

Documentação Externa - Projeto de SQL

melhorar o desempenho hoteleiro: **análises geral, financeira e temporal**

tendo como suporte a base de dados ***Hotel Bookings***

Realizado por

Luan Jonas

Duarte Cardoso

Maria João Alves

No âmbito do **trabalho final**,

para professora **Beatriz Duarte** e representante **Wawiwa**

Índice

Introdução.....	3
Metodologia.....	4
I. Análise geral	
II. Análise orientada para a receita	
III. Análise do tempo médio de reserva, antecedência e cancelamento	
Etapas do processo analítico.....	5
I. Limpeza e normalização dos dados, em Python	
II. Análise exploratória e agregação de métricas, em SQL	
Análise e resultados.....	7
I. Análise geral	
II. Análise financeira	
III. Análise temporal	
Conclusões e recomendações.....	14

Introdução

O presente trabalho foi desenvolvido no âmbito do projeto final da TechOf e tem como base a análise do conjunto de dados *Hotel Bookings*, um repositório com mais de 100 mil registos de reservas hoteleiras. Os hotéis analisados, designados por Resort Hotel e City Hotel, apresentam nomes fictícios e são, para efeitos de interpretação dos dados, considerados como unidades de gestão portuguesa.

O principal objetivo do projeto consiste na análise de padrões de reserva, cancelamento e comportamento dos hóspedes, com base em dados reais do setor hoteleiro, relativos ao período compreendido entre julho de 2015 e setembro de 2017. Através da exploração de métricas como o tempo médio de antecedência das reservas, a receita média, a sazonalidade e a segmentação de mercado, pretende-se identificar tendências relevantes e oportunidades de melhoria no desempenho e na tomada de decisão no contexto da gestão hoteleira.

Metodologia

De forma a assegurar a robustez e a relevância prática das conclusões obtidas, a metodologia adotada neste projeto foi estruturada em três eixos analíticos complementares, combinando técnicas de manipulação, limpeza e análise de dados. Para esse efeito, recorreu-se ao Python para o pré-processamento e mineração dos dados, ao SQL para a análise exploratória e agregação de métricas e o Power BI para a visualização e construção dos gráficos associados.

A abordagem metodológica incidiu sobre os seguintes pilares:

I. Análise geral

Numa primeira fase, foi realizada uma análise exploratória das reservas com o objetivo de compreender a estrutura do conjunto de dados e identificar padrões globais de comportamento. Esta etapa permitiu, em particular, comparar o desempenho e os perfis de reserva entre os dois tipos de hotel analisados, evidenciando diferenças relevantes ao nível da sazonalidade, procura e comportamento dos hóspedes.

II. Análise orientada para a receita

A segunda vertente da análise teve como foco a otimização da receita hoteleira. Foram avaliadas variações da taxa média diária (*ADR - Average Daily Rate*) em função de diferentes dimensões, nomeadamente segmentos de mercado, tipologia de quarto e mercados de reserva. Esta análise visou identificar potenciais perdas de receita associadas ao custo médio diário por quarto, contribuindo para uma melhor compreensão do impacto deste no risco de cancelamento.

III. Análise do tempo médio de reserva, antecedência e cancelamento

A terceira abordagem centrou-se numa análise estatística mais detalhada da relação entre a antecedência da reserva (*tempo médio de reserva*), as taxas de cancelamento e o perfil dos clientes. Através do cruzamento destas variáveis, foi possível quantificar o impacto da antecedência no risco de cancelamento, fornecendo uma base empírica para a formulação de recomendações relacionadas com políticas de depósito, gestão de risco e estratégias de overbooking.

Etapas do processo analítico

I. Limpeza e normalização dos dados, em Python

A fase inicial do projeto incidiu sobre a preparação, validação e normalização do conjunto de dados, com o objetivo de garantir a consistência, fiabilidade e adequação da informação às etapas subsequentes de análise.

Numa primeira fase, procedeu-se à importação do dataset original e à realização de uma análise exploratória preliminar, incluindo a verificação da dimensão do conjunto de dados (**119 390 registos e 32 variáveis**), visualização das primeiras observações, análise da estrutura das variáveis e cálculo de estatísticas descritivas gerais. Esta abordagem permitiu identificar padrões globais, potenciais anomalias e variáveis críticas para a análise.

Seguidamente, foi avaliada a integridade dos dados, com especial atenção à deteção de incoerências lógicas. Destacam-se a **identificação de hóspedes “fantasma”** (registos em que o número total de adultos, crianças e bebés era igual a zero), a **existência pontual de valores de ADR negativos** e a **verificação de possíveis valores negativos em variáveis sensíveis**, como cancelamentos prévios e dias em lista de espera.

Na **fase de limpeza dos dados**, os **registos associados a hóspedes “fantasma” foram removidos**, por não representarem estadias reais. Os **valores nulos foram tratados de forma diferenciada**, de acordo com a natureza das variáveis:

Nas **variáveis numéricas** (children, agent, company), os valores em falta foram **substituídos por zero**, interpretando-se como ausência da característica;

Na **variável country**, os valores nulos foram **substituídos pela categoria “undefined”**, evitando a eliminação de observações e preservando informação potencialmente relevante.

Posteriormente, **procedeu-se à correção e normalização dos tipos de dados** (data types).

Após a limpeza e normalização, o **conjunto de dados passou a integrar 119 210 registos válidos, mantendo as 32 variáveis originais**, agora consistentes e devidamente estruturadas. Por fim, os dados tratados foram exportados para o **ficheiro hotel_bookings_CLEAN.csv**, ficando prontos para utilização nas análises subsequentes, nomeadamente em ambiente SQL.

II. Análise exploratória e agregação de métricas, em SQL

Após a preparação e limpeza dos dados, foram desenvolvidas consultas em SQL com o objetivo de explorar o comportamento das reservas, identificar padrões de cancelamento e analisar o desempenho financeiro. Esta fase permitiu:

- **Analisar o volume de reservas e cancelamentos**, tanto de forma global como segmentada por tipo de hotel (City Hotel vs. Resort Hotel), incluindo o cálculo das taxas de cancelamento;
- **Estudar a sazonalidade**, através da agregação trimestral e anual das reservas e cancelamentos, evidenciando variações temporais relevantes;

- **Avaliar o comportamento dos diferentes segmentos de mercado**, comparando volume de reservas, taxas de cancelamento e contributo para a ocupação;
- **Analisar métricas financeiras baseadas no ADR**, como preço médio por hotel, tipo de quarto, tipo de depósito, tipo de estadia e segmento, bem como a estimativa de perdas de receita associadas a cancelamentos;
- **Explorar o impacto do *lead time***, analisando a sua distribuição, evolução temporal e relação com cancelamentos, segmentos de mercado, tipo de hotel, tipo de quarto, hóspedes repetidos e listas de espera;
- **Relacionar variáveis operacionais e financeiras**, como $ADR \times lead\ time$, $ADR \times cancelamento$ e $noites\ reservadas \times rendimento$, permitindo identificar padrões de risco e oportunidades de otimização da receita.

Esta abordagem possibilitou uma visão integrada do desempenho operacional e financeiro dos hotéis, suportando conclusões estratégicas relacionadas com gestão de risco, pricing, overbooking e planeamento de recursos.

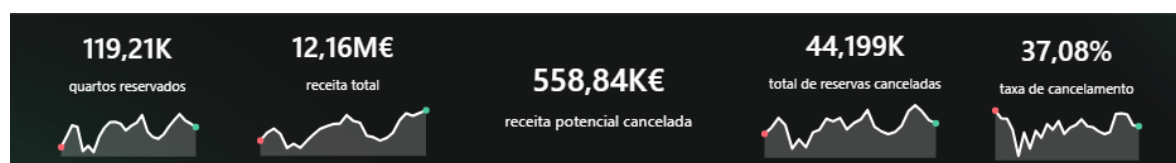
Análise e resultados

I. Análise geral

A análise global do conjunto de dados revela, aproximadamente, **119 mil quartos reservados** e uma **receita acumulada de 12,16 milhões de euros**. Apesar deste volume, o desempenho financeiro é significativamente impactado por uma **taxa de cancelamento global de 37,08%**, evidenciando um desafio estrutural de eficiência operacional (ver fig.1).

O *City Hotel* destaca-se como o principal foco de risco, apresentando uma taxa de cancelamento de **41,79%**, substancialmente superior à registrada no *Resort Hotel* (**27,77%**). Em termos financeiros, este volume de cancelamentos traduz-se numa **perda potencial estimada em 558,84 mil euros**, sublinhando o impacto direto do cancelamento na receita.

Figura 1 - Indicadores



A análise por segmentos de mercado indica que o mercado **Online TA (OTA)** concentra o maior volume de reservas (**56,4 mil**) e simultaneamente uma das taxas de cancelamento mais elevadas (**46,91%**) (ver fig. 2). No *City Hotel*, este mercado é responsável por **32,78% do total de cancelamentos**, configurando-se como um dos principais vetores de risco. Adicionalmente, o segmento **Groups** apresenta um comportamento particularmente instável no *City Hotel*, representando **21,77% dos cancelamentos**, aquando da totalidade de cancelamentos (ver fig. 3).

Figura 2 – Mercado *Online TA*, considerações gerais

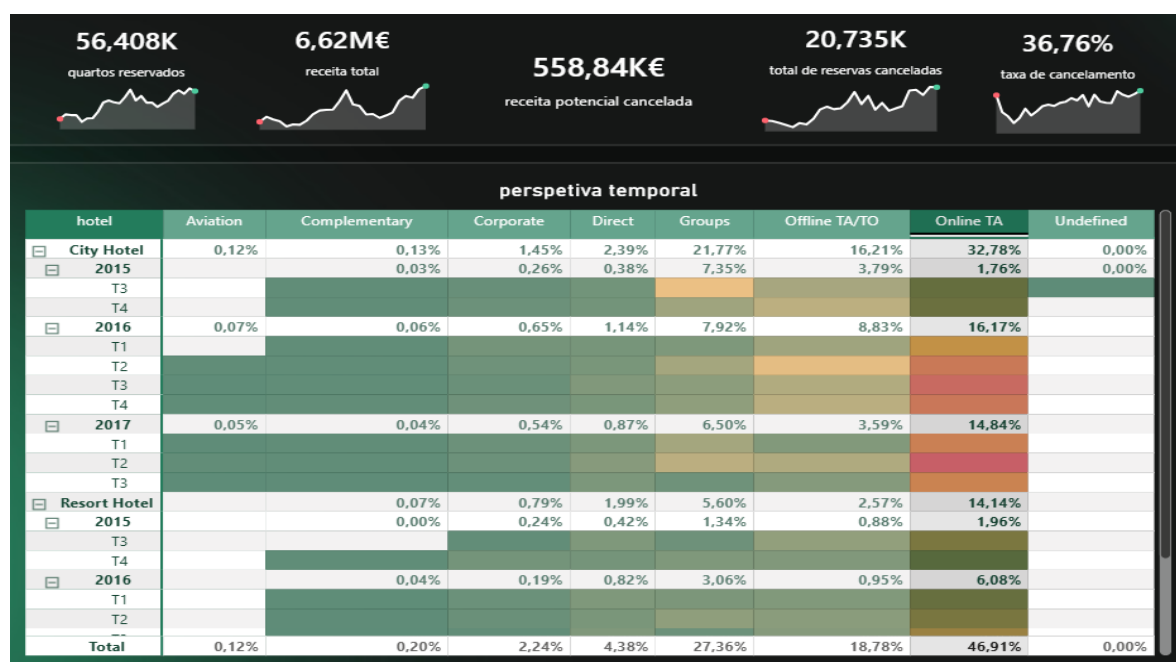
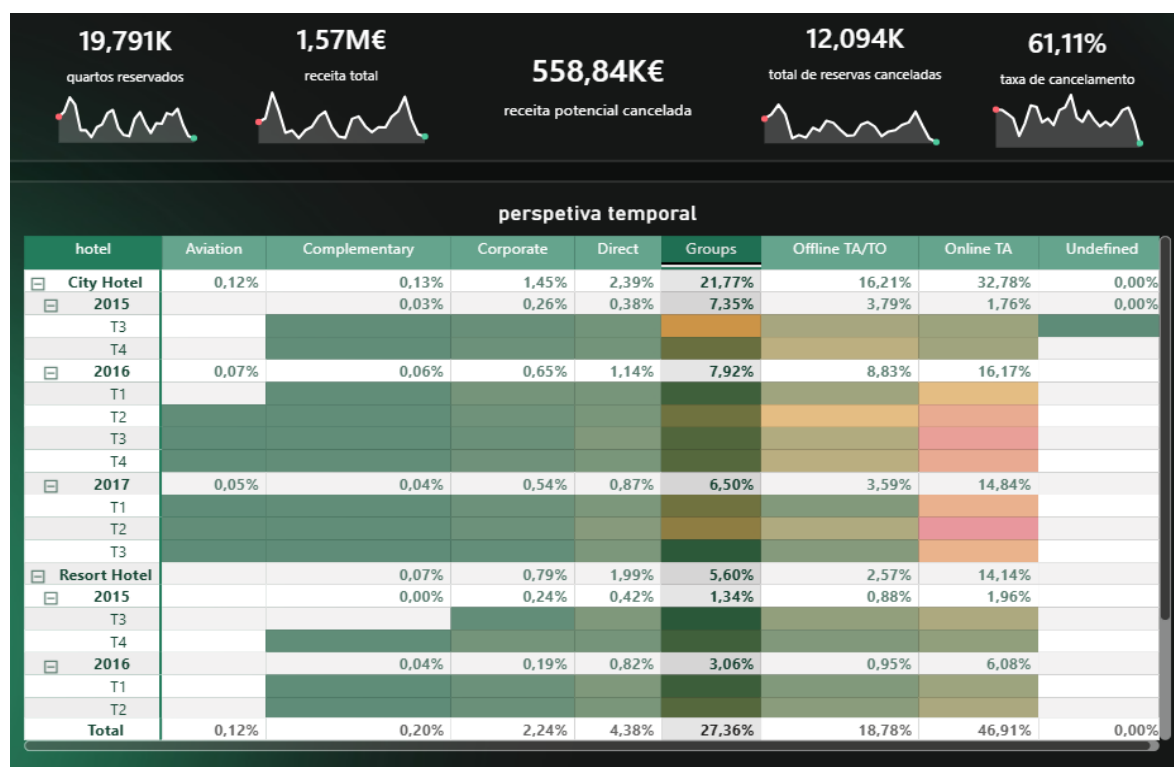


Figura 3 - Mercado *Groups*, considerações gerais



Do ponto de vista temporal, os cancelamentos revelam um padrão sazonal e uma concentração crescente em mercados específicos. O segundo trimestre de 2017 apresentou um volume atípico de desistências (ver fig. 4), enquanto o período de verão se mantém como o mais volátil, sobretudo no *Resort Hotel*, onde os cancelamentos via *OTA* ultrapassam os **56%** (ver fig. 5). Destaca-se ainda uma anomalia também no segmento *Groups* do *Resort Hotel* representando uma taxa de cancelamentos de **22,25%**, todos estes sugerindo a possível perda de contratos ou eventos corporativos de grande dimensão (ver, ainda, fig. 5).

Figura 4 – Cancelamentos por trimestre, considerações gerais

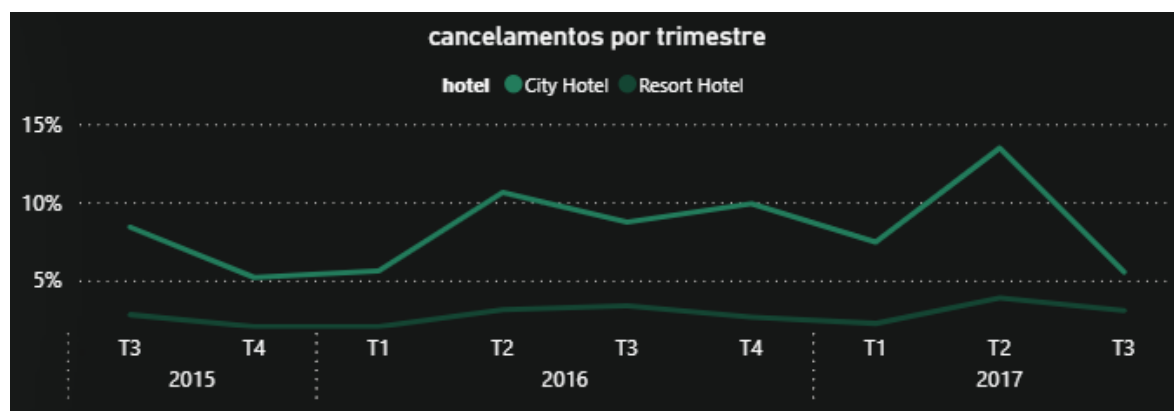


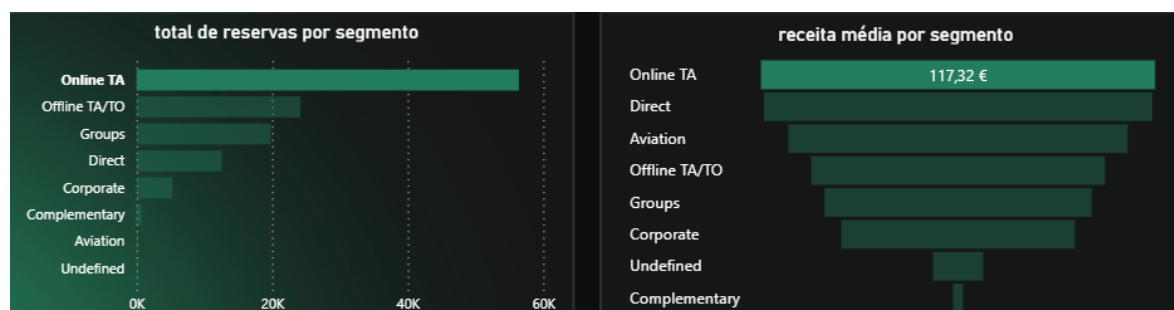
Figura 5 – Resort Hotel, considerações particulares

perspetiva temporal						
hotel	Complementary	Corporate	Direct	Groups	Offline TA/TO	Online TA
Resort Hotel	0,30%	3,16%	7,90%	22,25%	10,22%	56,19%
2015	0,01%	0,94%	1,65%	5,32%	3,51%	7,79%
T3						
T4						
2016	0,16%	0,77%	3,26%	12,17%	3,78%	24,17%
T1						
T2						
T3						
T4						
2017	0,13%	1,44%	2,98%	4,76%	2,93%	24,23%
T1						
T2						
T3						
Total	0,30%	3,16%	7,90%	22,25%	10,22%	56,19%

II. Análise financeira

A análise financeira comparativa das duas unidades hoteleiras revelou padrões relevantes e oportunidades estratégicas de otimização do desempenho financeiro. Apesar de o mercado *Online TA* concentrar o maior volume de reservas em ambos os hotéis, não apresenta a maior receita média por transação, evidenciando que volume não equivale necessariamente a rentabilidade (ver fig. 6).

Figura 6 – Online TA, considerações financeiras



Em contraste, o mercado *Direct* gera uma receita unitária significativamente superior (ver fig. 7), sobretudo no *City Hotel*, posicionando-se como um mercado estratégico prioritário para maximizar a receita por reserva. A análise temporal da receita, organizada por trimestres, evidenciou uma sazonalidade consistente, com maior concentração de receita nos trimestres 2 e 3 e desempenho inferior nos trimestres 1 e 4, bem como um crescimento financeiro sustentado ao longo dos anos em ambos os hotéis (ver fig. 8).

Figura 7 – Mercado *Direct*, considerações financeiras

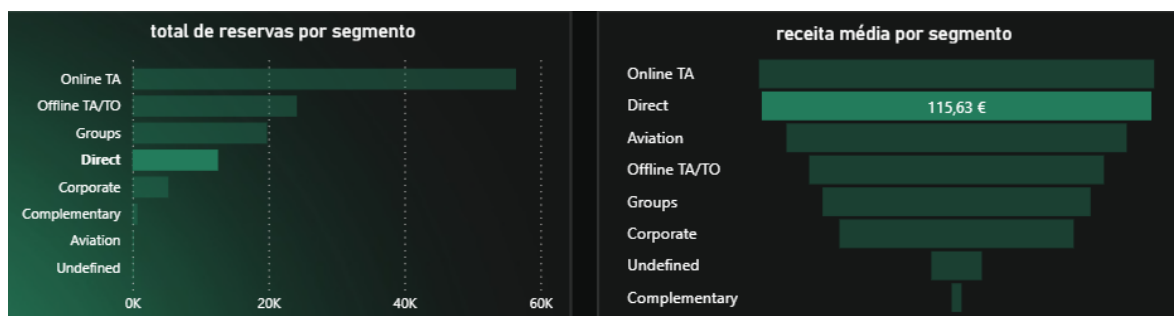
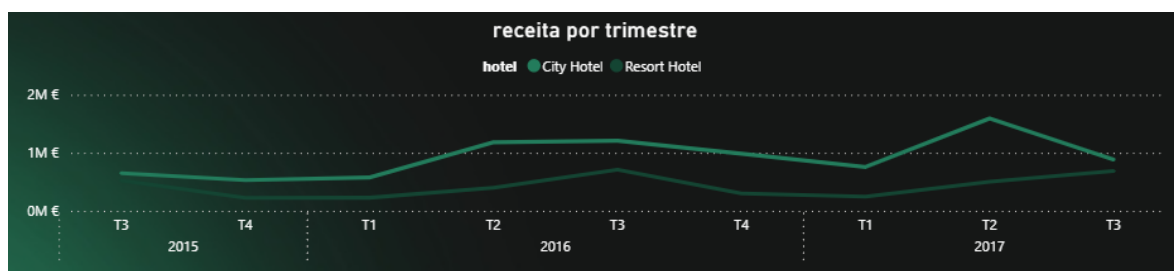


Figura 8 – Receita por trimestre



As elevadas taxas de cancelamento afetam a previsibilidade da receita e revelam comportamentos distintos entre as unidades. No *City Hotel*, o cliente demonstra baixa sensibilidade ao preço, com taxas de cancelamento a diminuírem à medida que a tarifa aumenta, sugerindo uma decisão baseada na qualidade do serviço. No *Resort Hotel*, observa-se maior sensibilidade ao preço até ao escalão dos 200 a 250€, após o qual a taxa de cancelamento volta a diminuir, indicando que, em tarifas mais elevadas, a perceção de valor e qualidade se sobrepõe ao fator preço (ver figs. 9a e 9b).

Figura 9a – Cancelamento por receita, *City Hotel*

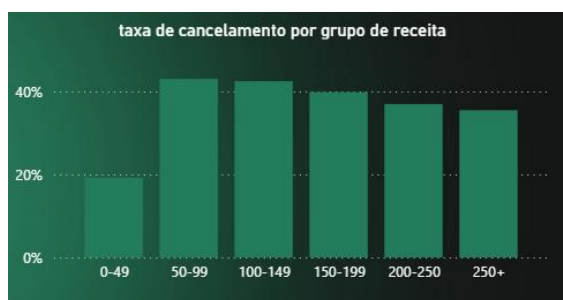
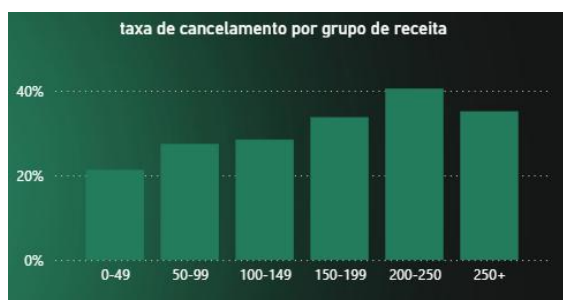
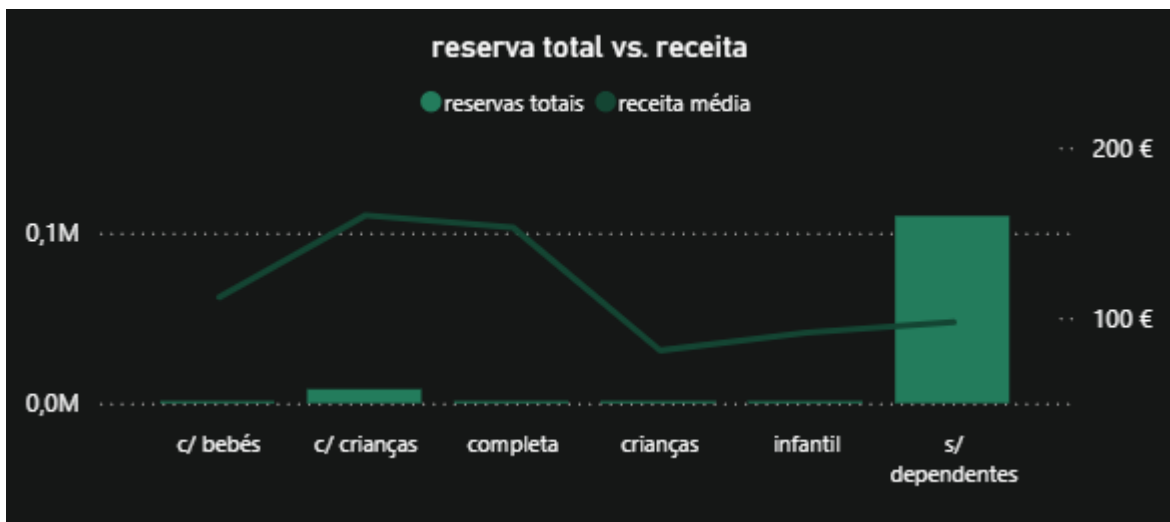


Figura 9b – Cancelamento por receita, *Resort Hotel*



Em termos de composição da receita, o segmento mais rentável corresponde a famílias (adultos com crianças ou bebés a partilhar quarto), apesar de o maior volume de reservas ser gerado por adultos sem dependentes, que apresentam uma receita média inferior por transação. Diferenças operacionais entre os hotéis, como a possibilidade de alojamento separado para crianças no *City Hotel*, influenciam estas dinâmicas.

Figura 10 – Grupos familiares, contabilização total de reservas e considerações financeiras



Globalmente, o *City Hotel* apresenta maior volume de reservas e maior receita total, sendo mais rentável, embora também mais volátil devido à sua maior exposição às variações de mercado. Os resultados indicam que a otimização financeira deve centrar-se no reforço do mercado *Direct*, na gestão estratégica dos cancelamentos para além da variável tarifária, na valorização dos segmentos de maior receita e na utilização da sazonalidade como suporte à gestão de fluxos de caixa e decisões operacionais.

III. Análise temporal

A análise temporal incidiu sobre o **tempo médio de reserva** (*lead time*) enquanto indicador central do comportamento de reserva, dada a sua influência direta na previsibilidade da procura, no risco de cancelamento e outras estratégias. Por conseguinte, foi avaliado o seu impacto para a gestão hoteleira, sendo que o comportamento do hóspede tende a ser regulado pelas épocas consideradas “altas e baixas” (ver fig. 11)

Figura 11 - Tempo médio de reservas, considerações temporais



Os resultados evidenciam diferenças estruturais entre os dois hotéis analisados. O *City Hotel* apresenta um **tempo médio de reserva mais elevado**, refletindo um padrão de reservas planeadas com maior antecedência, típico de turismo corporativo e estadias previsíveis. Em contraste, o *Resort Hotel* caracteriza-se por tempos médios de reserva mais curtos, com forte incidência de **reservas last-minute**, possivelmente associadas à sazonalidade, condições climáticas e decisões espontâneas.

A distribuição do tempo médio de reserva demonstra que reservas efetuadas com maior antecedência apresentam **taxas de cancelamento significativamente superiores**, enquanto reservas próximas da data de chegada tendem a ser mais estáveis. Este padrão confirma que **o risco de cancelamento**

aumenta proporcionalmente à antecedência da reserva, reforçando a necessidade de políticas diferenciadas para reservas de longo prazo (ver fig. 12).

Figura 12 – Cancelamentos por grupo temporal (tempo médio de reserva)



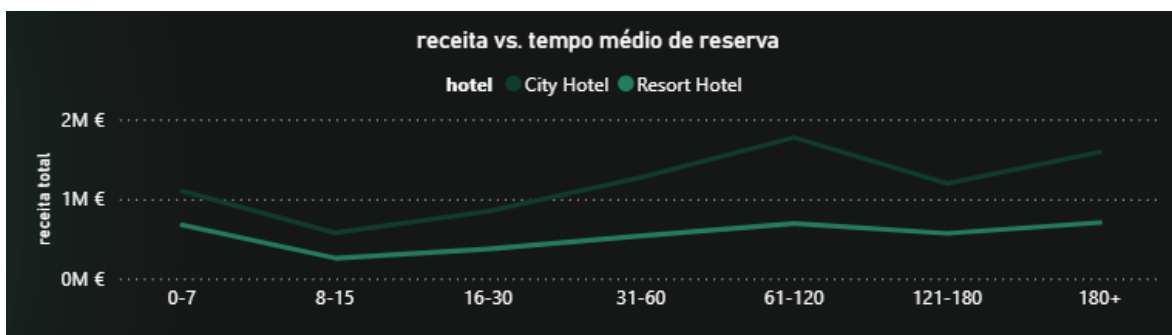
A segmentação da análise revela comportamentos distintos entre perfis de clientes. Segmentos como *Corporate* e *Complementary* reservam com menor antecedência, enquanto *Groups* e *Offline TA/TO* apresentam tempos de reserva mais elevados, refletindo um possível planeamento estruturado e contratos prévios (ver fig. 13). Adicionalmente, clientes internacionais tendem a reservar mais cedo do que clientes nacionais, enquanto hóspedes recorrentes realizam reservas mais próximas da data de estadia, demonstrando maior confiança e familiaridade com o hotel.

Figura 13 – Distribuição temporal por grupo de mercado



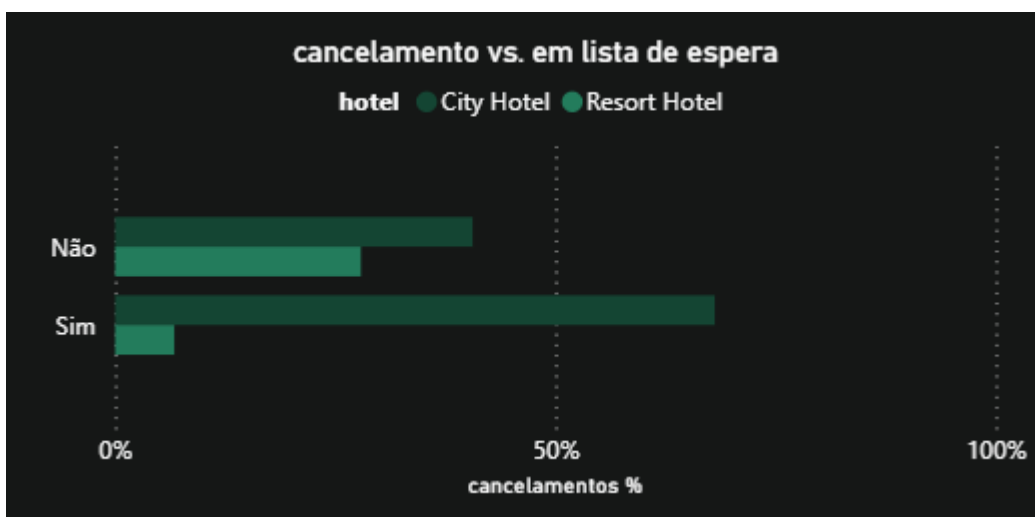
Relativamente ao preço, as tarifas mais elevadas concentram-se em **janelas intermédias de tempo médio de reserva**, identificando um intervalo temporal mais favorável à rentabilidade ([8-15] a [121-180]) (ver fig. 14).

Figura 14 – Receita por tempo médio de reserva



Por fim, a análise demonstra que reservas associadas à **lista de reserva** (*waiting list*) apresentam **maior probabilidade de cancelamento**, sublinhando a importância de estratégias de conversão ativa e comunicação prioritária (ver fig. 15).

Figura 15 – Cancelamentos em relação à lista de espera



Conclusões e recomendações

A análise global evidencia uma operação de elevada escala, mas marcada por uma taxa de cancelamento significativa, que afeta diretamente a eficiência operacional e a previsibilidade da receita, sobretudo no *City Hotel*. Os resultados demonstram que o risco financeiro está fortemente associado a determinados mercados (*Online TA e Groups*) e a reservas efetuadas com maior antecedência, confirmando a necessidade de uma gestão diferenciada por mercado, segmento e tempo médio de reserva.

Do ponto de vista financeiro, conclui-se que o volume de reservas não é sinónimo de rentabilidade. O mercado *Direct* apresenta uma receita média superior e maior estabilidade, devendo ser reforçado estrategicamente. A sazonalidade identificada, com concentração de receita nos trimestres 2 e 3, deve ser utilizada como suporte à gestão de fluxos de caixa e ao planeamento operacional. Verificam-se ainda comportamentos distintos quanto à sensibilidade ao preço: no *City Hotel*, a decisão é predominantemente orientada pela qualidade do serviço, enquanto no *Resort Hotel* o preço influencia o cancelamento até aos escalões intermédios.

A análise temporal confirma que reservas com maior antecedência apresentam maior risco de cancelamento, ao passo que reservas de curto prazo e clientes recorrentes revelam maior estabilidade. Adicionalmente, observa-se uma relação inversa entre tempo médio de reserva e ADR, indicando janelas temporais mais favoráveis à maximização da receita.

Com base nestes resultados, recomenda-se:

- **Implementação de políticas de mitigação de cancelamento**, como sinal não reembolsável no ato da reserva e estratégias de overbooking sustentadas em dados históricos;
- **Reforço do mercado Direct**, através de benefícios exclusivos, comunicação personalizada e estratégias de upselling;
- **Adoção de estratégias de pricing e condições diferenciadas por tempo médio de reserva**, penalizando reservas de longo prazo com maior risco;
- **Utilização do tempo médio de reserva como modelo preditivo**, permitindo antecipar níveis de ocupação e cancelamento e apoiar decisões de planeamento operacional, nomeadamente na alocação de recursos humanos, controlo de custos de staff e gestão de serviços;
- **Valorização de serviços diferenciados**, posicionando cada hotel como ponto de referência (p. ex.: eventos temáticos no *City Hotel*; espaços e serviços adaptados a perfis específicos no *Resort Hotel*);
- **Reaproveitamento de mercados menos rentáveis**, através de promoções e serviços complementares ajustados a cada segmento, como *transfers* fora de horários convencionais ou soluções digitais de acompanhamento de eventos.

Em síntese, a otimização do desempenho passa por uma abordagem integrada que combine gestão ativa do risco de cancelamento, reforço dos mercados mais rentáveis, segmentação eficiente dos hóspedes e adaptação estratégica às dinâmicas temporais da procura.