

OTIMIZAÇÃO DA LOCALIZAÇÃO DE PONTOS DE ABASTECIMENTO DE HIDROGÊNIO VERDE EM ÁREAS URBANAS

**José Augusto Santos Costa¹, Antony Gabriel Medeiros Dos Santos¹,
Renata Da Conceição Rocha¹, Gustavo Soares Bezerra², Iulle de Macedo Guerra Neves²,
Mauro Antonio Guimarães Clark²**

¹Universidade Estadual do Piauí (UESPI), Curso de Tecnologia em Energias Renováveis,
Teresina, Brasil (joseaug150405@gmail.com)

²Universidade Estadual do Piauí (UESPI), Curso de Engenharia Elétrica, Teresina,
Brasil

Este trabalho apresenta a aplicação prática do algoritmo K-Means para definição de locais estratégicos de abastecimento de hidrogênio verde em área urbana. A partir da coleta de coordenadas geográficas de pontos relevantes em Teresina, os dados foram processados por técnicas de agrupamento, permitindo a identificação de regiões prioritárias. A metodologia demonstrou ser eficiente, simples e replicável para planejamento energético urbano.

Palavras-chave: Hidrogênio verde; K-Means; Geolocalização; Otimização espacial; Planejamento urbano

INTRODUÇÃO

A crescente preocupação com os impactos ambientais associados ao uso de combustíveis fósseis tem impulsionado o desenvolvimento e a adoção de fontes alternativas de energia mais limpas e sustentáveis (JIA et al., 2026). Nesse cenário, o hidrogênio verde (H₂V) surge como uma das principais soluções energéticas do futuro, especialmente no setor de transportes, devido à sua capacidade de armazenamento energético e à ausência de emissão de poluentes durante sua utilização (IRENA, 2022; IEA, 2023).

O hidrogênio verde é produzido por meio da eletrólise da água utilizando energia proveniente de fontes renováveis, como a solar e a eólica, o que o torna uma alternativa estratégica para a descarbonização da matriz energética. Sua aplicação em sistemas de mobilidade urbana tem potencial para reduzir significativamente as emissões de gases de efeito estufa, contribuindo para o cumprimento de metas ambientais globais (ABDELGHANY et al., 2026).

No contexto brasileiro, o cenário é particularmente favorável à expansão dessa tecnologia, devido à abundância de recursos naturais renováveis e ao crescente interesse em soluções energéticas sustentáveis (EPE, 2023). Entretanto, apesar das condições favoráveis, a implementação de

uma infraestrutura eficiente de abastecimento ainda constitui um dos principais entraves para a consolidação do hidrogênio como vetor energético no país (NOGUEIRA et al., 2025).

A ausência de uma rede estruturada de pontos de abastecimento compromete diretamente a viabilidade operacional de veículos movidos a hidrogênio, impactando aspectos como autonomia, logística e custo de operação. Dessa forma, a definição adequada da localização desses pontos torna-se um problema relevante no planejamento urbano e energético, exigindo abordagens que considerem a distribuição espacial da demanda e a otimização da cobertura territorial (IEA, 2022).

No contexto das cidades inteligentes, a integração entre dados geográficos, algoritmos de análise e planejamento urbano tem se tornado cada vez mais relevante. A utilização de técnicas computacionais permite não apenas compreender padrões existentes, mas também propor soluções otimizadas para problemas complexos, como a distribuição de infraestrutura energética.

Nesse sentido, o uso de técnicas de análise de dados e métodos computacionais tem se destacado como uma ferramenta eficiente para apoiar processos decisórios em planejamento urbano. Em particular, algoritmos de agrupamento permitem identificar



padrões espaciais e segmentar regiões com base na proximidade entre pontos, facilitando a identificação de áreas estratégicas (ZHANG et al., 2021).

Entre esses métodos, o algoritmo *K-Means* destaca-se por sua simplicidade, eficiência computacional e ampla aplicação em problemas de clusterização (JULIE et al., 2025). A partir da definição de um número de agrupamentos, o algoritmo organiza os dados em clusters e calcula centroides que representam pontos centrais de cada região, possibilitando sua interpretação como locais estratégicos para instalação de infraestrutura.

Além disso, a utilização de bases de dados geográficas abertas, como o *OpenStreetMap*, aliada a ferramentas de programação e análise de dados, permite o desenvolvimento de soluções acessíveis e replicáveis. Essa integração entre dados abertos e métodos computacionais reforça o papel da ciência de dados no planejamento de cidades inteligentes (SANTOS; ALMEIDA, 2023).

Diante desse contexto, este trabalho tem como objetivo aplicar, de forma prática, o algoritmo *K-Means* sobre coordenadas geográficas obtidas na cidade de Teresina, visando identificar regiões prioritárias e propor localizações otimizadas para pontos de abastecimento de hidrogênio verde, contribuindo para o planejamento eficiente da infraestrutura energética urbana.

REFERENCIAL TEÓRICO

O hidrogênio verde (H_2V) tem se consolidado como uma alternativa promissora para a transição energética global, especialmente no setor de transportes. Produzido a partir da eletrólise da água utilizando fontes renováveis, como energia solar e eólica, o hidrogênio verde apresenta potencial significativo para redução das emissões de gases de efeito estufa, sendo considerado um vetor energético estratégico para sistemas sustentáveis (IRENA, 2022; IEA, 2023).

No contexto da mobilidade urbana, a utilização do hidrogênio como combustível depende diretamente da existência de uma infraestrutura adequada de abastecimento. A definição eficiente da localização de estações de abastecimento é um desafio complexo, que envolve variáveis espaciais, demanda energética e aspectos logísticos. Estudos recentes apontam que a otimização desses pontos é fundamental para garantir

viabilidade econômica e operacional da tecnologia (LI et al., 2022).

Nesse cenário, técnicas de análise de dados e inteligência computacional têm sido amplamente aplicadas para apoiar o planejamento de infraestrutura. Entre essas técnicas, destacam-se os algoritmos de agrupamento, que permitem identificar padrões em dados espaciais e segmentar regiões com características semelhantes (ZHANG et al., 2021).

O algoritmo *K-Means* é um dos métodos mais utilizados para esse tipo de aplicação, devido à sua simplicidade, eficiência e baixo custo computacional. O método consiste na divisão de um conjunto de dados em K grupos, de modo que cada ponto pertença ao cluster cujo centroide esteja mais próximo, minimizando a variância intra-cluster (SILVA et al., 2022).

Além disso, a aplicação do *K-Means* em problemas de planejamento urbano tem demonstrado resultados relevantes na identificação de regiões estratégicas para alocação de serviços, incluindo transporte, energia e logística. Sua capacidade de representar áreas por meio de centroides torna o algoritmo especialmente adequado para a definição de pontos de abastecimento (SANTOS; ALMEIDA, 2023).

A integração entre dados geográficos, algoritmos de agrupamento e ferramentas computacionais tem possibilitado avanços significativos no desenvolvimento de soluções para cidades inteligentes. Nesse contexto, o uso de bases abertas, como o *OpenStreetMap*, aliado a técnicas de análise espacial, permite a construção de modelos replicáveis e adaptáveis a diferentes realidades urbanas.

MATERIAL E MÉTODOS

A metodologia adotada neste trabalho pode ser visualizada de forma resumida por meio do fluxograma apresentado na Figura 1, no qual são destacadas as principais etapas do processo, desde a coleta dos dados até a definição dos pontos estratégicos.



Figura 1– Fluxograma

• Coleta De Dados Geográficos

Os dados utilizados consistem em coordenadas geográficas (latitude e longitude) de pontos relevantes ao estudo, incluindo postos de combustíveis, vias de grande circulação e regiões urbanas com potencial demanda. Essas informações foram obtidas por meio de ferramentas de geolocalização, com destaque para plataformas baseadas em dados abertos, como o *OpenStreetMap*, além de consultas em serviços de mapas digitais.

A escolha dessas fontes se justifica pela disponibilidade de dados atualizados, acessíveis e adequados para análises espaciais em ambiente urbano.

O conjunto de dados utilizado neste estudo foi composto por aproximadamente 230 coordenadas geográficas, selecionadas de forma a representar diferentes regiões da cidade de Teresina. A escolha desses pontos buscou abranger áreas com potencial demanda por abastecimento, incluindo regiões centrais e periféricas, permitindo uma análise mais abrangente da distribuição espacial.

Embora as coordenadas geográficas estejam representadas em latitude e longitude, foi adotada a distância euclidiana como aproximação para o cálculo de proximidade entre os pontos. Essa abordagem é amplamente utilizada em análises exploratórias e apresenta resultados satisfatórios em escalas urbanas, especialmente quando se busca uma modelagem simplificada do espaço geográfico (SILVA et al., 2022).

• Tratamento E Preparação Dos Dados

Após a coleta, os dados passaram por um processo de pré-processamento, que incluiu:

- Remoção de registros duplicados;
- Padronização das coordenadas geográficas;
- Organização em estrutura tabular (latitude e longitude);
- Conversão para formato compatível com bibliotecas computacionais.

Essa etapa foi fundamental para garantir consistência e qualidade na aplicação dos algoritmos.

• Aplicação Do Algoritmo K-Means

Para a análise dos dados, foi utilizado o algoritmo de agrupamento K-Means, amplamente empregado na segmentação de dados com base na proximidade entre pontos.

O algoritmo *K-Means* opera por meio de um processo iterativo de otimização, cujo objetivo é particionar o conjunto de dados em K grupos distintos, de forma que a variabilidade interna de cada grupo seja minimizada. Inicialmente, são definidos K centróides, que podem ser escolhidos aleatoriamente ou com base em critérios específicos.

A partir dessa inicialização, cada ponto do conjunto de dados é associado ao centróide mais próximo, com base em uma métrica de distância, geralmente a distância euclidiana. Essa etapa resulta na formação inicial dos clusters.

Em seguida, os centróides são recalculados com base na média das coordenadas dos pontos pertencentes a cada cluster. Esse novo conjunto de centróides passa a representar melhor a distribuição dos dados em cada grupo.

O processo de atribuição dos pontos e atualização dos centroides é repetido iterativamente até que não ocorram mais mudanças significativas na posição dos centroides ou na composição dos clusters, caracterizando a convergência do algoritmo.

Dessa forma, o *K-Means* encontra uma solução aproximada para o problema de agrupamento, minimizando a soma das distâncias entre os pontos e os centroides de seus respectivos clusters.

O objetivo do algoritmo é minimizar a soma das distâncias quadráticas entre os pontos e os centroides dos clusters, conforme a Equação (1):

$$J = \sum_{i=1}^K \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (1)$$

Em que K representa o número de clusters, C_i o conjunto de pontos pertencentes ao cluster i e μ_i o centróide correspondente.

O processo iterativo do algoritmo envolve as seguintes etapas:

- Inicialização dos centroides;
- Associação dos pontos ao centróide mais próximo;
- Recalculo dos centroides com base na média dos pontos;
- Repetição do processo até convergência.

• Pseudocódigo Do Algoritmo

O processo computacional adotado neste estudo pode ser sintetizado por meio do pseudocódigo apresentado no Algoritmo 1, que descreve as etapas de determinação do número de clusters por meio do método do cotovelo e a aplicação final do algoritmo *K-Means*.

Algoritmo 1 – Pseudocódigo do método *K-Means* com seleção do número de clusters

Entrada: Conjunto de pontos $D = \{(lat, lon)\}$, intervalo $[K_{min}, K_{max}]$

Saída: Centroides finais C e número ótimo de clusters K^*

Etapa 1: Método do cotovelo

para cada $K = K_{min}$ **até** K_{max} **faça**

Inicializar K centroides

enquanto não convergir **faça**

Atribuir cada ponto ao centróide mais próximo

Recalcular centroides

Calcular a inércia $J(K)$

Determinar K^* com base na curva $J(K)$

Etapa 2: *K-Means* final

Inicializar K^* centroides

enquanto não convergir **faça**

para cada ponto $p \in D$ **faça**

Calcular distância até cada centróide

Atribuir ao cluster mais próximo

para cada cluster i **faça**

Recalcular centróide μ_i

Retornar centroides C

• Justificativa Da Escolha Do Algoritmo

A escolha do algoritmo *K-Means* se justifica por sua simplicidade computacional, eficiência e ampla aplicação em problemas de agrupamento espacial. O método apresenta baixo custo computacional e fácil implementação, sendo adequado para análise de dados geográficos em ambientes urbanos.

Além disso, o algoritmo permite identificar regiões centrais representativas por meio dos centroides, os quais podem ser interpretados diretamente como localizações estratégicas para instalação de infraestrutura. Essa característica torna o *K-Means* particularmente relevante para problemas de planejamento urbano e distribuição de recursos.

Outra vantagem importante é sua compatibilidade com técnicas de validação, como o método do cotovelo, utilizado neste trabalho para definir o número ideal de agrupamentos.

- Definição Do Número De Clusters

A escolha do número de agrupamentos foi realizada por meio do método do cotovelo, que avalia a variação da inércia em função do número de clusters. Esse procedimento permite identificar um ponto de equilíbrio entre a complexidade do modelo e a qualidade da segmentação dos dados.

- Implementação Computacional

A implementação computacional foi realizada em ambiente de desenvolvimento baseado na linguagem *Python*, utilizando bibliotecas amplamente difundidas na análise de dados, como *pandas* para manipulação de dados, *scikit-learn* para aplicação do algoritmo *K-Means* e *matplotlib* para geração de visualizações gráficas.

Essas ferramentas possibilitaram a execução eficiente de todas as etapas do processo, desde a leitura e organização dos dados até a obtenção dos resultados finais. A utilização de bibliotecas consolidadas também contribui para a reprodutibilidade do estudo, permitindo que a metodologia seja aplicada em diferentes contextos com facilidade.

O processo computacional incluiu:

- Leitura e organização dos dados;
- Execução do algoritmo *K-Means*;
- Cálculo dos centroides;
- Geração de dados para visualização.

As visualizações foram produzidas com bibliotecas gráficas, permitindo a representação dos agrupamentos e sua interpretação espacial.

- Determinação Dos Pontos Estratégicos

Os centroides obtidos a partir do algoritmo foram considerados como estimativas das localizações ideais para instalação dos pontos de abastecimento. Esses pontos representam posições centrais dentro de cada agrupamento, contribuindo para maximizar a cobertura geográfica e a acessibilidade.

RESULTADOS E DISCUSSÃO

A aplicação do algoritmo *K-Means* aos dados geográficos coletados permitiu a identificação de padrões espaciais relevantes para o planejamento da infraestrutura de abastecimento de hidrogênio verde. Inicialmente, foi realizada a análise do número ideal de clusters por meio do método do cotovelo. A Figura 2 apresenta a variação da inércia em função do número de agrupamentos (*K*). Observa-se que a curva apresenta uma redução acentuada nos primeiros valores de *K*, seguida de uma diminuição progressivamente mais suave, caracterizando o ponto de inflexão conhecido como “cotovelo”.

Esse comportamento indica que, a partir desse ponto, o aumento no número de clusters não resulta em melhorias significativas na qualidade do agrupamento, justificando a escolha do valor de *K* adotado neste estudo.

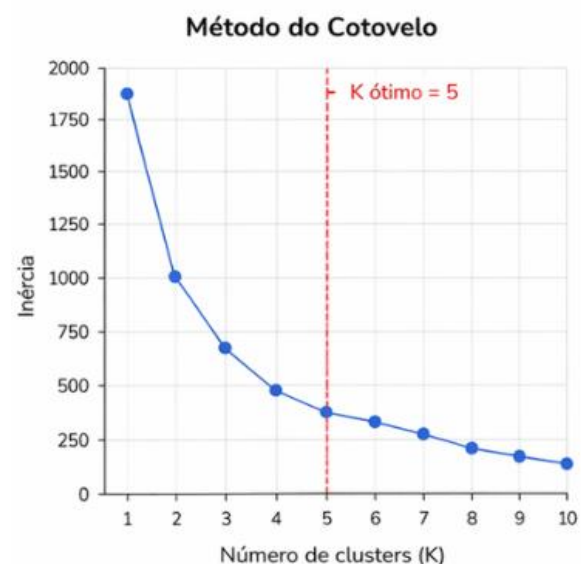


Figura 2 – Método do cotovelo

Com base na definição do número de clusters, o algoritmo foi aplicado aos dados geográficos, resultando na formação de agrupamentos espaciais bem definidos. A Figura 3 apresenta a distribuição dos pontos analisados, destacando os diferentes clusters por cores distintas, bem como a posição dos centroides.

A análise da Figura 3 permite observar que os agrupamentos acompanham a distribuição espacial dos dados, evidenciando regiões com maior concentração de pontos e, consequentemente, maior potencial de demanda por abastecimento. Clusters

mais densos indicam áreas com maior atividade, enquanto clusters mais dispersos refletem regiões com menor concentração de pontos.



Figura 3 – Mapa dos clusters

Os centroides obtidos para cada cluster representam as coordenadas médias dos pontos agrupados, sendo interpretados como possíveis localizações estratégicas para a instalação de pontos de abastecimento. A Tabela 1 apresenta os valores das coordenadas correspondentes a cada centroide.

Tabela 1 – Coordenadas dos centroides

Coordenadas dos centróides (K = 5)

Cluster	Latitude	Longitude
1	-5.078959	-42.803398
2	-5.067212	-42.781653
3	-5.097561	-42.797924
4	-5.115284	-42.816903
5	-5.122968	-42.789527

A análise dos centroides indica que estes se posicionam em regiões centrais dentro de cada agrupamento, contribuindo para a redução das distâncias médias entre os pontos atendidos e, conseqüentemente, para a melhoria da eficiência logística do sistema de abastecimento.

Do ponto de vista aplicado, os resultados obtidos indicam que a utilização de técnicas de agrupamento pode contribuir significativamente para a tomada de decisão no planejamento de infraestrutura energética. A definição de pontos baseada em centroides reduz a subjetividade do processo, permitindo uma abordagem orientada por dados e potencialmente mais eficiente do que métodos tradicionais.

Além disso, observa-se que os agrupamentos refletem a organização urbana da cidade, com clusters mais densos concentrados em áreas centrais e clusters mais dispersos em regiões periféricas. Esse resultado está em conformidade com padrões descritos na literatura de análise espacial (SILVA et al., 2022).

A metodologia demonstrou ser eficaz na identificação de regiões prioritárias, oferecendo uma abordagem baseada em dados para apoio ao planejamento urbano e energético. Sua simplicidade de implementação e adaptabilidade a diferentes conjuntos de dados reforçam sua aplicabilidade prática.

CONCLUSÃO

A aplicação do algoritmo *K-Means* demonstrou ser uma abordagem eficiente para a definição de localizações estratégicas de pontos de abastecimento de hidrogênio verde em áreas urbanas. A partir da análise de coordenadas geográficas, foi possível identificar regiões com maior potencial de demanda e propor soluções baseadas em agrupamento espacial. A metodologia adotada mostrou-se simples, de fácil implementação e capaz de gerar resultados consistentes, mesmo com um conjunto limitado de dados. A utilização do método do cotovelo permitiu uma escolha adequada do número de clusters, contribuindo para a qualidade da segmentação dos pontos analisados.

Os centroides obtidos representam alternativas viáveis para a instalação de infraestrutura de abastecimento, favorecendo a cobertura territorial e a acessibilidade. Dessa forma, o estudo contribui como uma ferramenta de apoio ao planejamento energético urbano, especialmente no contexto da transição para fontes de energia mais limpas.

Como trabalhos futuros, recomenda-se a incorporação de novas variáveis, como densidade populacional, fluxo de veículos e aspectos econômicos, a fim de refinar ainda mais a análise e ampliar a aplicabilidade da metodologia.



Dessa forma, o presente estudo reforça o potencial da aplicação de técnicas de ciência de dados no planejamento urbano, evidenciando que abordagens orientadas por dados podem contribuir de forma significativa para a transição energética e para o desenvolvimento de cidades mais sustentáveis.

AGRADECIMENTOS

Os autores deste presente artigo agradecem à Universidade Estadual do Piauí (UESPI) pelo suporte institucional e pela infraestrutura acadêmica disponibilizada para a realização deste estudo. Expressa-se também, um agradecimento especial ao Núcleo de Formação e Pesquisa em Energias Renováveis e Telecomunicações do Piauí (NUFPERPI), pelo apoio técnico e pelo ambiente de fomento à pesquisa e inovação tecnológica, fundamentais para a viabilização deste trabalho.

REFERÊNCIAS

- ABDELGHANY, M. B., SHAFIQUURRAHMAN, A., Al-Durra, A., MOHAMED, M., Hu, J., VACCARO, A., ... & El MOURSI, M. S. *Hydrogen energy systems for decarbonizing smart cities and industrial applications: A review. Renewable and Sustainable Energy Reviews*, v. 226, p. 116370, 2026.
- EPE – EMPRESA DE PESQUISA ENERGÉTICA. *Plano Nacional do Hidrogênio: Bases para a consolidação da economia do hidrogênio no Brasil*. Brasília: Ministério de Minas e Energia, 2023.
- IEA – INTERNATIONAL ENERGY AGENCY. *Global Hydrogen Review 2022*. Paris: IEA, 2022.
- IEA – INTERNATIONAL ENERGY AGENCY. *Global Hydrogen Review 2023*. Paris: IEA, 2023.
- IRENA – INTERNATIONAL RENEWABLE ENERGY AGENCY. *Green Hydrogen: A guide to policy making*. Abu Dhabi: IRENA, 2022.
- JIA, Wenhao; DING, Tao; HE, Yuhua. Synergistic integration of green hydrogen in renewable power systems: A comprehensive review of key technologies, research landscape, and future perspectives. *Renewable and Sustainable Energy Reviews*, v. 226, p. 116375, 2026.
- JULIE, J.; BALAJI, V. S.; AMUTHA YALINI, E. Optimizing Textual Data Categorization: An Evaluation of K-Means, Mean Shift, and Hierarchical Clustering Methods. In: **2025 International Conference on Innovative Trends in Information Technology (ICITIIT)**. IEEE, 2025.
- LI, X.; ZHANG, P.; WANG, Y. Optimization of hydrogen refueling station deployment using clustering techniques. *International Journal of Hydrogen Energy*, v. 47, n. 18, p. 11234–11245, 2022.
- SANTOS, F. R.; ALMEIDA, J. P. Planejamento de infraestrutura energética com apoio de ciência de dados. *Revista de Engenharia e Energia*, v. 29, n. 1, p. 77–91, 2023.
- SILVA, R. S.; OLIVEIRA, L. H.; COSTA, M. A. Aplicação de algoritmos de clusterização na análise espacial urbana. *Revista Brasileira de Computação Aplicada*, v. 14, n. 2, p. 45–58, 2022.
- NOGUEIRA, D. M., OLIVEIRA NETO, G. C., MATTOS, C. A. D., & SCUR, G. Pathways for hydrogen adoption in the Brazilian trucking industry: a low-carbon alternative to fossil fuels. *Processes*, v. 13, n. 10, p. 3240, 2025.
- ZHANG, Y.; WANG, J.; LIU, H. Spatial clustering analysis for urban infrastructure planning using K-Means algorithm. *Sustainable Cities and Society*, v. 74, p. 103–118, 2021.