

A topographic map of South America, showing the continent's diverse terrain with green lowlands, brown highlands, and blue rivers and lakes. The map is partially obscured by a teal box on the right.

BReATH

PRESENTATION

Brazilian
Research of
Atmosphere
Towards
Health

INTEGRANTES



233840

**Elton Cardoso do
Nascimento**

Implementação e
integração do Banco
de dados

234720

**Gabriel Costa
Kinder**

Limpeza, tratamento e
inserção do dataSUS

218733

**João Pedro de
Moraes Bonucci**

Limpeza, tratamento e
inserção dos dados
climáticos

240106

**Lucas Otávio
Nascimento de
Araújo**

Ferramentas de
análise sob o banco de
dados

01 PROBLEMA

**Relacionando o clima com
doenças respiratórias**

TEMA

O data set consiste em um banco de dados relacional que agrega

- Dados climaticos
- Dados de qualidade do ar
- Dados de doenças respiratórias

Do território Brasileiro entre os anos de 2000 e 2020 (nem todos os datasets contém todo o periodo). Com isso, ambicionamos encontrar ou fortalecer relações entre condições ambientais e doenças respiratórias no nosso país fornecendo dados específicos por cidade e mês ao longo de anos.

MOTIVAÇÃO E CONTEXTO GERADOR

Nossa motivação para o problema nasceu devido a duradoura estiagem que estamos passando somado ao contexto pandêmico onde doenças respiratórias são um tema de foco. Por isso queríamos trazer algo relacionado a saúde e ao tema debatido, mas contribuindo com bancos ainda não tão explorados.

Em 2017, dois problemas respiratórios estavam entre as dez maiores causas de morte do país. Considerando isso, escolhemos como objetivo prever a incidência de doenças respiratórias por meio de dados ambientais (clima, poluição).

Nosso objetivo se alinha com os Objetivos de Desenvolvimento Sustentável da ONU, mais especificamente com o objetivo 3, “Garantir o acesso à saúde de qualidade e promover o bem-estar para todos, em todas as idades”.



02

DATASET

- Fontes
- Modelo conceitual
- Modelo lógico

FONTES DE DADOS UTILIZADAS

DADOS CLIMÁTICOS

01

Climate Weather
Surface of Brazil
- Hourly

DADOS GEOGRÁFICOS

02

Banco de
Coordenadas
Geográficas das
Cidades
Brasileiras

03

Municípios do
IBGE

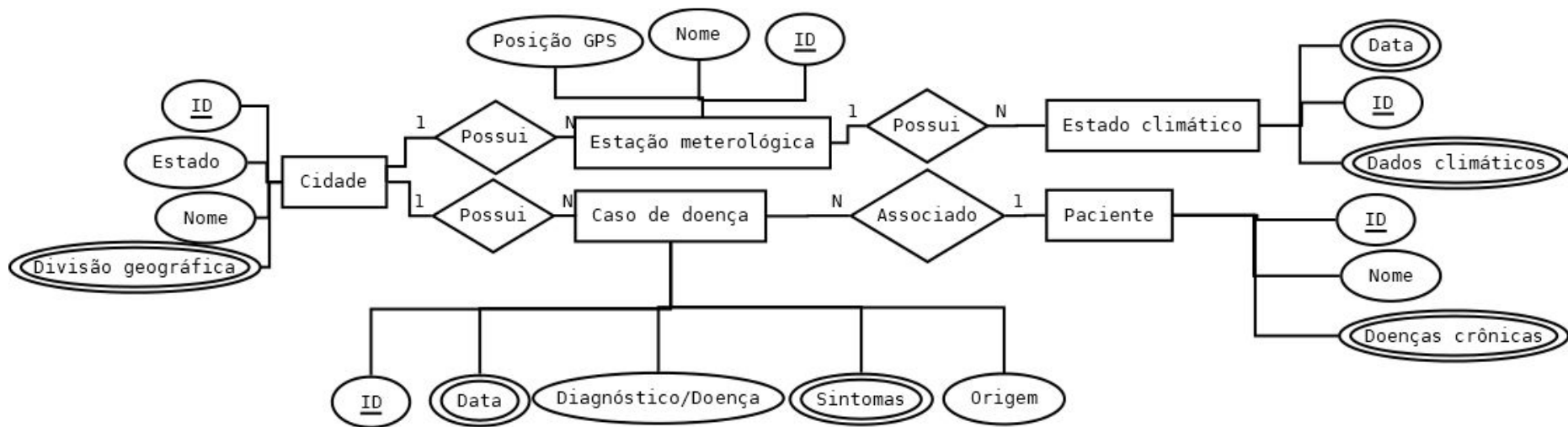
DADOS DE SAÚDE

04

SRGA
Banco de dados
síndrome
respiratória aguda
grave 2013 - 2018



MODELO CONCEITUAL



MODELO LÓGICO



SRAG(_id_, DT_NOTIFIC, ID_MUNICIP, SEM_NOT, SG_UF_NOT, DT_SIN_PRI, DT_NASC, NU_IDADE_N, CS_SEXO, CS_GESTANT, CS_RACA, CS_ESCOL_N, SG_UF, ID_MN_RESI, ID_OCUPA_N, VACINA, FEBRE, TOSSE, ...)

CLIMA(_id_, date, precipitacao, pressao_at_max, pressao_at_min, radiacao, temp_max, temp_min, umidade, max_vent, velocidade_vent, region, state, station , lat , lon , elvt)

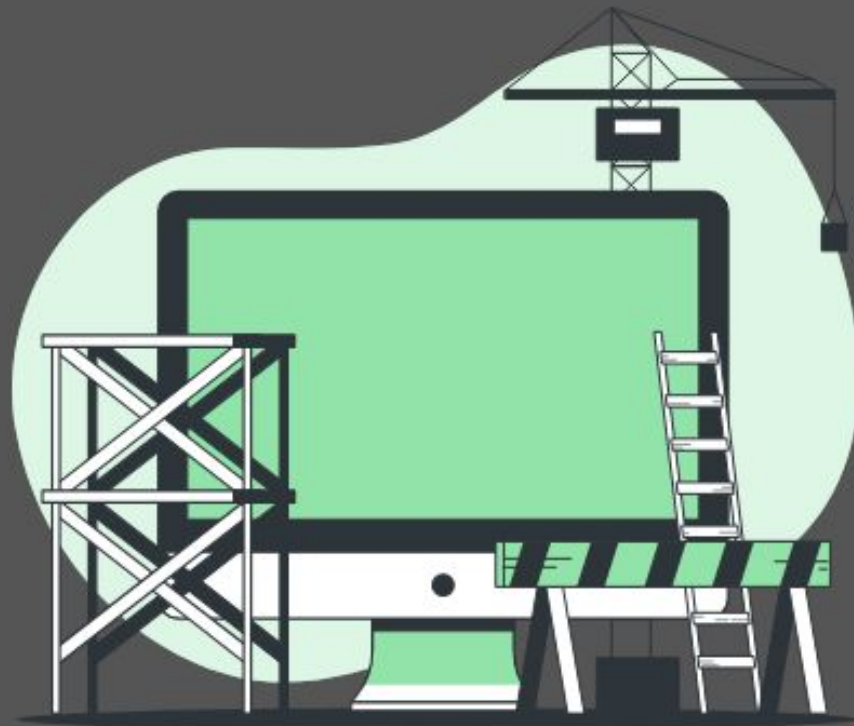
Estacoes(_Id_, Stacao, Regiao, UF,Codigo, Prim_data, alt, lon, lat)

Cidades(_Id_, UF, Nome_UF, Mesorregião Geográfica, Nome_Mesorregião, Microrregião Geográfica, Nome_Microrregião, Município, Código Município Completo, Nome_Município, lat, lon)

03

OPERACÕES

- Operações de preparo
- Implementação física
- Integralização



DATASUS

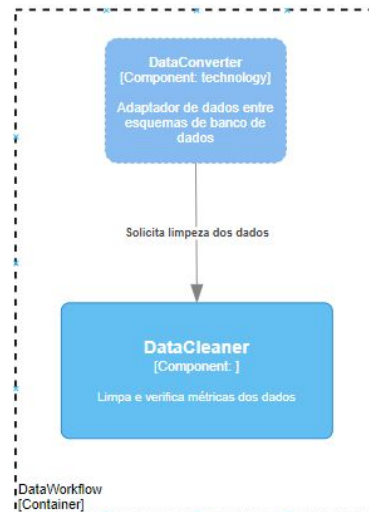
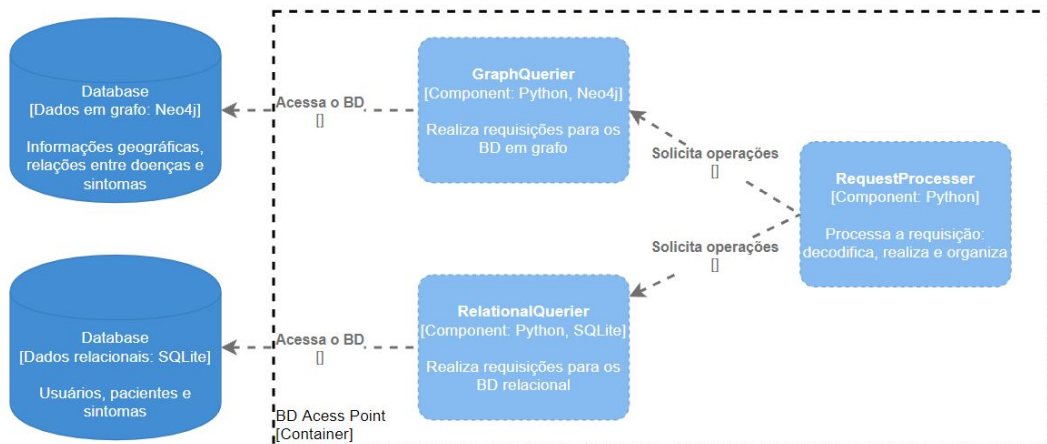
- Extração
 - Dados de diagnósticos referentes aos anos de 2013-2018 obtidos do dataSUS
- Agregação
 - Unificar as tabelas de diferentes anos
 - Juntar informações da tabela do IBGE (nome das Cidades)
 - Juntas informações de coordenadas geográficas (latitude e longitude)
- Tratamento
 - Remover cidades inválidas
 - Arrumar índices
- Integração
 - Integrar os dados do dataSUS com os outros dados de nosso BD

DADOS CLIMÁTICOS

- Extração
 - Extrair dados provenientes de estações climáticas
- Transformação
 - Padronização das colunas entre as tabelas de diferentes regiões do país
- Agregação
 - Junção dos dados climáticos de diferentes estações de coletas com suas respectivas estações e localidades
- Tratamento
 - Remover dados incompletos
 - Normalizar dados
 - Correção de dados com erros de digitação
 - Renomeação e exclusão de colunas do banco de dados
- Integração
 - Integração com os dados de saúde do banco SRAG

INTEGRALIZAÇÃO

- Acesso como serviço para outras aplicações



04

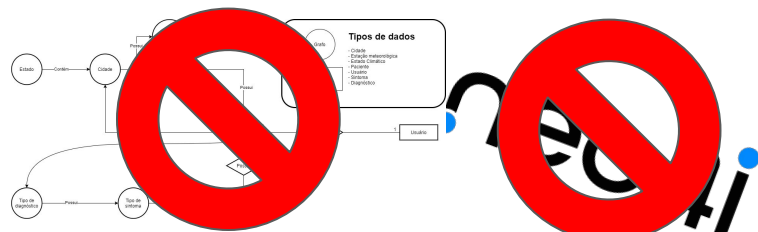
EVOLUÇÃO

Como o projeto
evoluiu no tempo



MODIFICAÇÕES NO PROJETO

- Alteração do período analisado: 2008-2018 → 2013-2018
 - Incompatibilidade entre os modelos de dados
- Foco apenas no modelo relacional
 - Dificuldades de integrar com o modelo de grafo
 - Falta de vantagens no modelo de grafo
 - Dependência no Neo4j - proprietário e necessita de instalação



+

10 GB



Dificuldade de processamento

Volume de dados muito grande

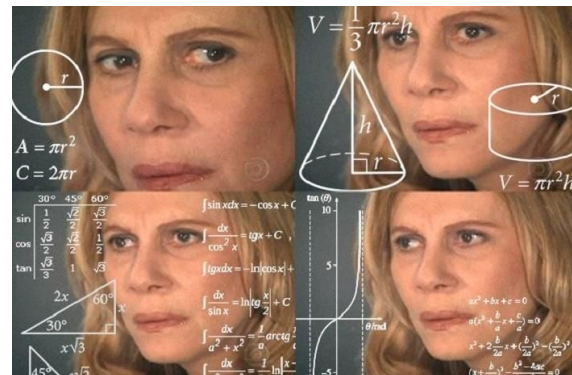
Falta de memória e lentidão

SGBD não permite acessos paralelos

Muitos dados
disponíveis, com
alguns bilhões a
mais no caminho



PLANEJAMENTO COMPLEXO



- Juntar domínios em um único projeto foi mais complexo que o esperado
 - Ideia inicial: juntar trabalhos → aplicação final com esforços somados
 - Resultado: juntar trabalhos → esforço multiplicado
- Dificuldade de integrar o BD com as outras partes do sistema
 - Volume de dados
 - Interfaces e operações

05

ANÁLISE

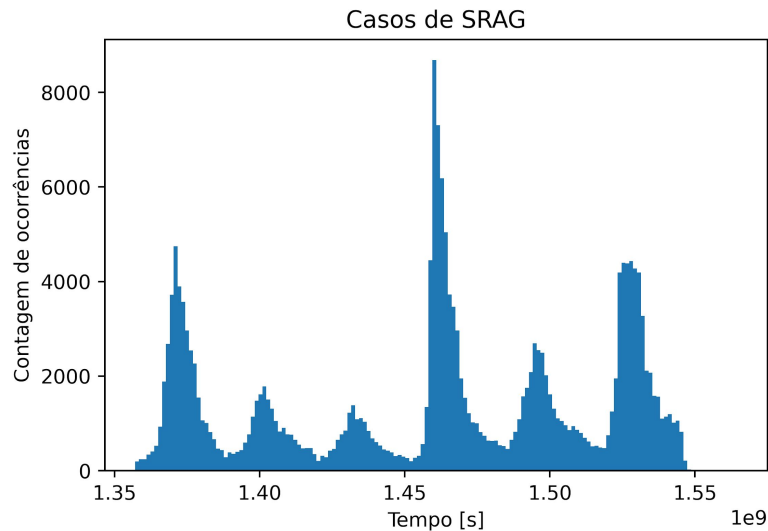
- Conjunto de perguntas
- Consultas iniciais
- Machine Learning



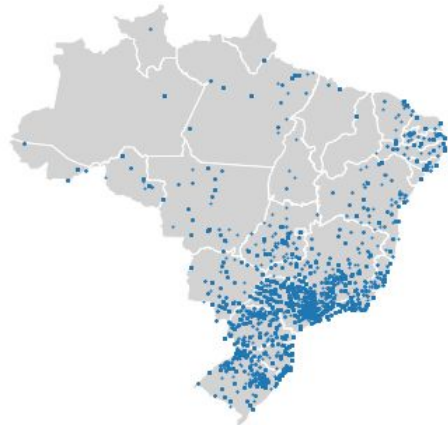
EXISTE ALGUMA SAZONALIDADE NOS CASOS DE SRAG?

Para cada estação, selecionamos os casos registrados entre suas datas de início e fim

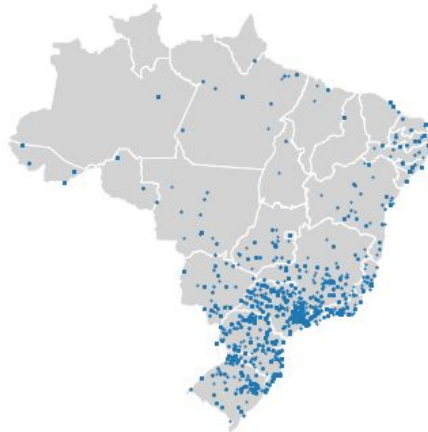
Sim, existe sazonalidade, e os casos se concentram nos períodos de outono e inverno



Outono de 2016



Inverno de 2016



Primavera de 2016



Verão de 2016-2017



QUAIS SÃO OS SINTOMAS MAIS COMUNS? (NÚMERO E INCIDÊNCIA)

Consulta

```
SELECT COUNT(FEBRE) FROM SRAG WHERE FEBRE = "1.0";  
SELECT COUNT(TOSSE) FROM SRAG WHERE TOSSE = "1.0";  
SELECT COUNT(GARGANTA) FROM SRAG WHERE GARGANTA = "1.0";  
SELECT COUNT(DISPNEIA) FROM SRAG WHERE DISPNEIA = "1.0";  
SELECT COUNT(MIALGIA) FROM SRAG WHERE MIALGIA = "1.0";  
SELECT COUNT(SATURACAO) FROM SRAG WHERE SATURACAO = "1.0";  
SELECT COUNT(DESC_RESP) FROM SRAG WHERE DESC_RESP = "1.0";
```

Análise

Tosse é o sintoma mais comum, seguido de febre, dispneia e desconforto respiratório. Baixa saturação também afeta a maioria dos pacientes

Sintoma	Casos	Incidência
Tosse	183.489	0,91
Febre	169.461	0,84
Dispneia	164.511	0,82
Desconforto Respiratório	159.059	0,81
Saturação de O ₂ abaixo de 95%	110.485	0,58
Mialgia (Dor Muscular)	052.384	0,30
Dor de Garganta	038.131	0,22

EXISTE ALGUMA SAZONALIDADE NOS SINTOMAS RELACIONADOS A PROBLEMAS RESPIRATÓRIOS?

Consulta

```
SELECT  
DT_NOTIFIC FROM  
SRAG WHERE  
TOSSE = "1.0"  
OR FEBRE =  
"1.0" OR  
GARGANTA =  
"1.0";
```

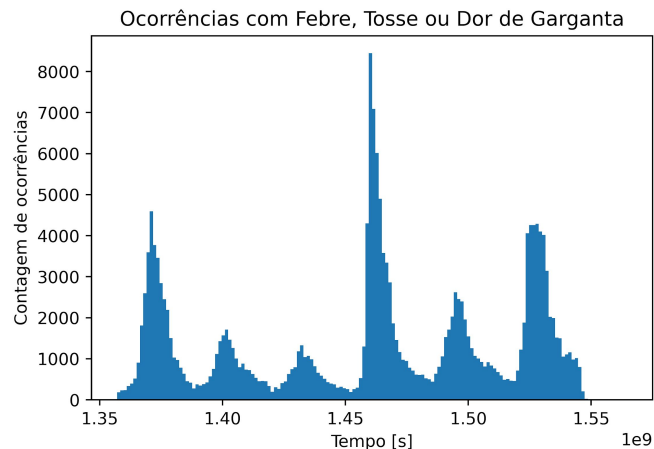
Após:
histograma

Análise

Sim, existe sazonalidade, e os sintomas se concentram nos períodos de outono e inverno



Resultado



OS CASOS SE CONCENTRAM EM ALGUMA REGIÃO GEOGRÁFICA?

Procedimento

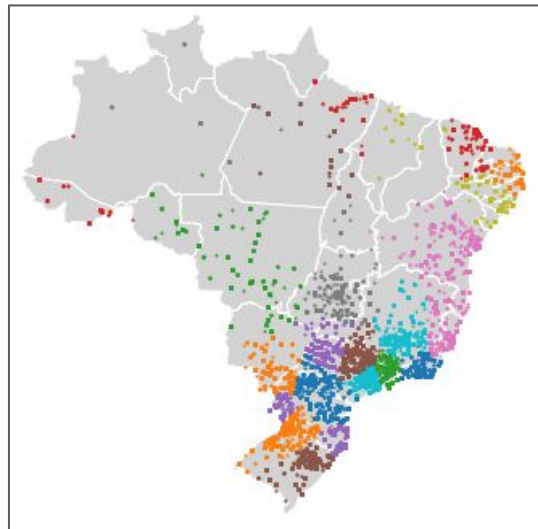
Clusterizamos por latitude e longitude todos os dados na tabela SRAG, utilizando Modelos de ML

Análise

Infelizmente, o resultado não parece ser significativo, já que parece apenas refletir a distribuição demográfica do Brasil



Resultado



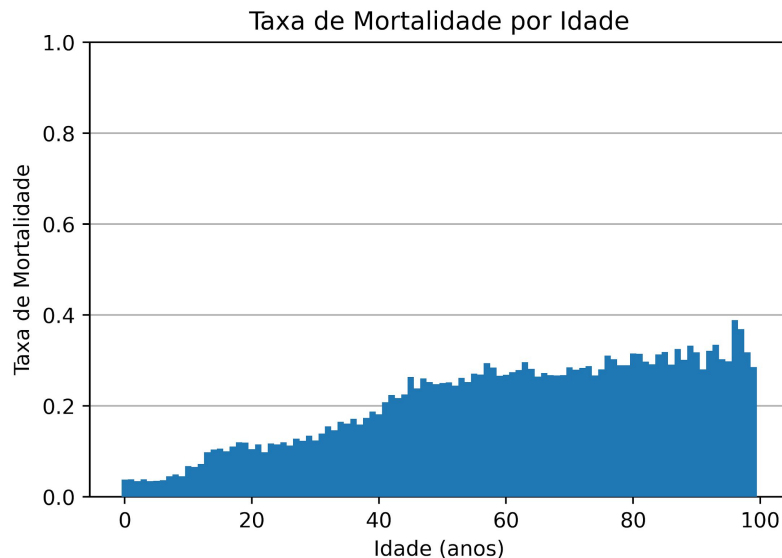
A IDADE INFLUENCIA NA TAXA DE MORTALIDADE?

Procedimento

Calculamos a porcentagem de casos em que a evolução foi registrada como óbito em relação ao total de casos em que a evolução foi registrada.

Análise

Sim, faixas etárias mais idosas apresentam maior taxa de mortalidade.



QUAIS SÃO AS PRINCIPAIS CAUSAS DE SRAG?

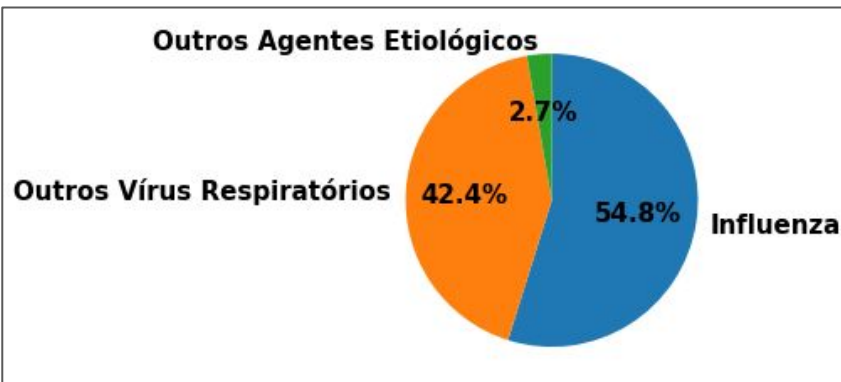
Procedimento

Inspecionar a coluna de classificação do BD, relevando os casos não especificados.

Análise

Influenza corresponde à maioria, seguida de outros vírus respiratórios.

Resultado



A VACINAÇÃO CONTRA INFLUENZA INFLUENCIA NA MORTALIDADE?

Procedimento

Dentre os casos causados por influenza, calculamos a taxa de mortalidade, separando o grupo vacinado e não vacinado

Análise

Sim, os casos de pessoas não vacinadas — embora sejam menos idosas — representaram uma taxa de mortalidade maior

Resultado

Grupo	Taxa de Mortalidade
Vacinados	0,11
Não vacinados	0,16

