

**FIAP - FACULDADE DE INFORMÁTICA E ADMINISTRAÇÃO
PAULISTA
DATA SCIENCE**

Diego Henrique Vieira – RM 557428

Felipe Arias – RM 99590

Giovanna Victoria Fernandes Zacarias – RM 554397

Gustavo Semenuk – RM 550472

Hugo de Oliveira Abdala Feital do Carmo – RM 557776

Ideação e Escopo do projeto

Challenge – Totvs

São Paulo
2025

IDEAÇÃO e ESCOPO

[IDEAÇÃO] do projeto para [TOTVS], sobre o desafio e problema proposto pela empresa.

Curso: Data Science
Turma: 2TSCOA

São Paulo
2025

Resumo

Neste projeto iremos apresentar o problema que pretendemos resolver, ideia da nossa solução em termos de negócios e tecnologia, escopo do projeto, aspectos técnicos e análises de mercado. Tudo sobre o projeto desenvolvido pelo nosso grupo que promete ser trazer uma solução que irá revolucionar a jornada do cliente nas empresas.

Palavras-chave: Ideação, Negócios, Jornada do Cliente.

Introdução:	5
Problema:	6
Ideia de Solução:	7
Minimum Viable Product (MVP):	9
Escopo do Projeto:	12
Engenharia de Requisitos:	14
Arquitetura de Solução:	15
Critérios de Sucesso:	17
Referências:	19

Introdução:

A Hermes AI surge com uma solução inovadora para solucionar os problemas relacionados a jornada do cliente dentro das empresas, trazendo tecnologias avançadas e uma inteligência de dados para otimizar, customizar, recomendar e prever tendências na jornada do seu cliente desde o onboarding até o pós-venda.

É importante citar que existe uma tendência para os próximos anos de uma “hiperpersonalização baseada na análise da jornada do cliente”, como diz Ricardo Pena, CEO da PeopleXperience. Pena afirma que nos próximos anos o mercado deve evoluir para níveis ainda mais segmentados e proativos, com o uso intensivo de IA e machine learning para antecipar comportamentos e oferecer recomendações em tempo real (Referência 1.1).

Com a crescente demanda e exigência dos clientes por uma experiência de excelência com as empresas que consomem, é fundamental a utilização das mais novas tecnologias para poder acompanhar o mercado e seguir inovando e fidelizando os clientes.

A nossa solução não só vai ajudar a resolver esse problema, mas também queremos mudar a forma como sua empresa interage com os seus clientes nos diversos níveis da sua jornada.

Neste documento iremos apresentar a seguir como a Hermes AI irá te ajudar e resolver os seus problemas relacionados a jornada do cliente e seguir a frente da concorrência nas novas tendências do mercado. Nos próximos passos, vamos explicar detalhes sobre como vamos fazer isso e como a nossa solução irá funcionar.

Problema:

O problema central que a nossa solução visa resolver é a falta e a dificuldade de personalização na jornada do cliente, que pode trazer diversos problemas para a empresa como: Queda na satisfação do cliente, aumento na taxa de abandono, baixa conversão de leads, perda de fidelização, falta de escalabilidade, entre tantos outros impactos.

A falta de personalização pode gerar problemas em diversos níveis da jornada do cliente, desde o primeiro contato aos leads e pré-venda até a fidelização dos clientes atuais. Fator que influencia no crescimento da empresa e deixando um passo atrás de concorrência.

De acordo com o relatório Next in Personalization 2021, 71% dos consumidores esperam ter experiências personalizadas e 76% se sentem frustrados quando isso não acontece. Já de acordo com o CX Trends Latam 2022, 90% dos clientes estão dispostos a gastar mais em experiências personalizadas e, empresas que personalizam experiência geram 40% mais receita (Referência 2.1).

Devido a isso diversas empresas veem enfrentando problemas para se adequar as novas necessidades e demandas no mercado, a personalização da jornada do cliente pode trazer grandes dificuldades como: A falta de dados precisos: **50% das empresas** enfrentam dificuldades para coletar dados precisos necessários à personalização eficaz, falta de integração de dados: Dados dispersos em diferentes sistemas dificultam a criação de uma visão unificada do cliente, comprometendo a eficácia da personalização, desafios na implementação de tecnologias avançadas: Embora **92% das organizações** afirmem adotar a personalização baseada em IA para adaptar experiências individuais, muitas ainda enfrentam desafios na implementação eficaz dessas tecnologias (Referência 2.6), escalabilidade e custo para grandes quantidades de clientes: O custo inicial das ferramentas e a atualização constante pode implicar em um custo considerável de acordo que a empresa cresça em número de clientes, pois personalizar para 100.000 ou mais exige automação, inteligência artificial e integração entre canais, muitas empresas não conseguem manter consistência na experiência do cliente em todos os pontos de contato à medida que escalam.

Ideia de Solução:

A nossa solução visa entregar uma plataforma de inteligência de dados voltada a customização e personalização da jornada do cliente utilizando machine learning, inteligência artificial e deep learning entre outras tecnologias para entregar valor com base em dados.

Como solução iremos disponibilizar uma plataforma web baseada no modelo de negócio PAAS, onde iremos disponibilizar a nossa plataforma para a empresa disponibilizar os seus dados e usufruir dos nossos modelos e AI para melhorar a jornada do cliente.

No primeiro contato com a empresa a nossa plataforma irá apresentar uma aba de de início (Home) para a empresa obter maiores informações sobre o nosso projeto, métodos e conceitos utilizados no desenvolvimento dele, tal como os benefícios da nossa solução. A nossa solução usará os dados fornecidos previamente pela empresa, e esses dados irão passar pelo processo de ingestão de dados batch de forma assíncrona, para que possamos realizar todo o tratamento e análise prévia dos dados e disponibilizar a análise em um dashboard, que ficará disponível para a empresa na aba do menu lateral “Análise”, ainda no menu a empresa poderá acessar de forma simples todas as nossas funcionalidades disponíveis. De posse dos dados fornecidos pelo cliente e seu processo de jornada do cliente, com os seus respectivos passos e etapas, usaremos ela para enriquecer e direcionar os insights e resultados extraídos da nossa solução, a empresa também pode optar por seguir sem a especificação da sua jornada do cliente.

Em relação aos nossos modelos ao modelo de clusterização será estudado e avaliado qual a melhor técnica a ser aplicada, utilizando python para o seu desenvolvimento e entrega via API. A nossa estratégia é estudar três modelos diferentes de clusterização avançada para quantificar os resultados e performance obtidos para obter o melhor modelo para a situação em que a empresa deseje, por isso implementaremos a introdução de viés para aperfeiçoar os resultados dos modelos, além dos métodos mencionados anteriormente pretendemos utilizar a teoria dos grafos, aonde utilizaremos o conceitos de relacionamentos para identificar relações mais fortes e mais fracas para melhorar consideravelmente os resultados extraídos.

Para popular e alimentar todo o sistema e trazer inteligência por meio de novas variáveis além das que foram apresentadas, visando aperfeiçoar os nossos modelos com dados relevantes, iremos buscar dados públicos e abertos para utilizar e

melhorar os resultados. Podendo já citar algumas bases de dados que mapeamos, estão dados demográficos, dados relacionados a economia dos estados, reclamações e registros do Reclame Aqui da empresa entre tantos outros dados abertos e públicos que iremos abordar para criar a nossa plataforma de dados e inteligência.

Após a construção dos cluster de clientes, a nossa solução irá apresentar recomendações para a jornada do cliente inserida, possibilitando direcionar cada cluster para cada trilha jornada do cliente. As recomendações visam trazer um outro nível de insight, já baseada com os dados disponibilizados dos clientes mais as informações inseridas, como o próprio fluxo da jornada do cliente da empresa.

A apresentação dos clusters e toda a análise será exibida por meio de painéis em uma página exclusiva para a visualização dos dados e exportação via pdf. Pensando no acompanhamento e transparência da efetividade dos modelos iremos disponibilizar também visões de desempenho dos modelos.

Minimum Viable Product (MVP):

1. Objetivo do MVP:

Para o nosso MVP (Minimum Viable Product) iremos desenvolver uma aplicação mais simples e de concepção rápida para testar, validar conceitos e hipóteses antes de acelerar o desenvolvimento da solução e implementação de novas funcionalidades ou melhoria das ideias validadas pelo MVP.

2. Público-alvo:

O público-alvo da nossa solução são empresas de médio e grande porte visto que a grande dificuldade de escalabilidade e acesso e/ou conhecimento em tecnologias avançadas acaba sendo uma barreira para a empresa implementar tais soluções. O MVP foi projetado para abranger uma amostra representativa e satisfazer as necessidades de nosso público-alvo desde o início do projeto. Optamos por concentrar nossos esforços inicialmente na base de informações disponibilizada pela Totvs, a qual representa perfeitamente o público que queremos atingir e nos traz uma grande variedade de caminhos em nosso sistema devido a grande diversidade de soluções e clientes que a Totvs possui.

3. Funcionalidades Principais:

- **Estrutura da solução e aplicação web:**

Iremos desenvolver a nossa aplicação em python utilizando a biblioteca de código aberto streamlit, no qual iremos estruturar a solução com os conceitos de MVC (Model-View-Controller), que é um padrão de arquitetura de software que separa um aplicativo em três componentes interconectados: o Modelo, a Visão e o Controlador, para melhorar a organização, a testabilidade e a escalabilidade, criando uma "separação de interesses", complementando esse conceito, iremos nos basear na arquitetura de micro serviços para separar as aplicações da solução de forma independente, que irão se comunicar via APIs. No qual a solução foi projetada para desde o MVP já possuir uma estrutura escalável e com as melhores práticas de desenvolvimento e arquitetura de software.

- **Análise de Dados (Descritiva e Exploratória):**

Com a base de dados disponibilizada pela empresa iremos tratá-la e realizar análises descritivas e exploratórias dos dados apresentados, com o intuito de extrair alguns insights e um resumo dos dados e suas características para auxiliar nos processos

seguintes da nossa solução e para poder apresentar ao usuário os insights extraídos dessa análise por meio de um dashboard.

- **Dashboard de visualização de dados:**

Conforme citado na análise de descritiva, iremos disponibilizar um dashboard para o usuário pode visualizar os resultados obtidos pelas análises realizadas. Este dashboard será desenvolvido utilizando uma ferramenta de mercado, chamada Power BI, no qual a nossa solução irá se integrar com o Power BI para apresentar o painel direto pela nossa solução, sem que o usuário tenha que ser direcionado para outra página web.

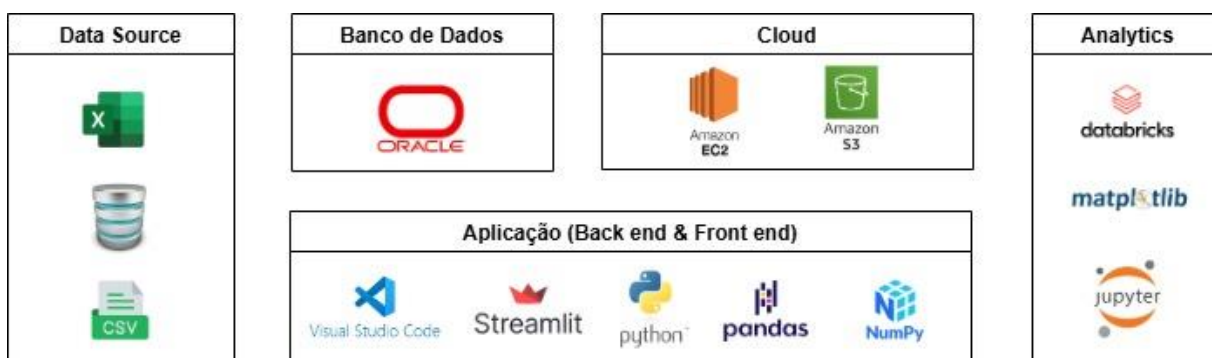
- **Modelo de Clusterização:**

Para realizar a clusterização nesta fase de projeto (MVP) iremos utilizar o algoritmo K-means devido a sua facilidade de aplicação e desenvolvimento para teste de variáveis e hipóteses. Porém com a evolução do projeto e seu desenvolvimento, iremos implementar modelos mais avançados de clusterização, no qual o usuário poderá escolher qual o modelo que ele quer obter o resultado, podendo assim comparar o desempenho entre os modelos.

4. Critérios de Sucesso do MVP:

- A aplicação está funcional e acessível via web.
- O pipeline de análise de dados roda com sucesso em pelo menos 80% dos testes com bases reais.
- O algoritmo de K-means roda sem erro e entrega resultados coerentes.
- O dashboard exibe corretamente os clusters gerados, com visualização clara dos dados.

5. Tecnologias e Recursos Envolvidos no Protótipo e MVP:



6. Limitações do MVP:

Funcionais:

- Não há autenticação ou controle de acesso.
- Apenas o algoritmo K-means estará disponível (sem testes com outros modelos durante o MVP).
- Não está planejado a ingestão de dados real time, somente batch.

Modelos e Inteligência:

- Não haverá personalização automática baseada em jornada do cliente (sem integração do modelo com trilhas específicas).
- Não será implementada a teoria dos grafos no MVP.

Interface e Experiência

- A interface será simples, baseada em Streamlit, com foco em funcionalidade e não em design final.
- O dashboard terá visualizações básicas, sem recursos interativos avançados ou filtros dinâmicos.

Performance e Escalabilidade

- A aplicação será implementada em cloud disponibilizada pela comunidade streamlit, então haverá possíveis limitações relacionada a acessos simultâneos, disponibilidade e capacidade de processamento.

Escopo do Projeto:

Para o desenvolvimento do projeto, optamos por utilizar o conceito de MVP (Produto Mínimo Viável) para entregar um protótipo simples em relação aos requisitos e objetivos descritos em nossa ideação. Isso nos permitirá validar nossa ideia e o conceito da solução, bem como lançar e desenvolver rapidamente, proporcionando tempo adequado para testes e aprimoramentos com base no feedback recebido.

Mesmo sendo uma versão inicial do nosso projeto, seu propósito é abranger uma amostragem considerável e atender às necessidades do nosso público-alvo desde os estágios iniciais. No desenvolvimento da solução iremos implementar técnicas de DevOps, desenvolvimento ágil e engenharia de software e produto para desenvolver uma aplicação escalável e uma infraestrutura para suportar as funcionalidades projetadas.

Com base nas informações fornecidas acima, descreveremos as principais etapas do desenvolvimento da inteligência da nossa solução a seguir:

Plataforma de Dados:

A plataforma de dados irá armazenar toda a nossa base de informações coletados e transformá-las em insights, gerando e entregando valor nos dados da base. Iremos o Databricks para ser a nossa plataforma de dados no qual será feita a ingestão de dados, transformação e disponibilização dos dados para todos os sistemas projetados.

1.1. Coleta de dados:

O processo de coleta de dados tem como insumo diversas fontes de dados e maneiras distintas de coleta para que possamos criar um data lake robusto como entrada da nossa solução, onde teremos processos de extração como web scrapping e coleta de arquivos via csv oriundos de data sets públicos que serão de extrema importância em nossa análise, pois será possível utilizar dados históricos e fatores externos, trazendo novas variáveis para o modelo. As principais fontes de dados serão de canais como: IBGE, Base dos dados, GOV etc. O formato dos dados coletados será em grande maioria arquivos csv, xlsx, json, que serão tratados utilizando tecnologias de Big Data e python. Outras fontes de dados também serão utilizadas como dados extraídos via web (Reclame Aqui e sites), além dos dados fornecidos pelo usuário.

4.2. Armazenamento, Tratamento, Processamento dos Dados: Conforme citado anteriormente iremos utilizar a plataforma do Databricks para e conjunto de algoritmos em python para a realização do ETL (Extract, Transform, Load). O tratamento dos dados coletados será feito no Databricks SQL para que ao término do processo os dados sejam disponibilizados em tabelas e sejam acessados via conexão com o cluster do Databricks. Tudo isso fazendo com tenhamos qualidade e escalabilidade dos dados para a aplicação de forma clara e rápida, na onde a aplicação e e seus serviços poderão consultar e trazer os dados por meio de uma conexão JDBC com o catalogo, bancos de dados e tabelas geradas.

Plataforma de Data Science e Machine Learning:

As análises que serão realizadas visam identificar padrões, resumir os dados, realizar agrupamentos e cálculos estatísticos para entregar os dados aos passos seguintes do processo. A nossa plataforma de Data Science e Machine Learning irá consumir os dados do Lake e irá disponibilizar os seus modelos e algoritmos via api, utilizando do conceito de arquitetura software baseada em microserviços.

4.3.1. Análise Descritiva e Exploratória: Para chegar ao nosso objetivo final, é necessário realizarmos uma exploração dos dados coletados, para podermos ordená-los e entendermos o que de fato conseguimos extrair de insight da base coletada, visto que teremos uma base de dados fixa (Proveniente de dados fornecidos pela empresa e dados coletados de fontes abertas). Por isso uma boa análise descritiva e exploratória faz jus a uma boa utilização dos dados e a extração máxima de insights possíveis.

4.3.2. Modelo de Clusterização: Uma análise descritiva serve para entendermos melhor da situação atual do usuário com base nas informações coletadas que serão utilizadas para identificar possíveis Outliers no monitoramento diário do usuário e também gerando um histórico de saúde do paciente para ser utilizado na análise preditiva.

Sistema de Automação e IA:

4.3.4. Inteligência Artificial: A utilização da IA terá como finalidade facilitar a compreensão dos resultados gerados pela solução para o usuário e também ser uma interface para o usuário gerar maiores insights, esta feature será uma interface no formato de chat para o usuário realizar suas consultas via texto sobre os resultados gerados, análises realizadas e os dados coletado pela nossa solução, facilitando a assim a UX dele dentro da plataforma e terá como finalidade de ser um componente

importante no contato do usuário e a transmissão das informações de forma clara e objetiva.

Painéis de Visualização:

4 .4. Visualização dos dados: Outra forma de visualização das informações para o usuário será por meio de um dashboard com infográficos e gráficos, trazendo de forma simples e clara dados estatísticos para a compreensão do usuário dos dados fornecidos e dos insights iniciais.

Engenharia de Requisitos:

Os requisitos descritos abaixo evidenciam somente os principais requisitos do sistema:

5.1. Requisitos funcionais:

- **RF 1 - Coleta de dados:** Coletar de dados de bases públicas e do usuário por meio de web scrapping.
- **RF 2 - Análise Descritiva e Exploratória:** Algoritmos para análise descritiva e exploratória dos dados fornecidos e coletados.
- **RF 3 - Data Science e Machine Learning:** Modelos de machine learning para realizar clusterização e classificações com base nos dados obtidos.
- **RF 4 - Inteligência Artificial:** A solução deve possibilitar ao usuário uma interface e tecnologia com inteligência artificial, para que ele possa realizar perguntas e extrair insights dos resultados gerados pelo modelo.
- **RF 5 - Visualização:** A visualização das informações e resultados obtidos disponibilizado para o usuário em forma de gráficos.

5.2. Requisitos não funcionais:

- **RNF 1 - Escalabilidade:** O sistema deve ser escalável para lidar com um aumento no número de usuários e grande volume de dados analisados.
- **RNF 2 - Segurança da Informação:** Os dados de saúde dos usuários devem ser criptografados e protegidos para garantir a privacidade de acordo com a Lei LGPD.
- **RNF 3 - Usabilidade:** A interface do usuário deve ser intuitiva e fácil de usar, considerando usuários de diferentes faixas etárias e grau de instrução.

- **RNF 4 - Compatibilidade e Acessibilidade:** O aplicativo deve ser compatível com uma ampla variedade de dispositivos móveis Android para atingir uma grande parte da população.

Arquitetura de Solução:

1 - Data source: Dados de formatos diversos, em suas maioria .csv e excel, e também textos extraídos via web scrapping;

2 - Data Lake: Criação do data lake utilizando o databricks (free edition) para ingerir todos os dados coletados e fornecidos;

3 - Analytics: Com os dados já dentro do Databricks, utilizei o Databricks SQL para tratar as tabelas, realizar análise prévia e criar uma validation para facilitar a utilização dos dados no Power BI e AI;

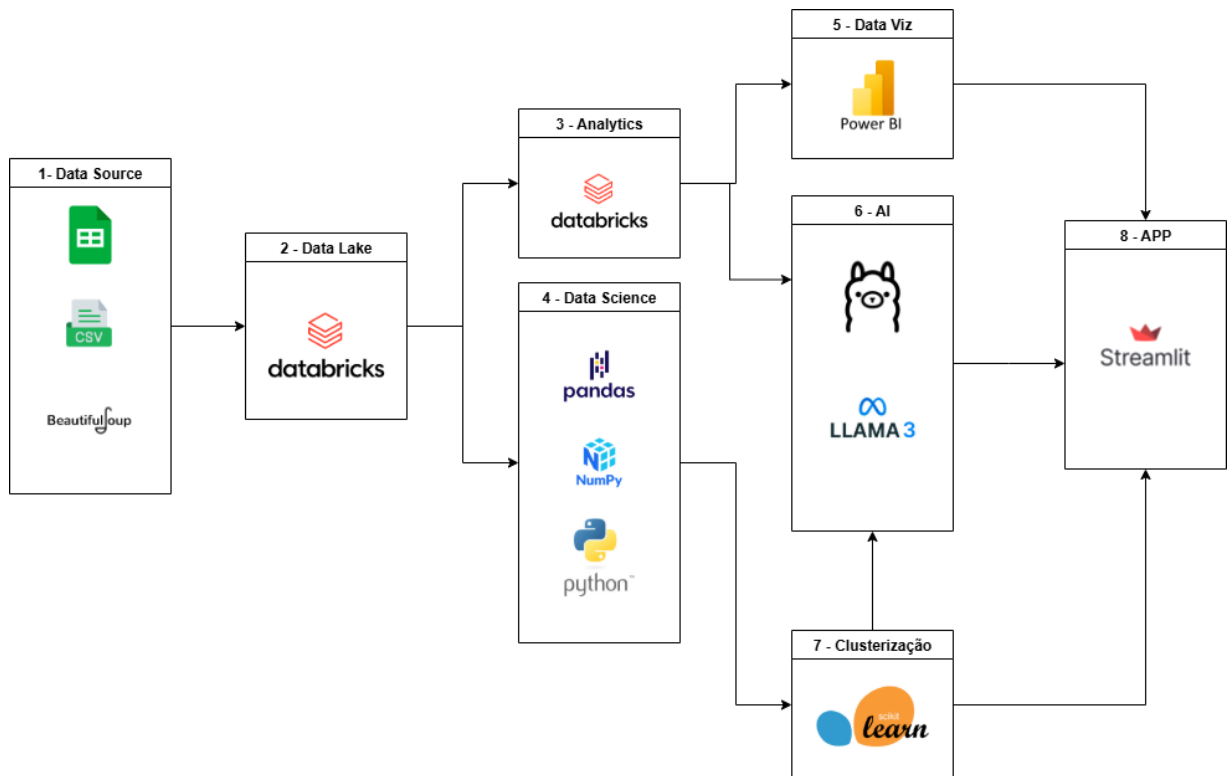
4 - Data Science: Utilizando python, conecto ao cluster do Databricks para acesso as tabelas e dados, e realizo a preparação dos dados para entrada no modelo de clusterização;

5 - Data Viz: Conexão do Power BI com o cluster do Databricks para acesso as tabelas e validation para a criação do dashboard;

6 - AI: Utilizando python para criar um agente de AI que irá utilizar como base de conhecimento os dados e tabelas do Databricks e os out put dos modelos de clusterização, para criar um chat integrado ao streamlit para interação do usuário;

7 - Clusterização: Criação dos modelos de clusterização utilizando bibliotecas python e o próprio scikit learn;

8 - Aplicação: Python e a biblioteca Streamlit como base da aplicação front e back end, com uso de HTML para personalização e customização do design.



Utilização dos Dados:

A utilização de dados públicos e de fontes abertas é fundamental para a nossa solução agregar valor e enriquecer os dados disponibilizados, citaremos abaixo algumas bases utilizadas e como elas foram tratadas.

Principais canais de busca dos dados:

- IBGE
- Reclame Aqui
- GOV
- Receita federal
- Base dos dados

Limitações do Projeto

Apesar de propor uma solução inovadora e escalável, o projeto **Hermes AI** não tem como objetivo contemplar todas as funcionalidades ou áreas relacionadas à jornada do cliente. Dessa forma, algumas limitações importantes devem ser consideradas:

- 1. Foco em análise e recomendações, não em execução operacional**
 - a. O escopo se restringe à **análise, clusterização e geração de insights/recomendações**.
- 2. Escopo de dados delimitado**
 - a. A plataforma se concentrará em **dados públicos, dados fornecidos pela empresa e bases previamente mapeadas**.
- 3. Interface com foco funcional, não em design final**
 - a. A aplicação prioriza usabilidade e clareza, mas **não busca ser uma plataforma de UI/UX avançada** ou competir em termos de design sofisticado.
 - b. O foco é entregar **insights acionáveis**.
- 4. Não é uma solução de Big Data em tempo real (streaming)**
 - a. Embora seja escalável, a solução não tem como foco inicial processar fluxos massivos de dados em tempo real.
 - b. O processamento é orientado a **lotes (batch)** para análises estratégicas e não para operações em tempo real.
- 5. Não cobre toda a jornada do cliente ponta a ponta**
 - a. A solução se concentra nas **etapas de análise, clusterização e recomendações personalizadas**.

Critérios de Sucesso:

Os critérios de sucesso abaixo foram definidos utilizando a metodologia de gestão baseada no método “Objetivos e resultados chave” (OKR) para setar os principais objetivos da nossa solução:

- **OKR1 - Desenvolver um app web responsivo:** Aplicativo web capaz de suportar e apresentar a solução proposta.
- **OKR2 - Conexão e integrações com diversos sistemas:** App capaz de se conectar com outras ferramentas sem causar indisponibilidade do app.
- **OKR3 - Solução com alta disponibilidade:** Desenvolver uma solução completa que consiga entregar alta disponibilidade.
- **OKR4 - Modelos com boa assertividade:** Modelos capaz de realizar as análises propostas com uma boa assertividade nos resultados.
- **OKR5 - Implementar Recursos de Inteligência Artificial no App:** Integrar inteligência artificial no aplicativo capaz de auxiliar e enriquecer resultados obtidos.

Referências:

As referências a seguir estão classificadas de acordo com o seu respectivo assunto e tópico do projeto a qual a referência foi utilizada.

1.0 Introdução:

1.1 - Disponível em: https://www.em.com.br/mundo-corporativo/2025/03/7089229-mercado-de-gestao-da-jornada-do-cliente-projeta-crescimento.html?utm_source=chatgpt.com. Acesso em: 14/04/25

1.2 - Disponível em: https://www.olos.com.br/como-a-digitalizacao-da-jornada-do-cliente-impacta-as-empresas/?utm_source=chatgpt.com. Acesso em: 14/04/25

1.3 - Disponível em: https://nuvy.com.br/jornada-do-cliente-a-importancia/?utm_source=chatgpt.com. Acesso em: 14/04/25

1.4 - Disponível em: https://blog.peopleexperience.com/pesquisa-mostra-que-87-valorizam-boa-experiencia-do-cliente/?utm_source=chatgpt.com. Acesso em: 14/04/25

2.0 Problema:

2.1 - Disponível em: https://harmo.me/blog/etapas-personalizadas-da-jornada-do-cliente?utm_source=chatgpt.com. Acesso em: 14/04/25

2.2 - Disponível em: <https://www.oracle.com/br/cx/customer-experience-challenges/>. Acesso em: 14/04/25

2.3 - Disponível em: https://www.em.com.br/mundo-corporativo/2025/03/7089229-mercado-de-gestao-da-jornada-do-cliente-projeta-crescimento.html?utm_source=chatgpt.com. Acesso em: 14/04/25

2.4 - Disponível em: <https://www.inovyo.com/2024/03/06/desafios-da-personalizacao-da-experiencia-do-cliente/>. Acesso em: 14/04/25

2.5 - Disponível em: <https://www.inovyo.com/2024/03/06/desafios-da-personalizacao-da-experiencia-do-cliente/>. Acesso em: 14/04/25

2.6 - Disponível em: <https://www.medallia.com/br/recurso/que-opinan-los-consumidores-de-la-personalizacion/>. Acesso em: 14/04/25

Arquitetura de Solução:

Disponível em: <https://blog.bytebytego.com/p/ep76-netflixs-tech-stack> . Acesso em: 20/04/25

Disponível em: https://www.linkedin.com/posts/brunosmm_idp-stack-tech-activity-7242893513426644992-apFO?utm_source=share&utm_medium=member_desktop&rcm=ACoAACSCzOcBXhpf0MxmkVakNxZF_bCyTSdstNk

. Acesso em: 20/04/25

Disponível em: <https://www.youtube.com/watch?v=zOAtcyDzYBs> . Acesso em: 21/04/25

Aspectos Técnicos:

Disponível em: <https://medium.com/@neboutben97/from-local-environment-to-microservices-deploying-a-streamlit-application-with-fastapi-and-9c9be653ff57>. Acesso em: 03/05/25

Disponível em: <https://charumakhijani.medium.com/machine-learning-model-deployment-as-a-web-app-using-streamlit-4e542d0adf15>. Acesso em: 03/05/25

Disponível em: https://www.alura.com.br/artigos/streamlit-compartilhando-sua-aplicacao-de-dados-sem-dor-de-cabeca?srsItd=AfmBOopDDCg9SyUQXTPS8WXB8386vfASYOuGKSxyFFqgb_vJVrie7RTz. Acesso em: 03/05/25

Disponível em: <https://blog.dp6.com.br/construindo-dashboards-e-data-apps-com-streamlit-afac87962ba0?gi=692359170c73>. Acesso em: 04/05/25

Disponível em: <https://medium.com/data-hackers/desenvolvimento-de-um-aplicativo-web-utilizando-python-e-streamlit-b929888456a5>. Acesso em: 10/05/25

Disponível em: <https://hub.asimov.academy/blog/streamlit-guia-completo/>. Acesso em: 10/05/25

Disponível em: <https://docs.streamlit.io/get-started/fundamentals/main-concepts>. Acesso em: 10/05/25

Disponível em: <https://docs.databricks.com/aws/pt/dev-tools/databricks-apps/tutorial-streamlit>. Acesso em: 14/05/25

Disponível em: <https://discuss.streamlit.io/t/embedding-bi-tool-in-streamlit/19030>. Acesso em: 20/06/25

Disponível em: <https://www.inf.ufpr.br/andre/files/Ribas2010.pdf>. Acesso em: 14/04/25

Disponível em: https://www.maxwell.vrac.puc-rio.br/24787/24787_5.PDF. Acesso em: 29/05/25

Disponível em: <https://medium.com/@anwarhermuche/tutorial-de-clusteriza%C3%A7%C3%A3o-de-dados-categ%C3%B3ricos-e-mistos-com-k-prototypes-98a2115a6b9f>. Acesso em: 11/05/25

Disponível em: <https://www.ic.unicamp.br/~wainer/cursos/2s2016/ml/aula-cluster-grafos.pdf>. Acesso em: 11/05/25

Disponível em: <https://www.teses.usp.br/teses/disponiveis/55/55134/tde-01042008-142253/publico/dissertacao.pdf>. Acesso em: 06/06/25

Disponível em: <https://pm3.com.br/blog/analise-de-cluster/>. Acesso em: 14/04/25

Disponível em: <https://scikit-learn.org/stable/modules/clustering.html>. Acesso em: 13/06/25

Disponível em: <https://ic.unicamp.br/~reltech/2005/05-03.pdf>. Acesso em: 17/06/25

Disponível em: <https://developers.google.com/machine-learning/clustering?hl=pt-br>. Acesso em: 15/07/25

Disponível em: <https://www.geeksforgeeks.org/machine-learning/clustering-in-machine-learning/>. Acesso em: 19/07/25

Disponível em: <https://medium.com/data-science/a-guide-to-clustering-algorithms-e28af85da0b7>. Acesso em: 19/07/25

Disponível em: <https://medium.com/cwi-software/entendendo-clusters-e-k-means-56b79352b452>. Acesso em: 19/07/25

Disponível em: <https://www.linkedin.com/pulse/clustering-e-arte-de-agrupar-dados-como-dominar-o-jonas-leonardo-i1itf/>. Acesso em: 19/07/25

Disponível em: <https://spark.apache.org/docs/latest/ml-clustering.html>. Acesso em: 20/07/25

Disponível em: <https://www.ime.usp.br/~yw/publications/books/TeoriaDosGrafos.pdf>. Acesso em: 25/07/25

Disponível em: <https://medium.com/@anwarhermuche/teoria-dos-grafos-introdu%C3%A7%C3%A3o-defini%C3%A7%C3%B5es-matriz-e-lista-de-adjac%C3%Aancia-2252d4800a44>. Acesso em: 25/07/25

Disponível em: <https://repositorio.ufpb.br/jspui/bitstream/tede/7549/5/arquivototal.pdf>. Acesso em: 08/08/25

Disponível em:
https://www.ic.unicamp.br/~ra063658/disciplinas/mat017_2022s1/mod15-handout.pdf. Acesso em: 10/08/25

Disponível em: https://www.ime.usp.br/~mota/bookFiles/livro_grafos.pdf. Acesso em: 10/08/25

Disponível em: <https://medium.com/xp-inc/grafos-teoria-e-aplica%C3%A7%C3%B5es-2a87444df855>. Acesso em: 14/04/25

Disponível em:
https://www.obm.org.br/content/uploads/2017/01/Nivel1_grafos_bruno.pdf. Acesso em: 14/08/25

Disponível em:
<https://www.inf.ufrgs.br/~prestes/Courses/Graph%20Theory/Capitulo%204.pdf>. Acesso em: 20/08/25

Disponível em: <https://www.inf.ufrgs.br/~prestes/Slides/GrafosA7.pdf>. Acesso em: 20/08/25

Utilização dos dados:

Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/outras-estatisticas-economicas/9123-as-micro-e-pequenas-empresas-comerciais-e-de-servicos-no-brasil.html?=&t=downloads>. Acesso em: 07/07/25

Disponível em: <https://www.ibge.gov.br/estatisticas/metodos-e-classificacoes/classificacoes-e-listas-estatisticas/9078-classificacao-nacional-de-atividades-economicas.html?=&t=downloads>. Acesso em: 07/07/25

Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9065-contagem-da-populacao.html?=&t=downloads>. Acesso em: 07/07/25

Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/industria/22649-demografia-das-empresas-e-estatisticas-de-empendedorismo.html?=&t=downloads>. Acesso em: 07/07/25

Disponível em: <https://dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica---cnpj>. Acesso em: 07/07/25

Disponível em:
https://arquivos.receitafederal.gov.br/dados/cnpj/dados_abertos_cnpj/2025-08/.
Acesso em: 07/07/25

Disponível em: <https://basedosdados.org/dataset/fcf025ca-8b19-4131-8e2d-5ddb12492347?table=93007431-7ce9-42ee-8740-8c2274d345ad>. Acesso em: 07/07/25

Disponível em: <https://basedosdados.org/dataset/d30222ad-7a5c-4778-a1ec-f0785371d1ca?table=b99f0017-e587-477e-8cfb-05fb5d1005b8>. Acesso em: 07/07/25

Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/comercio/9075-pesquisa-anual-de-comercio.html?=&t=downloads>. Acesso em: 07/07/25

Disponível em: . Acesso em: 07/07/25

Disponível em: . Acesso em: 07/07/25

Disponível em: . Acesso em: 07/07/25

Disponível em: . Acesso em: 07/07/25

Disponível em: . Acesso em: 07/07/25

Disponível em: . Acesso em: 07/07/25

Disponível em: . Acesso em: 07/07/25

