



Reduzindo prejuízos financeiros em hotéis: uma abordagem de análise de dados para prever cancelamentos de reservas

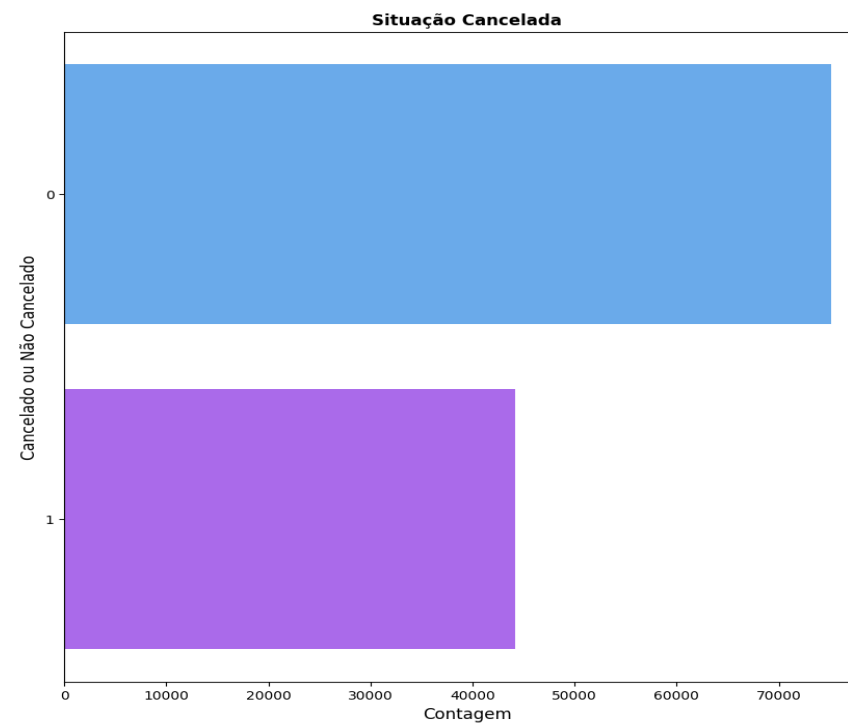
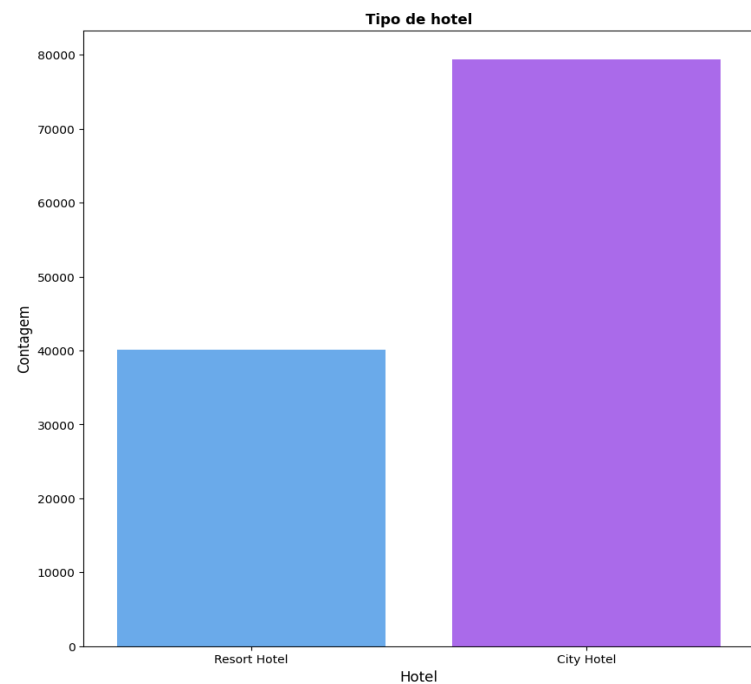
Gilson Francisco De Oliveira Castro

Problema de negócio & Perguntas

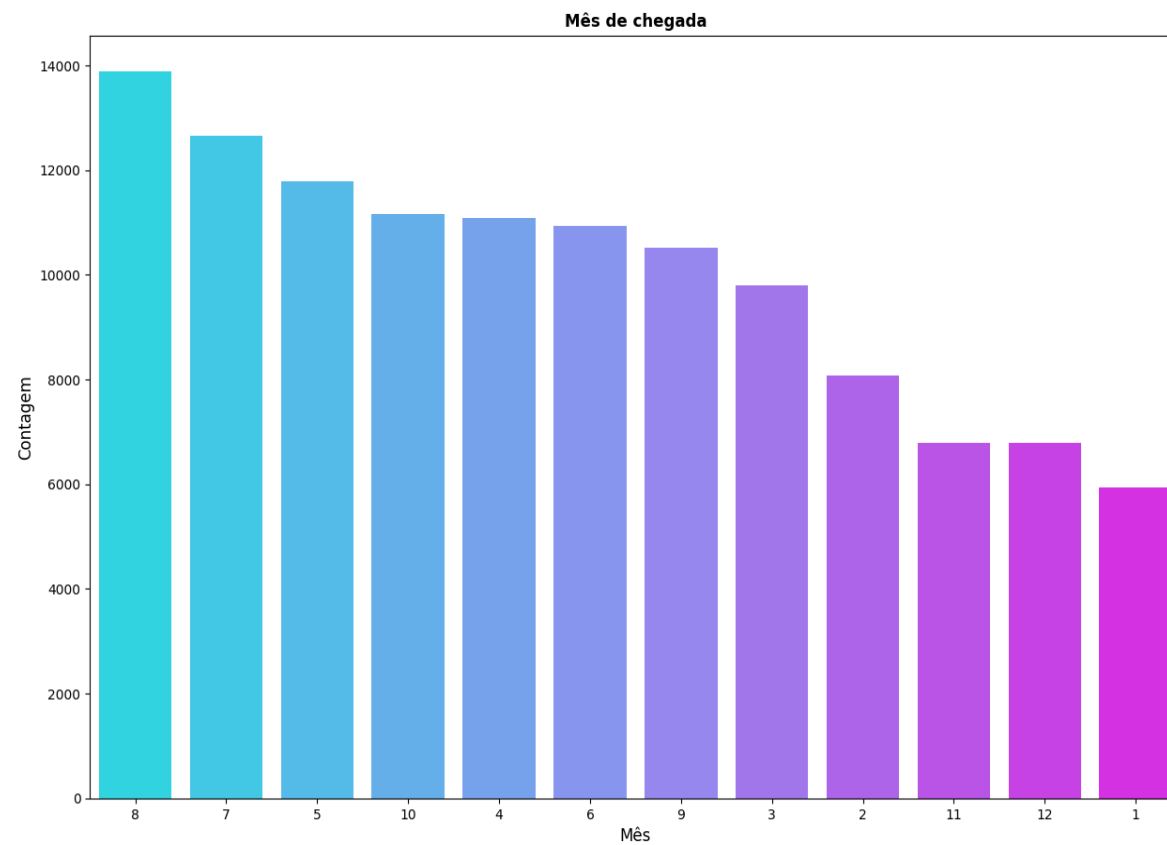
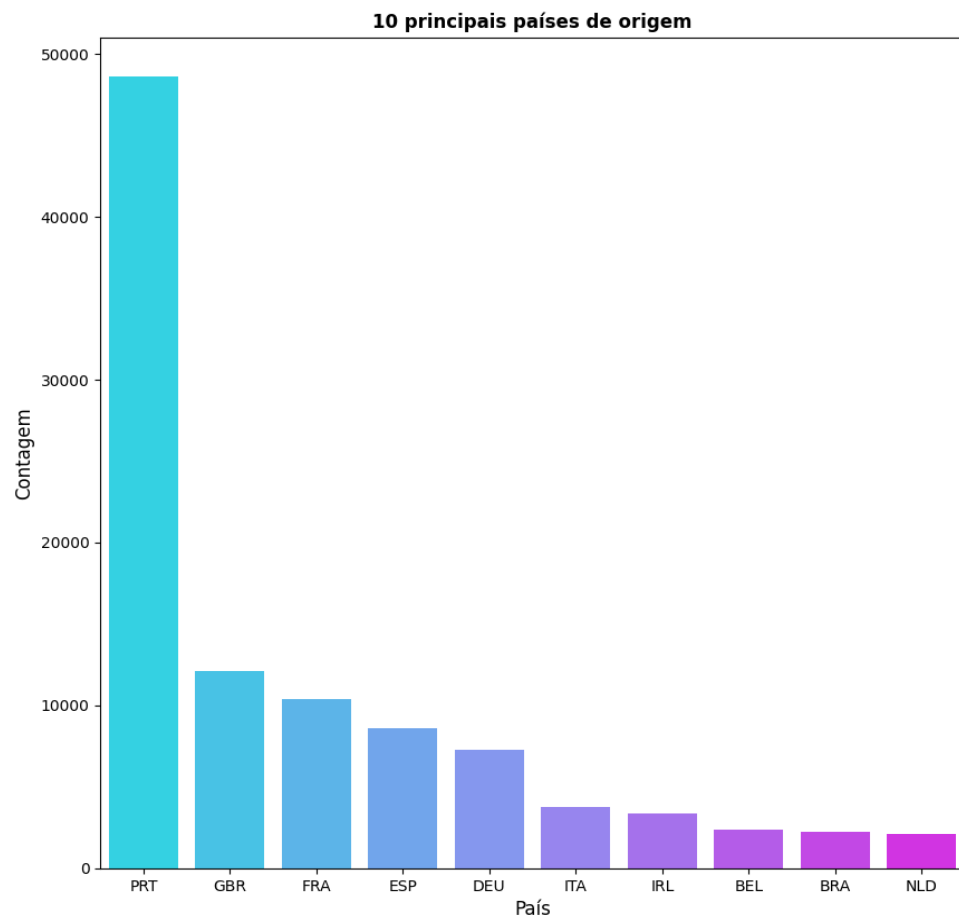
- A demanda hoteleira é gerada por uma variedade de viajantes, incluindo aqueles em viagens de negócios e turistas em busca de lazer.
- Existe a possibilidade de um cliente cancelar sua reserva?
- Se houver cancelamento, quais são os fatores determinantes que levam os clientes a cancelarem suas reservas?



Visualização dos dados



Visualização dos dados



Importância do problema

- A previsão da demanda hoteleira é um aspecto crucial da gestão eficaz de um hotel.
- Tornou-se cada vez mais importante estimar as taxas de cancelamento.

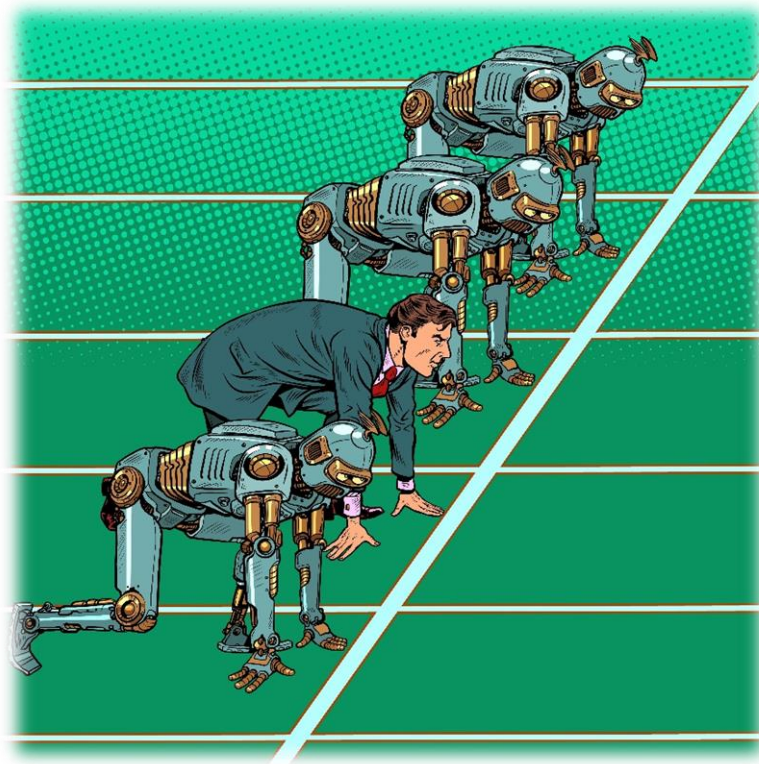


Solução

- Previsão de cancelamentos de reservas
- Utilização de algoritmos de machine learning
- O hotel pode adotar medidas preventivas, como oferecer promoções ou opções de reagendamento



Qual o melhor modelo para prever cancelamento ?



Modelos de classificação usados:

- **Regressão Logística** limita o valor entre 0 e 1, é simples e interpretável.
- **Árvore de decisão** é explicável, mas pode ter previsões grosseiras e ser suscetível a sobreajuste.
- **Floresta aleatória** fornece resultados mais precisos ao utilizar um conjunto de muitos modelos individuais com seleção aleatória de recursos.
- **CatBoost** é ideal para informações categóricas e foi criado para melhorar os resultados nesse tipo de dado.

Escolha da métrica de avaliação

- Considerando a possibilidade de prejuízos para os proprietários de hotéis em decorrência de excesso de reservas
- Foi priorizado um alto valor de Recall na avaliação do modelo

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)}$$

Comparação dos modelos de ML selecionados

Modelo	Precision	Recall	F1-Score
Logistic Regression	0.61	0.86	0.72
Decision Tree	0.63	0.93	0.75
Random Forest	0.66	0.93	0.77
CatBoost	0.68	0.90	0.78

Conclusão

- O modelo **Random Forest** é o mais adequado para a classificação de cancelamentos de reservas de hotéis.
- Reservas feitas através de agentes de viagens online e um tempo de antecedência maior são os fatores mais importantes na predição de cancelamentos
- Integração com sistemas de gerenciamento de reservas, permitindo que os hotéis monitorem automaticamente a probabilidade de cancelamento