

## Atividade Prática – BIG DATA

"Análise Integrada do Impacto da Pandemia COVID-19 nos Indicadores Socioeconômicos Globais"

### Visão Geral

Este projeto visa construir um **pipeline ETL** que integra dados oficiais da pandemia de COVID-19 com indicadores socioeconômicos globais, com o objetivo de **analisar como a pandemia afetou diferentes países e regiões**. A proposta é consolidar esses dados em arquivos prontos para uso em ferramentas como Excel e Power BI, possibilitando **a criação de dashboards interativos para tomada de decisão e estudos aprofundados**.

### APIs Utilizadas

[disease.sh](#) – dados históricos de COVID-19

[World Bank API](#) – dados socioeconômicos por país

### Etapas de Execução

Crie e ative o ambiente virtual:

```
python -m venv venv  
source venv/Scripts/activate  
pip install -r requirements.txt
```

Execute os scripts ETL:

```
python etl_covid.py  
python etl_worldbank.py  
python etl_analises.py  
python etl_tendencias.py  
python main.py
```

Os resultados estarão disponíveis na pasta /data em formato .csv

## ETL's

**etl\_covid.py:** Extrai dados de casos COVID-19 por país

**etl\_worldbank.py:** Extrai indicadores socioeconômicos do World Bank

**etl\_analises.py:** Une dados de COVID e economia para análise consolidada

**etl\_tendencias.py:** Gera séries temporais detalhadas para um país específico

**etl\_menu\_interativo.py:** Versão interativa para exportar dados por país

**mongodb\_connection.py:** Centraliza a conexão segura com o banco MongoDB

**Pasta data:** Armazena todos os CSVs gerados pelas análises

## Decisões Técnicas do Projeto

O projeto adotou decisões técnicas visando organização, escalabilidade e integração com ferramentas analíticas. Utilizou-se o MongoDB Atlas pela flexibilidade e fácil integração com Python, separando dados brutos e transformados. A linguagem escolhida foi o Python, com bibliotecas como pandas, requests e pymongo. O pipeline foi modularizado em scripts para coleta, transformação e geração de arquivos, facilitando manutenção e expansão. Os dados finais são exportados em CSV com separador ";", otimizando uso em Excel e Power BI para criação de dashboards.

## Resultados Esperados

Conjunto de dados cruzando COVID com PIB, desemprego, expectativa de vida etc.

Arquivos .csv que permitem visualização em dashboards

Possibilidade de análises históricas e comparativas por país

## Aprendizados

Criação de pipelines ETL reais

Integração de APIs públicas

Aplicação de transformações estatísticas (taxa de crescimento, média móvel)

## Melhorias Futuras

Visualização de Dashboards pelo PowerBI

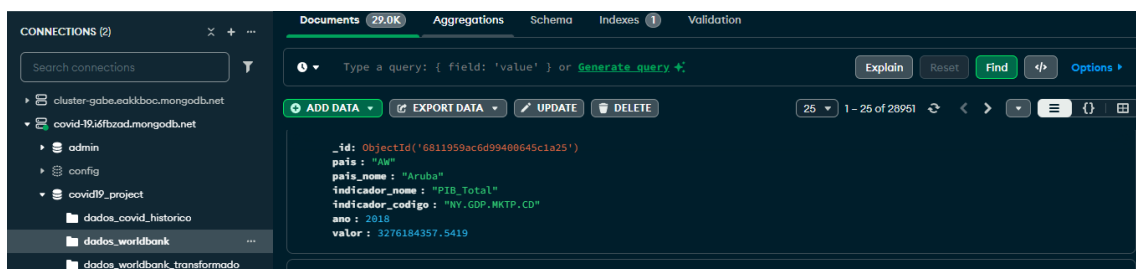
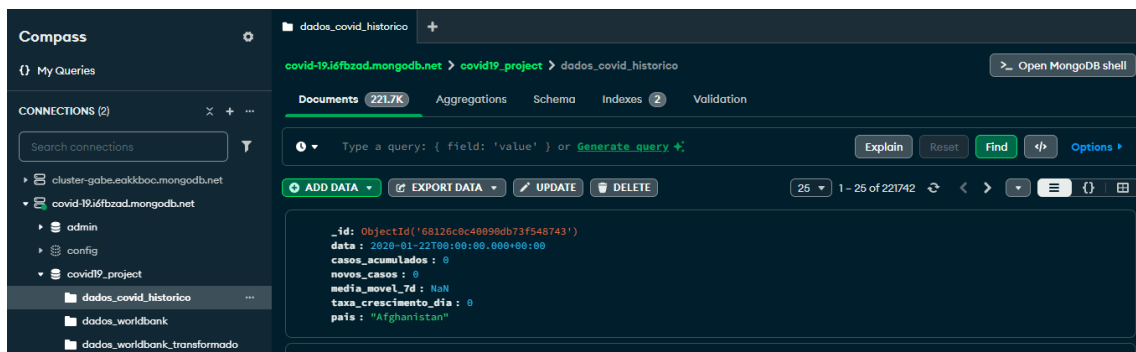
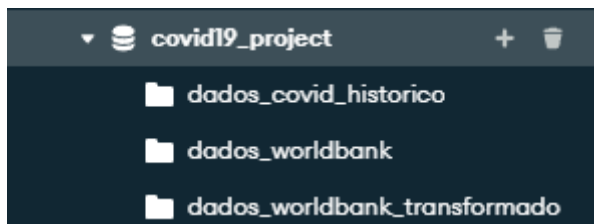
Novos indicadores: vacinação, inflação, mobilidade urbana

Armazenamento em Data Warehouse (BigQuery, PostgreSQL)

Análise de correlação e regressão estatística

Cache e paralelismo para otimizar coleta de dados

## Mongo Compass



CONNECTIONS (2)

Search connections

cluster-gabe.eakkboc.mongodb.net

covid-19.6fbzad.mongodb.net

admin

config

covid19\_project

dados\_covid\_historico

dados\_worldbank

dados\_worldbank\_transformado ...

Documents (29.0K)

Aggregations

Schema

Indexes (1)

Validation

Type a query: { field: 'value' } or [Generate query](#)

ADD DATA

EXPORT DATA

UPDATE

DELETE

251 - 25 of 28951

↺

↻

↷

⋮

⌕

🔗

📄

```
_id: ObjectId('681195acc6d99488645c8b43')
pais: "AD"
pais_nome: "Andorra"
indicador_nome: "Expectativa_vida"
indicador_codigo: "SP.DYN.LE00.IN"
ano: 2012
valor: 84.334
taxa_crescimento_pct: 0.85
media_movel_3anos: 84.2756666666667
```