

ANÁLISIS ESTADÍSTICO DEL HOMICIDIO EN COLOMBIA: UNA VISTA ENFOCADA A LOS PERFILES SOCIODEMOGRÁFICOS, EDUCATIVOS Y ECONÓMICOS

Cabarcas Mondol, Ricardo¹, Castañeda Correa, Mónica¹ & Van strahlen Martínez Esmeralda¹

¹Fundación Universitaria Los Libertadores

Aceptado: Junio 24, 2023

Resumen

Contexto. Esta investigación se centra en el análisis de los datos estadísticos de homicidios en Colombia entre los años 2016 y 2020. **Propósito.** El objetivo principal es examinar los perfiles sociodemográficos, educativos y económicos de la población con el fin de identificar patrones y tendencias que expliquen el actuar violento que desencadena los homicidios. **Metodología.** Esta investigación abordó la elevada tasa de homicidios en Colombia, mediante el análisis de la tasa por departamento y la exploración de variables relevantes. Se extrajo una metodología de minería de datos, incluyendo métodos estadísticos y técnicas de aprendizaje automático, para analizar los datos recopilados (11 variables, 1981 observaciones) empleando metodología CRISP DM. La información para el análisis se obtuvo de las páginas de la Policía Nacional, el DANE, la CEPAL, DNP (Departamento Nacional de Planeación) y HUMANITARIAN DATE EXCHANGE. **Resultados.** En los análisis realizados no se puede identificar de manera representativa una variable de las analizadas que aumente la tasa de homicidios, se debe realizar un análisis que agrupe más variables. **Conclusiones.** Este estudio sugiere que la educación, la pobreza y el desempleo pueden tener un papel importante en la determinación de la tasa de homicidios en la región estudiada. Sin embargo, se necesita un análisis más profundo y considerar otros factores para comprender completamente las dinámicas y causas subyacentes de los homicidios en la región. Estas evidencias determinaron una base inicial para futuras investigaciones que aborden el problema de la violencia y mejoren las condiciones socioeconómicas en la región.

Palabras clave— Desempleo, educación, homicidios, machine learning, pobreza, series de tiempo, violencia

Abstract

Context. This research focuses on the analysis of statistical data on homicides in Colombia between the years 2016 and 2020. **Purpose.** The main objective is to examine the sociodemographic, educational and economic profiles of the population in order to identify patterns and trends that explain the violent acts that trigger homicides. **Methodology.** This research addressed the high homicide rate in Colombia, by analyzing the rate by department and exploring relevant variables. A data mining methodology, including statistical methods and machine learning techniques, was extracted to analyze the collected data (11 variables, 1981 observations) using CRISP DM methodology. The information for the analysis was obtained from the pages of the National Police, DANE, ECLAC, and HUMANITARIAN DATE EXCHANGE. **Results.** In the analyzes carried out, it is not possible to identify in a representative way a variable of those analyzed that increases the homicide rate, an analysis must be carried out that groups more variables. **Conclusions.** This study suggests that education, poverty, and unemployment may play an important role in determining the homicide rate in the region studied. However, further analysis and consideration of other factors is needed to fully understand the dynamics and underlying causes of homicides in the region. These evidences determined an initial basis for future research that addresses the problem of violence and improves socioeconomic conditions in the region.

Keywords— Unemployment, education, homicides, machine learning, poverty, time series, violence

1. Introducción

Colombia se ha enfrentado a una triste realidad de homicidios y violencia durante muchas décadas. Este problema tiene sus raíces en la conquista española, los conflictos por la tierra y el poder político, y la influencia de las ideas marxistas y leninistas. La presencia de guerrillas y grupos paramilitares ha contribuido a aumentar la violencia, especialmente en las zonas rurales.

El narcotráfico ha sido una fuente enriquecedora de esta violencia, y el sicariato se ha convertido en

una herramienta común para resolver disputas entre narcotraficantes y grupos delictivos. Las riñas y los robos violentos también han provocado muchas muertes, agravados por la recesión económica y la llegada de migrantes venezolanos.

Las masacres atribuidas a grupos guerrilleros, disidencias de las FARC y bandas delincuenciales han causado gran conmoción en diferentes ciudades y pueblos. El feminicidio ha percibido en la última década, reflejando la desfiguración de la familia y la violencia intrafamiliar.

Los índices de homicidio en Colombia son alarmantes, ubicándose entre los más altos del mundo. Aunque ha habido algunos períodos de disminución de la violencia, como resultado de acuerdos de paz con las FARC y las autodefensas, el problema persiste.

Los acuerdos de paz son una necesidad para reducir los homicidios y construir una sociedad más justa. La falta de equidad y la falta de educación, especialmente en fortalecer los valores familiares y promover la tolerancia, contribuyen a la violencia. La inversión social y la educación en derechos humanos son fundamentales para prevenir la repetición de estos actos violentos.

En general, la situación de homicidios y violencia en Colombia es compleja y requiere un enfoque integral para abordar sus causas subyacentes y promover la paz duradera.

El contexto de la investigación o estudio. Esta investigación se enfoca en el análisis de los datos estadísticos de homicidios en Colombia entre los años 2016 y 2020, empleando técnicas de análisis exploratorios de datos y modelos de series temporales para realizar un análisis detallado y encontrar posibles relaciones entre los perfiles sociodemográficos, educativos y económicos, y el fenómeno de los homicidios en el país.

La importancia de la investigación. Esta investigación surge de la necesidad de comprender y abordar el problema de los homicidios en Colombia. El país ha enfrentado altos niveles de violencia homicida, lo cual ha generado un impacto negativo en la sociedad. Es importante entender las causas y factores asociados a estos actos violentos para poder implementar políticas públicas efectivas de prevención de la violencia. Mediante el análisis de los perfiles sociodemográficos, educativos y económicos, se busca identificar los factores de riesgo y vulnerabilidad que influyen en la comisión de homicidios, lo cual permitirá orientar las intervenciones y políticas de prevención hacia las poblaciones más afectadas.

Los antecedentes Las muertes por homicidio representan una realidad trágica y alarmante en todas las sociedades del mundo. Este fenómeno impacta de manera significativa la seguridad, la estabilidad y la calidad de vida de las comunidades, convirtiéndose en una gran crisis de salud pública. Los homicidios no solo generan pérdidas irreparables para las familias y seres queridos de las víctimas, sino que también tienen efectos devastadores en la sociedad en general. Las consecuencias de los homicidios van más allá del aspecto emocional y humano, ya que tienen un impacto directo en el desarrollo social y económico de los países. La presencia de asesinatos conlleva violencia e inseguridad, lo cual puede dificultar el desarrollo económico, desalentar la inversión extranjera, socavar la confianza de los ciudadanos en las instituciones y restringir el acceso a servicios esenciales como la educación y la salud. La capacidad de los responsables de formular políticas y los profesionales para tomar decisiones efectivas que ayuden a reducir la delincuencia depende de disponer de datos en tiempo real de manera amplia y rápida.

Es importante analizar y estudiar los casos de homicidios y los perfiles sociodemográficos, educativos y económicos asociados a ellos, porque nos brindan

una visión profunda de las causas y factores de riesgo que contribuyen a la violencia homicida. de Lima Friche (2023) describió las tasas de homicidio entre jóvenes y adultos jóvenes, así como su asociación con factores socioeconómicos y del entorno construido en 315 ciudades de ocho países de América Latina y el Caribe. La tasa de muerte por homicidio fue sustancialmente más alta en hombres que en mujeres, pero hubo una heterogeneidad importante entre ciudades y unidades de subciudades dentro de las ciudades. Para ambos sexos, los niveles más altos de educación en las subciudades y el PIB de la ciudad se asociaron con tasas de homicidio más bajas, mientras que una mayor desigualdad de ingresos en la ciudad y un mayor aislamiento del paisaje de la ciudad se asociaron con tasas de homicidio más altas. Se identifica que la falta de oportunidades económicas y laborales, el tráfico de drogas, el consumo abusivo de alcohol, la falta de actividades de ocio y las interacciones con las fuerzas del orden son algunos de los posibles determinantes de los homicidios en jóvenes.

Knight (2023) investigaron el vínculo entre los delitos violentos y la inmigración utilizando datos de los municipios colombianos durante el periodo de inmigración desde Venezuela. Los autores identifican que luego del cierre y la reapertura de la frontera en 2016, que precipitó una ola de inmigración masiva, los homicidios en Colombia aumentaron en áreas cercanas a los principales pasos fronterizos. Usando información sobre la nacionalidad de la víctima, se encontró que este aumento fue impulsado por homicidios que involucran víctimas venezolanas, que fueron victimizadas de manera desproporcionada en relación con el tamaño de su población.

Dare et al. (2019) analizó la variación geoespacial, racial y educativa en la mortalidad por armas de fuego en EE. UU., México, Brasil y Colombia entre 1990–2015. El homicidio fue la intención más común en México, Brasil y Colombia. Mientras que el suicidio fue la intención más común en los Estados Unidos. Alrededor de 1,4 millones de muertes por armas de fuego ocurrieron en hombres de 15 a 34 años. El lugar de residencia, la raza y los logros educativos influyen en el riesgo de mortalidad por armas de fuego de una persona en todos los países. En cada país, la mayoría de las muertes por armas de fuego ocurrieron en unos pocos estados con una carga alta, pero los estados con las tasas más altas de mortalidad por armas de fuego variaron en los intervalos de 5 años de nuestro análisis.

Zambrano Ricaurte et al. (2022) analizó los homicidios en Colombia de enero de 2012 a diciembre de 2021 y realizó comparaciones entre los sexos. En este sentido, los datos descriptivos muestran que, los hombres se involucran en mayores conductas extremas de agresividad y que son más los hombres víctimas de homicidios. Al realizar modelos ARIMA, evidencia que con las series estacionalizadas los parámetros estimados mejoran considerablemente. Además, muestra que este tipo de modelos se puede aplicar para analizar los homicidios en Colombia que suelen mostrar aumentos significativos en los meses de diciembre en cada año. Adicionalmente, se encontraron diferencias absolutas promedio pequeñas entre los valores pronosticados y los valores reales de las

series total y dividida por sexos.

Restrepo-Betancur (2022) compara la proporción de muertes por homicidio y la tasa de homicidios entre los países de Sudamérica durante los últimos treinta años. La hipótesis es que existen diferencias estadísticas en las cifras de homicidios entre estos países, relacionadas con condiciones sociales, políticas y económicas. El autor concluye que en los países donde hay grupos al margen de la ley y problemas de desigualdad y pobreza se da el mayor número de casos. Se resalta la necesidad de generar acciones y estrategias para prevenir los homicidios y proteger la seguridad de los ciudadanos.

Ariza García (2021) realizó una investigación para determinar si el proceso de paz con las FARC ha tenido efectos sobre los homicidios en Colombia en el periodo 2010-2019 mediante la metodología de Box y Jenkins (2015), para la selección del modelo que explique favorablemente el comportamiento del número de homicidios. Además, el autor predice el número de homicidios a diciembre de 2020 en Colombia. Seguidamente realiza el análisis de intervención teniendo como fecha clave mayo de 2014, debido a los diferentes acontecimientos que presentaron un impacto en la época.

Norza Céspedes, Molano, Harker, y Buitrago Cubides (2020) analizaron la asociación entre la capacidad del Estado y la violencia a partir del análisis de la evolución de la tasa de homicidio y el desempeño de los gobiernos municipales, para el periodo 2003-2014 en los 1102 municipios de Colombia. El estudio se basó en el análisis de datos públicos proporcionados por la Dirección de Investigación Criminal e INTERPOL de la Policía Nacional de Colombia. Los investigadores examinan las trayectorias de cambio en las tasas de homicidios municipales en Colombia desde 2003 hasta 2014, y exploran las posibles asociaciones entre estos patrones de cambio y el índice de desarrollo integral de los municipios. Los autores identificaron una importante variabilidad geográfica en los patrones de cambio en la criminalidad, evidenciando un conjunto de municipios que experimentaron mejorías en el desempeño Estatal y seguridad.

Chaparro-Narváez, Cotes-Cantillo, León-Quevedo, y Castañeda-Orjuela (2016), (2016) realizaron un estudio descriptivo basado en la información de las bases de datos sobre mortalidad y las proyecciones de población del Departamento Administrativo Nacional de Estadística, DANE. Calcularon las tasas de mortalidad específica y ajustada por edad y sexo. En Colombia se registraron 331.470 homicidios entre 1998 y 2012. La tasa cruda promedio fue de 51,5 por 100.000 habitantes: 95,9 en hombres y 8,2 en mujeres. A partir del 2003 se observó una disminución en el número de muertes y tasas de mortalidad. El 91,9% de las víctimas fueron hombres. Las tasas de mortalidad más altas se observaron en el grupo de 20 a 29 años. El mecanismo más frecuentemente involucrado fue el arma de fuego: en ocho de cada diez homicidios en hombres y en siete de cada diez en mujeres. De los 1.122 municipios, 186 estaban en el quintil más alto y acumularon el 50,1 % de las defunciones.

Bello Montes (2008) describe los diferentes periodos históricos en los que se ha desarrollado la violencia homicida en Colombia, destacando la influencia del

conflicto armado, las organizaciones guerrilleras, el narcotráfico y las políticas de seguridad en los niveles de violencia en el país. La autora sugiere que la tasa de homicidio en Colombia ha sido una de las más altas en América Latina como consecuencia del impacto del tráfico de drogas, las organizaciones criminales que delinquen en el territorio y del ciclo de violencia que ha caracterizado la historia colombiana.

de los Ángeles Rodríguez-Gázquez (2005) evaluó y describió la evolución de la mortalidad por homicidio en la ciudad de Medellín, Colombia, durante el periodo 1975-2003. Con ayuda del programa SSS1 efectuó un análisis de series temporales utilizando procedimientos iterativos de construcción de modelos ARIMA. Encontró una clara variación estacional anual en la ciudad de Medellín de la mortalidad por homicidio y observando que en el mes de diciembre se produjeron las tasas más altas.

Sánchez, Tejada, y Martínez (2005) analizó cuatro tipos de muerte violenta (homicidio, accidente de tránsito, suicidio y accidental diferente de accidente de tránsito). Los autores aplicaron técnicas de análisis de series de tiempo a 84 observaciones correspondientes a casos de mortalidad por causas violentas durante 7 años en Bogotá y efectuaron un modelamiento con la metodología de Box Jenkins. Los autores obtuvieron un pronóstico de la serie a 12 meses, identificando una tendencia en la reducción del número de muertes durante el periodo analizado, más marcada en los casos de homicidio y accidente de tránsito. En todos los casos, excepto en los de suicidio, evidenciaron un patrón estacional, con incremento de frecuencia de muertes en diciembre. Los pronósticos señalan un sostenimiento de las tendencias decrecientes de las series, más marcado en el caso de homicidios y accidentes de tránsito, que en el de suicidios y otras muertes accidentales.

Vizcaíno y Duca (2002) analizan los actos homicidas circunstanciales u ocasionales a partir del testimonio de los actores involucrados, centrándose en la comprensión de estos actos como expresiones de violencia societaria. Los autores examinan las motivaciones, las valoraciones morales y la relación de los actores con las instituciones intermedias para arrojar luz sobre los factores que pueden contribuir a la aparición de la violencia homicida. El estudio concluyó que la precariedad de las instituciones primarias y secundarias y la soledad de los actores en un tejido social frágil caracterizan al capital social como negativo en la medida en que la desconfianza en el otro, la desinstitucionalización social y la escasa eficacia del Estado muestran que no hemos sido capaces de fraguar un cemento para la sociedad.

El objetivo general El objetivo principal es examinar los perfiles sociodemográficos, educativos y económicos de la población con el fin de identificar patrones y tendencias que expliquen el actuar violento que desencadena en homicidios, aplicando una metodología de análisis mediante técnicas de Machine Learning y de series de tiempo, que permita identificar cuales de los perfiles en mención aumentan la tasa de homicios en el país.

2. Métodos (Materiales y Métodos)

Se empleó la metodología CRISP DM (Cross-Industry Standard Process for Data Mining), que consiste en seis fases: comprensión del negocio (problema), comprensión de los datos, preparación de los datos, modelado, evaluación y utilización.

A través de los métodos estadísticos y el uso de técnicas de aprendizaje automático, se generaron gráficos exploratorios y modelos de series de tiempo multivariadas. Estos modelos permitieron examinar las relaciones entre las variables seleccionadas y la tasa de homicidios, así como una comprensión más profunda de los factores que podrían influir en la problemática.

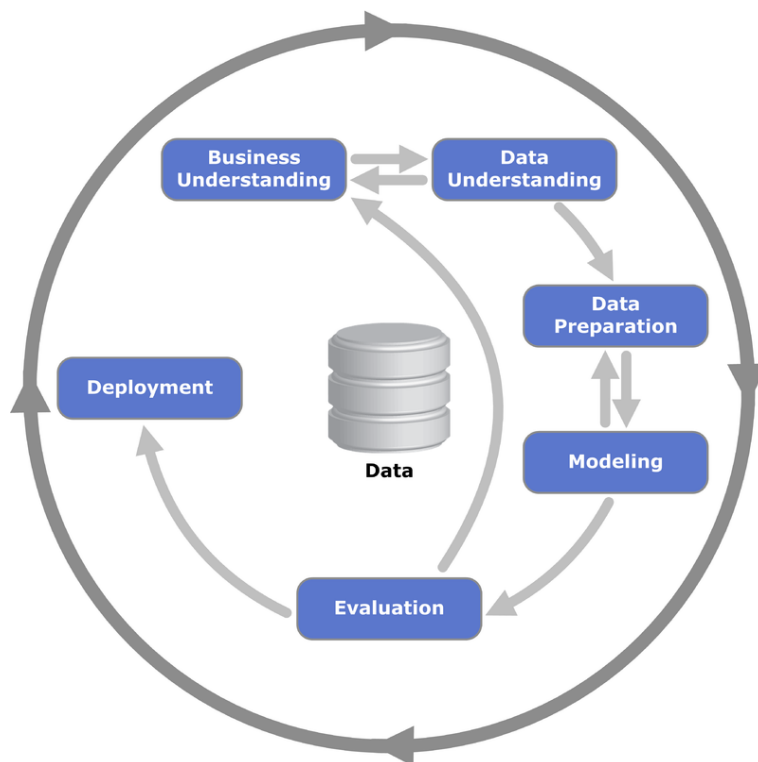


Figura 1: Metodología CRISP DM. Imagen tomada de Data mining.

2.1. Comprensión del problema

En esta investigación, se abordó la problemática de la alta tasa de homicidios en Colombia entre los años 2016 y 2020. El objetivo fue analizar este fenómeno por departamento y examinar algunas variables que podrían tener un impacto en él.

2.2. Comprensión de los datos

Los datos utilizados en esta investigación proporcionan información detallada sobre el número de homicidios en Colombia entre los años 2016 y 2020, y perfiles sociodemográficos, educativos y económicos relacionados. Para llevar a cabo este estudio, se recopilaban datos de varias fuentes como la Policía Nacional, el DANE (Departamento Administrativo Nacional de Estadísticas), la CEPAL (Comisión Económica para América Latina y el Caribe), el DNP (Departamento Nacional de Planeación) y HUMANITARIAN DATA EXCHANGE. Estas bases de datos proporcionan información relevante para el análisis.

Una vez recopilados los datos, se procedió a seleccionar y organizar la información en una tabla que agrupaba los aspectos a analizar (11 variables, 1981 observaciones). Posteriormente, se realizó un análisis de la tabla utilizando métodos estadísticos de aprendizaje

automático y series de tiempo multivariadas, empleando herramientas como Rstudio y Phytion.

La base de datos recopilada consta de 11 columnas que representan cada variable. Estas se clasifican en diferentes tipos, como se describen a continuación:

Año: indica el año de ocurrencia de los datos (2016 al 2020)

Mes: indica el mes de ocurrencia de los datos (los 12 meses del año)

Departamento: indica el departamento en el que se registraron los datos (32 departamentos).

Género Masculino: que representa el número de homicidios en hombres.

Género Femenino: representa el número de homicidios en mujeres.

Número de homicidios: representa la suma total de homicidios.

Tasa de homicidio *100.000 hts: representa la tasa de homicidios por cada 100,000 habitantes. Esta tasa se estimó así:

$$TH = \frac{NHM}{NHD} \times 100000 \quad (1)$$

Donde TH representa la Tasa de homicidio (por cada 100.000 habitantes), NHM representa el Número de

homicidios por mes y *NHD* representa el Número de habitantes por departamento (Censo 2018).

Cobertura bruta en educación - Total: que indica la cobertura bruta en educación. Se estimó así:

$$TTCE = \frac{NEE}{TND} \times 100000 \quad (2)$$

Donde *TTCE* representa la tasa total de cobertura en educación, *NEE*, representa niños que están estudiando, *TND* representa la población con edades entre 5 y 17 años.

Tasa de pobreza: representa la tasa de pobreza.

Tasa de desempleo: representa la tasa de desempleo.

Se procede a realizar un análisis estadístico básico, donde se evidencia una distribución de los datos mediante un violinplot, cada categoría se representa por un "violín" que muestra la densidad de probabilidad estimada de la variable en el eje vertical. La forma del violín muestra la distribución de los datos: es más ancho en las regiones donde hay más datos y más estrecho donde hay menos. Además, dentro de cada violín, se dibuja un diagrama de caja que muestra la mediana, los cuartiles y los posibles valores atípicos de los datos.

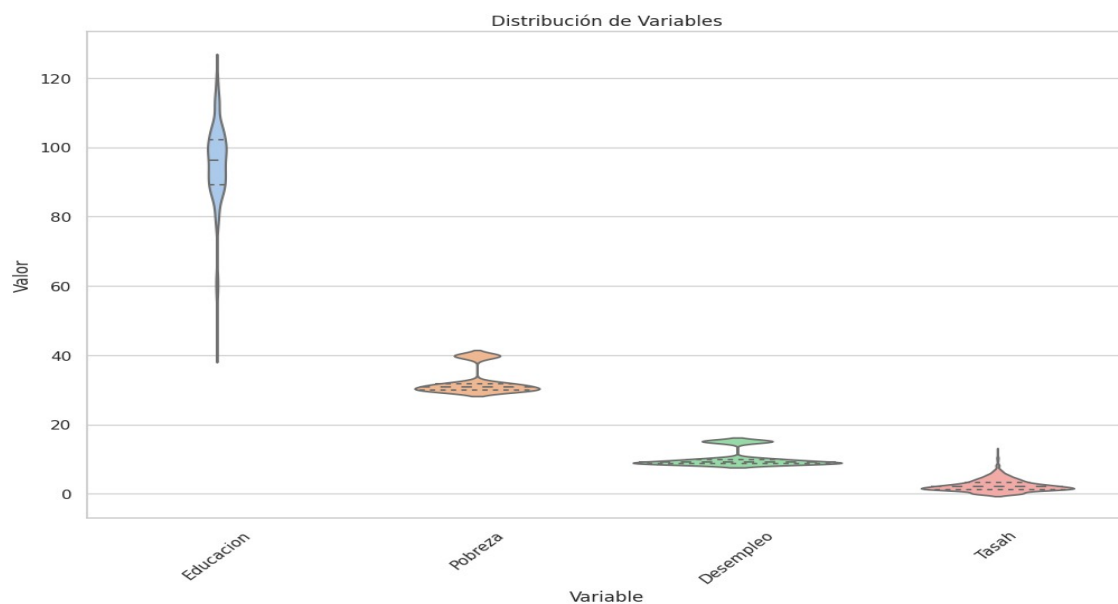


Figura 2: Análisis estadístico básico.

Realizado el análisis básico se puede observar en Educación (Cobertura bruta en educación - Total) que la distribución de esta variable muestra que la mayoría de los valores se encuentran en el rango de 100 a 110, hay cierta variabilidad en los valores superiores, lo que indica que algunos departamentos tienen una cobertura bruta en educación más alta que otros; en la variable Pobreza (Tasa de pobreza), la distribución muestra que la mayoría de los departamentos tienen una tasa de pobreza en el rango de 20 a 40, sin embargo, hay algunos departamentos con tasas de pobreza más altas, lo que indica la existencia de disparidades socioeconómicas entre las regiones; en la variable Desempleo (Tasa de desempleo), la distribución muestra que la mayoría de los departamentos tienen una tasa de desempleo en el rango de 6 a 10, al igual que con la variable de pobreza,

hay algunos departamentos con tasas de desempleo más altas, lo que refleja las diferencias en las condiciones económicas regionales, y en la variable Tasah (Tasa de homicidio *100.000 hts) la distribución de esta variable indica que la mayoría de los departamentos tienen tasas de homicidio bajas, en el rango de 0 a 5, sin embargo, también hay algunos departamentos con tasas más altas, lo que sugiere la existencia de áreas con problemas de seguridad más significativos.

2.3. Preparación de los datos

Se inicia la exploración de los datos con las gráficas de líneas separadas por tasa de homicidio, de cada una de las regiones del país.

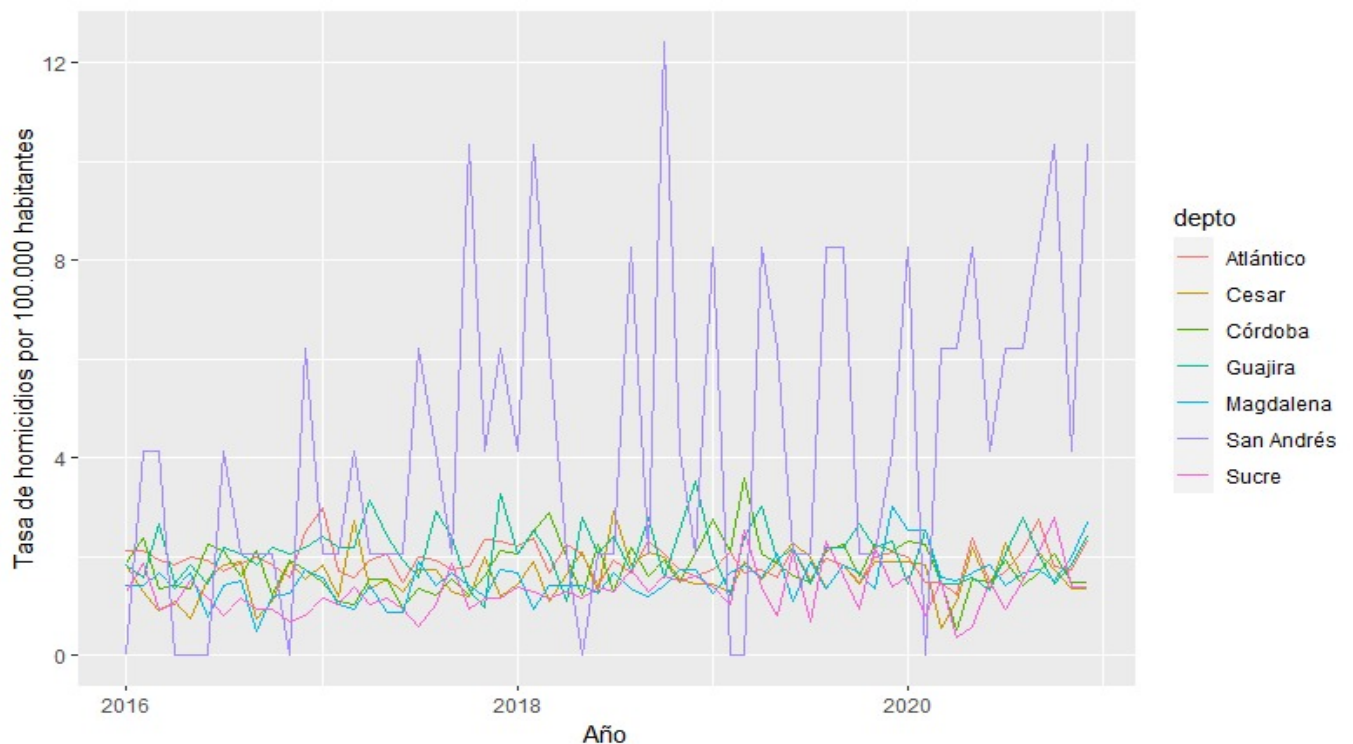


Figura 3: Gráfico de líneas separado por departamentos, tasa de homicidios. Región caribe.

Esta región consta de siete departamentos, en este grupo se incluyó San Andrés con una tasa de homicidios superior a 12 por cada 100.000 habitantes, siguiendo de manera descendente se encuentra a La Guajira, con picos en la tasa de homicidios cercanos a 4 por 100.000 habitantes, con valores elevados, pero este fenómeno se da debido al bajo censo poblacional, en donde un

homicidio impacta fuertemente a la población en términos cuantitativos.

De la gráfica de la región pacífica (figura 4), se puede resaltar que la tasa de homicidios en ella es relativamente heterogénea entre 2 y 7 homicidios por cada 100.000 habitantes, lo que hace que un departamento como Chocó alcance tasas superiores a 7 y Valle cercanas a él.

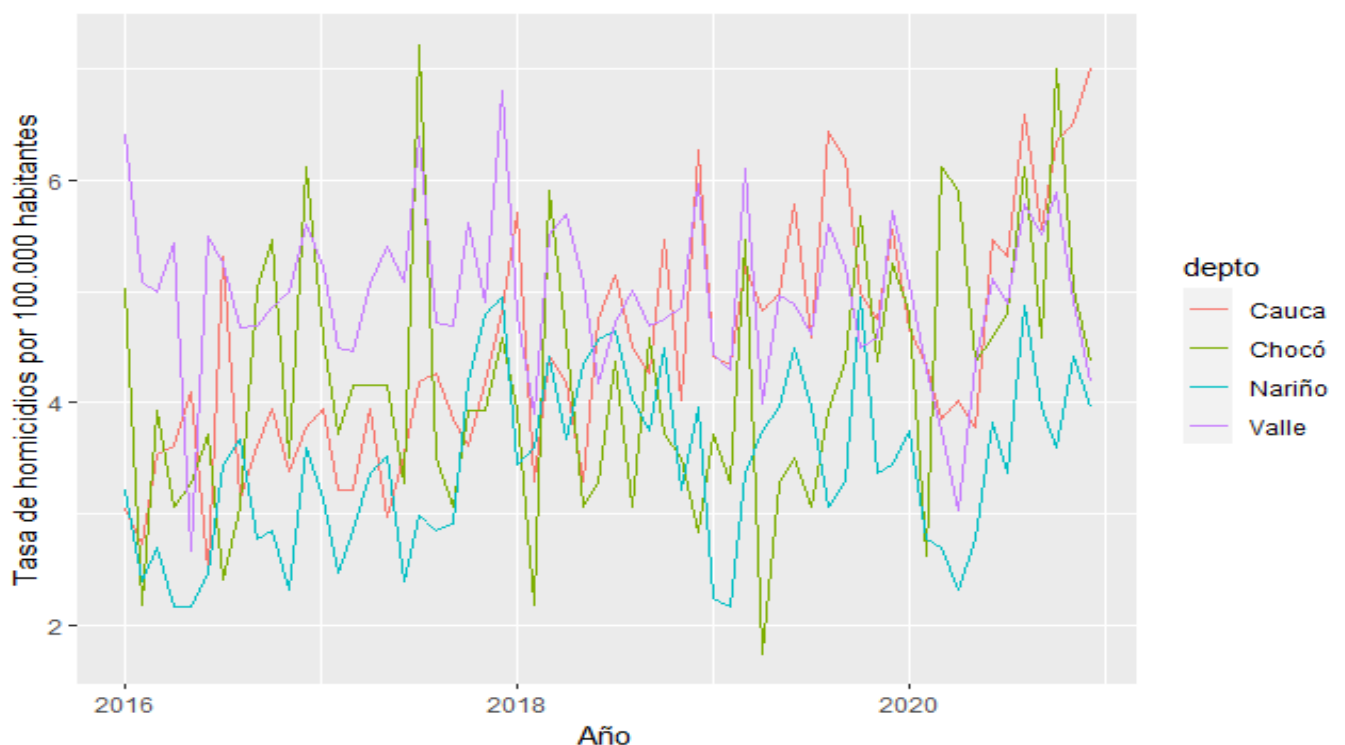


Figura 4: Gráfico de líneas separado por departamentos, tasa de homicidios. Región Pacífica.

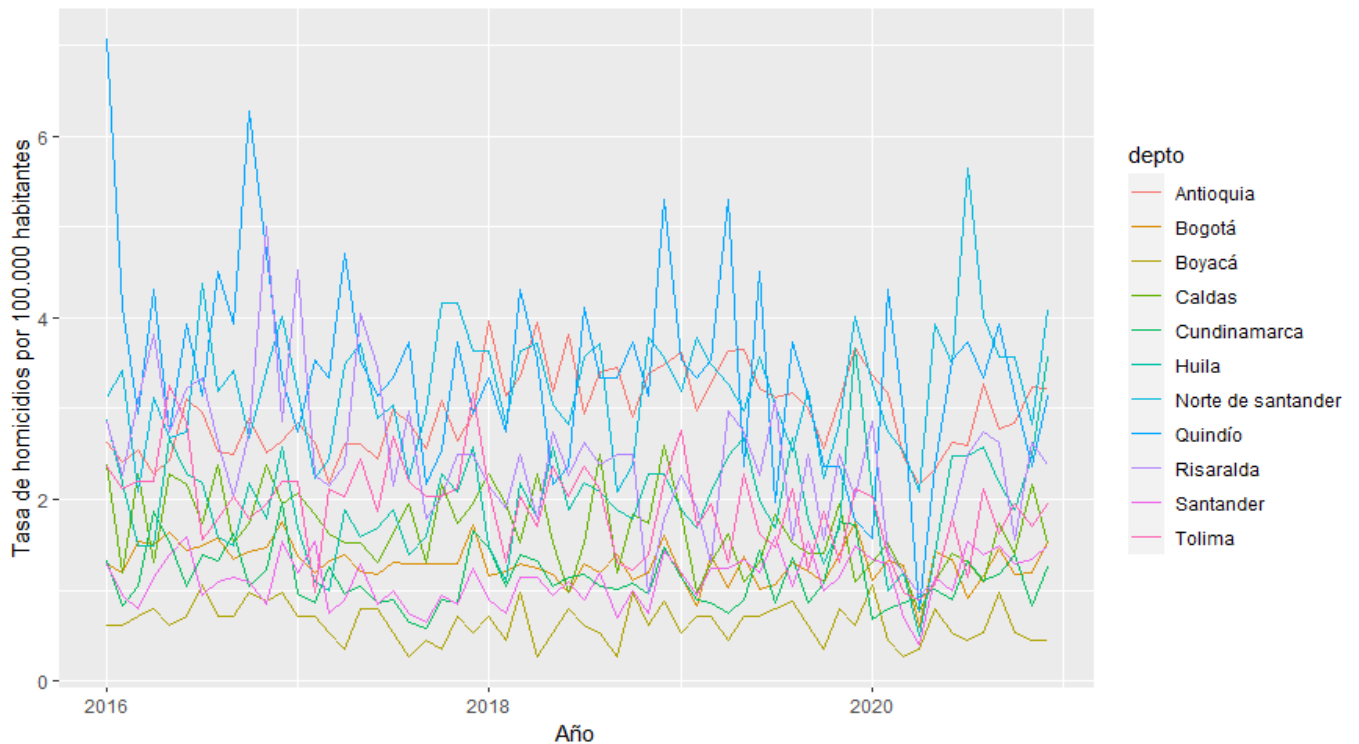


Figura 5: Gráfico de líneas separado por departamentos, tasa de homicidios. Región Andina.

La región Andina, una de las más grandes poblacionalmente, analizando el gráfico detalladamente, la tasa de homicidios no es tan alta como la de otras regiones, considerablemente con censo poblacional mas bajo, en esta región el departamento que tiene picos aislados es el Quindío, alcanzando valores entre 6 y 7

homicidios por cada 100.000 habitantes, pero en promedio la tasa de homicidios de la región se comporta entre 0.5 y 4.5, aún teniendo entre sus departamentos a Antioquia en el cual se ubica la ciudad Medellín, y la ciudad de Bogotá que también aparece en este departamento, las cuales están entre las principales del país.

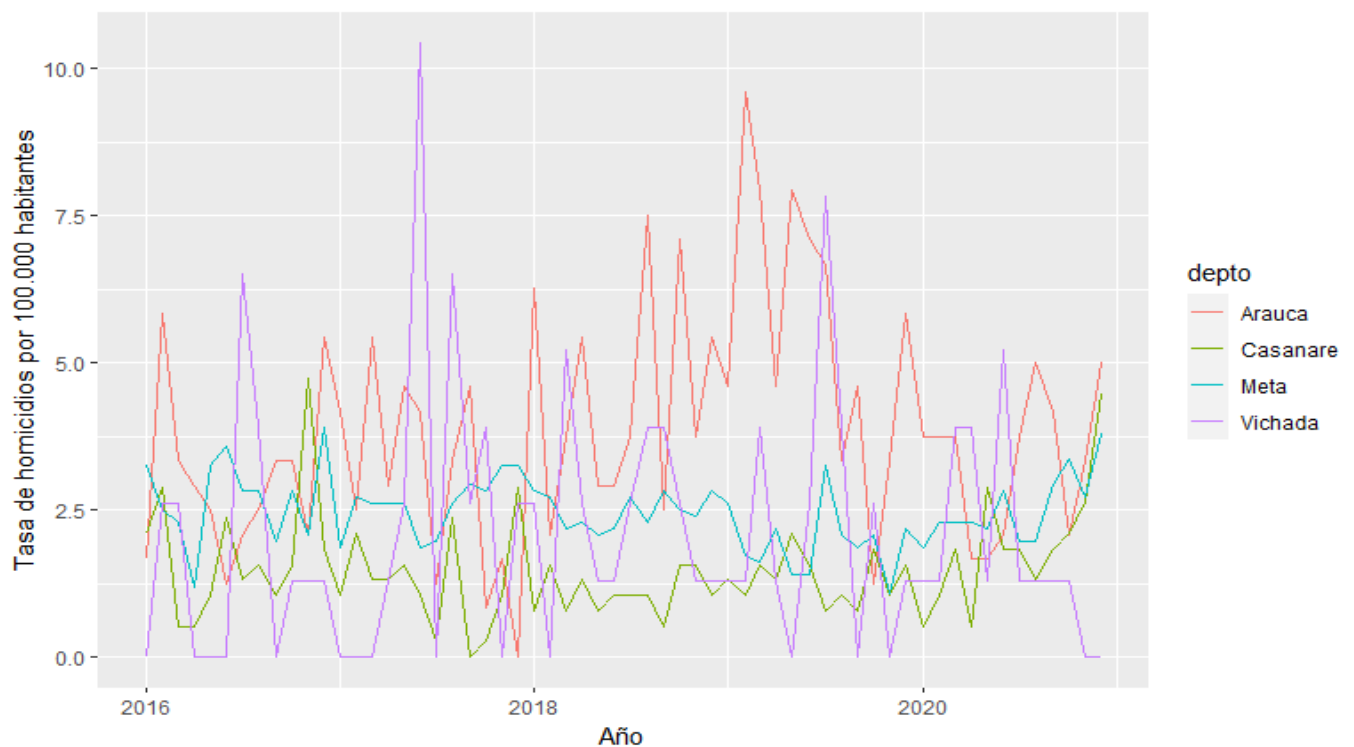


Figura 6: Gráfico de líneas separado por departamentos, tasa de homicidios. Región Orinoquía.

La región Orinoquía comprendida por cuatro departamentos: Arauca, Casanare, Meta y Vichada con tasas que superan los 10 homicidios por cada 100.000 habitantes, y Arauca que supera los 9,5 homicidios.

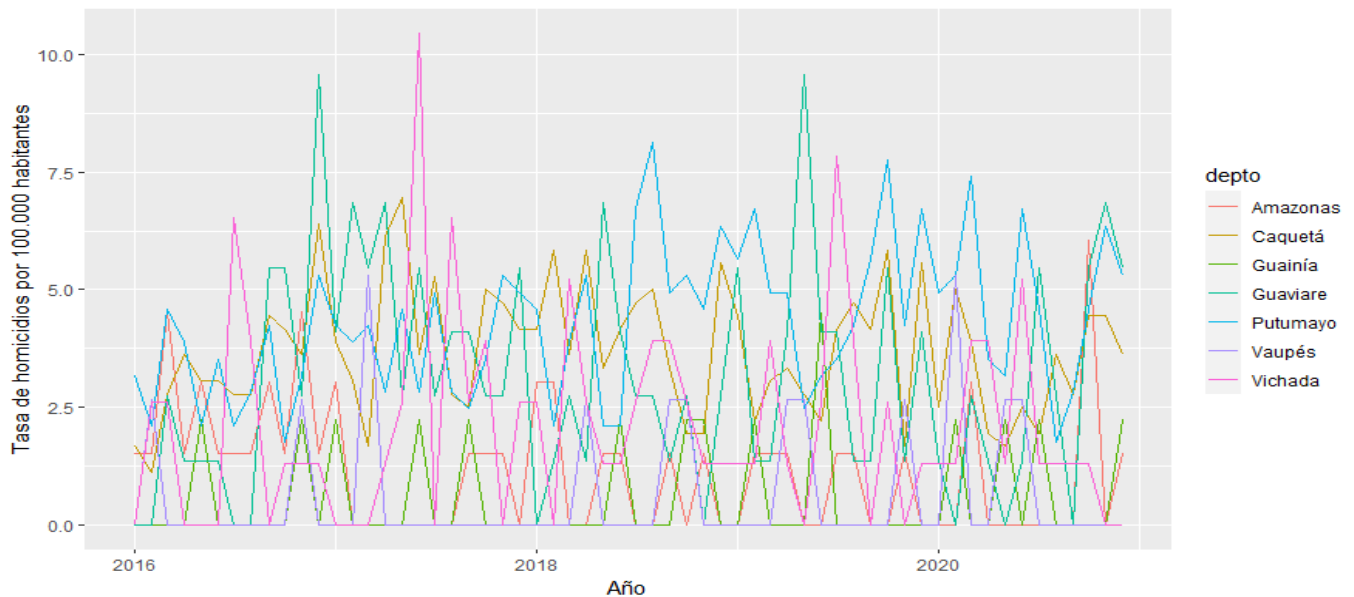


Figura 7: Gráfico de líneas separado por departamentos, tasa de homicidios. Región Amazonas.

La región Amazonía es la que sostiene una alta tasa de homicidios, en promedio se compara entre 2 y 10 homicidios por cada 100.000 habitantes, un valor muy alto respecto al número de pobladores.

El departamento del Vichada pertenece tanto a la región Amazonia como Orinoquia, para evitar duplicidad de datos se grafica en la región Orinoquía. Igualmente pasa con el departamento del Guaviare pertenece a la región Amazonia y a la Orinoquia, entonces se incluye en

el grafico de Amazonia.

Con el gráfico de heterogeneidad se puede identificar fácil la diferencia que hay entre el gráfico de tasa de homicidio y numero de homicidio, los departamentos que tienen más dispersión, no hay continuidad de los puntos, son los que en el periodo estudiado tuvieron tasas altas pero no frecuentes, mientras que los departamentos cuyos eventos son frecuente se ven como conglomerados y una tasa de homicidio más bajo.

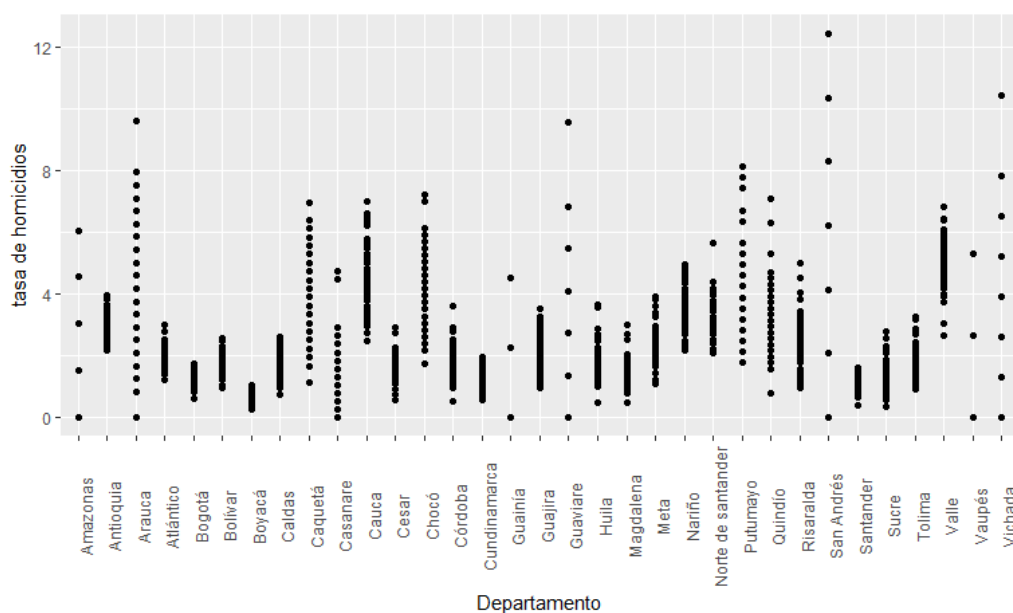


Figura 8: Heterogeneidad entre los departamentos.

Debido a que la información exploratoria evidenciada en los gráficos anteriores le dio al grupo de trabajo un resultado ajeno, se hace el gráfico con el número de

homicidios de los cinco departamentos donde se ubican las principales ciudades del país (Bogotá, Valle, Antioquia, Atlántico y Bolívar.)

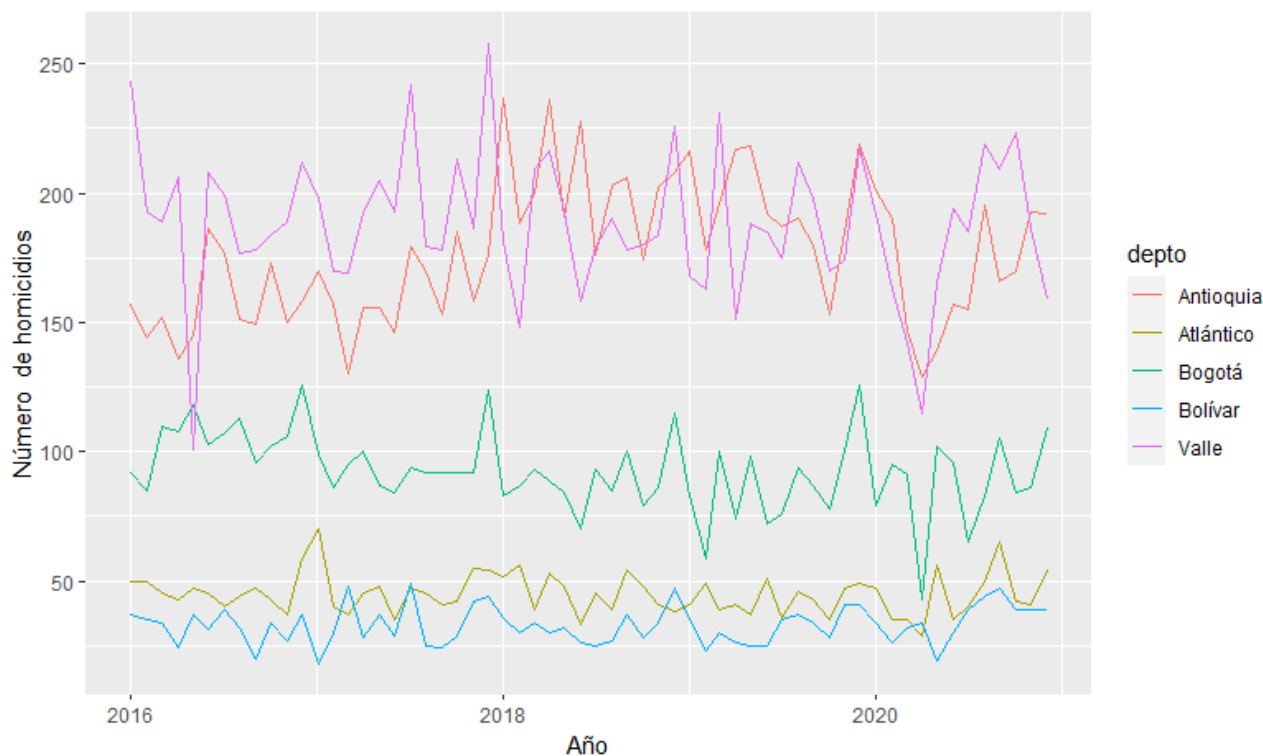


Figura 9: Número de homicios - cinco departamentos.

De este gráfico se puede concluir que los tres departamentos con número de homicidios más alto es Valle, Antioquia y Bogotá, en estos departamentos se encuentran las ciudades de Cali y Medellín.

2.4. Modelamiento

En este estudio, la variable dependiente es la tasa de homicidios en Colombia, las variables independientes son la pobreza, la educación, el desempleo. En Machine Learning, se utilizó el modelo de regresión de árbol de decisiones (DecisionTreeRegressor), utiliza la información contenida en las características para hacer predicciones sobre el valor numérico que se desea predecir. El árbol se construye dividiendo recursivamente los datos en función de las características más importantes y maximizando la reducción de la incertidumbre en la variable objetivo. Se empleó para predecir la variable objetivo de tasa de homicidios por cada 100.000 habitantes en Colombia.

La base de datos se dividió en dos conjuntos, un conjunto de entrenamiento (80 % de los datos) y un conjunto de prueba (20 % de los datos). El conjunto de entrenamiento se utilizó para ajustar el modelo y el conjunto de prueba se utilizó para evaluar su rendimiento. Se creó una instancia del modelo DecisionTreeRegressor y se ajustó utilizando el conjunto de entrenamiento. Durante el entrenamiento, el modelo aprendió a tomar decisiones basadas en las características para predecir la variable objetivo (ver anexo 1). La implementación se realizó en python, utilizando la librería sklearn.model.

La metodología utilizada para el análisis mediante series de tiempo multivariada es el método Panel – Modelo de efectos fijos, este método consiste en tener una variable dependiente, que es el elemento o la problemática a estudiar, y varias variables independientes que se asumen que están correlacionadas con la variable dependiente y esta correlación es fija durante todo el tiempo observado, y elimina las variables que son fijas en el tiempo, buscando las posibles dispersiones de las variables independientes. En el desarrollo de la aplicación de este método se ejecutó la prueba de mínimos cuadrados hallando que la variable que está correlacionada con la tasa de homicidios es cobertura en educación con un p-valor de 0.005, y cuyo coeficiente es negativo, si la cobertura en educación disminuye un punto porcentual la tasa de homicidios podría incrementar en 4,8 unidades al mes. Con el modelo de efectos fijos se evidencia que las variables independiente que impactan el modelo son el porcentaje de desempleo y el porcentaje de pobreza, cada una de ellas con un coeficiente de 0.17, -0.10, respectivamente, los departamentos con intercepto más altos son: Valle, Cauca, San Andrés, Putumayo y Arauca, lo que indica que en el caso que las variables incrementen una unidad, estos departamentos verán incrementada la tasa de homicidios en el valor que acompaña a cada departamento nombrado.

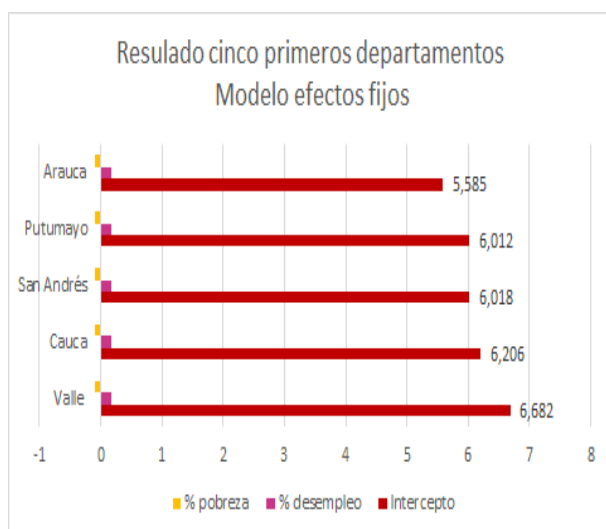


Figura 10: Intercepto por departamentos con tasas de homicidios más elevadas.

Del anterior gráfico se puede interpretar que en caso de que las variables pobreza y desempleo incrementen una unidad, estos departamentos verán incrementada la tasa de homicidios en valor de la etiqueta que acompaña a cada departamento nombrado, en atención a esto el departamento del Valle continúa siendo el más susceptible ante estas problemáticas con un posible incremento de 6,682 homicidios por 100.000 habitantes.

El modelo efectos aleatorios al contrario que el modelo descrito anteriormente, asume que las variables no están correlacionadas con la variable dependiente, al contrario, son aleatorias y se evalúan entre-individuos

e intra-individuos, esto significa que entre-individuos se evalúa un todo en conjunto e intra-individuos se analiza cada departamento, respecto a estos dos estadísticos hallamos que en la problemática que se está tratando, la variabilidad entre-individuos es del 45,5 % e intra-individuos de 54,5 %, aunque la varianza intra-departamento es más alta no podemos decir que la entre-departamentos es despreciable, por el contrario, es necesario integrar los dos aspectos. Además este estadístico suministra un intercepto de 4.118, las variables significativas son tasa de desempleo y tasa de pobreza, con p-valor de 0.012 y 0.021 respectivamente, y cada una con un coeficiente de 0.171 en pobreza y -0.103 de desempleo, lo que indica que por cada unidad o porcentaje que incremente alguna de las variables descritas la tasa de homicidio incrementaría en 4.118

Ahora se hace un juzgamiento de hipótesis para Datos Panel y se decide con cuál modelo se concluye el análisis, si es con modelo de efectos fijos (FE) o modelo de efectos aleatorios (RE), para esta decisión se utiliza el test de Hausman, su finalidad es evaluar la diferencia entre sus varianzas y los coeficiente, si la diferencia es pequeña entre ellos se puede concluir que no hay evidencia suficiente para rechazar la hipótesis nula (efectos aleatorios consistentes), pero si la diferencia entre los coeficientes y las varianzas es importante se concluye que el mejor modelo es el de FE.

En este test el p-valor es el dato que determina la escogencia del modelo, si este es $\leq$ a 0.05, se puede determinar que el modelo adecuado es del Efectos fijos, si es superior el mejor modelo es del RE. A continuación, presentamos el resultado de este análisis:

Tabla 1: Hausman Test

data	chisq	dg	p-value
tasa_hom~tasa_des + cob_edu	0.25129	3	0.9689

El resultado es un p-valor de 0.969, lo que nos lleva a tomar la decisión que el modelo mas apropiado es el RE, no es obligatoria la correlación y se debe tener en cuenta el porcentaje entre el intra-departamento y entre-departamento.

Dado el anterior resultado se describe la ecuación del modelo Efectos aleatorios, que fue el modelo escogido teniendo como referencia el resultado del test de Hausman.

La ecuación es:

$$Y_{it} = \alpha_i + \beta X_{it} + e_{it} \quad (3)$$

- α_i : Es el coeficiente del intercepto, que representa el valor esperado de cada una de las entidades cuando las variables independientes son cero y los efectos aleatorios también son cero.
- X_{it} : Es un vector de predictores (para la entidad i en el tiempo t).
- β : Es un coeficiente para cada entidad dada. Cuando un predictor cambia una unidad con el

tiempo, el resultado será aumentar/disminuir en β unidades (asumiendo que no se requiere ninguna transformación).

- e_{it} : Es el error residual para la unidad i en el tiempo t .

2.5. Evaluación

El siguiente gráfico representa la comparación entre los valores reales y los valores predichos por el modelo. En el eje x se encuentra el índice, que puede ser interpretado como una secuencia de puntos en el tiempo o en el orden en que aparecen los datos. En el eje y se encuentran los valores correspondientes a las mediciones de las tasas homicidios por cada 100.000 habitantes.

El siguiente gráfico muestra dos líneas: una que representa los valores reales y otra que representa los valores predichos. Ambas líneas están conectadas por segmentos rectos, lo que indica la tendencia general de los datos a medida que el índice aumenta.

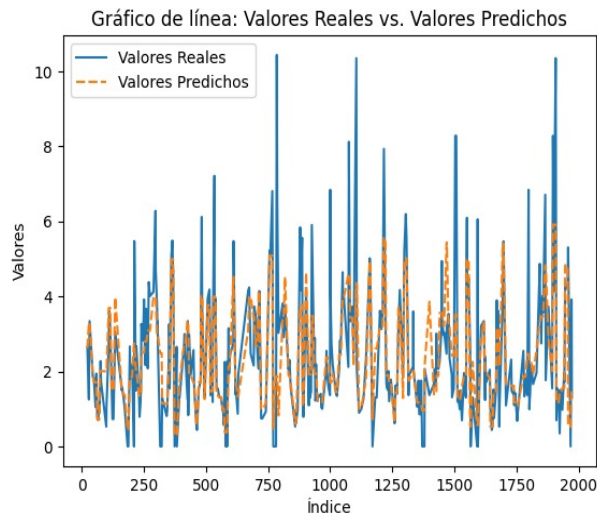


Figura 11: Valores reales y valores predichos.

Se puede observar que los valores predichos se ajustan bien a los valores reales, por la cercanía y la forma que siguen las líneas, se podría deducir que el modelo tiene un buen desempeño y es capaz de capturar la estructura y las tendencias de los datos reales.

A continuación se realiza un análisis de las métricas de evaluación del modelo para tener una comprensión más completa del rendimiento del modelo y la calidad de las predicciones.

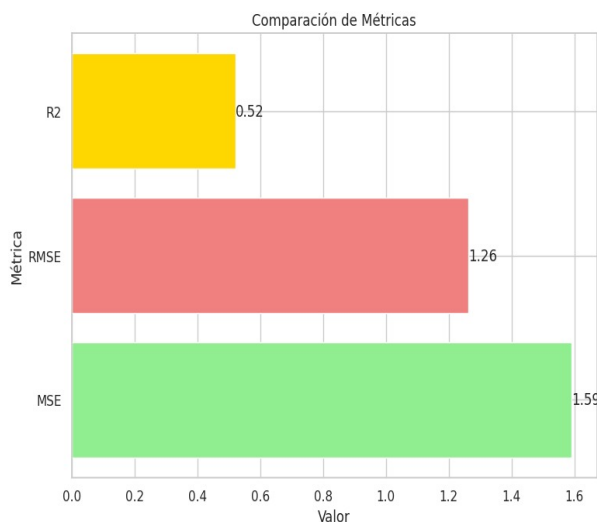


Figura 12: Comparación de métricas.

Para evaluar el modelo se realizaron predicciones utilizando el conjunto de prueba y se compararon con los valores reales de la variable objetivo. Se utilizaron varias métricas de evaluación para medir el

rendimiento del modelo, incluyendo el Error Cuadrático Medio ($MSE=1.591$), el Coeficiente de Determinación ($R^2=0.521$), el ($RMSE=1.26$) raíz cuadrada del Error Cuadrático Medio y el Error Porcentual Absoluto Medio ($MAPE=Inf$ debido a que hay valores en 0).

El MSE representa la media de los errores al cuadrado entre las predicciones y los valores reales. Un valor más bajo de MSE indica una mejor precisión del modelo. El MSE obtenido es de 1.5915, lo cual indica que el modelo tiene un nivel de precisión moderado. Esto significa que, en promedio, los errores al cuadrado entre las predicciones y los valores reales son relativamente altos.

El R2 mide la proporción de la varianza de la variable objetivo que es explicada por el modelo. Un valor más cercano a 1 indica un mejor ajuste del modelo a los datos. El valor de R2 es de 0.5210, lo que sugiere que aproximadamente el 52.10 % de la varianza en la variable objetivo puede ser explicada por el modelo. Esto indica un ajuste moderado del modelo a los datos, pero aún existe una cantidad significativa de variabilidad no explicada.

El RMSE es la raíz cuadrada del MSE y representa la desviación estándar de los errores del modelo. Proporciona una medida más interpretable del error en las mismas unidades que la variable objetivo. El RMSE obtenido es de 1.2616. Esta medida indica que, en promedio, los errores de predicción tienen una magnitud de aproximadamente 1.2616 en las mismas unidades que la variable objetivo.

El MAPE es el promedio del porcentaje absoluto de los errores entre las predicciones y los valores reales. Es una medida relativa que representa el error promedio como un porcentaje del valor real. Un valor más bajo de MAPE indica una mejor precisión del modelo. El MAPE calculado es infinito (inf), lo cual indica que hubo divisiones por cero debido a la presencia de cero valores reales en la muestra. Dado que el MAPE es una medida relativa que representa el error como un porcentaje, su cálculo no es válido en este caso.

Ajuste de parámetros: Se exploraron diferentes configuraciones de parámetros del modelo, como la profundidad máxima del árbol y el número mínimo de muestras requeridas para dividir un nodo, con el fin de encontrar la configuración que optimizara el rendimiento del modelo.

Dado que el MSE y el RMSE son relativamente altos y el R2 es moderado, se puede concluir que el modelo tiene una precisión promedio y aún queda margen para mejorar su ajuste a los datos. Es importante considerar que por la naturaleza específica del problema y las características de los datos para evaluar el modelo se considera que los resultados son aceptables.

PDP for feature "educacion"

Number of unique grid points: 10

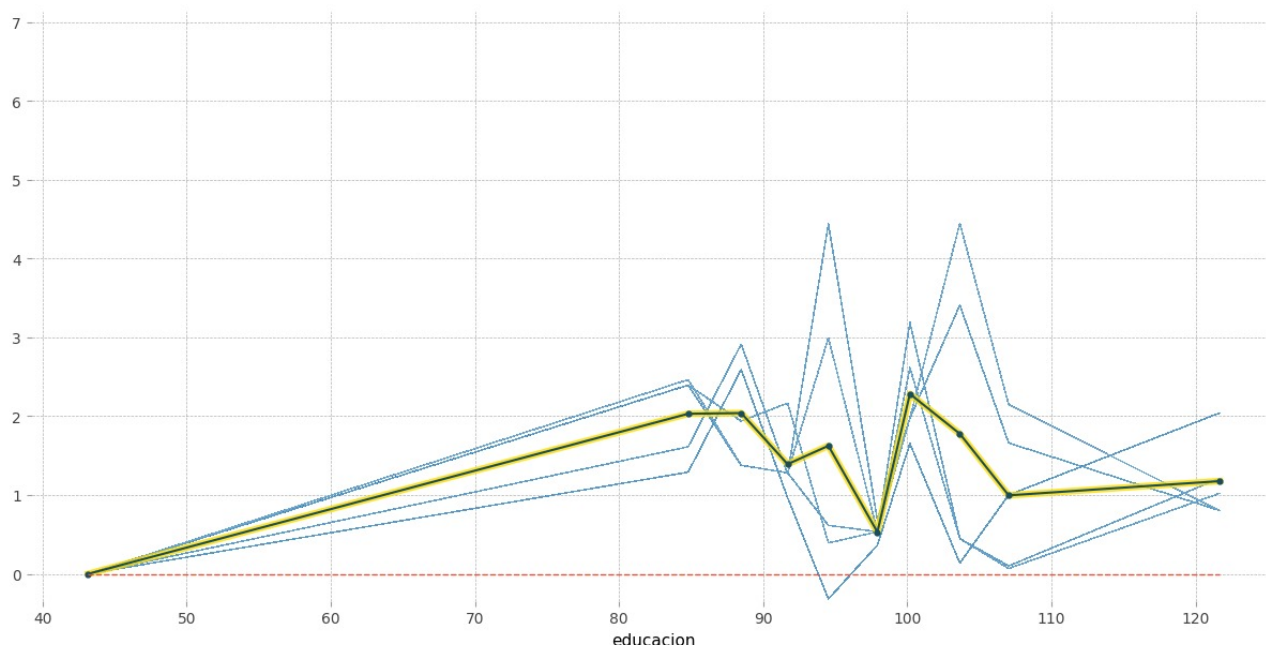


Figura 13: Dependencias parciales.

En el rango de valores desde 0 hasta aproximadamente el punto 85, se observa una relación positiva entre la cobertura bruta en educación y la tasa de homicidio. Esto significa que a medida que el nivel de educación aumenta en esa gama, la tasa de homicidio tiende a ser más baja. Esta relación inicial sugiere que una mayor cobertura educativa está asociada con una disminución en la tasa de homicidio, respaldando la idea de que la educación puede actuar como un factor protector contra la violencia.

Sin embargo, a partir del punto 85, la relación entre la educación y la tasa de homicidio se vuelve más compleja. Se observa que las líneas azules que ascienden por encima de la línea resaltada de amarillo indican que hay otros factores en juego que pueden influir en esta relación y comienzan a tener un impacto más significativo en la tasa de homicidio. Esto sugiere que la educación por sí sola puede no ser suficiente para explicar completamente las variaciones en la tasa de homicidio, y que existen factores adicionales que pueden contrarrestar los efectos positivos de la educación en la reducción de la violencia.

En cuanto a la Tasa de pobreza y la Tasa de desempleo, se observa que a medida que aumentan estos indicadores socioeconómicos, también aumenta la tasa de homicidio. Esto respalda la expectativa de que condiciones socioeconómicas desfavorables puedan contribuir al incremento de la violencia. Estos resultados sugieren que la pobreza y el desempleo pueden ser factores de riesgo que se asocian con una mayor tasa de homicidio.

Es importante destacar que en el punto de referencia

89 se observa nuevamente un aumento significativo en la importancia de la Cobertura bruta en educación, lo cual podría indicar una relación contradictoria con la variable objetivo. Esto podría deberse a la influencia de otros factores o dinámicas socioeconómicas complejas que interactúan con la educación y afectan la tasa de homicidio de manera no lineal. Estos hallazgos subrayan la necesidad de considerar la interacción entre múltiples variables y la complejidad de los factores que influyen en la tasa de homicidio.

En resumen, esta gráfica revela que la Cobertura bruta en educación, es la variable más relevante para predecir la tasa de homicidio. A medida que se avanza en los puntos de referencia, se observan cambios en la importancia relativa de estas características, lo que indica la existencia de relaciones complejas y no lineales entre las variables analizadas y la tasa de homicidio. Estos resultados destacan la importancia de realizar un análisis más detallado para comprender en mayor profundidad la influencia de cada variable y considerar otros posibles factores que puedan afectar la relación entre las variables analizadas y la tasa de homicidio.

El objetivo que se trazó al inicio de este análisis de datos, es identificar de las variables desempleo, educación y pobreza, cuál era la que influía significativamente en el alto número de homicidios en el país, pero usando las herramientas estadísticas que contiene el método panel y el test de Hausman el mejor modelo es el de Efectos Aleatorios, ya que de las tres variables analizadas la correlación es débil con la tasa de homicidios, lo que lleve a pensar que el estudio realizado por Restrepo

(2021), donde describe que la tasa de homicidios es tan alta debido a la disputa entre grupos armados, su actuar delictivo, aún está vigente, y sin olvidar que distintos gobiernos han sostenido diálogos de Paz con algunos actores del conflicto. Las manifestaciones de buena voluntad de los gobiernos no bastan para bajar la cultura homicida que se tiene, es un tema mas cultural, o heredado como lo recuerda Bello Montes (2008), donde menciona que comunidades que han vivido en conflicto tienen más probabilidad de continuar con este ciclo, en esos términos la inversión social debe ser enfocada en cambiar la cultura de los habitantes, tocar su psicología y forma de pensar.

3. Resultados y discusión

Se utilizó un modelo de regresión de árbol de decisiones y se realizaron diversos análisis y visualizaciones. Se obtuvieron los siguientes resultados de la evaluación del modelo:

MSE: 1.592, R2: 0.521, RMSE: 1.262, MAPE: inf.

Estos valores nos indican que el modelo presenta una capacidad moderada para explicar la variabilidad de la tasa de homicidios. El valor de R2 de 0.52 indica que aproximadamente el 52% de la variabilidad de la variable objetivo puede ser explicada por las variables independientes incluidas en el modelo.

Luego, se realizó un análisis de importancia de características para determinar la influencia de cada variable en la predicción de la tasa de homicidios. Los resultados indicaron que la variable Cobertura bruta en educación - Total^{es} la más importante, seguida por "Tasa de pobreza" "Tasa de desempleo". Esto sugiere que la educación, la pobreza y el desempleo pueden desempeñar un papel significativo en la determinación de la tasa de homicidios.

Además, se llevaron a cabo análisis de dependencias parciales para examinar las relaciones entre la variable objetivo (tasa de homicidios) y las variables de interés. Se observó que, en general, existe una tendencia creciente en la tasa de homicidios a medida que aumenta la cobertura bruta en educación. Sin embargo, a partir de cierto punto (aproximadamente el punto 85), las tendencias difieren entre las variables. Algunas variables presentan una disminución en la tasa de homicidios, mientras que otras continúan aumentando o tienen cambios abruptos en su comportamiento. Estas diferencias pueden indicar la presencia de interacciones complejas entre las variables o la influencia de factores adicionales no considerados en el modelo.

Es importante destacar que el análisis realizado se basa en los datos disponibles y en el modelo utilizado. Por

lo tanto, los resultados y conclusiones deben interpretarse con cautela y considerar las limitaciones del estudio. Es posible que existan otras variables no consideradas en el análisis que puedan influir en la tasa de homicidios y que podrían ser objeto de futuras investigaciones.

4. Conclusiones

Los resultados obtenidos en este estudio por aprendizaje automático, sugieren que la educación, la pobreza y el desempleo pueden desempeñar un papel importante en la determinación de la tasa de homicidios en la región estudiada. Sin embargo, se requiere un análisis más profundo y la consideración de otros factores para comprender completamente las dinámicas y causas subyacentes de los homicidios en Colombia. Estos hallazgos proporcionan una base inicial para futuras investigaciones orientadas a abordar el problema de la violencia y mejorar las condiciones socioeconómicas en el país.

Una vez terminados los análisis en series de tiempo, de esta problemática homicida en el país, se puede concluir que de las tres variables estudiadas (desempleo, educación y pobreza) ninguna es un detonante fuerte en el aumento de este delito, aunque el tratamiento de los datos estadísticos son entre los años 2016 y 2020, no distan mucho de la actualidad, mientras la tasa de homicidios en el año 2020 fue del 22,64 (fuente datosmacro Expansión), en el año 2021 la tasa homicida fue 26,8 (fuente infobae diario en línea Argentino), y un estudio realizado por la Universidad Externado de Colombia indica que la tasa de mortalidad por homicidio en Colombia en el año 2022, se situó en el 26,08 (las tasas anteriores todas calculadas por grupos de 100.000 habitantes) se puede concluir que el fenómeno observado no refleja un decrecimiento significativo, esto sin dejar de aclarar que el año 2020 fue uno donde las medidas de contingencia por la Covid 19, nos obligó a estar aislados, con cerramientos muy fuertes donde sólo teníamos acceso a lo vital (alimentación, medicamentos y educación remota), la tasa de homicidios en el país no tuvo una caída abrupta, hubo disminución significativa de los números de homicidios entre los meses abril, mayo y junio del año 2020, posteriormente la disminución no fue tan significativa. Teniendo en cuenta la investigación realizada por Melendez y de la Hoz (2019), en uno de sus apartes menciona que las manifestaciones de violencia intrafamiliar, riñas entre pandillas o riñas personales que se manifiestan como una expresión de la baja valoración de la tolerancia social en la ciudad de Cartagena, y ante el panorama de la investigación sería un buen punto evaluar si la tasa de homicidios es directamente proporcional con la cultura de las personas.

Referencias

- Ariza García, E. M. (2021). Homicidios y el proceso de paz en colombia. Descargado de <https://repositorio.udes.edu.co/server/api/core/bitstreams/7e484c45-86fc-41ee-8598-83e3d8b07e0c/content>
- Bello Montes, C. (2008). La violencia en colombia: Análisis histórico del homicidio en la segunda mitad del siglo xx. *Revista criminalidad*, 50(1), 73–84. Descargado de <http://www.scielo.org.co/pdf/crim/v50n1/v50n1a05.pdf>
- Chaparro-Narváez, P., Cotes-Cantillo, K., León-Quevedo, W., y Castañeda-Orjuela, C. (2016). Mortalidad por homicidios en colombia, 1998-2012. *Biomédica*, 36(4), 572–582. Descargado de <https://doi.org/10.7705/biomedica.v36i4.2811>
- Dare, A. J., Irving, H., Guerrero-López, C. M., Watson, L. K., Kolpak, P., Shigematsu, L. M. R., ... Gelband, P., Hellen y Jha (2019). Variación geoespacial, racial y educativa en la mortalidad por armas de fuego en los ee. uu., méxico, brasil y colombia, 1990–2015: un análisis comparativo de datos de estadísticas vitales. (6). Descargado de [https://doi.org/10.1016/S2468-2667\(19\)30018-0](https://doi.org/10.1016/S2468-2667(19)30018-0)
- de Lima Friche, U. M. y. B. U. y. S. O. L. y. d. S. D. M. A. y. P.-G. F. J. y. B.-L. R. y. A. M. y. D.-R. A. V. y. C. W. T., Amélia Augusta y Silva. (2023). Variación de las tasas de homicidio de jóvenes y adultos jóvenes y su asociación con las características de las ciudades en américa latina: el estudio salurbal. *The Lancet Regional Health–Américas*, 20. Descargado de <https://doi.org/10.1016/j.lana.2023.100476>
- de los Ángeles Rodríguez-Gázquez, M. (2005). Evolución de la mortalidad por homicidio en medellín (colombia): 1975-2003. *Gaceta sanitaria*, 19(3), 238–241. Descargado de <https://doi.org/https://doi.org/10.1157/13075958>
- Knight, A., Brian y Tribin. (2023). Inmigración y delitos violentos: Evidencia de la frontera entre colombia y venezuela. *Diario de Economía del Desarrollo*. Descargado de <https://doi.org/10.1016/j.jdeveco.2022.103039>
- Melendez, D. S., y de la Hoz, N. M. (2019). Aplicación de la política criminal en materia de homicidio en cartagena. *Revista Jurídica Mario Alario D'Filippo*, 11(22), 123–133.
- Norza Céspedes, E., Molano, A., Harker, A., y Buitrago Cubides, J. (2020). Trayectorias de la violencia homicida y desempeño estatal en colombia. *Colombia Internacional*(101), 91–120. Descargado de <https://doi.org/10.7440/colombiaint101.2020.04>
- Restrepo, A. M. (2021). Un acercamiento al estado actual de los estudios sobre genocidio. En *Revista forum* (pp. 50–65).
- Restrepo-Betancur, L. F. (2022). Evaluación estadística de los homicidios en sudamérica entre 1990 y 2020. *Revista Científica General José María Córdova*, 20(37), 187–205. Descargado de <https://doi.org/10.21830/19006586.852>
- Sánchez, R., Tejada, P., y Martínez, J. (2005). Comportamiento de las muertes violentas en bogotá, 1997-2003. *Revista de Salud Pública*, 7, 254–267. Descargado de <https://doi.org/10.1590/s0124-00642005000300002>
- Vizcaíno, M., y Duca, A. C. L. (2002). Homicidios: una mirada desde los actores. Descargado de <https://revistas.unal.edu.co/index.php/recs/article/view/11154/11820>
- Zambrano Ricaurte, D., y cols. (2022). Homicidios en colombia: un análisis comparativo entre sexos en el tiempo. Descargado de <http://hdl.handle.net/11371/4723>.