

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

NÚCLEO DE EDUCAÇÃO A DISTÂNCIA Pós-graduação *Lato Sensu* em Gestão e Análise Estratégica de Dados

RELATÓRIO TÉCNICO

ANÁLISE DO CENÁRIO DA INCLUSÃO DIGITAL NO BRASIL

Ana Claudia de Morais Rosa

Belo Horizonte 2023

SUMÁRIO

1. Introdução	3
1.1. Contexto	3
1.2. Objetivos	4
1.3. Público alvo	5
2. Modelos de Dados	5
2.1. Modelo Dimensional	5
2.2. Fatos e Dimensões	6
3. Integração, Tratamento e Carga de Dados	11
3.1. Fontes de Dados	11
3.2. Processos de Integração e Carga (ETL)	11
6. Links	17
REFERÊNCIAS	19

1. Introdução

1.1. Contexto

A inclusão digital e a universalização do acesso à internet são temas que tem sido debatidos há muito tempo no Brasil e no mundo, sendo inclusive pauta de políticas públicas e objeto de ações de diferentes agentes sociais — ONGs, Universidades, empresas e escolas. Todavia, por mais que as pesquisas atuais mostrem um aumento progressivo em relação à inclusão digital no Brasil, de acordo com dados divulgados pelo Instituto Brasileiro de Geografia e Estatística (IBGE) em 2021, cerca de 38% da população brasileira ainda não possui acesso à internet em casa. Além disso, a desigualdade de acesso à internet no Brasil é evidenciada pela diferença de acesso entre diferentes regiões do país, sendo a região Norte a que apresenta o menor índice de acesso (IBGE).

Essa falta de acesso à internet e às TICs pode prejudicar a formação do cidadão e a sua inclusão na sociedade, especialmente em um momento em que a tecnologia se tornou uma parte fundamental da vida cotidiana. Além disso, com o avanço da Indústria 4.0, a exclusão digital pode se tornar ainda mais prejudicial e aumentar as desigualdades sociais no país.

Um dos assuntos mais buscados no mundo, enquanto realizo esse trabalho, é em relação ao ChatGPT, um software de um chatbot com inteligência artificial, desenvolvido pela OpenAI, especializado em diálogo e que foi desenvolvido com técnicas de aprendizado supervisionado que possibilitam uma criação de conteúdo de altíssima qualidade, com respostas extremamente coerentes e coesas e que parecem ter sido escritas por um ser humano. Há quem diga que o impacto do ChatGPT na sociedade poderá ser comparável ao desenvolvimento do buscador Google. Mas, de que maneira poderíamos utilizar o ChatGPT como solução para a exclusão digital, por exemplo? De repente, ele poderia ser uma ferramenta útil para ampliar o acesso à informação e ao conhecimento para aqueles que não têm conexão com a internet. É importante que a sociedade entenda o avanço da tecnologia, que ora possa parecer assustador para uma parcela dela, não como um vilão simplesmente, mas sim, como um possível caminho que auxilie no combate a esse abismo digital.

Em resumo, a inclusão digital é um desafio importante para o desenvolvimento do Brasil, e requer esforços coletivos e coordenados de diferentes agentes sociais. Nesse contexto, é importante que sejam implementadas políticas públicas específicas para cada nicho social, que haja investimentos em infraestrutura e tecnologia, além de ações de conscientização e capacitação, na tentativa de garantir que todos os cidadãos tenham acesso à internet e às TICs, e que possam desfrutar dos benefícios da tecnologia.

1.2. Objetivos

O objetivo principal do presente trabalho é de conhecer a realidade da população brasileira quanto ao acesso e uso das TICs (Tecnologias da Informação e Comunicação) e, a partir deste entendimento, poder auxiliar os agentes interessados no estabelecimento de políticas públicas mais assertivas para as diversas realidades existentes. Este entendimento se dará a partir da análise dos dados da pesquisa realizada pelo Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br) e pelo departamento do Núcleo de Informação e Coordenação do Ponto BR (NIC.br), disponível publicamente no site da Cetic.br.

Para isso, propõe-se responder, inicialmente, à sete perguntas conforme seguem:

- 1) Qual o percentual de domicílios com computador de mesa e/ou notebook? Por área, região e classe social?
- 2) Qual o percentual de domicílios com acesso à internet e o principal tipo de conexão?
- 3) Dos domicílios sem acesso à internet, qual é o motivo principal?
- 4) Qual o perfil dos usuários de internet? (Por área, sexo, faixa etária, grau de instrução e classe social)
- 5) Qual a frequência de uso da internet nos últimos 3 meses?
- 6) Qual o perfil dos usuários que compraram produtos e serviços pela internet?

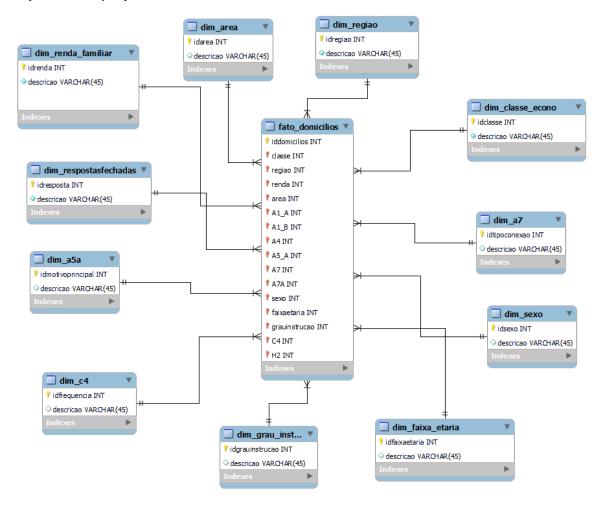
1.3. Público alvo

Este trabalho tem como público-alvo principal o governo, a fim de incentivar a elaboração de políticas públicas assertivas para realidade atual, mas almeja atingir também empresas privadas, Universidades e ONGs interessadas em atuar no âmbito da inclusão digital, além da mídia e da própria sociedade, já que se trata de um assunto de interesse público.

2. Modelos de Dados

2.1. Modelo Dimensional

Utilizando o software MySQL Workbench foi desenvolvido diagrama do modelo dimensional, que tem como finalidade oferecer uma visão sobre como os dados foram organizados para atender às necessidades identificadas por meio dos objetivos do projeto:



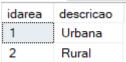
2.2. Fatos e Dimensões

Foram definidas 1 tabela Fato e 11 tabelas dimensão para o projeto, tal qual descritas abaixo:

DIM_AREA: tabela que separa as áreas em Urbana e Rural

Campos:

idarea(PK): tipo numérico descricao : tipo texto

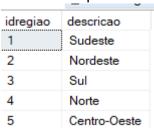


DIM_REGIAO: tabela que descreve as 5 regiões do Brasil

Campos:

idregiao(PK): tipo numérico

descricao: tipo texto



• DIM_CLASSE_ECONO: tabela que descreve a divisão de classes econômicas existentes no Brasil.

Campos:

idclasse(PK): tipo numérico

descricao: tipo texto

idclasse	descricao
1	Α
2	В
3	С
4	DE

• DIM_RENDA_FAMILIAR: tabela que apresenta a descrição de cada faixa da renda familiar, conforme o número de salários mínimos.

Campos:

idrenda(PK): tipo numérico

descricao: tipo texto

idrenda	descricao
1	Até 1 SM
2	Mais de 1 SM até 2 SM
3	Mais de 2 SM até 3 SM
4	Mais de 3 SM até 5 SM
5	Mais de 5 SM até 10 SM
6	Mais de 10 SM
7	Não tem renda
97	Não sabe
98	Não respondeu

• DIM_RESPOSTASFECHADAS: tabela que apresenta a descrição das respostas em relação às perguntas fechadas do questionário utilizado.

Campos:

idresposta(PK): tipo numérico

descricao: tipo texto

idresposta	descricao
0	Não
1	Sim
97	Não sabe
98	Não respondeu
99	Não se aplica
98	Sim Não sabe Não respondeu

Essa tabela dimensão possui as repostas para as perguntas fechadas e tem referencia com os campos A1_A, A1_B, A4, A5A, A7A e H2 da Tabela Fato.

• DIM_A5a: tabela que apresenta as respostas à seguinte pergunta do questionário utilizado: "Qual o principal motivo que o domicílio não tem acesso à internet?"

Campos:

idtipoconexao (PK): tipo numérico

descricao: tipo texto

idmotivoprincipal	descricao
1	Falta de computador
2	Falta de necessidade
3	Falta de interesse
4	Têm acesso em outro lugar
5	Acham muito caro
6	Não sabem usar Internet
7	Falta de disponibilidade
8	Preocupações com segurança ou privacidade
9	Evitam o contato com conteúdo perigoso
10	Outro motivo
97	Não sabe
98	Não respondeu
99	Não se aplica

• DIM_A7: tabela que apresenta as respostas à seguinte pergunta do questionário sobre o domicílio: "Qual o principal tipo de conexão utilizado para acessar a Internet no domicílio?"

Campos:

idtipoconexao (PK): tipo numérico

descricao: tipo texto

idtipoconexao	descricao
1	Conexão discada
2	Conexão DSL
3	Conexão via cabo ou fibra ótica
4	Conexão via Rádio
5	Conexão via Satélite
6	Conexão via modem (3G ou 4G)
97	Não sabe
98	Não respondeu
99	Não se aplica

• DIM_SEXO: tabela que apresenta qual o sexo do indivíduo respondente pela pesquisa.

Campos:

idsexo (PK): tipo numérico

descricao: tipo texto

idsexo	descricao
1	Masculino
2	Feminino

• DIM_FAIXA_ETARIA: tabela que apresenta qual a faixa etária do indivíduo respondente pela pesquisa, dentre 6 faixas.

Campos:

idfaixaetaria (PK): tipo numérico

descricao: tipo texto

idfaixaetaria	descricao
1	De 10 a 15 anos
2	De 16 a 24 anos
3	De 25 a 34 anos
4	De 35 a 44 anos
5	De 45 a 59 anos
6	60 anos ou mais

• DIM_GRAU_INSTRUCAO: tabela que apresenta qual o grau de instrução do indivíduo respondente pela pesquisa, dentre 12 classificações..

Campos:

idgrauinstrucao (PK): tipo numérico

descricao: tipo texto

idgrauinstrucao	descricao
1	Analfabeto
2	Sabe ler/escrever, mas não cursou escola
3	Até pré-escola incompleto
4	Pré-escola completo
5	1a até 3a série
6	4a até 5a série
7	5a até 7a série
8	8a série (9o ano)
9	1a e 2a série
10	3a série/vestibular
11	Superior incompleto
12	Superior completo

• DIM_C4: tabela que apresenta as respostas à seguinte pergunta do questionário sobre os indivíduos: "Em média, com que frequência o respondente usou a Internet nos últimos 3 meses?"

Campos:

idfrequencia (PK): tipo numérico

descricao: tipo texto

idfrequencia	descricao
1	Todos os dias ou quase todos os dias
2	Pelo menos uma vez por semana
3	Pelo menos uma vez por mês
4	Menos do que uma vez por mês
99	Não se aplica

• FATO_DOMICILIOS: tabela onde encontram-se as respostas às perguntas realizadas durante a pesquisa e que são determinantes para responder às perguntas elencadas no objetivo deste projeto.

A tabela possui um total de 23950 registros.

Cada domicílio possui um ID único de identificação e cada tabela dimensão relaciona-se com essa tabela fato pela primary key (pk) das dimensões.

Campos:

iddomicilio (pk): tipo numérico

classe: tipo numérico regiao: tipo numérico renda: tipo numérico area: tipo numérico sexo: tipo numérico

faixa_etaria: tipo numérico grau_instrucao: tipo numérico

A1_A: tipo numérico A1_B: tipo numérico A4: tipo numérico A5A: tipo numérico A7: tipo numérico A7A: tipo numérico C4: tipo numérico H2: tipo numérico

iddomicilio	classe	regiao	renda	area	sexo	faixa_etaria	grau_instrucao	A1_A	A1_B	A4	A5A	A7	A7A	C4	H2
1	2	4	5	1	1	5	12	0	1	1	99	3	1	1	1
2	4	4	1	1	2	3	12	0	0	1	99	3	1	1	0
3	3	4	3	1	2	5	7	0	0	1	99	6	0	1	0
4	3	4	1	1	2	2	10	0	0	1	99	1	1	1	0
5	4	4	2	1	2	3	9	0	0	1	99	6	0	1	0
6	4	4	2	1	1	2	9	0	0	1	99	5	1	1	0
7	4	4	7	1	1	2	7	0	0	1	99	3	1	1	1
8	4	4	2	1	2	5	6	0	0	1	99	3	1	1	0
9	3	4	5	1	1	1	7	0	1	1	99	4	1	1	0
10	3	4	2	1	1	4	6	0	0	1	99	3	1	1	0
11	3	4	4	1	2	5	10	0	1	1	99	3	1	1	0
12	3	4	2	1	2	5	11	0	0	1	99	5	1	1	0
13	4	4	2	1	2	2	10	0	0	1	99	6	1	1	0
14	2	4	2	1	2	3	10	0	0	1	99	3	1	1	1
15	3	4	2	1	1	6	1	0	0	1	99	3	1	99	99
16	4	4	2	1	1	6	10	1	0	1	99	3	1	1	0
17	3	4	1	1	1	5	6	1	0	1	99	3	1	1	0
18	4	4	1	1	2	4	10	0	0	1	99	6	0	1	0
19	3	4	2	1	2	3	9	0	0	1	99	2	1	1	1
20	3	4	3	1	2	2	10	0	1	1	99	2	1	1	1
21	3	4	1	1	NULL	NULL	NULL	0	0	1	99	6	1	NULL	NULL
22	4	4	2	1	1	5	6	0	0	0	4	99	99	2	1
23	3	4	1	1	2	3	9	0	0	1	99	6	0	1	0
24	3	4	4	1	1	5	7	0	0	1	99	2	1	1	0
25	3	4	4	1	1	5	9	1	0	1	99	2	1	1	1
26	4	4	2	1	1	3	9	0	0	0	4	99	99	1	0

3. Integração, Tratamento e Carga de Dados

3.1. Fontes de Dados

As bases de dados utilizadas foram retiradas do site CETIC.br (Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação), com a pesquisa TIC Domicílios de 2021, sendo elas:

- Bases de microdados:
 - 1 arquivos .csv com as respostas dos informantes sobre o domicílio que reside.
 - 1 arquivos .csv com as respostas pessoais dos indivíduos respondentes.
- Dicionário de dados para identificação das variáveis contidas nas bases:
 1 arquivo .csv para cada base de microdados (domicílio e indivíduos), contendo as perguntas realizadas no questionário, bem como a descrição das variáveis de cada campo.

3.2. Processos de Integração e Carga (ETL)

Para construção das tabelas Fato e Dimensões foi utilizado o banco Microsoft SQL Server Management Studio.

Pelo usuário master, criou-se um banco cujo nome é bd_tcc, para armazenar as tabelas que seriam utilizadas.

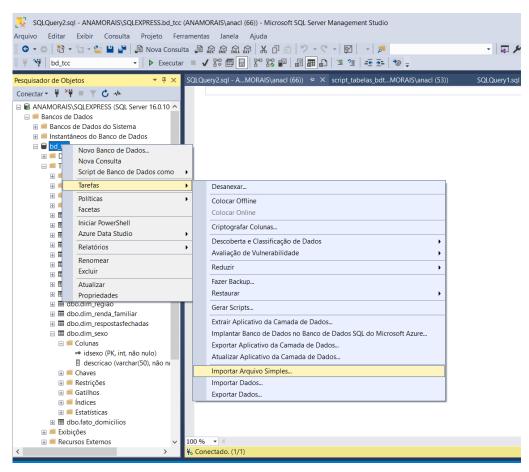
```
□USE MASTER;

CREATE DATABASE [bd_tcc];
```

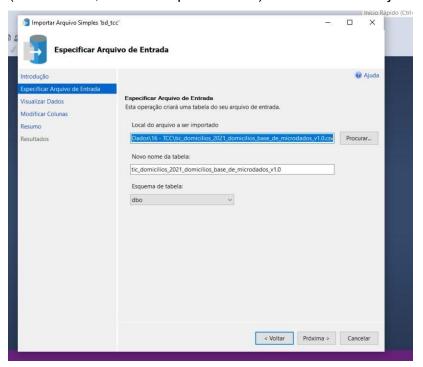
O segundo passo foi realizar o processo de ingestão no banco, dos arquivos .csv que contém os microdados dos questionários.

Para cada uma das bases, foi utilizada uma etapa manual e pré-existente no SQL Server de importação dos dados .csv, tal qual consta nas imagens dos fluxos abaixo:

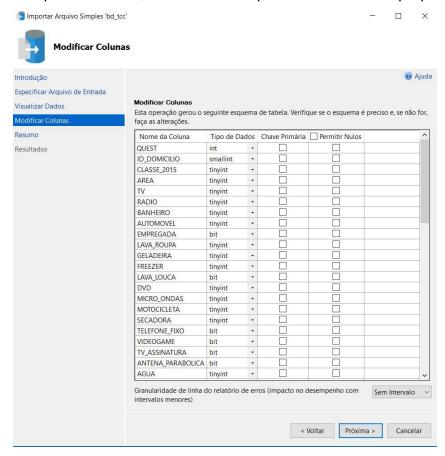
1 a etapa: Sobre o banco bd_tcc, clica-se com o botão da direita do mouse, seleciona o comando Tarefas >> Importar Arquivo Simples



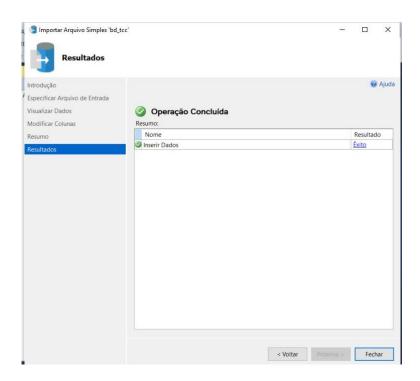
2 ^a **etapa**: Procurar no computador o local do arquivo de entrada, definir o nome que será dado à tabela (basedomicilios e baseindividuos), o schema onde será inserida (nesse caso, o schema padrão dbo) e clicar em avançar.



3 a etapa: Modificar as colunas, alterando os tipos de dados e/ou nomes dos campos da tabela, definir a chave prímária e se o campo permite ou não valor nulo.



4 ª etapa: Verificar os resultados e concluir a importação.



5 ^a etapa: Criar no banco todas as tabelas dimensão, descritas na Modelagem.

```
USE bd_tcc;
□CREATE TABLE [dbo] [dim_area](
      [idarea] [int] NOT NULL PRIMARY KEY,
      [descricao] [varchar] (50) NOT NULL
  );
CREATE TABLE [dbo].[dim_regiao](
      [idregiao] [int] NOT NULL PRIMARY KEY,
      [descricao] [varchar] (50) NOT NULL
  );
CREATE TABLE [dbo].[dim_classe_econo](
      [idclasse] [int] NOT NULL PRIMARY KEY,
      [descricao] [varchar] (50) NOT NULL
  );
□CREATE TABLE [dbo].[dim_renda_familiar](
      [idrenda] [int] NOT NULL PRIMARY KEY,
      [descricao] [varchar] (50) NOT NULL
  );
□CREATE TABLE [dbo].[dim_respostasfechadas](
      [idresposta] [int] NOT NULL PRIMARY KEY,
      [descricao] [varchar] (50) NOT NULL
  );
☐CREATE TABLE [dbo].[dim_a5a](
      [idmotivoprincipal] [int] NOT NULL PRIMARY KEY,
      [descricao] [varchar] (50) NOT NULL
  );
CREATE TABLE [dbo].[dim_a7](
      [idtipoconexao] [int] NOT NULL PRIMARY KEY,
      [descricao] [varchar] (50) NOT NULL
  );
 ĠCREATE TABLE [dbo].[dim_sexo](
      [idsexo] [int] NOT NULL PRIMARY KEY,
      [descricao] [varchar] (50) NOT NULL
 ĠCREATE TABLE [dbo].[dim_c4](
```

6 a etapa: inserir, nas tabelas dimensão, todos os valores referentes aos campos criados, conforme o dicionário de dados disponibilizado nos arquivos da pesquisa.

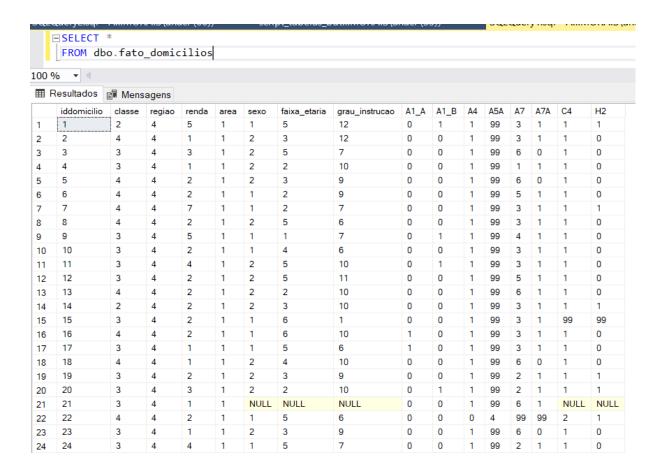
```
insert into dim area values (1,'Urbana')
insert into dim_area values (2,'Rural')
insert into dim regiao values (1,'Sudeste')
insert into dim_regiao values (2,'Nordeste')
insert into dim regiao values (3,'Sul')
insert into dim_regiao values (4,'Norte')
insert into dim_regiao values (5,'Centro-Oeste')
insert into dim classe econo values (1,'A')
insert into dim classe econo values (2,'B')
insert into dim_classe_econo values (3,'C')
insert into dim classe econo values (4,'DE')
insert into dim renda familiar values (1,'Até 1 SM')
insert into dim renda familiar values (2, 'Mais de 1 SM até 2 SM')
insert into dim renda familiar values (3, 'Mais de 2 SM até 3 SM')
insert into dim_renda_familiar values (4, 'Mais de 3 SM até 5 SM')
insert into dim_renda_familiar values (5, 'Mais de 5 SM até 10 SM')
insert into dim_renda_familiar values (6,'Mais de 10 SM')
insert into dim renda familiar values (7,'Não tem renda')
insert into dim_renda_familiar values (97,'Não sabe')
insert into dim_renda_familiar values (98,'Não respondeu')
insert into dim_respostasfechadas values (0,'Não')
insert into dim_respostasfechadas values (1,'Sim')
insert into dim respostasfechadas values (97, 'Não sabe')
insert into dim respostasfechadas values (98, 'Não respondeu')
insert into dim respostasfechadas values (99,'Não se aplica')
insert into dim_a5a values (1, 'Falta de computador')
insert into dim a5a values (2, 'Falta de necessidade')
insert into dim_a5a values (3,'Falta de interesse')
insert into dim a5a values (4, 'Têm acesso em outro lugar')
insert into dim_a5a values (5,'Acham muito caro')
insert into dim_a5a values (6,'Não sabem usar Internet')
insert into dim_a5a values (7,'Falta de disponibilidade')
insert into dim a5a values (8, 'Preocupações com segurança ou privacidade')
insert into dim_a5a values (9,'Evitam o contato com conteúdo perigoso')
insert into dim_a5a values (10,'Outro motivo')
insert into dim a5a values (97,'Não sabe')
```

7 ª etapa: Criar um script para construir a tabela FATO_DOMICILIOS onde será feito o join entre as bases de domicílio e base de indivídios, pelo campo comum iddomicilio e com apenas os outros campos necessários para responder às perguntas que encontram-se no item Objetivo desse Projeto.

Conferiu-se, também, se não houve a perda de nenhum registro quando da união entre as tabelas, já que nem todos os domicílios responderam as perguntas do questionário de indivíduos.

```
SELECT COUNT(*)
 FROM dbo.basedomicilios; -- 23950 REGISTROS
SELECT COUNT(*)
FROM dbo.baseindividuos; -- 21011 REGISTROS
SELECT d.ID_DOMICILIO
                          iddomicilio
, d.CLASSE_2015
 , d.COD_REGIAO_2
                              regiao
, d.RENDA_FAMILIAR_2
                            renda
, d.AREA
                             area
, i.SEXO
                             sexo
, i.FAIXA_ETARIA
                               faixa_etaria
, i.FAIXA_ETARIA faixa_etar:
, i.GRAU_INSTRUCAO grau_instrucao
, d.A1_A
, d.A1_B
, d.A4
, d.A5A
, d.A7
, d.A7A
, i.C4
 , i.H2
 INTO dbo.fato_domicilios
 FROM dbo.basedomicilios d
 LEFT JOIN dbo.baseindividuos i ON i.id_domicilio = d.id_domicilio; -- 23950 REGISTROS - TESTE QTDE REGISTROS OK
```

	EXEC sp_he	lp 'd	bo.fato_	domici	lios	' ;				
100 9	% •									
⊞ F	Resultados 🗐 N	/lensag	ens							
	Name Owner Type Created_datetime									
1										
	Column_name	Type	Computed	Length	Prec	Scale	Nullable	TrimTrailingBlanks	FixedLenNullInSource	Collation
1	iddomicilio	int	no	4	10	0	no	(n/a)	(n/a)	NULL
2	classe	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
3	regiao	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
4	renda	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
5	area	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
6	sexo	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
7	faixa_etaria	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
8	grau_instrucao	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
9	A1_A	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
10	A1_B	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
11	A4	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
12	A5A	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
13	A7	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
14	A7A	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
15	C4	int	no	4	10	0	yes	(n/a)	(n/a)	NULL
16	H2	int	no	4	10	0	yes	(n/a)	(n/a)	NULL



4. Camada de Apresentação (em desenvolvimento)

- 4.1 Dashboard
- 4.2 Análises avançadas
- 5. Registros de Homologação (em desenvolvimento)
- 5. Conclusões (em desenvolvimento)

6. Links

A fim de evitar qualquer inconveniente, os arquivos foram disponibilizados em dois repositórios diferentes, no GitHub e no Google Drive.

https://github.com/anacmoraes/TCC-PUC-Minas

 $\underline{\text{https://drive.google.com/drive/folders/1QMzVbPpCgit_rp8aGsQL1IP3I8WfRdh}}\\ \underline{\text{A?usp=share_link}}$

REFERÊNCIAS

Instituto Brasileiro de Geografia e Estatística (IBGE). Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua) - Acesso à Internet e à Televisão e Posse de Telefone Móvel Celular para Uso Pessoal. 2021. Disponível em: https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/31892-pnad-continua-2021-metade-dos-domicilios-brasileiros-tem-acesso-a-internet-por-computador. Acesso em 12 fev. 2023.

FERREIRA, T. O desafio da inclusão digital no Brasil: acesso, uso e apropriação das tecnologias. Observatório de Comunicação, Mídia e Política. 2019. Disponível em: http://www.observatoriodacomunicacao.org.br/wpcontent/uploads/2019/08/Livro-Observatorio_Cultura_Digital-ebook.pdf. Acesso em 12 fev. 2023.

Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (CETIC). **Pesquisa TIC Domicílios.** Disponível em: https://www.cetic.br/microdados/