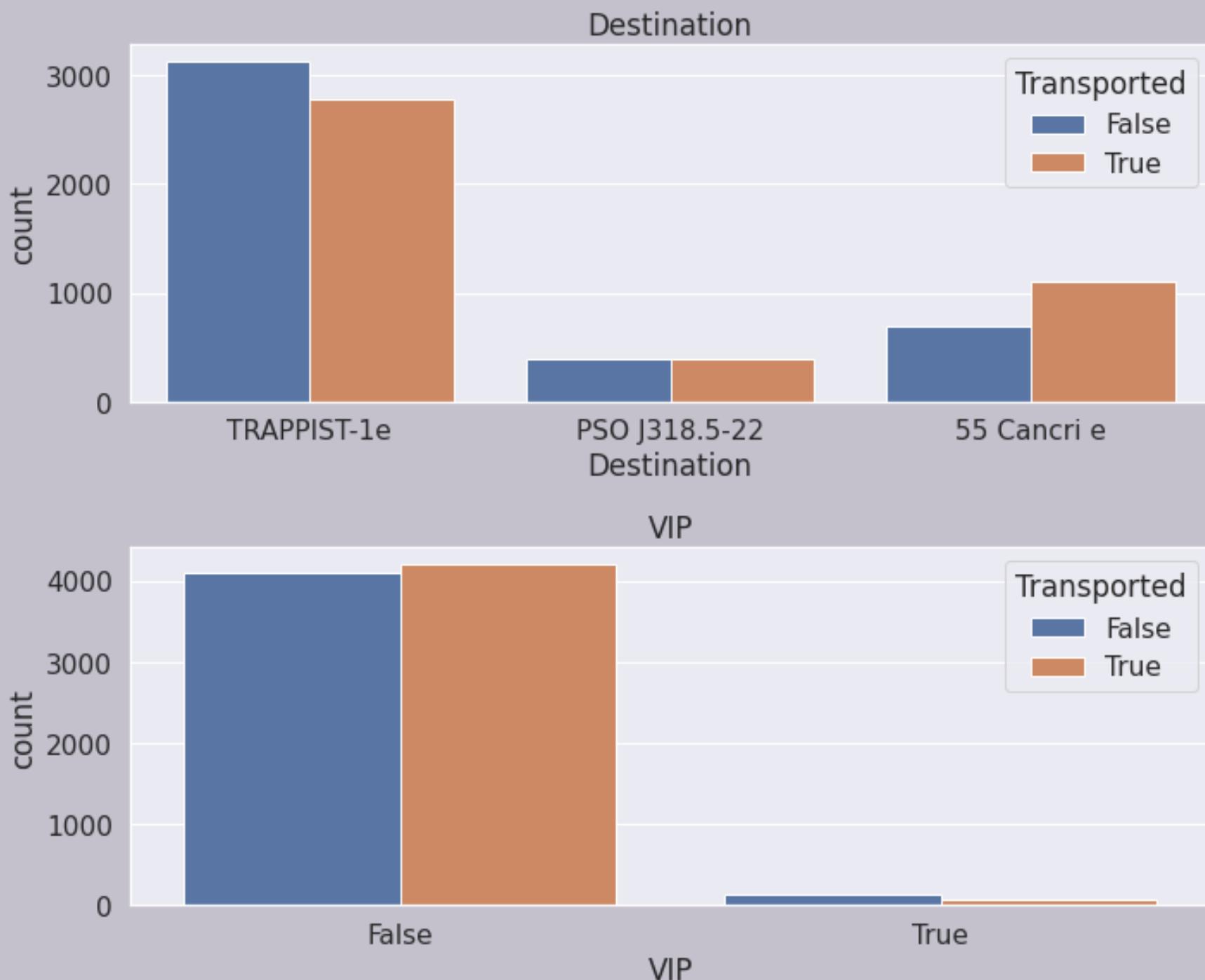
= DATOS



# PERFIL DE DATOS

TRAIN SET

= DATOS





# PREPROCESAMIENTO DE DATOS

# TRATAMIENTO E IMPUTACIÓN

## FASE 1

- Separación de cabina en: puerto (Deck) , número de puerto (deck\_number) y lado (Side)
  - De bool a int : Cryosleep, VIP
  - One hot encoding : HomePlanet, Destination, Side
  - Label encoding: Deck
  - Llenado de variables categóricas con la moda (train)
  - Llenado de variables numéricas con la mediana (train)

## FASE 2

- Separación por grupos de edad (age\_group)
  - Suma de gastos (total\_expenses)
  - Cualidad de haber gastado (no\_expenses)
  - Extracción del grupo a partir del Id del pasajero
  - Llenado de variables categóricas con la moda (train+test)
  - Llenado de variables numéricas con la mediana (train+test)
  - Estandarización de los features.

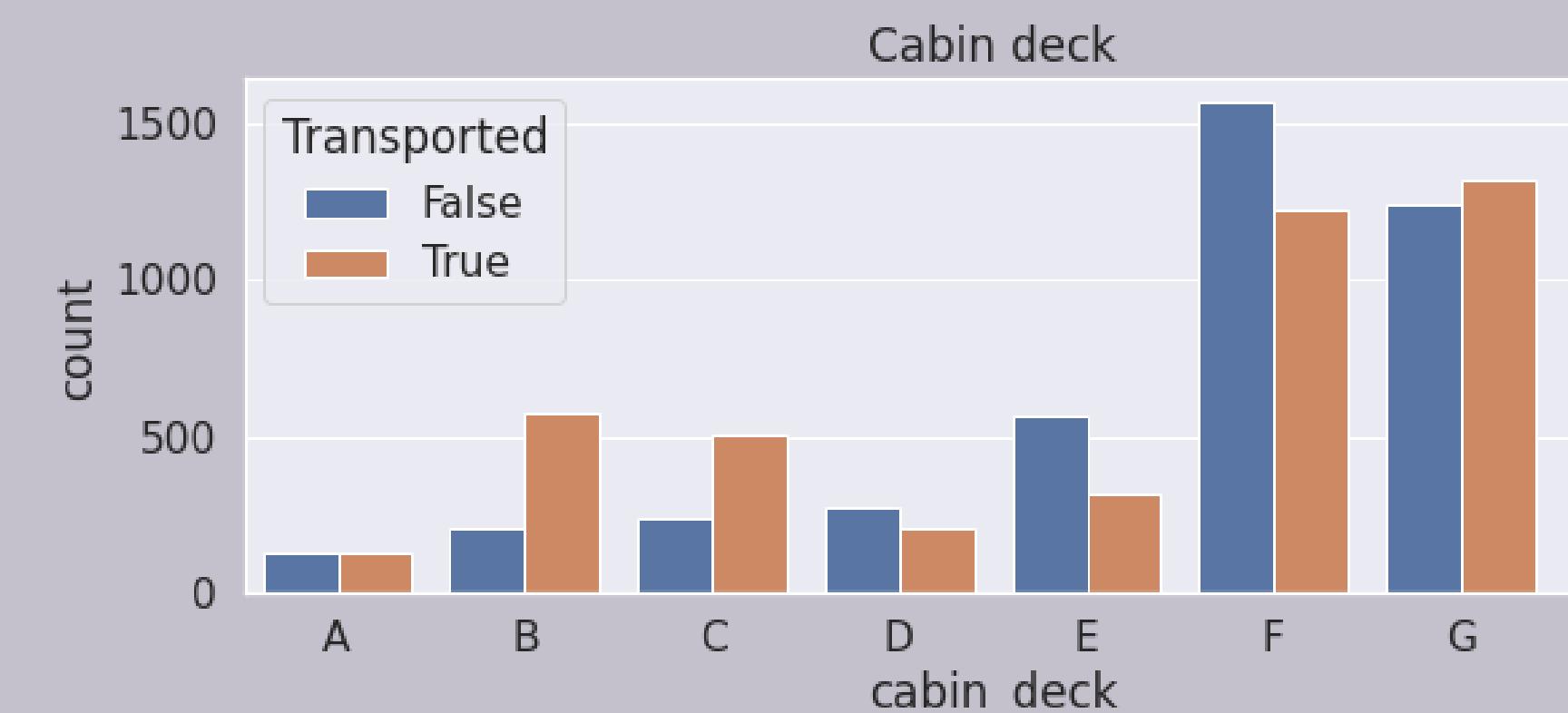
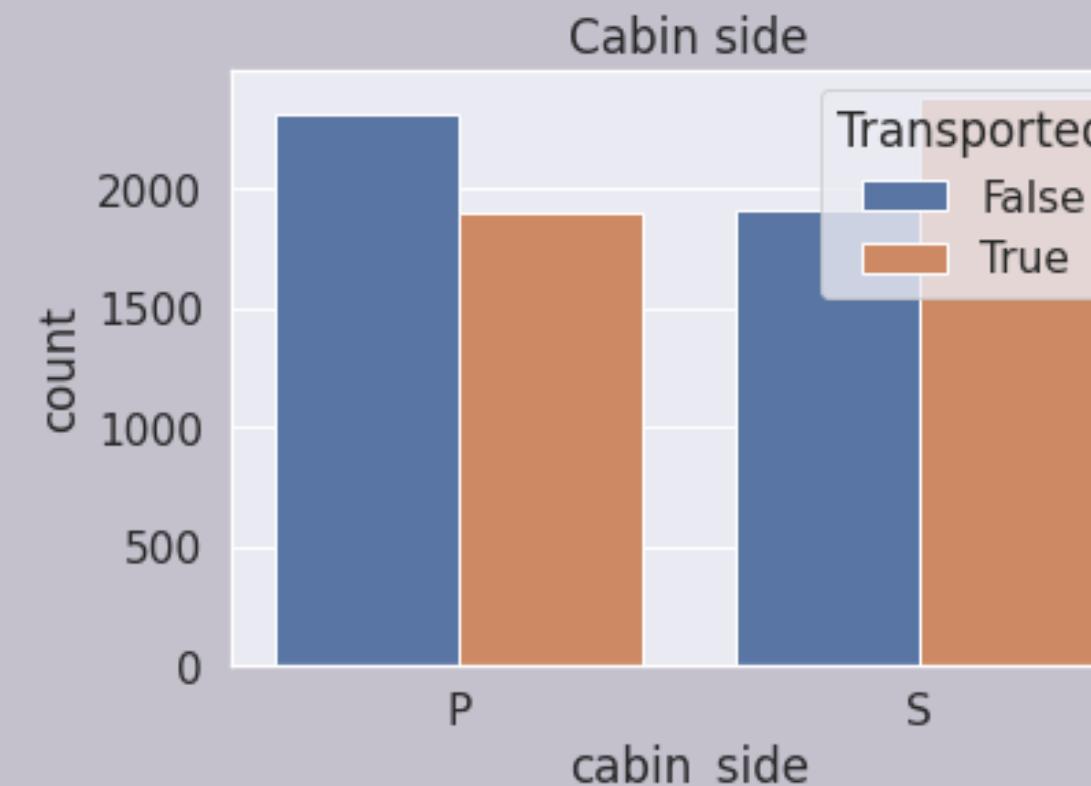
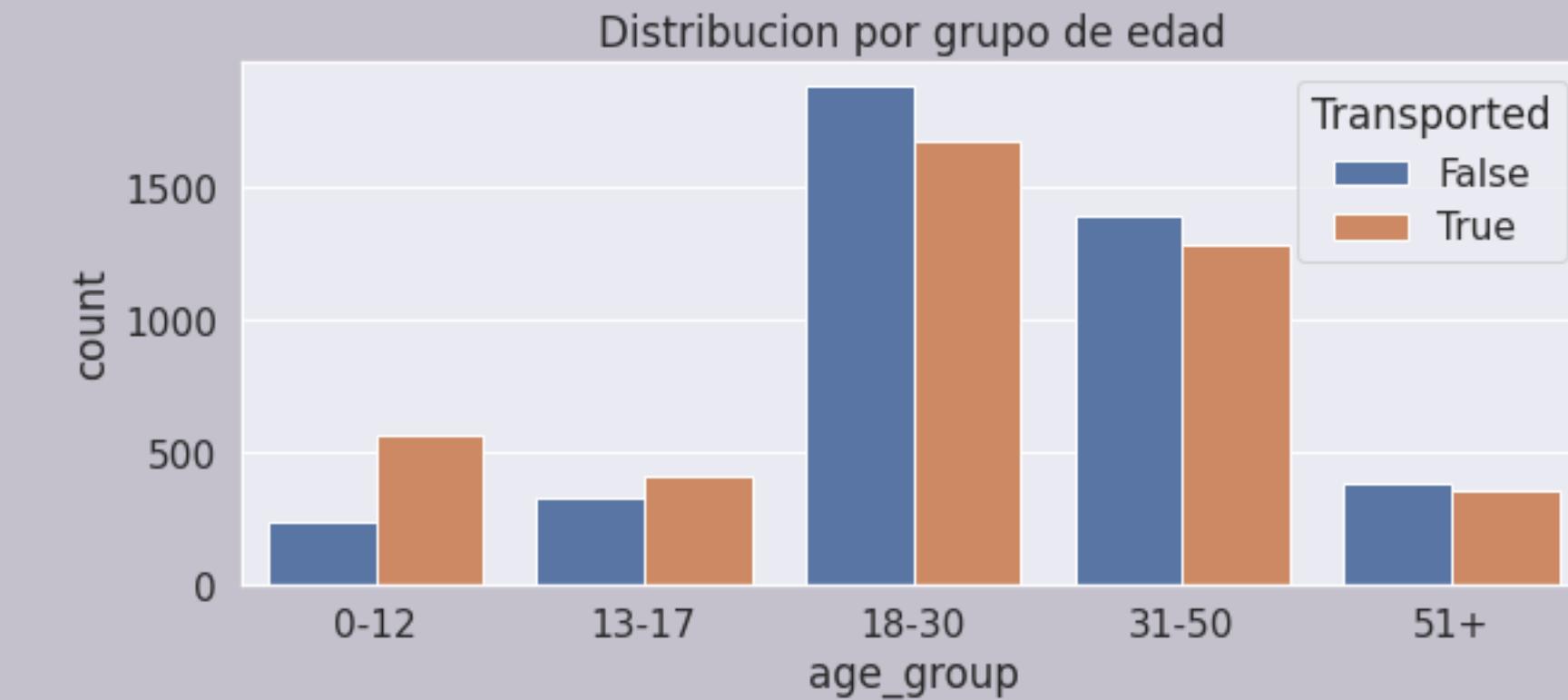
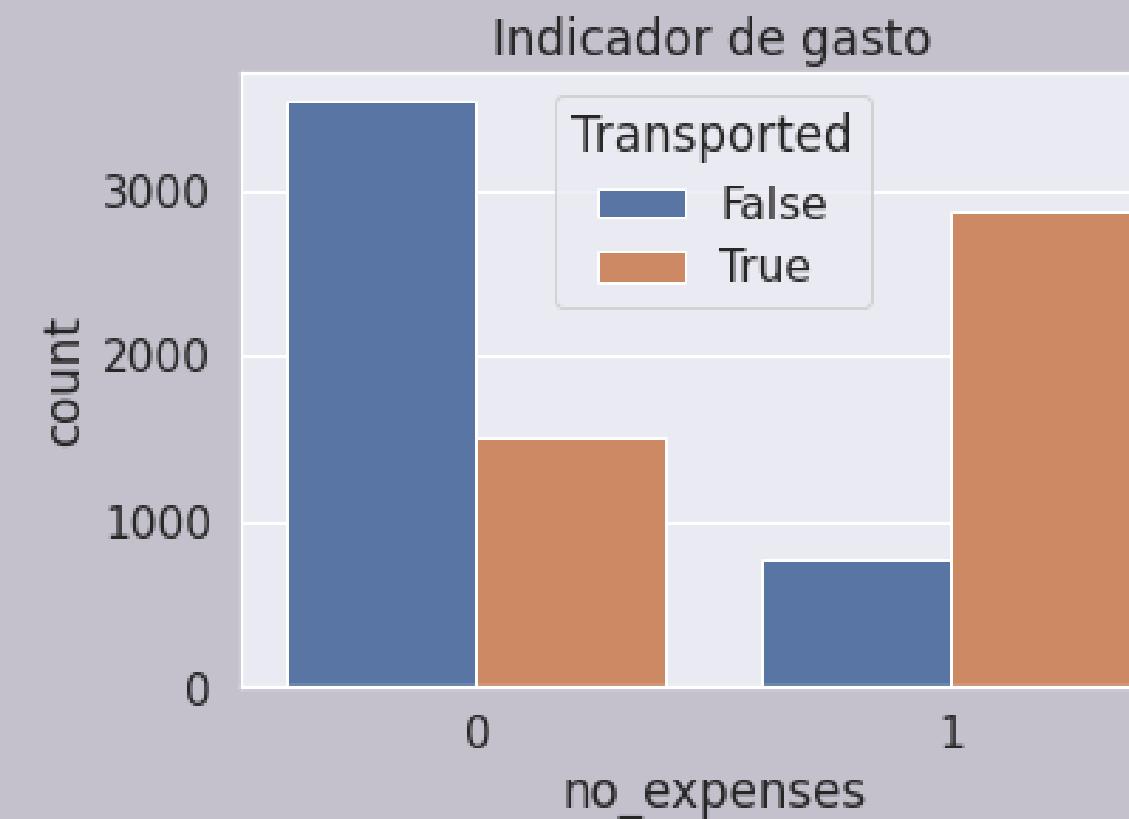
total_expenses	no_expenses	group	Deck	Side
0.0	1	1	B	R
736.0	0	2	F	S
10383.0	0	3	A	S
5176.0	0	3	A	S
1091.0	0	4	F	S

Destination_PS0	Destination_TRAPPIST-1e	Side_P	Side_S
J318.5-22	0	1	0
0	1	0	1
0	1	0	1
0	1	0	1
0	1	0	1

# PERFIL DE DATOS

TRAIN SET + CAMBIOS

= DATOS





# • MODELADO

## F A S E 1

Modelo inicial de prueba (Regresión Logística)

Score: 0.72340



## • F A S E 2

- Prueba de distintos modelos tras cambios adicionales en preprocessamiento de datos.
- 

Modelo	Logistic Regression	KNN	SVC	Random Forest	Decision Tree	XGBoost	LGMB	CatBoost	NaiveBayes
Score	0.7853	0.7842	0.7997	0.7741	0.7534	0.8141	0.8037	0.8037	0.7359





# MODELADO

## F A S E 4

- Estandarización de los datos y crossvalidation para probar distintos modelos.
- 
- 

Modelo	Logistic Reggresion	KNN	Random Forest
Score	[ 0.7996 0.7775 0.8066 0.8121 ]	[ 0.7665 0.7536 0.7826 0.7587 ]	[ 0.7683 0.7610 0.7826 0.7679 ]

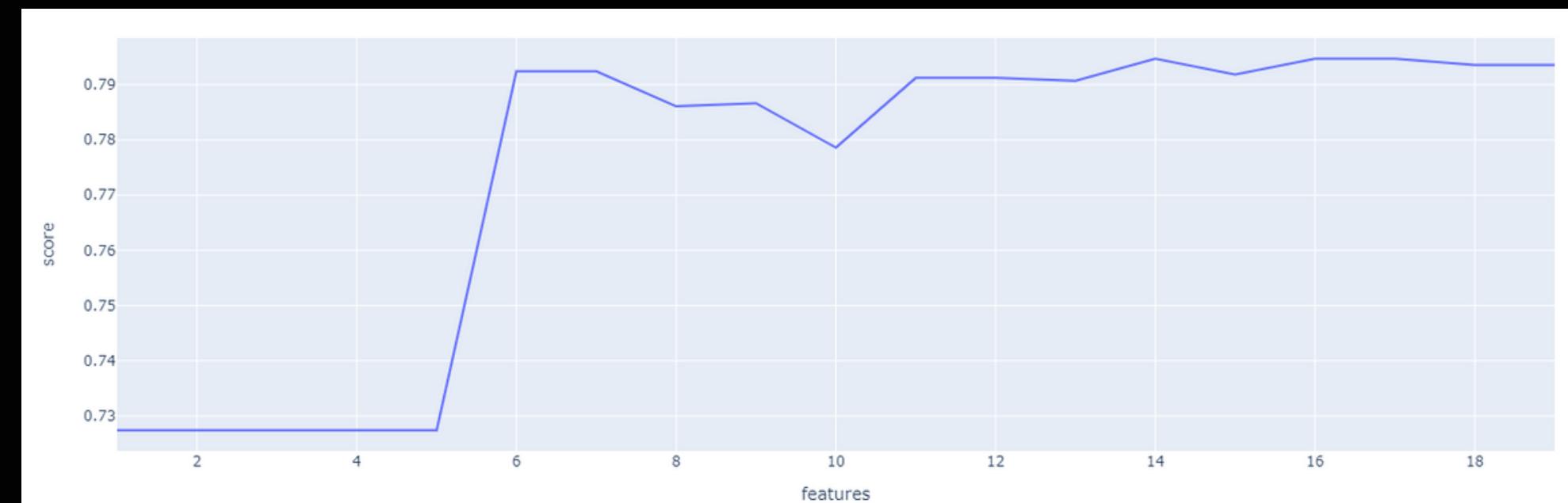
Todas las etiquetas	
PR AUC	0.911
ROC AUC	0.908
Log loss	0.379
F1 score	0.8106236
Precisión	81.1 %
Recuperación	81.1 %
Fecha de creación	9 sept 2022 19:31:52

## F A S E 3

Experimentación con AutoML  
Score: 0.58569

## F A S E 5

GridSearchCV y SelectKBest para ajuste de hiperparámetros



RESULTADOS EN

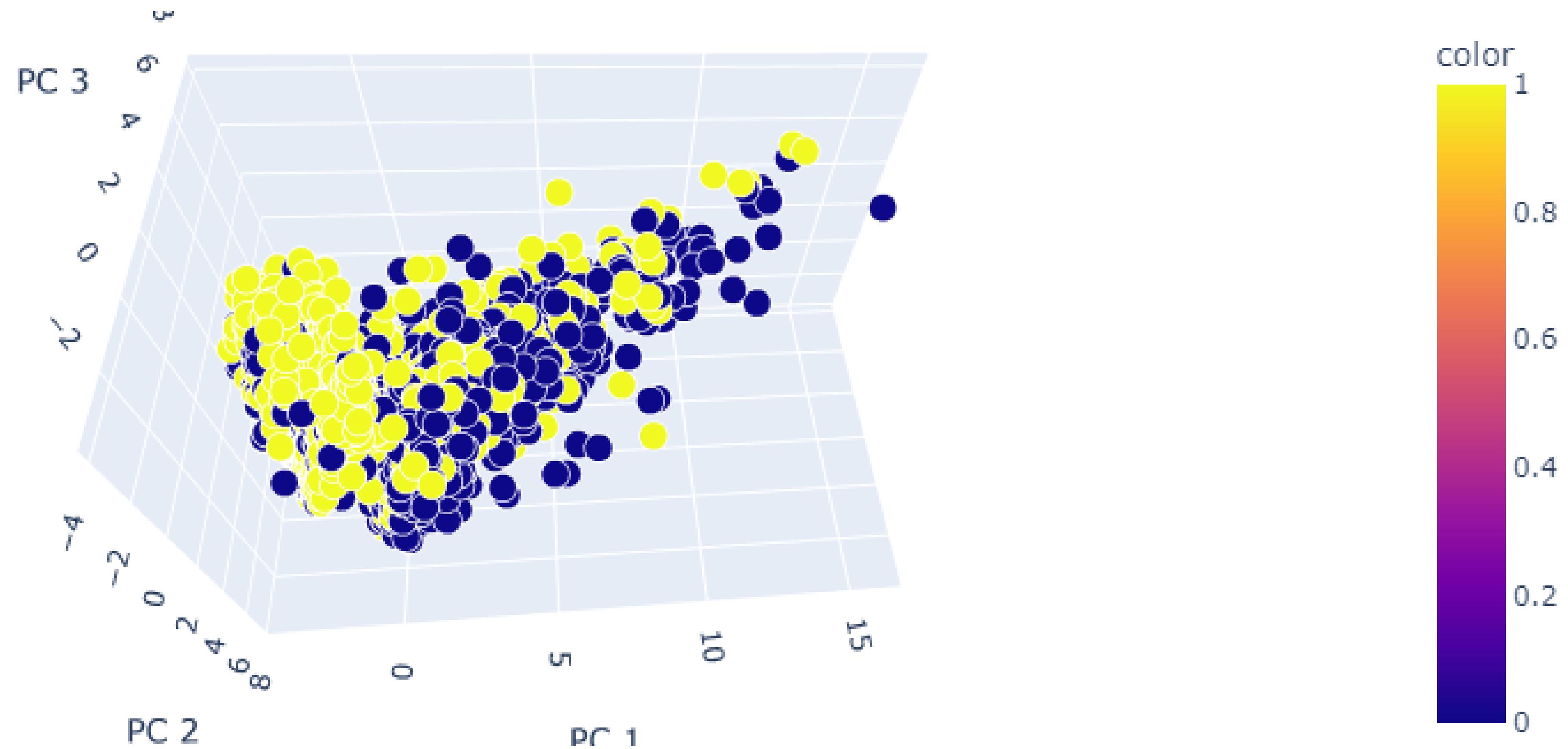
kaggle

SCORE:

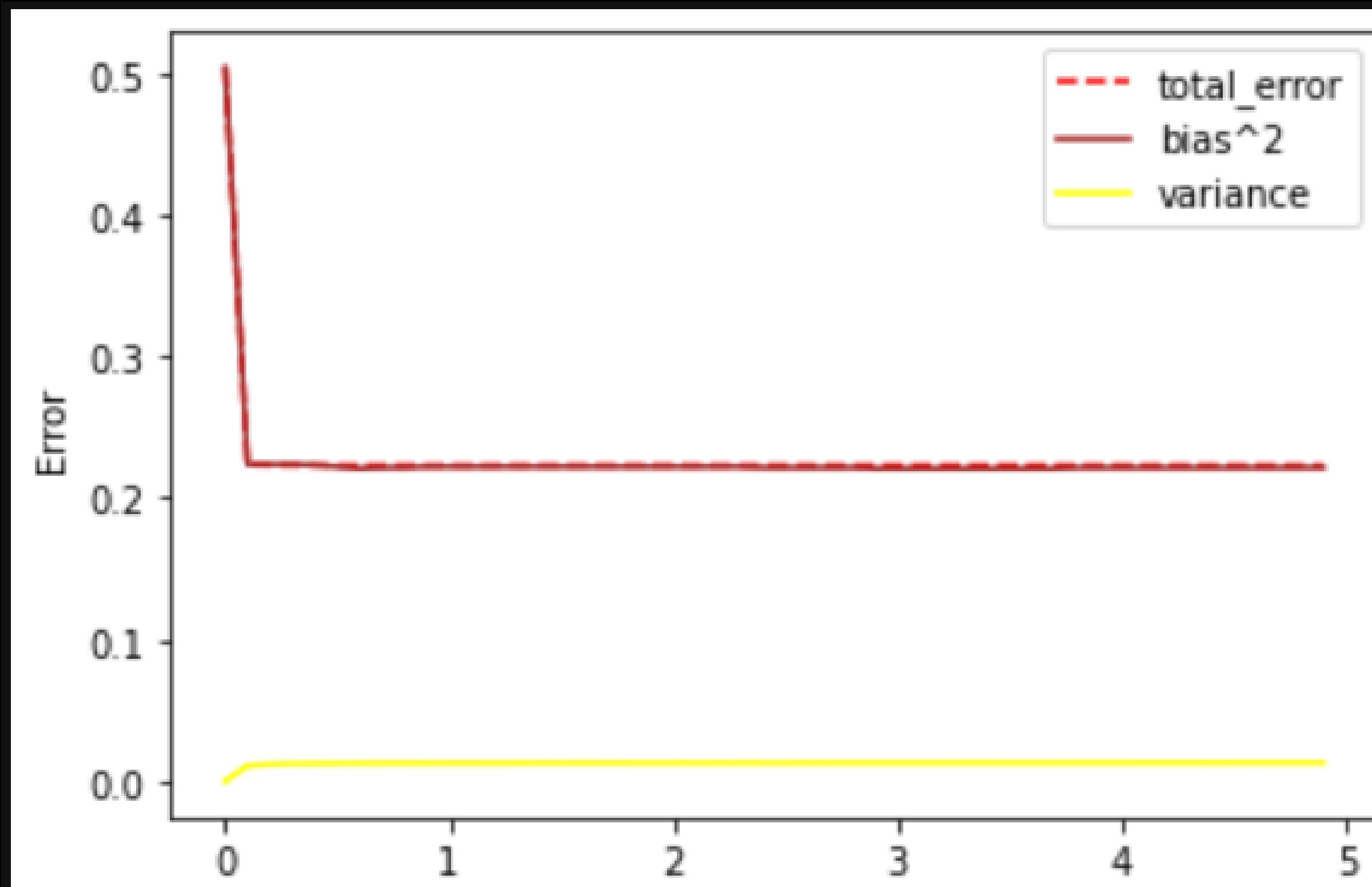
0.80009



Total Explained Variance: 42.04%



# BALANCE DEL MODELO



# Empresa de Grupo Ciete

Fundación: 14 de Septiembre 2022

## Misión:

Brindar la oportunidad de identificar  
y rescatar a las personas  
transportadas a otra dimensión

## Visión:

Completar 8 de cada 10 servicios de  
rescate contratados en los próximos  
2 años luz

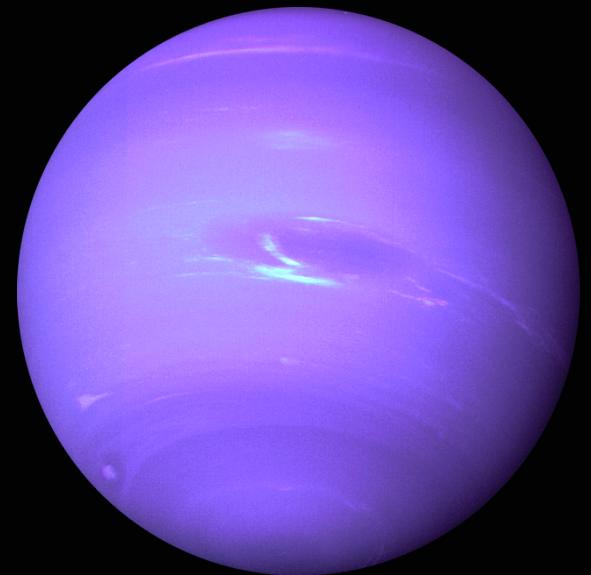




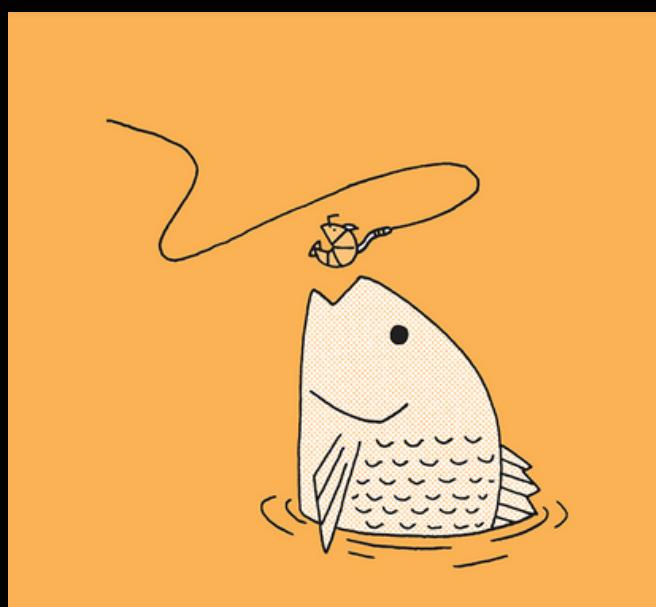
# DESPLIEGUE DEL MODELO



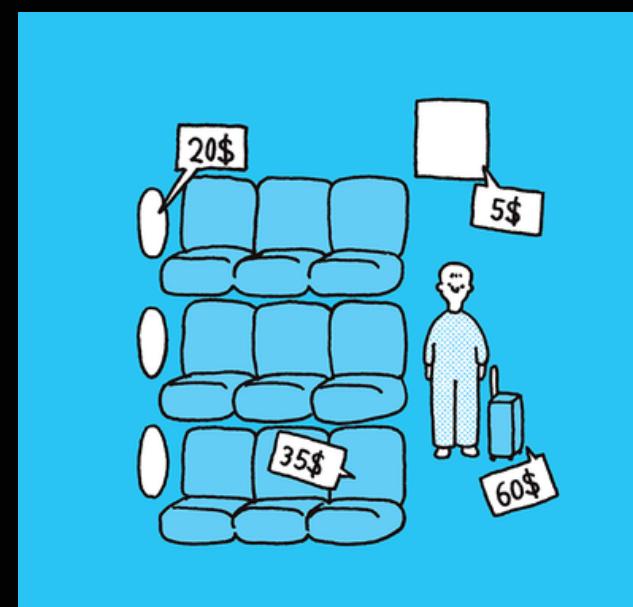
# MODELO DE NEGOCIO



- 
- 
- 
- 



RAZOR & BALDE



ADD ON

Entorno de océano rojo (alta competitividad)

Oferta de valor : Servicios adicionales al rescate del pasajero

Estrategia de atracción: Publicidad emocional

Dominio: Datos

KPI Sugerido: % de contratación de servicios de rescate por persona clasificada como 'Transported = True'



— GRACIAS