

Cinzas do Brasil

Análise Multidimensional de Dados de Queimadas

Projeto final da Disciplina Processamento Analítico de Dados

Realizado por:

Felipe Carneiro Machado - 14569373

Lívia Lelis - 12543822

Clara Ernesto de Carvalho - 14559479

Contexto do Problema

30,8 milhões de hectares queimados no Brasil em 2024

- Área maior que toda a Itália

Objetivo: Criar um Data Warehouse para análise de:

- Correlações entre queimadas e clima
- Padrões espaciais e temporais
- Impactos na qualidade do ar

Fontes de Dados

Queimadas - INPE

- Focos de incêndio detectados por satélite
- Localização geográfica (lat/long), bioma
- FRP, dias sem chuva, risco de fogo

Clima - SISAM/INPE

- Dados meteorológicos e qualidade do ar
- Temperatura, umidade, precipitação
- Poluentes: PM_{2.5}, CO, NO₂, O₃, SO₂

Fontes de Dados

Problema com horários

O Dataset de Clima possui dados apenas coletados com horas 0, 6, 12 e 18.

Sendo assim, quando juntamos os dados de Clima com os de Queimadas, perdemos muitos dados.

Outra opção seria interpolar os dados de Clima para os horários faltantes.

Decidimos manter apenas os horários disponíveis no Dataset de Clima, já que não impactava significativamente a análise.

Fontes de Dados

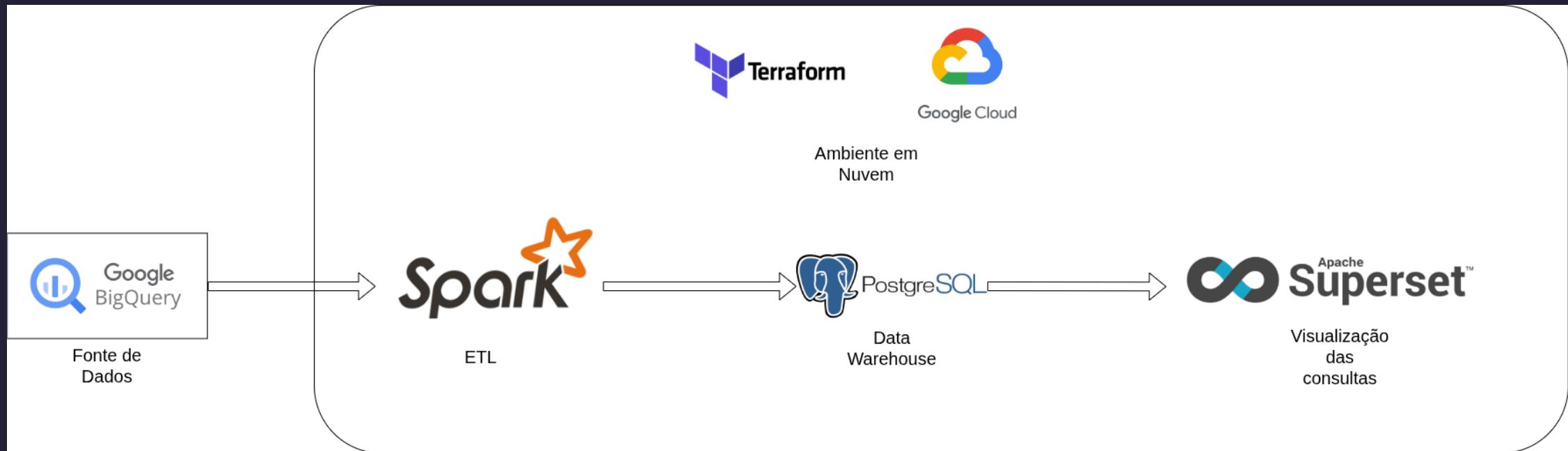
Geográficos - IBGE

- Diretórios de UFs e municípios
- Hierarquias administrativas e regiões

Objetivo

- Consolidação de dados do INPE (Instituto Nacional de Pesquisas Espaciais) relativos a focos de queimadas e clima
- Criação de um Data Warehouse com dados históricos
- Geração de visualizações baseadas em consultas analíticas para tomada estratégica de decisões

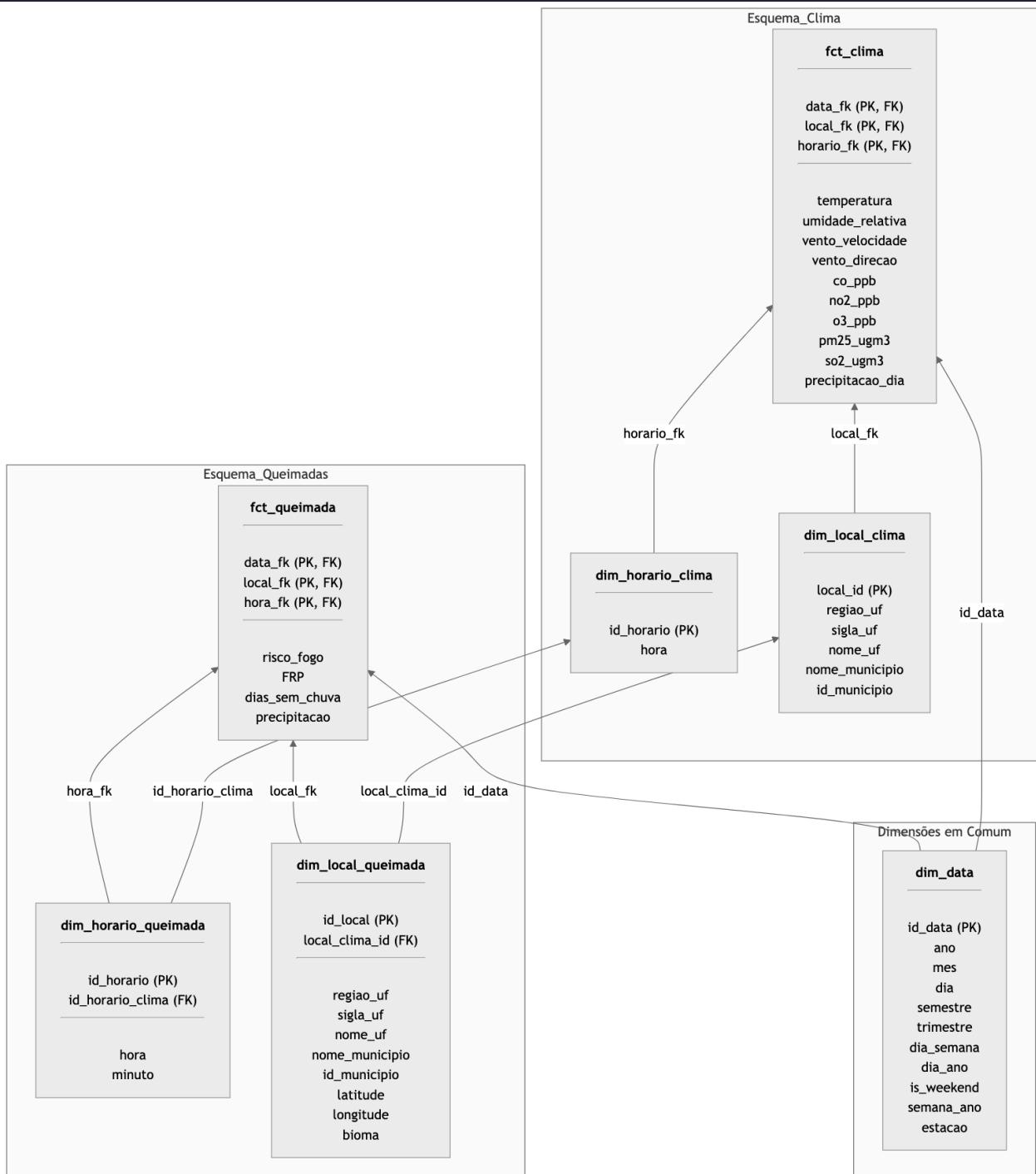
Visão geral da Arquitetura



Organização do Data Warehouse

Constelação de fatos corrigida

- Tabela bridge
 - Dimensão Data unificada
 - Dimensões do esquema
Queimada possuem Chaves Estrangeiras para as mesmas dimensões no esquema Clima

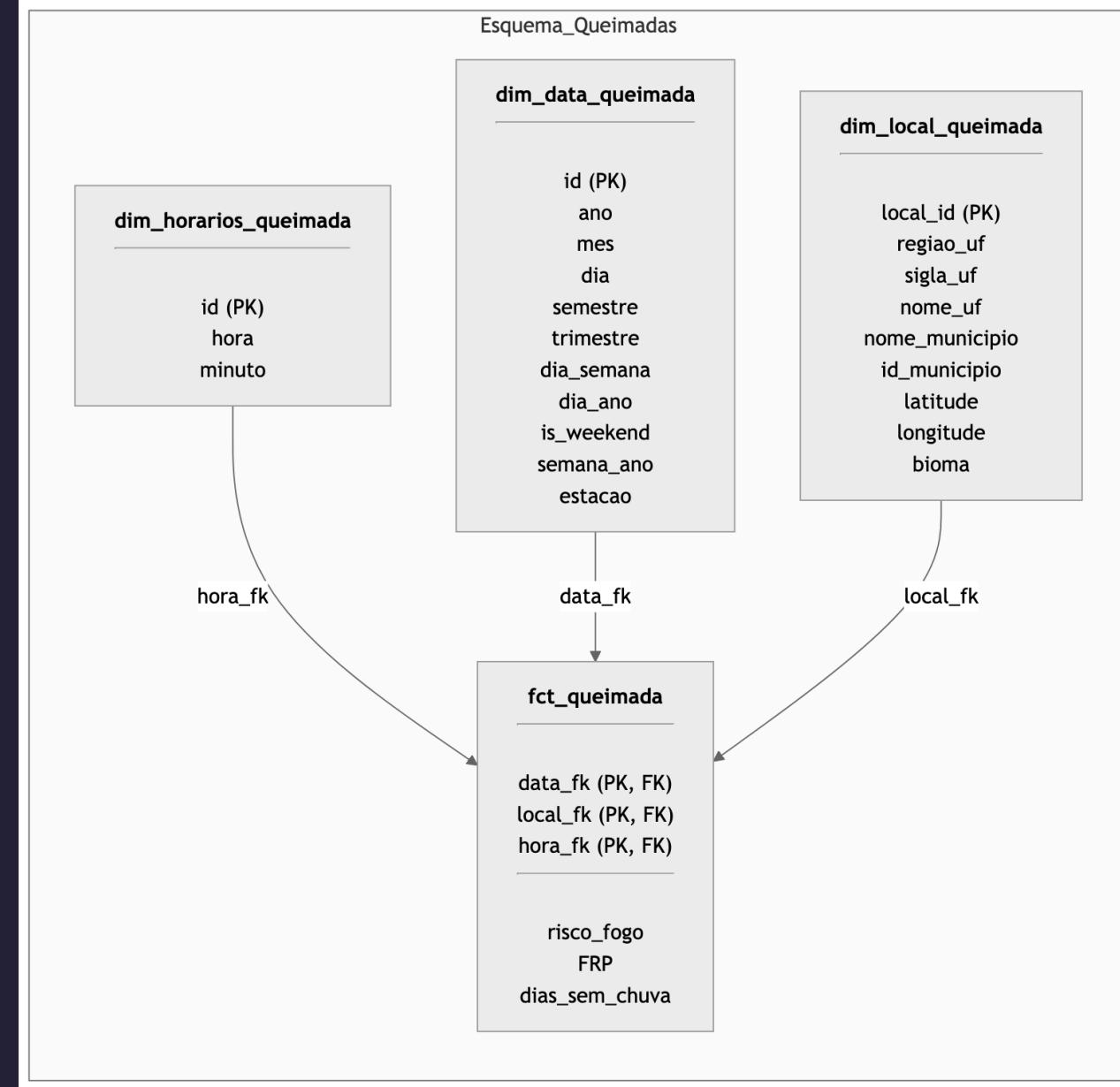


Organização do Data Warehouse

Queimadas

Esquema estrela corrigido

- Dessa vez, fizemos o esquema estrela sem pensar na posterior unificação

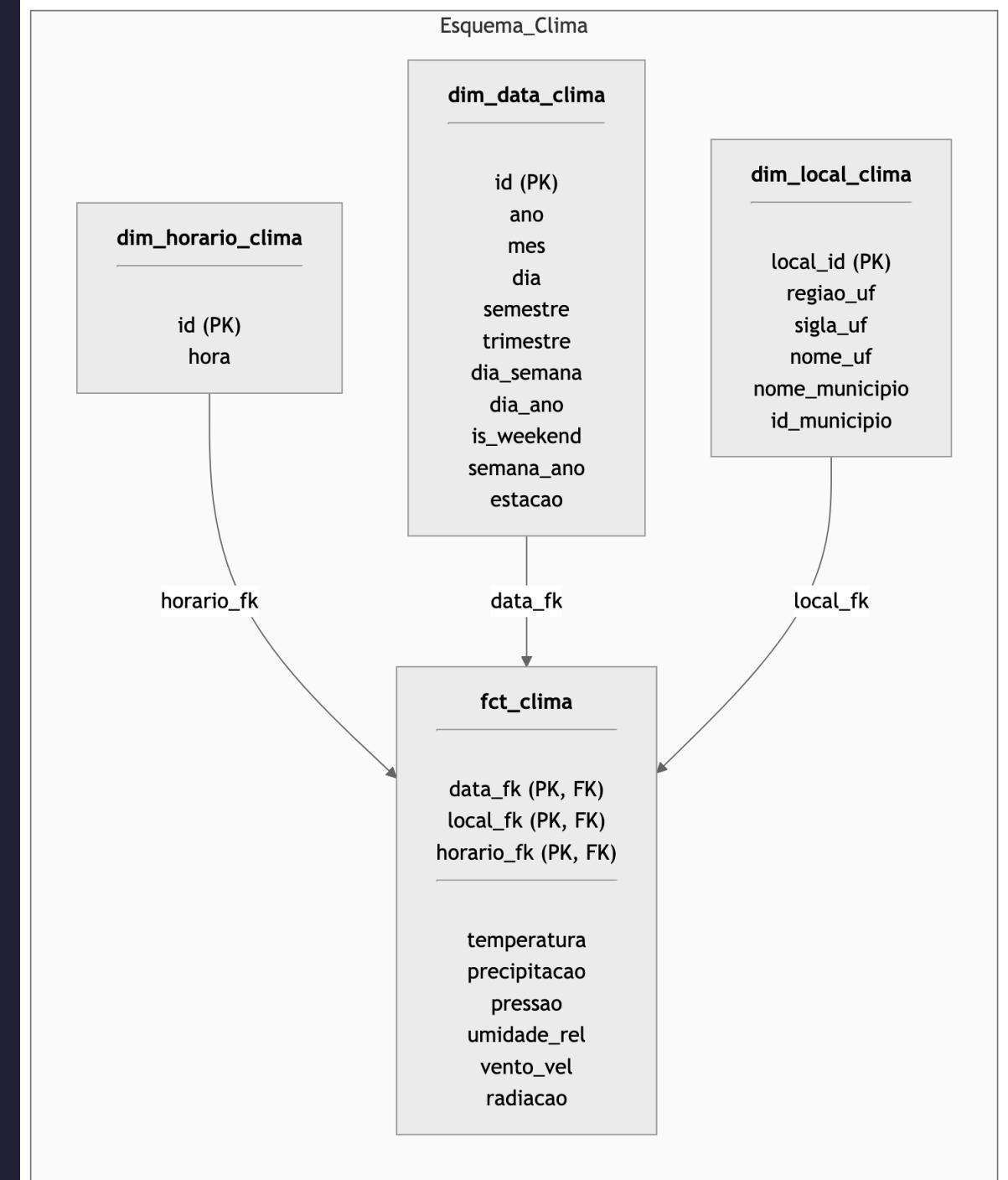


Organização do Data Warehouse

Clima

Esquema estrela corrigido

- Dessa vez, fizemos o esquema estrela sem pensar na posterior unificação

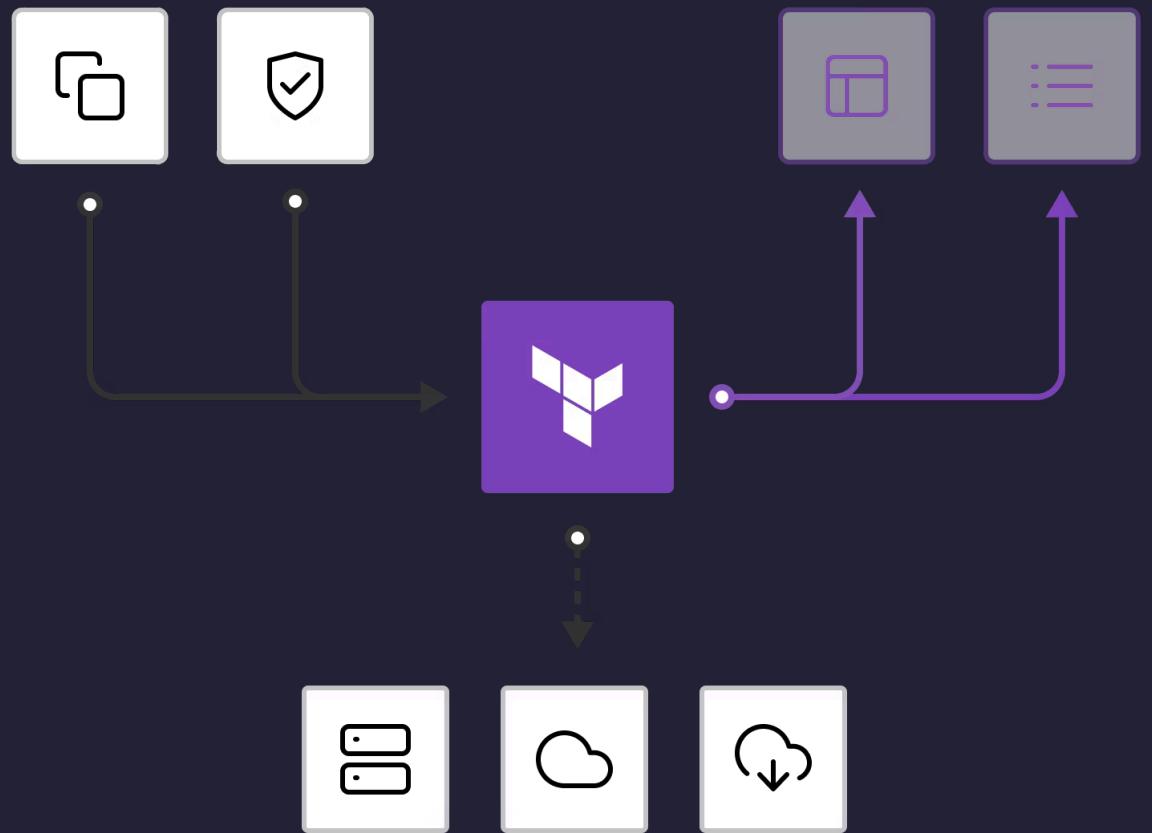


Infraestrutura

Provisionamento com Terraform

Provisionamos com Terraform:

- Base de dados PostgreSQL
- Jobs Spark para ETL
- Superset para visualização dos dados

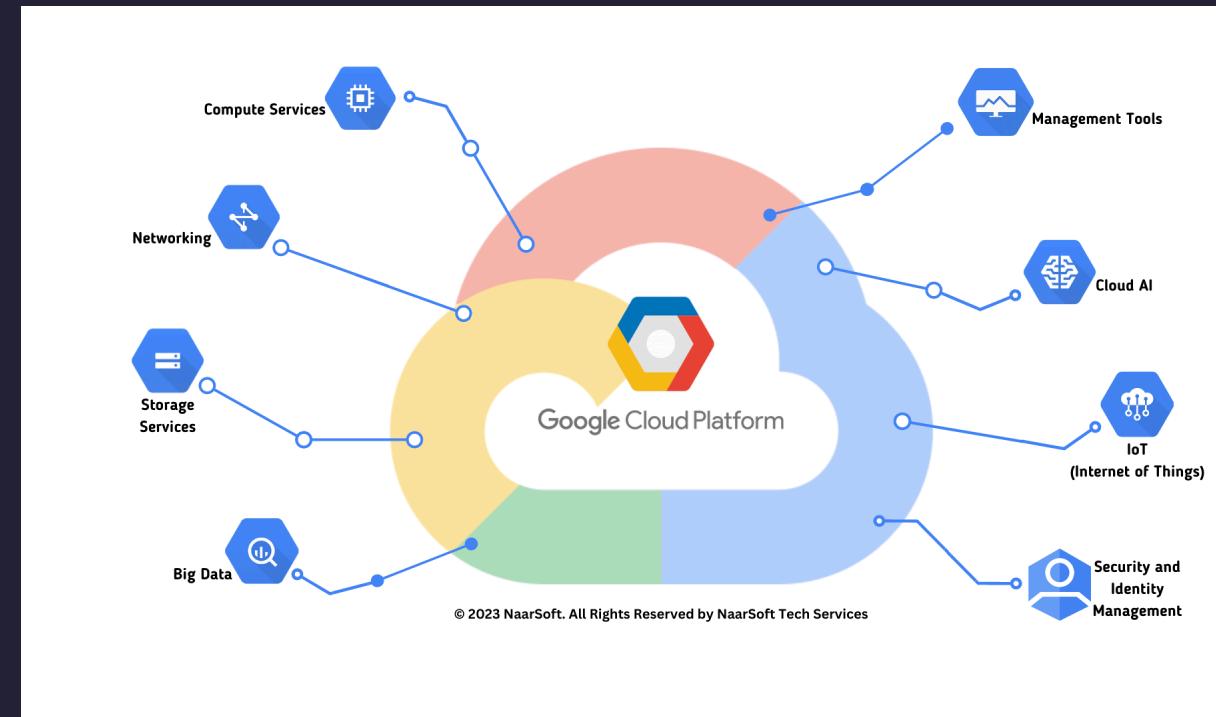


Infraestrutura

Google Cloud Platform

Conseguimos créditos para estudantes.

Algumas das VMs utilizadas eram Spot, portanto podem ser derrubadas a qualquer momento.



Base de Dados

PostgreSQL

Utilizamos Citus para
armazenamento colunar

| Ol'reliable



Extração dos Dados

Foram utilizadas 4 fontes:

- Dataset de focos de queimadas do INPE -> extraído diretamente do BigQuery
- Dataset de dados climáticos do SISAM -> extraído diretamente do BigQuery
- Relação de municípios e seus códigos pelo IBGE -> CSV obtido de fontes públicas
- Relação de estados e as regiões às quais pertencem -> CSV obtido de fontes públicas

Transformação

Pré-processamento dos dados:

- Tratamento de valores faltantes
- Computação de atributos derivados (ex: estação do ano)
- Atribuição de tipos de dados corretos

Transformação

Criação de dimensões e tabelas de fatos:

- Dimensão Data gerada a partir de união e projeção das tabelas de Queimadas e Clima
- Dimensões Local geradas a partir da junção e projeção de cada tabela com as relações de Municípios e regiões
- Dimensão Horário preenchida proceduralmente com todos os valores de horas e minutos
- Chaves Estrangeiras das tabelas de fatos preenchidas através de junção com as tabelas de dimensões

Carregamento

Dados carregados em um RSGBD (PostgreSQL)

Indíces criados para colunas de frequente acesso:

- Chaves estrangeiras nas tabelas de fatos e dimensões do esquema Queimadas
- Timestamp, Mes e ano para Data
- Hora para Horário
- UF e município para Local

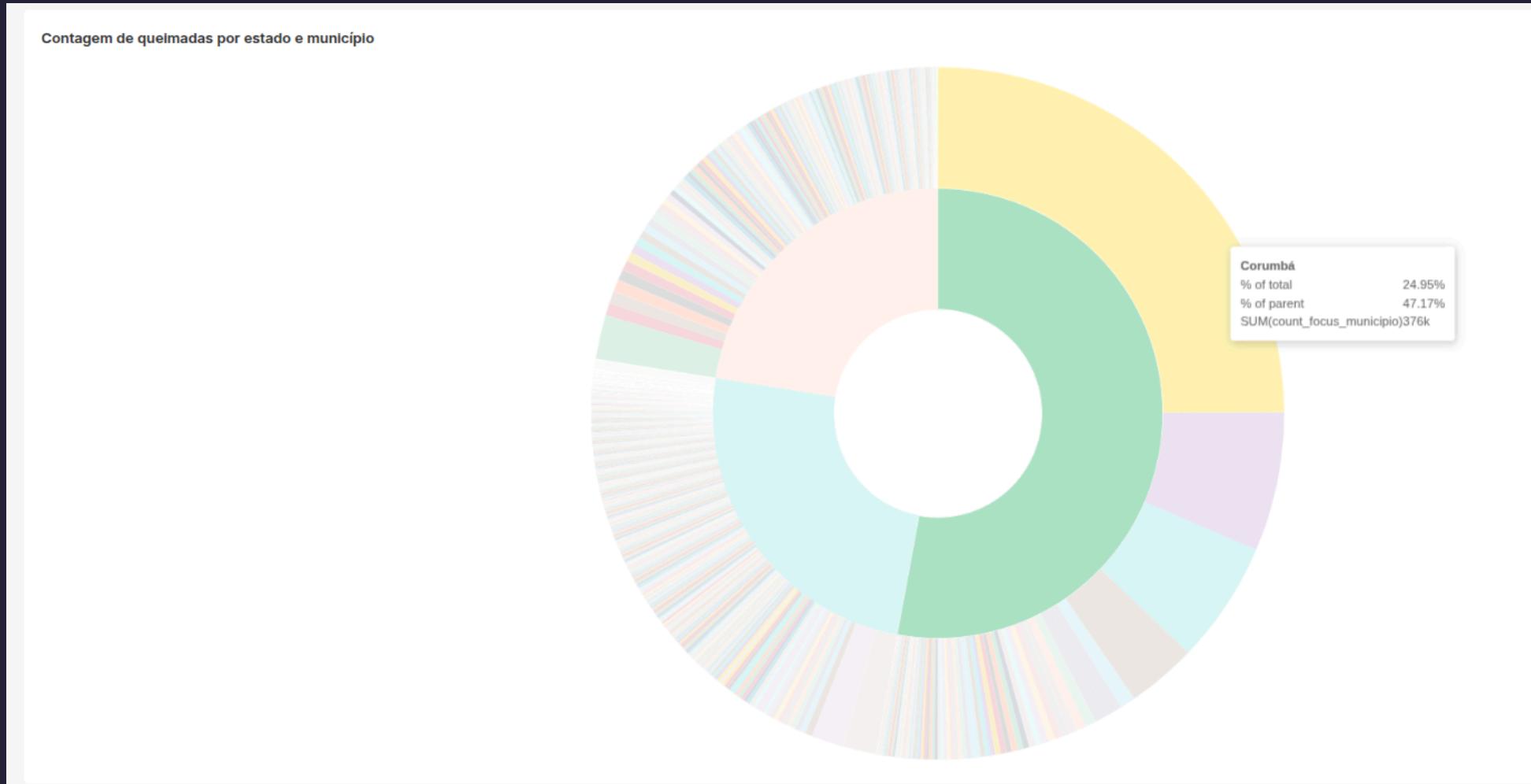
Consultas analíticas e visualização

- Consultas implementadas em SQL
- Visualizações construídas com Apache Superset

Queimadas por Estado e Município

```
SELECT DISTINCT
    dim_local_queimada.nome_uf,
    dim_local_queimada.nome_municipio,
    count(*) OVER (PARTITION BY id_municipio) as count_focus_municipio,
    count(*) OVER (PARTITION BY sigla_uf) as count_focus_uf
FROM fct_queimada
    INNER JOIN dim_local_queimada ON fct_queimada.id_local = dim_local_queimada.id_local;
```

Queimadas por Estado e Município



Risco de Fogo por Bioma e Estado

```
SELECT
    dd.ano,
    dlq.bioma,
    dlq.nome_uf,
    avg(fct_queimada.risco_fogo) as media_risco_fogo
FROM fct_queimada
    INNER JOIN dim_data dd on fct_queimada.id_data = dd.id_data
    INNER JOIN dim_local_queimada dlq on fct_queimada.id_local = dlq.id_local
WHERE fct_queimada.risco_fogo is not null
GROUP BY dd.ano, CUBE (dlq.nome_uf, dlq.bioma)
```

Risco de Fogo por Bioma e Estado

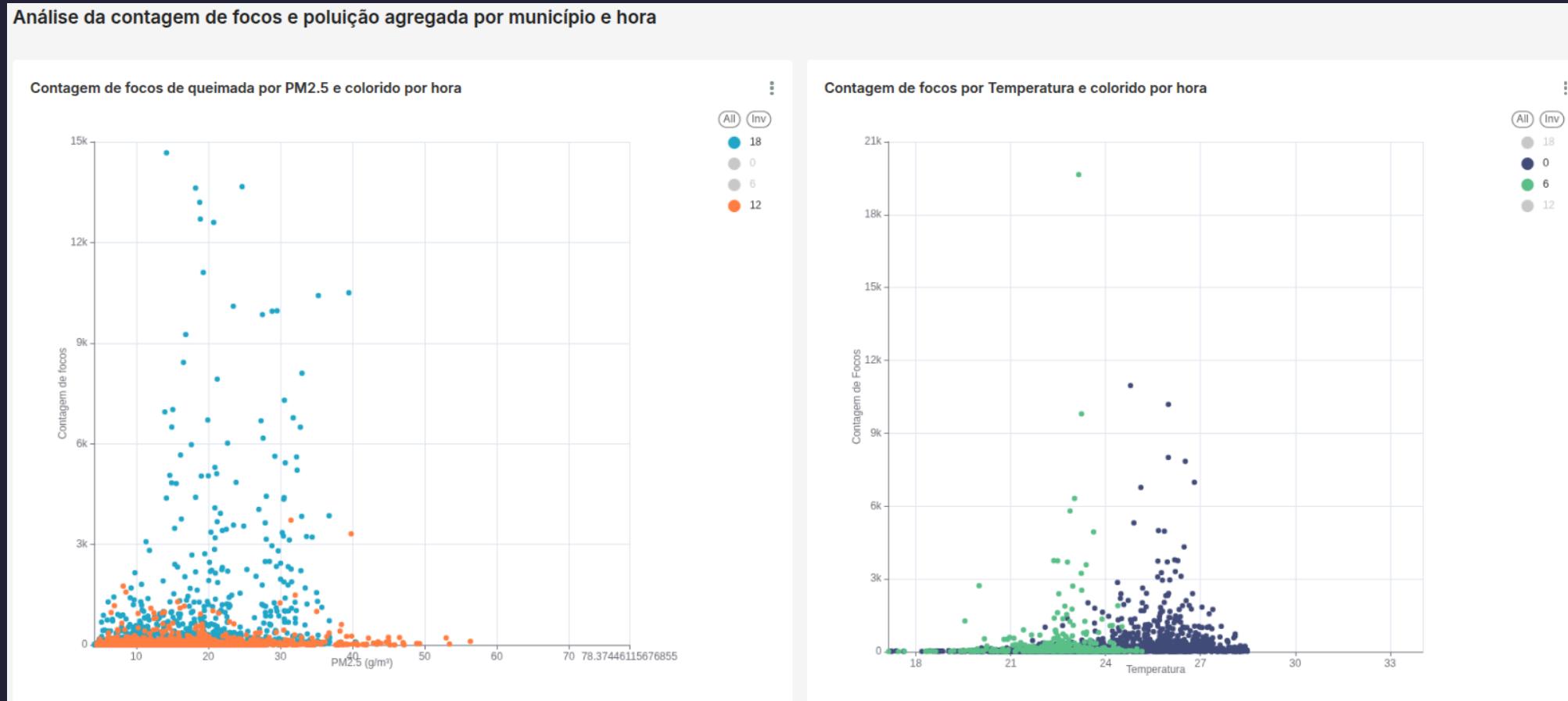


Risco de Fogo por Estado e Bioma



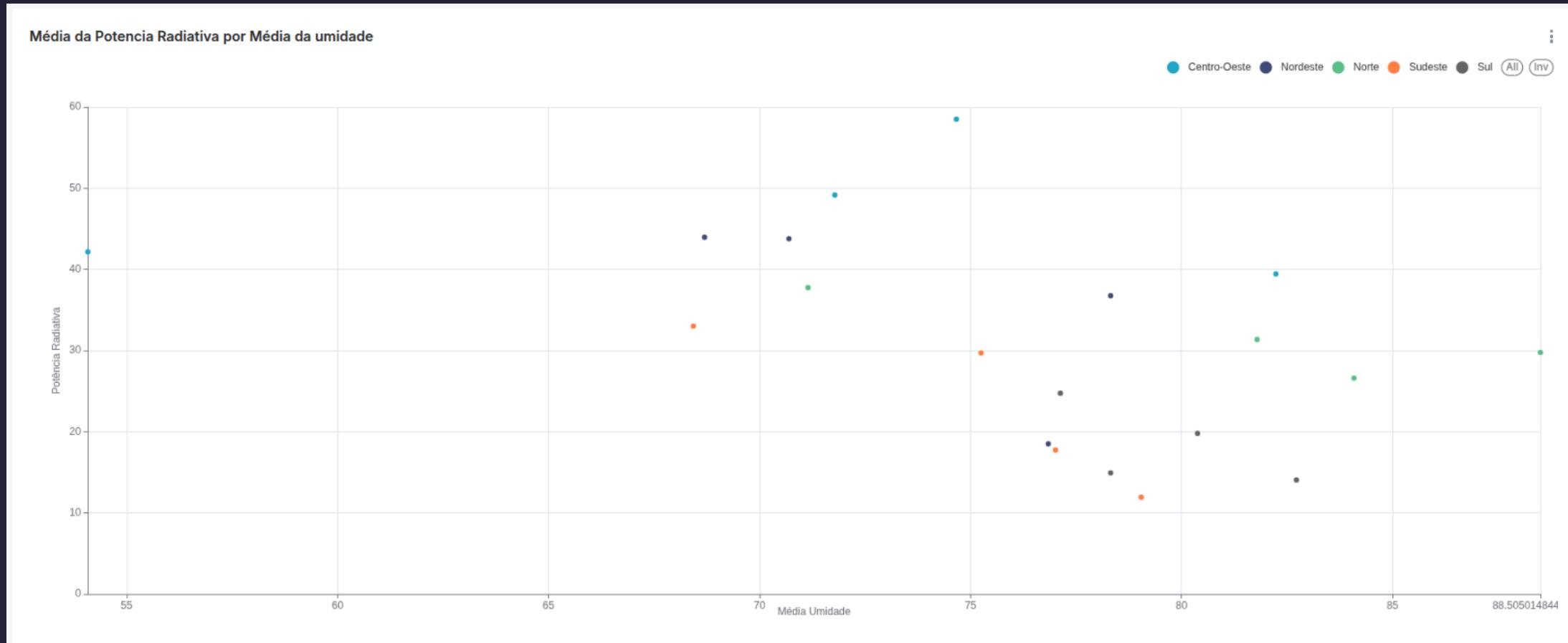
```
WITH joined_clima AS
  (SELECT
    dlc.id_local,
    dhc.id_horario,
    dlc.nome_municipio,
    dlc.sigla_uf,
    dhc.hora,
    avg(fct_clima.temperatura) AS media_temperature,
    avg(pm25_ugm3) AS media_pm25_ugm3
   FROM fct_clima
      INNER JOIN public.dim_local_clima dlc ON dlc.id_local = fct_clima.id_local
      INNER JOIN public.dim_horario_clima dhc ON dhc.id_horario = fct_clima.id_horario
   GROUP BY dlc.id_local, dhc.id_horario, dlc.nome_municipio, dlc.sigla_uf, dhc.hora
  ),
joined_queimada AS
  (SELECT
    dhq.id_horario_clima,
    dlq.id_local_clima,
    count(*) AS count_focus
   FROM fct_queimada
      INNER JOIN public.dim_horario_queimada dhq ON fct_queimada.id_horario = dhq.id_horario
      INNER JOIN public.dim_local_queimada dlq ON dlq.id_local = fct_queimada.id_local
   GROUP BY dhq.id_horario_clima, dlq.id_local_clima
  )
SELECT
  jc.hora,
  (jc.nome_municipio || '/' || jc.sigla_uf) AS nome_municipio,
  jq.count_focus,
  jc.media_pm25_ugm3,
  jc.media_temperature
FROM joined_clima jc
  INNER JOIN joined_queimada jq ON jc.id_local = jq.id_local_clima AND jc.id_horario = jq.id_horario_clima;
```

Focos de Incêndio relacionados à poluição e temperatura, por hora do dia

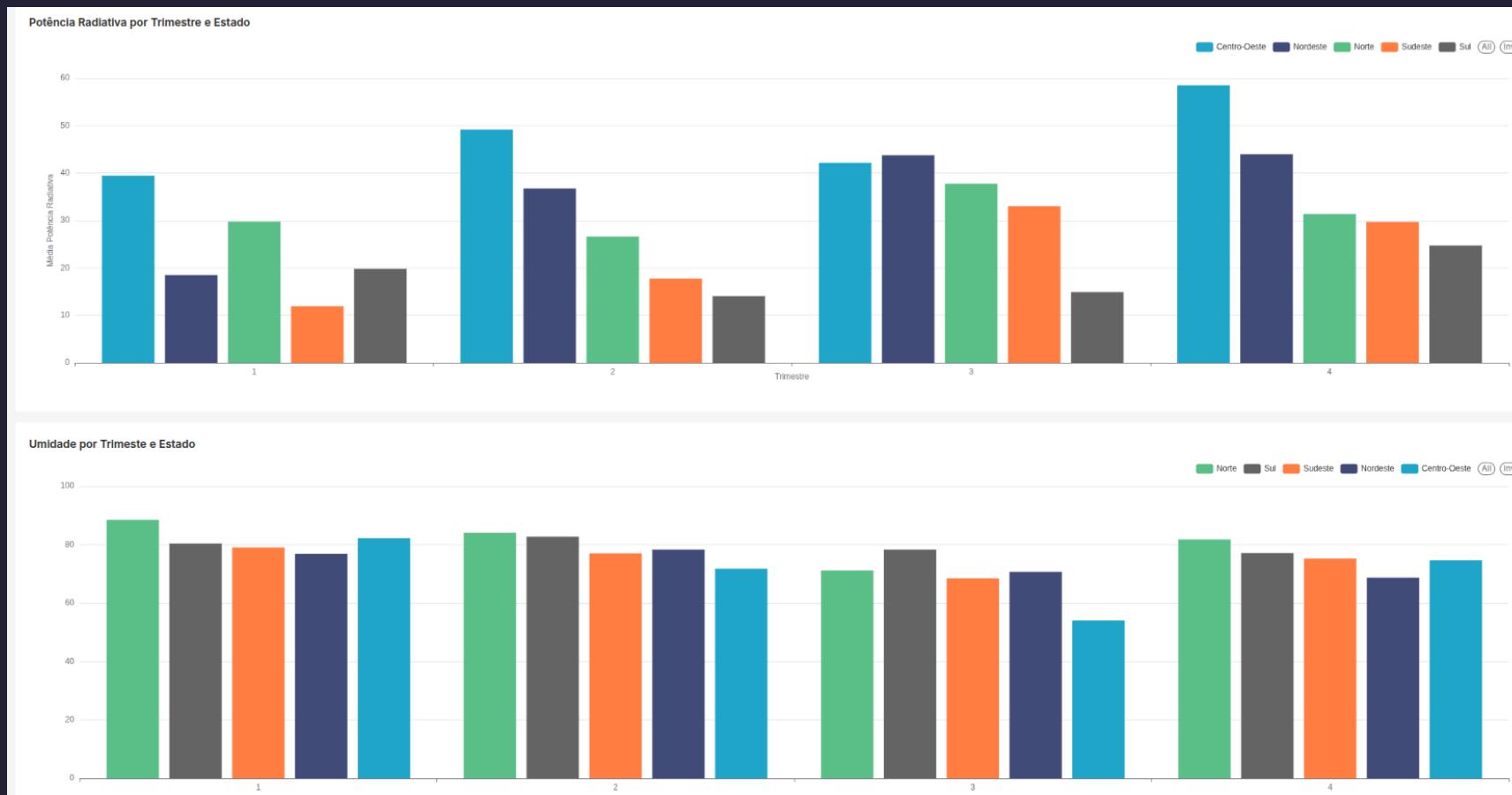


```
WITH joined_clima AS
  (SELECT
    dlc.regiao_uf,
    dd.trimestre,
    avg(fct_clima.umidade_relativa) media_umidade_relativa
  FROM fct_clima
  INNER JOIN public.dim_local_clima dlc ON dlc.id_local = fct_clima.id_local
  INNER JOIN public.dim_data dd ON fct_clima.id_data = dd.id_data
  GROUP BY dlc.regiao_uf, dd.trimestre
  ),
joined_queimada AS
  (SELECT
    d.trimestre,
    dlq.regiao_uf,
    avg(fct_queimada.potencia_radiativa_fogo) as media_potencia_radiativa_fogo
  from fct_queimada
  INNER JOIN public.dim_data d ON d.id_data = fct_queimada.id_data
  INNER JOIN public.dim_local_queimada dlq ON dlq.id_local = fct_queimada.id_local
  GROUP BY d.trimestre, dlq.regiao_uf
  )
SELECT
  jc.regiao_uf,
  jc.trimestre,
  jq.media_potencia_radiativa_fogo,
  jc.media_umidade_relativa
FROM joined_clima jc
  INNER JOIN joined_queimada jq ON jc.regiao_uf = jq.regiao_uf AND jc.trimestre = jq.trimestre
```

Potência Radiativa média por Umidade média para os Estados, com Região indicada



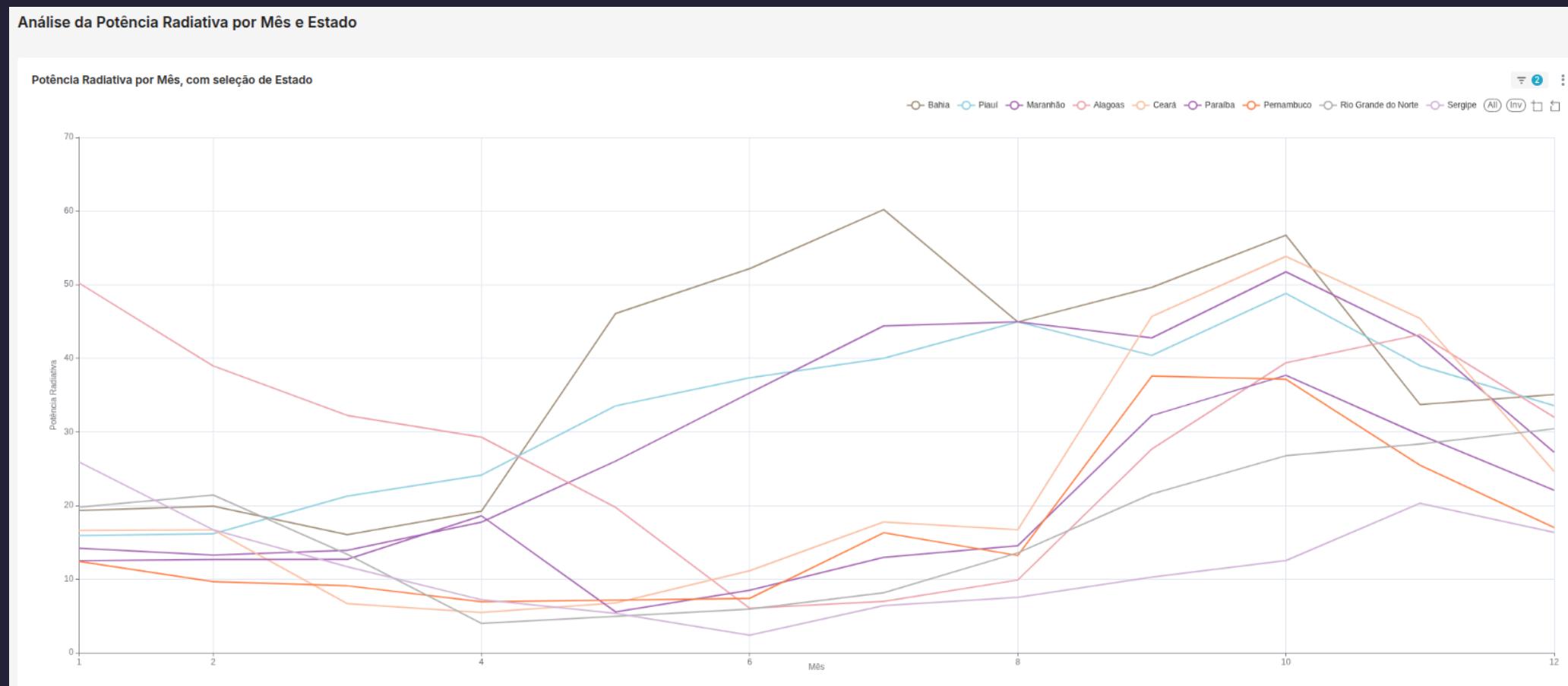
Potência Radiativa média e Umidade média por Trimestre e Região



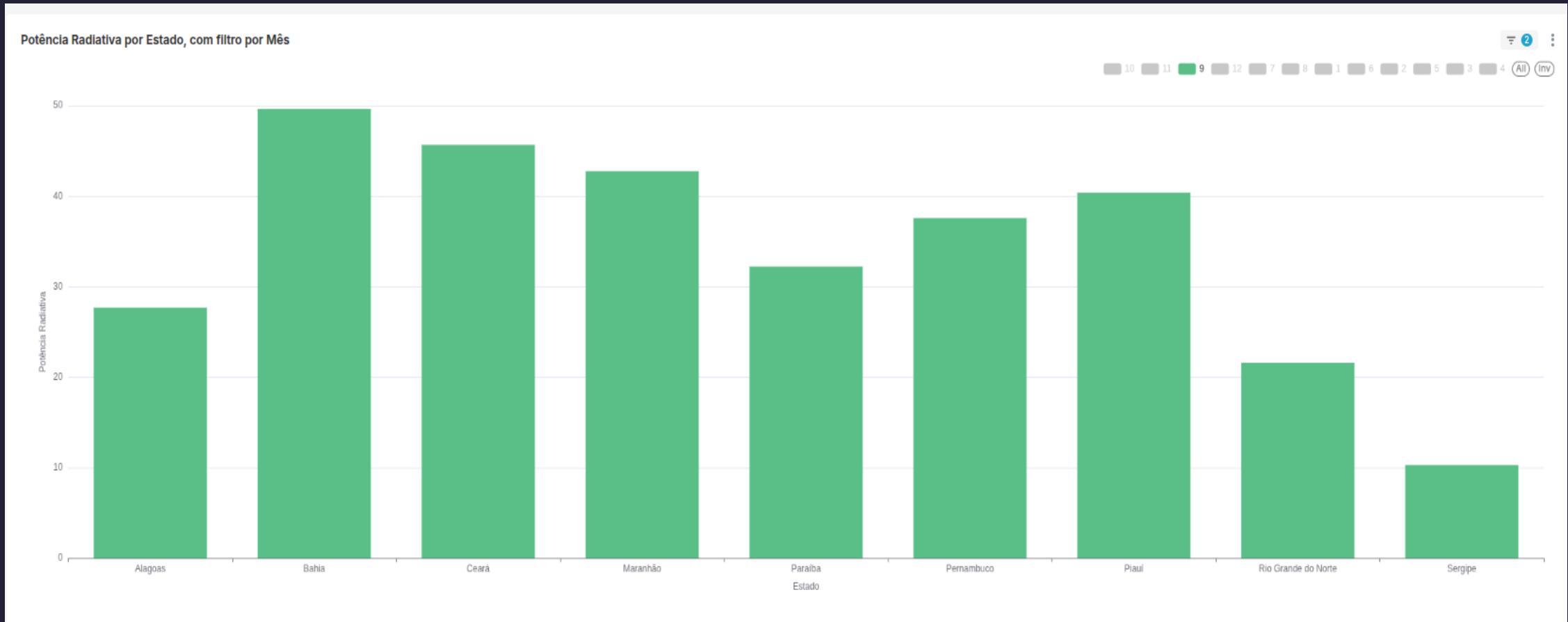
Potência Radiativa por Mês e Estado

```
SELECT
    dl.nome_uf,
    dd.mes,
    avg(fct_queimada.potencia_radiativa_fogo) as media_potencia
FROM fct_queimada
    INNER JOIN public.dim_local_queimada dl ON fct_queimada.id_local = dl.id_local
    INNER JOIN public.dim_data dd ON fct_queimada.id_data = dd.id_data
WHERE fct_queimada.potencia_radiativa_fogo is not null
GROUP BY (dl.nome_uf, dd.mes);
```

Potência Radiativa por Mês com filtro para Estado



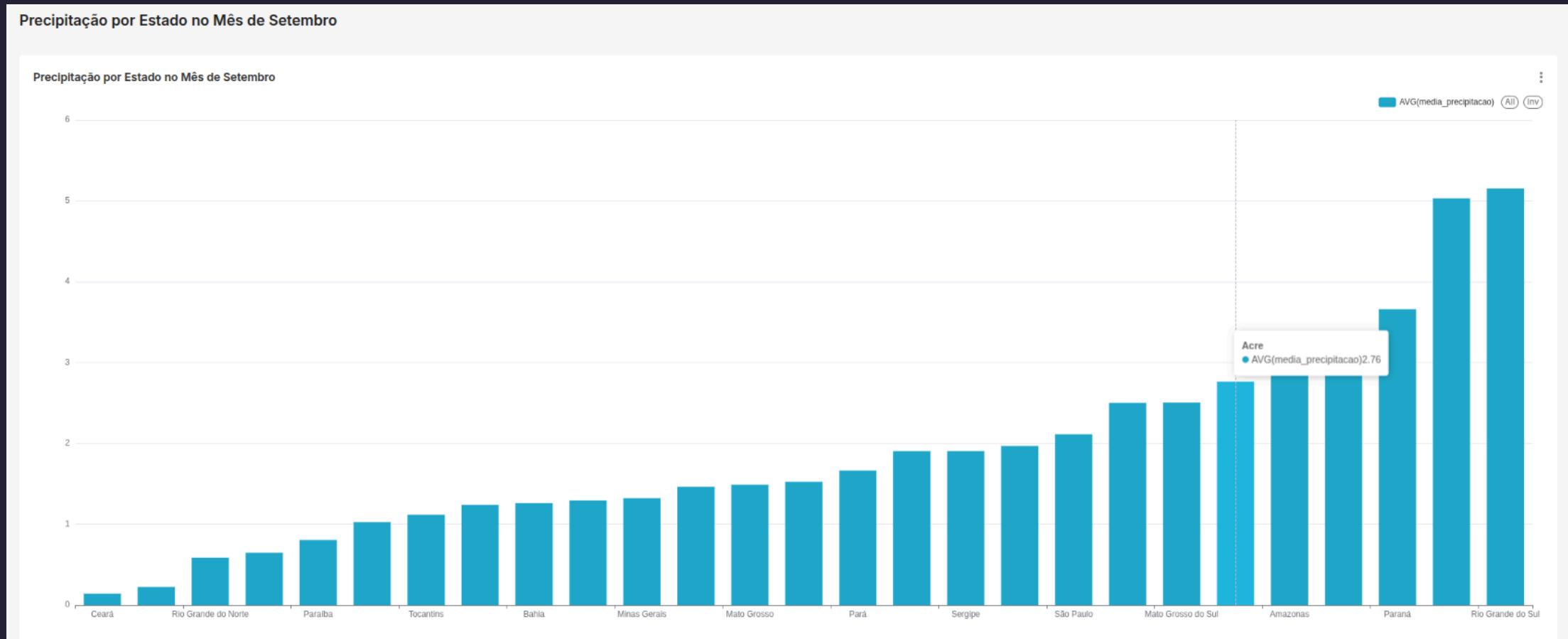
Potência Radiativa por Estado com filtro para Mês



Precipitação por Estado no Mês de Setembro

```
SELECT
    dd.mes,
    dlc.nome_uf,
    avg(fct_clima.precipitacao_dia) as media_precipitacao
FROM fct_clima
    INNER JOIN public.dim_data dd on dd.id_data = fct_clima.id_data
    INNER JOIN public.dim_local_clima dlc on dlc.id_local = fct_clima.id_local
WHERE dd.mes = 9
GROUP BY (dd.mes, dlc.nome_uf)
ORDER BY media_precipitacao DESC;
```

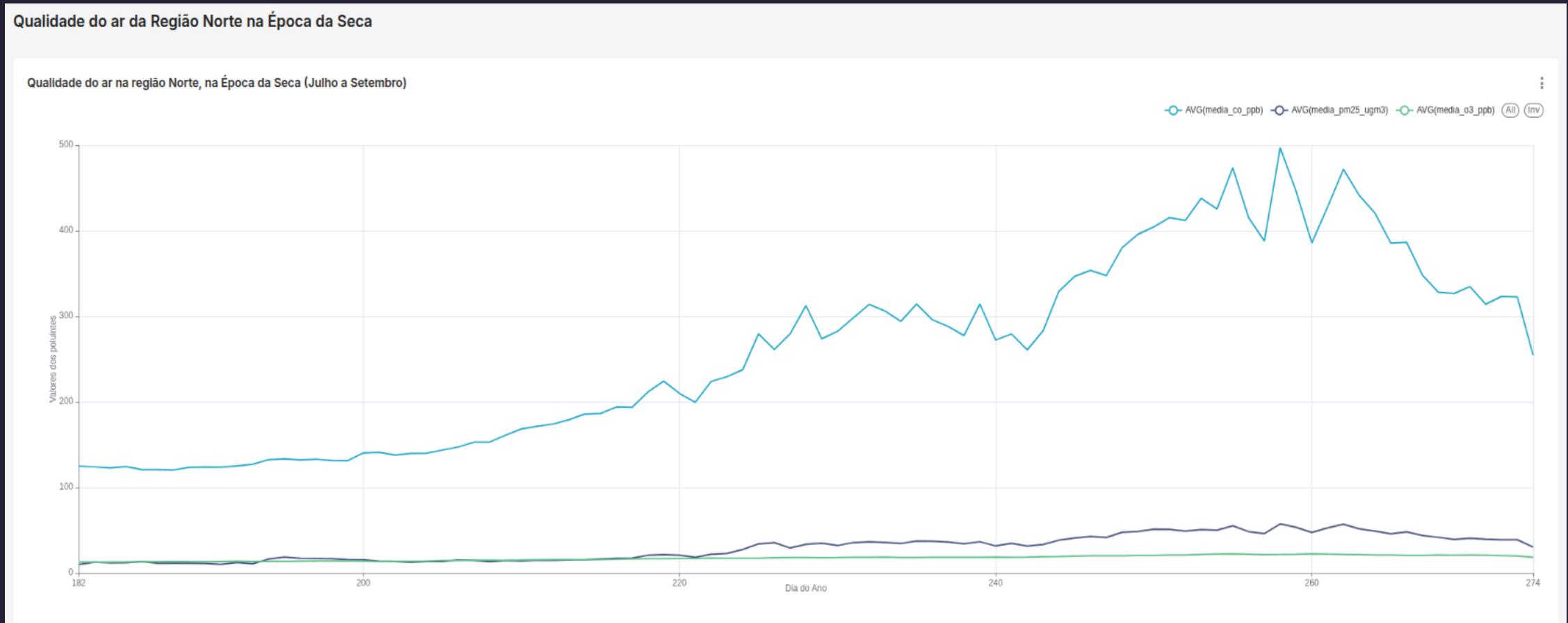
Precipitação por Estado no Mês de Setembro



Qualidade do ar na região Norte na época de secas

```
SELECT
    dd.dia_ano,
    dlc.nome_municipio,
    avg(fct_clima.co_ppb) as media_co_ppb,
    avg(fct_clima.pm25_ugm3) as media_pm25_ugm3,
    avg(fct_clima.o3_ppb) as media_o3_ppb
FROM fct_clima
    JOIN public.dim_data dd on dd.id_data = fct_clima.id_data
    JOIN public.dim_local_clima dlc on fct_clima.id_local = dlc.id_local
WHERE dd.mes >6 AND dd.mes < 10 AND regiao_ue = 'Norte'
GROUP BY dd.dia_ano, dlc.nome_municipio;
```

Qualidade do ar na região Norte na época de secas



RollUp Precipitação média por Ano, Região, mês e Estado

```
SELECT
    dd.ano,
    dd.mes,
    dl.nome_uf,
    dl.regiao_uf,
    avg(fct_clima.precipitacao_dia) as media_precipitacao

FROM fct_clima
    INNER JOIN dim_data dd on dd.id_data = fct_clima.id_data
    INNER JOIN dim_local_clima dl on dl.id_local = fct_clima.id_local
GROUP BY ROLLUP
    ((dd.ano, dl.regiao_uf), (dl.nome_uf, dd.mes))
```

Conclusões

- Correlações identificadas entre dados de queimadas e clima (Ex: FRP x Umidade)
- Identificação de biomas e estado em maior risco de focos de incêndios, além de estados e cidades com maior incidência
- Análise de comportamentos específicos para cada estado quanto a intensidade de queimadas

Referências

- Google Cloud Platform
- Terraform

Código Fonte

Github



Perguntas?

UwU