

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS**NÚCLEO DE EDUCAÇÃO A DISTÂNCIA

Pós-graduação *Lato Sensu* em Analytics e Business Intelligence

**RELATÓRIO TÉCNCO**

COLETA DE DADOS ABERTOS DA VACINAÇÃO DE COVID-19 E PAINEL DE COBERTURA VACINAL DO ESTADO DE SERGIPE

Marcos Rocha Tavares

Belo Horizonte

2023

**SUMÁRIO**

[1. Introdução 4](#_Toc74691365)

[1.1. Contexto 4](#_Toc74691366)

[1.2. Objetivos 5](#_Toc74691367)

[1.3. Público alvo 5](#_Toc74691368)

[2. Modelo de Dados 6](#_Toc74691369)

[2.1. Modelo Dimensional 6](#_Toc74691370)

[2.2. Fatos e Dimensões 7](#_Toc74691371)

[3. Integração, Tratamento e Carga de Dados 11](#_Toc74691372)

[3.1. Fontes de Dados 11](#_Toc74691373)

[3.2. Processos de Integração e Carga (ETL) 12](#_Toc74691374)

[4. Camada de Apresentação 20](#_Toc74691375)

[4.1 Dashboard 20](#_Toc74691376)

[4.2 Análises avançadas 20](#_Toc74691377)

[5. Registros de Homologação 21](#_Toc74691378)

[5. Conclusões 21](#_Toc74691379)

[6. Links 21](#_Toc74691380)

[REFERÊNCIAS 22](#_Toc74691381)

# 1. Introdução

## 1.1. Contexto

Desde 11 de Março de 2020 a SARS-CoV-2 a síndrome respiratória aguda grave, conhecida pela sigla SARS, que há alguns anos começou na China e se espalhou para países da Ásia, também é causada por um coronavírus que infecta humanos provocando a doença conhecida como Covid-19, e que surpreendeu cientistas do mundo todo devido seu alto nível de contágio, provocando um número alarmante de internações impactando o serviços de saúde mundial e mesmo tendo baixo nível de letalidade provocou milhões de mortes, além do surgimento de variantes com maior escape imunológico e ao aumento da transmissibilidade.

Segundo Ricardo et al (2020, p. 2): Este cenário complexo impõe desafios adicionais à vigilância epidemiológica, às relações internacionais e à programação de políticas públicas, sobretudo por meio de medidas que reduzam as desigualdades de acesso aos sistemas de saúde e a condições estruturais para o autocuidado. Atentar para o comportamento desta pandemia nas distintas regiões parece ser imprescindível para a atualização das estratégias de enfrentamento desta emergência global e suas repercussões no nível local.

Fernandes (2022, v.1, p. 19-20) Os impactos da pandemia são profundos, tanto no que se refere ao mais importante, as vidas perdidas, quanto no que tange às consequências econômicas e sociais geradas pelos procedimentos preventivos de distanciamento físico e isolamento social necessários para diminuir a velocidade de contágio do vírus.

Portanto é necessário levantamento de dados e análises que auxiliem na tomada de descrições para melhor reação a pandemia estabelecida

Conforme Coelho (2020, v. 9, p.186) A utilização de dados tem sido cada vez mais aplicada às estratégias no âmbito da epidemiologia, visto que o manejo de big data se tem mostrado efetivo em colaborar com estudos e previsões para o enfrentamento adequado de futuras situações de saúde.

Frente a este cenário está aplicação propõe a carga de dados automatizada dos arquivos delimitados CSV do Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19, Campanha Nacional de Vacinação contra Covid-19, SRAG 2021 a 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19 e Notificações de Síndrome Gripal 2020 a 2022, do Opendatasus.

Os dados populacionais foram extraídos da base de dados populacional do IBGE

Apresentar em uma aplicação do histórico vacinal da Covid-19 através de histogramas em diversas dimensões para auxiliar a análise estratégica com gráficos, tabelas com volume agregados do número de vacinados com indicadores que demonstrem a evolução das ações de vacinação, além de realizar projeções para nortear as futuras ações dos municípios e empresas.

## 1.2. Objetivos

Com objetivo de auxiliar na tomada de decisão para melhoria das políticas públicas municipais, no controle das infecções e nas ações de combate à disseminação do vírus e as empresas na retomada de suas atividades presencias e remanejamento de pessoal.

Oferecendo uma aplicação que pode ser utilizada pelas secretarias de saúde municipais e pelas mais diversas empresas do comércio e indústria e irá auxiliar no acompanhamento e evolução da pandemia da Covid-19 bem como nas ações estratégicas e proativas para orientar o planejamento estratégico de vacinação, restrição e retomada das atividades no estado de Sergipe podendo ser utilizado com dados de outros estados ou nacionalmente.

## 1.3. Público alvo

O público a que se destina o acesso a aplicação são prefeitos e secretários municipais que tem por competência prestar apoio e assistência direta e imediata ao chefe do Poder Executivo.

Aos secretários de saúde municipais que gerenciam o Sistema Único de Saúde das cidades e que são responsáveis por supervisionar e executar políticas de saúde pública, tais como atividades médicas, paramédicas, odontológicas e fornecimento gratuito de medicamentos básicos para a população. Também coordenar os serviços das vigilâncias sanitária e epidemiológica, bem como realizar pesquisas médicas-sanitárias e executar outras atividades de sua competência.

A secretários de assistência social que tem entre suas atribuições estimular e apoiar o desenvolvimento comunitário e ações socioeducativas.

A secretários municipais da Indústria, comércio e turismo que tem entre suas atribuições compreender ações de incremento e estímulo à indústria e ao comércio, viabilizar o desenvolvimento industrial e comercial, e respectivos incentivos visto que o comercio e indústria também foram impactados com a evolução da pandemia.

Ao demais funcionários públicos comissionados ou concursados que atuam como Analistas de MIS (Manangeament Information Sistem), Analistas de sistemas BI (Business Intelligence), Analistas de dados, Analista de Tecnologia da Informação responsáveis pelas análises e distribuição de relatórios as diversas áreas do setor público municipal, do comércio e indústria auxiliando as secretárias nas ações estratégicas e na tomada de decisão.

# 2. Modelo de Dados

## 2.1. Modelo Dimensional

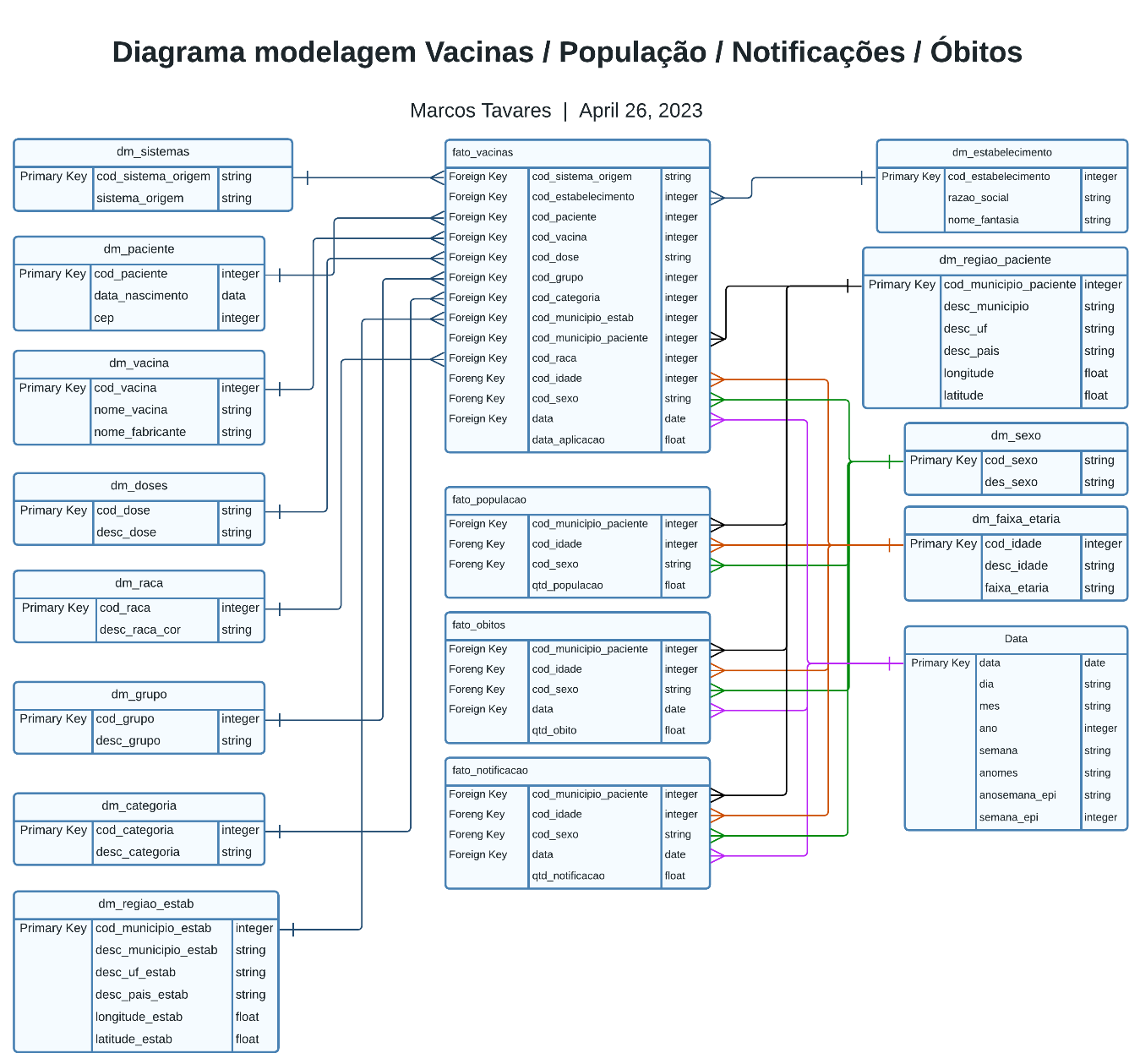
O diagrama do modelo dimensional das vacinas aplicadas nos pacientes do estado de Sergipe apresentado na figura 1 descreve como serão realizados os relacionamentos entre a fato e as onze dimensões.

As dimensões e fatos serão transformadas para seguir modelo multidimensional no esquema multi-estrela.

Ballard (2012), explica que existem três modelos básicos multidimensionais:

1. Esquema Estrela – Dimensões formadas por apenas uma tabela não normalizada, e existe uma quantidade menor de tabelas, mas em contrapartida pode existir redundância dos dados;
2. Esquema Floco de Neve – Nesse esquema, as tabelas são normalizadas evitando a redundância dos dados. nesse esquema os dados estão distribuídos várias tabelas;
3. Esquema Multi-estrela – No esquema as tabelas dimensão, estão não normalizadas, porém existem várias tabelas fato, unidas através das dimensões.

**Figura 1** – Apresentação do diagrama de modelagem das vacinas aplicadas, notificações, população e óbitos



Fonte: Própria (2023).

## 2.2. Fatos e Dimensões

## Fato vacinas

Conforme figura 2 a representação dos campos com volume de vacinas aplicadas, a partir da contagem distinta do campo “document\_id” não nulo, visto que não pode haver duplicidades neste número mesmo que haja duplicidade de registros.

**Figura 2** – Detalhamento dos dados da tabela fato vacinas



Fonte: Própria (2023).

Na estrutura da tabela fato existem chaves externas que serão relacionadas com as demais dimensões a fim de gerar um diagrama estrela visto na figura 1.

Os campos “sistema\_origem” e “dose” foram criptografados, eliminando quaisquer espaços ou acentos nas chaves, conforme demonstrado na figura 3.

**Figura 3** – Detalhamento dos dados da tabela fato\_vacinas

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Fonte: Própria (2023).

## Fato População

A tabela fato\_populacao contém a contagem de pessoas residentes no estado de Sergipe, segregada por sexo, idade, raça, situação do domicílio e município.

Porém foi definido como campos chaves para o relacionamento apenas os campos, sexo, idade, código do município. Ocorre que os dados de raça não são coincidentes nas demais fontes de dados, o dado de situação domiciliar não será utilizado nas análises. Por se tratar de dados populacionais do ano de 2010, foi necessário somar 9 anos a idade da população para considerar o envelhecimento até o ano de 2020 porém essa modificação não e a ideal visto que o seria necessário realizar a projeção populacional baseada na progressão geométrica, taxa de mortalidade, taxa de natalidade por sexo, taxa de migração e envelhecimento.

Na figura 4 está representado o volume populacional, a partir da soma do campo “Qtd”, originado da tabela de contagem populacional, encontrado no site do IBGE onde foram filtrados os dados do estado de Sergipe.

**Figura 4** – Detalhamento dos dados da tabela fato população



Fonte: Própria (2023).

As chaves externas que serão relacionadas com as demais dimensões serão geradas conforme a figura 5.

**Figura 5** – Detalhamento dos dados da tabela fato\_populacao



Fonte: Própria (2023).

## Fato notificação

A tabela fato\_notificacao contém os registros de notificação de síndrome aguda grave e detalha os dados com base no preenchimento de formulário padrão pelos agentes de saúde e pacientes conforme informado.

“Esta página tem como finalidade disponibilizar o legado dos bancos de dados (BD) epidemiológicos de SRAG, da rede de vigilância da Influenza e outros vírus respiratórios, desde o início da sua implantação (2009) até os dias atuais (2023), com a incorporação da vigilância da covid-19. Atualmente, o sistema oficial para o registro dos casos e óbitos por SRAG é o Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe).

O MS publicou o Guia de Vigilância Epidemiológica Emergência de Saúde Pública de Importância Nacional pela Doença pelo Coronavírus 2019 aonde estão disponíveis informações sobre definições de casos, critérios de confirmação e encerramento dos casos, dentre outros.

Ressaltamos que os dados da vigilância de SRAG no Brasil disponibilizados nesta página, estão sujeitos a alterações decorrentes da investigação, ou mesmo correções de erros de digitação, pelas equipes de vigilância epidemiológica que desenvolvem o serviço nas três esferas de gestão.”( SRAG 2021 a 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19, 2023)

Portanto algumas informações estão incompletas ou não foram informadas as secretarias municipais e estaduais de Sergipe, portanto foram selecionados apenas os campos chave com dados sobre o sexo, idade, município do paciente e data, a data escolhida foi a data da notificação, demais campos foram descartados para o desenvolvimento desse projeto.

Na figura 6 está representado o volume de notificações, a partir da contagem de registros, não foi encontrado na fonte de dados um campo identificador de notificação.

**Figura 6** – Detalhamento dos dados da tabela fato notificação



Fonte: Própria (2023).

As chaves externas que serão relacionadas com as demais dimensões serão geradas conforme a figura 7.

**Figura 7** – Detalhamento dos dados da tabela fato\_notificacao



Fonte: Própria (2023).

## Fato óbito

A tabela fato\_obito contém os registros de notificação de síndrome gripal e casos suspeitos de Covid-19, porém nesta fonte de dados o volume de óbitos registrados apresenta maior fidelidade com os dados acompanhados em boletins oficiais do estado de sergipe registrados.

“Descrição da informação disponibilizada: Os dados são oriundos do sistema e-SUS NOTIFICA, que foi desenvolvido para registro de casos de Síndrome Gripal suspeitos de Covid-19, e contém dados referentes ao local de residência do paciente (campos: estado, município), independentemente de terem sido notificados em outro estado ou município (Campos: estadoNotificação, municípioNotificação), além de demográficos e clínicos epidemiológicos dos casos.

Limitações dos dados: Os dados disponíveis não apresentam informações de estados e municípios que utilizam sistemas próprios de notificação de casos suspeitos de Covid-19 e, portanto, os dados para esses locais podem apresentar informações distintas.”( Notificações de Síndrome Gripal – 2022, 2023)

Algumas informações eram conflitantes com as demais fontes e não compreendia o mesmo número de casos identificados na fonte de notificações de síndrome gripal aguda grave.

Como foi verificado que o volume de óbitos encontrado na fonte de dados apresentou maior proximidade com os dados já publicados oficialmente, foi realizado um filtro da base da dados no campo evolução quando o valor for igual a 2, representando conforme dicionário de dados (Dicionário de Dados, Notificação de Síndrome Gripal – 2022, 2023)

E selecionados apenas os campos chave com dados sobre o sexo, idade, município do paciente e data, a data escolhida foi a data da notificação, demais campos foram descartados para o desenvolvimento desse projeto.

Na figura 8 está representado o volume de notificações, a partir da contagem de registros, não foi encontrado na fonte de dados um campo identificador de notificação.

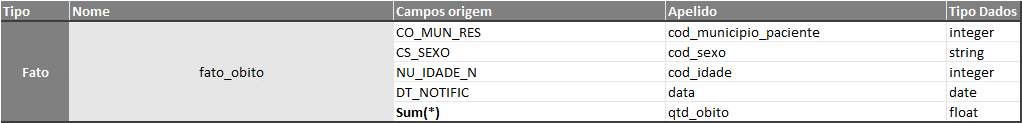
**Figura 8** – Detalhamento dos dados da tabela fato óbito



Fonte: Própria (2023).

As chaves externas que serão relacionadas com as demais dimensões serão geradas conforme a figura 9.

**Figura 9** – Detalhamento dos dados da tabela fato\_obito



Fonte: Própria (2023).

## Dimensões

As dimensões irão auxiliar na segregação dos dados e nas análises a serem realizadas pela aplicação, a partir dos dados e como se relacionam conforme figura 4.

**Figura 10** – Detalhamento dos dados nas dimensões

Tabela

Descrição gerada automaticamente

Fonte: Própria (2023).

A dimensão Estabelecimento descreve a razão social e nome fantasia do estabelecimento onde foi aplicada a dose da vacina.

**Figura 11** – Detalhamento da estrutura dimensão estabelecimento

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Fonte: Própria (2023).

Dimensão Paciente determina características do indivíduo como idade, sexo, nacionalidade, idade e cep tais dados serão mantidos nesta dimensão pois não possuem código para relacionamento com a tabela fato.

**Figura 12** – Detalhamento da estrutura dimensão paciente

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Fonte: Própria (2023).

A dimensão Vacina possuem entre os atributos nome da vacina e fabricante que apesar de haver um campo de referência para fabricante na base de dados de origem tal informação não agrega valor a análise sendo removida pois nome fabricante tem uma relação de unicidade com o código da vacina.

**Figura 13** – Detalhamento da estrutura dimensão vacina

Interface gráfica do usuário, Texto, Aplicativo, Tabela, Email

Descrição gerada automaticamente

Fonte: Própria (2023).

Foi definida uma dimensão de Dose das vacinas para impedir a duplicidade de dados na dimensão paciente e vacina, e para isso será gerado um código criptografado em hash256 a partir da descrição da dose.

**Figura 14** – Detalhamento da estrutura dimensão dose

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Fonte: Própria (2023).

A dimensão de Grupo descreve os grupos de atendimento para auxiliar a análise e nas medidas estratégicas.

**Figura 15** – Detalhamento da estrutura dimensão grupo

Interface gráfica do usuário, Texto, Email

Descrição gerada automaticamente

Fonte: Própria (2023).

Por sua vez a dimensão Categoria define em quais categorias estão os grupos de atendimento, e colabora nas análises estratégicas.

**Figura 16** – Detalhamento da estrutura dimensão categoria

Interface gráfica do usuário, Texto

Descrição gerada automaticamente

Fonte: Própria (2023).

A dimensão Raça segrega as vacinas aplicadas por raça/cor, para identificação do público de pacientes vacinados.

**Figura 17** – Detalhamento da estrutura dimensão raca

Interface gráfica do usuário, Texto

Descrição gerada automaticamente

Fonte: Própria (2023).

Foi definida duas dimensões Região Estabelecimento e Região Paciente que descrevem as regiões onde foram aplicadas as doses de vacina e dos pacientes vacinados.

**Figura 18** – Detalhamento da estrutura dimensão região do estabelecimento

Interface gráfica do usuário, Texto

Descrição gerada automaticamente

Fonte: Própria (2023).

**Figura 19** – Detalhamento da estrutura dimensão região do paciente

Texto

Descrição gerada automaticamente

Fonte: Própria (2023).

Foi definida a dimensão sexo para segregar os dados por sexo dos pacientes, e comparação com dados populacionais de óbito e notificações.

**Figura 20** – Detalhamento da estrutura dimensão sexo do paciente

Interface gráfica do usuário

Descrição gerada automaticamente

Fonte: Própria (2023).

Também com o mesmo propósito da dimensão de sexo a dimensão faixa etária segregará os dados por idade dos pacientes, porém aqui existe um objetivo mais estratégico, visto que é necessário observar os pacientes a depender de sua faixa etária.

**Figura 21** – Detalhamento da estrutura dimensão faixa etária do paciente

Tabela

Descrição gerada automaticamente

Fonte: Própria (2023).

Para desenvolvimento de análises com histogramas ou montagem de dados evolutivos e filtragem e seleção de períodos, a dimensão de Data a partir da data de aplicação da dose de vacina segrega por dia, mês, ano e semana os dados.

Realizada a composição do número da semana epidemiológica que conforme a convenção internacional as semanas epidemiológicas começam no domingo, indo até o sábado. Sendo a primeira semana epidemiológica do ano, por definição, termina no primeiro sábado de janeiro, devendo cobrir pelo menos quatro dias no mês, mesmo que comece em dezembro do ano anterior.

“Epidemiological weeks start on a Sunday and end on a Saturday; The first epidemiological week of the year ends, by definition, on the first Saturday of January, as long as it falls at least four days into the month, even if it means that this first week starts in December.” (PAHO, 2016, p. 1)

**Figura 22** – Detalhamento da estrutura dimensão data

Texto

Descrição gerada automaticamente

Fonte: Própria (2023).

Por fim a última dimensão Sistemas exige a geração da chave com base na descrição do sistema origem, para isso foi utilizado a função de criptografia do dado em hash 256 para gerar uma chave única, a descrição apresenta em qual sistema público foi aplicada a vacina.

**Figura 23** – Detalhamento da estrutura dimensão sistemas

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Fonte: Própria (2023).

# 3. Integração, Tratamento e Carga de Dados

## 3.1. Fontes de Dados

Dados são disponibilizados no portal do openDatasus na versão de arquivos delimitados .CSV “*comma-separated-values”* (valores separados por vírgulas) ou podem ser extraídos utilizando api em formato de arquivos semiestruturados JSON “*JavaScript Object Notation*”.

Para esta aplicação utilizaremos a versão do arquivo em delimitado por virgulas, disponível em: <https://opendatasus.saude.gov.br/dataset/covid-19-vacinacao>, arquivos do estado de Sergipe.

Os dados de óbitos serão extraídos do Opendatasus a partir da página de notificações de síndrome gripal, em arquivo delimitado por virgulas, disponível em: <https://s3.sa-east-1.amazonaws.com/ckan.saude.gov.br/SGL/2022/uf=SE/lote=1/part-00000-17377c38-e8b1-49d2-b8ad-0b83cf93227c.c000.csv>

As notificações de síndrome gripal aguda grave são e derivadas de 4 links para cada período de dados a serem coletados dos anos de 2020 disponível em: <https://s3.sa-east-1.amazonaws.com/ckan.saude.gov.br/SRAG/2020/INFLUD20-24-04-2023.csv>, 2021 disponível em: <https://s3.sa-east-1.amazonaws.com/ckan.saude.gov.br/SRAG/2021/INFLUD21-24-04-2023.csv>, 2022 disponível em: <https://s3.sa-east-1.amazonaws.com/ckan.saude.gov.br/SRAG/2022/INFLUD22-03-04-2023.csv>, 2023 disponível em: <https://s3.sa-east-1.amazonaws.com/ckan.saude.gov.br/SRAG/2023/INFLUD23-03-04-2023.csv>.

Dados populacionais serão extraídos a partir (Sistema IBGE de Recuperação Automática – SIDRA, Tabela de dados agregados: 3175 - População residente, por cor ou raça, segundo a situação do domicílio, o sexo e a idade, 2023), os dados foram baixados do site após selecionar todas os campos desejados e filtrar os municípios respectivos da unidade federativa.

Foi gerado um link temporário para realizar o download do arquivo <https://sidra.ibge.gov.br/geratabela/DownloadSelecaoComplexa/519096616> após realizar selecionar o download na página, conforme figura 24.

**Figura 24** – Download da Tabela 3175, do site Sistema IBGE de Recuperação Automática

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Fonte: Própria (2023).

## 3.2. Processos de Integração e Carga (ETL)

Conforme explicado por Ribeiro Homem, (2021 p.26) O processo de ETL é fundamental nas práticas implementadas pelo BI, visto que para realizar análises de dados de negócio é necessário dispor de informação no formato adequado. Dessa forma, é necessário diagnosticar e compreender o problema tratado para identificar qual a transformação de dados mais indicada, utilizando o sistema de ETL para gerar apenas informações pertinentes e completas.

Para realizar o processo de extração de dados, os arquivos em formato de texto separado por virgulas, serão extraídos do site do opendatasus através dos link (https://s3.sa-east-1.amazonaws.com/ckan.saude.gov.br/SIPNI/COVID/uf/uf%3DSE/) e transferido pela ferramenta KNIME para um “*bucket*”(contêiner de objetos) é como são chamados os repositórios de arquivos no S3 que são mapeados por região.

Explica Francisco (2021 p.31) No KNIME, é utilizado de forma intuitiva no qual é possível criar um workflow visual, pois conta com uma interface gráfica no estilo arrastar e soltar, sem a necessidade de codificação. Porém, caso o projeto necessite de alguma implementação extra, é possível a integração com as principais linguagens de programação utilizadas para ciência de dados, incluindo integrações com outras plataformas de Business Intelligence como Power BI, Tableau, Weka e outras.

Na figura 16 ilustra como foi realizado usando o Knime, um “Workflow”(fluxo de trabalho) com técnicas de “Webscraping”(raspagem de dados) no nó de XPath e posteriormente selecionar, filtrar e agrupar apenas as urls correspondentes as partes dos arquivos e dos estados, também foi realizado uma junção, com arquivo auxiliar contendo os estados e regiões a fim de elaborar a filtragem por região caso necessário apresentado na figura 25 e o resultado na figura 26.

**Figura 25** – Workflow no Knime para captura da URL para upload

Tabela

Descrição gerada automaticamente

Fonte: Própria (2023).

**Figura 26** – Url segregadas por parte, estado e região

Interface gráfica do usuário, Aplicativo, Word

Descrição gerada automaticamente

Fonte: Própria (2023).

Como apresentado na figura 27 temos o workflow que permite a filtragem das Url’s, em seguida a conversão para realizar o upload destes arquivos para S3 da Amazon após a conexão por chave de segurança na “Amazon Web Services” ( Serviços de Internet da Amazon),(AWS). E na figura 28 o workflow completo de todas as extrações.

**Figura 27** – Workflow no Knime para filtragem e upload ao S3

Diagrama

Descrição gerada automaticamente com confiança média

Fonte: Própria (2023).

**Figura 28** – Workflow no Knime completo.

Uma imagem contendo Diagrama

Descrição gerada automaticamente

Fonte: Própria (2023).

De acordo com Amazon Web Services, (2022) O Amazon S3 é um serviço de armazenamento de objetos que armazena dados como objetos em buckets. Um objeto é um arquivo e quaisquer metadados que descrevam o arquivo. Um bucket é um contêiner de objetos.

Os arquivos de texto armazenados no bucket do S3 demonstrado na figura 28 serão transformados usando o AWS Glue.

**Figura 28** – Arquivo .csv armazenado no bucket do AWS S3

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Fonte: Própria (2023).

Também é informado em Amazon Web Services, (2022) O AWS Glue é um serviço de ETL (extração, transformação e carregamento) totalmente gerenciado que torna mais fácil e econômico o processo de categorizar dados, limpá-los, aprimorá-los e movê-los de modo confiável entre vários armazenamentos e streams de dados. O AWS Glue consiste em um repositório de metadados central, conhecido como AWS Glue Data Catalog, um mecanismo de ETL que gera automaticamente um código Python ou Scala e um programador flexível que lida com resolução de dependências, monitoramento de trabalhos e novas tentativas.

Será Realizada a edição, depuração e teste do código de ETL Python com Apache Spark usando o ambiente de desenvolvimento com recursos auxiliares com Pandas e assim popular o S3 com as tabelas fatos e dimensões em arquivos formato parquet, que poderão ser consultados através do banco de dados Athena usando instruções em SQL e disponibilizará a informação para a aplicação.

Explicando PYTHON Ribeiro Homem, (2021) Python é uma das linguagens de programação mais utilizadas da atualidade, ela é versátil e simples de aprender, sendo considerada como uma linguagem de alto nível, cuja sintaxe é voltada para o entendimento humano. Por ser uma linguagem moderna ela foi construída para ser objetiva, assim, desenvolvedores conseguem definir instruções com menos linhas de código.

Ribeiro Homem complementa ainda, (2021) Uma das bibliotecas mais conhecidas é o Pandas. Ela é utilizada para realizar operações relacionadas a análise de dados. Suas funções permitem realizar a leitura e manipulação de dados, entregando estruturas e operações para manipular tabelas numéricas e séries temporais.

Quanto ao Spark de acordo com Garcia Enzo, (2021) O Spark permite que aplicações em clusters Hadoop executem até 100 vezes mais rápido em memória e até 10 vezes mais rápido em disco, desenvolver rapidamente aplicações em Java, Scala ou Python. Além disso, vem com um conjunto integrado de mais de 80 operadores de alto nível e pode ser usado de forma interativa para consultar dados diretamente do console. Além das operações de Map/Reduce, suporta consultas SQL, streaming de dados, aprendizado de máquina e processamento de grafos.

Conforme apresentado na figura 29 onde é possível realizar a extração dos dados e o tratamento por campos, definindo seus tipos e ajustando para ser mais intuitivo para elaboração dos relatórios, o “job”(trabalho) job\_raw\_to\_structured\_vacinacao realiza a criação da base de dados estruturada que será armazenada no s3 no bucket: s3://opendatasus.puc.prestructured/vacinação, demostrado na figura 30, para posterior uso pelo “job” job\_opendatasus\_vacinacao que irá produzir a tabela fato.

Logo no primeiro procedimento carga\_vacina são criados os campos chaves inclusive os campos que foram transformados usando criptografia para garantir a unicidade dos dados.

**Figura 29** – Extração do arquivo csv e estruturação dos dados pelo “job” job\_raw\_to\_structured\_vacinacao.

Texto

Descrição gerada automaticamente

Fonte: Própria (2023).

**Figura 30** – Dados estruturados no ”bucket” do S3.

Interface gráfica do usuário

Descrição gerada automaticamente com confiança média

Fonte: Própria (2023).

Utilizando funções do pyspark para executar comando em sql para o tratamento dos dados, e construção do esquema utilizado na modelagem o procedimento fato\_vacinas monta a tabela fato agrupando a contagem do “document\_id” pelas chaves que compõem o esquema de padrão estrela.

Na figura 31 são apresentados mais alguns exemplos de como foram criadas as dimensões e os tratamentos em campos como datas, chaves.

**Figura 31** – Exemplo de construção das dimensões usando pyspark e sql no “job” job\_opendatasus\_vacinacao.

Texto

Descrição gerada automaticamente

Fonte: Própria (2023).

Finalizado a construção do “job” conforme visualizado na figura 32, onde cada dimensão foi tratada em um processo, e para remoção de quaisquer discrepâncias foram aplicados filtros com as seguintes condições:

* Restringir pacientes residentes no estado de Sergipe;
* Restringir Pacientes com registro de Sexo (Masculino, Feminino) definido;
* Restringir Idade entre 10 e 110 anos público-alvo das análises;
* Datas igual ou superior a 19/01/2021 data da aplicação da primeira dose de vacina no estado;
* Remover dados de doses não informados.

**Figura 32** – Código da Tarefa que realiza o ETL dos arquivos e disponibiliza para visualização no Athena

Texto

Descrição gerada automaticamente

Fonte: Própria (2023).

Agora podem ser executados conforme gatilho parametrizado em um agendamento, que será disparado por função Lambda ou Glue Workflow, demonstrado na figura 33.

**Figura 33** – Exemplo de construção de um workflow no Glue, Agendamento Opendatasus.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Fonte: Própria (2023).

As dimensõe e fato são salvas em formato de arquivo parquet particionados no S3 em um bucket especifico para organizar e separar os dados brutos e transacionais dos dados que já foram estruturados , como pode ser visualizado na figura 33, e estes arquivos serão lidos pelo crawler para carregar no banco de dados Athena como visualizado na figura 34.

**Figura 33** – Arquivos salvos em formato parquet particionado no bucket opendatasus.puc.structured

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Fonte: Própria (2023).

**Figura 34** – Execução do crawler em cima dos dados transformados para subir os dados para o banco de dados do Athena

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Fonte: Própria (2023).

No banco de dados Athena é possivel verificar o resultado das extrações, realizar testes e a disposição dos dados nas tabelas e seus relacionamentos com a fato, através do console do athena apresentado na figura 35.

**Figura 35** – Consulta dos dados no Athena

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Fonte: Própria (2023).

A cadeia de processos de ETL se encerra com a disponibilização dos dados no Athena que permite consultas usando instruções em SQL e podera ser disponibilizado para diversas plataformas de BI como Qlikview, Microstrategy, Power BI, Tableu e demais outras conforme figura 36.

**Figura 36** – Diagrama de execução do ETL até a visualização dos dados

Tela de celular com publicação numa rede social

Descrição gerada automaticamente com confiança média

Fonte: Própria (2023).

## Preparação dos dados

O Microstrategy

Para preparação de dados foi utilizado a ferramenta Microstrategy Workstation. O MicroStrategy Workstation é uma ferramenta de desenvolvimento e análise de dados fornecida pela MicroStrategy, uma empresa que desenvolve software de business Intelligence e análise de dados, disponível para download de forma gratuita pelo link https://mci-prod-installs.s3.amazonaws.com/workstation-win-ent.zip., porém é necessário acessar usando uma conta coorporativa válida.

Microstrategy é o novo aplicativo de visualização de primeira geração que leva o poder das ferramentas analíticas diretamente para a área de trabalho, tudo em uma velocidade maior. Obtenha a experiência de resposta instantânea com novos recursos exclusivos do aplicativo, como o download inteligente e a atualização automática, que preveem e baixam proativamente o conteúdo de que você precisa tanto on-line quanto off-line. Com tempo de carregamento limitado ou nulo, otimizamos os fluxos de trabalho para aprimorar os processos de tomada de decisões e impulsionar a eficiência operacional. (Microstrategy, MicroStrategy Help, 2023).

Por que usar o aplicativo Microstrategy?

* Explore dados e colete insights com exibições criativas de layout de forma livre;
* Visualize documentos perfeitos em pixels ao lado de dossiês em sua biblioteca de conteúdo;
* Experimente a resposta instantânea com os recursos de Download Inteligente e Atualização Automática que preveem e baixam proativamente conteúdo para on-line e off-line;
* Salve rapidamente insights e visualizações pessoais com marcadores;
* Compartilhar conteúdo com outras pessoas diretamente por meio do aplicativo;
* Habilite a colaboração dinâmica e o sistema de mensagens para centralizar a comunicação;
* Navegue pelo conteúdo e obtenha as atualizações de dados mais recentes no menu principal da área de trabalho a qualquer momento. (Microstrategy, Introduction to the MicroStrategy App, 2023).

Preparação dos dados com Miscrostrategy

O Microstrategy permite que sejam realizadas diversas conexões com as mais diversas fontes de dados, para atender a necessidade do projeto existe a conexão com a base de dados athena, entretanto foi necessário baixar e incluir na pasta da aplicação o Amazon Athena JDBC Driver, conforme descrito no site da Microstrategy.

Para se conectar ao MicroStrategy Desktop em um sistema operacional Windows, o driver JDBC deve ser baixado e salvo na seguinte pasta MicroStrategy JDBC:C:\Arquivos de Programas\MicroStrategy\MicroStrategy Desktop\JDBC. (Microstrategy, How to connect to Amazon Athena, 2023)

Realizada essa configuração ao acessar a ferramenta, iniciando a criação de um novo dossier, no menu Inserir, adicionar novos dados e será disponibilizado na tela de fonte de dados a opção para conectar os dados no Amazon Athena, selecionar a opção importar tabelas, porém também é possível realizar consultas e gerar fontes de dados filtradas, agregadas e com campos tratados pelo usuário, conforme apresentado na figura 37 e a configuração da conexão com a Amazon na figura 38.

**Figura 37** – Inclusão de dados a partir da fonte de dados, Amazon Athena  
Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Fonte: Própria (2023).

**Figura 38** – Configurar a conexão com a AWS.

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Fonte: Própria (2023).

Após realizar a conexão devem ser selecionadas todas as tabelas disponíveis na base de dados do Athena, conforme figura 39.

**Figura 39** – Importando as tabelas disponíveis no banco de dados no Amazon Athena.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Fonte: Própria (2023).

As tabelas foram preparadas na tela de preparação de dados, onde foram definidos os atributos, métricas, e relacionamentos entre as tabelas, que por padrão quando os nomes dos campos têm a mesma nomenclatura tanto na tabela fato quanto na tabela dimensão esse relacionamento é feito de forma automática, mas também é possível relacionar campos com nomes diferentes, nessa tela também permite que sejam realizadas transformações, conforme figura 40.

**Figura 40** – Preparação dos dados, no Microstrategy.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Fonte: Própria (2023).

Após essa etapa de preparação das tabelas, os atributos e métricas serão disponibilizados no dossier, a partir desse ponto já se é possível construir os relatórios, gráficos, novos atributos ou novas métricas conforme funções disponíveis na ferramenta, como visualizado na figura 41.

**Figura 40** – Conjunto de dados disponibilizados no Microstrategy, dossier.

Interface gráfica do usuário

Descrição gerada automaticamente

Fonte: Própria (2023).

# 4. Camada de Apresentação

## 4.1 Dashboard

Para permitir que os usuários realizem análises confrontando os dados disponibilizados pelas fontes de dados do Opendatasus de forma intuitiva dispensando o uso de planilhas ou outras ferramentas, e então possam identificar novos insights, e assim orientar as tomadas de decisão das secretarias de saúde ou das melhores práticas na cobertura vacinal em todo estado, foi elaborado um dashboard contemplando diversas visões, com indicadores que apontem para as oportunidades estratégicas.

Análise Operacional

Na visão Painel Cobertura Vacinal – Covid 19 é possível acompanhar os principais indicadores, que permite uma visualização rápida, mas contempla todo o cenário da vacinação e já evidencia o resultado das ações de controle e estratégias de aplicação das doses de vacinas no estado.

**Figura 41** – Visão, de indicadores consolidados no Painel de Cobertura Vacinal – Covid 19.

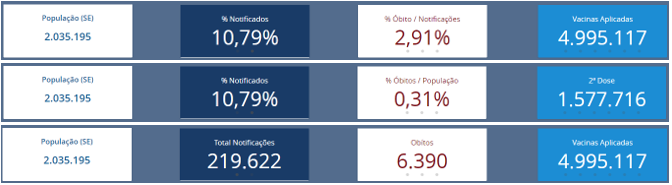
Interface gráfica do usuário, Site

Descrição gerada automaticamente

Fonte: Própria (2023).

Na parte superior da visão existem 4 indicadores que apresentam informações básicas como volume populacional fixo, quantidade de casos notificados, percentual de notificações frente ao volume populacional, quantidade de óbitos e percentual de óbitos frente ao volume populacional, o volume de vacinas aplicadas e volume por dose aplicada com efeito de carrossel permitindo assim a visualização destes indicadores sem interação do usuário, como mostrado na figura 42.

**Figura 42** – Principais indicadores, fixo ou com efeito carrossel.



Fonte: Própria (2023).

No centro da visão foi incluído um painel com 3 gráficos e um seletor de painéis para realizar essa transição, abaixo do painel existe um filtro temporal de ano e mês que permite ajustar a visão para um determinado momento no tempo.

O gráfico 1, cobertura vacinal apresenta no primeiro eixo horizontal o volume de vacinas aplicadas em barras e no segundo eixo horizontal o percentual de doses aplicadas frente o volume populacional do estado em linha, no eixo vertical a quebra por grupo de dose aplicado.

**Gráfico 1** – Cobertura vacinal por dose, cenário estadual.

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Fonte: Própria (2023).

No gráfico 2, com médias móveis de óbito e doses aplicadas, foram incluídas 5 médias moveis no período de 7 dias anteriores, quebra por data e classificação por semana epidemiológica no eixo horizontal, sendo uma delas a média móvel de óbitos no primeiro eixo horizontal e as demais no segundo eixo horizontal, no eixo vertical estão dispostos o ano e número da semana epidemiológica para identificação da evolução da vacinação e no impacto frente ao volume de óbitos.

**Gráfico 2** – No gráfico de evolução dos óbitos frente a aplicação das vacinas na população ao longo do tempo é um excelente indicador no auxílio da tomada de decisão.

Interface gráfica do usuário, Gráfico, Histograma

Descrição gerada automaticamente

Fonte: Própria (2023).

Incluído no gráfico 3, um histograma dos percentuais de médias moveis de notificações, óbitos e vacinação comparativo com o período anterior no eixo vertical e no eixo horizontal o mês e ano de acompanhamento.

O esse gráfico consegue informar que períodos em que houve um maior volume de aplicação das vacinas em quais quer doses aplicadas.

**Gráfico 3** – Acompanhamento de Notificações e óbitos, apresenta um histórico da evolução da.

Linha do tempo

Descrição gerada automaticamente

Fonte: Própria (2023).

Análise Estratégica

Conforme divulgado no portal gov.br o Ministério da Saúde e Fiocruz definiram estratégias para aumentar a cobertura vacinal.

O PNI estabelece que a cobertura vacinal adequada é de 80% para meningo e HPV; 90% para rotavírus, influenza e BCG, e 95% para as demais vacinas; e uma homogeneidade de no mínimo 70% entre os municípios. A ação estratégica da Pasta tem como meta que o aumento na cobertura vacinal seja homogêneo em todo o país até 2025. (gov.br, 2023)

Na parte superior da visão indicadores de cobertura vacinal, estão visíveis os percentuais totais de cobertura vacinal e o volume de vacinas aplicadas, também abrange a cobertura separando por sexo, como aparece na figura 44, com um “Threshold” nos percentuais onde de 0 a 70% é vermelho, 70,01% a 95% é amarelo e 95.01 a 100% é verde.

**Figura 44** – Sumario dos indicadores de cobertura vacinal total e por sexo.



Fonte: Própria (2023).

Não foi identificado um percentual específico para vacinas contra Sars-cov-2, porém é compreensível que o percentual aceitável de cobertura vacinal deve ser superior a 70% e 95% seria a cobertura ideal.

Com base nessa premissa foram criados velocímetros por faixa etária, dado que também foi evidenciado que as faixas etárias acima de 60 anos de idade são grupo de risco, e a distribuição e aplicação das vacinas também é por faixa etária, bem como a separação por sexo, servindo assim como indicadores para elaboração de táticas para o atingimento da cobertura vacinal ideal.

Como apresentado a figura 45, também foi criado níveis de coloração com base no percentual de cobertura vacinal e incluída a uma linha de referência em 95% como meta a ser alcançada. A coloração das bandas do velocímetro foi definida como sendo de 0 a 70% vermelho, de 70,01% a 80% laranja, 80,01% a 95% amarelo e 95,01% a 100% verde.

**Figura 45** – Coloração de bandas dos velocímetros de cobertura vacinal.

Interface gráfica do usuário

Descrição gerada automaticamente

Fonte: Própria (2023).

Foi inserido na visão o filtro por município e por dose de vacina, para auxiliar no detalhamento e acompanhamento

Como resultado temos no gráfico 4 a percepção se cobertura vacinal por dose aplicada está atendendo aos objetivos e estratégias elaboradas, bem como encontra qual público-alvo deve ser priorizado, e assim adotar medidas para realizar campanhas mais voltadas aquele público específico.

**Gráfico 4** – Distribuição dos indicadores de cobertura vacinal, em velocímetros e níveis de atingimento de cobertura.

Interface gráfica do usuário

Descrição gerada automaticamente

Fonte: Própria (2023).

Visões de Nível Tático

Foram criadas as visões Relatório AdHoc Cobertura Vacinal, Quadro Vacinal Estadual, Relatório Adhoc Vacinas e Cobertura Vacinal por Município.

É possível navegar a partir do menu lateral até o relatório Adhoc de cobertura vacinal, aqui o usuário conseguira ter a informação com um nível maior de detalhes, e dessa forma elaborar suas próprias estratégias e abordagens.

Nesse painel temos como um autosserviço com a possibilidade de o usuário escolher determinados atributos nos seletores de atributos e métricas, e ao selecioná-los serão inseridos na tabela logo abaixo conforme a disposição dos seletores de atributos e métricas permitir.

Na lateral é disponibilizado um painel com alguns filtros que podem ser utilizados para auxiliar na análise do usuário, como apresenta a figura 46.

**Figura 46** – Relatórios AdHoc para que o usuário possa mapear e filtrar.

Interface gráfica do usuário

Descrição gerada automaticamente

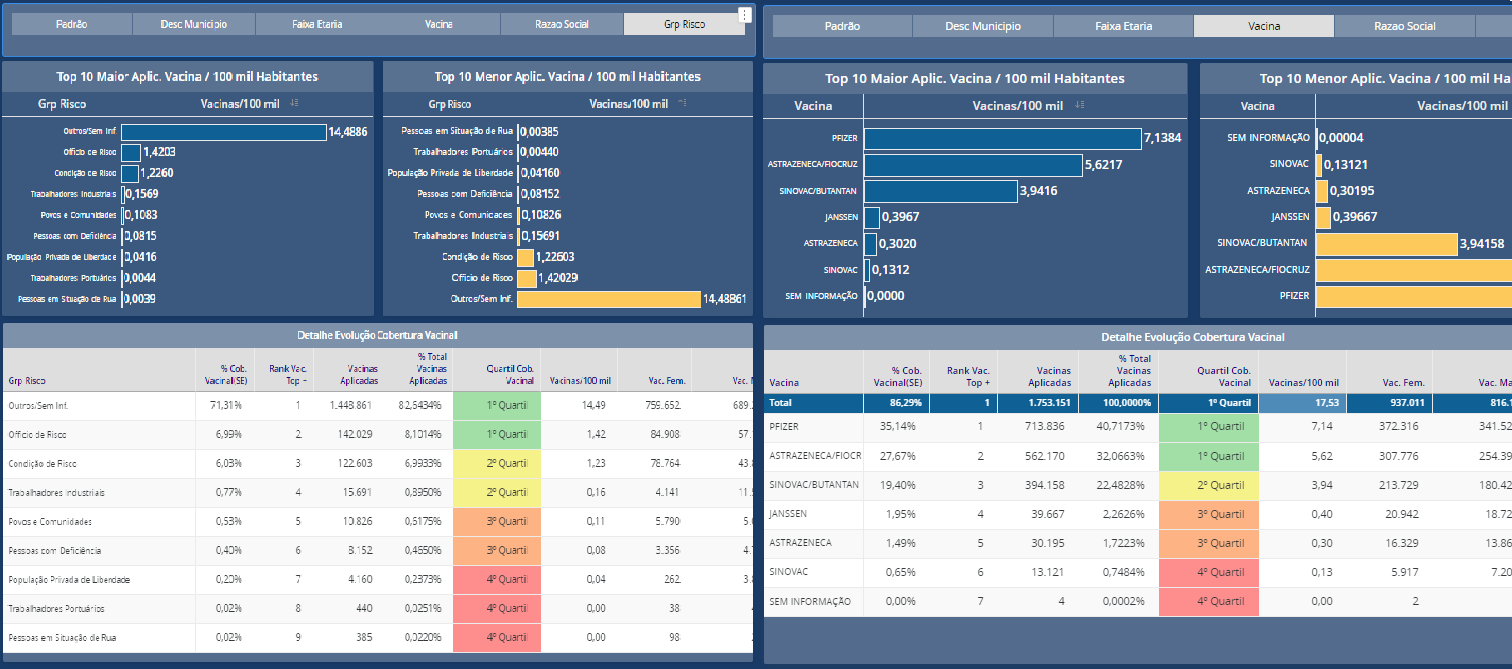
Fonte: Própria (2023).

No Quadro Vacinal Estadual são apresentados filtros a nível de município, doses aplicadas, faixa etária, vacinas distribuídas, razão social do estabelecimento que aplicou as doses. Foram inseridos 2 gráficos, ranqueados por Top 10 com maior aplicação de vacinas a cada 100 mil habitantes e Top 10 com menor aplicação de vacinas a cada 100 mil habitantes, e na parte inferior direita existe uma tabela com as métricas:

* % Cob. Vacinal (SE) – Percentual de Cobertura Vacinal frente a população do estado;
* Rank Vac. Top + - Rank de vacinas aplicadas decrescente;
* Vacinas Aplicadas – Qtd. de vacinas aplicadas;
* % Total Vacinas Aplicadas – Percentual de vacinas aplicadas pelo ao total de vacinas aplicadas;
* Quartil Cob. Vacinal – Quartil do percentil da % Cob. Vacinal, essa métrica possui “Threshold” quartilhado onde valores de 0 a 25% são vermelhos, valores de 25,01% a 50% são laranja, valores de 50,01% a 75% são amarelos e de 75,01% a 100% são verdes;
* Vacinas/100 mil – Vacinas aplicadas dividido por 100.000 habitantes;
* Vac. Fem. – Vacinas aplicadas no sexo feminino;
* Vac. Masc. – Vacinas aplicadas no sexo masculino.

O Atributo dos gráficos e do relatório pode ser alterado com base no seletor de atributos acima dos gráficos, entre Desc. Município que foi definido como padrão, Faixa etária, Vacina, Razão Social, Grp. Risco, conforme mostrado na figura 47.

**Figura 47** – Seletor de atributos da visão Quadro Vacinal Estadual.



Fonte: Própria (2023).

Esta visão nos apresenta dentre os atributos possíveis de serem selecionados quais os principais ofensores ou aqueles com maior desenvolvimento, bem como onde atuar de forma mais precisa a fim de conseguir atender de melhor forma a melhor cobertura, observando que o importante é seguir a tática de atingir os grupos de risco, mas com maior eficiência atuando no 2 e 3 quartil, como pontuado na figura 48.

**Figura 48** – Visão do Quadro Vacinal Estadual.

Interface gráfica do usuário, Site

Descrição gerada automaticamente

Fonte: Própria (2023).

Navegando a partir do menu lateral até o Relatório Adhoc Vacinas, aqui o usuário tem a possibilidade de complementar as análises com informações mais detalhadas.

Aqui o usuário tem a possibilidade de selecionar atributos e métricas, e adicioná-los na tabela abaixo conforme a disposição dos seletores de atributos e métricas permitir.

Na lateral estão disponíveis alguns filtros que podem ser utilizados para auxiliar na análise do usuário, como apresentado na figura 49.

**Figura 49** – Relatório Adhoc Vacinas, permite ao usuário uma maior amplitude nas análises.

Interface gráfica do usuário

Descrição gerada automaticamente

Fonte: Própria (2023).

A visão de cobertura vacinal por município torna-se uma opção a mais para complementar a analise, contém um filtro para doses aplicadas, possui 3 gráficos sendo um gráfico de anel com a cobertura vacinal por sexo, um gráfico de barras com os 10 municípios com a maior cobertura vacinal, logo abaixo um gráfico de barras com os 10 municípios com a menor cobertura vacinal, contém uma tabela com os indicadores de quartil de cobertura vacinal e percentual do total de vacinas aplicadas, em relação aos atributos faixa etária ou grupo de risco , conforme for selecionado através de um seletor de atributos acima.

Em destaque é possível identificar no Mapa do estado dimensionado pela quantidade de vacinas aplicadas cada município, em bolhas para identificação do volume de vacinas aplicas e com definição de limites com o padrão de cor definido como sendo de 0 a 70% vermelho, de 70,01% a 80% laranja, 80,01% a 95% amarelo e 95,01% a 100% verde, conforme pode ser visualizado na figura 50.

**Figura 50** – Cobertura Vacinal por Munícipio, apresenta um visual mais amigável para as análises e complementa os demais relatórios.

Interface gráfica do usuário

Descrição gerada automaticamente

Fonte: Própria (2023).

Ao selecionar quais quer um dos municípios do mapa, é ativado o filtro nos demais gráficos, permitindo um melhor acompanhamento para identificação dos indicadores por município, o mesmo ocorre ao selecionar algum dos municípios no gráfico Top 10 Maior Cob. Vacinal ou no gráfico Top 10 Menor Cob. Vacinal como aparece na figura 51.

**Figura 51** – Selecionando os municípios pelo gráfico de menor cobertura vacinal é possível filtrar os demais gráficos na visão.

Interface gráfica do usuário

Descrição gerada automaticamente

Fonte: Própria (2023).

# 5. Registros de Homologação

Para realização dos testes foi criado um dossiê a parte apenas para gerar as consultas sql dinamicamente nos dados presentes no banco de dados Amazon Athena, acessando os dados como demonstrado na figura 52.

**Figura 52** – Acesso a base de dados para geração de consultas em sql.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Fonte: Própria (2023).

Foi realizada a consulta para os principais indicadores, a fim de comprovar que os dados presentes nesses indicadores são os mesmos presentes no banco de dados. Para a consulta de validação dos dados das tabelas fato\_obitos, fato\_notificacao e fato\_populacao, realizada a consulta sumarizando os valores dos campos qtd\_obito, qtd\_notificacao e qtd\_populacao, e consecutivamente identificar os percentuais de notificação, realizando a divisão do total de notificações pelo total de população, os percentuais de óbitos por população realizando a divisão do total de óbitos pelo total de população e o percentual de óbitos por notificações realizando a divisão do total de óbitos pelo total de notificação. Como demonstrado na figura 53 todos os volumes realizados a consulta diretamente ao banco de dados estão em conformidade com o disponibilizado na ferramenta.

**Figura 53** – Consulta dos volumetria dos principais indicadores presentes no dashboard.

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Fonte: Própria (2023).

Como forma de verificação na figura 54 é apresentado o código SQL utilizado na consulta.

**Figura 54** – Consulta dos volumetria dos principais indicadores presentes no dashboard.



Fonte: Própria (2023).

Realizada outra consulta na base de dados para comprovar a volumetria das vacinas a partir da soma do campo qtd\_vacinas\_aplicadas bem como a consulta com base no grupo de doses aplicadas para demonstrar que os valores dos totais por grupo de dose aplicada, reflete o mesmo dado expressado nos dossier.

Para essa consulta foi aplicado o relacionamento com a dimensão de doses e utilizado a função case, para segregar os dados com base na composição de cada grupo de vacinas.

Como resultado encontramos tanto no resultado da consulta quanto nos valores presentes no conjunto de dados do dashboard, para as variáveis de doses e total de vacinas aplicadas, demonstrando que não existe divergência entre os dados presentes no banco de dados.

**Figura 55** – Consulta dos volumetria dos principais indicadores presentes no dashboard.

Tabela

Descrição gerada automaticamente

Fonte: Própria (2023).

Como forma de verificação na figura 56 é apresentado o código SQL utilizado na validação da fato\_vacinas.

**Figura 56** – Consulta volumétrica da fato\_vacinacao, por dose e totais.



Fonte: Própria (2023).

# 5. Conclusões

No Painel Cobertura Vacinal – Covid 19 no gráfico de cobertura vacinal por dose aplicada, é possível identificar a oportunidade de melhoria da cobertura vacinal para cada dose, e apresenta a necessidade estratégica de incentivar por meio de campanhas a vacinação das demais doses, apresentado no gráfico 1.

No gráfico 2 médias moveis de óbitos e doses aplicadas identifica-se que a vacinação já a partir da primeira dose reduz enormemente a os óbitos, porém a menor adesão pela segunda dose e surgimento de novas variantes apresentam um novo aumento na mortalidade. Até que com a aplicação da terceira dose e de reforço vacinal, ocorre novamente a redução na mortalidade.

O gráfico aponta os momentos em que se deve ampliar a aplicação das doses, aumentar e solicitar a aquisição de mais doses ao ministério da saúde, e ampliar os incentivos junto à população.

O gráfico 3 Mês anterior de notificações, óbitos e vacinados, mostra que períodos com aumento no volume de vacinados, são precedidos de redução de percentuais de mortalidade e de notificações frente aos meses anteriores, isso evidencia a eficácia da vacinação, também apresenta grandes picos de notificações e óbitos, com o surgimento de novas variantes e provável redução da adesão da população.

Na visão de indicadores de cobertura vacinal no gráfico 4, é possível explorar a cobertura vacinal por dose, faixas etárias e sexo e assim direcionar a atenção para atender os grupos de risco e atuar com ações mais eficazes.

Algumas das estratégias a serem criadas a partir desses insights são a realização de busca ativa para o grupo de maior idade, gravidas e puérperas, oferecer vacinação em asilos, maternidades, presídios. O acompanhamento por município permite identificar os municípios com maior deficiência em atender os percentuais de cobertura vacinal.

E nas demais visões as análises á nível tático permitem os usuários avançarem nos detalhes dos resultados obtidos e com base nesses dados encontrar formas de atuar, atendendo os munícipios mais carentes de atenção com resultados inferiores e nivelar esse resultado se utilizando de análises por quartil.

Para trabalhos futuros poderiam ser incluídas análises mais detalhadas, como óbitos, notificações, e vacinas aplicadas por raça ou por grupos de risco

Acompanhamento da distribuição das vacinas aos municípios e estabelecimentos e identificar os gaps, que impactaram no resultado da vacinação.

Também poderiam ser feitas visões para acompanhamento de casos que evoluíram para hospitalização ou internados em uti, este dado poderia ser incluído nas análises para identificar quantos óbitos ocorreram por falta de atendimento.

Entretanto foram encontrados alguns empecilhos, a princípio o censo demográfico encontrado está desatualizado, trazendo a população contada em 2010 e por falta de não haver a mesma base de dados com volume atual da população ocorreu que alguns percentuais de cobertura vacinal extrapolaram os 100%, dado a falta de volume populacional por município.

A base de dados disponibilizada pelos Opendatasus, apesar de ser rica em detalhes na informação, não apresenta consistência nos dados, dado que a base de síndrome aguda grave não traz a evolução completa dos pacientes notificados, e o volume de óbitos, hospitalização e internações em uti não reflete a realidade baseado nos boletins oficiais publicados pelas secretarias de saúde, sendo necessário incluir como fonte de dados de óbitos notificação de síndrome gripal e casos suspeitos de Covid-19 que apresentou o volume mais correto de óbitos.

# 6. Links

Segue abaixo os links com os artefatos criados no projeto.

Dashboard Painel Covid AWS, disponível em: <  
https://drive.google.com/drive/folders/18Zh\_vWzBhpKM2nUYUOEVM21oCuPGltxs?usp=sharing>

Codigo do processo de ETL, diponivel em: https://github.com/promarcosrt/TCC\_Pucminas.git

Vídeo de apresentação, disponível em: < <https://youtu.be/UPvkPVXhiAE>> ou <https://drive.google.com/drive/folders/1hp-ZYAn\_lmCcwjpIrIBliErgQBIUjWzp?usp=sharing>

# REFERÊNCIAS

Como um projeto aplicado, as referências, levantadas no processo de revisão bibliográfica, são opcionais e estimulamos que o responsável pelo projeto apresente aquilo que for relevante para complementar o entendimento do trabalho. Dessa forma, relacione-as de acordo com o modelo a seguir.

RAFAEL, Ricardo de Mattos Russo et al. Epidemiologia, políticas públicas e pandemia de Covid-19: o que esperar no Brasil?[Epidemiology, public policies and Covid-19 pandemics in Brazil: what can we expect?][Epidemiologia, políticas públicas y la pandémia de Covid-19 en Brasil: que podemos esperar?]. **Revista enfermagem UERJ**, v. 28, p. 49570, 2020.

FERNANDES, Antônio Sérgio Araújo; TUDE, João Martins. A pandemia de covid-19 no Brasil e o falso dilema economia x saúde.

COELHO, Akeni Lobo et al. A utilização de tecnologias da informação em saúde para o enfrentamento da pandemia do Covid-19 no Brasil. **Cadernos Ibero-Americanos de Direito Sanitário**, v. 9, n. 3, p. 183-199, 2020.

PAHO, Epidemiological Calendar 2016: A basic element for the use of the time variable in health surveillance, Disponível em: <https://www.paho.org/hq/dmdocuments/2016/2016-cha-epidemiological-calendar.pdf>. Acesso em: abril de 2023.

Documentação do AWS, **Amazon Web Services**, Inc. ou suas afiliadas, c2022, Disponível em: <https://docs.aws.amazon.com/pt\_br/index.html> Acesso em: março de 2022.

RIBEIRO HOMEM, Antonio Moreno, et al. Concepção e implementação de uma solução de Business Intelligence para gestão financeira de clientes em uma plataforma de e-commerce. 2021.

GARCIA, Enzo Viviani. Estrutura e características de um sistema de business intelligence baseado em data lake em uma empresa do setor de energia. 2021.

AMORIM, Leonardo Afonso. Introdução ao Spark com Pyspark. *Sociedade Brasileira de Computação*, 2021.

FRANCISCO, Klesley Goncalves. Análise de dados educacionais para determinação de perfis de evasão universitária. 2021.

SILVA, Douglas da. O que é ETL? [S.l.: s.n.], jan. 2021. Disponível em: <

https://www.zendesk.com.br/blog/o-que-e-etl/>. Acesso em: set. 2021

BALLARD, Chuck et al. **Dimensional Modeling:** **In a Business Intelligence Environment**. IBM Redbooks, 2012.

gov.br, Ministério da Saúde e Fiocruz traçam estratégias para aumentar coberturas vacinais no país, disponível em: <https://www.gov.br/saude/pt-br/assuntos/noticias/2021-1/dezembro/ministerio-da-saude-e-fiocruz-tracam-estrategias-para-aumentar-coberturas-vacinais-no-pais>, Acesso em janeiro 2023

SRAG 2020 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19, disponível em: <https://opendatasus.saude.gov.br/dataset/srag-2020>. Acesso em: janeiro 2023

SRAG 2021 a 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19, disponível em: < https://opendatasus.saude.gov.br/dataset/srag-2021-a-2023>. Acesso em: janeiro 2023.

Notificações de Síndrome Gripal – 2022, disponível em: < https://opendatasus.saude.gov.br/dataset/notificacoes-de-sindrome-gripal-leve-2022>. Acesso em: abril 2023.

Dicionário de Dados, Notificação de Síndrome Gripal - 2022, disponível em: <https://opendatasus.saude.gov.br/dataset/notificacoes-de-sindrome-gripal-leve-2022/resource/2fe23e2a-8f58-4ea6-8529-744e3f20ff2b>. Acesso em janeiro 2023.

Sistema IBGE de Recuperação Automática – SIDRA, Tabela de dados agregados: 3175 - População residente, por cor ou raça, segundo a situação do domicílio, o sexo e a idade, diponivel em: <https://sidra.ibge.gov.br/Tabela/3175>.Acesso em janeiro 2023.

MicroStrategy, Introduction to the MicroStrategy App, disponível em: <https://www2.microstrategy.com/producthelp/Current/MicroStrategy/en-us/Content/intro\_microstrategy.htm?tocpath=\_\_\_\_\_2>, Acesso em janeiro 2023.

MicroStrategy, MicroStrategy Help, disponível em: < https://www2.microstrategy.com/producthelp/Current/MicroStrategy/en-us/Content/home\_microstrategy.htm?tocpath=\_\_\_\_\_1>, Acesso em janeiro 2023.

Microstrategy, How to connect to Amazon Athena, disponível em: < https://community.microstrategy.com/s/article/How-to-connect-to-Amazon-Athena?language=en\_US#ConnMSTRDeskWin >, Acesso em janeiro 2023.