



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA

MARCOS EUGÊNIO BORGES DE OLIVEIRA JÚNIOR

**ANÁLISE DAS BASES DE DADOS ABERTOS DA ANEEL: ESTUDO DE RELAÇÃO
ENTRE OS CUSTOS DE TARIFICAÇÃO E A QUALIDADE COMERCIAL NO SERVIÇO
DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA DE CONCESSIONÁRIA DO
NORDESTE BRASILEIRO**

FORTALEZA

2024

MARCOS EUGÊNIO BORGES DE OLIVEIRA JÚNIOR

ANÁLISE DAS BASES DE DADOS ABERTOS DA ANEEL: ESTUDO DE RELAÇÃO
ENTRE OS CUSTOS DE TARIFICAÇÃO E A QUALIDADE COMERCIAL NO SERVIÇO DE
DISTRIBUIÇÃO DE ENERGIA ELÉTRICA DE CONCESSIONÁRIA DO NORDESTE
BRASILEIRO

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Engenharia Elétrica do
Centro de Tecnologia da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Engenharia Elétrica.

Orientador: Prof. Dr. Lucas Silveira
Melo

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- O48a Oliveira Júnior, Marcos Eugênio Borges de.
Análise das bases de dados abertos da ANEEL: estudo de relação entre os custos de tarifação e a qualidade comercial no serviço de distribuição de energia elétrica de concessionária do Nordeste brasileiro / Marcos Eugênio Borges de Oliveira Júnior. – 2024.
101 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia Elétrica, Fortaleza, 2024.
Orientação: Prof. Dr. Lucas Silveira Melo.
1. Distribuição de energia elétrica. 2. Análise exploratória. 3. ANEEL. 4. Tarifação . 5. Qualidade comercial de energia elétrica. I. Título.

CDD 621.3

MARCOS EUGÊNIO BORGES DE OLIVEIRA JÚNIOR

ANÁLISE DAS BASES DE DADOS ABERTOS DA ANEEL: ESTUDO DE RELAÇÃO
ENTRE OS CUSTOS DE TARIFICAÇÃO E A QUALIDADE COMERCIAL NO SERVIÇO DE
DISTRIBUIÇÃO DE ENERGIA ELÉTRICA DE CONCESSIONÁRIA DO NORDESTE
BRASILEIRO

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Engenharia Elétrica do
Centro de Tecnologia da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Engenharia Elétrica.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. Lucas Silveira Melo (Orientador)
Universidade Federal do Ceará (UFC)

Maria Ilana dos Santos Guedes
Universidade Federal do Ceará (UFC)

Prof. Dra. Raquel Cristina Filiagi Gregory
Universidade Federal do Ceará (UFC)

À vida, ao Universo e a tudo mais.

AGRADECIMENTOS

À minha avó, Lúcia, que aos meus 10 anos de idade, com seu amor incondicional em tentar sempre ver o neto feliz, atendeu ao meu pedido inocente de querer o livro "Harry Potter e as Relíquias da Morte". Hoje, se eu tivesse a capacidade de enxergar retrospectivamente duas linhas temporais da minha vida, uma com esse acontecimento e a outra sem, tenho certeza que o Marcos hoje seria completamente diferente. A leitura transformou minha vida. Sem perceber, naquele momento, ela acabou mudando o rumo da minha história.

À minha irmã gêmea, Marília, que desde que eu nasci compartilha comigo os melhores e piores momentos da minha jornada, que me entende como ninguém e que nunca me deixou se sentir sozinho no mundo. Gostaria que as pessoas pudessem ter a experiência de tê-la como irmã gêmea, por isso com certeza sou muito abençoado.

À família Oliveira, minha família, e em especial à minha mãe, Myrian, por acreditarem em mim e por, independente de qualquer coisa, terem proporcionado um terreno fértil para que eu pudesse crescer.

Aos meus amigos Ilana, Sarah e Lucas, que foram os primeiros que fizeram eu me sentir em casa na universidade. Sou extremamente grato por todos os momentos que vivemos juntos e fico muito orgulhoso de poder acompanhar de perto o crescimento de cada um.

Aos meus amigos Layson e Tales, que viraram verdadeiros irmãos para mim ao longo dessa jornada. Eu tenho certeza que ainda iremos dominar o mundo, foguete não dá ré meus manos!

Aos meus dois grandes amigos, e também meus veteranos, Alice e Levi, com os quais sinto muita admiração e foram verdadeiros achados para minha vida durante e após a graduação.

À minha namorada, Carol, que chegou em um momento conturbado da minha vida e trouxe felicidade, conforto e muito apoio. Talvez a pessoa que mais queria que eu terminasse logo este trabalho, finalmente veio aí amor. Obrigado por ser tão incrível comigo.

Aos meus amigos Kaio, Dudu, Luan, Navarro, Eidrian, Bia, Ally e Felipe, do grupo "Give a What?", que chegaram um pouco depois e mesmo assim hoje são uma parte essencial da minha vida. Ainda viveremos muitos outros momentos bons, estamos apenas no começo!

A Mariana, Leticia, Davi, Claudio, Emilly, Moacir e a todos os demais amigos que fiz durante a minha passagem pela UFC, com quem compartilhei momentos que para o resto da minha vida irei lembrar com carinho e com saudades.

Ao meu primeiro gestor, Engenheiro Fernando Vidal, que em todos os dias no qual trabalhamos juntos me ensinou sobre a vida, sobre a morte e sobre tudo que existe entre essas duas vertentes. Por mais breve que tenha sido nosso convívio, carrego comigo um respeito imenso pela pessoa e pelo profissional com que tive a oportunidade de trabalhar.

Aos amigos que fiz na Tecsys Jr., vocês foram verdadeiros aliados durante os quase 2 anos de convivência. Por onde eu olhava eu encontrava pessoas incríveis para me inspirar e que me ajudaram a ser quem sou hoje.

Ao meu orientador, Prof. Dr. Lucas Silveira Melo, que me ajudou a desenvolver este trabalho e teve paciência e compreensão das vezes em que eu não consegui cumprir com os prazos quando eu não conseguia tempo para conciliar meu emprego e o desenvolvimento do TCC.

Ao Doutorando em Engenharia Elétrica, Ednardo Moreira Rodrigues, e seu assistente, Alan Batista de Oliveira, aluno de graduação em Engenharia Elétrica, pela adequação do *template* utilizado neste trabalho para que o mesmo ficasse de acordo com as normas da biblioteca da Universidade Federal do Ceará (UFC).

“Eu não devo ter medo. Medo é o assassino da mente. Medo é a pequena morte que leva à aniquilação total. Eu enfrentarei meu medo. Permitirei que passe por cima e me atravesse. E, quando tiver passado, voltarei o olho interior para ver seu rastro. Onde o medo não estiver mais, nada haverá. Somente eu permanecerei”

(Litania Bene Gesserit - Frank Hebert)

RESUMO

No século XXI, os dados emergiram como recurso crucial, influenciando amplamente setores sociais e econômicos, impulsionados pela evolução tecnológica e capacidade de processamento de dados. O conceito de *data-driven* ganha destaque ao priorizar análises objetivas sobre intuições, abrangendo desde empresas até governos. Assim, os dados orientam decisões estratégicas e proporcionam inovações, redefinindo nossas interações sociais e laborais. Nesse sentido, em um primeiro plano, o presente trabalho explorou os dados do Sistema de Distribuição de Energia Elétrica disponibilizados pela Agência Nacional de Energia Elétrica (ANEEL) com a finalidade de entender a relação, caso existisse, entre a qualidade do serviço comercial oferecido pela concessionária de distribuição de energia elétrica e o custo repassado para os consumidores, por meio das tarifas, ao longo dos anos. Já em um segundo plano, o presente trabalho objetivou documentar e trazer luz à importância de se utilizar dados disponíveis, para promover entendimentos e tomadas de decisões na sociedade. Para isso, foram utilizadas ferramentas tecnológicas de análise exploratória de dados para trabalhar os conjuntos de dados escolhidos e trazer *insights* das possíveis conclusões sobre o tema. Além disso, alguns conceitos financeiros foram incorporados ao estudo, permitindo avaliar o aumento dos preços da tarifa ao longo do tempo em relação aos índices financeiros utilizados na indexação de valores contratuais. Desse modo, após a análise feita, foi possível entender o aumento da tarifação referente à distribuição de energia elétrica ao longo dos anos para os grupos de consumidores tipo A e tipo B. Para o primeiro grupo, a análise feita indica que o aumento dos preços ocorreu com ágio em comparação com a média dos índices financeiros de aumento dos preços no mesmo período. Já para o segundo grupo, a situação é o inverso: existe um deságio predominante. Isso pode ter impactado diretamente na capacidade da distribuidora de promover um serviço comercial com maior qualidade, como pode ser visto na queda da colocação da concessionária analisada, saindo do 5º lugar no *ranking* do "Índice ANEEL de Satisfação do Consumidor" (IASC) em 2013 para a 47ª posição em 2022, por exemplo. Por fim, foram propostos trabalhos futuros para se incorporar ao presente estudo e podermos, cada vez mais, tirar proveito dessa valiosa fonte de dados disponibilizada.

Palavras-chave: Distribuição de Energia. Tarifação. Qualidade Comercial de Energia Elétrica. Análise Exploratória de Dados. ANEEL. Inflação.

ABSTRACT

In the 21st century, data has emerged as a crucial resource, widely influencing social and economic sectors, driven by technological evolution and data processing capabilities. The concept of *data-driven* gains prominence by prioritizing objective analyses over intuitions, encompassing everything from businesses to governments. Thus, data guide strategic decisions and provide innovations, redefining our social and work interactions. In this context, initially, this work explored the data from the Electric Power Distribution System made available by ANEEL with the aim of understanding the relationship, if any, between the quality of the commercial service offered by the electricity distribution concessionaire and the cost passed on to consumers through tariffs over the years. Secondly, this work aimed to document and highlight the importance of using available data to promote understanding and decision-making in society. For this purpose, technological tools for exploratory data analysis were used to work on the chosen datasets and bring *insights* into possible conclusions on the subject. Additionally, some financial concepts were incorporated into the study, allowing an assessment of the increase in tariff prices over time in relation to the financial indices used in the indexing of contractual values. Thus, after the analysis, it was possible to understand the increase in tariffs for the distribution of electricity over the years for consumer groups type A and type B. For the first group, the analysis indicates that the price increase occurred with a premium compared to the average financial indices of price increases over the same period. For the second group, the situation is the opposite: there is a predominant discount. This may have directly impacted the distributor's ability to promote higher quality commercial service, as seen in the decline in the concessionaire's ranking from 5th place in the *ranking* of the IASC in 2013 to the 47th position in 2022, for example. Finally, future works were proposed to incorporate into the present study so we can increasingly take advantage of this valuable source of available data.

Keywords: Energy Distribution. Tariffing. Commercial Quality of Electric Power. Exploratory Data Analysis. ANEEL. Inflation.

LISTA DE FIGURAS

Figura 1 – Conjunto de Dados <i>Open Source</i> ANEEL	17
Figura 2 – Visão Geral do Caminho da Energia Até o Consumidor Final	20
Figura 3 – Descritivo Metodologia IASC	21
Figura 4 – Curva de Carga Horária Região Nordeste	23
Figura 5 – Estrutura Macro da Metodologia Adotada	34
Figura 6 – Estrutura Macro Expandida	35
Figura 7 – Configurando o ODBC	42
Figura 8 – Simulação Custos de Distribuição Grupo B	44
Figura 9 – Acumulado do Aumento Real nos Custos de Distribuição Grupo B	44
Figura 10 – Acumulado do índice IPCA no período	45
Figura 11 – Acumulado do índice IGPM-M no período	45
Figura 12 – Simulação Custos de Distribuição Grupo A1	46
Figura 13 – Simulação Custos de Distribuição Grupo A3	46
Figura 14 – Simulação Custos de Distribuição Grupo A4	47
Figura 15 – Acumulado do Aumento Real nos Custos de Distribuição Grupo A	47
Figura 16 – Inadimplência financeira por ano	48
Figura 17 – Reclamações procedentes	49
Figura 18 – Tempo de interrupção de energia por ano por grupo	50
Figura 19 – Prazo Médio Ligação Grupo A	51
Figura 20 – Prazo Médio Ligação Grupo B	51
Figura 21 – Quantidade Prazos Violados Grupo A	52
Figura 22 – Quantidade Prazos Violados Grupo B	52
Figura 23 – Indicador de Valor Gerado	53
Figura 24 – Indicador de Fidelidade	54
Figura 25 – Indicador de Satisfação	55
Figura 26 – Indicador de Confiança	56
Figura 27 – Indicador de Qualidade	56
Figura 28 – Ranking Anual IASC	57

LISTA DE TABELAS

Tabela 1 – Divisão Subgrupos Grupo A	22
Tabela 2 – Divisão Subgrupos Grupo B	22
Tabela 3 – Situação de contorno	39

LISTA DE ABREVIATURAS E SIGLAS

ANEEL	Agência Nacional de Energia Elétrica
CDE	Conta de Desenvolvimento Energético
CSV	<i>Comma-separated values</i>
DRA	Data de Referência Anterior
DRP	Data do Reajuste em Processamento
ETL	"Extrair, transformar e carregar"
FGV	Fundação Getúlio Vargas
IASC	"Índice ANEEL de Satisfação do Consumidor"
IBGE	Instituto Brasileiro de Geografia e Estatística
IBRE	Instituto Brasileiro de Economia
IGP	Índice Geral de Preços
IGP-M	Índice Geral de Preços - Mercado
INCC	Índice Nacional de Custo da Construção
IPA	Índice de Preços ao Produtor Amplo
IPC	Índice de Preços ao Consumidor
IPCA	Índice Nacional de Preços ao Consumidor Amplo
ONS	Operador Nacional do Sistema Elétrico
Proret	Procedimentos de Regulação Tarifária
PyPI	Python Package Index
PYPL	Popularity of Programming Language
RTA	Reajuste Tarifário Anual
RTP	Revisão Tarifária Periódica
SAMP	Sistema de Acompanhamento de Informações de Mercado para Regulação Econômica
SQL	<i>Structured Query Language</i>
TUSD	<i>Tarifa de Uso do Sistema de Distribuição</i>
TUST	<i>Tarifa de Uso do Sistema de Transmissão</i>

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Justificativa	16
1.2	Metodologia	18
1.3	Objetivos	18
<i>1.3.1</i>	<i>Objetivo geral</i>	<i>18</i>
<i>1.3.2</i>	<i>Objetivos específicos</i>	<i>19</i>
1.4	Organização do trabalho	19
2	DISTRIBUIÇÃO DE ENERGIA ELÉTRICA	20
2.1	Sistema de Distribuição	20
2.2	Qualidade Comercial da Distribuição de Energia Elétrica	20
<i>2.2.1</i>	<i>Índice ANEEL de Satisfação do Consumidor (IASC)</i>	<i>21</i>
2.3	Divisão dos Grupos Consumidores	22
2.4	Curva de Carga Horária	23
2.5	Modalidades de Tarifas Aplicadas aos Consumidores	24
2.6	Postos Tarifários	25
3	REAJUSTE TARIFÁRIO ANUAL DAS CONCESSIONÁRIAS DE DISTRIBUIÇÃO	26
3.1	Índices financeiros de evolução dos preços	26
<i>3.1.1</i>	<i>Índice Nacional de Preços ao Consumidor Amplo (IPCA)</i>	<i>26</i>
<i>3.1.2</i>	<i>Índice Geral de Preços - Mercado (IGP-M)</i>	<i>26</i>
3.2	Reajuste Tarifário Anual das Concessionárias de Distribuição	27
<i>3.2.1</i>	<i>Contexto geral</i>	<i>27</i>
<i>3.2.2</i>	<i>Receita Anual RA_0</i>	<i>27</i>
<i>3.2.3</i>	<i>Parcela A</i>	<i>28</i>
<i>3.2.4</i>	<i>Parcela B</i>	<i>28</i>
4	FERRAMENTAS UTILIZADAS	30
4.1	Linguagem de Programação Python	30
<i>4.1.1</i>	<i>Jupyter Notebook</i>	<i>30</i>
4.2	Biblioteca Pandas	31
<i>4.2.1</i>	<i>Processo de ETL</i>	<i>31</i>

4.2.2	<i>Banco de dados SQLite</i>	32
4.2.3	<i>Power BI</i>	32
5	METODOLOGIA	34
5.1	Base de dados analisada	35
5.1.1	<i>Escolha e obtenção dos dados</i>	36
5.1.2	<i>Estruturação dos dados no banco SQLite</i>	37
5.2	Algoritmo desenvolvido	37
5.2.1	<i>Análise tarifária</i>	38
5.2.1.1	<i>Premissas adotadas</i>	38
5.2.1.2	<i>Descrição do algoritmo</i>	39
5.2.2	<i>Análise da inadimplência</i>	40
5.2.3	<i>Análise da qualidade comercial</i>	40
5.2.4	<i>Análise das reclamações N1 e N2</i>	40
5.2.5	<i>Análise das interrupções</i>	41
5.2.6	<i>Análise do IASC</i>	41
5.3	Visualização dos dados modelados	42
6	RESULTADOS	43
6.1	Resultados da análise de tarifação aplicada ao longo dos anos	43
6.1.1	<i>Consumidores de baixa tensão (Grupo B)</i>	43
6.1.2	<i>Consumidores de Alta Tensão (Grupo A)</i>	46
6.2	Resultados da análise de inadimplência dos consumidores para a distri- buidora de energia elétrica analisada	47
6.3	Resultados da análise dos indicadores de qualidade	48
6.3.1	<i>Reclamações procedentes</i>	48
6.3.2	<i>Tempo de interrupção de energia</i>	49
6.3.3	<i>Indicadores relacionados a ligação de energia</i>	50
6.3.4	<i>Indicador de Satisfação IASC</i>	52
7	CONCLUSÕES E TRABALHOS FUTUROS	58
	REFERÊNCIAS	60
	APÊNDICES	62
	APÊNDICE A – Links dos conjuntos de dados utilizados	62
	ANEXOS	62

	ANEXO A – Jupyter Notebooks Utilizados	63
A.1	ANALISE_INADIMPLENCIA_NOTEBOOK	63
A.2	ANALISE_INTERRUPCAO_ENERGIA_NOTEBOOK	64
A.3	ANALISE_QUALIDADE_COMERCIAL_NOTEBOOK	67
A.4	ANALISE_RECLAMACOES_N1_N2_NOTEBOOK	74
A.5	ANALISE_SATISFACAO_CONSUMO_NOTEBOOK	77
A.6	ANALISE_TARIFA_NOTEBOOK	78
	ANEXO B – Estrutura Macro Expandida	101

1 INTRODUÇÃO

No século XXI, a importância dos dados atingiu um patamar sem precedentes, transformando-se em um recurso valioso que molda diversos setores da sociedade, desde a economia à pesquisa científica. O advento da tecnologia da informação, aliado ao crescimento exponencial da capacidade de armazenamento e processamento de dados, desencadeou uma revolução na forma como as organizações e indivíduos interagem com a informação (ROSS *et al.*, 2017).

O termo *data-driven* (orientado por dados) refere-se a uma abordagem ou método em que as decisões são fundamentadas na análise de dados e evidências objetivas, em vez de depender exclusivamente da intuição, experiência ou opinião subjetiva. É um conceito amplamente utilizado em diversos setores, desde negócios e ciência até governos e organizações sem fins lucrativos (PROVOST; FAWCETT, 2013).

Nesse sentido, neste século, os dados são um recurso crítico que impulsiona a inovação, orienta decisões estratégicas, promove avanços científicos e redefine a maneira como vivemos, trabalhamos e nos relacionamos. Cada vez mais, decisões baseadas em "achismos" estão sendo deixadas de lado, com uma crescente orientação para decisões baseadas em dados. Porém, para que essas análises sejam possível, é necessário antes de tudo existir a matéria-prima fundamental: o dado (HARARI, 2017).

1.1 Justificativa

Ao redor do mundo temos vários exemplos de nações desenvolvidas que se preocupam com a disponibilização de dados abertos, popularmente denominados por *open source*. Um bom exemplo são os Estados Unidos da América com sua lei *OPEN Government Data*, que estabelece um domínio público na internet uma exigência legal, em vez de uma política. Ela exige que as agências federais publiquem nesse domínio suas informações online como dados abertos, usando formatos de dados padronizados e legíveis por máquina (DATA.GOV,). Assim, qualquer pessoa ou organização pode utilizar essa fonte de dados para contribuir com o desenvolvimento de diversos setores da sociedade a que essas informações se relacionam.

No Brasil, a ANEEL disponibiliza dados *open source* referentes ao setor elétrico brasileiro, conforme seu Plano de Dados Abertos (PDA) para 2022-2024, de Portaria nº 6.785/2022. Esse plano está alinhado ao Planejamento Estratégico e ao Plano Estratégico de Tecnologia da

Informação (PETI) da agência. O objetivo do PDA é promover a abertura de dados, seguindo os princípios da publicidade, transparência e eficiência. Isso visa aumentar a disseminação de informações para a sociedade, melhorar a qualidade dos dados e apoiar a tomada de decisões pelos gestores públicos, além de favorecer o desenvolvimento social (ANEEL, 2023b).

Figura 1 – Conjunto de Dados *Open Source* ANEEL

The screenshot displays the ANEEL Open Source Data Portal interface. At the top, there is a navigation bar with the ANEEL logo and a search bar. Below the navigation bar, the main content area is titled 'Conjuntos de dados'. On the left, there is a sidebar with a tree view of categories: 'Organizações' (60), 'Grupos' (17), 'Distribuição' (12), 'Geração' (10), 'Fiscalização' (9), 'Tarifas' (3), 'Eficiência Energética' (3), 'Mercado' (3), 'Outros assuntos' (3), 'Pesquisa e Desenvol...' (3), and 'Transmissão' (3). The main content area features a search bar with the text 'Search datasets...' and a magnifying glass icon. Below the search bar, it displays '60 conjuntos de dados encontrados' and a dropdown menu for 'Ordenar por: Relevância'. Two dataset entries are visible: 'Relação de empreendimentos de Geração Distribuída' and 'SIGEC - Sistema de Gestão de Créditos'. Each entry includes a brief description and download options for PDF and CSV files.

Fonte: (ANEEL, 2023b).

Como pode-se perceber na figura 1, já há a disponibilização de 60 conjuntos de dados de diversas áreas do setor elétrico brasileiro. Vale ressaltar que, por ter sido uma iniciativa relativamente recente, esses dados ainda não são muito conhecidos no meio acadêmico. Em uma busca rápida por trabalhos e pesquisas que utilizassem de forma direta essas bases feita pelo autor, não foram encontrados nenhum estudo dessa natureza.

Nesse sentido, haja vista a oportunidade que se estabelece havendo de forma acessível dados sobre o setor elétrico brasileiro, o presente trabalho se concentrará em analisar os dados de distribuição de energia elétrica de uma concessionária do nordeste brasileiro buscando relacionar a qualidade comercial do serviço oferecido com o aumento dos preços no componente da tarifação praticada referente a distribuição de energia.

Encontrar essa relação, caso seja possível, será importante para entender a situação da concessionária de energia frente aos preços praticados: se as tarifas aumentaram menos do que deveriam e isso impactou a qualidade dos serviços oferecidos, se aumentaram mais do que deveriam e a qualidade não acompanhou esse aumento ou se, até mesmo, não tiver sido possível

encontrar relação entre essas questões.

Desse modo, ao fim desse tipo de análise, será possível entender padrões que antes não eram claros acerca da temática trabalhada. No caso do presente trabalho, essa análise permitirá que decisões *data-driven* sejam tomadas no que diz respeito a qualidade do serviço oferecido pela concessionária analisada.

1.2 Metodologia

Inicialmente, será feito um mapeamento dos conjuntos de dados disponibilizados pela ANEEL em que as características do dado permitam analisar a relação do custo *versus* a qualidade da prestação de serviço comercial da concessionária de distribuição de energia elétrica.

Em seguida, valendo-se do ferramental de programação computacional e de *software* de visualização de dados, será feita uma análise exploratória de dados nessas bases para entender, principalmente, como o aumento da tarifação de energia elétrica se relaciona com a qualidade comercial do serviço de distribuição de energia elétrica oferecida ao longo dos anos. Serão incorporados alguns conceitos econômicos de análise de dinheiro no tempo a fim de ajudar a comparação do ponto de vista financeiro.

1.3 Objetivos

1.3.1 Objetivo geral

Ao desenvolver uma análise exploratória dos dados abertos de distribuição de energia elétrica disponibilizados pela ANEEL, espera-se obter alguns *insights* sobre a forma como a qualidade no serviço de distribuição de energia elétrica de um estado do Nordeste analisado se relaciona com preço cobrado pela distribuidora ao consumidor final. Isto é, ao longo dos anos, se houve um aumento nos custos do serviço ao consumidor será que isso significou também uma melhora geral na qualidade da distribuição de energia?

Além disso, espera-se também que este estudo fomente e dissemine o interesse pela análise dos dados disponibilizados pela ANEEL ou por demais demais órgãos do setor em prol de um melhor entendimento dos cenários existentes, bem como de possíveis melhorias e/ou soluções.

1.3.2 *Objetivos específicos*

Para atingir o objetivo geral, os seguintes objetivos específicos foram estabelecidos:

- Examinar os dados do Sistema de Distribuição de Energia Elétrica de um estado do Nordeste brasileiro, disponibilizados pela ANEEL.
- Analisar o aumento dos custos das tarifas de distribuição para os consumidores da distribuidora de energia elétrica analisada.
- Correlacionar a variação de tarifas com os índices financeiros utilizados na indexação de valores contratuais.
- Avaliar a qualidade comercial da concessionária de energia ao longo dos anos.
- Correlacionar a variação das tarifas cobradas frente a qualidade comercial no serviço oferecido pela concessionária.
- Documentar e destacar a importância do uso de dados disponíveis para promover entendimentos e tomadas de decisões na sociedade.

1.4 Organização do trabalho

O presente trabalho é composto por este Capítulo 1 responsável pela Introdução do trabalho, três capítulos destinados a fundamentação teórica dos assuntos abordados, um capítulo para a metodologia utilizada, um capítulo de apresentação dos resultados obtidos e um capítulo final de conclusão dos resultados.

Na parte da fundamentação teórica, o Capítulo 2 trata do Sistema de Distribuição de Energia Elétrica. Já o Capítulo 4 fica responsável por apresentar todas as ferramentas utilizadas no trabalho. O Capítulo 3 fala sobre os reajustes tarifários anuais das concessionárias de distribuição e dos índices financeiros que serão utilizados em posteriores análises financeiras.

No Capítulo 5, referente à Metodologia, é apresentado todo o passo a passo adotado desde a obtenção dos conjuntos de dados da ANEEL, passando pelo processamento dos dados pelos algoritmos desenvolvidos e chegando até a geração dos gráficos finais.

Já no Capítulo 6 de Resultados são apresentados os gráficos finais da análise exploratória de dados e são tecidos alguns comentários do autor a partir de inferências feitas a partir dos resultados.

Por fim, a conclusão do trabalho como um todo é apresentada no Capítulo 7

2 DISTRIBUIÇÃO DE ENERGIA ELÉTRICA

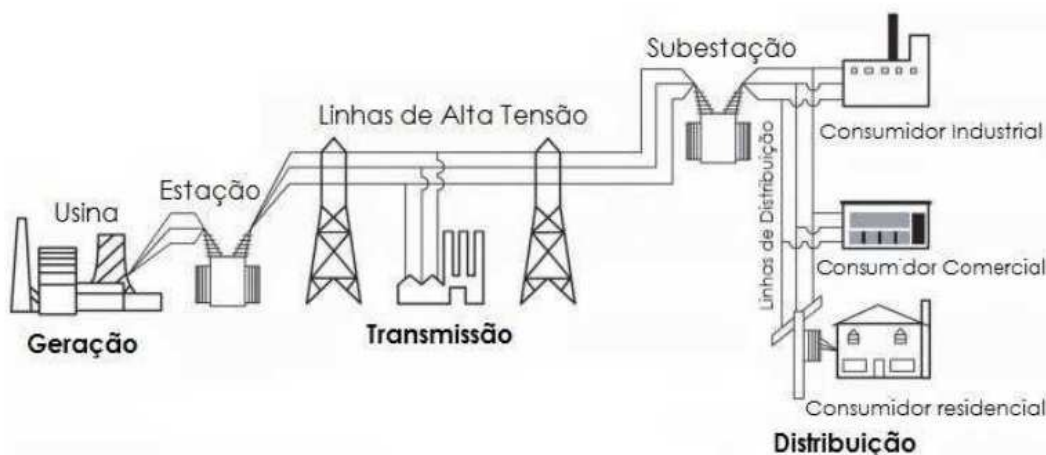
2.1 Sistema de Distribuição

O sistema de distribuição de energia elétrica brasileiro é o conjunto de redes elétricas que recebem, das linhas de transmissão, a energia gerada nas usinas e fica responsável por levá-la até o consumidor final. O sistema é composto por linhas de distribuição em alta, média e baixa tensão, além de subestações, transformadores e medidores. O sistema é operado e mantido pela concessionária de energia de cada região, que fica responsável, dentre outras coisas, pela manutenção e aprimoramento da rede, garantindo níveis de qualidade mínimos estabelecidos pelo órgão regulador, a ANEEL. Esta, por sua vez, é a entidade responsável pela regulação do setor elétrico, fiscalizando a qualidade do serviço oferecido e definindo as tarifas cobradas pelos serviços aos consumidores.

Por conseguinte, esse sistema opera em diferentes níveis de tensão, abrangendo desde alta tensão (69 kV ou superior e inferior a 230 kV), média tensão (2,3 kV ou superior e inferior a 69 kV) até baixa tensão (inferior a 2,3 kV) (ANEEL, 2023c).

Na figura 2 tem-se uma ilustração de todas as etapas que antecedem a distribuição de energia na rede elétrica ao consumidor final.

Figura 2 – Visão Geral do Caminho da Energia Até o Consumidor Final



Fonte: (BLUME, 2007)

2.2 Qualidade Comercial da Distribuição de Energia Elétrica

A avaliação da qualidade dos serviços fornecidos pelas empresas do setor elétrico de distribuição de energia pode ser realizada considerando três principais aspectos: a continuidade

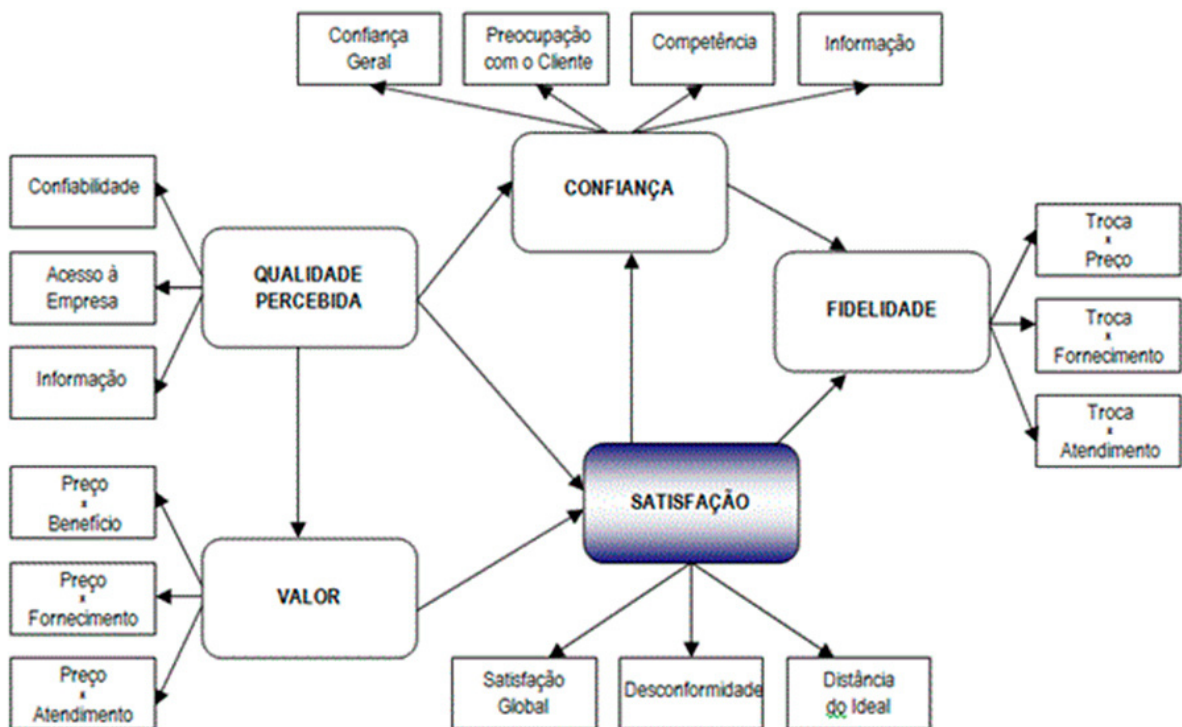
no fornecimento, a excelência no atendimento ao consumidor e a qualidade do produto, que, neste caso, refere-se à energia elétrica (BELISÁRIO *et al.*, 2003).

2.2.1 Índice ANEEL de Satisfação do Consumidor (IASC)

O IASC é um indicador, criado por uma equipe de especialistas da ANEEL, para avaliar o grau de satisfação do consumidor residencial com os serviços prestados pelas distribuidoras de energia elétrica.

Ele é obtido por meio de uma pesquisa amostral com aproximadamente 27.000 consumidores de todas as distribuidoras, concessionárias e permissionárias atuantes nacionalmente. Seus resultados são empregados para aprimorar instrumentos regulatórios e orientar a priorização das ações de fiscalização.

Figura 3 – Descritivo Metodologia IASC



Fonte: (ANEEL, 2024)

De forma resumida, tem-se as seguintes descrições:

- A avaliação do **Valor Percebido** analisa como o consumidor percebe, do ponto de vista econômico, as vantagens proporcionadas pela energia, considerando a qualidade do fornecimento e o atendimento ao cliente.

- A variável de **Satisfação** é avaliada por meio de três indicadores: satisfação global, distância em relação à empresa ideal e desconformidade global.
- A avaliação da **Confiança no Fornecedor** verifica se os consumidores percebem os provedores de energia elétrica como confiáveis, dedicados aos interesses dos consumidores, competentes na prestação de serviços e fornecedores de informações precisas.
- A avaliação da **Fidelidade** analisa a probabilidade de os consumidores trocarem de fornecedor com base em critérios como tarifas, qualidade do fornecimento e atendimento.

2.3 Divisão dos Grupos Consumidores

Conforme disposto na resolução normativa vigente (ANEEL, 2021), o Grupo A de consumidores é formado pela distribuição com base na tensão de conexão apresentada na Tabela 1:

Tabela 1 – Divisão Subgrupos Grupo A

Subgrupo	Tensão de conexão
A1	≥ 230 kV
A2	≥ 88 kV e ≤ 138 kV
A3	$= 69$ kV
A3a	≥ 30 kV e ≤ 44 kV
A4	$\geq 2,3$ kV e ≤ 25 kV

Fonte: elaborado pelo autor (2023)

Já o Grupo B apresenta-se dividido em unidades consumidoras com tensão de conexão menor que 2,3 kV conforme a Tabela 2:

Tabela 2 – Divisão Subgrupos Grupo B

Subgrupo	Tensão de conexão
B1	Residencial
B2	Rural
B3	Demais classes
B4	Iluminação Pública

Fonte: elaborado pelo autor (2023)

2.4 Curva de Carga Horária

A curva de carga horária representa a variação da carga elétrica em um sistema durante um período específico, geralmente ao longo de um dia ou de uma semana.

Ao analisar intervalos diários, a variabilidade na demanda de energia pode ser significativa, com amplitudes de 50 a 150% da média diária. Essa variação depende das características específicas de cada classe de consumo, incluindo fatores como jornada de trabalho, iluminação natural e hábitos de uso de equipamentos, levando a flutuações na necessidade de consumo ao longo do dia (FERREIRA *et al.*, 2020)

Para entendermos melhor a dinâmica de consumo na ponta do sistema de distribuição, podemos analisar de forma mais ampla o Sistema Interligado Nacional (SIN) já que a transmissão e a distribuição de energia são partes correlatas. O Operador Nacional do Sistema Elétrico (ONS) possui abordagem inovadora para apresentar os resultados passados da operação visando transparência e confiabilidade das informações. Em seu site (ONS, 2023) é possível traçar a Curva de Carga Horária para um determinado período. Por exemplo, para a região nordeste, no dia 16/11/2023, na Figura 4 temos a representação da curva de carga:

Figura 4 – Curva de Carga Horária Região Nordeste



Vale ressaltar que essa curva varia conforme as características do dia, se é um dia útil ou não, por exemplo, e até mesmo para períodos diferentes do ano, em casos de períodos mais quentes do ano que aumenta o consumo de energia por refrigeradores.

2.5 Modalidades de Tarifas Aplicadas aos Consumidores

Conforme a divisão de subgrupos de alta e média tensões disposta na Tabela 1, temos duas modalidades tarifárias existentes (ANEEL, 2023a):

- **Horária Azul:** Oferece tarifas distintas para o consumo de energia elétrica e a demanda de potência, variando de acordo com as diferentes horas do dia (postos tarifários). Aplica-se a todos os subgrupos do Grupo A.
- **Horária Verde:** Proporciona tarifas diferenciadas para o consumo de energia elétrica, levando em consideração as horas específicas do dia (postos tarifários), juntamente com uma única tarifa para a demanda de potência. Esta modalidade está disponível para os subgrupos A3a, A4 e AS.

Já para os subgrupos de baixa tensão dispostos na Tabela 2 (Grupo B), as modalidades tarifárias podem ser:

- **Convencional Monômnia:** Estabelece uma única tarifa para o consumo de energia elétrica, sem considerar as variações ao longo do dia.
- **Horária Branca:** Apresenta tarifas diferenciadas para o consumo de energia elétrica, ajustadas de acordo com os horários específicos do dia (postos tarifários). Importante observar que essa modalidade não está disponível para o subgrupo B4 e para a subclasse Baixa Renda do subgrupo B1.

Além disso, a grosso modo, a tarifa aplicada aos consumidores pode ser dividida em duas origens:

- **Tarifa de Uso do Sistema de Distribuição (TUSD):** é a tarifa cobrada pelo uso do sistema de distribuição de energia elétrica, distinguida por uma tarifa que varia de acordo com a demanda de potência e o consumo de energia no grupo A, enquanto no grupo B é aplicada uma tarifa única baseada no consumo de energia.
- **Tarifa de Uso do Sistema de Transmissão (TUST):** é a tarifa cobrada pelo uso do sistema de transmissão de energia elétrica, distinguida por uma tarifa única referente à demanda de potência.

2.6 Postos Tarifários

Os postos tarifários são estabelecidos por cada distribuidora durante seu processo regular de revisão tarifária, ocorrendo a cada 4 ou 5 anos, conforme estabelecido pela Resolução Normativa (REN) nº 1.000/2021 (ANEEL, 2021) e os Procedimentos de Regulação Tarifária (Proret).

No geral, a lógica segue as seguintes divisões:

- **Horário de ponta:** Período diário composto por 3h consecutivas em dias úteis;
- **Horário intermediário:** Exclusivamente aplicado a consumidores que escolham pela tarifa branca, é o período de 1h a 1h30 compreendido antes e depois do horário de ponta;
- **Horário fora de ponta:** Período complementar que engloba o restante das horas do dia.

3 REAJUSTE TARIFÁRIO ANUAL DAS CONCESSIONÁRIAS DE DISTRIBUIÇÃO

3.1 Índices financeiros de evolução dos preços

Os índices financeiros de evolução de preços são ferramentas essenciais para acompanhar e compreender as variações nos preços de bens e serviços ao longo do tempo. Eles desempenham um papel crucial na análise econômica e na tomada de decisões financeiras, pois fornecem uma previsão no aumento geral e contínuo dos preços de bens e serviços em uma economia.

Resumidamente, esse fenômeno financeiro relacionado ao aumento dos preços de produtos e serviços é chamado de inflação. Para avaliar essa variação dos preços, existem diversos índices, como o IPCA e o IGP-M. Esses índices acompanham a variação de preços de uma cesta de produtos e serviços representativa do consumo médio das famílias, incluindo alimentos, habitação, transporte, saúde, entre outros (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE), 2023a).

3.1.1 IPCA

O IPCA mede a variação dos preços de um conjunto de produtos e serviços consumidos pelas famílias brasileiras. Ele é calculado pelo IBGE, com base em uma cesta de itens que representa o consumo médio das famílias com renda entre 1 e 40 salários mínimos. O IPCA é divulgado mensalmente e serve como referência para as metas de inflação do governo e para a correção de alguns contratos (IBGE, 2023b).

3.1.2 IGP-M

O IGP, criado no final dos anos 1940, é uma média ponderada dos índices de preços Índice de Preços ao Produtor Amplo (IPA) - 60%, Índice de Preços ao Consumidor (IPC) - 30% e Índice Nacional de Custo da Construção (INCC) - 10%. Seu propósito original era ser uma medida ampla do movimento de preços no país, abrangendo não apenas diferentes atividades, mas também etapas distintas do processo produtivo. Essa concepção permitiu que o IGP fosse utilizado como deflator do índice de evolução dos negócios, tornando-se um indicador mensal do nível de atividade econômica. Os pesos de cada índice são determinados a partir das parcelas da despesa interna bruta, conforme calculadas nas Contas Nacionais. O IGP-M é o resultado

dessa pesquisa feita entre os dias 21 do mês anterior e 20 do mês de referência (FUNDAÇÃO GETÚLIO VARGAS (FGV), 2023).

3.2 Reajuste Tarifário Anual das Concessionárias de Distribuição

3.2.1 Contexto geral

No Reajuste Tarifário Anual (RTA), a ANEEL é responsável por homologar os reajustes das tarifas, observando estritamente as leis, normas e disposições dos contratos de concessão de distribuição de energia elétrica. As tarifas estabelecidas nos contratos, juntamente com as regras de reajuste e revisão, garantem a adequada prestação dos serviços e o equilíbrio econômico-financeiro dos contratos. As tarifas são reajustadas anualmente na data de aniversário do contrato de concessão, exceto nos anos em que ocorre a Revisão Tarifária Periódica (RTP). Para o reajuste tarifário, a receita da distribuidora, excluindo-se impostos, é dividida em duas parcelas: Parcela "A" e Parcela "B" (ANEEL, 2022).

A Parcela "A" inclui os custos de geração e transmissão de energia elétrica e encargos setoriais, que não são controlados pela distribuidora. Já a Parcela "B" abrange os custos gerenciáveis pela distribuidora, como operacionais, remuneração do capital e quota de reintegração, representando a receita remanescente após deduzida a Parcela "A" (ANEEL, 2022).

O "Mercado de Referência" é composto por montantes de energia, demanda de potência e uso do sistema faturados no "Período de Referência", que é o ciclo de 12 meses anteriores ao mês do reajuste. Se os índices IGP-M ou IPCA do último mês do Período de Referência não forem divulgados até 10 dias antes da DRP, será considerada a projeção mais recente informada pelo Banco Central, exceto se a divulgação oficial ocorrer antes da deliberação do processo tarifário (ANEEL, 2022).

3.2.2 Receita Anual RA_0

A Receita Anual (RA_0), conhecida como "Receita de Referência", é calculada usando as tarifas econômicas aprovadas na Data de Referência Anterior (DRA) e o "Mercado de Referência". Esses dados são obtidos pelo Sistema de Acompanhamento de Informações de Mercado para Regulação Econômica (SAMP) (ANEEL, 2022).

Para calcular a RA_0 , são usadas as tarifas aprovadas na DRA, sejam elas integrais ou com desconto, conforme a última revisão tarifária. O faturamento do mercado no último mês do

"Período de Referência" é estimado com base no mês anterior. Se a distribuidora vende energia por valores menores que os aprovados pela ANEEL, deve usar a tarifa econômica integral, sem ajustes compensatórios futuros (ANEEL, 2022).

A perda de receita devido a descontos concedidos é coberta pela Conta de Desenvolvimento Energético (CDE). Se usadas tarifas com desconto, estas são aumentadas para compensar a perda de receita, sem ajustes futuros por mudanças no mercado (ANEEL, 2022).

3.2.3 Parcela A

O Valor da Parcela "A" (VPA_0) é calculado com base nas condições vigentes na DRA e no "Mercado de Referência". Este cálculo envolve a compra de energia elétrica, conexão e uso dos sistemas de transmissão e outros itens tarifários (ANEEL, 2022).

A Energia Elétrica Comprada inclui o volume adquirido para atender consumidores, considerando perdas técnicas e não-técnicas e perdas associadas ao transporte. O Preço Médio de Repasse na DRA é o resultado da divisão entre o custo total e o montante total de energia comprada (ANEEL, 2022).

O Valor da Parcela "A" (VPA_1) é calculado com base nas condições vigentes na Data do Reajuste em Processamento (DRP) e inclui energia comprada por contratos anteriores e novos, energia de Itaipu, geração própria, PROINFA, Angra 1 e 2, usinas com concessões renovadas, e sistemas isolados. O cálculo também considera o uso dos sistemas de transmissão e os demais itens tarifários (ANEEL, 2022).

3.2.4 Parcela B

O Valor da Parcela "B" (VPB_0) é calculado considerando as condições vigentes na Data de Referência Anterior (DRA) e o Mercado de Referência (ANEEL, 2022):

$$VPB_0 = RA_0 - VPA_0 \quad (3.1)$$

Onde:

- RA_0 é a Receita Anual na DRA, também chamada de "Receita de Referência";
- VPA_0 é o Valor da Parcela "A" nas condições vigentes na DRA.

O Valor da Parcela "B" (VPB_1) na Data do Reajuste em Processamento (DRP) é calculado da seguinte forma (ANEEL, 2022):

$$VPB_1 = VPB_0 \times (IGPM - X) \quad (3.2)$$

Onde:

- VPB_0 é o Valor da Parcela "B" na DRA;
- IGPM é o Índice Geral de Preços do Mercado IGP-M
- X, ou Fator X, é subtraído do IGPM para compartilhar com os usuários e consumidores da distribuidora os ganhos de eficiência empresarial e competitividade, contribuindo para a modicidade tarifária.

4 FERRAMENTAS UTILIZADAS

4.1 Linguagem de Programação Python

Python é uma linguagem de programação conhecida por sua simplicidade e facilidade de aprendizado. Uma das grandes vantagens do Python é sua ampla adoção em diversos campos, desde desenvolvimento web até análise de dados e inteligência artificial. Além disso, a orientação a objetos em Python permite uma abordagem mais modular e organizada para o desenvolvimento de software, facilitando a reutilização de código e a manutenção de sistemas complexos.

A orientação a objetos, como paradigma de programação, oferece diversas vantagens em termos de estruturação de código e encapsulamento de dados. Em Python, a implementação de classes e objetos permite a criação de hierarquias de classes, herança, polimorfismo e encapsulamento, o que contribui para um código mais legível, modular e escalável. Isso também facilita a colaboração em equipe e o desenvolvimento de sistemas robustos e flexíveis.

Assim, ao combinar as vantagens do Python como linguagem de programação com os princípios da orientação a objetos, os desenvolvedores podem criar aplicações eficientes, com código mais limpo, estruturado e fácil de manter, promovendo uma maior produtividade e qualidade no desenvolvimento de software (LUTZ, 2013).

De acordo com o site do PYPL (POPULARITY OF PROGRAMMING LANGUAGE (PYPL), 2023), a linguagem de programação Python foi a primeira colocada no ranking de linguagens mais pesquisadas ao redor do mundo. Esse site analisa as tendências de busca no domínio da *Google* e plota uma tabela com as principais linguagens do mundo.

A comunidade de programadores em linguagem python já publicou mais de 490 mil projetos no repositório *open source* Python Package Index (PyPI) (PYPI, 2023). Nesse domínio, é possível encontrar uma infinidade de bibliotecas úteis para diversos trabalhos, inclusive na análise de dados.

4.1.1 Jupyter Notebook

O Jupyter Notebook é uma aplicação web de código aberto que permite a criação e compartilhamento de documentos que contêm código executável, visualizações e texto explicativo, como equações, descrições e visualizações de dados. Ele é amplamente utilizado em ciência de dados, aprendizado de máquina, análise de dados e ensino, permitindo que os usuários combinem código e narrativa em um único documento. Suporta várias linguagens de

programação, incluindo Python, R e Julia, e oferece uma interface interativa onde os resultados da execução do código são exibidos diretamente no navegador, facilitando a exploração e a visualização de dados (JUPYTER, 2024).

4.2 Biblioteca Pandas

O Pandas é uma biblioteca gratuita e de código aberto, que disponibiliza estruturas de dados eficientes e ferramentas de análise de dados acessíveis para a linguagem de programação Python (PANDAS, 2023).

É uma ferramenta poderosa e amplamente utilizada para análise de dados em Python. Ela oferece estruturas de dados flexíveis, como o *DataFrame*, que facilitam a manipulação e a organização de conjuntos de dados complexos. Uma das principais vantagens do Pandas é sua capacidade de lidar eficientemente com dados em formatos tabulares, permitindo operações como filtragem, seleção, agrupamento e transformação de dados de forma intuitiva e eficaz.

Além disso, o Pandas é especialmente útil para trabalhar com dados provenientes de diferentes fontes, como arquivos *Comma-separated values* (CSV), Excel, *Structured Query Language* (SQL) e até mesmo dados provenientes da web. Isso torna o Pandas uma ferramenta versátil para análise de dados em diversos contextos, desde análise exploratória até a preparação de dados para modelagem e machine learning.

Outra vantagem significativa do Pandas é sua documentação abrangente e uma comunidade ativa de usuários, o que facilita o aprendizado e a resolução de problemas. Com recursos poderosos para manipulação de dados e uma ampla gama de funcionalidades, o Pandas se tornou uma escolha popular entre cientistas de dados, analistas e desenvolvedores que trabalham com análise e manipulação de dados em Python (MCKINNEY, 2023).

4.2.1 Processo de ETL

O "Extrair, transformar e carregar" (ETL) é um processo de integração de dados que reúne informações de diversas fontes em um armazenamento único e consistente, geralmente em um armazém de dados. Surgido na década de 1970 com o aumento da popularidade dos bancos de dados, o ETL se tornou fundamental para projetos de armazenamento de dados, fornecendo a base para análise de dados e fluxos de trabalho de machine learning. Esse processo envolve a extração de dados de sistemas legados, a limpeza para melhorar a qualidade e consistência, e a

carga dos dados em um banco de dados de destino, sendo essencial para atender a necessidades específicas de negócios e realizar análises avançadas (IBM, 2023).

Para o presente estudo, utilizaremos a biblioteca Pandas (4.2) para desempenhar o ETL dos dados, quando necessário.

4.2.2 Banco de dados SQLite

O SQLite é uma biblioteca incorporada que oferece um mecanismo de banco de dados SQL autônomo, dispensando a necessidade de servidor, sem exigir configuração, e completamente transacional. Seu código está disponível no domínio público, o que significa que pode ser utilizado gratuitamente para qualquer propósito, seja ele comercial ou privado. O SQLite é um dos bancos de dados mais amplamente utilizados no mundo, presente em inúmeras aplicações, incluindo vários projetos de grande visibilidade (SQLITE, 2023).

Trabalhar com dados em um banco de dados SQLite em comparação com arquivos CSV, por exemplo, oferece várias vantagens, especialmente em cenários em que é necessário lidar com dados de forma mais estruturada e realizar consultas mais complexas. Algumas das vantagens de utilizá-lo são:

- **Índices e otimizações:** O SQLite suporta índices, o que pode acelerar significativamente as consultas, especialmente em grandes conjuntos de dados. Isso melhora o desempenho das operações de busca e filtragem.
- **Transações e atomicidade:** O SQLite oferece suporte a transações ACID (Atomicidade, Consistência, Isolamento e Durabilidade), garantindo que as operações de banco de dados sejam realizadas de maneira segura e confiável, mesmo em caso de falhas.
- **Escalabilidade:** Em cenários nos quais o volume de dados é grande e as operações são complexas, um banco de dados SQLite pode oferecer melhor escalabilidade do que manipular dados em arquivos CSV, especialmente quando se trata de consultas complexas e agregações.

4.2.3 Power BI

O Power BI é uma ferramenta de análise de dados que permite criar relatórios interativos e painéis dinâmicos. Com ele, é possível transformar dados brutos em insights valiosos e compartilhá-los com outras pessoas. O Power BI serve para facilitar a tomada de

decisões baseadas em dados, melhorar o desempenho dos processos e comunicar informações de forma eficaz (MICROSOFT, 2023).

Uma de suas principais funcionalidades é a capacidade de conectar-se a uma ampla variedade de fontes de dados, desde arquivos Excel e bancos de dados SQL até serviços em nuvem como *Azure* e *Google Analytics*. Essa flexibilidade facilita a consolidação de informações dispersas, proporcionando uma visão unificada e abrangente dos dados (MICROSOFT, 2023).

Além de suas capacidades de integração, o Power BI oferece ferramentas avançadas de visualização de dados. Usuários podem criar gráficos, tabelas, mapas geográficos e outros tipos de visualizações interativas que facilitam a compreensão e interpretação dos dados. A plataforma também suporta a criação de relatórios customizáveis, permitindo que os usuários ajustem visualizações e filtros de acordo com suas necessidades específicas. Com recursos como o Power Query, é possível realizar transformações e limpeza de dados de forma intuitiva, garantindo que as análises sejam baseadas em dados precisos e consistentes (MICROSOFT, 2023).

Por fim, essa é uma ferramenta que permite que profissionais de diversas áreas, sem necessidade de conhecimentos avançados em TI, possam explorar e interpretar dados de maneira eficiente. A interface amigável e intuitiva contribui para uma rápida curva de aprendizado, tornando a ferramenta acessível a um público amplo. Outrossim, o Power BI permite a colaboração em tempo real, possibilitando que equipes trabalhem conjuntamente na criação e análise de relatórios, promovendo uma cultura de dados mais integrada e colaborativa (MICROSOFT, 2023).

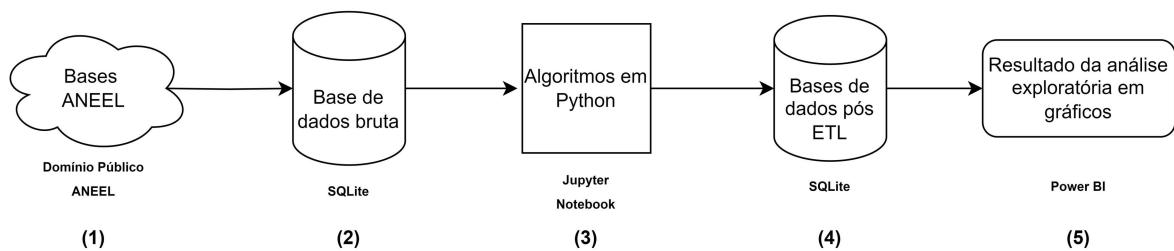
5 METODOLOGIA

Este capítulo apresenta a abordagem metodológica adotada para a realização do presente estudo, detalhando as etapas e procedimentos empregados na coleta, análise e interpretação dos dados. A escolha de uma metodologia adequada é fundamental para atingir os objetivos propostos e responder às questões de pesquisa formuladas. Neste contexto, são delineados os métodos e técnicas empregados, bem como a justificativa para suas escolhas.

Em resumo, a ideia metodológica envolve obter os dados, disponibilizá-los em um ambiente apropriado para serem trabalhados, fazer refinamentos por meio de ETLs tanto a nível de banco de dados como em processos intermediários utilizando Python, disponibilizar as bases refinadas em outro banco e, por fim, utilizar o *software* de visualização gráfica com os dados refinados afim de se obter conclusões sobre o tema debatido.

No diagrama da Figura 5 apresenta-se uma visão macro de toda a metodologia aplicada neste capítulo.

Figura 5 – Estrutura Macro da Metodologia Adotada



Fonte: Elaborado pelo autor.

Em resumo, cada uma das cinco etapas podem ser descritas por:

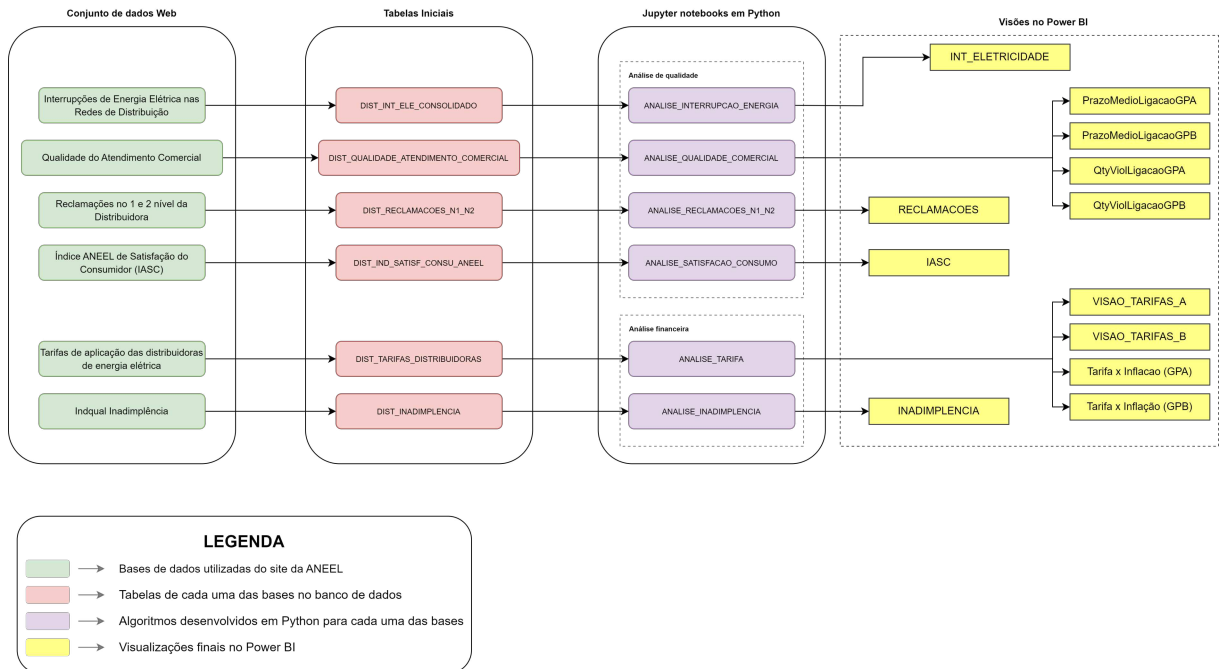
- (1) Dentre os mais de 60 conjuntos de dados disponibilizados pela ANEEL, foram escolhidos 6 conjuntos referentes ao sistema de distribuição de energia elétrica brasileiro que pudessem servir de base para analisarmos a qualidade e os custos envolvidos para o consumidor final;
- (2) Após o *download* das bases, foi feito o *upload* dessas informações para um ambiente de banco de dados. Além disso, como em algumas bases eram feitas segregações entre períodos de anos, alguns *scripts* em SQL foram utilizados para unificar em apenas uma tabela dados da mesma natureza mas com anos diferentes;
- (3) Foram desenvolvidos algoritmos em Python para cada uma das bases utilizadas para trabalhar os dados e disponibilizar uma nova base trabalhada pronta para alimentar posteriormente os gráficos finais do estudo;

- (4) As novas bases são salvas novamente em um banco de dados para se conectar com o Power BI;
- (5) Por fim, os gráficos finais da análise exploratória são construídos no *software* Power BI.

Em complemento a visão macro da Figura 5, na Figura 6 temos um diagrama que mostra o caminho seguido por cada um dos conjuntos de dados desde a origem até o resultado final. O mesmo diagrama está disponível em anexo (B) de forma a facilitar a visualização e o seu entendimento.

Os nomes utilizados são os mesmos encontrados nos arquivos disponibilizados.

Figura 6 – Estrutura Macro Expandida



Fonte: Elaborado pelo autor (2024).

Por fim, toda a parte de desenvolvimento dos algoritmos e suas dependências se encontram disponíveis em um repositório público do GitHub.

5.1 Base de dados analisada

Nesta seção será elucidado o processo de obtenção dos dados de Distribuição de Energia Elétrica do repositório *open source* da ANEEL.

5.1.1 Escolha e obtenção dos dados

A ANEEL disponibiliza 17 conjuntos de dados referentes ao tema de distribuição de energia elétrica (ANEEL, 2023b). Destes, para o desenvolvimento da temática do presente trabalho, foram escolhidas as bases dos seguintes conjuntos:

- **Tarifas de aplicação das distribuidoras de energia elétrica:** Apresenta os valores das Tarifas de Energia - TE e das Tarifas de Uso do Sistema de Distribuição - TUSD, resultantes dos processos de reajustes tarifários das distribuidoras de energia elétrica.
- **Interrupções de Energia Elétrica nas Redes de Distribuição:** Dados de todas as interrupções de energia elétrica ocorridas nas redes de distribuição de energia elétrica do país. Não constam as interrupções ocorridas em áreas sob gestão de permissionárias de serviço público (cooperativas).
- **Qualidade do Atendimento Comercial:** Com o objetivo de acompanhar o cumprimento dos prazos de prestação dos serviços estabelecidos nas Regras de Prestação do Serviço Público de Distribuição de Energia Elétrica (Resolução Normativa – REN nº 1.000, de 2021), a ANEEL recebe o relatório da apuração do cumprimento dos prazos de prestação dos serviços e das suspensões indevidas.
- **Índice de Inadimplência:** Os dados de inadimplência são disponibilizados para permitir a análise do *aging list* das distribuidoras por classe de consumo em relação à receita faturada além da quantidade de suspensões por inadimplemento.
- **Reclamações no 1º e 2º nível da Distribuidora:** Base de dados com a quantidade de reclamações dos consumidores às distribuidoras, por tipologia.
- **IASC:**
O Índice de Satisfação do Consumidor pela ANEEL.

Cada um dos conjuntos de dados supracitados são representados por uma tabela no banco de dados conforme o diagrama da Figura 6. No apêndice A estão dispostos os links dos dicionários de todos os conjuntos de dados utilizados, disponibilizados no site da ANEEL.

Já sobre a obtenção dos dados, o repositório permite diversas formas de obtenção destes. Primeiramente foi tentado extrair via requisição da API disponibilizada por eles por ser a solução mais robusta e que traria o maior grau de automatização na esteira da análise exploratória, mas até a data deste trabalho, 26 de dezembro de 2023, a API limitava a requisições de no máximo 30 mil linhas quando havia bases com mais de 1 milhão de linhas. Dessa forma,

tendo em vista que isso comprometeria a análise completa, foi optado por baixar os conjuntos de dados em arquivos do tipo CSV.

5.1.2 Estruturação dos dados no banco SQLite

Dado o volume extenso de todas as bases escolhidas, seria inviável continuar trabalhando com arquivos de tipo CSV. Dessa forma, foi optado por subir todos dados no banco de dados relacional SQLite.

Dessa forma, após concluído o *download* de todos os arquivos, foi utilizando um script em Python para subi-los como tabelas com nomes padronizadas no banco de dados.

Ademais, como alguns conjuntos de dados eram divididos por vários arquivos CSV com *ranges* de data e o ideal para o trabalho era que as informações se concentrassem em uma única tabela, alguns *scripts* em SQL foram desenvolvidos para agrupar e selecionar apenas as colunas consideradas úteis para a análise.

5.2 Algoritmo desenvolvido

Com os dados dispostos no banco de dados, foram desenvolvidos algoritmos para trabalhar cada uma das tabelas escolhidas. A ideia dos algoritmos montados é desempenhar a etapa de ETL descrita em 4.2.1 e, em seguida, popular o banco de dados que irá alimentar os *dashboards* do Power BI.

Para essa abordagem, optou-se por utilizar o ambiente de *Jupyter Notebook* por entender que esse formato seria didaticamente melhor tanto no desenvolvimentos dos scripts como na passagem deles para utilização futura em outros trabalhos ou análises.

Vale ressaltar também que a análise se concentra em compreender, dentre os diversos conjuntos de dados, a realidade da temática desenvolvida para um estado da região Nordeste, com a sua respectiva distribuidora de energia. Dessa forma, em todos os algoritmos desenvolvidos nos quais o campo "Agente" está disponível e tem significado prático, essa filtragem padrão é realizada concomitantemente às demais.

Nas subseções seguintes serão descritas as metodologias de ETL adotada para cada um dos conjuntos de dados utilizados.

5.2.1 *Análise tarifária*

Utilizando a tabela **DIST_TARIFAS_DISTRIBUIDORAS**, foram filtrados todos os editais de tarifação aprovados pela ANEEL para o estado analisado e aplicado um algoritmo de serialização dos valores em anos fechados para possibilitar uma análise anual.

5.2.1.1 *Premissas adotadas*

- Análise feita para cada tipo de grupo consumidor abarcando os diferentes tipos de modalidade tarifária disponíveis para cada um;
- A tarifa utilizada como referência para essa análise foi a TUSD disponibilizada na coluna "VirTUSD" da tabela da ANEEL. Como o estudo em questão se concentra no tema de "Distribuição de Energia", a parcela referente a geração de energia foi desconsiderada.
- Para grupos com tarifação diferenciada em horários de ponta e fora ponta, baseado na curva média de cargas da região Nordeste (4), foi adotado um peso de 60% para o valor da tarifa ponta e 40% para o da tarifa fora ponta.
- Foram desconsideradas, para esta análise, os seguintes postos tarifários: "Baixa Renda", "Serviço público de irrigação rural", "Cooperativa de eletrificação rural", "Geração", "Fora ponta seca", "Fora ponta úmida", "Ponta seca", "Ponta úmida", "Azul ABRACE CATIVO", "Verde ABRACE CATIVO", "Azul ABRACE LIVRE", "Convencional ABRACE" e "Verde ABRACE LIVRE"
- Para validar o aumento ou a redução, ano contra ano, das tarifações aplicadas para os consumidores da concessionária, foi utilizada uma situação de contorno arbitrária. Como a análise será feita tendo como base o aumento percentual entre os anos no período analisado, a situação de contorno é um valor hipotético base para cada um dos grupos consumidores para entender qual seria a realidade do preço da tarifa em cada ano. Vale ressaltar que o valor base escolhido de nada interfere no resultado final, mas sua definição é importante haja vista as especificidades de cada um dos grupos. A situação de contorno adotada está disposta na Tabela 3:

Tabela 3 – Situação de contorno

Grupo	Tensão kV	Consumo kWh
A1	230	100000
A3	69	50000
A4	25	10000
B1	-	10000
B2	-	10000
B3	-	10000
B4	-	10000

Fonte: elaborado pelo autor (2023)

5.2.1.2 Descrição do algoritmo

A ideia do algoritmo desenvolvido surge da necessidade de ser possível analisar o valor tarifário das resoluções homologatórias vigentes em cada um dos anos. Como o período de vigência de cada uma das resoluções não segue um padrão lógico, para que uma análise anual fosse possível, seria necessário distribuir, por meio de uma média ponderada, os valores aplicáveis por cada resolução dentro dos dias de cada um dos anos.

Nesse sentido, primeiramente, a lógica por trás do algoritmo é coletar, em cada uma das resoluções homologatórias, a quantidade de dias em que ficou vigente, os anos e o valor tarifário permitido. Com isso em mãos, para cada um dos anos, é feita a média ponderada dos valores homologados de tarifa baseado no número de dias vigentes de cada portaria. Dessa forma, é obtido um valor final anual da tarifa que, em média, foi aplicada para cada ano.

Além disso, com o valor de tarifa anual obtido, é aplicada uma situação de contorno arbitrária simulando um mesmo consumo de energia em cada um dos anos. Como a tarifa do Grupo A leva em consideração a tensão, foram também definidos valores padrões para cada um dos subgrupos. Vale ressaltar que, por mais que esses valores arbitrários não condizam com a realidade, como o intuito da análise é verificar o aumento ou a diminuição percentual dos custos referentes a distribuição ano contra ano, esse modelo é suficiente já que utiliza um mesmo parâmetro para toda a amostra de anos analisados.

Com a situação de contorno aplicada aos valores de tarifas obtidos, o algoritmo calcula o aumento percentual dos custos referentes a distribuição de energia elétrica no estado analisado, ano contra ano.

Por fim, em vias de refinar essa análise, o algoritmo também disponibiliza uma comparação anual com os índices financeiros que medem o aumento dos preços no mercado

(3.1).

5.2.2 *Análise da inadimplência*

Utilizando a tabela **DIST_INADIMPLENCIA**, foram filtrados os índices de inadimplência dos consumidores com a distribuidora de energia analisada ao longo dos anos. Além disso, o algoritmo calculou o aumento percentual desse montante ano contra ano.

5.2.3 *Análise da qualidade comercial*

Utilizando a tabela **DIST_QUALIDADE_ATENDIMENTO_COMERCIAL**, foram filtrados todos os indicadores de atendimento comercial da concessionária de energia. Após isso, o algoritmo agrupa, dentro de um mesmo indicador, os consumidores do grupo A e grupo B.

Dentre os indicadores disponibilizados na base, para incorporar no estudo de qualidade comercial da distribuição de energia, foram selecionados os seguintes tópicos:

- Quantidade de violações nos prazos de ligação de consumidores do Grupo B;
- Quantidade de violações nos prazos de ligação de consumidores do Grupo A;
- Prazo médio de ligação por consumidor do grupo B;
- Prazo médio de ligação por consumidor do grupo A;

5.2.4 *Análise das reclamações N1 e N2*

Utilizando a tabela **DIST_RECLAMACOES_N1_N2**, foram filtradas todas as reclamações feitas acerca dos serviços prestados pela concessionária. Após isso, o algoritmo agrupa, dentro de um mesmo indicador, os consumidores do grupo A e grupo B para uma visão do total das reclamações feitas por ano.

Como a ideia é analisar a qualidade comercial dos serviços de distribuição prestados, alguns motivos de reclamações foram removidos da base pelo algoritmo por se tratar de aspectos secundários dos serviços prestados, aqui com pouca relevância. A lógica utilizada para efetuar essa filtragem está disposta a seguir:

```

1 def removeMotivoInvalido(df): # Remove as reclamações com motivos sem
   aplicação para esta an lise
2 df['MotivoValido'] = ''
3 listaRemocao = ['Elogio', 'Informação', 'Alteração Cadastral', '
   Denúncia', 'Custo de Disponibilidade', 'Sugestão', 'Problemas
   de instalação interna na unidade consumidora', 'Prazos', '
   Indisponibilidade de Agência/Postos e Canais de atendimento/
   Atend. Telefônico/Serviço de Arrecadação', 'Cobrança Indevida
   de Atividade Acessória']
4
5 df['MotivoValido'] = ~df['DescReclamacao'].isin(listaRemocao)

```

5.2.5 Análise das interrupções

Utilizando a tabela **DIST_INT_ELE_CONSOLIDADO**, foram filtradas todas as interrupções de energia ocorridas aos consumidores de energia. Com isso, o algoritmo calcula, para cada um dos grupos consumidores disponíveis, a duração em horas de cada interrupção.

Ademais, um ponto importante no contexto dessa análise é que, de acordo com os dados da base, cada interrupção pode afetar mais de 1 consumidor. Dessa forma, para trazer luz ao impacto de interrupções que acometem um grande número de consumidores, o algoritmo calcula um indicador chamado "TempoInterrupcaoGlobal" que representa a multiplicação da quantidade de clientes impactados vezes o tempo que durou a falta de energia.

Novamente, mesmo que o significado prático desse indicador não faça tanto sentido ao analisarmos pontualmente caso a caso, como estamos analisando o dado na ótica de comparar ano contra ano, por essa perspectiva será possível analisar um aumento ou decréscimo percentual desse indicador, caso haja.

Um ponto negativo é que a ANEEL apenas disponibiliza esse dado a partir de 2018.

5.2.6 Análise do IASC

Utilizando a tabela **DIST_IND_SATISF_CONSU_ANEEL**, já são disponibilizados os valores obtidos (calculados pela metodologia) de todos os indicadores do índice (2.2.1).

Dessa forma, no *dataframe* final, o algoritmo deixa apenas as colunas com o valor

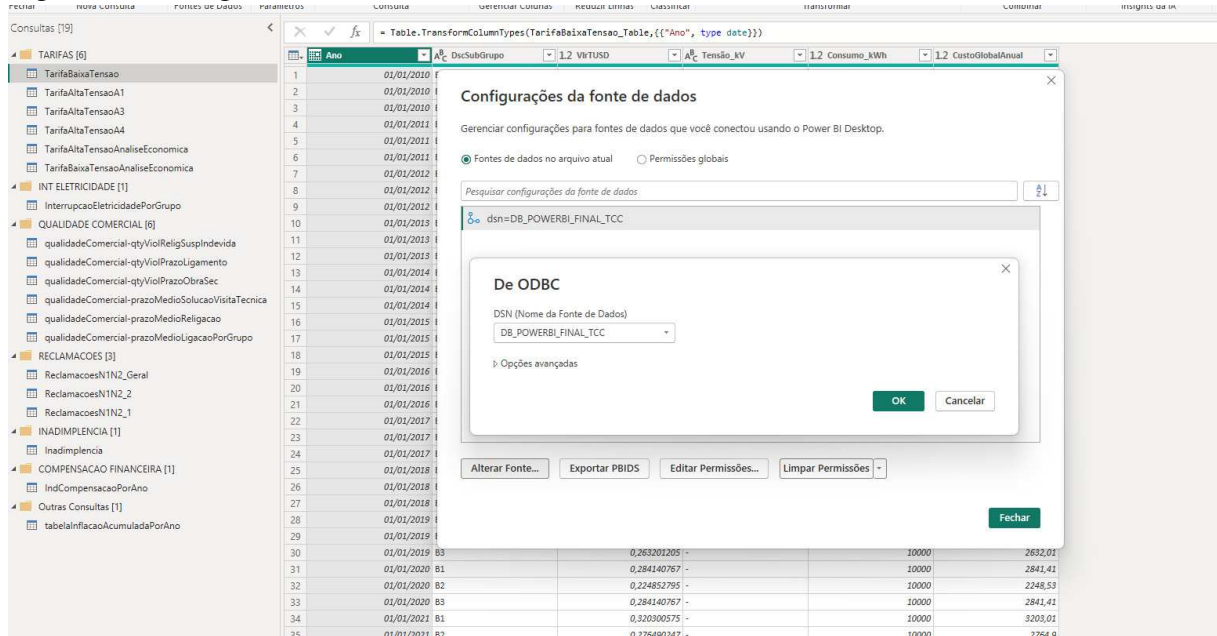
final de cada um dos indicadores, filtra o *range* de anos e separa a coluna referente a colocação IASC da concessionária de energia analisada frente as demais existentes no Brasil. Por fim, esse *dataframe* é levado para uma tabela da base de dados que alimenta o Power BI.

5.3 Visualização dos dados modelados

Ao terminar de rodar todo o algoritmo de processamento das bases brutas, são disponibilizadas tabelas refinadas no banco de dados "BD_POWER_BI_FINAL.db" que irá alimentar os dashboards no Power BI.

Para que fosse possível puxar essas tabelas direto do banco de dados para o Power Query, foi configurada uma conexão via ODBC conforme ilustra a Figura 7.

Figura 7 – Configurando o ODBC



Fonte: Próprio autor.

Com isso, foi possível obter os resultados da análise plotando os dados por meio dos gráficos disponíveis na ferramenta.

Por fim, alguns ajustes finos como a modificação do nome de algumas colunas e a formatação do tipo do dado ali presente foram feitos visando otimizar a visualização dos gráficos.

6 RESULTADOS

Nesta seção serão apresentados os resultados da análise exploratória de dados modelada no capítulo anterior (5).

Em linha com a ideia proposta nesse trabalho, a apresentação dos resultados será dividida nas seguintes etapas:

1. Análise financeira do preço tarifário praticado pela distribuidora de energia analisada ano a ano.
2. Situação de inadimplência dos consumidores de energia
3. Indicadores de qualidade no serviço prestado pela distribuidora.

Dessa forma, conseguiremos entender os possíveis impactos do aumento no preço das tarifas praticado pela concessionária na qualidade do serviço prestado, frente a diferentes níveis de inadimplência ao longo dos anos.

6.1 Resultados da análise de tarifação aplicada ao longo dos anos

Conforme a metodologia aplicada e as diferenças na natureza das tarifações dos consumidores do Grupo A e do Grupo B, os resultados também serão apresentados segregados afim de obter melhores *insights*.

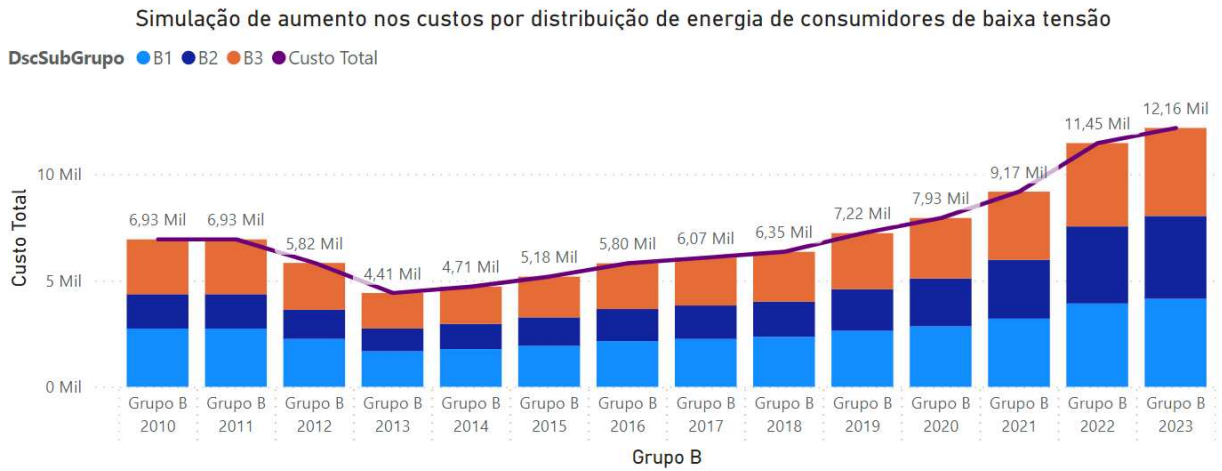
Vale ressaltar que os "custos de distribuição" considerados ao longo dessa seção são referentes ao componente TUSD, explicado na seção (2.5), que compõe o preço final da tarifa paga pelo consumidor final e que foi utilizada na metodologia como parâmetro para a análise no aumento do preços praticados pela concessionária de energia (5.2).

Essa é uma perspectiva na ótica do consumidor e não deve ser entendida com os custos que a concessionária de energia tem com a distribuição de energia, e sim o valor que ela cobra aos consumidores referente a essa atuação.

6.1.1 Consumidores de baixa tensão (Grupo B)

Conforme a Figura 8 abaixo, aplicando o modelo tarifário descrito em (5.2), é possível perceber uma queda no valor simulado dos custos envolvidos com distribuição de energia para o grupo de consumidores do tipo B apenas entre os anos 2012 e 2013. logo em seguida, de acordo com o modelo desenvolvido, os custos foram sempre crescentes.

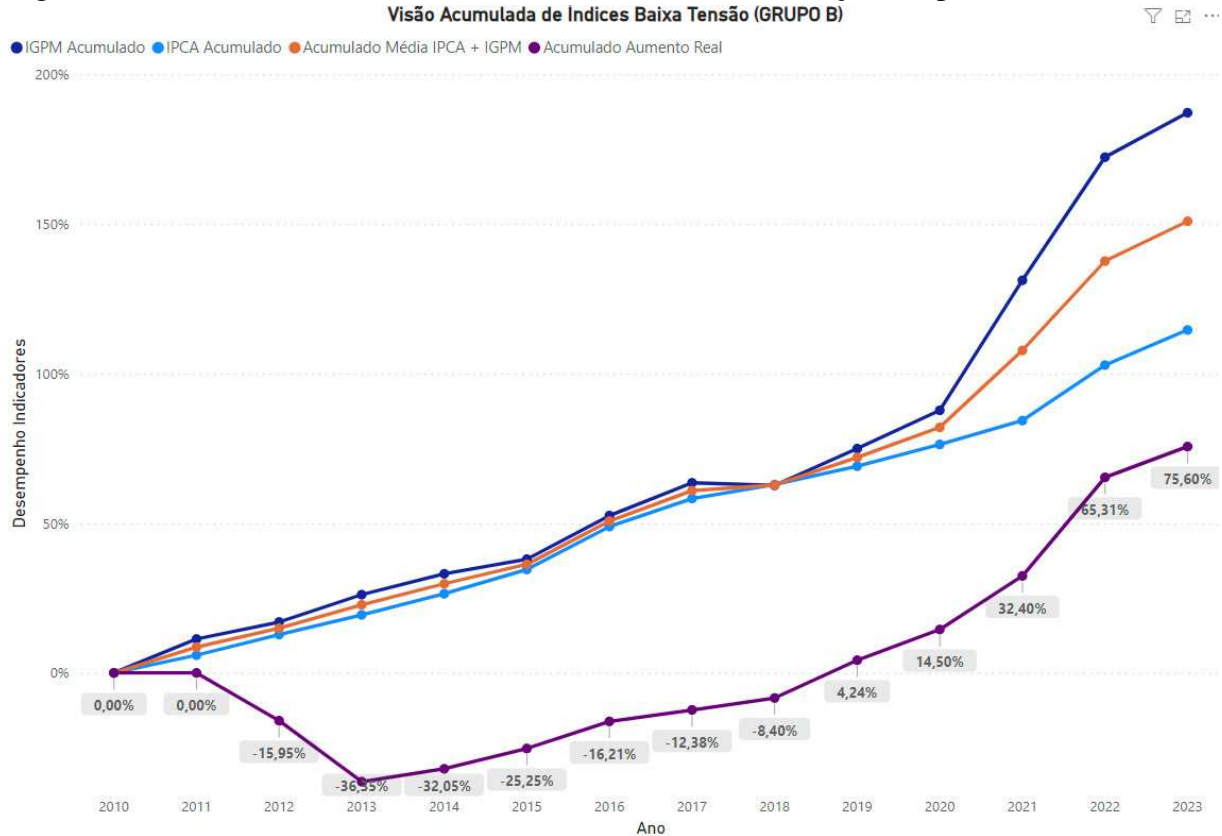
Figura 8 – Simulação Custos de Distribuição Grupo B



Fonte: Elaborado pelo autor (2023)

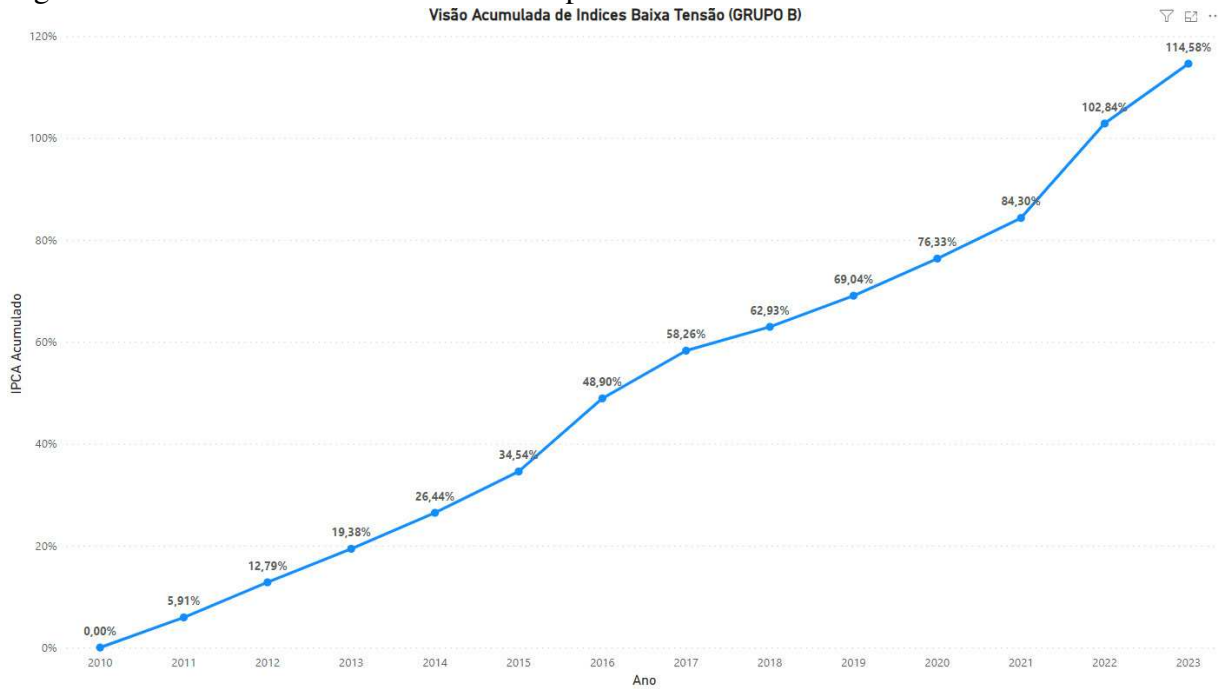
Além disso, fazendo uma análise do acumulado dos anos conforme as Figuras 9, 10 e 11, temos que, no período entre 2010 e 2023, os custos referentes a distribuição de energia, conforme o modelo, tiveram um aumento de 75,60% no período, enquanto o IPCA teve um aumento de 114,58% e o IGP-M de 187,15%.

Figura 9 – Acumulado do Aumento Real nos Custos de Distribuição Grupo B



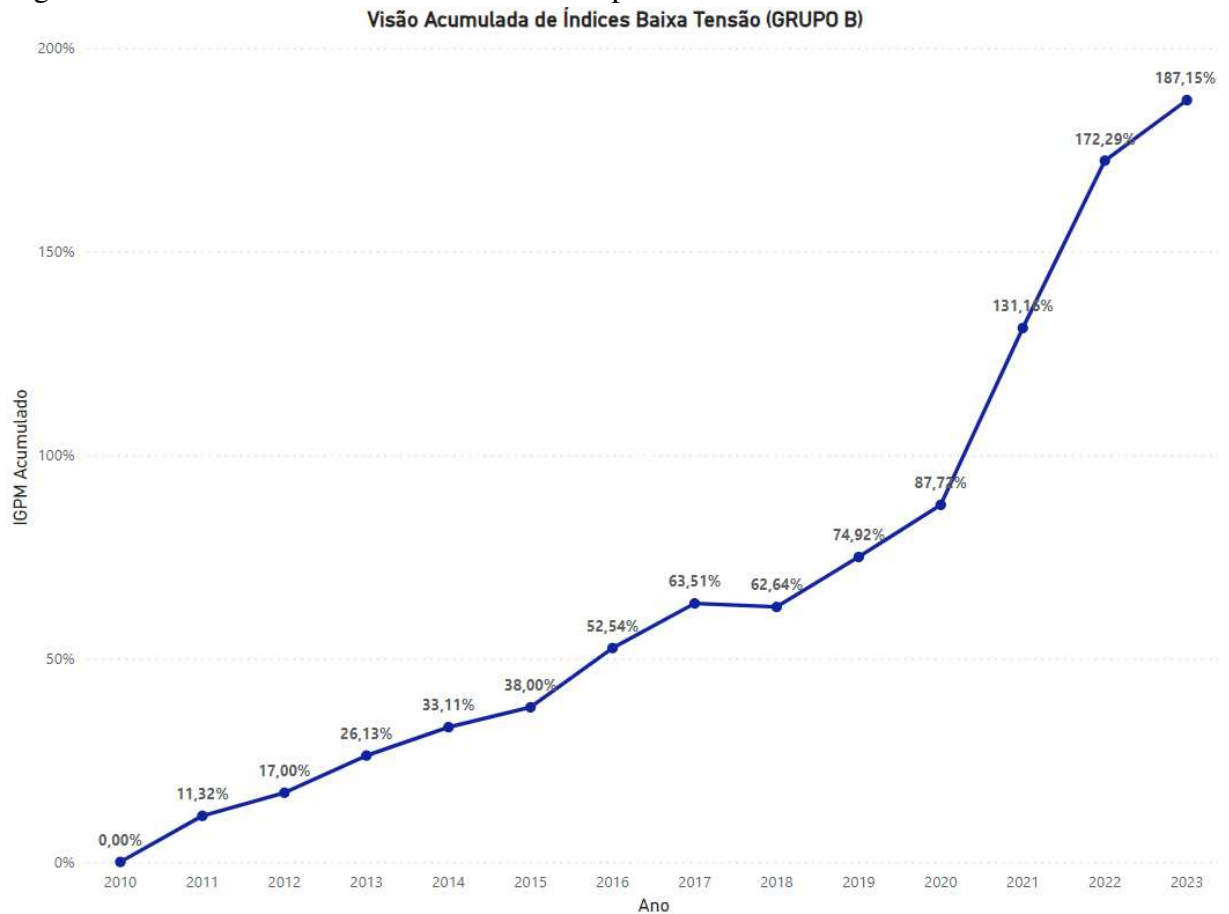
Fonte: Elaborado pelo autor (2023).

Figura 10 – Acumulado do índice IPCA no período



Fonte: Elaborado pelo autor (2023).

Figura 11 – Acumulado do índice IGPM-M no período



Fonte: Elaborado pelo autor (2023).

Na análise deste período, os custos de distribuição de energia para o grupo de consumidores tipo B permaneceram consistentemente inferiores aos índices financeiros acumulados de aumento de preços. Isso sugere um desafio para a concessionária em manter ou aprimorar o nível dos serviços prestados, especialmente diante do decréscimo real no valor cobrado.

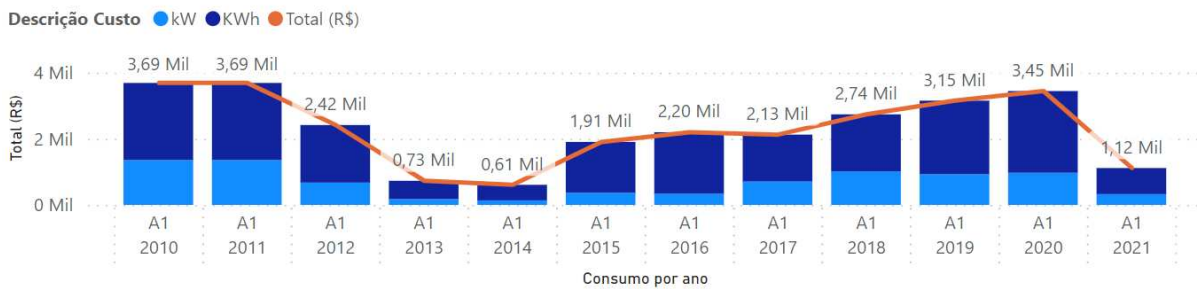
6.1.2 Consumidores de Alta Tensão (Grupo A)

Fazendo a mesma análise para o grupo de consumidores do tipo A, temos pelas Figuras 12 e 13 que os custos sofreram quedas, no modelo, entre 2012 e 2014, para os consumidores do grupo A1 e A3. Para os demais anos, sofreram um aumento.

Para os consumidores do grupo A4, Figura 14, os custos sofreram uma leve queda no ano de 2013, mas foram crescentes ao longo do período analisado.

Figura 12 – Simulação Custos de Distribuição Grupo A1

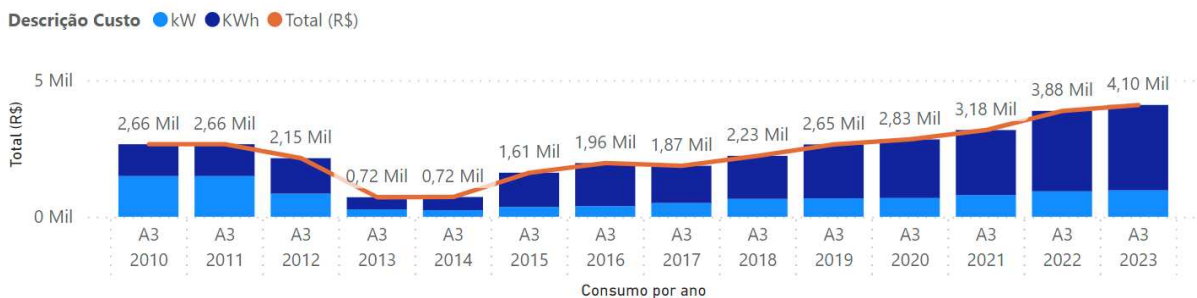
Simulação de aumento nos custos por distribuição de energia de consumidores de Alta Tensão (Grupo A1)



Fonte: Elaborado pelo autor (2023).

Figura 13 – Simulação Custos de Distribuição Grupo A3

Simulação de aumento nos custos por distribuição de energia de consumidores de Alta Tensão (Grupo A3)



Fonte: Elaborado pelo autor (2023).

Figura 14 – Simulação Custos de Distribuição Grupo A4

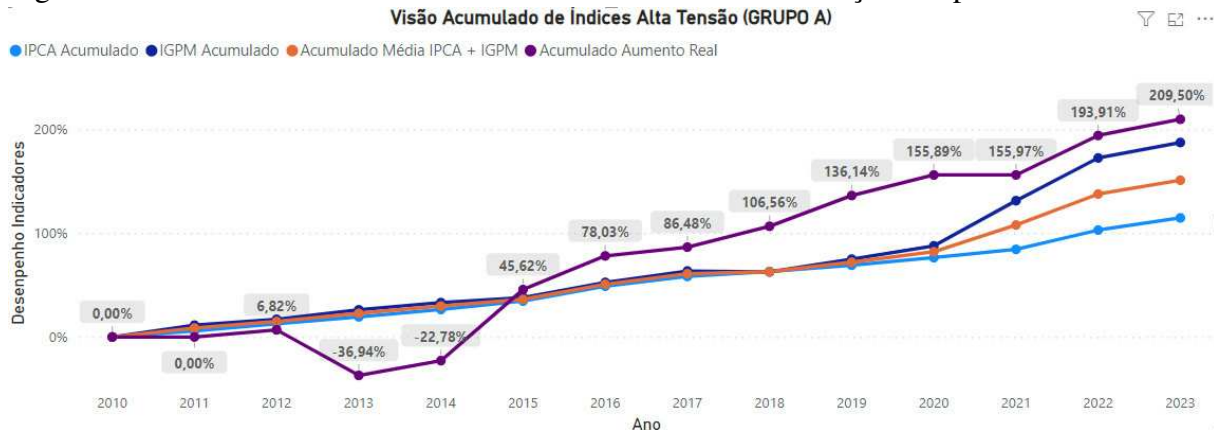


Fonte: Elaborado pelo autor (2023).

Analisando por uma perspectiva do aumento acumulado no período, Figura 15, nota-se que até o ano de 2023 houve um aumento de 209,50% na simulação dos custos referentes a distribuição de energia.

Além disso, percebe-se que apenas entre os anos 2011 e 2014 o indicador do acumulado dos custos ficou abaixo dos acumulados dos índices financeiros de aumento dos preços, indicando um aumento real no preço da energia ao consumidor, o que poderia acarretar, por sua vez, uma melhora na qualidade comercial do serviço oferecido.

Figura 15 – Acumulado do Aumento Real nos Custos de Distribuição Grupo A



Fonte: Elaborado pelo autor (2023).

6.2 Resultados da análise de inadimplência dos consumidores para a distribuidora de energia elétrica analisada

Antes de analisar os indicadores de qualidade dos serviços prestados pela concessionária, é interessante trazer luz ao índice de inadimplência dos consumidores.

Analisando a Figura 16 percebe-se que entre os anos de 2013 e 2014 o índice apresentava-se estável. Fatores políticos podem ter acarretado esse cenário já que, mesmo passando por uma crise hídrica nacional, o preço da energia foi mantido baixo por meio de

uma lei que permitia que as empresas de energia podiam antecipar a renovação das concessões programadas para os próximos anos, dispensando a necessidade de participar de licitações, mediante a condição de reduzir as tarifas (BRASIL, 2013).

Após esse período, os aumentos consecutivos nas tarifas de energia juntamente com a crise econômica que assolou o Brasil nesse período podem ter significado esse salto de 68,42% na inadimplência, que foi sendo reduzido até o menor patamar em 2020 com a melhora da economia do país como um todo.

A partir daí, a crise ocasionada pela epidemia do Coronavírus COVID-19 pode justificar o aumento de 272,28% no índice em 2021 e na manutenção em níveis elevados no ano seguinte.

Figura 16 – Inadimplência financeira por ano



Fonte: Elaborado pelo autor (2023).

6.3 Resultados da análise dos indicadores de qualidade

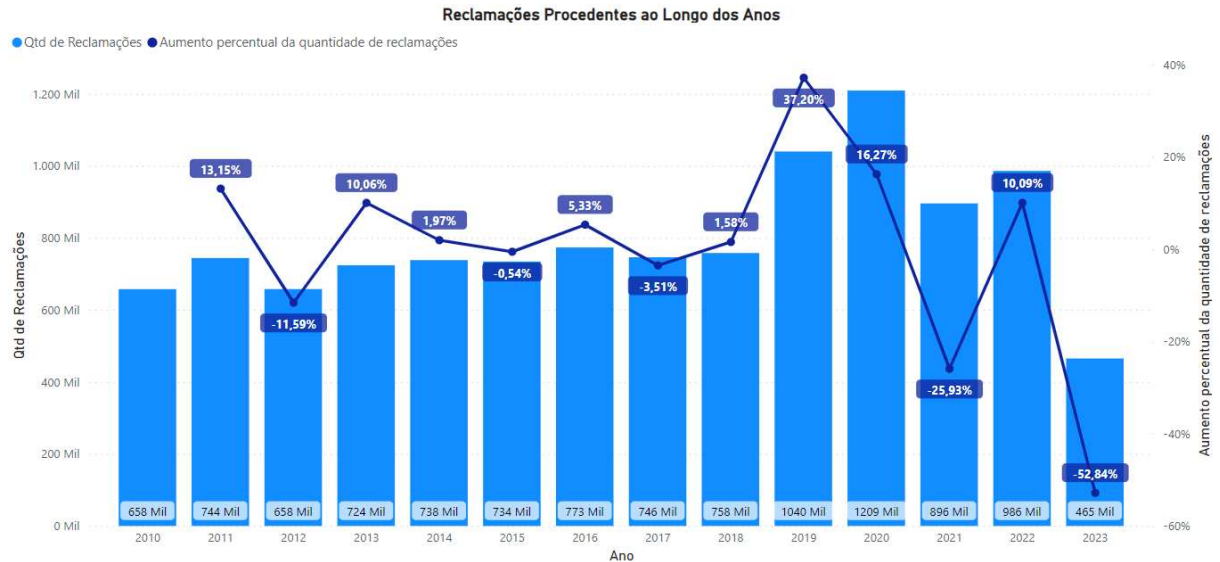
Baseado nos modelos definidos e aplicados no capítulo da Metodologia (5), para as demais bases de dados foram modelados indicadores de qualidade e o resultado é apresentado nas subseções seguintes.

6.3.1 Reclamações procedentes

Conforme a Figura 17, o número de reclamações procedentes feitas a concessionária de energia manteve-se constante principalmente entre os anos 2010 e 2018.

Entre os anos 2019 e 2023 houve uma variação maior, mas nada que se justificasse por um fator externo.

Figura 17 – Reclamações procedentes



Fonte: Elaborado pelo autor (2023).

6.3.2 Tempo de interrupção de energia

Nesta análise, a partir da Figura 18, observa-se que as interrupções programadas de ambos os grupos consumidores permaneceram estáveis dentro do período analisado. Já as que não foram programadas tiveram um aumento global ao analisarmos o Grupo A, esse mesmo aumento foi menor no Grupo B.

No entanto, como o período dos dados disponibilizados pela ANEEL compreende apenas 4 anos até o momento, este fato dificulta o estabelecimento de inferências sobre o assunto.

Figura 18 – Tempo de interrupção de energia por ano por grupo



Fonte: Elaborado pelo autor (2024).

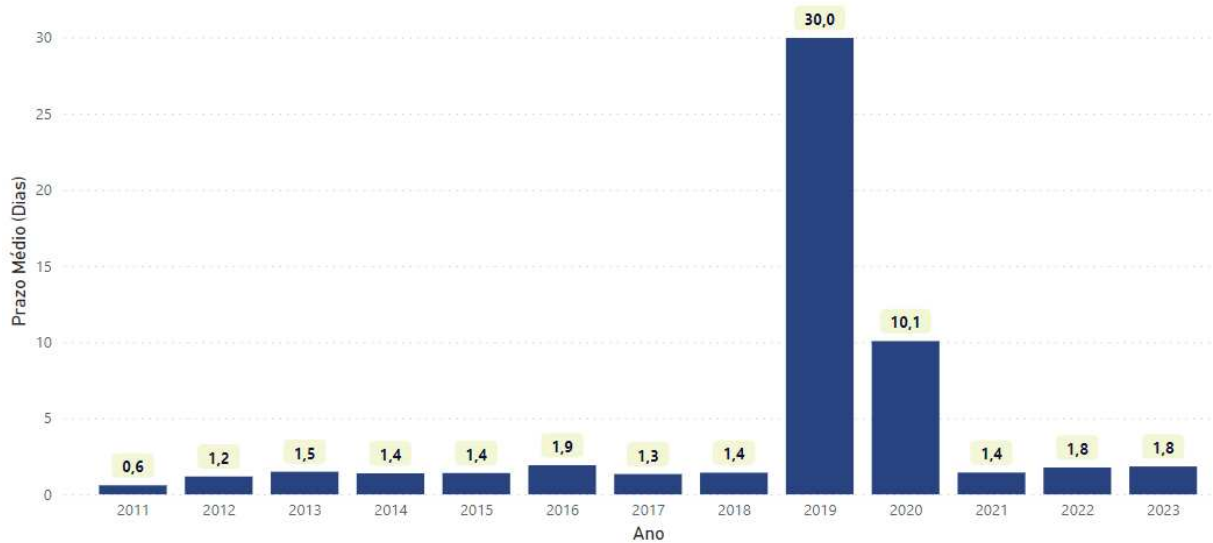
6.3.3 Indicadores relacionados a ligação de energia

O gráfico que representa o prazo médio de ligação de energia elétrica dos consumidores de alta tensão (Grupo A) entre 2011 e 2023, Figura 19, mostra um indicador abaixo de 2 dias até 2018. Nos anos de 2019 e 2020, houve uma variação acentuada, indicando possíveis desafios ou mudanças no sistema. Entre 2021 e 2023, o prazo médio retornou ao normal, sugerindo uma estabilização no processo de ligação de energia elétrica.

Figura 19 – Prazo Médio Ligação Grupo A

Prazo Médio de Ligação Por Ano

Indicador ● Prazo Médio de Ligação GRUPO A



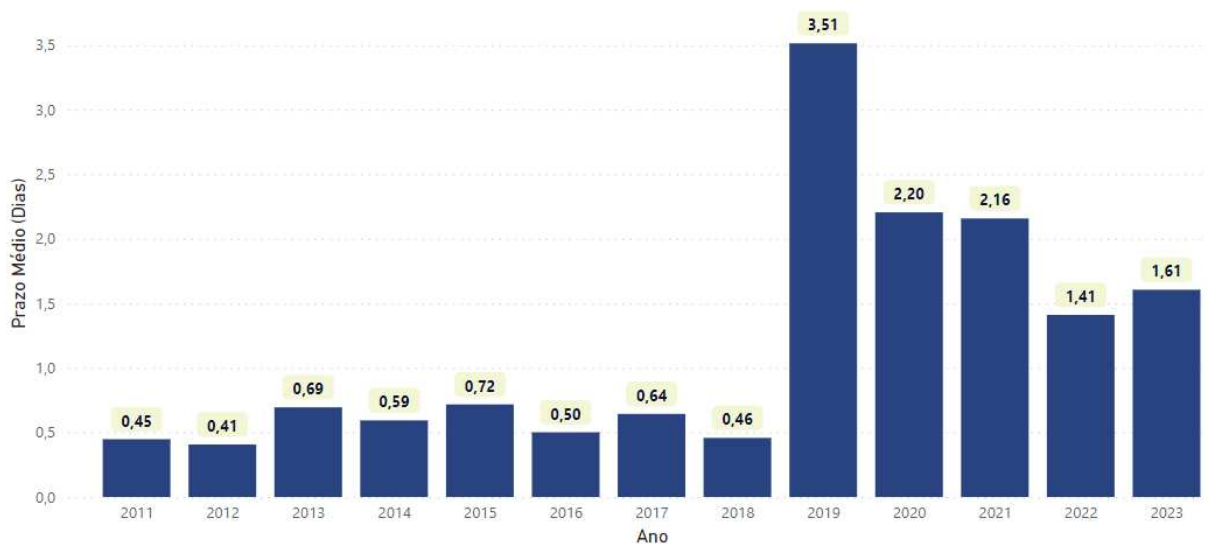
Fonte: Elaborado pelo autor (2023).

Já para o grupo de baixa tensão, Figura 20, para o mesmo período de anos, existe estabilidade no prazo médio entre os anos 2011 e 2018, porém após isso o prazo aumenta e não retorna para a condição inicial, mostrando uma piora no serviço prestado.

Figura 20 – Prazo Médio Ligação Grupo B

Prazo Médio de Ligação Por Ano (Grupo B)

Indicador ● Prazo Médio de Ligação GRUPO B



Fonte: Elaborado pelo autor (2023).

Dos prazos de ligação fornecidos pela concessionária, analisando as Figuras 21 e 22, vemos que para os consumidores de alta tensão houve uma certa variação mas com quantidade

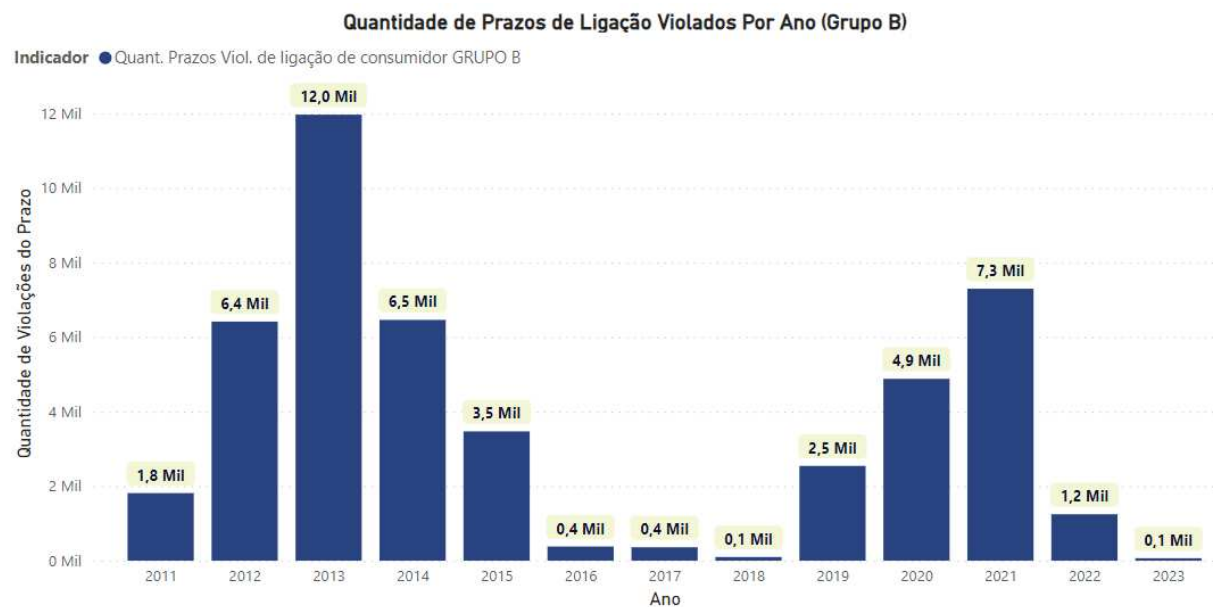
absoluta baixa. Já para os consumidores do Grupo B, os anos de 2013 e 2021 marcaram as piores quantidades de violação do prazo.

Figura 21 – Quantidade Prazos Violados Grupo A



Fonte: Elaborado pelo autor (2023).

Figura 22 – Quantidade Prazos Violados Grupo B



Fonte: Elaborado pelo autor (2023).

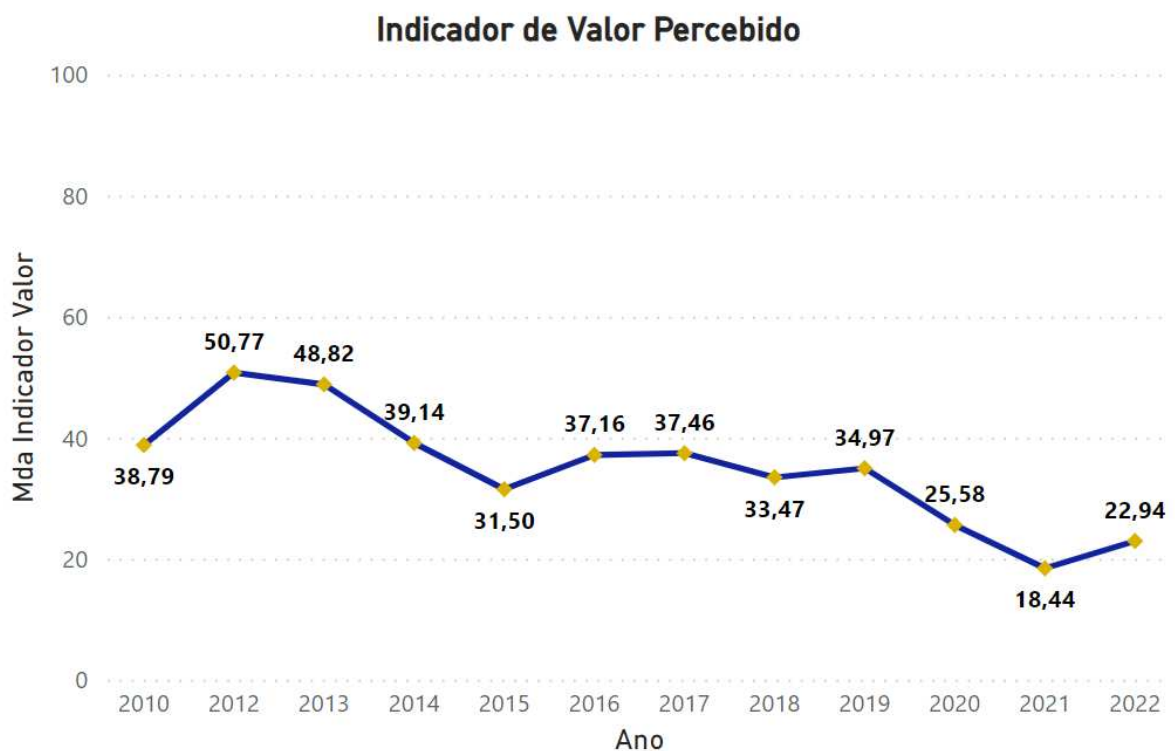
6.3.4 Indicador de Satisfação IASC

Para finalizar nossa análise de qualidade, debruçaremos sobre os indicadores que compõem o próprio indicador de satisfação do consumidor residencial da ANEEL.

Começando pelo indicador de Valor Percebido, Figura 23, percebe-se que entre 2010 e 2012 houve aumento expressivo no valor percebido pelos consumidores. Após esse período o indicador em um contexto geral manteve uma tendência de queda até o ano de 2021. Em 2022 houve uma leve melhora, mas no acumulado do período analisado, entre o melhor valor obtido em 2012 e o valor final, em 2022, nota-se uma queda de 54,81%.

Durante todo esse período de queda do indicador, a tarifa de energia estava sendo corrigida a valores menores que os indicadores padrões de aumento dos preços (9).

Figura 23 – Indicador de Valor Gerado



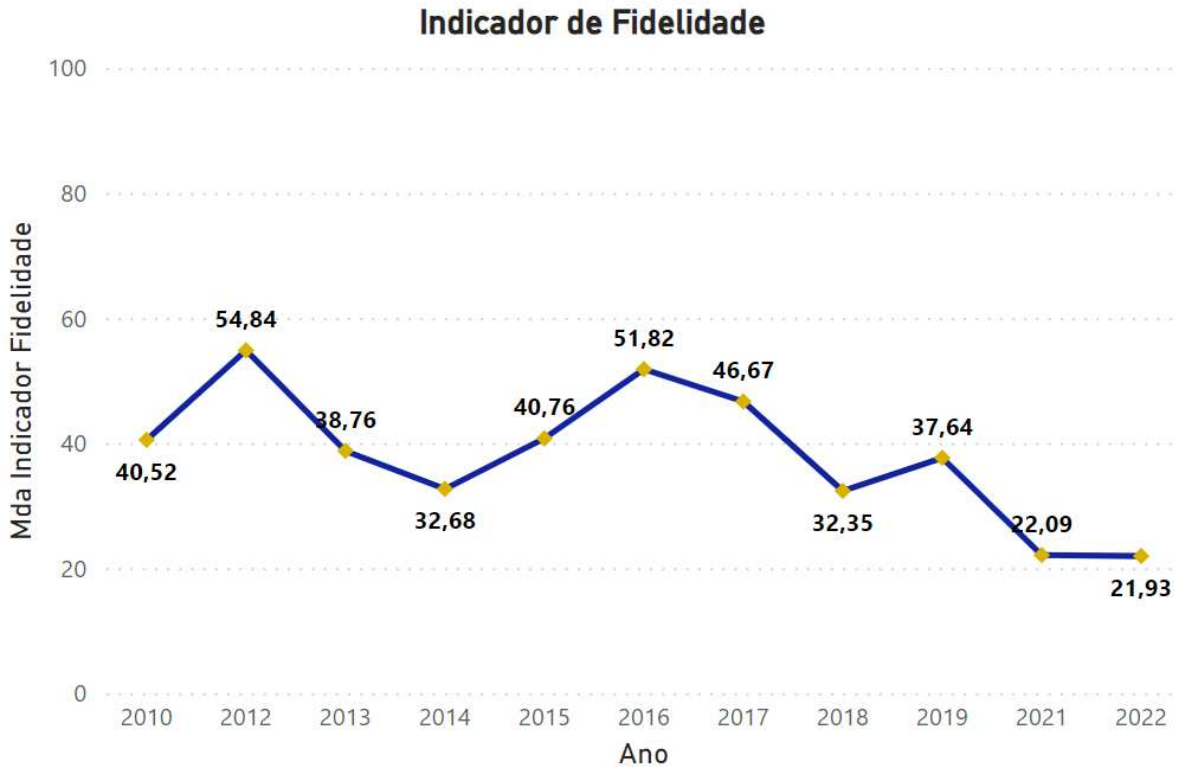
Fonte: Elaborado pelo autor (2024).

Em um cenário similar, analisando o indicador de Fidelidade (Figura 24), observa-se que talvez esse construto também possa ter sido impactado pela diferença entre aumento dos custos anuais e o repasse no valor das tarifas.

Em tese, a tarifa aumentar menos do que deveria poderia contribuir com esse indicador, a menos que isso também impacte na qualidade do serviço oferecido. Nesse sentido, analisando o Indicador de Fidelidade, parece haver uma relação entre o período em que o preço da energia elétrica foi mantido de forma artificial e os melhores momentos do indicador. Logo em seguida houve uma queda até o pior nível histórico, coincidindo com o período em que os aumentos foram novamente repassados ao consumidor.

Entretanto, é importante ressaltar que muitos outros fatores podem influenciar na qualidade do serviço prestado pela concessionária de energia. A variação do preço da tarifa em relação aos índices gerais de preços no mercado é apenas um deles.

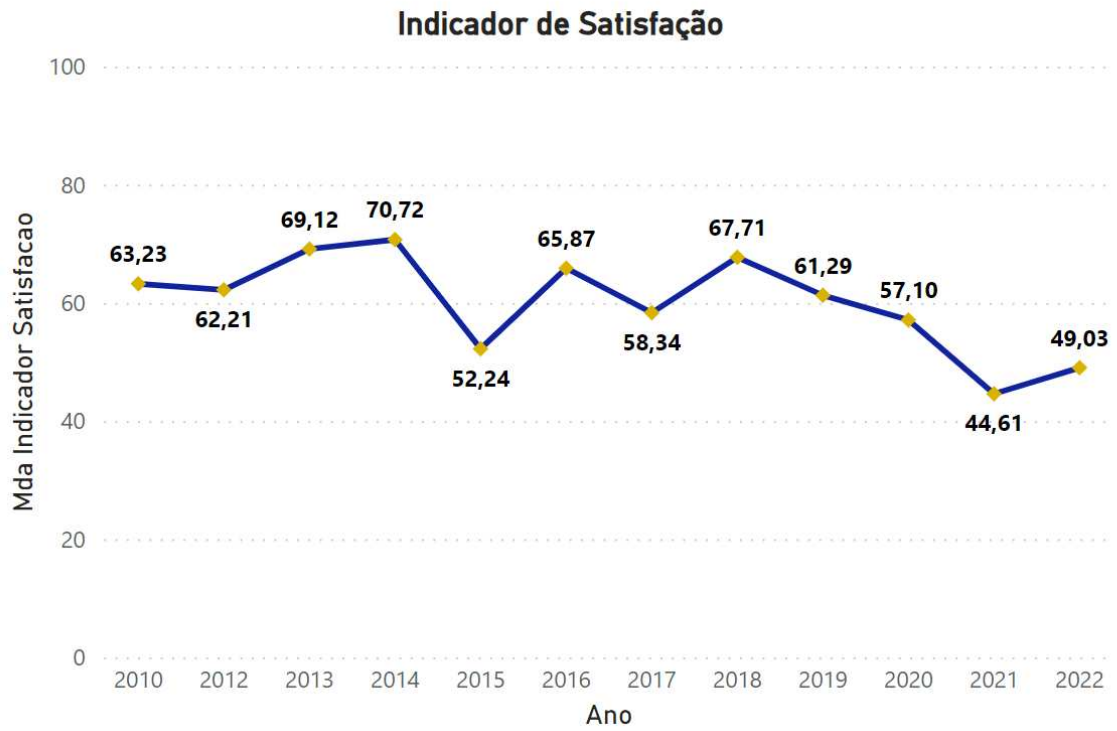
Figura 24 – Indicador de Fidelidade



Fonte: Elaborado pelo autor (2024).

Para o indicador que mede Satisfação (Figura 25), existe uma certa estabilidade nos valores do construto ao longo do período analisado. O ápice da satisfação é obtido em períodos em que a tarifação estava reduzindo de valor enquanto os custos, em situação inversa, estavam aumentando com inflação na casa dos 6% ao ano.

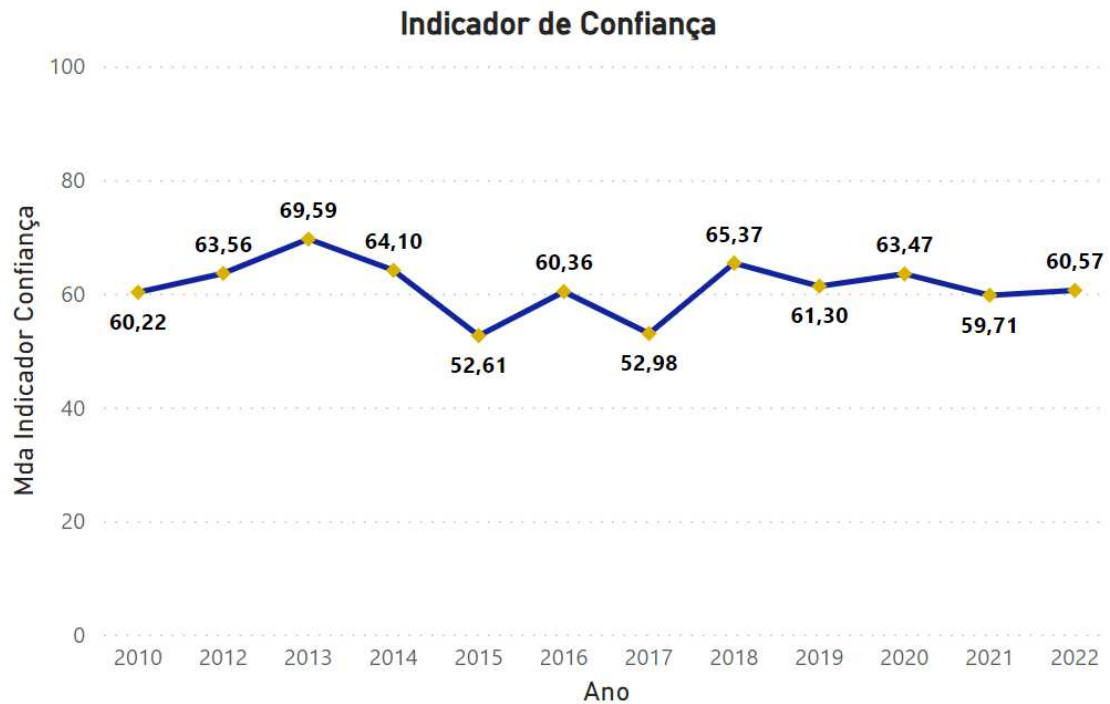
Figura 25 – Indicador de Satisfação



Fonte: Elaborado pelo autor (2024).

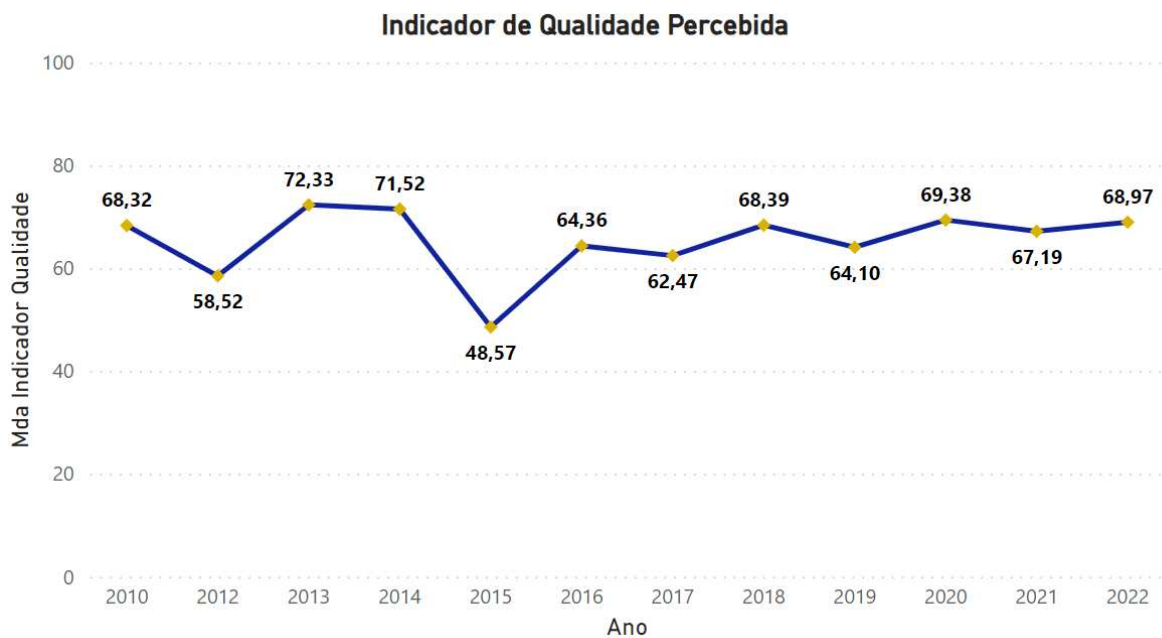
Finalmente, para os dois últimos indicadores que compõem o IASC, indicador de Confiança e de Qualidade Percebida, constatou-se uma estabilidade durante todo o período analisado.

Figura 26 – Indicador de Confiança



Fonte: Elaborado pelo autor (2024).

Figura 27 – Indicador de Qualidade



Fonte: Elaborado pelo autor (2024).

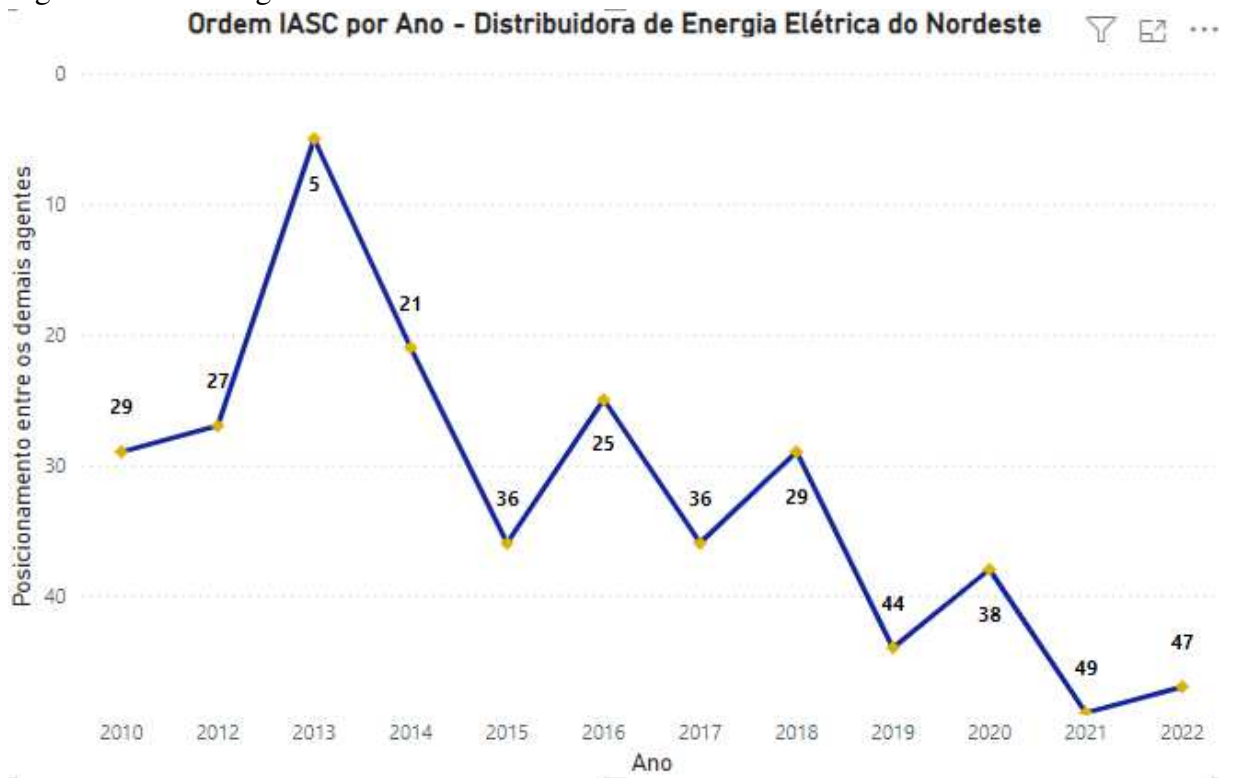
Por fim, a Figura 28 traz a visão da colocação da concessionária de energia durante o período analisado. O eixo das ordenadas foi invertido para que haja uma melhor compreensão da natureza do dado. Dessa forma, quanto mais próximo da primeira colocação do *ranking*, mais

alto o ponto se situará no gráfico.

Nesse sentido, percebe-se que o melhor ano para a concessionária, frente aos demais agentes, foi no ano de 2013, após isso sua colocação caiu gradualmente, mesmo que em alguns anos voltasse a subir algumas posições.

Infere-se, portanto, que a qualidade do serviço prestado pela concessionária de energia elétrica, baseado nessa métrica, caiu consideravelmente. Pode ter contribuído para esse cenário o fato de que, enquanto os indicadores de Qualidade Percebida e Confiança se mantiveram estáveis ao longo do período, os demais sofreram quedas.

Figura 28 – Ranking Anual IASC



Fonte: Elaborado pelo autor (2024).

7 CONCLUSÕES E TRABALHOS FUTUROS

Ao percorrer as análises e interpretações ao longo deste trabalho, ressalta-se uma visão mais aprofundada sobre a temática "Análise das bases de dados abertos da ANEEL: estudo de relação entre os custos de tarifação e a qualidade no serviço de distribuição de energia elétrica de concessionária do nordeste brasileiro". Ao longo da condução deste estudo, por meio dos dados utilizados, algoritmos e gráficos construídos e discussões feitas, enxergou-se não apenas tendências e padrões, mas também nuances e complexidades inesperadas.

Na análise crítica das informações, torna-se evidente a importância da análise exploratória de dados na compreensão da realidade. Este estudo não apenas fornece possíveis *insights* sobre como a qualidade da distribuição de energia relaciona-se com o preço base da tarifa cobrado, para o estado analisado, mas também aponta para possíveis implicações e áreas para futuras investigações.

Durante a análise financeira do preço tarifário praticado pela distribuidora de energia para os consumidores de Baixa Tensão, o modelo desenvolvido apontou para um deságio elevado entre o valor de correção da tarifa de distribuição de energia frente ao valor do aumento dos preços, no mesmo período, medido pelo IPCA e pelo IGP-M.

Esse cenário pode ter apresentado-se como um desafio para a concessionária no que tange a oferecer um serviço de qualidade. De fato, durante o período analisado de 2010 a 2022, apenas nos primeiros anos foi possível notar algum indicativo de melhora na situação do serviço. Excluindo-se esse momento, na maior parte do período avaliado houve piora ou estabilização dos indicadores, sobressaltando-se a piora no cenário.

Assim, a concessionária foi de 5º lugar no *ranking* do IASC, no ano de 2013, para 47º no último ano avaliado pela ANEEL, 2022.

A análise das quantidades de prazos violados e do prazo médio de ligação não foi conclusiva para incorporar ao estudo do impacto da tarifação no Grupo B.

Já para o grupo de consumidores de Alta Tensão, o modelo desenvolvido apontou um cenário favorável para a concessionária de energia. Apenas nos anos de 2013 e 2014 os consumidores do Grupo A tiveram suas tarifas com deságios elevados. Após isso, a curva subiu acima nos indicadores financeiros de aumento dos preços.

Nesse contexto, infere-se que essa situação beneficiaria a concessionária da rede elétrica no aprimoramento dos serviços prestados. Entretanto, os indicadores de qualidade disponibilizado para esse grupo não são suficientes para validar essa ideia, precisando assim que

mais dados sejam disponibilizados pela ANEEL.

Outrossim, duas análises de fatores importantes não se mostraram relevantes para a conclusão estudo. A análise de Reclamações Procedentes permaneceu praticamente estável durante todo o período, já a de Inadimplência Financeira dos consumidores não apresentou notória relação com os pontos debatidos.

Ademais, um ponto importante a ser ressaltado é que a base de dados disponibilizada pela ANEEL é relativamente nova. Como mencionado, ela atende a uma portaria que estabelece um plano de dados abertos para 2022-2024. Assim, existem muitos conjuntos de dados com períodos de abrangência ainda com poucos anos que com certeza terão muito mais impacto no futuro. Essa realidade trouxe um pouco de dificuldade para o presente trabalho, mas não impediu que tirássemos conclusões importantes que corroborassem com a tese defendida.

Em última análise, este trabalho não busca apenas documentar e analisar dados referentes a distribuição de energia em um estado do Nordeste brasileiro, e sim fornecer uma contribuição significativa para a compreensão da importância da cultura *data driven* no século em que vivemos. As descobertas aqui apresentadas não são meramente conclusivas, mas sim um ponto de partida para diálogos mais amplos e investigações mais aprofundadas tanto na área de distribuição de energia, como nas demais bases de dados, ainda inexploradas, disponibilizadas pela ANEEL.

Dessa forma, para futuros trabalhos, deixo como possíveis sugestões:

1. Fazer um estudo detalhado específico para cada um dos grupos consumidores;
2. Replicar a análise levando em consideração os pesos de cada um dos grupos consumidores na composição final dos consumidores da concessionária de energia;
3. Otimizar a lógica dos algoritmos aqui desenvolvidos a fim de obter resultados mais assertivos;
4. Buscar outros conjuntos de dados disponibilizados pela ANEEL que possam contribuir para a análise aqui desenvolvida;
5. Expandir a análise para outras concessionárias de outras regiões do Brasil;
6. Expandir a análise levando em consideração o processo de compra de energia elétrica para entender o impacto no preço ao consumidor final.

REFERÊNCIAS

- ANEEL. **Resolução Normativa ANEEL nº 1.000/2021**. [S.l.], 2021. Acessado em 19 nov. de 2023.
- BELISÁRIO, C. D. A.; BAHIENSE, D. A.; OLIVEIRA, G. M. Continuidade nos serviços de distribuição de energia elétrica. **Conjuntura & Planejamento**, n. 105, p. 36–40, 2003.
- BLUME, S. W. **Electric Power System Basics For the Nonelectrical Professional**. Piscataway, NJ: Wiley-IEEE Press, 2007. ISBN 978-0-471-74167-9.
- BRASIL. **Lei nº 12.783, de 11 de janeiro de 2013**. 2013. <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2013/lei/112783.htm>.
- DATA.GOV. **Catálogo de Dados**. Acesso em: novembro de 2023. Disponível em: <<https://catalog.data.gov/dataset>>.
- FERREIRA, T. V. B.; MACHADO, G. V.; ACHÃO, C. da C. L.; JUNIOR, A. dos S.; YUKIZAKI, A. Y. G. **Nota Técnica NT/EPE/DEA-005/2020**. [S.l.], 2020.
- ANEEL. **Procedimento de Regulação Tarifária (PRORET)**. 2022. Acesso em 01/06/2024. Disponível em: <<https://www.gov.br/aneel/pt-br/centrais-de-conteudos/procedimentos-regulatorios/proret>>.
- ANEEL. **Modalidades Tarifárias**. 2023. Acessado em 19 de novembro de 2023. Disponível em: <<https://www.gov.br/aneel/pt-br/assuntos/tarifas/entenda-a-tarifa/modalidades-tarifarias>>.
- ANEEL. **Portal de Dados Abertos da Agência Nacional de Energia Elétrica**. 2023. Acessado em 26 de novembro de 2023. Disponível em: <<https://dadosabertos.aneel.gov.br/>>.
- ANEEL. **Regulação da Distribuição de Energia**. 2023. Acessado em 19 de novembro de 2023. Disponível em: <<https://www.gov.br/aneel/pt-br/assuntos/distribuicao/regulacao>>.
- ANEEL. **Índice ANEEL de Satisfação do Consumidor (Iasc)**. 2024. Acessado em 04 de janeiro de 2024. Disponível em: <<https://www.gov.br/aneel/pt-br/consumidores/iasc>>.
- FGV. **Índice Geral de Preços (IGP) | Instituto Brasileiro de Economia (IBRE)**. 2023. Acessado em 25 de novembro de 2023. Disponível em: <<https://portalibre.fgv.br/igp>>.
- IBGE. **Inflação | IBGE**. 2023. Acessado em 25 de novembro de 2023. Disponível em: <<https://www.ibge.gov.br/explica/inflacao.php>>.
- IBGE. **IPCA | IBGE**. 2023. Acessado em 25 de novembro de 2023. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html>>.
- ONS. **CURVA DE CARGA HORÁRIA**. 2023. Acessado em 20 de novembro de 2023. Disponível em: <https://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/curva_carga_horaria.aspx>.
- PYPI. **PyPI - O Python Package Index**. 2023. Acessado em 25 de novembro de 2023. Disponível em: <<https://pypi.org/>>.

PYPL. **Popularity of Programming Language**. 2023. Acessado em 22 de novembro de 2023. Disponível em: <<https://pypl.github.io/PYPL.html>>.

HARARI, Y. N. Original articles. **New Perspectives Quarterly**, v. 34, n. 2, p. 36 – 43, 5 2017.

IBM. **O que é ETL (Extrair, Transformar e Carregar)?** 2023. Acessado em 03 de dezembro de 2023. Disponível em: <<https://www.ibm.com/br-pt/topics/etl>>.

JUPYTER, P. **Project Kupyter | Home**. 2024. Acessado em 28 de agosto de 2024. Disponível em: <<https://jupyter.org/>>.

LUTZ, M. **Learning Python: Powerful Object-Oriented Programming**. 5th. ed. [S.l.]: O'Reilly Media, 2013. 1643 p. ISBN 1449355730.

MCKINNEY, W. **Python Para Análise de Dados: Tratamento de Dados com Pandas, NumPy & Jupyter**. 3rd. ed. [S.l.]: Novatec Editora, 2023. 624 p. ISBN 8575228412.

MICROSOFT. **Visualização de Dados | Microsoft Power BI**. 2023. Acessado em 25 de novembro de 2023. Disponível em: <<https://powerbi.microsoft.com/pt-br/>>.

PANDAS. **Pandas Documentation**. 2023. Acessado em 25 de novembro de 2023. Disponível em: <<https://pandas.pydata.org/docs/>>.

PROVOST, F.; FAWCETT, T. Data science and its relationship to big data and data-driven decision making. **Big Data**, v. 1, n. 1, p. 51–59, 2013. PMID: 27447038. Disponível em: <<https://doi.org/10.1089/big.2013.1508>>.

ROSS, P.; RESSIA, S.; SANDER, E. Data driven management, artificial intelligence, and automation. In: **Work in the 21st Century (The Changing Context of Managing People)**. Leeds: Emerald Publishing Limited, 2017. p. 113–137. Disponível em: <<https://doi.org/10.1108/978-1-78714-577-120171008>>.

SQLITE. **About SQLite**. 2023. Acessado em 25 de novembro de 2023. Disponível em: <<https://www.sqlite.org/about.html>>.

APÊNDICE A – LINKS DOS CONJUNTOS DE DADOS UTILIZADOS

Abaixo constam os links utilizados neste trabalho para obtenção dos conjuntos de dados no domínio público da ANEEL além da documentação de cada um deles:

- <<https://dadosabertos.aneel.gov.br/dataset/tarifas-distribuidoras-energia-eletrica>>
- <<https://dadosabertos.aneel.gov.br/dataset/reclamacoes-no-1o-e-2o-niveis-da-distribuidora>>
- <<https://dadosabertos.aneel.gov.br/dataset/interruptoes-de-energia-eletrica-nas-redes-de-distribuicao>>
- <<https://dadosabertos.aneel.gov.br/dataset/qualidade-do-atendimento-comercial>>
- <<https://dadosabertos.aneel.gov.br/dataset/indqual-inadimplencia>>
- <<https://dadosabertos.aneel.gov.br/dataset/indice-aneel-de-satisfacao-do-consumidor-iasc>>

ANEXO A – JUPYTER NOTEBOOKS UTILIZADOS

A seguir estão dispostos os *Jupyter Notebooks* desenvolvidos e utilizados neste trabalho.

A.1 ANALISE_INADIMPLENCIA_NOTEBOOK

```
1 # %%
2 import pandas as pd
3 import numpy as np
4 import sqlite3
5 import matplotlib.pyplot as plt
6 import os
7 from datetime import datetime as dt
8 import datetime
9
10 # %%
11 #DEFINE FUNÇÕES PRINCIPAIS PARA ACESSO DO BANCO DE DADOS COM A BASE
12     INICIAL
13
14 conn = sqlite3.connect('DIST_DB_ANEEL.db')
15 cursor = conn.cursor()
16 conn_bi = sqlite3.connect('BD_POWER_BI_FINAL.db')
17 cursor_bi = conn_bi.cursor()
18
19 # %%
20 #PREPARA TABELA COM INFORMAÇÕES UTEIS PARA ANALISE
21
22 tableName1 = "DIST_INADIMPLENCIA"
23 columnsNecessary = f"SigAgente ,AnoIndice ,NumPeriodoIndice ,
24     VlrIndiceEnviado "
25
26 query = f"""
27     SELECT {columnsNecessary}
28     FROM {tableName1}
29     WHERE SigAgente = ""
```



```

30 df = pd.read_sql_query(query, conn)
31
32 # %%
33 #AGRUPA POR ANO O TOTAL EM INADIMPLENCIAS FINANCEIRAS
34
35 df_analise = df.copy()
36 df_analise['VlrIndiceEnviado'] = df_analise['VlrIndiceEnviado'].
    astype(str)
37 df_analise['VlrIndiceEnviado'] = df_analise['VlrIndiceEnviado'].apply
    (lambda x: float(x.replace(",",".")))
38 df_analise = df_analise.groupby(['SigAgente', 'AnoIndice'],)['
    VlrIndiceEnviado'].sum()
39 df_analise = df_analise.reset_index()
40 df_analise['AumentoPercentual'] = df_analise['VlrIndiceEnviado'].
    pct_change()
41 df_analise['refDate'] = df_analise['AnoIndice']
42 df_analise['refDate'] = df_analise['refDate'].apply(lambda x: "01-01-
    "+str(x))
43 df_analise['refDate'] = pd.to_datetime(df_analise['refDate'])
44
45 print(df_analise)
46
47 # %%
48 #SOBE PARA O BANCO DE DADOS
49 df_analise.to_sql('Inadimplencia', conn_bi, if_exists='replace', index
    =False)
50
51 conn.close()
52 conn_bi.close()

```

A.2 ANALISE_INTERRUPCAO_ENERGIA_NOTEBOOK

```

1 import pandas as pd
2 import numpy as np
3 import sqlite3
4 import matplotlib.pyplot as plt
5 import os
6 from datetime import datetime as dt

```

```
7 import datetime
8
9 # %%
10 #DEFINE FUNÇÕES PRINCIPAIS PARA ACESSO DO BANCO DE DADOS COM A BASE
    INICIAL
11
12 conn = sqlite3.connect('DIST_DB_ANEEL.db')
13 cursor = conn.cursor()
14 conn_bi = sqlite3.connect('BD_POWER_BI_FINAL.db')
15 cursor_bi = conn_bi.cursor()
16
17 # %%
18 #PREPARA TABELA COM INFORMAÇÕES UTEIS PARA ANALISE
19
20 tableName = "DIST_INT_ELE_CONSOLIDADO"
21 columnsNecessary = f'{tableName}.SigAgente, {tableName}.
    FatGeradorInterrupcao, {tableName}.DscTipoInterrupcao, \
22     {tableName}.IdeMotivoExpurgo, {tableName}.
    DatInicioInterrupcao, {tableName}.
    DatFimInterrupcao, {tableName}.NumNivelTensao
    , {tableName}.NumUnidadeConsumidora, \
23     {tableName}.NumAno '
24
25 query = f""" SELECT {columnsNecessary}
26     FROM {tableName}
27     WHERE SigAgente LIKE '
28
29 """
30
31 df = pd.read_sql_query(query, conn)
32
33
34 # %%
35 def changeTensaoToGroup(tensao): #Muda tensão para grupo
36     if tensao == 13800:
37         return "A"
38     elif tensao == 220:
39         return "B"
40     elif tensao == 69000:
```

```

41         return "A"
42
43
44 df_modificada = df.copy()
45 df_modificada['NumNivelTensao'] = df_modificada['NumNivelTensao'].
    apply(lambda x: changeTensaoToGroup(x))
46 df_modificada.rename(columns={'NumNivelTensao': 'GrupoConsumidor'},
    inplace=True)
47 df_modificada['DuracaoInterrupcaoHoras'] = ''
48 df_modificada['DatInicioInterrupcao'] = pd.to_datetime(df_modificada[
    'DatInicioInterrupcao'])
49 df_modificada['DatFimInterrupcao'] = pd.to_datetime(df_modificada[
    'DatFimInterrupcao'])
50
51 #Calcula a duração em horas de cada interrupção
52
53 for row in df_modificada.itertuples():
54     inicio = row.DatInicioInterrupcao
55     fim = row.DatFimInterrupcao
56     duracao = round(float(((fim - inicio).total_seconds())/3600),2)
57     df_modificada.at[row.Index, 'DuracaoInterrupcaoHoras'] = duracao
58
59 df_modificada['TempoInterrupcaoGlobal'] = df_modificada['
    NumUnidadeConsumidora']*df_modificada['DuracaoInterrupcaoHoras']
60
61
62 # for gp in set(df_modificada['GrupoConsumidor']):
63 #     df_temp = df_modificada[df_modificada['GrupoConsumidor']==gp]
64 #     df_temp.to_csv(rf"C:\Users\marco\Desktop\teste_parcial_{gp}.csv
    ", sep=';', index=False, encoding='ansi', decimal=',')
65
66 # %%
67 #Agrupa por ano e soma a duração de interrupção por tipo (programada/
    não-programada)
68
69 df_final_bi = df_modificada.copy()
70 df_final_bi = df_final_bi[['SigAgente', 'DscTipoInterrupcao', '
    NumUnidadeConsumidora', 'GrupoConsumidor', 'NumAno', '
    DuracaoInterrupcaoHoras', 'DatInicioInterrupcao', '

```

```

    TempoInterrupcaoGlobal']]
71 df_final_bi.to_sql('InterrupcaoEletricidadePorGrupo',conn_bi,
    if_exists='replace', index=False)
72
73 conn.close()
74 conn_bi.close()

```

A.3 ANALISE_QUALIDADE_COMERCIAL_NOTEBOOK

```

1 # %%
2 import pandas as pd
3 import numpy as np
4 import sqlite3
5 import matplotlib.pyplot as plt
6 import os
7 from datetime import datetime as dt
8 import datetime
9 import re
10
11 # %%
12 #DEFINE FUNÇÕES PRINCIPAIS PARA ACESSO DO BANCO DE DADOS COM A BASE
    INICIAL
13
14 conn = sqlite3.connect('DIST_DB_ANEEL.db')
15 cursor = conn.cursor()
16 conn_bi = sqlite3.connect('BD_POWER_BI_FINAL.db')
17 cursor_bi = conn_bi.cursor()
18
19 # %%
20 #PREPARA TABELA COM INFORMAÇÕES UTEIS PARA ANALISE
21
22 tableName1 = "DIST_QUALIDADE_ATENDIMENTO_COMERCIAL"
23 tableName2 = "DIST_IND_DOMINIOS"
24 columnsNecessary = f"{tableName1}.SigAgente,{tableName1}.AnoIndice,{
    tableName1}.NumPeriodoIndice,{tableName1}.VlrIndiceEnviado,{
    tableName2}.DscIndicador "
25
26 query = f"""

```

```
27     SELECT {columnsNecessary}
28     FROM {tableName1}
29     JOIN {tableName2} ON {tableName1}.SigIndicador = {tableName2
30         }.SigIndicador
31     WHERE SigAgente = ""
32     ""
33 df = pd.read_sql_query(query,conn)
34
35
36 # %%
37 #Dataframe resultado da query
38 df_analise = df.copy()
39 df_analise['AnoIndice'] = df_analise['AnoIndice'].astype(str)
40 df_analise['NumPeriodoIndice'] = df_analise['NumPeriodoIndice'].
41     astype(str)
42 df_analise['dateTimeRef'] = df_analise['AnoIndice'] + "-" +
43     df_analise['NumPeriodoIndice'] + "-" + "01"
44 df_analise['dateTimeRef'] = pd.to_datetime(df_analise['dateTimeRef'])
45 df_analise = df_analise.drop(columns=['AnoIndice','NumPeriodoIndice'
46     ])
47 df_analise['VlrIndiceEnviado'] = df_analise['VlrIndiceEnviado'].str.
48     replace(',','').astype(float)
49
50 #Listas com os tópicos analisados
51
52 quantityElements = [
53     "Quant. Prazos Viol. de vistoria de unidade consumidora de rea
54     urbana (art. 30)",
55     "Quantidade de religação por suspensão indevida do fornec. (art.
56     176)",
57     "Quant. Prazos Viol. de Conclusão das obras de até 1 km na rede
58     prim. (art. 34)",
59     "Quant. Prazos Viol. de religação de emergência em rea urbana (
60     art. 176)",
61     "Quant. Prazos Viol. de substituição do medidor com defeito (art.
62     115)",
63     "Quant. Prazos Viol. de religação em rea rural (art. 176)",
```

55 "Quant. Prazos Viol. de ressarc. por meio do pgto, conserto ou
subst. (art. 208)",

56 "Quant. Prazos Viol. de ligação de consumidor do grupo B em rea
rural (art. 31)",

57 "Quant. Prazos Viol. vistoria de unidade consumidora de rea
rural (art. 30)",

58 "Quant. Prazos Viol. de ligação de unidade consumidora do grupo A
(art. 31)",

59 "Quant. Prazos Viol. de religação de emergência em rea rural (
art. 176)",

60 "Quant. Prazos Viol. de verif. de equip. por ressarc. de dano elé
trico (art. 206)",

61 "Quant. Prazos Viol. de Conclusão das obras na rede tensão sec. (
art. 34)",

62 "Quant. Prazos Viol. de religação por suspensão indevida do
fornec. (art. 176)",

63 "Quant. Prazos Viol. de religação em rea urbana (art. 176)",

64 "Quant. Prazos Viol. de reclamação de cobrança ou devol. de
diferenças (art. 133)",

65 "Quant. Prazos Viol. de solicitações de aferição dos medidores (
art. 137)",

66 "Quant. Prazos Viol. de Início das Obras (art. 34)",

67 "Quant. Prazos Viol. de solução de reclamação com visita técnica
(art. 197)",

68 "Quant. Prazos Viol. de ligação de consumidor do grupo B em rea
urbana (art. 31)",

69 "Quant. Prazos Viol. de ligação de consumidor do grupo B em rea
rural (art. 31)"

70]

71

72 termAvarage = [
73 "Prazo médio de Conclusão das obras na rede tensão sec. (art. 34)
",
74 "Prazo Médio de religação de emergência em rea urbana (art.
176)",
75 "Prazo Médio de ligação de consumidor do grupo B em rea urbana
(art. 31)",
76 "Prazo Médio de religação de emergência em rea rural (art. 176)
",

```

77 "Prazo Médio de Início das Obras (art. 34)",
78 "Prazo Médio de vistoria de unidade consumidora de rea rural (
    art. 30)",
79 "Prazo Médio de religação por suspensão indevida do fornec. (art.
    176)",
80 "Prazo médio de vistoria de unidade consumidora de rea urbana (
    art. 30)",
81 "Prazo Médio de substituição do medidor com defeito (art. 115)",
82 "Prazo Médio de ressarc. por meio do pgto, conserto ou subst. (
    art. 208)",
83 "Prazo Médio de religação em rea urbana (art. 176)",
84 "Prazo Médio de religação em rea rural (art. 176)",
85 "Prazo Médio de solicitações de aferição dos medidores (art. 137)
    ",
86 "Prazo Médio de solução de reclamação com visita técnica (art.
    197)",
87 "Prazo médio de Conclusão das obras de até 1 km na rede prim. (
    art. 34)",
88 "Prazo Médio de ligação de consumidor do grupo B em rea rural",
89 "Prazo Médio de ligação de unidade consumidora do grupo A (art.
    31)",
90 "Prazo Médio de ligação de consumidor do grupo B em rea rural (
    art. 31)"
91 ]
92
93 #Divisão em dois dataframes com naturezas diferentes:
94 #1) Dataframe com as quantidades de prazos violados
95 #2) Dataframe com os prazos médios
96
97 quantidadePrazoDataFrame = df_analise[df_analise['DscIndicador'].isin
    (quantityElements)].copy()
98 prazoMedioDataFrame = df_analise[df_analise['DscIndicador'].isin(
    termAvarage)].copy()
99
100
101
102 # %%
103 #Criando tabelas
104

```

```
105
106 # 1)Quantidade Violação prazo para religamento de energia
107 filterList = ["Quant. Prazos Viol. de ligação de consumidor do grupo
108             B em rea urbana (art. 31)",
109             "Quant. Prazos Viol. de ligação de consumidor do grupo
110             B em rea rural (art. 31)",
111             "Quant. Prazos Viol. de ligação de unidade consumidora
112             do grupo A (art. 31)",
113 ]
114
115 qtyViolPrazoReligamento = quantidadePrazoDataFrame[
116     quantidadePrazoDataFrame['DscIndicador'].isin(filterList)].copy()
117
118 for linha in qtyViolPrazoReligamento.itertuples():
119     indicador = linha.DscIndicador
120     if indicador == "Quant. Prazos Viol. de ligação de consumidor do
121     grupo B em rea rural (art. 31)":
122         qtyViolPrazoReligamento.at[linha.Index,"DscIndicador"] = "
123             Quant. Prazos Viol. de ligação de consumidor GRUPO B"
124     elif indicador == "Quant. Prazos Viol. de ligação de consumidor
125     do grupo B em rea urbana (art. 31)":
126         qtyViolPrazoReligamento.at[linha.Index,"DscIndicador"] = "
127             Quant. Prazos Viol. de ligação de consumidor GRUPO B"
128     elif indicador == "Quant. Prazos Viol. de ligação de unidade
129     consumidora do grupo A (art. 31)":
130         qtyViolPrazoReligamento.at[linha.Index,"DscIndicador"] = "
131             Quant. Prazos Viol. de ligação de consumidor GRUPO A"
132     continue
133
134 qtyViolPrazoReligamento = pd.DataFrame(qtyViolPrazoReligamento.
135     groupby(['SigAgente', 'DscIndicador', 'dateTimeRef'])['
136     VlrIndiceEnviado'].sum())
137
138 qtyViolPrazoReligamento.reset_index(inplace=True)
139
140 # 2) Quantidade Violação Obras tensao sec
141 filterList = ["Quant. Prazos Viol. de Conclusão das obras na rede
142             tensão sec. (art. 34)"
143 ]
```



```
131
132 qtyViolPrazoObraSec = quantidadePrazoDataFrame[
    quantidadePrazoDataFrame['DscIndicador'].isin(filterList)].copy()
133 qtyViolPrazoObraSec['DscIndicador'] = "Quant. Prazos Viol. de Conclus
    ão das obras na rede tensão sec."
134
135
136 # 3) Quantidade Violação Relig Suspensao Indevida
137 filterList = ["Quant. Prazos Viol. de religação por suspensão
    indevida do fornec. (art. 176)"
138 ]
139
140 qtyViolReligSuspIndevida = quantidadePrazoDataFrame[
    quantidadePrazoDataFrame['DscIndicador'].isin(filterList)].copy()
141 qtyViolReligSuspIndevida['DscIndicador'] = "Quant. Prazos Viol. de
    religação por suspensão indevida do fornec."
142
143
144 # 4) Prazo Médio ligação unidade consumidora
145 filterList = ["Prazo Médio de religação em rea urbana (art. 176)",
    "Prazo Médio de religação em rea rural (art. 176)",
146 ]
147 ]
148
149 prazoMedioReligacao = prazoMedioDataFrame[prazoMedioDataFrame['
    DscIndicador'].isin(filterList)].copy()
150
151 for linha in prazoMedioReligacao.itertuples():
152     indicador = linha.DscIndicador
153     if indicador == "Prazo Médio de religação em rea urbana (art.
        176)":
154         prazoMedioReligacao.at[linha.Index, "DscIndicador"] = "Prazo M
            édio de religação urbana"
155     elif indicador == "Prazo Médio de religação em rea rural (art.
        176)":
156         prazoMedioReligacao.at[linha.Index, "DscIndicador"] = "Prazo M
            édio de religação rural"
157
158 # 5) Prazo Médio ligação por grupo
159 filterList = [
```

```

160     "Prazo Médio de ligação de consumidor do grupo B em rea rural (
        art. 31)",
161     "Prazo Médio de ligação de unidade consumidora do grupo A (art.
        31)",
162     "Prazo Médio de ligação de consumidor do grupo B em rea urbana
        (art. 31)",
163 ]
164
165 prazoMedioLigacaoPorGrupo = prazoMedioDataFrame[prazoMedioDataFrame['
        DscIndicador']].isin(filterList)].copy()
166
167 for linha in prazoMedioLigacaoPorGrupo.itertuples():
168     indicador = linha.DscIndicador
169     if indicador == "Prazo Médio de ligação de consumidor do grupo B
        em rea rural (art. 31)":
170         prazoMedioLigacaoPorGrupo.at[linha.Index, "DscIndicador"] = "
            Prazo Médio de Ligação GRUPO B"
171     elif indicador == "Prazo Médio de ligação de consumidor do grupo
        B em rea urbana (art. 31)":
172         prazoMedioLigacaoPorGrupo.at[linha.Index, "DscIndicador"] = "
            Prazo Médio de Ligação GRUPO B"
173     elif indicador == "Prazo Médio de ligação de unidade consumidora
        do grupo A (art. 31)":
174         prazoMedioLigacaoPorGrupo.at[linha.Index, "DscIndicador"] = "
            Prazo Médio de Ligação GRUPO A"
175     continue
176
177 prazoMedioLigacaoPorGrupo = pd.DataFrame(prazoMedioLigacaoPorGrupo.
        groupby(['SigAgente', 'DscIndicador', 'dateTimeRef'])['
            VlrIndiceEnviado'].sum())
178 prazoMedioLigacaoPorGrupo.reset_index(inplace=True)
179
180 #Faz a media dos prazos medios dos dois tipos de grupo B
181 for linha in prazoMedioLigacaoPorGrupo.itertuples():
182     if linha.DscIndicador == "Prazo Médio de Ligação GRUPO B":
183         prazoMedioLigacaoPorGrupo.at[linha.Index, "VlrIndiceEnviado"]
            = linha.VlrIndiceEnviado/2
184
185 # 6) Prazo Médio Solucao Visita Tecnica

```

```

186 filterList = [
187     "Prazo Médio de solução de reclamação com visita técnica (art.
188         197)"
189 ]
190 prazoMedioSolucaoVisitaTecnica = prazoMedioDataFrame[
191     prazoMedioDataFrame['DscIndicador'].isin(filterList)].copy()
192 prazoMedioSolucaoVisitaTecnica['DscIndicador'] = "Prazo Médio resoluç
193     ão com visita técnica"
194 # %%
195 #Sobe tabelas para banco POWER BI
196
197 listaTabelas = {
198     "qualidadeComercial-qtyViolPrazoReligamento":
199         qtyViolPrazoReligamento ,
200     "qualidadeComercial-qtyViolPrazoObraSec": qtyViolPrazoObraSec ,
201     "qualidadeComercial-qtyViolReligSuspIndevida":
202         qtyViolReligSuspIndevida ,
203     "qualidadeComercial-prazoMedioReligacao": prazoMedioReligacao ,
204     "qualidadeComercial-prazoMedioLigacaoPorGrupo":
205         prazoMedioLigacaoPorGrupo ,
206     "qualidadeComercial-prazoMedioSolucaoVisitaTecnica":
207         prazoMedioSolucaoVisitaTecnica
208 }
209
210 for k,v in listaTabelas.items():
211     df = v
212     df.to_sql(k,conn_bi, if_exists='replace', index=False)

```

A.4 ANALISE_RECLAMACOES_N1_N2_NOTEBOOK

```

1 # %%
2 import pandas as pd
3 import numpy as np
4 import sqlite3
5 import matplotlib.pyplot as plt

```

```

6 import os
7 from datetime import datetime as dt
8 import datetime
9
10 # %%
11 #DEFINE FUNÇÕES PRINCIPAIS PARA ACESSO DO BANCO DE DADOS COM A BASE
    INICIAL
12
13 conn = sqlite3.connect('DIST_DB_ANEEL.db')
14 cursor = conn.cursor()
15 conn_bi = sqlite3.connect('BD_POWER_BI_FINAL.db')
16 cursor_bi = conn_bi.cursor()
17
18 # %%
19 #PREPARA TABELA COM INFORMAÇÕES UTEIS PARA ANALISE
20
21 tableName1 = "DIST_RECLAMACOES_N1_N2"
22 columnsNecessary = f"SigAgente ,DatReferencia ,DescReclamacao ,
    QtdReclamacoesRecebidas ,QtdReclamacoesImprocedentes "
23
24 query = f"""
25     SELECT {columnsNecessary}
26     FROM {tableName1}
27     WHERE SigAgente = ""
28
29     """
30 df = pd.read_sql_query(query,conn)
31
32
33 # %%
34 def removeMotivoInvalido(df): #Remove as reclamações com motivos sem
    aplicação para esta análise
35     df['MotivoValido'] = ''
36     listaRemocao = ['Elogio','Informação','Alteração Cadastral','Denú
        ncia','Custo de Disponibilidade','Sugestão','Problemas de
        instalação interna na unidade consumidora',\
37         'Prazos','Indisponibilidade de Agência/Postos e
            Canais de atendimento/Atend. Telefônico/Serviço
            de Arrecadação','Cobrança Indevida de

```

```

    Atividade Acessória']
38
39     df['MotivoValido'] = ~df['DescReclamacao'].isin(listaRemocao)
40     df = df[df['MotivoValido']]
41
42     return df
43
44 df_analise = df.copy()
45 df_analise['QtdReclamacoesProcedentes'] = df_analise['
    QtdReclamacoesRecebidas'] - df_analise['
    QtdReclamacoesImprocedentes'] #Calcula a diferença entre reclamaçõ
    es recebidas e as improcedentes
46 df_analise = df_analise.drop(columns=['QtdReclamacoesRecebidas', '
    QtdReclamacoesImprocedentes']) # Deleta colunas não utilizadas
47 df_analise['Ano'] = df_analise['DatReferencia'].apply(lambda x: x
    [0:4]) #Cria coluna ano
48 df_analise['DatReferencia'] = pd.to_datetime(df_analise['
    DatReferencia'])
49 #print(set(df['DescReclamacao']))
50
51
52 df_analise = removeMotivoInvalido(df_analise) #Remove as reclamações
    com motivos sem aplicação para esta analise
53
54 #Agrupa por tipo e soma a quantidade de reclamações - Analise 1
55 df_analise1 = df_analise.groupby(['DescReclamacao'])['
    QtdReclamacoesProcedentes'].sum().reset_index().sort_values('
    QtdReclamacoesProcedentes',ascending=False)
56
57 #Agrupa por ano e soma a quantidade de reclamacoes - Analise 2
58 df_analise2 = df_analise.groupby(['Ano'])['QtdReclamacoesProcedentes'
    ].sum().reset_index().sort_values('Ano',ascending=True)
59 df_analise2['AumentoPercentual'] = df_analise2['
    QtdReclamacoesProcedentes'].pct_change()
60 df_analise2['DatReferencia'] = df_analise2['Ano']
61 df_analise2['DatReferencia'] = df_analise2['DatReferencia'].apply(
    lambda x: "01-01-"+str(x))
62 df_analise2['DatReferencia'] = pd.to_datetime(df_analise2['
    DatReferencia'])

```

```
63 print(df_analise2)
64
65
66
67
68 # %%
69 #SOBE PARA O BANCO DE DADOS
70
71 df_analise1.to_sql('ReclamacoesN1N2_1',conn_bi, if_exists='replace',
72                   index=False)
73 df_analise2.to_sql('ReclamacoesN1N2_2',conn_bi, if_exists='replace',
74                   index=False)
75 df_analise.to_sql('ReclamacoesN1N2_Geral',conn_bi, if_exists='replace
76                   ', index=False)
77
78 conn.close()
79 conn_bi.close()
```

A.5 ANALISE_SATISFACAO_CONSUMO_NOTEBOOK

```
1 # %%
2 import pandas as pd
3 import numpy as np
4 import sqlite3
5 import os
6 from datetime import datetime as dt
7 import datetime
8
9
10 # %%
11 #DEFINE FUNÇÕES PRINCIPAIS PARA ACESSO DO BANCO DE DADOS COM A BASE
12     INICIAL
13
14 conn = sqlite3.connect('DIST_DB_ANEEL.db')
15 cursor = conn.cursor()
16 conn_bi = sqlite3.connect('BD_POWER_BI_FINAL.db')
17 cursor_bi = conn_bi.cursor()
```

```

18 # %%
19 #PREPARA TABELA COM INFORMAÇÕES UTEIS PARA ANALISE
20
21 tableName1 = "DIST_IND_SATISF_CONSU_ANEEL"
22 columnsNecessary = f"{tableName1}.NumAno,{tableName1}.NumOrdemIASC,{
    tableName1}.MdaIndicadorQualidade,{tableName1}.
    MdaIndicadorSatisfacao, MdaIndicadorFidelidade,\
23         MdaIndicadorValor, MdaIndicadorConfianca "
24
25 query = f"""
26     SELECT {columnsNecessary}
27     FROM {tableName1}
28     WHERE SigAgente = ""
29
30     """
31 df = pd.read_sql_query(query,conn)
32 df['refDate'] = df['NumAno']
33 df['refDate'] = df['refDate'].apply(lambda x: "01-01-"+str(x))
34 df['refDate'] = pd.to_datetime(df['refDate'])
35
36 df = df[df['refDate'].dt.year>=2010]
37
38
39 # %%
40 #SOBE PARA O BANCO DE DADOS
41 df.to_sql('SatisfacaoConsumidor',conn_bi, if_exists='replace', index=
    False)
42
43 conn.close()
44 conn_bi.close()

```

A.6 ANALISE_TARIFA_NOTEBOOK

```

1 # %%
2 import pandas as pd
3 import numpy as np
4 import sqlite3
5 import matplotlib.pyplot as plt

```

```
6 import os
7 from datetime import datetime as dt
8
9 # %%
10 #DEFINE FUNÇÕES PRINCIPAIS PARA ACESSO DO BANCO DE DADOS COM A BASE
    INICIAL
11
12 conn = sqlite3.connect('DIST_DB_ANEEL.db')
13 cursor = conn.cursor()
14 conn_bi = sqlite3.connect('BD_POWER_BI_FINAL.db')
15 cursor_bi = conn_bi.cursor()
16
17 # %%
18 #TABELA DE POPULAÇÃO POR ESTADO ENTRE OS ANOS 2010 E 2022
19 #Baseado no censo do IBGE, para fins de simplificação, ser
    considerado que a população cresceu linearmente entre esses anos
20
21 def _split_state(df): #Recebe o dataframe com Estado no formato
    ACRE_AC e separa em duas colunas ACRE e AC
22
23     df_final = df.copy()
24
25     for row in df.itertuples():
26         old_column = row.Estado
27         old_column = old_column.split('_')
28         df_final.at[row.Index, 'Estado_nome'] = old_column[0]
29         df_final.at[row.Index, 'Estado_SIGLA'] = old_column[1]
30
31     df_final.drop(columns=['Estado'], inplace=True)
32     df_final.rename(columns={'Estado_nome': 'Estado', 'Estado_SIGLA': '
        SIGLA'}, inplace=True)
33     df_final = df_final[['Ano', 'Estado', 'SIGLA', 'Populacao']]
34
35     return df_final
36
37 #Consome planilha com dados do IBGE e estrutura dataframe
38 df_base = pd.read_excel(r'.\IBGE_POPULACAO_UF.xlsx')
39
40 anos = [i for i in range(2010, 2023, 1)]
```



```
41 estados = list(df_base.columns)
42 estados = estados[1:]
43
44 df_base = df_base.set_index('ANO')
45 df_base.index.name = None
46 tabela_censo = pd.DataFrame(columns=estados, index=anos)
47
48 #Interpolação linear para obter um numero de população por ano ao
    inves do range de 10 anos no censo
49 for estado in estados:
50     populacao_2010 = int(df_base.at[2010, estado])
51     populacao_2022 = int(df_base.at[2022, estado])
52
53     coeficiente_linear = (populacao_2022 - populacao_2010) / (2022 - 2010)
54     populacao_x = populacao_2010
55
56     for ano in anos: #Corrige possíveis números de população
        fracionados.
57
58         tabela_censo.at[ano, estado] = round(populacao_x)
59         populacao_x = populacao_x + coeficiente_linear
60
61 #Reestrutura dados para deixar em formato de subir para o banco de
    dados
62 tabela_censo.reset_index(inplace=True)
63 tabela_censo.rename(columns={'index': 'ANO'}, inplace=True)
64 tabela_censo = tabela_censo.transpose()
65
66 de_para = tabela_censo.iloc[0]
67 tabela_censo = tabela_censo.iloc[1:]
68
69 tabela_censo_final = pd.DataFrame(columns=['Ano', 'Estado', 'Populacao '
    ])
70 for index, ano in de_para.items(): #Loop Ano
71     for row2 in tabela_censo.itertuples(): #Loop Estado
72         _Ano = ano
73         Index_column = index + 1
74         Estado = row2.Index
75         Populacao = row2[Index_column]
```

```
76
77     dataset = {'Ano':_Ano, 'Estado':Estado, 'Populacao':Populacao}
78     tabela_transitoria = pd.DataFrame(data=[dataset])
79     tabela_censo_final = pd.concat([tabela_censo_final,
80                                     tabela_transitoria], ignore_index=True)
81
82
83 #Sobe tabela para o banco de dados
84 tabela_censo_final.to_sql('POPULACAO_CENSO_IBGE', conn, if_exists='
85     replace', index=False)
86
87
88 # %% [markdown]
89 # ### **1.0) VISAO PREÇO DE TARIFAS POR ANO AO LONGO DOS ANOS **
90 #
91 # * An lise feita por grupo de clientes (A e B)
92 # * Contratos com a ANEEL foram serializados em quantidades de dias
93   vigentes em um determinado ano
94 # * Tarifa de aplicação levada em consideração e a tarifa base
95   descartada
96 # * Fator de ponderamento do consumo Ponta e Fora Ponta (X e Y) levou
97   em consideração curva média das cargas
98 #
99
100 # %%
101 #Carrega tabela de inflação por ano
102
103 query = """
104
105     SELECT *
106     FROM INFLACAO_ACUMULADA_POR_ANO
107
108 """
109
```

```
110 tabelaInflacaoPorAno = pd.read_sql(query, conn)
111 tabelaInflacaoPorAno['acumuladoAnoAnterior'] = pd.to_datetime(
    tabelaInflacaoPorAno['acumuladoAnoAnterior'])
112 print(tabelaInflacaoPorAno)
113
114 #Função para acumular as as taxas
115
116 def aplicaValorAcumuladoIndice(df):
117
118     dfCopy = df.copy()
119
120     for linha in dfCopy.itertuples():
121         if linha.Index == 0:
122             df.at[linha.Index, 'acumuladoIPCA'] = float(1)
123             df.at[linha.Index, 'acumuladoIGPM'] = float(1)
124             df.at[linha.Index, 'acumuladoAumentoPercentualCustoGlobal'
125                 ] = float(1)
126
127             continue
128
129         elif linha.Index == 1:
130             ipca = linha.IPCA/100
131             ipca = ipca + 1
132
133             igpm = linha.IGPM/100
134             igpm = igpm + 1
135
136             aumentoGlobal = linha.AumentoPercentualCustoGlobal/100
137             aumentoGlobal = aumentoGlobal + 1
138
139             df.at[linha.Index, 'acumuladoIPCA'] = ipca
140             df.at[linha.Index, 'acumuladoIGPM'] = igpm
141             df.at[linha.Index, 'acumuladoAumentoPercentualCustoGlobal'
142                 ] = aumentoGlobal
143
144         elif linha.Index > 1:
145
146             indexAnterior = linha.Index
```

```

146         ipcaAnterior = df.at[(linha.Index-1), 'acumuladoIPCA']
147         igpmAnterior = df.at[(linha.Index-1), 'acumuladoIGPM']
148         aumentoGlobalAnterior = df.at[(linha.Index-1), '
           acumuladoAumentoPercentualCustoGlobal']
149
150         ipca = ((linha.IPCA)/100)
151         ipca = (ipca + 1)*ipcaAnterior
152
153         igpm = linha.IGPM/100
154         igpm = (igpm + 1)*igpmAnterior
155
156         aumentoGlobal = linha.AumentoPercentualCustoGlobal/100
157         aumentoGlobal = (aumentoGlobal + 1)*aumentoGlobalAnterior
158
159         df.at[linha.Index, 'acumuladoIPCA'] = ipca
160         df.at[linha.Index, 'acumuladoIGPM'] = igpm
161         df.at[linha.Index, 'acumuladoAumentoPercentualCustoGlobal'
           ] = aumentoGlobal
162
163     return df
164
165 # %%
166 #Gr fico por UF de financeiro inadiplente por estado
167
168 def _break_year_tax(df):
169
170     """Funcao que recebe datas de inicio e termino de vigencia da
           tarifa, calcula quantos dias
171     passou em cada ano e gera um novo dataframe serializado com
           dias por ano de cada preço de tarifa"""
172
173     def right_storage(ano, qty_dias, row):
174         dataset = {'Ano': ano,
175                   'Qty_dias': qty_dias,
176                   'DscBaseTarifaria': row.DscBaseTarifaria,
177                   'DscSubGrupo': row.DscSubGrupo,
178                   'NomPostoTarifario': row.NomPostoTarifario,
179                   'DscUnidadeTerciaria': row.DscUnidadeTerciaria,

```

```

180         'DscModalidadeTarifaria': row.
            DscModalidadeTarifaria,
181         'VlrTUSD': row.VlrTUSD,
182         'VlrTE': row.VlrTE,
183         'SigAgente': row.SigAgente,
184         'SigUF': row.SigUF,
185         'RegUF': row.RegUF}
186
187     return pd.DataFrame(data=[dataset]) #Retorna dataframe de uma
            linha no formato final
188
189
190 df['DatInicioVigencia'] = pd.to_datetime(df['DatInicioVigencia'])
191 df['DatFimVigencia'] = pd.to_datetime(df['DatFimVigencia'])
192
193 df_consolidado = pd.DataFrame(columns=['Ano', 'Qty_dias', '
            DscBaseTarifaria', 'DscSubGrupo', 'NomPostoTarifario', '
            DscUnidadeTerciaria', \
194                                     'DscModalidadeTarifaria',
            'VlrTUSD', 'VlrTE', '
            SigAgente', 'SigUF', '
            RegUF'])
195
196 for row in df.itertuples():
197     _data_inicio = row.DatInicioVigencia
198     _data_termino = row.DatFimVigencia
199     _delta_dias = (_data_termino - _data_inicio).days
200
201     _ano_inicio = (row.DatInicioVigencia).year
202     _mes_inicio = (row.DatInicioVigencia).month
203     _dia_inicio = (row.DatInicioVigencia).day
204
205     _ano_termino = (row.DatFimVigencia).year
206     _mes_termino = (row.DatFimVigencia).month
207     _dia_termino = (row.DatFimVigencia).day
208
209     #cria data de referência para anos fechados após a data
            inicio e anterior a data de termino

```

```

210     _ref_ano_inicio = dt.strptime(f'01/01/{_ano_inicio+1}', "%d/%
      m/%Y")
211     _ref_ano_termino = dt.strptime(f'31/12/{_ano_termino-1}', "%d
      /%m/%Y")
212
213     if _ano_inicio == _ano_termino:
214         df_consolidado = pd.concat([df_consolidado, right_storage(
      _ano_inicio, _delta_dias, row)], ignore_index=True)
215
216     else:
217         all_years_range = [x for x in range(_ano_inicio+1,
      _ano_termino)]
218
219         #Contabiliza os dias para fechar o ano na ponta inicial e
      final do range
220         dif_fisrt_year = (_ref_ano_inicio - _data_inicio).days
221         df_consolidado = pd.concat([df_consolidado, right_storage(
      _ano_inicio, dif_fisrt_year, row)], ignore_index=True)
222         dif_last_year = (_data_termino - _ref_ano_termino).days
223         df_consolidado = pd.concat([df_consolidado, right_storage(
      _ano_termino, dif_last_year, row)], ignore_index=True)
224
225         for ano in all_years_range:
226             df_consolidado = pd.concat([df_consolidado,
      right_storage(ano, 365, row)], ignore_index=True)
227
228     return df_consolidado
229
230
231 #Vari vel para as duas tabelas utilizadas na query
232 tb1 = "DIST_TARIFAS_DISTRIBUIDORAS"
233 tb_sigAg = "SIGAGENTE_UF"
234
235 #Select com as informações pertinentes para a an lise , filtrando
      pela modalidade tarifaria de Distribuição (propósito do estudo) e
      filtrando pelos postos tarifarios Ponta e Fora Ponta.
236 query = f""" SELECT {tb1}.DatInicioVigencia, {tb1}.DatFimVigencia, {
      tb1}.DscBaseTarifaria, {tb1}.DscModalidadeTarifaria, {tb1}.
      DscSubGrupo, {tb1}.NomPostoTarifario, {tb1}.DscUnidadeTerciaria,\

```

```

237         {tb1}.VlrTUSD, {tb1}.VlrTE, {tb_sigAg}.SigAgente, {
238             tb_sigAg}.SigUF, {tb_sigAg}.RegUF
239
240     FROM {tb1}
241     INNER JOIN {tb_sigAg} ON {tb1}.SigAgente = {tb_sigAg}.
242         SigAgente
243     WHERE {tb1}.DscBaseTarifaria = 'Tarifa de Aplicação'\
244         AND {tb1}.SigAgente = ''\
245         AND {tb1}.DscSubClasse <> 'Baixa Renda'\
246         AND {tb1}.DscSubClasse <> 'Serviço público de
247             irrigação rural'\
248         AND {tb1}.DscSubClasse <> 'Cooperativa de
249             eletrificação rural'\
250         AND {tb1}.DscModalidadeTarifaria <> 'Geração'\
251         AND {tb1}.NomPostoTarifario <> 'Fora ponta seca'\
252         AND {tb1}.NomPostoTarifario <> 'Fora ponta úmida
253             '\
254         AND {tb1}.NomPostoTarifario <> 'Ponta seca'\
255         AND {tb1}.NomPostoTarifario <> 'Ponta úmida'\
256         AND {tb1}.DscModalidadeTarifaria <> 'Azul ABRACE
257             CATIVO'\
258         AND {tb1}.DscModalidadeTarifaria <> 'Verde ABRACE
259             CATIVO'\
260         AND {tb1}.DscModalidadeTarifaria <> 'Azul ABRACE
261             LIVRE'\
262         AND {tb1}.DscModalidadeTarifaria <> 'Convencional
263             ABRACE'\
264         AND {tb1}.DscModalidadeTarifaria <> 'Verde ABRACE
265             LIVRE'\
266
267         ""
268
269     #({tb1}.DscModalidadeTarifaria = 'Distribuição' OR {tb1}.
270         DscModalidadeTarifaria = 'Convencional')
271
272     result = pd.read_sql_query(query,conn)
273     # result.to_excel(r".\APAGAR_tarifas.xlsx", index=False)
274     result = _break_year_tax(result)
275
276
277

```

```

265
266
267
268 # %%
269
270 result = result.copy()
271
272 #Remove linhas duplicadas e tira as tarifas de distribuição zeradas
    para não atrapalhar nos calculos estatísticos.
273 result = result.drop(result[result['VlrTUSD'] == ',00'].index)
274 result = result.drop_duplicates()
275
276 #Para valores dados por MWh, transforma a base para KWh
277 result['VlrTUSD'] = result['VlrTUSD'].str.replace(',','').astype(
    float)
278 for row in result.itertuples():
279     if row.DscUnidadeTerciaria == 'MWh':
280         result.at[row.Index, 'DscUnidadeTerciaria'] = 'KWh'
281         result.at[row.Index, 'VlrTUSD'] = (row.VlrTUSD/1000)
282
283 # result.to_excel(r"C:\Users\marco\Desktop\TCC\SEGREG_TARIFAS.xlsx",
    index = False)
284
285 #Consome condições de contorno para análise das tarifas
286 _inicial_condition = pd.read_excel(r"C:\Users\marco\Desktop\TCC\
    CONDICoes_CONTORNO_TARIFAS.xlsx")
287 result.to_excel(r"C:\Users\marco\Desktop\visao_trifas.xlsx", index=
    False)
288
289
290
291 # %% [markdown]
292 # ##### **1.1) An lise para consumidores de Baixa Tensão (GRUPO B)**
293
294 # %%
295 _df_baixa_tensao = result.loc[(result['DscSubGrupo']=='B1') | (result
    ['DscSubGrupo']=='B2') | (result['DscSubGrupo']=='B3')].copy() #
    Filtra grupos de baixa tensão

```



```

296 _df_baixa_tensao = _df_baixa_tensao.loc[_df_baixa_tensao['
      NomPostoTarifario'] == 'Não se aplica'] #Tira da analise a
      diferenciação de consumo ponta e fora ponta
297 _df_baixa_tensao = _df_baixa_tensao.loc[_df_baixa_tensao['
      DscModalidadeTarifaria'] == 'Convencional'] #Tira da analise os
      beneficios sociais nas tarifas
298
299 _df_bt_layout = pd.DataFrame(columns =_df_baixa_tensao.columns)
300 _df_bt_consolidado = pd.DataFrame(columns =_df_baixa_tensao.columns)
301
302 #Pega dias de cada ano
303 _df_dias_ano = _df_baixa_tensao[['Ano','Qty_dias']].copy()
304 _df_dias_ano.drop_duplicates(inplace=True)
305 _df_dias_ano = _df_dias_ano.groupby(by='Ano')['Qty_dias'].sum()
306 _df_dias_ano = _df_dias_ano.reset_index()
307
308 #Multiplica Qty_dias por preço da energia, gerando o numerador da
      media ponderada
309 _df_baixa_tensao['VlrTUSD'] = _df_baixa_tensao['VlrTUSD'] *
      _df_baixa_tensao['Qty_dias']
310 _df_baixa_tensao = _df_baixa_tensao.groupby(by=['Ano','
      DscBaseTarifaria','DscSubGrupo','DscModalidadeTarifaria'])['
      VlrTUSD'].sum()
311 _df_baixa_tensao = _df_baixa_tensao.reset_index()
312 # print(_df_baixa_tensao)
313
314 #Divide todos os valores pelo numero de dias do ano em questao,
      representando o denominador da media ponderada
315 _df_baixa_tensao = _df_baixa_tensao.merge(_df_dias_ano, how='left',
      on='Ano')
316 _df_baixa_tensao['VlrTUSD'] = _df_baixa_tensao['VlrTUSD'] /
      _df_baixa_tensao['Qty_dias']
317
318 #Mantem colunas importantes
319 _df_baixa_tensao = _df_baixa_tensao[['Ano','DscSubGrupo','VlrTUSD']]
320
321 #Cruza condições iniciais x resultado da media ponderada das tarifas
322 _df_baixa_tensao = _df_baixa_tensao.merge(_inicial_condition,how='
      left',on='DscSubGrupo')

```

```

323 _df_baixa_tensao['CustoGlobalAnual'] = (_df_baixa_tensao['VlrTUSD'] *
      _df_baixa_tensao['Consumo_kWh']).apply(lambda x: round(x,2))
324 _df_baixa_tensao['Ano'] = pd.to_datetime(_df_baixa_tensao['Ano'],
      format='%Y')
325
326 #Cria segunda tabela que aglutina todos os subgrupos do grupo de
      consumidor B + traz os indices de correção financeira
327 TarifaBaixaTensaoAnaliseEconomica = _df_baixa_tensao.copy()
328 TarifaBaixaTensaoAnaliseEconomica = pd.DataFrame(
      TarifaBaixaTensaoAnaliseEconomica.groupby(['Ano'])['
      CustoGlobalAnual'].sum())
329 TarifaBaixaTensaoAnaliseEconomica.reset_index(inplace=True)
330 TarifaBaixaTensaoAnaliseEconomica['AumentoPercentualCustoGlobal'] =
      TarifaBaixaTensaoAnaliseEconomica['CustoGlobalAnual'].pct_change()
      *100
331 TarifaBaixaTensaoAnaliseEconomica = TarifaBaixaTensaoAnaliseEconomica
      .merge(tabelaInflacaoPorAno, how='left', left_on='Ano', right_on='
      acumuladoAnoAnterior')
332 TarifaBaixaTensaoAnaliseEconomica = TarifaBaixaTensaoAnaliseEconomica
      .drop(columns=['acumuladoAnoAnterior'])
333 TarifaBaixaTensaoAnaliseEconomica.fillna(0, inplace=True)
334 TarifaBaixaTensaoAnaliseEconomica['mediaIndicesFinanceiros'] = (
      TarifaBaixaTensaoAnaliseEconomica['IPCA'] +
      TarifaBaixaTensaoAnaliseEconomica['IGPM'])/2
335
336 TarifaBaixaTensaoAnaliseEconomica['acumuladoIPCA'] = ""
337 TarifaBaixaTensaoAnaliseEconomica['acumuladoIGPM'] = ""
338 TarifaBaixaTensaoAnaliseEconomica['acumuladoMediaIndices'] = ""
339 TarifaBaixaTensaoAnaliseEconomica['
      acumuladoAumentoPercentualCustoGlobal'] = ""
340
341 TarifaBaixaTensaoAnaliseEconomica = aplicaValorAcumuladoIndice(
      TarifaBaixaTensaoAnaliseEconomica)
342 TarifaBaixaTensaoAnaliseEconomica['acumuladoMediaIndices'] = (
      TarifaBaixaTensaoAnaliseEconomica['acumuladoIPCA'] +
      TarifaBaixaTensaoAnaliseEconomica['acumuladoIGPM'])/2
343
344 TarifaBaixaTensaoAnaliseEconomica['acumuladoIPCA'] =
      TarifaBaixaTensaoAnaliseEconomica['acumuladoIPCA']-1

```

```

345 TarifaBaixaTensaoAnaliseEconomica['acumuladoIGPM'] =
      TarifaBaixaTensaoAnaliseEconomica['acumuladoIGPM']-1
346 TarifaBaixaTensaoAnaliseEconomica['acumuladoMediaIndices'] =
      TarifaBaixaTensaoAnaliseEconomica['acumuladoMediaIndices']-1
347 TarifaBaixaTensaoAnaliseEconomica['
      acumuladoAumentoPercentualCustoGlobal'] =
      TarifaBaixaTensaoAnaliseEconomica['
      acumuladoAumentoPercentualCustoGlobal']-1
348
349
350 #Sobe tabela final pro BD do power BI
351 _df_baixa_tensao.to_sql('TarifaBaixaTensao',conn_bi, if_exists='
      replace', index=False)
352 TarifaBaixaTensaoAnaliseEconomica.to_sql('
      TarifaBaixaTensaoAnaliseEconomica',conn_bi, if_exists='replace',
      index=False)
353
354 print('Atualizada base TarifaBaixaTensao no BD.')
355
356
357
358 # %% [markdown]
359 # ##### **1.2) An lise para consumidores de Alta Tensão (GRUPO A)**
360
361 # %%
362 _df_alta_tensao = result.loc[(result['DscSubGrupo']=='A1') | (result[
      'DscSubGrupo']=='A3') | (result['DscSubGrupo']=='A4')].copy() #
      Filtra grupos de baixa tensão
363 # print(_df_alta_tensao)
364
365 _df_at_layout = pd.DataFrame(columns =_df_alta_tensao.columns)
366 _df_at_consolidado = pd.DataFrame(columns =_df_alta_tensao.columns)
367
368 #Pega dias de cada ano
369 _df_dias_ano = _df_alta_tensao[['Ano', 'Qty_dias']].copy()
370 _df_dias_ano.drop_duplicates(inplace=True)
371 _df_dias_ano = _df_dias_ano.groupby(by='Ano')['Qty_dias'].sum()
372 _df_dias_ano = _df_dias_ano.reset_index()
373

```

```
374
375 #Multiplica Qty_dias por preço da energia, gerando o numerador da
      media ponderada
376 _df_alta_tensao['VlrTUSD'] = _df_alta_tensao['VlrTUSD'] *
      _df_alta_tensao['Qty_dias']
377 _df_alta_tensao = _df_alta_tensao.groupby(by=['Ano', 'DscBaseTarifaria',
      'DscSubGrupo', 'NomPostoTarifario', 'DscUnidadeTerciaria', 'DscModalidadeTarifaria'])['VlrTUSD'].sum()
378 _df_alta_tensao = _df_alta_tensao.reset_index()
379
380
381 #Divide todos os valores pelo numero de dias do ano em questao,
      representando o denominador da media ponderada
382 _df_alta_tensao = _df_alta_tensao.merge(_df_dias_ano, how='left', on=
      'Ano')
383 _df_alta_tensao['VlrTUSD'] = _df_alta_tensao['VlrTUSD'] /
      _df_alta_tensao['Qty_dias']
384 _df_alta_tensao.to_excel(r"C:\Users\marco\Desktop\teste_parcial.xlsx"
      ,index=False)
385
386
387
388 # %% [markdown]
389 # ##### **1.2.1) An lise para consumidores de Alta Tensão (GRUPO A1)
      **
390 #
391 # fator ponta = 60%
392 # fator fora ponta = 40%
393
394 # %%
395 _df_alta_tensao_a1 = _df_alta_tensao.loc[_df_alta_tensao['DscSubGrupo']
      == 'A1'].copy()
396 _df_alta_tensao_a1 = _df_alta_tensao_a1[['Ano', 'DscSubGrupo', '
      NomPostoTarifario', 'DscUnidadeTerciaria', 'DscModalidadeTarifaria',
      'VlrTUSD']]
397
398
399 #Guarda dataframe para anos de 2010 e 2011 por serem de
      caracteristicas distintas
```

```

400 _df_linhas_especiais = _df_alta_tensao_a1.loc[( _df_alta_tensao_a1['
      Ano'] == 2010) | ( _df_alta_tensao_a1['Ano'] == 2011)].copy()
401
402
403 #Aplica os fatores de ponta e fora ponta nas tarifas
404 _df_fora_ponta = _df_alta_tensao_a1[_df_alta_tensao_a1['
      NomPostoTarifario'] == 'Fora ponta'].copy()
405 _df_fora_ponta['VlrTUSD'] = _df_fora_ponta['VlrTUSD'].apply(lambda x:
      x*(0.4))
406 _df_ponta = _df_alta_tensao_a1[_df_alta_tensao_a1['NomPostoTarifario'
      ] == 'Ponta'].copy()
407 _df_ponta['VlrTUSD'] = _df_ponta['VlrTUSD'].apply(lambda x: x*(0.6))
408
409 _df_alta_tensao_a1 = pd.concat([_df_fora_ponta, _df_ponta],
      ignore_index=True)
410 _df_alta_tensao_a1 = _df_alta_tensao_a1.loc[( _df_alta_tensao_a1['Ano'
      ] != 2010) & ( _df_alta_tensao_a1['Ano'] != 2011)].copy() #Exclui
      anos com linhas especiais
411 _df_alta_tensao_a1 = pd.concat([_df_alta_tensao_a1,
      _df_linhas_especiais], ignore_index=True) #Adiciona linhas
      especiais do df antigo
412
413 #Soma tarifas de ponta e fora ponta com os fatores aplicados, para
      consumo de energia e consumo de tensão
414 _df_lista = []
415
416 for ano in set(_df_alta_tensao_a1['Ano']):
417     if ((ano == 2010 ) | (ano == 2011)): #Para os anos de linhas
      especiais, não trata o valor da tarifa, apenas replica a de
      ponta
418         df = _df_alta_tensao_a1.loc[( _df_alta_tensao_a1['Ano'] == ano)
      & ( _df_alta_tensao_a1['NomPostoTarifario'] == 'Ponta')]
419         df = df[['Ano', 'DscSubGrupo', 'DscUnidadeTerciaria', 'VlrTUSD'
      ]]
420         _df_lista.append(df)
421         continue
422
423     df = _df_alta_tensao_a1.loc[_df_alta_tensao_a1['Ano'] == ano]

```

```

424     df = df.groupby(by=['Ano', 'DscSubGrupo', 'DscUnidadeTerciaria'])[
         'VlrTUSD'].sum()
425     df = df.reset_index()
426     _df_lista.append(df)
427
428     #Concatena a lista de dfs por ano
429     _df_alta_tensao_a1 = pd.concat(_df_lista, ignore_index=True)
430
431     #Mantem colunas importantes para calculo consumo x ano e adiciona a
         chave para o merge
432     _df_alta_tensao_a1 = _df_alta_tensao_a1[['Ano', 'DscSubGrupo', '
         DscUnidadeTerciaria', 'VlrTUSD']]
433
434     tensao_base = _inicial_condition.iloc[0,1]
435     consumo_base = _inicial_condition.iloc[0,2]
436
437
438     #Cruza condicoes iniciais x resultado da media ponderada das tarifas
439     _df_a1_consumo = _df_alta_tensao_a1[_df_alta_tensao_a1['
         DscUnidadeTerciaria']=='KWh'].copy()
440     _df_a1_consumo['CustoGlobalAnual'] = _df_a1_consumo['VlrTUSD'].apply(
         lambda x: x*consumo_base)
441     _df_a1_tensao = _df_alta_tensao_a1[_df_alta_tensao_a1['
         DscUnidadeTerciaria']=='kW'].copy()
442     _df_a1_tensao['CustoGlobalAnual'] = _df_a1_tensao['VlrTUSD'].apply(
         lambda x: x*tensao_base)
443
444     _df_alta_tensao_a1 = pd.concat([_df_a1_consumo, _df_a1_tensao],
         ignore_index=True)
445
446     _df_alta_tensao_a1['CustoGlobalAnual'] = _df_alta_tensao_a1['
         CustoGlobalAnual'].apply(lambda x: round(x,2))
447     _df_alta_tensao_a1['Ano'] = pd.to_datetime(_df_alta_tensao_a1['Ano'],
         format='%Y')
448
449
450     _df_alta_tensao_a1.to_sql('TarifaAltaTensaoA1', conn_bi, if_exists='
         replace', index=False)
451     print(_df_alta_tensao_a1)

```

```
452
453 #
454
455 print('Atualizada base TarifaAltaTensaoA1 no BD.')
456
457
458
459
460
461
462
463 # %% [markdown]
464 # ##### **1.2.2) An lise para consumidores de Alta Tensão (GRUPO A3)
465 # **
466 #
467 # fator ponta = 60%
468 # fator fora ponta = 40%
469 # %%
470 _df_alta_tensao_a3 = _df_alta_tensao.loc[( _df_alta_tensao['
471     DscSubGrupo'] == 'A3') & ( _df_alta_tensao['DscModalidadeTarifaria'
472     ] != 'Convencional') ].copy()
473 _df_alta_tensao_a3 = _df_alta_tensao_a3[['Ano', 'DscSubGrupo', '
474     NomPostoTarifario', 'DscUnidadeTerciaria', 'DscModalidadeTarifaria',
475     'VlrTUSD']]
476
477 #Aplica os fatores de ponta e fora ponta nas tarifas
478 _df_fora_ponta = _df_alta_tensao_a3[_df_alta_tensao_a3['
479     NomPostoTarifario'] == 'Fora ponta'].copy()
480 _df_fora_ponta['VlrTUSD'] = _df_fora_ponta['VlrTUSD'].apply(lambda x:
481     x*(0.4))
482 _df_ponta = _df_alta_tensao_a3[_df_alta_tensao_a3['NomPostoTarifario'
483     ] == 'Ponta'].copy()
484 _df_ponta['VlrTUSD'] = _df_ponta['VlrTUSD'].apply(lambda x: x*(0.6))
485
486 _df_alta_tensao_a3 = pd.concat([_df_fora_ponta, _df_ponta],
487     ignore_index=True)
488
```

```
481 #Soma tarifas de ponta e fora ponta com os fatores aplicados, para
    consumo de energia e consumo de tensão
482 _df_lista = []
483
484 for ano in set(_df_alta_tensao_a3['Ano']):
485     df = _df_alta_tensao_a3.loc[_df_alta_tensao_a3['Ano']==ano]
486     df = df.groupby(by=['Ano', 'DscSubGrupo', 'DscUnidadeTerciaria'])['
        VlrTUSD'].sum()
487     df = df.reset_index()
488     _df_lista.append(df)
489
490
491 #Concatena a lista de dfs por ano
492 _df_alta_tensao_a3 = pd.concat(_df_lista, ignore_index=True)
493
494 #Mantem colunas importantes para calculo consumo x ano e adiciona a
    chave para o merge
495 _df_alta_tensao_a3 = _df_alta_tensao_a3[['Ano', 'DscSubGrupo', '
        DscUnidadeTerciaria', 'VlrTUSD']]
496
497 tensao_base = _inicial_condition.iloc[1,1]
498 consumo_base = _inicial_condition.iloc[1,2]
499
500 #Cruza condições iniciais x resultado da media ponderada das tarifas
501 _df_a3_consumo = _df_alta_tensao_a3[_df_alta_tensao_a3['
        DscUnidadeTerciaria']=='KWh'].copy()
502 _df_a3_consumo['CustoGlobalAnual'] = _df_a3_consumo['VlrTUSD'].apply(
    lambda x: x*consumo_base)
503 _df_a3_tensao = _df_alta_tensao_a3[_df_alta_tensao_a3['
        DscUnidadeTerciaria']=='kW'].copy()
504 _df_a3_tensao['CustoGlobalAnual'] = _df_a3_tensao['VlrTUSD'].apply(
    lambda x: x*tensao_base)
505
506 _df_alta_tensao_a3 = pd.concat([_df_a3_consumo, _df_a3_tensao],
    ignore_index=True)
507
508 _df_alta_tensao_a3['CustoGlobalAnual'] = _df_alta_tensao_a3['
    CustoGlobalAnual'].apply(lambda x: round(x,2))
```



```

509 _df_alta_tensao_a3['Ano'] = pd.to_datetime(_df_alta_tensao_a3['Ano'],
      format='%Y')
510
511
512 _df_alta_tensao_a3.to_sql('TarifaAltaTensaoA3',conn_bi, if_exists='
      replace', index=False)
513 print('Atualizada base TarifaAltaTensaoA3 no BD.')
514
515
516
517 # %% [markdown]
518 # ##### **1.2.3) An lise para consumidores de Alta Tensão (GRUPO A4)
      **
519 #
520 # fator ponta = 60%
521 # fator fora ponta = 40%
522
523 # %%
524 _df_alta_tensao_a4 = _df_alta_tensao.loc[( _df_alta_tensao['
      DscSubGrupo'] == 'A4') & ( _df_alta_tensao['DscModalidadeTarifaria'
      ] != 'Convencional') & ( _df_alta_tensao['NomPostoTarifario'] != '
      Não se aplica') ].copy()
525 _df_alta_tensao_a4 = _df_alta_tensao_a4[['Ano', 'DscSubGrupo', '
      NomPostoTarifario', 'DscUnidadeTerciaria', 'DscModalidadeTarifaria',
      'VlrTUSD']]
526
527 #Aplica os fatores de ponta e fora ponta nas tarifas
528 _df_fora_ponta = _df_alta_tensao_a4[_df_alta_tensao_a4['
      NomPostoTarifario']=='Fora ponta'].copy()
529 _df_fora_ponta['VlrTUSD'] = _df_fora_ponta['VlrTUSD'].apply(lambda x:
      x*(0.4))
530 _df_ponta = _df_alta_tensao_a4[_df_alta_tensao_a4['NomPostoTarifario'
      ]=='Ponta'].copy()
531 _df_ponta['VlrTUSD'] = _df_ponta['VlrTUSD'].apply(lambda x: x*(0.6))
532
533 _df_alta_tensao_a4 = pd.concat([_df_fora_ponta, _df_ponta],
      ignore_index=True)
534

```

```
535 #Soma tarifas de ponta e fora ponta com os fatores aplicados, para
    consumo de energia e consumo de tensão
536 _df_lista = []
537
538 for ano in set(_df_alta_tensao_a4['Ano']):
539     df = _df_alta_tensao_a4.loc[_df_alta_tensao_a4['Ano']==ano]
540     df = df.groupby(by=['Ano', 'DscSubGrupo', 'DscUnidadeTerciaria', '
        DscModalidadeTarifaria'])['VlrTUSD'].sum()
541     df = df.reset_index()
542     _df_lista.append(df)
543
544 #Concatena a lista de dfs por ano
545 _df_alta_tensao_a4 = pd.concat(_df_lista, ignore_index=True)
546
547 #Mantem colunas importantes para calculo consumo x ano e adiciona a
    chave para o merge
548 _df_alta_tensao_a4 = _df_alta_tensao_a4[['Ano', 'DscSubGrupo', '
        DscUnidadeTerciaria', 'DscModalidadeTarifaria', 'VlrTUSD']]
549
550
551 tensao_base = _inicial_condition.iloc[2,1]
552 consumo_base = _inicial_condition.iloc[2,2]
553
554 #Cruza condições iniciais x resultado da media ponderada das tarifas
555 _df_a4_consumo = _df_alta_tensao_a4[_df_alta_tensao_a4['
        DscUnidadeTerciaria']=='KWh'].copy()
556 _df_a4_consumo['CustoGlobalAnual'] = _df_a4_consumo['VlrTUSD'].apply(
    lambda x: x*consumo_base)
557 _df_a4_tensao = _df_alta_tensao_a4[_df_alta_tensao_a4['
        DscUnidadeTerciaria']=='kW'].copy()
558 _df_a4_tensao['CustoGlobalAnual'] = _df_a4_tensao['VlrTUSD'].apply(
    lambda x: x*tensao_base)
559
560 _df_alta_tensao_a4 = pd.concat([_df_a4_consumo, _df_a4_tensao],
    ignore_index=True)
561
562 _df_alta_tensao_a4['CustoGlobalAnual'] = _df_alta_tensao_a4['
    CustoGlobalAnual'].apply(lambda x: round(x,2))
```

```
563 _df_alta_tensao_a4['Ano'] = pd.to_datetime(_df_alta_tensao_a4['Ano'],
564     format='%Y')
565
566 _df_alta_tensao_a4.to_sql('TarifaAltaTensaoA4',conn_bi, if_exists='
567     replace', index=False)
568
569 # %%
570 def aplicaValorAcumuladoIndice(df):
571
572     dfCopy = df.copy()
573
574     for linha in dfCopy.itertuples():
575         if linha.Index == 0:
576             df.at[linha.Index, 'acumuladoIPCA'] = float(1)
577             df.at[linha.Index, 'acumuladoIGPM'] = float(1)
578             df.at[linha.Index, 'acumuladoAumentoPercentualCustoGlobal'
579                 ] = float(1)
580
581             continue
582
583         elif linha.Index == 1:
584             ipca = linha.IPCA/100
585             ipca = ipca + 1
586
587             igpm = linha.IGPM/100
588             igpm = igpm + 1
589
590             aumentoGlobal = linha.AumentoPercentualCustoGlobal/100
591             aumentoGlobal = aumentoGlobal + 1
592
593             df.at[linha.Index, 'acumuladoIPCA'] = ipca
594             df.at[linha.Index, 'acumuladoIGPM'] = igpm
595             df.at[linha.Index, 'acumuladoAumentoPercentualCustoGlobal'
596                 ] = aumentoGlobal
597
598         elif linha.Index > 1:
```

```

598     indexAnterior = linha.Index
599
600     ipcaAnterior = df.at[(linha.Index-1), 'acumuladoIPCA']
601     igpmAnterior = df.at[(linha.Index-1), 'acumuladoIGPM']
602     aumentoGlobalAnterior = df.at[(linha.Index-1), '
        acumuladoAumentoPercentualCustoGlobal']
603
604     ipca = ((linha.IPCA)/100)
605     ipca = (ipca + 1)*ipcaAnterior
606
607     igpm = linha.IGPM/100
608     igpm = (igpm + 1)*igpmAnterior
609
610     aumentoGlobal = linha.AumentoPercentualCustoGlobal/100
611     aumentoGlobal = (aumentoGlobal + 1)*aumentoGlobalAnterior
612
613     df.at[linha.Index, 'acumuladoIPCA'] = ipca
614     df.at[linha.Index, 'acumuladoIGPM'] = igpm
615     df.at[linha.Index, 'acumuladoAumentoPercentualCustoGlobal'
        ] = aumentoGlobal
616
617     return df
618
619
620
621 #Cria segunda tabela que aglutina todos os subgrupos do grupo de
        consumidor A + traz os indices de correçao financeira
622
623 listaTabelasConcatenadas = [_df_alta_tensao_a1, _df_alta_tensao_a3,
        _df_alta_tensao_a4]
624
625 TarifaAltaTensaoAnaliseEconomica = pd.concat(listaTabelasConcatenadas
        )
626 TarifaAltaTensaoAnaliseEconomica = pd.DataFrame(
        TarifaAltaTensaoAnaliseEconomica.groupby(['Ano'])['
        CustoGlobalAnual'].sum())
627 TarifaAltaTensaoAnaliseEconomica.reset_index(inplace=True)
628 TarifaAltaTensaoAnaliseEconomica['AumentoPercentualCustoGlobal'] =
        TarifaAltaTensaoAnaliseEconomica['CustoGlobalAnual'].pct_change()

```

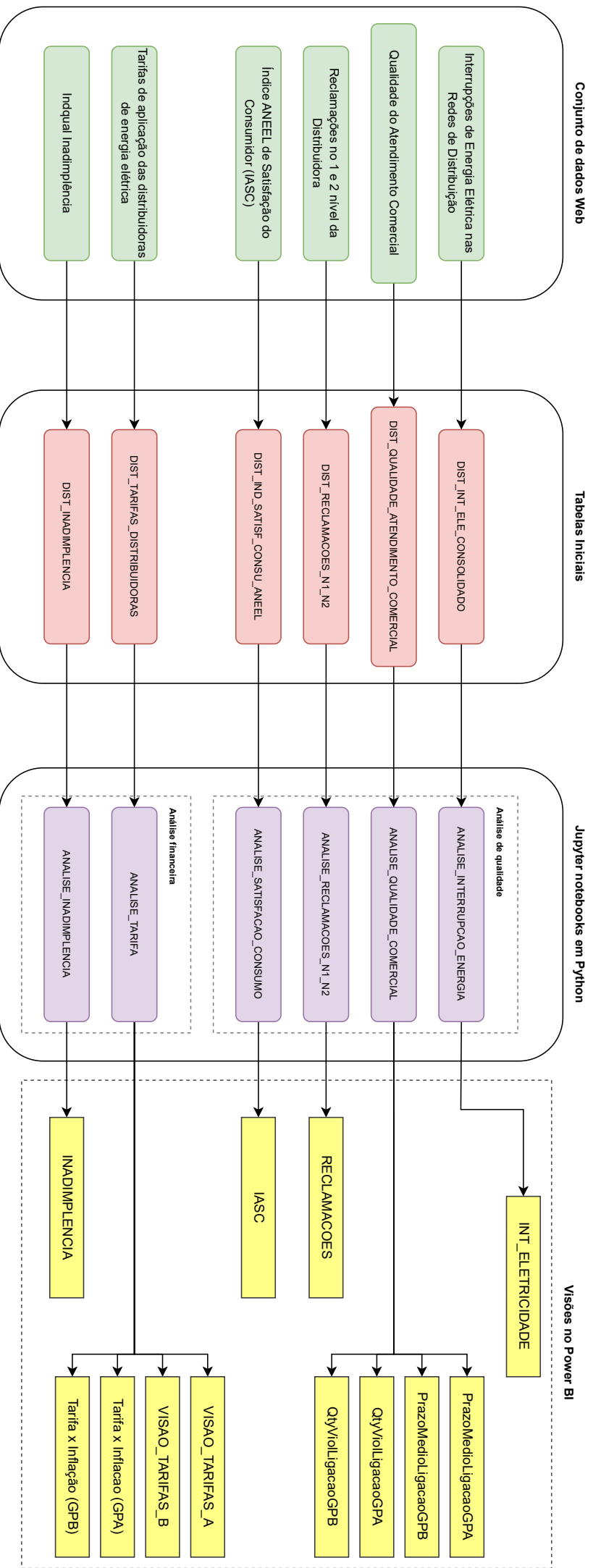
```

*100
629
630 TarifaAltaTensaoAnaliseEconomica = TarifaAltaTensaoAnaliseEconomica.
    merge(tabelaInflacaoPorAno, how='left', left_on='Ano', right_on='
        acumuladoAnoAnterior')
631 TarifaAltaTensaoAnaliseEconomica = TarifaAltaTensaoAnaliseEconomica.
    drop(columns=['acumuladoAnoAnterior'])
632 TarifaAltaTensaoAnaliseEconomica.fillna(0, inplace=True)
633 TarifaAltaTensaoAnaliseEconomica['mediaIndicesFinanceiros'] = (
    TarifaAltaTensaoAnaliseEconomica['IPCA'] +
    TarifaAltaTensaoAnaliseEconomica['IGPM'])/2
634
635 TarifaAltaTensaoAnaliseEconomica['acumuladoIPCA'] = ""
636 TarifaAltaTensaoAnaliseEconomica['acumuladoIGPM'] = ""
637 TarifaAltaTensaoAnaliseEconomica['acumuladoMediaIndices'] = ""
638 TarifaAltaTensaoAnaliseEconomica['
    acumuladoAumentoPercentualCustoGlobal'] = ""
639
640 TarifaAltaTensaoAnaliseEconomica = aplicaValorAcumuladoIndice(
    TarifaAltaTensaoAnaliseEconomica)
641 TarifaAltaTensaoAnaliseEconomica['acumuladoMediaIndices'] = (
    TarifaAltaTensaoAnaliseEconomica['acumuladoIPCA'] +
    TarifaAltaTensaoAnaliseEconomica['acumuladoIGPM'])/2
642
643 TarifaAltaTensaoAnaliseEconomica['acumuladoIPCA'] =
    TarifaAltaTensaoAnaliseEconomica['acumuladoIPCA']-1
644 TarifaAltaTensaoAnaliseEconomica['acumuladoIGPM'] =
    TarifaAltaTensaoAnaliseEconomica['acumuladoIGPM']-1
645 TarifaAltaTensaoAnaliseEconomica['acumuladoMediaIndices'] =
    TarifaAltaTensaoAnaliseEconomica['acumuladoMediaIndices']-1
646 TarifaAltaTensaoAnaliseEconomica['
    acumuladoAumentoPercentualCustoGlobal'] =
    TarifaAltaTensaoAnaliseEconomica['
    acumuladoAumentoPercentualCustoGlobal']-1
647
648 #Sobe tabela final pro BD do power BI
649 TarifaAltaTensaoAnaliseEconomica.to_sql('
    TarifaAltaTensaoAnaliseEconomica', conn_bi, if_exists='replace',
    index=False)

```

ANEXO B – ESTRUTURA MACRO EXPANDIDA

A seguir está disposta a estrutura macro desenvolvida na metodologia deste trabalho:



LEGENDA

- Bases de dados utilizadas do site da ANEEL
- Tabelas de cada uma das bases no banco de dados
- Algoritmos desenvolvidos em Python para cada uma das bases
- Visualizações finais no Power BI