

MC536-A

Poltrona de Dados

Olá! Tudo bem?

Somos o grupo
Poltrona de Dados



Tópicos

1

Relembrando o Tema

Tema abordado e o porquê dessa escolha

2

O que fizemos até aqui

O que já fizemos?

3

Próximos passos

O que vem a seguir?



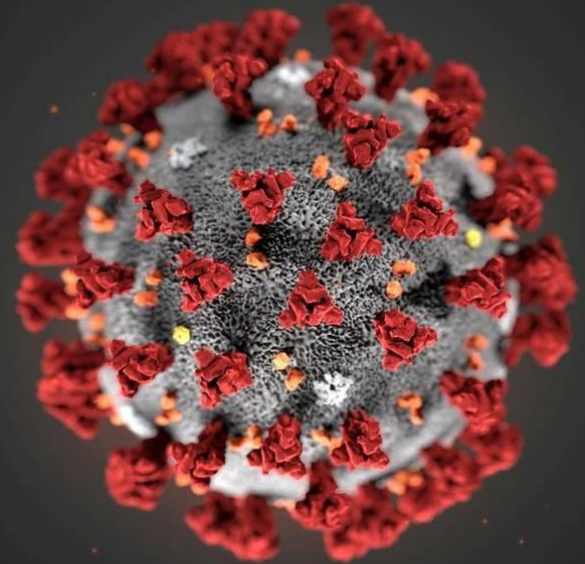
1. Relembrando o Tema

Qual tema iremos abordar?

Pandemia 2020

Dado o contexto atual da pandemia, achamos relevante abordar um tema relacionado ao coronavírus.

Sendo assim nosso objetivo é relacionar os dados sociais, econômicos e geográficos de países com seus respectivos dados da COVID-19.





2. O que fizemos até aqui?

Extração dos dados da COVID-19

Utilizamos a API <https://api.covid19api.com> para fazer a extração de alguns dados para a construção das primeiras tabelas com dados sobre a COVID-19. Segue os dados extraídos nesse processo:

- Novos casos, mortes e recuperados por país;
- Total de casos, mortes e recuperados por país;
- Países e sua sigla em ISO2;
- Data do primeiro caso confirmado em cada país.

Tratamento dos dados

Utilizando a biblioteca Pandas do Python foi possível fazer o tratamento e limpeza dos dados para inserir os mesmos em três tabelas:

Covid First Case

- CountryCode;
- DateOfFirstCase.

Covid Summary

- CountryCode (PK);
- TotalConfirmed;
- TotalDeaths;
- TotalRecovered.
- NewCases;
- NewDeaths;
- NewRecovered.

Countries

- CountryCode (PK);
- CountryName.

Desenvolvimento das primeiras consultas.

Subimos um banco em memória no jupyter para executar as primeiras consultas nas tabelas criadas.

Dado determinado país, qual seu nome e a data do primeiro caso registrado:

```
select c.country, f.first_case  
  from countries as c  
inner join  
  first_case as f on f.country_code = c.country_code
```

Dado determinado país, queremos saber um resumo dos dados de covid para o mesmo :

```
select c.country, f.first_case  
  from countries as c  
inner join  
  first_case as f on f.country_code = c.country_code
```



3. Próximos Passos

O que faremos agora?

Criação de novas tabelas!

A partir da base de dados 'World Development Dataset' montaremos novas tabelas contendo diversas informações para cada país.

Dados socioeconômicos

Tabela que contém diversos indicadores socioeconômicos de cada país. Por exemplo, taxa de desemprego.

Educação

Tabela que contém diversos indicadores do desenvolvimento educacional de cada país. Por exemplo, número de artigos científicos.

Economia

Tabela que contém indicadores econômicos de cada país. Por exemplo, renda per capita, PIB, entre outras.

Mais algumas...

Saúde

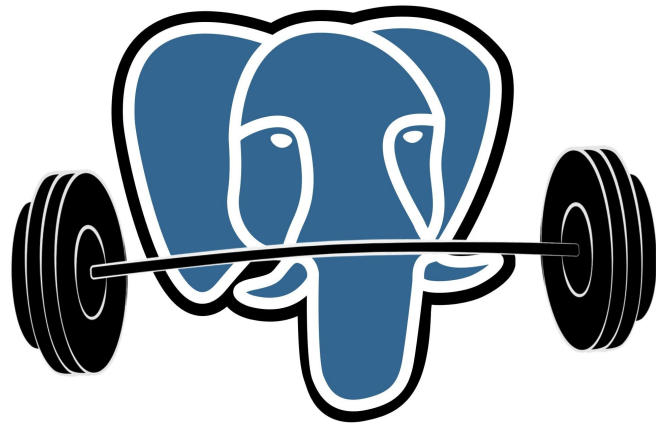
Tabela que contém indicadores do desenvolvimento da saúde em cada país. Por exemplo, número de profissionais da saúde, gastos com saúde, leitos em hospitais, entre outros.

Dados geográficos

Tabela que contém indicadores sobre os dados geográficos de cada país como, por exemplo, população urbana, densidade populacional, aglomeração urbana, entre outros.

Dados, dados e dados

Lidaremos com diversas fontes de dados em diferentes modelos lógicos. Não conseguimos fazer uma query com um join entre uma tabela do Postgres e uma coleção do Mongo, por exemplo.



Solução?

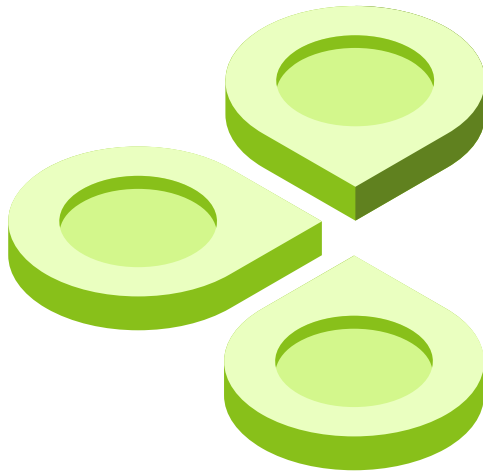
Bancos Online

Visando facilitar o setup do projeto, iremos subir instâncias de bancos online, para que as queries possam ser feitas a partir de um notebook python

Dessa forma, o resultado das queries será do formato de estruturas em Python, poderemos então cruzar informações armazenadas em datasets de modelos lógicos distintos

Conexões

JUPYTER NOTEBOOK
Queries usando bibliotecas em
python



POSTGRES
Instância online do
Postgres, utilizando
[elephantsql](#)

MONGODB
Cluster online utilizando
[Mongo Atlas](#)

Obrigado!

Leonardo Livrare - 220120
Pedro Pupo - 204729
Pedro Strambeck - 204759