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**

Base de dados de índice de reclamação da ANS

Rafael Rodrigues Nascimento Cardoso

Orientador(a): João Paulo Barbosa Nascimento

30/10/2022

****

**Rafael Rodrigues Nascimento Cardoso**

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

Base de dados de índice de reclamação da ANS

Relatório de Projeto Aplicado desenvolvido para fins de conclusão do curso MBA em Engenharia de Dados.  
  
Orientador (a): João Paulo Barbosa Nascimento

**São Paulo  
30/10/2022**

**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)

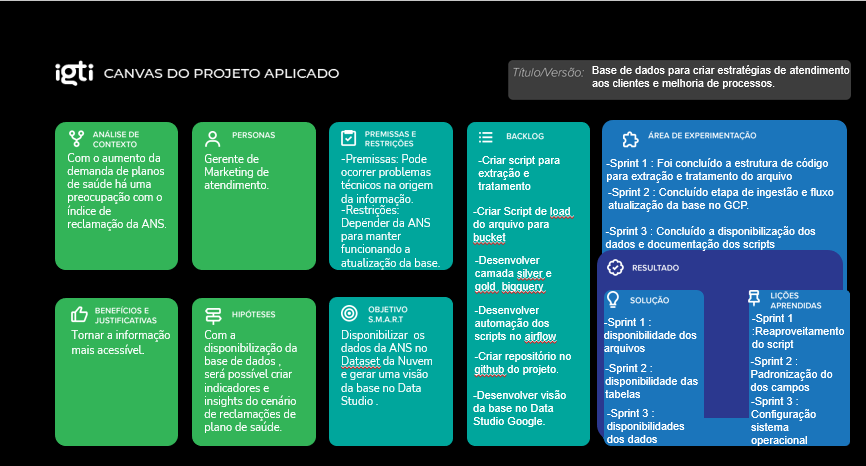
[3.1 Resultados 19](#_Toc101281540)

[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

O mercado de planos de saúde teve uma crescente de clientes nos anos recentes, devido a pandemia conforme informado pela ANS, na figura 1 e 2 ilustra esse cenário.

  
Figura 1: <https://www.poder360.com.br/saude/numero-de-pessoas-com-planos-de-saude-cresceu-5-na-pandemia/>

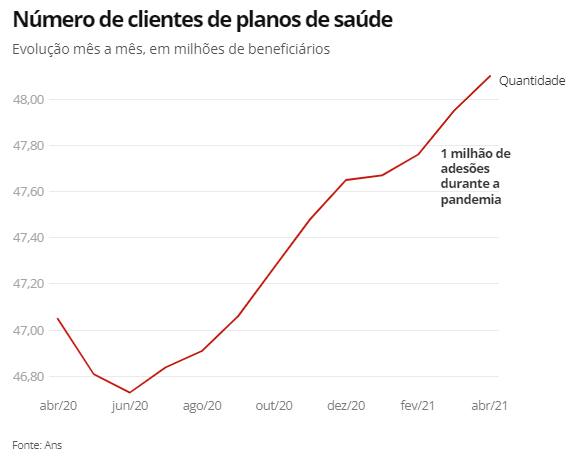


Figura 2: <https://g1.globo.com/economia/noticia/2021/05/21/planos-de-saude-ganham-mais-de-1-milhao-de-clientes-na-pandemia-e-atingem-maior-nivel-em-quase-5-anos.ghtml>

Houve também uma crescente de startups de saúde em busca de oferecer um serviço mais acessível aos clientes, portanto a categoria vem tendo investimentos superior com os demais setores, conforme mostra a figura 3.

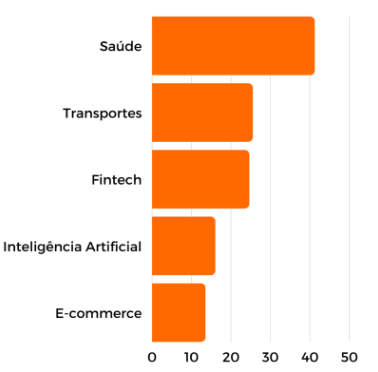


Figura 3: Os 5 segmentos de Startups com maior crescimento de investimentos <https://49educacao.com.br/startup/mercado-de-startups-no-brasil-2021/>

Portanto fica claro o aumento da demanda do setor de saúde e com isso as operadoras cada vez mais vem inovando e aprimorando formas de evitar problemas ao oferecer o serviço para o cliente. Onde preventivamente vem sendo necessário agir nas possíveis reclamações registradas na ANS (Agência Nacional de Saúde Suplementar). Considerando um exemplo de reclamações das operadoras de médio porte teve um aumento do índice no início de 2022, havendo um pico considerável em maio e com uma pequena baixa em julho conforme ilustrado na figura 4.

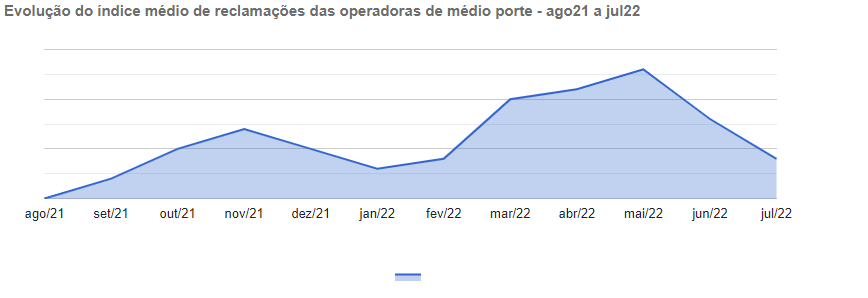


Figura 4: <https://www.ans.gov.br/indicedereclamacoes/indice_reclamacao_igr.asp>

Diante dos fatos mencionados, o desafio se dá através da necessidade de aumentar a capacidade analítica da área e agir preventivamente com analises preditivas.

Para o auxílio no mapeamento com detalhamento das possíveis soluções foi utilizado a matriz CSD e a observação tipo POEMS

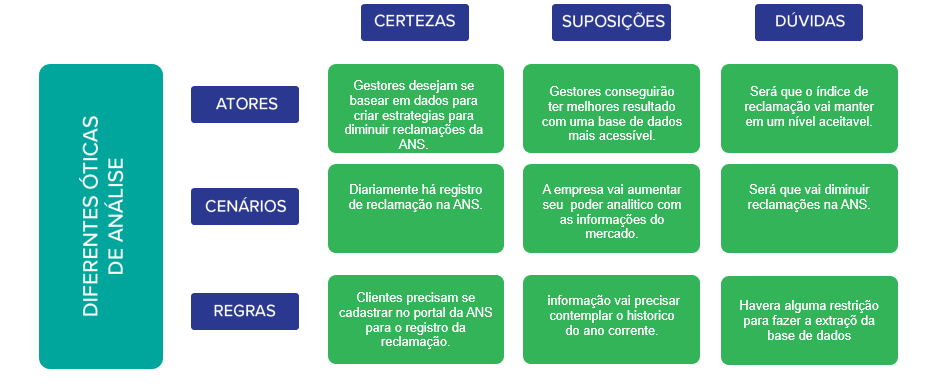


Figura 5:Matriz CSD.

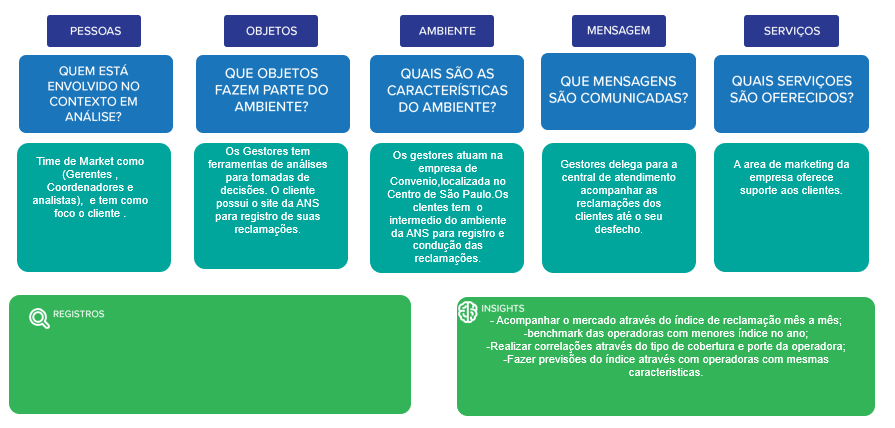


Figura 6:Analise POEMs.

### 1.1.2 Personas

Para identificação do perfil do usuário foi utilizado o recurso do Mapa de Empatia e o detalhamento extraído das etapas realizadas anteriormente.

Nome: Adolfo

Idade: 42 anos

Profissão: Gerente de marketing

Características comportamentais: Adolfo e um profissional experiente em áreas de marketing e atendimento com foco em resultados de satisfação do cliente.

Atualmente vem estudando formas para solucionar o aumento do índice de reclamações gerado pelos clientes no site da ANS.

Com um perfil inovador, Adolfo busca transformar sua área mais analítica e dinâmica diante das mudanças do mercador e surgimento de novas tecnologias.

Pessoa identificada nesse contexto do projeto, é um profissional de marketing, com cargo de gerente de atendimento. Responde pelos resultados de clientes na companhia, onde precisa decidir sobre estratégias de atendimento, e o retorno dessa atividade vem na satisfação dos clientes.

  
Figura 7:Mapa de empatia.

### 1.1.3 Benefícios e Justificativas

O projeto foi trabalhado em cima das dores que uma área de marketing sofre no contexto de insatisfação de seus clientes. E nesse sentido houve um indicativo que a melhor direção a tomar seria tornar a informação mais acessível e democrática para a área com o sentido de criar estratégias de manter um índice de reclamação aceito e coerente com ações desenvolvidas juntos aos clientes ao longo do tempo. Nessa direção foi utilizado a análise de Blue Print e Proposta de Valor.

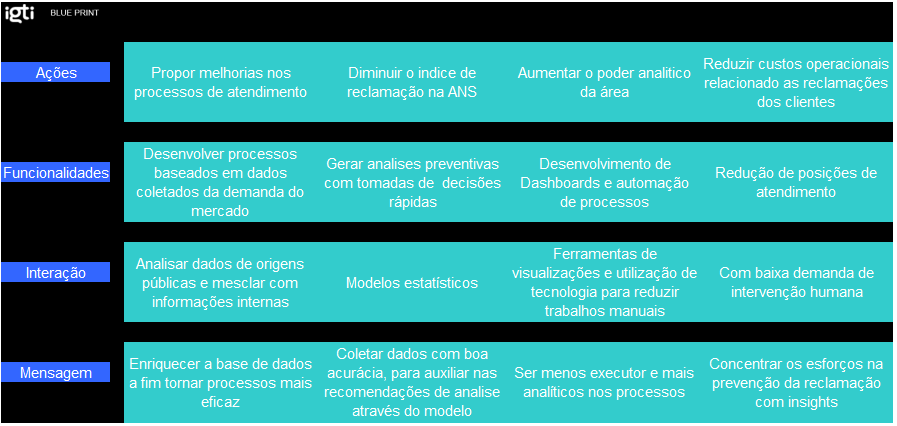
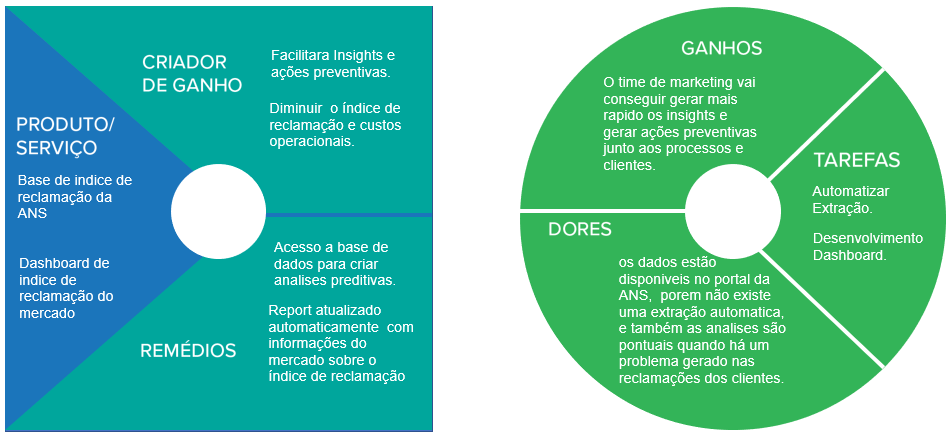


Figura 8:Blue Print

  
Figura 9: Explicação da proposta de valor

### 1.1.4 Hipóteses

Nesse ponto do projeto foi trabalhado a Matriz de observações para hipóteses e Priorização de Ideias para mapear e organizar as ideias diante hipóteses gerado nesse desafio.

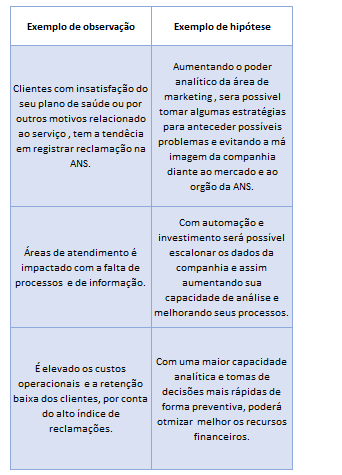


Figura 10: Matriz de observações para hipóteses.

Utilizando a matriz da figura 10, foi necessário priorizar como será inserido essas ideias no projeto, ilustrado na figura 11

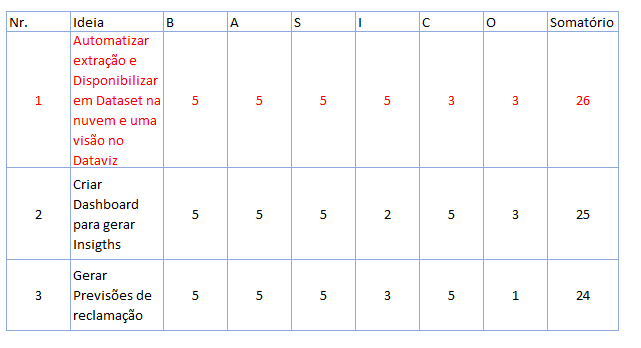
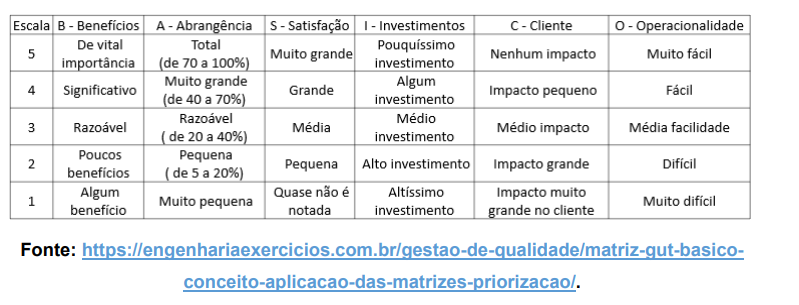


Figura 11: Priorização de Ideias.

O modelo da priorização de ideias foi baseado da tabela 1.

   
Tabela1:Matriz básico

## 1.2 Solução

### 1.2.1 Objetivo SMART

Essa fase do projeto foi escolhido um objetivo para condução da solução. Portanto, foi utilizado a definição do objetivo SMART que atende por S- eSpecífico, M – mensurável, A – Atingível, R – Relevante e T – temporal. Nesse sentido, o objetivo desse projeto e disponibilizar dados de índice de reclamação na ANS através de um banco de dados de forma autônoma e com atualização incremental para análise de histórico dos meses do ano vigente, os dados ficarão disponíveis no ambiente da Nuvem e também será criado uma visão dessa base em um Dataviz (Data Studio Google) que terá a conclusão no fim de Dezembro de 2022.

### Premissas e Restrições

Para definir as premissas e restrições, foi utilizado a matriz de riscos, ilustrado na figura 12

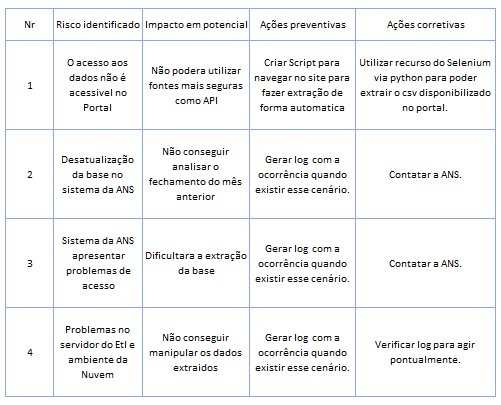


Figura 12:Matriz de Riscos.

### Backlog de Produto

Foi classificado nessa etapa os requisitos para execução do projeto ao longo do prazo determinado criando backlog com 4 cards e distribuindo estes nas sprints a seguir conforme ilustrado na figura 13.

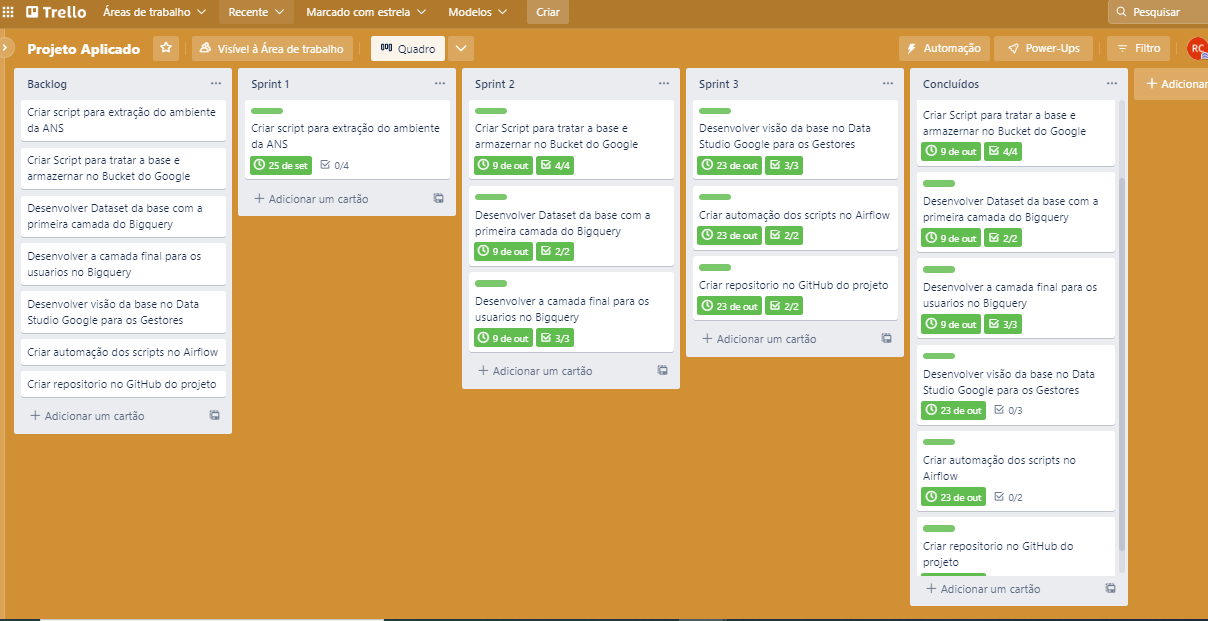


Figura 13:Trello do Projeto.

# 2. Área de Experimentação

O objetivo dessa etapa e formalizar e documentar toda as evidencias do proposto no mapeamento do Backlog de Produto e demonstrar os resultados evidenciando todas etapas e registrando as lições aprendidas ao longo desse desenvolvimento da sprint 1.

## 2.1 Sprint 1

A etapa da sprint 1, contextualiza a forma que se deu a captura dos dados, para ser utilizado na construção da solução. Demonstrando as evidências das execuções e resultados obtidos durante a etapa e naturalmente identificar as lições aprendidas no proposito da etapa programada.

### 2.1.1 Solução

O produto final do proposto seria a criação da infraestrutura do código possível de realizar a extração do arquivo no site e conseguir tratar o formato desse arquivo e disponibilizando em um diretório local de forma padronizada e automática, contudo essa solução vai agregar valor na estratégia definida para entrega do projeto completo envolvendo as necessidades da persona descrita na etapa do desafio e solução.

#### Evidência do planejamento:

A figura 14 ilustra as tarefas que foram concluídas da programação da sprint 1, no termino dessa etapa foi contemplado a entrega final da extração, capturando a informação de forma automática. Dessa forma no contexto geral do projeto, foi útil essa sprint no sentido de entregar a etapa inicial que seria o tratamento dos dados extraídos da origem.

A importância da evidência do planejamento colaborou para documentar o checklist proposto nessa etapa inicial da sprint 1.

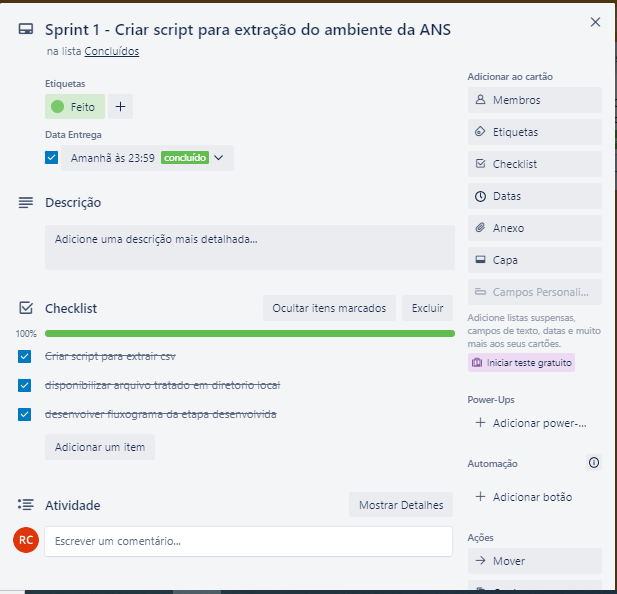


figura 14: Atividades do card sprint 1

#### Evidência da execução de cada requisito:

O vídeo 1 a seguir, demonstra a execução da extração e tratamento do arquivo e a disponibilização em um diretório local. Com isso evidencia as seguintes tarefas.

- Criar script para extrair csv  
 - disponibilizar arquivo tratado em diretório local



Video1: Evidência das etapas de execução

A figura 15 demonstra em destaque o que a sprint 1 entregou na visão geral do fluxograma do projeto. Com isso evidencia a atividade seguinte.

- Desenvolver fluxograma das etapas desenvolvidas

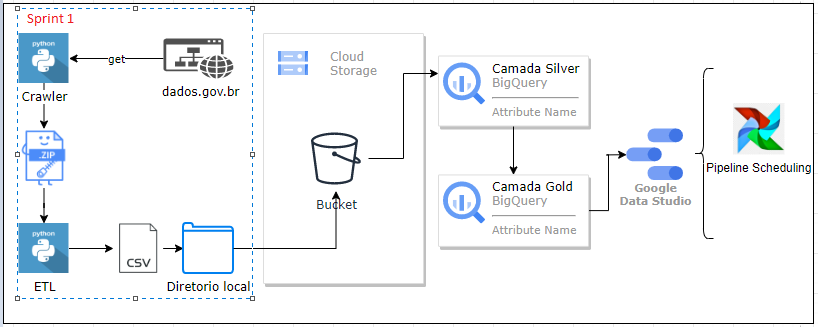
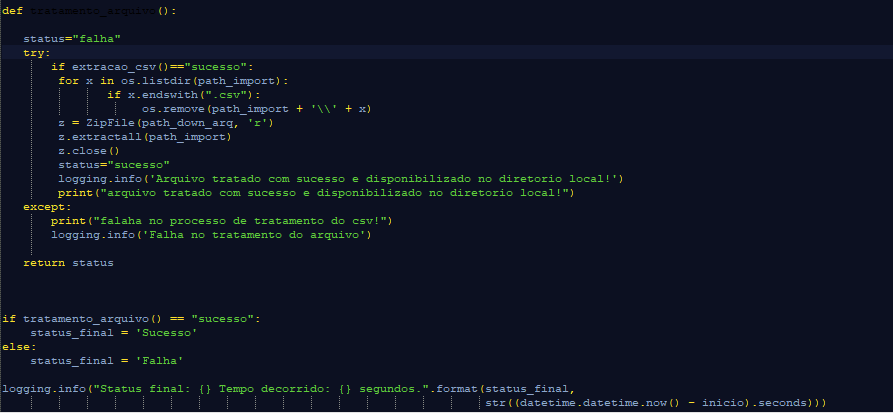


Figura 15: Fluxograma do Processo da solução

#### Evidência dos resultados:

Como infraestrutura de código, a figura 16 e 17 demonstra o resultado do script desenvolvido que foi proposto nessa etapa.



  
 Figura 16: Script Python arquivo extracao\_indice\_reclamacao.py

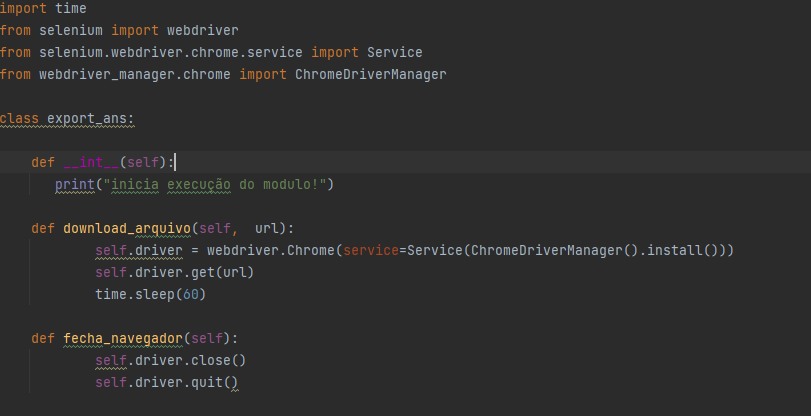
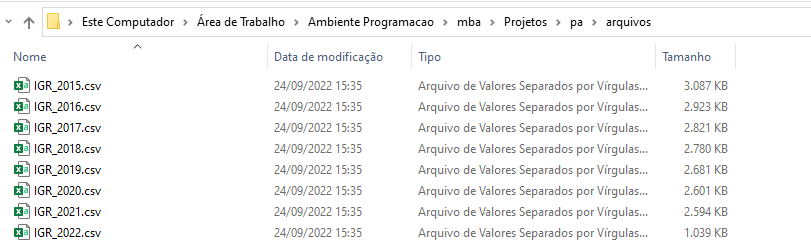


Figura 17: Script Python arquivo modulo\_selenium\_ans.py

Considerando os dados, os resultados extraídos da sprint 1 contempla a disponibilidade dos arquivos extraídos em um diretório local, contendo um histórico desde 2015.A figura 18 demonstra esse resultado.

  
 Figura 18: diretório local

### 2.1.2 Experiências vivenciadas

A infraestrutura do código teve a separação em dois arquivos Python sendo o primeiro com as funções de manipulação da extração e tratamento dos arquivos e o segundo com as configurações do navegador do Chrome. Pois como há mudanças constantemente em versões do driver do Chrome, fica melhor em dá manutenção nesse segundo arquivo. Caso também há necessidade de implantar mais alguma extração porem com outro link de arquivo, poderá utilizar esse segundo arquivo Python para aproveitar a estrutura de código desenvolvida. A sprint 1 encerra dessa forma inicial de disponibilidade dos arquivos e na seguinte sprint vamos trabalhar a etapa de load e armazenamento desses arquivos no ambiente de nuvem utilizando os serviços do Google GCP como Cloud Storage(Bucket) e Bigquery para disponibilização das tabelas em camadas Silver e Gold para uma melhor pratica e performance para tornar esses dados mais acessível para a persona descrita na etapa do desafio e solução proposto nesse projeto.

## 2.2 Sprint 2

A etapa da sprint 2, vai tratar os métodos e forma que foram aplicadas na ingestão e armazenamento dos dados que foram extraídos e tratados na sprint1, portanto será evidenciado as execuções e resultados desse processo.

### 2.2.1 Solução

A solução dessa etapa vai agregar ao projeto toda a garantia de atualização e ingestão das camadas proposta no backlog ao ponto de entregar velocidade e performance do processo.

#### Evidência do planejamento:

As figuras expostas nesse tópico, ilustra os cards definidos na sprint 2 e o checklist de cada um dos cards, a figura 19 ilustra o card de criação e ingestão do bucket.

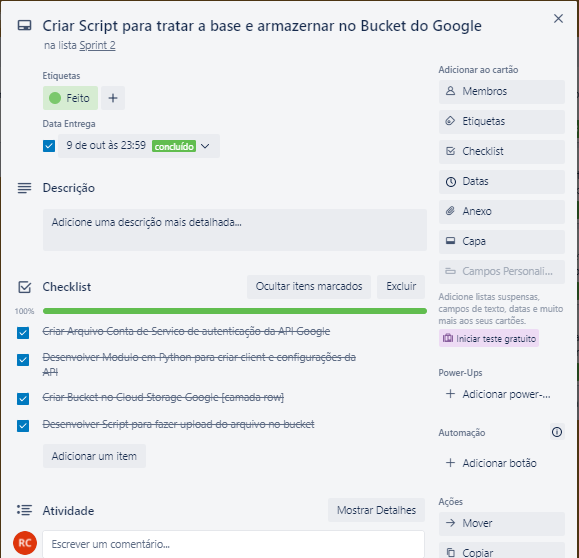


Figura 19: Card de criação do bucket e ingestão.

A figura 20 ilustra o planejamento da criação da camada silver no Bigquery, contendo a criação do dataset e a tabela conectada nos arquivos armazenados no bucket.

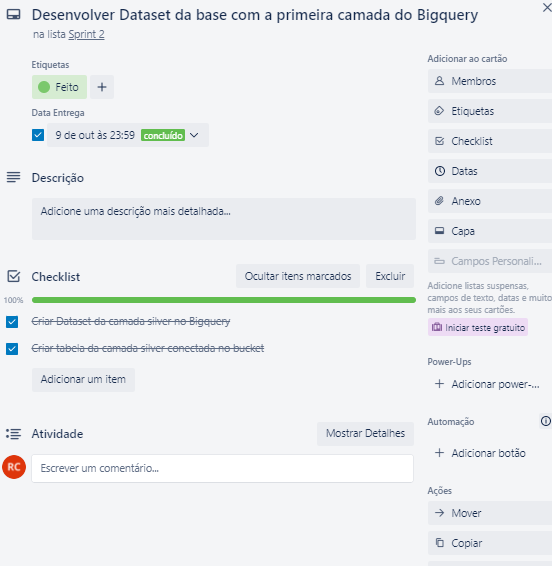


Figura 20: Card de criação do dataset silver no Bigquery.

A figura 21 ilustra o planejamento da criação da camada gold no Bigquery, contendo a criação do dataset e a tabela conectada na camada silver.

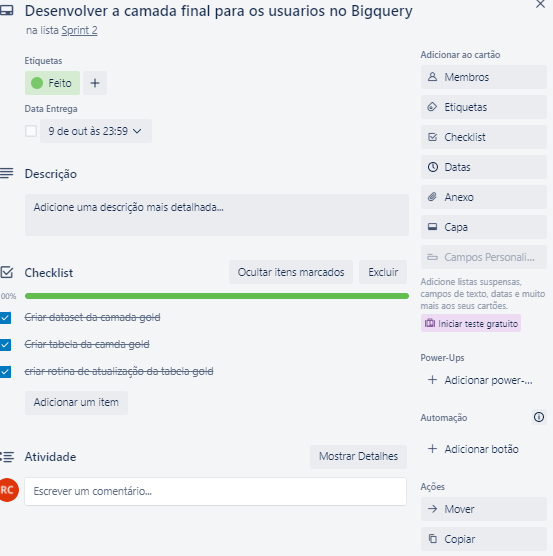


Figura 21: Card de criação do dataset gold no Bigquery.

#### Evidência da execução de cada requisito:

A figura 22 e 23 ilustra a etapa da criação da conta de serviço para os scripts consiga interagir via API com os recursos do Google (GCP).

Com isso evidencia a seguinte etapa do checklist:

-Criar Arquivo Conta de Servico de autenticação da API Google

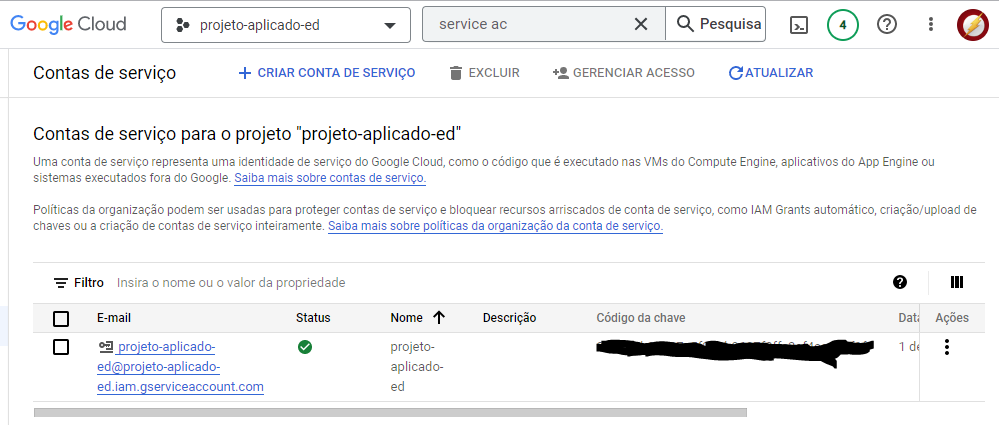


Figura 22: Criação da conta de serviço no Google(GCP).



Figura 23: Arquivo Json da conta de serviço.

A figura 24 ilustra o arquivo Python do modulo e o trecho do código que demonstra a criação da autenticação com a API do Google(GCP) para que seja possível realizar os processos proposto na sprint 2 .

Com isso evidencia a seguinte etapa do checklist:

- Desenvolver Modulo em Python para criar client e configurações da API

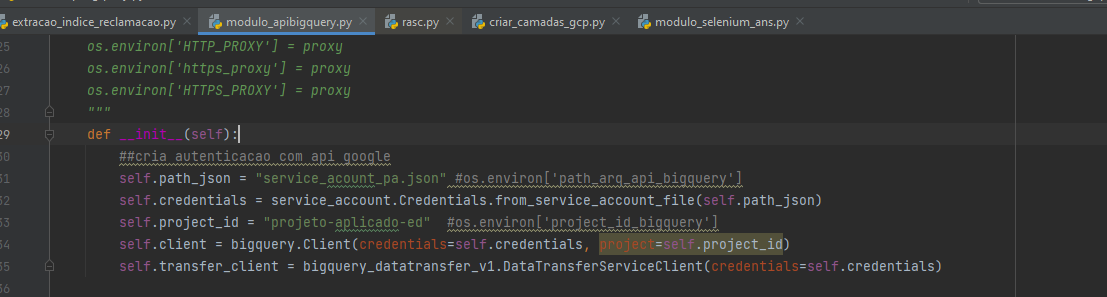


Figura 24: Trecho do código que cria o client.

A figura 25 ilustra a função do código que cria a camada row no bucket do Cloud Storage, para receber os arquivos tratados da extração.

Com isso evidencia a seguinte etapa do checklist:

- Criar Bucket no Cloud Storage Google [camada row]

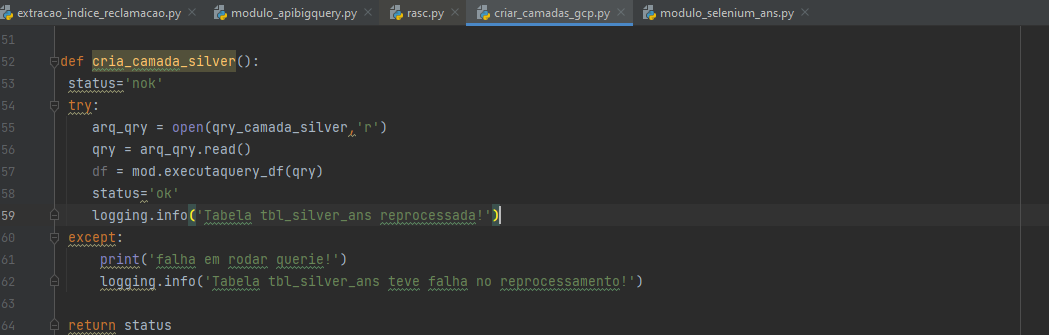


Figura 25: Trecho do código que cria o bucket no Cloud Storage

A figura 26 ilustra a função do código que faz o upload dos arquivos no bucket. Esse trecho está conectando ao modulo que contém os recursos do bigquery conforme mencionado nas etapas anteriores.

Com isso evidencia a seguinte etapa do checklist:

- Desenvolver Script para fazer upload do arquivo no bucket

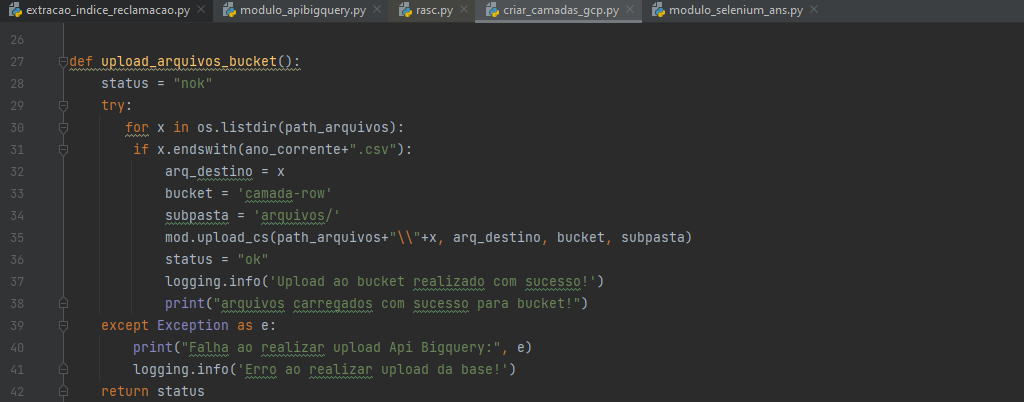


Figura 26: Trecho do código que faz upload dos arquivos no diretório local para o bucket

A figura 27 ilustra a função do código que cria dataset no Bigquery da camada silver.Esse trecho está conectando ao modulo que contém os recursos do bigquery conforme mencionado nas etapas anteriores.

Com isso evidencia a seguinte etapa do checklist:

- Criar Dataset da camada silver no Bigquery

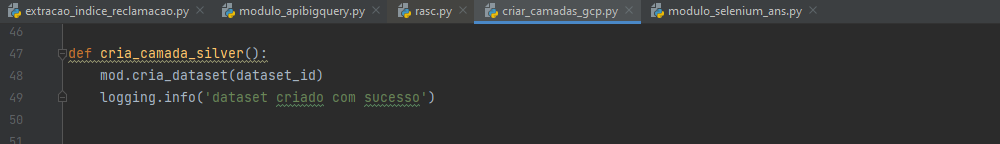


Figura 27: Trecho do código que cria dataset

A figura 28 e 29 ilustra a função do código que cria a tabela no Bigquery da camada silver, com a opção da função externa conectada no bucket dos arquivos para deixar dinâmico a atualização, pois qualquer alteração no bucket vai refletir automaticamente nessa tabela com isso a figura 29 ilustra a query que e executado na função da figura 28 para criar a tabela externa.Esse trecho está conectando ao modulo que contém os recursos do bigquery conforme mencionado nas etapas anteriores.

Com isso evidencia a seguinte etapa do checklist:

- Criar tabela da camada silver conectada no bucket

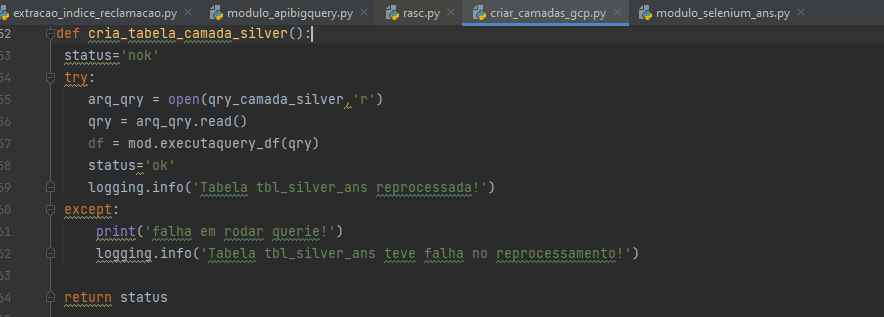


Figura 28: Trecho do código que cria tabela



Figura 29: Query que cria a tabela externa

A figura 30 ilustra a função do código que cria a camada Gold no Bigquery.Esse trecho está conectando ao modulo que contém os recursos do bigquery conforme mencionado nas etapas anteriores.

Com isso evidencia a seguinte etapa do checklist:

- Criar dataset da camada gold

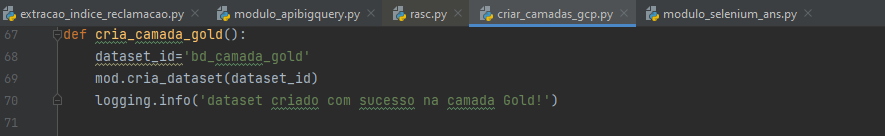


Figura 30: Trecho do código que cria dataset

A figura 31 ilustra a função do código que cria a tabela da camada Gold no Bigquery.Esse trecho está conectando ao modulo que contém os recursos do bigquery conforme mencionado nas etapas anteriores. A figura 32 ilustra a query que é executada nessa função para criar a tabela.

Com isso evidencia a seguinte etapa do checklist:

- Criar tabela da camada gold

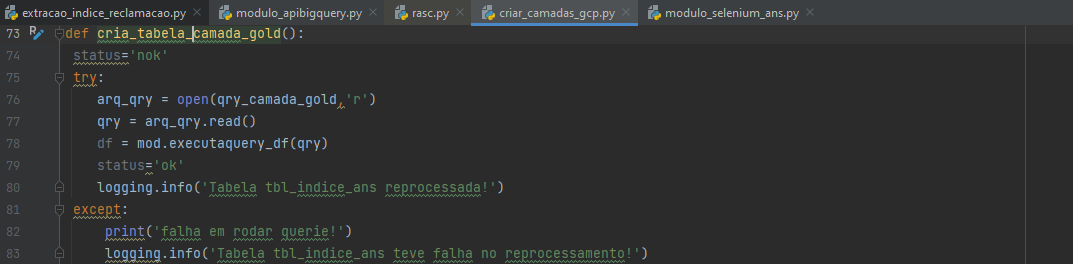


Figura 31: Trecho do código que cria a tabela na camada Gold

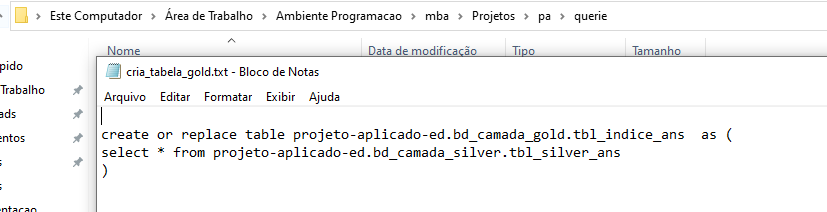


Figura 32: Trecho do código que cria a tabela na camada Gold

A figura 33 ilustra a função do código que faz a atualização da tabela da camada gold no Bigquery.Esse trecho está conectando ao modulo que contém os recursos do bigquery conforme mencionado nas etapas anteriores.A figura 34 ilustra a query e executado nessa função para realizar a ingestão da tabela.

Com isso evidencia a seguinte etapa do checklist:

- Criar rotina de atualização da tabela gold

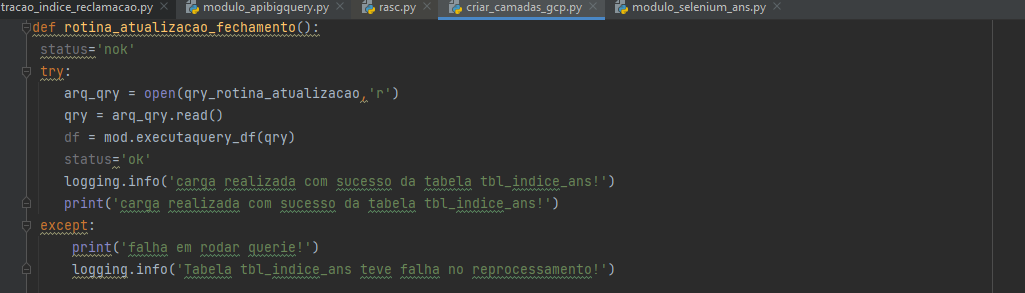


Figura 33: Trecho do código que faz a rotina de ingestão

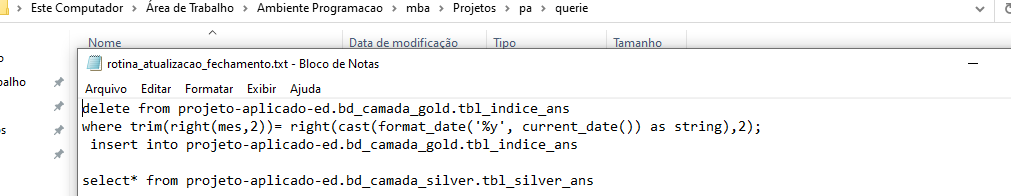


Figura 34: Query que atualiza a tabela

#### Evidência dos resultados:

A figura 35 evidencia a ingestão do arquivo do ano corrente no ano na camada row do bucket do Cloud Storage.

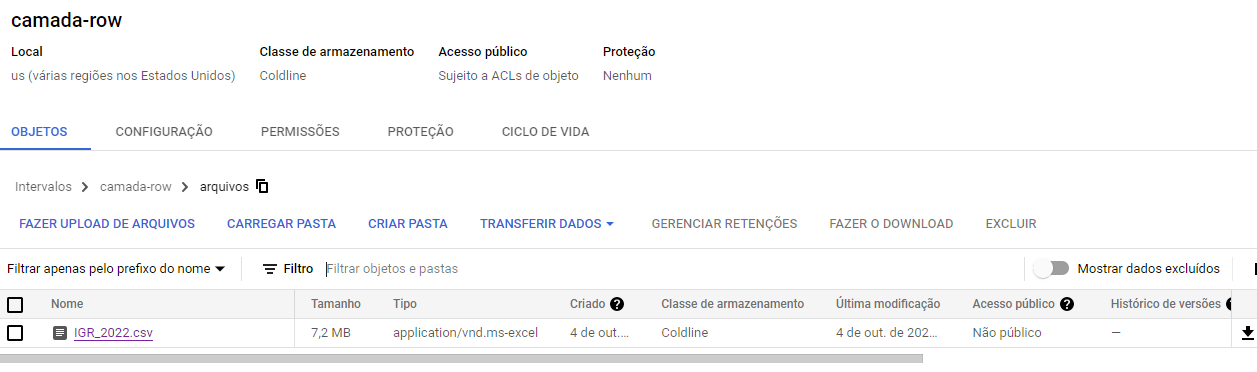


Figura 35: Trecho do código que faz a rotina de ingestão

A figura 36 evidencia a criação do dataset no Bigquery da camada silver.

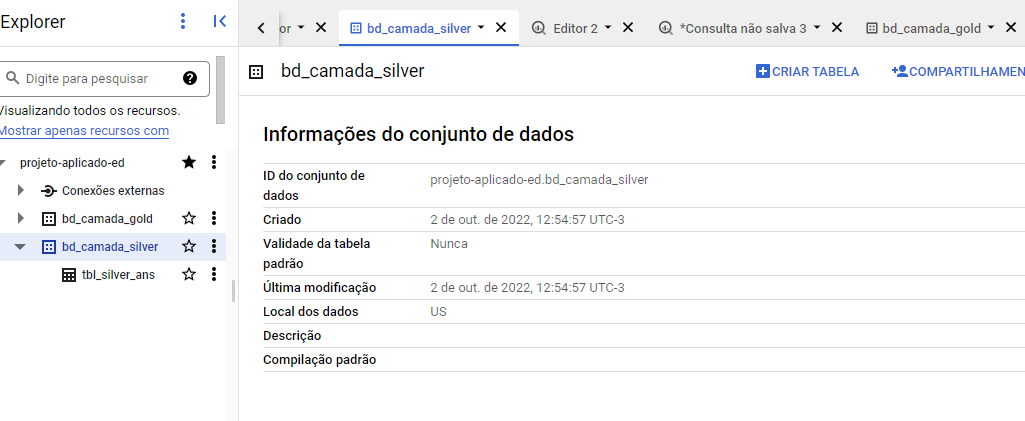


Figura 36: Dataset Silver Bigquery

A figura 37 e 38 evidencia o resultado da tabela da camada silver.

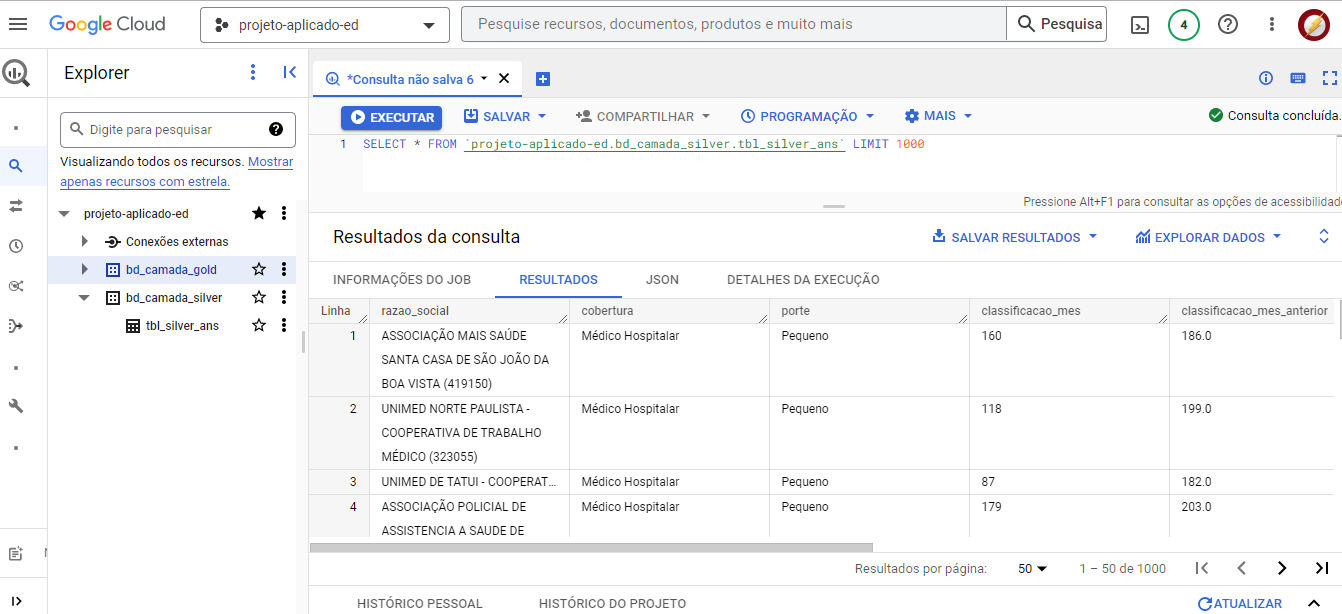


Figura 37: Tabela camada silver



Figura 38: Campos da tabela

A figura 39 evidencia o resultado do dataset da camada gold.

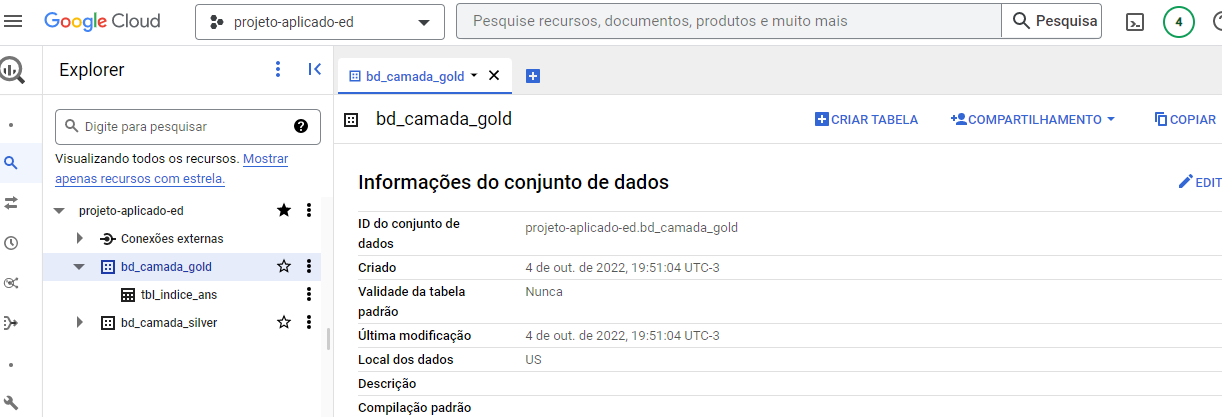


Figura 39: Dataset da camada gold

A figura 40 evidencia o resultado da tabela da camada gold.

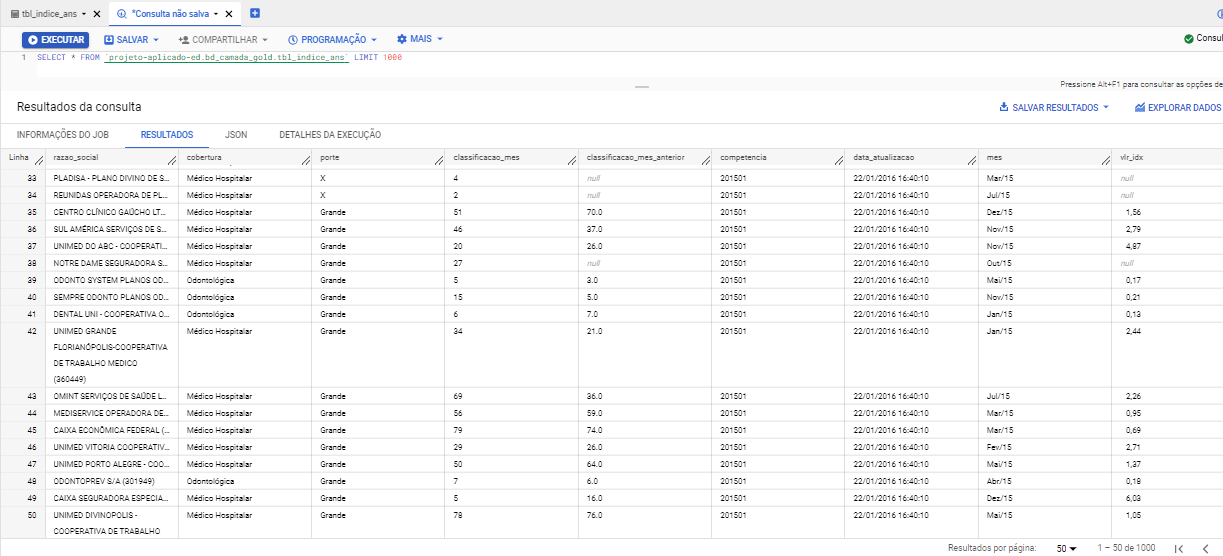


Figura 40: Tabela da camada gold

### 2.2.2 Experiências vivenciadas

A construção dessa etapa da sprint 2 foi bem interessante na questão dos contornos diante dos problemas que surgiram ao longo da sprint, tais como formato da base que e extraído na origem, conforme mostrado na figura 41 , os campos de meses ficam como coluna, isso dificultou a padronização do template da base , dessa forma foi necessário transformar essas colunas como linha para ser possível fazer um append dessas informações do ano corrente com o histórico desde 2015, dessa forma foi utilizado uma função no python para contornar essa situação antes de subir o arquivo no bucket conforme ilustrado na figura 42.

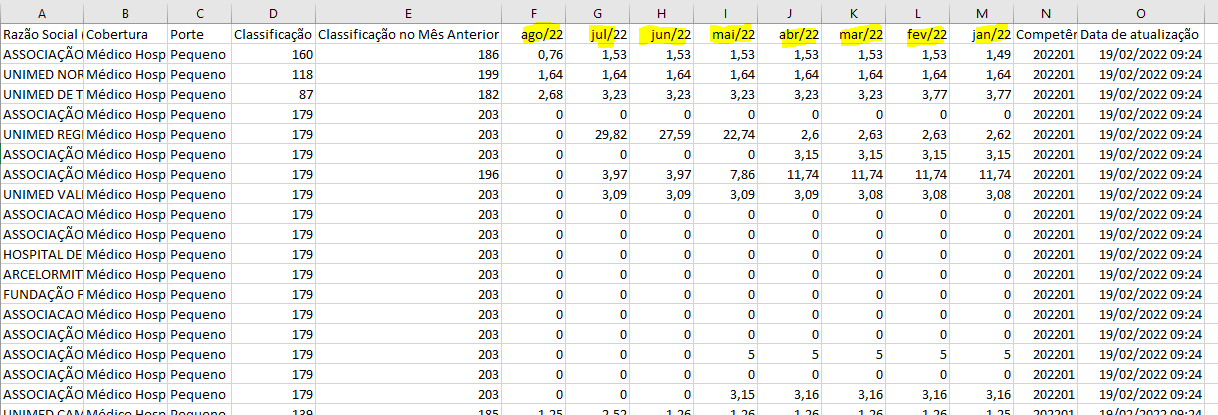


Figura 41: Formato do arquivo na origem

### 

Figura 42: Trecho do código da função de pivot

Outro ponto foi focar na performance de atualização dos dados, onde a camada gold ficou parametrizada para receber dados do ano corrente somente, os demais meses ficaram estáticos nela , e dessa forma não vai ser necessário processar sempre o histórico da base inteira que seria desde 2015 contém informações, conforme ilustrado na figura 43.

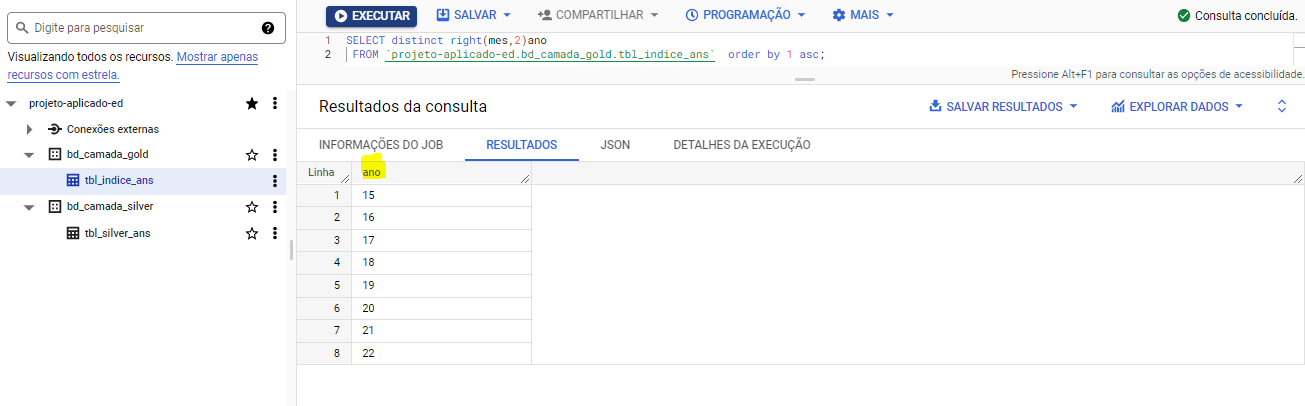


Figura 43: Anos históricos que contem na tabela final

Para concluir essa sprint, a figura 44 ilustra o fluxograma do projeto destacando o que a sprint 2 representou no contexto geral da entrega.

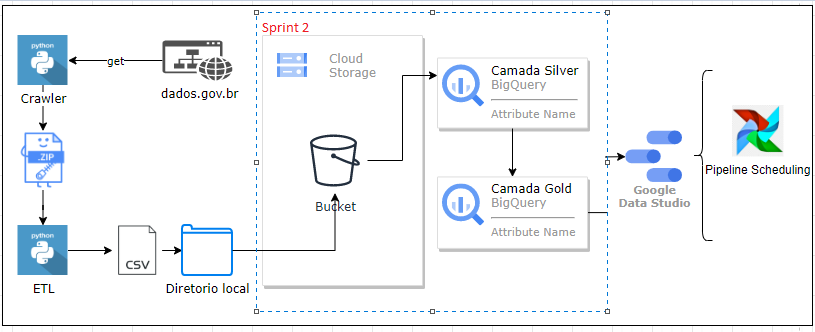


Figura 44: Fluxograma do Projeto

### 2.3 Sprint 3

A etapa da sprint 3, vai tratar os métodos e forma que foram aplicadas na disponibilização dos dados e a documentação no Github, portanto será evidenciado as execuções e resultados desse processo.

### 2.3.1 Solução

A solução dessa sprint vai agregar ao projeto toda a garantia de tornar os dados acessível para as áreas de negócio utilizar e tirar insights do relatório.

#### Evidência do planejamento:

As figuras expostas nesse tópico, ilustra os cards do Trello definidos no sprint 3 e o checklist de cada um dos cards, a figura 45 ilustra o card de criação do relatório no Data Studio.

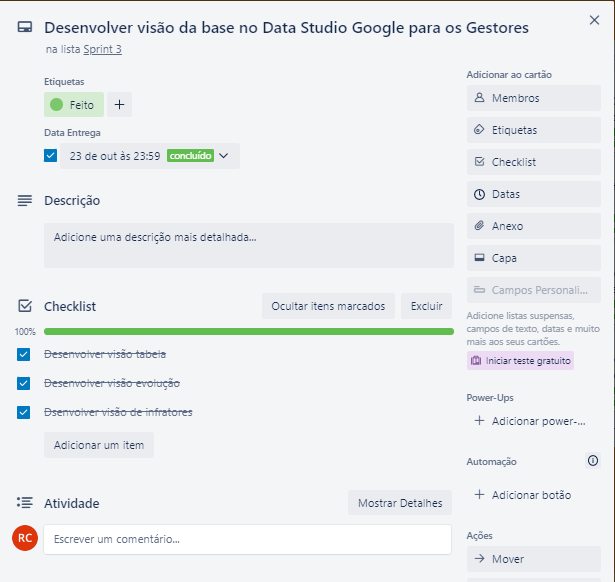


Figura 45: Card Data Studio

A figura 46 ilustra o card de automação dos scripts no Airflow.

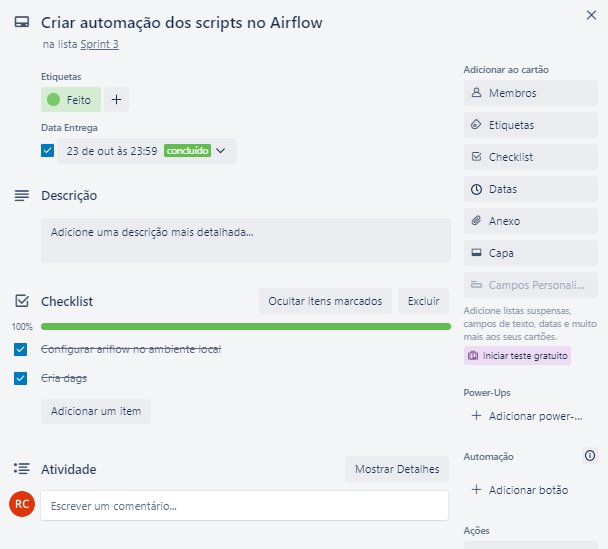


Figura 46: Card Airflow

A figura 47 demonstra o planejamento da tarefa de criar repositório do projeto no Github.

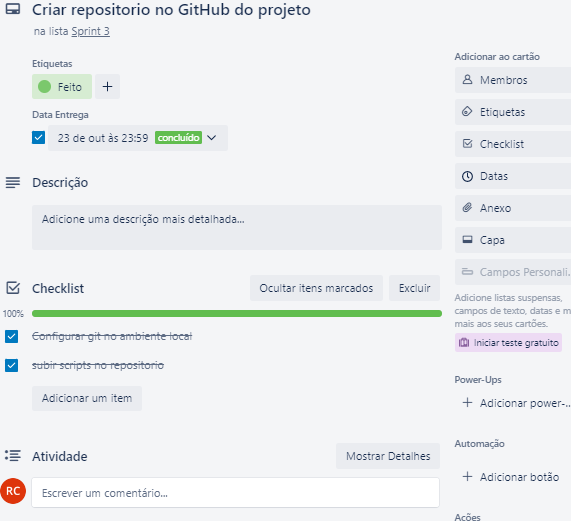


Figura 47: Card Github

#### Evidência da execução de cada requisito:

A figura 48 ilustra a visão de tabela no Data Studio

Com isso evidencia a seguinte etapa do checklist:

- Desenvolver visão tabela

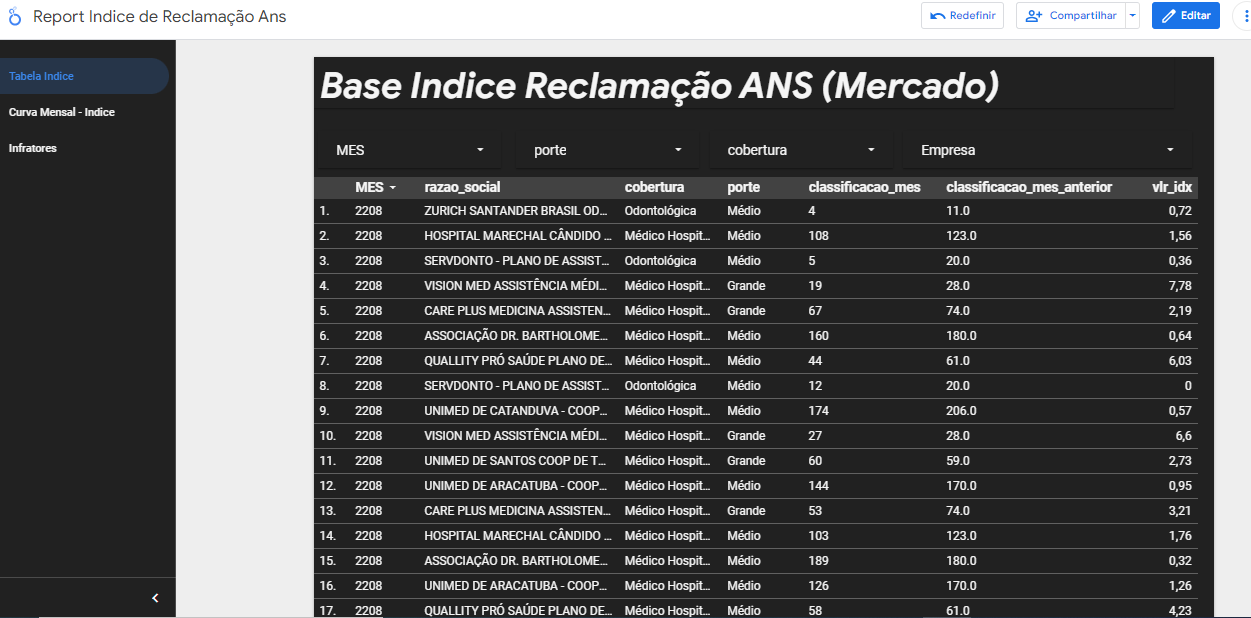


Figura 48: Visão tabela no Data Studio

A figura 49 ilustra a visão evolutiva no Data Studio

Com isso evidencia a seguinte etapa do checklist:

- Desenvolver visão evolução

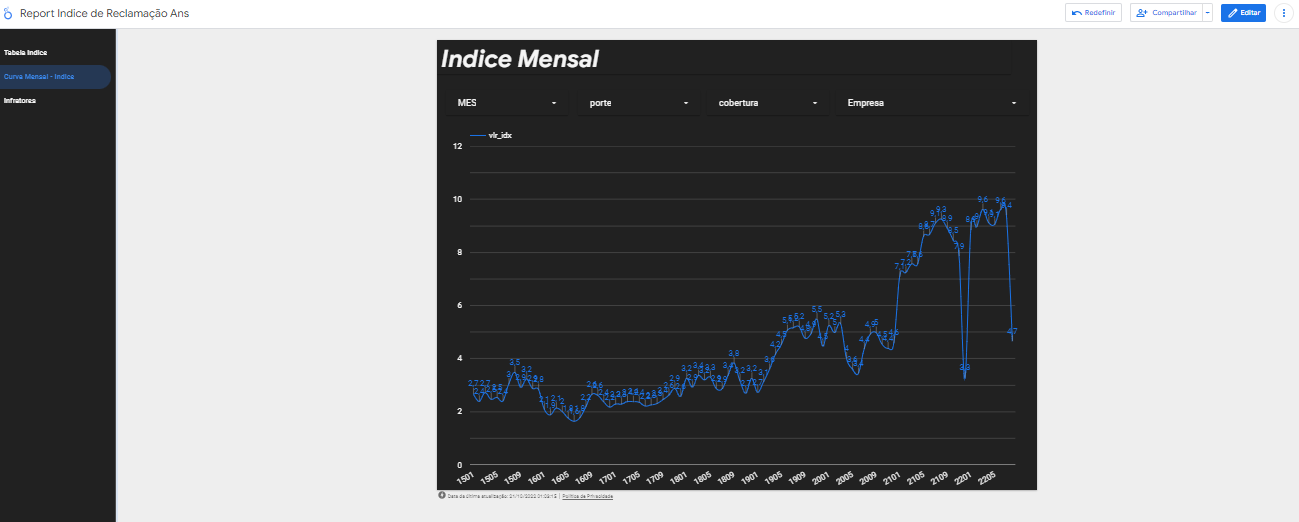


Figura 49: Visão Evolutiva

A figura 50 ilustra a visão de infratores no Data Studio

Com isso evidencia a seguinte etapa do checklist:

- Desenvolver visão de infratores

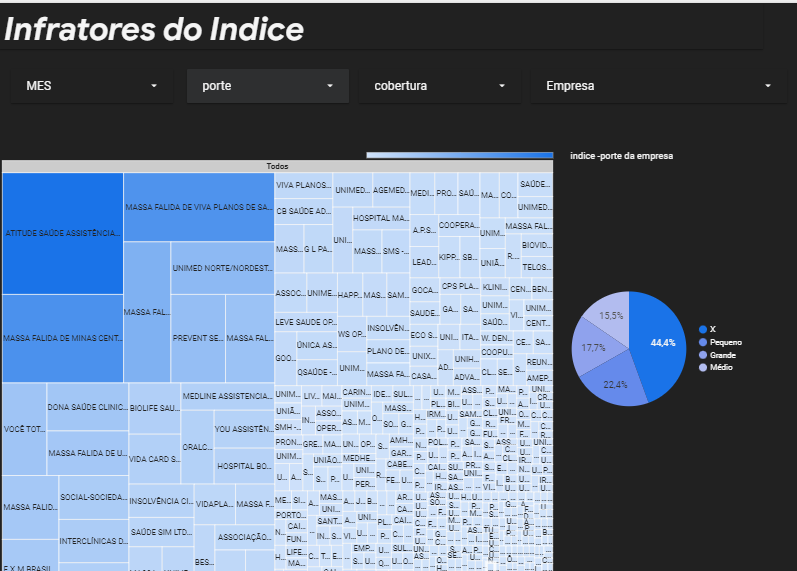


Figura 50: Visão de Infratores

A figura 51 ilustra a configuração do Airflow rodando no ambiente local

Com isso evidencia a seguinte etapa do checklist:

- Configurar ariflow no ambiente local

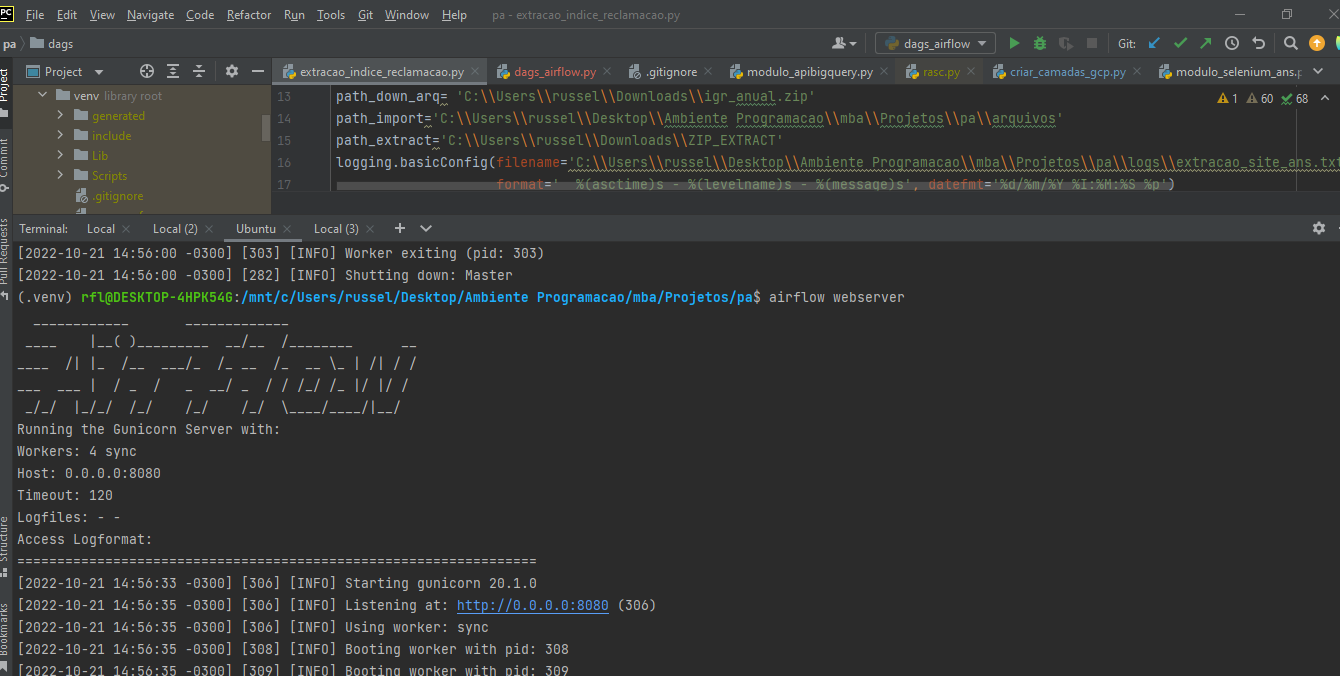


Figura 51: Airflow rodando no ambiente local

A figura 52 e 53 ilustra o script contendo a criação das dags para executar os scripts python desenvolvido nas sprints anteriores para extração e armazenamento dos dados.

Com isso evidencia a seguinte etapa do checklist:

- Cria dags

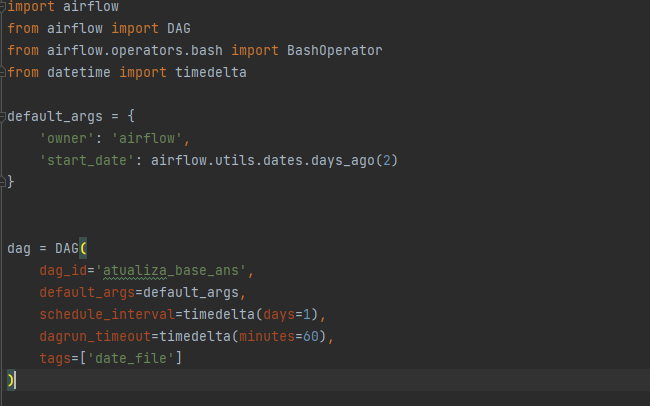


Figura 52: Primeiro trecho do script

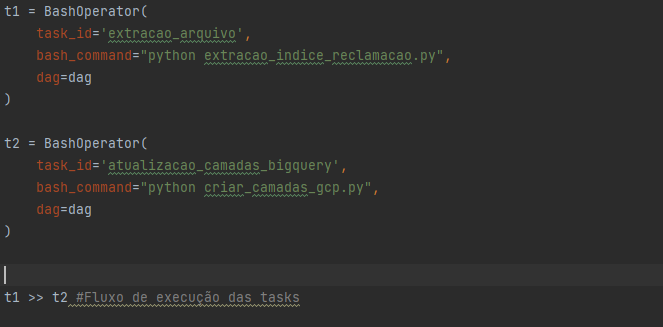


Figura 53: Segundo trecho do script

A figura 54 demonstra a versão do Github que está instalada no ambiente local.

Com isso evidencia a seguinte etapa do checklist:

- Configurar git no ambiente local



Figura 54: Versão do git instalado no ambiente local

A figura 55 demonstra os arquivos python no diretório criado para o projeto.

Com isso evidencia a seguinte etapa do checklist:

- subir scripts no repositório

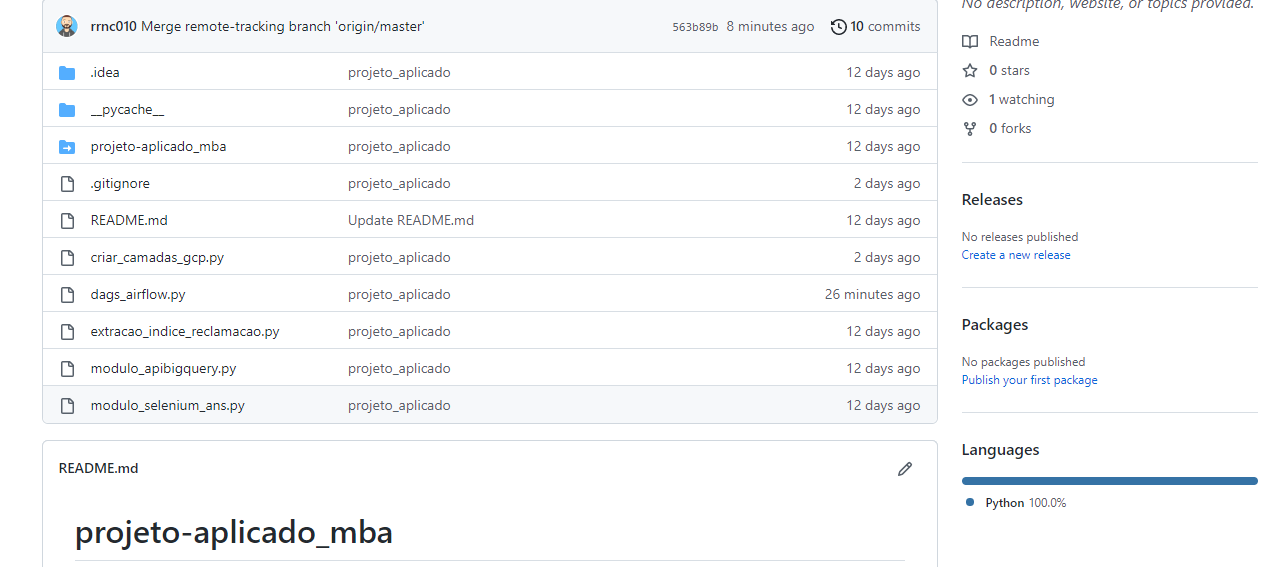
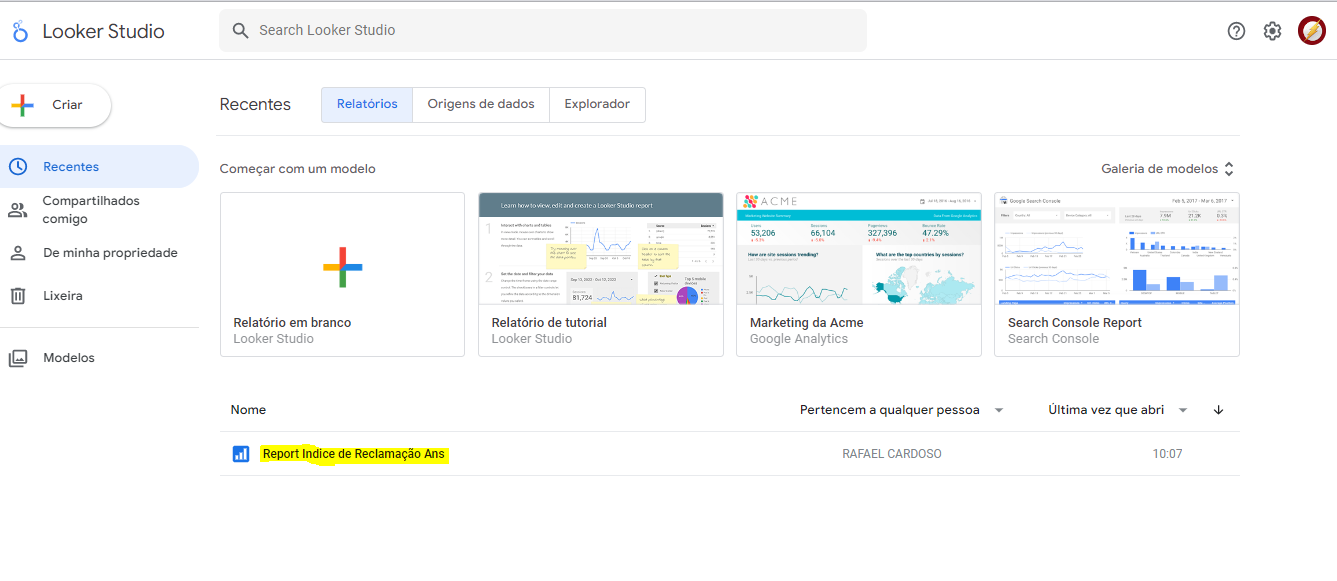


Figura 55: Arquivos adicionados no repositório

#### Evidência dos resultados:

A figura 56 evidencia o resultado da criação do relatório no Data studio conforme proposto no planejamento do backlog, onde este relatório está conectado na base do Bigquery que seria a camada gold sendo a tabela final para ser utilizada pelas áreas de negócio e este report auxiliara as áreas ter o acesso mais acessível as informações coletadas do índice de reclamação da Ans.

  
 Figura 56: Relatório no Data Studio

A figura 57 evidencia o resultado da criação das dags no Airflow conforme o proposto no planejamento do backlog, onde auxiliara na automação da execução dos scripts em python para processar os dados.

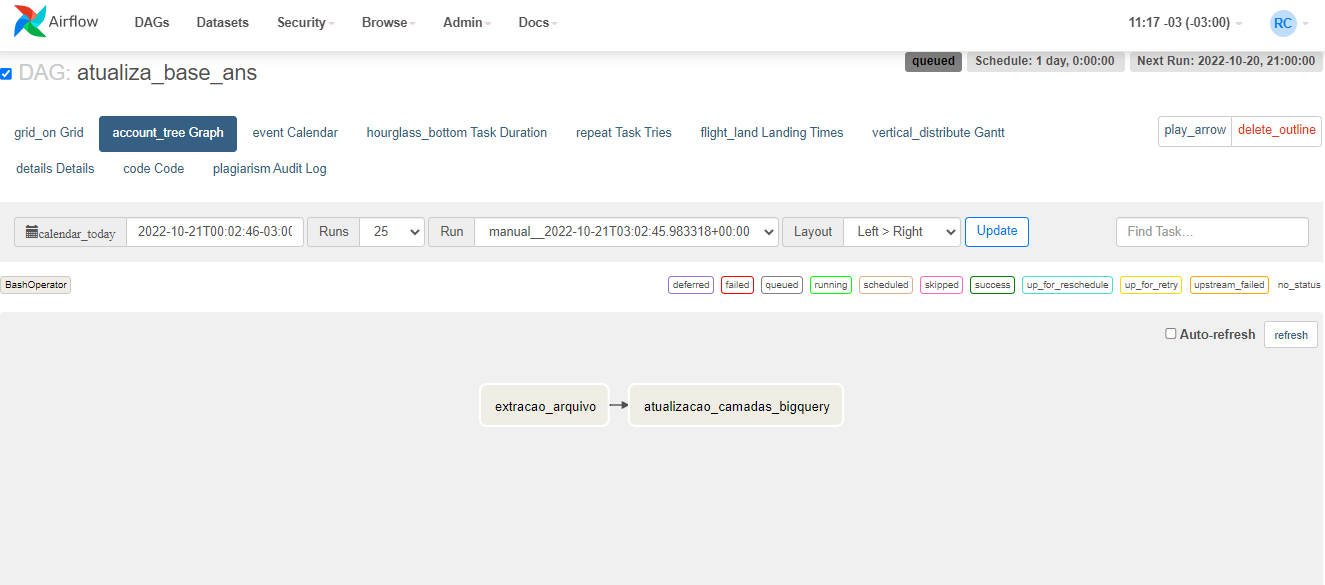


Figura 57: Dags Webserver Airflow

As figuras 58,59 e o link (<https://github.com/rrnc010/projeto-aplicado_mba>), evidencia o resultado da disponibilização dos scripts no Github conforme o proposto no planejamento do backlog.

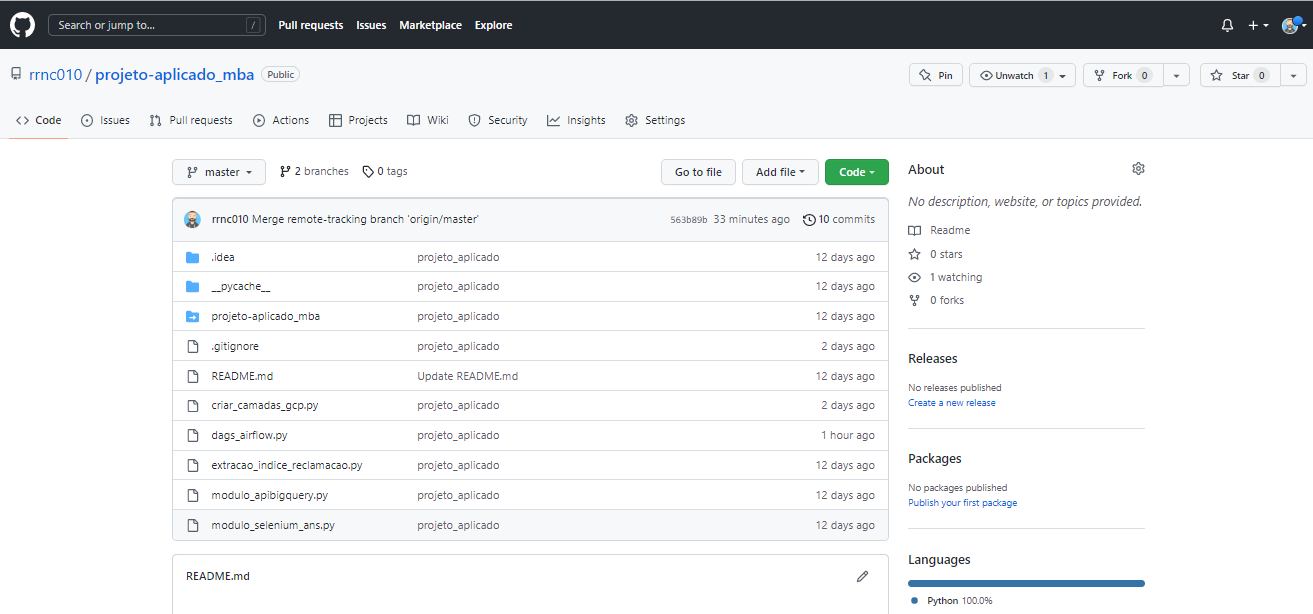


Figura 58: Parte 1 do github do projeto



Figura 59: Parte 2 do github do projeto

### 

### 2.3.2 Experiências vivenciadas

O desenvolvimento dessa etapa da sprint 3 foi bem interessante na questão dos contornos diante dos problemas que surgiram ao longo da sprint, tais como a configuração do Airflow no ambiente local, uma vez que foi em todo projeto um ambiente Windows, e no momento da configuração do Airflow de forma mais intuitiva e rápida seria necessário utilizar o Linux, o contorno desse problema foi adaptar o Linux no Windows para utilizar no projeto conforme mencionado na figura 60.



Figura 60: Utilização do Ubuntu no projeto

Para concluir essa sprint, a figura 61 ilustra o fluxograma do projeto destacando o que a sprint 3 representou no contexto geral da entrega

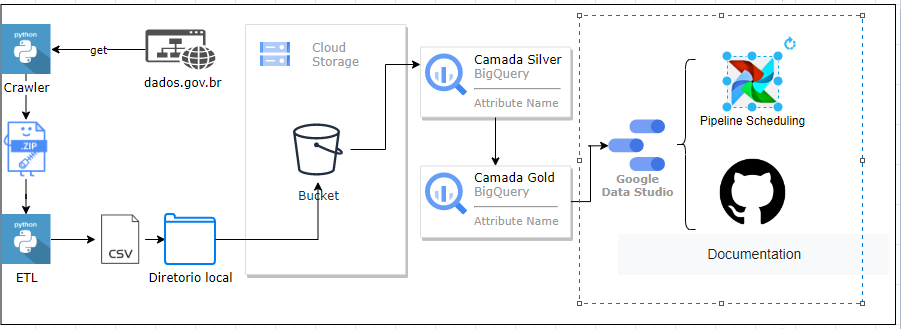


Figura 61: Fluxograma do projeto após sprint 3

# 3. Considerações Finais

A etapa da consideração final tem como demonstrar resultado do produto que foi proposto nesse projeto, contendo suas contribuições e a perspectiva das próximas etapas ao longo do amadurecimento do produto.

## 3.1 Resultados

Principais resultados alcançados:

* Organização dos dados em ambiente de Nuvem;
* Automatização de todo o processo;
* Performance, escalabilidade e disponibilidade dos dados;
* Tornar a informação acessível para as áreas de negócios.

A Organização dos dados em ambiente de Nuvem proporcionou um maior controle da base de dados, contendo um histórico desde 2015 com um processo de atualização incremental sendo armazenado em camadas no Bigquery.

Automatização de todo o processo colaborou para o projeto a garantia de entregar um produto para o cliente de forma que o mesmo vai apenas consumir as informações sem se preocupar em fazer tarefas manuais e desgastante onde pouparia tempo e teria os esforços melhor aproveitados em outras tarefas e ficando somente o papel de analisar os dados.

Performance, escalabilidade e disponibilidade dos dados garantiram ao projeto tornar um produto solido e útil através do uso das ferramentas e boas praticas que foi aplicado ao longo do projeto aplicado.

Tornar a informação acessível para as áreas de negócios foi um desafio desde o inicio para desenhar um produto final que tornaria útil para todos os envolvidos no contexto dessa base de dados, e nesse sentido agregou muito valor pois através de uma tabela no Bigquery e de um Dataviz no Data Studio foi possível proporcionar aos usuários diversas formas para o mesmo manipular esses dados para tirar um maior aproveitamento possível da base.

Os pontos a destacar que foi considerado positivo e negativo, foi o andamento e o corpo que se foi criando o projeto ao longo do desenvolvimento, por exemplo a questão de criar um padrão no layout da base depois que a mesma e extraída, facilitou a ingestão das camadas do Bigquery, outra dificuldade encontrada nas sprints de desenvolvimento foi a configuração de algumas ferramentas sendo necessário adaptações no sistema operacional. Ao todo a experiência vivenciada do projeto contribuiu para um maior aprendizado, aplicando novas metodologias de projeto e técnicas de pipeline de dados que foi possível conectar lições aprendidas com experiência de outros projetos e cursos para colocar em pratica todo o desenvolvimento do projeto aplicado.

## 3.2 Contribuições

Em todo o projeto teve o foco em conseguir tornar uma entrega com valor, dessa forma era necessária ser algo que seria útil e mudaria a rotina de trabalho de uma determinada área para melhor, dessa forma foi atacado as dores das áreas de negocio de modo que foi mapeado a dificuldade de controle de informações sobre as reclamações de plano de saúde registrado no site da Ans.

Portanto como contribuição que este projeto trouxe para esse contexto, foi de tornar acessível a base de dados para as áreas de forma totalmente automática com a garantia de tornar esse processo solido. Dessa forma os usuários vão conseguir acessar os dados atualizados frequentemente e conseguir acompanhar a evolução das reclamações de seus clientes e também podendo agir preventivamente. Com isso a vantagem foi tornar as pessoas mais analíticas e menos executoras de atividades manuais que tornou os times mais maduros diante dos resultados que o projeto proporcionou através das inovações como a usabilidade de ferramentas em cloud e com toda a infraestrura de código conectando as melhores práticas. A figura 62 demonstra o contexto final dessa contribuição que o projeto proporcionou

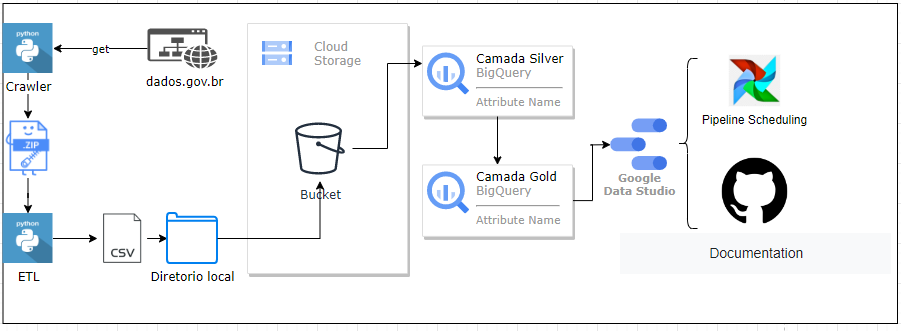


Figura 62: Fluxograma do projeto

## 3.3 Próximos passos

Seguindo com o amadurecimento do produto, e necessário a continuação da usabilidade de ferramentas com melhores práticas de pipeline para manter e melhorar o fluxo de atualização da base de dados.

Dessa forma para atender esse contexto, será necessário rodar esse pipeline em um ambiente de container como Docker, para poder garantir uma maior performance e também um respaldo em caso de necessidade de escalonar os dados dessa base como um maior volume com a computação distribuída.