

**MODELOS DE MACHINE LEARNING APLICADOS A UN SISTEMA DE
INDICADORES QUE EVALÚEN LA SOSTENIBILIDAD EN LA MOVILIDAD URBANA
DEL VALLE DE ABURRÁ**

Luisa Fernanda Montoya Salazar

Anteproyecto del Proyecto Aplicado para optar al título de

Magister en Ciencia de Datos

Gustavo Adolfo Arteaga Botero

FACULTAD DE INGENIERÍA Y CIENCIAS

MAESTRÍA EN CIENCIA DE DATOS

SANTIAGO DE CALI, JUNIO 28 DE 2024

TABLA DE CONTENIDO

1.	INTRODUCCIÓN	1
2.	DEFINICIÓN DEL PROBLEMA.....	2
2.1.	PLANTEAMIENTO DEL PROBLEMA	2
2.2.	FORMULACIÓN DEL PROBLEMA	4
3.	OBJETIVOS DEL PROYECTO	5
3.1.	OBJETIVO GENERAL	5
3.2.	OBJETIVOS ESPECIFICOS	5
4.	MARCO TEÓRICO Y ANTECEDENTES	6
4.1.	MARCO TEÓRICO	6
4.1.1.	Sostenibilidad en la Movilidad Urbana	6
4.1.1.1.	Sostenibilidad	6
4.1.1.2.	Movilidad sostenible	8
4.1.1.3.	Movilidad sostenible en el Valle de Aburrá.....	12
4.1.2.	Indicadores de sostenibilidad	13
4.1.2.1.	Definición y tipología de indicadores	13
4.1.2.2.	Relevancia de los indicadores para la de movilidad urbana.....	15
4.1.2.3.	Indicadores aplicados a la movilidad urbana sostenible	16
4.1.3.	Machine Learning en la Movilidad Urbana	18

4.1.3.1.	Introducción a Machine Learning	18
4.1.3.1.1.	Aprendizaje supervisado (Supervised learning).....	19
4.1.3.1.2.	Aprendizaje no supervisado (Unsupervised learning)	19
4.1.3.1.3.	Aprendizaje por refuerzo (Reinforcement learning)	20
4.1.3.2.	Enfoque en el Aprendizaje no supervisado	20
4.1.3.2.1.	Métodos de segmentación para datos numéricos	20
4.1.3.2.2.	Métodos de Partición.....	20
4.1.3.2.3.	Métodos de Densidad	21
4.1.3.2.4.	Métodos Difusos	21
4.1.3.3.	Métricas de Evaluación Comunes	22
4.1.3.4.	Aplicación de métodos de clustering para la movilidad	22
4.1.3.4.1.	Segmentación de usuarios de transporte	22
4.1.3.4.2.	Identificación de áreas con alta demanda de transporte público	23
4.1.3.4.3.	Análisis de patrones de tráfico y congestión.....	23
4.1.3.4.4.	Segmentación de áreas urbanas según su nivel de sostenibilidad.....	23
5.	ANTECEDENTES.....	24
5.1.	Locales	24
5.1.1.	Área Metropolitana del Valle de Aburrá.....	24
5.2.	Nacionales	25

5.2.1.	Santiago de Cali, Colombia.....	25
5.2.2.	Bogotá, Colombia.....	28
5.3.	Internacionales	30
5.3.1.	Madrid, España.....	30
5.3.2.	Lima, Perú.....	31
6.	DESARROLLO	34
6.1.	ENTENDIMIENTO DEL NEGOCIO	35
6.2.	ENTENDIMIENTO DE LOS DATOS	36
6.2.1.	Descripción de las Encuestas	36
6.2.2.	Carga y Transformación Inicial.....	37
6.2.2.1.	Transformación de Columnas	37
6.2.2.2.	Selección y Renombrado de Variables	37
6.2.2.3.	División y Agrupamiento de Datos	38
6.2.2.4.	Transposición y Combinación de Datos.....	38
6.2.3.	Integración de Encuestas.....	38
6.2.3.1.	Combinar Encuestas.....	38
6.2.3.2.	Enriquecimiento con Información Adicional	38
6.2.4.	Resultados y Almacenamiento	38
6.3.	PREPARACIÓN DE LOS DATOS.....	43

6.3.1.	Datos faltantes	44
6.3.2.	Análisis exploratorio de variables	46
6.3.2.1.	Teletrabajo	47
6.3.2.2.	Estrato socioeconómico.....	47
6.3.2.3.	Género	48
6.3.2.4.	Edad.....	49
6.3.2.5.	Tipo de empleado	50
6.3.2.6.	Gasto de desplazamiento	52
6.3.2.7.	Modo de llegada al trabajo	55
6.3.2.8.	Razón de uso del modo de llegada al trabajo	57
6.3.2.9.	Frecuencia de uso semanal del modo de transporte	59
6.3.2.10.	Duración de la jornada laboral	59
6.3.2.11.	Hora de inicio laboral.....	60
6.3.2.12.	Hora de salida laboral.....	62
6.3.2.13.	Duración del desplazamiento al trabajo	63
6.3.3.	Análisis exploratorio de indicadores	65
6.4.	MODELADO	69
6.4.1.	Selección del número de clúster.....	69
6.4.2.	Algoritmo K-Means	70

6.4.3.	Algoritmo Hierarchical Agglomerative Clustering (HAC).....	73
6.4.4.	Algoritmo K-Medoids.....	75
6.4.5.	Selección del mejor algoritmo de clustering.....	78
6.5.	EVALUACIÓN.....	80
6.5.1.	Caracterización de los grupos de clustering.....	80
6.5.2.	Huella de carbono (CO ₂).....	81
6.5.3.	Huella de PM _{2,5}	84
6.5.4.	Huella energética.....	86
6.5.5.	Huella económica.....	88
6.5.6.	Huella tiempo dedicado a la movilidad.....	91
6.5.7.	Mapas de calor de variables por clúster.....	95
6.6.	ANÁLISIS DEL MODELO.....	106
7.	CONCLUSIONES Y TRABAJOS FUTUROS.....	111
7.1.	CONCLUSIONES.....	111
7.1.1.	Conclusiones sobre objetivos de la investigación.....	111
7.1.2.	Conclusiones sobre el tipo de modelo de ciudad.....	113
7.2.	TRABAJOS FUTUROS.....	115
8.	REFERENCIAS BIBLIOGRÁFICAS.....	117

LISTA DE FIGURAS

Figura 1. Esquema de sostenibilidad.....	7
Figura 2. Principios modelo ciudad compacta	11
Figura 3. Esquema del sistema de indicadores para la sostenibilidad.....	14
Figura 4. Emisiones de PM 2.5. en el Valle de Aburrá	25
Figura 5. Contribución de CO2 por modo, en la ciudad de Cali.....	26
Figura 6. Contribución de PM 2.5 por modo, en la ciudad de Cali	27
Figura 7. Comparación de emisiones y uso de espacio por modos.....	29
Figura 8. Emisiones contaminantes PM2,5 y CO2, Madrid	31
Figura 9. Propósito de viaje por modo de transporte	32
Figura 10. Base de datos entregada por el AMVA	37
Figura 11. Análisis de datos faltantes	44
Figura 12. Tipo de datos.....	46
Figura 13. Frecuencia de empleados por modalidad de trabajo.....	47
Figura 14. Frecuencia de empleados por estrato socioeconómico	47
Figura 15. Distribución de empleados por estrato socioeconómico	48
Figura 16. Frecuencia de empleados por género.....	48
Figura 17. Distribución de empleados por género	49
Figura 18. Frecuencia de empleados por edad	49
Figura 19. Distribución de empleados por edad.....	50
Figura 20. Frecuencia de empleados por tipo de cargo.....	50
Figura 21. Frecuencia de empleados por tipo de cargo.....	51

Figura 22. Frecuencia de empleados que no relacionan gasto	52
Figura 23. Resumen estadístico gasto de desplazamiento	53
Figura 24. Boxplot gasto de desplazamiento	53
Figura 25. Boxplot sin datos atípicos del gasto de desplazamiento	54
Figura 26. Resumen estadístico sin datos atípicos del gasto de desplazamiento	55
Figura 27. Frecuencia de empleados por modo de transporte.....	56
Figura 28. Frecuencia de empleados por razón de uso modo de transporte	57
Figura 29. Porcentaje acumulado razón de uso modo de transporte.....	57
Figura 30. Frecuencia ajustada razón de uso modo de transporte.....	58
Figura 31. Frecuencia de uso semanal del modo de transporte.....	59
Figura 32. Resumen estadístico duración jornada laboral	59
Figura 33. Boxplot duración jornada laboral	60
Figura 34. Frecuencia de empleados por hora de inicio laboral	61
Figura 35. Gráfico de línea para la hora de inicio laboral.....	61
Figura 36. Frecuencia de empleados por hora de salida laboral	62
Figura 37. Gráfico de línea para la hora de salida laboral	62
Figura 38. Boxplot duración del desplazamiento al trabajo.....	63
Figura 39. Resumen estadístico de la duración del desplazamiento al trabajo	64
Figura 40. Boxplot sin datos atípicos de la duración del desplazamiento al trabajo	64
Figura 41. Resumen estadístico sin datos atípicos de la duración del desplazamiento al trabajo.....	65
Figura 42. Boxplot de cada uno de los indicadores de movilidad sostenible	66
Figura 43. Resumen estadístico de cada uno de los indicadores de movilidad sostenible.....	66

Figura 44. Selección del número de clúster	69
Figura 45. Score Silhouette para cada clúster K-Means	71
Figura 46. Gráfico de dispersión clúster K-Means	72
Figura 47. Gráfico de dispersión Hierarchical Agglomerative Clustering.....	74
Figura 48. Score Silhouette para cada clúster K-Medoids.....	76
Figura 49. Gráfico de dispersión K-Medoids.....	78
Figura 50. Frecuencia de empleados por grupo de clustering.....	80
Figura 51. Distribución de empleados por grupo de clustering	81
Figura 52. Resumen estadístico Huella de carbono (CO2) por clúster	81
Figura 53. Boxplot Huella de carbono (CO2) por clúster	82
Figura 54. Resumen estadístico Huella de PM 2.5 por clúster	84
Figura 55. Boxplot Huella de PM 2.5 por clúster	84
Figura 56. Resumen estadístico Huella energética por clúster.....	86
Figura 57. Boxplot Huella energética por clúster	87
Figura 58. Resumen estadístico Huella económica por clúster.....	88
Figura 59. Boxplot Huella económica por clúster	89
Figura 60. Resumen estadístico Huella de tiempo dedicado a la movilidad por clúster.....	91
Figura 61. Boxplot Huella de tiempo dedicado a la movilidad por clúster.....	92
Figura 62. Mapa de calor de frecuencia de género, edad y modalidad de trabajo por clúster	95
Figura 63. Mapa de calor porcentual de género, edad y modalidad de trabajo por clúster.....	96
Figura 64. Mapa de calor de frecuencia de hora inicio y salida laboral por clúster.....	98
Figura 65. Mapa de calor porcentual hora de inicio y salida laboral	99

Figura 66. Boxplot duración jornada laboral por clúster	99
Figura 67. Mapa de calor de frecuencia del modo de transporte por clúster	101
Figura 68. Mapa de calor porcentual del modo de transporte por clúster.....	102
Figura 69. Mapa de calor de frecuencia de tipo labor y cantidad de uso por clúster	104
Figura 70. Mapa de calor porcentual de tipo labor y cantidad de uso por clúster	104
Figura 71. Dashboard indicadores movilidad sostenible	109

LISTA DE TABLAS

Tabla 1. Definición indicadores movilidad urbana sostenible.....	16
Tabla 2. Descripción de las variables entregadas por el negocio	39
Tabla 3. Modos de transporte reconocidos para cálculo de emisiones	41
Tabla 4. Métricas de evaluación Algoritmo K-Means.....	70
Tabla 5. Métricas de evaluación Algoritmo HAC	73
Tabla 6. Métricas de evaluación Algoritmo K-Medoids	75

1. INTRODUCCIÓN

La movilidad diaria es un fenómeno extraordinariamente complejo, en el que influyen dimensiones sociales, culturales e incluso emocionales y que las acciones en torno a esta situación tienen implicaciones que pueden ir desde afectaciones en la salud pública, aumento de los niveles de accidentalidad, ineficiente ocupación del espacio público e incremento de recursos hídricos y energéticos.

Por lo tanto, para disminuir los efectos anteriores, es importante garantizar una movilidad urbana sostenible que sea eficaz y equitativa, que favorezca la accesibilidad a los recursos y que garantice la conexión entre los desplazamientos intra e interurbanos y la intermodalidad. Para el logro de lo anterior, es necesario conocer las dinámicas actuales de movilidad, entender de qué manera se mide este proceso y conocer el modelo de ciudad por el que se apueste. Sin embargo, los actuales instrumentos de planeación para políticas públicas están desactualizados respecto a la velocidad con la que evolucionan las dinámicas urbanas, ejemplo de esto son las Encuestas Origen Destino, donde su última versión para los diez (10) municipios que componen el Área Metropolitana del Valle de Aburrá corresponde al año 2017, es decir, que se están planeando y ejecutando proyectos con información que no representan las dinámicas actuales.

En este sentido, la presente investigación tiene como objetivo construir modelos de Machine Learning aplicados a un sistema de indicadores que evalúen la sostenibilidad en la movilidad urbana del Valle de Aburrá, con el propósito de identificar patrones de comportamiento y agrupaciones significativas en los datos relacionados con la movilidad urbana. Estos insights pueden ser utilizados por instituciones, entidades y autoridades como instrumentos de

planificación que aporten a la toma de decisiones al momento de ejecutar proyectos de movilidad urbana en el territorio.

Para alcanzar estos objetivos, se llevaron a cabo cinco etapas, que incluyeron desde la búsqueda bibliográfica de indicadores de sostenibilidad para evaluar la movilidad urbana, hasta el análisis de los resultados y las conclusiones derivadas de un modelo de Machine Learning basado en clustering. Este modelo permitió identificar patrones de comportamiento y agrupaciones significativas en datos relacionados con la movilidad urbana y sus indicadores, lo cual fue fundamental para la planificación estratégica y la toma de decisiones informadas en este campo, particularmente para la entidad que proporcionó los datos, los cuales serán comparados con resultados de otros modelos realizados por diferentes universidades y consultorías.

2. DEFINICIÓN DEL PROBLEMA

Este capítulo se divide en dos (2) secciones que hablan sobre la problemática asociada al modelo actual de movilidad en la región y las preguntas esenciales que guiarán el estudio. El objetivo de este trabajo es abordar de manera rigurosa estos desafíos y presentar enfoques innovadores basados en indicadores y modelos de Machine Learning, con el propósito de elevar la calidad de vida y fomentar la competitividad en la región del Valle de Aburrá.

2.1. PLANTEAMIENTO DEL PROBLEMA

La movilidad urbana representa uno de los grandes retos a corto y mediano plazo para hacer sostenibles a las ciudades contemporáneas. En Latinoamérica el nivel creciente de urbanización y

la mayor productividad asociada al proceso de aglomeración económica, han inducido una mayor demanda de vehículos por parte de los hogares y de los negocios [1].

El modelo de movilidad actual del Área Metropolitana del Valle de Aburrá (AMVA) se basa en el transporte individual y motorizado; esto hace que el sistema sea insostenible, ineficiente, inequitativo, riesgoso e improductivo, y que afecte la sostenibilidad y la competitividad de la región. Lo anterior se evidencia en el crecimiento del parque automotor en un 186% en los últimos 10 años, lo que aumenta la congestión vial y, por consiguiente, el tiempo promedio de viaje de los ciudadanos. Adicional a esto, también se generan efectos en la contaminación atmosférica, los índices de accidentalidad y el consumo energético, siendo el sector transporte de los mayores consumidores de energía con una participación del 45.4% del total de la energía consumida en el AMVA [2].

Para contribuir con la planeación de la movilidad y el transporte en el territorio, existen herramientas como estudios longitudinales o encuestas de movilidad, que generan grandes volúmenes de datos y que implican un alto costo de recopilación y levantamiento de información. Sin embargo, en muchos casos estos estudios tienen un bajo impacto debido a la calidad de los datos, el desconocimiento de la información generada por los estudios, la “rápida” pérdida de vigencia de la información, el cambio de administraciones, que provoca la “pérdida” de la información [3].

En este sentido, diseñar un sistema de indicadores que evalúen la sostenibilidad de la movilidad urbana y aplicar modelos de Machine Learning que permitan identificar patrones de comportamiento y agrupaciones significativas en los datos representan un punto de partida valioso

para que instituciones, entidades y autoridades lo consideren como instrumento que aporta a la toma de decisiones al momento de planear proyectos de movilidad.

2.2. FORMULACIÓN DEL PROBLEMA

A continuación, se presentan las preguntas problemáticas que guiarán esta investigación y que serán validadas a lo largo del desarrollo de este estudio.

- ¿Cuáles son las condiciones de sostenibilidad que tiene la movilidad urbana del Valle de Aburrá?
- ¿Los territorios a nivel local cuentan con indicadores de sostenibilidad en común para evaluar la movilidad urbana?
- ¿La base de datos utilizada cumplió con los parámetros adecuados para la observación?
- ¿El modelo de Machine Learning seleccionado generó conclusiones respecto a patrones y posibles comportamientos en la movilidad urbana?
- ¿El visualizador de información contempla todos los posibles datos de interés para la comunidad?

3. OBJETIVOS DEL PROYECTO

3.1. OBJETIVO GENERAL

Construir modelos de Machine Learning aplicados a un sistema de indicadores que evalúen la sostenibilidad en la movilidad urbana del Valle de Aburrá.

3.2. OBJETIVOS ESPECIFICOS

- Elaborar el estado del arte de los indicadores de sostenibilidad más utilizados para analizar la movilidad urbana en diferentes contextos locales.
- Diseñar las estrategias de recolección de información y depuración de bases de datos para la observación.
- Seleccionar el modelo de Machine Learning adecuado para evaluar agrupaciones y patrones de comportamiento de la movilidad urbana en el Valle de Aburrá.
- Formular los parámetros para un visualizador de información que contenga resultados de interés para la comunidad.

4. MARCO TEÓRICO Y ANTECEDENTES

4.1. MARCO TEÓRICO

4.1.1. Sostenibilidad en la Movilidad Urbana

4.1.1.1. Sostenibilidad

La sostenibilidad, como principio fundamental, aborda la interconexión entre el desarrollo económico social y ambiental. Surge en respuesta a la creciente preocupación por la degradación del medio ambiente ocasionada por el modelo de crecimiento económico sin límites, que ignoraba las consecuencias ambientales. Su objetivo principal es asegurar que las necesidades presentes sean satisfechas sin poner en riesgo los recursos y oportunidades de las generaciones futuras. Este enfoque implica una transformación profunda en la forma en que se gestionan los recursos naturales, se desarrolla la tecnología y se estructuran las instituciones, con el propósito de garantizar un desarrollo equitativo y sostenible a largo plazo [4].

Además, la sostenibilidad implica una relación armoniosa entre la sociedad y el entorno natural, preservando la salud y la diversidad de los ecosistemas que sustentan la vida en la Tierra. Este concepto va más allá de la simple conservación del medio ambiente, involucrando también aspectos sociales y económicos. Se trata de promover prácticas y políticas que permitan mejorar la calidad de vida de las personas, al mismo tiempo que se protege y regenera el entorno natural. Por tanto, la sostenibilidad se presenta como un paradigma indispensable para abordar los desafíos ambientales y sociales del siglo XXI, orientando hacia un desarrollo más justo, equitativo y respetuoso con el planeta y sus habitantes [5].



Figura 1. Esquema de sostenibilidad

Fuente: Motos G., 2019 [6]

En términos generales, la sostenibilidad se concibe como el enfoque para abordar los desafíos originados por la actividad humana, tales como la contaminación del aire, los cambios climáticos, la disminución de la diversidad biológica y el agotamiento de los recursos naturales. Tal como se observa en la **Figura 1** este concepto se fundamenta en tres pilares principales:

- **Economía:** Crucial para el desarrollo de proyectos, debe generar beneficios económicos a corto o largo plazo para ser viable.
- **Social:** Esencial para que los proyectos sostenibles beneficien a la sociedad en su conjunto, promoviendo la equidad y la inclusión.
- **Medio ambiente:** Debido a la limitación de recursos naturales, cualquier actividad que afecte al medio ambiente debe abordarse con precaución para preservar los recursos para las generaciones futuras.

En resumen, una buena definición de sostenibilidad podría ser la "evolución de la población hacia las necesidades actuales respetando el entorno natural y sin comprometer los recursos para las generaciones futuras". Este enfoque implica un desarrollo consciente del medio ambiente y una

consideración equilibrada de las necesidades de las próximas generaciones. Los pilares fundamentales de la economía, el medio ambiente y la sociedad son igualmente importantes en este contexto [7].

4.1.1.2.Movilidad sostenible

El concepto de movilidad urbana sostenible se fundamenta en la integración de diversas dimensiones del desarrollo sostenible, que como se mencionó en el concepto dado para la sostenibilidad, incluyen aspectos ambientales, económicos y sociales [8]. En primer lugar, la movilidad urbana sostenible busca garantizar la protección del medio ambiente, mediante la reducción de la contaminación y la minimización del uso de recursos no renovables en los desplazamientos dentro de las ciudades. Esto implica promover formas de transporte más eficientes y menos contaminantes, como el transporte público y el uso de medios no motorizados como la bicicleta. Además, se busca mantener una buena calidad de vida para los ciudadanos, facilitando el acceso equitativo a servicios y actividades urbanas, lo cual mejora la accesibilidad y la inclusión social [9].

En segundo lugar, la movilidad urbana sostenible se orienta hacia el desarrollo económico, mediante la maximización de los beneficios y la minimización de los costos externos asociados a la movilidad. Esto implica la eficiente gestión de los recursos y la infraestructura de transporte, así como la promoción de inversiones sostenibles que contribuyan al desarrollo económico de las ciudades. Se buscan políticas que fomenten la innovación en tecnologías limpias y la creación de empleo en sectores relacionados con la movilidad sostenible [10].

Por otro lado, las políticas de movilidad urbana sostenible buscan promover la equidad y la justicia

social, garantizando que los beneficios de la movilidad sean distribuidos de manera equitativa entre todos los ciudadanos. Esto implica eliminar barreras físicas y sociales que limiten el acceso a los servicios y oportunidades urbanas, así como promover la participación ciudadana en la toma de decisiones relacionadas con la movilidad, presentándose como un enfoque integral que busca armonizar las necesidades de movilidad de la población con la protección del medio ambiente y el desarrollo económico y social de las ciudades [8].

Así, la movilidad sostenible, vista desde la perspectiva de la calidad de vida urbana y una adecuada gestión municipal, implica un cambio en las prioridades de las políticas públicas. Destacando como acciones prioritarias, las siguientes:

- Rehabilitar el espacio público urbano para el disfrute de los ciudadanos.
- Moderar el tráfico para mejorar la convivencia y la seguridad vial.
- Incrementar la eficiencia y la accesibilidad del transporte colectivo.
- Promover los desplazamientos a pie como medio de transporte preferente.
- Fomentar el uso de la bicicleta como alternativa sostenible y saludable.
- Implementar un conjunto de medidas que desalienten comportamientos no deseables y promuevan aquellos socialmente beneficiosos [2].

Por último, se destaca la importancia de incorporar las percepciones y emociones de los ciudadanos en el marco teórico de la planificación de la movilidad urbana. Se cuestiona si la participación ciudadana puede no solo identificar aspectos objetivos, sino también aquellos subjetivos que influyen en la experiencia de movilidad, lo que plantea la necesidad de revisar las políticas y procedimientos actuales para asegurar una planificación más inclusiva y centrada en el

bienestar emocional de los ciudadanos [11].

Algunos lugares como Copenhague, Chile, Teherán, Vancouver y México han logrado mejorar las condiciones de movilidad mediante un nuevo diseño urbano que fomenta la sustentabilidad, promoviendo el uso de bicicletas, así como la transitabilidad (viajes a pie o caminata activa) y la utilización de transporte público. Enfocando sus esfuerzos también en la preservación de espacios verdes y reconstrucción de espacios culturales [12].

La movilidad sostenible es clave para integrar los principios definidos por Salvador Rueda desde 1999, ya que se enfoca en desarrollar sistemas de transporte que no solo sean eficientes, sino también respetuosos con el medio ambiente y equitativos para todos los ciudadanos. Su estado del arte, conocido como principios e indicadores para ciudades más sostenibles, ha subrayado la importancia de indicadores de sostenibilidad urbana que consideren el impacto ambiental, social y económico de las políticas de movilidad, haciendo hincapié en la necesidad de ciudades que favorezcan el uso de medios de transporte no contaminantes y la planificación urbana orientada a reducir la huella ecológica [13]

Incorporar la movilidad sostenible en el diseño urbano, significa adoptar una visión holística donde se prioriza la creación de entornos que reduzcan la necesidad de transporte motorizado, mejoren la calidad de vida y contribuyan a la lucha contra el cambio climático. Esto resulta en ciudades más habitables, inclusivas y preparadas para las generaciones futuras.

Este enfoque está en línea con el concepto de ciudad compacta, que se caracteriza por un desarrollo urbano denso, accesible y eficiente en el uso del suelo. A diferencia de la ciudad difusa, la ciudad compacta promueve la proximidad, la mezcla de usos y la conectividad, lo que facilita la

implementación de principios de movilidad sostenible. Al adoptar estos principios, se minimiza la expansión urbana descontrolada, se optimizan los recursos y se crean entornos más saludables y equitativos, donde la movilidad activa y el transporte público son las opciones preferidas. De esta manera, la ciudad compacta se convierte en un modelo ideal para alcanzar los objetivos de sostenibilidad y calidad de vida urbana, como lo proponen los ocho principios del urbanismo, los cuales de manera grafica se pueden observar en la **Figura 2**.



Figura 2. Principios modelo ciudad compacta

Fuente: Rueda, et. al, 1999 [14]

La **Figura 2** representa el concepto de ciudad compacta y visibiliza la importancia de incorporar los ocho principios del urbanismo a la construcción de movilidad sostenible. Este modelo de ciudad promueve un entorno urbano donde caminar, pedalear y usar el transporte público son

opciones viables y preferidas. En una ciudad compacta, la proximidad de servicios y la mezcla de usos reducen la necesidad de transporte motorizado, mientras que la densificación y la concentración del desarrollo en áreas urbanas existentes optimizan el uso del suelo y los recursos. Además, al minimizar el espacio destinado a automóviles y priorizar redes de transporte bien conectadas, se crea una ciudad más eficiente, equitativa y resiliente, alineada con los objetivos de sostenibilidad y calidad de vida urbana [14].

4.1.1.3. Movilidad sostenible en el Valle de Aburrá

El Plan Maestro de Movilidad para el Valle de Aburrá es una herramienta de planeación que establece directrices para implementar un modelo de movilidad sostenible y seguro hasta el año 2030, basado en las perspectivas de desarrollo del territorio definidas en el Plan Estratégico Metropolitano de Ordenamiento Territorial (PEMOT) y los planes de ordenamiento territorial municipales.

Su modelo actual está centrado en el transporte individual y motorizado, afectando la competitividad y la sostenibilidad regional por el aumento significativo del parque automotor, la congestión vial, la contaminación y los accidentes de tránsito. Por lo tanto, el Área Metropolitana del Valle de Aburrá (AMVA) en pro de mejorar la sostenibilidad y competitividad del territorio, establece el objetivo de promover una movilidad sostenible y segura mediante una serie de proyectos y acciones que abordan áreas clave de la movilidad en la región.

Una de ellas es la creación de una región segura y amigable para la movilidad activa, promoviendo el desplazamiento peatonal y en bicicleta a través de proyectos como la construcción de centralidades caminables y la consolidación de la red ciclista. Otra acción importante es el

fortalecimiento del transporte público, haciendo énfasis en la accesibilidad para todos mediante la implementación de corredores de movilidad sostenible y la mejora de la infraestructura del SITVA. Además, se busca una infraestructura equitativa para todos los ciudadanos, con medidas como la implementación de restricciones al uso del vehículo privado y el fortalecimiento de los Planes de Movilidad Empresarial Sostenible (Planes MES). Estos pilares se complementan con acciones para promover el respeto por la vida en las vías y fortalecer la gestión institucional de la movilidad, junto con la divulgación de los beneficios de una movilidad sostenible [2].

4.1.2. Indicadores de sostenibilidad

4.1.2.1. Definición y tipología de indicadores

La inclusión de indicadores para el análisis de la movilidad urbana sostenible refleja la complejidad y relevancia que el tema debe tener en la planificación del territorio. Estas medidas cuantitativas permiten evaluar el desempeño de políticas, programas y acciones destinadas a promover una movilidad urbana más sostenible. Los indicadores se clasifican en distintas categorías según los aspectos que midan, abarcando desde la infraestructura vial hasta las emisiones de gases de efecto invernadero [7].

En la evaluación de la sostenibilidad en la movilidad urbana, los indicadores son esenciales. Se dividen en dos categorías principales: ambientales y de desarrollo sostenible. Los primeros se centran en comprender los estados y dinámicas ambientales, mientras que los segundos integran aspectos económicos, sociales y ambientales. La selección y construcción de estos indicadores implica un proceso riguroso que incluye la revisión de documentación, la recopilación de información a través de diversas fuentes y el análisis histórico de datos. Estas herramientas son

fundamentales para una toma de decisiones informada y para impulsar el desarrollo sostenible de las ciudades [15].

Como se mencionó en líneas anteriores, para evaluar la sostenibilidad en la movilidad urbana sostenible existe una amplia gama de indicadores que abarcan desde la cantidad de infraestructura vial dedicada al transporte público y a las bicicletas hasta la seguridad vial y las emisiones de gases contaminantes. Dichos indicadores son herramientas fundamentales para medir el progreso hacia una movilidad urbana más sostenible. Por ejemplo, el número de kilómetros de ciclovías puede indicar el nivel de accesibilidad y seguridad para los ciclistas, mientras que la cantidad de accidentes de tránsito por cada 100,000 habitantes refleja la seguridad vial en la ciudad. Asimismo, el porcentaje de viajes en transporte público per cápita y la edad de la flota vehicular son indicadores importantes para evaluar la eficiencia del transporte público y la calidad del aire [7].

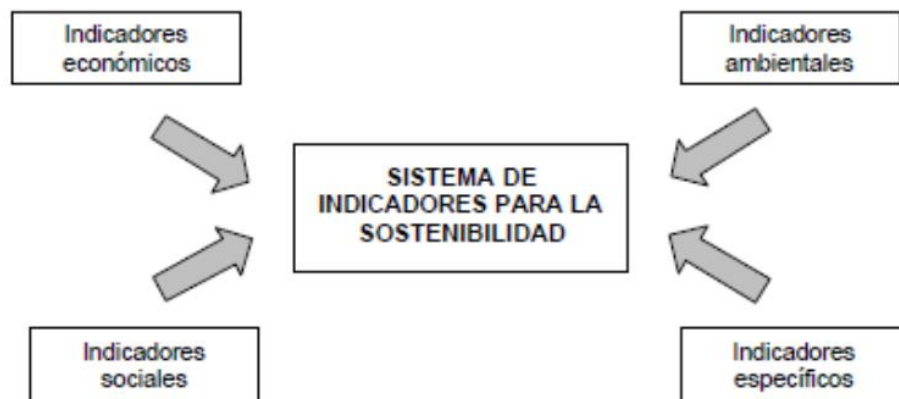


Figura 3. Esquema del sistema de indicadores para la sostenibilidad

Fuente: Motos G., 2019 [7]

La **Figura 3** consolida el estado del arte realizado por Salvador Rueda, un pionero en el desarrollo de modelos e indicadores para ciudades sostenibles que propuso una serie de indicadores que van

desde la infraestructura vial y la calidad del aire hasta la seguridad vial y el uso de transporte público, proporcionando así una base sólida para medir el progreso hacia la movilidad sostenible. Dichos indicadores pueden quedar discriminados en cuatro frentes importantes: económicos, sociales, ambientales y específicos.

4.1.2.2.Relevancia de los indicadores para la de movilidad urbana

La selección meticulosa y la vigilancia constante de indicadores son fundamentales para dirigir la planificación de políticas de movilidad urbana sostenible. Estos indicadores son instrumentos que permiten identificar áreas de mejora y priorizar acciones para impulsar un transporte más sostenible en entornos urbanos. Por ejemplo, si un indicador revela un déficit de infraestructura ciclista en comparación con estándares recomendados, esto podría señalar la necesidad de invertir en la expansión de las ciclovías para fomentar el uso de la bicicleta como medio de transporte sostenible. Asimismo, si los índices de accidentes de tránsito superan ciertos límites establecidos, esto podría requerir la implementación de medidas adicionales de seguridad vial, como la creación de zonas peatonales seguras o la mejora de la señalización vial [7].

Los indicadores son herramientas fundamentales para la planificación de políticas de movilidad urbana sostenible, dado que posibilitan la evaluación, el control y la divulgación de la evolución de factores sociales, ambientales y económicos relevantes para el bienestar de la población y el medio ambiente. Estos indicadores ofrecen datos clave para identificar áreas de mejora, priorizar acciones y monitorear el impacto de las políticas implementadas en el ámbito de la movilidad urbana. Además, facilitan la comunicación con el público y fomentan la transparencia en la toma de decisiones, contribuyendo así a la edificación de ciudades más sostenibles y habitables. La

inclusión de indicadores en cualquier proceso de planificación urbana posibilita la cuantificación y comparación de resultados para adoptar decisiones fundamentadas y promover el desarrollo sostenible de las urbes[6].

4.1.2.3.Indicadores aplicados a la movilidad urbana sostenible

El Plan Integral de Movilidad Sostenible (PIMS) es un instrumento de planeación que permite a las organizaciones gestionar estrategias dirigidas a promover la movilidad sostenible y mitigar las externalidades negativas asociadas a los viajes realizados por motivos laborales o de estudio. En el marco del PIMS, se establecen indicadores clave que permiten cuantificar aspectos como la huella de carbono, energética, calidad de vida, económica, equidad y sedentarismo, proporcionando así una visión integral de la movilidad dentro de una organización.

Estos indicadores son fundamentales para evaluar el impacto de las acciones implementadas y orientar la toma de decisiones hacia prácticas más sostenibles. La normatividad relacionada con el PIMS, emanada por la Administración Distrital y respaldada por decretos, acuerdos y leyes, brinda un marco legal para el desarrollo e implementación efectiva de este tipo de planes en el ámbito institucional [16]. La **Tabla 1**, muestra convergencia con los indicadores que se proponen en la **Figura 3**, dado que se consideran indicadores de tipo social, ambiental, económicos y específicos en el ámbito de la movilidad sostenible.

Tabla 1. Definición indicadores movilidad urbana sostenible

Huella	Definición	Unidad de Medida	Impacto
Huella de	Mide las emisiones	gr CO2/día	Contribuye al calentamiento global y

Huella	Definición	Unidad de Medida	Impacto
Carbono (CO2)	de gases de efecto invernadero, principalmente CO2		cambio climático por efecto invernadero. Además de la combustión de combustibles fósiles, las emisiones de CO2 también pueden ser indirectas, como las relacionadas con la generación de electricidad para vehículos eléctricos.
Huella de PM2,5	Mide las emisiones de partículas finas PM2,5 en el aire	gr PM2,5/día	Impacta la salud al penetrar en los pulmones y torrente sanguíneo, causando problemas respiratorios y cardiovasculares. Además de la combustión de motores, las partículas PM2,5 pueden originarse del desgaste de neumáticos y frenos, así como de actividades industriales y residenciales.
Huella Energética	Mide el consumo de energía relacionado con la movilidad	Gal/día	Uso de recursos energéticos no renovables y emisiones asociadas. Considera tipos de combustibles utilizados como gasolina, diesel, entre otros.
Huella Económica	Mide los costos monetarios	\$ COP/día	Carga económica directa e indirecta, incluyendo costos externos como

Huella	Definición	Unidad de Medida	Impacto
	asociados con la movilidad		mantenimiento de infraestructura y congestión.
Huella	Mide el tiempo empleado diariamente en desplazamientos	min/día	Impacto en calidad de vida, bienestar personal y oportunidades de participación en otras actividades.

Fuente: Elaboración propia

Por lo tanto, la **Tabla 1** combinada con el enfoque integral propuesto por Salvador Rueda, quien ha establecido una base sólida para la medición de ciudades sostenibles mediante una serie de indicadores económicos, sociales, ambientales y específicos, forman un estado del arte robusto para definir y analizar la movilidad sostenible. El cálculo de estos indicadores es viable a partir de los datos que recopila el Área Metropolitana del Valle de Aburrá en sus encuestas de movilidad, lo que los convierte en una excelente oportunidad para analizar de manera holística la movilidad urbana sostenible en la región. Esta integración proporciona una visión comprehensiva y multidimensional que permitirá una evaluación más precisa y efectiva de las estrategias de movilidad en el contexto de esta investigación.

4.1.3. Machine Learning en la Movilidad Urbana

4.1.3.1. Introducción a Machine Learning

El Machine Learning (ML) es una rama de la inteligencia artificial que capacita a las máquinas

para aprender de los datos y tomar decisiones sin intervención humana directa. A través de modelos estadísticos, las computadoras pueden identificar patrones en grandes conjuntos de datos, adaptarse a medida que se exponen a nuevos datos y mejorar con el tiempo. Este enfoque se ha vuelto fundamental debido al aumento en la disponibilidad de datos, el procesamiento computacional mejorado y el almacenamiento accesible. El ML se clasifica en diferentes tipos, como aprendizaje supervisado, no supervisado y por esfuerzo, cada uno con aplicaciones únicas que abordaremos en este trabajo, incluyendo su relevancia en la movilidad urbana sostenible [17].

4.1.3.1.1. Aprendizaje supervisado (Supervised learning)

En este enfoque, el modelo se entrena con datos etiquetados, es decir, datos que ya tienen la respuesta correcta. En el contexto de la movilidad urbana sostenible, un ejemplo sería la predicción de la demanda de transporte público en función de variables como el día de la semana, la hora del día y las condiciones climáticas. Otro caso podría ser la clasificación de rutas de transporte en eficientes o no eficientes según el nivel de congestión observado.

4.1.3.1.2. Aprendizaje no supervisado (Unsupervised learning)

En este enfoque, el modelo se entrena con datos sin etiquetas con el objetivo de encontrar patrones y estructuras ocultas en los datos. En el ámbito de la movilidad urbana sostenible, un ejemplo sería la segmentación de viajes de acuerdo con las características de los usuarios y sus patrones de movilidad, lo que podría ayudar a diseñar políticas de transporte más adaptadas a las necesidades de la población. Otro caso podría ser la identificación de áreas urbanas con problemas de congestión vehicular basándose en datos de tráfico y movilidad.

4.1.3.1.3. Aprendizaje por refuerzo (Reinforcement learning)

En este enfoque, el modelo aprende a través de la interacción con un entorno, recibiendo recompensas o castigos por sus acciones. En el contexto de la movilidad urbana sostenible, un ejemplo sería el diseño de algoritmos para la gestión del tráfico que aprenden a optimizar la fluidez del tráfico y reducir la congestión a partir de la retroalimentación obtenida en tiempo real de sensores de tráfico y sistemas de control de semáforos. Otro caso podría ser el desarrollo de sistemas de gestión de flotas de vehículos eléctricos que aprenden a optimizar las rutas y horarios de carga para maximizar la eficiencia energética y reducir las emisiones de gases de efecto invernadero [18].

4.1.3.2. Enfoque en el Aprendizaje no supervisado

4.1.3.2.1. Métodos de segmentación para datos numéricos

Son técnicas de análisis de datos que se utilizan para agrupar observaciones en conjuntos homogéneos o clústers basados en características numéricas o cuantitativas. Estos métodos son fundamentales en el campo del aprendizaje automático y la minería de datos, ya que permiten identificar patrones y estructuras ocultas en conjuntos de datos complejos. La segmentación es especialmente útil cuando se trabaja con grandes volúmenes de datos numéricos y se busca comprender la estructura subyacente de los datos sin la necesidad de etiquetas predefinidas [19].

4.1.3.2.2. Métodos de Partición

- **K-Means:** Un algoritmo de clustering que divide un conjunto de datos en K grupos basados en las características de las observaciones. Funciona iterativamente asignando puntos de datos a los grupos más cercanos y recalculando los centroides de los grupos.

- **K-Medoides:** Similar a K-Means, pero utiliza puntos de datos reales en lugar de centroides para representar cada grupo. Minimiza la distancia total de cada punto de datos al medoide (punto central) de su respectivo grupo.
- **CLARA:** Abreviación de Clustering Large Applications, es una variante de K-Medoides diseñada para conjuntos de datos grandes. Utiliza una muestra aleatoria para construir un subconjunto más pequeño que se utiliza para la agrupación.

4.1.3.2.3. Métodos de Densidad

- **DBSCAN:** Density-Based Spatial Clustering of Applications with Noise es un algoritmo que agrupa puntos de alta densidad en clústers, ignorando regiones de baja densidad. Es especialmente útil para datos con clústers de formas arbitrarias y para identificar puntos ruidosos o outliers.
- **DENCLUE:** Density Estimation-based Clustering of Applications with Noise utiliza una técnica de estimación de densidad para identificar clústers en datos multidimensionales. Tiene en cuenta la distribución de densidades locales para determinar los límites de los clústers.

4.1.3.2.4. Métodos Difusos

- **Método Fuzzy C-Means:** Una extensión del algoritmo K-Means que permite la pertenencia parcial de los puntos de datos a múltiples clústers. Asigna a cada punto un grado de pertenencia a cada clúster en función de su proximidad a los centroides.

4.1.3.3. Métricas de Evaluación Comunes

- **Inercia:** Es la suma de las distancias cuadráticas entre cada punto y el centroide de su clúster. Se utiliza para medir la compactación de los clústeres.
- **WCSS (Within-Cluster Sum of Squares):** Mide la variación interna dentro de cada clúster. Es una métrica importante para evaluar la calidad de los clústeres, ya que valores más bajos indican una mejor calidad de agrupación.
- **Coefficiente de Silueta:** Evalúa la calidad de la agrupación de un punto en su clúster en comparación con otros clústeres. Los valores del coeficiente de silueta oscilan entre -1 y 1, donde valores cercanos a 1 indican que el punto está bien agrupado, mientras que valores cercanos a -1 sugieren que el punto podría estar mal asignado.
- **Índice Silueta:** Mide la cohesión y separación de los clústeres, proporcionando una visión general de la calidad del clustering. Un valor promedio alto indica clústeres bien definidos.
- **Índice Davies-Bouldin (DB):** Evalúa la relación entre la distancia entre clústeres y la dispersión interna de los clústeres. Un valor bajo de DB sugiere clústeres más distintos entre sí y más compactos internamente.
- **Índice de Dunn:** Mide la separación entre los clústeres y la compactación interna. Un valor alto indica una buena separación entre clústeres y una alta cohesión interna.

4.1.3.4. Aplicación de métodos de clustering para la movilidad

4.1.3.4.1. Segmentación de usuarios de transporte

Utilizando algoritmos de clustering, es posible segmentar a los usuarios de transporte público en diferentes grupos en función de sus patrones de viaje, preferencias y comportamientos. Esto puede

ayudar a diseñar servicios de transporte más personalizados y adaptados a las necesidades específicas de cada segmento de usuarios, lo que a su vez puede fomentar el uso del transporte público y reducir la congestión vehicular [20].

4.1.3.4.2. Identificación de áreas con alta demanda de transporte público

Mediante el análisis de datos de movilidad, como la cantidad de pasajeros en diferentes estaciones o paradas de autobús, y utilizando técnicas de clustering, es posible identificar áreas urbanas con una alta demanda de transporte público. Esto puede ser útil para planificar la expansión de la infraestructura de transporte público en áreas donde sea más necesaria, mejorando así la accesibilidad y la movilidad sostenible en la ciudad [21].

4.1.3.4.3. Análisis de patrones de tráfico y congestión

Al aplicar algoritmos de clustering a datos de tráfico en tiempo real, como la velocidad de los vehículos y la densidad de tráfico en diferentes puntos de la ciudad, es posible identificar patrones de congestión y flujo de tráfico. Esto puede ayudar a las autoridades de tráfico a tomar medidas proactivas para mitigar la congestión, como ajustar los tiempos de los semáforos o desviar el tráfico hacia rutas alternativas, contribuyendo así a una movilidad más fluida y sostenible [22].

4.1.3.4.4. Segmentación de áreas urbanas según su nivel de sostenibilidad

Utilizando datos ambientales, como la calidad del aire y las emisiones de carbono, junto con datos de movilidad, es posible aplicar técnicas de clustering para segmentar áreas urbanas en función de su nivel de sostenibilidad. Esto puede ayudar a identificar áreas que requieren medidas adicionales para mejorar su sostenibilidad y reducir su impacto ambiental, promoviendo así prácticas de movilidad más ecológicas y sostenibles [23].

5. ANTECEDENTES

5.1. Locales

5.1.1. Área Metropolitana del Valle de Aburrá

El Valle de Aburrá, una región geográfica conformada por diez municipios ha experimentado un notable crecimiento urbano y poblacional en los últimos años. Este crecimiento se ha traducido en una mayor concentración de personas y vehículos, especialmente en torno a la ciudad de Medellín, considerada el eje central de desarrollo de la zona. Con una población que alcanza los 4.119.008 habitantes, distribuidos mayormente en áreas urbanas, el valle enfrenta desafíos significativos en términos de movilidad y calidad del aire [2].

El aumento del parque automotor ha sido notable, con un incremento del 2.9% en menos de un año, según datos de la Secretaría de Movilidad de Medellín. Esta situación se ve agravada por la topografía desfavorable del valle, que dificulta la circulación del aire y contribuye al fenómeno de inversión térmica, generando problemas de contaminación atmosférica.

Para abordar estos desafíos, se han implementado diversas medidas, como el fortalecimiento del sistema de transporte público, la promoción de la movilidad sostenible en organizaciones públicas y privadas, y la adopción de planes para reducir la contaminación del aire. Estas acciones están respaldadas por iniciativas como el Plan Integral de Gestión de la Calidad del Aire en el Valle de Aburrá y los Planes de Movilidad Empresarial Sostenible (Planes MES), que buscan gestionar la demanda de transporte y promover alternativas más sostenibles.

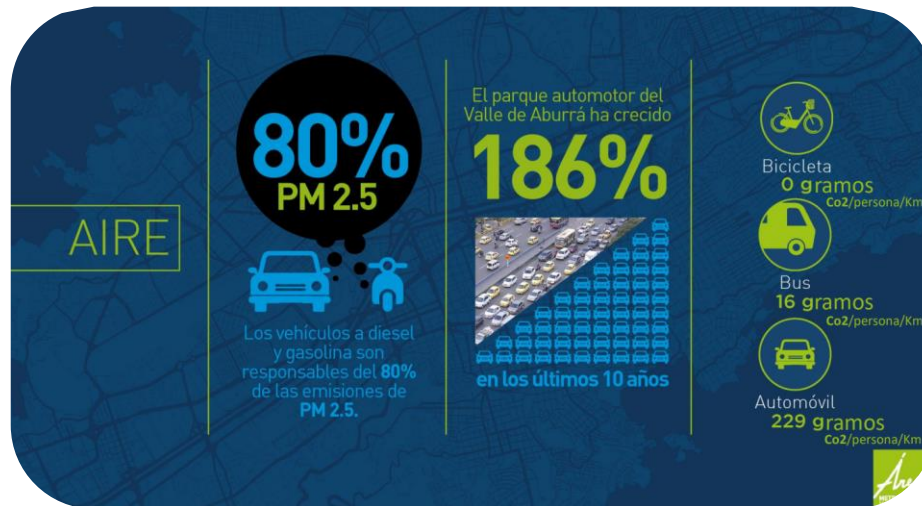


Figura 4. Emisiones de PM 2,5. en el Valle de Aburrá

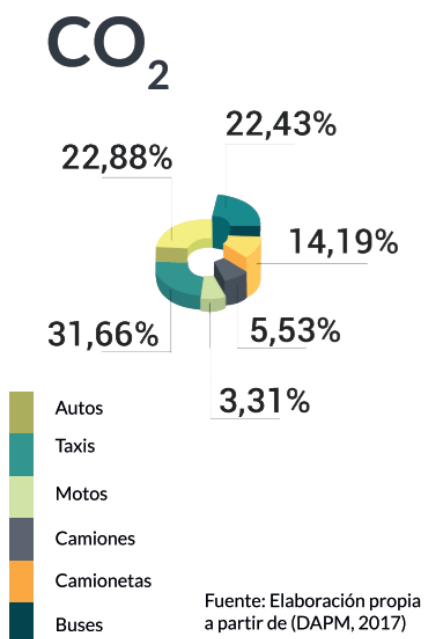
Fuente: AMVA, 2022

La **Figura 4**, muestra el contexto de movilidad en el Valle de Aburrá, demostrando que el 80% de las emisiones de PM2,5, provienen de vehículos motorizados, principalmente a diésel y gas.

5.2. Nacionales

5.2.1. Santiago de Cali, Colombia

En la ciudad de Santiago de Cali, la movilidad se caracteriza por una alta cantidad de viajes diarios, con un total de 3.603.297 viajes registrados en un día típico según la Encuesta de Movilidad de 2015. La mayoría de estos viajes se realizan en modos motorizados, siendo el transporte público responsable del 21,4% de los desplazamientos, con el Sistema Integrado de Transporte Masivo (SITM-MIO) como el principal medio utilizado. Por otro lado, el uso de vehículos particulares ha experimentado un notable aumento en los últimos años, con un incremento significativo en la tasa de motorización, que se ha más que duplicado en un período de 12 años, especialmente en el caso de las motocicletas.



***Figura 5.** Contribución de CO₂ por modo, en la ciudad de Cali*

La **Figura 5**, permite observar las emisiones de CO₂ provenientes de los diferentes modos de transporte en la ciudad de Cali, evidenciando que los modos que más generan emisiones son los taxis, autos y buses, respectivamente.

Fuente: Alcaldía de Santiago de Cali, 2021

Material particulado PM

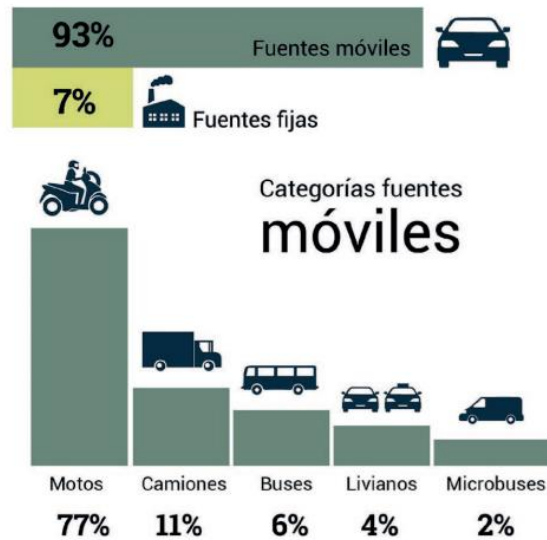


Figura 6. Contribución de PM 2.5 por modo, en la ciudad de Cali

Fuente: Alcaldía de Santiago de Cali, 2021

La **Figura 6** demuestra que frente a las emisiones de PM_{2,5} para la ciudad de Cali el 93% es generado por las fuentes móviles, que incluye categorías como: motos, camiones, buses, livianos y microbuses.

Por otro lado, se observa que la tasa de viajes por persona al día es de 1,52 viajes, siendo el retorno al hogar y los desplazamientos relacionados con trabajo o estudio las razones más comunes para viajar diariamente. Estos datos reflejan la complejidad y la diversidad de los patrones de movilidad en la ciudad, así como la necesidad de abordar los retos asociados al aumento del uso de vehículos particulares y promover alternativas de transporte más sostenibles y eficientes. Para lograr este objetivo, se plantea un cambio de paradigma en la planificación e inversión del transporte, que incluye acciones como evitar viajes innecesarios, promover medios de transporte más eficientes y

mejorar la tecnología y operación de los servicios de transporte [24].

5.2.2. Bogotá, Colombia

La movilidad en Bogotá enfrenta diversos desafíos que afectan tanto a los ciudadanos como al medio ambiente. Uno de los principales problemas es el tiempo excesivo en los desplazamientos, con un promedio de 65,44 minutos para llegar al lugar de trabajo, lo que equivale a casi 22 días al año. Esta congestión vehicular se suma a las largas distancias y las dificultades para acceder al transporte público, generando un impacto significativo en la calidad de vida de los habitantes de la ciudad.

Además, la contaminación atmosférica es una preocupación importante en Bogotá, con las fuentes móviles, especialmente los vehículos, contribuyendo en gran medida a las emisiones de contaminantes. Según la Secretaría Distrital de Ambiente, el transporte terrestre es responsable del 99% de las emisiones de carbono negro y el 98% de las emisiones totales de monóxido de carbono en la ciudad. Estas emisiones tienen efectos adversos en la salud de los ciudadanos, aumentando el riesgo de enfermedades respiratorias, cardiovasculares y diversos tipos de cáncer.

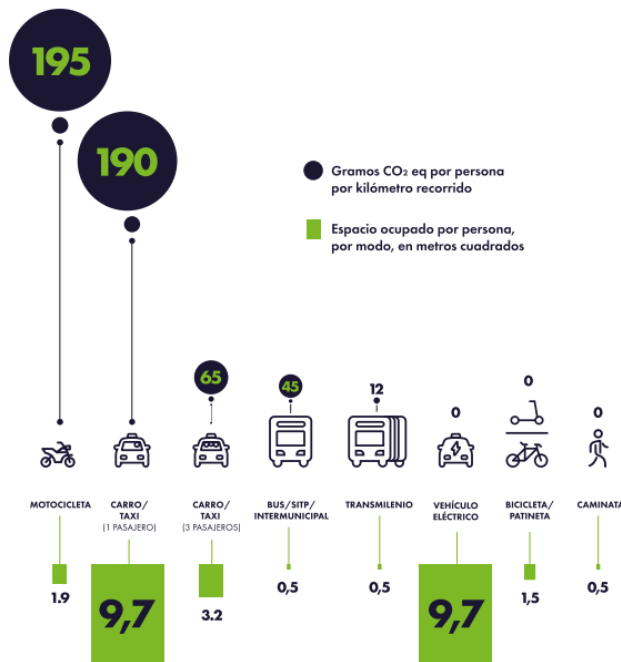


Figura 7. Comparación de emisiones y uso de espacio por modos

Fuente: López, C., 2023 [16].

La **Figura 7**, permite observar la proporción de emisiones de CO₂ eq que emite una persona de acuerdo con el modo de transporte en la ciudad de Bogotá, en la cual se evidencia que la motocicleta y el carro representan la mayor proporción.

Por otro lado, la siniestralidad vial es otro aspecto preocupante de la movilidad en Bogotá. En el año 2022, 558 personas perdieron la vida en siniestros viales y más de 21.000 resultaron lesionadas, principalmente debido a la velocidad de los vehículos motorizados. Estos accidentes representan una amenaza constante para la seguridad de los ciudadanos y generan impactos económicos y sociales significativos en la ciudad. En resumen, mejorar la movilidad en Bogotá no solo requiere medidas para abordar la congestión y mejorar la infraestructura vial, sino también

acciones para reducir la contaminación, mejorar la seguridad vial y promover formas más sostenibles de transporte.

5.3. Internacionales

5.3.1. Madrid, España

La Organización Mundial de la Salud reporta que aproximadamente 3 millones de muertes están relacionadas con la contaminación en todo el mundo. En el caso de España, según datos recopilados por la Agencia Europea del Medio Ambiente (AEMA), en el año 2013 fallecieron 23.940 personas debido a enfermedades relacionadas con las partículas contaminantes presentes en el aire. Además, 4.280 personas murieron por exposición al dióxido de nitrógeno (NO₂) y 1,760 más por sobreexposición al ozono (O₃), lo que suma un total de 29.980 muertes. Es importante destacar que la contaminación no causa directamente estas muertes, sino que agrava enfermedades preexistentes como las cardiovasculares, respiratorias y diversos tipos de cáncer [7].

La AEMA también informa que el 13% de las partículas contaminantes en los 28 países de la Unión Europea provienen del transporte por carretera. Extrapolando este porcentaje al número de muertes en España relacionadas con la contaminación, se estima que el transporte por carretera estaría implicado en 3.897 muertes. En contraste, según la Dirección General de Tráfico (DGT), el año pasado fallecieron 1.160 personas en accidentes de tráfico.

Además, se presenta la distribución modal del transporte en España, utilizando las ciudades de Madrid, Murcia y Vitoria como ejemplos representativos. Estas ciudades se seleccionaron por su diversidad en tamaño y sistemas de transporte público. Por ejemplo, Madrid, a pesar de tener una densidad poblacional alta y una extensión media, muestra un porcentaje

significativo de uso del transporte público. Sin embargo, como se observa en la **Figura 8** el parque automotor es el principal responsable de las emisiones contaminantes en Madrid, de las cuales, el 55% corresponden a PM_{2,5}. y el 40% a CO₂.

El tráfico rodado es el principal responsable de las emisiones de contaminantes atmosféricos más relevantes en Madrid:

51.4 % de NOx

55.1 % de PM_{2.5}

40.3 % de CO₂

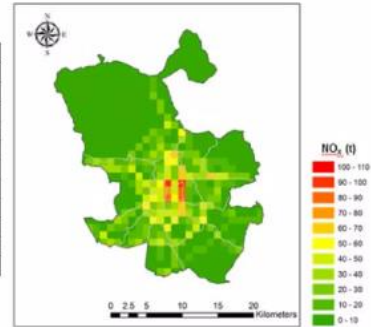
Seguido por otros modos de transporte

21,5% of NOx

Y combustión no industrial (RCI)

18% de NOx

Pollutant	Emission 2014	Ud.
CH ₄	133	t
CO	6.234	t
CO ₂	2.213	Mt
COVNM	2.863	t
N ₂ O	81	t
NH ₃	122	t
NO _x	7.012	t
PM ₁₀	572	t
PM _{2,5}	428	t
PST	754	t
SO ₂	14	t



Área de Medio Ambiente y Movilidad | MADRID

Figura 8. Emisiones contaminantes PM_{2,5} y CO₂, Madrid

Fuente: Motos, G., 2019

5.3.2. Lima, Perú

La movilidad en Lima y Callao presenta desafíos significativos debido a la congestión del tráfico y la falta de infraestructura adecuada. A pesar de contar con un sistema de transporte público compuesto por buses y minibuses, conocidos como "combis", este enfrenta problemas de saturación y falta de mantenimiento. Además, la presencia de taxis y servicios de transporte por aplicaciones móviles ofrece alternativas, aunque estos también se ven afectados por la congestión. La construcción de nuevas líneas de metro, como la Línea 2, busca aliviar la carga del tráfico y mejorar la movilidad dentro de la ciudad. Sin embargo, estos proyectos enfrentan desafíos de financiamiento y plazos de ejecución. En el caso del Callao, el tráfico se ve agravado por la actividad portuaria y la concentración de vehículos de carga, lo que hace que los desplazamientos

sean más complejos. A pesar de estos obstáculos, se están implementando medidas para mejorar la movilidad, como la promoción del uso de bicicletas y la expansión de carriles exclusivos para transporte público.

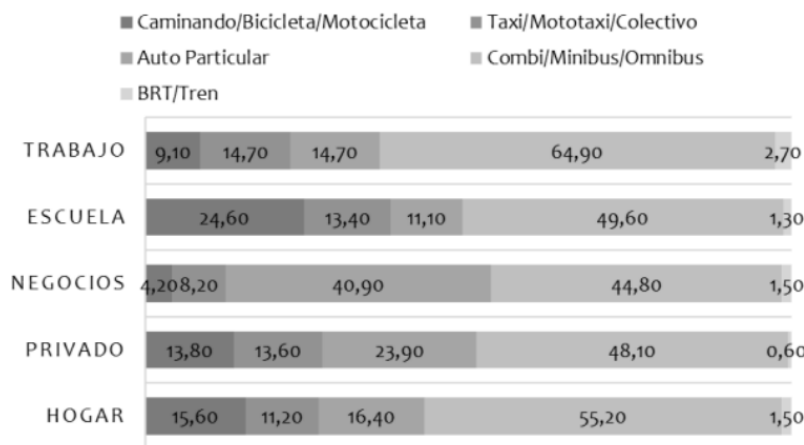


Figura 9. Propósito de viaje por modo de transporte

Fuente: Córdova, et. al., 2021[25]

El estudio realizado por Córdova-Chavez y Santa María-Dávila sobre los Planes de Desplazamiento de Empresa (PDE) en el Área Metropolitana de Lima y Callao y que se muestra en la **Figura 9**, se enmarca en un contexto global donde estas iniciativas han ganado relevancia como herramientas para gestionar la demanda de movilidad laboral de manera más sostenible.

El origen de los PDE se remonta a Estados Unidos durante la crisis del combustible en los años 70 y se extendió a Europa en los años 90, donde países como Holanda y Reino Unido adoptaron estas medidas como políticas oficiales para gestionar la movilidad urbana. En Europa, muchos países cuentan con una metodología normada para la implementación de los PDE, con requisitos específicos para empresas de diferentes tamaños y sectores.

En Latinoamérica, también se están realizando esfuerzos para gestionar la demanda de movilidad, como en Bogotá, donde se establecieron lineamientos para la formulación, adopción e implementación de los Planes Integrales de Movilidad Sostenible (PIMS) en organizaciones.

En el contexto de Lima y Callao, se destaca que el área de estudio concentra el mayor número de empresas del país, siendo el principal centro generador de empleo. Además, se evidencia una significativa movilidad diaria, con más de 22 millones de viajes diarios, de los cuales más del 65% se realizan en transporte público colectivo. Esto sugiere una alta dependencia del transporte público en la región, especialmente para los desplazamientos al trabajo. Este estudio involucra a 168 empleados distribuidos en 31 organizaciones del sector de la construcción en Lima y Callao. Se utilizaron encuestas y entrevistas para recopilar información tanto cuantitativa como cualitativa sobre los hábitos y comportamientos de movilidad de los empleados [25].

De acuerdo con la información anterior, se observa que a nivel local, nacional e internacional las ciudades enfrentan retos comunes que les permita garantizar su sostenibilidad, algunos de ellos son:

Contaminación y Emisiones: Los municipios que componen al Área Metropolitana del Valle de Aburrá, Santiago de Cali, Bogotá, Madrid y Lima/Callao enfrentan problemas graves de contaminación atmosférica debido al aumento del parque automotor. Las emisiones de CO₂ y partículas como PM_{2,5} son significativas, afectando la salud pública con enfermedades respiratorias y cardiovasculares.

Modos de Transporte: Hay una fuerte dependencia del transporte privado motorizado en todas las regiones estudiadas, con un crecimiento acelerado del parque vehicular, especialmente de

motocicletas. A pesar de los esfuerzos por mejorar el transporte público, como sistemas integrados y expansión de metros, persisten desafíos en capacidad y mantenimiento.

Desafíos Comunes: La congestión vehicular es una preocupación generalizada que afecta la calidad de vida y la eficiencia económica. También existe una alta incidencia de accidentes viales, destacándose cifras alarmantes en ciudades como Bogotá.

Iniciativas y Políticas: Se están implementando políticas para abordar estos desafíos, incluyendo planes de movilidad sostenible y regulaciones para controlar las emisiones contaminantes. Sin embargo, se requieren esfuerzos coordinados y sostenibles para mejorar la movilidad urbana y mitigar los impactos ambientales negativos.

Por lo tanto, este análisis destaca la necesidad urgente de soluciones efectivas y colaborativas para mejorar la calidad del aire, reducir la congestión y promover modos de transporte más sostenibles en estas áreas urbanas.

6. DESARROLLO

Esta investigación se basó en el método CRISP-DM (Cross Industry Standard Process for Data Mining) para desarrollar un modelo de aprendizaje no supervisado, específicamente modelos de clustering. Este enfoque estructurado comprendió diversas etapas. Se inició con la comprensión del problema y la identificación clara de los objetivos del proyecto. Luego, se procedió a reunir los datos necesarios para el entrenamiento del modelo, seguido de una fase de preparación de datos que incluyó la limpieza, normalización y selección de características relevantes. Posteriormente, se eligió el modelo de clustering más apropiado para abordar el problema específico y se procedió

a entrenarlo utilizando el conjunto de datos preparado. Una vez entrenado, se evaluó el modelo utilizando datos no vistos previamente para medir su rendimiento. Este proceso garantizó que el modelo desarrollado esté bien ajustado y pueda proporcionar insights valiosos sobre la movilidad urbana, como patrones de comportamiento, agrupaciones de usuarios o áreas con características similares en términos de movilidad y contaminación.

6.1. ENTENDIMIENTO DEL NEGOCIO

Los PLANES MES del Valle de Aburrá tienen como objetivo principal promover prácticas de movilidad sostenible entre las organizaciones, contribuyendo a la gestión integral de la calidad del aire y la movilidad urbana. Estos planes se enfocan en cinco áreas clave: movilidad activa, transporte público, reducción de viajes, racionalización del vehículo privado y cultura de movilidad sostenible.

La iniciativa ha logrado una alta participación del sector privado, con empresas de diferentes sectores económicos y con más de 200 empleados. Las encuestas realizadas a los trabajadores de cada empresa son esenciales para recopilar información detallada sobre sus patrones de movilidad, incluyendo preferencias de transporte, frecuencia de desplazamientos y modalidades utilizadas como transporte público, bicicleta o vehículo pTrivado. Sin embargo, se han identificado desafíos en la implementación, como la falta de infraestructura para peatones y ciclistas, así como deficiencias en el transporte público.

En términos de resultados objetivos, los PLANES MES han demostrado reducir las emisiones de CO₂ per cápita en el Valle de Aburrá. Además, han promovido el uso de la bicicleta y el teletrabajo, contribuyendo a la optimización del espacio público y la reducción del tráfico vehicular.

Las organizaciones participantes han experimentado beneficios como mejoras en la productividad, imagen corporativa, ambiente laboral y programas de sostenibilidad. Estos beneficios se ven respaldados por incentivos como descuentos en impuestos locales para aquellas organizaciones que cumplen con las metas de reducción de emisiones.

Finalmente, el modelo de clustering aplicado en la investigación permitió identificar patrones de comportamiento y agrupaciones significativas en relación con la movilidad urbana, facilitando la toma de decisiones estratégicas basadas en datos concretos. Sin embargo, las lecciones aprendidas sugieren que la replicación de estos planes en otras ciudades requiere un marco normativo claro, colaboración interinstitucional, apoyo técnico continuo y la implementación de incentivos adecuados para asegurar la adopción y el éxito del programa.

6.2. ENTENDIMIENTO DE LOS DATOS

6.2.1. Descripción de las Encuestas

Se utilizaron dos encuestas principales denominadas "EncuestasTrabajadores2022_1" y "EncuestasTrabajadores2022_2", cada una con su propia estructura y conjunto de preguntas. La *Figura 10*, muestra la estructura de la base de datos inicial entregada por el Área Metropolitana del Valle de Aburrá.

ID_TERCERO	ENCUESTA	ENCUESTA_NO	GRUPO	NO_PREGUNTA	PREGUNTA	VALOR_NUMERO	S_VALOR	X_VALOR	Y_VALOR	OPCION	OPCIONES
175	ENCUESTA DE TRABAJADORES-2022-1	Colaborador 0001	1. DATOS GENERALES	1	Nombre		Astrid carolina				
175	ENCUESTA DE TRABAJADORES-2022-1	Colaborador 0001	1. DATOS GENERALES	2	Apellido		Alvarez Montoya				
175	ENCUESTA DE TRABAJADORES-2022-1	Colaborador 0001	1. DATOS GENERALES	3	Tipo de documento					CÉDULA DE CIUDADANÍA (CC)	
175	ENCUESTA DE TRABAJADORES-2022-1	Colaborador 0001	1. DATOS GENERALES	4	Número de documento	1,15E+09					
175	ENCUESTA DE TRABAJADORES-2022-1	Colaborador 0001	1. DATOS GENERALES	5	Área o departamento de la organización a la que pertenece						
175	ENCUESTA DE TRABAJADORES-2022-1	Colaborador 0001	1. DATOS GENERALES	6	Tipo de trabajador					ADMINISTRATIVO	
175	ENCUESTA DE TRABAJADORES-2022-1	Colaborador 0001	1. DATOS GENERALES	7	Municipio						
175	ENCUESTA DE TRABAJADORES-2022-1	Colaborador 0001	1. DATOS GENERALES	8	Dirección		Calle 45c 6c 60;MEDELLIN	- 0.00003989438 519040332	4.83028180530 1505e-7		
175	ENCUESTA DE TRABAJADORES-2022-1	Colaborador 0001	2. INFORMACIÓN SOCIOECONÓMICA	1	Sexo					FEMENINO	
175	ENCUESTA DE TRABAJADORES-2022-1	Colaborador 0001	2. INFORMACIÓN SOCIOECONÓMICA	2	Si se encuentra usted en situación de discapacidad, por favor reporte cuál tiene (en caso de contar con más de dos discapacidades, favor reportar la predominante):					NINGUNA	
175	ENCUESTA DE TRABAJADORES-2022-1	Colaborador 0001	2. INFORMACIÓN SOCIOECONÓMICA	3	Edad					25 A 34 AÑOS	

Figura 10. Base de datos entregada por el AMVA

Fuente: Elaboración propia

6.2.2. Carga y Transformación Inicial

Los datos son analizados inicialmente a través de R Studio, utilizando librerías como readxl, dplyr, reshape2, gtools, stringr y scipen=999, con el propósito de generar una estructura más clara para la preparación de los datos.

6.2.2.1. Transformación de Columnas

Se realizaron transformaciones específicas en la columna "Dirección", concatenando valores adicionales como longitud, latitud y NIT.

6.2.2.2. Selección y Renombrado de Variables

Se seleccionaron y renombraron las variables de interés para ambas encuestas. Este paso fue

crucial para mantener consistencia y claridad en los datos.

6.2.2.3.División y Agrupamiento de Datos

Se implementó una función para dividir los dataframes en grupos según cambios en la variable "GRUPO", lo cual permitió manejar los datos de manera más estructurada.

6.2.2.4.Transposición y Combinación de Datos

Se transpusieron los dataframes para facilitar la consolidación y se combinaron utilizando la función smartbind para obtener un solo dataframe consolidado para cada encuesta.

6.2.3. Integración de Encuestas

6.2.3.1.Combinar Encuestas

Se unieron los dataframes consolidados de las encuestas "EncuestasTrabajadores2022_1" y "EncuestasTrabajadores2022_2" en un solo dataframe final.

6.2.3.2.Enriquecimiento con Información Adicional

Se integraron los datos de la "EncuestaSitio" para añadir información sobre la cantidad de colaboradores y la infraestructura de teletrabajo.

6.2.4. Resultados y Almacenamiento

El resultado final de este proceso es un dataframe consolidado y enriquecido, que se muestra en la **Tabla 2**, almacenado en un archivo de Excel denominado "Base_final.xlsx". Este archivo contiene:

- **DF_ResultadoFinal:** La base de datos final con la información consolidada de todas las encuestas.
- **Base_Cias:** Información específica de las compañías extraída de la "EncuestaSitio".

Tabla 2. Descripción de las variables entregadas por el negocio

Variable	Descripción
NIT	NIT de la empresa
Nombre	Nombre del empleado
Apellido	Apellido del empleado
TipoDoc	Tipo de documento
NumDoc	Número de documento
TipoTrabajador	Tipo de trabajador
Dirección	Dirección del empleado
Género	Género del empleado
TipoDiscapacidad	Si se encuentra usted en situación de discapacidad por favor reporte cuál tiene (en caso de contar con más de dos discapacidades favor reportar la predominante):
Edad	Edad del empleado
Estrato	Estrato socio-económico
HoraInicioLabor	Seleccione la hora de inicio de su jornada laboral
HoraSalidaLabor	Seleccione la hora de salida de su jornada laboral
TeletrabajoCompleto	¿Realiza usted teletrabajo o trabajo en casa durante TODA la semana laboral?
TeletrabajoHibrido	¿Realiza usted teletrabajo o trabajo en casa con alternancia durante su semana laboral?
FreqTeletrabajo	Con qué frecuencia durante la semana laboral realiza trabajo en casa con

Variable	Descripción
	alternancia
HoraInicioViaje	Seleccione la hora de inicio de su viaje al trabajo
HoraFinViaje	Seleccione la hora de llegada a la organización (oficina/lugar de trabajo)
ModoLlegadaLabor	Seleccione todos aquellos modos que usted utiliza para cada etapa o tramo del viaje hasta su lugar de trabajo
OcupantesAuto	Seleccione el número de ocupantes que viajan en el automóvil
CombustibleAuto	Seleccione el combustible del automóvil
CilindrajeAuto	Seleccione el cilindraje del automóvil
ModeloAuto	Seleccione el modelo del automóvil
CilindrajeMoto	Seleccione el cilindraje de la motocicleta
ModeloMoto	Seleccione el modelo de la motocicleta
FreqSem.ModosLlegadaLabor	Durante una semana típica laboral por favor conteste cuántas veces o días a la semana utiliza el o los modos seleccionados anteriormente para llegar a su lugar de trabajo
RazonModoLlegadaLabor	Por favor Seleccione las razones más importantes para elegir su modo de transporte actual para ir a trabajar.
GastoDiarioDesplazamiento	¿Cuánto dinero (en pesos) se gasta en promedio al día en sus desplazamientos desde el hogar hasta el trabajo y viceversa?
OtroModoTransporte	¿Qué otras modalidades de transporte / prácticas de trabajo está interesado en usar aparte de lo que usa actualmente? por favor Seleccione tantas como

Variable	Descripción
	deseo.
Part.EstrategiasMovilidad	De las estrategias de movilidad sostenible ejecutadas por la organización por favor indique en cuales participó
Longitud	Coordenadas espaciales
Latitud	Coordenadas espaciales
CantColaboradores	Número total de colaboradores en la empresa
Infra.TeleTrabajo	¿Qué infraestructura tiene la compañía para permitir el teletrabajo?

Fuente: Elaboración propia

Teniendo en cuenta que el objetivo de esta tesis es analizar la sostenibilidad en la movilidad urbana del Valle de Aburrá, se incluyó también una base de datos proporcionada por la subdirección ambiental del AMVA, que normalmente es utilizada por la entidad para estimar las emisiones generadas por fuentes móviles. Esta base de datos incluye variables como el combustible utilizado, cilindraje, modelo, factores de emisión de CO₂ y PM 2.5, velocidad promedio, rendimiento del combustible y costo del viaje. Los modos de transporte cubiertos en esta base de datos incluyen los que se relacionan en la **Tabla 3**.

Tabla 3. *Modos de transporte reconocidos para cálculo de emisiones*

Modos de transporte con emisiones relacionadas
Motocicleta 2t individual
Motocicleta 4t individual
A pie/caminata

Modos de transporte con emisiones relacionadas

Alimentador del metro

Automóvil compartido

Automóvil individual

Bicicleta eléctrica

Bicicleta particular

Bicicleta pública

Bus/colectivo

Metro

Metrocable

Metroplús

Motocicleta 2t compartida

Motocicleta 4t compartida

Motocicleta eléctrica

Patinetas, scooters o similares

Ruta empresarial

Taxi o similares

Teletrabajo

Tranvía

Fuente: Elaboración propia

El dataframe consolidado ahora contiene columnas adicionales que capturan estos factores de

emisión y otros detalles relevantes, proporcionando una visión más completa del impacto ambiental y económico del transporte utilizado por los trabajadores. La estructura de la base de datos final incluye las siguientes variables adicionales:

- **Factor de emisión (g CO₂/km) per cápita**
- **Factor de emisión (g PM_{2,5}/km) per cápita**
- **Velocidad promedio (km/min)**
- **Rendimiento del combustible (Gal/Km)**

6.3. PREPARACIÓN DE LOS DATOS

Para la preparación de los datos, se incluyeron únicamente las variables de interés para el modelo de clustering, las cuales se relacionan en el numeral 6.3.1.

6.3.1. Datos faltantes

	N_nas	Porcentaje
FreqTeletrabajo	32726	85.98
Rendimiento del combustible (Gal/Km)	15097	39.66
Factor de emisión (g PM 2.5/km)	15097	39.66
Velocidad promedio (km/min)	15097	39.66
Factor de emisión (g CO2/km)	15096	39.66
FreqSem.ModollegadaLabor	3178	8.35
HoraFinViaje	3168	8.32
HorainicioViaje	3167	8.32
OtroModoTransporte	498	1.31
RazonModollegadaLabor	333	0.87
GastoDiarioDesplazamiento	179	0.47
TipoTrabajador	158	0.42
Sexo	154	0.40
Estrato	153	0.40
Edad	152	0.40
HoraSalidaLabor	151	0.40
HorainicioLabor	150	0.39
TeletrabajoHibrido	139	0.37
TeletrabajoCompleto	139	0.37
ModollegadaLabor	0	0.00

Figura 11. Análisis de datos faltantes

Fuente: Elaboración propia

La *Figura 11* permite observar que la variable "FreqTeletrabajo" presenta un 85,98% de datos faltantes. Debido a las siguientes razones:

- 1. Trabajadores 100% Presenciales:** Una gran cantidad de personas en la muestra trabajan de forma completamente presencial. Para ellos, la pregunta sobre la frecuencia de teletrabajo no aplica, por lo que no respondieron.
- 2. Trabajadores 100% en Teletrabajo:** De manera similar, quienes trabajan de forma

completamente remota tampoco respondieron, ya que la pregunta se enfoca en la alternancia entre teletrabajo y trabajo presencial.

Estas dos situaciones explican el elevado porcentaje de registros faltantes en "FreqTeletrabajo". Dado el volumen de datos ausentes, se decidió excluir esta variable del análisis de clustering, ya que su inclusión podría sesgar los resultados.

Análisis de datos faltantes en otras variables:

Medio de Transporte, Frecuencia de Uso, y Horarios de Viaje: Las variables relacionadas con el transporte también muestran datos faltantes, aunque en menor medida. Estos faltantes se asocian con personas que declararon realizar teletrabajo completo. En estos casos, los valores faltantes se reemplazaron con "No Aplica", y sí se incluirán en el análisis.

Gasto Diario de Desplazamiento: Para esta variable, el 0,47% (179 casos) son datos faltantes, los cuales fueron reemplazados con "0", indicando que la persona no incurre en gastos de desplazamiento.

Otras Variables: Para variables con un porcentaje de faltantes menor al 1,31%, se decidió omitir estos registros por su baja representación en el total de datos.

Variables excluidas:

Otro Modo de Transporte: Esta variable fue excluida del análisis debido a la gran diversidad de respuestas (4.186 categorías diferentes) y su poca relevancia para el objetivo principal del estudio. Por lo tanto, después del pre-procesamiento de los datos faltantes, la base de datos para el análisis de clúster queda compuesta por 37.290 registros y 18 atributos, tal cual como se puede observar en la *Figura 12*.

▼ Tipo de datos

Base_cluster.dtypes		
🔍	TipoTrabajador	object
	Sexo	object
	Edad	object
	Estrato	float64
	HoraInicioLabor	object
	HoraSalidaLabor	object
	TeletrabajoCompleto	object
	TeletrabajoHibrido	object
	HoraInicioViaje	object
	HoraFinViaje	object
	ModoLlegadaLabor	object
	FreqSem.ModoLlegadaLabor	object
	RazonModoLlegadaLabor	object
	GastoDiarioDesplazamiento	float64
	Velocidad promedio (km/min)	float64
	Factor de emisión (g CO2/km)	float64
	Factor de emisión (g PM 2.5/km)	float64
	Rendimiento del combustible (Gal/Km)	float64

Figura 12. Tipo de datos

Fuente: Elaboración propia

6.3.2. Análisis exploratorio de variables

En este análisis exploratorio de datos, se analiza detalladamente la distribución de diversas características de los empleados encuestados, especialmente en relación con sus modalidades de trabajo y atributos socio demográficos.

6.3.2.1. Teletrabajo

TeletrabajoHibrido	NO	SI	Total
TeletrabajoCompleto			
NO	29430	5196	34626
SI	2935	16	2951
Total	32365	5212	37577

Figura 13. Frecuencia de empleados por modalidad de trabajo

Fuente: Elaboración propia

De acuerdo con las frecuencias cruzadas, se puede destacar que hay 29.184 personas que no tienen ni teletrabajo completo ni híbrido, por lo que se supondrá que esta cantidad de personas, que representan el 78,26% del total, tienen trabajo 100% presencial.

Por su parte, 2.937 (7,88%) registran teletrabajo completo y 5.185 (13,91%) teletrabajo híbrido.

6.3.2.2. Estrato socioeconómico

	Estrato	Cantidad	Porcentaje
0	3.00	13598	36.19
1	2.00	13550	36.06
2	4.00	4316	11.49
3	1.00	2840	7.56
4	5.00	2270	6.04
5	6.00	1003	2.67

Figura 14. Frecuencia de empleados por estrato socioeconómico

Fuente: Elaboración propia

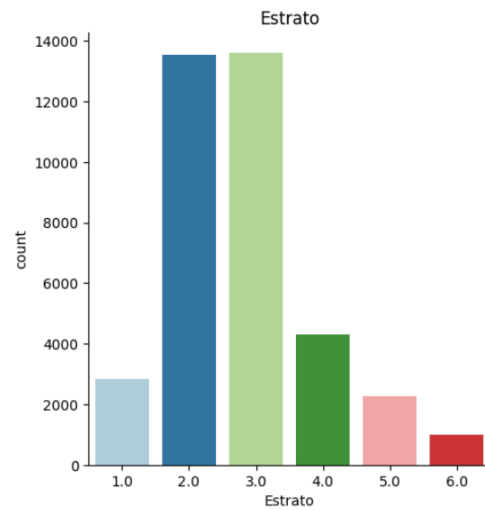


Figura 15. Distribución de empleados por estrato socioeconómico

Fuente: Elaboración propia

6.3.2.3. Género

	Género	Cantidad	Porcentaje
0	MASCULINO	21200	56.42
1	FEMENINO	16377	43.58

Figura 16. Frecuencia de empleados por género

Fuente: Elaboración propia

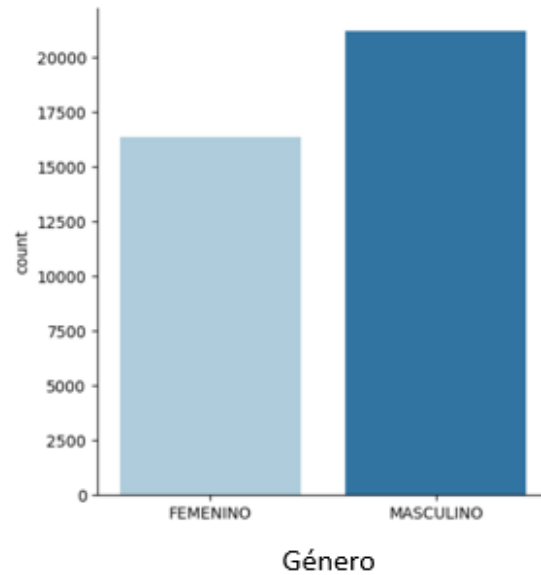


Figura 17. Distribución de empleados por género

Fuente: Elaboración propia

6.3.2.4. Edad

	Edad	Cantidad	Porcentaje
0	25 A 34 AÑOS	13168	35.04
1	35 A 44 AÑOS	10836	28.84
2	45 A 54 AÑOS	6249	16.63
3	MENOR DE 25 AÑOS	4447	11.83
4	55 O MÁS	2877	7.66

Figura 18. Frecuencia de empleados por edad

Fuente: Elaboración propia

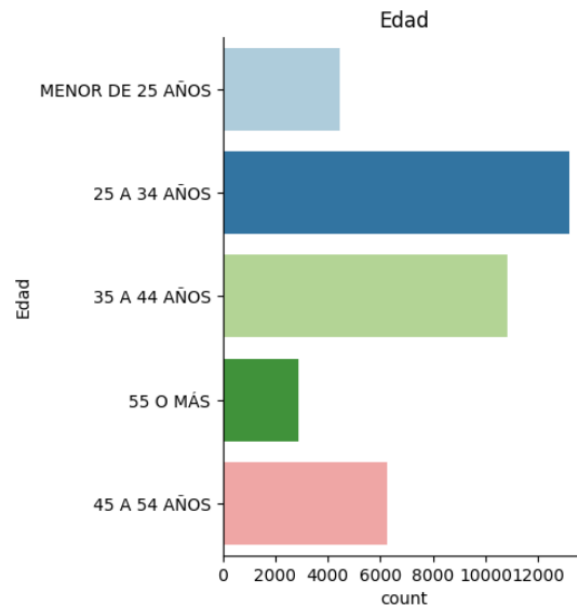


Figura 19. Distribución de empleados por edad

Fuente: Elaboración propia

6.3.2.5. Tipo de empleado

	TipoTrabajador	Cantidad	Porcentaje
0	OPERATIVO	21715	57.79
1	ADMINISTRATIVO	15862	42.21

Figura 20. Frecuencia de empleados por tipo de cargo

Fuente: Elaboración propia

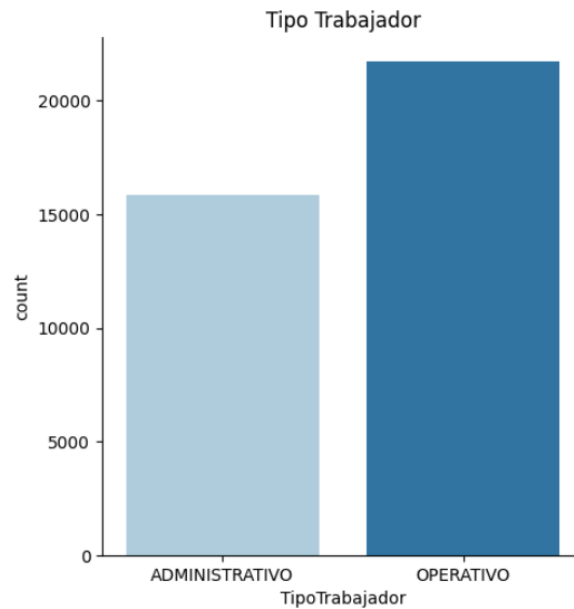


Figura 21. Frecuencia de empleados por tipo de cargo

Fuente: Elaboración propia

En principio, se destaca que los atributos correspondientes a tipo de cargo, estrato socioeconómico, género y edad del empleado no presentan inconsistencias en sus categorías. De esta forma, se puede observar que:

- La mayoría de los empleados pertenecen a los estratos socioeconómicos 2 y 3, representando aproximadamente el 72,22% del total de registros. Y tan solo el 8,72% pertenecen a los estratos 5 y 6.
- Más de la mitad de los empleados son hombres, precisamente el 56,48%. El otro 43,52% son mujeres.

- La mayor parte de empleados, el 35,1% tienen edades entre 25 a 34 años, seguido de los que tienen entre 35 y 44 años, con una representación del 28,84%. Tan solo el 7,63% de todos los empleados tienen 55 años o más.
- La mayoría de los empleados, precisamente, el 57,77%, tienen un trabajo operativo y el otro 42,23% un trabajo administrativo.

6.3.2.6. Gasto de desplazamiento

	TeletrabajoHibrido	NO	SI	Total
TeletrabajoCompleto				
NO		4991	389	5380
SI		1573	13	1586
Total		6564	402	6966

Figura 22. Frecuencia de empleados que no relacionan gasto

Fuente: Elaboración propia

La cantidad de empleados que no registran gasto diario en su desplazamiento corresponde a 6.966 de ellos, que representan el 19% del total, de estos, hay 1.573 que tienen teletrabajo completo, y 4.991 que trabajan de forma presencial.

GastoDiarioDesplazamiento	
count	37577.00
mean	9659.40
std	268969.95
min	-32000.00
25%	2000.00
50%	5500.00
75%	8000.00
max	51605160.00

Figura 23. Resumen estadístico gasto de desplazamiento

Fuente: Elaboración propia

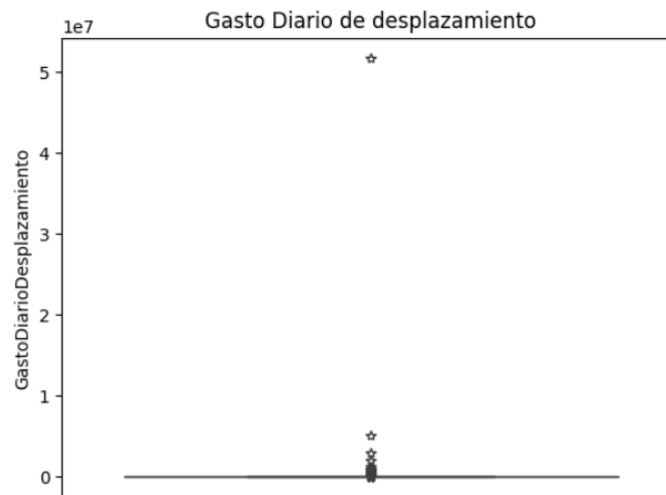


Figura 24. Boxplot gasto de desplazamiento

Fuente: Elaboración propia

El gasto diario de desplazamiento presenta valores negativos en 54 casos, los cuales son reemplazados por cero. Así mismo hay presencia de valores atípicos que superan los \$200.000

COP diarios, incluso llegando a un registro de más de \$50'000.000.000 COP, los cuales son reemplazados con el límite de acuerdo con el criterio de Tukey que dice que:

Un dato se considera atípico si está por debajo de $Q1 - 1,5 * IQR$ o por encima de $Q3 + 1,5 * IQR$, donde Q1 es el primer cuartil, Q3 es el tercer cuartil y IQR es la diferencia entre ellos: $IQR = Q3 - Q1$. La fórmula para detectar valores atípicos según el criterio de Tukey es:

- **Límite inferior = $Q1 - 1,5IQR$**
- **Límite superior = $Q3 + 1,5IQR$**

Por lo tanto, después de aplicar este criterio, se obtiene que el límite considerado por el criterio de Tukey es un gasto de \$17.000 COP diarios, después de este monto se considera atípico. Con respecto a esto, hay 2.379 registros que representan el 6,38%, al ser atípicos estos son reemplazados por el límite hallado, es decir, \$17.000 COP y de esta forma evitar atipicidades en este atributo.

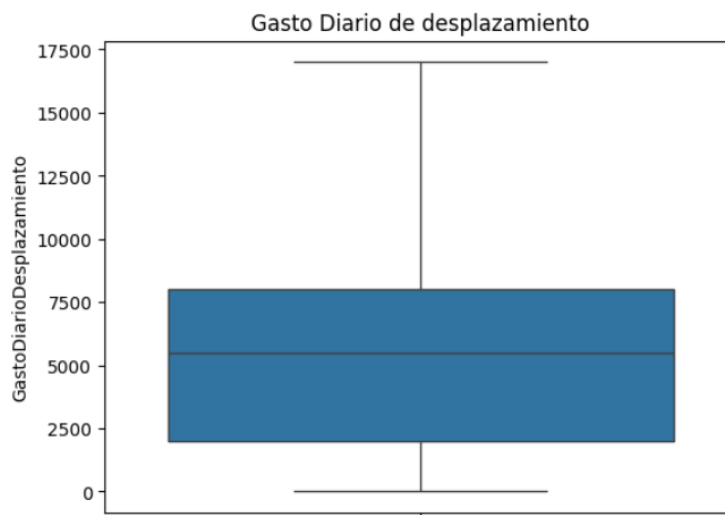


Figura 25. *Boxplot sin datos atípicos del gasto de desplazamiento*

Fuente: Elaboración propia

GastoDiarioDesplazamiento	
count	37577.00
mean	5950.64
std	4785.22
min	0.00
25%	2000.00
50%	5500.00
75%	8000.00
max	17000.00

Figura 26. Resumen estadístico sin datos atípicos del gasto de desplazamiento

Fuente: Elaboración propia

En la Figura 25 y Figura 26, se observa que ya no hay presencia de registros atípicos en el gasto diario de desplazamiento, y que el máximo gasto corresponde a \$17.000 COP, con un gasto promedio de aproximadamente \$6.000 COP, y tan solo el 25% de los empleados gastan más de \$8.000 COP pesos diarios por trayecto o desplazamiento.

6.3.2.7. Modo de llegada al trabajo

La variable correspondiente al modo de transporte utilizado para llegar al trabajo contiene 1.272 categorías diferentes. Las cuales deben normalizarse con respecto a las categorías utilizadas por la subdirección ambiental del AMVA de modo que puedan ser cruzadas con la información del inventario de emisiones que utiliza la entidad. Por lo tanto, teniendo en cuenta dicha condición y que, además, existe un alto número de categorías, se decide agrupar esta variable utilizando las 21 categorías entregadas por el AMVA, las demás combinaciones de modos de llegada al trabajo son recategorizadas como "Otros".

	ModoLlegadaLabor	Cantidad	Porcentaje
0	OTROS	14969	39.84
1	MOTOCICLETA 4T INDIVIDUAL	5194	13.82
2	AUTOMÓVIL INDIVIDUAL	3088	8.22
3	TELETRABAJO	2976	7.92
4	BUS/COLECTIVO	2888	7.69
5	A PIE/CAMINATA	2185	5.81
6	METRO	1476	3.93
7	AUTOMÓVIL COMPARTIDO	1197	3.19
8	MOTOCICLETA 4T COMPARTIDA	928	2.47
9	RUTA EMPRESARIAL	846	2.25
10	BICICLETA PARTICULAR	820	2.18
11	TAXI O SIMILARES	311	0.83
12	MOTOCICLETA 2T INDIVIDUAL	238	0.63
13	ALIMENTADOR DEL METRO	126	0.34
14	METROPLÚS	89	0.24
15	MOTOCICLETA 2T COMPARTIDA	87	0.23
16	BICICLETA ELÉCTRICA	75	0.20
17	MOTOCICLETA ELÉCTRICA	27	0.07
18	PATINETAS, SCOOTERS O SIMILARES	25	0.07
19	BICICLETA PÚBLICA	14	0.04
20	METROCABLE	11	0.03
21	TRANVÍA	7	0.02

Figura 27. Frecuencia de empleados por modo de transporte

Fuente: Elaboración propia

La variable correspondiente al medio de transporte para la llegada al trabajo queda ya con 22 categorías y se resalta una buena representación del uso de motocicleta y automóvil individual, representando el 13,85% y 8,22% del total de empleados respectivamente. Quedan 39,81% bajo la categoría de "Otros" modos de transporte o etapas de viaje.

6.3.2.8. Razón de uso del modo de llegada al trabajo

	RazonModoLlegadaLabor	Cantidad	Porcentaje	Porcentaje acumulado
0	COMODIDAD , RAPIDEZ, COSTO	3790	10.09	10.09
1	COMODIDAD , RAPIDEZ	2549	6.78	16.87
2	RAPIDEZ, COSTO	2436	6.48	23.35
3	RAPIDEZ	2223	5.92	29.27
4	COMODIDAD , RAPIDEZ, COSTO, CONVENIENCIA	2118	5.64	34.91
5	CONVENIENCIA	1885	5.02	39.93
6	COSTO	1833	4.88	44.81
7	COMODIDAD	1612	4.29	49.10
8	OTRO	1431	3.81	52.91
9	COMODIDAD , RAPIDEZ, SEGURIDAD PERSONAL	1010	2.69	55.60

Figura 28. Frecuencia de empleados por razón de uso modo de transporte

Fuente: Elaboración propia

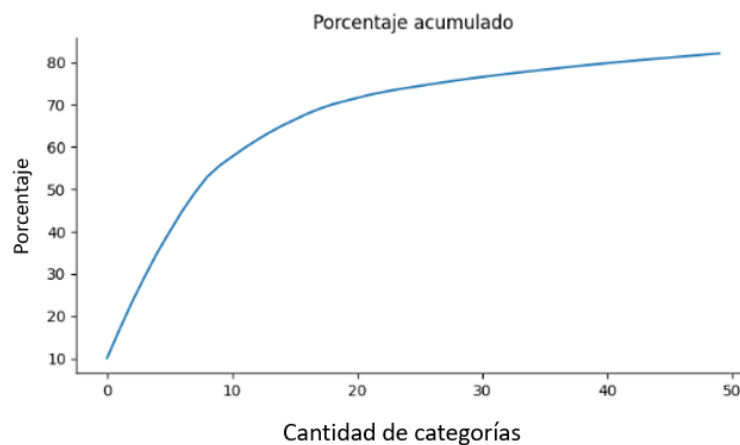


Figura 29. Porcentaje acumulado razón de uso modo de transporte

Fuente: Elaboración propia

La variable correspondiente a las razones de uso del modo de transporte cuenta con 807 categorías diferentes. Por este motivo se toman únicamente las primeras 19 categorías más frecuentes, dado

que acumulan aproximadamente el 70% de todos los registros y a partir de 18 categorías la ganancia de representación porcentual no es considerable. Las demás categorías se recategorizan como "Otros".

RazonModoLlegadaLabor	Cantidad	Porcentaje
0 OTROS	11248	29.93
1 COMODIDAD , RAPIDEZ, COSTO	3790	10.09
2 COMODIDAD , RAPIDEZ	2549	6.78
3 RAPIDEZ, COSTO	2436	6.48
4 RAPIDEZ	2223	5.92
5 COMODIDAD , RAPIDEZ, COSTO, CONVENIENCIA	2118	5.64
6 CONVENIENCIA	1885	5.02
7 COSTO	1833	4.88
8 COMODIDAD	1612	4.29
9 OTRO	1431	3.81
10 COMODIDAD , RAPIDEZ, SEGURIDAD PERSONAL	1010	2.69
11 COMODIDAD , RAPIDEZ, CONVENIENCIA	793	2.11
12 COSTO, CONVENIENCIA	774	2.06
13 RAPIDEZ, COSTO, CONVENIENCIA	719	1.91
14 COMODIDAD , COSTO	666	1.77
15 COMODIDAD , RAPIDEZ, COSTO, SEGURIDAD PERSONAL	596	1.59
16 RAPIDEZ, CONVENIENCIA	530	1.41
17 COMODIDAD , RAPIDEZ, COSTO, CONVENIENCIA, SEGU...	527	1.40
18 COMODIDAD , CONVENIENCIA	452	1.20
19 COMODIDAD , SEGURIDAD PERSONAL	385	1.02

Figura 30. Frecuencia ajustada razón de uso modo de transporte

Fuente: Elaboración propia

La variable razón del uso del modo de transporte para la llegada al trabajo queda con 20 categorías, destacando las principales razones, las cuales Comodidad, Rapidez y Costo corresponden a la mayor representación porcentual.

6.3.2.9. Frecuencia de uso semanal del modo de transporte

	FreqSem.ModoLlegadaLabor	Cantidad	Porcentaje
0	SEIS (6) VECES A LA SEMANA	13723	36.52
1	CINCO (5) VECES A LA SEMANA	9308	24.77
2	TODOS (7) LOS DÍAS DE LA SEMANA	4292	11.42
3	No Aplica	2966	7.89
4	CUATRO (4) VECES A LA SEMANA	2908	7.74
5	TRES (3) VECES A LA SEMANA	1751	4.66
6	DOS (2) VECES A LA SEMANA	1742	4.64
7	UNA (1) VEZ A LA SEMANA	887	2.36

Figura 31. Frecuencia de uso semanal del modo de transporte

Fuente: Elaboración propia

Durante una semana típica laboral, es decir 6 veces a la semana, la mayoría de los empleados, precisamente el 36,6% usan el o los modos seleccionados anteriormente para llegar a su lugar de trabajo. Esto seguido de 5 veces a la semana con el 24,68%. Tan solo el 2,36% utiliza los medios de desplazamiento una vez a la semana.

6.3.2.10. Duración de la jornada laboral

DuraciónJornadaLaboral	
count	37577.00
mean	9.30
std	2.35
min	0.00
25%	8.00
50%	10.00
75%	10.00
max	22.00

Figura 32. Resumen estadístico duración jornada laboral

Fuente: Elaboración propia

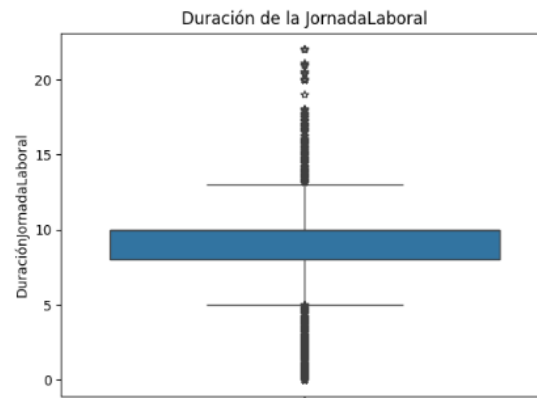


Figura 33. *Boxplot duración jornada laboral*

Fuente: Elaboración propia

Para este caso de valores atípicos, se opta por no eliminar los datos que representan esta condición dado que algunos empleados que registren menos de 5 horas podrían estar trabajando a medio tiempo o bajo un esquema por horas. Del mismo modo, aquellos que trabajen más de 13 horas podrían estar realizando turnos dobles o desempeñando labores que requieren una mayor cantidad de horas.

6.3.2.11. Hora de inicio laboral

Dada la alta cantidad de categorías en la hora de inicio laboral, se decide recategorizar estas horas en intervalos de tiempo de una hora.

	HoraInicioLabor	Cantidad	Porcentaje
0	07:00 - 08:00	15099	40.18
1	06:00 - 07:00	9399	25.01
2	08:00 - 09:00	3674	9.78
3	14:00 - 15:00	2097	5.58
4	05:00 - 06:00	2074	5.52
5	13:00 - 14:00	1232	3.28
6	10:00 - 11:00	902	2.40
7	12:00 - 13:00	626	1.67
8	11:00 - 12:00	565	1.50
9	09:00 - 10:00	499	1.33

Figura 34. Frecuencia de empleados por hora de inicio laboral

Fuente: Elaboración propia

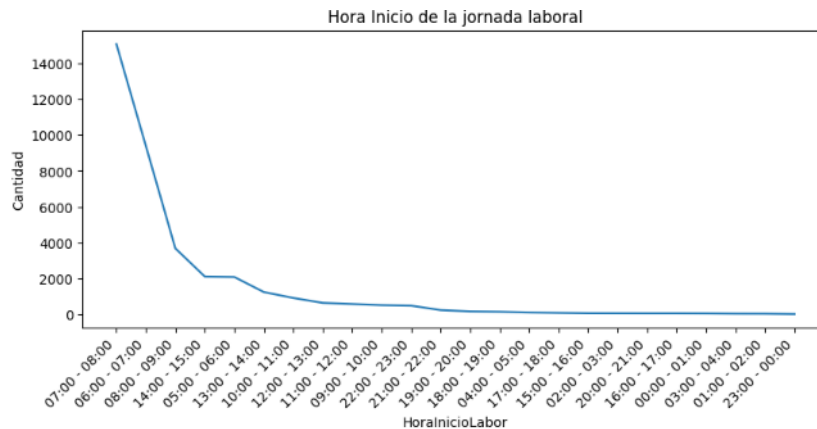


Figura 35. Gráfico de línea para la hora de inicio laboral

Fuente: Elaboración propia

La gran mayoría de empleados, el 40,19%, inician su jornada laboral entre las 7 y 8 de la mañana, seguido del 25% de empleados que inician entre las 6 y 7 am. Que corresponden a las jornadas de inicio laboral más frecuentes.

6.3.2.12. Hora de salida laboral

	HoraSalidaLabor	Cantidad	Porcentaje
0	17:00 - 18:00	12375	32.93
1	16:00 - 17:00	5083	13.53
2	14:00 - 15:00	4205	11.19
3	18:00 - 19:00	3818	10.16
4	22:00 - 23:00	2547	6.78
5	19:00 - 20:00	1791	4.77
6	21:00 - 22:00	1782	4.74
7	13:00 - 14:00	1342	3.57
8	05:00 - 06:00	1027	2.73
9	15:00 - 16:00	941	2.50

Figura 36. Frecuencia de empleados por hora de salida laboral

Fuente: Elaboración propia



Figura 37. Gráfico de línea para la hora de salida laboral

Fuente: Elaboración propia

Por su parte, la gran mayoría de empleados, precisamente el 32,91%, terminan su jornada laboral

entre las 5:00 y 6:00 pm, seguido del 13,54% que terminan su trabajo entre las 4:00 y 5:00 pm.

De acuerdo con los resultados, se puede concluir que las jornadas laborales más comunes para los empleados del análisis se encuentran entre las 7:00 am y 6:00 pm.

6.3.2.13. Duración del desplazamiento al trabajo

Dentro de esta categoría se analiza implícitamente las variables de Hora inicio viaje y Hora fin viaje, dado que la duración es calculada a partir del tiempo que toma ambas horas.

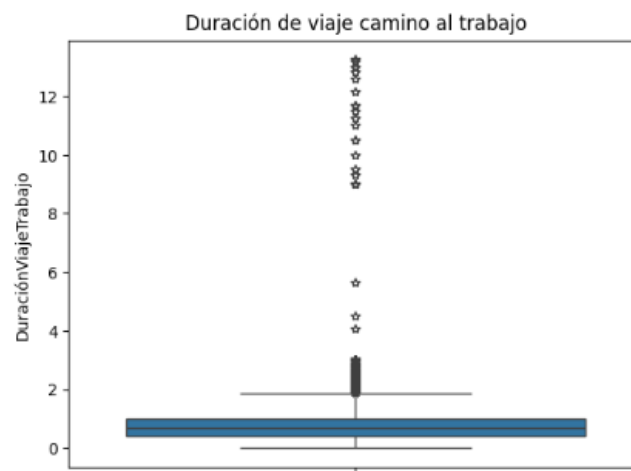


Figura 38. Boxplot duración del desplazamiento al trabajo

Fuente: Elaboración propia

DuraciónViajeTrabajo	
count	37577.00
mean	0.70
std	0.50
min	0.00
25%	0.40
50%	0.67
75%	1.00
max	13.25

Figura 39. Resumen estadístico de la duración del desplazamiento al trabajo

Fuente: Elaboración propia

Aplicando el criterio de Tukey, se considera que la duración de un viaje camino al trabajo es atípica si está por encima de 2 horas. El 0,498 % (187 empleados) del total de empleados superan este límite, por lo que son reemplazados por el límite hallado.

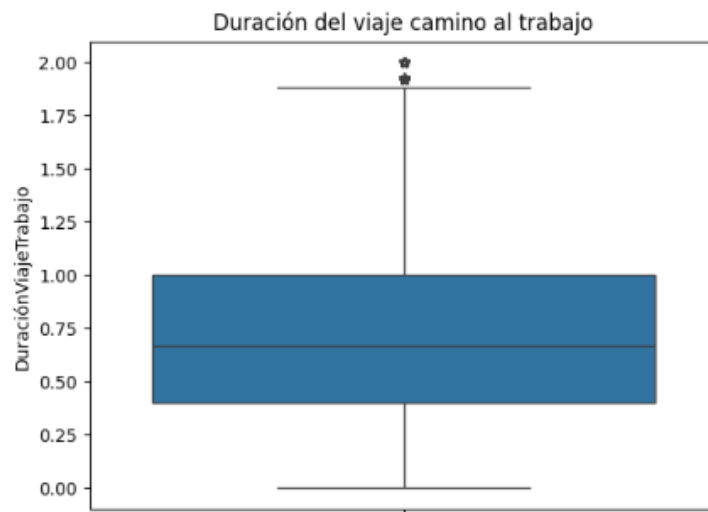


Figura 40. Boxplot sin datos atípicos de la duración del desplazamiento al trabajo

Fuente: Elaboración propia

DuraciónViajeTrabajo	
count	37577.00
mean	0.69
std	0.43
min	0.00
25%	0.40
50%	0.67
75%	1.00
max	2.00

Figura 41. Resumen estadístico sin datos atípicos de la duración del desplazamiento al trabajo

Fuente: Elaboración propia

Se puede observar que la gran mayoría de empleados, aproximadamente el 75% de estos tardan una hora o menos en llegar a sus trabajos, y por su parte, aproximadamente el 6% de los empleados gastan una hora y media o más para llegar a sus puestos de trabajo.

6.3.3. Análisis exploratorio de indicadores

Teniendo en cuenta que el objetivo de esta investigación es construir modelos de Machine Learning aplicados a un sistema de indicadores que evalúen la sostenibilidad en la movilidad urbana del Valle de Aburrá y que para el cálculo de dichos indicadores se debe tener en cuenta la información de factores de emisión suministrada por la subdirección ambiental del AMVA, se deben excluir aquellas categorías de modo de transporte que no pueden ser relacionadas con estos factores, en este caso, y como se evidenció en el análisis exploratorio de variables, son 22 categorías de modos de transporte, de las cuales quedan 21 de ellas y se excluye la

categoría “Otros” para el siguiente análisis de indicadores, obteniendo una nueva base de datos con 22.607 registros y 31 campos.

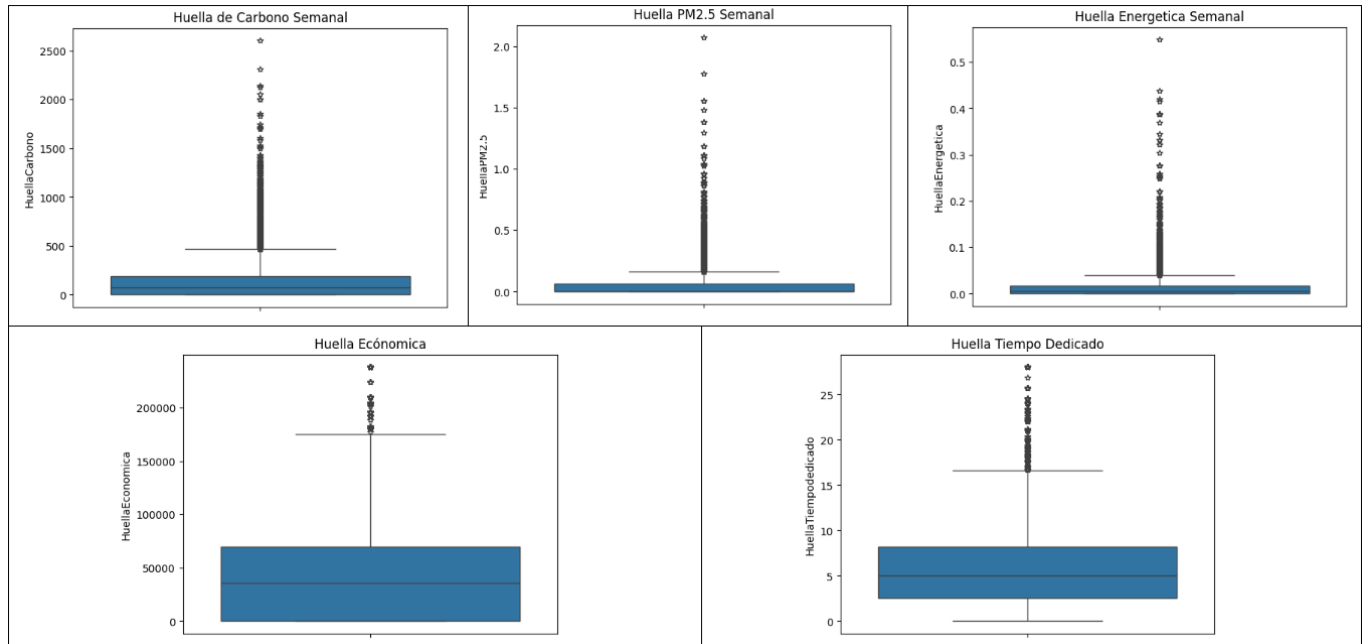


Figura 42. Boxplot de cada uno de los indicadores de movilidad sostenible

Fuente: Elaboración propia

	HuellaEnergetica	HuellaPM2.5	HuellaCarbono	HuellaTiempodedicado	HuellaEconomica
count	22607.00	22607.00	22607.00	22607.00	22607.00
mean	0.01	0.05	135.33	5.86	45448.30
std	0.02	0.11	192.83	4.68	49334.07
min	0.00	0.00	0.00	0.00	0.00
25%	0.00	0.00	0.00	2.50	0.00
50%	0.00	0.00	68.80	5.00	36000.00
75%	0.02	0.06	186.16	8.17	70000.00
max	0.55	2.07	2602.49	28.00	238000.00

Figura 43. Resumen estadístico de cada uno de los indicadores de movilidad sostenible

Fuente: Elaboración propia

La Huella de Carbono tiene un promedio de 135,33 gr CO₂/semana y una desviación estándar de 192,83, reflejando una amplia variabilidad en las emisiones de gases de efecto invernadero. El valor mínimo es 0,00 gr CO₂/semana, indicando que algunas personas no generan emisiones de CO₂, mientras que el máximo de 2602,49 gr CO₂/semana sugiere emisiones muy altas en ciertos casos. Respecto a la mediana, se observa que el 50% de los registros tienen valores de hasta 68,80 gr CO₂/semana. Estas cifras resaltan la necesidad de políticas que incentiven la reducción de emisiones a través de tecnologías más limpias y la promoción de transporte público o no motorizado.

La Huella de PM_{2,5} presenta un promedio de 0,05 gr/semana con una desviación estándar de 0,11, lo que refleja variaciones en la cantidad de partículas finas emitidas por los diferentes modos de transporte. El valor mínimo de 0,00 gr/semana sugiere la existencia de modos de transporte sin emisiones de PM_{2,5}, mientras que el máximo de 2,07 gr/semana indica que algunos desplazamientos generan una cantidad considerable de estas partículas. La mediana de 0,00 gr/semana destaca que el 50% de los desplazamientos no contribuyen significativamente a las emisiones de PM_{2,5}, subrayando la necesidad de fomentar modos de transporte más limpios para mejorar la calidad del aire y reducir los problemas de salud relacionados.

La Huella Energética, con un promedio de 0,01 gal/semana y una desviación estándar de 0,02, indica un consumo moderado de energía en los desplazamientos semanales. El valor mínimo es 0,00, lo que sugiere que hay individuos que no consumen energía en su movilidad, mientras que el máximo de 0,55 gal/semana señala un consumo significativamente alto en algunos casos. La mediana en 0,00 gal/semana resalta que exactamente el 50% de los empleados tienen un consumo

superior a cero, reflejando así el uso de medios de transporte no motorizados o eléctricos. Esta distribución muestra la importancia de promover alternativas energéticamente eficientes para reducir el uso de recursos no renovables y minimizar las emisiones asociadas.

En cuanto a la Huella Económica, con un promedio de \$ 45.448,30 COP/semana y una desviación estándar de \$49.334,07, muestra un amplio rango de costos asociados a la movilidad. El valor mínimo de \$0,00 COP/semana indica que algunos individuos no incurren en costos de movilidad, mientras que el máximo de \$238.000,00 COP/semana refleja gastos extremadamente altos en ciertos casos. La mediana sugiere que exactamente el 50% de la población enfrenta costos de \$36.000,00 COP/semana, lo cual puede representar una carga económica considerable para los empleados del Valle de Aburrá. Estos datos subrayan la importancia de políticas que reduzcan los costos de transporte y mejoren la accesibilidad económica a modos de transporte sostenibles.

Finalmente, la Huella de Tiempo Dedicado a la Movilidad, el promedio es de 5,86 horas/semana con una desviación estándar de 4,68 horas. El tiempo mínimo dedicado es 0,00 horas/semana, lo que podría indicar el uso de teletrabajo o la proximidad a lugares de interés, mientras que el máximo es de 28,00 horas/semana, lo que implica un tiempo considerable invertido en desplazamientos para algunos individuos. La mediana de 5,00 horas/semana muestra que exactamente el 50% de las personas dedica una cantidad significativa de tiempo a la movilidad, afectando su calidad de vida y limitando el tiempo disponible para otras actividades. Esta variabilidad sugiere la necesidad de mejorar la infraestructura y los servicios de transporte para reducir el tiempo de desplazamiento.

6.4. MODELADO

Para el proceso de modelado, fue esencial preparar y transformar los datos, asegurando que todas las variables fueran numéricas y compatibles con el algoritmo de clustering que sería elegido más adelante. Para cumplir con esto, las variables categóricas fueron transformadas en numéricas utilizando métodos como Label Encoding y One-Hot Encoding. Una vez logrado esto, se procede a seleccionar el modelo adecuado.

6.4.1. Selección del número de clúster

Para esto, se utiliza la medida de WCSS (Within-Cluster Sum of Squares) utilizada en algoritmos de clustering para evaluar la calidad de la agrupación realizada. El WCSS mide la suma de las distancias al cuadrado entre cada punto de datos en un clúster y su centroide. La idea es que los puntos dentro de un clúster deben estar cercanos entre sí y lejos de los puntos en otros clusters.

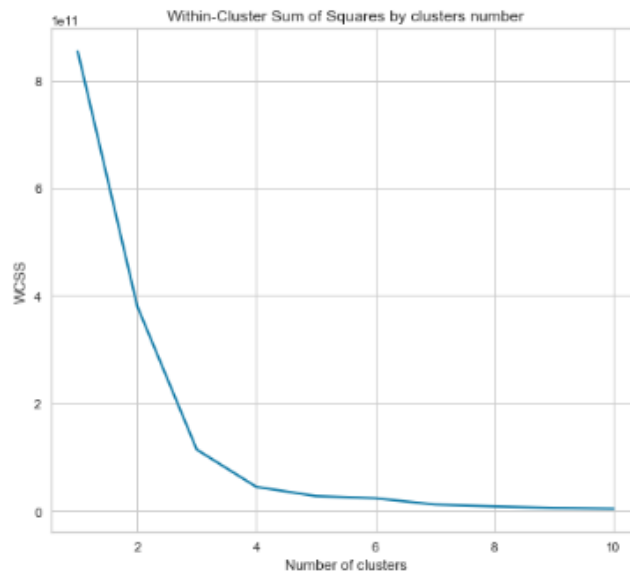


Figura 44. Selección del número de clúster

Fuente: Elaboración propia

De acuerdo con la **Figura 44** y con base al método de codo (Número de clúster óptimo donde la varianza intra-cluster disminuya considerablemente) se decide dividir a los empleados en 4 grupos.

6.4.2. Algoritmo K-Means

En esta sección, implementamos el algoritmo K-Means para agrupar los datos en 4 clusters distintos. Se detalla el proceso de ajuste del modelo, la evaluación de la calidad de los clusters mediante métricas como la inercia y el coeficiente de silueta, y la visualización de los resultados obtenidos. Además, se explora la relación entre diferentes variables dentro de los clusters formados para obtener insights valiosos sobre el comportamiento de los datos.

Tabla 4. Métricas de evaluación Algoritmo K-Means

Evaluación intracluster		Evaluación intercluster	
Inercia K-Means	25173832995,0195	Indice Davies Bouldin	0,3396
Silueta K-Means	0,7126	Indice Dunn	1,4347
Distancia promedio	834,5785	Distancia promedio	8751,845

Fuente: Elaboración propia

La **Tabla 4**, contiene las métricas de evaluación que indican que el modelo K-Means ha logrado un buen equilibrio entre la cohesión intracluster y la separación intercluster. La alta silueta y el bajo índice Davies-Bouldin sugieren que los clústers están bien definidos, con una buena separación entre ellos. La inercia y la distancia promedio intracluster apuntan a clústers internamente cohesionados, mientras que las distancias promedio intercluster confirman que estos clústers están adecuadamente separados en el espacio de características.

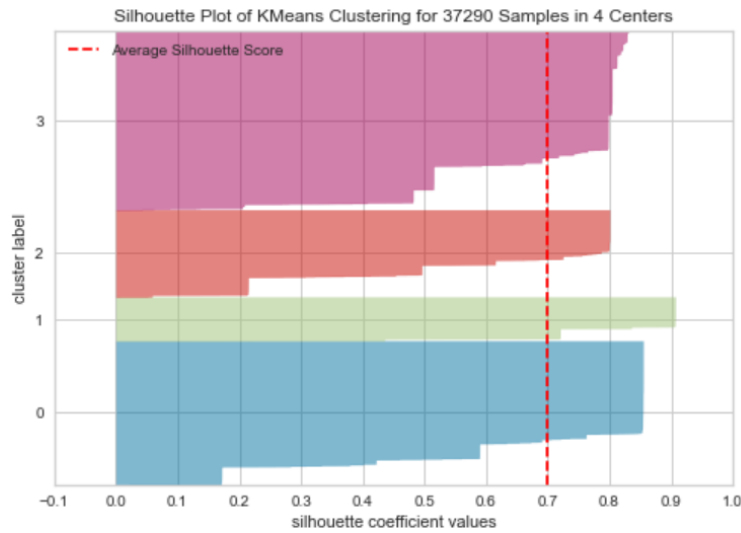


Figura 45. Score Silhouette para cada clúster K-Means

Fuente: Elaboración propia

Referente a la **Figura 45**, un puntaje promedio de silueta de aproximadamente 0,7 sugiere que, en general, los clústeres están bien definidos y que la mayoría de los puntos están correctamente asignados a sus respectivos clústeres. La presencia de puntajes negativos en los clústeres indica que algunos puntos están probablemente asignados al clúster incorrecto.

Clúster 0 (azul): Este clúster muestra una buena consistencia interna, con la mayoría de los puntos teniendo valores de silueta altos. Sin embargo, también hay variabilidad, lo que sugiere que algunos puntos están más cerca de los bordes del clúster.

Clúster 1 (verde): Este clúster tiene pocos elementos que lo componen, y estos pocos están bien asignados pues todos los puntos se encuentran por encima de 0,7.

Clúster 2 (salmón): Aunque la mayoría de los puntos tienen coeficientes de silueta positivos y mayores a 0,7, hay algunos puntos con valores por debajo de este valor, lo que sugiere que estos

puntos están en el límite entre clústeres y podrían no estar tan bien definidos.

Clúster 3 (rosado): Este clúster tiene una gran cantidad de puntos con valores de silueta positivos y por encima de 0,71, lo que indica que los puntos dentro de este clúster están bien agrupados y separados de los puntos en otros clústeres.

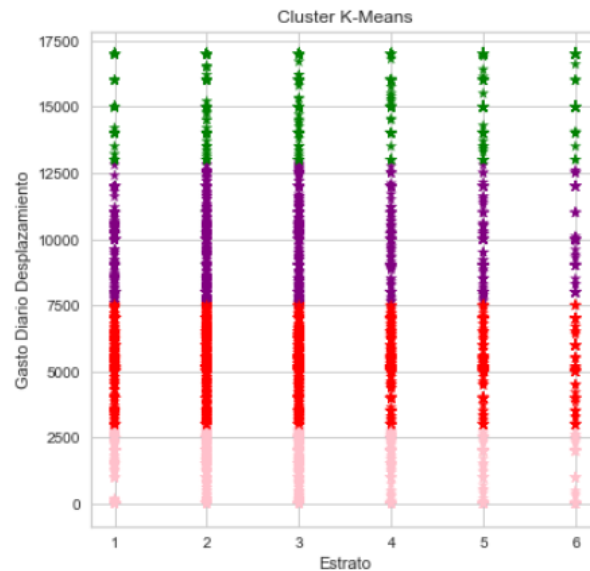


Figura 46. Gráfico de dispersión clúster K-Means

Fuente: Elaboración propia

La **Figura 46** muestra la distribución de los cuatro clústeres en relación con cada estrato y el gasto diario de desplazamiento, para evaluar la capacidad discriminadora del algoritmo. El gráfico revela una diferenciación clara de los clústeres en función de estas variables, lo que confirma la efectividad del modelo K-Means en identificar grupos con características distintas. Los clústeres están bien definidos, y los valores de silueta cercanos a 0,7 refuerzan la cohesión interna y la correcta asignación de la mayoría de los puntos, validando la solidez del análisis realizado.

6.4.3. Algoritmo Hierarchical Agglomerative Clustering (HAC)

En esta sección, se analiza el algoritmo Hierarchical Agglomerative Clustering para agrupar los datos en 4 clusters. Se utiliza el método de Ward para el criterio de enlace, ya que es eficaz en la formación de clusters compactos y bien definidos.

Tabla 5. Métricas de evaluación Algoritmo HAC

Evaluación intracluster		Evaluación intercluster	
Silueta HAC	0,6716761017595367	Indice Davies Bouldin	0,3677

Fuente: Elaboración propia

Las métricas de evaluación que se observan en la **Tabla 5**, demuestran un buen rendimiento en la formación de clusters para el algoritmo Hierarchical Agglomerative Clustering (HAC), El coeficiente de silueta de 0,6717 indica una adecuada cohesión intracluster y separación entre clusters, aunque este valor es ligeramente inferior al obtenido con K-Means, sigue siendo positivo. Por otro lado, el índice Davies-Bouldin de 0,3677 refleja que los clusters son compactos y bien definidos, aunque con una separación ligeramente menor en comparación con los resultados de K-Means, lo que sugiere que, aunque HAC es efectivo, K-Means ofrece una mejor distinción entre clusters.

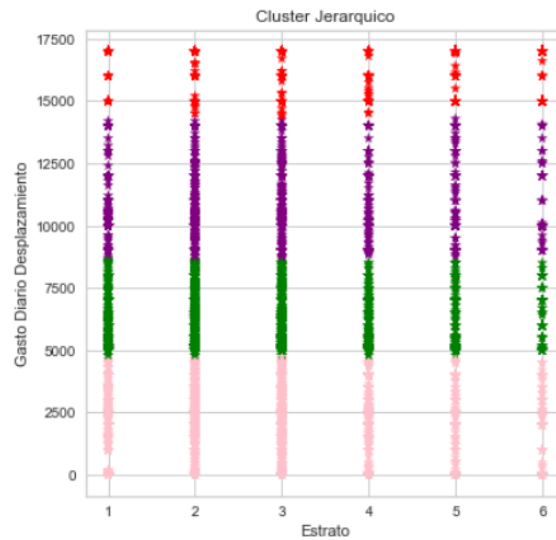


Figura 47. Gráfico de dispersión Hierarchical Agglomerative Clustering

Fuente: Elaboración propia

En la **Figura 47** se presenta la distribución de los clústeres obtenidos mediante el algoritmo *Hierarchical Agglomerative Clustering* con el método de enlace de Ward, cuyo objetivo es maximizar la compacidad y definición de los clústeres. Se observa cómo los grupos se distribuyen en función del Estrato y el Gasto Diario de Desplazamiento.

Cohesión de los clústeres: El puntaje de silueta de 0,6767 obtenido para este método sugiere una cohesión interna adecuada dentro de los clústeres, lo que implica que la mayoría de los puntos están correctamente asignados a sus respectivos grupos, aunque con una ligera variabilidad comparado con el análisis de K-Means. Esto es esperable, ya que el algoritmo jerárquico tiende a generar agrupaciones más compactas, pero a veces menos flexibles en comparación con K-Means.

Distribución y separación de clústeres: Al igual que en el análisis previo con K-Means, se evidencia que los clústeres, representados por diferentes colores, se concentran en ciertas áreas del

gráfico, reflejando una clasificación aceptable. Sin embargo, la superposición observada en algunos estratos indica que la variabilidad en el Gasto Diario de Desplazamiento no proporciona suficiente información para una segmentación altamente diferenciada de la población solo con estas dos variables.

Importancia del Método de Ward: El uso del método de Ward en el criterio de enlace ha contribuido a la formación de clústeres más compactos, lo cual es favorable para asegurar que los datos dentro de cada grupo sean lo más homogéneos posible. Sin embargo, la superposición mencionada sugiere que, aunque el algoritmo ha logrado una buena compactación, algunas diferencias sutiles en la población podrían no estar bien capturadas únicamente con estas variables

6.4.4. Algoritmo K-Medoids

Tabla 6. Métricas de evaluación Algoritmo K-Medoids

Evaluación intracluster		Evaluación intercluster	
Inercia K-Medoids	20.888.628,4253	Indice Davies Bouldin	0,4597
Silueta K- Medoids	0,663	Indice Dunn	0,3675
Distancia promedio	1.052,1829	Distancia promedio	5.515,1678

Fuente: Elaboración propia

Las métricas de evaluación del algoritmo K-Medoids muestra un rendimiento sólido en la formación de clusters. La inercia intracluster de 20.888.628,4253 sugiere que los puntos de datos dentro de cada cluster están relativamente cerca de sus respectivos medoids, lo que indica una buena cohesión interna. El coeficiente de silueta de 0,663 refuerza esta cohesión, mostrando que

la mayoría de los puntos están bien agrupados dentro de sus clusters. El índice Davies-Bouldin de 0,4597 indica que los clusters son compactos, aunque con una separación moderada entre ellos. Además, la distancia promedio intracluster de 1.052,1829 respalda la proximidad de los puntos dentro de los clusters, mientras que la distancia promedio intercluster de 5.515,1678 refleja una separación razonable entre los clusters formados.

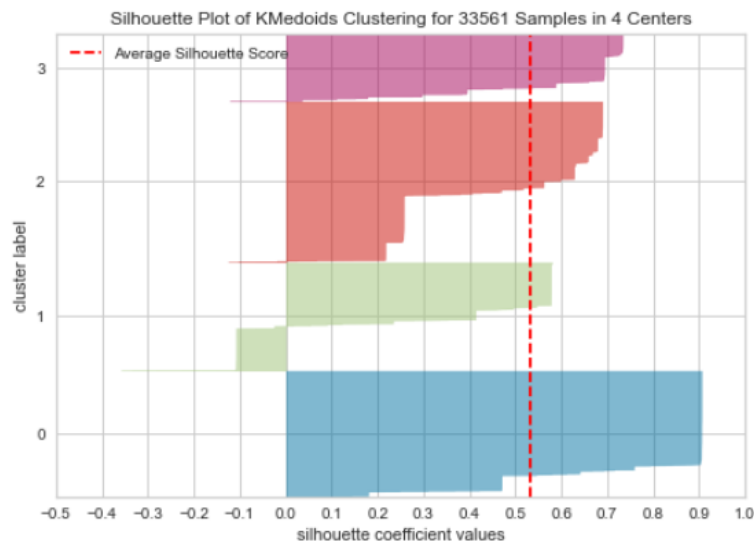


Figura 48. Score Silhouette para cada clúster K-Medoids

Fuente: Elaboración propia

Un puntaje promedio de silueta de aproximadamente 0,7 sugiere que, en general, los clústeres están bien definidos y que la mayoría de los puntos están correctamente asignados a sus respectivos clústeres. La presencia de puntajes negativos en los clústeres indica que algunos puntos están probablemente asignados al clúster incorrecto. Respecto a este contexto, la **Figura 48**, permite observar lo siguiente:

Clúster 0 (azul): Este clúster tiene una distribución amplia de coeficientes de silueta, con muchos

puntos por encima de 0,7. Esto indica una consistencia buena dentro del clúster, lo que sugiere que los puntos están razonablemente bien agrupados.

Clúster 1 (verde): Este clúster muestra una amplia variabilidad en los coeficientes de silueta, con muchos puntos cerca de 0 e incluso varios negativos. Esto sugiere que este clúster puede tener problemas con la consistencia interna y la separación de otros clústeres, indicando que algunos puntos podrían estar mal asignados y pertenecer a otros clústeres.

Clúster 2 (salmón): Los puntos en este clúster tienen una mejor consistencia interna, con la mayoría de los coeficientes de silueta en el rango positivo, aunque con algunos puntos cercanos a 0. Esto sugiere que, en general, los puntos están bien agrupados, aunque hay cierta incertidumbre en la asignación de algunos puntos.

Clúster 3 (rosado): Este clúster muestra una buena consistencia interna, con la mayoría de los puntos teniendo altos valores de silueta positivos. Esto indica una excelente separación y consistencia interna, sugiriendo que los puntos en este clúster están muy bien agrupados y claramente diferenciados de los puntos en otros clústeres.

En conclusión, el algoritmo de K-Medoides ha realizado una agrupación moderadamente efectiva. Con un puntaje promedio de silueta de aproximadamente 0,5, hay una buena separación y consistencia en general, aunque hay variabilidad significativa dentro de los clústeres, especialmente en el clúster 1. La presencia de algunos valores negativos y cercanos a 0 sugiere que podría haber espacio para mejorar la asignación de ciertos puntos.

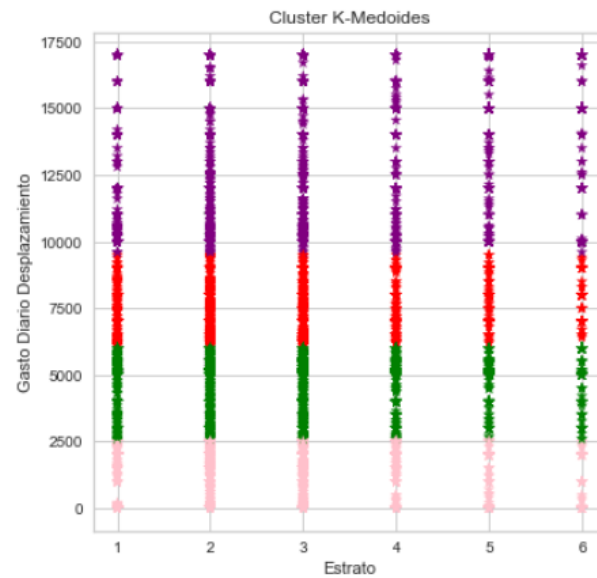


Figura 49. Gráfico de dispersión K-Medoids

Fuente: Elaboración propia

La **Figura 49** confirma que el modelo K-Medoids ha logrado una agrupación efectiva en términos generales. Aunque se observa alguna variabilidad en la cohesión interna de ciertos clústeres, la diferenciación clara entre los clústeres sugiere que el algoritmo ha sido capaz de identificar grupos con características distintas. Esto valida la capacidad del K-Medoids para agrupar datos de manera efectiva, a pesar de la presencia de algunas áreas que podrían beneficiarse de ajustes adicionales.

6.4.5. Selección del mejor algoritmo de clustering

La selección del mejor algoritmo de clustering en este estudio se fundamenta en una evaluación exhaustiva de las métricas de desempeño clave, con un enfoque particular en el coeficiente de silueta, la inercia intracluster, y los índices Davies-Bouldin y Dunn. Estos indicadores proporcionan una visión integral sobre la cohesión interna de los clústers y la separación entre

ellos, aspectos críticos para la interpretación de los resultados en el contexto de la movilidad sostenible.

El algoritmo K-Means se destaca como el método más adecuado para la segmentación de los datos en este estudio, respaldado por un coeficiente de silueta de 0,7126. Este valor indica una cohesión y separación excepcional, lo que refleja que los puntos de datos dentro de cada cluster están bien agrupados, con una clara distinción entre los clusters formados. Además, la inercia de 25.173.832.995,0195 refuerza la estabilidad y consistencia de los clusters, indicando que los puntos están suficientemente cerca del centroide de cada cluster, lo cual es crucial en la interpretación de patrones en los datos de movilidad urbana.

Por otra parte, aunque el algoritmo Hierarchical Agglomerative Clustering (HAC) utilizando el método de Ward también mostró un rendimiento sólido con un coeficiente de silueta de 0,6717 y un índice Davies-Bouldin de 0,3677, que refleja clusters compactos y bien separados, estos resultados fueron marginalmente inferiores en comparación con K-Means. Asimismo, el algoritmo K-Medoids, aunque robusto en su estructura de clustering, presentó un coeficiente de silueta de 0,663 y un índice Davies-Bouldin de 0,4597, lo que indica un desempeño adecuado, pero no superior al de K-Means.

Dado el propósito de este estudio, que es identificar patrones de movilidad urbana que puedan ser utilizados para formular políticas sostenibles, la precisión y claridad en la segmentación de los datos son esenciales. K-Means ofrece la mejor combinación de cohesión intracluster y separación intercluster, factores determinantes para la interpretación efectiva de los resultados. Este nivel de rendimiento en el clustering es fundamental para aplicaciones en movilidad sostenible, donde la

correcta identificación de patrones puede influir directamente en la toma de decisiones para mejorar la infraestructura y reducir el impacto ambiental.

Por lo tanto, la decisión de continuar el análisis con el algoritmo K-Means no solo está justificada por su superioridad en las métricas evaluadas, sino también por su capacidad para proporcionar insights precisos y accionables en el contexto del urbanismo y la movilidad sostenible, haciendo de este algoritmo la opción óptima para el objetivo de este estudio.

6.5. EVALUACIÓN

6.5.1. Caracterización de los grupos de clustering

Cluster	Cantidad	Porcentaje
0	11051	48.88
1	7887	34.89
2	901	3.99
3	2768	12.24

Figura 50. Frecuencia de empleados por grupo de clustering

Fuente: Elaboración propia

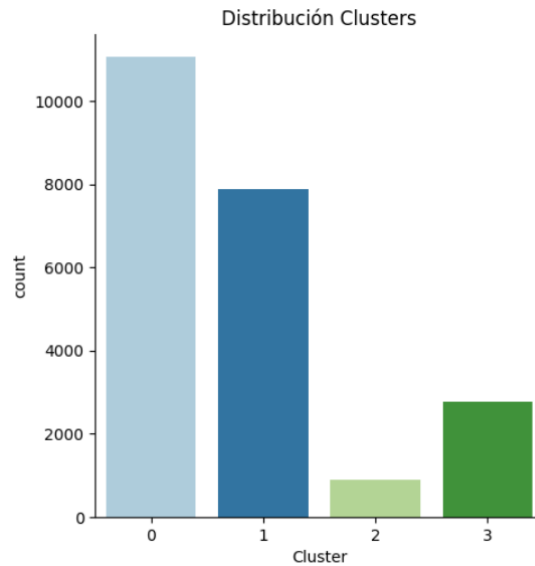


Figura 51. Distribución de empleados por grupo de clustering

Fuente: Elaboración propia

La mayoría de los empleados, precisamente 11.051 (48,9%) se consolidaron dentro del grupo 0, seguido del grupo 1 que cuenta con 7.887 empleados (34,89%), luego el grupo 3 agrupó 2.768 (12,24%) y el grupo 2 tan solo 901 empleados (9,66%).

6.5.2. Huella de carbono (CO₂)

	Promedio	Mediana	Std Dev	Min	Max	Q1	Q3
Cluster							
0	65.53	0.00	124.27	0.00	2121.46	0.00	93.08
1	160.35	108.59	170.92	0.00	2314.32	47.61	225.98
2	372.38	321.68	320.55	0.00	1999.53	117.06	551.03
3	265.54	186.16	262.58	0.00	2602.49	71.67	392.61

Figura 52. Resumen estadístico Huella de carbono (CO₂) por clúster

Fuente: Elaboración propia

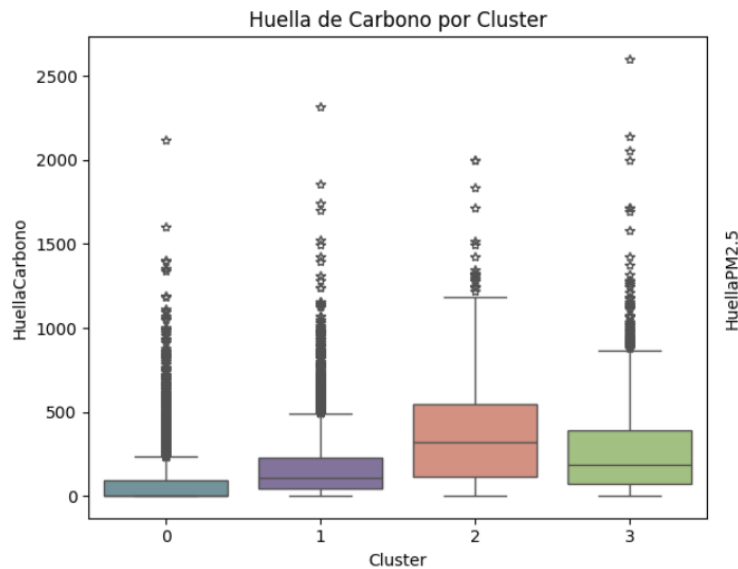


Figura 53. Boxplot Huella de carbono (CO₂) por clúster

Fuente: Elaboración propia

Cluster 0:

El grupo de empleados del cluster 0 presenta la huella de carbono más baja entre todos los grupos, con un promedio de 65,53 gramos de CO₂ por semana y una mediana de 0,00 gramos de CO₂. Esto sugiere que muchos empleados en este grupo no generan emisiones significativas de carbono, probablemente debido a que utilizan modos de transporte sin emisiones, como caminar o andar en bicicleta. La desviación estándar de 124,27 gramos indica cierta variabilidad, pero en general, los empleados de este cluster tienen un impacto ambiental muy bajo en términos de emisiones de carbono.

Cluster 1:

El cluster 1 tiene una huella de carbono promedio de 160,35 gramos de CO₂ por semana, con una mediana de 108,59 gramos de CO₂. Esto indica que los empleados en este grupo generan más

emisiones de carbono que el cluster 0, probablemente debido al uso de vehículos motorizados, aunque en niveles aún moderados. La desviación estándar de 170,92 gramos sugiere una variabilidad considerable en las emisiones, con algunos empleados generando cantidades significativas de CO₂.

Cluster 2:

El grupo de empleados del cluster 2 presenta la huella de carbono más alta entre todos los grupos, con un promedio de 372,38 gramos de CO₂ por semana y una mediana de 321,68 gramos. Esto sugiere que estos empleados dependen en gran medida de vehículos personales que generan altas emisiones de carbono. La desviación estándar de 320,55 gramos muestra una considerable variabilidad, pero en general, las emisiones son bastante altas.

Cluster 3:

El cluster 3 tiene una huella de carbono promedio de 265,54 gramos de CO₂ por semana, con una mediana de 186,16 gramos. Esto indica que los empleados en este grupo generan emisiones de carbono intermedias, posiblemente debido a una combinación de transporte individual y público. La desviación estándar de 262,58 gramos sugiere una variabilidad considerable en las emisiones, con algunos empleados generando cantidades significativas de CO₂.

6.5.3. Huella de PM_{2,5}.

	Promedio	Mediana	Std Dev	Min	Max	Q1	Q3
Cluster							
0	0.04	0.00	0.08	0.00	1.55	0.00	0.03
1	0.07	0.01	0.12	0.00	1.77	0.00	0.11
2	0.06	0.01	0.13	0.00	1.29	0.00	0.01
3	0.06	0.01	0.13	0.00	2.07	0.00	0.01

Figura 54. Resumen estadístico Huella de PM 2.5 por clúster

Fuente: Elaboración propia

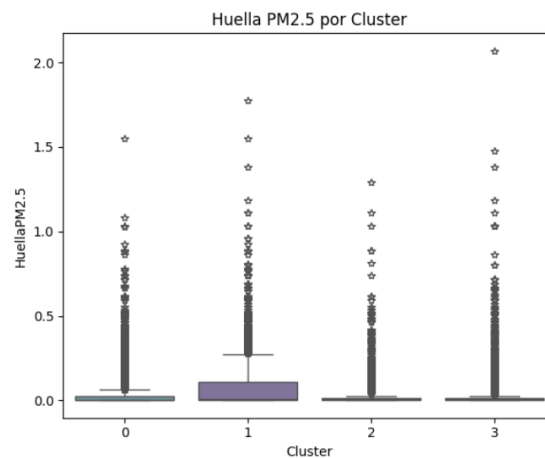


Figura 55. Boxplot Huella de PM 2.5 por clúster

Fuente: Elaboración propia

Cluster 0:

En cuanto a la huella de PM_{2,5}, para el grupo de cluster 0, el promedio es de 0,04 gramos por semana, con una mediana de 0,00 gramos. Esto refuerza la idea de que los empleados del cluster 0 utilizan modos de transporte que no generan emisiones significativas de partículas finas. La mayoría de los empleados tienen valores de PM_{2,5} muy bajos, lo que sugiere un impacto mínimo en la calidad del aire y, por ende, en la salud pública.

Por lo tanto, los empleados del cluster 0 tienen tanto una huella de carbono como una huella de PM_{2,5} muy bajas. Esto indica que utilizan modos de transporte altamente sostenibles y amigables con el medio ambiente. Este perfil de empleados no solo beneficia al medio ambiente al reducir las emisiones de gases de efecto invernadero y partículas finas, sino que también mejora su calidad de vida al evitar la exposición a contaminantes del aire.

Cluster 1:

La huella de PM_{2,5} para el cluster 1 es de 0,07 gramos por semana en promedio, con una mediana de 0,01 gramos. Aunque estos valores son más altos que los del cluster 0, siguen siendo relativamente bajos. Sin embargo, la mayor variabilidad en las emisiones de partículas finas indica que algunos empleados podrían estar utilizando vehículos personales o modos de transporte que generan más contaminantes del aire.

En resumen, los empleados del cluster 1 tienen una huella de carbono y una huella de PM_{2,5} moderadas, reflejando un uso mixto de transporte motorizado y posiblemente público. Estos empleados tienen un impacto ambiental mayor que el cluster 0, pero aún mantienen un nivel relativamente bajo de emisiones, lo cual es positivo para el medio ambiente y la salud pública.

Cluster 2:

La huella de PM_{2,5} para el cluster 2 es de 0,06 gramos por semana en promedio, con una mediana de 0,01 gramos. Aunque el promedio es similar al del cluster 1, la desviación estándar y el rango de emisiones indican que algunos empleados generan una cantidad significativa de partículas finas. Esto podría estar relacionado con el uso de vehículos personales y la exposición a contaminantes del aire.

En conclusión, los empleados del cluster 2 tienen la huella de carbono más alta y una huella de PM2,5 significativa, lo que sugiere una fuerte dependencia de vehículos motorizados. Esto no solo contribuye al calentamiento global y al cambio climático, sino que también puede tener efectos adversos en la salud pública debido a la mayor exposición a contaminantes del aire.

Cluster 3:

La huella de PM2,5 para el cluster 3 es de 0,06 gramos por semana en promedio, con una mediana de 0,01 gramos. Estos valores son similares a los del cluster 2, lo que indica que los empleados de este grupo también tienen un impacto notable en la calidad del aire. La variabilidad en las emisiones de partículas finas sugiere que algunos empleados utilizan modos de transporte que generan más contaminantes del aire.

En resumen, los empleados del cluster 3 tienen una huella de carbono y una huella de PM2,5 intermedias, reflejando un uso mixto de transporte individual y público. Aunque su impacto ambiental no es tan alto como el del cluster 2, sigue siendo significativo, lo que sugiere que podrían beneficiarse de estrategias para reducir sus emisiones y mejorar la calidad del aire.

6.5.4. Huella energética

	Promedio	Mediana	Std Dev	Min	Max	Q1	Q3
Cluster							
0	0.01	0.00	0.01	0.00	0.44	0.00	0.01
1	0.01	0.01	0.02	0.00	0.39	0.00	0.02
2	0.04	0.03	0.05	0.00	0.55	0.01	0.06
3	0.03	0.02	0.03	0.00	0.39	0.00	0.04

***Figura 56.** Resumen estadístico Huella energética por clúster*

Fuente: Elaboración propia

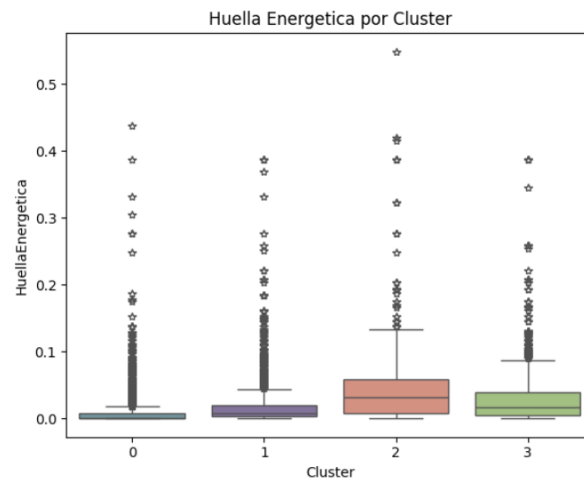


Figura 57. *Boxplot Huella energética por clúster*

Fuente: Elaboración propia

Cluster 0:

El grupo del cluster 0 se caracteriza por tener la huella energética más baja entre todos los grupos, con un promedio de 0,01 galones por semana, incluso el 75% de sus empleados registran una huella nula. Este bajo consumo energético sugiere que estos empleados probablemente tienen trabajo desde casa o utilizan modos de transporte más eficientes o menos dependientes de combustibles fósiles, como el transporte público, bicicletas, o caminatas. La mediana de la huella energética es de 0,00 galones por semana, lo cual refuerza la idea de un uso muy limitado de energía en sus desplazamientos.

Cluster 1:

El grupo de empleados del cluster 1 muestra una huella energética ligeramente superior a la del cluster 0, con un promedio de 0,01 galones por semana. Aunque esta cifra es aún baja, sugiere que hay una ligera variación en el uso de energía entre los empleados de este grupo. La mediana de la

huella energética también es de 0,01 galones por semana, indicando que la mayoría de los empleados tienen un consumo energético similar.

Cluster 2:

El cluster 2 se distingue por tener la huella energética más alta entre todos los grupos, con un promedio de 0,04 galones por semana. Esta cifra, aunque no excesivamente alta, indica un uso de energía más significativo relacionado con la movilidad. La mediana de 0,03 galones por semana sugiere que la mitad de los empleados en este grupo utiliza modos de transporte más intensivos en energía, posiblemente vehículos personales.

Cluster 3:

El grupo de empleados del cluster 3 tiene una huella energética promedio de 0,03 galones por semana, ligeramente inferior a la del cluster 2 pero aún más alta que la de los clusters 0 y 1. La mediana de la huella energética es de 0,02 galones por semana, lo que indica un consumo energético moderado entre los empleados de este grupo. Esta cifra sugiere que los modos de transporte utilizados son variados, pero con una tendencia hacia un mayor uso de energía.

6.5.5. Huella económica

	Promedio	Mediana	Std Dev	Min	Max	Q1	Q3
Cluster							
0	6749.71	0.00	10752.97	0.00	32400.00	0.00	15000.00
1	58381.15	60000.00	14437.03	33000.00	86800.00	48000.00	70000.00
2	190783.01	180000.00	26023.40	153600.00	238000.00	170000.00	204000.00
3	115791.59	119980.00	17956.87	87600.00	153280.00	100000.00	132000.00

Figura 58. Resumen estadístico Huella económica por clúster

Fuente: Elaboración propia

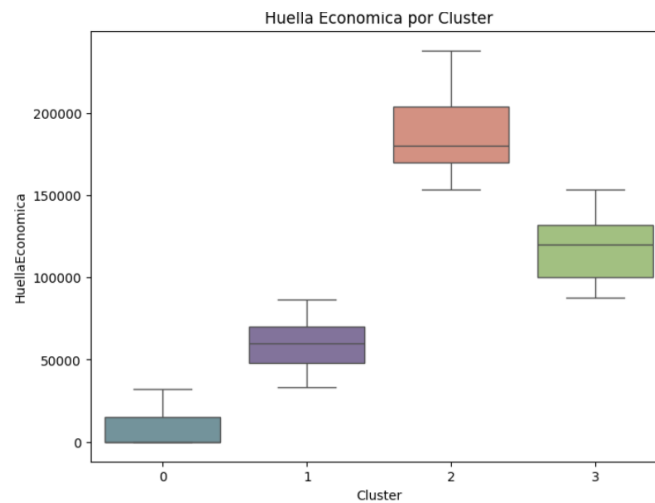


Figura 59. *Boxplot Huella económica por clúster*

Fuente: Elaboración propia

Cluster 0:

El grupo de empleados del cluster 0 gasta muy poco dinero en su movilidad semanal, con un promedio de apenas \$6.749,71 COP. Muchos de ellos no gastan nada en absoluto, lo que sugiere que pueden estar trabajando desde casa o usan formas de transporte gratuitas como caminar o andar en bicicleta. Incluso los que gastan algo, tienden a mantener sus gastos bastante bajos, generalmente por debajo de los \$15.000 COP por semana.

Este bajo gasto en movilidad sugiere que los empleados del cluster 0 tienen una ventaja económica significativa, ya que no necesitan destinar parte de su presupuesto semanal a desplazamientos. Además, es probable que disfruten de una mejor calidad de vida debido a la menor necesidad de viajar largas distancias cada día. En general, el cluster 0 está compuesto por empleados que han optimizado sus rutas de viaje para minimizar los costos.

Cluster 1:

El cluster 1 está compuesto por empleados que gastan más en su movilidad, con un promedio semanal de \$58.381,15 COP. Estos empleados probablemente utilizan medios de transporte que requieren un gasto constante, como el transporte público o vehículos personales. Aunque sus gastos varían, la mayoría gasta entre \$48.000 y \$70.000 COP por semana.

Estos empleados probablemente viven a una distancia moderada de su lugar de trabajo y tienen que destinar una parte considerable de su presupuesto a la movilidad. Aunque sus gastos son más altos que los del cluster 0, siguen siendo manejables. Sin embargo, esta inversión regular en transporte podría impactar su economía personal y su calidad de vida, ya que implica un gasto continuo y significativo cada semana.

Cluster 2:

El cluster 2 tiene los gastos de movilidad más altos de todos, con un promedio de \$190.783,01 COP por semana. Esto indica que los empleados de este grupo probablemente dependen en gran medida de vehículos personales, con todos los costos asociados como combustible, mantenimiento y posiblemente peajes. Sus gastos son consistentemente altos, y la mayoría gasta entre \$170.000 y \$204.000 COP semanalmente.

Este nivel de gasto sugiere que los empleados del cluster 2 enfrentan una carga económica significativa debido a sus desplazamientos. Es probable que vivan lejos de su lugar de trabajo, lo que les obliga a invertir mucho dinero en movilidad. Esta situación puede tener un impacto negativo en su presupuesto personal y en su calidad de vida, ya que gran parte de su ingreso se destina al transporte.

Cluster 3:

Los empleados del cluster 3 también tienen gastos de movilidad relativamente altos, con un promedio de \$115.791,59 COP por semana. Este grupo gasta menos que el cluster 2, pero, aun así, sus costos son significativos. La mayoría de los empleados en este grupo gasta entre \$100.000 y \$132.000 COP por semana, lo que sugiere una combinación de uso de transporte público y vehículos personales.

Estos empleados probablemente viven a una distancia considerable de su lugar de trabajo y utilizan una mezcla de opciones de transporte que resultan en costos moderadamente altos. Aunque no gastan tanto como el cluster 2, la inversión en movilidad sigue siendo una parte importante de su presupuesto semanal. Este gasto considerable puede afectar su economía personal y su bienestar, ya que deben equilibrar estos costos con otras necesidades financieras.

6.5.6. Huella tiempo dedicado a la movilidad

	Promedio	Mediana	Std Dev	Min	Max	Q1	Q3
Cluster							
0	3.78	3.00	3.86	0.00	28.00	0.00	6.00
1	7.36	6.67	4.24	0.00	28.00	4.00	10.00
2	9.20	8.00	5.02	0.83	28.00	5.00	12.00
3	8.78	8.00	4.91	0.17	28.00	5.00	11.67

Figura 60. Resumen estadístico Huella de tiempo dedicado a la movilidad por clúster

Fuente: Elaboración propia

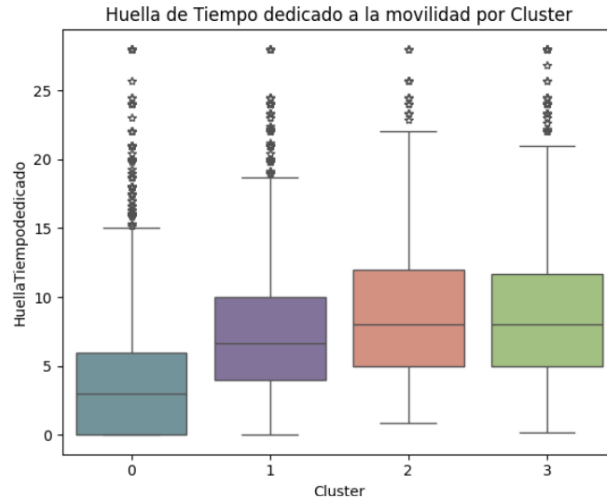


Figura 61. *Boxplot Huella de tiempo dedicado a la movilidad por clúster*

Fuente: Elaboración propia

Cluster 0:

En cuanto a la huella de tiempo dedicado a la movilidad, los empleados del cluster 0 dedican en promedio 3,78 horas por semana a sus desplazamientos y el 75% de los empleados incluso tardan menos de 6 horas. Esta cantidad de tiempo es significativamente menor comparada con los otros clústeres, lo que podría indicar que estos empleados tienen teletrabajo o viven más cerca de sus lugares de trabajo o tienen opciones de transporte más rápidas y eficientes. A Esta combinación de baja huella energética y menor tiempo dedicado a la movilidad sugiere un perfil de empleados que podrían estar disfrutando de una mayor calidad de vida y bienestar personal, gracias a menores tiempos de desplazamiento y un impacto ambiental reducido.

En comparación con otras huellas, se observa que los empleados del cluster 0 parecen ser aquellos que han optimizado sus desplazamientos tanto en términos de tiempo como de consumo energético. Estos empleados posiblemente tengan modalidad de trabajo desde casa o residen cerca

de sus lugares de trabajo o utilizan medios de transporte muy eficientes, lo cual les permite reducir tanto su huella energética como el tiempo dedicado a la movilidad.

Cluster 1:

En términos de tiempo dedicado a la movilidad, los empleados del cluster 1 dedican en promedio 7,36 horas por semana a sus desplazamientos. Esta cantidad de tiempo es significativamente mayor que la del cluster 0, lo que sugiere que estos empleados podrían estar viviendo más lejos de sus lugares de trabajo o enfrentando condiciones de tráfico menos favorables. Adicionalmente el 75% de los empleados registran una duración del viaje de máximo 10 horas semanales, lo que muestra una distribución más amplia de los tiempos de desplazamiento dentro de este grupo. Además, cabe resaltar que el rango de tiempo dedicado a la movilidad en el cluster 1 varía de 0 a 28 horas por semana, con una mayor concentración en el rango de 4 a 10 horas (cuartil inferior y superior). Esto indica que aunque algunos empleados tienen desplazamientos relativamente largos, la mayoría se encuentra dentro de un rango moderado de tiempo dedicado a la movilidad. Este patrón sugiere que los empleados de este grupo podrían estar utilizando una combinación de transporte individual y público.

En resumen, los empleados del cluster 1 tienen una huella energética baja pero dedican más tiempo a sus desplazamientos comparado con el cluster 0. Esto podría estar relacionado con la distancia a sus lugares de trabajo o las opciones de transporte disponibles. A pesar de esto, mantienen un consumo energético eficiente, lo cual es positivo desde una perspectiva ambiental.

Cluster 2:

En términos de tiempo dedicado a la movilidad, los empleados del cluster 2 dedican en promedio

9,20 horas por semana a sus desplazamientos. Este es el tiempo más alto dedicado a la movilidad entre todos los clústeres, lo que sugiere que estos empleados probablemente viven más lejos de sus lugares de trabajo y/o enfrentan mayores tiempos de viaje debido al tráfico. La mediana indica que la mitad de los empleados pueden llegar a dedicar más de 8 horas semanales a su movilidad hacia el trabajo. Adicionalmente, este grupo presenta una alta concentración en el rango de 5 a 12 horas (cuartil inferior y superior). Esta amplia variabilidad indica que hay empleados con desplazamientos muy largos, lo cual podría estar impactando negativamente su calidad de vida y tiempo disponible para otras actividades. La combinación de una huella energética más alta y tiempos de desplazamiento más largos sugiere que este grupo podría beneficiarse de medidas para mejorar la eficiencia de sus desplazamientos.

Por lo tanto, los empleados del cluster 2 tienen la mayor huella energética y dedican más tiempo a la movilidad, lo que indica un potencial impacto negativo en su bienestar personal y ambiental. Este grupo podría beneficiarse significativamente de iniciativas que promuevan el uso de transporte más eficiente y opciones de trabajo remoto o flexibilidad laboral para reducir sus tiempos de desplazamiento.

Cluster 3:

En cuanto al tiempo dedicado a la movilidad, los empleados del cluster 3 dedican en promedio 8,78 horas por semana a sus desplazamientos. Esta cantidad de tiempo es similar a la del cluster 2, lo que sugiere que estos empleados también podrían estar enfrentando largos desplazamientos. Este grupo presenta una mayor concentración en el rango de 5 a 11,67 horas. Esta distribución sugiere que, aunque algunos empleados tienen desplazamientos largos, la mayoría se encuentra

dentro de un rango moderado. La combinación de una huella energética moderada y tiempos de desplazamiento más largos indica que estos empleados podrían estar utilizando una mezcla de transporte individual y público, con una eficiencia energética intermedia.

Por lo tanto, los empleados del clúster 3 presentan una huella energética y tiempos de desplazamiento moderados, indicando una combinación de modos de transporte que no optimizan completamente ni el tiempo ni el consumo energético. Este grupo podría beneficiarse de estrategias que promuevan el uso de transporte más eficiente y que reduzcan los tiempos de desplazamiento, mejorando así su calidad de vida y contribuyendo a la sostenibilidad ambiental.

6.5.7. Mapas de calor de variables por clúster

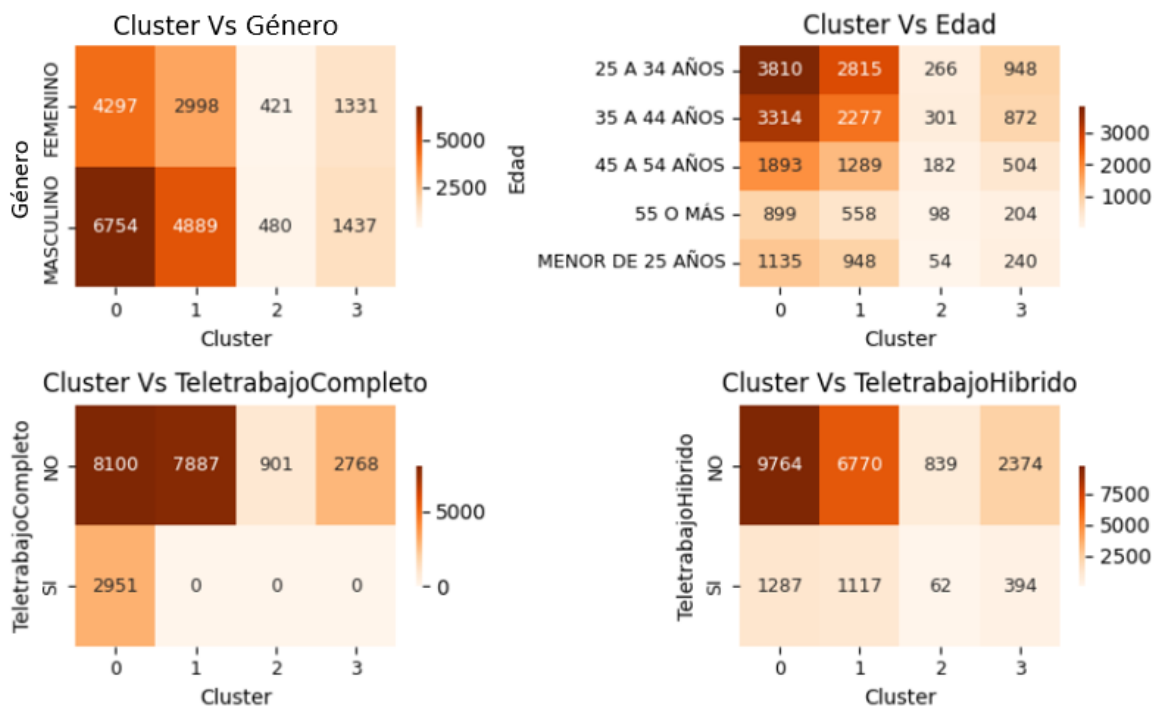


Figura 62. Mapa de calor de frecuencia de género, edad y modalidad de trabajo por clúster

Fuente: Elaboración propia



Figura 63. Mapa de calor porcentual de género, edad y modalidad de trabajo por clúster

Fuente: Elaboración propia

Cluster 0:

El Cluster 0 tiene una mayor proporción de empleados masculinos (61,1%) en comparación con las empleadas femeninas (38,9%). En términos de edad, este grupo tiene una mayor concentración de empleados jóvenes, con el 34,5% de los empleados en el rango de 25 a 34 años y el 30% en el rango de 35 a 44 años. Además, un 17,1% de los empleados tiene entre 45 y 54 años, mientras que solo el 8,1% tiene 55 años o más, y el 10,3% son menores de 25 años. En cuanto al teletrabajo completo, el 73,3% de los empleados de este grupo no realizan teletrabajo completo, mientras que el 26,7% sí lo hace, lo que indica una cierta flexibilidad laboral en este grupo. Finalmente, el teletrabajo híbrido es menos común, con el 88,4% de los empleados no participando en esta

modalidad y solo el 11,6% realizándolo.

Cluster 1:

El Cluster 1 presenta una distribución de género similar a la del Cluster 0, con una mayoría de empleados masculinos (62,0%) y un 38,0% de empleadas femeninas. Este grupo también está compuesto principalmente por empleados jóvenes, con el 35,7% en el rango de 25 a 34 años y el 28,9% en el rango de 35 a 44 años. Un 16,3% de los empleados tiene entre 45 y 54 años, mientras que el 7,1% tiene 55 años o más, y el 12% son menores de 25 años. En términos de teletrabajo completo, todos los empleados de este grupo (100%) no realizan teletrabajo completo. En cuanto al teletrabajo híbrido, el 85,8% de los empleados no participan en esta modalidad, mientras que el 14,2% sí lo hacen, lo que muestra una ligera inclinación hacia esta forma de trabajo.

Cluster 2:

El Cluster 2 tiene una distribución de género más equilibrada, con el 53,3% de los empleados siendo masculinos y el 46,7% femeninos. En términos de edad, este grupo tiene una mayor proporción de empleados de mediana edad, con el 29,5% en el rango de 25 a 34 años, el 33,4% en el rango de 35 a 44 años, y el 20,2% en el rango de 45 a 54 años. Solo el 10,9% de los empleados tiene 55 años o más, y el 6% son menores de 25 años. En lo que respecta al teletrabajo completo, ninguno de los empleados de este grupo (100%) realiza teletrabajo completo. Además, el teletrabajo híbrido es poco común, con el 93,1% de los empleados no participando en esta modalidad y solo el 6,9% realizándolo, lo que sugiere una tendencia hacia formas de trabajo más tradicionales.

Cluster 3:

El Cluster 3 tiene una distribución de género similar a la del Cluster 2, con el 51,9% de los empleados siendo masculinos y el 48,1% femeninos. Este grupo también tiene una mayor proporción de empleados en los rangos de edad de 25 a 34 años (34,2%) y 35 a 44 años (31,5%), mientras que el 18,2% de los empleados tiene entre 45 y 54 años, el 7,4% tiene 55 años o más, y el 8,7% son menores de 25 años. En términos de teletrabajo completo, todos los empleados de este grupo (100%) no realizan teletrabajo completo. En cuanto al teletrabajo híbrido, el 85,8% de los empleados no participan en esta modalidad, mientras que el 14,2% sí lo hacen, mostrando una ligera preferencia por esta forma de trabajo combinada.

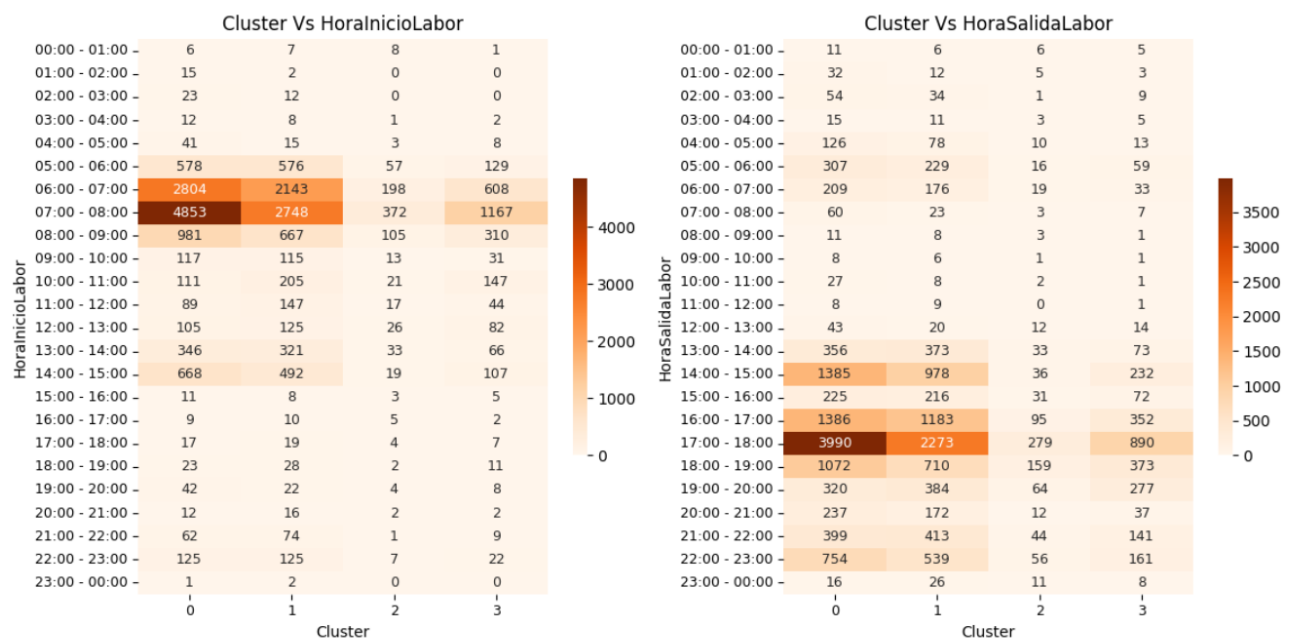


Figura 64. Mapa de calor de frecuencia de hora inicio y salida laboral por clúster

Fuente: Elaboración propia

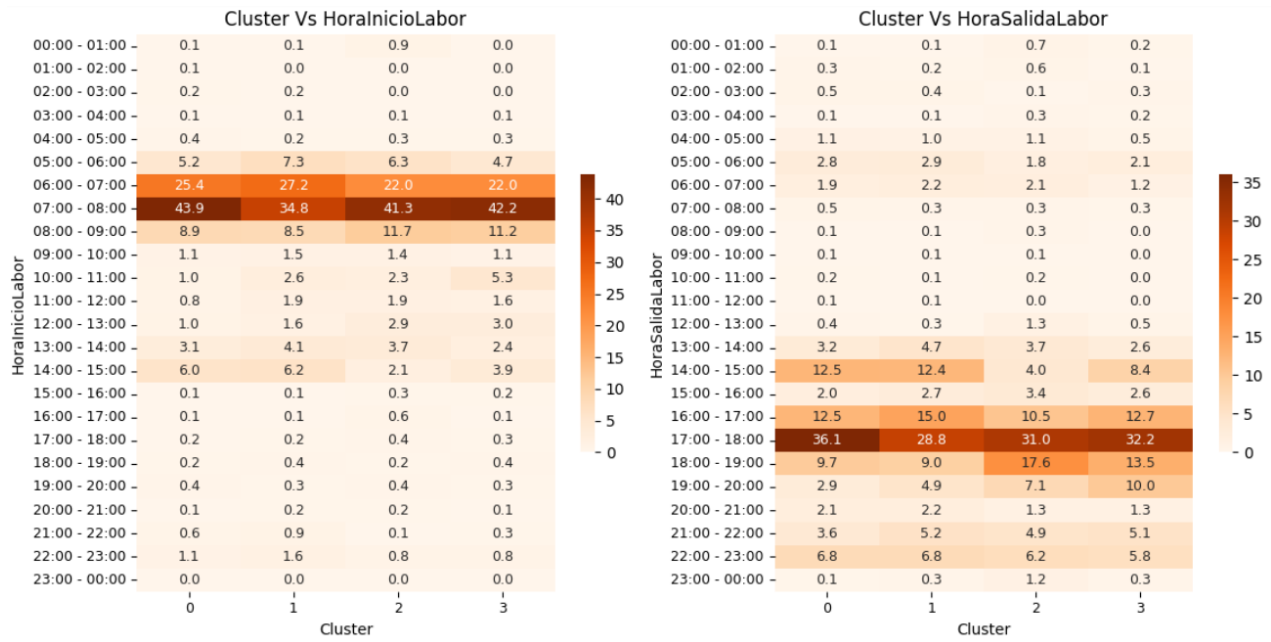


Figura 65. Mapa de calor porcentual hora de inicio y salida laboral

Fuente: Elaboración propia

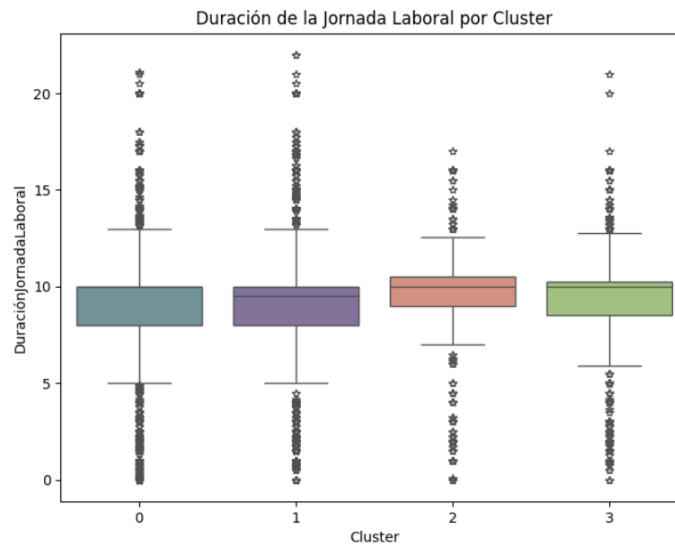


Figura 66. Boxplot duración jornada laboral por clúster

Fuente: Elaboración propia

Cluster 0:

Los empleados del cluster 0 generalmente comienzan su jornada laboral temprano en la mañana, con un pico significativo entre las 6:00 y las 8:00 de la mañana. La mayor parte de estos empleados salen del trabajo entre las 16:00 y las 18:00 horas, con un pico notable a las 17:00. La duración de la jornada laboral para este grupo es bastante variada, pero la mayoría trabaja alrededor de 8 a 9 horas al día. Este patrón sugiere que los empleados de este grupo probablemente prefieren un horario de trabajo tradicional, comenzando y terminando sus días de trabajo en horarios convencionales.

Cluster 1:

Los empleados del cluster 1 también tienden a iniciar su jornada laboral temprano, con el mayor número comenzando entre las 6:00 y las 9:00 de la mañana. Sin embargo, tienen una mayor concentración de inicios de jornada a las 7:00 y 8:00 de la mañana. En cuanto a las salidas, estos empleados suelen terminar su jornada laboral entre las 15:00 y las 18:00, con un pico significativo a las 17:00. La duración de la jornada laboral en este grupo es un poco más larga, en promedio, con muchos empleados trabajando alrededor de 9 a 10 horas diarias. Esto sugiere que los empleados del cluster 1 podrían estar trabajando horas extras o tener jornadas más intensas.

Cluster 2:

El cluster 2 muestra una distribución más uniforme en los horarios de inicio y salida laboral. La mayoría de los empleados comienzan su jornada entre las 6:00 y las 9:00 de la mañana, con un pico alrededor de las 7:00 y 8:00. Las horas de salida están más dispersas, pero predominan entre las 15:00 y las 18:00, con el pico más alto a las 17:00. La duración de la jornada laboral para este

grupo es considerablemente larga, con muchos empleados trabajando entre 9 y 10 horas diarias, y algunos incluso más. Este patrón indica que los empleados del cluster 2 podrían tener responsabilidades laborales que requieren horarios extendidos.

Cluster 3:

Los empleados del cluster 3 también inician su jornada laboral temprano, principalmente entre las 6:00 y las 8:00 de la mañana, con un notable pico a las 7:00. En cuanto a la hora de salida, la mayoría termina su jornada entre las 16:00 y las 18:00 horas, con un pico a las 17:00. La duración de la jornada laboral en este grupo es similar a la de los clusters 1 y 2, con la mayoría de los empleados trabajando alrededor de 9 horas diarias. Esto sugiere que los empleados del cluster 3 tienen horarios de trabajo relativamente estables y tradicionales, pero con una tendencia a jornadas completas y productivas.

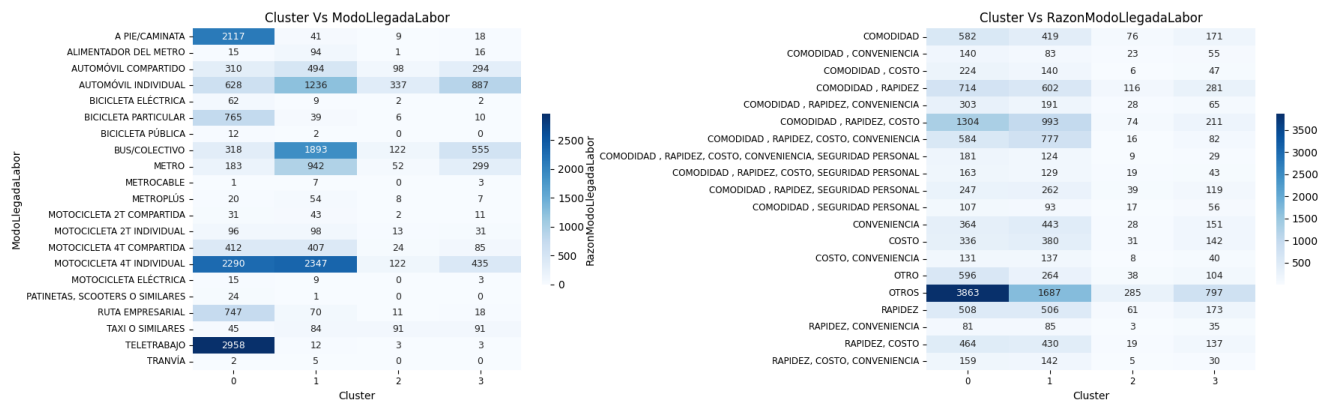


Figura 67. Mapa de calor de frecuencia del modo de transporte por clúster

Fuente: Elaboración propia

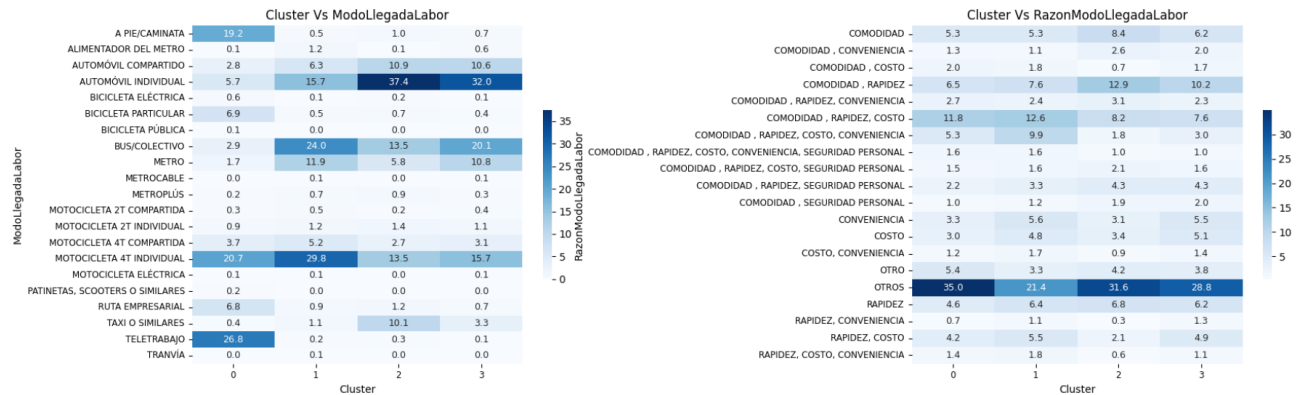


Figura 68. Mapa de calor porcentual del modo de transporte por clúster

Fuente: Elaboración propia

Cluster 0:

El Cluster 0 presenta una diversidad de modos de transporte utilizados para llegar al trabajo, con una alta prevalencia de empleados que caminan o utilizan el transporte público. Un gran número de empleados (2.117) llega al trabajo a pie, reflejando una posible proximidad a sus lugares de trabajo. El uso del bus o colectivo también es común, con 318 empleados utilizando este medio. El transporte privado, como automóviles individuales (628) y motocicletas de 4T individuales (2.290), también tiene una presencia notable. Las razones principales para elegir estos modos de transporte incluyen la comodidad, la rapidez y otras razones diversas, con un número significativo de empleados (3.863) citando estas motivaciones. Esto sugiere que el Cluster 0 valora altamente la conveniencia y la eficiencia en sus desplazamientos diarios, lo que se alinea con sus bajas huellas de carbono y energéticas.

Cluster 1:

El Cluster 1 muestra un uso más equilibrado de diferentes modos de transporte, con un número

considerable de empleados utilizando automóviles individuales (1.236) y motocicletas de 4T individuales (2.347). El transporte público también es popular, con 942 empleados usando el metro y 1893 utilizando el bus o colectivo. La comodidad, combinada con la rapidez y el costo, son las razones más citadas para sus elecciones de transporte. Por ejemplo, 1.304 empleados mencionan la combinación de estos factores como sus principales razones. Este patrón sugiere que los empleados del Cluster 1 buscan un equilibrio entre la comodidad y la economía en sus opciones de transporte, lo que se refleja en sus huellas moderadas de carbono y tiempo dedicado a la movilidad.

Cluster 2:

El Cluster 2 tiene una menor cantidad de empleados en comparación con los otros clusters y muestra una preferencia significativa por el uso de automóviles individuales (337) y motocicletas de 4T individuales (122). Aunque el transporte público como el bus o colectivo es utilizado, es en menor medida (122 empleados). Las razones para la elección de transporte varían, pero la comodidad y la rapidez son predominantes, con 285 empleados citando estas motivaciones. La menor proporción de empleados utilizando transporte público sugiere que el Cluster 2 podría tener un enfoque más individualista en sus desplazamientos, alineándose con sus huellas energéticas y de carbono más altas.

Cluster 3:

El Cluster 3 muestra una alta dependencia del transporte privado, con 887 empleados utilizando automóviles individuales y 435 usando motocicletas de 4T individuales. El transporte público es menos común, con 299 empleados utilizando el metro y 555 el bus o colectivo. Las razones

principales para elegir su modo de transporte incluyen la comodidad, junto con la rapidez y el costo, aunque en menor grado que en otros clusters. Un número significativo de empleados (797) menciona razones variadas para su elección de transporte, indicando una diversidad en las prioridades y necesidades de transporte dentro de este grupo. Este perfil de transporte privado y las razones para su elección se reflejan en sus huellas energéticas y económicas moderadamente altas, así como en los tiempos de desplazamiento más largos.

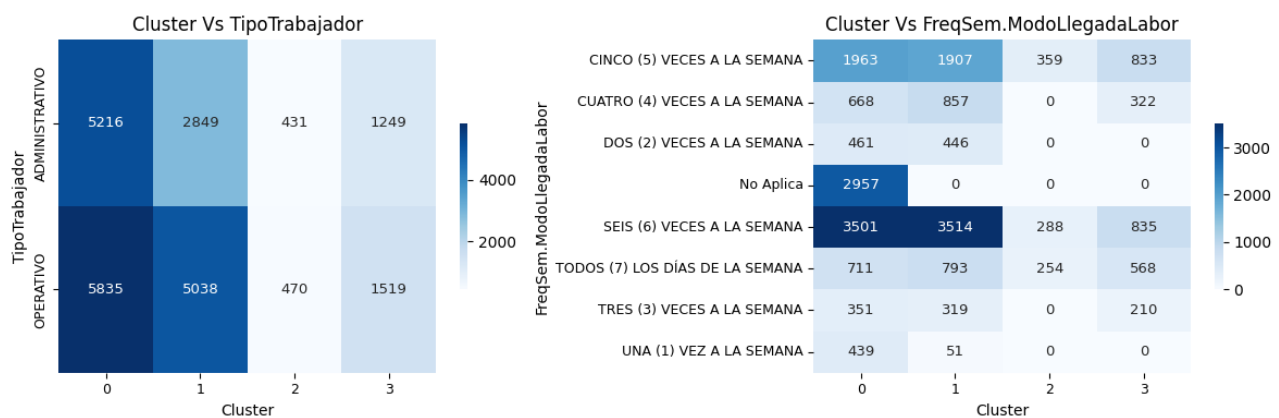


Figura 69. Mapa de calor de frecuencia de tipo labor y cantidad de uso por clúster

Fuente: Elaboración propia

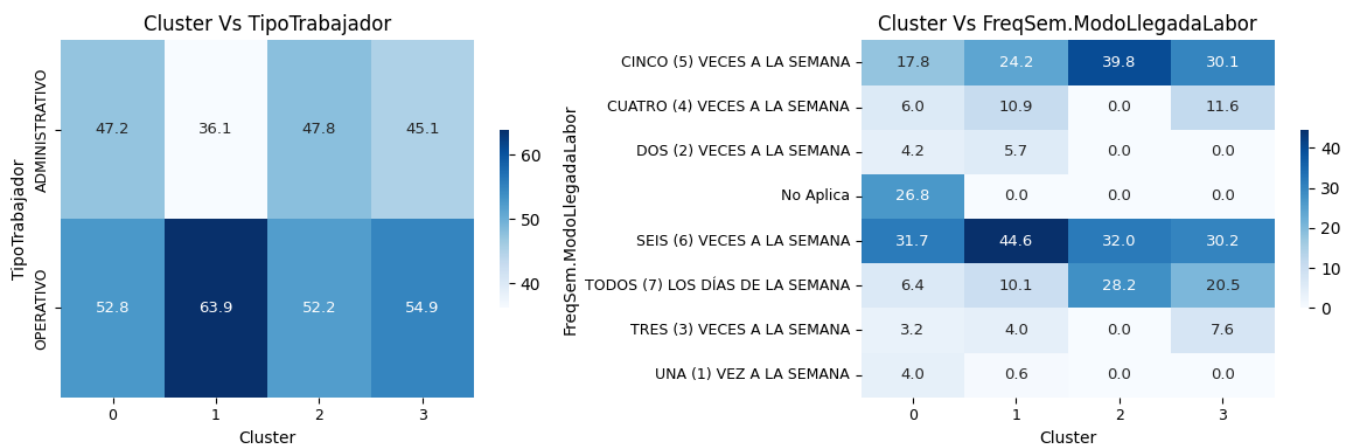


Figura 70. Mapa de calor porcentual de tipo labor y cantidad de uso por clúster

Fuente: Elaboración propia

Cluster 0:

El Cluster 0 está compuesto en gran parte por empleados operativos (52,8%) y administrativos (47,2%). En términos de frecuencia del uso del transporte, una gran cantidad de empleados utiliza el transporte seis veces a la semana (31,7%) o cinco veces a la semana (17,8%). Un grupo significativo también utiliza el transporte todos los días (6,4%) o cuatro veces a la semana (6,0%). Esta alta frecuencia de uso del transporte sugiere que los empleados de este grupo dependen en gran medida de sus modos de transporte para llegar al trabajo casi todos los días laborales, lo cual está alineado con sus bajas huellas económicas y de tiempo, posiblemente debido a la eficiencia en sus elecciones de transporte.

Cluster 1:

El Cluster 1 tiene una mayor proporción de empleados operativos (63,9%) en comparación con los administrativos (36,1%). La mayoría de los empleados utiliza el transporte seis veces a la semana (44,6%) o cinco veces a la semana (24,2%), lo que indica una fuerte dependencia del transporte para llegar al trabajo casi todos los días laborales. Un número significativo también utiliza el transporte todos los días (10,1%). La alta frecuencia de uso del transporte y la mezcla de tipos de empleados sugiere que este grupo valora tanto la conveniencia como la eficiencia en sus desplazamientos, reflejado en sus huellas de carbono y energéticas moderadas.

Cluster 2:

El Cluster 2 está equilibrado entre empleados operativos (52,2%) y administrativos (47,8%). La mayoría de los empleados en este grupo utiliza el transporte seis veces a la semana (32,0%) o cinco

veces a la semana (39,8%). Un número notable de empleados también utiliza el transporte todos los días (28,2%). Esta alta frecuencia de uso del transporte sugiere que los empleados del Cluster 2 dependen fuertemente de sus modos de transporte, lo cual se alinea con sus huellas de carbono y energéticas más altas, probablemente debido a un uso intensivo de vehículos personales.

Cluster 3:

El Cluster 3 presenta una proporción de empleados operativos (54,9%) y administrativos (45,1%) similar a la de otros clusters. La frecuencia de uso del transporte es alta, con la mayoría utilizando el transporte seis veces a la semana (30,2%) o cinco veces a la semana (30,1%). Un número significativo también usa el transporte todos los días (20,5%). Esta dependencia del transporte refleja una rutina diaria consistente de desplazamientos al trabajo, alineándose con sus huellas económicas y de tiempo relativamente altas, indicativas de una mayor distancia o tiempo de viaje diario o semanal (27,5%).

6.6. ANALISIS DEL MODELO

El **cluster 0** se caracteriza por un bajo gasto en movilidad, con un promedio de \$6.749,71 COP por semana, sugiriendo que muchos empleados trabajan desde casa o utilizan medios de transporte gratuitos como caminar o andar en bicicleta. Tienen la huella energética más baja, con un promedio de 0,01 galones por semana y una huella de carbono mínima de 65,53 gramos de CO₂ por semana. En promedio, dedican 3,78 horas por semana a la movilidad, generalmente en un horario de trabajo tradicional de 8 a 9 horas diarias, comenzando entre las 6:00 y 8:00 am y terminando entre las 4:00 y 6:00 pm. Los modos de transporte preferidos incluyen caminar y el uso del bus o colectivo, motivados por la comodidad y la rapidez. La mayoría de los empleados son operativos (52,8%) y

utilizan el transporte cinco o seis veces a la semana. En resumen, los empleados del Cluster 0 optimizan sus desplazamientos en términos de tiempo y consumo energético, valorando la conveniencia y la eficiencia, lo que les permite disfrutar de una mejor calidad de vida y un menor impacto ambiental.

El **Cluster 1** gasta más en movilidad, con un promedio semanal de \$58.381,15 COP, utilizando una combinación de transporte público y vehículos personales. Presenta una huella energética ligeramente superior a la del Cluster 0, con un promedio de 0,01 galones por semana y una huella de carbono moderada de 160,35 gramos de CO₂ por semana. Dedicán en promedio 7,36 horas por semana a la movilidad, trabajando alrededor de 9 a 10 horas diarias, comenzando entre las 6:00 y 9:00 am y terminando entre las 3:00 y 6:00 pm. Utilizan automóviles individuales, motocicletas y transporte público, motivados por la comodidad, la rapidez y el costo. La mayoría son empleados operativos (63,9%) que utilizan el transporte seis veces a la semana. Este grupo busca un equilibrio entre la comodidad y la economía en sus opciones de transporte, manteniendo huellas de carbono y energéticas moderadas.

El **Cluster 2** tiene los gastos de movilidad más altos, con un promedio de \$190.783,01 COP por semana, indicando una fuerte dependencia de vehículos personales. Este grupo presenta la huella energética más alta, con un promedio de 0,04 galones por semana, y la huella de carbono más alta, con 372,38 gramos de CO₂ por semana. Dedicán en promedio 9,20 horas por semana a la movilidad, sugiriendo que viven lejos de su lugar de trabajo y tienen jornadas laborales largas de 9 a 10 horas diarias. Prefieren automóviles individuales y motocicletas, motivados principalmente por la comodidad y la rapidez. El grupo está equilibrado entre empleados operativos (52,2%) y

administrativos (47,8%), utilizando el transporte cinco o seis veces a la semana. Los empleados del Cluster 2 enfrentan un impacto negativo en su bienestar personal y ambiental, beneficiándose de iniciativas que promuevan el uso de transporte más eficiente y opciones de trabajo remoto.

El **Cluster 3** gasta una cantidad significativa en movilidad, con un promedio de \$115.791,59 COP por semana, utilizando una mezcla de transporte público y vehículos personales. Tiene una huella energética moderada de 0,03 galones por semana y una huella de carbono intermedia de 265,54 gramos de CO₂ por semana. Dedican en promedio 8,78 horas por semana a la movilidad, trabajando alrededor de 9 horas diarias, comenzando principalmente entre las 6:00 y 8:00 am y terminando entre las 4:00 y 6:00 pm. Utilizan principalmente automóviles individuales y motocicletas, motivados por la comodidad, la rapidez y el costo. El grupo está compuesto por un 54,9% de empleados operativos y un 45,1% de administrativos, utilizando el transporte cinco o seis veces a la semana. Los empleados del Cluster 3 tienen huellas energéticas y tiempos de desplazamiento moderados, beneficiándose de estrategias que promuevan el uso de transporte más eficiente y reduzcan los tiempos de desplazamiento, mejorando así su calidad de vida y contribuyendo a la sostenibilidad ambiental.

Finalmente, se construyó un dashboard en PowerBI que surgió como una necesidad adicional al análisis del modelo de clustering, con el objetivo de proporcionar un análisis detallado y discriminado de los indicadores de movilidad para cada una de las empresas vinculadas al programa de Planes MES. Mientras que el modelo de clustering se centra en identificar patrones y agrupaciones generales en los datos, el dashboard se orienta a ofrecer una visión más granular y accesible para los diferentes grupos de interés. Este enfoque permite a las empresas consultar y

filtrar la información según las distintas condiciones de sus empleados, facilitando así una comprensión más precisa del comportamiento diario de los indicadores de movilidad.

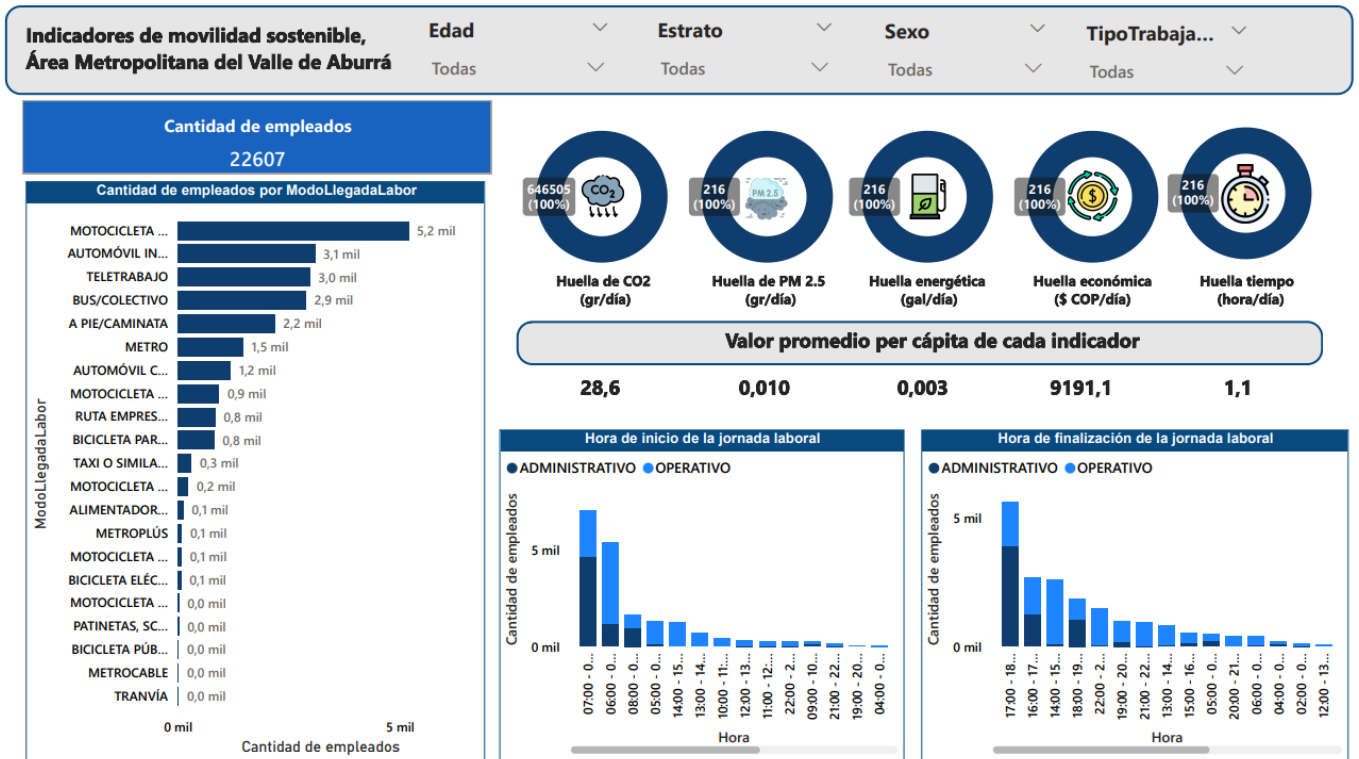


Figura 71. Dashboard indicadores movilidad sostenible

Fuente: Elaboración propia

La estructura del dashboard se definió en colaboración con funcionarios del Área Metropolitana del Valle de Aburrá para garantizar que la información sea valiosa y relevante. Además, se optó por presentar los indicadores per cápita, ya que este enfoque permite una comparación más equitativa entre diferentes grupos de empleados y empresas. El análisis per cápita ofrece una perspectiva más clara sobre el impacto individual de la movilidad y facilita la identificación de áreas específicas para la mejora. En el contexto del dashboard, esto resulta crucial para ajustar las

estrategias de movilidad y sostenibilidad de manera precisa y efectiva, alineando los esfuerzos con las necesidades particulares de cada empresa y sus empleados.

7. CONCLUSIONES Y TRABAJOS FUTUROS

7.1. CONCLUSIONES

7.1.1. Conclusiones sobre objetivos de la investigación

La investigación confirmó la relevancia de una variedad de indicadores de sostenibilidad para la evaluación de la movilidad urbana. Inicialmente, el análisis se centraba únicamente en la huella de carbono; sin embargo, la integración de indicadores como la huella de PM_{2,5}, la huella energética, la huella económica y el tiempo dedicado a la movilidad ha enriquecido considerablemente la comprensión de los impactos asociados con el transporte urbano.

Estos indicadores adicionales permiten una visión más holística de la sostenibilidad en la movilidad, abarcando aspectos ambientales, sociales y económicos. Por ejemplo, la huella de PM_{2,5} ofrece una evaluación de la calidad del aire, mientras que la huella energética y económica proporcionan una perspectiva sobre el consumo de recursos y el impacto financiero. El tiempo dedicado a la movilidad, por su parte, refleja cómo los desplazamientos afectan la calidad de vida. Este enfoque integral facilita una toma de decisiones más informada y adaptable, contribuyendo a la formulación de políticas públicas más efectivas y sostenibles.

Por otro lado, el diseño de estrategias para la recolección y depuración de datos reveló desafíos significativos y lecciones importantes. Inicialmente, se identificó la falta de claridad en los datos proporcionados por el Área Metropolitana del Valle de Aburrá y más adelante, se enfrentaron problemas adicionales relacionados con la estructura de los datos, lo cual dificultó su análisis y uso de estos para Modelos de Machine Learning. Estos desafíos permitieron validar que la limpieza y preparación meticulosa de datos son aspectos críticos en proyectos de ciencia de datos, resaltando

la importancia de una colaboración efectiva con especialistas y expertos de cada negocio para asegurar una base de datos confiable y útil en proyectos de ciencia de datos.

Garantizar dicha limpieza y transformación correcta de los datos aseguró la compatibilidad con los algoritmos de clustering. La selección del número de clústeres utilizando la medida de WCSS y el método del codo permitió identificar que cuatro grupos eran óptimos para la agrupación de los datos. Adicional a esto, la transformación de variables categóricas a numéricas mediante Label Encoding y One-Hot Encoding fue determinante para asegurar que los datos fueran compatibles con los algoritmos de clustering.

Se implementaron y compararon tres algoritmos de clustering: K-Means, Hierarchical Agglomerative Clustering y K-Medoids. Cada algoritmo fue evaluado utilizando métricas como la inercia y el coeficiente de silueta. Los resultados mostraron que el algoritmo K-Means presentó el mejor desempeño con un puntaje de silueta de 0,6926, indicando clusters bien definidos y densos. La selección del modelo K-Means, basado en la evaluación exhaustiva de la calidad de los clusters, subraya la importancia de utilizar métodos cuantitativos y visualización para tomar decisiones informadas.

Es así como el objetivo general de construir modelos de Machine Learning aplicados a un sistema de indicadores que evalúen la sostenibilidad en la movilidad urbana del Valle de Aburrá se cumplió de manera efectiva y significativa. A través del análisis de clustering, se identificaron distintos patrones de comportamiento en la movilidad urbana, reflejados en variables como el gasto en transporte, las huellas de carbono y energéticas, y los tiempos de desplazamiento. Estos hallazgos proporcionan una comprensión profunda de cómo diferentes grupos de población se movilizan en

la región, lo cual es crucial para diseñar políticas públicas que promuevan prácticas de transporte más sostenibles y eficientes.

La implementación de un dashboard en PowerBI ha sido crucial para visualizar los indicadores de sostenibilidad derivados del tratamiento y transformación de la base de datos. Este dashboard no solo facilita la visualización clara de los datos, sino que también permite compartir resultados con las empresas vinculadas a la estrategia PMES, un aspecto que previamente no era posible debido a la falta de una plataforma de visualización adecuada. El impacto de esta herramienta en el cumplimiento del objetivo es significativo, ya que proporciona a los ciudadanos y responsables de políticas una herramienta accesible para tomar decisiones basadas en datos, promoviendo así prácticas de transporte más sostenibles y eficientes en la región.

En conclusión, El análisis de clustering no solo ayuda a comprender y segmentar los datos, sino que también sienta las bases para desarrollar modelos supervisados que puedan predecir, clasificar y evaluar de manera más precisa. La identificación de patrones y características a través del clustering facilita la construcción de modelos más específicos y dirigidos, mejorando tanto la precisión como la utilidad de los análisis posteriores.

7.1.2. Conclusiones sobre el tipo de modelo de ciudad

Caminar y Pedalear:

Los datos muestran que los empleados en el cluster 0, con menores gastos y tiempos de desplazamiento, podrían estar beneficiándose de un entorno más cercano y accesible, posiblemente alineado con los principios de caminabilidad. Sin embargo, para los clusters 1, 2 y 3, donde los

tiempos y gastos de movilidad son más altos, se identifican oportunidades para mejorar la infraestructura peatonal y ciclista. La promoción de redes seguras y cómodas para peatones y ciclistas puede reducir la dependencia del transporte motorizado y fomentar una mayor integración de estos modos en la vida diaria.

Conectar y Transportar:

La conectividad de las rutas peatonales y ciclistas con el transporte público de calidad es esencial para una ciudad compacta. La investigación sugiere que, para los clusters con mayores tiempos de desplazamiento, es crucial garantizar que las rutas de transporte público estén bien conectadas con las zonas residenciales y que las redes sean más rápidas y accesibles en comparación con los trayectos en automóvil. Esto no solo facilitaría el acceso al transporte público, sino que también incentivaría su uso.

Mezclar y Densificar:

La mezcla de usos urbanos y la densificación son fundamentales para reducir la necesidad de desplazamientos largos. Los clusters que muestran altos gastos y tiempos de movilidad podrían beneficiarse de una mayor densificación de servicios y viviendas. Integrar una variedad de servicios y viviendas asequibles dentro de un área compacta puede promover una vida urbana vibrante y reducir la dependencia de transporte motorizado.

Compactar y Cambiar:

El Valle de Aburrá, en su estado actual, puede ser caracterizado en parte como una ciudad difusa, con áreas extensas de desplazamiento que reflejan la necesidad de un desarrollo más compacto.

Avanzar hacia un modelo de ciudad compacta implicará enfocar el desarrollo en centros urbanos existentes, minimizando la expansión urbana y promoviendo alternativas al uso del automóvil. La reducción del espacio ocupado por vehículos y la promoción de opciones de transporte sostenibles son pasos cruciales hacia este objetivo.

La investigación ha confirmado que la segmentación en cuatro clústeres a través de modelos de Machine Learning ofrece una perspectiva integral y valiosa para la evaluación de la sostenibilidad en la movilidad urbana del Valle de Aburrá. Cada clúster revela patrones distintivos en el comportamiento de movilidad, reflejando variaciones en las huellas de carbono, PM2,5, energética, económica y el tiempo dedicado a la movilidad. Estos hallazgos proporcionan una visión detallada de cómo los distintos grupos de población impactan y experimentan la sostenibilidad, permitiendo una toma de decisiones más informada y dirigida.

Finalmente, este estudio establece un precedente al identificar patrones de movilidad y necesidades de infraestructura que son cruciales para el desarrollo hacia una ciudad compacta. La investigación subraya la necesidad de ajustar las políticas urbanas y de transporte para alinearse con los principios del modelo de ciudades compactas, promoviendo entornos que reduzcan los desplazamientos extensos y fomenten la sostenibilidad. Los cuatro clústeres no solo ilustran comportamientos clave, sino que también proporcionan una base sólida para el desarrollo de estrategias que mejoren la calidad de vida y promuevan una movilidad urbana más sostenible.

7.2. TRABAJOS FUTUROS

Refinamiento de Indicadores de Sostenibilidad: Continuar desarrollando y refinando indicadores que abarquen no solo aspectos ambientales como las emisiones de CO2 y la huella

energética, sino también considerar otros factores clave como la infraestructura vial, la seguridad, la equidad en el acceso al transporte y la calidad de vida.

Desarrollo de herramientas de visualización: Continuar desarrollando herramientas de visualización accesibles para facilitar la toma de decisiones informadas por parte de ciudadanos y planificadores urbanos.

Validación y actualización de datos: Establecer procesos robustos para la recolección, validación y actualización continua de datos relacionados con la movilidad urbana. Esto incluye trabajar en colaboración con expertos y entidades relevantes para asegurar la calidad y la integridad de los datos utilizados en análisis futuros.

Investigación en nuevas tecnologías: Explorar y aplicar tecnologías avanzadas y métodos analíticos en ciencia de datos y Machine Learning para mejorar la precisión y eficiencia de los modelos de sostenibilidad en la movilidad urbana. Además, desde la subdirección de cooperación y convenios se están desarrollando modelos utilizando la misma información, que serán comparados con técnicas de Machine Learning en futuras investigaciones. Estas iniciativas son fundamentales para fortalecer los resultados obtenidos y contribuir a ciudades más habitables, sostenibles y resilientes frente a los desafíos urbanos contemporáneos.

Estas recomendaciones apuntan a fortalecer y ampliar el impacto del estudio actual, contribuyendo a la planificación urbana sostenible y la mejora continua de la calidad de vida en áreas urbanas.

8. REFERENCIAS BIBLIOGRÁFICAS

- [1] D. Trocha Guzmán and J. Camilo Sánchez Gómez, “Análisis de la elección modal en el Área Metropolitana del Valle de Aburrá,” vol. 4, no. 2, pp. 60–63, 2021, doi: 10.32997/RJIA-vol.4.
- [2] Área Metropolitana del Valle de Aburrá, “Plan Maestro de Movilidad para el Valle de Aburrá,” Medellín, 2020.
- [3] The City Fix Learn, “Encuesta Origen-Destino EOD: Planeación y Aprovechamiento.” Accessed: Mar. 25, 2024. [Online]. Available: <https://thecityfixlearn.org/resource/modulo-3-encuesta-origen-destino-eod-planeacion-y-aprovechamiento/>
- [4] Sostenibilidad para todos, “¿Cómo conseguirá la sostenibilidad salvar el planeta?” Accessed: Mar. 25, 2024. [Online]. Available: https://www.sostenibilidad.com/desarrollo-sostenible/que-es-la-sostenibilidad/?_adin=02021864894
- [5] E. Flores, J. García, J. Chica, and E. Mora, “Identificación y análisis de indicadores de sostenibilidad para la movilidad,” *Revista de la Facultad de Arquitectura y Urbanismo de la Universidad de Cuenca*, vol. 6, no. 11, pp. 99–109, 2017, doi: 10.18537/est.v006.n011.a07.
- [6] C. F. Guerrero Dumas, C. Maldonado Noboa, D. K. Reyes Rodas, and M. Ávila Calle, “Propuesta, indicadores de movilidad sostenible para la ciudad de Cuenca, Ecuador,” *ConcienciaDigital*, vol. 5, no. 1.2, 2022, doi: 10.33262/concienciadigital.v5i1.2.2085.
- [7] G. Motos, “Análisis de indicadores de movilidad urbana sostenible,” 2019.
- [8] C. E. Grande Ayala, “Movilidad urbana socialmente sostenible, un modelo no explorado

- para las capitales del Triángulo Norte,” *ECA: Estudios Centroamericanos*, vol. 76, no. 767, 2021, doi: 10.51378/eca.v76i767.6472.
- [9] E. González, “Movilidad Urbana Sostenible,” *COMISIÓN NACIONAL PARA EL USO EFICIENTE DE LA ENERGIA*, 2018, [Online]. Available: <http://www.naibooksellers.nl/urbanism/>
- [10] S. Rueda Palenzuela, A. Santasusagna Riu, B. Cormenzana Izquierdo, J. Tort Donada, and X. Úbeda, “Understanding Urban Complexity via the Spatial Diversity of Activities: An Application to Barcelona (Spain),” *Sustainability (Switzerland)*, vol. 14, no. 3, 2022, doi: 10.3390/su14031298.
- [11] H. Uribe Bedoya, A. Valencia Arias, and S. Ramos Y Yovera, “Tendencias y evolución investigativa sobre la movilidad sostenible: una aproximación bibliométrica,” *Producción + Limpia*, vol. 14, no. 2, 2020, doi: 10.22507/pml.v14n2a5.
- [12] G. Nuri Barón, “La transición urbana y social hacia un paradigma de movilidad sostenible,” *Cuadernos del Centro de Estudios de Diseño y Comunicación*, no. 80, Aug. 2020, doi: 10.18682/cdc.vi80.3701.
- [13] S. Rueda Palenzuela, “Principios e indicadores para ciudades más sostenibles,” *Inguruak: Soziologia eta zientzia politikoaren euskal aldizkaria = Revista vasca de sociología y ciencia política*, no. 24, 1999.
- [14] S. Rueda-Palenzuela, “La complejidad urbana y su relación con la morfología de los tejidos urbanos y la proximidad,” *Ciudad y Territorio Estudios Territoriales*, vol. 54, no. M, 2022, doi: 10.37230/cytet.2022.m22.10.

- [15] M. P. León Andrade and V. V. Carriel Bustamante, “Indicadores para medir la movilidad sostenible en espacios urbanos de la sierra ecuatoriana. Caso de estudio el cantón Azogues,” *ConcienciaDigital*, vol. 4, no. 1.2, 2021, doi: 10.33262/concienciadigital.v4i1.2.1738.
- [16] C. López, “Cómo construir e implementar un Plan Integral de Movilidad Sostenible PIMS en organizaciones,” Bogotá D.C., 2023.
- [17] S. M. Palacio, “Machine Learning Forecasts of Public Transport Demand: a comparative analysis of supervised algorithms using smart card data,” *Documentos de trabajo (XREAP)*, N^o. 3, 2018, no. 3, p. 1, 2018, Accessed: Nov. 08, 2023. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=7312241&info=resumen&idioma=ENG>
- [18] B. Romero, “Una introducción a los modelos de Machine Learning,” 2020.
- [19] L. Fernanda, P. Ramírez, and S. Gongora Aya, “ALGORITMO DE SELECCIÓN Y VALIDACIÓN DEL MÉTODO DE CLUSTERIZACIÓN ÓPTIMO PARA DATOS NO SUPERVISADOS,” 2021.
- [20] M. Khakpour, “Análisis de los patrones de movilidad residencial en España mediante árboles de clasificación,” *Universidad de Cantabria*, Jul. 2021, Accessed: Nov. 08, 2023. [Online]. Available: <https://repositorio.unican.es/xmlui/handle/10902/24929>
- [21] R. López Nácher, “Diseño de un método de evaluación y selección de cartera de proyectos, basado en análisis de interesados y Proceso Analítico en Red (ANP), que incluyen: inversiones con actuaciones de energías renovables, hidrógeno renovable, movilidad sostenible y gestión de la demanda,” Oct. 2022, Accessed: Apr. 15, 2024. [Online]. Available: <https://riunet.upv.es:443/handle/10251/187182>

- [22] E. E. Morales Martínez, “Técnicas de Machine Learning y desarrollo de modelos predictivos aplicados en la movilidad de Bogotá DC,” *Universidad Tecnológica de Bolívar*, 2020.
- [23] C. Engel and L. Braganca, “Versao digital_comunidades urbanas energeticamente eficientes,” *Editora filiada à Associação Brasileira das Editoras Universitárias (Abeu)*, 2016.
- [24] Alcaldía de Santiago de Cali, “Plan de movilidad sectorial Sector Empresarial,” Santiago de Cali, Oct. 2021.
- [25] L. S. Córdova-Chavez and E. Santa María-Dávila, “Planes de desplazamiento de empresa: una propuesta para mejorar la movilidad en el área metropolitana de Lima y Callao,” *TECNIA*, vol. 31, no. 2, pp. 11–21, Jun. 2021, doi: 10.21754/tecnia.v21i2.1037.