

UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
INSTITUTO DE CIÊNCIAS EXATAS  
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**Uso de *Data Warehouse* para Análise da  
Eficiência Energética Residencial:  
Interpretação e Armazenamento de Dados  
para Tomada de Decisões**

**Lásaro de Almeida Deodoro**

JUIZ DE FORA  
SETEMBRO, 2024

# Uso de *Data Warehouse* para Análise da Eficiência Energética Residencial: Interpretação e Armazenamento de Dados para Tomada de Decisões

LÁSARO DE ALMEIDA DEODORO

Universidade Federal de Juiz de Fora

Instituto de Ciências Exatas

Departamento de Ciência da Computação

Bacharelado em Ciência da Computação

Orientador: Victor Ströele de Andrade Menezes

JUIZ DE FORA

SETEMBRO, 2024

USO DE *Data Warehouse* PARA ANÁLISE DA EFICIÊNCIA  
ENERGÉTICA RESIDENCIAL: INTERPRETAÇÃO E  
ARMAZENAMENTO DE DADOS PARA TOMADA DE  
DECISÕES

Lásaro de Almeida Deodoro

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS  
EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTE-  
GRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE  
BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Victor Ströele de Andrade Menezes  
Doutor em Engenharia de Sistemas e Computação (UFRJ)

Luciana Conceição Dias Campos  
Doutora em Engenharia Elétrica (PUC-Rio)

Ronney Moreira de Castro  
Doutor em Informática (UNIRIO)

JUIZ DE FORA  
26 DE SETEMBRO, 2024

*Aos meus amigos e irmãos.*

*Aos pais, pelo apoio e sustento.*

## Resumo

A busca por eficiência no sistema energético brasileiro e mundial tem crescido ano após ano. O setor residencial é um importante consumidor de energia elétrica e pode ser analisado para se entender o nível de conscientização e hábitos de consumo. Com base nessas informações, consumidores podem entender maneiras de economizar energia e empresas fornecedoras podem desenvolver estratégias para conscientização de seus consumidores e para a melhoria dos serviços da empresa. Uma maneira eficiente de disponibilização desses dados de hábitos de consumo é um *Data Warehouse* que vai permitir que os dados de diversas origens e formatos sejam armazenados e modelados de forma que a busca por informações seja otimizada. Com isso, um *Data Warehouse* é apresentado como uma solução importante para o processo de tomada de decisões, permitindo análises e gerações de relatórios de maneira eficiente. Assim, este trabalho busca utilizar dados de hábitos de consumo de energia elétrica para construção de um *Data Warehouse* para análise da eficiência energética residencial.

**Palavras-chave:** *Data Warehouse*, ETL, Consumo de energia, Apoio à decisão.

## Abstract

The search for efficiency in the Brazilian and global energy system has been growing year after year. The residential sector is an important consumer of electricity and can be analyzed to understand the level of awareness and consumption habits. Based on this information, consumers can understand ways to save energy, and supplying companies can develop strategies to raise awareness among their consumers and improve company services. An efficient way to make these consumption habit data available is through a Data Warehouse, which will allow data from various sources and formats to be stored and modeled so that information retrieval is optimized. Therefore, a Data Warehouse is presented as an important solution for the decision-making process, enabling analyses and report generation efficiently. Thus, this work aims to use data on electricity consumption habits to build a Data Warehouse for the analysis of residential energy efficiency.

**Keywords:** *Data Warehouse*, ETL, Energy consumption, Decision support.

## Agradecimentos

A todos os meus parentes, principalmente meus pais e minha irmã, pelo encorajamento, apoio e sustento.

A minha namorada pela companhia e apoio. Esteve sempre lá nos momentos de dificuldade durante a realização do trabalho.

Aos amigos que fiz durante esses anos, principalmente àqueles que já considero como irmãos. Sempre estiveram lá para ajudar a superar os desafios e tornar a caminhada mais fácil.

Ao professor Victor Stroele, pela paciência e apoio neste desafio. E os professores do Departamento de Ciência da Computação pelos seus ensinamentos e aos funcionários do curso, que durante esses anos, contribuíram de algum modo para o nosso enriquecimento pessoal e profissional.

*“Faça o teu melhor na condição que você tem, enquanto você não tem condições melhores para fazer melhor ainda!”*

*Mario Sergio Cortella*

# Conteúdo

<b>Lista de Figuras</b>	<b>7</b>
<b>Lista de Tabelas</b>	<b>8</b>
<b>Lista de Abreviações</b>	<b>9</b>
<b>1 Introdução</b>	<b>10</b>
1.1 Contextualização . . . . .	12
1.2 Descrição do Problema . . . . .	13
1.3 Justificativa . . . . .	14
1.4 Objetivos . . . . .	14
1.5 Metodologia . . . . .	15
1.6 Organização do Trabalho . . . . .	15
<b>2 Fundamentação Teórica</b>	<b>17</b>
2.1 Sistemas de apoio à decisão . . . . .	17
2.2 Extração, transformação e carga . . . . .	18
2.3 Data Warehouse . . . . .	19
2.4 Considerações finais . . . . .	21
<b>3 Trabalhos Relacionados</b>	<b>22</b>
3.1 The Application of Data Warehouse in Decision Support System . . . . .	22
3.2 Energy Provider Data Warehouse Design and Implementation - Case Study	23
3.3 Data Warehouse and Decision Support on Integrated Crop Big Data . . . .	24
3.4 COVID-WAREHOUSE: A Data Warehouse of Italian COVID-19, Pollution, and Climate Data . . . . .	26
3.5 Estimativa da evolução do uso final de energia elétrica no setor residencial do Brasil por região geográfica . . . . .	27
3.6 Uso de Data Warehouse para Análise do Consumo de Energia no Brasil . .	28
3.7 Conclusão . . . . .	29
<b>4 Materiais e Métodos</b>	<b>31</b>
4.1 Sobre os dados . . . . .	31
4.2 Extração, Transformação e Carga . . . . .	32
4.3 Modelagem do <i>Data Warehouse</i> . . . . .	34
4.4 Resultados . . . . .	38
4.5 Limitações . . . . .	48
<b>5 Conclusões e Trabalhos Futuros</b>	<b>49</b>
<b>Bibliografia</b>	<b>51</b>

## Lista de Figuras

2.1	Exemplo de um esquema de constelação de fatos (GOEL, 2023). . . . .	20
4.1	Diagrama de relacionamento entre as tabelas do <i>Data Warehouse</i> . . . . .	37
4.2	Número de entrevistas por região. . . . .	38
4.3	Número de entrevistas por classe social. . . . .	39
4.4	Número de Entrevistas por Região e Classe Social . . . . .	39
4.5	Quantidade de eletrodomésticos distintos nos domicílios por classe social. .	40
4.6	Consumo anual por classe social. . . . .	42
4.7	Quantidade de eletrodomésticos em funcionamento por hora. . . . .	42
4.8	Quantidade de eletrodomésticos de climatização em funcionamento por hora	44
4.9	Média da quantidade de ar-condicionados em domicílios por região. . . . .	44
4.10	Média da quantidade de ar-condicionados em domicílios por classe social. .	45

## Lista de Tabelas

3.1	Comparação Trabalhos Relacionados . . . . .	29
4.1	Tabelas presentes no DW e seus números de linhas . . . . .	36
4.2	Top 10 Eletrodomésticos mais comuns por classe social . . . . .	41
4.3	Tradução dos nomes das perguntas da Tabela 4.4 . . . . .	46
4.4	Distribuição de Respostas para Práticas de Economia de Energia . . . . .	47

## Lista de Abreviações

DCC	Departamento de Ciência da Computação
DW	<i>Data Warehouse</i>
ETL	Extração, transformação e carga
HQL	<i>Hibernate Query Language</i>
N/A	Não se aplica
N.I	Não Informado
PPH	Pesquisa de Posse e Hábitos de Uso de Equipamentos Elétricos na Classe Residencial
SQL	Linguagem de consulta estruturada
TCC	Trabalho de Conclusão de Curso
UFJF	Universidade Federal de Juiz de Fora

# 1 Introdução

Nosso planeta sempre passou por mudanças naturais do clima e de temperatura, mas desde o século XVIII, com o advento da Primeira Revolução Industrial, essas mudanças têm se intensificado devido aos aumentos da emissão de gases do efeito estufa na atmosfera (CINTRA, 2023). Esses gases são predominantemente eliminados a partir da queima de combustíveis fósseis, pelo desmatamento e pelas queimadas (LEGNAIOLI, 2024). Segundo o Painel Intergovernamental sobre Mudanças Climáticas (IPCC), estima-se que até o final do século a temperatura média do planeta possa aumentar de 3.3 °C a 5.7 °C se medidas de redução da emissão de gases de efeito estufa não forem tomadas (BOEHM; SCHUMER, 2023).

Entre os efeitos causados pelas mudanças climáticas, temos as fortes ondas de calor, chuvas acentuadas e grandes períodos de estiagem. Para o Brasil, um país cuja matriz elétrica é composta por cerca de 55% pela energia hidrelétrica (SPUTNIK, 2024), os períodos de estiagem acabam se tornando um problema devido às baixas nos volumes dos reservatórios. As regiões Sudeste e Centro-Oeste são responsáveis por cerca de 70% da geração de energia elétrica no país e, em 2021, foram forçadas a operar com cerca de 22.5% de suas capacidades devido a uma das maiores crises hídricas enfrentadas nos últimos 100 anos (MACKENZIE, 2022). Esta crise fez com que o país precisasse utilizar termelétricas para suprir o abastecimento de energia, entretanto, essa medida fez com que os custos com a conta de energia tivessem um aumento de quase 25% no ano, devido ao alto custo para geração de energia neste segmento (MALAR, 2021). Outras duas crises hídricas registradas no país foram a crise de 2014 e 2015, que prejudicou o abastecimento de água em cidades do estado de São Paulo, devido à baixa nos reservatórios do sistema Cantareira (MARTÍN, 2015). A segunda foi a crise de 2001, uma das mais marcantes do país, marcada por apagões resultantes da baixa nos reservatórios das hidrelétricas, da alta dependência do país nesta fonte de geração de energia e da falta de investimentos no setor elétrico (MACEDO et al., 2014).

Na contramão do que foi dito anteriormente, existem períodos em que o Brasil

produz mais energia do que é capaz de consumir, tendo assim que jogar fora parte desta energia já que não é possível armazená-la para consumo posterior. Esse processo se dá pela abertura das comportas de usinas hidrelétricas, fazendo com que a água escoe e não passe pelas turbinas de geração de energia e, pelo desligamento de fontes solares e eólicas (CARREGOSA, 2024). É possível observar também que a produção de energia varia em razão das fontes solares e eólicas. A produção de energia elétrica é mais abundante durante o dia devida a participação da energia solar e, em períodos com menos vento, a energia eólica será produzida em menor quantidade. Para atender horários de pico e que a produção de energia não seja suficiente, as usinas termelétricas precisam ser ativadas para garantir o abastecimento (BRAZ, 2024).

Neste contexto, uma questão que também merece atenção é a rede de distribuição energética brasileira. A distribuição de eletricidade no Brasil enfrenta vários desafios, incluindo desigualdades sociais e espaciais, dificuldades para alcançar regiões remotas como o Pantanal Sul-Mato-Grossense, questões regulatórias e falta de gerenciamento para conectar fontes de geração distribuída. As disparidades sociais e espaciais resultam em menor acesso à eletricidade nas regiões norte e nordeste (THIVES; GHISI; JÚNIOR, 2022), enquanto as áreas remotas, como o Pantanal, lutam com o reconhecimento dos custos operacionais e de manutenção pelas agências reguladoras (PAULA; MEDEIROS; SALAS, 2018). As questões regulatórias abrangem revisões de tarifas, intervenção estatal e proteção ao consumidor, impactando o setor de distribuição (FERNANDES, 2018). Além disso, a relutância das empresas de distribuição em conectar novas fontes de geração, como o biogás, dificulta o investimento em geração distribuída, afetando a qualidade geral da eletricidade fornecida (BRIGNOL et al., 2017).

Por meio do entendimento por parte das empresas de como a energia é consumida por seus clientes e quais os horários de maior consumo, decisões podem ser tomadas para melhoria na infraestrutura e da confiabilidade da rede de distribuição. Uma infraestrutura mais moderna pode garantir que a energia seja melhor distribuída entre as regiões que geram mais e menos energia em determinados períodos, podendo assim reduzir a utilização das termelétricas e, conseqüentemente, a geração de gases do efeito estufa.

Um tema muito citado nos últimos anos e que pode nos ajudar a enfrentar mo-

mentos de escassez hídrica, questões relacionadas a distribuição de energia e, principalmente, reduzir a emissão de gases de efeito estufa é a eficiência energética. A eficiência energética está ligada aos padrões de geração e de consumo de energia, tratando de como gerar energia de maneira mais eficiente, consumindo menos recursos para obter os mesmos resultados (ENGIE, 2022).

Assim, a população também pode contribuir conhecendo as boas práticas para economizar energia, mas para isso deve ser conscientizada por governos e empresas de energia que conheçam seus perfis de consumo. Para obter informações importantes sobre o consumo de energia em domicílios, é preciso que tenhamos dados reais e estes dados precisam ser coletados através de pesquisas e armazenados de maneira simples e segura para que possam ser posteriormente analisados. Dito isso, este trabalho busca mostrar que *Data Warehouses* podem ser uma alternativa eficiente para o armazenamento de dados de consumo de energia elétrica no Brasil, além de facilitar a realização de análises para obtenção de informações acerca destes dados.

## 1.1 Contextualização

O aumento gradual do consumo energético mundial, atrelado a preservação dos recursos naturais e a redução da emissão dos gases de efeito estufa, são os principais motivadores destas discussões (EDP, 2023). Como dito anteriormente, o Brasil é um país que possui a matriz elétrica composta, predominantemente, pela energia hidrelétrica, o que faz com que o país deva se preocupar ainda mais com as mudanças climáticas geradas pelo efeito estufa, já que isso pode resultar em alterações na intensidade e na distribuição das chuvas durante os meses do ano e acabar afetando negativamente a geração de energia (GEOGRAPHIC, 2022).

Para que as empresas fornecedoras de energia possam garantir uma distribuição de energia estável para o abastecimento da população, planejar investimentos, garantir a eficiência energética e o atendimento das demandas de seus clientes, é importante que elas conheçam os hábitos de consumo de seus consumidores (CASAGRANDE; RODRIGUES; GUIMARÃES, 2023). Consumidores que tenham conhecimento desses dados podem aprender como mudar seus hábitos, reduzir o consumo de energia e, por con-

sequência, reduzir os custos com as contas de energia elétrica, além de preservar o meio ambiente.

## 1.2 Descrição do Problema

Poucas são as empresas e órgãos que disponibilizam dados sobre o consumo de energia elétrica no Brasil. Podemos citar a Agência Nacional de Energia Elétrica (ANEEL), a Empresa de Pesquisa Energética (EPE), o Programa Nacional de Conservação de Energia Elétrica (Procel) e Operador Nacional do Sistema Elétrico (ONS). Geralmente, os dados que esses órgãos disponibilizam sobre consumo de energia elétrica residencial são mais gerais, tratando do consumo de energia nas residências de uma forma regional.

Para que possamos entender como as residências podem reduzir o consumo de eletricidade, se faz necessário entender como a energia é utilizada dentro delas. São necessários dados sobre quais e quantos eletrodomésticos são usados, a frequência de uso, o tempo de utilização, o consumo de equipamentos de climatização (ar-condicionado, por exemplo) e de aquecimento de água, entre outros.

Uma das poucas pesquisas sobre o consumo final de eletricidade é a Pesquisa de Posse e Hábitos de Uso de Equipamentos Elétricos na Classe Residencial de 2019 (PPH-2019)<sup>1</sup>, realizada pela Eletrobrás, em parceria com o Procel e, teve com objetivo conhecer os perfis de posse e hábitos de uso de equipamentos eletrônicos nas residências. Em posse destes dados, é importante que possamos interpretá-los para gerarmos informações importantes na tomada de decisões.

Quando disponibilizados, os dados são apresentados no formato de planilhas, dificultando a extração de informações e análise global do consumo de energia. Assim, o problema abordado nesse trabalho de conclusão de curso é: *como organizar os dados de consumo de energia elétrica no Brasil de forma a facilitar sua análise e auxiliar de alguma forma na redução do consumo de eletricidade?*

---

<sup>1</sup><https://eletrobras.com/pt/AreasdeAtuacao/PPH%202019%20-%20Question%C3%A1rio%20da%20Pesquisa.pdf>

## 1.3 Justificativa

Diante da necessidade de entender os hábitos de consumo de energia descritos nos dados da pesquisa PPH-2019, uma solução para o armazenamento desses dados é um *Data Warehouse* (DW). Por meio de um DW, dados provenientes de diferentes origens podem ser armazenados e organizados em uma estrutura que facilita análises e consultas de informações que apoiem a tomada de decisão (FERREIRA et al., 2010).

Ao utilizar um Data Warehouse, podemos realizar consultas para gerar visualizações gráficas e tabelas capazes de apresentar informações que descrevem o cenário atual de consumo nos domicílios. Além disso, essas informações podem auxiliar empresas e órgãos, como a Eletrobrás e o Procel, a desenvolverem estratégias eficientes para conscientizar a população em relação a economia no consumo de energia. Assim, todos podem contribuir significativamente na preservação do meio ambiente, auxiliar a superar momentos de crises na geração de energia, além de auxiliar as empresas fornecedoras de energia a entenderem seus consumidores para poderem melhorar sua infraestrutura para garantir uma rede de distribuição mais eficiente e que possa atender um maior número de consumidores.

## 1.4 Objetivos

O principal objetivo deste trabalho é modelar e construir um Data Warehouse a partir dos dados coletados pela pesquisa PPH-2019 <sup>2</sup>, da Eletrobras e, a partir desse DW, realizar a obtenção de informações importantes acerca de padrões de consumo dos usuários. Em posse destes dados é possível extrair informações relevantes para processos de tomada de decisão. Para isso, os seguintes objetivos secundários precisam ser alcançados:

- obter os dados da pesquisa;
- analisar e entender o conjunto de dados em busca das variáveis de interesse;
- selecionar quais são os dados possíveis de serem usados para a pesquisa;
- criar um tipo de vocabulário para a redução dos nomes das colunas das tabelas;

---

<sup>2</sup><https://eletrobras.com/pt/AreasdeAtuacao/PPH%202019%20-%20Banco%20de%20Dados%20V2.xlsx>

- realizar a extração, transformação e carga dos dados no modelo do banco de dados (processo ETL);
- explorar os dados e obter os resultados;
- realizar a disponibilização de conjuntos de dados, códigos e de resultados encontrados.

## 1.5 Metodologia

Para o desenvolvimento deste trabalho, o primeiro passo foi a busca por informações na literatura, que fossem úteis para a tomada de decisões ao longo do desenvolvimento do projeto. Em seguida, foi obtida a base de dados referente à pesquisa PPH-2019, permitindo o entendimento dos dados e suas representações, o que possibilitou a definição das variáveis de interesse para a construção do projeto e definição do modelo do banco de dados.

A etapa seguinte envolveu a modelagem dos dados, onde foram definidos quais tabelas seriam geradas e seus respectivos relacionamentos. Simultaneamente, foram desenvolvidos códigos para a extração e transformação dos dados, preparando-os conforme os padrões estabelecidos durante a modelagem. *Scripts* em linguagem Python e SQL foram construídos para a inserção dos dados no *Data Warehouse*.

A última fase do trabalho consistiu na exploração dos dados em busca de informações relevantes sobre a posse e os hábitos de uso de equipamentos elétricos. Com essas informações, deve ser possível gerar gráficos e indicadores significativos para a tomada de decisões e para a compreensão do cenário atual do uso de energia elétrica em domicílios e a posse de equipamentos elétricos.

## 1.6 Organização do Trabalho

Além desta introdução, este trabalho de conclusão de curso está organizado da seguinte forma: no Capítulo 2, temos a apresentação dos fundamentos teóricos e conceitos importantes, que serão essenciais para o entendimento do trabalho. O Capítulo 3 apresenta os

---

trabalhos relacionados e uma breve comparação entre eles. O Capítulo 4 detalha todo o processo de desenvolvimento do trabalho, as análises realizadas e os resultados encontrados. Por fim, no Capítulo 5, são apresentadas a conclusão e possíveis trabalhos futuros.

## 2 Fundamentação Teórica

Este capítulo propõe apresentar alguns conceitos fundamentais para o desenvolvimento deste trabalho. A seção 2.1 apresenta um pouco do que são os Sistemas de Apoio à Decisão, como os dados são importantes para eles e a importância desses sistemas para as empresas. Na seção 2.2 é apresentado o processo de extração, transformação e carga e como ele se relaciona com os *Data Warehouses*. Em seguida, temos a seção 2.3 que apresenta o que são os *Data Warehouses*, a importância, aplicação, a maneira que são modelados e que estruturas os compõem. A partir destes conceitos, é possível entender mais sobre o processo de construção de um DW para armazenamento de dados e auxílio na tomada de decisões.

### 2.1 Sistemas de apoio à decisão

Com a globalização, nos últimos anos o número de dados gerados e coletados a partir das mais diversas fontes tem se tornado cada vez maior. Para o matemático Clive Humby: *“Os dados são o novo petróleo. Como o petróleo, os dados são valiosos, mas se não refinados não podem realmente ser usados. Tem de ser transformado em gás, plástico, produtos químicos, etc. para criar uma entidade valiosa que impulse a atividade lucrativa. Então, os dados devem ser discriminados, analisados para que tenham valor.”* (TALAGALA, 2022). Isso nos mostra que possuir dados e realizar boas análises sobre eles, é algo de grande valor e que pode ajudar empresas a desenvolverem estratégias que minimizem custos e aumentem seus lucros. Para isso, é importante a utilização de sistemas que apoiem os processos de tomada de decisão.

Sistemas de Apoio à Decisão são sistemas computacionais que combinam dados e informações com o objetivo de auxiliar na resolução de problemas (CASTOR et al., 2022). São essenciais para reduzir o tempo e o custo em processos de tomada de decisão, além de aumentar o grau de assertividade em relação às decisões tomadas, mesmo que os usuários não conheçam totalmente o problema que estejam tentando resolver (ÖZTÜRK

et al., 2024). Como componentes desses sistemas, estão incluídos os bancos de dados, bases de modelos, interfaces com o usuário e o acesso à internet (BATISTA, 2016).

A análise de dados está fortemente ligada aos sistemas de apoio à tomada de decisão. É com ela que os dados serão tratados, analisados e transformados em informações relevantes para o negócio. Para isso, é importante que os dados das mais diversas fontes possam ser coletados e armazenados de maneira eficiente e centralizada para facilitar as análises. Uma abordagem que pode ser usada é a construção de *Data Warehouses*.

## 2.2 Extração, transformação e carga

O ETL, do inglês *Extract Transform Load*, é o processo de coletar dados de diversas fontes e formatos, realizar transformações com base em regras pré-definidas e carregá-los em um novo ambiente (ZENDESK, 2023).

De acordo com a documentação da Oracle<sup>3</sup>, a etapa de Extração é responsável por realizar a identificação dos dados e realizar a cópia a partir de suas origens para um sistema intermediário de destino. Os dados podem vir de diversas fontes e formatos, podendo ser estruturados e não estruturados. O processo de Transformação é responsável por capturar os dados em seu formato original e os transformar por meio de operações de remoção de duplicidades, validações, modelagens e agregações para que respeitem as regras definidas no sistema de destino. Já a etapa de Carga é composta pelo processo de inserir os dados transformados no sistema de destino. É um processo que pode ser feito em tempo real, em lotes, pode ser feito a partir da carga total dos dados vindos das origens ou pode contar com processos de carga incrementais, que levam os dados mais recentes para o novo armazenamento e realiza o incremento da base ao longo do tempo.

O uso de ETLs é fundamental no processo de criação de um *Data Warehouse*, é por meio dele que poderemos centralizar os dados em um sistema que facilite a realização de consultas e análises sobre esses dados.

---

<sup>3</sup><https://www.oracle.com/br/integration/what-is-etl/>

## 2.3 Data Warehouse

Os *Data Warehouses*, armazéns de dados, são repositórios centrais de dados construídos para dar suporte à processos de análises de dados (TORANZO, 2022). São destinados para realização de consultas e análises sobre um grande volume de dados, permitindo que sejam extraídas informações úteis para auxiliar nas tomadas de decisões. Podem ser formados por um ou mais sistemas de gerenciamento de banco de dados que centralizam e armazenam grandes quantidades de dados vindos das mais diversas fontes, além de otimizar os processos de busca de informações.

Entre as principais qualidades dos *Data Warehouses*, podemos citar:

- **orientados por assunto:** a modelagem se orienta a partir dos assuntos principais da empresa, assim dados de mesmos assuntos tender a ser agrupados;
- **integração:** visa unificar os dados vindos das diversas áreas e sistemas da empresa, com objetivo de padronização entre dados de fontes e tipos distintos;
- **não volátil:** ao contrários de sistemas transacionais onde os dados sofrem muitas alterações por inclusão, alteração e exclusão, em um DW os dados são trazidos de outros ambientes e passam por processos de filtragem, limpeza e transformação para garantir a qualidade das informações. Assim, os dados ficam disponíveis para consultas até que não sejam mais necessários;
- **dados não variam com o tempo:** bancos transacionais armazenam o valor atual das informações, podendo assim variar com o tempo. Entretanto, em um *Data Warehouse* é realizada a manutenção de um histórico desses dados, permitindo a visualização de períodos de tempo de acordo com o momento em que foram capturados (MACHADO, 2004).

A técnica usada para modelar de maneira otimizada a estrutura de uma *Data Warehouse* é a Modelagem Dimensional (KIMBALL; ROSS, 2011). O objetivo desse tipo de modelagem é a melhoria do processo de recuperação dos dados, facilitando as análises e obtenção de informações. Os modelos dimensionais são estruturas desnormalizadas e que foram projetadas para a recuperação de dados, algo diferente do modelo entidade-

relacionamento que tenta eliminar redundâncias e são voltados para operações de inserção, atualização e exclusão (AHMED, 2024).

O Modelo Dimensional é composto por tabelas de dimensão e tabelas de fato.

- **fato:** são as tabelas centrais que armazenam as medidas ou métricas do processo de negócio. São cercadas por tabelas de dimensão, com as quais se relacionam a partir de chaves estrangeiras;
- **dimensão:** são as tabelas compostas por dados textuais que tem o objetivo de caracterizar as medidas presentes nas tabelas fato.

A organização das tabelas fato e tabelas dimensão pode ter vários nomes de acordo com a quantidade de tabelas fato. Quando há a presença de somente uma tabela fato central e que se relaciona com tabelas de dimensão, é chamado de esquema estrela (*star schema*). Quando temos a presença de tabelas de dimensão normalizadas em outras tabelas, temos o esquema de flocos de neve (*snowflake schema*). Por último, quando existem várias tabelas fatos e elas compartilham de tabelas de dimensão, chamamos de constelação de fatos ou esquema de galáxia (*fact constellation*) (GOEL, 2023). A figura 2.1 apresenta um exemplo de um esquema de constelação de fatos, nele podemos ver duas tabelas fato compartilhando de tabelas dimensão e seus relacionamentos.

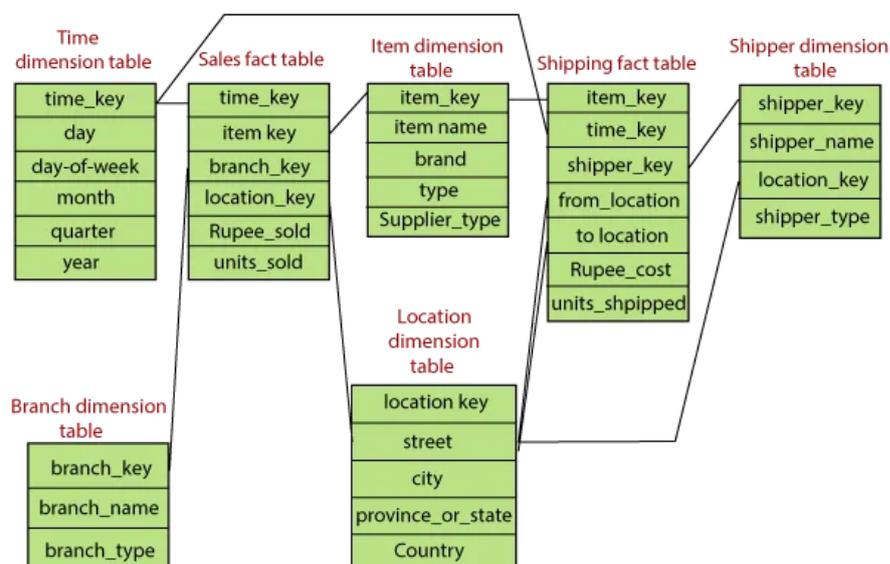


Figura 2.1: Exemplo de um esquema de constelação de fatos (GOEL, 2023).

---

## 2.4 Considerações finais

Neste capítulo foram apresentados conceitos importantes que serão usados durante todo o desenvolvimento deste trabalho. Inicialmente foi introduzida a ideia de sistemas de apoio à decisão, o que são e como estão fortemente ligados a processos de análises de dados. Em seguida, temos a apresentação do processo de extração, transformação e carga, que é responsável por obter dados de vários tipos e origens e aplicar transformações para poderem efetuar a carga desses dados em um ambiente novo. Por último, o conceito de *Data Warehouse* foi apresentado, o seu objetivo dentro das organizações e algumas de suas qualidades. O modelo dimensional foi descrito, sua diferença para o modelo entidade-relacionamento, o que são as tabelas de dimensão e fato que o compõe e quais os esquemas usados para representar a maneira como as tabelas fato e dimensão devem se relacionar.

## 3 Trabalhos Relacionados

Neste capítulo serão apresentados alguns trabalhos que, de alguma maneira, se relacionam com este trabalho de monografia. São apresentados trabalhos que tratam de sistemas que auxiliam no apoio a decisões e que se baseiam em *Data Warehouses* e um trabalho que trata sobre aplicações de técnicas estatísticas para análise do consumo de energia em residências. Para cada trabalho foi apresentada uma breve descrição sobre o projeto e os resultados que foram apresentados, e por fim, foram comparados com o presente trabalho de monografia. Ao final do capítulo, foi construída uma tabela que apresenta algumas diferenças e semelhanças entre todos os trabalhos e este trabalho de monografia.

### 3.1 The Application of Data Warehouse in Decision Support System

Em Liu e Qi (2012), os autores discutem sobre o crescimento e desenvolvimento de tecnologias de banco de dados, e a geração e acumulação de um volume cada vez maior de dados. É apresentado que o uso de bancos de dados tradicionais para extração de informações para tomada de decisão, acaba sendo uma fonte pouco eficiente devido à limitações na capacidade de análise dos dados, dificuldades na construção de consultas complexas e o baixo desempenho dessas consultas.

O artigo apresenta o que são *Data Warehouses*, e discute a importância de serem orientados à assuntos e a capacidade de integração de informações vindas de diversas fontes. É apresentado um pouco do que são os sistemas de apoio à decisão e sua importância na geração e apresentação de informações precisas nos processos de tomada de decisão. Também é abordado um processo de extração de informações implícitas a partir dos dados, chamado escavação de dados.

O trabalho fez uso de um sistema de gestão de supermercado como exemplo para a aplicação de sistemas de apoio à decisão baseado em *Data Warehouse*, em conjunto

com processos de escavação de dados. Como isso, foi descrito a possibilidade de se obter informações sobre os hábitos de compra e as preferências de seus clientes, podendo assim identificar padrões de mercadorias que sejam vendidas juntas e isso pode ser utilizado para tomar decisões de como organizar os produtos nas prateleiras de forma a otimizar o tempo de compra dos clientes e, também fazer com que eles sejam influenciados a comprar determinadas mercadorias.

Também foi apresentado que o uso desses sistemas em supermercados pode auxiliar na busca por informações sobre os produtos em estoque e tomar decisões para redução dos custos de armazenagem. Pode-se também gerar informações sobre cada filial, para ser possível entender sobre a situação das vendas e do estoque de cada uma. Além disso, os supermercados podem realizar análises para identificar a eficiência de propagandas e promoções, a fim economizar recursos da empresa.

O artigo apresentou diversas vantagens do uso de *Data Warehouses* como sistemas de apoio à decisão e suas aplicações. Este Trabalho de Conclusão de Curso tentou fazer uso dessas vantagens para gerar dados relevantes no contexto da eficiência energética residencial.

## **3.2 Energy Provider Data Warehouse Design and Implementation - Case Study**

O artigo de Kryeziu et al. (2019) fornece um estudo de caso no projeto e implementação de um *Data Warehouse* para construção de relatórios para a KEDS (Empresa Fornecedora de Energia de Kosovo, país do sudeste da Europa). O objetivo do trabalho era demonstrar que um DW pode ser usado para integrar dados de diferentes departamentos da empresa. Segundo o artigo, os dados da empresa estavam distribuídos em sete bancos de dados transacionais diferentes e que podiam se conectar ao DW. O processo de extrair, transformar os dados em modelos lógicos e fazer a carga dos dados no DW, foi a principal etapa do processo de construção do DW.

O data warehouse apresentado foi construído usando o SQL Server 2012. Como ferramenta de ETL foi usado o *SQL Server Integration Services* (SSIS), *SQL Server*

*Analysis Services* (SSAS) como ferramenta de criação de cubos e, o *SQL Server Reporting Services* (SSRS) para construção de relatórios. O projeto deste DW faz uso de vários esquemas em estrela, onde tabelas de dimensão e fato são criadas e relacionadas usando chaves primárias e estrangeiras.

Com o *Data Warehouse* criado, foram geradas consultas para comparação de desempenho entre o esquema estrela do DW e as instâncias dos bancos transacionais. A expectativa era que o DW se saísse melhor no tempo de consulta já que, em bancos transacionais, o número de junções entre tabelas seria maior e haveria uma complexidade maior na geração dos resultados. Utilizando a linguagem T-SQL (*Transact-SQL*), as mesmas consultas foram executadas nos dois modelos de banco e constatou-se que o tempo de execução de consultas no DW foram consideravelmente menores em todos os casos testados, se mostrando assim uma maneira mais rápida e eficiente na geração de resultados para a empresa.

Neste trabalho de TCC, não foram usadas nenhuma dessas ferramentas citadas no artigo e deu-se preferência por ferramentas gratuitas. Como o artigo mostrou, um *Data Warehouse* é uma maneira bastante eficiente para armazenamentos dos dados de forma a otimizar as consultas, essa ideia é fundamental na construção do presente trabalho, na busca por uma solução eficiente na tomada de decisão acerca da eficiência energética.

### **3.3 Data Warehouse and Decision Support on Integrated Crop Big Data**

Em Ngo, Khac e Kechadi (2020), é discutida a necessidade do aumento da produção de alimentos devido ao crescimento da população mundial e ao grande número de pessoas que enfrentam problemas de insegurança alimentar. Para o aumento da produção, se faz importante o crescimento da produtividade, além da modernização das técnicas de cultura. Para auxiliar nesse processo, sistemas de apoio à decisão com base em *Big Data*, são usados para fornecer informações importantes sobre os mais diversos assuntos, como meteorologia, sistemas de irrigação, solo, uso de inseticidas e herbicidas, e muitos outros.

Nesses sistemas de apoio à decisão são usados dados vindos de diversas origens e que passam por processos de mineração de dados para coleta, modelagem, armazenamento e análises.

Com isso, os autores julgaram como importante o projeto de um *Data Warehouse*, chamado de Armazém de Dados Agrícolas (ADW), para armazenar os dados de maneira eficiente e uniforme para representar os diversos tipos de dados oriundos de diversas origens como imagens de drones e satélites, dados de sensores, bancos de dados e muitos outros. O processo de ETL não foi especificado, então não foram informadas quais as ferramentas usadas. Foi dito que o projeto desse DW de grande porte é algo que apresenta muitos desafios, devido a presença de dados espaciais, temporais, complexos, de alta dimensão e não padronizados.

Os dados usados foram obtidos de uma empresa de agronomia e são dados de alguns países da Europa. Foram coletados cerca de 29 conjuntos de dados, com média de 18 tabelas e 1.4GB cada. Para modelar essa grande quantidade de dados, foi usado o esquema de constelação, devido a presença de várias tabelas de fatos. Os dados brutos foram armazenados no banco de dados NoSQL Cassandra e, os dados são importados para o Hive, um sistema de *Data Warehouse* construído para análises de grandes conjuntos de dados armazenados em arquivos, processados e enviados para o MongoDB (banco de dados NoSQL). Todo o projeto foi desenvolvido para ser executado em sistemas distribuídos ou em nuvem.

O banco de dados do ADW foi copiado para o MySQL para que fosse feita uma comparação de desempenho para operações de leitura entre bancos SQL e NoSQL. Foram combinados comandos semelhantes de HQL/SQL, e foram gerados 10 grupos com 5 consultas cada. Como resultado, foi constatado que o ADW baseado em NoSQL foi mais eficiente em 46 das 50 consultas realizadas. Em apenas 3 consultas o MySQL foi mais eficiente e em uma consulta houve um empate entre eles.

O artigo mostra a importância de um DW no setor da agricultura, devido ao grande número de dados e a necessidade de tomar decisões importantes para melhoria da produção. Em comparação com o presente trabalho, temos como semelhança a apresentação e construção de um DW para armazenar informações de maneira que facilite a

extração de informações. Temos que a maior diferença é o uso de banco de dados NoSQL no artigo, e SQL neste trabalho.

### 3.4 COVID-WAREHOUSE: A Data Warehouse of Italian COVID-19, Pollution, and Climate Data

O pesquisa de Agapito, Zucco e Cannataro (2020), apresentou como a pandemia da COVID-19 gerou desafios nunca antes vistos em relação aos diversos campos e como eles poderiam se beneficiar com a tecnologia, principalmente para integrar dados sobre o avanço da doença. O artigo utilizou dados da COVID na Itália e esses dados eram obtidos de administrações de províncias, regiões e do país como um todo. Esses dados eram fornecidos por meio de arquivos CSV pelo Departamento de Proteção Civil da Itália, mas também haviam dados em HTML que eram disponibilizados por algumas regiões provinciais.

Foi proposto então a criação de um DW que era formado por um *framework* construído em Python, que simulava as características de um DW tradicional e foi adaptado para trabalhar com pequenos volumes de dados. Esse DW foi denominado como COVID-WAREHOUSE. Além disso, foi integrado à esse DW, dados sobre poluição e clima que foram disponibilizados por duas regiões da Itália, e que foram coletados a partir de páginas HTML por meio das bibliotecas Python Beautiful Soup e Requests, bibliotecas utilizadas para realização de acesso e manipulação de páginas web na linguagem Python. Todos os dados eram tratados e convertidos em CSV e o *Data Warehouse* funcionava por meio de relacionamentos entre *dataframes* da biblioteca Pandas. Fazendo uso de funções OLAP (Processamento Transacional Online), foi possível estudar possíveis correlação entre os dados de poluição e clima com os dados da COVID no país.

Fazendo o uso de mapas de calor para representar a correlação entre variáveis, puderam constatar que a pandemia se espalhou mais em regiões que possuíam grandes concentrações de partículas no ar, além de possuírem baixas ocorrências de ventos e chuvas.

Podemos ver neste trabalho a importância de utilizar um *Data Warehouse* para

armazenar dados de diversas origens e de uma maneira que facilite a geração de consultas. Com dados como esses, é possível que órgãos do governo tomem decisões sobre campanhas de conscientização e vacinação e que sejam mais efetivas, atingindo diretamente as regiões mais afetadas. É um trabalho construído de maneira bem diferente do convencional, e deste TCC, já que não faz uso de nenhum sistema gerenciador de banco de dados, apenas utiliza do Python e arquivos CSV para abstrair e representar as operações necessárias que seriam realizadas por um banco de dados.

## **3.5 Estimativa da evolução do uso final de energia elétrica no setor residencial do Brasil por região geográfica**

O trabalho de Abrahão e Souza (2021) buscou analisar o consumo de energia elétrica em domicílios para 2005 e 2019, fazendo uso de métodos estatísticos. Para isso, foram usados os dados da Pesquisa Posse e Hábitos de Uso de Equipamentos Elétricos na Classe Residencial referente aos anos de 2004-2006 e 2018-2019. Os dados foram analisados por região geográfica e, os eletrodomésticos foram categorizados como conforto ambiental, conservação de alimentos, iluminação, lazer, serviços gerais e aquecimento de água. Foram utilizados também dados sobre a potência média dos eletrodomésticos mas, para isso foi necessário o levantamento destes dados a partir de outras fontes.

Várias equações foram apresentadas para representar as análises efetuadas. Entre essas análises, temos o consumo de energia elétrica estimado equipamento por região, consumo de energia elétrica por região, uso final de energia elétrica por equipamento por região e, ainda foi estipulado uma maneira de calcular o erro global considerando todo o processo.

Foi observado que todas as regiões apresentaram uma redução no uso de equipamentos de aquecimento de água, exceto na região Sul. Os equipamentos de conservação de alimentos apresentaram uma participação menor no total em decorrência do crescimento do uso de equipamentos de outras categorias. As categorias que mais consomem energia

são as de aquecimento de água, conforto ambiental e conservação de alimentos. A categoria de conforto ambiental apresentou crescimento entre 2005 e 2019, com destaque para as regiões Nordeste e Sudeste, devido ao aumento do uso de ar-condicionado e ventiladores. A categoria de aquecimento de água apresentou queda em todas as regiões, exceto a Sul e Centro-Oeste. Logo, foi verificado que o consumo de energia elétrica em domicílios mudou ao longo dos anos e que essas mudanças foram diferentes para cada região.

Pode-se concluir pelo artigo que entender mais sobre os hábitos de consumo de energia elétrica, além de entender esses hábitos por região, contribui para o desenvolvimento de políticas de conservação de energia, e também ajuda empresas de distribuição de energia a tomarem decisões importantes sobre melhorias que devem ser feitas para suprir as necessidades dos usuários e evitar desabastecimento.

Este Trabalho de Conclusão de Curso abordou algumas informações semelhantes, principalmente o uso de eletrodomésticos para cada categoria. A construção do *Data Warehouse* fez com que pudéssemos gerar consultas e extrair informações detalhadas sobre os usos dos eletrodomésticos e também contribuir em processos de tomada de decisão, algo que poderia ter sido implementado no artigo citado.

## 3.6 Uso de Data Warehouse para Análise do Consumo de Energia no Brasil

Em ARAGÃO (2023), um *Data Warehouse* é construído a partir dos dados da pesquisa PPH-2019. O processo foi dividido em três fases, começando pela pré-análise e organização dos dados da pesquisa, além da execução de processos de limpeza e extração de dados.

Para construção do DW foi utilizado o SQLite, DBeaver para visualização e manipulação dos dados, e gráficos gerados a partir do Matplotlib, além de todo o processo de ETL ser construído utilizando a linguagem Python com auxílio da biblioteca Pandas. Foram construídas as tabelas de dimensão residência, classe social, data, região e conscientização e a tabela fato consumo, formando assim um modelo estrela.

Como resultados, foram apresentadas informações sobre o consumo de energia por região, por classe econômica, por quantidade de moradores nas residências, por metro

quadrado e por nível de conscientização.

Realizando uma pré-análise dos dados da pesquisa PPH-2019, podemos perceber que o artigo fez uso de uma parte pequena entre todos os dados disponíveis. Logo, este TCC tentou explorar dados que não foram explorados pelo autor. Um DW bem maior em relação ao número de tabelas foi construído para armazenar também as informações de eletrodomésticos e extrair informações sobre eles.

## 3.7 Conclusão

Neste capítulo foram apresentados alguns trabalhos que de alguma forma se relacionam com este Trabalho de Conclusão de curso. Dois deles também abordam o tema de energia elétrica residencial, e os outros quatro apresentam como os *Data Warehouses* podem ser usados para auxílio na tomada de decisões e que podem ser aplicados nas mais diversas áreas. A tabela 3.1 apresenta algumas diferenças e semelhanças entre os trabalhos.

Tabela 3.1: Comparação Trabalhos Relacionados

Trabalho	Solução	Tipo de dados	Alvo do apoio à decisão	Tecnologia de armazenamento	Tecnologia usada na construção do DW
Este trabalho	DW	planilhas excel	eficiência energética	PostgreSQL	Python, Pandas e SQL
Liu e Qi (2012)	DW	relatórios de vendas e estoque	gerenciamento de supermercados	N.I	N.I
Kryeziu et al. (2019)	DW	bancos de dados transacionais	geração de <i>insights</i>	SQL Server 2012	SSIS e SSAS
Ngo, Khac e Kechadi (2020)	DW	imagens, dados de sensores, bancos de dados	eficiência agrícola	MongoDB e Hive	Cassandra para armazenar os dados brutos*
Agapito, Zucco e Cannataro (2020)	DW*	CSV e HTML	ações de saúde pública	arquivos CSV	Python, Pandas, Beautiful Soup e Requests
Abrahão e Souza (2021)	N.I	documentos e planilhas excel	política energética brasileira	planilhas excel	N/A
ARAGÃO (2023)	DW	planilhas excel	eficiência energética	SQLite	Python, Pandas e SQL

Em relação à solução utilizada nos trabalhos, podemos notar que apenas Abrahão e Souza (2021) não utilizou de *Data Warehouses* e também não informou (N.I) a solução aplicada. O trabalho de Agapito, Zucco e Cannataro (2020) está marcado com um destaque na coluna de solução pelo fato de construírem um DW de uma maneira diferente da convencional, que seria por meio de um banco de dados. O *Data Warehouse* construído foi representado por um *framework* construído com Python e Pandas que simulava um DW, utilizando arquivos CSVs como tabelas e *dataframes* Pandas para efetuar os

relacionamentos e consultas.

Analisando os tipos de dados usados nos trabalhos, podemos notar que bancos de dados e planilhas excel foram fontes bastante comuns. O trabalho de Liu e Qi (2012) não especificou os formatos dos arquivos usados, apenas que eram relatórios de vendas e estoque. Entre os trabalhos, Ngo, Khac e Kechadi (2020) se destaca por usar dados complexos, como imagens de satélite e drone e também dados de sensores.

Os alvos no apoio à decisão dos trabalhos foram bem diversos, de acordo com o tema de cada um. No geral, todos foram usados para facilitar análises de dados por meio de um repositório centralizado e otimizado, para que informações importantes pudessem ser geradas para agregar de alguma forma em processos de tomada de decisão.

Comparando as tecnologias de armazenamento utilizadas, temos uma boa diversidade de soluções. Temos bancos de dados relacionais (PostgreSQL, SQL Server 2012 e SQLite), bancos não relacionais (MongoDB e Hive), arquivos CSV foram usadas no trabalho que implementou o *framework* que simulava um DW, Liu e Qi (2012) não informou a tecnologia usada e, em Abrahão e Souza (2021), os dados foram armazenados em planilhas excel.

Por último, em relação às tecnologias usadas para construir o DW, vemos que o Python e suas bibliotecas estão presentes em três dos trabalhos. Kryeziu et al. (2019) fez uso dos sistemas *SQL Server Integration Services* (SSIS) e *SQL Server Analysis Services* (SSAS), enquanto Ngo, Khac e Kechadi (2020) não especificou o que foi usado para extrair os dados brutos de suas origens, apenas informou que os dados foram armazenados no banco Cassandra antes de serem levados para o DW. O trabalho de Abrahão e Souza (2021) ficou com status de não se aplica (N/A) já que não utilizou de DW no desenvolvimento da solução e o trabalho de Liu e Qi (2012) não informou sobre as tecnologias usadas. Agapito, Zucco e Cannataro (2020) fizeram uso do Python, Pandas e também de ferramentas usadas para fazer raspagem em páginas da HTML na internet (*Beautiful Soup e Requests*).

## 4 Materiais e Métodos

Neste capítulo são apresentadas as etapas de desenvolvimento do projeto<sup>4</sup> até a apresentação dos resultados encontrados. Inicialmente, foi apresentada a maneira de como a base de dados original estava organizada e alguns detalhes sobre ela. O processo de ETL foi descrito a partir da apresentação dos filtros, transformações e ferramentas utilizadas. Posteriormente, foi apresentada a descrição da modelagem do *Data Warehouse*, as tabelas criadas e suas descrições. Por último, foram apresentados os resultados encontrados a partir de análises no DW construído.

### 4.1 Sobre os dados

A Pesquisa de Posse e Hábitos de Uso de Equipamentos de Uso Residencial (PPH-2019) foi realizada pela Eletrobrás, em parceria com o Procel, entre 2018 e 2019. A pesquisa teve o objetivo de obter informações sobre a conscientização, hábitos e posse de equipamentos elétricos nas residências. O estudo contou com uma amostra de 18.775 entrevistas, que foram aplicadas em 26 estados e no Distrito Federal.

O questionário da pesquisa era composto por 411 perguntas e resultou em uma base com pouco mais de 2300 colunas, onde cada linha representa uma das entrevistas. É possível notar que o número de colunas é significativamente maior que o número de perguntas do questionário. Isso se deve ao fato de que muitas perguntas podem assumir múltiplas respostas. Com isso, a base tem uma coluna para cada possível resposta, quando necessário. Também há casos de perguntas que se referem a meses do ano ou horários, o que fez com que houvesse uma coluna referente a cada mês e horário.

A base de dados da pesquisa é apresentada como uma planilha do Excel<sup>5</sup>, onde existem três abas. A aba “Banco de Dados” apresenta os resultados da pesquisa, no formato descrito anteriormente, onde a maioria das colunas é representada por um código (ex.: “P10.9.6”) e os valores das variáveis estão representados com números. A aba “Texto

<sup>4</sup><https://github.com/LasaroAlmeida/LasaroAlmeidaTCC>

<sup>5</sup><https://eletrobras.com/pt/AreasdeAtuacao/PPH%202019%20-%20Banco%20de%20Dados%20V2.xlsx>

das perguntas” possui uma coluna que se refere ao número da pergunta (representa o número da pergunta mas não possui a letra “P” como na aba de “Banco de Dados”) e uma para a descrição, que contém os textos das perguntas no questionário. Já a aba “Texto das alternativas” possui uma linha para cada coluna da primeira planilha, indicando o que cada uma das colunas representa e uma tradução do que as respostas numéricas significam. Ex.: A variável classe social é apresentada como a coluna “CLASSE” e seus possíveis valores são do 1 ao 6, onde 1 significa A1, 2 a classe B1, 3 a classe B2, ... e 6 as classes D e E.

## 4.2 Extração, Transformação e Carga

A primeira etapa do processo de Extração Transformação e Carga (*Extract, Transform, and Load* - ETL) se deu pelo entendimento dos dados por meio da exploração da base e pela leitura do formulário que foi utilizado na pesquisa. Desta forma, foi possível definir quais seriam as variáveis de interesse para a construção do trabalho. Dentre elas, temos variáveis relacionadas às características do domicílio, consumo de energia elétrica mensal, conscientização, classe social, município e eletrodomésticos. A pesquisa também coletou informações sobre domicílios comerciais, mas esses dados não foram selecionados como um dos alvos de análise neste trabalho. Assim, os dados analisados neste trabalho e referentes a eletrodomésticos são somente da parte residencial dos domicílios.

Com base em notícias e nos sites da Eletrobrás e Procel, foi verificado que 18.775 entrevistas haviam sido realizadas, porém a base de dados encontrada possui 27.826 registros. Com base na coluna “ENTREVISTA”, que contém um identificador único para as entrevistas, foi possível observar que haviam entrevistas repetidas e que havia um número bem maior de respostas vazias para essas linhas de mesmo identificador. Também foi possível observar que a base de dados possuía uma coluna chamada “LOOP” e não foi encontrada nenhuma documentação que indicasse o que esta coluna significasse. A coluna podia assumir valores inteiros entre 1 e 9 e foi observado que a quantidade de linhas em que seu valor era 1 resultava em 18.775. Para os demais valores desta variável, o identificador da entrevista passava a ter algum identificador já existente. Considerou-se, então, que essas linhas onde “LOOP” possuía o valor 1 eram as linhas originais da pesquisa, e

essas foram selecionadas para a realização deste trabalho.

Realizadas as primeiras definições, iniciou-se a construção dos códigos de ETL utilizando a linguagem Python e a biblioteca para análise e manipulação de dados, Pandas. Foi utilizado o Google Colab como plataforma de desenvolvimento do trabalho, uma plataforma online para desenvolvimento no formato de Jupyter Notebooks. A base de dados da pesquisa foi lida, as colunas que não seriam usadas foram removidas e as linhas foram filtradas conforme a variável “LOOP” citada anteriormente. Manualmente, foram criadas duas planilhas de parâmetros para auxiliar na filtragem e nomenclatura das colunas. A primeira planilha de parâmetros contém o nome da coluna na base de dados, o novo nome que terá no *Data Warehouse*, o nome da tabela que essa coluna vai fazer parte e, caso seja uma coluna referente a eletrodoméstico, há também uma coluna que indica qual é o eletrodoméstico referenciado. A segunda planilha contém o nome da primeira coluna referente a um eletrodoméstico (geralmente, a primeira coluna se refere à quantidade daquele equipamento) e o nome do eletrodoméstico a que se refere.

Para a nomenclatura das colunas do *Data Warehouse*, foi necessária a criação de uma maneira de colocar nomes pequenos nas colunas e que pudessem representar o texto das perguntas da pesquisa. Desta forma, foi criado um vocabulário controlado, onde muitas palavras foram abreviadas em siglas únicas, de forma que palavras com mais de 5 letras fossem representadas com 4 letras ou 2, em alguns casos. Por exemplo, a palavra “eletrodoméstico” passou a ser “ETDC”. A nomenclatura das colunas seguiu a tradução pelo vocabulário e a separação entre as palavras utilizou “\_”.

Antes de serem inseridos no banco de dados, os dados passaram por várias transformações. A principal delas foi o mapeamento dos valores presentes nas colunas para o seu significado em texto, como foi citado na Seção 4.1. Com base nas planilhas de parâmetros, os dados foram selecionados de acordo com a tabela que fariam parte no banco de dados. Os dados relacionados às tabelas de dimensões foram inseridos primeiro no banco e, só depois, nas tabelas fato.

Para a inserção dos dados das tabelas fato, foi necessária a inclusão de novas linhas nesse grupo de colunas. Onde havia uma coluna para cada mês do ano, foi criada uma linha referente a cada mês, assim as informações de meses poderiam estar dispostas

em somente uma coluna. Em relação ao conjunto de colunas referentes a eletrodomésticos, temos 62 conjuntos de colunas, um para cada eletrodoméstico. Foi necessário transformar esses conjuntos em linhas e criar um conjunto de colunas padrão que pudesse conter todos os dados. Entre essas colunas, também havia um caso onde o eletrodoméstico apresentava uma informação para cada mês do ano. Dito isso, cada linha referente aos eletrodomésticos foi transformada em 73 linhas (uma para cada eletrodoméstico e 12 para o caso onde havia informações sobre os meses do ano). Assim, cerca de 600 colunas e as 18.775 linhas referentes a eletrodomésticos se transformaram em 29 colunas (incluindo as chaves estrangeiras para relacionamento com outras tabelas) e 1.370.575 linhas. Já para as colunas de consumo de energia elétrica mensal, onde havia uma coluna para cada mês, cada linha foi separada em 12 (uma para cada mês), assim apenas uma coluna precisaria ser usada para representação da informação, gerando uma tabela com 225.300 linhas.

### 4.3 Modelagem do *Data Warehouse*

O *Data Warehouse* foi construído utilizando o banco de dados PostgreSQL, um sistema gerenciador de banco de dados relacional gratuito e de código aberto. O banco de dados foi instalado no *cluster* do Google Colab, que utiliza o sistema Linux Ubuntu 22.04 e permite a instalação de ferramentas via linha de comando. Para que o Python pudesse acessar e manipular o banco de dados, foi usada a biblioteca SQLAlchemy.

A modelagem do *Data Warehouse* teve como ponto de partida o projeto apresentado por ARAGÃO (2023). Inicialmente, foram adicionadas novas colunas às tabelas propostas por ele e criadas novas tabelas relacionadas a eletrodomésticos. O projeto original continha apenas seis tabelas e utilizava um modelo dimensional em estrela, com uma tabela fato. No entanto, devido à necessidade de inclusão de novas tabelas e colunas, bem como ao aumento da complexidade, o modelo evoluiu para uma arquitetura composta por dez tabelas, definindo-se como um modelo multidimensional de constelação de fatos (DIANA; KARDIANA, 2015). Esse modelo é caracterizado por tabelas fato que compartilham tabelas de dimensão. No presente trabalho, são utilizadas duas tabelas fato que compartilham cinco tabelas de dimensão e o modelo está apresentado na Figura 4.1.

A seguir, apresentamos algumas informações sobre as tabelas usadas:

- **D\_DMCL** (Dimensão Domicílio): Armazena diversas informações relacionadas às características das residências, como a quantidade de banheiros, automóveis, motocicletas, tipo de pavimentação da rua, área construída, número de moradores, entre outros. Serve como uma fonte primária para a construção de filtros quando relacionada às tabelas fato. Está presente como chave estrangeira nas tabelas fato existentes no projeto.
- **D\_CTZC** (Dimensão Conscientização): Contém diversas informações relacionadas à conscientização, escolaridade e alguns hábitos relacionados à economia de energia elétrica. Nesta tabela, temos dados como a escolaridade do chefe da família, se há preferência por equipamentos que consomem menos energia, se evitam equipamentos em *stand by*, se evitam ligar as lâmpadas durante o dia, se recebem informações sobre como economizar energia elétrica, etc. Em conjunto com a tabela de Fato Consumo Energia Elétrica Mensal, podemos responder questões sobre hábitos de consumo para os domicílios que mais consomem energia.
- **D\_CLSS\_SOCL** (Dimensão Classe Social): Possui as possíveis classes econômicas e sociais às quais os moradores dos domicílios podem pertencer. É uma tabela importante para a construção de filtros em consultas relacionadas a indicadores sociais.
- **D\_MNCP** (Dimensão Município): Guarda as informações do nome do município, região e sigla do estado a que pertence. Pode ser usada como um filtro para a geração de informações regionais.
- **D\_DT\_MES** (Dimensão Data Mês): Tabela usada para análises temporais dos dados. Armazena a lista de meses do ano e também os valores “N/A”, para casos onde a análise temporal não se aplica, e “Indefinido” para casos onde não foi coletado o mês em que determinado evento aconteceu.
- **D\_ETDC** (Dimensão Eletrodoméstico): Contém a lista com os 62 eletrodomésticos apresentados na pesquisa. É uma tabela importante para a realização de filtros sobre características de um eletrodoméstico e vários outros.

- **D\_FRQC\_USO\_ETDC** (Dimensão Frequência Uso Eletrodoméstico): Nesta tabela, temos as possíveis frequências de uso que um eletrodoméstico pode ter. Entre elas, temos as frequências intensa, média, raramente, não utiliza, “N/A” para os casos onde um eletrodoméstico não tem informações de frequência de uso, e vazio caso o eletrodoméstico tenha frequência de uso mas não houve resposta.
- **D\_HRAR\_USO\_ETDC** (Dimensão Horário Uso Eletrodoméstico): Tabela que contém os possíveis horários de uso de um eletrodoméstico, de 0h até 23h. Também temos uma linha que indica o uso eventual e com as colunas vazias para indicar que um eletrodoméstico não tem informações de horário de uso.
- **F\_CSMO\_EE\_MSAL** (Fato Consumo Energia Elétrica Mensal): Relaciona-se com as dimensões de data mês, classe social, município, conscientização e domicílio, apresentando a quantidade de energia, em kWh, consumida por mês. Com base nesta tabela, é possível realizar o relacionamento entre as dimensões citadas e extrair diversas informações, contendo ou não a quantidade de energia elétrica consumida.
- **F\_ETDC** (Fato Eletrodoméstico): É a maior tabela do *Data Warehouse*, relacionando-se com todas as tabelas de dimensão apresentadas. É responsável por registrar diversas informações observadas sobre os eletrodomésticos, como tipo, quantidade, capacidade de armazenamento, tempo que possui o aparelho, tempo de uso diário, possui relacionamento com a frequência de uso e os horários de utilização, entre outros.

Tabela 4.1: Tabelas presentes no DW e seus números de linhas

Tabela	Número de linhas
d_cls_socl	6
d_frqc_uso_etdc	9
d_dt_mes	14
d_etdc	62
d_mncp	166
d_hrar_uso	5.539
d_dmcl	11.457
d_ctzc	16.024
f_csmo_ee_msal	225.300
f_etdc	1.370.575

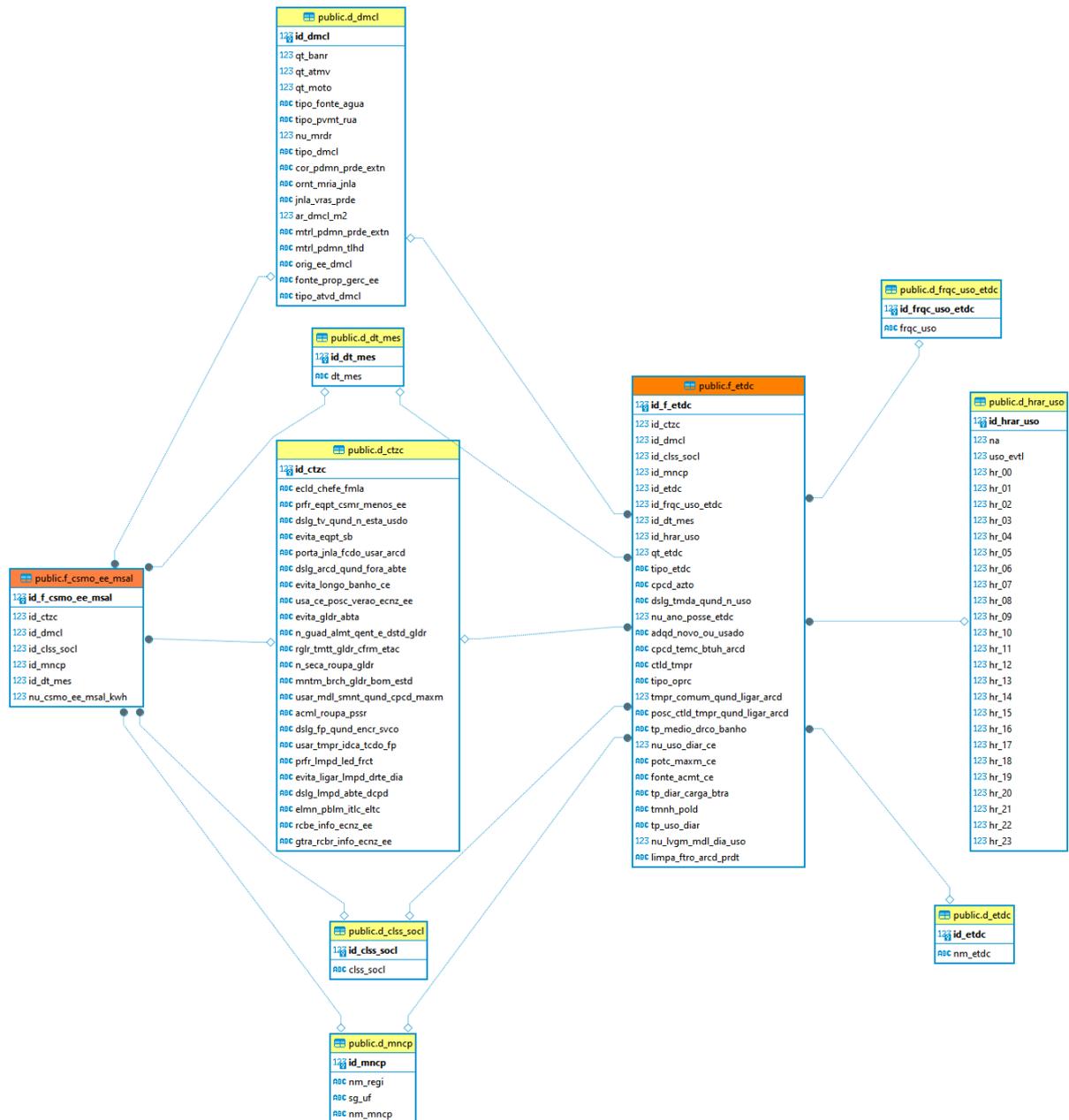


Figura 4.1: Diagrama de relacionamento entre as tabelas do *Data Warehouse*.

De acordo com a Figura 4.1, podemos ver que as tabelas fato F\_CSMO\_EE\_MSAL e F\_ETDC (em laranja) compartilham algumas tabelas de dimensão (em amarelo). Em relação ao relacionamento entre as tabelas, é possível notar que as linhas que as conectam possuem uma extremidade preenchida e outra não. A extremidade preenchida indica que cada registro da tabela de dimensão está relacionado a vários registros da tabela fato. Por outro lado, quando a extremidade do relacionamento não é preenchida, significa que vários registros da tabela fato se relacionam a um único registro da tabela de dimensão.

## 4.4 Resultados

Para o desenvolvimento das análises, foram considerados somente os domicílios cuja energia elétrica era proveniente da rede geral de distribuição, desconsiderando, assim, os casos onde a energia era gerada no próprio domicílio, mista (rede geral de distribuição e gerada no próprio domicílio), e também os casos onde o entrevistado optou pela opção de “não sabe/não responde”. Dentre os 18.775 domicílios entrevistados, 99,48% informaram usar a energia proveniente da rede geral de distribuição.

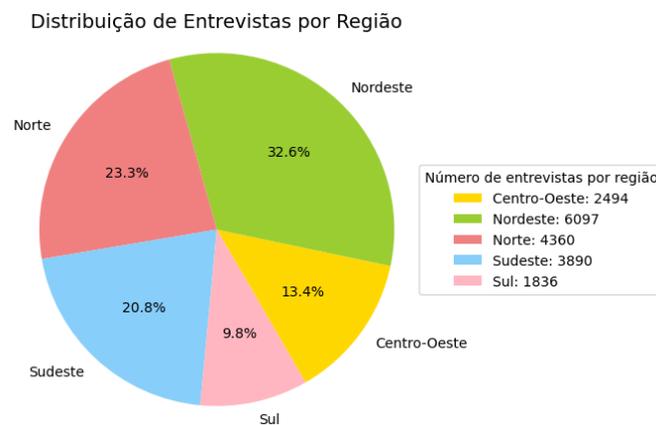


Figura 4.2: Número de entrevistas por região.

O gráfico de setores da Figura 4.2 apresenta como as entrevistas estão distribuídas por região geográfica do país. É possível notar que a região Nordeste concentrou o maior número de entrevistas, se comparada às demais. Já a região Sul foi a região com a menor participação absoluta na pesquisa.

O gráfico da Figura 4.3 apresenta a distribuição da quantidade de entrevistas por classe social. A partir do gráfico, podemos observar que uma grande porcentagem das pessoas, que participaram da entrevista, pertenciam a classes sociais mais baixas. Mais de 50% dos participantes pertenciam as classes C2 ou D-E e, apenas 2,5% à classe A.

Distribuição de Entrevistas por Classe Social

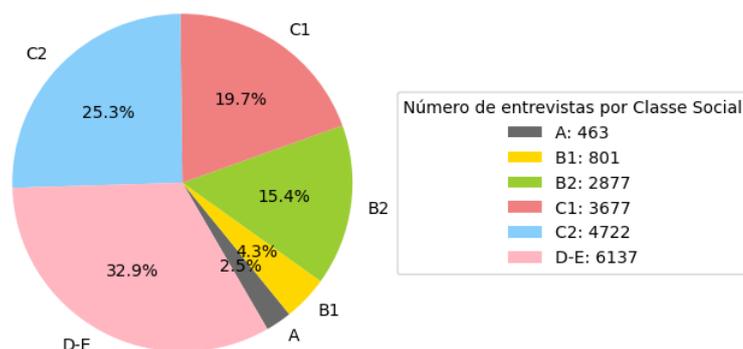


Figura 4.3: Número de entrevistas por classe social.

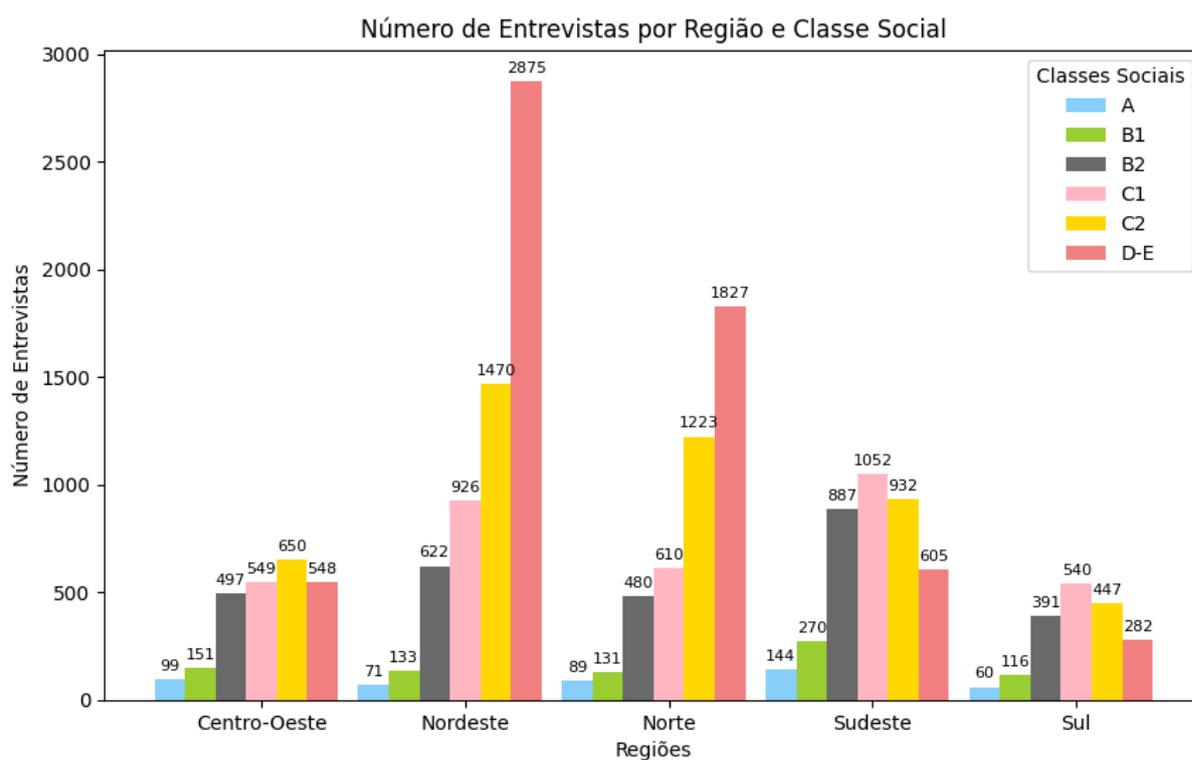


Figura 4.4: Número de Entrevistas por Região e Classe Social

De acordo com a Figura 4.4, é apresentada a distribuição do número de entrevistas por região e classe social. É possível notar que a Região Nordeste apresentou um número consideravelmente mais alto de entrevistas para as classes D-E em comparação com as demais classes da região e também em comparação com as demais regiões.

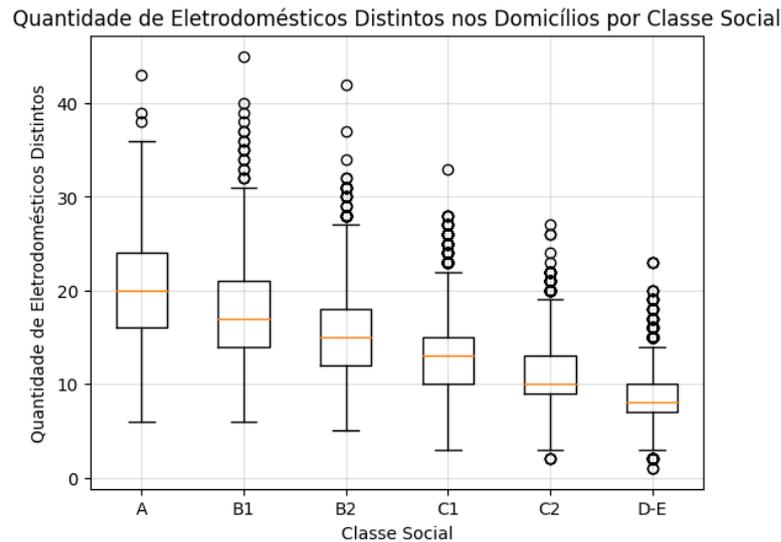


Figura 4.5: Quantidade de eletrodomésticos distintos nos domicílios por classe social.

O Boxplot da Figura 4.5 apresenta, em seu eixo X, as classes sociais e econômicas e, no eixo Y, o número de eletrodomésticos diferentes que um domicílio possuía no momento da entrevista. É possível notar que o valor da mediana da quantidade de eletrodomésticos distintos na classe A é mais de duas vezes o valor da mediana da classe D-E. Todas as classes apresentaram valores discrepantes em relação ao intervalo de análise, conhecidos como *outliers*. Todas as classes sociais apresentaram *outliers* maiores que o intervalo típico. Entretanto, as classes C2 e D-E apresentaram também valores discrepantes abaixo do intervalo analisado, indicando que alguns dos domicílios possuíam menos de 5 eletrodomésticos diferentes.

Por meio deste gráfico, é possível notar que o número de eletrodomésticos em uma residência está diretamente ligado à classe social e econômica dos moradores. Isso pode estar relacionado ao fato de que famílias com maior poder aquisitivo têm a possibilidade de adquirir um maior número de eletrodomésticos para a melhoria de seu conforto, enquanto residências de classes sociais mais baixas tendem a ter menos eletrodomésticos e, principalmente, aqueles que são mais necessários (ver Tabela 4.2). Assim, a diferença entre o número de eletrodomésticos entre as classes sociais torna-se uma métrica capaz de representar a desigualdade social e econômica existente no país.

Tabela 4.2: Top 10 Eletrodomésticos mais comuns por classe social

Porcentagem de domicílios, por classe, que possuem os eletrodomésticos											
A		B1		B2		C1		C2		D-E	
Eleto.	%	Eleto.	%	Eleto.	%	Eleto.	%	Eleto.	%	Eleto.	%
Chuveiro	100,0	Televisor	100,0	Chuveiro	100,0	Chuveiro	100,0	Chuveiro	100,0	Chuveiro	100,0
Refrigerador	100,0	Chuveiro	100,0	Refrigerador	99,0	Refrigerador	99,0	Refrigerador	98,0	Refrigerador	96,0
Televisor	100,0	Refrigerador	99,0	Televisor	99,0	Televisor	98,0	Televisor	97,0	Televisor	95,0
M. L. Roupas	97,0	Celular	96,0	Celular	96,0	Celular	96,0	Celular	95,0	Celular	93,0
Celular	95,0	M. L. Roupas	95,0	M. L. Roupas	92,0	Liquidificador	88,0	Liquidificador	84,0	Ventilador	77,0
Micro-ondas	93,0	Liquidificador	90,0	Liquidificador	89,0	M. L. Roupas	86,0	Ventilador	78,0	Liquidificador	76,0
Liquidificador	91,0	Micro-ondas	88,0	Micro-ondas	81,0	Ventilador	75,0	M. L. Roupas	68,0	Ferro a seco	38,0
Notebook	79,0	Notebook	72,0	Ventilador	74,0	Micro-ondas	67,0	Ferro a seco	45,0	M. L. Roupas	33,0
Modem	71,0	Modem	70,0	Modem	61,0	Ferro a seco	46,0	Micro-ondas	41,0	DVD/Video/Blu-Ray	21,0
Ventilador	64,0	Ventilador	69,0	Notebook	53,0	Modem	45,0	Modem	31,0	Conversor digital	20,0

Por meio da Tabela 4.2, podemos observar os 10 eletrodomésticos mais comuns nos domicílios de cada classe social. O cálculo foi baseado na quantidade de domicílios que possuíam cada eletrodoméstico, expressando a porcentagem em relação ao total de domicílios da respectiva classe social. Para a construção da tabela, os nomes de alguns eletrodomésticos foram abreviados devido à limitação de espaço: “Modem” refere-se a “Modem para internet com função Roteador Wi-Fi (pertencente ao provedor de internet)”, “M. L. Roupas” representa “Máquina de Lavar Roupas”, “Ventilador” corresponde a “Ventilador ou Circulador de ar”, e “Conversor Digital” refere-se a “Conversor digital externo para TV aberta”. É possível observar que “Chuveiro”, “Refrigerador” e “Televisor” são os três eletrodomésticos mais comuns em todas as classes sociais, estando presentes em 100% dos domicílios de classe A. Somente o chuveiro está presente em todos os domicílios das classes D-E. Nota-se também que os aparelhos celulares ocupam a quinta posição na classe A e a quarta posição nas demais classes. Além disso, alguns aparelhos, como “micro-ondas” e “modens”, estão presentes nas classes de A a C2, mas nem sequer aparecem entre os 10 primeiros na classe D-E. Um ponto interessante é que o ar-condicionado não figura entre os 10 eletrodomésticos mais comuns em nenhuma das classes sociais.

Com base na Figura 4.6, é possível observar que o número de eletrodomésticos em um domicílio de uma determinada classe social está relacionado à média anual de consumo de energia elétrica para essa classe social. Os *outliers*, valores que se destacam significativamente em relação à maioria dos dados analisados, representam domicílios que demandam maior atenção em relação ao consumo de energia elétrica.

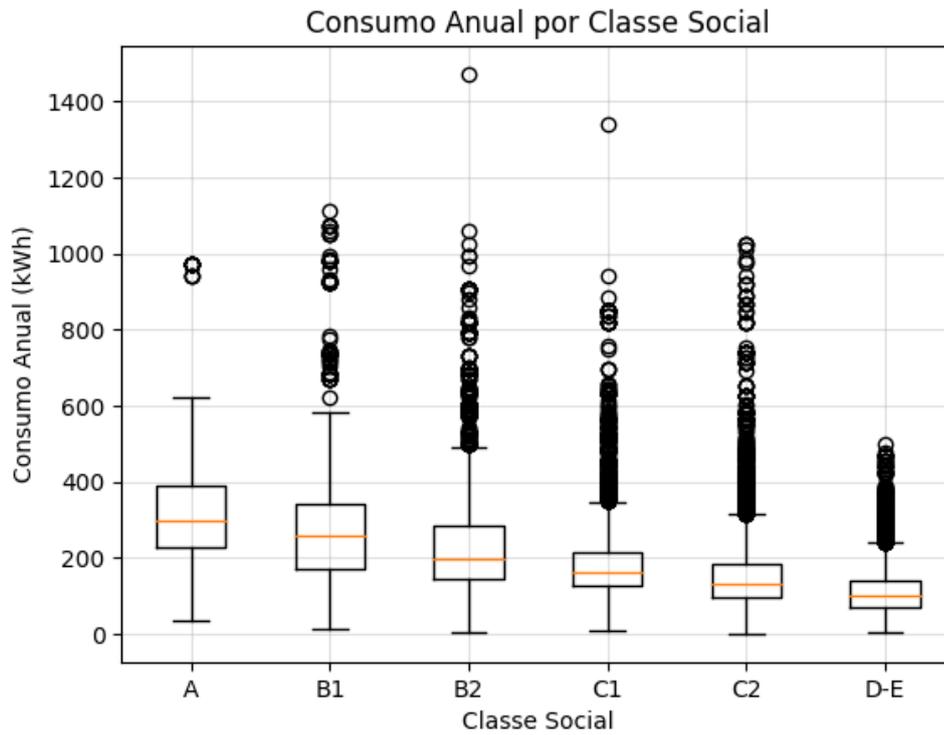


Figura 4.6: Consumo anual por classe social.

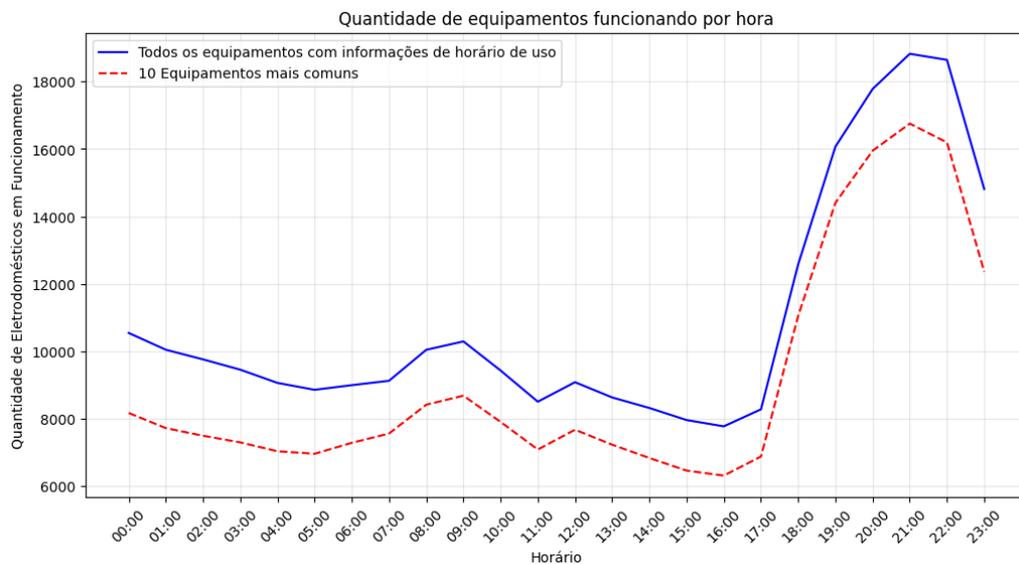


Figura 4.7: Quantidade de eletrodomésticos em funcionamento por hora.

Na Figura 4.7, apresentamos um gráfico de linhas que mostra as quantidades de eletrodomésticos em funcionamento por hora ao longo do dia no país. Na base de dados utilizada, apenas 20 dos 62 eletrodomésticos tiveram informações coletadas sobre

o horário de utilização, desconsiderando as lâmpadas. Para a linha representada em vermelho (linha tracejada), foram selecionados os 10 eletrodomésticos mais comuns entre esses 20 com dados de horário de uso. São eles: “Aparelho de som/rádio (ligados na tomada quando utilizados)”, “Chuveiro Elétrico”, “Computador”, “Ferro elétrico a seco”, “Máquina de Lavar Roupas”, “Notebook”, “Televisor”, “Ventilador de teto”, “Ventilador ou Circulador de ar” e “Videogame”.

De acordo com a linha azul (linha contínua), que se refere aos horários de uso de todos os 20 equipamentos que possuem esse dados, é possível notar que o pico de equipamentos em funcionamento ocorre entre às 17:00 e 23:00 horas. Isso é válido, já que é de conhecimento comum que esse é um período em que grande parte das pessoas chega do trabalho, toma banho, prepara refeições e assiste à TV ou outro entretenimento antes de dormir. Outros dois momentos de crescimento da curva são entre as 06:00 e 09:00 horas, quando as pessoas estão acordando e se arrumando para ir ao trabalho ou para a escola, e das 11:00 às 13:00 horas, horário em que, em geral, as pessoas param para almoçar e, também, é o período em que as crianças chegam em casa depois da escola. Observando o indicador para os eletrodomésticos mais comuns vemos que seu comportamento é o mesmo do comportamento total.

Uma análise específica sobre os horários de uso para os equipamentos de climatização, responsáveis por controlar a temperatura do ambiente em que estão inseridos, também foi realizada e é apresentada na Figura 4.8. Entre os equipamentos com informações de horário de uso e que são do tipo climatização, incluem-se o ar-condicionado, aquecedor de ambiente, ventilador de teto e ventilador ou circulador de ar. É possível observar que a maior utilização desses equipamentos ocorre entre as 17:00 e 08:00 horas, ficando ativos cerca de 15h das 24h horas do dia. Este período coincide com o momento em que as pessoas acabaram de chegar do trabalho ou estão se preparando para ele, sendo também o momento em que as pessoas estão dormindo e desejam um ambiente confortável em termos de temperatura.

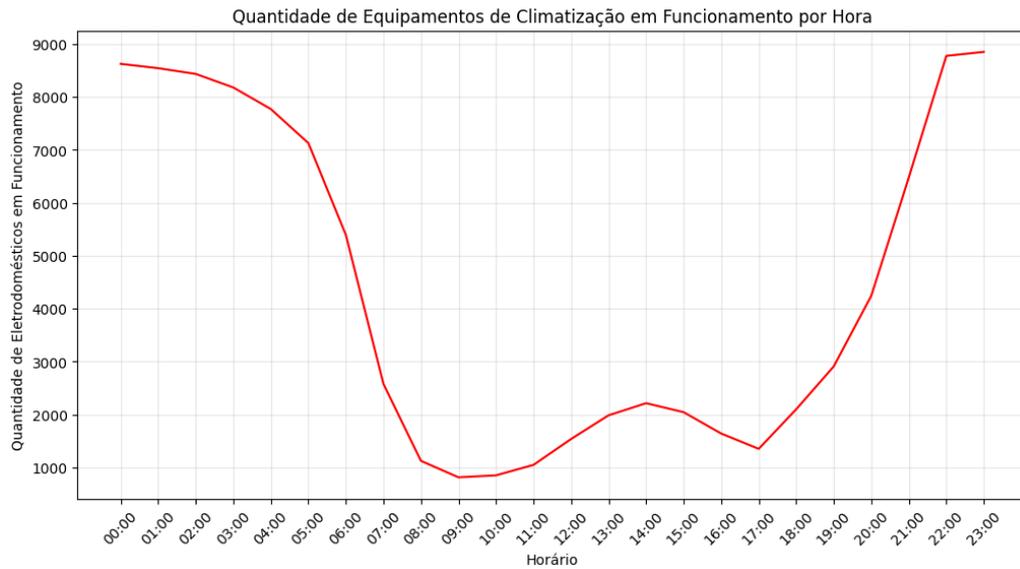


Figura 4.8: Quantidade de eletrodomésticos de climatização em funcionamento por hora

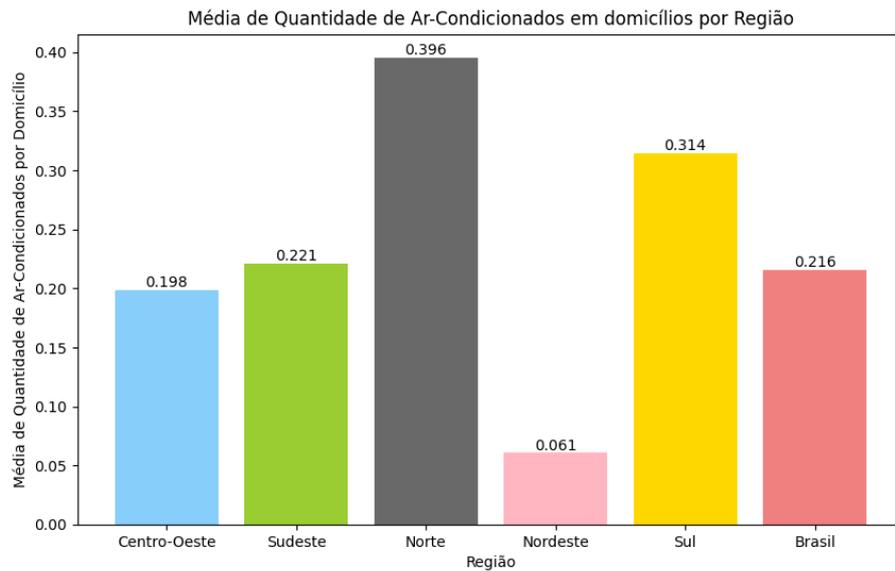


Figura 4.9: Média da quantidade de ar-condicionados em domicílios por região.

O gráfico de barras da Figura 4.9 apresenta a quantidade média de aparelhos de ar-condicionado em domicílios para cada região do país. O ar-condicionado foi selecionado como alvo para essa análise por ser um equipamento que possui um custo consideravelmente alto para a aquisição e que consome bastante energia elétrica, resultando em um aumento das contas de energia. É possível observar no gráfico que a região com o menor número de aparelhos de ar-condicionado por domicílio é a Nordeste, que apresenta cerca

de 1 aparelho para cada 16 domicílios, mesmo sendo uma das regiões com as maiores temperaturas do país, como a região Norte. Já a região Norte se destacou por ser a região com o maior número de aparelhos por domicílio, chegando a uma marca de 1 aparelho para cada 3 domicílios. A segunda região com o maior número de aparelhos foi a região Sul e, isso pode estar relacionado com o uso de certos aparelhos de ar-condicionado para o aquecimento de ambientes, já que a região apresenta as menores temperaturas médias do país.

A quantidade de aparelhos de ar-condicionado em um domicílio pode estar relacionada com a classe social, como indicada o gráfico da Figura 4.10. É possível observar que para a classe A, a média de aparelhos é maior do que um aparelho por domicílio, e essa quantidade caí de acordo com o poder aquisitivo de cada classe. A classe D-E apresentou uma média de aproximadamente 1 aparelhos para cada 17 residências.

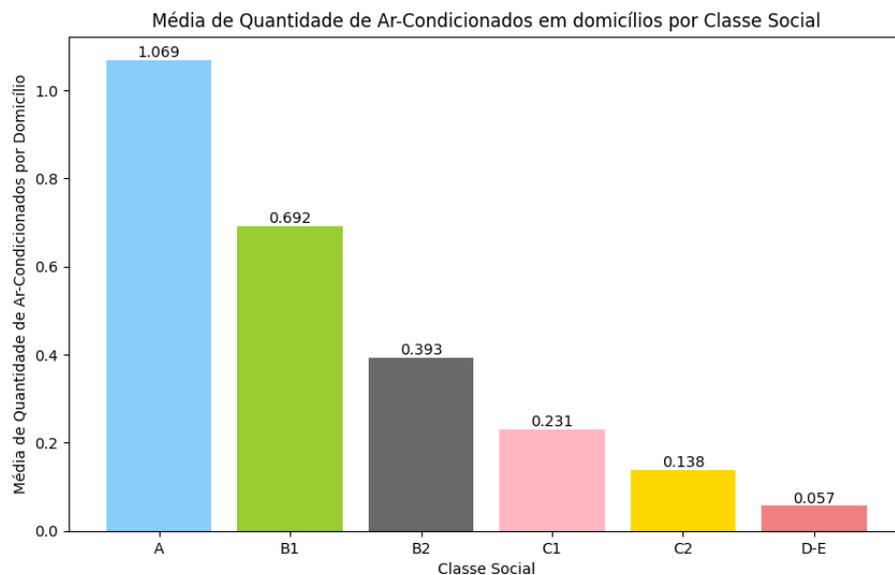


Figura 4.10: Média da quantidade de ar-condicionados em domicílios por classe social.

Por meio dessas informações, é possível ver os impactos da desigualdade social e da distribuição de renda no país. Segundo Salim et al. (2021), as regiões Norte e Nordeste possuem os menores Índices de Desenvolvimento Humano (IDH) do país e os dados mostram que a região Nordeste é a que apresentou o menor número de aparelhos de ar-condicionado, corroborando com Salim et al. (2021). Isso também poder estar relacionado ao grande número de entrevistas aplicadas em classes sociais mais baixas na

região, conforme a Figura 4.4, fazendo com que os resultados sejam mais tendenciosos a apresentarem as características dessas classes.

Muitos são os problemas trazidos pela desigualdade social, e o acesso a aparelhos de ar-condicionado é apenas um deles. A falta desses aparelhos em domínios de classes sociais mais baixas pode resultar em problemas de desconforto térmico e noites mal dormidas para os moradores, especialmente em regiões de temperaturas extremas (frias ou quentes). O motivo da região Norte ter se mostrado como a região com o maior número de aparelhos por domicílio, mesmo sendo uma das regiões com os menores IDHs e maiores temperaturas, pode estar associado a fatores regionais e culturais. Uma explicação pode ser o pequeno número de entrevistas aplicadas sobre a classe A na região, fazendo com que os dados representem algo que pode não ser a realidade.

Tabela 4.3: Tradução dos nomes das perguntas da Tabela 4.4

Id Pergunta	Texto Original
1	Prefere equipamentos que consomem menos energia elétrica
2	Desliga a TV quando não está usando
3	Evita deixar equipamentos em <i>Stand By</i> .
4	Portas e janelas fechadas quando usa o ar-condicionado
5	Desliga o ar-condicionado quando está fora do ambiente
6	Evita longos banhos no chuveiro elétrico
7	Usa chuveiro elétrico na posição verão para economizar energia
8	Evita deixar a geladeira aberta
9	Não guarda alimentos quentes e destampados na geladeira
10	Regula termostato da geladeira conforme as estações do ano.
11	Não seca roupa atrás da geladeira
12	Montém a borracha da geladeira em bom estado
13	Usa máquina de lavar somente quando capacidade máxima
14	Acumula roupa para passar
15	Desliga ferro de passar quando encerra o serviço (entre peças)
16	Usa temperatura indicada por tecido no ferro de passar
17	Prefere lâmpada LED ou fluorescentes
18	Evita ligar lâmpadas durante o dia
19	Desliga lâmpadas em ambientes desocupados
20	Elimina problemas na instalação elétrica

Tabela 4.4: Distribuição de Respostas para Práticas de Economia de Energia

Id _ Nome coluna DW	Sempre(%)	Normalmente(%)	Raramente(%)	Nunca(%)	N/A(%)
1_prfr_eqpt_csmr_menos_ee	75.64	18.69	4.49	0.91	0.27
2_dslg_tv_qund_n_esta_usdo	66.04	24.13	6.91	1.48	1.43
3_evita_eqpt_sb	58.71	24.68	10.09	4.44	2.08
4_porta_jnla_fcdo_usar_arcd	72.57	20.02	2.05	0.64	4.72
5_dslg_arcd_qund_fora_abte	69.87	21.53	2.5	1.25	4.84
6_evita_longo_banho_ce	55.02	33.57	7.43	1.32	2.66
7_usa_ce_posc_verao_ecnz_ee	63.13	22.65	8.72	3.06	2.44
8_evita_gldr_abta	63.53	23.63	4.92	5.87	2.04
9_n_guad_almt_qent_e_dstd_gldr	58.32	23.69	6.48	10.34	1.18
10_rglr_tmmt_gldr_cfrm_etac	49.07	25.0	12.93	11.24	1.77
11_n_seca_roupa_gldr	52.28	19.18	7.11	20.08	1.35
12_mntm_brch_gldr_bom_estd	68.55	20.87	5.04	4.29	1.25
13_usar_mdsl_smmt_qund_cpdc_maxm	50.3	18.25	2.62	1.67	27.16
14_acml_roupa_pssr	36.07	21.03	7.44	2.74	32.73
15_dslg_fp_qund_encr_svco	28.29	23.88	9.29	2.32	36.22
16_usar_tmpr_idca_tcdo_fp	37.62	17.78	6.41	4.43	33.75
17_prfr_lmpd_led_frct	61.56	28.66	6.86	1.53	1.39
18_evita_ligar_lmpd_drte_dia	70.13	24.69	3.92	0.76	0.49
19_dslg_lmpd_abte_dcpd	72.15	23.17	3.67	0.66	0.35
20_elmn_pblm_itlc_eltc	74.7	17.82	4.21	1.92	1.35

Para a geração da Tabela 4.4, foram excluídos os dados onde as respostas às perguntas estavam vazias.

Segundo o site da CEMIG<sup>6</sup> (Companhia Energética de Minas Gerais), existem diversas ações simples que podemos tomar dentro de nossas casas para economizar energia elétrica. A Tabela 4.4 apresenta 20 perguntas abordadas na pesquisa e que se referem a essas ações e hábitos de conscientização e economia de energia, com a porcentagem de cada uma das cinco possíveis respostas.

Como base na tabela, é possível observar que o hábito mais comum entre os entrevistados é a preferência por equipamentos (eletrodomésticos e lâmpadas) que consumem menos energia. Por outro lado, a pergunta com o menor percentual de respostas “sempre”, é sobre regular a temperatura do termostato da geladeira conforme as estações

<sup>6</sup><https://www.cemig.com.br/usina-do-conhecimento/>

do ano, uma ação simples, mas que não é sempre seguida.

Todos esses hábitos contribuem significativamente para eficiência energética dentro dos domicílios e, conseqüentemente, na redução dos valores das contas de energia. Por meio da análise das métricas de adoção desses hábitos, percebe-se a necessidade de uma maior aceitação e valorização dos consumidores acerca das práticas de economia de energia

## 4.5 Limitações

Os dados da pesquisa PPH-2019 são os mais recentes em relação ao cenário de hábitos de consumo de energia elétrica no Brasil. No entanto, dois grandes eventos ocorreram desde a realização da pesquisa e podem ter mudado o cenário em estudo: a pandemia da COVID-19 e uma crise hídrica.

Durante a pandemia da COVID-19, as medidas de isolamento social fizeram com que muitas pessoas passassem a trabalhar de suas casas, aumentando o consumo residencial. Apesar disso, houve uma redução significativa no consumo de energia elétrica no país devido às restrições nos setores de comércio e indústria (SILVA, 2023). Entre 2020 e 2021, o Brasil enfrentou uma de suas maiores crises hídricas, como mencionado anteriormente neste trabalho, o que levou à ativação de usinas termelétricas para suprir a demanda nacional. Isso resultou em maiores custos de geração de energia no momento em que o consumo residencial estava elevado devido à pandemia, levando muitas pessoas a adotarem hábitos de economia de energia para reduzir os custos (SOLFUS, 2022).

Um problema encontrado diz respeito aos dados de residências que exerciam alguma atividade comercial. Como a pesquisa inclui dados referentes a eletrodomésticos das residências com ou sem atividade comercial, é difícil distinguir se um equipamento citado para a parte comercial é o mesmo da parte residencial. Por isso, este trabalho se concentrou nos dados exclusivamente da parte residencial dos domicílios.

Por último, os dados alvo deste trabalho foram referentes aos domicílios cuja fonte de energia era exclusivamente a rede geral de distribuição. Isso se deve ao fato de que não há informações sobre as demais fontes de energia, como solar, eólica, combustíveis fósseis e mista, não indicando qual a participação de cada uma no consumo residencial final.

## 5 Conclusões e Trabalhos Futuros

Os resultados apresentados neste trabalho são simples se comparados à capacidade que os *Data Warehouses* possuem quando utilizados, porém, conseguem alcançar os objetivos definidos. A modelagem de dados por meio do modelo dimensional permite o armazenamento e relacionamento de dados de maneira simples, facilitando o desenvolvimento de consultas e manipulações sobre os dados. Mesmo não sendo o único modelo para o armazenamento de dados, devido aos seus benefícios, os DWs e seu modelo dimensional são a base utilizada em arquiteturas e sistemas mais modernos.

Informações como as apresentadas por este trabalho podem ser utilizadas para a tomada de decisão em diversos setores. Empresas de distribuição de energia podem entender melhor os hábitos de consumo e a posse de eletrodomésticos de seus consumidores, usando essas informações para auxiliar na tomada de decisão em projetos de melhoria de seus serviços e entender como o poder aquisitivo de seus clientes impacta em seus consumos. O setor público pode utilizar esses dados de consumo relacionados às classes sociais para a criação de políticas de incentivo para a aquisição de equipamentos domésticos pelas classes sociais mais baixas, incentivar que indústrias produzam equipamentos mais eficientes no consumo de energia por preços acessíveis, além da criação de campanhas de conscientização apresentando ações e hábitos que as pessoas podem seguir para a eficiência energética. A população, em posse dessas informações, também pode tomar suas próprias decisões e realizar ações para a redução de seu consumo de energia e, conseqüentemente, dos custos atrelados a ela.

Com base neste trabalho e em tudo que foi discutido, é possível citar alguns pontos que poderiam ser usados em trabalhos futuros para melhores análises e interpretações. São eles:

- buscar por fontes que apresentem os consumos médios de cada eletrodoméstico presente nesta pesquisa para a geração de informações relacionadas ao consumo durante os horários do dia;

- buscar por informações de pesquisas anteriores ou posteriores, e semelhantes a essa, para a geração de informações sobre a mudança de padrões e hábitos de consumo ao longo dos anos;
- modelar e adicionar ao DW as colunas referentes aos eletrodomésticos da parte não residencial dos domicílios;
- verificar a possibilidade e necessidade de modelar tabelas de dimensão relacionadas às lâmpadas, devido à presença de um grande número de colunas referentes a esse assunto na base, tratando da quantidade de lâmpadas por cômodo e também dos tipos dessas lâmpadas (LED, fluorescente, etc);
- buscar também por informações de consumo energético dos setores comerciais e industriais e realizar a comparação com o setor residencial;
- realizar novas pesquisas de hábitos de uso, com menos perguntas e com alvos bem definidos. Um exemplo seria tratar dos hábitos e ações que poderiam ser tomadas para a redução do consumo de energia em prédios públicos;
- pesquisar sobre o uso de fontes de energia limpa, como energia solar, e sua participação no consumo domiciliar;
- cruzar informações deste trabalho com informações do CENSO mais recente.

## Bibliografia

ABRAHÃO, K. C. d. F. J.; SOUZA, R. G. V. d. Estimativa da evolução do uso final de energia elétrica no setor residencial do Brasil por região geográfica. *Ambiente Construído*, SciELO Brasil, v. 21, p. 383–408, 2021.

AGAPITO, G.; ZUCCO, C.; CANNATARO, M. Covid-warehouse: A data warehouse of Italian COVID-19, pollution, and climate data. *International Journal of Environmental Research and Public Health*, v. 17, n. 15, 2020. ISSN 1660-4601. Disponível em: <https://www.mdpi.com/1660-4601/17/15/5596>.

AHMED, I. *Modelagem dimensional: Visão Geral, benefícios E dicas de design*. 2024. Disponível em: <https://www.astera.com/pt/knowledge-center/dimensional-modeling-guide/>.

ARAGÃO, L. N. *Uso de Data Warehouse para Análise do Consumo de Energia no Brasil*. Tese (Doutorado) — Federal University of Juiz de Fora, 2023. Available at <http://monografias.ice.ufjf.br/tcc-web/tcc?id=836>.

BATISTA, F. *Sistemas de Apoio A decisão em grupo (SADG)*. Medium, 2016. Disponível em: <https://medium.com/@flailtonbatista/sistemas-de-apoio-a-decis%C3%A3o-em-grupo-sadg-d2be2f2a4d54#:~:text=Os%20componentes%20de%20um%20SAD,usu%C3%A1rio%20ou%20gerenciador%20de%20di%C3%A1logo.>

BOEHM, S.; SCHUMER, C. *10 Conclusões do relatório do IPCC sobre Mudanças Climáticas de 2023*. 2023. Disponível em: <https://www.wribrasil.org.br/noticias/10-conclusoes-do-relatorio-do-ipcc-sobre-mudancas-climaticas-de-2023#:~:text=Em%20um%20cen%C3%A1rio%20t%C3%A3o%20intensivo,de%20tr%C3%AAs%20milh%C3%B5es%20de%20anos.>

BRAZ, L. *Energia Renovável: O Paradoxo da Sobra de Energia Limpa no país*. 2024. Disponível em: <https://brasil61.com/n/energia-renovavel-o-paradoxo-da-sobra-de-energia-limpa-no-pais-bras2411167>.

BRIGNOL, W. S.; AZEVEDO, R. M. de; CANHA, L. N.; FARIAS, C.; BERTINETI, D.; NETO, N. K. Technical-commercial management to increase the participation of micro and mini generation in the Brazilian energy sector. In: *2017 52nd International Universities Power Engineering Conference (UPEC)*. IEEE, 2017. Disponível em: <http://dx.doi.org/10.1109/UPEC.2017.8231938>.

CARREGOSA, L. *Oferta de Energia cresce mais que consumo, e Brasil “joga fora” excesso; Entenda*. 2024. Disponível em: <https://g1.globo.com/economia/noticia/2024/03/04/oferta-de-energia-cresce-mais-que-consumo-e-brasil-joga-fora-excesso-entenda.gh.html>.

CASAGRANDE, C. G.; RODRIGUES, A. C. C.; GUIMARÃES, T. C. d. C. Smart home IoT: Uma análise acerca da redução do consumo de energia elétrica residencial com dispositivos smart acessíveis. *VERNÁCULA - Territórios Contemporâneos*, v. 1, n. 1, out. 2023. Disponível em: <https://periodicos.univag.com.br/index.php/Vernacula/article/view/2305>.

- CASTOR, E. C. S.; FERNANDES, A. L.; LAIA, M. M. de; MENEZES, L. M. B. de; AMORIM, M. C. Sistemas de apoio à decisão. *P2P E INOVAÇÃO*, 2022. Disponível em: [⟨https://api.semanticscholar.org/CorpusID:252674170⟩](https://api.semanticscholar.org/CorpusID:252674170).
- CINTRA, L. A. *O que É o efeito estufa e Como Ele Influencia as Mudanças Climáticas?* 2023. Disponível em: [⟨https://globorural.globo.com/clima/noticia/2023/11/o-que-e-o-efeito-estufa-e-que-como-ele-influencia-as-mudancas-climaticas.ghtml⟩](https://globorural.globo.com/clima/noticia/2023/11/o-que-e-o-efeito-estufa-e-que-como-ele-influencia-as-mudancas-climaticas.ghtml).
- DIANA, N. E.; KARDIANA, A. Comprehensive centralized-data warehouse for managing malaria cases. *International Journal of Advanced Computer Science and Applications*, v. 6, 2015. Disponível em: [⟨https://api.semanticscholar.org/CorpusID:2478612⟩](https://api.semanticscholar.org/CorpusID:2478612).
- EDP. 2023. Disponível em: [⟨https://empresas.edp.com.br/blog/o-que-e-o-consumo-consciente/⟩](https://empresas.edp.com.br/blog/o-que-e-o-consumo-consciente/).
- ENGIE. *O que É Eficiência Energética e quais São Seus Benefícios?* 2022. Disponível em: [⟨https://www.alemdaenergia.engie.com.br/o-que-e-eficiencia-energetica/⟩](https://www.alemdaenergia.engie.com.br/o-que-e-eficiencia-energetica/).
- FERNANDES, W. Legal regulation of the distribution of electricity in brazil. In: \_\_\_\_\_. *Energy Law and Regulation in Brazil*. Springer International Publishing, 2018. p. 137–158. ISBN 9783319734569. Disponível em: [⟨http://dx.doi.org/10.1007/978-3-319-73456-9\\_7⟩](http://dx.doi.org/10.1007/978-3-319-73456-9_7).
- FERREIRA, J.; MIRANDA, M.; ABELHA, A.; MACHADO, J. O processo etl em sistemas data warehouse. In: SN. *INForum*. [S.l.], 2010. p. 757–765.
- GEOGRAPHIC, N. *Mudanças climáticas: Como O Aquecimento Global Afeta a vida no Brasil*. 2022. Disponível em: [⟨https://www.nationalgeographicbrasil.com/ciencia/2022/02/mudancas-climaticas-como-o-aquecimento-global-afeta-a-vida-no-brasil⟩](https://www.nationalgeographicbrasil.com/ciencia/2022/02/mudancas-climaticas-como-o-aquecimento-global-afeta-a-vida-no-brasil).
- GOEL, A. *Star, snowflake, and Fact Constellation(Galaxy Schema) schemas*. Medium, 2023. Disponível em: [⟨https://medium.com/@adityagoel2397/star-snowflake-and-fact-constellation-galaxy-schema-schemas-a6434dc5325e⟩](https://medium.com/@adityagoel2397/star-snowflake-and-fact-constellation-galaxy-schema-schemas-a6434dc5325e).
- KIMBALL, R.; ROSS, M. *The data warehouse toolkit: the complete guide to dimensional modeling*. [S.l.]: John Wiley & Sons, 2011.
- KRYEZIU, N.; ISMAILI, F.; AJDARI, J.; RAUFI, B.; ZENUNI, X. Energy provider data warehouse design and implementation - case study. In: *2019 International Conference on Information Technologies (InfoTech)*. [S.l.: s.n.], 2019. p. 1–5.
- LEGNAIOLI, S. *O que são os gases do Efeito Estufa?* 2024. Disponível em: [⟨https://www.ecycle.com.br/gases-do-efeito-estufa/⟩](https://www.ecycle.com.br/gases-do-efeito-estufa/).
- LIU, G.-x.; QI, Z.-h. The application of data warehouse in decision support system. In: *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*. [S.l.: s.n.], 2012. p. 1373–1376.
- MACEDO, H. R.; MACEDO, H. R.; SANTOS, C. S.; SOUZA, K. T. d.; MOURA, R. F. d. Expansão e perspectivas do sistema de transmissão brasileiro após a crise energética de 2001. In: *5ª JICE-JORNADA DE INICIAÇÃO CIENTÍFICA E EXTENSÃO*. [S.l.: s.n.], 2014.
- MACHADO, F. N. R. *Tecnologia e projeto de Data Warehouse*. [S.l.]: Saraiva Educação SA, 2004.

MACKENZIE. *Crise Hídrica no Brasil: Quais São as Causas e Os Impactos?* 2022. Disponível em: <https://blog.mackenzie.br/vestibular/atualidades/crise-hidrica-no-brasil-quais-sao-as-causas-e-os-impactos/>.

MALAR, J. P. *Energia Elétrica Tem Alta Acumulada de Quase 25% EM 2021, Diz Ibge*. 2021. Disponível em: <https://www.cnnbrasil.com.br/economia/energia-eletrica-tem-alta-acumulada-de-quase-25-em-2021-diz-ibge/>.

MARTÍN, M. *Alertas ignorados, Seca e Eleições: A receita ideal para o Desastre Hídrico*. 2015. Disponível em: [https://brasil.elpais.com/brasil/2015/02/09/politica/1423507719\\_304859.html](https://brasil.elpais.com/brasil/2015/02/09/politica/1423507719_304859.html).

NGO, V. M.; KHAC, N. A. L.; KECHADI, M. T. Data warehouse and decision support on integrated crop big data. *International Journal of Business Process Integration and Management*, Inderscience Publishers, v. 10, n. 1, p. 17, 2020. ISSN 1741-8771. Disponível em: <http://dx.doi.org/10.1504/IJBPIM.2020.113115>.

PAULA, G. B. de; MEDEIROS, L. de; SALAS, C. S. S. Study of the opex benchmarking methodology of pantanal sul-mato-grossense. In: *2018 Simposio Brasileiro de Sistemas Eletricos (SBSE)*. IEEE, 2018. Disponível em: <http://dx.doi.org/10.1109/SBSE.2018.8395799>.

SALIM, T. R.; ANDRADE, T. M.; KLEIN, C. H.; OLIVEIRA, G. M. M. d. Idh, recursos tecnológicos e humanos para diagnóstico e tratamento das malformações do aparelho circulatório no brasil. *Arquivos Brasileiros de Cardiologia*, SciELO Brasil, v. 117, p. 63–71, 2021.

SILVA, J. P. Efeitos da covid-19 no setor elétrico brasileiro. 2023.

SOLFUS. *Pandemia impacta no consumo de energia do Brasileiro*. 2022. Disponível em: <https://solfus.com.br/pandemia-impacta-no-consumo-de-energia-do-brasileiro/>.

SPUTNIK. *Matriz Elétrica brasileira Passa de 200 GW com 84% da produção a partir de Fontes renováveis*. Sputnik Brasil, Rosiya Segodnya 25260, 2024. Disponível em: <https://sputniknewsbr.com.br/20240307/matriz-eletrica-brasileira-passa-de-200gw-com-84-da-producao-a-partir-de-fontes-renovaveis-334495.html>.

TALAGALA, N. *Data as the new oil is not enough: Four principles for avoiding data fires*. Forbes Magazine, 2022. Disponível em: <https://www.forbes.com/sites/nishatalagala/2022/03/02/data-as-the-new-oil-is-not-enough-four-principles-for-avoiding-data-fires/?sh=422ba7d6c208>.

THIVES, L. P.; GHISI, E.; JÚNIOR, J. J. T. Regional inequalities in electricity access versus quality of life in brazil. *Ambiente Construído*, FapUNIFESP (SciELO), v. 22, n. 3, p. 47–65, set. 2022. ISSN 1415-8876. Disponível em: <http://dx.doi.org/10.1590/s1678-86212022000300609>.

TORANZO, B. *Data Warehouse: O que É, para que serve e Como Funciona*. 2022. Disponível em: <https://www.insper.edu.br/noticias/data-warehouse-o-que-e-para-que-serve-e-como-funciona/>.

ZENDESK. *O que É etl? Conheça as 7 principais ferramentas!* 2023. Disponível em: <https://www.zendesk.com.br/blog/o-que-e-etl/>.

---

ÖZTÜRK, E. G.; ROCHA, P.; RODRIGUES, A. M.; Soeiro Ferreira, J.; LOPES, C.; OLIVEIRA, C.; NUNES, A. C. D3s: Decision support system for sectorization. *Decision Support Systems*, v. 181, p. 114211, 2024. ISSN 0167-9236. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167923624000447>).