Spaceship Titanic

Integrantes:

Alline Ferreira Rávilla Moreira

Visão geral

O Spaceship Titanic é uma competição fictícia hospedada no <u>Kaggle</u>, inspirada no famoso dataset Titanic, mas ambientada em uma nave espacial. Essa nave espacial, viajava com destino a três exoplanetas recém-habitáveis que orbitam estrelas próximas ao nosso sistema solar. Mas houve um acidente e que levou a metade dos 13.000 passageiros serem transportados para uma dimensão alternativa. Para ajudar a equipe de resgate a recuperar os passageiros perdidos, temos o objetivo de prever qual passageiro foi transportado para outra dimensão (Transported = True/False).

Objetivos e Metas



Preparação dos dados

Meta 2

Elaborar um algoritmo que seja capaz de identificar o destino dos passageiros.

Meta 3

Mostrar os resultados da execução

Descrição dos dados

Treino: 8.693 linhas × 14 colunas

Teste: 4.277 linhas × 13 colunas

Variáveis numéricas: Age, RoomService, FoodCourt, ShoppingMall, Spa, VRDeck

Variáveis categóricas: HomePlanet, CryoSleep, Cabin, Destination, VIP Target: Transported (booleano)

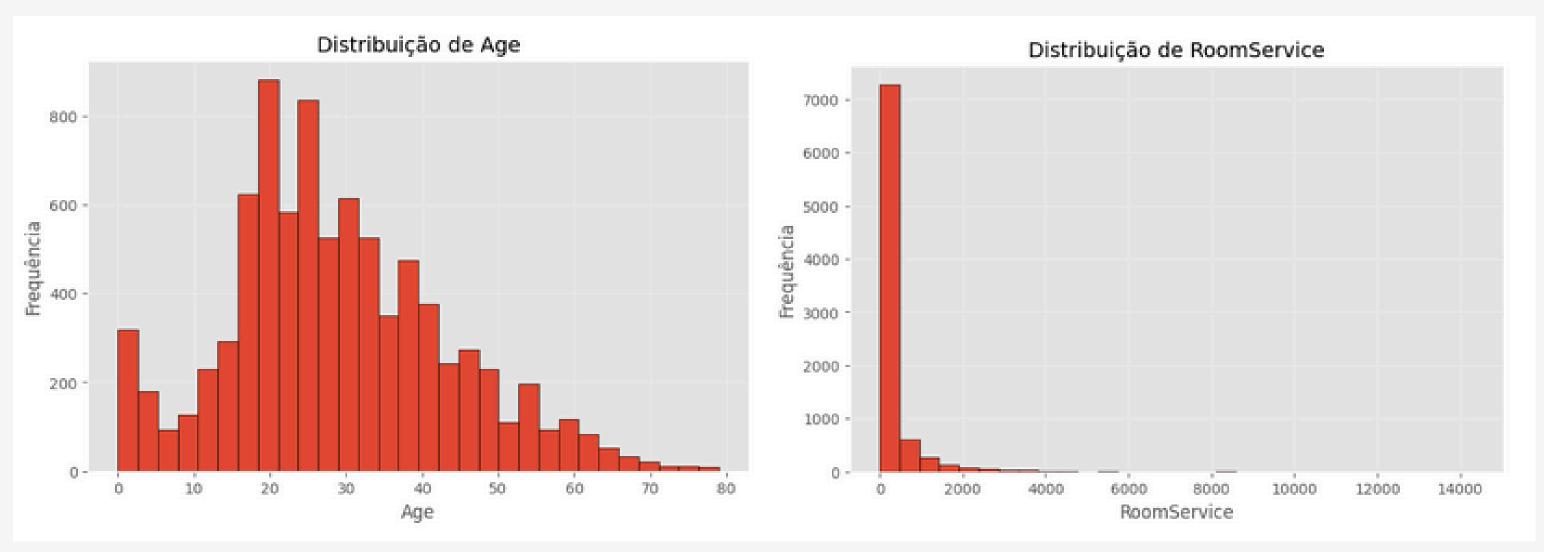
Valores ausentes:

Variável	% Missing	
Cabin	20%	
HomePlanet	2%	
Destination	2%	
Age	5%	
RoomService	7%	
FoodCourt	7%	
ShoppingMall	7%	
Spa	7%	
VRDeck	7%	

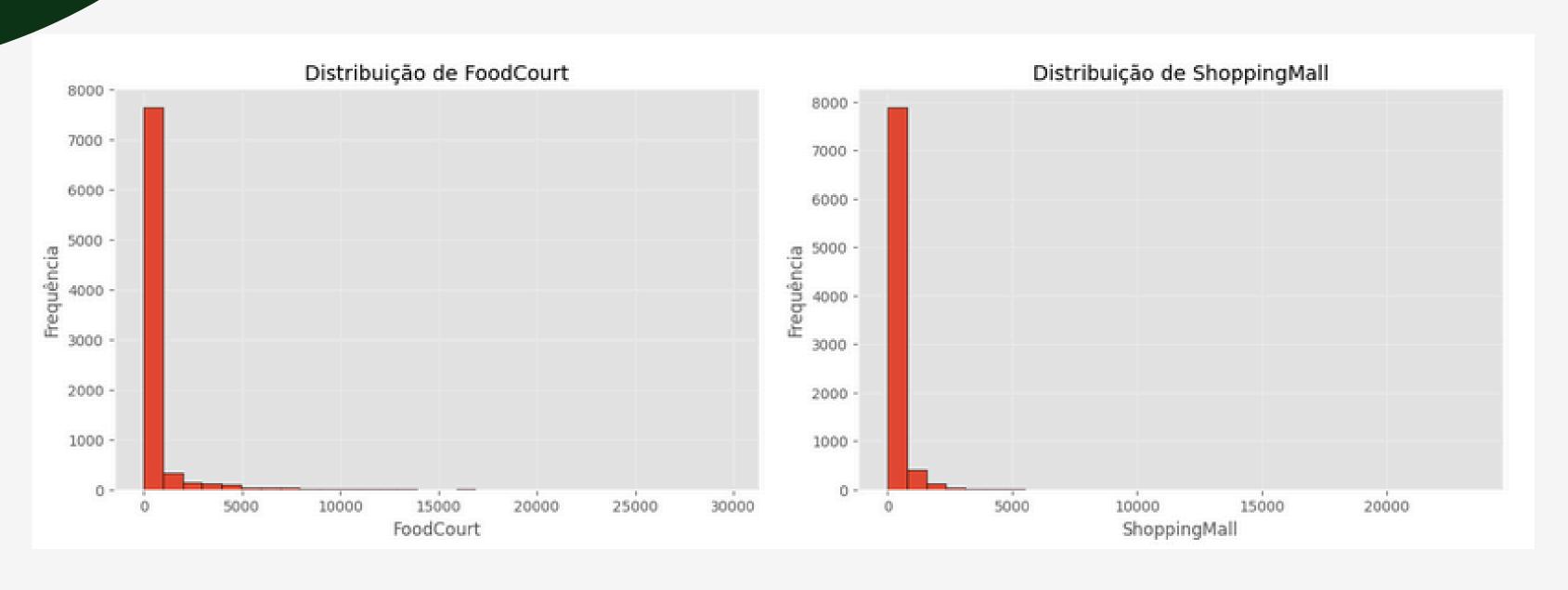
Análise Exploratória

- Idade: maioria entre 20 e 40 anos.
- Gastos (RoomService, FoodCourt, ShoppingMall, Spa e VRDeck): altamente assimétricos, com muitos passageiros gastando pouco ou nada.

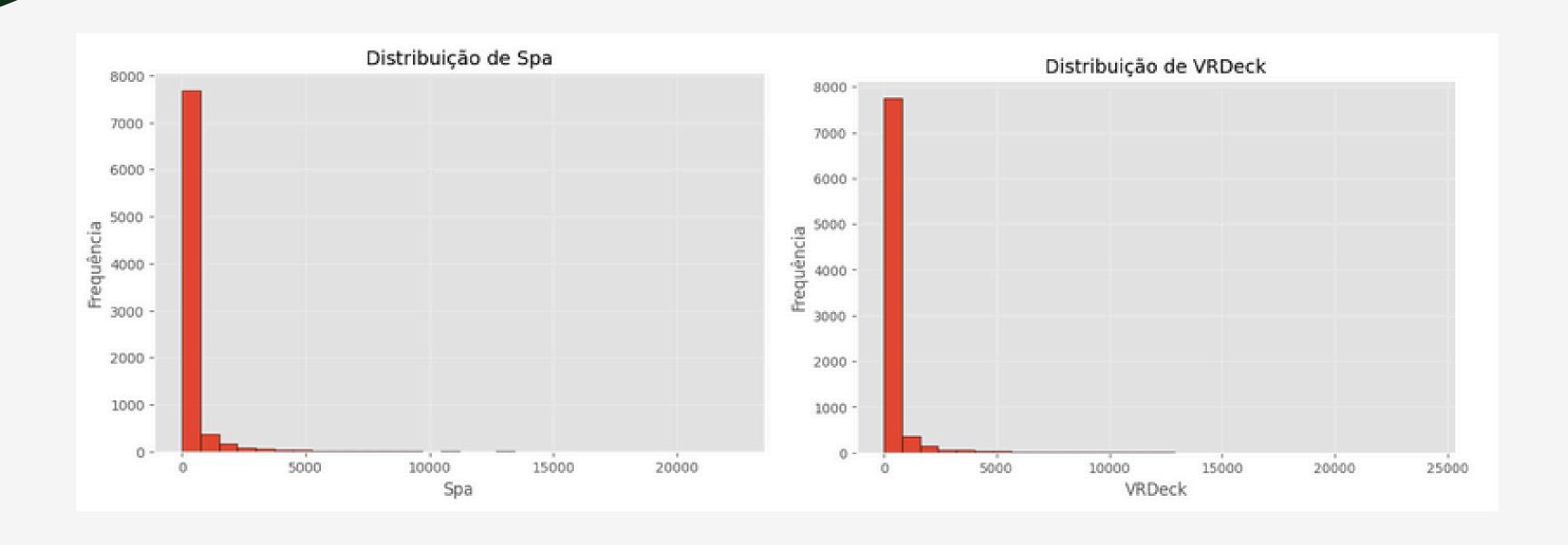
Distribuição das idades e Gastos com RomService



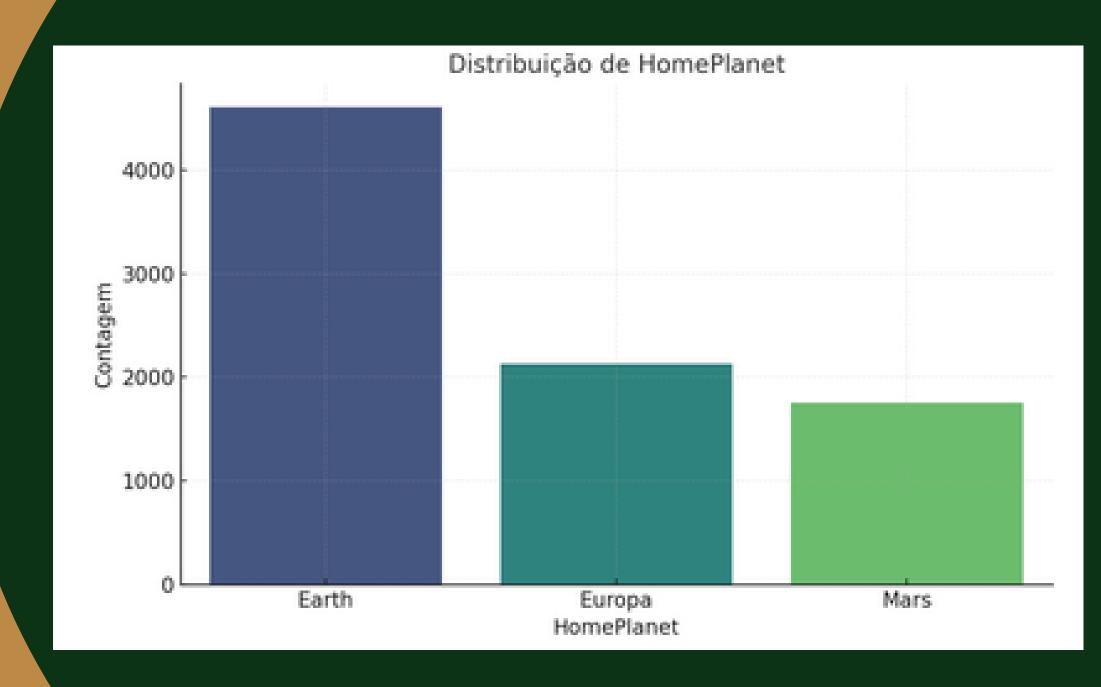
Gastos com FoodCourt e ShoppingMall



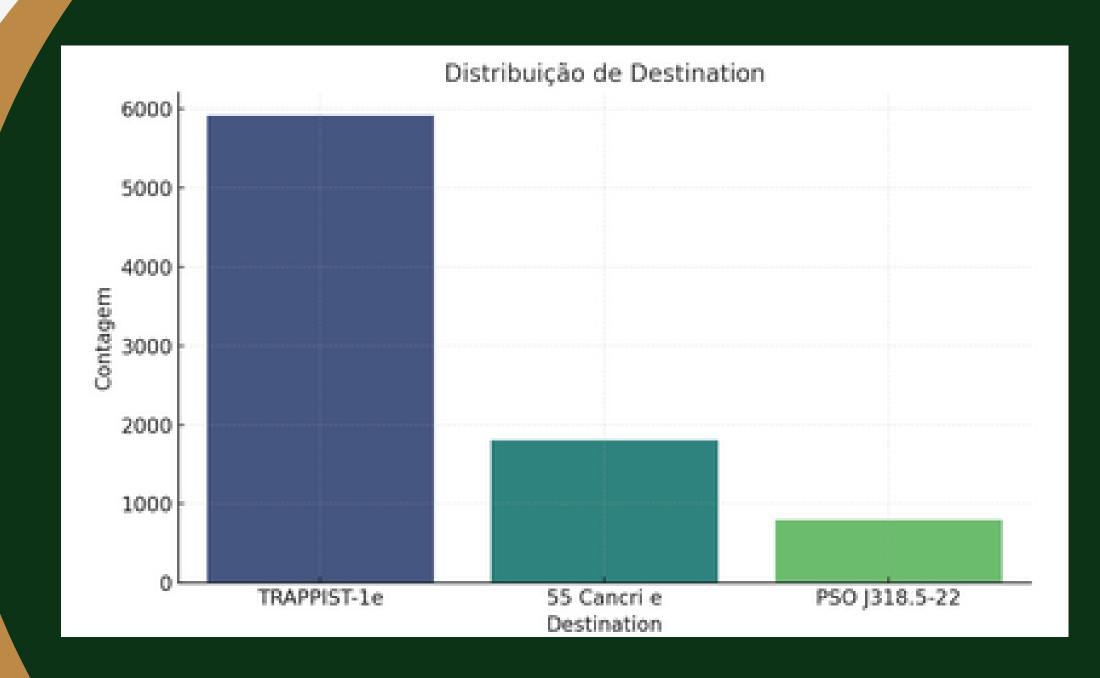
Gastos com Spa e VRDeck



 HomePlanet: maioria de passageiros vem de Earth.



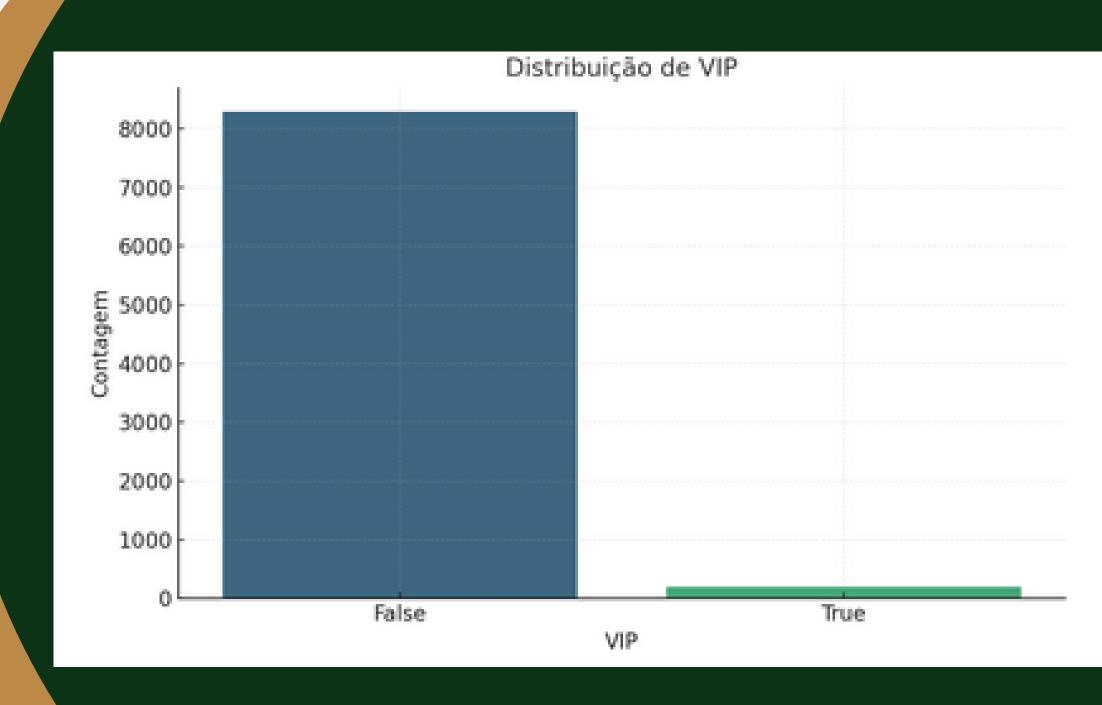
Destination: destino mais comum é TRAPPIST-1e.



CryoSleep: cerca de 30% dos passageiros estavam em sono criogênico.

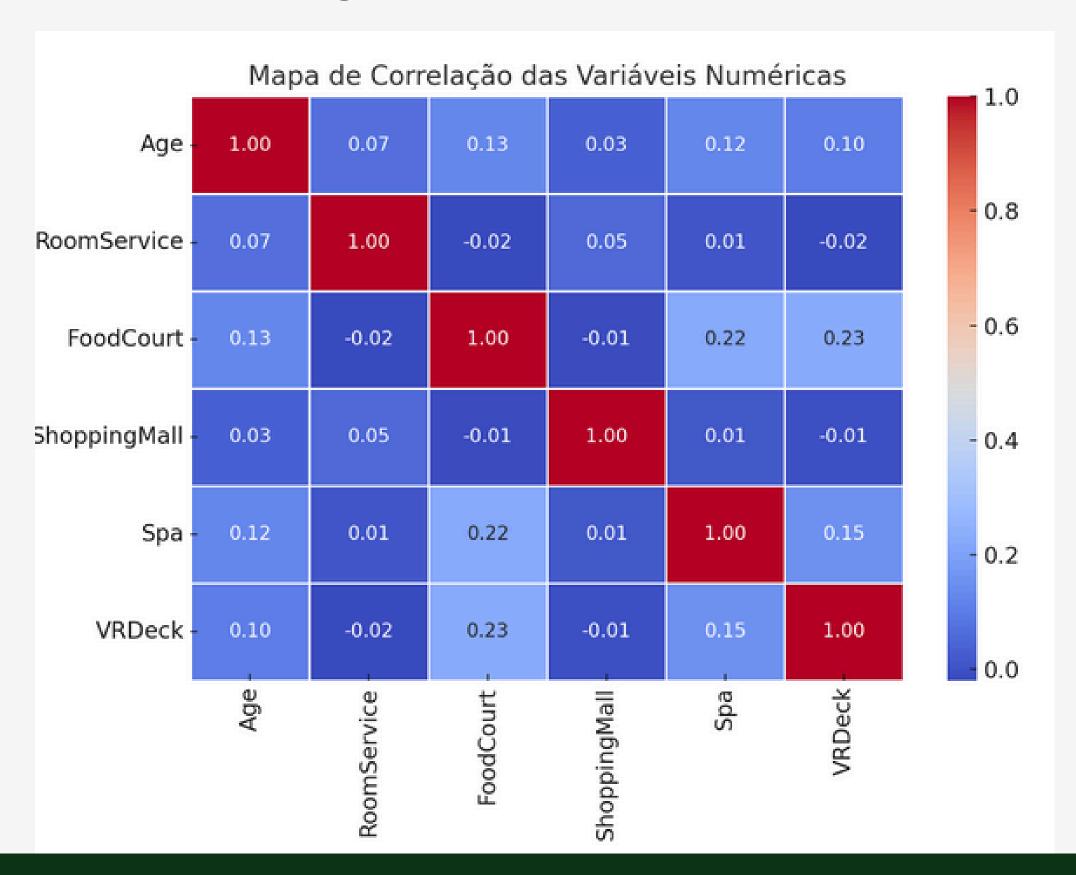


VIP: apenas uma pequena fração dos passageiros era VIP.



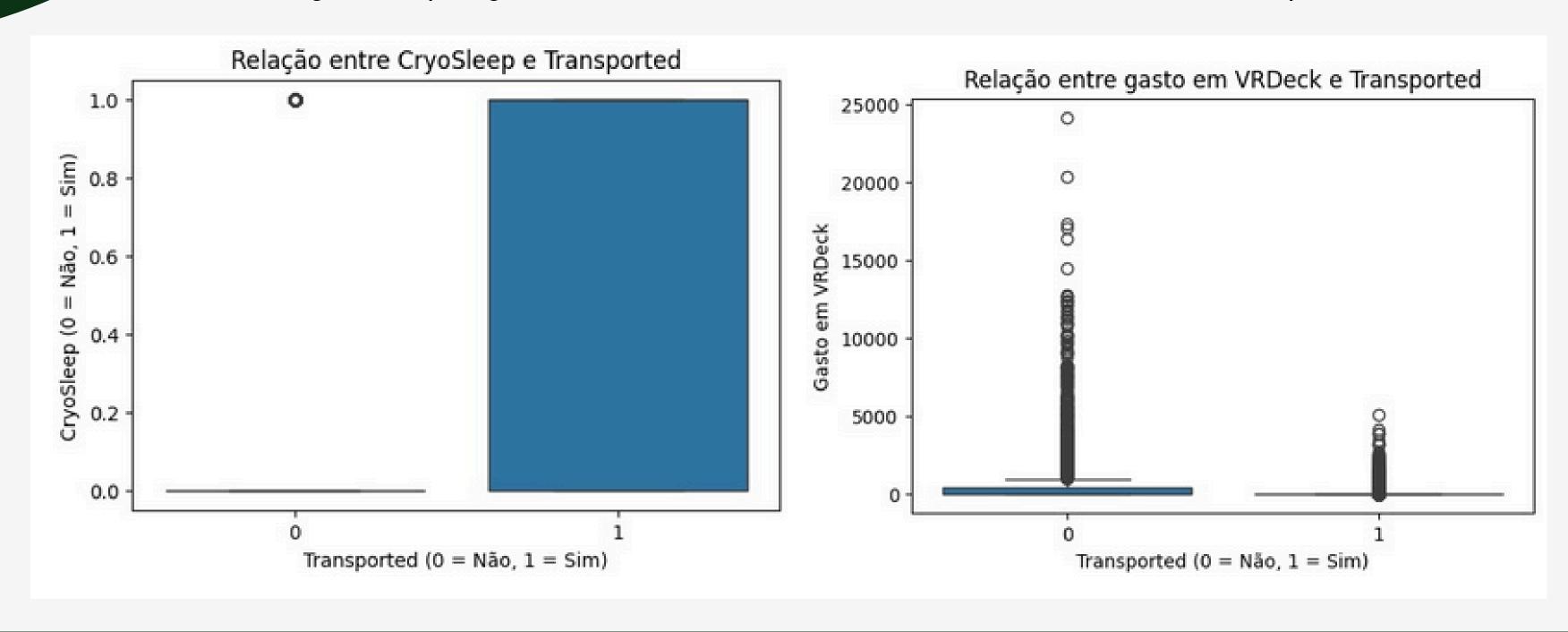
Correlação

- Gastos em serviços (Spa, VRDeck) têm correlação positiva moderada entre si.
- Idade não apresenta correlação forte com o target.



Relação com Target

- Passageiros em CryoSleep têm maior probabilidade de serem transportados.
- Passageiros que gastaram mais em VRDeck tendem a não ser transportados.



Trabalhos semelhantes

Os autores ZHUCUI JING e XIAOLI YIN, em seu o artigo: "Neural Network-Based Prediction Model for Passenger Flow in a Large Passenger Station: An Exploratory Study" apresenta a aplicação de redes neurais para prevê a entrada e saída de passageiros de uma estação, com objetivo de fornecer a garantia de segurança da estação, alocação de recursos e distribuição de pessoal.

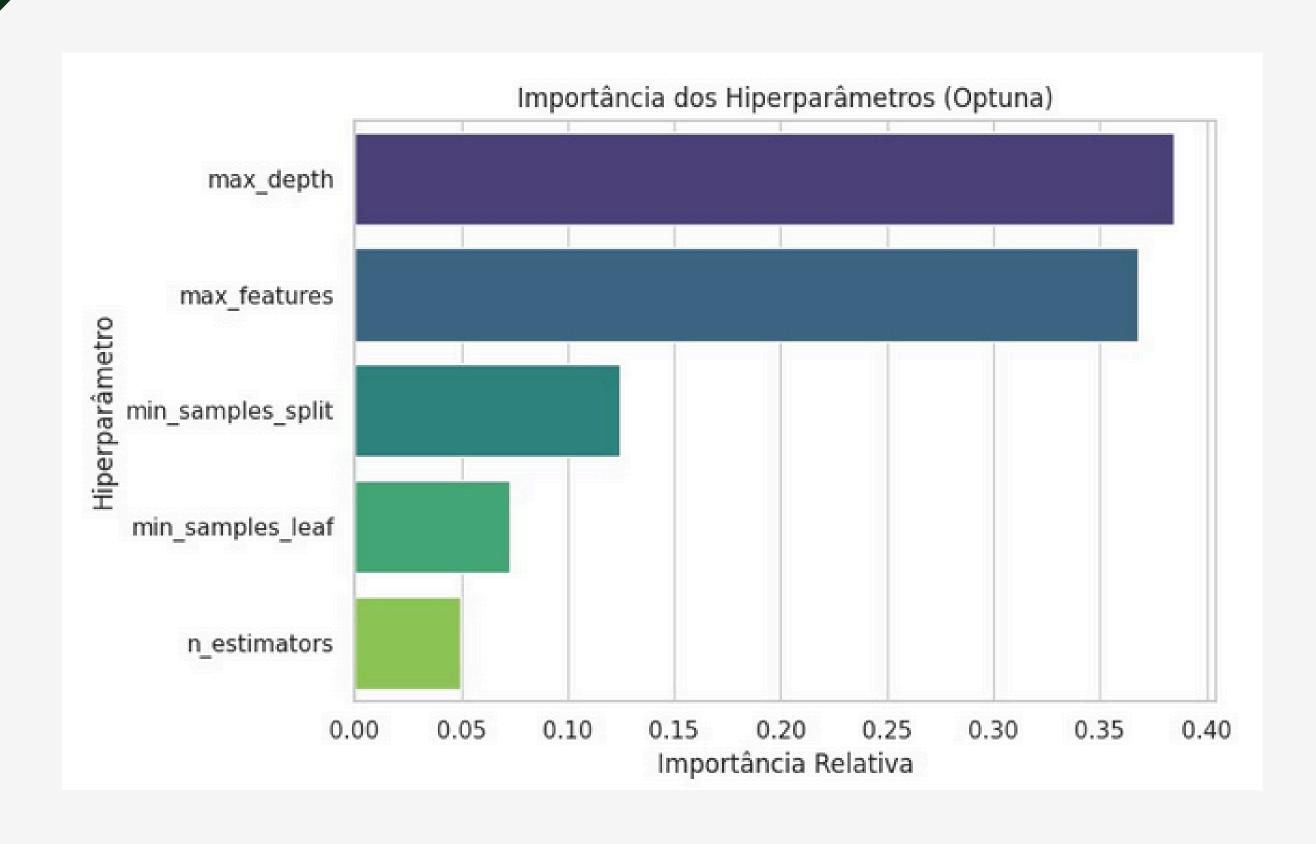
No artigo "Prospects and challenges of Metaverse application in data-driven intelligent transportation systems", de acordo com os autores Judith Nkechinyere Njoku, Cosmas Ifeanyi Nwakanma, Gabriel Chukwunonso Amaizu, Dong-Seong Kim, o Metaverso é um conceito usado para referir a um mundo virtual que existe paralelamente ao mundo físico.

Modelos Testados

Modelo	Accuracy (CV)	
RandomForest	0.78	
XGBoost	0.79	
KNN	0.74	

• VotingClassifier (soft voting) → Accuracy: 0.80

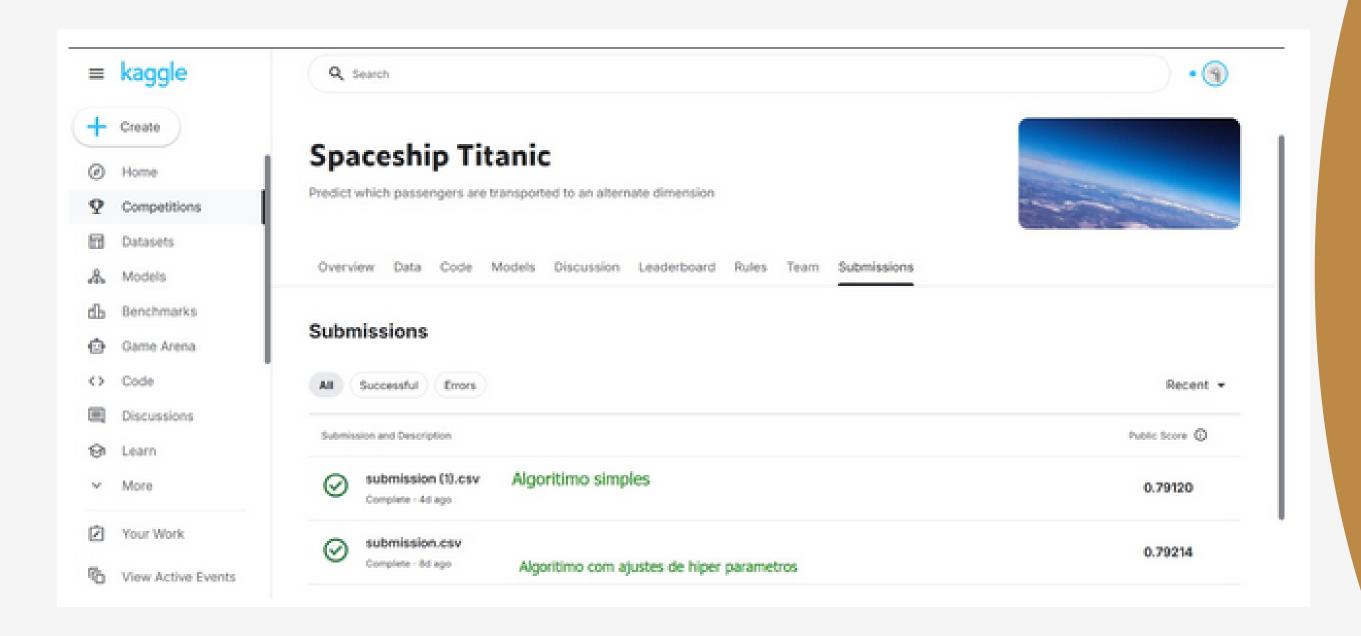
Otimização de Hiperparâmetros



Resultados

Modelo	Accuracy (CV)	Melhor Parâmetro	Observações
Baseline	0.50	_	Referência mínima
RandomForest	0.78	_	Bom desempenho
XGBoost	0.79	_	Estável
KNN	0.74	_	Sensível ao escalonamento
VotingClassifier	0.80	_	Ensemble de 3 modelos
RandomForest Tunado	0.81	Optuna params	Melhor resultado

Resultados



Conclusão

- O RandomForest Tunado apresentou o melhor desempenho (Accuracy = 0.81).
- O VotingClassifier também foi competitivo, mostrando que ensembles podem ajudar.
- Ao submeter-mos na competição o modelo RandomForest e RandomForest Tunado, notamos que a potuação teve pouca diferença mas o tempo de execução foi maior para o RandomForest.

Obrigado a todos .

Dúvidas?

