igti

RELATÓRIO

PROJETO APLICADO

Instituto de Gestão e Tecnologia da Informação Relatório do Projeto Aplicado

Disponibilização de dados analíticos e recomendações para clientes no Varejo E-commerce

Gabriel Lima Novais

Orientador(a):

Murillo Barbosa

14/01/23





GABRIEL LIMA NOVAIS

INSTITUTO DE GESTÃO E TECNOLOGIA DA INFORMAÇÃO RELATÓRIO DO PROJETO APLICADO

Disponibilização de dados analíticos e recomendações para clientes no Varejo E-commerce

Relatório de Projeto Aplicado desenvolvido para fins de conclusão do curso MBA em Engenharia de Dados.

Orientador (a): Murillo Barbosa

Niterói 14/01/23



Sumário

1. CANVAS do Projeto Aplicado	4
1.1 Desafio	5
1.1.1 Análise de Contexto	5
1.1.2 Personas	9
1.1.3 Benefícios e Justificativas	12
1.1.4 Hipóteses	14
1.2 Solução	17
1.2.1 Objetivo SMART	17
1.2.2 Premissas e Restrições	18
1.2.3 Backlog de Produto	21
2. Área de Experimentação	24
2.1 Sprint 1	26
2.1.1 Solução	26
 Evidência do planejamento: 	26
 Evidência da execução de cada requisito: 	26
 Evidência dos resultados: 	26
2.1.2 Experiências vivenciadas	26
2.2 Sprint 2	27
2.2.1 Solução	27
 Evidência do planejamento: 	27
 Evidência da execução de cada requisito: 	27
 Evidência dos resultados: 	27
2.2.2 Experiências vivenciadas	27
2.3 Sprint 3	28
2.3.1 Solução	28
 Evidência do planejamento: 	28
 Evidência da execução de cada requisito: 	28
Evidência dos resultados:	28
2.3.2 Experiências vivenciadas	28
3. Considerações Finais	29
3.1 Resultados	29
3.2 Contribuições	29
3.3 Próximos passos	29



1. CANVAS do Projeto Aplicado

Segue a Figura 01 conceitual canvas do projeto aplicado, que representa todas as etapas do projeto.



Figura 01 - Canvas do Projeto Aplicado



1.1 Desafio

1.1.1 Análise de Contexto

O varejo, principalmente em países em desenvolvimento como o Brasil, é de suma importância, por ser uma fonte provedora de empregos. No Brasil corresponde cerca de 47,4% do PIB, apresentando um crescimento expressivo e sempre está na mira de empreendedores em busca de oportunidades.

O setor consiste no processo de compra de produtos em quantidades relativamente grande dos produtores atacadistas e outros fornecedores e posteriormente vendidos em quantidades menores ao consumidor final. Além disso, pode ser considerado como atividade responsável por providenciar mercadorias e serviços desejados pelos consumidores.

O varejo é definido, segundo Philip Kotler como método comercial que engloba todas as atividades relativas à venda direta de produtos ou serviços ao consumidor final. No entanto, Parente (2000), conceitua o varejo como todas as atividades que englobam o processo de venda de produtos e serviços para atender a uma necessidade pessoal do consumidor final. Las Casas (2004, p. 17) cita uma definição de varejo de acordo a American Marketing Association, onde considera o varejo como "uma unidade de negócio que compra mercadorias de fabricantes, atacadistas e outros distribuidores e vende diretamente a consumidores finais e eventualmente aos outros consumidores".

Então qualquer empresa que forneça um produto ou serviço para o consumidor final está praticando varejo. O comércio varejista conta com processos sistematizados, a fim de atender bem às demandas dos clientes. É importante apresentar uma cadeia de suprimentos bem definida para corresponder às necessidades do mercado e demonstrar competitividade. O comércio varejista vem assumindo uma importância crescente no panorama empresarial do Brasil e do mundo.

À medida que as empresas varejistas se expandem, passam a adotar tecnologias avançadas de informação e de gestão e desempenham papel cada



vez mais importante na modernização do sistema de distribuição e da economia brasileira. Para que o setor seja bem-sucedido, é preciso estar alerta e pronto para se adaptar aos desafios de mercado.

A tecnologia da informação desempenha papel primordial nesse processo de evolução, visto que o comércio virtual (e-commerce) é a grande tendência do setor de varejo. O varejo online fornece ao consumidor informações, agilidade na entrega e apresenta descontos atraentes. Não apenas a tecnologia servirá como meio que possibilita a realização da venda, através das plataformas digitais, mas como uma forma essencial de permitir análises de planejamento estratégico, marketing, performance e muitas outras de acentuada importância para tomadas de decisão.

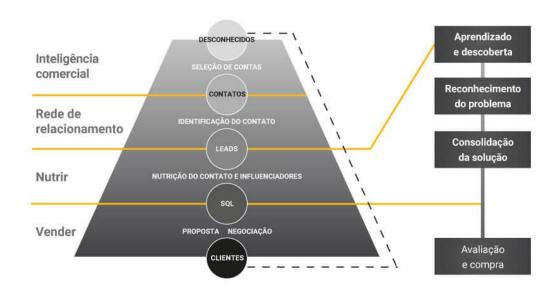


Figura 02 - Account Based Marketing

Fonte: https://layerup.com.br/account-based-marketing/

Segundo afirma a Associação Brasileira de Comércio Eletrônico, as vendas online correspondem a 11,6% do setor varejista do país. Contudo, toda esta evolução causa um impacto no comportamento do consumidor.

A tendência do consumidor moderno é ser mais exigente, mais direcionado por questões de tempo, mais necessitado de informação e



altamente individualista. Ações que visam a construção de um ambiente de vendas mais personalizado e com um conforto maior para o consumidor, garantem elevação de rentabilidade ao negócio.



Figura 03 - Modelo de Matriz CSD

Dessa forma, após elucidação do contexto e da exposição do problema, é possível defender cada vez mais o uso intensivo de dados. A necessidade de se construir maneiras mais eficientes de garantir uma experiência do usuário, ao mesmo tempo que se permite um ferramental adequado para acompanhar a performance das vendas, preços e visitas do site e dos desdobramentos de tais monitorias, se baseiam quase que completamente nessa hipótese.

Para estruturar as idéias mencionadas foram realizadas a matriz e a observação contida na Figura 03 acima. Adicionalmente, para realizar a análise do contexto do problema, foi utilizado o canvas abaixo, contido na Figura 04.





Figura 04 - Análise do Contexto do problema



1.1.2 Personas

No desenvolvimento de uma solução, um comum exercício que busca identificar o perfil daqueles que serão os usuários é a criação de personas, não necessariamente representando alguém do público alvo que saiba tudo do negócio em questão. Especificamente para este projeto, foram utilizadas as informações detalhadas nas seções anteriores para criar 3 perfis de personas. Esses perfis são de duas "pontas" completamente distintas do negócio em questão.

Enquanto a primeira e a segunda persona que são funcionários da empresa, do local onde a solução será empreendida, sendo os dois analistas de áreas diferentes, a terceira persona representa o consumidor final dos serviços apresentados. De forma resumida, listam-se essas personas:

- Analistas de BI e de Marketing dentro da empresa, na qual a solução será utilizada como um guia para tomada de decisão.
- Clientes, consumidores dos produtos vendidos pelo e-commerce, que desfrutarão de uma experiência personalizada de compra na plataforma.

Persona 1

Nome: Pedro (colocar sobrenome)

Idade: 30 anos

Profissão: Analista de BI

Características comportamentais: Pedro é engenheiro de produção e possui uma formação mais analítica. Possui um perfil bem pragmático, gosta de resolver problemas gerando formas de solução que funcionem no dia a dia, não precisando recorrer a complexidades desnecessárias. Ele precisa de dados para acompanhar de maneira mais eficiente as métricas de vendas e visitas da empresa e entender como esses indicadores refletem o valor gerado para companhia. Ele deseja rapidez e facilidade para gerar os reports e visualizações de dados para garantir tomadas de decisões mais concretas sobre preços e atingimento de metas.



Persona 2

Nome: Maria

Idade: 28 anos

Profissão: Analista de Marketing

Características comportamentais: Maria é formada em administração e possui pós-graduação em marketing digital. Ela possui um perfil mais criativo, com boas análises sobre tendências de mercado e conhecimentos sobre ações de marketing e segmentação de audiências, o que lhe confere resultados bons nas suas campanhas de marketing. Ela possui uma dificuldade de entender como melhorar os produtos recomendados para seus clientes e como segmentá-los para melhorar suas campanhas de push (envio de mensagens e acionamento de clientes).

Persona 3

Nome: Sérgio

Idade: 54 anos

Profissão: Carpinteiro

Características comportamentais: Sérgio possui curso técnico em carpintaria, e no seu negócio, prioriza pela qualidade do acabamento dos seus móveis. Entretanto precisa de ferramentas e outros insumos. Como possui muitas demandas, não pode perder tempo procurando esses recursos e para isso realiza compras online. Além disso, gosta de pescar e fazer trilhas, como forma de lazer, e compra todas as suas iscas e materiais pela internet. Se ele tiver o que precisa e gosta, em um só lugar, com certeza compraria e otimizaria seu tempo tão valioso.

As personas listadas foram construídas após feito o mapa da empatia mostrado na Figura 05. Note que nesse mapa as dores mostradas são compartilhadas pelas 3 personas e a solução a ser proposta visa sanar essa dor. De modo que, é criada a seguinte relação lógica: Uma vez disponibilizado dados para consulta sobre indicadores e métricas de avaliação do negócio e ao mesmo tempo recomendações de produtos, os usuários terão uma experiência personalizada de consumo, aumentando a rentabilização que poderá ser



acompanhada nos dados da outra tabela, permitindo que os analistas tomem decisões estratégicas para aproveitar não apenas essa nova maneira de exposição de conteúdo ou de seleção de portfólio de produtos a serem vendidos, como aperfeiçoar as campanhas de marketing e melhor interpretar as audiências e seus perfis.



Figura 05 - Mapa de Empatia



1.1.3 Benefícios e Justificativas

Conforme apresentado nas seções anteriores, o problema está bem delineado, ou seja, conforme o comércio varejista se expandiu em especial por conta de fatores externos, como a pandemia e evolução tecnológica, o volume de dados, produtos e serviços se expandiram também. Para que seja possível melhorar o acompanhamento e availação de métricas e operações com a finalidade de proporcionar tomadas de decisões assertivas, é necessário intensificar o uso de tecnologias relacionadas a big data, processamento e armazenamento de dados em cloud e afins.

O objetivo do projeto é o de proporcionar o tratamento de dados, armazenamento de dados processados e disponibilização de dados para duas finalidades distintas: uma focada nas análises de suporte de tomada de decisão, e outra com a finalidade de melhorar a personalização da experiência do usuário.

Os benefícios auferidos dessa empreitada podem ser observados nas facilidades de obtenção de dados, que estarão disponibilizados em formato de tabela SQL, contendo métricas de avaliação, que inclusive dada a natureza da infraestrutura aplicada, permite expansão de funcionalidades e adição de colunas. Além desses benefícios, existirão dados também disponibilizados em bases que permitem consultas em SQL, que fornecem a reelações de recomendações de produtos e usuários que facilitarão o uso de vitrines, ações de acionamento de clientes, segmentação de audiências e outras finalidades (inclusive mensuradas pelos dados da outra tabela)

O Blue print (Figura 06), divide a experiência em ação, funcionalidade, interação e mensagem. Deve-se entender a ação da persona ao buscar uma solução da sua dor. Para cada ação, a persona verá funcionalidades para as soluções possíveis, como interagir com as possíveis soluções e qual mensagem ela absorve durante aquela ação.





Figura 06 - Blueprint

A ferramenta da proposição de valor (Figura 07) detalha como que a persona enxerga o valor na solução a ser desenvolvida diante das dores que enfrenta. A reflexão de como a persona irá extrair valor ajuda a tornar a proposta mais objetiva. Como resultado, obteve-se que a solução mais adequada está direcionada em fornecer uma fonte de dados que auxilia a tomada de decisão.



Figura 07 - Explicação de Proposição de Valor



1.1.4 Hipóteses

As hipóteses seguem expostas na Figura 08, nela as observações feitas pelas personas em uma coluna e na coluna ao lado as hipóteses criadas para desenvolver o produto. Com essa informação foi possível estabelecer especificações mais direcionadas e detalhadas que eventualmente serão consideradas no processo macro de desenvolvimento do projeto.



Figura 08 - Matriz de observações para hipóteses

Verifica-se que as observações da matriz de observações para hipóteses direcionam o projeto na busca por formas de auxiliar tomada de decisões de negócios por meio de dados e as tecnologias associadas ao uso intesivo de dados.

Para elucidar as ideias que originaram a solução do projeto final usou-se um método de priorização de ideias. A matriz de priorização de idieas trata-se de uma ferramenta simples na qual costuma-se inserir as ideias validadas na vertical e os critérios de avaliação (de acordo com as principais necessidades identificadas) na horizontal.



Uma vez construída a matriz, a análise indica a ideia mais provável de sucesso e que poderá trazer maior resultado efetivo para o projeto. Dessa maneira ela serve para fornecer uma melhor visão das ideias e apoiar o processo de tomada de decisão. As pontuações atribuídas na matriz estão de acordo com a tabela na Figura 09 abaixo.

Escala	B - Benefícios	A - Abrangência	S - Satisfação	I - Investimentos	C - Cliente	O - Operacionalidade
5	De vital	Total	Muito grande	Pouquíssimo	Nenhum impacto	Muito fácil
	importância	(de 70 a 100%)		investimento		
4	Significativo Muito grande (de 40 a 70%)	Grande	Algum Impacto pequeno	Fácil		
		(de 40 a 70%)	Grande	investimento impacto pequeno	FaCII	
3	Razoável	Razoável	Média	Médio	Médio impacto	Média facilidade
		(de 20 a 40%)		investimento		
2	Poucos	Pequena	Pequena	Alto investimento	Immaeta granda	Difícil
	benefícios	(de 5 a 20%)		Pequena	Aito investimento	Impacto grande
1	Algum	Muito negueno	Quase não é	Altíssimo	Impacto muito	Muito diffail
	benefício	Muito pequena nota	notada	investimento	grande no cliente	Muito difícil

Figura 09 - Matriz da priorização de ideias.

Fonte: Apostila do MBA em Egenharia de Dados IGTI.

Usou-se a classificação conhecida como BASICO para se fazer a priorização de ideias. A sigla nos remete aos seguintes pontos, que devem ser pontuados:

- B (Benefícios): quais são os benefícios para a organização caso a solução seja adotada?
- A (Abrangência): quantas pessoas (clientes internos e externos) serão beneficiadas com essa decisão?
- S (Satisfação): qual é a satisfação das pessoas com a solução a ser adotada?
- I (Investimento): qual será o investimento necessário para a aplicação da solução?
- C (Cliente): Qual o impacto que o cliente sofrerá com a mudança?
- O (Operações): O quão complexa é a implantação da solução?



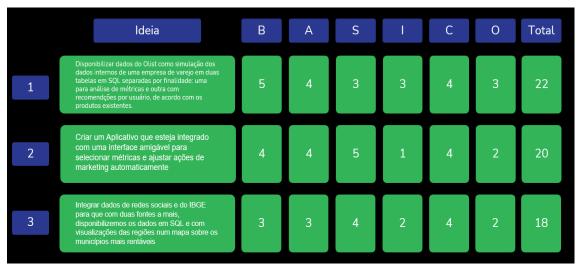


Figura 10 - Matriz da priorização de ideias.

Antes para desenvolver as ideias da matriz acima na Figura 10, foi realizado um Brainstorm levantando os aspectos que envolvem a solução final, algumas de suas características e as fontes de informação. A Figura 11 mostra um conjunto de ideias separadas por finalidade e contexto. Nessa etapa considera-se que soluções coletivas trazem resultados melhores e mais criativos do que ações individuais e isoladas.

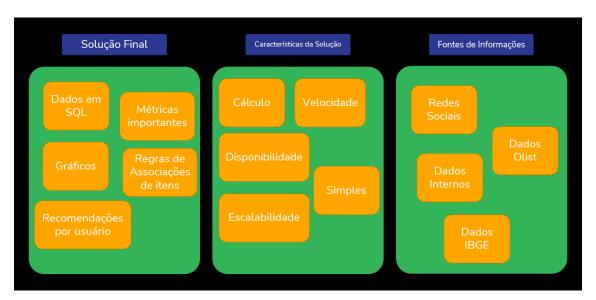


Figura 11 - Brainstorm



1.2 Solução

1.2.1 Objetivo SMART

"Objetivou-se fornecer dados tratados e com valor adicionado que possa estar acessível por um DataWarehouse disponível em um serviço de Cloud permitindo extrair análises de duas naturezas distintas.

A primeira natureza é um acompanhamento das métricas de desempenho do negócio. E a segunda é disponibilizar para clientes do Varejo E-commerce recomendações personalizadas. O usuário terá acesso a pelo menos uma plataforma para dar entrada a rotinas SQL para extrair informação dos dados. A maior parte da infraestrutura será construída em forma de código e estará em nuvem. O projeto deve ser concluído até Março de 2023."



1.2.2 Premissas e Restrições

As premissas e restrições do projeto são importantes, uma vez que conhecendo os impactos gerados caso algo ocorra fora do esperado é possível buscar ações efetivas para mitigar os riscos. Dessa forma criou-se uma Matriz de Riscos, conforme apresentado na Figura 11. Os riscos identificados são os seguintes:

- Preços, quantidades vendidas e visitas difíceis de agrupar: significa dados muito separados e dispersos. Os dados que devem idealmente ser utilizados são dados internos, mas para fins de generalização da solução optou-se por utilizar dados do Olist que representam com louvor os dados comumente encontrados nas empresas de varejo e-commerce. O impacto postencial seria a redução de velocidade de consulta.
- Volume muito grande de produtos: significa que caso haja uma variedade muito grande de produtos, diversidade de portfólio elevada, os modelos que calculam as regras de associação entre os dados de compra ou visitas tendem a apresentar resultados de qualidade não tão bons.
- Custo elevado de infraestrutura: refere-se ao custo total necessário para conseguir implementar a solução na cloud da AWS. Esse fator de custo se torna um risco em especial pela necessidade de processamento computacional pelos clusters kuberenetes que utilizam máquinas EC2.
 Uma boa medida preventida adotada seria a utilização de máquinas spot ao invés de máquina on demand.





Figura 11 - Matriz de Risco

A arquitetura construída para a solução, destaca-se na Figura 12 abaixo. Nela pode-se verificar que os dados serão distribuídos em 3 buckets no S3: um para ingestão, outro para processamento e outro para consumo.

Dessa forma a estrutura de processamento formada pelo cluster Kebernetes consegue organizar os inputs e outputs e garante a transferência desses dados de maneira rápida e fácil através do AWS Glue Crawler para o serviço de DataWarehouse da AWS, o Athena, local onde os analistas poderão realizar as consultas necessárias e eventuais diligências de ações para rentabilização.

A forma de construção da parte de buckets e serviços do Glue Crawler serão efetivadas por meio de estratégia IaC, conforme ferramental do Terraform. Todos os códigos, comandos, arquivos e instruções necessárias serão disponibilizadas no GitHub, conforme boas práticas de versionamento, no link a seguir: https://github.com/NovaisGabriel/MBA_project



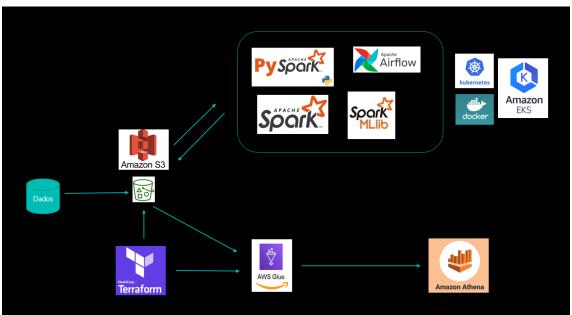


Figura 12 - Arquitetura na AWS



1.2.3 Backlog de Produto

O backlog compreende uma lista com todas as tarefas a serem desenvolvidas. As tarefas são mensuradas com prioridades, indivíduos que executarão a tarefa e instruções específicas dos requisitos necessários para que ela seja entendida como concluída. No backlog desse projeto, primeiramente, foram listados os requisitos, e conforme na Figuira 13, são descritos como cada requisito será coberto ao longo dos sprints.

Na sprint 3, com sua conclusão, espera-se que a solução companha um MVP, que esteja pronto e possa ser usada pelos usuários definidos. Alguns exemplos e testes serão executados para direcionar e simular a maneira como as personas do projeto extrairão o valor da solução.

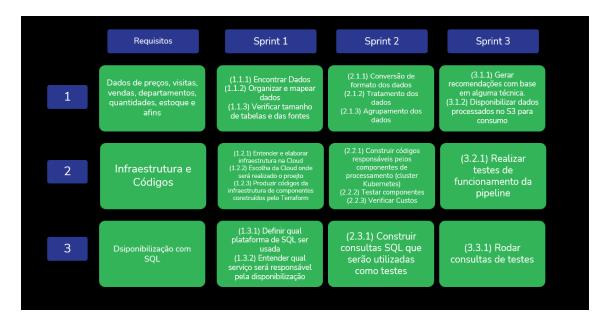


Figura 13 - Requisitos separados por sprints



A materialização do backlog descrito na Figura 13 foi registrado em ferramenta de planejamento Trello ¹. As atividades delineadas na figura acima foram organizadas nos quadros do Trello, descritos como na Figura 14. Conforme forem executadas as tarefas em andamento das Sprints do projeto, as atividades destacadas do backlog serão dadas como concluídas.

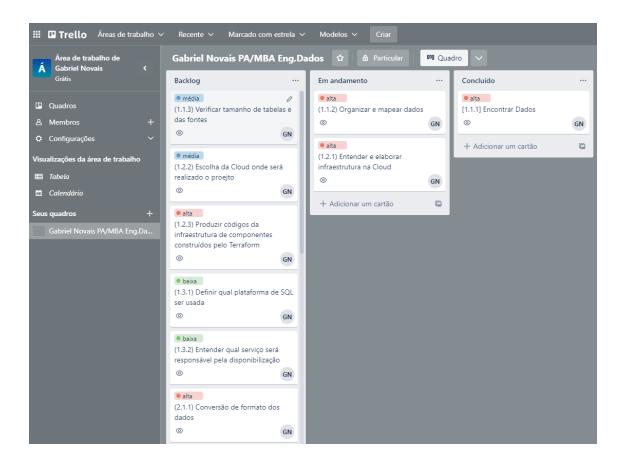


Figura 14 - Sprints, tarefas e tags no Trello

As tarefas estão numeradas conforme código, descrevendo, primeiro a Sprint, depois o requisito e terceiro a etapa. Elas são as que se seguem:

- 1.1.1) Encontrar Dados
- 1.1.2) Organizar e mapear dados
- 1.1.3) Verificar tamanho de tabelas e das fontes

¹ https://trello.com/b/U8By4GoR/gabriel-novais-pa-mba-engdados



- 1.2.1) Entender e elaborar infraestrutura na Cloud
- 1.2.2) Escolha da Cloud onde será realizado o projeto
- 1.2.3) Produzir códigos da infraestrutura de componentes construídos pelo Terraform
- 1.3.1) Definir qual plataforma de SQL ser usada
- 1.3.2) Entender qual serviço será responsável pela disponibilização
- 2.1.1) Conversão de formato dos dados
- 2.1.2) Tratamento dos dados
- 2.1.3) Agrupamento dos dados
- 2.2.1) Construir códigos responsáveis pelos componentes de processamento (cluster Kubernetes)
- 2.2.2) Testar componentes
- 2.2.3) Verificar Custos
- 2.3.1) Construir consultas SQL que serão utilizadas como testes
- 3.1.1) Gerar recomendações com base em alguma técnica.
- 3.1.2) Disponibilizar dados processados no S3 para consumo
- 3.2.1) Realizar testes de funcionamento da pipeline
- 3.3.1) Rodar consultas de testes



2. Área de Experimentação

O que significa esta seção?

Esta seção tem o objetivo de apresentar as evidências do planejamento dos requisitos selecionados do Backlog de Produto, além de mostrar a maneira como eles foram desenvolvidos e registrar os resultados alcançados.

É necessário expor a execução e a validação dos experimentos relacionados ao desenvolvimento da solução, ou seja, testar se você está no caminho certo ou se algo precisa ser modificado (pivotar).

Quais etapas já devem estar finalizadas no momento do preenchimento desta seção? (Pré-requisitos)

No momento do preenchimento, é esperado que você já tenha cursado a disciplina de Inovação e Design Thinking, em especial as etapas do processo de Design Thinking, além de estar se preparando para desenvolver a solução idealizada no seu Projeto Aplicado.

Você também já deve ter preenchido o primeiro capítulo deste relatório (CANVAS do Projeto Aplicado).

Como esta seção deve ser preenchida?

Esta seção é a área mais dinâmica do CANVAS do Projeto Aplicado. Nela você deverá inserir os experimentos necessários para desenvolver e validar cada Sprint. Ao final do experimento, você deverá preencher o item "Solução" da seguinte maneira:

- Evidência do Planejamento: comprove que os requisitos referentes à Sprint foram efetivamente planejados. Para isso, utilize o Trello e adicione, neste campo, uma cópia da tela da ferramenta com a Sprint planejada.
- Evidência da Execução de cada Requisito: para cada requisito planejado, adicione um artefato que comprove o cumprimento da etapa. Podem ser anexados, por exemplo, códigos, documentos, modelos, scripts, capturas de tela, entre outros. Importante: o número de artefatos adicionados deve ser o mesmo que o número de requisitos planejados.
- Evidência da Solução: os requisitos implementados contribuem para o alcance de um resultado geral, que deverá ser comprovado neste campo. Isso será feito



por meio de capturas de tela, gráficos, modelos, textos, figuras, tabelas, testes, entre outros.

Para cada Sprint, cite no item "Experiências vivenciadas" o que não foi validado, mas forneceu insights para ajuste da rota.

Quais ferramentas devem ser utilizadas?

Obs.: Para realização desta seção você deverá utilizar o Trello.



- 2.1 Sprint 1
- 2.1.1 Solução
 - Evidência do planejamento:
 - Evidência da execução de cada requisito:
 - Evidência dos resultados:
- 2.1.2 Experiências vivenciadas



2.2 Sprint 2

2.2.1 Solução

- Evidência do planejamento:
- Evidência da execução de cada requisito:
- Evidência dos resultados:

2.2.2 Experiências vivenciadas



2.3 Sprint 3

2.3.1 Solução

- Evidência do planejamento:
- Evidência da execução de cada requisito:
- Evidência dos resultados:

2.3.2 Experiências vivenciadas



3. Considerações Finais

3.1 Resultados

Por meio de um texto detalhado, apresente os principais resultados alcançados pelo seu Projeto Aplicado.

Cite os pontos positivos e negativos, as dificuldades enfrentadas e as experiências vivenciadas durante todo o processo.

3.2 Contribuições

Apresente quais foram as contribuições que o seu Projeto Aplicado trouxe para que o Desafio proposto fosse solucionado.

Cite, por exemplo, as inovações, as vantagens sobre os similares, as melhorias alcançadas, entre outros.

3.3 Próximos passos

Descreva quais são os próximos passos que poderão contribuir com o aprimoramento da solução apresentada pelo seu Projeto Aplicado.