

# *Introdução a Ciência de Dados*



Professor: Alex Pereira

# *Escopo*

- Pivotar tabela com CASE
  - Contar doses de vacinas
- Custom query
- Média móvel de 7 dias
- Record Linkage
  - Auxilio emergencial e socios
- Contar intercambialidade por combinação de imunizantes

# Join com dimensão de menor granularidade

- Adicionar ao modelo de dados o consumo de Energia por UF
  - [basedosdados.br\\_mme\\_consumo\\_energia\\_eletrica.uf](#)
- Seu modelo de dados tem granularidade por município
  - Se repetirmos o valor do consumo para cada registro
    - ✓ Não conseguiremos calcular o consumo do Brasil no Data Studio
      - A soma seria muito maior do que o valor real
- Como resolver ?

**Tabela Fato**

ano	sigla_uf	id_municipio	populacao	nome_municipio	pib
2002	RO	1100023	78039.0	Ariquemes	449592816.0
2003	RO	1100023	79680.0	Ariquemes	539636214.0
2004	RO	1100023	86901.0	Ariquemes	657193231.0
2005	RO	1100023	85031.0	Ariquemes	749021187.0
2006	RO	1100023	86924.0	Ariquemes	790696634.0

**Tabela da Dimensão de Consumo de Energia (MWh)**

ano	mes	sigla_uf	tipo_consumo	consumo	numero_consumidores
2004	1	RO	Total	112812.0	<i>null</i>
2004	1	AC	Total	34840.05	<i>null</i>
2004	1	AM	Total	274773.0	<i>null</i>
2004	1	RR	Total	31695.63	<i>null</i>
2004	1	PA	Total	1011353.04	<i>null</i>

# *Join com dimensão de menor granularidade*

- Solução

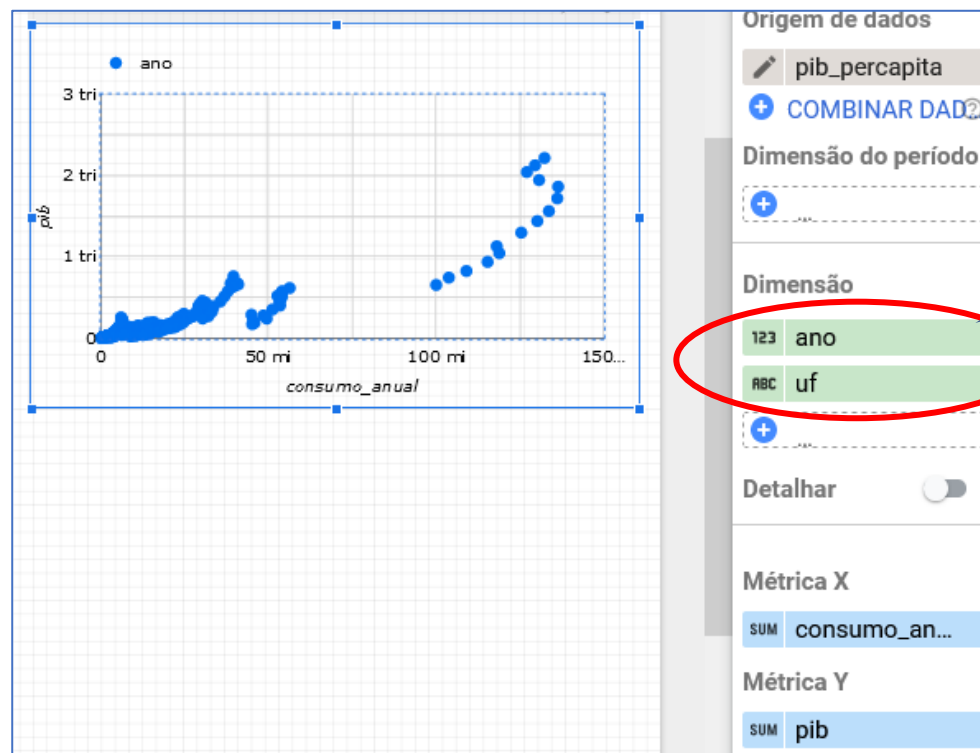
- Crie um dataframe com o consumo repetido para apenas um dos municípios
  - ✓ Não utilizar esta métrica como consumo de energia de municípios
- Existe mais alguma solução ?





# Atividade 8.1

- Adicione ao seu modelo de dados
  - Os dados do consumo de Energia por UF
    - ✓ Use o [caderno colab](#) disponibilizado pelo professor
- Criar no Data Studio um Scatter Plot com os valores
  - do consumo de energia e do PIB dos Estados



A ordem faz  
diferença

# *Pivotar tabela usando a função CASE*

- Calcular a quantidade de doses 1ª, 2ª, Única, Adicional e Reforço de vacina do COVID-19
  - Para cada UF, Semana, Imunizante
- Classificação das doses (do professor)
  - 1ª Dose
    - ✓ 1ª Dose, Dose, Dose Inicial
  - 2ª Dose
  - Reforço
    - ✓ Qualquer contendo a palavra Reforço
  - Adicional
    - ✓ Doses Adicional e 3ª Dose
  - Única

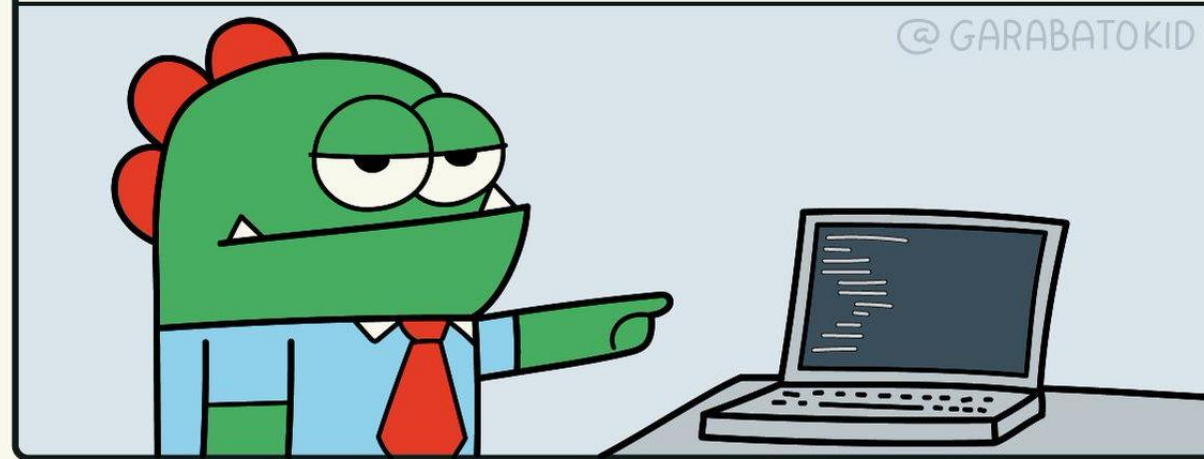
Linha	dose	Qtd
1	Única	224783
2	1º Reforço	33188
3	3º Reforço	2
4	Tratamento com dezessete doses	1
5	Revacinação	3
6	1ª Dose	154696905
7	2º Reforço	1879
8	Reforço	11565304
9	Dose Adicional	511211
10	Dose Inicial	1378
11	Tratamento com uma dose	2
12	1ª Dose Revacinação	759
13	2ª Dose Revacinação	862
14	Dose	4353953
15	2ª Dose	118663607
16	3ª Dose	308020

[Diferença entre 3ª Dose \(Adicional\) e Reforço](#)

# *Se entender, já está falando a língua dos nerds*

## HOW TO REGEX

STEP 1: OPEN YOUR FAVORITE EDITOR



STEP 2: LET YOUR CAT PLAY ON YOUR KEYBOARD



# Regex para detectar tipos de doses

- Usar a função REGEXP\_CONTAINS(value, regexp) do BigQuery

- REGEXP\_CONTAINS(dose, regexp)

- ✓ 1ª Dose

- 1ª Dose, Dose, Dose Inicial

- '1ª Dose\$|^Dose\$|Inicial'

- ✓ '2ª Dose\$'

- ✓ Reforço

- Qualquer contendo a palavra Reforço

- 'Reforço'

- ✓ Adicional

- Doses Adicional e 3ª Dose

- 'Dose Adicional|3ª Dose'

- ✓ 'Única'

- Solução

Linha	dose	Qtd
1	Única	224783
2	1º Reforço	33188
3	3º Reforço	2
4	Tratamento com dezessete doses	1
5	Revacinação	3
6	1ª Dose	154696905
7	2º Reforço	1879
8	Reforço	11565304
9	Dose Adicional	511211
10	Dose Inicial	1378
11	Tratamento com uma dose	2
12	1ª Dose Revacinação	759
13	2ª Dose Revacinação	862
14	Dose	4353953
15	2ª Dose	118663607
16	3ª Dose	308020

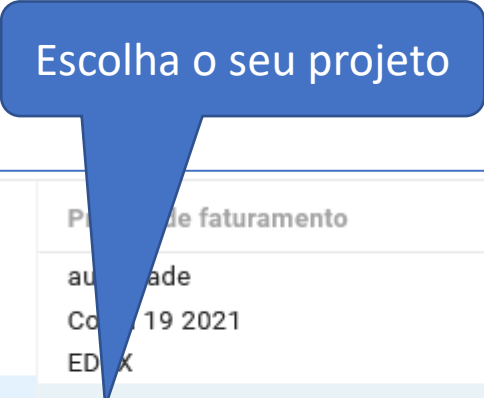


# *Custom query no Data Studio e BigQuery*

- Simulação de projeção de demanda de 2ª e 3ª Dose
  - A partir de input do usuário
    - ✓ no Data Studio
- Custom Query com Parâmetro
  - na [Documentação do BigQuery](#)

## Atividade 8.2

- Simulação de projeção de demanda de 2ª
  - A partir de input do usuário
    - ✓ no Data Studio e query no BigQuery
- Criar uma Consulta Personalizada
  - Escolha um projeto **SEU**

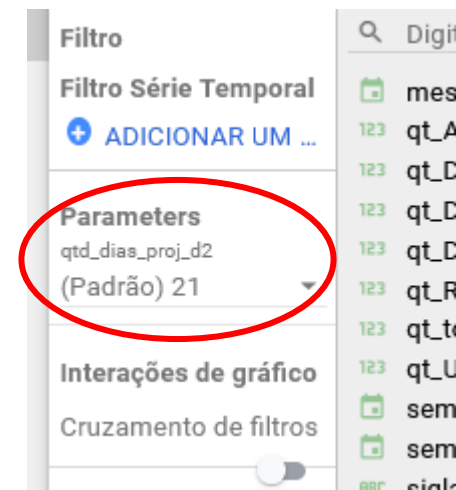


PROJETOS RECENTES	Projeto de faturamento
MEUS PROJETOS	aula de
PROJETOS COMPARTILHADOS	Curso 19 2021
CONSULTA PERSONALIZADA	EDX
CONJUNTOS DE DADOS PÚBLICOS	enap
	Google Play Android Developer
	IDP-MBA
	mscovid

```
SELECT v.sigla_uf, v.vacina_apelido, v.semana, v.mes, v.qt_total, v.qt_D1, v.qt_D2, v.qt_Reforco, v.qt_Adicional, v.qt_Unica,
vp.qt_D2_Proj, vp.semana_proj, vp.sigla_uf_proj, vp.vacina_apelido_proj
FROM `enap-331414.enapdatasets.vacinacao` v
JOIN (
  SELECT sigla_uf as sigla_uf_proj, vacina_apelido as vacina_apelido_proj, qt_D1 as qt_D2_Proj, DATE_ADD(semana, INTERVAL
@qtd_dias_proj_d2 DAY) as semana_proj
  FROM `enap-331414.enapdatasets.vacinacao`
) as vp
ON v.sigla_uf=vp.sigla_uf_proj and v.vacina_apelido=vp.vacina_apelido_proj and vp.semana_proj=v.semana
order by v.sigla_uf, v.vacina, v.semana, vp.semana_proj, vp.sigla_uf_proj, vp.vacina_apelido_proj
```

## Atividade 8.2

- Simulação de projeção de demanda de 2ª
  - A partir de input do usuário
    - ✓ no Data Studio e query no BigQuery
- Criar um gráfico de Série Temporal
  - Eixo x: data (**semana**)
    - ✓ Ajuste para o tipo semana ano
  - Eixo y: Projeção da 2ª dose (**qt\_D2\_Proj**)
  - Na métrica detalhada: nome da vacina (**vacina\_apelido**)
  - Ordenação: pelo campo semana
    - ✓ Crescente
  - Ative a opção Cumulativo na aba estilos do gráfico
    - ✓ Para as 4 séries
- Teste vários valores para o parâmetro **qtd\_dias\_proj\_d2**



## Atividade 8.3 – Visualizar a projeção futura no gráfico

- Simulação de projeção de demanda de 2ª
  - Mude a query para FULL OUTER JOIN

- Alterar a query para ficar assim

```
SELECT v.sigla_uf, v.vacina_apelido, v.semana, v.mes, v.qt_total, v.qt_D1, v.qt_D2, v.qt_Reforco, v.qt_Adicional, v.qt_Unica,  
vp.qt_D2_Proj, vp.semana_proj, vp.sigla_uf_proj, vp.vacina_apelido_proj
```

```
FROM `enap-331414.enapdatasets.vacinacao` v
```



```
FULL OUTER JOIN (
```

```
SELECT sigla_uf as sigla_uf_proj, vacina_apelido as vacina_apelido_proj, qt_D1 as qt_D2_Proj, DATE_ADD(semana,  
INTERVAL @qtd_dias_proj_d2 DAY) as semana_proj
```

```
FROM `enap-331414.enapdatasets.vacinacao`
```

```
) as vp ON v.sigla_uf=vp.sigla_uf_proj and v.vacina_apelido=vp.vacina_apelido_proj and vp.semana_proj=v.semana  
order by v.sigla_uf, v.vacina, v.semana, vp.semana_proj, vp.sigla_uf_proj, vp.vacina_apelido_proj
```

- Criar **2** (o da UF é opcional) campos calculado com as fórmula
  - IFNULL(semana, semana\_proj)
  - IFNULL(vacina\_apelido, vacina\_apelido\_proj)
  - IFNULL(sigla\_uf, sigla\_uf\_proj) – não será usado na série temporal
- Adicionar os 2 campos ao gráfico de Série Temporal

## *Atividade 8.4 (até o final da aula) - Difícil*

- Calcular a quantidade de doses de intercambialidade das vacinas da COVID-19
  - 1ª Dose / Dose de Reforço
    - ✓ Agregado por UF e tipo de combinação de imunizante



## Exercício 8.1

- Faça um dashboard no Google Data Studio, contendo
  - Uma página de índices das UFs
  - Na página detalhada de cada UF, apresente pelo menos
    - ✓ Um Scatter plot com 4 cores ajustadas dinamicamente
    - ✓ Um mapa com as mesmas cores do Scatter plot
- Sinta-se a vontade para colocar outros gráficos, KPIs, tabelas e outros elementos gráficos
- Recomendação:
  - Use o mesmo dataset, campos, parâmetros e gráficos
    - ✓ das atividades
- Submeta [aqui](#) um PDF e o link do seu dashboard.

