

Percepción de Seguridad en Colombia: Un Análisis Exploratorio con Técnicas de Aprendizaje No Supervisado

Grupo 4: Daniela Castillo Tellez, Angie Lorena Hidalgo Duran, Katherine Russi Parra, Andrés Santiago Serna Tangarife

Resumen.

En Colombia, la seguridad ciudadana se ha convertido en una preocupación primordial del gobierno debido al aumento de la violencia y la delincuencia, especialmente en áreas urbanas, desde la década de 1980. Esta problemática ha dejado profundas huellas en la sociedad colombiana, lo que ha llevado al Departamento Administrativo Nacional de Estadística (DANE) a medir y analizar la seguridad en el país en diversos aspectos que abarcan la economía, la sociedad y el territorio.

La "Encuesta de Convivencia y Seguridad Ciudadana" (Dane,2021), arroja datos reveladores, donde indica que el 52.9% de las personas mayores de 15 años mencionan sentirse inseguros(as), los principales generadores de inseguridad se encuentran relacionados al hurto, las riñas y la extorsión.

En este contexto, este proyecto tiene como objetivo identificar patrones en grupos de personas que puedan estar relacionados con una percepción negativa de la seguridad en Colombia. Utilizaremos técnicas de modelado no supervisado, como el Análisis de Componentes Principales (PCA) y el clustering, para comprender los factores que influyen en la percepción de seguridad.

El propósito principal de este estudio es aportar al entendimiento de los factores clave que influyen en la percepción de seguridad entre los colombianos. Esto no solo podría servir como referencia para la formulación de políticas y estrategias de seguridad ciudadana por parte del Ministerio de Defensa del gobierno colombiano, sino también como punto de partida para el desarrollo de medidas más efectivas que aborden las preocupaciones de seguridad de la población.

1. Introducción.

La seguridad ciudadana en Colombia es un tema de gran relevancia que afecta a la vida de millones de colombianos a diario. La pregunta que nos guía en esta investigación es ¿Cuáles son los factores que influyen en la percepción de seguridad de los colombianos? Este interrogante surge en un contexto enmarcado en la preocupación por la seguridad ciudadana en Colombia. El cliente potencial interesado en abordar este problema pertenece al sector público, específicamente el Ministerio de Defensa y el Gabinete de Seguridad del país, que incluye figuras como el General de las Fuerzas Militares, los Comandantes de Fuerza, el Director General de la Policía Nacional y el Viceministro, donde la necesidad de entender mejor los factores que afectan la percepción de seguridad tiene implicaciones significativas en la formulación de políticas, generar estrategias para reducir la delincuencia y mejorar la sensación de seguridad entre los colombianos.

La motivación para resolver este problema es la creación de un entorno más seguro y confiable para la población colombiana. Para abordar esta cuestión, se propone la utilización de modelos de aprendizaje no supervisado, específicamente técnicas de agrupación (clustering) y la reducción de dimensionalidad (PCA). Estas técnicas se aplicarán a la Encuesta de Convivencia y Seguridad Ciudadana - ECSC - 2021. Esta encuesta contiene información relevante sobre la percepción de seguridad y convivencia de los colombianos. El propósito es identificar patrones de datos que podrían pasar desapercibidos mediante enfoques tradicionales además de comprender mejor el comportamiento de la delincuencia en Colombia.

El aprendizaje no supervisado tiene el potencial de agrupar personas según su confianza en la policía, satisfacción con el sistema de justicia penal y miedo a ser víctima de un delito. La aplicación de estas técnicas permitirá a las autoridades comprender mejor los factores subyacentes que afectan la percepción de seguridad en Colombia, y así desarrollar estrategias más efectivas para mejorarla.

2. Revisión preliminar de la literatura.

La exploración de la literatura en torno a problemas similares proporciona valiosos marcos de referencia para abordar la percepción de seguridad en Colombia. Se observan enfoques similares y divergentes, lo que contribuye a contextualizar y diferenciar el enfoque del presente estudio.

En el estudio realizado por Pacini, B. (2019) utiliza la base de datos de victimización 2017 de Argentina, abordando técnicas de PCA y LASSO para buscar caracterizar la percepción de la seguridad de los ciudadanos. La base de datos describe la percepción de los ciudadanos frente al accionar de las autoridades y la seguridad en general, recoge además información de características socioeconómicas del hogar, perfil del encuestado, medidas de seguridad adoptadas y sensaciones respecto a la seguridad. Esta encuesta es muy similar a la del DANE utilizada en el presente trabajo. A diferencia de nuestro enfoque, se utilizaron los datos para predecir denuncias y no tanto para entender patrones o relaciones relevantes frente a la percepción de la seguridad.

Tejeda Parra, G., & Palafox Moyers, C. G. (2021), enfocan su estudio en Sonora México donde se aborda la percepción de seguridad ciudadana utilizando un enfoque supervisado mediante el algoritmo logit dicotómico. Aunque difiere en la metodología, este estudio también explora variables de entorno, satisfacción y opinión relacionadas con la percepción de seguridad.

En un enfoque multinivel, Murrià Sangenís, M. (2021) utilizaron modelos de regresión logística multinivel, con el fin de interpretar la percepción de la seguridad y sus variables relacionadas en más de un nivel de agregación (individual, barrio, distrito), cuentan con datos detallados de la seguridad en los barrios, con los cuales entrenan varios modelos donde en cada uno van incluyendo grupos de variables relacionadas con varios aspectos de la percepción de la seguridad, con el fin de analizar cómo se ven afectadas las probabilidades de experimentar poca seguridad en los barrios.

Moreno Vera, F. A. (2022) en su tesis de maestría, aborda la percepción de la seguridad urbana en base a la percepción visual que puedan tener las personas sobre la apariencia de lugares de la ciudad, utilizando la base de datos de imágenes Place Pulse 2.0. Para esto utilizaron técnicas supervisadas y semi-supervisadas con aprendizaje profundo para el análisis de imágenes.

De acuerdo a Bermúdez García, A. (2023) cuantifica la percepción de seguridad por medio de clasificadores de aprendizaje supervisado, analizando textos de twitter para cuantificar el nivel de percepción de seguridad asociada a estos textos. Utiliza un enfoque de análisis de sentimientos con NLP y técnicas de interpretación de modelos como LIME y Shapley, buscando identificar cómo el uso de algunas palabras contribuyen a la percepción de seguridad.

Velandia, R., Guataquira, A. y Prada, J. (2018). Utiliza datos de una encuesta realizada a ciudadanos de Bogotá con algunas preguntas relacionadas con percepción de seguridad y delincuencia, con el fin de aplicar técnicas de clustering con el algoritmo k-means, buscando obtener patrones relacionados con la criminalidad. Además, se evalúan factores que a simple vista no son perceptibles para identificar las causas de este tipo de actos delictivos. Dicha encuesta, fue realizada a un grupo

pequeño de personas (49) y con algunas preguntas muy básicas que no logran cubrir los mismos aspectos que cubre la encuesta del DANE.

Finalmente, un informe en Antioquia Centro de Análisis Político, Universidad EAFIT. (2014) presenta los resultados de una encuesta en Antioquia en 2013 sobre seguridad, convivencia y victimización, con el objetivo de mejorar sistemas de seguridad. Analiza literatura relevante y contextos. Se encuestaron 2,408 personas en 9 subregiones, evaluando delitos, percepciones, confianza y características. Se resumen hallazgos y proponen recomendaciones. Se basó en encuestas anteriores y preguntas de fuentes como DANE.

3. Descripción de los datos.

Los datos utilizados en este estudio se recopilaban a partir de la Encuesta de Convivencia y Seguridad Ciudadana - ECSC - 2021 Colombia. (2022) elaborada por el DANE. El objetivo de esta encuesta es recopilar indicadores que proporcionen información sobre la población que ha sido afectada por actividades delictivas como hurto, riñas, peleas o extorsión. Además, la encuesta indaga sobre la percepción de seguridad que tienen los ciudadanos. Esta encuesta se lleva a cabo anualmente y los datos están disponibles públicamente sin restricciones de acceso.

Análisis descriptivo de los datos:

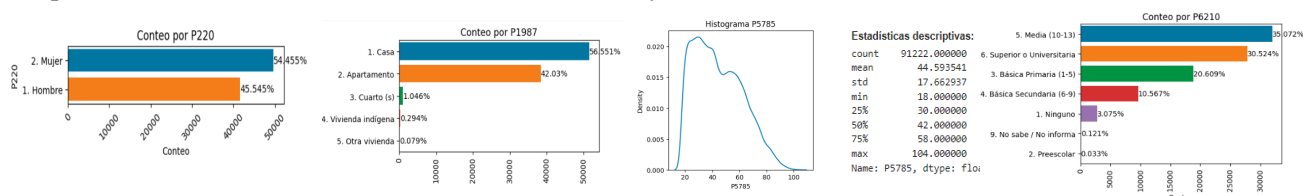
Grupo	Votacion entre opciones	Puntaje	Percepcion	Si No	total variables
variables generales			x	x	9
limitaciones fisicas	x		x		9
propietario de vehículo	x			x	4
ha sufrido agresión				x	3
ha sufrido delito				x	10
ha sufrido delito general				x	12
problemas en barrio	x			x	21
presencia grupos en barrio	x			x	5
se siente inseguro en barrio porque	x			x	16
se siente inseguro en ciudad porque	x			x	15
ha experimentado situación incómoda				x	8
ha experimentado situación inseguridad				x	7
ha tomado medidas	x				14
sentimientos con respecto a seguridad			x		7
seguridad en lugares			x		8
cree será víctima	x				16
justificaría violencia				x	6
percepción aumento o disminución situaciones			x		12
confianza instituciones		x	x		12
qué mejoraría	x				12

La encuesta cuenta con 122.251 registros y un total de 1384 variables, de estas seleccionamos 206 variables que consideramos relevantes para responder la pregunta, las cuales agrupamos por factores según la orientación de las preguntas. Dichos factores se clasificaron para entender mejor el tipo de respuestas por pregunta: **Votación entre opciones** son preguntas con opciones a,b,c,h de las cuales 1 es la selecciona y missing lo contrario. **Puntaje** son calificaciones de 1 a 5 donde 5 es excelente las cuales se trataran como variables ordinales, y **Binarias** de respuesta Sí/No.

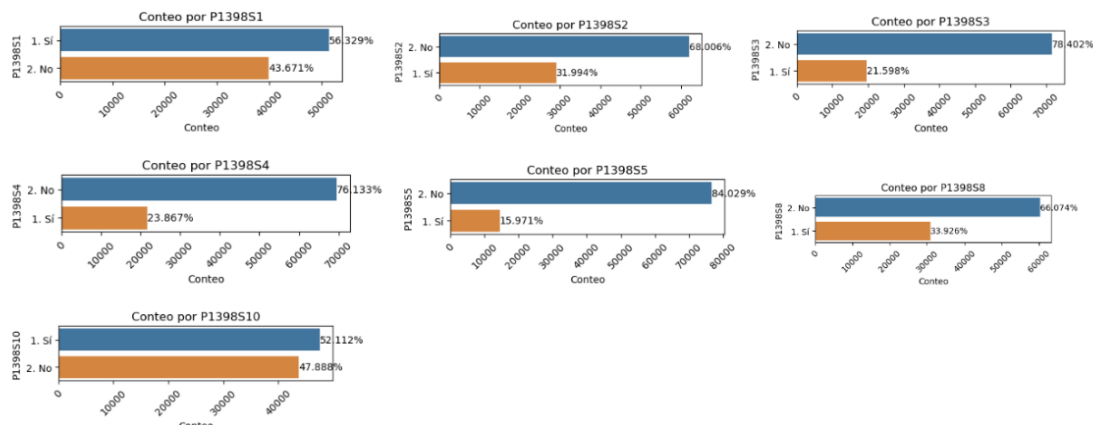
- Diccionario de datos: [link](#)
- Análisis exploratorio: [link notebook](#)
- Repositorio en Github [link_README](#)

Análisis exploratorio por agrupación de variables:

Caracterización de la población: Las siguientes variables son preguntas asociadas a la caracterización de la población, ahora bien de acuerdo a los valores de cada variable, se definieron cuales son numéricas, discretas u ordinales, y se asignaron los valores correspondientes a las variables discretas. En general la base de datos cuenta con 54% de registros en género mujeres, 56% viven en casa, 46% viven en apartamento, frente a las edades están acumuladas entre los 20 y 40 años, 35% de la población encuestada cuenta con educación media y 30% Universitaria.



El 56% de los encuestados concuerdan que el delito que se presenta en el barrio es hurto a personas, 32% hurto residencias, 22% hurto comercio, 23% hurto vehículos, 16% homicidios, riñas 33%, consumo drogas 52%.

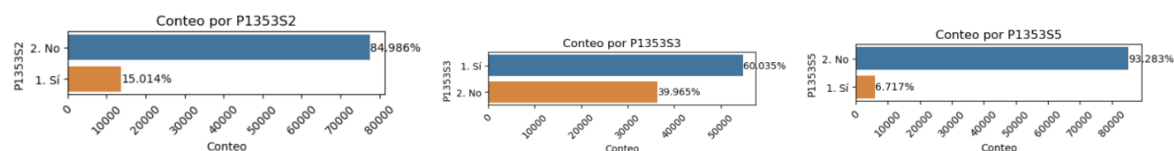


La presencia de grupos y/o pandillas no es algo común para el 73% de la población, lo mismo se habla de guerrillas, expendedores de drogas, barras bravas o grupos armados.

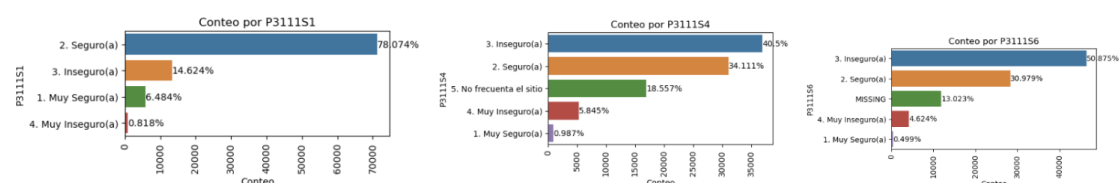
Análisis por grupos de variables: Se identificaron los siguientes grupos de variables de tipo binario, debido a que dentro de este grupo de variables se presentan altas concentraciones en respuesta "No" se genera una dispersión en los datos y por ende poca variación, se decide cambiar el sentido de las variables y generar una sola variable para cada grupo, donde se realiza un conteo para las respuestas "Si". Así mismo se asume que los nulos, respuesta "2" que corresponde a "NO" y respuesta "9" que corresponde a dato faltante o no respondido, se asignan con valor de 0 para los respectivos conteos.

Dentro de los grupos identificados se describen varias categorías de variables relacionadas con diferentes aspectos de seguridad y delincuencia, incluyendo limitaciones físicas, propiedad de vehículos, experiencias de agresión y delitos, problemas en el vecindario, presencia de grupos delictivos, sentimientos de seguridad, medidas tomadas, creencia de ser víctima, justificación de la violencia, percepción de cambios en situaciones de inseguridad y acciones de mejora. Las variables en cada categoría tienen diferentes valores y se tratan como binarias en su mayoría, aunque hay algunas excepciones. Estas variables se utilizan para recopilar información sobre la seguridad y la delincuencia en una población.

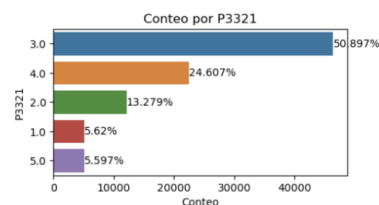
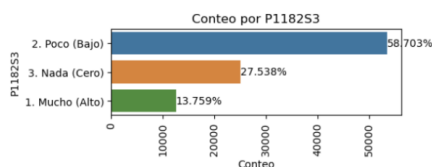
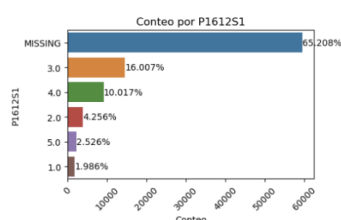
En preguntas que hacen referencia a justificar la violencia, 85% no utilizaría violencia para castigar a un delincuente, 60% justificaría la violencia si es en defensa propia, 93% no justifican en caso de defender a alguien desconocido de una agresión.



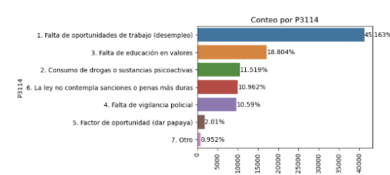
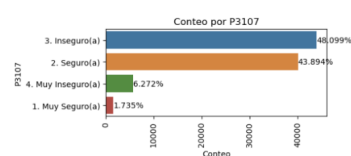
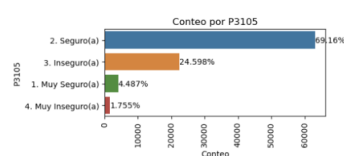
El 78% de encuestados se siente seguro en el lugar donde realiza su actividad principal, 40% se siente inseguro en transporte público, 51% se siente inseguro en vía pública.



Variables "confianza instituciones": Las variables P1612S1 a P1182S3 preguntan sobre la contribución de instituciones como la Policía, las Fuerzas Militares, la Alcaldía y los jueces a la seguridad, con valores de 1 (Mucho), 2 (Poco) y 3 (Nada), tratándose todas como ordinales. La Policía tiene un puntaje de 3 en un 16% de casos, con muchos datos faltantes, y alrededor del 60% considera que estas instituciones contribuyen "poco" a la seguridad.



El 69% de los encuestados se sienten seguros en su barrio, 48% se siente inseguro en la ciudad y considera que dicho nivel sigue igual desde hace 6 meses, 45% considera que la delincuencia se debe al desempleo.



4. Propuesta metodológica.

La pregunta de investigación se centra en identificar los factores que influyen en la percepción de seguridad de los colombianos. Dado que la percepción de seguridad es un concepto multidimensional y complejo, el uso de técnicas analíticas es esencial para abordar esta cuestión de manera comprensiva y rigurosa. En este contexto, se justifica la elección de PCA (Análisis de Componentes Principales) y clustering como enfoques adecuados para llevar a cabo el análisis.

PCA permite reducir la dimensionalidad de los datos mientras conserva la mayor parte de la variabilidad original. En el caso de la percepción de seguridad, es probable que haya múltiples indicadores o variables que contribuyan a la percepción general. PCA puede ayudar a identificar patrones y relaciones subyacentes entre estas variables, resumiéndose en componentes principales que explican la mayor parte de la varianza. Esto facilita la identificación de los factores latentes que impactan en la percepción de seguridad.

Dado que estamos interesados en identificar patrones de percepción de seguridad, usar clustering ayuda a agrupar individuos que comparten características similares. Con dicha agrupación podríamos identificar subgrupos de la población con percepciones similares. Esto puede revelar diferencias en la percepción de seguridad basadas en ciertas características demográficas, geográficas o socioeconómicas. Además, el clustering puede ayudar a segmentar la población en función de la intensidad de su percepción de seguridad y revelar posibles subconjuntos de preocupaciones de seguridad.

La hipótesis inicial es que la percepción de seguridad entre diferentes grupos de personas puede estar relacionada con la edad, género, nivel educativo, o ubicación geográfica. Estas diferencias son útiles para comprender las desigualdades en la seguridad ciudadana además de conocer diferencias en los factores que pueden ser más importantes para los ciudadanos a la hora de evaluar su seguridad, factores como la delincuencia, confianza en la policía, percepción de la eficacia de las medidas de seguridad.

5. Bibliografía

- Bermúdez García, A. (2023). Interpretabilidad categórica de clasificadores automáticos sobre contenido relacionado a la percepción de la seguridad. Universidad Nacional de Colombia. Recuperado de <https://repositorio.unal.edu.co/handle/unal/84030>
- Centro de Análisis Político, Universidad EAFIT. (2014). Encuesta de Percepción de Seguridad, Convivencia y Niveles de Victimización en Antioquia: Informe final. Colombia. Recuperado de https://www.eafit.edu.co/centros/analisis-politico/documents/informe%20final%20epscv%20_formato.pdf
- Departamento Administrativo Nacional de Estadística (DANE). (2021). Boletín Técnico: Encuesta de Convivencia y Seguridad Ciudadana 2021. Recuperado de: https://www.dane.gov.co/files/investigaciones/poblacion/convivencia/2021/Bol_ECSC_2021.pdf
- Encuesta de Convivencia y Seguridad Ciudadana - ECSC - 2021 Colombia. (2022). Data Dictionary. Departamento Administrativo Nacional de Estadística (DANE). Recuperado de <https://microdatos.dane.gov.co/index.php/catalog/790/data-dictionary>
- Informe de Convivencia y Seguridad Ciudadana del DANE: Departamento Administrativo Nacional de Estadística (DANE). (2021). Encuesta de Convivencia y Seguridad Ciudadana. Recuperado de: https://www.dane.gov.co/files/investigaciones/poblacion/convivencia/2021/Bol_ECSC_2021.pdf
- Microdatos del DANE: Departamento Administrativo Nacional de Estadística (DANE). (2021). DANE Microdatos. Recuperado de: <https://microdatos.dane.gov.co/index.php/catalog/790/get-microdata>
- Moreno Vera, F. A. (2022). Técnicas de Aprendizaje Profundo para el Análisis de la Percepción de la Seguridad Urbana (Tesis de maestría). Universidad Católica San Pablo, Departamento de Ciencia de la Computación, Arequipa, Perú. Recuperado de <https://repositorio.ucsp.edu.pe/bitstreams/4cb5ba6a-c710-44b1-b437-69542998101c/download>
- Murrià Sangenís, M. (2021). El estudio sobre la percepción de inseguridad en los barrios de Barcelona. Minipapers, Instituto de Estudios Regionales y Metropolitanos de Barcelona (IERMB), Universidad de Barcelona, España. Recuperado de

<https://postc.umh.es/minipapers/el-estudio-sobre-la-percepcion-de-inseguridad-en-los-barrios-de-barcelona/>

- Pacini, B. (2019). Reducción de dimensionalidad y regularización en Encuesta Nacional de Victimización (Tesis de maestría). Universidad de San Andrés, Departamento de Economía, Maestría en Economía, Buenos Aires. Recuperado de:
<https://repositorio.udesar.edu.ar/jspui/bitstream/10908/17697/1/%5BP%5D%5BW%5D%20M.%20Eco.%20Pacini%2C%20Bianca.pdf>
- Tejeda Parra, G., & Palafox Moyers, C. G. (2021). Percepción sobre la seguridad ciudadana en Sonora. Revista Región y Sociedad, Universidad de Sonora, Departamento de Economía, México. Recuperado de:
<https://regionysociedad.colson.edu.mx:8086/index.php/rys/article/view/1438>
- Velandia, R., Guataquira, A. y Prada, J. (2018). Prototipo de un modelo clasificador para la toma de decisiones a partir de la percepción actual de la seguridad en Bogotá. TIA, 6(2), pp. 18-24. Universidad Distrital Francisco Jose de Caldas. Colombia. Recuperado de
<https://revistas.udistrital.edu.co/index.php/tia/article/download/9744/14689/72561>