**Logotipo

Descripción generada automáticamente**

**Alcances del fenómeno de la población habitante de calle desde   
una perspectiva social y de salud pública en Colombia.**

Juan Diego Villa Franco

Laura Estefanía Noriega Perlaza

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor  
Sin asignar

Universidad de Antioquia

Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2024

|  |  |
| --- | --- |
| **Cita** | (Villa Franco & Noriega Perlaza, 2024) |
| **Referencia**  **Estilo APA 7 (2020)** | Villa Franco, J. Noriega Perlaza, L. (2024). Alcances del fenómeno de la población habitante de calle desde una perspectiva social y de salud pública en Colombia. Universidad de Antioquia, Medellín, Colombia. |

**** 

Especialización en Analítica y Ciencia de Datos, CohorteV.

Centro de Investigación Ambientales y de Ingeniería (CIA).

|  |  |
| --- | --- |
|  | Diagrama  Descripción generada automáticamente con confianza media |

Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** http://bibliotecadigital.udea.edu.co

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Diego José Luis Botia Valderrama

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

**Dedicatoria**

Texto de dedicatoria centrado.

**Agradecimientos**

Texto de agradecimientos centrado.

Tabla de contenido

[Resumen 8](#_Toc167737169)

[Abstract 9](#_Toc167737170)

[1. Descripción del problema 10](#_Toc167737171)

[1.1. Problema de negocio 11](#_Toc167737172)

[1.2. Aproximación desde la analítica de datos 11](#_Toc167737173)

[1.3. Origen de los datos 11](#_Toc167737174)

[1.4. Métricas de desempeño 11](#_Toc167737175)

[2. Objetivos 13](#_Toc167737176)

[2.1. Objetivo general 13](#_Toc167737177)

[2.2. Objetivos específicos 13](#_Toc167737178)

[3. Datos 14](#_Toc167737179)

[3.1. Datos originales 14](#_Toc167737180)

[3.2. Dataset 14](#_Toc167737181)

[3.3. Analítica descriptiva 14](#_Toc167737182)

[4. Referencias 22](#_Toc167737183)

**Lista de tablas**

[**Tabla 1.** Descripción de depuración de base de datos. 16](#_Toc167737167)

[**Tabla 2**. Tabla de depuración de variables 17](#_Toc167737168)

**Lista de figuras**

[**Gráfica 1**. Distribución de la edad de los habitantes de calle 18](#_Toc167737184)

[**Gráfica 2.** Distribución de frecuencias de habitantes de calle según departamento 18](#_Toc167737185)

[**Gráfica 3.** Diagrama de cajas y bigotes de edad de inicio de consumo de sustancias psicoactivas 19](#_Toc167737186)

[**Gráfica 4.** Distribución de habitantes de calle según enfermedad diagnosticada 20](#_Toc167737187)

[**Gráfica 5.** Distribución porcentual de alfabetización en habitantes de calle 20](#_Toc167737188)

[**Gráfica 6.** Distribución de frecuencias de habitantes de calle que temen por su vida y han recibido alguna ayuda 21](#_Toc167737189)

**Siglas, acrónimos y abreviaturas**

**CSV** Comma Separated Values (Valores separados por coma)

**DANE**  [Departamento Administrativo Nacional de Estadística](https://www.bing.com/ck/a?!&&p=af5ee19e3217b1faJmltdHM9MTcxNjY4MTYwMCZpZ3VpZD0xMjlhMWNjNS05MGJjLTY2ZmItMTNhOS0wODRmOTExMzY3NDYmaW5zaWQ9NTIwNg&ptn=3&ver=2&hsh=3&fclid=129a1cc5-90bc-66fb-13a9-084f91136746&psq=dane&u=a1aHR0cHM6Ly93d3cuZGFuZS5nb3YuY28v&ntb=1)

**ETS** Enfermedad de transmisión sexual

**JSON** JavaScript object notation (Tipo de extensión de archivo)

**MB** Mega Bytes

**NaN** Not At Number (No es un número)

**SPA**  Sustancias psicoactivas

# Resumen

El proyecto busca segregar la población de habitantes de calle en Colombia con metadatos anonimizados del DANE, los cuales fueron recolectados a través de encuestas aplicadas en el año 2019. Todo esto con el fin de identificar patrones de comportamiento que faciliten la toma de decisiones sobre sus necesidades específicas. Dentro de las iteraciones para lograr este objetivo se optó por recopilar, preparar y analizar los datos de la población habitantes de calle; estas iteraciones implicaron ajustes en los métodos de análisis como imputar de valores para algunas variables, asignar etiquetas a los códigos por variable para precisar los análisis, reemplazar variables de tipo NaN por None, convertir variables cuantitativas, concatenar variables en una, verificar registros duplicados, verificar valores faltantes (nulos) y eliminación de encuestas incompletas. En cuanto a los obstáculos se pone en consideración los desafíos de las diferentes metodologías usadas al momento de realizar la preparación de los datos por el tiempo que conlleva.

*Palabras clave*: Habitante de calle, población, encuesta, datos, metodología, metadatos.

Repositorio GitHub: [**https://github.com/juandiegovillaf/seminario\_esp\_acyd\_udea/**](https://github.com/juandiegovillaf/seminario_esp_acyd_udea/)

# Abstract

The project seeks to segregate the population of street dwellers in Colombia with anonymized metadata from DANE, which were collected through surveys applied in 2019. All this in order to identify behavioral patterns that facilitate decision-making on their specific needs. Within the iterations to achieve this objective, it was chosen to collect, prepare and analyze the data of the homeless population; these iterations involved adjustments in the analysis methods such as imputation of values for some variables, assigning labels to the codes per variable to refine the analysis, replacing NaN type variables by None, converting quantitative variables, concatenating variables into one, verifying duplicate records, verifying missing values (null) and eliminating incomplete surveys. In terms of obstacles, the challenges of the different methodologies used at the time of data preparation due to the time involved are taken into consideration.

*Keywords***:** street dwellers, population, survey, data, methodology, metadata.

# Descripción del problema

En la actualidad los habitantes de calle representan una problemática tanto social como de salud pública no solo en el contexto local sino también a nivel nacional y mundial, esta situación es algo que se ha venido acentuando cada vez más con el pasar del tiempo. Este es un escenario que se viene determinando por diversos factores estructurales y sociales que influyen en el hecho de que individuos habiten la calle de forma transitoria o permanente.

Según una caracterización aplicada a este grupo poblacional realizada entre 2017 y 2021 en Colombia, “la mayoría de las personas habitantes de la calle se ubican en la región andina; son jóvenes y adultas entre los 25 y 44 años sin una pertenencia étnica definida y con limitado acceso a la educación, en general cuentan solo con educación básica primaria. En gran porcentaje son hombres, y los principales motivos para iniciar la vida en calle son: el consumo de sustancias psicoactivas (SPA) en los hombres, los conflictos y dificultades familiares en las mujeres. Aproximadamente dos (2) de cada cinco (5) habitantes de la calle llevan 10 o más años en esta situación y generalmente inician la vida en calle entre los 14 y 29 años. Se considera que cerca de la tercera parte de esta población se encuentra en un alto grado de exclusión ya que no cuentan con: redes de apoyo de personas ni de instituciones, limitado acceso a servicios de salud y sin un enfoque diferencial. La evidencia corrobora que la población habitante de la calle hace parte de un fenómeno complejo que articula las dificultades para el ejercicio de los derechos, el acceso a bienes y servicios, las desigualdades, especialmente la pobreza; la coexistencia de situaciones que afectan la salud, y dinámicas sociales que propician y mantienen las desigualdades y la exclusión, tales como la estigmatización y discriminación” (Ministerio de Salud y Protección Social, 2022)

Por lo anterior, es considerable indagar e investigar acerca de este evento y sus causas subyacentes como: la falta de vivienda asequible, problemas de salud mental, abuso de sustancias, desempleo, violencia familiar y demás factores que están relacionados con este fenómeno social y sanitario; todo esto con el fin de proporcionar información y resultados de que permitan abordar las raíces de esta problemática en materia de salud pública.

## Problema de negocio

Explorar la relación entre la situación de habitar en la calle y diversos factores como la falta de vivienda asequible, problemas de salud mental, abuso de sustancias, desempleo y violencia familiar. Estos datos fueron recopilados con encuestas del DANE (2019). “Las problemáticas sociales como la marginalidad, exclusión social y disfunción familiar modifican las dinámicas políticas, económicas y culturales de las ciudades y la vida urbana. Los habitantes de calle son un resultado de estas dinámicas, bien sea porque el sujeto decida elegir ese tipo de vida o porque la sociedad lo empuje a esa condición”. (Restrepo, 2016). El objetivo es identificar patrones y con esto clasificarlos en grupos, con el fin de diseñar propuestas específicas adaptadas a las necesidades particulares de cada grupo identificado

## Aproximación desde la analítica de datos

Se comienza por un análisis de componentes principales, que busca reducir la dimensionalidad de variables, encontrando relación entre ellas. Luego, se hará una clasificación no supervisada para encontrar grupos de individuos con comportamientos similares; y para terminar, un análisis de correspondencia múltiple, que pretende identificar características comunes entre las categorías de las variables y entre los grupos resultantes de la clasificación.

## Origen de los datos

El dataset analizado se obtiene del repositorio del DANE llamado microdatos, este tiene 13252 registros de encuestas realizadas en cinco diferentes departamentos de Colombia a personas que habitan la calle.

## Métricas de desempeño

* Análisis de componentes principales:
  + la variabilidad explicada acumulativa, lo ideal es que represente una proporción alta (>70%) con un número mínimo de componentes.
* Análisis de correspondencia:
  + la proporción de inercia explicada
* Modelo de clasificación no supervisado K-Means:
  + la inercia intra-cluster, que representa la dispersión de los datos dentro de cada grupo
  + el punto de inflexión en la gráfica de la inercia intra-cluster (método del codo)

# Objetivos

## Objetivo general

Segregar la población de los habitantes de calle de las principales ciudades de Colombia para determinar patrones de comportamiento, y así contribuir a la toma de decisiones con respecto a las necesidades que se visibilizan dentro de cada grupo.

## Objetivos específicos

* Reducir la dimensionalidad de las variables con el fin de lograr una representación más concisa y comprensible de las características distintivas de los grupos identificados.
* Identificar características comunes de las variables y los grupos resultantes de la clasificación.
* Identificar grupos de individuos con comportamientos similares, para posteriormente proponer los planes de acción según las necesidades específicas de cada grupo.

# Datos

## Datos originales

El dataset se encuentra en un repositorio del DANE llamado microdatos en el que se puede consultar diferentes temáticas como sociales, económicas y del territorio. Los datos tienen 13252 registros de encuestas realizadas como: entrevista directa y observación, pesa 2.6MB y la extensión del archivo es CSV. Los datos fueron anonimizados y codificados; para este último se cuenta con un archivo extra llamado “metadata\_habitantes\_calle.json” el cual contiene la descripción y los diccionarios de cada una de las variables. El acceso a estos microdatos es de carácter público y su estructura es estática. (DANE, 2020)

## Dataset

A partir de un análisis no supervisado se generan grupos, y a estos grupos resultantes se les aplicarán análisis supervisados de clasificación para evaluar la efectividad y veracidad del primer análisis.

## Analítica descriptiva

**Preparación de datos.**

Antes de comenzar con los análisis, se hizo una primera depuración de variables, mostrando la razón de su exclusión de cada una de ellas, en las que se toma la decisión de eliminarlas por razones no analíticas, sino de su naturaleza. Esto se muestra a continuación:

|  |  |  |
| --- | --- | --- |
| **Código** | **Variable** | **Razón de eliminación** |
| DIRECTORIO | DIRECTORIO | Identificación de la encuesta |
| TIP\_FOR | Formulario aplicado en | No aporta información |
| P2 | 2. Clase | Solo hay una categoría |
| CTL\_1 | Tipo de diligenciamiento del cuestionario | No aporta información |
| P8R | 8. ¿Cuántos años cumplidos tiene usted? | Recodificación en otra variable |
| P9 | 9. ¿Usted es hombre o mujer? | Recodificación en otra variable |
| P15R | 15. ¿De acuerdo con su CULTURA, PUEBLO, O RASGOS FISICOS, usted es o se reconoce como: | Mala codificación |
| P20 | 20. ¿Usted ha sido diagnosticado con alguna de las siguientes enfermedades: | Encabezado de sección de respuestas |
| P23 | 23. ¿Cuánto tiempo lleva usted viviendo en la calle? Años | Encabezado de sección de respuestas |
| P26 | 26. ¿Usted recibe algún tipo de ayuda: | Encabezado de sección de respuestas |
| P30 | 30. Actualmente, ¿usted consume: | Encabezado de sección de respuestas |
| P33 | 33. Su seguridad en la calle se ha visto afectada por | Encabezado de sección de respuestas |
| P33\_2 | 33.2 ¿En los últimos 30 días, usted ha sido víctima de: | Encabezado de sección de respuestas |
| P35 | 35. Sexo: | Recodificación en otra variable |
| P36R | 36. Edad estimada | Recodificación en otra variable |
| P37S1 | 37. La entrevista no se realizó porque la persona estaba: Muy alterada por el efecto de sustancias psicoactivas | Se eliminarán las encuestas incompletas, por lo que esta variable pierde relevancia |
| P37S2 | 37. La entrevista no se realizó porque la persona estaba: Dormida | Se eliminarán las encuestas incompletas, por lo que esta variable pierde relevancia |
| P37S3 | 37. La entrevista no se realizó porque la persona estaba: Con actitud agresiva | Se eliminarán las encuestas incompletas, por lo que esta variable pierde relevancia |
| P37S4 | 37. La entrevista no se realizó porque la persona estaba: Aparentemente con problemas de salud mental | Se eliminarán las encuestas incompletas, por lo que esta variable pierde relevancia |
| P37S5 | 37. La entrevista no se realizó porque la persona estaba: Totalmente desinteresada | Se eliminarán las encuestas incompletas, por lo que esta variable pierde relevancia |
| P37S6 | 37. La entrevista no se realizó porque la persona estaba: Hay condiciones de riesgo para los encuestadores | Se eliminarán las encuestas incompletas, por lo que esta variable pierde relevancia |
| P37S7 | 37. La entrevista no se realizó porque la persona estaba: Otra | Se eliminarán las encuestas incompletas, por lo que esta variable pierde relevancia |
| COMPLETA | Identificador de finalización de la encuesta | Luego de filtrar las encuestas completas, se elimina por falta de información |

**Tabla 1.** Descripción de depuración de base de datos.

Luego de que se hace esta primera depuración de variables, se ejecuta una segunda eliminación, en la que se toma la decisión de descartar a cada una de ellas por razones analíticas y de creación de nuevas variables. Se puede observar en la siguiente tabla cuáles fueron y su razón de eliminación:

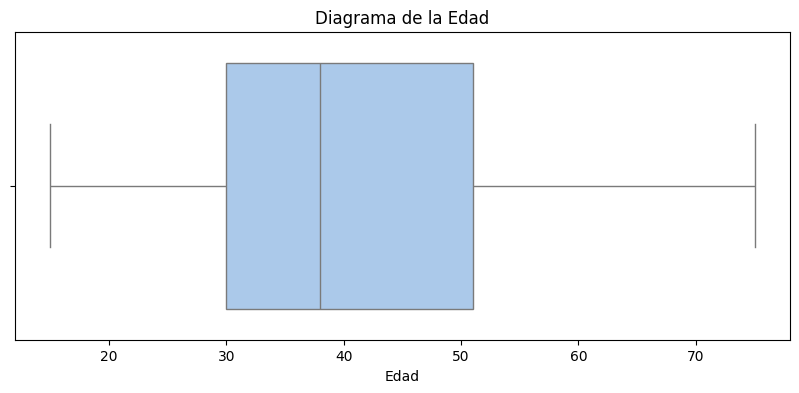
|  |  |  |
| --- | --- | --- |
| **Código** | **Variable** | **Razón de eliminación** |
| P12 | 12. ¿En qué municipio duerme usted habitualmente? | La concentración de la respuesta está sobre una categoría |
| P17S4 | 17.1 ¿Cuál(es): Intento de suicidio? | la respuesta se centra en una categoría específica |
| P17S7 | 17.1 ¿Cuál(es): Enfermedades de transmisión sexual (venéreas)? | La respuesta se centra en una categoría específica |
| P19 | 19. ¿Lo atendieron? | Es variable dependiente de la P18 y la frecuencia de ella es baja |
| P20S3 | 20. ¿Usted ha sido diagnosticado con alguna de las siguientes enfermedades: 3. Cáncer? | La respuesta se centra en una categoría específica |
| P20S5 | 20. ¿Usted ha sido diagnosticado con alguna de las siguientes enfermedades: 5. VIH- SIDA | La respuesta se centra en una categoría específica |
| ^P32S | Categorías de por qué no utiliza los programas donde se atiende a los habitantes de la calle | Por frecuencias bajas en las personas que sí conoce estos programas. |

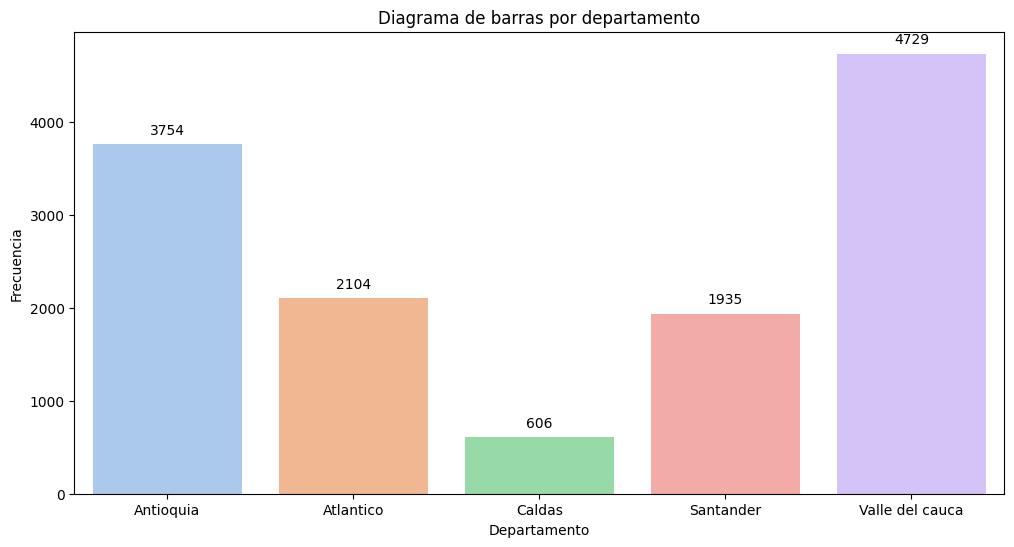
**Tabla 2**. Tabla de depuración de variables

Al terminar esta depuración, se resuelve un dataset con 95 variables y con 13128 registros con el que se harán los posteriores análisis

**Análisis descriptivo**

Se tuvo una distribución de la edad que va desde los 15 hasta los 75 años, con una mediana alrededor de los 38 años.

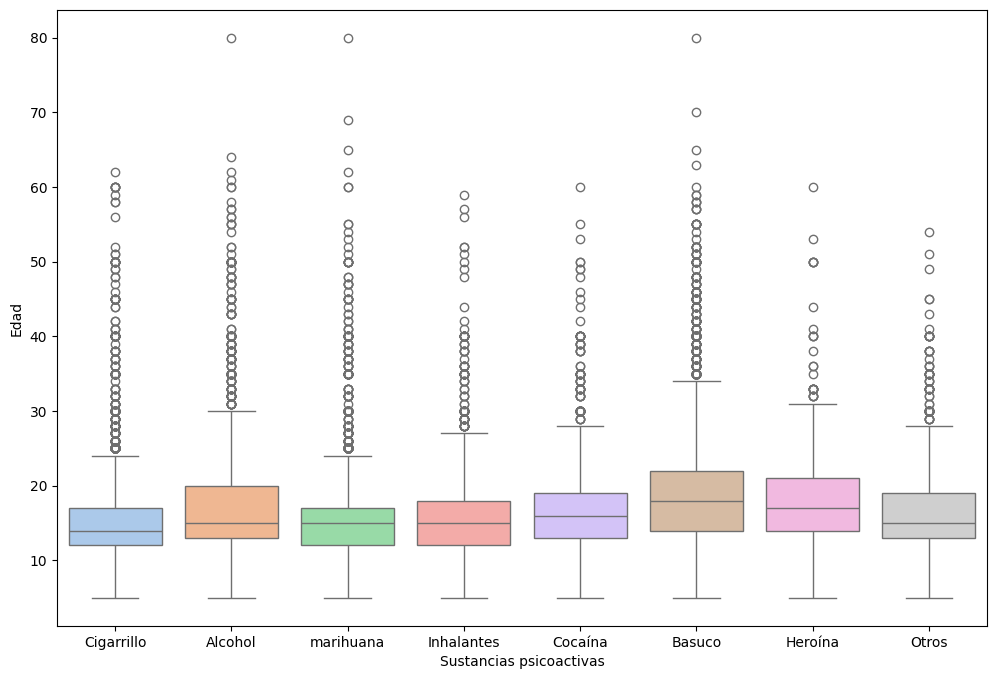
**Gráfica 1**. Distribución de la edad de los habitantes de calle

Las encuestas se llevaron a cabo en 5 departamentos distribuidos de la siguiente manera:

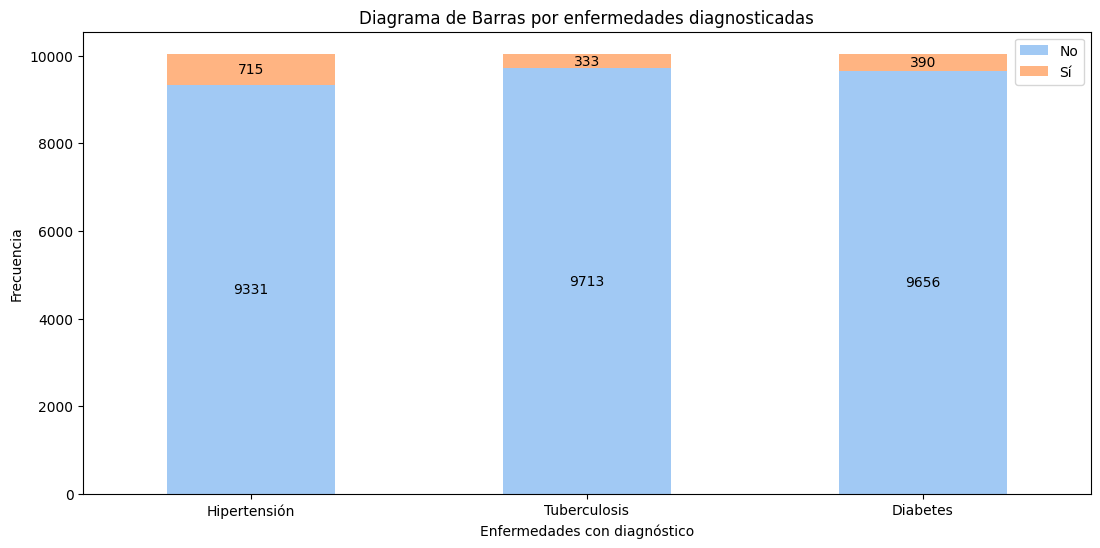
**Gráfica 2.** Distribución de frecuencias de habitantes de calle según departamento

En donde los departamentos con mayor encuesta fueron el Valle del Cauca y Antioquia.

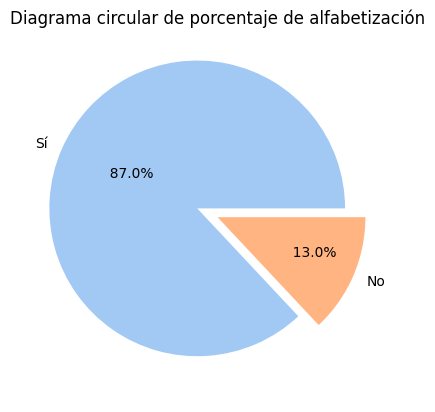
Se quiso indagar por diferentes variables para entender la población encuestada, y es por eso que se busca visualizar el consumo de sustancias psicoactivas y su edad en que lo hizo por primera vez. En lo que se observa, cada una de las sustancias tiene la concentración en una edad temprana por debajo de los 20 años alrededor de todas aquellas preguntadas en la encuesta, esto puede corroborar que las consecuencias que puede traer consumir sustancias psicoactivas pueden desembocar en enfermedades y riesgos para jóvenes, tales como la habitabilidad en la calle o una disminución de la calidad de vida, en especial de su salud. (Vega, 2021)



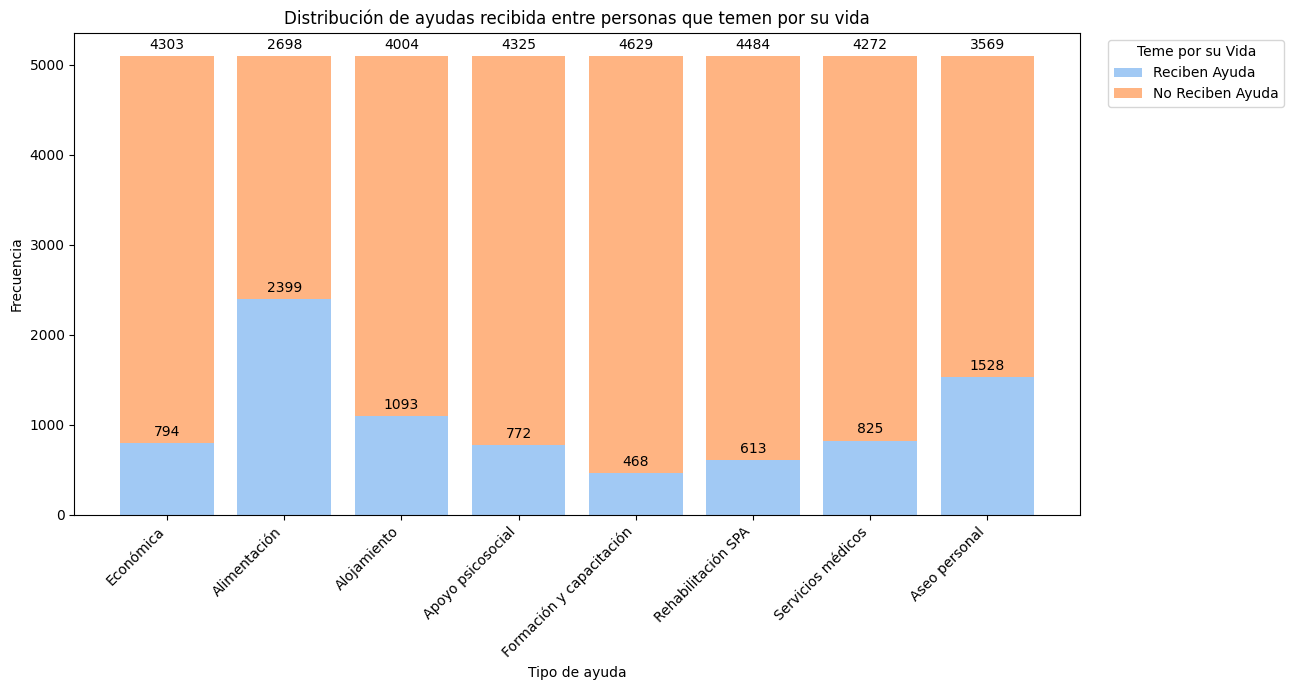
**Gráfica 3.** Diagrama de cajas y bigotes de edad de inicio de consumo de sustancias psicoactivas

Cuando se refiere enfermedades, algunas de ellas fueron eliminadas, como enfermedades de transmisión sexual (ETS), y por la poca participación de respuestas en ellas. Sin embargo, para las enfermedades como: la hipertensión, la tuberculosis y la diabetes, muestra que pueden ser patologías comunes entre los individuos. No obstante, esto puede estar sujeto a investigación, de que no haya disposición de parte de los encuestados en ver su salud como una prioridad.

**Gráfica 4.** Distribución de habitantes de calle según enfermedad diagnosticada

A los individuos se les preguntó sí sabían leer, a lo que el 87% de ellos, contestaron que sí sabían, frente a un 13% que no.

**Gráfica 5.** Distribución porcentual de alfabetización en habitantes de calle

Para terminar, se analizó en las personas que mencionaron temer por su vida, y sí estos han recibido algún tipo de apoyo como se enlista en el siguiente diagrama, en el que se evidencia que la ayuda más frecuente que pueden recibir estos individuos es alimenticia, seguido de aseo personal frente a los demás tipos de ayudas.

**Gráfica 6.** Distribución de frecuencias de habitantes de calle que temen por su vida y   
han recibido alguna ayuda

# Referencias

Ministerio de Salud y Protección Social. (2022). Caracterización de personas habitantes de la calle en Colombia entre 2017 y 2021: Actualización de línea de base de la política pública social para habitantes de la calle.

Restrepo, A. (2016). *El ser humano al límite: una mirada reflexiva al habitante de calle. Drugs and Addictive Behavior, 1(1), 89-100.*

Vega, N. L. (2021). Jóvenes en riesgo de habitar la calle y sus redes de apoyo: la experiencia desde IDIPRON UPI Perdomo. *Universidad Nacional de Colombia*.