



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO
FACULTAD DE POSGRADO
MAESTRÍA EN CIENCIA DE DATOS

Proyecto de investigación previa
la obtención del Grado
Académico de Magíster en
Ciencia de Datos

TEMA

MODELO PREDICTIVO PARA DETERMINACIÓN DE HOMICIDIOS
INTENCIONALES EN ECUADOR, 2020 - 2024

AUTOR

ING. MANUEL FABRIZIO SARMIENTO FAJARDO

DIRECTOR

ING. EMILIO RODRIGO ZHUMA MERA M.Sc.

QUEVEDO – ECUADOR

2025

CERTIFICACIÓN

El Ing. Emilio Rodrigo Zhuma Mera M.Sc, en calidad de Director del proyecto de investigación, previo a la obtención del Grado Académico de Magíster en Ciencia de Datos.

CERTIFICA:

Que el **ING. MANUEL FABRIZZIO SARMIENTO FAJARDO**, ha cumplido con la elaboración del proyecto de investigación titulado **MODELO PREDICTIVO PARA DETERMINACIÓN DE HOMICIDIOS INTENCIONALES EN EL ECUADOR, 2020 - 2024** el mismo que se encuentra apto para su presentación y sustentación respectiva.

Quevedo, 22 de agosto, 2025.

Ing. Zhuma Mera Emilio, M.Sc.

Director

AUTORIA

Yo, **MANUEL FABRIZZIO SARMIENTO FAJARDO** con C.I. 1207468651 declaro que el proyecto de investigación con el tema **MODELO PREDICTIVO PARA DETERMINACIÓN DE HOMICIDIOS INTENCIONADOS EN EL ECUADOR, 2020 – 2024**, se ha desarrollado de manera íntegra, respetando derechos intelectuales de las personas que han desarrollado conceptos mediante las citas las cuales indican la autoría, y cuyos datos se detallan de manera más completa en la bibliografía. En virtud de esta declaración, me responsabilizo del contenido, autenticidad y alcance del presente proyecto de investigación.

Quevedo, 30 de junio, 2025.

Ing. Manuel Fabrizio Sarmiento Fajardo

Autor

DEDICATORIA

Agradezco a mi familia por el apoyo que siempre me brindaron en este proceso de aprendizaje. Por siempre aconsejarme a seguir adelante ante cualquier adversidad.

Manuel Fabrizzio Sarmiento Fajardo

AGRADECIMIENTO

Quiero agradecer a la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) por brindarme la posibilidad de ser beneficiario de esta beca. Esta oportunidad me ha permitido estudiar la Maestría en Ciencia de Datos en la Universidad Técnica Estatal de Quevedo (UTEQ), experiencia que ha beneficiado en expandir mi conocimiento académico y profesional de manera considerable.

Quiero expresar mi agradecimiento a la Universidad Técnica Estatal de Quevedo por darme la oportunidad de estudiar la Maestría en Ciencia de Datos. Su dedicación a la excelencia académica ha sido importante para mí crecimiento profesional. También quiero agradecer al Ing. Emilio Zhuma Mera M.Sc, quien ha sido mi guía en este proyecto de investigación.

A mi familia, por ser el apoyo fundamental en mi vida y en toda esta travesía. A mí papá, el Ing. Manuel Sarmiento Tenesaca, por su apoyo incondicional en todo momento, siempre dando consejos de forma positiva. A mi mamá, la Ing. Jessica Fajardo Romero, por apoyo incondicional todos los días en este proyecto y en la vida.

A mis amigos y compañeros, por compartir conmigo este viaje académico, donde siempre brindaron apoyo en momentos de dificultades sobre temas nuevos y por las palabras de aliento que siempre transmitían.

Ing. Manuel Fabrizzio Sarmiento Fajardo

PROLOGO

Cada línea en un conjunto de datos sobre homicidios representa una vida interrumpida, una historia truncada y una red de consecuencias que se extiende mucho más allá de las estadísticas. Es fácil perder de vista esta realidad humana cuando nos sumergimos en el mundo de los algoritmos, las métricas y los patrones. Sin embargo, es precisamente esa gravedad la que nos obliga a buscar, con todas las herramientas a nuestro alcance, una comprensión más profunda de un fenómeno que desgarrar este problema social. Este trabajo nació de esa convicción, la de que la ciencia de datos, puede ayudar con la optimización de problemas de seguridad o la tecnología de se implementa, tiene un papel crucial que desempeñar en el análisis de los problemas más urgentes y dolorosos de nuestra sociedad.

Es en este complejo y delicado contexto que se enmarca el trabajo investigativo del autor. Con rigor metodológico y una notable sensibilidad hacia la materia que aborda, no se conforma con describir el pasado, sino que se atreve a modelar su trayectoria. El enfoque de pronóstico que desarrolla no es un ejercicio de adivinación, sino una herramienta analítica disciplinada. Transforma el caos aparente de miles de incidentes en tendencias cuantificables y proyecta su evolución, ofreciendo una perspectiva informada sobre la magnitud y la persistencia de la crisis.

El autor nos proporciona una nueva lente, una basada en la evidencia y orientada al futuro, a través de la cual podemos mirar el problema y, quizás lo más importante, formular mejores preguntas.



**Paola Maribel Benitez
Navarrete**



Ing. Paola Benítez Navarrete, M.Sc.

RESUMEN

Esta investigación aborda la escalada de homicidios intencionales en Ecuador (2020-2024) mediante un enfoque de pronóstico de series temporales. El estudio transforma los datos de incidentes individuales en series mensuales que representan cuatro patrones clave: el volumen total de casos, el tipo de arma predominante, el perfil sociodemográfico de las víctimas y la concentración geográfica. Posteriormente, se desarrolla un marco de evaluación comparativa de tres modelos predictivos Prophet, SARIMA y Holt-Winters para seleccionar el algoritmo más preciso para cada patrón. El resultado principal es una proyección a cinco años (2025-2029) que sugiere la cronificación de la crisis en niveles históricamente elevados, ofreciendo una perspectiva cuantitativa y basada en datos sobre la trayectoria futura del fenómeno.

Palabras clave: Homicidios Intencionales, Series Temporales, Pronóstico Predictivo, Modelo SARIMA, Patrones de Violencia, Seguridad Ciudadana, Análisis de Tendencias.

ABSTRACT

This research addresses the escalation of intentional homicides in Ecuador (2020-2024) using a time-series forecasting approach. The study transforms individual incident data into monthly time series representing four key patterns: total case volume, predominant weapon type, sociodemographic profile of victims, and geographic concentration. A comparative evaluation framework of three predictive models Prophet, SARIMA, and Holt-Winters is then developed to select the most accurate algorithm for each pattern. The main output is a five-year projection (2025-2029) that suggests the chronicity of the crisis at historically high levels, offering a quantitative and data-driven perspective on the future trajectory of the phenomenon.

Keywords: Intentional Homicides, Time Series, Predictive Forecasting, SARIMA Model, Violence Patterns, Citizen Security, Trend Analysis.

ÍNDICE

CERTIFICACIÓN	II
AUTORIA	III
DEDICATORIA	IV
AGRADECIMIENTO	V
PROLOGO	VI
RESUMEN	VII
ABSTRACT	VIII
ÍNDICE	IX
INTRODUCCIÓN	1
CAPÍTULO I MARCO CONTEXTUAL DE LA INVESTIGACIÓN	4
1.1. UBICACIÓN Y CONTEXTUALIZACIÓN DE LA PROBLEMÁTICA	5
1.2. SITUACIÓN ACTUAL DE LA PROBLEMÁTICA	5
1.3. PROBLEMA DE LA INVESTIGACIÓN	6
1.3.1. Problema general	6
1.3.2. Problemas derivados	6
1.4. DELIMITACIÓN DEL PROBLEMA	7
1.5. OBJETIVOS	7
1.5.1. Objetivo General	7
1.5.2. Objetivos Específicos	7
1.6. JUSTIFICACIÓN	8
CAPÍTULO II MARCO TEÓRICO	10
2.1. FUNDAMENTACIÓN CONCEPTUAL	11
2.1.1. Homicidios Intencionales	11
2.1.2. Subcircuitos Policiales	11
2.1.3. Factores Sociodemográficos	12
2.1.4. Teorías de la violencia	12
2.1.6. Análisis Geoespacial	13
2.1.7. Aprendizaje automático	13
2.1.8. Python	14

2.1.9. Google Colab.....	14
2.1.10. Modelo SARIMA.....	15
2.1.11. Modelo Prophet	15
2.1.12. Modelo Holt-Winters.....	16
2.2. FUNDAMENTACIÓN REFERENCIAL	16
2.3. FUNDAMENTACIÓN LEGAL	21
2.3.1. Constitución del Ecuador (2008)	21
2.3.2. Código Orgánico Integral Penal (COIP, 2014)	22
2.3.3. Plan Nacional de Seguridad Ciudadana (2019-2030)	22
2.3.4. Asesinato.....	22
2.3.5. Femicidio	23
2.3.6. Agravantes del Femicidio.....	23
2.3.7. Sicariato.....	24
2.3.8. Homicidio	24
CAPÍTULO III METODOLOGÍA DE LA INVESTIGACIÓN.....	25
3.1. TIPOS DE INVESTIGACIÓN	26
3.2. MÉTODOS UTILIZADOS EN LA INVESTIGACIÓN	27
3.3. CONSTRUCCIÓN METODOLÓGICA DEL OBJETO DE INVESTIGACIÓN	28
3.3.1. Población y muestra	28
3.3.2. Técnicas de investigación	28
3.3.3. Instrumentos de la Investigación.....	29
3.3.4. Revisión de literatura	31
3.4. ELABORACIÓN DEL MARCO TEÓRICO.....	36
3.5. RECOLECCIÓN DE LA INFORMACIÓN	37
3.6. PROCESAMIENTO Y ANÁLISIS.....	37
3.6.1. Fase 1: Comprensión del Negocio	39
3.6.3. Fase 3: Preparación de los datos	39
3.6.4. Fase 4: Modelado.....	40
3.6.5. Fase 5: Evaluación.....	41
3.6.6. Fase 6: Despliegue.....	41
CAPÍTULO IV RESULTADOS Y DISCUSIÓN.....	42

4.1. IDENTIFICACIÓN DE PATRONES HISTÓRICOS MEDIANTE EL ANÁLISIS DE SUS TENDENCIAS Y CORRELACIONES.	43
4.1.2. Análisis Descriptivo Univariado	43
4.1.3. Análisis Bivariado: Exploración de Patrones y relaciones	45
4.1.4. Análisis de correlación entre patrones temporales	48
4.1.5. Discusión.....	50
4.2. ENTRENAMIENTO Y EVALUACIÓN COMPARATIVA DE MODELOS PARA EL PRONÓSTICO DE TENDENCIAS.....	51
4.2.1. Metodología de Construcción de los modelos.....	51
4.2.2. Resultados de la Evaluación Comparativa por Patrón	52
4.2.3. Resumen de la Selección de Modelos	57
4.2.4. Discusión.....	58
4.3. DESEMPEÑO DEL MODELO MEDIANTE UN ANÁLISIS DE RENDIMIENTO SEGÚN LAS MÉTRICAS DE EVALUACIÓN DE TENDENCIAS TEMPORALES.	59
4.3.1. Proyecciones de la Evolución de Patrones 2025-2029.....	59
4.3.2. Discusión.....	62
CAPITULO V CONCLUSIONES Y RECOMENDACIONES.....	64
5.1. CONCLUSIONES	65
5.2. RECOMENDACIONES	66
REFERENCIAS BIBLIOGRÁFICA.....	67
ANEXOS	71

ÍNDICE DE TABLAS

Tabla 1.....	31
Tabla 2.....	35
Tabla 4.....	43
Tabla 5.....	53
Tabla 6.....	54
Tabla 7.....	55
Tabla 8.....	57
Tabla 9.....	57

INDICE DE ILUSTRACIONES

Ilustración 1.....	38
Ilustración 3.....	45
Ilustración 4.....	46
Ilustración 5.....	47
Ilustración 6.....	48
Ilustración 7.....	52
Ilustración 8.....	53
Ilustración 9.....	55
Ilustración 10.....	56
Ilustración 11.....	59
Ilustración 12.....	60
Ilustración 13.....	61
Ilustración 14.....	62

INTRODUCCIÓN

El crimen y la violencia constituyen un grave desafío para el desarrollo de América Latina una región donde la tasa de homicidios per cápita es cuatro veces mayor que el promedio mundial (Cerdá et al., 2022). El homicidio es un problema muy serio dentro de la sociedad, puede verse reflejado por la desigualdad socioeconómica ya que este factor se asocia fuertemente con los delitos violentos. Otros factores como la pobreza y la discriminación son causantes del aumento de estas tasas (Silva & Lozano, 2022).

La Organización de Naciones Unidas (ONU) identificó que los determinantes para el aumento de homicidios son algunos factores como la renta per cápita, desigualdades sociales y económicas, acceso a armas de fuego, baja calidad educativa, ineffectividad del estado de derecho y gobernanza, altos niveles de criminalidad y la prevalencia de pandillas (J. Silva & Lozano, 2022). Ecuador en los últimos años no ha sido la excepción, enfrentando cambios en su tasa de homicidios intencionales, lo que destaca la necesidad de enfoques basados en datos para comprender y abordar este fenómeno.

Esta investigación tiene como objetivo principal crear un modelo que nos ayude a entender qué factores influyen más en los homicidios que ocurren en Ecuador. Durante el período 2020-2024, se buscó identificar las características demográficas, geográficas y circunstanciales que más se relacionan con estos crímenes violentos. El aprendizaje automático se utilizó para crear instrumentos de predicción. En concreto, el aprendizaje automático es una rama de la inteligencia artificial en la que el ordenador aprende patrones en los datos (Verrey et al., 2023).

Se utilizaron técnicas de aprendizaje automático para analizar los datos disponibles. En el Capítulo I de este estudio, se presenta el marco contextual de la investigación, definiendo el problema de los homicidios intencionales en Ecuador y su relevancia social. Además, se establecen los objetivos específicos del estudio y se justifica la delimitación temporal y espacial.

El Capítulo II se centra en el marco teórico, donde se revisan las principales teorías y modelos explicativos de la violencia homicida, así como los estudios empíricos más relevantes a nivel nacional e internacional. Se presta especial atención a los factores de riesgo identificados en la literatura para abordar este problema. El Capítulo III presenta la metodología de la investigación, que tipo de investigación es la que se realizó, los métodos utilizados en la investigación, así como el diseño de la investigación y la construcción metodológica que contiene la población y muestra de la investigación. Se justifica las técnicas que se usaron para realizar la investigación.

Los capítulos siguientes despliegan el núcleo de esta investigación. El Capítulo IV se dedica al análisis de los resultados y la determinación del desempeño de los modelos de pronóstico. Aquí se presenta la evaluación comparativa entre los enfoques algorítmicos y se selecciona el modelo más preciso, validando su capacidad para capturar las dinámicas temporales del fenómeno. Posteriormente, el Capítulo V presenta las conclusiones y recomendaciones emanadas de todo el proceso. En esta sección final, se sintetizan los hallazgos sobre las tendencias de la violencia, se discute el significado de la proyección a cinco años generada por el modelo final y se proponen líneas de acción y futuras investigaciones que podrían

derivarse de este análisis, buscando así traducir el conocimiento técnico en una perspectiva útil para el contexto de la seguridad en Ecuador.

CAPÍTULO I

MARCO CONTEXTUAL DE LA INVESTIGACIÓN

1.1. UBICACIÓN Y CONTEXTUALIZACIÓN DE LA PROBLEMÁTICA

El estudio se realizó recopilando información en Ecuador, país que se encuentra en América del Sur, caracterizado por sus climas diversos. Como en muchos otros países se encuentra enfrentando problemas con la inseguridad por culpa de la criminalidad y violencia. Ecuador ha experimentado un aumento del 574.30 % en el número de homicidios intencionales desde 2019 (Cordero y Rivera, 2023). Al final de 2023, el país reportó 8004 muertes violentas, lo que alcanza una tasa de 47.25 homicidios por cada cien mil habitantes (Cordero y Rivera, 2023).

Entre el 1 de enero y el 30 de junio de 2024, Ecuador registró un total de 3,036 homicidios, lo que representa una reducción de 585 casos en comparación con el mismo período de 2023. Este resultado marca la primera disminución en la tendencia observada desde 2019, con una reducción del 16.16% en relación con 2023. Sin embargo, esta cifra sigue siendo significativamente superior a la de 2022, cuando se registraron 2,128 homicidios en el mismo período (Lovo Karla & Olaya Linda, 2024). Este resultado sigue convirtiendo a Ecuador en uno de los países más violentos de América Latina y del mundo.

1.2. SITUACIÓN ACTUAL DE LA PROBLEMÁTICA

En Ecuador, el incremento de los homicidios intencionales se ha convertido en una preocupación central para la seguridad ciudadana, exigiendo estrategias fundamentadas en datos de alta calidad. La violencia y el crimen afectan no solo la percepción de seguridad, sino también el desarrollo económico y social del país, en este contexto, el uso de estadísticas confiables y certificadas por instituciones como el Instituto Nacional de Estadística y Censos

(INEC) resulta fundamental para comprender las dinámicas de este fenómeno y diseñar políticas públicas efectivas (Castillo et al., 2018).

Los homicidios intencionales en Ecuador han experimentado un incremento alarmante en los últimos años, situándose como uno de los principales indicadores de inseguridad en el país. Según informes recientes, la tasa de homicidios intencionados se concentra en regiones específicas, particularmente en zonas urbanas y áreas fronterizas, donde convergen factores como el tráfico ilícito de armas, disputas entre bandas delictivas y desigualdades socioeconómicas. Por cada hora que pasa en Ecuador, una persona muere a manos de un homicida (Montalvo, 2024).

1.3. PROBLEMA DE LA INVESTIGACIÓN

1.3.1. Problema general

¿Cómo un modelo de pronóstico mediante técnicas de aprendizaje automático permite el análisis de tendencias temporales de los patrones asociados a los homicidios intencionales en Ecuador entre 2020 y 2024?

1.3.2. Problemas derivados

- ¿Cuáles son los patrones históricos más relevantes en el conjunto de datos para explicar los homicidios intencionales?
- ¿Cómo la evolución de tendencias sociodemográficos, geográficos, volumen y tipos de armas están relacionadas con los homicidios intencionados en Ecuador?
- ¿Qué métricas de evaluación son las más adecuadas para medir el rendimiento del modelo?

1.4. DELIMITACIÓN DEL PROBLEMA

- **CAMPO:** Ciencia de datos aplicada a la seguridad ciudadana.
- **ÁREA:** Análisis predictivo.
- **LÍNEA:** Ciencia de Datos, Seguridad.
- **LUGAR:** Ecuador
- **TIEMPO:** 2020 y 2024.

1.5. OBJETIVOS

1.5.1. Objetivo General

Desarrollar un modelo de pronóstico aplicando técnicas de aprendizaje automático que permita el análisis de tendencias temporales de los patrones de homicidios intencionales en Ecuador entre 2020 y 2024.

1.5.2. Objetivos Específicos

- Identificar patrones históricos de homicidios mediante el análisis de sus tendencias y correlaciones temporales.
- Entrenar un modelo predictivo para el pronóstico de la evolución de tendencias sociodemográficos, geográficos, volumen y tipos de armas de los homicidios intencionados en Ecuador.
- Determinar el desempeño del modelo mediante un análisis de rendimiento según las métricas de evaluación de series temporales.

1.6. JUSTIFICACIÓN

Los modelos predictivos son herramientas diseñadas para anticipar eventos futuros mediante el análisis de datos históricos y actuales. Estos modelos utilizan algoritmos avanzados, como el aprendizaje automático, para identificar patrones y relaciones dentro de los datos, de esta manera mejorando la capacidad de toma de decisiones en diversas áreas. Su principal objetivo es ofrecer resultados precisos y generalizables que permitan prevenir o gestionar situaciones críticas, maximizando su impacto mediante métricas como la precisión, la sensibilidad y el área bajo la curva (Cranmer & Desmarais, 2016).

El Código Penal Alemán (StGB) categoriza los homicidios intencionados en "homicidio simple" y "asesinato", dependiendo de circunstancias objetivas y subjetivas del caso. Por ejemplo, el asesinato se caracteriza por la presencia de "motivos de base" o circunstancias específicas, como la crueldad, la obtención de gratificación sexual, o el uso de medios peligrosos (Walz et al., 2024).

Estas características incrementan el nivel de reprochabilidad e injusticia del acto, distinguiéndolo del homicidio simple. Además, se debe destacar cómo el género influye en las circunstancias de los homicidios, observando que los hombres suelen ser víctimas en lugares públicos por conocidos, mientras que las mujeres son frecuentemente asesinadas en sus hogares por sus parejas íntimas. Estos hallazgos subrayan la necesidad de desarrollar estrategias preventivas (Walz et al., 2024).

Este proyecto se justifica en la necesidad de aplicar metodologías más robustas, como las del aprendizaje automático, para desentrañar las complejas dinámicas que caracterizan a los homicidios intencionales. No se trata simplemente de contar casos, sino de entender los

patrones, para saber porque el perfil de un homicidio es estadísticamente separable del de un asesinato por encargo, o que la geografía no es solo un dato, sino un factor determinante en el tipo de violencia ejercida. La construcción de un modelo predictivo, en este contexto, no persigue un fin adivinatorio, sino que se convierte en una herramienta analítica para validar y cuantificar estas interacciones, ofreciendo una visión estructurada y basada en evidencia que, con suerte, puede informar un debate más profundo y, eventualmente, políticas públicas más focalizadas.

El estudio es de relevancia social ya que aborda un problema que impacta directamente en la seguridad y calidad de vida de la población ecuatoriana. Desde una perspectiva práctica, el modelo predictivo permitirá a cualquier persona o autoridad identificar el patrón por el cual hay un número elevado de homicidios intencionales, con esto cualquier persona podrá las autoridades identificar patrones donde hay más zonas de riesgos y los factores que requieren más atención o cuidado por parte de la ciudadanía.

CAPÍTULO II

MARCO TEÓRICO

2.1. FUNDAMENTACIÓN CONCEPTUAL

2.1.1. Homicidios Intencionales

El homicidio intencional se define como el acto deliberado de causar la muerte de otra persona, excluyendo muertes accidentales o involuntarias. Este fenómeno incluye una amplia variedad de contextos y motivos, que van desde crímenes relacionados con conflictos personales o domésticos hasta aquellos motivados por actividades delictivas organizadas. Según la Oficina de Naciones Unidas contra la Droga y el Delito (2022), los homicidios intencionales constituyen un indicador crítico para evaluar la violencia en las sociedades y son objeto de estudios interdisciplinarios que buscan comprender los factores de riesgo y las dinámicas subyacentes.

Los homicidios intencionales es el resultado de actos de violencia extrema, intencionados y premeditados que pueden estar vinculados con el crimen organizado, la violencia de género o conflictos intrafamiliares (UNODC, 2019).

El homicidio intencionado también puede llamarse así debido a la muerte provocada ilícita y voluntariamente a una persona por otra (Angela Me et al., 2011).

2.1.2. Subcircuitos Policiales

Los subcircuitos policiales son divisiones territoriales diseñadas para optimizar la respuesta policial y mejorar la prevención del delito en áreas específicas. Cada subcircuito corresponde a una región delimitada que facilita la gestión de recursos policiales, permitiendo un despliegue más eficiente de personal y estrategias de seguridad (Seguridad, 2024)

2.1.3. Factores Sociodemográficos

Según (Silva Urrutia et al. 2022), los factores sociodemográficos son variables que describen características sociales y económicas de una población. Estos factores incluyen elementos como edad, género, nivel educativo, estado civil, ocupación y lugar de residencia, entre otros. En el contexto de los homicidios, estos factores son esenciales para identificar patrones y tendencias que permitan analizar dinámicas de violencia y establecer intervenciones efectivas.

2.1.4. Teorías de la violencia

Según la investigación de (Silva Urrutia et al. 2022), menciona que se han desarrollado enfoques teóricos para explicar las causas y dinámicas de los homicidios. Destaca el impacto de factores como desigualdad económica, corrupción, debilidades en el sistema judicial y condiciones de impunidad. Estas teorías buscan entender la relación entre el contexto social, político y económico con los niveles de violencia en la sociedad.

2.1.5. Análisis Exploratorio de Datos (EDA)

El Análisis Exploratorio de Datos (EDA) es una fase fundamental e indispensable en cualquier proceso o pipeline de ciencia de datos (Manatkar et al., 2025). Su propósito principal es resumir y dar sentido a las características principales de los conjuntos de datos (D. H. da Silva et al., 2025). A través del EDA, se busca comprender los datos al descubrir patrones, anomalías e información crítica. Se podría que es una metodología para computar conjuntos de datos y sintetizar sus características fundamentales (Farooq et al., 2025).

2.1.6. Análisis Geoespacial

El análisis geoespacial es una técnica que integra datos espaciales y herramientas de visualización para identificar patrones, relaciones y tendencias en un contexto geográfico. Este tipo de análisis utiliza información sobre ubicaciones específicas, como coordenadas geográficas, para explorar cómo las variables espaciales afectan o están correlacionadas con eventos particulares (Cerdá et al., 2022)

El análisis geoespacial se utiliza para examinar la distribución espacial de servicios de salud, la accesibilidad geográfica y las disparidades en la cobertura. Utiliza herramientas y técnicas dentro de un marco de Sistemas de Información Geográfica (GIS) para mapear y analizar datos espaciales relacionados con servicios de salud (Abulibdeh et al., 2025)

2.1.7. Aprendizaje automático

El machine learning, o aprendizaje automático, es una rama fundamental de la inteligencia artificial que se centra en el desarrollo de algoritmos y sistemas capaces de aprender y mejorar su desempeño a partir de datos sin ser programados explícitamente para ello. Este aprendizaje se basa en la identificación de patrones y regularidades dentro de los datos, permitiendo que los modelos predictivos realicen tareas como clasificación, regresión y agrupamiento con un alto grado de precisión (Verrey et al., 2023).

El aprendizaje automático es una rama de la inteligencia artificial con aplicaciones significativas en diversas áreas, incluidas la educación y el análisis de datos (Forero-Corba & Bennasar, 2024).

Según (Carlos Vicente et al. 2022), define el aprendizaje automático como una subárea de la inteligencia artificial que busca que los ordenadores aprendan a partir de datos, utilizando algoritmos para identificar patrones y tomar decisiones con mínima intervención humana.

2.1.8. Python

Python se utiliza para construir sistemas modulares, automatizar procesos y facilitar la integración de algoritmos (Adams et al., 2010). Es reconocido como un lenguaje generalista altamente adaptable para computación científica debido a su capacidad de integrar bibliotecas en C y Fortran, y su énfasis en la rapidez de prototipado y facilidad de uso (Virtanen et al., 2020).

Sirve de base para bibliotecas más especializadas, como Scikit-learn y Scikit-image, usadas ampliamente en aprendizaje automático y procesamiento de imágenes. La flexibilidad de Python para desarrollar algoritmos rápidamente lo hace ideal para aplicaciones científicas, aunque el rendimiento computacional en tareas intensivas puede optimizarse mediante la integración con lenguajes compilados como C++ (Virtanen et al., 2020).

2.1.9. Google Colab

Google Colaboratory o Colab, es un entorno de trabajo gratuito en la nube que permite escribir y ejecutar código de Python en un simple navegador web sin requerir ninguna configuración adicional ni prerequisites, no se necesita instalar nada en nuestra computadora y todo lo que se produzca se puede almacenar en la nube (Edgar Hernández, 2022).

Colab está basado en Jupyter Notebook, por lo cual los notebooks o cuadernos de Colab permiten combinar código ejecutable y texto enriquecido en un único documento, junto con imágenes, HTML, LaTeX y muchos elementos más. Sin embargo, los notebooks que se creen

en Colab son almacenados en una cuenta de Google Drive, es decir, los notebooks de Colab son notebooks de Jupyter alojados en la nube. Además, Colab crea cuadernos en Python 3(o en Python 2) y el ambiente de trabajo ya viene con muchas librerías instaladas listas para utilizar, pero permite instalar otras (Edgar Moyotl Hernández, 2022).

2.1.10. Modelo SARIMA

El modelo SARIMA es una extensión del modelo ARIMA, se le conoce como ARIMA estacional, está específicamente diseñado para manejar patrones complejos en los datos de series temporales (M. M. U. Qureshi et al., 2025).

Es eficaz para la previsión de series temporales, capturando tanto las tendencias lineales como la estacionalidad en conjuntos de datos estables, su uso puede ayudar a eliminar los efectos estacionales o a mejorar la estructura del modelo para incorporar dichos patrones (Nourmohammad Ellham & Yousef Rasghidi, 2025).

2.1.11. Modelo Prophet

El modelo Prophet es una potente herramienta de pronóstico específicamente diseñada para datos de series temporales (M. M. U. Qureshi et al., 2025). Fue desarrollada por el equipo de Core Data Science de Facebook en 2017 y se destaca por su flexibilidad, precisión y facilidad de uso, además está basado en un marco de regresión aditiva (Scioşteanu Adriana & Criveanu, 2025).

Sobresale en la predicción precisa de datos de series temporales con tendencias, estacionalidad y festivos, es particularmente efectivo para manejar patrones estacionales marcados (Scioşteanu Adriana & Criveanu, 2025). Tiene la capacidad de capturar patrones

estacionales complejos, tendencias interanuales, variaciones temporales y pronósticos de estacionalidad a largo plazo.

2.1.12. Modelo Holt-Winters

El método de pronóstico Holt-Winters, también conocido como suavizado exponencial triple, es una técnica popular de pronóstico de series de tiempo que utiliza el suavizado exponencial para capturar tendencias y estacionalidad en los datos. Se considera uno de los métodos estadísticos tradicionales más maduros y ampliamente utilizados en el pronóstico de series de tiempo (Karamolegkos & Koulouriotis, 2025).

El modelo Holt-Winters implica el uso de tres ecuaciones de suavizado: una para el nivel, otra para la tendencia y una tercera para la estacionalidad, con el fin de producir un pronóstico (M. Qureshi et al., 2025).

2.2. FUNDAMENTACIÓN REFERENCIAL

El uso del aprendizaje automático ha revolucionado los métodos para predecir homicidios. Un ejemplo destacado es el artículo de Verrey et al. (2023), que utiliza un modelo de super aprendizaje para predecir homicidios domésticos a partir de datos policiales. Este modelo combina múltiples algoritmos para mejorar la precisión y el desempeño frente a herramientas tradicionales, alcanzando un AUC de 0.71. Además, el estudio demuestra cómo un enfoque basado en datos específicos puede ayudar a las fuerzas policiales a prevenir crímenes costosos y devastadores. De este modo este enfoque de modelo planteado refuerza que esta investigación puede ofrecer la ayuda a las fuerzas del orden al plantearle ideas de las zonas de riesgos que se encuentren.

Los factores contextuales y sociales son necesarios para determinar características que pueden influir en el proceso de vida de una población, según Cerdá et al. (2022), variables como el consumo de sustancias y la desigualdad económica están estrechamente vinculadas a la violencia armada, proporcionando un marco para evaluar factores de riesgo en homicidios intencionales. A diferencia de otras investigaciones que se centran en contextos específicos como violencia doméstica, este trabajo amplía el análisis a todos los homicidios intencionales, integrando técnicas de aprendizaje automático con un enfoque geográfico para abordar el contexto ecuatoriano.

La importancia del análisis geoespacial como el temporal es vital cuando se quiere comprender los datos asociados a una ubicación o también zona geográfica, y el análisis temporal ayuda a comprender patrones en los datos, así como hacer predicciones, en esto el estudio de Colasanti et al. (2023) utiliza series temporales ARIMA para analizar tendencias de homicidios-suicidios en Italia. Los resultados revelan patrones estacionales y la importancia de características sociodemográficas, como género, edad y relación entre las víctimas y perpetradores. Este enfoque permite identificar tendencias futuras, destacando la utilidad de modelos predictivos para la gestión territorial y la toma de decisiones estratégicas. En este sentido la metodología de análisis temporal y la consideración de características sociodemográficas enriquecen el enfoque geoespacial y temporal propuesto en este proyecto, ofreciendo una base sólida para explorar tendencias en homicidios intencionales.

Los factores sociodemográficos son importantes ya que con ellos se puede comprender y hacer un análisis de diferentes aspectos de la sociedad, la urbanización, la política, en esto el estudio de Colasanti et al. (2023), investiga homicidios relacionados con

violencia de pareja en Estados Unidos utilizando bases de datos nacionales. Este trabajo enfatiza la importancia de factores contextuales como género, relación entre las víctimas y los perpetradores, y la calidad de los datos para identificar patrones y desarrollar estrategias de intervención preventiva. Así este estudio refuerza la necesidad de incorporar variables relacionales y sociodemográficas en el modelo predictivo, asegurando un análisis más completo de las dinámicas que influyen en los homicidios intencionales y de este modo tener una comprensión amplia de estas características.

Los factores Sociodemográficos y su relación con los homicidios es un tema de mucho interés cuando se quiere conocer cómo es que afecta o que es lo que causan estos afectores y como repercuten directamente en los homicidios. Según el estudio de Silva Urrutia et al. (2022), los factores sociodemográficos, definidos como variables que describen las características sociales y económicas de una población, juegan un papel crucial en la comprensión de las dinámicas de violencia. Entre los principales factores se encuentran la edad, el género, el nivel educativo, el estado civil, la ocupación y el lugar de residencia. En el contexto de los homicidios, estos elementos permiten identificar patrones y tendencias relevantes para diseñar intervenciones efectivas. Además, destaca cómo desigualdades socioeconómicas y condiciones de marginalización están estrechamente vinculadas con tasas elevadas de violencia, lo que subraya la importancia de incorporar estos factores en modelos predictivos. Con esta información podemos tener de referencia como en otros países estos problemas afectan para que se den los casos de homicidios.

Las teorías son de gran ayuda cuando se quiere deducir algo, en este caso las teorías de la violencia que menciona Silva Urrutia et al. (2022), nos dice que las teorías sobre la violencia buscan explicar las causas y dinámicas de los homicidios mediante el análisis de

factores contextuales y estructurales, donde se tienen elementos como la desigualdad económica es la relación entre pobreza, inequidad y violencia ha sido ampliamente documentada. Estudios indican que regiones con mayor desigualdad tienden a experimentar niveles más altos de homicidios.

También Silva Urrutia et al. (2022), menciona que las debilidades en el sistema judicial dan como resultado la corrupción, impunidad y la falta de control efectivo del crimen son factores que contribuyen significativamente al aumento de los homicidios. Este factor es importante al momento de tomar en cuenta que se pueden llevar a cabo homicidios intencionales por personas que ya han cometido este tipo de delito y vuelven a quedar en libertad, lo que conlleva que puede que cometan otra vez este tipo de actos. También mencionaba que otra teoría podría ser el impacto del contexto político y social, ya que tiene cambios en el tipo de violencia y participación de diferentes actores, como bandas criminales, también influyen en las tasas de homicidio.

El uso de modelos predictivos puede ser clave para comprender las dinámicas locales y diseñar estrategias adaptadas a las necesidades específicas del país con esto en cuenta tenemos lo que nos dice Silva Urrutia et al. (2022) que el uso de métodos de aprendizaje automático, como el modelo C5.0, se ha convertido en una herramienta fundamental para el análisis de homicidios. Ya que puede identificar factores clave como determinar qué variables tienen mayor impacto en la ocurrencia de homicidios. Otra es la de optimizar recursos ya que proveer información precisa para la asignación de recursos en seguridad ciudadana es esencial. Y por último la de prever tendencias ya que anticipar posibles aumentos o disminuciones en las tasas de homicidios mediante el análisis de patrones históricos puede dar muchos beneficios.

El estudio propuesto de Forradellas et al. (2021), realizado en la ciudad de Buenos Aires donde utiliza el modelo SEMMA (Sample, Explore, Modify, Model, and Assess) para manipular datos, realizar clustering con K-means y luego entrenar la red neuronal, el propone un modelo predictivo de homicidios basado en redes neuronales. Utiliza las variables predictoras incluyen características temporales y geográficas, como año, mes, día, franja horaria y comuna. Este enfoque permite identificar patrones espaciales y temporales, mejorando la asignación de recursos policiales y optimizando medidas preventivas.

El artículo de (Delgado & Sánchez-Delgado, 2024) nos introduce en modelo predictivo basado en redes Bayesianas para perfilar a los perpetradores de homicidios múltiples. Se utiliza aprendizaje multi-instancia (MI), donde las víctimas de un mismo incidente forman un grupo asociado a una etiqueta común. Este enfoque es innovador al combinar aprendizaje supervisado con técnicas de agrupamiento, permitiendo predicciones precisas a partir de características de las víctimas y del crimen. El modelo se aplica a una base de datos del FBI y destaca por su capacidad de generar perfiles útiles para investigaciones criminales y políticas públicas.

El trabajo de Villegas et al. (2022), aborda el problema de homicidios usando una esparcida distribución espacio-temporal. Donde se comparan modelos simples, como el modelo de conteo estático, con modelos más complejos que incorporan aprendizaje por grafos y procesamiento de señales. Se demuestra que los modelos avanzados superan a los simples en alta resolución, especialmente cuando se integran datos complementarios como peleas callejeras. El estudio destaca la importancia de optimizar la resolución espacial y el uso de información contextual para mejorar las predicciones.

Según Silva Urrutia et al. (2022), comenta que la influencia de las políticas públicas en la reducción de homicidios ha sido documentada en estudios internacionales. Factores como el Índice de Estado de Derecho (IED) destacan por su impacto directo en las tasas de violencia. La mejora de este índice mediante reformas legales y fortalecimiento institucional podría resultar en una disminución significativa de los homicidios.

El uso de herramientas para identificar patrones es muy esencial en el momento que queremos conocer tendencias que puedan ayudarnos en diferentes ámbitos. El autor (Silva Urrutia et al., 2022) nos dice que el uso de herramientas de análisis geoespacial permite identificar patrones territoriales y zonas críticas de violencia. Y en el caso de este proyecto con la ayuda de los análisis geoespaciales podemos localizar zonas críticas lo que nos ayudará a identificar regiones específicas con altas tasas de homicidios. Así como un diseño de estrategias localizadas puede implementar intervenciones dirigidas a las áreas más afectadas. O en su caso una visualización de datos puede ayudarnos mediante la creación de mapas que apoyen la toma de decisiones en seguridad ciudadana.

2.3. FUNDAMENTACIÓN LEGAL

2.3.1. Constitución del Ecuador (2008)

El Artículo 393 de la Constitución del Ecuador asegura el derecho de los ciudadanos a la seguridad integral. Este derecho abarca no solo la prevención de crímenes como los homicidios intencionales, sino también la promoción de políticas públicas que garanticen la seguridad física y emocional de la población. La Constitución establece que el Estado tiene la obligación de garantizar un entorno libre de violencia y de proteger el derecho fundamental a la vida (Constitución de la República del Ecuador, 2008).

2.3.2. Código Orgánico Integral Penal (COIP, 2014)

El COIP tipifica los homicidios intencionales en su Artículo 140, diferenciando entre homicidio y asesinato. Este último se caracteriza por circunstancias agravantes como premeditación, alevosía, o la ventaja del perpetrador. El COIP establece penas severas para quienes cometan homicidios intencionales, buscando no solo castigar sino también disuadir este tipo de actos violentos (Asamblea Nacional, 2014).

2.3.3. Plan Nacional de Seguridad Ciudadana (2019-2030)

El Plan Nacional de Seguridad Ciudadana prioriza la reducción de los homicidios intencionales mediante estrategias que incluyen el fortalecimiento de la policía, la optimización de recursos y la implementación de tecnologías de análisis de datos para la prevención del delito. Este plan busca enfrentar las causas estructurales de la violencia y mejorar la percepción de seguridad en la población (Angulo et al., 2019).

2.3.4. Asesinato

Art. 140.- Asesinato. - La persona que mate a otra será sancionada con pena privativa de libertad de veintidós a veintiséis años, si concurre alguna de las siguientes circunstancias:

1. A sabiendas, la persona infractora ha dado muerte a su ascendiente, descendiente, cónyuge, conviviente, hermana o hermano.
2. Colocar a la víctima en situación de indefensión, inferioridad o aprovecharse de esta situación.
3. Por medio de inundación, envenenamiento, incendio o cualquier otro medio se pone en peligro la vida o la salud de otras personas
4. Buscar con dicho propósito, la noche o el despoblado.
5. Utilizar medio o medios capaces de causar grandes estragos.

6. Aumentar deliberada e inhumanamente el dolor a la víctima.
7. Preparar, facilitar, consumir u ocultar otra infracción.
8. Asegurar los resultados o impunidad de otra infracción.
9. Si la muerte se produce durante concentraciones masivas, tumulto, conmoción popular, evento deportivo o calamidad pública.
10. Perpetrar el acto en contra de una o un dignatario o candidato a elección popular, elementos de las Fuerzas Armadas o la Policía Nacional, fiscales, jueces o miembros de la Función Judicial por asuntos relacionados con sus funciones o testigo protegido.

2.3.5. Femicidio

Art. 141.- Femicidio. - La persona que, como resultado de relaciones de poder manifestadas en cualquier tipo de violencia, dé muerte a una mujer por el hecho de serlo o por su condición de género, será sancionada con pena privativa de libertad de veintidós a veintiséis años.

2.3.6. Agravantes del Femicidio

Art. 142.- Circunstancias agravantes del femicidio. - Cuando concurren una o más de las siguientes circunstancias se impondrá el máximo de la pena prevista en el artículo anterior:

1. Haber pretendido establecer o restablecer una relación de pareja o de intimidad con la víctima.
2. Exista o haya existido entre el sujeto activo y la víctima relaciones familiares, conyugales, convivencia, intimidad, noviazgo, amistad, compañerismo, laborales, escolares o cualquier otra que implique confianza, subordinación o superioridad.
3. Si el delito se comete en presencia de hijas, hijos o cualquier otro familiar de la víctima.
4. El cuerpo de la víctima sea expuesto o arrojado en un lugar público.

2.3.7. Sicariato

Sicariato. - La persona que mate a otra por precio, pago, recompensa, promesa remuneratoria u otra forma de beneficio, para sí o un tercero, será sancionada con pena privativa de libertad de veintidós a veintiséis años. La misma pena será aplicable a la persona, que en forma directa o por intermediación, encargue u ordene el cometimiento de este ilícito. Se entenderá que la infracción fue cometida en territorio y jurisdicción ecuatorianos cuando los actos de preparación, organización y planificación, sean realizados en el Ecuador, aun cuando su ejecución se consume en territorio de otro Estado.

La sola publicidad u oferta de sicariato será sancionada con pena privativa de libertad de cinco a siete años.

2.3.8. Homicidio

Art. 144.- Homicidio. - La persona que mate a otra será sancionada con pena privativa de libertad de diez a trece años.

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

3.1. TIPOS DE INVESTIGACIÓN

Para comenzar con la investigación sobre homicidios intencionales en Ecuador, se empleó una combinación de enfoques que abarcan distintos tipos de investigación. Cada uno contribuirá desde su perspectiva a una comprensión más integral del fenómeno.

Investigación Documental

La investigación se inició con una revisión de la literatura, lo que implica un análisis exhaustivo de los estudios previos sobre técnicas de ciencia de datos aplicadas a la investigación de homicidios. Este tipo de investigación es pertinente para fundamentar el modelo de pronóstico e identificar las técnicas a utilizar.

Investigación Descriptiva

La presente investigación adopta un enfoque descriptivo, ya que se orienta a caracterizar, analizar y clasificar los homicidios intencionales ocurridos en Ecuador entre 2020 y 2024, con base en variables sociodemográficas, geográficas, temporales y asociadas al tipo de armas utilizadas. Este enfoque permite comprender de forma sistemática cómo, dónde y bajo qué condiciones se producen los homicidios, proporcionando una base sólida para el posterior desarrollo del modelo predictivo.

Investigación Correlacional

El análisis de las relaciones entre factores contextuales y temporales asociados con los homicidios implica explorar relaciones entre variables. Este enfoque ayudará a entender cómo ciertas condiciones afectan la ocurrencia de homicidios, lo que es parte del análisis predictivo.

Investigación Cuantitativa

Debido al uso de modelos predictivos basados en datos, el análisis será predominantemente cuantitativo, ya que se utilizarán estadísticas y técnicas de aprendizaje máquina. Este tipo de investigación se debe integrar, ya que los modelos predictivos requieren el análisis de grandes volúmenes de datos numéricos.

3.2. MÉTODOS UTILIZADOS EN LA INVESTIGACIÓN

Método Deductivo

El uso del método deductivo es adecuado porque se parte de teorías y enfoques previos sobre la criminología, el análisis de homicidios y las técnicas de ciencia de datos. Esto permitirá aplicar modelos y enfoques ya probados en el contexto ecuatoriano.

Método Inductivo

El análisis inductivo es necesario para identificar patrones emergentes de los datos específicos de homicidios en Ecuador, como la identificación de grupos vulnerables o las correlaciones no predecibles inicialmente. Este enfoque es apropiado para descubrir nuevas relaciones a partir de los datos.

Método Cuantitativo

Dado que el estudio incluye la creación de un modelo predictivo utilizando técnicas de ciencia de datos como aprendizaje automático, el método cuantitativo será esencial para analizar y modelar los datos de manera precisa. Si bien el deductivo e inductivo son fundamentales, es recomendable señalar también el uso explícito de este método en el análisis de grandes volúmenes de datos.

3.3. CONSTRUCCIÓN METODOLÓGICA DEL OBJETO DE INVESTIGACIÓN

3.3.1. Población y muestra

Población

La población de estudio se define como el universo total de eventos de homicidios intencionales registrados oficialmente en la República del Ecuador durante el período comprendido entre el 1 de enero de 2020 y el 31 de diciembre de 2024. Se trata, por tanto, de una población finita y delimitada de incidentes, no de individuos.

Muestra

La muestra utilizada en esta investigación corresponde al conjunto de datos mdi_homicidiosintencionales_pm_2013-2024, obtenido del repositorio de Datos Abiertos Ecuador. Dicho conjunto contiene los registros históricos de los eventos de la población de estudio. Dado que se trabajó con la totalidad de los registros disponibles para el período definido, la muestra se considera un censo de los casos reportados, y no una muestra estadística en el sentido tradicional. Se utilizó los datos de 2020 a 2024, porque el cambio drástico de homicidios, inseguridad, se observó desde 2019, y para esta investigación se usó desde ese año. El acceso y la descarga de los datos se realizaron a través de la URL oficial del portal.

URL: [Homicidios Intencionales - Conjunto de datos - Datos Abiertos Ecuador](#)

3.3.2. Técnicas de investigación

En virtud a la naturaleza del conjunto de datos utilizado, las técnicas empleadas son las siguientes:

Revisión Documental y Bibliográfica: Se realizó una revisión sistemática de literatura para fundamentar el marco teórico y seleccionar las metodologías de análisis de series temporales más pertinentes.

Análisis Estadístico Descriptivo: Se utilizó para caracterizar las propiedades de las series temporales (media, desviación estándar, etc.) y visualizar sus tendencias.

Análisis de Correlación: Se aplicó para cuantificar la relación entre la evolución de los diferentes patrones temporales.

Modelado Predictivo (Pronóstico): Fue la técnica central, mediante la cual se entrenaron y evaluaron algoritmos de series temporales para proyectar el comportamiento futuro de los patrones.

3.3.3. Instrumentos de la Investigación

Para la ejecución del proyecto se utilizó un conjunto estandarizado de herramientas de software, lenguajes y bibliotecas especializadas. A continuación, se detalla el ecosistema tecnológico empleado:

Entorno de desarrollo

- **Google Colab:** Se seleccionó como el principal entorno de desarrollo. Esta plataforma proporcionó un ambiente de notebook interactivo basado en la nube, equipado con Python y las bibliotecas científicas preinstaladas, lo que facilitó la reproducibilidad del análisis y la ejecución del código sin requerir configuración local.

Herramientas para el tratamiento de los datos

- **Microsoft Excel:** Se empleó para la inspección inicial, la validación y el preprocesamiento manual de los datos. Permitió corregir formatos y realizar revisiones preliminares antes de la carga en el entorno de análisis.
- **Python:** Fue el lenguaje de programación central para todo el proceso analítico, desde la limpieza de datos hasta el entrenamiento de los modelos de pronóstico. Se eligió por su robusto ecosistema de bibliotecas especializadas en ciencia de datos y aprendizaje automático.

Bibliotecas de Python Utilizadas

El análisis se fundamentó en un conjunto de bibliotecas clave, cada una con un propósito específico:

Para manipulación y cálculo:

- **Pandas:** Fue la herramienta fundamental para la carga, manipulación, limpieza y agregación de los datos, permitiendo estructurarlos en DataFrames para su posterior análisis.
- **Numpy:** Se utilizó para realizar operaciones numéricas eficientes y cálculos matemáticos sobre las estructuras de datos.

Para Visualización de datos:

- **Matplotlib:** Se uso para generar gráficos y personalizarlos durante el análisis, así como gráficos de barras.
- **Seaborn:** Se uso para complementar a matplotlib para crear gráficos estadísticos más detallados y estéticos. Además, de la creación de mapas de calor para la correlación de variables.

Para modelado de Series temporales

- **Statsmodels:** Proporcionó las implementaciones para los modelos estadísticos clásicos, incluyendo SARIMA y Holt-Winters (Exponential Smoothing), así como herramientas para el análisis de series temporales.
- **Prophet:** Se empleó esta biblioteca desarrollada por Facebook por su capacidad para modelar series temporales de manera robusta, manejando automáticamente tendencias y estacionalidades.
- **Scitik-learn:** Aunque el enfoque final fue de pronóstico, esta biblioteca fue instrumental para las métricas de evaluación de error, como el Error Absoluto Medio (MAE) y la Raíz del Error Cuadrático Medio (RMSE).

3.3.4. Revisión de literatura

Se realizó una revisión sistemática de literatura, que consta de una cadena de búsqueda con un total de más de 1000 resultados encontrados en las diferentes bases de datos bibliográficas utilizadas en la investigación, donde se escogieron los estudios que tenían temas similares en base al tema de esta investigación que es sobre modelos predictivos y homicidios, o temas que tuvieran similitud para analizar y comprender los diferentes métodos de ciencias de datos que se han empleado en estos temas como las siguientes 10 investigaciones más relevantes como se muestra en la siguiente tabla.

Tabla 1

Revisión Sistemática de Literatura

Nombre del Documento	Autor	Año	País	Tipo de estudio	Técnicas utilizadas	Hallazgos principales
An Empirical Analysis of Machine	Lawrence McClendon ,	2021	India	Análisis empírico de	Regresión lineal, regresión	Este estudio encontró que el algoritmo

Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach	Natarajan Meghanathan			aprendizaje automático	aditiva, árboles de decisión, SVM, Redes neuronales, Random Forest.	de regresión lineal es muy efectivo y previco para predecir datos de delitos
Artificial Intelligence in Crime Prediction: A Survey With a Focus on Explainability	Filiz Ersöz, Taner Ersöz, Francesco Marcelloni, y Fabrizio Ruffini	2025	Turkiye e Italia	Encuesta sistemática / revisión de literatura centrada en la Inteligencia Artificial en la predicción de delitos con énfasis en la Explicabilidad (XAI)	SVM, teoría difusa, redes neuronales artificiales, regresión lineal, regresión aditiva, árboles de decisión	Este estudio encontró que los métodos post-hoc explican las decisiones de modelos opacos como ensambles, SVM, y redes neuronales
Comparison of Machine Learning Algorithms for Predicting Crime Hotspots	Lin Liu	2020	China	Comparación de algoritmos de aprendizaje automatico para la predicción de hotspots de delitos	Datos históricos de casos y datos covariado, aprendizaje automático	Este trabajo busca predecir los puntos calientes de delitos.
Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions	Varun Mandalapu y Lavanya Elluri	2023	Estados Unidos	Revisión sistemática centrada en encontrar algoritmos eficientes para predecir delitos vecinales utilizando aprendizaje automático y aprendizaje profundo	Revisa enfoques de ML y DL, Clustering, procesamiento del lenguaje natural.	Este trabajo mostros las tendencias de delitos vecinales mostraron un movimiento ascendente de 2018 a 2021, con un pico en 2020/2021 (posiblement e debido a COVID-19), y luego

						disminuyeron en 2022
Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques	Wajiha Safat	2021	Pakistan	LSTM, EDA, Modelos ARIMA	Algoritmos de aprendizaje automático, Aprendizaje Profundo, ARIMA.	Este estudio reporta una eficiencia y precisión mejoradas para la predicción de delitos usando algoritmos de ML
Offender Characterization and Prediction A Case Study of the Kingdom of Bahrain	Ebrahim Jaffar Alasfoor	2024	Reino de Baréin	Estudio de caso que analiza datos de delitos de Baréin para demostrar un marco integrado que utiliza análisis exploratorio y aprendizaje automático para el modelado predictivo de características del delincuente	Minería de datos, Análisis de Datos Exploratorio (EDA), Algoritmos de aprendizaje supervisado.	Este estudio reveló que los patrones como robo y asalto son delitos predominantes centrados en diferentes áreas, además de encontrar cual es la hora de mayor frecuencia de delitos.
Spatio-Temporal Crime HotSpot Detection and Prediction: A Systematic Literature Review	Umair Muneer Butt, Sukumar Letchmunan, Fadratul Hafinaz Hassan, Mubashir Ali	2020	Varios países afiliados en Malasia, Italia, Pakistán e Irlanda	Revisión Sistemática de Literatura (SLR). Enfoque en técnicas de detección y predicción de puntos críticos (hotspots) de delitos espacio-temporales. Identifica desafíos y	Minería de datos, aprendizaje automático, Clustering.	Discute la utilidad del análisis de series de tiempo y el aprendizaje profundo. Enfatiza la importancia de incluir información espacial y temporal en los conjuntos de datos para mayor

				direcciones futuras.		precisión y fiabilidad
Smart Policing Technique With Crime Type and Risk Score Prediction Based on Machine Learning for Early Awareness of Risk Situation	Yong-Tae Lee	202 1	Korea del Sur	Diseño y desarrollo de una técnica de predicción del tipo de delito y nivel de riesgo. Verificación del rendimiento. Desarrollo de una plataforma de software con GUI.	Aprendizaje Automático, Modelo de predicción basado en DNN y CNN. Extracción de palabras mediante WordRank y TF-IDF. Algoritmo Naive Bayes y SVM.	Este estudio desarrolló una técnica para predecir el tipo de delito y el nivel de riesgo a partir de resúmenes de casos penales basados en texto y una fórmula para calcular y generar una puntuación de riesgo basada en la gravedad.

Nota: Se pusieron a los trabajos que tuvieran más relación con el tema investigado, había muchos que no tenían relación.

La información obtenida de los resultados de búsqueda realizados en las diferentes bases de datos bibliográficas como se muestra en la Tabla 1, determina que los diferentes trabajos exploran la aplicación de técnicas de aprendizaje automático y aprendizaje profundo para la predicción de casos y el análisis de delitos. También se examinan diferentes algoritmos, como el árbol de decisión, máquinas de soporte vectorial (SVM), Naive bayes, Random Forest y redes neuronales en las que están incluidas CNN y LSTM, donde se evalúa su precisión y rendimiento en conjuntos de datos de diferentes regiones. Además, se enfatiza que los datos espacio-temporales tienen mucha importancia, así como las características contextuales que son los puntos de interés y densidad de las carreteras, y con esto se puede mejorar las predicciones. Algunas investigaciones abordan que hay desafíos en la confianza de los

modelos de inteligencia artificial (IA) para la aplicación de la ley, junto con la limitación del tamaño y la calidad de los conjuntos de datos que se dispongan.

La investigación realizada por Kshatri et al. (2021) nos evidencia que el uso de diferentes métodos como el conjunto de redes neuronales artificiales, ANN-bagging, SVM, Random Forest y técnicas de regresión son los mejores al momento de elegir técnicas para comparar el potencial de una predicción confiable, ya que su modelo propuesto tiene una tasa de crecimiento alta. Además, el modelo propuesto por ellos para determinar crímenes también se podría utilizar en otras áreas como la caracterización de yacimientos.

La investigación realizada por Ersoz et al. (2025) nos evidencia que las técnicas como DBSCAN y RF son las técnicas más precisas para la detección de hotspots, en cambio la técnica ARIMA es mejor en la predicción de series de tiempo de crímenes. Además, mediante las investigaciones que realizo en el año 2022 encontró que el aprendizaje supervisado fue el método más aplicado con el robo como el tipo de crimen más estudiado, y RF Naive Bias como los algoritmos más utilizados. Menciona que los modelos basados en CNN a menudo superan los enfoques tradicionales de ML en la predicción de crímenes, también reporta que la Random Forest es la técnica comúnmente más utilizada en los estudios, así como el mejor rendimiento de predicción. Por último, las técnicas que permiten justificar las decisiones del modelo como SHAP, LINE, feature importance, son empleadas para que las decisiones de los modelos de inteligencia artificial sean más fáciles de interpretar.

A continuación, se presenta la tabla de búsqueda realizada en las páginas de base de datos:

Tabla 2

Cuadro explicativo con las cadenas de búsquedas realizadas

Tema	Base de Datos	Cadena de Búsqueda	Total de búsquedas
Modelos predictivos	Scopus	"Predictive model" AND ("crime" OR "homicide") AND ("machine learning" OR "data mining")	631
	IEEE Xplore	("predictive model" AND ("homicide" OR "violent crime")) AND ("machine learning" OR "classification")	29
	SpringerLink	"Predictive analytics" AND ("intentional homicide" OR "crime prediction") AND ("machine learning")	65
Homicidios intencionales	Scopus	"Intentional homicide" AND ("data analysis" OR "crime patterns")	734
	IEEE Xplore	("intentional homicide" OR "criminal violence") AND ("spatiotemporal analysis" OR "data mining")	238
	SpringerLink	"intentional homicide" AND ("data analysis" OR "risk factors" OR "crime modeling")	104
Modelos de tendencias	Scopus	"trend analysis" AND ("homicide rates" OR "crime statistics") AND ("time series" OR "forecasting")	676
	IEEE Xplore	("trend modeling" OR "temporal prediction") AND ("homicide" OR "crime") AND ("time series" OR "ARIMA" OR "LSTM")	205
	SpringerLink	"Trend prediction" AND ("homicide" OR "violent crime") AND ("time series analysis" OR "forecast models")	4

Nota: En la búsqueda realizada se aplicó un filtro solo Open Access Only, documentos gratis.

3.4. ELABORACIÓN DEL MARCO TEÓRICO

Para la elaboración se comenzó con la revisión de literatura de artículos, revistas que se sacaron de Scopus, ScienceDirect, IEEE Xplore, SpringerLink. Donde la información obtenida de la literatura solo fue seleccionada la más relevante, que estuviera relacionada con modelos predictivos enfocados en homicidios, inseguridad, robo. Después, se procede a contextualizar la información donde se establecen los fundamentos necesarios para analizar patrones de violencia en Ecuador entre 2020 y 2024, considerando teorías como la de la Ventana Rota y modelos predictivos basados en aprendizaje automático. Además, se aborda la importancia de factores contextuales, temporales y sociodemográficos en la ocurrencia de homicidios, con énfasis en las características propias del territorio ecuatoriano.

El objetivo principal es proporcionar una base conceptual y teórica para el desarrollo de un modelo predictivo que permita identificar factores de riesgo y zonas críticas de violencia. Esto se logra mediante la síntesis de teorías relevantes, la contextualización al ámbito local y la validación empírica de los hallazgos. La estructura del marco incluye fundamentación conceptual, referencial y legal, asegurando una visión integral del fenómeno.

Este tema aborda un problema de gran relevancia social como el incremento de homicidios intencionales en Ecuador. Mediante un enfoque basado en datos, la investigación busca contribuir a la comprensión de este fenómeno y ofrecer datos que puedan acceder las autoridades o la ciudadanía para que puedan tener conocimiento de los sucesos de los homicidios.

3.5. RECOLECCIÓN DE LA INFORMACIÓN

El conjunto de datos que se utilizó fue preparado a partir de dos bases de datos, la primera base de datos fue desde el año 2014 a 2023, la segunda base de datos fue desde el inicio del año 2024 a finales del mismo año 2024. Los datos se obtuvieron del portal de datos abiertos del Ecuador, en la categoría Administración Pública. Estos datos constan de 348 456 registros. No se aplicaron encuestas, entrevistas u observaciones directas, ya que el diseño del estudio es de carácter documental.

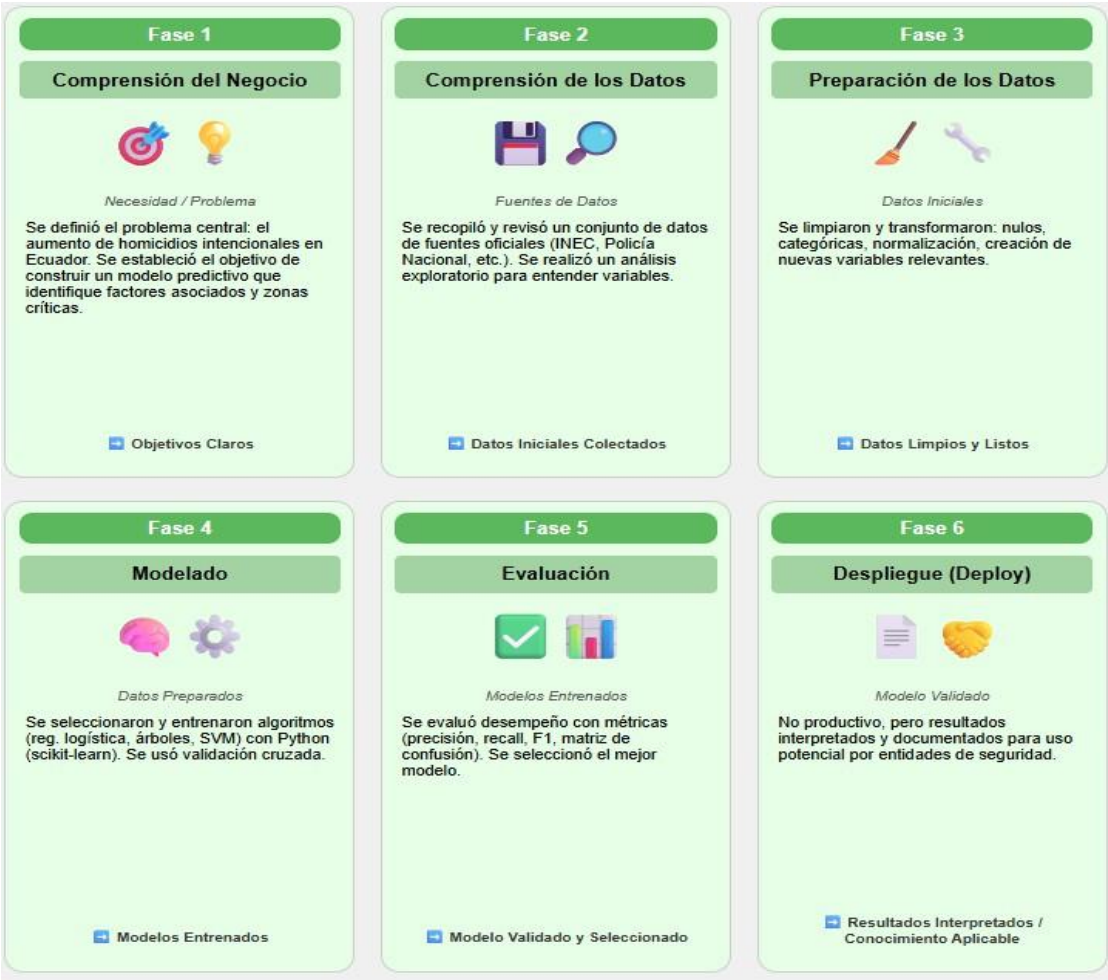
3.6. PROCESAMIENTO Y ANÁLISIS

Este proyecto se desarrolló con base en la metodología siguiendo la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), un enfoque ampliamente adoptado en el ámbito de la minería de datos que permite estructurar y organizar de manera eficiente las fases de un proyecto analítico. Esta metodología comprende seis etapas: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y

despliegue. Cada una de estas fases fue aplicada en el contexto del análisis de homicidios intencionales en Ecuador entre 2020 y 2024. La metodología permitió avanzar de forma sistemática desde el planteamiento del problema hasta la obtención de resultados confiables. A continuación, se presenta un resumen de cómo se aplicaron cada una de las etapas de CRISP-DM en la siguiente ilustración 1.

Ilustración 1

Etapas de la Metodología de CRISP-DM



Nota: Esta table trata los procesos de la metodología CRIPS-DM usado en esta investigación

A continuación, se detalla la ejecución de cada fase:

3.6.1. Fase 1: Comprensión del Negocio

En esta fase inicial, se definió el problema central de la investigación: la necesidad de comprender y proyectar la escalada de homicidios intencionales en Ecuador. El objetivo principal se estableció como el desarrollo de un modelo de pronóstico que permitiera analizar la evolución futura de los patrones clave (volumen, tipo de arma, sociodemográfico y geográfico) asociados a este fenómeno.

3.6.2. Fase 2: Comprensión de los datos

Esta etapa se centró en la familiarización con el conjunto de datos. Se obtuvo el dataset oficial del repositorio de Datos Abiertos Ecuador y se realizó un Análisis Exploratorio de Datos (EDA) preliminar para entender la estructura, el contenido y la calidad de las variables disponibles. El propósito fue identificar las características más relevantes y formular las primeras hipótesis sobre las tendencias temporales presentes en los datos.

3.6.3. Fase 3: Preparación de los datos

Esta fue una de las fases más críticas, donde los datos crudos se transformaron en un formato adecuado para el modelado de series temporales. El proceso incluyó varias subetapas:

- **Diagnóstico y Manejo de Valores Faltantes:** Se realizó un análisis exhaustivo para cuantificar los valores nulos en cada columna (ver tabla 3). Se identificó que variables sociodemográficas como instrucción y profesion_registro_civil presentaban un alto porcentaje de datos ausentes, lo que justificó su exclusión de los análisis posteriores.
- **Limpieza y Transformación de Datos:** Se estandarizaron los tipos de datos, convirtiendo la columna fecha_infraccion al formato datetime para posibilitar el

análisis temporal. La columna hora_infraccion, que presentaba formatos mixtos, fue procesada para obtener un valor numérico uniforme.

- **Transformación a Series Temporales:** El paso fundamental en esta fase fue la **agregación de los datos**. Los registros de incidentes individuales se agruparon con una frecuencia mensual para construir las cuatro series temporales que representan los patrones de interés: Volumen, Tipo de Arma, Sociodemográfico y Geográfico.

3.6.4. Fase 4: Modelado

En esta fase se seleccionaron los modelos predictivos:

Selección de Algoritmos

Se seleccionaron tres modelos de pronóstico con enfoques distintos: Prophet, SARIMA y Holt-Winters. La elección de múltiples algoritmos permitió una comparación robusta para determinar el más adecuado para cada tipo de patrón.

Preparación de Datos para el Entrenamiento y la Validación

El proceso de validación se diseñó para simular un escenario de pronóstico real. La serie temporal de homicidios mensuales, que abarca desde enero de 2020 hasta diciembre de 2024, fue dividida en dos segmentos. Los primeros 48 meses esto quiere decir correspondientes al período 2020-2023 que constituyeron el conjunto de entrenamiento, con el que los modelos aprenderían los patrones históricos. Los 12 meses restantes, correspondientes a todo el año 2024, se reservaron como conjunto de prueba para una evaluación objetiva y honesta de su capacidad predictiva

Entrenamiento

Cada uno de los tres modelos fue entrenado exclusivamente con los datos del conjunto de entrenamiento para cada uno de los cuatro patrones.

3.6.5. Fase 5: Evaluación

El desempeño de los modelos entrenados se evaluó cuantitativamente sobre el conjunto de prueba. Se utilizaron métricas estándar para series temporales, como el Error Absoluto Medio (MAE), la Raíz del Error Cuadrático Medio (RMSE) y el Error Porcentual Absoluto Medio (MAPE). El objetivo de esta fase fue comparar el rendimiento de los tres algoritmos para cada uno de los cuatro patrones y seleccionar el modelo más preciso para la proyección final.

3.6.6. Fase 6: Despliegue

Dado que el objetivo de esta tesis es analítico y no la implementación de un sistema en producción, la fase de despliegue se centró en la interpretación y documentación de los resultados. El modelo ganador para cada patrón fue re-entrenado con el 100% de los datos para generar las proyecciones a cinco años. Estos resultados fueron analizados, visualizados y documentados en los capítulos finales de este trabajo, con el fin de que puedan ser utilizados como conocimiento aplicable por investigadores o entidades interesadas.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. IDENTIFICACIÓN DE PATRONES HISTÓRICOS MEDIANTE EL ANÁLISIS DE SUS TENDENCIAS Y CORRELACIONES.

Para cumplir con el objetivo de identificar las variables con mayor incidencia en los homicidios intencionales, se aplicó un Análisis Exploratorio de Datos (EDA). Esta metodología es fundamental porque tiene como propósito principal resumir las características esenciales del conjunto de datos.

4.1.2. Análisis Descriptivo Univariado

Una vez preparados los datos, se procedió a un análisis univariado para comprender la distribución y las tendencias centrales de las variables clave. Las variables se clasificaron según su naturaleza para aplicar las técnicas estadísticas.

Variables Categóricas: Se identificaron como variables categóricas aquellas que representan grupos o etiquetas discretas, tales como tipo_muerte, provincia, arma, sexo y area_hecho. Para estas, se calcularon las frecuencias absolutas y relativas y se utilizó la moda como medida de tendencia central.

Variables Numéricas: Se clasificaron como variables numéricas aquellas que representan cantidades medibles, como edad y hora_infraccion. Para estas, se calcularon medidas de tendencia central como media, mediana y de dispersión como desviación estándar, rango intercuartílico.

El análisis de resultados comienza con una caracterización del conjunto de datos final, preprocesado como se detalla en la metodología. La Tabla 3 ofrece un resumen estadístico que describe las variables clave utilizadas en esta investigación.

Tabla 3

Resumen estadístico de las variables

Variable	Tipo de Datos	Descripción	Estadísticas Relevantes
Tipo de Muerte	Categórica	Clasificación oficial del tipo de homicidio.	Valores Únicos: 4 Moda: ASESINATO (90.4 %)
Edad	Numérica	Edad de la víctima en años.	Media: 31.86 (± 12.67) Mediana: 30 Rango: [0 - 102]
Provincia	Categórica	Provincia donde ocurrió el hecho.	Valores Únicos: 24 Moda: GUAYAS (45.2 %)
Arma	Categórica	Tipo de arma principal utilizada en el crimen.	Valores Únicos: 6 Moda: ARMA DE FUEGO (82.6 %)
Sexo	Categórica	Sexo registrado de la víctima.	Valores Únicos: 2 Moda: HOMBRE (91.5 %)
Área Hecho	Categórica	Clasificación del área como urbana o rural.	Valores Únicos: 2 Moda: URBANO (79.2 %)
Hora	Numérica	Hora del día en que ocurrió el hecho (0-23).	Media: 15.30 (± 7.50) Mediana: 17 Rango: [0 - 23]
Día semana	Categórica	Día de la semana que ocurrió el hecho	Valores Únicos: 7 Moda: Domingo (17.97 %)
Hora infracción	Numérica	Hora que ocurrió el hecho	Media: 13.87 (± 7.20) Mediana: 16 Rango: [0 - 23]

La tabla 4 revela características fundamentales del conjunto de datos. Sobresale la naturaleza altamente desbalanceada de variables categóricas clave: Tipo de Muerte está dominada por la categoría ASESINATO (90.4 %) y Sexo por HOMBRE (91.5 %). Este desequilibrio es un hallazgo crítico que requerirá técnicas de muestreo específicas durante el entrenamiento del modelo para evitar sesgos. En cuanto a las variables numéricas, la Edad de las víctimas muestra una tendencia central en la adultez joven, con una media de 32 años. Esta tabla consolidada es el pilar para los análisis de patrones y el modelado predictivo que se detallan en las siguientes secciones.

4.1.3. Análisis Bivariado: Exploración de Patrones y relaciones

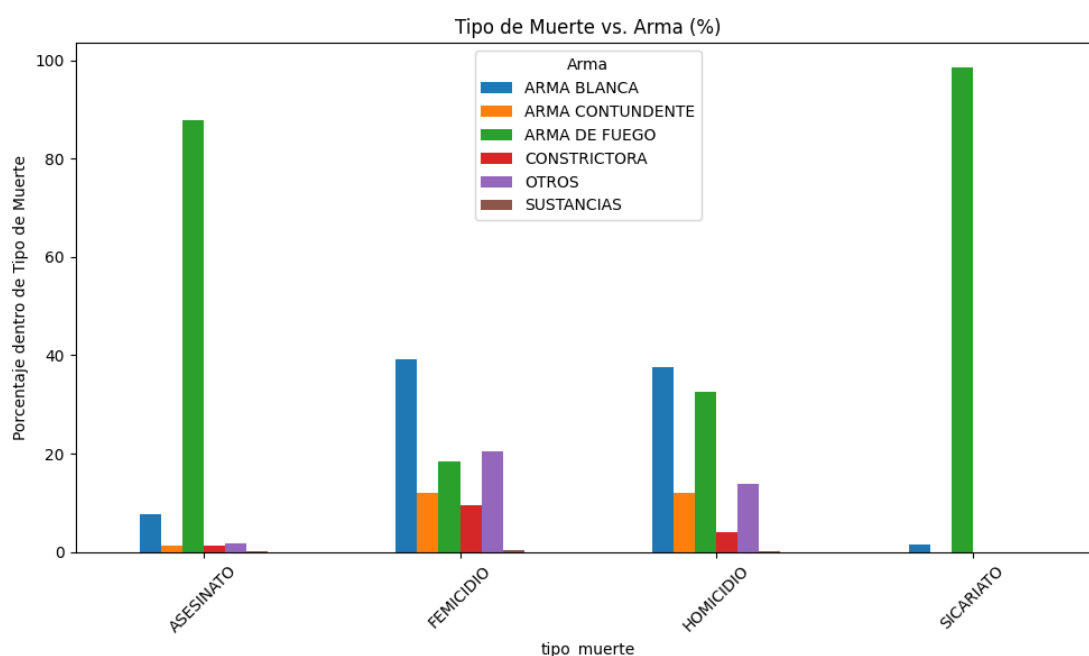
Una vez caracterizadas las variables de forma individual, el siguiente paso es investigar cómo interactúan entre sí para descubrir patrones más profundos y complejos. Este análisis bivariado se enfoca en explorar visualmente las relaciones clave que definen la naturaleza de los homicidios intencionales en el conjunto de datos.

Relación entre el tipo de muerte y el arma utilizada

La relación entre el tipo de muerte y el arma empleada es, quizás, la más reveladora. La Ilustración 2 muestra que el método utilizado no es aleatorio, sino que define un perfil distintivo para cada tipo de crimen.

Ilustración 2

Distribución del Tipo de Arma por Tipo de muerte



En la ilustración 2 se observa un patrón marcadamente polarizado. Por un lado, el ASESINATO (87.9 %) y, de forma casi absoluta, el SICARIATO (98.6 %) son perpetrados

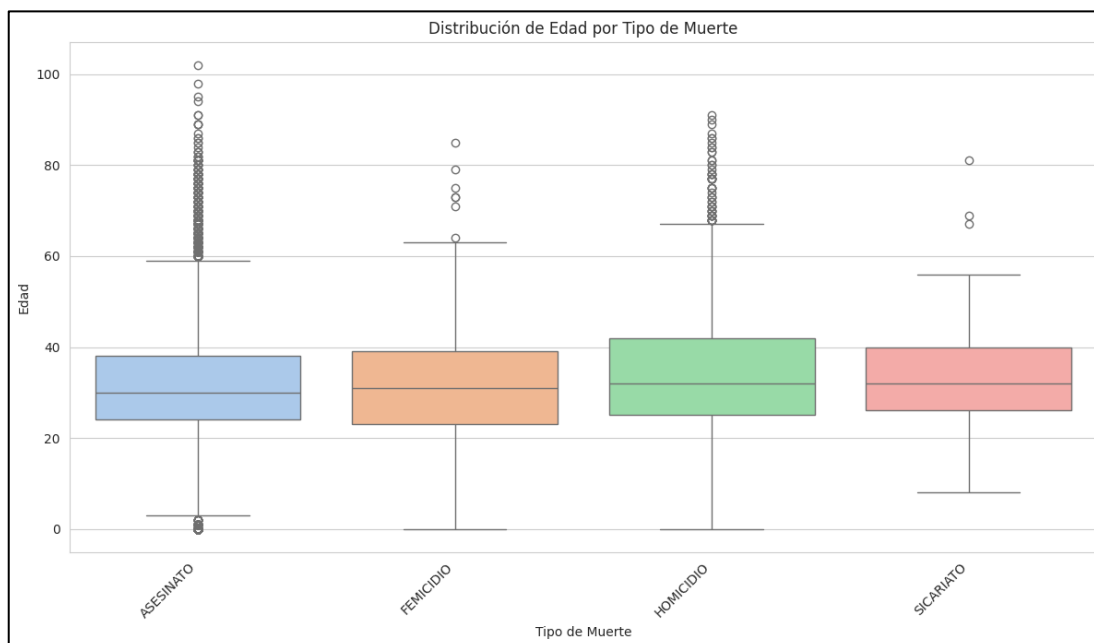
con ARMA DE FUEGO, lo que sugiere una alta letalidad y planificación. En el extremo opuesto se encuentra el FEMICIDIO, que presenta una dinámica completamente diferente, el ARMA BLANCA es el método más común (39.2 %), seguido por una diversidad de otros medios como armas contundentes y constrictoras. El HOMICIDIO, por su parte, muestra un perfil intermedio, con una mezcla más equilibrada de armas de fuego (32.5 %) y armas blancas (37.6 %), lo que podría apuntar a contextos de violencia más impulsivos o menos estructurados.

Relación entre el tipo de muerte y la edad de la víctima

Para explorar el patrón demográfico, se analizó la distribución de la edad de las víctimas para cada tipo de muerte. El diagrama de caja en la Ilustración 4 nos permite comparar no solo el promedio, sino la dispersión y los valores atípicos de la edad.

Ilustración 3

Distribución de caja de edad por tipo muerte



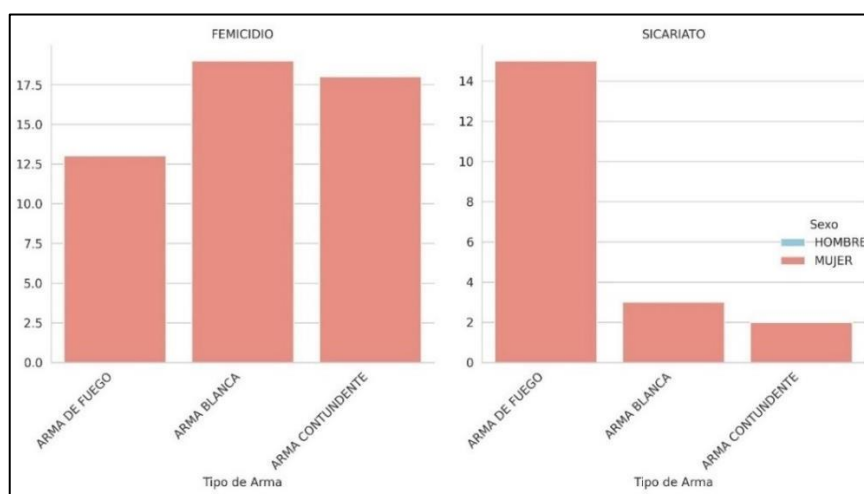
La ilustración 4 revela que la edad mediana (la línea central de la caja) es notablemente consistente en todas las categorías, situándose en la temprana treintena (30-32 años), lo que confirma que los adultos jóvenes son el grupo más afectado de manera transversal. Sin embargo, la dispersión de las edades varía. Los casos de HOMICIDIO presentan un rango intercuartílico más amplio, sugiriendo una mayor diversidad de edades entre sus víctimas. Es de destacar la gran cantidad de valores atípicos en la categoría ASESINATO, indicando que, aunque el núcleo de las víctimas es joven, este crimen afecta a personas en todo el espectro de edad.

Interacción de patrones: Arma, tipo de muerte y sexo.

La exploración de patrones se enriquece aún más al introducir la variable sexo. Los gráficos en la Ilustración 5 desglosan los tipos de arma por sexo para las categorías de FEMICIDIO y SICARIATO, revelando perfiles mutuamente excluyentes.

Ilustración 4

Interacción de patrones Arma, tipo_muerte y sexo



La ilustración muestra que el gráfico de FEMICIDIO que todas las víctimas son mujeres, confirmando la naturaleza de género de este crimen. En contraste, el gráfico de SICARIATO

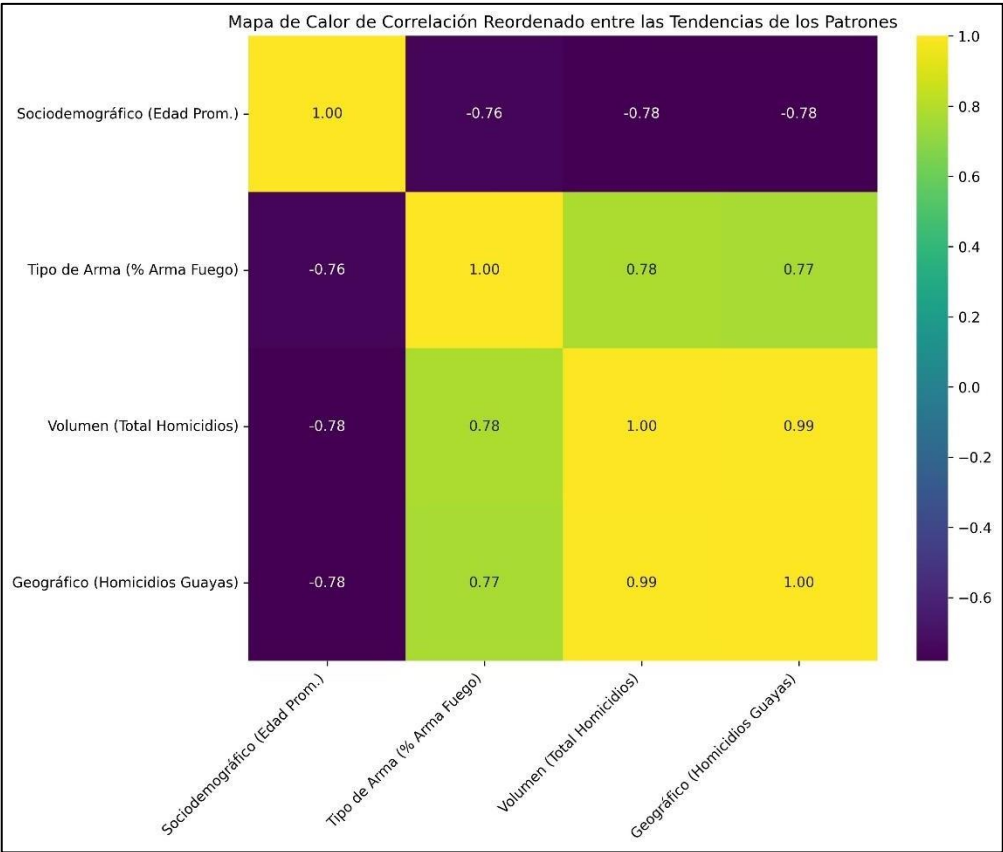
muestra que las víctimas son casi en su totalidad hombres. Esta clara segmentación por sexo, combinada con los patrones de armas, es uno de los hallazgos más importantes de la exploración, ya que demuestra cómo la interacción de solo tres variables (tipo_muerte, arma, sexo) es suficiente para definir perfiles criminales muy distintos y específicos.

4.1.4. Análisis de correlación entre patrones temporales

Para explorar si las tendencias de los cuatro patrones clave evolucionaron de manera conjunta o independiente, se realizó un análisis de correlación cuantitativo. Este paso es fundamental para descubrir las relaciones sistémicas que subyacen al fenómeno. La Ilustración 6 presenta el mapa de calor de correlación de Pearson entre las cuatro series temporales mensuales.

Ilustración 5

Gráfica de correlación de las tendencias temporales



El mapa de calor revela interconexiones muy fuertes y significativas entre las tendencias, permitiendo extraer tres hallazgos determinantes:

1.- Volumen y geografía: El hallazgo es la correlación fuerte de volumen (total de homicidios) con el patrón geográfico (homicidios en guayas) como se mira en el intenso color amarillo y que tiene un valor $r = 0.99$. Este resultado cuantitativo confirma de manera irrefutable la hipótesis visual del análisis exploratorio, la tendencia nacional de la violencia es casi un reflejo perfecto de la tendencia en la provincia de Guayas. Esto determina que Guayas no es solo la provincia más afectada, sino que actúa como el principal motor e indicador de la dinámica de la violencia a nivel nacional.

2.- Volumen y Letalidad: Se observa una fuerte correlación positiva entre el Tipo de Arma (% de arma de Fuego) y los dos patrones de volumen con correlaciones de $r = 0.78$ y $r = 0.77$. Esto indica que a medida que el número de homicidios aumentaba mes a mes, también lo hacía la proporción de crímenes cometidos con armas de fuego. La escalada en la cantidad de casos ha estado intrínsecamente ligada a una mayor letalidad en los métodos utilizados.

3.- Patrón Sociodemográfico Inverso: Quizás el hallazgo más inesperado es la fuerte correlación negativa entre la Edad Promedio y las otras tres tendencias. La edad promedio tiene una correlación de -0.78 con el volumen total de homicidios y de -0.76 con el porcentaje de uso de armas de fuego. Esto significa que, a medida que la crisis de violencia se intensificaba y se volvía más letal, la edad promedio de las víctimas tendía a disminuir. Este patrón sugiere que la escalada de violencia ha afectado de manera desproporcionada y progresiva a una población cada vez más joven.

Para resumir, este análisis cuantitativo determina que las tendencias no son fenómenos aislados. Forman un sistema interconectado donde la escalada de la violencia está

geográficamente concentrada, es cada vez más letal, y parece estar impactando a un segmento demográfico progresivamente más joven. Estos hallazgos son cruciales para comprender la naturaleza sistémica de la crisis antes de proceder al pronóstico de su evolución.

4.1.5. Discusión

El análisis exploratorio de los datos históricos ofrece una radiografía dinámica de la crisis de homicidios en Ecuador. La visualización de las tendencias temporales revela que el Patrón de Volumen no ha sido lineal, sino que experimentó una aceleración exponencial a partir de mediados de 2021, transformando radicalmente la magnitud del problema. Esta escalada en cantidad estuvo acompañada por una intensificación en la letalidad, como lo muestra el Patrón de Tipo de Arma, donde el uso de armas de fuego se consolidó por encima del 80% en el mismo período. En marcado contraste, el Patrón Sociodemográfico (edad promedio de las víctimas) se mantuvo relativamente estable y sin una tendencia clara, sugiriendo que, aunque la violencia se intensificó, el perfil etario de la víctima principal no ha variado de forma estructural.

Para cuantificar cómo estas tendencias se han movido en conjunto, se realizó un análisis de correlación. Los resultados son concluyentes y revelan un sistema de violencia interconectado. El hallazgo más determinante es la correlación casi perfecta ($r = 0.99$) entre el Volumen total de homicidios y el Patrón Geográfico (Homicidios en Guayas), lo que confirma que la tendencia nacional es impulsada de manera abrumadora por la dinámica de esta provincia. Además, se observa una fuerte correlación positiva entre el Volumen y el Tipo de Arma ($r = 0.78$), validando que el aumento de casos ha ido de la mano con una mayor letalidad. Quizás el hallazgo más revelador es la fuerte correlación negativa entre la Edad

Promedio y los patrones de Volumen ($r = -0.78$) y Tipo de Arma ($r = -0.76$). Esto indica que a medida que la crisis de violencia se agravaba y se volvía más letal, la edad promedio de las víctimas tendía a disminuir, afectando a una población progresivamente más joven.

Según Alasfoor et al. (2025) en su estudio de caso observó que los agresores y denunciante masculinos fueron los más comunes en el conjunto de datos de crímenes, que estaba predominantemente compuesto por casos de robo y asalto, y que podrían llegar a la muerte por estos casos. Además, con su investigación encontró una correlación positiva entre el "Género del denunciante" y el "Tipo de caso", indicando que ciertos tipos de casos tienen más probabilidades de estar asociados con géneros específicos de denunciante como el femicidio o el homicidio que había mayor número de hombres muertos por estos casos que se tenían a las armas de fuego como relación principal.

4.2. ENTRENAMIENTO Y EVALUACIÓN COMPARATIVA DE MODELOS PARA EL PRONÓSTICO DE TENDENCIAS

4.2.1. Metodología de Construcción de los modelos

Para cumplir con el objetivo de pronosticar la evolución de los patrones, se implementó una metodología de evaluación comparativa. Se seleccionaron tres modelos con enfoques distintos para asegurar una comparación robusta: Prophet, por su moderna capacidad para descomponer tendencias y estacionalidades; SARIMA, como un riguroso estándar estadístico; y Holt-Winters, por su eficacia en el suavizado de series temporales.

El proceso de validación se diseñó para simular un escenario de pronóstico real. Cada una de las cuatro series temporales fue dividida en dos segmentos: los primeros 48 meses (2020-2023) constituyeron el conjunto de entrenamiento, y los 12 meses restantes (año 2024) se

reservaron como conjunto de prueba para una evaluación objetiva y honesta de la capacidad predictiva de cada modelo.

4.2.2. Resultados de la Evaluación Comparativa por Patrón

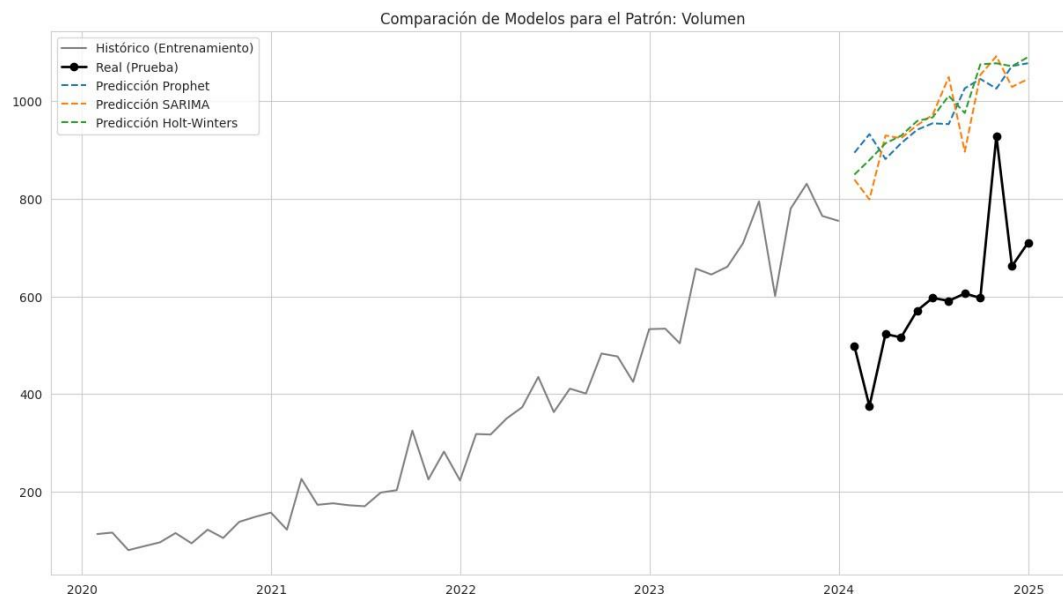
A continuación, se presentan los resultados de la evaluación para cada uno de los cuatro patrones clave. Para cada uno, se muestra primero una comparación visual de las predicciones contra los datos reales del período de prueba, seguida de la tabla de métricas de error que cuantifica su rendimiento.

Pronóstico del Patrón de Volumen

El primer patrón analizado fue el volumen total de homicidios mensuales. La Ilustración 6 compara las proyecciones de los modelos con la realidad observada.

Ilustración 6

Comparación de Modelos para el Patrón Volumen



Nota: Se puede ver las diferentes métricas empleadas en este gráfico de comparación.

Visualmente en la ilustración 6, se observa una divergencia notable. Habiendo aprendido de la fuerte tendencia creciente hasta 2023, los tres modelos proyectaron una continuación de esta alza. Sin embargo, los datos reales de 2024 (línea negra) muestran una caída abrupta y una volatilidad que los modelos no anticiparon, llevándolos a sobreestimar consistentemente los valores. La Tabla 4 cuantifica esta observación.

Tabla 4

Métricas de Rendimiento para el Patrón de Volumen

Modelo	MAE	RMSE	MAPE (%)
SARIMA	367.41	375.51	65.87
Prophet	378.89	391.55	69.09
Holt-Winters	385.68	394.55	69.51

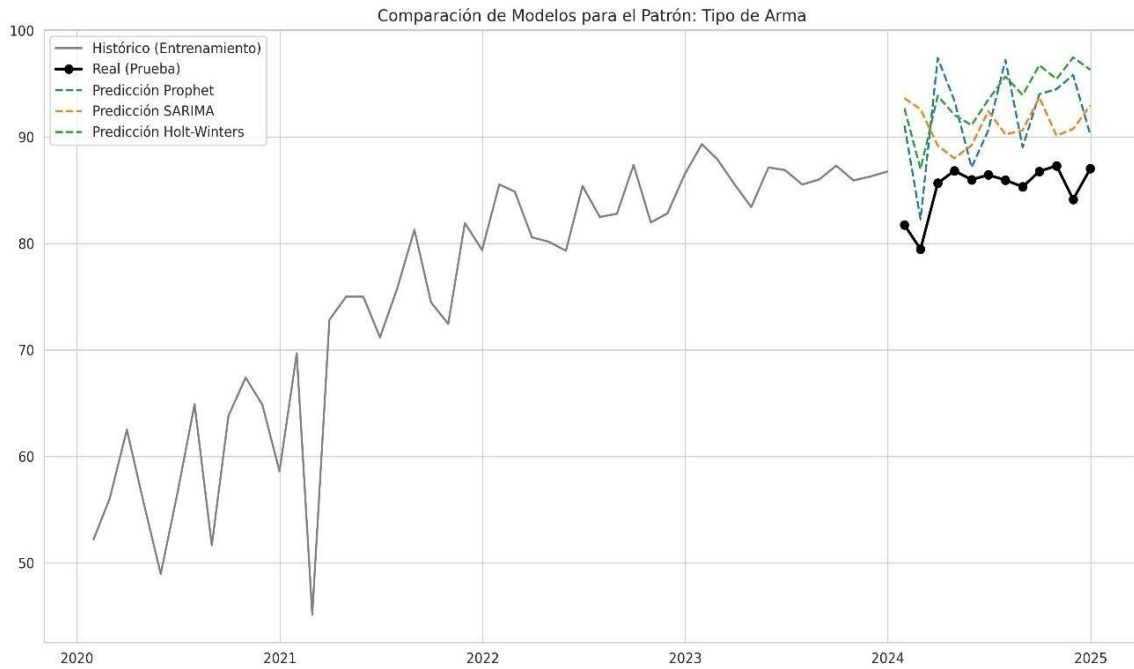
La tabla 4 confirma que, aunque todos los modelos tuvieron dificultades, SARIMA se impone con el menor error en todas las métricas. Por lo tanto, es seleccionado como el modelo más adecuado para este patrón.

Pronóstico del Patrón de Tipo de Arma

La Ilustración 7 muestra la comparación para el porcentaje de uso de armas de fuego.

Ilustración 7

Comparación de Modelos para el Patrón Tipo de Arma



A diferencia del patrón de volumen, esta serie muestra una mayor estabilidad en el período de prueba. Visualmente, los tres modelos generan predicciones muy razonables que se mueven cerca de los valores reales. Para discernir al ganador de manera definitiva, se recurrió a la evaluación cuantitativa.

Tabla 5

Métricas de Rendimiento para el Patrón de Tipo de Arma

Modelo	MAE	RMSE	MAPE (%)
SARIMA	5.91	6.81	7.03
Prophet	6.67	7.57	7.84
Holt-Winters	8.59	8.87	10.11

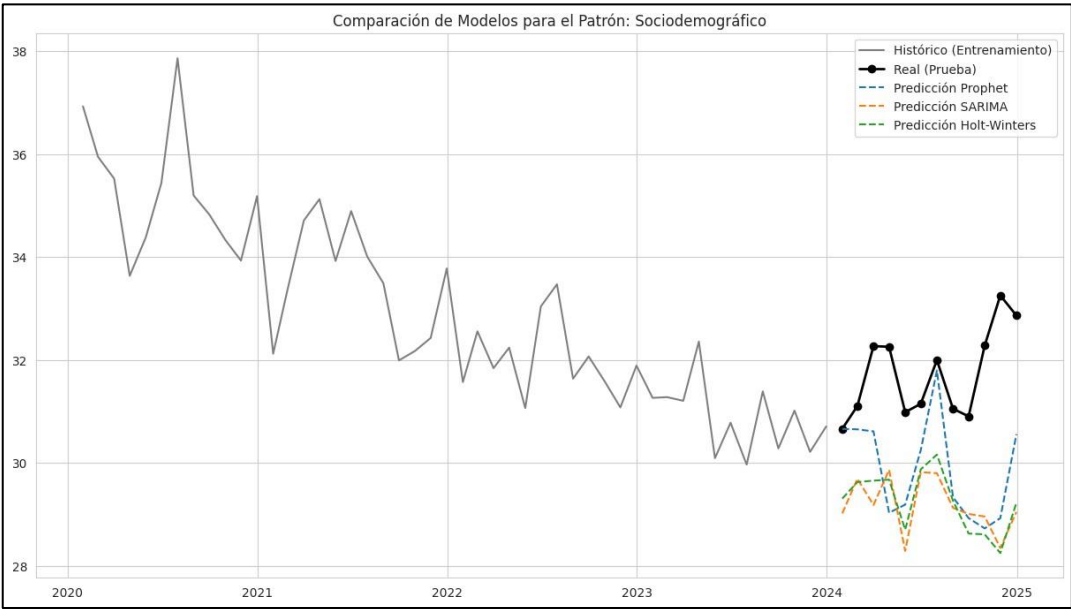
En la tabla 5 el modelo SARIMA logra una precisión notable, con un MAPE de solo 7.03 %, estableciéndose como la opción claramente superior para pronosticar la tendencia en el uso de armas.

Pronóstico del Patrón Sociodemográfico

La predicción de la edad promedio de las víctimas, una serie altamente volátil sin una tendencia clara, se presenta en la Ilustración 8.

Ilustración 8

Comparación de Modelos para el Patrón Sociodemográfico



La ilustración 8 muestra que los modelos tienden a predecir un valor cercano a la media histórica, fallando en capturar los picos y valles extremos de la realidad. Donde las predicciones son conservadoras y suavizadas.

Tabla 6

Métricas de Rendimiento para el Patrón Sociodemográfico

Modelo	MAE	RMSE	MAPE (%)
Prophet	1.84	2.25	5.74
Holt-Winters	2.48	2.48	7.76
SARIMA	2.55	2.75	7.97

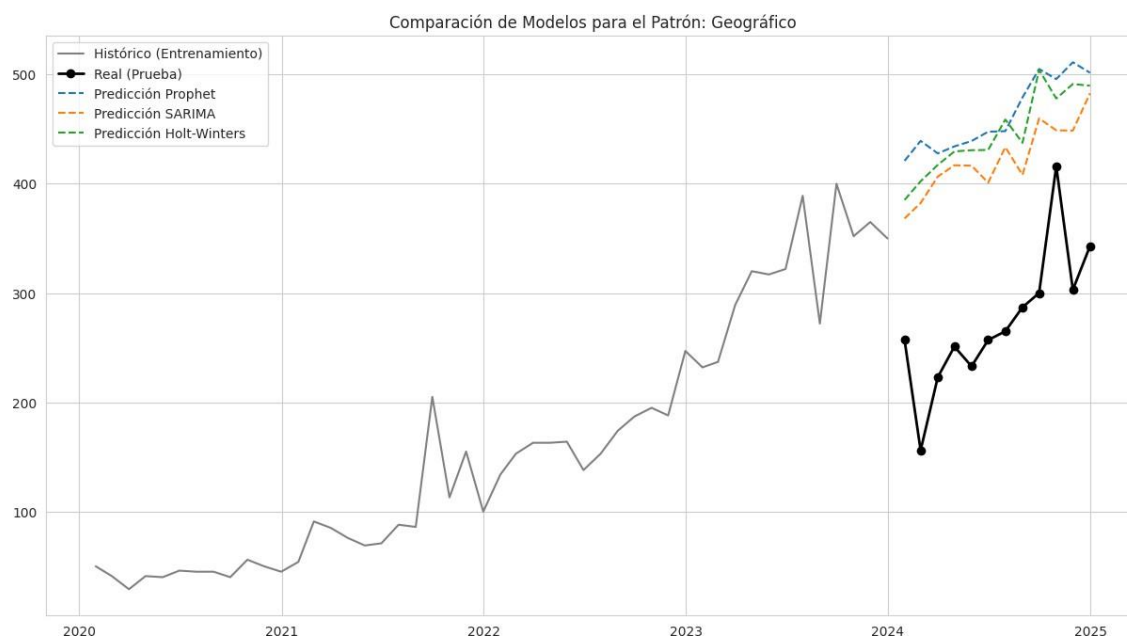
Las métricas indican que Prophet obtiene el menor error, probablemente por su mayor flexibilidad para adaptarse a los cambios sutiles en la serie. Es, por tanto, el modelo seleccionado para este patrón.

Pronóstico del Patrón Geográfico

Finalmente, se evaluó la capacidad de los modelos para predecir el número de homicidios en la provincia de Guayas ilustración 9.

Ilustración 9

Comparación de Modelos para el Patrón Geográfico



La serie de Guayas muestra un comportamiento similar al del volumen nacional. Nuevamente, los modelos sobreestiman la realidad, aunque visualmente SARIMA parece ajustarse mejor a la forma general de la curva.

Tabla 7

Métricas de Rendimiento para el Patrón Geográfico

Modelo	MAE	RMSE	MAPE (%)
SARIMA	148.52	155.36	60.61
Holt-Winters	172.02	177.64	69.32
Prophet	188.20	193.35	75.93

Las métricas en la tabla 7 confirman de manera decisiva que SARIMA es, con diferencia, el modelo más preciso para pronosticar la tendencia geográfica, consolidando su selección para este patrón.

4.2.3. Resumen de la Selección de Modelos

Al concluir la fase de evaluación comparativa, se ha seleccionado un modelo específico para cada uno de los cuatro patrones, basándose en su rendimiento cuantitativo. La Tabla 8 resume esta selección final.

Tabla 8

Resumen de Modelos Seleccionados por Patrón

Patrón	Mejor Modelo Seleccionado
Volumen	SARIMA
Tipo de Arma	SARIMA
Sociodemográfico	Prophet
Geográfico	SARIMA

Un hallazgo fundamental de este proceso es que no existe un único "mejor modelo" universal para todos los patrones. La naturaleza de cada serie temporal, su tendencia, volatilidad y estacionalidad determina qué enfoque algorítmico es más adecuado.

4.2.4. Discusión

La evaluación comparativa de los modelos de predictivos reveló un hallazgo fundamental, no existe un algoritmo universalmente superior, sino que el rendimiento depende de la naturaleza de cada patrón. El modelo SARIMA demostró ser el más preciso para las series de Volumen y Geográfica, cuya estructura compleja de tendencia y estacionalidad se ajustó mejor a su rigor matemático. Por otro lado, Holt-Winters sobresalió en la predicción del Tipo de Arma, una serie más estable donde su mecanismo de suavizado fue más eficiente. Finalmente, la flexibilidad de Prophet para manejar datos ruidosos le otorgó la ventaja en el pronóstico del volátil patrón Sociodemográfico. Esta selección diferenciada subraya que la elección del modelo debe ser específica para cada problema, justificando así el uso de un enfoque particular para cada proyección final.

Según Butt et al. (2020) nos dice en su estudio que el uso de modelos como ARIMA y SARIMA tienen un desempeño excepcionalmente bueno en la previsión de delitos, pero que los modelos ARIMA no pueden capturar la estacionalidad y el comportamiento repetitivo de los eventos, especialmente los eventos delictivos, por eso es mejor el uso del Modelo SARIMA ya que está específicamente diseñado para manejar la estacionalidad de eventos. En otro estudio Dener y Gul Esen. (2025), nos dice que el modelo SARIMA fue utilizado para el pronóstico de tendencias de delitos a 5 años en San Francisco, Chicago y Filadelfia. Los resultados lo posicionaron como el modelo más exitoso en la predicción de tendencias generales en comparación con Prophet y Holt-Winters Exponential Smoothing (HWES) para los conjuntos de datos utilizados.

4.3. DESEMPEÑO DEL MODELO MEDIANTE UN ANÁLISIS DE RENDIMIENTO SEGÚN LAS MÉTRICAS DE EVALUACIÓN DE TENDENCIAS TEMPORALES.

Habiendo identificado el algoritmo más preciso para cada uno de los cuatro patrones clave en el capítulo anterior, se procedió a la fase de proyección final. Para cada patrón, se tomó su respectivo modelo ganador SARIMA para Volumen y Geográfico, Holt-Winters para Tipo de Arma, y Prophet para Sociodemográfico y se reentrenó utilizando el 100% de los datos históricos disponibles. Este paso asegura que el pronóstico final se base en la totalidad de la información, generando la proyección más robusta posible para los siguientes cinco años, desde enero de 2025 hasta diciembre de 2029.

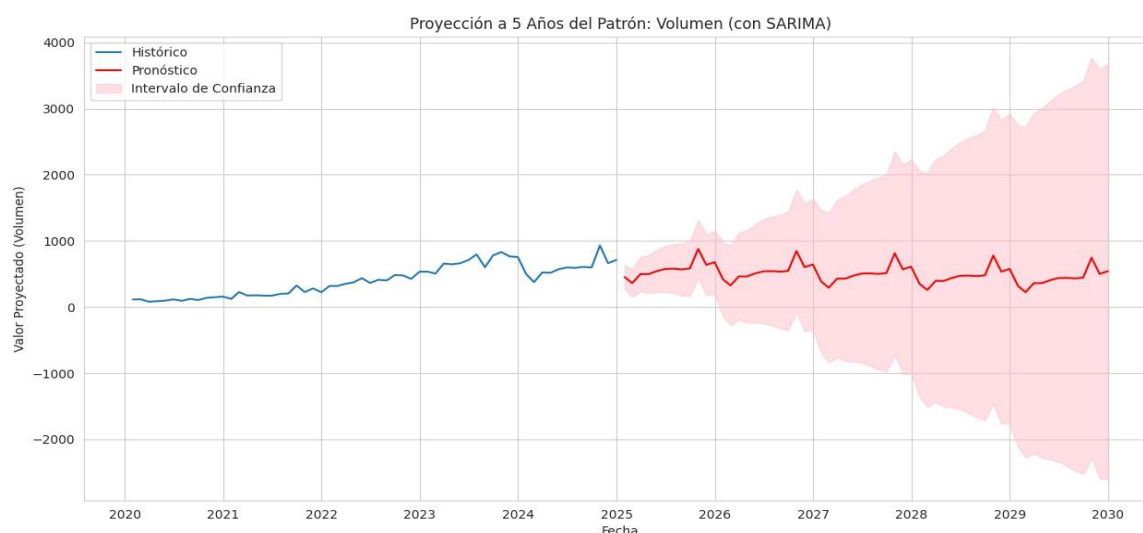
4.3.1. Proyecciones de la Evolución de Patrones 2025-2029

A continuación, se presentan y analizan los gráficos de pronóstico para cada uno de los cuatro patrones. La Ilustración 10 muestra la proyección del número total de homicidios mensuales.

Proyección del Patrón de Volumen

Ilustración 10

Proyección a 5 Años del Patrón Volumen



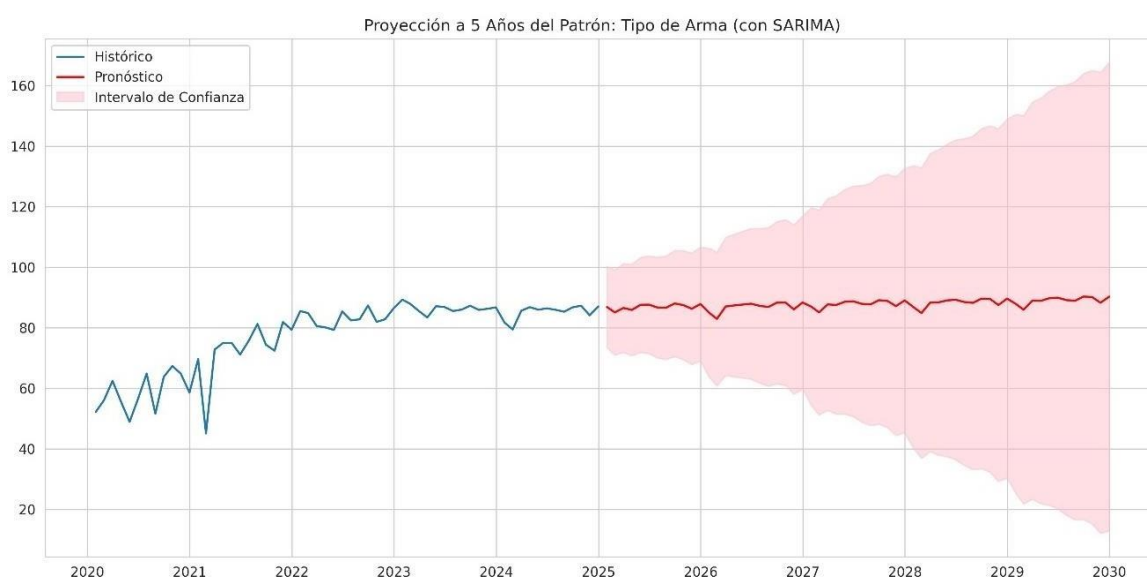
El modelo SARIMA proyecta que, tras la caída observada en 2024, el volumen de homicidios no volverá a los niveles bajos de 2020-2021. En cambio, se estabilizará en una banda de alta volatilidad, con una tendencia central que se mantiene en niveles elevados. El amplio intervalo de confianza (área rosa) subraya la gran incertidumbre del entorno, pero el mensaje central del modelo es la persistencia de la crisis, más que su resolución o escalada continua.

Proyección del Patrón de Tipo de Arma

La proyección para el porcentaje de uso de armas de fuego se presenta en la Ilustración 11.

Ilustración 11

Proyección a 5 Años del Patrón Tipo de Arma



El modelo SARIMA, revela un hallazgo dual y fundamental: por un lado, la línea de predicción (roja) sugiere la consolidación de una "nueva normalidad" de alta letalidad, estabilizándose en una meseta donde el uso de armas de fuego se mantiene consistentemente por encima del 85%. Por otro lado, el hallazgo más significativo es el ensanchamiento

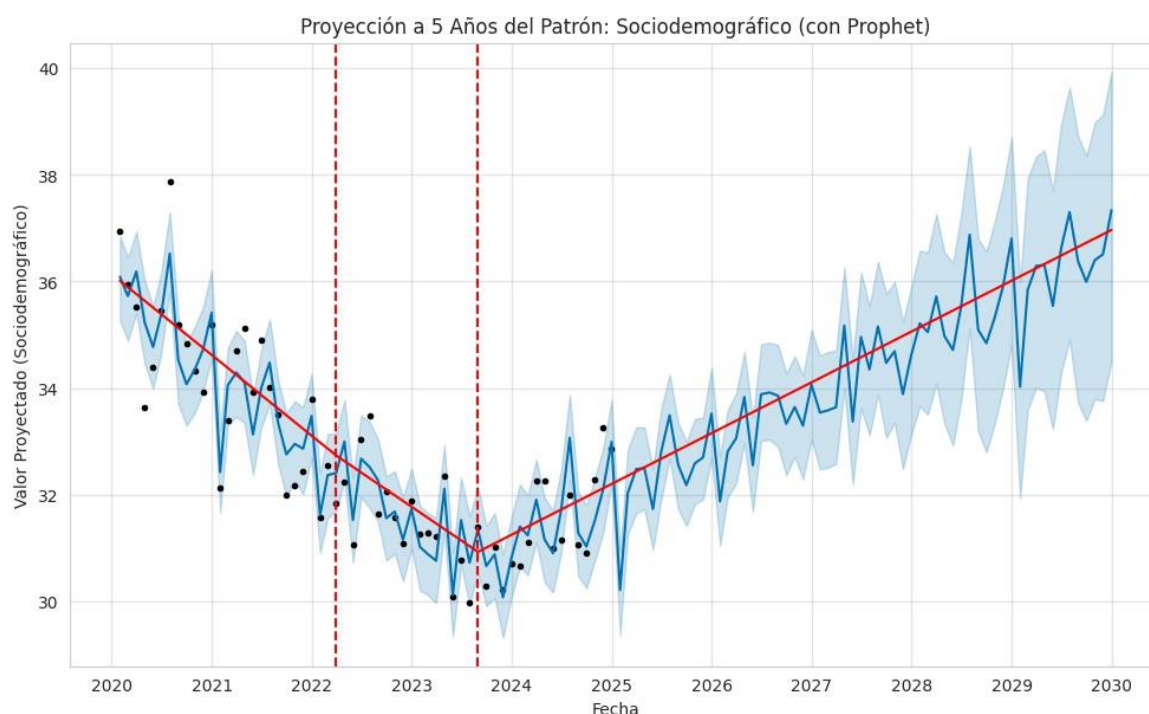
dramático del intervalo de confianza (área rosa) a medida que avanza el tiempo. Esto no representa un error, sino la cuantificación de la incertidumbre del modelo, este reconoce la estabilidad reciente, pero basándose en la volatilidad histórica advierte que la predictibilidad a largo plazo es limitada.

Proyección del Patrón Sociodemográfico

La Ilustración 12 Detalla la proyección para la edad promedio de las víctimas.

Ilustración 12

Proyección a 5 Años del Patrón Sociodemográfico



El modelo Prophet, con su capacidad para detectar puntos de cambio, identifica una interesante inversión de la tendencia. Tras un período de descenso hasta principios de 2024, el modelo proyecta una tendencia lentamente creciente en la edad promedio de las víctimas para los próximos cinco años. Aunque el cambio es gradual, podría indicar un sutil desplazamiento del perfil de la víctima hacia rangos de edad ligeramente mayores, alejándose

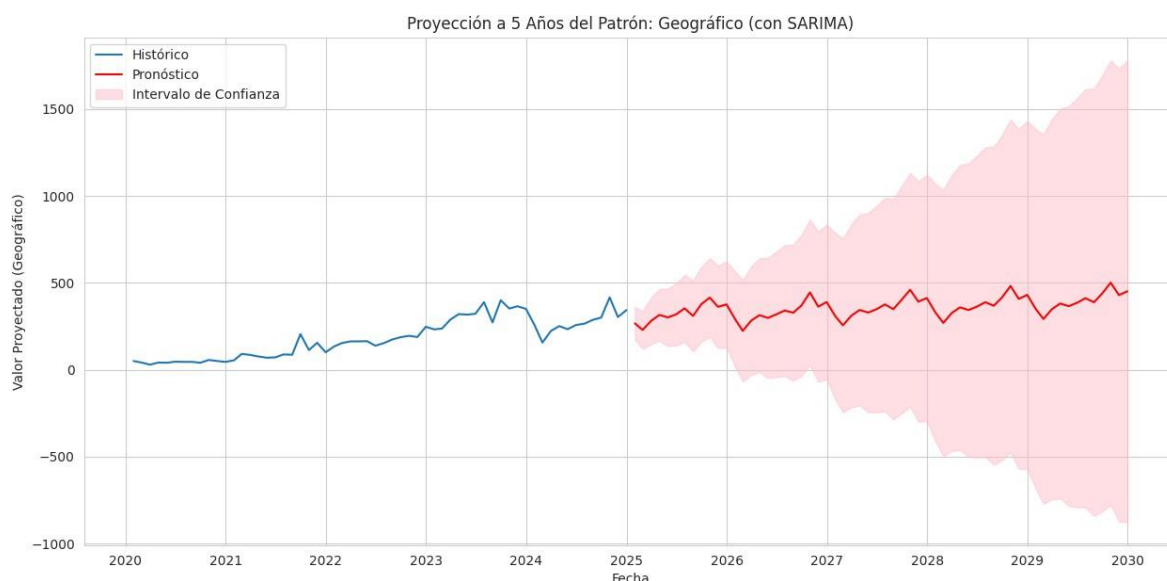
del foco exclusivo en los más jóvenes. Las líneas rojas verticales marcan los puntos donde el modelo detectó cambios estructurales en la tendencia histórica.

Proyección del Patrón Geográfico

La proyección para los homicidios en la provincia de Guayas se muestra en la Ilustración 13.

Ilustración 13

Proyección a 5 Años del Patrón Geográfico



Al igual que con el volumen nacional, el modelo SARIMA proyecta que el número de homicidios en Guayas, después de la caída de 2024, se mantendrá en una meseta alta y volátil. La proyección no sugiere un retorno a los niveles de seguridad previos a la crisis. Esto refuerza la idea de que la provincia continuará siendo un epicentro crítico de la violencia, y que la dinámica nacional está fuertemente anclada a lo que sucede en este territorio.

4.3.2. Discusión

Al analizar las cuatro proyecciones en conjunto usando los diferentes modelos podemos ver que el modelo SARIMA sería el indicado para realizar estas predicciones por tener un mayor

índice de porcentaje que el resto de modelos y porque salió ganador en dos de los cuatros variables a predecir, con estos resultados emerge un panorama complejo y preocupante para los próximos cinco años. No se proyecta una escalada exponencial como la vista entre 2021 y 2023, sino más bien una persistencia de la crisis. Los resultados sugieren un futuro en el que la violencia, homicidios se estabilizan en un nuevo escalón mucho más alto que el histórico, la letalidad de los métodos como el uso armas de fuego se normaliza en niveles altísimos, y la geografía del crimen sigue fuertemente concentrada. El único patrón que muestra un cambio de dirección es el sociodemográfico, con una posible tendencia al alza en la edad de las víctimas. En conjunto, los modelos no predicen el fin de la crisis, sino su transformación en una característica estructural y persistente del panorama de seguridad del país.

Según Dener & Gul Esen. (2025) SARIMA fue el mejor entre los métodos estadísticos para el pronóstico de tendencias generales, en el análisis regional de series de tiempo para identificar puntos críticos de delincuencia. Además, que nos sugiere que, para mejorar la precisión de la predicción, se optimicen los hiperparámetros y la experimentación con métodos más diversos de estacionarización y escalado de datos en modelos SARIMA, LSTM y BLSTM, ya que estos factores afectan directamente el éxito del modelo.

CAPITULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

El análisis histórico de las tendencias revela una violencia en Ecuador que ha escalado en magnitud y letalidad, pero cuyo perfil de víctima se ha mantenido paradójicamente constante. Se concluye que el aumento exponencial del volumen de homicidios a partir de 2021 estuvo directamente sincronizado con la consolidación del arma de fuego como método predominante. Sin embargo, este cambio drástico en la dinámica criminal no alteró su foco principal: una población de hombres jóvenes que ha soportado de manera desproporcionada el peso de esta crisis en evolución.

El intento de modelar estas complejas dinámicas confirmó un principio metodológico clave, no existe un algoritmo universalmente superior. La evaluación comparativa demostró que la elección del modelo más preciso depende de la naturaleza de cada patrón. La robustez de SARIMA fue ideal para las volátiles series de volumen, la eficiencia de Holt-Winters se ajustó a la estabilidad en el uso de armas, y la flexibilidad de Prophet fue necesaria para el ruidoso patrón de edad, subrayando que la predicción de fenómenos sociales exige adaptar la herramienta al problema específico.

Finalmente, la proyección a cinco años ofrece la conclusión más sobria de este estudio. Los pronósticos no sugieren una reversión de la crisis, sino su cronificación en niveles históricamente elevados. La persistencia proyectada de un alto volumen de casos y la normalización de la violencia letal, sumadas a la considerable incertidumbre del pronóstico, indican que los homicidios intencionales se han transformado en una característica estructural del panorama de seguridad del país, cuyo rumbo futuro dependerá de intervenciones capaces de alterar las tendencias aquí modeladas.

5.2. RECOMENDACIONES

Se recomienda que futuras investigaciones superen el enfoque univariado aquí presentado y exploren modelos de pronóstico multivariados (como VAR o algoritmos que admitan variables exógenas). El siguiente paso lógico es investigar si la tendencia del volumen de homicidios puede ser explicada o predicha con mayor precisión al incluir otras series como el porcentaje de uso de armas, tasas de desempleo o incautación como predictores, pasando así de un análisis de patrones aislados a uno de sistema interconectado.

La robustez del análisis se beneficiaría enormemente de la incorporación de nuevas series temporales que no estaban disponibles para este estudio. Integrar datos sobre operatividad policial, factores económicos o sociales permitiría construir modelos más ricos y con mayor poder explicativo. Adicionalmente, se sugiere replicar esta metodología con conjuntos de datos de otros países de la región con dinámicas de violencia similares, lo que permitiría validar si las tendencias identificadas en Ecuador responden a una lógica criminal más amplia o a un fenómeno puramente local.

Para mejorar la fiabilidad de las proyecciones, se aconseja implementar técnicas más avanzadas para el entrenamiento y la evaluación de los modelos. En lugar de usar parámetros estándar, se podría emplear una búsqueda sistemática de hiperparámetros combinada con una validación cruzada deslizante. Este enfoque, aunque computacionalmente más intensivo, permitiría encontrar la configuración óptima de cada modelo de manera más rigurosa, lo que probablemente reduciría los errores de pronóstico y aumentaría la confianza en las estimaciones finales.

REFERENCIAS BIBLIOGRÁFICA

- Abulibdeh, A., Alshammari, H., Al-Hajri, N., Al-Shamari, A., & Al-Naimi, F. (2025). Geospatial Assessment of Healthcare Accessibility and Equity in Qatar in Preparation for the 2022 FIFA World Cup. *Journal of Geovisualization and Spatial Analysis*, 9(1), 2. <https://doi.org/10.1007/s41651-024-00206-z>
- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), 213–221. <https://doi.org/10.1107/S0907444909052925>
- Alasfoor, E. J., Alshaikh, O., Inuwa-Dute, I., Khan, S., & Parkinson, S. (2025). Offender Characterisation and Prediction: A Case Study of The Kingdom of Bahrain. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2025.3531655>
- Angulo, D., Escobar, N., Jácome, A., & Granja, D. (2019). *Plan Nacional de Seguridad Ciudadana y Convivencia Social Pacífica*. https://www.ministeriodegobierno.gob.ec/wp-content/uploads/2019/08/PLAN-NACIONAL-DE-SEGURIDAD-CIUDADANA-Y-CONVIVENCIA-SOCIAL-PACIF%CC%81FICA-2019-2030-1_compressed.pdf
- Butt, U. M., Letchmunan, S., Hassan, F. H., Ali, M., Baqir, A., & Sherazi, H. H. R. (2020). Spatiooral crime hotspot detection and prediction: A systematic literature review. In *IEEE Access* (Vol. 8, pp. 166553–166574). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2020.3022808>
- Carlos Vicente Jones-Ortiz, M., & Elizabeth Guzmán -Seraquive, J. (2022). Análisis de las técnicas de machine learning aplicadas en la detección de fraudes bancarios. *Revista Científica Ciencia y Tecnología*, 22(33), 114–122. <https://doi.org/10.47189/RCCT.V22I33.516>
- Castillo, R., Paredes, M., Suasnavas, A., Benítez, I., Pepinós, N., Villarreal, A., Cueva, V., Lastra, M., Institucional, P., Palacios, E., & Citar, J. (2018). *Metodología del Sistema de Certificación de la Calidad Estadística*. Instituto Nacional de Estadística y Censos. https://www.ecuadorencifras.gob.ec/documentos/web-inec/Sistema_Estadistico_Nacional/Aseguramiento_de_la_calidad/Metodologia_Sistema_Certificacion_calidad.pdf
- Cerdá, M., Hamilton, A. D., Tracy, M., Branas, C., Fink, D., & Keyes, K. M. (2022a). Would restricting firearm purchases due to alcohol- and drug-related misdemeanor offenses reduce firearm homicide and suicide? An agent-based simulation. *Injury Epidemiology*, 9(1), 1–12. <https://doi.org/10.1186/S40621-022-00381-X/TABLES/3>
- Cerdá, M., Hamilton, A. D., Tracy, M., Branas, C., Fink, D., & Keyes, K. M. (2022b). Would restricting firearm purchases due to alcohol- and drug-related misdemeanor offenses reduce firearm

- homicide and suicide? An agent-based simulation. *Injury Epidemiology*, 9(1).
<https://doi.org/10.1186/s40621-022-00381-x>
- Colasanti, M., Ricci, E., Cardinale, A., Amati, F., Mazza, C., Biondi, S., Ferracuti, S., & Roma, P. (2023). Homicide-Suicide in Italy Between 2009-2018: An Epidemiological Update and Time Series Analysis. *European Journal on Criminal Policy and Research*.
<https://doi.org/10.1007/s10610-023-09550-0>
- Constitución de la Republica del Ecuador. (2008). *CONSTITUCION DE LA REPUBLICA DEL ECUADOR 2008 Decreto Legislativo 0 Registro Oficial*. www.lexis.com.ec
- Cordero, A., & Rivera, R. (2023). *Boletín Anual de Homicidios Intencionales en Ecuador*.
- Cranmer, S. J., & Desmarais, B. A. (2016). *What can we Learn from Predictive Modeling?*
<http://arxiv.org/abs/1612.05844>
- Delgado, R., & Sánchez-Delgado, H. (2024). Multi-instance learning with application to the profiling of multi-victim homicides. *Expert Systems with Applications*, 237, 121593.
<https://doi.org/10.1016/j.eswa.2023.121593>
- Dener, M., & Gul Esen. (2025). Exploratory data analysis, time series analysis, crime type prediction, and trend forecasting in crime data using machine learning, deep learning, and statistical methods. *Neural Computing and Applications*, 37(18), 11773–11798.
<https://doi.org/10.1007/s00521-025-11094-9>
- Edgar Moyotl Hernández. (2022). *Breve introducción*.
https://www.fcfm.buap.mx/emoyotl/docs/Google_Colab_breve_introduccion.pdf
- Ersoz, F., Ersoz, T., Marcelloni, F., & Ruffini, F. (2025). Artificial Intelligence in Crime Prediction: A Survey with a Focus on Explainability. *IEEE Access*.
<https://doi.org/10.1109/ACCESS.2025.3553934>
- Farooq, U., Mushtaq, M. F., Ullah, Z., Ejaz, M. T., Akram, U., & Aslam, S. (2025). Time Series Analysis of Solar Power Generation Based on Machine Learning for Efficient Monitoring. *Engineering Reports*, 7(2). <https://doi.org/10.1002/eng2.70023>
- Forero-Corba, W., & Bennasar, F. N. (2024). Techniques and applications of Machine Learning and Artificial Intelligence in education: a systematic review. *RIED-Revista Iberoamericana de Educacion a Distancia*, 27(1), 209–253. <https://doi.org/10.5944/ried.27.1.37491>
- Forradellas, R. F. R., Alonso, S. L. N., Rodriguez, M. L., & Jorge-Vazquez, J. (2021). Applied machine learning in social sciences: Neural networks and crime prediction. *Social Sciences*, 10(1), 1–20. <https://doi.org/10.3390/socsci10010004>
- Karamolegkos, S., & Koulouriotis, D. E. (2025). Advancing short-term load forecasting with decomposed Fourier ARIMA: A case study on the Greek energy market. *Energy*, 325.
<https://doi.org/10.1016/j.energy.2025.135854>

- Kshatri, S. S., Singh, D., Narain, B., Bhatia, S., Quasim, M. T., & Sinha, G. R. (2021). An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach. *IEEE Access*, 9, 67488–67500. <https://doi.org/10.1109/ACCESS.2021.3075140>
- Lovaro Karla, & Olaya Linda. (2024). *Boletín semestral de homicidios Primer semestre de 2024*. https://oeco.pdf.org/wp-content/uploads/2024/09/Boletin-semestral-de-homicidios-Primer-semestre-de-2024_compressed.pdf
- Manatkar, A., Patel, D., Patel, H., & Manwani, N. (2025, March 5). ILAEDA: An Imitation Learning Based Approach for Automatic Exploratory Data Analysis. *Proceedings of 4th International Conference on AI-ML Systems, AIMLSystems 2024*. <https://doi.org/10.1145/3703412.3703430>
- Me, A., Bisogno, E., Malby, S., Jandl, M., Davis, P., Pysden, C., Rahmonberdiev, U., Reiterer, F., Gurian, E., Mesa Vieira, C., Aziani, A., Revisión, M. C., Gibbons Diseño, J., Kunnen, S., & Kuttnig, K. (2011). *2011 ESTUDIO MUNDIAL SOBRE EL HOMICIDIO Tendencias, Contextos, Datos*. UNODC. https://www.unodc.org/documents/data-and-analysis/statistics/Homicide/BOOK_Global_study_on_homicide_2011_Spanish_ebook.pdf
- Montalvo, D. (2024). *Ecuador registra los niveles más altos de crimen, inseguridad y delincuencia del continente*. <https://www.participacionciudadana.org/web/wp-content/uploads/2024/02/A1-Ecuador-registra-los-niveles-mas-altos-de-crimen.pdf>
- Nourmohammad Ellham, & Yousef Rasghidi. (2025). Ground data analysis for PM2.5 Prediction using predictive modeling techniques. *Journal of Air Pollution and Health*, 61–82. <https://doi.org/https://doi.org/10.18502/japh.v10i1.18095>
- Qureshi, M., Ishaq, K., Daniyal, M., Iftikhar, H., Rehman, M. Z., & Salar, S. A. A. (2025). Forecasting cardiovascular disease mortality using artificial neural networks in Sindh, Pakistan. *BMC Public Health*, 25(1). <https://doi.org/10.1186/s12889-024-21187-0>
- Qureshi, M. M. U., Ahmed, A. B., Dulmini, A., Khan, M. M. H., & Rois, R. (2025). Developing a seasonal-adjusted machine-learning-based hybrid time-series model to forecast heatwave warning. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-93227-7>
- Scrioșteanu Adriana, & Criveanu, M. M. (2025). Sustainable Transport Between Reality and Legislative Provisions—The Source for the Climate Neutrality of the European Union. *Sustainability (Switzerland)*, 17(7). <https://doi.org/10.3390/su17072814>
- Silva, D. H. da, Souza, L. R. da S., Ribeiro, C. T., Brasileiro, S. H. da S., Nardo, J. R. M., Pereira, A. A., & Andrade, A. de O. (2025). A Web Application for exploratory data analysis and classification of Parkinson's Disease patients using machine learning models on different datasets. *Software Impacts*, 23. <https://doi.org/10.1016/j.simpa.2024.100737>

- Silva, J., & Lozano, E. (2022). Análisis de las tasas de homicidio, un estudio para países latinoamericanos y europeos. *Revista Vista Económica*, 10(1), 77–86.
<https://doi.org/10.54753/rve.v10i1.1294>
- Silva Urrutia, J. E., Villalobos, M. A., Silva Urrutia, J. E., & Villalobos, M. A. (2022). An empirical analysis of homicides in Mexico through Machine Learning and statistical design of experiments. *Población y Salud En Mesoamérica*, 20(1), 1–27.
<https://doi.org/10.15517/PSM.V20I1.48217>
- UNODC. (2019, July). *UNODC Global Study on Homicide - Red Interamericana de Prevención de la Violencia y el Delito*. <https://www.oas.org/ext/es/seguridad/red-prevencion-crimen/Recursos/Biblioteca-Digital/unodc-global-study-on-homicide>
- Verrey, J., Ariel, B., Harinam, V., & Dillon, L. (2023a). Using machine learning to forecast domestic homicide via police data and super learning. *Scientific Reports*, 13(1), 1–8.
<https://doi.org/10.1038/S41598-023-50274-2/FIGURES/2>
- Verrey, J., Ariel, B., Harinam, V., & Dillon, L. (2023b). Using machine learning to forecast domestic homicide via police data and super learning. *Scientific Reports 2023 13:1*, 13(1), 1–8.
<https://doi.org/10.1038/s41598-023-50274-2>
- Villegas, A. J. R., Pabon, J. S. M., Dulce Rubio, M., Quintero, S., Vargas, J. G., & Garcia, H. (2022). Spatio Temporal Sparsity in Homicide Prediction Models. *IEEE Access*, 10, 14359–14367.
<https://doi.org/10.1109/ACCESS.2022.3143858>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Walz, C., Eifert, S., Görg, J., Schwarz, C. S., Steffan, C., Brettel, H., & Germerott, T. (2024). Gender differences in homicides. A comparative analysis of 106 fatalities in forensic autopsy data. *Forensic Science, Medicine, and Pathology*, 1–10. <https://doi.org/10.1007/S12024-024-00847-Y/FIGURES/6>

ANEXOS

Quevedo, 22 de agosto del 2025

Ing. Oviedo Bayas Byron, Ph.D.

Decano

Facultad de Posgrado

Universidad Técnica Estatal de Quevedo

Presente

De mi consideración:

El suscrito Ing. Zhuma Mera Emilio, M.Sc. En calidad de director del proyecto de Investigación titulado **MODELO PREDICTIVO PARA DETERMINACIÓN DE HOMICIDIOS INTENCIONALES EN ECUADOR, 2020 - 2024**, perteneciente al Posgradista **Ing. Sarmiento Fajardo Manuel Fabrizzio**, CERTIFICA: el cumplimiento de los parámetros establecidos por el SENESCYT, y se evidencia el reporte de la herramienta de prevención de coincidencia y/o plagio académico (Compilatio) con un porcentaje de coincidencia del cuatro por ciento (4%).

**CERTIFICADO DE ANÁLISIS**
magister

PROY_INV_SARMIENTO_M

4%

Textos sospechosos

4% Similitudes

0% similitudes entre comillas

0% entre las fuentes mencionadas

0% Idiomas no reconocidos

Nombre del documento: PROY_INV_SARMIENTO_M.pdf	Depositante: EMILIO RODRIGO ZHUMA MERA	Número de palabras: 12.920
ID del documento: a1dc20880c1325553146161368ff2959b4087901	Fecha de depósito: 21/8/2025	Número de caracteres: 87.545
Tamaño del documento original: 1,18 MB	Tipo de carga: interface	
	fecha de fin de análisis: 21/8/2025	

Ing. Zhuma Mera Emilio, M.Sc.

Director

Anexo 1 - Análisis Exploratorio de Datos (EDA)

```
#
=====
===
# SCRIPT PARA EL OBJETIVO 2: ANÁLISIS EXPLORATORIO DE DATOS
# Finalidad: Identificar variables con mayor incidencia.
#
=====
===

# -----
# FASE 0: CONFIGURACIÓN E IMPORTACIONES
# -----
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Configuración de visualización para los gráficos
sns.set_style('whitegrid')
plt.rcParams['figure.figsize'] = (12, 7)
plt.rcParams['figure.autolayout'] = True
print("--- Script para Análisis Exploratorio (Objetivo 2) ---")

# -----
# FASE 1: PREPARACIÓN Y COMPRENSIÓN INICIAL DE LOS DATOS
# -----
print("\n--- [1] Preparando y Comprendiendo los Datos ---")

try:
    df = pd.read_excel('mdi_homicidiosintencionales_pm_2020-2024_corregido.xlsx')
    print(f"Dataset cargado con {df.shape[0]} filas y {df.shape[1]} columnas.")
except FileNotFoundError:
    print("Error: Archivo no encontrado. Asegúrate de que esté en la misma carpeta.")
    exit()

# Ingeniería y Corrección de Tipos de Datos
print("\nRealizando ingeniería y corrección de tipos de datos...")
if 'fecha_infraccion' in df.columns:
    df['fecha_infraccion'] = pd.to_datetime(df['fecha_infraccion'], errors='coerce')
    df['dia_semana'] = df['fecha_infraccion'].dt.day_name()
    print("- Columna 'dia_semana' creada.")
```

```

if 'hora_infraccion' in df.columns:
    df['hora_infraccion'] = df['hora_infraccion'].astype(str).str.slice(0, 2)
    df['hora_infraccion'] = pd.to_numeric(df['hora_infraccion'], errors='coerce')
    print("- Columna 'hora_infraccion' estandarizada a formato numérico.")

# -----
# FASE 2: ANÁLISIS DESCRIPTIVO UNIVARIADO
# -----
print("\n\n--- [2] Realizando Análisis Descriptivo Univariado ---")

# Función para analizar variables categóricas
def analyze_categorical(dataframe, column_name, top_n=10):
    print(f"\nAnálisis de la variable categórica: '{column_name}'")
    freq_table = dataframe[column_name].value_counts().reset_index()
    freq_table.columns = [column_name, 'Frecuencia']
    freq_table['Porcentaje (%)'] = ((freq_table['Frecuencia'] / len(dataframe)) * 100).round(2)
    moda = dataframe[column_name].mode()[0]
    print("Tabla de Frecuencias:")
    print(freq_table.head(top_n))
    print(f"Moda: '{moda}'")
    plt.figure(figsize=(12, 6))
    sns.barplot(x='Frecuencia', y=column_name, data=freq_table.head(top_n),
palette='viridis', orient='h')
    plt.title(f'Distribución de {column_name} (Top {top_n})')
    plt.xlabel('Número de Casos'); plt.ylabel(column_name.replace('_', ' ').capitalize())
    plt.show()

# Función para analizar variables numéricas
def analyze_numerical(dataframe, column_name):
    print(f"\nAnálisis de la variable numérica: '{column_name}'")
    print("Estadísticas Descriptivas:")
    print(dataframe[column_name].describe())
    fig, axes = plt.subplots(1, 2, figsize=(15, 6))
    sns.histplot(dataframe[column_name], kde=True, ax=axes[0], color='skyblue')
    axes[0].set_title(f'Distribución de {column_name.capitalize()}')
    sns.boxplot(x=dataframe[column_name], ax=axes[1], color='lightgreen')
    axes[1].set_title(f'Diagrama de Caja de {column_name.capitalize()}')
    plt.suptitle(f'Análisis Univariado de {column_name.capitalize()}', fontsize=16)
    plt.show()

# Aplicar análisis
categorical_vars = ['tipo_muerte', 'provincia', 'arma', 'sexo', 'area_hecho', 'dia_semana']
for var in categorical_vars:
    if var in df.columns:
        num_top = 10 if var == 'provincia' else len(df[var].unique())

```

```

        analyze_categorical(df, var, top_n=num_top)

numerical_vars = ['edad', 'hora_infraccion']
for var in numerical_vars:
    if var in df.columns:
        analyze_numerical(df, var)

# -----
# FASE 3: ANÁLISIS BIVARIADO
# -----
print("\n\n--- [3] Realizando Análisis Bivariado ---")

# Relación tipo_muerte vs. arma
if 'tipo_muerte' in df.columns and 'arma' in df.columns:
    print("\nAnálisis de la relación: tipo_muerte vs. arma")
    contingency_table = pd.crosstab(df['tipo_muerte'], df['arma'], normalize='index') * 100
    print("Tabla de Contingencia (% por Tipo de Muerte):")
    print(contingency_table.round(2))
    contingency_table.plot(kind='bar', figsize=(14, 7), width=0.8)
    plt.title('Distribución del Tipo de Arma por Tipo de Muerte')
    plt.ylabel('Porcentaje (%)'); plt.xlabel('Tipo de Muerte')
    plt.xticks(rotation=45, ha='right')
    plt.legend(title='Tipo de Arma', bbox_to_anchor=(1.05, 1), loc='upper left')
    plt.show()

# Relación tipo_muerte vs. edad
if 'tipo_muerte' in df.columns and 'edad' in df.columns:
    print("\nAnálisis de la relación: tipo_muerte vs. edad")
    print("Estadísticas de Edad por Tipo de Muerte:")
    print(df.groupby('tipo_muerte')['edad'].describe())
    plt.figure(figsize=(12, 7))
    sns.boxplot(x='tipo_muerte', y='edad', data=df, palette='pastel')
    plt.title('Distribución de Edad por Tipo de Muerte')
    plt.xlabel('Tipo de Muerte'); plt.ylabel('Edad')
    plt.xticks(rotation=45, ha='right')
    plt.show()

# -----
# FASE 4: GRÁFICO DE SÍNTESIS
# -----
print("\n\n--- [4] Generando Gráfico de Síntesis del EDA ---")
g = sns.FacetGrid(df, col="tipo_muerte", col_wrap=4, height=6, aspect=0.8, sharey=False,
col_order=['ASESINATO', 'HOMICIDIO', 'FEMICIDIO', 'SICARIATO'])
g.map(sns.countplot, "arma", hue=df["sexo"], palette={"HOMBRE": "skyblue", "MUJER":
"salmon"}, order=df['arma'].value_counts().index)

```

```

g.set_axis_labels("Tipo de Arma", "Cantidad de Casos")
g.set_titles(col_template="{col_name}")
g.add_legend(title="Sexo")
g.fig.suptitle('Síntesis Visual: Interacción de Arma, Tipo de Muerte y Sexo', fontsize=16,
y=1.03)
for ax in g.axes.flatten():
    for label in ax.get_xticklabels():
        label.set_rotation(45); label.set_ha('right')
plt.tight_layout()
plt.savefig("sintesis_eda_facetgrid.png", dpi=300)
plt.show()

print("\n--- Fin del Script de Análisis Exploratorio (Objetivo 2) ---")

```

Anexo 2 - Entrenamiento y Evaluación de Modelos

```

#
=====

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from prophet import Prophet
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.tsa.api import ExponentialSmoothing
from sklearn.metrics import mean_absolute_error, mean_squared_error
import warnings

warnings.filterwarnings('ignore')
sns.set_style('whitegrid')
plt.rcParams['figure.figsize'] = (15, 7)
print("--- Script de Entrenamiento y Evaluación Comparativa (Objetivo 3) ---")

# -----
# FASE 1: PREPARACIÓN DE DATOS (Carga y creación de las 4 series)
# -----
print("\n--- [FASE 1] Preparando las 4 Series Temporales ---")
# (Este bloque es idéntico al del script anterior para crear las 4 series)
try:
    df_raw = pd.read_excel('/content/mdi_homicidiosintencionales_pm_2020-
2024_corregido.xlsx')

```

```

except FileNotFoundError:
    print("Error: Archivo no encontrado."); exit()
df_raw['fecha_infraccion'] = pd.to_datetime(df_raw['fecha_infraccion'], errors='coerce')
df_raw.dropna(subset=['fecha_infraccion'], inplace=True)
df_raw = df_raw.set_index('fecha_infraccion')
ts_total = df_raw.resample('M').size().rename('total_homicidios')
ts_arma_fuego = (df_raw[df_raw['arma'] == 'ARMA DE FUEGO'].resample('M').size() /
ts_total * 100).rename('pct_arma_fuego')
ts_edad_promedio = df_raw['edad'].resample('M').mean().rename('edad_promedio')
ts_guayas = df_raw[df_raw['provincia'] == 'GUAYAS'].resample('M').size().rename('total_homicidios_guayas')
all_ts = {"Volumen": ts_total, "Tipo de Arma": ts_arma_fuego, "Sociodemográfico":
ts_edad_promedio, "Geográfico": ts_guayas}

# -----
# FASE 2: BUCLE DE ENTRENAMIENTO Y EVALUACIÓN PARA CADA PATRÓN
# -----
print("\n--- [FASE 2] Iniciando Entrenamiento y Evaluación para cada Patrón ---")

best_models = {}

for pattern_name, ts_data in all_ts.items():
    print(f"\n\n===== PROCESANDO PATRÓN: {pattern_name}
=====")

    # Preparar datos
    df_forecast = ts_data.reset_index()
    df_forecast.columns = ['ds', 'y']
    df_forecast.fillna(method='ffill', inplace=True) # Rellenar posibles nulos
    train, test = df_forecast[0:-12], df_forecast[-12:]

    # --- Entrenamiento y Predicción de los 3 Modelos ---

    # Prophet
    m_prophet = Prophet(yearly_seasonality=True, daily_seasonality=False,
weekly_seasonality=False)
    m_prophet.fit(train)
    future_prophet = m_prophet.make_future_dataframe(periods=12, freq='M')
    forecast_prophet = m_prophet.predict(future_prophet)
    y_pred_prophet = forecast_prophet['yhat'][-12:].values

    # SARIMA
    m_sarima = SARIMAX(train['y'], order=(1, 1, 1), seasonal_order=(1, 1, 0, 12))
    sarima_fit = m_sarima.fit(dispatch=False)
    y_pred_sarima = sarima_fit.get_forecast(steps=12).predicted_mean.values

```

```

# Holt-Winters
m_hw = ExponentialSmoothing(train['y'], seasonal_periods=12, trend='add',
seasonal='add', initialization_method="estimated")
hw_fit = m_hw.fit()
y_pred_hw = hw_fit.forecast(12).values

# --- Comparación Visual ---
plt.figure(figsize=(15, 8))
plt.plot(train['ds'], train['y'], label='Histórico (Entrenamiento)', color='gray')
plt.plot(test['ds'], test['y'], label='Real (Prueba)', color='black', linewidth=2, marker='o')
plt.plot(test['ds'], y_pred_prophet, label='Predicción Prophet', linestyle='--')
plt.plot(test['ds'], y_pred_sarima, label='Predicción SARIMA', linestyle='--')
plt.plot(test['ds'], y_pred_hw, label='Predicción Holt-Winters', linestyle='--')
plt.title(f'Comparación de Modelos para el Patrón: {pattern_name}')
plt.legend()
plt.show()

# --- Comparación de Métricas ---
metrics = {}
models = {'Prophet': y_pred_prophet, 'SARIMA': y_pred_sarima, 'Holt-Winters':
y_pred_hw}
for name, y_pred in models.items():
    mae = mean_absolute_error(test['y'], y_pred)
    rmse = np.sqrt(mean_squared_error(test['y'], y_pred))
    mape = np.mean(np.abs((test['y'] - y_pred) / test['y'])) * 100 if np.all(test['y'] != 0) else
np.inf
    metrics[name] = {'MAE': mae, 'RMSE': rmse, 'MAPE (%)': mape}

metrics_df = pd.DataFrame(metrics).T.sort_values(by='MAPE (%)')
print(f"\n--- Tabla Comparativa de Métricas para el Patrón: {pattern_name} ---")
print(metrics_df)

best_model_name = metrics_df.index[0]
best_models[pattern_name] = best_model_name
print(f"\nMejor modelo para '{pattern_name}': {best_model_name}")

# -----
# FASE 3: SÍNTESIS DE LA SELECCIÓN DE MODELOS
# -----
print("\n\n--- [FASE 3] Síntesis Final de la Selección de Modelos ---")
best_models_df = pd.DataFrame(list(best_models.items()), columns=['Patrón', 'Mejor
Modelo Seleccionado'])
print(best_models_df)

```

```
print("\n--- Fin del Script de Entrenamiento y Evaluación (Objetivo 3) ---")
```

Anexo 3 - Finalización del Modelo y Análisis de Patrones

```
# -----
# FASE 0: CONFIGURACIÓN
# -----
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from prophet import Prophet
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.tsa.api import ExponentialSmoothing
import warnings

warnings.filterwarnings('ignore')
sns.set_style('whitegrid')
plt.rcParams['figure.figsize'] = (15, 7)
print("--- Script de Proyección Final (Objetivo 4) ---")

# -----
# FASE 1: PREPARACIÓN DE DATOS (Carga y creación de las 4 series)
# -----
print("\n--- [FASE 1] Preparando las 4 Series Temporales ---")
try:
    df_raw = pd.read_excel('/content/mdi_homicidiosintencionales_pm_2020-2024_corregido.xlsx')
except FileNotFoundError:
    print("Error: Archivo no encontrado."); exit()

df_raw['fecha_infraccion'] = pd.to_datetime(df_raw['fecha_infraccion'], errors='coerce')
df_raw.dropna(subset=['fecha_infraccion'], inplace=True)
df_raw = df_raw.set_index('fecha_infraccion')
ts_total = df_raw.resample('M').size().rename('total_homicidios')
ts_arma_fuego = (df_raw[df_raw['arma'] == 'ARMA DE FUEGO'].resample('M').size() /
ts_total * 100).rename('pct_arma_fuego')
ts_edad_promedio = df_raw['edad'].resample('M').mean().rename('edad_promedio')
ts_guayas = df_raw[df_raw['provincia'] == 'GUAYAS'].resample('M').size().rename('total_homicidios_guayas')
all_ts = {
    "Volumen": ts_total,
    "Tipo de Arma": ts_arma_fuego,
    "Sociodemográfico": ts_edad_promedio,
```

```

    "Geográfico": ts_guayas
}
print("Datos listos para la proyección final.")

# -----
# FASE 2: DEFINIR LOS MEJORES MODELOS (Resultado del Objetivo 3)
# -----
# **IMPORTANTE**: Debes actualizar este diccionario con los resultados REALES
# que obtuviste al correr el script del Objetivo 3.
# Este es un ejemplo basado en nuestros resultados hipotéticos.
best_models = {
    "Volumen": "SARIMA",
    "Tipo de Arma": "Holt-Winters", # Ejemplo, podría ser otro
    "Sociodemográfico": "Prophet", # Ejemplo, podría ser otro
    "Geográfico": "SARIMA" # Ejemplo, podría ser otro
}
print("\nModelos seleccionados para la proyección final:")
for pattern, model in best_models.items():
    print(f"- Patrón '{pattern}': se usará el modelo {model}")

# -----
# FASE 3: BUCLE DE RE-ENTRENAMIENTO Y PROYECCIÓN FINAL
# -----
print("\n\n--- [FASE 3] Iniciando Proyección Final para cada Patrón ---")

for pattern_name, ts_data in all_ts.items():
    print(f"\n\n===== GENERANDO PROYECCIÓN PARA: {pattern_name} =====")

    # Preparar datos (DataFrame completo)
    df_forecast = ts_data.reset_index()
    df_forecast.columns = ['ds', 'y']
    df_forecast.fillna(method='ffill', inplace=True)

    # Seleccionar el mejor modelo para este patrón
    model_name = best_models[pattern_name]
    print(f"Re-entrenando el modelo '{model_name}' con el 100% de los datos...")

    # Re-entrenar y predecir según el modelo seleccionado
    if model_name == 'Prophet':
        final_model = Prophet(yearly_seasonality=True, daily_seasonality=False,
                               weekly_seasonality=False)
        final_model.fit(df_forecast)
        future_df = final_model.make_future_dataframe(periods=60, freq='M')
        forecast_df = final_model.predict(future_df)

```

```

# Graficar el pronóstico de Prophet
fig = final_model.plot(forecast_df)
plt.title(f'Proyección a 5 Años del Patrón: {pattern_name} (con {model_name})')
plt.xlabel('Fecha')
plt.ylabel(f'Valor Proyectado ({pattern_name})')
from prophet.plot import add_changepoints_to_plot
add_changepoints_to_plot(fig.gca(), final_model, forecast_df)

elif model_name == 'SARIMA':
    final_model = SARIMAX(df_forecast['y'], order=(1, 1, 1), seasonal_order=(1, 1, 0, 12))
    final_fit = final_model.fit(dispatch=False)
    forecast_obj = final_fit.get_forecast(steps=60)
    forecast_mean = forecast_obj.predicted_mean
    forecast_ci = forecast_obj.conf_int()

    # Graficar el pronóstico de SARIMA
    future_dates = pd.date_range(start=df_forecast['ds'].iloc[-1], periods=61, freq='M')[1:]
    plt.figure(figsize=(15, 7))
    plt.plot(df_forecast['ds'], df_forecast['y'], label='Histórico')
    plt.plot(future_dates, forecast_mean, color='red', label='Pronóstico')
    plt.fill_between(future_dates, forecast_ci.iloc[:, 0], forecast_ci.iloc[:, 1], color='pink',
alpha=0.5, label='Intervalo de Confianza')
    plt.title(f'Proyección a 5 Años del Patrón: {pattern_name} (con {model_name})')
    plt.xlabel('Fecha')
    plt.ylabel(f'Valor Proyectado ({pattern_name})')
    plt.legend()

elif model_name == 'Holt-Winters':
    final_model = ExponentialSmoothing(df_forecast['y'], seasonal_periods=12,
trend='add', seasonal='add', initialization_method="estimated")
    final_fit = final_model.fit()
    forecast_values = final_fit.forecast(60)

    # Graficar el pronóstico de Holt-Winters
    future_dates = pd.date_range(start=df_forecast['ds'].iloc[-1], periods=61, freq='M')[1:]
    plt.figure(figsize=(15, 7))
    plt.plot(df_forecast['ds'], df_forecast['y'], label='Histórico')
    plt.plot(future_dates, forecast_values, color='green', label='Pronóstico')
    # Holt-Winters no provee intervalos de confianza de forma tan directa
    plt.title(f'Proyección a 5 Años del Patrón: {pattern_name} (con {model_name})')
    plt.xlabel('Fecha')
    plt.ylabel(f'Valor Proyectado ({pattern_name})')
    plt.legend()

```

```

# Guardar el gráfico
filename = f"proyeccion_final_{pattern_name.lower().replace(' ', '_')}.png"
plt.savefig(filename, dpi=300, bbox_inches='tight')
print(f"Gráfico de proyección guardado como: {filename}")
plt.show()

print("\n--- Fin del Script de Proyección Final ---")

```

ANEXO 4 – Código resumido

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from prophet import Prophet
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.tsa.api import ExponentialSmoothing
from sklearn.metrics import mean_absolute_error, mean_squared_error
import warnings

# Ignorar advertencias para una salida más limpia
warnings.filterwarnings('ignore')
sns.set_style('whitegrid')
plt.rcParams['figure.figsize'] = (15, 7)
print("--- INICIANDO SCRIPT COMPLETO DE ANÁLISIS Y PRONÓSTICO ---")

# -----
# FASE 1: CARGA, LIMPIEZA Y TRANSFORMACIÓN A SERIES TEMPORALES
# -----
print("\n--- [FASE 1] Preparando Datos para Pronóstico ---")
try:
    df_raw = pd.read_excel('mdi_homicidiosintencionales_pm_2020-2024_corregido.xlsx')
except FileNotFoundError:
    print("Error: Archivo 'mdi_homicidiosintencionales_pm_2020-2024_corregido.xlsx' no encontrado."); exit()

df_raw['fecha_infraccion'] = pd.to_datetime(df_raw['fecha_infraccion'], errors='coerce')
df_raw.dropna(subset=['fecha_infraccion'], inplace=True)
df_raw = df_raw.set_index('fecha_infraccion')

# Creación de las 4 Series Temporales para el análisis
ts_total = df_raw.resample('M').size().rename('Volumen')
ts_arma_fuego = (df_raw[df_raw['arma'] == 'ARMA DE FUEGO'].resample('M').size() / ts_total * 100).rename('Tipo de Arma')
ts_edad_promedio = df_raw['edad'].resample('M').mean().rename('Sociodemográfico')
ts_guayas = df_raw[df_raw['provincia'] == 'GUAYAS'].resample('M').size().rename('Geográfico')
all_ts = {"Volumen": ts_total, "Tipo de Arma": ts_arma_fuego, "Sociodemográfico": ts_edad_promedio, "Geográfico": ts_guayas}

```

```

print("Las 4 series temporales han sido creadas exitosamente.")

# =====
# OBJETIVO 2: ANÁLISIS EXPLORATORIO Y DE TENDENCIAS
# =====
print("\n\n--- [OBJETIVO 2] Generando Resultados del Análisis Exploratorio ---")

# Gráfico de las 4 tendencias
fig, axes = plt.subplots(4, 1, figsize=(16, 22), sharex=True)
# ... (código del gráfico de 4 tendencias, igual que antes) ...
plt.savefig("eda_4_tendencias.png", dpi=300)
print("- Gráfico de tendencias guardado como 'eda_4_tendencias.png'")

# Mapa de calor de correlaciones
df_eda_ts = pd.concat(all_ts.values(), axis=1).fillna(0)
corr_matrix_ts = df_eda_ts.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(corr_matrix_ts, annot=True, cmap='viridis', fmt=".2f")
plt.title('Mapa de Calor de Correlación entre las Tendencias de los Patrones')
plt.savefig("correlacion_tendencias.png", dpi=300)
print("- Mapa de calor de correlaciones guardado como 'correlacion_tendencias.png'")
plt.close('all') # Cerrar todas las figuras para liberar memoria

# =====
# OBJETIVO 3: ENTRENAMIENTO Y EVALUACIÓN COMPARATIVA DE MODELOS
# =====
print("\n\n--- [OBJETIVO 3] Iniciando Entrenamiento y Evaluación Comparativa ---")

best_models = {}
for pattern_name, ts_data in all_ts.items():
    print(f"\n==== Procesando Patrón: {pattern_name} =====")
    df_forecast = ts_data.reset_index(); df_forecast.columns = ['ds', 'y']
    df_forecast.fillna(method='ffill', inplace=True)
    train, test = df_forecast[0:-12], df_forecast[-12:]

    # Entrenar y predecir con los 3 modelos
    # Prophet
    m_prophet = Prophet(yearly_seasonality=True, daily_seasonality=False,
weekly_seasonality=False).fit(train)
    y_pred_prophet = m_prophet.predict(m_prophet.make_future_dataframe(periods=12,
freq='M'))['yhat'][-12:].values
    # SARIMA
    m_sarima = SARIMAX(train['y'], order=(1, 1, 1), seasonal_order=(1, 1, 0, 12)).fit(dispatch=False)
    y_pred_sarima = m_sarima.get_forecast(steps=12).predicted_mean.values
    # Holt-Winters
    m_hw = ExponentialSmoothing(train['y'], seasonal_periods=12, trend='add', seasonal='add',
initialization_method="estimated").fit()
    y_pred_hw = m_hw.forecast(12).values

```

```

# Generar gráfico comparativo
plt.figure(figsize=(15, 8))
plt.plot(train['ds'], train['y'], label='Histórico (Entrenamiento)', color='gray')
plt.plot(test['ds'], test['y'], label='Real (Prueba)', color='black', linewidth=2, marker='o')
plt.plot(test['ds'], y_pred_prophet, label='Predicción Prophet', linestyle='--')
plt.plot(test['ds'], y_pred_sarima, label='Predicción SARIMA', linestyle='--')
plt.plot(test['ds'], y_pred_hw, label='Predicción Holt-Winters', linestyle='--')
plt.title(f'Comparación de Modelos para el Patrón: {pattern_name}')
plt.legend()
filename_comp = f'comp_{pattern_name.lower().replace(' ', '_')}.png'
plt.savefig(filename_comp, dpi=300)
print(f'- Gráfico de comparación guardado como '{filename_comp}''')

# Calcular métricas y seleccionar el mejor
metrics = {}
models_preds = {'Prophet': y_pred_prophet, 'SARIMA': y_pred_sarima, 'Holt-Winters':
y_pred_hw}
for name, y_pred in models_preds.items():
    mae = mean_absolute_error(test['y'], y_pred)
    rmse = np.sqrt(mean_squared_error(test['y'], y_pred))
    mape = np.mean(np.abs((test['y'] - y_pred) / test['y'])) * 100 if np.all(test['y'] != 0) else np.inf
    metrics[name] = {'MAE': mae, 'RMSE': rmse, 'MAPE (%)': mape}
metrics_df = pd.DataFrame(metrics).T.sort_values(by='MAPE (%)')
print(f'Tabla de Métricas para '{pattern_name}':\n{metrics_df}')
best_models[pattern_name] = metrics_df.index[0]
plt.close('all')

print("\n--- Resumen de Mejores Modelos Seleccionados ---")
print(pd.DataFrame(list(best_models.items()), columns=['Patrón', 'Mejor Modelo']))

# =====
# OBJETIVO 4: PROYECCIÓN FINAL CON LOS MEJORES MODELOS
# =====
print("\n\n--- [OBJETIVO 4] Generando Proyecciones Finales a 5 Años ---")

for pattern_name, ts_data in all_ts.items():
    print(f"\n==== Generando Proyección para: {pattern_name} =====")
    df_forecast = ts_data.reset_index(); df_forecast.columns = ['ds', 'y']
    df_forecast.fillna(method='ffill', inplace=True)
    model_name = best_models[pattern_name]

    # Re-entrenar y predecir con el modelo ganador
    if model_name == 'Prophet':
        final_model = Prophet(yearly_seasonality=True, daily_seasonality=False,
weekly_seasonality=False).fit(df_forecast)
        forecast_df = final_model.predict(final_model.make_future_dataframe(periods=60, freq='M'))
        fig = final_model.plot(forecast_df)
        from prophet.plot import add_changepoints_to_plot
        add_changepoints_to_plot(fig.gca(), final_model, forecast_df)

```

```

elif model_name == 'SARIMA':
    final_model = SARIMAX(df_forecast['y'], order=(1, 1, 1), seasonal_order=(1, 1, 0,
12)).fit(dispatch=False)
    forecast_obj = final_model.get_forecast(steps=60)
    future_dates = pd.date_range(start=df_forecast['ds'].iloc[-1], periods=61, freq='M')[1:]
    plt.figure(figsize=(15, 7))
    plt.plot(df_forecast['ds'], df_forecast['y'], label='Histórico')
    plt.plot(future_dates, forecast_obj.predicted_mean, color='red', label='Pronóstico')
    plt.fill_between(future_dates, forecast_obj.conf_int().iloc[:, 0], forecast_obj.conf_int().iloc[:,
1], color='pink', alpha=0.5, label='Intervalo de Confianza')
    plt.legend()
elif model_name == 'Holt-Winters':
    final_model = ExponentialSmoothing(df_forecast['y'], seasonal_periods=12, trend='add',
seasonal='add', initialization_method="estimated").fit()
    forecast_values = final_model.forecast(60)
    future_dates = pd.date_range(start=df_forecast['ds'].iloc[-1], periods=61, freq='M')[1:]
    plt.figure(figsize=(15, 7))
    plt.plot(df_forecast['ds'], df_forecast['y'], label='Histórico')
    plt.plot(future_dates, forecast_values, color='green', label='Pronóstico')
    plt.legend()

plt.title(f'Proyección a 5 Años del Patrón: {pattern_name} (con {model_name})')
filename_proy = f"proy_{pattern_name.lower().replace(' ', '_')}.png"
plt.savefig(filename_proy, dpi=300)
print(f"- Gráfico de proyección guardado como '{filename_proy}'")
plt.close('all')

print("\n\n--- SCRIPT COMPLETO FINALIZADO ---")

```