



# Cluster analysis applied to fuel price data in Campina Grande

Richard Silva <sup>1\*</sup>   
Thaís Azevedo <sup>2</sup>   
Hiago Martins <sup>3</sup>   
Lucas Cardoso Pereira <sup>4</sup> 

<sup>1 2 3 4 5</sup> Departamento de Estatística, State University of Paraíba, Campina Grande, Brazil

Emails: <sup>1</sup> richard.silva@aluno.uepb.edu.br ; <sup>2</sup> thaís.azevedo@aluno.uepb.edu.br;

<sup>3</sup> hiago.martins@servidor.uepb.edu.br; patricia.lima@aluno.eupb.edu.br; <sup>4</sup> lucascardoso@servidor.uepb.edu.br.

\* Corresponding author

**This article has been retracted by the publisher. Readers are encouraged to consult the retraction notice for further details.**

## RESEARCH ARTICLE

### Socioeconomic Analytics

<https://periodicos.ufpe.br/revistas/SECAN/>

ISSN Online: 2965-4661

**Submitted on:** 18.11.2024

**Accepted on:** 15.12.2024.

**Published on:** 27.12.2024.

**Retracted on:** 09.05.2025

Copyright © 2024 by author(s).

This work is licensed under the Creative Commons Attribution International License CC BY-NC-ND 4.0

<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>



## Abstract

This work uses cluster analysis to analyze similarities in the prices of regular gasoline and ethanol in Campina Grande-PB, Brazil. This multivariate statistical technique groups similar elements into distinct clusters using the non-hierarchical method (k-means). The analysis reports three clusters: cluster 1 has the lowest prices, while cluster 3 has the highest. We report descriptive statistics and a discussion based on the empirical local context.

## Keywords

Cluster Analysis, K-means, fuel price fuel, Exploratory Data Analysis, transport, Data Analytics.

## 1. Introdução

Conforme PETRÓLEO (2017) e a ANP, os postos de combustíveis são estabelecimentos responsáveis pela venda de combustíveis para veículos automotores, sendo os mais comuns a gasolina e o diesel. Alguns postos também oferecem combustíveis alternativos, como o etanol e o gás natural. Para que um posto de combustíveis seja estabelecido e comece a operar, é necessário que ele passe por uma inspeção detalhada realizada pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP).

A análise de *cluster* é o processo de dividir um conjunto de dados em grupos que refletem padrões consistentes de comportamento, organizando o banco de dados de maneira que cada grupo ou partição seja semelhante segundo um critério ou métrica específica (Doni, 2004). Após a identificação dos padrões, esses grupos podem ser usados para segmentar os dados em subconjuntos mais facilmente compreensíveis, facilitando análises posteriores.

Assim, pretendesse verificar as similaridades entre os preços dos seguintes combustíveis: gasolina comum e etanol, referente ao mês de outubro de 2024 no município de Campina Grande–PB. Para isto, foi utilizada a técnica multivariada de análise de *cluster*, que consiste em formar grupos semelhantes, no caso do trabalho comparar os postos com preços mais parecidos.

## 2. Metodologia

O objetivo deste trabalho é verificar as similaridades dos preços para os seguintes combustíveis, gasolina comum e etanol referente ao mês de outubro de 2024 no município de Campina Grande– PB. Os dados foram fornecidos pela pesquisa de preços de combustíveis que é realizada mensalmente pelo Fundo Municipal de Defesa dos Direitos Difusos (PROCON) de Campina Grande- PB, relativas ao mês de outubro no ano de 2024. O conjunto de dados é composto por 61 postos de combustíveis distribuídos na cidade, suas respectivas variáveis são: posto de combustível, bandeira, bairro, gasolina comum, gasolina aditivada diesel, s-10 e etanol.

Para análise utilizamos os combustíveis gasolina comum e etanol, uma vez que para os demais tipos de combustíveis não tinha fornecimento dos mesmos em todos os postos.

Análise de *cluster* consiste em agrupar observações de modo que cada observação seja o mais semelhante possível segundo suas características, os grupos tem que ter elevada homogeneidade interna e heterogeneidade externa (Hair et al., 2009). Desta forma, foi realizada a análise de *cluster* não hierárquico, o método *k-means*, depois dos *clusters* formados, foi realizada a análise descritiva para observar o comportamento dos preços em cada um dos *clusters*. Os procedimentos estatísticos, cálculos e gráficos executados neste trabalho foram realizados no software R

## 3. Aplicação

Para os 61 postos os quais participaram da pesquisa feito pelo PROCON, tem-se que o preço médio para a gasolina comum foi de R\$ 6,04, já o etanol apresentou uma média de R\$3,97.

Ainda pode-se destacar que o etanol apresentou uma variabilidade maior que a gasolina comum.

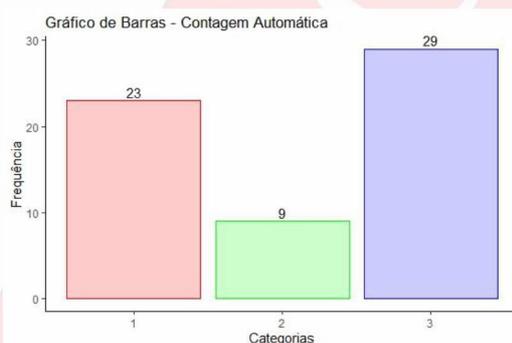
Tabela 1 – Sumário das estatísticas para gasolina comum por *clusters*

Combustível	Média	D.P.	Mín.	1º Q.	2º Q.	3º Q.	Máx.
Gasolina Comum	6,04	0,07	5,90	5,98	6,09	6,09	6,19
Etanol	3,97	0,09	3,79	3,89	3,99	3,99	4,29

Obtendo o número de *clusters* através de método de *Elbow* (cotovelo), foi verificado o total da variância explicada, após de realizar testes, optou-se pelo número de 3 *cluster* o qual obteve um total de variância explicada de 71,5%.

Utilizando o método *K-means* podemos observar a quantidade de postos de combustíveis em cada *cluster*, na figura 1.

Figura 1 – Frequência de postos por *cluster*



Observa-se que o *cluster* 1 ficou um total de 23 postos de combustíveis, o *cluster* 2 com 9 e por último o *cluster* 3 com 29, totalizando 61 postos de combustíveis.

No intuito de se fazer um levantamento sobre algumas características importantes, a Tabela 2 apresenta algumas estatísticas descritivas obtidas para os clusters em relação a gasolina comum.

Tabela 2 – Sumário das estatísticas para gasolina comum por *clusters*

Estatísticas	Cluster 1	Cluster 2	Cluster 3
Mínimo	5,90	5,99	6,09
Média	5,96	6,06	6,09
Mediana	5,95	6,09	6,09
Máximo	5,99	6,09	6,19
Desvio Padrão	0,03	0,05	0,02

Fonte: Elaborada pelos autores, (2024).

Observando as estatísticas para cada *cluster*, o que chama atenção é o *cluster 3* que tem um total de 29 postos de combustíveis e teve a maior média de preço, indicando que esses 29 postos vendiam a gasolina comum com os preços mais elevados durante o período de estudo. Os valores da média indicam que os postos com os melhores preços são os do *cluster 1* (um total de 23 postos), pois apresenta a menor média de preço para a gasolina comum, o *cluster 2* com um total de 9 tem a média entre o *cluster 1* e 3. A Tabela 3 apresenta estatísticas descritivas para os *clusters* em relação ao etanol.

Tabela 3 - Sumário das estatísticas para etanol por *clusters*.

Estatísticas	Cluster 1	Cluster 2	Cluster 3
Mínimo	3,79	3,88	4,09
Média	3,93	3,94	4,13
Mediana	3,95	3,99	4,09
Máximo	4,08	3,99	4,29
Desvio Padrão	0,06	0,05	0,07

Fonte: Elaborada pelos autores, (2024).

A Tabela 4 detalha os estabelecimentos com a respectiva localização e cluster pertencente.

Tabela 4 – Informações dos estabelecimentos.

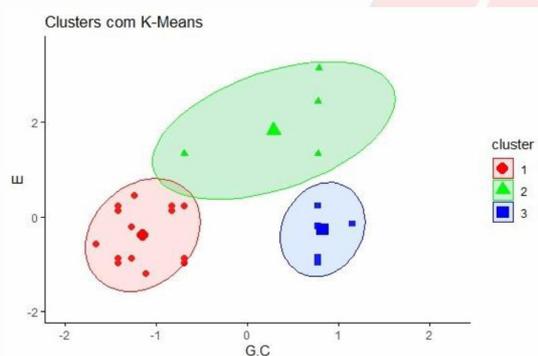
Cluster	Estabelecimento	Bandeira	Endereço	Bairro
1	Anel do Brejo	Branca	Rua Dr. Vasconcelos, 127.	Alto Branco
1	MS	Branca	Rua Avani Casemiro de Albuquerque, 20.	Alto Branco
1	Santo Antônio	Branca	Rua Tavares Cavalcante, 655.	Centro
1	Posto Roma	Alê	Rua Quebra Quilos 47	Centro
1	Santa Terezinha	Ipiranga	Av. Prof. Severino Bezerra Cabral, s/n, BR 230.	Vila Cabral
1	GS 3	Branca	Av. João Wallig,160	Itararé
1	Posto GS	Ipiranga	R. Santa Catarina, 135	Vila Cabral
1	Posto Santo Antônio	Branca	BR, 230	
1	Posto Marília	Ipiranga	Av. Prof. Almeida Barreto 201	Estação Velha
1	São Marcos	Ipiranga	Av. Assis Chateaubriand, 878.	Liberdade
1	Paulistano	Ipiranga	Av.Jorn.Assis chateaubriand, 2292	Jardim Paulistano
1	Master Gás	Branca	Av, Assis Chateaubriand, 2675 - Tambor	Tambor
1	Amigão	Branca	Av. Assis Chateaubriand, 2980.	Distrito industrial
1	Distrito	Branca	Av. João Wallig, 2304.	Distrito industrial
1	Dallas	Petrobrás	Rua Tomaz Soares 170	Catolé
1	Posto Sudoeste	Branca	BR-230, 1000	Velame
1	Domingo	Setta	Av. Dep. Raimundo Asfora, 1699.	Três irmãs
1	Posto Santo Antônio 2	Branca	Rodovia BR 230	Três irmãs
1	Posto Fechine	Shell	Av. Mal. Floriano Peixoto 1350	Centro

1	Norberto	Branca	Rua Lourival de Andrade, 1001.	Bodocongó
1	Nova Campina		R. Três Irmãs, 1613	Acácio Figueiredo
1	Posto Canal		R. Elizabeth Arruda, 310	Bodocongó
1	Posto GS 2	Branca	Av. Sen. Argemiro de Figueiredo, 2458	Vila Cabral
2	Bandeirantes	Branca	R, Eptácio Pessoa, 385	Centro
2	Posto PSR	Branca	R, João Suassuna, 579	Centro
2	BR Mania	Petrobrás	R. João Quirino, 895	Catolé
2	Posto Shopping	Shell	Av. Pref. Severino Bezerra Cabral 1225	Mirante
2	Rodo Paraíba	Petrobrás	Acácio Figueiredo	Velame
2	São José	Branca	Av. Francisco Lopes de Almeida, 5	Três Irmãs
2	Posto Portal	Branca	Av. Francisco Lopes de Almeida, 2000	Malvinas
2	Posto Unigás	Branca	Av. Joaquim Caroca 517	Bodocongó
2	Posto RP Petróleo	Petrobrás	R. João Quirino, 895	Catolé
3	Posto São Luís II	Ipiranga	R, Raimundo Alves da Silva, 378	Centro
3	Posto São Vicente	Ipiranga	Rua Getúlio Vargas, 684.	Centro
3	Posto FRS	Branca	Av. Pres. Getúlio Vargas, 1367	Bela Vista
3	Posto Cem réis	Petrobrás	Rua João Alves de Oliveira, 327.	Centro
3	Posto Opção	Branca	Rua Almirante Barroso, 284	Alto Branco
3	São Luís VII	Shell	Av. Manoel Tavares, 710	Alto Branco
3	Posto (Viadulto)	Ipiranga	R. Giló Guedes, Sn	Santo Antônio
3	Posto JE	Ipiranga	R. Francisco Antonio Nascimento, 584	Santo Antônio
3	Posto Monumento	Ipiranga	Av. Pref. Severino Bezerra Cabral 38	José Pinheiro
3	São Luís V	Shell	Rua Inácio Pedro Diniz, 115.	Catolé
3	Santa Ana	Ipiranga	Av. Pref. Severino Cabral 950	Catolé
3	São Luís XIX	Shell	Prata	Prata
3	Padre Cícero	Ipiranga	Av. Assis Chateaubriand, 44.	Liberdade
3	Posto (Rede Opção)	Ipiranga	Av. Pres. Juscelino Kubitschek, 1500	Cruzeiro
3	Almirante	Ipiranga	Alm. Barroso 634	Quarenta
3	Avenida	Ipiranga	Av. Assis Chateaubriand, 1600.	Liberdade
3	São Luís XXI	Ipiranga		
3	Posto Alternativa	Petrobrás	Av, Pres. Juscelino Kubitschek, 1500	Cruzeiro
3	Posto Jardins	Branca	Rua Aprígio Pereira Nepomuceno	Jardim Paulistano
3	Posto Rodoviário	Branca	Av. Dom Pedro II 148	Prata
3	Posto Centenário	Ipiranga	Av. Prof. Almeida Barreto	Centenário
3	Posto São Luís I	Ipiranga	Rua Pres. Eptácio Pessoa 350	Centenário
3	Posto Laís XII	Alê	Rua Dinâmica Alves Correia 325	Centenário
3	Posto Sertões	Ipiranga	Avenida Marechal Floriano Peixoto, 2770	Centenário
3	Posto Universitário	Petrobrás	Rua Aprígio Veloso 625	Bodocongó
3	São Luís XIV	Ipiranga	R. Aprígio Veloso, 625	Bodocongó
3	Padre Cícero II	Ipiranga	Rua Manoel Mota, 635.	Bodocongó
3	São Joaquim	Alê	Av. Mal. Floriano Peixoto, 525	São José
3	São Luís III	Ipiranga	Rua Consul Joseph Noujain Habbi, 1001.	

Portanto, o que foi observado para a gasolina comum acontece com o etanol, o *cluster 3* tem o maior preço médio para o etanol, além dos maiores valores para todas as estatísticas apresentadas, como para a gasolina comum o menor preço médio se encontra no *cluster 1*.

A Figura 2 apresenta graficamente os *clusters* e onde está localizado cada um dos postos de combustíveis.

Figura 2 – Clusters



A Figura 2 indica a localização de cada posto de combustível nos *clusters*. Caso uma pessoa queira abastecer seu veículo com um desses dois combustíveis ela iria optar pelo os postos que estão no *cluster 1* onde estão os preços mais acessíveis. Pode-se observar que tem alguns postos do *cluster 2* que ficam bem próximo ao *cluster 1*.

#### 4. Conclusões

De maneira geral, após todas as análises feitas, pode-se concluir que o *cluster 1* com um total de 23 postos de combustíveis tiveram os preços mais acessíveis para a gasolina comum e o etanol, já o *cluster 3* apresentou os preços mais elevados, ou seja, os 29 postos que fazem parte deste *cluster* tinham os preços mais caros no em outubro de 2024, por fim o *cluster 2* composto por 9 postos esteve com os preços entre os *clusters 1* e 3. Assim, os postos que compõe o *cluster 3* são os que devem ser evitados para aqueles que pretendem economizar na hora de abastecer seu veículo.

Considera-se relevante explorar o seguinte ponto: o Fundo Municipal de Defesa dos Direitos Difusos (PROCON) de Campina Grande – PB poderia adotar a análise de cluster nas pesquisas mensais de preços de combustíveis, auxiliando a população a identificar os postos com os preços mais baixos. Essa abordagem contribuiria para que os consumidores economizassem nos gastos com o abastecimento de gasolina comum e etanol.

#### Referências

1. Campina Grande (2024). Pesquisa de combustíveis – outubro 2024. Procon Campina Grande. Disponível em: <https://procon.campinagrande.pb.gov.br/pesquisa-de-combustiveis-outubro-2024/>. Acessado em 15/12/2024
2. Doni, M. V. (2004). Análise de cluster: métodos hierárquicos e de particionamento. Universidade Presbiteriana Mackenzie.

3. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis*. *New Jersey*, 5(3), 207-219.
4. Petróleo, G. N. e. B. Agência Nacional do. (2017). *Cartilha do posto revendedor de combustíveis*.

RETRACTED