

***Business Intelligence e Ciência de Dados Aplicados
ao Mercado de Arrendamento Estudantil - Uniplaces***

Mestrado em Ciência de Dados

Pedro Alexandre Pereira Fernandes, 2240277

Leiria, junho de 2026

***Business Intelligence e Ciência de Dados Aplicados
ao Mercado de Arrendamento Estudantil - Uniplaces***

Mestrado em Ciência de Dados

Pedro Alexandre Pereira Fernandes, 2240277

Estágio realizado sob a orientação da Doutora Anabela Moreira Bernardino, Professora da Escola Superior de Tecnologia e Gestão do Politécnico de Leiria e da Doutora Eugénia Moreira Bernardino, Professora da Escola Superior de Tecnologia e Gestão do Politécnico de Leiria

Leiria, junho de 2026

Originalidade e Direitos de Autor

O presente relatório de estágio é original, elaborado unicamente para este fim, tendo sido devidamente citados todos os autores cujos estudos e publicações contribuíram para o elaborar.

Reproduções parciais deste documento serão autorizadas na condição de que seja mencionado o Autor e feita referência ao ciclo de estudos no âmbito do qual o mesmo foi realizado, a saber, mestrado em Ciência de Dados, no ano letivo 2025/2026 da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, Portugal, e, bem assim, à data das provas públicas que visaram a avaliação destes trabalhos.

Agradecimentos

A realização deste relatório e do respetivo estágio só foi possível graças ao apoio e incentivo de várias pessoas que acompanharam de perto este percurso. Em primeiro lugar, quero expressar o meu profundo agradecimento à minha família, pelo apoio incondicional, compreensão e motivação constantes ao longo de toda a minha formação académica. Agradeço igualmente à minha namorada, pela paciência, carinho e encorajamento em todos os momentos, bem como aos meus amigos, pela amizade, apoio e pelos momentos de descontração que contribuíram para equilibrar esta etapa exigente. Aos meus colegas de curso, deixo também uma palavra de apreço pela entreatajuda e pelo espírito de companheirismo que tornou este percurso mais enriquecedor.

Quero ainda agradecer aos meus professores, pelo conhecimento transmitido e pela orientação ao longo do curso, que foram fundamentais para a minha formação académica e profissional. No contexto do estágio, manifesto o meu sincero agradecimento ao Pedro Matos, pela partilha de conhecimentos, pelos ensinamentos e pela visão demonstrada ao longo do acompanhamento do trabalho desenvolvido. Agradeço também aos meus colegas Moe e Alyssa, pela colaboração e boa disposição, assim como à Catarina Coelho, pela ajuda na integração no departamento e pela disponibilidade demonstrada desde o primeiro momento, contribuindo para uma adaptação mais rápida e positiva ao ambiente de trabalho.

Resumo

O presente relatório descreve o trabalho desenvolvido no âmbito do Estágio Curricular do Mestrado em Ciência de Dados, realizado na empresa *Uniplaces*, integrando o departamento financeiro. O estágio teve como principal propósito a aplicação prática de competências analíticas na função de *Data Analyst Intern*, assumindo um papel transversal na otimização dos fluxos de informação e no suporte à tomada de decisão estratégica de todas as equipas da organização.

A crescente importância da Ciência de Dados no panorama empresarial contemporâneo reside na capacidade de transformar vastos volumes de dados brutos em conhecimento estratégico. Através da implementação de processos analíticos e da automatização de fluxos de informação, torna-se possível mitigar a ambiguidade na tomada de decisão, otimizando recursos e identificando oportunidades de crescimento baseadas em evidências empíricas, fundamentais para a competitividade em mercados dinâmicos.

As atividades técnicas centraram-se na gestão do ciclo de vida dos dados, desde a extração até à visualização final, servindo as necessidades de múltiplos departamentos. O trabalho envolveu a utilização de linguagem SQL (*Structured Query Language*) para a manutenção de bases de dados, bem como a criação de *scripts* de automação e lógicas complexas em folhas de cálculo (*Google Sheets*). Complementarmente, foram desenvolvidas soluções de *Business Intelligence* recorrendo a ferramentas como *Looker Studio* e *Tableau*. Destaca-se a intervenção na correção de *scripts* de extração, na resolução de inconsistências de dados e na estruturação de mecanismos automáticos para a priorização de tarefas.

Os contributos do estágio traduziram-se num aumento significativo da integridade da informação corporativa e na eficiência dos processos de análise. Através da centralização de métricas e da automatização de relatórios, foi possível transformar grandes volumes de dados brutos em *insights* acionáveis para as diversas áreas de negócio. Este percurso permitiu consolidar conhecimentos técnicos em manipulação e visualização de dados, demonstrando o valor acrescentado da Ciência de Dados num contexto empresarial dinâmico.

Palavras-chave: Ciência de Dados, *Business Intelligence*, SQL, Visualização de Dados, Automação.

Abstract

This report describes the work carried out as part of the Master's in Data Science placement, undertaken at *Uniplaces* within the finance department. The main purpose of the placement was the practical application of analytical skills in the role of Data Analyst Intern, taking on a cross-functional role in optimising information flows and supporting strategic decision-making across all teams within the organisation.

The growing importance of data science in today's business landscape lies in its ability to transform vast amounts of raw data into strategic insights. Through the implementation of analytical processes and the automation of information flows, it becomes possible to reduce uncertainty in decision-making, optimise resources and identify growth opportunities based on empirical evidence, which are fundamental to competitiveness in dynamic markets.

The technical activities focused on data lifecycle management, from extraction to final visualisation, serving the needs of multiple departments. The work involved using SQL for database maintenance, as well as creating automation *scripts* and complex logic in spreadsheets (*Google Sheets*). In addition, *Business Intelligence* solutions were developed using tools such as *Looker Studio* and *Tableau*. Notable contributions included correcting extraction *scripts*, resolving data inconsistencies and structuring automated mechanisms for task prioritisation.

The contributions made during the internship resulted in a significant improvement in the integrity of corporate information and the efficiency of analytical processes. Through the centralisation of metrics and the automation of *reporting*, it was possible to transform large volumes of raw data into actionable insights for various business areas. This experience enabled the consolidation of technical knowledge in data manipulation and visualisation, demonstrating the added value of Data Science in a dynamic business environment.

Keywords: Data Science, *Business Intelligence*, SQL, Data Visualization, Automation.

Índice

Originalidade e Direitos de Autor	iii
Agradecimentos.....	iv
Resumo.....	v
Abstract	vi
Lista de Figuras	ix
Lista de Tabelas	xii
Lista de Siglas e Acrónimos	xiii
1. Introdução	1
1.1. Relevância do estágio.....	1
1.2. Contextualização da área da Ciência de Dados	1
1.3. Cronograma do estágio.....	2
1.4. Metodologias utilizadas	4
1.4.1. Metodologia CRISP-DM	4
1.4.2. Princípios do CRISP-DM	5
1.4.2.1. Compreensão do Negócio	5
1.4.2.2. Preparação e Integridade de Dados	6
1.4.2.3. Avaliação e Refinamento de Modelos.....	6
1.4.2.4. Transparência na Comunicação de Resultados	6
1.5. Estrutura do documento	6
2. Enquadramento	8
2.1. Caracterização da Entidade de Acolhimento	8
2.1.1. Identificação e Evolução Histórica	8
2.1.2. Missão, Visão e Valores	10
2.1.3. Estrutura Organizacional e Recursos Humanos.....	10
2.1.3.1. Departamento Financeiro	11
2.1.4. Caracterização dos Clientes	13
2.1.5. Modelo de Negócio e Evolução Estratégica	13
2.1.6. Análise SWOT	14
2.2. Ciência de Dados - Conceito e Relevância	15
2.3. Fases do Processo de Ciência de Dados.....	15
2.4. ETL - <i>Extract, Transform, Load</i>.....	16

2.5. <i>Text Mining</i> - Conceito e Importância Empresarial	17
2.6. <i>Business Intelligence</i> e <i>Dashboards</i>	18
2.7. Inteligência Artificial e Modelos Preditivos	19
3. Programa de Estágio	22
3.1. Rotina Diária e Dinâmica de Trabalho no Estágio	22
3.2. Ferramentas e tecnologias utilizadas	24
3.2.1. <i>Slack</i>	24
3.2.2. <i>Google Sheets</i>	25
3.2.3. <i>Python</i>	26
3.2.4. <i>Looker Studio</i>	27
3.2.5. <i>Tableau</i>	28
3.2.6. <i>MySQL</i>	28
3.2.7. <i>Confluence</i>	30
3.3. Atividades desenvolvidas.....	31
3.3.1. Semana 1: Integração e Mapeamento da Infraestrutura de Dados.....	32
3.3.2. Otimização e Automatização do <i>Reporting</i> de Marketing.....	37
3.3.3. Relatório de Desempenho da Ferramenta "Talk to the Landlord"	40
3.3.4. Relatório de Reservas de Último Minuto (<i>Bookings Last Minute</i>).....	49
3.3.5. Migração de <i>Dashboards</i> B2B para <i>Looker Studio</i>	52
3.3.6. Auditoria e Regularização do Processo de Atribuição de Portefólio (<i>Account Management</i>).....	56
3.3.7. Automatização da Extração de Avaliações do <i>Trustpilot</i> com Menção a Agentes.....	61
3.3.8. Desenvolvimento de um Modelo de Previsão de Preços	67
3.3.9. Reestruturação e Automação da Matriz de Análise de Cancelamentos em <i>Tableau</i>	71
3.3.10. Análise de Fatores de Conversão de Ofertas e Modelo Preditivo de Sucesso	74
3.4. Análise crítica e proposta de melhorias	85
4. Conclusão	88
Bibliografia.....	90

Lista de Figuras

Figura 1 - Cronograma do estágio	2
Figura 2 - Estrutura hierárquica principal da <i>Uniplaces</i>	10
Figura 3 - Departamento Financeiro	12
Figura 4 - Processo de ETL.....	16
Figura 5 - Integração do <i>Slack</i> na <i>Uniplaces</i>	25
Figura 6 - Interface Inicial do <i>MySQL</i>	30
Figura 7 - <i>Request</i> feito no <i>Confluence</i>	31
Figura 8 - Principais tabelas de dados da <i>Uniplaces</i>	33
Figura 9 - Ciclo de vida de um Booking Request	35
Figura 10 - <i>Dashboard</i> antes da automatização	38
Figura 11 - <i>Dashboard</i> depois da automatização.....	38
Figura 12 - <i>Script</i> da extração dos dados no <i>Apps Script</i>	39
Figura 13 - Dados referentes à <i>Revenue</i> e <i>Acquisition Cost</i> da <i>Uniplaces</i>	40
Figura 14 - Pedido do Relatório de Desempenho "Talk to the Landlord"	41
Figura 15 - Consulta inicial no SQL	42
Figura 16 - Mensagens recebidas por IA e pelo Agente	43
Figura 17 - Perspetiva semanal entre IA e Agente.....	43
Figura 18 - Configuração de um <i>Trigger</i>	44
Figura 19 - Automatização de cinco <i>Triggers</i> no <i>AppScript</i>	44
Figura 20 - Ligação do <i>Tableau</i> à base de dados da <i>Uniplaces</i>	45
Figura 21 - Consulta SQL no <i>Tableau</i>	46
Figura 22 - <i>Heatmap</i> semanal	47
Figura 23 - Primeira versão do <i>Dashboard</i> Final.....	47
Figura 24 - <i>Dashboard</i> Final para a Ferramenta "Talk To The Landlord"	48
Figura 25 - Reunião Semanal	49
Figura 26 - Consulta SQL desenvolvida para a extração dos dados de reservas de último minuto	50
Figura 27 - Primeira versão do relatório <i>Bookings Last Minute</i> no <i>Google Sheets</i>	51
Figura 28 - Resultado final do relatório <i>Bookings Last Minute</i> com painel de indicadores de síntese e formatação condicional automatizada	52
Figura 29 - Pedido formal de migração e levantamento de requisitos no <i>Confluence</i>	53

Figura 30 - Interface de seleção de conectores nativos no <i>Looker Studio (MySQL)</i>	54
Figura 31 - Configuração da <i>Query SQL</i> customizada para extração de dados B2B	54
Figura 32 - Painel de edição e gestão da funcionalidade <i>Extract Data</i>	55
Figura 33 - <i>Dashboard</i> finalizado no <i>Looker Studio</i> com integração de filtros dinâmicos e métricas de parceiros	56
Figura 34 - Consulta SQL para identificação de senhorios sem agente atribuído no mercado de Portugal	57
Figura 35 - Estrutura da folha de cálculo para o processo de <i>Bulk Update</i> com mapeamento de <i>IDs</i> de senhorios e agentes	58
Figura 36 - Interface do sistema OPS para carregamento e atualização massiva de dados de senhorios (<i>Landlords</i>)	58
Figura 37 - Verificação pós-atualização via <i>MySQL</i> confirmando a atribuição correta dos agentes ao portefólio	59
Figura 38 - Consulta SQL ao esquema <i>agency_staff</i> para obtenção do identificador único (<i>ID</i>) da agente Diana Santos	60
Figura 39 - Interrogação à base de dados para extração da listagem integral de senhorios associados à conta da agente Juliana Silva.....	60
Figura 40 - Estrutura do ficheiro de <i>Bulk Update</i> no <i>Google Sheets</i> com a substituição dos <i>IDs</i> para reequilíbrio de portefólio entre agentes.....	61
Figura 41 - Pedido formal de BI relativo à extração de avaliações do <i>Trustpilot</i> com menção a agentes	62
Figura 42 - Ficheiro de inserção manual de avaliações utilizado anteriormente ao desenvolvimento da solução automatizada	63
Figura 43 - Script inicial de ligação à <i>Trustpilot</i> API desenvolvido em <i>Python</i> no <i>Visual Studio Code</i>	64
Figura 44 - Resultado da extração de avaliações de cinco estrelas convertido em <i>DataFrame</i> através da biblioteca <i>Pandas</i>	64
Figura 45 - Script desenvolvido em <i>Google Apps Script</i> para extração automatizada das avaliações do <i>Trustpilot</i> com definição do intervalo temporal mensal.....	65
Figura 46 - Dicionário de variantes ortográficas dos nomes dos agentes implementado no <i>script</i> de identificação	65
Figura 47 - Função <i>findMentionedAgent</i> responsável pela identificação do agente mencionado no texto da avaliação	65

Figura 48 - Implementação da iteração multilingue sobre as avaliações extraídas da <i>Trustpilot</i> API	66
Figura 49 - Resultado final do relatório mensal de avaliações com menção a agentes no <i>Google Sheets</i>	66
Figura 50 - Gráfico de barras com a contagem de menções por agente no período analisado	67
Figura 51 - Interface de demonstração (<i>Proof of Concept</i>) desenvolvida em <i>Streamlit</i> para a inserção manual de características e comodidades do imóvel	68
Figura 52 - Output do modelo de <i>Machine Learning</i> com a apresentação do preço alvo sugerido e o respetivo intervalo de mercado para otimização da conversão	69
Figura 53 - Auditoria ao comportamento do consumidor em 2025	70
Figura 54 - Distribuição da eficiência de preços no inventário de 2026.....	70
Figura 55 - Matriz estática semanal em folha de cálculo	72
Figura 56 - Ambiente de desenvolvimento e agregação de dados no <i>Tableau</i>	73
Figura 57 - Visualização final do <i>Dashboard</i> mensal de cancelamentos	73
Figura 58 - Consulta SQL desenvolvida para a extração do conjunto de dados de ofertas ativas nas cidades prioritárias a partir de 2024.....	75
Figura 59 - Impacto da presença de planta do imóvel no volume médio de pedidos de reserva	76
Figura 60 - Volume médio de pedidos de reserva por política de cancelamento.....	77
Figura 61 - Impacto da reserva instantânea e da inclusão de despesas no volume médio de pedidos de reserva	78
Figura 62 - Volume médio de pedidos de reserva por cidade e tipologia de alojamento	78
Figura 63 - Relação entre o preço de arrendamento e o volume de pedidos de reserva por tipologia de alojamento	79
Figura 64 - Matriz de correlação entre as variáveis numéricas do conjunto de dados e o volume de pedidos de reserva	80
Figura 65 - Importância relativa das variáveis preditoras no modelo <i>Random Forest Regressor</i>	82
Figura 66 - Extração SQL do volume total de pedidos agregado por mês e tipologia de alojamento	84
Figura 67 - Evolução sazonal do volume de pedidos de reserva por tipologia de alojamento ao longo do ano.....	84

Lista de Tabelas

Tabela 1 - Análise SWOT <i>Uniplaces</i>	14
--	----

Lista de Siglas e Acrónimos

API - *Application Programming Interface*
BI - *Business Intelligence*
CEO - *Chief Executive Officer*
CFO - *Chief Financial Officer*
CRISP-DM - *Cross-Industry Standard Process for Data Mining*
CSV - *Comma-Separated Values*
DDDM - *Data-Driven Decision Making*
EDA - *Exploratory Data Analysis*
ESTG - *Escola Superior de Tecnologia e Gestão*
ETL - *Extract, Transform, Load*
GET - *Método de requisição HTTP para obtenção de dados*
HR - *Human Resources*
HTTP - *Hypertext Transfer Protocol*
IA - *Inteligência Artificial*
ID - *Identification*
JSON - *JavaScript Object Notation*
KPI - *Key Performance Indicator*
MAE - *Mean Absolute Error*
NLP - *Natural Language Processing*
OPS - *Operations System (Uniplaces)*
RDBMS - *Relational Database Management System*
SQL - *Structured Query Language*
SWOT - *Strengths, Weaknesses, Opportunities, Threats*
TI - *Tecnologias de Informação*
URL - *Uniform Resource Locator*

1. Introdução

Este relatório tem como finalidade descrever, de modo organizado e pormenorizado, as atividades realizadas durante o estágio curricular, com duração total de 1040 horas, desenvolvido no âmbito da unidade curricular de Estágio do Mestrado em Ciência de Dados, ministrado pela Escola Superior de Tecnologia e Gestão (ESTG) do Instituto Politécnico de Leiria.

O presente relatório traduz a concretização dos conhecimentos e metodologias adquiridos ao longo do percurso académico, evidenciando a sua aplicação em contexto profissional. A *Uniplaces* serviu como entidade de acolhimento e ambiente propício ao aperfeiçoamento de competências técnicas, analíticas e profissionais no domínio da Ciência de Dados.

Para além da extração, transformação e visualização de dados estruturados, o estágio incluiu também a aplicação de técnicas de *Text Mining* sobre conteúdo textual não estruturado, nomeadamente avaliações de utilizadores recolhidas a partir da plataforma *Trustpilot*, com o objetivo de identificar menções a agentes de apoio ao cliente e estruturar essa informação para análise operacional.

1.1. Relevância do estágio

O estágio curricular constitui uma etapa essencial do percurso académico, permitindo articular os conhecimentos teóricos adquiridos ao longo do curso com a sua aplicação prática em ambiente profissional. No contexto do Mestrado em Ciência de Dados, esta experiência assume particular importância, uma vez que a área da ciência de dados desempenha atualmente um papel estratégico nas organizações modernas, apoiando a transformação digital, a otimização de processos e a tomada de decisões baseadas em informação.

Assim, o estágio confere ao estudante a oportunidade de vivenciar desafios reais, promovendo uma aprendizagem ativa e a consolidação de competências técnicas, analíticas e organizacionais indispensáveis ao exercício profissional.

1.2. Contextualização da área da Ciência de Dados

A Ciência de Dados é uma área multidisciplinar que integra conceitos de estatística, programação, gestão e análise de informação, com o objetivo de extrair conhecimento útil e

gerar valor a partir de grandes volumes de dados. No contexto empresarial, a sua aplicação permite compreender padrões de comportamento, prever tendências e apoiar decisões estratégicas.

Neste enquadramento, o estágio na *Uniplaces* (empresa tecnológica dedicada ao arrendamento de alojamentos a nível internacional) proporcionou um contexto real para a aplicação destes princípios, possibilitando o contacto direto com dados relevantes e desafios concretos associados ao negócio digital.

1.3. Cronograma do estágio

O estágio curricular teve uma duração total de 1040 horas, decorrendo entre 15 de setembro de 2025 e 20 de março de 2026, na empresa *Uniplaces*, em Lisboa. Ao longo deste período, as atividades foram organizadas e distribuídas de forma a garantir uma evolução gradual das tarefas, acompanhada por uma aprendizagem progressiva e autónoma das ferramentas de análise e gestão de dados utilizadas pela organização.

O cronograma do estágio, ilustrado na Figura 1, apresenta a sequência temporal das principais fases de trabalho desenvolvidas, desde a integração inicial e familiarização com o ambiente empresarial até à realização de projetos técnicos e à transmissão final de conhecimento interno.

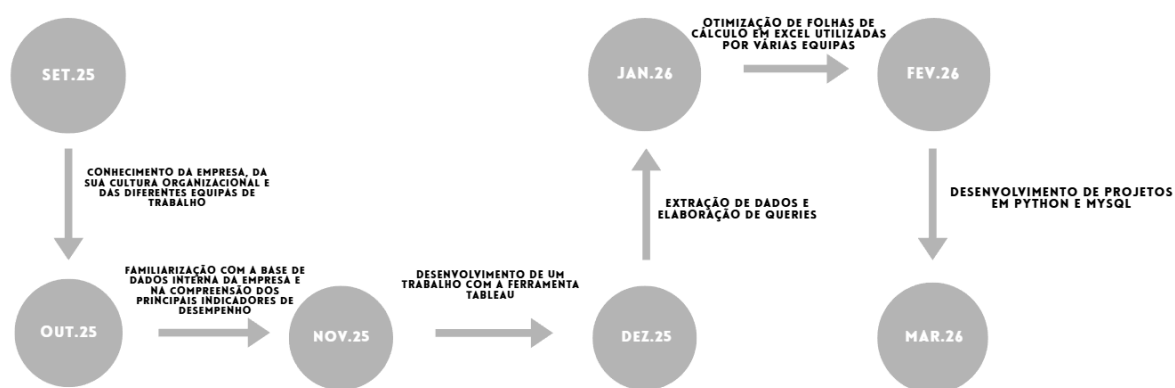


Figura 1 - Cronograma do estágio

O estágio iniciou-se com a integração na cultura organizacional e equipas (setembro-outubro), fase que permitiu avançar para a análise de bases de dados e *Key Performance Indicators* - KPI (outubro-novembro). Esta base teórica e exploratória viabilizou o desenvolvimento em *Tableau* (novembro-dezembro) e a subsequente extração de dados via

queries (dezembro-janeiro), culminando na otimização de processos em *Excel* (janeiro-fevereiro) e na execução de projetos técnicos em *Python* e *MySQL* (fevereiro-março).

A primeira fase, compreendida entre setembro e outubro de 2025, foi dedicada ao conhecimento da empresa, da sua cultura organizacional e das diferentes equipas de trabalho. Este período de integração permitiu compreender a dinâmica interna da *Uniplaces*, o fluxo de informação entre departamentos e o papel do setor de dados no apoio às decisões operacionais e estratégicas.

Entre outubro e novembro de 2025, o foco incidiu na familiarização com a base de dados interna da empresa e na compreensão dos principais indicadores de desempenho KPI utilizados para monitorizar as operações. Esta etapa foi essencial para adquirir autonomia na consulta e manipulação de dados, assegurando uma utilização eficiente dos recursos disponibilizados.

A fase seguinte, de novembro a dezembro de 2025, consistiu no trabalho com a ferramenta *Tableau*, em colaboração com a equipa de *Product*, com o objetivo de desenvolver um *dashboard* analítico orientado para o acompanhamento e avaliação da implementação de uma solução de Inteligência Artificial na plataforma da *Uniplaces*. Este projeto inicial permitiu aplicar metodologias de visualização e comunicar resultados de forma clara e interativa.

Entre dezembro de 2025 e janeiro de 2026, as atividades centraram-se na extração de dados e na elaboração de *queries*, com o propósito de gerar *insights* para a equipa de *marketing*. O objetivo principal foi identificar os *promocodes* mais eficazes e os embaixadores com melhor desempenho, contribuindo para a definição de estratégias de *marketing* mais direcionadas.

Durante o período de janeiro a fevereiro de 2026, foi dada atenção à otimização de folhas de cálculo em *Excel* utilizadas por várias equipas, cujas dimensões e complexidade estavam a comprometer o desempenho. Paralelamente, houve um processo de aprendizagem e implementação da ferramenta *Looker Studio*, com vista à migração de tabelas provenientes do *Google Sheets*, garantindo maior eficiência e estabilidade na apresentação dos dados.

Posteriormente, entre fevereiro e março de 2026, foram desenvolvidos projetos em *Python* e *MySQL*, com ênfase na otimização dos preços de arrendamento submetidos por senhorios e na identificação das características com maior impacto nas reservas de alojamentos nas principais cidades onde a *Uniplaces* atua. Estes projetos, de natureza exploratória e preditiva, reforçaram a componente técnica do estágio e evidenciaram a aplicabilidade prática dos modelos de ciência de dados no contexto do negócio.

Finalmente, na fase terminal, decorreu a transferência de conhecimento e documentação para outra colega estagiária, assegurando a continuidade dos projetos iniciados e a manutenção da coerência metodológica nos processos internos. Esta etapa constituiu uma oportunidade para consolidar as aprendizagens e refletir sobre o percurso desenvolvido ao longo dos seis meses de estágio.

1.4. Metodologias utilizadas

A execução das atividades ao longo do estágio baseou-se na aplicação de metodologias estruturadas de análise de dados, com especial enfoque no modelo CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Esta abordagem foi adaptada à realidade dinâmica da *Uniplaces*, permitindo organizar o fluxo de trabalho de forma cíclica e orientada a objetivos de negócio, garantindo que cada intervenção técnica estivesse alinhada com as necessidades transversais da organização.

A adoção desta metodologia revelou-se fundamental para assegurar a integridade e a fiabilidade da informação, desde a extração de dados via *Structured Query Language* (SQL) até à visualização final em ferramentas de *Business Intelligence* (BI). A flexibilidade do modelo permitiu uma preparação de dados rigorosa e a incorporação de melhorias contínuas, assegurando entregas incrementais que responderam com precisão às exigências operacionais e financeiras de diversos departamentos.

1.4.1. Metodologia CRISP-DM

A metodologia *Cross-Industry Standard Process for Data Mining* (CRISP-DM) foi adotada como o pilar estrutural para a organização de todas as fases do projeto. Ao contrário de metodologias focadas exclusivamente no desenvolvimento de produto, esta estrutura foi concebida especificamente para responder a projetos de análise de dados, onde a incerteza reside na natureza da informação e na descoberta de padrões [1]. O processo organiza-se de forma cíclica, permitindo que a compreensão do negócio oriente a extração e a preparação técnica dos dados, garantindo que as soluções finais, como os *dashboards* analíticos e os relatórios financeiros, possuam relevância estratégica para a organização.

O fluxo de trabalho foi dividido em etapas iterativas que facilitaram a adaptação a novos requisitos e a correção de inconsistências em tempo real. Cada ciclo iniciava-se com a exploração das bases de dados através de consultas SQL, seguida de uma fase rigorosa de

tratamento e limpeza de dados, essencial para assegurar a fiabilidade dos indicadores apresentados. Esta abordagem permitiu que o desenvolvimento não fosse linear, mas sim um processo de refinamento contínuo, onde o *feedback* recebido após a visualização dos dados em ferramentas como *Tableau* ou *Looker Studio* permitia retroceder e ajustar as lógicas de cálculo sempre que necessário.

A colaboração transversal com os diferentes departamentos da *Uniplaces* foi outro aspeto central desta metodologia. Embora não tenha sido seguida uma estrutura rígida de rituais formais, o espírito de agilidade esteve presente através de uma comunicação transparente e de inspeções constantes aos resultados parciais. Esta abertura permitiu identificar precocemente erros de lógica ou sobreposição de dados, como verificado na reestruturação de ficheiros financeiros, assegurando que o progresso fosse visível para todos os envolvidos e que as entregas finais estivessem perfeitamente alinhadas com as prioridades operacionais da empresa.

Finalmente, a adoção desta metodologia favoreceu uma entrega de valor incremental, em que cada fase concluída representava um avanço concreto na qualidade da informação corporativa. A flexibilidade do modelo CRISP-DM permitiu gerir de forma independente tarefas distintas, desde a auditoria de códigos promocionais até à manutenção de *dashboards* de *performance*, garantindo que, mesmo perante a diversidade de pedidos, se mantivessem o rigor técnico e a eficácia na transformação de dados brutos em conhecimento útil para a tomada de decisão.

1.4.2. Princípios do CRISP-DM

A implementação desta metodologia foi guiada por pilares essenciais, tais como a compreensão do negócio (Secção 1.4.2.1), a preparação e integridade de dados (Secção 1.4.2.2), a avaliação e refinamento de modelos (Secção 1.4.2.3) e a transparência na comunicação de resultados (Secção 1.4.2.4). Estes elementos garantiram que as soluções desenvolvidas, desde a automação de relatórios até à centralização de métricas de performance, estivessem em permanente alinhamento com os objetivos estratégicos da empresa.

1.4.2.1. Compreensão do Negócio

A aplicação do CRISP-DM iniciou-se com a imersão nas necessidades estratégicas da *Uniplaces*, convertendo requisitos operacionais em objetivos de análise de dados. Este princípio garantiu que todas as intervenções técnicas, desde o suporte ao departamento financeiro até ao auxílio às equipas de operações e *marketing*, estivessem alinhadas com as metas da

organização. Ao definir claramente os indicadores de desempenho (KPI) antes da fase de execução, assegurou-se que as soluções desenvolvidas respondessem a problemas reais, como a necessidade de monitorização de reservas e a gestão de embaixadores.

1.4.2.2. Preparação e Integridade de Dados

Um dos pilares centrais do trabalho consistiu na exploração exaustiva e no tratamento rigoroso das bases de dados através de consultas SQL. Esta fase foi determinante para garantir a fiabilidade da informação, envolvendo a limpeza de dados, a resolução de inconsistências e a consolidação de fontes dispersas em repositórios únicos. Ao estabelecer uma "fonte única de verdade", foi possível eliminar redundâncias e garantir que as análises subsequentes, tanto em folhas de cálculo como em *dashboards* de BI, refletissem com precisão o estado real da operação.

1.4.2.3. Avaliação e Refinamento de Modelos

Dada a natureza iterativa da Ciência de Dados, o processo baseou-se num ciclo de *feedback* constante e refinamento das soluções implementadas. Este princípio permitiu que, após a visualização inicial dos dados em ferramentas como *Tableau* ou *Looker Studio*, se procedesse ao ajuste de lógicas de cálculo ou à reestruturação de tabelas dinâmicas sempre que se detetassem erros de sobreposição ou lacunas na informação. Esta validação contínua assegurou que o produto final não fosse apenas tecnicamente correto, mas também funcional e intuitivo para os utilizadores finais de todos os departamentos.

1.4.2.4. Transparência na Comunicação de Resultados

A transparência e a entrega incremental foram fundamentais para manter os diversos intervenientes informados sobre o progresso dos projetos. Em vez de uma entrega única e final, optou-se pela disponibilização faseada de relatórios e automações, permitindo uma inspeção regular do trabalho por parte da supervisão e das equipas afetadas. Esta abordagem facilitou a deteção precoce de problemas e a incorporação imediata de melhorias, garantindo que as ferramentas de suporte à decisão evoluíssem de forma natural e estivessem em permanente sintonia com as prioridades dinâmicas da empresa.

1.5. Estrutura do documento

O presente documento encontra-se dividido em quatro capítulos principais, encadeados de forma a guiar o leitor através do ciclo de vida das atividades realizadas durante o período de estágio.

O Capítulo 2 faz a caracterização da empresa de estágio e apresenta um enquadramento teórico aos principais conceitos abordados no estágio. O capítulo inicia com o contexto organizacional, onde se explora a identidade da *Uniplaces* e o seu modelo operativo no setor do alojamento universitário. Nesta fase, é analisada a relevância estratégica da análise de dados para a sustentabilidade do negócio, abordando como a extração e o tratamento rigoroso da informação permitem otimizar processos em múltiplos departamentos, desde as Finanças até às Operações.

No Capítulo 3, detalha-se o núcleo operacional do estágio, descrevendo o ecossistema tecnológico utilizado. Este abrange desde a interrogação de bases de dados via SQL e o desenvolvimento de *scripts* de automação em *Python* (recorrendo ao *Visual Studio Code*), até à gestão avançada de dados em *Google Sheets* e à criação de *dashboards* em *Tableau* e *Looker Studio*. O capítulo aprofunda as metodologias de limpeza e consolidação de dados, ilustrando como estas soluções contribuíram para a eliminação de redundâncias e para a estruturação de mecanismos de priorização de tarefas. O capítulo finaliza com uma reflexão analítica sobre o ecossistema de dados da empresa. Em vez de descrever tarefas executadas, este espaço foca-se na identificação de pontos de melhoria e na proposta de recomendações estratégicas que visam fortalecer a governação de dados e a escalabilidade das ferramentas de reporte no futuro.

Por fim, o Capítulo 4 sintetiza as conclusões retiradas desta experiência. É realizada uma avaliação do cumprimento dos objetivos inicialmente traçados e uma reflexão sobre a evolução das competências técnicas e interpessoais adquiridas, evidenciando o contributo deste estágio para o percurso académico no Mestrado em Ciência de Dados.

Esta sequência foi planeada para demonstrar de que forma o rigor técnico na manipulação de dados brutos se traduz em valor acrescentado para uma organização moderna e orientada por dados.

2. Enquadramento

O presente capítulo estabelece o enquadramento teórico e metodológico fundamental para a compreensão das soluções desenvolvidas ao longo do projeto. Inicialmente é caracterizada a empresa de estágio (Secção 2.1). Posteriormente são explorados o conceito e a relevância da Ciência de Dados (Secção 2.2) no panorama atual, seguindo-se uma análise detalhada das fases do processo de Ciência de Dados (Secção 2.3) que estruturaram a abordagem técnica.

Adicionalmente, aborda-se o fluxo de *Extract, Transform, Load* (ETL) (Secção 2.4), crucial para a integridade e preparação dos dados, bem como os fundamentos de *Text Mining* (Secção 2.5). Este enquadramento prossegue com a apresentação de conceitos de *Business Intelligence* e *dashboards* (Secção 2.6). Por fim, abordam-se os fundamentos de Inteligência Artificial e a aplicação de modelos preditivos no contexto do estágio (Secção 2.7).

2.1. Caracterização da Entidade de Acolhimento

A presente secção é dedicada à caracterização da entidade de acolhimento do estágio curricular, a *Uniplaces*. Numa primeira instância, é apresentada uma ficha de identificação da empresa (Secção 2.1.1), sendo igualmente abordados a sua evolução histórica e o contexto em que a organização foi fundada. Posteriormente, são descritos a missão, a visão e os valores que norteiam a atividade da empresa (Secção 2.1.2), bem como a estrutura organizacional vigente à data de início do estágio. É ainda realizada uma caracterização dos recursos humanos (Secção 2.1.3), seguindo-se a caracterização e segmentação dos seus clientes (Secção 2.1.4). Adicionalmente, são analisados o modelo de negócio e a evolução estratégica da empresa, com particular enfoque nos mercados considerados prioritários para a organização (Secção 2.1.5). A secção é encerrada com uma análise SWOT (*Strengths, Weaknesses, Opportunities, Threats*) (Secção 2.1.6), através da qual são identificados os principais pontos fortes, pontos fracos, oportunidades e ameaças associados à posição competitiva da *Uniplaces* no mercado atual.

2.1.1. Identificação e Evolução Histórica

A entidade de acolhimento do presente estágio é a *Uniplaces*, uma empresa tecnológica portuguesa com sede em Lisboa, que atua no setor da intermediação digital no mercado de arrendamento habitacional, com enfoque no segmento estudantil. Do ponto de vista jurídico, a empresa opera sob a forma de sociedade por quotas, enquadrada no setor dos serviços de tecnologia e informação.

A *Uniplaces* foi fundada no ano de 2012 por três co-fundadores de diferentes nacionalidades: Miguel Santo Amaro (Portugal), Ben Grech (Reino Unido) e Mariano Kostelec (Argentina), tendo o conceito da plataforma surgido ainda em 2011, durante o período em que os fundadores frequentavam instituições de ensino superior no Reino Unido. A motivação subjacente à criação da plataforma decorreu das dificuldades enfrentadas pelos próprios fundadores na procura de alojamento enquanto estudantes no estrangeiro, tendo sido identificada uma lacuna no mercado com potencial para servir milhões de estudantes a nível global.

O projeto foi o primeiro a ser incubado na *Startup Lisboa*, em 2012, tendo, desde então, registado um crescimento assinalável. Em 2013, a empresa expandiu a sua presença para o Porto e para Londres, alargando posteriormente a sua operação a Madrid em 2014. Na sequência de uma ronda de investimento concluída em outubro do mesmo ano, a *Uniplaces* expandiu-se rapidamente para 39 cidades distribuídas por 8 países europeus.

No que respeita ao financiamento, a empresa angariou um total de investimento na ordem dos 28 milhões de euros, incluindo uma ronda inicial de 3,2 milhões de euros proveniente da *Octopus Investments*, seguida de uma ronda de 22 milhões de euros liderada pela *Atomico*, com a participação da *Caixa Capital*, da *Shilling Capital* e da *Octopus Ventures*.

Atualmente, a plataforma disponibiliza mais de 90.000 propriedades verificadas em mais de 30 cidades europeias, estando a operação concentrada estrategicamente em três mercados considerados prioritários: Portugal, Espanha e Itália. Estes mercados, designados internamente como *key cities*, representam o núcleo da atividade comercial da empresa e constituem os territórios onde a presença da *Uniplaces* é mais consolidada e onde os esforços de crescimento são mais intensivos.

Um marco recente relevante na evolução da *Uniplaces* prende-se com a diversificação do seu modelo de negócio. No início de 2026, a empresa iniciou a gestão direta de imóveis próprios disponibilizados para arrendamento na sua plataforma, representando uma mudança estratégica significativa face ao modelo tradicional de intermediação. Esta iniciativa traduz-se numa maior integração vertical da cadeia de valor, permitindo à organização exercer um controlo mais direto sobre a qualidade da oferta e a experiência proporcionada ao utilizador final, reduzindo simultaneamente a dependência exclusiva de proprietários terceiros.

2.1.2. Missão, Visão e Valores

A missão da *Uniplaces* assenta em dois eixos complementares: por um lado, apoiar os estudantes na identificação e reserva de alojamento adequado para as suas experiências de mobilidade internacional; por outro, auxiliar os proprietários no arrendamento dos seus imóveis de forma eficiente e alinhada com os valores de referência do mercado.

No que concerne à visão estratégica, a empresa ambiciona consolidar-se como uma marca tecnológica de referência global no mercado de alojamento universitário, perspetivando a expansão progressiva para segmentos complementares que respondam às necessidades dos estudantes ao longo da sua experiência académica.

A cultura organizacional da *Uniplaces* encontra-se sustentada em quatro valores nucleares: confiança, paixão, impacto e colaboração.

2.1.3. Estrutura Organizacional e Recursos Humanos

A estrutura hierárquica da *Uniplaces*, vigente à data de início do estágio curricular, encontra-se representada na Figura 2. É possível verificar que a liderança máxima da organização é assegurada pelo *Chief Executive Officer* (CEO), Cyril Jessua, ao qual reportam diretamente sete responsáveis de área, cobrindo as principais funções da empresa:

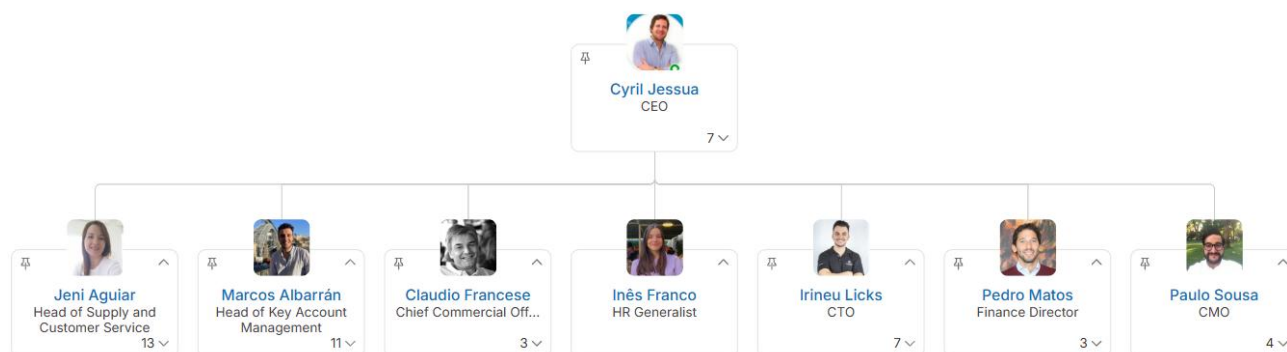


Figura 2 - Estrutura hierárquica principal da *Uniplaces*

Os principais departamentos da *Uniplaces* são:

- *Supply and Customer Service* (Jeni Aguiar) - área responsável pela gestão da oferta de alojamento disponível na plataforma, bem como pelo apoio prestado aos clientes ao longo de todo o processo de reserva e arrendamento;

- *Key Account Management* (Marcos Alberran) - departamento dedicado à gestão das parcerias estratégicas com os principais proprietários e agências imobiliárias presentes na plataforma;
- Comercial (Claudio Francese, *Chief Commercial Officer*) - área responsável pela definição e execução da estratégia comercial da empresa, com foco na captação e retenção de clientes nos mercados onde a *Uniplaces* opera;
- Recursos Humanos (Inês Franco, *HR Generalist*) - departamento encarregue da gestão do capital humano da organização, incluindo processos de recrutamento, integração e desenvolvimento de colaboradores;
- Tecnologias de Informação (Irineu Licks, *Chief Technology Officer*) - área responsável pelo desenvolvimento e manutenção da infraestrutura tecnológica da plataforma, sendo transversal a toda a operação da empresa;
- Financeiro (Pedro Matos, *Finance Director*) - departamento responsável pela gestão financeira e contabilística da organização, incluindo o controlo orçamental e o reporte financeiro;
- Marketing (Paulo Sousa, *Chief Marketing Officer*) - área encarregue da definição da estratégia de comunicação e *marketing* da marca, com foco na captação de novos utilizadores e no reforço da notoriedade da *Uniplaces* nos mercados-chave.

Esta estrutura reflete uma organização de cariz funcional, em que cada área de negócio dispõe de autonomia de gestão sob supervisão direta da administração.

À data do estágio, a *Uniplaces* contava com cerca de 60 colaboradores, distribuídos pela sede em Lisboa e por escritórios presentes em vários países europeus. Aproximadamente 40% dos recursos humanos eram de nacionalidade estrangeira, o que espelha o carácter marcadamente internacional da organização. A empresa apresenta uma estrutura etária predominantemente jovem, com forte representação de perfis especializados nas áreas de tecnologia, produto, *marketing* e análise de dados. O estágio curricular foi integrado na equipa de *Data Analytics*, área de crescente relevância estratégica no seio da organização.

2.1.3.1. Departamento Financeiro

O departamento financeiro da *Uniplaces* encontra-se sob a responsabilidade do *Finance Director*, Pedro Matos, ao qual toda a equipa reporta em matéria financeira. A estrutura interna deste departamento, representada na Figura 3, é composta por três elementos, cada um com responsabilidades bem delimitadas.

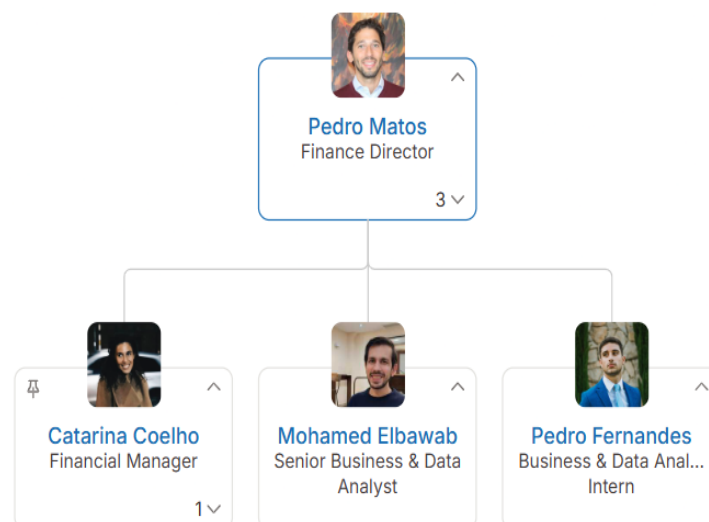


Figura 3 - Departamento Financeiro

A gestão contabilística da organização é assegurada pela *Financial Manager*, Catarina Coelho, sendo da sua responsabilidade a supervisão e validação de todos os processos contabilísticos da empresa. No exercício das suas funções, a *Financial Manager* conta com o apoio de uma colaboradora adicional, Ana Chainho, cuja função consiste no registo e reporte contabilístico corrente, com reporte direto à Catarina Coelho.

No que respeita à vertente de análise de dados, o departamento integra uma equipa dedicada constituída por dois elementos. O Sênior *Business & Data Analyst*, Mohamed Elbawab, assume a liderança técnica desta subequipa, sendo responsável pela condução das análises de maior complexidade e pela supervisão das tarefas de âmbito analítico desenvolvidas no seio do departamento. No âmbito do estágio curricular, o autor do presente relatório integrou esta equipa na qualidade de *Business & Data Analyst Intern*, reportando diretamente ao Sênior *Data Analyst*. A colaboração entre ambos os elementos configurou uma equipa de dados autónoma dentro do departamento financeiro, com foco na produção de análises, indicadores e relatórios de suporte à tomada de decisão financeira e operacional da organização.

Esta estrutura dual, contabilística e analítica, reflete a crescente importância atribuída pela *Uniplaces* à integração entre a gestão financeira tradicional e a análise de dados como ferramenta de apoio à gestão, evidenciando uma abordagem orientada para a tomada de decisão baseada em evidências (*data-driven decision making*).

2.1.4. Caracterização dos Clientes

A caracterização de uma plataforma de intermediação exige, primeiramente, o entendimento da sua natureza enquanto mercado multilateral, onde o valor gerado depende da interação coordenada entre diferentes grupos de utilizadores. Este tipo de modelo de negócio floresce ao minimizar os custos de transação e ao promover efeitos de rede que beneficiam simultaneamente a oferta e a procura [2]. No caso da *Uniplaces*, esta dualidade é evidente na forma como a plataforma serve de ponte entre o ecossistema de alojamento e as necessidades específicas de mobilidade académica e profissional.

O segmento da procura é predominantemente constituído por estudantes internacionais, com um foco muito acentuado em participantes do programa *Erasmus+* e outros regimes de mobilidade académica. Estes clientes, tipicamente na faixa etária dos 18 aos 25 anos, possuem elevada literacia digital e procuram um processo de reserva totalmente online que ofereça garantias de segurança e verificação das propriedades, eliminando a necessidade de deslocações prévias para visitas. Contudo, tem-se verificado um alargamento deste perfil para incluir nómadas digitais e jovens profissionais em início de carreira, que necessitam de estadias de média duração (entre 3 a 9 meses) e valorizam características como a localização central, conectividade de alta velocidade e a inclusão de despesas mensais no valor do aluguer.

Do lado da oferta, a plataforma agrega dois tipos fundamentais de parceiros que garantem a diversidade do inventário. Por um lado, encontram-se os proprietários individuais, que possuem quartos em apartamentos partilhados e procuram rentabilizar o seu imóvel de forma simples e segura através da marca *Uniplaces*. Por outro lado, o segmento profissional é composto por agências imobiliárias, gestores de ativos e operadores de residências universitárias ou espaços de *co-living*. Estes parceiros utilizam a plataforma como um canal de distribuição global estratégico, permitindo-lhes alcançar uma audiência internacional de estudantes que, de outra forma, teria dificuldade em aceder ao mercado de arrendamento local de forma independente.

2.1.5. Modelo de Negócio e Evolução Estratégica

A inovação e a adaptação do modelo de negócio são processos vitais para a sobrevivência de empresas tecnológicas num mercado globalizado e altamente competitivo. A estratégia de uma organização deve ser dinâmica, permitindo o ajuste contínuo dos mecanismos de captura de valor face às mudanças nas preferências dos consumidores e nas estruturas de mercado [3]. No contexto das plataformas digitais, esta evolução traduz-se frequentemente na transição de

modelos puramente transacionais para modelos híbridos que procuram um maior controlo sobre a experiência do utilizador e a qualidade da oferta.

O modelo de negócio da *Uniplaces* baseia-se num sistema de comissionamento sobre as reservas confirmadas, operando historicamente sob uma lógica *asset-light*, na qual a empresa gere a intermediação financeira e logística sem deter os ativos imobiliários. Recentemente, a empresa tem passado por marcos evolutivos significativos que refletem a sua maturidade no mercado europeu. Um exemplo relevante foi a tentativa de expansão para o mercado chinês em 2023, visando captar o vasto fluxo de estudantes desse país. No entanto, a operação enfrentou desafios complexos, nomeadamente barreiras culturais, regulatórias e a forte concorrência de plataformas domésticas já estabelecidas, o que levou a empresa a reorientar o seu foco para o reforço da liderança nos mercados europeus de alta densidade universitária.

Olhando para o futuro próximo, especificamente para 2026, a *Uniplaces* planeia introduzir uma vertente estratégica complementar através da aquisição e gestão de propriedades próprias. Esta nova linha de negócio não substitui o modelo de *marketplace* existente, mas adiciona-lhe uma camada de estabilidade e controlo total sobre o inventário em localizações *premium*. Esta transição para um modelo híbrido permitirá à empresa atuar simultaneamente como intermediário e como operador imobiliário, garantindo maior estabilidade na oferta perante a escassez habitacional europeia, ao mesmo tempo que aumenta as margens de lucro diretas e a capacidade de oferecer serviços personalizados aos seus clientes.

2.1.6. Análise SWOT

A análise SWOT apresentada na Tabela 1 sintetiza os principais fatores internos e externos que condicionam a posição competitiva da *Uniplaces* no mercado atual.

Tabela 1 - Análise SWOT *Uniplaces*

	Fatores Positivos	Fatores Negativos
Ambiente Interno	Forças: Plataforma totalmente digital e escalável; marca reconhecida no mercado europeu; equipa multicultural e especializada; modelo de negócio baseado em comissionamento sem necessidade de gestão direta de imóveis	Fraquezas: Elevada dependência de mercados com forte pressão concorrencial; margens comerciais condicionadas pela necessidade de manter comissões competitivas; vulnerabilidade à reputação online

Ambiente Externo	Oportunidades: Crescimento contínuo da mobilidade estudantil internacional; digitalização progressiva do mercado de arrendamento; potencial de expansão para mercados extra-europeus	Ameaças: Concorrência de plataformas consolidadas como o Airbnb e portais imobiliários locais; crescente regulação do mercado de arrendamento de curta duração; conjuntura macroeconómica instável com impacto na mobilidade estudantil
------------------	---	--

2.2. Ciência de Dados - Conceito e Relevância

A Ciência de Dados (*Data Science*) constitui uma área interdisciplinar que combina métodos e técnicas provenientes da estatística, da matemática, da ciência da computação e do conhecimento de domínio, com o propósito de extrair conhecimento útil e acionável a partir de conjuntos de dados estruturados e não estruturados [4]. Num contexto empresarial cada vez mais orientado para a evidência empírica, a Ciência de Dados assume um papel central na transformação de dados brutos em informação estratégica, contribuindo diretamente para a otimização de processos, a melhoria da experiência do utilizador e o suporte à tomada de decisão [5].

No caso concreto da *Uniplaces*, empresa cuja operação assenta numa plataforma digital que gera e processa diariamente um volume considerável de dados, nomeadamente transacionais, comportamentais e relacionais, a Ciência de Dados revela-se um instrumento indispensável para a monitorização da performance operacional, a avaliação da qualidade do serviço prestado e a identificação de oportunidades de melhoria.

2.3. Fases do Processo de Ciência de Dados

O processo de Ciência de Dados não se circunscreve à análise estatística isolada, sendo antes composto por um conjunto sequencial e iterativo de fases, frequentemente designado na literatura como o *Data Science Lifecycle* [6]. A fase inicial consiste na definição do problema, ou seja, na formulação clara da questão analítica a responder, sendo fundamental o alinhamento entre os objetivos da análise e as necessidades da organização. Segue-se a recolha de dados, que corresponde à identificação e obtenção das fontes de dados relevantes, podendo estas ser internas como bases de dados organizacionais e sistemas de informação ou externas, como API (*Application Programming Interfaces*), plataformas digitais e ficheiros de acesso público. A

qualidade e a representatividade dos dados recolhidos condicionam diretamente a fiabilidade dos resultados obtidos nas fases subsequentes [7].

Após a recolha, procede-se ao processamento e à limpeza dos dados, etapa que engloba a identificação e tratamento de valores omissos, a remoção de duplicados, a normalização de formatos e a transformação de variáveis, com o objetivo de garantir a integridade e consistência do conjunto de dados. Esta fase, frequentemente designada por *Data Wrangling*, é considerada uma das mais morosas e determinantes de todo o processo [7]. Concluída a preparação dos dados, é realizada uma análise exploratória (*Exploratory Data Analysis* - EDA) que, através de técnicas estatísticas e de visualização, permite identificar padrões, tendências, anomalias e relações entre variáveis, orientando as decisões analíticas subsequentes [8].

As fases de modelação e avaliação envolvem, respetivamente, a aplicação de algoritmos estatísticos ou de aprendizagem automática adequados ao problema em estudo e a validação dos modelos com base em métricas de desempenho apropriadas [9]. O processo culmina na comunicação e visualização dos resultados, fase em que as conclusões obtidas são apresentadas de forma clara e acessível aos diferentes *stakeholders* da organização, recorrendo frequentemente a *dashboards*, relatórios e representações gráficas [10].

2.4. ETL - *Extract, Transform, Load*

O processo de ETL (ver Figura 4) constitui uma das componentes fundamentais da engenharia de dados e da gestão de informação nas organizações [11]. Este processo descreve o fluxo pelo qual os dados são extraídos de uma ou mais fontes heterogéneas, transformados de acordo com regras e critérios pré-definidos, e subsequentemente carregados num sistema de destino, como uma base de dados, um armazém de dados (*data warehouse*) ou uma folha de cálculo.



Figura 4 - Processo de ETL

A fase de extração corresponde à recolha de dados a partir das fontes de origem, que podem incluir bases de dados relacionais, API externas ou ficheiros nos mais diversos formatos. A fase de transformação engloba o conjunto de operações aplicadas aos dados extraídos,

nomeadamente a limpeza, a normalização, a agregação e o enriquecimento da informação, assegurando a sua adequação ao formato e às regras de negócio do sistema de destino. Por fim, a fase de carregamento consiste na inserção dos dados transformados no sistema de destino, podendo este processo ser realizado de forma integral (*full load*) ou incremental (*incremental load*), conforme a frequência e o volume de dados envolvidos [11].

No contexto do estágio realizado na *Uniplaces*, o processo de ETL foi amplamente aplicado, designadamente através da extração de dados por via de consultas em SQL sobre as bases de dados internas da organização, seguida da transformação e do carregamento dos dados em folhas de cálculo no *Google Sheets*, com recurso a automatizações que garantiam a atualização periódica da informação disponibilizada às equipas.

2.5. Text Mining - Conceito e Importância Empresarial

O *Text Mining*, também designado por mineração de texto, consiste no processo de extração automatizada de informação relevante e padrões significativos a partir de dados textuais não estruturados [12]. Enquanto os dados estruturados se encontram organizados em formatos tabelares e são facilmente processáveis por sistemas de informação tradicionais, os dados não estruturados como comentários de clientes, avaliações online ou registos de interações encerram um elevado valor informacional que apenas pode ser acedido através de técnicas especializadas de processamento de linguagem natural (*Natural Language Processing* - NLP) [13].

Do ponto de vista empresarial, o *Text Mining* assume particular relevância no que respeita à análise da voz do cliente (*Voice of the Customer*), permitindo às organizações monitorizar a sua reputação online, identificar tendências de satisfação e insatisfação, detetar menções a colaboradores ou serviços específicos e fundamentar decisões operacionais e estratégicas com base em evidências qualitativas sistematizadas [12].

No âmbito da *Uniplaces*, a aplicação do *Text Mining* revestiu-se de especial importância na análise das avaliações submetidas pelos utilizadores na plataforma *Trustpilot*. A extração e análise destes comentários permitiram identificar menções nominais a agentes de apoio ao cliente, possibilitando uma avaliação individualizada da performance dos colaboradores com base no *feedback* orgânico dos utilizadores, sendo esta uma abordagem que alia a objetividade da análise quantitativa à riqueza informacional do conteúdo textual não estruturado.

2.6. *Business Intelligence e Dashboards*

O conceito de *Business Intelligence* (BI) é definido como um sistema interativo baseado em computador que auxilia os decisores na utilização de tecnologias de dados para identificar e resolver problemas de negócio de elevada complexidade [14]. Esta disciplina evoluiu de sistemas de reporte estáticos para ecossistemas dinâmicos que integram o armazenamento de dados e ferramentas de análise, permitindo o que se descreve como a democratização da informação no seio das empresas [15]. Na *Uniplaces*, esta evolução é crítica, uma vez que a agilidade na resposta ao mercado depende da capacidade de consolidar dados provenientes de fontes heterogêneas num repositório centralizado. A base de um sistema de BI robusto reside na criação de um ambiente de dados que garante a consistência e a orientação por temas necessária para que a análise deixe de ser reativa e passe a ser proativa [16]. Esta transformação digital exige que os dados não sejam apenas armazenados, mas convertidos em ativos estratégicos que fundamentam a vantagem competitiva da organização [17].

A fundação técnica do BI sustenta-se na fiabilidade do processamento, onde a arquitetura de informação dita a integridade dos resultados finais apresentados ao utilizador. Durante o estágio, a necessidade de automatizar processos manuais baseou-se na premissa de que a dependência de sistemas de suporte à decisão fragmentados aumenta exponencialmente o risco de erro humano e reduz a eficácia operacional. Uma estrutura de dados bem delineada permite que a organização transcenda a mera recolha de informação e atinja um estado de inteligência organizacional, onde a análise de fluxos complexos se torna sistemática e reproduzível [18]. Ao implementar fluxos de dados centralizados, assegura-se que a comunicação entre os departamentos técnicos e financeiros ocorre sem ruído, promovendo uma cultura verdadeiramente *data-driven*. Este rigor na preparação da informação constitui o passo prévio indispensável para que qualquer análise de métricas, como os indicadores de desempenho (KPI), possua validade estatística e operacional [19].

A face visível deste ecossistema são os *dashboards*, cujo design deve respeitar princípios cognitivos fundamentais para garantir uma comunicação de dados eficiente. Uma interface de visualização eficaz deve obrigatoriamente seguir a taxonomia de "visão geral primeiro, zoom e filtro, e detalhes a pedido", permitindo que o decisor navegue pela informação de forma intuitiva [20]. No desenvolvimento dos painéis na *Uniplaces*, a aplicação destes princípios permitiu que métricas complexas fossem interpretadas de forma célere, evitando a sobrecarga cognitiva dos utilizadores. A visualização de dados não deve ser encarada apenas como uma

componente estética, mas como um elemento de suporte técnico que permite identificar anomalias em conjuntos de dados de grande dimensão através da percepção visual humana [21]. Assim, a clareza visual dos painéis em ferramentas como o *Tableau* ou *Looker Studio* é o que garante que os *insights* gerados são traduzíveis em ações de negócio concretas.

A relevância dos *dashboards* estende-se ao alinhamento estratégico da empresa, atuando como um elo de ligação entre a performance operacional e os objetivos financeiros. Argumenta-se que um painel de bordo bem desenhado não serve apenas para documentar o que aconteceu no passado, mas sim para impulsionar a ação imediata através da monitorização de desvios em tempo real [22]. Ao transpor esta teoria para o contexto da *Uniplaces*, a criação de soluções automatizadas visou colmatar a lacuna existente entre os dados brutos e a estratégia de crescimento, permitindo um acompanhamento rigoroso do *Booking Request Lifecycle*. Este alinhamento permite monitorizar as variações de performance mensal e o cumprimento do orçamento com uma precisão que a literatura identifica como o fator diferenciador entre empresas estagnadas e organizações resilientes [23].

Por fim, a maturidade da infraestrutura de BI e a correta implementação de visualizações de dados constituem o suporte indispensável para a transição para métodos de análise avançada. A visualização de dados funciona como a linguagem fundamental para comunicar ciência de dados, sendo essencial tanto para a exploração inicial de variáveis como para a validação final de resultados complexos [24]. Assim, o rigor teórico aplicado no desenvolvimento do BI durante este projeto não representou apenas uma melhoria na eficiência do reporte; funcionou também como o alicerce necessário para a aplicação de algoritmos de Inteligência Artificial. Desta forma, garante-se que as previsões geradas por modelos avançados assentam em dados historicamente validados e visualmente auditáveis pelos diversos intervenientes organizacionais [25].

2.7. Inteligência Artificial e Modelos Preditivos

A Inteligência Artificial (IA) é definida como o campo da ciência que procura desenvolver sistemas capazes de executar tarefas que, tradicionalmente, exigiriam a aplicação de processos cognitivos humanos, tais como a aprendizagem, o raciocínio e a resolução de problemas [26]. Dentro deste vasto domínio, destaca-se o *Machine Learning* (ML), que se foca na criação de algoritmos que permitem ao computador aprender padrões a partir de dados, sem ser explicitamente programado para cada tarefa específica [27]. No contexto da *Uniplaces*, a

transição para a análise preditiva representa um salto qualitativo em relação ao *Business Intelligence* tradicional, pois permite que a organização deixe de apenas descrever eventos passados e comece a antecipar tendências futuras com base em modelos matemáticos rigorosos [28]. Esta capacidade de previsão é fundamental para otimizar processos financeiros e operacionais, transformando a incerteza dos dados históricos em estimativas de probabilidade com elevado valor estratégico [29].

No âmbito das técnicas de *Supervised Learning*, a regressão assume um papel central quando o objetivo é prever uma variável contínua, como os valores de faturação ou a variação de preços no mercado imobiliário. O sucesso de um modelo preditivo depende da sua capacidade de generalização, ou seja, da sua eficácia ao processar dados novos que não foram utilizados durante a fase de treino [30]. Um dos principais desafios neste processo é o equilíbrio entre o *bias* e a *variance*, procurando evitar que o modelo sofra de *overfitting*, em que o algoritmo memoriza o ruído dos dados em vez de aprender a lógica subjacente [31]. Para mitigar estes riscos, a literatura científica sugere a utilização de métodos de validação cruzada e a seleção criteriosa de atributos, garantindo que o modelo preditivo possui robustez estatística para apoiar a tomada de decisão em ambientes dinâmicos e incertos [32].

Entre os algoritmos de maior destaque para tarefas de regressão complexas encontra-se o *Random Forest Regressor*, uma técnica de *ensemble learning* que combina o poder preditivo de múltiplas árvores de decisão. Este método baseia-se no princípio de *Bootstrap Aggregating*, onde diversas árvores são treinadas em subconjuntos diferentes dos dados, sendo a previsão final obtida através da média aritmética de todas as árvores individuais [33]. A grande vantagem da utilização do *Random Forest* reside na sua resistência a *outliers* e na sua capacidade de lidar com relações não lineares entre as variáveis, características comuns em dados provenientes de plataformas de reservas como a *Uniplaces* [34]. Além disso, este algoritmo fornece métricas de importância das variáveis, permitindo aos analistas identificar quais os fatores que mais influenciam os resultados de negócio, conferindo assim uma camada adicional de interpretabilidade técnica ao processo de modelação [35].

A aplicação prática desta teoria no departamento financeiro permite a criação de modelos de previsão de receita e de comportamento de mercado que reduzem a margem de erro nas projeções anuais. Ao integrar modelos preditivos com a infraestrutura de dados existente, a organização consegue automatizar a geração de *insights* que anteriormente exigiam análises manuais extensas e propensas a falhas [36]. Esta abordagem não só valoriza o capital intelectual

da empresa, como também assegura que a estratégia de crescimento é sustentada por modelos de inteligência computacional que evoluem à medida que novos dados são inseridos no sistema [37].

3. Programa de Estágio

O Capítulo 3 constitui a componente central e operacional deste relatório, servindo como o registo detalhado da transposição dos conhecimentos académicos para o contexto profissional. Numa primeira instância, é traçado um panorama sobre a rotina diária e a dinâmica de trabalho no estágio (Secção 3.1), permitindo compreender o ecossistema e o fluxo operacional que sustentou o percurso na organização.

Posteriormente, são detalhadas as ferramentas e tecnologias utilizadas (Secção 3.2), evidenciando a forma como a infraestrutura de dados e os instrumentos técnicos foram orquestrados para responder a desafios complexos e garantir a integridade da informação. O capítulo é encerrado com a descrição das atividades desenvolvidas (Secção 3.3), onde a autonomia técnica e a capacidade de análise crítica foram aplicadas em cenários reais, demonstrando o impacto direto das soluções implementadas na dinâmica corporativa da *Uniplaces*.

3.1. Rotina Diária e Dinâmica de Trabalho no Estágio

A rotina diária no contexto do estágio desenvolvido na *Uniplaces* iniciava-se, de forma sistemática, às 09:00 horas, com a verificação dos principais canais de comunicação e organização de trabalho. Esta atividade incluía a análise de mensagens de correio eletrónico, a consulta de comunicações internas na plataforma *Slack* e a validação da agenda de reuniões no *Google Meet*. Este procedimento inicial é consistente com boas práticas de gestão de tarefas e comunicação em ambientes organizacionais digitais, permitindo uma priorização eficiente das atividades e uma rápida adaptação a eventuais mudanças no planeamento diário [38].

No início de cada semana, especificamente às segundas-feiras, realizava-se uma reunião de alinhamento com o departamento financeiro, com o objetivo de definir as prioridades e objetivos a atingir ao longo da semana. Este tipo de reunião enquadra-se nas metodologias ágeis de gestão de projetos, particularmente nos princípios do *Sprint Planning*, onde se estabelecem metas claras e se promove a coordenação entre equipas [39]. Complementarmente, às sextas-feiras, ocorria uma reunião de revisão semanal com o mesmo departamento, focada na análise dos resultados obtidos, identificação de constrangimentos e avaliação do progresso face aos objetivos definidos. Esta prática é análoga ao *Sprint Review*, promovendo uma cultura de melhoria contínua e transparência organizacional.

Adicionalmente, às terças-feiras durante o período da tarde, realizava-se uma reunião técnica envolvendo a equipa de *Business Intelligence* e o *Chief Financial Officer* (CFO). Estas sessões tinham como principal finalidade a discussão de tópicos técnicos, incluindo o desenvolvimento de novos projetos, resolução de problemas analíticos e validação de abordagens metodológicas. A interação direta com perfis de decisão estratégica, como o CFO, evidencia a importância crescente da função de *Data Analyst* na criação de valor organizacional, nomeadamente através da transformação de dados em *insights* acionáveis [4].

Para além das reuniões estruturadas, verificava-se a realização pontual de reuniões informais de curta duração (*fast calls*) com elementos de diferentes departamentos. Estas interações tinham como objetivo esclarecer dúvidas específicas ou alinhar rapidamente requisitos de negócio, sendo fundamentais para reduzir ambiguidades e acelerar o ciclo de desenvolvimento de soluções analíticas. Este tipo de comunicação síncrona e direta é apontado na literatura como um fator crítico para o sucesso de equipas multidisciplinares [40].

A partir do mês de janeiro, e na sequência de alterações na estrutura de liderança com a entrada de um novo CFO, foi introduzida uma reunião semanal adicional às quintas-feiras, dedicada à revisão de código (*Code Review*). Esta sessão era conduzida em conjunto com uma Sênior *Data Analyst* e tinha como objetivo avaliar a qualidade do código desenvolvido, identificar possíveis melhorias e garantir a adoção de boas práticas de programação. O processo de *Code Review* é amplamente reconhecido como uma prática essencial no desenvolvimento de software, contribuindo para a deteção precoce de erros, melhoria da legibilidade do código e partilha de conhecimento entre membros da equipa [41].

No que concerne às atividades operacionais, a semana de trabalho era maioritariamente dedicada à resolução de problemas provenientes de diferentes departamentos da organização, bem como ao desenvolvimento de novos projetos analíticos. Estas tarefas incluíam, entre outras, a extração, transformação e análise de dados, alinhando-se com o paradigma de *Data-Driven Decision Making* (DDDM), onde as decisões organizacionais são suportadas por evidência empírica e análise quantitativa [42].

Em síntese, a rotina de estágio caracterizava-se por uma forte componente colaborativa, estruturada em torno de reuniões regulares e interações interdepartamentais, complementada por períodos de trabalho autónomo focados no desenvolvimento de soluções analíticas. Esta dinâmica permitiu não só a aquisição de competências técnicas, mas também o

desenvolvimento de capacidades de comunicação, gestão de tempo e adaptação a contextos organizacionais dinâmicos.

3.2. Ferramentas e tecnologias utilizadas

Esta secção apresenta as principais ferramentas utilizadas ao longo do estágio, destacando o seu papel no desenvolvimento das atividades realizadas. Serão descritas as soluções que permitiram analisar, tratar e visualizar os dados, bem como apoiar a execução dos diferentes projetos desenvolvidos.

3.2.1. *Slack*

A eficácia de qualquer projeto de análise de dados depende diretamente da fluidez e da transparência na comunicação entre os diversos intervenientes da organização. A utilização de ferramentas de colaboração social em contextos tecnológicos não serve apenas para a troca de mensagens, mas atua como um repositório de conhecimento e um catalisador para a coordenação de tarefas complexas em equipas de engenharia e dados [43]. O *Slack* é definido como uma plataforma de comunicação baseada em canais, concebida para substituir o e-mail tradicional por um fluxo de trabalho mais dinâmico, permitindo a integração de notificações automáticas e a centralização de discussões técnicas num ambiente acessível e organizado.

No universo da Ciência de Dados, o *Slack* desempenha um papel que transcende a simples conversa, funcionando como uma ponte entre os silos técnicos e os decisores de negócio. A sua utilidade reflete-se na capacidade de criar canais dedicados a projetos específicos ou fluxos de dados, facilitando a partilha de *scripts*, a discussão de discrepâncias em *dashboards* e a receção de alertas sobre a saúde das bases de dados. A capacidade de uma organização se tornar verdadeiramente orientada por dados (*data-driven*) está intrinsecamente ligada à rapidez com que as informações circulam entre os analistas e os gestores, sendo o *Slack* uma ferramenta que potencia esta agilidade operacional [44].

Ao longo do estágio na *Uniplaces*, a utilização do *Slack* foi fundamental para a integração no departamento financeiro e para o suporte transversal prestado às restantes equipas (ver Figura 5). Através da plataforma, foi possível gerir o fluxo de pedidos de análise, participar em canais específicos da equipa de dados para a resolução conjunta de problemas e utilizar a funcionalidade de *huddles* (chamadas rápidas) para o esclarecimento imediato de dúvidas técnicas ou alinhamento de requisitos. Além disso, a ferramenta foi utilizada para o

acompanhamento de anúncios globais da empresa e eventos internos, permitindo uma visão clara das prioridades da organização e garantindo que o trabalho de análise de dados estivesse sempre contextualizado com as dinâmicas mais recentes do negócio.

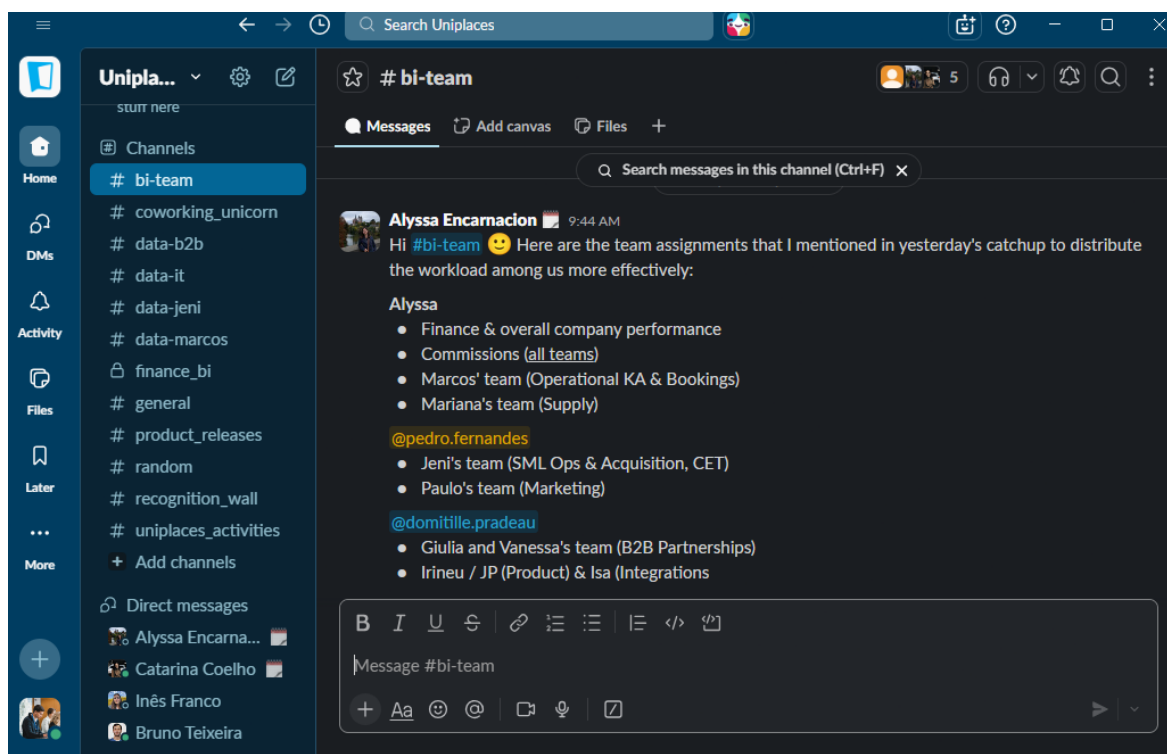


Figura 5 - Integração do Slack na Uniplaces

3.2.2. Google Sheets

As folhas de cálculo baseadas na nuvem transformaram a forma como as organizações colaboram e processam informação em tempo real, superando as limitações de isolamento dos ficheiros tradicionais. As folhas de cálculo continuam a ser uma das ferramentas de suporte à decisão mais ubíquas e poderosas, permitindo que utilizadores com diferentes níveis de literacia técnica possam manipular dados, realizar modelações financeiras e visualizar indicadores de desempenho de forma ágil [45]. No contexto da Ciência de Dados, o *Google Sheets* atua frequentemente como uma interface de "última geração", onde os dados extraídos de sistemas complexos são refinados e apresentados para consumo imediato pelos decisores estratégicos.

A utilidade desta tecnologia no mundo dos dados reside na sua flexibilidade para a prototipagem de relatórios e na facilidade de integração com outras fontes através de *scripts* e API. A capacidade de automatizar tarefas repetitivas e de estruturar lógicas de cálculo complexas dentro de uma grelha familiar permite que as empresas mantenham processos ágeis sem a necessidade imediata de desenvolver aplicações customizadas para cada necessidade

analítica [46]. Esta ferramenta é particularmente valiosa para a limpeza de dados e para a criação de painéis dinâmicos que exigem atualizações constantes e acesso colaborativo simultâneo.

Ao longo do estágio na *Uniplaces*, o domínio avançado do *Google Sheets* foi essencial para a manutenção e otimização de ferramentas críticas como o *dashboard* de Finanças e o sistema de acompanhamento de parceiros. A experiência adquirida permitiu o desenvolvimento de lógicas de filtragem e ordenação complexas para gerir volumes de dados superiores a 9.000 linhas, garantindo a performance e a usabilidade dos ficheiros. Entre as competências consolidadas, destaca-se a resolução de conflitos estruturais em tabelas dinâmicas, a criação de métricas agregadoras para quantificar volumes de trabalho pendente e a utilização das funções *Filter* e *Sort* para automatizar listas de tarefas inteligentes. Esta tecnologia revelou-se, assim, um pilar fundamental para garantir que os *insights* gerados fossem acessíveis e acionáveis por todos os departamentos da empresa.

3.2.3. Python

O *Python* é uma linguagem de programação de alto nível, interpretada e de uso geral, que se tornou a linguagem de referência no domínio da Ciência de Dados e da análise de dados [47]. A sua popularidade neste contexto deve-se, em grande medida, à simplicidade sintática, à versatilidade e ao vasto ecossistema de bibliotecas disponíveis. No âmbito do estágio, o *Python* foi utilizado como linguagem principal para o desenvolvimento de *scripts* de extração, transformação e análise de dados, nomeadamente através da *Trustpilot* API. Esta API disponibiliza um conjunto de *endpoints* que permitem aceder programaticamente às avaliações e metadados associados às páginas de empresas registadas na plataforma, possibilitando a extração automatizada de conteúdo textual para posterior processamento e análise. No decorrer do estágio, a versão do *Python* utilizada foi a 3.13 (versão mais recente) que possibilitou o desenvolvimento das tarefas propostas.

A biblioteca *Pandas* constitui um instrumento fundamental para a manipulação e análise de dados tabulares em *Python* [47]. Disponibiliza estruturas de dados de alto desempenho, como o *DataFrame*, que permitem realizar operações de filtragem, agregação, junção e transformação de dados de forma eficiente e intuitiva. A sua utilização foi transversal a praticamente todas as tarefas de tratamento de dados realizadas durante o estágio. A biblioteca *Requests* destina-se à realização de pedidos HTTP (*Hypertext Transfer Protocol*) em *Python*, sendo amplamente utilizada na comunicação com API externas [48], tendo sido empregue na interação com a

Trustpilot API para a extração dos comentários pretendidos. O módulo *Re*, nativo do *Python* é dedicado ao processamento de expressões regulares, foi utilizado na identificação e extração de padrões específicos em cadeias de texto, revelando-se particularmente útil na deteção de menções nominais nos comentários textuais extraídos [49].

3.2.4. *Looker Studio*

A visualização de dados é uma etapa crítica no processo de Ciência de Dados, funcionando como o elo de ligação entre a análise técnica e a interpretação estratégica por parte dos decisores. A capacidade de transformar dados complexos em representações visuais eficazes não é apenas uma competência técnica, mas uma forma de narrativa que permite destacar o que é realmente importante num vasto oceano de informação [50]. O *Looker Studio* é uma ferramenta de BI baseada na nuvem que permite a criação de painéis de controlo e relatórios dinâmicos, facilitando a monitorização de indicadores de desempenho em tempo real através de uma interface intuitiva e colaborativa.

A utilidade desta tecnologia no ecossistema de dados reside na sua integração nativa com o ambiente *Google* e na sua capacidade de consolidar fontes de dados heterogéneas num único ponto de consulta. A implementação de *dashboards* inteligentes é essencial para que as organizações possam passar de uma postura reativa para uma cultura proativa, onde as métricas de performance são visíveis e compreensíveis para todos os colaboradores, independentemente do seu nível técnico [51]. Esta ferramenta permite a automatização do *reporting*, garantindo que os dados brutos, após o tratamento em SQL ou *Python*, cheguem aos gestores de forma visual e acionável, reduzindo o tempo despendido em tarefas manuais de atualização.

No decorrer do estágio, a utilização do *Looker Studio* foi fundamental para converter o trabalho de processamento de dados em ferramentas visuais de suporte à decisão para as diversas equipas da *Uniplaces*. A aprendizagem focou-se na conexão de diversas fontes de dados, nomeadamente tabelas de *Google Sheets* e resultados de interrogações SQL, para construir *dashboards* que monitorizassem o sucesso de parcerias e a performance operacional. Através desta ferramenta, foi possível implementar filtros interativos e visualizações de séries temporais que permitiram aos diversos departamentos analisar tendências e identificar desvios de forma autónoma. Esta experiência consolidou a capacidade de traduzir requisitos de negócio em métricas visuais, reforçando a transparência e a eficiência na comunicação de resultados dentro da organização.

3.2.5. Tableau

O *Tableau* é amplamente reconhecido como uma das plataformas líderes em *Visual Analytics*, permitindo a exploração de dados através de uma interface altamente intuitiva e baseada no princípio de arrastar e largar. Esta ferramenta diferencia-se pela sua capacidade de permitir aos utilizadores realizar uma análise exploratória de dados (EDA) de forma fluida, facilitando a descoberta de padrões e tendências que poderiam permanecer ocultos em formatos tabulares tradicionais [52]. Ao contrário de outras ferramentas de reporte estático, o *Tableau* foca-se na interação direta com os dados, permitindo que o utilizador "questiona" a informação em tempo real e obtenha respostas visuais imediatas.

No domínio da Ciência de Dados, a utilidade do *Tableau* reside na sua elevada capacidade de processamento de grandes volumes de informação e na sua versatilidade para se ligar a múltiplas fontes de dados, desde simples ficheiros locais até bases de dados relacionais complexas. O sucesso de uma solução de BI depende da sua capacidade de reduzir a carga cognitiva do utilizador, apresentando a informação de forma clara e rigorosa [53]. O *Tableau* cumpre este propósito ao oferecer uma vasta gama de visualizações avançadas e cálculos analíticos integrados, permitindo que equipas de diversos departamentos possam monitorizar métricas complexas de forma simples e integrada, promovendo uma cultura de agilidade na resposta aos desafios de mercado.

Durante o estágio na *Uniplaces*, o contacto com o *Tableau* foi fundamental para elevar o nível de sofisticação dos relatórios apresentados. A aprendizagem centrou-se no desenvolvimento de painéis de controlo detalhados, onde foi possível integrar dados provenientes de diversas áreas de negócio para criar uma visão holística da performance da empresa. A ferramenta foi utilizada para construir visualizações que permitissem cruzar informações de reservas com métricas financeiras, facilitando a identificação de correlações e a monitorização de KPI críticos de forma dinâmica. Esta experiência permitiu consolidar competências em *Data Visualization* e perceber como a escolha correta de elementos visuais pode transformar a interpretação de dados complexos em decisões estratégicas fundamentadas.

3.2.6. MySQL

O *MySQL* é um sistema de gestão de bases de dados relacionais (RDBMS – *Relational Database Management System*) de código aberto que utiliza a linguagem SQL para a

manipulação e consulta de dados. A importância dos sistemas relacionais reside na sua capacidade de organizar vastos volumes de informação em estruturas tabulares ligadas por chaves lógicas, garantindo a integridade referencial e a consistência dos dados através das propriedades: Atomicidade, Consistência, Isolamento e Durabilidade (ACID) [54]. No mundo da Ciência de Dados, o domínio do SQL é considerado uma competência fundamental, pois constitui a linguagem primária de comunicação com os repositórios de dados corporativos, permitindo transformar repositórios brutos em conjuntos de dados estruturados para análise.

A utilidade desta tecnologia transcende a mera armazenagem, sendo o motor que viabiliza a escalabilidade e a rapidez no acesso à informação em organizações de base tecnológica. A eficácia de uma base de dados relacional é medida pela sua capacidade de executar interrogações complexas (*queries*) que cruzam múltiplas dimensões de negócio de forma performante [55]. Para um analista de dados, o *MySQL* funciona como a "fonte única de verdade", onde a extração rigorosa de informação é o primeiro passo essencial de qualquer fluxo de trabalho, precedendo todas as fases de tratamento em *Python* ou de visualização em plataformas de BI.

Para a *Uniplaces*, o *MySQL* representa o coração informacional da empresa, onde está registado todo o ciclo de vida da operação, desde os detalhes das propriedades e o histórico de reservas até às transações financeiras e interações dos utilizadores (ver Figura 6). A importância desta ferramenta no contexto organizacional é absoluta, dado que a precisão das decisões estratégicas tomadas pelas equipas de Finanças e Operações depende inteiramente da correção das consultas efetuadas a este sistema. Sem a robustez do *MySQL*, não seria possível garantir a fiabilidade dos indicadores apresentados nos diversos *dashboards* da empresa, tornando esta tecnologia o elo indispensável que sustenta a cultura orientada por dados da organização.

O contato intensivo com o *MySQL* durante o estágio permitiu consolidar uma compreensão profunda da arquitetura de dados da *Uniplaces* e da complexidade das suas relações internas. A aprendizagem focou-se no desenvolvimento de consultas otimizadas para extrair métricas de performance e na manutenção da integridade dos dados face a alterações operacionais, como a redistribuição de portfólios ou a auditoria de ativos. Esta experiência demonstrou que a capacidade de navegar com fluidez em bases de dados relacionais é o que permite a um analista garantir que a informação que chega aos decisores é não só atualizada, mas tecnicamente irrepreensível, reforçando o papel do *MySQL* como o pilar mais influente no sucesso das atividades desenvolvidas.

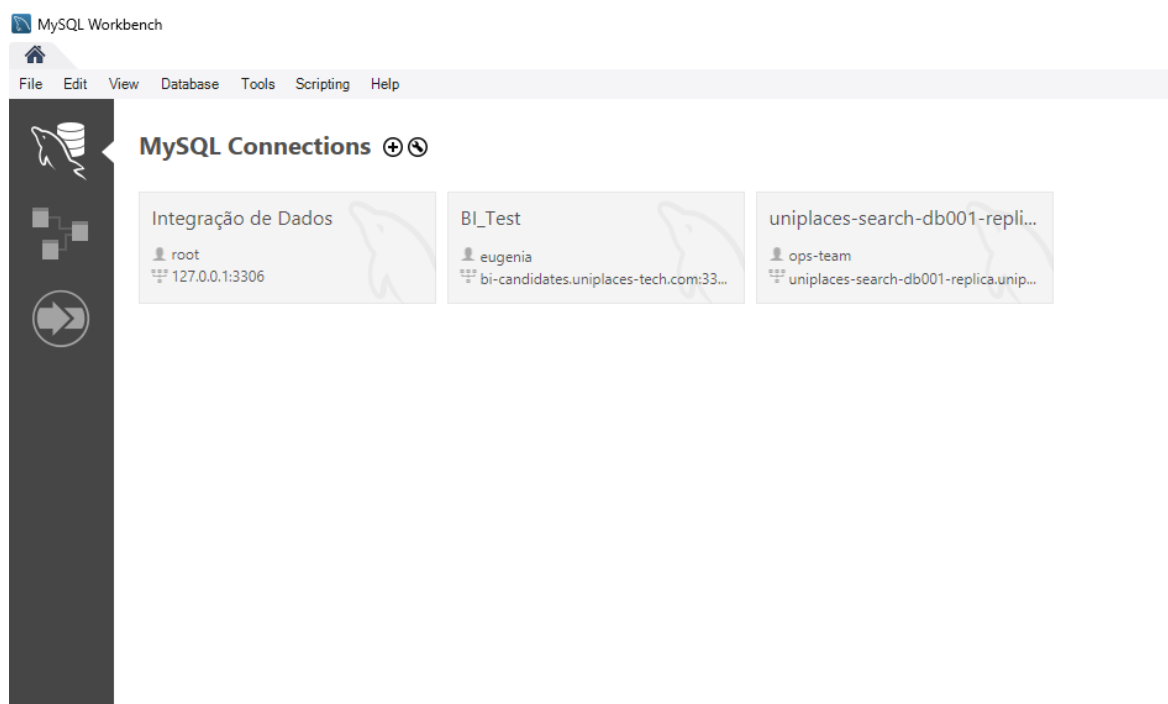


Figura 6 - Interface Inicial do MySQL

3.2.7. Confluence

A gestão do conhecimento e a documentação técnica são pilares fundamentais para a continuidade e eficiência de qualquer equipa de Ciência de Dados, especialmente em ambientes onde a colaboração interdepartamental é intensa. A introdução de plataformas de "Enterprise 2.0", como o *Confluence*, revolucionou a forma como as organizações capturam e partilham conhecimento, substituindo silos de informação e cadeias de e-mail por plataformas colaborativas que promovem a transparência e a inteligência coletiva [56]. O *Confluence* atua como um sistema de gestão de conhecimento (*Knowledge Management System*), permitindo que processos complexos, requisitos de negócio e especificações técnicas sejam documentados de forma centralizada e acessível a todos os intervenientes.

A utilidade desta tecnologia no domínio dos dados reside na sua capacidade de atuar como uma interface de governação e uma ponte de comunicação entre equipas técnicas e equipas de negócio. O conhecimento organizacional é um ativo valioso que deve ser estruturado para evitar a perda de contexto crítico durante o ciclo de vida dos projetos [57]. No contexto de BI, o *Confluence* permite documentar as definições de KPI, as origens de dados e as lógicas de cálculo, garantindo que qualquer membro da organização consiga interpretar os relatórios produzidos com o mesmo rigor. Além disso, a plataforma facilita a gestão de fluxos de trabalho através da criação de pedidos formais (*requests*), assegurando que o desenvolvimento de novas análises esteja alinhado com as prioridades estratégicas da empresa (ver Figura 7).

[BI Request] 2025-09-26 - Leads per Offer received through Idealista PT

Por Pedro Matos 3 Adicionar reação

Requester	@Pedro Matos
Team	fin
Urgency	High
Status	To be filled by BI
Project Responsible	To be filled by BI
Time needed	To be filled by BI
Description	@Moe bawab @pedro.fernandes Please provide the number of leads received though Idealista PT per offer ID listing, per month with city, neighbourhood, offer type, monthly rent, number of pics, bills included (Y/N).

Figura 7 - Request feito no Confluence

Durante o estágio na *Uniplaces*, o *Confluence* foi a plataforma central para a gestão de pedidos provenientes da equipa B2B (*Business-to-Business*), funcionando como o canal oficial de comunicação entre departamentos. A utilização da ferramenta permitiu estruturar cada solicitação de análise ou automação de forma rigorosa, definindo parâmetros críticos como o identificador do requerente, a equipa de origem, o grau de urgência, o responsável pelo desenvolvimento e o tempo estimado de execução. Esta estrutura possibilitou monitorizar em tempo real a fase de cada pedido, desde o levantamento de requisitos até à entrega final, garantindo um acompanhamento transparente do progresso. A plataforma serviu ainda como repositório central para documentar o *feedback* das equipas e arquivar ficheiros de suporte, revelando-se indispensável para a organização do trabalho e para a manutenção de um histórico detalhado de todas as intervenções técnicas realizadas.

3.3. Atividades desenvolvidas

A presente secção é dedicada à descrição e análise das atividades desenvolvidas ao longo do estágio curricular realizado na *Uniplaces*. As tarefas aqui documentadas refletem o trabalho efetivamente realizado no seio da equipa de dados, abrangendo um conjunto diversificado de projetos que contemplam desde a análise e extração de dados até ao desenvolvimento de modelos preditivos e à construção de *dashboards* interativos.

Na secção 3.3.1 é descrito o processo de integração inicial na organização, com particular enfoque no mapeamento da infraestrutura de dados existente e na familiarização com os sistemas e ferramentas utilizados pela equipa. A secção 3.3.2 documenta o trabalho de otimização e automatização do *reporting* de *marketing*, onde foram desenvolvidas soluções

para agilizar a produção e atualização de relatórios periódicos. A secção 3.3.3 descreve a construção de um relatório de desempenho associado à ferramenta interna *Talk to the Landlord*, desenvolvido com o objetivo de avaliar o impacto desta funcionalidade junto da equipa de gestão. Na secção 3.3.4 é apresentado o relatório de reservas de último minuto, orientado para a identificação e monitorização de reservas com janelas temporais críticas entre o pedido e a data de entrada no alojamento.

A secção 3.3.5 aborda o processo de migração de *dashboards* B2B para o *Looker Studio*, envolvendo a transposição e reestruturação de relatórios previamente desenvolvidos noutras plataformas. A secção 3.3.6 documenta uma auditoria ao processo de atribuição de portefólio no âmbito da gestão de contas, com vista à identificação e correção de irregularidades no sistema de alocação de clientes. Na secção 3.3.7 é descrito o desenvolvimento de um sistema automatizado de extração de avaliações do *Trustpilot* em que os agentes da *Uniplaces* eram explicitamente mencionados, com recurso à API da plataforma e ao *Google Apps Script*. A secção 3.3.8 apresenta o desenvolvimento de um modelo de previsão de preços, onde foram aplicadas técnicas de aprendizagem automática para estimar o valor de arrendamento de ofertas com base nas suas características. Na secção 3.3.9 é descrita a reestruturação e automatização da matriz de análise de cancelamentos em *Tableau*, com o objetivo de melhorar a legibilidade e a atualização automática deste instrumento de análise operacional. Por fim, a secção 3.3.10 documenta o projeto de maior abrangência analítica do estágio, que consistiu na análise dos fatores de conversão das ofertas publicadas na plataforma e no desenvolvimento de um modelo preditivo de sucesso baseado nas características de cada anúncio.

3.3.1. Semana 1: Integração e Mapeamento da Infraestrutura de Dados

A fase inicial do estágio foi dedicada à aculturação tecnológica e à compreensão profunda do ecossistema de dados da *Uniplaces*. Esta etapa revelou-se fundamental para estabelecer a base cognitiva necessária para a execução das tarefas subsequentes, garantindo que a manipulação técnica dos dados estivesse sempre contextualizada com as regras de negócio da organização. O foco central desta semana residiu na exploração do esquema relacional das bases de dados através do *MySQL*, onde foi possível identificar a segmentação da informação em diferentes ambientes. O trabalho foi predominantemente desenvolvido nos esquemas *prod_reporting*, que armazena os dados transacionais e de inventário consolidados, e *prod_reporting_ops*, focado na vertente operacional e financeira, permitindo uma análise granular da performance e rentabilidade do negócio.

Através da análise do diagrama físico da base de dados (ver Figura 8), foi possível mapear as tabelas *core* que sustentam a operação do *marketplace* e compreender a hierarquia do inventário disponível. As entidades *accommodation_provider*, *property* e *accommodation_offer* formam a base logística da plataforma. A primeira armazena informações críticas sobre os senhorios e parceiros, incluindo a gestão de conta e o seu estado de atividade. A tabela *property* detalha as especificidades físicas e geográficas dos imóveis, enquanto a *accommodation_offer* descreve as condições comerciais e a tipologia de cada oferta publicada. A relação entre estas tabelas é crucial, pois permite que uma única propriedade possa conter múltiplas ofertas, como quartos individuais num apartamento partilhado, refletindo a complexidade da oferta habitacional da *Uniplaces*.

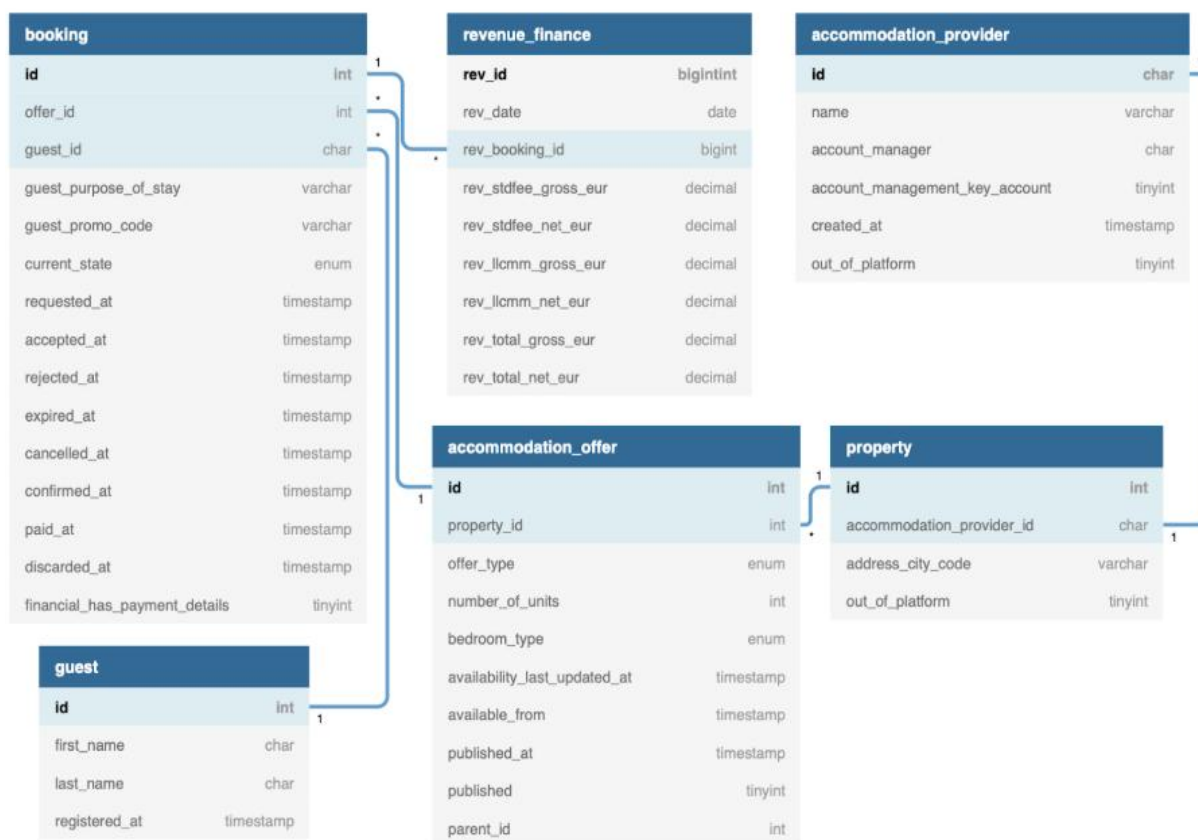


Figura 8 - Principais tabelas de dados da *Uniplaces*

No que concerne à vertente transacional, a tabela *booking* assume o papel de entidade nuclear, registrando todo o ciclo de vida de uma reserva, desde os identificadores do inquilino e da oferta até aos carimbos temporais de confirmação ou cancelamento. Esta informação é complementada pela tabela *guest*, que armazena os perfis dos inquilinos, e pela tabela *revenue_finance*, situada no esquema de operações. Esta última é vital para a saúde financeira da empresa, pois é nela que se observa o lucro efetivo de cada transação, discriminando valores

brutos e líquidos em euros e as respetivas datas de liquidação. A integração destes dados permite cruzar o volume de reservas com a margem financeira real gerada em cada período.

A compreensão da integridade referencial entre estas entidades, visível através das ligações de chaves externas no diagrama, foi determinante para o sucesso das tarefas de extração de dados. O triângulo relacional entre as reservas, os parceiros e as ofertas constitui a espinha dorsal das operações da organização. Foi nesta interseção que se concentrou a exploração inicial, uma vez que a capacidade de interrogar estas bases de dados de forma otimizada permite monitorizar a eficiência da conversão entre a oferta disponível e a procura dos inquilinos, garantindo que os indicadores apresentados nos *dashboards* futuros fossem tecnicamente irrepreensíveis e alinhados com as prioridades estratégicas da *Uniplaces*. A qualidade e a credibilidade dos dados que sustentam estes indicadores constituem, aliás, um fator crítico para a fiabilidade de qualquer processo analítico e de suporte à decisão organizacional [58].

Para complementar a análise técnica da infraestrutura de dados, revelou-se igualmente essencial a compreensão do ciclo de vida das reservas (*Booking Request Lifecycle*), conforme ilustrado na Figura 9. Este fluxo operacional é o que dita a lógica de transição de estados na base de dados, permitindo a um analista interpretar corretamente os carimbos temporais e os indicadores de performance. De acordo com os princípios do processo de descoberta de conhecimento em bases de dados, a correta interpretação dos estados e transições dos dados constitui uma etapa prévia indispensável a qualquer tarefa de extração e análise [59]. O processo inicia-se quando um potencial inquilino submete um pedido de reserva, entrando no estado inicial de *Requested*. Nesta fase, existe uma janela temporal crítica de 24 horas para que o utilizador submeta os detalhes de pagamento ou associe um método de liquidação válido. Caso este requisito não seja cumprido dentro do prazo estipulado, o sistema descarta automaticamente o pedido, que transita para o estado de *Discarded*, interrompendo o fluxo comercial antes mesmo da interação com o proprietário.

Uma vez validados os detalhes de pagamento, o pedido evolui para o estado de *Awaiting*, momento em que a responsabilidade de progressão recai sobre o senhorio. Durante este período de espera, o pedido pode seguir três trajetórias distintas baseadas na resposta do parceiro: se o senhorio não responder dentro do tempo de aceitação configurado, o pedido transita para *Expired*; se houver uma recusa explícita, o estado passa a *Rejected*; e, caso o senhorio aceite as condições propostas, a reserva avança para o estado de *Accepted*. Esta fase intermédia de aceitação é particularmente complexa, pois ainda está sujeita a variáveis de comportamento do

utilizador e a falhas técnicas de transação. A modelação de fluxos de estados desta natureza, onde cada transição é condicionada por múltiplos fatores externos e internos, é reconhecida na literatura como um dos desafios centrais da engenharia de dados aplicada a plataformas de *marketplace* [60]. Se o inquilino rejeitar ou ignorar eventuais sugestões de datas alternativas propostas pelo senhorio, ou se ocorrer uma falha crítica no processamento do pagamento, a reserva retrocede para estados de descarte ou expiração de pagamento (*Payment Expired*).

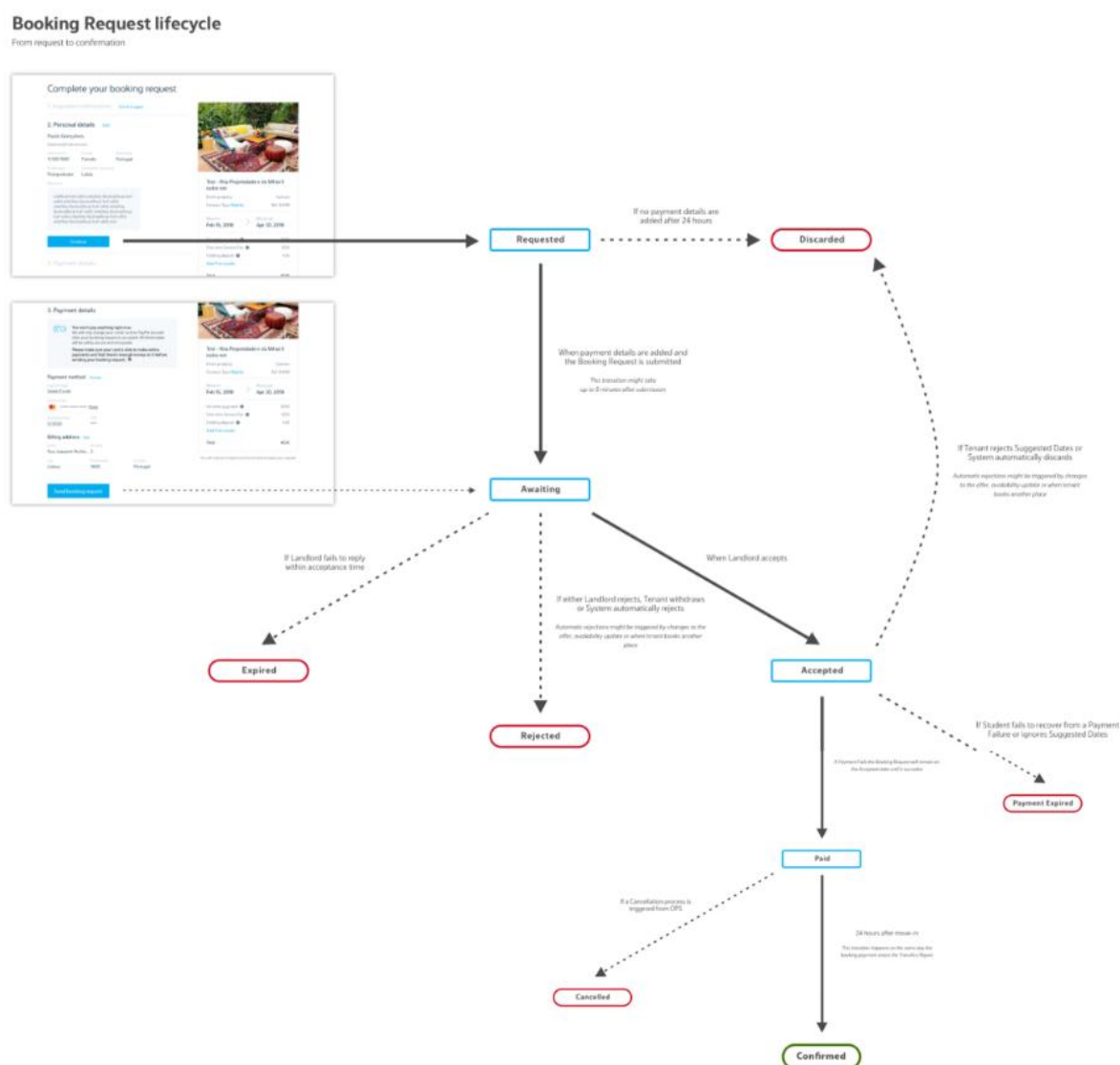


Figura 9 - Ciclo de vida de um Booking Request

A transição para a fase final do ciclo de vida ocorre quando o pagamento é processado com sucesso, colocando a reserva no estado de *Paid*. A partir deste ponto, o acompanhamento operacional é assegurado através da plataforma OPS (*Operations System*), a ferramenta interna utilizada pela *Uniplaces* para a gestão integrada de *bookings* e ofertas. É nesta etapa que podem ocorrer cancelamentos manuais ou automáticos (*Cancelled*) caso surjam inconformidades

logísticas ou problemas reportados pela equipa de operações. Se o processo decorrer sem incidentes até 24 horas após a data de entrada (*move-in*), a reserva atinge o seu estado definitivo de *Confirmed*. Para a Ciência de Dados, este mapeamento é vital, pois permite analisar o funil de conversão da plataforma, identificar gargalos operacionais e garantir que a extração de dados no *MySQL* reflita fielmente a realidade do negócio em cada etapa do processo. A análise de funis de conversão e a identificação de pontos de fricção ao longo do ciclo de vida transacional constituem, aliás, uma das aplicações mais relevantes da Ciência de Dados no contexto de plataformas digitais de dois lados, onde a eficiência do processo de correspondência entre oferta e procura determina diretamente a performance do negócio [61].

Por fim, a conclusão desta etapa inicial envolveu a aprendizagem e consolidação dos principais indicadores de desempenho (KPI) que regem a análise de performance da *Uniplaces*. Estes indicadores são fundamentais para traduzir o ciclo de vida das reservas em métricas quantificáveis, permitindo avaliar a saúde do *marketplace* sob diferentes perspetivas. Grande parte destes rácios incide sobre o funil de conversão inicial, tendo como denominador comum o volume total de pedidos realizados. Neste âmbito, o *Acceptance Rate* assume um papel preponderante ao medir a receptividade da oferta, sendo calculado através da divisão entre as reservas aceites e o total de pedidos. Complementarmente, o *Rejection Rate* e o *Expired Rate* permitem diagnosticar perdas de eficiência nesta fase; enquanto o primeiro reflete a percentagem de recusas explícitas por parte dos senhorios, o segundo quantifica as reservas que expiram por inação, servindo ambos para identificar fricções ou desajustes entre as expectativas dos inquilinos e a disponibilidade real dos parceiros.

No que concerne à conversão final e à rentabilidade do negócio, são utilizados indicadores que monitorizam a transição para o estado de pagamento e a efetivação da reserva. O *Paid Rate* representa a métrica global de conversão, calculada pela razão entre as reservas pagas e o total de pedidos submetidos. Já o *Success Rate* introduz uma camada de análise focada no comportamento do utilizador, ao relacionar o número de reservas pagas com o total de pedidos efetuados por cada inquilino único (*unique guest*), permitindo compreender a eficácia da plataforma em converter um utilizador interessado num cliente efetivo. Por último, o *Cancellation Rate* atua como um indicador de risco e qualidade pós-transação, sendo determinado pela divisão entre as reservas canceladas e o total de reservas pagas. Esta métrica é crítica para a equipa de operações, pois permite monitorizar o *churn* após a fase de pagamento e identificar eventuais falhas logísticas ou insatisfações que ocorram antes do período de

entrada no imóvel. O domínio destes conceitos permitiu assegurar que todas as análises subsequentes estivessem alinhadas com a linguagem estratégica da organização.

3.3.2. Otimização e Automatização do *Reporting* de Marketing

No âmbito da gestão financeira e do controlo de performance, o "Online Marketing Target" constitui um instrumento de reporte estratégico que permite monitorizar o desempenho dos diversos canais de aquisição face aos objetivos orçamentais definidos pela organização. Esta ferramenta é essencial para avaliar a eficiência de rubricas como o *Paid Search* (incluindo campanhas de *Bespoke* e *Remarketing*), *Classified* e as comissões de B2B, cruzando o volume de faturação (*Revenue*) com os respetivos custos de aquisição (*Acquisition Cost*). A modelação de dados em ambientes de folha de cálculo deve privilegiar a integridade e a flexibilidade, permitindo que os decisores identifiquem desvios de performance em tempo real para ajustar as estratégias de investimento de forma ágil e fundamentada [62].

Antes da intervenção realizada, a estrutura de reporte disponível baseava-se numa lógica estática de *Year To Date* (YTD), conforme ilustrado na Figura 10. Esta abordagem apresentava limitações operacionais significativas, uma vez que apenas permitia visualizar os valores acumulados até ao período corrente, dificultando a análise de sazonalidade e a extração de *insights* sobre meses específicos sem a necessidade de reestruturação manual das fórmulas. A dependência de atualizações manuais para cada novo período de reporte não só aumentava a carga de trabalho administrativo, como elevava o risco de erro humano na manipulação de grandes volumes de dados financeiros, comprometendo a fiabilidade da "fonte única de verdade" da empresa.

Para mitigar estas limitações, foi desenvolvida uma solução de automatização que transformou o documento num *dashboard* dinâmico e interativo, visível na Figura 11. A implementação técnica baseou-se na criação de seletores de dados através de validação de dados, permitindo ao utilizador filtrar instantaneamente o ano e o mês pretendidos. A lógica de cálculo foi integralmente reestruturada utilizando funções de agregação dinâmica e interrogações que comunicam com as bases de dados centralizadas. Esta alteração permitiu que as métricas de *Actuals*, *Budget* e *Prior Year* (PY), bem como as respetivas variações absolutas e percentuais (YoY - *Year over Year*), passassem a ser processadas de forma automática em função da seleção temporal, garantindo uma visualização imediata da performance mensal e acumulada.

	Actuals	Budget	Dev.	Dev. (%)	PY	Var.	YoY
Revenue	€4,556,784	€5,185,317	-€628,533	-12%	€4,322,425	€234,359	5.4%
Marketing	€3,951,203	€4,709,751	-€758,548	-16%	€3,865,293	€85,910	2%
B2B	€605,581	€475,566	€130,015	27%	€457,132	€148,449	32%
Acquisition Cost	€1,587,283	€1,416,364	€170,919	12%	€1,289,868	€297,415	23%
Paid Search (incl. Bespoke, Remkt)	€1,119,460	€995,581	€123,879	12%	€857,334	€262,125	31%
Other	€75,216	€80,000	-€4,784	-6%	€85,708	-€10,492	-12%
Classified	€218,042	€200,000	€18,042	9%	€256,386	-€38,344	-15%
B2B Commissions	€174,565	€140,783	€33,782	24%	€90,439	€84,126	93%
Gross Margin	€2,969,501	€3,768,953	-€799,452	-21%	€3,032,557	-€63,056	-2%
Total Marketing Cost (%)	34.8%	27.3%	7.5 p.p.	7.5 p.p.	29.8%	5.0 p.p.	5.0 p.p.
Total Marketing excl B2B Comm (%)	31.0%	24.6%	6.4 p.p.	6.4 p.p.	27.7%	3.3 p.p.	3.3 p.p.

Figura 10 - Dashboard antes da automatização

	Actuals	Budget	Dev.	Dev. (%)	PY	Var.	YoY
Revenue	€460,702	€455,000	€5,702	1%	€428,170	€32,532	7.6%
Marketing	€405,135	€404,731	€404	0%	€386,242	€18,893	5%
B2B	€55,567	€50,269	€5,298	11%	€41,928	€13,639	33%
Acquisition Cost	€101,745	€142,468	-€40,723	-29%	€89,341	€12,404	14%
Paid Search (incl. Bespoke, Remkt)	€58,610	€87,360	-€28,750	-33%	€46,106	€12,504	27%
Other	€10,092	€10,000	€92	1%	€4,203	€5,889	140%
Classified	€22,620	€25,000	-€2,380	-10%	€30,256	-€7,636	-25%
B2B Commissions	€10,423	€20,108	-€9,685	-48%	€8,776	€1,647	19%
Gross Margin	€358,957	€312,532	€46,425	15%	€338,830	€20,128	6%
Total Marketing Cost (%)	22.1%	31.3%	-9.2 p.p.	-9.2 p.p.	20.9%	1.2 p.p.	1.2 p.p.
Total Marketing excl B2B Comm (%)	19.8%	26.9%	-7.1 p.p.	-7.1 p.p.	18.8%	1.0 p.p.	1.0 p.p.

Figura 11 - Dashboard depois da automatização

A transição de um modelo estático para uma ferramenta de *reporting* automatizada resultou num incremento substancial na eficiência operacional das equipas de Finanças e Marketing. A automatização de relatórios recorrentes em plataformas de BI é um passo crítico para a maturidade analítica de uma organização, pois liberta os analistas para tarefas de interpretação estratégica em detrimento da mera compilação de dados [63]. Esta otimização permitiu à *Uniplaces* monitorizar os alvos de *marketing* com uma precisão temporal superior, facilitando a identificação imediata de desvios orçamentais e potenciando uma tomada de decisão baseada em dados rigorosos e atualizados.

Para viabilizar a automatização do ficheiro "Online Marketing Target" descrita anteriormente, foi necessário desenvolver uma camada de integração técnica que permitisse a comunicação direta entre a folha de cálculo e os repositórios de dados da *Uniplaces*. Esta

solução foi implementada através do *Google Apps Script*, uma plataforma de *scripting* baseada em *cloud* que permite estender as funcionalidades das aplicações *Google Workspace* através de uma linguagem baseada em *JavaScript*. A utilização de ferramentas de *scripting* em ambiente corporativo é fundamental para criar fluxos de trabalho personalizados que eliminam a necessidade de intervenções manuais, garantindo que a informação flua de forma íntegra entre sistemas complexos e interfaces de utilizador simples [64].

A componente técnica desta implementação pode ser observada no ambiente de desenvolvimento do projeto (ver Figura 12), onde foi estruturado o código para a extração automática de rendimentos. O *script* foi desenhado para executar uma interrogação SQL específica na tabela *revenue_finance* do esquema *prod_reporting_ops*. A lógica da consulta foca-se na agregação do rendimento líquido total (*rev_total_net_eur*), devidamente arredondado e agrupado por ano e mês para o período de 2025. A utilização de comandos de agregação e ordenação em SQL é o que permite transformar registos transacionais brutos em métricas financeiras consolidadas, prontas para serem consumidas por ferramentas de BI ou folhas de cálculo estratégicas [65].

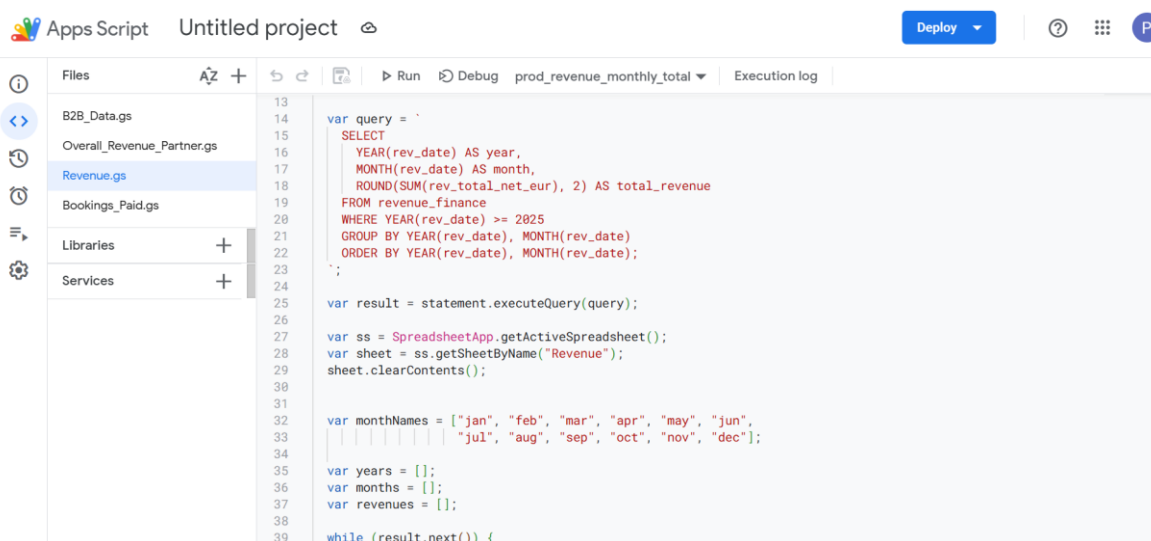


Figura 12 - Script da extração dos dados no Apps Script

O impacto direto desta automação reflete-se na atualização dinâmica da folha de cálculo (ver Figura 13), onde os dados de *Revenue* e *Acquisition Cost* são injetados automaticamente nas respetivas colunas mensais. O código inclui funções críticas de gestão de dados, como o *clearContents()*, que assegura que as células são limpas antes de cada nova extração, evitando a persistência de valores obsoletos ou erros de cálculo. Esta abordagem garantiu que o *dashboard* de Marketing deixasse de ser um documento estático dependente de inserções

manuais para se tornar uma "fonte única de verdade" dinâmica, permitindo à equipa financeira monitorizar as margens de lucro e os gastos por canal de forma instantânea e rigorosa.

	2025	2025	2025	2025	2025	2025	2025	2025	2025	2025	2025	2025
Uniplaces	ACT	ACT	ACT	ACT	ACT	ACT	ACT	ACT	ACT	ACT	ACT	ACT
	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
Revenue	€460,702	€308,437	€434,451	€448,149	€601,149	€666,499	€883,636	€753,761	€447,413	€424,964	€458,268	€454,351
Marketing	€405,135	€275,020	€396,885	€389,924	€510,386	€582,739	€751,044	€640,070	€447,413	€424,964	€458,268	€454,351
B2B	€55,567	€33,417	€37,566	€58,225	€90,763	€83,760	€132,592	€113,691	€0	€0	€0	€0
Acquisition Cost	€101,745	€111,465	€140,434	€128,583	€206,080	€335,876	€279,931	€283,169	€133,343	€138,265	€146,104	€145,182
Paid Search (incl. Social, PMax)	€58,610	€76,513	€100,931	€82,772	€148,281	€229,698	€207,990	€214,664	€95,106	€100,028	€107,867	€106,945
Agencies												
Classified	€22,620	€17,740	€20,334	€25,026	€29,565	€31,952	€38,335	€32,471	€32,471	€32,471	€32,471	€32,471
SEO	€4,833	€4,833	€4,833	€4,833	€4,833	€0	€0	€4,833	€0	€0	€0	€0
Attribution	€5,000	€5,000	€5,000	€5,000	€5,000	€5,000	€5,000	€5,000	€5,000	€5,000	€5,000	€5,000
Other	€258	€1,308	€1,195	€1,004	€1,219	€208	€258	€765	€765	€765	€765	€765
B2B Commissions	€10,423	€6,071	€8,141	€9,947	€17,182	€69,018	€28,348	€25,435	€0	€0	€0	€0
Gross Margin	€358,957	€196,972	€294,017	€319,566	€395,069	€330,623	€603,705	€470,592	€314,071	€286,699	€312,164	€309,169
Total Marketing Cost (%)	22.1%	36.1%	32.3%	28.7%	34.3%	50.4%	31.7%	37.6%	29.8%	32.5%	31.9%	32.0%
Total Marketing excl B2B Comm (%)	19.8%	34.2%	30.5%	26.5%	31.4%	40.0%	28.5%	34.2%	29.8%	32.5%	31.9%	32.0%

Figura 13 - Dados referentes à *Revenue* e *Acquisition Cost* da *Uniplaces*

Esta intervenção técnica no ficheiro de alvos de *marketing* demonstra a aplicação prática dos princípios de engenharia de dados no suporte à gestão. A maturidade analítica de uma organização é atingida quando a latência dos dados é reduzida através da automação, permitindo que o foco dos analistas mude da compilação de dados para a interpretação de resultados [44]. Ao centralizar a extração do rendimento via *Apps Script*, foi possível assegurar um sistema de suporte à decisão resiliente, onde a precisão dos dados financeiros é garantida por processos automatizados, reforçando o controlo orçamental da *Uniplaces*.

3.3.3. Relatório de Desempenho da Ferramenta "Talk to the Landlord"

A primeira tarefa de maior relevo desenvolvida no âmbito do estágio curricular consistiu na criação de um relatório de desempenho associado a uma nova funcionalidade lançada pela *Uniplaces*, denominada *Talk to the Landlord* (ver Figura 14). Esta ferramenta, desenvolvida pela equipa de *Product Engineering & Design* (PED) em conjunto com a equipa de Operações, tem como propósito facilitar a comunicação direta entre inquilinos e proprietários, permitindo que questões, informações ou problemas relativos a um alojamento sejam tratados de forma mais ágil e centralizada dentro da própria plataforma.

À data em que a tarefa foi solicitada, a funcionalidade havia sido lançada há apenas dez dias, tendo registado aproximadamente 160 conversações. Dado o carácter estratégico da iniciativa, foi considerado prioritário (com urgência classificada como *High* no sistema interno de gestão de pedidos de BI) produzir um relatório que permitisse analisar o impacto da ferramenta e apresentar os resultados de desempenho diretamente ao CEO da *Uniplaces*, Cyril Jessua. O pedido foi formalizado por Marcos Garcia Albarrán, *Head of Key Account*

Management, contando ainda com o envolvimento de Renato Duarte, João Pereira e Florin Frunza, todos pertencentes às equipas de PED e Operações.

[BI Request] 2025-10-13 - Talk to the Landlord

Por Marcos Garcia Albarrán

6

Adicionar reação

Requester	@Marcos Garcia Albarrán @Renato Duarte @João Pereira @Florin Frunza
Team	PED, Operations
Urgency	High
Status	To be filled by BI
Project Responsible	To be filled by BI
Time needed	To be filled by BI
Description	Hello team, PED launched a new tool (Talk to the Landlord) 10 days ago an we already received +160 conversations. We need to create a report to analyse the impact and show performance to Cyril. You can find ALL info in MySQL > prod_reporting > conversation 1. Conversations received a. Conversations dealt by APs > X% b. Conversations dealt by AMs > Y% 2. Messages received a. Mark as safe i. by agent > X% ii. by AI > X% b. Flagged i. by agent > Y% ii. by AI > Y% 3. Who is dealing with those messages? entity_type a. GUEST_QUESTION b. AP_REPLY c. AGENT_REPLY 4. Connect guest_id to check a. ISC (Initiated Shopping Carts) in the offers they started the conversation VS in a different offer i. Before conversation (check dates) ii. After conversation (check dates) b. Unique Requests in the offers they started the conversation VS in a different offer i. Paid Bookings ii. Rejected iii. Etc c. Revenue We don't need to see the conversations nor messages, we just need the OVERALL data from here. If you have any questions, please refer to the reporters of the BI request. This is quite important so give it the maximum priority possible. Our PED team invested many many hours. Also from OPS, to decide and see what was needed and right now since all members from Bookings Team are dealing with. Thanks in advance 😊

Figura 14 - Pedido do Relatório de Desempenho "Talk to the Landlord"

Os requisitos definidos para o relatório abrangiam diversas dimensões de análise. No que respeita às conversações recebidas, era necessário apurar o número total de conversações tratadas por *Account Providers* (AP) e por *Account Managers* (AM). Relativamente às mensagens recebidas, pretendia-se distinguir as que haviam sido classificadas como seguras (*Mark as Safe*) das que haviam sido sinalizadas (*Flagged*), em ambos os casos discriminando entre intervenção humana, por parte dos agentes, e intervenção automatizada por Inteligência Artificial. Era igualmente necessário identificar quem estava a lidar com as mensagens, com base no campo *entity_type*, que contemplava as categorias *Guest Question*, *Ap Reply* e *Agent Reply*. Por fim, era requerida uma análise cruzada com o *guest_id*, com o objetivo de verificar o comportamento dos utilizadores nas ofertas em que as conversações foram iniciadas, comparativamente a outras ofertas, considerando métricas como *Initiated Shopping Carts* (ISC), reservas pagas, reservas rejeitadas e receita gerada.

Toda a informação necessária à construção do relatório encontrava-se armazenada na base de dados interna da *Uniplaces*, acessível através do *schema prod_reporting*, na tabela *conversation*. A fase de extração de dados foi iniciada com a formulação de uma consulta em SQL (ver Figura 15), que permitiu identificar e relacionar os dados relevantes para a análise. A consulta desenvolvida recorreu a junções entre as tabelas *booking*, *accommodation_offer* e *conversation*, estabelecendo a ligação entre os identificadores de hóspede (*guest_id*) e de oferta (*offer_id*), e filtrando os registos com base no campo *financial_has_payment_details*, de forma a garantir a inclusão apenas de reservas com informação financeira associada. O resultado desta consulta permitiu obter um conjunto de dados estruturado, contendo o estado atual de cada reserva (*current_state*), as datas de pedido e pagamento e o identificador da oferta correspondente.

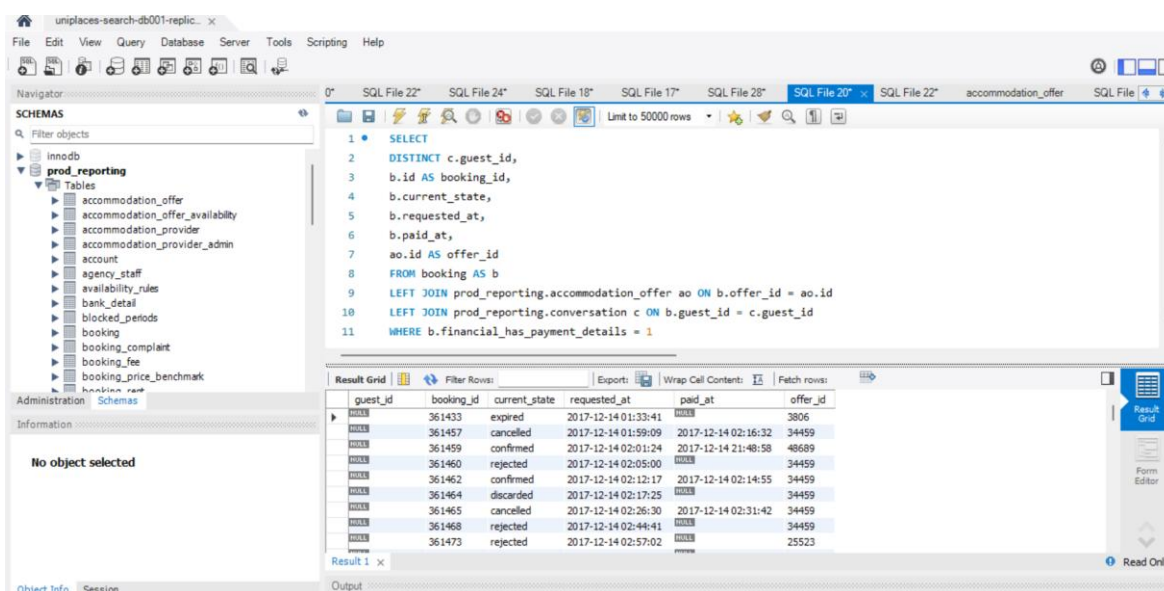


Figura 15 - Consulta inicial no SQL

Após a extração dos dados para o *Google Sheets*, procedeu-se à organização e estruturação da informação em tabelas de análise. Numa primeira fase, os dados foram agregados de forma a distinguir as interações processadas por Inteligência Artificial das tratadas por agentes humanos, cruzando esta distinção com métricas como o número de conversações, o número de hóspedes únicos (*Unique Guests*), o número de AP únicos (*Unique AP*) e o número de ofertas envolvidas (*Offers*). Para cada uma destas dimensões, foram calculados os totais correspondentes às classificações *BYPASS*, *NULL* e *SAFE*, conforme ilustrado na Figura 16. Esta segmentação revelou-se particularmente relevante para avaliar a eficácia do sistema de moderação automática implementado na ferramenta, permitindo quantificar em que medida a

IA estava a classificar corretamente as mensagens recebidas comparativamente à intervenção humana.

	Check	BYPASS	NULL	SAFE	TOTAL
Conversations	AI	546	315	13	874
Unique Guests	AI	265	268	12	268
Unique APs	AI	120	122	12	122
Offers	AI	234	240	13	240
Conversations	HUMAN	119	336	419	874
Unique Guests	HUMAN	99	268	178	268
Unique APs	HUMAN	62	122	100	122
Offers	HUMAN	97	240	174	240

Figura 16 - Mensagens recebidas por IA e pelo Agente

Numa fase subsequente, os mesmos dados foram reapresentados numa perspetiva temporal semanal, distribuindo os valores por semana (da semana 40 à semana 50), o que permitiu observar a evolução da adoção e utilização da ferramenta ao longo do tempo, conforme demonstrado na Figura 17. Esta abordagem revelou-se de elevado valor analítico, na medida em que possibilitou identificar tendências de crescimento ou decrescimento na utilização da funcionalidade, bem como detetar eventuais anomalias ou quebras no volume de interações em determinados períodos.

		WEEK 40	WEEK 41	WEEK 42	WEEK 43	WEEK 44	WEEK 45	WEEK 46	WEEK 47	WEEK 48	WEEK 49	WEEK 50
	Messages received											
AI	Conversations	34	218	361	306	0	0	0	0	0	0	0
AI	Unique Guests	23	122	194	164	0	0	0	0	0	0	0
AI	Unique APs	11	65	105	108	0	0	0	0	0	0	0
AI	Unique Offers	20	115	191	171	0	0	0	0	0	0	0
HUMAN	Conversations	14	180	344	342	0	0	0	0	0	0	0
HUMAN	Unique Guests	9	92	174	178	0	0	0	0	0	0	0
HUMAN	Unique APs	6	49	97	115	0	0	0	0	0	0	0
HUMAN	Unique Offers	7	83	171	179	0	0	0	0	0	0	0
	Who is dealing with those messages?											
	Guest_Question	25	150	238	173	0	0	0	0	0	0	0
	AP_REPLY	9	68	123	133	0	0	0	0	0	0	0

Figura 17 - Perspetiva semanal entre IA e Agente

No decurso do desenvolvimento desta tarefa, foi ainda adquirido conhecimento sobre uma funcionalidade do *Google Apps Script* até então não utilizada: os *triggers*. Esta funcionalidade permite configurar a execução automática de funções previamente definidas em intervalos de tempo pré-estabelecidos, sem necessidade de intervenção manual. No caso concreto, foi configurado um *trigger* do tipo *Time-driven* com um intervalo de execução horária, conforme ilustrado na Figura 18, garantindo que os dados presentes nas folhas de cálculo eram atualizados de forma contínua e automática. Foram criados cinco *triggers* distintos, correspondentes a

funções específicas do relatório, nomeadamente *extract_conversation_weekly_data*, *exportDistinctGuests*, *message_received*, *extract_AI_HUMAN_data* e *entity_type*, todos com uma taxa de erro de 0%, conforme visível na Figura 19, o que atesta a estabilidade e fiabilidade das automatizações implementadas.

Add Trigger for Untitled project

Select event source
Time-driven

Select type of time based trigger
Hour timer

Select hour interval
Every hour

Cancel Save

Figura 18 - Configuração de um *Trigger*

Triggers Showing 5 triggers

+ Add a filter

Owned by	Last run	Deployment	Event	Function	Error rate
Me	Oct 24, 2025, 4:18:00PM	Head	Time-based	<i>extract_conversation_weekly_data</i>	0%
Other user	Oct 24, 2025, 5:46:54AM	Head	Time-based	<i>exportDistinctGuests</i>	0%
Me	Oct 24, 2025, 4:26:16PM	Head	Time-based	<i>message_received</i>	0%
Me	Oct 24, 2025, 4:55:47PM	Head	Time-based	<i>extract_AI_HUMAN_data</i>	0%
Me	Oct 24, 2025, 4:40:51PM	Head	Time-based	<i>entity_type</i>	0%

Figura 19 - Automatização de cinco *Triggers* no *AppScript*

A utilização dos *triggers* representou um contributo significativo para a qualidade e utilidade do relatório produzido, na medida em que eliminou a necessidade de atualizações manuais periódicas, assegurando que a informação disponibilizada às equipas refletia sempre os dados mais recentes. Esta abordagem é particularmente relevante num contexto empresarial em que a tomada de decisão depende da disponibilidade de informação atualizada em tempo útil.

Concluída a fase de extração e estruturação dos dados no *Google Sheets*, o processo evoluiu para a construção de um *dashboard* interativo em *Tableau*, ferramenta de visualização de dados amplamente utilizada no contexto empresarial para a criação de relatórios visuais e interativos de suporte à tomada de decisão [10]. Esta representou a primeira experiência com a ferramenta

no âmbito do estágio, tendo sido desenvolvida com o apoio da colega Alyssa Encarnacion, Sênior *Data Analyst* da equipa.

O primeiro passo consistiu na ligação do *Tableau* à base de dados interna da *Uniplaces*. Para tal, foi estabelecida uma conexão direta ao servidor *MySQL* da organização (*Uniplaces-search-db001-replica.Uniplaces.com*), utilizando as credenciais de acesso da equipa de operações, conforme ilustrado na Figura 20. Esta ligação permitiu aceder diretamente ao *schema prod_reporting*, eliminando a necessidade de exportações intermédias de dados e garantindo que a informação presente no *dashboard* reflita sempre os dados mais atualizados da base de dados.

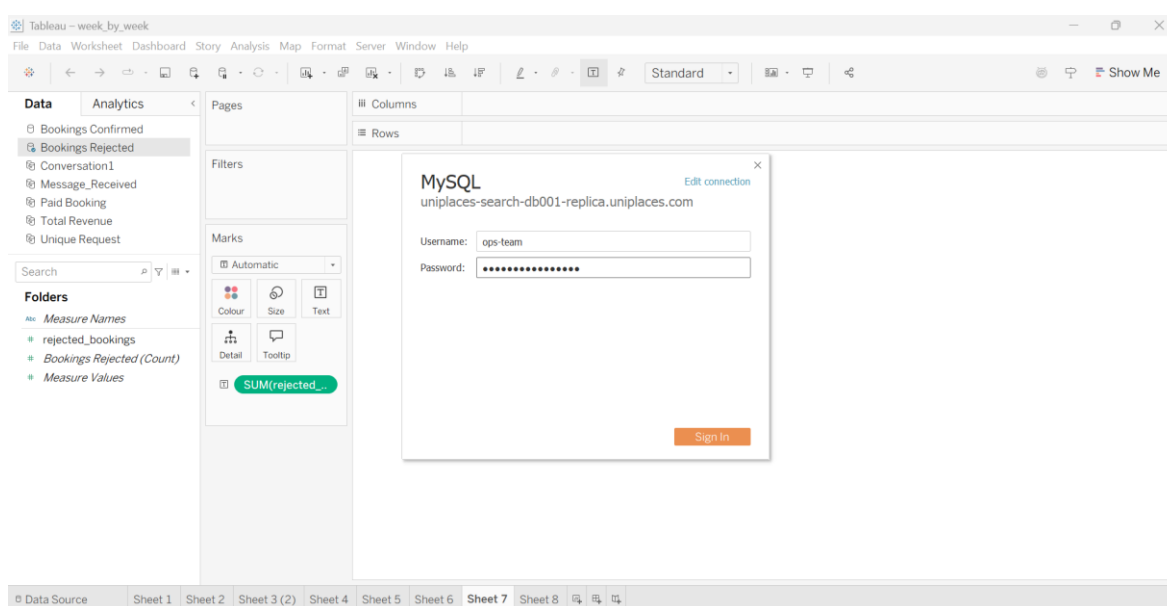


Figura 20 - Ligação do *Tableau* à base de dados da *Uniplaces*

Após o estabelecimento da ligação, procedeu-se à definição das fontes de dados através da escrita de consultas SQL personalizadas (*Custom SQL*) diretamente na interface do *Tableau*, conforme demonstrado na Figura 21. Esta abordagem revelou-se mais adequada do que a utilização direta das tabelas disponíveis, na medida em que permitiu definir com precisão os campos, agregações e filtros necessários para cada componente do *dashboard*. A consulta desenvolvida para a análise semanal das interações por tipo de moderação recorreu à cláusula *UNION ALL* para combinar os resultados relativos às interações processadas por Inteligência Artificial e por agentes humanos numa única fonte de dados, discriminando métricas como o número de conversações, hóspedes únicos, *providers* únicos e ofertas únicas, agrupadas por semana. O campo *ai_mod_updated_at_week* foi utilizado como critério temporal de agrupamento, sendo os dados filtrados para o intervalo de semanas relevante para a análise. A construção destas consultas revelou-se particularmente desafiante, dado que os valores obtidos

nem sempre correspondiam ao esperado, exigindo múltiplas iterações de teste e validação até à obtenção de resultados consistentes e fiáveis.

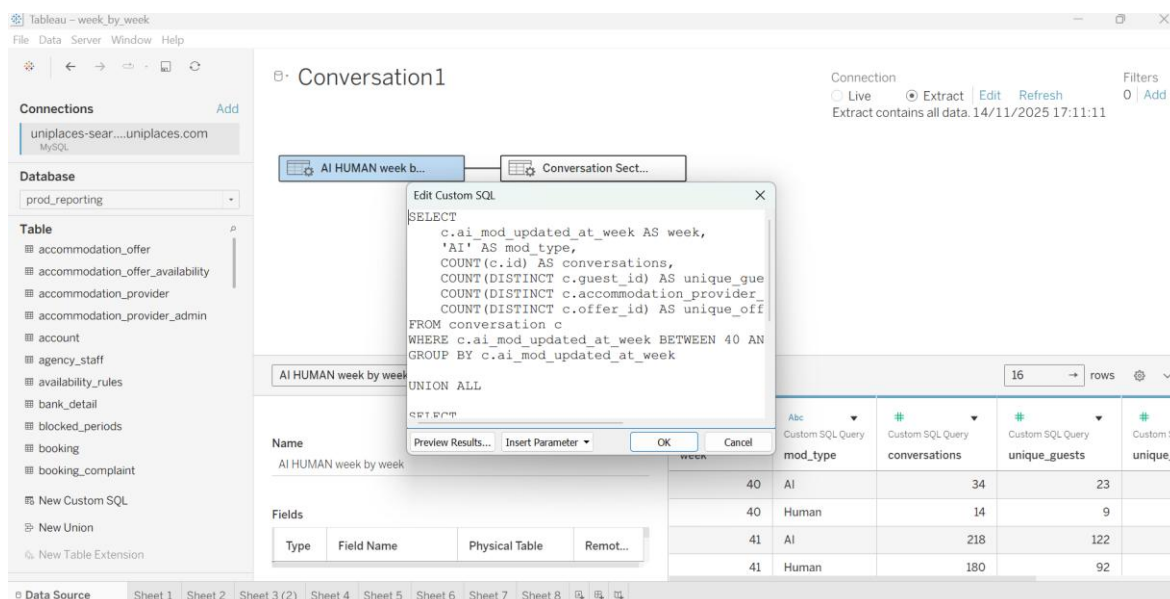


Figura 21 - Consulta SQL no Tableau

A primeira visualização desenvolvida no *Tableau*, ilustrada na Figura 22, consistiu numa tabela de calor (*heatmap*) que apresentava, semana a semana, o número de conversações, hóspedes únicos, *providers* únicos e ofertas únicas, segmentados por tipo de moderação (IA e *Human*). A opção por esta tipologia de visualização justifica-se pela sua capacidade de transmitir, de forma simultânea, a magnitude dos valores e a sua evolução temporal, através de uma escala de cor que permite identificar visualmente os períodos de maior e menor atividade sem necessidade de análise numérica detalhada. Este tipo de representação é particularmente eficaz quando se pretende comparar múltiplas métricas ao longo de vários períodos temporais numa única visualização [10].

Uma das componentes que apresentou maior grau de complexidade técnica foi a implementação do filtro temporal *Timespan*, que permitia ao utilizador selecionar o período de análise pretendido (como a última semana ou a totalidade do histórico disponível). A operacionalização desta funcionalidade exigiu a criação de vários campos calculados no *Tableau*, que relacionavam as datas presentes nos dados com o intervalo temporal selecionado pelo utilizador. Esta lógica de filtragem dinâmica, embora de implementação morosa e exigente, conferiu ao *dashboard* uma flexibilidade analítica significativa, permitindo que o mesmo instrumento fosse utilizado tanto para análises de curto prazo como para uma visão acumulada de toda a atividade registada desde o lançamento da ferramenta.

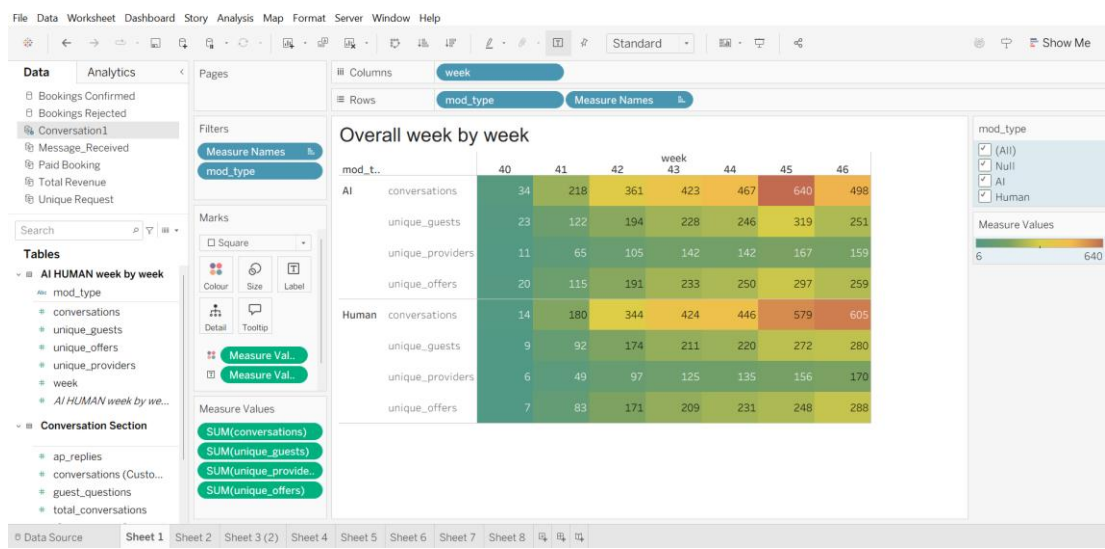


Figura 22 - Heatmap semanal

O *dashboard* final, apresentado na Figura 23 e na Figura 24, agregou o conjunto de visualizações desenvolvidas numa interface coesa e estruturada, incorporando os seguintes elementos.

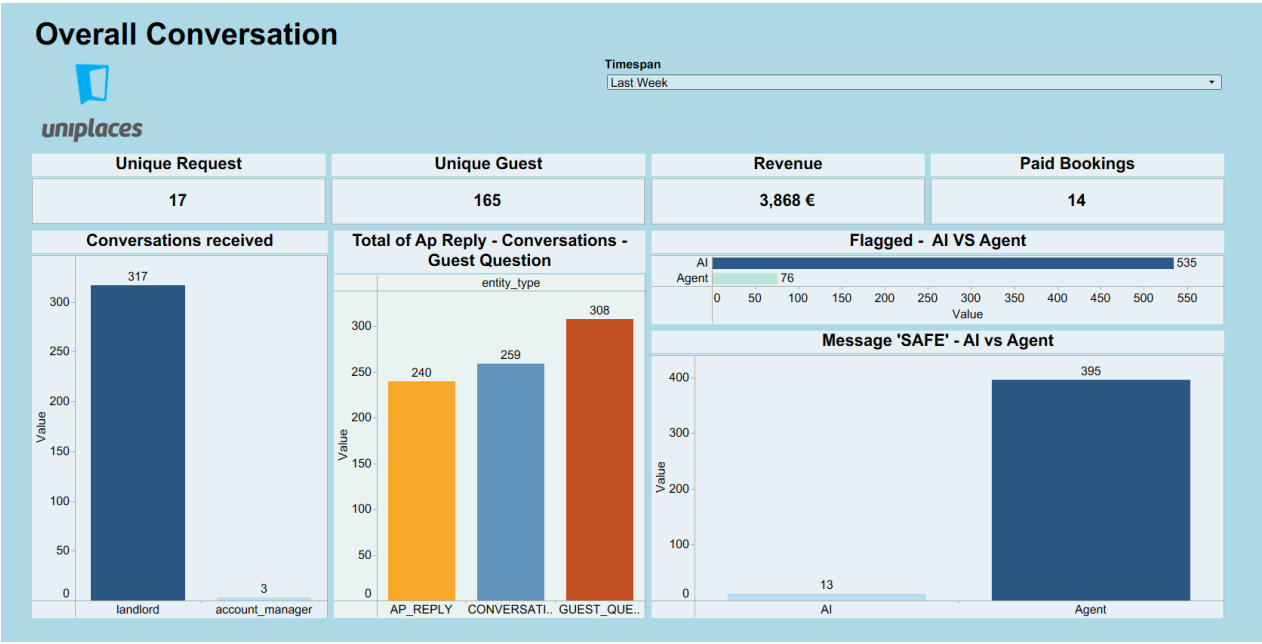


Figura 23 - Primeira versão do Dashboard Final

Os indicadores de topo - *Unique Request*, *Unique Guest*, *Revenue*, *Paid Bookings*, *Conversion in Unique Request* e *Conversion in Paid Bookings* - foram apresentados sob a forma de cartões de métrica (*KPI cards*), por se tratar da forma mais direta e legível de comunicar valores absolutos e taxas a um público executivo, sem necessidade de interpretação visual adicional.

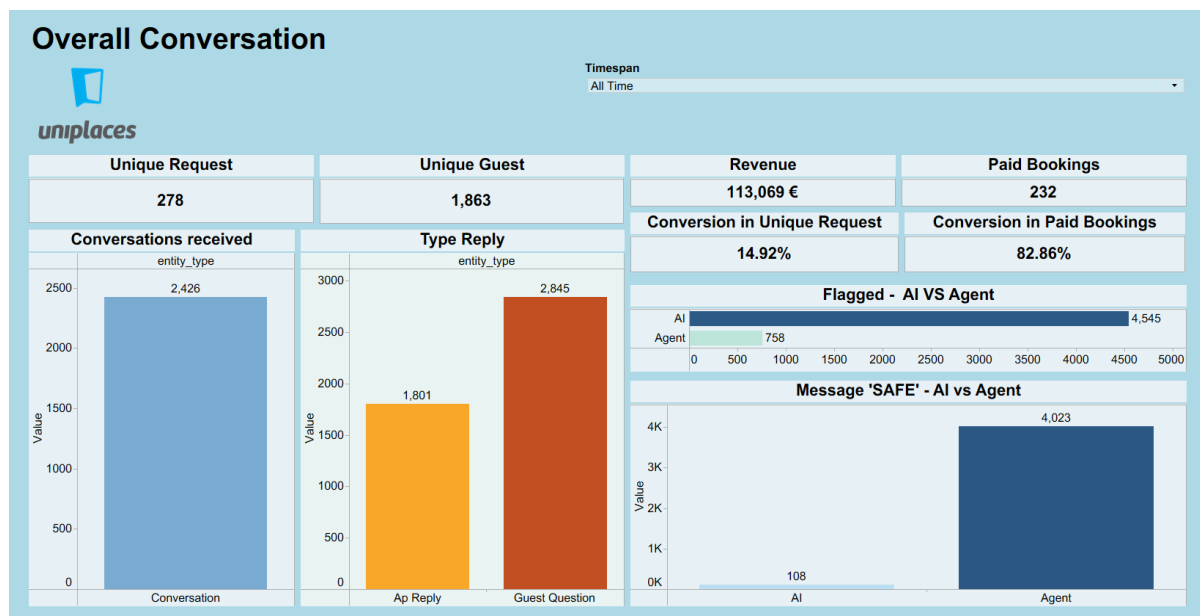


Figura 24 - Dashboard Final para a Ferramenta “Talk To The Landlord”

O gráfico de barras verticais relativo às conversações recebidas foi adotado por constituir a representação mais intuitiva para a comparação de volumes entre categorias discretas, neste caso entre conversações iniciadas por proprietários (*landlord*) e por gestores de conta (*account_manager*), permitindo uma leitura imediata das proporções relativas [66]. A mesma lógica presidiu à escolha do gráfico de barras para a representação do *Type Reply*, onde se distinguem as respostas por *AP_Reply* e *Guest_Question*, facilitando a comparação direta entre os dois tipos de interação. Para a análise comparativa entre a Inteligência Artificial e os agentes humanos nas dimensões *Flagged* e *Message 'SAFE'*, foi igualmente utilizado o gráfico de barras horizontais, tipologia especialmente adequada quando os valores a comparar são de grandezas muito distintas, pois permite uma leitura mais clara das diferenças de magnitude entre as categorias.

No decurso do projeto, foi realizada uma reunião com Renato Duarte, responsável pela equipa de Produto e um dos solicitantes originais do pedido de BI, com o objetivo de validar o trabalho desenvolvido e identificar eventuais melhorias a introduzir, conforme ilustrado na Figura 25. A reunião revelou-se produtiva, tendo permitido clarificar a lógica de análise pretendida pela equipa de Produto e identificar ajustes a realizar tanto ao nível das visualizações como das consultas subjacentes. Neste contexto, Renato Duarte colaborou ativamente na reformulação de algumas das consultas SQL, contribuindo para a sua otimização em termos de eficiência e precisão dos resultados obtidos.

pedido (*requested_at*), o dia da semana correspondente, calculado através da função *DATE_FORMAT* com o parâmetro *%W*, a data de *move-in* pretendida pelo hóspede, o dia da semana do *move-in*, a data de aceitação da reserva, o número de horas de antecedência entre o pedido e o *move-in* (*hours_until_movein*), o tempo de espera até à aceitação em horas (*awaiting_time_hours*) e o estado atual da reserva (*booking_status*). A definição do critério de *last minute* foi operacionalizada através de um filtro aplicado ao campo *hours_until_movein*, restringindo os resultados aos casos em que o intervalo entre o pedido e o *move-in* era igual ou inferior a 72 horas, correspondendo ao limiar de três dias estabelecido como referência para esta categoria de reservas.

```

1 • SELECT
2     b.id AS booking_id,
3     CONCAT(g.first_name, ' ', g.last_name) AS TT_Name,
4     ap.name AS AP_Name,
5     b.requested_at AS 'Requested date-time of booking',
6     DATE_FORMAT(b.requested_at, '%W') AS requested_day,
7     b.guest_move_in AS 'Move-in date',

```

Requested date-time of booking	requested_day	Move-in date	movein_day	Accepted Date	hours_until_movein	awaiting_time_hours	booking_status
2020-02-18 09:55:08	Tuesday	2020-02-20 00:00:00	Thursday	2020-02-18 10:31:48	62	0	confirmed
2020-02-01 11:44:02	Saturday	2020-02-01 00:00:00	Saturday	NULL	12	NULL	rejected
2020-01-31 09:22:11	Friday	2020-02-01 00:00:00	Saturday	NULL	38	NULL	discarded
2020-01-26 14:17:25	Sunday	2020-01-26 00:00:00	Sunday	NULL	9	NULL	expired
2020-01-26 12:13:10	Sunday	2020-01-26 00:00:00	Sunday	2020-01-26 13:58:05	11	1	confirmed

Figura 26 - Consulta SQL desenvolvida para a extração dos dados de reservas de último minuto

As duas métricas calculadas no âmbito desta consulta assumem particular relevância do ponto de vista operacional. O campo *hours_until_movein* quantifica a janela de tempo efetivamente disponível para que a reserva seja processada e aceite antes da data de entrada do hóspede, constituindo um indicador direto da urgência associada a cada caso. O campo *awaiting_time_hours*, por sua vez, mede o intervalo de tempo decorrido entre a submissão do pedido e a sua aceitação por parte do agente responsável, permitindo avaliar a capacidade de resposta da equipa operacional face a situações de elevada urgência. A relação entre estas duas métricas é determinante para a identificação dos casos críticos: sempre que o tempo de espera até à aceitação é superior à antecedência disponível, a reserva é classificada como problemática, sinalizando uma situação em que a equipa não terá conseguido processar o pedido em tempo útil.

Numa fase inicial, os dados extraídos foram carregados para o *Google Sheets*, dando origem a uma primeira versão do relatório, visível na Figura 27, que permitiu validar a estrutura e o conteúdo da informação obtida através da consulta SQL. Esta fase exploratória revelou-se

importante para identificar padrões relevantes nos dados, nomeadamente a recorrência de reservas de último minuto associadas a períodos de fim de semana. Este padrão é explicável pelo facto de os serviços de apoio ao cliente operarem com menor capacidade durante o fim de semana, o que aumenta o risco de atrasos no processamento de reservas urgentes. Com base nesta observação, foram adicionadas ao relatório colunas adicionais com o dia da semana correspondente tanto à data de pedido como à data de *move-in*, bem como uma *flag* específica para identificar os casos em que o pedido havia sido submetido numa sexta-feira com *move-in* previsto para segunda-feira, situação particularmente crítica, dado que implica que a reserva necessita de ser processada durante o fim de semana para garantir a entrada atempada do hóspede.

		C	D	E	F	G	H	I	J	K
		AP Name	Requested at	Requested day	Move-in date	Move-in day	Accepted at	Hours until move-in	Awaiting time (h)	Booking status
8	xhi		2024-05-30 6:28:24	Thursday	2024-05-31 0:00:00	Friday		41		discarded
9			2024-05-17 15:12:03	Friday	2024-05-17 0:00:00	Friday		8		rejected
10	DN		2024-05-13 13:15:13	Monday	2024-05-15 0:00:00	Wednesday		58		discarded
11	g		2024-05-06 20:45:10	Monday	2024-05-06 0:00:00	Monday		3		discarded
12			2024-05-02 22:42:28	Thursday	2024-05-05 0:00:00	Sunday		73		discarded
13			2024-05-01 14:35:52	Wednesday	2024-05-02 0:00:00	Thursday		33		discarded
14			2024-04-30 10:28:41	Tuesday	2024-05-01 0:00:00	Wednesday		37		discarded
15	do		2024-03-22 16:46:48	Friday	2024-03-22 0:00:00	Friday		7		rejected
16			2024-03-08 14:10:58	Friday	2024-03-08 0:00:00	Friday		9		discarded
17			2024-02-27 2:00:32	Tuesday	2024-03-01 0:00:00	Friday		93		discarded
18	ary		2024-02-20 10:08:34	Tuesday	2024-02-20 0:00:00	Tuesday	2024-02-20 17:13:33	13	7	payment-expired
19	iti		2024-02-04 15:34:46	Sunday	2024-02-05 0:00:00	Monday		32		discarded
20	i		2024-02-01 22:47:05	Thursday	2024-02-03 0:00:00	Saturday		49		discarded
21	alez		2023-12-23 9:18:00	Saturday	2023-12-23 0:00:00	Saturday		14		discarded
22	z		2023-12-13 15:33:07	Wednesday	2023-12-16 0:00:00	Saturday	2023-12-13 15:39:49	80	0	confirmed
23	sz		2023-12-03 11:32:56	Sunday	2023-12-05 0:00:00	Tuesday		60		discarded
24	ul		2023-11-22 8:55:49	Wednesday	2023-11-22 0:00:00	Wednesday		15		expired
25			2023-10-03 15:51:09	Tuesday	2023-10-06 0:00:00	Friday		80		discarded
26			2023-10-01 16:26:36	Sunday	2023-10-01 0:00:00	Sunday	2023-10-01 17:18:20	7	0	confirmed
27	se		2023-09-29 18:16:41	Friday	2023-10-02 0:00:00	Monday		77		discarded
28	v		2023-09-29 14:21:02	Friday	2023-10-02 0:00:00	Monday		81		discarded
29	la		2023-09-20 20:28:10	Wednesday	2023-09-22 0:00:00	Friday		51		discarded
30	Palma Barao		2023-09-07 6:17:50	Thursday	2023-09-09 0:00:00	Saturday		65		discarded

Figura 27 - Primeira versão do relatório *Bookings Last Minute* no *Google Sheets*

Após a validação da estrutura do relatório, procedeu-se à sua automatização através do *Google Apps Script*. O *script* desenvolvido foi configurado para executar, de forma periódica e sem necessidade de intervenção manual, as seguintes operações: extração dos dados da base de dados com base na consulta SQL previamente definida, carregamento dos registos obtidos na folha de cálculo correspondente e aplicação de formatação condicional automática para destacar visualmente os casos que requeriam atenção imediata. A lógica de formatação implementada seguiu um código de cores estruturado: a cor laranja foi aplicada às células correspondentes a reservas com *move-in* em dias de fim de semana, a cor vermelha foi reservada para os casos de maior criticidade, a cor verde foi utilizada para assinalar os pedidos submetidos à sexta-feira com *move-in* na segunda-feira, e a cor amarela foi aplicada especificamente à célula do tempo de espera (*awaiting_time_hours*) nos casos em que este valor excedia a antecedência

disponível, sinalizando as situações em que a capacidade de resposta da equipa havia ficado aquém do necessário.

O resultado final do relatório, ilustrado na Figura 28, apresenta uma visão consolidada das reservas de último minuto, complementada por um painel de indicadores de síntese que discrimina o total de reservas analisadas por estado, nomeadamente pagas, rejeitadas, confirmadas, canceladas e expiradas, bem como a distinção entre reservas aceites e não aceites. Esta estrutura permitiu à gestão operacional da *Uniplaces* dispor de uma ferramenta de monitorização clara e atualizada, capaz de identificar rapidamente os casos prioritários e apoiar a tomada de decisão em situações de urgência.

Booking ID	TT Name	Requested at	Requested day	Move-in date	Move-in day	Accepted at	Hours until move-in	Awaiting time (h)	Booking status
1126370		2025-10-29 12:31:28	Wednesday	2025-11-01 0:00:00	Saturday	2025-10-29 12:20:50	83	0	rejected
1126364		2025-10-29 12:04:26	Wednesday	2025-11-01 0:00:00	Saturday	2025-10-29 4:00:46	93	1	paid
1126305		2025-10-29 2:53:54	Wednesday	2025-11-01 0:00:00	Saturday	2025-10-28 19:30:30	4	0	paid
1126244		2025-10-28 19:26:03	Tuesday	2025-10-28 0:00:00	Tuesday	2025-10-28 14:09:28	9	0	cancelled
1126237		2025-10-28 18:31:11	Tuesday	2025-10-28 0:00:00	Tuesday	2025-10-28 9:23:03	14	0	rejected
1126163		2025-10-28 14:06:50	Tuesday	2025-10-28 0:00:00	Tuesday		15		rejected
1126095		2025-10-28 9:22:47	Tuesday	2025-10-28 0:00:00	Tuesday		29		rejected
1126080		2025-10-28 8:08:36	Tuesday	2025-10-28 0:00:00	Tuesday		29		rejected
1127983		2025-10-27 18:18:27	Monday	2025-10-28 0:00:00	Tuesday		12		expired
1127977		2025-10-27 18:05:07	Monday	2025-10-28 0:00:00	Tuesday	2025-10-27 16:01:40	9	1	paid
1127912		2025-10-27 14:34:39	Monday	2025-10-27 0:00:00	Monday		50		expired
1127890		2025-10-27 14:54:31	Monday	2025-10-27 0:00:00	Monday	2025-10-26 21:53:06	27	1	paid
1127796		2025-10-26 21:52:58	Sunday	2025-10-28 0:00:00	Tuesday	2025-10-26 13:04:16	83	0	paid
1127788		2025-10-26 20:50:16	Sunday	2025-10-27 0:00:00	Monday	2025-10-24 8:53:52	15	0	confirmed
1127719		2025-10-26 12:49:06	Sunday	2025-10-29 0:00:00	Wednesday		16		rejected
1127417		2025-10-24 8:53:51	Friday	2025-10-24 0:00:00	Friday	2025-10-23 18:44:47	6	0	confirmed
1127404		2025-10-24 7:56:39	Friday	2025-10-24 0:00:00	Friday		47		rejected
1127334		2025-10-23 17:58:17	Thursday	2025-10-23 0:00:00	Thursday		14		rejected
1127150		2025-10-23 16:05:13	Thursday	2025-10-24 0:00:00	Friday	2025-10-22 7:56:11	27	11	cancelled
1126955		2025-10-22 9:28:07	Wednesday	2025-10-22 0:00:00	Wednesday		30		rejected
1126854		2025-10-21 20:40:41	Tuesday	2025-10-22 0:00:00	Wednesday		83	0	paid
1126559		2025-10-20 17:54:44	Monday	2025-10-21 0:00:00	Tuesday	2025-10-21 7:19:45	36	19	confirmed
1126480		2025-10-20 12:53:27	Monday	2025-10-23 0:00:00	Thursday		14		rejected
1126456		2025-10-20 11:26:54	Monday	2025-10-21 0:00:00	Tuesday		4	0	confirmed
1126280		2025-10-19 9:05:54	Sunday	2025-10-19 0:00:00	Sunday	2025-10-18 19:35:39	8		rejected
1126222		2025-10-18 19:15:02	Saturday	2025-10-18 0:00:00	Saturday		27		rejected
1126189		2025-10-18 15:29:27	Saturday	2025-10-18 0:00:00	Saturday		32		rejected
1126092		2025-10-17 20:59:01	Friday	2025-10-18 0:00:00	Saturday		34		rejected
1126017		2025-10-17 15:12:35	Friday	2025-10-18 0:00:00	Saturday		34		rejected
1125985		2025-10-17 13:10:29	Friday	2025-10-18 0:00:00	Saturday				rejected
1125982		2025-10-17 13:04:43	Friday	2025-10-18 0:00:00	Saturday				rejected

Figura 28 - Resultado final do relatório *Bookings Last Minute* com painel de indicadores de síntese e formatação condicional automatizada

Do ponto de vista do desenvolvimento técnico, esta tarefa apresentou um grau de complexidade considerável, particularmente na definição da lógica de cálculo das métricas temporais em SQL e na implementação da formatação condicional automatizada no *Google Apps Script*, que exigiu o tratamento de múltiplas condições em simultâneo. O resultado obtido constituiu um contributo relevante para a operação da empresa, ao disponibilizar um instrumento de gestão que tornou visível e mensurável um problema operacional previamente identificado de forma apenas qualitativa pelas equipas.

3.3.5. Migração de *Dashboards B2B* para *Looker Studio*

A transição das ferramentas de produtividade para plataformas dedicadas de BI constitui um passo crítico na maturidade analítica de organizações que lidam com volumes crescentes de

informação. A migração para ecossistemas de visualização avançada é frequentemente motivada pela necessidade de reduzir a latência na tomada de decisão e garantir a escalabilidade dos processos que, em ferramentas tradicionais como as folhas de cálculo, tendem a tornar-se instáveis e propensos a erros de processamento [45]. Na *Uniplaces*, este cenário levou a equipa de dados a iniciar um projeto de migração de *dashboards* estratégicos para o *Looker Studio*, uma plataforma que, embora funcione como uma extensão das ferramentas *Google*, oferece capacidades de agregação e interatividade superiores, permitindo transformar dados brutos em ativos de conhecimento acionáveis.

O projeto teve o seu início formal com o levantamento de requisitos documentado através do *Confluence* (ver Figura 29), onde se estabeleceu uma colaboração estreita com a equipa de B2B para definir a estrutura e as funcionalidades necessárias para o novo sistema de reporte. No pedido formal, ficou estipulado o desenvolvimento de quatro ficheiros centrais, nomeadamente o acompanhamento de reservas com Universidades, parcerias com Instituições, uma visão global de dados (*Overall Data*) e o documento "Recap" focado nos objetivos estratégicos para 2026. Esta fase de planeamento foi determinante para alinhar as expectativas sobre os filtros dinâmicos e as dimensões a incluir, garantindo que a transição tecnológica não resultasse em perda de contexto operacional para os utilizadores finais.

Publicado há jan. 08 Editar Compartilhar

[BI Request] 2026-01-08 - Transfer 3.0 file to Looker

Por alyssa.encarnacion 4 1

Requester	@Giulia Provenza @vanessa.bunga
Team	
Urgency	High
Status	
Project Responsible	@pedro.fernandes
Time needed	To be filled by BI
Description	As discussed, move these 4 tabs in the 3.0 File to Looker 1. Overall Data (follow same format as in the file) 2. Bookings w/ Universities (follow same format as in the file) 3. Bookings w/ Institutions (follow same format as in the file) 4. Recap (if not possible to follow the same format, feel free to propose a different layout as long as the same data is visible) @pedro.fernandes Ensure they can filter by requested_at , partner name , promo code, and partner type. @Giulia Provenza / @vanessa.bunga , kindly advise if there are other tabs or tables or reports you want to add here. Also if there are other filters / functionalities you need

Figura 29 - Pedido formal de migração e levantamento de requisitos no *Confluence*

Do ponto de vista técnico, a arquitetura da solução baseou-se na extração direta de dados através dos conectores nativos do *Looker Studio* (ver Figura 30), especificamente a ligação ao servidor *MySQL*.

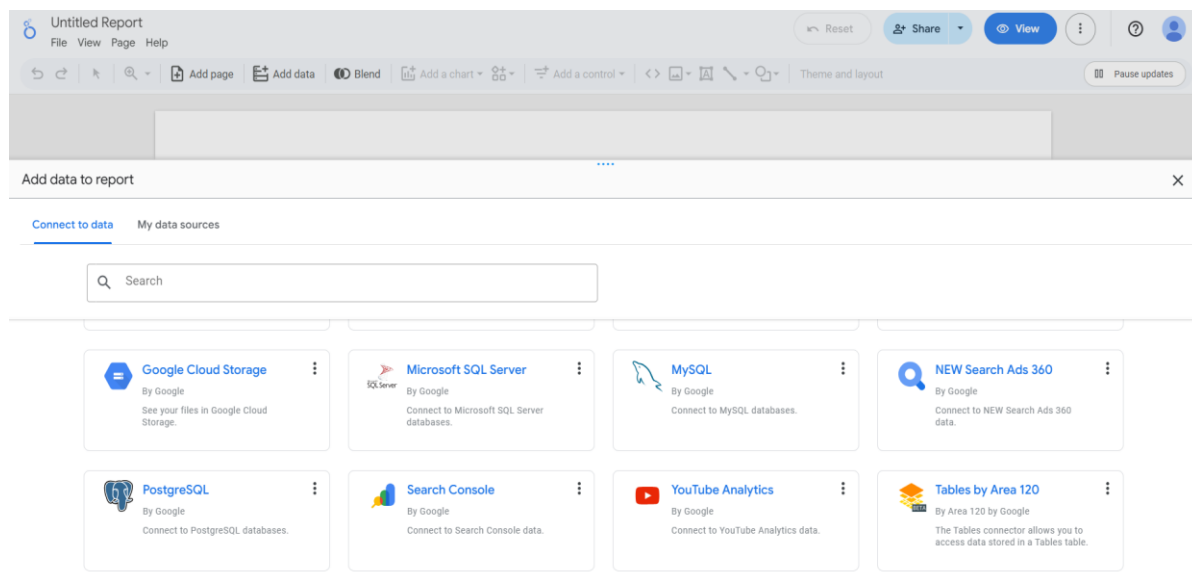


Figura 30 - Interface de seleção de conectores nativos no *Looker Studio (MySQL)*

Para assegurar que apenas a informação pertinente fosse migrada, foi implementada uma interrogação SQL customizada no *Data Source* (ver Figura 31), focada no esquema *prod_reporting_ops*.

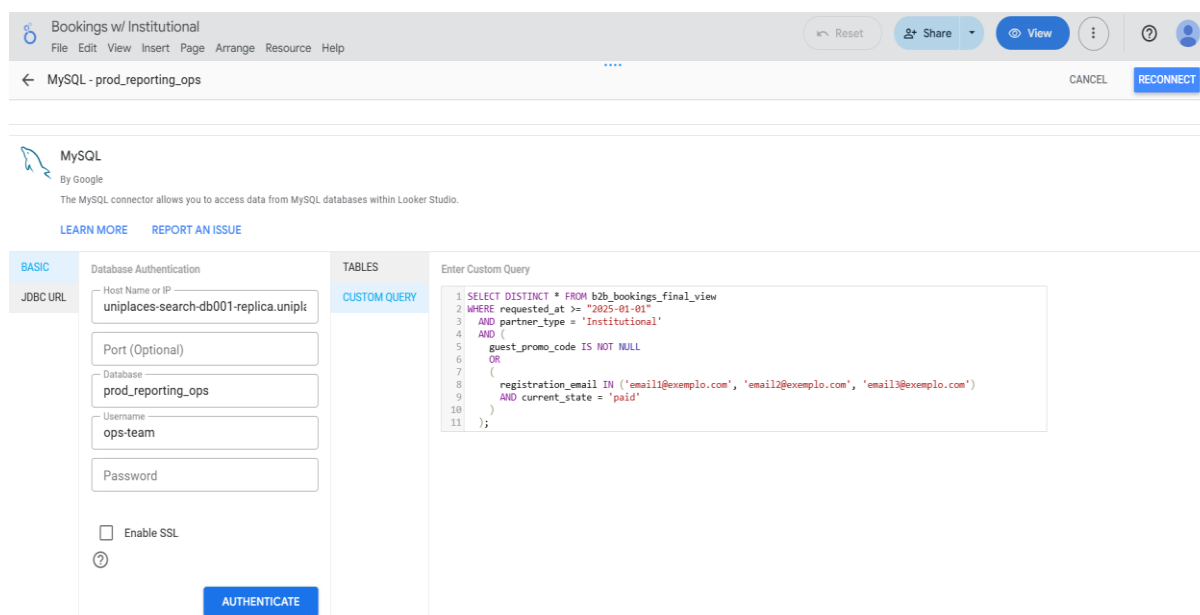


Figura 31 - Configuração da *Query* SQL customizada para extração de dados B2B

No entanto, a elevada densidade da informação exigiu uma abordagem focada na eficiência, optando-se pela funcionalidade de *Extract Data*, visível nas propriedades de configuração da tabela (ver Figura 32). Este método realiza o processamento da *query* central e armazena os resultados na infraestrutura da *Google*, otimizando a rapidez de consulta e a experiência de navegação, priorizando a performance do *dashboard* em detrimento de uma ligação constante e pesada à base de dados de produção.

a migração para plataformas analíticas de maior custo e complexidade até à refatorização dos processos originais no *Google Sheets*, demonstrando a necessidade de equilibrar constantemente as capacidades das ferramentas de BI com as exigências de tempestividade e orçamento da organização.

Bookings w/ Institutional

Reset

Share

Edit

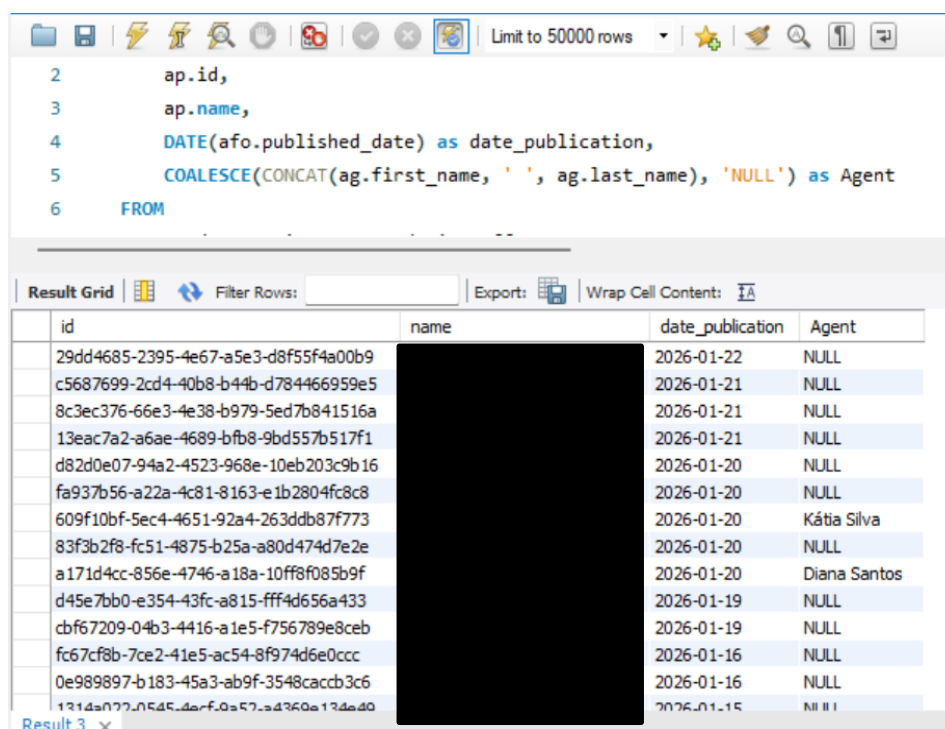
partner	partner_simplified	id	guest_name	requested_at	status_date	current_state	reason	accepted_at	paid_at	ap_id	guest_id	account_type	market	address_city_code	financial_has_payment_details	registration_email	
Select date range																	
1 University Living (new) (Off...	Unive...	1148025	Lutèce Agounne	Jan 15, 2026	Jan 15, 2026	paid	null	Jan 15, 2026	Jan 15, 2...	83d7...	50b0b0c3...	KA	ES	ES-barcelona	1	lutecia.agn@gm...	partner
2 University Living (new) (Off...	Unive...	1145994	Lutèce Agounne	Jan 15, 2026	null	requested	null	null	null	83d7...	50b0b0c3...	KA	ES	ES-barcelona	null	lutecia.agn@gm...	id
3 Castla (Offline Leads)	Castla	1145871	Mohammad Ja...	Jan 15, 2026	Jan 15, 2026	paid	null	Jan 15, 2026	Jan 15, 2...	58d5...	9d0ed095...	KA	DE	DE-dortmund	1	aktibipat024q...	current_state
4 University Living (new) (Off...	Unive...	1145882	Damian Kofi M...	Jan 15, 2026	null	requested	null	null	null	c9c5...	2952b79b...	SML	ES	ES-barcelona	null	damianp024q...	damianp024q
5 University Living (new) (Off...	Unive...	1145805	Damian Kofi M...	Jan 14, 2026	Jan 15, 2026	discarded	missing...	null	null	623c...	2952b79b...	SML	ES	ES-barcelona	0	damianp024q...	damianp024q
6 University Living (new) (Off...	Unive...	1145791	Damian Kofi M...	Jan 14, 2026	Jan 15, 2026	discarded	missing...	null	null	c9c5...	2952b79b...	SML	ES	ES-barcelona	null	damianp024q...	damianp024q
7 Castla (Offline Leads)	Castla	1145693	Mohammad Ja...	Jan 14, 2026	Jan 15, 2026	rejected	suggest...	null	null	58d5...	9d0ed095...	KA	DE	DE-dortmund	1	aktibipat024q...	registration_email
8 Castla (Offline Leads)	Castla	1145645	Mohammad Ja...	Jan 14, 2026	Jan 15, 2026	discarded	no-longer...	null	null	58d5...	9d0ed095...	KA	DE	DE-dortmund	null	aktibipat024q...	accepted_at
9 Castla (Offline Leads)	Castla	1145598	Fezzani Safia	Jan 14, 2026	Jan 15, 2026	discarded	missing...	null	null	a6c...	8a44b73e...	KA	ES	ES-barcelona	null	fezzanisafia5...	financial_has_pay...
1 University Living (new) (Off...	Unive...	1145597	Mahmoud Zorkot	Jan 14, 2026	Jan 15, 2026	discarded	missing...	null	null	5c0a...	db7f909a...	KA	ES	ES-malaga	null	mahmoudzorko...	mahmoudzorko
1 University Living (new) (Off...	Unive...	1145590	Mahmoud Zorkot	Jan 14, 2026	Jan 15, 2026	discarded	missing...	null	null	5c0a...	db7f909a...	KA	ES	ES-malaga	null	mahmoudzorko...	mahmoudzorko
1 Assignment Helping Service	Assig...	1145575	Rachele Gallina	Jan 14, 2026	Jan 14, 2026	discarded	suggest...	Jan 14, 2026	null	fee3...	30efaa853...	KA	ES	ES-madrid	1	rachelegallina2	Requested Week
1 Assignment Helping Service	Assig...	1145505	Rachele Gallina	Jan 13, 2026	Jan 14, 2026	rejected	suggest...	null	null	fee3...	30efaa853...	KA	ES	ES-madrid	1	rachelegallina2	Request Month
1 University Living (new) (Off...	Unive...	1145424	Lutèce Agounne	Jan 13, 2026	Jan 14, 2026	discarded	missing...	null	null	83d7...	50b0b0c3...	KA	ES	ES-barcelona	null	lutecia.agn@gm...	Request Year
1 Castla (Offline Leads)	Castla	1144890	Jeevika Hajit...	Jan 12, 2026	Jan 13, 2026	expired	null	null	null	4802...	49ba5d27...	SML	DE	DE-frankfurt	1	ipeevka@gmail	ipeevka@gmail
1 University Living (new) (Off...	Unive...	1144807	Vatsal Agarwal	Jan 12, 2026	Jan 13, 2026	rejected	offer-uns...	null	null	4967...	17b9f95-a...	KA	ES	ES-barcelona	1	vatsal@unibab...	Paid Week
1 University Living (new) (Off...	Unive...	1144764	Yael DiPietra	Jan 11, 2026	Jan 12, 2026	rejected	other...	null	null	e5f54...	149b96-4...	SML	IT	IT-rome	1	dipietrayael@p...	Paid Month
1 University Living (new) (Off...	Unive...	1144754	Yael DiPietra	Jan 11, 2026	Jan 12, 2026	discarded	missing...	null	null	e5f54...	149b96-4...	SML	IT	IT-rome	null	dipietrayael@p...	Paid Year
1 University Living (new) (Off...	Unive...	1144340	Vatsal Agarwal	Jan 9, 2026	Jan 9, 2026	rejected	tenant-o...	null	null	4967...	17b9f95-a...	KA	ES	ES-barcelona	1	vatsal@unibab...	vatsal@unibab
2 University Living (new) (Off...	Unive...	1144338	Vatsal Agarwal	Jan 9, 2026	Jan 10, 2026	discarded	missing...	null	null	4967...	17b9f95-a...	KA	ES	ES-barcelona	null	vatsal@unibab...	vatsal@unibab
2 University Living (new) (Off...	Unive...	1144281	Lutèce Agounne	Jan 9, 2026	Jan 10, 2026	discarded	missing...	null	null	83d7...	50b0b0c3...	KA	ES	ES-barcelona	null	lutecia.agn@gm...	lutecia.agn@gm
2 University Living (new) (Off...	Unive...	1144278	Vatsal Agarwal	Jan 9, 2026	Jan 9, 2026	rejected	offer-uns...	null	null	4967...	17b9f95-a...	KA	ES	ES-barcelona	1	vatsal@unibab...	Status Week
2 University Living (new) (Off...	Unive...	1144246	Jesse Emil	Jan 9, 2026	Jan 9, 2026	paid	null	Jan 9, 2026	Jan 9, 2026	89c1...	73caed0...	KA	DE	DE-berlin	1	emjesset@gp...	Status Month
2 MyStudentDigs.com (Offlin...	MySt...	1144170	Ayman Ibrahim	Jan 8, 2026	Jan 9, 2026	discarded	missing...	null	null	1628...	02a23a8f...	SML	PT	PT-lisbon	null	aymanicham7d...	Status Year
2 University Living (new) (Off...	Unive...	1144116	Vatsal Agarwal	Jan 8, 2026	Jan 9, 2026	rejected	offer-uns...	null	null	4967...	17b9f95-a...	KA	ES	ES-barcelona	1	vatsal@unibab...	vatsal@unibab
2 Student.com (Offline Leads)	Stude...	1144104	moustafa emre...	Jan 8, 2026	Jan 9, 2026	discarded	missing...	null	null	c181...	5993b32...	SML	IT	IT-rome	null	moustafa.emre.k	moustafa.emre.k

Figura 33 - Dashboard finalizado no Looker Studio com integração de filtros dinâmicos e métricas de parceiros

3.3.6. Auditoria e Regularização do Processo de Atribuição de Portefólio (*Account Management*)

Dando continuidade ao suporte analítico às diversas áreas de negócio, surgiu a necessidade de intervir num processo crítico para a equipa de *Account Management*. Após uma reunião com a Jeniffer (Responsável pelos *Account Managers*), foi reportada uma anomalia na distribuição automática de novos senhorios (*Accommodation Providers*) pelos agentes. O problema era particularmente evidente pela ausência de novas atribuições a vários gestores de conta num período superior a dois meses, o que comprometia a capacidade de acompanhamento e retenção destes parceiros na plataforma.

A primeira etapa da intervenção consistiu numa auditoria técnica via *MySQL* para validar a extensão da falha. Ao analisar o mercado de Portugal, conforme ilustrado na Figura 34, confirmou-se que uma fatia substancial de senhorios recém-publicados apresentava o campo *Agent* com o valor *NULL*. Esta ausência de atribuição significava que estes parceiros estavam "orfãos" no sistema, sem um ponto de contacto direto na *Uniplaces* para gerir as suas propriedades e reservas.



```

2      ap.id,
3      ap.name,
4      DATE(afo.published_date) as date_publication,
5      COALESCE(CONCAT(ag.first_name, ' ', ag.last_name), 'NULL') as Agent
6  FROM

```

id	name	date_publication	Agent
29dd4685-2395-4e67-a5e3-d8f55f4a00b9		2026-01-22	NULL
c5687699-2cd4-40b8-b44b-d784466959e5		2026-01-21	NULL
8c3ec376-66e3-4e38-b979-5ed7b841516a		2026-01-21	NULL
13eac7a2-a6ae-4689-bfb8-9bd557b517f1		2026-01-21	NULL
d82d0e07-94a2-4523-968e-10eb203c9b16		2026-01-20	NULL
fa937b56-a22a-4c81-8163-e1b2804fc8c8		2026-01-20	NULL
609f10bf-5ec4-4651-92a4-263ddb87f773		2026-01-20	Kátia Silva
83f3b2f8-fc51-4875-b25a-a80d474d7e2e		2026-01-20	NULL
a171d4cc-856e-4746-a18a-10ff8f085b9f		2026-01-20	Diana Santos
d45e7bb0-e354-43fc-a815-fff4d656a433		2026-01-19	NULL
cbf67209-04b3-4416-a1e5-f756789e8ceb		2026-01-19	NULL
fc67cf8b-7ce2-41e5-ac54-8f974d6e0ccc		2026-01-16	NULL
0e989897-b183-45a3-ab9f-3548cacb3c6		2026-01-16	NULL
121da072-05d5-dae-f-0a57-a4360a13da00		2026-01-15	NULL

Figura 34 - Consulta SQL para identificação de senhorios sem agente atribuído no mercado de Portugal

Para solucionar esta lacuna, procedeu-se à preparação de um processo de *Bulk Update*. Esta fase envolveu a extração manual dos *IDs* dos senhorios não atribuídos e o seu cruzamento num ficheiro de gestão de *Account Managers*. Na aba "*Bulk Update*" deste ficheiro (ver Figura 35), foi realizado o mapeamento entre o *ID* do senhorio e o *ID* do agente correspondente, respeitando a regra de especialização geográfica: agentes específicos para o mercado de Portugal, Espanha e Itália. Este rigor na correspondência de *IDs* é fundamental para garantir a integridade da base de dados e evitar erros de comunicação entre agentes e senhorios de diferentes idiomas.

Após a estruturação dos dados, o ficheiro foi exportado em formato CSV e submetido no OPS (*Operations Management System*), a plataforma interna que atua como interface de gestão da base de dados da *Uniplaces*. A Figura 36 demonstra a interface de carregamento, onde o pedido de atualização massiva é processado. Este procedimento requer coordenação com a equipa de Produto para assegurar que a inserção dos dados não causa conflitos com as rotinas automáticas do sistema.

A eficácia da intervenção foi confirmada através de uma nova consulta SQL de verificação. Como se observa na Figura 37, os senhorios anteriormente sem gestor passaram a apresentar a atribuição correta aos agentes (neste caso, a agente Kátia Silva), normalizando o fluxo de trabalho. Esta regularização é um fator determinante para a experiência do utilizador na plataforma; um senhorio bem acompanhado traduz-se numa maior qualidade de serviço, o que

impacta diretamente a satisfação final do inquilino. Atualmente, encontra-se em fase de estudo uma solução para automatizar este diagnóstico, permitindo à equipa de BI identificar e corrigir estas quebras de fluxo com maior prontidão.

Figura 35 - Estrutura da folha de cálculo para o processo de *Bulk Update* com mapeamento de *IDs* de senhorios e agentes

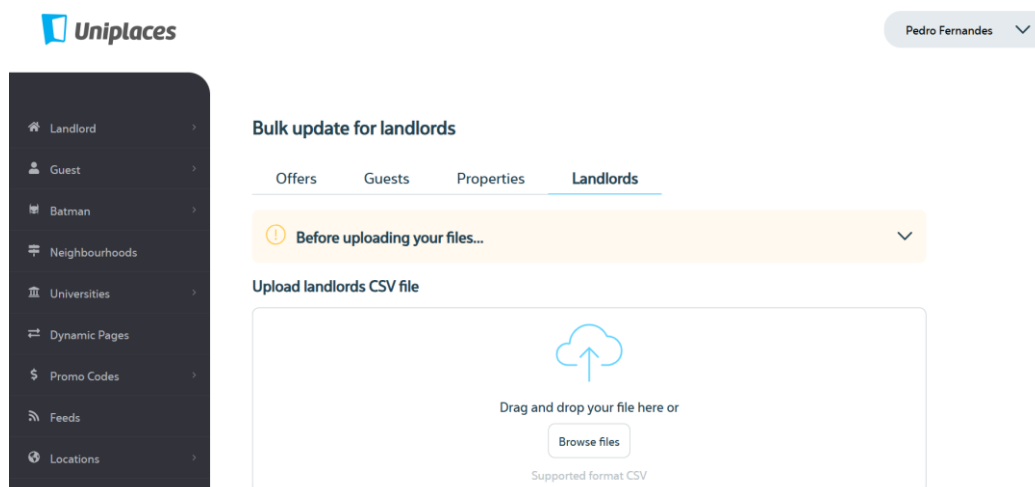


Figura 36 - Interface do sistema OPS para carregamento e atualização massiva de dados de senhorios (*Landlords*)

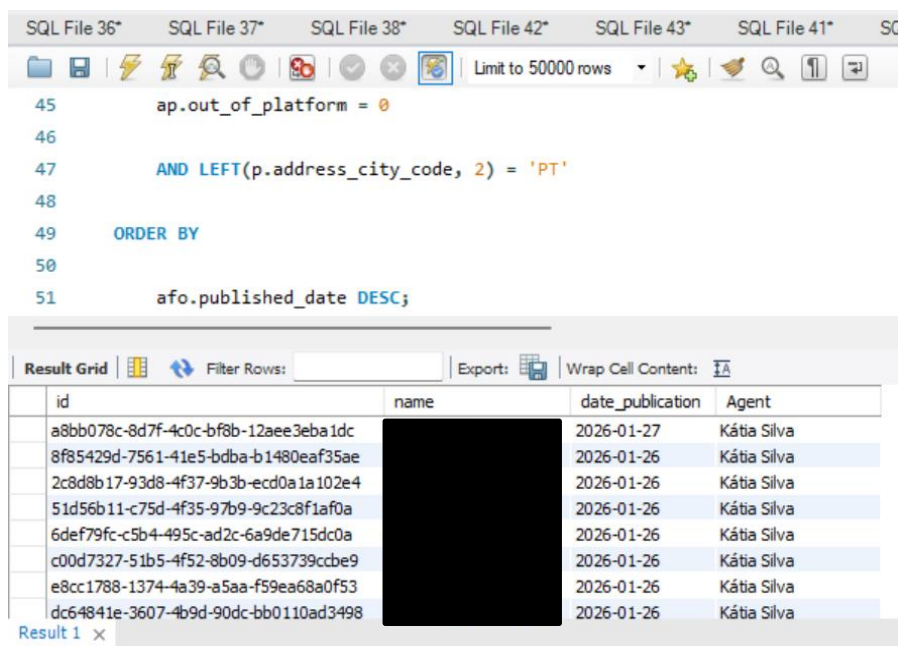


Figura 37 - Verificação pós-atualização via *MySQL* confirmando a atribuição correta dos agentes ao portefólio

A utilização de *SQL* foi a peça fundamental para o sucesso da tarefa, demonstrando por que razão é uma ferramenta indispensável no quotidiano de um analista de dados.

Em suma, o *SQL* funcionou como o "raio-x" do problema operacional, permitindo que uma falha invisível na interface do utilizador fosse detetada, quantificada e, posteriormente, validada após a correção.

Algumas semanas após a intervenção inicial de regularização de portefólios, surgiu um novo requisito operacional submetido pela responsável de *Account Management*, motivado pela necessidade de reequilibrar a carga de trabalho da equipa. Devido à sobrecarga de acompanhamento detetada em certos elementos, foi solicitada a transferência massiva de senhorios entre agentes específicos. Embora o objetivo fosse o reajuste de carteiras em vez da atribuição de perfis "órfãos", a arquitetura da solução revelou-se análoga ao processo de *bulk update* desenvolvido anteriormente, validando a escalabilidade e a reutilização do método criado.

A execução desta sub-tarefa exigiu uma manipulação precisa dos identificadores relacionais na base de dados. O primeiro passo consistiu em interrogar o esquema *prod_reporting* (especificamente a tabela *agency_staff*) através do *MySQL*, com o intuito de isolar o *ID* exato da agente recetora, neste caso prático, a agente Diana Santos (ver Figura 38).

1 •
2
3

```
SELECT id, first_name, last_name
FROM prod_reporting.agency_staff
WHERE first_name IN ('Diana');
```

Result Grid	Filter Rows:	Edit:
id	first_name	last_name
382a262e-c939-44a0-bd07-5fb21878c9dc	Diana	Santos
6fef8b62-da0d-4768-82dc-9494623f16ba	Diana	Pinto
718ffc9-3b7a-409f-9644-f15c68e0b1eb	Diana	Saraiva
d19ba62d-5773-4bdb-a5ec-2a86df3346df	Diana	Loureiro
NULL	NULL	NULL

Figura 38 - Consulta SQL ao esquema *agency_staff* para obtenção do identificador único (ID) da agente Diana Santos

Após a deteção do identificador, procedeu-se a uma nova consulta SQL complexa, cruzando tabelas para extrair a lista integral de senhorios (*Accommodation Providers*) atualmente associados à agente de origem, Juliana Silva (ver Figura 39). Esta extração rigorosa garantiu que a totalidade da carteira de clientes fosse identificada para transição, sem falhas de integridade.

8
9
10
11
12
13
14
15

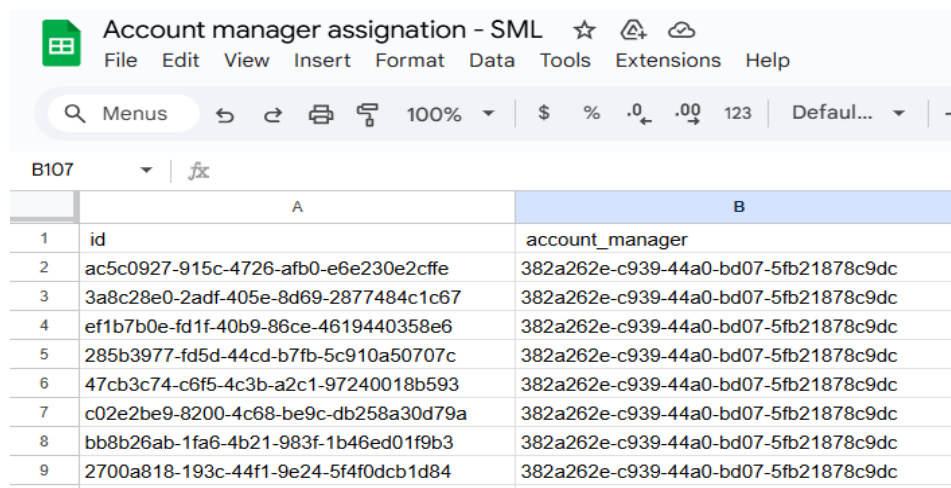
```
LEFT JOIN
prod_reporting.agency_staff AS ag
ON ag.id = ap.account_manager
WHERE
ap.account_manager = 'd600cb1c-5bf6-47ca-a945-eade33cecc2f'
AND ap.created_at >= '2025-01-01'
ORDER BY
ap.created_at DESC;
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
provider_id	provider_name	data_criacao	agent_name
ac5c0927-915c-4726-afb0-e6e230e2cffe		2025-09-30 22:30:00	Juliana Silva
3a8c28e0-2adf-405e-8d69-2877484c1c67		2025-09-18 10:15:04	Juliana Silva
ef1b7b0e-fd1f-40b9-86ce-4619440358e6		2025-09-17 08:29:09	Juliana Silva
285b3977-fd5d-44cd-b7fb-5c910a50707c		2025-09-11 19:05:38	Juliana Silva
47cb3c74-c6f5-4c3b-a2c1-97240018b593		2025-09-03 13:18:47	Juliana Silva
c02e2be9-8200-4c68-be9c-db258a30d79a		2025-08-16 22:01:34	Juliana Silva
bb8b26ab-1fa6-4b21-983f-1b46ed01f9b3		2025-08-04 17:26:47	Juliana Silva

Result 4 x

Figura 39 - Interrogação à base de dados para extração da listagem integral de senhorios associados à conta da agente Juliana Silva

Com os dados devidamente extraídos, os *IDs* dos senhorios foram transpostos para a aba de *Bulk Update* no *Google Sheets*. Nesta matriz de transição, a chave estrangeira referente ao agente foi substituída, emparelhando cada *ID* de senhorio com o novo identificador da agente recetora (ver Figura 40). A etapa de consolidação passou pela exportação deste ficheiro em formato *CSV* e pelo respetivo carregamento na plataforma interna OPS. Esta ação acionou a reescrita dos dados na base central da *Uniplaces*, transferindo o portefólio de forma automática e massiva.



	A	B
1	id	account_manager
2	ac5c0927-915c-4726-afb0-e6e230e2cffe	382a262e-c939-44a0-bd07-5fb21878c9dc
3	3a8c28e0-2adf-405e-8d69-2877484c1c67	382a262e-c939-44a0-bd07-5fb21878c9dc
4	ef1b7b0e-fd1f-40b9-86ce-4619440358e6	382a262e-c939-44a0-bd07-5fb21878c9dc
5	285b3977-fd5d-44cd-b7fb-5c910a50707c	382a262e-c939-44a0-bd07-5fb21878c9dc
6	47cb3c74-c6f5-4c3b-a2c1-97240018b593	382a262e-c939-44a0-bd07-5fb21878c9dc
7	c02e2be9-8200-4c68-be9c-db258a30d79a	382a262e-c939-44a0-bd07-5fb21878c9dc
8	bb8b26ab-1fa6-4b21-983f-1b46ed01f9b3	382a262e-c939-44a0-bd07-5fb21878c9dc
9	2700a818-193c-44f1-9e24-5f4f0dc1d84	382a262e-c939-44a0-bd07-5fb21878c9dc

Figura 40 - Estrutura do ficheiro de *Bulk Update* no *Google Sheets* com a substituição dos *IDs* para reequilíbrio de portefólio entre agentes

A eficiência deste procedimento técnico demonstrou o seu valor para a agilidade organizacional da empresa. O mesmo fluxo de trabalho foi posteriormente replicado noutras reestruturações de equipa, nomeadamente na transferência integral da carteira de senhorios da agente Jennifer para a agente Matilde Alcobia. Em suma, a capacidade de cruzar SQL com ferramentas de atualização massiva permitiu transformar um problema logístico e de recursos humanos (que manualmente demoraria dias a resolver), num processo de execução rápida e isento de erros, assegurando a continuidade do apoio prestado aos parceiros da plataforma.

3.3.7. Automatização da Extração de Avaliações do *Trustpilot* com Menção a Agentes

A presente secção descreve o desenvolvimento de um sistema automatizado de extração e organização das avaliações de cinco estrelas publicadas na plataforma *Trustpilot* em que os agentes da *Uniplaces* eram explicitamente mencionados pelos clientes. Esta tarefa surgiu de um pedido formal de BI submetido por Jennifer Aguiar, com urgência classificada como alta, conforme ilustrado na Figura 41.

[BI Request] 2025-10-21 - <Trustpilot Commission Plan - file needed>

Atualizado out. 21



Requester	@Jeniffer Aguiar
Team	Ops
Urgency	High
Status	To be filled by BI
Project Responsible	To be filled by BI
Time needed	To be filled by BI
Description	We need a file to gather the Trustpilot 5 stars reviews where agents are mentioned and to assign the 10€ bonus per review.

Hi team,

We'd like to request a file to track commissions based on reviews mentioning Uniplaces agents. Here's some information:

1. Data Scope:

- Include all reviews where an agent is mentioned.
- Include reviews from all dates, with an option to filter by date range.
- Include both Automatic Invitation, Basic Invitation, Organic, and Generated Link sources.

2. Required Columns (exact order):

- Review Date – Date of the review.
- Agent Name – Name of the agent mentioned in the review.
- Reviewer Name – Customer who left the review.
- Review Comment – Only the sentence mentioning the agent, maximum 15 words, enclosed in quotation marks, in the original language.
- Bonus (€) – Always €10.
- Source – Review type/source.

Figura 41 - Pedido formal de BI relativo à extração de avaliações do *Trustpilot* com menção a agentes

Esta atividade enquadrou-se também no domínio do *Text Mining*, uma vez que implicou a extração e o processamento automático de texto não estruturado proveniente de avaliações da *Trustpilot*. Em concreto, o objetivo consistiu em identificar, no conteúdo textual dos comentários, menções nominais a agentes de apoio ao cliente, recorrendo a regras de correspondência textual baseadas num dicionário de variantes ortográficas e a uma função de deteção designada *findMentionedAgent*.

O objetivo central consistia em disponibilizar, para cada mês, uma folha de cálculo estruturada que permitisse à responsável pela equipa de agentes monitorizar a frequência com que cada agente era mencionado em avaliações positivas, servindo igualmente de base para o cálculo de um bônus de 10 euros por cada avaliação de cinco estrelas em que o agente fosse identificado.

Até à data de início desta tarefa, o processo de recolha destas avaliações era realizado de forma manual, com inserção direta dos dados numa folha de cálculo, conforme visível na Figura 42, o que tornava o processo moroso, suscetível a erros humanos e de difícil escalabilidade. A automatização deste fluxo de trabalho representou, por isso, uma melhoria operacional significativa para a equipa.

	Review Date	Agent Name	Reviewer Name	Review Comment	Bonus (€)	Source	Agent	Reviews	Bonus	Duplicated	Bonus	Total Bonus
4	2025-10-01	Pablo		"Estoy muy contento con Pablo, mi asesor"	€10	Automatic Invitation	Adriana	19	€190.00	0	€0.00	€190.00
5	2025-10-10	Mattia		"Ormai Mattia è il mio punto di riferimento"	€10	Basic Invitation	Claudia	4	€40.00	0	€0.00	€40.00
6	2025-10-10	Pablo		"Pablo nuestro ejetuivo muy atento"	€10	Basic Invitation	Diana	7	€70.00	0	€0.00	€70.00
7	2025-10-10	Pablo		"son Pablo in particular, is fantastico"	€10	Basic Invitation	Federico	5	€50.00	0	€0.00	€50.00
8	2025-10-10	Pablo		"Pablo siempre atende al momento las comunicaciones"	€10	Basic Invitation	Filippo	1	€10.00	0	€0.00	€10.00
9	2025-10-10	Pablo		"I would like to highlight Pablo, who has been professional"	€10	Basic Invitation	Gabriela	4	€40.00	0	€0.00	€40.00
10	2025-10-10	Mattia		"In account, Mattia is amazing"	€10	Basic Invitation	Joana	3	€30.00	0	€0.00	€30.00
11	2025-10-10	Pablo		"We really appreciate your dedication Pablo"	€10	Basic Invitation	Juliana	1	€10.00	0	€0.00	€10.00
12	2025-10-10	Mattia		"son mattia ohe è gentilissimo e professionale"	€10	Basic Invitation	Kátia	1	€10.00	0	€0.00	€10.00
13	2025-10-10	Lorey		"Lorey es única"	€10	Basic Invitation	Leticia	6	€60.00	0	€0.00	€60.00
14	2025-10-10	Mattia		"nel mio caso MATTIA - lui è il mio angelo custode"	€10	Basic Invitation	Lorey	49	€490.00	1	-€10.00	€480.00
15	2025-10-10	Pablo		"quiero destacar especialmente el trabajo de Pablo"	€10	Organic	Mattia	10	€100.00	0	€0.00	€100.00
16	2025-10-10	Pablo		"Pablo nos ha ayudado demasiado"	€10	Automatic Invitation	Pablo	32	€320.00	0	€0.00	€320.00
17	2025-10-10	Pablo		"Muchas gracias a Pablo por estar siempre al pendiente"	€10	Basic Invitation	Paula	9	€90.00	1	-€10.00	€80.00
18	2025-10-10	Joana		"Thank you to Joana"	€10	Basic Invitation	Rodrigo	2	€20.00	0	€0.00	€20.00
19	2025-10-10	Lorey		"ayuda siempre atenta y eficaz de Lorey"	€10	Basic Invitation	Sara	1	€10.00	0	€0.00	€10.00
20	2025-10-10	Lorey		"Well i will say Lorey makes things easy"	€10	Basic Invitation	Victor	2	€20.00	0	€0.00	€20.00
21	2025-10-10	Pablo		"Queremos destacar especialmente a Pablo"	€10	Basic Invitation						
22	2025-10-10	Mattia		"son Mattia, sempre disponibile e professionale"	€10	Basic Invitation						
23	2025-10-10	Paula		"Paula la señoita peña atendida"	€10	Basic Invitation						
24	2025-10-10	Adriana		"Adriana, who took care of me"	€10	Basic Invitation						
25	2025-10-11	Joana		"Joana foi excelente"	€10	Basic Invitation						
26	2025-10-13	Pablo		"mencionó a Pablo, es una persona súper atenta"	€10	Basic Invitation						
27	2025-10-13	Pablo		"profesionalidad de nuestro gestor Pablo y su eficiencia"	€10	Basic Invitation						
28	2025-10-13	Mattia		"especialmente Mattia sempre disponibile"	€10	Basic Invitation						
29	2025-10-13	Kátia		"Excelente atendimento, da Kátia"	€10	Automatic Invitation						
30	2025-10-13	Pablo		"profesionalidad de PABLO, que está disponible"	€10	Basic Invitation						
31	2025-10-13	Mattia		"Servizio ottimo. Seguiti da Mattia"	€10	Basic Invitation						
32	2025-10-13	Mattia		"son polite and pro-active Mattia"	€10	Basic Invitation						
33	2025-10-13	Diana		"Diana is extremely professional"	€10	Basic Invitation						
34	2025-10-13	Adriana		"Thank Adriana for her prompt services"	€10	Basic Invitation						
35	2025-10-13	Diana		"falar com a Sra. D. Diana"	€10	Basic Invitation						
36	2025-10-13	Mattia		"Mattia: il mio manager su Uniplaces"	€10	Basic Invitation						
37	2025-10-13	Lorey		"siempre cuento con la ayuda inestimable de Lorey"	€10	Basic Invitation						
38	2025-10-14	Pablo		"Pablo es un gran gestor, siempre está pendiente."	€10	Basic Invitation						
39	2025-10-14	Paula		"VALORO MUY BIEN A PAULA DE UNIPLACES"	€10	Basic Invitation						
40	2025-10-14	Lorey		"Mi comunicación con Lorey ha sido siempre fluida"	€10	Basic Invitation						
41	2025-10-14	Lorey		"I have been supervised and helped by my account manager Lorey."	€10	Basic Invitation						
42	2025-10-14	Lorey		"especialmente Lorey muy buen trato, estoy muy contenta"	€10	Basic Invitation						
43	2025-10-14	Federico		"I have been in contact with Federico who has dealt with"	€10	Basic Invitation						

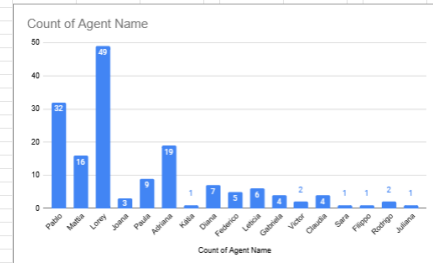


Figura 42 - Ficheiro de inserção manual de avaliações utilizado anteriormente ao desenvolvimento da solução automatizada

O processo de desenvolvimento teve início com o estudo da *Trustpilot* API, cuja documentação fornecia as instruções necessárias para a realização de pedidos de extração de dados. A utilização de API como mecanismo de acesso programático a dados externos constitui uma prática consolidada no desenvolvimento de pipelines de dados modernos, permitindo às organizações integrar fontes de informação heterogêneas de forma estruturada, segura e escalável [67]. Para aceder aos dados da página da *Uniplaces* no *Trustpilot*, foi necessário obter o identificador único da empresa na plataforma (*business unit ID*) bem como uma chave de autenticação (*API key*), que foram incorporados nos pedidos HTTP efetuados ao *endpoint* da API.

A fase exploratória foi desenvolvida em *Python*, através do ambiente de desenvolvimento *Visual Studio Code*, com recurso às bibliotecas *requests* e *json*, conforme ilustrado na Figura 43. O *script* inicial realizava um pedido *GET* ao *endpoint* com os parâmetros de número de avaliações a extrair e língua de filtragem, e validava a resposta recebida através do código de estado HTTP, imprimindo o conteúdo em formato JSON (*JavaScript Object Notation*) em caso de sucesso ou uma mensagem de erro em caso de falha.

O resultado desta fase exploratória, visível na Figura 44, consistiu num *DataFrame* estruturado em *Pandas* que agregava as avaliações extraídas com os campos necessários: data da avaliação, *username* que comentou, comentário, bônus atribuído e fonte da avaliação.

Depois do trabalho desenvolvido em *Visual Studio Code*, o trabalho prosseguiu para a fase de implementar no *Google Apps Script* (ver Figura 45).

```

EXPLORER
TRUSTPILOT
API curl.ipynb
API_trustpilot.ipynb
trustpilot_5star_reviews.csv
trustpilot_reviews_5stars_bonus.csv
trustpilot_reviews.csv
trustpilot_uniplaces_5star_agent_reviews.csv
trustpilot_uniplaces_5star_reviews_copilot.csv

API curl.ipynb > import requests
Generate + Code + Markdown | ▶ Run All | Clear All Outputs | Outline ... Python 3.14.0

import requests
import json

# Trustpilot API endpoint
url = "https://api.trustpilot.com/v1/reviews/latest"

# Query parameters
params = {
    "count": 3,      # Number of reviews to fetch
    "language": "en" # Language of the reviews
}

# HTTP headers (replace with your actual API key)
headers = {
    "apikey": "apikey"
}

# Send GET request
response = requests.get(url, headers=headers, params=params)

# Check the result
if response.status_code == 200:
    data = response.json()
    print("✅ Success! Here are the latest reviews:\n")
    print(json.dumps(data, indent=2)) # Pretty-print JSON response
else:
    print(f"❌ Error {response.status_code}: {response.text}")
  
```

Figura 43 - Script inicial de ligação à *Trustpilot* API desenvolvido em *Python* no *Visual Studio Code*

... ✅ 5-star reviews fetched and converted to DataFrame:

	Review Date	Agent Name	Reviewer Name \
0	2025-11-06T07:09:28Z		
1	2025-11-05T19:16:24Z		
2	2025-11-04T20:33:04Z		
3	2025-11-04T18:37:31Z		
4	2025-11-04T14:34:06Z		
5	2025-11-04T13:49:43Z		
6	2025-11-04T12:40:20Z		
7	2025-11-04T11:18:47Z		
8	2025-11-03T23:59:48Z		
9	2025-11-03T09:37:18Z		
10	2025-11-02T18:17:49Z		
11	2025-11-01T08:08:43Z		
12	2025-10-31T22:27:10Z		
13	2025-10-31T22:00:56Z		
14	2025-10-31T21:05:40Z		
15	2025-10-31T18:58:06Z		
16	2025-10-31T15:53:29Z		
17	2025-10-31T15:48:05Z		
18	2025-10-30T22:38:01Z		
19	2025-10-30T22:06:40Z		

	Review Comment	Bonus (€)	Source
...			
18	Fast responses, clear instructions and kind se...	10	

Figura 44 - Resultado da extração de avaliações de cinco estrelas convertido em *DataFrame* através da biblioteca *Pandas*

Um dos desafios de maior complexidade técnica prendeu-se com a necessidade de garantir que o sistema era capaz de identificar corretamente o nome do agente mencionado no texto da avaliação, independentemente da forma como esse nome era escrito pelo cliente. Com efeito, verificou-se que os clientes recorriam frequentemente a variações ortográficas dos nomes dos agentes, omitindo acentuação, utilizando letras minúsculas ou referindo apenas o apelido. Para

resolver esta problemática, foi construído um dicionário de variantes por agente, no qual cada nome era associado a um conjunto de formas alternativas aceites pelo sistema. O exemplo ilustrado na Figura 46 demonstra esta lógica aplicada ao agente de nome Flávio, cujas variantes consideradas incluíam "flávio", "flavio", "Miller", "miler", "Miler" e "miller".

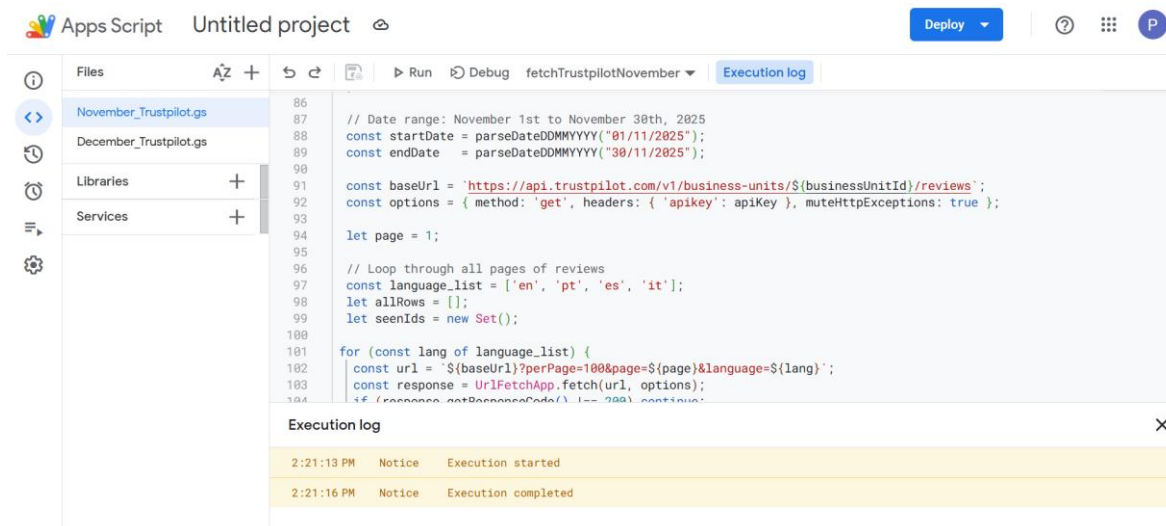


Figura 45 - Script desenvolvido em *Google Apps Script* para extração automatizada das avaliações do *Trustpilot* com definição do intervalo temporal mensal

"Flávio": ["flávio", "flavio", "Miller", "miler", "Miler", "miller"],

Figura 46 - Dicionário de variantes ortográficas dos nomes dos agentes implementado no *script* de identificação

A função responsável pela identificação do agente no texto da avaliação, designada *findMentionedAgent* e visível na Figura 47, operava convertendo o texto do comentário para letras minúsculas e verificando, para cada agente registrado no dicionário, se alguma das suas variantes estava contida no texto. Em caso de correspondência positiva, a função retornava o nome canónico do agente; caso contrário, retornava um valor nulo, excluindo a avaliação do relatório final.

```
// Detect which agent is mentioned in the review comment
function findMentionedAgent(comment) {
  const lowerComment = comment.toLowerCase();
  for (const agent in agentVariants) {
    if (agentVariants[agent].some(v => lowerComment.includes(v))) return agent;
  }
  return null;
}
```

Figura 47 - Função *findMentionedAgent* responsável pela identificação do agente mencionado no texto da avaliação

Um segundo desafio identificado no decurso do desenvolvimento foi a constatação de que o volume de dados extraídos era significativamente inferior ao esperado. Após análise, concluiu-se que a causa residia na configuração inicial do parâmetro de língua, que se

encontrava restrito ao inglês, excluindo assim todas as avaliações redigidas em português, espanhol e italiano, línguas prevalentes nos mercados prioritários da *Uniplaces*. A solução implementada consistiu na definição de uma lista de línguas a percorrer iterativamente, composta pelos códigos *en*, *pt*, *es* e *it*, conforme visível na Figura 48. Para cada língua da lista, era construído um URL de pedido específico e os resultados obtidos eram agregados numa lista global, com recurso a um conjunto de identificadores únicos já processados (*seenIds*) para evitar a duplicação de registos que pudessem ser devolvidos em múltiplas línguas.

```
// Loop through all pages of reviews
const language_list = ['en', 'pt', 'es', 'it'];
let allRows = [];
let seenIds = new Set();

for (const lang of language_list) {
  const url = `${baseUrl}?perPage=100&page=${page}&language=${lang}`;
  const response = UrlFetchApp.fetch(url, options);
  if (response.getResponseCode() !== 200) continue;
}
```

Figura 48 - Implementação da iteração multilingue sobre as avaliações extraídas da *Trustpilot* API

O resultado final do sistema, ilustrado na Figura 49 e na Figura 50, consistiu numa folha de cálculo estruturada com as colunas definidas no pedido original: data da avaliação, nome do agente mencionado, *username* de quem comentou, excerto do comentário com menção ao agente, bónus atribuído em euros e fonte da avaliação. A folha incluía ainda, do lado direito, uma tabela de consolidação por agente com o total de avaliações e o bónus acumulado, bem como um gráfico de barras que permitia visualizar de forma imediata a distribuição das menções pelos diferentes agentes ao longo do mês.

Review Date	Agent Name	Reviewer Name	Review Comment	Bonus (€)	Source	Agent	Reviews
10/11/2025	Leticia		I have just booked an apartment in Milan for my son. So far everything is great, although we haven't seen	€10	Basic Invitation	Adriana Jesus	3
07/11/2025	Adriana		Adriana was amazing, answered all my questions in a second!	€10	Basic Invitation	Ana Oliveira	0
07/11/2025	Leticia		UniPlaces is such a good website if you want to book a room. Whether it is a university student or a worki	€10	Automatic Invitation	Ana Guedes	0
07/11/2025	Leticia		It was a good experience especially Leticia as she helped me alot during recommendations, I am thankful	€10	Basic Invitation	Andrea Pesaresi	2
07/11/2025	Victor		I was adviced by Victor who contacted me and really helped me find a suitable room for my stay.	€10	Basic Invitation	Benian Bartals	0
07/11/2025	Gabriela		He is really friendly and responsivel	€10	Basic Invitation	Claudia Caldarola	0
			Gabriela has been a great help. She responded quickly to my questions, gave me the necessary informat	€10	Basic Invitation		
			Manager Adriana was very patient and helped for me a lot			Davide Levi	0
05/11/2025	Adriana		Supported	€10	Basic Invitation		
04/11/2025	Lorey		Highly recommend ❤️	€10	Basic Invitation		
04/11/2025	Lorey		My experience with uniplaces was wonderful especially with Lorey that personally kept tabs on the proper	€10	Basic Invitation	Diana Santos	2
04/11/2025	Adriana		Adriana was incredibly helpful and friendly. There were some technical issues and she immediately conta	€10	Basic Invitation	Edgar Parreira	0
04/11/2025	Lidia		What an amazing and friendly service	€10	Basic Invitation	Fausto Lelles	0
04/11/2025	Gabriela		Gabriela was so helpful and nice , responded to all my requests correctly and efficiently thank you	€10	Basic Invitation	Filippo Casamassima	0
03/11/2025	Juliana		When I needed some help on reactivating one ancient announcement and adapting it to my present cond	€10	Basic Invitation	Flávio Miller	0

Figura 49 - Resultado final do relatório mensal de avaliações com menção a agentes no *Google Sheets*

A opção pelo gráfico de barras verticais justifica-se pela sua adequação à comparação de valores discretos entre categorias, neste caso o número de menções por agente, permitindo uma leitura intuitiva das diferenças de desempenho entre os membros da equipa.

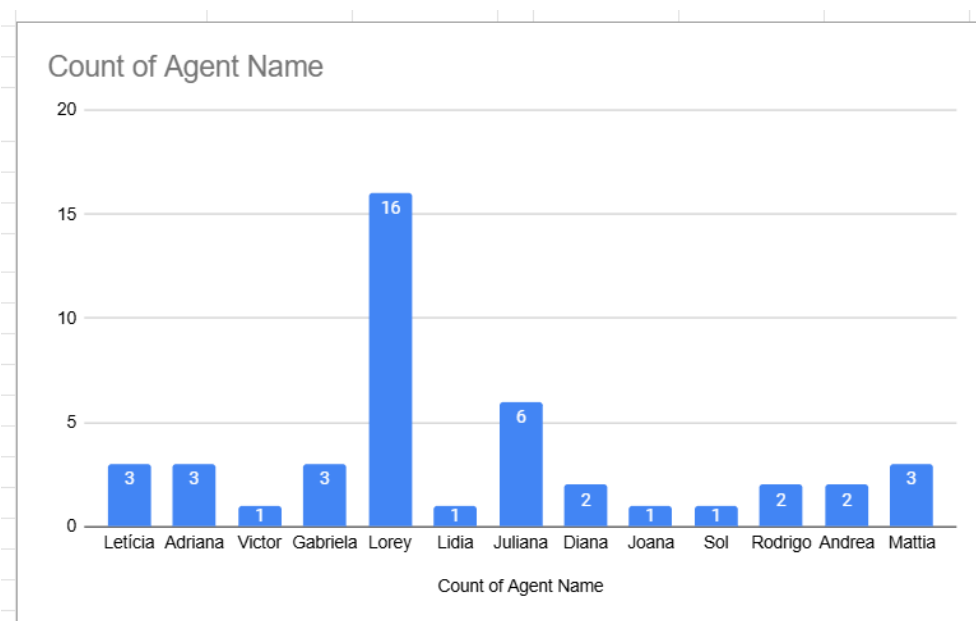


Figura 50 - Gráfico de barras com a contagem de menções por agente no período analisado

O trabalho foi concluído com a configuração de folhas automatizadas para os meses subsequentes, cobrindo o período até março de 2026, garantindo assim a continuidade do processo sem necessidade de intervenção técnica adicional. Esta tarefa assumiu particular relevância formativa, na medida em que representou o primeiro contacto com a integração de uma API externa no âmbito do estágio, exigindo a articulação de conhecimentos em duas linguagens de programação distintas, a compreensão do funcionamento de pedidos HTTP autenticados e a resolução iterativa de problemas técnicos identificados ao longo do desenvolvimento, desde a gestão de variantes ortográficas até à cobertura multilingue das avaliações extraídas.

3.3.8. Desenvolvimento de um Modelo de Previsão de Preços

O projeto culminante desta tarefa incidiu no desenvolvimento de um modelo de *Machine Learning* concebido para prever o preço justo de mercado dos imóveis listados na plataforma da *Uniplaces*. A génese desta iniciativa baseou-se na identificação de uma ineficiência estrutural do negócio: a definição arbitrária de preços por parte dos senhorios. Valores excessivamente altos (*Overpriced*) geram um excesso de inventário estático que não converte e consome recursos da plataforma, enquanto preços artificialmente baixos (*Underpriced*) traduzem-se numa perda direta de margem comissionável para a empresa. Para mitigar este problema, a solução passou pela criação de um algoritmo inteligente capaz de avaliar as características do imóvel e sugerir um intervalo de preço otimizado, maximizando tanto a taxa de conversão como o retorno financeiro.

A criação do algoritmo exigiu um pré-processamento rigoroso, recorrendo a um *dataset* histórico de aproximadamente 59.000 reservas confirmadas. Para garantir que o modelo refletisse a realidade económica atual, aplicou-se um filtro temporal estrito, confinando o treino apenas a dados posteriores a 2024. Esta decisão foi crucial para isolar o ruído provocado pela elevada inflação e pela crise habitacional europeia, que tornaram os preços de 2022 e 2023 obsoletos. Do ponto de vista da engenharia de dados, desenvolveu-se um algoritmo de extração baseada em expressões regulares para processar as comodidades dos imóveis, que se encontravam encapsuladas em formatos JSON complexos. Adicionalmente, durante a Análise Exploratória de Dados (EDA), detetou-se que comodidades essenciais como "Internet" e "TV" apresentavam taxas de preenchimento residuais. Concluiu-se que estas *features* são atualmente assumidas como dados adquiridos pelos inquilinos, sendo frequentemente omitidas pelos senhorios. Para evitar um enviesamento prejudicial, estas variáveis foram excluídas do modelo de treino.

Após sucessivas iterações, o algoritmo estabilizou com uma Margem de Erro Absoluto Médio (MAE) na ordem dos 133 euros. Face à elevada volatilidade do mercado imobiliário e à ausência de variáveis qualitativas determinantes como a área exata em metros quadrados ou a qualidade fotográfica do anúncio. Esta margem foi validada como altamente realista, fornecendo um intervalo de mercado fiável e operacional. Com o intuito de tangibilizar a componente analítica, desenvolveu-se uma aplicação web interativa através da biblioteca *Streamlit*, concebida como uma Prova de Conceito (*Proof of Concept*).

Conforme ilustrado na Figura 51 e na Figura 52, este simulador permite a inserção manual das características de um imóvel, devolvendo instantaneamente a estimativa do preço justo.

The image shows a web application interface for estimating rental prices. At the top, a welcome message in Portuguese says: "Bem-vindo(a) ao avaliador de rendas da Uniplaces. Insira os detalhes do imóvel abaixo para obter a estimativa do Preço Justo de mercado." Below this, the interface is divided into two main sections: "Detalhes Principais" (Main Details) and "Comodidades Premium" (Premium Amenities). The "Detalhes Principais" section contains four dropdown menus: "Cidade" (set to "PT-lisbon"), "Tipo de Oferta" (set to "bedroom"), "Tipo de Quarto" (set to "single"), and "Tipo de Senhorio (Gestão)" (set to "SML"). The "Comodidades Premium" section contains a list of amenities with checkboxes: "Cozinha Equipada" (checked), "Máquina de Lavar/Secar" (unchecked), "Ar Condicionado" (checked), "Varanda/Terraço" (checked), "Elevador" (unchecked), "Segurança/Vigilância" (unchecked), and "Equipa de Suporte/Manutenção" (unchecked).

Figura 51 - Interface de demonstração (*Proof of Concept*) desenvolvida em *Streamlit* para a inserção manual de características e comodidades do imóvel

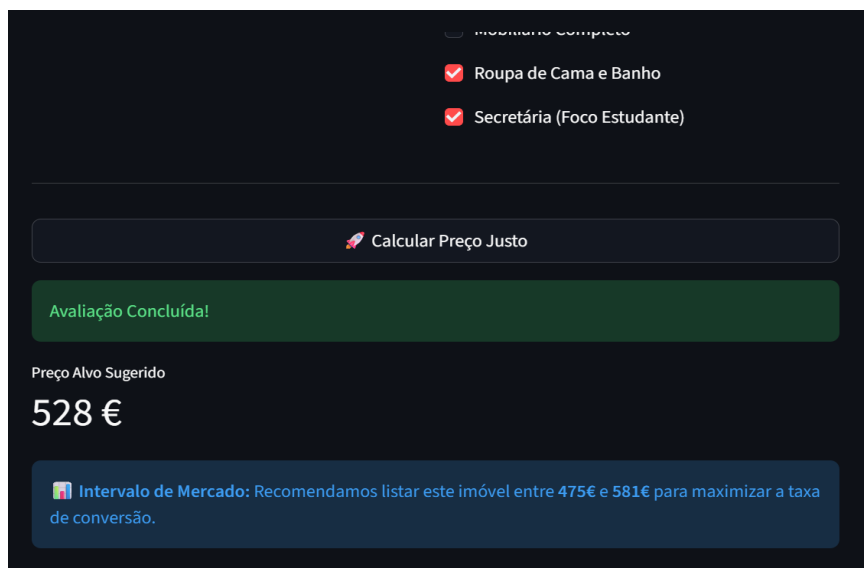


Figura 52 - Output do modelo de *Machine Learning* com a apresentação do preço alvo sugerido e o respetivo intervalo de mercado para otimização da conversão

Esta interface revelou-se um instrumento crucial para demonstrar à gestão que a solução não era um mero exercício teórico de *Data Science*, possuindo potencial para ser integrada no quotidiano da equipa comercial como ferramenta de simulação e apoio à negociação de rendas.

A validação do modelo materializou-se através de uma auditoria financeira às reservas da zona euro em 2025, período em que a *Uniplaces* transacionou mais de seis milhões de euros em cerca de 13.000 reservas. A classificação destas transações pela Inteligência Artificial (ver Figura 53) demonstrou um forte alinhamento com o comportamento do consumidor: a esmagadora maioria (11.607) das reservas ocorreu em imóveis com preços justos. Em contrapartida, registaram-se apenas 660 reservas em propriedades classificadas como *Overpriced*, confirmando empiricamente que os estudantes rejeitam rendas inflacionadas. Nas 825 casas *Underpriced*, evidenciou-se a perda de capital direto, uma vez que a comissão da empresa incidia sobre valores artificialmente abaixo do mercado, deixando margem financeira por rentabilizar em transações que ocorreriam de qualquer forma.

Ciente das limitações inerentes à avaliação imobiliária automatizada, a recomendação estratégica do projeto não passou pela alteração impositiva de preços, mas sim pela adoção de uma abordagem *Human-in-the-Loop*. O modelo funcionaria como um radar interno, gerando relatórios periódicos para capacitar os *Account Managers* com dados concretos durante a gestão de portefólio. Para sustentar esta visão perante a gestão, analisou-se o inventário recente de 2026, composto por 7.474 novos anúncios. A distribuição resultante (ver Figura 54) expôs a ineficiência latente do *stock* atual: 2.872 imóveis encontram-se listados com valores mais de 20% acima do recomendável, representando "inventário fantasma" com conversão quase nula,

enquanto 980 propriedades subvalorizadas geram lucro cessante imediato. Esta auditoria provou que a ausência de uma monitorização inteligente resulta em perdas financeiras diárias, justificando a adoção da ferramenta como um mecanismo de recuperação de receita.

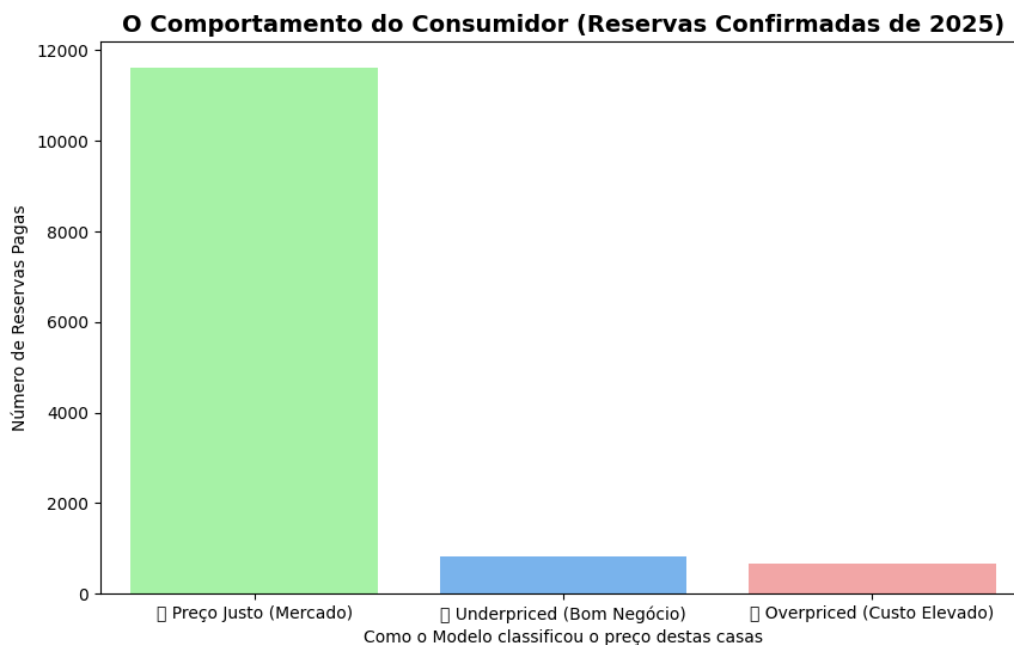


Figura 53 - Auditoria ao comportamento do consumidor em 2025

Distribuição da Eficiência de Preços na Uniplaces

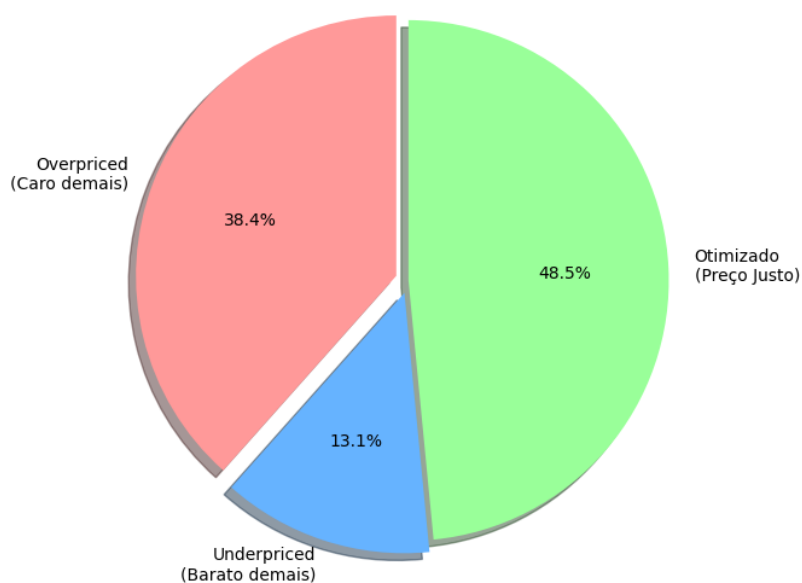


Figura 54 - Distribuição da eficiência de preços no inventário de 2026

O projeto culminou com uma apresentação formal à CFO da organização, Catarina Coelho, englobando a fundamentação teórica, o impacto financeiro e uma simulação em tempo real. Embora a receptividade tenha sido francamente positiva, o enquadramento macroeconómico e

as prioridades do departamento de Produto ditaram o adiamento da sua implementação. A quebra geral na taxa de conversão enfrentada pela empresa nesse período tornou desaconselhável a sugestão de intervenção nos imóveis *Underpriced*, uma vez que qualquer inflacionamento de preços poderia retrainir ainda mais a procura. Este projeto em concreto proporcionou uma aprendizagem inestimável sobre a realidade corporativa. Compreendeu-se que o desenvolvimento tecnológico tem de estar invariavelmente alinhado com o momento estratégico do negócio e que o adiamento de boas ideias é uma contingência natural em ambientes empresariais dinâmicos. O percurso efetuado, desde a limpeza de dados brutos até à defesa de um *business case* estruturado perante a administração, revelou-se uma experiência de maturação profissional e técnica extremamente gratificante.

3.3.9. Reestruturação e Automação da Matriz de Análise de Cancelamentos em Tableau

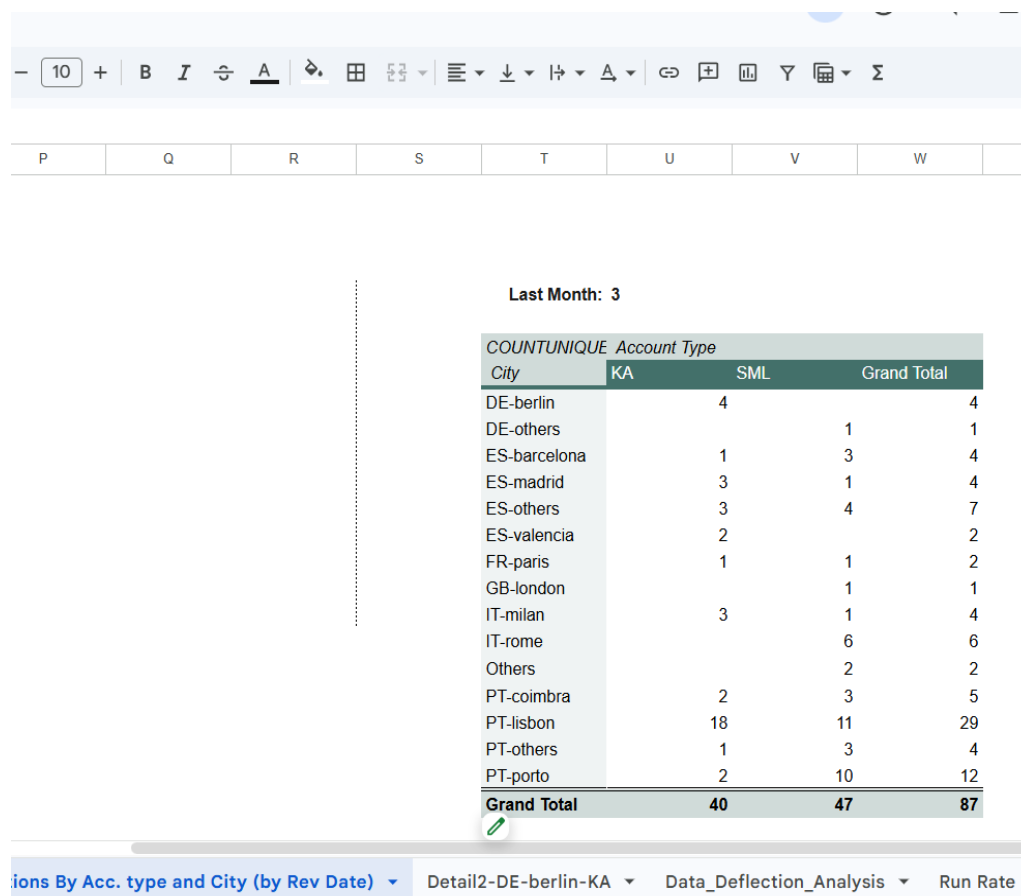
Na sequência das iniciativas de otimização operacional, foi reestruturada e automatizada a matriz de análise de cancelamentos. O processo analítico anterior assentava num modelo baseado em folhas de cálculo, cujo funcionamento exigia extrações e atualizações manuais recorrentes. Esta abordagem tradicional, além de consumir tempo útil e representar um esforço repetitivo para a equipa, apresentava uma elevada suscetibilidade a erros operacionais, nomeadamente através da sobreposição indevida de dados transacionais ou da aplicação inconsistente de filtros ocultos na folha de cálculo.

A dependência de processos manuais de atualização de dados em contextos analíticos é amplamente reconhecida como uma fonte crítica de degradação da qualidade da informação disponibilizada às equipas de gestão, constituindo um dos principais obstáculos à adoção de uma cultura organizacional verdadeiramente orientada para os dados [68]. A motivação central para esta intervenção prendeu-se, assim, com a urgência de mitigar ineficiências operacionais e transitar o modelo de reporte para um ambiente analítico centralizado, robusto e governado, utilizando o *Tableau*.

A intervenção técnica focou-se primariamente na evolução da granularidade temporal e na exatidão do processamento de dados. O relatório original consistia numa vista estática semanal, visível na Figura 55, que limitava a análise de tendências de médio e longo prazo.

No ambiente de desenvolvimento do *Tableau* (ver Figura 56), a arquitetura foi redesenhada para suportar um modelo dinâmico com granularidade mensal. Para o efeito, otimizou-se a lógica de agregação da base de dados, empregando funções de contagem distinta,

especificamente o cálculo CNTD (*count_distinct*) sobre o campo *booking_id* para garantir o apuramento exato dos eventos de cancelamento, eliminando o risco de duplicação de registos. Esta nova infraestrutura foi suportada pela implementação de filtros temporais hierárquicos (como o ano e mês de faturação) e segmentações parametrizadas por tipo de conta (*Key Accounts* - KA e *Small/Medium Landlords* - SML) e localização geográfica (*city*). Este rigor técnico assegurou a total integridade dos dados, replicando as regras de negócio originais numa interface analítica livre de ruído visual.



Formula bar: =10

Columns: P, Q, R, S, T, U, V, W

Filter: Last Month: 3

COUNTUNIQUE City	Account Type		Grand Total
	KA	SML	
DE-berlin	4		4
DE-others		1	1
ES-barcelona	1	3	4
ES-madrid	3	1	4
ES-others	3	4	7
ES-valencia	2		2
FR-paris	1	1	2
GB-london		1	1
IT-milan	3	1	4
IT-rome		6	6
Others		2	2
PT-coimbra	2	3	5
PT-lisbon	18	11	29
PT-others	1	3	4
PT-porto	2	10	12
Grand Total	40	47	87

Filters: By Acc. type and City (by Rev Date) | Detail2-DE-berlin-KA | Data_Deflection_Analysis | Run Rate

Figura 55 - Matriz estática semanal em folha de cálculo

O produto final desta migração materializou-se num *dashboard* interativo e limpo, ilustrado na Figura 57.

À medida que o volume de reservas e transações aumenta, a infraestrutura desenvolvida suporta o crescimento contínuo dos dados sem qualquer degradação de performance ou necessidade de reconfigurações manuais. A implementação de sistemas de reporte automatizados e escaláveis que eliminam a necessidade de intervenção humana recorrente constitui uma das práticas mais recomendadas no desenvolvimento de pipelines de dados em contexto empresarial, na medida em que a acumulação de decisões técnicas de curto prazo pode

gerar uma dívida técnica significativa que compromete a manutenibilidade e a fiabilidade dos sistemas a longo prazo [69]. Olhando para o futuro da arquitetura de dados da *Uniplaces*, esta automação representa um passo fundamental para consolidar uma "única fonte de verdade" (*Single Source of Truth*) dentro do ecossistema da empresa.

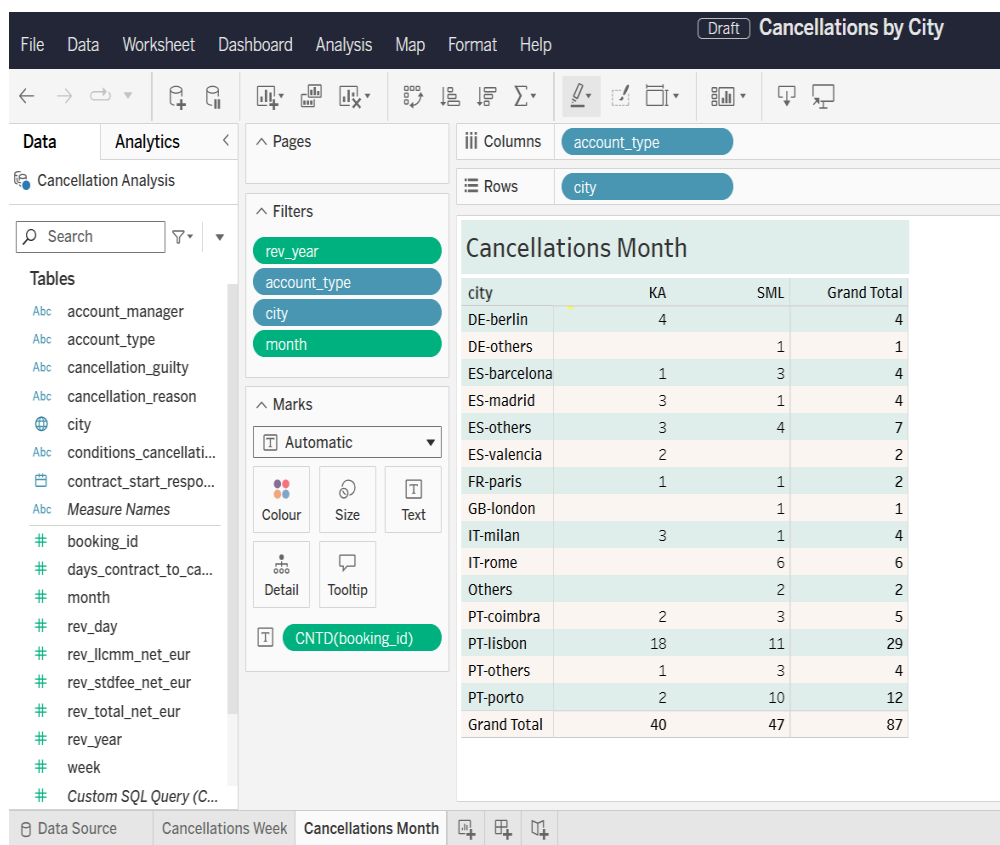


Figura 56 - Ambiente de desenvolvimento e agregação de dados no *Tableau*

Cancellations by City			
city	KA	SML	Grand Total
DE-berlin	3		3
ES-barcelona		1	1
ES-madrid	2		2
ES-others	1	1	2
IT-rome		1	1
PT-coimbra		2	2
PT-lisbon	5	2	7
PT-others		1	1
PT-porto		1	1
Grand Total	11	9	20

Figura 57 - Visualização final do *Dashboard* mensal de cancelamentos

Mais importante ainda do ponto de vista organizacional, a nova solução democratiza o acesso à informação. A gestão e os diversos *stakeholders* passam a deter total autonomia para extraírem *insights* históricos e analisarem tendências mensais de cancelamento de forma

instantânea e intuitiva. Consequentemente, esta transição tecnológica liberta as equipas analíticas e técnicas das tarefas rotineiras de manutenção de relatórios, permitindo a realocação de tempo e recursos para o desenvolvimento de análises preditivas e projetos de maior valor acrescentado para a estratégia da empresa.

3.3.10. Análise de Fatores de Conversão de Ofertas e Modelo Preditivo de Sucesso

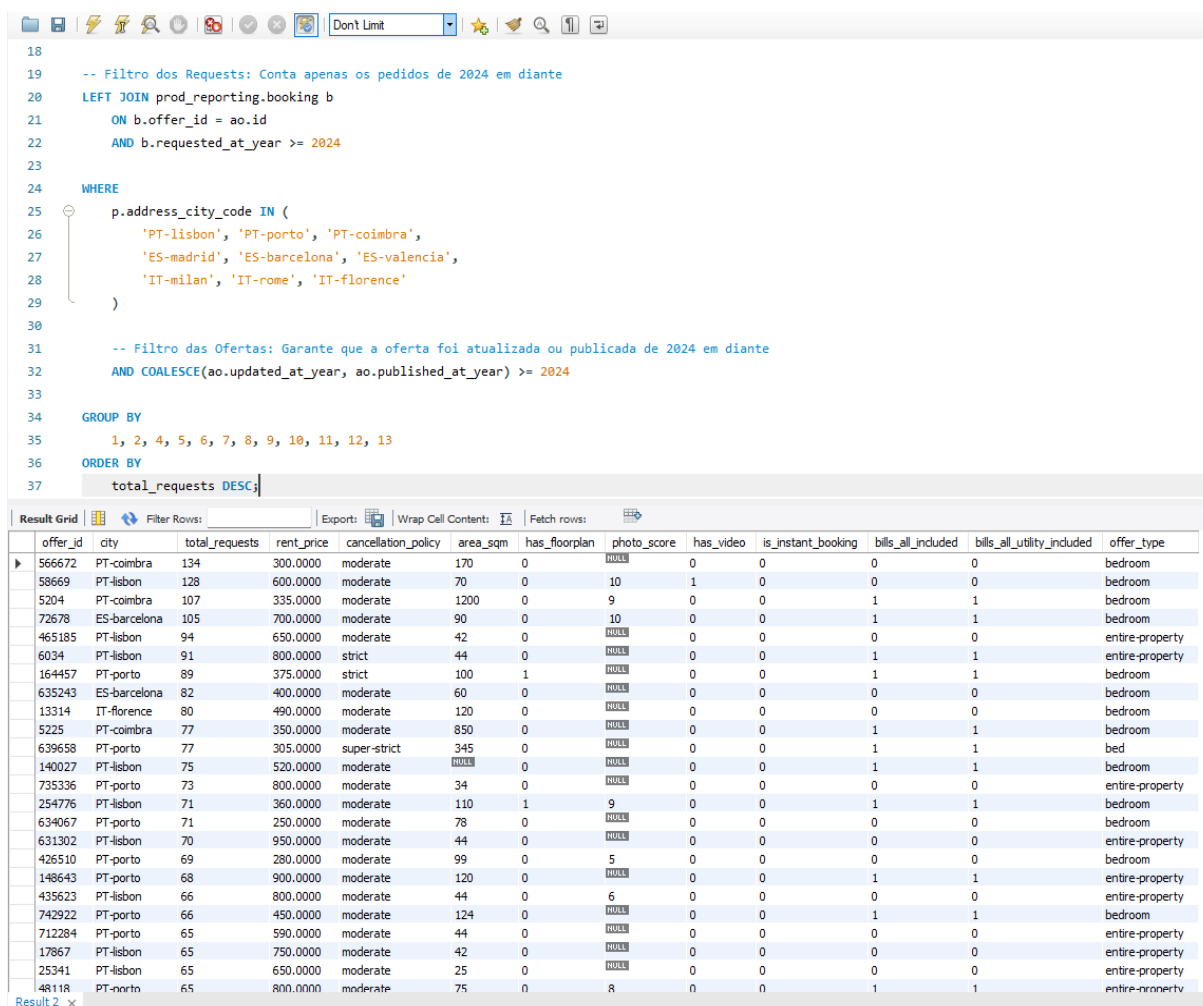
A presente secção descreve o desenvolvimento de um projeto de análise de dados de cariz científico, proposto pela Sênior *Data Analyst Alyssa Encarnacion*, com o objetivo de identificar os fatores que mais influenciam o volume de pedidos de reserva nas principais cidades de operação da *Uniplaces*. Este projeto representou o trabalho de maior complexidade técnica e analítica desenvolvido ao longo do estágio, na medida em que conjugou análise exploratória de dados, técnicas de estatística descritiva e inferencial e, numa fase mais avançada, o desenvolvimento de um modelo preditivo de aprendizagem automática, integrando assim múltiplas dimensões da Ciência de Dados numa única iniciativa.

A motivação subjacente ao projeto decorreu da necessidade de substituir o conhecimento empírico e intuitivo do mercado por uma abordagem rigorosa e baseada em evidências, capaz de responder a questões concretas de negócio: as ofertas mais baratas geram mais conversão? A disponibilização de planta do imóvel é determinante para a captação de pedidos? Uma política de cancelamento mais flexível atrai mais estudantes? De que forma o preço por metro quadrado influencia a decisão do utilizador? Estas perguntas, frequentemente respondidas com base no julgamento subjetivo das equipas comerciais, careciam de uma validação quantitativa que apenas uma abordagem de Ciência de Dados seria capaz de fornecer de forma robusta e escalável.

Para garantir a relevância temporal da análise face às dinâmicas atuais do mercado, foi desenvolvida uma consulta SQL sobre a base de dados interna da organização, visível na Figura 58, que permitiu extrair um conjunto de dados composto por mais de 124.000 ofertas publicadas ou atualizadas a partir do ano de 2024.

A consulta foi estruturada com recurso a uma junção entre a tabela de ofertas de alojamento (*accommodation_offer*) e a tabela de reservas (*booking*), estabelecida através do campo *offer_id*, e filtrada geograficamente para as cidades consideradas prioritárias pela organização: Lisboa, Porto e Coimbra em Portugal; Madrid, Barcelona e Valência em Espanha; e Roma,

Milão e Florença em Itália. Este filtro foi implementado através da cláusula *WHERE* aplicada ao campo *address_city_code*, com os códigos de cidade correspondentes a cada mercado.



```

18
19 -- Filtro dos Requests: Conta apenas os pedidos de 2024 em diante
20 LEFT JOIN prod_reporting.booking b
21   ON b.offer_id = ao.id
22   AND b.requested_at_year >= 2024
23
24 WHERE
25   p.address_city_code IN (
26     'PT-lisbon', 'PT-porto', 'PT-coimbra',
27     'ES-madrid', 'ES-barcelona', 'ES-valencia',
28     'IT-milan', 'IT-rome', 'IT-florence'
29   )
30
31 -- Filtro das Ofertas: Garante que a oferta foi atualizada ou publicada de 2024 em diante
32 AND COALESCE(ao.updated_at_year, ao.published_at_year) >= 2024
33
34 GROUP BY
35   1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
36 ORDER BY
37   total_requests DESC;

```

offer_id	city	total_requests	rent_price	cancellation_policy	area_sqm	has_floorplan	photo_score	has_video	is_instant_booking	bills_all_included	bills_all_utility_included	offer_type
566672	PT-coimbra	134	300.0000	moderate	170	0	NULL	0	0	0	0	bedroom
58669	PT-lisbon	128	600.0000	moderate	70	0	10	1	0	0	0	bedroom
5204	PT-coimbra	107	335.0000	moderate	1200	0	9	0	0	1	1	bedroom
72678	ES-barcelona	105	700.0000	moderate	90	0	10	0	0	1	1	bedroom
465185	PT-lisbon	94	650.0000	moderate	42	0	NULL	0	0	0	0	entire-property
6034	PT-lisbon	91	800.0000	strict	44	0	NULL	0	0	1	1	entire-property
164457	PT-porto	89	375.0000	strict	100	1	NULL	0	0	1	1	bedroom
635243	ES-barcelona	82	400.0000	moderate	60	0	NULL	0	0	0	0	bedroom
13314	IT-florence	80	490.0000	moderate	120	0	NULL	0	0	0	0	bedroom
5225	PT-coimbra	77	350.0000	moderate	850	0	NULL	0	0	1	1	bedroom
639658	PT-porto	77	305.0000	super-strict	345	0	NULL	0	0	1	1	bed
140027	PT-lisbon	75	520.0000	moderate	NULL	0	NULL	0	0	1	1	bedroom
735336	PT-porto	73	800.0000	moderate	34	0	NULL	0	0	0	0	entire-property
254776	PT-lisbon	71	360.0000	moderate	110	1	9	0	0	1	1	bedroom
634067	PT-porto	71	250.0000	moderate	78	0	NULL	0	0	0	0	bedroom
631302	PT-lisbon	70	950.0000	moderate	44	0	NULL	0	0	0	0	entire-property
426510	PT-porto	69	280.0000	moderate	99	0	5	0	0	0	0	bedroom
148643	PT-porto	68	900.0000	moderate	120	0	NULL	0	0	1	1	entire-property
435623	PT-lisbon	66	800.0000	moderate	44	0	6	0	0	0	0	entire-property
742922	PT-porto	66	450.0000	moderate	124	0	NULL	0	0	1	1	bedroom
712284	PT-porto	65	590.0000	moderate	44	0	NULL	0	0	0	0	entire-property
17867	PT-lisbon	65	750.0000	moderate	42	0	NULL	0	0	0	0	entire-property
25341	PT-lisbon	65	650.0000	moderate	25	0	NULL	0	0	0	0	entire-property
48118	PT-north	65	800.0000	moderate	75	0	8	0	0	1	1	entire-property

Figura 58 - Consulta SQL desenvolvida para a extração do conjunto de dados de ofertas ativas nas cidades prioritárias a partir de 2024

A restrição temporal foi assegurada através de dois critérios combinados: a data de submissão dos pedidos de reserva deveria ser igual ou posterior ao ano de 2024, e as ofertas incluídas deveriam ter sido publicadas ou atualizadas nesse mesmo período, com recurso à função *COALESCE* aplicada aos campos *updated_at_year* e *published_at_year*, garantindo, assim, que apenas o inventário ativo e recente fosse considerado. Os resultados foram ordenados de forma decrescente pelo número total de pedidos recebidos por oferta, permitindo identificar imediatamente as ofertas com maior tração na plataforma. O conjunto de dados final contemplou as três tipologias de alojamento disponíveis na plataforma, nomeadamente quartos privados (*bedroom*), apartamentos inteiros (*entire-property*) e camas em quartos partilhados (*bed*), e incluiu variáveis relativas ao preço de arrendamento, área em metros quadrados,

política de cancelamento, pontuação das fotografias, presença de vídeo, disponibilidade de reserva instantânea, inclusão de despesas e presença de planta do imóvel.

A fase de análise exploratória foi conduzida em *Python*, com recurso às bibliotecas *Pandas*, para a manipulação e agregação dos dados, e *Matplotlib* e *Seaborn*, para a produção das representações gráficas. Esta fase teve como propósito isolar o efeito individual de cada variável sobre o volume de pedidos recebidos, permitindo desconstruir um conjunto de pressupostos frequentemente assumidos pelas equipas comerciais sem suporte quantitativo.

A primeira variável analisada foi a presença de planta do imóvel (*has_floorplan*), cujo impacto se revelou de magnitude considerável. Conforme ilustrado na Figura 59, as ofertas que disponibilizavam planta do alojamento receberam, em média, 2,4 pedidos de reserva, face a uma média de 1,1 pedidos nas ofertas sem planta, o que representa um acréscimo de aproximadamente 118% no volume médio de conversão. Este resultado é interpretável à luz do contexto específico do mercado de arrendamento estudantil a distância: na ausência de visitas presenciais ao imóvel, a clareza espacial proporcionada pela planta traduz-se num aumento significativo da confiança do utilizador no momento da decisão, reduzindo a incerteza associada à escolha de um alojamento que não foi fisicamente visitado. A tipologia de gráfico de barras foi adotada nesta e nas análises subsequentes por constituir o formato mais adequado para a comparação de médias entre categorias discretas e mutuamente exclusivas, permitindo uma leitura direta das diferenças de magnitude entre grupos [66].

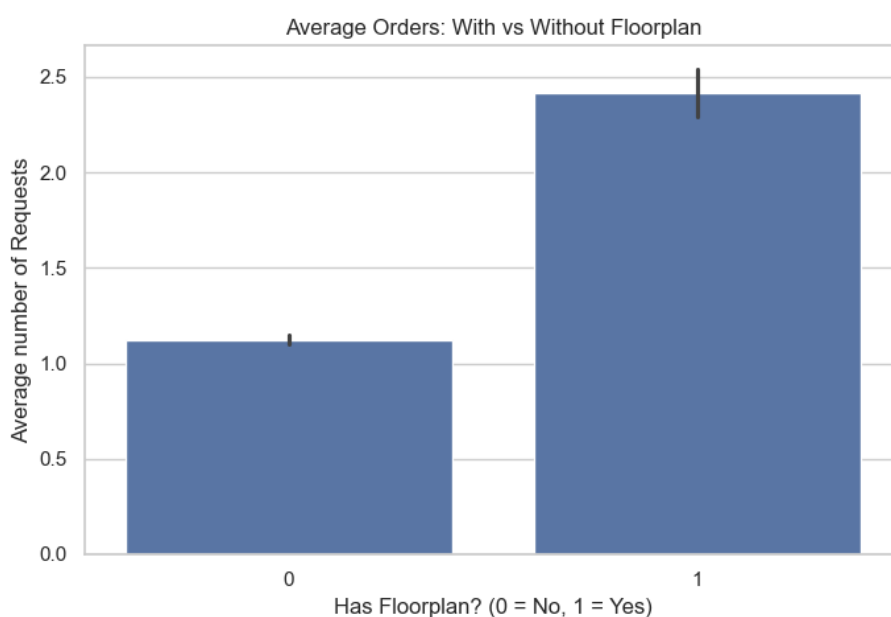


Figura 59 - Impacto da presença de planta do imóvel no volume médio de pedidos de reserva

A análise da política de cancelamento, ilustrada na Figura 60, produziu um resultado que contraria a intuição comercial predominante. Contrariamente ao que seria expectável, a política de cancelamento flexível registou o menor volume médio de pedidos de todo o sistema, com uma média inferior a 0,8. A política moderada emergiu como o ponto de equilíbrio preferido pelo mercado, liderando com uma média de 1,3 pedidos por oferta. Este padrão sugere que, no segmento de arrendamento de médio e longo prazo predominante na plataforma da *Uniplaces*, os utilizadores não valorizam a flexibilidade de cancelamento da mesma forma que em contextos de arrendamento de curta duração. Uma hipótese explicativa adicional reside no facto de as políticas mais flexíveis poderem estar associadas a preços base mais elevados, na medida em que os proprietários tendem a incorporar o risco acrescido de cancelamento no valor da renda, penalizando, assim, a visibilidade da oferta nas pesquisas ordenadas por preço crescente.

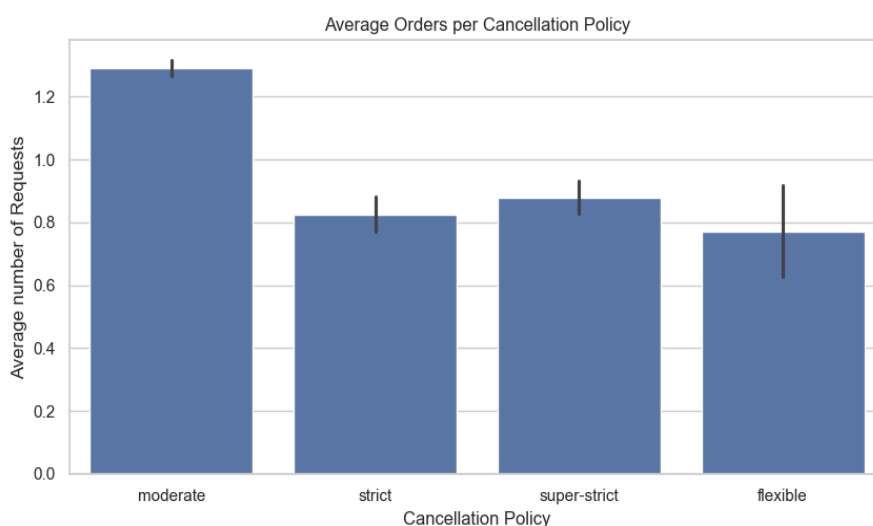


Figura 60 - Volume médio de pedidos de reserva por política de cancelamento

A análise do impacto das condições financeiras, designadamente a inclusão de despesas e a disponibilidade de reserva instantânea, produziu resultados igualmente contraintuitivos, conforme demonstrado na Figura 61. As ofertas sem despesas incluídas registaram um volume médio de pedidos superior a 1,5, enquanto as ofertas com todas as despesas incluídas ficaram abaixo de 1,0. A justificação para este paradoxo reside no comportamento de pesquisa dos utilizadores: ao incluir as despesas no valor da renda, os proprietários incrementam o preço base apresentado na plataforma, o que penaliza a posição da oferta nas listagens ordenadas por preço crescente. Como os estudantes tendem a iniciar a sua pesquisa pelo critério do menor preço, as ofertas com despesas incluídas perdem o clique inicial, independentemente do seu valor intrínseco. A variável de reserva instantânea revelou, por sua vez, um impacto positivo marginal e sem significado estatístico relevante no volume médio de pedidos.

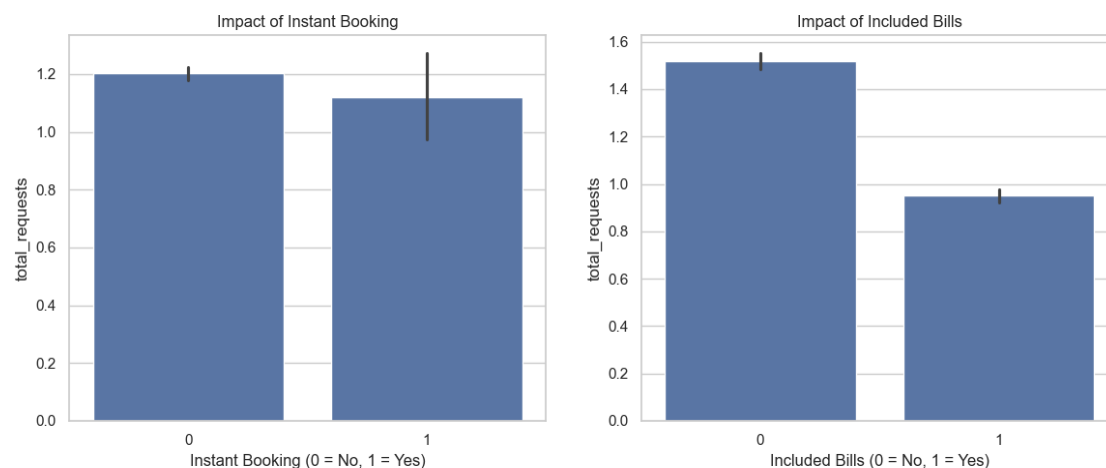


Figura 61 - Impacto da reserva instantânea e da inclusão de despesas no volume médio de pedidos de reserva

O cruzamento das variáveis de cidade e tipologia de alojamento, ilustrado na Figura 62 através de um gráfico de barras agrupadas, permitiu identificar padrões geográficos de relevância estratégica. O mercado português, com destaque para Coimbra, Lisboa e Porto, domina o volume médio de pedidos por oferta, com uma concentração marcante na tipologia de quarto privado. A exceção mais notável no contexto europeu analisado é Milão, onde a tipologia de cama em quarto partilhado supera, em média, a procura por quartos privados e apartamentos inteiros, fenómeno explicável pelos elevados custos de habitação na cidade italiana, que tornam as soluções de alojamento partilhado a opção mais acessível para o segmento estudantil.

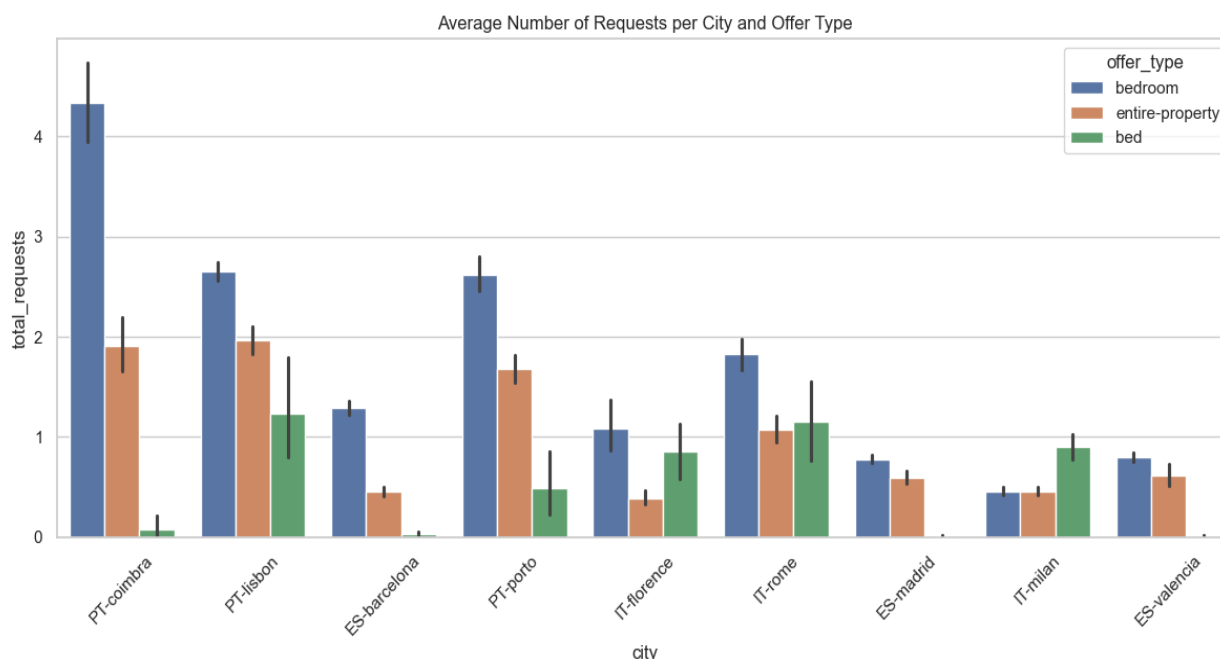


Figura 62 - Volume médio de pedidos de reserva por cidade e tipologia de alojamento

A relação entre o preço de arrendamento e o volume de pedidos foi analisada através de um gráfico de dispersão, tipologia escolhida por ser a mais indicada para identificar padrões e concentrações numa distribuição bivariada com elevado número de observações [10], conforme visível na Figura 63. A análise revelou a existência de um pico de procura pronunciado para quartos privados na faixa de preços compreendida entre os 300 e os 600 euros mensais. A partir do limiar de 750 a 800 euros, o volume de pedidos para quartos privados colapsa de forma abrupta, sendo o mercado nessa faixa de preços composto exclusivamente por procura de apartamentos inteiros. Este padrão constitui uma referência de posicionamento de preço de elevado valor estratégico para os proprietários que operam na plataforma.

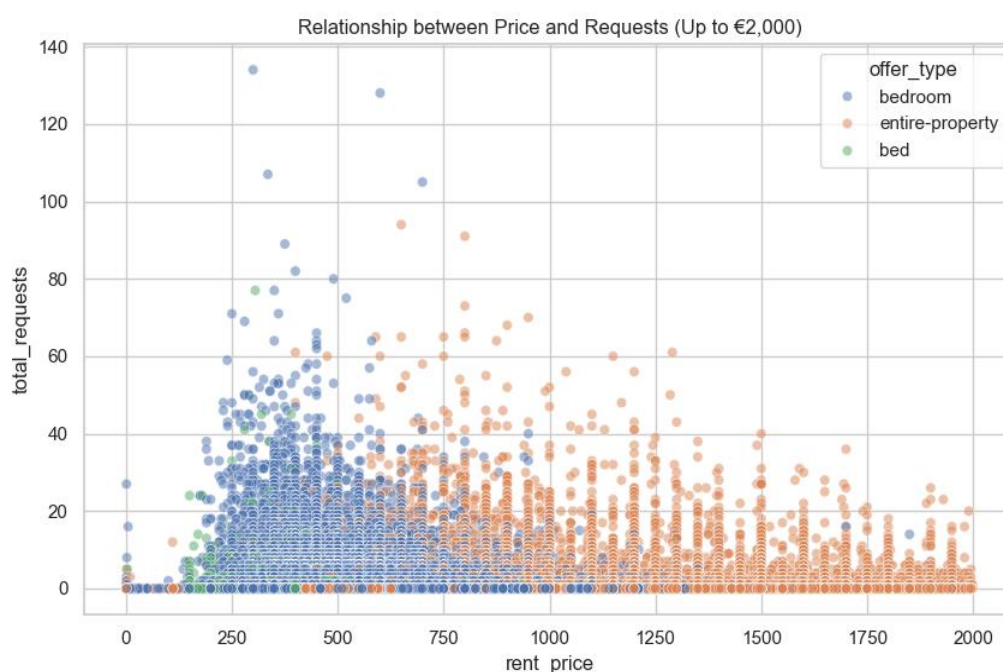


Figura 63 - Relação entre o preço de arrendamento e o volume de pedidos de reserva por tipologia de alojamento

Para validar estatisticamente as observações resultantes da análise exploratória, foi produzida uma matriz de correlação de *Pearson*, visível na Figura 64, que quantificou o grau de associação linear entre cada variável numérica do conjunto de dados e o volume de pedidos recebidos. A matriz de correlação é representada sob a forma de um mapa de calor (*heatmap*), formato que permite uma leitura visual imediata da intensidade e direção das associações entre variáveis, sendo os tons avermelhados indicativos de correlações positivas e os tons azulados indicativos de correlações negativas [9].



Figura 64 - Matriz de correlação entre as variáveis numéricas do conjunto de dados e o volume de pedidos de reserva

Da análise da matriz extraíram-se três conclusões de particular relevância. Em primeiro lugar, as variáveis *has_floorplan* e *has_video* partilham a correlação positiva mais elevada com o volume de pedidos, ambas com um coeficiente de 0,08, confirmando quantitativamente o papel determinante da qualidade e completude do conteúdo visual na captação de interesse por parte dos utilizadores.

Foi ainda identificada uma correlação significativa entre estas duas variáveis entre si, com um coeficiente de 0,38, sugerindo que os proprietários com maior grau de profissionalização tendem a adotar ambas as práticas em simultâneo. Em segundo lugar, a variável *bills_all_included* apresentou uma correlação negativa de 0,07 com o volume de pedidos, validando matematicamente a hipótese de que o modelo de preço com todas as despesas incluídas prejudica a visibilidade e a atratividade inicial da oferta. Em terceiro lugar, a correlação isolada do preço bruto de arrendamento com o volume de pedidos revelou-se praticamente nula, com um coeficiente de 0,00, o que não deve ser interpretado como ausência de impacto do preço, mas antes como indicativo de que o efeito do preço é fortemente mediado pela tipologia e área do imóvel, tornando a análise univariada insuficiente para captar esta dinâmica e justificando o avanço para um modelo de aprendizagem automática capaz de tratar múltiplas variáveis em simultâneo.

A fase de modelação teve início com a definição de uma nova métrica de negócio, designada *Success Score*, que traduzia o desempenho de cada oferta numa escala percentual de 0 a 100%, onde 100 pedidos ou mais correspondiam a uma conversão total de 100%. Esta transformação permitiu reenquadrar o problema como uma tarefa de regressão, tornando o modelo mais interpretável do ponto de vista do negócio.

No âmbito da engenharia de variáveis (*Feature Engineering*), foi identificada a necessidade de criar uma variável derivada que capturasse de forma mais fiel o processo de decisão do consumidor. A análise isolada do preço de arrendamento ou da área do imóvel revelou-se insuficiente, dado que o valor percebido pelo utilizador decorre da relação entre ambas as dimensões. Uma renda de 800 euros pode representar uma oferta apelativa para um apartamento espaçoso, mas ser completamente desadequada para um quarto de dimensão reduzida. Para colmatar esta lacuna analítica, foi criada a variável *price_per_sqm*, correspondente ao rácio entre o preço de arrendamento e a área do imóvel em metros quadrados, que passou a integrar o conjunto de variáveis preditoras do modelo.

O algoritmo selecionado para o desenvolvimento do modelo preditivo foi o *Random Forest Regressor*, um método de aprendizagem automática baseado em conjuntos de árvores de decisão (*ensemble learning*) que se distingue pela sua robustez face à presença de variáveis com escalas distintas, pela sua capacidade de captar relações não lineares entre variáveis e pela interpretabilidade dos resultados através da análise de importância das variáveis (*feature importance*) [9]. O modelo foi treinado sobre o conjunto de dados extraído, após a aplicação de técnicas de pré-processamento que incluíram a codificação das variáveis categóricas através de *one-hot encoding*, nomeadamente as variáveis de cidade e tipologia de alojamento, e a remoção de registos com valores omissos nas variáveis preditoras.

A seleção exclusiva do *Random Forest Regressor*, prescindindo de um *benchmark* comparativo alargado com outros algoritmos, justificou-se pela natureza híbrida e complexa do conjunto de dados da *Uniplaces*. A presença simultânea de variáveis categóricas de elevada cardinalidade (como a cidade e a tipologia) e de métricas numéricas não lineares inviabiliza a aplicação eficaz de modelos paramétricos tradicionais, como a Regressão Linear, sem transformações severas que comprometeriam a interpretabilidade. Adicionalmente, o contexto prático de estágio priorizou a entrega de uma solução fim-a-fim que garantisse interpretabilidade imediata para o negócio através do cálculo de *feature importance*, um

requisito estratégico para o departamento financeiro que modelos do tipo *black-box* (como redes neurais complexas) não conseguiriam oferecer com a mesma transparência operacional.

Os resultados obtidos na avaliação do modelo revelaram um desempenho satisfatório para o problema em causa. O erro absoluto médio (*Mean Absolute Error*, MAE) registado foi de 1,44, indicando que, em média, as previsões do modelo se desviam do valor real do *Success Score* em menos de 1,5 pontos percentuais. O coeficiente de determinação R^2 situou-se em 0,21, o que significa que o modelo é capaz de explicar 21% da variância total observada no comportamento de procura. Este valor, embora moderado, é interpretável no contexto da complexidade inerente à modelação do comportamento humano, onde uma proporção significativa da variância é determinada por fatores subjetivos e contextuais não capturáveis pelas variáveis disponíveis, como a qualidade da decoração, a exata proximidade a instituições de ensino ou a reputação do proprietário. A obtenção de valores moderados de R^2 em modelos preditivos aplicados ao comportamento humano é um resultado expectável e amplamente documentado na literatura, na medida em que os modelos de aprendizagem automática, independentemente da sua sofisticação, são intrinsecamente limitados pela informação disponível nas variáveis observadas, não conseguindo capturar a totalidade dos fatores latentes que influenciam a decisão individual [70].

A análise da importância das variáveis no modelo, ilustrada na Figura 65, permitiu ordenar os fatores determinantes do sucesso de uma oferta pelo seu peso preditivo efetivo.

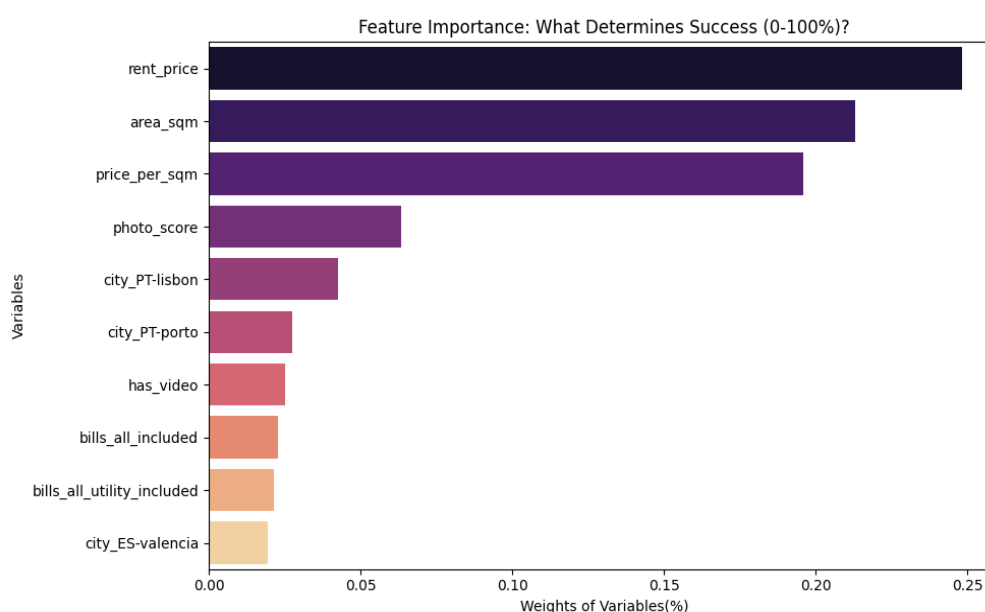


Figura 65 - Importância relativa das variáveis preditoras no modelo *Random Forest Regressor*

O preço de arrendamento (*rent_price*), a área do imóvel (*area_sqm*) e o preço por metro quadrado (*price_per_sqm*) representam conjuntamente cerca de 70% do peso total do modelo, confirmando que o mercado de arrendamento estudantil é fundamentalmente racional e orientado pela relação entre qualidade e custo. A análise da importância relativa das variáveis preditoras constitui uma das principais vantagens dos modelos baseados em árvores de decisão, na medida em que permite identificar, de forma interpretável, quais os fatores que mais contribuem para a capacidade preditiva do modelo [71]. Imediatamente a seguir ao fator financeiro, a pontuação das fotografias (*photo_score*) surge como a terceira variável mais importante de todo o sistema, seguida pela presença de vídeo (*has_video*), demonstrando que a qualidade do conteúdo visual de um anúncio supera, em termos preditivos, a localização geográfica da oferta. As variáveis de localização premium, nomeadamente Lisboa e Porto, surgem no grupo intermédio de importância, enquanto as condições financeiras adicionais, como a inclusão de despesas, ocupam posições residuais na hierarquia de fatores determinantes da decisão do utilizador. A interpretação da importância das variáveis em modelos de aprendizagem automática deve ser sempre complementada com uma análise crítica do contexto de negócio, de forma a distinguir relações causais de meras correlações estatísticas [72].

Para complementar a análise estática das características das ofertas com uma perspetiva temporal do comportamento da procura, procedeu-se a uma análise de sazonalidade baseada na agregação mensal do volume total de pedidos de reserva. Os dados foram extraídos via SQL, agrupados pelo mês de submissão do pedido e segmentados pela tipologia de alojamento, conforme ilustrado na Figura 66.

A representação visual desta série temporal foi desenvolvida em *Python* com recurso a um gráfico de linhas, tipologia estatisticamente mais adequada para ilustrar a evolução contínua de variáveis ao longo do tempo e identificar tendências, ciclos e anomalias sazonais [10], conforme visível na Figura 67.

A análise da série temporal revelou padrões comportamentais distintos consoante a tipologia de alojamento. No que respeita aos quartos privados, tipologia dominante na plataforma, verificou-se que a procura opera com forte antecipação relativamente ao início do ano letivo, atingindo o seu pico máximo em julho, com aproximadamente 14.000 pedidos registados. Este padrão demonstra que os estudantes tendem a assegurar o seu alojamento com vários meses de antecedência relativamente à data de entrada. Foi igualmente identificado um

ciclo secundário de procura em janeiro, atribuível à chegada de estudantes em programas de mobilidade internacional para o segundo semestre letivo.

mes_pedido	offer_type	total_requests
1	bedroom	9567
1	entire-property	4427
1	bed	328
2	bedroom	7278
2	entire-property	4063
2	bed	202
3	bedroom	6723
3	entire-property	3500
3	bed	135
4	bedroom	6716
4	entire-property	3109
4	bed	148
5	bedroom	9283
5	entire-property	4157
5	bed	197
6	bedroom	10126
6	entire-property	3862
6	bed	292
7	bedroom	14025
7	entire-property	4742
7	bed	518
8	bedroom	12243
8	entire-property	5118
8	bed	447
9	bedroom	7968
9	entire-property	3363
9	bed	368
10	bedroom	6251

Figura 66 - Extração SQL do volume total de pedidos agregado por mês e tipologia de alojamento

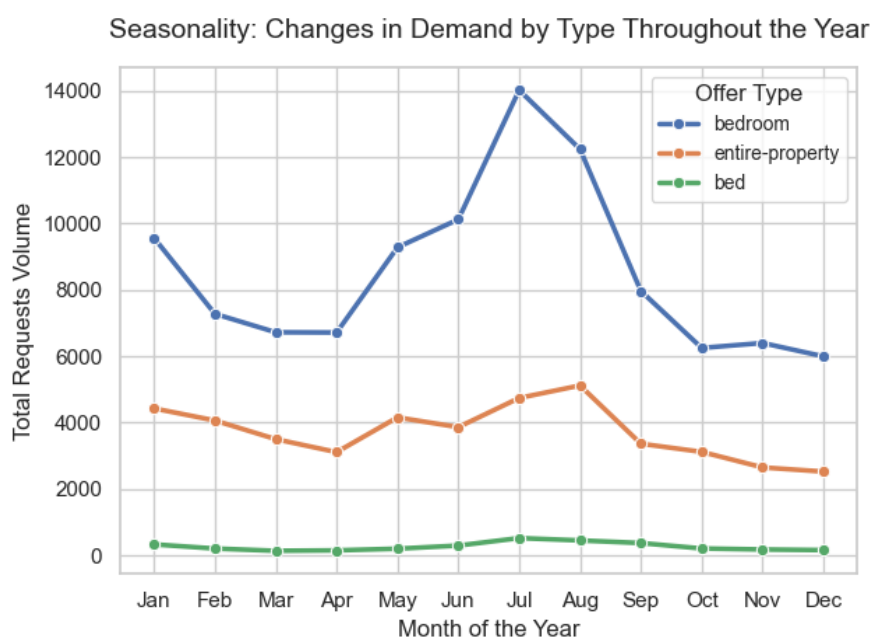


Figura 67 - Evolução sazonal do volume de pedidos de reserva por tipologia de alojamento ao longo do ano

A identificação de padrões sazonais em séries temporais de procura constitui uma das aplicações mais relevantes da aprendizagem estatística aplicada ao comportamento do consumidor, permitindo às organizações antecipar flutuações de mercado e ajustar a sua estratégia operacional em conformidade [73]. O comportamento dos apartamentos inteiros

apresenta uma dinâmica distinta, com o pico de procura a ocorrer em agosto, sugerindo um processo de decisão mais tardio e concentrado, característico de um segmento de utilizadores com perfil diferente do estudante universitário típico, como jovens profissionais ou famílias em situação de relocação.

Com base no conjunto de evidências quantitativas produzidas ao longo do projeto, foram formuladas recomendações estratégicas destinadas à responsável pelo departamento de dados. Dado que variáveis como a área do imóvel e o preço de arrendamento são determinadas pelo proprietário e não são diretamente controláveis pela plataforma, a alavanca de maior retorno imediato identificada consiste no investimento na qualidade visual dos anúncios. Recomendou-se assim a criação de campanhas de incentivo dirigidas aos proprietários para a inclusão de plantas do imóvel, vídeos e fotografias de maior qualidade, dado que estas variáveis apresentam um impacto preditivo superior ao da localização geográfica no modelo desenvolvido. Foi igualmente recomendado sensibilizar os proprietários para os efeitos negativos da estratégia de preço com despesas incluídas, demonstrando com dados que esta prática reduz a visibilidade da oferta nas pesquisas por preço crescente. Em termos de posicionamento de preço, os dados indicam que os quartos privados devem ser mantidos preferencialmente abaixo do limiar dos 700 a 750 euros mensais, com foco na janela de maior conversão situada entre os 300 e os 600 euros. Por fim, no que respeita à estratégia de expansão da tipologia de camas em quartos partilhados, os dados apontam Milão como o mercado com maior aptidão para absorver este tipo de oferta, sendo a única cidade analisada onde esta tipologia supera a procura por quartos privados e apartamentos inteiros.

Este projeto constituiu a experiência formativa mais abrangente do estágio, por ter exigido a mobilização integrada de competências em extração de dados via SQL, análise exploratória em *Python*, visualização de dados, estatística inferencial e desenvolvimento de modelos de aprendizagem automática, culminando na produção de recomendações estratégicas com impacto direto na operação comercial da *Uniplaces*.

3.4. Análise crítica e proposta de melhorias

Ao longo do estágio, foi possível observar que, apesar da elevada competência técnica e da cultura de inovação associadas à *Uniplaces*, existiam ainda vários aspetos organizacionais e operacionais suscetíveis de melhoria. A análise crítica destas dimensões é particularmente relevante, uma vez que permite enquadrar a experiência de estágio não apenas como um

momento de aprendizagem prática, mas também como uma oportunidade para refletir sobre a eficiência dos processos e sobre a forma como a estrutura interna de uma organização pode influenciar o desempenho das equipas.

Um dos aspetos que mais se destacou foi o facto de grande parte do estágio ter decorrido em regime remoto. Embora este modelo de trabalho apresente vantagens evidentes, nomeadamente em termos de flexibilidade e conciliação entre vida pessoal e profissional, também limita, em certa medida, a construção de um ambiente de proximidade entre colegas. No caso concreto da *Uniplaces*, existia um espaço de *coworking* em Lisboa, porém a presença física era reduzida na maioria das equipas. Esta realidade dificultava a interação espontânea, o fortalecimento de relações interpessoais e a criação de um verdadeiro espírito de equipa. Na prática, a ausência de contacto presencial frequente pode também repercutir-se na dinâmica de trabalho, uma vez que a troca informal de ideias e a resolução rápida de dúvidas tendem a ocorrer com maior fluidez em ambiente presencial.

Outro ponto identificado relaciona-se com a existência de diversos processos ainda baseados em intervenção manual. Apesar de a empresa se posicionar como uma organização tecnológica, foi possível constatar que algumas tarefas continuavam a depender de procedimentos repetitivos e pouco automatizados, o que implicava um dispêndio considerável de tempo por parte das equipas. Esta situação não só afetava a produtividade, como também aumentava a probabilidade de erro humano e reduzia a disponibilidade para tarefas de maior valor acrescentado. Numa perspetiva de melhoria contínua, seria desejável investir de forma mais consistente na automatização de fluxos de trabalho, sobretudo em tarefas recorrentes e de carácter operacional.

A gestão e organização dos ficheiros de trabalho constituiu igualmente uma área com margem para evolução. Durante o estágio, verificou-se que a multiplicidade de ficheiros em *Google Sheets*, associada ao elevado volume de informação processada, originava dificuldades de navegação, tempos de carregamento elevados e alguma dispersão documental. Em determinados casos, os ficheiros tornavam-se demasiado pesados, comprometendo a sua utilização eficiente. Esta realidade evidencia a necessidade de repensar a arquitetura de armazenamento e de migração de dados, privilegiando ferramentas mais robustas e escaláveis, capazes de responder de forma mais adequada às exigências analíticas da empresa. Para além disso, uma maior centralização da informação permitiria reduzir redundâncias, melhorar a rastreabilidade dos dados e facilitar o acesso por parte das diferentes equipas.

Entre as abordagens exploradas durante o estágio, destaca-se o recurso a técnicas elementares de *Text Mining* aplicadas a avaliações da *Trustpilot*, que permitiram estruturar *feedback* textual e associá-lo a agentes específicos. Embora a solução implementada tenha assentado numa lógica baseada em correspondência por variantes nominais, esta experiência evidenciou o potencial de evolução para abordagens mais avançadas, como análise de sentimento, reconhecimento automático de entidades ou classificação temática de comentários.

Uma outra observação relevante prende-se com a necessidade de reforçar uma cultura organizacional mais orientada para a otimização de processos e para a digitalização de tarefas internas. Sendo a *Uniplaces* uma empresa com forte componente tecnológica, seria expectável uma maior aposta na sistematização e automação de operações rotineiras. A persistência de muitos procedimentos manuais sugere, por isso, não apenas limitações técnicas pontuais, mas também a existência de hábitos de trabalho que podem exigir uma mudança cultural mais ampla. Nesse sentido, a adoção de ferramentas de automação, a revisão periódica dos processos internos e a definição de boas práticas de gestão da informação poderiam contribuir significativamente para o aumento da eficiência global da organização.

Em síntese, apesar da relevância formativa da experiência de estágio e da oportunidade de contactar com contextos reais de trabalho em ciência de dados, foi possível identificar várias áreas com potencial de melhoria. Entre as principais sugestões destacam-se o reforço da colaboração presencial ou híbrida, a automatização de processos repetitivos, a reorganização e centralização dos ficheiros de trabalho e a transição gradual para ferramentas mais adequadas à escala dos dados tratados. Estas medidas, além de contribuírem para um funcionamento mais eficiente da empresa, poderiam também potenciar uma melhor experiência de trabalho para os colaboradores e estagiários.

4. Conclusão

O presente relatório teve como propósito documentar e refletir sobre o trabalho desenvolvido ao longo do estágio curricular realizado na *Uniplaces*, no âmbito do Mestrado, com o objetivo central de aplicar em contexto profissional real os conhecimentos teóricos e metodológicos adquiridos durante o percurso académico. A análise do conjunto de tarefas realizadas permite concluir que este objetivo foi alcançado, tendo o estágio proporcionado uma imersão efetiva nas práticas e dinâmicas de trabalho de uma equipa de dados inserida numa empresa tecnológica de dimensão europeia.

Ao longo do período de estágio, foi possível intervir em projetos de natureza e complexidade diversas, desde a construção de relatórios operacionais automatizados até ao desenvolvimento de um modelo preditivo de aprendizagem automática aplicado ao mercado de arrendamento estudantil. Esta diversidade de tarefas permitiu não apenas consolidar competências previamente adquiridas, nomeadamente ao nível da manipulação de dados em SQL e *Python*, mas também desenvolver proficiência em ferramentas que até então não faziam parte do repertório técnico, como o *Tableau*, o *Google Apps Script* e a integração com API externas. A autonomia progressivamente conquistada ao longo do estágio, à medida que o domínio das ferramentas e dos processos internos se foi consolidando, constituiu igualmente um elemento de desenvolvimento pessoal e profissional de relevo, na medida em que exigiu capacidade de iniciativa, gestão de prioridades e responsabilidade pelos resultados produzidos.

Não obstante o balanço globalmente positivo, importa reconhecer que o estágio apresentou também aspetos menos favoráveis. A fase inicial revelou-se particularmente exigente, dado que o acompanhamento disponibilizado não foi suficiente para uma integração plena e imediata nos processos e ferramentas da equipa, o que implicou um período de adaptação mais longo e autónomo do que o desejável. Adicionalmente, algumas das tarefas desenvolvidas ao longo do estágio revestiram-se de um carácter mais rotineiro e repetitivo, afastando-se pontualmente do trabalho analítico de maior valor acrescentado que constituía a principal motivação para a realização do estágio. Ainda assim, estas circunstâncias contribuíram para o desenvolvimento de resiliência e capacidade de resolução autónoma de problemas, competências igualmente valorizadas no mercado de trabalho.

Do ponto de vista das consequências para o percurso futuro, o estágio na *Uniplaces* consolidou a orientação profissional para a área de Ciência de Dados e Análise de Dados. A

experiência de trabalhar com conjuntos de dados reais, em ambiente de pressão e com impacto direto na tomada de decisão organizacional, tornou evidente a relevância e a aplicabilidade das competências desenvolvidas ao longo do Mestrado, reforçando a motivação para aprofundar o conhecimento nesta área e para prosseguir uma carreira profissional orientada para a análise e interpretação de dados como suporte à estratégia empresarial.

A sólida formação teórica e prática providenciada pelo Mestrado em Ciência de Dados revelou-se um fator estruturante para o sucesso de todas as atividades desenvolvidas. Unidades curriculares fundamentais como Integração de Dados, *Text Mining*, *Business Intelligence* e *Deep Learning* foram competências essenciais para modelar a infraestrutura de dados da *Uniplaces*, desenhar os *dashboards* operacionais de reporte financeiro e implementar o algoritmo preditivo de regressão. Esta ligação direta entre o plano de estudos académico e as exigências do ambiente empresarial dinâmico permitiu consolidar os conceitos teóricos através da sua aplicação direta na resolução de problemas reais de negócio, validando o impacto prático do percurso académico.

Em suma, o estágio curricular realizado na *Uniplaces* representou uma etapa determinante no processo de transição entre o percurso académico e o mundo profissional, tendo proporcionado uma experiência formativa abrangente que combinou aprendizagem técnica, desenvolvimento de autonomia e contacto com os desafios reais de uma organização orientada para os dados.

Bibliografia

- [1] K. J. Cios, W. Pedrycz, R. W. Swiniarski e L. A. Kurgan, *Data Mining: A Knowledge Discovery Approach*, New York: Springer, 2007.
- [2] A. Osterwalder e Y. Pigneur, *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*, John Wiley & Sons, 2010.
- [3] D. Teece, “Business models, business strategy and innovation,” *Long Range Planning*, vol. 43, n° 2-3, pp. 172-194, 2010.
- [4] F. Provost e T. Fawcett, *Data Science for Business*, O'Reilly Media, 2013.
- [5] J. Han, M. Kamber e J. Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2011.
- [6] D. Donoho, “50 years of data science,” *Journal of Computational and Graphical Statistics*, vol. 26, n° 4, p. 745–766, 2017.
- [7] H. Wickham e G. Grolemund, *R for Data Science*, O'Reilly Media, 2017.
- [8] J. Tukey, *Exploratory Data Analysis*, Addison-Wesley, 1977.
- [9] T. Hastie, R. Tibshirani e J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.
- [10] S. Few, *Now You See It: Simple Visualization Techniques for Quantitative Analysis*, Analytics Press, 2009.
- [11] R. Kimball e M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, Wiley, 2013.
- [12] R. Feldman e J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, 2007.
- [13] C. Manning, P. Raghavan e H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.
- [14] D. J. Power, *Decision Support Systems: Concepts and Resources for Managers*, Westport: Quorum Books, 2002.
- [15] H. J. Watson e B. H. Wixom, “The Current State of Business Intelligence,” *IEEE Computer*, vol. 40, n° 9, pp. 96-99, 2007.
- [16] W. H. Inmon, *Building the Data Warehouse*, Indianapolis: John Wiley & Sons, 2005.

-
- [17] D. L. Rogers, *The Digital Transformation Playbook*, New York: Columbia Business School Publishing, 2016.
- [18] C. Anderson, *Creating a Data-Driven Organization*, Sebastopol: O'Reilly Media, 2015.
- [19] P. C. Zikopoulos, *Harnessing the Power of Big Data*, New York: McGraw-Hill Education, 2013.
- [20] B. Shneiderman, “The Eyes Have It: A Data Type by Data Task Taxonomy for Information Visualizations,” em *Proceedings of the IEEE Symposium on Visual Languages (VL'96)*, 1996.
- [21] C. O. Wilke, *Fundamentals of Data Visualization*, Sebastopol: O'Reilly Media, 2019.
- [22] S. Wexler, J. Cotgreave e A. Jones, *The Big Book of Dashboards*, Hoboken: Wiley, 2017.
- [23] W. W. Eckerson, *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*, Hoboken: Wiley, 2010.
- [24] T. Munzner, *Visualization Analysis and Design*, Boca Raton: CRC Press, 2014.
- [25] J. D. Kelleher, B. Mac Namee e A. D'Arcy, *Fundamentals of Machine Learning for Predictive Data Analytics*, Cambridge: MIT Press, 2015.
- [26] S. Russell e P. Norvig, *Artificial Intelligence: A Modern Approach*, Hoboken: Pearson, 2021.
- [27] T. M. Mitchell, *Machine Learning*, New York: McGraw-Hill, 1997.
- [28] E. Alpaydin, *Introduction to Machine Learning*, Cambridge: MIT Press, 2020.
- [29] I. Goodfellow, Y. Bengio e A. Courville, *Deep Learning*, Cambridge: MIT Press, 2016.
- [30] C. M. Bishop, *Pattern Recognition and Machine Learning*, New York: Springer, 2006.
- [31] K. P. Murphy, *Probabilistic Machine Learning: An Introduction*, Cambridge: MIT Press, 2022.
- [32] P. Flach, *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*, Cambridge: Cambridge University Press, 2012.
- [33] T. G. Dietterich, “Ensemble Methods in Machine Learning,” em *Multiple Classifier Systems (Lecture Notes in Computer Science)*, Berlin, 2000.
- [34] A. Geron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, Sebastopol: O'Reilly Media, 2019.
- [35] B. Efron e T. Hastie, *Computer Age Statistical Inference*, Cambridge: Cambridge University Press, 2016.

-
- [36] N. Silver, *The Signal and the Noise: Why So Many Predictions Fail-but Some Don't*, New York: Penguin Press, 2012.
- [37] M. I. Jordan e T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” *Science*, vol. 349, n° 6245, pp. 255-260, 2015.
- [38] T. Davenport e J. Harris, *Competing on Analytics: The New Science of Winning*, Harvard Business School Press, 2007.
- [39] K. Schwaber e J. Sutherland, “The Scrum Guide,” 2020. [Online]. Available: <https://scrumguides.org/scrum-guide.html>.
- [40] E. Meyer, *The Culture Map: Breaking Through the Invisible Boundaries of Global Business*, PublicAffairs, 2014.
- [41] A. Bacchelli e C. Bird, “Expectations, outcomes, and challenges of modern code review,” em *Proceedings of the 2013 International Conference on Software Engineering (ICSE)*, 2013.
- [42] G. Shmueli, P. Bruce, I. Yahav, N. Patel e K. Lichtendahl Jr, *Data Mining for Business Analytics: Concepts, Techniques, and Applications*, Wiley, 2020.
- [43] M.-A. Storey, L. Singer, B. Cleary, F. Filho e A. Zagalsky, “The Role of Social Media in Software Engineering,” *IEEE Software*, p. 39–45, 2017.
- [44] T. Davenport, *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*, Harvard Business Review Press, 2014.
- [45] R. Sharda, D. Delen e E. Turban, *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*, Pearson, 2020.
- [46] J. Walkenbach, *Excel 2016 Bible*, John Wiley & Sons, 2015.
- [47] W. McKinney, *Python for Data Analysis*, O'Reilly Media, 2017.
- [48] R. Chandra e B. Varanasi, *Python Requests Essentials*, Packt Publishing, 2015.
- [49] J. Friedl, *Mastering Regular Expressions*, O'Reilly Media, 2006.
- [50] C. Knafllic, *Storytelling with Data: A Data Visualization Guide for Business Professionals*, John Wiley & Sons, 2015.
- [51] B. Marr, *Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance*, John Wiley & Sons, 2015.
- [52] D. Murray, *Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software*, John Wiley & Sons, 2013.

-
- [53] S. Few, *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, Analytics Press, 2012.
- [54] A. Silberschatz, H. Korth e S. Sudarshan, *Database System Concepts*, McGraw-Hill Education, 2020.
- [55] R. Elmasri e S. Navathe, *Fundamentals of Database Systems*, Pearson, 2015.
- [56] A. McAfee, *Enterprise 2.0: New Collaborative Tools for Your Organization's Toughest*, Harvard Business Press, 2009.
- [57] T. Davenport e L. Prusak, *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, 1998.
- [58] T. Redman, "Data's Credibility Problem," *Harvard Business Review*, vol. 91, nº 12, pp. 84-88, 2013.
- [59] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer e R. Wirth, "CRISP-DM 1.0: Step-by-step data mining guide," SPSS Inc., Chicago, 2000.
- [60] E. Gamma, R. Helm, R. Johnson e J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software*, Boston: Addison-Wesley Professional, 1994.
- [61] C. O'Neil e R. Schutt, *Doing Data Science: Straight Talk from the Frontline*, Sebastopol: O'Reilly Media, 2013.
- [62] M. Alexander e D. Kusleika, *Excel Dashboards and Reports*, John Wiley & Sons, 2018.
- [63] J. Pauley, *Dashboards for Excel*, Apress, 2021.
- [64] L. Ferreira, *Google Apps Script: Automatização e Integração*, Edições Sílabo, 2020.
- [65] A. Beaulieu, *Learning SQL: Generate, Manipulate, and Retrieve Data*, O'Reilly Media, 2020.
- [66] A. Cairo, *The Truthful Art: Data, Charts, and Maps for Communication*, New Riders, 2016.
- [67] D. Jacobson, G. Brail e D. Woods, *APIs: A Strategy Guide*, Sebastopol: O'Reilly Media, 2012.
- [68] B. Lorica e P. Nathan, "The State of Data Quality in 2020," O'Reilly Media, Califórnia, 2020.
- [69] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J. Crespo e D. Dennison, "Hidden Technical Debt in Machine Learning Systems," em *Advances in Neural Information Processing Systems*, 2015.

- [70] F. Chollet, Deep Learning with Python, Shelter Island: Manning Publications, 2018.
- [71] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, n° 1, pp. 5-32, 2001.
- [72] M. Kuhn e K. Johnson, Applied Predictive Modeling, New York: Springer, 2013.
- [73] G. James, D. Witten, T. Hastie e R. Tibshirani, An Introduction to Statistical Learning: with Applications in R, New York: Springer, 2013.
- [74] J. VanderPlas, Python Data Science Handbook, O'Reilly Media, 2016.
- [75] W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Jupyter, O'Reilly Media, 2022.