

Identificação de cancelamentos de reservas de uma rede hoteleira a fim de reduzir sua ocorrência

Conclusões e planos de ação para aumentar a receita

Lucas Silva Marcondes

smarcondes.lucas@gmail.com

Introdução

Utilizando os dados de reservas de quartos referentes a duas unidades de uma empresa hoteleira de Portugal, foi feita a presente análise, tendo como principal objetivo ajudar a empresa a reduzir o número de cancelamentos de reservas, aumentando assim a taxa de ocupação dos quartos.

Para esse objetivo, foram criados modelos a fim de prever quando uma reserva poderia acarretar em um cancelamento, fornecendo ao hotel uma base para poder tomar medidas que diminuam essas ocorrências.

A base original utilizada para análise e criação dos modelos pode ser acessada aqui: [link externo](#)

Sobre o trabalho

Neste presente documento, temos um resumo de todo o trabalho realizado, concentrado nos principais pontos observados. O trabalho completo consiste em uma análise mais profunda e sequencial, uma jornada pela exploração dos dados, testes e experimentações durante a construção dos modelos preditivos.

Foram utilizados modelos supervisionados para classificação dos dados, além de análise de componentes principais para ajudar no processo. Os cadernos detalhados podem ser acessados pelos links abaixo:

- Início da Análise e Árvores de Decisão: [link externo](#)
- PCA: [link externo](#)
- SVM: [link externo](#)
- Random Forest: [link externo](#)
- Gradient Boosting: [link externo](#)

Conclusões resumidas

A seguir, descrevemos resumidamente as principais conclusões e planos de ações do trabalho:

- A taxa de **cancelamento** do hotel, segundo os dados analisados, é de cerca de **37%**.
- Elaboramos uma lista com os **países** de origem dos hóspedes que mais **cancelaram** reservas, com taxas acima de **50%**. ([link](#))
- Construímos um modelo que prevê os **cancelamentos** de reservas com **81,69%** de precisão. ([link](#)).
- Construímos um modelo que prevê os **não cancelamentos** de reserva (ou quando os hóspedes irão **comparecer**) com **89,54%** de precisão. No entanto, esse não foi o foco do trabalho, e se desejável, essa acurácia pode ser melhorada para mais de **90%**. ([link](#))

- As previsões de **comparecimento** (ou não cancelamentos) podem ser úteis para garantir **previsibilidade de receita e fluxo de caixa e facilitar** a gestão do hotel. (mais detalhes [aqui](#))
- As previsões de **cancelamentos** podem fornecer ao hotel a possibilidade de realização de campanhas mais **assertivas e efetivas** a hóspedes com grandes probabilidades de não comparecimento. (mais detalhes [aqui](#))
- Essas campanhas, se bem sucedidas, podem render ao hotel um **aumento** na receita de até **23%**. (mais detalhes [aqui](#))

Os detalhes podem ser acessados através dos links acima, ou através das seções a seguir.

Detalhes do trabalho

Países com maiores taxas de cancelamento

Temos abaixo a lista com os países de origem dos hóspedes que tiveram 100% de cancelamento:

Nome País	Sigla País	Percentual de cancelamento	Número cancelamentos
Fiji	FJI	100%	1
Guernsey	GGY	100%	3
Guadalupe	GLP	100%	2
Honduras	HND	100%	1
Ilha de Man	IMN	100%	2
Jersey	JEY	100%	8
Camboja	KHM	100%	2
Mayotte	MYT	100%	2
Nicarágua	NIC	100%	1
Ilhas Menores Distantes dos Estados Unidos	UMI	100%	1
Ilhas Virgens (Reino Unido)	VGB	100%	1

E abaixo a lista com os países de origem dos hóspedes que tiveram mais de 50% de cancelamento:

Nome País	Sigla País	Percentual de cancelamento	Número cancelamentos
Macau	MAC	93,75%	16
Hong Kong	HKG	89,66%	29
Tajiquistão	TJK	88,89%	9
Emirados Árabes Unidos	ARE	84,31%	51
Bahrein	BHR	80,00%	5
Ilhas Faroe	FRO	80,00%	5
Bangladesh	BGD	75,00%	12
Maldivas	MDV	75,00%	12
Qatar	QAT	73,33%	15
Senegal	SEN	72,73%	11
Andorra	AND	71,43%	7
Arábia Saudita	SAU	68,75%	48
Indonésia	IDN	68,57%	35
Geórgia	GEO	68,18%	22
Paquistão	PAK	64,29%	14
Filipinas	PHL	62,50%	40
Nigéria	NGA	61,76%	34
Gibraltar	GIB	61,11%	18
Tanzânia	TZA	60,00%	5
República Dominicana	DOM	57,14%	14
Portugal	PRT	56,64%	48.590
Angola	AGO	56,63%	362
Azerbaijão	AZE	52,94%	17

O intuito desta lista é fornecer ao hotel informações sobre os países que mais cancelam reservas, a fim de que possa ser tomada alguma providência, desde que de maneira responsável, como campanhas especiais para esses países (promoções, por exemplo), ou até mesmo investigações mais profundas sobre os motivos dos cancelamentos.

É importante notar que a taxa média de cancelamentos dos hotéis é de 37%, portanto 50% é uma taxa bastante relevante e passível de ser investigada mais a fundo.

Além disso, essa lista foi disponibilizada porque não utilizamos os dados de país de origem nos nossos modelos, para evitar possíveis discriminações referentes à nacionalidade dos hóspedes. Nossa intenção é construir um modelo que foca simplesmente nas características das reservas e não das pessoas que reservaram.

Dado potencialmente inconsistente

Durante a análise, foi encontrado um campo potencialmente inconsistente: `deposit_type`.

Esse campo, teoricamente, se refere ao tipo de depósito no ato da reserva, e indica três cenários: 1 - Depósito não realizado, 2 - Depósito não reembolsável e 3 - Depósito reembolsável.

Foi observado que mais de **99%** dos agendamentos **não reembolsáveis** foram **cancelados**, enquanto os outros apresentaram cancelamentos próximos de 20%. Por ser uma informação contra intuitiva e por não ter sido possível verificar a veracidade desse fato, a variável não foi utilizada para treinar os modelos e fica aqui o alerta para possível verificação deste campo.

Melhor modelo para previsão de cancelamentos

O melhor modelo encontrado para prever os cancelamentos das reservas dos hotéis, foi o **Gradient Boosting** (link para o caderno detalhado:)

O modelo construído conseguiu rotular corretamente **81,69%** dos **cancelamentos** de reservas e **83,75%** dos **não cancelamentos**.

Melhor modelo para previsão de não cancelamentos

O melhor modelo encontrado para prever o não cancelamento das reservas dos hotéis, foi o **Random Forest** (link para o caderno detalhado:)

O modelo construído conseguiu rotular corretamente **89,54%** dos **não cancelamentos** de reservas e **79,79%** dos **cancelamentos**. É importante notar que nosso foco no trabalho foi prever os cancelamentos e em consequência disso, as taxas de previsão de não cancelamento não foram otimizadas. No entanto, caso seja de interesse da empresa, com alguns ajustes, essa taxa pode ultrapassar facilmente **90%** de acurácia.

Apesar de não ter sido o foco, prever os não cancelamentos pode ser útil para estimar a possível taxa de ocupação mínima do hotel em um período e facilitar atividades administrativas.

Por exemplo: se o modelo prevê que **100** reservas não serão canceladas, com uma acurácia de **89%**, temos a certeza de que por volta de 89 famílias ou indivíduos que realizaram reservas estarão presentes, facilitando assim a organização das atividades necessárias nos estabelecimentos.

Sugestões de planos de ação

Temos duas situações a serem exploradas com os modelos desenvolvidos: cancelamentos e não cancelamentos.

Para os **não cancelamentos** (ou seja, o hóspede efetivamente comparece ao hotel), temos o potencial observado no tópico anterior, que é ajudar o hotel a organizar melhor sua infraestrutura durante determinados períodos.

Nosso modelo consegue prever com uma precisão de quase **90%** quando um hóspede irá comparecer, ou seja, conseguimos estabelecer um número **mínimo** de hóspedes por período.

Com esses dados, a **gestão** da operação fica mais **assertiva**, pois tem maior consciência do **número de funcionários** que precisa, além da **quantidade de produtos** que precisa adquirir para prover seus serviços com qualidade, garantindo que não faltará nada e **diminuindo** também o **desperdício**.

Além disso, o hotel consegue ter maior **previsibilidade** de **receita**, principalmente quando trabalha com uma política de reserva sem depósito. Tendo um valor aproximado da receita por período, o hotel consegue planejar e trabalhar melhor com seu fluxo de caixa, garantindo uma **melhor saúde financeira** ao longo do tempo.

Já para o principal objetivo de nosso trabalho, os **cancelamentos**, temos um grande potencial de aumento de receita, pois podemos utilizar estratégias para reduzir esses cancelamentos e garantir a presença de mais hóspedes. Algumas dessas estratégias podem incluir:

- Entrar em contato com os clientes de maneiras estratégicas, confirmando o agendamento;
- Enviar e-mails e mensagens periodicamente, lembrando o cliente do agendamento realizado e adicionando informações sobre serviços oferecidos pelo hotel ou atividades nos seus arredores que vão incentivar o hóspede a garantir sua estadia;
- Oferecer descontos e promoções para clientes que possuam possibilidade de cancelar, caso façam o pagamento antecipado e sem reembolso;

Possíveis consequências dos planos de ação sugeridos

As consequências dos planos de ação referentes aos **não cancelamentos** são difíceis de serem medidas sem maiores informações sobre custos atuais das operações dos hotéis.

No entanto, é de se imaginar que a **otimização** da **gestão**, além de **aumentar o lucro** e **facilitar** o trabalho dos administradores e funcionários, tem grande potencial de impactar positivamente as finanças, principalmente devido à maior **previsibilidade** do fluxo de caixa.

Já no que diz respeito aos cancelamentos, embora não tenhamos dados sobre as receitas atuais, podemos fazer uma estimativa simples (que precisará ser confirmada com dados reais no decorrer do tempo), utilizando porcentagens.

Considerando que cerca de **37%** das reservas são canceladas, o hotel fica com apenas **63%** do seu potencial de receita.

Sem os dados do modelo, seria possível aos administradores realizarem campanhas anti-cancelamento para todas as reservas. No entanto, dos custos dessas campanhas, **63%** são **desnecessários**, já que apenas **37%** dos hóspedes iriam **cancelar**. Ou seja, focando apenas nos hóspedes que vão cancelar, já temos uma **redução** de **63%** nos **custos** de campanhas anti-cancelamento. Na verdade, como nosso modelo prevê cerca de **80%** dos cancelamentos, teremos uma redução de aproximadamente **70%**, pois iremos focar em

apenas **30%** dos cancelamentos (os outros 7% seriam ignorados devido à acurácia do nosso modelo).

Além dessa redução nos custos das campanhas e consequentemente maior foco nos clientes que podem cancelar, podem ser realizadas campanhas mais assertivas, como promoções somente para esses clientes, o que apesar de reduzir o valor total da receita com esses hóspedes em 10%, 20%, 30% (dependendo da promoção), seria compensado pela presença deles. Ou seja, onde antes a receita era garantidamente 0, agora pode ser maior, sem arriscar prejuízos com clientes que já compareceriam de qualquer maneira.

Para calcular o quão maior essa receita poderia ser, vamos utilizar uma estimativa simples: se conseguirmos converter, com diferentes campanhas, **metade** dos **30%** de hóspedes que identificamos que cancelariam, teremos mais **15%** acrescentados à receita inicial de **63%**, o que indica um aumento total de **23%** na receita base.

Além disso, campanhas anti cancelamento que antes precisavam rodar para **todas** as reservas (na nossa base são aproximadamente 120.000 durante 3 anos) agora rodarão para cerca de **30%** disso (aproximadamente 36.000), o que pode propiciar novas estratégias e ideias que não eram possíveis ou viáveis para um número 3 vezes maior de hóspedes.