



Programa de Pós-Graduação

# Seminário

## SQL para Ciência de Dados

Prof. Msc. Gustavo Alexandre



Faculdade de Administração  
Universidade Federal Fluminense  
Laboratório de Inovação, Tecnologia e  
Sustentabilidade- LITS





## Ênfase

- Contexto de Ciência de Dados
- Orientação de Projeto de Dados
- Estruturação da informação da origem até destino
- Aplicabilidade baseada em soluções Python



# CONTEÚDO



## 01 PROJETO DE BANCO DE DADOS

*Modelando dados em SQL*

## 02 ESTRUTURAÇÃO DE DADOS

*Dados sociodemográficos*

## 03 MANIPULAÇÃO DE DADOS

*Cadastro de Atividade Econômicas*

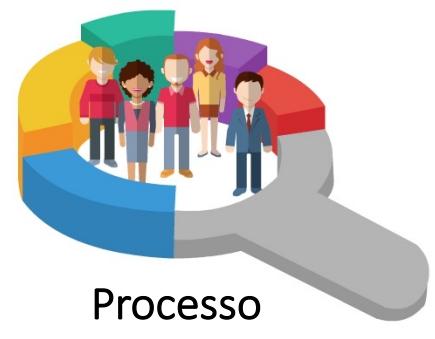
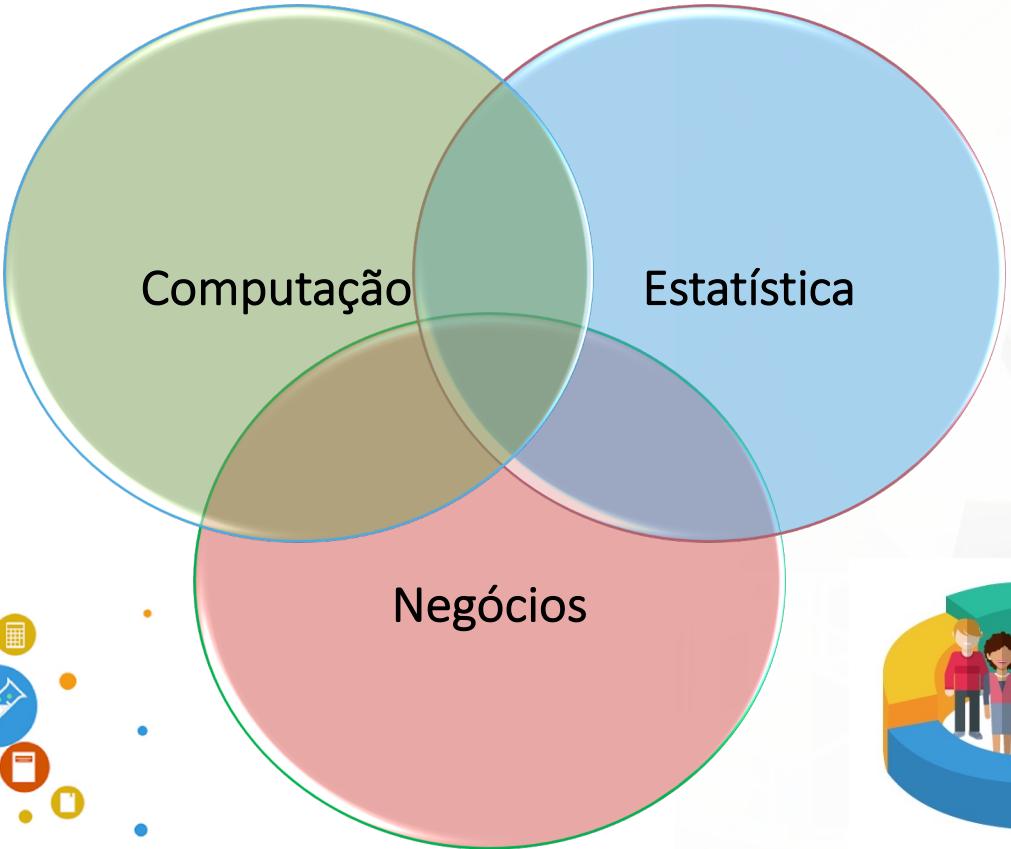
# CIÊNCIA DE DADOS



Pessoas



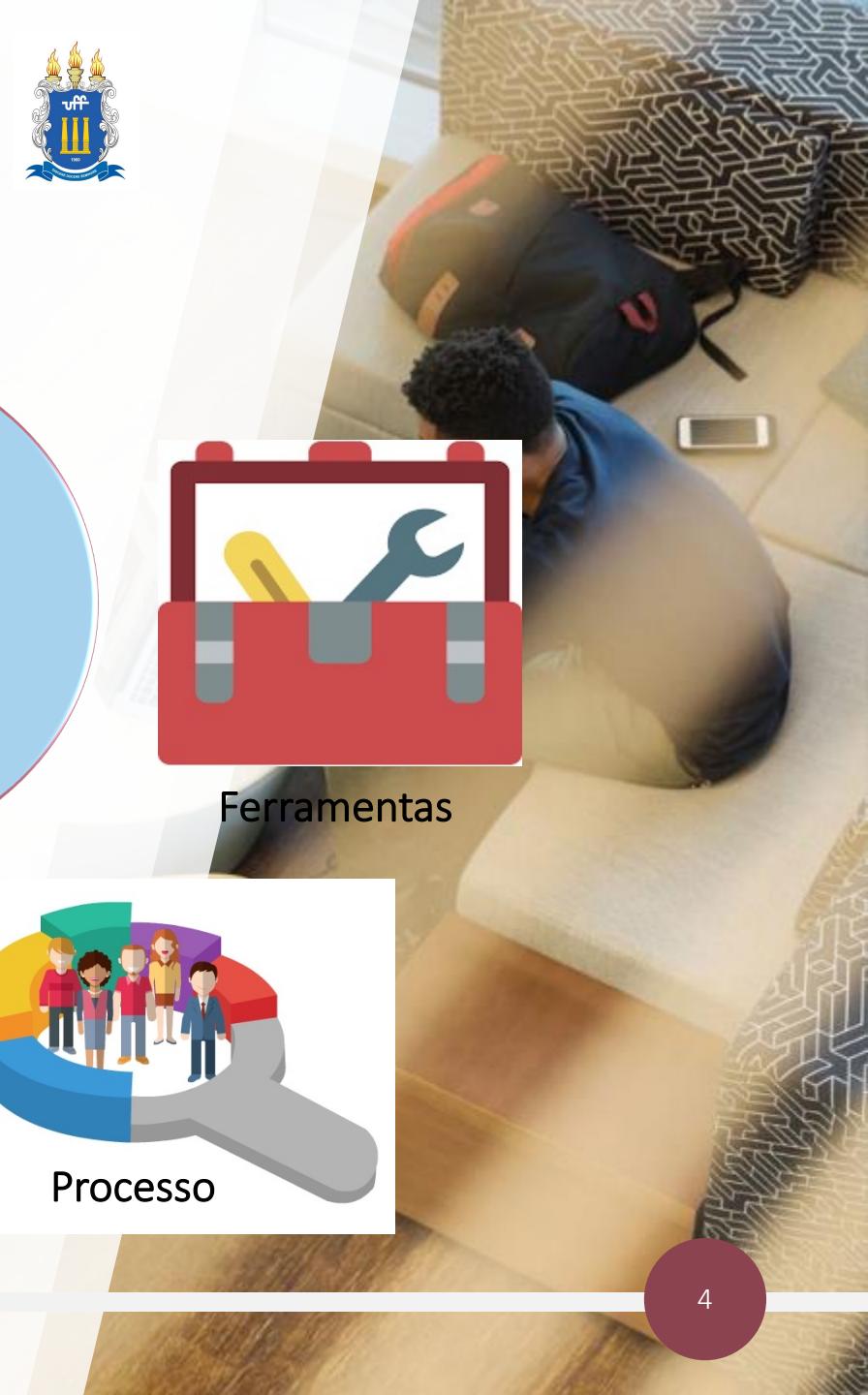
Competências



Processo



Ferramentas



# CIÊNCIA DE DADOS



- Manipulação de dados: Excel, CSV, MySQL, SQL Server, Oracle, MongoDB, etc.
- Programação: Python, SQL, Scala, Java, R, SAS, etc.
- Estatística: Testes estatísticos, Abordagem Frequentista e Bayesiana, etc.
- Aprendizado de Máquina: NLP, Classificação, Clusters, Deep Learning, etc.
- Visualização de dados: PowerBI, Tableau, SAS, D3.js, Python, R, etc.

Fonte: [Data science sareers shaping our future](#)

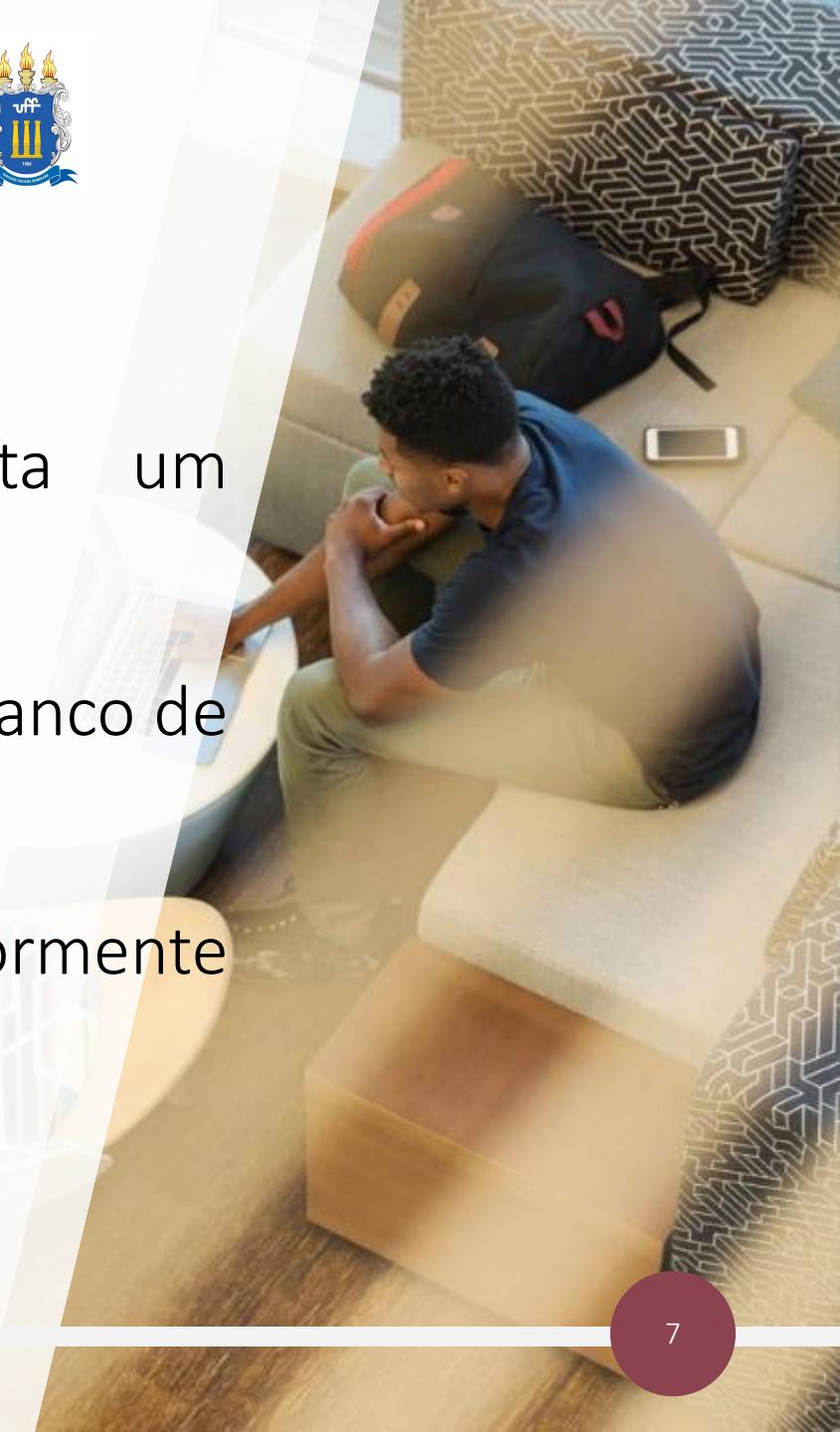
# PROJETO DE BANCO DE DADOS

- *Modelando dados em SQL*

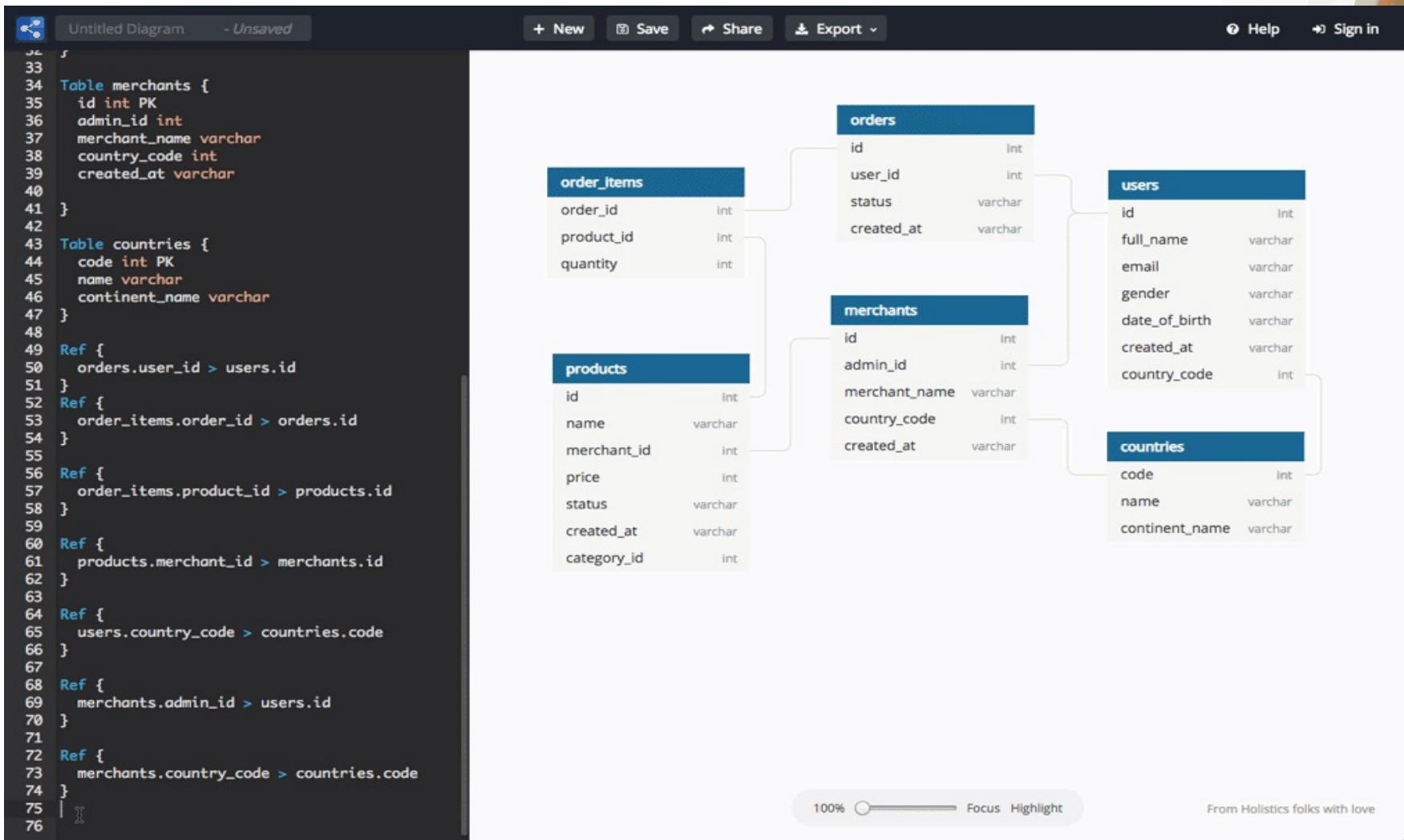
# PROJETO DE BANCO DE DADOS



- Conceitual: modelo conceitual, representa um diagrama entidade-relacionamento.
- Lógico: a partir do conceitual, define como o banco de dados será implementado em um SGBD.
- Físico: persiste o modelo descrito anteriormente fisicamente no banco de dados.



# PROJETO DE BANCO DE DADOS



# ESTRUTURAÇÃO DE DADOS

- *Dados sociodemográficos*

# ESTRUTURAÇÃO DE DADOS



COVID-19 ➔ Processos de fiscalização ➔ Comunicados importantes ➔ ENEM ➔ Contatos TCE-RJ

RESPOSTAS A CONSULTAS	SÚMULAS	PESQUISA TEXTUAL	PEDIDO DE DEFESA ORAL	MURAL DE EDITAIS	BANCO DE PREÇOS	INDICADOR DE TRANSPARÊNCIA	e-TCERJ
PAUTAS	BOLETINS DE JURISPRUDÊNCIA	CALENDÁRIO	SICODI	PORTAL DE DADOS - SERJ	ESCOLA DE CONTAS	PROTÓCOLO ELETRÔNICO	OVIDORIA

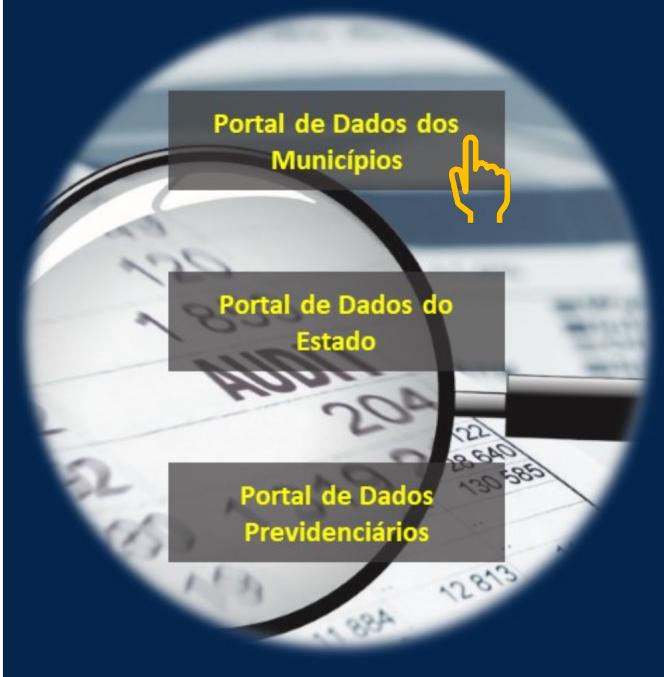
# ESTRUTURAÇÃO DE DADOS



TRIBUNAL DE CONTAS DO ESTADO DO RIO DE JANEIRO

## Bem-vindo ao TCE-RJ - Portal de Dados

O Tribunal de Contas do Estado do Rio de Janeiro é responsável pela fiscalização dos recursos públicos por parte de 91 municípios fluminenses, com exceção da capital, cuja administração é fiscalizada pelo Tribunal de Contas do Município.



Pelo Portal de Dados do TCE-RJ você pode acompanhar como se dá a aplicação de recursos dos órgãos e entidades jurisdicionados ao Tribunal. A ferramenta disponibiliza informações fornecidas pelo Governo do Estado e pelas prefeituras de cada um dos municípios sob jurisdição do TCE-RJ.

A disponibilização dos dados previdenciários tem como objetivo permitir a verificação sobre se o Regime Próprio de Previdência Social (RPPS) está cumprindo as exigências dispostas na legislação previdenciária, de acordo com a Secretaria Especial de Previdência e Trabalho.

A implementação do Portal de Dados é mais uma ação do TCE-RJ para aumentar a transparência e incentivar o controle social.



# ESTRUTURAÇÃO DE DADOS



**TCE RJ**  
TRIBUNAL DE CONTAS DO ESTADO DO RIO DE JANEIRO

**Bem-vindo** ao Portal de Dados dos Municípios TCE-RJ

**Selezione o Município**

ANGRA DOS REIS



**ANGRA DOS REIS**

O município tem uma área total de 825,1 quilômetros quadrados, correspondentes a 39,2% da área da Região da Costa Verde. Os limites municipais, no sentido horário, são: Estado de São Paulo, Rio Claro, Mangaratiba, Oceano Atlântico e Paraty.

2020	207.044	População Estima...	<a href="#">i</a>
2018	8.936.325,00	PIB	<a href="#">i</a>
2010	0,724	IDHM	<a href="#">i</a>

**Compras Diretas Geral** **Compras Diretas Covid: Del. 313** **Contratos** **Despesas com Educação** **Despesas com Saúde** **Licitações** **Obras Paralisadas** **Prestação Contas** [Saiba mais](#)

# ESTRUTURAÇÃO DE DADOS



**TCE RJ**  
TRIBUNAL DE CONTAS DO ESTADO DO RIO DE JANEIRO

[Retornar ao Portal](#)

[Detalhar Dados](#)

**Período**  
Todos

**Selecionar Município**  
ANGRA DOS REIS

**Quantidade Dispensa(s) no período no Município**  
20

**Valor Total Dispensa(s) no período no Município**  
R\$ 18,38 Mi

**Comparativo entre todos os Municípios**  
R\$ 1,19 Bi Valores Dispensados

© 2021 TomTom, © 2021 Microsoft Corporation. Terms

**Quantidade Dispensas por Período**

Possui Contrato: • Não • Sim

Mês	Não (Contrato)	Sim (Contrato)	Total
2020-04	6	0	6
2020-05	4	2	6
2020-06	3	0	3
2020-07	2	0	2
2020-10	1	0	1
2020-11	1	0	1
2021-03	1	0	1

**Relação municípios que não informaram gastos**

- CACHOEIRAS DE MACACU
- MACUCO
- SAO SEBASTIAO DO ALTO

**Top 5 Unidades Atos Emergenciais**

FUNDO MUN SAÚDE ANGRA DOS REIS	R\$ 14,63 Mi
FUNDO MUN ASSIST SOCIAL ANGRA DO...	R\$ 2,09 Mi
FUNDO MUN DIR CRI ADOLES ANGRA D...	R\$ 1,18 Mi
PREFEITURA ANGRA DOS REIS	R\$ 0,42 Mi
FUNDO MUN IDOSO ANGRA DOS REIS	R\$ 0,06 Mi

**Top 5 Fornecedores Atos Emergenciais**

LEMARC PRODUTOS MEDICOS HOSPITALARE...	R\$ 8,38 Mi
CONTRATE DE ANGRA CONSTRUOES LTDA	R\$ 2,89 Mi
V B SERVICOS COMERCIO E ADMINISTRACA...	R\$ 2,36 Mi
PHARMTECH DISTRIBUIDORA DE MEDICAMENT...	R\$ 1,95 Mi
LUMEN COMERCIO DE MATERIAL MEDICO E ...	R\$ 1,38 Mi

# ESTRUTURAÇÃO DE DADOS



**TCE RJ**  
TRIBUNAL DE CONTAS DO ESTADO DO RIO DE JANEIRO

Portal de Dados dos Municípios TCE-RJ - Detalhes COVID Deliberação nº 313/20

Retornar

Período: Todos | Selecionar Município: ANGRA DOS REIS | Valor Total Dispensa(s) no período no Município: R\$ 18,38 Mi | Quantidade Dispensa(s) no período no Município: 20

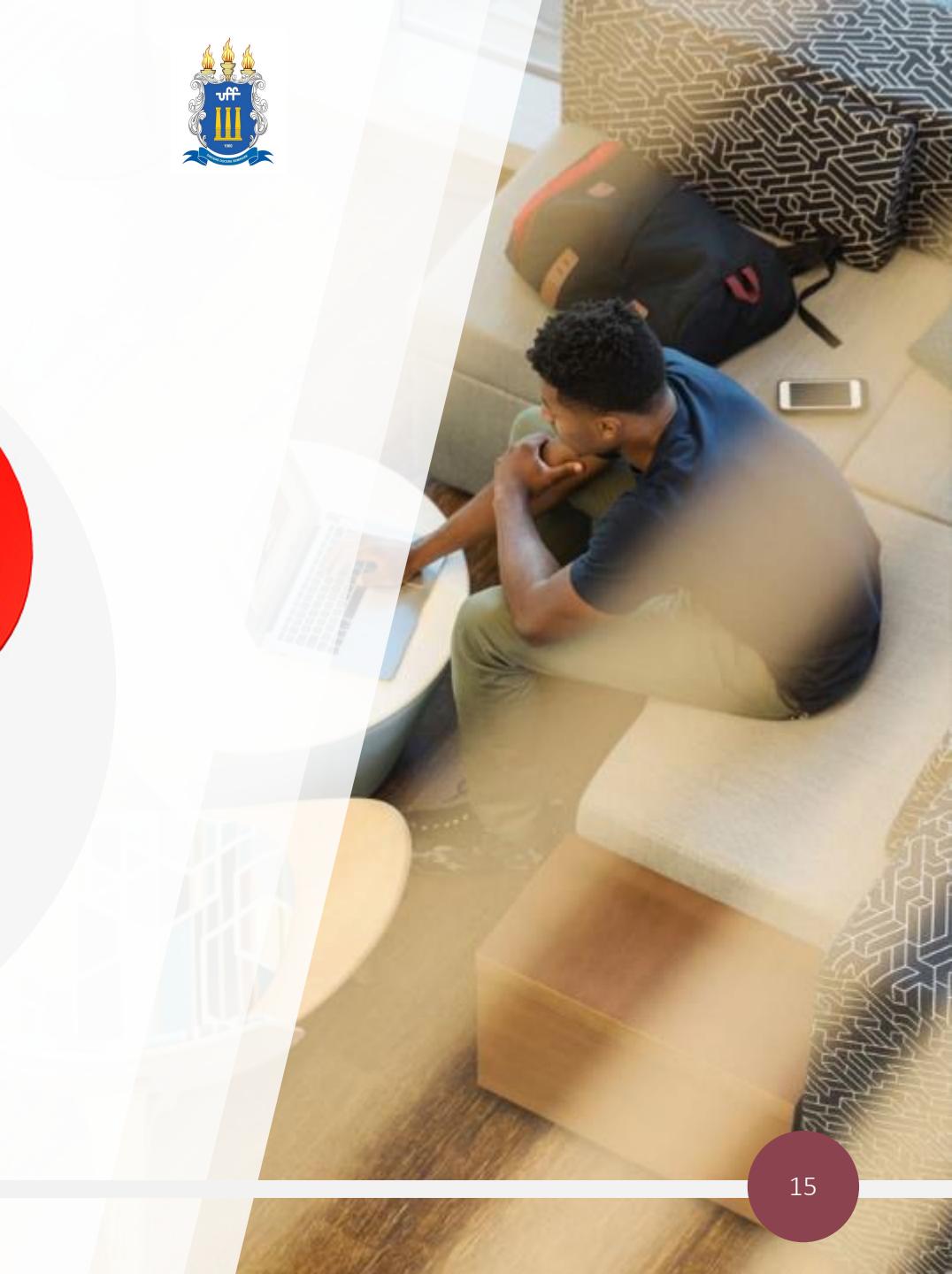
Pesquisa Fornecedor: Search

**Download dos Dados de Todos Municípios**

Informações da(s) Dispensa(s)										
Período	Número Interno	Ente	Unidade	Processo Administrativo	Protocolo Integração	Data Ato	Valor	CPF Ordenador	Nome Ordenador	CPF/ CNPJ Fornecedor
2020-05	284	ANGRA DOS REIS	FUNDO MUN SAÚDE ANGRA DOS REIS	2020006029	410452-7/2020	07/05/2020	R\$ 6.178.000,00	02796197743	RODRIGO DE ARAÚJO MUCHELI	093003000
2020-05	288	ANGRA DOS REIS	FUNDO MUN SAÚDE ANGRA DOS REIS	2020006077	410453-1/2020	07/05/2020	R\$ 2.200.000,00	02796197743	RODRIGO DE ARAÚJO MUCHELI	093003000
2020-05	290	ANGRA DOS REIS	FUNDO MUN SAÚDE ANGRA DOS REIS	2020006239	410454-5/2020	07/05/2020	R\$ 1.948.200,00	02796197743	RODRIGO DE ARAÚJO MUCHELI	2516538900
2020-04	371	ANGRA DOS REIS	FUNDO MUN ASSIST SOCIAL ANGRA DOS REIS	2020006229	410596-9/2020	16/04/2020	R\$ 88.740,00	79621066700	CELIA CRISTINA AMORIM SILVA JORDÃO	3075660400

Informações dos Itens da(s) Dispensa(s)				
Período	Número Interno	Unidade	Número Item	Descrição Item
2020-05	284	FUNDO MUN SAÚDE ANGRA DOS REIS	1	CAPOTE DESCARTÁVEL – CAPOTE DESCARTÁVEL, CONFECCIONADO EM SMS MANGAS LONGAS, BRANCO OU AMARELO COM BARREIRA PROTET COM AMARRAS CERVICAL E LOMBAR, COMPRIMENTO APROXIMADO 120CM E LARGURA DE 150CM, GRAMATURA DE 25G/M2 OU SUPERIOR.
2020-05	284	FUNDO MUN SAÚDE ANGRA DOS REIS	2	PROTETOR FACIAL – MÁSCARA, VISOR INCOLOR DE 6", COM APROXIMADAMENTE 150MM DE ALTURA E 195MM DE LARGURA, PLÁSTICO.
2020-05	284	FUNDO MUN SAÚDE ANGRA DOS REIS	3	MÁSCARA N95 OU FFP2 – COM FILTRAÇÃO ACIMA DE 95% PARA PARTÍCULAS DE 0,3 MICRON, COMPOSTA POR 04 CAMADAS, SENDO DUAS DE MELTBLOW E UM FILME LONCET DE POLIETILENO, QUE BLOQUEIA A ENTRADA DE FLUIDOS CORPÓREOS, COM DUPLO SISTEMA DE TIRES ELÁSTICAS.
2020-05	284	FUNDO MUN SAÚDE ANGRA DOS REIS	4	TOALHAS ECOHIGIENICAS DE USO HOSPITALAR PARA USO GETAL, NA COR BRANCA, MEDINDO 33 X 35CM, , ACONDICIONADO EM PACOTES C
2020-05	284	FUNDO MUN SAÚDE ANGRA DOS REIS	5	UAVA FARRICADA EM BORRACHA SINTÉTICA (NITRILO). CORANTE FOTOSENSIBILIZADOR COM AÇÃO ANTIMICROBIANA EFICIÊNCIA BACTERICIDA 99,99%

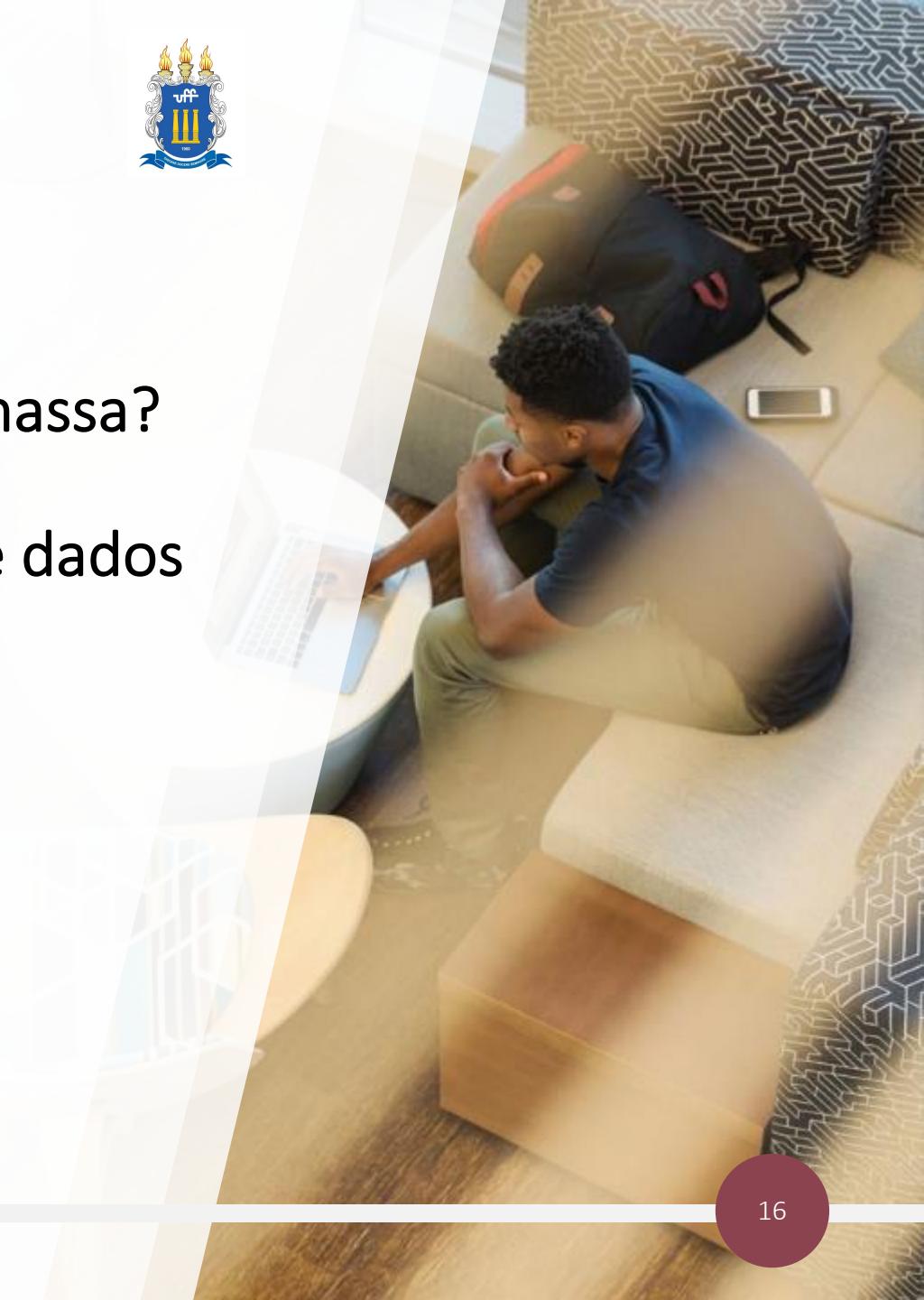
# DÚVIDAS



# ESTRUTURAÇÃO DE DADOS



- Que tal colocarmos um pouco a mão na massa?
- Vamos conhecer um grande arcabouço de dados públicos.



# ESTRUTURA E MANIPULAÇÃO (DW)



The screenshot shows the homepage of the BD+ website. At the top, there is a navigation bar with links for 'Dados', 'Documentação', 'Contato', 'Sobre', 'Newsletter', 'Apoie', a search bar, 'Entrar', and 'Cadastrar'. Below the navigation bar, there is a large green text block: 'Um único lugar para buscar e acessar os dados que você precisa'. To the right of this text is a search bar placeholder: 'Pesquisar palavras-chave, instituições ou tema →'. Below the search bar, there is a section for 'Termos populares' with links to 'lei', 'mortalidade', and 'COVID19'. At the bottom of the page, there are three sections with icons and descriptions: 1. A magnifying glass icon with a funnel and a building icon, describing the search mechanism. 2. A circular arrow icon with a download icon and a database icon, describing the availability of treated and updated data from a public datalake. 3. An R icon with a Python icon and code snippets, describing the programming packages available for Python and R.

B D

Dados Documentação Contato Sobre Newsletter Apoie Entrar Cadastrar

Um único lugar para buscar e acessar os dados que você precisa

Pesquisar palavras-chave, instituições ou tema →

Termos populares: lei mortalidade COVID19

Com o mecanismo de busca é possível descobrir informações sobre mais de 900 bases de dados de diversos temas e organizações.

Disponibilizamos o download dos dados tratados e atualizados direto do nosso *datalake* público num só click.

Através dos nossos pacotes de programação você pode acessar o *datalake* público BD+ em Python, R ou pela linha de comando.

# ESTRUTURA E MANIPULAÇÃO (DW)



B D Documentação

Buscar

basedosdados/mais  
v1.5.7 ⭐239 50

Documentação

Home

Aprenda sobre a BD

**BigQuery**

Primeiros passos

- Antes de começar: Crie o seu projeto no Google Cloud
- Acessando o datalake da [basedosdados](#)
- Faça sua primeira consulta!

Tutoriais

- Como navegar pelo BigQuery
- Entenda os dados
- Conectando com o PowerBI
- Manuais e Cursos de SQL

Pacotes

Contribua

## BigQuery

O BigQuery é um serviço de banco de dados em nuvem da Google. Você faz consultas ao banco em SQL direto do navegador com:

- Rapidez:** Mesmo queries muito longas demoram apenas minutos para serem processadas.
- Escala:** O BigQuery escala magicamente para hexabytes se necessário.
- Economia:** Todo usuário possui 1 TB gratuito por mês para consulta aos dados.

Pronto(a) para começar? Nesta página você encontra:

- [Primeiros passos](#)
- [Tutoriais](#)
- [Manuais e Cursos de SQL](#)

### Primeiros passos

Antes de começar: Crie o seu projeto no Google Cloud

Para criar um projeto no Google Cloud basta ter um email cadastrado no Google. É necessário ter um projeto seu, mesmo que vazio, para você fazer queries em nosso *datalake* público.

1. [Acesse o Google Cloud](#). Caso for a sua primeira vez, aceite o Termo de Serviços.



# ESTRUTURA E MANIPULAÇÃO (DW)



Acessando o datalake da basedosdados

O botão abaixo irá te direcionar ao nosso projeto no Google BigQuery:

[Ir para BigQuery →](#)

Na sua tela deverá aparecer o projeto fixado no menu lateral esquerdo, como na imagem abaixo.

A screenshot of the Google BigQuery web interface. On the left, there is a sidebar titled 'Resources' with a search bar above it. Below the search bar, there are three items: 'basedosdados' (which is highlighted with a red box), 'pr-educacao-rio', and 'robusta-lab'. At the top right of the main area, there are several buttons: 'Run', 'Save query', 'Save view', 'Schedule query', and 'More'. Below these buttons, the text 'basedosdados' is displayed. Underneath the dataset name, the text 'Datasets and tables available' is shown, followed by the instruction 'Use the Resources tree to view your data, or create a new dataset using the controls above.' In the bottom right corner of the main area, there is a circular icon containing a magnifying glass over a document.



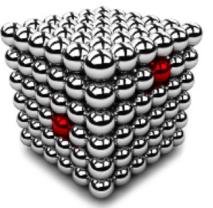
## MANIPULAÇÃO DE DADOS

- *Cadastro de Atividade Econômicas*

# MANIPULAÇÃO DE DADOS



API de serviço de dados



## Agregados

Análise multidimensional ao seu alcance.

3.0



## Calendário

Conheça o cronograma de ações e publicações do IBGE

3.0



## CNAE

API referente à Classificação Nacional de Atividades Econômicas

2.0



## Localidades

Obtenha os dados sobre as divisões administrativas do Brasil



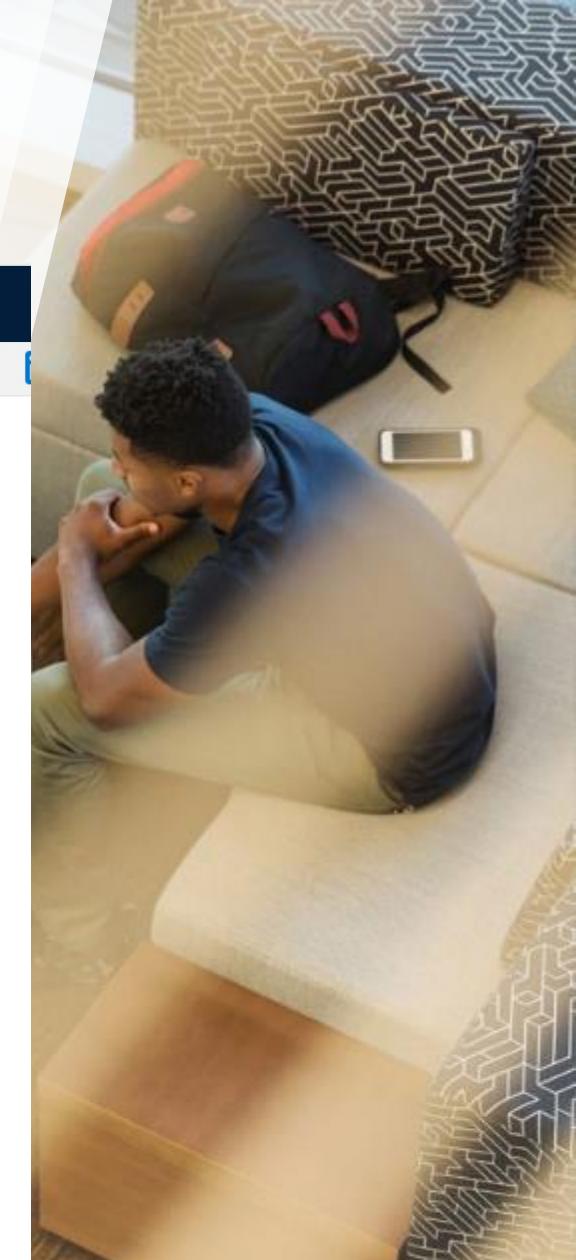
## Malhas geográficas

Malhas com as mais diversas resoluções e formatos para você liberar a sua imaginação



## Metadados

Metadados associados às pesquisas



# MANIPULAÇÃO DE DADOS



**SUMÁRIO**

**CLASSES**

- Classe por identificador
- Classes
- Classes por divisão
- Classes por grupo
- Classes por seção

**DIVISÕES**

- Divisão por identificador
- Divisões
- Divisões por seção

**GRUPOS**

- Grupos por divisão
- Grupos
- Grupo por identificador
- Grupos por seção

**SEÇÕES**

- Seções
- Seção por identificador

**SUBCLASSES**

- Subclasses por classe
- Subclasses por divisão
- Subclasses por grupo
- Subclasses por seção
- Subclasses
- Subclasse por identificador

**NOTAS DE LIBERAÇÃO**

## API CNAE - Cadastro Nacional de Atividades Econômicas

### API e documentação

Versão: 2.0.0

A CNAE, Classificação Nacional de Atividades Econômicas, é a classificação oficialmente adotada pelo Sistema Estatístico Nacional na produção de estatísticas por tipo de atividade econômica, e pela Administração Pública, na identificação da atividade econômica em cadastros e registros de pessoa jurídica. A CNAE não faz distinção entre tipo de propriedade, natureza jurídica, tamanho do negócio e modo de operação, uma vez que tais critérios não interferem na caracterização da atividade em si.

O IBGE é o órgão gestor da CNAE, responsável pela documentação da classificação, desenvolvimento dos instrumentos de apoio, disseminação e atendimento aos usuários sobre a aplicação da classificação. Compete-lhe ainda a manutenção da classificação e a condução dos processos de revisão.

Exceptuando as subclasses, que estão na versão 2.2, a presente versão 2.0, resultante da revisão 2007, substituiu a 1.0. Essa versão é uma classificação organizada de forma hierárquica com cinco níveis, 21 seções, 87 divisões, 285 grupos, 673 classes e 1301 subclasses. As subclasses são definidas para uso da Administração Pública.

- 1. Seção
- 2. Divisão
- 3. Grupo
- 4. Classe
- 5. Subclasse
- 6. Atividade econômica

Para mais informações, acesse [https://cnae.ibge.gov.br/images/concla/documentacao/CNAE20\\_Introducao.pdf](https://cnae.ibge.gov.br/images/concla/documentacao/CNAE20_Introducao.pdf)

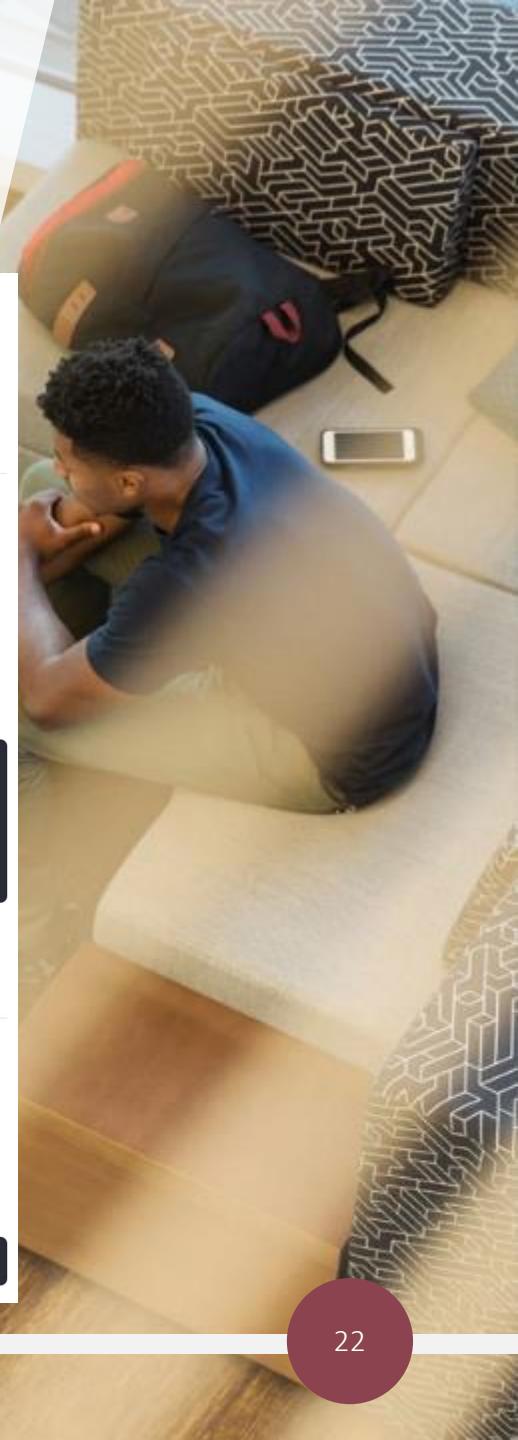
**obs:** a partir do dia primeiro de janeiro de 2019, as subclasses, e somente as subclasses, serão migradas para a versão 2.3. Portanto, a partir da referida data, a versão 2.2 será descontinuada.

### Classe por identificador

Obtém o conjunto de classes a partir dos respectivos identificadores

**GET**

`https://servicodados.ibge.gov.br/api/v2/cnae/classes/{classe}`



# PERGUNTAS



# OBRIGADO



[Prof. Gustavo Alexandre](#)



+55 21 98097-1668



[gassantos@id.uff.br](mailto:gassantos@id.uff.br)



[github.com/gassantos](https://github.com/gassantos)

