

Análise dos Dados Ambulatoriais em Pernambuco

Alan Victor Da Silva Torres
av.torres@ufrpe.br

Fonte dos Dados: <https://datasus.saude.gov.br/>

Aplicação OLAP: <https://encurtador.com.br/2nap1>

Github: [AlannTorres/OLAP_DadosAmbulatorios \(github.com\)](https://github.com/AlannTorres/OLAP_DadosAmbulatorios)



ETAPA 1 - PLANEJAMENTO





1. Contextualização

Os dados se tornaram extremamente importantes na sociedade atual. Quando falamos de dados relacionados à saúde, essa importância se torna ainda mais evidente. No entanto, nem sempre esses dados são utilizados de maneira eficiente. Por isso, este projeto visa criar um data mart com informações ambulatoriais do estado de Pernambuco. O objetivo é obter insights relevantes, como identificar épocas com maior incidência de ocorrências e regiões com maiores necessidades. Essas informações são essenciais para aprimorar a gestão da saúde nos municípios pernambucanos.

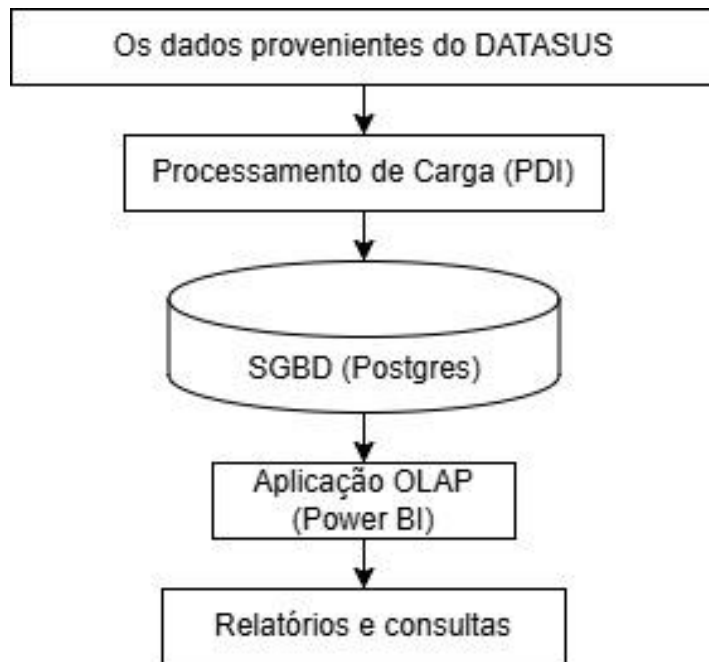


2. Escopo/objetivo do Data Mart

Como objetivo principal, este data mart busca criar insights que auxiliem na gestão estadual do sistema público de saúde, permitindo a formulação de estratégias mais eficientes para aprimorar a saúde pública no estado. Para isso, foram analisados dados ambulatoriais do estado de Pernambuco, gerados pelo DATASUS no período de 2020 a 2023.



3. Arquitetura Tecnológica





4. Processo

Para construir o Data Mart, seguiremos um processo estruturado. Primeiro, definiremos os requisitos provenientes dos dados obtidos no DATASUS. Em seguida, extraímos os dados relevantes para a análise, selecionando as informações específicas da região de Pernambuco. Realizaremos transformações para limpeza e padronização, como a exclusão de registros nulos, a fim de obter uma base de dados mais consistente. Após aplicar o processo ETL, os dados poderão ser carregados no Data Mart. Em seguida, utilizaremos esses dados na aplicação OLAP para gerar análises e métricas necessárias, avaliando os relatórios.



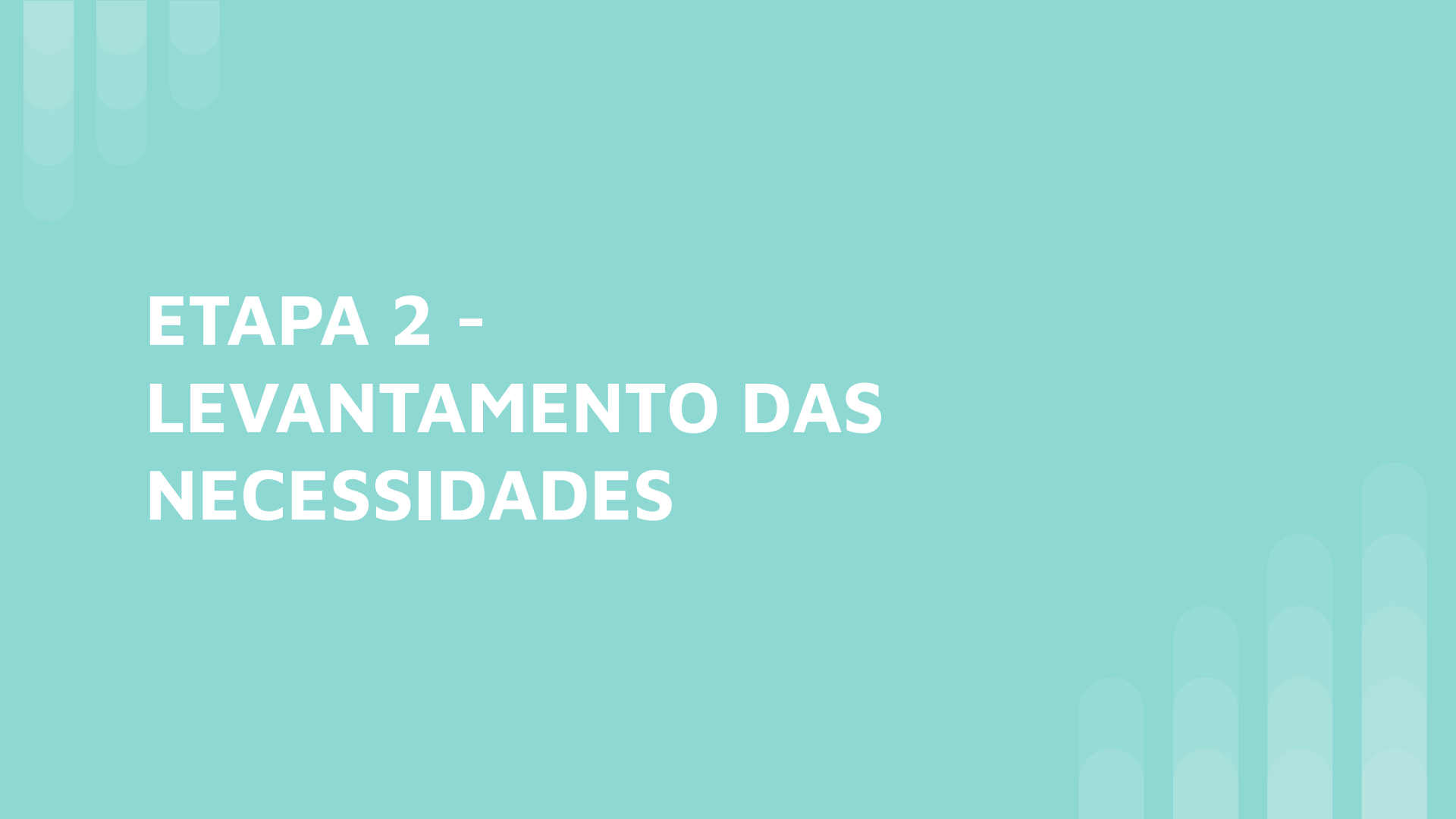
4. Abordagem

Para a abordagem, optamos pelo método Bottom-Up, proposto por Ralph Kimball, pois ele proporciona uma visão mais específica do estado de Pernambuco, com a possibilidade de expansão para uma perspectiva regional e nacional. Além disso, escolhemos o esquema “Star Schema” por permitir o uso de SGBDs, simplificando assim o processo de consultas. Esse esquema possibilita a criação de uma tabela de fatos central, contendo métricas relevantes, e várias tabelas de dimensões conectadas a ela. Essa estrutura facilita a análise dos dados e a geração de relatórios significativos para gestores e analistas.




5. Usuários

Dessa forma, esse projeto se torna útil tanto para gestores de saúde estadual, profissional de saúde dos municípios e gestores de políticas públicas, que podem utilizar esses dados para tomar decisões estratégicas, quanto para análises, que têm a oportunidade de explorar os dados e identificar tendências em diversos fatores.



ETAPA 2 - LEVANTAMENTO DAS NECESSIDADES





6. Consultas de Apoio à Decisão

Para esta etapa, realizou-se um estudo para identificar os indicadores mais relevantes para a gestão da saúde pública em documentos governamentais e trabalhos relacionados à vivência dos profissionais de saúde nos pontos de atendimento. Sendo exemplo de possíveis consultas:

- Quantidade de pacientes mulheres atendido no município de Recife.
- Quais CIDs mais atendido no período do 1 trimestre de 2020 em Serra Talhada
- Qual CID mais presente em pacientes femininos, acima de 50 anos.
- Qual município registra o maior número de atendimentos.



7. Indicadores do <negócio>

- Quantidade de procedimentos realizados, para medir o total de procedimentos realizados em um período específico. Ajuda a monitorar a carga de trabalho e eficiência operacional e avalia a produtividade e a capacidade de atendimento da organização.
- Número de pacientes atendidos, para avaliar a demanda e a carga de trabalho.
- Percentual de procedimentos por Sexo, Faixa Etária, Raça, para Identificar tendências de gênero, idade e possíveis necessidades específicas de atendimento. Além de ajudar a entender quais grupos etários são mais atendidos e pode orientar políticas de atendimento.

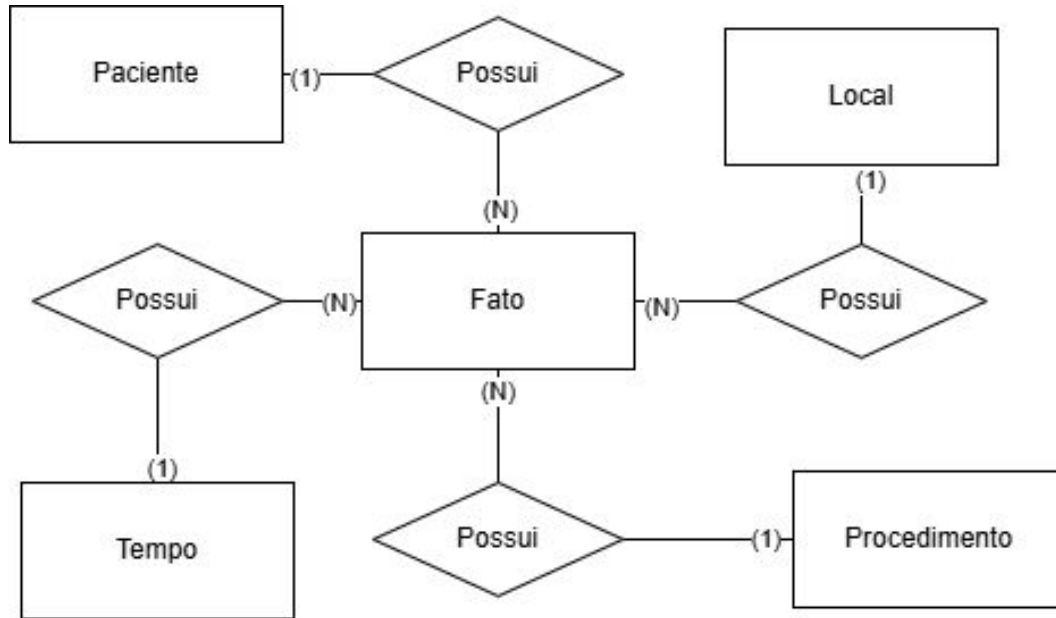
ETAPA 3 - MODELAGEM



9. Modelo Relacional

Para gerar o modelo relacional, foram identificadas quatro tabelas de dimensões essenciais. A **dim_local** captura as informações geográficas e de localização dos estabelecimentos de saúde, com o atributo principal sendo o `cod_municipio`. A **dim_tempo** cobre a dimensão temporal, permitindo análises ao longo do tempo, com atributos como `mes` e `ano`. A **dim_procedimento** detalha os procedimentos médicos realizados, categorizados por regras contratuais e códigos de doenças, com atributos como `cid` e `descrição`. Por fim, a **dim_paciente** abrange as características dos pacientes, com atributos como `idade`, `raça` e `sexo` e `faixa etaria`. Além disso, foi criada a tabela de fatos **fato_atendimento**, que armazena métricas quantitativas fundamentais para a geração de relatórios significativos.

9. Modelo Relacional





10. Modelo Dimensional

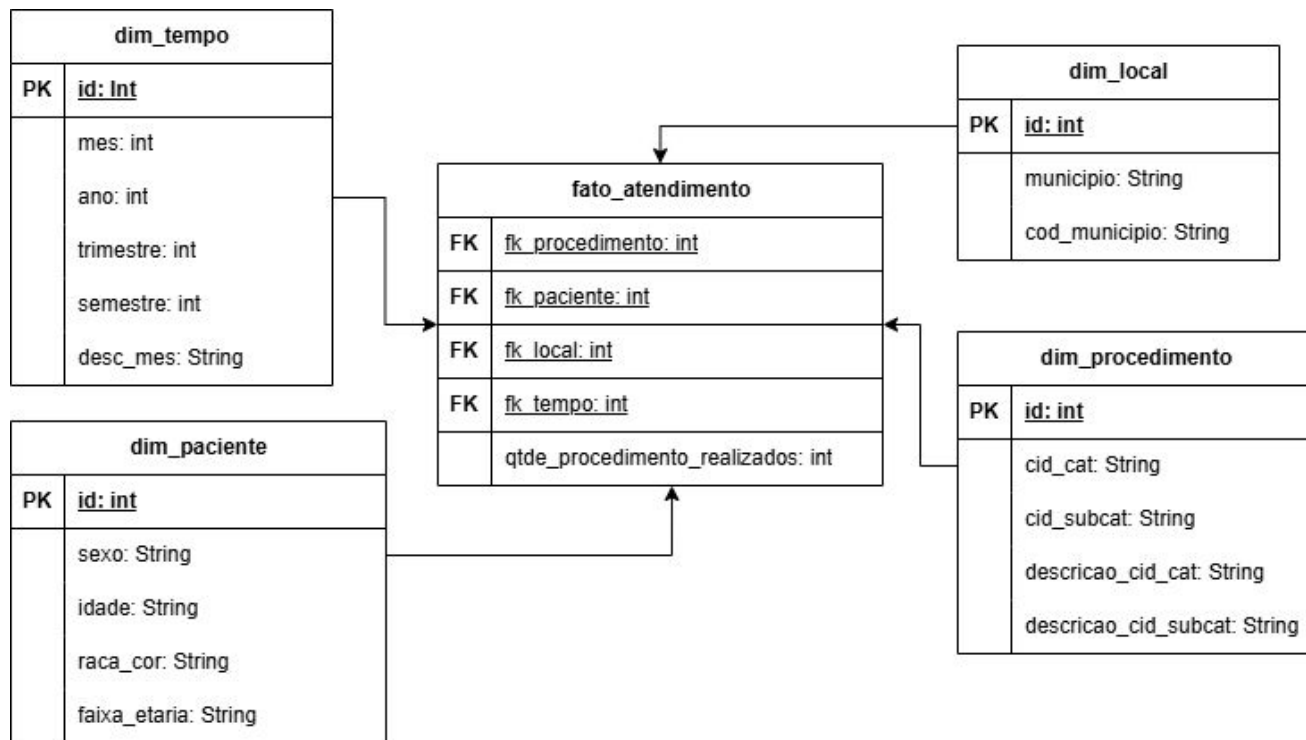
Esse Modelo Dimensional foi desenvolvido especificamente para a gestão de saúde pública, visando facilitar a tomada de decisões estratégicas e operacionais. A área de negócios engloba o gerenciamento de atendimentos e procedimentos médicos, com foco em otimizar os recursos e melhorar a qualidade dos serviços de saúde. O processo central é a análise granular dos atendimentos, onde cada fato registrado corresponde a uma interação única no sistema de saúde. A granularidade é definida no nível mais detalhado possível: o atendimento individual.



10. Modelo Dimensional

As hierarquias dentro das dimensões são claras na dimensão Tempo, a hierarquia é **ano > mês > trimestre > semestre** na Paciente, a hierarquia pode ser **idade > faixa etária**, procedimento **sendo categoria CID e Sub-categoria CID**. Além de outras combinações de características como sexo, etnia e raça. Essas estruturas permitem uma análise flexível e detalhada, suportando uma gestão eficiente e orientada a dados.

10. Modelo Dimensional



10. Modelo Dimensional



	id_tempo	id_local	id_procedimento	id_paciente	qtde_procedimentos_realizados
1	6	33	1	127	1
2	1	13	1	79	1
3	5	13	548	820	1
4	6	13	145	194	1
5	6	115	23	103	1
6	6	33	55	31	1
7	6	133	67	684	1



10. Modelo Dimensional do Data Mart (lógico)

Tabela Fato:

Quantidade de linhas da fato = 41.972.585 Linhas

Quantidade de colunas Colunas = 5 (Bytes por linha = 4 bytes)

$5 * 4 = 20 * 41.972.585 = 839.451.700$ bytes => **839,4 MB**

Tabelas Dimensões(25% da Fato) = 209,8 MB

Equivalente a: **$209,8 + 839,4 = 1,04$ GB**

ETAPA 4 - PROJETO FÍSICO DO BD



11. Modelo Relacional do Data Mart (físico)

```
CREATE TABLE dim_tempo (  
  id SERIAL PRIMARY KEY,  
  mes INT,  
  ano INT,  
  trimestre INT,  
  semestre INT,  
  descricao_mes VARCHAR(255),  
  UNIQUE (mes, ano),  
  CONSTRAINT unique_mes_ano  
UNIQUE (mes, ano));
```

```
CREATE TABLE dim_local (  
  id SERIAL PRIMARY KEY,  
  cod_municipio VARCHAR(255)  
UNIQUE,  
  municipio VARCHAR(255));
```

```
CREATE TABLE dim_paciente (  
  id SERIAL PRIMARY KEY,  
  idade VARCHAR(255),  
  sexo VARCHAR(255),  
  raca_cor VARCHAR(255),  
  faixa_etaria VARCHAR(255),  
  CONSTRAINT unique_combination  
UNIQUE (idade, sexo, raca_cor,  
  faixa_etaria));
```

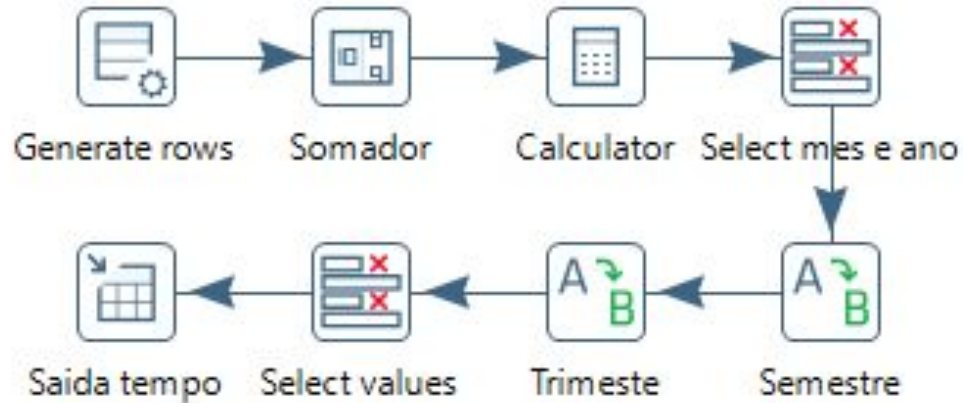
```
CREATE TABLE dim_procedimento (  
  id SERIAL PRIMARY KEY,  
  descricao_cid_cat TEXT,  
  descricao_cid_subcat TEXT,  
  cid_cat VARCHAR(255),  
  cid_subcat VARCHAR(255),  
  CONSTRAINT  
  unique_cid_cat_cid_subcat UNIQUE  
  (cid_cat, cid_subcat));
```

```
CREATE TABLE dim_fato_atendimento (  
  id_tempo INT,  
  id_local INT,  
  id_procedimento INT,  
  id_paciente INT,  
  qtde_procedimentos_realizados INT,  
  CONSTRAINT fk_id_tempo FOREIGN KEY (id_tempo)  
REFERENCES dim_tempo(id),  
  CONSTRAINT fk_id_local FOREIGN KEY (id_local)  
REFERENCES dim_local(id),  
  CONSTRAINT fk_id_procedimento FOREIGN KEY  
  (id_procedimento) REFERENCES dim_procedimento(id),  
  CONSTRAINT fk_id_paciente FOREIGN KEY  
  (id_paciente) REFERENCES dim_paciente(id),  
  CONSTRAINT unique_combination_fato UNIQUE  
  (id_tempo, id_local, id_procedimento, id_paciente));
```

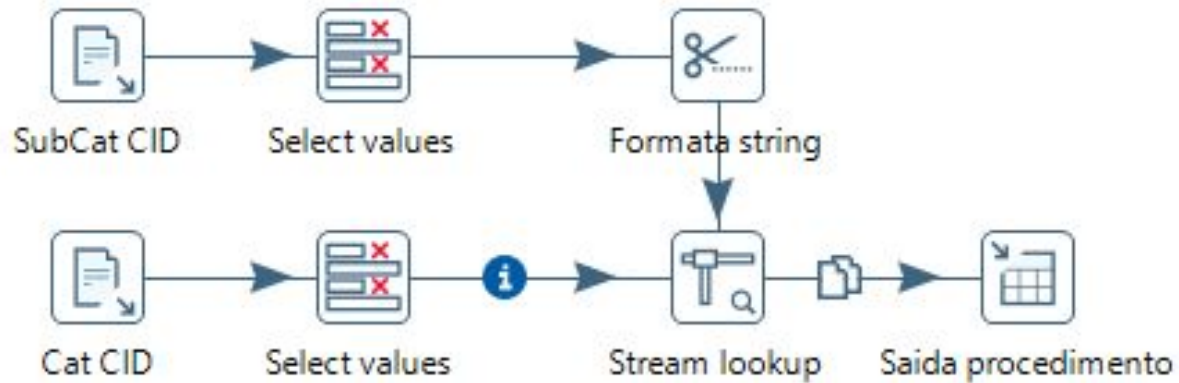


ETAPA 5 - EXTRAÇÃO, TRANSFORMAÇÃO E CARGA

12. Plano de Carga da Dimensão Tempo



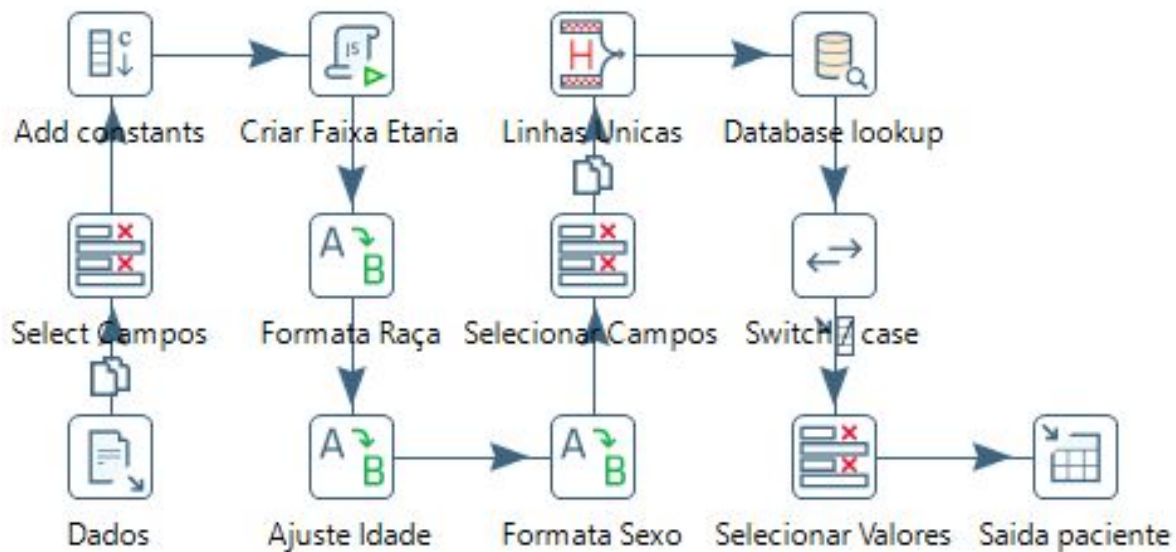
13. Plano de Carga da Dimensão Procedimento



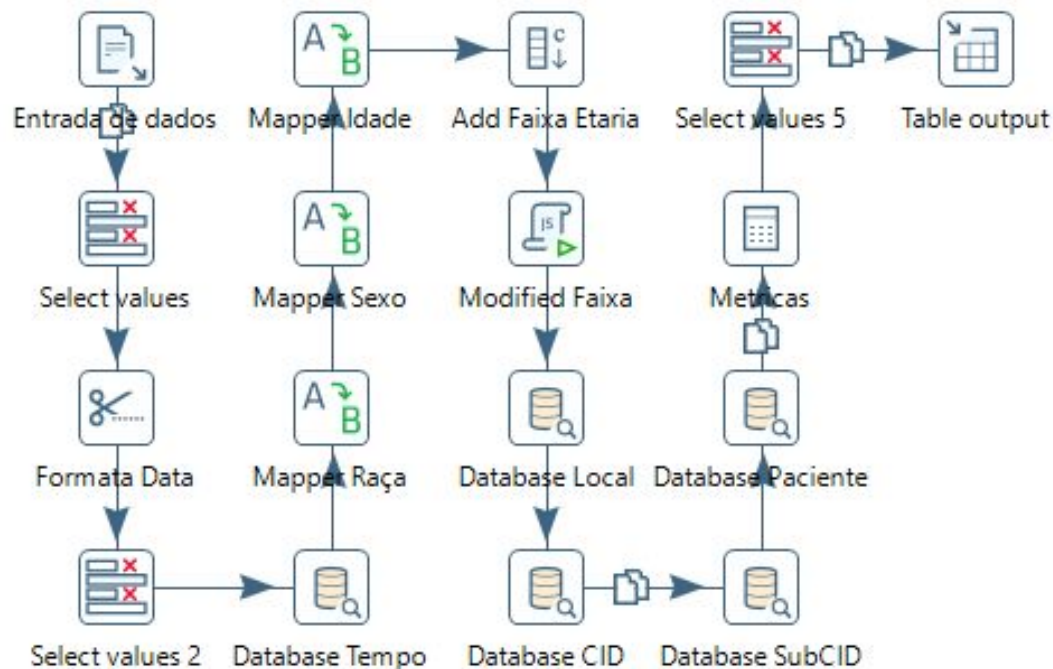
14. Plano de Carga da Dimensão Local



14. Plano de Carga da Dimensão Paciente



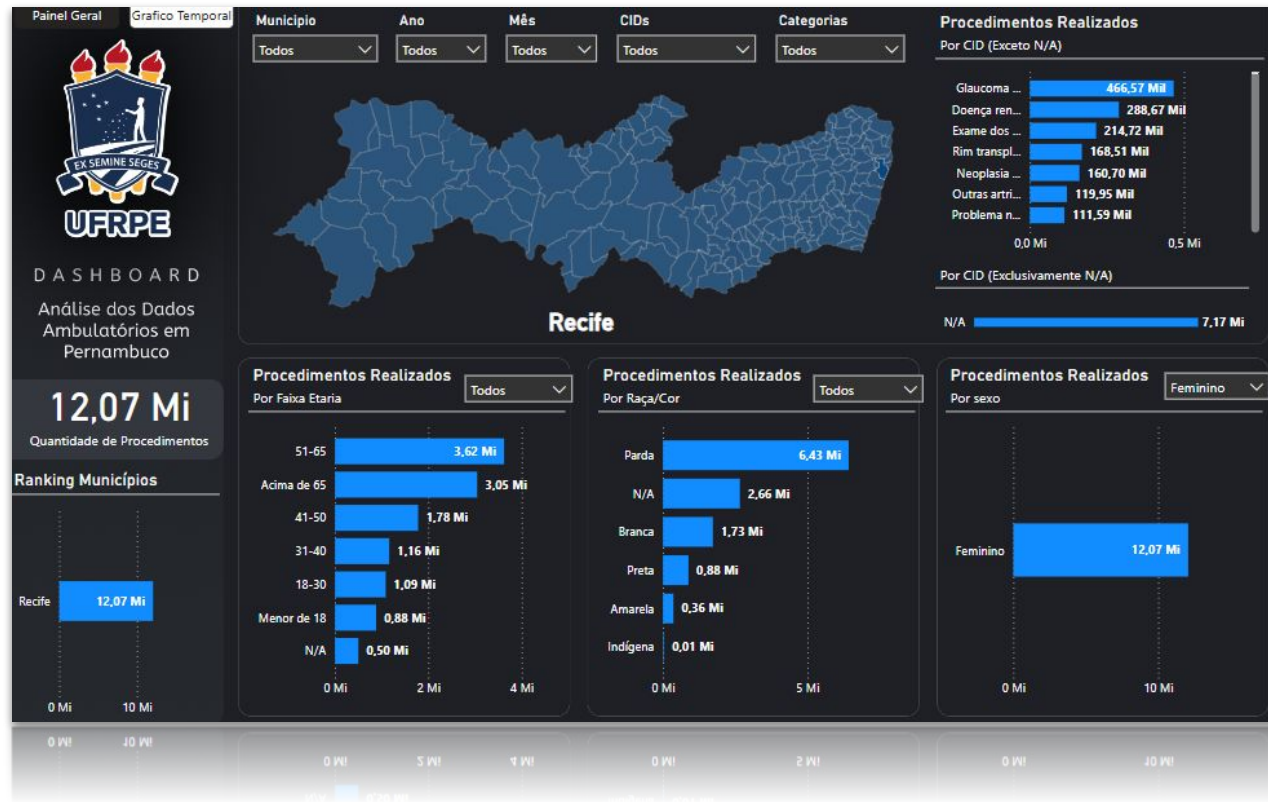
15. Plano de Carga da Fato



ETAPA 6 - APLICAÇÃO OLAP e PAINEL DE BORDO

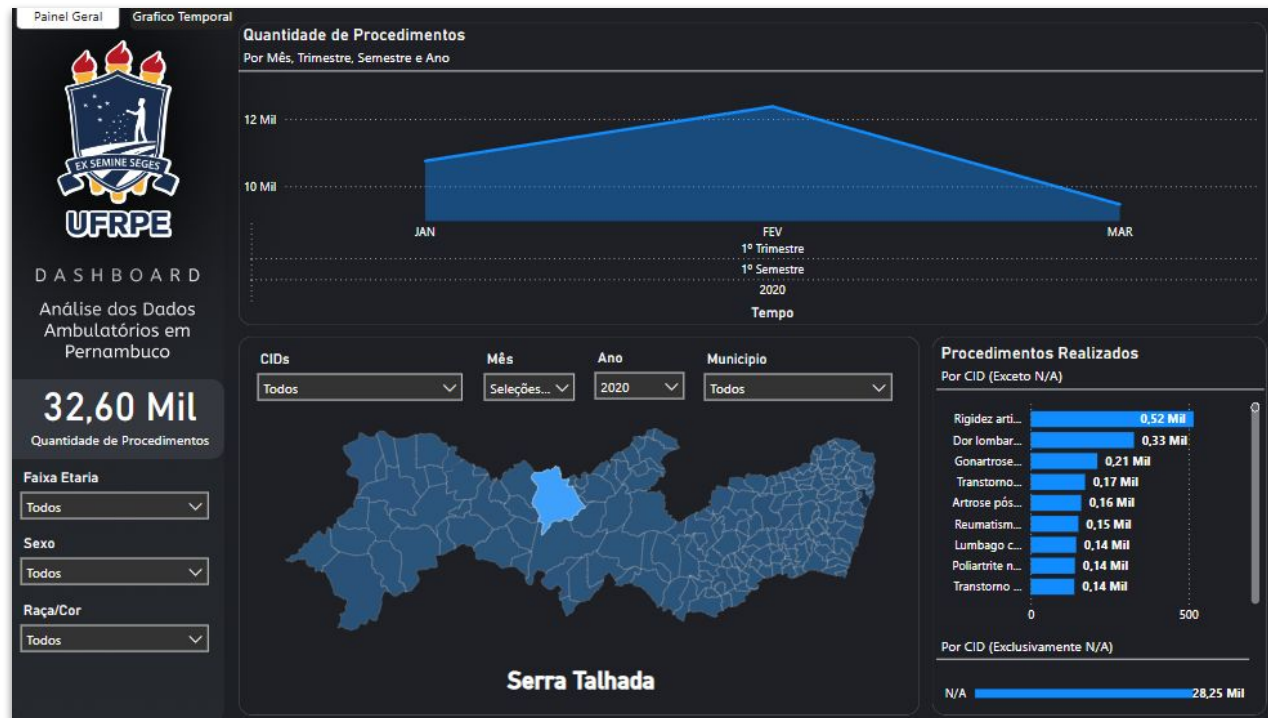
16. Consulta OLAP 1

Quantidade de
pacientes mulheres
atendido no município
de recife?



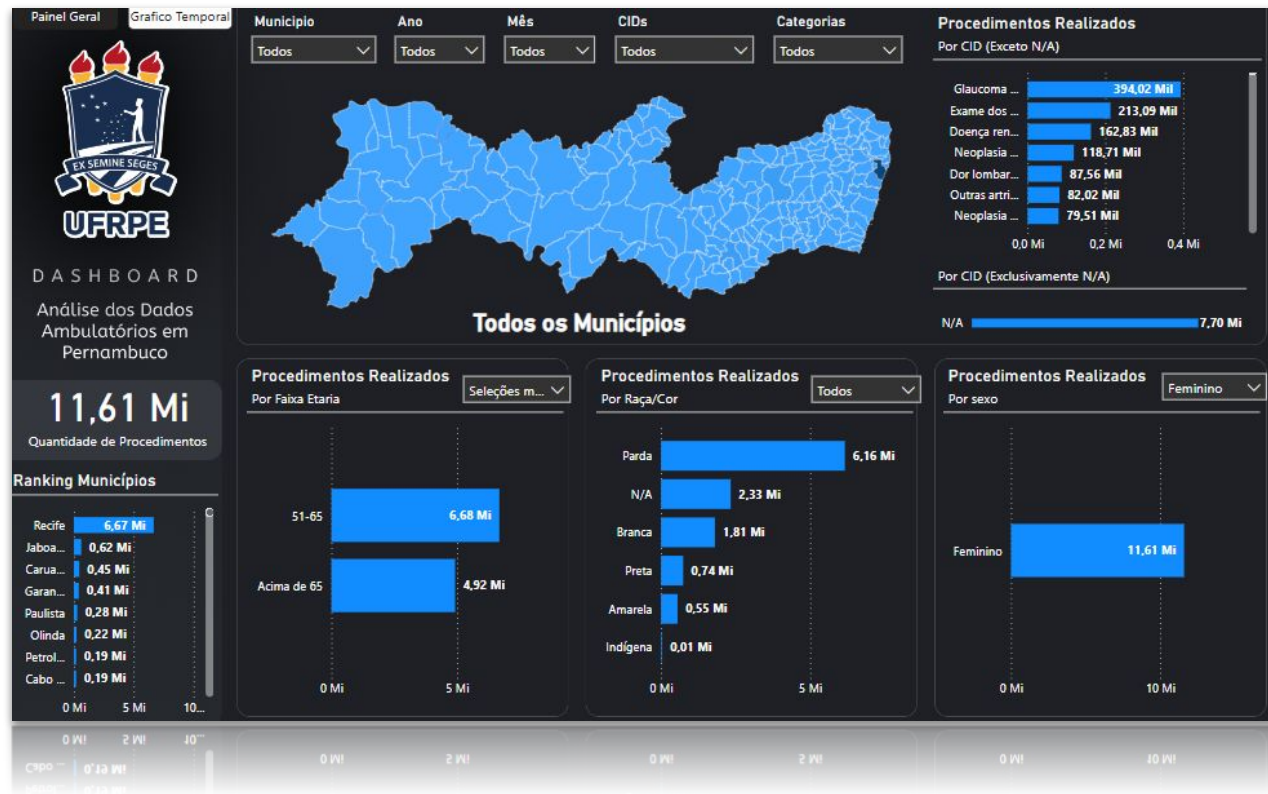
16. Consulta OLAP 2

Quais CIDs mais atendido no período do 1 trimestre de 2020 em Serra Talhada?



16. Consulta OLAP 3

Qual CID mais presente
em pacientes
femininos, acima de 50
anos.





Referências

BRASIL. Ministério da Saúde. DATASUS. Transferência de Arquivos. Disponível em: <https://datasus.saude.gov.br/transferencia-de-arquivos/>. Acesso em: 28 set. 2024.