



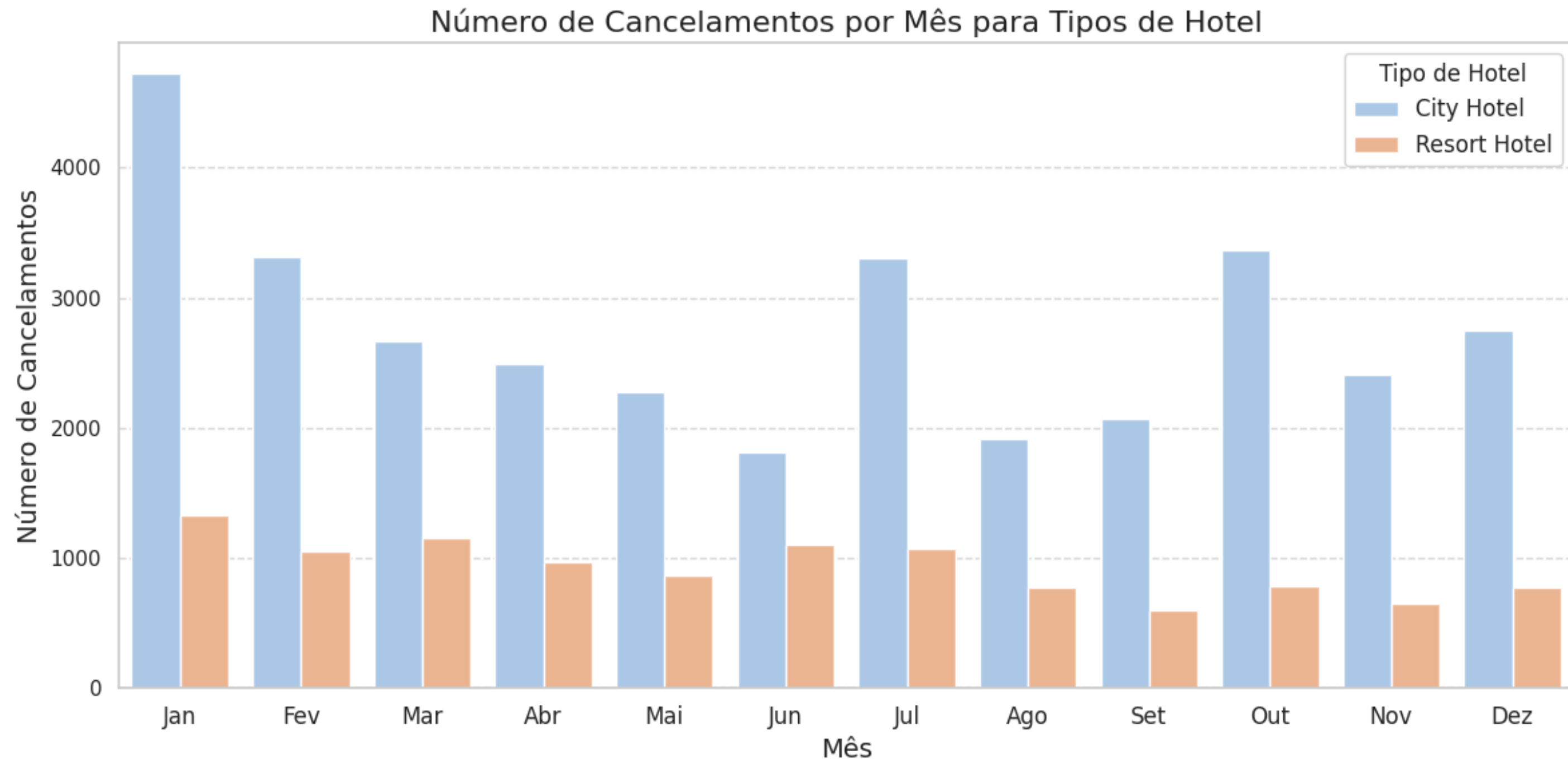
ANÁLISE DE DADOS PARA HOTÉIS: ESTRATÉGIAS DE PREVISÃO PARA REDUÇÃO DE CANCELAMENTOS E PREJUÍZOS

Luís Felipe Barros Pacheco
Rodrigo Moreira Marinho,
Erick Keven da Silva Alves
Másio César

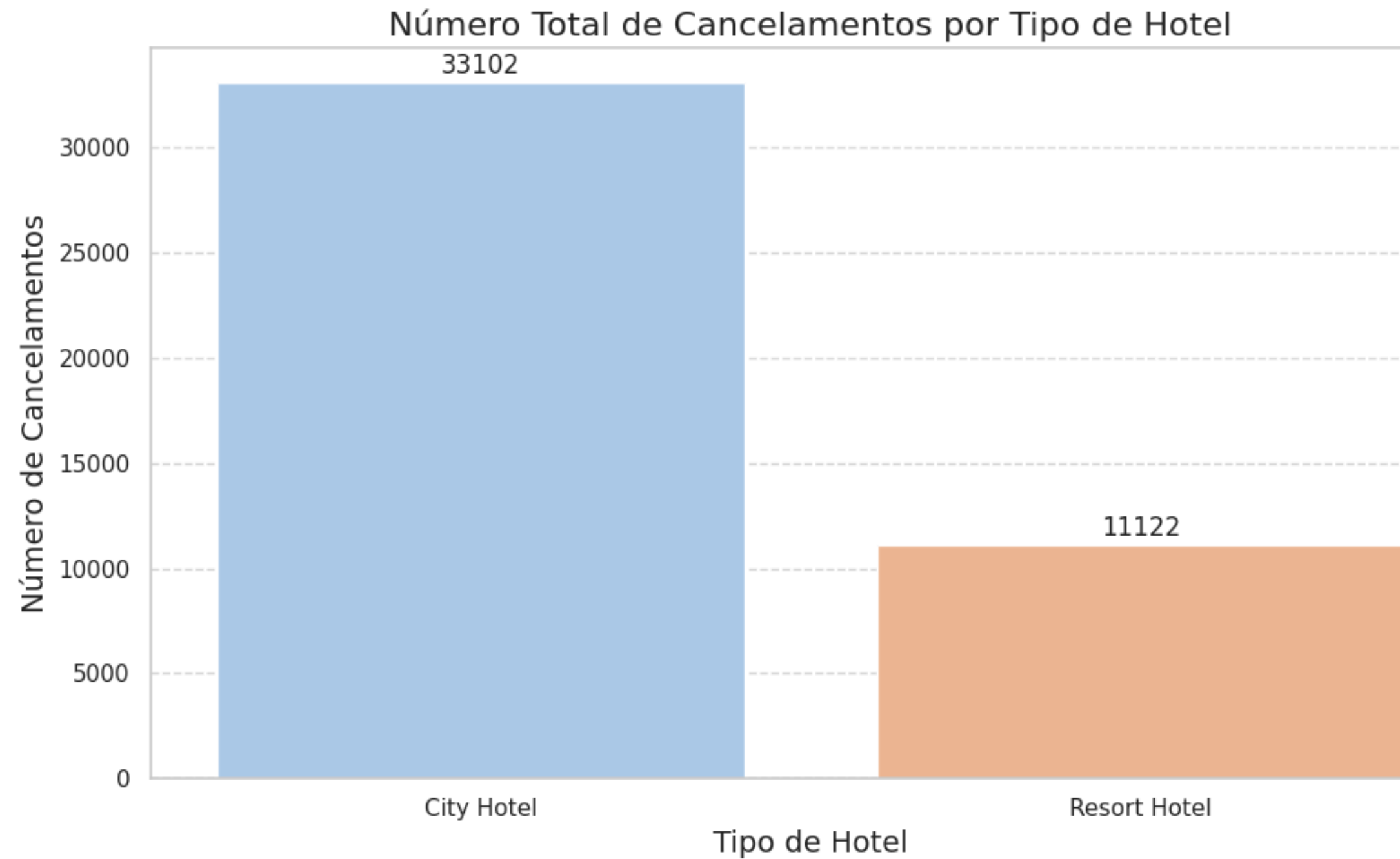
PROBLEMA DE NEGÓCIO

A instabilidade nas reservas causa um problema contínuo: **sem previsões precisas**, a gestão do hotel enfrenta dificuldades para **otimizar** a ocupação, resultando em uma perda significativa de **receita** e em falhas nas estratégias de atendimento ao cliente.

Visualização de dados



Visualização de dados



Observações

City Hotel:

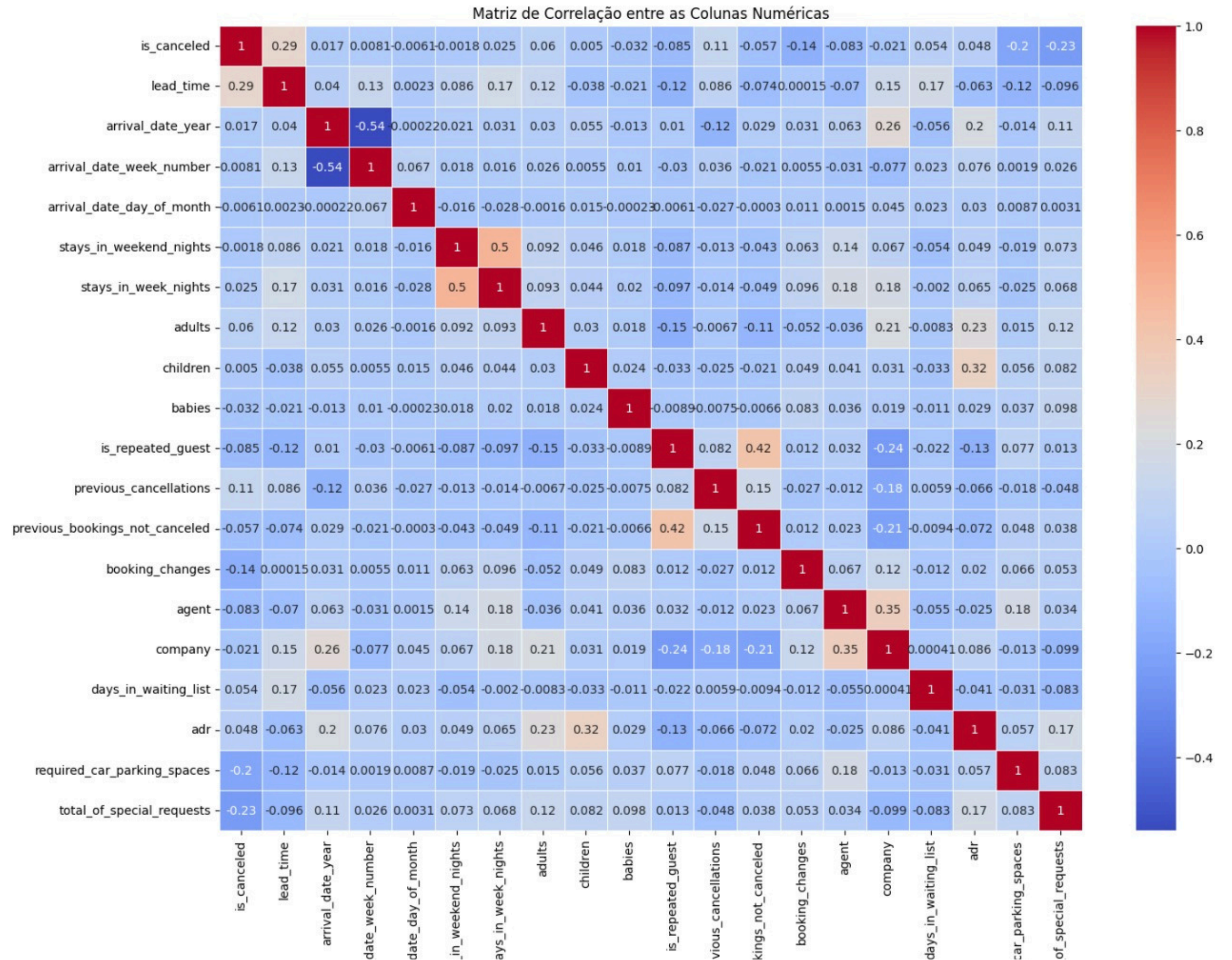
Apresenta um número consistentemente maior de cancelamentos em comparação com o Resort Hotel. Janeiro é o mês com o maior número de cancelamentos, seguido por outubro e julho.

Resort Hotel:

Tem um padrão mais estável de cancelamentos ao longo do ano, com picos menores em meses específicos. Janeiro e julho são os meses com os maiores números de cancelamentos.

Essas informações podem ser utilizadas para ajustar estratégias de marketing e políticas de reserva, visando reduzir a taxa de cancelamento em períodos críticos e melhorar a ocupação dos hotéis.

Matriz de Correlação



Estratégias para city Hotel

1. Status da Reserva

Estratégia: Melhorar a comunicação com os clientes antes do check-out para aumentar a satisfação e incentivar feedbacks positivos.

2. Tempo de Antecedência da Reserva (Lead Time)

Estratégia: Oferecer descontos para reservas feitas com bastante antecedência incentiva ocupação antecipada e melhora o planejamento.

3. Clientes de Portugal

Estratégia: Criar pacotes promocionais específicos para clientes portugueses, adaptando ofertas que atraiam esse público-alvo.

4. Depósito Não Reembolsável

Estratégia: Adotar uma política de cancelamento rígida para reservas com depósito não reembolsável, garantindo segurança financeira.

5. Estacionamento

Estratégia: Oferecer benefícios extras para clientes que reservam vagas de estacionamento, tornando o serviço mais atraente e valorizado.

Estratégias para Resort Hotel

1. Lead Time (Antecedência da Reserva)

Estratégia: Para reservas com muita antecedência, oferecer promoções especiais incentiva a ocupação antecipada e melhora a previsibilidade de receita.

2. Cancelamentos Anteriores

Estratégia: Adotar depósitos não reembolsáveis reduz o impacto de cancelamentos de última hora, garantindo maior segurança financeira.

3. ADR (Tarifa Média Diária)

Estratégia: Criar pacotes promocionais atrativos em períodos de baixa demanda ajuda a elevar a taxa diária média e a ocupação.

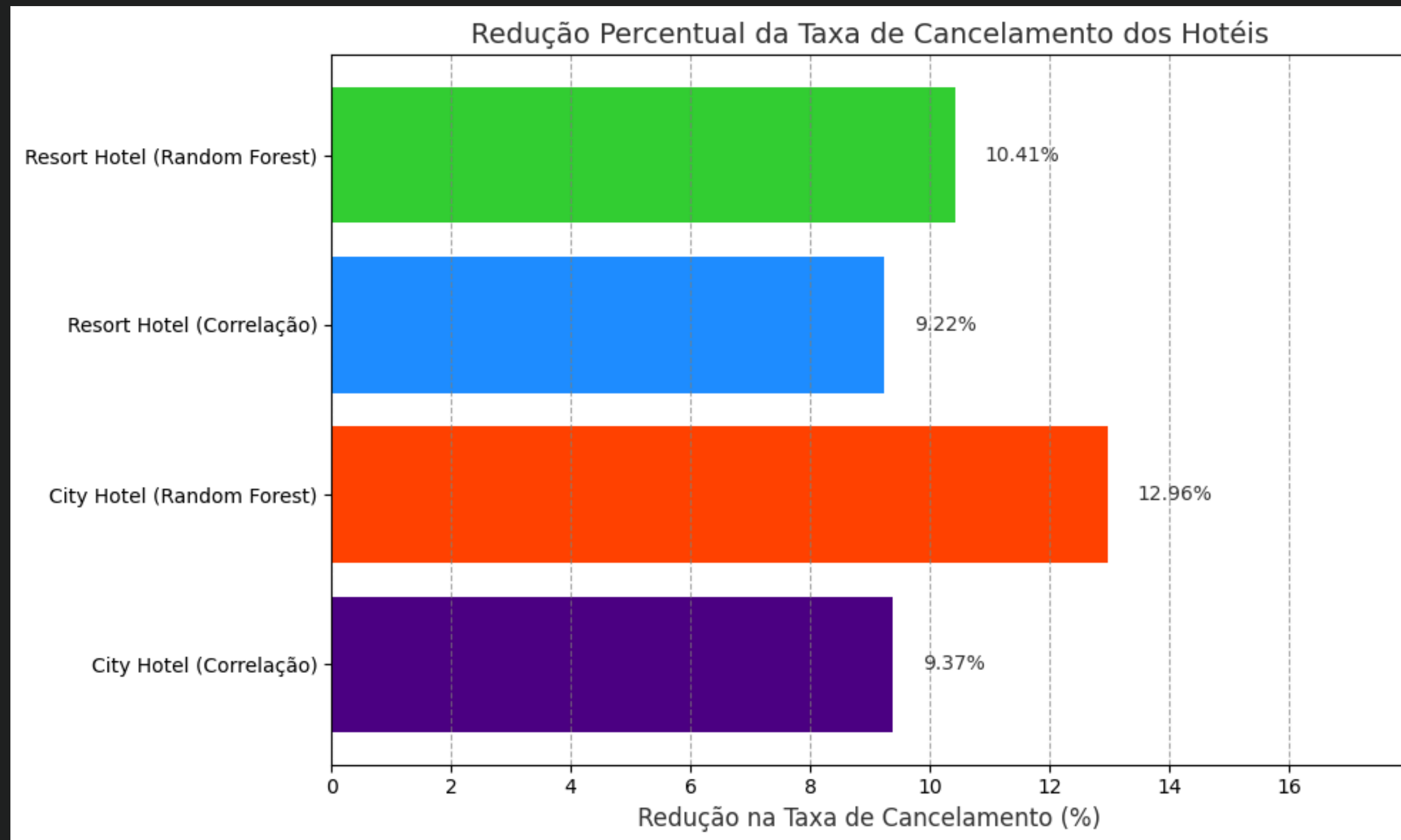
4. Crianças

Estratégia: Oferecer pacotes familiares com benefícios para crianças torna o local mais atrativo para famílias, incentivando reservas.

5. Número de Adultos

Estratégia: Pacotes para grupos maiores de adultos atraem famílias grandes e amigos em viagem, otimizando a ocupação e aumentando a receita.

Resultado das Estratégias Aplicadas



Métricas Usadas para Avaliar os Métodos

F1 Score

O F1 Score combina precisão e revocação, sendo especialmente útil para dados desbalanceados. Ele é calculado como a média harmônica entre precisão e revocação. Um F1 Score próximo de 1 indica eficácia em prever a classe positiva, enquanto valores próximos de 0 indicam baixa eficácia.

AUC (Area Under the ROC Curve)

A AUC mede a capacidade do modelo de discriminar entre classes. A curva ROC plota a taxa de verdadeiros positivos contra a taxa de falsos positivos. A AUC varia de 0,5 (modelo sem discriminação) a 1,0 (modelo perfeito). Quanto maior a AUC, melhor o modelo discrimina entre classes.

Accuracy

A precisão (accuracy) é a proporção de previsões corretas em relação ao total. Embora comum, pode ser enganosa em dados desbalanceados. Um valor próximo de 1 indica que a maioria das previsões está correta, enquanto valores baixos indicam falhas na previsão correta das instâncias

Métodos de Aprendizagem de Máquina

Árvore de Decisão

Utiliza uma estrutura em forma de árvore para modelar decisões, dividindo os dados em subconjuntos baseados em perguntas sobre os atributos. Cada nó interno representa uma pergunta e cada folha representa um resultado.

Random Forest

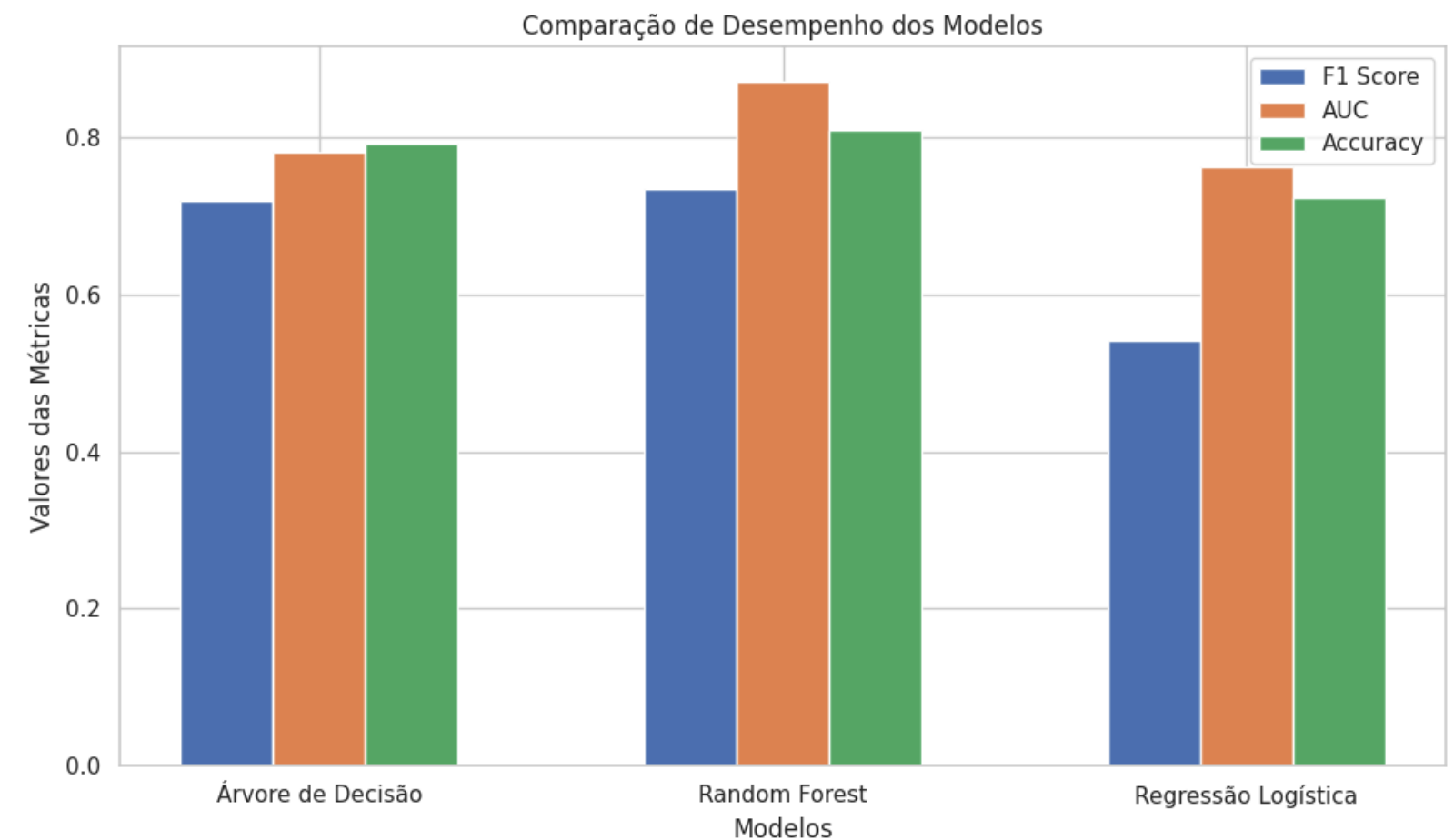
É um conjunto de múltiplas árvores de decisão, onde cada árvore é treinada em uma amostra aleatória do conjunto de dados. As previsões são feitas pela média das previsões de todas as árvores, aumentando a robustez e precisão do modelo.

Regressão Logística

Modelo estatístico que usa uma função logística para modelar a probabilidade de uma classe ou evento, como a classificação binária (por exemplo, se uma reserva será cancelada ou não).

Comparação dos modelos utilizados

Com base nos resultados, o modelo Random Forest é o melhor a ser escolhido. Ele apresentou os melhores desempenhos em todas as métricas avaliadas, sugerindo que é o mais eficaz para a tarefa de previsão de cancelamento de reservas no hotel. Além disso, devido à sua natureza de ensemble, o Random Forest tende a ser mais robusto e menos suscetível ao overfitting, especialmente em conjuntos de dados complexos.



Uso de Amostragem para Avaliar os Modelos

K-Fold Cross-Validation

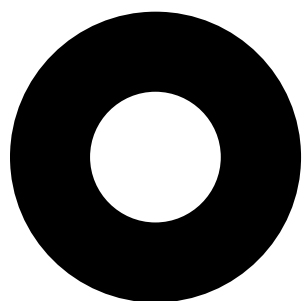
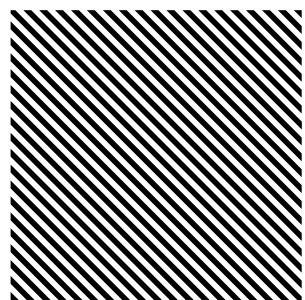
Divide os dados em k partes, treina o modelo em $k-1$ partes e testa na parte restante. Repete o processo k vezes para uma avaliação mais robusta.

LOOCV (Leave-One-Out Cross-Validation)

Para cada iteração, usa uma única observação como teste e o restante como treino. Fornece uma avaliação detalhada, mas é computacionalmente caro.

Bootstrap

Gera múltiplas amostras com reposição do conjunto de dados original, treina o modelo em cada uma e avalia a incerteza nas previsões.



Validação Estatística

O teste t para comparar os F1 Scores entre o Random Forest e a Árvore de Decisão resultou em:

T-statistic: 8.93

P-value: 4.95×10^{-8}

O p-value muito abaixo de 0.05 indica que a diferença é estatisticamente significativa, confirmando que o Random Forest tem desempenho superior e é a melhor escolha.

Conclusão

A análise mostra que o k-fold cross-validation foi eficaz na avaliação dos modelos, fornecendo uma estimativa robusta de seu desempenho. Com base nas métricas superiores e na validação positiva, o Random Forest foi escolhido como o modelo mais adequado para prever cancelamentos.

OBRIGADO!