

Poltrona de Dados

MC536

Olá, Tudo certo?

Somos o grupo
Poltrona de Dados



Tópicos

1. Introdução ao Tema
Qual Tema será abordado?

2. Datasets
Datasets utilizados

3. Modelos Relacionais e Lógicos
Modelos criados

4. Tratamento dos Dados
Organização dos dados colhidos

5. Bancos online
Utilização de bancos em cloud

6. Queries e Resultados
Consultas feitas e seus vereditos

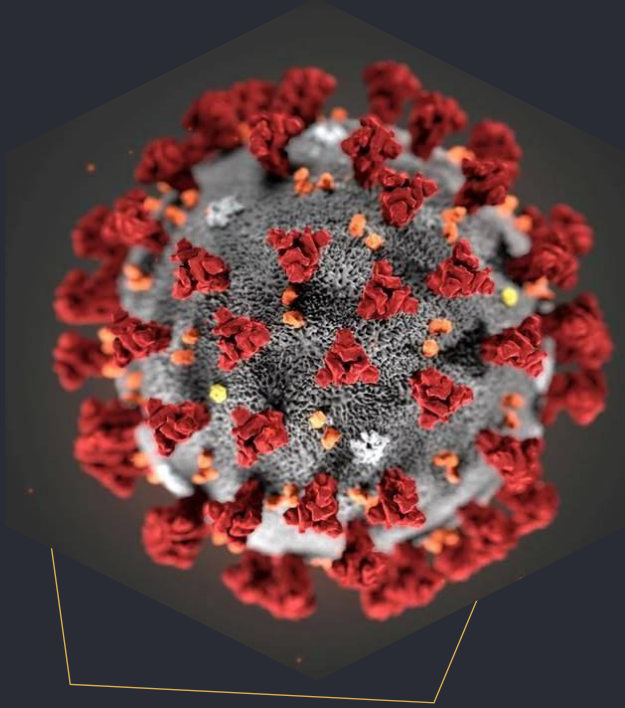


1

Introdução ao Tema

Qual Tema será abordado?

Pandemia COVID-19



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

Sendo assim nosso objetivo é relacionar os dados sobre a covid-19 a dados socioeconômicos, educacionais, geográficos, entre outros.



2

Datasets

Quais datasets serão utilizados?

Datasets

World Bank
Socio Economics

World Development

COVID19

COVID19 API

World Bank Education
Education

COVID-19 Dataset

Nesse dataset, com estrutura hierarquica, temos as seguintes informações agrupadas por país:

- Total de casos;
- Total de mortes;
- Total de recuperados;
- Novos casos;
- Novas mortes;
- Novos Recuperados;
- Data;
- Posição geográfica do país.

```
{
  "Message": "",
  "Global": {
    "NewConfirmed": 443698,
    "TotalConfirmed": 41220369,
    "NewDeaths": 6668,
    "TotalDeaths": 1131337,
    "NewRecovered": 214218,
    "TotalRecovered": 28116565
  },
  "Countries": [
    {
      "Country": "Afghanistan",
      "CountryCode": "AF",
      "Slug": "afghanistan",
      "NewConfirmed": 153,
      "TotalConfirmed": 40510,
      "NewDeaths": 2,
      "TotalDeaths": 1501,
      "NewRecovered": 34,
      "TotalRecovered": 33824,
      "Date": "2020-10-22T22:47:30Z",
      "Premium": {}
    },
    {
      "Country": "Albania",
      "CountryCode": "AL",
      "Slug": "albania",
      "NewConfirmed": 297,
      "TotalConfirmed": 17948,
      "NewDeaths": 4,
      "TotalDeaths": 462,
      "NewRecovered": 116,
      "TotalRecovered": 10341,
      "Date": "2020-10-22T22:47:30Z",
      "Premium": {}
    }
  ],
}
```


Brazil	2011	2012
Adolescent fertility rate (births per 1,000 women ages 15-19)	65.5	64.1
Agriculture, forestry, and fishing, value added (% of GDP)	4.3	4.2
Annual freshwater withdrawals, total (% of internal resources)	..	1.3
Births attended by skilled health staff (% of total)	99.0	99.0
CO2 emissions (metric tons per capita)	2.2	2.4
Contraceptive prevalence, any methods (% of women ages 15-49)
Domestic credit provided by financial sector (% of GDP)
Electric power consumption (kWh per capita)	2,430.8	2,501.5
Energy use (kg of oil equivalent per capita)	1,367.2	1,413.7
Exports of goods and services (% of GDP)	11.6	11.9
External debt stocks, total (DOD, current US\$)	404,046,105,34	440,515,187,63
Fertility rate, total (births per woman)	1.8	1.8
Foreign direct investment, net inflows (BoP, current US\$)	102,427,229,78	92,568,379,494
Forest area (sq. km)	4,974,740.0	4,964,900.0
GDP (current US\$)	2,616,200,980,3	2,465,188,674,4

World Bank Socio Economics

Dataset tabular contendo diversas informações e indicadores sobre a economia, saúde, aspectos da população em geral, entre outras áreas. É possível traçar uma relação entre os dados contidos neste banco e dados provenientes do dataset referentes ao coronavírus. O banco de dados possui ainda informações de anos diferentes, possibilitando ainda que análises distintas sejam feitas.

World Bank Education

Nesse dataset, com estrutura tabular, temos diversas informações sobre dados relacionados a educação. A ideia é relacionar como o acesso a educação teve efeito nas medidas tomadas pelos governos e nas consequências da pandemia

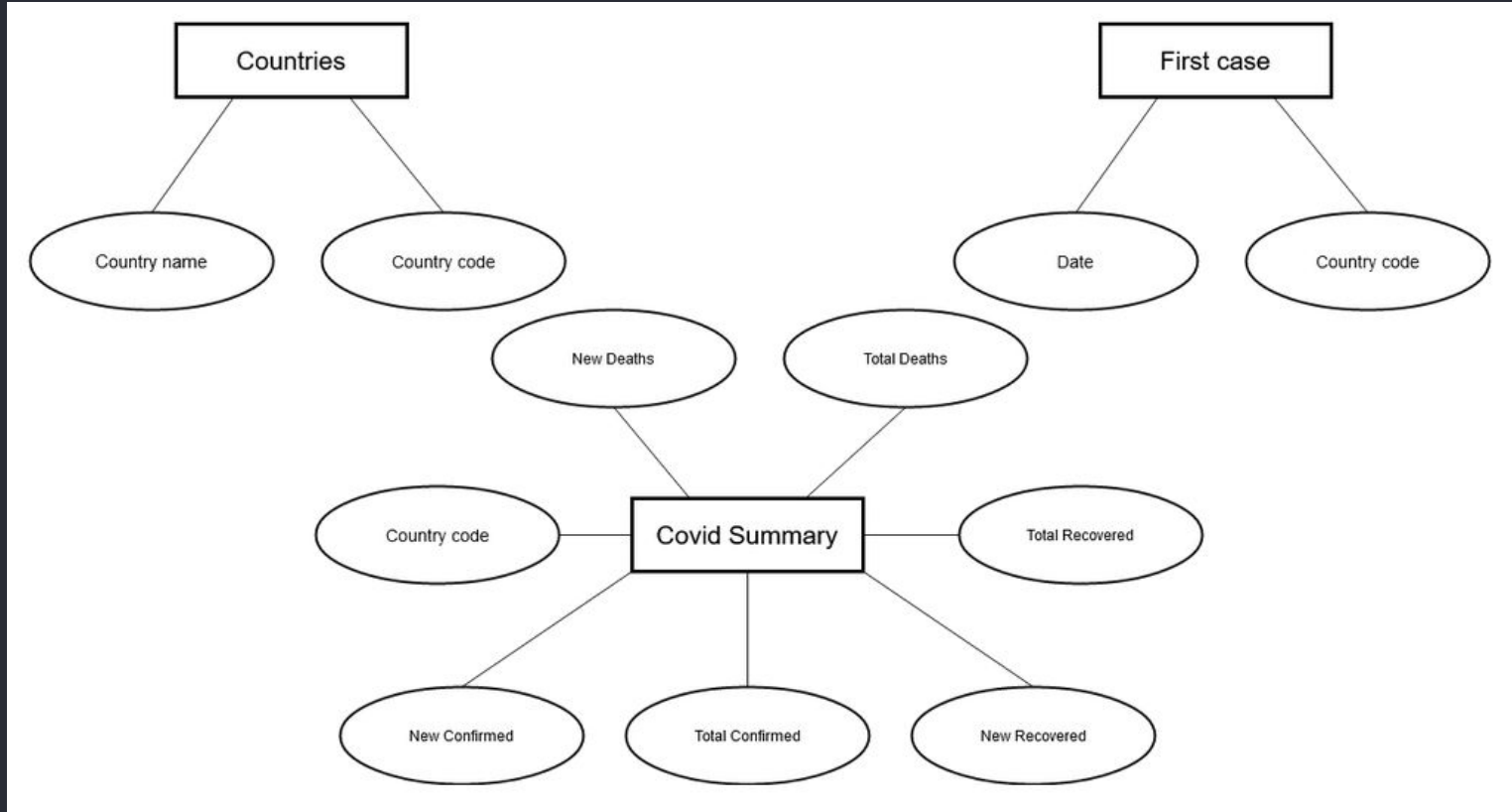
	2017	2018	2019
Adjusted net intake rate to Grade 1 of primary education, gender parity index (GPI)	1.0	..	
Adjusted net intake rate to Grade 1 of primary education, male (%)	87.6	..	
Adult illiterate population, 15+ years, % female	50.1	49.8	
Adult illiterate population, 15+ years, both sexes (number)	11,276,037.0	11,167,773.0	
Adult illiterate population, 15+ years, female (number)	5,644,716.0	5,562,972.0	
Adult illiterate population, 15+ years, male (number)	5,631,321.0	5,604,800.0	
Adult literacy rate, population 15+ years, both sexes (%)	93.1	93.2	
Adult literacy rate, population 15+ years, female (%)	93.2	93.4	
Adult literacy rate, population 15+ years, gender parity index (GPI)	1.0	1.0	
Adult literacy rate, population 15+ years, male (%)	92.9	93.0	
Africa Dataset: Average number of grades per multigrade class in primary schools (number of grades)	

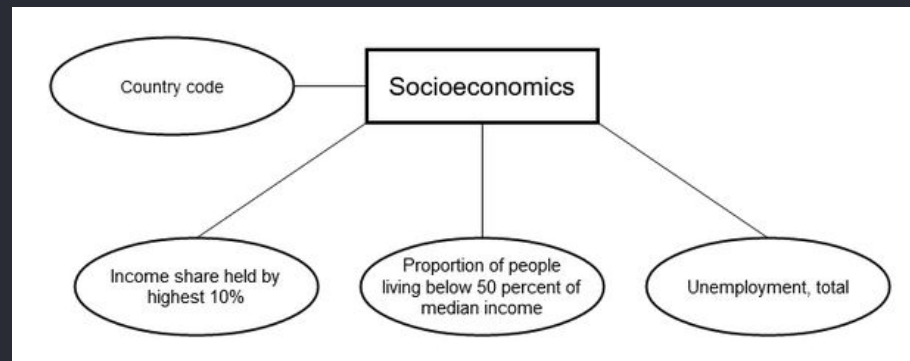
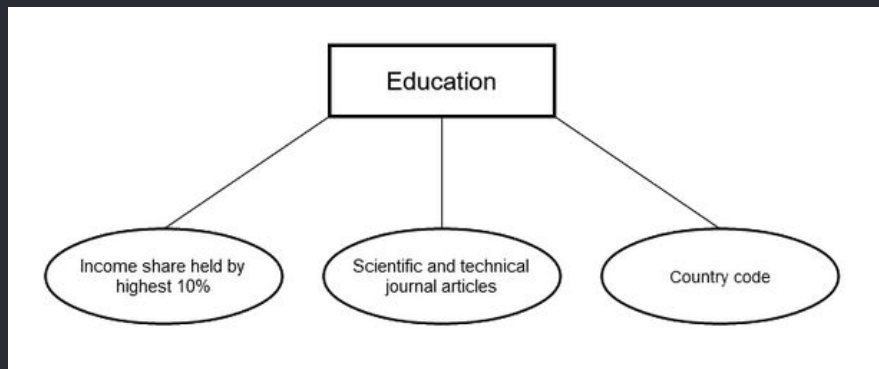
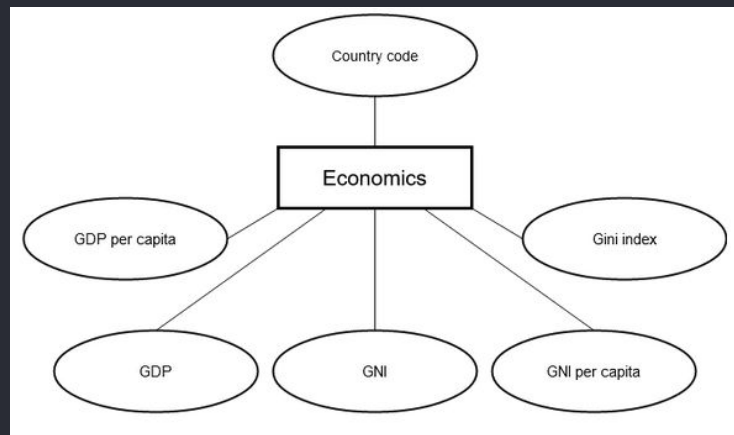
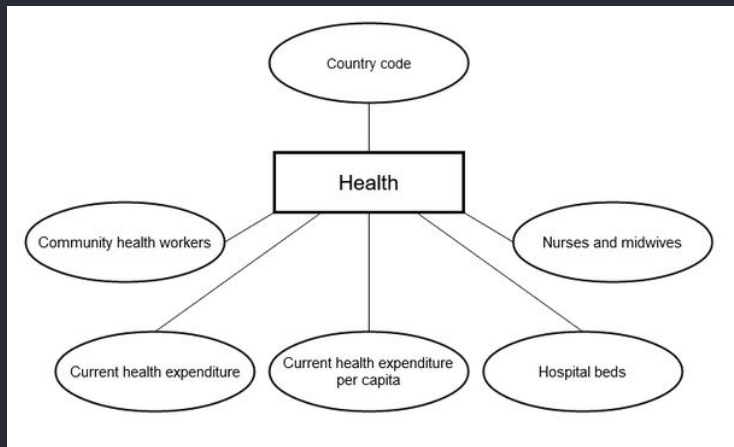
3

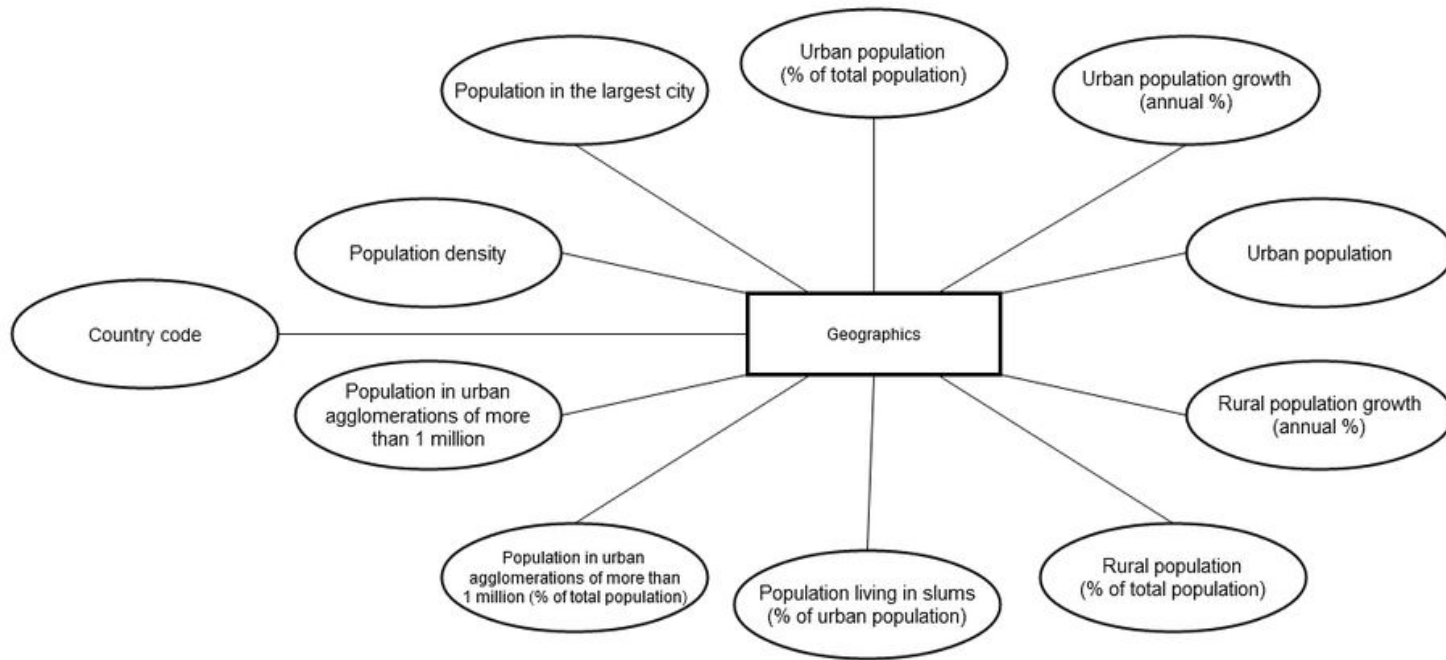
Modelos Relacionais e Lógicos

Quais foram os modelos criados?

Modelos Conceituais







Modelos Lógicos

Covid_Summary

<u>Country_Code</u>	New_confirmed	Total_confirmed	New_deaths	Total_deaths	New_recovered	Total_recovered
---------------------	---------------	-----------------	------------	--------------	---------------	-----------------

Countries

<u>Country_Code</u>	Country_name
---------------------	--------------

First_Case

<u>Country_Code</u>	Country_name
---------------------	--------------

Health

<u>Country_Code</u>	Com_H_Workers	Current_H_exp	Current_H_exp_pc	Hospital_beds	Nurses_midwives
---------------------	---------------	---------------	------------------	---------------	-----------------

Economics

<u>Country_Code</u>	GDP	GDP_pc	Gini_index	GNI	GNI_pc
---------------------	-----	--------	------------	-----	--------

Socioeconomics			
<u>Country_Code</u>	Income_share_10p	living_below_50p	Unemployment

Education		
<u>Country_Code</u>	Scientific_journals	Research_exp

Geographics				
<u>Country_Code</u>	Urban_Pop	Pop_density	Urban_Pop_p	Urban_Pop_growth_p

Geographics					
Pop_urban_ag_1mi	Pop_urban_ag_1mi_p	Pop_slums	Rural_Pop_p	Rural_Pop_growth	Pop_largest_city

4

Tratamento dos Dados

Como se deu o tratamento dos dados?

Dados COVID-19

Utilizando a biblioteca Pandas do Python foi possível fazer o tratamento e limpeza dos dados referentes ao COVID-19 para inseri-los em três tabelas:

First_case

- countries
- first_case

Summary

- CountryCode (PK)
- NewConfirmed
- NewDeaths
- NewRecovered
- TotalDeaths
- TotalConfirmed
- TotalRecovered

Countries

- CountryCode (PK)
- CountryName

Novas Tabelas

Criadas utilizando dados referentes aos anos de 2017, 2018 e 2019 presentes nos datasets Socio Economics e Education.

socioEconomics_ANO

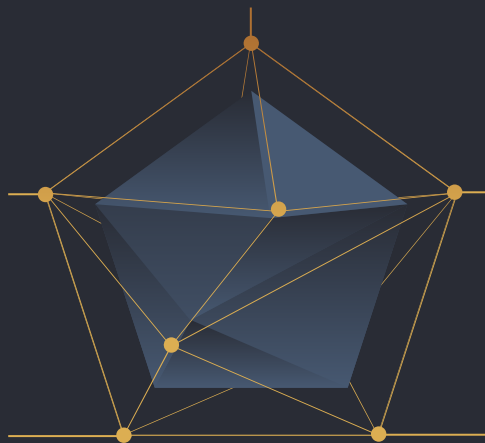
Dados socioeconômicos dos países.

economics_ANO

Dados econômicos dos países

education_ANO

Indicadores e dados educacionais



health_ANO

Indicadores e dados gerais sobre a saúde em cada país.

geographics_ANO

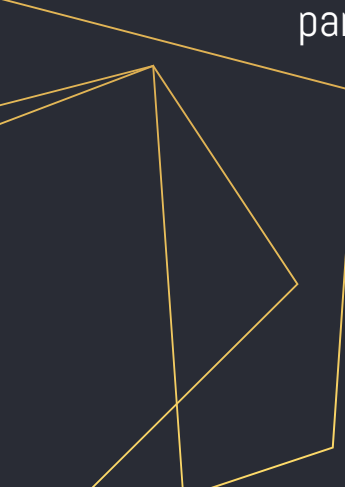
Dados geográficos dos países



5

Bancos Online

Cloud



Visando facilitar o setup do projeto, subimos instâncias de bancos online, para que as queries pudessem ser feitas a partir de um notebook python.

Dessa forma, o resultado das queries se deu no formato de estruturas em Python, podemos então cruzar informações armazenadas em datasets de modelos lógicos distintos

Conexões

Jupyter Notebook

Queries usando bibliotecas
em python

MongoDB

Cluster online utilizando
[Mongo Atlas](#)

Postgres

Instância online do
Postgres, utilizando
[Elephantsql](#)

6

Queries e Resultados

Quais queries foram realizadas e quais
foram seus resultados?

Países x Taxa de mortalidade

	Country	alpha_2	Slug	NewConfirmed	TotalConfirmed	NewDeaths	TotalDeaths	NewRecovered	TotalRecovered	first_case	DeathRatio
0	France	FR	france	248	2275677	1	52822	0	167915	2020-01-24T00:00:00Z	0.023212
1	Morocco	MA	morocco	4346	364190	70	5985	4044	314237	2020-03-02T00:00:00Z	0.016434
2	Papua New Guinea	PG	papua-new-guinea	0	669	0	7	0	597	2020-03-20T00:00:00Z	0.010463
3	Portugal	PT	portugal	3384	303846	68	4645	2569	223446	2020-03-02T00:00:00Z	0.015287
4	Andorra	AD	andorra	52	6842	0	76	48	5988	2020-03-02T00:00:00Z	0.011108
...
186	Bahrain	BH	bahrain	133	87270	0	341	106	85463	2020-02-24T00:00:00Z	0.003907
187	Saudi Arabia	SA	saudi-arabia	249	357872	12	5919	337	347513	2020-03-02T00:00:00Z	0.016539
188	Republic of Kosovo	XK	kosovo	0	40117	0	1026	0	25586	2020-03-14T00:00:00Z	0.025575
189	Ghana	GH	ghana	0	51667	0	323	0	50547	2020-03-14T00:00:00Z	0.006252
190	Japan	JP	japan	2456	153432	32	2141	1844	127148	2020-01-22T00:00:00Z	0.013954

Taxa de mortalidade média dos 20 países com maior PIB, 20 países com menor PIB e Geral

```
: df_covid_ratio_pib = df_pib_per_capita.merge(df_covid_ratio, how="inner", on="alpha_2")
```

```
: top20_pib_death_ratio = df_covid_ratio_pib.head(20)["DeathRatio"].mean()  
last20_pib_death_ratio = df_covid_ratio_pib.tail(20)["DeathRatio"].mean()  
death_ratio = df_covid_ratio_pib["DeathRatio"].mean()
```

```
: print("top20: ", top20_pib_death_ratio)  
print("last20: ", last20_pib_death_ratio)  
print("All: ", death_ratio)
```

```
top20: 0.015978299079689017  
last20: 0.02473783886622265  
All: 0.02048843412682874
```

Investimento em saúde x Taxa de mortalidade

```
df_covid_ratio_health = df_health_expenditure_per_capita.merge(df_covid_ratio, how="inner", on="alpha_2")
```

```
df_covid_ratio_health
```

```
top10_health_death_ratio = df_covid_ratio_health.head(10)["DeathRatio"].mean()  
last10_health_death_ratio = df_covid_ratio_health.tail(10)["DeathRatio"].mean()  
health_death_ratio = df_covid_ratio_health["DeathRatio"].mean()
```

```
print("Top 10: ", top10_health_death_ratio)  
print("Last 10: ", last10_health_death_ratio)  
print("All: ", health_death_ratio)
```

```
Top 10:  0.017152070771987122  
Last 10:  0.023677886835821914  
All:  0.01907830705728344
```

Investimento em educação x Taxa de mortalidade

```
df_reasearch_death = df_research.merge(df_covid_ratio, how="inner", on="alpha_2")  
df_reasearch_death
```

```
top_10_research = df_reasearch_death.head(10) ["DeathRatio"].mean()  
last_10_research = df_reasearch_death.tail(10) ["DeathRatio"].mean()  
total_research = df_reasearch_death["DeathRatio"].mean()
```

```
print("Top 10: ", top_10_research)  
print("Last 10: ", last_10_research)  
print("All: ", total_research)
```

```
Top 10:  0.015175292505589511  
Last 10:  0.020476933583673147  
All:  0.02030516412427076
```

Outras

Foram feitas várias outras queries relacionando dados internos de cada tabela, como, por exemplo:

Queries relacionadas a Socio Economics

```
unemployment_query = """
select
cc.country_name,
se.country_code,
cc.alpha_2,
unemployment
from "socioEconomics_{" se
inner join
country_codes as cc on cc.alpha_3 = se.country_code
where unemployment is not null
order by unemployment desc
"""
```

Queries relacionadas a Geografia

```
pop_density_query = """
select
cc.country_name,
e.country_code,
cc.alpha_2,
pop_density
from geographics_{" e
inner join
country_codes as cc on cc.alpha_3 = e.country_code
where pop_density is not null
order by pop_density desc
"""
```

Obrigado!

Leonardo Livrare - 220120

Pedro Pupo - 204729

Pedro Strambeck - 204759

CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), infographics & images by [Freepik](#).