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.

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

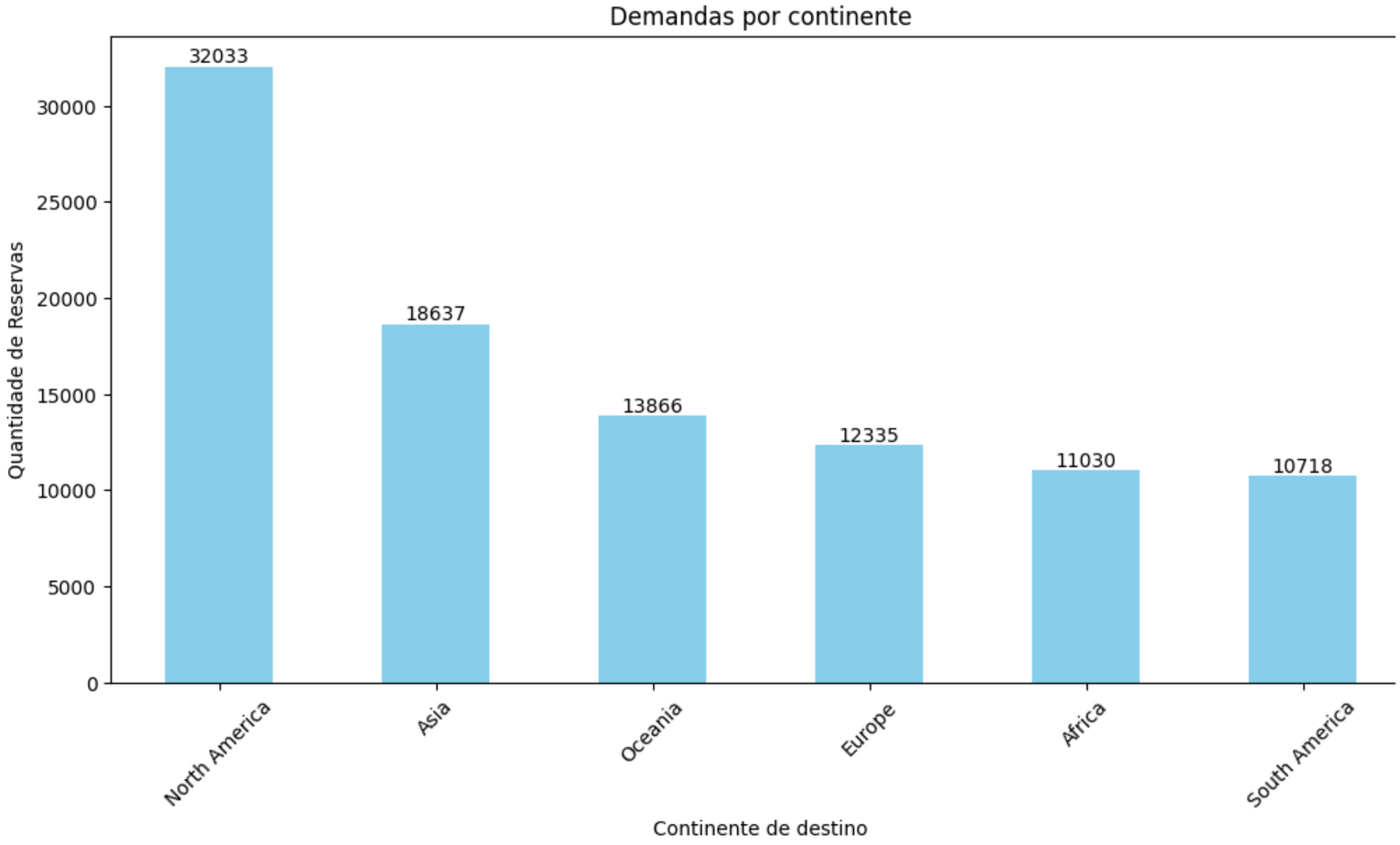
Análise Exploratória dos Dados

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

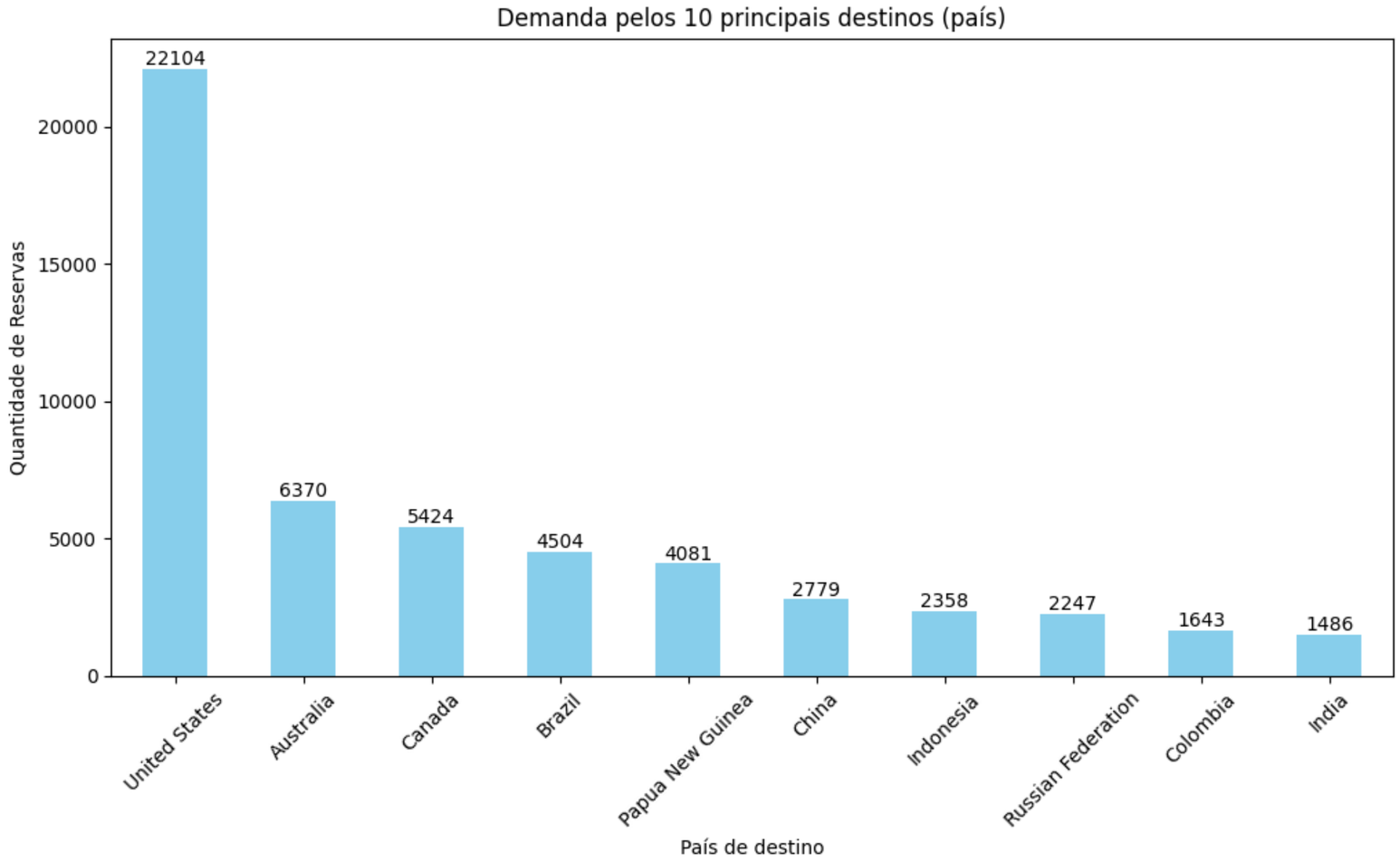
Insights



Insights

Adentrando ao conjunto de dados, descobrimos que a agência "Boa Viagem" atua como uma bússola, direcionando viajantes por oceanos de informações. Em nossa recente análise, encontramos um ponto de convergência fascinante: a América do Norte emerge como o epicentro pulsante da demanda por viagens.

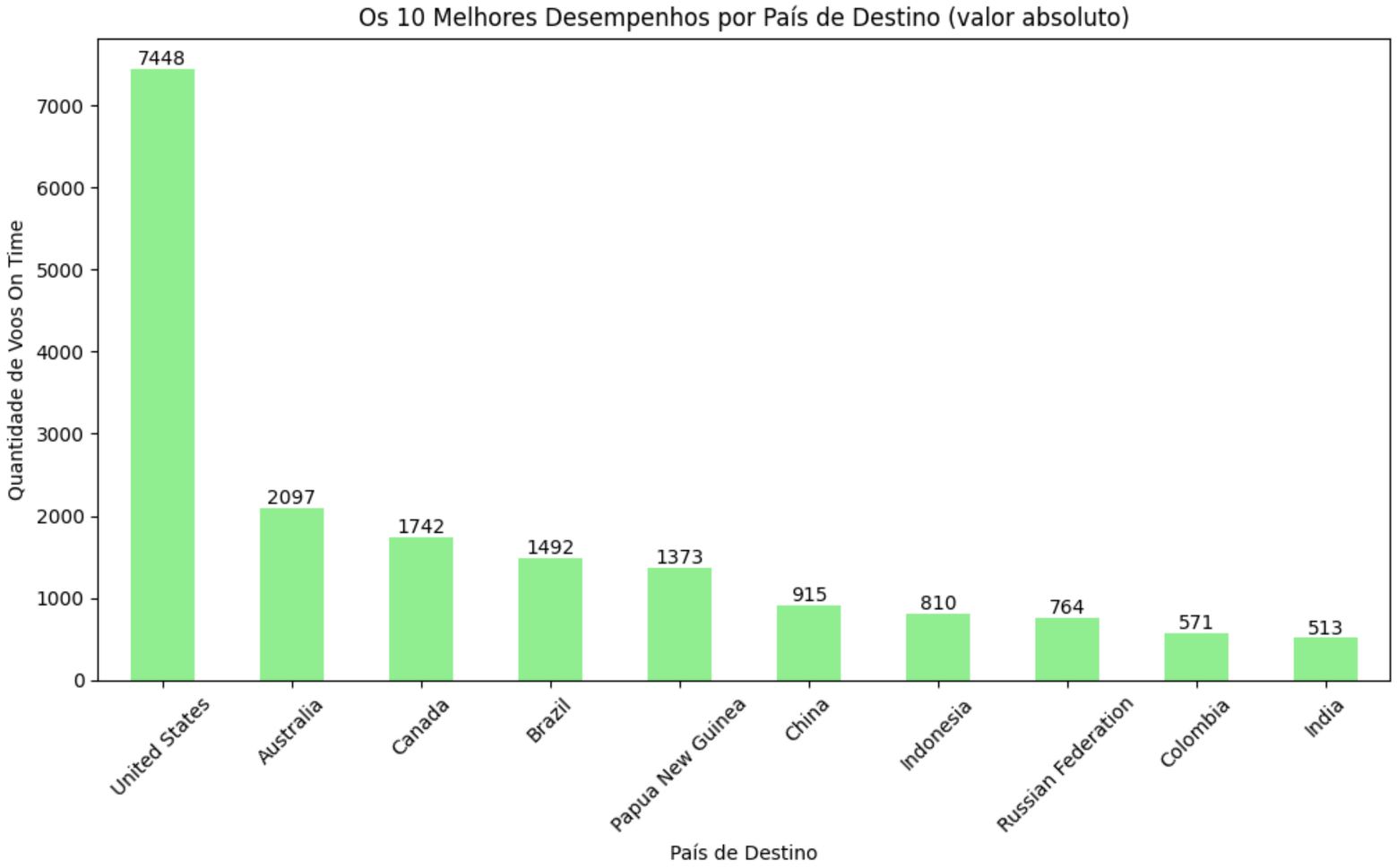
Insights



Insights

O país que se destaca como o principal destino dos voos operados pela "Boa Viagem" é os Estados Unidos. O top 3 de destinos é notavelmente composto por dois países situados no continente Norte Americano, que, como previamente destacado, lidera em termos de demanda continental. Adicionalmente, essa análise revela que os Estados Unidos apresentam uma demanda avassaladora em comparação com outros destinos, solidificando sua posição proeminente no panorama das preferências de viagem.

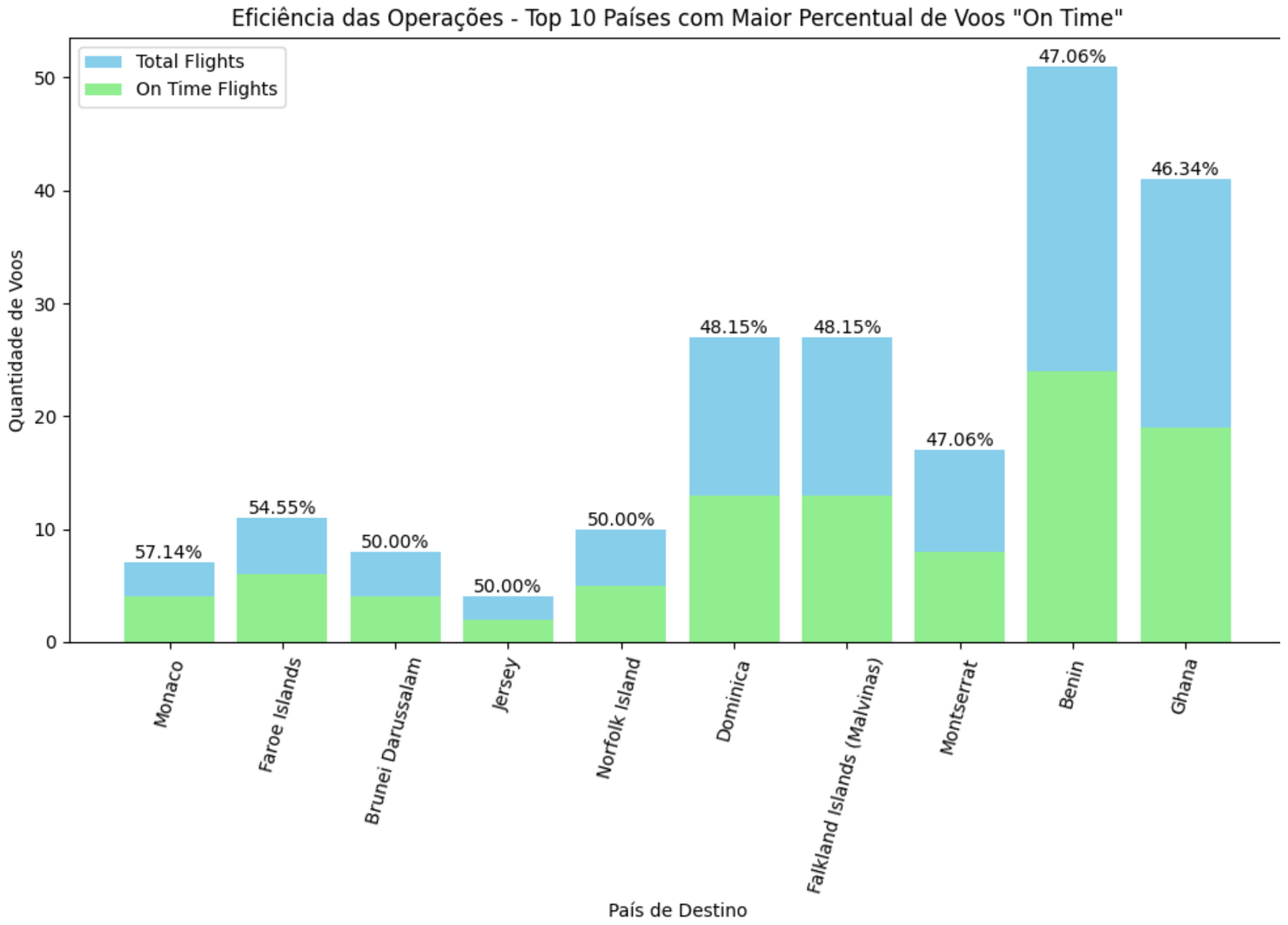
Insights



Insights

No quesito eficiência de voos, apresentamos os países com as maiores quantidades de voos "On Time", isto é, voos que não foram cancelados e transcorreram sem atrasos. Esses destinos destacam-se não apenas pela popularidade, mas também pela consistência e pontualidade nas operações aéreas, proporcionando aos passageiros uma experiência de viagem ainda mais confiável e suave.

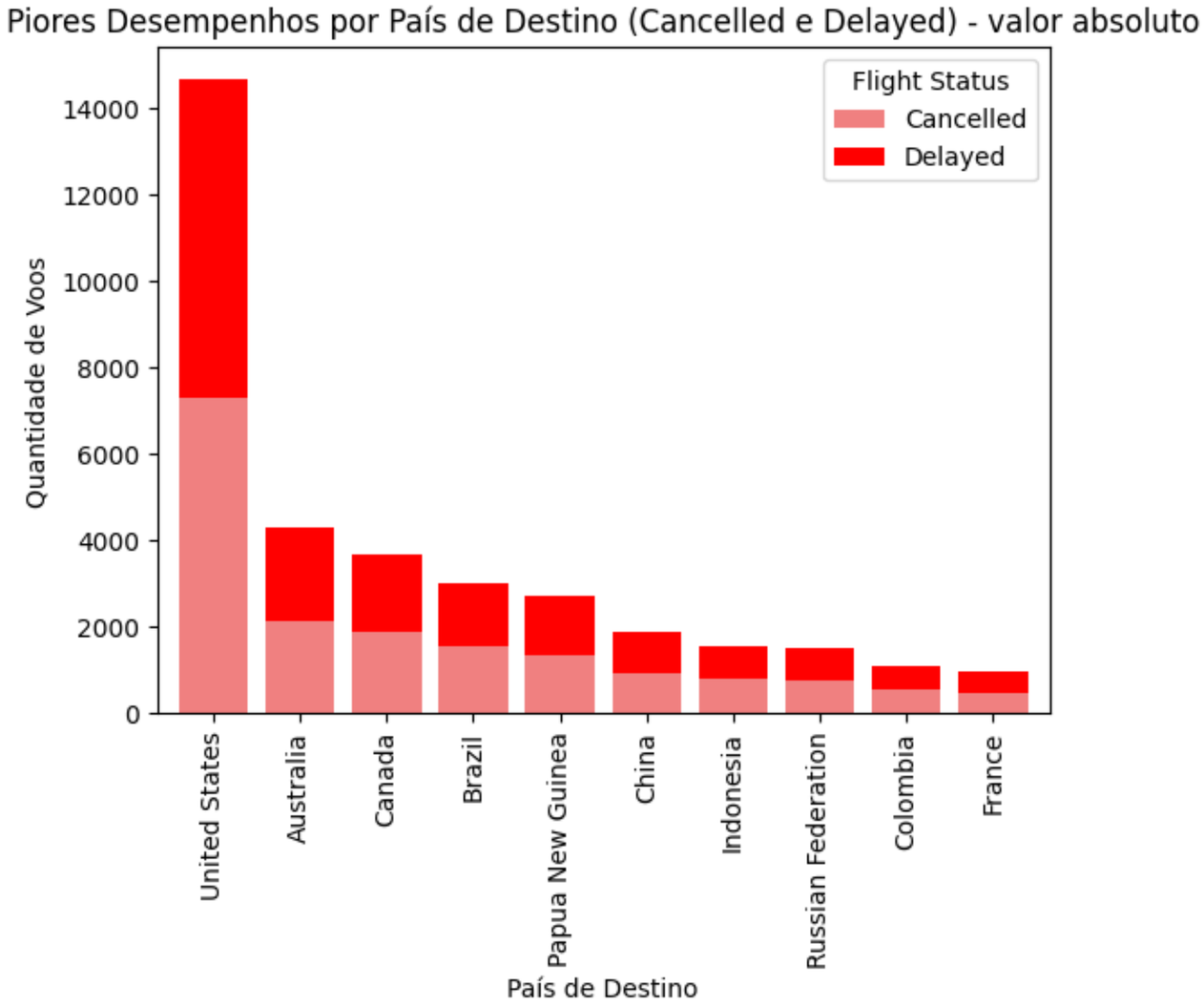
Insights



Insights

Optamos por uma análise proporcional da eficiência dos voos, considerando que os Estados Unidos, por terem uma quantidade maior de voos, não necessariamente indicavam uma eficiência superior com base nos dados "on time". Nessa investigação, revelou-se que Mônaco surge como o país com a maior eficiência em suas operações aéreas. Esta abordagem mais refinada nos permitiu identificar não apenas a quantidade de voos pontuais, mas também avaliar a eficiência relativa de cada país, destacando Mônaco como um modelo exemplar neste quesito.

Insights



Insights

Embora os Estados Unidos liderem em termos absolutos com a maior quantidade de voos "On Time", proporcionalmente, destacam-se como o país com a maior incidência de ineficiência, representada pelos status "Cancelled" (cancelados) ou "Delayed" (com atrasos). Esta análise mais aprofundada revela que, apesar da expressiva quantidade de voos pontuais, a proporção de voos afetados por cancelamentos ou atrasos destaca uma faceta adicional a ser considerada na avaliação da eficiência operacional nos voos para este destino.

Modelo ML para Análise de Demanda por Destino

Modelo: Regressão Linear e Regressão Polinomial

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

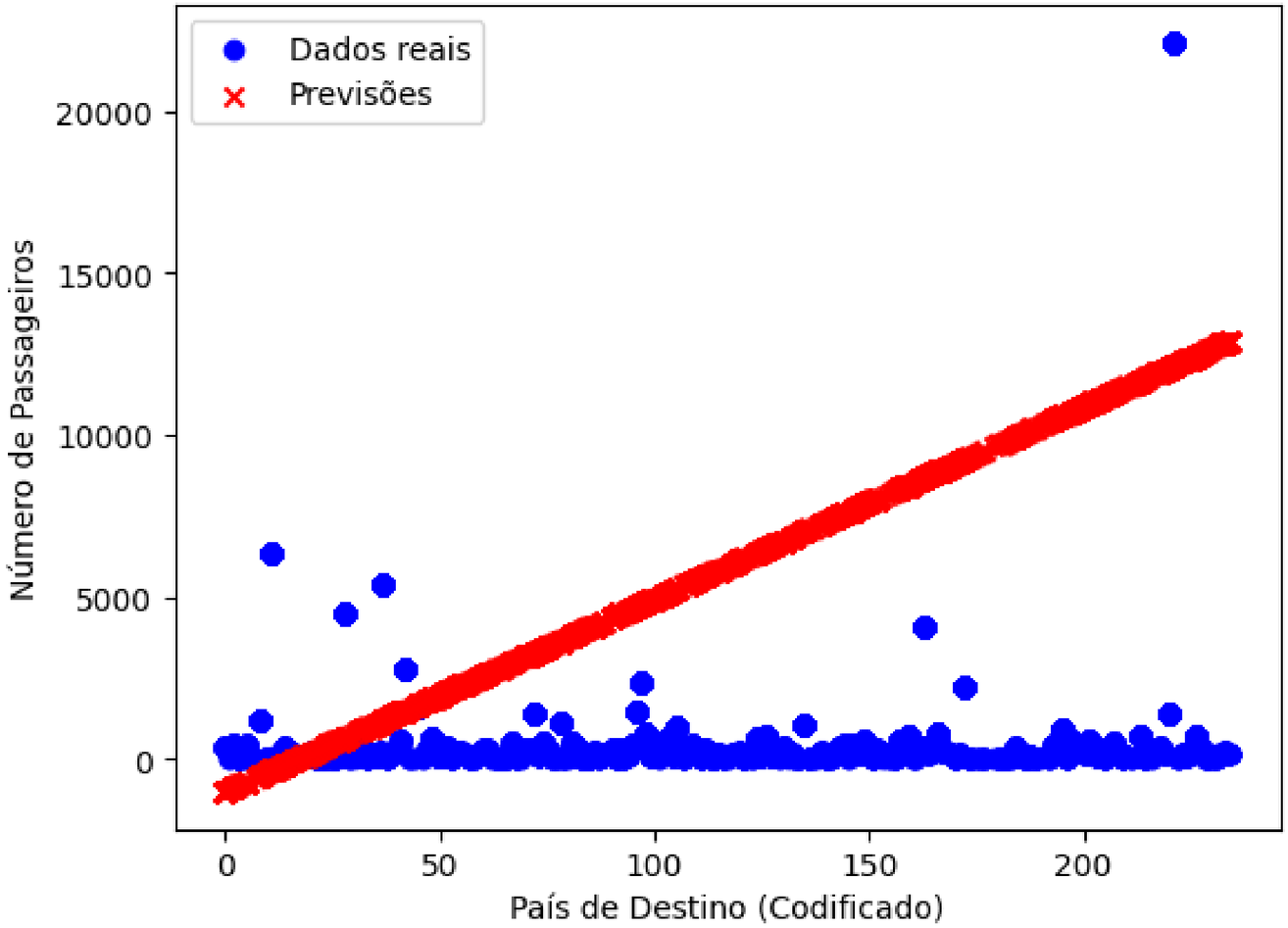
Codificação e Contagem do Modelo:

Inicialmente, para aplicação do medelo 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

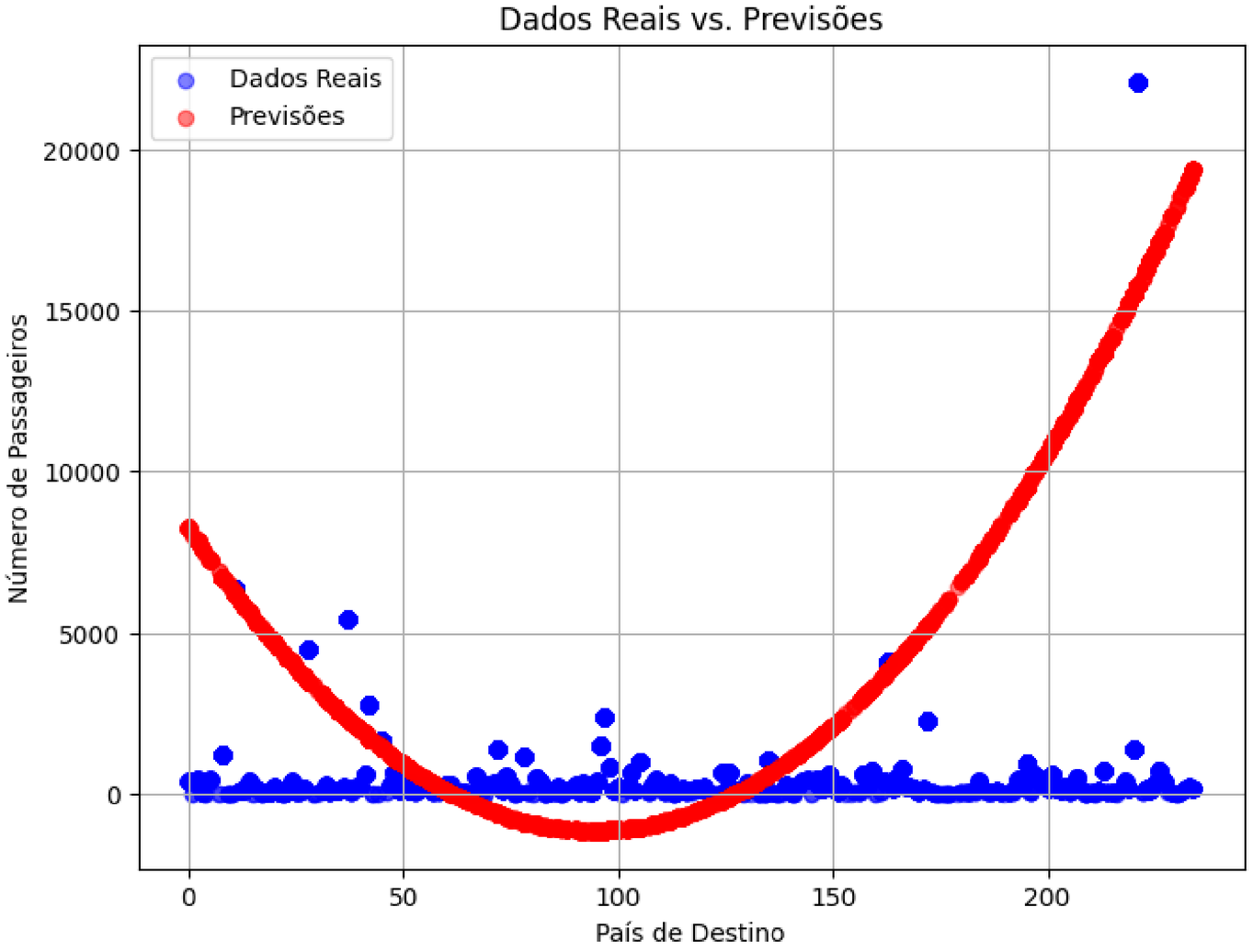


Resultado de acurácia:

Erro Quadrático Médio (MSE) = 51689893.02
Coeficiente de Determinação (R^2) = 0.30

Diante da constatação de que as acurácias não atenderam aos padrões de aceitabilidade desejados, tomamos a decisão estratégica de implementar um modelo de regressão polinomial. Esta abordagem visa aprimorar a precisão e o desempenho das análises, proporcionando uma maior flexibilidade na adaptação aos padrões complexos presentes nos dados. Dessa forma, almejamos elevar a qualidade das conclusões extraídas, reforçando nosso compromisso com a excelência na interpretação dos resultados.

Resultado da Regressão Polinomial



Resultado de acurácia:

Erro Quadrático Médio (MSE) = 32125604.94
Coeficiente de Determinação (R^2) = 0.56

A aplicação do modelo de regressão polinomial resultou em uma significativa melhoria nos resultados, evidenciada pelo Erro Quadrático Médio (MSE) de 32,125,604.94. Este valor reflete uma redução notável na discrepância entre as previsões do modelo e os dados reais, indicando uma melhor capacidade de ajuste aos padrões subjacentes nos dados. Além disso, o Coeficiente de Determinação (R^2) de 0.56 corrobora a eficácia do modelo na explicação da variabilidade dos dados. Esse índice, que varia de 0 a 1, sugere que aproximadamente 56% da variação nos resultados pode ser explicada pelas variáveis consideradas no modelo de regressão polinomial. Essa métrica robusta fortalece nossa confiança nos insights gerados e reforça a utilidade desse modelo aprimorado na análise dos dados em questão.

Obrigado