**INTELIGENCIA ARTIFICIAL**

**Informe Técnico**

**LatAm sobre ruedas verdes: inteligencia artificial para mapear el camino hacia la neutralidad de carbono**

**Autores:**

**Daniel Vallejo Soto**

**Valeria Flórez Murillo**

**Diana Cristina Cadavid Agudelo**

**Amalia Del Pilar Castro Areiza**

**MEDELLÍN**

**2025**

**Página**

Contenido

[1.](#_heading=h.u0b5blylbs8g)  CASO DE APLICACIÓN 4

[1.1](#_heading=h.klyl1d3npbk) Entendimiento del Caso: 4

[1.2](#_heading=h.h4vsq2nwb8i4) Entendimiento de los Datos. 5

[1.3](#_heading=h.hmxbpnvd0leu) Preparación de los Datos 8

[2.](#_heading=h.n7uyi0ur7t9c) MODELAMIENTO 13

[2.1](#_heading=h.q79brzrcayxk) Justificación de la Técnica 13

[2.2](#_heading=h.bpilulj8ixee) Consideraciones al Aplicar 13

[2.3](#_heading=h.2tfpd71mku0u) Parametrización de Método seleccionado 14

[3.](#_heading=h.k800k9xs3xcq) EVALUACIÓN E IMPLEMENTACIÓN 15

[3.1](#_heading=h.a4a9fbr95818) Evaluación 15

[3.2](#_heading=h.3wyfgj9sr39e) Implementación 18

[3.3](#_heading=h.2mrblz1zy7pz) Plataforma y Herramientas 18

[4.](#_heading=h.z86tty5q7qrg) CONCLUSIONES 20

LISTA DE FIGURAS

**Página**

[Figura 1 Esquema países Latinoamericanos 4](#_heading=h.nmxv9tv4v58h)

[Figura 2 Variables utilizadas en el análisis 8](#_heading=h.wsotvym26lkq)

[Figura 3 Limpieza de datos 9](#_heading=h.s67hzg9vl9x3)

[Figura 4 Tendencia datos de Población 10](#_heading=h.s9mhnusvzcof)

[Figura 5 Tendencia datos PIB per cápita 10](#_heading=h.os0srmd1v2i4)

[Figura 6 Tendencia datos densidad 11](#_heading=h.7fe52opt0gnk)

[Figura 7 Tendencia datos emisiones de Co2 11](#_heading=h.fejwsuph2sq6)

[Figura 8 Método para la creación de índices 12](#_heading=h.hwbxhcz7xbz3)

[Figura 9 Análisis método Elbow 15](#_heading=h.q5kb1jwjwfgl)

[Figura 10 Resultado Cluster k-means 16](#_heading=h.7ix5rhu2w013)

[Figura 11 Resultado análisis DB SCAN 17](#_heading=h.yq6j5un1aoi8)

[Figura 12 Herramientas Utilizadas 19](#_heading=h.xj6maa2ackdq)

# 1. CASO DE APLICACIÓN

## Entendimiento del Caso:

La creciente necesidad de soluciones sostenibles para la movilidad urbana en América Latina ha impulsado el interés por los sistemas de transporte eléctrico de pequeña escala, como bicicletas, scooters eléctricos, motos y vehículos de baja velocidad (VLV). Estos modos de transporte ofrecen beneficios ambientales, sociales y económicos al reducir las emisiones de gases de efecto invernadero, descongestionar las vías, promover el acceso equitativo a la movilidad (ONU-Hábitat, 2020). Ante la ausencia de una clasificación estandarizada que indique qué ciudades son más viables para su implementación, se propone una solución basada en aprendizaje no supervisado mediante técnicas de clustering, que permiten identificar agrupaciones de ciudades con condiciones estructurales, ambientales y socioeconómicas similares.

Este estudio se sitúa en un contexto regional marcado por la liberalización del transporte, el debilitamiento de las empresas públicas y la proliferación de servicios informales, todos fenómenos que han profundizado la fragmentación urbana, la segregación socioespacial y la desigualdad en el acceso a la movilidad (Figueroa, 2005). Las ciudades han evolucionado hacia formas más dispersas y motorizadas, en las que el transporte formal es insuficiente o excluyente, especialmente para las poblaciones periféricas y de bajos ingresos. En este escenario, la micromovilidad eléctrica puede actuar como solución intermedia que complementa la red de transporte, reduce la dependencia del automóvil y mejora la accesibilidad urbana en territorios desatendidos.

Más allá de su función de movilidad, este proyecto se enmarca en la línea de transición energética justa, al analizar la adopción de energías limpias, promover la democratización del consumo energético, y abrir espacios para la organización de comunidades energéticas vinculadas al uso compartido de vehículos eléctricos de baja escala. Esta transición no se limita a la sustitución tecnológica, sino que busca impulsar el desarrollo comunitario y la equidad territorial, en armonía con el uso responsable de minerales y con un modelo de ciudad menos dependiente de los combustibles fósiles. Por tanto, analizar la viabilidad de implementar micro-movilidad eléctrica en ciudades latinoamericanas no solo responde a criterios técnicos y ambientales, sino que se inscribe en una apuesta estructural por la justicia energética y territorial en la región.

## Entendimiento de los Datos.

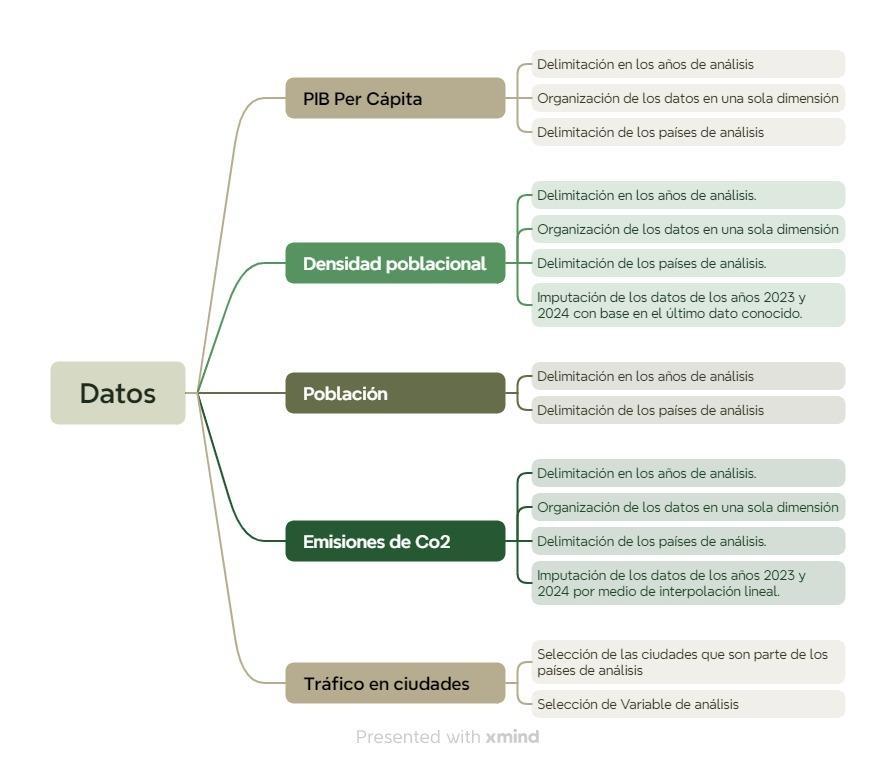
Para desarrollar el modelo de clustering, se requiere una base de datos multivariable que contenga indicadores representativos de las condiciones técnicas, ambientales y sociales de las ciudades. A continuación, se presenta un resumen de cada uno de los datos que se consideraron para el proyecto según las variables definidas:

| **Variable** | **Qué se desea conocer** | **Descripción** |
| --- | --- | --- |
| **Crecimiento PIB Per Cápita** | Capacidad económica para adquirir sistemas de transporte eléctrico a pequeña escala | Dataset consultado en el Banco Mundial, presenta una tabla principal en la que cada registro corresponde a un país y al indicador PIB per cápita seguido de una columna por cada año desde 1960 hasta 2024 con el porcentaje de variación del PIB per cápita. |
| **Emisiones de CO2 per cápita** | Emisiones de CO2 por persona en países de América Latina. | Dataset consultado en Kaggle. Está integrado por cuatro columnas en las cuales se describe el país, código del país, el año (1949 a 2021) y las emisiones anuales de dióxido de carbono expresadas en toneladas métricas por persona. Permite analizar tendencias históricas de emisiones por persona, comparar países entre sí o estudiar la evolución a lo largo del tiempo |
| **Población** | Analizar la cantidad de habitantes de un país para conocer sus posibilidades de mercado. | El dataset consultado en Our World in Data consta de tres columnas en las cuales se define el país, el año (entre 1950 y 2022) y la cantidad de población. Permite estudiar la evolución demográfica de cada país a lo largo del tiempo, ver cómo crece o decrece la población de un país año tras año y comparar diferentes naciones en un mismo año o calcular agregados regionales. |
| **Densidad** | Cantidad de habitantes por m2 | Consultado en el Banco Mundial, muestra una tabla donde cada fila corresponde a un país (identificado por su nombre y su código ISO de tres letras) y cada columna, un año entre aproximadamente 1960 y 2024, mostrando el número de personas que, en promedio, viven en cada kilómetro cuadrado de tierra en ese país durante ese año; si una casilla está en blanco, es que no se dispone de dato para ese periodo. |
| **Tráfico.** | Índice de tráfico en diferentes ciudades del mundo. | Consultado en Kaggle, recoge más de un millón de registros de mediciones de tráfico en 60 países, con datos tomados entre el 1 de octubre de 2024 y el 13 de enero de 2025. Nos permite comparar la congestión en una ciudad “hoy” vs. “hace una semana”, ver cuántos kilómetros de atascos había y cuánto han retrasado de media los desplazamientos en periodos concretos. |

*Figura 2 Variables utilizadas en el análisis*

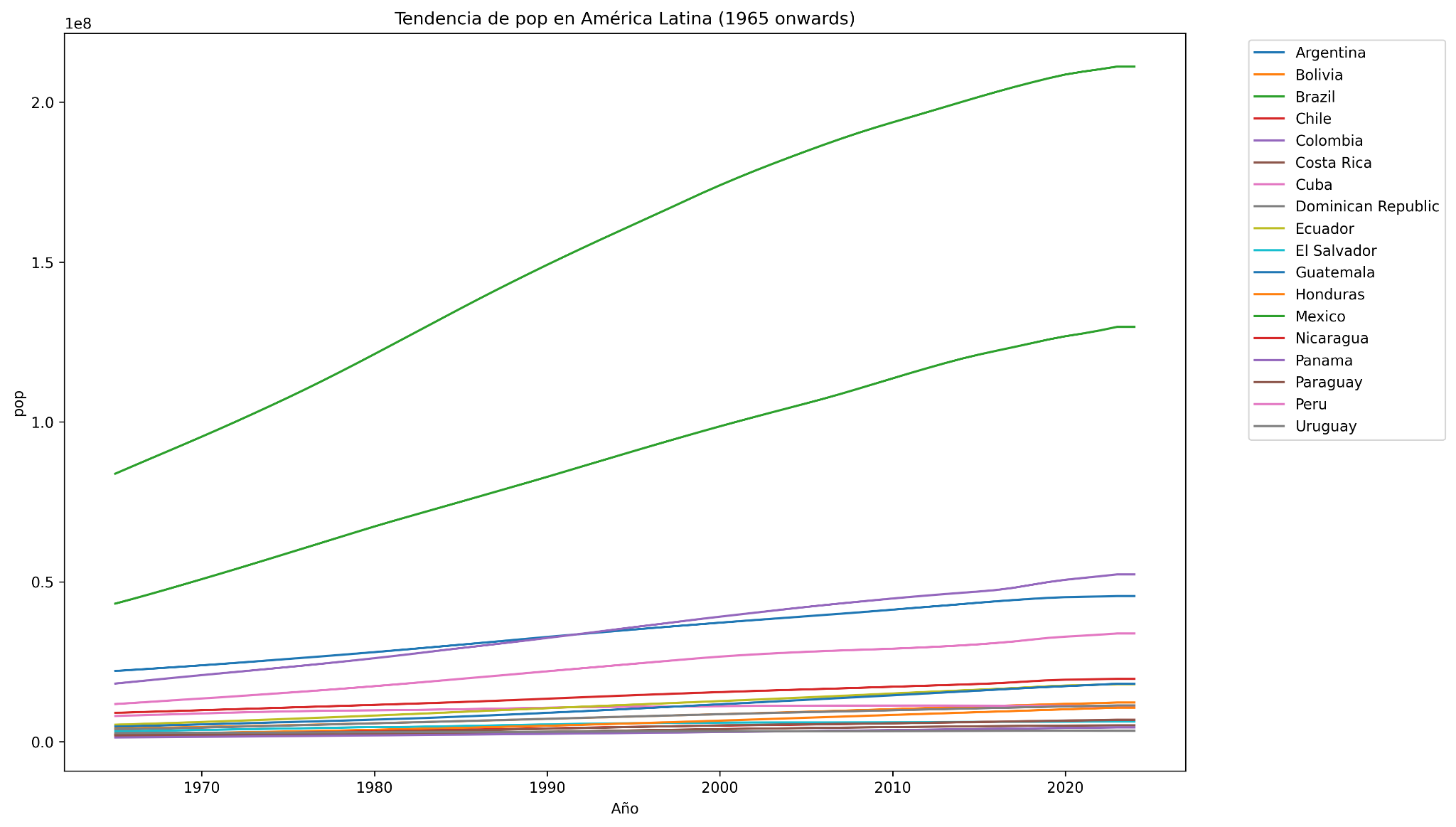
## Preparación de los Datos

Para la preparación de los datos se desarrollaron los siguientes procedimientos. en cada una de las fuentes de datos consultadas:

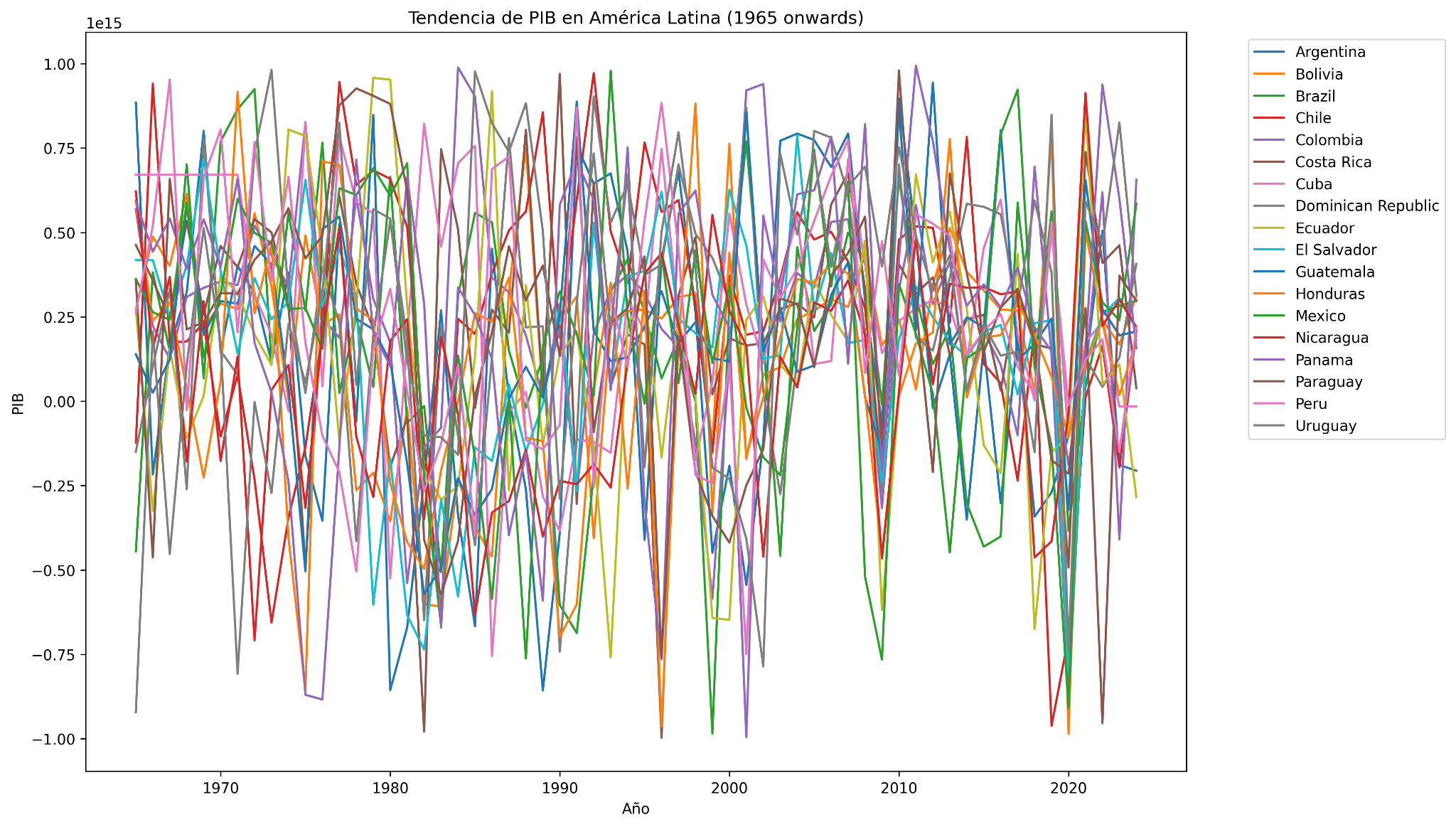


*Figura 3 Limpieza de datos*

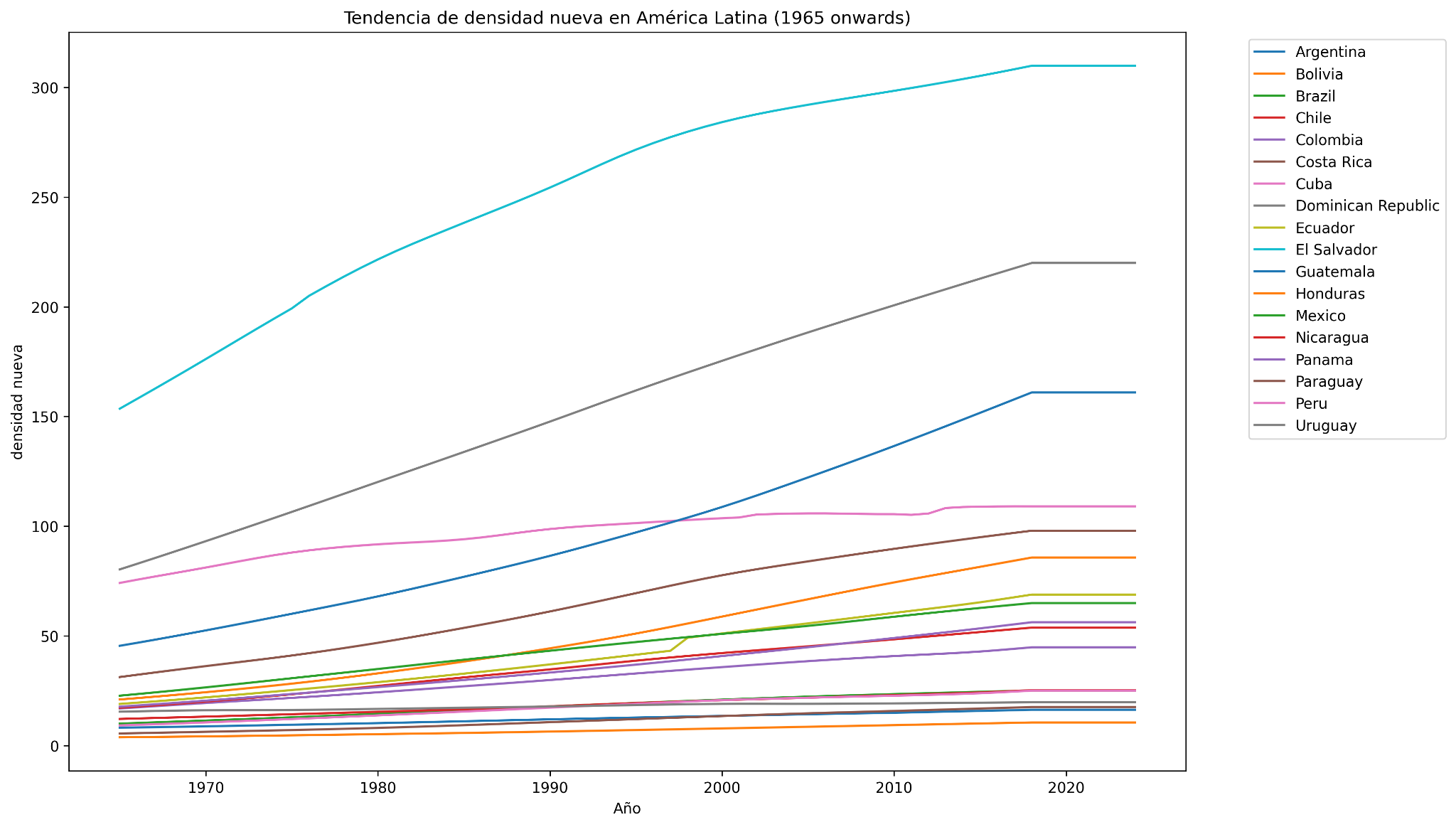
A continuación, se muestran las gráficas lineales de los datos de: PIB per cápita, densidad poblacional, población y emisiones de Co2.



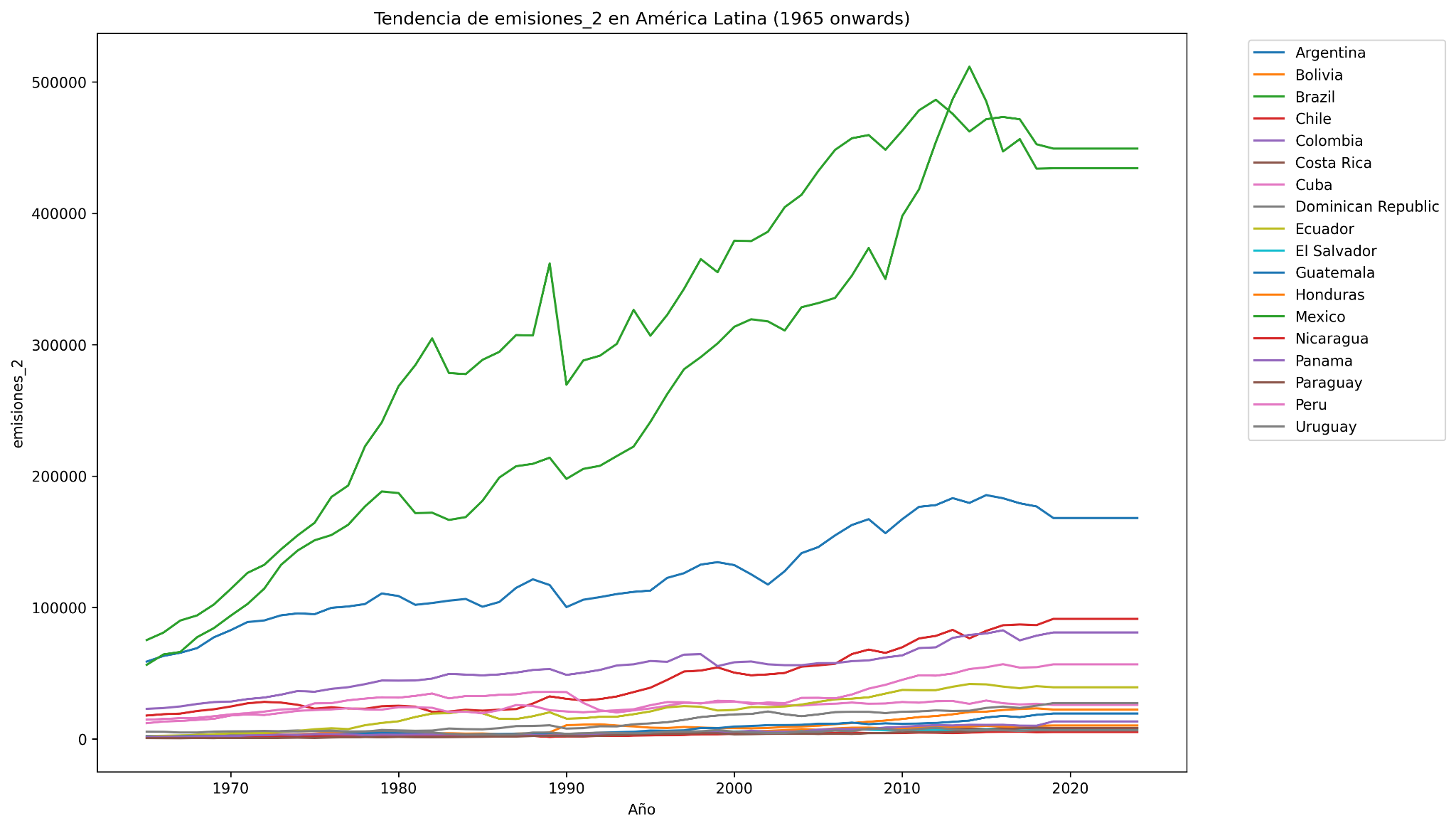
*Figura 4 Tendencia datos de Población*



*Figura 5 Tendencia datos PIB per cápita*



*Figura 6 Tendencia datos densidad*



*Figura 7 Tendencia datos emisiones de Co2*

A partir de la depuración de las distintas bases de datos se construyeron dos índices que se contrastará en el análisis de clustering. El primero mide la viabilidad de implementar la tecnología en cada país de Latinoamérica; el segundo cuantifica el nivel de tráfico en algunas de sus ciudades. La Figura 4 Método para la creación de índices sintetiza los pasos seguidos para consolidar cada índice.



*Figura 8 Método para la creación de índices*

Una vez desarrollados estos índices se une toda la información en una sola base de datos para el proceso de clusterización.

# MODELAMIENTO

## Justificación de la Técnica

La aplicación del *clustering* en contextos urbanos latinoamericanos exige una cuidadosa consideración de la naturaleza de los datos y del tipo de agrupamiento buscado. K-Means requiere datos numéricos continuos y su rendimiento óptimo se da cuando los clústeres son convexos y de tamaño similar. Este método no gestiona bien el ruido ni los datos desbalanceados, por lo que debe utilizarse cuando los indicadores presentan distribución homogénea. Se determinará el número ideal de clústeres mediante métodos como el *elbow* y el *silhouette score*.

DBSCAN, por su parte, no requiere especificar el número de clústeres y permite identificar agrupamientos con densidades variables, lo cual es útil en territorios con urbanización dispersa, como las áreas periféricas donde proliferan servicios informales (Figueroa, 2005). Sin embargo, su sensibilidad a los parámetros (*epsilon* y *minPts*) implica realizar varias pruebas para evitar sobreajuste o agrupamientos arbitrarios.

## Consideraciones al Aplicar

La aplicación del *clustering* exige una cuidadosa consideración de la naturaleza de los datos y del tipo de agrupamiento buscado. Dado que se trabajará con variables de distinta escala, será necesario normalizar los datos antes de aplicar cualquier algoritmo.

**K-Means** requiere datos numéricos continuos y su rendimiento óptimo se da cuando los clústeres son convexos y de tamaño similar. Este método no gestiona bien el ruido ni los datos desbalanceados, por lo que debe utilizarse cuando los indicadores presentan distribución homogénea. Se determinará el número ideal de clústeres mediante métodos como el *elbow* y el *silhouette score*.

**DBSCAN**, por su parte, no requiere especificar el número de clústeres y permite identificar agrupamientos con densidades variables, lo cual es útil en territorios con urbanización dispersa, como las áreas periféricas donde proliferan servicios informales (Figueroa, 2005). Sin embargo, su sensibilidad a los parámetros (*epsilon* y *minPts*) implica realizar varias pruebas para evitar sobreajuste o agrupamientos arbitrarios.

## Parametrización de Método seleccionado

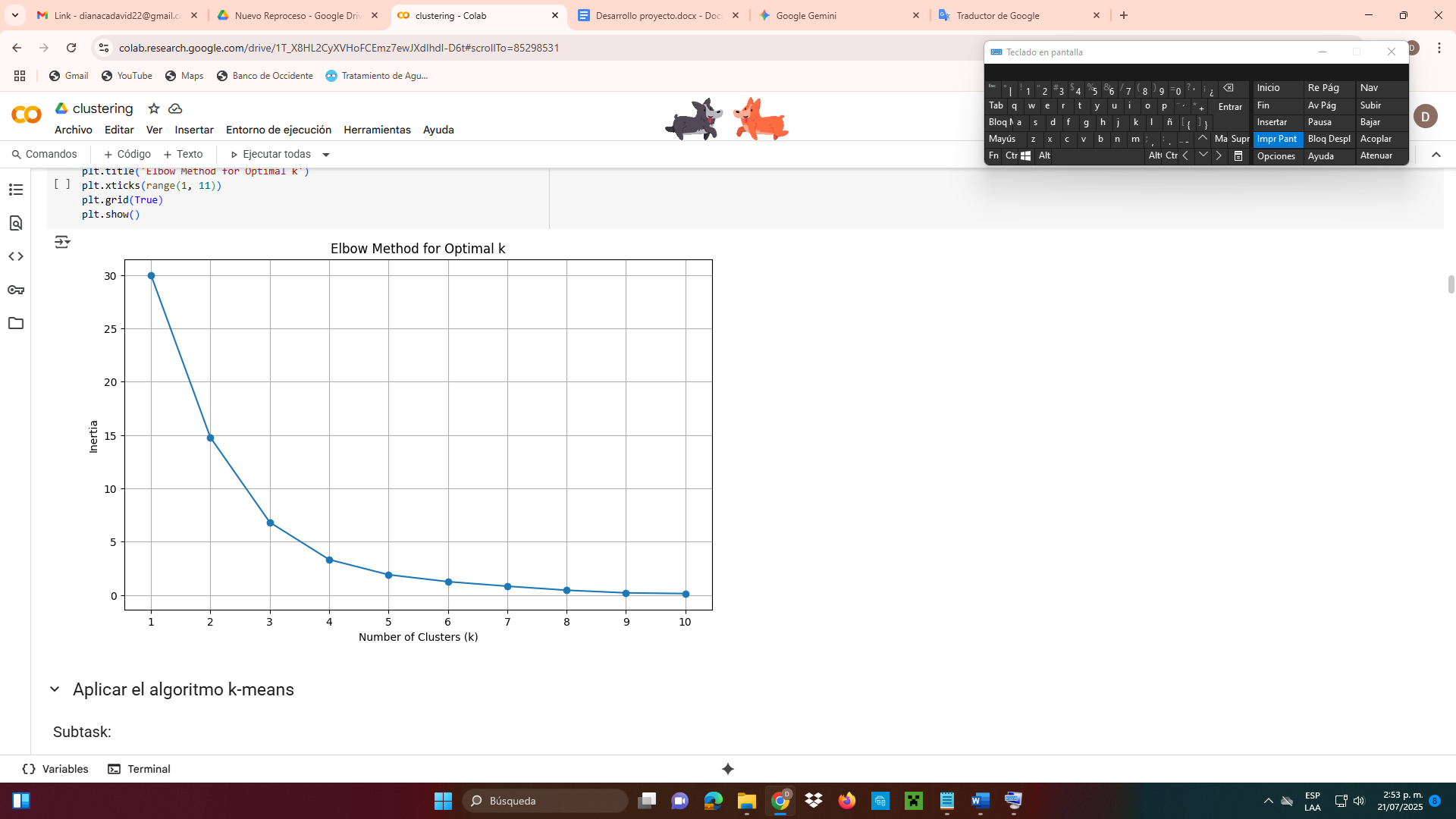
Los algoritmos de clustering seleccionados (K-Means y DBSCAN) requieren como entradas principales: datos normalizados y una métrica de distancia (usualmente euclidiana). En el caso de K-Means, las salidas típicas incluyen la asignación de etiquetas a cada ciudad según su grupo y los centroides que representan el perfil promedio de cada clúster. Para DBSCAN, las salidas son las etiquetas de clúster y la detección de ciudades atípicas (outliers), que no pertenecen a ningún grupo. Estas salidas permiten identificar tanto patrones dominantes como excepciones territoriales en la región.

# EVALUACIÓN E IMPLEMENTACIÓN

## Evaluación

Para nuestro proyecto se utilizaron los modelos de clustering (K-Means y DBSCAN), para lo cual utilizamos dataset de Crecimiento PIB Per Cápita, Emisiones de CO2 per cápita, Población, Densidad y Tráfico de ciudades en América Latina para identificar en qué ciudades de América Latina se cumplen mejor las condiciones para aumentar o implementar la venta de vehículos eléctricos de pequeña escala.

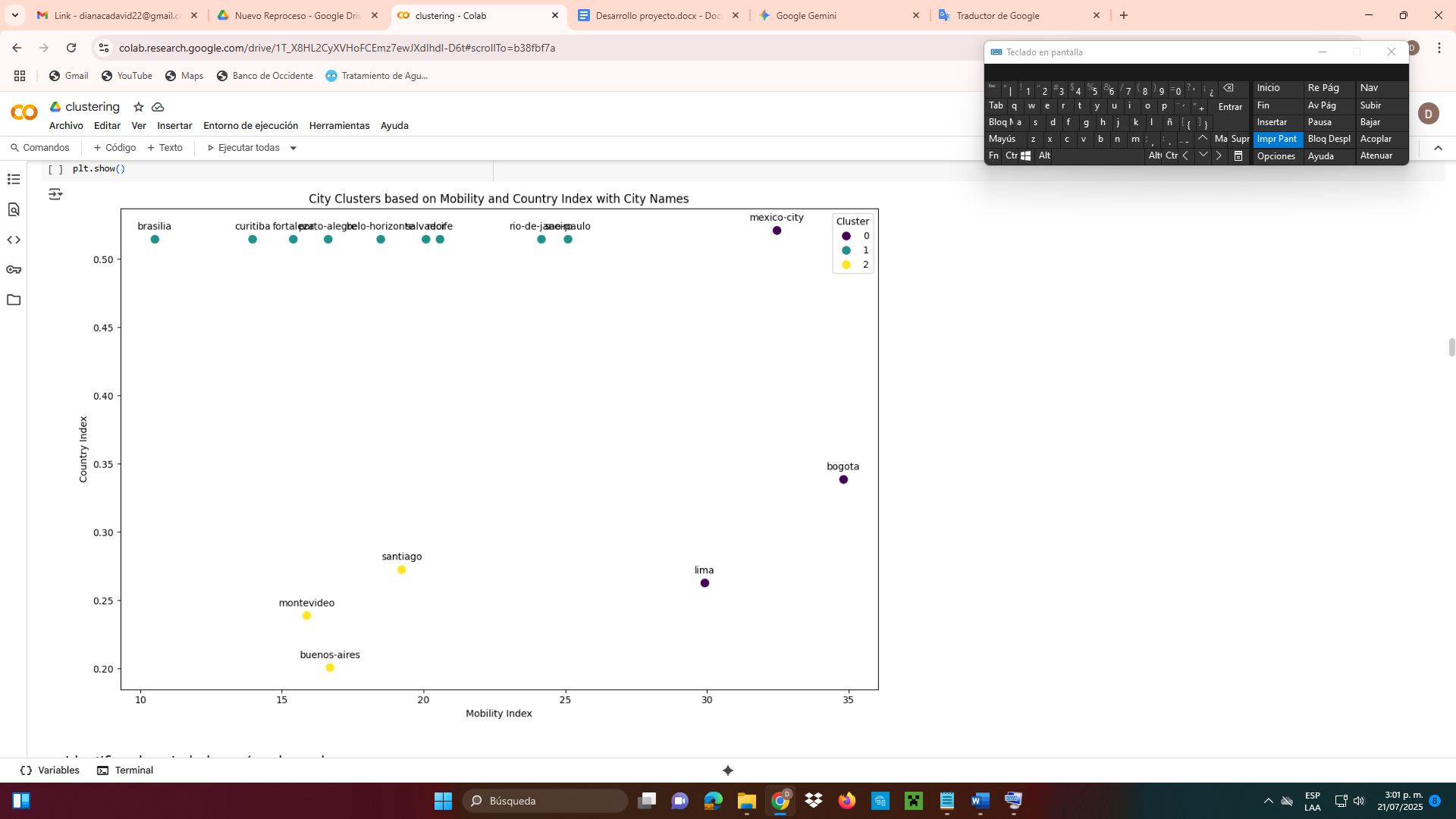
Se Determinó el número óptimo de clusters (k), utilizando un método como el método del codo (Elbow Method) para encontrar un número apropiado de clusters para K-Means:



*Figura 9 Análisis método Elbow*

Procedimos a ejecutar la agrupación en clústeres K-Means con la cantidad óptima de clústeres (k = 3, según el método Elbow del paso anterior) y se agregaron las etiquetas de los clústeres al marco de datos original.

Se obtuvo como resultado el siguiente diagrama de dispersión para visualizar los clústeres y las ciudades asignadas en función de la movilidad y el índice de país, coloreando puntos por clúster y agregando etiquetas y un título para mayor claridad:



*Figura 10 Resultado Cluster k-means*

Al comparar K-Means y DBSCAN, K-Means dividió todas las ciudades en 3 clústeres que representan diferentes niveles de potencial, mientras que DBSCAN destacó un grupo denso específico de ciudades y clasificó otras como valores atípicos. Según el análisis de DBSCAN, las ciudades del Clúster 0, se consideran las más prometedoras debido a que conforman un clúster denso con características favorables (movilidad moderada e índice de país alto).

Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.

*Figura 11 Resultado análisis DB SCAN*

K-Means identificó tres clústeres distintos basándose en la minimización de la inercia, intentando particionar todos los puntos de datos.

DBSCAN, con los parámetros seleccionados (eps=0.5, min\_samples=5), identificó un clúster principal y trató el resto como ruido.

Los clústeres de K-Means tienden a ser más esféricos y de tamaño similar, mientras que DBSCAN puede encontrar clústeres con formas arbitrarias e identificar valores atípicos.

En este caso, K-Means agrupó las ciudades según un equilibrio entre la movilidad y el índice de país, lo que generó clusters que representan diferentes niveles de potencial (movilidad alta/país moderado, movilidad moderada/país alto, movilidad baja/país bajo).

DBSCAN, por otro lado, encontró una región densa de ciudades con movilidad moderada e índices de país altos, clasificando las ciudades fuera de esta región densa como ruido. Esto sugiere que solo un grupo específico de ciudades forma un clúster densamente agrupado en este espacio de características con los parámetros seleccionados. Para identificar ciudades prometedoras para la venta de vehículos eléctricos, K-Means ofrece una categorización más amplia en niveles potenciales.

DBSCAN destaca un conjunto muy específico y estrechamente agrupado de ciudades similares entre sí en ambos índices, a la vez que indica que otras ciudades son menos similares a este grupo principal o entre sí.

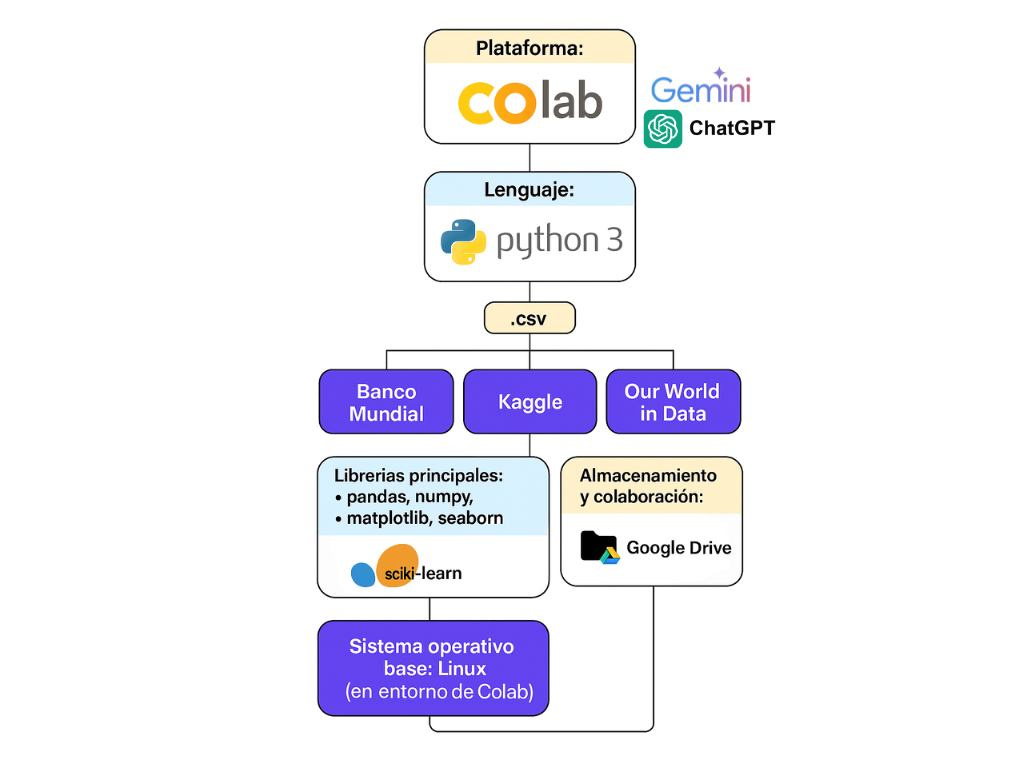
La definición de "prometedor" depende del método de agrupamiento: K-Means sugiere que las ciudades de los clústeres 0 y 1 son prometedoras, mientras que DBSCAN destaca las ciudades de su único clúster (Clúster 0) como similares y potencialmente prometedoras, pero no ofrece un potencial escalonado para los puntos de ruido.

Tras el desarrollo de los dos métodos implementados se encuentra que K-means es la técnica que representa mejor las realidades territoriales para la implementación de medios de transporte eléctricos de pequeña escala.

## Implementación

https://github.com/vallejosoto/LatAm-sobre-ruedas-verdes

## 3.3 Plataforma y Herramientas



*Figura 12 Herramientas Utilizadas*

# CONCLUSIONES

**Conclusiones del caso de aplicación:**

* La micro movilidad eléctrica representa una alternativa viable y sostenible para mejorar la movilidad urbana en América Latina, especialmente en contextos donde el transporte público es insuficiente o excluyente.
* El análisis permitió identificar qué ciudades presentan mayores condiciones para implementar este tipo de transporte, considerando factores como PIB per cápita, densidad poblacional, nivel de tráfico, y emisiones de CO₂.
* El enfoque del estudio se alinea con los principios de una transición energética justa, contribuyendo al desarrollo comunitario, la equidad territorial y la reducción del uso de combustibles fósiles.

**Conclusiones sobre la técnica seleccionada (Clustering):**

* El uso de técnicas de aprendizaje no supervisado, como el clustering, es apropiado para clasificar ciudades sin requerir etiquetas predefinidas, permitiendo revelar patrones naturales en los datos.
* El clustering permite no solo segmentar las ciudades según su viabilidad para micro movilidad, sino también destacar desigualdades y oportunidades de intervención en zonas específicas.

**Conclusiones del método seleccionado (K-Means y DBSCAN):**

* **K-Means** resultó útil para clasificar las ciudades en tres niveles de potencial (alto, medio y bajo), facilitando una interpretación escalonada para la toma de decisiones.
* **DBSCAN** permitió detectar un grupo denso de ciudades con características particularmente favorables y trató otras como ruido, lo cual es valioso para identificar núcleos altamente viables sin suponer homogeneidad en los datos.
* La combinación de ambos métodos ofreció una visión más integral: K-Means brindó una visión estructurada y escalonada, mientras que DBSCAN ayudó a identificar agrupamientos naturales con formas más flexibles.

**Consideraciones personales**:

Consideramos que el método más adecuado para este estudio fue K-Means, ya que permitió clasificar de manera clara y escalonada las ciudades latinoamericanas según su potencial para implementar micro movilidad eléctrica. Su estructura facilitó la interpretación de los resultados y la toma de decisiones estratégicas. No obstante, la falta de datos completos y actualizados limitó la posibilidad de incluir un mayor número de ciudades en el análisis, lo cual podría haber enriquecido y ampliado el alcance del estudio. A pesar de esta limitación, los resultados obtenidos fueron significativos: Ciudad de México se posicionó como la ciudad con las mejores condiciones para la implementación de micro movilidad eléctrica, seguida por Bogotá y posteriormente Lima, lo cual permite orientar futuras inversiones e intervenciones hacia estos territorios con alto potencial de adopción.