Exploración y visualización de datos para lo socioeconómico

Miguel Andrés Garzón

|  |
| --- |
| Proyecto 1: EDA, visualización y estadística para respaldar hipótesis**[[1]](#footnote-1)** |

**ELABORADO POR:**

Laura Sarif Rivera Sanabria

María Camila Caraballo Candela

Javier Antonio Amaya Nieto

**Enunciado del taller- situación**

El propósito de este proyecto es consolidar lo aprendido en la primera parte del curso, integrando el análisis exploratorio de datos (EDA) con la formulación de una prueba de hipótesis que respalde un hecho estilizado. El proyecto busca fortalecer tanto la coherencia conceptual del análisis como la argumentación basada en evidencia. Para esto, vamos a retomar lo desarrollado en la Actividad 2 y profundizaremos en la generación de evidencia sobre los hallazgos encontrados inicialmente, o su verificación y ajuste.

Su grupo de trabajo hace parte de un equipo técnico que apoya a una alcaldía en un programa de mejora de la calidad de vida de hogares venezolanos. La alcaldía quiere mejorar la focalización de su respuesta en vivienda y servicios básicos, pero no sabe qué factores son los más importantes para identificar condiciones de vida deficientes.

A partir del análisis exploratorio previamente realizado (EDA y visualización ajustado), el objetivo ahora es formular y poner a prueba una hipótesis estadística que respalde un hecho estilizado del contexto estudiado. Deberá identificar una relación o diferencia entre variables observadas en los datos que pueda expresarse mediante una hipótesis estadística contrastable. Esto es, una frase concreta, respaldada con evidencia.

# Parte 1: Análisis Exploratorio de Datos

## Formulación del problema de indagación y la claridad del contexto analítico

## El fenómeno migratorio que atraviesa Colombia, aunque no es nuevo, ha marcado un precedente significativo para el Estado en términos de articulación institucional, al configurar un nuevo mapa social, económico y político que requiere atención frente a la llegada masiva de migrantes. En los últimos años, Colombia se ha consolidado como el principal país receptor de venezolanos en la región: para el año 2025, se estima que cerca de 2,8 millones de personas han ingresado al territorio nacional.

## Este movimiento humano sin precedentes, motivado por la búsqueda de mejores condiciones de vida, plantea desafíos de gran magnitud para la capacidad institucional y la respuesta de las autoridades tanto locales como nacionales. Se requiere no solo de voluntad política para diseñar e implementar respuestas integrales e incluyentes, sino también de recursos materiales y técnicos que permitan dar soluciones efectivas y sostenibles a este reto (Migración Colombia, 2023). Por esta razón, garantizar condiciones mínimas de bienestar resulta especialmente complejo ante un flujo tan voluminoso en un periodo tan corto y una alternativa para agilizar la asistencia es recurrir a herramientas estadísticas que permitan caracterizar con mayor precisión a los hogares más vulnerables y tomar decisiones basadas en evidencia. De este modo, es posible brindar soluciones oportunas y contribuir a una integración efectiva de la población migrante. Así, las autoridades locales estarían en mejores condiciones de atender las necesidades básicas de estas familias, sin que su calidad de vida se vea comprometida durante el proceso de adaptación.

## No obstante, el flujo migratorio continúa siendo dinámico y cambiante, lo que representa un desafío no solo para la regularización de las personas, sino también para su integración socioeconómica en Colombia. Según Migración Colombia, en 2023 la cifra de migrantes en el país se redujo en 1,10 %, independientemente de si ello obedeció al retorno a Venezuela o a que muchos utilizan a Colombia como territorio de tránsito (Migración Colombia, 2024). En el marco del trabajo técnico para la Alcaldía, la implementación de políticas de vivienda y acceso a servicios básicos se plantea como una prioridad estratégica. Un criterio clave para orientar estas políticas consiste en focalizar esfuerzos en ciudades donde los hogares migrantes presentan menores incentivos de movilidad posterior[[2]](#footnote-2), pues esto permite optimizar los recursos en poblaciones más estables y menos transitorias, aunque no necesariamente en mejores condiciones socioeconómicas. Este enfoque busca no solo atender necesidades inmediatas, sino también generar bases más sólidas para la integración de largo plazo.

## A partir de los datos de la Evaluación Conjunta de Necesidades del año 2025 (ECN2025), se observa que Cartagena de Indias es la ciudad con menor proporción de hogares que expresan intención de mudarse fuera de su territorio. Este hallazgo contrasta con municipios fronterizos como Pasto, Arauca, Riohacha y Maicao, caracterizados por la presencia de población migrante en tránsito, y con grandes urbes como Bogotá, Medellín o Pereira, que funcionan como destinos temporales donde los migrantes suelen instalarse inicialmente para generar ingresos y, posteriormente, desplazarse hacia otros lugares (Figura 1). A ello se suma la existencia de redes comunitarias que brindan apoyo en empleo y vivienda, consolidando dinámicas de arraigo que permiten aproximarse a una política pública más eficiente y de mayor impacto en términos de integración.

Gráfico

El contenido generado por IA puede ser incorrecto.

**Figura 1.** Ranking de ciudades con mayor intención de movilidad. Fuente: Evaluación Conjunta de Necesidades. Elaboración propia

## Este patrón es consistente con la literatura sobre migraciones, que muestra cómo la combinación de oportunidades laborales inmediatas y redes de apoyo locales tiende a disminuir la movilidad secundaria y a favorecer la consolidación de proyectos de vida en el lugar de asentamiento (Massey, et al., 1993). Por ello, la decisión de focalizar el análisis en Cartagena no solo responde a la evidencia estadística sobre la intención de permanencia, sino también a un marco conceptual que entiende la integración como un proceso que requiere estabilidad territorial (Figura 2). Esta perspectiva contribuye a situar la política pública en un camino más eficiente respondiendo a necesidades inmediatas de alojamiento y servicios, mientras se sientan bases más sólidas para la inclusión social de la población migrante y refugiada.

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

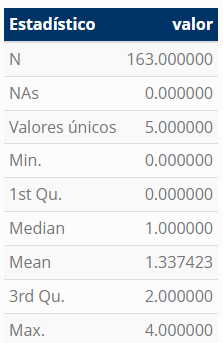
**Figura 2.** Ranking de ciudades con menor intención de movilidad. Fuente: Evaluación Conjunta de Necesidades. Elaboración propia

Finalmente, dado el objetivo de orientar políticas públicas de alto impacto y la disponibilidad de datos de la ECN2025, el presente análisis se focalizará geográficamente en la ciudad de Cartagena de Indias, selección que responde a la baja intención de movilidad posterior[[3]](#footnote-3) de los hogares migrantes, sugiriendo una mayor estabilidad residencial propicia para intervenciones a largo plazo. En este contexto, **este trabajo busca responder si las carencias en el acceso a servicios básicos de los hogares migrantes en Cartagena difieren según su nivel de vulnerabilidad.** Para ello, se utilizará como unidad de análisis los hogares de migrantes provenientes de Venezuela residentes en la ciudad de Cartagena encuestadas en la ECN2025. De esta manera, la variable dependiente en este análisis es la carencia de acceso a servicios públicos y la variable independiente es el nivel de vulnerabilidad considerando situaciones de vulnerabilidad las siguientes: (i) Jefatura del hogar femenina, ii) hogar monoparental, iii) antecedente de violencia. Los hallazgos de este análisis permitirán a la Alcaldía de Cartagena diseñar y focalizar políticas públicas más eficientes, asegurando que los recursos se dirijan a los hogares migrantes con mayores niveles de vulnerabilidad que al final representan mayores barreras para su integración a la sociedad.

## La definición de las variables utilizadas y la justificación de su selección

Considerando la pregunta de investigación planteada, en esta sección se definen y describen las variables centrales que orientan el análisis. La variable dependiente corresponde a la carencia de acceso a servicios públicos, entendida como un indicador sintético de las condiciones de vivienda. Por su parte, la variable independiente se construye como un índice de vulnerabilidad a nivel de hogar, elaborado a partir de tres dimensiones estructurales que reflejan riesgos sociales específicos. El objetivo de esta sección es caracterizar ambas variables, identificando su composición y lógica de construcción como paso previo al análisis comparativo de su relación.

**1.2.1 Variable dependiente-carencia de acceso a servicios públicos:** La variable dependiente del estudio corresponde a la carencia de acceso a servicios públicos, concebida como un indicador sintético de las condiciones básicas de habitabilidad del hogar. Su construcción integra cinco dimensiones esenciales: electricidad, gas, alcantarillado, acueducto y recolección de basuras. A cada hogar se le asigna un punto por cada servicio del que carece, por lo que la variable puede tomar valores entre 0 (sin carencias) y 5 (carencia total de servicios básicos). Este indicador permite cuantificar de manera integral la privación material relacionada con la infraestructura básica, constituyéndose en un componente clave del índice de vulnerabilidad de los hogares migrantes en Cartagena.

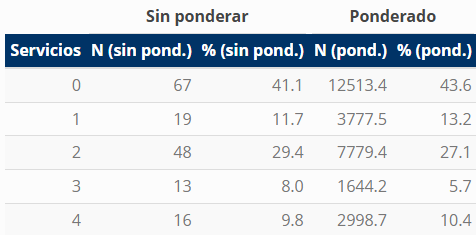


**Tabla 1.** Estadísticas descriptivas de la variable dependiente-carencia de servicios públicos.

**Notas de método.** La variable servicios mide el número de carencias en el acceso a servicios básicos (0 a 5). No se identificaron valores fuera del rango 0–5. Los valores NA se reportan explícitamente y no se imputaron.

De la tabla 1, se evidencia que la variable servicios es de tipo numérico discreto y cuenta con 163 observaciones sin valores faltantes. Representa el número de servicios disponibles en cada hogar, con un rango de 0 a 4, de manera que ningún hogar alcanza el máximo teórico de cinco. Los estadísticos descriptivos muestran un mínimo de 0, un primer cuartil de 0, una mediana de 1, una media de 1.34, un tercer cuartil de 2 y un máximo de 4, lo que evidencia que la distribución está sesgada hacia valores bajos y que la mayoría de los hogares cuentan con pocos servicios. El sesgo hacia los valores bajos sugiere una heterogeneidad limitada en el acceso, con escasos hogares en situación de privación extrema.

En términos de vulnerabilidad, este patrón puede interpretarse como una acumulación moderada de desventajas estructurales: la presencia simultánea de múltiples carencias incrementa la exposición de los hogares a condiciones de precariedad. No obstante, se identifica un resultado relevante: algunos hogares presentan mayor número de servicios básicos, pero mantienen niveles altos del índice agregado de vulnerabilidad, lo que sugiere que el acceso a infraestructura no siempre se traduce en menor vulnerabilidad multidimensional, posiblemente debido a limitaciones en ingresos, empleo o capital humano.



**Tabla 2.** Estadísticas descriptivas básicas de la variable dependiente-carencia de servicios públicos. Representación categórica de la variable

La Tabla 2 presenta la distribución de los hogares según el número de servicios con los que cuentan, distinguiendo entre proporciones sin ponderar y ponderadas con los factores de expansión. Los resultados muestran que la mayor concentración corresponde a los hogares sin acceso a ningún servicio básico, que representan el 41,1% de la muestra sin ponderar y el 43,6% al incorporar la ponderación.

Estas diferencias entre los resultados ponderados y no ponderados indican que, al ajustar por la estructura poblacional, las carencias tienden a ser más prevalentes, reflejando una mayor proporción de hogares con acceso limitado dentro del universo total de hogares migrantes. En conjunto, los datos evidencian una marcada desigualdad en la cobertura de servicios públicos y una alta concentración de hogares en condiciones de precariedad infraestructural. Este patrón sugiere que el déficit en servicios básicos es uno de los principales componentes estructurales de la vulnerabilidad, aunque, como se analizará más adelante, no siempre existe una correspondencia lineal entre el número de carencias y el nivel general de vulnerabilidad, lo que puede responder a la interacción con otros factores socioeconómicos (como empleo, ingresos o tamaño del hogar).

Imagen que contiene Logotipo

El contenido generado por IA puede ser incorrecto.

**Figura 3.** Proporción de hogares con servicios públicos faltantes en la ciudad de Cartagena utilizando el factor de expansión (color azul) y sin utilizar (color gris). Fuente: Evaluación Conjunta de Necesidades-2025. Elaboración propia.

La Figura 3 muestra la distribución porcentual de hogares según el número de carencias en servicios públicos (0 a 4), comparando dos grupos (barras azul oscuro-usando el factor de expansión oscuro y gris-sin el factor de expansión). Se observa que en ambos grupos predomina la ausencia de carencias: más del 40 % de los hogares no presenta ninguna. En segundo lugar, cerca de un 25 %–30 % tiene carencia de dos servicios públicos, mientras que los hogares con una, tres o cuatro carencias representan proporciones menores. Estos resultados sugieren que no cambian de forma considerable al usar el factor de expansión y que cerca del 60 % de los hogares tienen problemas de acceso a alguno de los servicios públicos mencionados.

**1.2.3 Variable independiente- índice de vulnerabilidad:** La variable independiente corresponde a un índice de vulnerabilidad que sintetiza condiciones estructurales que incrementan el riesgo social de los hogares y que pueden afectar su acceso a servicios públicos. Este índice se construye a partir de cuatro dimensiones: (i) jefatura femenina del hogar, (ii) condición monoparental, (iii) presencia de personas con discapacidad y (iv) antecedentes de violencia o amenazas. Cada condición suma un punto, por lo que el índice varía entre 0 y 3. De esta manera, el indicador permite clasificar a los hogares de acuerdo con su nivel de vulnerabilidad acumulada y constituye la base para analizar en qué medida estas diferencias estructurales se relacionan con las carencias en servicios básicos.

**a) Jefatura femenina del hogar:** Identifica si la persona cabeza de hogar es mujer.

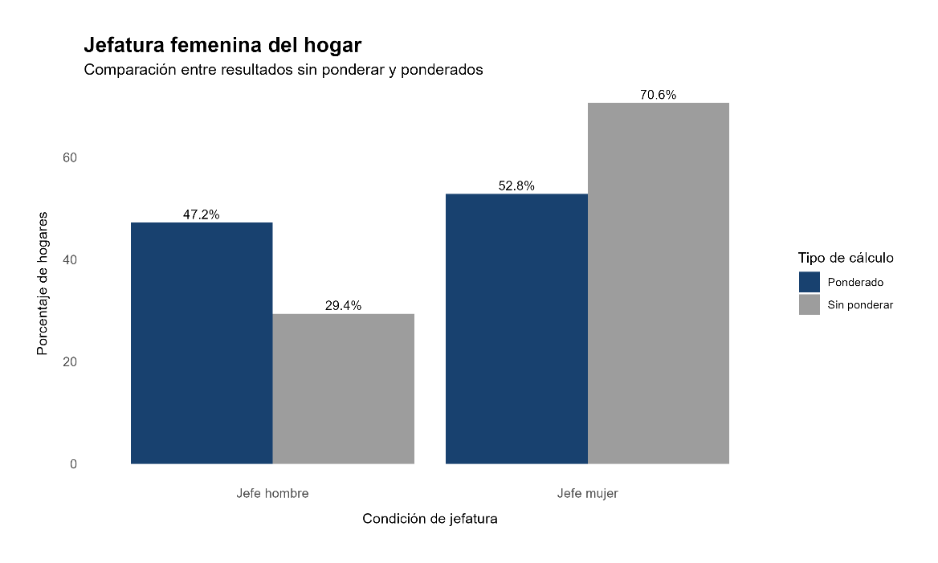
Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

**Tabla 4.** Estadísticas descriptivas básicas de la variable jefatura de hogar

**Notas de método:** La variable 'head\_female' identifica si el hogar tiene jefatura femenina (1 = mujer, 0 = hombre). No se identificaron valores perdidos, no se imputaron.

De acuerdo con la Tabla 3, la variable *jefatura femenina del hogar* es de tipo binaria, con dos categorías posibles: 0 para hogares con jefe hombre y 1 para hogares con jefa mujer. Se registran 163 observaciones válidas y sin valores faltantes. Los estadísticos descriptivos muestran un valor mínimo de 0, un máximo de 1, una media de 0.70 y una mediana de 1, lo que indica que aproximadamente el 70% de los hogares son encabezados por mujeres. La concentración de valores en 1 confirma que la jefatura femenina es predominante dentro de la muestra. Este resultado refuerza la idea de que, en la población analizada, las mujeres desempeñan un rol central como proveedoras y responsables principales del hogar, especialmente en contextos de movilidad y vulnerabilidad. La sobrerrepresentación de hogares con jefatura femenina puede estar relacionada con procesos de desplazamiento o separación familiar, que conllevan una mayor carga económica y de cuidado sobre las mujeres.



**Figura 4.** Proporción de hogares con jefatura del hogar masculina y femenina en la ciudad de Cartagena utilizando el factor de expansión (color azul) y sin utilizar este factor (color gris). Fuente: Evaluación Conjunta de Necesidades-2025. Elaboración propia.

La Figura 4 ilustra estas diferencias: al incorporar los factores de expansión (barras azules), la proporción de jefatura masculina aumenta de 29,4% a 47,2%, mientras que la jefatura femenina disminuye de 70,6% a 52,8%. Este patrón sugiere que, en el universo poblacional de hogares migrantes en Cartagena, la jefatura masculina es más común que lo que reflejaba la muestra bruta, lo cual tiene implicaciones importantes para la interpretación del índice de vulnerabilidad. La mayor representación de hogares con jefatura femenina en la muestra original podría estar vinculada a condiciones de mayor exposición socioeconómica, dado que las mujeres jefas suelen enfrentar mayores restricciones de ingreso, empleo y redes de apoyo. Por tanto, la ponderación permite ajustar esta sobrerrepresentación y obtener una estimación más realista del peso relativo de cada tipo de hogar dentro del total.

**b) Hogar monoparental:** Determina si el hogar está conformado por un padre o madre sin cónyuge y con hijos, o si incluye un “Papá/Mamá” con estado civil soltero, separado, divorciado o viudo. En relación con esta variable, se observa que no presenta valores faltantes y que sus únicos valores posibles corresponden a 0 y 1.

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

**Tabla 5.** Estadísticas descriptivas básicas de la variable hogar monoparental

**Notas de método**: La variable identifica si el hogar es monoparental bajo una definición relajada (1 = sí, 0 = no). No se identificaron valores perdidos o inconsistentes en la variable.

Imagen que contiene Logotipo

El contenido generado por IA puede ser incorrecto.

**Figura 5**. Proporción de hogares con jefatura del hogar monoparental en la ciudad de Cartagena utilizando el factor de expansión (color azul) y sin utilizar este factor (color gris). Fuente: Evaluación Conjunta de Necesidades-2025. Elaboración propia.

En un primer análisis sin ponderar, se observa que el 30.7% de los hogares corresponde a estructuras monoparentales, es decir, aquellos en los que un solo adulto asume la responsabilidad del hogar y del cuidado de los hijos, mientras que el 69.3% restante cuenta con dos o más adultos a cargo. Sin embargo, al incorporar el factor de expansión, que ajusta los datos para reflejar la estructura real de la población, la proporción de hogares monoparentales disminuye a 24.1%, en tanto que los hogares con dos o más adultos aumentan a 75.9%.

Este cambio revela que la muestra original sobreestimaba la presencia de hogares monoparentales y que, en términos poblacionales, su peso relativo es menor. No obstante, su relevancia social continúa siendo significativa, dado que concentran mayores niveles de vulnerabilidad frente a los hogares biparentales o extendidos. La existencia de un único adulto proveedor y cuidador suele implicar limitaciones en la generación de ingresos, mayor carga de trabajo doméstico y de cuidado, y mayores dificultades para acceder a servicios esenciales como educación o salud infantil. Finalmente, la ponderación no solo permite una lectura más representativa de esta variable, sino que también refuerza la necesidad de diseñar políticas públicas diferenciadas para estos hogares, orientadas a fortalecer la conciliación entre trabajo y cuidado, promover el acceso a servicios de apoyo familiar y reducir las brechas de bienestar asociadas a la estructura monoparental.

**c) Presencia de personas con discapacidad:** Esta variable identifica si al menos un miembro del hogar reporta alguna limitación o barrera física. Es de tipo entero y toma los valores 0 (sin discapacidad) y 1 (al menos un miembro con discapacidad).

Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.

**Tabla 6.** Estadísticas descriptivas básicas personas con discapacidad

**Notas de método:** La variable 'household\_has\_disability' identifica si en el hogar hay al menos una persona con discapacidad (1 = sí, 0 = no). La mayoría de los hogares no reportan personas con discapacidad, por lo que la incidencia de esta condición es baja.

Imagen que contiene Gráfico

El contenido generado por IA puede ser incorrecto.

**Figura 6.** Proporción de hogares con alguna persona con discapacidad física en la ciudad de Cartagena utilizando el factor de expansión (color azul) y sin utilizar este factor (color gris). Fuente: Evaluación Conjunta de Necesidades-2025. Elaboración propia.

En la figura 6 sin ponderar, se identificó que el 99,4% de los hogares no presenta ningún integrante con discapacidad, mientras que menos del 1% reporta al menos una persona con esta condición. Al incorporar los factores de expansión, los resultados permanecen prácticamente inalterados, confirmando que la presencia de hogares con personas con discapacidad es marginal en la población analizada. Además, no se registraron valores faltantes, lo que garantiza la completitud de la variable, aunque no su capacidad discriminante.

La escasa variabilidad de esta variable (concentrada casi por completo en una sola categoría) limita su utilidad analítica, ya que no permite distinguir diferencias significativas entre hogares. En términos metodológicos, su inclusión dentro del índice de vulnerabilidad podría introducir ruido estadístico o distorsiones en la medición del constructo, al otorgar peso a una dimensión que no aporta información diferenciadora en el contexto de la muestra, por esa razón se elimina del índice.

**d) Antecedentes de violencia o amenazas:** Registra si algún miembro del hogar ha sido víctima de violencia o amenazas.

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

**Tabla 7.** Estadísticas descriptivas básicas de la variable violencia

**Notas de método:** La variable 'household\_survivor\_violence' identifica si en el hogar hay al menos una persona sobreviviente de violencia (1 = sí, 0 = no). Se validó que la codificación sea binaria (0 y 1). Los valores NA se reportan explícitamente y no fueron imputados. No se aplican barras de error porque la variable es categórica y discreta.

La variable de violencia es una variable de tipo dicotómico, codificada como entero con valores 0 y 1. Su propósito es registrar de manera agregada si en el hogar hay al menos una persona que ha sido víctima de violencia o amenazas. El valor 1 identifica la presencia de al menos un sobreviviente de violencia o amenazas en el hogar, mientras que el valor 0 indica la ausencia de este antecedente. En una muestra de 163 hogares, la media de esta variable se ubicó en 0.0736, lo que se traduce en una **prevalencia del 7.36%**. Esto significa que la condición de violencia o amenazas a un miembro del hogar es un evento relativamente infrecuente. La distribución está fuertemente sesgada hacia el valor 0, tal como confirman la Mediana y el Tercer Cuartil, lo que subraya la baja incidencia de este antecedente en el grupo de estudio.

Imagen que contiene Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.

**Figura 7.** Proporción de hogares con antecedentes de violencia o amenazas en la ciudad de Cartagena utilizando el factor de expansión (color azul) y sin utilizar este factor (color gris). Fuente: Evaluación Conjunta de Necesidades-2025. Elaboración propia.

En el análisis sin ponderar se identifican que el 92.6% y 7.4 % de la muestra, respectivamente. Una vez aplicado el factor de expansión el porcentaje de hogares sin violencia se reduce en 8 puntos. La variable no presenta valores faltantes. Este resultado reviste especial relevancia desde una perspectiva de política pública, ya que los hogares que han experimentado situaciones de violencia suelen enfrentar mayores riesgos de vulnerabilidad social y económica, así como dificultades en el acceso a redes de apoyo institucional y comunitario.

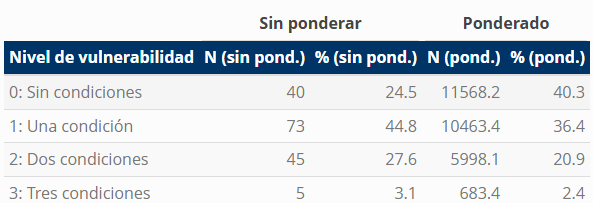
**e) Índice de vulnerabilidad:** Síntesis de condiciones estructurales que incrementan el riesgo social de los hogares y que pueden afectar su acceso a servicios públicos.

Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.

**Tabla 7.** Estadísticas descriptivas de la variable índice de vulnerabilidad

El análisis descriptivo del índice de vulnerabilidad evidencia que los valores oscilan entre 0 y 3, donde 0 corresponde a hogares no vulnerables y 3 a los más vulnerables. La media se ubica en 1,09, muy cercana a la mediana (1), lo que indica una distribución concentrada en los niveles bajos de vulnerabilidad. El primer cuartil y la mediana coinciden en 1, mientras que el tercer cuartil se sitúa en 2, mostrando que al menos el 75% de los hogares presentan niveles de vulnerabilidad de 2 o inferiores. Finalmente, estas estadísticas sugieren que la mayoría de los hogares se ubican en condiciones de baja a media vulnerabilidad, con una proporción reducida alcanzando el nivel máximo.



**Tabla 8.** Estadísticas descriptivas de la variable dependiente-carencia de servicios públicos. Descripción categórica de la variable.

La distribución del índice de vulnerabilidad muestra que, al aplicar ponderaciones, se incrementa de manera significativa la proporción de hogares clasificados como no vulnerables (categoría 0), pasando de 24,5% sin ponderar a 40,3% ponderado. En contraste, la categoría 1, que concentra la mayor proporción de hogares en la estimación simple (44,8%), se reduce a 36,4% con ponderaciones de la variable de expansión. De igual forma, la proporción de hogares con vulnerabilidad media (categoría 2) disminuye de 27,6% a 20,9%, mientras que los altamente vulnerables (categoría 3) mantienen una presencia baja en la muestra, oscilando entre 3,1% y 2,4%. Estos resultados evidencian que, al considerar los factores de expansión, el peso poblacional de los hogares no vulnerables es mayor al que sugiere la muestra sin ponderar, lo que implica una menor incidencia relativa de la vulnerabilidad en la población total.

Al examinar en detalle el índice de vulnerabilidad (Figura 8), se observa que corresponde a una variable categórica con valores que van de 0 a 3, donde cada número refleja la cantidad de condiciones de vulnerabilidad presentes en el hogar. Se trata de una variable entera que no presenta valores faltantes, lo que asegura la integridad de la información para el análisis. En términos generales, la distribución ponderada muestra que el 40,3 % de los hogares no registra ninguna vulnerabilidad, mientras que el 36,4 % presenta al menos una. Por su parte, el 20,9 % acumula dos condiciones de vulnerabilidad y un 2,4 % alcanza tres. Aunque los hogares con niveles más altos de vulnerabilidad representan un porcentaje reducido, su identificación resulta clave, ya que son los más propensos a enfrentar mayores limitaciones en el acceso a servicios básicos y a requerir intervenciones más focalizadas.

Gráfico

El contenido generado por IA puede ser incorrecto.

**Figura 8.** Proporción de hogares con un índice de vulnerabilidad entre 0 y 4, en la ciudad de Cartagena utilizando el factor de expansión (color azul) y sin utilizar este factor (color gris). Fuente: Evaluación Conjunta de Necesidades-2025. Elaboración propia

**1.2.4 Análisis agregado y relación entre variables**

Con el propósito de representar de manera integral las distintas combinaciones de carencias en servicios públicos a través de los niveles de vulnerabilidad, se elaboró un gráfico de barras apiladas que muestra la distribución de privaciones en cada una de las categorías del índice de vulnerabilidad. Esta visualización resulta especialmente útil, pues permite observar con mayor claridad las diferencias en la incidencia de carencias según las condiciones estructurales de los hogares y, de este modo, avanzar en la respuesta a la pregunta de investigación planteada.

El análisis gráfico revela que la distribución de carencias en servicios públicos varía de manera sustancial entre los distintos grupos de hogares vulnerables, lo que permite identificar perfiles de mayor riesgo. Por ejemplo, entre los hogares sin vulnerabilidad, el 36,8 % no reporta carencias y apenas un 11,2 % acumula cuatro, mientras que en aquellos con jefatura femenina o en condición monoparental la proporción de hogares sin carencias disminuye de manera importante, observándose una mayor concentración en niveles intermedios de privación. Esto sugiere que las características estructurales de los hogares inciden en la probabilidad de enfrentar limitaciones en el acceso a servicios básicos, aunque la intensidad de las carencias no sea homogénea entre todos los hogares vulnerables.

Por otro lado, los hogares víctimas de violencia muestran una distribución más dispersa: aunque el 54,1 % no presenta carencias, un 15,1 % acumula cuatro, lo que indica un mayor riesgo de privaciones extremas en este grupo. Estos resultados ponen de relieve la heterogeneidad entre los hogares vulnerables y la necesidad de que las políticas públicas no se enfoquen únicamente en la cobertura general, sino que prioricen a aquellos grupos con mayor exposición a múltiples carencias, tales como hogares monoparentales, o víctimas de violencia.

Gráfico

El contenido generado por IA puede ser incorrecto.

**Figura 9.** Distribución de carencias en servicios por condición de vulnerabilidad del hogar. Fuente: Evaluación Conjunta de Necesidades. Elaboración propia.

El Gráfico 10 presenta la distribución porcentual de los hogares según su nivel de vulnerabilidad y el número de carencias en servicios básicos. El índice de vulnerabilidad varía entre 0 (hogares sin condiciones de vulnerabilidad) y 3 (hogares con tres condiciones simultáneas). En primer lugar, se observa que la ausencia de vulnerabilidades no garantiza un acceso pleno a los servicios básicos. Entre los hogares sin vulnerabilidad (índice 0), solo el 36,8 % no presenta carencias, mientras que un 29,3 % tiene al menos una y un 11,2 % acumula cuatro. Este hallazgo evidencia que, aun en contextos de baja vulnerabilidad, persisten brechas estructurales en la provisión de servicios públicos.

Entre los hogares con una condición de vulnerabilidad (índice 1), la proporción sin carencias aumenta a 42,8 %, lo que podría indicar cierta capacidad de resiliencia. Sin embargo, también crece la participación de aquellos con cuatro carencias (13,2 %), lo que revela una mayor polarización interna: mientras algunos hogares mantienen acceso adecuado a los servicios, otros enfrentan múltiples privaciones.

En el caso de los hogares con dos condiciones de vulnerabilidad (índice 2), predomina una acumulación intermedia de carencias: el 61 % presenta al menos una y el 21,7 % acumula dos. Esto sugiere que, conforme aumenta la vulnerabilidad, la exposición a limitaciones en servicios esenciales se vuelve más generalizada. Finalmente, los hogares con tres condiciones de vulnerabilidad (índice 3) exhiben el panorama más crítico. Solo el 17,6 % no presenta carencias, mientras que el 57,5 % enfrenta al menos una y un 13,4 % acumula cuatro. Este patrón refuerza la necesidad de focalizar las intervenciones públicas en los hogares con mayor acumulación de vulnerabilidades, que son los más propensos a situaciones de exclusión en el acceso a servicios básicos.

Aunque se esperaría una relación estrictamente creciente entre el número de vulnerabilidades y las carencias, los resultados no muestran un patrón completamente monotónico. Esto podría explicarse por el tamaño reducido de los grupos con mayor vulnerabilidad, que limita la variabilidad estadística, o por la existencia de apoyos institucionales y comunitarios que mitigan parcialmente las privaciones. Por lo tanto, los resultados sugieren que la vulnerabilidad social no se traduce linealmente en privación material, sino que depende de factores contextuales y de protección social que ameritan un análisis más detallado.

Gráfico

El contenido generado por IA puede ser incorrecto.

**Figura 10.** Distribución porcentual de hogares según la vulnerabilidad y el número de carencias en servicios básicos. Fuente: Evaluación Conjunta de Necesidades. Elaboración propia.

# Parte 2: Evidencia empírica y estimación

## Diseño del contraste estadístico

La prueba estadística seleccionada fue la **Correlación de Spearman ()**, dado que es la más adecuada para medir el grado de asociación entre dos variables ordinales: el "índice de vulnerabilidad" (con 5 categorías ordenadas de 0 a 3) y el "número de carencias de servicios básicos" (5 niveles ordenados de 0 a 4). Esta prueba no paramétrica evalúa la fuerza y la dirección de una relación monotónica (es decir, si cuando una variable crece, la otra tiende a crecer o decrecer consistentemente). Específicamente, la prueba compara los *rangos* asignados a los valores de ambas variables para cada observación, determinando si las posiciones jerárquicas más altas en vulnerabilidad tienden a coincidir con las posiciones jerárquicas más altas en el número de carencias.

La expectativa teórica, basada en el hecho estilizado de que la vulnerabilidad socioeconómica incrementa la probabilidad de sufrir privaciones en términos del acceso a servicios públicos, es que existe una asociación positiva y directa entre ambas variables. Se espera que los individuos u hogares con mayores índices de vulnerabilidad presenten, en promedio, un mayor número de carencias de servicios básicos. Debido a que la revisión teórica y los hallazgos iniciales de la parte 1 sugieren una dirección clara *exante*, se plantea una prueba de hipótesis unilateral ().

**Planteamiento de la hipótesis**

* No existe asociación entre el índice de vulnerabilidad y el número de carencias.
* Existe una asociación positiva entre el índice de vulnerabilidad y el número de carencias.

**Formulación matemática de la hipótesis:**

* (No hay correlación monotónica entre la vulnerabilidad y las carencias).
* (Hay una correlación monotónica positiva. A más vulnerabilidad, más carencias).

Para responder a la pregunta de investigación se plantea una hipótesis unilateral, ya que la literatura teórica y empírica sugiere una dirección esperada en la relación entre el crecimiento económico, medido a través del ingreso per cápita, y las emisiones de CO₂ per cápita, de acuerdo con la teoría de Kuznets. La evidencia indica que las economías más desarrolladas, al alcanzar niveles más altos de ingreso, tienden también a registrar mayores emisiones por persona, especialmente cuando su matriz energética depende de los combustibles fósiles y su estructura productiva conserva sectores industriales con un alto consumo de energía.

En este contexto, una prueba bilateral no resulta adecuada, pues implicaría evaluar si las diferencias entre grupos podrían presentarse en cualquier dirección. Dado que la literatura anticipa un vínculo positivo entre el ingreso y las emisiones, el uso de un contraste unilateral se considera más consistente y metodológicamente coherente con los fundamentos teóricos y la evidencia empírica que sustentan este análisis.

Es así como para operacionalizar la pregunta, se deben clasificar los países en grupos discretos como ‘altos ingresos’ y ‘bajos ingresos’ y la medida que proponemos para determinar un punto de corte sería la mediana, considerando la distribución sesgada a la derecha de la variable de ingreso per cápita.

Dado que el interés del contraste es determinar si los países de altos ingresos presentan mayores emisiones de CO₂ per cápita que los de bajos ingresos, la prueba se formula como una prueba de una cola (superior). Por lo tanto, con base en este objetivo, se formulan las siguientes hipótesis:

Donde:

* representa la media de emisiones de CO₂ per cápita en los países de altos ingresos
* representa la media correspondiente a los países de bajos ingresos.

Siendo la hipótesis nula () la representación de un ‘no efecto’ o ‘no diferencia’ pues es bajo la suposición de que esto se mantiene a menos de que exista evidencia estadística que sea lo suficientemente significativa para rechazarla. Es así como la hipótesis nula en este caso establece que la media de las emisiones per cápita de los países con ingreso alto es menor o igual a la del grupo con ingresos bajos, determinando que la riqueza no conduce a mayores emisiones. Por su parte, la hipótesis alternativa () representa la afirmación que como investigadores buscamos respaldar con evidencia estadística, siendo en este caso la afirmación de que los países de ingreso alto tienen en efecto una media de emisiones de CO₂ per cápita significativamente mayor que la de los países de ingreso bajo.

## Revisar los supuestos y ejecutar la prueba

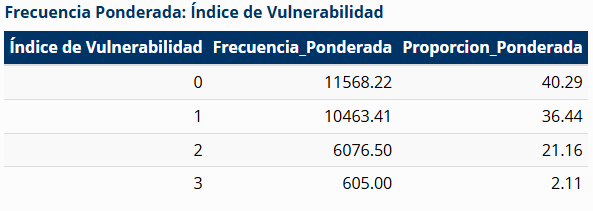
### Planteamiento teórico de los supuestos de la prueba

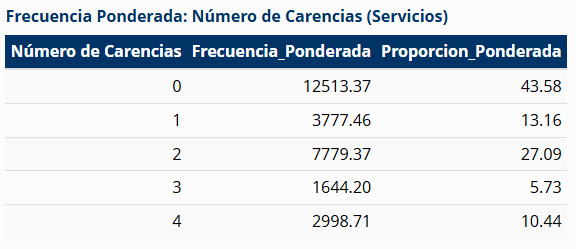
Teniendo en cuenta que en la sección 2.1 se explicaron las razones de seleccionar la prueba de Correlación de Spearman () que es un coeficiente no paramétrico diseñado para medir la fuerza y la dirección de la asociación entre variables. De esta manera, para que su aplicación sea válida, la prueba se fundamenta en varios supuestos teóricos. En primer lugar, se requiere que ambas variables de interés, el índice de vulnerabilidad () y el número de carencias (), estén medidas a un nivel ordinal o superior. Este es el supuesto más importante, ya que la prueba opera sobre los rangos u órdenes jerárquicos de las observaciones, no sobre sus valores brutos. En segundo lugar, se asume que la relación entre las variables, si existe, es monotónica. Esto implica que, a medida que los valores de aumentan, los de deben mostrar una tendencia consistente a aumentar (monotónica positiva) o a disminuir (monotónica negativa), sin la exigencia de una relación lineal estricta. El tercer supuesto implica independencia de las observaciones, lo que significa que la medición de una unidad muestral no debe influir ni estar relacionada con la medición de otra. Finalmente, es importante medir el número de empates (múltiples observaciones con el mismo valor), ya que esto puede reducir la precisión del coeficiente () estimado.

### Evaluación y argumentación de los supuestos

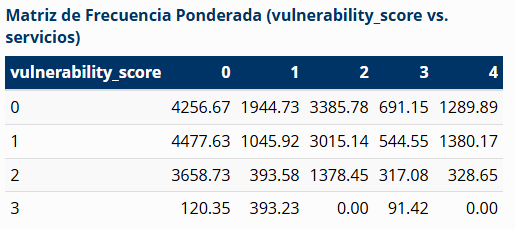
El supuesto fundamental que rige la elección de esta prueba es el del nivel de medición ordinal. Este requisito se cumple por diseño, ya que tanto el índice de vulnerabilidad como el número de carencias de servicios están constituidos como escalas categóricas discretas y ordenadas. De esta manera, la naturaleza no paramétrica de Spearman es adecuada para este tipo de variables. En relación con el supuesto de independencia de las observaciones, se considera cumplido de forma metodológica en este análisis. Si bien el origen de los datos, provenientes de una encuesta con un diseño de muestreo probabilístico, introduce una dependencia estructural entre las unidades muestrales, esta limitación se subsana mediante la incorporación de los factores de expansión al análisis.

Respecto a la evaluación de empates, la naturaleza discreta de las variables resultó en una alta concentración de observaciones en categorías específicas. La categoría más frecuente del Índice de Vulnerabilidad (X) concentró el 40.29% de las observaciones, mientras que la del Número de Carencias (Y) concentró el 43.58%. Si bien esta alta proporción podría volver el coeficiente () una estimación más conservadora que si las variables fuesen continuas, la Correlación de Spearman es robusta y su cálculo incorpora de manera estándar el ajuste de rangos promedio para manejar esta situación. En la siguiente tabla se muestran las frecuencias absolutas





Finalmente, la evaluación del supuesto de monotonicidad se realizó mediante la inspección de la tabla de contingencia ponderada y el estadístico () preliminar. El coeficiente de ( reveló una asociación monotónica negativa y débil, lo que indica una ligera tendencia a que la vulnerabilidad y las carencias se muevan en direcciones opuestas. La Matriz de Contingencia Ponderada reflejó esta debilidad y dispersión, sin mostrar una concentración dominante en la diagonal. Esta magnitud marginal es lo que el análisis de hipótesis debe validar como estadísticamente significativo o no, utilizando el error estándar ajustado por la ponderación.



El valor observado del coeficiente de correlación de Spearman ponderado fue . El signo negativo indica que la relación monotónica, aunque muy débil, se presenta en la dirección opuesta a la esperada (es decir, una tendencia a que una mayor vulnerabilidad se asocie con un menor número de carencias ponderado). El p-valor obtenido fue de 0.91481 (obtenido mediante la prueba de permutación) que es mayor que el nivel de significancia ; por lo tanto, se falla en rechazar la hipótesis nula (). Adicionalmente, el intervalo de confianza al 95% fue de . Dado que el límite inferior es negativo y el intervalo contiene valores negativos e incluye el cero (, esto confirma la ausencia de evidencia de una relación monotónica positiva en la población. En conclusión, considerando el efecto de los factores de expansión poblacional, no existe evidencia estadística significativa para afirmar que el índice de vulnerabilidad y el número de carencias de servicios presentan una relación monotónica positiva en la población migrante de Cartagena.

# Parte 3: Interpretación y comunicación de los resultados

## Interpretación

* Verificar si el resultado de la prueba y el intervalo conducen a la misma conclusión
* Interpretar el resultado en relación con la hipótesis nula y con el contexto del análisis
* Explicar qué implican los resultados en términos sustantivos: ¿qué nos dicen sobre el fenómeno estudiado?

## Comunicación

Acompañe el resultado del contraste con una o más visualizaciones que permitan interpretar la diferencia o relación observada (por ejemplo, barras con intervalos de confianza, densidades comparadas, gráficos de dispersión con tendencia, etc.) Si ya los ya creados en el EDA son oportunos, compleméntelos e integrarlos a la narrativa que permite realizar la prueba de hipótesis. La narrativa es un breve texto de apoyo que explique lo que muestra el gráfico y cómo respalda (o no) la hipótesis planteada.

# Consistencia entre visualizaciones y la respuesta a la pregunta planteada

# Precisión y pertenencia de la respuesta planteada de acuerdo con la situación del contexto

# Coherencia narrativa y argumentativa en las conclusiones

# Añadir una breve descripción de los principales ajustes realizados a partir de la retroalimentación recibida.

En respuesta a la retroalimentación recibida en el Taller 2, se implementaron los siguientes ajustes metodológicos y de formato, enfocados en robustecer la validez estadística y la presentación del análisis:

**1. Racionalización de Variables y Metodología**

* Exclusión de Variable "Discapacidad": Se verificó el análisis univariado de la variable "Hogar con Discapacidad". Dado que los resultados mostraron una fuerte predominancia de hogares sin esta condición, se determinó que no existía un aporte significativo al Índice de Vulnerabilidad. Por consiguiente, la variable fue eliminada de la construcción final del índice.
* Profundización del Análisis de Vulnerabilidad: Se amplió el desarrollo interpretativo para cada variable, estableciendo explícitamente la relación entre el incremento en el número de carencias y el consecuente aumento en la vulnerabilidad cuantificada.
* Inclusión de Notas Metodológicas: Se integraron notas de método detalladas en la presentación de cada estadística descriptiva del análisis univariado. Estas notas especifican las reglas aplicadas para el tratamiento de valores perdidos (NA) y valores atípicos (outliers), garantizando la transparencia metodológica.

**2. Estandarización de la Presentación de Resultados**

* Consistencia de Formato: Se realizaron ajustes en la mayoría de las tablas y gráficos para lograr una homogeneidad visual y estructural con las secciones iniciales del trabajo (Partes 1 a 3). (Nota: La aplicación de este ajuste a las tablas del análisis univariado está pendiente de implementación).
* Validación Estadística (Barras de Error): Se estandarizó la presentación gráfica mediante la inclusión de barras de error. Esta medida robustece la visualización de los resultados, confirmando que las diferencias observadas entre las condiciones de los hogares son diferencias reales y estadísticamente significativas (evidenciado por la ausencia de solapamiento en los intervalos de confianza).

# Referencias

Massey, D., Arango, J., Graeme, H., Kouaouci, A., Pellegrino, A., & Taylor, E. (1993).

Theories of International Migration: A Review and Appraisal. Population and Development

Review.

Migración Colombia . (2023). Colombia, el país más solidario con la migración venezolana.

Migración Colombia. (2024). Informe de migrantes venezolanos en Colombia en febrero de 2024.

1. https://github.com/Laurarivera34/Proyecto-1-EDA-estad-stica [↑](#footnote-ref-1)
2. Para efectos de este análisis y considerando la naturaleza de los datos proporcionados, se considera movilidad posterior como intención de movilidad fuera de la ciudad en los siguientes 6 meses. [↑](#footnote-ref-2)
3. Para efectos de este análisis y considerando la naturaleza de los datos proporcionados, se considera movilidad posterior como intención de movilidad fuera de la ciudad en los siguientes 6 meses. [↑](#footnote-ref-3)