



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 Moraes Rosa

Belo Horizonte

2023

SUMÁRIO

- 1. Introdução3**
 - 1.1. Contexto3
 - 1.3. Público alvo5
- 2. Modelos de Dados5**
 - 2.1. Modelo Dimensional5
 - 2.2. Fatos e Dimensões6
- 3. Integração, Tratamento e Carga de Dados11**
 - 3.1. Fontes de Dados.....11
 - 3.2. Processos de Integração e Carga (ETL).....11
- 4. Camada de Apresentação18**
 - 4.1 Dashboard18
 - 4.2 Análises avançadas28
- 5. Registros de Homologação28**
- 6. Conclusões31**
- 7. Links32**
- Referências33**

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.

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?

Para tanto, serão elaborados três modelos de dashboards:

- Operacional: com métricas de controle de curto prazo.
- Tático: com indicadores que já possam permitir a tomada de decisão e alcance dos objetivos em médio prazo.
- Estratégico: apresenta KPIs relevantes para a validação da estratégia.

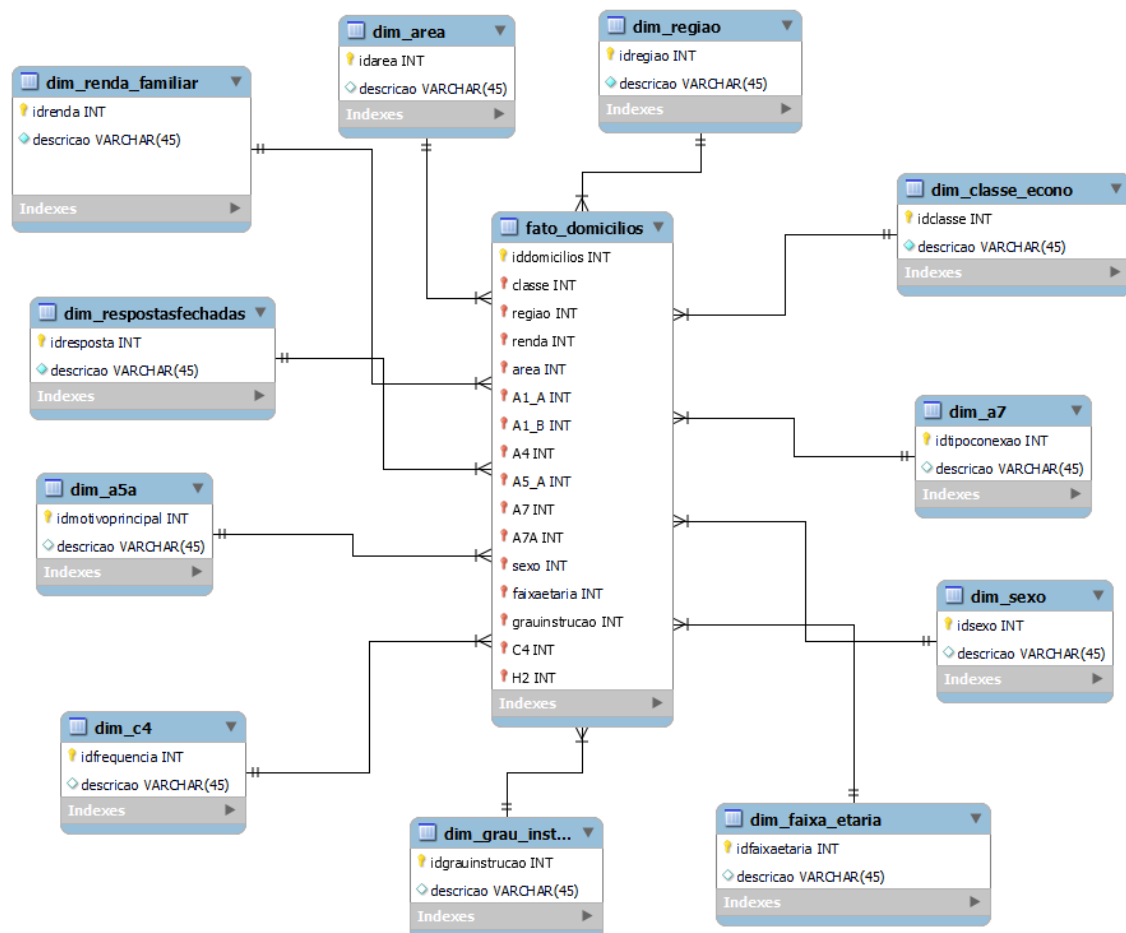
1.3. Público alvo

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

idarea	descricao
1	Urbana
2	Rural

- DIM_REGIAO: tabela que descreve as 5 regiões do Brasil

Campos:

idregiao(PK): tipo numérico

descricao: tipo texto

idregiao	descricao
1	Sudeste
2	Nordeste
3	Sul
4	Norte
5	Centro-Oeste

- 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	A
2	B
3	C
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

Essa tabela dimensão possui as repostas para as perguntas fechadas e tem referência os campos A1_A, A1_B, A4, A7A e H2 da Tabela Fato, cujas perguntas são:

A1_A: Neste domicílio tem computador de mesa?

A1_B: Neste domicílio tem notebook?

A4: Este domicílio tem acesso à Internet?

A7A: Neste domicílio tem Wi-Fi?

H2: Nos últimos 12 meses, o respondente comprou ou encomendou produtos ou serviços pela Internet, mesmo que o pagamento não tenha sido feito pela Internet?

- 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

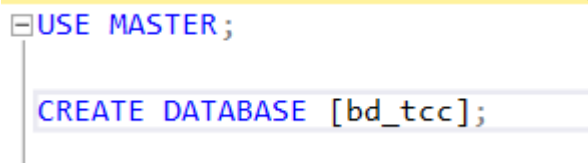
As bases de dados utilizadas foram retiradas do site CETIC.br (Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação), com a pesquisa TIC Domicílios de 2021, sendo elas:

- 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.

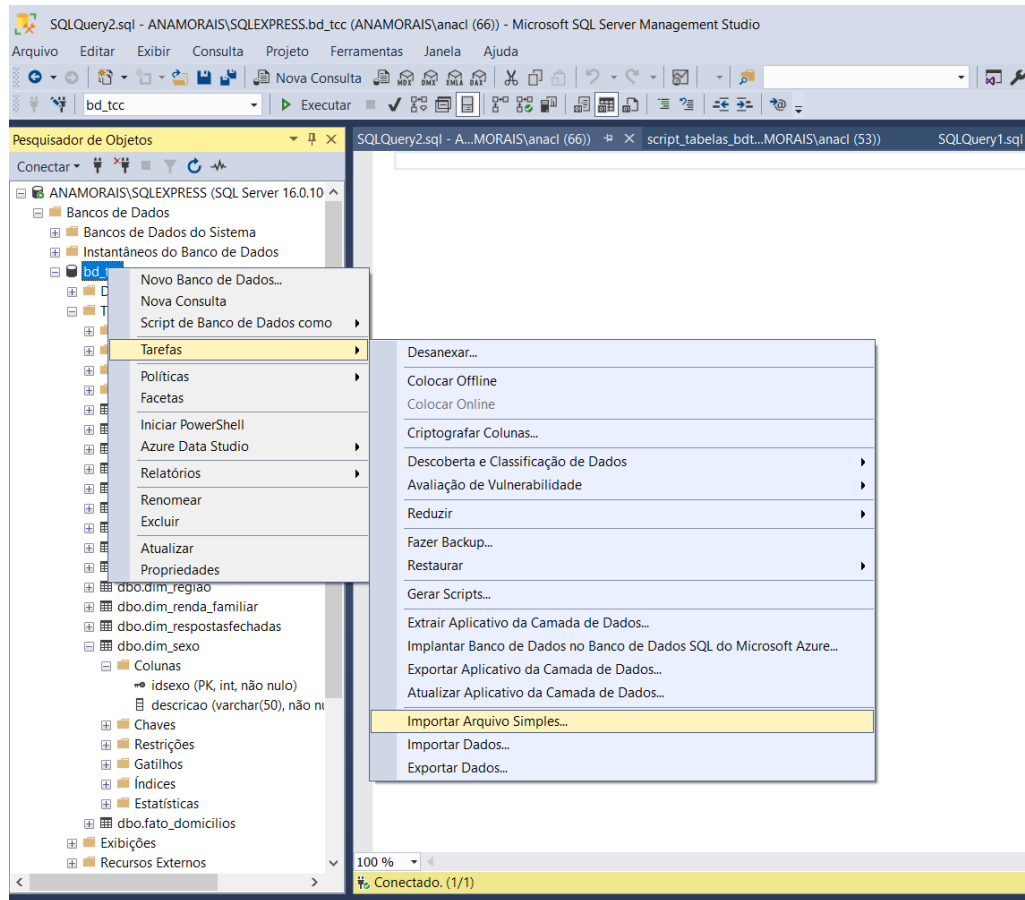


```
USE MASTER;  
  
CREATE DATABASE [bd_tcc];
```

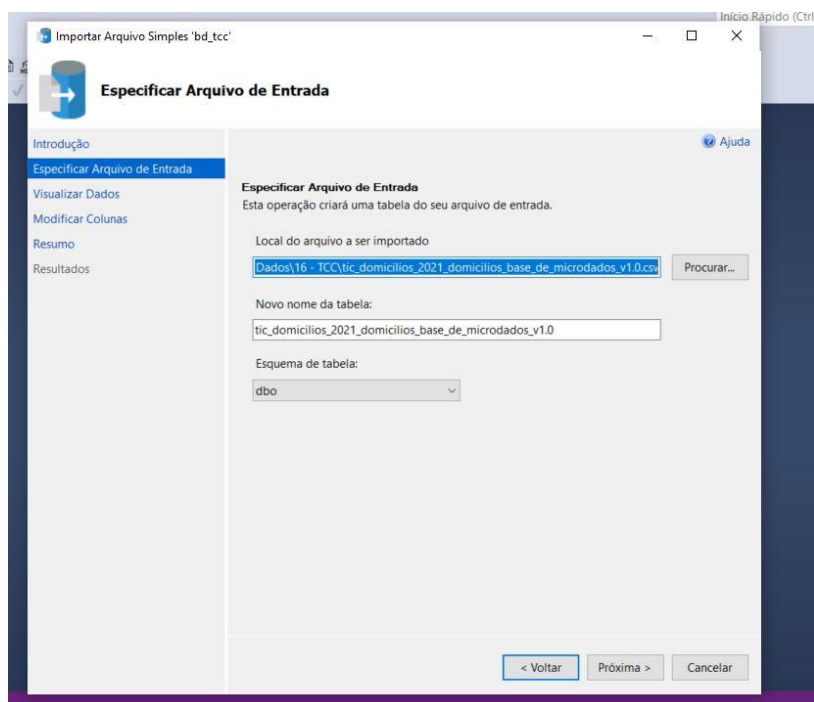
O segundo passo foi realizar o processo de ingestão no banco, dos arquivos .csv que contém os microdados dos questionários.

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:

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



2ª 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.



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

Importar Arquivo Simples 'bd_tcc'

Modificar Colunas

Introdução | Especificar Arquivo de Entrada | Visualizar Dados | **Modificar Colunas** | Resumo | Resultados

Modificar Colunas

Esta operação gerou o seguinte esquema de tabela. Verifique se o esquema é preciso e, se não for, faça as alterações.

Nome da Coluna	Tipo de Dados	Chave Primária	<input type="checkbox"/> Permitir Nulos
QUEST	int	<input type="checkbox"/>	<input type="checkbox"/>
ID_DOMICILIO	smallint	<input type="checkbox"/>	<input type="checkbox"/>
CLASSE_2015	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
AREA	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
TV	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
RADIO	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
BANHEIRO	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
AUTOMOVEI	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
EMPREGADA	bit	<input type="checkbox"/>	<input type="checkbox"/>
LAVA_ROUPA	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
GELADEIRA	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
FREEZER	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
LAVA_LOUCA	bit	<input type="checkbox"/>	<input type="checkbox"/>
DVD	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
MICRO_ONDAS	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
MOTOCICLETA	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
SECADORA	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
TELEFONE_FIXO	bit	<input type="checkbox"/>	<input type="checkbox"/>
VIDEOGAME	bit	<input type="checkbox"/>	<input type="checkbox"/>
TV_ASSINATURA	bit	<input type="checkbox"/>	<input type="checkbox"/>
ANTENA_PARABOLICA	bit	<input type="checkbox"/>	<input type="checkbox"/>
AGUA	tinyint	<input type="checkbox"/>	<input type="checkbox"/>

Granularidade de linha do relatório de erros (impacto no desempenho com intervalos menores) Sem Intervalo

< Voltar **Próxima >** Cancelar

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

Importar Arquivo Simples 'bd_tcc'

Resultados

Introdução | Especificar Arquivo de Entrada | Visualizar Dados | Modificar Colunas | Resumo | **Resultados**

Operação Concluída

Resumo:

Nome	Resultado
✓ Inserir Dados	Éxito

< Voltar Próxima > **Fechar**

5ª 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
);

CREATE TABLE [dbo].[dim_sexo](
    [idsexo] [int] NOT NULL PRIMARY KEY,
    [descricao] [varchar] (50) NOT NULL
);

CREATE TABLE [dbo].[dim_c4](
```

6ª 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')

```

7ª etapa: Criar um script para construir a tabela FATO_DOMICILIOS onde será feito o join entre as bases de domicílio e base de indivíduos, 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.baseindivíduos; -- 21011 REGISTROS

SELECT d.ID_DOMICILIO      iddomicilio
, d.CLASSE_2015            classe
, d.COD_REGIAO_2           regiao
, d.RENDA_FAMILIAR_2       renda
, d.AREA                  area
, i.SEXO                  sexo
, i.FAIXA_ETARIA           faixa_etaria
, 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.baseindivíduos i ON i.id_domicilio = d.id_domicilio; -- 23950 REGISTROS - TESTE QTDE REGISTROS OK

```

```
EXEC sp_help 'dbo.fato_domicilios';
```

100 %

Resultados Mensagens

	Name	Owner	Type	Created_datetime						
1	fato_domicilios	dbo	user table	2023-03-12 15:45:15.093						

	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

SELECT *
FROM dbo.fato_domicilios

100 %

Resultados Mensagens

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

8ª etapa: Realizar a ingestão dos dados no Power BI, software escolhido para a visualização dos dados.

Navegador

Opções de Exibição

10W10-2432\SQLEXPRESS [1]

bd_tcc [14]

- ☐ basedomicilios
- ☐ baseindividuos
- ☐ dim_a5a
- ☐ dim_a7
- ☐ dim_area
- ☐ dim_c4
- ☐ dim_classe_econo
- ☐ dim_faixa_etaria
- ☐ dim_grau_instrucao
- ☐ dim_regiao
- ☐ dim_renda_familiar
- ☐ dim_respostasfechadas
- ☐ dim_sexo
- ☐ fato_domicilios

Nenhum item selecionado para visualização

Selecionar Tabelas Relacionadas

Carregar

Transformar Dados

Cancelar

4. Camada de Apresentação

A camada de apresentação foi realizada no Power BI Desktop e publicada na aplicação da ferramenta e dividida em três painéis:

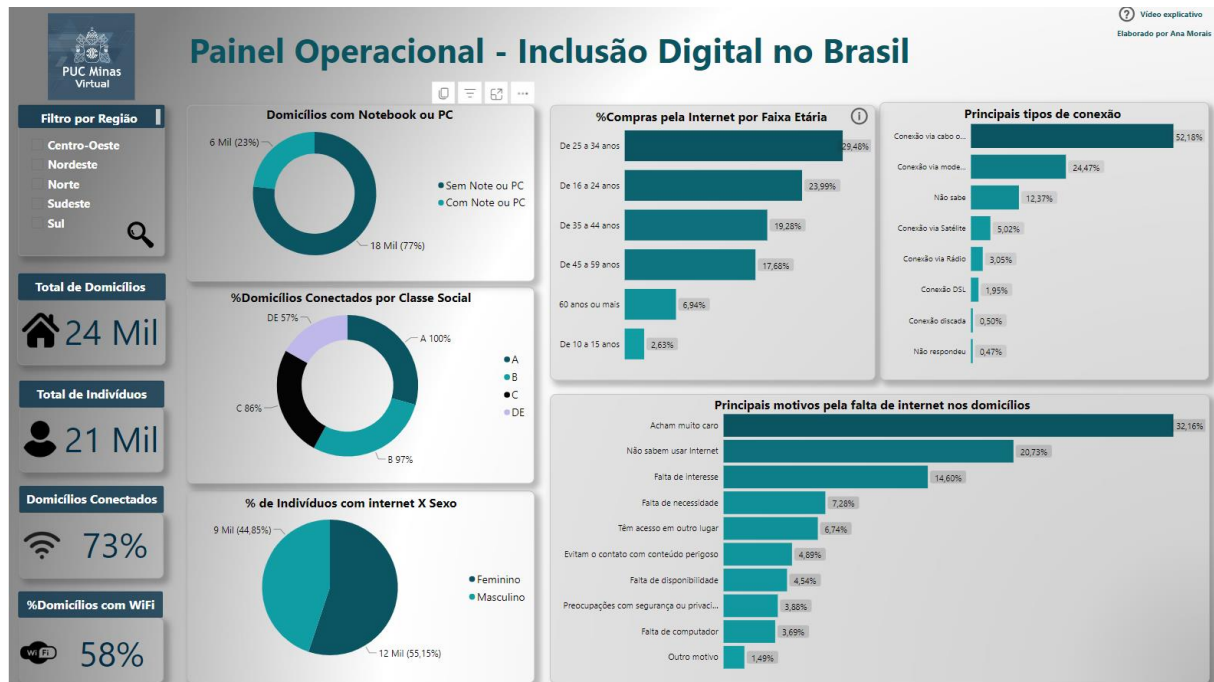
- **Operacional:** com informações mais diretas e que pode ter como público-alvo a sociedade e mídia, a fim de que se atentem a fatos relevantes quanto à Inclusão Digital no Brasil.
- **Tático:** com informações mais aprofundadas, pensando em um público que possa tomar decisões em curto prazo ou definir algumas ações e metas frente ao exposto, como as Universidades, organizações e governo.
- **Estratégico:** painel com indicadores e metas que podem contribuir para atender às demandas sociais de combate à desigualdade digital e atingir os Objetivos propostos pela ONU na Agenda 2030 para o Desenvolvimento Sustentável (ODS), relacionados ao tema. Dentre esses objetivos, destacam-se principalmente:
 - **Objetivo 10:** Reduzir a desigualdade dentro dos países e entre eles.
 - **Objetivo 4:** Assegurar a educação inclusiva e equitativa e de qualidade, e promover oportunidades de aprendizagem ao longo da vida para todos;
 - **Objetivo 11:** Tornar as cidades e os assentamentos humanos inclusivos, seguros, resilientes e sustentáveis;

4.1 Dashboard

Os painéis podem ser acessados pelo link a seguir:
<https://app.powerbi.com/reportEmbed?reportId=4c35979f-2dfd-40e3-8b7c-82d05a5bb650&autoAuth=true&ctid=14cbd5a7-ec94-46ba-b314-cc0fc972a161>

4.1.1 Painel operacional

O painel operacional contempla 1 filtro, 4 cards informativos e 6 gráficos, entre rosca, pizza e de coluna.



O filtro imediato está por região do Brasil, mas é possível filtrar os gráficos entre si, utilizando um ao outro, conforme será demonstrado a seguir.

Filtro por Região

☐ Centro-Oeste

☐ Nordeste

☐ Norte

☐ Sudeste

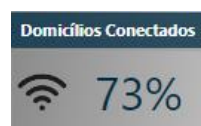
☐ Sul

Os cards criados trazem as informações da amostragem da pesquisa:

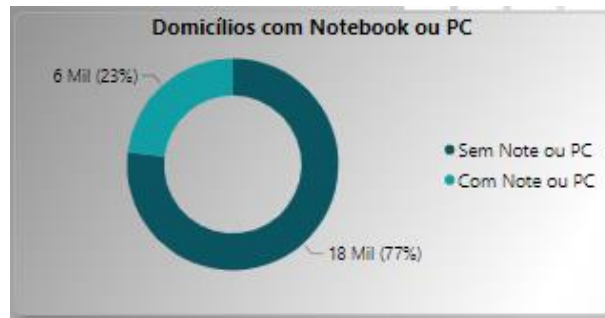
- 24 mil domicílios respondentes;
- 21 mil indivíduos respondentes;



E dos percentuais de domicílios que possuem internet (73%) e que possuem Wi-Fi (58%).



O primeiro gráfico de rosca traz a informação da quantidade e percentual de domicílios que possuem Notebook ou Computador de Mesa (perguntas A1_A e A1_B). O resultado foi de 18 mil domicílios sem notebook ou computador de mesa (77% do total) e 6 mil domicílios com (23% do total).



O próximo gráfico de rosca traz o percentual de conectados por cada Classe Social, evidenciando que a desigualdade social influencia totalmente no acesso à internet.



Podemos verificar que 100% dos domicílios que se declararam pertencentes à classe A possuem acesso à internet, enquanto dos domicílios da classe D e E, apenas 57%. Quando se trata da classe B, 97% dos domicílios estão conectados e para a classe C, 86% deles.

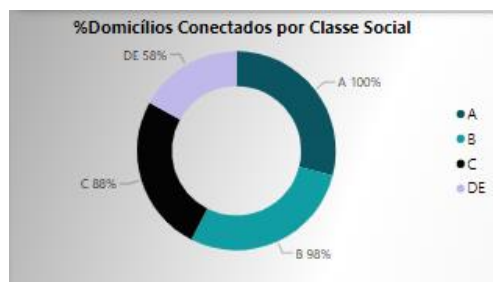
O painel permite verificar também a variação desses percentuais, de Domicílios com Notebook ou PC e de % de Conectados por Classe Social, conforme a região brasileira, utilizando o filtro disponibilizado no dashboard.

Pode-se observar que, em todas as regiões, 100% dos domicílios que se declararam classe A possuem internet. Assim, temos:

- Para a região Centro-Oeste:



- Nordeste:



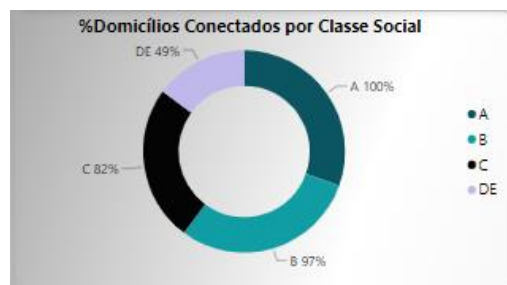
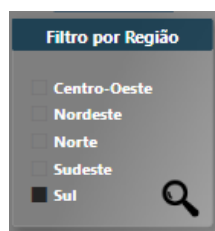
- Norte:



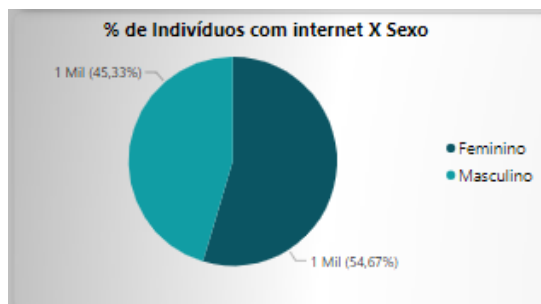
- Sudeste:



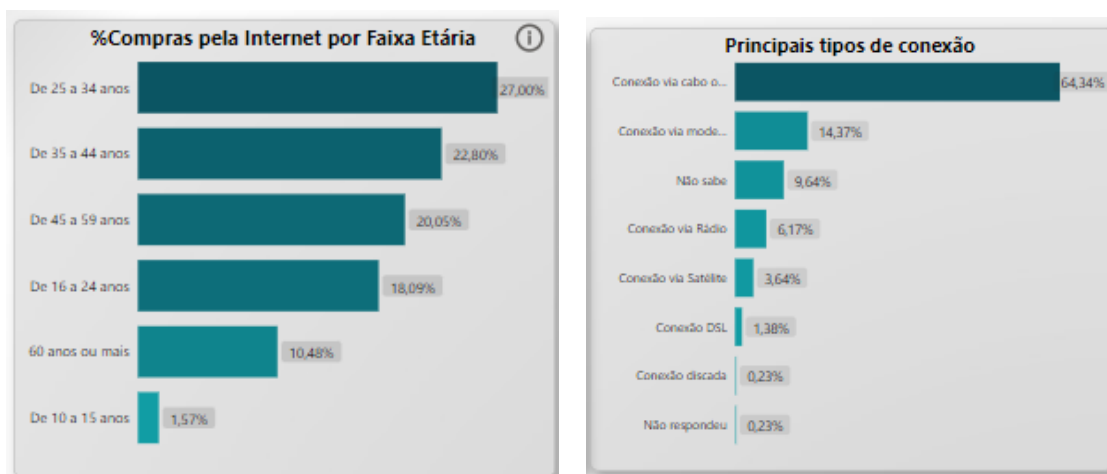
- Sul



O gráfico de pizza traz o percentual de indivíduos que possuem internet conforme o sexo. A divisão é quase igualitária, havendo uma pequena diferença de 5% a mais para o sexo feminino.



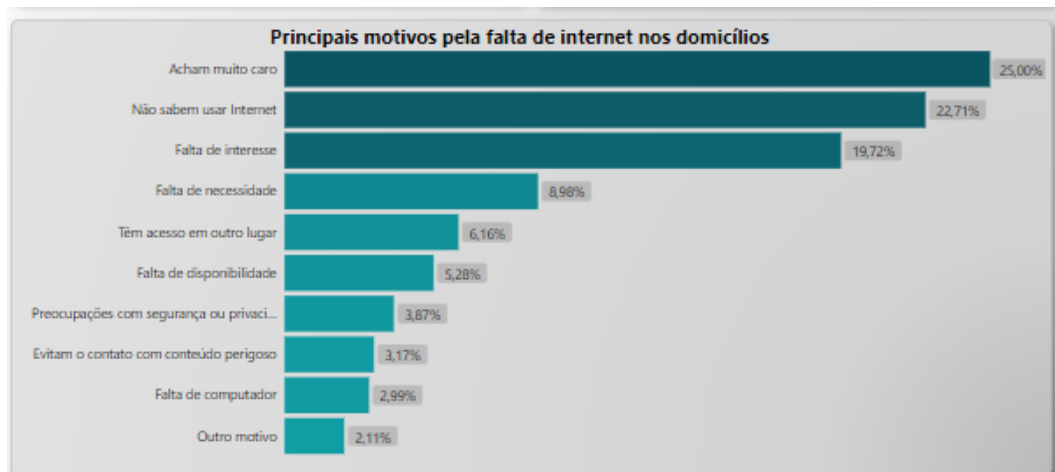
Os próximos 3 gráficos de colunas trazem as informações, respectivamente, sobre o % de compras pela internet conforme a faixa etária, os principais tipos de conexão utilizados e os principais motivos alegados pelos respondentes, pela falta de internet nos domicílios.



Dos indivíduos que compram pela internet, a faixa etária de 25 a 34 anos é a que possui o maior percentual com 27%, seguido de 35 a 44 anos, com 22,8%.

Já o principal tipo de conexão dos domicílios que possuem internet é a conexão via cabo ou fibra ótica, com quase 65% das respostas.

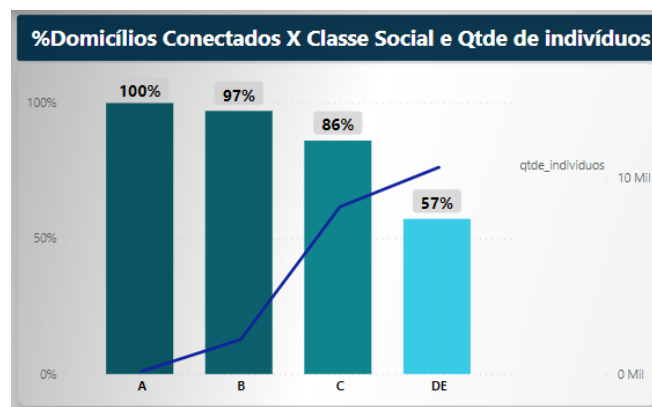
Por fim, no último gráfico disponibilizado nesse painel operacional, temos como fatores preponderantes pela falta de internet nos domicílios, o preço, seguido da falta de conhecimento em se utilizar a internet.



4.1.2 Painel Tático

O painel táctico possui gráficos que possibilitam análises mais aprofundadas, por não trazer apenas números absolutos e análises pontuais. Para isso, foram realizados alguns cruzamentos de dados a fim de entender como as variáveis se relacionam. Por exemplo, há um gráfico de linha e coluna que relaciona o percentual de domicílios que possuem internet versus a classe social, porém traz também no gráfico de linha, a quantidade de indivíduos dentro de cada classe. Resultando numa análise que possibilita verificar que quanto mais é aumentada a classe social, maior é o acesso à internet e menor é a quantidade de pessoas que possuem esse privilégio, confirmando o que já se imaginava sobre o fato de a desigualdade social estar intimamente ligada à exclusão digital.

Abaixo a imagem desse gráfico mencionado:

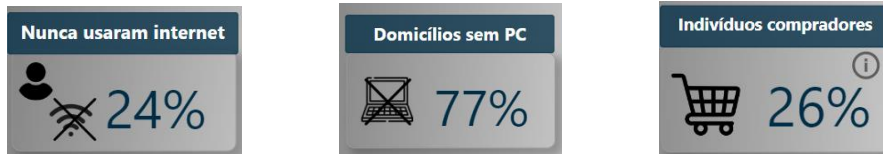


Neste painel, foi mantido o filtro por Região, mas também inserido o filtro por Classe Social.

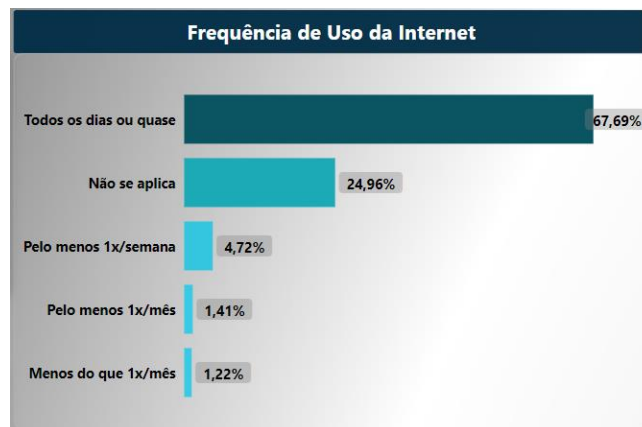
Filtro por Classe

- ☐ A
- ☐ B
- ☐ C
- ☐ DE

Foram criados três cards com as seguintes informações: percentual de indivíduos que nunca usaram a internet, percentual de domicílios sem computador ou notebook e percentual de indivíduos que realizaram compra pela internet nos últimos 3 meses.

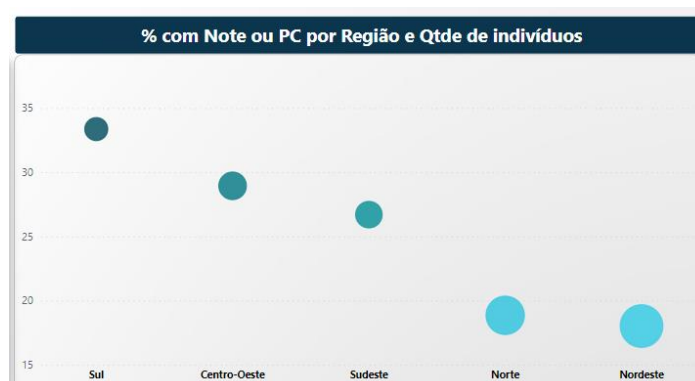


Em um dos gráficos, optou-se por trazer a informação de frequência de uso da internet, que pode ser analisado pelo cenário geral ou utilizando os filtros de Região e/ou Classe Social.



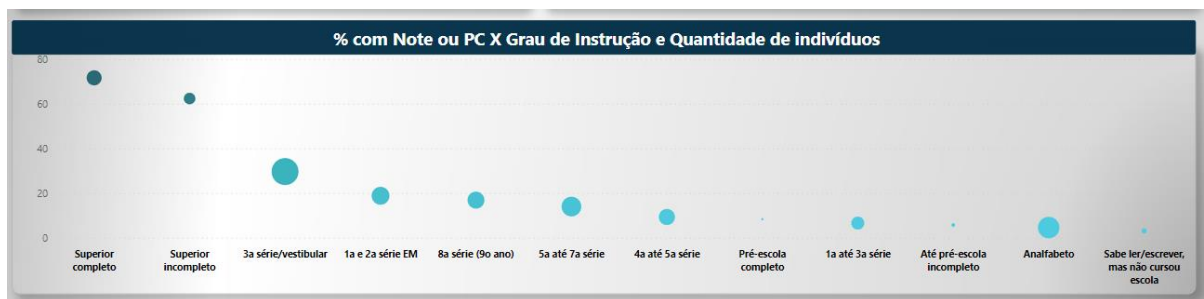
Percebe-se que, daqueles indivíduos que utilizam a internet, quase 68% deles têm uma frequência diária de uso, porém, há um percentual considerável de quase 25% a quem não se aplica essa pergunta pois, muito provavelmente, sequer já utilizaram a internet alguma vez na vida (pelo menos não diretamente).

Um dos gráficos de bolha trazidos no painel táctico, apresenta a informação de percentual de domicílios que possui notebook ou computador de mesa (PC) por Região do Brasil, sendo o tamanho da bolha a representação da quantidade de indivíduos envolvidos em cada cenário.



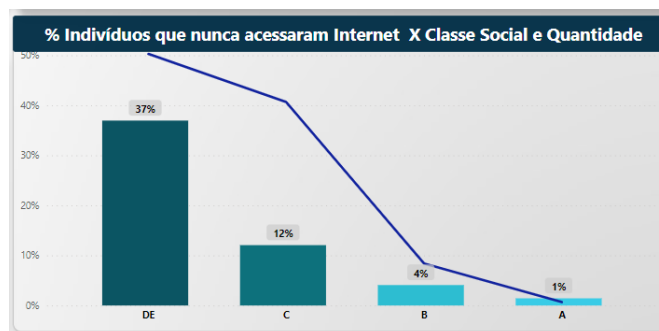
Assim, analisando o gráfico, percebe-se que há um maior percentual de domicílios com notebook ou computador de mesa na região Sul (33%), porém com uma menor quantidade de indivíduos, sendo que, no Nordeste ocorre o inverso, temos o mais baixo percentual de computadores de todas as regiões com uma maior quantidade de pessoas nessa situação.

Ainda em relação ao percentual de notebook ou PC, o painel exibe um novo gráfico de bolhas que compara essa variável com o grau de instrução dos indivíduos respondentes, mantendo o tamanho da bolha para representar a quantidade de indivíduos.



Os maiores percentuais de notebook e computadores de mesa pertencem aos indivíduos que possuem ensino superior completo ou incompleto. A maior quantidade de indivíduos está naqueles que finalizaram a 3ª série do ensino médio, estes com o terceiro maior índice de detentores de notebook ou PC. Há uma quantidade considerável de analfabetos com baixíssimo acesso à computadores em suas residências.

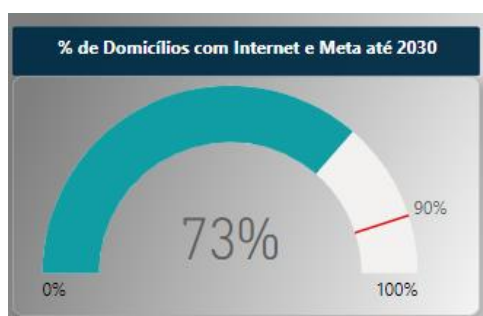
O último gráfico deste painel possui o formato de colunas e linha e demonstra o percentual de indivíduos que nunca utilizaram a internet em relação à classe social, sendo que a linha representa a quantidade de indivíduos. Observa-se que há uma maior quantidade de indivíduos nas classes D e E, as quais apresentam o maior percentual de pessoas que nunca acessaram a internet. À medida que a classe social vai aumentando, a quantidade de indivíduos diminui, até tender a zero na Classe A, onde praticamente todas as pessoas tiveram a oportunidade de acessar a internet.



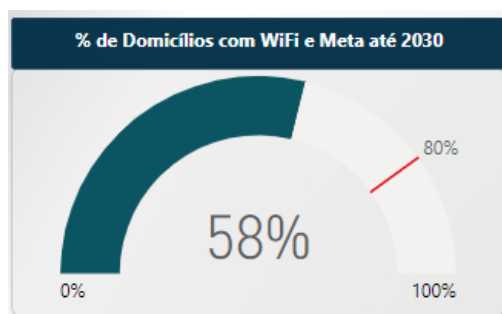
4.1.3 Painel Estratégico

O painel estratégico traz uma visão geral das análises realizadas ao longo do relatório técnico, sendo estabelecidos KPI's e sugeridas metas com o objetivo de auxiliar no alcance de melhores índices quanto à Inclusão Digital no Brasil.

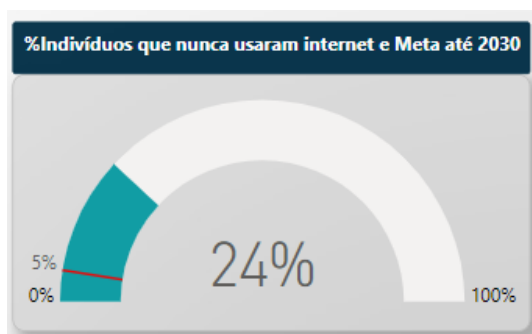
Com relação ao percentual de domicílios com internet e sua respectiva meta até 2030, temos o gráfico a seguir mostrando que há 73% de domicílios com Internet hoje e que se pretende atingir o percentual de 90%.



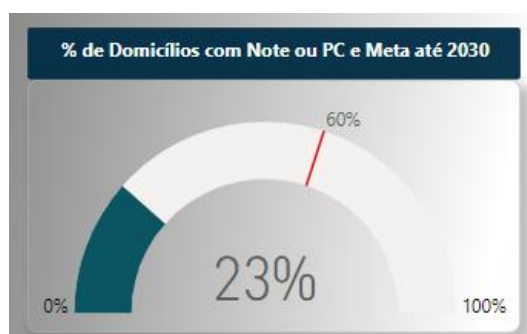
Quanto ao percentual de domicílios que possuem WiFi, a pesquisa trouxe que apenas 58% deles respondeu positivamente. A meta sugerida para este KPI é de atingir 80% até 2030.



Olhando para os indivíduos que nunca usaram a internet, esse KPI tem sentido para baixo, ou seja, quanto menor for o percentual, melhor. Atualmente, tem-se o resultado de 24% e pretende-se atingir, até 2030, apenas 5% de indivíduos que nunca usaram a internet.



A pesquisa demonstrou que há um percentual muito baixo de domicílios que possuem notebook ou computador de mesa, com apenas 23%. E a meta sugerida para este KPI é de 60%.



Neste dashboard estratégico, optou-se por trazer o cenário de domicílios com internet em relação à classe social, para que possam ser estabelecidas metas específicas para as classes mais carentes, C e DE.



A classe DE possui um percentual de 57% dos domicílios com internet, assim, pretende-se chegar à 75% até 2030. Já a classe C, possui 86% dos domicílios com internet e até 2030, pretende-se que essa classe social se iguale ao que é a classe A, atualmente, com 100% dos domicílios conectados.

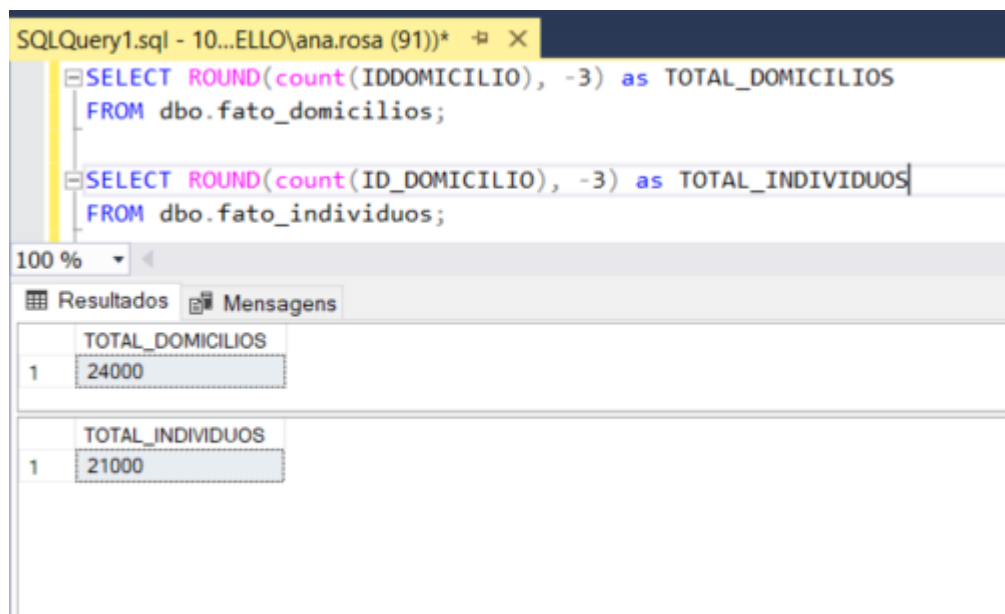
4.2 Análises avançadas

Este relatório não contemplará análises avançadas, porém, é possível realizar em um projeto futuro, análise preditiva a partir da inserção de outros campos da base original, que trazem informações específicas dos indivíduos respondentes como, por exemplo, compra pela internet, uso de rede social, uso de streaming de filmes, músicas e/ou jogos, possibilitando realizar uma clusterização e traçar os perfis dos usuários.

5. Registros de Homologação

A homologação dos resultados apresentados nos painéis foi realizada no SQL Server, utilizando T-SQL.

Para os cards de total de domicílios e total de indivíduos, foram realizadas as queries abaixo, sendo comprovados os valores de 24mil e 21mil, respectivamente:



```
SQLQuery1.sql - 10...ELLO\ana.rosa (91))* -# X
SELECT ROUND(count(IDDOMICILIO), -3) as TOTAL_DOMICILIOS
FROM dbo.fato_domicilios;

SELECT ROUND(count(ID_DOMICILIO), -3) as TOTAL_INDIVDUOS
FROM dbo.fato_individuos;
```

TOTAL_DOMICILIOS	
1	24000

TOTAL_INDIVDUOS	
1	21000

Para o card de percentual de conectados, foi realizada a query abaixo a fim de homologar o resultado de 73%.

```
-- HOMOLOGAÇÃO CARD % DE CONECTADOS

select
c.quantidade as Qtde,
t.total as Total,
CEILING(c.quantidade / (t.Total +0.) * 100) as PERC_CONECTADOS
from
(
SELECT
count(iddomicilio) as total
FROM dbo.fato_domicilios
) t,
(
SELECT
count(iddomicilio) as quantidade
FROM dbo.fato_domicilios
WHERE A4 = 1
) c
```

1 %

Resultados Mensagens

Qtde	Total	PERC_CONECTADOS
17404	23950	73

Para o card de percentual de domicílios com WiFi, foi realizada a query abaixo a fim de homologar o resultado de 58%.

```
SQLQuery1.sql - 10...ELLO\ana.rosa (73))* X
-- HOMOLOGAÇÃO CARD % com WiFi (58%)

select
c.quantidade as Qtde,
t.total as Total,
CEILING(c.quantidade / (t.Total +0.) * 100) as PERC_WIFI
from
(
SELECT
count(iddomicilio) as total
FROM dbo.fato_domicilios
) t,
(
SELECT
count(iddomicilio) as quantidade
FROM dbo.fato_domicilios
WHERE A7A = 1
) c
```

100 %

Resultados Mensagens

Qtde	Total	PERC_WIFI
13801	23950	58

Para o card de percentual de indivíduos que nunca usaram Internet, foi realizada a query abaixo a fim de homologar o resultado de 24%.

```

SQLQuery1.sql - 10...ELLO\ana.rosa (73))*
-- HOMOLOGAÇÃO CARD % Nunca usaram Internet (24%)
select
c.quantidade as Qtde,
t.total as Total,
CEILING(c.quantidade / (t.Total +0.) * 100) as PERC_WIFI
from
(
SELECT
count(id_domicilio) as total
FROM dbo.fato_individuos
) t,
(
SELECT
count(id_domicilio) as quantidade
FROM dbo.fato_individuos
WHERE C1 = 0
) c

```

100 %

Resultados Mensagens

	Qtde	Total	PERC_WIFI
1	5008	21011	24

Para o card de percentual de domicílios sem PC, foi realizada a query abaixo a fim de homologar o resultado de 77%.

```

SQLQuery1.sql - 10...ELLO\ana.rosa (73))*
-- HOMOLOGAÇÃO CARD % Domicílios sem PC (77%)
SELECT
c.quantidade as Qtde,
t.total as Total,
CEILING(c.quantidade / (t.Total +0.) * 100) as PERC_WIFI
FROM
(
SELECT
count(iddomicilio) as total
FROM dbo.fato_domicilios
) t,
(
SELECT
count(iddomicilio) as quantidade
FROM dbo.fato_domicilios
WHERE A1_A = 0 AND A1_B = 0
) c

```

100 %

Resultados Mensagens

	Qtde	Total	PERC_WIFI
1	18377	23950	77

Para o card de percentual de indivíduos compradores, foi realizada a query abaixo a fim de homologar o resultado de 26%.

```

SQLQuery1.sql - 10...ELLO\ana.rosa (73))*
t.total as Total,
FLOOR(c.quantidade / (t.Total +0.) * 100) as PERC_WIFI
FROM
(
SELECT
count(id_domicilio) as total
FROM dbo.fato_individuos
) t,
(
SELECT
count(id_domicilio) as quantidade
FROM dbo.fato_individuos
WHERE H2 = 1
) c

```

100 %

Resultados Mensagens

	Qtde	Total	PERC_WIFI
1	5549	21011	26

6. Conclusões

A base de dados estudada permitiu confirmar alguns pontos que já se imaginava com relação à inclusão digital no Brasil, como o fato de que as classes C e DE estão muito longe de atingirem um patamar aceitável quanto ao acesso à internet. Em todas as regiões do Brasil, a classe A tem 100% dos domicílios conectados, e o percentual de indivíduos que nunca usaram a internet só aumenta quando a diminuição da classe social, chegando a 37% na classe DE.

Quanto ao motivo predominante para os domicílios que não possuem acesso à internet, o valor cobrado por ela obteve 32% das respostas. Ou seja, precisam-se construir projetos de política pública para aumentar o acesso à internet das classes mais carentes. O percentual geral de 73% de domicílios conectados é muito baixo tendo em vista que a pesquisa foi realizada em 2021, quando já estávamos há pelo menos 1 ano na pandemia, momento em que a internet passou a ser mais preponderante ainda para o próprio exercício da cidadania.

O segundo motivo informado por não haver internet nos domicílios, com 20% das respostas, é o desconhecimento quanto ao uso. Com isso, entende-se que um

projeto de aumento do acesso à internet precisa estar alinhado a um projeto de educação inclusiva.

Ainda com relação à educação, verificou-se pela análise que o grau de instrução do indivíduo é determinante para que ele possua ou não notebook ou computador de mesa, os maiores índices estão com aqueles que tem ao menos Superior incompleto.

A frequência de uso da internet, para aqueles que informaram utilizá-la é de praticamente todos os dias, com quase 68%, demonstrando o quanto a internet é fundamental para o dia a dia das pessoas, porém, o segundo maior percentual de frequência, com quase 25% é daqueles a quem essa pergunta não se aplica, ou seja, aqueles que não tem acesso algum a internet.

Esse relatório preocupou-se em demonstrar quanto o abismo social impacta no uso da internet, porém a base de dados utilizada é muito rica, podendo ser aprofundado o perfil do usuário de internet no Brasil e entendermos outros pontos ligados ao assunto.

Com essa análise, cabe ao poder público e agentes interessados criarem programas sociais e de educação a fim de diminuir a desigualdade digital no Brasil e auxiliar no atendimento aos objetivos propostos pela ONU na Agenda 2030 no tocante ao assunto, sobretudo o objetivo 10 que pretende reduzir a desigualdade dentro dos países e o objetivo 4, que trata de assegurar a educação inclusiva e equitativa e de qualidade, e promover oportunidades de aprendizagem ao longo da vida para todos.

E nós, como sociedade civil devemos também entender qual é o nosso papel nessa jornada e estabelecer, sobretudo na comunidade onde vivemos e trabalhamos, ações que possam auxiliar no combate à desigualdade social e digital. É fundamental que trabalhemos juntos na busca por um país mais justo e inclusivo.

7. Links

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

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

https://drive.google.com/drive/folders/1QMzVbPpCgit_rp8aGsQL1IP3l8WfRdhA?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/wp-content/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://cetic.br/pt/arquivos/domicilios/2021/domicilios/>

Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (CETIC). **Pesquisa TIC Indivíduos**. Disponível em: <https://cetic.br/pt/arquivos/domicilios/2021/individuos/>

SITE FREEPIK. Ícones gratuitos dos dashboards. Disponível em:

<https://www.flaticon.com/br/icones-gratis/sinal-wi-fi>

<https://www.flaticon.com/br/icones-gratis/do-utilizador>

<https://www.flaticon.com/br/icones-gratis/casa>