

Apresentação do Case

Este case foi desenvolvido com o intuito de utilizar as ferramentas do Google Cloud Platform (GCP) para a sua solução, dessa forma, recomendamos a utilização dessas ferramentas nessa etapa.

A utilização de ferramentas de outras plataformas Cloud não é de caráter eliminatório.

Importante:

Para Engenheiro de Dados o case deve contemplar a Etapa 1 e Etapa 2 obrigatoriamente, sendo a Etapa 3 facultativa.

Para Analista de BI o case deve contemplar a Etapa 2 (os dados podem ser carregados de forma manual no Bigquery) e Etapa 3 obrigatoriamente , sendo a Etapa 1 facultativa.

Estudo de Caso COVID-19

Boletins informativos e casos do coronavírus por município por dia

Fonte original: [Secretarias de Saúde estaduais](#)

Libertado por: [Álvaro Justen e dezenas de colaboradores](#)

Código-fonte: <https://github.com/turicas/covid19-br>

Licença: [Creative Commons Attribution-ShareAlike 4.0 International \(CC BY-SA 4.0\)](#)

fonte: <https://brasil.io/dataset/covid19>

Conjunto de dados

Tabelas: caso, obito_cartorio.

Informações úteis

Essa tabela possui os casos confirmados e óbitos obtidos dos boletins das Secretarias Estaduais de Saúde (SES). Os dados foram enriquecidos, de forma que a

partir do momento em que um município confirma um caso, ele sempre aparecerá nessa tabela (mesmo que para uma determinada data a SES não tenha liberado o boletim - nesse caso é repetido o dado do dia anterior).

Base de dados:

Arquivo	Tamanho	SHA512SUM
caso.csv.gz	30.25MB	c6f6d8659f7e8aa177e33620c4ead67da09f5df089c7c31b52acee54b426f165cbfeacca2b51874b84c25247a63cda5dbf27c8824254aca500e50faa87f18650
obito_cartorio.csv.gz	434.18kB	94f91c9e05516b8f72419a4937d3d4aa27989d64e76b9badabec0b496e6ac8b96c6dccf4ab116f79d175eae93ac93fb1436006013f515034feb8600dd8c940e0

Baixe os arquivos acima citados em: <https://brasil.io/dataset/covid19/files/>

Dicionário de Dados

Base de Dados : caso

Colunas:

- **date:** data de coleta dos dados no formato YYYY-MM-DD.
- **state:** sigla da unidade federativa, exemplo: SP.
- **city:** nome do município (pode estar em branco quando o registro é referente ao estado, pode ser preenchido com Importados/Indefinidos também).
- **place_type:** tipo de local que esse registro descreve, pode ser city ou state.
- **order_for_place:** número que identifica a ordem do registro para este local. O registro referente ao primeiro boletim em que esse local aparecer será contabilizado como 1 e os demais boletins incrementarão esse valor.
- **is_last:** campo pré-computado que diz se esse registro é o mais novo para esse local, pode ser True ou False (caso filtre por esse campo, use is_last=True ou is_last=False, não use o valor em minúsculas).
- **city_ibge_code:** código IBGE do local.
- **confirmed:** número de casos confirmados.
- **Deaths:** número de mortes.
- **estimated_population:** população estimada para esse município/estado em 2020, segundo o IBGE. (acesse o script que faz o download e conversão dos dados de população).
- **estimated_population_2019:** população estimada para esse município/estado em 2019, segundo o IBGE. ATENÇÃO: essa coluna possui valores desatualizados, prefira usar a coluna estimated_population.
- **confirmed_per_100k_inhabitants:** número de casos confirmados por 100.000 habitantes (baseado em estimated_population).
- **death_rate:** taxa de mortalidade (mortes / confirmados).

Base de Dados: *obito_cartorio*

Colunas:

- **date**: Data no formato YYYY-MM-DD
- **state**: Sigla da unidade federativa com dois dígitos, exemplo: SP
- **epidemiological_week_2019**: Número da semana epidemiológica para essa data em 2019.
- **epidemiological_week_2020**: Número da semana epidemiológica para essa data em 2020.
- **deaths_total_2019**: Quantidade de óbitos totais para o estado state acumulados no ano de 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).
- **deaths_total_2020**: Quantidade de óbitos totais para o estado state acumulados no ano de 2020 (de 1 de janeiro de 2020 a date).
- **new_deaths_total_2019**: Quantidade de óbitos totais para o estado state ocorridos no dia/mês de date, porém em 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).
- **deaths_covid19**: Quantidade de óbitos em decorrência de suspeita ou confirmação de covid19 para o estado state acumulados no ano de 2020 (de 1 de janeiro de 2020 a date).
- **new_deaths_total_2020**: Quantidade de óbitos totais para o estado state ocorridos na data date (em 2020).
- **deaths_indeterminate_2019**: Quantidade de óbitos de motivo indeterminado para o estado state acumulados no ano de 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).
- **deaths_indeterminate_2020**: Quantidade de óbitos de motivo indeterminado para o estado state acumulados no ano de 2020 (de 1 de janeiro de 2020 a date).
- **deaths_others_2019**: Quantidade de óbitos em decorrência de outros motivos para o estado state acumulados no ano de 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).
- **deaths_others_2020**: Quantidade de óbitos em decorrência de outros motivos para o estado state acumulados no ano de 2020 (de 1 de janeiro de 2020 a date).
- **deaths_pneumonia_2019**: Quantidade de óbitos em decorrência de pneumonia para o estado state acumulados no ano de 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).
- **deaths_pneumonia_2020**: Quantidade de óbitos em decorrência de pneumonia para o estado state acumulados no ano de 2020 (de 1 de janeiro de 2020 a date).
- **deaths_respiratory_failure_2019**: Quantidade de óbitos em decorrência de insuficiência respiratória para o estado state acumulados no ano de 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).
- **deaths_respiratory_failure_2020**: Quantidade de óbitos em decorrência de insuficiência respiratória para o estado state acumulados no ano de 2020 (de 1 de janeiro de 2020 a date).
- **deaths_sars_2019**: Quantidade de óbitos em decorrência de SRAG para o estado state acumulados no ano de 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).
- **deaths_sars_2020**: Quantidade de óbitos em decorrência de SRAG para o estado state acumulados no ano de 2020 (de 1 de janeiro de 2020 a date).
- **deaths_septicemia_2019**: Quantidade de óbitos em decorrência de septicemia para o estado state acumulados no ano de 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).

- **deaths_septicemia_2020**: Quantidade de óbitos em decorrência de septicemia para o estado state acumulados no ano de 2020 (de 1 de janeiro de 2020 a date).
- **new_deaths_covid19**: Quantidade de óbitos em decorrência de suspeita ou confirmação de covid19 para o estado state ocorridos na data date (em 2020).
- **new_deaths_indeterminate_2019**: Quantidade de óbitos de motivo indeterminado para o estado state ocorridos no dia/mês de date, porém em 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).
- **new_deaths_indeterminate_2020**: Quantidade de óbitos de motivo indeterminado para o estado state ocorridos na data date (em 2020).
- **new_deaths_others_2019**: Quantidade de óbitos em decorrência de outros motivos para o estado state ocorridos no dia/mês de date, porém em 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).
- **new_deaths_others_2020**: Quantidade de óbitos em decorrência de outros motivos para o estado state ocorridos na data date (em 2020).
- **new_deaths_pneumonia_2019**: Quantidade de óbitos em decorrência de pneumonia para o estado state ocorridos no dia/mês de date, porém em 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).
- **new_deaths_pneumonia_2020**: Quantidade de óbitos em decorrência de pneumonia para o estado state ocorridos na data date (em 2020).
- **new_deaths_respiratory_failure_2019**: Quantidade de óbitos em decorrência de insuficiência respiratória para o estado state ocorridos no dia/mês de date, porém em 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).
- **new_deaths_respiratory_failure_2020**: Quantidade de óbitos em decorrência de insuficiência respiratória para o estado state ocorridos na data date (em 2020).
- **new_deaths_sars_2019**: Quantidade de óbitos em decorrência de SRAG para o estado state ocorridos no dia/mês de date, porém em 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).
- **new_deaths_sars_2020**: Quantidade de óbitos em decorrência de SRAG para o estado state ocorridos na data date (em 2020).
- **new_deaths_septicemia_2019**: Quantidade de óbitos em decorrência de septicemia para o estado state ocorridos no dia/mês de date, porém em 2019 (de 1 de janeiro de 2019 a dia/mês de date em 2019).
- **new_deaths_septicemia_2020**: Quantidade de óbitos em decorrência de septicemia para o estado state ocorridos na data date (em 2020).

fonte: https://github.com/turicas/covid19-br/blob/master/api.md#caso_full

ETAPA 1 - Cloud Function e Cloud Storage

Fazer a ingestão dos dados no Storage e depois via Cloud Function adicionar os dados a uma tabela no Bigquery.

1 - Crie um Bucket no Cloud Storage chamado upload_covid_data, este bucket será utilizado para receber os arquivos baixados anteriormente.

2 - Crie uma Cloud Function que carregue o arquivo adicionado ao bucket, criado no passo anterior, ao término do upload do arquivo (gatilho de atualização/adicion de um determinado arquivo no bucket), faça a carga dos dados para uma tabela (conjunto de dados) com os respectivos nomes dos arquivos (*caso* e *obito_cartorio*) dentro de um dataset chamado covid no Bigquery.

NOTA: Os dados devem ser decodificados corretamente e a tabela deve ser sobre-escrita (não pode haver duplicidade de dados) em caso de um novo arquivo para: *caso.csv* ou *obito_cartorio.csv* ser adicionado ao bucket.

Disponibilizar o código da Cloud function e as saídas dos logs da função.

ETAPA 2 - Bigquery

Utilizando o ambiente do Bigquery e a linguagem SQL, responda os seguintes questionamentos do cliente sobre os dados:

NOTA: Para cada uma das perguntas descritas abaixo apresentar o código SQL e o resultado (dados) gerado pela consulta.

1 - Qual o total de Casos Confirmados?

Query:

```
SELECT sum(confirmed) as total_confirmados
FROM `kabum-case-project.covid.caso`
WHERE is_last=True AND place_type = 'state'
```

Resultado:

Row	total_confirmados
1	29849740

2 - Qual a quantidade de Casos confirmados por Estado, classificando os 5 primeiros estados com mais casos?

Query:

```
WITH
  casos_confirmados_cte AS (
    SELECT
      SUM(confirmed) AS total_confirmados,
      state,
      RANK() OVER(ORDER BY SUM(confirmed) DESC) AS ranking
    FROM
      `covid.caso`
    WHERE
      is_last=TRUE
      AND place_type = 'state'
    GROUP BY
      state)
SELECT
  state,
  casos_confirmados_cte.total_confirmados,
  ranking
FROM
  casos_confirmados_cte
WHERE
  ranking <= 5
ORDER BY
  casos_confirmados_cte.total_confirmados DESC
```

Resultado:

Row	ranking	state	total_confirmados
1	1	SP	5232374
2	2	MG	3317401
3	3	PR	2407960
4	4	RS	2263880
5	5	RJ	2078817

3 - Qual a Letalidade em % (mortes/casos confirmados) por Estado, classificando os 5 primeiros estados com maior letalidade ?

Query:

```
WITH
  letalidade_cte AS (
    SELECT
      SUM(deaths)/SUM(confirmed) AS letalidade,
      state,
      RANK() OVER(ORDER BY SUM(deaths)/SUM(confirmed) DESC) AS ranking
    FROM
      `covid.caso`
    WHERE
      is_last=TRUE AND place_type = 'state'
    GROUP BY
      state)
SELECT
  state,
  ROUND(letalidade_cte.letalidade * 100, 5) AS letalidade_em_percent,
  ranking
FROM
  letalidade_cte
WHERE
  ranking <= 5
ORDER BY
  letalidade_cte.letalidade DESC
```

Resultado:

Row	ranking	state	letalidade_em_percent
1	1	RJ	3.49694
2	2	SP	3.19377
3	3	MA	2.56224
4	4	AM	2.43533
5	5	PA	2.40638

4 - Qual a Taxa de Óbitos por cada mil habitantes, por estado , listar os 5 primeiros estados com maior concentração de óbitos por cada mil habitantes (população) ?

Query:

```
WITH
  covid_caso_cte AS (
    SELECT
      SUM(estimated_population)/1000 AS population,
      state
    FROM
      `covid.caso`
    WHERE
      is_last = TRUE
      AND place_type = 'state'
    GROUP BY
      state )
SELECT
  *
FROM (
  SELECT
    DENSE_RANK() OVER(ORDER BY `covid.obito_cartorio`.deaths_total_2020/
population DESC) AS ranking,
    `covid.obito_cartorio`.state as Estado,
    ROUND(`covid.obito_cartorio`.deaths_total_2020/(population), 5) AS
taxa_de_obitos_1k,
    `covid.obito_cartorio`.deaths_total_2020 AS obitos_2020,
    population AS population_1k
  FROM
    `covid.obito_cartorio`
  JOIN
    covid_caso_cte
  ON
    `covid.obito_cartorio`.state = covid_caso_cte.state
  WHERE
    `covid.obito_cartorio`.date = '2020-12-31'
  ORDER BY
    ranking )
WHERE
  ranking <= 5
```

Resultado:

Row	ranking	Estado	taxa_de_obitos_1k	obitos_2020	population_1k
1	1	RJ	9.41519	163506	17366.189
2	2	RS	7.54217	86154	11422.973
3	3	SP	7.31706	338702	46289.333
4	4	PE	6.99435	67262	9616.621
5	5	ES	6.66232	27076	4064.052

5 - Qual a porcentagem de municípios que registraram óbito em relação ao total de municípios da amostra?

(pode-se obter os municípios brasileiros através do link:
<https://www.ibge.gov.br/explica/codigos-dos-municipios.php> /
<https://www.ibge.gov.br/cidades-e-estados>)

Query:

```
WITH
municipios_sem_obito AS (
SELECT
CAST(COUNT(city) AS int64) AS count_municipios_sem_obito
FROM
`covid.caso`
WHERE
deaths = 0
AND is_last = TRUE
AND city_ibge_code IS NOT NULL
AND place_type = 'city' ),
total_municipios AS (
SELECT
CAST(COUNT(city) AS int64) AS count_municipios
FROM
`covid.caso`
WHERE
is_last = TRUE
AND city_ibge_code IS NOT NULL
AND place_type = 'city' )
SELECT
total_municipios.count_municipios,
municipios_sem_obito.count_municipios_sem_obito,
ROUND((total_municipios.count_municipios -
municipios_sem_obito.count_municipios_sem_obito) /
(total_municipios.count_municipios) * 100, 2) AS percent_municipios
FROM
municipios_sem_obito,
total_municipios
```

Resultado:

Row	count_municipios	count_municipios_sem_obito	percent_municipios
1	5570	25	99.55

6 - Qual a população total por estado , o município mais populoso de cada estado e a representatividade de concentração populacional em porcentagem deste município em relação ao total de habitantes do estado ?

Query:

```
WITH
  pop_por_estado AS (
    SELECT
      SUM(estimated_population) AS pop_estado,
      state
    FROM
      `covid.caso`
    WHERE
      is_last = TRUE
      AND place_type = 'state'
    GROUP BY
      state ),
  pop_por_municipio AS (
    SELECT
      SUM(estimated_population) AS pop_municipio,
      state,
      city,
      DENSE_RANK() OVER(PARTITION BY state ORDER BY (SUM(estimated_population))
DESC) AS ranking
    FROM
      `covid.caso`
    WHERE
      is_last = TRUE
      AND city_ibge_code IS NOT NULL
      AND city IS NOT NULL
      AND place_type = 'city'
    GROUP BY
      city,
      state )
SELECT
  pop_por_municipio.state AS Estado,
  pop_por_estado.pop_estado AS total_estado,
  pop_por_municipio.city AS municipio_mais_populoso,
  pop_por_municipio.pop_municipio AS total_municipio,
  ROUND((pop_por_municipio.pop_municipio / pop_por_estado.pop_estado) * 100, 2) AS
representatividade_percent
FROM
  pop_por_municipio
JOIN
  pop_por_estado
ON
  pop_por_estado.state = pop_por_municipio.state
WHERE
  pop_por_municipio.ranking = 1
ORDER BY
  Estado ASC
```

6 - Qual a população total por estado , o município mais populoso de cada estado e a representatividade de concentração populacional em porcentagem deste município em relação ao total de habitantes do estado ?

Resultado:

Row	Estado	total_estado	municipio_mais_populoso	total_municipio	representatividade_percent
1	AC	894470	Rio Branco	413418	46.22
2	AL	3351543	Maceió	1025360	30.59
3	AM	4207714	Manaus	2219580	52.75
4	AP	861773	Macapá	512902	59.52
5	BA	14930634	Salvador	2886698	19.33
6	CE	9187103	Fortaleza	2686612	29.24
7	DF	3055149	Brasília	3055149	100.0
8	ES	4064052	Serra	527240	12.97
9	GO	7113540	Goiânia	1536097	21.59
10	MA	7114598	São Luís	1108975	15.59
11	MG	21292666	Belo Horizonte	2521564	11.84
12	MS	2809394	Campo Grande	906092	32.25
13	MT	3526220	Cuiabá	618124	17.53
14	PA	8690745	Belém	1499641	17.26
15	PB	4039277	João Pessoa	817511	20.24
16	PE	9616621	Recife	1653461	17.19
17	PI	3281480	Teresina	868075	26.45
18	PR	11516840	Curitiba	1948626	16.92
19	RJ	17366189	Rio de Janeiro	6747815	38.86
20	RN	3534165	Natal	890480	25.2
21	RO	1796460	Porto Velho	539354	30.02
22	RR	631181	Boa Vista	419652	66.49
23	RS	11422973	Porto Alegre	1488252	13.03
24	SC	7252502	Joinville	597658	8.24
25	SE	2318822	Aracaju	664908	28.67
26	SP	46289333	São Paulo	12325232	26.63
27	TO	1590248	Palmas	306296	19.26

ETAPA 3 - DashBoard

Utilizando o Data Studio, crie um dashboard com os seguintes dados:

1 - FILTROS :

Data (dd/mm/yyyy) ,
Cidade ,
Estado

2 - TOTALIZADORES:

Total de casos Confirmados,
Total de Óbitos Confirmados,
Total de Municípios atingidos (que possuem pelo menos um caso)
Total de População
Total de Municípios com Óbitos

Exemplo:



3 - TABELA:

Crie uma tabela com:

a Data da ocorrência,
Município,
Estado (UF) ,
Quantidade de Confirmados,
Quantidade de confirmados por 100k habitantes,
Quantidade de Óbitos,
Taxa de letalidade e óbitos por 100k habitantes.

Adicione um filtro avançado ou do tipo caixa de entrada.

Exemplo:

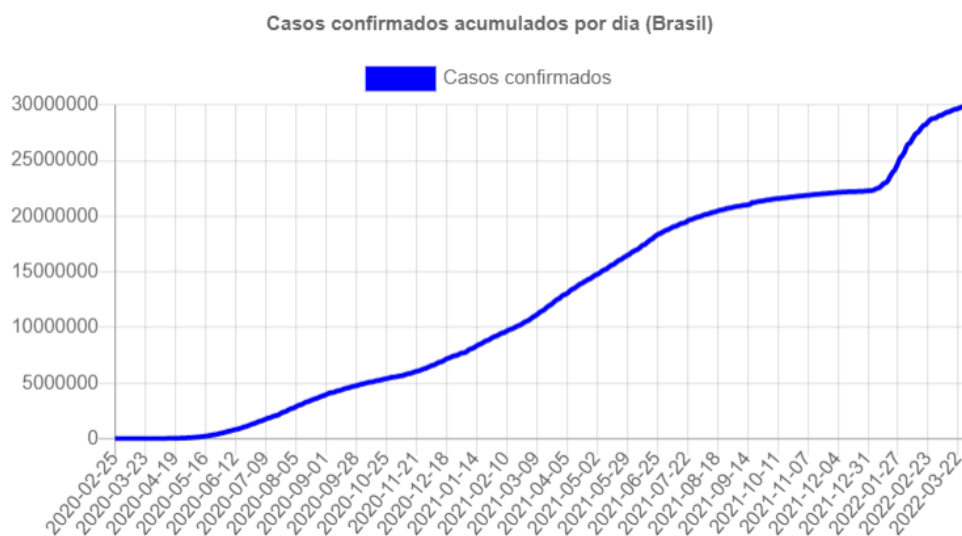
Buscar:

Data	Município	UF	Confirmados	Confirmados por 100k hab.	Óbitos	Letalidade	Óbitos por 100k hab.
06/12/2021	São Paulo	SP	976.214	7.920,45	39.352	4,03%	319,28
18/02/2022	Brasília	DF	597.871	19.569,29	10.363	1,73%	339,20
02/10/2021	Rio de Janeiro	RJ	484.433	7.179,11	34.102	7,04%	505,38
22/09/2021	Belo Horizonte	MG	278.370	11.039,58	6.636	2,38%	263,17
16/09/2021	Fortaleza	CE	256.441	9.545,14	9.654	3,76%	359,34
24/03/2022	Porto Alegre	RS	245.594	16.502,18	6.180	2,52%	415,25

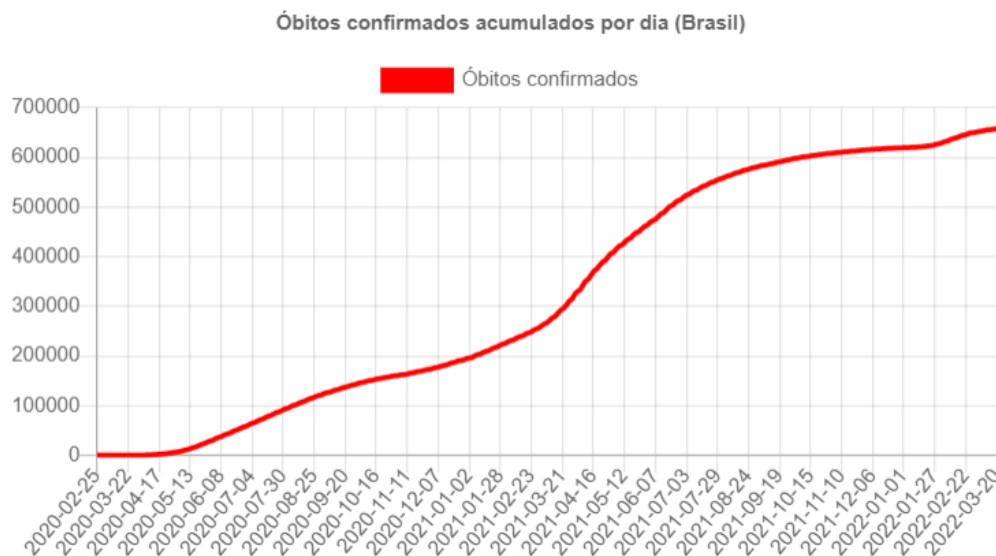
4 - GRÁFICO:

Crie um gráfico de linhas que representam os casos confirmados acumulados dia a dia no Brasil.

Exemplo:



Crie um gráfico de linhas que representam os óbitos acumulados dia a dia no Brasil.



fonte: <https://brasil.io/covid19/>

5 - De acordo com sua análise apresente pelo menos um gráfico/tabela que traga uma visão sua dos dados e explique o porquê de ter apresentado essa análise.

NOTA: Disponibilizar o link público do dashboard para análise. Pode-se inserir prints também como nos exemplos acima, mas o link do dashboard é obrigatório. O Layout do Dash também será levado em consideração.