



**Evolución Y Análisis de la Adopción de Vehículos Eléctricos e Híbridos en
Colombia: Un Enfoque de Machine Learning**

Diego Fernando Alfaro Quintero

Universidad EAN

Facultad de Ingeniería

Maestría en Ciencia de Datos

Bogotá, Colombia

22/01/2026

**Evolución Y Análisis de la Adopción de Vehículos Eléctricos e Híbridos en Colombia:
Un Enfoque de Machine Learning**

Diego Fernando Alfaro Quintero

Trabajo de grado presentado como requisito para optar al título de:

Magister en Ciencia de Datos

Directora:

Carolina María Luque Zabala

Modalidad:

Monografía

Universidad EAN

Facultad de Ingeniería

Maestría en Ciencia de Datos

Bogotá, Colombia

22/01/2026

EVOLUCIÓN Y ANÁLISIS DE LA ADOPCIÓN DE VEHÍCULOS
ELÉCTRICOS E HÍBRIDOS EN COLOMBIA: UN ENFOQUE DE
MACHINE LEARNING

3

Nota de aceptación:

Firma del jurado

Firma del jurado

Firma del director del trabajo de grado

Bogotá, día/mes/año

EVOLUCIÓN Y ANÁLISIS DE LA ADOPCIÓN DE VEHÍCULOS
ELÉCTRICOS E HÍBRIDOS EN COLOMBIA: UN ENFOQUE DE
MACHINE LEARNING

4

La información es el petróleo del siglo XXI
y el análisis es el motor de combustión.

Peter Sondergaard

Agradecimientos

A Dios, por darme fortaleza, sabiduría y constancia para completar este camino de formación y aprendizaje.

A la universidad EAN, por ofrecer formación de calidad y con visión de futuro, y a los profesores de la Maestría en Ciencia de Datos, cuyo conocimiento y compromiso fueron fundamentales para este recorrido formativo.

Un agradecimiento especial a la profesora Carolina María Luque Zabala, por su invaluable orientación, su disposición permanente y su exigencia académica, que guiaron el presente trabajo en cada una de sus etapas.

Finalmente, a mi familia, por caminar a mi lado en cada desafío y celebrar conmigo cada logro alcanzado. Esta meta cumplida también les pertenece.

Resumen

Esta monografía analiza la adopción de vehículos eléctricos e híbridos en Colombia utilizando técnicas de aprendizaje automático aplicadas a datos vehiculares y sociodemográficos de fuentes oficiales. Se lleva a cabo una segmentación no supervisada de registros de la flota actual de vehículos de este tipo y se aplican métodos de interpretabilidad para identificar los factores más determinantes en cada grupo. Adicionalmente, se proyecta la adopción futura de cada segmento. El estudio evidencia patrones en la adopción de la movilidad sostenible asociados a características técnicas y sociodemográficas que pueden apoyar la toma de decisiones que impulsen la transición energética en el sector transporte y, por tanto, la disminución de emisiones de carbono.

Palabras clave: Vehículos eléctricos, vehículos híbridos, aprendizaje automático, segmentación, métodos de interpretabilidad, transición energética.

Abstract

This monograph analyzes the adoption of electric and hybrid vehicles in Colombia using machine learning techniques applied to vehicle and sociodemographic data from official sources. An unsupervised segmentation of the current fleet records for these types of vehicles is conducted, and interpretability methods are applied to identify the most influential factors within each group. Additionally, the future adoption of each segment is projected. The study reveals patterns in the adoption of sustainable mobility associated with technical and sociodemographic characteristics, which can support decision-making processes aimed at promoting the energy transition in the transportation sector and, consequently, reducing carbon emissions.

Keywords: Electric vehicles, hybrid vehicles, machine learning, segmentation, interpretability methods, energy transition.

Contenido

Resumen	6
Abstract	7
Contenido	8
Lista de Figuras	10
Lista de Tablas	11
1. Introducción	13
2. Objetivos	17
<i>2.1. Objetivo general</i>	<i>17</i>
<i>2.2. Objetivos específicos</i>	<i>17</i>
3. Justificación	19
4. Marco Teórico	22
<i>4.1. Vehículos eléctricos e híbridos</i>	<i>22</i>
4.1.1. Vehículos eléctricos.....	22
4.1.2. Vehículos híbridos	22
4.1.3. Efectos de los vehículos eléctricos e híbridos	23
<i>4.2. Dificultades para la adopción de vehículos eléctricos e híbridos</i>	<i>24</i>
<i>4.3. Adopción de tecnologías limpias en Colombia</i>	<i>24</i>

4.3.1. Ley 1964 de 2019.....	24
4.3.2. Estrategia Nacional de Movilidad Eléctrica.....	25
4.3.3. Panorama Reciente de la Adopción de Vehículos Eléctricos e Híbridos en Colombia.....	25
4.3.4. Estudios en Colombia.....	27
4.4. <i>Introducción a la Segmentación de mercado</i>	28
4.5. <i>Criterios de Segmentación Clásicos</i>	28
4.5.4. Limitaciones de los Métodos Tradicionales	29
4.6. <i>Introducción a Aprendizaje Automático</i>	30
4.6.1. Beneficios de Usar Aprendizaje Automático para Segmentación	30
4.6.2. Segmentación con Aprendizaje Automático	31
4.6.3. Análisis de Segmentos con Aprendizaje Automático	31
4.6.4. Similitud Basada en Distancia.....	32
4.6.4.1. Distancia Euclidiana.	32
4.6.4.2. Distancia Manhattan.....	32
4.6.4.3. Distancia Gower.....	33
4.7. <i>Técnicas de Aprendizaje Automático Para la Segmentación No Supervisada</i>	33
4.7.1. K-means.....	33
4.7.1.1. CatBoost con MiniBatch K-Means.....	34
4.7.2. K-prototypes	35
4.7.3. Agrupamiento Jerárquico	36

EVOLUCIÓN Y ANÁLISIS DE LA ADOPCIÓN DE VEHÍCULOS ELÉCTRICOS E HÍBRIDOS EN COLOMBIA: UN ENFOQUE DE MACHINE LEARNING	10
4.7.4. Supuestos estadísticos de los algoritmos utilizados	36
4.8. Métricas utilizadas en la segmentación de datos.....	37
4.8.1. Ley del codo	37
4.8.2. Puntaje de silueta	37
4.8.3. índice de Davies-Bouldin.....	38
4.9. Interpretabilidad de Modelos con SHAP.....	38
4.10. Reducción y visualización de datos con PCA-Biplot.....	38
4.11. Predicción de tendencias con Prophet.....	39
4.12. Desafíos de la Aplicación del Aprendizaje Automático en Segmentación de Mercado.....	40
4.13. Trabajos relacionados.....	41
5. Hipótesis	44
5.1. Hipótesis Sobre Brechas Socioeconómicas y Territoriales.....	44
5.2. Hipótesis Sobre Segmentación Algorítmica y Caracterización Técnica	45
6. Variables	46
7. Metodología	49
7.1. Enfoque y alcance de la investigación.....	49
7.2. Población y muestra.....	50
7.3. Técnicas para el análisis de la información.....	50

8. Trabajo de Campo.....	57
8.1. <i>Procesamiento de Datos.....</i>	57
8.2. <i>Análisis Exploratorio de Datos a Nivel Nacional</i>	59
8.3. <i>Formación de Subgrupos.....</i>	64
8.4. <i>Segmentación de los Subgrupos.....</i>	78
8.4.1. Resultados de Segmentación en Subgrupo Motos	82
8.4.2. Resultados de Segmentación en Subgrupo Particular Oficial Ligero ...	85
8.4.3. Resultados de Segmentación en Subgrupo Público Ligero	87
8.4.4. Resultados de Segmentación en Subgrupo Público Pesado.....	90
8.4.5. Resultados de Segmentación en Subgrupo Carga	91
8.5. <i>Evolución de los Segmentos 2010 a 2022</i>	93
8.6. <i>Proyección de los Segmentos a 2028</i>	98
8.6.1. Proyección del Subgrupo Motos	100
8.6.2. Proyección del Subgrupo Particular Oficial Ligero	105
8.6.3. Proyección del Subgrupo Público Ligero	111
8.6.4. Proyección del Subgrupo Público Pesado.....	113
8.6.5. Proyección del Subgrupo Carga	116
8.7. <i>Análisis Enriquecido con Variables Socioeconómicas y de Infraestructura</i>	120
8.7.1. Análisis de Valores SHAP	122
8.7.2. Análisis de Componentes Principales con Biplot.....	126

EVOLUCIÓN Y ANÁLISIS DE LA ADOPCIÓN DE VEHÍCULOS ELÉCTRICOS E HÍBRIDOS EN COLOMBIA: UN ENFOQUE DE MACHINE LEARNING	12
8.8. Verificación de Hipótesis y Propuestas Sobre los Resultados	132
9. Discusión	134
10. Conclusiones y Trabajo Futuro.....	137
10.1. Conclusiones.....	137
10.2. Trabajo futuro	139
Referencias.....	142

Lista de Figuras

Figura 1. Venta histórica total de vehículos eléctricos e híbridos en Colombia en unidades 2014-2024.....	26
Figura 2. Clústeres formados por K-means.....	34
Figura 3. División en subgrupos de acuerdo a clase de vehículo.	51
Figura 4. Proceso de consolidación de dataframe unificado para análisis integral.	54
Figura 5. Distribución del parque automotor eléctrico e híbrido por clase....	60
Figura 6. Tendencia por tipo de servicio 2010-2022 en el parque automotor eléctrico e híbrido por clase.....	61
Figura 7. Evolución del parque automotor eléctrico e híbrido 2010-2022.....	62
Figura 8. Porcentaje de mercado en subgrupo de motos.....	65
Figura 9. Porcentaje de mercado en subgrupo de particular y oficial ligero.	66
Figura 10. Porcentaje de mercado en subgrupo de público ligero.	67
Figura 11. Porcentaje de mercado en subgrupo de público pesado.....	68
Figura 12. Porcentaje de mercado en subgrupo de carga.....	69
Figura 13. Distribución departamental del subgrupo de motos.	71

Figura 14. Distribución departamental del subgrupo de particular y oficial ligero.....	72
Figura 15. Distribución departamental del subgrupo de público ligero.	74
Figura 16. Distribución departamental del subgrupo de público pesado.	75
Figura 17. Distribución departamental del subgrupo de carga.....	77
Figura 18. Proporción de fuente de energía por clúster.	83
Figura 19. Proporción de clase por clúster.	83
Figura 20. Proporción de marca por clúster.....	84
Figura 21. Proporción de fuente de energía por clúster.	86
Figura 22. Proporción de clase por clúster.	86
Figura 23. Proporción de marca por clúster.....	86
Figura 24. Proporción de fuente de energía por clúster.	88
Figura 25. Proporción de clase por clúster.	88
Figura 26. Proporción de marca por clúster.....	89
Figura 27. Proporción de fuente de energía por clúster.	90
Figura 28. Proporción de clase por clúster.	90

Figura 29. Proporción de marca por clúster.....	91
Figura 30. Proporción de fuente de energía por clúster.	92
Figura 31. Proporción de clase por clúster.	92
Figura 32. Proporción de marca por clúster.....	93
Figura 33. Evolución de registros por año de cada segmento en el subgrupo motos.....	94
Figura 34. Evolución de registros por año de cada segmento en el subgrupo particular y oficial ligero.	95
Figura 35. Evolución de registros por año de cada segmento en el subgrupo público ligero.	96
Figura 36. Evolución de registros por año de cada segmento en el subgrupo público pesado.....	97
Figura 37. Evolución de registros por año de cada segmento en el subgrupo carga.....	97
Figura 38. Pronóstico Prophet para subgrupo motos segmento 0.	101
Figura 39. Pronóstico Prophet para subgrupo motos segmento 3.	102
Figura 40. Pronóstico Prophet para subgrupo motos segmento 1.	103

Figura 41. Pronóstico Prophet para subgrupo motos segmento 4.	103
Figura 42. Pronóstico Prophet para subgrupo motos segmento 5.	104
Figura 43. Pronóstico Prophet para subgrupo particular y oficial ligeros segmento 0.	105
Figura 44. Pronóstico Prophet para subgrupo particular y oficial ligeros segmento 1.	106
Figura 45. Pronóstico Prophet para subgrupo particular y oficial ligeros segmento 2.	107
Figura 46. Pronóstico Prophet para subgrupo particular y oficial ligeros segmento 3.	108
Figura 47. Pronóstico Prophet para subgrupo particular y oficial ligeros segmento 4.	109
Figura 48. Pronóstico Prophet para subgrupo particular y oficial ligeros segmento 5.	110
Figura 49. Pronóstico Prophet para subgrupo público ligero segmento 3. .	112
Figura 50. Pronóstico Prophet para subgrupo público ligero segmento 4. .	112
Figura 51. Pronóstico Prophet para subgrupo público pesado segmento 0.	114

Figura 52. Pronóstico Prophet para subgrupo público pesado segmento 1.	115
Figura 53. Pronóstico Prophet para subgrupo público pesado segmento 2.	115
Figura 54. Pronóstico Prophet para subgrupo carga segmento 0.	117
Figura 55. Pronóstico Prophet para subgrupo carga segmento 1.	118
Figura 56. Pronóstico Prophet para subgrupo carga segmento 2.	118
Figura 57. Pronóstico Prophet para subgrupo carga segmento 3.	119
Figura 58. Análisis con SHAP para el subgrupo motos.....	123
Figura 59. Análisis con SHAP para el subgrupo particular y oficial ligero. .	123
Figura 60. Análisis con SHAP para el subgrupo público ligero.....	124
Figura 61. Análisis con SHAP para el subgrupo público pesado.....	125
Figura 62. Análisis con SHAP para el subgrupo carga.	125
Figura 63. PCA Biplot para el subgrupo motos.	127
Figura 64. PCA Biplot para el subgrupo particular y oficial ligero.	127
Figura 65. PCA Biplot para el subgrupo público ligero.	128

Figura 66. PCA Biplot para el subgrupo público pesado..... 129

Figura 67. PCA Biplot para el subgrupo carga..... 130

Lista de Tablas

Tabla 1. Variables utilizadas en la presente monografía.	46
Tabla 2. Conteo de registros para base final consolidada.	59
Tabla 3. Medidas estadísticas del grupo de datos utilizado.....	63
Tabla 4. Conteo de registros de cada subgrupo generado.	64
Tabla 5. Registros per cápita a nivel departamental del subgrupo de motos.	72
Tabla 6. Registros per cápita a nivel departamental del subgrupo de particular y oficial ligero.....	73
Tabla 7. Registros per cápita a nivel departamental del subgrupo de público ligero.....	74
Tabla 8. Registros per cápita a nivel departamental del subgrupo de público pesado.....	76
Tabla 9. Registros per cápita a nivel departamental del subgrupo de carga.	77
Tabla 10. Técnicas de segmentación utilizadas por cada subgrupo.	79
Tabla 11. Resultados de las técnicas de segmentación no supervisada ejecutadas en cada subgrupo.	81
Tabla 12. Segmentación de subgrupo moto.....	83

Tabla 13. Segmentación de subgrupo particular y oficial ligero.	85
Tabla 14. Segmentación de subgrupo público ligero.....	88
Tabla 15. Segmentación de subgrupo público pesado.	90
Tabla 16. Segmentación de subgrupo carga.....	91
Tabla 17. Métricas de proyecciones para subgrupo motos.	100
Tabla 18. Métricas de proyecciones para subgrupo particular oficial ligero.	105
Tabla 19. Métricas de proyecciones para subgrupo público ligero.	111
Tabla 20. Métricas de proyecciones para subgrupo público pesado.....	113
Tabla 21. Métricas de proyecciones para subgrupo carga.....	116
Tabla 22. Estructura de datasets de análisis socioeconómico y de infraestructura departamental.....	121

1. Introducción

La descarbonización del transporte constituye uno de los principales desafíos en la transición energética global, dado el imperativo de reducir la dependencia de los combustibles fósiles y mitigar el cambio climático (Kang, 2025). En este proceso, los vehículos eléctricos e híbridos han adquirido un rol estratégico; estos no solo ofrecen una eficiencia superior en la conversión de energía en movimiento frente a los motores de combustión (Thomas, 2009), sino que permiten una reducción significativa de las emisiones de gases de efecto invernadero y la contaminación auditiva, especialmente en entornos urbanos (Un-Noor et al., 2017). Para impulsar su penetración en el mercado y superar barreras económicas como los altos costos iniciales, diversos países han implementado incentivos económicos y marcos regulatorios, lo que ha generado un creciente interés académico en comprender los factores determinantes de esta transición tecnológica desde una perspectiva multidisciplinar apoyada en el análisis de datos (Bas et al., 2021).

En Colombia, se han implementado diversas medidas de fomento, como la Ley 1964 de 2019 y la Estrategia Nacional de Movilidad Eléctrica, que incluyen incentivos tributarios y beneficios arancelarios. Sin embargo, la efectividad de estas políticas enfrenta un obstáculo mayor: la fragmentación y calidad de la información disponible. Estudios previos advierten que ignorar la interacción compleja entre factores de adopción, por ejemplo, realidades socioeconómicas territoriales, puede derivar en políticas públicas poco efectivas (Lopez-Arboleda et al., 2021), una situación agravada por las dificultades inherentes a la gestión de datos desconectados o incompletos que limitan la caracterización del mercado (Reddy et al., 2025). Esta falta de integración impide identificar con claridad qué factores

locales están frenando o impulsando la adopción en las distintas regiones del país. No obstante, los registros oficiales de vehículos constituyen una fuente valiosa para caracterizar la evolución de la flota y, al complementarse con información sociodemográfica y geoespacial, permiten analizar de manera integral las dinámicas de adopción tecnológica en el país y sus particularidades regionales.

La literatura internacional ha documentado que factores como el ingreso, la densidad urbana y la infraestructura son determinantes clave (Hardaway et al., 2025) (Ramadoss et al., 2025). No obstante, en Colombia la investigación se ha enfocado principalmente en proyecciones generales de ventas o análisis normativos (Campo & Pino, 2024) (Lopez-Arboleda et al., 2021), sin explorar sistemáticamente la relación entre las especificaciones técnicas de la flota y el perfil sociodemográfico de los departamentos. Asimismo, la aplicación de técnicas de aprendizaje automático con enfoque interpretativo es incipiente en el contexto nacional, limitando la capacidad de los tomadores de decisiones para entender las causas raíz detrás de los patrones de adopción.

Para abordar esta brecha, este estudio propone una metodología para analizar la movilidad híbrida y eléctrica en Colombia mediante la construcción de un pipeline de procesamiento de datos que depura, estandariza y enriquece los registros históricos del Registro Único Nacional de Tránsito (RUNT) para el periodo 2010-2022, integrándolos con variables sociodemográficas y de infraestructura a nivel departamental. La contribución de este trabajo radica en la aplicación de algoritmos de segmentación no supervisada para identificar nichos de mercado (Venkata Reddy Konasani & Shailendra Kadre, 2021), complementada con técnicas de interpretabilidad como SHAP (Marcilio & Eler, 2020) para jerarquizar los factores de influencia. Adicionalmente, se incorpora el modelo Prophet

(Zherlitsyn et al., 2025), una herramienta de series de tiempo que ya ha sido validada en estudios similares en el contexto colombiano (Campo & Pino, 2024), para proyectar los escenarios de adopción hasta el año 2028. De esta forma, la investigación realiza un aporte al campo de la Ciencia de Datos aplicada a la movilidad sostenible, evidenciando cómo el análisis de datos puede favorecer la comprensión del fenómeno de adopción de vehículos híbridos y eléctricos en Colombia y así mismo la toma de decisiones para la transición energética.

La pregunta que guía esta investigación es: ¿Qué factores sociodemográficos y técnicos influyen en la adopción de vehículos eléctricos e híbridos en Colombia y cómo varían entre segmentos identificados mediante técnicas de aprendizaje automático? El objetivo es analizar los patrones de composición técnica, distribución geográfica y evolución temporal de la flota eléctrica e híbrida en Colombia para identificar los factores que impulsan su adopción, contribuyendo así con evidencia empírica que apoye la toma de decisiones tanto en políticas públicas como en estrategias de mercado orientadas a una movilidad sostenible.

El presente documento se organiza en 10 capítulos: introducción, donde se da contexto de la problemática expuesta anteriormente; objetivos, que describen el alcance del estudio; justificación, que expone la relevancia de este análisis en diferentes sectores; marco teórico, que explica los principales conceptos involucrados; hipótesis, en el que se establecen los supuestos a contrastar con los resultados; variables, que detalla los datos utilizados; metodología, que describe el pipeline y las técnicas utilizadas; trabajo de campo, que muestra los resultados obtenidos tras la preparación de la información y aplicación de los algoritmos de aprendizaje automático; discusión, donde se interpretan los resultados;

conclusiones y líneas de trabajo futuro, que sintetiza los hallazgos más relevantes relacionados a cada objetivo fijado y plantea posibles evolutivos de la investigación.

2. Objetivos

2.1. Objetivo general

Analizar los patrones en la composición técnica, la distribución geográfica y la evolución temporal de los vehículos eléctricos e híbridos en Colombia, identificando los factores que impulsan su adopción mediante el uso de técnicas de aprendizaje automático orientadas a segmentación no supervisada, modelado de proyecciones y análisis interpretativo de variables socioeconómicas.

2.2. Objetivos específicos

- Realizar una revisión de estudios nacionales e internacionales que empleen registros vehiculares para analizar la adopción de vehículos eléctricos e híbridos, con el fin de identificar metodologías de segmentación y modelos interpretativos que orienten la formulación de una propuesta metodológica adaptada al contexto colombiano.
- Desarrollar un pipeline de procesamiento de datos que garantice la depuración y estandarización de los registros automotores oficiales de vehículos eléctricos e híbridos en Colombia, y que permita su integración con variables sociodemográficas, económicas y geoespaciales a nivel departamental.
- Implementar y comparar algoritmos de aprendizaje automático no supervisado con el fin de identificar segmentos de adopción vehicular, y caracterizar cada segmento según su composición técnica y evolución temporal.
- Construir perfiles detallados de los segmentos de adopción vehicular identificados y derivar sus implicaciones para políticas públicas y estrategias comerciales, así

como proponer líneas futuras de investigación que fortalezcan el monitoreo y la proyección de la adopción en Colombia.

3. Justificación

La adopción de vehículos eléctricos e híbridos es un tema clave en el que intervienen factores socioeconómicos, políticos, ambientales y empresariales (Naseri et al., 2024). Colombia, al igual que todo el mundo, enfrenta desafíos complejos respecto al manejo del cambio climático, la contaminación y el uso de combustibles fósiles. El análisis de la transición hacia tecnologías más limpias es crucial para entender el avance que se ha logrado hasta la actualidad y para diseñar políticas públicas y privadas que sigan apoyando este proceso.

Desde el punto de vista socioeconómico, el uso creciente de esta tecnología puede generar impactos positivos, como creación de puestos de trabajo en sectores afines y reducción de enfermedades provocadas por la contaminación. No obstante, existen retos importantes, como la inversión inicial de alto costo, la infraestructura limitada y la brecha tecnológica entre regiones (Shareef et al., 2016). Por esto, se hace necesaria la identificación de patrones de adopción y características socioeconómicas con el objetivo de comprender mejor el panorama.

En el ámbito político, Colombia tiene leyes y planes que fomentan el uso de vehículos eléctricos e híbridos, y que se alinean con sus compromisos a nivel internacional para cumplir con la agenda de lucha contra el cambio climático (Ministerio del Ambiente y Desarrollo Sostenible et al., 2019). El seguimiento y la adaptación de estas políticas deben fundamentarse en un conocimiento profundo y actualizado sobre el progreso de la adopción en el país, lo que permitirá diseñar incentivos y regulaciones más eficaces y adaptados a las particularidades de cada región.

Desde la perspectiva empresarial, la movilidad limpia plantea una oportunidad estratégica para industrias como la automotriz, proveedores de tecnología y servicios relacionados. La adopción creciente de esta tecnología impulsa proyectos de innovación, inversión en infraestructura y desarrollo de negocios entorno a la movilidad sostenible. El conocimiento detallado del mercado ayuda a las empresas a enfocarse en segmentos clave, generar estrategias comerciales efectivas y crear alianzas que impulsen el crecimiento del sector (Pyzdek & Keller, 2013).

Desde el punto de vista ambiental, el uso de vehículos eléctricos e híbridos ayuda a reducir las emisiones contaminantes y la huella de carbono generada por el sector transporte (Un-Noor et al., 2017). Con la generación de estrategias que fomenten la adopción, se pueden maximizar los beneficios ecológicos de esta tecnología.

Técnicamente, el uso de segmentación no supervisada y otras técnicas de aprendizaje automático permiten crear grupos con características similares según variables diversas (Jauregui Romero et al., 2024). Esto facilita la identificación de regiones con adopción avanzada o incipiente y factores socioeconómicos influyentes. El uso de aprendizaje automático proporciona hallazgos clave que los métodos tradicionales no pueden, mejorando el análisis y convirtiéndose en un insumo crucial para tomar decisiones (Bogacki et al., 2024).

Esta monografía se alinea con la Política de Sostenibilidad y Emprendimiento Sostenible de la Universidad EAN al proponer un análisis técnico de los datos disponibles, integrando variables de tipo social, ambiental y económico mediante aprendizaje automático, con la visión de favorecer la toma de decisiones sostenibles y apoyar el crecimiento de la movilidad limpia en el contexto colombiano.

EVOLUCIÓN Y ANÁLISIS DE LA ADOPCIÓN DE VEHÍCULOS
ELÉCTRICOS E HÍBRIDOS EN COLOMBIA: UN ENFOQUE DE
MACHINE LEARNING

21

Este trabajo de grado se desarrolla en la línea de investigación del grupo de Ciencias básicas - Estadística aplicada y ciencia de datos, dirigido por la profesora Carolina María Luque Zabala.

4. Marco Teórico

4.1. Vehículos eléctricos e híbridos

En esta sección se hará una introducción a los conceptos fundamentales de los vehículos eléctricos e híbridos, su funcionamiento y tecnología.

4.1.1. Vehículos eléctricos

Tal como explica Sadeq (2024), los vehículos eléctricos funcionan mediante uno o más motores eléctricos impulsados por baterías recargables. Tienen la capacidad de convertir más del 60% de la energía eléctrica en movimiento, lo cual supera la eficiencia de vehículos con motor de combustión interna. Destacan por ser una opción ecológica y con costos operativos menores debido a su bajo mantenimiento. Estos vehículos usan un tren motriz totalmente eléctrico. Su autonomía ha mejorado con los avances en baterías, lo que ha mejorado su desempeño en viajes largos.

4.1.2. Vehículos híbridos

Los vehículos híbridos usan dos o más fuentes de energía. La combinación más común es un motor de combustión interna con un motor eléctrico y una batería. Existen dos configuraciones básicas: híbrido tipo serie, en el que el vehículo es impulsado únicamente por el motor eléctrico; e híbrido tipo paralelo, en el que el vehículo puede ser impulsado por el motor eléctrico, el de combustión interna o ambos de forma simultánea, según las exigencias y condiciones del trayecto. El tipo serie es más adecuado para la ciudad, y el tipo paralelo es más eficiente en carretera y a altas velocidades. Los híbridos también se pueden clasificar en no recargables, que dependen del combustible fósil, y recargables, que tienen baterías más grandes para viajar distancias considerables solo

con energía eléctrica. Ambos tipos pueden usar freno regenerativo, que recupera energía durante la desaceleración para recargar la batería (Larminie & Lowry, 2012).

4.1.3. Efectos de los vehículos eléctricos e híbridos

Desde el punto de vista ambiental, los vehículos eléctricos ayudan a disminuir significativamente las emisiones de gases de efecto invernadero y la contaminación auditiva, especialmente en zonas urbanas (Un-Noor et al., 2017). Aunque su carga puede generar emisiones si la electricidad proviene de fuentes fósiles (Ma et al., 2012), en general producen menos emisiones que los vehículos de combustión interna, especialmente cuando se integran con energías renovables. A lo largo de su vida útil, los vehículos eléctricos son menos contaminantes, aunque el reciclaje de las baterías sigue representando un reto importante (Shareef et al., 2016).

Bajo la óptica económica, los vehículos eléctricos ofrecen menores costos operativos debido a su mayor eficiencia (Thomas, 2009). Actualmente, el costo de los vehículos eléctricos es alto, pero probablemente disminuirá debido a la producción en masa y a mejores políticas energéticas, esto podría aumentar los beneficios económicos para los propietarios (Forero & Mihet-Popa, 2014).

Respecto a la red eléctrica, los vehículos eléctricos son cargas de alta potencia, que pueden desestabilizar la red de distribución, especialmente componentes como transformadores, cables y fusibles. Efectos adversos como sobrecargas, inestabilidad del voltaje y deterioro prematuro de los equipos puede suceder principalmente con la carga en horas pico, por lo que es recomendable realizar la carga en horarios de baja demanda (Un-Noor et al., 2017).

Por su parte, los vehículos híbridos también tienen beneficios en el entorno ambiental como la reducción del consumo de combustibles y emisiones contaminantes; y en el entorno económico, como el retorno de la inversión para los usuarios al ahorrar en combustibles costosos. Además, su desarrollo ha servido como base en la transición a los vehículos totalmente eléctricos y otras tecnologías de movilidad sostenible (Prakash Kumbar, 2020).

4.2. Dificultades para la adopción de vehículos eléctricos e híbridos

Pese a presentar varios beneficios, la adopción de vehículos eléctricos e híbridos presenta varios retos. La autonomía limitada y tiempos de carga prolongados generan dudas entre los usuarios, en especial en lugares con infraestructura de recarga insuficiente (Shareef et al., 2016). El temor por quedarse sin carga persiste y el desarrollo de tecnologías de carga rápida está en proceso, lo que mantiene esta desventaja frente a vehículos de combustión interna (Un-Noor et al., 2017).

También existen factores sociales y económicos que se deben tener en cuenta. La transición hacia la movilidad sostenible implica cambios en hábitos de uso y recarga, y mayores costos en algunos casos, lo que genera resistencia al cambio (Un-Noor et al., 2017). En conjunto, estos desafíos muestran que la transición a vehículos eléctricos e híbridos no depende solo de características técnicas, sino también de infraestructura, economía y apertura de los usuarios hacia la tecnología.

4.3. Adopción de tecnologías limpias en Colombia

4.3.1. Ley 1964 de 2019

Fue promulgada el 11 de julio de 2019 para promover el uso de vehículos eléctricos y de cero emisiones en Colombia, alineándose al objetivo de fomentar la movilidad

sostenible y disminuir las emisiones contaminantes. Dispone medidas como el límite al impuesto automotor para este tipo de vehículos que no podrá superar el 1 % de su valor comercial, la concesión de descuentos en la revisión técnico-mecánica y el SOAT (Seguro Obligatorio de Accidentes de Tránsito), la exclusión del pico y placa y día sin carro, y la obligación de reservar al menos el 2 % de parqueaderos para estos vehículos. Además, regula la cuota de adopción de vehículos eléctricos en transporte masivo, y exige la instalación de estaciones de carga rápida según categoría de municipio (Congreso de la República de Colombia, 2019).

4.3.2. Estrategia Nacional de Movilidad Eléctrica

Fue desplegada en 2019 con el objetivo de acelerar la transición hacia la movilidad eléctrica en Colombia, para mejorar la calidad del aire, reducir el consumo de combustibles fósiles y cumplir con el Acuerdo de París. Propone la incorporación de 600000 vehículos eléctricos para 2030 y se articula a través de cuatro objetivos específicos: fortalecer el marco regulatorio y de política, diseñar mecanismos económicos de mercado, impulsar el desarrollo tecnológico e infraestructura de carga, y articular la planificación territorial (Ministerio del Ambiente y Desarrollo Sostenible et al., 2019).

4.3.3. Panorama Reciente de la Adopción de Vehículos Eléctricos e Híbridos en Colombia

En agosto de 2024, FENALCO (Federación Nacional de Comerciantes) y ANDI (Asociación Nacional de Empresarios de Colombia) publicaron el informe de Registro de Vehículos Eléctricos e Híbridos, en el que se destaca un importante crecimiento en la adopción de estas tecnologías de movilidad en Colombia. Según datos del RUNT, en julio de 2024 los vehículos eléctricos representaron el 4,7 % del total de registros, mientras que

los híbridos alcanzaron el 21 %. Las ventas de vehículos eléctricos ese mes aumentaron un 267 % comparado a julio de 2023 (778 unidades), marcando un récord, mientras que los híbridos aumentaron un 50 % (3.467 unidades).

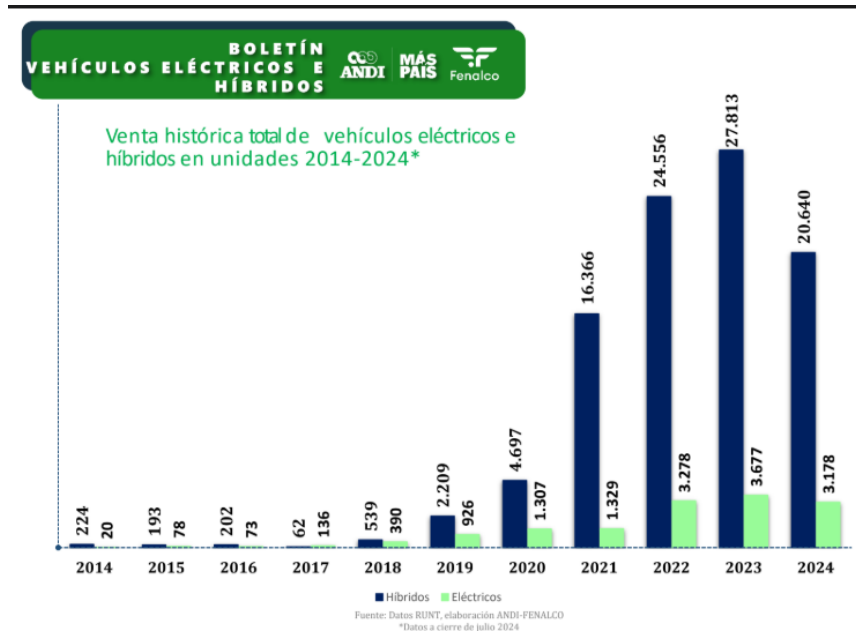
Los segmentos de mayor demanda fueron los utilitarios y automóviles. El crecimiento en registros de vehículos eléctricos se vio principalmente en ciudades como Rionegro, Cali, Chía, Bogotá y Funza, mientras que Cota, Neiva, Madrid y Bucaramanga destacaron en vehículos híbridos.

Entre enero y julio de 2024, se contabilizaron 3.178 vehículos eléctricos, lo que representa un aumento del 73,6 % en comparación con el mismo período de 2023. Asimismo, se registraron 20.640 vehículos híbridos, con un crecimiento del 41,6 %. Estas cifras demuestran la tendencia hacia una movilidad más sostenible en el país (ANDI, 2024).

Las cifras de registros de vehículos eléctricos e híbridos entre 2014 y 2024 (parcial) se pueden observar en la figura 1.

Figura 1.

Venta histórica total de vehículos eléctricos e híbridos en Colombia en unidades 2014-2024.



Nota. Tomado del Boletín Vehículos Eléctricos e Híbridos Julio 2024 (ANDI, 2024)

4.3.4. Estudios en Colombia

Se han realizado algunos estudios sobre movilidad limpia en Colombia, que han reflejado el avance y el impacto de estas tecnologías en el contexto nacional.

Cuéllar-Álvarez et al. (2024) hizo un análisis integral del ciclo de vida de los vehículos eléctricos y de combustión en Bogotá, teniendo en cuenta emisiones y costo total de propiedad para poder compararlos. Con esto, pudo evidenciar que, aunque la inversión inicial para un vehículo eléctrico puede ser mayor, en menos de tres años resultan ser más económicos que los de combustión. También, resaltó la reducción significativa de emisiones lograda con el uso de vehículos eléctricos, especialmente cuando se combinan con una matriz energética limpia.

La investigación realizada por Lopez-Arboleda et al. (2021) analiza como la adopción de vehículos eléctricos en Colombia se relaciona con diferentes factores como

infraestructura y reducción de emisiones. El análisis enfatiza en las interacciones mutuas entre estos factores y como estas no deben ser ignoradas, para evitar la implementación de políticas poco efectivas.

4.4. Introducción a la Segmentación de mercado

Tal como lo indica Kotler (2000), el mercado no es homogéneo, por lo que segmentarlo permite identificar grupos y ofrecer productos o servicios adaptados a cada uno. Este proceso ayuda a enfocar las estrategias comerciales y publicitarias, así como las políticas públicas en los segmentos que representan mayores oportunidades de satisfacción y rentabilidad.

La segmentación de mercado implica dividir un grupo de registros en varios clústeres donde los datos al interior presenten un alto grado de similitud y sean diferentes a los datos de otros clústeres (Kuo et al., 2020). Esto puede ser aplicado a productos para segmentación basada en sus características, como especificaciones de vehículos, para crear perfiles técnicos (Andrusyk & Guryanova, 2025). Este enfoque permite la identificación de potenciales nichos de mercado interesados en características específicas de un producto, que pueden ser tenidas en cuenta para planes de mercadeo y de desarrollo. Asimismo, un gobierno puede impulsar políticas que ayuden a la adopción de tecnologías con incentivos adaptados a segmentos específicos.

4.5. Criterios de Segmentación Clásicos

Los métodos de segmentación tradicionales están principalmente caracterizados por su dependencia en criterios fácilmente medibles para definir los subgrupos del mercado estudiado (Dumitrescu, 2023). De acuerdo a Drummond et al. (2008), existen

principalmente tres categorías para los criterios de segmentación: variables de perfil, variables de conducta y variables psicográficas.

Las variables de perfil comprenden información demográfica, socioeconómica y geográfica. Las variables demográficas incluyen datos como la edad, el género, el nivel de ingresos y la educación (Gaona, 2024). Las necesidades y preferencias de los compradores suelen cambiar en torno a estas variables, pero pueden ser insuficientes para una segmentación efectiva (Drummond et al., 2008).

Las variables socioeconómicas incluyen ingresos, educación y percepciones sociales, como estatus y clase social (American Psychological Association, 2025). El uso de estas variables presenta algunas dificultades debido a su naturaleza cambiante (Drummond et al., 2008).

Las variables geográficas han sido utilizadas en el pasado para segmentar por áreas como países, regiones o ciudades. El lugar de residencia de un comprador puede influir en sus decisiones (8th ICSIMAT, 2020).

La segmentación geo demográfica integra la ubicación de cada registro con datos demográficos y socioeconómicos. Puede ser útil para identificar una oferta de productos y servicios que tengan un recibimiento sobresaliente en un segmento comercial específico (Drummond et al., 2008).

Cualquiera de estos enfoques explicados anteriormente, deben ser puestos en el contexto del análisis, dado que no se pueden aplicar de la misma forma a todas las dinámicas de negocio o mercados (Drummond et al., 2008).

4.5.4. Limitaciones de los Métodos Tradicionales

Los principales retos de la segmentación tradicional son sus limitaciones inherentes en el procesamiento de grandes cantidades de datos dispares generados desde varias fuentes (Osakwe et al., 2023). Adicionalmente, tienen problemas con la naturaleza dinámica de los mercados llevando a que segmentaciones estáticas queden rápidamente obsoletas (Reddy et al., 2025).

Por su parte, los modelos de aprendizaje automático cuentan con la capacidad de procesar grandes cantidades de datos y detectar relaciones complejas y no lineales (Bogacki et al., 2024). Esta capacidad avanzada de análisis permite crear perfiles con alto nivel de detalle y dinámicos, que superan clasificaciones estáticas y dan paso a una segmentación eficaz que evoluciona con el tiempo (Bhagat et al., 2024).

4.6. Introducción a Aprendizaje Automático

Deepti Chopra & Roopal Khurana (2023) definen aprendizaje automático como una subdisciplina de la inteligencia artificial, que involucra el análisis de estructuras y patrones detectados en un conjunto de datos, que pueden ser utilizados por una máquina para aprender y tomar decisiones, permitiendo realizar funciones sin participación humana. Mediante el aprendizaje automático, se pueden utilizar grandes cantidades de datos para entrenar, probar y generar resultados automáticos.

Paparrizos et al. (2024) resalta que una de las aplicaciones más extendidas del aprendizaje automático es la segmentación de mercado, en la que los algoritmos de agrupamiento forman clústeres similares de objetos de acuerdo a varios atributos.

4.6.1. Beneficios de Usar Aprendizaje Automático para Segmentación

Desde la perspectiva de Andrusyk & Guryanova (2025), el surgimiento del aprendizaje automático ha impulsado la integración de herramientas analíticas más

sofisticadas en el análisis de mercados. Asimismo, Jauregui Romero et al. (2024) explica que los algoritmos de aprendizaje automático analizan y procesan cantidades considerables de datos a gran velocidad y con mayor precisión que las técnicas utilizadas tradicionalmente. Esto aumenta la efectividad de las estrategias aplicadas tanto a nivel público como privado. Asimismo, el uso de aprendizaje automático permite encontrar patrones y correlaciones, que no son fácilmente detectables con métodos tradicionales.

No obstante, no todos los datos o aplicaciones pueden usar el mismo algoritmo. Se deben evaluar varias opciones dependiendo de factores como las variables consideradas, el tipo de problema a resolver, el ajuste de los modelos generados, entre otros (Jauregui Romero et al., 2024).

4.6.2. Segmentación con Aprendizaje Automático

Como lo plantea Venkata Reddy Konasani & Shailendra Kadre (2021), dependiendo de la existencia de una variable objetivo en el conjunto de datos, se pueden clasificar los algoritmos de aprendizaje automático en dos grupos: los de aprendizaje supervisado, en los que se dispone de una lista de variables predictoras y una variable objetivo; y los de aprendizaje no supervisado, en los que la variable objetivo puede no estar presente en todos los casos.

Para la segmentación de una flota de vehículos eléctricos e híbridos, el análisis de Andrusyk & Guryanova (2025) señala que variables como potencia, autonomía, capacidad y tiempo de carga constituyen criterios relevantes para establecer perfiles técnicos claros.

4.6.3. Análisis de Segmentos con Aprendizaje Automático

Esta es una de las aplicaciones más comunes con aprendizaje no supervisado. En esta se utilizan un grupo de variables de entrada para detectar similitudes y dividir el

conjunto de registros en subconjuntos denominados clústeres, de tal forma que en cada clúster queden elementos similares entre sí y diferentes de los registros de otros clústeres. A nivel técnico, se busca que la distancia intraclúster sea mínima, y la distancia interclúster, máxima. Estas distancias se relacionan a la disimilitud de registros, y se pueden calcular con diferentes métricas descritas a continuación (Venkata Reddy Konasani & Shailendra Kadre, 2021).

4.6.4. Similitud Basada en Distancia

De acuerdo a Andrusyk & Guryanova (2025), las métricas de distancia son usadas para evaluar la proximidad o diferencia de los registros, lo cual es fundamental para agrupar observaciones similares en segmentos. Hay diferentes métricas, pero las más utilizadas son las siguientes.

4.6.4.1. Distancia Euclidiana.

Es una métrica ampliamente utilizada en el análisis de clústeres, particularmente para datos continuos, donde se mide la distancia directa entre dos puntos en un espacio multidimensional, representando la diferencia entre registros (France & Ghose, 2018). Se calcula con la siguiente fórmula:

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

Donde n es el número de dimensiones, x_i es la coordenada del punto A y y_i es la coordenada del punto B.

4.6.4.2. Distancia Manhattan.

Suwanda et al. (2020) describe que esta métrica es la suma de la distancia de todos los atributos. Para dos puntos x y en un espacio de m dimensiones, esta métrica se calcula con la siguiente fórmula.

$$d(x, y) = \left(\sum_{i=1}^m |x_i - y_i| \right)$$

Es la métrica más adecuada para aplicaciones en las que existen variables discretas o binarias. En caso de que los registros tengan valores numéricos con cifras decimales, se recomienda utilizar distancia Euclidiana por rendimiento (Turing, 2025).

4.6.4.3. Distancia Gower.

Tal como lo explicó Gower (1971), esta distancia permite calcular la similitud entre registros que contienen variables tanto categóricas como continuas, asignando a cada par de observaciones un valor de similitud promedio entre 0 y 1. En variables continuas, el cálculo se hace mediante la siguiente fórmula:

$$S_{ijk} = 1 - \frac{|x_i - x_j|}{R_k}$$

Donde R_k es el rango de la variable. Para variables categóricas se asigna un valor de 1 si coinciden, de lo contrario se asigna 0. Esta combinación permite el uso de esta métrica con métodos basados en distancias euclidianas multidimensionales.

4.7. Técnicas de Aprendizaje Automático Para la Segmentación No Supervisada

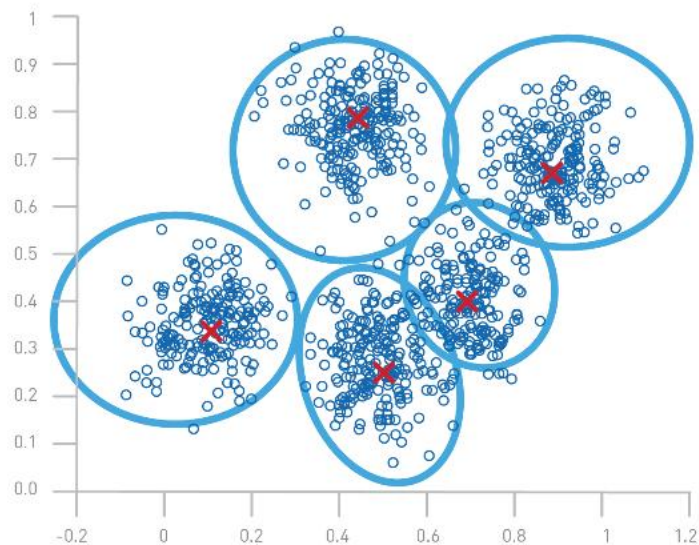
A continuación, se describen las técnicas de segmentación no supervisada que se utilizaron en este estudio.

4.7.1. K-means

Método ampliamente utilizado para agrupación de datos que funciona asignando cada registro al grupo cuyo centro sea más cercano y actualizando esos centros hasta lograr una buena segmentación como se muestra en la figura 2. Esta técnica requiere definir previamente cuántos clústeres se deben formar y, frecuentemente, probar varias opciones hasta encontrar el número óptimo (Oti et al., 2021).

Figura 2.

Clústeres formados por K-means.



Nota. Tomado de K-Means Clustering Algorithm (NVIDIA, 2025)

Sus principales ventajas son su simplicidad y eficiencia computacional para grandes grupos de datos. En contraste, se tiene que lidiar con su sensibilidad en la selección inicial de centroides y el requerimiento de especificar el número de clústeres (Ramos Salinas et al., 2023).

4.7.1.1. CatBoost con MiniBatch K-Means.

Los embeddings de Catboost son un método para codificar variables categóricas que mitiga las limitaciones del encoding tradicional generando representaciones numéricas que capturan relaciones intrínsecas en los datos (Geiler et al., 2022).

Por su parte MiniBatch K-Means es una variante más eficiente de K-Means al trabajar con volúmenes grandes de datos. Mientras K-Means necesita iterar sobre el conjunto de datos completo, MiniBatch K-Means los divide en lotes aleatorios pequeños que se pueden mantener en memoria y procesarse por iteraciones actualizando los clústeres. Gracias a este enfoque se logra reducir significativamente el tiempo de procesamiento en la segmentación (Kamal et al., 2025). En cada iteración se toma un nuevo lote con el que se ajustan los segmentos aplicando una tasa de aprendizaje descendiente conforme al aumento de las iteraciones (Béjar, 2013).

Utilizando las dos técnicas en un pipeline se pueden segmentar grupos de datos con variables numéricas y categóricas. Este enfoque combina las fortalezas de MiniBatch K-Means para segmentación eficiente de grandes grupos de datos y el manejo robusto de CatBoost de variables categóricas, ofreciendo exactitud mejorada y velocidad computacional (Dumitrache et al., 2020).

4.7.2. K-prototypes

Es una extensión de K-Means para trabajar con datos mixtos, es decir numéricos y categóricos. Es ampliamente utilizado debido a sus ventajas como la escalabilidad, simplicidad y convergencia (Ben Ncir et al., 2022). K-prototype combina dos tipos de distancias: euclidiana para los números y otra para las categorías, de forma que cuando los registros tienen categorías distintas, la diferencia entre ellos se hace mayor, mejorando así la clasificación en grupos más homogéneos (Akay & Yüksel, 2018).

Sin embargo, como explican Balaji & Lavanya (2018) este algoritmo es no incremental, porque necesita cargar todos los datos en memoria desde el inicio para su procesamiento. Por tanto, exige mucha memoria RAM, lo cual puede ser limitante al trabajar con conjuntos grandes de datos.

4.7.3. Agrupamiento Jerárquico

Tan et al. (2014) indica que las técnicas de agrupamiento jerárquico tienen dos enfoques básicos:

Aglomerativo, que comienza con cada registro como un clúster individual y, en cada etapa, va combinando los dos grupos más cercanos. Para esto, es necesario definir un criterio de cercanía entre grupos.

Divisivo, que comienza con todos los registros en un clúster y, en cada etapa, va dividiéndolo hasta llegar a clústeres de un registro.

A su vez, Tan et al. (2014) señala que las más comunes son las de enfoque aglomerativo. Gráficamente, se puede representar mediante un diagrama con forma de árbol denominado dendograma, que ilustra las relaciones entre clústeres.

Este tipo de algoritmos es utilizado principalmente en aplicaciones que tienen implícitamente una jerarquía en sus agrupaciones. Tienen como desventaja su alto costo computacional y de almacenamiento, además de su susceptibilidad a datos ruidosos y registros con muchas variables (Tan et al., 2014).

4.7.4. Supuestos estadísticos de los algoritmos utilizados

K-prototypes opera con los supuestos estadísticos de sus componentes de K-means y K-modes. Esta técnica asume que los atributos numéricos dentro de los clústeres están distribuidos normalmente con varianzas iguales, mientras que los atributos categóricos

conforman una distribución multinomial (Saligkaras & Papageorgiou, 2023). Asimismo, Catboost con minibatch K-means, hereda los supuestos ya explicados de K-means y agrega la premisa de que el submuestreo provee suficiente representación estadística del grupo de datos completo como para garantizar la convergencia a centroides significativos (Imani et al., 2025).

Por su parte, el agrupamiento jerárquico opera bajo el supuesto de que los datos pueden ser organizados significativamente en una jerarquía anidada de grupos sin asumir una forma específica de distribución dentro de cada grupo (Mussabayev & Mussabayev, 2024).

4.8. Métricas utilizadas en la segmentación de datos

Existen indicadores que evalúan los resultados de un proceso de segmentación midiendo la cohesión dentro de los clústeres y la separación entre ellos. A continuación, se describen los indicadores utilizados en este análisis.

4.8.1. Ley del codo

Es un método utilizado para encontrar el número óptimo de segmentos en los algoritmos de segmentación no supervisada. Consiste en analizar la variación de la inercia, que es la suma de las distancias al cuadrado entre los puntos y el centroide dentro de los segmentos, a medida que aumenta el número de clústeres. Cuando se grafican estos valores, se puede identificar el codo donde la reducción de la inercia es menos pronunciada, y que corresponde al número óptimo de segmentos que garantiza equilibrio entre precisión y simplicidad (Velasco Rebolledo, 2024).

4.8.2. Puntaje de silueta

Soria Olivas et al. (2023) explica que el puntaje de silueta evalúa la calidad de la segmentación midiendo el ajuste de cada elemento a su clúster. En cada segmento se calcula la cohesión, que es la distancia media entre los elementos internos, y luego, la separación, que es la distancia media al clúster más cercano. Con estos cálculos se obtiene un puntaje entre -1 y 1 , en el que se espera obtener valores cercanos a 1 , que indican grupos bien definidos y separados.

4.8.3. Índice de Davies-Bouldin

Como lo plantea Velasco Rebolledo (2024), este índice también mide la calidad de un proceso de segmentación evaluando cohesión interna y separación entre clústeres. Se enfoca en la relación entre la dispersión interna y la distancia entre los centroides de cada grupo. Para esta métrica un valor bajo (idealmente 0) indica una segmentación efectiva, con clústeres compactos y bien separados, mientras valores altos evidencian solapamiento o alta dispersión de los grupos.

4.9. Interpretabilidad de Modelos con SHAP

De acuerdo con Marcilio & Eler (2020), SHAP (SHapley Additive exPlanations) es un método de explicación que asigna a cada variable predictora un valor de contribución. Estos valores indican la importancia de cada variable en la predicción, permiten ordenar las variables por relevancia y seleccionar las más influyentes. Al sumar la contribución individual de cada variable se obtiene la predicción del modelo, siendo así que las variables menos importantes hacen variar ligeramente el valor de la predicción.

4.10. Reducción y visualización de datos con PCA-Biplot

Según Velasco Rebolledo (2024), el análisis de componentes principales (PCA) es una técnica muy utilizada para simplificar datos complejos al reducir dimensionalidad sin

descartar información importante. Transforma linealmente las variables originales para formar un sistema de componentes principales que, a su vez, son combinaciones lineales de las variables originales.

PCA ha sido utilizado extensamente por investigadores gracias a su alta precisión de clasificación y su baja complejidad de procesamiento (Saranya & Poonguzhali, 2024). Sin embargo, como señala Chisimkwuo et al. (2019) los nuevos componentes generados no ofrecen explicabilidad. Los biplots de PCA ofrecen la solución a esta limitación, mediante la visualización de la influencia de las variables originales en los componentes principales. Estas gráficas tienen vectores de cada variable original donde la magnitud indica la varianza que aporta cada variable y la cercanía entre los vectores revela correlaciones, que ayudan a seleccionar las variables más relevantes para mejorar la interpretabilidad del modelo.

4.11. Predicción de tendencias con Prophet

Prophet es un modelo de pronóstico de series temporales usado ampliamente dentro de Meta Open Source gracias a su precisión, velocidad y confiabilidad. Trabaja combinando la tendencia general con patrones repetitivos del año, la semana o día. Una de sus ventajas es que es totalmente automático y soporta datos incompletos, ruidosos o con cambios bruscos, entregando resultados en segundos sin ajustes complejos (Zherlitsyn et al., 2025). Esto se acopla al tipo de datos con los que se trabajó en este análisis.

El modelo Prophet predice un valor superior y un valor inferior que definen el intervalo de incertidumbre de la predicción. El ancho de este intervalo ofrece una medición robusta de la variabilidad inherente de la predicción (Mboko et al., 2024) y es representado

visualmente por un área sombreada en los gráficos Prophet, ofreciendo a los tomadores de decisiones una indicación clara de los riesgos asociados (Lee et al., 2025).

4.12. Desafíos de la Aplicación del Aprendizaje Automático en Segmentación de Mercado

Pitafi et al. (2023) argumenta la existencia de varios desafíos que se comentan a continuación:

Determinar el número de clústeres a priori es una de las principales dificultades para varios algoritmos.

Varios métodos de agrupamiento son muy costosos a nivel computacional con conjuntos de datos a gran escala y con muchas variables. Esto puede requerir el uso de una GPU (Unidad de Procesamiento Gráfico) de alta capacidad.

La utilización de métricas de similitud inadecuadas puede resultar en agrupamientos incorrectos. Se puede requerir de un refinamiento de los resultados con otras técnicas adicionales.

Registros con datos incompletos o faltantes pueden empeorar el rendimiento de los algoritmos de agrupamiento. La posibilidad de que existan estos registros se aumenta cuando hay un proceso de recopilación de una gran variedad de fuentes.

Los algoritmos utilizados deben ser lo suficientemente robustos para manejar el flujo de datos de su aplicación. Esto debe ir apoyado por capacidades de cómputo adecuadas y procesos de optimización de memoria.

Los grandes conjuntos de datos, del orden de Terabytes o Petabytes, también pueden representar un desafío al momento de implementar el agrupamiento con aprendizaje automático (Pitafi et al., 2023).

4.13. Trabajos relacionados

El estudio de la adopción de tecnologías de movilidad limpia utilizando herramientas de aprendizaje automático, se ha analizado con diversos enfoques, mediante diferentes algoritmos y acotados a diferentes ciudades o países.

El estudio presentado por Liu et al. (2017) utiliza la información tomada de la Oficina de Vehículos Motorizados de Ohio, sobre características de los vehículos; y de la Oficina del Censo de los Estados Unidos, sobre características socioeconómicas, dividiendo el análisis en 2943 sectores censales. Con esto crean una matriz de contigüidad y analizan el efecto de cada sector en la adopción de vehículos eléctricos e híbridos de los otros sectores.

La revisión de Bas et al. (2021), busca identificar los factores clave en la adopción de vehículos eléctricos utilizando diversas técnicas de segmentación supervisada como Máquinas de Vectores de Soporte (SVM), Redes Neuronales Artificiales (ANN), Redes Neuronales Profundas (DNN), Impulso por Gradiente (GB), Bosques Aleatorios Distribuidos (DRF) y Bosques Extremadamente Aleatorizados (ERF) entrenadas para clasificar si una persona es un comprador potencial de un vehículo eléctrico en función de una variedad de factores como características socioeconómicas, información sobre relaciones sociales, propiedad de automóvil, información sobre viajes y actitudes hacia la tecnología y el medio ambiente.

Choi & Jiao (2024) presentan los resultados de segmentación no supervisada mediante K-means con información del uso del suelo, terreno, factores sociodemográficos, medidas del entorno construido, proyección energética y factores relacionados con los

vehículos eléctricos. Esto ayudó a descubrir disparidades espaciales en la adopción de vehículos eléctricos.

En la investigación de Hardaway et al. (2025), se implementaron modelos de regresión incluyendo 39 variables sociodemográficas, sobre tecnología de vehículo y entorno construido. Dentro de las técnicas utilizadas están modelos con Modelo Lineal Generalizado (GLM), Modelo Lineal Generalizado transformado, regresión Ridge, regresión Lasso, Modelo Aditivo Generalizado (GAM) saturado, Modelo Aditivo Generalizado (GAM) reducido, árboles de regresión y bosques aleatorios.

En el artículo de Kang (2025), se utilizó *Big Data* geoespacial recolectada de los smartphones, cruzada con datos de entorno construido y aplicando Árbol de Decisión con Impulso por Gradiente (GBDT).

El análisis desarrollado en Jiang et al. (2024) utilizó un Modelo de Ecuaciones Estructurales (SEM) para encontrar las variables más decisivas en la adopción de vehículos eléctricos dentro de un grupo de variables de entorno construido, infraestructura, sociodemográficas y sobre comportamientos de viaje.

En su trabajo, Ramadoss et al. (2025) identificó que el impulso que se le ha dado a la adopción total de vehículos eléctricos en EE.UU. es notable, pero se presentan tasas menores en comunidades con inequidades sociales y geográficas. El estudio aplica segmentación a los usuarios de vehículos eléctricos y modela su adopción mediante el modelo de difusión de Bass, con el objetivo de fomentar una adopción más equitativa hacia las tecnologías de movilidad sostenible.

La propuesta metodológica de Naseri et al. (2024) examinó qué factores influyen más en la intención de compra de vehículos eléctricos en Canadá. Utilizó una encuesta

con 1077 participantes y 83 variables de elección discreta. Se aplicaron modelos de aprendizaje automático junto con SHAP para identificar las variables más influyentes, y PDP para entender su efecto sobre la intención de compra. Los resultados mostraron que las emisiones, factores financieros y psicológicos son claves para orientar políticas efectivas de adopción de opciones de movilidad eléctrica.

La investigación de Kumar et al. (2024) utilizó el método DEMATEL (*Decision Making Trial and Evaluation Laboratory*) para identificar y ponderar los factores más influyentes en las ventas de vehículos eléctricos y sus relaciones intrínsecas causa-efecto. Estos pesos son utilizados con modelos como ARIMA (Media móvil integrada autorregresiva), XGBoost, LSTM (Memoria a Largo Corto Plazo) y *Random Forest*. Esta metodología es útil para prever tendencias de ventas y apoyar decisiones en el mercado de vehículos eléctricos.

En el contexto de Colombia, Campo & Pino (2024) analizaron el crecimiento de la flota de vehículos eléctricos e híbridos, y proyectaron las ventas hasta 2030 usando modelos ARIMA y Prophet aplicados a nivel de marcas y regiones. Se utilizó la información almacenada en la plataforma de Datos Abiertos Colombia para el año 2022. Los resultados predictivos mostraron una tendencia creciente hacia 2030, con impulso mayor en regiones como Bogotá.

5. Hipótesis

5.1. Hipótesis Sobre Brechas Socioeconómicas y Territoriales

La revisión de la literatura internacional evidencia que la adopción de vehículos eléctricos no se distribuye de manera uniforme en el territorio (Choi & Jiao, 2024). Diversos estudios han señalado que las barreras económicas, derivadas principalmente del alto precio de las baterías y la tecnología, limitan el acceso inicial a estos vehículos, haciéndolos más costosos que los de combustión interna (Shareef et al., 2016). Investigaciones recientes en Estados Unidos han identificado disparidades espaciales significativas, donde la adopción se correlaciona positivamente con comunidades que poseen mejores condiciones socioeconómicas e infraestructura, dejando rezagadas a zonas con inequidades sociales y geográficas (Ramadoss et al., 2025).

En el contexto colombiano, donde existen marcadas diferencias de desarrollo e infraestructura entre el centro del país y las regiones periféricas, es esperable que estos patrones globales de desigualdad tecnológica se repliquen o acentúen. Investigaciones locales enfatizan que la interacción entre factores como la infraestructura y el mercado es crítica, sugiriendo que ignorar estas relaciones puede llevar a políticas poco efectivas (Lopez-Arboleda et al., 2021). Si la infraestructura de carga y el poder adquisitivo son determinantes, la penetración de la movilidad eléctrica debería mostrar una fuerte centralización.

Con base en lo anterior, se plantea la siguiente hipótesis:

H1: La adopción de vehículos eléctricos e híbridos en Colombia se concentra en las regiones con mayor poder adquisitivo, lo que evidencia una brecha socioeconómica en el acceso a tecnologías de movilidad sostenible.

5.2. Hipótesis Sobre Segmentación Algorítmica y Caracterización Técnica

El análisis de la adopción de nuevas tecnologías, como los vehículos eléctricos e híbridos, genera datos que integran variables de diversa naturaleza (categóricas y numéricas). En este contexto, los métodos de aprendizaje automático presentan una oportunidad significativa para profundizar en el entendimiento del mercado, ya que sus algoritmos están diseñados para analizar estructuras y detectar patrones intrínsecos en los datos que permiten generar nuevo conocimiento (Deepti Chopra & Roopal Khurana, 2023). Específicamente, la aplicación de técnicas de segmentación no supervisada ofrece la ventaja de procesar datos para descubrir correlaciones y perfiles de adopción que no son evidentes en una primera exploración (Jauregui Romero et al., 2024).

La oportunidad clave de estos métodos radica en su capacidad para gestionar la multidimensionalidad de la flota vehicular. Al utilizar métricas de similitud avanzadas, los algoritmos pueden agrupar registros maximizando la cohesión interna de cada grupo y su diferenciación respecto a los demás (Venkata Reddy Konasani & Shailendra Kadre, 2021). Esto permite trascender las clasificaciones generales y avanzar hacia una segmentación donde los vehículos se organizan por afinidades técnicas. De esta forma, el aprendizaje automático se convierte en una herramienta para construir una caracterización detallada y evolutiva que se ajuste a la realidad cambiante de la transición energética.

A partir de este potencial metodológico para revelar cómo se estructura la flota, se formula la segunda hipótesis:

H2: El uso de técnicas de segmentación con aprendizaje automático permite identificar grupos de vehículos con atributos técnicos específicos, facilitando una caracterización tipológica de la flota eléctrica e híbrida en Colombia.

6. Variables

En la tabla 1, se presentan las variables utilizadas en este estudio con su respectiva descripción, fuente y tipo.

Tabla 1.

Variables utilizadas en la presente monografía.

Variable	Descripción	Fuente	Tipo
Fuente de energía	Tipo de recurso energético que impulsa el vehículo.	Registro Único Nacional de Tránsito	Categórica nominal.
Modelo	Año de fabricación del vehículo.	Registro Único Nacional de Tránsito	Categórica ordinal.
Año de registro	Año en que el vehículo fue oficialmente registrado.	Registro Único Nacional de Tránsito	Categórica ordinal.
Clase	Subcategoría más específica del vehículo.	Registro Único Nacional de Tránsito	Categórica nominal.
Servicio	Tipo de uso autorizado del vehículo.	Registro Único Nacional de Tránsito	Categórica nominal.
Marca	Fabricante del vehículo.	Registro Único Nacional de Tránsito	Categórica nominal.
Línea	Nombre comercial del modelo dentro de la marca.	Registro Único Nacional de Tránsito	Categórica nominal.
Municipio	División territorial y administrativa de primer nivel en Colombia.	Registro Único Nacional de Tránsito	Categórica nominal.
Código Municipio	Código DANE para el municipio.	DANE	Categórica nominal.

Departamento	División territorial y administrativa de segundo nivel en Colombia.	Registro Único Nacional de Tránsito	Categórica nominal.
Código Departamento	Código DANE para el departamento.	DANE	Categórica nominal.
Cilindraje	Volumen total de los cilindros de un motor (Alcaldía de Popayán, n.d.). Expresado en centímetros cúbicos (cc).	Portales web de cada marca.	Numérica discreta.
Capacidad de carga	Peso máximo permitido que un vehículo puede transportar (Alcaldía de Popayán, n.d.) . Expresado en Kg.	Portales web de cada marca.	Numérica discreta.
Capacidad de pasajeros	Número máximo de personas que pueden viajar en un vehículo (Alcaldía de Popayán, n.d.).	Portales web de cada marca.	Numérica discreta.
Potencia	Cantidad de energía mecánica que el motor es capaz de generar en un período determinado. Influye directamente en la aceleración del vehículo, su capacidad para remolcar peso y su rendimiento en pendientes (Renault Group, 2021).	Portales web de cada marca.	Numérica continua.
PIB per cápita departamental	Resultado final de la actividad productiva dividida entre la población total (DANE, n.d.) .	DANE	Numérica discreta.
Número de matrículas de educación superior por departamento	Número de matriculados para formación técnica, tecnológica, pregrado y posgrado a nivel departamental.	Ministerio de Educación de Colombia	Numérica discreta.

Índice de Brecha Digital (IBD) departamental	Desigualdades en el acceso y uso de las TIC (Tecnologías de la Información y la Comunicación) entre los habitantes de cada departamento (MINTIC Colombia, 2023). Porcentaje de la fuerza de trabajo que no logra encontrar empleo (Banco de la República de Colombia, n.d.).	Ministerio TIC	Numérica continua.
Tasa de desocupación	Número total de personas que viven en el departamento.	DANE	Numérica continua.
Población total departamental	Número de personas que residen dentro de la zona urbana de la unidad espacial de referencia (Cifuentes, n.d.).	DANE	Numérica discreta.
Población urbana departamental	Cantidad de puntos de carga para vehículos eléctricos e híbridos por departamento.	DANE	Numérica discreta.
Número de puntos de carga por departamento		Aplicaciones con mapas de puntos de carga	Numérica discreta.

Nota. Elaboración propia.

7. Metodología

7.1. Enfoque y alcance de la investigación

Esta monografía tiene un enfoque cuantitativo, basado en datos estructurados se realizaron análisis y segmentación mediante aprendizaje automático de los vehículos eléctricos e híbridos registrados en Colombia. Para ello, se integraron a los análisis variables sociodemográficas, regionales, económicas y de infraestructura con el fin de conocer cómo varía la adopción en cada departamento y cómo esta se relaciona con dichas variables contextuales. Adicionalmente, se identificaron cuáles son los factores que más influencia tienen en la adopción mediante la variable objetivo, que es el número de vehículos eléctricos e híbridos registrados por año, lo que aporta información clave para adaptar estrategias comerciales diferenciadas y para el diseño de políticas públicas focalizadas que promuevan una adopción más efectiva y equitativa en el territorio.

Para la segmentación, se agruparon los datos en categorías específicas: motocicletas, vehículos ligeros de uso particular y oficial, vehículos ligeros de uso público, vehículos pesados de uso público y vehículos de carga. Esta segmentación permitió identificar patrones de adopción diferenciados según características técnicas como potencia, cilindraje, capacidad de carga y capacidad de pasajeros.

El tipo de investigación es descriptivo-correlacional, al buscar caracterizar los segmentos identificados y explorar la relación entre los patrones de adopción y diversas condiciones socioeconómicas y territoriales.

Este enfoque permitió generar una visión integral y basada en evidencia del proceso de adopción de movilidad sostenible en Colombia, aportando insumos valiosos para los responsables de tomar decisiones en los sectores público y privado.

7.2. Población y muestra

La población de este estudio está conformada por los registros de vehículos eléctricos e híbridos en Colombia disponibles hasta 2022, complementados con información socioeconómica, geográfica y demográfica correspondiente al mismo periodo para garantizar consistencia temporal. Estos registros abarcan vehículos de uso particular, oficial y público en los distintos departamentos del país.

Se utilizó exclusivamente la información disponible para consulta pública de fuentes oficiales como el Registro Único Nacional de Tránsito (RUNT), Ministerio de Educación, el Departamento Administrativo Nacional de Estadística (DANE) y otras entidades. Estos datos constituyen una buena base accesible para análisis y diagnósticos sobre tendencias históricas y patrones de adopción en el proceso de transición energética.

En línea con esto, los registros debieron cumplir con criterios de integridad y consistencia de datos, con variables que aporten al análisis como características técnicas de los vehículos eléctricos e híbridos registrados, año y lugar de registros y factores socioeconómicos y de infraestructura de la región analizada.

7.3. Técnicas para el análisis de la información

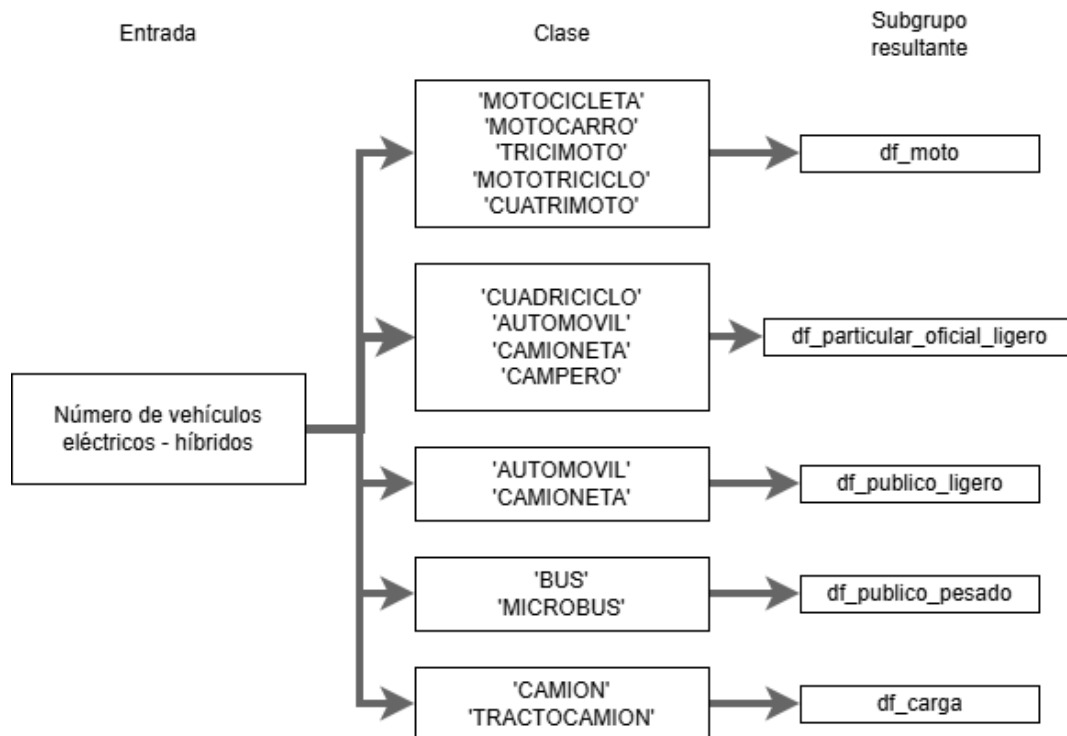
Para iniciar se recopiló la información de las variables utilizadas en el estudio, a partir de fuentes confiables como el RUNT, portales web de las marcas, aplicaciones de puntos de carga, Ministerio TIC, Ministerio de Educación y DANE. Los datos se descargaron de las fuentes oficiales o se construyeron de acuerdo a disponibilidad, y se adaptaron a archivos en formato CSV (*Comma Separated Values*) que pueden ser importados fácilmente al proyecto.

Teniendo en cuenta el enfoque en identificación de patrones de adopción de vehículos eléctricos e híbridos en Colombia a partir de datos estructurados, se emplearon técnicas de Análisis Exploratorio de Datos (EDA), que iniciaron con un proceso de limpieza y preprocesamiento que incluyó tratamiento de registros con valores faltantes para consolidar un grupo de datos de los vehículos registrados, que contenga la información técnica contrastada con datos publicados por los fabricantes.

Los datos se dividieron en subgrupos con categorías específicas: motocicletas, vehículos ligeros de uso particular y oficial, vehículos ligeros de uso público, vehículos pesados de uso público y vehículos de carga, según su clase registrada, como se muestra en la figura 3.

Figura 3.

División en subgrupos de acuerdo a clase de vehículo.



Nota. Elaboración propia.

Para la segmentación no supervisada se seleccionaron tres técnicas: K-prototypes, por estar diseñada para trabajar con variables numéricas y categóricas simultáneamente, sin necesidad de usar variables dummy; CatBoost embedding con minibatch K-means, por su codificación de variables categóricas que captura relaciones intrínsecas entre los datos, además de su eficiencia para trabajar con grandes volúmenes de información; y agrupamiento jerárquico, por tener un enfoque diferente al utilizado en las otras técnicas, lo que permite contrastar resultados y añadir robustez al análisis.

En la metodología propuesta, se aplicó el algoritmo de segmentación no supervisada K-prototypes a cada uno de los subgrupos de datos, y de forma complementaria se emplearon agrupamiento jerárquico y Catboost embedding con minibatch K-means, dependiendo del volumen de registros a procesar. El número óptimo de segmentos se definió mediante el método del codo, mientras que la combinación del puntaje silueta y el índice Davies Bouldin permitió seleccionar los mejores resultados entre los algoritmos utilizados evaluando simultáneamente cohesión interna y separación entre clústeres. Se optó por utilizar el índice de Davies Bouldin sobre el de Calinski Harabasz, dado que este último tiende a favorecer segmentaciones con mayor número de clústeres (Wani, 2024), lo cual no se alinea con el enfoque del presente análisis. En este estudio se prioriza una cantidad más reducida de segmentos, desde que estos tengan buena cohesión interna y caracterización técnica diferencial. Utilizando estas técnicas se logró identificar grupos difíciles de evidenciar con métodos tradicionales.

Adicionalmente, se llevó a cabo una proyección de la adopción futura por cada segmento identificado hasta el año 2028, utilizando Prophet de Meta Open Source, con el

propósito de anticipar cómo evolucionará la adopción de este tipo de vehículos en los próximos años. Para esta etapa del análisis, se seleccionaron solo los segmentos que contaban con registros de al menos cinco años, para asegurar una cantidad mínima de datos que el modelo pudiera utilizar para una proyección fiable. Luego se generó el pronóstico de adopción para el periodo estipulado anteriormente haciendo calibración de hiperparámetros basado en la métrica de MAE (Error Absoluto Medio) y se estimó la incertidumbre promedio de las predicciones calculando el promedio de los intervalos de incertidumbre correspondientes a cada año proyectado por cada segmento, de acuerdo con la siguiente fórmula:

$$Incertidumbre\ promedio = \frac{1}{6} \sum_{t=2023}^{2028} (y_t^{sup} - y_t^{inf})$$

Asimismo, se calculó la incertidumbre relativa con el objetivo de evaluar que tan confiable es cada predicción en proporción a su magnitud proyectada. Esto permite evaluar en los mismos términos la fiabilidad de la proyección para clústeres que tienen decenas o miles de unidades. Este indicador se calculó con la siguiente fórmula:

$$Incertidumbre\ relativa = \frac{Incertidumbre\ promedio}{Valor\ proyectado\ 2028}$$

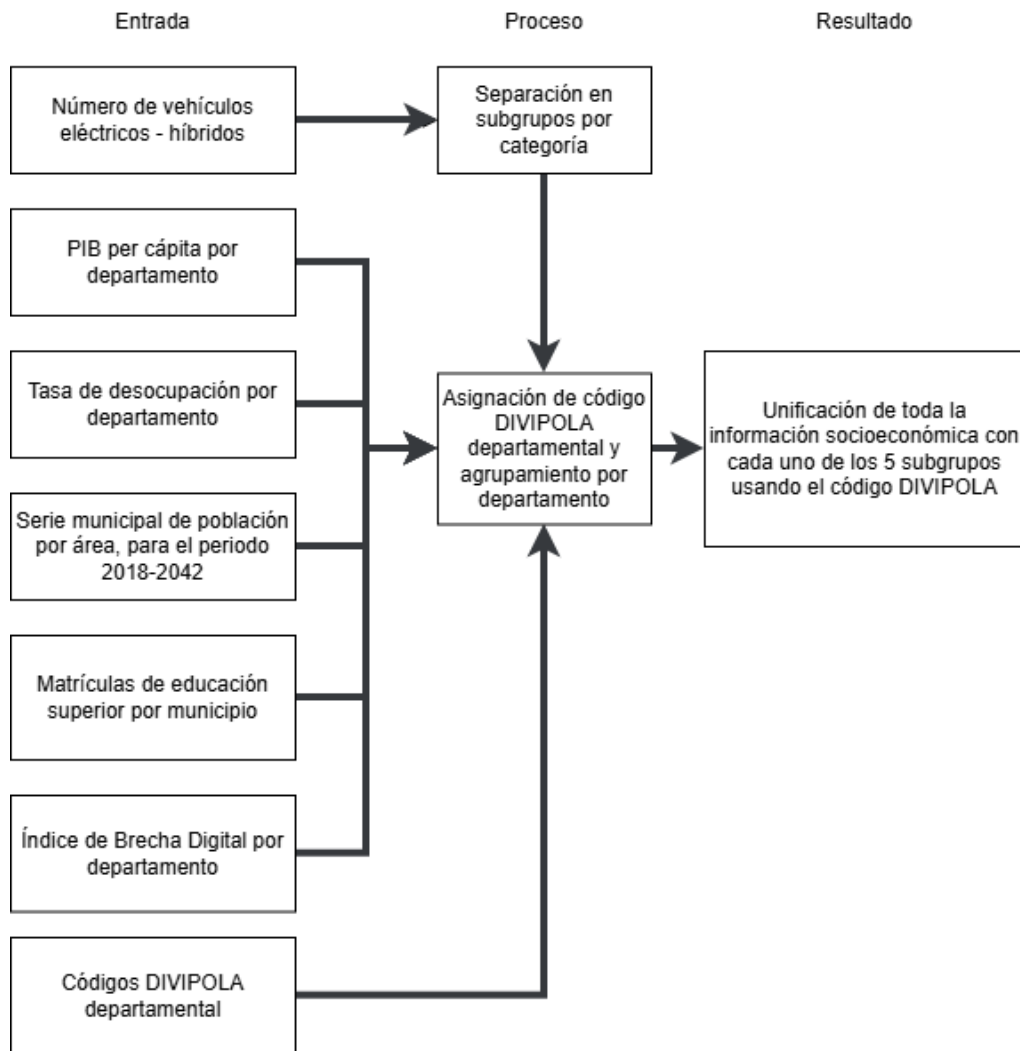
Con estos indicadores se realiza un análisis de las tendencias para cada segmento teniendo en cuenta las gráficas de Prophet, los valores proyectados y la incertidumbre calculada.

Posteriormente, se integraron los registros de vehículos eléctricos e híbridos con información sociodemográfica de cada departamento, utilizando el código DIVIPOLA

(División Político-Administrativa) para asegurar una correcta vinculación entre bases de datos como se ilustra en la figura 4.

Figura 4.

Proceso de consolidación de dataframe unificado para análisis integral.



Nota. Elaboración propia.

Con los 5 subconjuntos complementados con la información integrada se realizó un análisis de las variables socioeconómicas que más influyen en la cantidad de registros por

departamento, empleando SHAP (SHapley Additive exPlanations) junto con sus gráficas interpretativas, sobre un modelo RandomForestRegressor. Además, se utilizaron gráficas PCA biplot para complementar la interpretación y comprender de forma visual las dinámicas socioeconómicas asociadas a la adopción.

Todo el proceso descrito anteriormente se implementó en Jupyter Notebook con ambientes manejados por Conda y los siguientes paquetes estadísticos de Python 3.10.19:

- Pandas 2.3.3 y numpy 1.26.4, para manejo y cálculo de datos numéricos.
- Matplotlib 3.10.7 y seaborn 0.13.2, para graficación de datos.
- Kprototypes (kmodes 0.12.2), para segmentación de variables mixtas (numéricas y categóricas).
- StandardScaler (scikit-learn 1.7.2) y MinMaxScaler (scikit-learn 1.7.2), para normalización de variables numéricas.
- silhouette_score (scikit-learn 1.7.2) y gower 0.1.2, para medir resultados de segmentación no supervisada.
- AgglomerativeClustering (scikit-learn 1.7.2), linkage (scipy 1.15.2) y dendrogram (scipy 1.15.2); para segmentación y visualización por agrupamiento jerárquico.
- PCA (scikit-learn 1.7.2), para reducción de dimensionalidad.
- Geopandas 1.1.1, multipolygon (shapely 2.1.2) y polygon (shapely 2.1.2); para análisis y geometría de datos geoespaciales.
- CatBoostEncoder (category-encoders 2.8.1), para transformación de variables categóricas en embeddings numéricos.
- MiniBatchKMeans (scikit-learn 1.7.2), para segmentación de grandes volúmenes de datos.

- RandomForestRegressor (scikit-learn 1.7.2) y train_test_split (scikit-learn 1.7.2), para modelado y partición de predicción supervisada.
- Shap 0.49.1, para explicación de la importancia de las variables en un modelo.
- Prophet 1.1.7, para pronóstico de series temporales.

Este análisis comprende la exploración de datos, segmentación no supervisada, graficación geoespacial, explicabilidad de modelos y proyección de registros, para ofrecer una visión general del panorama de adopción de la movilidad sostenible en Colombia. Puede ser útil para el sector público y privado en la creación de estrategias comerciales enfocadas en cada segmento, en el fortalecimiento de políticas e incentivos de cada región y en las decisiones sobre inversiones en infraestructura.

8. Trabajo de Campo

8.1. Procesamiento de Datos

El proceso inició con la recolección de datos provenientes de fuentes oficiales. El conjunto principal corresponde a la base “Número de Vehículos Eléctricos - Híbridos” disponible en la plataforma de Datos Abiertos Colombia ([https://www.datos.gov.co/Transporte/Numero-de-Veh-culos-El-ctricos-Hibridos/7qfh-
tkr3](https://www.datos.gov.co/Transporte/Numero-de-Veh-culos-El-ctricos-Hibridos/7qfh-
tkr3)), asociada al Ministerio de Transporte, que reúne los registros de este tipo de vehículos desde 2010 hasta 2022.

Complementariamente, se integró información del año 2022 sobre el PIB per cápita por departamento (<https://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales/cuentas-nacionales-trimestrales/historicos-producto-interno-bruto-pib>), la tasa de desocupación (<https://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral/empleo-y-desempleo/geih-historicos>); https://www.dane.gov.co/files/investigaciones/boletines/ech/ech_regiones/bol_regiones_1_s_22.pdf), y la población total y urbana por departamento, tomada de la “Serie municipal de población por área, para el periodo 2018-2042” (<https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion>). También se incorporaron el número de matrículas en educación superior (año 2020) disponibles en Datos Abiertos Colombia (https://www.datos.gov.co/Educaci-n/MEN_MATRICULA_ESTADISTICA_ES/5wck-szir/about_data) y el Índice de Brecha Digital 2023 (https://colombiatic.mintic.gov.co/679/articles-396961_recurso_1.pdf). Asimismo, el

número de puntos de carga se recolectó manualmente a partir de mapas alimentados por la comunidad en varias aplicaciones.

Posteriormente, se realizó un análisis exploratorio de la base “Número de Vehículos Eléctricos - Híbridos”, identificando inconsistencias y vacíos significativos en las columnas de información técnica. Con el objetivo de asegurar la integridad y exactitud de los datos técnicos, garantizando que corresponden a las especificaciones reales y que todos los registros cuentan con la información relevante completa, se extrajeron todas las combinaciones únicas de marca y línea, para luego consultar de manera manual y sistemática los sitios web oficiales de cada fabricante. Con esta información se construyó un diccionario técnico que incluye atributos como cilindraje, potencia, capacidad de pasajeros y capacidad de carga.

La construcción de este diccionario se realizó según la naturaleza de cada vehículo: para camiones y tractocamiones se registró únicamente la capacidad de carga, dejando vacía la capacidad de pasajeros; mientras que para los otros vehículos se completó la capacidad de pasajeros, omitiendo la capacidad de carga porque no es relevante. En el caso de vehículos completamente eléctricos, no se incluyó el cilindraje, dado que esta característica no aplica.

El diccionario técnico fue posteriormente subido a un repositorio en GitHub (<https://github.com/AlfaroDev/VEH-adopcion-Col/tree/main>) e importado al proyecto en Python para alimentar la base de datos principal. Para aquellas referencias sin información disponible en línea, especialmente motocicletas eléctricas, se aplicó un proceso de imputación: se calculó el promedio del valor técnico correspondiente entre otras referencias de la misma marca y clase; y cuando estos datos tampoco estaban disponibles, se utilizó

el promedio general de la misma clase de vehículo. Adicionalmente, se realizó una corrección manual de la clase en algunos registros, con el fin de garantizar la coherencia de la información.

Con estas validaciones y depuraciones, se consolidó una base final compuesta por 56545 registros descritos en la tabla 2. Las variables con el sufijo _R corresponden a los campos que fueron rectificadas utilizando el diccionario técnico previamente construido.

Tabla 2.

Conteo de registros para base final consolidada.

Columna	Conteo de no nulos	Tipo
Combustible	56545	object
Año_Registro	56545	int64
Clase	56545	object
Servicio	56545	object
Marca	56545	object
Departamento	56545	object
Cilindraje_R	56545	float
Capacidad_Carga_R	56545	float
Capacidad_Pasajeros_R	56545	float
Potencia_R	56545	float

Nota. Elaboración propia.

8.2. Análisis Exploratorio de Datos a Nivel Nacional

Con esta información, se realizó una exploración inicial de los datos mediante visualizaciones estadísticas, con el objetivo de identificar patrones clave en la adopción de

vehículos eléctricos e híbridos en Colombia. Los gráficos permitieron analizar aspectos como:

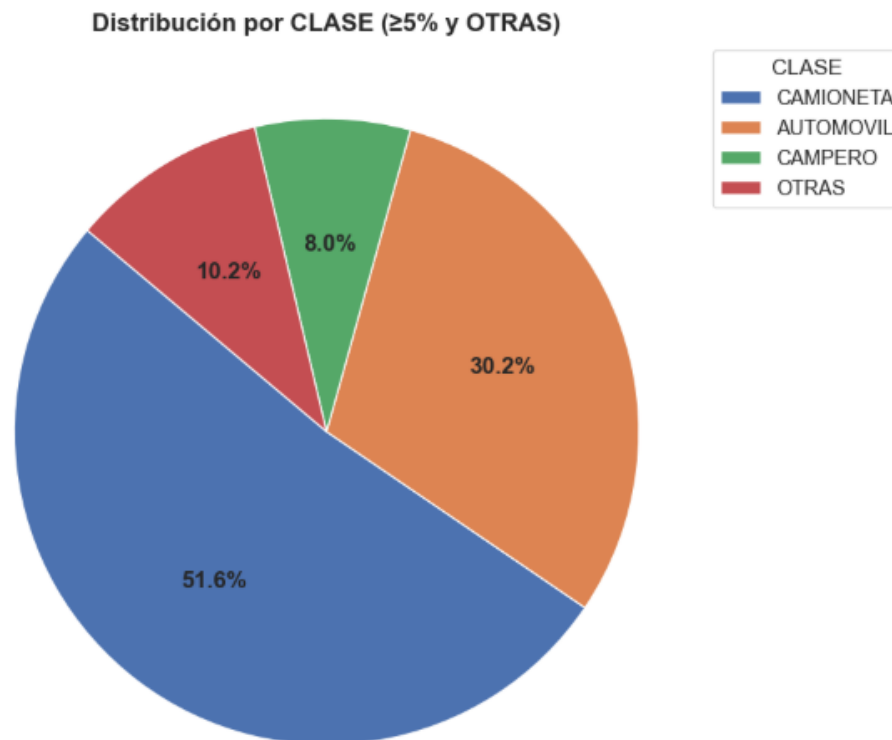
- Distribución de la adopción por clase de vehículo, así como su porcentaje dentro del parque automotor eléctrico e híbrido.
- Tendencia temporal de registros según tipo de uso (particular, público y oficial).
- Evolución de la adopción por fuente de energía (eléctrico, híbrido gas-eléctrico e híbrido diésel-eléctrico).

Este análisis exploratorio permitió contextualizar el comportamiento del mercado y orientar las etapas posteriores de segmentación y modelado.

La figura 5 muestra que aproximadamente la mitad del parque automotor de vehículos eléctricos e híbridos está dominado por vehículos de clase camioneta (51.6%), lo que indica una marcada preferencia por este tipo de vehículos que puede tener un uso familiar o utilitario. Lo siguen los automóviles (30.2%) y camperos (8%). El 10.2% restante se divide en otras denominaciones. El mercado presenta una concentración alta en estas clases, de lo que se puede inducir que las estrategias comerciales actuales están enfocadas en este tipo de vehículos.

Figura 5.

Distribución del parque automotor eléctrico e híbrido por clase.

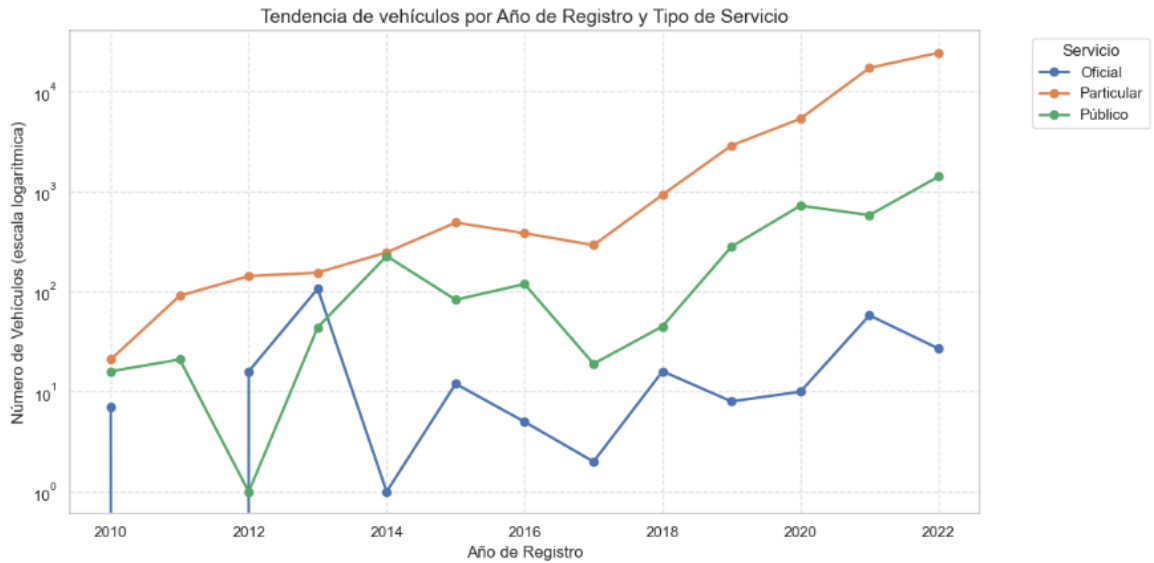


Nota. Elaboración propia.

Por su parte, la figura 6 muestra un dominio marcado del servicio particular, con un crecimiento exponencial, especialmente desde 2018. El servicio público tiene un crecimiento gradual pero sostenido, con subidas marcadas desde 2019. El servicio oficial registra muy pocas unidades, lo que hace que su relevancia en el mercado sea baja, por eso en este estudio se analizaron de forma unificada el sector particular y el oficial, aprovechando que sus características técnicas y patrones de uso son similares.

Figura 6.

Tendencia por tipo de servicio 2010-2022 en el parque automotor eléctrico e híbrido por clase.

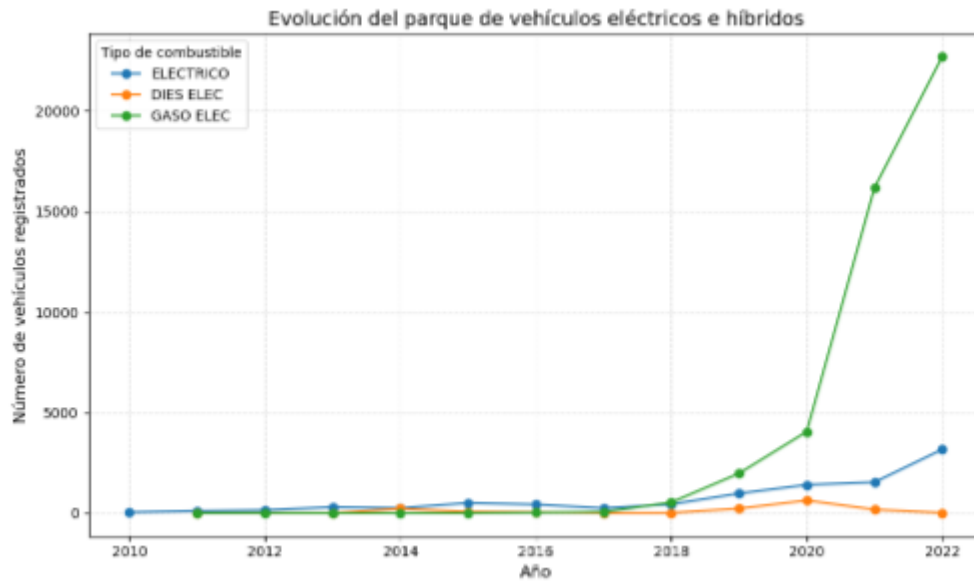


Nota. Elaboración propia.

La figura 7 muestra el crecimiento exponencial en los registros de vehículos gaso-eléctricos desde 2019, y su amplio dominio sobre otras opciones de fuente de energía sostenible. Los vehículos eléctricos también muestran crecimiento, pero mucho más gradual, mientras los diésel-eléctricos tienen presencia marginal. De aquí se puede inferir que el mercado colombiano favorece soluciones híbridas en la transición energética sobre opciones totalmente eléctricas, probablemente por costos, percepción negativa de tecnologías en surgimiento y construcción lenta de infraestructura de carga.

Figura 7.

Evolución del parque automotor eléctrico e híbrido 2010-2022.



Nota. Elaboración propia.

También se realizaron cálculos estadísticos que permitieron conocer mejor la composición general de las características técnicas de la flota.

Tabla 3.

Medidas estadísticas del grupo de datos utilizado.

Variable	Media	Mínimo	Máximo	P75	P90
Cilindraje (cc)	1620.38	0	6000	2000	2500
Potencia (KW)	135.72	0.18	735	163	235
Capacidad de pasajeros	7.03	1	160	5	5
Capacidad de carga	2468.5	280	25000	4115	4115

Nota. Elaboración propia.

En la tabla 3 se evidencia la heterogeneidad en características técnicas de los vehículos con cilindrajes desde 0 cc (100% eléctricos) hasta 6000 cc. Un rango igualmente

amplio se ve en la variable de potencia (0.18 a 735 kW), lo que sugiere la presencia de segmentos especializados en ambos extremos. Respecto a la capacidad de pasajeros la media es de 7, pero conociendo el valor máximo y los percentiles se puede intuir la existencia de transporte colectivo en la flota. Asimismo, la capacidad de carga, muestra unos valores máximos elevados, que obviamente están ligados a vehículos de carga pesada.

8.3. Formación de Subgrupos

A continuación, se procedió a segmentar la base en subgrupos específicos, con el objetivo de enfocar el análisis y permitir una identificación más precisa de patrones en cada categoría. Los subgrupos definidos fueron:

- Motocicletas (df_moto)
- Vehículos ligeros de uso particular y oficial (df_particular_oficial_ligero)
- Vehículos ligeros de uso público (df_publico_ligero)
- Vehículos pesados de uso público (df_publico_pesado)
- Vehículos de carga (df_carga)

La tabla 4 muestra cómo quedó estructurado cada uno de los subgrupos.

Tabla 4.

Conteo de registros de cada subgrupo generado.

Subgrupo	Cantidad de registros
df_moto	1862
df_particular_oficial_ligero	51167
df_publico_ligero	631

df_publico_pesado	1967
df_carga	918

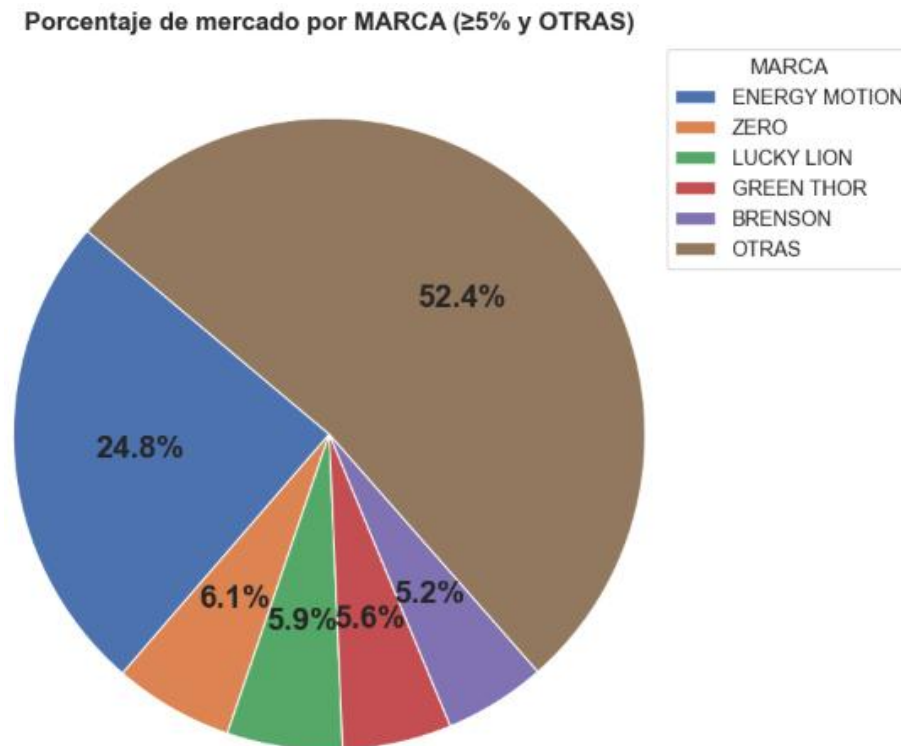
Nota. Elaboración propia.

Asimismo, se exploró cómo está dividido el mercado por marcas en cada subgrupo. Las marcas con menos del 5% de participación se agruparon como “Otras”.

En el subgrupo de motos, como se observa en la figura 8, el mercado está muy fragmentado. Aproximadamente la mitad de los registros (52.4 %) está repartido entre marcas con menos del 5% del mercado. La marca líder Energy Motion tiene el 24.8%, seguida de lejos por la marca Zero con 6.1%. Se plantea como un mercado con espacio para nuevos actores, pero con dificultad para escalar con tanta competencia.

Figura 8.

Porcentaje de mercado en subgrupo de motos.

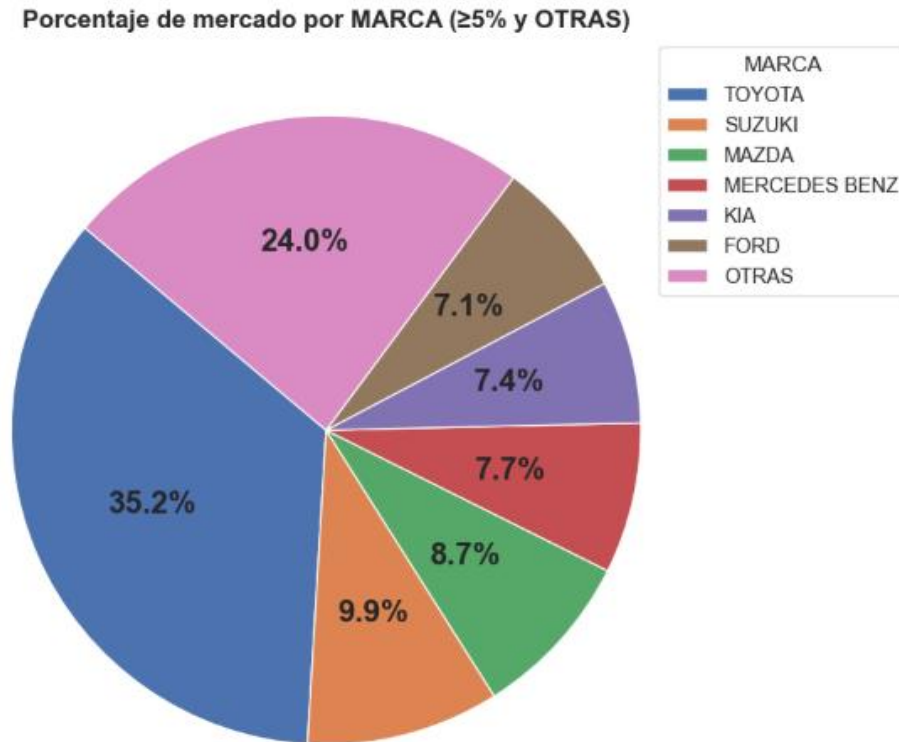


Nota. Elaboración propia.

En el subgrupo de particulares y oficiales ligeros, según se evidencia en la figura 9, se muestra el dominio de Toyota con aproximadamente un tercio del mercado (35.2 %) seguido por Suzuki (9.9 %) y Mazda (8.7 %). Destaca que hay un cuarto del mercado manejado por marcas con poca participación.

Figura 9.

Porcentaje de mercado en subgrupo de particular y oficial ligero.

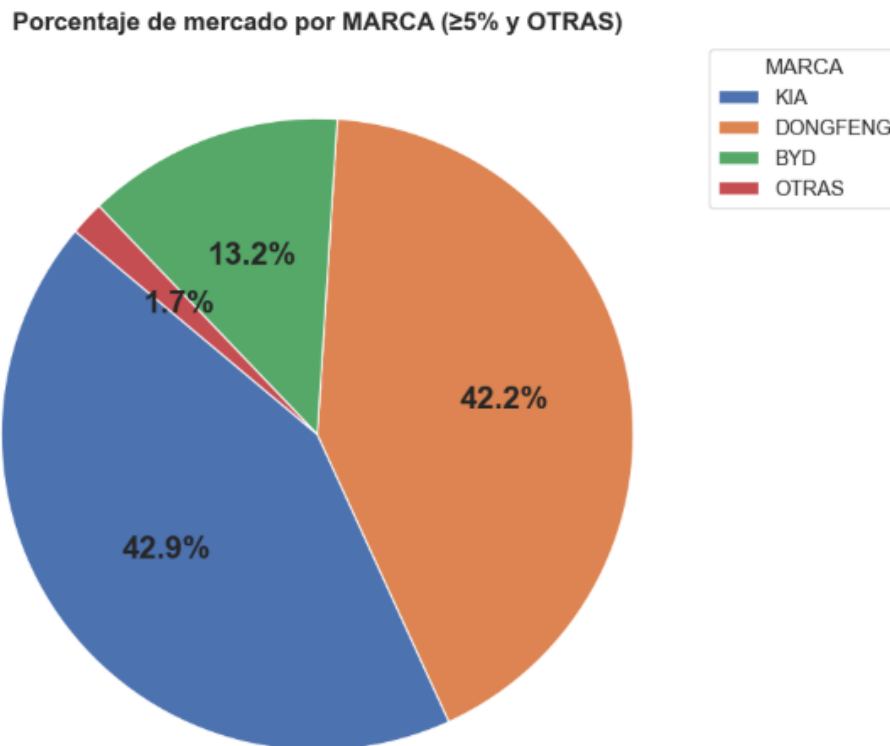


Nota. Elaboración propia.

Dentro del subgrupo de uso público ligero, como se aprecia en la figura 10, hay un duopolio entre Kia y Dongfeng, cada una con más del 40% de participación. BYD en tercer lugar puede ser un disruptor en este mercado, pero dependerá de decisiones comerciales hechas a nivel de flotillas de este sector.

Figura 10.

Porcentaje de mercado en subgrupo de público ligero.

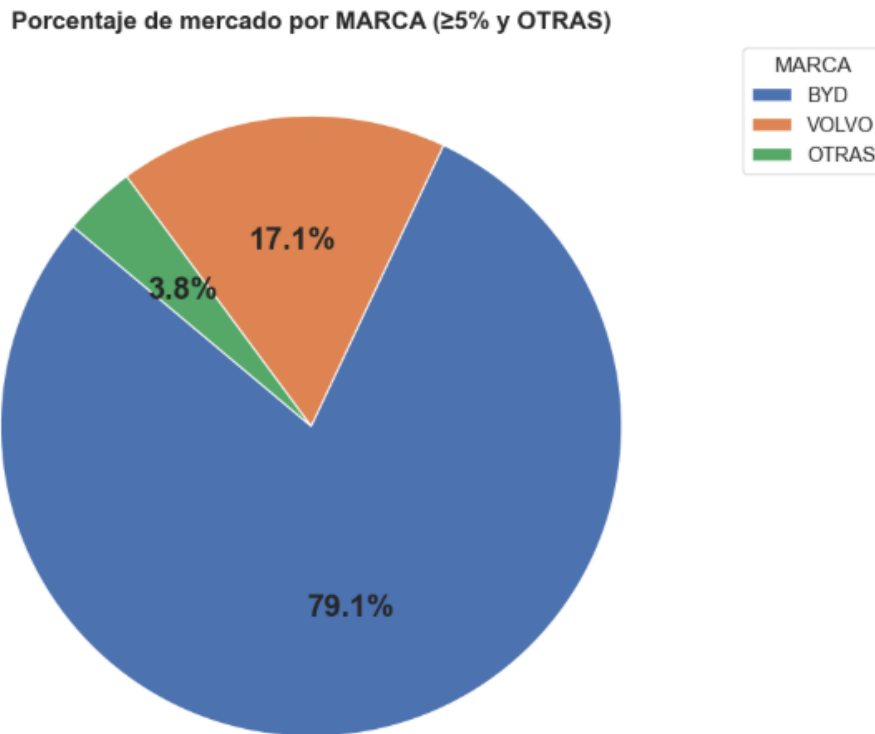


Nota. Elaboración propia.

La figura 11 del subgrupo de uso público pesado, evidencia que BYD domina ampliamente con el 79.1 %, probablemente gracias a alianzas comerciales y su liderazgo tecnológico. Se presenta como un mercado cerrado para nuevos competidores.

Figura 11.

Porcentaje de mercado en subgrupo de público pesado.



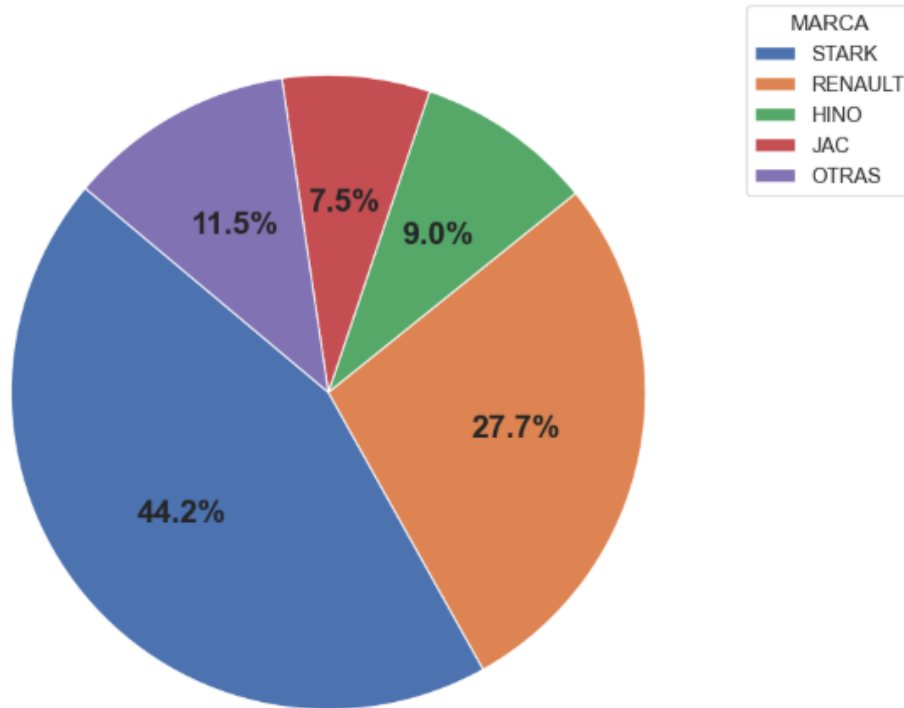
Nota. Elaboración propia.

Como se presenta en la figura 12, dentro del subgrupo de carga, la marca Stark lidera (44.2 %) seguida de cerca por Renault (27.7 %). Es un mercado en crecimiento que puede estar abierto a nuevos competidores que intenten diferenciarse en servicio, financiamiento o innovaciones tecnológicas.

Figura 12.

Porcentaje de mercado en subgrupo de carga.

Porcentaje de mercado por MARCA (≥5% y OTRAS)



Nota. Elaboración propia.

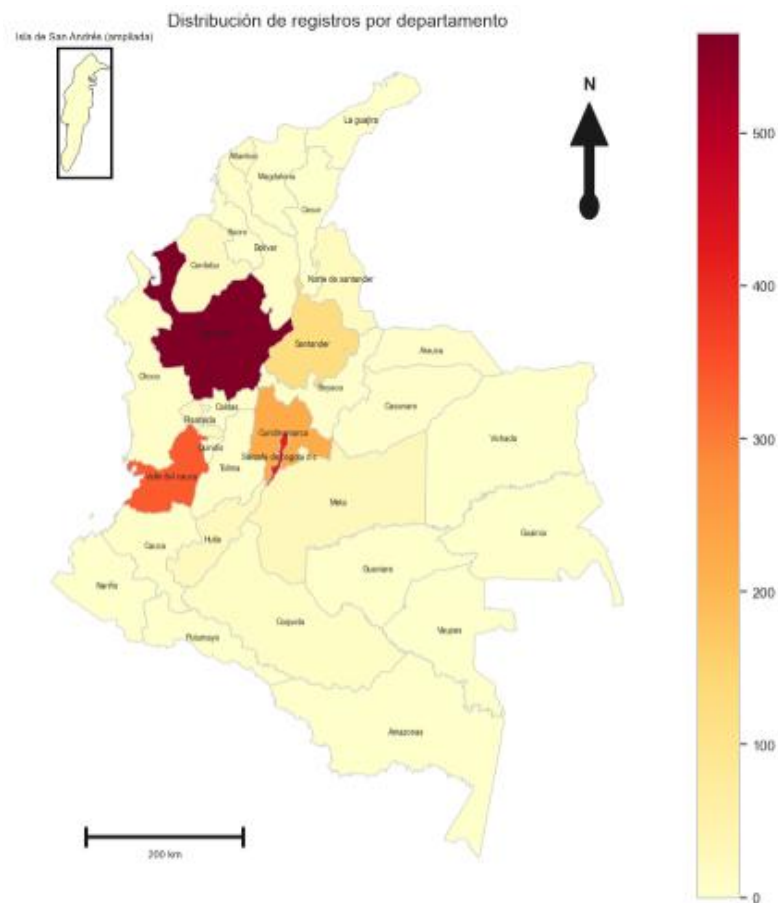
También se graficaron mapas de calor sobre la división departamental de Colombia por cada uno de los subgrupos para ver el nivel de adopción en cada región y poder identificar las zonas más avanzadas, las zonas más atrasadas y la desigualdad entre regiones.

Adicionalmente, se calculó la relación entre el total de vehículos eléctricos e híbridos registrados y la población total de cada departamento, con el objetivo de incorporar al análisis la proporcionalidad esperada entre ambas variables. Los resultados se presentan en las tablas 5 a 9 ubicadas después de cada mapa, ordenados de mayor a menor, mostrando los cinco departamentos con los valores más altos para facilitar la comparación.

En el mapa del subgrupo de motos, presentado en la figura 13, hay una concentración de registros en Bogotá, Antioquia y Valle del Cauca. Esto indica que la adopción se concentra fuertemente en los principales centros urbanos y regiones con mayor poder económico. Hay algunos departamentos como Cundinamarca y Santander, que muestran una participación significativa, pero los demás presentan una adopción muy baja, reflejando importantes brechas territoriales.

Figura 13.

Distribución departamental del subgrupo de motos.



Nota. Elaboración propia.

Tabla 5.

Registros per cápita a nivel departamental del subgrupo de motos.

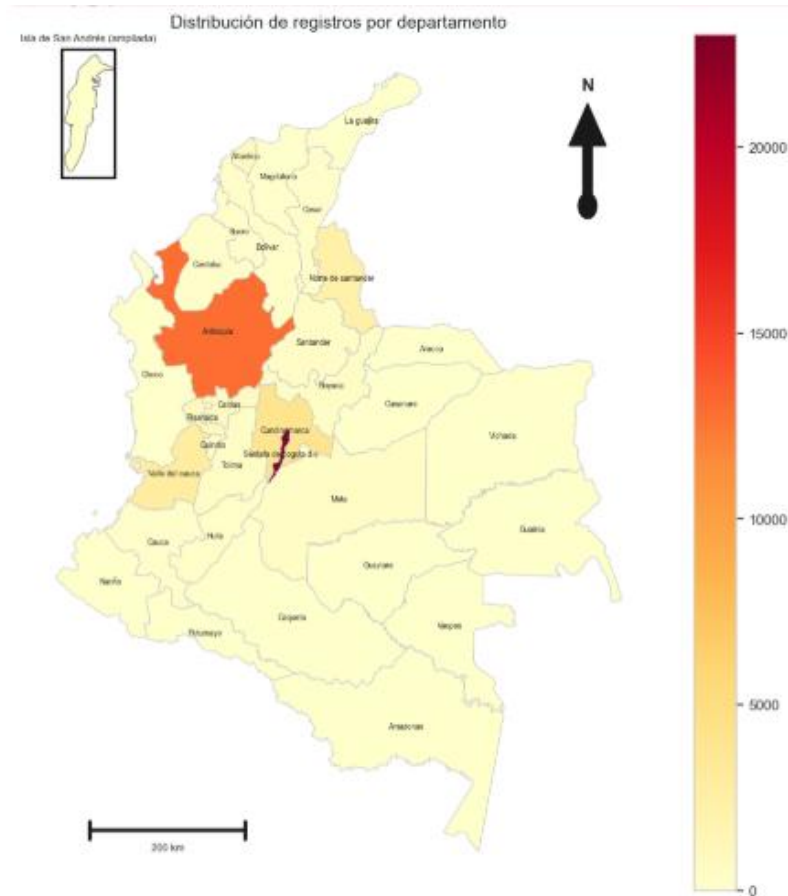
Departamento	Cantidad de registros	Población total	Registros per cápita
Antioquia	565	6777249	0,000083
Valle del Cauca	339	4658811	0,000073
Cundinamarca	226	3316583	0,000068
Santander	125	2341058	0,000053
Bogotá DC	418	7849206	0,000053

Nota. Elaboración propia.

El panorama se presenta muy similar en los mapas 14 a 17 de los otros subgrupos, con una concentración muy fuerte en los departamentos principales y en la ciudad capital, una adopción incipiente en departamentos medios y mínima en los sectores históricamente rezagados.

Figura 14.

Distribución departamental del subgrupo de particular y oficial ligero.



Nota. Elaboración propia.

Tabla 6.

Registros per cápita a nivel departamental del subgrupo de particular y oficial ligero.

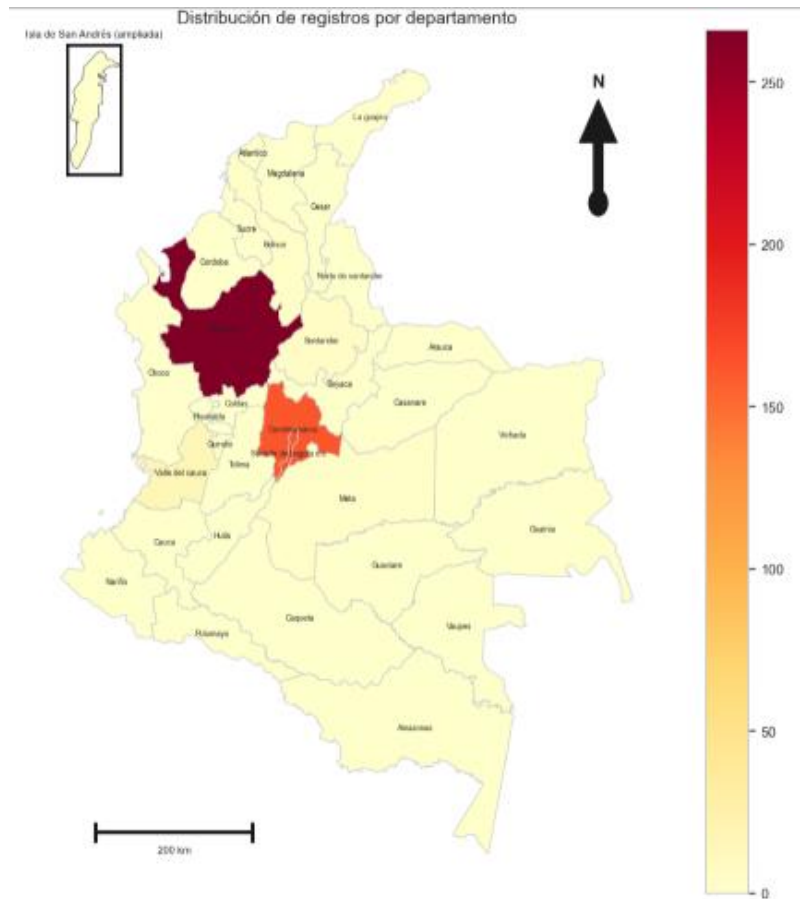
Departamento	Cantidad de registros	Población total	Registros per cápita
Bogotá DC	23023	7849206	0,002933
Antioquia	12863	6777249	0,001898
Norte de Santander	2327	1659814	0,001402
Cundinamarca	3983	3316583	0,001201

Risaralda	933	986996	0,000945
-----------	-----	--------	----------

Nota. Elaboración propia.

Figura 15.

Distribución departamental del subgrupo de público ligero.



Nota. Elaboración propia.

Tabla 7.

Registros per cápita a nivel departamental del subgrupo de público ligero.

Departamento	Cantidad de registros	Población total	Registros per cápita
Cundinamarca	162	3316583	0,000049

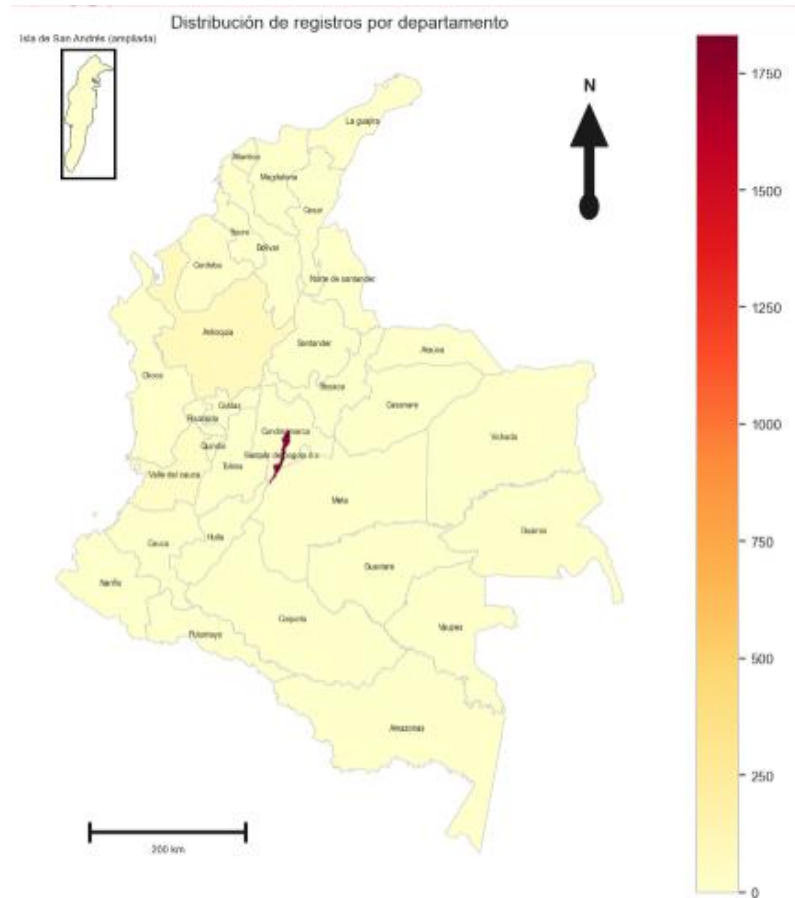
Antioquia	266	6777249	0,000039
Bogotá DC	165	7849206	0,000021
Meta	4	1118349	0,000004
Valle del Cauca	16	4658811	0,000003

Nota. Elaboración propia.

Resalta que en el mapa 16 del subgrupo de público pesado, Bogotá DC tiene una ventaja muy marcada respecto al resto del país, gracias a ser pionera en este tipo de movilidad para el transporte público con grandes implementaciones en su sistema de transporte SITP (Sistema Integrado de Transporte Público).

Figura 16.

Distribución departamental del subgrupo de público pesado.



Nota. Elaboración propia.

Tabla 8.

Registros per cápita a nivel departamental del subgrupo de público pesado.

Departamento	Cantidad de registros	Población total	Registros per cápita
Bogotá DC	1831	7849206	0,000233
Antioquia	72	6777249	0,000011
Valle del Cauca	38	4658811	0,000008
Córdoba	12	1917742	0,000006

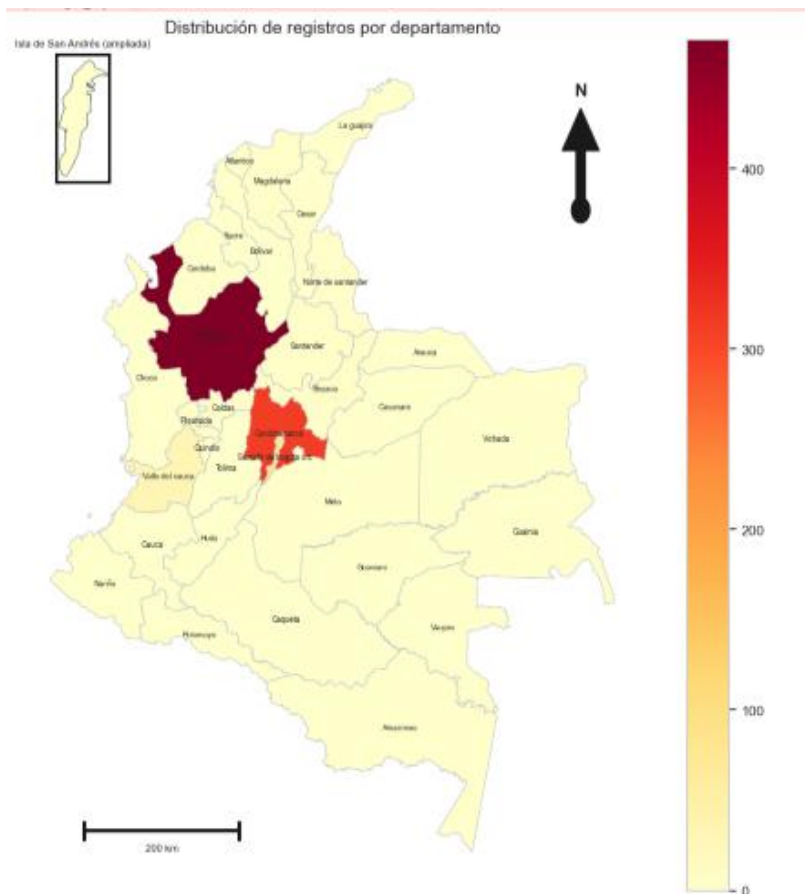
Bolívar 3 2195275 0,000001

Nota. Elaboración propia.

En el único subgrupo en el que Bogotá DC no domina es en el de carga, en donde el líder es Antioquia, como se puede observar en el mapa de la figura 17.

Figura 17.

Distribución departamental del subgrupo de carga.



Nota. Elaboración propia.

Tabla 9.

Registros per cápita a nivel departamental del subgrupo de carga.

Departamento	Cantidad de registros	Población total	Registros per cápita
Cundinamarca	314	3316583	0,000095
Antioquia	471	6777249	0,000069
Bogotá DC	70	7849206	0,000009
Valle del Cauca	32	4658811	0,000007
Risaralda	4	986996	0,000004

Nota. Elaboración propia.

En conclusión, los mapas de calor muestran la marcada centralización geográfica de la adopción de vehículos eléctricos e híbridos. Asimismo, las tablas de registros per cápita muestran la misma brecha con ligeros cambios en el orden de los departamentos. En conjunto, tanto mapas de calor como registros per cápita, evidencian la desigualdad en la penetración tecnológica, influenciada por varios factores socioeconómicos y de infraestructura que se exploraron más adelante en este análisis.

8.4. Segmentación de los Subgrupos

Posteriormente, se realizó la segmentación de cada subgrupo utilizando los algoritmos más adecuados según el tamaño y naturaleza de los datos:

Para los subgrupos `df_moto`, `df_publico_ligero`, `df_publico_pesado` y `df_carga`, se emplearon K-Prototypes, CatBoost con MiniBatch K-Means y agrupamiento jerárquico, dado que el volumen de registros en estos casos es manejable para trabajar con matrices de distancias completas, lo que permite capturar relaciones combinadas entre variables categóricas y numéricas de manera precisa.

En cambio, para el subgrupo `df_particular_oficial_ligero` (con 51167 registros), el tamaño del conjunto de datos hace inviable la construcción de una matriz de distancias tradicional que, al almacenar la distancia entre cada par de registros, genera que su procesamiento sea muy complejo debido al costo computacional y al consumo de memoria. Por ello, se optó por utilizar K-Prototypes y una estrategia computacionalmente eficiente que combina embeddings generados con CatBoost y MiniBatch K-Means, logrando un enfoque adecuado para grandes volúmenes de datos. La tabla 10 resume las técnicas de segmentación utilizadas para cada subgrupo.

Tabla 10.

Técnicas de segmentación utilizadas por cada subgrupo.

Subgrupo	K-Prototypes	Agrupación jerárquica	CatBoost MB K-means
Motos	X	X	X
Particular Oficial ligero	X	NA	X
Público ligero	X	X	X
Público pesado	X	X	X
Carga	X	X	X

Nota. Elaboración propia.

Las variables técnicas utilizadas para la segmentación incluyeron combustible (fuente de energía), clase, marca, cilindraje, potencia, capacidad de pasajeros y capacidad de carga. Sin embargo, la selección se ajustó a las particularidades de cada subgrupo. Para el subgrupo de motos, no se consideraron ni combustible ni cilindraje, dado que todos los registros son de tipo eléctrico. En el caso de los vehículos de carga, se priorizó la

capacidad de carga sobre la capacidad de pasajeros, mientras que en los demás subgrupos se aplicó el criterio inverso, de acuerdo con su uso principal.

Luego, se ejecutó un ciclo de segmentación con el algoritmo K-Prototypes, variando el número de segmentos (`n_clusters`) para generar una gráfica que permitiera identificar el número óptimo de clústeres utilizando el método del codo. Una vez definido ese número óptimo, se evaluó el resultado mediante los índices de silueta y Davies Bouldin.

Asimismo, se aplicaron los algoritmos complementarios correspondientes a cada caso, siguiendo el mismo procedimiento (búsqueda del número óptimo de clústeres y evaluación con índices silueta y Davies Bouldin). Posteriormente, se compararon los puntajes de las técnicas utilizadas y se seleccionó la segmentación con la mejor combinación de puntaje de silueta alto e índice Davies Bouldin bajo, dando prioridad a silueta en caso de empate, garantizando cohesión interna de los segmentos y separación más clara entre los grupos.

El ajuste fino de K-Prototypes se hizo principalmente con los parámetros `n_clusters`, cuyo valor óptimo se encontró mediante el método del codo e `init = 'Cao'`, que establece el método de cálculo de los centroides iniciales. Por su parte, `minibatch K-means` fue parametrizado principalmente mediante `n_clusters`; `batch_size=2048`, que configura la cantidad de observaciones en cada minibatch de entrenamiento; `init='k-means++'`, que inicializa los centroides con un algoritmo que prioriza la distribución y `reassignment_ratio=0.01`, que ayuda a evitar clústeres vacíos o mal situados. Finalmente, agrupación jerárquica fue afinado mediante `n_clusters`; `metric = 'precomputed'`, porque se le envió la matriz de distancias calculada previamente y `linkage = 'average'`, que configura la forma de medición promedio de distancias entre grupos.

Finalmente, los segmentos seleccionados fueron caracterizados estadísticamente mediante el cálculo de moda, promedio de las variables relevantes y mapas de calor de las variables categóricas, con el objetivo de interpretar su comportamiento y relacionarlos con dinámicas reales del mercado.

En la tabla 11 se muestran los resultados de los procesos de segmentación con las tres técnicas utilizadas y sus puntajes de silueta y Davies Bouldin, así como la selección de la mejor técnica (resaltada en negrita) y el número de segmentos identificado en el proceso. Como se explicó en el marco teórico los mejores resultados para puntaje silueta son cercanos a 1 (Soria Olivas et al., 2023), y para puntaje Davies Bouldin son cercanos a 0 (Velasco Rebolledo, 2024).

Tabla 11.

Resultados de las técnicas de segmentación no supervisada ejecutadas en cada subgrupo.

Subgrupo	Técnica [Puntaje Silueta, Puntaje Davies Bouldin]	Número de segmentos
	K-Prototypes [0.296,1.697]	
Motos	Agrupación jerárquica [0.51,0.762]	6
	CatBoost MB K-means [0.857,0.267]	
Particular	K-Prototypes [0.282,1.51]	6
Oficial ligero	CatBoost MB K-means [0.457,1.254]	
	K-Prototypes [0.904,1.876]	
Público ligero	Agrupación jerárquica [0.853,0.297]	5
	CatBoost MB K-means [0.894,0.397]	

	K-Prototypes [0.837,0.376]	
Público		
pesado	Agrupación jerárquica [0.906,0.103]	3
	CatBoost MB K-means [0.927,0.119]	
	K-Prototypes [0.828,0.897]	
Carga	Agrupación jerárquica [0.776,0.285]	4
	CatBoost MB K-means [0.905,0.395]	

Nota. Elaboración propia. La técnica con mejores resultados por cada subgrupo se resalta en negrita.

Los mejores resultados de partición se obtuvieron en los subgrupos de motos y uso público pesado en donde se lograron puntajes de silueta altos y Davies Bouldin bajos, lo que evidencia una segmentación sólida, con clústeres bien definidos y con buena separación entre ellos. En el subgrupo de carga también se obtuvo una segmentación robusta con buena separación. Por su parte, el subgrupo de uso público ligero tiene una segmentación con excelente cohesión interna, pero con menor separación estructural entre clústeres. Finalmente, el subgrupo de uso particular oficial ligero presenta una segmentación con separación moderada pero aceptable, por lo que el perfilamiento técnico de estos clústeres se realizó considerando que las diferencias entre ellos son más graduales que totalmente marcadas.

8.4.1. Resultados de Segmentación en Subgrupo Motos

En la tabla 12 se muestran los segmentos generados para el subgrupo de motos. Como se había mencionado anteriormente y se muestra en la figura 18, todos los registros de este subgrupo son eléctricos. En complemento con los mapas de calor de las figuras

19 y 20, se realizó la caracterización basada en frecuencia de variables categóricas y promedio de variables numéricas.

Tabla 12.

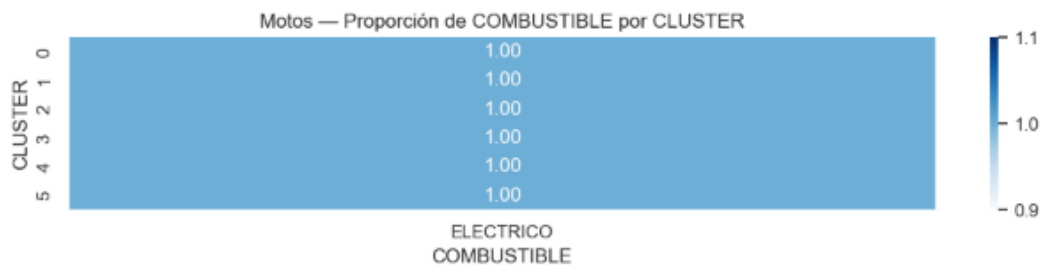
Segmentación de subgrupo moto.

Segmento	Capacidad pasajeros	Potencia (KW)	Marca	Clase	Número de registros
0	1	4.5	Energy Motion	Motocicleta	738
1	2	2.2	Sunra	Motocicleta	298
2	0	1.3	Winlee	Tricimoto	67
3	2	11.2	Zero	Motocicleta	119
4	1.9	29.4	BMW	Motocicleta	15
5	1	1.2	Green Thor	Motocicleta	625

Nota. Elaboración propia.

Figura 18.

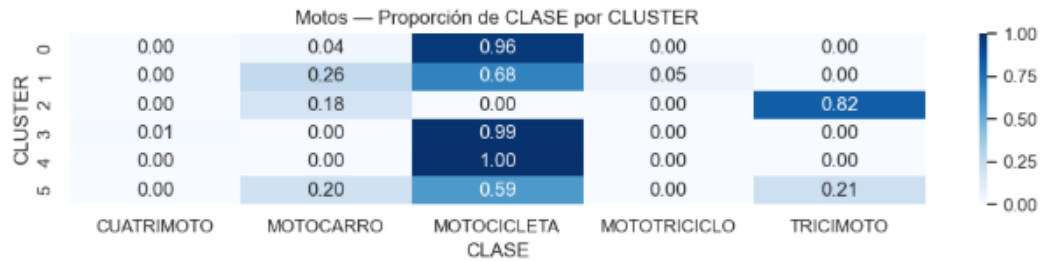
Proporción de fuente de energía por clúster.



Nota. Elaboración propia.

Figura 19.

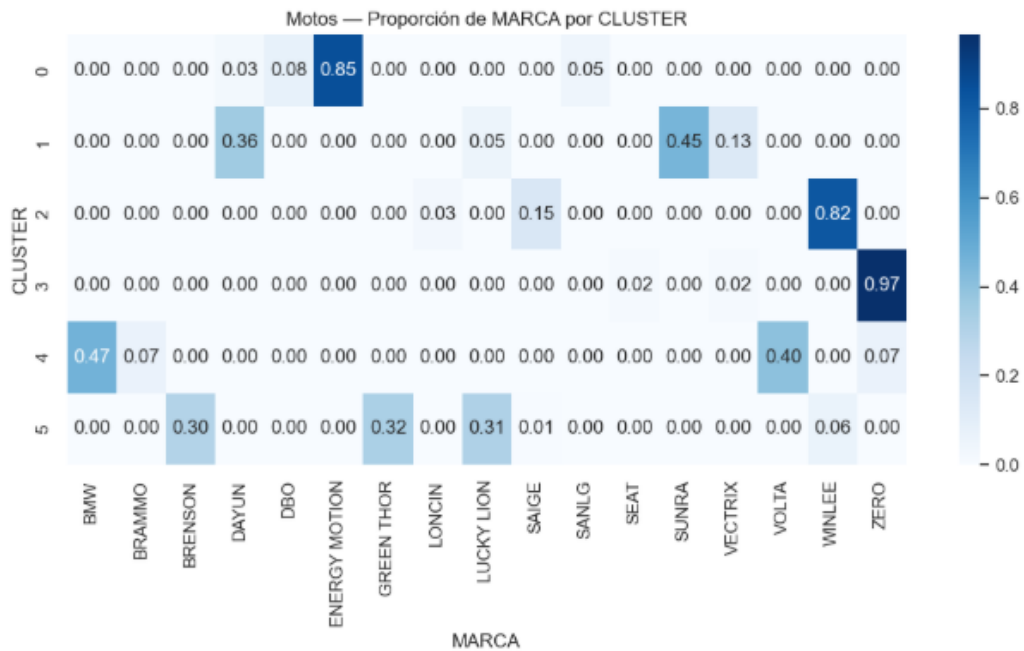
Proporción de clase por clúster.



Nota. Elaboración propia.

Figura 20.

Proporción de marca por clúster.



Nota. Elaboración propia.

Los segmentos resultantes se describen a continuación.

Segmento 0: Motocicletas de potencia media con marca líder Energy Motion. Es el segmento más grande.

Segmento 1: Motocicletas de baja potencia, repartidas principalmente entre marcas Sunra y Dayun.

Segmento 2: Tricimotos con marca líder Winlee.

Segmento 3: Motocicletas de alta potencia, de marca Zero principalmente.

Segmento 4: Motocicletas de altísima potencia repartidas principalmente entre marcas BMW y Volta. Segmento de nicho

Segmento 5: Motocicletas de baja potencia, con un mercado dividido entre las marcas Green Thor, Brenson y Lucky Lion.

8.4.2. Resultados de Segmentación en Subgrupo Particular Oficial Ligero

En el subgrupo más numeroso, que es el de uso particular y oficial ligero, se identificaron los segmentos registrados en la tabla 13, y se caracterizaron con apoyo de los mapas de calor de las figuras 21, 22 y 23.

Tabla 13.

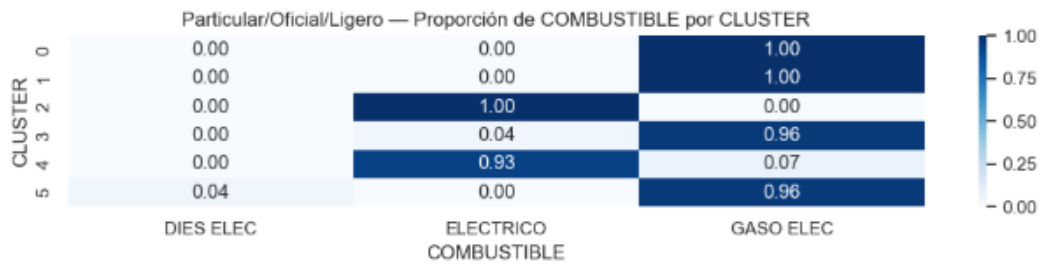
Segmentación de subgrupo particular y oficial ligero.

Segmento	Fuente de energía	Clase	Marca	Cilindraje (cc)	Capacidad pasajeros	Potencia (KW)	Número de registros
0	Gasol	Automóvil	Suzuki	1222.2	5	66.4	7563
1	Gasol	Camioneta	Ford	2461.9	5	200.12	13485
2	Eléctrico	Cuadriciclo	Renault	0.6	1.9	12.3	1332
3	Gasol	Camioneta	Mercedes Benz	2519.8	6.9	310.3	1277
4	Eléctrico	Automóvil	BMW	0.2	4.7	131	3776
5	Gasol	Camioneta	Toyota	1833.9	5	124.1	23734

Nota. Elaboración propia.

Figura 21.

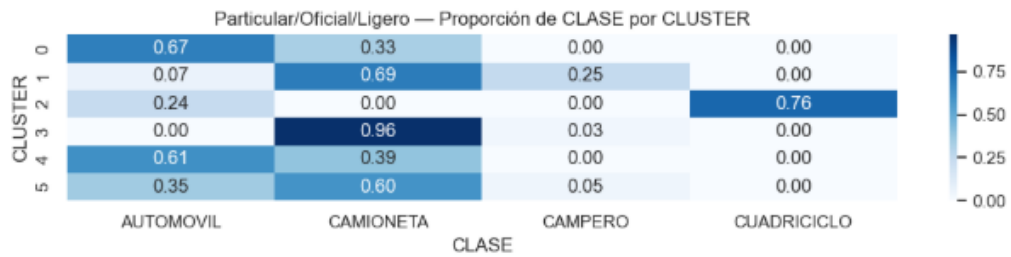
Proporción de fuente de energía por clúster.



Nota. Elaboración propia.

Figura 22.

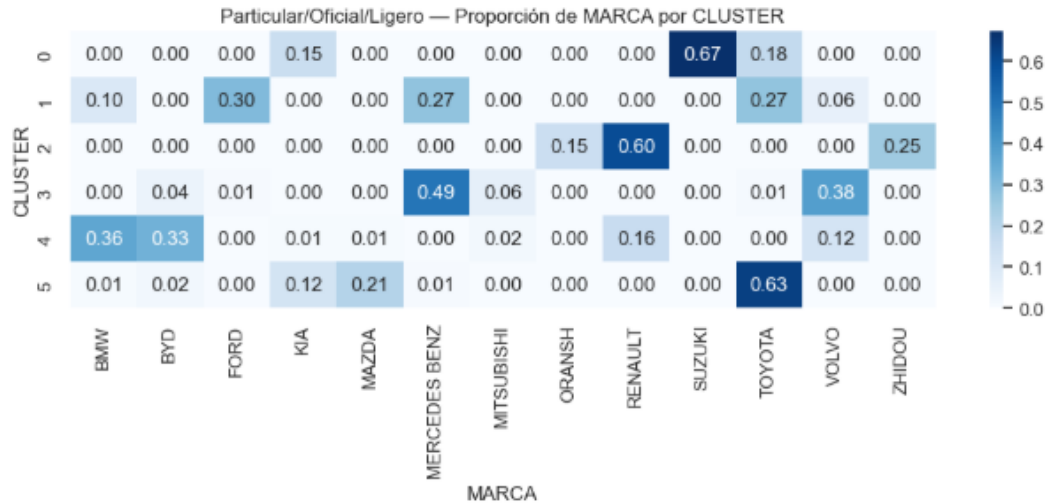
Proporción de clase por clúster.



Nota. Elaboración propia.

Figura 23.

Proporción de marca por clúster.



Nota. Elaboración propia.

Los segmentos resultantes se describen a continuación.

Segmento 0: Automóviles híbridos, marca líder Suzuki.

Segmento 1: Camionetas híbridas de alta potencia con capacidad estándar de pasajeros (5), con mercado bastante dividido entre marcas Ford, Mercedes Benz y Toyota.

Segmento 2: Ultralivianos urbanos, con marca líder Renault. Resalta su potencia baja y capacidad para 1 o 2 pasajeros.

Segmento 3: Camionetas híbridas de más alta potencia con capacidad ampliada de pasajeros (7) con marca líder Mercedes Benz, seguida de cerca por Volvo.

Segmento 4: Automóviles eléctricos, con marcas principales BMW y BYD.

Segmento 5: Camionetas híbridas de potencia media con capacidad estándar de pasajeros (5), marca líder Toyota. Es el segmento más grande.

8.4.3. Resultados de Segmentación en Subgrupo Público Ligero

Para el subgrupo de uso público ligero, la tabla 14 muestra los segmentos generados por el proceso de segmentación.

Tabla 14.

Segmentación de subgrupo público ligero.

Segmento	Fuente de energía	Clase	Marca	Cilindraje (cc)	Capacidad pasajeros	Potencia (KW)	Número de registros
0	Eléctrico	Camioneta	Dongfeng	0	5	120	266
1	Eléctrico	Camioneta	BYD	400	7	114.7	6
2	Gasol Eléctrico	Camioneta	Kia	1586.8	5	101.32	272
3	Eléctrico	Automóvil	BYD	0	5	95.7	61
4	Eléctrico	Automóvil	BYD	138.5	5	162.4	26

Nota. Elaboración propia.

Figura 24.

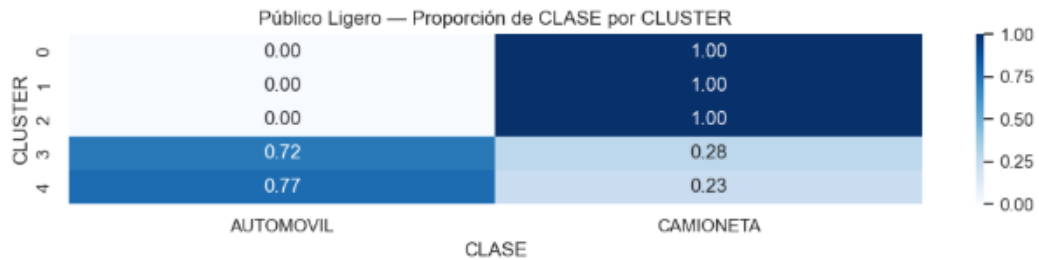
Proporción de fuente de energía por clúster.



Nota. Elaboración propia.

Figura 25.

Proporción de clase por clúster.



Nota. Elaboración propia.

Figura 26.

Proporción de marca por clúster.



Nota. Elaboración propia.

A continuación, la caracterización de cada segmento identificado con ayuda de los mapas de calor de las figuras 24, 25 y 26.

Segmento 0: Camionetas eléctricas de marca Dongfeng.

Segmento 1: Camionetas eléctricas de capacidad de pasajeros ampliada (7), con marca líder BYD. Segmento marginal.

Segmento 2: Camionetas híbridas de marca Kia.

Segmento 3: Automóviles eléctricos de baja potencia de marca BYD.

Segmento 4: Automóviles eléctricos de media potencia, con marca líder BYD.

8.4.4. Resultados de Segmentación en Subgrupo Público Pesado

Dentro del subgrupo de público pesado, como se observa en la tabla 15 y los mapas de calor de las figuras 27, 28 y 29, se encontraron los siguientes segmentos con su respectiva caracterización.

Tabla 15.

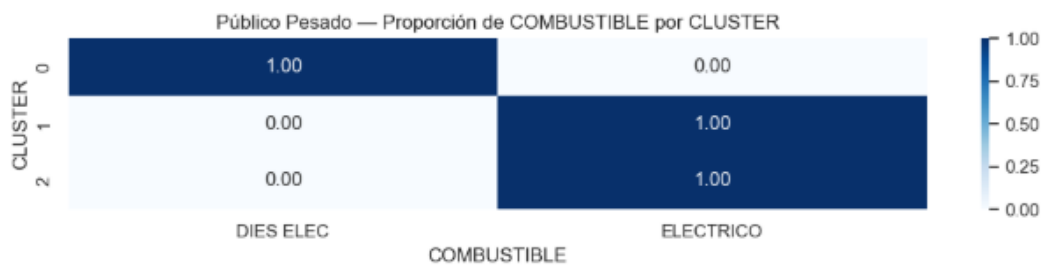
Segmentación de subgrupo público pesado.

Segmento	Fuente de energía	Clase	Marca	Cilindraje (cc)	Capacidad pasajeros	Potencia (KW)	Número de registros
0	Diésel Eléctrico	Bus	Volvo	4760.7	76.1	162	340
1	Eléctrico	Bus	BYD	0	51.7	201.4	875
2	Eléctrico	Bus	BYD	0	80.4	303.6	752

Nota. Elaboración propia.

Figura 27.

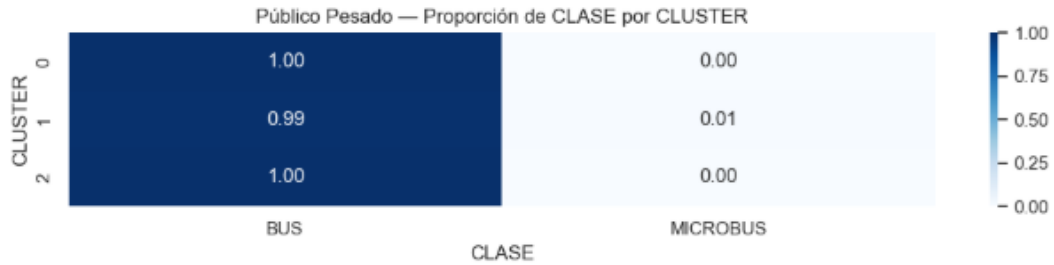
Proporción de fuente de energía por clúster.



Nota. Elaboración propia.

Figura 28.

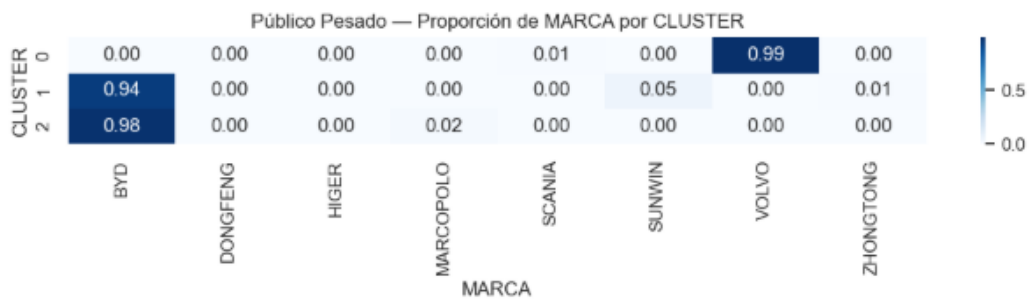
Proporción de clase por clúster.



Nota. Elaboración propia.

Figura 29.

Proporción de marca por clúster.



Nota. Elaboración propia.

Segmento 0: Buses híbridos de baja potencia y capacidad promedio de pasajeros de 76 de marca Volvo.

Segmento 1: Buses eléctricos de mediana potencia y capacidad promedio de pasajeros de 52 de marca BYD. Es el segmento más grande.

Segmento 2: Buses eléctricos de alta potencia de marca BYD.

8.4.5. Resultados de Segmentación en Subgrupo Carga

Por último, los resultados de la segmentación del subgrupo de carga están registrados en la tabla 16 y en los mapas de calor de las figuras 30, 31 y 32.

Tabla 16.

Segmentación de subgrupo carga.

Segmento	Fuente de energía	Clase	Marca	Cilindraje (cc)	Capacidad carga (Kg)	Potencia (KW)	Número de registros
0	Diésel Eléctrico	Camión	Hino	4000	2570	97.7	95
1	Eléctrico	Camión	JAC	0	1974.5	103.4	118
2	Eléctrico	Camión	Renault	0	763.9	40.8	296
3	Eléctrico	Camión	Stark	0	4166.2	121	409

Nota. Elaboración propia.

Figura 30.

Proporción de fuente de energía por clúster.



Nota. Elaboración propia.

Figura 31.

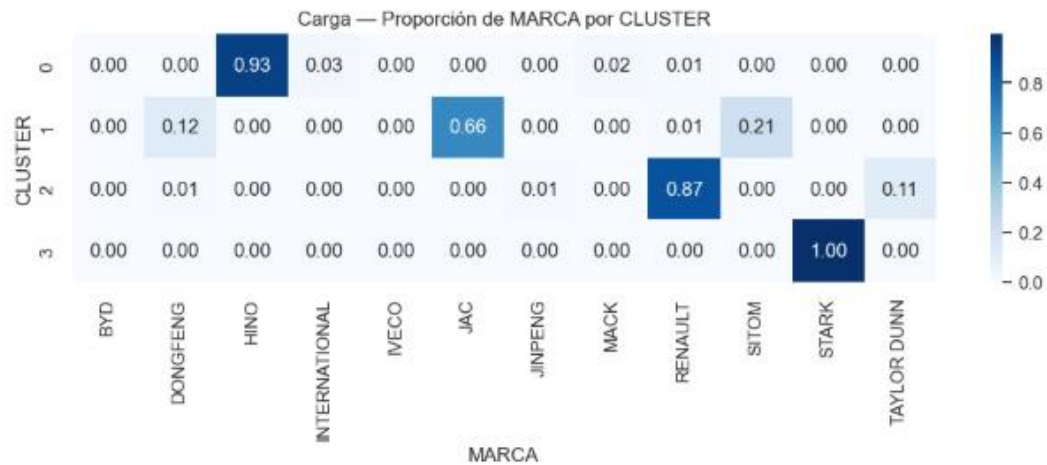
Proporción de clase por clúster.



Nota. Elaboración propia.

Figura 32.

Proporción de marca por clúster.



Nota. Elaboración propia.

La caracterización de los segmentos se muestra a continuación.

Segmento 0: Camión híbrido con media capacidad de carga de marca Hino.

Segmento 1: Camión eléctrico con media capacidad de carga, con marca líder JAC.

Segmento 2: Camión eléctrico con baja capacidad de carga, con marca líder Renault.

Segmento 3: Camión eléctrico con alta capacidad de carga de marca Stark.

8.5. Evolución de los Segmentos 2010 a 2022

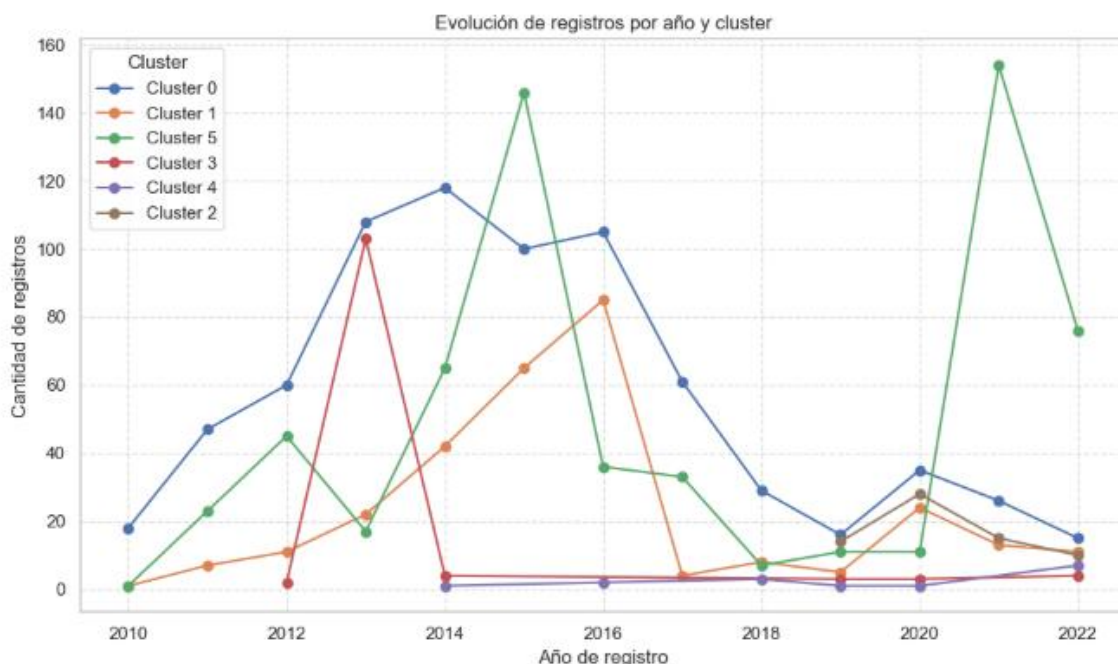
Como complemento al análisis, se graficó la evolución temporal de cada segmento por año, con el fin de identificar cuáles presentan un mayor crecimiento en el tiempo.

Dentro del subgrupo de motos, según lo observado en la figura 33, los segmentos de motocicletas de potencia media y baja tuvieron un comportamiento positivo hasta 2015,

año en el cual inició una caída marcada de la cual solo repuntó con fuerza el segmento de baja potencia dominado por las marcas Green Thor, Brenson y Lucky Lion (segmento 5). Los otros segmentos no tienen una evolución notable.

Figura 33.

Evolución de registros por año de cada segmento en el subgrupo motos.



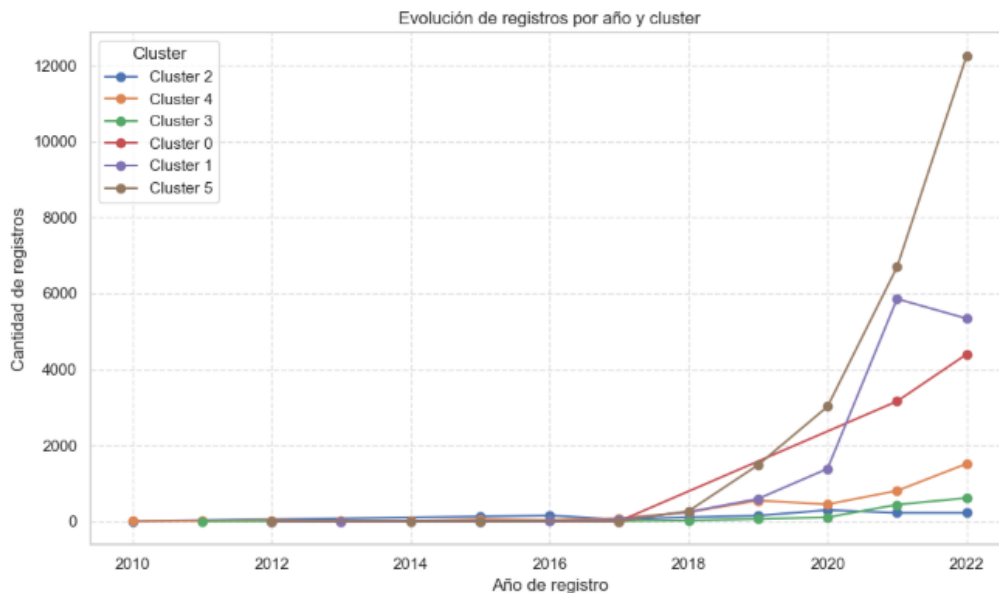
Nota. Elaboración propia.

En este subgrupo, la tendencia general no evidencia un crecimiento tan pronunciado como podría sugerir la percepción general. Este comportamiento puede explicarse en línea con la alerta realizada por la Asociación Nacional de Empresarios Colombianos (ANDI) sobre el crecimiento de la importación y comercialización informal e ilegal de ciclomotores en Colombia (Asociación Nacional de Empresarios de Colombia, 2025), que muestra una brecha significativa entre las unidades importadas y aquellas registradas. En este aspecto, el presente análisis se limita exclusivamente a las cifras oficiales de vehículos registrados.

En cuanto al subgrupo de uso particular y oficial ligero, de acuerdo a la figura 34, el segmento de camionetas híbridas de potencia media con capacidad estándar de pasajeros (5) es el que muestra un crecimiento más fuerte y prometedor. Los segmentos de automóvil híbrido y camioneta híbrida de alta potencia con capacidad estándar de pasajeros presentan también tendencias positivas que pueden seguir creciendo y que ya tienen una buena presencia en este subgrupo. El segmento de automóvil eléctrico tiene un porcentaje de mercado pequeño, pero con buen crecimiento desde el año 2020, por lo que puede ser una buena oportunidad de mercado. Los segmentos de ultraliviano urbano y camioneta híbrida de más alta potencia con capacidad ampliada de pasajeros (7) parecen ser segmentos de nicho.

Figura 34.

Evolución de registros por año de cada segmento en el subgrupo particular y oficial ligero.

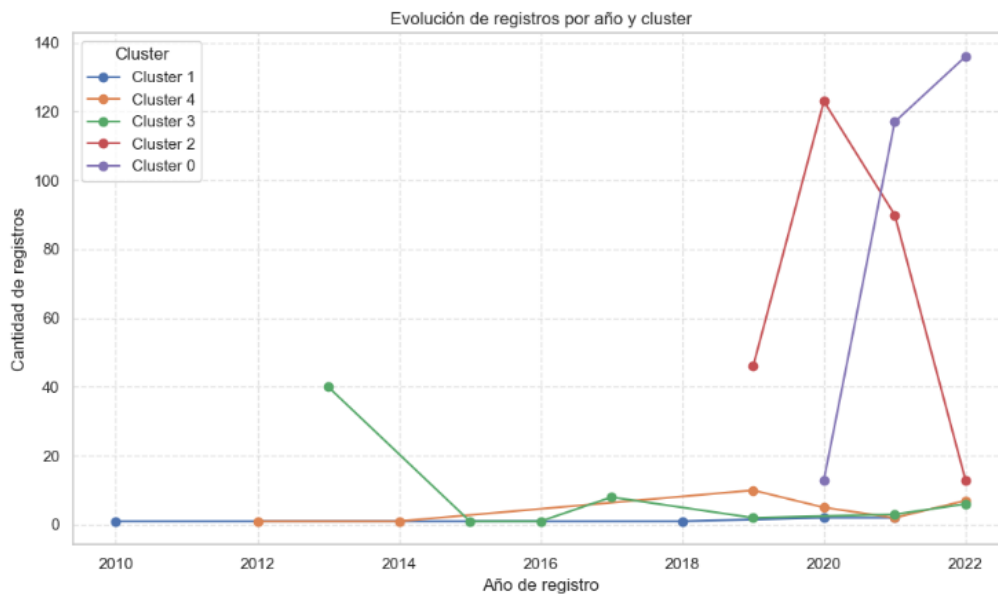


Nota. Elaboración propia.

Como lo muestra la figura 35, en el subgrupo de uso público ligero, el segmento de camioneta híbrida tuvo un incremento significativo en 2020, pero una fuerte caída en los dos años siguientes dejándolo en 2022 con una cantidad de vehículos registrados por debajo de 20. El segmento de camioneta eléctrica ha surgido con fuerza desde 2020 a 2022. Los otros segmentos presentan una participación baja y su evolución no parece ser prometedora.

Figura 35.

Evolución de registros por año de cada segmento en el subgrupo público ligero.



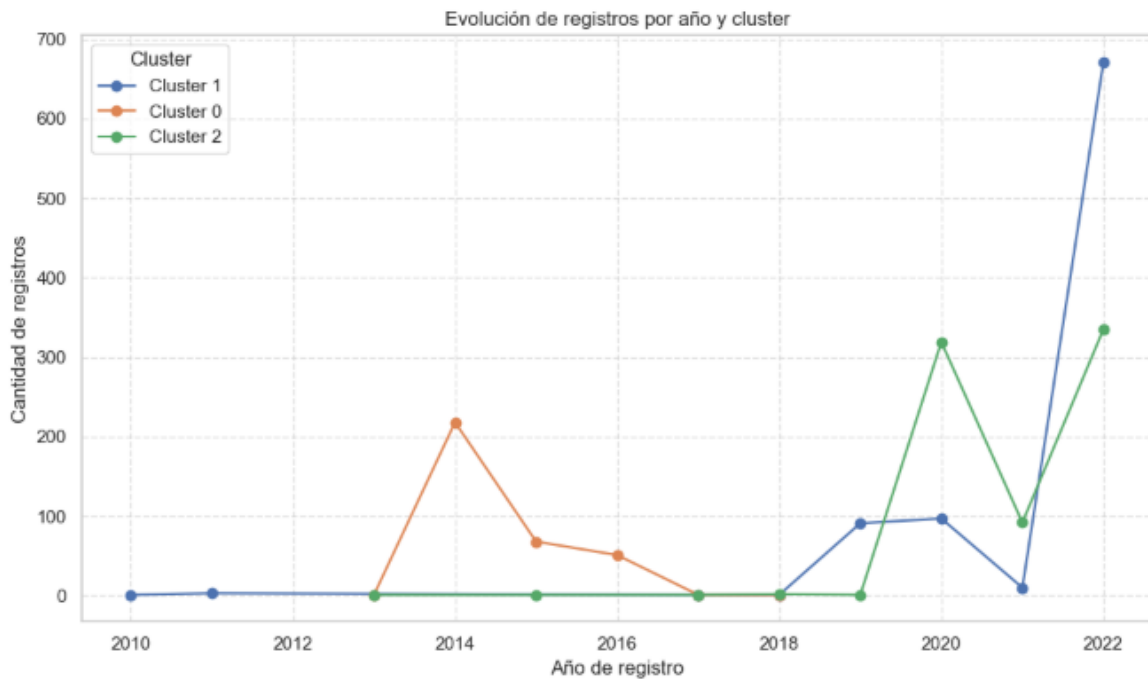
Nota. Elaboración propia.

Evaluando el subgrupo de uso público pesado, tal como se detalla en la figura 36, el segmento con la mayor tendencia positiva es el de bus eléctrico de mediana potencia y capacidad promedio de pasajeros de 52, que ha presentado un crecimiento fuerte especialmente en 2022. El segmento de buses eléctricos de alta potencia ha tenido un

comportamiento oscilante desde 2019 y el segmento de buses híbridos ha caído desde 2015 hasta el punto de no tener registros desde 2018.

Figura 36.

Evolución de registros por año de cada segmento en el subgrupo público pesado.

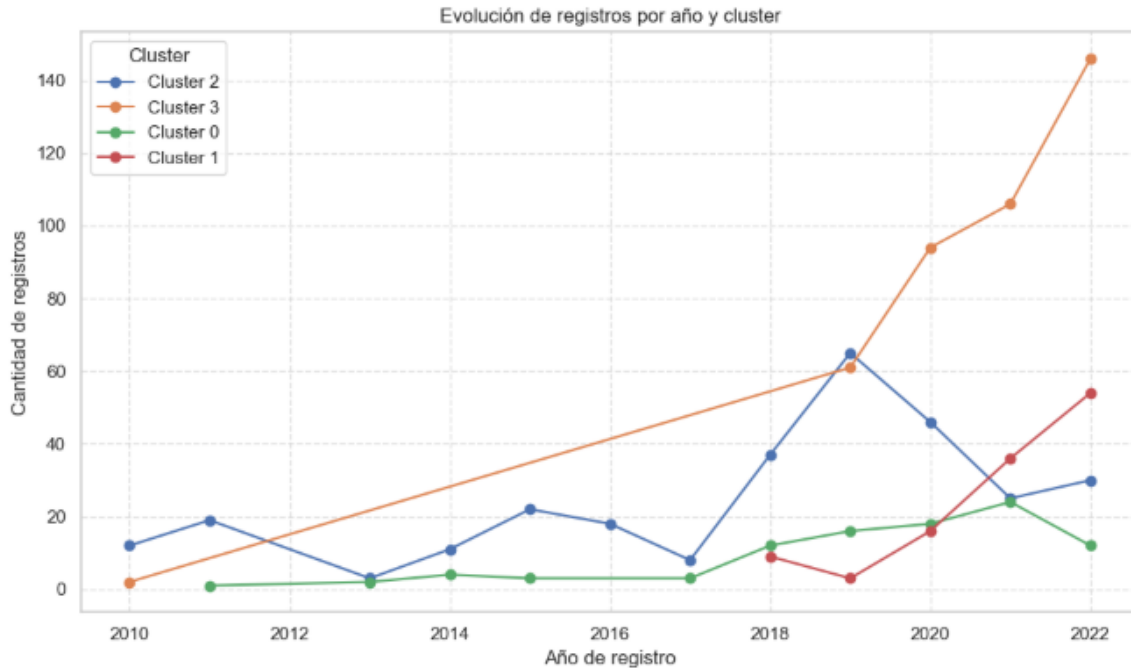


Nota. Elaboración propia.

Por último, en el subgrupo de carga, cuyos resultados se presentan en la figura 37, los segmentos de camión eléctrico de alta capacidad de carga y camión eléctrico de media capacidad de carga son los que muestran un crecimiento reciente. El segmento de camión eléctrico de baja capacidad de carga tuvo un pico alto en 2019 pero después decayó. Los camiones híbridos tienen una tendencia creciente pero moderada.

Figura 37.

Evolución de registros por año de cada segmento en el subgrupo carga.



Nota. Elaboración propia.

8.6. Proyección de los Segmentos a 2028

Finalmente, se proyectó la adopción hasta 2028 utilizando Prophet, modelo de series de tiempo desarrollado por Meta Open Source. Se escogió este modelo por su robustez al trabajar con datos incompletos, ruidosos o con cambios bruscos, como los que se pudieron observar en la sección anterior, con segmentos cuyo comportamiento histórico muestra años sin registros y puntos de quiebre. Otros modelos requieren la cadena de registros completa por lo que no son compatibles con los datos disponibles. Además, entrega gráficas de tendencia con rango de incertidumbre que facilitan la interpretación.

Este proceso permitió obtener proyecciones para cada segmento identificado en cada subgrupo a partir de sus datos históricos, con el fin de facilitar la anticipación del

comportamiento a futuro y tener insumos que ayuden a tomar decisiones en el sector de política pública y estrategias comerciales.

Como se explicó en la metodología solo se tuvieron en cuenta los segmentos con registros de al menos 5 años, y a cada proyección se le calculó la incertidumbre promedio y relativa de los años pronosticados.

La preparación de los datos de entrada consistió en agrupar los registros de cada subgrupo por clúster y por año, obteniendo un *dataset* con las variables clúster, año y número de registros. Posteriormente, mediante un ciclo, se generó un grupo de datos filtrado de cada clúster con el fin de analizar su evolución histórica hasta 2022 y se utilizó *cross-validation* temporal, para entrenar el modelo con los datos del pasado y evaluar su desempeño al predecir los últimos tres años históricos, usando la métrica de rendimiento MAE (Error Absoluto Medio) como criterio de calibración. Se optó por esta métrica por su facilidad de interpretación al expresarse en las mismas unidades de las variables. El ajuste del modelo se realizó con los siguientes hiperparámetros fijos: *growth* lineal, que evita proyecciones aceleradas poco realistas y favorece cambios graduales; y *yearly_seasonality* false, puesto que la estacionalidad (patrones que se repiten en periodos fijos) no aplica a este análisis con registros anuales. Adicionalmente, se evaluaron distintos valores de los hiperparámetros *changeoint_prior_scale* de 0.01, 0.05 y 0.1, donde valores menores reducen la sensibilidad del modelo a variaciones pequeñas; y *n_changepoints* de 1, 2 y 3, que restringe el número de cambios bruscos en la tendencia estimada. El modelo con el menor MAE de validación fue reentrenado y utilizado para las proyecciones finales.

En general, los valores de MAE obtenidos son relativamente altos frente al nivel promedio de registros por año. Esto es consistente con las características de la información

disponible, que tiene registros anuales de máximo 12 observaciones históricas y sin estacionalidad, lo que limita significativamente la capacidad de proyección. En este contexto, el modelo capta la tendencia general, pero se ve limitado por la alta variabilidad y cantidad escasa de datos. En algunos segmentos se observa una diferencia significativa entre el MAE de entrenamiento y el MAE de validación, que indica sobreajuste del modelo. Por tanto, las proyecciones de estos grupos son menos confiables.

La salida del modelo para el número de registro de vehículos eléctricos e híbridos se compone de tres estimaciones, calculadas a partir de la calibración de los datos históricos 2010-2022: \hat{y} , que corresponde a la predicción más probable; \hat{y}_{lower} , que representa un escenario conservador; y \hat{y}_{upper} , que refleja un escenario optimista.

La incertidumbre de cada predicción anual se obtiene como el ancho del intervalo de predicción, que corresponde a la diferencia entre \hat{y}_{upper} y \hat{y}_{lower} . Para evaluar la fiabilidad global del modelo, se calcula la incertidumbre promedio de todos los años proyectados y se normaliza dividiéndola por el último valor predicho, obteniendo así una medida adimensional de incertidumbre relativa. Este indicador permite interpretar las proyecciones para el periodo 2023-2028 en función de la magnitud histórica característica de cada segmento.

8.6.1. Proyección del Subgrupo Motos

En el subgrupo de motos se encuentran los resultados descritos en la tabla 17.

Tabla 17.

Métricas de proyecciones para subgrupo motos.

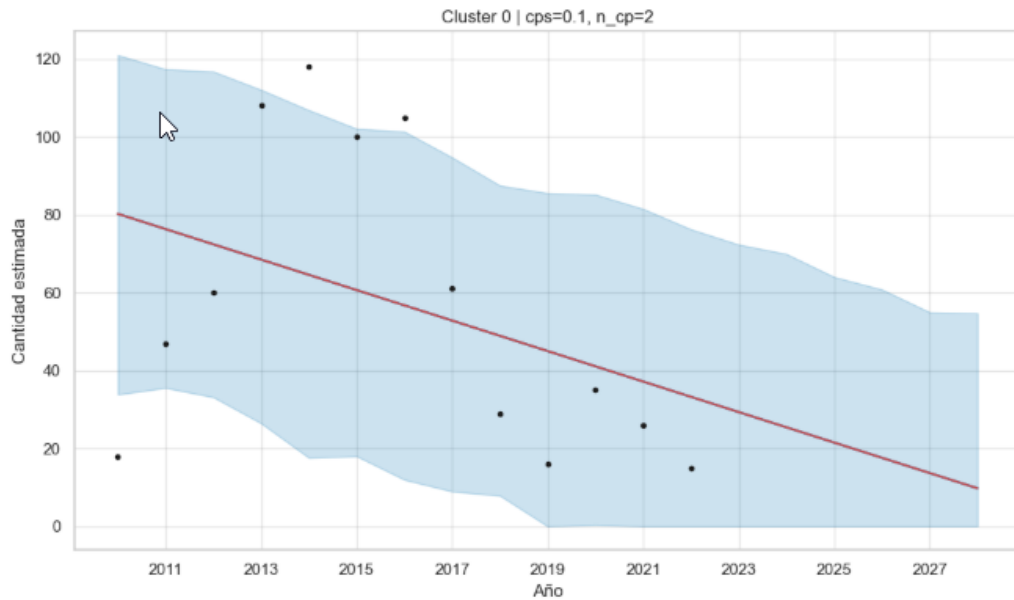
Clúster	Último valor	Último valor	Incertidumbre		MAE de	MAE de
	histórico (unidades registradas)	proyectado (unidades registradas)	promedio (unidades registradas)	Incertidumbre relativa	entrenamiento (unidades registradas)	validación (unidades registradas)
0	33.26	9.78	77.44	7.92	31.86	25.11
1	21.71	20.47	53.09	2.59	21.47	13.53
3	0.0	0.0	42.94	NA	35.19	14.09
4	4.42	7.42	4.07	0.55	0.48	3.02
5	71.83	95.57	108.57	1.14	30.9	52.86

Nota. Elaboración propia.

De acuerdo a las figuras 38 y 39, en los segmentos de motocicletas de potencia media (clúster 0) y alta (clúster 3), se proyecta que continúe su caída. Para el segmento de motocicletas de potencia media la cifra proyectada para 2028 es de 9.78 registros y la incertidumbre promedio es de 77.44, lo que resulta en una incertidumbre relativa de 7.92 derivada principalmente de subidas y caídas frecuentes con cambios de tendencia fuertes. Si bien, la incertidumbre es alta la proyección para este segmento es negativa. La proyección de caída del segmento de motocicletas de potencia alta presenta mayor confiabilidad.

Figura 38.

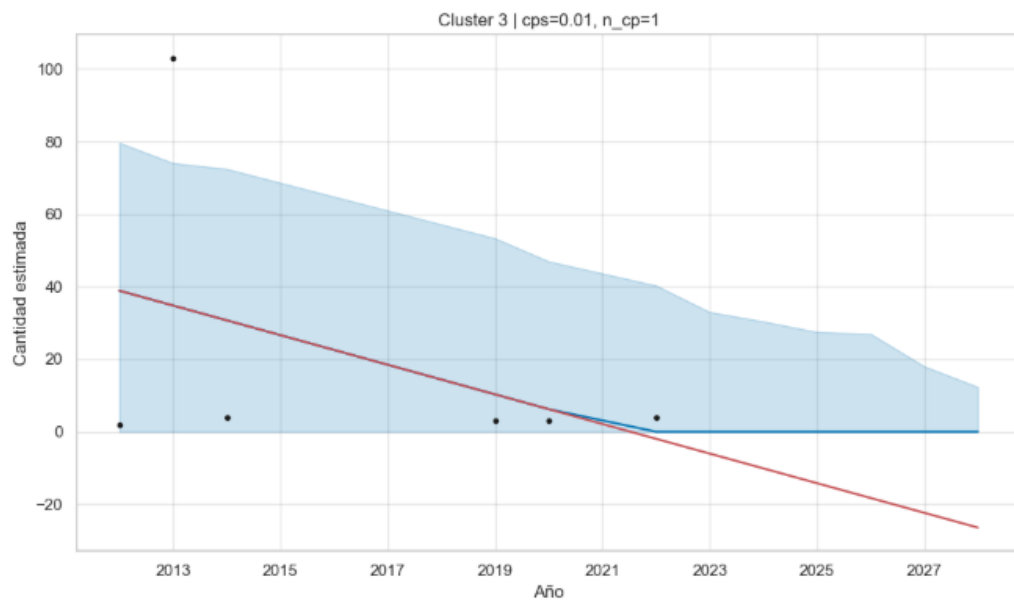
Pronóstico Prophet para subgrupo motos segmento 0.



Nota. Elaboración propia.

Figura 39.

Pronóstico Prophet para subgrupo motos segmento 3.



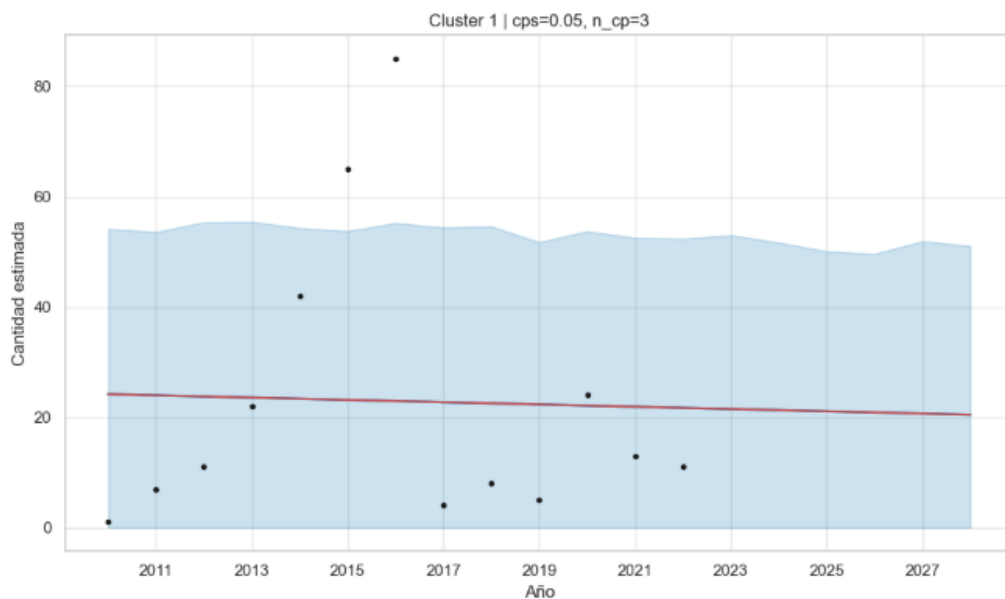
Nota. Elaboración propia.

El segmento de tricimotos (clúster 2) tiene menos de 5 años con registros por lo que no fue considerado para la proyección.

Según lo expuesto en la figura 40, el segmento de motocicletas de baja potencia de marca Sunra y Dayún no tiene una tendencia marcada, la proyección es de 20 registros por año aproximadamente, pero con una incertidumbre considerable.

Figura 40.

Pronóstico Prophet para subgrupo motos segmento 1.

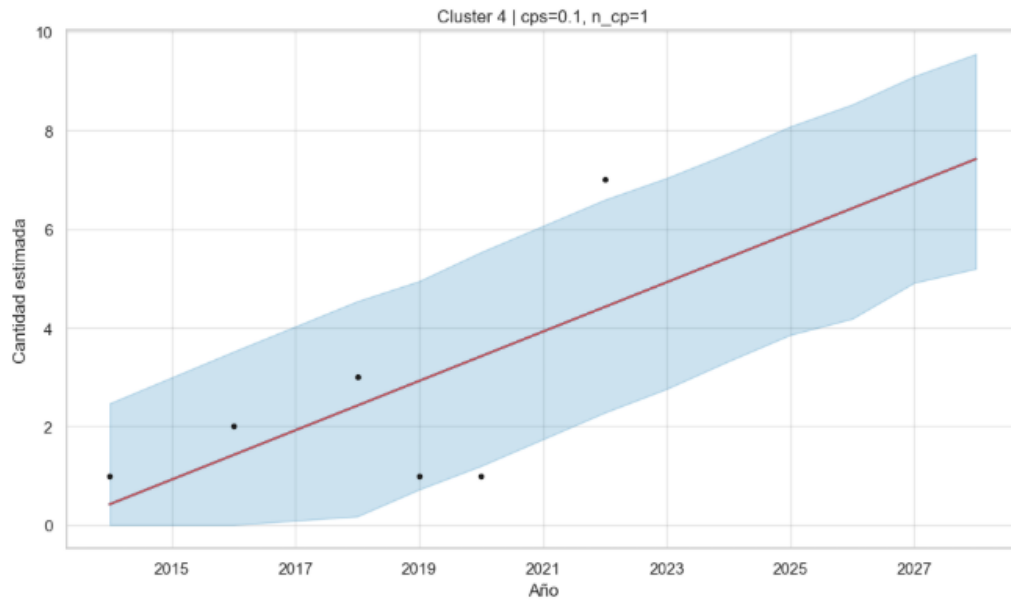


Nota. Elaboración propia.

Según lo expuesto en las figuras 41 y 42, en los demás segmentos se proyecta que mantengan un crecimiento suave. Resalta la incertidumbre del segmento de baja potencia de marcas Green Thor, Brenson y Lucky Lion debido a comportamientos fluctuantes con picos pronunciados y cambios bruscos en su tendencia.

Figura 41.

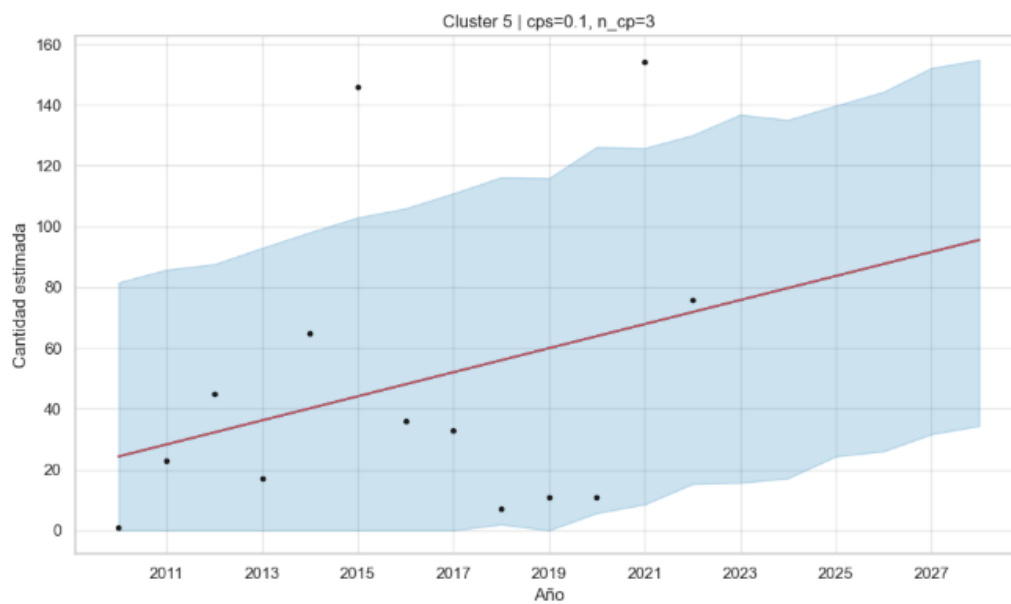
Pronóstico Prophet para subgrupo motos segmento 4.



Nota. Elaboración propia.

Figura 42.

Pronóstico Prophet para subgrupo motos segmento 5.



Nota. Elaboración propia.

8.6.2. Proyección del Subgrupo Particular Oficial Ligero

En el subgrupo de uso particular y oficial ligero se encuentran los resultados descritos en la tabla 18.

Tabla 18.

Métricas de proyecciones para subgrupo particular oficial ligero.

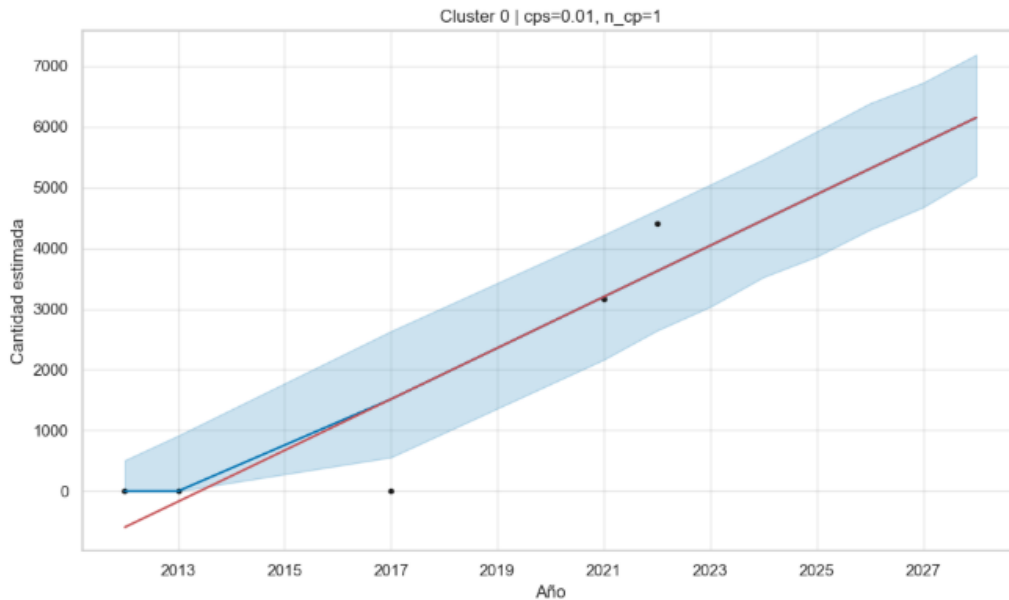
Clúster	Último valor	Último valor	Incertidumbre		MAE de	MAE de
	histórico (unidades registradas)	proyectado (unidades registradas)	promedio (unidades registradas)	Incertidumbre relativa	entrenamiento (unidades registradas)	validación (unidades registradas)
0	3621.6	6149.23	1790.74	0.29	154.91	1577.49
1	3767.61	6852.34	3026.53	0.44	403.68	2740.55
2	237.48	358.18	123.89	0.35	41.52	55.18
3	334.16	586.76	309.9	0.53	32.78	276.67
4	845.15	1399.3	545.29	0.39	84.61	478.17
5	6951.49	12747.29	5670.42	0.44	582.06	4879.13

Nota. Elaboración propia.

En el segmento de automóviles híbridos (clúster 0), que corresponde a la figura 43, se observa una tendencia creciente sostenida. Esto evidencia que es un segmento en plena expansión y consolidación dentro del mercado. La predicción de Prophet sugiere un incremento significativo de la flota (6149) para el año 2028. La incertidumbre relativa de esta predicción es de 0.29, lo que brinda confiabilidad en estos resultados.

Figura 43.

Pronóstico Prophet para subgrupo particular y oficial ligeros segmento 0.

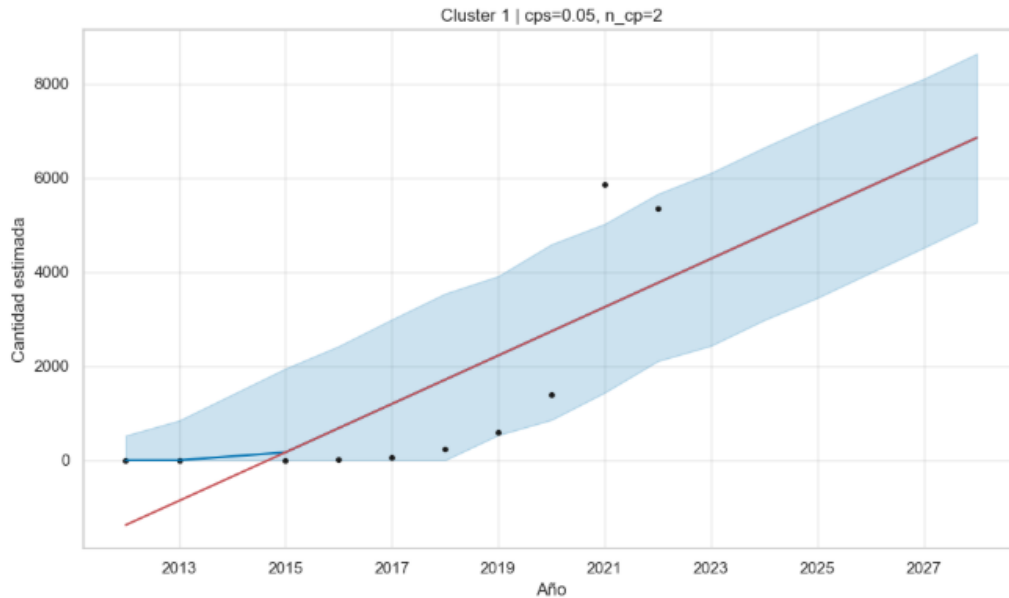


Nota. Elaboración propia.

En el segmento de camionetas híbridas de alta potencia con capacidad estándar de pasajeros (clúster 1), que se observa en la figura 44, se pronostica un crecimiento importante, que podría llegar a registrar 6852 unidades en 2028. El escenario menos optimista sigue siendo bueno con un crecimiento aproximado de 5000 unidades registradas en 2028.

Figura 44.

Pronóstico Prophet para subgrupo particular y oficial ligeros segmento 1.

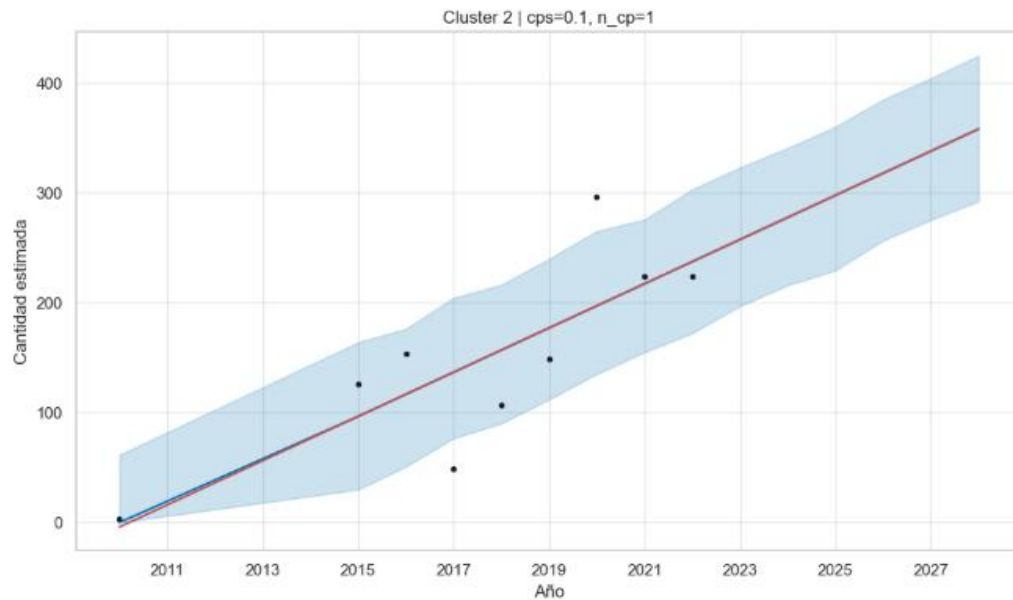


Nota. Elaboración propia.

Como se observa en la figura 45, en el segmento de ultraliviano urbano (clúster 2), se proyecta un crecimiento continuo hasta llegar a registrar 358 vehículos en 2028, manteniendo su expansión con una base estable. Con una incertidumbre relativa de 0.35, la proyección para este segmento no da margen a un cambio en la tendencia positiva.

Figura 45.

Pronóstico Prophet para subgrupo particular y oficial ligeros segmento 2.

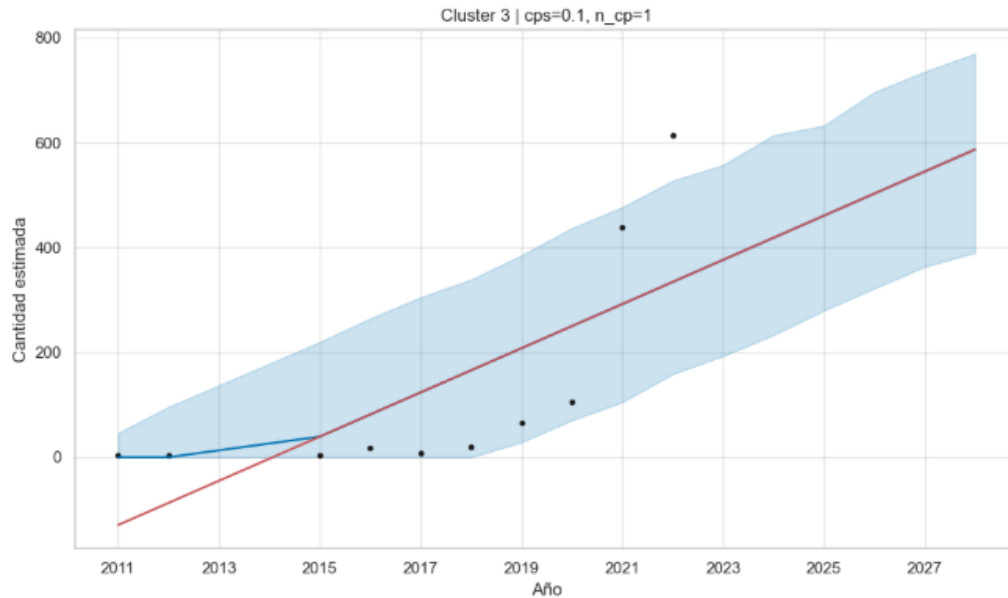


Nota. Elaboración propia.

En el segmento de camionetas híbridas de más alta potencia con capacidad ampliada de pasajeros (clúster 3), que se puede detallar en la figura 46, se presenta una predicción positiva con un registro esperado de 586 unidades para 2028.

Figura 46.

Pronóstico Prophet para subgrupo particular y oficial ligeros segmento 3.

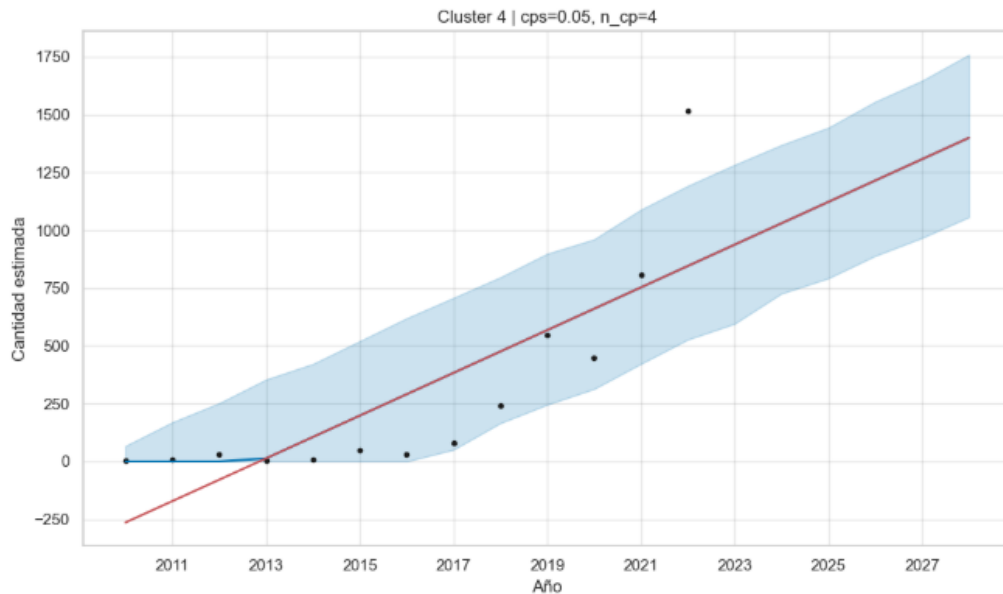


Nota. Elaboración propia.

Para el segmento de automóvil eléctrico (clúster 4), la figura 47 evidencia una tendencia de crecimiento sostenido hasta 2028, con un registro esperado para ese año de 1399 vehículos. La incertidumbre relativa es de 0.39 lo que sugiere estabilización incluso en el escenario menos optimista.

Figura 47.

Pronóstico Prophet para subgrupo particular y oficial ligeros segmento 4.

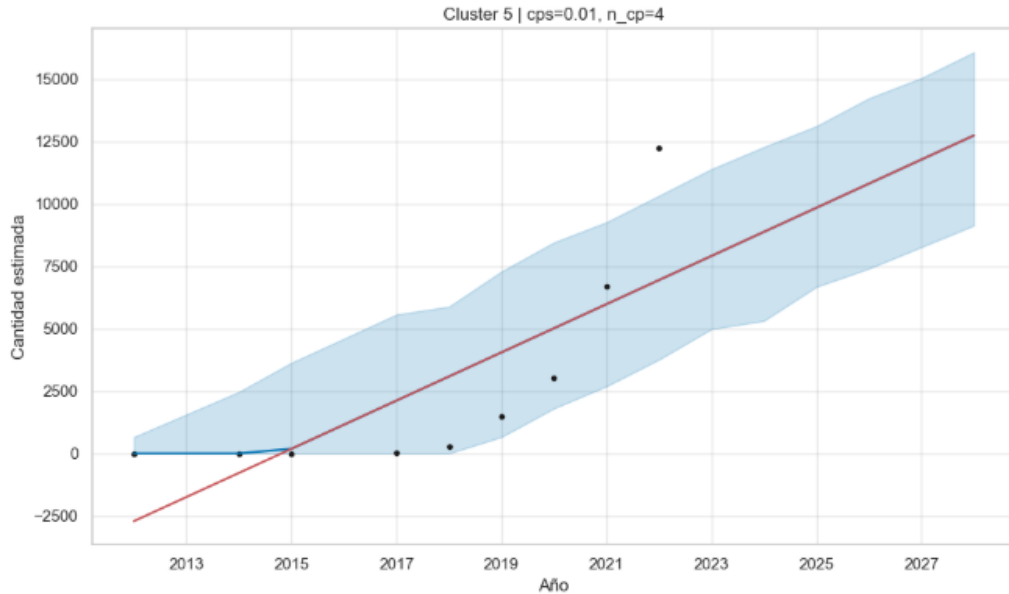


Nota. Elaboración propia.

En el segmento de camionetas híbridas de potencia media con capacidad estándar de pasajeros (clúster 5), según se observa en la figura 48, hay un crecimiento sostenido y acelerado desde 2018, con proyecciones favorables para 2028 de 12747 unidades registradas. Esta tendencia refleja una expansión constante, posicionando a este segmento como uno de los de mayor potencial de desarrollo en el futuro.

Figura 48.

Pronóstico Prophet para subgrupo particular y oficial ligeros segmento 5.



Nota. Elaboración propia.

8.6.3. Proyección del Subgrupo Público Ligero

En el subgrupo de uso público ligero se encuentran los resultados descritos en la tabla 19.

Tabla 19.

Métricas de proyecciones para subgrupo público ligero.

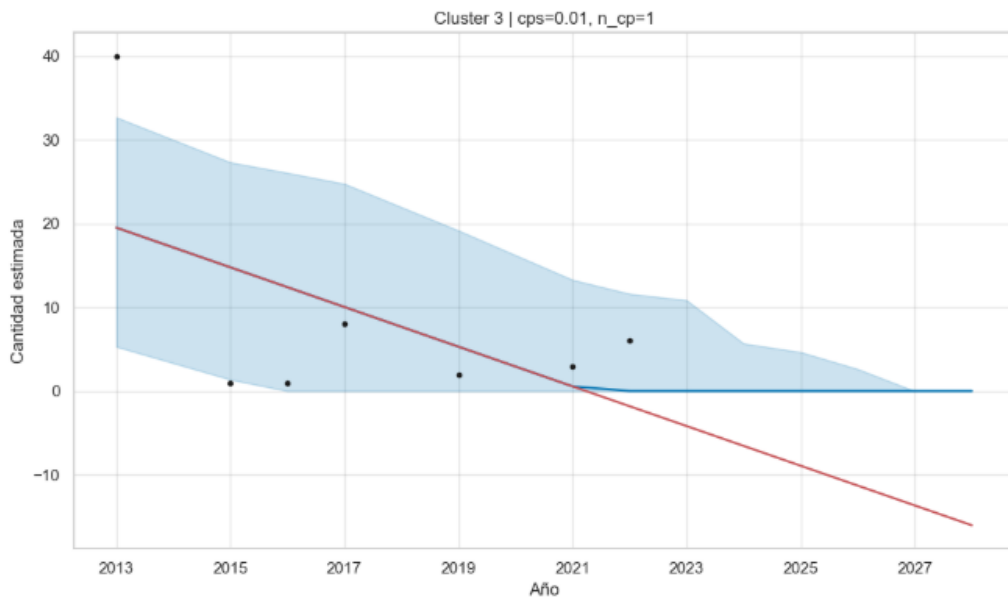
Clúster	Último valor	Último valor	Incertidumbre		MAE de	MAE de
	histórico (unidades registradas)	proyectado (unidades registradas)	promedio (unidades registradas)	Incertidumbre relativa	entrenamiento (unidades registradas)	validación (unidades registradas)
3	0.0	0.0	13.24	NA	8.37	3.67
4	6.49	9.71	6.37	0.66	1.69	4.4

Nota. Elaboración propia.

El segmento de automóviles eléctricos de baja potencia (clúster 3), según la figura 49, proyecta un descenso total en toda la predicción. Posiblemente el segmento esté siendo desplazado por otras alternativas.

Figura 49.

Pronóstico Prophet para subgrupo público ligero segmento 3.

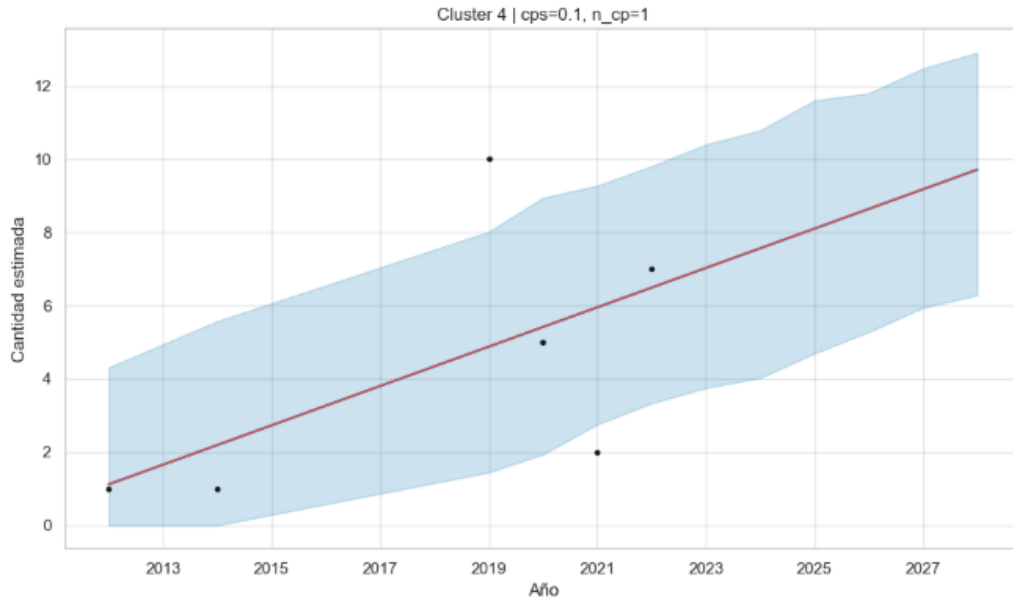


Nota. Elaboración propia.

En el segmento de automóviles eléctricos de media potencia (clúster 4), la figura 50 muestra una tendencia creciente a mediano plazo, aunque es un segmento muy pequeño dentro de su grupo.

Figura 50.

Pronóstico Prophet para subgrupo público ligero segmento 4.



Nota. Elaboración propia.

Los demás segmentos no tienen proyección de Prophet por no tener registros de al menos 5 años. Dentro de estos está el segmento de camionetas eléctricas de marca Dongfeng (clúster 0) que tiene registros únicamente de 2019 a 2022, pero por su comportamiento histórico, parece el más prometedor del grupo de uso público ligero.

8.6.4. Proyección del Subgrupo Público Pesado

En el subgrupo de uso público pesado se encuentran los resultados descritos en la tabla 20.

Tabla 20.

Métricas de proyecciones para subgrupo público pesado.

Clúster	Último valor histórico	Último valor proyectado	Incertidumbre promedio	Incertidumbre relativa	MAE de entrenamiento	MAE de validación

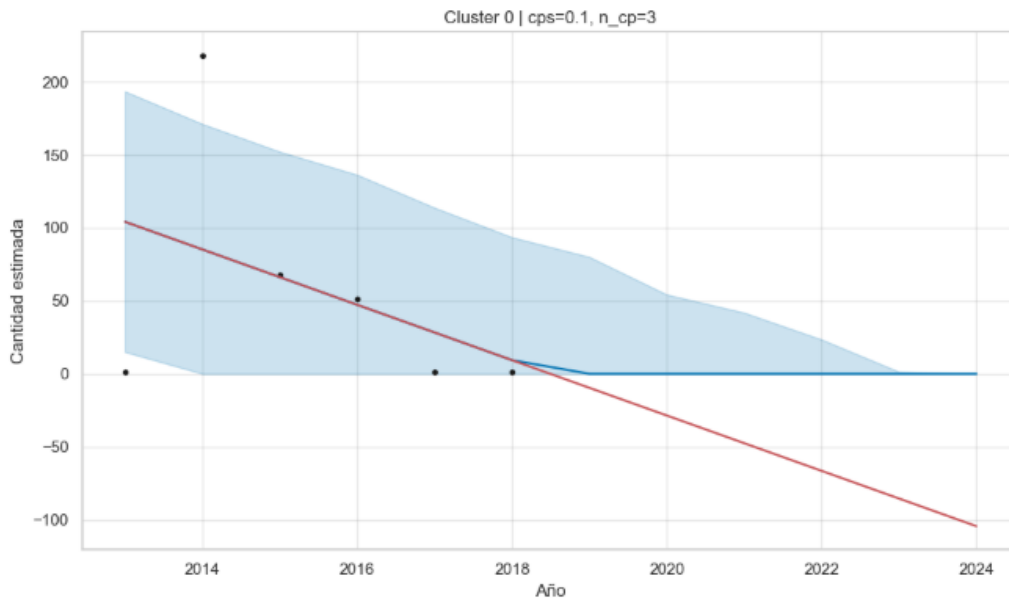
	(unidades registradas)	(unidades registradas)	(unidades registradas)		(unidades registradas)	(unidades registradas)
0	9.31	0.0	87.23	NA	67.27	70.71
1	234.57	390.7	406.12	1.04	26.76	242.45
2	222.1	420.06	219.22	0.52	45.26	188.3

Nota. Elaboración propia.

Para el segmento de buses híbridos con baja potencia y capacidad promedio de pasajeros de 76 (clúster 0), la figura 51 muestra una proyección de comportamiento con tendencia de descenso a cero.

Figura 51.

Pronóstico Prophet para subgrupo público pesado segmento 0.

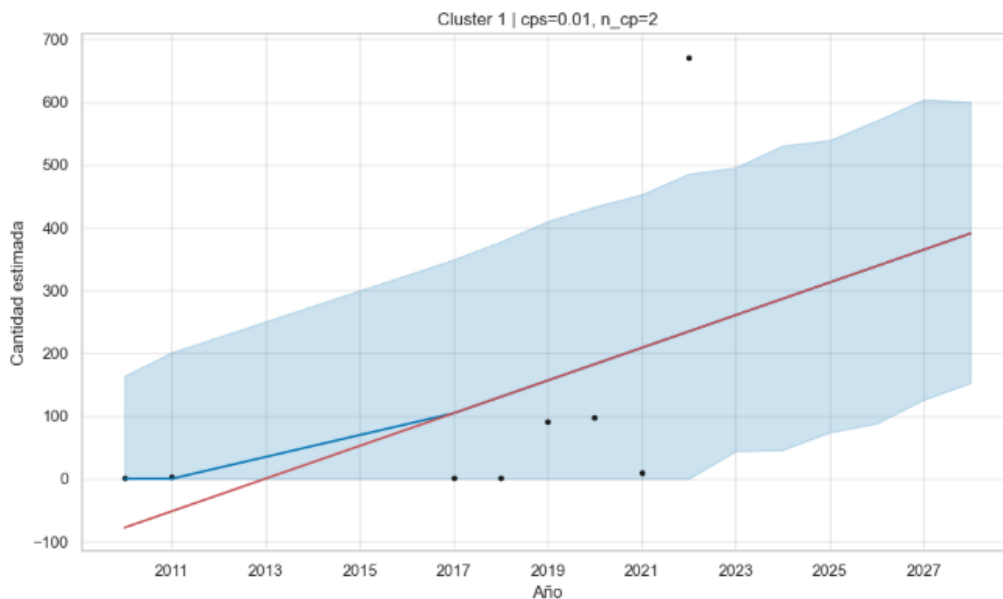


Nota. Elaboración propia.

En el segmento de buses eléctricos con mediana potencia y capacidad promedio de pasajeros de 52 (clúster 1), según se observa en la figura 52, se proyecta un crecimiento sostenido hasta los 391 registros en 2028, pero con bastante incertidumbre relativa (1.04).

Figura 52.

Pronóstico Prophet para subgrupo público pesado segmento 1.

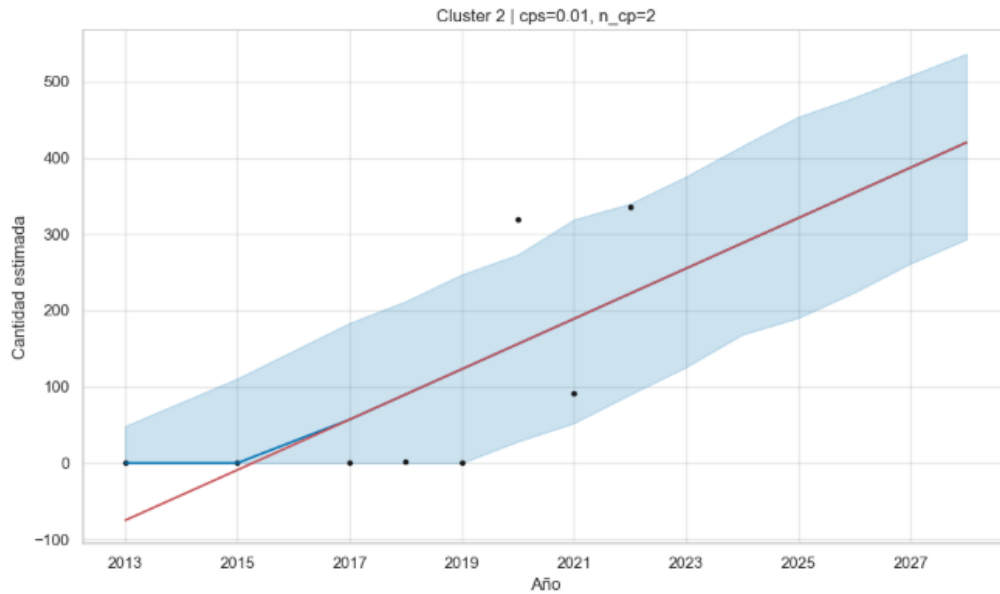


Nota. Elaboración propia.

Como se muestra en la figura 53, la proyección para el segmento de buses eléctricos de alta potencia (clúster 2) tiene una incertidumbre relativa media (0.52) y muestra un crecimiento sostenido hasta 2028 con 420 registros en ese año. Este es un segmento impulsado por políticas públicas que seguramente seguirán apoyando su consolidación.

Figura 53.

Pronóstico Prophet para subgrupo público pesado segmento 2.



Nota. Elaboración propia.

8.6.5. Proyección del Subgrupo Carga

En el subgrupo de carga se encuentran los resultados descritos en la tabla 21.

Tabla 21.

Métricas de proyecciones para subgrupo carga.

	Último valor histórico (unidades registradas)	Último valor proyectado (unidades registradas)	Incertidumbre promedio (unidades registradas)	Incertidumbre relativa	MAE de entrenamiento (unidades registradas)	MAE de validación (unidades registradas)
0	18.72	29.78	9.88	0.33	2.1	6.89
1	48.19	121.94	17.51	0.14	3.57	16.76
2	40.04	56.3	33.92	0.6	11.01	13.37

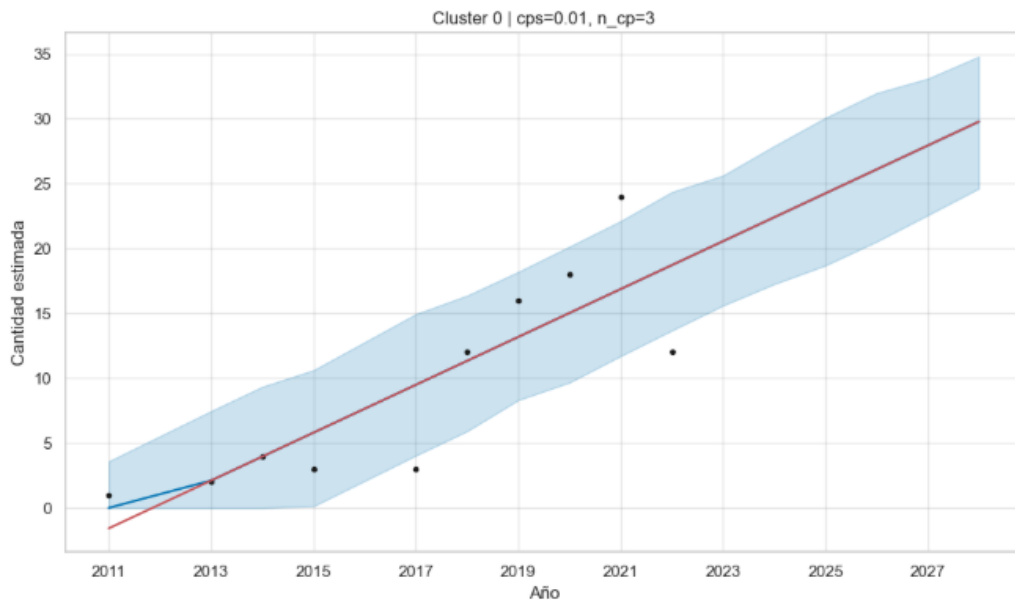
3 118.96 180.88 42.59 0.24 5.83 27.42

Nota. Elaboración propia.

Para el segmento de camiones híbridos con media capacidad de carga (clúster 0), en la figura 54 se evidencia un crecimiento hacia 30 registros en 2028. La incertidumbre relativa es de 0.33, lo que da fiabilidad sobre esta proyección. Este es un segmento incipiente, con buenas perspectivas de crecimiento.

Figura 54.

Pronóstico Prophet para subgrupo carga segmento 0.

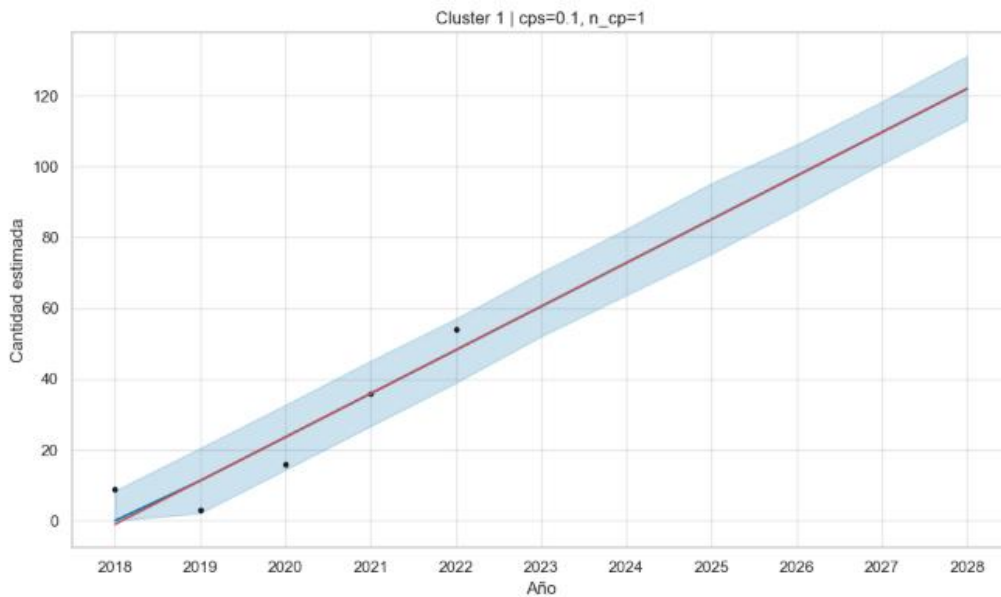


Nota. Elaboración propia.

En el segmento de camiones eléctricos con media capacidad de carga (clúster 1), la figura 55 expone una proyección de crecimiento positiva con una banda de incertidumbre pequeña (incertidumbre relativa de 0.14). Parece un segmento con comportamiento prometedor a mediano plazo.

Figura 55.

Pronóstico Prophet para subgrupo carga segmento 1.

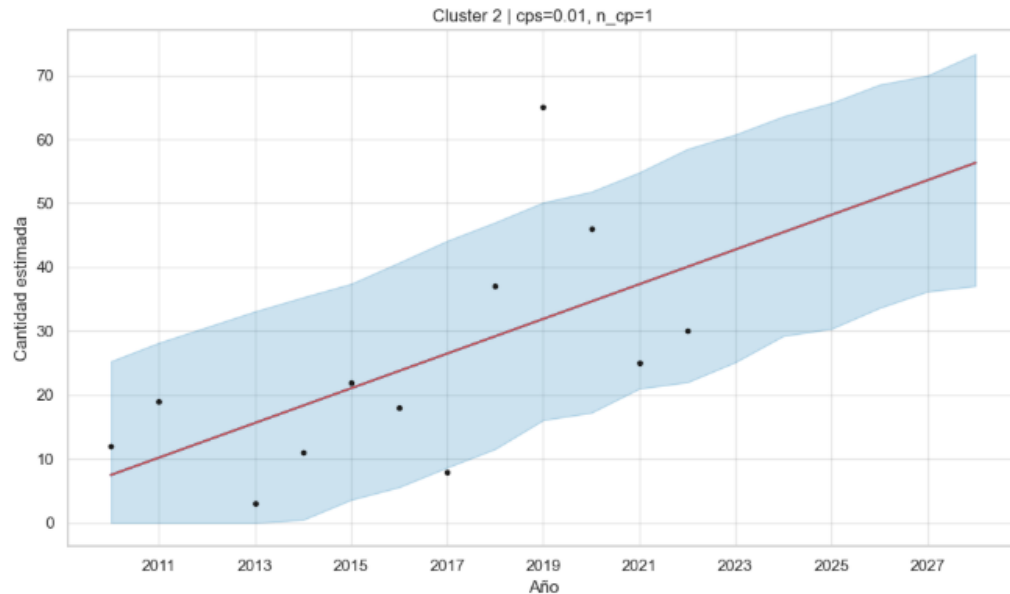


Nota. Elaboración propia.

Como se aprecia en la figura 56, el segmento de camiones eléctricos con baja capacidad de carga (clúster 2) ha tenido fluctuaciones interanuales, pero en general muestra una proyección positiva hacia 2028 con una incertidumbre relativa media de 0.6 y 56 registros esperados para ese año.

Figura 56.

Pronóstico Prophet para subgrupo carga segmento 2.

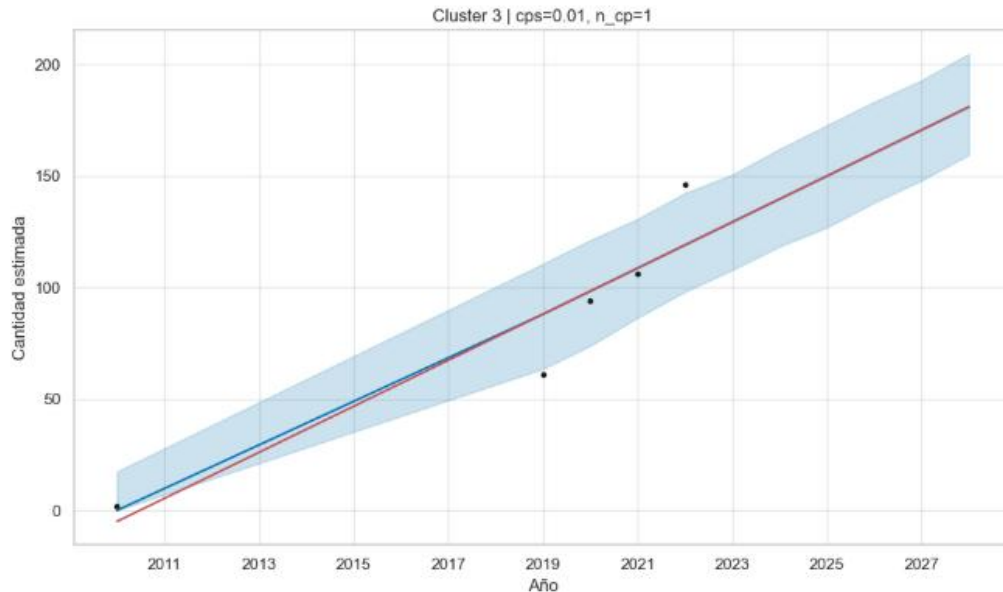


Nota. Elaboración propia.

En el segmento de camiones eléctricos con alta capacidad de carga (clúster 3), la figura 57 muestra una tendencia creciente con incertidumbre relativa de 0.24 y 181 registros en 2028. Se pronostica que continuará siendo una parte importante del grupo de carga.

Figura 57.

Pronóstico Prophet para subgrupo carga segmento 3.



Nota. Elaboración propia.

8.7. Análisis Enriquecido con Variables Socioeconómicas y de Infraestructura

Posteriormente, se avanzó hacia un análisis enriquecido incorporando variables socioeconómicas para cada departamento, con el objetivo de identificar las que mayor efecto tienen en la adopción de vehículos eléctricos e híbridos en Colombia utilizando análisis de valores SHAP y de Componentes Principales (PCA) con Biplot. Para ello, se construyeron grupos de registros que toman como base cada subgrupo utilizado en la segmentación (motos, uso particular y oficial ligeros, uso público ligeros, uso público pesados y uso de carga). Cada uno pasó por un proceso en el que se consolidó el número de registros por departamento en una nueva columna mediante una agrupación.

El nombre de los departamentos fue normalizado, retirando caracteres especiales y unificando versiones de nombres como las muchas variaciones del Archipiélago de San Andrés, Providencia y Santa Catalina, para poder integrar el código DIVIPOLA (División Político-Administrativa de Colombia) desde el grupo de registro de estos códigos del

DANE. Con este código integrado a los subgrupos y teniendo en cuenta que todos los grupos de registros de variables socioeconómicas y de infraestructura tienen el código DIVIPOLA y/o el nombre de departamento normalizado, se procedió a integrar la información utilizando alguno de los dos como llave para la operación de unificación.

Como se ilustró anteriormente en la figura 4, las variables externas integradas fueron: PIB per cápita, indicador del nivel de desarrollo económico, para detectar si los departamentos con mayor capacidad adquisitiva lideran la adopción; matrículas de educación superior, para evaluar la relación de la adopción con el nivel educativo de la región; índice de brecha digital, para ver cómo el acceso a herramientas TIC puede ser barrera o facilitador en la adopción de opciones sostenibles de movilidad; tasa de desocupación, para evaluar si la estabilidad económica del territorio influye en la adopción; población total y porcentaje de población urbana, que puede ayudar a identificar si territorios densamente poblados o urbanizados tienen mayor adopción dado que la movilidad eléctrica suele estar más enfocada a estos entornos, y número de puntos de carga, que es una variable clave de infraestructura para la transición energética, recopilada manualmente a partir de mapas alimentados por la comunidad en varias aplicaciones.

Cada subgrupo para análisis socioeconómico y de infraestructura departamental quedó conformado con la estructura descrita en la tabla 22.

Tabla 22.

Estructura de datasets de análisis socioeconómico y de infraestructura departamental.

Columna	Tipo
Departamento_Normalizado	object

Codigo_Depto	int64
Cantidad_Registros	int64
PIB_per_capita	int64
Matriculados_Edu_Superior	int64
IBD	float64
Tasa_Desocupación	float64
Porcentaje_Urbano	float64
Población_Total	int64
Numero_Puntos_Carga	int64

Nota. Elaboración propia.

Los resultados del análisis integrado utilizando análisis de valores SHAP y de Componentes Principales (PCA) con Biplot se presentan en las siguientes secciones.

8.7.1. Análisis de Valores SHAP

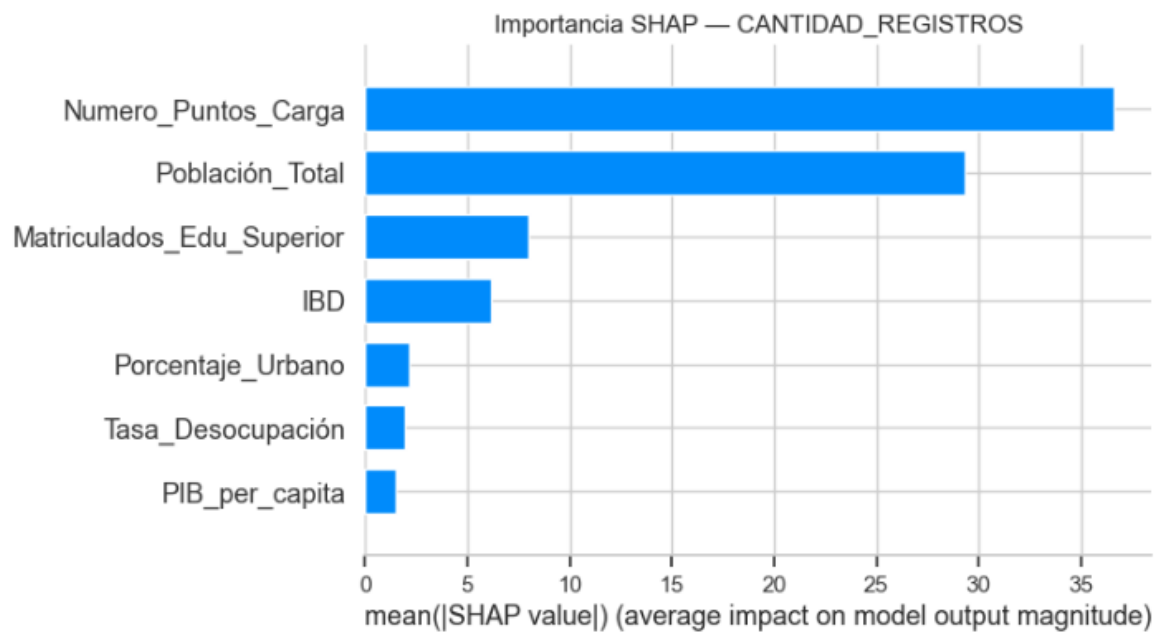
Con esta información, se aplicó un análisis con SHAP (SHapley Additive exPlanations) sobre cada subgrupo, utilizando un modelo tipo RandomForestRegressor para estimar qué variables socioeconómicas tienen mayor influencia en la adopción. SHAP permite interpretar el modelo de regresión y visualizar el aporte individual positivo o negativo de cada variable sobre los registros de vehículos eléctricos e híbridos, de esta forma se logró obtener una explicación de los resultados que pueda ayudar a los diferentes actores a generar accionables.

En los gráficos SHAP de todos los subgrupos excepto público pesado, mostrados en las figuras 58, 59, 60 y 62, las variables más influyentes son el número de puntos de carga y la población total, lo que indica que el nivel de desarrollo de la infraestructura de

carga eléctrica de cada departamento condiciona la decisión de optar por opciones eléctricas e híbridas de movilidad, y que a su vez existe una correlación positiva entre la población total y la adopción, como es previsible en mercados donde la demanda se concentra en territorios más poblados.

Figura 58.

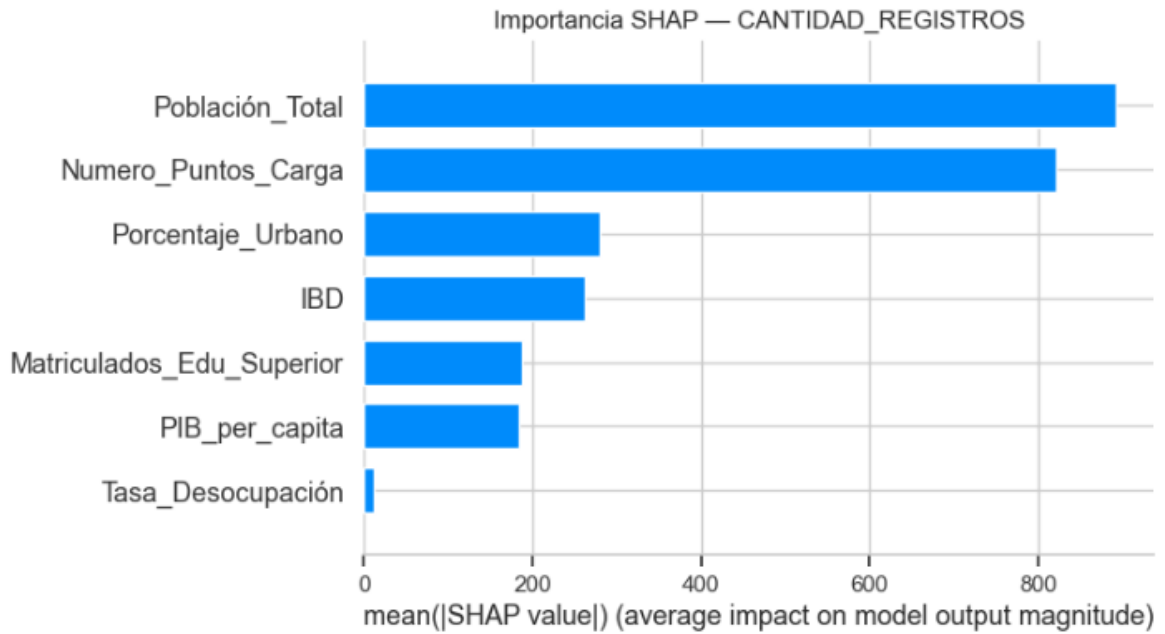
Análisis con SHAP para el subgrupo motos.



Nota. Elaboración propia.

Figura 59.

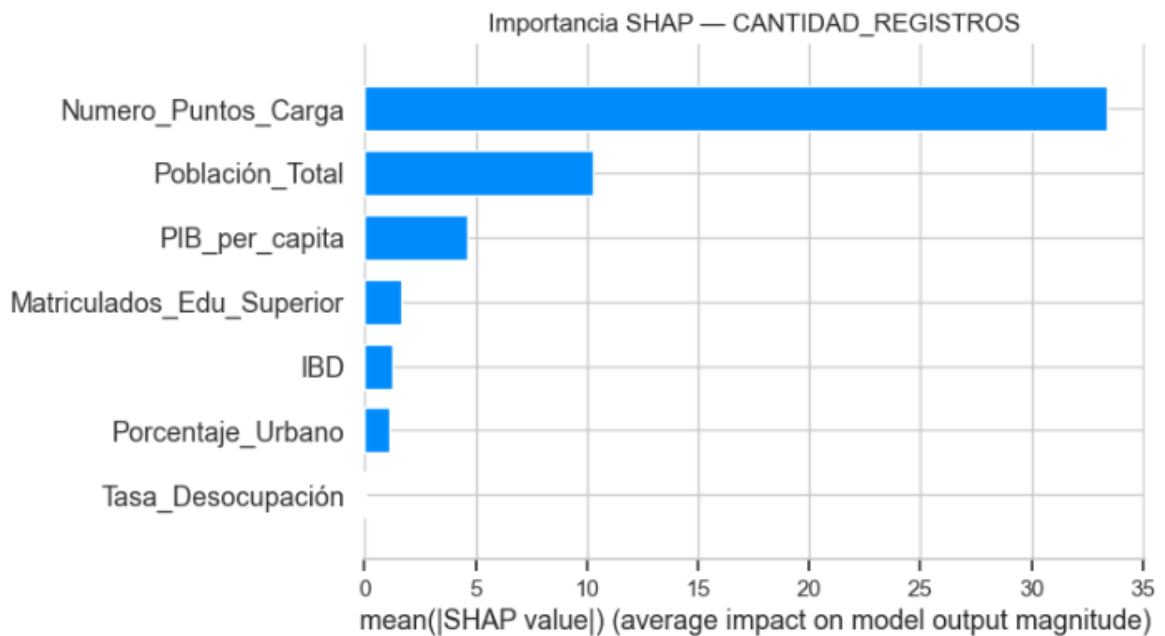
Análisis con SHAP para el subgrupo particular y oficial ligero.



Nota. Elaboración propia.

Figura 60.

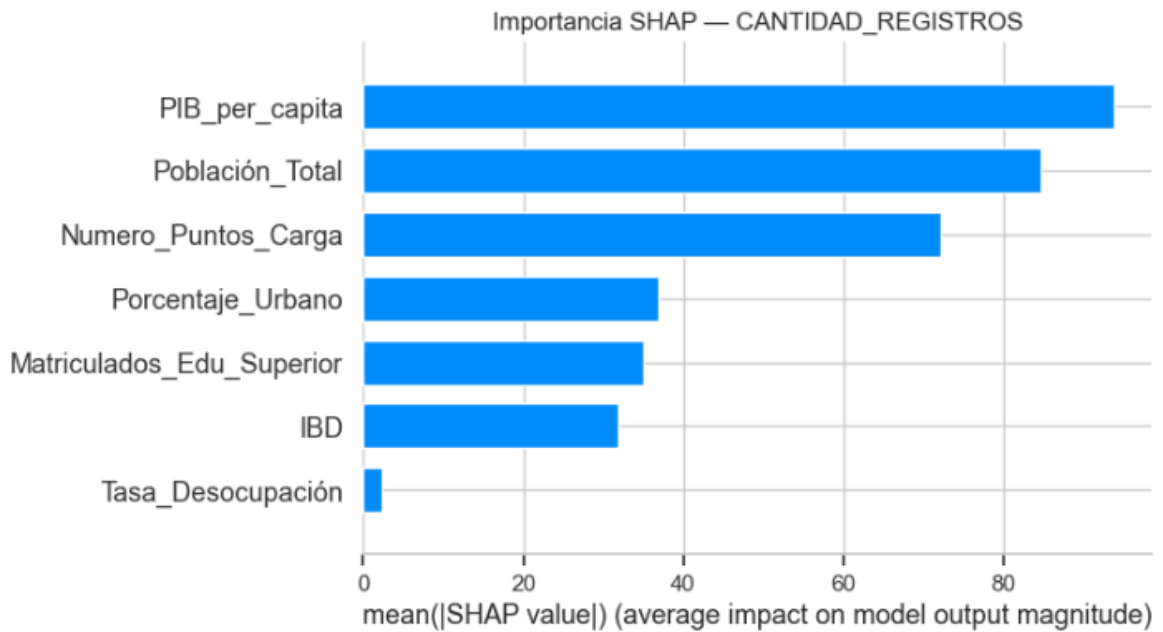
Análisis con SHAP para el subgrupo público ligero.



Nota. Elaboración propia.

Figura 61.

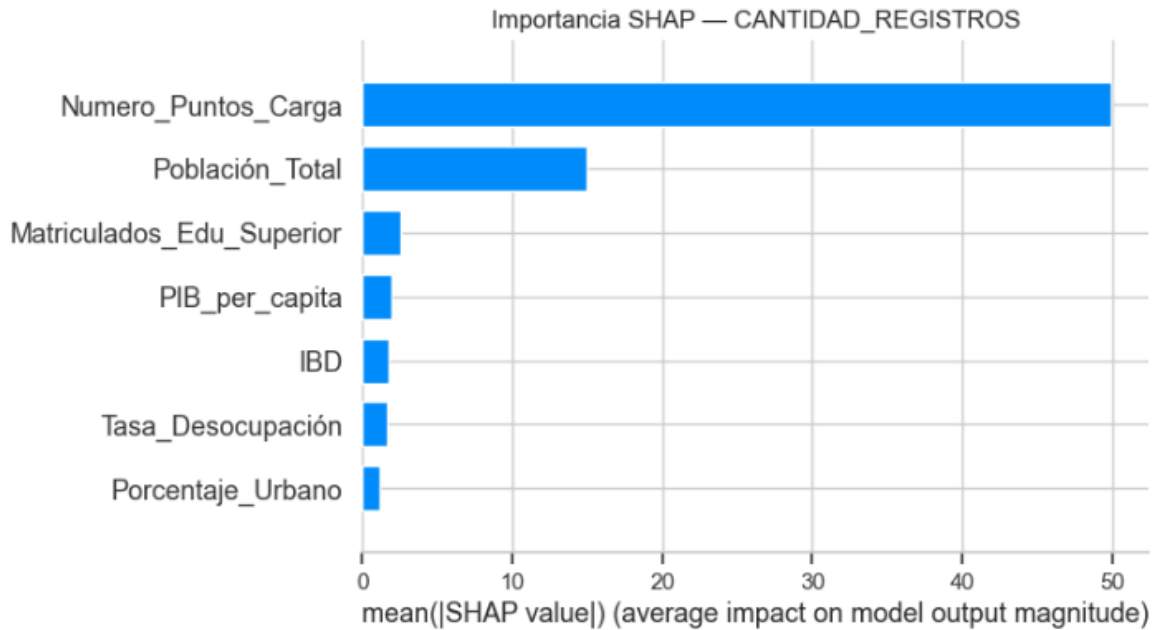
Análisis con SHAP para el subgrupo público pesado.



Nota. Elaboración propia.

Figura 62.

Análisis con SHAP para el subgrupo carga.



Nota. Elaboración propia.

Destaca que, para el subgrupo de público pesado que se muestra en la figura 61, el PIB per cápita es la variable más influyente. Esto puede estar vinculado a que las regiones con mayor poder económico tienen la posibilidad de ser las primeras en iniciar proyectos a gran escala de transformación tecnológica, como lo hace Bogotá D.C., que es la líder absoluta en este subgrupo. Por otro lado, las variables del sector educativo y laboral parecen no tener tanto impacto en la adopción de ningún subgrupo.

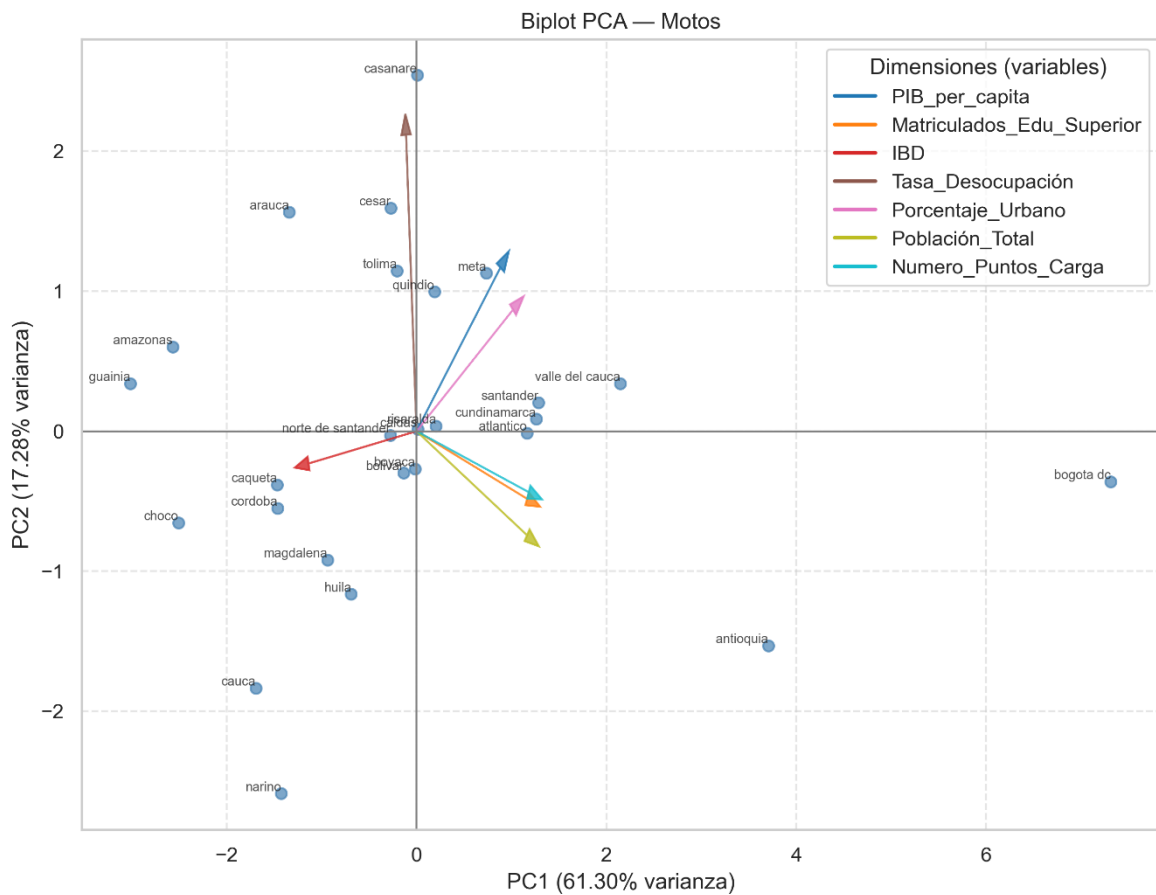
8.7.2. Análisis de Componentes Principales con Biplot

Este análisis se complementó con gráficos PCA biplot, que reducen la dimensionalidad del conjunto de variables y permiten visualizar los departamentos y las variables en un mismo plano. Gracias a esto, se logra una visión general sobre cómo se agrupa la adopción por territorios y qué variables impulsan más este proceso, permitiendo la formación de perfiles regionales.

Como se aprecia la figura 63, para el subgrupo de motos, PC1 y PC2 explican el 78.58 % de la varianza total, lo que indica que el modelo recoge la mayor parte de la información relevante de los datos. De igual forma sucedió en las demás gráficas 64 a 67, en las que ambos componentes principales explican por lo menos el 70% de la varianza total.

Figura 63.

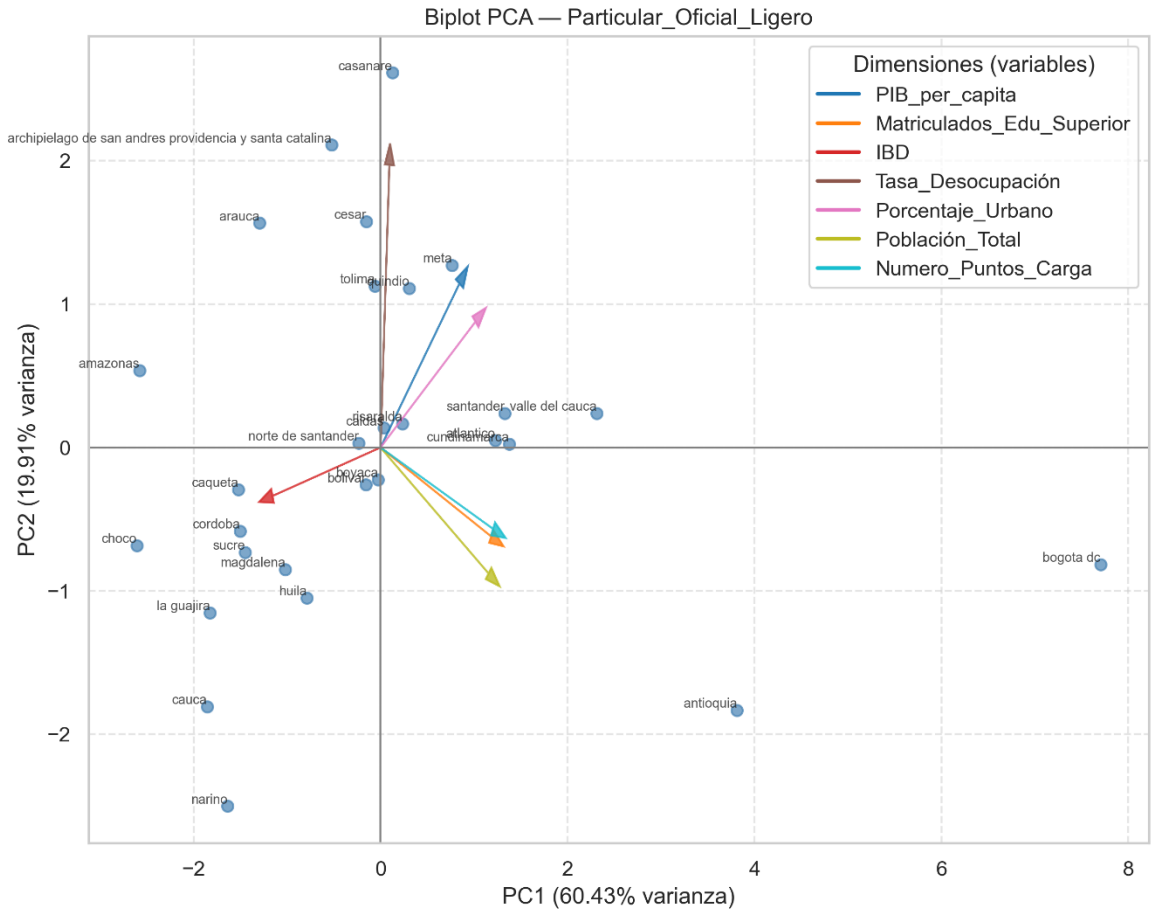
PCA Biplot para el subgrupo motos.



Nota. Elaboración propia.

Figura 64.

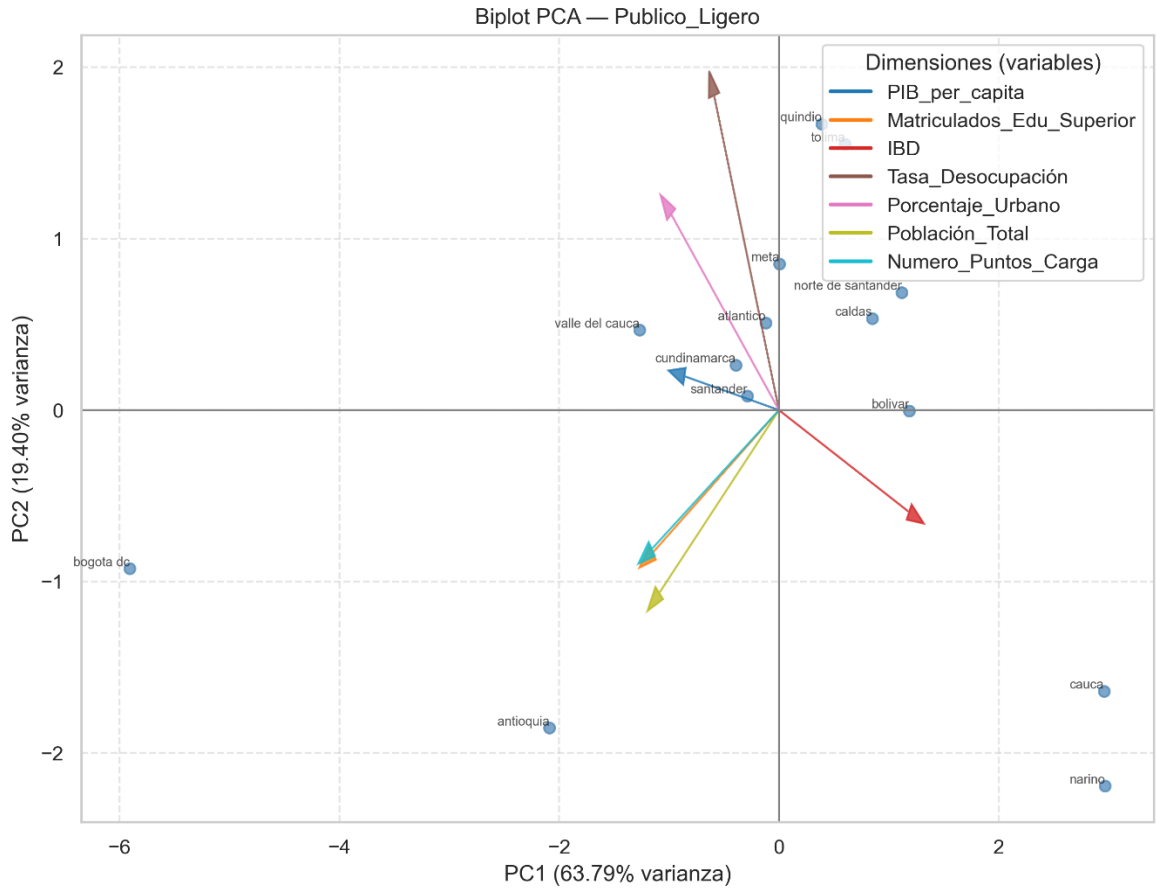
PCA Biplot para el subgrupo particular y oficial ligero.



Nota. Elaboración propia.

Figura 65.

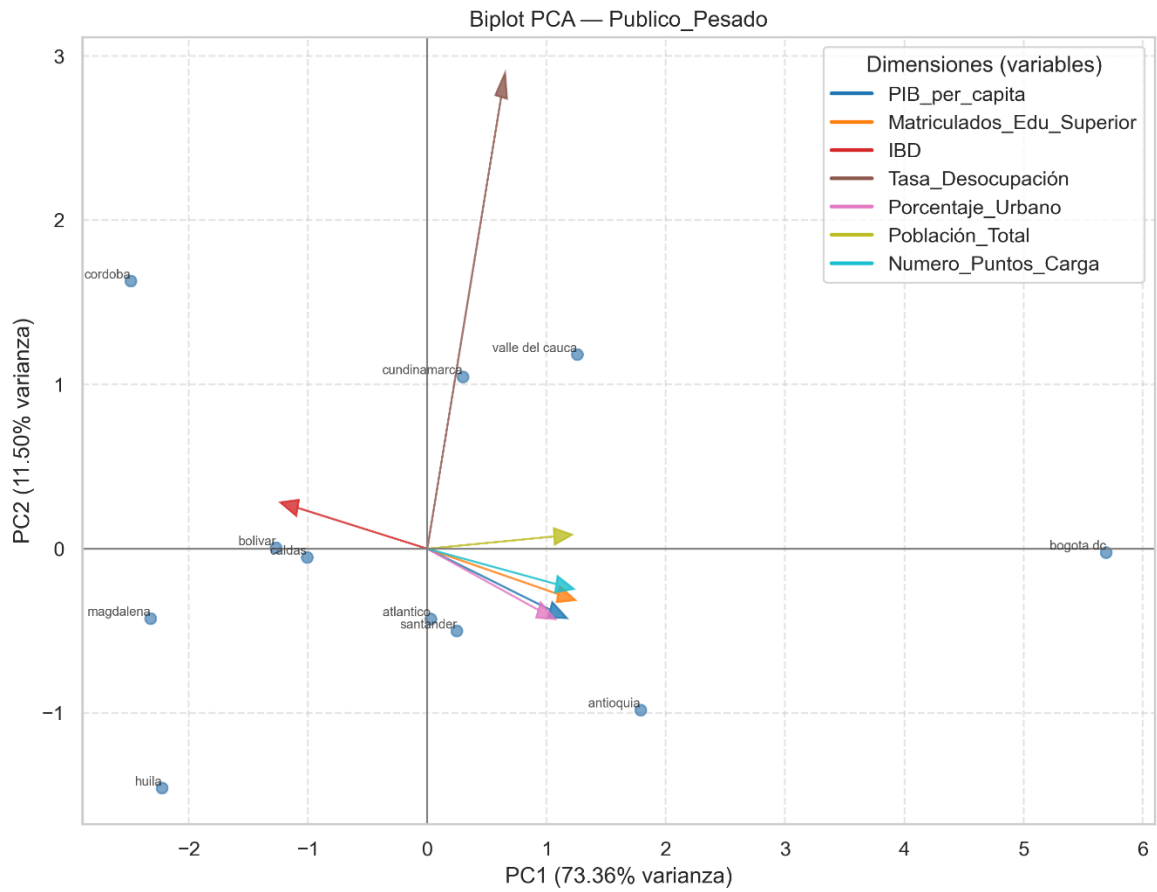
PCA Biplot para el subgrupo público ligero.



Nota. Elaboración propia.

Figura 66.

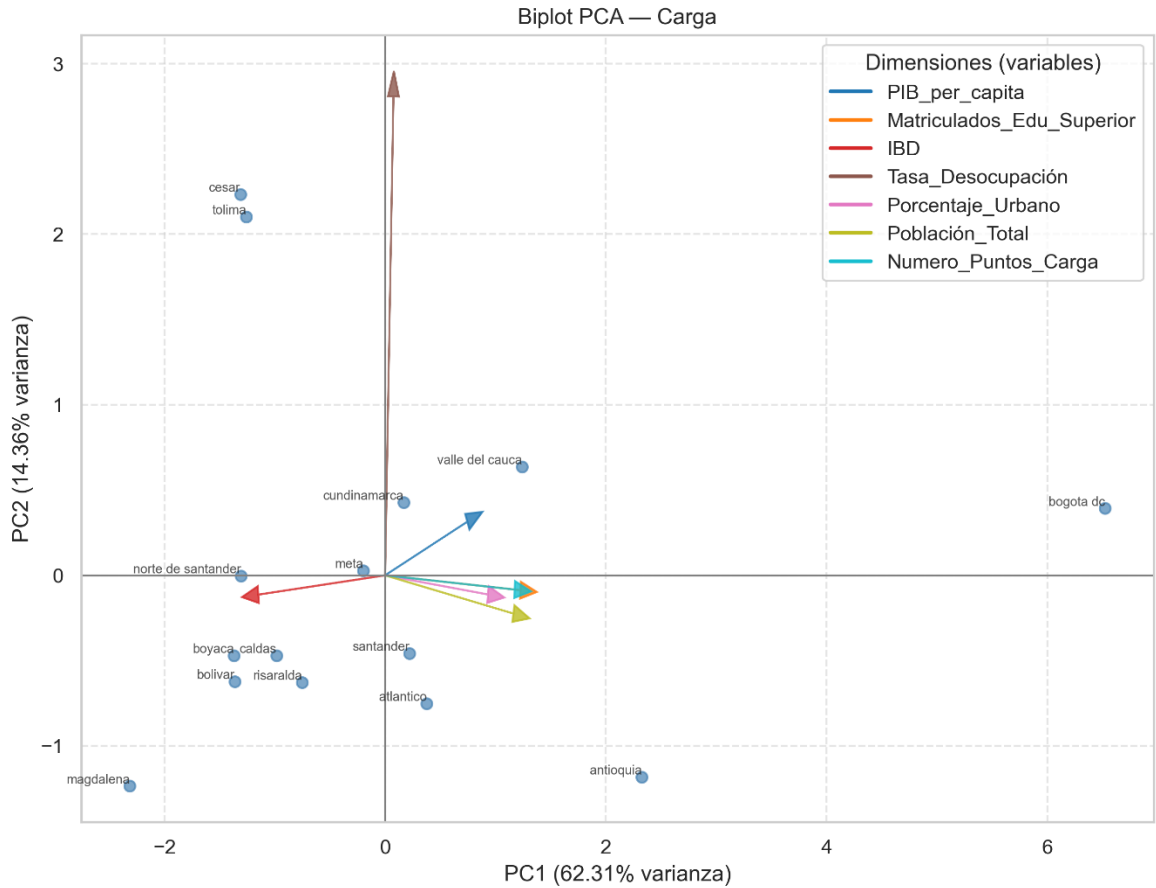
PCA Biplot para el subgrupo público pesado.



Nota. Elaboración propia.

Figura 67.

PCA Biplot para el subgrupo carga.



Nota. Elaboración propia.

En las gráficas de los cinco subgrupos (figuras 63 a 67) se observa que Bogotá D.C. y los departamentos con mayor adopción se ubican en un sector extremo del plano. Las variables originales que más se relacionan a la ubicación en este sector son población total, número de puntos de carga y matrículas de educación superior. El PIB per cápita y el porcentaje de población urbana también tienen un efecto importante a considerar en los subgrupos de público pesado y carga. Por el contrario, la variable que más aleja a los departamentos del grupo de líderes en adopción es el Índice de Brecha Digital.

De esta forma, se identificaron correlaciones positivas y negativas entre las variables consideradas en el estudio y las regiones de mayor adopción de movilidad eléctrica e híbrida.

8.8. Verificación de Hipótesis y Propuestas Sobre los Resultados

El presente análisis permite confirmar la hipótesis de que la adopción de vehículos eléctricos e híbridos en Colombia se concentra en las regiones con mayor poder adquisitivo y desarrollo económico. Tal como se puede ver en las gráficas 13 a 17, Bogotá, Antioquia y Cundinamarca lideran la adopción tecnológica. Asimismo, los resultados obtenidos con SHAP y PCA Biplot muestran la separación marcada entre los departamentos dominantes y el resto en factores como el PIB per cápita, la densidad urbana y la disponibilidad de puntos de carga que presentan una relación positiva con la cantidad de registros de vehículos eléctricos e híbridos. Estos resultados se alinean con el análisis de Salazar-Serna et al. (2024) que expone esta misma disparidad en países en vía de desarrollo, en donde se evidencia la distribución desigual en infraestructura e incentivos, los cuales se concentran en los centros urbanos más importantes. Por su parte, también afirman el estudio de Soto et al. (2021) que señala la relevancia del PIB como un factor que influencia la capacidad gubernamental para implementar medidas y proyectos que faciliten la adopción.

Como lo sugiere Díaz Merchán et al. (2015) para reducir esta brecha es importante implementar políticas y estrategias de inversión dirigidas específicamente a las necesidades y retos de las zonas más atrasadas en la adopción. En línea con el presente análisis de las variables más influyentes, es necesario fortalecer la infraestructura de carga en departamentos intermedios y con mayoría rural, así como aumentar las políticas de

incentivo económico que impulsen el cambio del parque automotor de energía 100% combustible a opciones híbridas y eléctricas. También son necesarias alianzas público-privadas que ayuden a expandir la red de puntos de carga, a financiar y facilitar la adquisición de nuevas unidades de este tipo de vehículos y ofrecer beneficios tributarios o exenciones a medidas como el pico y placa en todo el país. Gobierno y empresas privadas pueden utilizar estos mecanismos para equilibrar la adopción tecnológica y apuntar a una transición más rápida y equitativa en todas las regiones de Colombia.

Respecto a la segunda hipótesis, Jauregui Romero et al. (2024) resalta que los algoritmos de aprendizaje automático permiten encontrar patrones y correlaciones, que no son fácilmente detectables. Esta ventaja se aprovechó en el presente estudio demostrando que la utilización de técnicas de segmentación no supervisada de aprendizaje automático permite la detección de segmentos con características particulares, que puede ayudar a identificar nichos y segmentos en crecimiento. Mediante la aplicación de los algoritmos k-prototypes, agrupación jerárquica y CatBoost embeddings con minibatch k-means se pudo perfilar nichos y segmentos interesantes como el de motocicletas eléctricas de altísima potencia, camionetas híbridas de potencia media con capacidad de 5 pasajeros para uso particular, camionetas eléctricas Dongfeng para uso público, buses eléctricos con capacidad aproximada de 50 pasajeros y camiones eléctricos con alta capacidad de carga. El conocimiento a fondo de estos segmentos puede ayudar a las empresas a ajustar su estrategia comercial y a los actores del sector público a encaminar políticas de acuerdo a la demanda y características de su región, como también lo aconseja Díaz Merchán et al. (2015) en su estudio.

9. Discusión

En el mercado colombiano la clase más común dentro de la flota eléctrica e híbrida es la camioneta. Esto puede estar relacionado con la versatilidad que ofrece este tipo de vehículos, que tiene usos tanto familiares como en actividades productivas (Cabarrubias De La Cruz & Santos Tolentino, 2023). Las condiciones de la malla vial del país pueden condicionar también la predilección por las camionetas, dado que son más adecuadas para vías en mal estado y terrenos rurales debido a características técnicas como la suspensión y la tracción. Asimismo, como explica Soto et al. (2021), la propiedad de este tipo de vehículos proyecta percepción de estatus y seguridad, que es un factor significativo en países en vía de desarrollo como Colombia, en el que los fabricantes han implementado esta imagen en su estrategia comercial.

Se identificó también que los colombianos tienen predilección por los vehículos híbridos sobre los eléctricos, posiblemente por factores como el costo inicial menor y la conveniencia al no depender totalmente de una infraestructura de carga en desarrollo (Espinosa Ramírez & Ozgur Kayalica, 2023).

Respecto al análisis de marcas dominantes en cada subgrupo estudiado, se encontraron varios casos de marcas totalmente dominantes y de algunos duopolios. Uy et al. (2024) en su análisis del mercado filipino resaltó que esta concentración de mercado es especialmente pronunciada en naciones en vía de desarrollo, donde los compradores confían en marcas establecidas y en redes de mantenimiento disponibles para adquirir nuevas tecnologías.

En relación con los resultados de la segmentación, en el subgrupo de motos, los segmentos de baja y media potencia son los más populares. Esta tendencia es relevante

en el contexto colombiano, donde, como lo señala Cadavid & Salazar-Serna (2021), las motocicletas tienen un papel importante que cubre vacíos provocados por deficiencias en los sistemas de transporte público formales y una planificación urbana insuficiente.

En el grupo de uso particular y oficial ligeros, las camionetas híbridas de capacidad estándar (5 pasajeros) de media y alta potencia, y los automóviles híbridos dominan el mercado, probablemente por la popularidad de este tipo de vehículos que como explica Soto et al. (2021), puede estar basada en la prioridad que le dan los usuarios a la versatilidad, practicidad y seguridad. En línea con lo anterior, para el grupo de uso público ligero, las camionetas eléctricas de capacidad estándar (5 pasajeros) representan el segmento con mayor crecimiento, impulsado también por los beneficios técnicos expuestos anteriormente.

Respecto al grupo de uso público pesado, los buses eléctricos con capacidad aproximada de 50 pasajeros presentan un crecimiento importante, mientras los buses híbridos han tenido una caída pronunciada. Esto va en línea con el análisis realizado en países en vía desarrollo de Shamsuddoha & Nasir (2025), en el que el sector público está llevando a cabo importantes iniciativas de descarbonización y mitigación de la contaminación del aire en zonas densamente pobladas.

Por su parte en la flotilla de carga, el segmento de camiones de alta capacidad es el dominante. Esto puede estar influenciado por incentivos públicos y metas a nivel corporativo del sector logístico (Soto et al., 2021).

En cuanto al análisis de variables que potencian la transición energética en la movilidad de Colombia, se identifica que el número de puntos de carga es la variable más influyente en la adopción en la mayoría de subgrupos. Tal como lo expone Díaz Merchán

et al. (2015), la implementación de una red de puntos de carga a nivel nacional es crucial para impulsar la adopción masiva de alternativas de movilidad sostenible en Colombia.

El PIB per cápita tiene influencia considerable en la adopción del subgrupo de uso público pesado. Esto también fue resaltado por Espinosa Ramírez & Ozgur Kayalica (2023), explicando que este indicador económico se relaciona directamente con la capacidad de una unidad territorial para invertir en infraestructura y dar incentivos a la adopción de tecnologías sostenibles.

En la misma línea, se evidencia que los departamentos con mayor adopción son los que están en mejores condiciones socioeconómicas y por tanto el avance en la transición energética está muy centralizado. Esto se articula con el análisis de Soto et al. (2021) que en su estudio señala que esta concentración se alinea con inequidades socioeconómicas como la propiedad de vehículos que es más común en clases medias y altas, que suelen residir en las zonas más desarrolladas.

10. Conclusiones y Trabajo Futuro

10.1. Conclusiones

La revisión de literatura permitió conocer enfoques y metodologías utilizadas por varios autores para el análisis de la adopción de vehículos eléctricos e híbridos en diferentes contextos. Esta recopilación de información sirvió como base teórica para formular el presente análisis en el contexto colombiano. Uno de los puntos clave extraídos de esta investigación y utilizados posteriormente fue el aprovechamiento de variables de diferentes campos que permiten encontrar factores explicativos de la adopción y entender mejor la dinámica de la transición a la movilidad sostenible.

Asimismo, se pudo identificar un vacío en estudios multivariados sobre este proceso de adopción en Colombia impulsados por segmentación no supervisada, métodos explicativos y proyecciones con aprendizaje automático, que permitieran tener un panorama acorde a las características de los datos disponibles del mercado nacional.

Con el pipeline desarrollado para realizar este análisis se pudieron corregir inconsistencias en la información y estandarizar variables, lo que garantizó un grupo de datos confiables para llevar a cabo el estudio. Además, la aplicación de técnicas de segmentación no supervisada, junto con la integración de variables socioeconómicas y de infraestructura, permitió identificar patrones relevantes, algunos de los cuales se comentan a continuación.

Se detectó que las camionetas son el tipo de vehículo más representativo en la flota eléctrica e híbrida de Colombia, lo que evidencia una acogida en el mercado local y una proyección positiva y sostenida. De aquí se puede inferir que el mercado de nuestro país prioriza opciones con mayor capacidad, buena autonomía y versatilidad, lo que puede

enfocar estrategias comerciales de fabricantes e importadores, y decisiones gubernamentales sobre este segmento.

Además, el análisis mostró que los vehículos híbridos tienen un crecimiento mayor que los eléctricos puros en Colombia. Sin embargo, desde el punto de vista ambiental, los vehículos eléctricos ofrecen una reducción mayor de emisiones y un impulso más fuerte a la descarbonización del transporte. Por tanto, las políticas públicas deberían priorizar el fortalecimiento de la infraestructura de carga y los incentivos exclusivos para apoyar la adopción de vehículos 100% eléctricos.

Es importante señalar que los resultados evidencian una brecha regional marcada en la adopción de vehículos eléctricos e híbridos en Colombia. Regiones como Bogotá y Antioquia concentran la mayor parte de los registros, mientras que zonas muy amplias del país presentan una adopción incipiente o incluso nula. Esta diferencia sugiere que las políticas de movilidad sostenible y la transición tecnológica están fuertemente centralizadas, por lo que se hace evidente la necesidad de estrategias diferenciadas de acuerdo al contexto socioeconómico y territorial de cada región.

Asimismo, utilizando técnicas de segmentación no supervisada de aprendizaje automático, se identificaron grupos de vehículos con características similares dentro de cada subgrupo analizado. Las proyecciones realizadas con Prophet hasta 2028 indican que hay varios segmentos con alto potencial de crecimiento, lo cual abre oportunidades de negocio para empresas del sector automotor y para gobiernos de diferente orden interesados en impulsar la movilidad sostenible. Estas proyecciones deben considerarse prudentemente teniendo en cuenta la incertidumbre y limitaciones de información explicadas anteriormente.

Finalmente, el análisis de variables socioeconómicas en Colombia encontró que, en la adopción de vehículos eléctricos e híbridos por parte de particulares, la infraestructura de carga y la población total son los principales impulsores. En el sector público, el PIB per cápita y la infraestructura de carga son las variables más influyentes. Estos hallazgos pueden servir de base para el diseño de políticas públicas y planificación a la transición energética.

10.2. Trabajo futuro

El presente estudio enfrentó una serie de limitaciones que afectan los resultados obtenidos y su aplicabilidad. Estas limitaciones están principalmente relacionadas a la calidad, disponibilidad y estandarización de los datos analizados.

La principal limitación fue el acceso limitado a datos actualizados, ya que la información disponible solo llega hasta 2022, lo que no permite capturar cambios y tendencias recientes. Por tanto, los resultados obtenidos son un análisis histórico y exploratorio más que una representación de la situación actual.

Otra limitación del estudio está relacionada con la calidad de los registros originales, en los que se identificaron inconsistencias y valores erróneos en variables técnicas, lo que obligó a realizar un proceso de corrección y estandarización mediante un diccionario técnico construido manualmente. Aunque este procedimiento permitió mejorar la coherencia e integridad de los datos, conlleva un margen de error y pérdida de precisión en algunos registros debido a la imputación de valores aproximados cuando la información técnica no estaba disponible.

A pesar de las limitaciones expuestas, este estudio brinda un aporte metodológico y exploratorio en la integración de información dispersa en diferentes bases de datos,

construcción y estandarización de un diccionario técnico, procesamiento de variables técnicas y socioeconómicas, y aplicación de modelos de aprendizaje automático en el entendimiento integral de un proceso de transformación tecnológica en el entorno de un país en vía de desarrollo.

Asimismo, la descripción de las limitaciones enfrentadas refuerza la necesidad de seguir consolidando fuentes de datos de libre acceso que mantengan sus registros actualizados para el análisis del parque automotor en el futuro con resultados cada vez más precisos. Este trabajo puede ser tomado como punto de partida para el análisis integral de la adopción de movilidad sostenible en Colombia y su metodología puede aplicarse al estudio de dinámicas de transformación en otras áreas.

Como línea de trabajo futuro, el pipeline propuesto es apto para la inclusión de nuevas referencias de vehículos mediante la adecuada actualización del diccionario técnico. Asimismo, la segmentación y el análisis extendido de adopción con variables socioeconómicas, puede ser actualizado con información más reciente, lo que podría llevar a la identificación de nuevos segmentos y factores adicionales que influyan en la adopción departamental. Esto también permitiría el seguimiento de los hallazgos del presente análisis y la comparación entre las proyecciones realizadas y los resultados obtenidos en los años ya transcurridos.

Asimismo, sería interesante integrar otras variables como las utilizadas por Choi & Jiao (2024) que incluyen uso del suelo y proyección energética, o factores como el precio de la energía y el marco regulatorio contemplado en el análisis de Díaz Merchán et al. (2015). También podría aplicarse la segmentación a nivel de clientes potenciales mediante encuestas que permitan identificar los factores con mayor incidencia en la intención de

compra, como lo planteó Naseri et al. (2024) en su propuesta metodológica. Adicionalmente, se puede profundizar el análisis en zonas o departamentos específicos, lo que permitiría identificar patrones particulares que pueden no ser evidentes en un análisis de escala nacional.

Con esta información ampliada se podrían mejorar los resultados de los modelos implementados y expandir el análisis general de la transición energética en el país.

También, se plantea integrar el análisis de la flota eléctrica e híbrida con un estudio de impacto ambiental como el realizado por Alfonso A. et al. (2025). Esto permitiría conectar el análisis del mercado de este tipo de vehículos con el cumplimiento de metas a nivel ecológico que Colombia y cada una de sus regiones debe cumplir.

Referencias

8th ICSIMAT. (2020). *Strategic Innovative Marketing and Tourism* (A. Kavoura, E. Kefallonitis, & P. Theodoridis, Eds.). Springer International Publishing. <https://doi.org/10.1007/978-3-030-36126-6>

Akay, Ö., & Yüksel, G. (2018). Clustering the mixed panel dataset using Gower's distance and k-prototypes algorithms. *Communications in Statistics: Simulation and Computation*, 47(10), 3031–3041. <https://doi.org/10.1080/03610918.2017.1367806>

Alcaldía de Popayán. (n.d.). *Glosario de Secretaria de Tránsito y Transporte de Popayán*. Retrieved November 3, 2025, from <https://www.popayan.gov.co/SecretariasyEntidades/sectransito/Paginas/Glosario.aspx#gsc.tab=0>

Alfonso A., J. V., Bermúdez, D., Mangones, S. C., Rojas, N. Y., Rivera, C., Darghan, A., & Osses, M. (2025). Spatial and Temporal Analysis of Road Transport Emissions in Colombia. *Ingeniería e Investigación*, 45(1). <https://doi.org/10.15446/ing.investig.113104>

American Psychological Association. (2025). *socioeconomic status (SES)*. <https://dictionary.apa.org/socioeconomic-status>.

ANDI, M. P. F. (2024). *Boletín Vehículos Eléctricos e Híbridos Julio 2024*.

Andrusyk, Y. V., & Guryanova, L. S. (2025). The Hybrid Market Segmentation of Electric Vehicles in Ukraine Using Data Science Methods. *The Problems of Economy*, 2(64), 212–226. <https://doi.org/10.32983/2222-0712-2025-2-212-226>

Asociación Nacional de Empresarios de Colombia. (2025). *Cámara de la Industria de Motocicletas de la ANDI alerta sobre aumento de la ilegalidad en la venta de ciclomotores.*

Balaji, K., & Lavanya, K. (2018). Clustering Algorithms for Mixed Datasets: A Review. *International Journal of Pure and Applied Mathematics*, 118, 547–556. <http://www.ijpam.eu>

Banco de la República de Colombia. (n.d.). *Glosario de Conceptos Económicos del Banco de la República de Colombia*. Retrieved November 3, 2025, from <https://www.banrep.gov.co/es/glosario/tasa-desempleo>

Bas, J., Cirillo, C., & Cherchi, E. (2021). Classification of potential electric vehicle purchasers: A machine learning approach. *Technological Forecasting and Social Change*, 168. <https://doi.org/10.1016/j.techfore.2021.120759>

Béjar, J. (2013). *K-means vs Mini Batch K-means: A comparison.*

Ben Ncir, C.-E., Alyoubi, B., Alharbi, I., & Jarboui, A. (2022). *Machine Learning and Data Analytics for Solving Business Problems Methods, Applications, and Case Studies*. Springer.

Bhagat, A., Bhandari, P., Yadav, A. L., & Singh Talwandi, N. (2024). *AI-Powered Customer Segmentation For Marketing*. <https://ssrn.com/abstract=4811907>

Bogacki, S., Rymarczyk, P., Smutek, T., Rutkowski, M., & Chmielowska-Marmucka, A. (2024). Advanced methods for target audience identification: enhancing marketing strategies through machine learning and data analytics. *Journal of Modern Science*, (3), 417–435. <https://doi.org/10.13166/jms/191180>

Cabarrubias De La Cruz, K. P., & Santos Tolentino, L. K. (2023). Unlocking the market potential of electric vehicles in the Philippines: A statistical and neural network approach to customer willingness to purchase electric vehicles. *International Journal of Innovative Research and Scientific Studies*, 6(4), 888–902. <https://doi.org/10.53894/ijirss.v6i4.2088>

Cadavid, L., & Salazar-Serna, K. (2021). Mapping the Research Landscape for the Motorcycle Market Policies: Sustainability as a Trend—A Systematic Literature Review. In *Sustainability* (Vol. 13, Number 19). MDPI. <https://doi.org/10.3390/su131910813>

Campo, J., & Pino, J. (2024). *Análisis de tendencia de vehículos híbridos y eléctricos en Colombia*. Universidad Nacional Abierta y a Distancia UNAD.

Chisimkwuo, J., Chukwuemeka, O., & Godwin, O. (2019). The Principal Component Analysis Biplot Predictions versus the Ordinary Least Squares Regression Predictions: The Anthropometric Case Study. *Asian Journal of Probability and Statistics*.

Choi, S. J., & Jiao, J. (2024). Uncovering electric vehicle ownership disparities using K-means clustering analysis: A case study of Austin, Texas. *Journal of Computational Social Science*, 7, 2403–2456. <https://doi.org/10.1007/s42001-024-00310-6>

Cifuentes, M. (n.d.). *Ficha Técnica Población Urbana*. Retrieved http://www.dane.gov.co/daneweb_V09/index.php?option=com_

Congreso de la República de Colombia. (2019). *Ley 1964 del 11 de julio de 2019*.

Cuéllar-Álvarez, Y., Clappier, A., Rojas-Roa, N. Y., Thunis, P., Mangones, S., & Belalcázar-Cerón, L. C. (2024). Emissions and ownership-cost of conventional and electric passenger vehicles in Bogotá, Colombia. *Transportation Research Part D*, 128. <https://doi.org/10.1016/j.trd.2024.104105>

DANE. (n.d.). *Producto Interno Bruto Preguntas frecuentes*. Retrieved September 1, 2025, from https://www.dane.gov.co/files/faqs/faq_pib.pdf

Deepti Chopra, & Roopal Khurana. (2023). *Introduction to Machine Learning with Python*. Bentham Science Publishers.

Díaz Merchán, M. V., Rosero García, J., & Prías Caicedo, O. F. (2015). Métodos Prospectivos Implementados en la Vigilancia Tecnológica y Prospectiva de Vehículos Eléctricos (EV) y Tecnologías Periféricas en Colombia. *Inge CUC*, 11(2), 59–67. <https://doi.org/10.17981/ingecuc.11.2.2015.06>

Drummond, G., Ensor, J., & Ashford, R. (2008). *Strategic Marketing: Planning and Control* (Elsevier, Ed.; Third edition). Elsevier.

Dumitrache, A., Melian, D., Balacian, D., Nastu, A., & Stancu, S. (2020). Churn prepaid customers classified by HyperOpt techniques. *Proceedings of the 14th International Conference on Applied Statistics*, 139–151. <https://doi.org/10.2478/icas-2021-0013>

Dumitrescu, A. (2023). Market Segmentation Based on Materialistic, Conspicuous, and Traditionalist Goals in Support of an Effective Industrial Design. *Bulletin of the Transilvania University of Brasov. Series V: Economic Sciences*, 25–36. <https://doi.org/10.31926/but.es.2023.16.65.2.3>

Espinosa Ramírez, R. S., & Ozgur Kayalica, M. (2023). Strategic environmental policies: Electric vehicles vs internal combustion engine vehicles. *Contaduría y Administración*, 68(4), 215–240. <https://doi.org/10.22201/fca.24488410e.2023.3234>

Forero, O. M., & Mihet-Popa, L. (2014). Fast Charging and Smart Charging Tests for Electric Vehicles Batteries Using Renewable Energy. *Oil and Gas Science and Technology*, 71. <https://doi.org/10.2516/ogst/2014001>

France, S. L., & Ghose, S. (2018). *Marketing Analytics: Methods, Practice, Implementation, and Links to Other Fields*.

Gaona, A. (2024). *Estrategias de Segmentación de Mercados*. Corporación Universitaria de Asturias.

Geiler, L., Affeldt, S., & Nadif, M. (2022). An effective strategy for churn prediction and customer profiling. *Data and Knowledge Engineering*, 142. <https://doi.org/10.1016/j.datak.2022.102100>

Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics International Biometric Society*, 27(4), 857–871. <https://about.jstor.org/terms>

Hardaway, K., Genc, U., Cai, H., & Nateghi, R. (2025). Electric vehicle adoption and planning: The increasing importance of the built environment. *Journal of Transport Geography*, 123. <https://doi.org/10.1016/j.jtrangeo.2025.104115>

Imani, M., Beikmohammadi, A., & Arabnia, H. R. (2025). Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE, ADASYN, and GNUS Under Varying Imbalance Levels. *Technologies*, 13. <https://doi.org/10.3390/technologies13030088>

Jauregui Romero, E. R., Alca Gomez, J., Vilca Tantapoma, M. E., & Llanos Gonzales, O. T. (2024). Artificial intelligence in potential customer segmentation: machine learning approach. *Data and Metadata*, 3. <https://doi.org/10.56294/dm2024305>

Jiang, H., Xu, H., Liu, Q., Ma, L., & Song, J. (2024). An urban planning perspective on enhancing electric vehicle (EV) adoption: Evidence from Beijing. *Travel Behaviour and Society*, 34. <https://doi.org/10.1016/j.tbs.2023.100712>

Kamal, H., Marfani, H., Hussain, A. S., & Syed, D. (2025). A comparison of k-means and mini-batch k-means algorithm for customer segregation analysis. *Societal Transformation: AI and Big Data Journal*, 3(1), 27–49. <https://doi.org/10.20547/aibd.253103>

Kang, J. (2025). Spatial coordination of electric vehicle adoption within the urban built environment using machine learning. *Sustainable Futures*, 10. <https://doi.org/10.1016/j.sftr.2025.100807>

Kotler, P. (2000). *Marketing Management, Millenium Edition* (Pearson Education Company, Ed.; Tenth edition). Prentice-Hall Inc. www.pearsoncustom.com

Kumar, S., Singh, V., & Goel, R. (2024). Strategic Forecasting for Electric Vehicle Sales: A Cutting Edge Holistic Model Leveraging Key Factors and Machine Learning Technique. *Transportation in Developing Economies*. <https://doi.org/10.1007/s40890-024-00213-1>

Kuo, R. J., Amornnikun, P., & Nguyen, T. P. Q. (2020). Metaheuristic-based possibilistic multivariate fuzzy weighted c-means algorithms for market segmentation. *Applied Soft Computing Journal*, 96. <https://doi.org/10.1016/j.asoc.2020.106639>

Larminie, J., & Lowry, J. (2012). *Electric Vehicle Technology Explained* (Segunda). John Wiley & Sons, Ltd. www.wiley.com.

Lee, D., Perez Tellez, F., & Jaiswal, R. (2025). Predicting fire incidents with ML: an XAI approach. *AI and Ethics*, 5(5), 4429–4439. <https://doi.org/10.1007/s43681-025-00683-y>

Liu, X., Roberts, M. C., & Sioshansi, R. (2017). Spatial effects on hybrid electric vehicle adoption. *Transportation Research Part D*, 52, 85–97. <https://doi.org/10.1016/j.trd.2017.02.014>

Lopez-Arboleda, E., Sarmiento, A. T., & Cardenas, L. M. (2021). Understanding synergies between electric-vehicle market dynamics and sustainability: Case study of Colombia. *Journal of Cleaner Production*, 321. <https://doi.org/10.1016/j.jclepro.2021.128834>

Ma, H., Balthasar, F., Tait, N., Riera-Palou, X., & Harrison, A. (2012). A new comparison between the life cycle greenhouse gas emissions of battery electric vehicles and internal combustion vehicles. *Energy Policy*, 44, 160–173. <https://doi.org/10.1016/j.enpol.2012.01.034>

Marcilio, W. E., & Eler, D. M. (2020). From explanations to feature selection: Assessing SHAP values as feature selection mechanism. *Proceedings - 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2020*, 340–347. <https://doi.org/10.1109/SIBGRAPI51738.2020.00053>

Mboko, H. M., Balde, M. A. M. T., & Ndiaye, B. M. (2024). *Mobility-based Traffic Forecasting in a Multimodal Transport System*. <http://arxiv.org/abs/2411.08052>

Ministerio del Ambiente y Desarrollo Sostenible, Ministerio de Minas y Energía, Ministerio de Transporte, & Unidad de Planeación Minero Energético. (2019). *Estrategia Nacional de Movilidad Eléctrica*. PuntoAparte Bookvertising. www.minambiente.gov.co

MINTIC Colombia. (2023). *Índice de Brecha Digital IBD 2022*. <https://colombiatic.mintic.gov.co/679/articles->

Mussabayev, R., & Mussabayev, R. (2024). *Comparative Analysis of Optimization Strategies for K-means Clustering in Big Data Contexts: A Review*. <http://arxiv.org/abs/2310.09819>

Naseri, H., Waygood, E. O. D., Patterson, Z., & Wang, B. (2024). Which variables influence electric vehicle adoption? *Transportation*. <https://doi.org/10.1007/s11116-024-10525-1>

NVIDIA. (2025). *K-Means Clustering Algorithm*. <https://www.nvidia.com/en-us/glossary/k-means/>.

Osakwe, J., Shilongo, A., & Ziezo, M. (2023). Optimising Customer Segmentation in Digital Marketing Using Predictive Analytics: A Review of Literature. *Proceedings of International Conference on Information Systems and Emerging Technologies*.

Oti, E. U., Olusola, M. O., Eze, F. C., & Enogwe, S. U. (2021). Comprehensive Review of K-Means Clustering Algorithms. *International Journal of Advances in Scientific Research and Engineering*, 07(08), 64–69. <https://doi.org/10.31695/ijasre.2021.34050>

Paparrizos, J., Yang, F., & Li, H. (2024). Bridging the Gap: A Decade Review of Time-Series Clustering Methods. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)* (Vol. 1).

Pitafi, S., Anwar, T., & Sharif, Z. (2023). A Taxonomy of Machine Learning Clustering Algorithms, Challenges, and Future Realms. *Applied Sciences (Switzerland)*, 13(6), 12–14. <https://doi.org/10.3390/app13063529>

Prakash Kumbar. (2020). Impact of Hybrid Vehicles on Fuel Economy and Emissions: A Comprehensive Analysis. *World Journal of Advanced Research and Reviews*, 8(1), 307–319. <https://doi.org/10.30574/wjarr.2020.8.1.0347>

Pyzdek, T., & Keller, P. (2013). *Handbook for Quality Management: A Complete Guide to Operational Excellence* (2nd Edition). McGraw-Hill Education.

Ramadoss, T. V., Lee, J. H., Davis, A. W., Hardman, S., & Tal, G. (2025). Classifying electric vehicle adopters and forecasting progress to full adoption. *Npj Sustainable Mobility and Transport*. <https://doi.org/10.1038/s44333-025-00049-1>

Ramos Salinas, M. J., Villegas Pancca, F. M., Cordova Chipa, B. B., Cano Quito, S. P., & Lezama Gonzales, P. M. (2023). Análisis de patrones de morbilidad por anemia mediante algoritmos no supervisados: un enfoque basado en datos de establecimientos de salud a nivel nacional. *Revista de Investigación de Sistemas e Informática*, 16(2), 15–24. <https://doi.org/10.15381/risi.v16i2.25776>

Reddy, N., Reddy, S., & Adapa, M. (2025). Beyond Demographics: How Artificial Intelligence Redefines Customer Segmentation in Digital Marketing. *Computer Fraud and Security*.

Renault Group. (2021, February 23). Understanding the power output of an electric motor. *Renault Group Magazine*.
<https://www.renaultgroup.com/en/magazine/energy-and-motorization/understanding-the-power-output-of-an-electric-motor/>

Sadeq, A. M. (2024). *Hybrid and Electric Vehicles-A Comprehensive Guide* (Primera). Research Gate. <https://doi.org/10.13980/RG.2.2.33942.85975>

Salazar-Serna, K., Cadavid, L., & Franco, C. J. (2024). *Analyzing Transport Policies in Developing Countries with ABM*.

Saligkaras, D., & Papageorgiou, V. E. (2023). Seeking the Truth Beyond the Data. An Unsupervised Machine Learning Approach. In *AIP Conference Proceedings* (Vol. 2812, Number 1). American Institute of Physics Inc. <https://doi.org/10.1063/5.0161454>

Saranya, S., & Poonguzhali, S. (2024). Principal component analysis biplot visualization of electromyogram features for submaximal muscle strength grading. *Computers in Biology and Medicine*, 182. <https://doi.org/10.1016/j.compbiomed.2024.109142>

Shamsuddoha, M., & Nasir, T. (2025). The Road Ahead for Hybrid or Electric Vehicles in Developing Countries: Market Growth, Infrastructure, and Policy Needs. *World Electric Vehicle Journal*, 16(3). <https://doi.org/10.3390/wevj16030180>

Shareef, H., Islam, M. M., & Mohamed, A. (2016). A review of the stage-of-the-art charging technologies, placement methodologies, and impacts of electric vehicles. In *Renewable and Sustainable Energy Reviews* (Vol. 64, pp. 403–420). Elsevier Ltd. <https://doi.org/10.1016/j.rser.2016.06.033>

Soria Olivas, E., Sánchez, M. A., Gamero Cruz, R., Castillo Caballero, B., & Cano Michelena, P. (2023). *Sistemas de Aprendizaje Automático* (Ediciones de la U, Ed.; 1st ed.). Ediciones de la U.

Soto, J. J., Cantillo, V., & Arellana, J. (2021). Market segmentation for incentivising sustainable transport policies. *Transportation Research Part D: Transport and Environment*, 99. <https://doi.org/10.1016/j.trd.2021.103013>

Suwanda, R., Syahputra, Z., & Zamzami, E. M. (2020). Analysis of Euclidean Distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K. *Journal of Physics: Conference Series*, 1566(1). <https://doi.org/10.1088/1742-6596/1566/1/012058>

Tan, P.-Nin., Steinbach, Michael., & Kumar, Vipin. (2014). *Introduction to data mining* (Pearson Education Limited, Ed.). Pearson Education Limited.

Thomas, C. E. (2009). Fuel cell and battery electric vehicles compared. *International Journal of Hydrogen Energy*, 34, 6005–6020. <https://doi.org/10.1016/j.ijhydene.2009.06.003>

Turing. (2025). *How to Decide the Perfect Distance Metric For Your Machine Learning Model*. <https://www.turing.com/kb/how-to-decide-perfect-distance-metric-for-machine-learning-model>.

Un-Noor, F., Padmanaban, S., Mihet-Popa, L., Mollah, M. N., & Hossain, E. (2017). A comprehensive study of key electric vehicle (EV) components, technologies, challenges, impacts, and future direction of development. *Energies*, 10. <https://doi.org/10.3390/en10081217>

Uy, J. R. R., Ong, A. K. S., & German, J. D. (2024). Marketing Strategy and Preference Analysis of Electric Cars in a Developing Country: A Perspective from the Philippines. *World Electric Vehicle Journal*, 15(3).
<https://doi.org/10.3390/wevj15030111>

Velasco Rebolledo, J. (2024). *Machine learning: fundamentos, algoritmos y aplicaciones para los negocios, industria y finanzas* (1st ed.). Ediciones Díaz de Santos.

Venkata Reddy Konasani, & Shailendra Kadre. (2021). *Machine Learning and Deep Learning Using Python and TensorFlow* (McGraw Hill, Ed.; First edition). McGraw Hill.

Wani, A. A. (2024). Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions. *PeerJ Computer Science*, 10, 1–45.
<https://doi.org/10.7717/PEERJ-CS.2286>

Zherlitsyn, D., Kharchenko, V., & Kharchenko, H. (2025). Enhancing Wheat Price Forecasting Accuracy through Prophet Based Models. *International Workshop on Computational Intelligence, Co-Located with the IV International Scientific Symposium "Intelligent Solutions."*