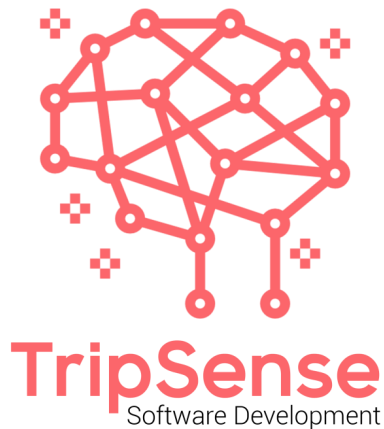


Sentiment Analysis - Turismo e Hotelaria



A Nossa **Equipa** e Competências



Pedro Cunha

20 anos – Finalista de IG

Bastante empenhado, persistente e dedicado nas tarefas que lhe são propostas, bem como nos seus objetivos. Gosta de trabalhar em equipa e é bastante flexível e compreensivo nas tomadas de decisão.

Possui nível C1 de Inglês e interesse nas áreas de marketing, automóvel e de data analytics.

SQL



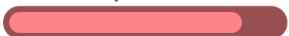
Python



Power BI



Photoshop



João Ramos

20 anos - Finalista de IG

Líder criativo muito organizado, com talento para apresentar ideias e soluções únicas. Eficiente a planear e gerir tarefas o que torna um ativo valioso em trabalho de equipa.

Possui ainda nível 4 em manutenção de equipamentos informáticos e certificação de reparador Apple.

SQL



Python



Power BI



UML





Introdução ao Projeto

- Plataforma Online de **Listening/ Sentiment Analysis** dos consumidores enquadrado ao setor do turismo e hotelaria.
- Análise baseada nas *Reviews* dos clientes em websites de reservas hoteleiras.
- Recurso a **Python (Análise)** e **PowerBI (Visualização)**.
- Dashboards/Insights de apoio à Gestão

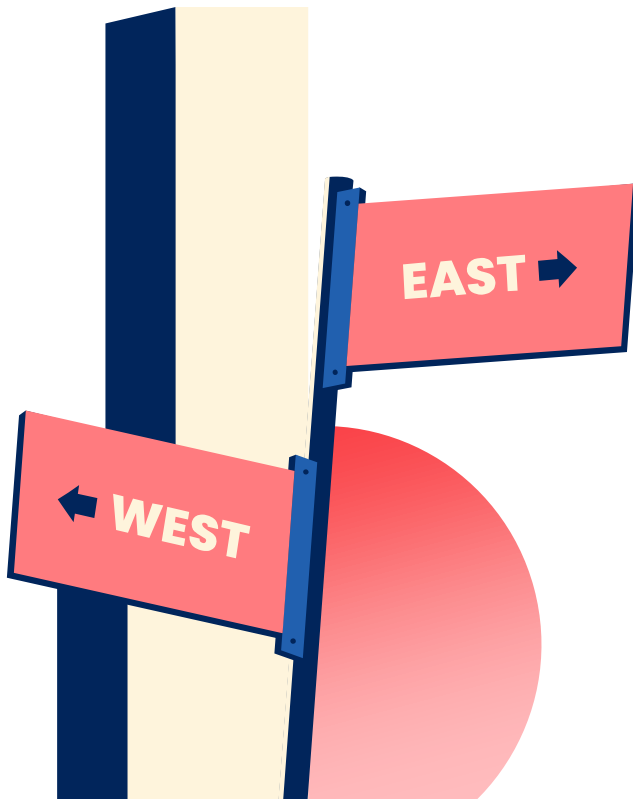
Motivações e Objetivos

- Setor em crescimento e com potencial
- Interpretar e analisar os dados recorrendo a Python e apresentar a informação de fácil interpretação com recurso a PowerBI
- Fornecer informações valiosas para a tomada de decisões
- Fornecer insights para a indústria hoteleira como um todo, incluindo tendências e padrões emergentes
- Quais são as percepções/opiniões dos clientes sobre hotéis específicos?
- Perceber quais são as tendências gerais no setor hoteleiro
- É possível melhorar a satisfação dos clientes identificando os pontos fortes e fracos dos hotéis analisando a sua opinião?
- Como ajudar os hotéis a identificar áreas de melhoria, com base nas percepções e opiniões dos clientes?

Quem são os Stakeholders?

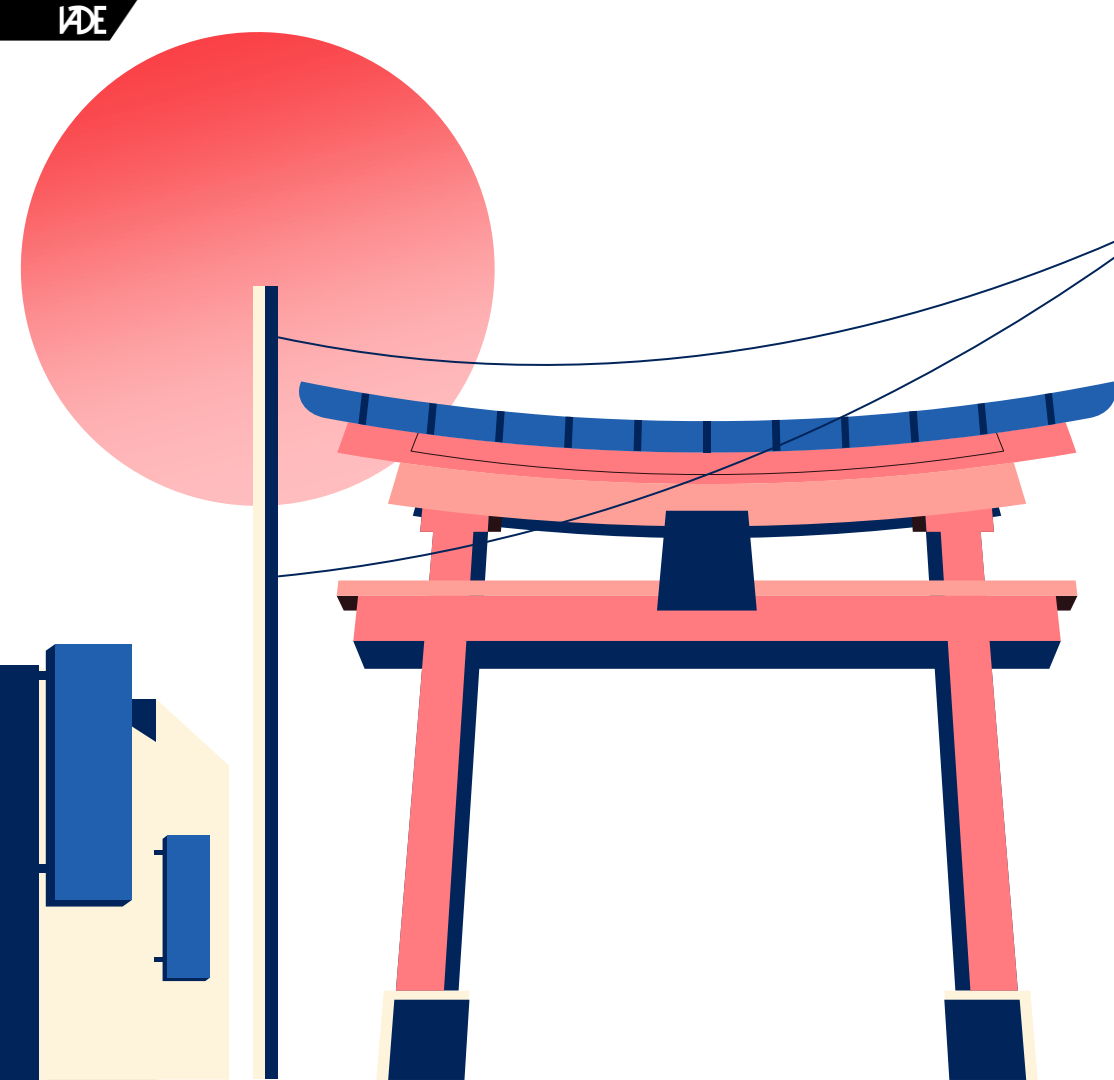
Equipa Projeto

- Desenvolver e documentar todo o projeto
- Conceber e implementar as regras de negócio
- Avaliar os resultados da análise de sentimento
- Apresentar as conclusões de forma gráfica ao cliente
- Arquitetura e Implementação



Gestor Projeto

- Moderar o desenvolvimento do projeto
- Auxiliar na realização da documentação
- Orientação das melhores práticas a seguir e implementar
- Avaliar os resultados do projeto em relação aos objetivos estabelecidos



01 Análise de Situação

Quem é o Grupo Alvo?

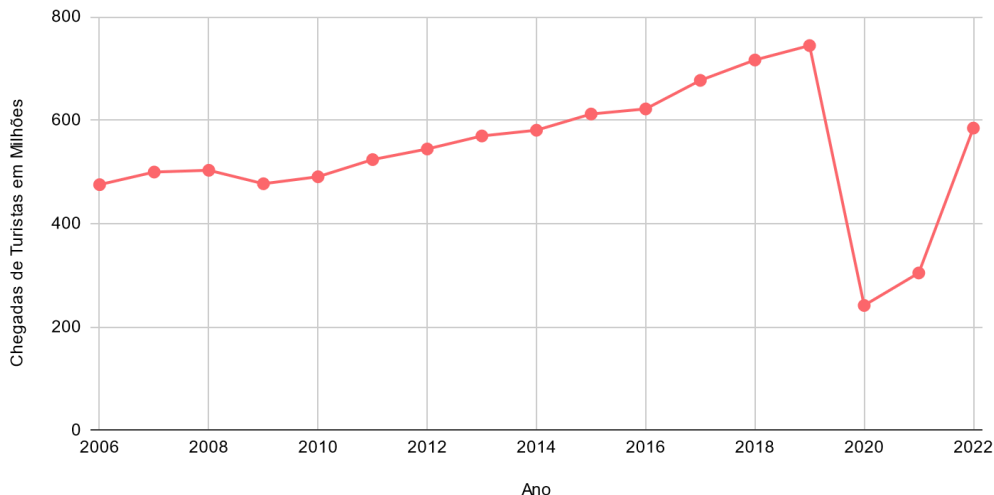
- **Proprietários e Gerentes de Empreendimentos Turísticos – Grupo Alvo a alcançar numa fase inicial**

Grupos-Alvo a atingir no futuro:

- Investidores e analistas de mercado
- Profissionais de marketing digital



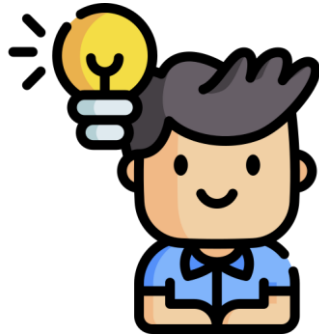
Número de Chegadas de Turistas Internacionais à Europa



- Em 2019, segundo dados do Statista, **cerca de 744.5 milhões de turistas internacionais viajaram para a Europa**. Desta maneira, numa fase mais avançada, pretendemos implementar o projeto ao nível europeu.

Uma **Pergunta** para a Plateia!

- **Alguém sabe quais são os destinos mais populares entre os turistas?**



Quem é o Grupo Alvo?

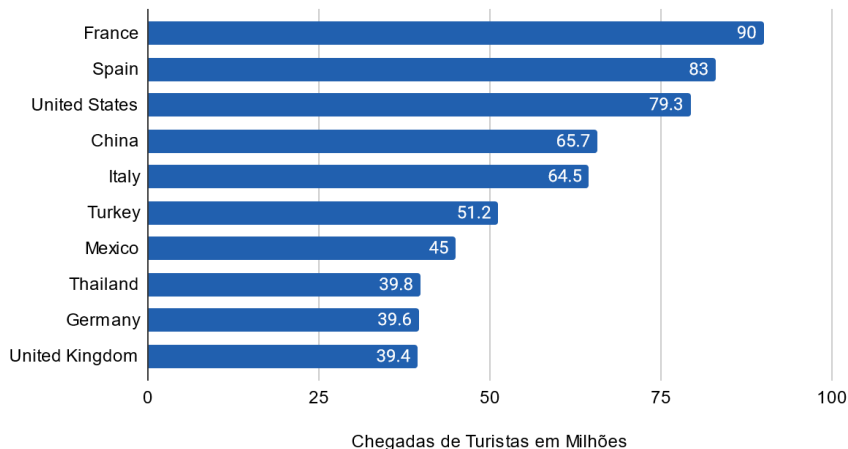
- **Proprietários e Gerentes de Empreendimentos Turísticos – Grupo Alvo a alcançar numa fase inicial**

Grupos-Alvo a atingir no futuro:

- Investidores e analistas de mercado
- Profissionais de marketing digital



Países Mais Visitados em 2019



- **6 dos 10 países mais visitados são europeus** - sustentando o objetivo de implementar o **projeto a nível europeu**. Contudo, antes de lançar oficialmente o projeto, seria ideal realizar um **projeto piloto a nível nacional**.

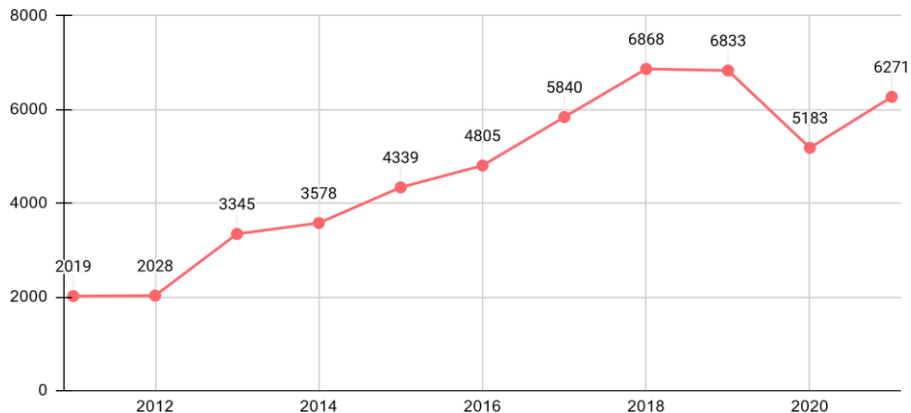
Quem é o Grupo Alvo?

- **Proprietários e Gerentes de Empreendimentos Turísticos – Grupo Alvo a alcançar numa fase inicial**

Grupos-Alvo a atingir no futuro:

- Investidores e analistas de mercado
- Profissionais de marketing digital

Número de Alojamentos Turísticos por Ano em Portugal



Dados: [Pordata](#)



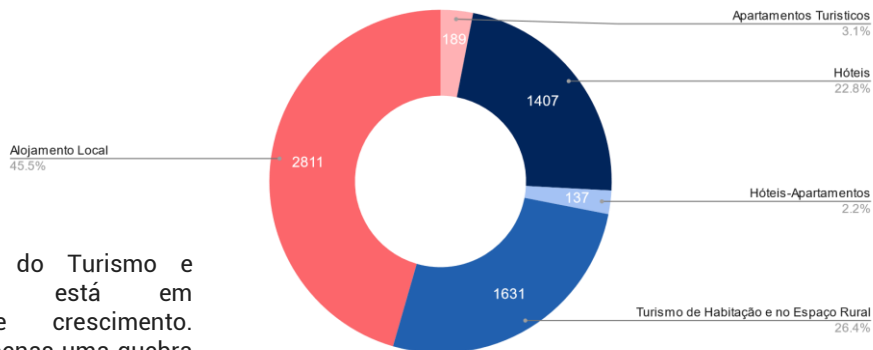
Quem é o Grupo Alvo?

- **Proprietários e Gerentes de Empreendimentos Turísticos – Grupo Alvo a alcançar numa fase inicial**

Grupos-Alvo a atingir no futuro:

- Investidores e analistas de mercado
- Profissionais de marketing digital

Número de Alojamentos Turísticos por Categoria



- O setor do Turismo e Hotelaria está em constante crescimento. Houve apenas uma quebra na altura do Covid-19, devido às regras que foram impostas.














Qual é o nosso Modelo de Negócio?

- **Software as a Service (SaaS):** utilizadores acedem a apps pela Internet e pagam uma taxa de subscrição, em vez de comprar uma licença vitalícia.
- **Vendas de acesso à plataforma:** Cobrar uma taxa mensal de acesso à plataforma (vários planos de subscrição).
- **Relatórios/insights personalizados:** Ajustado às necessidades de cada cliente. Podem ser vendidos separadamente ou como complemento.

Apresentação do Lean Canvas

| | | | | |
|---|---|--|--|---|
| <div>Parcerias Principais</div> <div></div> <div><ul style="list-style-type: none">• Booking• Trivago• Pestana Hotéis• Accor• Vila Galé• Ibis• Marriott• SANA</div> | <div>Atividades-chave</div> <div></div> <div><ul style="list-style-type: none">• Análise de Sentimento• Desenvolver KPIs• Representação visual em Dashboard</div> <div>Recursos Principais</div> <div></div> <div><ul style="list-style-type: none">• Reviews em sites de Reservas Hoteleiras</div> | <div>Proposta de Valor</div> <div></div> <div><ul style="list-style-type: none">• Oferecer uma solução que permita que os nossos clientes (hotéis) identifiquem os pontos a melhorar de forma eficaz• Ajudar na tomada de decisão• Potenciar crescimento dos clientes• Personalização face às necessidades</div> | <div>Relacionamento com Clientes</div> <div></div> <div><ul style="list-style-type: none">• Relação de longo prazo com os clientes fornecendo atualizações regulares e análises precisas e relevantes.</div> <div>Canais</div> <div></div> <div><ul style="list-style-type: none">• Redes sociais• Eventos e Feiras• Website• Pitch Formal Direto</div> | <div>Segmentos de Clientes</div> <div></div> <div><ul style="list-style-type: none">• Hotéis na Europa interessados em melhorar a satisfação do cliente para potenciar o seu crescimento e que pretendem obter insights para ajudar a tomada de decisões</div> |
| <div>Estrutura de Custo</div> <div></div> <div><ul style="list-style-type: none">• Documento Excel em anexo</div> | | <div>Fontes de Receita</div> <div></div> <div><ul style="list-style-type: none">• Planos de Subscrição (Individual, Multiple, BigChains)• Serviços adicionais• + Detalhes documento Excel anexado</div> | | |

Planos de **Subscrição**

| Plano | Recursos do Plano | Target | Preço |
|--------------|--|--|-------|
| Individual | Dashboard para único estabelecimento | Hotéis que possuam um único estabelecimento | 950€ |
| Individual + | Dashboard para único estabelecimento + Dashboard global zona (cidade) | | 1100€ |
| Multiple | Dashboard para mais do que um estabelecimento | Pequenas cadeias hoteleiras nacionais | 1800€ |
| Multiple + | Dashboard para mais do que um estabelecimento + Dashboard global zonas (cidades) | | 1950€ |
| Big Chains | Dashboard para cadeias de hotéis | Grandes cadeias hoteleiras sediadas em vários países | 3600€ |
| Big Chains + | Dashboard para cadeias de hotéis + Dashboard Global (países) | | 3750€ |

Serviços **Adicionais**

| Serviço | Recursos do Plano | Target | Preço |
|-------------------|---|--|---------|
| Análise de cidade | Relatório de uma cidade específica | Hotel que queira expandir para outra cidade | 1200€ |
| Análise de país | Relatório de uma país específico | Cadeia que queira entrar noutro mercado | 2400€ |
| Personalizado | Relatório personalizado com base nos pedidos do cliente | Hotéis que queiram analisar pontos muito específicos | 10-20k€ |

Demonstração de Resultados

Demonstração de Resultados Previsional

| | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 |
|---|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Vendas e serviços prestados | 231,613,050 | 233,037,722 | 237,864,710 | 252,992,960 | 254,938,085 | 255,738,365 |
| Subsídios à Exploração | | | | | | |
| Ganhos/perdas imputados de subsidiárias, associadas e empreendimentos conjuntos | | | | | | |
| Variação nos inventários da produção | | | | | | |
| Trabalhos para a própria entidade | | | | | | |
| CMVMC | | | | | | |
| Fornecimento e serviços externos | 508,409 | 518,881 | 529,614 | 539,147 | 548,043 | 554,620 |
| Gastos com o pessoal | 8,054,022 | 9,251,840 | 10,389,440 | 12,029,600 | 13,482,424 | 14,265,632 |
| Imparidade de inventários (perdas/reversões) | | | | | | |
| Imparidade de dívidas a receber (perdas/reversões) | 5,104,585 | 5,139,632 | 5,258,375 | 5,630,530 | 5,669,433 | 5,685,439 |
| Provisões (aumentos/reduções) | | | | | | |
| Imparidade de investimentos não depreciáveis/amortizáveis (perdas/reversões) | | | | | | |
| Aumentos/reduções de justo valor | | | | | | |
| Outros rendimentos e ganhos | | | | | | |
| Outros gastos e perdas | | | | | | |
| EBITDA (Resultado antes de depreciações, gastos de financiamento e impostos) | 217,946,035 | 218,127,370 | 221,687,281 | 234,793,683 | 235,238,185 | 235,232,675 |
| Gastos/reversões de depreciação e amortização | 95,500 | 191,000 | 286,500 | 372,500 | 470,000 | 567,500 |
| Imparidade de activos depreciáveis/amortizáveis (perdas/reversões) | | | | | | |
| EBIT (Resultado Operacional) | 217,850,535 | 217,936,370 | 221,400,781 | 234,421,183 | 234,768,185 | 234,665,175 |
| Juros e rendimentos similares obtidos | | | | | | |
| Juros e gastos similares suportados | | | | | | |
| RESULTADO ANTES DE IMPOSTOS | 217,850,535 | 217,936,370 | 221,400,781 | 234,421,183 | 234,768,185 | 234,665,175 |
| Imposto sobre o rendimento do período | 54,462,634 | 54,484,093 | 55,350,195 | 58,605,296 | 58,692,046 | 58,666,294 |
| RESULTADO LÍQUIDO DO PERÍODO | 163,387,901 | 163,452,278 | 166,050,586 | 175,815,887 | 176,076,139 | 175,998,881 |

Concorrentes Diretos



- Fundada em 1998
- Sede em Nova Iorque
- Soluções de Market Intelligence e gestão de social media para empresas
- Ajuda empresas a monitorizar presença na web
- Acompanhar a atividade dos concorrentes
- Entender tendências de mercado e identificar oportunidades de negócio
- Entender sentimento do consumidor e tomar medidas para melhorar o engagement



- Fundada em 2004
- Sede na Califórnia
- Soluções de Market Intelligence e análise de dados
- Recolher e analisar grandes volumes de dados para ajudar empresas a entender melhor clientes, concorrentes e mercados
- Compreender melhor a opinião sobre marca, produtos ou serviços
- Identificar oportunidades de mercado e melhorar eficácia de campanhas de marketing e comunicação



Brandwatch

- Fundada em 2005
- Sede no Reino Unido
- Plataforma de Market Intelligence e gestão de social media
- Ajuda as empresas a monitorizar e analisar a presença da marca nas redes sociais
- Fornecer informações úteis sobre a opinião do consumidor, concorrentes e tendências de mercado
- Grandes volumes de dados
- Fornecer insights valiosos para a tomada de decisões estratégicas de negócios

Modelo Negócio – Concorrência



Brandwatch

- Software as a Service (SaaS)
- Venda de subscrições da sua plataforma
- Vários planos de subscrição consoante as necessidades dos clientes
- Serviços de consultoria para ajudar os clientes a aproveitar a plataforma ao máximo (integração, treino e suporte técnico)
- Personalização da plataforma para atender às necessidades específicas dos clientes
- Ampla gama de setores (tecnologia, finanças, bens de consumo, saúde, etc)
- Modelo de negócio escalável



Preços – Concorrência



- Varia dependendo do plano e dos recursos específicos
- **Starter** começa em torno de \$1.000/mês
- **Pro** começa em torno de \$2.000/mês
- **Ultimate** preços **sob orçamento** conforme as necessidades específicas do cliente



- Varia dependendo do plano e dos recursos específicos
- **Basic** cerca de \$1.000/mês
- **Standart** cerca de \$2.000/mês
- **Pro** cerca de \$4.000/mês
- **Enterprise** preço **sob orçamento** e varia de acordo com as necessidades do cliente



Brandwatch

- Varia dependendo do plano e dos recursos específicos
- **Essentials** começa em torno de \$800/mês
- **Pro** começa em torno de \$1.200/mês
- **Enterprise** tem **preços sob orçamento** dependendo das necessidades específicas do cliente

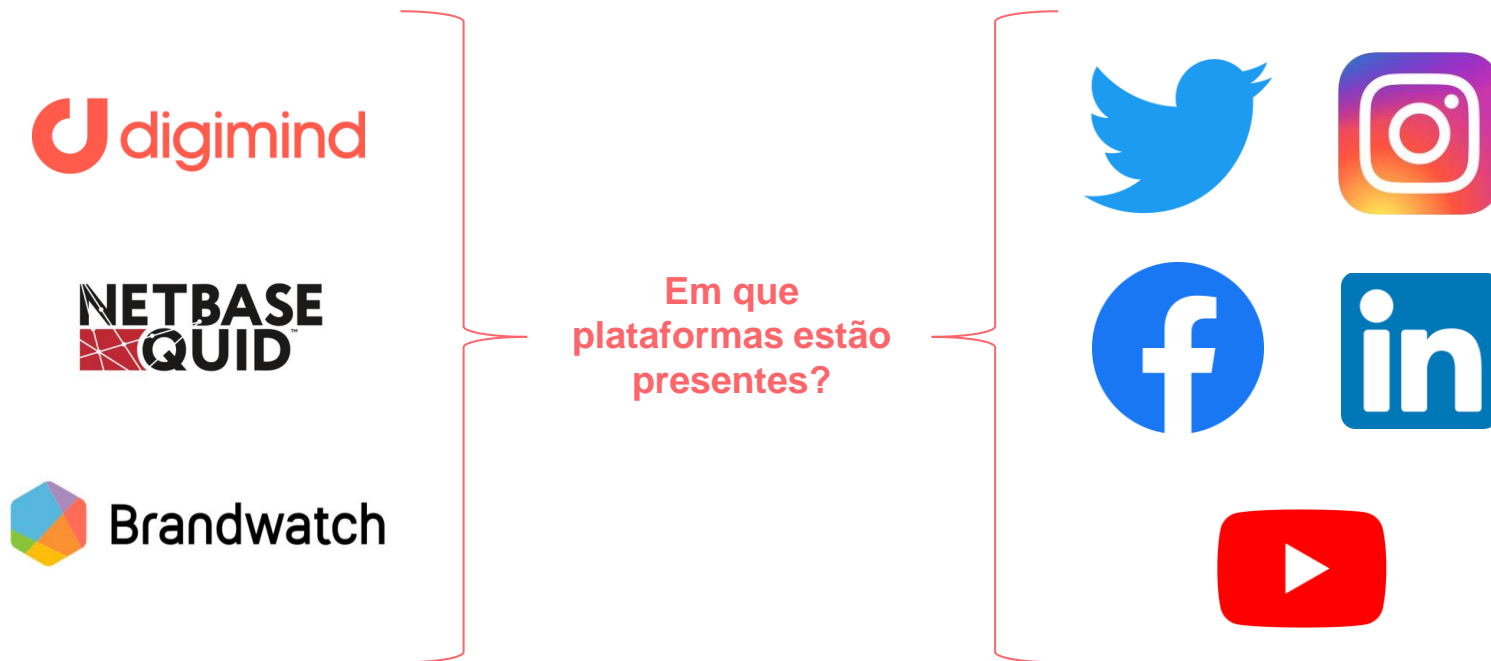
Canais Distribuição - Concorrência



Brandwatch

- **Site da Plataforma:** Principal fonte de informação e aquisição de novos clientes
- **Redes Sociais:** Para partilhar conteúdo relacionado e para promover a empresa e seus produtos
- **Parcerias:** Agências de marketing digital e de gestão de social media. Permitem alcançar público mais amplo e oferecer soluções integradas aos clientes
- **Eventos e Conferências:** Promover a empresa e seus produtos

Social Media - Concorrência



- Presentes nas principais plataformas, **usando cada uma delas de maneira estratégica e adaptada ao público-alvo específico** de cada rede social para se comunicarem com os seguidores, clientes.

Benchmark - Concorrência

Ao efetuar o Benchmark, foi possível verificar que os nossos pontos de diferenciação relativamente à concorrência são:

- **Monitorização de Sites de Reviews**
- **Foco num Único Setor**
- **Otimização de Algoritmos para o setor**
- **Oferta de serviços adicionais sem subscrição**
- **Controlo Humano sobre a Tecnologia**

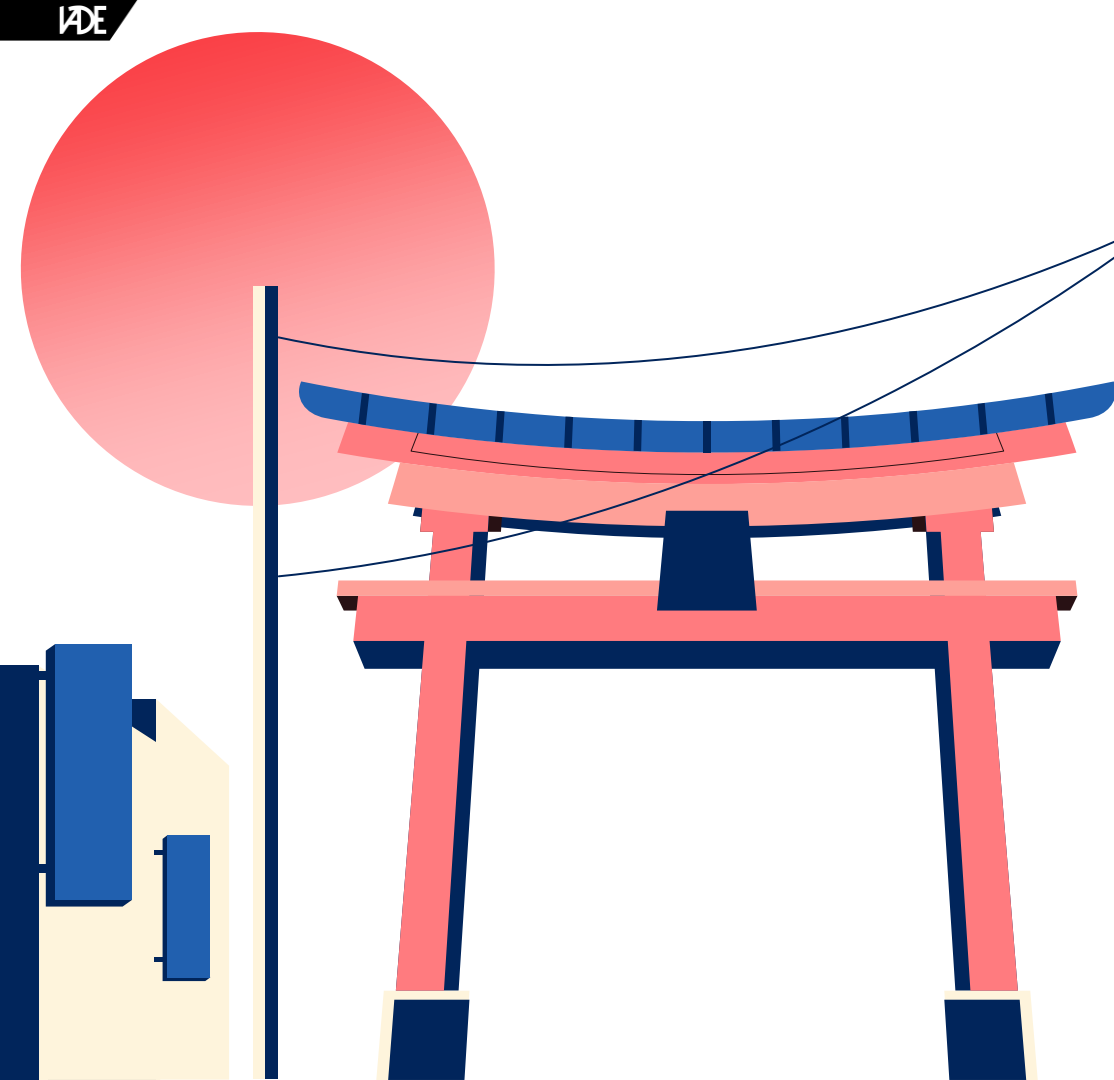
Podemos verificar os aspetos mencionados na tabela de Benchmark abaixo:

| # | Funcionalidade | Digimind | NetBased Quid | BrandWatch | TripSense | Total |
|----|---|----------|---------------|------------|-----------|-------|
| | Empresa | | | | | |
| 8 | Monitorização de Sites de Reviews | 0 | 1 | 0 | 1 | 2 |
| 24 | Foco num único setor | 0 | 0 | 0 | 1 | 1 |
| 36 | Algoritmos otimizados para cada setor | 0 | 0 | 0 | 1 | 1 |
| 41 | Oferta de serviços adicionais sem necessidade de subscrição | 0 | 0 | 0 | 1 | 1 |
| 43 | Controlo Humano sobre a Tecnologia | 0 | 0 | 0 | 1 | 1 |



02

Arquitetura da Solução



Regras de Negócio

| #RN | Descrição da Regra de Negócio |
|------|--|
| RN01 | Quando acaba a subscrição, a plataforma fica disponível para consulta durante mais 1 semana, passado este tempo: o cliente continua a ter acesso, mas a informação fica desfocada e é mostrado um pop up para renovar a subscrição. |
| RN02 | Uma das visualizações será um conjunto de <i>cards</i> com positivo/negativo/neutro com as palavras mais relevantes por baixo dos mesmos e ainda o ranking do cliente em relação aos demais. |
| RN03 | Se a data dos dados presentes na base de dados for superior a um trimestre é efetuada uma nova requisição de dados. |
| RN04 | O cliente não necessita de ter um plano de subscrição para requisitar um serviço adicional, pois estes são pagos separadamente. |
| RN05 | Fazer upgrade de plano de subscrição é possível a qualquer momento, tendo apenas de pagar a diferença entre os planos. |
| RN06 | A plataforma deve garantir a privacidade e a segurança dos dados recolhidos, em conformidade com as leis e regulamentações aplicáveis, como o Regulamento Geral de Proteção de Dados (PGPR), e implementar medidas de segurança adequadas para proteger os dados contra acesso não autorizado. |
| RN07 | A plataforma deve permitir que os utilizadores personalizem os alertas e notificações com base em critérios específicos, como palavras-chave relevantes, avaliações com determinadas classificações, entre outros, para que possam ser prontamente informados sobre eventos relevantes e agir de forma proativa. |
| RN08 | Caso a subscrição seja anulada, o cliente perde imediatamente o acesso à plataforma. |

Levantamento de Requisitos

| #RF | Requisito Funcional | Descrição |
|------|-------------------------|---|
| RF01 | Registrar na plataforma | <p>Permite que novos utilizadores criem uma conta e acessem aos recursos e serviços oferecidos pela plataforma.</p> <p>A interface de registo deve ser intuitiva e fácil de usar, permitindo que os utilizadores insiram as informações necessárias para criar uma conta na plataforma, como o nome, endereço de e-mail e password, bem como outros detalhes relevantes, como data de nascimento, localização ou preferências de conta.</p> <p>É importante fornecer informações claras sobre os termos e condições da plataforma, políticas de privacidade e consentimento de utilização de dados durante o processo de registo. Os utilizadores devem ser informados sobre como os seus dados serão utilizados e ter a opção de consentir ou recusar a partilha de informações pessoais.</p> <p>Após o registo, os utilizadores devem receber uma confirmação de conta, através de um e-mail, garantindo a segurança da conta e a validade do endereço de e-mail fornecido. É importante também permitir que os utilizadores possam verificar a sua conta através de outras opções, como número de telefone ou autenticação de dois fatores.</p> <p>Além disso, a plataforma deve ter uma política de proteção de dados robusta, garantindo a segurança e a privacidade das informações dos utilizadores. Isso pode incluir medidas como criptografia de dados, acesso seguro às informações do utilizador e conformidade com as regulamentações de proteção de dados aplicáveis, como o Regulamento Geral de Proteção de Dados (RGPD).</p> <p>A funcionalidade de registo pode ser integrada com outras funcionalidades da plataforma, como a capacidade de personalizar o perfil do utilizador e subscrever a newsletter. É importante fornecer uma experiência de registo otimizada, que permita aos utilizadores criar uma conta de forma rápida e eficiente, e começar a desfrutar dos recursos da plataforma sem problemas.</p> |

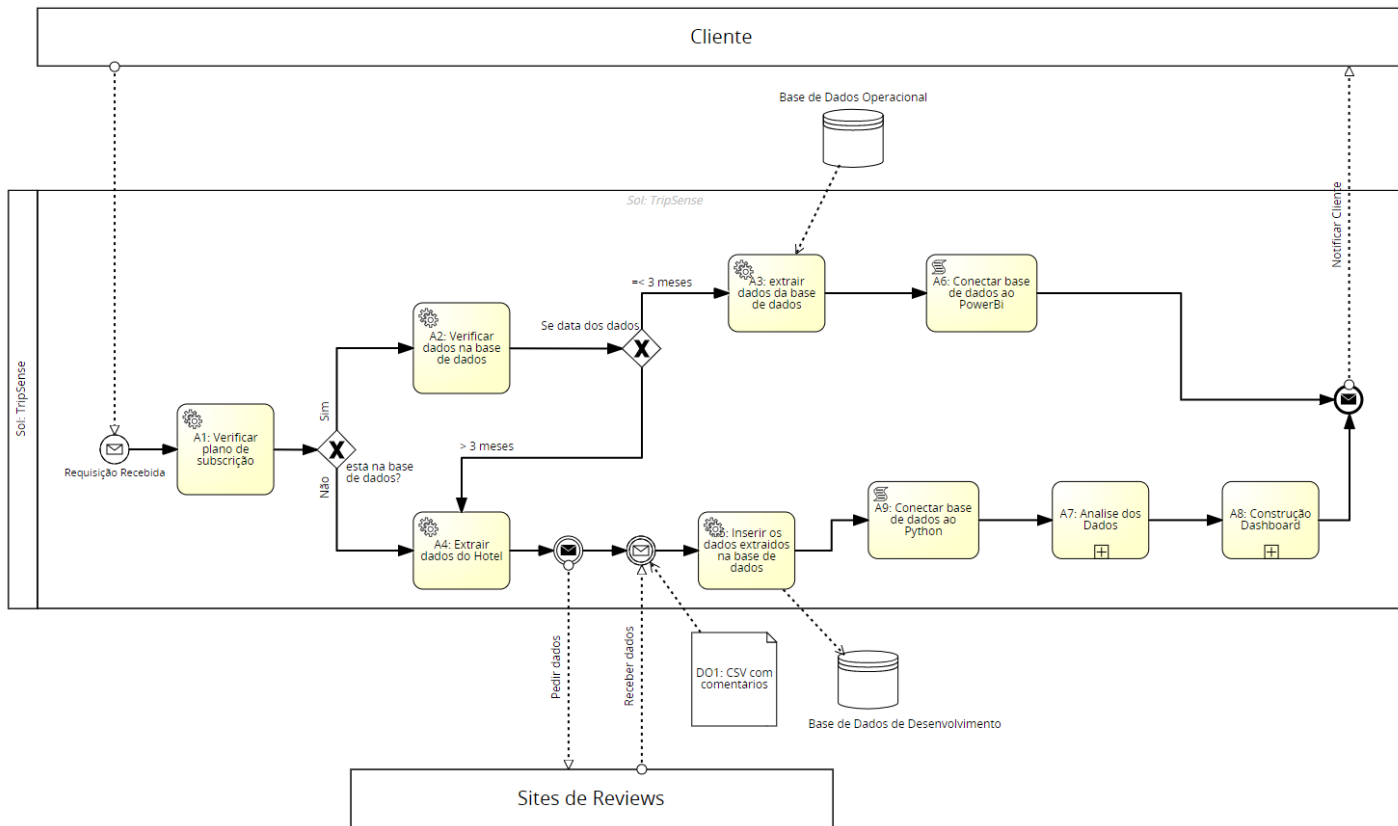
Processo de Negócio #1

Criação de Report para Cliente

- Uma requisição de report é recebida, sistema verifica o plano de subscrição do cliente e se os dados necessários estão disponíveis na base de dados. Se sim deve-se verificar a data dos dados – se for inferior a 3 meses deverá extrair-se os dados da base de dados operacional e seguidamente conectar a base de dados ao powerbi e fornecer o dashboard ao cliente, caso contrário, deve-se extrair os dados do hotel fazendo a requisição de dados à fonte externa com comentários em formato CSV.
- Seguidamente, estes dados são inseridos na base de dados de desenvolvimento, conecta-se ao Python e é efetuado o subprocesso de Análise dos Dados, que por sua vez leva ao subprocesso de construção Dashboard para que o mesmo seja entregue ao cliente.

Processo de Negócio #1

Criação de Report para Cliente



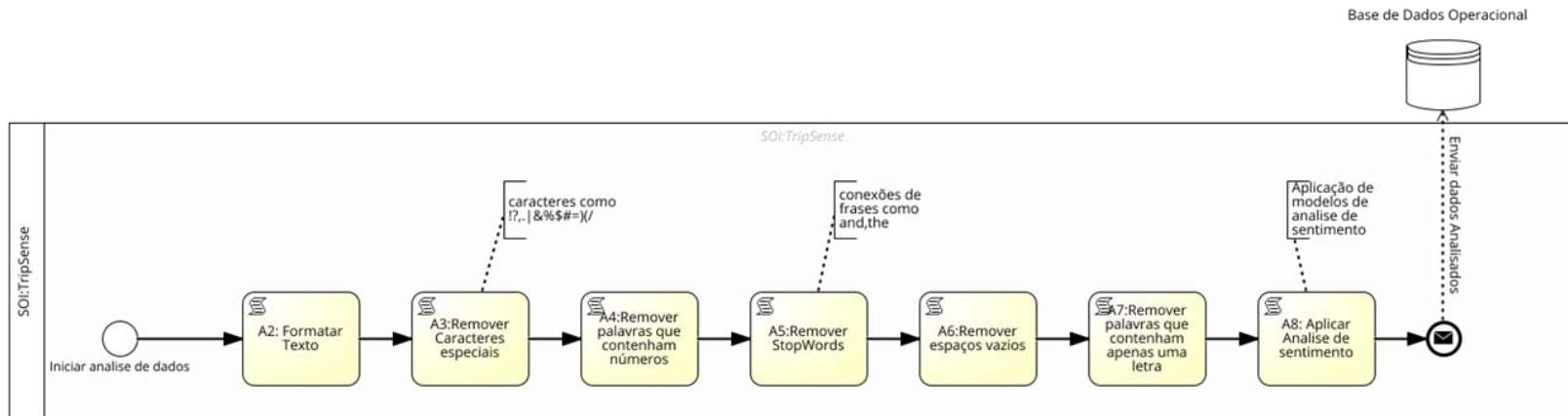
Processo de Negócio #2

Tratamento dos Dados

- Inicia-se o subprocesso de análise dos dados com a atividade de formatação do texto, ou seja, colocar as palavras com maiúscula no início e o resto em minúsculas, colocar acentuação correta e encontrar palavras com erros ortográficos e corrigir os mesmos. De seguida remove-se caracteres especiais como por exemplo: "!.?.,|&%\$#=(/\" e também as stopwords que se tratam de conexões de frases tais como "and" e "the", entre outras.
- Além disto, são removidos os espaços vazios, remove-se palavras que contenham apenas uma letra, aplica-se o modelo de análise de sentimento e após isto envia-se os dados analisados para a base de dados operacional para poderem ser utilizados no Python e no PowerBi.

Processo de Negócio #2

Tratamento dos Dados



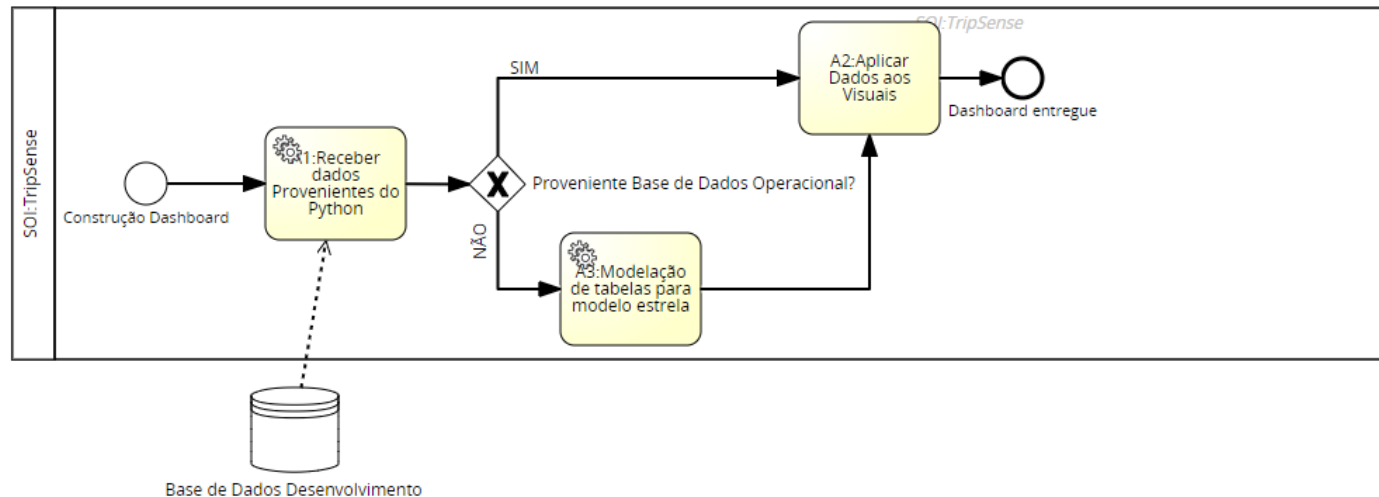
Processo de Negócio #3

Construção do Dashboard

- O processo de construção do Dashboard inicia quando os dados provenientes do Python são recebidos, se estes dados não vierem da base de dados operacional faz-se a criação do modelo em PowerBI e aplica-se os visuais caso sejam provenientes da base de dados operacional, de seguida são aplicados os visuais e o dashboard é entregue ao cliente.

Processo de Negócio #3

Construção do Dashboard



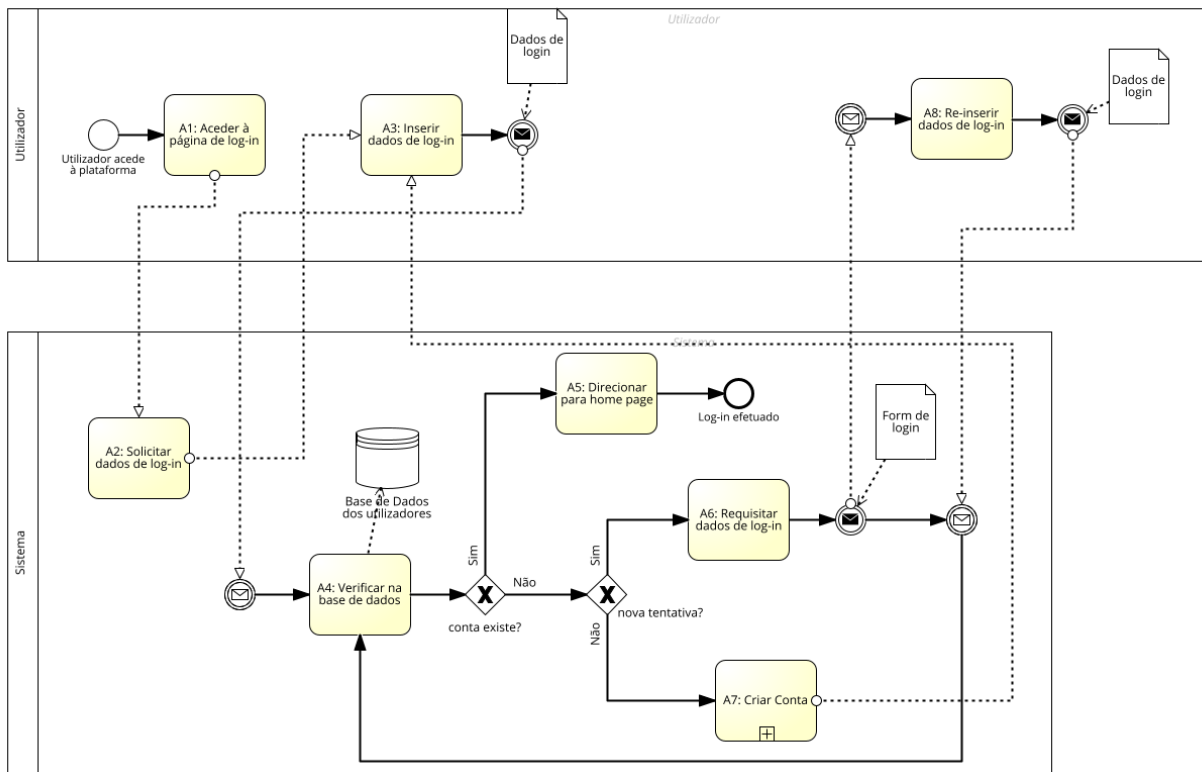
Processo de Negócio #4

Autenticação do utilizador

- O processo de efetuar log-in inicia quando o utilizador acessa a plataforma, de seguida deverá aceder à página de log-in onde o sistema solicita ao utilizador que insira as suas credenciais de log-in, como nome de utilizador ou endereço de e-mail, password e código 2FA, uma vez inseridos, o sistema recebe as credenciais de log-in inseridas pelo utilizador.
- Após isto, o sistema verifica se as informações introduzidas pelo utilizador estão corretas e se correspondem a uma conta existente na base de dados de seguida direciona o utilizador para a sua home page do dashboard.
- Caso o utilizador não possua uma conta na plataforma, o processo segue para um subprocesso de criar conta.

Processo de Negócio #4

Autenticação do utilizador



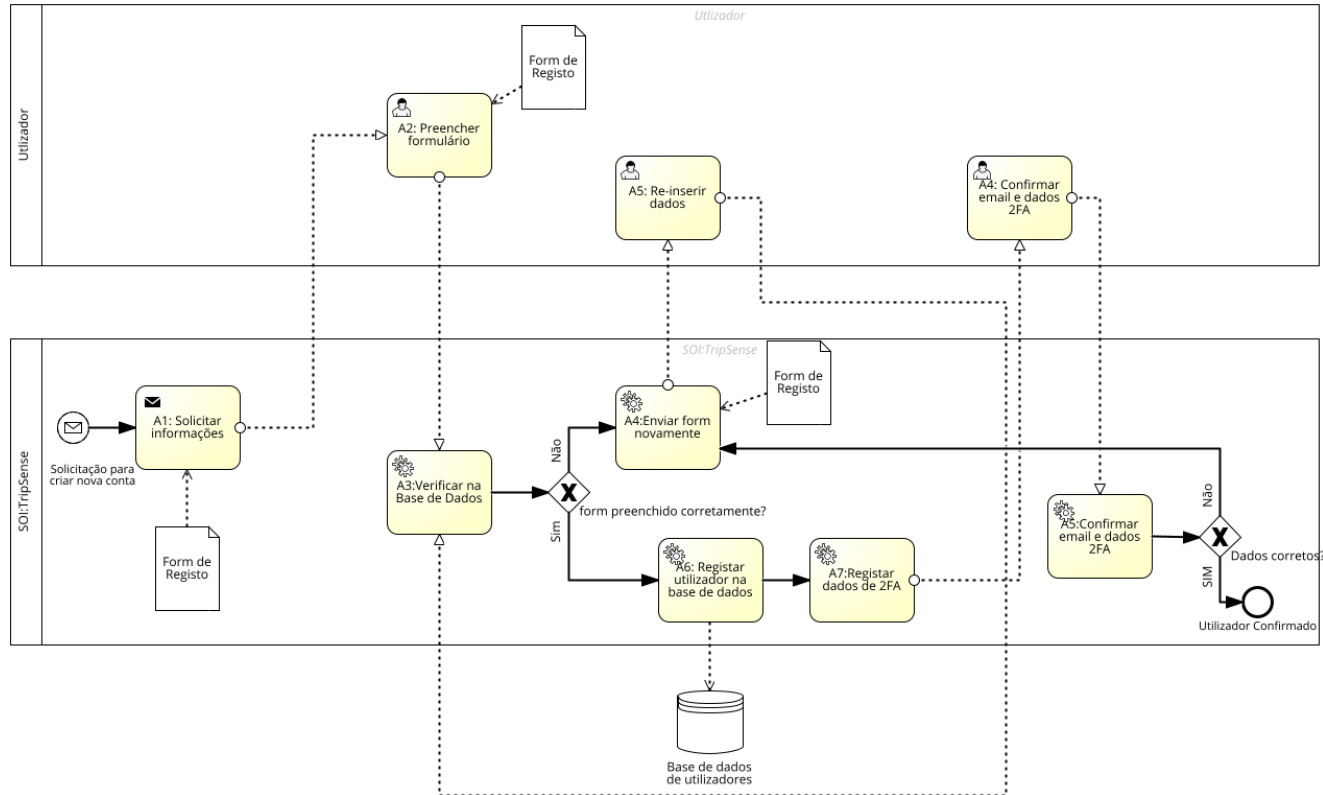
Processo de Negócio #5

Criar Conta

- O subprocesso de criar conta solicita ao utilizador as informações necessárias para criação da mesma, como nome, endereço de e-mail, senha, idade, nome do hotel, localização, entre outras, que são recebidas pelo sistema assim que o formulário é submetido. Depois de este passo ser concluído com sucesso, a conta é assim criada e registada na base de dados da plataforma.
- No entanto, para que a conta seja confirmada e o processo de registo seja concluído, é necessário que o utilizador verifique a sua caixa de e-mail e clique no link em anexo, para que desta maneira seja possível ter a certeza que os dados inseridos são verdadeiros, desta maneira, o subprocesso de criar conta é concluído e o processo de efetuar log-in retorna ao ponto de verificação de conta existente e se o utilizador possui uma conta existente ou criou uma nova com sucesso, o sistema concede acesso à plataforma.
- Caso as credenciais de log-in do utilizador não sejam autenticadas ou o utilizador não forneça as informações necessárias para criar uma conta, o processo de efetuar log-in pode falhar e ser repetido até que as informações correspondam.

Processo de Negócio #5

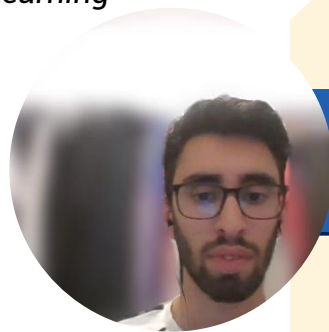
Criar Conta



Dataset Utilizado - Kaggle

515K Hotel Reviews Data in Europe disponível no Kaggle

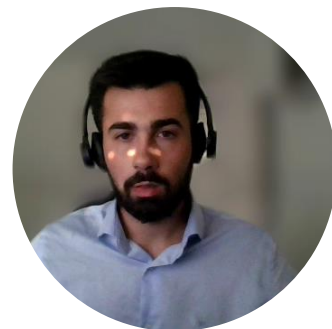
- Cerca de 515.000 avaliações de hotéis em toda a Europa, recolhidas através de diversas fontes online, incluindo o Booking.com e o TripAdvisor.
- Nome do hotel, localização, preço, categoria, as comodidades oferecidas, além de avaliações de utilizadores sobre a sua experiência no hotel.
- Limpeza rigorosa dos dados
- Modelos de *Sentiment Analysis* com base em algoritmos de *Machine Learning*



Frameworks Utilizadas

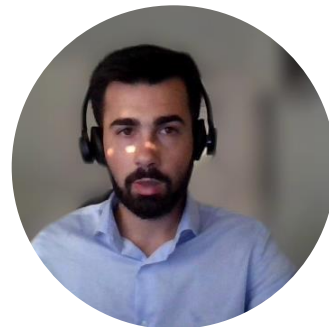
Python

- Principal tecnologia para limpar e analisar o dataset.
- O Python é uma linguagem de programação de alto nível bastante conhecida pela sua simplicidade e facilidade de uso, tornando-a uma escolha cada vez mais popular para muitos projetos de análise de dados.
- A limpeza de dados é uma etapa crítica em qualquer projeto desta natureza, e o Python oferece uma ampla gama de bibliotecas para auxiliar nesta tarefa.



Python - Bibliotecas Utilizadas

| Biblioteca | Versão | Use Case |
|---------------------------------|--------|---|
| Pandas | 1.1.5 | Carregar o dataset, selecionar colunas revelantes, exportar o dataframe como um arquivo csv |
| Re | 3.8.1 | Remover números |
| nltk.corpus | 3.8.2 | Importar stopwords |
| nltk.tokenize | 3.8.3 | Tokenizar palavras |
| nltk.stem | 3.8.4 | Lematizar palavras |
| textblob | 0.17.1 | Adicionar coluna de polaridade e de sentimento |
| sklearn.model_selection | 1.0.2 | Dividir o conjunto de dados em um conjunto de treino e teste |
| sklearn.feature_extraction.text | 1.0.3 | Vetorizar as avaliações positivas e negativas |
| sklearn.metrics | 1.0.4 | Avaliar a precisão do modelo |
| sklearn.naive_bayes | 1.0.5 | Treinar o modelo com o conjunto de dados de treino |
| Time | N/A | Medir o tempo de execução |
| string | N/A | Remover pontuação |



Frameworks Utilizadas

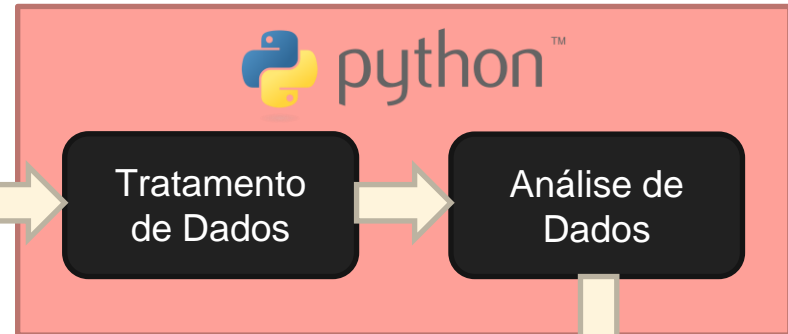
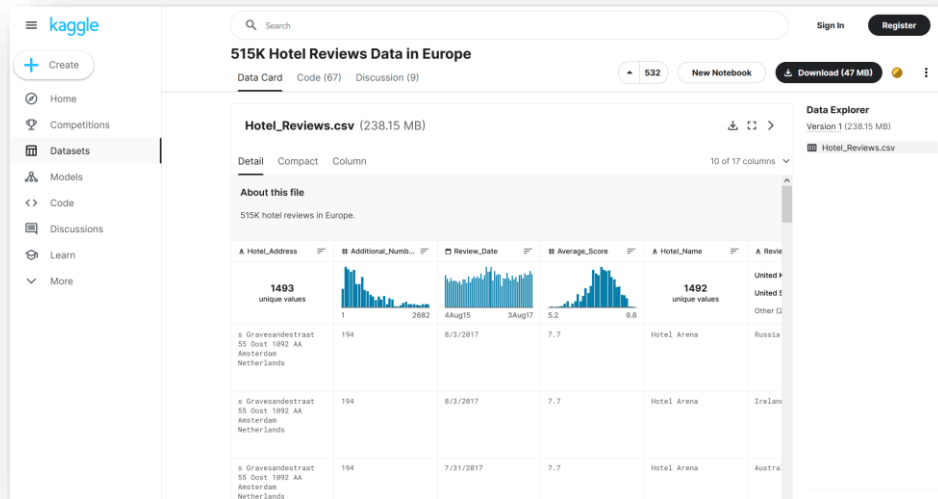
PowerBI

- Utilizado para desenvolver a vertente gráfica do projeto.
- O PowerBI é uma ferramenta de visualização de dados que permite criar dashboards interativos e relatórios a partir de dados de diversas fontes
- Depois de analisar os dados com o Python, os dados foram importados para o PowerBI para construir um dashboard que apresentasse os insights necessários ao cliente - incluindo gráficos e tabelas interativos que permitiram ao utilizador explorar e analisar os dados de diversas perspetivas.



Demonstração - Metodologia

Recolha de Dados – Dataset Kaggle



Apresentação
de Dados

Demonstração – Tratamento de dados

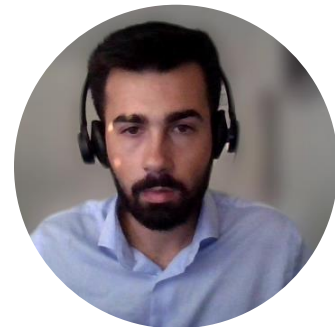
```
import pandas as pd
import re
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from textblob import TextBlob
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics import accuracy_score, classification_report
from sklearn.naive_bayes import MultinomialNB
import time
import string
start_time = time.time()

# Carregar o dataset
df = pd.read_csv('Hotel_Reviews.csv')

# Selecionar as colunas relevantes
df = df[['Hotel_Address', 'Review_Date', 'Hotel_Name', 'Reviewer_Nationality', 'Review', 'Reviewer_Score', 'Total_Number_of_Reviews_Reviewers']]

def clean_text(text):
    # Converter para minúsculas
    text = text.lower()
    # Remover números
    text = re.sub(r'\d+', '', text)
    # Remover pontuação
    text = text.translate(str.maketrans('', '', string.punctuation))
    # Remover stopwords e palavras com menos de 3 caracteres
    stop_words = set(stopwords.words('english'))
    words = word_tokenize(text)
    tokens = [word for word in words if len(word) > 2 and word not in stop_words]
    lemmatizer = WordNetLemmatizer()
    text = " ".join([lemmatizer.lemmatize(word) for word in tokens])
    # Remover duplicados
    df.drop_duplicates(inplace=True)

    return text
```

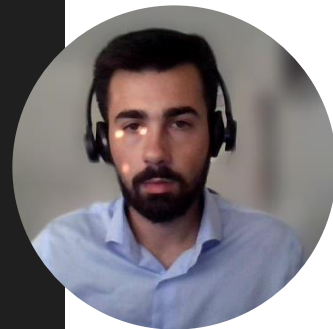


Demonstração – Categorizar as informações



```
# Definir as categorias
categories = {
    'room': ['room', 'bed', 'pillow', 'blanket', 'linen', 'towel'],
    'bathroom': ['bathroom', 'shower', 'bathtub', 'toilet', 'sink'],
    'cleanliness': ['clean', 'dirty', 'dust', 'stain', 'smell'],
    'service': ['service', 'staff', 'reception', 'concierge', 'bellboy'],
    'food': ['food', 'restaurant', 'breakfast', 'dinner', 'lunch', 'menu']
}

def get_categories(text):
    result = []
    for category, keywords in categories.items():
        for keyword in keywords:
            if keyword in text and category not in result:
                result.append(category)
                break
    return result
```



Demonstração – Algoritmo de Sentimental Analysis

```
# Adicionar coluna de sentimento
def get_sentiment(sentimo_num):
    if sentimo_num < 0:
        return 'Negativo'
    elif sentimo_num == 0:
        return 'Neutro'
    else:
        return 'Positivo'

df['Sentiment'] = df['Sentimo_num'].apply(get_sentiment)

# Remover linhas neutras
df_clean = df[df['Sentiment'] != 'Neutro']

# Dividir o conjunto de dados em conjunto de treinamento e teste
train_data, test_data = train_test_split(df_clean, test_size=0.3, random_state=42)

# Vetorizar as avaliações negativas e positivas
vectorizer = CountVectorizer(stop_words='english')
train_features = vectorizer.fit_transform(train_data['Negative_Review'])
test_features = vectorizer.transform(test_data['Negative_Review'])

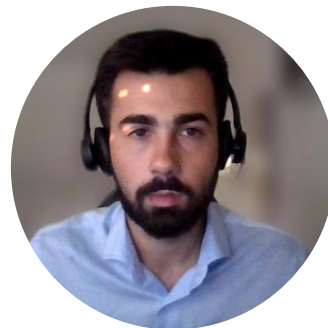
# Treinar o modelo com o conjunto de treino
nb = MultinomialNB()
nb.fit(train_features, train_data['Sentiment'])

# Testar o modelo com o conjunto de teste e avaliar sua precisão
predictions = nb.predict(test_features)
accuracy = accuracy_score(test_data['Sentiment'], predictions)
print("Precisão do modelo: {:.2f}%".format(accuracy * 100))
print(classification_report(test_data['Sentiment'], predictions))

# Mostrar o relatório de classificação
print(classification_report(test_data['Sentiment'], predictions))

# Exportar dataframe como um arquivo CSV
df_clean.to_csv('hotel_reviews_clean.csv', index=False)

end_time = time.time()
total_time = end_time - start_time
print("Tempo de execução do programa: {} minutos e {:.2f} segundos".format(int(total_time//60), total_time%60))
```



Nome do Hotel

Nacionalidade do Reviewer

Score do Reviewer

Num. Reviews do Reviewer

Data

88 Studios

All

All

1

355

Multiple selections

Total Reviews

Sentimento Positivo

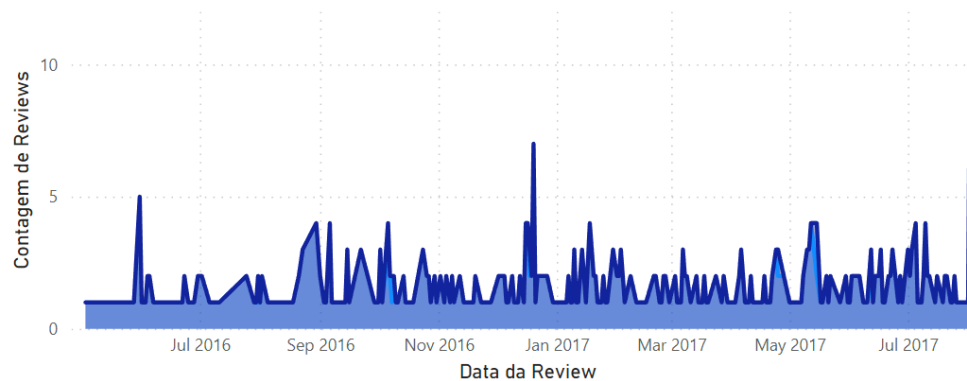
Sentimento Negativo

423

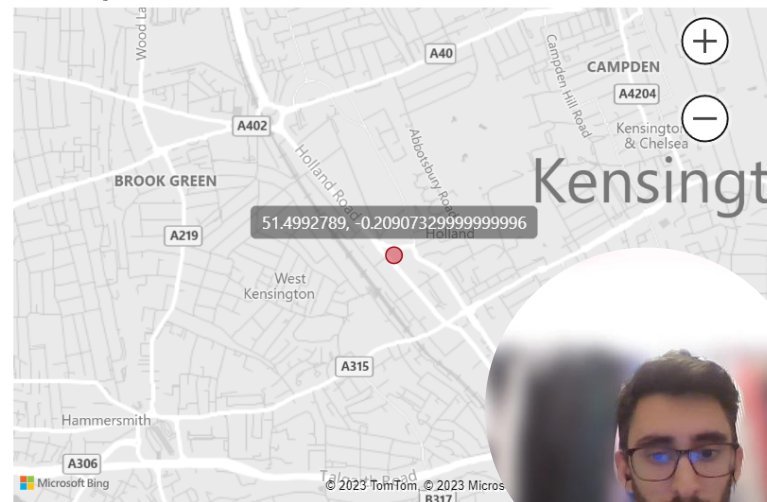
413

10

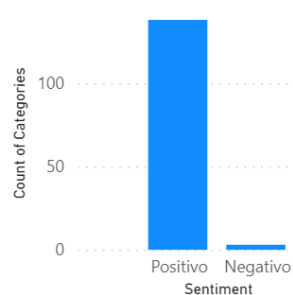
Total Reviews ao longo do Tempo



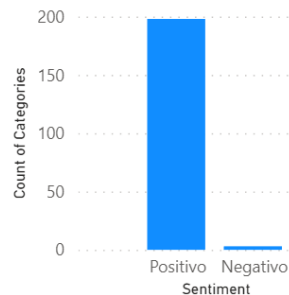
Localização do Hotel



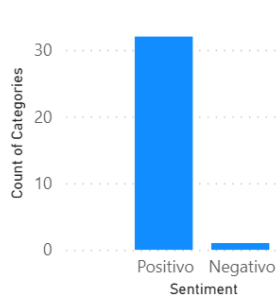
Casa de Banho



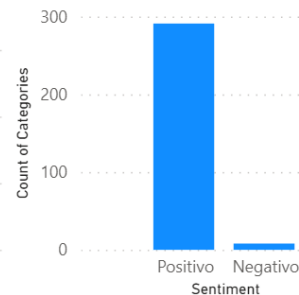
Limpeza



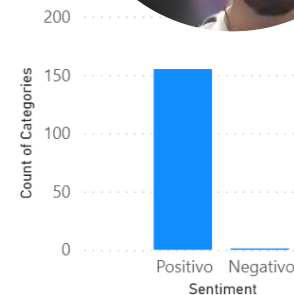
Comida



Quartos



Serviço

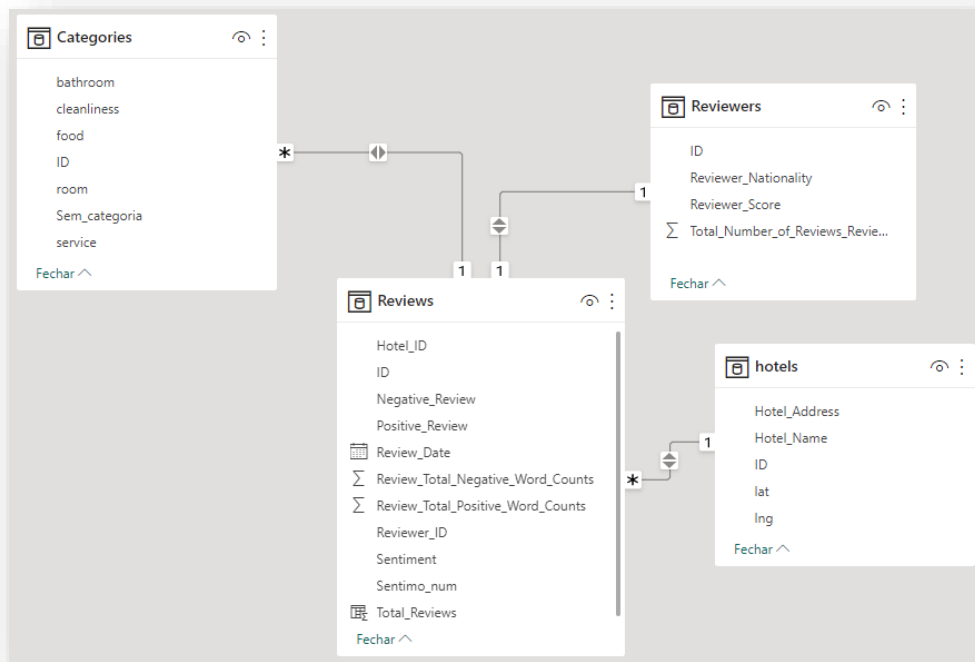


TripSense Demo, individual

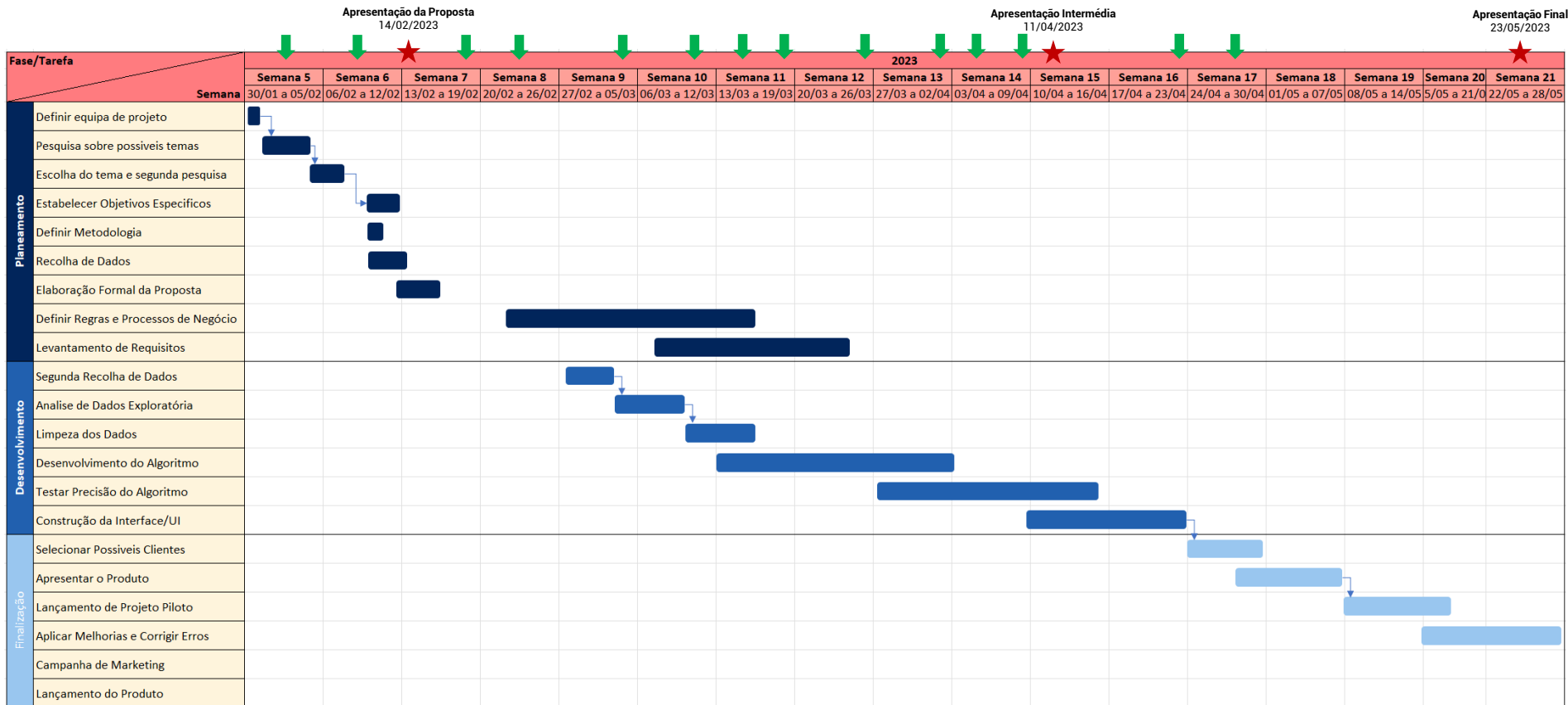
Data updated on 24/05/23, 18:53

Filters

PowerBI – Modelo de Dados ERD



Planeamento - Fases



Reuniões



Milestones

Referências Bibliográficas

1. V. Elango and G. Narayanan, "Sentiment Analysis for Hotel Reviews," Stanford University. [Online]. Available: <https://cs229.stanford.edu/proj2014/Vikram%20Elango,%20Govindrajan%20Narayanan,%20Sentiment%20Analysis%20for%20Hotel%20Reviews.pdf>. [Accessed: 14-Feb-2023].
2. "International Tourist Arrivals in Europe 2022," *Statista*, 02-Feb-2023. [Online]. Available: <https://www.statista.com/statistics/273598/international-tourist-arrivals-in-europe/>. [Accessed: 14-Feb-2023].
3. "World tourism rankings," *Wikipedia*, 04-Feb-2023. [Online]. Available: https://en.wikipedia.org/wiki/World_Tourism_rankings. [Accessed: 14-Feb-2023].
4. "Most Visited Countries 2023," *Most visited countries 2023*. [Online]. Available: <https://worldpopulationreview.com/country-rankings/most-visited-countries>. [Accessed: 14-Feb-2023].
5. "47+ Europe Travel & Tourism Statistics (2023)," *dream big travel far blog*. [Online]. Available: <https://www.dreambigtravelfarblog.com/blog/europe-tourism-travel-statistics>. [Accessed: 14-Feb-2023].
6. W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014.



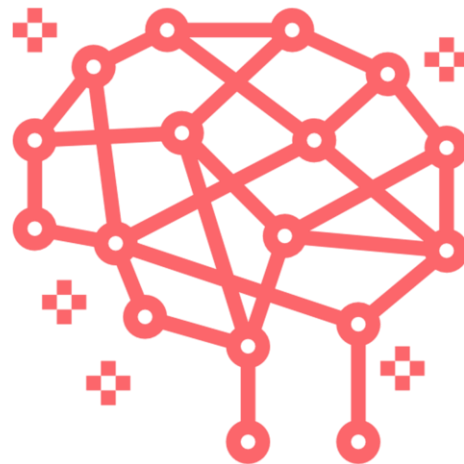
Obrigado pela Atenção!

Gestão de Projetos Informáticos

João Ramos (20200255)

Pedro Cunha (20200908)

07/03/2023



TripSense
Software Development