

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

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

Leonardo Nunes Aragão

JUIZ DE FORA
DEZEMBRO, 2023

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

LEONARDO NUNES ARAGÃO

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
em Ciência da Computação

Orientador: Victor Stroele de Andrade Menezes

JUIZ DE FORA
DEZEMBRO, 2023

USO DE DATA WAREHOUSE PARA ANÁLISE DO CONSUMO DE ENERGIA NO BRASIL

Leonardo Nunes Aragão

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

Aprovada por:

Victor Stroele de Andrade Menezes
Doutor em Engenharia de Sistemas e Computação pela UFRJ

Regina Maria Maciel Braga Villela
Doutor em Engenharia de Sistemas e Computação pela UFRJ

Mário Antônio Ribeiro Dantas
Doutor em Ciência da Computação pela University of Southampton (UK)

JUIZ DE FORA
12 DE DEZEMBRO, 2023

Resumo

Ultimamente a preocupação com o consumo de energia elétrica e como seus gastos impactam diretamente na vida dos consumidores tem aumentado. Para iniciar estudos nessa área existe uma grande necessidade de dados, principalmente aqueles que são confiáveis e bem organizados. Ao trabalhar com uma base de dados extensa e pouco organizada, realizar buscas simples podem ser custosas. Neste sentido, criar um Data Warehouse ajuda a resolver esse problema e agilizar os estudos realizadas sobre essa base de dados, sendo essa a proposta desse trabalho. Inicialmente propõe um modelo de arquitetura para criação do Data Warehouse, em seguida mostra os métodos utilizados para extração, limpeza e organização da base de dados, e por fim apresenta análises de alguns perfis de consumo, que exemplificam o uso e apoiem o tema de consumo e economia de energia elétrica no setor residencial.

Palavras-chave: Big data, Data Warehouse, Consumo de energia.

Abstract

Lately, concern about electricity consumption and how its costs directly impact consumers' lives has increased. To begin studies in this area, there is a great need for data, especially those that are reliable and well-organized. When working with an extensive and poorly organized database, simple searches can be costly, creating a Data Warehouse helps solve this problem and speed up the studies carried out on this database, and this is the proposal of this work. Initially, it proposes a model of architecture for the creation of the Data Warehouse, then it shows the methods used for extraction, cleaning, and organization of the database, and finally, it presents analyses of some consumption profiles, which exemplify the use and support the theme of consumption and energy saving in the residential sector.

Keywords: Big data, Data Warehouse, Energy consumption.

Agradecimentos

A minha amada esposa e amigos por todo apoio, aos professores do Departamento de Ciência da Computação pelos seus ensinamentos e aos funcionários da Universidade, que durante esses anos, contribuíram de algum modo para o nosso enriquecimento pessoal e profissional.

*“Só se vê bem com o coração, o essencial
é invisível aos olhos”.*

*Antoine de Saint-Exupéry (O pequeno
Príncipe)*

Conteúdo

Lista de Abreviações	6
1 Introdução	7
1.1 Organização do trabalho	9
2 Trabalhos Relacionados	10
2.1 Designing and Implementing Data Warehouse for Agricultural Big Data . .	10
2.2 Good practices for clinical data warehouse implementation: A case study in France	11
2.3 Data Warehouse to Support Creation of Key Performance Indicators (KPIs) in Electricity Supply	12
2.4 A Data Warehouse Architecture supporting Energy Management of Intel- ligent Electricity System	12
2.5 Estimativa da evolução do uso final de energia elétrica no setor residencial do Brasil por região geográfica	13
3 Materiais e Métodos	14
3.1 Sobre os dados	14
3.2 Criação do Data WareHouse	15
3.2.1 Trabalhando com os dados	16
3.3 O Modelo Dimensional	17
3.3.1 Tabelas dimensão	18
3.3.2 Sobre a tabela de fatos	19
4 Análise dos resultados	20
4.1 Análise do Consumo por Região	20
4.2 Análise do Consumo por Classe Econômica	22
4.3 Análise do Consumo por Quantidade de pessoas	24
4.4 Análise do Consumo por m ²	25
4.5 Análise do Consumo por nível de conscientização	25
4.6 Limitações	28
5 Conclusão e Trabalhos Futuros	30
Bibliografia	31

Lista de Abreviações

DCC Departamento de Ciência da Computação

UFJF Universidade Federal de Juiz de Fora

DW Data Warehouse

ETL Extract, Transform e Load

1 Introdução

Com o acúmulo do aumento do preço da energia chegando a 25% em algumas regiões do Brasil (MALAR, 2021), e com a frequente dificuldade na geração de energia elétrica no país, em 2021 “... o Brasil enfrenta uma das piores estiagens da história, o que ocasionou uma crise hídrica, principalmente na região Sudeste”, relata (BUCCO, 2021). Em meados de 2022, o mundo passou por sua maior crise energética jamais registrada (UNEP, 2022), e em 2023 a matriz energética brasileira sofreu com as mudanças nos regimes de chuvas e secas mais intensas (SCHUCK, 2023). O problema enfrentado pelo Brasil também ocorre em outras regiões do planeta, o custo atribuído a geração de energia elétrica e as tarifas referentes a seu consumo são altas. Reduzir o consumo de energia elétrica e criar hábitos inteligentes no uso de eletrodoméstico e dispositivos eletroeletrônicos pode ser parte da solução, e como consequência ainda ajudar o meio-ambiente.

Segundo o Balanço Energético Nacional 2023 (EPE, 2023), publicado pela Empresa de Pesquisa Energética vinculada ao Ministério de Minas e Energia, o setor residencial é responsável por 10,7% do consumo total da energia gerada no país, e a que por sua vitalidade tem a maior parcela de usuários, o que mostra a relevância do setor.

Diversos estudos estão sendo realizados focados no área de Gerenciamento e Controle do Consumo de energia elétrica. Para realização desses estudos são necessárias fontes com grande quantidade de dados. Por esse motivo, a maioria utiliza o auxílio de ferramentas para análises Big Data como Data Warehouses, alguns desses estudos são abordados mais a frente nesse trabalho. Muitos desses estudos mostram a grande e crescente preocupação sobre o consumo e geração de energia elétrica no planeta, sendo que seus impactos e questões ecológicas também devem ser considerados. A redução do consumo de energia elétrica é uma estratégia eficaz para diminuir a pegada de carbono, além disso, a conscientização sobre o consumo de energia pode ter impactos significativos na economia e na sociedade (NS, 2021). De acordo com (DRUCKMAN; JACKSON, 2008), para elaborar e revisar políticas visadas para uma sociedade com baixa poluição de carbono é necessário a compreensão do consumo de energia elétrica em diferentes categorias de residências.

No cenário encontrado por este Trabalho de Conclusão de Curso, existem poucas fontes de dados sobre o consumo de energia no Brasil, o Programa Nacional de Conservação de Energia Elétrica (Procel) é um dos poucos que fornece esse tipo de informação, mesmo existindo mais de 30 empresas que atuam em serviços de distribuição de energia elétrica (ANEEL, 2022) pelo país. Foi encontrado também ausência de análises e pesquisas que avaliem os perfis de consumo de energia nas residências brasileiras.

Após estudos iniciais, constatou-se a falta de pesquisas na área de consumo e economia de energia no Brasil. Com o intuito de incentivar e facilitar o surgimento de novos estudos que possam identificar, avaliar e descrever os perfis de consumo, é essencial melhorar a qualidade dos dados disponíveis sobre o consumo de energia elétrica no setor residencial brasileiro. Este é o primeiro passo nesse processo.

Assim, este Trabalho de Conclusão de Curso propõe a criação de um Data Warehouse utilizando o modelo dimensional para integrar dados sobre o consumo de energia elétrica nas residências brasileiras. Apesar da existência de outras tecnologias para concentração e distribuição de dados, o Data Warehouse é o método mais indicado quando se deseja definir um modelo padrão para análise dos dados (VAISMAN; ZIMÁNYI, 2014). O objetivo é fornecer uma fonte de dados confiável e acessível para facilitar futuras pesquisas na área, permitindo a condução de análises multidimensionais sobre o setor energético. A solução será avaliada por meio de cenários encontrados durante a pré-avaliação dos dados, como a variação do consumo de energia por região, o consumo de energia por classe econômica, o acesso e interesse de informação sobre energia e redução no consumo, entre outros. O Data Warehouse será construído com base na Pesquisa de Posse e Hábitos de Uso de Equipamentos Elétricos na Classe Residencial de 2019, fornecida pela Eletrobras. Além disso, será demonstrado como o Data Warehouse pode ser utilizado para pesquisas de diversas áreas, via exemplos de análises de perfis de consumo de energia.

Para alcançar o objetivo definido, a metodologia do trabalho foi dividida nas seguintes etapas: (1) revisão da literatura para compreensão dos temas, construção da fundamentação teórica e identificação dos trabalhos relacionados; (2) pré-análise dos dados da pesquisa para definir melhores abordagens e auxiliar o processo de modelagem; (3) criação do esquema do banco de dados e do modelo dimensional; (4) desenvolvimento dos

processos de ETL sobre os dados da pesquisa; (5) definição dos pontos a serem analisados e realização das análises; (6) avaliação descritiva dos gráficos gerado pelas análises.

Esse trabalho apoia-se na metodologia de criação de Data Warehouse do Livro *The Data Warehouse Toolkit* (KIMBALL; ROSS, 2011), que defende que Data Warehouses devem ser projetados para serem compreensíveis e rápidos, e enfatiza o quanto o uso correto da modelagem de dados com o modelo Dimensional traz melhorias de desempenho, velocidade de buscas no banco e simplificação do entendimento das informações.

1.1 Organização do trabalho

Além da introdução, o trabalho está organizado da seguinte forma: os trabalhos relacionados são apresentados no Capítulo 2. Os Materiais e Métodos são encontrados no Capítulo 3. O Capítulo 4 contém as discussões das análises geradas e seus resultados e por último, o Capítulo 5 apresenta a conclusão do trabalho e discute trabalhos futuros.

2 Trabalhos Relacionados

Neste capítulo, são apresentados alguns trabalhos relacionados ao tema proposto neste TCC, que implementam a utilização da tecnologia de Data Warehouse em diversos setores. Com isso, fica claro ao observar a praticidade, versatilidade e importância do uso de Data WareHouses para diversas áreas. Por fim, é apresentado um trabalho que fez uso da mesma pesquisa utilizada de fonte para esse trabalho, porém, com métodos estatísticos e foco na evolução do uso final de energia elétrica no setor residencial do Brasil por região geográfica. É importante ressaltar que atualmente existem poucos trabalhos e estudos realizados na área de consumo de energia elétrica no setor residencial brasileiro.

2.1 Designing and Implementing Data Warehouse for Agricultural Big Data

O artigo (NGO; LE-KHAC; KECHADI, 2019), inicialmente, compara e analisa alguns dos populares DWs de código aberto existentes no contexto do Big Data agrícola. Ao analisar outros trabalhos é verificado que nenhuma das soluções propostas satisfazem o problema de Big Data Agrícola, além disso, em alguns é proposto um "smart framework agrícola", mas não é discutido como construir ou implementar um. Os dados são inconsistentes, ambíguos e extremamente volumosos, por serem coletados de várias fontes heterogêneas, por esse motivo exigem uma série de padrões que não são atendidos pelos trabalhos citados pelo artigo.

Depois desenvolvem e implementam sua própria solução de DW para Big Data agrícola combinando Apache Hive, os bancos de dados não-relacionais MongoDB e Apache Cassandra, e o modelo multidimensional constelação de fatos. Além disso, é realizada uma avaliação experimental para apresentar o desempenho do DW desenvolvido por eles. O esquema foi projetado como uma constelação, sendo flexível para se adaptar a outros conjuntos de dados agrícolas e critérios de qualidade do Big Data agrícola. Finalmente,

por meio de consultas de leitura específicas usando comandos populares de HQL/SQL, o artigo mostra que seu DW consegue de superar o banco MySQL.

Como mostrado, o Data Warehouse com modelo multidimensional é uma abordagem altamente eficaz, contudo, não foi escolhido o modelo multidimensional devido à proposta do trabalho, foi cogitado utilizar o modelo de dados Snowflake (LEVENE; LOIZOU, 2003) caso fosse adicionado os eletrodomésticos e suas muitas categorias existentes na pesquisa.

2.2 Good practices for clinical data warehouse implementation: A case study in France

O artigo (DOUTRELIGNE et al., 2023) discute sobre boas práticas de implementação de Data Warehouse Clínico (CWD) na França e faz recomendações focadas a consolidar e aumentar o potencial dos dados de cuidados rotineiros para melhorar os cuidados de saúde. Focando nos tópicos de governança, transparência, tipos de dados, reutilização dos dados, ferramentas e a qualidade dos processos de controle dos dados.

Ele mostra também a importância de tratar os dados conforme os cenários de cada caso e ainda ressalta a importância da padronização desse tipo de dado que pode ser benéfico para diversas organizações e estudos.

”A reutilização de dados de cuidados rotineiros não é gratuita. Deve-se prestar atenção a todo o ciclo de vida dos dados para criar conhecimento robusto e desenvolver inovação” (DOUTRELIGNE et al., 2023).

O artigo também cita a importância de ter dados transparentes e com algumas melhorias a possibilidade de conseguir reutilizar os dados para permitir que diversas áreas se aproveitem dele, o que é um dos pontos-chave almejados por esse TCC.

2.3 Data Warehouse to Support Creation of Key Performance Indicators (KPIs) in Electricity Supply

Em (KRSTEV; KRNETA, 2022) é mostrado como criar, monitorar e analisar indicadores chaves de desempenho (KPIs) no fornecimento de energia com um Data Warehouse usando o modelo dimensional. Também é discutido o funcionamento das KPIs para ajudar em tomadas de decisões e o setor de fornecimento de energia. O Data warehouse foi criado utilizando SQL Server Database Engine tool e o sistema de *Business Intelligence*(BI) foi feito utilizando também a tecnologia SQL Server Data Tools da empresa Microsoft.

Nesse trabalho, não foram utilizados os softwares citados da Microsoft ou demais softwares proprietários de outras empresas. Ao invés disso, foi dada preferência em utilizar softwares de código aberto, pois essas ferramentas são geralmente projetadas para permitir maior interação com outras ferramentas, além disso, utilizar software código aberto representa uma redução de custos com licenças e burocracia, e ainda permitem a personalização e adaptação do código-fonte quando necessário.

2.4 A Data Warehouse Architecture supporting Energy Management of Intelligent Electricity System

Em (LI et al., 2013/03) é apresentado o problema que a China enfrentou, do processo de industrialização e urbanização se desenvolvendo rapidamente, causando um aumento da demanda de energia elétrica constante. Para solucionar isso, foi proposto um sistema de gerenciamento de energia utilizando um Data Warehouse para guardar os dados relevantes para a pesquisa. O propósito do sistema desenvolvido é armazenar, integrar e analisar datasets complexos de múltiplas fontes de informação.

Concluíram que o sistema proposto provê as ferramentas necessárias para um gerenciamento e monitoramento de energia eficiente, por permitir combinar dados de diferentes fontes. Além disso, implementaram um aplicativo web que representa os dados agregados com a granularidade desejada para alcançar requisitos de usuários diferentes.

Atualmente no Brasil, infelizmente, não existe essa quantidade de dados dis-

poníveis como mostrado na China e a visão de centralizar informações úteis em um só lugar também é uma ideia bem remota em nosso setor de consumo de energia doméstica. Por esse motivo, incentivar pesquisas nessas áreas torna-se necessário, a criação de Data WareHouses é benéfico para diversas áreas de interesse.

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

O Artigo (ABRAHÃO; SOUZA, 2021) analisou a estrutura de consumo de energia elétrica no setor residencial brasileiro por região geográfica, com foco em usos finais, entre os anos de 2005 e 2019. A pesquisa utilizou dados da pesquisa Posse e Hábitos de Uso de Equipamentos Elétricos na Classe Residencial dos anos 2004-2006 e 2018-2019, além de técnicas estatísticas.

Como resultado, identificaram que a estrutura de consumo de energia evoluiu de maneira diferente em cada região do Brasil. Entre as mudanças, as principais evoluções na estrutura estimada foram a ampliação do consumo por uso final dos equipamentos como ar-condicionado, ventilador e aquecedores. Foi identificado também a permanência de uma elevada participação no consumo dos equipamentos geladeira e freezer. Além disso, foi observada uma redução no consumo de energia para aquecimento de água em todas as regiões, exceto na região Sul.

As conclusões do estudo indicam um alto potencial para a implementação de medidas de eficiência energética. No entanto, salientam a importância em considerar as diferenças regionais no padrão de consumo para garantir a efetividade dessas medidas. A pesquisa ressalta a importância de políticas de eficiência energética que levem em conta as especificidades regionais do Brasil.

3 Materiais e Métodos

Para atingir os objetivos propostos e testar a praticidade gerada pelo Data Warehouse, realizou-se um estudo em três fases distintas.

A primeira fase consistiu na análise e organização dos dados brutos da pesquisa, incluindo documentos, planilhas e arquivos no formato CSV. Esta fase também envolveu a limpeza e transformação desses dados para garantir sua qualidade e consistência, com a remoção de colunas e adição de outras

Na segunda fase, procedeu-se à criação do Data Warehouse. Isso envolveu a modelagem dos dados coletados na primeira fase em um formato que facilita a análise e a geração de relatórios. Os dados foram organizados em tabelas de dimensões e fato ¹, permitindo uma análise multidimensional. Além disso, foram implementadas medidas de segurança para garantir a privacidade e a integridade dos dados.

Finalmente, foram desenvolvidos consultas voltadas para os cenários escolhidos inicialmente e Scripts para geração de gráficos que permitem a análise dos dados de maneira eficiente e intuitiva. Além do foco na questão energética, o estudo demonstra a importância da criação de um Data Warehouse, facilitando a extração de informações e auxiliando futuras análises dos dados.

3.1 Sobre os dados

A Pesquisa de Posse e Hábitos de Uso de Equipamentos Elétricos na Classe Residencial de 2019 foi realizada pela Eletrobras em parceria com o Programa Nacional de Conservação de Energia Elétrica (Procel) ² e o Ministério de Minas e Energia. O objetivo da pesquisa era levantar informações sobre a posse e o uso de equipamentos elétricos nas residências brasileiras, bem como os hábitos de consumo e as práticas de eficiência energética dos consumidores.

¹<https://github.com/LeonardoN-Aragao/TCC>

²<https://eletrobras.com/pt/AreasdeAtuacao/PPH%202019%20-%20Question%C3%A1rio%20da%20Pesquisa.pdf>

O arquivo resultante³ foi estruturado em um formato onde cada coluna representa uma variável da pesquisa. Já as entrevistas (questionários aplicados) estão referenciadas às colunas. Em geral, um questionário está totalmente representado em uma única linha. Contém um total de 27.826 linhas (entrevistas), registros que abrangem todas as regiões do País, e 2.319 colunas, que traz separado por categoria questões sobre informações gerais da moradia e da família, posse e uso de equipamentos elétricos, etc.

3.2 Criação do Data Warehouse

Data Warehouse na tradução literal significa "armazém de dados", mas ele não os guarda de qualquer forma. Os dados são armazenados de forma organizada, seguindo estruturas pré-definida que facilitam sua manipulação. Apesar de existirem outras tecnologias que possam ser utilizadas para concentração e distribuição de dados, o Data Warehouse, embora exija um grande trabalho com sua modelagem, é o método mais fácil para extração de dados. Com os dados já organizados, limpos e integrados ainda facilita a reutilização dos mesmo em outros modelos.

O Data Warehouse é criado passando pelas camadas de processamento dos dados, tais como a camada de extração, transformação e carga (ETL), podendo ser adicionadas mais camadas conforme a necessidade do projeto. Cada camada tem uma função específica e requer ferramentas e técnicas adequadas para garantir a qualidade e a consistência dos dados. Essas camadas fazem o processo de integração de dados que pode combinar dado de diversas fontes em um armazenamento de dados único, carregado em um DW ou outro sistema de destino.

A Figura 3.1 apresenta o processo de criação do Data Warehouse e os passos utilizados nesse trabalho. No primeiro passo temos a seleção ou coleta dos dados relevantes da pesquisa fonte; no segundo passo temos o processamento dos dados que será detalhado na seção 3.2.1; no terceiro passo foi realizado a modelagem, integração e armazenamento dos dados do DW, processo detalhado nas seções 3.3.1 e 3.3.2; como último passo foi feita a análise de dados e interpretação dos dados, os resultados podem ser encontrados no capítulo 4.

³<https://eletrobras.com/pt/AreasdeAtuacao/PPH%202019%20-%20Banco%20de%20Dados%20V2.xlsx>

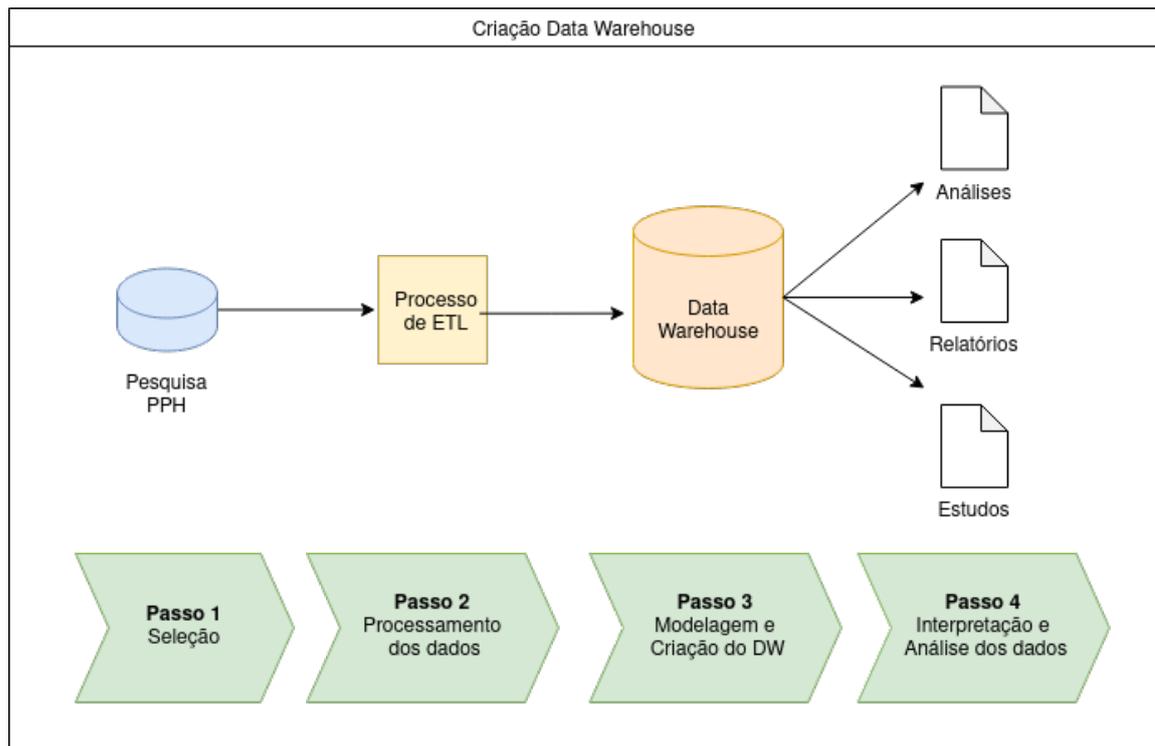


Figura 3.1: Criação do Data Warehouse

3.2.1 Trabalhando com os dados

Para trabalhar sobre uma grande quantidade de dados existem limitações do hardware que devem ser levadas em consideração, como a falta de memória RAM. Assim, realizar uma operação em todas as linhas ou em algumas colunas tem um peso grande e pode ser demorado e custoso. Por esse motivo, antes de iniciar o design do modelo dimensional e organizar os dados para análise foi necessária uma pré-análise do documento com intuito de reduzir informações desnecessárias. Após leitura de todo o documento de questionário da pesquisa ¹ procurando as informações que seriam relevantes para as análises foram removidas 530 colunas.

Posteriormente, foi realizado um refinamento para acrescentar legibilidade aos dados. Originalmente, cada coluna da base de dados tinha o nome no padrão 'P. N^o Sessão. N^o Pergunta' (Ex: P5.15.2), os nomes foram alterados para que se torne mais claro e de fácil entendimento. Embora na pesquisa existam as opções, as resposta ou linhas em uma mesma coluna são numéricas de extensão variada entre colunas e seguem o número de alternativas de respostas possíveis no questionário da pesquisa, foi realizada substituição

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

dos valores por textos correspondentes as possíveis alternativas de cada coluna (Ex: 1: Sim, 2: Não, 9: Não sabe/não responde).

Foi efetuada a operação de remover as informações de consumo mensal e transformadas em duas novas colunas "Mês" e "Consumo" visando facilitar a geração do Data Warehouse e da Dimensão Data. Essas operações foram implementadas com algoritmo no formato dividir para conquistar para atender os limites do hardware utilizado, o procedimento removeu os consumos mensais apresentados inicialmente em linha e gerou uma nova linha para cada mês do ano, com isso o novo número de linhas da base de dados após todo o processo foi de 333.912 linhas e 1.791 colunas.

A escolha de utilizar Softwares de código aberto possibilitou a resolução de um problema encontrado durante avaliação dos dados. Para facilitar a geração do esquema do banco de uma tabela tão grande utilizando o SQLite foi preciso modificar parâmetros no código-fonte e recompilar o código do programa, que não permitia carregar um arquivo .csv contendo a quantidade de colunas existentes no arquivo-fonte.

Para o processo de ETL o código foi implementado utilizando a linguagem Python v3.8.10 com o auxílio da biblioteca de análise de dados Pandas e Jupyter Notebook. Foi utilizado para implementação do banco de dados o SQLite, visualização e manipulação o Software DBeaver Community Edition, gráficos gerados com uso da biblioteca Matplotlib e foram executados em um processador AMD Ryzen TM 5 3400G 4x 3.7GHz, 16 GiB memória RAM, 8 GiB de memória Swap, no sistema operacional Linux Mint 20.2.

3.3 O Modelo Dimensional

O Modelo Dimensional criado por Ralph Kimball é uma metodologia de modelagem de dados que permite representar informações de negócios em um formato fácil de entender e acessar. Ele define os processos de negócios em tabelas fatos (dimensões de medida) e tabelas dimensões que contêm informações descritivas sobre os dados modelados nas tabelas fato, que podem ser usadas para filtrar e agrupar os fatos. As colunas contidas nas tabelas fato e tabelas dimensão são nomeadas a partir de nomenclaturas padronizadas e de fácil entendimento.

3.3.1 Tabelas dimensão

As tabelas dimensão são as que armazenam os atributos descritivos dos dados, usadas normalmente para filtrar e agrupar os fatos nas consultas. Essas tabelas devem ter uma coluna de chaves identificadoras para cada linha podendo também serem relacionadas com uma ou mais tabelas fato por meio desses identificadores como chave estrangeira, nesse trabalho as tabelas são relacionadas com somente uma tabela fato. As dimensões apresentadas a seguir foram selecionadas para otimizar as análises que levantaram interesse durante a pré-avaliação dos dados da pesquisa e alcançar o objetivo desejado por esse trabalho. No que diz respeito às dimensões:

- **Dimensão Data** - é uma estrutura que armazena informações sobre os meses para facilitar a análise temporal dos dados, permitindo filtrar, agrupar e comparar registros por diferentes períodos, e.g., "Energia consumida por uma residência específica no mês de janeiro"
- **Dimensão Classe** - armazena informações sobre a Classe Econômica de cada família, permitindo a visualização dos resultados por indicadores sociais, e.g., "Preocupação com o tempo de banho por classe econômica"
- **Dimensão Região** - armazena informações sobre região, estado e cidade, permitindo organizar e analisar os dados de diferentes perspectivas geográficas, e também comparar os registros entre diferentes regiões, e.g., "Média de consumo de energia por Região Brasileira"
- **Dimensão Residência** - armazena informações sobre tipo, área construída, quantidade de pessoas que moram na residência, material utilizado na construção, etc., permitindo assim analisar a situação de moradia da família que se encontra num determinado tipo de residência, e.g., "Número de residências de classe econômica B que sejam feitas de alvenaria"
- **A Dimensão de Conscientização** - armazena informações sobre a consciência e os hábitos energéticos, permitindo conhecer mais sobre quais hábitos que geram impacto no consumo de energia, medido por escolaridade, a escolha de comprar

produtos econômicos, se o tempo de banho é uma preocupação e a regulação do controle de temperatura da geladeira..., e.g. "Porcentagem de residências da classe econômica A que se preocupam com o tempo de banho"

3.3.2 Sobre a tabela de fatos

A tabela de fatos armazena os fatos que são métricas geradas pelas dimensões. Para isso, ela deve guardar as chaves estrangeiras das tabelas dimensionais. Para poder realizar as análises desejadas foi criada uma Tabela Fato Consumo onde um fato nessa tabela representa o consumo(kWh) individual de uma residência por mês. Os grãos que formam a granularidade são: mensal, regional, classe econômica, residência, e informações de consciência energética, relacionado ao consumo(KWh) de energia. A Figura 3.2 apresenta a estrutura do modelo dimensional proposto.

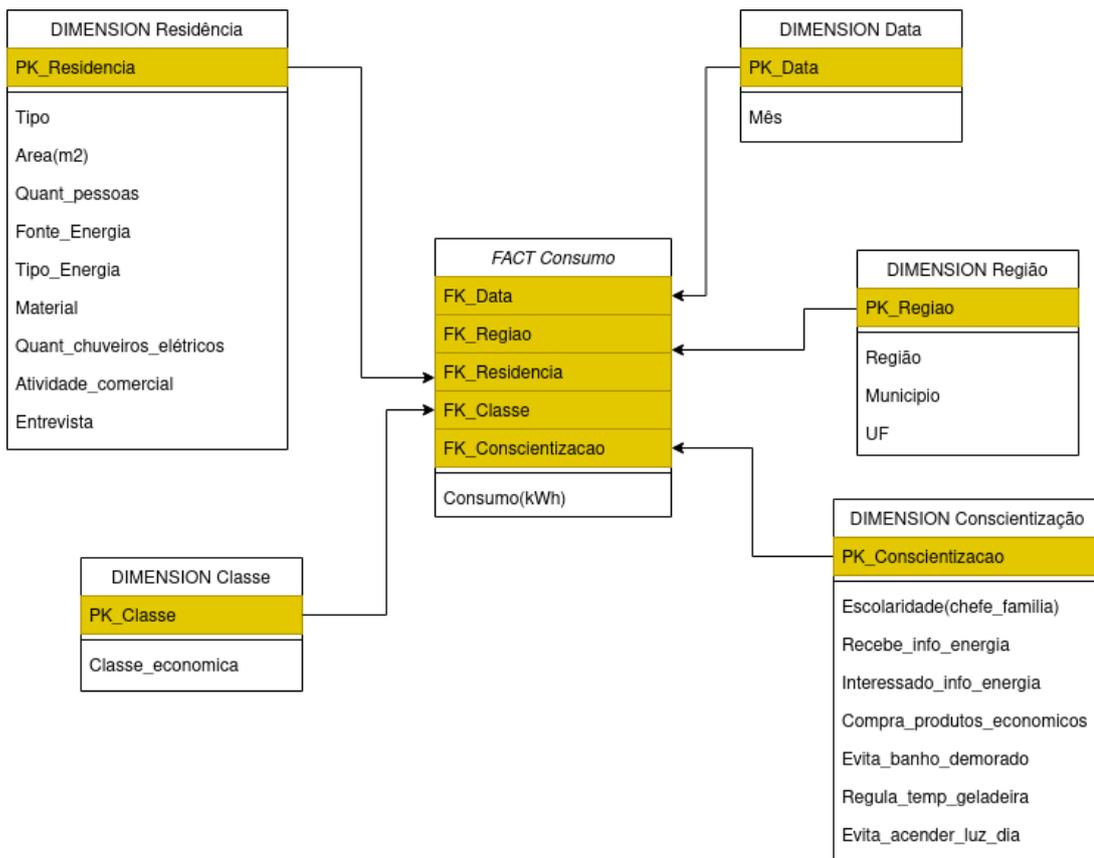


Figura 3.2: Modelo Dimensional

4 Análise dos resultados

Para a geração das análises manuais foram removidos das consultas todos os fatos cujas residências com o atributo consumo fossem menores que 10 KWh, tivessem valores nulos ou com campo vazio; processo detalhado na Seção 3.2.1. Além disso, conforme explicado na Seção 4.6, os fatos cuja dimensão de Residência possuía o atributo “Tipo de energia” com valor “Gerada em domicílio” ou “Mista” também foram excluídos das consultas. Isso resultou na remoção de todos os fatos onde as limitações da pesquisa poderiam introduzir erros ou informações falsas. Para o gráfico da Figura 4.7, todos os fatos que não continham informações sobre a área da residência foram removidos.

4.1 Análise do Consumo por Região

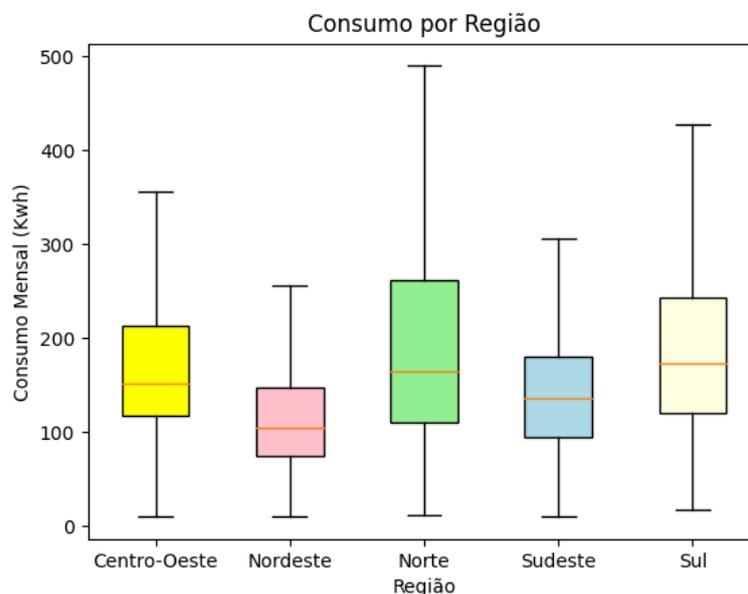


Figura 4.1: Consumo mensal por residência por Região

O Boxplot da Figura 4.1 apresenta em seu eixo X as Regiões Brasileiras e no eixo Y o Consumo Mensal em kWh. O gráfico da figura mostra que o maior consumo mensal está na região Norte do país (na cor bege), enquanto o menor consumo mensal está na região Nordeste (em verde). É possível observar que as regiões com maior faixa de área litorânea

e com maior densidade populacional possuem uma média de gasto energético menor do que as outras regiões, mesmo considerando que as metrópoles são maiores consumidoras de energia do que as áreas rurais e menos industrializadas. Isso não significa que as pessoas nessas cidades sejam menos eficientes ou conscientes no uso da energia. Esse fenômeno pode estar relacionado a fatores econômicos e culturais que influenciam o uso da eletricidade nas diferentes regiões do país.

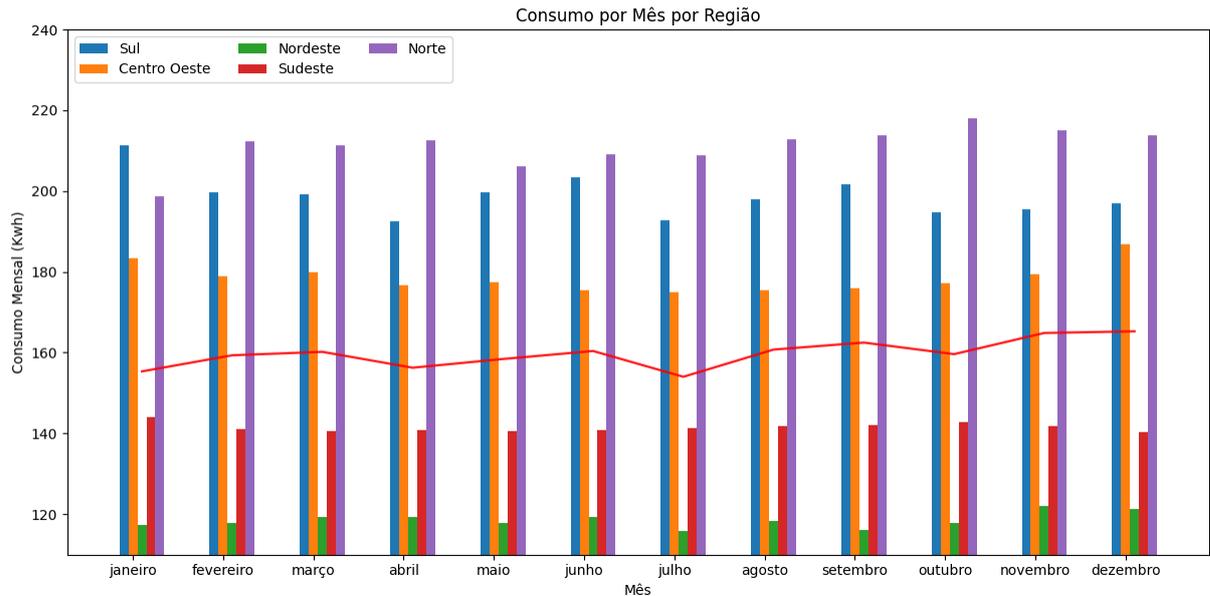


Figura 4.2: Consumo por Mês por Região

Na Figura 4.2 temos um gráfico de barras agrupadas respectivamente com as regiões: Sul (cor azul), Centro-oeste (cor laranja), Nordeste (cor verde), Sudeste (cor vermelha) e Norte (cor roxa); seu eixo X representa os meses do ano e seu eixo Y o Consumo Mensal por kWh. Com exceção do mês de janeiro, o gráfico mostra que a relação encontrada na Figura 4.1 se mantém. A região Norte com o maior consumo durante todo o ano e as regiões litorâneas com menor consumo. Essa região do Brasil é caracterizada por apresentar uma temperatura média-alta durante todo o ano, isso leva a um maior uso de aparelhos de ar-condicionado e sistemas de refrigeração, o que pode estar resultando em um maior consumo de eletricidade residencial.

4.2 Análise do Consumo por Classe Econômica

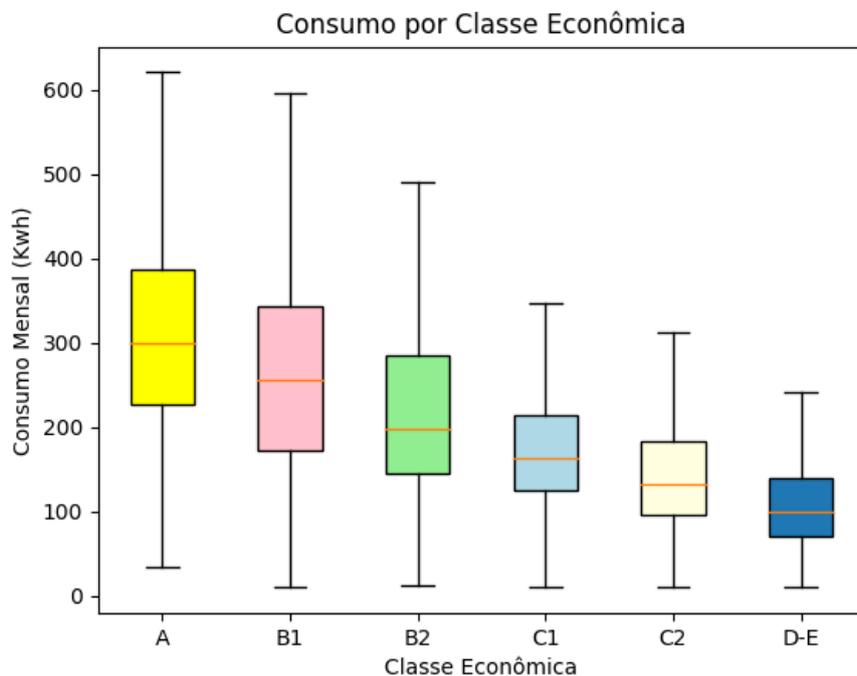


Figura 4.3: Consumo mensal por Classe Econômica

No Boxplot da Figura 4.3 temos em seu eixo X representado a classe econômica e no eixo Y o Consumo Mensal em kWh. A figura mostra que o valor da mediana do consumo de uma residência classe A equivale a 3 residências da Classe D-E, mesmo normalmente morando mais pessoas na residência de classe D-E. A causa pode ser a maior quantidade de aparelhos elétricos nas residências da Classe A ou a diferença no uso de aparelhos que consomem mais energia, como ar condicionado e sistemas de aquecimento.

O Boxplot da Figura 4.4 representa em seu eixo X o nível de escolaridade do chefe de família da residência e no eixo Y o Consumo Mensal em kWh. Observa-se abaixo na Figura 4.4 que indivíduos com maior nível de escolaridade têm um consumo maior de energia, o que pode estar relacionado ao fato de que indivíduos com maior nível de instrução tendem a ter empregos melhor remunerados, o que pode resultar em um consumo maior de energia, conforme evidenciado na Figura 4.3.

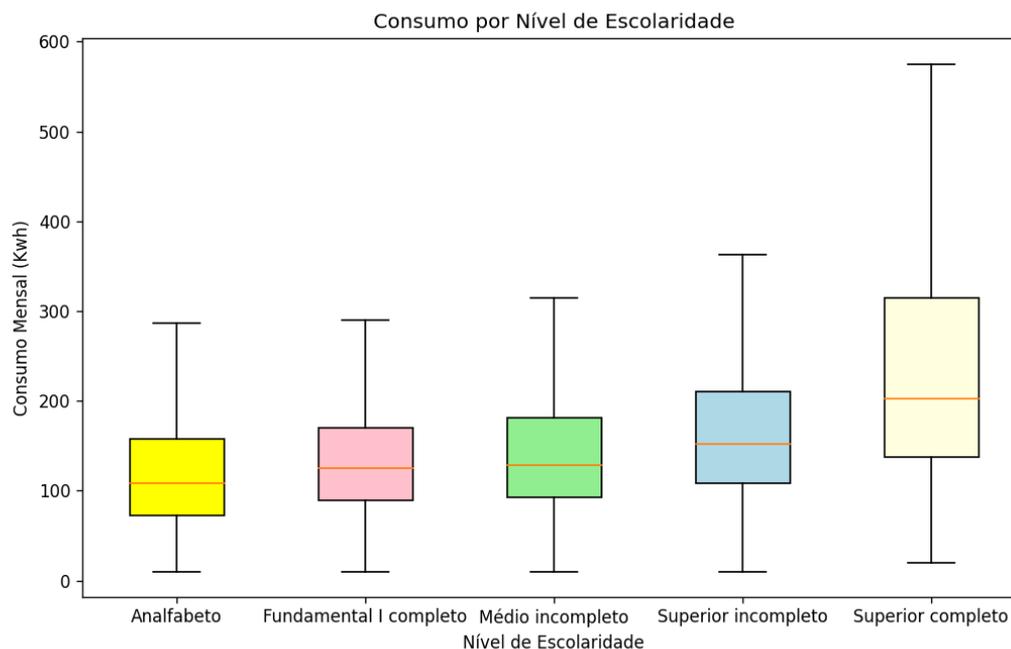


Figura 4.4: Consumo mensal por Nível de Escolaridade

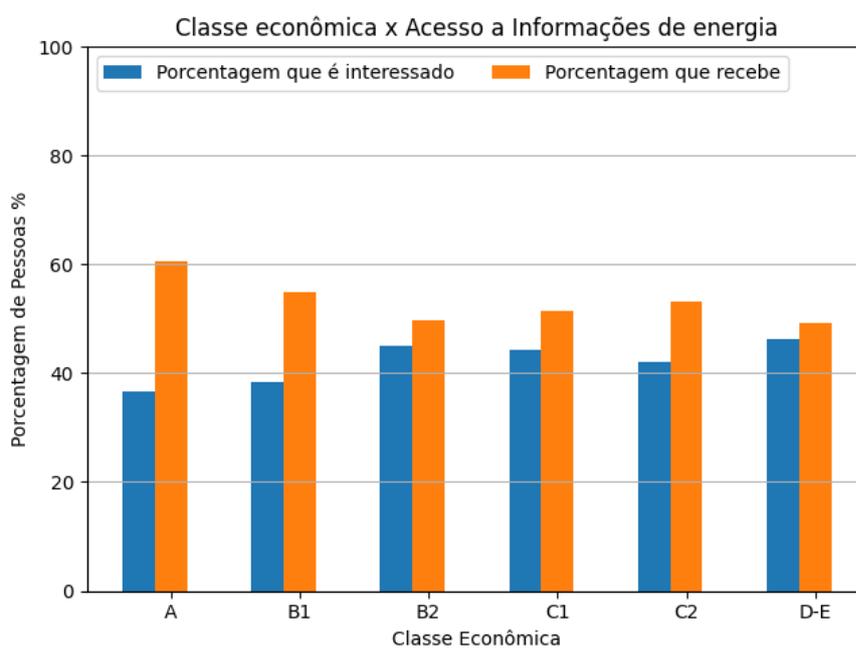


Figura 4.5: Acesso à Informação por Classe Econômica

O gráfico em barras da Figura 4.5 representa em seu eixo x a classe econômica e no eixo y a porcentagem de pessoas, em sua primeira barra (cor azul) mostra as pessoas interessadas em receber informações de energia e na segunda barra (cor laranja) mostra a porcentagem que recebe informações de energia. Conforme a figura, as pessoas de classe econômica mais baixa, por terem um gasto maior como mostrado na Figura 4.3, apresen-

tam mais interesse em receber informações sobre energia e como economizar. Enquanto pessoas em classes econômicas mais altas têm mais acesso a informações sobre energia. Além disso, elas não têm tanto interesse nessas informações por já receberem elas ou ter algum conhecimento sobre o assunto. Apesar das pessoas das classes econômicas mais baixas terem menos acesso a essas informações, provavelmente devido à falta de recursos, despertando maior interesse nelas por questões econômicas, o que não é acompanhado por uma maior compreensão de que essas informações são relevantes para suas vidas diárias.

4.3 Análise do Consumo por Quantidade de pessoas

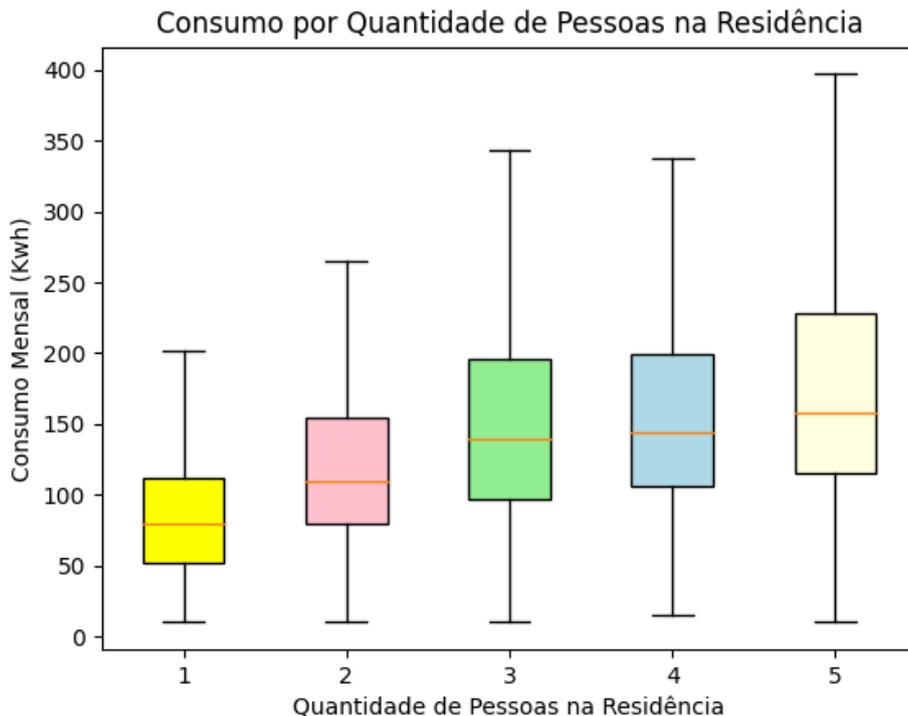


Figura 4.6: Consumo por Quantidade de pessoas na residência

O Boxplot da Figura 4.6 representa em seu eixo x a quantidade de pessoas por residência e no eixo y o Consumo Mensal em kWh, o gráfico demonstra que as residências com maior número de moradores tendem a consumir mais eletricidade. Além disso, a partir de três moradores na residência, o consumo de energia elétrica por pessoa adicionada permanece estável, sem um aumento significativo.

4.4 Análise do Consumo por m²

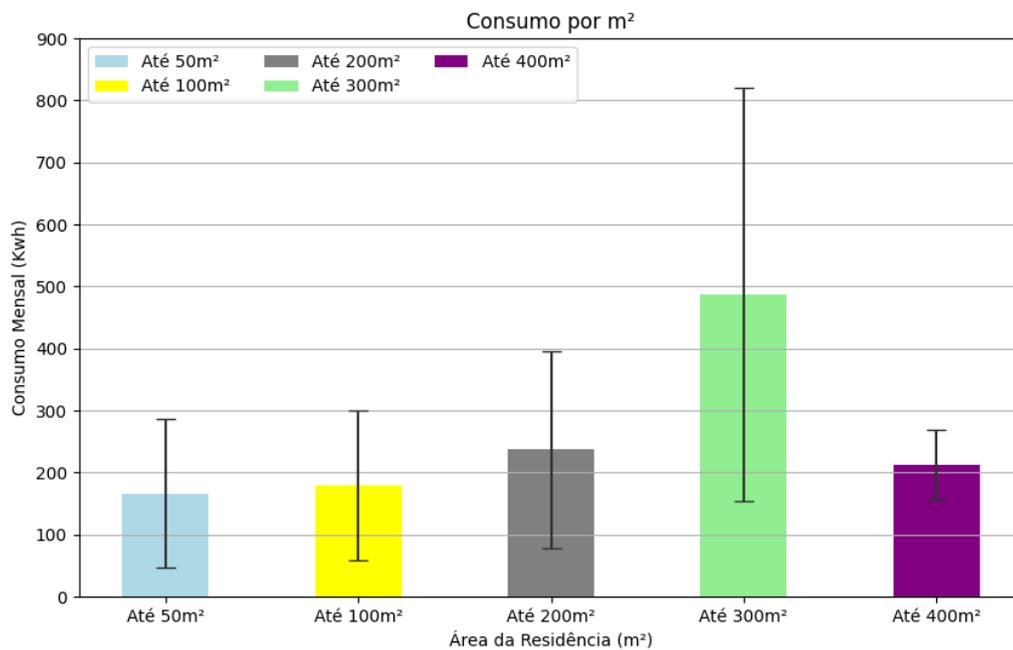


Figura 4.7: Consumo por tamanho da residência (m²)

Observa-se no gráfico em barras da figura 4.7 onde no eixo x representa a área da residência por metro quadrado agrupados por residências com até 50m², 100m², 200m², 300m² e 400m²; e no eixo y o consumo mensal em kWh, que o consumo de eletricidade aumenta à medida que o tamanho da residência aumenta. O maior consumo é visto em residências de 201 a 300 m², enquanto o menor consumo é visto em residências que têm 50m² ou menos.

4.5 Análise do Consumo por nível de conscientização

Para realizar a categorização por nível de conscientização foi criado um sistema de pontuação que indica em qual nível de conscientização energética a residência se encontra. A pontuação é calculada com base nas colunas "Recebe Informação de Energia", "Interesse em receber Informação de Energia", "Compra Produtos Econômicos", "Evita Banho Demorado", "Regula Temperatura da Geladeira" e "Evita Acender Luz de Dia". As colunas foram inicialmente categorizadas com pesos baseados na Economia de Energia que cada ação gera.

Conforme os valores apresentados na Tabela de equipamentos disponibilizada no site da Procel, o consumo médio estimado dos principais aparelhos elétricos em residências foi utilizado para dar informar a economia de energia e os pesos distribuídos segundo a tabela 4.1, em seguida, a tabela 4.2 apresenta os possíveis valores e como são calculadas suas pontuações, e por fim na tabela 4.3 é apresentado os intervalos de cada nível, definido pela soma da pontuação de cada coluna para uma determinada residência.

Tabela de Referência		
Item	Economia de Energia	Peso
Compra Produtos Econômicos	Alta	4
Evita Banho Demorado	Alta	4
Regula Temperatura da Geladeira	Média	3
Evita Acender Luz de dia	Baixa	2
Recebe Informação de Energia	Indireta	1
Interesse em receber Informação de Energia	Indireta	1

Tabela 4.1: Tabela de Referência

Tabela de Pontuação	
Valor	Pontuação
Sempre	1.0 * Peso
Normalmente	0.5 * Peso
Não se aplica ou não respondeu	0.0 * Peso
Raramente	-0.5 * Peso
Nunca	-1.0 * Peso
Sim	1.0 * Peso
Não	-1.0 * Peso

Tabela 4.2: Tabela de pontuação

Tabela de Níveis de Conscientização	
Nível	Intervalo
Nada Consciente	[-15,-9]
Pouco Consciente	(-9,-3)
Neutro	[-3, 3]
Consciente	(3, 9)
Muito Consciente	[9, 15]

Tabela 4.3: Tabela de Níveis de Conscientização

A tabela 4.4 mostra um exemplo de como ocorre o cálculo dos atributos de uma residência com o nível de conscientização "Neutro" por apresentar uma pontuação total -1.0.

Tabela de Exemplo		
Coluna	Valor	Pontuação
Compra Produtos Econômicos	Normalmente	2
Evita Banho Demorado	Normalmente	2
Regula Tem Geladeira	Nunca	-3
Evita Acender Luz Dia	Nunca	-2
Recebe Informação de Energia	Sim	+1.0
Interesse Informação de Energia	Não	-1.0
Pontuação total		-1.0

Tabela 4.4: Tabela de Exemplo

A análise foi proposta com o intuito de verificar a diferença do Consumo mensal que ocorre entre residências que possuem hábitos energéticos nada conscientes e muito conscientes, verificando assim o consumo por níveis de conscientização.

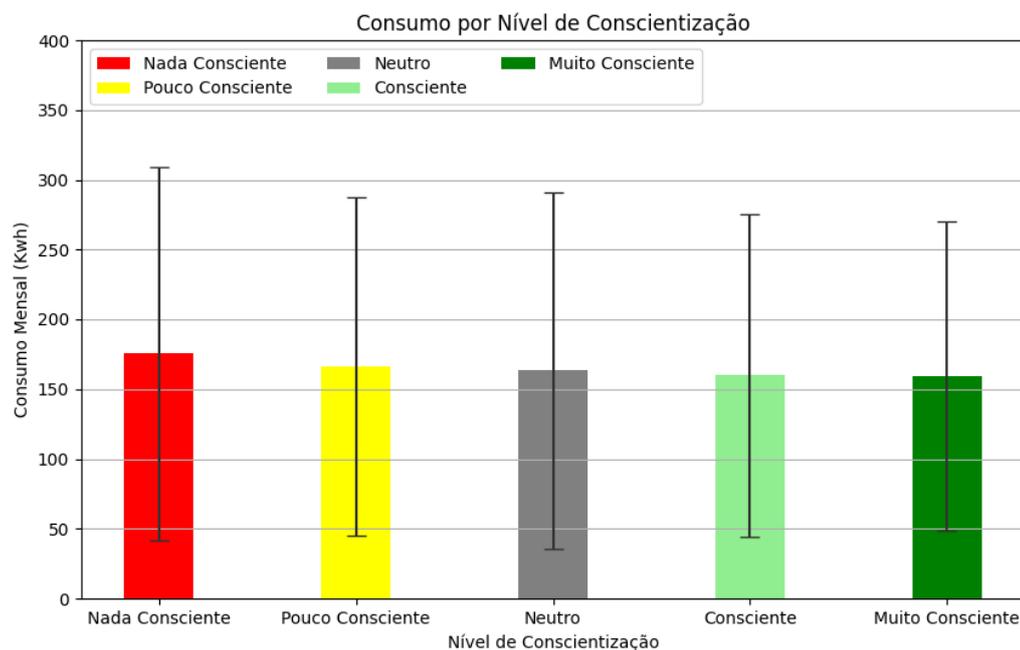


Figura 4.8: Consumo por Nível de Conscientização

O gráfico de barras da figura 4.8 apresenta no eixo x o nível de conscientização e no eixo y o consumo mensal em kWh. A análise do gráfico mostra que o maior consumo é visto na categoria do nível de conscientização "Nada Consciente", e o menor nível de consumo

é visto na categoria “Muito Consciente”. Embora esperado o resultado, vale ressaltar que por ser uma pesquisa com inúmeras perguntas e as referentes a informações sobre conscientização ficarem no final, poucas pessoas entrevistadas pela pesquisa responderam às perguntas sobre conscientização, talvez por esse motivo poucas pessoas se encontrem na categoria “Muito Consciente”.

4.6 Limitações

A pesquisa utilizada é a mais recente focada nesse cenário, rica em informações e muitos aspectos que facilitam análises em várias áreas de conhecimento. No entanto, a pesquisa tem algumas limitações que devem ser consideradas na interpretação dos resultados. Uma delas é que os dados são referentes ao período de janeiro de 2019 a dezembro de 2019, o que pode não refletir as tendências mais recentes ou as mudanças provocadas pela pandemia de COVID-19.

De acordo com (ANIBOLETE, 2021), durante a pandemia em 2021 ocorreu o efeito em algumas regiões do Brasil, onde empresas gastaram menos com energia nos momentos de lockdown. Por outro lado, nas residências brasileiras dessas regiões, houve um aumento considerável do gasto de energia, pessoas viram o valor de sua conta de luz aumentar como consequência do lockdown e distanciamento social. Por esse motivo, e com a falta de dados de uma nova pesquisa mais recente, o período de pandemia não será levado em consideração nesse trabalho. Assim, recomenda-se cautela na generalização dos achados deste artigo e sugere-se a realização de estudos futuros que incorporem outras fontes de dados e períodos mais longos de análise.

Outra limitação, é que os dados da pesquisa não incluem informações claras sobre outras fontes de energia, principalmente de geração própria como energia solar, eólica, biomassa, etc. Área pouco explorada pela pesquisa, que deixa margem para perguntas “Como o consumo apresentado é dividido(%) pelas fontes?”, “O valor apresentado é da geração potencial de energia ou do real consumo da residência?”, “Quanto é a média de energia gerada por tipo de geração utilizada?”, “Quanta energia extra é gerada? E quanto dessa energia é perdida?”. Essas fontes podem representar uma parcela significativa do consumo energético e geração de alguns setores ou regiões, mas não são capturadas pelos

dados oficiais. Portanto, as análises realizadas neste artigo somente leva em consideração residências que possuem o "Tipo de Energia" como "Rede geral de distribuição".

Além disso, apesar da pesquisa focar na categoria residência e não comercial, existe uma lacuna sobre a quantidade e qualidade dos dados de residências onde também ocorrem atividades comerciais, o que impede uma análise mais detalhada desse tipo de residência.

5 Conclusão e Trabalhos Futuros

As análises dos perfis de consumo de energia geradas são simplórias, comparado com o potencial que Data Mining pode alcançar, alcançam perfeitamente o objetivo idealizado por esse TCC, e ainda conseguem demonstrar algumas das muitas possibilidades de estudos que um mesmo dado pode oferecer, diferentes visões atribuídas a esses dados podem gerar inovações e valor para diversas áreas.

Apesar de não ser a única opção, o Data Warehouse com suas tabelas de fatos e dimensões, facilita a compreensão dos dados manipulados, em comparação a outras tecnologias utilizadas para armazenar e processar dados, o que o torna uma ótima ferramenta para compartilhar e tornar os dados mais acessíveis, e ainda pode ser facilmente exportado para outros modelos.

Um desafio enfrentado por todos os países é o custo atribuído a geração de energia elétrica, por isso para reduzir o consumo e o custo pago pelo consumidor final do setor residencial, poderia ser compartilhado com residências que pertencem às classes econômicas mais baixas e moram em locais de maior pobreza, panfletos com infográficos ou algo como o documento da Procel "Dicas da Eletrobras Procel para economizar energia elétrica em residências e condomínios". Pois, como mostrado nas análises, existe uma relação importante nesse indicador.

Existem muitos pontos encontrados durante a realização desse trabalho que poderiam ser utilizados para trabalhos futuros, afinal é um tema muito rico e pouco explorado no Brasil, alguns são citados a seguir: adicionar outras fontes de dados ao Data Warehouse proposto ou uma nova e atualizada versão dessa mesma pesquisa (quando lançada); analisar o desempenho e comparar com outros modelos de armazenamento de dados; modelar e adicionar a dimensão de eletrodomésticos, por existir mais de 1000 colunas só para esse assunto; criar ferramentas para auxiliar na visualização dos dados, desenvolver análises automatizadas com ajuda de IA para encontrar *insights* significativos sobre os dados e ainda reutilizando os dados trabalhar com outros perfis de consumo energéticos para extrair valor para outras áreas.

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*, Associação Nacional de Tecnologia do Ambiente Construído - ANTAC, v. 21, n. 2, p. 383–408, Apr 2021. ISSN 1678-8621. Disponível em: <https://doi.org/10.1590/s1678-86212021000200532>.

ANEEL. Aneel divulga desempenho e ranking das distribuidoras sobre fornecimento de energia em 2021. *Portal ANEEL*, 2022. Disponível em: <https://www.gov.br/aneel/pt-br/assuntos/noticias/2022/aneel-divulga-desempenho-e-ranking-das-distribuidoras-sobre-fornecimento-de-energia-em-2021>.

ANIBOLETE, D. Estudo analisa aumento do consumo residencial de energia durante a pandemia. *Procel Info*, 2021. Disponível em: <https://ccs2.ufpel.edu.br/wp/2021/04/07/estudo-analisa-aumento-do-consumo-residencial-de-energia-durante-a-pandemia/>.

BUCCO, E. Usinas hidrelétricas da região já apresentam dificuldades na geração de energia. *Serra Nossa*, 2021. Disponível em: <https://www.serranossa.com.br/usinas-hidreletricas-da-regiao-ja-apresentam-dificuldades-na-geracao-de-energia/>.

DOUTRELIGNE, M. et al. Good practices for clinical data warehouse implementation: A case study in France. *PLOS Digital Health*, Public Library of Science San Francisco, CA USA, v. 2, n. 7, p. e0000298, 2023.

DRUCKMAN, A.; JACKSON, T. Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model. *Energy Policy*, Elsevier, v. 36, n. 8, p. 3177–3192, 2008.

EPE. *Balanco Energético Nacional 2023*. Empresa de Pesquisa Energética, 2023. Disponível em: <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/balanco-energetico-nacional-2023>.

KIMBALL, R.; ROSS, M. *The data warehouse toolkit: the complete guide to dimensional modeling*. [S.l.]: John Wiley & Sons, 2011.

KRSTEV, S.; KRNETA, D. Data warehouse to support creation of key performance indicators (KPIs) in electricity supply. In: *2022 21st International Symposium INFOTEH-JAHORINA (INFOTEH)*. [S.l.: s.n.], 2022. p. 1–4.

LEVENE, M.; LOIZOU, G. Why is the snowflake schema a good data warehouse design? *Information Systems*, Elsevier, v. 28, n. 3, p. 225–240, 2003.

LI, Y. et al. A data warehouse architecture supporting energy management of intelligent electricity system. In: *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*. Atlantis Press, 2013/03. p. 696–699. ISBN 978-90-78677-61-1. ISSN 1951-6851. Disponível em: <https://doi.org/10.2991/iccsee.2013.177>.

MALAR, J. P. Energia elétrica tem alta acumulada de quase 25% em 2021, diz ibge. *Cnn Brasil*, 2021. Disponível em: <https://www.cnnbrasil.com.br/business/energia-eletrica-tem-alta-acumulada-de-quase-25-em-2021-diz-ibge/>).

NGO, V. M.; LE-KHAC, N.-A.; KECHADI, M.-T. Designing and implementing data warehouse for agricultural big data. In: CHEN, K.; SESHADRI, S.; ZHANG, L.-J. (Ed.). *Big Data – BigData 2019*. Cham: Springer International Publishing, 2019. p. 1–17. ISBN 978-3-030-23551-2.

NS, R. *Energia elétrica e gases poluentes: qual é a relação?* 2021. Disponível em: <https://www.noticiasustentavel.com.br/energia-eletrica-gases-poluentes/>).

SCHUCK, S. "Brasil no escuro": crise climática ameaça produção de energia elétrica. 2023. <https://umsoplaneta.globo.com/clima/noticia/2023/05/27/brasil-no-escuro-crise-climatica-ameaca-producao-de-energia-eletrica-aponta-relatorio.ghml>).

UNEP. *Relatório da Situação Global das Renováveis 2022*. 2022. <https://www.unep.org/pt-br/resources/relatorios/relatorio-da-situacao-global-das-renovaveis-2022>).

VAISMAN, A.; ZIMÁNYI, E. Data warehouse systems. *Data-Centric Systems and Applications*, Springer, 2014.