Interface gráfica do usuário, Texto, Aplicativo, chat ou mensagem de texto

Descrição gerada automaticamente

**Instituto de Gestão e Tecnologia da Informação**

**Relatório do Projeto Aplicado**

Lives analytics para negócios

Pedro Henrique Faria Teixeira

Orientador(a):

Daniel Viana

01/06/2022

****

**Pedro Henrique Faria Teixeira**

**INSTITUTO DE GESTÃO E TECNOLOGIA DA INFORMAÇÃO**RELATÓRIO DO PROJETO APLICADO

Lives analystics para negócios

Relatório de Projeto Aplicado desenvolvido para fins de conclusão do curso [...].  
  
Orientador (a): Daniel Viana

**Araguari  
01/06/2022**

**Sumário**

[1. CANVAS do Projeto Aplicado 4](#_Toc101281510)

[1.1 Desafio 5](#_Toc101281511)

[1.1.1 Análise de Contexto 5](#_Toc101281512)

[1.1.2 Personas 6](#_Toc101281513)

[1.1.3 Benefícios e Justificativas 7](#_Toc101281514)

[1.1.4 Hipóteses 8](#_Toc101281515)

[1.2 Solução 9](#_Toc101281516)

[1.2.1 Objetivo SMART 9](#_Toc101281517)

[1.2.2 Premissas e Restrições 11](#_Toc101281518)

[1.2.3 Backlog de Produto 13](#_Toc101281519)

[2. Área de Experimentação 14](#_Toc101281520)

[2.1 Sprint 1 16](#_Toc101281521)

[2.1.1 Solução 16](#_Toc101281522)

[● Evidência do planejamento: 16](#_Toc101281523)

[● Evidência da execução de cada requisito: 16](#_Toc101281524)

[● Evidência dos resultados: 16](#_Toc101281525)

[2.1.2 Experiências vivenciadas 16](#_Toc101281526)

[2.2 Sprint 2 17](#_Toc101281527)

[2.2.1 Solução 17](#_Toc101281528)

[● Evidência do planejamento: 17](#_Toc101281529)

[● Evidência da execução de cada requisito: 17](#_Toc101281530)

[● Evidência dos resultados: 17](#_Toc101281531)

[2.2.2 Experiências vivenciadas 17](#_Toc101281532)

[2.3 Sprint 3 18](#_Toc101281533)

[2.3.1 Solução 18](#_Toc101281534)

[● Evidência do planejamento: 18](#_Toc101281535)

[● Evidência da execução de cada requisito: 18](#_Toc101281536)

[● Evidência dos resultados: 18](#_Toc101281537)

[2.3.2 Experiências vivenciadas 18](#_Toc101281538)

[3. Considerações Finais 19](#_Toc101281539)

[3.1 Resultados 19](#_Toc101281540)

[3.2 Contribuições 19](#_Toc101281541)

[3.3 Próximos passos 19](#_Toc101281542)

## 1. CANVAS do Projeto Aplicado

Figura conceitual, que representa todas as etapas do Projeto Aplicado.



## Desafio

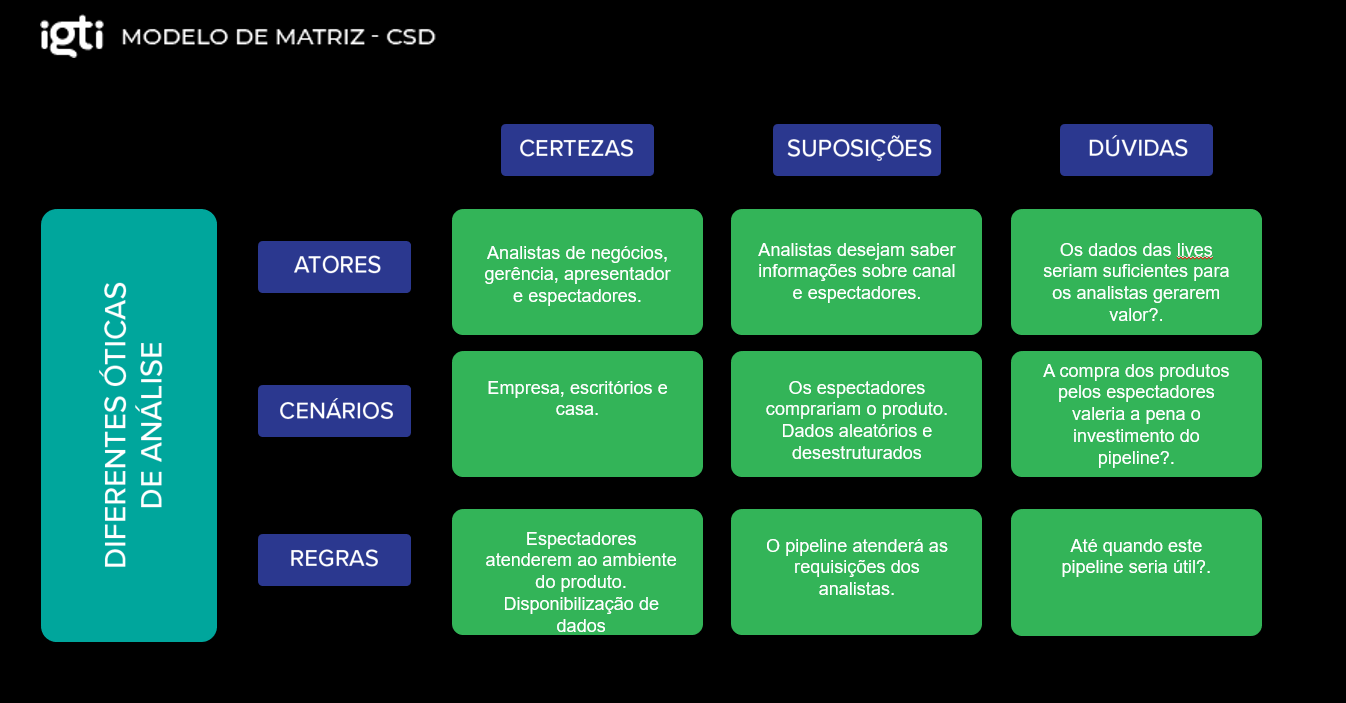
### 1.1.1 Análise de Contexto

Uma empresa possui um produto e quer encontrar uma boa maneira de fazer divulgações a fins de aumentar suas vendas. Dado que hoje estamos vivenciando dias onde podcasts e lives estão sendo realizados diariamente e em praticamente qualquer horário do dia, esta empresa então optou por divulgar seus produtos por meio dessas lives, porém ela não sabe como escolher o melhor canal, apresentador e nicho de espectadores.

Sendo assim, a empresa precisa construir uma aplicação para fazer a captação dos dados dessas lives e disponibilizar em um datalake para que os seus analistas de dados possam conectar suas ferramentas de BI a fim de fazer as análises e relatórios para entregar a gerência e assim estes realizarem o levantamento de negócio e avaliação de viabilidade, para então contratarem o canal ou apresentador para representarem a sua marca.

Dado isso, o desafio para resolver esse problema é construir um pipeline para extrair e concentrar as informações dos espectadores das lives/podcasts em um datalake, disponibilizando esses dados para os analistas fazerem as análises, aplicando regras de disponibilização que deve ser documentada e seguir todos os padrões de LGPD para que assim a gerência possa analisar a viabilidade do anuncio do produto em determinado canal.

Matriz CSD:



Observação POEMS:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PESSOAS | OBJETOS | AMBIENTE | MENSAGEM | SERVIÇOS |
| Quem está presente no contexto em análise? | Que objetos fazem parte do ambiente? | Quais são as características do ambiente? | Que mensagens são comunicadas? | Quais serviços são oferecidos? |
| Apresentador | Computadores, microfones, fones de ouvidos, assentos e mesas | Ambiente descontraído onde se pode conversar sobre assuntos diversos | Mensagem verbal informal ou formal sobre um assunto ou acontecimento | Comunicação com pessoas por meio de uma plataforma online |
| Espectador | Computadores, notebooks, mesas, assentos e smartphones | Ambiente diverso, formal ou informal. Pode ser sua casa, carro ou trabalho. | Mensagem de texto informal sobre um assunto ou acontecimento | Comentários por meio da plataforma online |
| Analista de dados | Notebooks, telas, mesas e assentos | Ambiente formal ou informal podendo ser a sede da empresa ou sua casa | Relatórios com informações sobre as lives | Análises e dashboards sobre os comentários dos espectadores |
| Gerência | Notebooks, telas, mesas e assentos | Ambiente formal ou informal podendo ser a sede da empresa ou sua casa | Relatórios sobre viabilidade do negócio | Análise decisória para a contratação de um canal ou apresentador |
| Registros  - Falta de dados  - Não existem relatórios nem análise de negócio | | | Insights  - Hoje as lives atraem uma alta taxa de espectadores e isso pode ser utilizado para vendas de produtos  - Criar pipeline de dados para automatizar a captação de dados | |

### 1.1.2 Personas



* + Nome: Laysa Vieira
  + Idade: 28 anos
  + Gênero: Feminino
  + Estado civil: Solteira
  + Educação: Pós-graduada em Gestão de Negócios e superior completo em Ciência da Computação
  + Profissão: Gerente de Projetos
  + Características comportamentais: É uma pessoa séria e formal, não gosta de bebidas alcoólicas e por isso frequenta somente ambientes reservados. Gosta bastante de sair com amigos para jantar. Seu hobby favorito é dormir.
  + Objetivos: Ter indicadores de espectadores que sejam capazes e tenham interesse em comprar produtos pela internet.



### 1.1.3 Benefícios e Justificativas

Os benefícios desse desafio são os seguintes: Criar uma estrutura de dados consistentes que será utilizada pela área de negócio da empresa para acelerar o processo de vendas de seus produtos, isso inclui, agilização do processo de encontrar público alvo, aumento de capital em vendas, maior porcentagem de venda por publicação, mitigação de insatisfação de compra do produto e impacto financeiro negativo. Com os dados disponíveis no datalake a equipe de analista de dados poderá focar somente na parte de criação de dashboards e relatórios pois toda a infraestrutura dos dados já estará pronta para ser utilizada.

Outros benefícios que podem ser citados sobre a criação do piepline de dados:

- Centralização dos dados;

- Segurança e backup de dados;

- Automatização de processos;

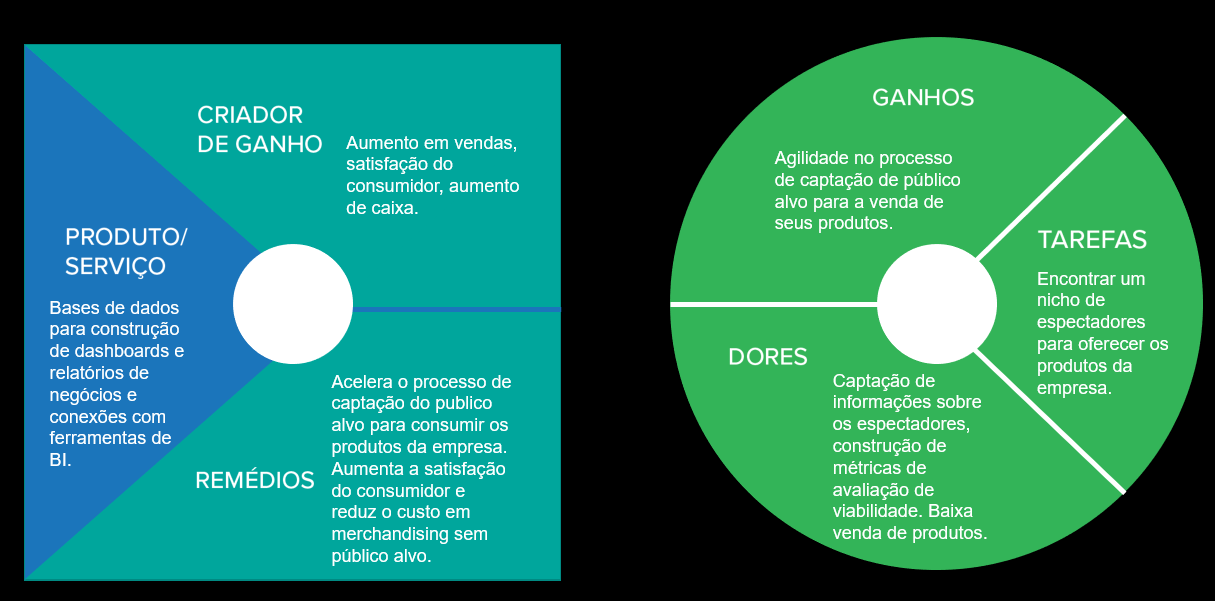
- Compartilhamento de informações

- Aumento de produtividade da equipe;

- Redução de esforços desnecessários.

**BluePrint:**

|  |  |
| --- | --- |
| Itens | Detalhamento |
| Objetivos | Conhecer o perfil do espectador, encontrar indicadores de adesão ao produto, realizar vendas de produtos por meio de canais que façam lives. |
| Atividades | Procurar indicações, procura opções de canais, procura pelos cronogramas das lives, encontrar um assunto de seu agrado, assistir ao programa analisar dashboards e relatórios, relatar viabilidade de merchandising do produto. |
| Questões | O que eu quero assistir? Quais os assuntos possíveis? Será que vai ser divertido ou entediante? Vai agregar ao meu conhecimento? Será que meus amigos também assistirão? O que eu irei comentar no chat da live? |
| Barreiras | Acesso a internet, local apropriado, forma de assistir ou ouvir, dados desatualizados, falta de relatórios, falta de espectadores e canais. |
| Ações do cliente | Quer assistir a um podcast, procura uma plataforma, faz o login, procura um canal que goste, espera o cronograma, começa a assistir. |
| Funcionalidades | Captação de público alvo para a venda de produtos, possibilidade criação de dashboards e relatórios analíticos, concentração de massa de dados para análise, retenção de dados históricos. |
| Interação | Possibilitar análise de negócio, diversificação de dados para análise, conexões com ferramentas de BI, classificação dos dados de espectadores |
| Mensagem | Descubra o melhor canal para divulgar seus produtos, encontre o nicho de pessoas para vender seus produtos, aumente suas vendas com a análise dos nossos dados, veja qual o ambiente seu produto se adapta as vendas. |
| Onde ocorre | Casa, escritório, site, aplicativos, internet. |
| Tarefas aparentes | Dashboards, relatório de análises, propagandas dos produtos. |
| Tarefas escondidas | Bancos de dados locais, esteiras de pipeline, conexões de ferramentas. |
| Processos de suporte | Documentação dos dashboards, documentação das bases de dados, documentação de relatórios, suporte ao consumidor, reembolso da compra. |
| Saída desejável | Aumentar as vendas da empresa, compras de produtos por espectadores de lives, aumento do caixa da empresa, satisfação de quem compra os produtos. |



### 1.1.4 Hipóteses

|  |  |  |
| --- | --- | --- |
| Nº | Observação | Hipótese |
| 1 | Dados são gerados aleatoriamente em tempo real/quase real nas plataformas | Identificar a melhor forma de extrai-los e enriquecer os dados |
| 2 | Os dados estão desestruturados | Identificar locais de armazenamento |
| 3 | Dependência de interação de espectadores para geração de dados | Captar canais com maior volume de acesso |
| 4 | Baixo nível de informação dos dados | Criar mecanismo de classificação |
| 5 | Acessibilidade aos dados pelo pessoal de analytics | Disponibilizar os dados em uma ferramenta de BI |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Ideias | Critérios de comparação | | | | | | Somatório |
| C1 | C2 | C3 | C4 | C5 | C6 |
| I1 | 1 | 4 | 3 | 5 | 2 | 3 | 18 |
| I2 | 2 | 3 | 1 | 2 | 5 | 2 | 15 |
| I3 | 3 | 4 | 5 | 2 | 1 | 1 | 16 |
| I4 | 4 | 2 | 3 | 2 | 3 | 5 | 19 |
| I5 | 2 | 3 | 5 | 2 | 3 | 2 | 17 |

## 1.2 Solução

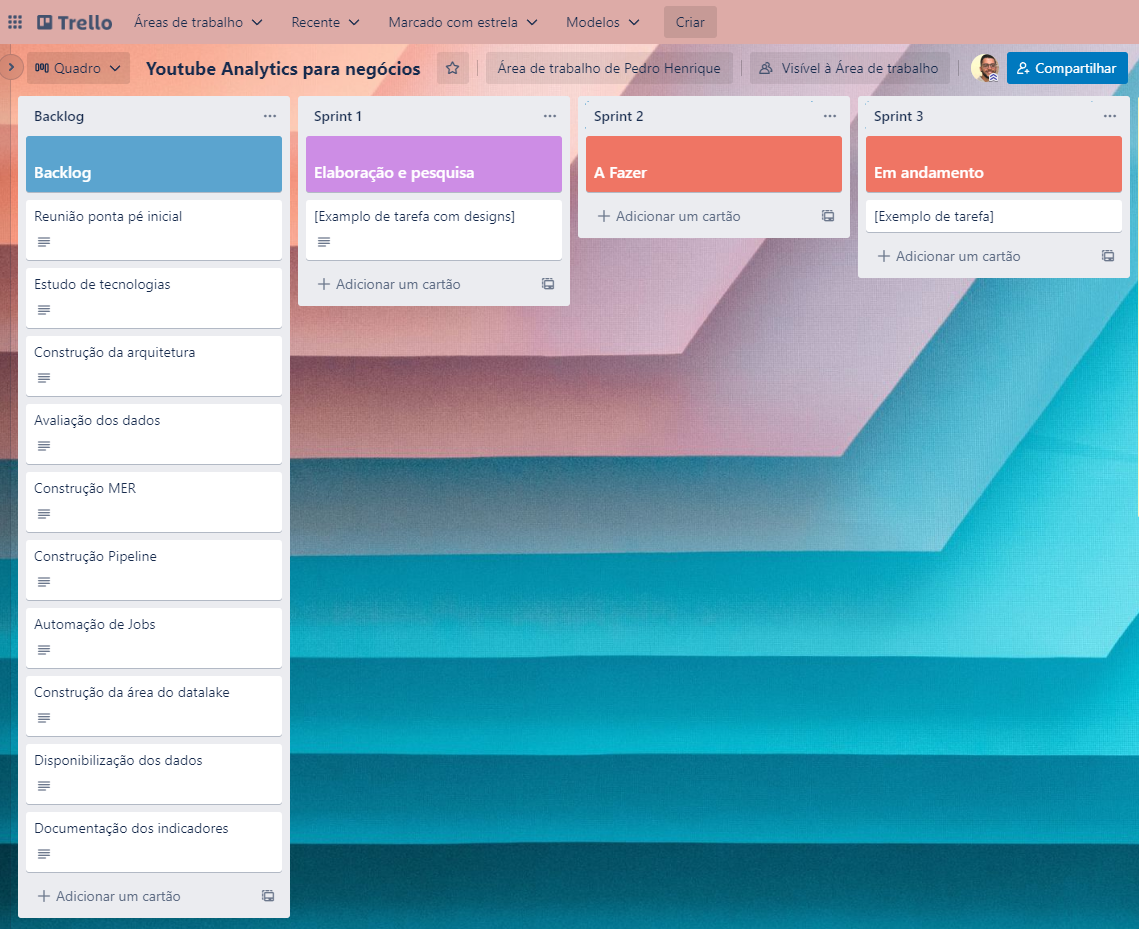
### Objetivo SMART

Construir até o encerramento da disciplina um pipeline de dados que atenda a demanda dos analistas de dados e gerência da empresa onde conseguiam extrair informações para realizar vendas de produtos da empresa para 10% dos espectadores de cada live/podcast realizada onde apareça o anuncio. Aumentar em 5% a taxa de venda dos produtos mensalmente. Encontrar canais onde os espectadores pertençam ao nicho do nosso produto ou então espectadores que estão aptos a pagarem pelo produto se divulgado de maneira correta pelo apresentador. Considerando a atração que as lives e podcasts estão recebendo do público, se torna um meio mais atual de se realizar merchandising, atingindo mais rápido o público alvo.

### 1.2.2 Premissas e Restrições

|  |  |  |  |
| --- | --- | --- | --- |
| Risco Identificado | Impacto Potencial | Ações preventivas | Ações corretivas |
| Dados aleatórios | Alto | Criar processos de classificação de dados | Classificar os dados |
| Dados inconsistentes ou duplicados | Alto | Criar processos de validação antes do armazenamento | Identificar e eliminar dados duplicados ou inconsistentes |
| Dados desestruturados | Alto | Criar estruturação de dados para o armazenamento | Estruturar os dados |
| Dados indisponíveis ou ausentes | Alto | Disponibilizar os dados para acesso dos analistas | Regras de disponibilização imediata dos dados mediante solicitação |
| Dados sensíveis de espectadores | Alto | Criar sistema de identificação de dados sensíveis | Deixar os dados sensíveis anônimos e aplicar LGPD |

### 1.2.3 Backlog de Produto



# 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:

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

#### Evidência da execução de cada requisito:

**Reunião ponta pé inicial:**

Foi realizada duas reuniões com a gerência e a equipe que será responsável por fazer a construção do projeto *Lives* *analytics* para negócios, com o intuito de discutir qual o problema a ser solucionado, quais os passos para solucionar os problemas, quais seriam os possíveis impeditivos e quais os são os principais indicadores para o negócio. Nessa reunião foi discutido também as possíveis tecnologias a serem utilizadas e o porquê de serem utilizadas, além do formato de extração e armazenamento dos dados. Foi definido que o banco de dados a ser utilizado para consulta será o MySQL devido a fácil acessibilidade pelos membros da equipe e por ser um banco de dados relacional onde os analistas e cientistas de dados já possuem certa facilidade em consultar com o SQL. Porém as tecnologias posteriores que levaram os dados até o banco foram deixadas em aberto para uma avaliação e estudo das tecnologias para levantar aspectos de performance, manutenção e custo. Além do banco de dados os dados serão armazenados em um *datalake* com a estrutura de pastas (*bronze*, *silver*, *gold*), para manter dos arquivos segmentada de menos tratado e mais tratado, facilitando também o backup do banco de dados.

|  |  |
| --- | --- |
| **Reunião 18/06/2022** | |
| Participantes | Laysa Vieira, Pedro Teixeira, Marcos Vilela |
| Tema | Ponta pé inicial |
| Assuntos tratados | Primeiros passos, entendimento do problema, ferramentas, disponibilização dos dados |
| Comentários importantes | Os dados são gerados em real-time, assim temos que criar uma estrutura que suporte essa coleta e armazene de forma adequada. Temos que criar uma estrutura que suporte análise em *batchs* para manter dados históricos e backups. |

|  |  |
| --- | --- |
| **Reunião 20/06/2022** | |
| Participantes | Laysa Vieira, Pedro Teixeira, Marcos Vilela |
| Tema | Discussão de indicadores |
| Assuntos tratados | Estrutura do *schema*, relevância dos indicadores, quais os mais importantes |
| Comentários importantes | Os dados são coletados das *lives* no formato *json*, o que já é bem estruturado para dados crus, é um ponto positivo para a evolução do pipeline  Levantamento das principais *features* e hipóteses que podem ser coletadas a partir dos indicadores presentes no *schema* |

**Estudo de tecnologias:**

**Construção da arquitetura:**

Essa arquitetura foi projetada pensando em 3 aspectos, disponibilidade dos dados, custo, escalabilidade.

Escalável pois neste primeiro momento será construído em cima de ferramentas locais que são facilmente escaláveis para a nuvem, o que também já impacta na redução de custos, pois são ferramentas ***opensources*** ou então possuem um acesso grátis inicial.

O aspecto de custo se dá por conta de as ferramentas em sua maioria serem de acesso livre e que quando é escalada para a nuvem o custo aumenta, porém o retorno de valor ao negócio é na maioria das vezes maior do que é gasto com as ferramentas.

Já em relação a disponibilidade dos dados, tem se um datalake onde é possível armazenar os dados desde quando são “crus” e até quando são mais performáticos e estruturados, tendo possibilidade de versionamento e backup, junto com o banco de dados MySQL que servirá como ferramenta de consulta ao datalake.

**Tabela

Descrição gerada automaticamente com confiança baixa**

**Construir uma POC para viabilizar a coleta de dados:**

Código responsável por fazer a captura dos dados do Youtube/Twitch por meio da biblioteca **pytchat** e disponibilizar em um tópico do Kafka no formato **json**.

Texto

Descrição gerada automaticamente

Código responsável por ler os dados do tópico e assim salvá-lo em um arquivo **json** para ser feito a análise inicial da estrutura do **schema**.

Texto

Descrição gerada automaticamente

**Avaliação dos Dados:**

Exemplo de arquivo **json** que é enviado pelo *KafkaProducer*, nele podemos perceber os indicadores que aparecem quando um autor faz uma ação no *chat* da *live*. Aqui temos diversas informações, sendo elas relevantes e não relevantes para o projeto que será desenvolvido. Julgando o problema a ser solucionado que é encontrar possíveis autores a comprarem um produto, os indicadores que elevam essa possibilidade são os seguintes:

- i***d***: ID do autor que fez a interação com a *live*;

- ***name***: Nome do autor;

- ***message***: Mensagem em texto que o autor enviou;

- ***datetime***: Data da interação que o autor fez na *live*;

- ***isVerified***: Indicador se o autor possui um canal verificado pelo Youtube;

- ***isChatSponsor***: Se o autor é patrocinador do *live* chat;

- ***amountValue:*** Indicador do valor da doação para o canal;

- ***amountString:*** Indicador do valor da doação para o canal;

- ***type***: Esse indicador possui dois valores que podem ser considerados monetários, sendo eles: “***superChat***” que indica que um autor doou dinheiro para fazer alguma propaganda ou interagir na *live* e “***newSponsor***” que indica que um novo autor se tornou um patrocinador.

Com esses indicadores os analistas e cientistas de dados conseguiram levantar informações relevantes como a participação do autor, tanto monetária quanto proativa. Se o autor é um usuário ativo em *lives* ou não, se o autor tem disposição de comprar um produto ou até mesmo levantar hipóteses de sentimentalidade em cima das mensagens de texto.

Texto

Descrição gerada automaticamente

#### 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.