



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

NÚCLEO DE EDUCAÇÃO A DISTÂNCIA

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

RELATÓRIO TÉCNICO

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. Introdução4**
 - 1.1. Contexto4
 - 1.2. Objetivos.....5
 - 1.3. Público alvo5
- 2. Modelo de Dados6**
 - 2.1. Modelo Dimensional6
 - 2.2. Fatos e Dimensões8
- 3. Integração, Tratamento e Carga de Dados12**
 - 3.1. Fontes de Dados.....16
 - 3.2. Processos de Integração e Carga (ETL).....17
- 4. Camada de Apresentação28**
 - 4.1 Dashboard33
 - 4.2 Análises avançadas Erro! Indicador não definido.
- 5. Registros de Homologação43**
- 5. Conclusões47**
- 6. Links51**
- REFERÊNCIAS53**

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 presenciais 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 comércio e indústria também foram impactados com a evolução da pandemia.

Ao demais funcionários públicos comissionados ou concursados que atuam como Analistas de MIS (Management Information System), 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.

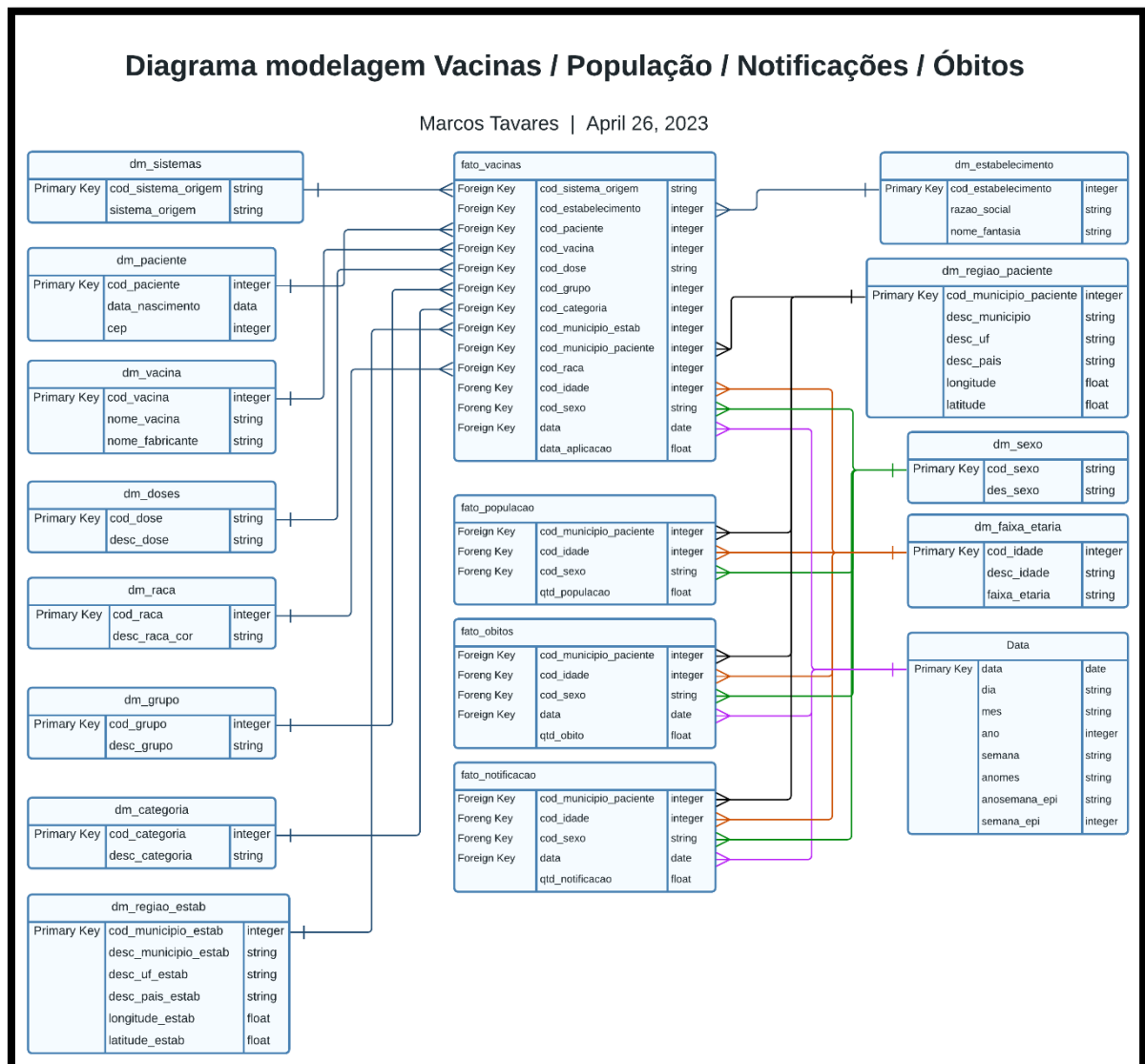
As dimensões e fatos serão transformadas para seguir modelo multidimensional no esquema multi-estrela.

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

2.2.1 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

Nome	Tipo	Descrição	Exemplo
fato_vacinas	fato	Quantidade de vacinas aplicadas	qtd_vacinas_aplicadas: 1,1001, 5099100

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Fato	fato_vacinas	sistema_origem (gerado hash256)	cod_sistema_origem	string
		estabelecimento_valor	cod_estabelecimento	integer
		vacina_grupoAtendimento_codigo,	cod_grupo	integer
		vacina_categoria_codigo,	cod_categoria	integer
		vacina_codigo,	cod_vacina	integer
		paciente_id	cod_paciente	integer
		paciente_racaCor_codigo,	cod_raca	integer
		dose (gerado hash256)	cod_dose	string
		paciente_endereco_colbgeMunicipio,	cod_municipio_estab	integer
		estabelecimento_municipio_codigo,	cod_municipio_paciente	integer
		paciente_enumSexoBiologico	cod_sexo	string
		paciente_idade	cod_idade	integer
		vacina_dataAplicacao,	data	date
		Count(document_id)	QtdVacinasAplicadas	float

Fonte: Própria (2023).

2.2.2 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 é 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

Nome	Tipo	Descrição	Exemplo
fato_populacao	fato	Quantidade da população	qtd_populacao: 2,3,100,253500

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Fato	fato_populacao	Cod	cod_municipio_paciente	integer
		Sexo	cod_sexo	string
		Idade	cod_idade	integer
		Sum(Qtd)	qtd_populacao	float

Fonte: Própria (2023).

2.2.3 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

Nome	Tipo	Descrição	Exemplo
fato_notificacao	fato	Quantidade de Notificações	qtd_notificacao: 2,1,15233

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Fato	fato_notificacao	municipioIBGE	cod_municipio_paciente	integer
		Sexo	cod_sexo	string
		Idade	cod_idade	integer
		dataNotificacao	data	date
		Sum(*)	qtd_notificacao	float

Fonte: Própria (2023).

2.2.4 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

Nome	Tipo	Descrição	Exemplo
fato_obitos	fato	Quantidade de obitos	qtd_obito: 3,4,100,100

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Fato	fato_obito	CO_MUN_RES	cod_municipio_paciente	integer
		CS_SEXO	cod_sexo	string
		NU_IDADE_N	cod_idade	integer
		DT_NOTIFIC	data	date
		Sum(*)	qtd_obito	float

Fonte: Própria (2023).

2.2.5 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

Nome	Tipo	Descrição	Exemplo
dm_sistemas	Dimensão	Sistemas usados para registros das doses	cod_sistemas:IDS Saúde, Novo PNI, Esus
dm_paciente	Dimensão	Dados do paciente	cod_paciente: 9d731cac7dedfb95894a667035cc58e9817c83 341563204519fba52505c056b1; data_nascimento: 01/02/1998; CEP: 4012352;
dm_vacinas	Dimensão	Dados das vacinas	cod_vacina: 85,88; desc_vacina: COVID-19 PFIZER-CORMIRNATY; Fabricante: Aztazenica,Jansen;
dm_doses	Dimensão	Descrição da dose aplicada	cod_dose: 1ª Dose, 2ª Dose, Reforço;desc_dose: 1ª Dose, 2ª Dose, Reforço;
dm_grupo	Dimensão	Grupo profissional atendido	cod_grupo: 201,230; desc_grupo: Biologo, Caminhoneiro, Cozinheiro;
dm_data	Dimensão	Data da aplicação para acompanhamento	Data da Aplicação: 19/01/2021; dia: 01,02; mês: Março, Abril; Ano: 2021,2022; Semana : 21, 22; mês_ano: 012021;
dm_estabelecimento	Dimensão	Local da Aplicação	razao_social: MUNICIPIO DE ARACAJU, PREFEITURA MUNICIPAL DE UMBAUBA; nome_fantasia: CENTRO DE SAUDE 1, UBS GETULIO SAVIO SOBRAL;
dm_regiao_estab	Dimensão	Região onde foram aplicadas as doses	cod_municipio_estab: 280010, 280100; desc_municipio: ARAUA, ARACAJU; desc_uf: SE; desc_pais: BRASIL;
dm_regiao_paciente	Dimensão	Região origem do paciente	cod_municipio_estab: 110010, 120100; desc_municipio: ARAUA, ARACAJU; desc_uf: SE; desc_pais: BRASIL, ALEMANHA;
dm_raca	Dimensão	Descrição da raça dos pacientes	cod_raca: 01,02; desc_raca_cor: BRANCA,INDIGENA;
dm_idade	Dimensão	Idade dos Pacientes e Faixa etaria	cod_idade: 1,110; desc_idade: "1 ano", "10 anos"; faixa_etaria: "De 10 a 19 anos", "Mais de 90 anos"
dm_sexo	Dimensão	Sexo dos Pacientes	cod_sexo: M, F; desc_sexo: Feminino, Masculino;
dm_categoria	Dimensão	Categoria condicional do paciente	cod_categoria: 1,2,3; desc_categoria: Gestantes, Comorbidades;

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Dimensão	dm_estabelecimento	estabelecimento_valor	cod_estabelecimento	integer
		estabelecimento_razaoSocial	razao_social	string
		estalecimento_noFantasia	nome_fantasia	string

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Dimensão	dm_paciente	paciente_id	cod_paciente	integer
		paciente_dataNascimento	data_nascimento	date
		paciente_endereco_cep	cep	integer

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Dimensão	dm_vacina	vacina_codigo	cod_vacina	integer
		vacina_nome	nome_vacina	string
		vacina_fabricante_nome	nome_fabricante	string

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Dimensão	dm_doses	cod_dose(gerado) hash256	cod_dose (gerado)	string
		vacina_descricao_dose	desc_dose	string

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Dimensão	dm_grupo	vacina_grupoAtendimento_codigo	cod_grupo	integer
		vacina_grupoAtendimento_nome	desc_grupo	string

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Dimensão	dm_categoria	vacina_categoria_codigo	cod_categoria	integer
		vacina_categoria_nome	desc_categoria	string

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 raça

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Dimensão	dm_raca	paciente_racaCor_codigo	cod_raca	integer
		paciente_racaCor_valor	desc_raca	string

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Dimensão	dm_regiao_estab	estabelecimento_municipio_codigo	cod_municipio_estab	integer
		estabelecimento_municipio_nome	desc_municipio_estab	string
		estabelecimento_uf	desc_uf_estab	string
		estabelecimento_pais	desc_pais_estab	string
		longitude	longitude_estab	float
		latitude	latitude_estab	float

Fonte: Própria (2023).

Figura 19 – Detalhamento da estrutura dimensão região do paciente

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Dimensão	dm_regiao_paciente	paciente_endereco_colbgeMunicipio	cod_municipio	integer
		paciente_endereco_nmMunicipio	desc_municipio	string
		paciente_endereco_uf	desc_uf	string
		paciente_endereco_nmPais	desc_pais	string
		longitude	longitude	float
		latitude	latitude	float

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Dimensão	dmsexo	paciente_enumSexoBiologico	codsexo	string
		dessexo	dessexo	string

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Dimensão	dmidade	paciente_idade	cod_idade	integer
		desc_idade	desc_idade	string
		faixa_etaria	faixa_etaria	string

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Dimensão	dm_data	vacina_dataAplicacao,	data_aplicacao	date
		dia (transformado)	dia	string
		mes (transformado)	mes	string
		ano (transformado)	ano	string
		semana (transformado)	semana	string
		anomes(transformado)	anomes	date
		semana_epi(transformado)	semana_epi	string
		anosemana_epi(transformado)	anosemana_epi	string
		mes_ano(transformado)	mes_ano	string

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

Tipo	Nome	Campos origem	Apelido	Tipo Dados
Dimensão	dm_sistemas	sistema_origem (gerado) hash256	cod_sistema_origem	string
		sistema_origem	desc_sistema	string

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

Download

* 100.800 valores na seleção

Nome do arquivo:

Formato: ☒ Comprimir (.zip)

Ranquear valores: Caso a tabela possua mais de uma coluna, o ranking será dado pelos valores da primeira coluna

☒ Exibir siglas de níveis territoriais
☒ Exibir códigos de territórios
☒ Exibir nomes de territórios
☒ Exibir unidades de medida como coluna

☒ Imediato (até 200.000 valores) ☐ A Posteriori (até 3.000.000 valores)

[Veja as gravações a posteriori efetuadas nos últimos 60 dias](#)

Fonte: Própria (2023).

3.2. Processos de Integração e Carga (ETL)

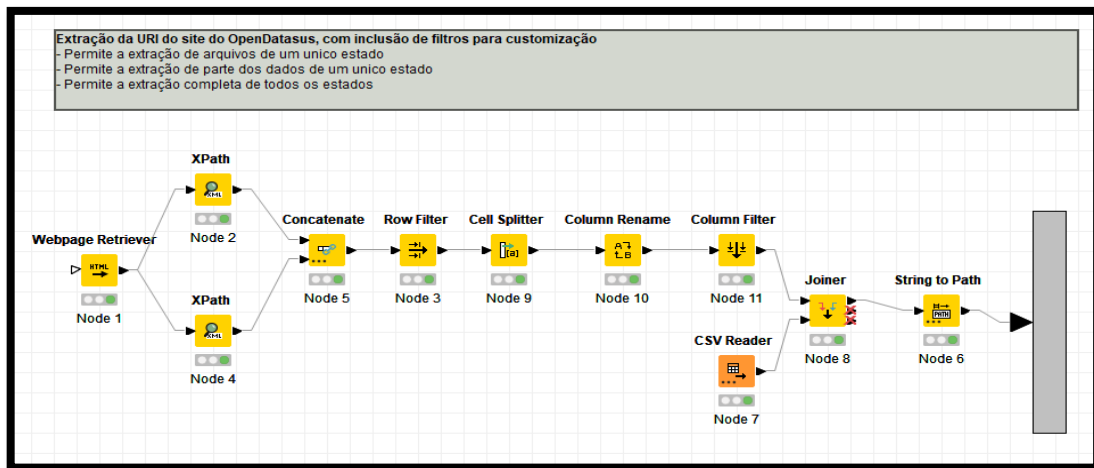
Conforme explicado por Ribeiro Homem, (2021 p.26) O processo de ETL é fundamental nas práticas implementadas pelo BI, visto que para realizar análises de dados de negócio é necessário dispor de informação no formato adequado. Dessa forma, é necessário diagnosticar e compreender o problema tratado para 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 vírgulas, 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 “Web scraping”(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



Fonte: Própria (2023).

Figura 26 – Url segregadas por parte, estado e região

Join result - 0:7:8 - Joiner

File Edit Hilit Navigation View

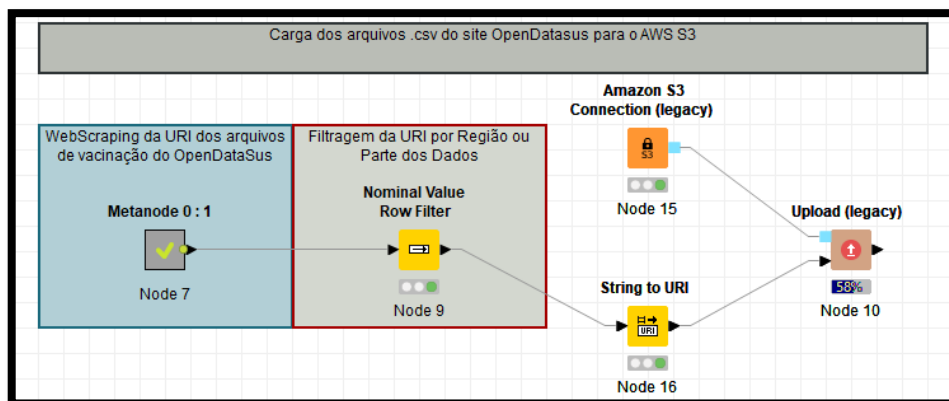
Table "default" - Rows: 3 Spec - Columns: 4 Properties Flow Variables

Row ID	href	parte	estado	regiao
Row0_24_Ro...	https://s3.sa-east-1.amazonaws.com/dkan.saude.gov.br/SIPNI/COVID/uf/uf%3DSE/part-00000-26f414ff-4c20-4f7d-bfda-a89d003a154c.c000.csv	1	Sergipe	Nordeste
Row0_25_Ro...	https://s3.sa-east-1.amazonaws.com/dkan.saude.gov.br/SIPNI/COVID/uf/uf%3DSE/part-00002-26f414ff-4c20-4f7d-bfda-a89d003a154c.c000.csv	3	Sergipe	Nordeste
Row0_13_du...	https://s3.sa-east-1.amazonaws.com/dkan.saude.gov.br/SIPNI/COVID/uf/uf%3DSE/part-00001-26f414ff-4c20-4f7d-bfda-a89d003a154c.c000.csv	2	Sergipe	Nordeste

Fonte: Própria (2023).

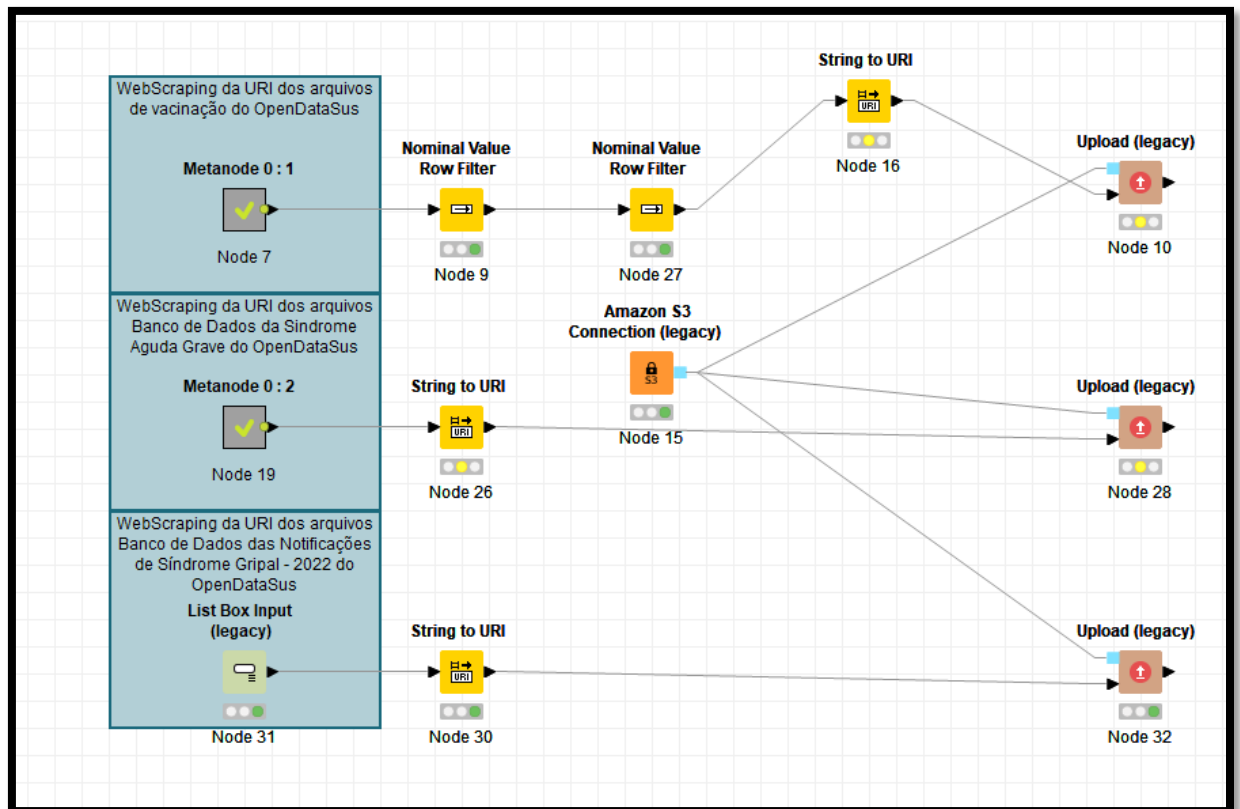
Como apresentado na figura 27 temos o workflow que permite a filtragem das Url's, em seguida a conversão para realizar o upload destes arquivos para S3 da Amazon após a conexão por chave de segurança na "Amazon Web Services" (Serviços de Internet da Amazon),(AWS). E na figura 28 o workflow completo de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

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



Fonte: Própria (2023).

Figura 28 – Workflow no Knime completo.

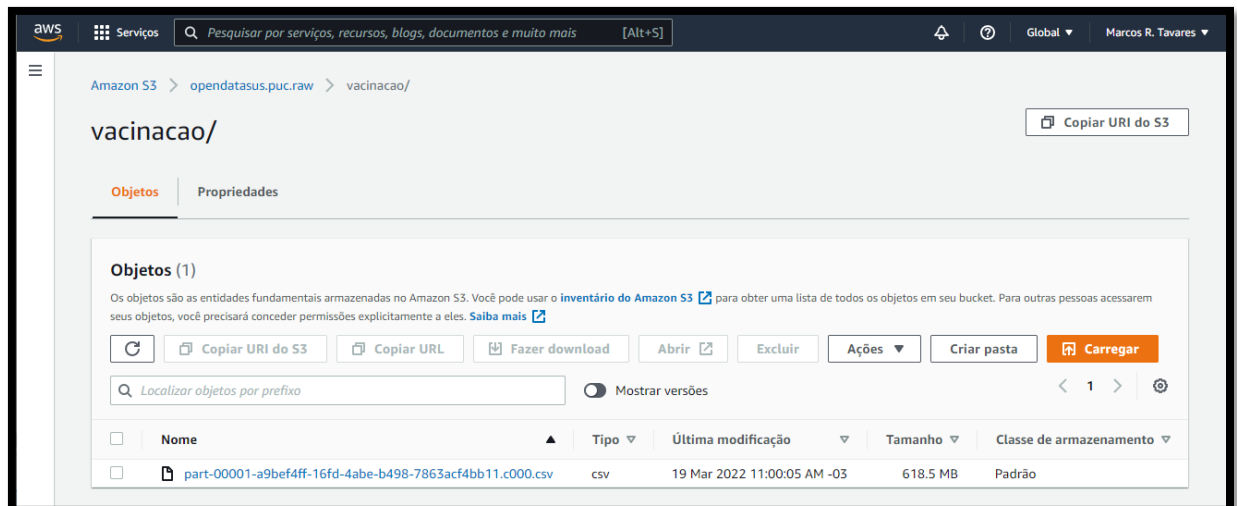


Fonte: Própria (2023).

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



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`, demonstrado 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`.

```

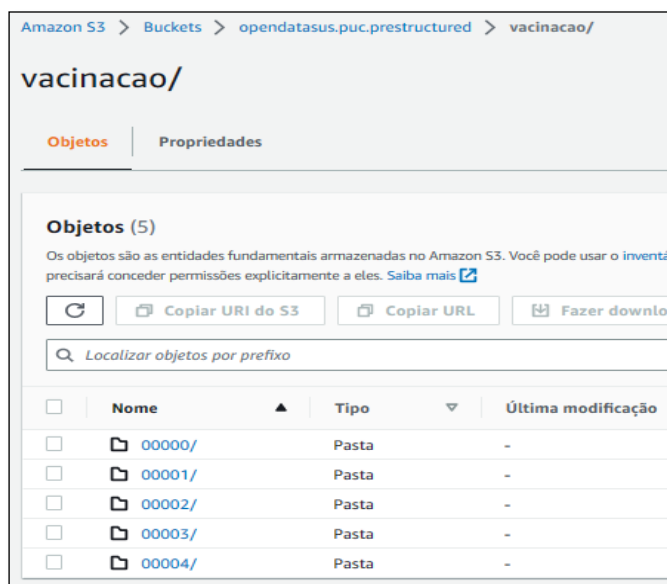
45     print(f"ERROR: {e}")
46
47 def carga_vacinas(prefixo,path_prestr):
48
49     lista = listar_arquivos_s3(prefixo)
50     print(lista)
51
52     for path in lista:
53         pstr = path.split("-")
54
55         parte = str(pstr[1])
56         print(f"| parte: {parte} e path: {path} |")
57
58     try:
59         print("Inicio dados Brutos")
60
61         df = spark.read.option("header","True").option("delimiter",";").csv(path)
62         df = df.withColumn("cod_sistema_origem",substring(upper(trim(col("sistema_origem"))),1,3))
63         df = df.withColumn("cod_dose",sha2(trim(col("vacina_descricao_dose")),256))
64         df = df.withColumn("cod_estabelecimento",trim(col("estabelecimento_valor")))
65         df = df.withColumn("cod_grupo",trim(col("vacina_grupoAtendimento_codigo")))
66         df = df.withColumn("cod_categoria",trim(col("vacina_categoria_codigo")))
67         df = df.withColumn("cod_vacina",trim(col("vacina_codigo")))
68         df = df.withColumn("cod_paciente",trim(col("paciente_id")))

```

Python Ln 1, Col 1 Errors: 0 Warnings: 0

Fonte: Própria (2023).

Figura 30 – Dados estruturados no "bucket" do S3.



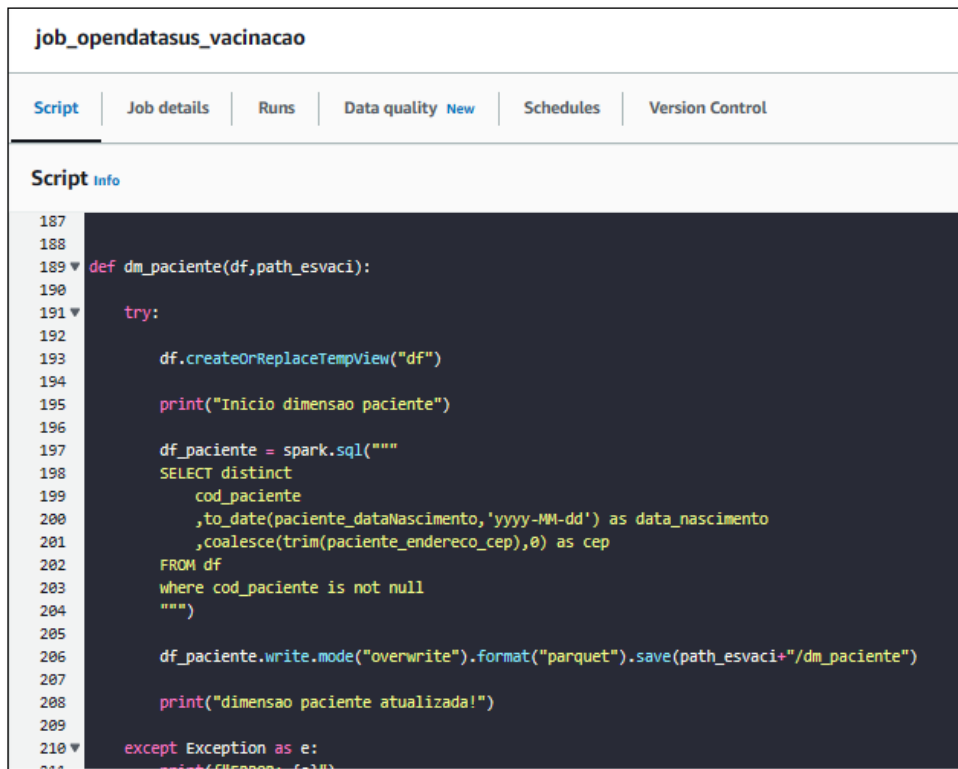
Fonte: Própria (2023).

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.



```

187
188
189 def dm_paciente(df,path_esvaci):
190
191     try:
192
193         df.createOrReplaceTempView("df")
194
195         print("Inicio dimensao paciente")
196
197         df_paciente = spark.sql("""
198         SELECT distinct
199             cod_paciente
200             ,to_date(paciente_dataNascimento,'yyyy-MM-dd') as data_nascimento
201             ,coalesce(trim(paciente_endereco_cep),0) as cep
202         FROM df
203         where cod_paciente is not null
204         """)
205
206         df_paciente.write.mode("overwrite").format("parquet").save(path_esvaci+"/dm_paciente")
207
208         print("dimensao paciente atualizada!")
209
210     except Exception as e:
211         print(f"Erro: {e}")

```

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

```

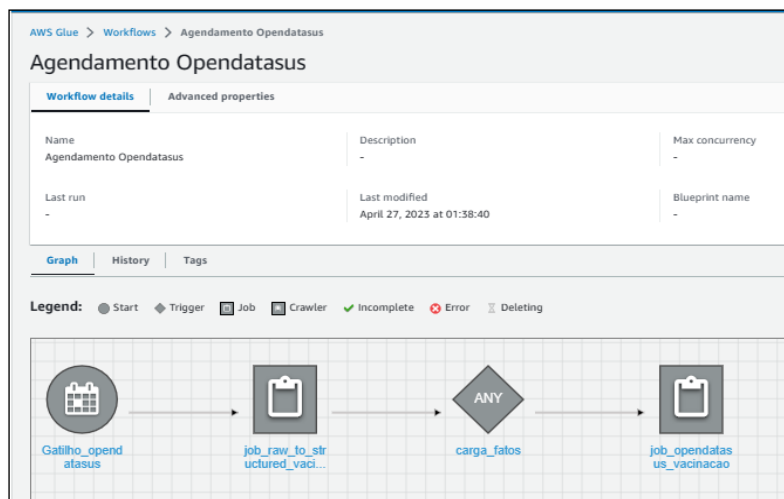
13 from awsglue.utils import getResolvedOptions
14 from pyspark.sql import SparkSession
15 from datetime import date, timedelta, datetime
16 from pyspark.sql.window import Window
17 import os
18 import datetime
19 from dateutil.relativedelta import relativedelta
20
21 # Setting Glue Context
22 sc = SparkContext()
23 glueContext = GlueContext(sc)
24 spark = glueContext.spark_session
25
26
27 path_vacina = "s3://opendatasus.puc.raw/vacinacao/"
28 path_prestr = "s3://opendatasus.puc.prestructured/vacinacao/"
29 path_esvac1 = "s3://opendatasus.puc.structured/vacinacao/"
30 path_estemp = "s3://opendatasus.puc.structured/tempo/"
31 path_popula = "s3://opendatasus.puc.structured/populacao/fato_populacao/"
32 path_locali = "s3://opendatasus.puc.raw/auxiliares/BR_Localidades_2010_v1.csv"
33
34 def cordenadas_municipios(path_locali):
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53 def carga_vacinas(path_prestr):
54
55
56
57
58
59
60
61
62
63
64
65 def fato_vacinas(df, path_esvac1, path_popula):
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108 def dm_sistemas(df, path_esvac1):
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141 def dm_doses(df, path_esvac1):
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164 def dm_sexo(df, path_esvac1):
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186 def dm_paciente(df, path_esvac1):
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210 def dm_estabelecimento(df, path_esvac1):
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234 def dm_vacina(df, path_esvac1):
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258 def dm_grupo(df, path_esvac1):
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281 def dm_categoria(df, path_esvac1):
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304 def dm_raca(df, path_esvac1):
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327 def dm_regiao_estab(df, df_cord, path_esvac1):
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356 def dm_regiao_paciente(df, df_cord, path_esvac1):
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386 def dm_data_aplicacao(df, path_estemp):
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414 def main():
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436

```

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.

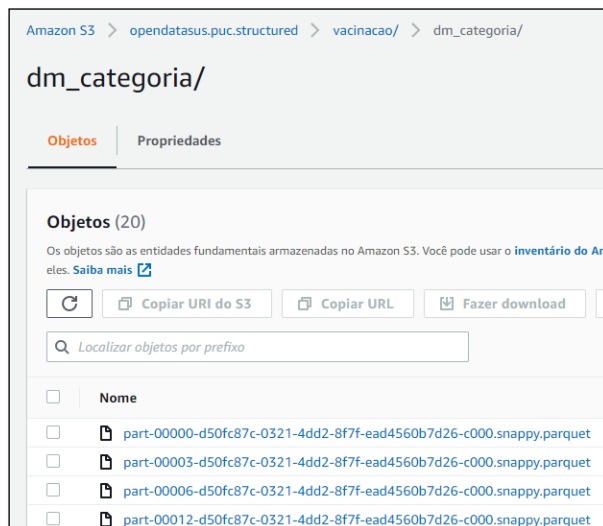
Figura 33 – Exemplo de construção de um workflow no Glue, Agendamento Opendatasus.



Fonte: Própria (2023).

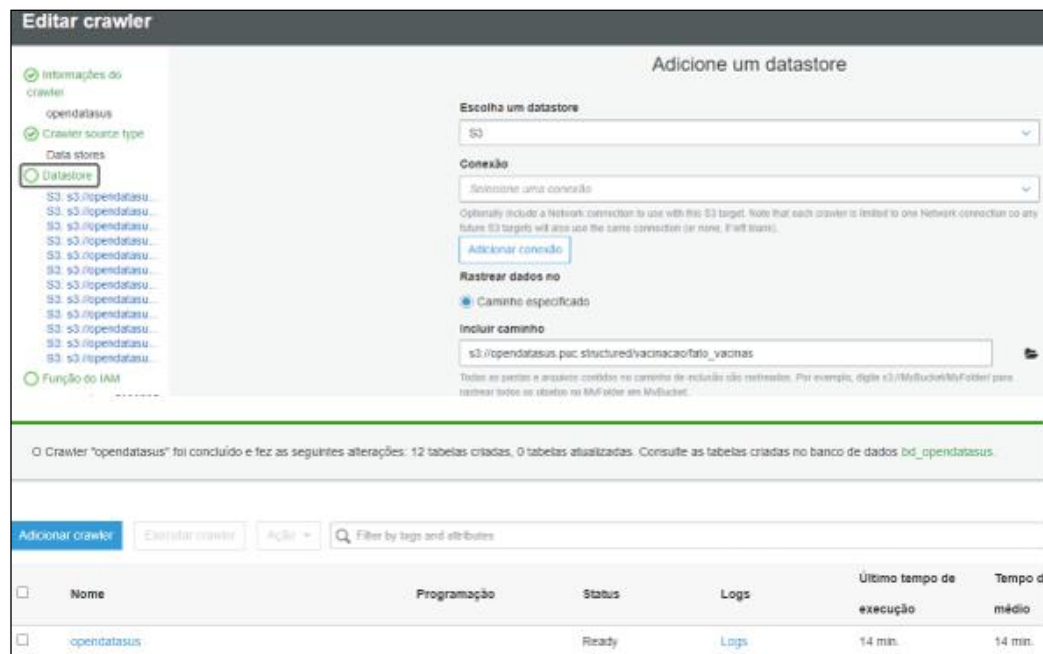
As dimensão e fato são salvos em formato de arquivo parquet particionados no S3 em um bucket específico para organizar e separar os dados brutos e transacionais dos dados que já foram estruturados, como pode ser visualizado na figura 33, e estes arquivos serão lidos pelo crawler para carregar no banco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

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



Fonte: Própria (2023).

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



Fonte: Própria (2023).

No banco de dados Athena é possível 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

Query 1 X Query 2 X

```

1 SELECT
2   D.ano,
3   O.vacina_descricao_dose,
4   sum(F.qtd_vacinas_aplicadas) as Qtd_Vacinas_Aplicadas
5 FROM "bd_opendatasus"."fato_vacinas" F
6 JOIN "bd_opendatasus"."dm_data" D on f.data_aplicacao = D.data_aplicacao
7 JOIN "bd_opendatasus"."dm_doses" O on f.cod_dose = O.cod_dose
8 group by
9   D.ano
10  ,O.vacina_descricao_dose
11 limit 10;

```

SQL Ln 11, Col 29

Run again Cancel Save Clear Create

Completed Time in queue: 0.134 sec

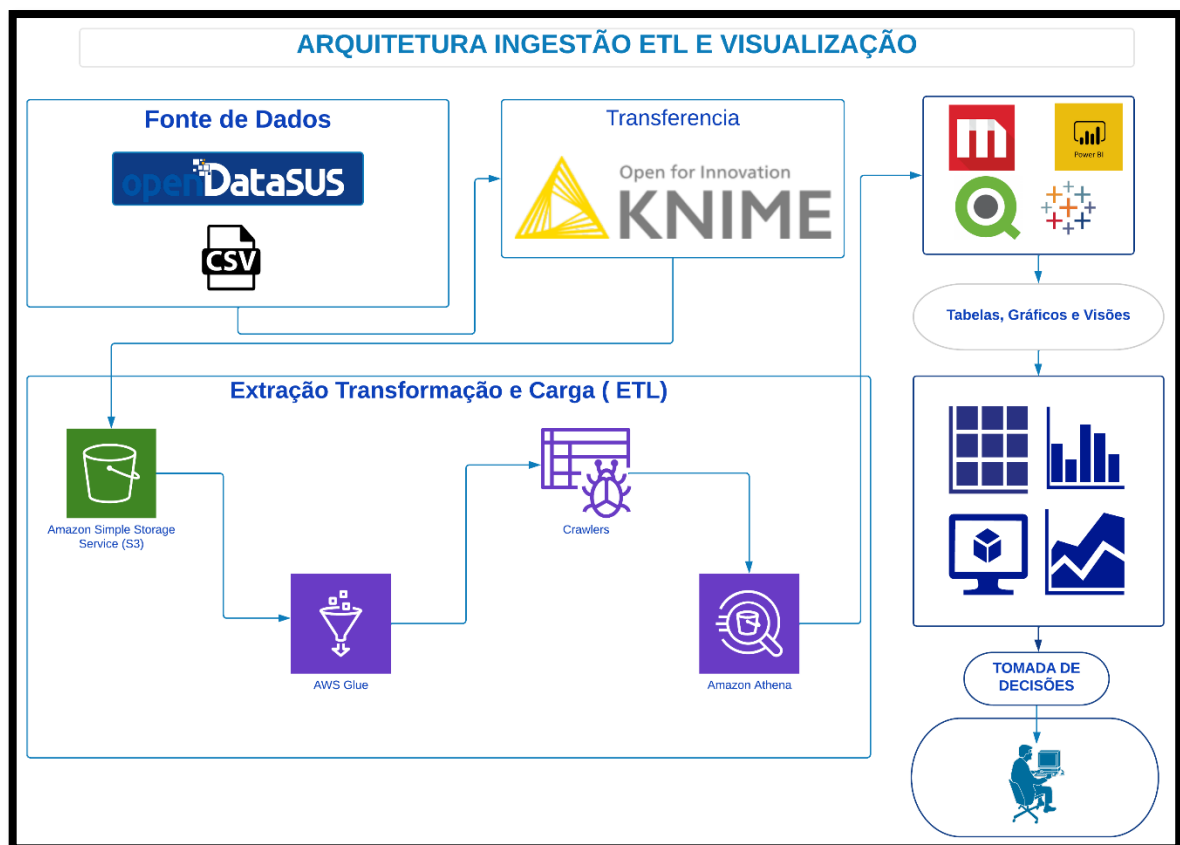
Results (10)

#	ano	vacina_descricao_dose	Qtd_Vacinas_Aplicadas
1	2021	Reforço	80450
2	2022	1ª Dose	35370
3	2021	Dose Adicional	2115
4	2022	2ª Dose	24266
5	2022	Dose Adicional	2688

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, Tableau e demais outras conforme figura 36.

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



Fonte: Própria (2023).

3.3. 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 corporativa 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 Microstrategy

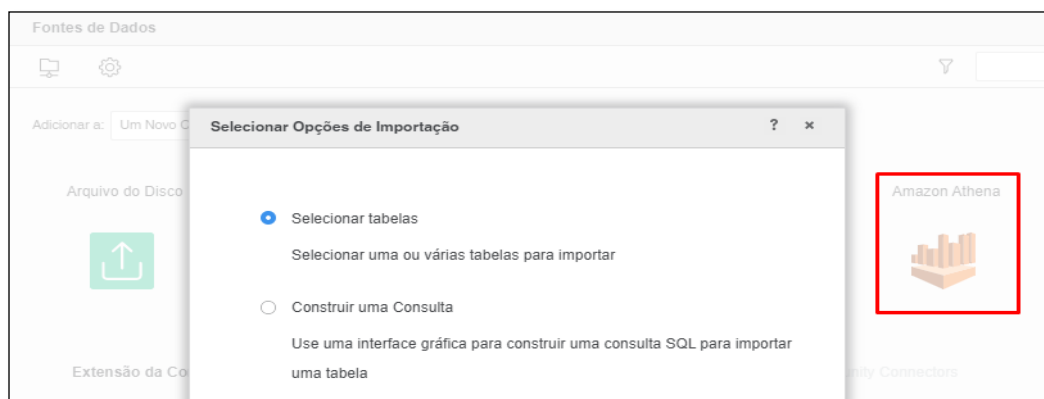
O Microstrategy permite que sejam realizadas diversas conexões com as mais diversas fontes de dados, para atender a necessidade do projeto existe a conexão com a base de dados athena, entretanto foi necessário baixar e incluir na pasta da aplicação o Amazon Athena JDBC Driver, conforme descrito no site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

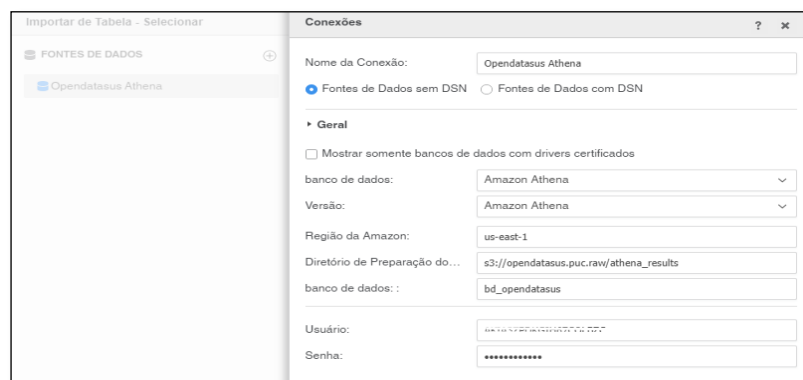
Realizada essa configuração ao acessar a ferramenta, iniciando a criação de um novo dossier, no menu Inserir, adicionar novos dados e será disponibilizado na tela de fonte de dados a opção para conectar os dados no Amazon Athena, selecionar a opção importar tabelas, porém também é possível realizar consultas 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



Fonte: Própria (2023).

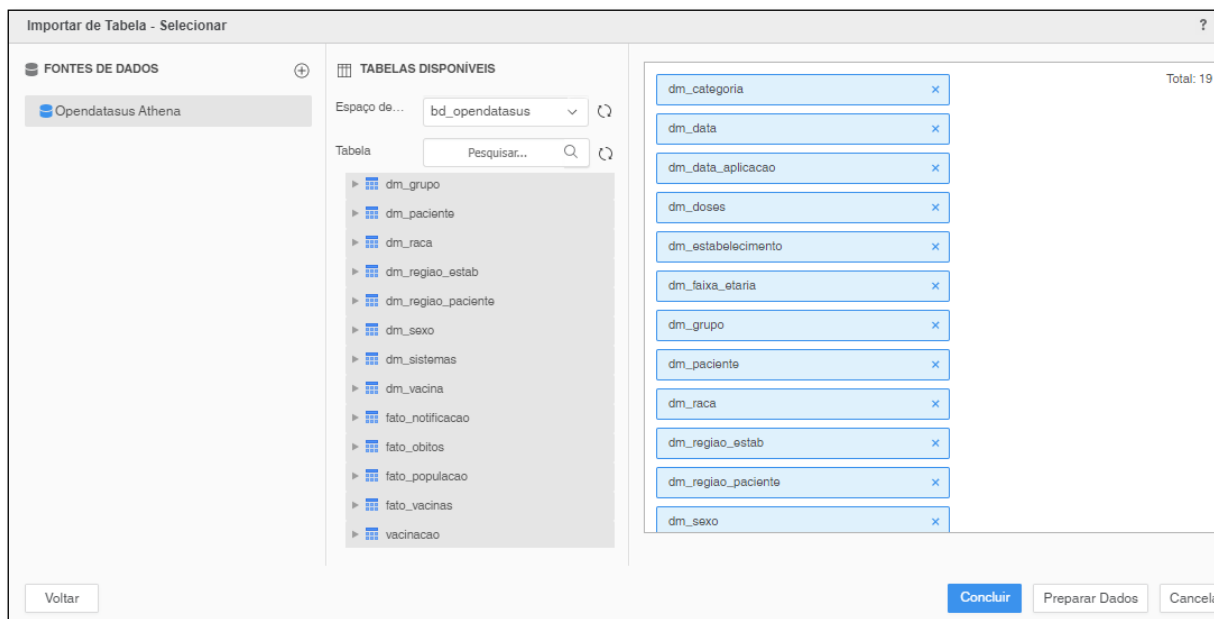
Figura 38 – Configurar a conexão com a AWS.



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.



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.

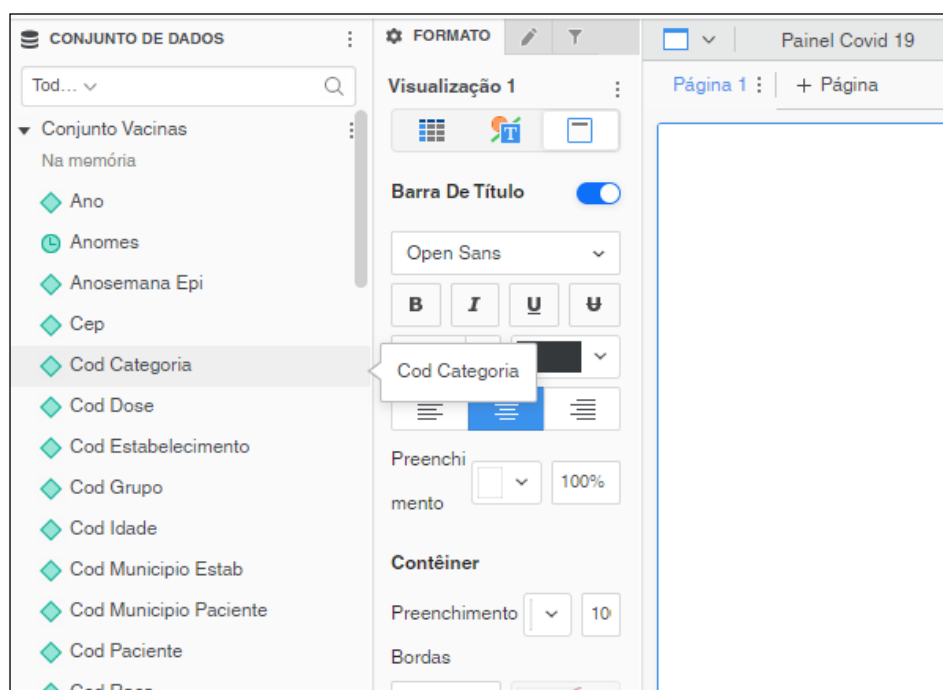
Figura 40 – Preparação dos dados, no Microstrategy.

Cod Dose	Cod Estabelecimento	Cod Grupo	Cod Idade	Cod Municipio Estab	Cod
0705c575ea2f4bcafb8571a9eb71972...	2380	201	18	280030	280030
0705c575ea2f4bcafb8571a9eb71972...	2542	201	19	280030	280030
0705c575ea2f4bcafb8571a9eb71972...	2526	201	20	280030	280670
0705c575ea2f4bcafb8571a9eb71972...	2600366	201	20	280030	280350
0705c575ea2f4bcafb8571a9eb71972...	2600366	201	20	280030	280350

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.



Fonte: Própria (2023).

4. Camada de Apresentação

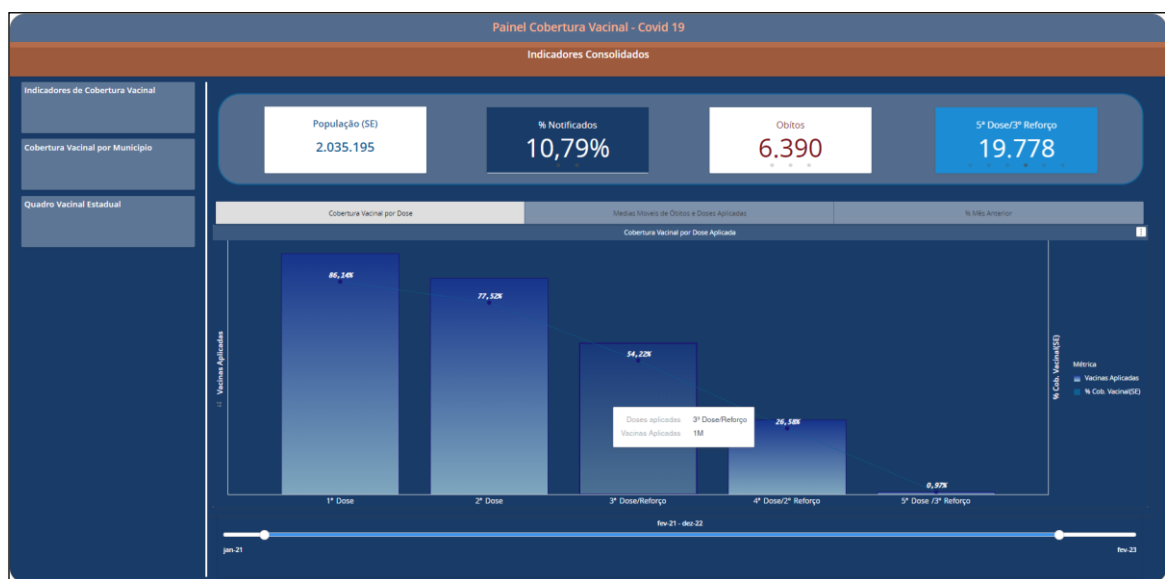
4.1 Dashboard

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 Tático

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.



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.

População (SE) 2.035.195	% Notificados 10,79%	% Óbito / Notificações 2,91%	Vacinas Aplicadas 4.995.117
População (SE) 2.035.195	% Notificados 10,79%	% Óbitos / População 0,31%	2ª Dose 1.577.716
População (SE) 2.035.195	Total Notificações 219.622	Óbitos 6.390	Vacinas Aplicadas 4.995.117

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.

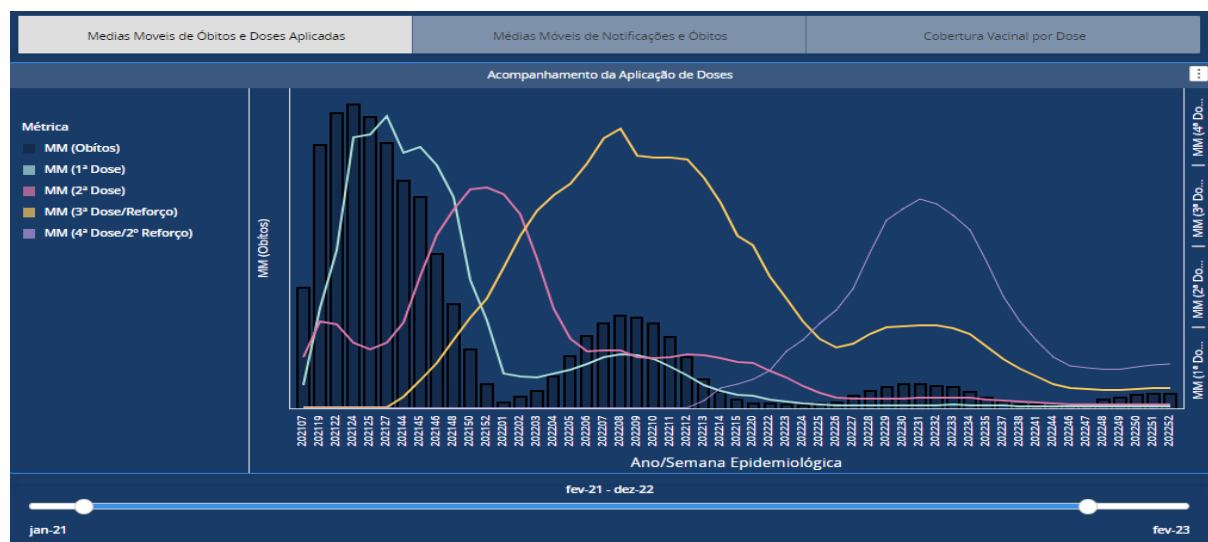
Gráfico 1 – Cobertura vacinal por dose, cenário estadual.



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.

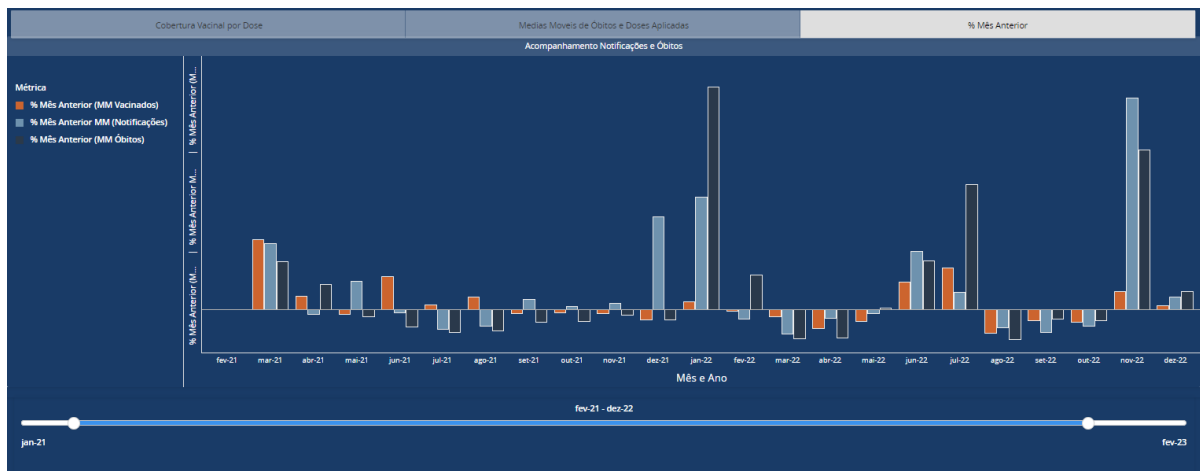


Fonte: Própria (2023).

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 doses aplicadas.

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



Fonte: Própria (2023).

Análise Estratégico

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 meningite 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.

Vacinas Aplicadas	% Cob. Vac. Idade	Vac. Homens	% Cob. Vac. Homens	Vac. Mulheres	% Cob. Vac. Mulheres
1.753.147	86,29%	816.138	82,72%	937.009	89,66%

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.

▼ Bandas

Tipo: % 123

☒ Faixa 4

Min. 70.01 Máx. 80

Cor: [Laranja]

☒ Faixa 3

Min. 95.01 Máx. 100

Cor: [Verde]

☒ Faixa 2

Min. 80.01 Máx. 95

Cor: [Amarelo]

☒ Faixa 1

Min. 0 Máx. 70

Cor: [Vermelho]

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.



Fonte: Própria (2023).

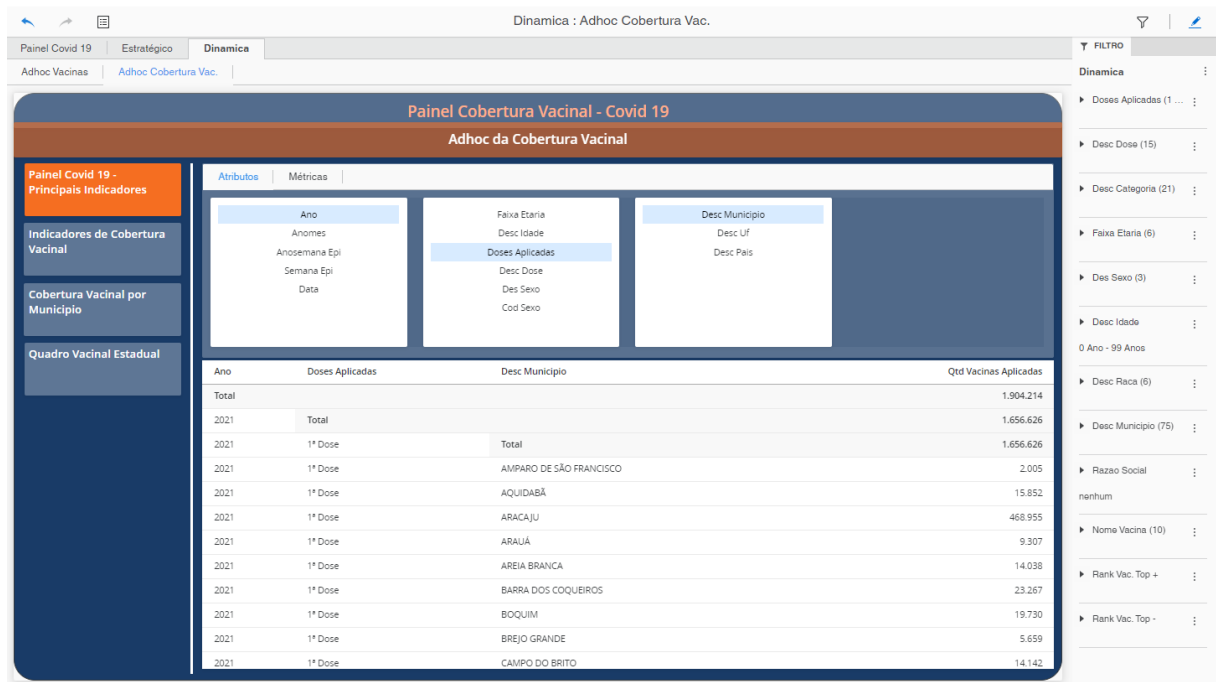
Análise Operacional

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.

Fonte: Própria (2023).

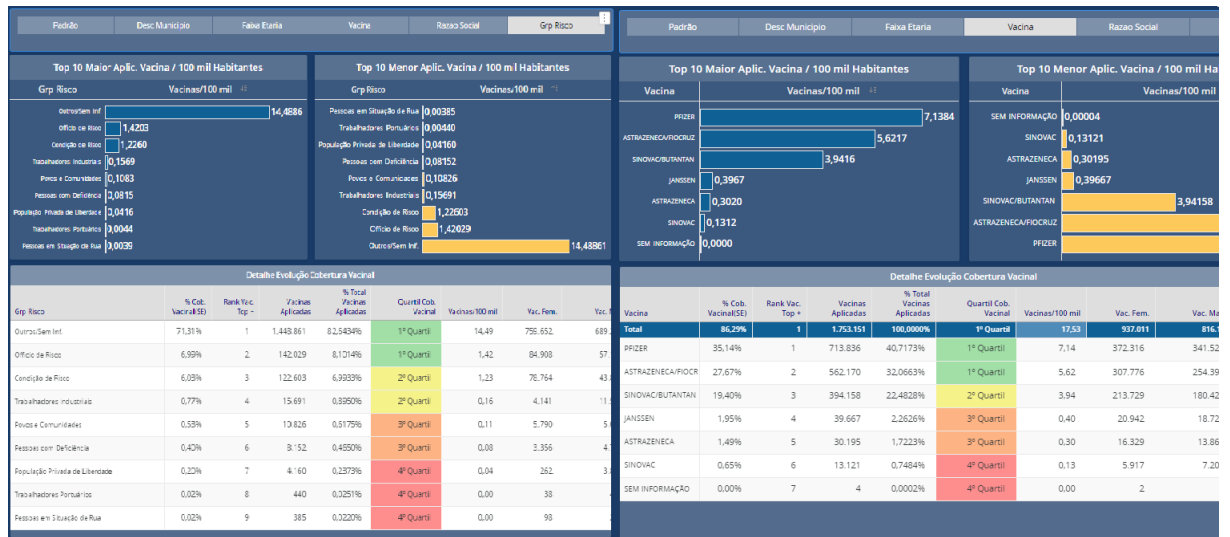
No Quadro Vacinal Estadual são apresentados filtros a nível de município, doses aplicadas, faixa etária, vacinas distribuídas, razão social do estabelecimento que aplicou as doses. Foram inseridos 2 gráficos, ranqueados por Top 10 com maior aplicação de vacinas a cada 100 mil habitantes e Top 10 com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

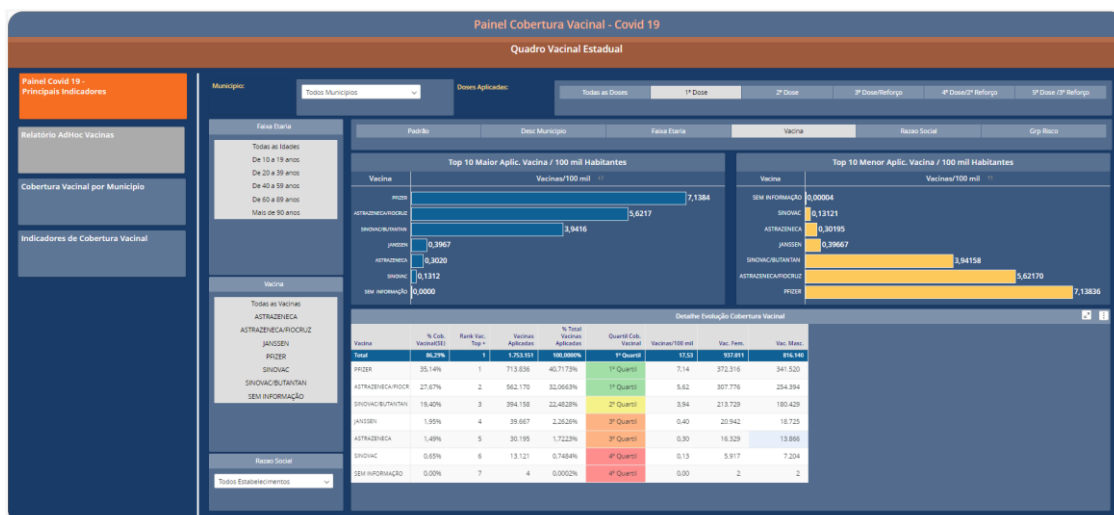
Figura 47 – Seletor de atributos da visão Quadro Vacinal Estadual.



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.



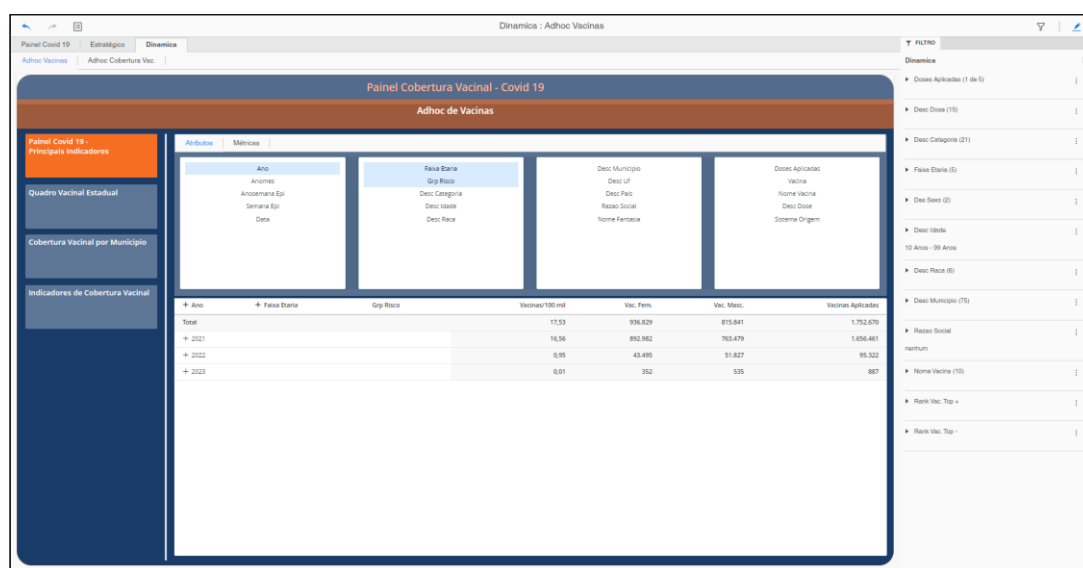
Fonte: Própria (2023).

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.

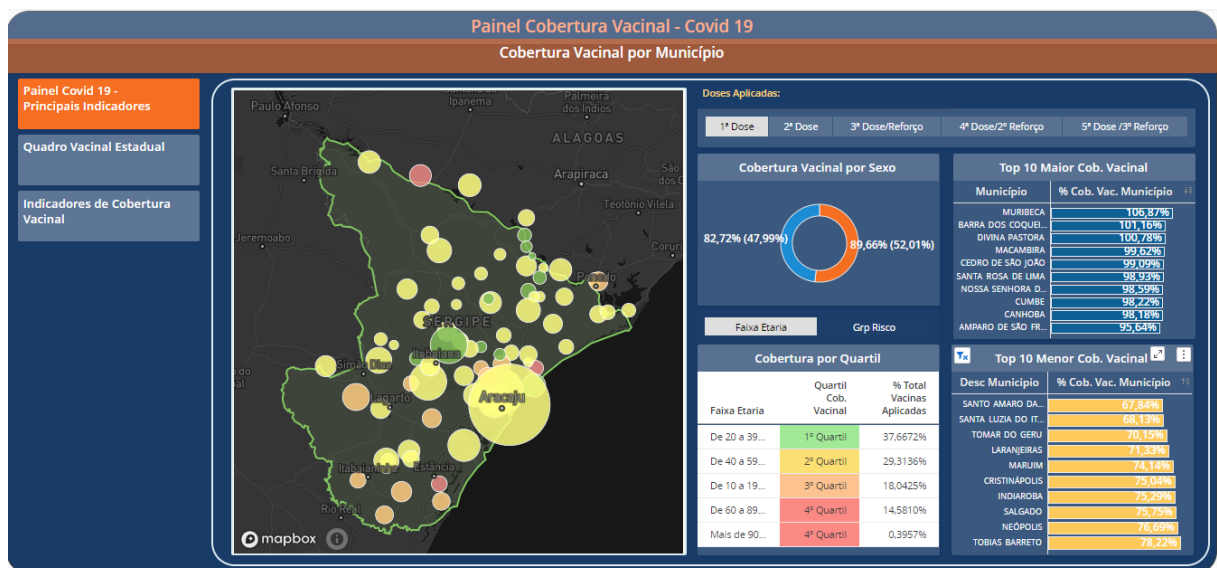


Fonte: Própria (2023).

A visão de cobertura vacinal por município torna-se uma opção a mais para complementar a análise, 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 aplicadas 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.

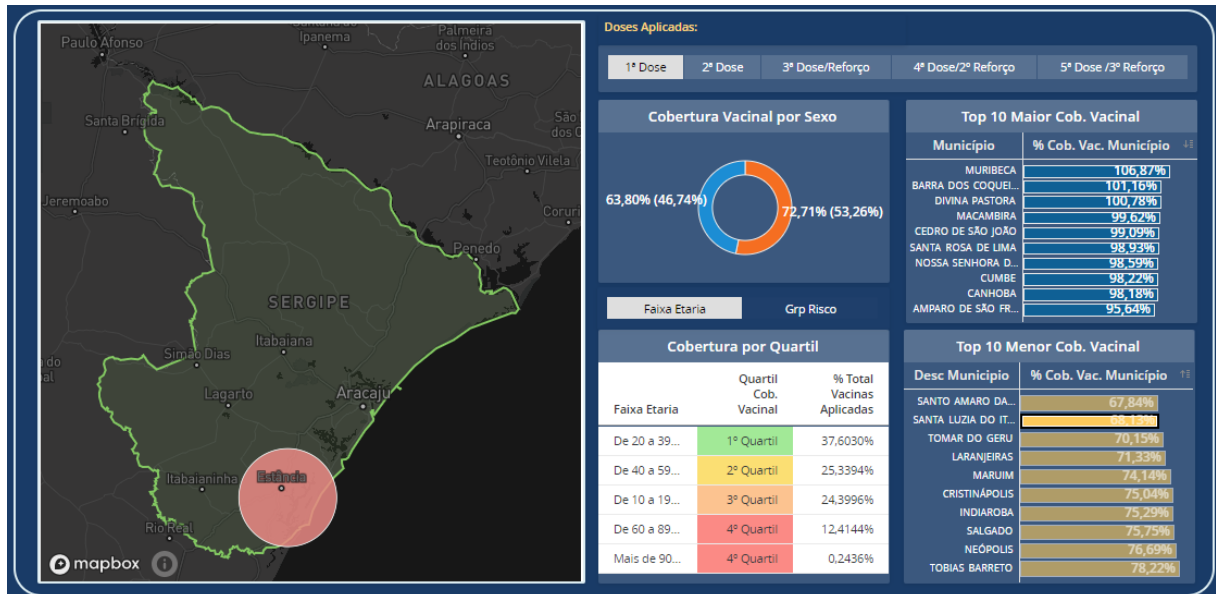
Figura 50 – Cobertura Vacinal por Município, apresenta um visual mais amigável para as análises e complementa os demais relatórios.



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.

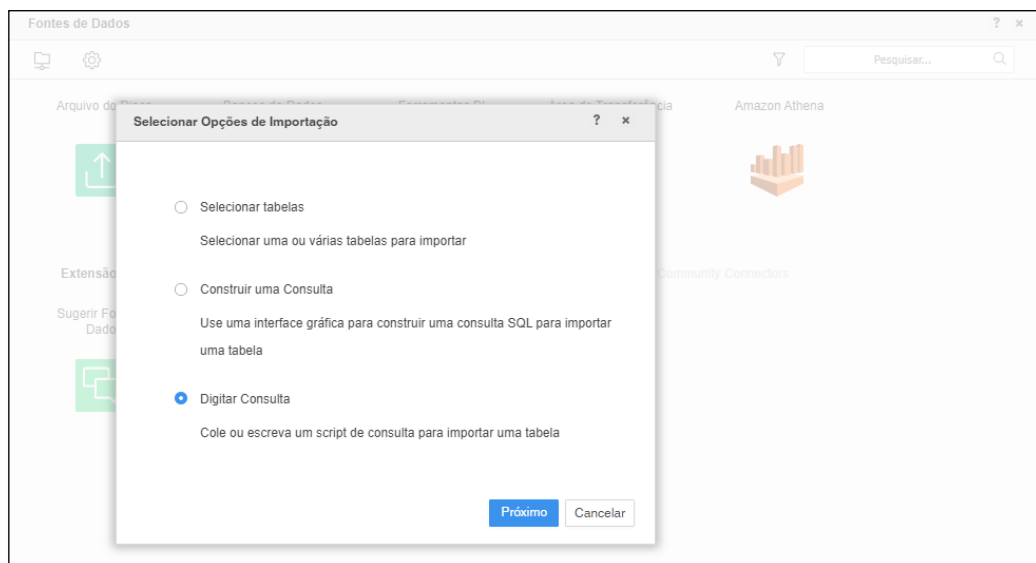


Fonte: Própria (2023).

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.

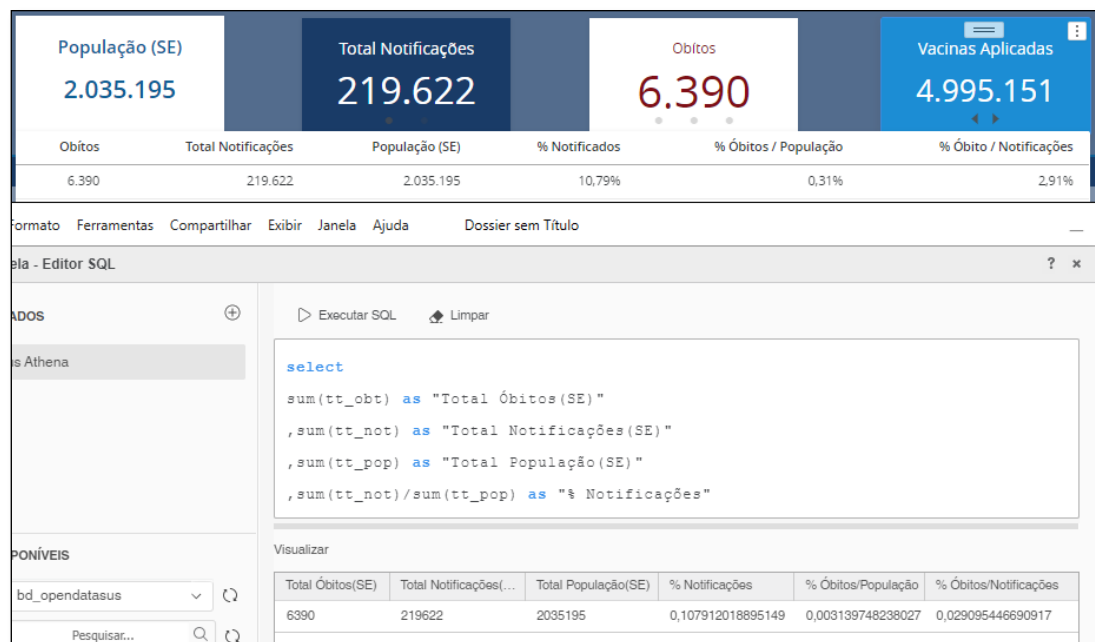
Figura 52 – Acesso a base de dados para geração de consultas em sql.



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.



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.

```

select
sum(tt_obt) as "Total Óbitos(SE)"
,sum(tt_not) as "Total Notificações(SE)"
,sum(tt_pop) as "Total População(SE)"
,sum(tt_not)/sum(tt_pop) as "% Notificações"
,sum(tt_obt)/sum(tt_pop) as "% Óbitos/População"
,sum(tt_obt)/sum(tt_not) as "% Óbitos/Notificações"
from(
select
n.cod_idade
,n.cod_sexo
,n.cod_municipio_paciente
,0 as tt_obt
,sum(n.qtd_notificacao) as tt_not
,0 as tt_pop
from bd_opendatasus.fato_notificacao n
group by
n.cod_idade
,n.cod_sexo
,n.cod_municipio_paciente
union
select
o.cod_idade
,o.cod_sexo
,o.cod_municipio_paciente
,sum(o.qtd_obito) as tt_obt
,0 as tt_not
,0 as tt_pop
from bd_opendatasus.fato_obitos o
group by
o.cod_idade
,o.cod_sexo
,o.cod_municipio_paciente
union
select
p.cod_idade
,p.cod_sexo
,p.cod_municipio_paciente
,0 as tt_obt
,0 as tt_not
,sum(p.qtd_populacao) as tt_pop
from bd_opendatasus.fato_populacao p
group by
p.cod_idade
,p.cod_sexo
,p.cod_municipio_paciente
) ;

```

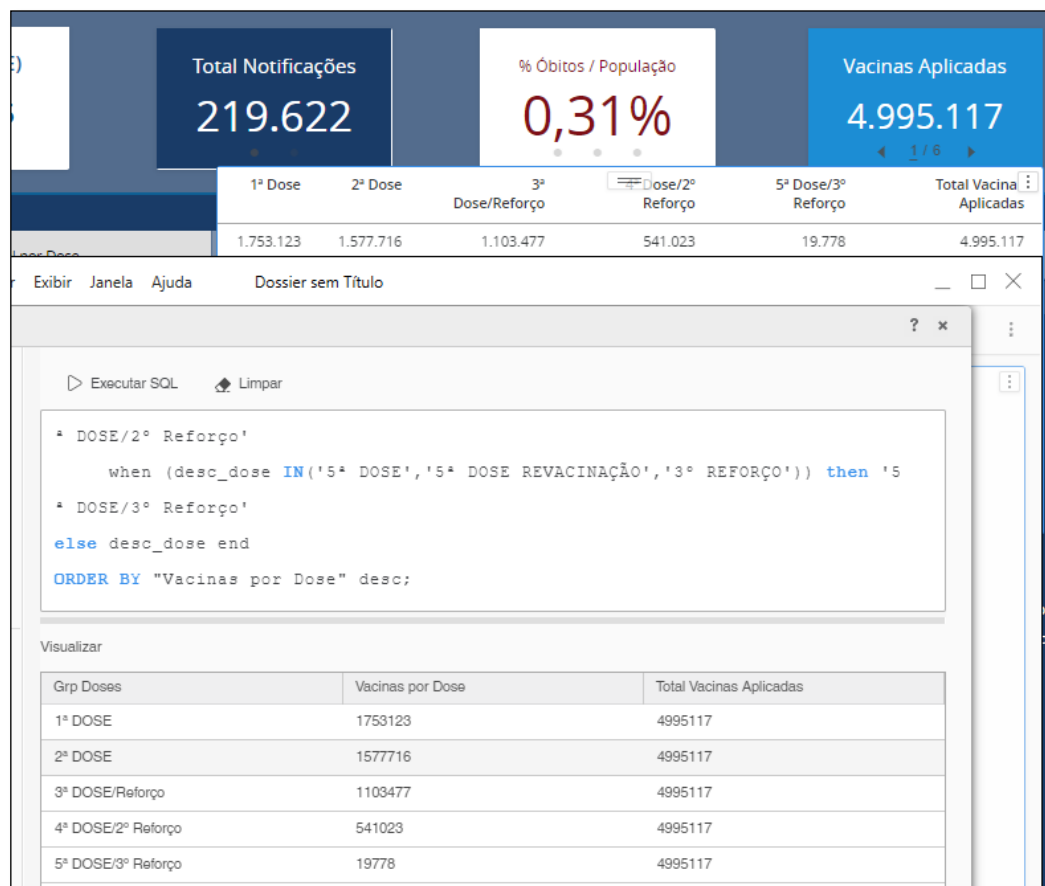
Fonte: Própria (2023).

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.



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.

```

select
case when (desc_dose IN('1ª DOSE','1ª DOSE REVACINAÇÃO','DOSE','REVACINAÇÃO')) then '1ª DOSE'
  when (desc_dose IN('2ª DOSE','2ª DOSE REVACINAÇÃO')) then '2ª DOSE'
  when (desc_dose IN('3ª DOSE','3ª DOSE REVACINAÇÃO','REFORÇO','DOSE ADICIONAL')) then '3ª
DOSE/Reforço'
  when (desc_dose IN('4ª DOSE','4ª DOSE REVACINAÇÃO','2º REFORÇO')) then '4ª DOSE/2º Reforço'
  when (desc_dose IN('5ª DOSE','5ª DOSE REVACINAÇÃO','3º REFORÇO')) then '5ª DOSE/3º Reforço'
else desc_dose end "Grp Doses",
sum(v.qtd_vacinas_aplicadas) as "Vacinas por Dose",
(select sum(qtd_vacinas_aplicadas) as total from bd_opendatasus.fato_vacinas) as "Total Vacinas Aplicadas"
from bd_opendatasus.fato_vacinas v
join bd_opendatasus.dm_doses d
on v.cod_dose = d.cod_dose
group by
case when (desc_dose IN('1ª DOSE','1ª DOSE REVACINAÇÃO','DOSE','REVACINAÇÃO')) then '1ª DOSE'
  when (desc_dose IN('2ª DOSE','2ª DOSE REVACINAÇÃO')) then '2ª DOSE'
  when (desc_dose IN('3ª DOSE','3ª DOSE REVACINAÇÃO','REFORÇO','DOSE ADICIONAL')) then '3ª
DOSE/Reforço'
  when (desc_dose IN('4ª DOSE','4ª DOSE REVACINAÇÃO','2º REFORÇO')) then '4ª DOSE/2º Reforço'
  when (desc_dose IN('5ª DOSE','5ª DOSE REVACINAÇÃO','3º REFORÇO')) then '5ª DOSE/3º Reforço'
else desc_dose end
ORDER BY "Vacinas por Dose" desc;

```

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, grávidas 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 a nível tático permitem os usuários avançarem nos detalhes dos resultados obtidos e com base nesses dados encontrar formas de atuar, atendendo os municípios 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.

Lista de atributos criados no Microstrategy

Foi necessário a criação de novos atributos para uso na construção das análises e relatórios, listado na figura 57.

Figura 57 – Atributos criados no Microstrategy, para compor os gráficos.

Grupo de Atributos Criados	Fórmulas	Objetivo
Doses Aplicadas	Grupo de Atributos	Agrupar por descrição as doses de vacina
Grp Risco	Grupo de Atributos	Agrupar por categoria em grupos de risco
Município	[Desc Município]@ID ; Latitude@ID ; Longitude@ID	Inclusão de Longitude e Latitude ao Município para Mapeamento
Vacina	Replace(IF([Nome Vacina]@ID="SEM INFORMACAO");"SEM INFORMAÇÃO";SubStr([Nome Vacina]@ID;9;(Position(" - ";[Nome Vacina]@ID)-9)));;"PEDIÁTRICA";"PFIZER")	Ajuste da Descrição da Vacina

Fonte: Própria (2023).

Lista de métricas criadas no Microstrategy

Abaixo todas as métricas criadas para utilização nos gráficos e relatórios, com detalhamento das fórmulas e objetivos das métricas, como listado nas figuras 58 e 59.

Figura 58 – Métricas criadas no Microstrategy, para compor os gráficos parte 1.

Grupo de métricas criadas	Fórmulas	Objetivo
Vacinas/100 mil	[Vacinas Aplicadas]/100000	Vacinas aplicadas a cada 100 mil habitantes
Vacinas Aplicadas	Sum([Qtd Vacinas Aplicadas]){~+}	Soma de qtd_vacinas_aplicadas, nivelado por visualização
Vac. Sexo	Sum([Vacinas Aplicadas])([Des Sexo]+, [Desc Municipio]+, [Doses Aplicadas]+)	Vacinas aplicadas nivelado por sexo, município e doses aplicadas
Vac. Município	Sum([Qtd Vacinas Aplicadas])([Cod Idade]+, [Cod Sexo]+, [Desc Municipio]+, [Doses Aplicadas]+, Municipio+)	Vacinas aplicadas nivelado por município e doses aplicadas, idade, sexo
Vac. Masc.	Sum([Qtd Vacinas Aplicadas]){~+}< ;@2;->	Vacinas aplicadas , filtrado por sexo Masculino
Vac. Fem.	Sum([Qtd Vacinas Aplicadas]){~+}< ;@2;->	Vacinas aplicadas , filtrado por sexo Feminino
Total Vacinas Aplicadas	Sum([Qtd Vacinas Aplicadas])(@p+)	Total Geral Vacinas aplicadas
Total Notificações	Sum([Qtd Notificacao])(@p+)	Total Geral Notificações
Rank Vac. Top +	Rank<ASC=False, BreakBy={@auto}>([Vacinas Aplicadas])	Rank de vacinação aplicadas descendente
Rank Vac. Top -	Rank<BreakBy={@auto}>([Vacinas/100 mil])	Rank de vacinação aplicadas ascendente
Rank % Cob. Top +	Rank<ASC=False, BreakBy={@auto}>([% Cob. Vac. Municipio])	Rank de % Cobertura Vacinal descendente
Rank % Cob Top -	Rank<BreakBy={@auto}>([% Cob. Vac. Municipio])	Rank de % Cobertura Vacinal ascendente
Quartil Cob. Vacinal	Case([([Percentil Vacinal]>=0) And ([Percentil Vacinal]<=0,25));4;([([Percentil Vacinal]>0,25) And ([Percentil Vacinal]<=0,5));3;([([Percentil Vacinal]>0,5) And ([Percentil Vacinal]<=0,75));2;1)	Quartil do percentil de Vacinas Aplicadas
População (SE)	Sum([Distinct=True]>([Qtd Populacao Se])(@p+)	Total geral População
Pop. Sexo	Sum([Qtd Populacao])([Des Sexo]+, [Desc Municipio]+)	População nivelado por sexo e município
Pop. Município	Sum([Qtd Populacao])([Cod Idade]+, [Cod Sexo]+, [Desc Municipio]+, Municipio+)	População nivelado por município, sexo, idade
Pop. Mulheres	Sum([Qtd Populacao])([Desc Municipio]+, [Faixa Etaria]+, [Des Sexo]+)< ;@2;->	População nivelado por sexo, município, faixa etária e filtrado pelo sexo feminino
Pop. Idade	Sum([Qtd Populacao])([Cod Idade]+, [Cod Sexo]+, [Desc Municipio]+, [Faixa Etaria]+)	População nivelado por Idade, município, faixa etária
Pop. Homens	Sum([Qtd Populacao])([Cod Idade]+, [Cod Sexo]+, [Desc Municipio]+, [Faixa Etaria]+)< ;@2;->	População nivelado por sexo, município, faixa etária e filtrado pelo sexo masculino
Percentil Vacinal	PercentRank<BreakBy={@auto}>([Vacinas Aplicadas])	Percentil de vacinas aplicadas
Óbitos	Sum([Qtd Obito])(~+)	Soma de qtd_obito
Notificações Anomes	Sum([Qtd Notificacao])(Anomes+)	Notificações por anomes
MM (Vacinados)	ExpWghMovingAvg<BreakBy={Data}, SortBy={([Anosemana Epi]@ID)>([Qtd Vacinas Aplicadas];7;1)	Média Móvel de vacinas aplicadas, quebrada por data, 7 dias anteriores e classificado por anosemana epidemiologica
MM (Óbitos)	ExpWghMovingAvg<BreakBy={Data}, SortBy={([Anosemana Epi]@ID)>([Qtd Obito];7;1)	Média Móvel de óbitos, quebrada por data, 7 dias anteriores e classificado por anosemana epidemiologica
MM (Notificações)	ExpWghMovingAvg<BreakBy={Data}, SortBy={([Anosemana Epi]@ID)>([Qtd Notificacao];7;1)	Média Móvel de notificação, quebrada por data, 7 dias anteriores e classificado por anosemana epidemiologica
MM (5ª Dose/3º Reforço)	ExpWghMovingAvg<BreakBy={Data}, SortBy={([Anosemana Epi]@ID)>([5ª Dose/3º Reforço];7;1)	Média Móvel de 5ª dose, quebrada por data, 7 dias anteriores e classificado por anosemana epidemiologica
MM (4ª Dose/2º Reforço)	ExpWghMovingAvg<BreakBy={Data}, SortBy={([Anosemana Epi]@ID)>([4ª Dose/2º Reforço];7;1)	Média Móvel de 4ª dose, quebrada por data, 7 dias anteriores e classificado por anosemana epidemiologica
MM (3ª Dose/Reforço)	ExpWghMovingAvg<BreakBy={Data}, SortBy={([Anosemana Epi]@ID)>([3ª Dose/Reforço];7;1)	Média Móvel de 3ª dose, quebrada por data, 7 dias anteriores e classificado por anosemana epidemiologica
MM (2ª Dose)	ExpWghMovingAvg<BreakBy={Data}, SortBy={([Anosemana Epi]@ID)>([2ª Dose];7;1)	Média Móvel de 2ª dose, quebrada por data, 7 dias anteriores e classificado por anosemana epidemiologica
MM (1ª Dose)	ExpWghMovingAvg<BreakBy={Data}, SortBy={([Anosemana Epi]@ID)>([1ª Dose];7;1)	Média Móvel de 1ª dose, quebrada por data, 7 dias anteriores e classificado por anosemana epidemiologica
Evol. Vac. Mulheres	OLAPSum<OLAPWinStType=0, OLAPWinEndType=2, BreakBy={([Faixa Etaria], [Desc Municipio], Vacina, [Cod Sexo]), SortBy={(@auto)>([Vac. Fem.]	Evolutivo vacinas no sexo feminino, por faixa etária, município, vacina, sexo
Evol. Vac. Idade	OLAPSum<OLAPWinStType=0, OLAPWinEndType=2, BreakBy={([Faixa Etaria], [Desc Municipio], [Cod Sexo], Vacina), SortBy={(@auto)>([Qtd Vacinas Aplicadas])	Evolutivo vacinas por idade, por faixa etária, município, vacina, sexo
Evol. Vac. Homens	OLAPSum<OLAPWinStType=0, OLAPWinEndType=2, BreakBy={([Faixa Etaria], [Desc Municipio], Vacina, [Cod Sexo]), SortBy={(@auto)>([Vac. Masc.]	Evolutivo vacinas no sexo masculino, por faixa etária, município, vacina, sexo
5ª Dose/3º Reforço	Sum([Qtd Vacinas Aplicadas])(~+)< ;@2;->	Vacinas aplicadas filtrada pela 5ª Dose
4ª Dose/2º Reforço	Sum([Qtd Vacinas Aplicadas])(~+)< ;@2;->	Vacinas aplicadas filtrada pela 4ª Dose
3ª Dose/Reforço	Sum([Qtd Vacinas Aplicadas])(~+)< ;@2;->	Vacinas aplicadas filtrada pela 3ª Dose
2ª Dose	Sum([Qtd Vacinas Aplicadas])(~+)< ;@2;->	Vacinas aplicadas filtrada pela 2ª Dose
1ª Dose	Sum([Qtd Vacinas Aplicadas])(~+)< ;@2;->	Vacinas aplicadas filtrada pela 1ª Dose

Fonte: Própria (2023).

Figura 59 – Métricas criadas no Microstrategy, para compor os gráficos parte 2.

Grupo de métricas criadas	Fórmulas	Objetivo
% Total Vacinas Aplicadas	$\frac{[Vacinas\ Aplicadas]}{Sum<UseLookupForAttributes=False>([Vacinas\ Aplicadas])\{ @p \}}$	Percentual de vacinas pelo Total vacinas aplicadas
% Óbitos / População	$\frac{Obitos}{[População\ (SE)]}$	Percentual de obitos pela população
% Óbito / Notificações	$\frac{Obitos}{[Total\ Notificações]}$	Percentual de obito pela notificações
% Notificados	$\frac{[Notificações\ Anomes]}{[População\ (SE)]}$	Percentual de Notificações pela população
% Mês Anterior MM (Notificações)	$\frac{([MM\ (Notificações)] - Lag<BreakBy=\{ @auto \}, SortBy=(@auto)>([MM\ (Notificações)]); 1; 0])}{Lag<BreakBy=\{ @auto \}, SortBy=(@auto)>([MM\ (Notificações)]); 1; 0])}$	Percentual de Medias moveis em notificações pelas Médias moveis em notificações do mês anterior
% Mês Anterior (MM Vacinados)	$\frac{([MM\ (Vacinados)] - Lag<BreakBy=\{ @auto \}, SortBy=(@auto)>([MM\ (Vacinados)]); 1; 0])}{Lag<BreakBy=\{ @auto \}, SortBy=(@auto)>([MM\ (Vacinados)]); 1; 0])}$	Percentual de Medias moveis Vacinados em mm vacinados pelas Médias moveis em mm vacinados do mês anterior
% Mês Anterior (MM Óbitos)	$\frac{([MM\ (Óbitos)] - Lag<BreakBy=\{ @auto \}, SortBy=(@auto)>([MM\ (Óbitos)]); 1; 0])}{Lag<BreakBy=\{ @auto \}, SortBy=(@auto)>([MM\ (Óbitos)]); 1; 0])}$	Percentual de Medias moveis em Óbitos pelas Médias moveis em mm óbitos do mês anterior
% Cob. Vacinal(SE)	$\frac{Sum<UseLookupForAttributes=False>([Qtd\ Vacinas\ Aplicadas])\{ [Doses\ Aplicadas] \}}{[População\ (SE)]}$	Percentual Vacinas Aplicadas pela População total
% Cob. Vac. Município	$\frac{[Vac.\ Município]}{[Pop.\ Município]}$	Percentual Vacinas Aplicadas Municipal pela População do município
% Cob. Vac. Mulheres	$\frac{[Evol.\ Vac.\ Mulheres]}{[Pop.\ Mulheres]}$	Percentual Evolutivo Vacinas Aplicadas Mulheres pela População Feminina
% Cob. Vac. Idade	$\frac{[Evol.\ Vac.\ Idade]}{[Pop.\ Idade]}$	Percentual Evolutivo Vacinas Aplicadas Idade pela População por Idade
% Cob. Vac. Homens	$\frac{[Evol.\ Vac.\ Homens]}{[Pop.\ Homens]}$	Percentual Evolutivo Vacinas Aplicadas Homens pela População Masculina
% Cob Vac. Sexo	$\frac{[Vac.\ Sexo]}{[Pop.\ Sexo]}$	Percentual Evolutivo Vacinas Aplicadas Sexo pela População Sexo

Fonte: Própria (2023).

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_vWzBhpKM2nUYUOEVM21oCuPGItxs?usp=sharing>

Codigo do processo de ETL, diponivel em:
 <https://github.com/promarcosrt/TCC_Pucminas.git>

Vídeo de apresentação, disponível em: < <https://youtu.be/28LzNQQrRYE> > ou
 <https://drive.google.com/drive/folders/1hp-ZYAn_lmCcwjplrlBliErgQBIUjWzp?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, disponível 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.