



UNIVERSIDADE FEDERAL DE CAMPINA GRANDE  
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA  
CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA  
TRABALHO DE CONCLUSÃO DE CURSO

**APLICAÇÃO DE *ADVANCED ANALYTICS* NA GESTÃO DA PERDA  
NÃO TÉCNICA DE ENERGIA EM SISTEMA ELÉTRICO DE  
DISTRIBUIÇÃO**

Odenilson Santa Brígida Leite

Orientador: Prof Dr. Karcus Marcelus Colaço Dantas

Campina Grande, PB.

Agosto, 2022

ODENILSON SANTA BRÍGIDA LEITE

**APLICAÇÃO DE *ADVANCED ANALYTICS* NA GESTÃO DA PERDA  
NÃO TÉCNICA DE ENERGIA EM SISTEMA ELÉTRICO DE  
DISTRIBUIÇÃO**

Trabalho de Conclusão de Curso  
submetido à Coordenação de Graduação em  
Engenharia Elétrica da Universidade Federal  
de Campina Grande, Campus Campina  
Grande, como parte dos requisitos  
necessários para obtenção do título de  
Graduação em Engenharia elétrica.

Campina Grande, PB.

Agosto, 2022

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE  
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA  
CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA  
TRABALHO DE CONCLUSÃO DE CURSO

FOLHA DE APROVAÇÃO

**APLICAÇÃO DE *ADVANCED ANALYTICS* NA GESTÃO DA PERDA  
NÃO TÉCNICA DE ENERGIA EM SISTEMA ELÉTRICO DE  
DISTRIBUIÇÃO**

---

Odenilson Santa Brígida Leite

Orientando

---

Prof. Karcus Marcelus Colaço Dantas

Orientador

Campina Grande

2022

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus pelo milagre da vida e permitir ter pais que sacrificaram muito dos seus sonhos para me orientar, guiar e tendo como prioridade minha educação.

Agradeço a meus avós Maria Francisca e Edilberto Galucio que em vários momentos eram como luzes me motivando a sempre seguir em frente, acreditando nas minhas habilidades e potencial.

Agradeço a minha irmã Nanny Victoria e meus familiares por toda a torcida e apoio em muitos momentos difíceis, me proporcionando suporte na minha jornada.

Agradeço aos amigos que fiz ao longo do curso e que nunca me negaram ajuda e sempre foram prestativos, sempre juntos em diversos momentos, nos desenvolvendo e crescendo juntos. Obrigado meus amigos Mariana Dourado, Luan Bastos, Thiago Moura, Daniel Loula, Everaldo Deocleciano, Henrique Brazil, Tais Lima, Savio Ferreira, Rhavél Batista, Matheus Morais, Andhré de Souza e Samuel Cesarino.

Agradeço a Universidade Federal de Campina Grande, junto com todos seus funcionários, o que permitiu poder está realizando meu sonho.

Agradeço ao meu orientador Karcus Marcelus, pelo seu trabalho e paciência comigo, além de orientação para que pudesse obter um melhor desempenho na fase final do curso.

## RESUMO

O objetivo deste trabalho é analisar a fraude ou furto de energia elétrica por ação do consumidor relacionando com a violência urbana. Para tanto, serão abordadas técnicas de *Advanced Analytics*; extração, transformação, aplicação das informações em algoritmo para agrupamento ou *clusters* de dados e em seguida aplicar em um modelo de regressão linear. Os dados são disponibilizados pela distribuidora de energia responsável pela concessão do Estado do Pará e pelo Centro Integrado de Operações (CIOP) de Belém – PA. Esses dados são indicadores de aferições de equipes de fiscalização na identificação de fraude ou furto de energia e de ocorrências policiais seguindo as naturezas de crimes pré-definidas. Para compreensão desse tema é desenvolvido uma abordagem de visão macro sobre perda não técnica no país e em seguida uma análise micro sobre a distribuição de energia elétrica no Estado do Pará, mas específico em Belém.

Palavras-chaves: *Advanced Analytics*; *regressão linear*; *clusters*.

## ABSTRACT

The objective of this work is to analyze the fraud or theft of electricity by the consumer related to urban violence. To this end, *Advanced Analytics* techniques will be addressed; creation, transformation and application of information in data *clusters* and then apply in a *linear regression* model. The data are made available by the energy distributor responsible for the concession of the state of Pará and by the Integrated Operations Center (CIOP) of Belém - PA, bases are indicators of inspection measurements in the identification of fraud or theft of energy and of police occurrences following the nature of pre-defined crimes. In order to understand this theme, a macro view approach is developed on non-technical loss in the country and then a micro analysis on the distribution of electric energy in the state of Pará, but specifically in Belém.

Keywords: *Advanced Analytics; Linear regression; clusters.*

## LISTA DE FIGURAS

Figura 1: Notícia de furto de energia na cidade de Nova York em 1886. ....	15
Figura 2: Sede da empresa Equatorial Energia Pará. ....	18
Figura 3: Distribuição da Equatorial Energia Pará por Regionais. ....	19
Figura 4: Participação da energia requerida por regional.....	20
Figura 5: Distribuição de clientes da Equatorial Energia Pará. ....	20
Figura 6: Municípios com maior concentração de clientes no Pará. ....	21
Figura 7: Dados Trimestrais de PNT do mercado BT da Equatorial Energia Pará. ....	22
Figura 8: Medidor quebrado.....	25
Figura 9: Frequência de códigos 115 por bairros de Belém PA.....	25
Figura 10: Frequência de código 154 por bairros de Belém PA. ....	26
Figura 11: Ligação direta na rede de distribuição. ....	27
Figura 12: Frequência de código 165 por bairros de Belém PA .....	27
Figura 13: Equipamentos de blindagem da rede de baixa tensão. 13. a) Capa protetora de conexões. 13. b) Protetor de cabo BT multiplexado. ....	28
Figura 14: Equipamentos de blindagem da rede de baixa tensão. 14. a) Manta asfáltica. 14. b) Tela ciranda. 14. c) Seal tubo. 14. d) Tela ciranda entre duas mantas. ....	29
Figura 15: Exemplo de sistema de medição centralizado ..... 30	30
Figura 16: Detalhes dos equipamentos. 18.a: CP, CS e TLI. 18.b:Modulo de medição E13.....	30
Figura 17: Estrutura de poste e cruzeta. ....	31
Figura 18: Tentativa de ligação clandestina na rede de distribuição.....	31
Figura 19: Fluxo de projeto de previsão de possíveis alvos que estão fraudando ou furtando energia elétrica. ....	32
Figura 20: Etapas do Trabalho ..... 33	33
Figura 21: Comando SQL ..... 34	34
Figura 22: Distribuição de dados preditores..... 39	39
Figura 23: Distribuição de dados target. .... 40	40
Figura 24: Digrama de caixas e outliers..... 40	40
Figura 25: Fluxo no Knime para organizar as informações . .... 42	42
Figura 26: Base de dados para treinar o modelo K-means..... 43	43
Figura 27: Base de dados com bairros agrupados. .... 44	44
Figura 28: Histogramas dos bairros agrupados como grupo 1. .... 45	45
Figura 29: Histogramas dos bairros agrupados como grupo 2. .... 46	46
Figura 30: Histogramas dos bairros agrupados como grupo 3. .... 46	46

Figura 31: Histogramas dos bairros agrupados como grupo 4. ....	47
Figura 32: Normalizando os dados.....	48
Figura 33: Chamando o modelo e treinando. ....	48
Figura 34: Modelo de Regressão Linear do grupo 4. ....	49
Figura 35: Quantidade de ocorrências policial em 2021 .....	51
Figura 36: Quantidade ações contra fraude e furto de energia.....	51



## LISTA DE TABELAS

Tabela 1: Ranking de complexidade das distribuidoras.....	17
Tabela 2: Descrição dos códigos de irregularidade relacionados a fraude e furto. ....	23
Tabela 3: Tabela dinâmica dos dados relacionados a PNT. ....	34
Tabela 4: Tabela com dados relacionados a criminalidade. ....	35
Tabela 5: Dados relacionados a criminalidade. ....	36
Tabela 6: Dados de ocorrências de criminalidade e de código 100. Tabela 6.a: Dados de criminalidade ou ocorrências policiais. Tabela 6.b: Dados de ocorrências de código 100. ....	36
Tabela 7: Dados de ocorrências de criminalidade, de código 100 e percentual de SMC por bairro. Tabela 7.a: Dados de criminalidade ou ocorrências policiais. Tabela 7.b: Dados de ocorrências de código 100. Tabela 7.c: Percentual de SMC por bairro. ....	38
Tabela 8: Dataset finalizado. ....	42
Tabela 9: Coeficientes de correlação. ....	47
Tabela 10: Coeficiente $R^2$ . ....	49

## LISTA DE SIGLAS

ETL	<i>Extract Transform Load</i>
ANEEL	Agência Nacional de Energia Elétrica
ABRADEE	Associação Brasileira de Distribuidores de Energia Elétrica
PNT	Perda Não Técnica
BT	Baixa Tensão
PA	Pará
CIOP	Centro Integrado de Operações
CELPA	Centrais Elétricas do Pará
TC	Transformador de Corrente
TP	Transformador de Potência
INMETRO	Instituto Nacional de Metrologia, Qualidade e Tecnologia
SMC	Sistema de Medição Centralizado
TLI	Terminal de Leitura
CP	Concentrador Primário
CS	Concentrador Secundário
P&D	Pesquisa e Desenvolvimento
CITENEL	Congresso de Inovação Tecnologia em Energia Elétrica
DF	Distrito Federal
SENDI	Seminário Nacional de Distribuição de Energia Elétrica
ELM	<i>Extreme Learning Machine</i>
SMC	Sistema de Medição Centralizado
TLI	Terminal de Leitura
CP	Concentrador Primário
CS	Concentrador Secundário
P&D	Pesquisa e Desenvolvimento
CITENEL	Congresso de Inovação Tecnológica de Energia Elétrica
DF	Distrito Federal
SENDI	Seminário Nacional de Distribuição de Energia Elétrica

## SUMÁRIO

1.	INTRODUÇÃO .....	13
1.1.	OBJETIVOS.....	14
1.1.1.	OBJETIVO GERAL .....	14
1.1.2.	OBJETIVOS ESPECÍFICOS.....	14
1.2.	CONTRIBUIÇÕES.....	14
2.	<b>REVISÃO BIBLIOGRÁFICA</b> .....	15
3.	<b>FUNDAMENTAÇÃO TEÓRICA</b> .....	16
3.1.	PERDAS COMERCIAIS DE ENERGIA ELÉTRICA .....	16
3.1.1.	Perdas Não Técnicas Equatorial Energia Pará.....	18
3.1.2.	Perdas Comerciais por Ação do Consumidor .....	23
3.1.2.1.	Código 115.....	24
3.1.2.2.	Código 154.....	26
3.1.2.3.	Código 165.....	26
3.2.	COMBATE A PNT POR AÇÃO DO CONSUMIDOR .....	28
3.2.1.	Equipamentos de Blindagem da Rede de Distribuição.....	28
3.2.2.	Sistema de Medição Centralizado (SMC).....	29
3.2.3.	Poste e Cruzeta.....	30
3.2.4.	Geração de Alvos.....	32
4.	<b>METODOLOGIA</b> .....	32
4.2.	COMPREENSÃO DOS DADOS .....	33
4.3.	PREPARAÇÃO DE DADOS .....	39
4.4.	MODELAGEM.....	43
4.4.1.	<i>K-means</i> .....	43
4.4.2.	Regressão Linear.....	44
4.5.	AVALIAÇÃO .....	48
5.	<b>CONCLUSÃO</b> .....	50
	REFERÊNCIAS.....	53

ANEXO A - TIPOS DE NATUREZA DE OCORRÊNCIA POLICIAL .....	55
ANEXO B – BAIROS IDENTIFICADOS NAS BASES DA DISTRIBUIDORA DE ENERGIA ELÉTRICA E DO CENTRO INTEGRADO DE OPERAÇÕES .....	58

## 1. INTRODUÇÃO

Segundo (ANEEL,2021), a distribuição de energia elétrica é caracterizada como um segmento do setor elétrico que se dedica ao rebaixamento da tensão proveniente do sistema de transmissão, gerando a interligação de centrais geradoras e o fornecimento de energia elétrica ao consumidor.

No Brasil, segundo a (ABRADEE,2021), o setor de distribuição é composto por 53 concessionárias, que são responsáveis pela operação e distribuição de linhas de transmissão de tensão abaixo de 230 mil Volts.

Um dos grandes desafios das distribuidoras de energia elétrica é a redução de perdas totais, que de acordo com (ANEEL,2021), essas perdas representaram aproximadamente 14,8% do mercado consumidor em 2020, e esse mesmo valor representou o consumo de energia elétrica nas regiões Norte e Centro-Oeste do país em 2018.

As perdas de energia elétrica totais, ou perdas globais, podem ser divididas em dois componentes: técnicos e não técnicos. As perdas técnicas ocorrem naturalmente e consistem principalmente na dissipação de energia em componentes do sistema elétrico, como linhas de transmissão e distribuição, transformadores e sistemas de medição (ANTMANN, 2009). Entretanto, as perdas não técnicas, ou perdas comerciais, são caracterizadas por ações externas ao sistema de energia e consistem principalmente em furto de eletricidade (denominado popularmente “gato”), fraudes, erros de medição, etc.

Dentre as distribuidoras brasileiras, o Grupo Equatorial Energia é atualmente responsável por 9,07% do mercado de baixa tensão, possuindo os contratos de concessão de distribuição de energia nos Estados do Pará, Maranhão, Piauí, Alagoas, Rio Grande do Sul e Amapá

Neste trabalho, é apresentado um método para identificação e correlação de quantidade de PNT no mercado BT (baixa tensão) por índice de criminalidade nos bairros. Sendo assim, para o desenvolvimento do mesmo foi levado em consideração os dados disponibilizados e relacionados à PNT da distribuidora de energia do Estado do Pará e dados de criminalidade dos bairros de Belém-PA cedido pelo CIOP (Centro Integrado de Operações).

## 1.1. OBJETIVOS

### 1.1.1. OBJETIVO GERAL

Este trabalho tem como objetivo desenvolver uma análise para estudo de perdas não técnicas em baixa tensão por ação do consumidor em Belém - PA. Para isso, será utilizado conhecimentos de *Advanced Analytics*, dados relacionados à PNT e informações de criminalidades nos bairros de Belém-PA.

### 1.1.2. OBJETIVOS ESPECÍFICOS

- Observar a correlação entre PNT em baixa tensão por ação do consumidor em bairros com alto índice de violência urbana;
- A abordagem de algoritmo *K-means* para agrupamento dos bairros e identificar os que possuem grande índice de violência urbana;
- A utilização do algoritmo de regressão linear que correlacione perdas não técnicas com índice de criminalidade, além de estimar a quantidade de PNT;

## 1.2. CONTRIBUIÇÕES

O trabalho contribui para elucidar o tema sobre PNT no país e as ações da distribuidora responsável pela concessão do Estado do Pará. Destarte, aborda o assunto desde o combate em campo com equipes de fiscalização, a identificação para distintos tipos de fraude e furto de energia elétrica, equipamentos para blindagem, automatização da rede de distribuição e a utilização de *Advanced Analytics* como ferramenta para orientar na geração de possíveis alvos que estão fraudando ou furtando energia elétrica.

Ademais, contribui também demonstrando algumas técnicas de *Advanced Analytics*, desde extração, transformação e carregamento dos dados.

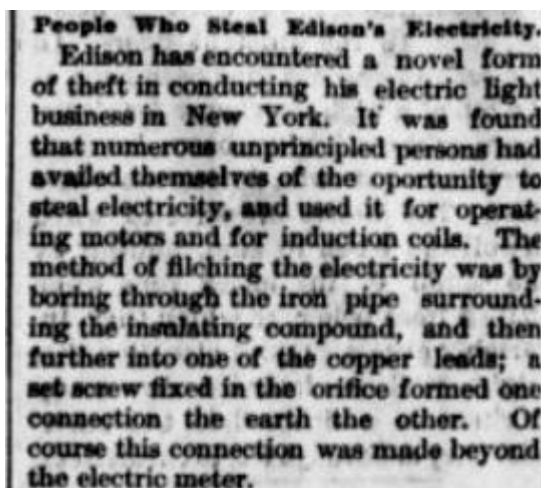
Outrossim, demonstra a correlação entre a violência urbana e perdas não técnicas em baixa tensão. Nesse sentido, a identificação de bairros que possuem grande quantidade de PNT por ação de violência urbana pode ajudar na orientação de outros projetos, como: blindagem da rede de distribuição, orientar as equipes de fiscalização sobre a

periculosidade da área, orientação em projetos que necessitam mapeamento do território, entre outros.

## 2. REVISÃO BIBLIOGRÁFICA

O combate ao furto de energia não é um problema recente e muito menos que ocorre só no Brasil. Podemos evidenciar, na Figura 1 uma notícia de furto de energia em 1886, na cidade de Nova York, contra uma rede de distribuição instalada pela *Edison General Electric Company* para iluminação pública. De acordo com a notícia a solução da empresa foi aplicar pulsos de corrente na rede elétrica, que danificaria as cargas clandestinas e apenas causaria oscilação na iluminação das lâmpadas incandescentes.

Figura 1: Notícia de furto de energia na cidade de Nova York em 1886.



Fonte: Knight e Gordon (1886).

Atualmente essa estratégia seria muito arriscada, uma vez que há vários equipamentos além de lâmpadas incandescentes ligadas à rede de distribuição, que podem ser danificados com uma pequena variação de corrente. Ademais, a ANEEL, é o órgão que fiscaliza e tabela o permitido de variação da energia fornecida pelas distribuidoras, afim de evitar quaisquer irregularidades e garantir a proteção efetiva dos equipamentos dos usuários da rede de distribuição.

Além disso, segundo Barros (2021), no final do século XX foi publicado o primeiro trabalho de destaque que trata de problemática da perda não técnica de energia elétrica, no qual é relatado por Dick (1995), o caso de furto de energia elétrica no Reino Unido, contra as ações realizadas pela companhia de distribuição, gerando um prejuízo de £50 milhões por ano à empresa. Nesse sentido, foi demonstrado que a principal forma

da fraude era a interferência no medidor de energia que com o simples ato de inclina-lo poderia alterar o valor de energia medida. Esse tipo de fraude ainda ocorre nos dias atuais, sendo uma das formas de solucionar o problema a realização de troca dos equipamentos por uma versão mais eficiente no combate de tentativa de furto ou fraude de energia; o que foi realizado pela Distribuidora do Reino Unido.

Outrossim, de acordo com Ramos (2014), algoritmos híbridos foram desenvolvidos baseados em técnicas evolutivas com objetivo de realizar a seleção de características no âmbito da caracterização de perdas não técnicas, verificando a taxa de acerto e as características selecionadas.

Ademais, segundo Filho, Galas e Junior (2018), utilizaram técnicas de ciência de dados como árvore de decisão, classificador de rede neural, *random forest* e *gradiente boosting*; na concessão do Ceará pela distribuidora Enel Brasil. Obtiveram uma recuperação de energia de 87,3 GWh faturados.

De acordo Alves (2019), foi analisado a contribuição de variáveis do histórico de consumo de energia elétrica para identificação de fraude de energia elétrica pelo o usuário. As variáveis foram desenvolvidas parti das séries de consumo mensal, levando em consideração informações como: amplitude, sazonalidade e frequência. Portanto, esse trabalho demonstra que com a criação de novas variáveis a partir das variáveis originais, possibilita a melhora no modelo preditivo.

A partir dessa linha de raciocínio, observa-se que com a grande quantidade de clientes e, portanto, de informações, vem tendo destaque a área de ciência de dados, no que concerne à expansão de algoritmos de aprendizado supervisionado ou não supervisionado, inteligência artificial, etc. Dessa maneira, a partir do desenvolvimento e treinamento de modelos para previsão de PNT, as distribuidoras conseguem gerir seus recursos de forma eficiente e obter resultados positivos.

### **3. FUNDAMENTAÇÃO TEÓRICA**

#### **3.1. PERDAS COMERCIAIS DE ENERGIA ELÉTRICA**

As perdas não técnicas ou comerciais são apuradas a partir da diferença entre as perdas totais e as perdas técnicas; têm como alguns tipo de causas:

- 1) Fraude de energia onde utiliza-se de artifícios ilícitos para redução de energia elétrica consumida pela instalação, através da adulteração dos



equipamentos de medição ou da utilização de condutores derivativos para captação de energia elétrica antes da passagem pelo medidor.

- 2) Ligação clandestina é um tipo de furto, no qual há ligação à rede elétrica existente, porém o usuário não possui medidor ou registro cadastral.
- 3) Gambiarra é um furto no qual há ligação, mas o usuário criou uma estrutura com postes e cabos para alcançar a rede elétrica da distribuidora. O usuário não possui registro cadastral e nem medidor.
- 4) Erros de medição, leitura e faturamento.
- 5) Defeito técnicos nos equipamentos de medição.

Os tópicos 1, 2 e 3 estão relacionados a furto de energia elétrica e se caracterizam como crime, com pena de reclusão de 1 a 8 anos e multa, segundo os ditames dos §3º e §4º do Artigo 155 do Código Penal Brasileiro. Ademais, os tópicos 4 e 5 estão relacionados à problemas de gestão da distribuidora e podem ser denominados como perda administrativa.

Segundo a (ANEEL, 2021), as perdas comerciais realizada pela ação do consumidor, mais especificamente no mercado de baixa tensão, representa um grande impacto financeiro para as distribuidoras, uma vez que, o nível de perdas não técnicas possui influência da gestão da concessionária (investimentos em fiscalização, equipamentos de monitoramento e blindagem mais sofisticados), das características socioeconômicas e de aspectos culturais existentes nas áreas de concessão.

Como as distribuidoras atuam em áreas de concessão heterogêneas, a comparação entre elas leva em consideração um *ranking* de complexidade socioeconômica, que é calculado considerando de um modelo econométrico (modelo de análise de regressão que utiliza variáveis como: violência, desigualdade, precariedade e infraestrutura).

Tabela 1: Ranking de complexidade das distribuidoras.

<b>DISTRIBUIDORAS GRANDES (GRUPO 1)</b>	<b>ÍNDICE DE COMPLEXIDADE</b>
CELPA	0,503
LIGHT	0,377
AMAZONAS ENERGIA	0,364
CEMAR	0,315
CELPE	0,313
COELBA	0,284

CEAL	0,266
ELETROPAULO	0,265
CEPISA	0,257
COELCE	0,253
ELETROACRE	0,243
ESCELSA	0,235
ESSE	0,224
AMPLA	0,218
EPB	0,197
CERON	0,191
CEEE – D	0,179
COSERN	0,177
BANDEIRANTE	0,172
CPFL PIRATININGA	0,170
CEB	0,166
CEMIG	0,147
CELTINS	0,139
CEMAT	0,122
ELKTRO	0,106
COPEL	0,105
RGE	0,092
EMG	0,091
AES SUL	0,086
CPFL – PAULISTA	0,080
CELESC – D	0,077
CELG – D	0,075
ENERSUL	0,063

Fonte: ANEEL 2021

O Estado que possui a maior complexidade socioeconômica é o Estado do Pará, com índice de 0,503, demonstrado na Tabela 1, no qual tem como distribuidora responsável a CELPA (Centrais Elétricas do Pará), que atualmente tem o nome de Equatorial Energia Pará.

### 3.1.1. Perdas Não Técnicas Equatorial Energia Pará

A única empresa responsável pela distribuição de energia elétrica em todo território do Estado do Pará pela ANEEL é a Equatorial Energia Pará (Figura 2), administrada pelo Grupo Equatorial Energia desde 3 de novembro de 2012.

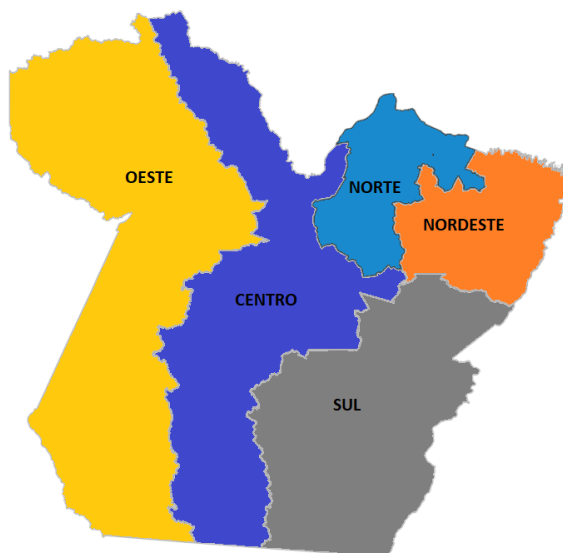
Figura 2: Sede da empresa Equatorial Energia Pará.



Fonte: Site Equatorial Energia (2022).

A distribuidora para atender o território do Pará de 1.248.000 km<sup>2</sup>, com 144 municípios, equivalente a 2,6 milhões de clientes, dividiu o Estado em cinco Regionais (Figura 3), e tendo como sedes as cidades Belém, Castanhal, Marabá, Santarém e Altamira.

Figura 3: Distribuição da Equatorial Energia Pará por Regionais.

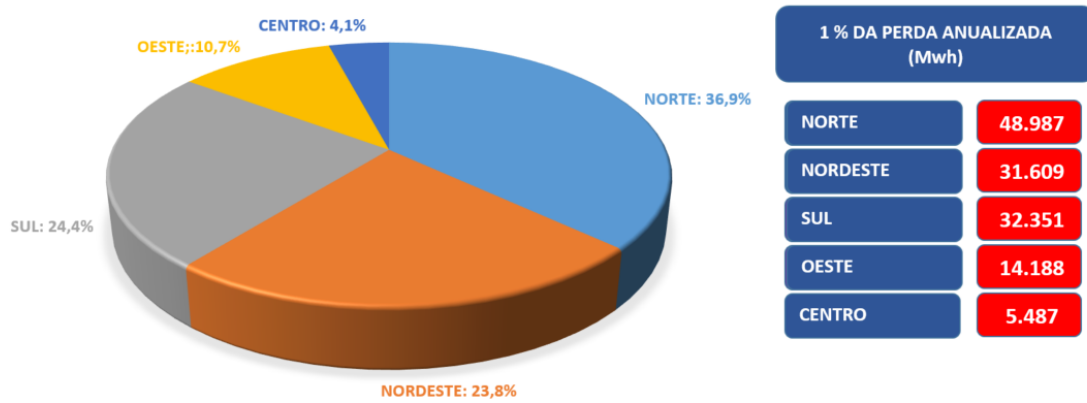


Fonte: Autoria própria.

Entretanto, o tamanho territorial das regionais não necessariamente está diretamente ligado ao consumo de energia e pode-se constatar pela Figura 4. Uma vez que, demonstra a quantidade de energia requerida por regional em um período de doze meses entre dezembro de 2020 até novembro de 2021. Essa análise ajuda a identificar as

regiões que possuem mais contribuição no consumo de energia elétrica, ou seja, possuem maior densidade demográfica e portando muitas das vezes uma maior PNT.

Figura 4: Participação da energia requerida por regional.



Fonte: Equatorial Energia Pará (2022)

Portanto, parti dos dados da Figura 5 pode-se perceber que as três regionais que possuem maior consumo de energia são: Norte, Sul e Nordeste. Isso posto, podemos confirmar essa informação a partir da distribuição de clientes pelo Estado do Pará (Figura 7), onde é exposto a subdivisão das regionais e os dos pontos vermelhos representam os usuários.

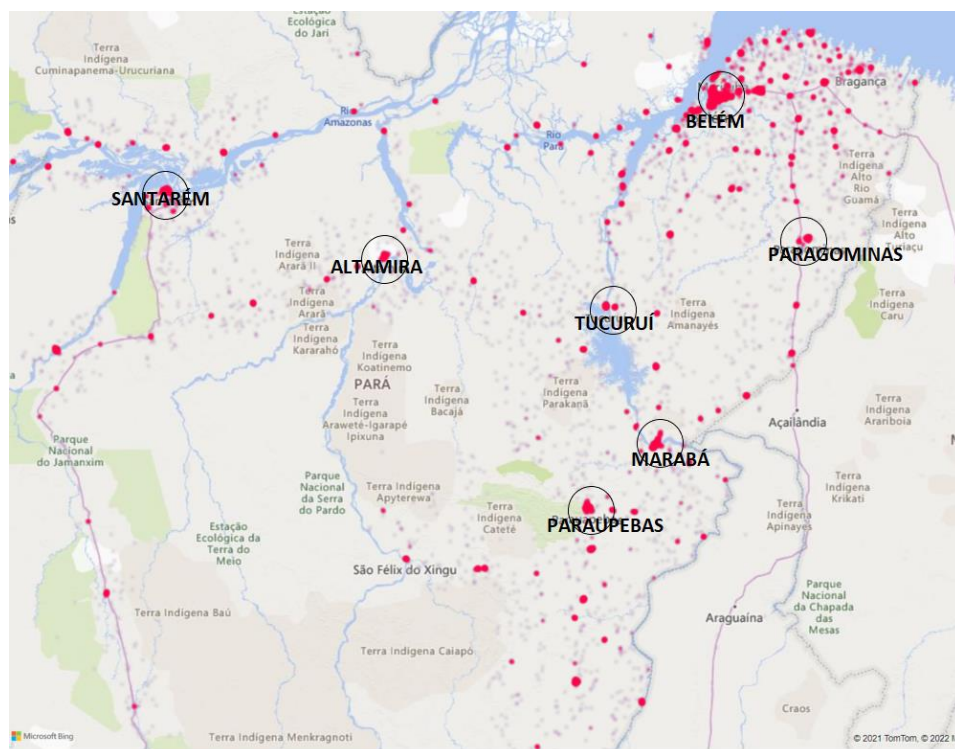
Figura 5: Distribuição de clientes da Equatorial Energia Pará.



Fonte: Autoria própria.

Contudo, a maior densidade de clientes se localiza na região metropolitana de Belém demonstrado na Figura 6.

Figura 6: Municípios com maior concentração de clientes no Pará.

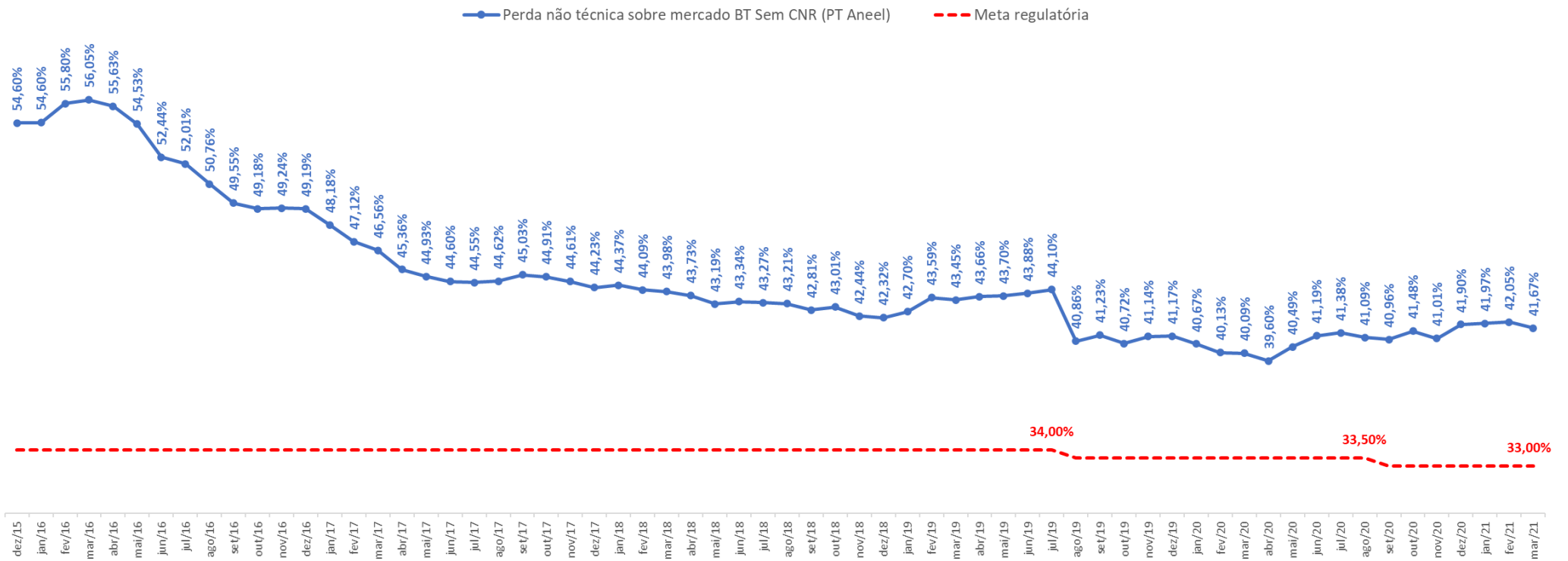


Fonte: Autoria própria.

Nesse cenário, é importante destacar o combate na redução de PNT empregado pela Equatorial Energia, desde 2012, com objetivo de alcançar a meta regularia definida pela ANEEL. Sendo que, a meta regulatória é um valor que a concessionaria define como aceitável de perda de energia elétrica e que muitas das vezes é menor que a perda real da distribuidora, como demonstrado na Figura 7.

. Destarte, em novembro de 2012, foi quando a Equatorial Energia conseguiu a concessão para distribuir energia no Estado do Pará, de tal forma que a porcentagem de PNT era de 57,06%, da energia injetado na rede de distribuição. Logo, com investimento em gestão de combate a perdas comerciais ao longo dos anos esse indicador foi reduzindo e em março de 2021 a porcentagem era de 41,67% como demonstrado na Figura 7.

Figura 7: Dados Trimestrais de PNT do mercado BT da Equatorial Energia Pará.



Fonte: Equatorial Energia Pará (2022).

### 3.1.2. Perdas Comerciais por Ação do Consumidor

O combate a perdas comerciais ou perdas não técnicas por ação do consumidor é um dos desafios que toda distribuidora brasileira enfrenta, problema esse complexo e difícil de ser combatido. Dentro aos tipos de fraudes e furtos por ação do consumidor o Grupo Equatorial Energia identifica essas ações com uma descrição simplificada, subdividida por códigos e tipo, onde é demonstrado na Tabela 2.

Tabela 2: Descrição dos códigos de irregularidade relacionados a fraude e furto.

DESCRIÇÃO CÓDIGOS DE IRREGULARIDADES		
TIPO	CÓDIGO	DESCRIÇÃO
CÓDIGO_100	109	CIRCUITO POTENCIAL INTERROMPIDO
	115	MEDIDOR AVARIADO
	129	MEDIDOR SEM ALIMENTAÇÃO/NEUTRO
	131	IRREGULARIDADE NO TC
	132	TC/TP COM LIGAÇÃO INVERTIDA
	136	CIRCUITO TC/TP INTERROMPIDO
	143	IRREGULARIDADE NO TP
	146	TP QUEIMADO
	154	LIGAÇÃO CLANDESTINA
	156	IRREGULARIDADE NA CHAVE DE AFERIÇÃO
	164	DESVIO EMBUTIDO
	165	DESVIO ANTES DA MEDIÇÃO
	168	PONTE NO BLOCO DE TERMINAIS
	171	LIGAÇÃO A REVELIA COM CONTRATO ATIVO
	172	LIGAÇÃO INVERTIDA
	174	MEDIDOR INCLINADO/DEITADO

Autor: Grupo Equatorial Energia (2022).

Portanto, a natureza de cada tipo de irregularidade do tipo código 100 e sua frequência vai variar de região para região, e essa variância de acordo com (ANEEL,2021), depende de certos indicadores:

- Renda da população;
- Nível de escolaridade;
- Complexidade socioeconômica da região;
- Valor da tarifa de energia elétrica em relação ao salário médio;
- Gestão da concessionária;
- Cultura de aceitação da fraude.

Levando em consideração que as informações do tipo código 100 demonstrada na Tabela 2, foi disponibilizada pela distribuidora do Estado do Pará, da cidade de Belém - PA, mais especificamente dos bairros e entre os anos de 2011 e 2016. Então pode-se identificar a frequência de ocorrências do tipo código 100 que foram executadas pelas equipes de fiscalização.

Sendo que, o código 100 contem subcódigos que podem ser identificados na Tabela 2, e que possuem diferenças relacionadas ao tipo de crime (furto ou fraude) ou ao tipo de cliente: ligado, desligado e cortado. Visto que o cliente ligado é aquele que possui vínculo contratual com a distribuidora. Contudo, no caso de cliente desligado é aquele que não possui nenhum vínculo contratual com a distribuidora, sendo por que pediu desligamento ou simplesmente possui um imóvel novo e não realizou o contrato de fornecimento de energia. Logo, o cliente cortado possui vínculo contratual com a empresa, mas por causa de atraso de pagamento ou furto de energia a distribuidora cancela o fornecimento de energia elétrica, até que o cliente normalize sua situação.

Posteriormente será abordado alguns tipos de códigos, levando em consideração a descrição e a quantidade de execução.

#### 3.1.2.1. Código 115

O código 115 representa casos de medidores avariados com intervenção interna, deixando de registrar corretamente a energia elétrica consumida. O tipo de avaria pode ser:

- Circuito interno de potência interrompido;
- Relação da registradora trocada;
- Mancal desregulado ou solto;
- Disco parado, empenado ou travado (preso);
- Redução dos dentes da engrenagem;
- Engrenagem danificada ou trocada;
- Bloqueio do registrador;
- Bloqueio da engrenagem;
- Registrador desacoplado ou danificado;
- Registrador trocado;
- Bulbo furado;



- Selo violado;
- Medidor quebrado/danificado.

Portanto, uma forma de visualizar esse tipo de fraude é a partir da Figura 8, onde visivelmente o medidor se encontra danificado. Além do que, pode-se classificar o código 115 como de natureza técnica, pois em alguns casos como de manipulação interno do equipamento o responsável deve ter um conhecimento do funcionamento, para que seja possível realizar a fraude.

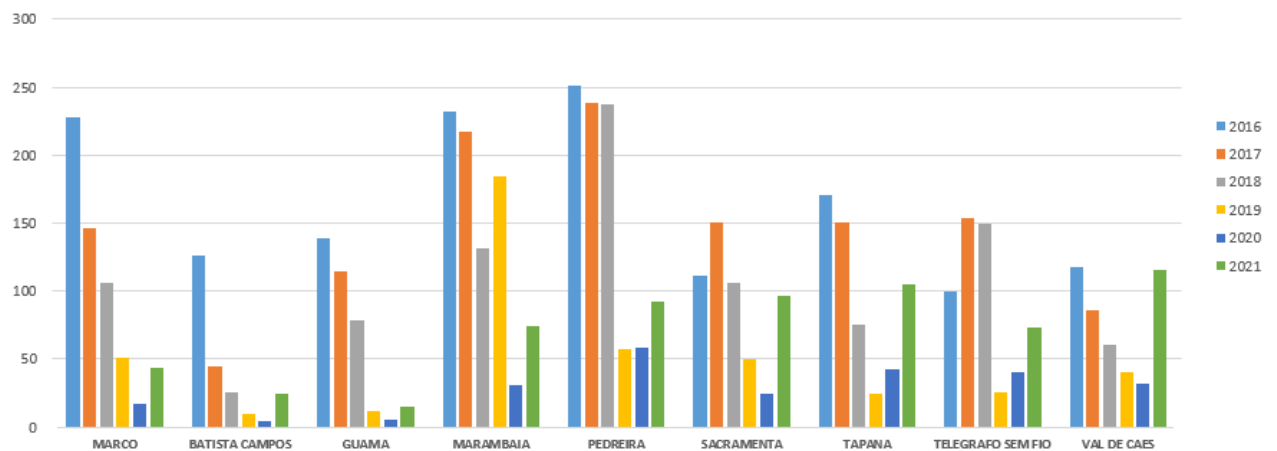
Figura 8: Medidor quebrado.



Fonte: Equatorial Energia Pará (2022)

Levando em consideração a Figura 9, é possível realizar uma análise quantitativa dos casos de código 115, no qual é demonstrado alguns bairros de Belém - PA com maior incidência e entre os anos de 2016 e 2021.

Figura 9: Frequência de códigos 115 por bairros de Belém PA.



Fonte: Autoria própria.

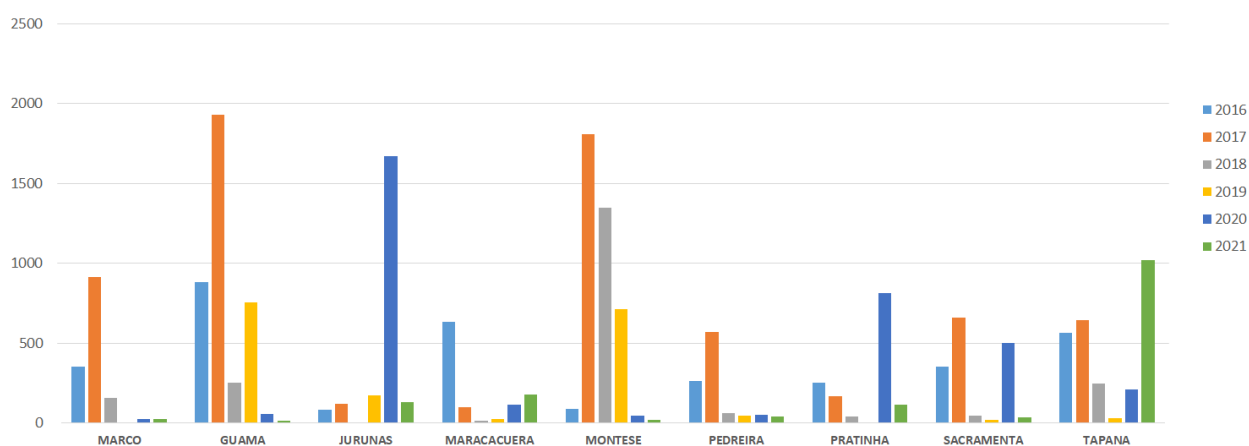
Além disso, casos de fraude como o código 115, possuem um processo no qual a equipe de fiscalização deve identificar o mal funcionamento do medidor ou qualquer aparência fora do padrão do equipamento, em seguida recolher da casa do cliente e enviar para o INMETRO (Instituto Nacional de Metrologia, Qualidade e Tecnologia), para que assim seja realizado uma análise do equipamento e por fim determinar se o mesmo sofreu alguma alteração indevida.

### 3.1.2.2. Código 154

O código 154 representa ligações a rede de distribuição de forma clandestina. São casos de furto de energia elétrica, uma vez que o usuário não possui histórico de vínculo contratual com a distribuidora e realiza uma ligação direta à rede, ou seja, à revelia da Equatorial Energia, deixando de registrar a energia elétrica consumida.

Sendo assim, na Figura 10 é possível identificar uma grande quantidade desse tipo de ocorrência ao longo dos anos, variando entre os bairros e tendo como destaque o bairro do Tapanã que em 2021 teve mais de 1000 ocorrências.

Figura 10: Frequência de código 154 por bairros de Belém PA.



Fonte: Autoria própria.

### 3.1.2.3. Código 165

O código 165 representa casos de desvio antes da medição, mais conhecido popularmente como “gato”, que é visivelmente constado na Figura 11, no qual há vários cabos elétricos ligados diretamente nas fazes da rede de distribuição da concessionária. Esse caso ocorre quando há uma derivação antes da medição, tais como: saindo do borne

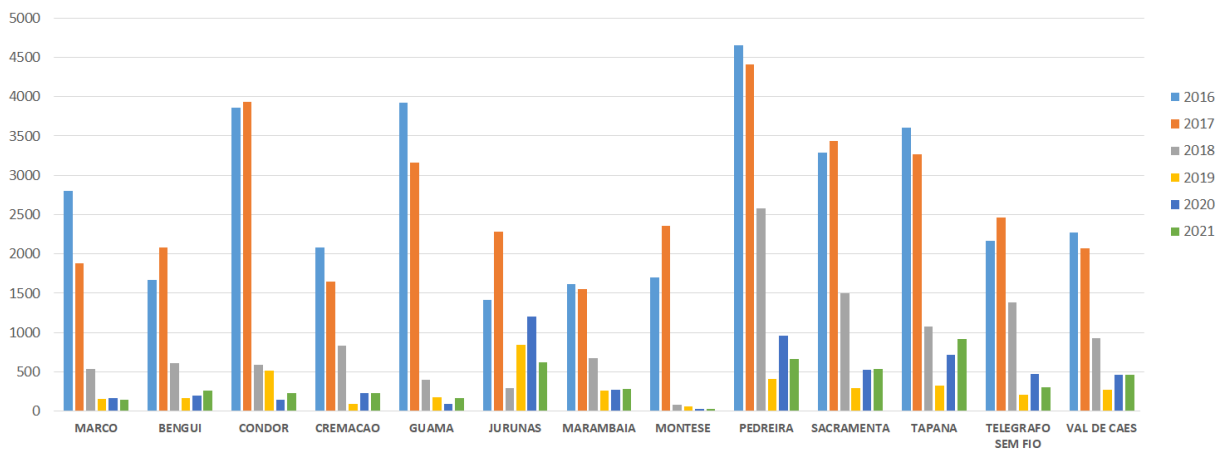
de linha, direto do poste, beiral, barramento, bucha do transformador, transformador ligado direto e nos casos de medidor duplê. Desde de 2016 é a ocorrência que é mais encontrada e na Figura 12 é possível ver a alta incidência ao longo dos anos.

Figura 11: Ligação direta na rede de distribuição.



Fonte: Equatorial Energia Pará (2022).

Figura 12: Frequência de código 165 por bairros de Belém PA



Fonte: Autoria própria.

### 3.2. COMBATE A PNT POR AÇÃO DO CONSUMIDOR

Toda distribuidora brasileira possui uma gestão de combate às perdas comerciais, com objetivo de reduzir a porcentagem de PNT real, tentando muitas das vezes alcançar a meta regulatória determinada pela ANEEL, que normalmente é um valor inferior ao determinada pela concessionária. Ademais, a meta regulatória é um meio que a ANEEL utiliza para estimular as distribuidoras a investir no combate de PNT, uma vez que os valores regulatórios serão os reconhecidos nas tarifas de energia e a diferença entre o custo do valor regulatório e o real é de responsabilidade da distribuidora, portanto, as distribuidoras tentam reduzir ao máximo esse custo.

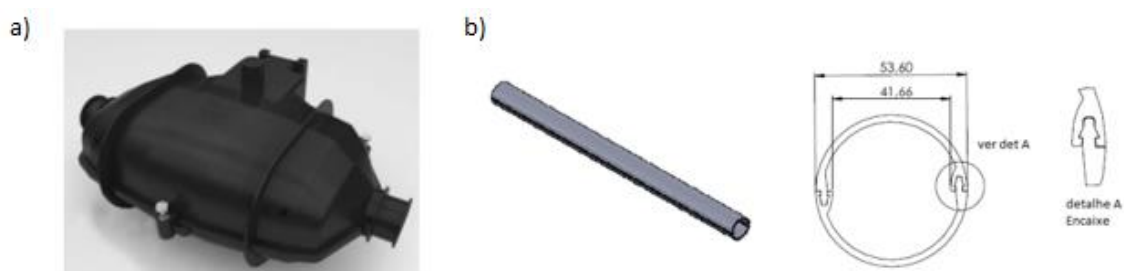
Portanto, várias são as ações utilizadas pelas distribuidoras para reduzir o índice de perda não técnica, tais como: campanhas de conscientização, instalação de equipamentos de blindagem da rede de distribuição, equipes de fiscalização e projetos mais sofisticados que muitas das vezes age no sentido de gerenciar equipes de combate a PNT.

Na seção seguinte será abordado algumas técnicas e equipamentos utilizados para o combate de PNT.

#### 3.2.1. Equipamentos de Blindagem da Rede de Distribuição

Os equipamentos para blindagem da rede de distribuição atuam de forma a prevenir e/ou dificultar possíveis tentativa de furto ou fraude de energia elétrica. Esses equipamentos podem variar entre: capa protetora de conexões (Figura 13. a), protetor para cabo de baixa tensão multiplexado (Figura 13. b), manta asfáltica (Figura 13. a), seal tubo (Figura 13. c), tela cirando (Figura 13. b), etc.

Figura 13: Equipamentos de blindagem da rede de baixa tensão. 13. a) Capa protetora de conexões. 13. b) Protetor de cabo BT multiplexado.



Fonte: Grupo Equatorial Energia.

Figura 14: Equipamentos de blindagem da rede de baixa tensão. 14. a) Manta asfáltica. 14. b) Tela ciranda. 14. c) Seal tubo. 14. d) Tela ciranda entre duas mantas.

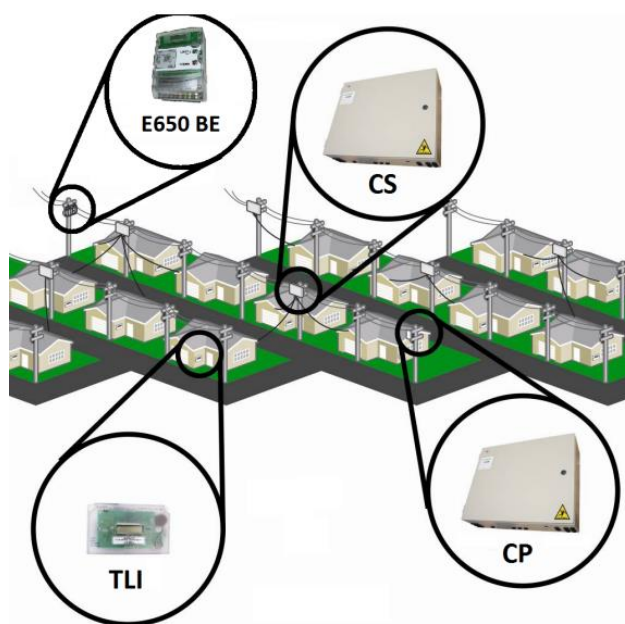


Fonte: Grupo Equatorial Energia.

### 3.2.2. Sistema de Medição Centralizado (SMC)

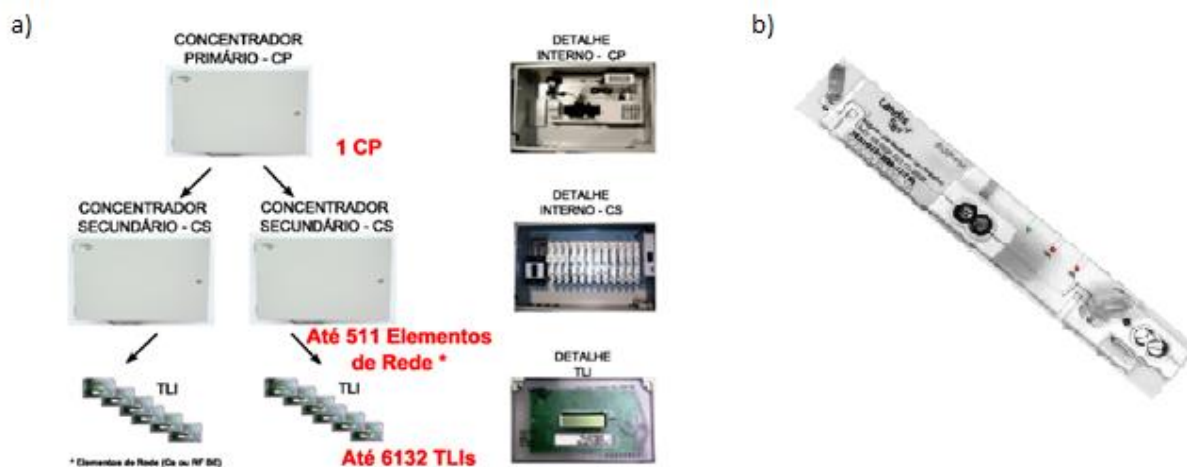
O SMC é um sistema de gerenciamento de perdas e medição, com funcionalidade de corte e religamento de usuário, alarmes, leitura e comandos remotos. Ele possui a composição dos equipamentos, no qual seguem um fluxo de funcionamento, onde em cada residência existe um terminal de leitura (TLI), que recebe informações de módulos de medição (E13) concentrados num equipamento chamado de concentrado secundário (CS). Diversos CS se comunicam com um concentrador Primário (CP), o qual tem a responsabilidade pela troca de dados e comunicação entre o centro de operação da distribuidora e os equipamentos instalados; essa comunicação é feita via a rádio ou GPRS. Outro equipamento monitorado pelo CP é o medidor E60 BE, que realiza a medição nos transformadores de distribuição, permitindo estimar a perda técnica na rede. Todo a estrutura desse sistema e o fluxo de funcionamento é demonstrado nas Figuras 15 e 16.

Figura 15: Exemplo de sistema de medição centralizado



Fonte: Manual da Landis + Gyr sobre Sistema de Medição Centralizada - AMI

Figura 16: Detalhes dos equipamentos. 18.a: CP, CS e TLI. 18.b:Modulo de medição E13.



Fonte: Manual da Landis + Gyr sobre Sistema de Medição Centralizada – AMI

### 3.2.3. Poste e Cruzeta

O poste e a cruzeta são estruturas para dificultar o acesso do fraudador na linha de baixa tensão e no SMC no sentido de o indivíduo tentar escalar. O poste de distribuição em baixa tensão pode variar seu comprimento de 9 a 13 m e a cruzeta tem com intuito afastar em sentido vertical o SMC da linha do poste, demonstrado na Figura 17.

Figura 17: Estrutura de poste e cruzeta.



Fonte: Autoria própria.

Entretanto, mesmo com o poste tendo em sua estrutura linhas de telecomunicação, poste de iluminação, rede de alta tensão e em outros casos até transformadores; ainda existe tentativas de fraudes como representado na Figura 18.

Figura 18: Tentativa de ligação clandestina na rede de distribuição.



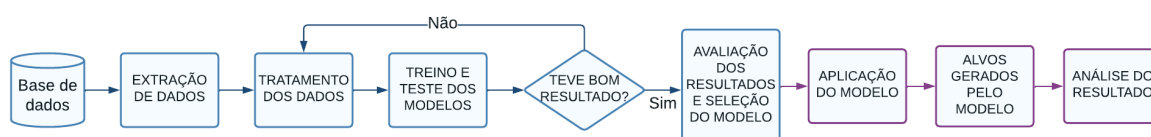
Fonte: extra.globo.com (2022).

### 3.2.4. Geração de Alvos

A ideia de combater a PNT prevendo possíveis alvos que estão fraudando ou furtando energia elétrica, tem como base utilizar dados dos clientes, variando desde informações como: característica de consumo, histórico de fraudes, informações de equipamentos técnicos instalado para fornecer e medir energia elétrica, etc.

O fluxo básico de funcionamento de um projeto de previsão de possíveis alvos que fraudam ou furtam energia elétrica pode ser identificado na Figura 19. Destarte, o funcionamento do projeto se inicia extraíndo as informações do banco de dados levando em consideração a importância das informações e compreensão do negócio. Em seguida, é realizado um tratamento desses dados com intuito de normaliza-los ou padroniza-los, para posteriormente aplicá-los em modelos de *Machine Learning*. Após verificar o desempenho dos modelos é selecionado o modelo ou os modelos que tiveram melhor resultados para posteriormente gerar alvos. A seguir, as coordenadas dos alvos são enviadas para os funcionários de campo que irão realizar as fiscalizações e por fim é verificado a assertividade do modelo.

Figura 19: Fluxo de projeto de previsão de possíveis alvos que estão fraudando ou furtando energia elétrica.



Fonte: Autoria própria.

Portanto, cada distribuidora possui uma estratégia de como aplicar o conceito de geração de alvos para tentar alcançar a maior assertividade, podendo variar na seleção de modelos, a tipo de dados que devem ser extraídos, tipo de clientes fraudadores e entre outros.

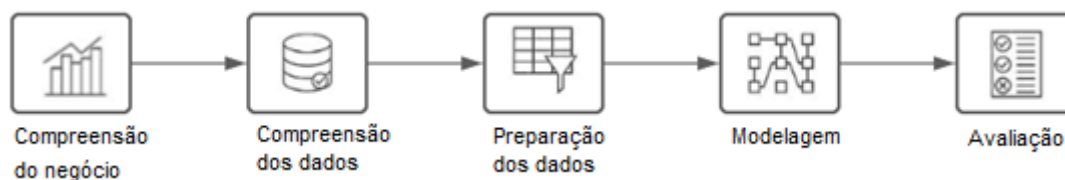
## 4. METODOLOGIA

Para poder alcançar o objetivo geral do trabalho, que é desenvolver um modelo que identifique a relação entre PNT e índice de criminalidade e violência nos bairros de Belém – PA, é aplicado a metodologia da Figura 20, na qual em cada etapa utilizamos



ferramentas para auxiliar na análise e desenvolvimento do projeto, tais como: Oracle SQL, Excel, Knime e Spyder.

Figura 20: Etapas do Trabalho



Fonte: Autoria própria

Nas seções seguintes serão abordados cada uma das etapas apresentadas na Figura 20, explanando cada uma delas e como esses conceitos foram utilizados.

#### 4.1. COMPREENSÃO DO NEGÓCIO

Os dados para realizar o trabalho vem de duas entidades distintas, a primeira da distribuidora de energia elétrica do Estado do Pará e a segunda do Centro Integrado de Operações de Belém do Pará.

Portanto, as informações da distribuidora são ações que foram executadas por equipes de fiscalizações no qual foi identificado algum tipo de fraude ou furto de energia elétrica. Em seguida, esse tipo de ocorrência é classificado de acordo com um código que faz parte de um conjunto denominado código 100. Além disso, a base de informação disponibilizada é de 2016 até 2021.

Consequentemente, os dados do Centro Integrado de Operações são ocorrências policiais que foram realizadas nos bairros, essas ocorrências variam de acordo com sua natureza, desde assalto até homicídio. Entretanto, os dados disponibilizados são apenas de 2021.

Outrossim, como os dados do Centro Integrado de Operações são apenas do ano de 2021, houve a necessidade de usar os dados de caso de código 100 apenas de 2021.

Na seção seguinte será abordado mais detalhado as informações de cada organização.

#### 4.2. COMPREENSÃO DOS DADOS

As informações da distribuidora de energia elétrica foram baixadas via comando SQL demonstrado na Figura 21. Entretanto, as únicas ocorrências que serão contabilizadas

para análise, são as que foram finalizadas e executadas, ou seja, comando STATUS\_CCS = 'FINL' e F\_FISC\_EXECUTADA = 1.

Figura 21: Comando SQL

```

SELECT CAD.MUNICIPIO,
       CAD.BAIRRO,
       CAD.COD_IRREG,
       TRUNC(A.DTA_ORIGEM_NS,'MM') AS CMPT_ORIGEM,
       A.COD_IRREG,
       COUNT(A.NOTIFICATN) AS QTD_FS,
       SUM(NVL(A.COD_100,0)) AS QTD_COD_100,
       SUM(NVL(A.COD_200,0)) AS QTD_COD_200,
       SUM(NVL(A.COD_310,0)) AS QTD_COD_310,
       SUM(NVL(A.TROCA_MD,0)) AS QTD_TROCA_MD,
       SUM(A.NORMALIZACAO) AS QTD_NORMALIZACAO
FROM OWRCPLP.GP_FISCALIZACOES A
LEFT JOIN CELPA.TAB_CADASTRO CAD ON A.INSTALACAO = CAD.INSTALACAO
WHERE 1=1
AND CAD.MUNICIPIO = 'BELEM'
AND A.STATUS_CCS = 'FINL'
AND A.F_FISC_EXECUTADA = 1

GROUP BY CAD.MUNICIPIO,
         CAD.BAIRRO,
         TRUNC(A.DTA_ORIGEM_NS,'MM'),
         A.COD_IRREG
    
```

Fonte: Autoria própria

Dentre os dados extraídos via Excel, é realizado uma filtragem demonstrado na Tabela 3, com apenas as informações necessárias: bairros, quantidade de ocorrência de código 100 (QTD\_COD\_100) e o mês da ação (CMPT\_ORIGEM).

Tabela 3: Tabela dinâmica dos dados relacionados a PNT.

Rótulos de Linha	2021												Total Geral
	Trim1			Trim2			Trim3			Trim4			
	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez	
MARCO	30	24	22	34	28	22	15	31	32	23	33	14	308
AEROPORTO	3	4	6	2	5	4	8	11	3	9	5	2	62
AGUA BOA	32	25	16	12	27	68	36	55	39	32	17	55	414
AGUA CRISTALINA	0				1	2		0	0	0		0	3
AGUAS LINDAS	11	3	19	23	0	15	7	4	4	1	1	2	90
AGUAS NEGRAS	27	23	7	3	27	14	4	26	24	12	8	11	186
AGULHA	24	36	6	6	24	10	22	32	34	32	12	10	248
ARIRAMBA	0	4	2	8	8	11	17	14	9	18	11	9	111
ATALAIA		0	0	0		0		1	1		0	0	2
BAIA DO SOL	0	6	6	3	7	7	7	7	4	3	5	7	62
BARREIRO	5	7	2	19	6	2	8	5	4	5	3	5	71
BARRO BRANCO						1						0	1
BATISTA CAMPOS	15	6	20	4	14	7	14	6	24	5	12	11	138
BENGUI	121	49	54	61	18	38	16	11	23	42	39	34	506
BRASILIA	24	13	8	13	7	12	5	12	6	22	27	6	155
CABANAGEM	64	20	15	34	69	74	77	67	58	44	35	126	683
CAMPINA	0	2	7	8	8	11	5	10	9	6	2	4	72
CAMPINA DE ICOARACI	19	73	30	14	67	47	41	28	34	17	40	35	445
CANUDOS	16	5	0	2	20	2	2	6	5	4	1	6	69
CARANANDUBA	7	8	18	17	7	22	31	30	10	15	14	29	208
CARMELENDIA						0							0

Autoria: Fonte própria.

Consequentemente, na Tabela 3, é possível verificar a distribuição de casos de código 100 por bairros ao longo dos meses de 2021. Além disso, pode-se identificar situações que em certos bairros existem meses que não há nenhuma ocorrência de código 100, sendo representadas na Tabela 3 como valor 0 ou sem nenhum valor, como é o caso do bairro de Água Cristalina no mês de janeiro e fevereiro.

Para conseguir as informações relacionadas a criminalidade foi necessário o preenchimento de uma tabela identificada no Anexo A, que possui as naturezas das ocorrências policiais. Em seguida, foi disponibilizado pelo CIOP uma planilha em Excel, demonstrado na Tabela 4, que possui a quantidade de ocorrências policiais que foram realizadas em um determinado período levando em consideração apenas as naturezas dos crimes solicitados, os bairros de Belém – PA e no decorrer dos meses de 2021.

Tabela 4: Tabela com dados relacionados a criminalidade.

01 - Quantitativo de ocorrências registradas por bairro, levando em consideração as naturezas solicitadas para análise. Registro dividido por meses de janeiro a dezembro no ano de 2021.	
BAIRRO	QUANTIDADE
janeiro	6946
localidades sem registro no sistema	299
40 Horas	14
Aeroporto	4
Água Boa	28
Água Boa (Outeiro)	1
Água Boasao João do Outeiro (Outeiro)	1
Águas Brancas	89
Águas Lindas	123
Águas Negras	33
Agulha	38
Almir Gabriel	21
Arirambá	9
Ariramba (Mosqueiro)	4
Atalaia	98
Aura	9
Aurá	3
Baía do Sol	4
Baía do Sol	10
Baía do Sol (Mosqueiro)	7
Bairro Centro	6

Autoria: Fonte própria.

Portanto, para tornar a tabela mais visível foi reformulada sua estrutura para que seja possível ver todas as ocorrências nos bairros ao longo dos meses, havendo ou não ação policial, demonstrado na Tabela 5.

Tabela 5: Dados relacionados a criminalidade.

BAIROS	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
40Horas	14	13	19	26	29	12	14	14	5	15	26	17
Aeroporto	4	6	6	10	14	7	7	11	4	19	10	13
Água Boa	28	35	38	39	37	33	56	31	35	39	34	33
Águas Lindas	123	124	130	116	115	100	135	121	117	110	124	163
Águas Negras	33	22	19	22	25	13	22	28	18	17	15	20
Agulha	38	36	56	53	33	54	53	71	98	71	44	41
Arirambá	13	20	16	22	17	12	24	18	15	27	13	24
Atalaia	98	72	59	96	76	59	75	67	72	81	77	112
Baía do Sol	21	8	13	13	11	7	3	5	9	4	6	8
Barreiro	39	36	68	57	39	29	47	39	49	53	59	41
Batista Campos	70	64	45	43	54	52	67	83	94	112	86	67
Bengui	93	62	79	107	104	85	82	107	103	102	80	106
Bonfim	7	2	2	2	0	2	1	3	4	4	5	4
Brasília	15	17	16	22	27	23	18	16	29	23	22	26
Cabanagem	70	42	47	64	59	51	59	52	68	80	68	65
Campina	87	52	84	60	71	68	87	69	89	120	96	121
Campina de Icoaraci	82	67	81	77	64	47	84	79	72	79	97	68
Canudos	55	42	55	41	51	56	39	51	60	53	57	45
Carananduba	137	109	136	118	115	103	123	130	132	132	154	113
Caruara	1	1	4	2	0	0	9	0	2	0	0	1

Autoria: Fonte própria.

Após os dados de casos de código 100 e ocorrências policiais serem estruturados no mesmo formato, como demonstrado nas Tabelas 3 e 5, foi desenvolvido um fluxo para identificar os bairros existentes em ambas as bases de dados, uma vez que por serem instituições com áreas de atuações distintas possuem nomenclaturas diferentes para alguns bairros. Com isso, foi possível identificar 74 bairros existentes.

Ademais, após identificar os bairros existentes nas tabelas de casos de código 100 e ocorrência policial, foi possível identificar os bairros que não agregam informação, ou seja, que a quantidade de ocorrência de código 100 e ocorrência policial é ínfima ou nula. Logo, esses bairros são caracterizados como *outliers* e por possuírem muitos valores nulos são prejudiciais para os algoritmos que são abordados no projeto. Portanto, a técnica utilizada para tratar esses *outliers* foi a de exclusão. Desse modo, formado a base de dados inicial para o desenvolvimento do trabalho, demonstrado na Tabela 6.

Tabela 6: Dados de ocorrências de criminalidade e de código 100. Tabela 6.a: Dados de criminalidade ou ocorrências policiais. Tabela 6.b: Dados de ocorrências de código 100.

a)

	CRIMINALIDADE				
MES	AEROPORTO	AGUA_BOA	AGUAS_LINDAS	AGUAS_NEGRAS	AGULHA
JAN	4	28	123	33	38
FEV	6	35	124	22	36
MAR	6	38	130	19	56
ABR	10	39	116	22	53
MAI	14	37	115	25	33
JUN	7	33	100	13	54
JUL	7	56	135	22	53
AGO	11	31	121	28	71
SET	4	35	117	18	98
OUT	19	39	110	17	71
NOV	10	34	124	15	44
DEZ	13	33	163	20	41

b)

	CÓDIGO_100				
MES	AEROPORTO	AGUA_BOA	AGUAS_LINDAS	AGUAS_NEGRAS	AGULHA
JAN	3	32	11	27	27
FEV	4	25	3	23	23
MAR	6	16	19	7	7
ABR	2	12	23	3	3
MAI	5	27	0	27	27
JUN	4	68	15	14	14
JUL	8	36	7	4	4
AGO	11	55	4	26	26
SET	3	39	4	24	24
OUT	9	32	1	12	12
NOV	5	17	1	8	8
DEZ	2	55	2	11	11

Autoria: Fonte própria.

Outrossim, é possível inicialmente observar nos dados da Tabela 6, uma relação entre as variáveis de quantidade de ocorrência policial e quantidade de caso de fraude ou furto de energia elétrica (código 100). Contudo, há bairros que possuem uma demasiada quantidade de ocorrências policiais e encontra partida uma baixa parcela de ações contra fraude e furto de energia. Logo, os bairros que quebram a relação entre as variáveis foram identificados e demarcados com a cor vermelha como demonstrado na Tabela 6. Por conseguinte, esses bairros possuem áreas que a distribuidora instalou um projeto de blindagem de rede denominado de SMC (Sistema de Medição Centralizado), que é o responsável pela redução de casos de código 100.

O projeto SMC ou projeto de blindagem de rede é aplicado em áreas complexas, ou seja, localidades onde as equipes de fiscalização retiram a fraude ou furto de energia normalizando assim o cliente, porém, após serem normalizados, os mesmos escolhem retornar para ilegalidade. Além disso, a estrutura SMC nos bairros pode variar de 2% até 100%, em outras palavras, existem casos no qual todo bairro ou parte do seu território é considerado como área complexa.

Portanto, na Tabela 7.c, é descrito a quantidade de clientes que recebem energia elétrica através de uma estrutura SMC, a quantidade total de clientes no bairro e o percentual de SMC por bairro. Além disso, bairros que possuem percentual de SMC acima 80% são demarcados com a cor vermelha, uma vez que, são os casos no qual o projeto SMC teve um grande impacto na redução de fraude ou furto de energia, ou seja, *outliers*. Sendo então necessário a exclusão desses dados.

Tabela 7: Dados de ocorrências de criminalidade, de código 100 e percentual de SMC por bairro. Tabela 7.a: Dados de criminalidade ou ocorrências policiais. Tabela 7.b: Dados de ocorrências de código 100. Tabela 7.c: Percentual de SMC por bairro.

a)

MES	AEROPORTO	AGUA_BOA	AGUAS_LINDAS	AGUAS_NEGRAS	AGULHA
JAN	4	28	123	33	38
FEV	6	35	124	22	36
MAR	6	38	130	19	56
ABR	10	39	116	22	53
MAI	14	37	115	25	33
JUN	7	33	100	13	54
JUL	7	56	135	22	53
AGO	11	31	121	28	71
SET	4	35	117	18	98
OUT	19	39	110	17	71
NOV	10	34	124	15	44
DEZ	13	33	163	20	41

b)

MES	AEROPORTO	AGUA_BOA	AGUAS_LINDAS	AGUAS_NEGRAS	AGULHA
JAN	3	32	11	27	27
FEV	4	25	3	23	23
MAR	6	16	19	7	7
ABR	2	12	23	3	3
MAI	5	27	0	27	27
JUN	4	68	15	14	14
JUL	8	36	7	4	4
AGO	11	55	4	26	26
SET	3	39	4	24	24
OUT	9	32	1	12	12
NOV	5	17	1	8	8
DEZ	2	55	2	11	11

c)

	AEROPORTO	AGUA_BOA	AGUAS_LINDAS	AGUAS_NEGRAS	AGULHA
QTD_SMC	0	0	5033	156	68
QTD_INSTALACAO	1322	2974	5126	2213	2665
%	0,00%	0,00%	98,19%	7,05%	2,55%

Autoria: Fonte própria

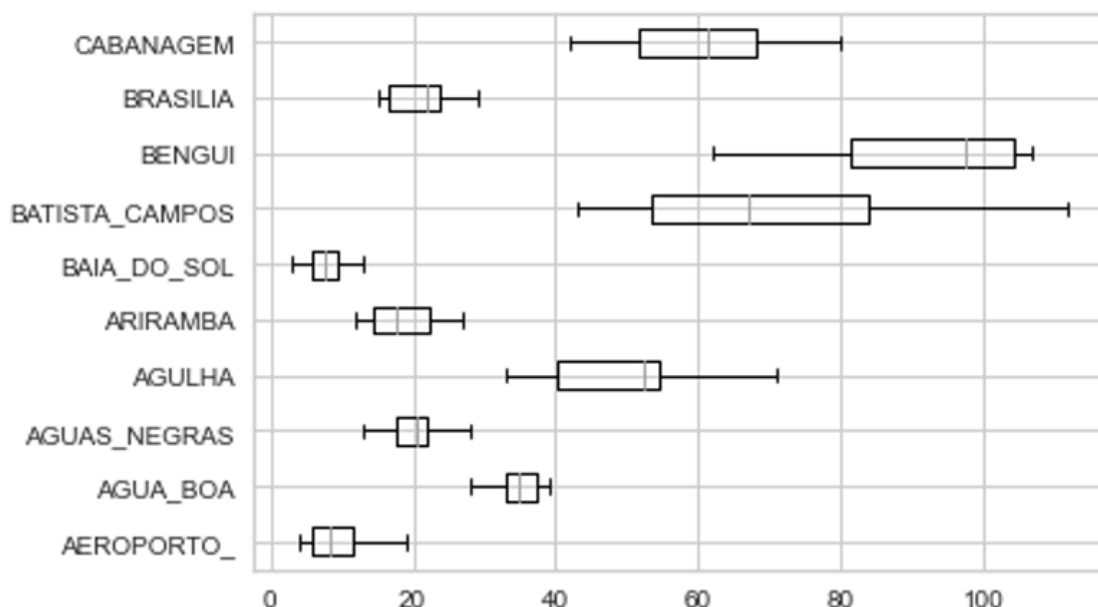
Dessa forma, dos 74 bairros identificados em ambas as bases de dados (código100 e ocorrência policial), houve uma redução de 10 bairros, sendo esses identificados como *outliers*.

Na seção seguinte será abordado de forma mais específica o tratamento de todos os dados e ferramentas utilizadas. Além disso, será denominado variável preditora ou X os dados dos bairros relacionados a ocorrência policial e variável *target* ou Y as informações dos bairros relacionados a ocorrência de código 100 (fraude ou furto de energia elétrica).

#### 4.3. PREPARAÇÃO DE DADOS

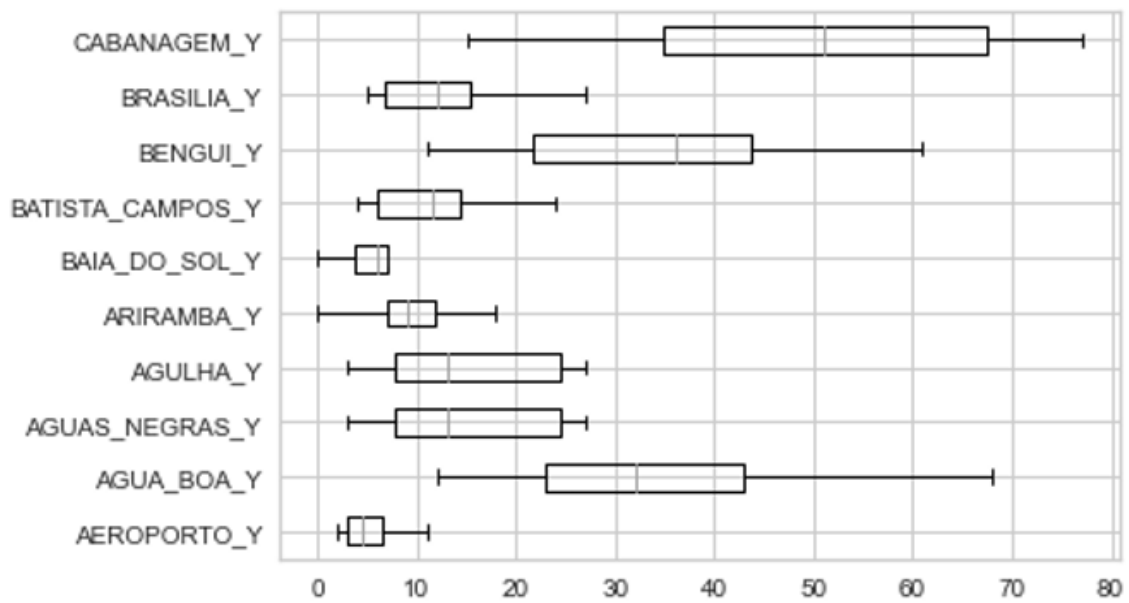
A preparação dos dados inicia-se com uma base constituída de 64 bairros de variáveis predictoras (ocorrência policial) e variáveis *target* (casos de código 100). Logo, para verificar a distribuição dessas informações, utilizou-se o *software Spyder*, que é um ambiente gratuito de código aberto escrito em *Python*. Sendo assim, foram criados gráficos de *boxplot* ou diagrama de caixas, que são gráficos que representa a composição das informações via quartis. Desse modo, na Figura 22 e 23 é possível visualizar o ordenamento de dados de 10 bairros, de variáveis predictoras e variáveis *target*.

Figura 22: Distribuição de dados preditores.



Autoria: Fonte própria

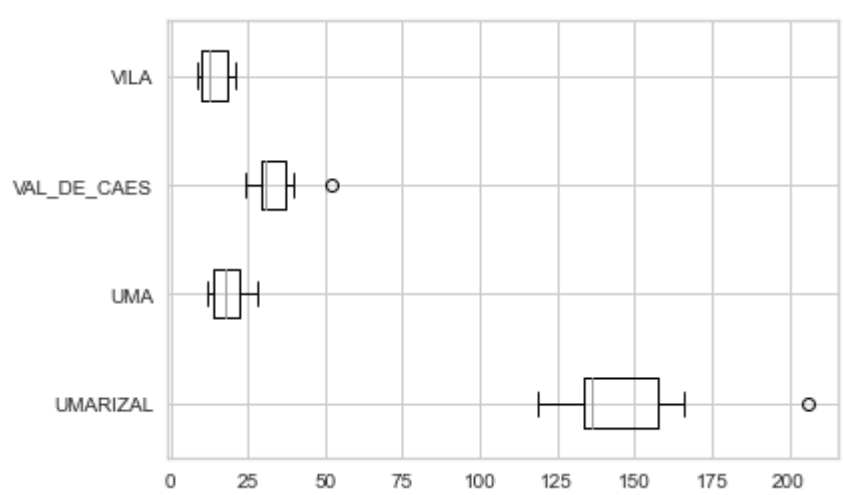
Figura 23: Distribuição de dados *target*.



Autoria: Fonte própria

Outrossim, nos diagramas em caixa é possível identificar os *outliers*. Eles são demonstrados como pontos fora do limite do digrama de caixa, como demonstrado na Figura 24.

Figura 24: Digrama de caixas e outliers



Autoria: Fonte própria



Além disso, outro ponto em questão são os meses que há 0 ocorrências. Essas duas situações ocorrem com as variáveis preditora e *target*. Contudo, a variável *target* (código 100) possui dados de anos anteriores enquanto a variável preditora (ocorrência policial) só tem informações de 2021.

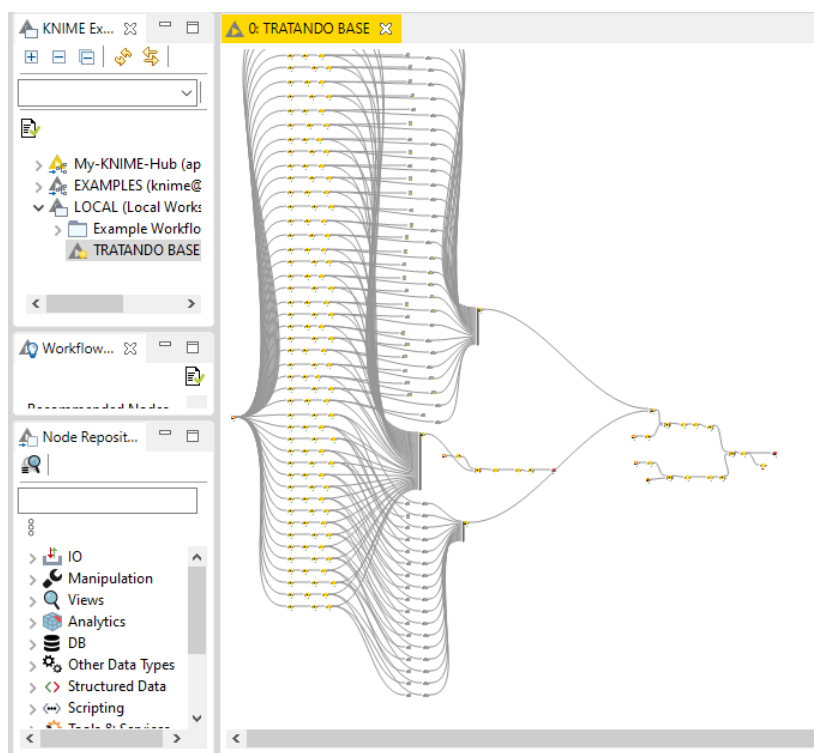
Portanto, a forma de tratar os *outliers* e valores nulos das variáveis *target*, leva em consideração os dados dos anos antecedentes os bairros analisados, possibilitando identificar a variação de: média, mediana, média harmônica, tendência e sazonalidade. Logo, permitindo alterar os valores dos *outliers* e nulos.

Destarte, o tratamento dos valores nulos e *outliers* das variáveis preditoras, por não possuir uma base dos anos precedentes foi realizado um processo de distribuição dos dados ao longo dos meses, levando em consideração a média, mediana e média harmônica.

Após o tratamento dos *outliers* e valores nulos de cada bairro das variáveis preditora e *target*. Foi necessário estruturar as 64 variáveis *target* e 64 variáveis preditoras em apenas duas variáveis identificadas como X e Y na Tabela 8, ou seja, foi transformado 128 colunas em apenas 2 colunas (X e Y). Portanto, todo esse processo foi utilizando o *software Knime*, uma plataforma livre e de código aberto de análise, construção de relatórios e integração de dados, todo construído via diagrama de blocos. Na Figura 25, é demonstrado o fluxo de trabalho no *Knime*, necessário para organizar as variáveis.

Outrossim, parti das variáveis X e Y foi possível obter o percentual de cada uma criando assim duas variáveis (VIOLENCIA\_PERCENTIL e COD100\_PERCENTIL) demonstrado na Tabela 8. Além disso, houve necessidade de inserir mais uma informação, o percentual de clientes identificados pela empresa como baixa renda, demonstrado na Tabela 8 como BAIXA\_RENDA\_PERCENT.

Figura 25: Fluxo no *Knime* para organizar as informações .



Autoria: Fonte própria

Tabela 8: *Dataset* finalizado.

	A	B	C	D	E	F	G	H	I	J
1	MES	X	Y	X_NORM	Y_NORM	BAIXA_RENDA_PERCENT	VIOLENCIA_PERCENTIL	COD100_PERCENTIL	BAIRRO	
2	JAN		4	3 -4.5	-1.5	0.322213798	0.002303	0.000168236877523553	AEROPORTO	
3	FEV		6	4 -2.5	-0.5	0.322213798	0.002303	0.000224315836698071	AEROPORTO	
4	MAR		6	6 -2.5	1.5	0.322213798	0.002303	0.000336473755047106	AEROPORTO	
5	ABR		10	2 1.5	-2.5	0.322213798	0.002303	0.000112157918349035	AEROPORTO	
6	MAI		14	5 5.5	0.5	0.322213798	0.002303	0.000280394795872589	AEROPORTO	
7	JUN		7	4 -1.5	-0.5	0.322213798	0.002303	0.000224315836698071	AEROPORTO	
8	JUL		7	8 -1.5	3.5	0.322213798	0.002303	0.000448631673396142	AEROPORTO	
9	AGO		11	11 2.5	6.5	0.322213798	0.002303	0.000616868550919695	AEROPORTO	
10	SET		4	3 -4.5	-1.5	0.322213798	0.002303	0.000168236877523553	AEROPORTO	
11	OUT		19	9 10.5	4.5	0.322213798	0.002303	0.000504710632570659	AEROPORTO	
12	NOV		10	5 1.5	0.5	0.322213798	0.002303	0.000280394795872589	AEROPORTO	
13	DEZ		13	2 4.5	-2.5	0.322213798	0.002303	0.000112157918349035	AEROPORTO	
14	JAN		28	31	-7 -0.5	0.392811297	0.008672559	0.00173844773441005	AGUA BOA	
15	FEV		34	25	-1 -6.5	0.392811297	0.008672559	0.00140197397936294	AGUA BOA	
16	MAR		38	16	3 -15.5	0.392811297	0.008672559	0.000897263346792284	AGUA BOA	
17	ABR		39	12	4 -19.5	0.392811297	0.008672559	0.000672947510094213	AGUA BOA	
18	MAI		37	27	2 -4.5	0.392811297	0.008672559	0.00151413189771198	AGUA BOA	
19	JUN		33	68	-2 36.5	0.392811297	0.008672559	0.00381336922386721	AGUA BOA	
20	JUL		36	36	1 4.5	0.392811297	0.008672559	0.00201884253028264	AGUA BOA	
21	AGO		31	55	-4 23.5	0.392811297	0.008672559	0.00308434275459847	AGUA BOA	
22	SET		36	39	1 7.5	0.392811297	0.008672559	0.00218707940780619	AGUA BOA	
23	OUT		39	32	4 0.5	0.392811297	0.008672559	0.00179452669358457	AGUA BOA	

Autoria: Fonte própria

Na seção seguinte é abordado o tema sobre modelagem e aplicação dos algoritmos *K-means* e Regressão Linear.

#### 4.4. MODELAGEM

O processo de modelagem possui duas etapas, na primeira realizamos o processo de agrupamento de dados com o algoritmo *K-means*, que agrupa as informações BAIXA\_RENDA\_PERCENT, VIOLENCIA\_PERCENTIL e COD100\_PERCENTIL. Em seguida, a segunda etapa é aplicar o algoritmo de Regressão Linear, utilizando os dados X e Y. Todos os dados utilizados para treinar esses modelos são identificados na Tabela 8.

##### 4.4.1. *K-means*

O *K-means* é um algoritmo de clusterização que separa dados agrupando-os em grupos distintos, levando em consideração suas semelhanças. É um algoritmo de aprendizagem não supervisionada que tem sua lógica baseado em distância euclidiana entre o conjunto de dados e uma determinada centroide. Portanto, o objetivo é agrupar o máximo de dados em um menor valor de centroides ou K. Segundo Fávero e Belfiore (2021, p.337), o algoritmo *K-means* é o mais utilizado por pesquisadores devido a possibilidade de o usuário definir de forma preliminar a quantidade de *clusters*.

Sendo assim, para aplicar o algoritmo foi necessário utilizar três variáveis como: BAIXA\_RENDA\_PERCENT, VIOLENCIA\_PERCENTIL e COD100\_PERCENTIL. Logo, a base utilizada para aplicação do *K-means* está demonstrado na Figura 26.

Figura 26: Base de dados para treinar o modelo *K-means*.

Index	BAIXA_RENDA_PERCENT	VIOLENCIA_PERCENTIL	COD100_PERCENTIL
0	0.322214	0.002303	0.000168237
1	0.322214	0.002303	0.000224316
2	0.322214	0.002303	0.000336474
3	0.322214	0.002303	0.000112158
4	0.322214	0.002303	0.000280395
5	0.322214	0.002303	0.000224316
6	0.322214	0.002303	0.000448632
7	0.322214	0.002303	0.000616869
8	0.322214	0.002303	0.000168237
9	0.322214	0.002303	0.000504711
10	0.322214	0.002303	0.000280395
11	0.322214	0.002303	0.000112158
12	0.392811	0.00867256	0.00173845
13	0.392811	0.00867256	0.00140197

Autoria: Fonte própria.

Para treinar o modelo, valores de  $K$  ou *clusters* distintos foram aplicados, mais específico  $K$  igual a 5, 7, 9, 11 e 13. Dessa forma, os melhores resultados foram para  $K$  igual 5 e 13. Logo, para verificar o desempenho do agrupamento foi inserindo uma coluna *Kmeans-classes* contendo as *labels* do modelo, e assim possível verificar quais valores de  $K$ , agrupavam mais quantidades de dados de um bairro em um único grupo (Figura 27).

Figura 27: Base de dados com bairros agrupados.

Index	MES	X	Y	X NORM	Y NORM	BAIYA RENDA PERCENT	VIOLENCIA PERCENTII	COD100 PERCENTII	BAIRRO	kmeans-classes
44	SET	52	24	-0.5	11	0.363104	0.0124901	0.0013459	AGULHA	2
45	OUT	71	12	18.5	-1	0.363104	0.0124901	0.000672948	AGULHA	2
46	NOV	44	8	-8.5	-5	0.363104	0.0124901	0.000448632	AGULHA	2
47	DEZ	41	11	-11.5	-2	0.363104	0.0124901	0.000616869	AGULHA	2
48	JAN	13	9	-4.5	0	0.206664	0.00458525	0.000504711	ARIRAMBA	4
49	FEV	20	4	2.5	-5	0.206664	0.00458525	0.000224316	ARIRAMBA	4
50	MAR	16	2	-1.5	-7	0.206664	0.00458525	0.000112158	ARIRAMBA	4
51	ABR	22	8	4.5	-1	0.206664	0.00458525	0.000448632	ARIRAMBA	4
52	MAI	17	8	-0.5	-1	0.206664	0.00458525	0.000448632	ARIRAMBA	4
53	JUN	12	11	-5.5	2	0.206664	0.00458525	0.000616869	ARIRAMBA	4
54	JUL	24	17	6.5	8	0.206664	0.00458525	0.000953342	ARIRAMBA	4
55	AGO	18	14	0.5	5	0.206664	0.00458525	0.000785105	ARIRAMBA	4
56	SET	15	9	-2.5	0	0.206664	0.00458525	0.000504711	ARIRAMBA	4
57	OUT	27	18	9.5	9	0.206664	0.00458525	0.00100942	ARIRAMBA	4
58	NOV	13	11	-4.5	2	0.206664	0.00458525	0.000616869	ARIRAMBA	4
59	DEZ	24	9	6.5	0	0.206664	0.00458525	0.000504711	ARIRAMBA	4
60	JAN	7	5	-0.5	-0.5	0.272923	0.00195029	0.000280395	BAIA DO SOL	1
61	FEV	8	7	0.5	1.5	0.272923	0.00195029	0.000392553	BAIA DO SOL	1
62	MAR	13	5	5.5	-0.5	0.272923	0.00195029	0.000280395	BAIA DO SOL	1
63	ABR	13	3	5.5	-2.5	0.272923	0.00195029	0.000168237	BAIA DO SOL	1
64	MAI	11	6	3.5	0.5	0.272923	0.00195029	0.000336474	BAIA DO SOL	1
65	JUN	7	8	-0.5	2.5	0.272923	0.00195029	0.000448632	BAIA DO SOL	1
66	JUL	3	8	-4.5	2.5	0.272923	0.00195029	0.000448632	BAIA DO SOL	1
67	AGO	5	7	-2.5	1.5	0.272923	0.00195029	0.000392553	BAIA DO SOL	1
68	SET	9	4	1.5	-1.5	0.272923	0.00195029	0.000224316	BAIA DO SOL	1
69	OUT	4	3	-3.5	-2.5	0.272923	0.00195029	0.000168237	BAIA DO SOL	1
70	NOV	6	5	-1.5	-0.5	0.272923	0.00195029	0.000280395	BAIA DO SOL	1
71	DEZ	8	7	0.5	1.5	0.272923	0.00195029	0.000392553	BAIA DO SOL	1
72	JAN	70	15	3	3.5	0.0683223	0.0173659	0.000841184	BC	3
73	FEV	64	6	-3	-5.5	0.0683223	0.0173659	0.000336474	BC	3

Autoria: Fonte própria.

Após a modelagem pelo *K-means* o melhor valor escolhido para *clusters* foi de  $K$  igual a 5, ou seja, todos os bairros foram agrupados em 5 grupos, além que os dados de um bairro não foram agrupados em diferentes grupos. Em seguida será abordado na seção seguinte a aplicação do modelo de Regressão Linear.

#### 4.4.2. Regressão Linear

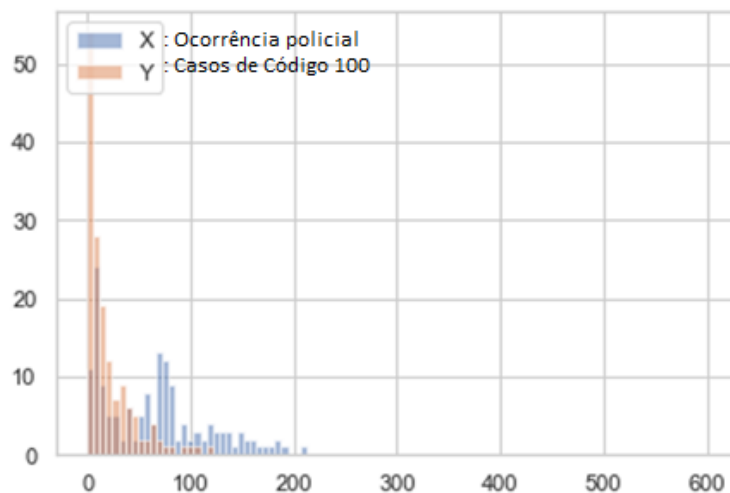
De acordo com Fávero e Belfiore (2021, p.505), a técnica de regressão linear é provavelmente uma técnica de análise de dados muito utilizada para entender a relação entre o comportamento de determinado fenômeno e o comportamento de uma ou mais

variáveis potencialmente preditoras, sem que haja, contudo, uma relação obrigatória de causa e efeito.

O agrupamento dos bairros realizado pelo algoritmo *K-means* foi dividido em 5 grupos, onde varia do 0 até 4, essa variação está ligada à relação entre quantidade de ocorrências policiais e ocorrência de fraude ou furto de energia. Entretanto, o grupo 0 representa o caso de bairros que não conseguiram ser agrupados.

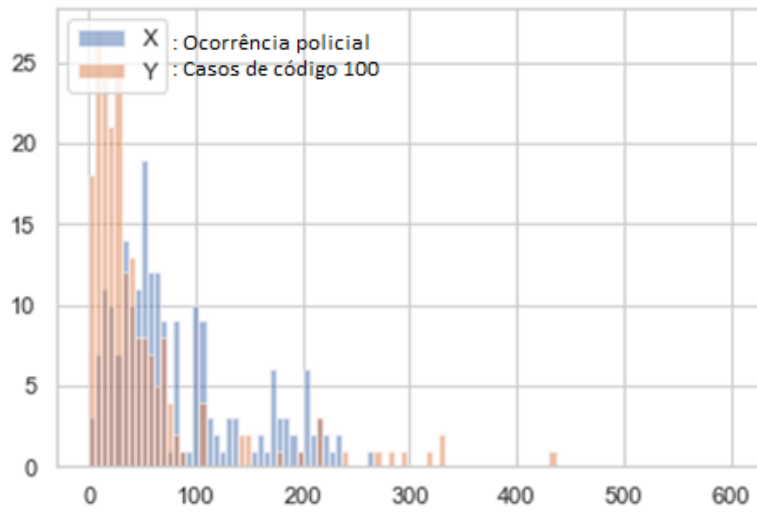
Em seguida, para realizar o processo de normalização ou padronização é necessário observar a distribuição dos dados por grupos. Portanto, foi gerado histogramas que são gráficos em barras que representam a distribuição da frequência das variáveis X (ocorrências policiais) e Y (casos de código 100), subdivididos por grupos e demonstrado nas Figuras 28 até 31.

Figura 28: Histogramas dos bairros agrupados como grupo 1.



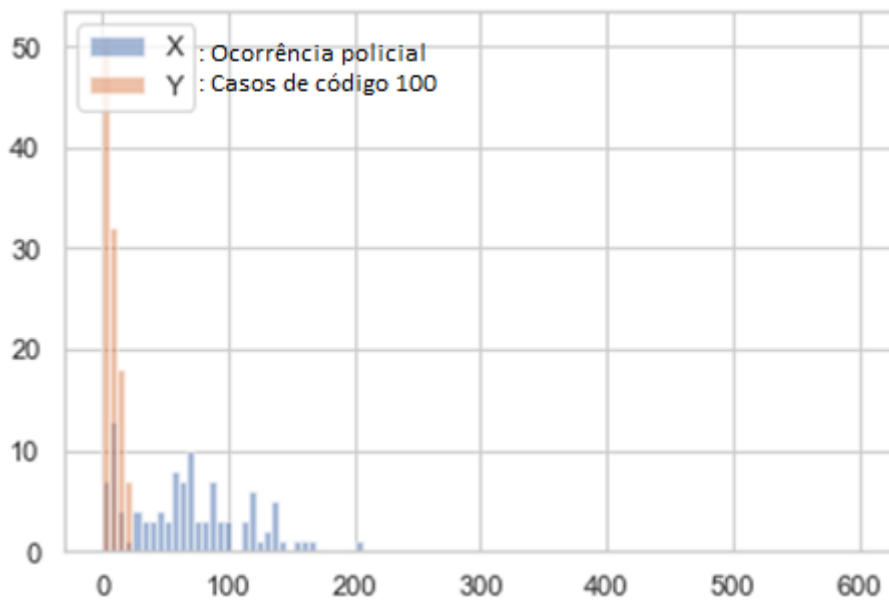
Autoria: Fonte própria.

Figura 29: Histogramas dos bairros agrupados como grupo 2.



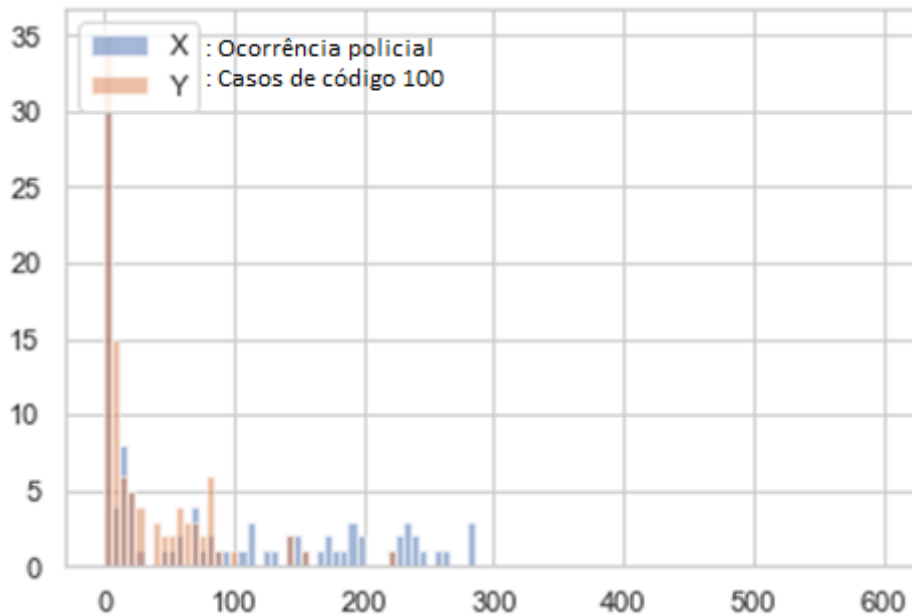
Autoria: Fonte própria.

Figura 30: Histogramas dos bairros agrupados como grupo 3.



Autoria: Fonte própria.

Figura 31: Histogramas dos bairros agrupados como grupo 4.



Autoria: Fonte própria.

Após visualizar os histogramas dos grupos aplicamos uma função de correlação, que aplica o conceito de correlação de Pearson entre a variável X e Y. De acordo com Fávero e Belfiore (2021, p.382), a correlação de Pearson é uma medida do grau da relação linear entre duas variáveis, podendo variar de -1 a 1, um valor mais próximo de um desses extremos indica a existência de relação linear entre duas variáveis, que podem contribuir significativamente para a extração de um único fator. Contudo, um valor com correlação de Pearson muito próximo de 0 indica que a relação linear entre as duas variáveis é praticamente inexistente; portanto, diferentes fatores podem ser extraídos. Na Tabela 9 é possível identificar os valores de correlação de cada grupo.

Tabela 9: Coeficientes de correlação.

<b>GRUPO</b>	<b>CORRELAÇÃO</b>
1	0,32
2	0,33
3	0,81
4	0,47

Autoria: Fonte própria.

Posteriormente, os dados dos grupos são normalizados utilizando a biblioteca *StandardScaler* do *Sklearn*, o que faz com que, a média seja igual a zero e desvio padrão

igual a 1 demonstrado na Figura 32. E por fim os dados normalizados são inseridos no modelo podendo ver o código na Figura 33.

Figura 32: Normalizando os dados.

```
322 #STANDARDSCALER
323 num_observ = len(df_1)
324 XSL_k5b1 = df_1['X'].values.reshape((num_observ,1))
325 YSL_k5b1 = df_1['Y'].values.reshape((num_observ,1))
326
327 scaler = StandardScaler().fit(XSL_k5b1)
328 normal_xls_k5b1 = scaler.transform(XSL_k5b1)
329
330 scaler = StandardScaler().fit(YSL_k5b1)
331 normal_yls_k5b1 = scaler.transform(YSL_k5b1)
332
```

Autoria: Fonte própria.

Figura 33: Chamando o modelo e treinando.

```
516 # CRIANDO MODELO
517 modelo = LinearRegression(normalize = False, fit_intercept = True)
518
519 # TREINANDO MODELO
520
521 modelo_geral = modelo.fit(normal_xls_k5,normal_yls_k5)
522 modelo_0 = modelo.fit(normal_xls_k5b0,normal_yls_k5b0)
523 modelo_1 = modelo.fit(normal_xls_k5b1,normal_yls_k5b1)
524 modelo_2 = modelo.fit(normal_xls_k5b2,normal_yls_k5b2)
525 modelo_3 = modelo.fit(normal_xls_k5b3,normal_yls_k5b3)
526 modelo_4 = modelo.fit(normal_xls_k5b4,normal_yls_k5b4)
527
528
```

Autoria: Fonte própria.

Na seção posterior será discutido os resultados do modelo para os quatro agrupamentos de bairros.

#### 4.5. AVALIAÇÃO

Para avaliação do desempenho dos modelos de cada grupo, se utiliza a métrica  $R^2$ . Segundo Fávero et al. (2009), a capacidade explicativa do modelo é analisada pelo  $R^2$  da regressão, conhecido também como coeficiente de ajuste ou de explicação. Para o modelo de regressão simples, essa medida indica quanto do comportamento da variável Y é explicado pelo comportamento da variável X, sabendo que, não é indicativo de efeito ou causalidade entre as variáveis.



Ainda de acordo com Fávero et al. (2009), o  $R^2$  pode variar entre 0 e 1, mas é praticamente impossível a obtenção de um  $R^2$  igual a 1 e quanto mais próximo de 0 menos as variáveis X e Y se relacionarão.

Após calcular os valores de  $R^2$  os melhores resultados são os bairros do grupo 2 e 4, demonstrado na Tabela 10.

Tabela 10: Coeficiente  $R^2$

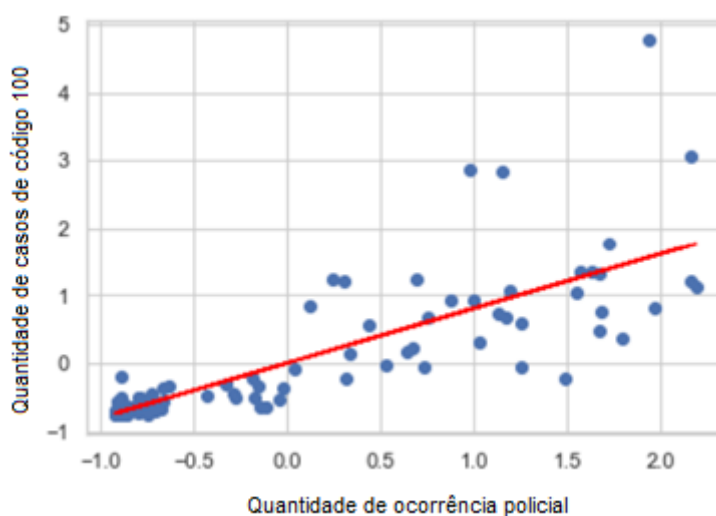
<b>GRUPO</b>	<b><math>R^2</math></b>
2	0,35
4	0,64

Autoria: Fonte própria.

O resultado do  $R^2$  demonstrado na Tabela 10, é bastante interessante, principalmente o do grupo 4. Sendo que, o grupo 4 representa bairros com grande quantidade de ocorrências policiais e ações de combate à fraude e furto de energia elétrica. Portanto, por ter um  $R^2$  igual a 0,64, indica que o comportamento da variável Y (fraude ou furto de energia elétrica) é explicado 64% pelo comportamento de variação da variável X (ocorrências policiais), nos casos dos bairros que foram agrupados como grupos 4.

Outrossim, com relação ao modelo demonstrado na Figura 34, seu comportamento demonstra o acompanhamento da distribuição dos dados. Logo, pode ser utilizado para estimar aproximadamente a quantidade de casos de código 100, de bairros que são agrupados com as mesmas características e semelhanças que os bairros do grupo 4.

Figura 34: Modelo de Regressão Linear do grupo 4.



Autoria: Fonte própria.

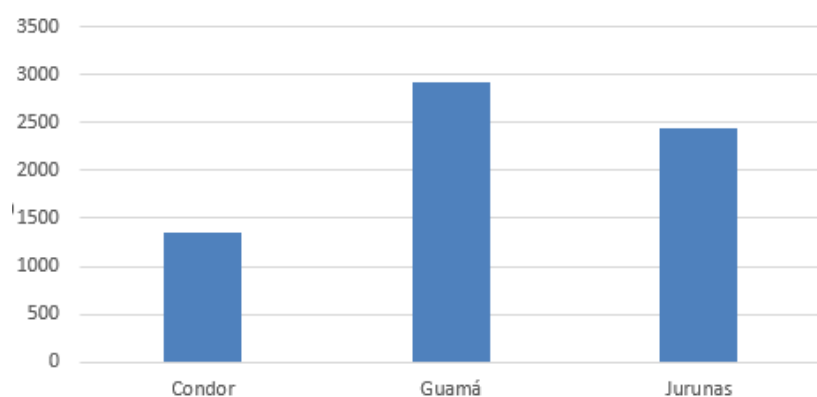
Em relação aos resultados dos demais grupos, dentre os fatores que podem indicar um valor baixo do  $R^2$ , está a baixa relação nesses agrupamentos de bairros de ocorrências policiais e fraude ou furto de energia. Portanto, existe outros fatores para ação de furto de energia pelo consumidor e além que não são bairros com muita ação policial. Existe também a questão do tratamento dos dados o que pode ter impactado negativamente sendo necessário uma outra abordagem.

## **5. CONCLUSÃO**

A partir dos resultados obtidos nos modelos de Regressão Linear, mais especificamente do agrupamento dos bairros denominado grupo 4, onde possui bairros com grande quantidade de chamado policial e de fraude e furto de energia elétrica. É possível perceber que bairros com auto índice de violência urbana possui uma cultura ou facilidade para ter ações de fraude ou furto de energia elétrica.

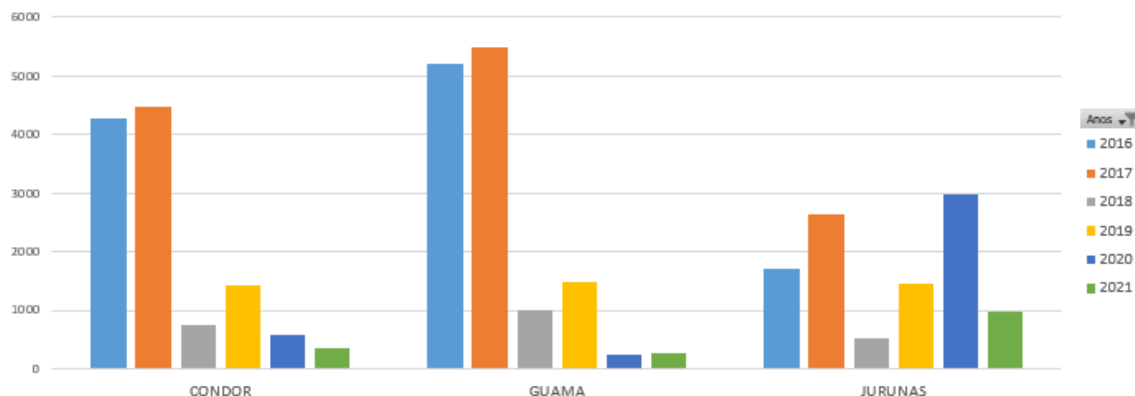
Além dos resultados do modelo podemos aferir a relação de violência urbana e quantidade de ações de combate à fraude e furto de energia, parti dos dados de bairros com SMC (blindagem da rede de distribuição). Esse projeto é definido pela distribuidora como uma forma de combater PNT de baixa tensão em bairros alta complexidade, ou seja, regiões onde mesmo com grande quantidade de fiscalizações o usuário retorna para ilegalidade. Logo, selecionando os bairros que possuem um percentual de mais de 80% do território com SMC e os mesmos sendo de senso comum da população que são áreas muito perigosas, temos que, em 2021, esses bairros tiveram mais de 1000 ocorrências policiais demonstrado na Figura 41. Além disso, antes do projeto SMC ser implantado nessas regiões, ou seja, anterior ao ano de 2018, esses bairros possuem muitas ações de combate à fraude e furto de energia elétrica demonstrado na figura 35.

Figura 35: Quantidade de ocorrências policial em 2021



Autoria: Fonte própria.

Figura 36: Quantidade ações contra fraude e furto de energia



Autoria: Fonte própria.

Sendo assim, é possível constatar a relação entre criminalidade urbana e fraude ou furto de energia elétrica. Portanto, esse trabalho pode auxiliar na identificação de áreas complexas, parti do uso do algoritmo de *K-means*, em novas concessões do Grupo Equatorial Energia.

Além disso, como o modelo de regressão linear do grupo 4 é possível estimar com dados de ocorrência policial quantos casos de código 100, possivelmente pode haver em um determinado bairro, em casos que os bairros tenham semelhança com os bairros do grupo 4.

Todo o desenvolvimento desse trabalho foi um estudo analisando os bairros de Belém do Pará, fazendo com que seja possível identificar os bairros que possuem um alto grau de violência urbana e ao mesmo tempo uma grande quantidade de casos de fraude e

furto de energia elétrica, fazendo com que a empresa tenha investido em projetos como blindagem da rede de distribuição. Portanto esse, estudo leva em consideração as informações de Belém com objetivo de aplicar em outras cidades em outras concessões da empresa e assim identificando bairros que possuem as mesmas características que os de Belém – PA.

O mais interessante seria se o trabalho fosse uma análise consideração os dados de violência urbana de 2016, uma vez que no ano de 2016 o grupo equatorial energia ainda não havia implantado projetos de blindagem de rede. Com isso, seria possível identificar com mais eficiência os bairros que a empresa classificou como complexos e possuem blindagem de rede. Como o Centro Integrado de Operações teve problemas em sua base de dados, só dispunham de informações do ano de 2021, sendo assim foi necessário excluir muitas informações que prejudicariam na análise do trabalho.

## REFERÊNCIAS

ANTMANN, Pedro - **Reducing Technical and Non-Technical Losses in the Power Sector** - Paper for the World Bank Group Energy Sector Strategy - 2009. Disponível <[openknowledge.worldbank.org](http://openknowledge.worldbank.org)> Acesso em: 08/11/2021

ANEEL. **Regulação dos Serviços de Distribuição**, 2015 , Agência Nacional de Energia Elétrica. Disponível em: <[www.aneel.com.br](http://www.aneel.com.br)> Acesso em: 08/11/2011.

ANEEL. **Resolução Normativa nº 414**, Agência Nacional de Energia Elétrica. Brasília, 2010.

ANEEL. **Perda de Energia Elétrica na Distribuição**, Agência Nacional de Energia Elétrica. Brasília, p. 2-9, 2021.

ANEEL. **Programa de Pesquisa e Desenvolvimento Tecnológico do Setor de Energia Elétrica**, Agência Nacional de Energia Elétrica. Disponível em: <<https://antigo.aneel.gov.br/programa-de-p-d> > Acesso em: 31/11/2020.

ABRADEE. **Setor Elétrico**, Associação Brasileira de Distribuidoras de Energia Elétrica, 2021. Disponível em: <[www.abradee.com.br](http://www.abradee.com.br)> Acesso em: 07/11/2020.

ALVES, H. M. D. M. **Análise da Contribuição de Atributos Derivados do Histórico de Consumo para Detecção de Perdas Não Técnicas**. Univ. Fed. de Campina Grande. Dep. de Eng. Elétrica. Dissertação de Mestrado. Campina Grande ,p.65-67.2019

BARROS, R.M.R. **Advanced Analytics Aplicada à Gestão da Perda Não Técnica de Energia em Sistemas Elétricos de Distribuição**. Univ. Fed. de Campina Grande. Dep. de Eng. Elétrica. Tese de doutorado. Campina Grande,202

DICK, A.J. **Theft of electricity-how UK electricity companies detect and deter**. Proceedings of the European Convention on Security and Detection. Brighton: IET. 1995.p. 85-97.

Distribuição, **Equatorial Energia**, 2022. Disponível em:<<https://www.equatorialenergia.com.br/nossos-negocios/distribuicao/>>. Acesso em: 01/12/2021.

FÁVERO, L. P.; BELFIORE, P.; SILVA, F.L; CHAN, B.L. **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: Campus Elsevier,2009.

FÁVERO, L. P.; BELFIORE, P. **Manual de Análise de Dados Estatística e Modelagem Multivariada com Excel, SPSS e Stata**; Rio de Janeiro; LTC, 2021.

HARRISON, M. **Machine Learning Guia de Referência Rápida Trabalhando com dados estruturados em Python**. São Paulo: O' Reilly, 2020.

SENDI. **Trabalhos Técnicos**, Seminário Nacional de Distribuição de Energia Elétrica, 2018. Disponível em: <<https://sendi.org.br/sendi2018/trabalhos-tecnicos.html>> Acesso em: 12/12/2020.

SENDI. **Trabalhos Técnicos**, Seminário Nacional de Distribuição de Energia Elétrica, 2016. Disponível em: <<http://abradee03.org/sendi2016/trabalhos/>> Acesso em : 11/12/2020.

RAMOS, C. C. O. et al. A New Approach for nntechnical Losses Detection Based on Optimum-Path Forest. **IEEE Transaction on Power Systems**, v. 26, n.1, p. 180-188, Febuary 201. ISSN 0885-8950.

ZHENG, A.; CASARI, A. **Feature Engineering for Machine Learning**. Boston: O' Reilly ,2018

## ANEXO A - TIPOS DE NATUREZA DE OCORRÊNCIA POLICIAL

NATUREZAS
Abordagem a pessoas em atitude suspeita
Abordagem a veículo
Ameaça
Arremesso de projétil contra veículo
Assédio sexual
Ato Infracional (Furto)
Ato Infracional (Roubo)
Ato Infracional (Tentativa de Furto)
Choque elétrico
Comercialização e tráfico de animais
Comércio ilegal de arma de fogo
Comunicação de aliciamento
Comunicação de aliciamento de menor
Comunicação de pedofilia
Comunicação de privação da liberdade
Comunicação de tráfico de pessoas
Comunicação de violação
Crueldade contra os animais
Desordem
Disparo de arma de fogo
Estelionato
Estupro
Estupro de vulnerável
Exterminio de Animal
Furto
Furto à Transeunte
Furto de Animal em Área Urbana
Furto de Cabos de Transmissão de Energia
Furto de Coisa Comum
Furto de Energia Elétrica
Furto de Objeto
Furto de veículo
Furto de Veículo Automotor
Furto de Veículo de Propulsão Humana
Furto em Estabelecimento Comercial ou de Serviços
Furto em Estabelecimento de Ensino
Furto em Estabelecimento Público Estadual
Furto em Estabelecimento Público Municipal
Furto em Instituição Financeira
Furto em Residência
Furto em Veículo
Homicídio
Homicídio Contra a Mulher
Lesão corporal

Lesão corporal culposa
Lesão corporal culposa no trânsito
Lesão corporal decorrente de intervenção policial
Lesão corporal dolosa praticada contra agente de segurança pública ou seus familiares
Morte acidental provocada por eletroplessão (choque elétrico)
Morte por intervenção de agente do estado
Paralisação de trabalho com violência
Pichar ou conspurcar edificação ou monumento urbano
Policiamento ostensivo em manifestação - rebelião
Porte de arma branca
Porte de drogas para consumo pessoal
Porte ilegal de arma de fogo
Posse irregular de arma de fogo
Posse ou porte ilegal de arma de fogo de uso restrito
Receptação
Rixa
Roubo
Roubo - Fios e cabos elétricos
Roubo à Transeunte
Roubo com Restrição da Liberdade (Sequestro Relâmpago)
Roubo de Veículo Automotor
Roubo de Veículo de Propulsão Humana
Roubo em Estabelecimento Comercial e de Serviços
Roubo em Estabelecimento Público Municipal
Roubo em Residência
Roubo em Veículo
Roubo Seguido de Lesões
Roubo seguido de morte - latrocínio
Sequestro e cárcere privado
Simulacro de arma de fogo
Tentativa de estelionato
Tentativa de estupro
Tentativa de extorsão
Tentativa de fraude
Tentativa de furto
Tentativa de Furto à Transeunte
Tentativa de Furto de Veículo
Tentativa de Furto de Veículo de Propulsão Humana
Tentativa de Furto em Estabelecimento Comercial
Tentativa de Furto em Estabelecimento Público Municipal
Tentativa de Furto em Estabelecimento Religioso
Tentativa de Furto em Residência
Tentativa de Furto em Veículo
Tentativa de homicídio
Tentativa de Homicídio Contra a Mulher



Tentativa de roubo
Tentativa de Roubo a Estabelecimento Comercial
Tentativa de Roubo a Transeunte
Tentativa de Roubo de Veículo
Tentativa de suicídio
Tentativa de violação de domicílio
Tortura
Tráfico de drogas
Uso de documento falso
Usurpação - esbulho possessório
Vadiagem
Vender a criança ou adolescente, produtos que causem dependência física ou psíquica
Vias de fato
Violação de domicílio
Violência doméstica

ANEXO B – BAIRROS IDENTIFICADOS NAS BASES DA  
DISTRIBUIDORA DE ENERGIA ELÉTRICA E DO CENTRO INTEGRADO DE  
OPERAÇÕES

<b>BAIRROS</b>
<b>AEROPORTO</b>
<b>AGUA_BOA</b>
<b>AGUAS_LINDAS</b>
<b>AGUAS_NEGRAS</b>
<b>AGULHA</b>
<b>ARIRAMBA</b>
<b>ATALAIA</b>
<b>BAIA_DO_SOL</b>
<b>BARREIRO</b>
<b>BATISTA_CAMPOS</b>
<b>BENGUI</b>
<b>BRASILIA</b>
<b>CABANAGEM</b>
<b>CAMPINA</b>
<b>CAMPINA_DE_ICORACI</b>
<b>CANUDOS</b>
<b>CARANANDUBA</b>
<b>CARUARA</b>
<b>CASTANHEIRA</b>
<b>CENTRO</b>
<b>CHAPEU_VIRADO</b>
<b>CIDADE_VELHA</b>
<b>CONDOR</b>
<b>COQUEIRO</b>
<b>CREMACAO</b>
<b>CRUZEIRO_ICORACI</b>
<b>CURIO_UTINGA</b>
<b>DISTRITO_INDUSTRIAL</b>
<b>FAROL</b>
<b>FATIMA</b>
<b>GUAMA</b>
<b>GUANABARA</b>
<b>ICORACI</b>
<b>ITAITEUA</b>
<b>JURUNAS</b>
<b>MAGUARI</b>
<b>MANGUEIRAO</b>

<b>MANGUEIRAS</b>
<b>MARACACUERA</b>
<b>MARACAJA</b>
<b>MARACANGALHA</b>
<b>MARAHU</b>
<b>MARAMBAIA</b>
<b>MIRAMAR</b>
<b>MONTENSE_TERRA_FIRME</b>
<b>MURUBIRA</b>
<b>NATAL_DO_MURUBIRA</b>
<b>NAZARE</b>
<b>OUTEIRO</b>
<b>PARAISO</b>
<b>PARQUE_GUAJARA</b>
<b>PARQUE_VERDE</b>
<b>PEDREIRA</b>
<b>PONTA_GROSSA</b>
<b>PORTO_ARTHUR</b>
<b>PRAIA_GRANDE</b>
<b>PRATINHA</b>
<b>REDUTO</b>
<b>SACRAMENTO</b>
<b>SANTA_ROSA</b>
<b>SAO_BRAS</b>
<b>SAO_CLEMENTE</b>
<b>SAO_FRANCISCO</b>
<b>SAO_JOAO_DO_OUTEIRO</b>
<b>SOUZA</b>
<b>SUCURIJUQUARA</b>
<b>TAPANA</b>
<b>TELEGRAFO</b>
<b>TENONE</b>
<b>TERRA_FIRME</b>
<b>UMARIZAL</b>
<b>UnA</b>
<b>VAL_DE_CAES</b>
<b>VILA</b>