

# Projeto Aplicado II



Análise de dados para a agência de viagens "Boa Viagem"

<https://github.com/OhashiMarina/Projeto-Aplicado-II>

Andrei Souza de Oliveira - TIA: 22520600

Daniele dos Santos Rosa - TIA: 22510631

Gabriela Ohashi de Souza - TIA: 22521097

Marina Ohashi de Souza - TIA: 22520971

Miguel Maurício T. Pitali da Silva - TIA: 22507310



# Introdução

A agência de viagens "Boa Viagem" está empenhada em melhorar seus serviços e aumentar a satisfação do cliente por meio da análise de dados relacionados às reservas de passagens aéreas. Com um extenso conjunto de dados que abrange informações sobre passageiros, voos, destinos e comportamento de reserva, o projeto visa aplicar técnicas como análise descritiva, segmentação de clientes, análise de tendências temporais, previsão de demanda por destino e avaliação do desempenho de voos.





# Premissas

Com uma bordagem data-driven, permite que a "Boa Viagem" tome melhores decisões, com propósito de antecipar as necessidades dos clientes e se destacar no cenário dinâmico do setor.

Airline dataset, disponível em:

<https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset>



# Objetivos e Metas

01

Análise de Demanda por Destino: Analisar a demanda por destinos específicos com base nas reservas de passagens aéreas, identificando os destinos mais populares e os segmentos de mercado mais relevantes.

02

Segmentação de Clientes: Segmentar os clientes com base em critérios como idade, gênero, nacionalidade e preferências de viagem para oferecer pacotes de viagens mais personalizados.

03

Análise de Desempenho de Voos: Avaliar o desempenho de voos com base no status do voo e identificar áreas de melhoria na eficiência das operações de voo.

04

Análise de Desempenho de Voos: Avaliar o desempenho de voos com base no status do voo e identificar áreas de melhoria na eficiência das operações de voo.

# Aquisição do Dataset

## Base de Dados:

<https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset>

01

Shape: 98619 linhas, 15 colunas

02

235 países avaliados

03

6 Continentes

# Análise Exploratória dos Dados



## Avaliação Primária

### Identificou-se as colunas:

**Passenger ID:** identificação de cada passageiro

**First Name:** Primeiro nome do passageiro

**Last Name:** Último nome do passageiro

**Gender:** Gênero

**Age:** Idade do passageiro

**Nationality:** Nacionalidade do passageiro

**Airport Name:** Nome do aeroporto

**Airport Country Code:** Código do país do aeroporto

**Country Name:** Nome do país que o voo pertence

**Airport Continent:** Abreviação do continente que País pertence

**Continents:** Continente do País pertence

**Depature Date:** Data de partida do voo

**Arrival Airport:** Abreviação do aeroporto de chegada

**Pilot Name:** Nome do piloto responsável pelo voo

**Flight Status:** Situação/condição do voo

# Análise Exploratória dos Dados



## Insights

Através de linhas de códigos para geração de gráficos, conseguimos identificar algumas informações importantes para a agência de viagens “Boa Viagem.

As demandas por “Continentes” e pelos principais “Países” de destino foram investigadas, com a utilização de gráficos de barras, proporcionando insights sobre as preferências dos passageiros e auxiliando nas estratégias de marketing e expansão.



Demanda por continentes



Demanda pelos 10 principais destinos (país)



10 Melhores Desempenhos por País de Destino (valor absoluto)

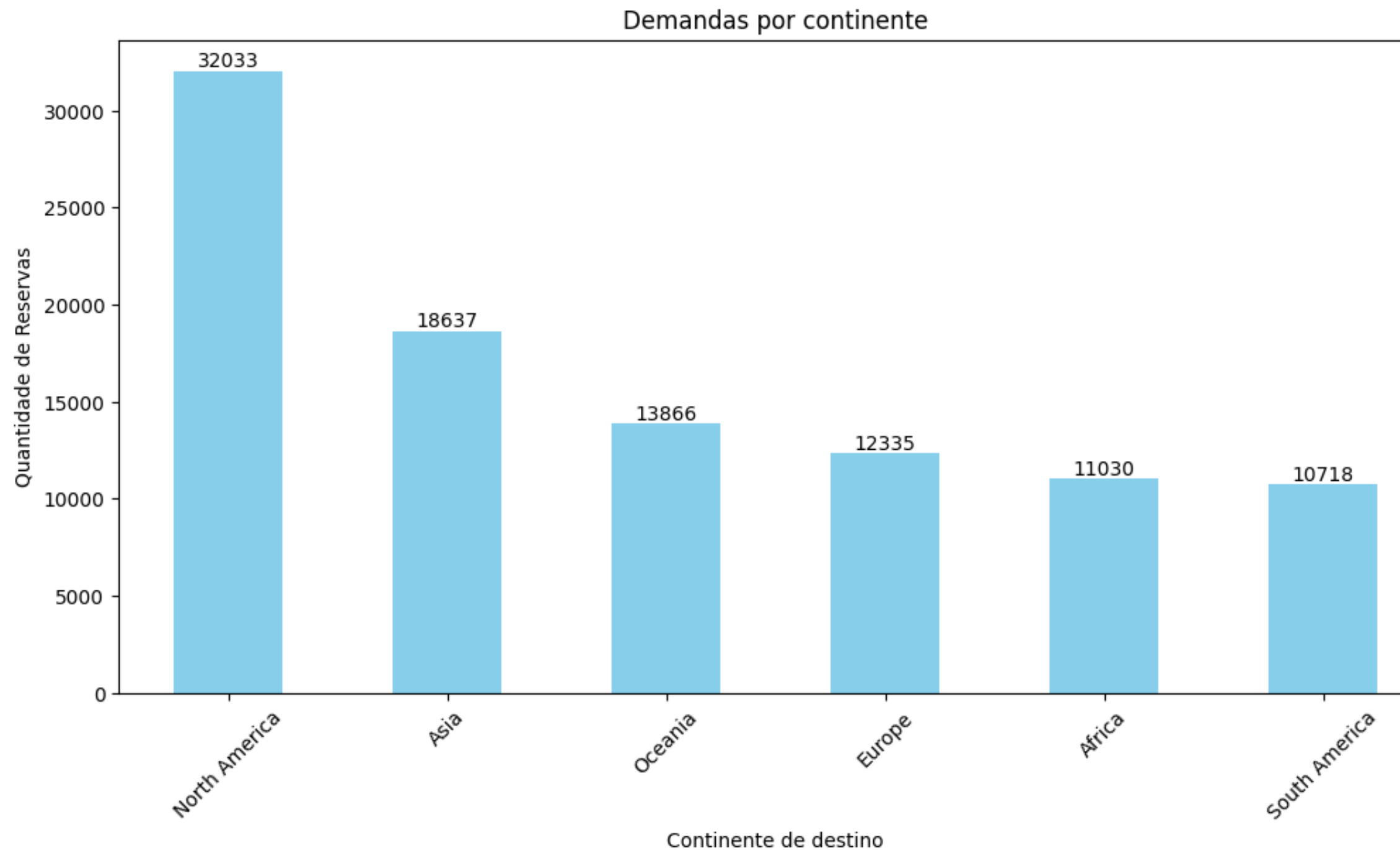


Eficiência das Operações



Piores Desempenhos por País de Destino

# Insights

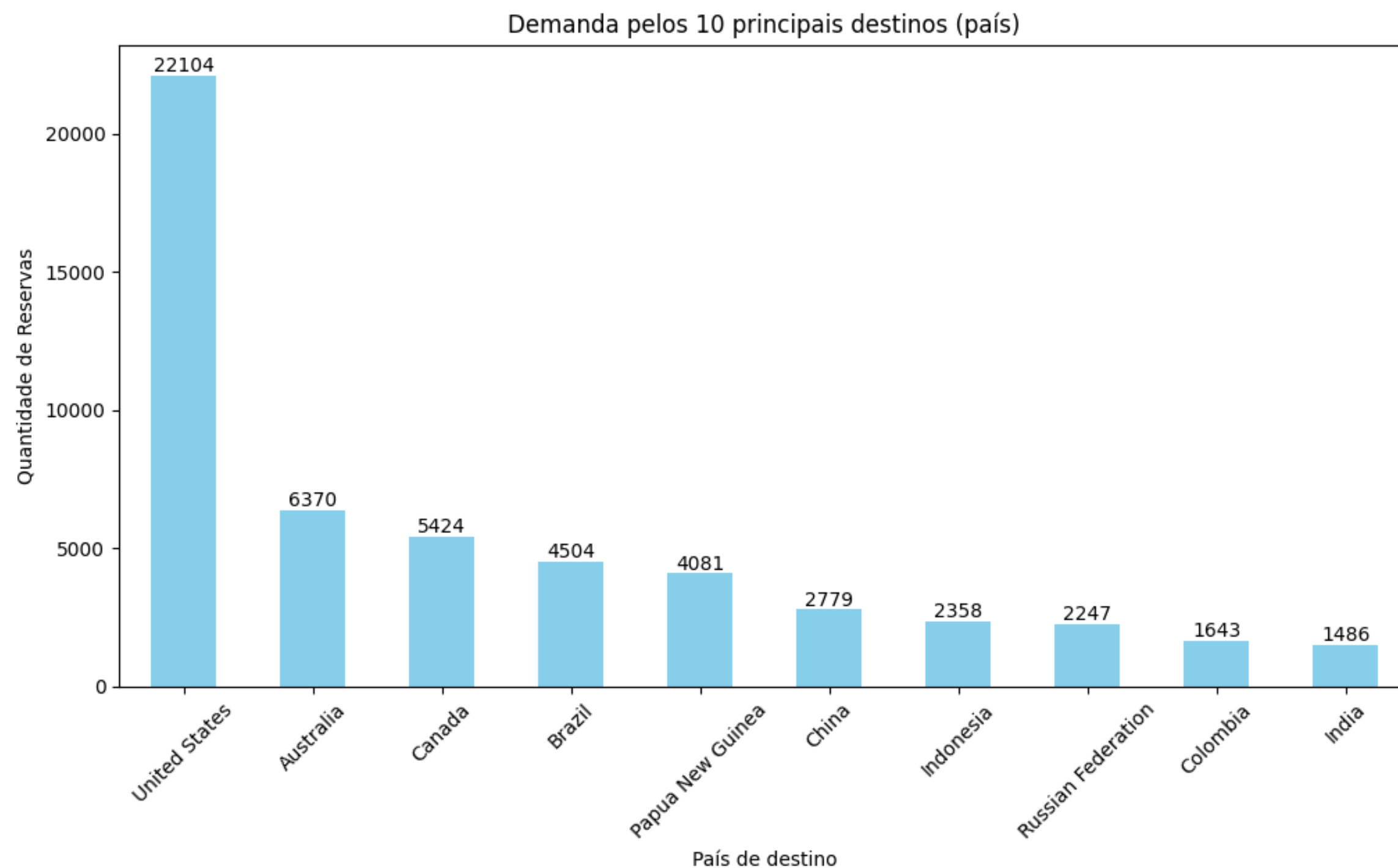


## Insights

Através dos dados e gráfico notamos que o Continente de maior demanda para a agência "Boa viagem" é North America (América do Norte).



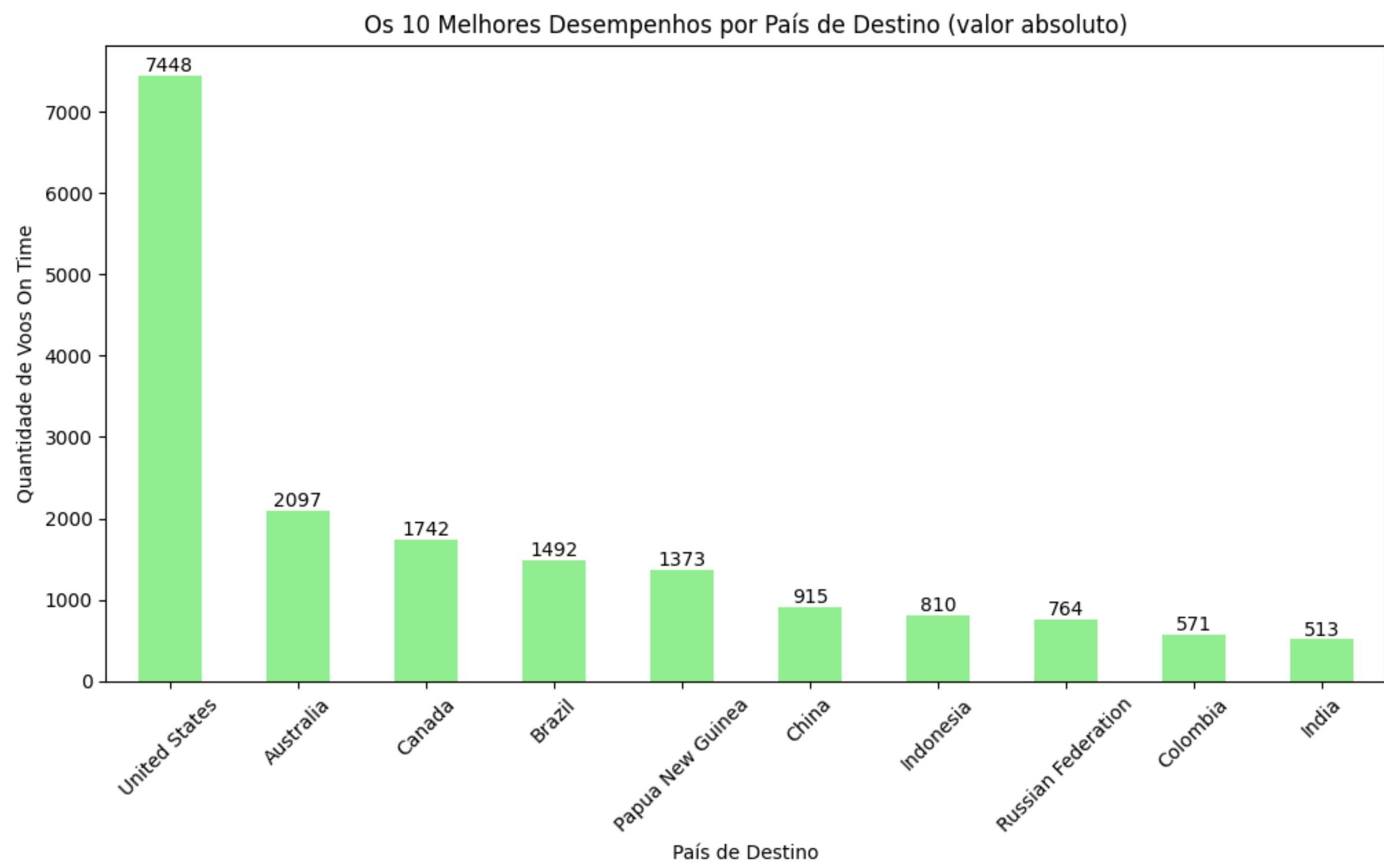
# Insights



## Insights

O principal destino dos voos da “Boa viagem” no aspecto país é Estados Unidos. O top 3 destinos é composto por dois países do continente Norte Americano (líder em demanda no aspecto continetal)

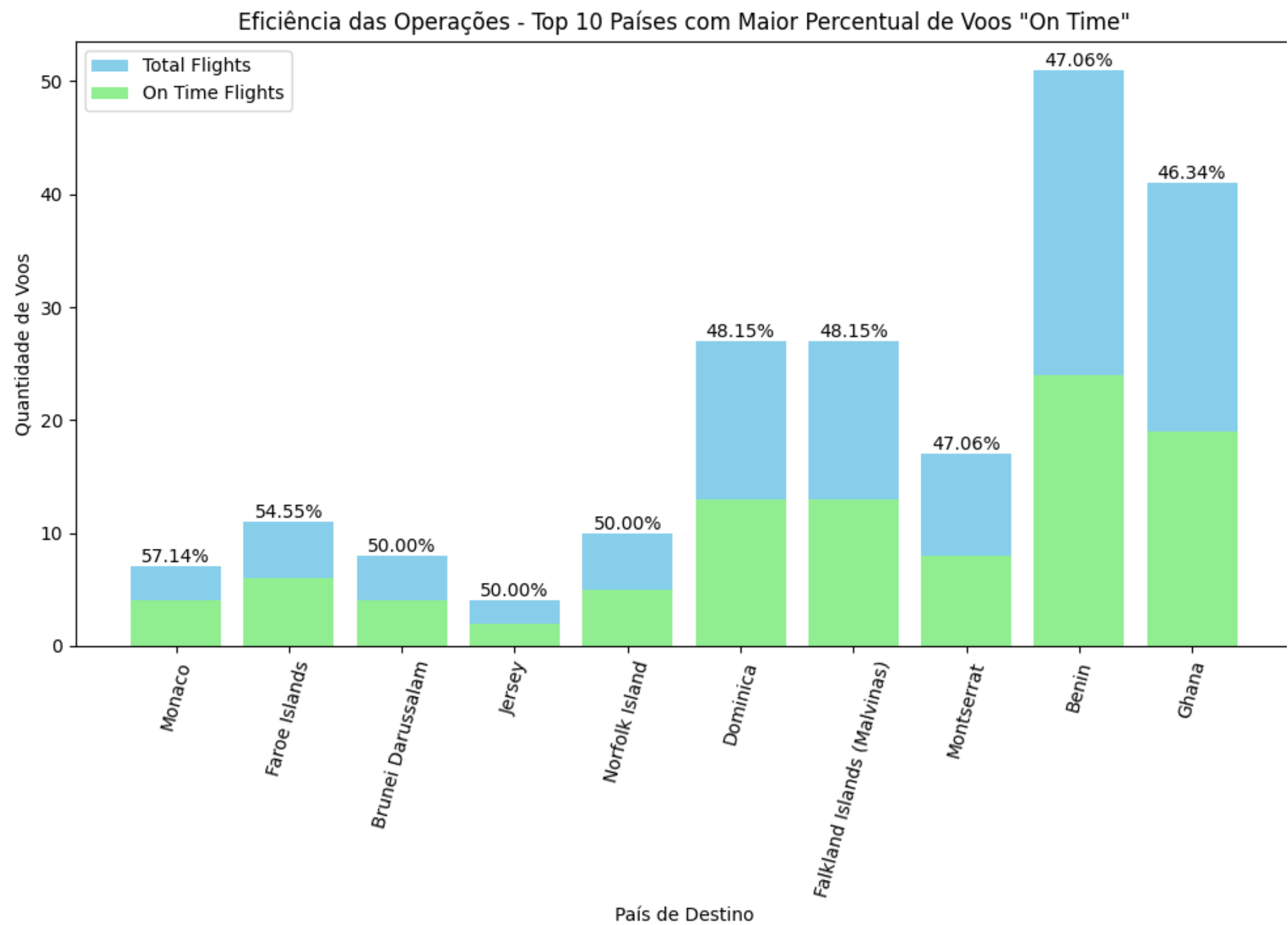
# Insights



## Insights

No aspecto eficiência de voos, esses são os países com maiores quantidades de voos “On Time”, ou seja, voos que não foram cancelados e nem houveram atrasos.

# Insights

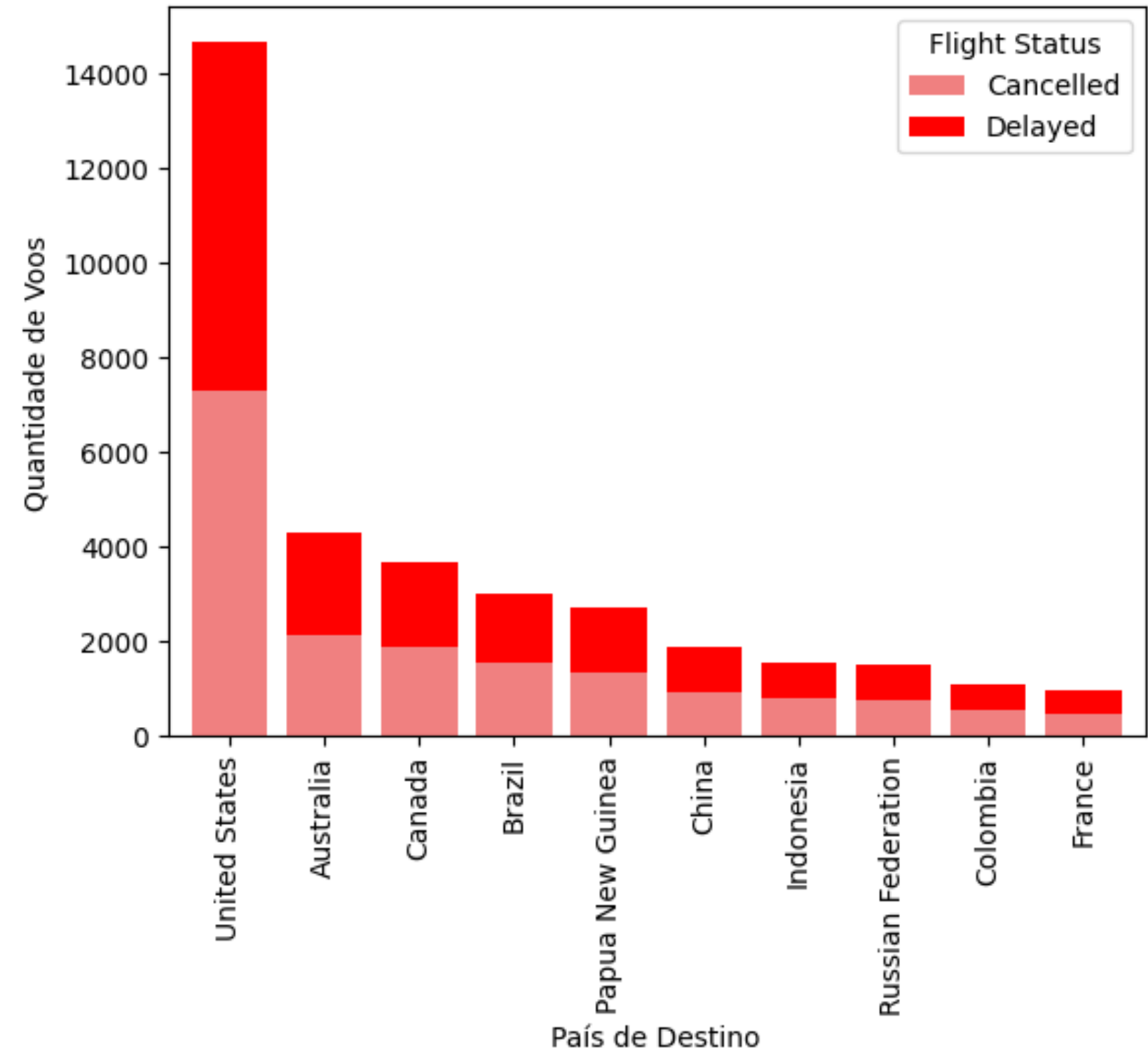


## Insights

Decidimos analisar a eficiência dos voos proporcionalmente, visto que EUA tinha mais voos e isso não significava que por terem mais dados "on time" eram o país com mais voos eficiente. Com isso, descobrimos que o país com maior eficiência é Mônaco

# Insights

Piores Desempenhos por País de Destino (Cancelled e Delayed) - valor absoluto



## Insights

Apesar dos EUA apresentar a maior quantidade de voos "On time", em proporção é o país com maior proporção no quesito ineficiência, ou seja, status de voos "Cancelled ou "Delayed" (cancelados ou com atrasos



# Modelo ML para Análise de Demanda por Destino

Modelo: Regressão Linear

Medidas de acurácia: Erro Quadrático Médio (MSE), Coeficiente de Determinação ( $R^2$ )





## ❖ Codificação e Contagem do Modelo:

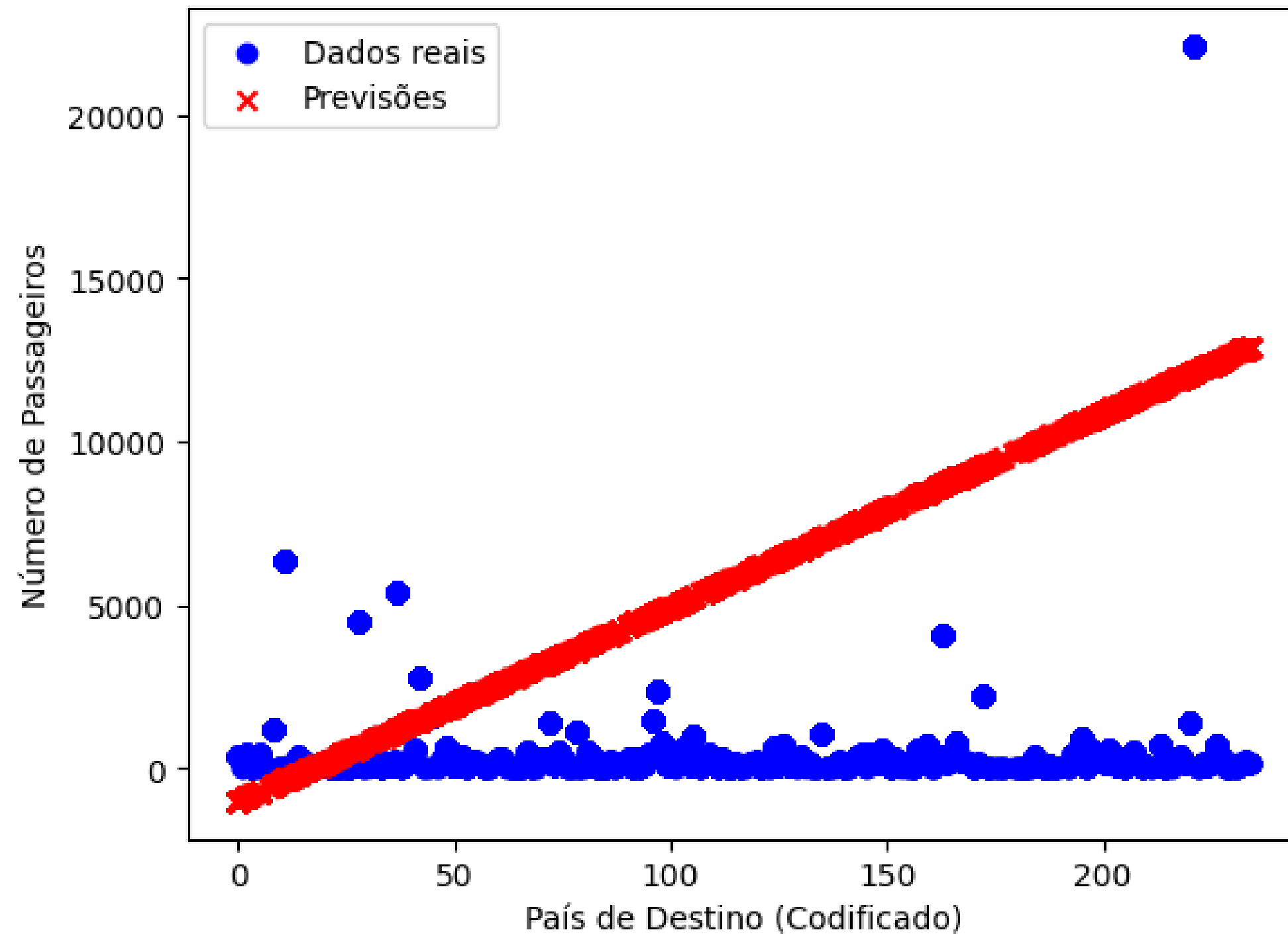
Inicialmente, para aplicação do modelo a codificação da variável categórica "Nome do País" foi realizada usando Label Encoding. Em seguida, o número de registros para cada país de destino foi calculado, resultando em um DataFrame com as colunas "Nome do País" e "Contagem de Passageiros".

## ❖ Preparação e Treinamento do Modelo:

Os dados foram preparados para o treinamento do modelo de regressão linear. A variável independente ( $X$ ) é definida como "Nome do País", enquanto a variável dependente ( $y$ ) é "Contagem de Passageiros". A eficácia do modelo foi avaliada com a divisão dos dados em conjuntos de treinamento (80%) e teste (20%). O modelo foi treinado usando a biblioteca scikit-learn, permitindo previsões com base nos dados de teste. Visualmente, um gráfico de dispersão foi gerado para comparar dados reais (azul) e previsões do modelo (vermelho, marcadas com 'x').



# Resultado da Regressão Linear



## Resultado de acurácia:

Erro Quadrático Médio (MSE) -

Coeficiente de Determinação ( $R^2$ ) -

# Obrigado

---

