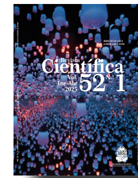
UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Revista Científica

http://revistas.udistrital.edu.co/ojs/index.php/revcie/index



Artículo de investigación científica

Account-Based Marketing in the Railway Sector: An Optimization Approach with K-means 9++w and Linear Programming

Marketing basado en cuentas en el sector ferroviario: un enfoque de optimización con K-means9++w y programación lineal

Marketing Baseado em Contas no Setor Ferroviário: Uma Abordagem de Otimização com K-means 9++w e Programação Linear

Carlos Eduardo Díaz Peñuela[1]
Javier-Antonio Ballesteros-Ricaurte[2]
Gustavo Cáceres-Castellanos[3]

Recibido: 10 de diciembre de 2024 **Aceptado:** 18 de septiembre de 2025

Para citar este artículo: Díaz, C. E., Ballesteros-Ricaurte, J. A. y Cáceres-Castellanos, G. (2025). Account-Based Marketing in the Railway Sector: An Optimization Approach with K-means 9++w and Linear Programming. *Revista Científica*, *52*(1), 51-72. https://doi.org/10.14483/23448350.23083

Resumen

Este artículo propone un enfoque optimizado para el *marketing* basado en cuentas (ABM) en el sector ferroviario, utilizando una variante del algoritmo K-means, denominada *K-means 9++w*, y la programación lineal. La metodología integra el modelo CRISP-DM y la estrategia de inicialización K-means++ para superar la ineficiencia computacional y la inicialización subóptima de los algoritmos tradicionales. Este estudio analizó datos de diversas fuentes, como la longitud de las vías férreas, el transporte de pasajeros y carga, y el PIB per cápita. Los resultados demuestran la efectividad de las distancias Chebyshov y Euclidiana ponderada en el procesamiento y la evaluación de variables. De manera crucial, la distancia Euclidiana ponderada propuesta mostró un rendimiento superior, logrando el coeficiente de silueta más alto (0.3807) y el índice Davies-Bouldin más bajo (0.8174) en comparación con las variantes de distancia tradicionales. La optimización de ABM a través de reglas automáticas produjo un ahorro de tiempo, mientras que la normalización de datos dispersos mejoró la coherencia y el rendimiento computacional. Las implicaciones del estudio indican el potencial de la campaña de construcción de vías férreas a nivel global y un potencial significativo para el sector ferroviario en Colombia. El análisis de los hallazgos, realizado con un conjunto de datos global, valida la aplicabilidad del modelo en la expansión de una empresa B2B más allá de su mercado local. Las limitaciones incluyen la disponibilidad de datos públicos y las restricciones de tiempo y presupuesto.

Palabras clave: K-means, programación lineal, industria ferroviaria, *marketing* basado en cuentas, gestión de *marketing*, optimización

1. Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia. carlos.diazpenuela@uptc.edu.co
2. Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia. javier.ballesteros@uptc.edu.co
3. Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia. gustavo.caceres@uptc.edu.co

Abstract
This article proposes an optimized approach for account-based marketing (ABM) in the railway sector, utilizing a variant of the K-means algorithm, dubbed *K-means 9++w*, as well as linear programming. The methodology integrates the CRISP-DM model and the K-means++ initialization strategy to overcome computational inefficiency and suboptimal initialization of traditional algorithms. This study analyzed data from various sources, such as railway track length, passenger and freight transport, and *per capita* GDP. The results demonstrate the effectiveness of Chebyshov and weighted Euclidean distance metrics in data processing and variable evaluation. Crucially, the proposed weighted Euclidean distance showed a superior performance, achieving the highest silhouette coefficient (0.3807) and the lowest Davies-Bouldin index (0.8174) in comparison with traditional distance variants. ABM optimization through automatic rules yielded time savings, while the normalization of scattered data improved coherence and computational performance. The implications of this study indicate the global potential of the railway construction campaign and a significant potential for the railway sector in Colombia. Our analysis of the findings, carried out with a global dataset, validates the model's applicability to the expansion of a B2B company beyond its local market. The limitations include the availability of public data and time and budget constraints.
Keywords: K-means, linear programming, railway industry, Account Based Marketing, ABM, marketing management, optimization.

Resumo
Este artigo propõe uma abordagem otimizada para o Marketing Baseado em Contas (ABM) no setor ferroviário, empregando uma variante do algoritmo K-means, denominada K-means 9++w, e a programação linear. A metodologia integra o modelo CRISP-DM e a estratégia de inicialização K-means++ para superar a ineficiência computacional e a inicialização subótima dos algoritmos tradicionais. O estudo analisou dados provenientes de diversas fontes, incluindo a extensão da malha ferroviária, o transporte de passageiros e carga, e o PIB per capita. Os resultados demonstram a eficácia das distâncias Chebyshev e Euclidiana ponderada no processamento e na avaliação de variáveis. De modo crucial, a distância Euclidiana ponderada proposta exibiu um desempenho superior, alcançando o coeficiente de silhueta mais alto (0.3807) e o índice Davies-Bouldin mais baixo (0.8174) em comparação com as variantes de distância convencionais. A otimização do ABM por meio de regras automáticas gerou uma economia de tempo, enquanto a normalização de dados dispersos aprimorou a coerência e o desempenho computacional. As implicações do estudo indicam o potencial da campanha de construção de vias férreas em nível global e um potencial significativo para o setor ferroviário na Colômbia. A análise dos achados, realizada com um conjunto de dados global, valida a aplicabilidade do modelo na expansão de uma empresa B2B para além do seu mercado local. As limitações incluem a disponibilidade de dados públicos e as restrições de tempo e orçamento.
Palavras-chaves: K-means, programação linear, setor ferroviário, Marketing Baseado em Contas (ABM), gestão de *marketing*, otimização

INTRODUCTION

Account-based marketing (ABM) is a strategic approach that focuses on identifying and strengthening relationships with individual, high-value customer accounts. The goal is to create mutual benefit and build long-term relationships, rather than focusing on individual transactions with a large number of customers (Gene Day & Wei Shi, 2020). It is a strategic approach that starts by defining a clear business ambition for an account. It then analyzes the account's needs to prioritize the best opportunities, identifies target audience segments, and develops a compelling proposition and messaging to attract them. Finally, it involves designing the right marketing mix, including campaigns and experiences, and collaborating with sales and other teams to deliver on the business ambition (Burgess, 2025). In the business to business (B2B) context, these consumers are usually companies or governments, which implies larger sales when compared to marketing aimed at end consumers.

This relational approach is particularly relevant in the B2B field, where strategic decision-making requires agility and is favored by the existence of trusting relationships. However, the construction of these relationships faces several challenges that, according to contemporary authors (Saura *et al.*, 2021; Verma *et al.*, 2021; Ling & Weiling, 2025), have not yet been sufficiently explored. Among them, the complexity derived from the number of customers and employees stands out, which can prolong sales times and hinder access to technological tools, especially in companies with limited resources (Gene Day & Wei Shi, 2020). In this scenario, the adoption of information technologies and artificial intelligence (AI) has transformed the B2B ecosystem (Saura *et al.*, 2021). AI, through clustering algorithms like K-means 9+ (Abdulnassar & Nair, 2023), allows automating processes such as customer segmentation. By grouping customers into clusters based on similarities, this algorithm facilitates the implementation of differentiated marketing strategies. Linear programming (Vaquer, 2023) complements the use of clusters by optimizing input variables in order to maximize return on investment (ROI). This study focuses on the application of variants of the K-means 9++w algorithm and linear programming in the B2B railway industry, with special emphasis on the weighted Euclidean distance for segmentation. This distance stands out for its ability to correct errors in the data, as well as for its suitability for principal components analysis (Buzzell *et al.*, 2022). Moreover, the use of K-means++ (Makarychev *et al.*, 2020; Ling & Weiling, 2025) is suggested, as it can optimize the results thanks to a better initialization of the centroids. This approach is particularly relevant given the need to use distance metrics that account for correlations between variables, as highlighted in recent works (Li *et al.*, 2025). The weighted Euclidean distance is a special case of the Mahalanobis distance, an advanced metric that considers the correlation between variables. Our methodology combines three key aspects: the initialization of centroids with K-means++ (++), the optimization of cluster assignment with the K-means 9+ logic (9+), and the use of a weighted Euclidean distance (w) to represent the relative importance of business variables. By integrating these improvements, we propose the K-means9++w algorithm to improve customer segmentation. Considering the high market potential for electric locomotives in Colombia and the existence of a niche rail market in Boyacá (Gobierno de Colombia, 2020), this study is relevant in the current context, where the search for more sustainable transportation solutions is crucial. While based on a global dataset, our analysis demonstrates the optimization model's potential for expanding a B2B company beyond its local market. In the railway sector, where inventory turnover is very low, a company's long-term sustainability depends on its ability to seek and capture international markets. Therefore, the use of a global dataset should not be regarded as a limitation, but as a virtue that validates the model's applicability in the context of a real business with a view towards expansion.

LITERATURE REVIEW

Drawing on the literature and studies mentioned in the introduction section, we can infer a clear research gap in the application of data science methodologies for quantitative optimization in niche B2B markets. While advancements in clustering algorithms and frameworks for AI in marketing have been documented, there is a lack of integrated approaches that combine a specific clustering method with linear programming in order to optimize customer segmentation. Furthermore, existing models often overlook the need for a weighted distance metric that can account for the varying strategic importance of client features. Our study directly addresses these gaps by proposing an innovative methodology that integrates the K-means9++w algorithm and linear programming. The introduction of a weighted distance metric is a key contribution, allowing our model to create customer segments that are not only statistically valid, but also strategically relevant. Table 1 serves as a structured overview of the key literature and highlights how our research fills these specific gaps.

Table 1. *Literature review and positioning of this study*

Author-year	**Methodology/key concept**	**Limitations**	**Contributions**
Gene Day and Wei Shi (2020)	Account-based marketing (ABM): A strategic approach for B2B marketing	It focuses on qualitative analysis and best practices, without a detailed quantitative optimization methodology for customer segmentation.	Proposes ABM as a key strategy for the growth of B2B startups, justifying the importance of focusing resources on high-value accounts.
Gobierno de Colombia (2020)	Rail market analysis: A strategic government document	The information is contextual and does not present a clustering or AI model optimization methodology.	Provides the socioeconomic and political context for the research, justifying the relevance of studying the electric locomotive market in the Colombian rail industry.
Makarychev *et al.* **(2020)**	K-means++: Advanced algorithm analysis	This work focuses on a theoretical analysis. The base algorithm is still limited by inefficiency in cluster assignment for large datasets.	Provides a more robust theoretical justification and improved approximation guarantees for the algorithm.
Saura *et al.* **(2021)**	Artificial intelligence in B2B marketing: Literature review on the impact of AI on CRM systems	This study focuses on reviewing the literature and does not present a new methodology; it does not address a specific optimization problem in a practical way.	Provides an overview and a framework for future research on how AI is transforming B2B marketing.
Verma *et al.* **(2021)**	AI in marketing: Systematic review of AI in marketing	This study is theoretical and does not present a practical implementation. Its findings are based on a systematic review and not on experimentation.	Offers a comprehensive view of the trends and challenges of using AI in marketing, identifying areas for future research.
Abdulnassar and Nair (2023)	K-means9+: A variant of the cluster assignment step	The benefits presented herein are significant only for a large number of clusters (k). For small values, the additional complexity is not justified.	Increases the computational efficiency of the K-means algorithm, allowing for better performance in scenarios with a large number of clusters.
Vaquer (2023)	Linear programming: An optimization method for decision-making.	The methodology is not directly related to customer segmentation or marketing strategies.	Provides a solid foundation for resource optimization and decision-making, maximizing return on investment (ROI) in production contexts.

Lew Sook Ling *et al***.** **(2025)**	K-Means++: A comparative study to improve customer segmentation in e-marketing	This study is limited to a selection of clustering methods and does not explore the generalization of its findings to other industries. Scalability is a limitation for large datasets.	The study demonstrates the superior performance of K-means++ in segmentation accuracy when compared to other methods like K-means, K-medoids, and Gaussian mixture models (GMM). It emphasizes the importance of validation metrics such as the silhouette score and the Davies-Bouldin index.
Li *et al***. (2025)**	Mahalanobis distance: A method that uses Mahalanobis distance for fault localization in smart electrical substations	The methodology is focused on a very specific domain (electrical engineering) and not directly on marketing or customer segmentation.	It demonstrates the ability of the Mahalanobis distance to manage the correlation between variables and improve accuracy in fault localization, resulting in a high accuracy value (92%).
This study	K-means (K-means++, K-means 9+) with weighted Euclidean distance: A comprehensive approach that combines efficient initialization and assignment with a personalized distance metric	Performance depends on the accuracy of the weights vector, which must be defined by means of business knowledge.	It introduces a computationally efficient variant of the Mahalanobis distance, optimized for normalized data; combines K-means++ and K-means9+ for fast execution; and allows for clustering to reflect the importance of business features.

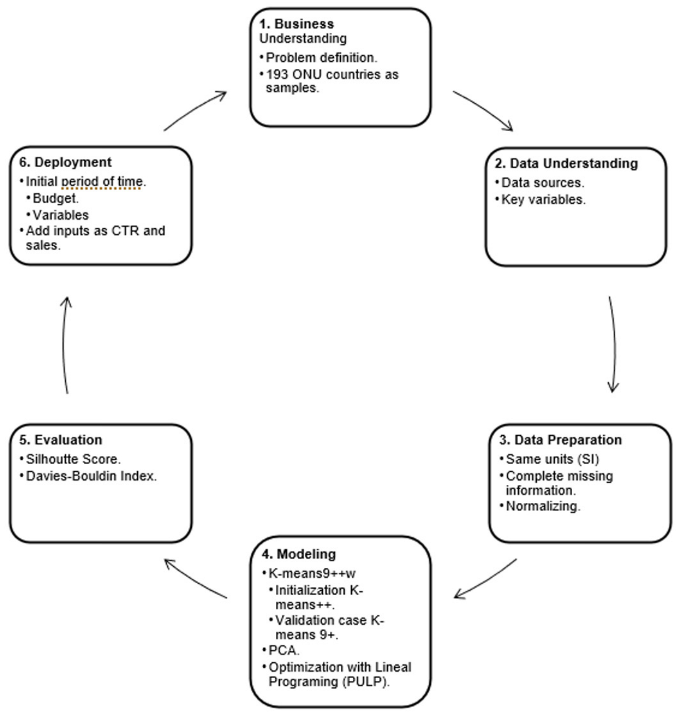
Although our contribution is focused on a practical advance rather than a groundbreaking innovation, the following points highlight the uniqueness and relevance of this study:

- **Methodological integration.** We propose a novel combination of methodologies, integrating an original variant of the K-means algorithm (K-means 9++w) with linear programming. This approach provides a new framework for optimizing segmentation in ABM in the B2B sector.
- **Computational efficiency.** Our results demonstrate the effectiveness and efficiency of this methodology. By using automatic, rule-based optimization and normalizing disparate data, we not only achieve greater consistency in clustering; we also offer a solution that saves time and resources, which are critical for companies operating in niche markets.
- **Practical applicability.** A key contribution of this study is the validation of the Chebyshov and weighted Euclidean distances, proving their effectiveness for data handling and variable evaluation. This methodological contribution not only strengthens the robustness of our clustering model but also makes it more intuitive and adaptable to the subjective needs of a business. It is important to note that our weighted Euclidean distance serves as a computationally efficient variant of the Mahalanobis distance, allowing variables to be weighted based on their strategic importance without the high computational cost that is typically involved in this process.

METHODOLOGY

This research adapted the methodology proposed by Hernández Sampieri *et al.* (2014), which is framed within the phases of the cross-industry standard process for data mining (CRISP-DM) (Ayele, 2020; Figueroa *et al.*, 2023). This methodology is widely used in data mining projects. Its phases include (i) business understanding, (ii) data understanding, (iii) data preparation, (iv) modeling, (v) evaluation, and (vi) deployment.

Figure 1. *Proposed methodological architecture for customer segmentation and marketing optimization*



RESULTS

Business understanding

Our research design was based on an experiment with controlled independent variables, groups, and measurable dependent variables. We employed a sample of 193 member countries of the United Nations (UN, 2024), representing a broad universe of potential customers. For each country, three sets of information were collected: railway data, GDP *per capita*, and percentage of Internet access. The sample selection was non-probabilistic, according to the criteria established for this research. The reason for choosing these countries was the public availability of most of their information, including that related to the railway industry. The central hypothesis of this work was as follows:

It is possible to define the ABM process of a multinational railway company through variants of the K-means algorithm and linear programming optimization.

Data understanding

Collection and definition of variables

The independent variables were obtained from government sources such as the CIA (2024), the UNECE (2024), the World Bank (2024a, 2024b, 2024c), and commercial databases such as Word Population Review (2024). For the specific case of Colombia, additional sources such as the 2020 Railway Master Plan (Gobierno de Colombia, 2020) and the Ministry of Transportation (2019) were used.

Independent variables

The following independent variables were considered:

- **a. CIA_WFB_leng_railway_km:** length in km of the country's railways according to CIA (2024).
- **b. UNECE_avg_delta_per_year_2005-2022_leng_railway_km:** delta total, calculated using the difference between the minimum and maximum number of railways per year over the total number of years. This was calculated for each country, in km, and according to UNECE (2024).
- **c. WPR_rail_passengers_2022_M:** passengers per train in 2022, in millions, according to World Population Review (2024).
- **d. WPR_rail_passenger-kilometers_transported_2022_M:** unit of measurement used in transportation that represents a passenger transported for one kilometer (pkm)—in this case, during the year 2022. The units of this variable are in millions of pkm, according to that reported by World Population Review (2024).
- **e. UNECE_avg_2005-2022_goods_carriage_rail_M_tkm:** unit of measurement representing a ton transported per km. We used the average for the 2005-2022 period, as reported by UNECE (2024).
- **f. UNECE_avg_2005-2022_passenger_transport_rail_M_pkm:** unit of measurement representing the number of passengers transported per km. We used the average for the 2005-2022 period, as reported by UNECE (2024).
- **g. UNECE_avg_2005-2022_railway_density_km_/_square-km:** railway density in $\frac{km}{km^2}$, as reported by UNECE (2024). For the countries without information in UNECE (2024) we used the length in km of the country's railways CIA (2024) over the country surface in km^2, as reported by World Bank (2024a).
- **h. WB_most_recent_PBIPC:** GDP *per capita* for each country (in dollars), according to the World Bank (2024b).
- **i. WB_use_of_Intenet_population_%:** percentage of Internet use among the country's inhabitants according to the World Bank (2024c).

Campaigns and budget

Four marketing campaigns were defined: electric locomotives, passenger cars, freight cars, and railroad track construction (km). The budget range for the campaigns was established in dollars, based on research conducted by Forrester Research (2019).

Profit and utility

The profit of each campaign was defined using approximate prices for electric locomotives, freight cars, km of railroad track built, and passenger cars, as obtained from various sources such as RENFE (2023), the Colombian National Planning Department (2020) and Findeter (2020).

Variable control

Variable control focused on the algorithm’s execution environment, keeping the same source code and railway data file in all executions.

K-means 9++w clustering algorithm

The K-means 9++w clustering algorithm was selected because of its ability to group the countries database into clusters with similar characteristics, which allowed designing personalized marketing campaigns. In addition, K-means facilitated the experimental modification of the distance variable. Through iterations, the algorithm minimized the distance between the data and their assigned centroids.

This study addresses two key limitations of the traditional K-means algorithm: suboptimal initialization and computational inefficiency. To overcome the first issue, we adopted the K-means++ initialization strategy, which is widely recognized for its ability to select well-dispersed initial centroids, thereby accelerating convergence and leading to more robust clustering outcomes. The effectiveness of this approach has been thoroughly analyzed in the literature, demonstrating its superiority in terms of solution quality and stability (Ling & Weiling, 2025).

As for the second limitation, we integrated a highly efficient cluster assignment step inspired by the K-means 9+ algorithm, introduced by Abdulnassar and Nair (2023), who extensively analyzed its performance in cases with a large number of clusters ($k>9$). While our specific implementation may not utilize a high cluster count, the proposed K-means 9++w algorithm was designed to conditionally adapt to large numbers of clusters. When $k \leq 9$, the algorithm executes the traditional K-means assignment, but it is architecturally prepared to handle higher cluster counts by adapting the K-means 9+ logic.

The steps of the algorithm are as follows:

1. Centroid initialization, which utilizes K-means++
2. Each data point is assigned to the nearest centroid, forming clusters
3. The centroids are recalculated as the average of all points in the cluster
4. Steps 2 and 3 are repeated iteratively until the centroids no longer change or a maximum number of iterations is reached

For total utility optimization, we employed linear programming based on the results of each variant of the K-means algorithm.

Variants of distance calculation within the K-means algorithm

The following variants were used for distance calculation:

- **Euclidean distance**, based on the Pythagorean theorem
- **Chebyshov distance**, which measures the maximum absolute value of the difference between point coordinates
- **Manhattan distance**, which measures the sum of the absolute differences between the corresponding coordinates of two points
- **Weighted Euclidean distance**, a variant of Euclidean distance that introduces a vector with weights for each independent variable, with the aim of prioritizing certain variables

While traditional K-means algorithms rely on Euclidean distance, this study highlights the need for advanced metrics. Recent research shows that Mahalanobis distance effectively accounts for the correlations between variables, improving clustering accuracy (Li *et al.,* 2025). Building on this concept, our methodology's innovation lies in its use of weighted Euclidean distance, which allows the model to reflect the subjective importance of business variables. Mathematically, this metric is a special case of the Mahalanobis distance from a matrix perspective, but it is innovative in its use of a subjective parameter defined by business rules. The metric, formally defined with a weight vector *w*, can be generalized using matrix notation. However, for the sake of simplicity, an example with two dimensions is shown in Equation (1):

$$D(x, y, w) = \sqrt{\sum_{i=1}^{n} w_i \left(x_i - y_i\right)^2} \tag{1}$$

Our approach simplifies the Mahalanobis distance by applying a diagonal weight matrix, making it computationally lighter and more interpretable while still addressing the need to account for variable correlations.

Computational and operational complexity

The K-means 9++w algorithm was designed for both theoretical rigor and practical scalability. The time complexity of the traditional K-means algorithm is $O(n \cdot k \cdot I \cdot d)$, where n is the number of data points, k is the number of clusters, I is the number of iterations, and d denotes dimensionality. Our implementation optimizes this performance as follows:

- **Initialization.** The K-means++ strategy significantly reduces I, leading to a net reduction in computation time.
- **Cluster management.** The algorithm is designed to efficiently handle a large number of clusters. Although k is less than 9 in this study, the complexity of the cluster assignment adapts to k and is architecturally prepared for cases where $k > 9$. This is achieved through the conditional adaptation of the K-means 9+ logic.
- **Weighted distance.** The weighted Euclidean distance, a special case of the Mahalanobis distance, reduces the computational complexity of the original metric. The Mahalanobis distance requires a full covariance matrix and has a complexity of $O(d^3)$, but our metric avoids this operation with a simple element-wise multiplication by the weight vector, making it viable for business applications.

Therefore, this design ensures that our methodology is not only theoretically valid but also scalable for large datasets and viable for real-world business applications

Data preparation

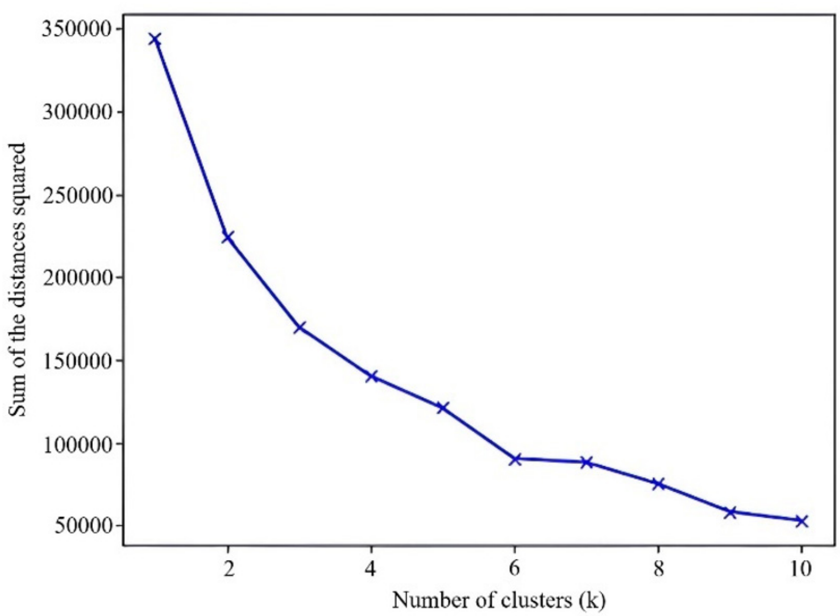
This phase involved several steps aimed at ensuring that the data were suitable for clustering analysis. First, the data, collected from various sources, were standardized to the International System of Units (SI)—any measurements in miles were converted to kilometers. This was done to ensure uniformity and consistency across all variables. A key challenge was handling missing values, which were numerous in some cases (*e.g.*, railway length in countries without a railway system). Instead of using a traditional imputation method, which could have introduced noise, these null values were intentionally converted to zeros, as this accurately represented the absence of a given feature in a country. Next, the data were normalized by means of a *min-max* scaling technique, transforming all numerical features to a common range of [0, 100]. This step was critical for preventing variables with larger scales, such as the GDP *per capita*, from disproportionately influencing the clustering algorithm, thereby ensuring that each variable contributed equally to the distance calculations.

Modeling

Determining the optimal number of clusters

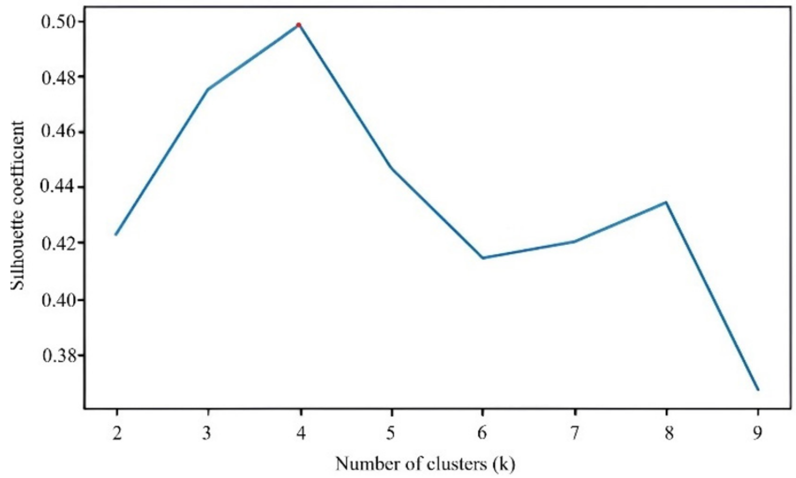
Initially, an attempt was made to use the Jambú elbow method (Chikumbo & Granville, 2019) to determine the optimal number of clusters. However, this technique was discarded due to subjectivity in the interpretation of *inertia deceleration* and the variability of the results in different runs on the same data. Figure 2 shows the results of the test performed on the Jambú elbow method applied to the normalized data.

Figure 2. *Jambú elbow method applied to normalized data*



Finally, an algorithm was implemented in Python, using the *K-means* class from the *sklearn.cluster* package and the *silhouette_score* class from *sklearn.metrics*, in order to find the silhouette coefficient (Lenssen & Schubert, 2023), varying the number of clusters between 2 and 10. The results indicated that the optimal number of clusters is 4, with a maximum average silhouette coefficient of 0.498282. Figure 3 depicts the result of this analysis.

Figure 3. Silhouette coefficient method applied to normalized data



Initializing the centroids

In our methodology, the cluster centroids were initialized using the K-means++ strategy. Unlike random initialization, which can lead to suboptimal results and slow convergence, K-means++ selects initial centroids that are well-dispersed across the dataset. This smart seeding process, which first selects a random point and then chooses subsequent centroids based on their distance from existing ones, helps the algorithm to converge more quickly and consistently to a better solution (Ling & Weiling, 2025).

Implementation of the K-means algorithm

The K-means algorithm was implemented in Python and executed four times, once for each distance variant. Table 2A shows an example of the resulting assignment of clusters per country, while Tables 1B to 1E show the count of countries per cluster for each distance variant (B: Euclidean, C: Chebyshov, D: Manhattan, E: weighted Euclidean).

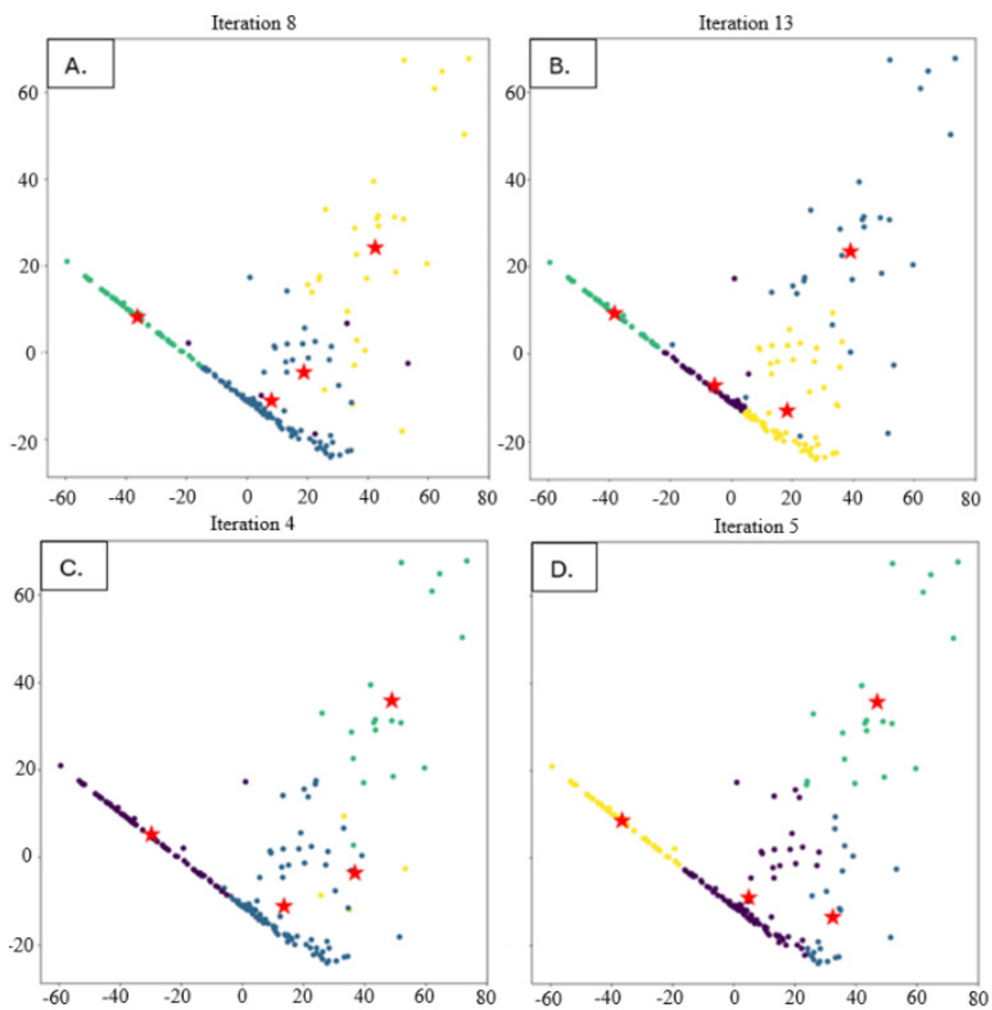
Table 2. *Cluster assignment and country counts per cluster in each variant*

A.		B.		C.		D.		E.	
Country	Cluster	Cluster	Country count	Cluster	Country count	Cluster	Country count	Cluster	Country count
0	0	1	102	3	62	1	94	0	94
1	3	2	58	2	52	0	77	3	57
...	...	3	28	0	50	2	18	1	23
192	2	0	5	1	29	3	4	2	19

Principal components analysis

With the clusters assigned by each K-means variant, a principal components analysis (PCA) (Buzzell et al., 2022) was performed using the *PCA* library of the *sklearn.decomposition* Python package. This process identifies the key directions in the data that capture the maximum variance. These directions are new, uncorrelated axes that represent the most significant features of the dataset, enabling an efficient representation of the data in fewer dimensions. This analysis transformed the initial nine dimensions of each country into two components, with the aim of plotting the relationships between the countries and their clusters in two dimensions. The results of the PCA are shown in Figure 4.

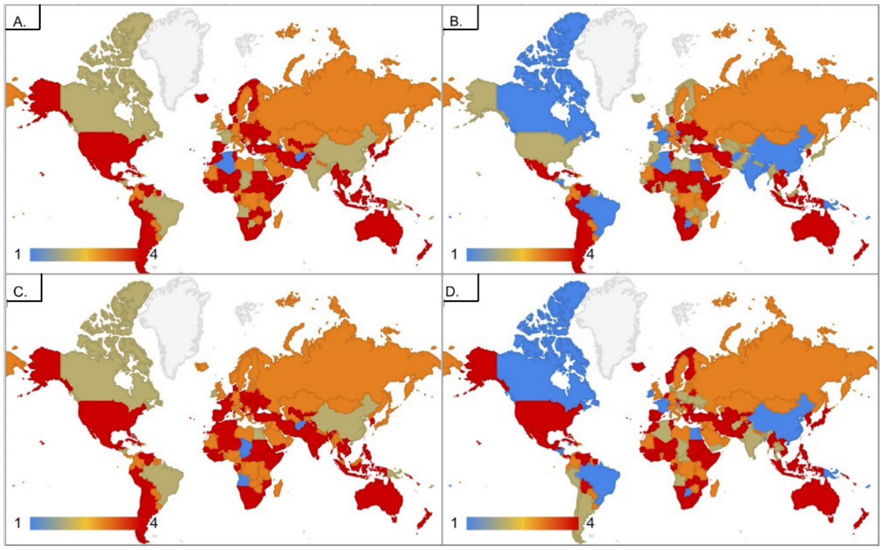
Figure 4. *PCA, result plots of the resulting clusters (distance variants – A: Euclidean, B: Chebyshov, C: Manhattan, D: weighted Euclidean)*



Visualizing the clusters in world maps

To visualize the clusters for each K-means variant, world maps were created (Figure 5). The clusters were ordered ascendingly by the number of countries, and a color was assigned to each cluster in order to facilitate further analysis (1: blue, 2; green, 3: orange, 4: red; A: Euclidean, B. Chebyshov, C. Manhattan, D. weighted Euclidean).

Figure 5. *World maps by cluster*



Allocation of marketing campaigns

In our method, a Python algorithm adds the mean values of the variables of each cluster to a row of a matrix. These rows are arranged in descending order according to their priority, *i.e.*, GDP, railroad length, railroad density, passenger transport, freight transport, and percentage of Internet usage. Then, in order to assign a campaign, the algorithm compares the average value of the variables of each cluster against the average value of the variables of all the countries (Table 3).

Table 3. *Average values for the matrix of country characteristics*

a.	b.	c.	d.	e.	f.	g.	h.	i.
3.25	3.34	2.63	2.26	2.25	3.31	9.32	8.63	63.79

Campaign allocation was prioritized as follows:

1. Electric locomotives
2. Passenger cars
3. Freight cars
4. Construction of railroad tracks (km)

The algorithm searches for the first cluster that meets the conditions defined for each campaign and assigns it accordingly. To ensure a transparent and reproducible prioritization, the logical rules shown in Table 4 were applied based on the average characteristics of each cluster in comparison with the global mean. This approach clearly delineates the decision-making process for assigning a specific campaign to a cluster.

Table 4. *Logical rules for marketing campaign allocation*

Condition (cluster characteristic>global mean)	**Assigned marketing campaign**
High rail density & high GDP *per capita*	Railway construction
High passenger transport & high GDP *per capita*	Passenger railcars
High freight transport & low GDP *per capita*	Freight railcars
Low rail density & low GDP *per capita*	Electric locomotives

Linear programming optimization

The budget range for the campaigns was defined in dollars (Forrester Research, 2019): maximum spend = $600 000; minimum spend = $350 000; minimum spend per cluster = (minimum spend/4) - $50 000. The campaign profit per dollar invested was calculated as 1% of the selling price of each product (Findeter, 2020; RENFE, 2023; *Colombian National Planning Department, 2020*), which yielded the following values: electric locomotives = $2 500 000 * 1%; passenger cars = $1 508 000 * 1%; freight cars = $303 000 * 1%; railroad construction = $48 280 000 * 1%. To perform linear programming optimization, the *PuLP* Python library was used (Universidad Oberta de Catalunya, 2024). The optimization problem consisted of predicting and guiding the expected utility maximization of ABM by varying the investments of each campaign. The constraints included the minimum and maximum total expenditure as well as the minimum expenditure per cluster.

Evaluation

To quantitatively assess the performance of the K-means 9++w algorithm and its distance variants, we used two distinct metrics for internal cluster validation: the silhouette coefficient (higher is better) (Lenssen & Schubert, 2023) and the Davies-Bouldin index (lower is better) (Rubiños *et al.*, 2024). These two metrics measure the cluster compactness and separation. Both metrics were seamlessly adapted to the weighted Euclidean distance by replacing the standard distance formula with our custom weighted expression. The results are summarized in Table 5.

Table 5. *Results by distance variant for the metrics used for internal cluster validation*

Distance variant	**Silhouette** Coefficient	**Davies-Bouldin Index**
Euclidean	0.2796	1.1098
Chebyshov	0.1749	1.5424
Manhattan	0.1983	1.4870
Weighted Euclidean (our proposal)	**0.3807**	**0.8174**

Our proposed weighted Euclidean distance exhibited a superior performance, with the highest silhouette coefficient and the lowest Davies-Bouldin index. This quantitative evidence supports our claim that business-driven variable weighting significantly improves segmentation quality. Our findings are consistent with those of Ling and Weiling (2025), who also found K-means++ to be the best-performing algorithm for segmentation tasks.

Euclidean distance

The largest cluster, with 102 countries, including the United States of America and Uzbekistan, was formed using the Euclidean distance. Despite being the variant that exhibited the highest average standard deviation, with 11.95, this figure indicates a good similarity between the countries in the cluster. The largest deviations within this cluster correspond to the freight and passenger transport variables, and, in general, the countries in this group have a high railway density. The other clusters generated with the Euclidean distance are analyzed below.

- **Cluster 0.** Composed of 5 countries, it exhibits the highest standard deviation and contains the highest values in the aforementioned variables. Order by number of countries: 1, standard deviation: 30.24, campaign assigned: electric locomotives.
- **Cluster 1.** It gathers 102 countries and is characterized by a low standard deviation and a high average GDP per capita. Order by number of countries: 4, standard deviation: 3.75, campaign assigned: passenger railcars.
- **Cluster 2.** With 58 countries, it also has a low standard deviation and a high average GDP per capita. Order by number of countries: 3, standard deviation: 1.55, campaign assigned: railway construction.
- **Cluster 3.** Formed by 28 countries, it exhibits a medium to high standard deviation and a high average rail density. Order by number of countries: 2, standard deviation: 12.26, campaign assigned: railway construction.

Optimizing investment in ABM campaigns to maximize utility yielded the following distribution: $183 333.33 in electric locomotives, $116 666.67 in passenger cars, $0.0 in freight cars, and $300 000.0 in railroad track construction.

Chebyshov distance

This variant generated the most even distribution of clusters, with China and Burundi grouped in the smallest cluster size. With an average standard deviation of 7.48, the lowest of the four distances, this variant exhibited the highest similarity among the clustered countries. The largest deviations were observed for the railroad density and passenger transport variables in one of the clusters. In general, the clusters formed with this distance showed a high rail density. Each cluster generated with the Chebyshov distance is described below.

- **Cluster 0.** With 50 countries, it has small deviations and a high average GDP *per capita*. Order by number of countries: 2, standard deviation: 2.1, campaign assigned: electric locomotives.
- **Cluster 1.** The least numerous cluster, with 29 countries. It shows a high standard deviation compared to the other three clusters and a high average value for railroad density. Order by number of countries: 1, standard deviation: 20.48, campaign assigned: railway construction.

- **Cluster 2.** It comprises 52 countries and is characterized by a low standard deviation and a high average GDP *per capita*. Order by number of countries: 3, standard deviation: 1.36, campaign assigned: railway construction.
- **Cluster 3.** It groups 62 countries, has a medium to low standard deviation, and a high average GDP *per capita*. Order by number of countries: 4, standard deviation: 5.96, campaign assigned: railway construction.

To maximize profit, the investment in ABM campaigns should be distributed as follows: \$300 000.0 in electric locomotives, \$0.0 in passenger cars, \$0.0 in freight cars, and \$300 000.0 in railroad track construction.

Manhattan distance

The Manhattan distance produced the smallest cluster among the four variants, which includes countries such as Belgium and Afghanistan. With an average standard deviation of 10.15, the second highest value, this distance still indicates good similarity between the clustered countries. The largest deviations were observed for railroad length and freight transport in one of the clusters. In general, the clusters formed with this distance were characterized by high average values for the variables related to railroad length and its annual variation. Each cluster generated with the Manhattan distance is described below.

- **Cluster 0**. Composed of 77 countries, it shows small deviations and a high average value for the passenger transport variable. Order by number of countries: 3, standard deviation: 5.07, campaign assigned: electric locomotives.
- **Cluster 1**. The largest cluster, with 94 countries, exhibits small deviations and a high average GDP *per capita*. Order by number of countries: 4, standard deviation: 8.66, campaign assigned: passenger railcars.
- **Cluster 2**. With 18 countries, it is characterized by a low standard deviation and a high average value for railroad density. Order by number of countries: 2, standard deviation: 9.15, campaign assigned: freight railcars.
- **Cluster 3**. With only four countries, it is the smallest cluster in the whole experiment. It has a high standard deviation compared to the other three clusters and a high average value for the length of railroad tracks. Order by number of countries: 1, standard deviation: 17.71, campaign assigned: railway construction.

The optimization carried out to maximize utility suggested the following distribution of investment in ABM campaigns: \$183 333.0 in electric locomotives, \$116 667.0 in passenger cars, \$116 667.0 in freight cars, and \$183 333.0 in railroad track construction.

Weighted Euclidean distance

The weighting vector prioritized the variables with available information, counting only non-null values in each variable. This strategy minimized the impact of missing data on cluster formation and enabled a more accurate analysis of the variables with higher weights when running the K-means algorithm. Specifically, increasing the weight of a variable increased its influence on the calculation of the distance between countries and their centroids. For example, if two countries coincided in all but two variables, one with a low weight and the other with a high weight, the latter would have a more significant impact on cluster assignment. This favors the formation of clusters specializing in variables with high weights.

Table 6 shows the vector of weights used in this study for all the independent variables listed in the data understanding phase of the methodology section.

Table 6. *Values for the vector of weights per independent variable*

a.	b.	c.	d.	e.	f.	g.	h.	i.
132	44	64	65	49	49	132	193	192

This weighted Euclidean distance generated the second most equal distribution of countries per cluster among the four variants. The smallest cluster included countries such as China and Fiji. With an average standard deviation of 8.56, the second lowest value, this distance indicated a good similarity between the clustered countries. The largest deviations were observed for the variables related to the annual change in railroad track length and freight transport in one of the clusters. In general, the clusters formed with this distance exhibited a high average value in the variable related to railroad density. Each cluster generated with the weighted Euclidean distance is described below.

- **Cluster 0.** The largest cluster, with 94 countries, has small deviations and a high average GDP *per capita*. Order by number of countries: 4, standard deviation: 10.54, campaign assigned: electric locomotives.
- **Cluster 1.** It comprises 23 countries and is characterized by a high mean, high standard deviations, and a high average GDP *per capita*. Order by number of countries: 2, standard deviation: 19.15, campaign assigned: passenger railcars.
- **Cluster 2.** The least numerous cluster, with 19 countries, it shows medium to low standard deviations and a high average value for railroad density. Order by number of countries: 1, standard deviation: 22.55, campaign assigned: railway construction.
- **Cluster 3.** It groups 57 countries and exhibits a low standard deviation and a high average value for pkm. Order by number of countries: 3, standard deviation: 4.54, campaign assigned: railway construction.

To maximize profit, the investment in ABM campaigns should be distributed as follows: $183 333.0 in electric locomotives, $116 667.0 in passenger cars, $0.0 in freight cars, and $300 000.0 in railroad track construction.

Deployment

The proposed model is ready for use in a real railway ABM environment. To this effect, we suggest that it be deployed in an initial time period, with its own budget and initial variables defined. As periods are iterated, the model should be complemented with the results obtained, such as the click-through rate (CTR) and sales results. Our analysis of the international market, with an emphasis on the case of a Colombian company, constitutes a real starting point for a railway ABM strategy and its optimization.

DISCUSSION

This study explored the application of variants of the K-means algorithm and linear programming to optimize ABM in a hypothetical railroad company. The results obtained, based on the analysis of data from 193 countries, allowed for a dialogue with previous research and provided new insights into the application of these techniques in the context of B2B ABM marketing.

The importance of data normalization confirms what was pointed out by Cuevas-Díaz Durán *et al.* (2024), who emphasized its impact on the comparison of data from heterogeneous sources. In this study, normalization was especially relevant due to the heterogeneity of the variables used.

The effectiveness of the Chebyshov distance in error correction coincides with that reported by Bereg *et al*. (2024), who used this variant for noise correction in communications. However, this study extends this perspective by demonstrating its usefulness in customer segmentation for ABM, especially in the correction of unimportant or missing data. The weighted Euclidean distance stands out as an unexplored alternative in the existing literature. The effectiveness of our weighted distance approach in improving segmentation precision suggests that, similar to the findings of Li *et al*. (2025) regarding the Mahalanobis distance, incorporating a more sophisticated view of data structure is crucial for achieving better clustering results.

Beyond its technical rigor, a key contribution of our proposed methodology is its significant enhancement of operational efficiency in B2B marketing. Traditional ABM relies on manual, time-intensive processes for customer segmentation and budg*et al*location, which are often subject to human bias and delayed decision-making (Saura *et al*., 2021). Our integrated approach, which automates these tasks by combining K-means clustering and linear programming, drastically reduces the time required for strategic planning. The system can process large datasets and generate an optimized, data-driven budg*et al*location plan in a matter of minutes, a process that would otherwise take days or even weeks for marketing teams. This not only allows companies to respond to market changes with greater agility but also ensures that marketing resources are allocated for maximum utility, thereby increasing the potential for return on investment.

The global potential of the railroad track construction campaign contrasts with the more limited potential of the freight railcar campaign. This difference can be attributed to distinct market drivers. Railway construction represents a long-term strategic investment, often initiated and funded by government policies and public-private partnerships aimed at developing the national infrastructure (Gobierno de Colombia, 2020). This type of investment is driven by a country's need for efficient and sustainable public transport and trade routes (Fabri *et al*., 2024; Ren *et al*., 2024). In contrast, the acquisition of freight railcars is typically a more specific, short-term investment decision made by private companies. This decision is highly sensitive to market competition from other transportation modes, such as road freight, which often exhibit lower barriers to entry and greater flexibility (Karam *et al*., 2023). The opportunities for the rail sector in Colombia and Boyacá, which were identified in this study, are framed within a political-environmental context that is favorable for investment in rail infrastructure (Gobierno de Colombia, 2020; Ministry of the Environment, 2024). However, it is necessary to consider the challenges faced by the sector, such as the need to modernize the existing infrastructure and improve competitiveness vis-à-vis other modes of transportation (Karam *et al*., 2023).

CONCLUSIONS

This study faced some limitations, such as the absence of data for certain countries in the openly available sources used due to budget constraints. Additionally, the selection of algorithms and variables could have been broader in order to provide more robust results, a limitation imposed by time.

Our analysis of country data, utilizing variables such as railroad track length, passenger and freight transportation, Internet penetration, and GDP, led to relevant conclusions regarding customer segmentation and the efficient allocation of marketing resources. The main findings are presented below.

Normalization and consistency

The normalization of sparse data proved crucial for improving the consistency of the analysis and the computational performance of the K-means algorithm. By reducing the difference in variable magnitudes, normalization facilitated the formation of more homogeneous clusters and allowed for a more accurate analysis of the relationships between countries.

Distance metrics and performance

The Chebyshov and weighted Euclidean variants performed better in correcting data errors and assessing the relevance of specific variables. The weighted Euclidean distance stood out for its ability to adjust to the PCA, allowing for a better visualization of the relationships between countries and their clusters. Our evaluation, using metrics such as the silhouette score, indicates the good performance of the algorithm. The effectiveness of the weighted Euclidean distance suggests that incorporating the importance of variables is essential. Future work could explore the full implementation of the Mahalanobis distance to assess whether its benefits in handling variable correlations outweigh its increased computational complexity in the context of B2B customer segmentation.

Quantitative results

Our experimental analysis yielded significant quantitative findings that validate the proposed methodology. The weighted Euclidean distance reported a silhouette coefficient of 0.3807 and a Davies-Bouldin index of 0.8174, indicating the most robust and well-defined clusters. These scores significantly outperform those of the traditional variants (Euclidean: 0.2796; Manhattan: 0.1983; Chebyshov: 0.1749). Furthermore, linear programming optimization provided tangible financial insights. For example, our model consistently allocated a significant portion of the marketing budget—approximately $300 000—to the railroad track construction campaign, particularly for clusters identified using the Chebyshov and weighted Euclidean variants. This contrasts with the Manhattan distance, which distributed the investment more evenly, suggesting a less focused and potentially less effective strategy. These results demonstrate the value of a data-driven approach to budg*et al*location and confirm the superiority of specific distance metrics in identifying high-potential market segments.

Operational efficiency

Beyond its technical rigor, a key contribution of our proposed methodology is its significant enhancement of operational efficiency in B2B marketing. Our integrated approach, which automates tasks by combining K-means clustering and linear programming, drastically reduces the time required for strategic planning and ensures that marketing resources are allocated for maximum utility.

Market potential

Our analysis revealed a high potential for the global railroad track construction campaign, while the freight railcar campaign showed a more limited potential. We suggest that researchers consider the specific characteristics of each market and product when designing marketing campaigns. This study also identified a high market potential for electric locomotives in Colombia and a niche railroad market in Boyacá. The railroad sector can leverage the current political-environmental situation, which favors the search for more sustainable transportation solutions, and contribute to the country's development. Our analysis, which employed a global dataset, demonstrates the model's potential for railway ABM in international markets. Researchers can replicate this model on a small scale as a pilot for a specific region (e.g., Colombia) by replacing the country's data with regional client information.

AUTHOR CONTRIBUTIONS

Carlos Eduardo Díaz Peñuela: Writing of original draft, writing-review, editing and software.
Javier Antonio Ballesteros Ricaurte: Writing-review and editing.
Gustavo Cáceres Castellanos: Writing-review and editing.

ACKNOWLEDGMENTS

The authors would like to thank Universidad Pedagógica y Tecnológica de Colombia's Information Management research group (GIMI) for supporting this work.

REFERENCES

Abdulnassar, A. A., & Nair, L. R. (2023). Performance analysis of Kmeans with modified initial centroid selection algorithms and developed Kmeans9+ model. *Measurement: Sensors*, *25*, 100666. https://doi.org/10.1016/J.MEASEN.2023.100666

Account-based marketing budgets demystified: Findings from SiriusDecisions' latest study. (2019). https://www.forrester.com/blogs/abm-budgets-study-findings/

Ayele, W. Y. (2020). *Adapting CRISP-DM for idea mining a data mining process for generating ideas using a textual dataset*. *International Journal of Advanced Computer Science and Applications*, *11*(6), 3. https://doi.org/10.14569/IJACSA.2020.0110603

Bereg, S., Haghpanah, M., Malouf, B., & Sudborough, I. H. (2024). Improved bounds for permutation arrays under Chebyshev distance. *Designs, Codes, and Cryptography*, *92*(4), 1023-1039. https://doi.org/10.1007/S10623-023-01326-1/TABLES/5

Burgess, B. (2025). *Account-based marketing: The definitive handbook for B2B marketers*. Kogan Page.

Buzzell, G. A., Niu, Y., Aviyente, S., & Bernat, E. (2022). A practical introduction to EEG time-frequency principal components analysis (TF-PCA). *Developmental Cognitive Neuroscience*, *55*, 101114. https://doi.org/10.1016/J.DCN.2022.101114

Central Intelligence Agency (CIA) (2024). *The world factbook*. https://www.cia.gov/the-world-factbook/

Chikumbo, O., & Granville, V. (2019). Optimal clustering and cluster identity in understanding high-dimensional data spaces with tightly distributed points. *Machine Learning and Knowledge Extraction, 1*(2), 715-744. https://doi.org/10.3390/MAKE1020042

Colombian National Planning Department (2020). *APP-Sector férreo (2020)*. https://colaboracion.dnp.gov.co/CDT/Participacin%20privada%20en%20proyectos%20de%20infraestructu/Gu%C3%ADa%20APP%20Sector%20F%C3%A9rreo.pdf

Cuevas-Díaz Durán, R., Wei, H., & Wu, J. (2024). Data normalization for addressing the challenges in the analysis of single-cell transcriptomic datasets. *BMC Genomics*, *25*(1), 1-18. https://doi.org/10.1186/S12864-024-10364-5

Fabri, G., Ometto, A., Li, H., & d'Ovidio, G. (2024). Redesign of a non-electrified urban railway line with hydrogen-fuelled trains. *Lecture Notes in Civil Engineering*, *526*, 640-648. https://doi.org/10.1007/978-981-97-4355-1_62

Figueroa, A. E. N., Castellanos, G. C., & Sanabria, J. S. G. (2023). Modelo de *machine learning* para la clasificación de municipios por cultivos ilícitos en Colombia de 2010 a 2020. *Inge CuC*, *19*(1), 47-60. https://doi.org/10.17981/INGECUC.19.1.2023.05

Findeter (2020). *Metro de Medellín tendrá dos nuevos trenes y llega el primero de los 20 que habían sido comprados | Findeter*. https://www.findeter.gov.co/noticias/infraestructura/metro-de-medell%C3%ADn-tendr%C3%A1-dos-nuevos-trenes-y-llega-el-primero-de-los-20-que-hab%C3%ADan-sido-comprados#

Forrester Research (2019). *Account-based marketing budgets demystified: Findings from SiriusDecisions' latest study*. https://www.forrester.com/blogs/abm-budgets-study-findings/

Gene Day, D., & Wei Shi, S. (2020). Automated and scalable: Account-based B2B marketing for startup companies. *Journal of Business Theory and Practice*, *8*(2), 16-23. https://doi.org/10.22158/jbtp.v8n2p16

Gobierno de Colombia. (2020, November). *Plan maestro ferroviario*. https://colaboracion.dnp.gov.co/CDT/Prensa/Plan-Maestro-Ferroviario.pdf

Hernández Sampieri, R., Feránadez Collado, C., & Baptista Lucio, M. D. P. (2014). *Metodología de la investigación*. McGraw Hill España. https://dialnet.unirioja.es/servlet/libro?codigo=775008&info=resumen&idioma=SPA

Karam, A., Jensen, A. J. K., & Hussein, M. (2023). Analysis of the barriers to multimodal freight transport and their mitigation strategies. *European Transport Research Review*, *15*(1), 1-16. https://doi.org/10.1186/S12544-023-00614-0/METRICS

Lenssen, L., & Schubert, E. (2023). Medoid silhouette clustering with automatic cluster number selection. *Information Systems*, *120*, 102290. https://doi.org/10.1016/j.is.2023.102290

Li, X., Shi, H., Yang, K., Dou, Q., & Jia, N. (2025). Ground fault insulation monitoring method for smart substation based on Mahalanobis distance and automatic code generation. *Energy Informatics*, *8*(1), 22. https://doi.org/10.1186/S42162-025-00470-3

Ling, L. S., & Weiling, C. T. (2025). Enhancing segmentation: A comparative study of clustering methods. *IEEE Access*, *13*, 47418-47439. https://doi.org/10.1109/ACCESS.2025.3550339

Makarychev, K., Reddy, A., & Shan, L. (2020). Improved Guarantees for k-means++ and k-means++ Parallel. In H. Larochelle, M. Ranzato, R. Hadzel, & H. Lin (Eds.), *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)* (pp. 1-11). NEurIPS. https://proceedings.neurips.cc/paper_files/paper/2020/hash/ba304f3809ed31d0ad97b5a2b5df2a39-Abstract.html

Ministry of Transportation (2019). *En Colombia, cada año cerca de 670 mil personas usan el tren como medio de transporte*. https://mintransporte.gov.co/publicaciones/7286/en-colombia-cada-ano-cerca-de-670-mil-personas-usan-el-tren-como-medio-de-transporte/

Ministry of the Environment (2024). *MinAmbiente presenta borrador de decreto para impulsar transporte ferroviario en ciudades del país*. https://www.minambiente.gov.co/minambiente-presenta-borrador-de-decreto-para-impulsar-transporte-ferroviario-en-ciudades-del-pais/

Ren, Y., Yang, M., Chen, E., Cheng, L., & Yuan, Y. (2024). Exploring passengers' choice of transfer city in air-to-rail intermodal travel using an interpretable ensemble machine learning approach. *Transportation*, *51*(4), 1493-1523. https://doi.org/10.1007/S11116-023-10375-3/FIGURES/7

RENFE (2023). *Renfe licita la compra de 149 vagones para mercancías*. https://www.renfe.com/es/es/grupo-renfe/comunicacion/renfe-al-dia/sala-de-prensa/renfe-licita-la-compra-de-vagones-para-mercancias

Rubiños, M., Díaz-Longueira, A., Timiraos, M., Michelena, Á., García-Ordás, M. T., & Alaiz-Moretón, H. (2024). A Comparative Analysis of Algorithms and Metrics to Perform Clustering. *Lecture Notes in Networks and Systems*, *1173*, 63-72. https://doi.org/10.1007/978-3-031-73910-1_7

Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). Setting B2B digital marketing in artificial intelligence-based CRMs: A review and directions for future research. *Industrial Marketing Management*, *98*, 161-178. https://doi.org/10.1016/J.INDMARMAN.2021.08.006

United Nations (UN) (2024). *Member States | United Nations*. https://www.un.org/en/about-us/member-states

United Nations Economic Commission for Europe (UNECE) (2024). *Data portal*. https://w3.unece.org/PXWeb/en

Universidad Oberta de Catalunya (2024). *Espacio de recursos de ciencia de datos*. https://datascience.recursos.uoc.edu/es/optimizacion-con-pulp/

Vaquer, A. J. (2023). Aplicación de la programación lineal en el planeamiento de la producción de una acería. *South Florida Journal of Development*, *4*(4), 1623-1638. https://doi.org/10.46932/sfjdv4n4-015

Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021). Artificial intelligence in marketing: Systematic review and future research direction. *International Journal of Information Management Data Insights*, *1*(1), 100002. https://doi.org/10.1016/J.JJIMEI.2020.100002

World Bank (2024a). *Superficie (kilómetros cuadrados) | Data*. https://datos.bancomundial.org/indicador/AG.SRF.TOTL.K2

World Bank (2024b). *PIB per cápita (US$ a precios actuales) | Data*. https://datos.bancomundial.org/indicador/NY.GDP.PCAP.CD

World Bank (2024c). *Individuos que utilizan Internet (% de la población) | Data*. https://datos.bancomundial.org/indicador/IT.NET.USER.ZS

World Population Review (2024). *Rail usage by country 2024*. https://worldpopulationreview.com/country-rankings/rail-usage-by-country