Proyecto - Titanic Espacial

MAT281 - Aplicaciones de la Matemática en Ingeniería

David Contreras

Universidad Técnica Federico Santa María

4 de Diciembre, 2023

Presentation Overview

Contexto

La nave espacial Titanic fue un transatlántico de pasajeros interestelar lanzado hace un mes. Con casi 13.000 pasajeros a bordo, la nave emprendió su viaje inaugural transportando emigrantes de nuestro sistema solar a tres exoplanetas recientemente habitables que orbitan estrellas cercanas.

Carga de datos

Se importan las librerías y datos a manejar, los cuales están explicitados a continuación.

Librerías

- Numpy
- Pandas
- Matplotlib
- Seaborn
- Sklearn
- Missingno

Datos a manejar

- Train
- Test

Estadística Descriptiva

Dimensiones

- Entrenamiento $\rightarrow$ Train:(8693,14)
- Prueba $\rightarrow$ Tes: (4277,13)

Columnas

- PassengerId: object HomePlanet: object
- CryoSleep: object Cabin: object
- Destination: object Age: float64
- VIP: object RoomService: float64
- FoodCourt: float64 ShoppingMall: float64
- Spa: float64 VRDeck: float64
- Name: object Transported: bool

Visualización Descriptiva

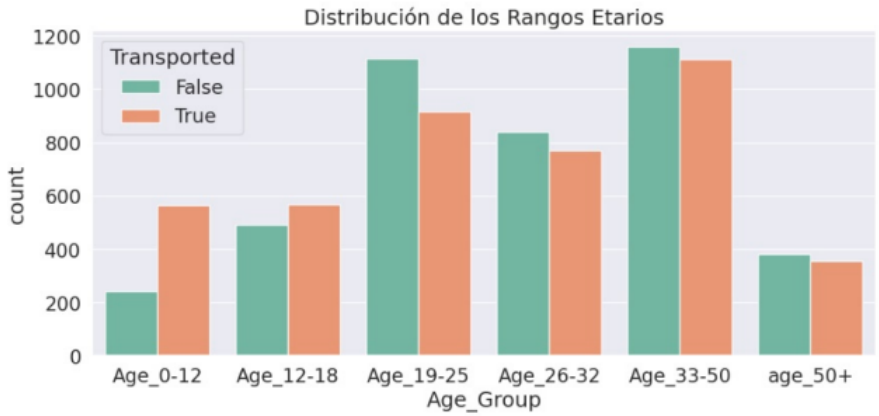


Figure: Porcenatje de Transportados

Visualización Descriptiva

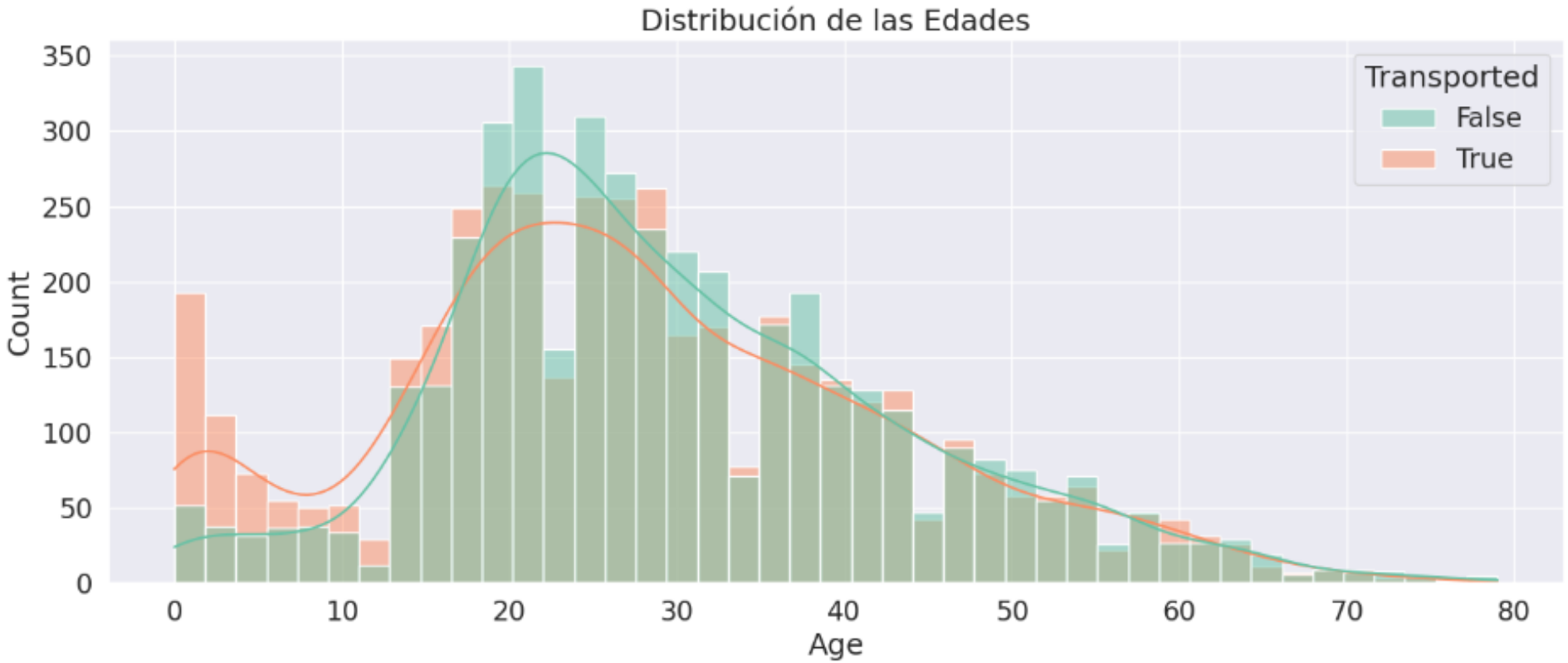


Figure: Distribución de edades

Visualización Descriptiva

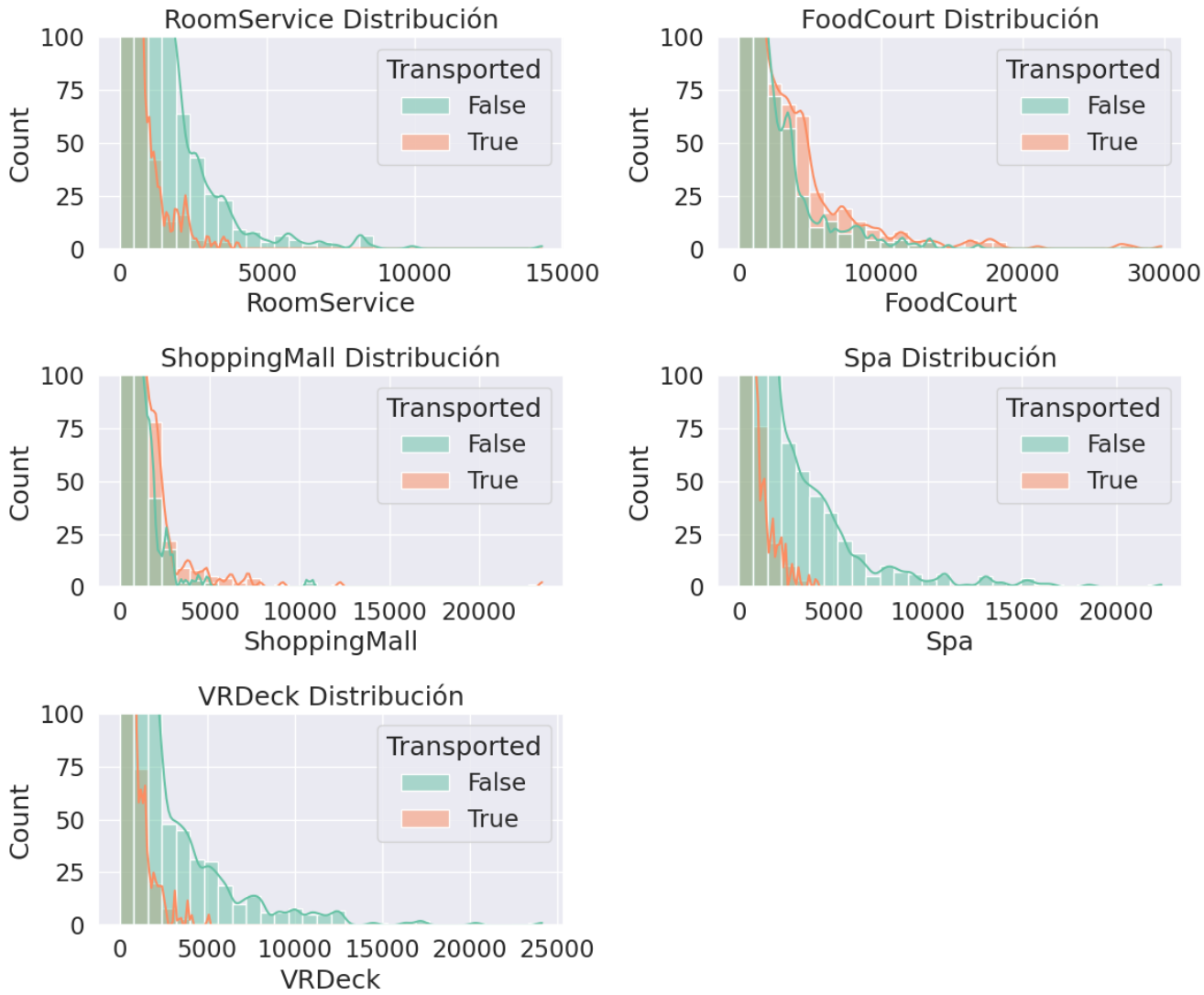


Figure: Distribución de gastos

Visualización Descriptiva

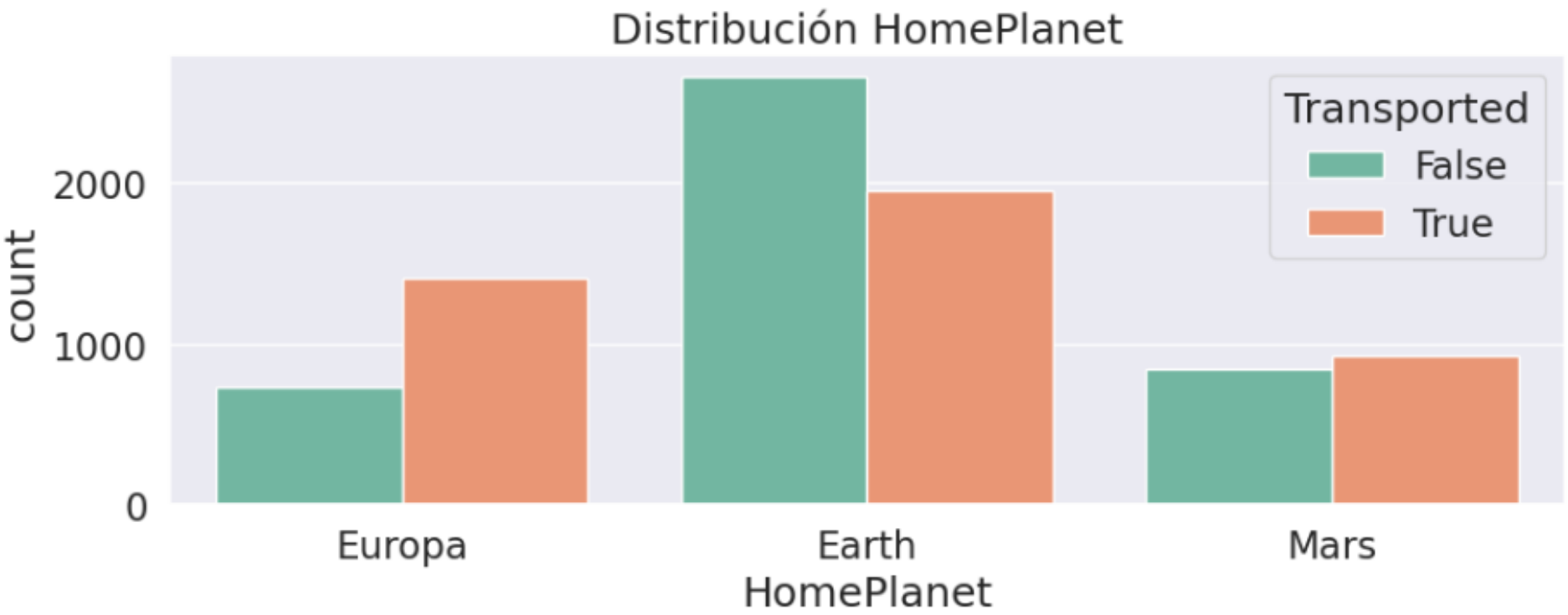


Figure: Distribución lugar de origen

Visualización Descriptiva

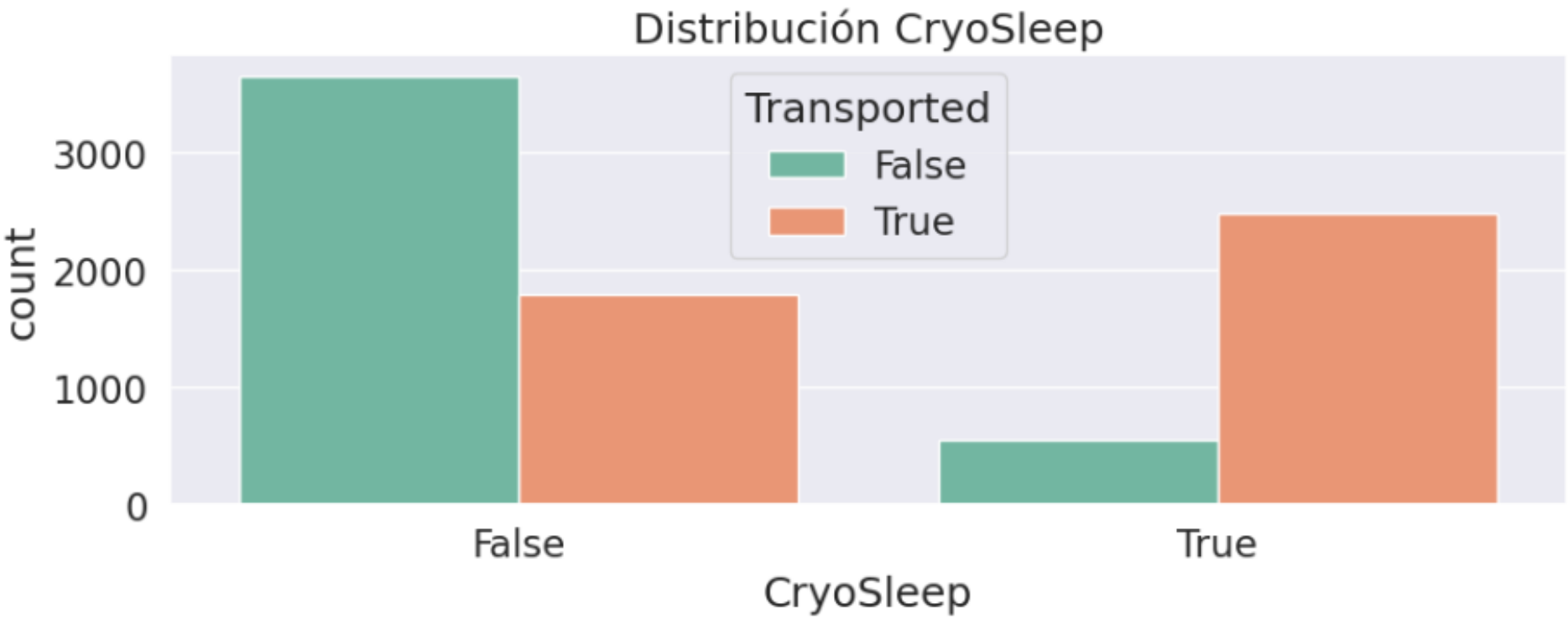


Figure: Distribución CryoSleep

Visualización Descriptiva

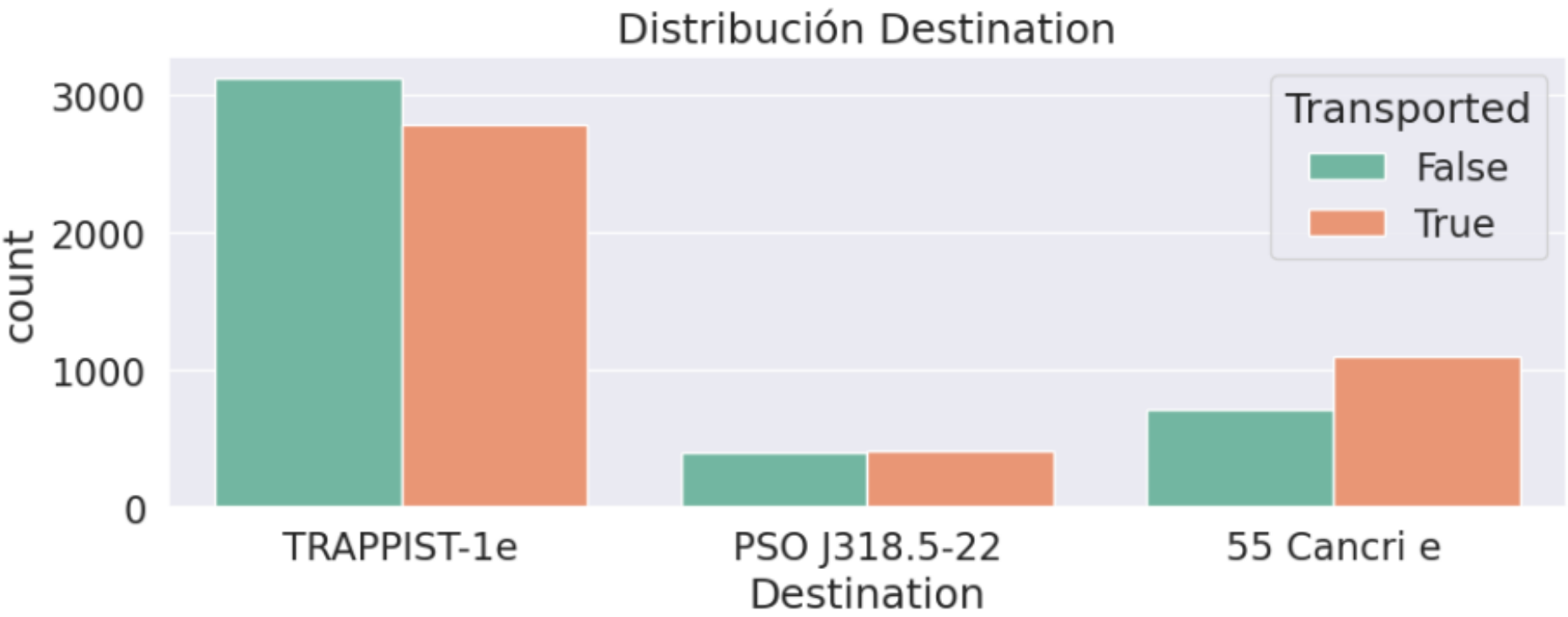


Figure: Distribución de destino

Visualización Descriptiva

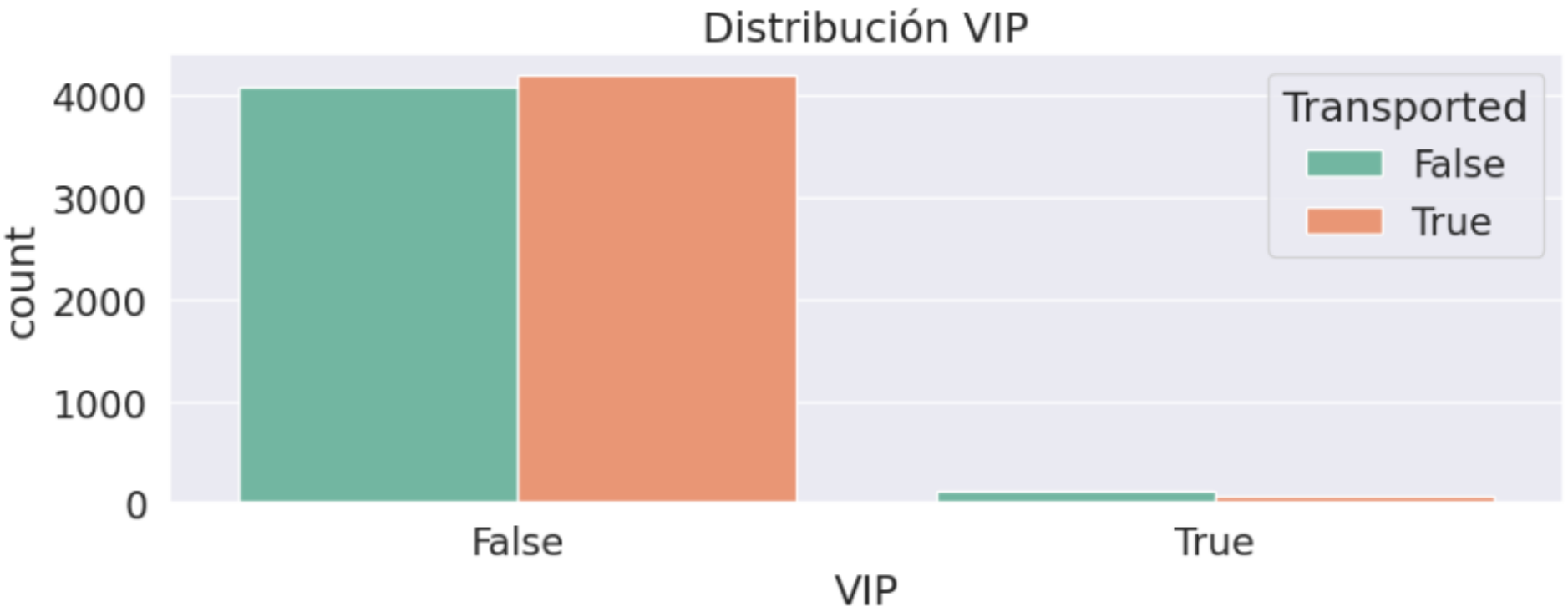


Figure: Distribución pasajeros VIP

Ingeniería de atributos

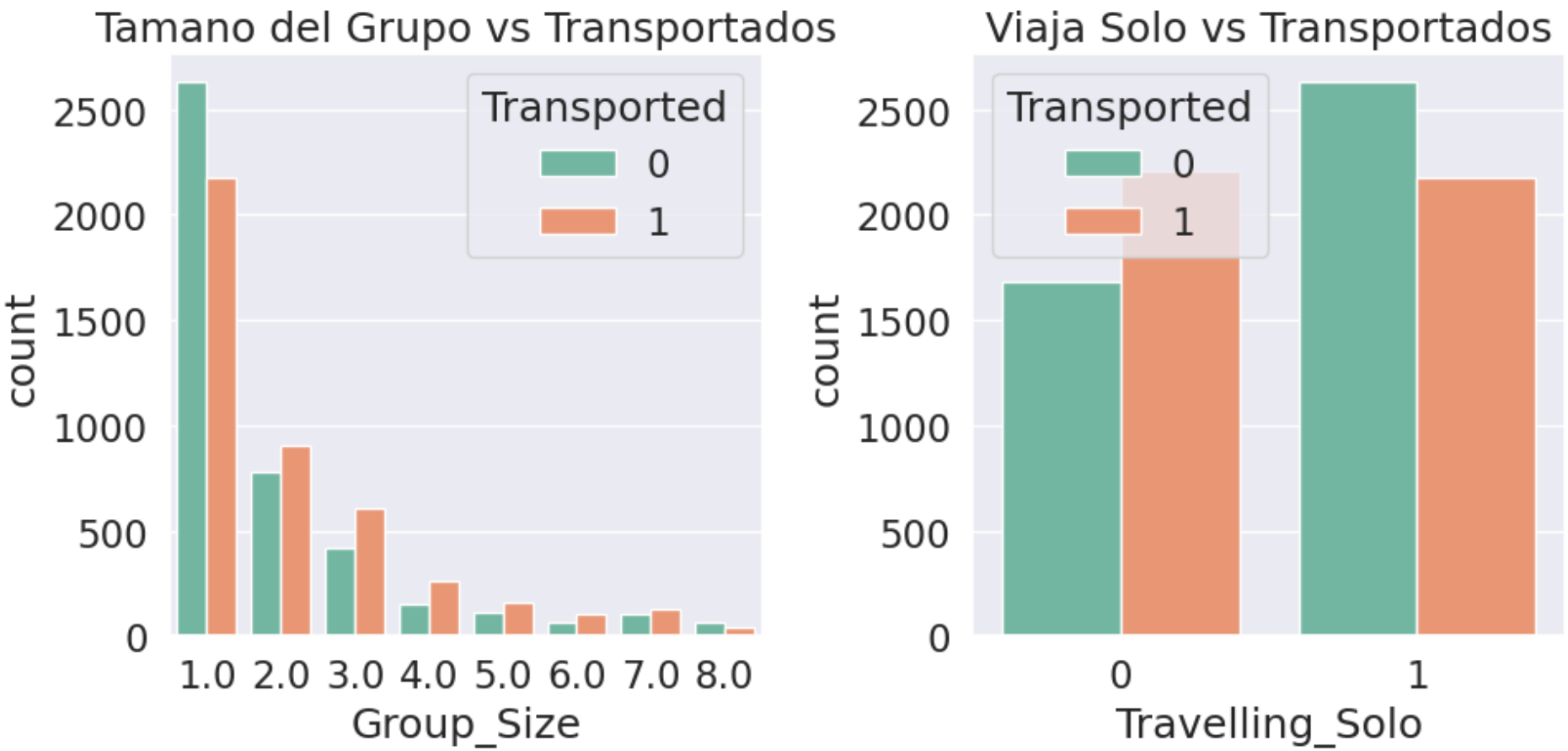


Figure: Resultados

Ingeniería de atributos

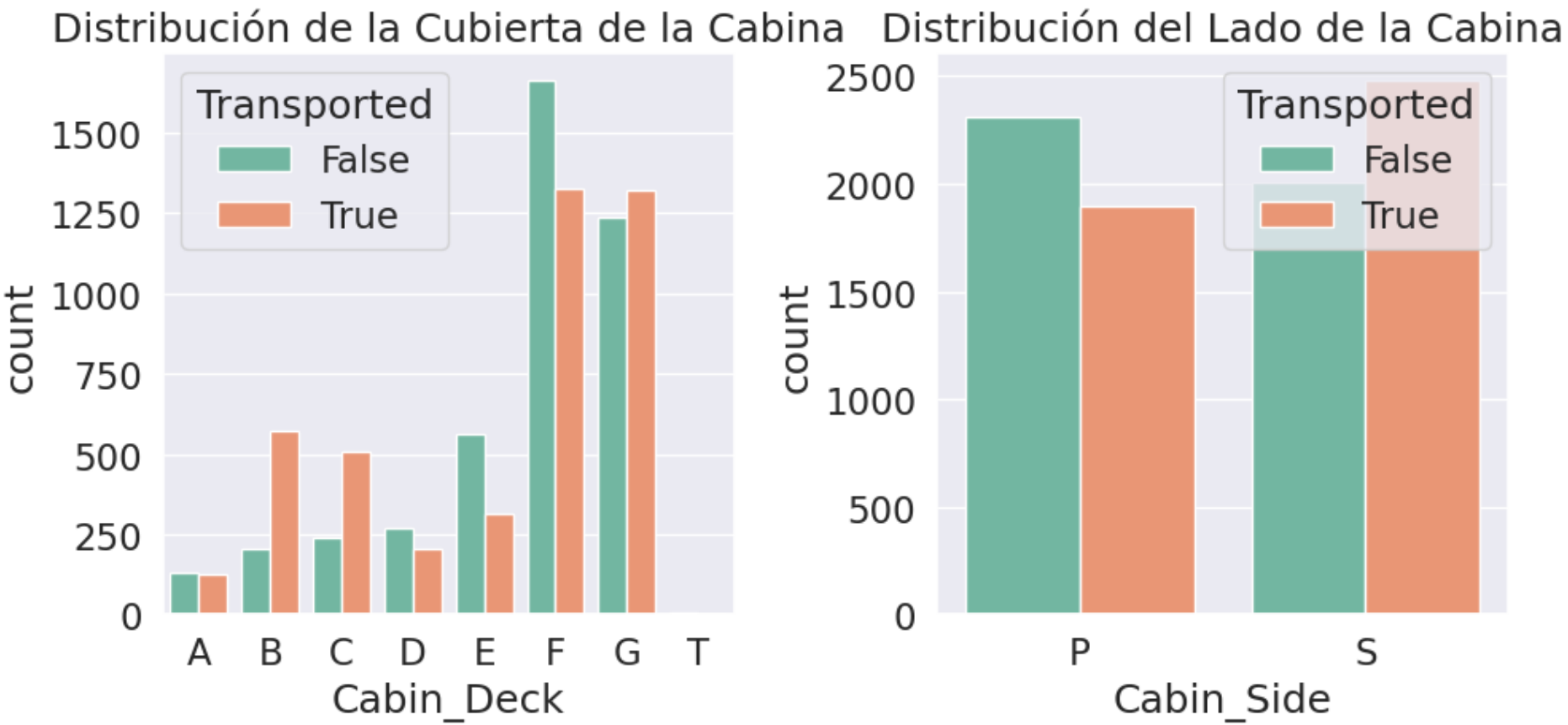


Figure: Resultados

Creación de variables

- Se distribuyen las cabinas por regiones
- Se distribuyen las edades por rangos etarios
- Se fusionan los gastos

Ingeniería de atributos

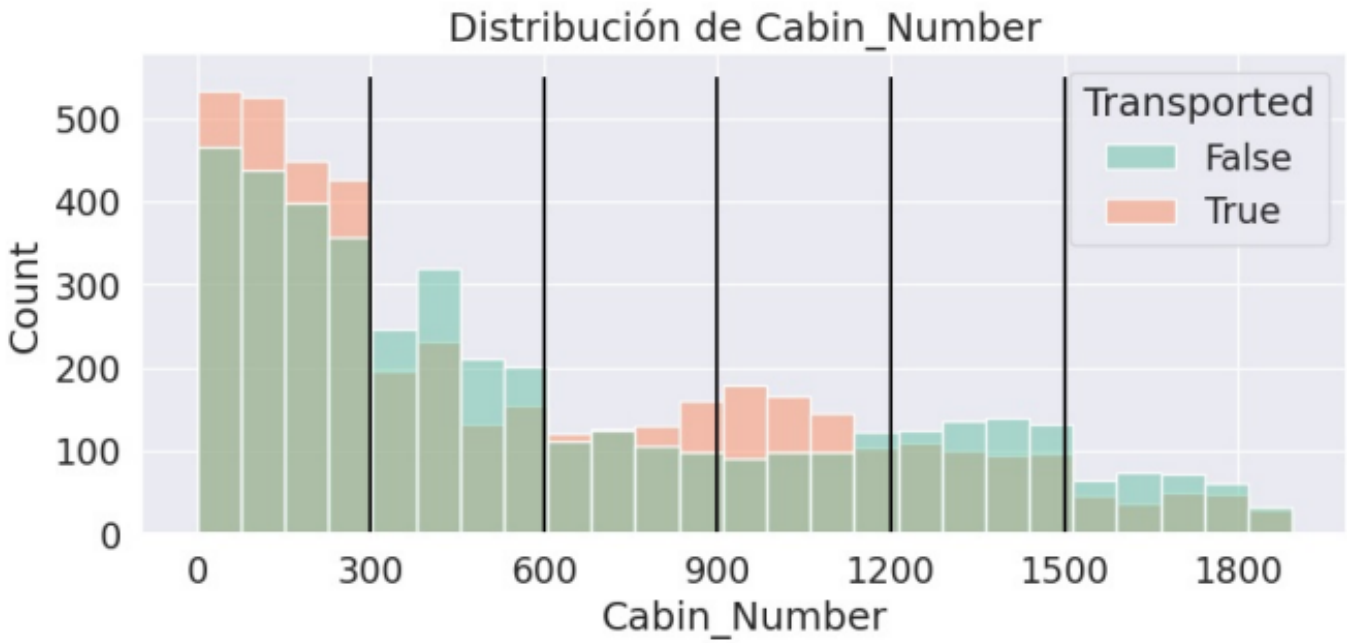


Figure: Distribución Cabina

Ingeniería de atributos

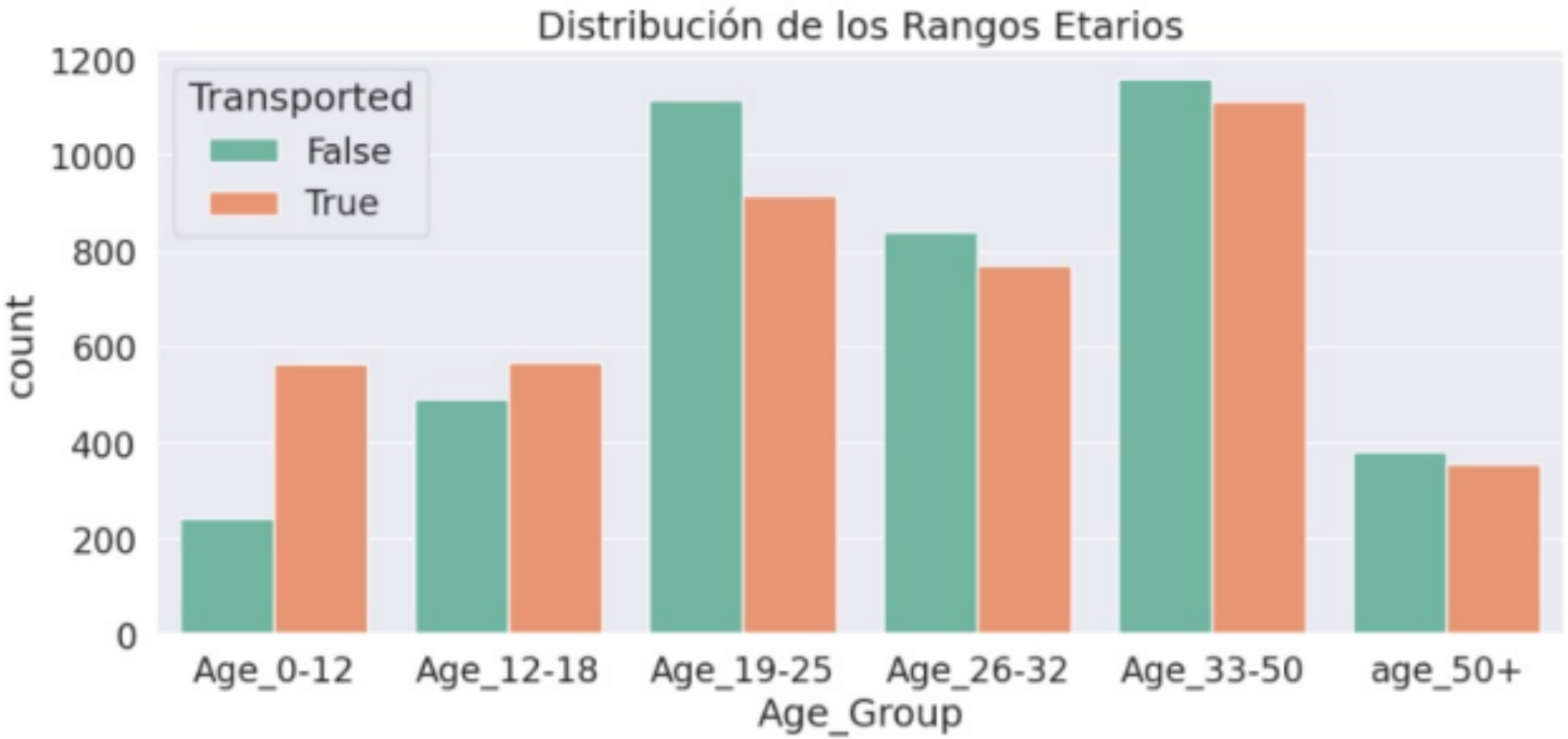


Figure: Distribución rangos etarios

Ingeniería de atributos

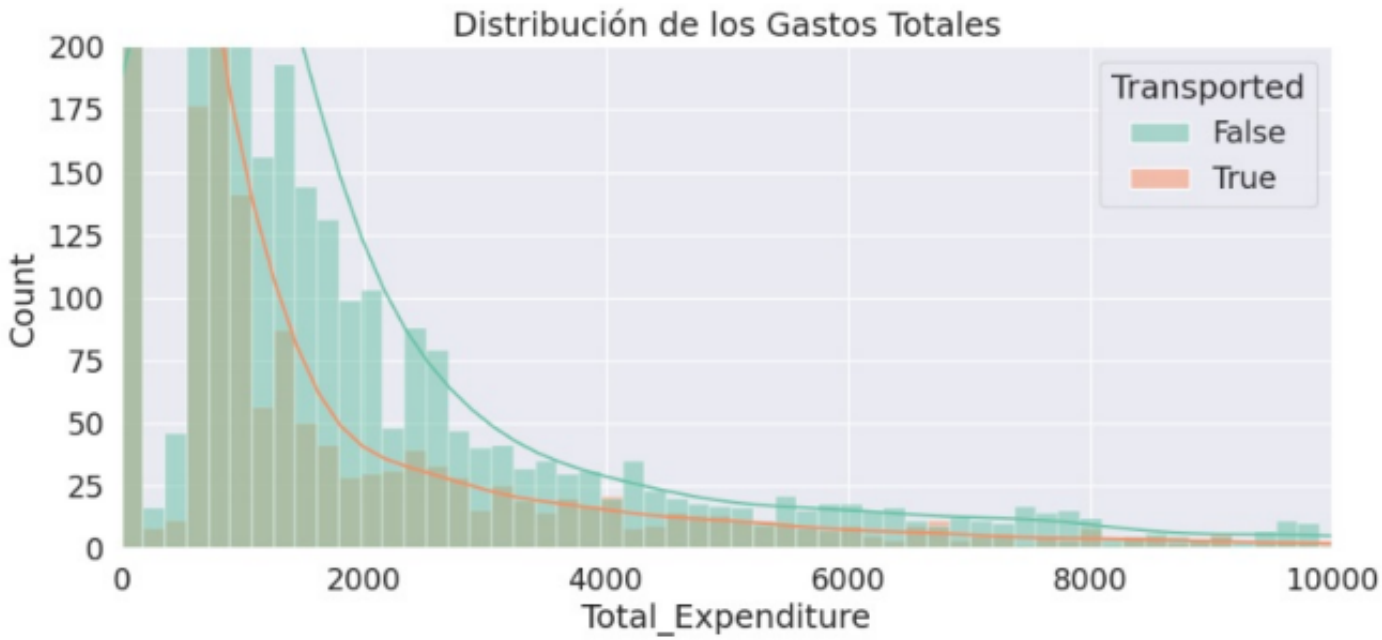


Figure: Gastos totales

Pre procesamiento

Información

- Cambiamos funciones de tipo objeto a tipo booleano
- Cambiamos funciones de tipo Int a tipo Float

Con missingno se completaron datos faltantes.

Pre procesamiento

Además se aplicó una transformación logarítmica a las variables de gasto.

- One Hot Encoding para variables categóricas nominales.
- LabelEncoding para variables categóricas ordinales.

Modelo para Datos Estandarizados

Pasos:

- Estandarizar con StandardScaler
- Train Test Split
- Crear dunción que entregará métricas

Métricas y Análisis de resultados

Regresión Logística

- Accuracy Score del conjunto de Entrenamiento es: 77.86
- Accuracy Score del conjunto de Testeo es: 77.17
- Precision Score es: 0.75
- Recall Score es: 0.80
- F1 Score es: 0.78

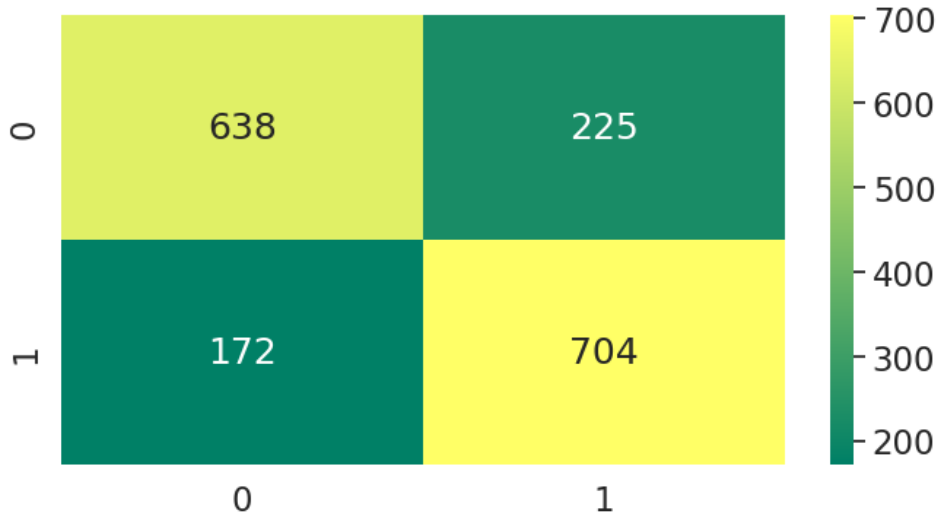


Figure: Regresión Logística

Métricas y Análisis de resultados

SVC

- Accuracy Score del conjunto de Entrenamiento es: 81.82
- Accuracy Score del conjunto de Testeo es: 79.70
- Precision Score es: 0.79
- Recall Score es: 0.79
- F1 Score es: 0.79

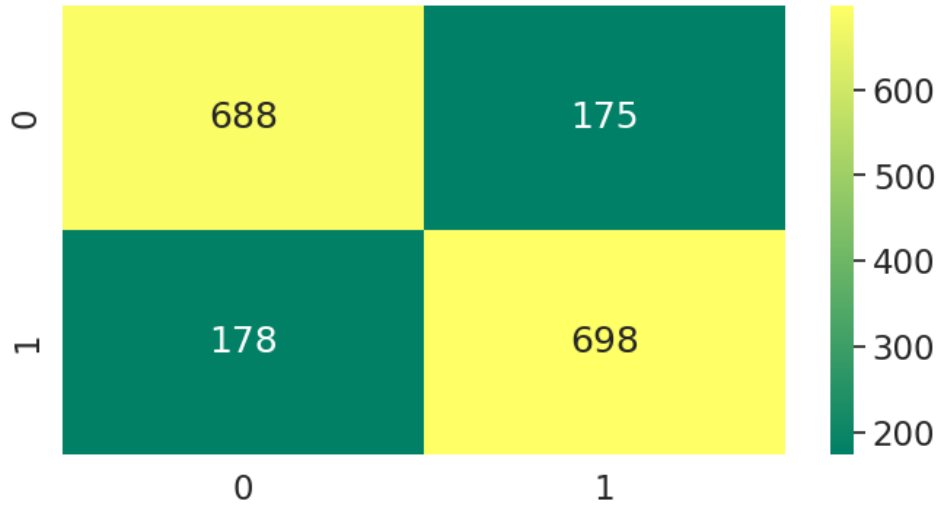


Figure: SVC

Modelo para Datos no Estandarizados

Pasos

- Train Test Split
- Creamos una función que entrega las métricas

Los modelos por aplicar son Random Forest y Gradient Boosting.

Métricas y Análisis de resultados

Random Forest

- Accuracy Score del conjunto de Entrenamiento es: 98.51
- Accuracy Score del conjunto de Testeo es: 80.56
- Precision Score es: 0.82
- Recall Score es: 0.77
- F1 Score es: 0.80

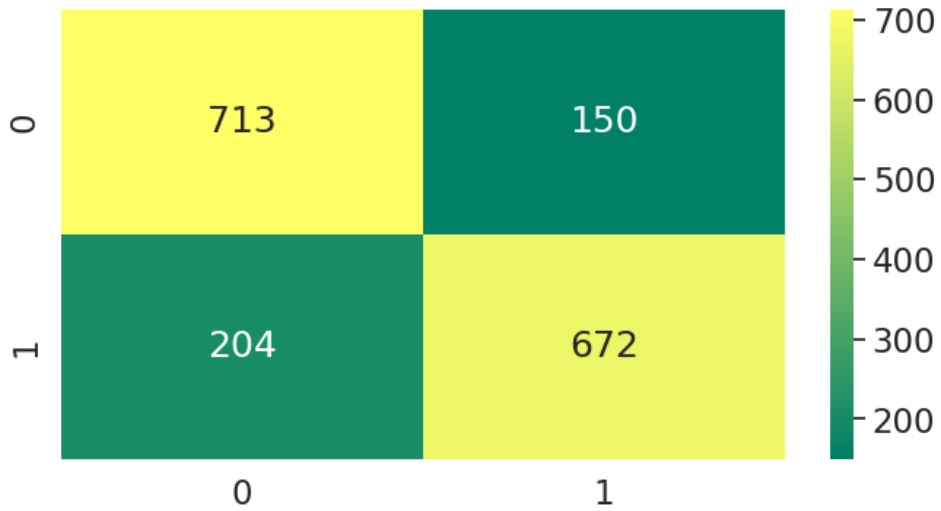


Figure: Random Forest

Métricas y Análisis de resultados

Gradient Boosting

- Accuracy Score del conjunto de Entrenamiento es: 82.05
- Accuracy Score del conjunto de Testeo es: 79.29
- Precision Score es: 0.77
- Recall Score es: 0.83
- F1 Score es: 0.80

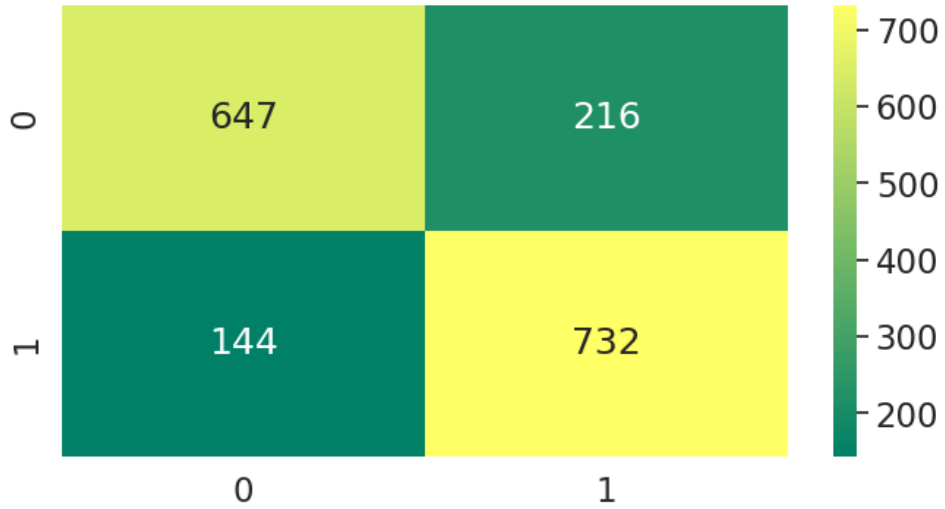
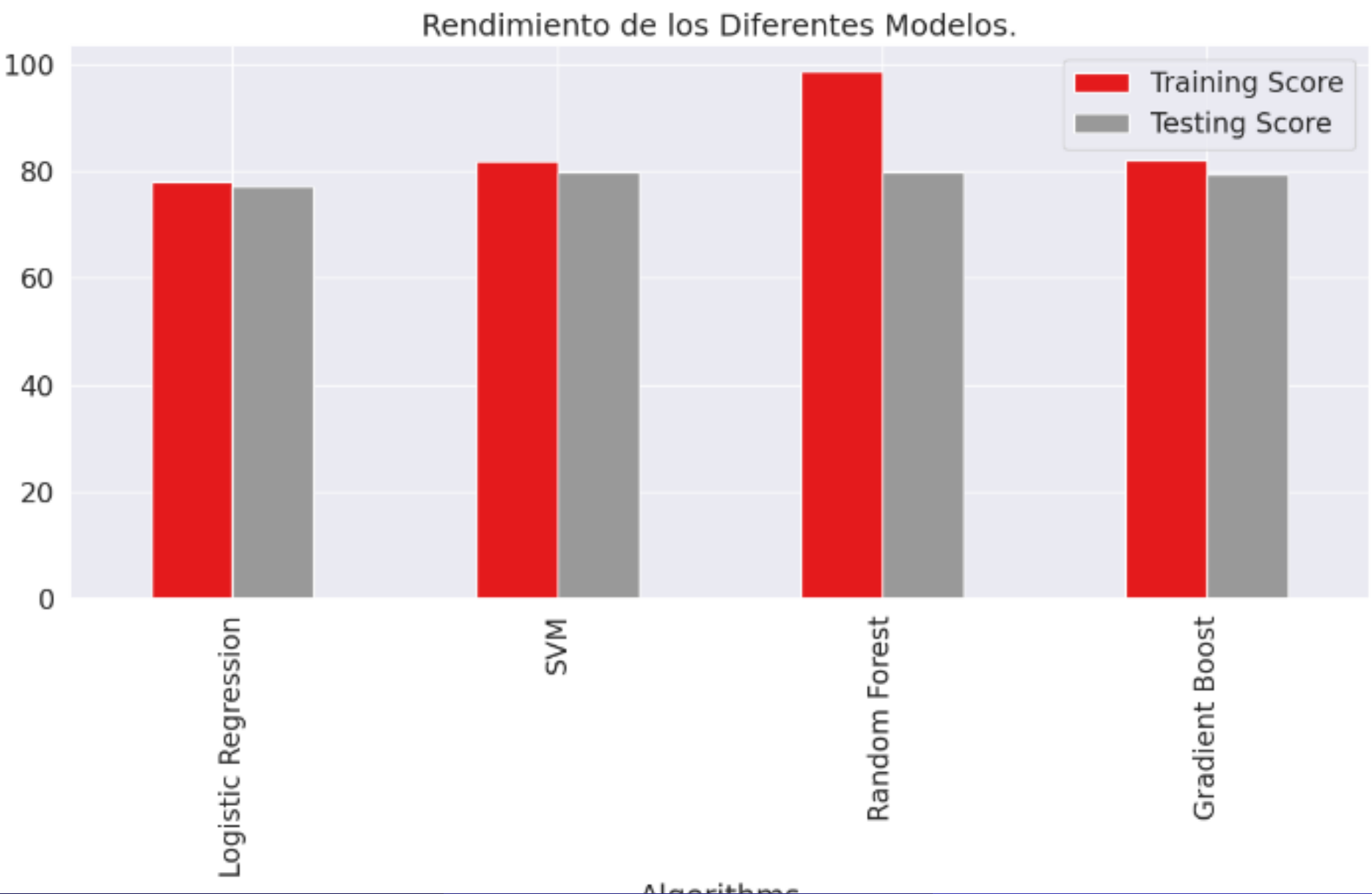


Figure: Gradient Boosting

Métrica y Análisis de resultados



Conclusiones

Modelo	Training Score	Testing Score
Regresion Logistica	77.868852	77.170788
SVM	81.823411	79.700978
Random Forest	98.518838	80.563542
Gradient Boost	82.053494	79.298447

Figure: Resultados