

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: PM2.5, CO, NO₂, O₃, SO₂

Fontes de Dados

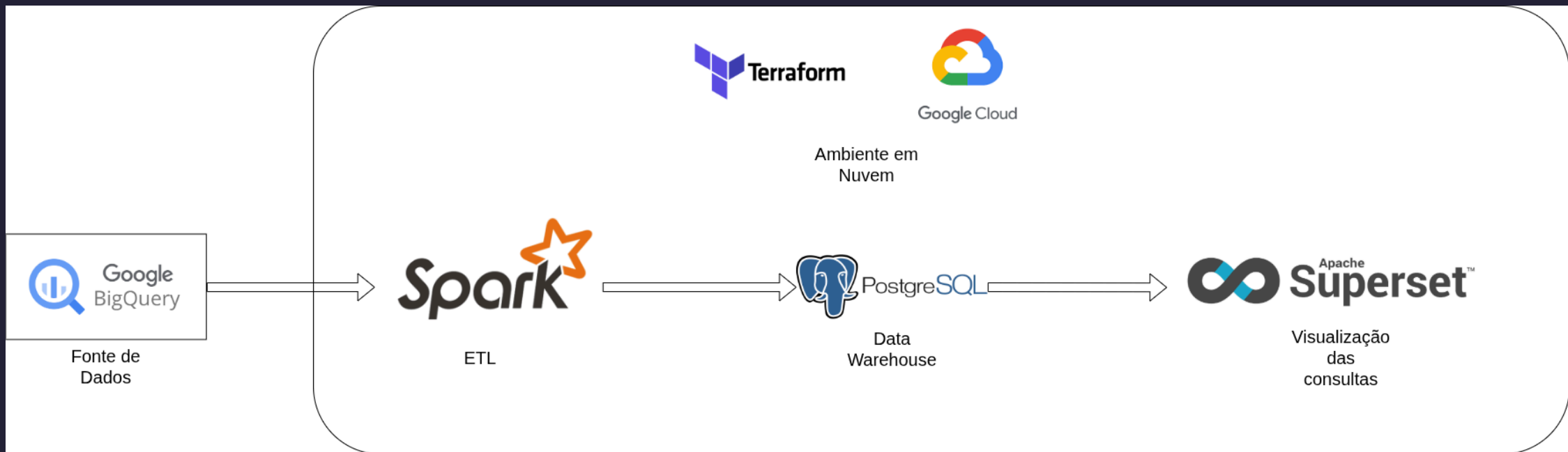
Geográficos - IBGE

- 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 (desde 2003 (alguém corrige essa data))
- Geração de visualizações baseadas em consultas analíticas para tomada estratégica de decisões

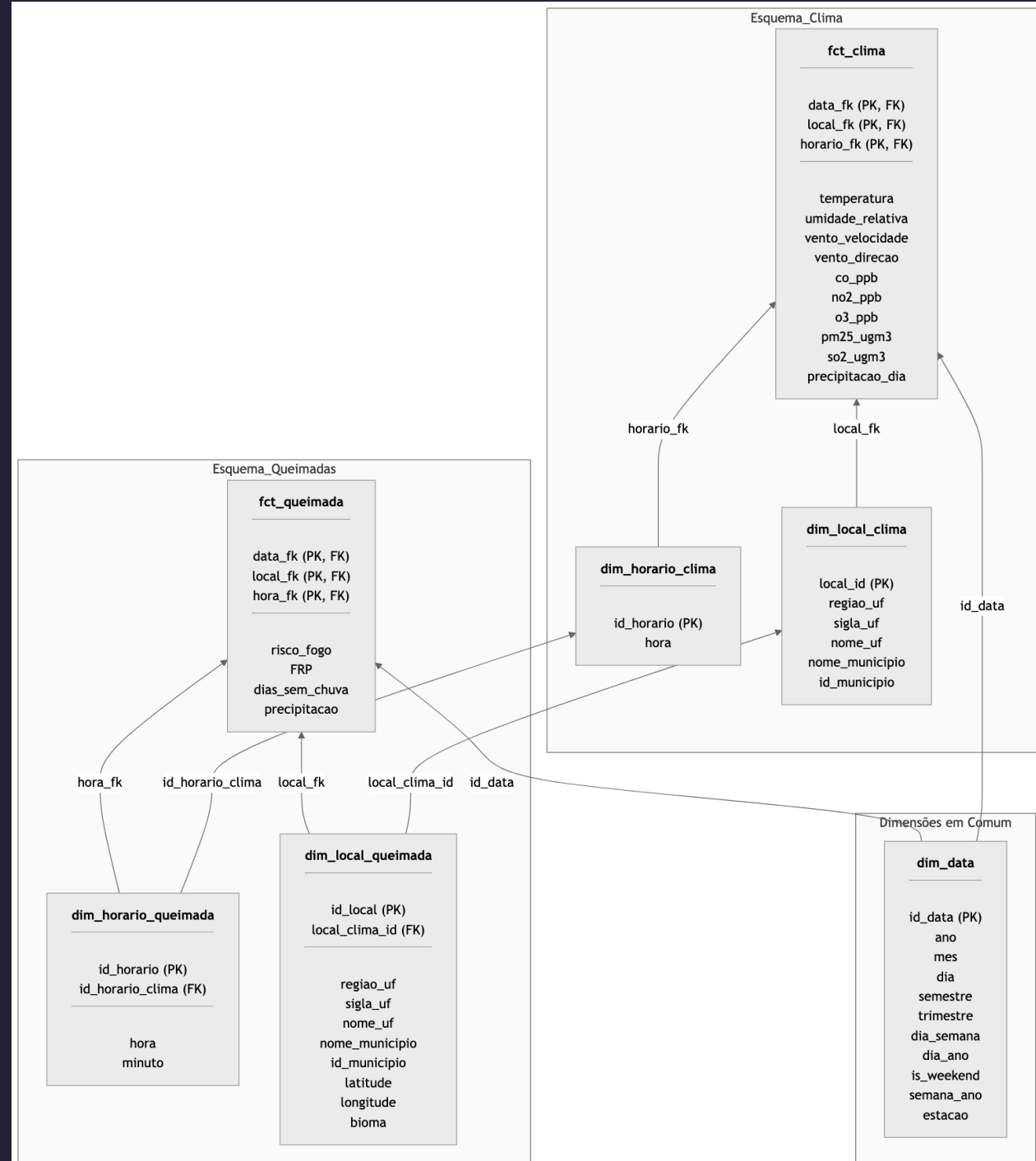
Visão geral da Arquitetura



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

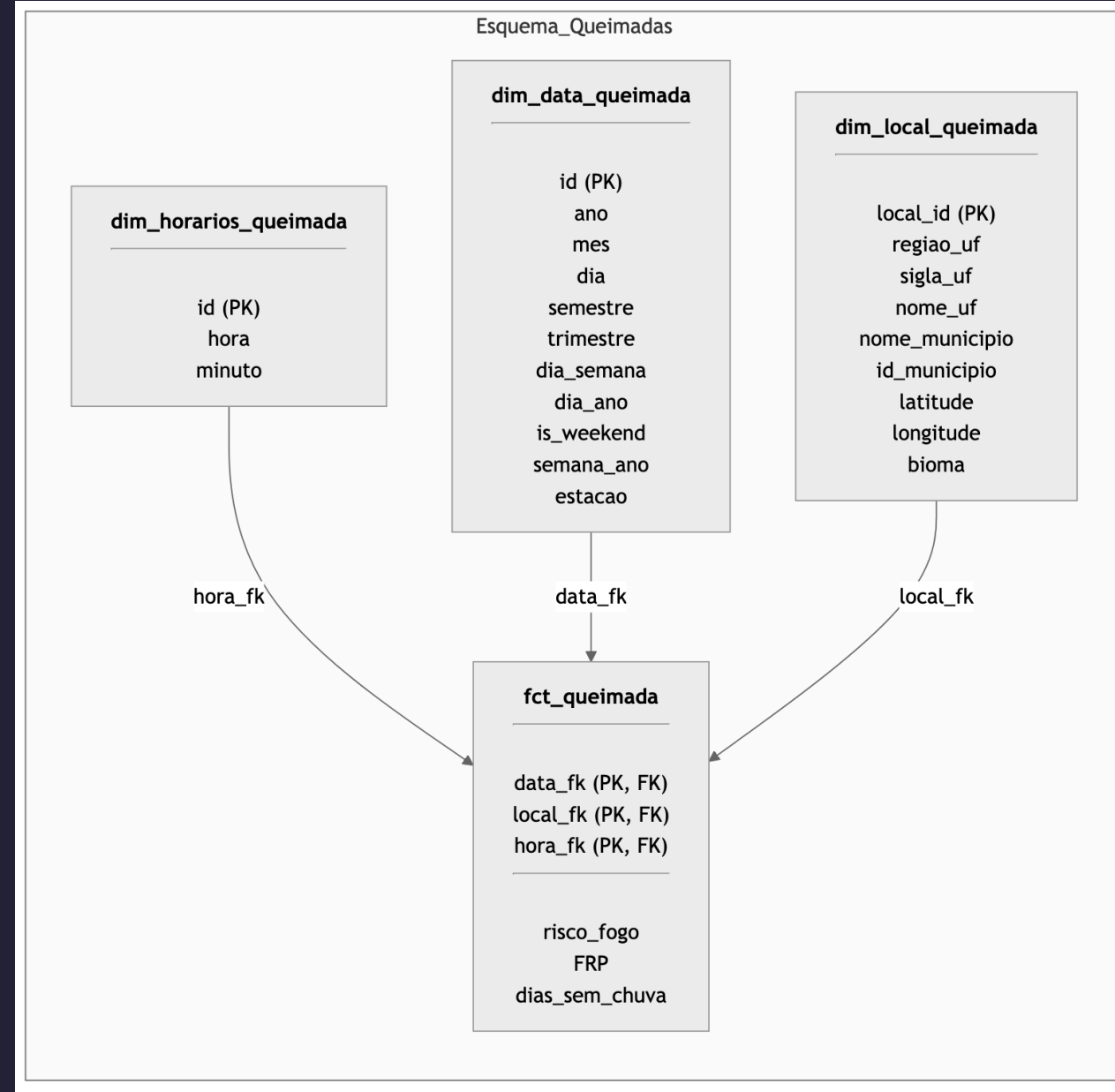


Organização do Data Warehouse

Queimadas

Esquema estrela corrigido

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

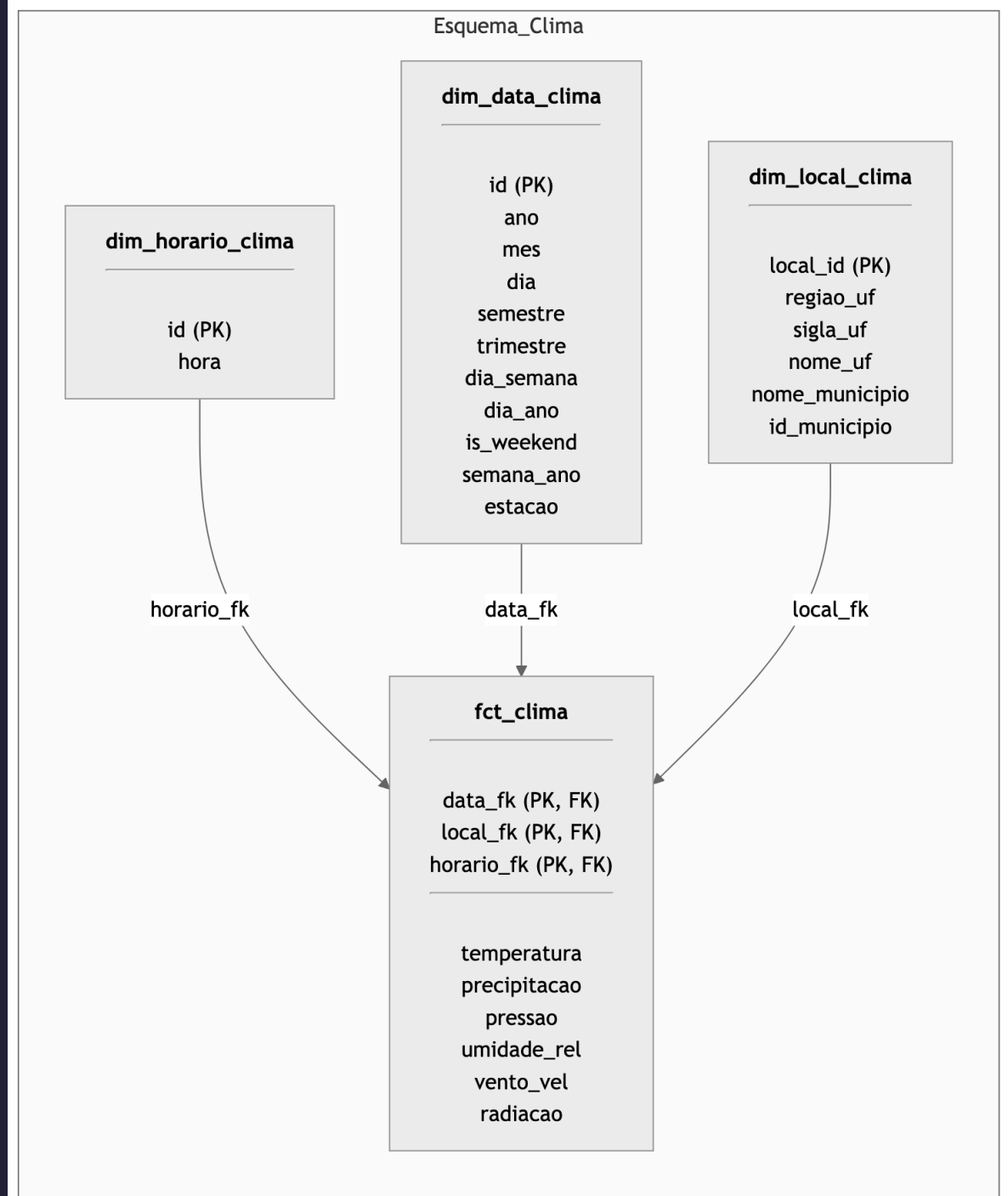


Organização do Data Warehouse

Clima

Esquema estrela corrigido

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

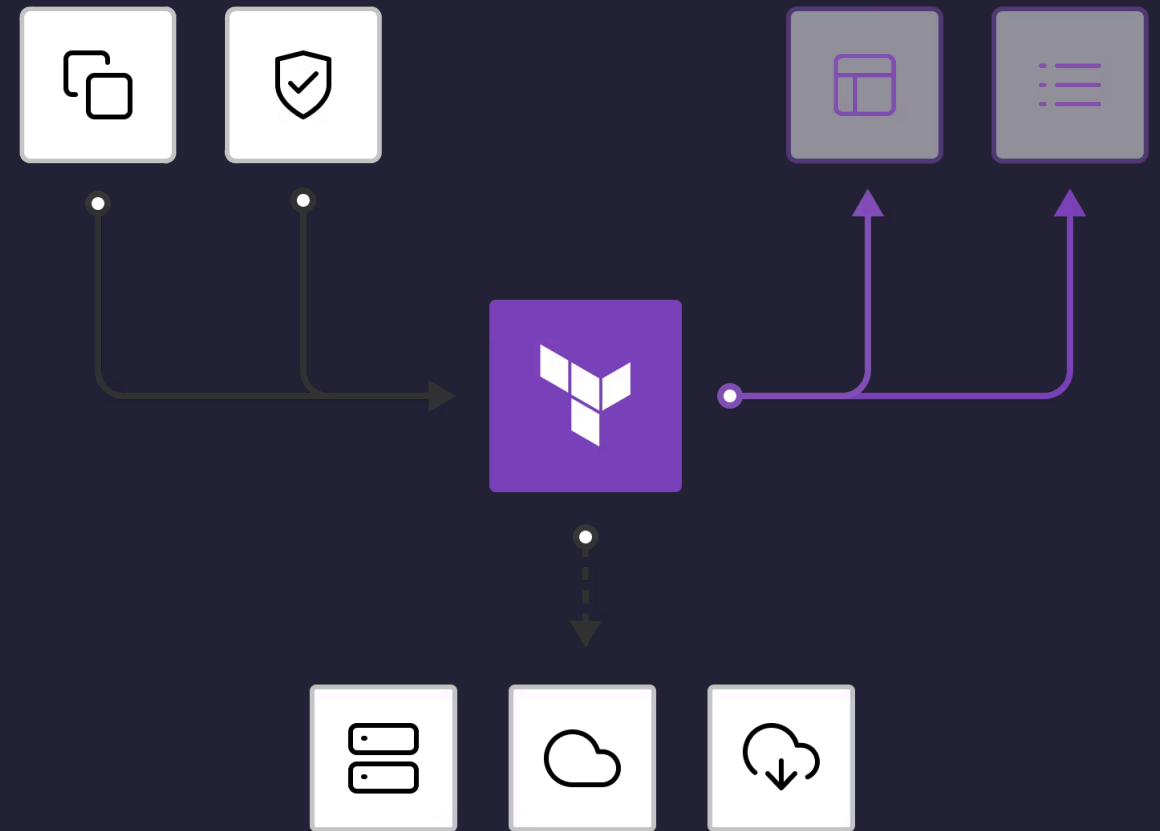


Infraestrutura

Provisionamento com Terraform

Provisionamos com Terraform:

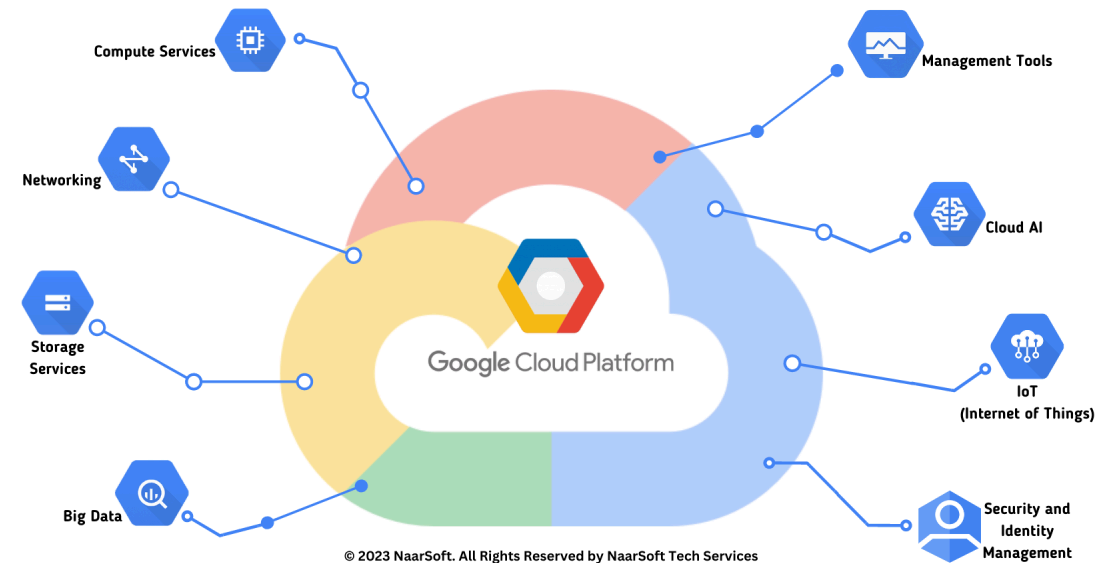
- Base de dados PostgreSQL
- Jobs Spark para ETL
- Superset para visualização dos dados



Infraestrutura

Google Cloud Platform

Tava de graça 🙏 (Deram 1800 reais de crédito pra gente)



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)

Índices criados para colunas de frequente acesso:

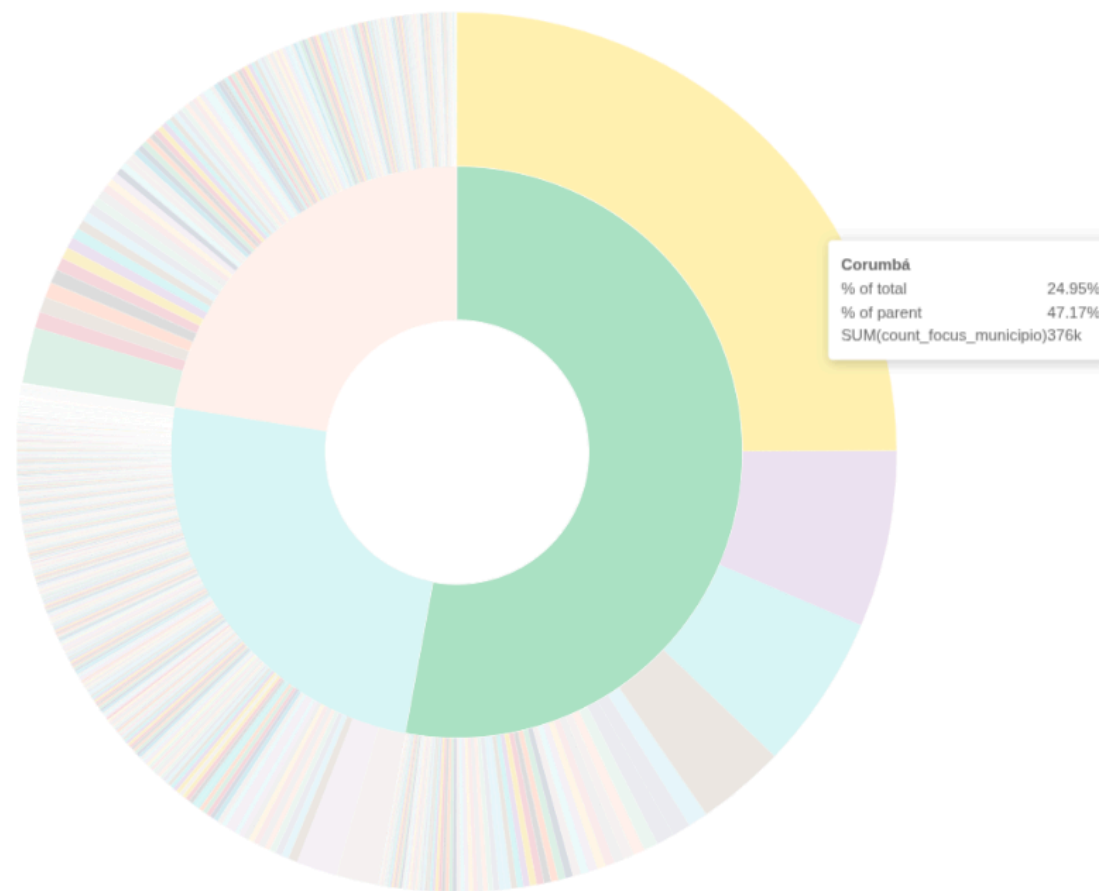
- Chaves estrangeiras nas tabelas de fatos e dimensões do esquema Queimadas
- Timestamp, Mes e ano para Data
- Hora para Horário
- UF e município para Local

Consultas analíticas e visualização

- Consultas implementadas em SQL
- Visualizações construídas com Apache Superset

Queimadas por Estado e Município

Contagem de queimadas por estado e município



Risco de Fogo por Bioma e Estado



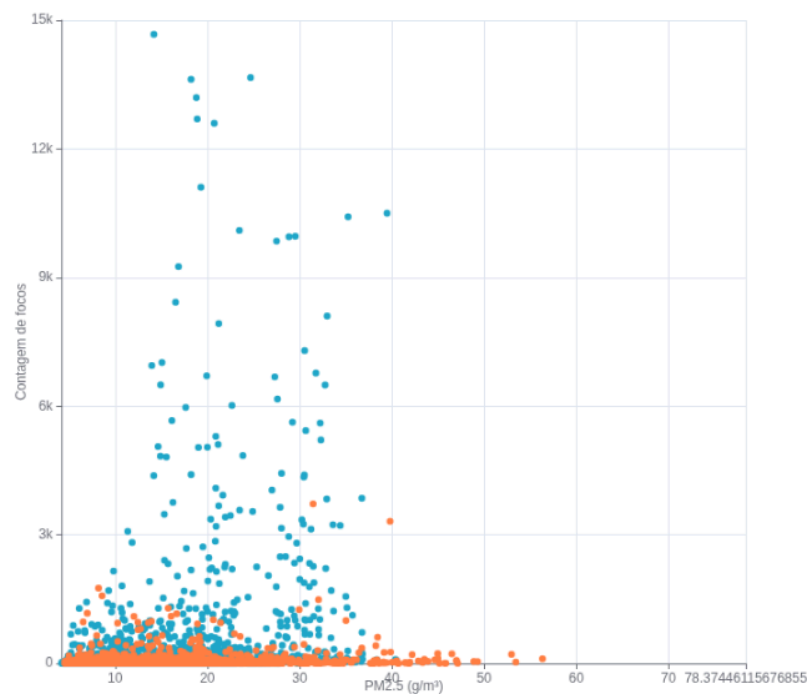
Risco de Fogo por Estado e Bioma



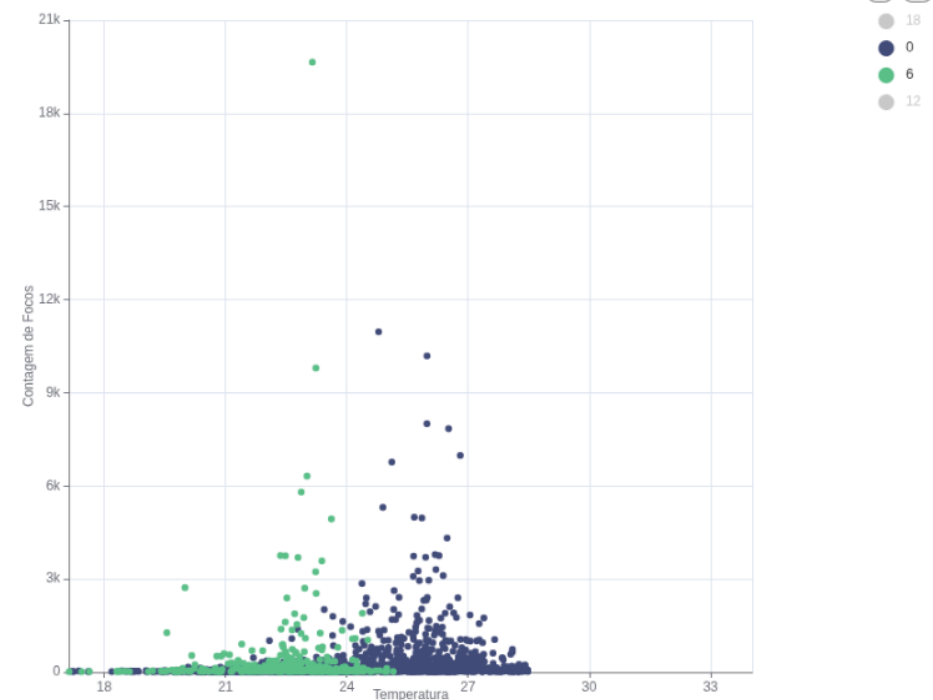
Focos de Incêndio relacionados à poluição e temperatura, por hora do dia

Análise da contagem de focos e poluição agregada por município e hora

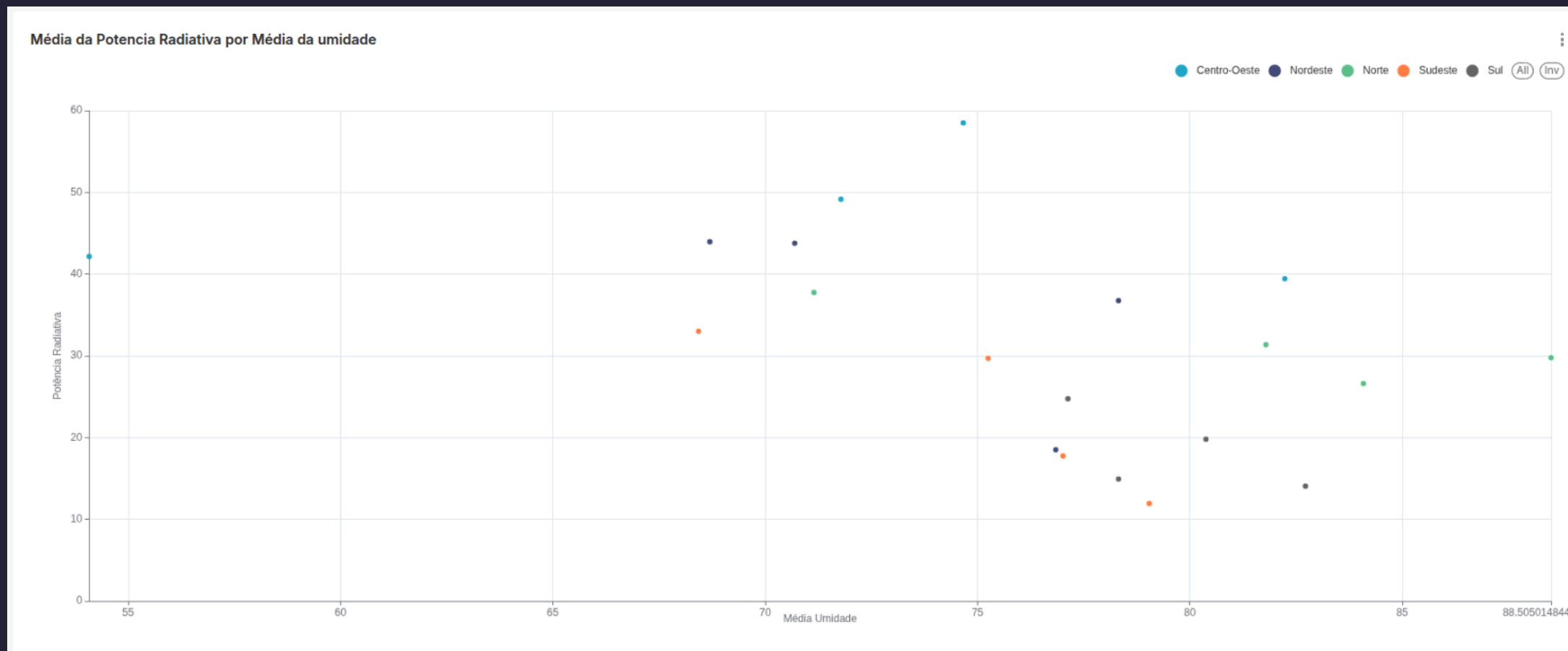
Contagem de focos de queimada por PM2.5 e colorido por hora



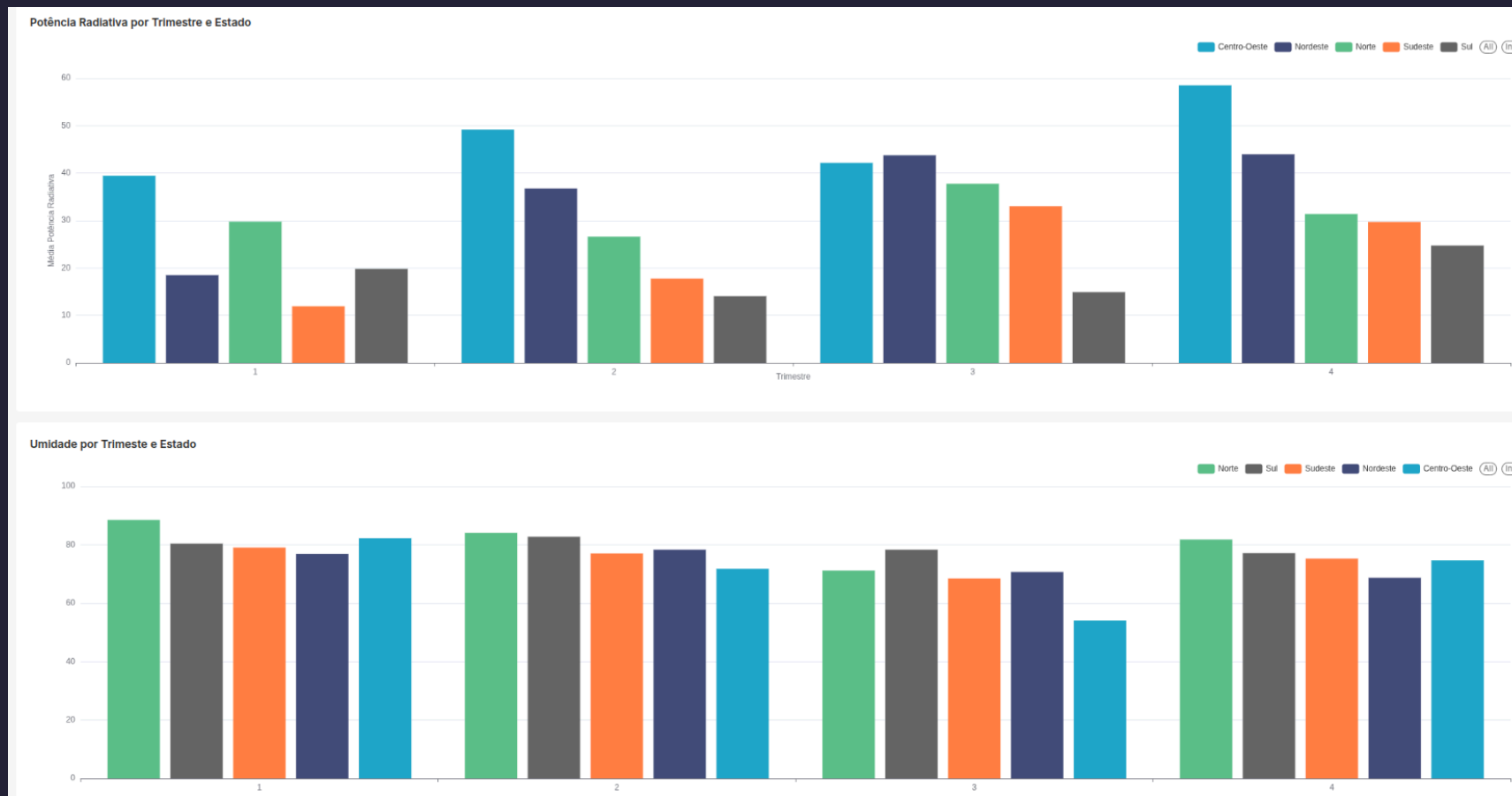
Contagem de focos por Temperatura e colorido por hora



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



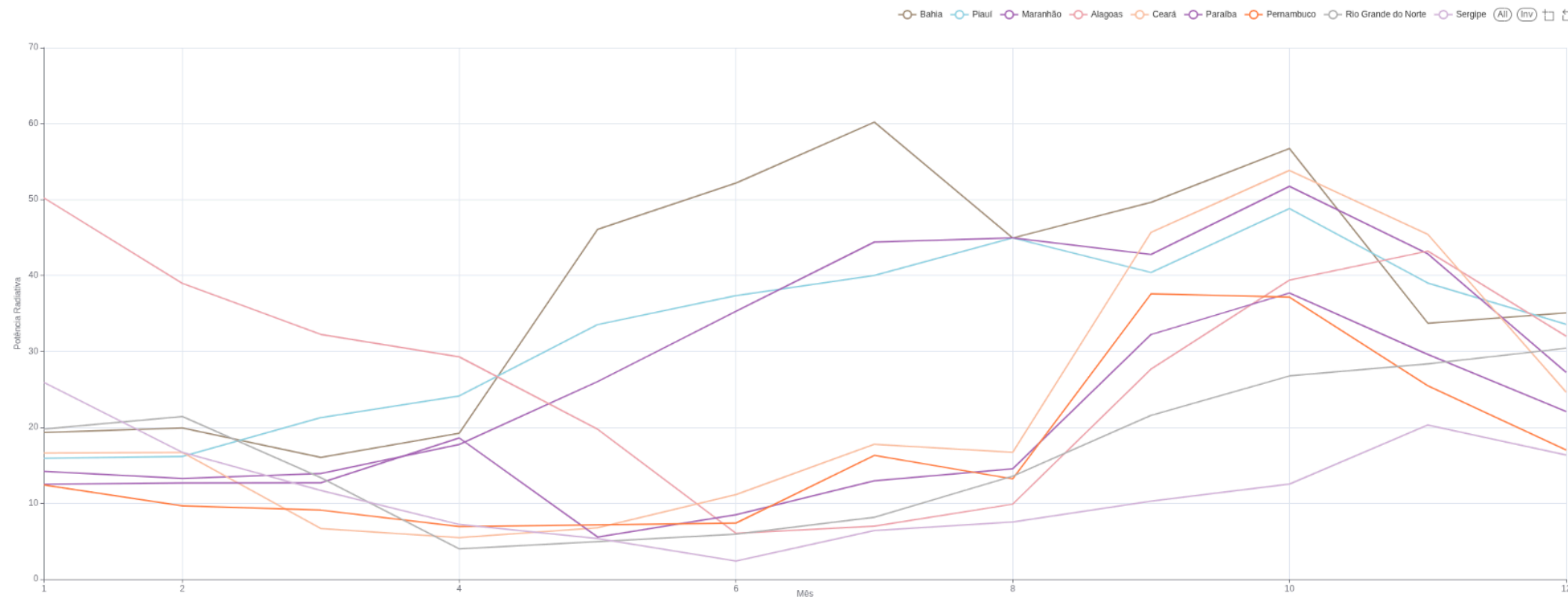
Potência Radiativa média e Umidade média por Trimestre e Região



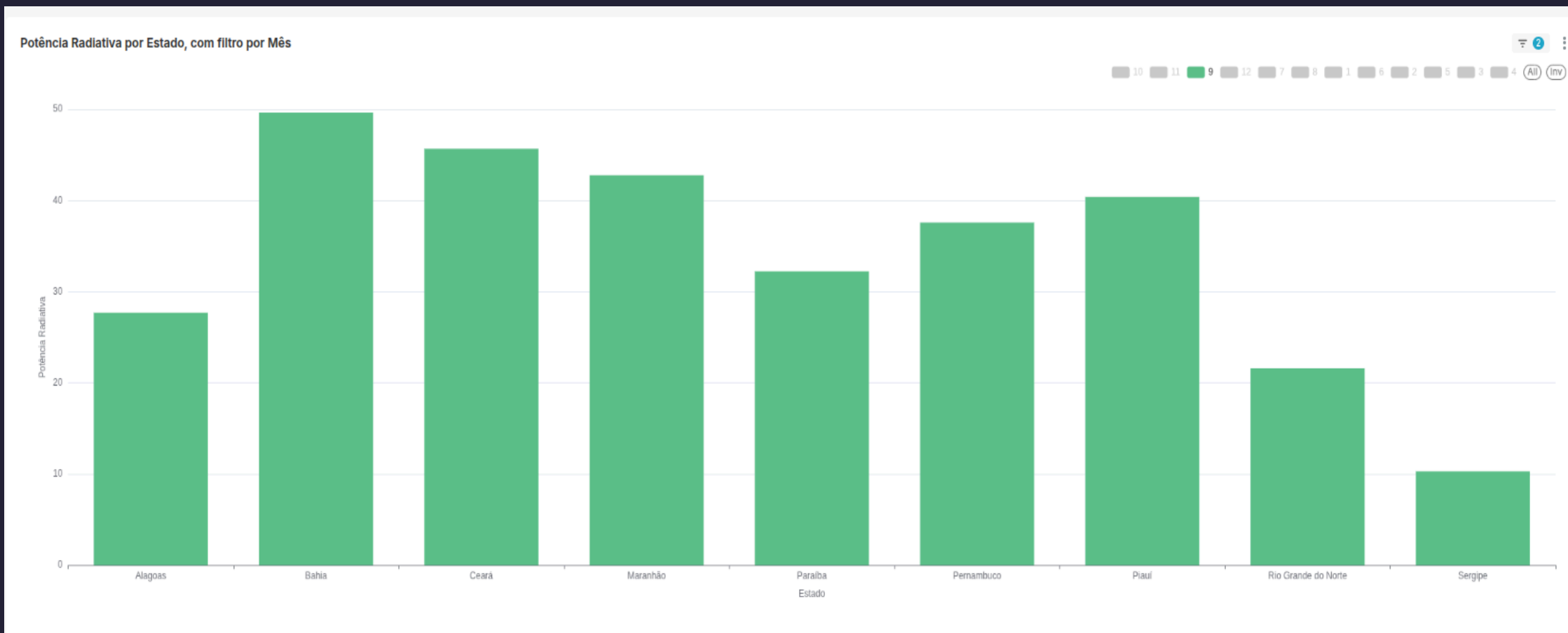
Potência Radiativa por Mês e Estado

Análise da Potência Radiativa por Mês e Estado

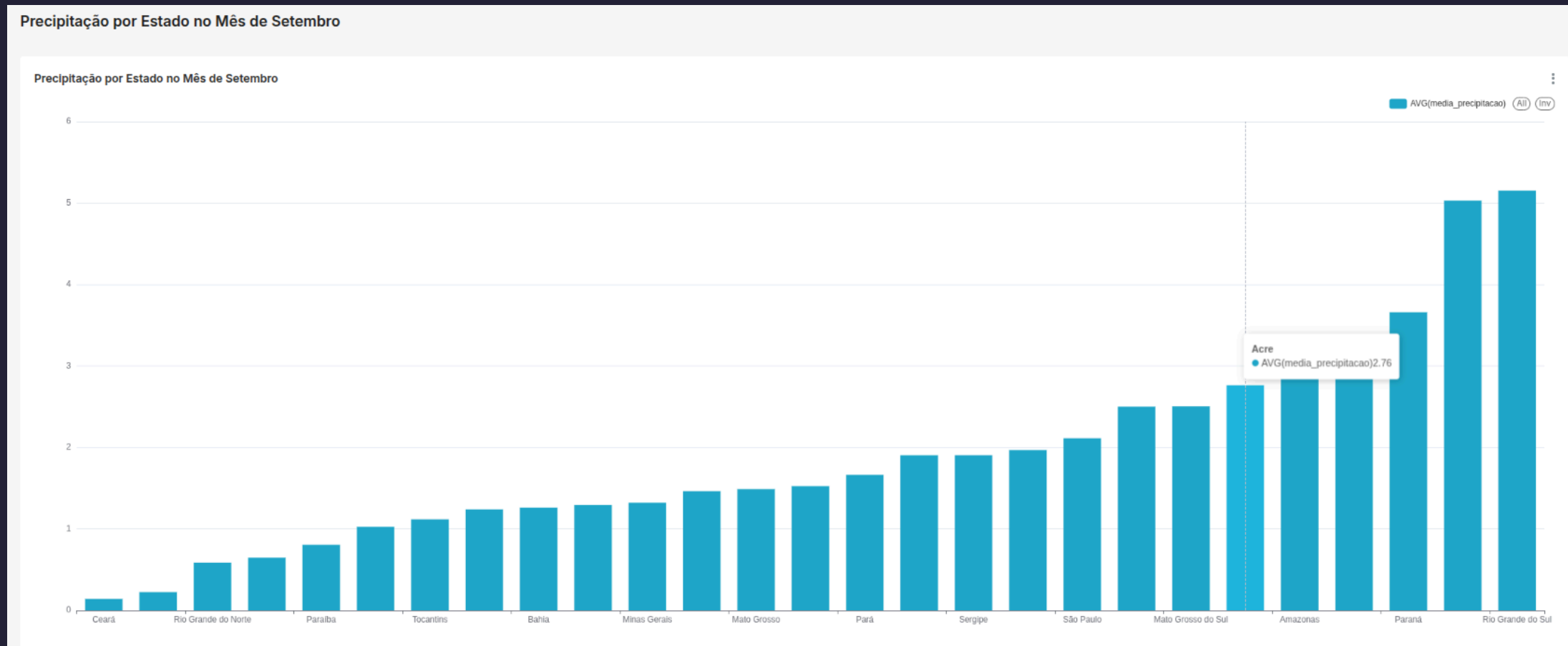
Potência Radiativa por Mês, com seleção de Estado



Potência Radiativa por Estado com filtro para Mês



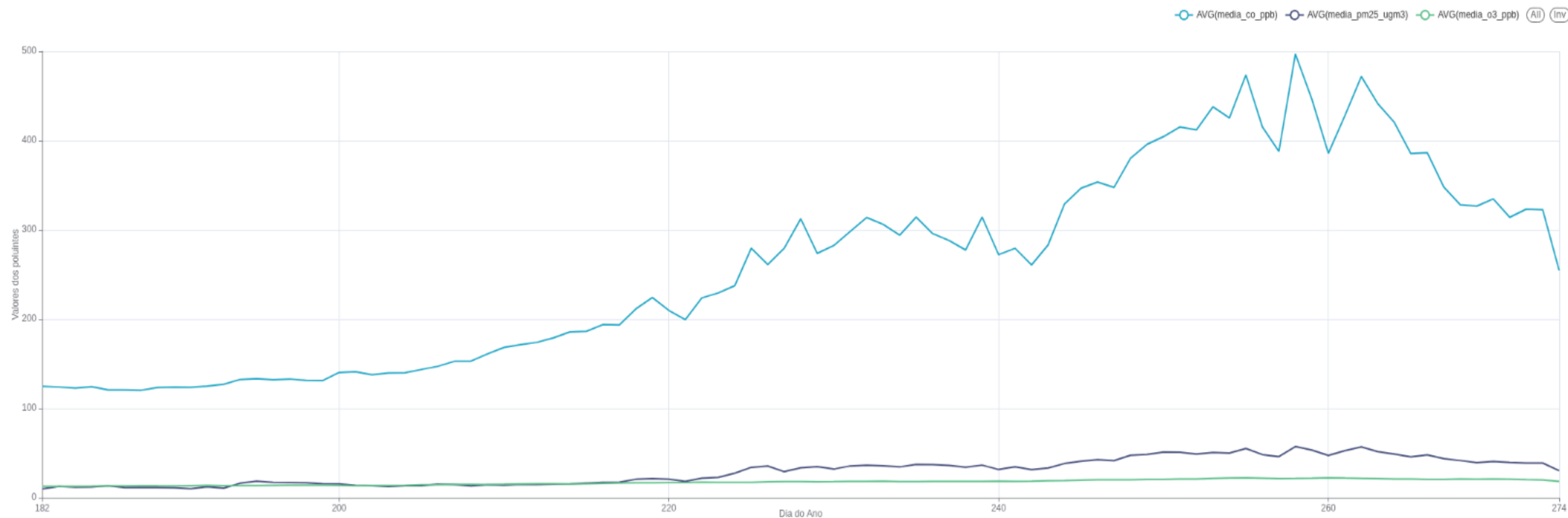
Precipitação por Estado no Mês de Setembro



Qualidade do ar na região Norte na época de secas

Qualidade do ar da Região Norte na Época da Seca

Qualidade do ar na região Norte, na Época da Seca (Julho a Setembro)



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