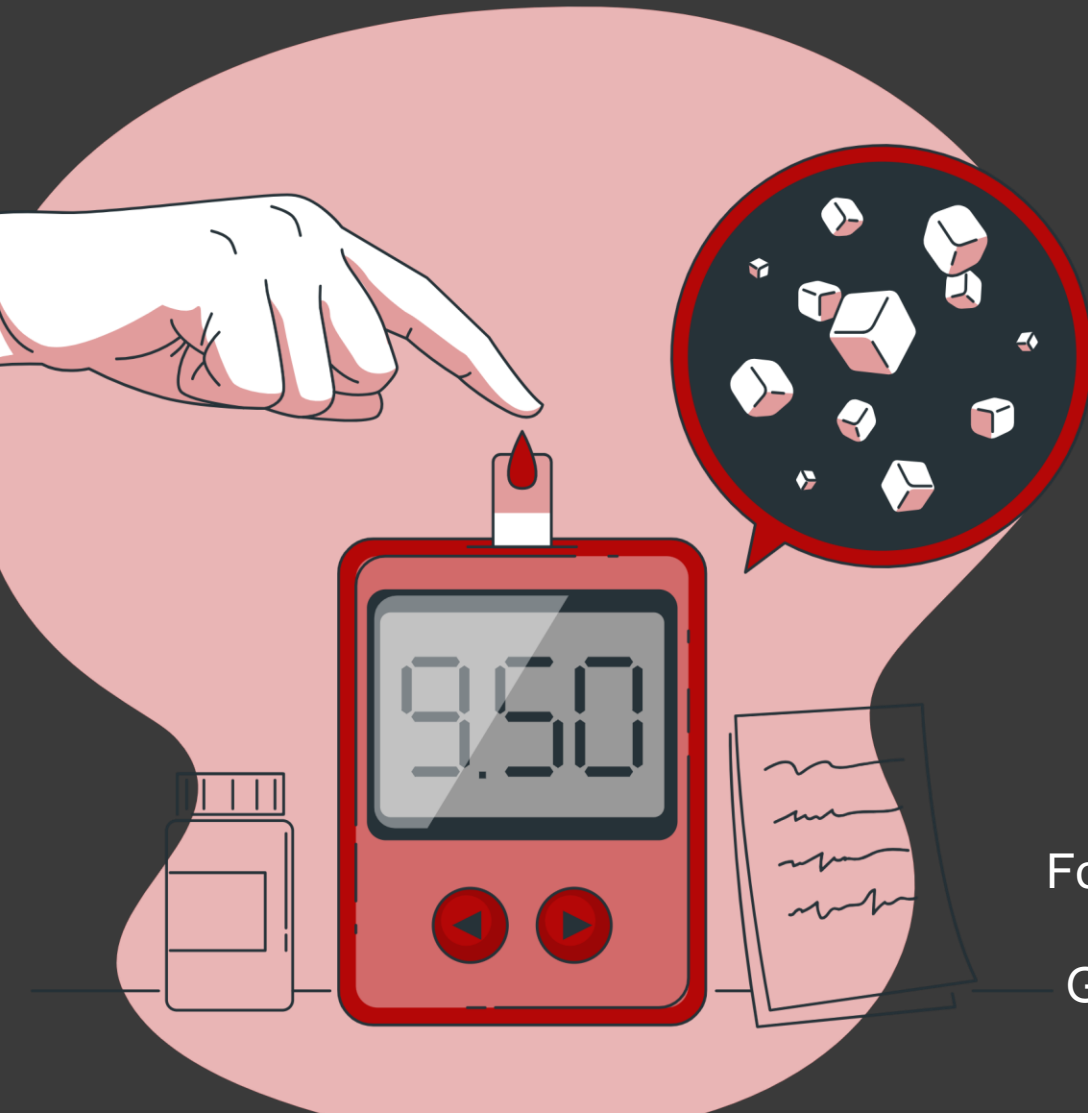


Análise Diabetes

Matheus Caldas Carvalho
matheus.ccarvalho@ufrpe.br



Fonte dos Dados: <<https://diabetesatlas.org/data/en/indicators/21/>>

Aplicação OLAP: <<http://bit.ly/3qKqgoF>>

Github: <<https://github.com/matheuscarvalho16/diabetes-data-mart>>

1

Etapa 1 - Planejamento

2

3

4

5

6

7



1 Contextualização

1

O Diabetes Mellitus (DM) é uma síndrome do metabolismo, de origem múltipla, decorrente da falta de insulina e/ou da incapacidade de a insulina exercer adequadamente seus efeitos.

2

3

De acordo com o Atlas do Diabetes da Federação Internacional de Diabetes (IDF) o Brasil é o 5º país em incidência de diabetes no mundo, com 16,8 milhões de doentes adultos (20 a 79 anos).

4

5

Os tipos mais conhecidos de Diabetes são: Tipo 1 causado pela destruição das células produtoras de insulina, em decorrência de defeito do sistema imunológico em que os anticorpos atacam as células que produzem a insulina.

6

Tipo 2 que resulta da resistência à insulina e de deficiência na sua secreção.

7

Diabetes Gestacional que é a diminuição da tolerância à glicose, diagnosticada pela primeira vez na gestação, podendo ou não persistir após o parto.

2 Escopo/objetivo do Data Mart

1

2

3

4

5

6

7

O objetivo do Data Mart a ser desenvolvido é fornecer uma análise detalhada da incidência de diabetes em diferentes países e faixas etárias ao longo do tempo. O Data Mart permitirá extrair insights valiosos sobre as tendências de prevalência da doença, identificar fatores de risco associados e avaliar a eficácia de medidas de prevenção e controle.

Os dados a serem analisados serão os Países, a faixa etária e a quantidade de pessoas com diabetes. E o período de análise são os anos 2000, 2011 e 2021.



3 Arquitetura Tecnológica

1

2

3

4

5

6

7

Fonte dos dados



ETL



Pentaho

Data Mart



PostgreSQL

Aplicação OLAP



Power BI

4 Processo

1

2

3

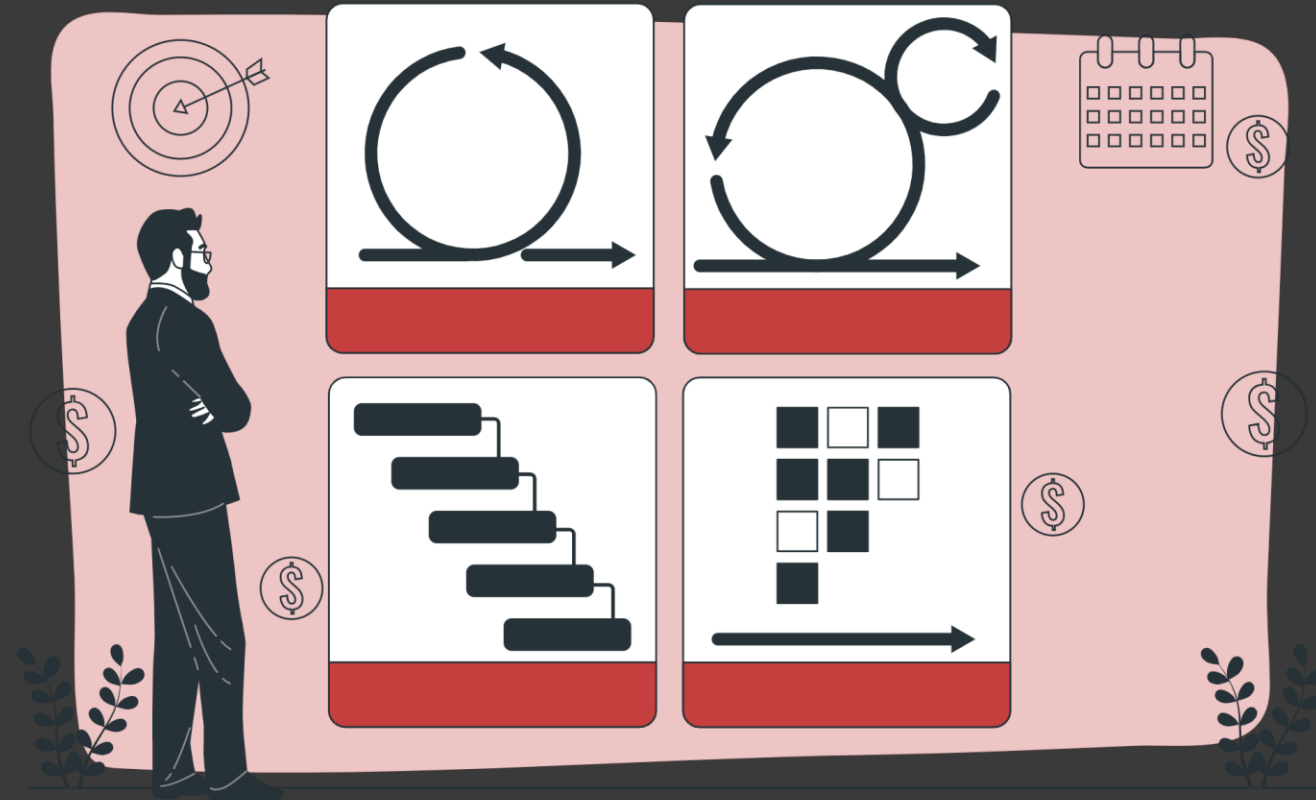
4

5

6

7

- Planejamento
- Levantamento das Necessidades
- Modelagem dimensional
- Projeto Físico do BD
- Projeto ETC
- Aplicação OLAP



5 Abordagem

1

2

3

4

5

6

7

Para o desenvolvimento deste Data Mart foi utilizada a abordagem Bottom-up, pois ela é uma abordagem evolutiva, com disponibilização imediata dos produtos, e por se tratar de um Data Mart específico, focando de partes para o todo, usa-se a abordagem Bottom-up.

E foi utilizado o “Star Schema”, pois é um estilo incremental do Bottom-up e é focada na construção de Data Mart, pois é uma abordagem descentralizada do conceito de DW, trabalhando na construção de tabelas dimensões e fatos.

6 Usuários

1

2

3

4

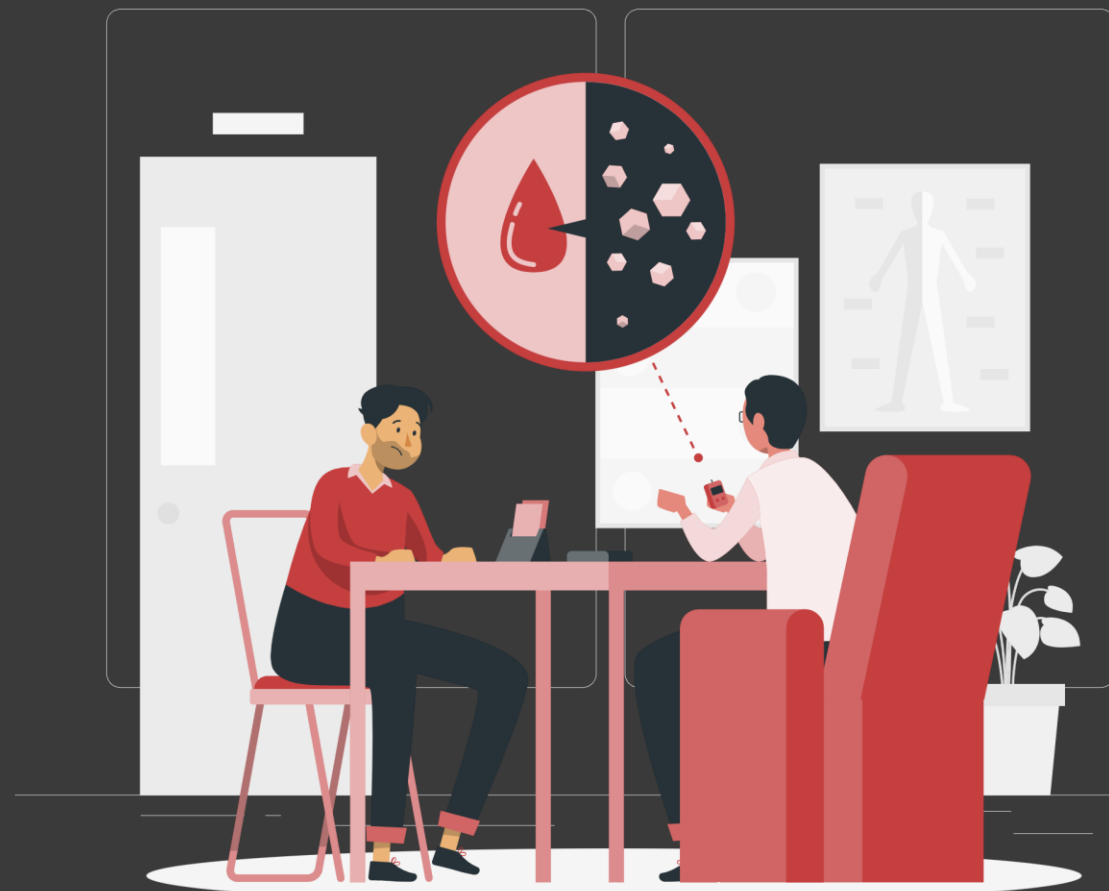
5

6

7

Os possíveis usuários para esse Data Mart são:

- Órgãos governamentais
- Pesquisadores e acadêmicos
- Profissionais de saúde
- Organizações de Saúde
- Diabéticos



1

2

Etapa 2 – Levantamento das Necessidades

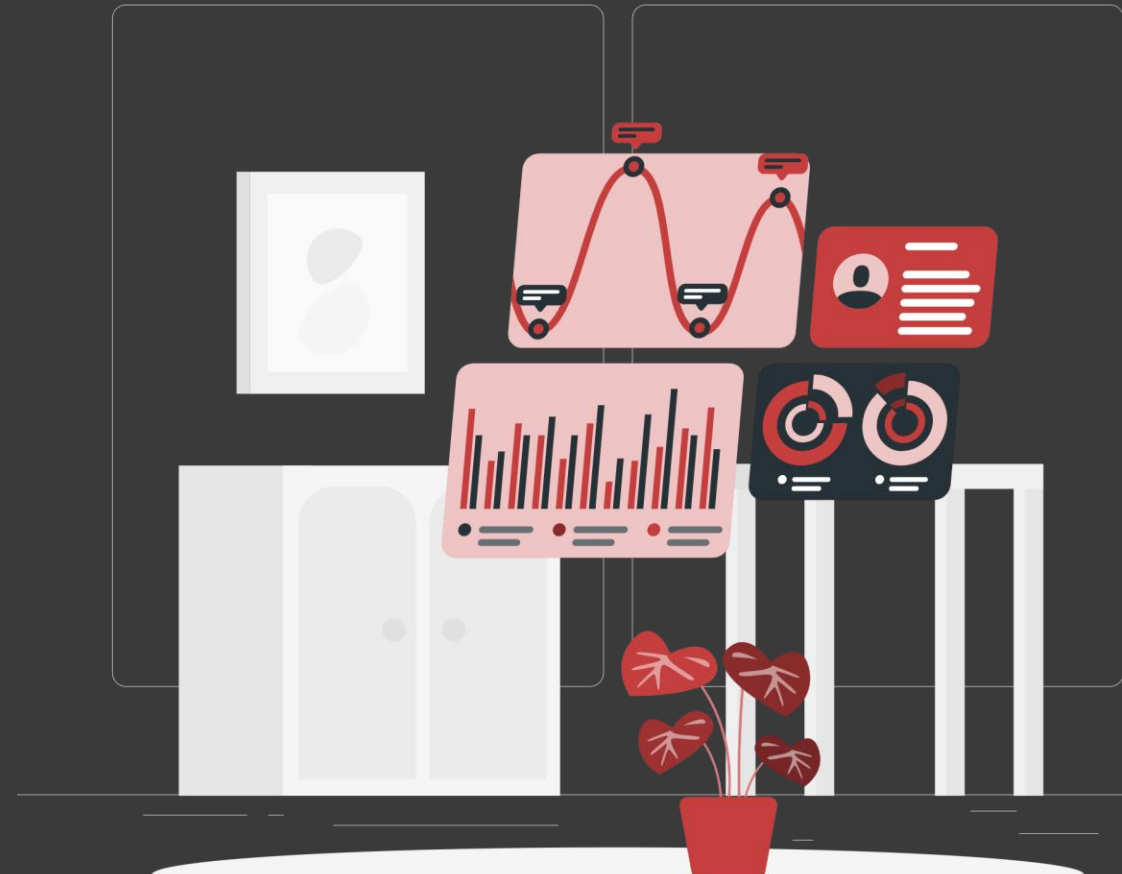
3

4

5

6

7



7 Consultas de Apoio à Decisão

1

2

3

4

5

6

7

- Crescimento de diabéticos entre os anos;
- Quantidade de diabético por região e ano;
- Comparação da quantidade de diabéticos entre países por ano;
- Distribuição da quantidade de diabéticos por faixa etária e ano;



8 Indicadores de análise

1

2

3

4

5

6

7

- Quantidade de diabéticos por região
- Quantidade de diabéticos por país
- Quantidade de diabéticos por faixa etária
- Quantidade de diabéticos a cada 10 anos



1

2

3

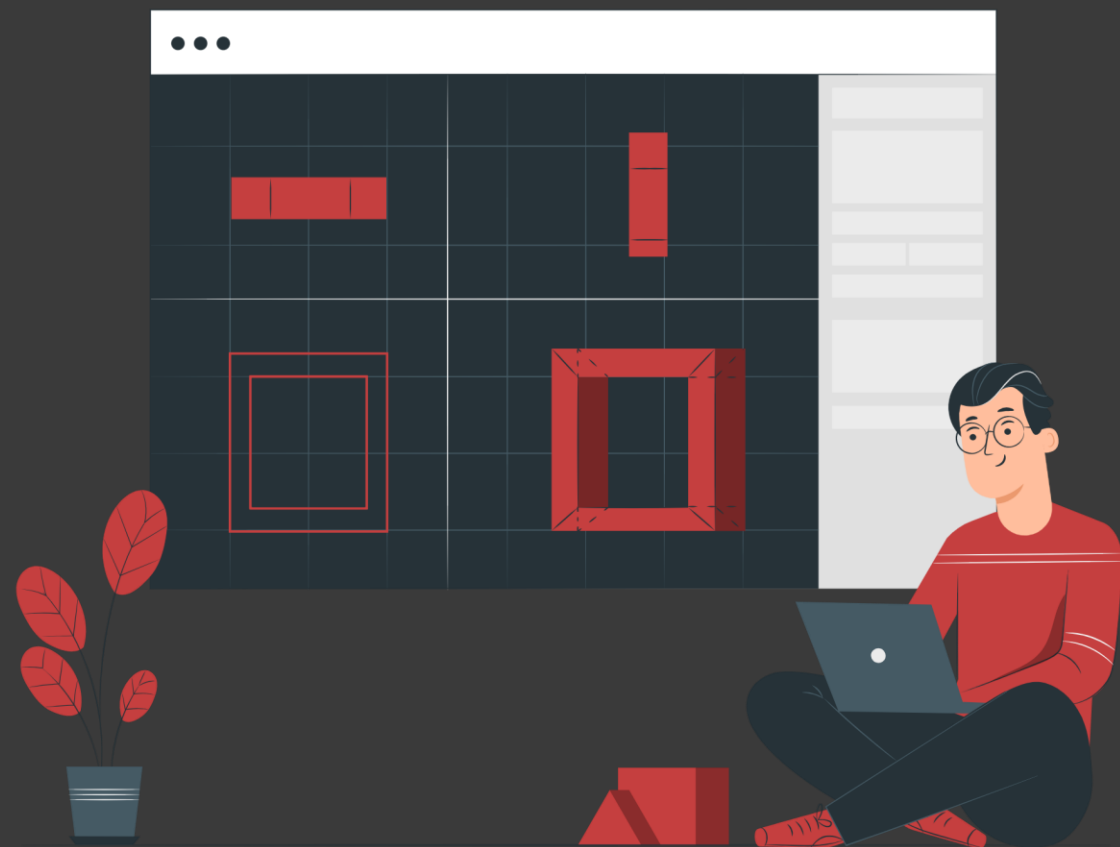
Etapa 3 - Modelagem

4

5

6

7



9 Modelo Relacional

1

2

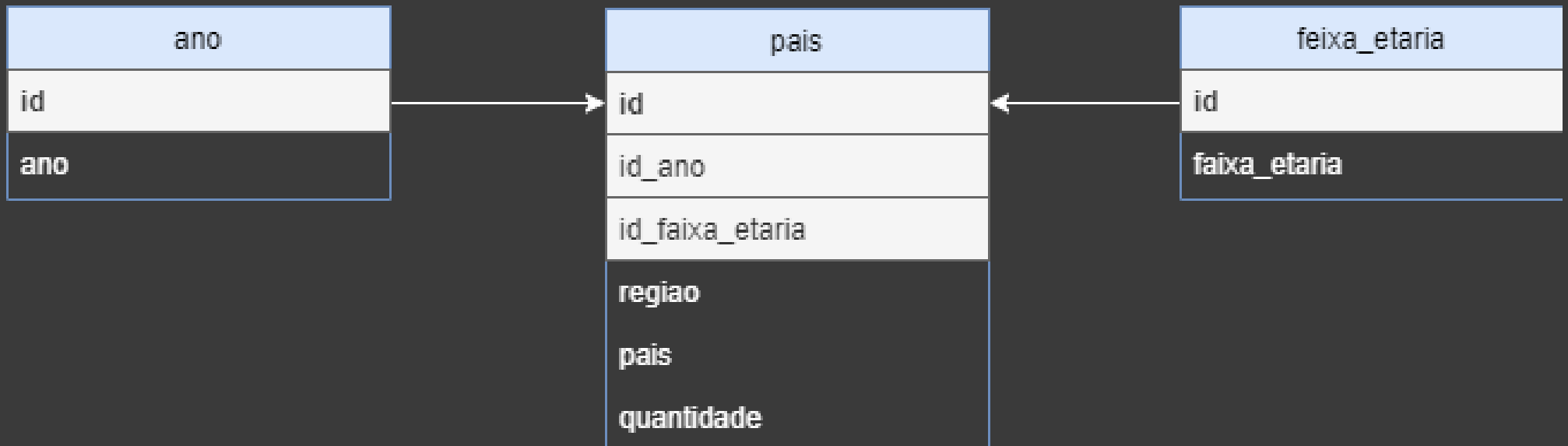
3

4

5

6

7



10 Modelo Dimensional

1

A - Área de Negócios

- Saúde

2

3

B - Processo

- Análise da quantidade de diabéticos no mundo/país a cada 10 anos

4

5

C - Granularidade

- Região X País X Ano X Faixa etária

6

7

10 Modelo Dimensional

1

2

3

4

5

6

7

D - Atributos e Hierarquia das Dimensões

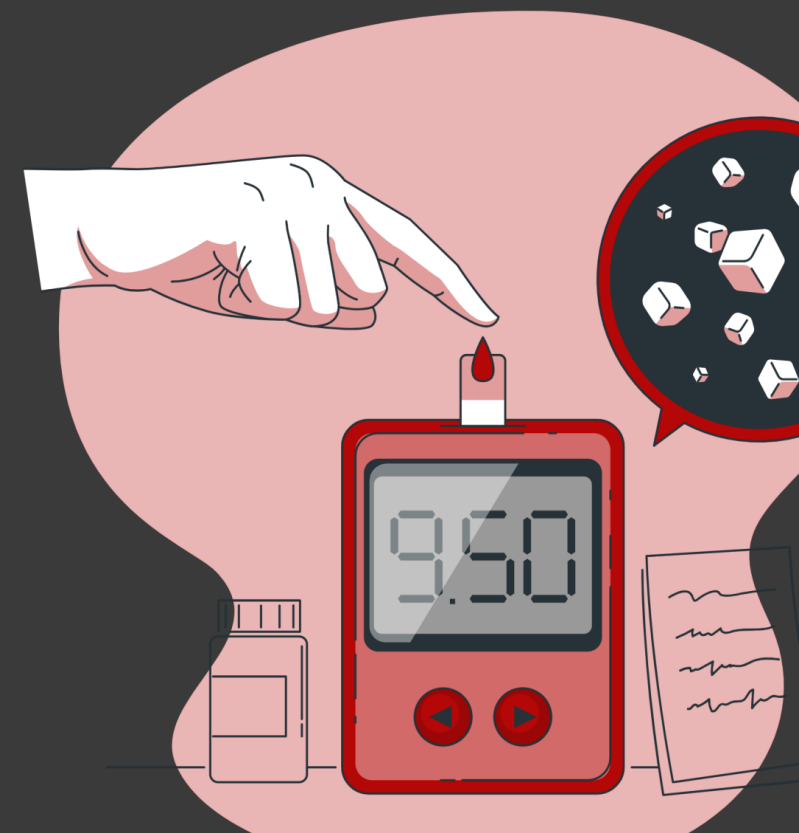
(DIM_Local: região, país)

(DIM_Tempo: ano)

(DIM_Faixa: faixa etária)

E - Métricas da Fato

- quantidade_diabetes (aditiva)



10 Modelo Dimensional

1

2

3

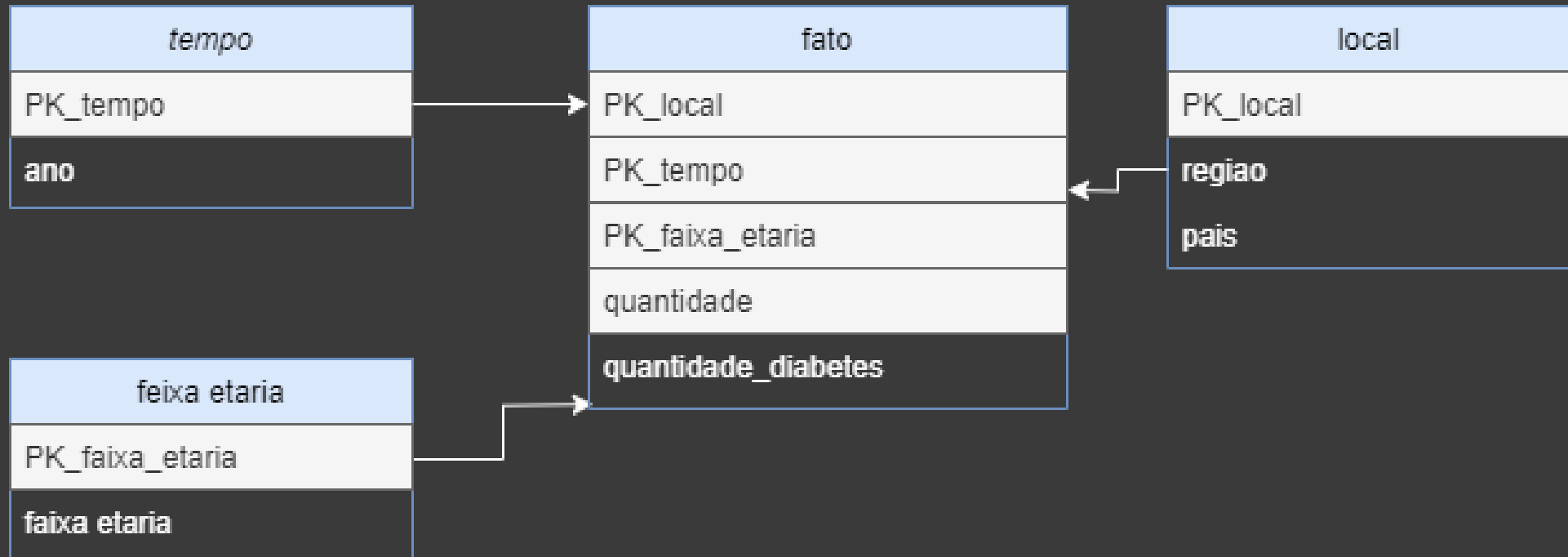
4

5

6

7

F – Esquema Estrela



10 Modelo Dimensional

1

2

3

4

5

6

7

G - Simulação de inserção de 10 “fatos”

tempo	
PK_tempo	ano
1	2000
2	2011
3	2021

faixa etaria	
PK_faixa_etaria	faixa etaria
1	0-14
2	0-19
3	20-79

local		
PK_local	regiao	pais
1	América do Sul e Central	Brazil
2	América do Norte e Caribe	Canada
3	Africa	Angola

fato			
id_tempo	id_local	id_faixa_etaria	quantidade
1	1	1	85654
1	1	2	81300
1	2	3	90865
2	1	2	114966
2	1	3	84474
3	3	3	90825
2	3	3	137730
2	2	2	99326
1	3	3	118160
3	3	1	83912

10 Modelo Dimensional

1

H - Estimativa de espaço

2

Supondo uma média de 225 países analisados por ano, com 3 faixas etárias diferentes para o período de 3 anos diferentes temos:

4

$225 \times 3 \times 3 = 2025$ registros;

5

6

Considerando ainda 3 chaves (pk_tempo, pk_local, pk_feica_etaria), cada uma com 4 bytes e 1 valor numérico de 4 bytes, temos;

7

$2025 \times ((3 \times 4) + (1 \times 4)) = 32,4 \text{ bytes}$

1

2

3

4

Etapa 4 – Projeto Físico do BD

5

6

7



1

2

3

4

5

6

7

11 Modelo Relacional do Data Mart (físico)



```
1 CREATE TABLE dim_tempo (id SERIAL PRIMARY KEY,  
2                             ano INTEGER NOT NULL);  
3  
4 CREATE TABLE dim_local (id SERIAL PRIMARY KEY,  
5                             regiao VARCHAR(255),  
6                             pais VARCHAR(255));  
7  
8 CREATE TABLE dim_faixa_etaria (id SERIAL PRIMARY KEY,  
9                                 faixa_etaria VARCHAR(10));  
10  
11 CREATE TABLE fato (id_tempo INTEGER REFERENCES dim_tempo(id),  
12                     id_local INTEGER REFERENCES dim_local(id),  
13                     id_faixa_etaria INTEGER REFERENCES dim_faixa_etaria(id),  
14                     quantidade INTEGER, PRIMARY KEY (id_tempo, id_local, id_faixa_etaria));
```

1

2

3

4

6

7

5

Etapa 5 - Extração, Transformação e Carga



12 Plano de Carga da Dimensão Tempo

1

2

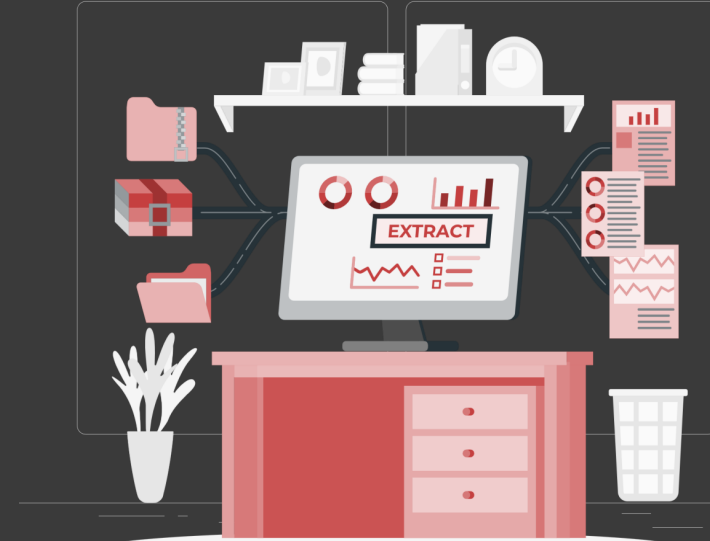
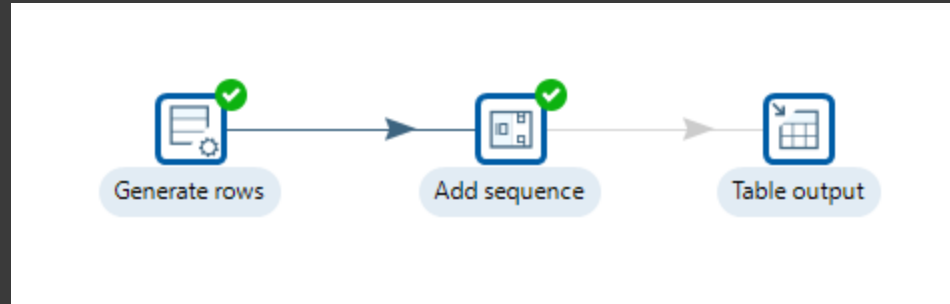
3

4

5

6

7



13 Plano de Carga da Dimensão Local

1

2

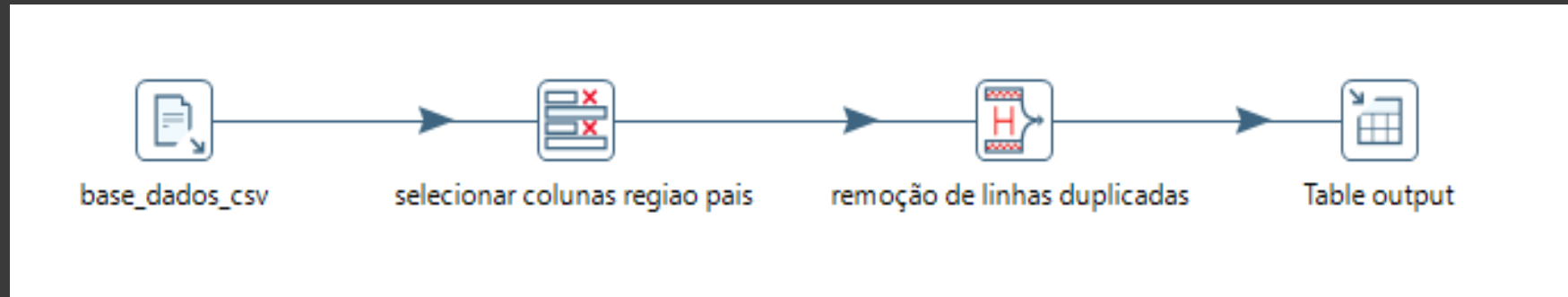
3

4

5

6

7



14 Plano de Carga da Dimensão Faixa Etária

1

2

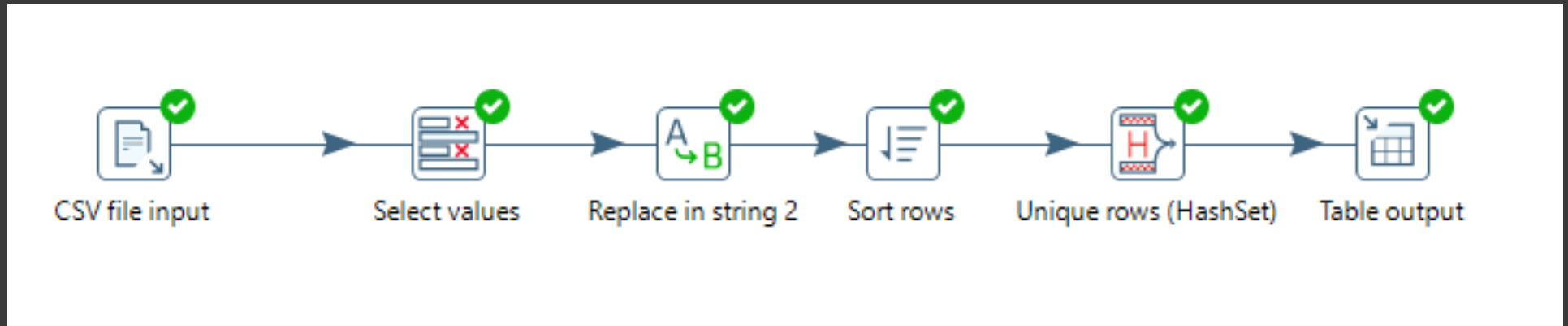
3

4

5

6

7



15 Plano de Carga da Fato

1

2

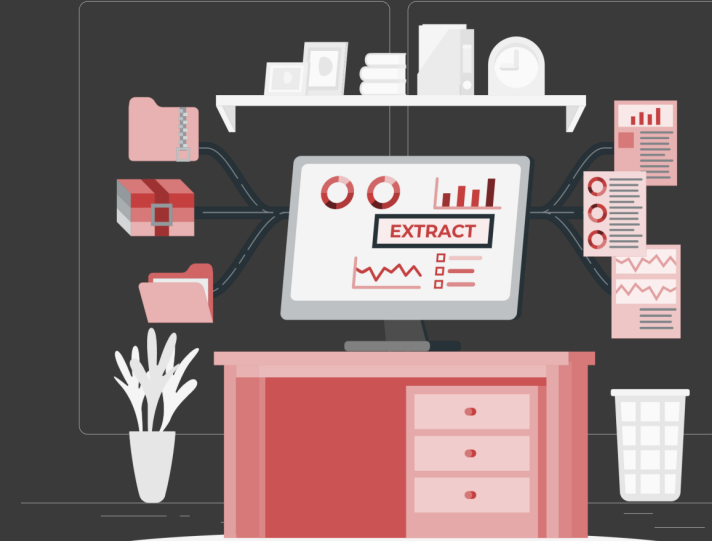
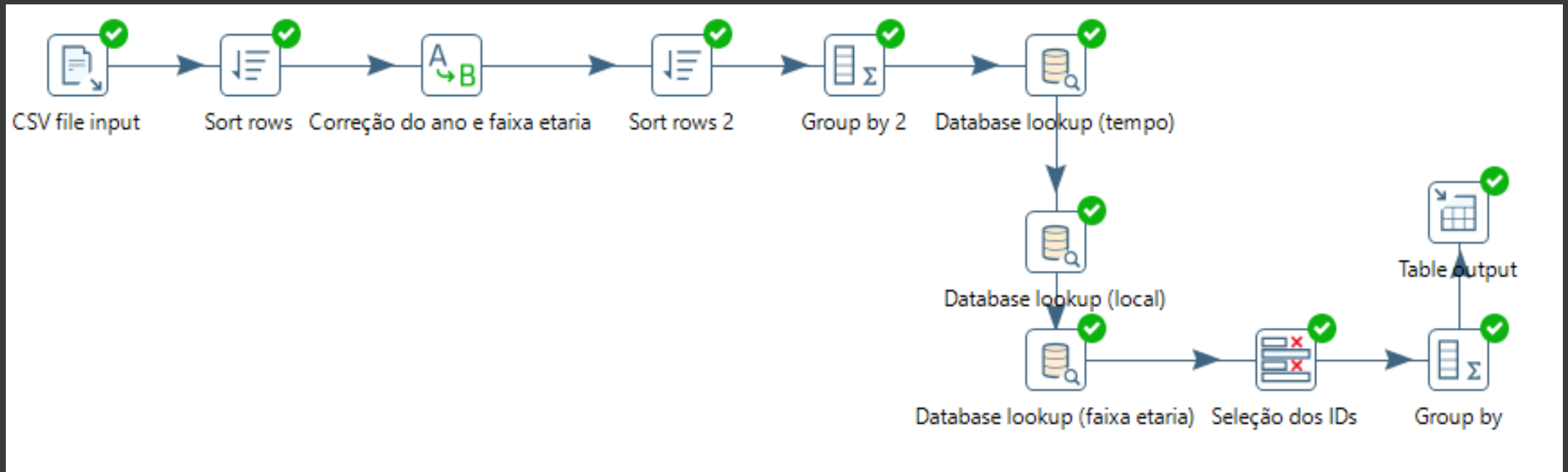
3

4

5

6

7



1

2

3

4

5

7

6

Etapa 6 - Aplicação OLAP e Painel de Bordo



16 Consulta OLAP - 1: Crescimento x ano



1

2

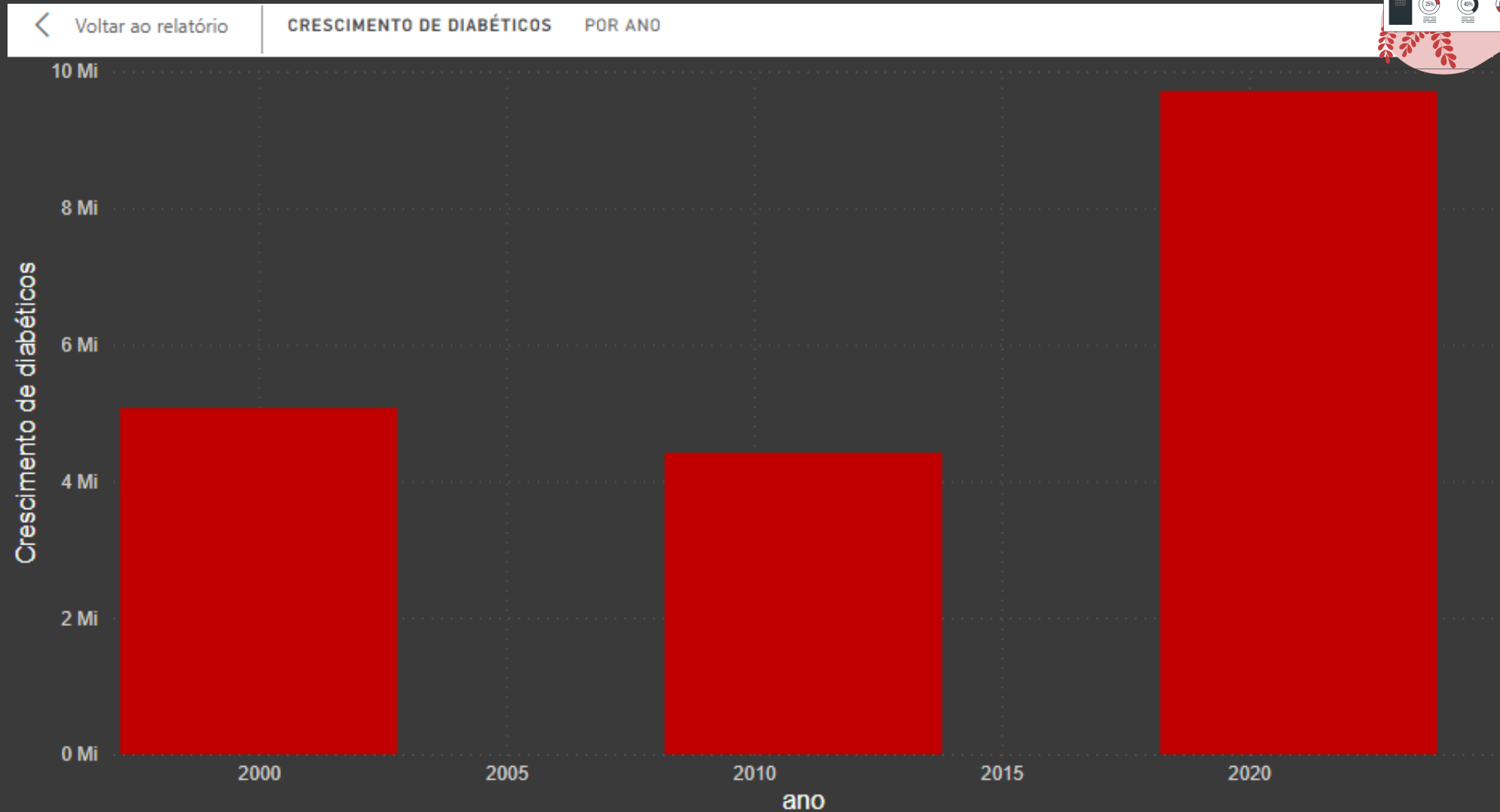
3

4

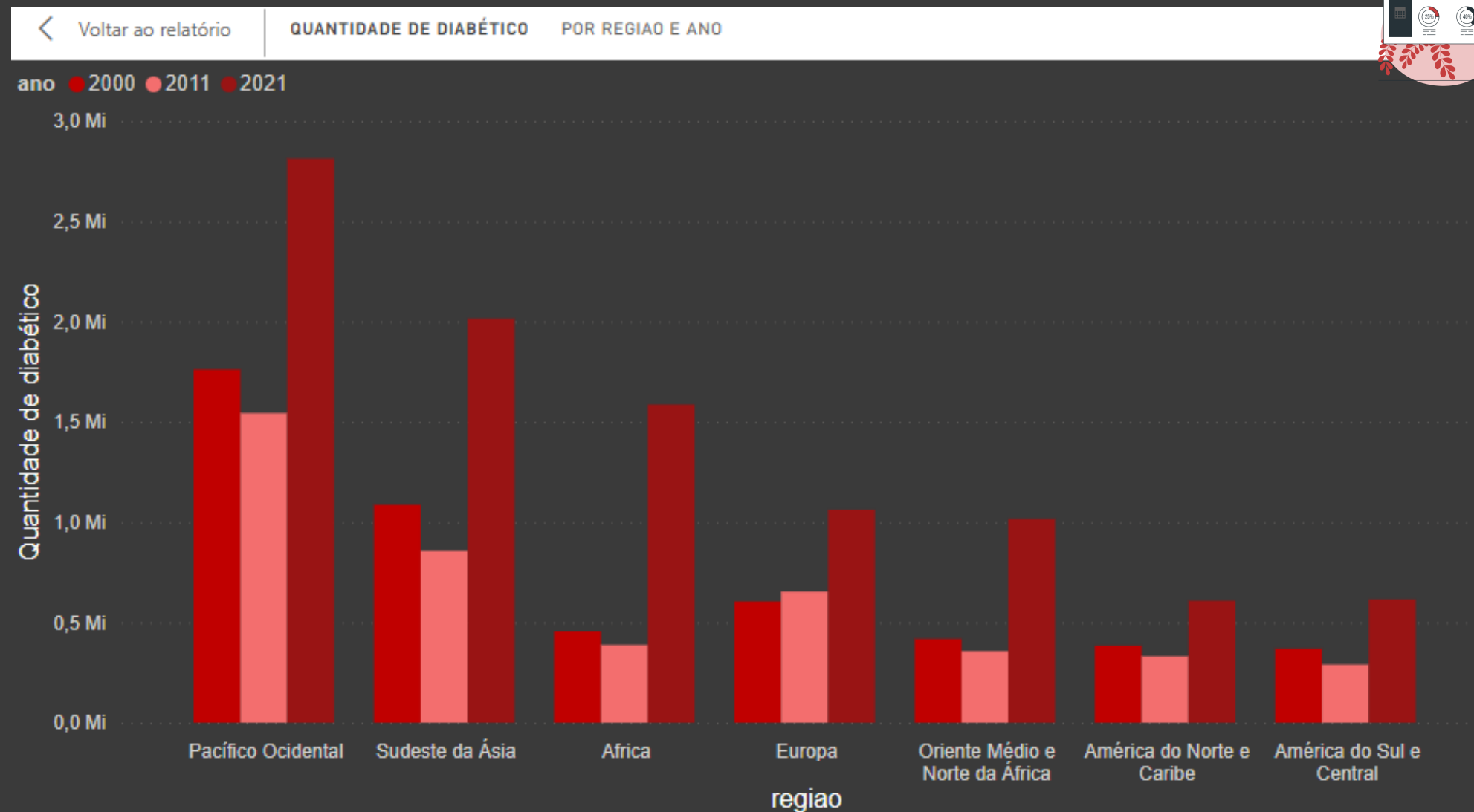
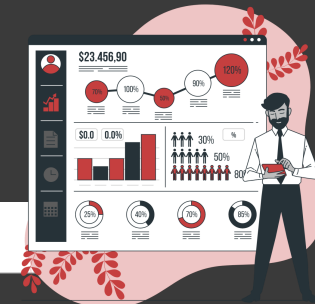
5

6

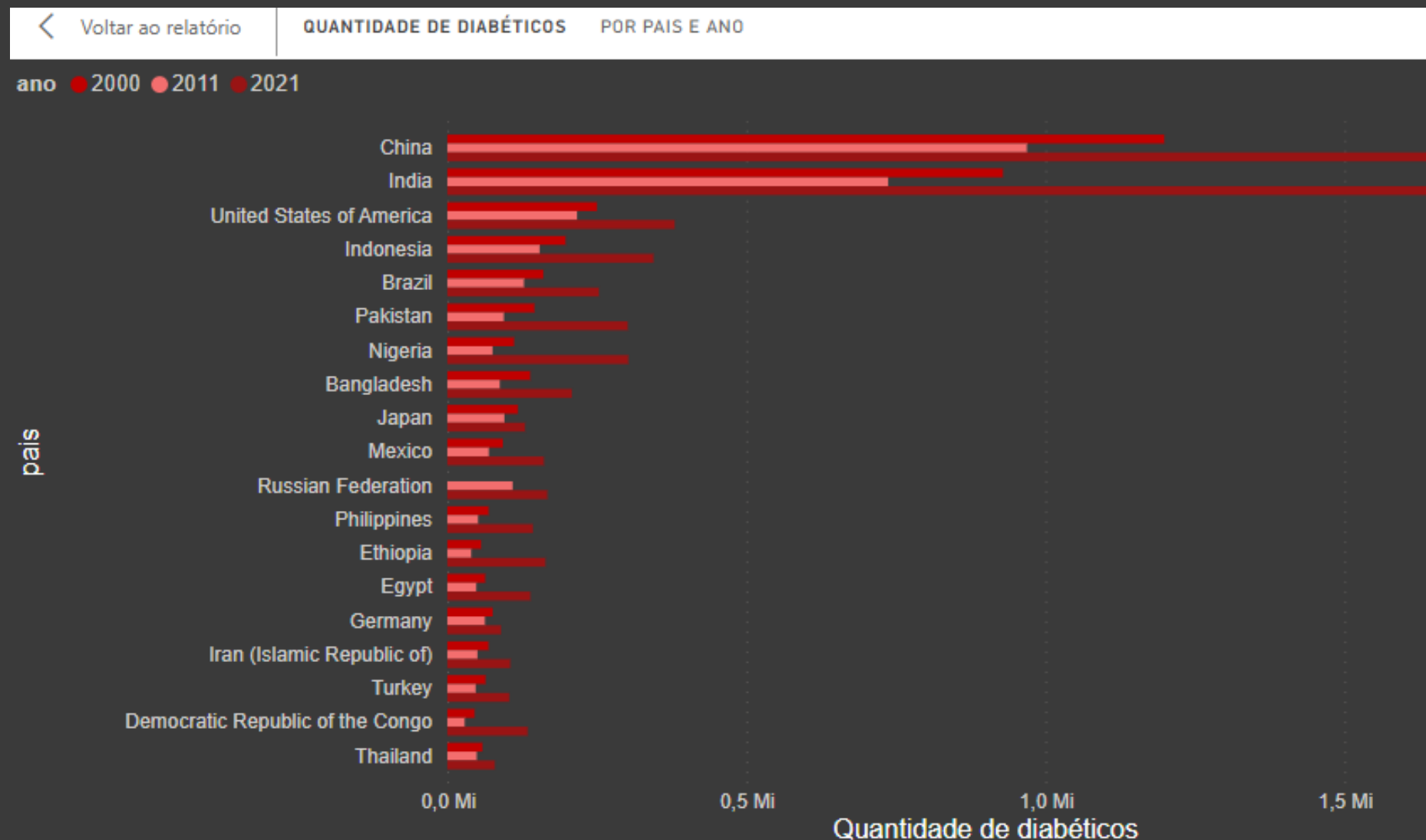
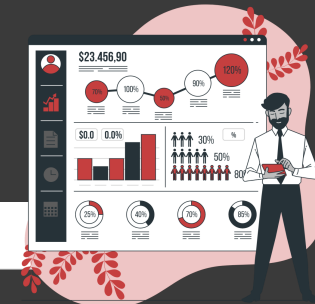
7



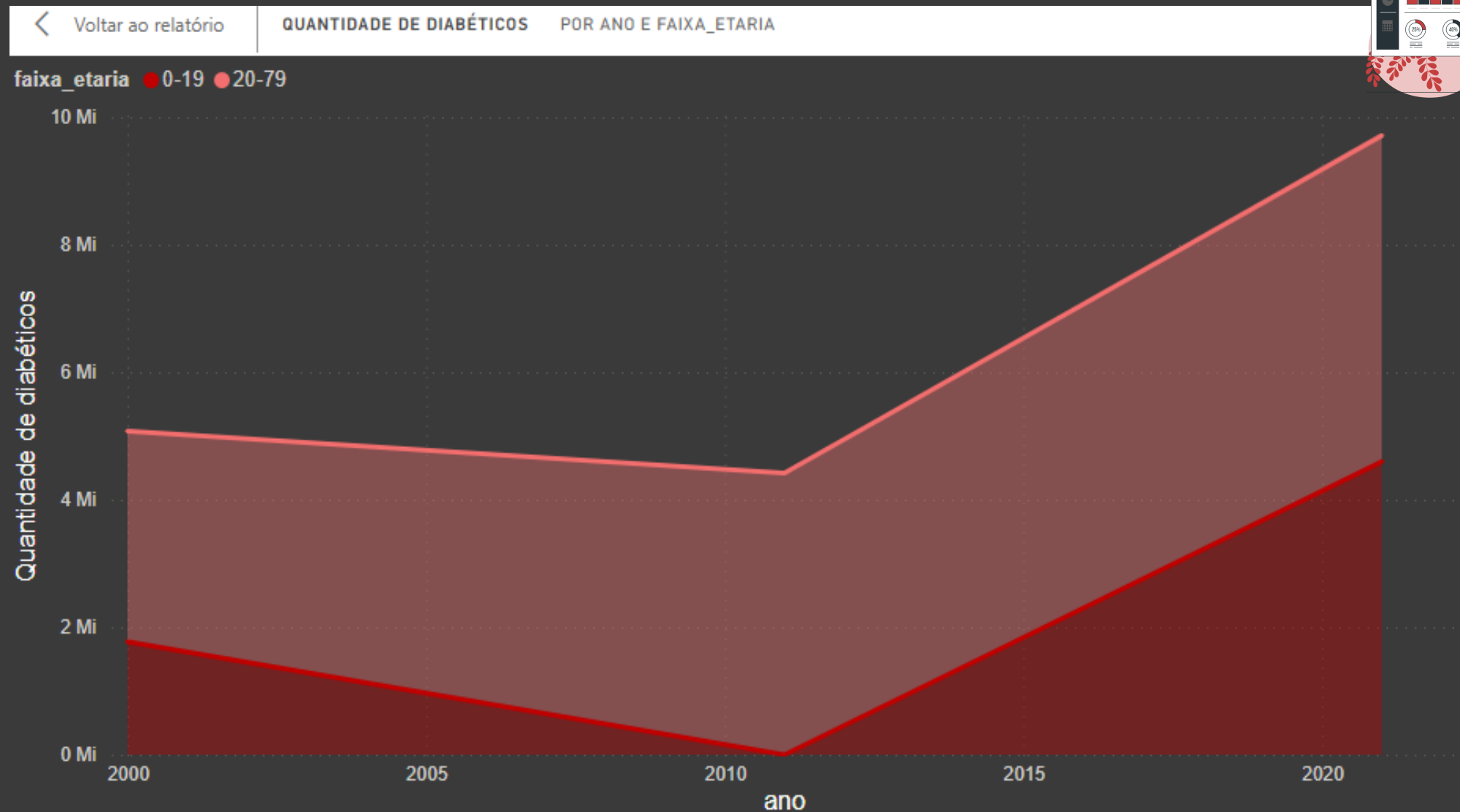
16 Consulta OLAP - 2: Quantidade região x ano



16 Consulta OLAP - 3: Quantidade pais x ano



16 Consulta OLAP - 4: Quantidade ano x faixa etaria



16 Painel de Bordo

1

2

3

4

5

6

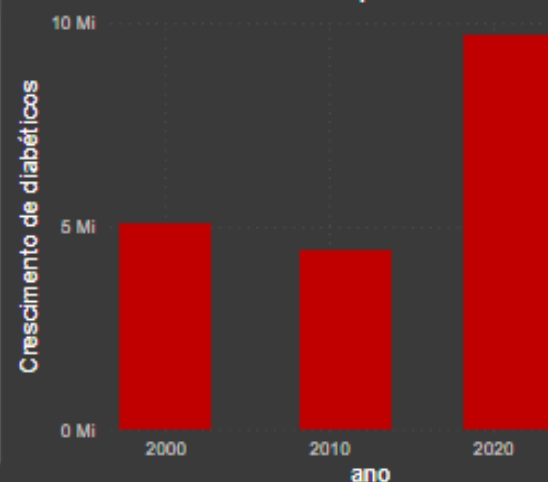
7

Análise Diabética no Mundo nos Anos de 2000, 2011 e 2021

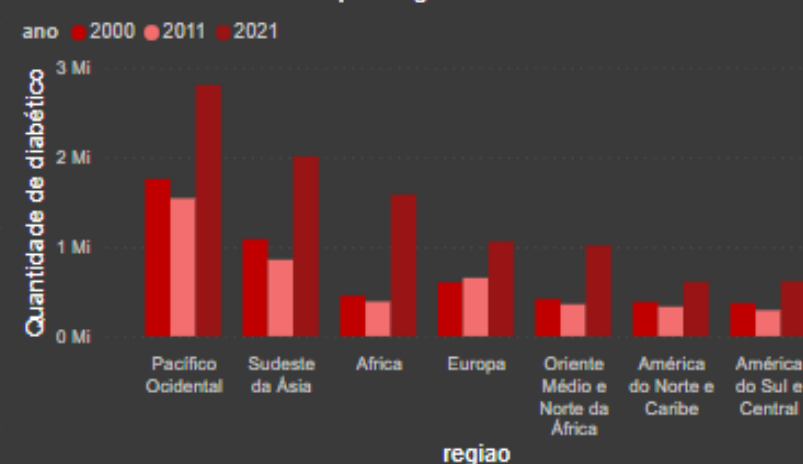
pais

Afghanistan
Albania
Algeria
American Samoa
Andorra
Angola
Anguilla
Antigua and Barbuda
Argentina
Armenia
Aruba
Australia
Austria
Azerbaijan
Bahamas
Bahrain

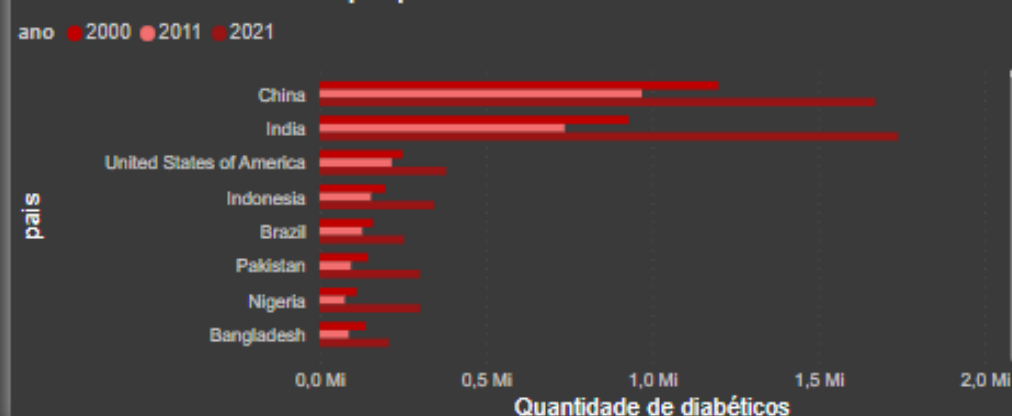
Crescimento de diabéticos por ano



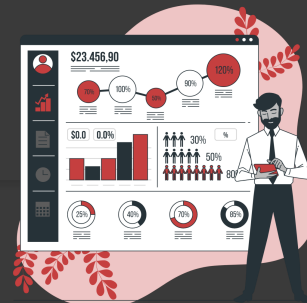
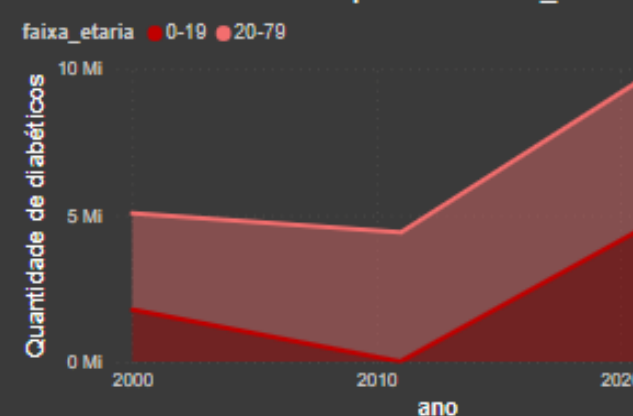
Quantidade de diabético por regioao e ano



Quantidade de diabéticos por pais e ano



Quantidade de diabéticos por ano e faixa_etaria



1

2

3

4

5

6

7

Referências

Dados fontes (abertos): Os dados utilizados para a criação do Data Mart são da IDF Diabetes Atlas (International Diabetes Federation) entre os anos 2000, 2011 e 2021.

Estão disponíveis no site da IDF, através do endereço:
<https://diabetesatlas.org/data/en/indicators/21/>