Análise de Agrupamentos de Atributos de um Sistema de Armazenamento de Energia por Bateria de Chumbo-Carbono

**Álefe Cruz** ¹

**Andrea Sarmento Maia** ²,³

**Pedro Henrique Paschoal Travassos** ¹

**Rodrigo Simões** ²

1 Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil,

2 Pós-graduação em Engenharia de Sistemas, Escola Politécnica de Pernambuco, Pernambuco, Brasil,

3 Instituto de Tecnologia Edson Mororó Moura, ITEMM, Pernambuco, Brasil.

**E-mail do autor principal: Andrea Sarmento** [asm@ecomp.poli.br](mailto:asm@ecomp.poli.br)

**Resumo**



Sistema de armazenamento de energia por baterias é um conjunto de elementos de potência e de baterias, comumente, denominado de BESS, do inglês, *Battery Energy Storage System* cuja principal finalidade é suportar o setor elétrico nos diversos âmbitos de atuação: geração, transmissão, distribuição e consumidor final. Para operar de forma ótima, tem-se a necessidade de interpretação do atributo do estado de carga, do inglês, *State of Charge*. É com esse indicador que é possível indicar se o sistema tem ou não energia suficiente para determinada atuação. A partir de agrupamentos pela metodologia K-means desse atributo é possível indicar diferentes comportamentos do SoC combinado com a tensão, corrente e temperatura do BESS.

**Palavras-Chave*:***Mineração de dados; Agrupamento; BESS; SoC; K-means

***Abstract***



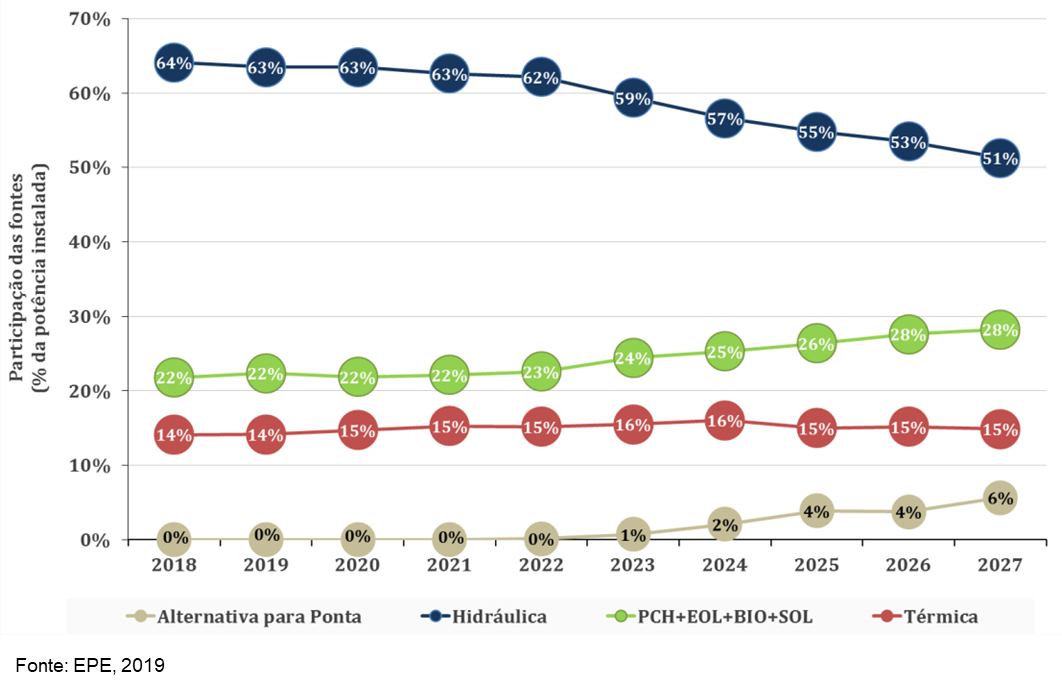
*Battery Energy Storage System is a set of power and battery elements, commonly called BESS, whose main purpose is to support the electrical sector in the various fields of activity: generation, transmission, distribution and final consumer. In order to operate optimally, it is necessary to interpret the state of charge attribute. It is with this indicator that it is possible to indicate whether or not the system has enough power for a certain activity. By using the K-means methodology of this attribute, it is possible to indicate different behaviors of the SoC combined with the voltage, current and temperature of the BESS.*

***Key-words:*** *Data mining; Grouping; BESS; SoC; K-means.*

# **Introdução**

## **Contextualização**

A utilização de fontes renováveis para a geração de energia vem crescendo nos últimos anos com destaque para a energia eólica e solar, as quais são fontes limpas e inesgotáveis **[1, 2].** Estima-se que até 2030 a energia total de usinas eólicas e solares existentes no mundo seja de 7,7 TWh **[3]**. No Brasil, portanto, a expectativa é que a representatividade dessas fontes renováveis seja superior a 28% na matriz energética a partir de 2027, como apresentado na **Figura 1** **[4]**.



**Figura 1 -** Estimativa da participação das fontes energéticas no Brasil até 2027

Apesar do uso de fontes renováveis trazer vantagens ao meio ambiente, a sua inserção intensa ao sistema elétrico ainda passa por dificuldades, justificado por grandezas não-controláveis (vento e da radiação solar) que geram flutuações de energia elétrica, por isso, garantir segurança, estabilidade e confiabilidade ao sistema elétrico se torna um desafio. A mitigação da intermitência dessas fontes, portanto, é crucial e mandatório para o setor elétrico.

No Brasil, a inserção de soluções contemplando armazenamento de energia por baterias ainda é pouco representativa, mas tem ganhado visibilidade a partir dos resultados da Chamada de Projetos de Pesquisa e Desenvolvimento Estratégico 21/2016 da Agência Nacional de Energia Elétrica (ANEEL), focada na inserção desses sistemas no setor elétrico nacional, tanto em aspectos técnicos, como econômicos e regulatórios. Ademais, no desenvolvimento da cadeia produtiva nacionais desses sistemas.

Em resposta a esse movimento de P&D e da inserção de novos produtos, o Instituto Edson Mororó Moura (ITEMM) está suportando o desenvolvimento dessas soluções. Os dados a serem apresentados nesse artigo, são de sistemas em operação, especificado e desenvolvido pelo Instituto.

## **Descrição do Problema**

No BESS, o sistema de gerenciamento das baterias é um elemento essencial, pois permite a operação adequada dos equipamentos e, consequentemente, do sistema. Esse sistema tem como objetivo:

* Operar as baterias de acordo com um Estado de Carga (SoC, *State of Charge*) delimitado;
* Controlar a corrente e tensão máxima de carga/descarga;
* Definir uma profundidade de descarga adequada (DoD, *Depth of Discharge*);
* Implementar um procedimento apropriado de carga/descarga de baterias que garante as condições operacionais, conforme especificado pelo fabricante;
* Garantir que as baterias estão operando corretamente, no intervalo definido pelo fabricante.

A qualidade do controle realizado por esse gerenciado está diretamente relacionada à assertividade dos modelos embarcados. Portanto, é essencial realizar correta parametrização das baterias e controlar o processo de produção para reduzir os desvios entre amostras, permitindo um modelo mais eficaz.

O SoC da bateria, o qual é utilizado para descrever a capacidade remanescente de um acumulador, é um parâmetro muito importante para a estratégia de controle. Logo, há uma necessidade de verificar o intervalo de operação do SoC com precisão, de modo a permitir uma maior proteção da bateria e prevenir situações de utilização que ofereçam risco à vida útil. Além disso, os dados de medição do SoC, da tensão das *strings* e das células e das temperaturas da célula tornam possível a operabilidade adequada.

## **Objetivo**

Nesse contexto, esse artigo tem o objetivo de apresentar a metodologia e os resultados obtidos no desenvolvimento de algoritmo de agrupamento de operabilidade do BESS, indicando possíveis melhorias nas baterias, a partir dos intervalos de atuação.

## **Justificativa**

Atualmente, não há uma rotina de verificação e/ou identificação da operabilidade das baterias do BESS. Logo, identifica-se como uma oportunidade técnica e financeira para a tomada de decisão do desenvolvedor de produto. Reduzindo, consequentemente, custos atrelados a operação e manutenção do produto.

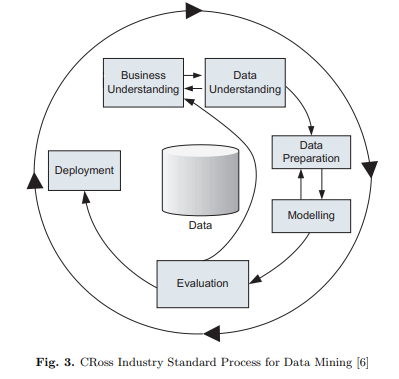
## **Escopo Negativo**

Nesse desenvolvimento, não será considerada a análise de estado de saúde das baterias e, nem, o perfil de consumo por aplicação da solução.

# **Fundamentação Teórica**

Um dos maiores desafios dos gestores e operadores do produto BESS está na usabilidade dos dados gerados pelo sistema, assim é essencial o desenvolvimento de um algoritmo de mineração de dados para extrair e relacionar informações dos diversos sensores e medidores (tensão, temperatura, corrente, umidade, fumaça etc.) que há no BESS.

À vista disso, para suportar o entendimento dos dados e avaliação do negócio, utilizou-se a metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*), criada em 1996 com o objetivo de padronizar os processos de mineração de dados a partir de seis fases, **Figura 2**.



**Figura 2 -** Cross Industry Standard Process for Data Mining (CRISP-DM) **[3]**

Cada etapa tem finalidade distinta e é essencial para o sucesso da mineração de dados. Abaixo, tem-se um resumo de cada fase **[3]**:

* *Business Understanding* (Entendimento do negócio) – fase de entendimento do negócio, dos objetivos e requerimentos do projeto;
* *Data Understanding* (Entendimento dos dados) – fase de compreensão, verificação de problemas e os primeiros *insights* nos dados;
* *Data Preparation* (Preparação dos dados)– fase responsável pelo tratamento dos dados (seleção, limpeza e transformação) para a fase posterior;
* *Modeling* (Modelagem) – fase para aplicação das técnicas (modelagem e algoritmos) de mineração de dados;
* *Evaluation* (Avaliação) – fase de construção de um modelo para análise dos resultados;
* Deployment (Implantação) – fase da aplicação do conhecimento gerado na tomada de
* Decisões.

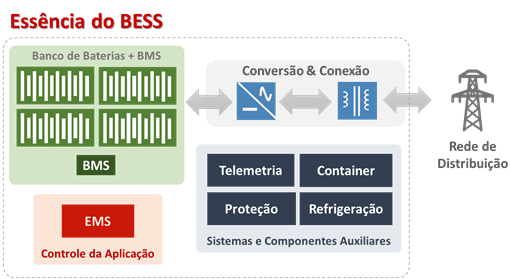
Nesse artigo, seguiremos a metodologia para desenvolvimento do algoritmo de agrupamento dos dados do BMS.

## **Área do Negócio**

A primeira etapa, segundo a metodologia CRISP-DM, é a de Entendimento do Negócio cuja principal finalidade é entender, estrategicamente, a solução e os dados fornecidos. Abaixo, portanto, detalha-se os elementos, a sinergia com o setor elétrico e as aplicações mais comuns.

O BESS, portanto, converte energia elétrica em eletroquímica para armazená-la, quando necessário, converter-se de volta em energia elétrica **[4]**. Esse sistema é dividido em grandes partes, conforme apresentado na **Figura 3**:

1. Banco de baterias – conjunto de células de armazenamento de energia;
2. Sistema de gerenciamento de baterias (*Battery Management System* - BMS);
3. Sistema de conversão de energia – (*Power Conversion System* - PCS);
4. Sistema de gerenciamento de energia (*Energy Management System* - EMS);
5. Sistemas e componentes auxiliares – composto por contêiner, sistema HVAC (*Heating Ventilation and Air Conditioning*) e sistema de segurança.



**Figura 3 –** Ilustração do Sistema de Armazenamento de Energia

Existem quatro formas básicas para suavizar os problemas ocasionados pela inclusão das fontes renováveis: (i) uso de centrais geradoras flexíveis de rápida entrada em operação; (ii) aumento de interconexão com regiões vizinhas – aumento da potência de curto circuito (rede mais forte); (iii) aplicação de controle pelo lado da carga (resposta do lado de demanda, que requer a opção pelo cliente e rede de comunicação forte) e, finalmente; iv) uso de sistemas de armazenamento por baterias (BESS, *Battery* *Energy Storage System*).

As aplicações desses sistemas podem acontecer em todos os âmbitos do sistema elétrico (geração, transmissão, distribuição e consumidor final) e são divididas em cinco categorias: suprimento de energia, serviços ancilares, serviços de infraestrutura e suporte da transmissão e distribuição, serviços atrás do medidor e serviços de integração a renováveis. Na **Tabela 1**, apresenta-se as categorias e suas referentes aplicações.

**Tabela 1 –** Categorias e aplicações do BESS no setor elétrico



## **Mineração de Dados**

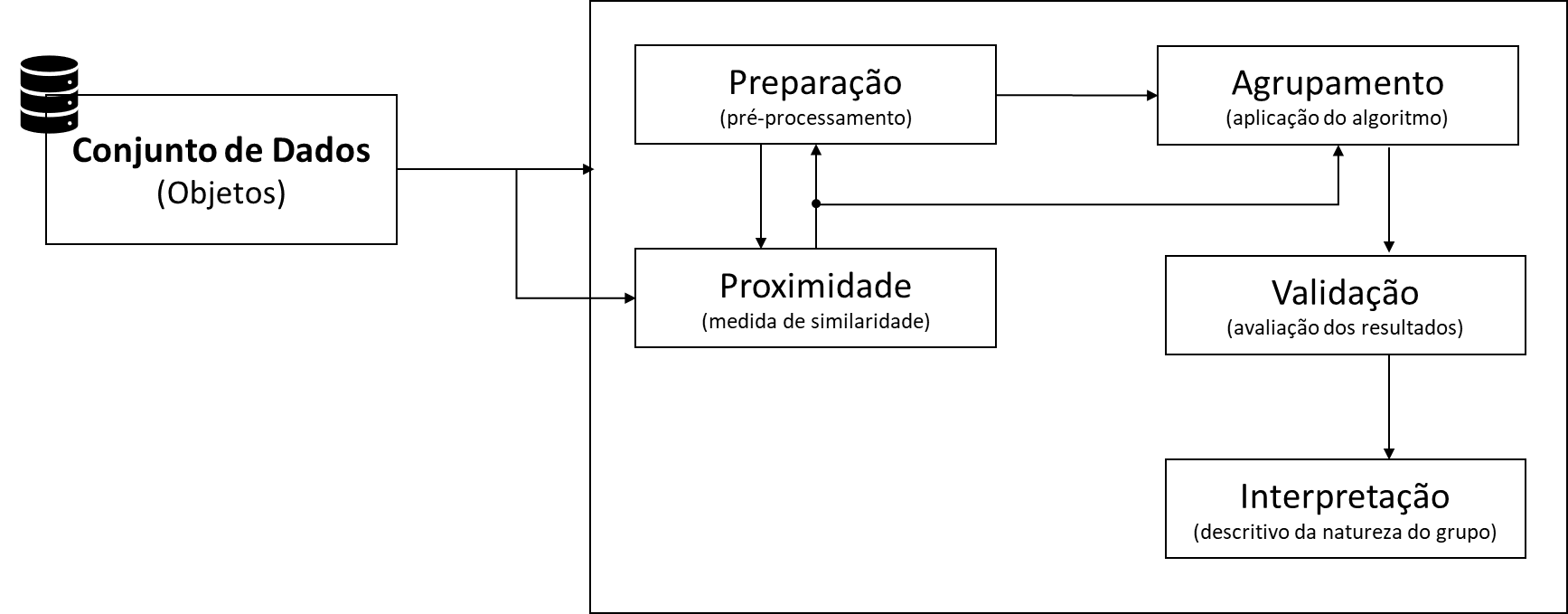
Há diversas metodologias, técnicas de mineração de dados na literatura, como: descrição, classificação, estimação, predição, associação, sumarização, regressão, agrupamento (segmentação) e análise de sequência, mas a que mais se enquadra no perfil do banco de dados que está em análise nesse artigo, é o agrupamento, pois é de grande utilidade na análise exploratória de dados sobre os quais existe pouco ou nenhum conhecimento prévio disponível **[5]**.

## **2.2.1 Definição**

O objetivo de uma técnica de agrupamento é encontrar uma estrutura de grupos (*clusters*) nos dados, em que os atributos pertencentes a cada grupo compartilhem alguma característica relevante para o domínio do problema em estudo **[5]**.

A subjetividade, por conseguinte, está presente em diversos aspectos, entre eles nas hipóteses estabelecidas sobre os dados, a definição da medida de proximidade, a determinação do número de grupos, a seleção do algoritmo de agrupamento e a determinação dos índices de validação **[6]**. Além disso, para o mesmo conjunto de dados, objetivos diferentes geralmente levam a diferentes partições.

O processo de agrupamento compreende diversas etapas que vão desde a preparação dos atributos, até a interpretação dos grupos obtidos. Na **Figura 4**, tem-se o resumo adaptado da construção do modelo, utilizando essa técnica **[7] [8]**.



**Figura 4 -** Etapas de agrupamento de dados

**Etapa de Preparação**: faz-se o pré-processamento (normalizações, conversão de tipos e redução de atributos) e se define a forma apropriada de representação para utilização do algoritmo de agrupamento **[7] [8]**.

**Etapa de Proximidade**: define-se a medida de proximidade apropriada ao domínio da aplicação do conjunto de dados, podendo ser uma similaridade ou dissimilaridade entre os atributos. Em geral, as medidas de similaridade usadas são as medidas de distâncias tradicionais (distância de Hamming, euclidiana, de Manhattan, de Minkowski, de Canberra, de Pearson, do Cosseno, matriz de confusão etc.) **[5]**.

**Etapa de Agrupamento**: aplica-se um algoritmo de agrupamento apropriado para os dados e para o objetivo específico.

**Etapa de Validação**: avalia-se os resultados do agrupamento e se a solução é representativa para o conjunto de dados analisado **[5].**

**Etapa de Interpretação**: examina-se cada grupo com relação aos atributos para rotulá-los, descrevendo a natureza do grupo **[9]**.

Algumas das dificuldades dos algoritmos de agrupamento são **[5][9]**:

* Adequação a conjuntos de dados restritos;
* Restrição dos formatos da estrutura que pode ser encontrada;
* Necessidade de conhecimento prévio do número de clusters presentes nos dados ou o difícil ajuste de parâmetros;
* Instabilidade dos resultados obtidos.

## **2.2.2 Técnicas de Agrupamento**

Os algoritmos de agrupamento podem ser classificados em métodos hierárquicos, particionais, densidade e grade **[10].** Os algoritmos de *clustering* particionais implementam os conceitos de maior interesse desse trabalho. Portanto, a descrição da abordagem desse método é apresentada com maiores detalhes em relação aos outros.

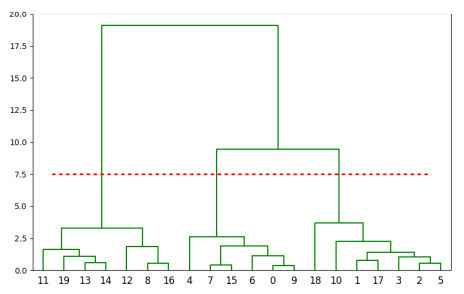
## **Método Hierárquico**

Algoritmos hierárquicos produzem uma sequência de partições rígidas aninhadas, cada uma contendo uma quantidade diferente de grupos. Esses algoritmos podem seguir duas abordagens **[10] [11]**:

**Aglomerativa (*bottom-up*)**: o algoritmo inicia com n grupos, cada um formado por um atributo diferente do conjunto de dados. Em cada passo, os dois grupos mais próximos, segundo algum critério pré-estabelecido, são unidos. Desse modo, o procedimento é repetido até que reste apenas um único grupo contendo todos os atributos.

**Divisiva (*top-down*)**: o algoritmo inicia com apenas um grupo contendo todos os atributos. Em cada passo, algum dos grupos é dividido em dois novos grupos, segundo algum critério pré-estabelecido. Todo o procedimento é realizado até que sejam formados n grupos, cada um contendo apenas um atributo do conjunto de dados original.

Nesse método, por conseguinte, o processo de identificação de grupos é geralmente realimentado recursivamente, utilizando tanto atributos quanto grupos já identificados previamente como entrada para o processamento. Constrói-se, portanto, uma hierarquia de grupos de atributos, no estilo de uma árvore, conhecida como dendrograma que é uma das principais vantagens desse método, **Figura 5** **[12]**.



**Figura 5 -** Exemplo de dendrograma – (i) eixo horizontal: índice de cada atributo; (ii) eixo vertical: valor de distância quando cada par de objetos foi unido; (iii) linha horizontal pontilhada – possível corte. Fonte: edisciplinas.usp.br

Os métodos hierárquicos não requerem que seja definido um número *a priori* de grupos. Por meio da análise do dendograma, pode-se inferir no número de agrupamentos adequados.

## **Método Particional ou de Otimização**

Um agrupamento particional é uma divisão do conjunto de objetos de dados em subconjuntos (grupos) não interseccionados de modo que cada objeto de dado esteja exatamente em um subconjunto.

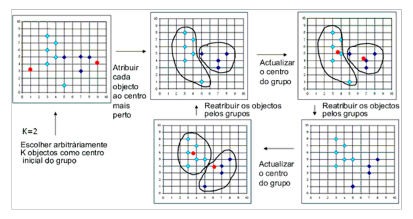
Dado um conjunto D de dados com n registros e k o número de agrupamentos desejados, os algoritmos de particionamento organizam os objetos em k agrupamentos, tal que k <= n. Os algoritmos mais comuns de agrupamento são: K-Means e K-Medoids.

**Algoritmo K-means**

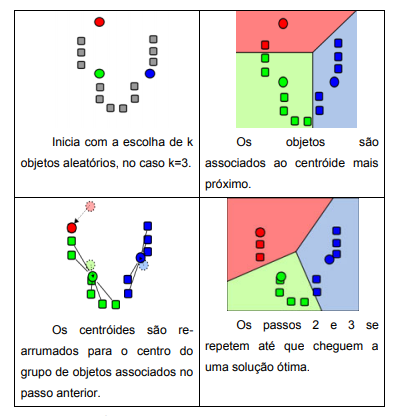
O algoritmo K-Means apresentado em 1967 **[13][14]** também chamado de K-Médias é uma das técnicas de agrupamento particional mais populares, justificado por sua simplicidade e facilidade de implementar em linguagens computacionais.

A concepção da técnica K-Means é fornecer uma classificação de informações de acordo com os próprios dados. Essa classificação, baseia-se em análise e comparações entre os valores numéricos dos dados. Assim, o algoritmo automaticamente vai fornecer uma classificação automática sem a necessidade de nenhuma supervisão humana, ou seja, sem nenhuma pré-classificação existente.

Além disso, norteia-se a técnica por meio de protótipos (centros). O K-Means define um protótipo em termos de um centróide, que normalmente corresponde à média dos padrões em um grupo. K-Means inicia escolhendo K centróides iniciais, em que K é um parâmetro definido pelo usuário por meio do método de Elbow **[15]**, que representa o número de grupos desejados. Cada padrão é então atribuído ao centróide mais próximo, e cada coleção de padrões atribuída ao centróide forma um grupo. O centróide de cada grupo é então atualizado baseado nos padrões atribuídos ao grupo. O processo de atribuição dos padrões e a atualização dos centróides se repete até que os centróides permaneçam inalterados, **Figura 6** e **Figura 7.**



**Figura 6 -** Processo de aplicação da técnica K-means. Fonte: https:goo.gl/G1NNyF



**Figura 7 -** Exemplificação da técnica K-means. Fonte: PUC-Rio

**Algoritmo K-medoids**

Esse algoritmo é uma variação do algoritmo K-means e, ao contrário desse, o K-medoids escolhe atributos existentes como centróides. Ao final do agrupamento, obtém-se um atributo como o elemento central, normalmente, classificado como o protótipo do agrupamento, o medóide.

Uma vantagem desse algoritmo em relação ao K-means é sobre os ruídos ou *outliers* contidos no agrupamento, pois as estratégias na escolha do centróide e do medóide são diferentes. No K-means, o centróide é dado pela média de todos os atributos dentro de um agrupamento. Dessa maneira, se o agrupamento possuir um objeto muito distante dos outros, o centróide será influenciado erroneamente. Já no K-medoids, o algoritmo utiliza a média do erro quadrado para validar a escolha de um medóide então ao escolher um medóide que esteja mais próximo a um ruído ou *outlier*, a média do erro quadrado irá aumentar e esta escolha será descartada, dado que o objetivo é minimizar a média do erro quadrado. Com isso, o algoritmo obtém o elemento mais representativo ou central do agrupamento.

## **Método Baseado em Densidade**

Os métodos de particionamento e hierárquicos geram agrupamentos de formato esféricos (distribuição dos valores dos dados é mais esparsa). No entanto, existem situações em que essa distribuição é mais densa e que tais métodos não apresentam resultados satisfatórios. Os métodos baseados na densidade conseguem melhores resultados. Destacamos os algoritmos: DBSCAN (*A Density-Based Clustering Method Based on Connected Regions with Sufficiently High Density),* OPTICS (*Ordering Points to Identify the Clustering Structure*) e DENCLUE (*DENsity-based CLUstEring*). Em **[12]**, é proposta uma técnica usando a estratégia do *Simulated Annealing*.

## **Método Baseado em Grade**

Métodos baseado em grades, utilizam-se da estrutura de grades e dividem os registros nas células desta grade. Apresentam um tempo de processamento bem rápido. Os principais algoritmos são o STING (*STatistical INformation Grid*) e o WaveCluster (*Clustering UsingWavelet Transformation*).

Apesar de cada método possuir suas peculiaridades e apresentar melhor resultado com um certo tipo de dado, não existe uma classificação única para a escolha e aplicação destes métodos **[16]**.

Para esse desenvolvimento, escolheu-se a metodologia particional e o algoritmo *K-means*.

# **MATERIAIS E MÉTODOS**

Dando continuidade a metodologia CRISP-DM, nesse item será descrito como foi o decorrer da fase de Entendimento e Preparação do Dados **[4]**.

## **Descrição da Base de Dados**

Nesse artigo, utilizou-se dados reais de 71 dias de operação de um sistema de armazenamento de energia por baterias de chumbo-carbono do ITEMM no consumidor final cuja finalidade é fornecer qualidade de energia e segurança energética. Esse sistema está instalado em Belo Jardim/PE, localizado na unidade matriz do Instituto. Por ser um produto em desenvolvimento, não é possível compartilhar o banco de dados.

Trata-se de dados extraídos do banco de dados do EMS do ITEMM, especificamente do BMS, em que são coletadas todas as informações do banco de baterias. Nesse período de composição do banco, tem-se 23.939 instâncias e 1.208 atributos¹.

A base de dados desse artigo, portanto, fornece informações reais do estado de carga das baterias e das medições instantâneas de tensão, corrente e temperatura, tanto do banco de baterias, como de cada célula do sistema. Na **Tabela 2**, apresenta-se o dicionário de todos os atributos de leitura do BMS.

**Tabela 2 -** Dicionário de atributos do banco de dados do BMS do BESS de PbC

|  |  |  |
| --- | --- | --- |
| **Atributo** | **Descrição** | **Domínio** |
| Timestamp | Tempo de coleta de dado | timestamp |
| Bstring1\_Battery\_Voltage | Tensão Total | float64 |
| Bstring1\_Battery\_Current | Corrente Total | float64 |
| Bstring1\_Battery\_Soc | SoC Calculado | float64 |
| Bstring1\_Battery\_Soh | SOH Calculado | int64 |
| Bstring1\_Statistic\_VoltageMean | Tensão Média | float64 |
| Bstring1\_Statistic\_VoltageMax | Tensão Máxima | float64 |
| Bstring1\_Statistic\_VoltageMin | Tensão Mínima | float64 |
| Bstring1\_Statistic\_TemperatureMean | Temperatura Média | float64 |
| Bstring1\_Statistic\_TemperatureMax | Temperatura Máxima | int64 |
| Bstring1\_Statistic\_TemperatureMin | Temperatura Mínima | int64 |
| Bstring1\_Bmu\_Voltage\_0 a 287² | Tensão Célula 0 a 287 | float64 |
| Bstring1\_Bmu\_Temperature\_0 a 287² | Temperatura Célula 0 a 287 | int64 |

## **Análise Descritiva dos Dados**

Ao analisar os dados reais de operação extraídos pelo BMS, identificou-se, juntamente com os *stakeholders* que alguns atributos são considerados essenciais para a análise do estado sistêmico. Na **Tabela 3**, apresenta-se a distribuição dos atributos fundamentais para análise e mapeamentos da mineração de dados.

**Tabela 3 –** Descritivo da quantidade de Atributos

|  |  |
| --- | --- |
| **Dado** | **Quantidade de Atributos** |
| Tempo de coleta de dado | 1 |
| Tensão Total | 1 |
| Corrente Total | 1 |
| SoC Calculado | 1 |
| Tensão Célula | 288 |
| Temperatura Célula | 288 |

Além disso, foi necessário definir as condições de contorno permitido de operação desses atributos, garantindo, portanto, a operabilidade, segurança e funcionamento ótimo do sistema. Na **Tabela 4 e Tabela 5**, pode-se observar o intervalo de operação de cada atributo definido e o de operação, respectivamente.

**Tabela 4 –** Definição dos intervalos de operação de cada atributo

|  |  |
| --- | --- |
| **Dado** | **Intervalo de Operação** |
| Tensão Total (V) | [510; 730] |
| Corrente Total (A) | [-100; 100] |
| SoC Calculado (%) | [20; 90] |
| Tensão Célula (V) | [1,75; 2,7] |
| Temperatura Célula (°C) | [15; 50] |

**Tabela 5 –** Intervalos dos dados de operação de cada atributo do banco de dados

|  |  |
| --- | --- |
| **Dado** | **Intervalo de Operação** |
| Tensão Total (V) | [680,79; 556,79] |
| Corrente (A) | [116,20; -79,09] |
| SoC Calculado (%) | [97,40; 38,20] |
| Tensão Célula (V) | [2,43;1,99] |
| Temperatura Média (°C) | [25,02;23,34] |

## **Pré-processamento dos Dados**

Com o objetivo de encontrar um conjunto mínimo de atributos, de tal forma a preservar a informação original, ao analisar o banco de dados, identificou-se a necessidade de estruturar o pré-processamento dos dados. A técnica definida foi a de redução vertical, já que elimina características irrelevantes, consequentemente se, reduz a dimensionalidade e os ruídos.

Com isso, os atributos com valores que se mantinham constantes, bem como os que não foram selecionados como essenciais foram removidos. Após o pré-processamento desses dados, desconsiderou-se atributos com valores constantes e/ou irrelevantes, consolidando para processamento, portanto, um total de 587 atributos e 23.939 instâncias.

Dentre esses atributos, os que serão avaliados nesse documento são: (i) SoC Calculado, (ii) Tensão Total, (iii) Corrente Total, (iv) Tensão das Células e (v) Temperatura das Células. Em todos esses atributos, distribui-se os dados em 5 grupos cujo objetivo foi facilitar a interpretação dos dados.

Na distribuição dos dados, por conseguinte, utilizou-se os limites inferiores e superiores de apresentados nas instâncias para encontrar os limites dos grupos, bem como o ponto médio, conforme equação **(1)**:

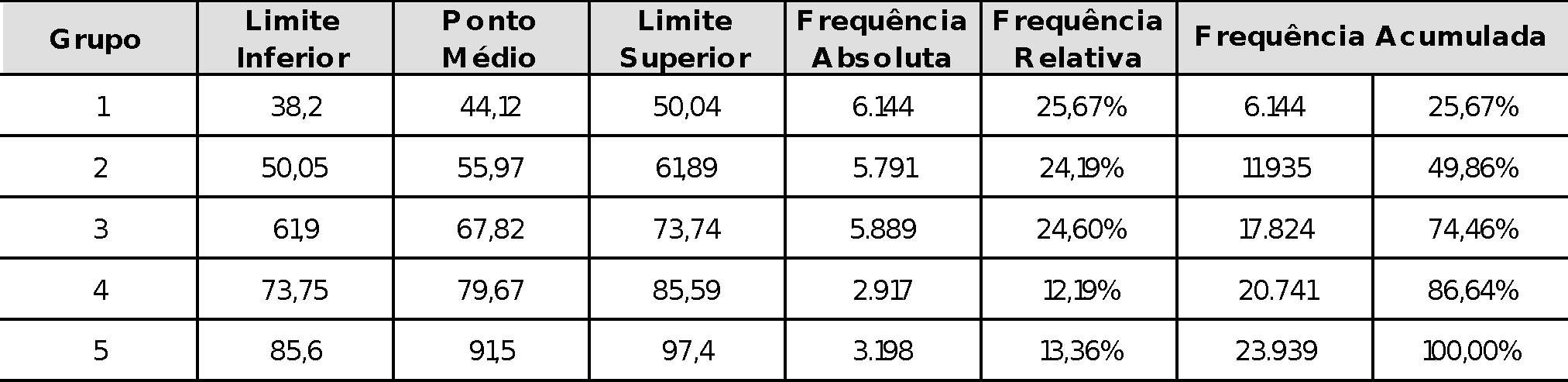
em que, é o valor da amostra, o valor máximo superior e, por fim, o o mínimo inferior. Além disso, ainda no pré-processamento dos dados, fez-se análises gráficas de cada atributo, conforme detalhado nos próximos itens.

1. **SoC Calculado – “*Bstring1\_Battery\_SoC”***

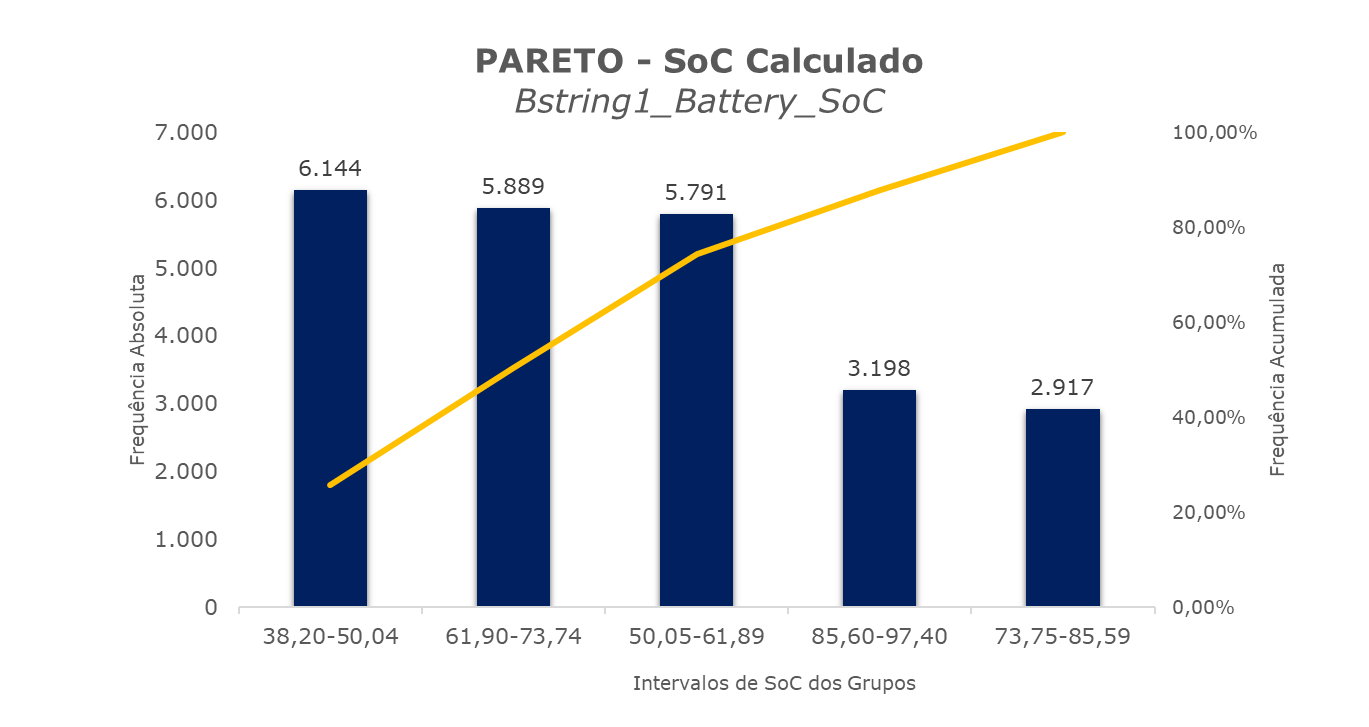
Atributo relevante para a avaliação da capacidade de um acumulador de energia, suportando decisões essenciais em baterias e sistemas de baterias, como é o caso do BESS.

Na **Tabela 6**, apresenta-se a distribuição dos dados do SoC Calculado. Percebe, assim, que a maior frequência de operação do SoC está no Grupo 1 e de menor no Grupo 4, 25,67% e 12,19%, respectivamente.

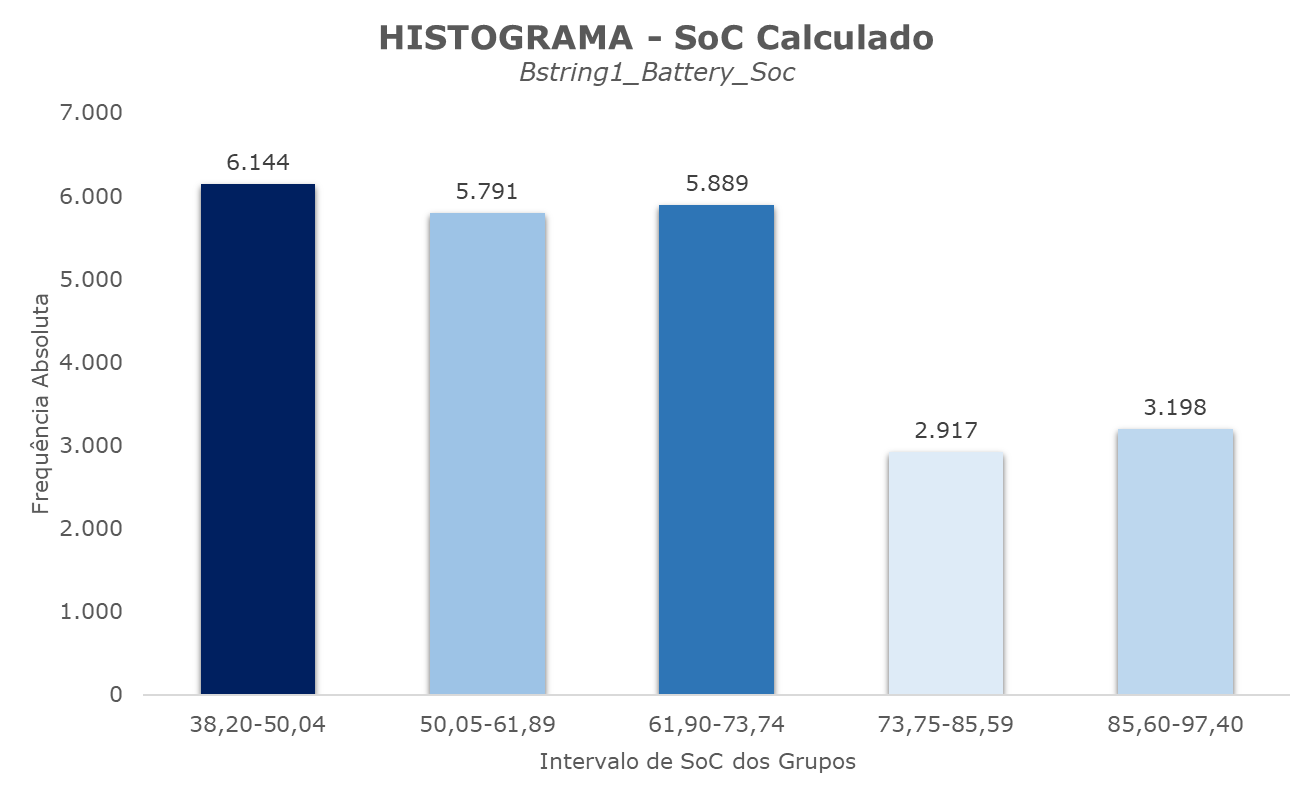
**Tabela 6 -** Distribuição de Frequência do Atributo SoC Calculado - *Bstring1\_Battery\_SoC*"



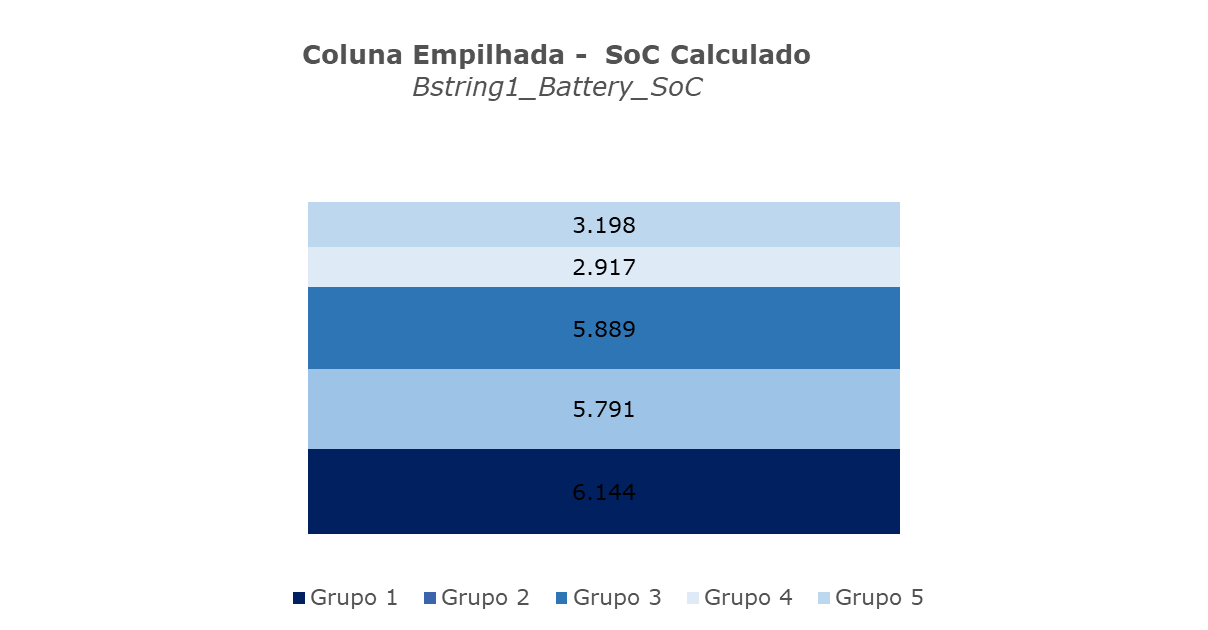
Nas **Figuras 8, 9, 10** e **11**, pode-se observar a análise gráfica desse atributo, ratificando a maior frequência no Grupo 1 e a menor no Grupo 4.



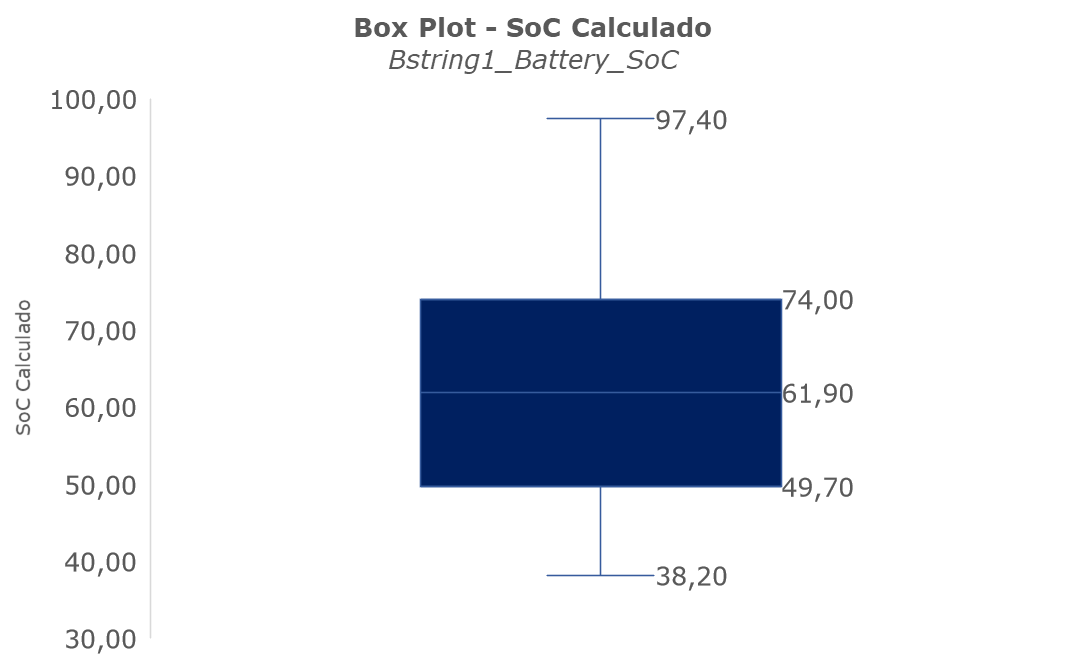
**Figura 8 -** Pareto do Atributo SoC Calculado



**Figura 09 -** Histograma do Atributo SoC Calculado



**Figura 10 –** Coluna Empilhada do Atributo SoC Calculado



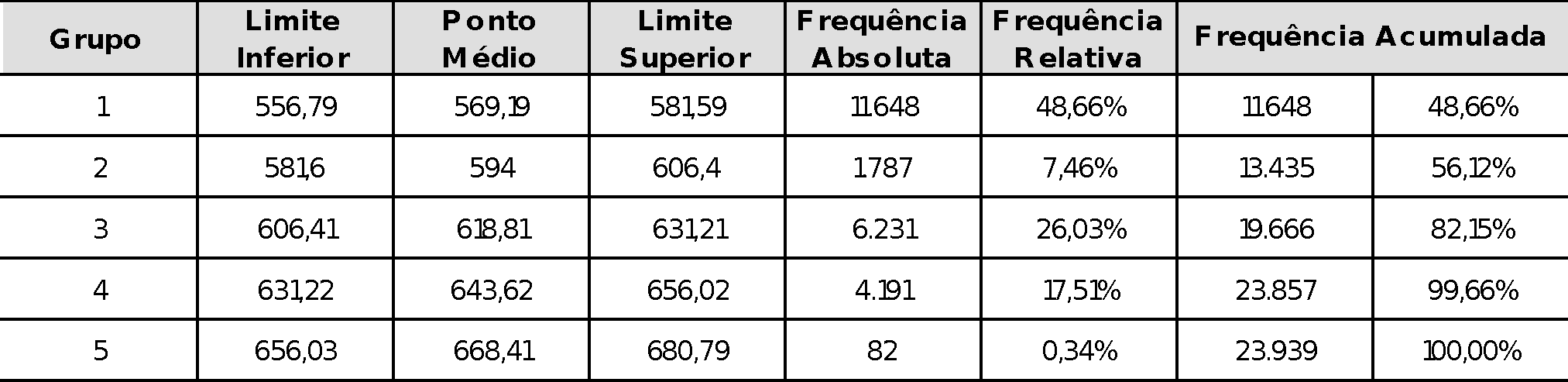
**Figura 11 –** Gráfico Box Plot do Atributo SoC Calculado

Em suma, pode-se afirmar que o sistema em análise possui um comportamento arrojado, pois está operando no SoC mínimo no maior intervalo de tempo.

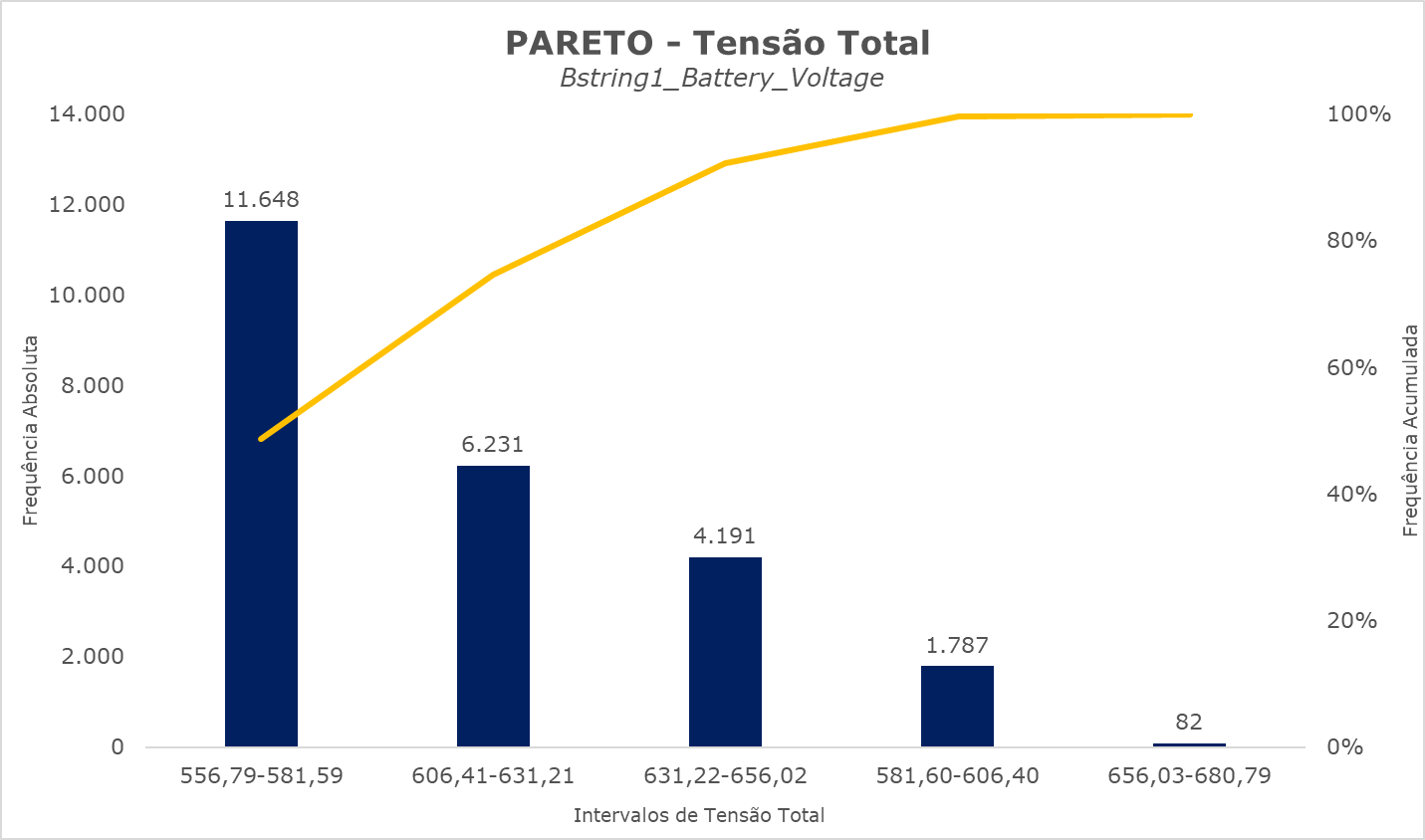
1. **Tensão Total – “*Bstring1\_Battery\_Voltage”***

Atributo essencial para operabilidade adequada do sistema de armazenamento de energia por baterias, já que garantem as aplicações efetivas do sistema. Analogamente, na **Tabela 7**, apresenta-se a distribuição desse atributo.

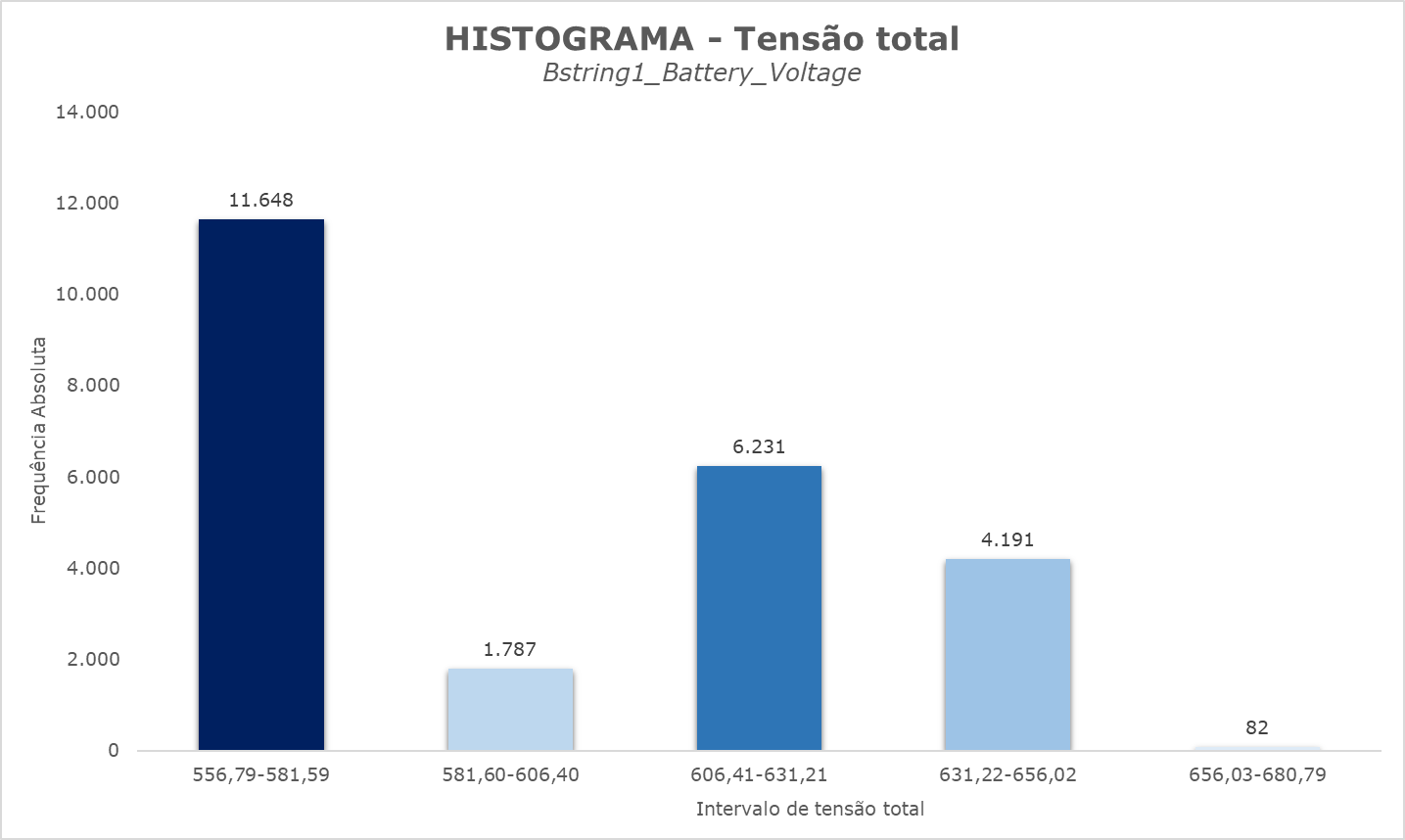
**Tabela 7 -** Distribuição de Frequência do Atributo Tensão Total -"*Bstring1\_Battery\_Voltage*"



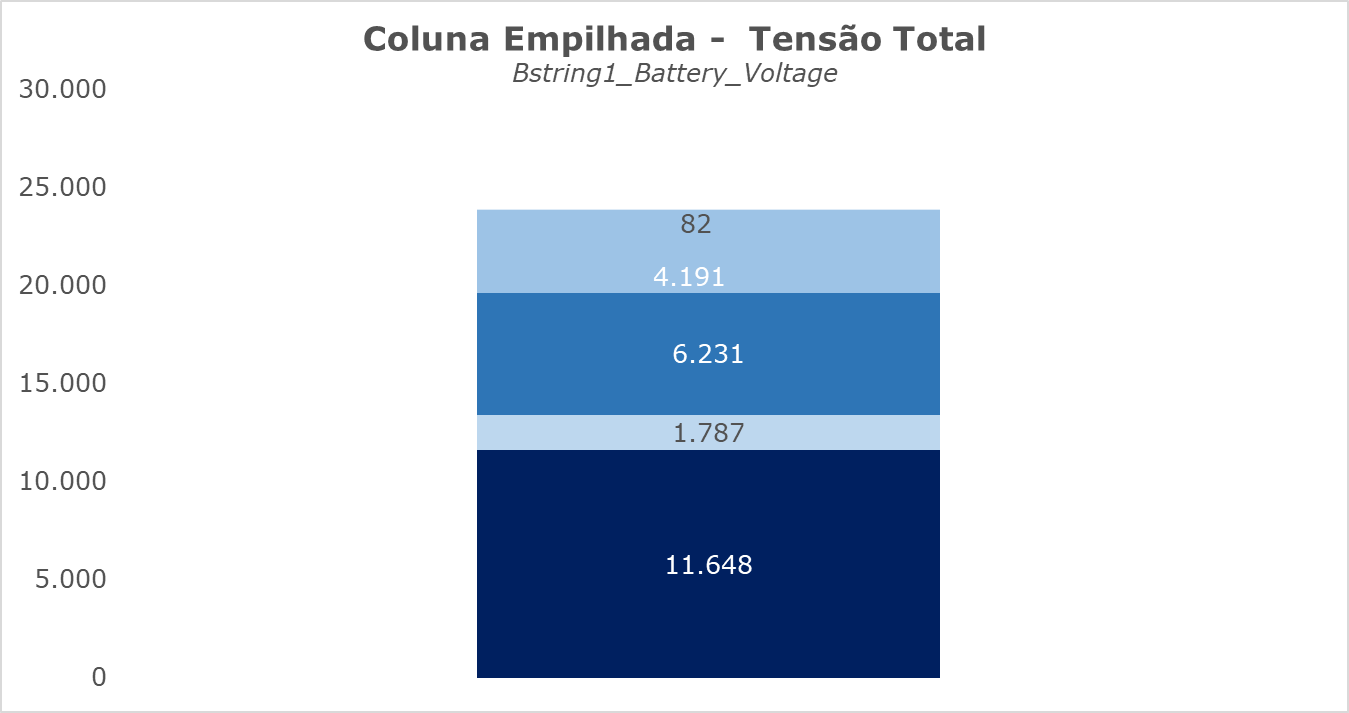
Diante do explanado, a maior frequência de operação da Tensão Total está no Grupo 1 e de menor no Grupo 5, 48,66% e 0,34%, respectivamente. Garantindo, portanto, que o sistema não trabalha em sobretensão. Abaixo, semelhantemente, serão apresentadas as análises gráficas, **Figuras 12, 13, 14** e **15**.



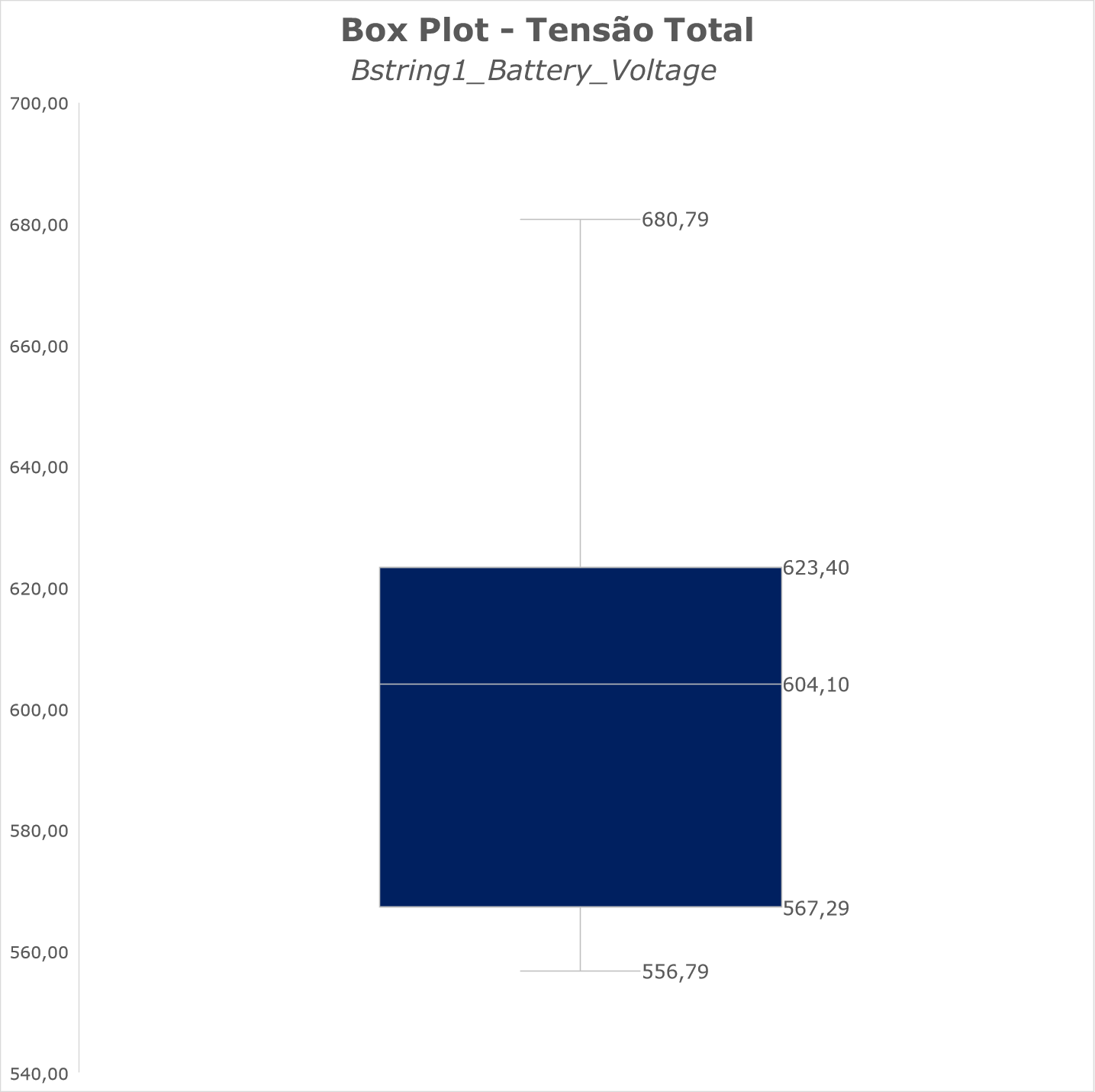
**Figura 12 -** Pareto do Atributo Tensão Total



**Figura 13 -** Histograma do Atributo Tensão Total



**Figura 14 -** Coluna Empilhada do Atributo Tensão Total

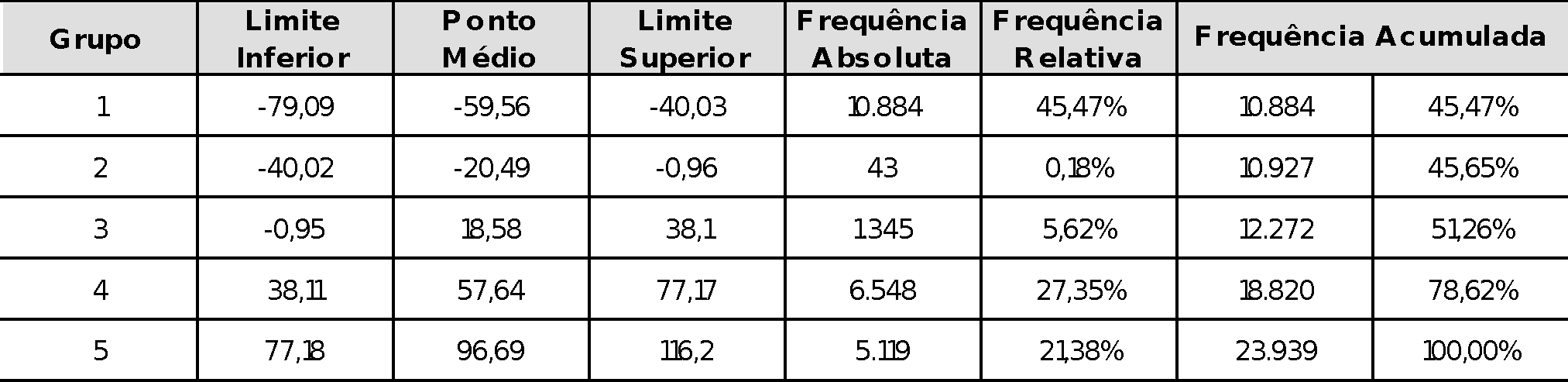


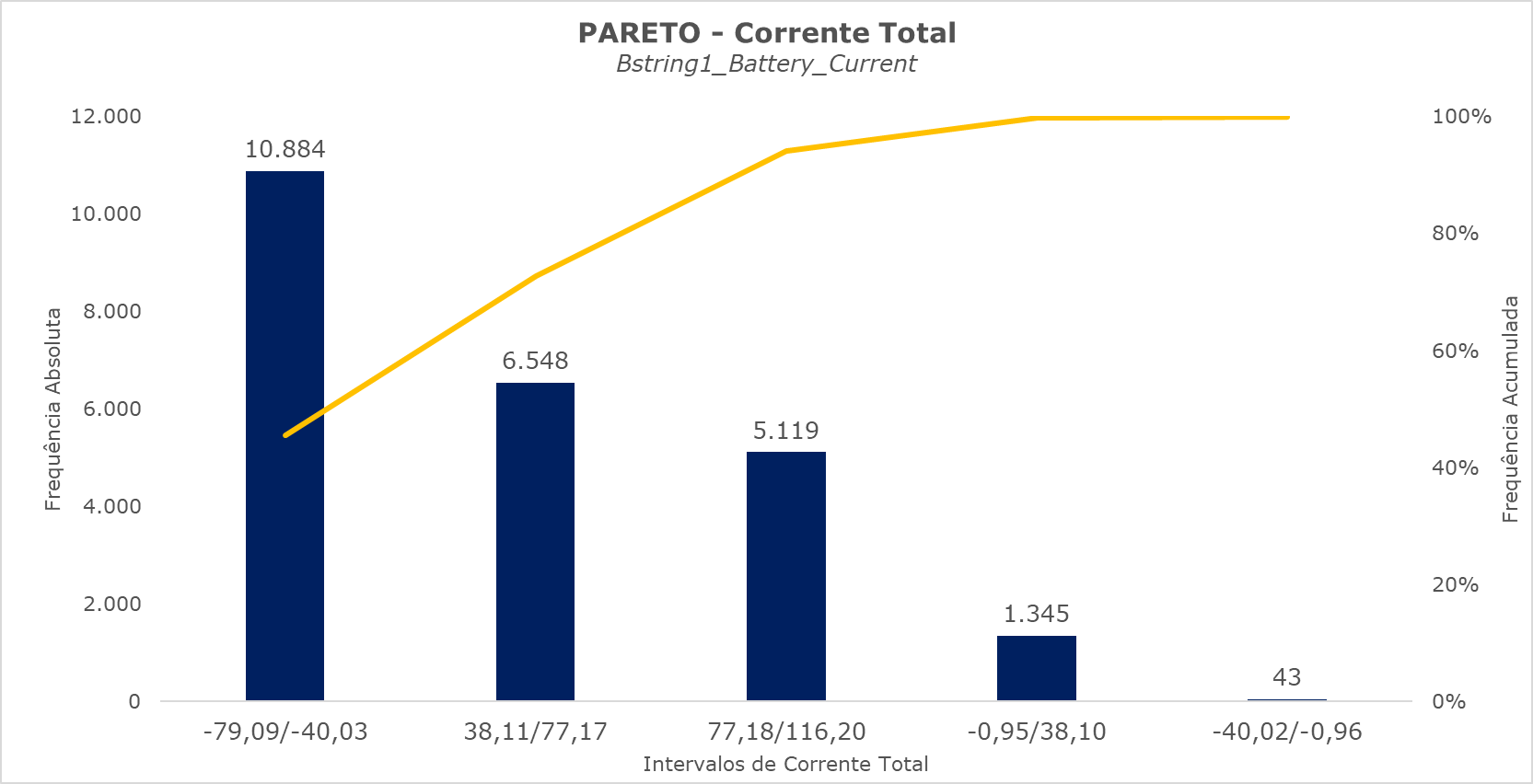
**Figura 15** - Gráfico Box Plot do Atributo Tensão Total

1. **Corrente Total – “*Bstring1\_Battery\_Current”***

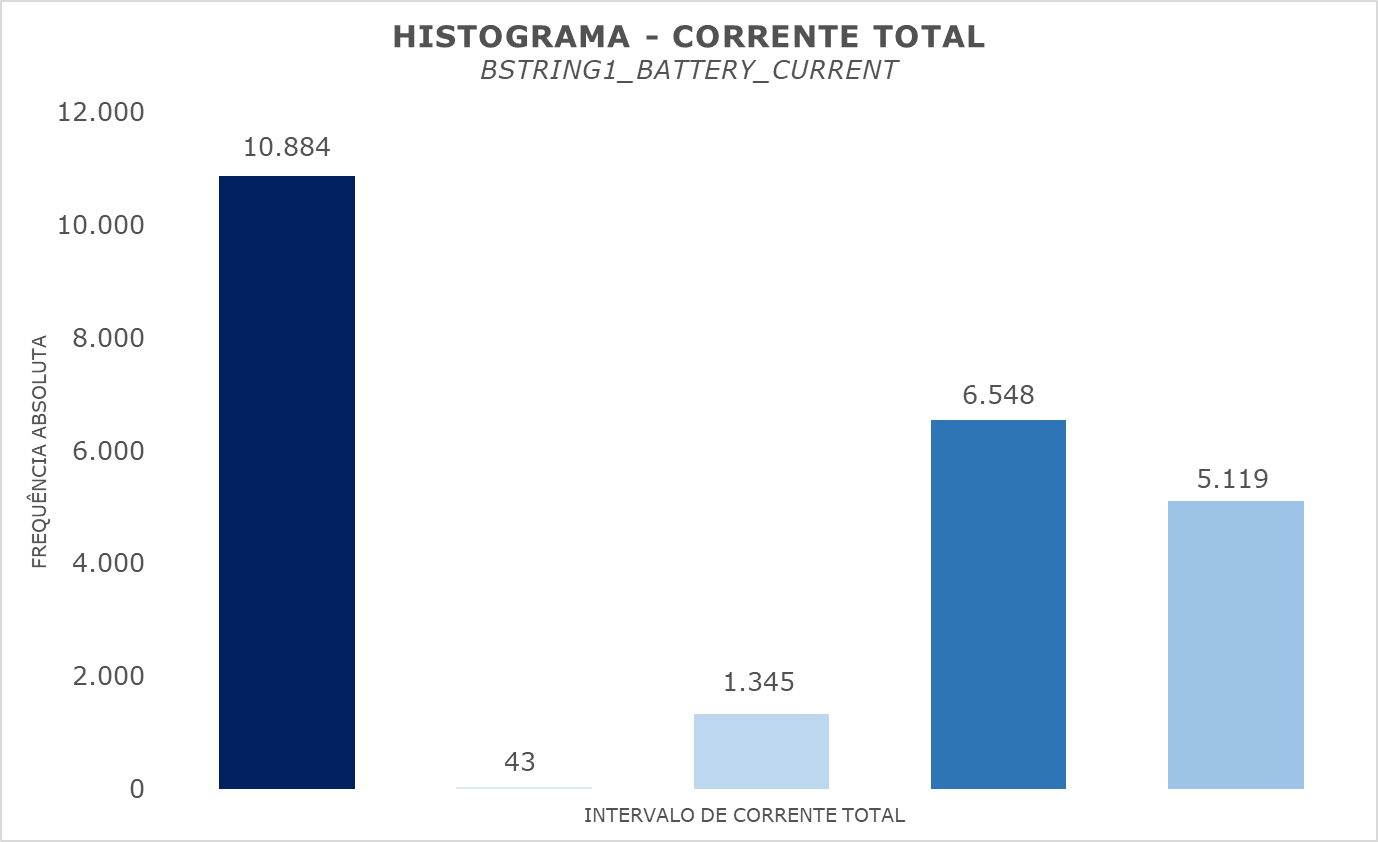
Equivalente a tensão, a corrente também garante a operabilidade efetiva do BESS. Se estiver fora do intervalo permitido, o sistema de monitoramento deve desarmar, por motivos de segurança. A **Tabela 8** e as **Figuras de 16** a **19**, por conseguinte, detalham a distribuição dos dados medidos pelo BMS. Os valores negativos indicam que o sistema está descarregando, consequentemente, positivos recarregando.

**Tabela 8 -** Distribuição de Frequência do Atributo Corrente Total - "*Bstring1\_Battery\_Current*"

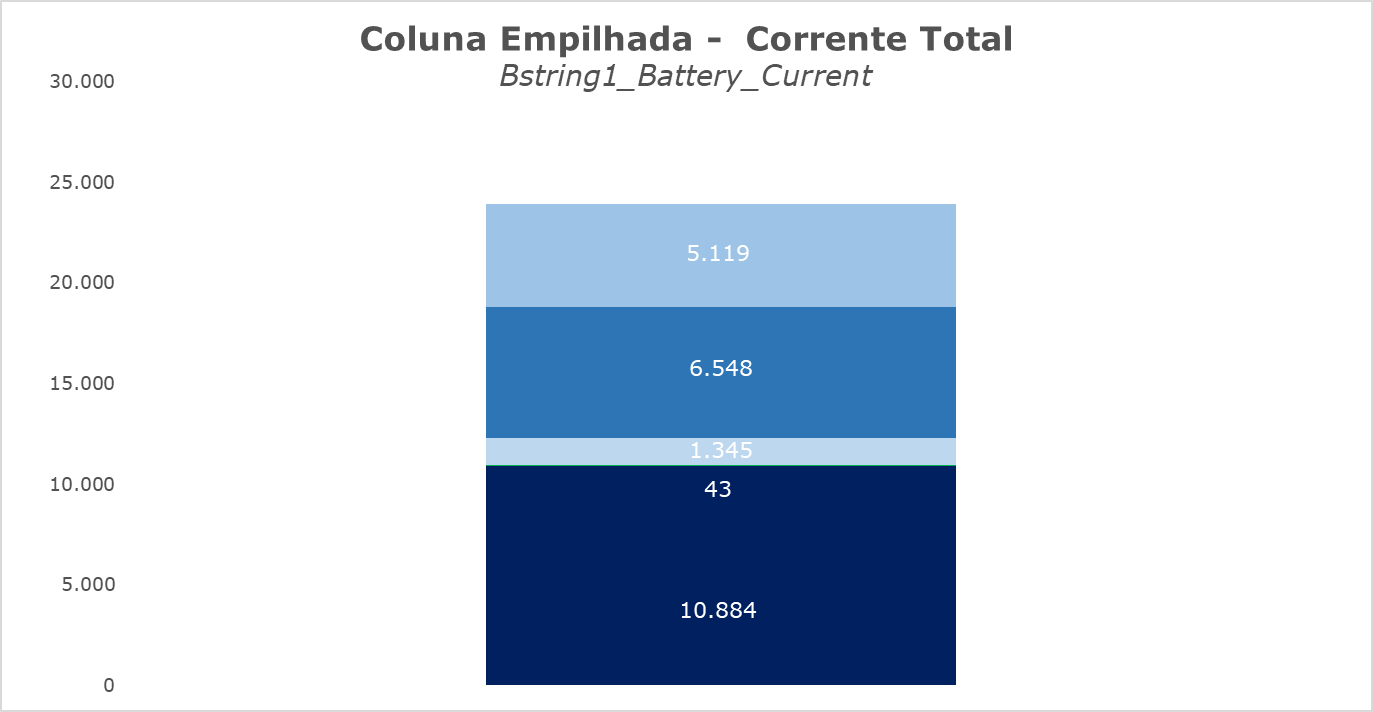




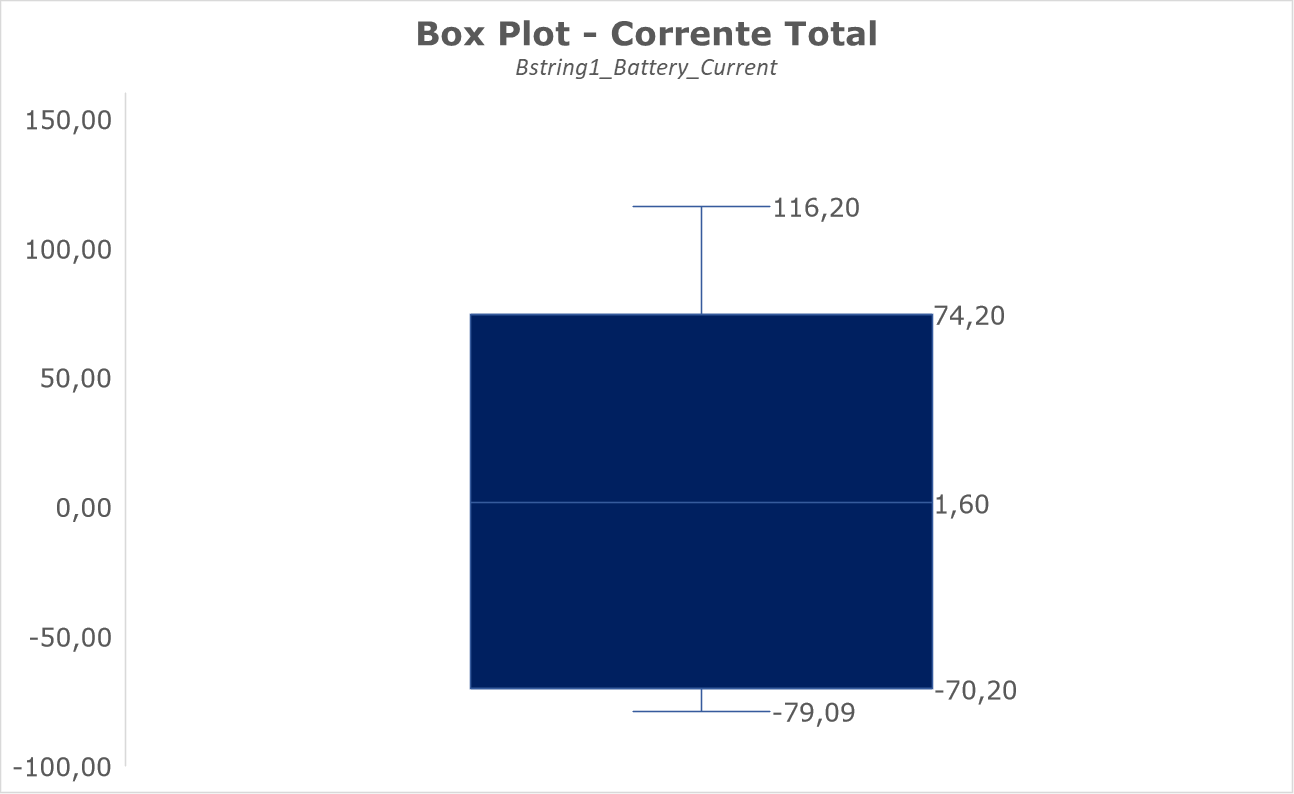
**Figura 16 –** Pareto do Atributo Corrente Total



**Figura 17** –Histograma do Atributo Corrente Total



**Figura 18** – Coluna Empilhada do Atributo Corrente Total



**Figura 19 –** Gráfico Box Plot do Atributo Corrente Total

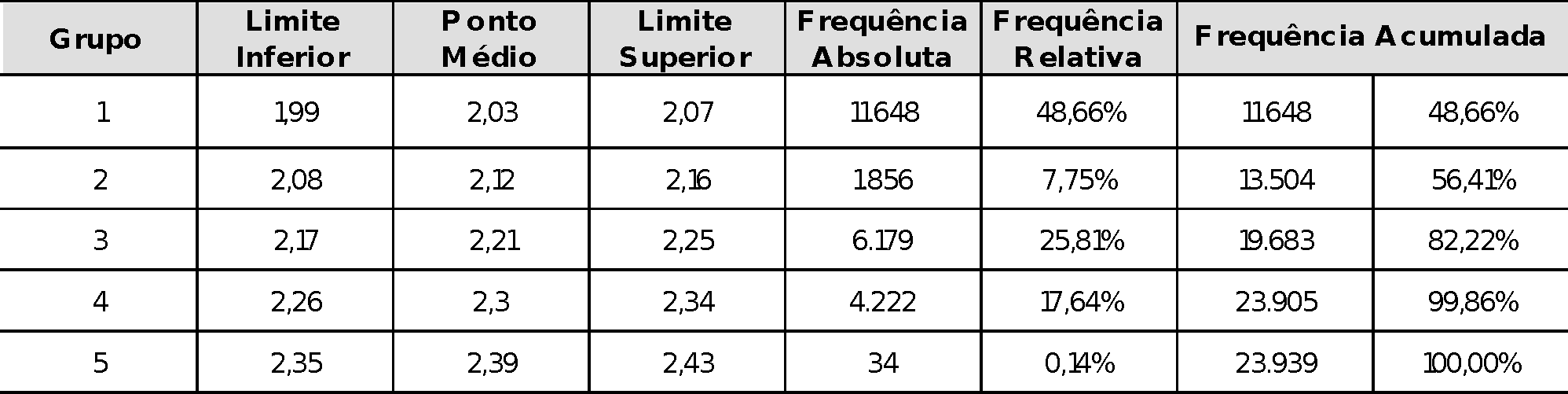
Indica-se, assim, que na maior parte do tempo, o BESS está descarregando com corrente elevada (demanda intensa da carga), já que a maior frequência de uso está no Grupo 1, 45,47%, em que os limites de corrente são o mais negativo dentre os grupos [-79,09; -40,03].

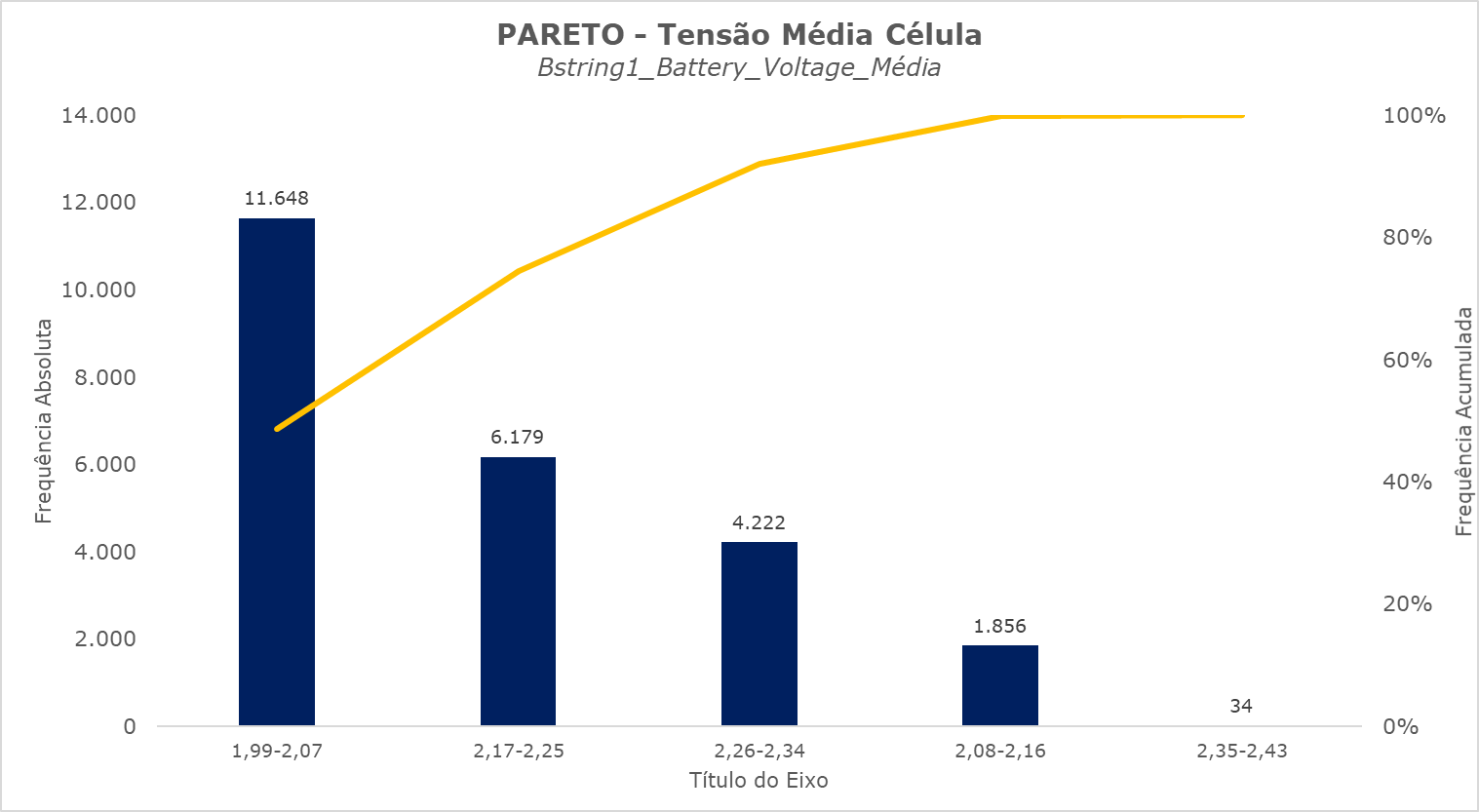
1. **Tensão Célula – “*Bstring1\_Bmu\_Voltage\_n”***

Neste atributo, podemos avaliar a média das tensões individuais de cada célula, igualmente essencial para operação adequada do BESS, mas, nesse caso, há como se observar o comportamento individual e agir de maneira preventiva ou corretiva, pontualmente na célula defeituosa. Utilizou-se média, pois a diferença máxima entre a mediana e a moda das 287 células é pouco representativa, 0,008 e 0,01, respectivamente.

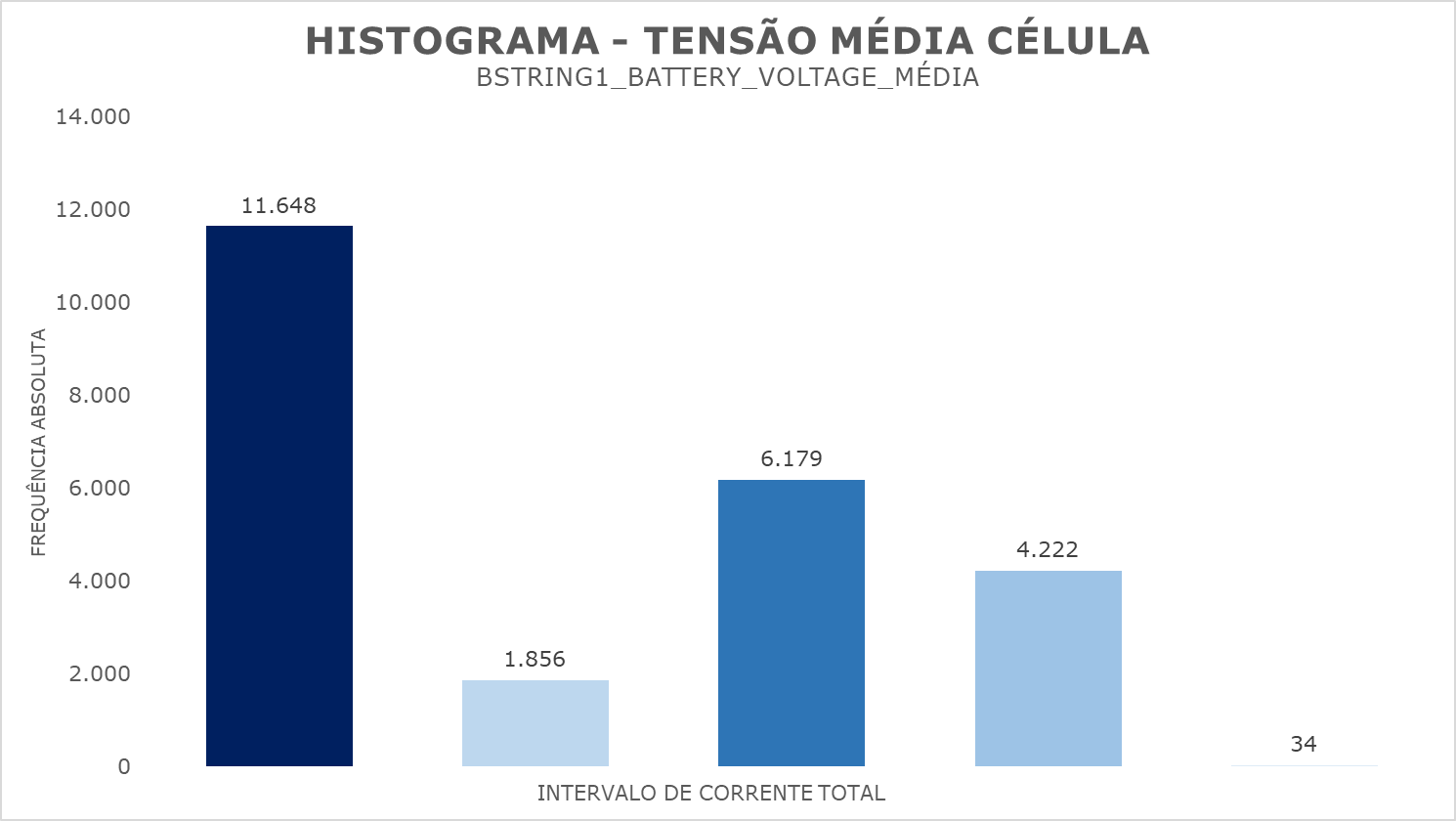
Na **Tabela 9**, observa-se a distribuição média das células do BESS. Percebe, assim, que a maior frequência está no Grupo 1 e de menor no Grupo 4, 48,66% e 0,14%, respectivamente. Comprovando a sinergia entre os outros atributos essenciais, como SoC e Tensão Total.

**Tabela 9** - Distribuição de Frequência do Atributo Tensão Média Célula - "*Bstring1\_BMU\_Voltage\_Média*"

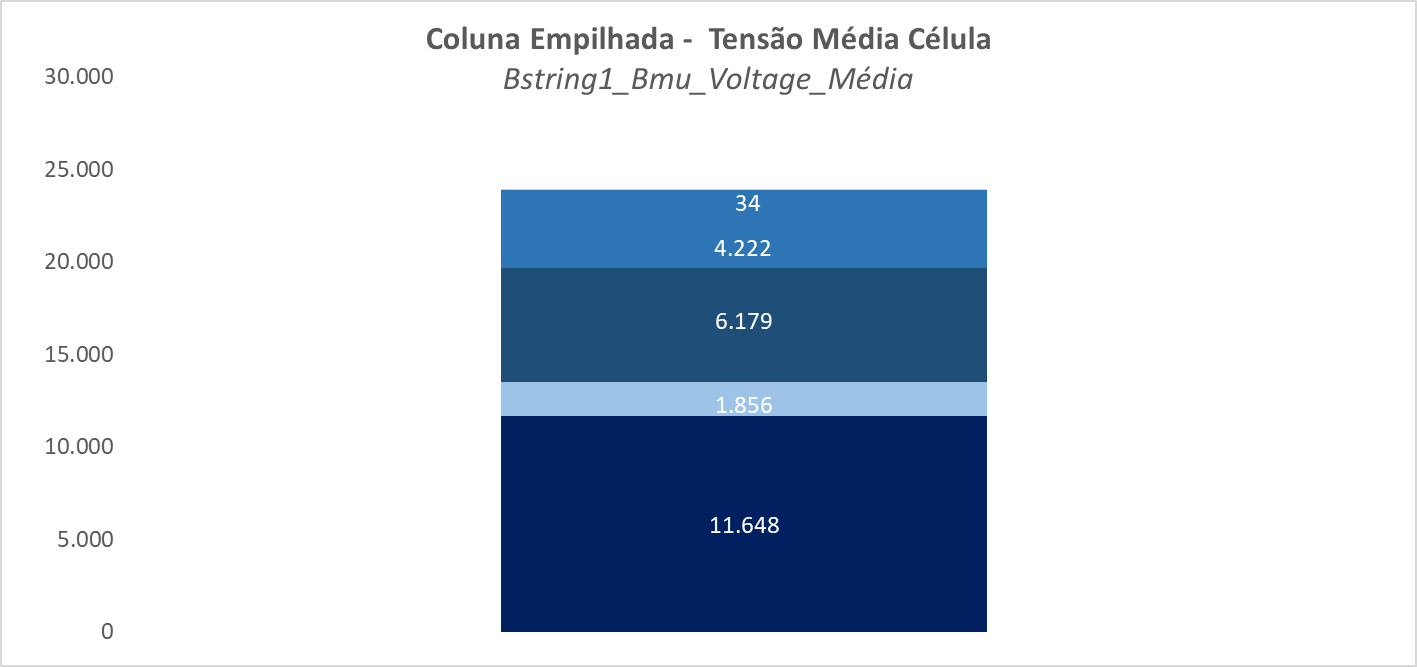




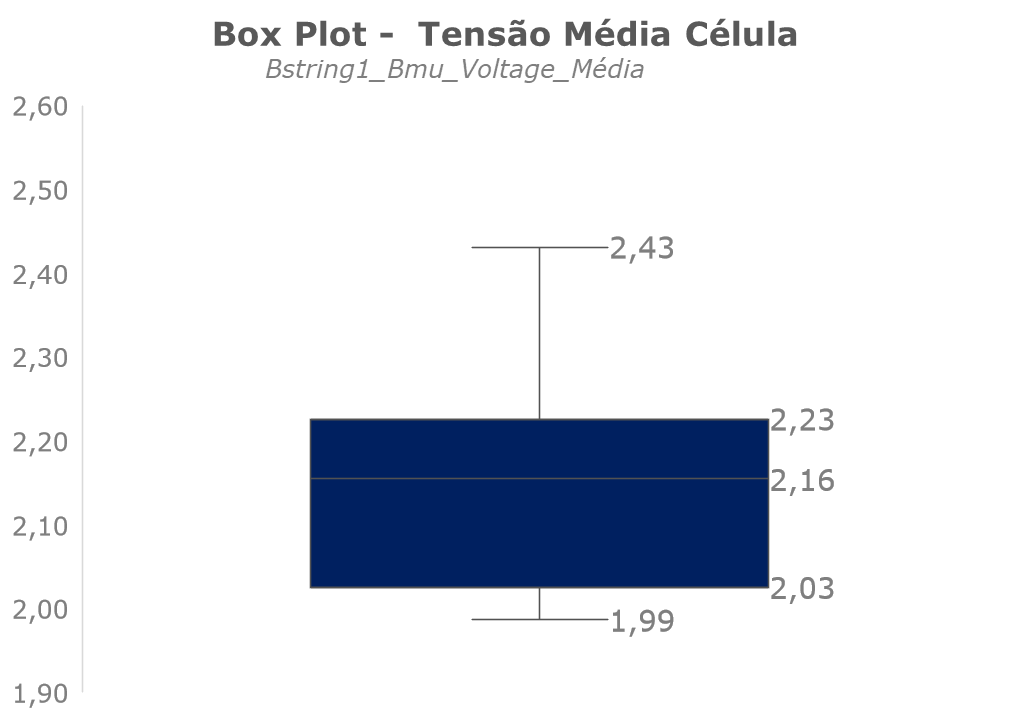
**Figura 20 –** Pareto do Atributo Tensão Média Célula



**Figura 21** –Histograma do Atributo Tensão Média Célula



**Figura 22** –Coluna Empilhada do Atributo Tensão Média Célula



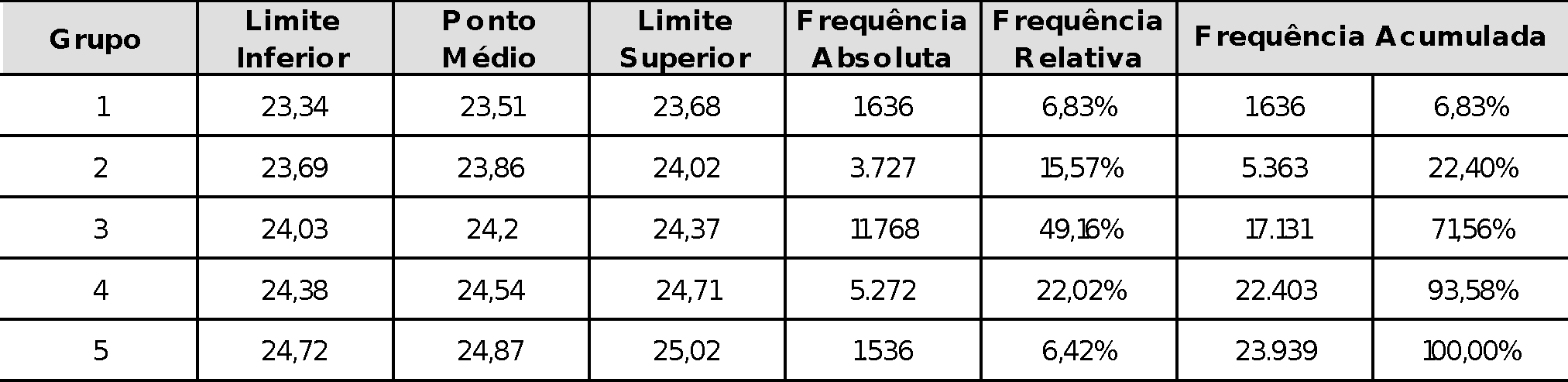
**Figura 23 –** Gráfico Box Plot do Atributo Tensão Média Célula

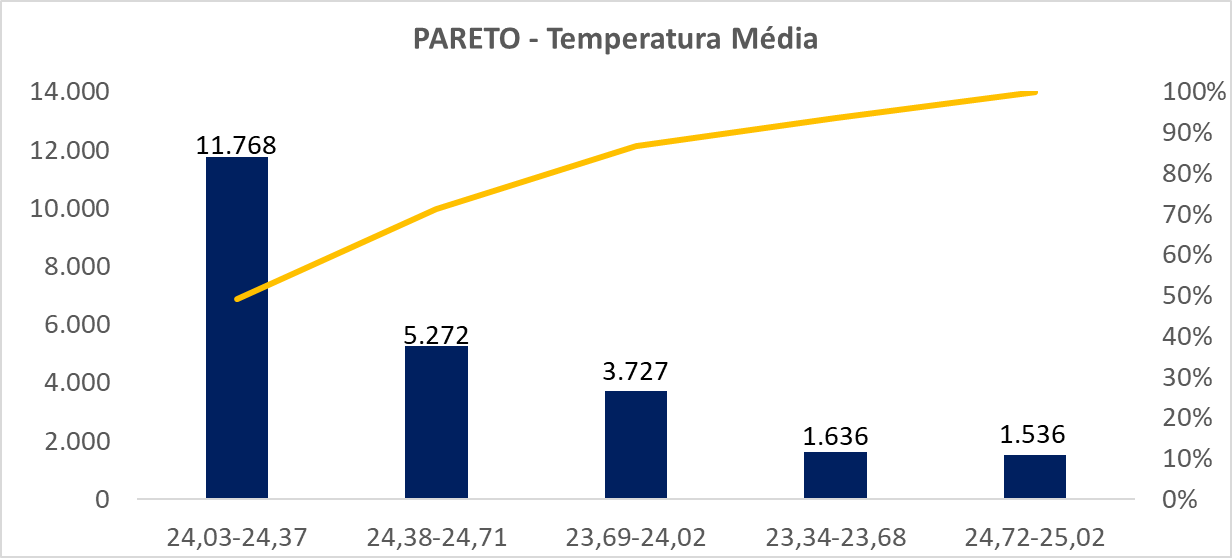
1. **Temperatura Célula**

É um atributo essencial, pois as baterias possuem um limite de operação de temperatura. Se há uma extrapolação da temperatura, pode-se acontecer acidentes não planejados e estimados. Diferentemente da tensão, utilizou-se a mediana das células.

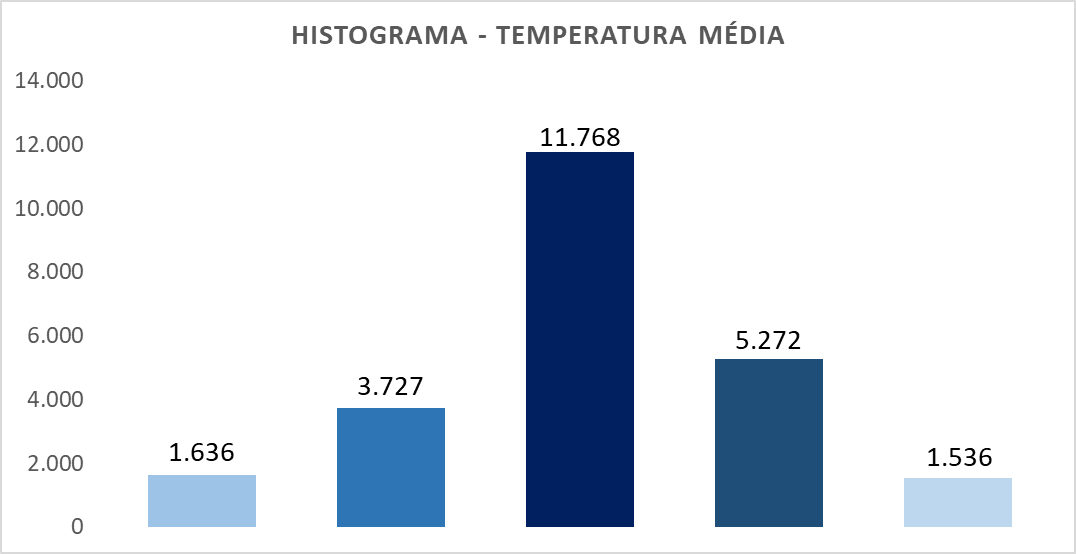
A **Tabela 10** e os gráficos entre as **Figuras 24** e **27** representam que não há muita variação da temperatura nesse dia de operação. Provavelmente, é um dia de inverno. Ver-se, também, que o grupo de maior frequência é o 3 e o menor 5.

**Tabela 10** – Distribuição de Frequência do Atributo Temperatura Média Célula

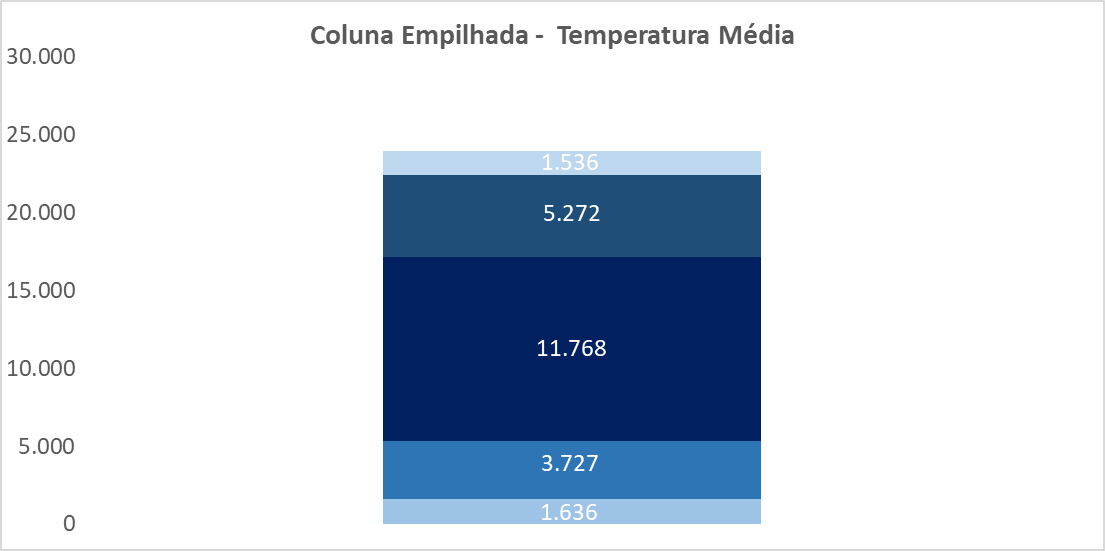




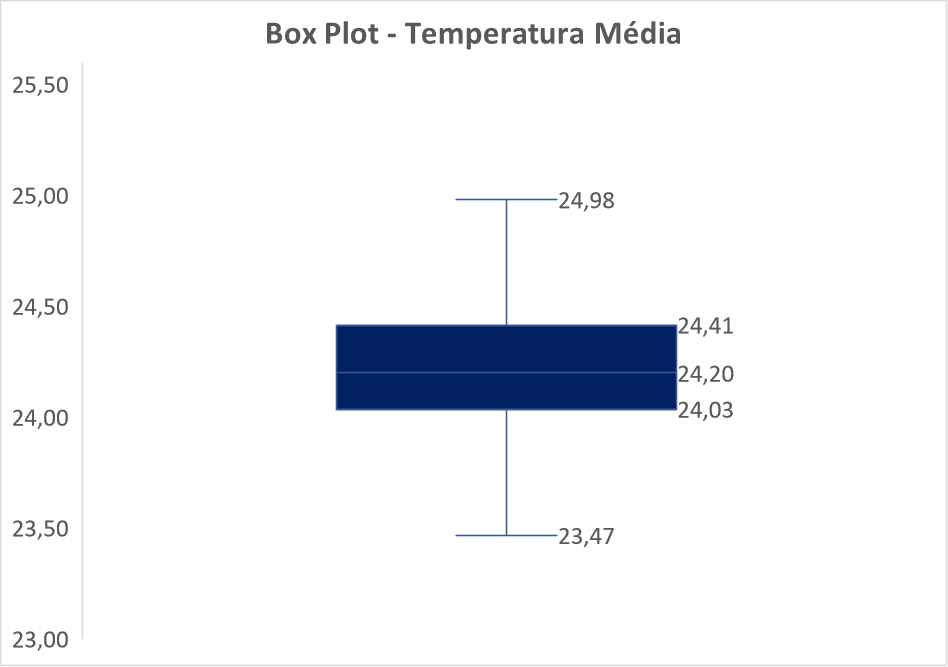
**Figura 24** - Pareto do Atributo Temperatura Célula



**Figura 25** -Histograma do Atributo Tensão Célula



**Figura 26 –** Coluna Empilhada do Atributo Tensão Célula



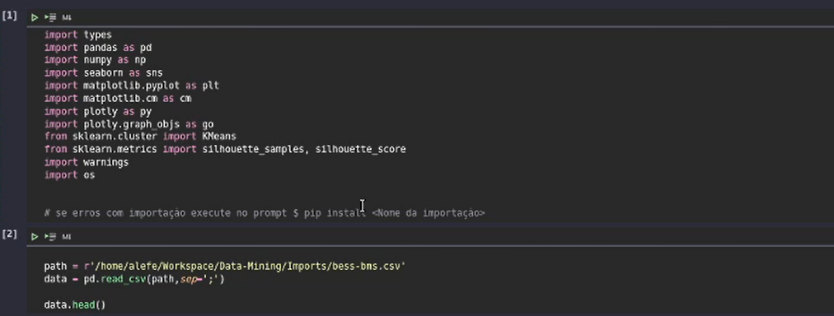
**Figura 27 –** Gráfico Box Plot da Tensão Célula

## **Metodologia Experimental**

Após o processamento e a análise descritiva dos dados, pela metodologia CRISP-DM, faz-se a aplicação das técnicas (modelagem e algoritmos) de mineração de dados. Nesse trabalho, utilizou-se a metodologia de agrupamento com o algoritmo *K-means*, implementado com a linguagem de programação Python 3.8*.*

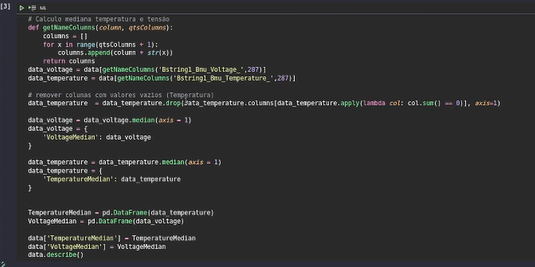
O desenvolvimento do agrupamento foi realizado utilizando o método K-means da biblioteca sklearn no módulo clusters, em que, para definição dos grupos, foram processados os dados de atributos.

Na **Figura 28**, tem-se a importação de todas as bibliotecas e os dados utilizados no desenvolvimento do agrupamento, representando as etapas 1 e 2.



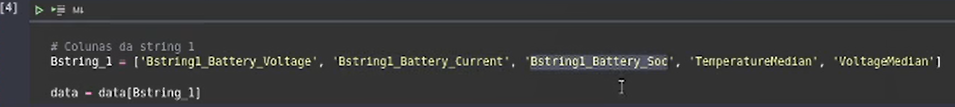
**Figura 28 –** (i) Etapa 1 - Importação das bibliotecas (ii) Etapa 2 - Importação dos dados utilizados no agrupamento desenvolvido

Ao analisar os dados, percebeu-se que havia colunas com valores nulos de temperatura, pois não possuíam sensores atuantes, então foi necessário remover essas colunas zeradas, para não interferir no resultado. Além disso, aprimorou-se a análise, calculando a mediana da tensão e mediana das células de baterias, ao invés da média. Na **Figura 29**, tem-se o código, da etapa 3.



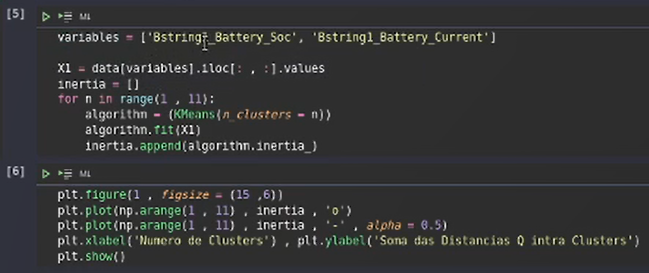
**Figura 29 –** Etapa 3 – Remoção das colunas zeradas de temperatura e cálculo da mediana da temperatura e tensão das células no agrupamento desenvolvido

Em seguida, juntamente com os stakeholders, selecionou-se as principais variáveis essenciais para realizar o agrupamento com K-means desse banco de dados: tensão total da *string1* (atributo - *Bstring1\_Battery\_Voltage*), corrente total da *string1* (atributo - *Bstring1\_Battery\_Currente*), estado de carga da *string1* (atributo - *Bstring1\_Battery\_SoC*), mediana da temperatura das células (atributo – *TemperatureMedian*) e mediana da tensão das células (atributo – *TemperatureMedian*), **Figura 30**.



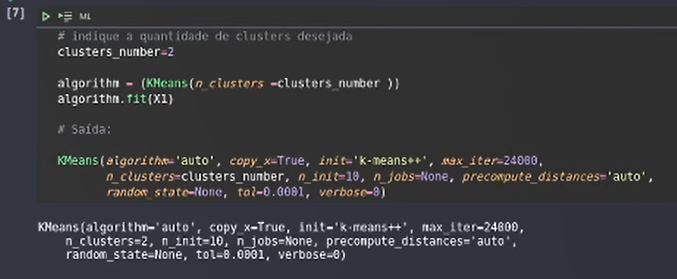
**Figura 30 –** Etapa 4 – Seleção das colunas do agrupamento

Na etapa seguinte, a partir da combinação de duas variáveis, implementou-se o algoritmo de Elbow para identificar inicialmente o número de *clusters* ideal para realizar o agrupamento, **Figura 31**.

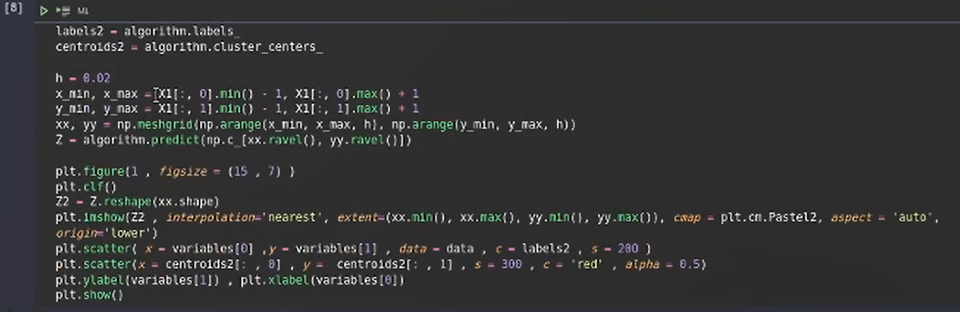


**Figura 31 –** Etapa 5 – Código do algoritmo de Elbow e da plotagem dos gráficos com duas variáveis

Na Etapa 6 e 7, utiliza-se o resultado obtido do algoritmo de Elbow para executar o algoritmo de K-means e exibir o gráfico do agrupamento resultante, **Figuras 32** e **33**, respectivamente.

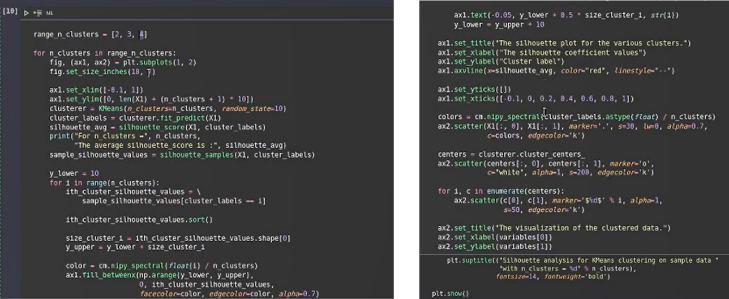


**Figura 32 –** Etapa 6 – Código do algoritmo K-means



**Figura 33 –** Etapa 7 – Código do gráfico resultante do agrupamento de K-means

Por fim, utiliza-se o método de interpretação e validação da consistência dentro dos clusters de dados, o coeficiente Silhouette. É uma técnica que oferece uma representação gráfica sucinta do quão bem cada objeto se situa dentro do seu *cluster*. Um valor elevado entre [-1;1], portanto, que o objeto apresenta um bom casamento com o seu *cluster*. Na **Figura 34**, apresenta-se o código implementado da Etapa 8.



**Figura 34 –** Etapa 8 – Gráfico resultante do agrupamento de K-means

# **ANÁLISE E DISCUSSÃO DOS RESULTADOS**

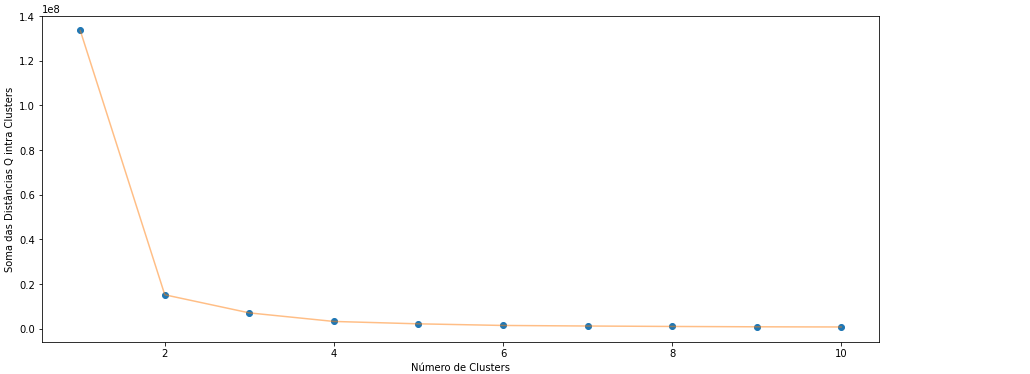
Apresenta-se, portanto, os resultados e as discussões das etapas dos seguintes agrupamentos:

1. SoC & Corrente Total;
2. SoC & Tensão Total;
3. SoC & Temperatura Célula;
4. SoC & Tensão Célula.

## **Resultados**

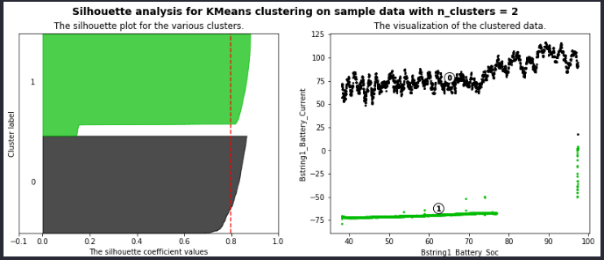
1. **SoC & Corrente Total**

Conforme mencionado anteriormente, a etapa inicial do agrupamento K-means é definir o K (número de *clusters*) que iremos avaliar. Para isso, plotou-se o gráfico de Elbow, **Figura 35**.

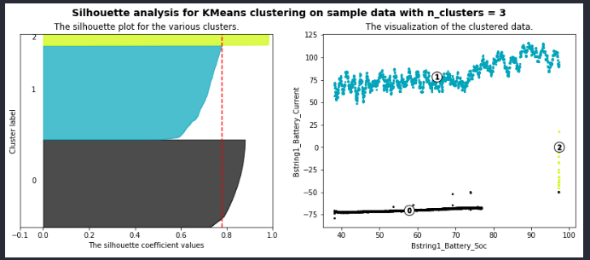


**Figura 35 –** Gráfico de Elbow para SoC e Corrente Total

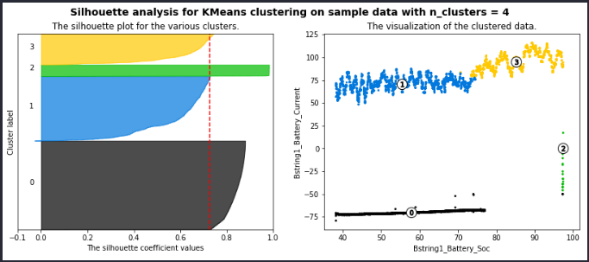
A partir desse resultado, supõe que o K ideal é o 2. Mas, para comprovar, analisou-se os resultados do coeficiente Sillhouette para K = 2, 3 e 4. Nas **Figuras 36, 37** e **38**, há os gráficos do coeficiente com K =2, K=3 e K=4, respectivamente.



**Figura 36 –** Plotagem do coeficiente Silhouette para K = 2 dos atributos SoC e Corrente Total



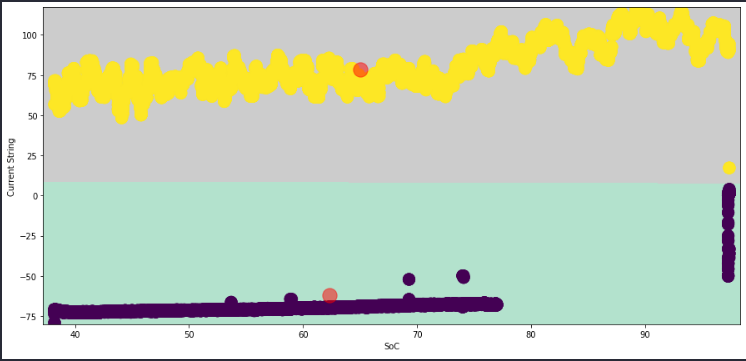
**Figura 37 –** Plotagem do coeficiente Silhouette para K = 3 dos atributos SoC e Corrente Total



**Figura 38 –** Plotagem do coeficiente Silhouette para K = 4 dos atributos SoC e Corrente Total

Observa-se, portanto, assim como o algoritmo de Elbow, o do coeficiente Silhouette também recomenda o K = 2 para os atributos corrente total e SoC, pois é coeficiente de maior valor, dentre os 3.

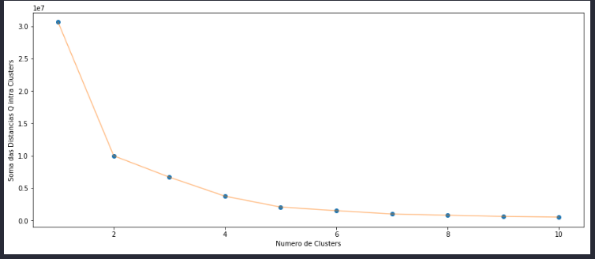
Por fim, na **Figura 39**, tem-se o agrupamento K-means para K = 2. Claramente, ver-se a distinção entre as regiões de recarga (corrente < 0) e descarga (corrente > 0) das baterias. Ademais, na recarga, o SoC máximo tende a 80% e o nível de corrente não varia tanto.



**Figura 39 –** Agrupamento K-means ideal dos atributos SoC e Corrente Total - K = 2

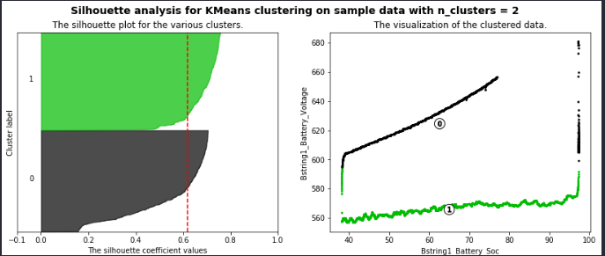
1. **SoC & Tensão Total**

Analogamente, a partir do resultado do gráfico de Elbow, **Figura 40**, estima-se que o K deve ser 2.

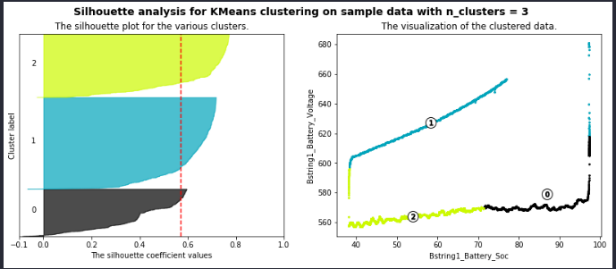


**Figura 40 –** Gráfico de Elbow para SoC e Tensão Total

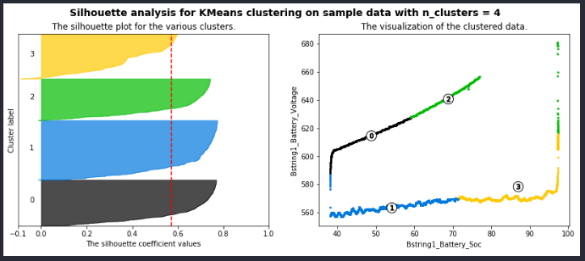
Então, analisa-se os resultados do coeficiente Silhouette para K = 2, K = 3 e K = 4, **Figuras 41, 42** e **43**, respectivamente.



**Figura 41 –** Plotagem do coeficiente Silhouette para K = 2 dos atributos SoC e Tensão Total

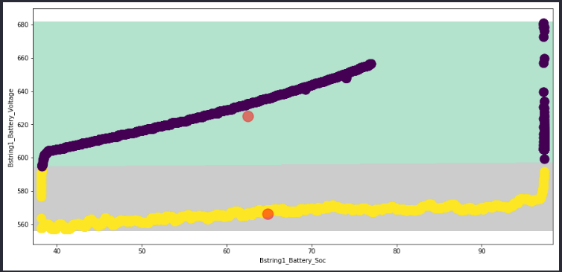


**Figura 42 –** Plotagem do coeficiente Silhouette para K = 3 dos atributos SoC e Tensão Total



**Figura 43 –** Plotagem do coeficiente Silhouette para K = 4 dos atributos SoC e Tensão Total

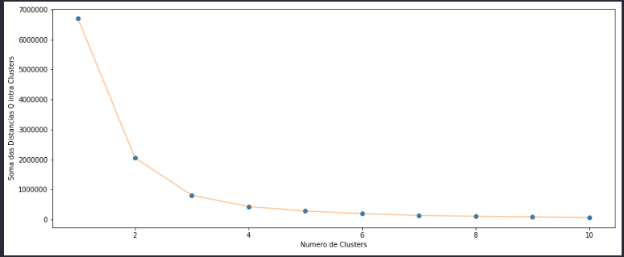
Novamente, percebe-se que os métodos de Elbow e Silhouette possuem o mesmo resultado, indicando K = 2, como *cluster* ideal para o agrupamento do SoC e a Tensão total. Portanto, na **Figura 44**, tem-se o agrupamento resultante de K = 2.



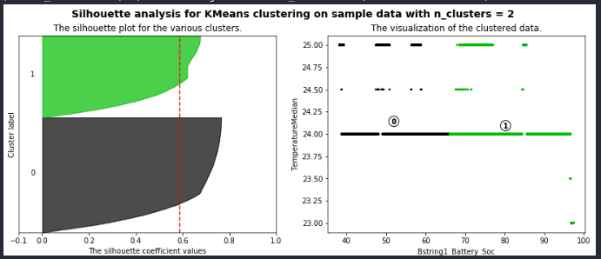
**Figura 44 –** Agrupamento K-means ideal dos atributos SoC e Tensão Total - K = 2

1. **SoC & Temperatura Célula**

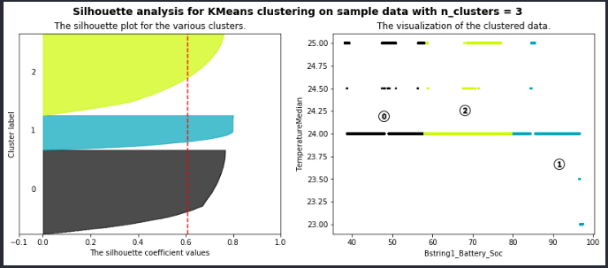
Diferente das outras combinações de atributos, a indicação do gráfico de Elbow (**Figura 45**) e do coeficiente Silhouette foram distintos, K = 2 e K = 3, respectivamente. Nas **Figuras 46, 47 e 48**, pode-se observar os gráficos Silhouette para K = 2, K = 3 e K = 4, respectivamente.



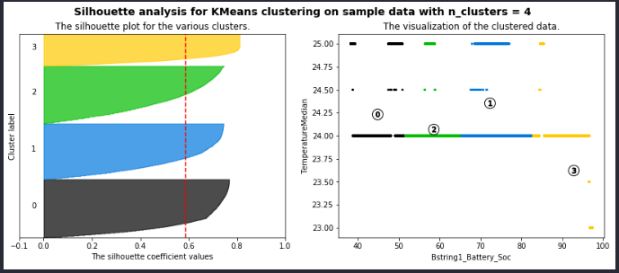
**Figura 45 –** Gráfico de Elbow para SoC e Temperatura Célula



**Figura 46 –** Plotagem do coeficiente Silhouette para K = 2 dos atributos SoC e Temperatura Célula

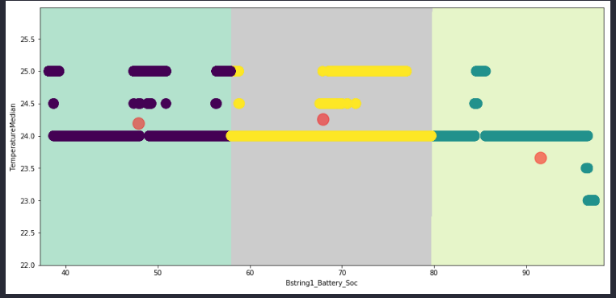


**Figura 47 –** Plotagem do coeficiente Silhouette para K = 3 dos atributos SoC e Temperatura Célula



**Figura 48 –** Plotagem do coeficiente Silhouette para K = 4 dos atributos SoC e Temperatura Célula

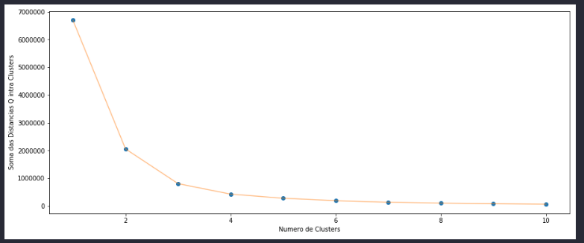
Assim, pode-se concluir que o agrupamento é ideal é com o K = 3, o, na **Figura 49**.



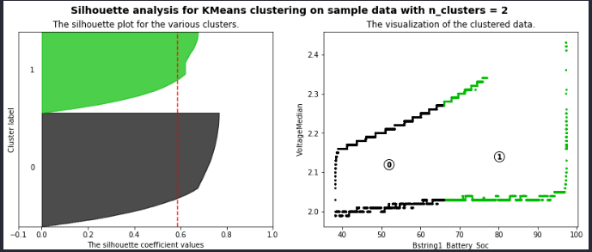
**Figura 49 –** Agrupamento K-means ideal dos atributos SoC e Temperatura Célula - K = 3

1. **SoC & Tensão Célula**

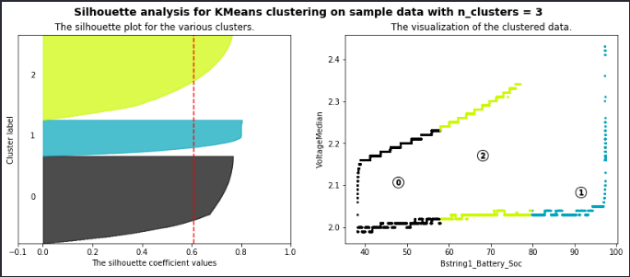
Analogamente ao SoC e Temperatura Célula, a combinação SoC e Tensão Célula indicou *clusters* ideais distintos K = 2 e K = 3 nas metodologias de Elbow e Silhouette, respectivamente. Nas **Figuras 50, 51, 52** e **53**, observa-se esses gráficos.



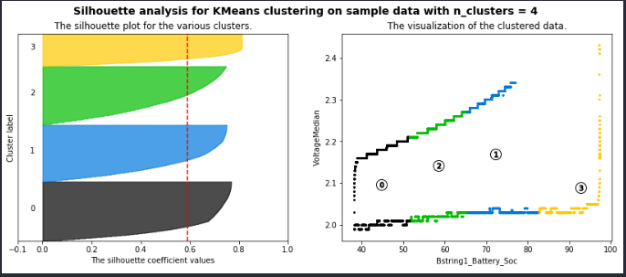
**Figura 50 –** Gráfico de Elbow para SoC e Tensão Célula



**Figura 51 –** Plotagem do coeficiente Silhouette para K = 2 dos atributos SoC e Tensão Célula

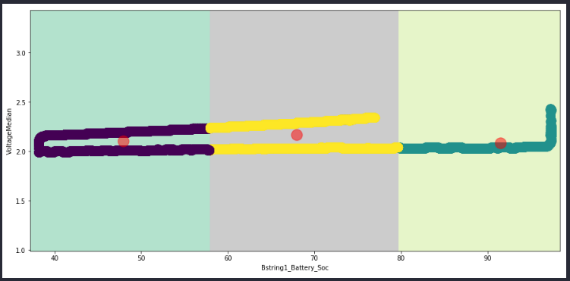


**Figura 52 –** Plotagem do coeficiente Silhouette para K = 3 dos atributos SoC e Tensão Célula



**Figura 53 –** Plotagem do coeficiente Silhouette para K = 4 dos atributos SoC e Tensão Célula

O resultado do agrupamento ideal, portanto, é para K = 3, **Figura 54**.



**Figura 54 –** Agrupamento K-means ideal dos atributos SoC e Tensão Célula - K = 3

## **Discussão**

Diante dos resultados apresentados, é possível concluir que os atributos Tensão e Temperatura Célula possuem o mesmo comportamento, quando agrupado com o SoC. Indica-se, portanto, que a temperatura e a tensão sofrem estresses, aproximadamente, em três níveis de SoC: [0; 60], ]60;80] e ]80;100].

Além disso, os agrupamentos de Tensão e Corrente Total possuem agrupamentos equivalentes com o SoC, conforme esperado, já que o SoC é dependente dessas duas variáveis por operar em modo bidirecional, como fonte de tensão e fonte de corrente, a depender da necessidade do consumidor.

Ficou claro, portanto, que o sistema opera em estado parcial de carga na maior do tempo o que é importante, pois ele foi desenvolvido para realizar ciclos diários e em horários específicos.

# **CONCLUSÕES E TRABALHOS FUTUROS**

Pode-se concluir que os agrupamentos são relevantes por identificarem regiões e equivalências ainda não avaliados pela engenharia de produto do BESS, visto que é possível dividir o perfil tanto na recarga, como na descarga. A depender da demanda da carga e/ou da exigência da aplicação do BESS, o aumento de *clusters* suporta a engenharia de produto na compreensão de operação das baterias. Promove-se, portanto, o desenvolvimento de baterias otimizadas por aplicação. Ratifica, portanto, a pluralidade do BESS e relevância para o setor elétrico.

Os próximos passos desse trabalho são avaliar novos agrupamento de atributos, principalmente, os vinculados as células, pois esses poderão identificar falhas previamente, garantindo a segurança do produto e reduzindo o custo de trocas indevidas.

# **TRABALHOS RELACIONADOS**

|  |
| --- |
| **[1]** M. Beudin, H. Zareipour, A. Schellenberglabe e W. Rosehart, “Energy storage for mitigating the variability of renewable electricity sources: An updated review,” *Energy for sustainable development,* nº 14, pp. 302-314, 2010.  **[2]** MINISTÉRIOS DE MINAS E ENERGIA. Secretaria de Planejamento e Desenvolvimento Estratégico Energético. Plano Nacional de Energia 2050 – PNE 2050. Consulta Pública. 2020.  **[3]** Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0 – Step-by-step data mining guide, The CRISPDM Consortium / SPSS Inc., 2000, Availabe on <http://www.crisp-dm.org>  **[4]** O. A. C. D. AMARANTE, Atlas do Potencial Eólico Brasileiro, 2001.  **[5]** JAIN, A.K.; DUBES, R. C. *Algorithms for clustering data.* [S.I.]: Prentice-Hall, Inc., 1988.  **[6]** XU, R.; WUNSCH, D. C. Clustering. Hoboken. [S.l.]: NJ: Wiley, 2009.  **[7]** JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. ACM computing surveys (CSUR), Acm, v. 31, n. 3, p. 264–323, 1999.  **[8]** BARBARA, D. An introduction to cluster analysis for data mining. Retrieved November, v. 12, p. 2003, 2000.  **[9]** HANDL, J.; KNOWLES, J.; KELL, D. B. Computational cluster validation in post-genomic data analysis. Bioinformatics, Oxford University Press, v. 21, n. 15, p. 3201–3212, 2005.  **[10]** Faceli, K., Lorena, A. C., Gama, J., e Carvalho, A. C. P. L. F. (2011). Inteligência artificial: Uma abordagem de aprendizado de máquina. Rio de Janeiro: LTC, 2:192  **[11]** TAN, P.-N.; STEINBACH, M.; KUMAR, V. Introdução ao datamining: mineração de dados.[S.l.]: Ciência Moderna, 2009.  **[12]** DINIZ, C. A. R.; NETO, F. L. *Data mining: uma introdução*. [S.l.]: ABE, 2000.  **[13]** MACQUEEN, J. et al. Some methods for classification and analysis of multivariate observations. In: OAKLAND, CA, USA. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. [S.l.], 1967. v. 1, n. 14, p. 281–297.  **[14]** DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern classification. 2nd. Edition. New York, p. 55, 2001.  **[15]** See, e.g., Ketchen, Jr, David J.; Shook, Christopher L. (1996). "The application of cluster analysis in Strategic Management Research: An analysis and critique". Strategic Management Journal. 17 (6): 441–458.  **[16]** FAYYAD, U; PIATETSKY-SHAPIRO, G; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence, 1996. |