

AÑO 30 No. 110, 2025  
ABRIL-JUNIO



No. 110, 2025

ABRIL-JUNIO



# Revista Venezolana de Gerencia



UNIVERSIDAD DEL ZULIA (LUZ)  
Facultad de Ciencias Económicas y Sociales  
Centro de Estudios de la Empresa

ISSN 1315-9984

Esta obra está bajo una licencia de Creative Commons  
Reconocimiento-NoComercial-CompartirIgual 3.0 Unported.  
[http://creativecommons.org/licenses/by-nc-sa/3.0/deed.es\\_ES](http://creativecommons.org/licenses/by-nc-sa/3.0/deed.es_ES)

Como citar: Solano, G. D. C., López, J. F., Pomaquero, J. C., y Tobar, M. G. (2025). Patrones de Comportamiento en usuarios de transporte interprovincial en Ecuador mediante Técnicas de Machine Learning. *Revista Venezolana De Gerencia*, 30(110), 1047-1061. <https://doi.org/10.52080/rvgluz.30.110.17>

Universidad del Zulia (LUZ)  
Revista Venezolana de Gerencia (RVG)  
Año 30 No. 110, 2025, 1047-1061  
Abril-Junio  
ISSN 1315-9984 / e-ISSN 2477-9423



# Patrones de Comportamiento en usuarios de transporte interprovincial en Ecuador mediante Técnicas de Machine Learning

Solano Aguilar, Gabriela del Cisne \*  
López Aguirre, José Fernando \*\*  
Pomaquero Yuquilema, Juan Carlos \*\*\*  
Tobar Ruiz, María Gabriela \*\*\*\*

## Resumen

Este estudio tiene como objetivo analizar y predecir patrones de comportamiento de los usuarios de transporte interprovincial en Ecuador mediante técnicas de aprendizaje automático. Se utilizó un conjunto de datos proporcionado por la Unión de Cooperativas de Transporte Interprovincial de Ecuador que abarca viajes realizados entre 2022 y 2024. La metodología incluyó la implementación de K-means para la segmentación de usuarios y PCA para la reducción dimensional. Inicialmente, K-means identificó cuatro clústeres, pero el solapamiento entre grupos motivó la aplicación de PCA, mejorando la separación. Los resultados revelaron cuatro grupos: Ritmo Diario, Exploradores de Fin de Semana, Nómadas de Eventos y Viajeros Flexibles. Esta segmentación ofrece información clave para optimizar los servicios de transporte y mejorar la experiencia del usuario al ajustar recursos a las necesidades de cada grupo.

**Palabras clave:** Transporte interprovincial; machine learning; análisis de patrones; clustering; reducción dimensional.

---

Recibido: 02.12.24

Aceptado: 10.02.25

\* Ingeniera en electrónica, control y redes industriales. Investigador independiente. Email: [gabys\\_9308@hotmail.com](mailto:gabys_9308@hotmail.com). ORCID: <https://orcid.org/0009-0007-7565-4702>

\*\* Ingeniero en industrias pecuarias – Máster en administración de empresas – Máster en big data. Escuela Superior Politécnica de Chimborazo. Email: [josef.lopez@esepoch.edu.ec](mailto:josef.lopez@esepoch.edu.ec). ORCID: <https://orcid.org/0000-0001-9706-5115>

\*\*\* Máster en Políticas Públicas y Sociales. Ingeniero en Administración de Empresas. Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador. Email: [carlos.pomaquero@esepoch.edu.ec](mailto:carlos.pomaquero@esepoch.edu.ec). ORCID: <https://orcid.org/0000-0003-0952-943X>

\*\*\*\* Ingeniera en administración de empresas y negocios - Magíster en gestión de marketing y servicio al cliente. Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador. Email: [maria.tobar@esepoch.edu.ec](mailto:maria.tobar@esepoch.edu.ec). ORCID: <https://orcid.org/0000-0002-3796-0545>

# Behavioral Patterns in Interprovincial Transport Users in Ecuador Using Machine Learning Techniques

## Abstract

This study aims to analyze and predict behavior patterns of interprovincial transport users in Ecuador using machine learning techniques. A dataset provided by the Union of Interprovincial Transport Cooperatives of Chimborazo, covering trips from 2022 to 2024, was used. The methodology involved K-means for user segmentation and PCA for dimensionality reduction. Initially, K-means identified four clusters, but group overlap led to the application of PCA, enhancing cluster separation. Results revealed four distinct groups: Daily Rhythm, Weekend Explorers, Event Nomads, and Flexible Travelers. This segmentation provides valuable insights to optimize transport services and enhance user experience by tailoring resources to the needs of each group.

**Keywords:** Interprovincial transport; machine learning; pattern analysis; clustering; dimensionality reduction.

## 1. Introducción

El transporte interprovincial es una pieza clave para la movilidad de las personas y el desarrollo de actividades socioeconómicas en Ecuador (Argüello Erazo et al., 2020), pero enfrenta desafíos en la optimización del servicio y la satisfacción de los usuarios. Aunque las organizaciones generan grandes volúmenes de datos, estos no se aprovechan completamente para tomar decisiones estratégicas. Según investigaciones recientes, la integración del transporte fomenta el desarrollo económico de las regiones y actúa como un motor de cohesión territorial (Jalolova et al., 2022; Taghvaei et al., 2022; Yarushkina et al., 2022), lo que resalta la necesidad de aplicar tecnologías avanzadas para mejorar la gestión del transporte.

La inteligencia artificial (IA), definida como la capacidad de los sistemas para interpretar datos, aprender de ellos y adaptarse (Kaplan & Haenlein, 2019), se ha consolidado como una herramienta poderosa para analizar grandes volúmenes de datos, identificar patrones y hacer predicciones. Por ejemplo, se han empleado algoritmos avanzados para predecir tiempos de viaje, capturando dinámicas complejas del tráfico (Abdollahi et al., 2020), así como para optimizar el uso de infraestructuras conectadas, como sistemas de bicicletas compartidas y estaciones de metro (Lv et al., 2021). Además, el éxito de estas técnicas en sectores como el comercio electrónico destaca su potencial en el ámbito del transporte (Bandyopadhyay et al., 2020).

En el sector ferroviario, la IA ha optimizado áreas como mantenimiento,

planificación y control autónomo (Tang et al., 2022). De manera similar, el transporte interprovincial podría beneficiarse de estas herramientas, aunque persiste una brecha en el conocimiento sobre su implementación en regiones con infraestructura limitada. Factores como la falta de alfabetización en IA, la desconfianza tecnológica y las limitaciones organizacionales dificultan su adopción (Cubric, 2020; Herath & Mittal, 2022), lo que resalta la necesidad de un enfoque estructurado para integrar la IA en este sector.

Además, gran parte de las investigaciones actuales se centran en la movilidad urbana (Brůhová Foltýnová et al., 2020; Miskolczi et al., 2021) o en vehículos autónomos (Golbabaie et al., 2020; Narayanan et al., 2020; Rahman & Thill, 2023), dejando de lado la aplicación de IA en el transporte interprovincial. Este estudio aborda esta brecha utilizando técnicas de segmentación de mercado para comprender el comportamiento de los usuarios.

El enfoque metodológico se basa en el Análisis de Componentes Principales (PCA) y el algoritmo K-means. PCA reduce la dimensionalidad de los datos, facilitando el análisis de patrones complejos (Anowar et al., 2021; Lee & Jemain, 2021). Luego, K-means agrupa a los usuarios según sus hábitos de viaje, proporcionando una base sólida para decisiones estratégicas. Este algoritmo, eficaz y de baja complejidad computacional, ha demostrado ser útil para gestionar grandes volúmenes de datos y segmentar de manera efectiva (Ezugwu et al., 2022; Ikotun et al., 2023), optimizando decisiones estratégicas en contextos complejos (de Oliveira et al., 2023; Zhang et al., 2021).

## 2. Aplicación de aprendizaje automático para el análisis de patrones de comportamiento: Revisión de literatura

El análisis de patrones de comportamiento de usuarios mediante aprendizaje automático ha ganado relevancia en sectores como el transporte, comercio electrónico y gestión de relaciones con clientes. En el transporte, entender los patrones de movilidad es esencial para mejorar la planificación y eficiencia de los servicios.

El algoritmo K-means es ampliamente reconocido por su simplicidad y eficiencia computacional, lo que lo convierte en una herramienta esencial para la segmentación de grandes volúmenes de datos (Gbadoubissa et al., 2020; Ikotun et al., 2023; Javidan et al., 2023; F. Khan et al., 2024; Mussabayev et al., 2023; Ragunthar et al., 2021; Ran et al., 2024). Sin embargo, presenta limitaciones como la necesidad de especificar a priori el número de clústeres y su sensibilidad a la inicialización de centroides, lo que puede afectar su rendimiento en conjuntos de datos complejos (Alguliyev et al., 2021; Bai et al., 2020). Para superar estas limitaciones, se han desarrollado métricas de validación como el Silhouette Score, el índice de Calinski-Harabasz y el índice de Davies-Bouldin. Estas métricas permiten evaluar la calidad del agrupamiento al medir la cohesión interna y la separación entre clústeres (Gagolewski et al., 2021; Halim et al., 2021; Hassan et al., 2024; I. K. Khan et al., 2024).

El Silhouette Score evalúa la similitud de una observación con su

propio clúster en comparación con otros, ofreciendo valores entre -1 y 1, donde valores más altos indican una mejor agrupación (Bagirov et al., 2023; Halim et al., 2021; Kim et al., 2020; Mehedi Hassan et al., 2022). Por otro lado, el índice de Calinski-Harabasz mide la relación entre la varianza entre clústeres y dentro de ellos, con valores más altos representando clústeres más compactos y bien separados (Hajihosseini et al., 2024; Junjie et al., 2024; Ning et al., 2022; Wei, 2024; Zhao et al., 2023). Finalmente, el índice de Davies-Bouldin calcula la media de las similitudes máximas entre cada clúster y el más cercano, donde valores más bajos reflejan una mejor separación (I. K. Khan et al., 2024; Naghizadeh & Metaxas, 2020; Nowak-Brzezinska & Horyn, 2020; Ros et al., 2023).

El Análisis de Componentes Principales es una técnica de reducción de dimensionalidad que permite eliminar redundancias y resaltar las características más relevantes de los datos (Anowar et al., 2021; El-Rawy et al., 2024; Lee & Jemain, 2021; Schreiber, 2021). Su integración con K-means ha demostrado mejorar significativamente la segmentación al facilitar la identificación de patrones ocultos y reducir el solapamiento entre clústeres (Bandyopadhyay et al., 2020; Jafarzadegan et al., 2019; Jansson et al., 2022; Pellegrino et al., 2023; Troccoli et al., 2022; Zhu et al., 2022). Por ejemplo, (Jafarzadegan et al., 2019) propusieron un método que combina PCA y clustering jerárquico, mejorando la precisión del agrupamiento en datos complejos. Asimismo, esta combinación ha sido aplicada en sectores como la interpretación sísmica (Troccoli et al., 2022), la predicción de congestión del tráfico (Chiabaut & Faitout, 2021),

la gestión de relaciones con clientes (Sun et al., 2023) y en la detección no supervisada de patrones en series temporales InSAR, identificando deformaciones del terreno (Festa et al., 2023).

En el transporte, estas técnicas han permitido optimizar servicios y comprender patrones de movilidad. (Fabre et al., 2024) utilizaron K-means y datos de Wi-Fi para estimar la movilidad de pasajeros de autobuses, mientras que (Yarlagadda et al., 2021) aplicaron PCA y K-means para identificar patrones de conducción agresiva en conductores de vehículos pesados. Chun et al., 2023 clasificaron estaciones de metro en Seúl utilizando PCA y K-means, revelando patrones de uso similares entre estaciones. (Ma et al., 2021) desarrollaron un marco para clasificar estilos de conducción en aplicaciones móviles de transporte, aplicando PCA y K-means para analizar datos de comportamiento de conductores y destacando cómo las tareas de conducción generan variaciones en el estilo. (Güller & Varol, 2024) analizaron cómo el entorno construido afecta los patrones de movilidad, destacando su impacto en la planificación urbana.

En Ecuador, se ha explorado el uso de K-means en el transporte interprovincial, segmentando clientes y analizando la demanda para mejorar decisiones estratégicas. Este enfoque es particularmente relevante en contextos con recursos tecnológicos limitados, donde la optimización de servicios puede contribuir al desarrollo socioeconómico (López Aguirre et al., 2024).

A pesar de estos avances, persisten desafíos en sectores específicos como el transporte interprovincial, donde las limitaciones de infraestructura tecnológica dificultan la

implementación de estas técnicas. Este estudio busca llenar este vacío mediante la aplicación de PCA y K-means para analizar patrones de comportamiento en usuarios de transporte interprovincial en Ecuador, con el objetivo de optimizar los servicios y mejorar la toma de decisiones operativas.

### 3. Consideraciones metodológicas para el análisis de patrones de comportamiento

Este estudio se centró en la implementación de técnicas de aprendizaje automático para analizar los patrones de comportamiento de los usuarios de transporte interprovincial en Chimborazo, Ecuador. El proceso metodológico incluyó la recolección, procesamiento y análisis de datos mediante técnicas de segmentación y reducción dimensional.

El conjunto de datos, proporcionado por la Unión de Cooperativas de Transporte Interprovincial de Chimborazo (UCTCH), contiene registros de ventas de boletos de diferentes cooperativas entre 2022 y 2024. Debido a la inconsistencia de los archivos, se realizó un proceso de estandarización, unificando la información y seleccionando los campos comunes:

- Fecha: Indica el día en que se realizó el viaje.
- Frecuencia: Tiempo específico de salida del bus de la terminal.
- Ruta: Descripción del trayecto cubierto por el viaje.
- Valor: Ingresos generados por cada viaje.

El conjunto de datos unificado contiene 139,150 filas, correspondientes a viajes individuales. Posteriormente,

se enriqueció con datos temporales adicionales, como los días de la semana y del mes, para facilitar un análisis detallado de la demanda de transporte.

El proceso de unificación redujo el conjunto de datos a 10,423 registros, al consolidar entradas duplicadas o fragmentadas, sumando los ingresos por cada viaje para mejorar la precisión y claridad del análisis.

Para la segmentación de usuarios, se utilizó una combinación de PCA y K-means para reducir la dimensionalidad y agrupar los datos. Aunque se probaron otros algoritmos de clustering como el jerárquico, GMM y DBSCAN, estos no ofrecieron resultados satisfactorios. Las métricas de evaluación, como el Silhouette Score y los índices de Calinski-Harabasz y Davies-Bouldin, indicaron una mayor calidad en los resultados obtenidos con PCA y K-means, que finalmente se eligieron para la segmentación.

El estudio se fundamenta en datos de una región y período específico (2022-2024), lo que puede limitar su aplicabilidad en otros contextos. Futuras investigaciones podrían extender el análisis a distintas regiones y considerar otras variables. Además, la integración de técnicas avanzadas de clustering o algoritmos híbridos permitiría mejorar la precisión de los resultados. A pesar de estas limitaciones, la metodología propuesta es replicable y adaptable a diferentes entornos.

### 4. Análisis de Resultados: Comparación entre la Aplicación Inicial de K-means y PCA + K-means

En el análisis de los datos de transporte interprovincial, se aplicó

inicialmente el algoritmo K-means, utilizando el método del codo para determinar el número óptimo de clústeres. Este método indicó que el valor óptimo de  $k=4$ , lo que sugiere una partición adecuada de los datos. Para evaluar la calidad del agrupamiento, se utilizaron métricas de validación: el Silhouette Score para  $k=4$  fue 0.1881, indicando un solapamiento moderado entre los clústeres; el índice de Calinski-Harabasz fue de 2072.33, sugiriendo buena separación; y el índice de Davies-Bouldin, con un valor de 1.5344, indicó cierta superposición entre los grupos. A pesar de que se observaron diferencias entre los clústeres en variables como el día del mes, la ruta y el valor del boleto, los resultados mostraron que la segmentación no fue completamente precisa, lo que subraya la necesidad de aplicar técnicas adicionales para mejorar la calidad del agrupamiento.

A continuación, se realizó el análisis de PCA sobre los datos preprocesados para identificar las principales fuentes de variabilidad relacionadas con la venta de

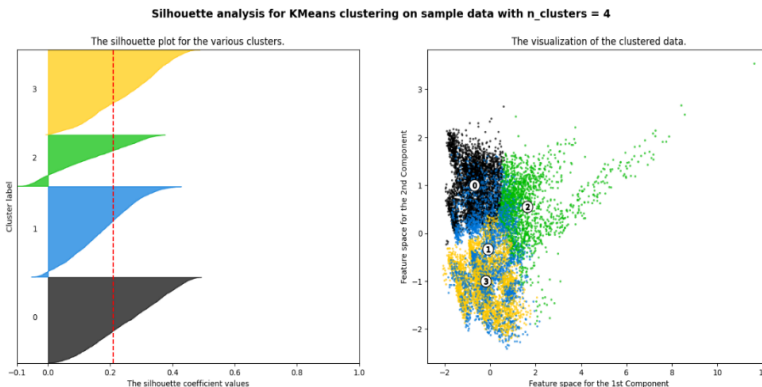
boletos, rutas y horarios de transporte. Los primeros cuatro componentes principales capturaron el 86.46% de la variabilidad acumulada, lo que es suficiente para representar la mayor parte de la información de los datos originales.

Tras la reducción de dimensionalidad con PCA, se aplicó K-means para el agrupamiento. La combinación de PCA y K-means mejora la interpretación y separabilidad de los clústeres al reducir la correlación entre las variables. La reducción de correlación entre las variables mediante PCA permite que los clústeres formados sean generalmente más compactos y separables (Bandyopadhyay et al., 2020; Jafarzadegan et al., 2019).

La efectividad de la combinación de PCA y K-means se evaluó mediante varias métricas de validación interna. El Silhouette Score para  $k = 4$  aumentó a 0.2097, lo que indicó una mejora en la cohesión interna y en la separación entre clústeres respecto al análisis previo (ilustración 1).

### Ilustración 1

## Análisis de silueta para K-means con 4 clústeres, luego del PCA. El promedio del silhouette\_score es 0.2097

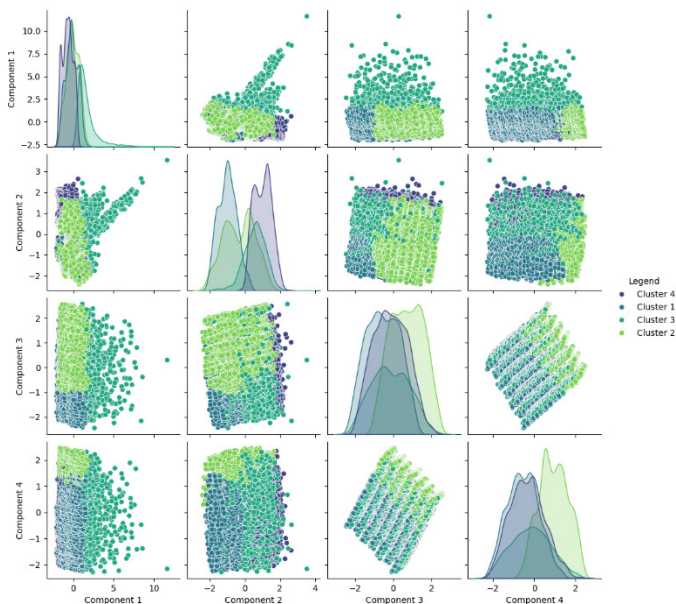


El Índice de Calinski-Harabasz incrementó de 2072.33 a 2526.07, sugiriendo que los clústeres están más dispersos y bien definidos en el espacio reducido, mientras que el Índice de Davies-Bouldin disminuyó de 1.5344 a 1.4049, reflejando una menor superposición entre clústeres y una mejora en la calidad del agrupamiento (Anowar et al., 2021; Ikotun et al., 2023).

Una vez realizado el procedimiento,

se observa la distribución de los clústeres luego de la aplicación conjunta de PCA y K-Means (ilustración 2). Esta figura muestra la matriz de dispersión de los primeros cuatro componentes principales con los clústeres generados por K-means, representados en diferentes colores. En la diagonal, se aprecian las distribuciones univariadas de cada componente.

### Ilustración 2 Matriz de dispersión de los cuatro primeros componentes principales con clústeres K-means



Se destaca que el Componente 1 muestra una mayor separación entre los clústeres, lo que sugiere que captura una variabilidad significativa para distinguir los grupos. En cambio, los Componentes 3 y 4 presentan una mayor superposición, indicando

que las variables asociadas a estos componentes no segmentan claramente a los usuarios en esos ejes. Aunque los clústeres no están perfectamente separados, la combinación de PCA y K-means mejora la segmentación en comparación con el análisis inicial,



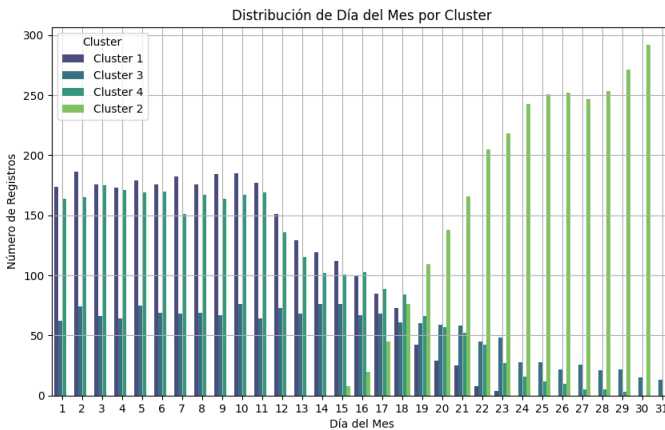
facilitando una interpretación más precisa de los patrones de comportamiento de los usuarios.

La combinación de PCA y K-means permitió segmentar a los usuarios del transporte interprovincial en cuatro clústeres, cada uno con patrones de comportamiento distintivos. Para evaluar la calidad de la segmentación, se emplearon tres métricas de validación: Silhouette Score, índice Calinski-Harabasz e índice Davies-Bouldin. Los resultados de estas métricas mostraron una mejora en la segmentación tras

la aplicación combinada de PCA y K-means, en comparación con el uso exclusivo de K-means sin reducción de dimensionalidad.

Una vez validada la segmentación, los clústeres fueron analizados a través de diversas gráficas que permitieron estudiar su distribución según diferentes características, lo que facilitó la identificación de patrones de viaje específicos. Por ejemplo, una de las gráficas analizadas (Gráfico 1) muestra cómo los clústeres se distribuyen a lo largo del mes.

**Gráfico 1**  
**Distribución de Día del Mes por Clúster**



A partir de este análisis, y profundizando en características como el día de la semana, las rutas y los horarios, se pudo definir con mayor precisión los grupos y sus respectivos patrones de viaje, identificándose los siguientes clústeres: Clúster 1 (Ritmo Diario), que agrupa a usuarios con viajes regulares durante la semana, especialmente al inicio del mes, como

trabajadores y estudiantes; clúster 2 (Exploradores de Fin de Semana), que incluye a quienes prefieren viajar durante los fines de semana, especialmente los domingos, para actividades recreativas o visitas familiares; clúster 3 (Nómadas de Eventos), formado por usuarios que viajan de forma esporádica, principalmente los viernes y domingos, debido a eventos o actividades de ocio; y

clúster 4 (Viajeros Flexibles), que agrupa a personas con viajes dispersos a lo largo del mes, sin patrones fijos.

Este estudio demuestra la eficacia de combinar PCA con K-means para segmentar a los usuarios del transporte interprovincial en Ecuador. La identificación de cuatro clústeres—Ritmo Diario, Exploradores de Fin de Semana, Nómadas de Eventos y Viajeros Flexibles—ofrece una comprensión detallada de los patrones de comportamiento de los usuarios, lo que facilita la optimización de los servicios de transporte.

Inicialmente, el uso de K-means sobre los datos originales mostró solapamiento entre los clústeres, lo cual es común debido a la sensibilidad del algoritmo a la inicialización de centroides, como indican Ikotun et al. (2023). La incorporación de PCA para reducir la dimensionalidad mejoró la separación entre clústeres, como se ha demostrado en estudios previos (Chun et al., 2023).

Los clústeres identificados coinciden con patrones de movilidad observados en estudios previos. Por ejemplo, el clúster Ritmo Diario refleja desplazamientos regulares por motivos laborales o educativos, similar a lo descrito por Chun et al. (2023). Los Exploradores de Fin de Semana y Nómadas de Eventos coinciden con usuarios cuya demanda fluctúa según eventos o fines de semana, un patrón encontrado por Fabre et al. (2024).

Los Viajeros Flexibles, con patrones impredecibles, representan un desafío para la planificación del transporte, destacando la importancia de usar técnicas de aprendizaje automático para anticipar demandas y ajustar los servicios, como sugieren (Güller & Varol, 2024).

Las métricas de validación, como el índice de Calinski-Harabasz y Davies-Bouldin, indican una segmentación más precisa tras la aplicación de PCA. El Silhouette Score de 0.2097 muestra una mejora en la separación de los clústeres respecto al análisis inicial, en línea con estudios similares (Anowar et al., 2021; Hassan et al., 2024).

Comprender estos clústeres permite a las cooperativas de transporte adaptar horarios, frecuencias y rutas para satisfacer las necesidades de cada grupo. Por ejemplo, se podrían aumentar las frecuencias durante las horas pico para el clúster Ritmo Diario y planificar servicios especiales para los otros segmentos. Esto coincide con lo propuesto por López et al. (2024), que abogan por el análisis de datos en la toma de decisiones de transporte, con un enfoque de economía social y solidaria.

Estos resultados también tienen implicaciones para el desarrollo económico y social, mejorando la asignación de recursos y la eficiencia operativa, lo que favorece el desarrollo territorial, como señalan (Yarushkina et al., 2022).

## 5. Conclusiones

La investigación demuestra la efectividad de combinar el análisis de componentes principales (PCA) con el algoritmo de agrupamiento K-means para segmentar a los usuarios del transporte interprovincial en Ecuador. La aplicación de PCA permitió reducir la dimensionalidad de los datos, optimizando la separación entre los clústeres y mejorando la precisión de la segmentación. Como resultado, se identifican cuatro segmentos diferenciados de usuarios: Ritmo Diario, compuesto por trabajadores

y estudiantes con viajes frecuentes y regulares; Exploradores de Fin de Semana, caracterizados por desplazamientos recreativos y visitas familiares; Nómadas de Eventos, que viajan en función de eventos específicos, principalmente los fines de semana; y Viajeros Flexibles, sin un patrón fijo de desplazamiento.

Estos hallazgos coinciden con estudios previos en movilidad urbana y demuestran el potencial de la analítica de datos para optimizar la planificación del transporte. Asimismo, la metodología utilizada es replicable y adaptable a distintos contextos, lo que puede contribuir significativamente a la mejora de la eficiencia operativa y a la toma de decisiones estratégicas en el sector del transporte interprovincial.

Las métricas de validación, como el Silhouette Score, el índice de Calinski-Harabasz y el índice de Davies-Bouldin, confirman que la combinación de PCA y K-means mejora significativamente la segmentación en comparación con el uso exclusivo de K-means. La reducción de dimensionalidad mediante PCA optimiza la separación entre los clústeres, minimizando el solapamiento y permitiendo una clasificación más precisa de los usuarios del transporte interprovincial.

Esta metodología no solo proporciona una base sólida para el análisis de patrones de movilidad, sino que también representa una herramienta estratégica para la planificación del transporte. Su implementación permite a las cooperativas y operadores ajustar rutas, optimizar frecuencias y asignar recursos de manera eficiente, garantizando un servicio más alineado con las necesidades reales de los pasajeros. Además, la capacidad de replicación de este enfoque lo convierte

en un modelo.

El estudio evidencia cómo la aplicación de PCA y K-means mejora la segmentación de usuarios del transporte interprovincial, proporcionando información clave para la optimización del servicio. La identificación de patrones de movilidad permite a los operadores ajustar rutas, horarios y recursos de manera eficiente. Igualmente, esta metodología es replicable en distintos contextos, facilitando la planificación estratégica y contribuyendo al desarrollo de un sistema de transporte más sostenible, accesible y basado en datos.

## Referencias

- Abdollahi, M., Khaleghi, T., & Yang, K. (2020). An integrated feature learning approach using deep learning for travel time prediction. *Expert Systems with Applications*, 139, 112864. <https://doi.org/10.1016/J.ESWA.2019.112864>
- Alguliyev, R. M., Aliguliyev, R. M., & Sukhostat, L. V. (2021). Parallel batch k-means for Big data clustering. *Computers & Industrial Engineering*, 152, 107023. <https://doi.org/10.1016/J.CIE.2020.107023>
- Anowar, F., Sadaoui, S., & Selim, B. (2021). Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Computer Science Review*, 40, 100378. <https://doi.org/10.1016/J.COSREV.2021.100378>
- Argüello Erazo, S. E., Villa Uvidia, R. N., & Palahuachi Sumba, J. P. (2020). *Historia y evolución de la gestión del transporte público urbano en la provincia de Chimborazo*.
- Bagirov, A. M., Aliguliyev, R. M., &

- Sultanova, N. (2023). Finding compact and well-separated clusters: Clustering using silhouette coefficients. *Pattern Recognition*, 135, 109144. <https://doi.org/10.1016/J.PATCOG.2022.109144>
- Bai, L., Liang, J., & Cao, F. (2020). A multiple k-means clustering ensemble algorithm to find nonlinearly separable clusters. *Information Fusion*, 61, 36–47. <https://doi.org/10.1016/J.INFFUS.2020.03.009>
- Bandyopadhyay, S., Thakur, S. S., & Mandal, J. K. (2020). Product recommendation for e-commerce business by applying principal component analysis (PCA) and K-means clustering: benefit for the society. *Innovations in Systems and Software Engineering*, 17(1), 45–52. <https://doi.org/10.1007/S11334-020-00372-5>
- Brůhová Foltýnová, H., Vejchodská, E., Rybová, K., & Květoň, V. (2020). Sustainable urban mobility: One definition, different stakeholders' opinions. *Transportation Research Part D: Transport and Environment*, 87, 102465. <https://doi.org/10.1016/J.TRD.2020.102465>
- Chiabaut, N., & Faitout, R. (2021). Traffic congestion and travel time prediction based on historical congestion maps and identification of consensual days. *Transportation Research Part C: Emerging Technologies*, 124, 102920. <https://doi.org/10.1016/J.TRC.2020.102920>
- Chun, K. C., Bahk, J., Kim, H., Jeong, H. C., & Kim, G. (2023). Classification of the metropolitan subway stations and spheres of influence of main commercial areas in Seoul. *Physica A: Statistical Mechanics and Its Applications*, 609, 128387. <https://doi.org/10.1016/J.PHYSA.2022.128387>
- Cubic, M. (2020). Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study. *Technology in Society*, 62, 101257. <https://doi.org/10.1016/J.TECHSOC.2020.101257>
- de Oliveira, M. S., Steffen, V., de Francisco, A. C., & Trojan, F. (2023). Integrated data envelopment analysis, multi-criteria decision making, and cluster analysis methods: Trends and perspectives. *Decision Analytics Journal*, 8, 100271. <https://doi.org/10.1016/J.DAJOUR.2023.100271>
- El-Rawy, M., Wahba, M., Fathi, H., Alshehri, F., Abdalla, F., & ElAttar, R. M. (2024). Assessment of groundwater quality in arid regions utilizing principal component analysis, GIS, and machine learning techniques. *Marine Pollution Bulletin*, 205, 116645. <https://doi.org/10.1016/J.MARPOLBUL.2024.116645>
- Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence*, 110, 104743. <https://doi.org/10.1016/J.ENGAPPAI.2022.104743>
- Fabre, L., Bayart, C., Bonnel, P., & Mony, N. (2024). Estimating Bus Passenger Mobility with Wi-Fi Data and Clustering. *Transportation Research Procedia*, 76, 445–457. <https://doi.org/10.1016/J.TRPRO.2023.12.067>
- Festa, D., Novellino, A., Hussain, E., Bateson, L., Casagli, N., Confuorto, P., Del Soldato, M., & Raspini, F. (2023). Unsupervised detection of InSAR time series patterns based on PCA and K-means clustering. *International Journal*

of Applied Earth Observation and Geoinformation, 118, 103276. <https://doi.org/10.1016/J.JAG.2023.103276>

Gagolewski, M., Bartoszek, M., & Cena, A. (2021). Are cluster validity measures (in) valid? *Information Sciences*, 581, 620–636. <https://doi.org/10.1016/J.INS.2021.10.004>

Gbadoubissa, J. E. Z., Ari, A. A. A., & Gueroui, A. M. (2020). Efficient k-means based clustering scheme for mobile networks cell sites management. *Journal of King Saud University - Computer and Information Sciences*, 32(9), 1063–1070. <https://doi.org/10.1016/J.JKSUCI.2018.10.015>

Golbabaie, F., Yigitcanlar, T., Paz, A., & Bunker, J. (2020). Individual Predictors of Autonomous Vehicle Public Acceptance and Intention to Use: A Systematic Review of the Literature. *Journal of Open Innovation: Technology, Market, and Complexity*, 6(4), 106. <https://doi.org/10.3390/JOITMC6040106>

Güller, C., & Varol, C. (2024). Unveiling the daily rhythm of urban space: Exploring the influence of built environment on spatiotemporal mobility patterns. *Applied Geography*, 170, 103366. <https://doi.org/10.1016/J.APGEOG.2024.103366>

Hajihosseini, M., Maghsoudi, A., & Ghezlbash, R. (2024). A comprehensive evaluation of OPTICS, GMM and K-means clustering methodologies for geochemical anomaly detection connected with sample catchment basins. *Geochemistry*, 84(2), 126094. <https://doi.org/10.1016/J.CHEMER.2024.126094>

Halim, Z., Sargana, H. M., Adam, Uzma, & Waqas, M. (2021). Clustering of graphs using pseudo-guided random walk. *Journal of Computational*

*Science*, 51, 101281. <https://doi.org/10.1016/J.JOCS.2020.101281>

Hassan, B. A., Tayfor, N. B., Hassan, A. A., Ahmed, A. M., Rashid, T. A., & Abdalla, N. N. (2024). From A-to-Z review of clustering validation indices. *Neurocomputing*, 601, 128198. <https://doi.org/10.1016/J.NEUCOM.2024.128198>

Herath, H. M. K. K. M. B., & Mittal, M. (2022). Adoption of artificial intelligence in smart cities: A comprehensive review. *International Journal of Information Management Data Insights*, 2(1), 100076. <https://doi.org/10.1016/J.JJIMEI.2022.100076>

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhajja, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210. <https://doi.org/10.1016/J.INS.2022.11.139>

Jafarzadegan, M., Safi-Esfahani, F., & Beheshti, Z. (2019). Combining hierarchical clustering approaches using the PCA method. *Expert Systems with Applications*, 137, 1–10. <https://doi.org/10.1016/J.ESWA.2019.06.064>

Jalolova, M., Amirov, L., Askarova, M., & Zakhidov, G. (2022). Territorial features of railway transport control mechanisms. *Transportation Research Procedia*, 63, 2645–2652. <https://doi.org/10.1016/J.TRPRO.2022.06.305>

Jansson, N. F., Allen, R. L., Skogsmo, G., & Tavakoli, S. (2022). Principal component analysis and K-means clustering as tools during exploration for Zn skarn deposits and industrial carbonates, Sala area, Sweden. *Journal of Geochemical Exploration*, 233,

106909. <https://doi.org/10.1016/J.GEXPLO.2021.106909>
- Javidan, S. M., Banakar, A., Vakilian, K. A., & Ampatzidis, Y. (2023). Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning. *Smart Agricultural Technology*, 3, 100081. <https://doi.org/10.1016/J.ATECH.2022.100081>
- Junjie, J., Wenhao, S., & Yuan, W. (2024). A risk assessment approach for road collapse along tunnels based on an improved entropy weight method and K-means cluster algorithm. *Ain Shams Engineering Journal*, 15(7), 102805. <https://doi.org/10.1016/J.ASEJ.2024.102805>
- Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25. <https://doi.org/10.1016/J.BUSHOR.2018.08.004>
- Khan, F., Khan, O., Parvez, M., Ahmad, S., Yahya, Z., Alhodaib, A., Kumar Yadav, A., & Ağbulut, Ü. (2024). K-means clustering optimization of various quantum dots and nanoparticles-added biofuels for engine performance, emission, vibration, and noise characteristics. *Thermal Science and Engineering Progress*, 54, 102815. <https://doi.org/10.1016/J.TSEP.2024.102815>
- Khan, I. K., Daud, H. B., Zainuddin, N. B., Sokkalingam, R., Farooq, M., Baig, M. E., Ayub, G., & Zafar, M. (2024). Determining the optimal number of clusters by Enhanced Gap Statistic in K-mean algorithm. *Egyptian Informatics Journal*, 27, 100504. <https://doi.org/10.1016/J.EIJ.2024.100504>
- Kim, H., Kim, H. K., & Cho, S. (2020). Improving spherical k-means for document clustering: Fast initialization, sparse centroid projection, and efficient cluster labeling. *Expert Systems with Applications*, 150, 113288. <https://doi.org/10.1016/J.ESWA.2020.113288>
- Lee, L. C., & Jemain, A. A. (2021). On overview of PCA application strategy in processing high dimensionality forensic data. *Microchemical Journal*, 169, 106608. <https://doi.org/10.1016/J.MICROC.2021.106608>
- López, J. F., Sánchez, M. E., Pomaquero, J. C., & Vasco, J. A. (2024). Regulaciones en la ley de economía social del sector transporte-Ecuador. *Revista Venezolana de Gerencia*, 29(Especial 11), 279-292. <https://doi.org/10.52080/rvgluz.29.e11.16>
- Lv, Y., Zhi, D., Sun, H., & Qi, G. (2021). Mobility pattern recognition based prediction for the subway station related bike-sharing trips. *Transportation Research Part C: Emerging Technologies*, 133, 103404. <https://doi.org/10.1016/J.TRC.2021.103404>
- Ma, Y., Li, W., Tang, K., Zhang, Z., & Chen, S. (2021). Driving style recognition and comparisons among driving tasks based on driver behavior in the online car-hailing industry. *Accident Analysis & Prevention*, 154, 106096. <https://doi.org/10.1016/J.AAP.2021.106096>
- Mehedi Hassan, M., Mollick, S., & Yasmin, F. (2022). An unsupervised cluster-based feature grouping model for early diabetes detection. *Healthcare Analytics*, 2, 100112. <https://doi.org/10.1016/J.HEALTH.2022.100112>
- Miskolczi, M., Földes, D., Munkácsy, A., & Jászberényi, M. (2021). Urban mobility scenarios until the 2030s. *Sustainable Cities and Society*, 72, 103029. <https://doi.org/10.1016/J.SCS.2021.103029>

- Mussabayev, R., Mladenovic, N., Jarboui, B., & Mussabayev, R. (2023). How to Use K-means for Big Data Clustering? *Pattern Recognition*, 137, 109269. <https://doi.org/10.1016/J.PATCOG.2022.109269>
- Naghizadeh, A., & Metaxas, D. N. (2020). Condensed Silhouette: An Optimized Filtering Process for Cluster Selection in K-Means. *Procedia Computer Science*, 176, 205–214. <https://doi.org/10.1016/J.PROCS.2020.08.022>
- Narayanan, S., Chaniotakis, E., & Antoniou, C. (2020). Shared autonomous vehicle services: A comprehensive review. *Transportation Research Part C: Emerging Technologies*, 111, 255–293. <https://doi.org/10.1016/J.TRC.2019.12.008>
- Ning, Z., Chen, J., Huang, J., Sabo, U. J., Yuan, Z., & Dai, Z. (2022). WeDIV – An improved k-means clustering algorithm with a weighted distance and a novel internal validation index. *Egyptian Informatics Journal*, 23(4), 133–144. <https://doi.org/10.1016/J.EIJ.2022.09.002>
- Nowak-Brzezinska, A., & Horyn, C. (2020). Outliers in rules - the comparison of LOF, COF and KMEANS algorithms. *Procedia Computer Science*, 176, 1420–1429. <https://doi.org/10.1016/J.PROCS.2020.09.152>
- Pellegrino, N., Fieguth, P. W., & Haji Reza, P. (2023). K-Means for noise-insensitive multi-dimensional feature learning. *Pattern Recognition Letters*, 170, 113–120. <https://doi.org/10.1016/J.PATREC.2023.04.009>
- Ragunthar, T., Ashok, P., Gopinath, N., & Subashini, M. (2021). A strong reinforcement parallel implementation of k-means algorithm using message passing interface. *Materials Today: Proceedings*, 46, 3799–3802. <https://doi.org/10.1016/J.MATPR.2021.02.032>
- Rahman, M. M., & Thill, J. C. (2023). Impacts of connected and autonomous vehicles on urban transportation and environment: A comprehensive review. *Sustainable Cities and Society*, 96, 104649. <https://doi.org/10.1016/J.SCS.2023.104649>
- Ran, X., Suyaraj, N., Tepsan, W., Ma, J., Zhou, X., & Deng, W. (2024). A hybrid genetic-fuzzy ant colony optimization algorithm for automatic K-means clustering in urban global positioning system. *Engineering Applications of Artificial Intelligence*, 137, 109237. <https://doi.org/10.1016/J.ENGAPAI.2024.109237>
- Ros, F., Riad, R., & Guillaume, S. (2023). PDBI: A partitioning Davies-Bouldin index for clustering evaluation. *Neurocomputing*, 528, 178–199. <https://doi.org/10.1016/J.NEUCOM.2023.01.043>
- Schreiber, J. B. (2021). Issues and recommendations for exploratory factor analysis and principal component analysis. *Research in Social and Administrative Pharmacy*, 17(5), 1004–1011. <https://doi.org/10.1016/J.SAPHARM.2020.07.027>
- Sun, Y., Liu, H., & Gao, Y. (2023). Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model. *Heliyon*, 9(2), e13384. <https://doi.org/10.1016/J.HELIVON.2023.E13384>
- Taghvaei, V. M., Nodehi, M., Saber, R. M., & Mohebi, M. (2022). Sustainable development goals and transportation modes: Analyzing sustainability pillars of environment, health, and economy. *World Development Sustainability*, 1, 100018. <https://doi.org/10.1016/J.WDS.2022.100018>

[org/10.1016/J.WDS.2022.100018](https://doi.org/10.1016/J.WDS.2022.100018)

- Tang, R., De Donato, L., Bešinović, N., Flammini, F., Goverde, R. M. P., Lin, Z., Liu, R., Tang, T., Vittorini, V., & Wang, Z. (2022). A literature review of Artificial Intelligence applications in railway systems. *Transportation Research Part C: Emerging Technologies*, 140, 103679. <https://doi.org/10.1016/J.TRC.2022.103679>
- Troccoli, E. B., Cerqueira, A. G., Lemos, J. B., & Holz, M. (2022). K-means clustering using principal component analysis to automate label organization in multi-attribute seismic facies analysis. *Journal of Applied Geophysics*, 198, 104555. <https://doi.org/10.1016/J.JAPGEO.2022.104555>
- Wei, Q. (2024). Accounting Data Encryption Processing Based on K-Means Clustering Algorithm. *Procedia Computer Science*, 247, 819–825. <https://doi.org/10.1016/J.PROCS.2024.10.099>
- Yarlagadda, J., Jain, P., & Pawar, D. S. (2021). Assessing safety critical driving patterns of heavy passenger vehicle drivers using instrumented vehicle data – An unsupervised approach. *Accident Analysis & Prevention*, 163, 106464. <https://doi.org/10.1016/J.AAP.2021.106464>
- Yarushkina, N., Matyugina, E., & Vanina, I. (2022). Transport integration in providing the economic development of the territory. *Transportation Research Procedia*, 63, 486–494. <https://doi.org/10.1016/J.TRPRO.2022.06.039>
- Zhang, C., Lasaulce, S., Hennebel, M., Saludjian, L., Panciatici, P., & Poor, H. V. (2021). Decision-making oriented clustering: Application to pricing and power consumption scheduling. *Applied Energy*, 297, 117106. <https://doi.org/10.1016/J.APENERGY.2021.117106>
- Zhao, W., Ma, J., Liu, Q., Song, J., Tysklind, M., Liu, C., Wang, D., Qu, Y., Wu, Y., & Wu, F. (2023). Comparison and application of SOFM, fuzzy c-means and k-means clustering algorithms for natural soil environment regionalization in China. *Environmental Research*, 216, 114519. <https://doi.org/10.1016/J.ENVRES.2022.114519>
- Zhu, J., Ji, S., Yu, J., Shao, H., Wen, H., Zhang, H., Xia, Z., Zhang, Z., & Lee, C. (2022). Machine learning-augmented wearable triboelectric human-machine interface in motion identification and virtual reality. *Nano Energy*, 103, 107766. <https://doi.org/10.1016/J.NANOEN.2022.107766>