



**ESCUELA DE NEGOCIOS**

**MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA DE DATOS**

**DESARROLLO DE UN MODELO PREDICTIVO QUE CUANTIFIQUE LOS EVENTOS  
PELIGROSOS POR PROVINCIA**

**Profesor**

**Víctor Manuel Gonzales Holguin**

**Autor**

**Wilmer Santiago Crisanto Ñacata**

**2024**

## RESUMEN

Este estudio aborda la gestión de desastres desde una perspectiva integral, enfatizando la necesidad de mejorar la predicción y respuesta ante eventos adversos. Se destaca la importancia de integrar tecnologías avanzadas, como las redes sociales y la inteligencia artificial, para optimizar la gestión de desastres. A través de una revisión de literatura, identificamos carencias en la preparación y recuperación, sugiriendo que métodos específicos de aprendizaje automático pueden ser claves en las diferentes fases de gestión. Este trabajo propone un enfoque colaborativo interdisciplinario y el uso de datos de calidad para fortalecer las estrategias de respuesta ante desastres.

## ABSTRACT

Disaster management presents a complex challenge that requires a multidisciplinary approach to mitigate impacts on vulnerable societies and economies. This research explores the critical dimensions of disaster management, focusing on prediction and response strategies. Through comprehensive literature review, we reveal a significant gap in preparation and recovery phases, highlighting the potential of social media and artificial intelligence to enhance prediction and response capabilities. We advocate for the adoption of phase-specific machine learning methods, emphasizing the need for interdisciplinary collaboration and high-quality data. This study aims to contribute to a more robust and adaptable disaster management framework, ultimately aiming to save lives and minimize damages in future events.

## ÍNDICE DE CONTENIDO

ÍNDICE DE CONTENIDO.....	1
ÍNDICE DE FIGURAS.....	2
1. INTRODUCCIÓN.....	3
2. REVISIÓN DE LITERATURA.....	4
3. IDENTIFICACIÓN DEL OBJETO DE ESTUDIO.....	10
4. PLANTEAMIENTO DEL PROBLEMA.....	11
5. OBJETIVO GENERAL.....	12
6. OBJETIVOS ESPECÍFICOS.....	13
7. JUSTIFICACIÓN Y APLICACIÓN DE LA METODOLOGÍA.....	14
7.1. Modelo ARIMA.....	15
7.2. Metodología.....	16
7.2.1. Guayas.....	22
7.2.2. Pichincha.....	29
7.2.3. Azuay.....	30
7.2.4. El Oro.....	31
8. RESULTADOS.....	33
9. DISCUSIÓN DE LOS RESULTADOS Y PROPUESTA DE SOLUCIÓN.....	36
9.1. Análisis de Tendencias Temporales.....	36
9.2. Comparación por Provincia.....	36
9.3. Evaluación de la Incertidumbre.....	37
9.4. IMPLICACIONES SOBRE INNOVACION EMPRESARIAL.....	37
10. CONCLUSIONES Y RECOMENDACIONES.....	40
11. REFERENCIAS.....	42

## ÍNDICE DE FIGURAS

Figura 1. Visualización General Dataset.....	17
Figura 2. Distribuciones variables cuantitativas.....	18
Figura 3. Distribución por cantón.....	18
Figura 4. Distribución por parroquia. ....	19
Figura 5. Distribución de eventos peligrosos por provincia 2022.....	20
Figura 6. Tendencia mensual de eventos peligrosos por provincia. ....	21
Figura 7. Cantidad de eventos peligrosos por periodo. ....	22
Figura 8. Estadísticas de estacionariedad. ....	22
Figura 9. Autocorrelación ACF.....	24
Figura 10. Autocorrelación parcial PACF.....	24
Figura 11. Resumen evaluación del modelo. ....	25
Figura 12. Resumen evaluación modelo 2. ....	26
Figura 13. Comparación parámetros AIC y BIC. ....	26
Figura 14. Gráfico de residuos del modelo ARIMA. ....	27
Figura 15. Autocorrelación de residuos del modelo ARIMA.....	27
Figura 16. Datos reales vs predicciones de Guayas. ....	28
Figura 17. Predicciones por periodo de Guayas.....	29
Figura 18. Datos reales vs predicciones de Pichincha.....	30
Figura 19. Predicciones por periodo de Pichincha.....	30
Figura 20. Datos reales vs predicciones de Azuay.....	31
Figura 21. Predicciones por periodo de Azuay.....	31
Figura 22. Datos reales vs predicciones de El Oro. ....	32
Figura 23. Predicciones por periodo de El Oro.....	32
Figura 24. Vista general de predicciones por periodo y provincia.....	33
Figura 25. Comparación de predicciones por provincia.....	34
Figura 26. Comportamiento de los límites y predicciones.....	35

# 1. INTRODUCCIÓN

La gestión eficaz de desastres es un desafío global que requiere un enfoque multidisciplinario para mitigar los impactos negativos en las sociedades y economías vulnerables. Este estudio explora las dimensiones críticas de la gestión de desastres, centrándose en la importancia de la predicción, evaluación de vulnerabilidades, y estrategias de respuesta. A través de una revisión exhaustiva de la literatura, identificamos una brecha significativa en la fase de preparación y recuperación, destacando la necesidad de integrar tecnologías avanzadas como las redes sociales y la inteligencia artificial para fortalecer las capacidades de predicción y respuesta. Este trabajo subraya la relevancia de adoptar métodos de aprendizaje automático específicos para cada fase del ciclo de gestión de desastres, enfatizando la importancia de la colaboración interdisciplinaria y el acceso a datos de alta calidad. Al hacerlo, se espera contribuir al desarrollo de un marco más robusto y adaptable para la gestión de desastres, que pueda minimizar daños ante eventos futuros.

## 2. REVISIÓN DE LITERATURA

Los estudios propuestos tenían como su principal objetivo abordar diversas áreas dentro de la gestión de desastres. Estas áreas incluyen la predicción de desastres y riesgos, la evaluación de la vulnerabilidad, la detección de desastres, la implementación de sistemas de alerta temprana, el monitoreo de desastres, la evaluación de daños y la respuesta después de un desastre. Dentro de estos estudios de investigación, se observó que la mayoría se enfocaron en las operaciones de respuesta, superando en número a aquellos centrados en otras etapas de la gestión de desastres. La mayoría de estos estudios se dirigieron específicamente a la fase de respuesta después de un desastre, representando un 38,2% del total. A esto le siguió la evaluación de daños con un 20%, la evaluación de riesgos y vulnerabilidad con un 14,5%, la predicción de desastres y peligros con un 9,09%, los sistemas de alerta temprana, la detección de desastres y el monitoreo de desastres, cada uno con un 5,45%. En resumen, más de la mitad de los estudios de investigación se enfocaron en la respuesta a desastres, seguida por la mitigación de desastres y, finalmente, la preparación para desastres.

La disponibilidad de información sobre desastres a través de las redes sociales facilita a las entidades gubernamentales la preparación y asignación de recursos necesarios para abordar situaciones de emergencia, lo que posibilita una comprensión más completa de la extensión y gravedad de dichos eventos (Sit, Koylu, & Demir, 2019). Por otro lado, diferentes investigaciones se encuentran especialmente centrados en accidentes de tráfico y emplean un enfoque de modelado predictivo de accidentes basado en macrodatos, siendo un componente esencial para la formulación de estrategias en materia de seguridad vial, además, destacan la posibilidad de conexión con modelos de predicción de la gravedad de lesiones (Yasir, Hussain, & Haque, 2024). Esto se considera un punto de partida, ya que dichos modelos de predicción pueden ser utilizados para la predicción de estrategias para la gestión de desastres naturales. Adicionalmente y enfocados en IA, señalan que la inteligencia artificial tiene el potencial de mejorar la gestión de desastres naturales, no obstante, es esencial

comprender y abordar sus limitaciones para aprovechar plenamente sus beneficios, y para lograrlo, se destaca la necesidad de una colaboración interdisciplinaria entre diversas partes interesadas. Además, se indica que la utilización de algoritmos basados en inteligencia artificial para la detección o predicción de eventos puede lograrse con conjuntos de datos de alta calidad, mediante la combinación de diversas fuentes de datos o técnicas de modelado (Kuglitsch, Pelivan, Ceola, Menon, & Xoplaki, 2022). Es importante mencionar que los métodos de aprendizaje automático se han aplicado para reconocer el comportamiento general de una multitud durante evacuaciones. Estos métodos permiten detectar el estado del flujo de multitudes y estimar la aparición de anomalías, como congestión en las rutas de evacuación o movimientos direccionales concentrados. La capacidad del aprendizaje automático para calcular las rutas óptimas y desarrollar soluciones matemáticas aborda eficazmente problemas asociados con diversos parámetros de evacuación (Lu, Christie, Nguyen, Freeman, & Hsu, 2022).

Al profundizar en el análisis de los enfoques utilizados en las fuentes revisadas y su precisión, se pueden identificar ciertos métodos que se destacan por su rendimiento preciso y que arrojan resultados superiores dependiendo del tipo de desastre que se aborde. En el contexto de la preparación para desastres, se logró una asombrosa precisión del 99,93% al emplear el método de regresión logística (LR) en el aprendizaje automático (ML). En cuanto a la mitigación de desastres, el método más efectivo de ML resultó ser KNN, con una precisión del 92,74%. Finalmente, en la fase de respuesta a desastres, se alcanzó la máxima precisión, llegando al 99%, mediante el método de Máquinas de Vectores de Soporte (SVM) en ML.

Es crucial destacar que, en todos los estudios analizados, los métodos de ML obtuvieron sus resultados más precisos cuando se utilizaron datos estructurados. Esto subraya la eficacia del ML en entornos con datos estructurados. Existen varios factores que influyen en la efectividad general de estos métodos. Por ejemplo, los datos no estructurados suelen ser más desafiantes de analizar, procesar previamente y modelar, lo que conlleva a un rendimiento deficiente y la posibilidad de un sobreajuste cuando se aplican en



sistemas basados en ML. Un sistema entrenado con conjuntos de datos de baja calidad podría generar confusiones al ser utilizado en situaciones del mundo real debido a sus resultados sesgados, que son una consecuencia de los datos de entrenamiento deficientes, por lo tanto, es apropiado comparar casos específicos, como se presenta en parte de la literatura revisada, para determinar qué método se adecua mejor a cada escenario (Fan, Wu, & Mostafani, 2020).

Enfocados en ML y DL (Linardos, Drakaki, Tzionas, & Karnavas, 2022), muestran un mejor rendimiento en datos estructurados. Aunque este análisis destaca los métodos de ML/DL más efectivos en términos de precisión en cada fase del desastre, varios factores influyen en la eficacia global de los métodos utilizados. Por ejemplo, los datos no estructurados suelen ser más desafiantes de analizar, procesar y modelar, lo que puede resultar en un rendimiento inferior y posiblemente sobreajuste al integrarlos en un sistema basado en ML/DL. Un sistema entrenado con conjuntos de datos de baja calidad podría generar resultados sesgados en aplicaciones del mundo real debido a la baja calidad de los datos de entrenamiento. Por otro lado, (Kameswari, 2023) indican que el manejo de desastres naturales y el cambio climático representan desafíos globales significativos, y el empleo de métodos de aprendizaje automático (ML) para abordar estos problemas ofrece perspectivas alentadoras. Sin embargo, es importante destacar que el aprendizaje automático no es una solución rápida para combatir el cambio climático y los desastres naturales, sino que debe ser parte de un enfoque más amplio que incluya métodos científicos probados, toma de decisiones informada y participación comunitaria. La implementación de sistemas de aprendizaje automático debe abordar consideraciones éticas, protección de datos y reducción de sesgos.

Estudios enfocados en clasificación de tweets (Iparraguirre, Melgarejo, Castro, & Olaya, 2023) definen que el análisis exploratorio de datos es esencial en el proceso de análisis de datos, ya que ofrece una comprensión inicial completa de los datos y establece una base sólida para análisis subsiguientes. Este proceso implica la visualización de datos previa al entrenamiento de modelos, la aplicación de estadísticas descriptivas y el estudio de tendencias. A través de esta técnica, es viable detectar valores atípicos, patrones y relaciones en los

datos. Algunos ejemplos centrados en desastres naturales, como lo menciona (Huang, Jin, Sheng Zhao, & Yan Huang, 2018) dónde indican que la exactitud en la predicción de la precipitación durante ciclones tropicales es crucial para la gestión de desastres y la planificación de emergencias. En este sentido, se han creado y examinado diversos métodos con el fin de perfeccionar la precisión de estas proyecciones. Dentro de este abanico de técnicas, la Fuzzy Neural Network (FNN) y el Algoritmo LLE han surgido como alternativas prometedoras, ya que resaltan por su habilidad para modelar la intrincada relación entre las variables meteorológicas y la precipitación en estos fenómenos extremos. Por otro lado, (Sankaranarayanan, Prabhakar, Satish, & Jain, 2020) emplearon una red neuronal profunda para pronosticar la ocurrencia de inundaciones en función de la temperatura y las precipitaciones. Sobre la base de la predicción de la probabilidad de que se produzcan inundaciones, se puede proporcionar una alerta para la evacuación que puede salvar vidas humanas y propiedades. El conjunto de datos constaba de grandes cantidades de datos de lluvia observados y de factores recopilados como la temperatura mínima y máxima y la ocurrencia de inundaciones. Estudios sobre terremotos, permiten también predecir el número de víctimas suscitadas por este desastre natural, como lo señala (Huang, Song, & Jin, 2020) en dónde indican que el enfoque tradicional para predecir víctimas de desastres sísmicos requiere datos de muestra rigurosos y la configuración manual de numerosos parámetros, lo que conlleva a una precisión deficiente y una velocidad de aprendizaje lenta. En este artículo, se presenta la introducción de la Extreme Learning Machine (ELM) para mejorar la precisión en la predicción de víctimas de terremotos. Mediante el entrenamiento de datos, se establece la estructura de la red ELM para la predicción de víctimas de terremotos, determinando el número de nodos en la capa oculta y la función de excitación, lo que asegura la confiabilidad de los resultados de predicción obtenidos por la red ELM. En contraste con el estudio de (Amin & Ahn, 2021) cuyos hallazgos experimentales resaltan la relevancia del trabajo de proporcionar a las personas información sobre los objetos en sus salas de estar mediante la asignación de etiquetas de riesgo, con el objetivo de garantizar la seguridad humana. La anotación de etiquetas de riesgo en los objetos se

presenta como un indicador efectivo para determinar si un objeto representa algún riesgo para los seres humanos, facilitando la comprensión de las precauciones de seguridad necesarias. Otros estudios centrados en la predicción de tormentas de arena, como el caso de (Shaiba, Alaashoub, & Alzahrani, 2018), en dónde comparan los resultados de los modelos en Riad, Dammam y Jeddah y seleccionamos el modelo con mejor rendimiento, que resultó ser el modelo CART para predicciones en tiempo real. Luego, implementaron este modelo en su página web para mostrar predicciones de tormentas de arena en tiempo real con hasta 24 horas de antelación. Crearon un modelo para cada ventana de tiempo, generando así 24 modelos para cada una de las tres ciudades. El resultado final es un sistema capaz de predecir tormentas de arena y alertar a las personas antes de su llegada, permitiéndoles tomar las precauciones necesarias. Por otra parte, Los resultados del estudio de (Ataollah, Dieu, Binh, Karim, & Kamran, 2017) mencionan que la confiabilidad y efectividad del clasificador NBT propuesto y del conjunto de modelos RS-NBT para mejorar la precisión tanto del clasificador NBT como de los mapas de susceptibilidad a deslizamientos de tierra. Determinaron, además, que el mapa de susceptibilidad a deslizamientos de tierra elaborado mediante este enfoque posee un potencial prometedor para la evaluación de regiones propensas a estos eventos. Por lo tanto, esta técnica utilizada por sus autores podría incrementar la capacidad predictiva no solo para eventos de deslizamientos de tierra, sino también para otros estudios de peligros geo ambientales como inundaciones, surgimientos de manantiales, terremotos y erosión del suelo. Finalmente, algo importante que se debe destacar en base a (Deparday, Gevaert, Molinario, Soden, & Balog-Way, 2019) cuyo estudio indica que es importante señalar que algunos algoritmos de ML, especialmente los de aprendizaje profundo, no nos dan una idea de qué variables de entrada son importantes o qué relaciones entre variables condujeron a un resultado específico. Por el contrario, cuando se utilizan regresión lineal de mínimos cuadrados ordinarios o árboles de decisión, por ejemplo, queda claro qué características explican mejor un resultado específico del modelo. Además, la mejora del rendimiento de los métodos desarrollados

puede lograrse mediante la integración de grandes conjuntos de datos procedentes de diversas fuentes.

Para el presente proyecto se la base de datos que se va a utilizar proviene de la Secretaria de Gestión de Riesgos del Ecuador, la cual es una entidad que es responsable de administrar, anticipar y abordar las repercusiones adversas derivadas de los eventos de riesgo que suceden en Ecuador (Secretaria de Gestion de Riesgos, 2023). El dataset contiene información de los eventos peligrosos a nivel nacional, provincia, cantón y parroquia reportados en los periodos de 2010 al 2022.

### 3. IDENTIFICACIÓN DEL OBJETO DE ESTUDIO

Este proyecto reside en su enfoque innovador para abordar la creciente problemática de los accidentes en regiones específicas, una cuestión que no solo representa un desafío de salud pública, sino que también implica un significativo impacto económico. La aplicación de técnicas de aprendizaje automático para predecir y cuantificar los daños de futuros accidentes ofrece una metodología más precisa y avanzada, superando las limitaciones de los métodos tradicionales de gestión de riesgos. Este enfoque no solo beneficiará a las organizaciones en términos de planificación y asignación eficiente de recursos, sino que también contribuirá significativamente a la seguridad y bienestar de la comunidad, posicionando al proyecto como una herramienta clave en la estrategia de prevención y respuesta a accidentes.

## 4. PLANTEAMIENTO DEL PROBLEMA

La alta incidencia y severidad de eventos o desastres en ciertas regiones conlleva a significativos costos económicos y humanos, estos eventos afectan no solo a individuos, sino también a la infraestructura, economía y bienestar social, por lo cual, el uso de análisis avanzado y aprendizaje automático puede proporcionar predicciones más precisas y estrategias de mitigación efectivas, lo cual es vital para la planificación y respuesta organizacional.

## 5. OBJETIVO GENERAL

Desarrollar un modelo predictivo utilizando aprendizaje automático para cuantificar los eventos futuros por región, con el fin de mejorar las estrategias de prevención y respuesta como prepararse de mejor manera para enfrentar los eventos o asignar más recursos a las regiones donde se espera que la frecuencia de eventos sea mayor.

## 6. OBJETIVOS ESPECÍFICOS

Procesar los datos para garantizar su calidad, incluyendo la limpieza de datos, el manejo de valores faltantes, y la normalización o estandarización de variables según sea necesario.

Crear el modelo y validar las predicciones del modelo con los datos reales para determinar su precisión y fiabilidad en diferentes periodos de tiempo.

Proponer recomendaciones basadas en los resultados del modelo para las autoridades provinciales y otros stakeholders interesados en la gestión de eventos.



## 7. JUSTIFICACIÓN Y APLICACIÓN DE LA METODOLOGÍA

La Secretaría de Gestión de Riesgos de Ecuador desempeña un papel crucial en la coordinación y ejecución de la gestión integral de riesgos en el país, enfocándose en minimizar la vulnerabilidad de la población y del territorio frente a fenómenos naturales o provocados por el ser humano (Secretaría de Gestión de Riesgos, 2023). La entidad trabaja en la prevención, mitigación, respuesta y rehabilitación ante desastres, colaborando estrechamente con instituciones gubernamentales, locales, nacionales, organizaciones no gubernamentales y comunidades.

La adopción de un modelo predictivo que emplea aprendizaje automático para cuantificar eventos futuros por región representaría un avance significativo para la Secretaría. Este enfoque mejoraría la preparación y la capacidad de respuesta ante desastres, al permitir una asignación más precisa de recursos y personal basada en la probabilidad y severidad potencial de eventos en distintas áreas. Además, facilitaría la optimización de recursos, concentrándolos en las zonas más susceptibles, lo cual es vital para una gestión de riesgos eficaz en un contexto de amenazas naturales diversas.

La implementación de medidas de mitigación específicas se vería potenciada por la capacidad del modelo de identificar regiones de alto riesgo. Esto posibilitaría la realización de proyectos de infraestructura resistente, la reubicación de comunidades vulnerables y programas de educación y capacitación dirigidos a fortalecer la resiliencia local.

La comunicación de riesgos basada en predicciones confiables también juega un papel fundamental en el aumento de la conciencia y participación comunitaria, aspecto esencial para la preparación ante desastres y la reducción de

vulnerabilidades. La innovación y mejora continua que trae consigo la aplicación del aprendizaje automático en la gestión de riesgos promete una comprensión más profunda de los patrones de riesgo y mejores prácticas en mitigación y respuesta.

### 7.1. Modelo ARIMA

El método ARIMA (AutoRegressive Integrated Moving Average) es una técnica muy comúnmente utilizada para modelar y predecir datos de series temporales (Cheema, Saqib, & Iqbal, 2023). Este método se compone principalmente de tres partes fundamentales. En primer lugar, el componente autoregresivo detecta la conexión entre los valores previos y los actuales. En segundo lugar, el componente de integración se emplea para hacer que los datos adquieran una forma estacionaria. Por último, el componente de media móvil identifica la relación entre los errores previos y los valores actuales.

La implementación del modelo ARIMA para el análisis del conjunto de datos de eventos se demuestra a través de una serie de consideraciones específicas relacionadas con la naturaleza de los datos y los objetivos del análisis.

Primero, se consideró la estructura inherente del conjunto de datos. Dado que los datos consisten en registros temporales de eventos, presentan una naturaleza de serie temporal. El modelo ARIMA, siendo uno de los enfoques más establecidos para el análisis de series temporales, es particularmente adecuado para este tipo de datos. Este modelo es capaz de manejar datos temporales discretos, como los registros mensuales disponibles en el conjunto de datos, lo que lo hace idóneo para el análisis propuesto.

En segundo lugar, el modelo ARIMA es efectivo en la identificación y modelado de tendencias y patrones estacionales, que son aspectos cruciales en la predicción de eventos futuros. En caso de que el conjunto de datos muestre características de no estacionariedad, como tendencias o ciclos, el componente integrado de ARIMA permite transformar estos datos en una serie estacionaria a través de la diferenciación (Ospina, Gondim, Leiva, & Castro, 2023). Esto es

especialmente relevante si se observan tendencias o patrones estacionales en la frecuencia de los eventos registrados.

El objetivo principal del análisis de estos datos es predecir la frecuencia futura de eventos. ARIMA, con su capacidad para analizar y proyectar valores basados en patrones históricos, se alinea bien con este objetivo. Puede proporcionar pronósticos precisos sobre la frecuencia de eventos futuros, basándose en su comportamiento pasado.

Desde el punto de vista de la complejidad del modelo, ARIMA ofrece una ventaja significativa. A diferencia de los modelos de aprendizaje automático más complejos, no requiere grandes cantidades de datos y es relativamente sencillo de implementar y entender. Esto lo hace no solo práctico sino también eficiente para conjuntos de datos de tamaño moderado, como el proporcionado.

Finalmente, la interpretabilidad de los resultados es un aspecto clave en cualquier análisis de datos. Los parámetros del modelo ARIMA son intuitivos y proporcionan información directa sobre la naturaleza de la serie temporal analizada. Esto facilita la comprensión y comunicación de los resultados, lo que es especialmente valioso en contextos donde los tomadores de decisiones pueden no estar familiarizados con técnicas de modelado estadístico avanzadas.

## 7.2. Metodología

Para desarrollar un modelo predictivo que estime la cantidad de eventos futuros por provincia, primero se necesitó preparar y analizar los datos disponibles para esto se usó el lenguaje Python gestionada dentro de la plataforma Anaconda.

Con la carga inicial del dataset de eventos peligrosos se identificó que el dataset contiene 56986 filas y 56 columnas.

Posteriormente se procedió con el procesamiento de los datos, esto incluyó la gestión de valores faltantes, la codificación de variables categóricas y la normalización o estandarización de los datos, para luego analizar las tendencias, patrones y relaciones en los datos. Esto con la finalidad de entender mejor los datos para la selección de características del modelo.

Se detectó valores faltantes en los campos **COMUNIDAD/BARRIO/SECTOR**, **NUMERO DE INFORME PROVINCIAL**, **CATEGORÍA DE VÍA**, **DESCRIPCIÓN GENERAL DEL EVENTO**, **NIVEL DEL INF**, **Macroevento**, **Estado actual de la vía o del INF** y **Vía Alternativa**, estos campos contaban con un porcentaje entre 75 y 90% de valores faltantes, por lo cual se procedió a eliminarlos del dataset.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 56986 entries, 0 to 56985
Data columns (total 56 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   ID                                    56986 non-null  int64
 1   PROVINCIA                            56986 non-null  object
 2   CODIFICACION PROVINCIAL              56986 non-null  int64
 3   CANTON                               56986 non-null  object
 4   CODIFICACION CANTONAL               56986 non-null  int64
 5   PARROQUIAS                          56986 non-null  object
 6   CODIFICACION PARROQUIAL             56986 non-null  int64
 7   COMUNIDAD/BARRIO/SECTOR             56986 non-null  object
 8   LATITUD                             56986 non-null  float64
 9   LONGITUD                             56986 non-null  float64
10  EVENTO                              56986 non-null  object
11  CAUSA                               56986 non-null  object
12  CATEGORIA DEL EVENTO                56986 non-null  object
13  FECHA DEL EVENTO                    56986 non-null  datetime64[ns]
14  MES                                  56986 non-null  int64
15  NUMERO DE INFORME PROVINCIAL        44757 non-null  object
16  FALLECIDAS                          56986 non-null  int64
17  HERIDAS                              56986 non-null  int64
18  PERSONAS DESAPARECIDAS              56986 non-null  int64
19  FAMILIAS AFECTADAS                  56986 non-null  float64
20  PERSONAS AFECTADAS DIRECTAMENTE     56986 non-null  int64
21  AFECTADAS INDIRECTAS               56985 non-null  float64
22  FAMILIAS DAMNIFICADAS               56986 non-null  int64
23  PERSONAS DAMNIFICADAS               56985 non-null  float64
24  PERSONAS EVACUADAS                  56986 non-null  int64
25  PERSONAS ALBERGADAS                 56986 non-null  int64
26  P. EN FAMILIAS ACOGIENTES           56986 non-null  int64
27  PERSONAS EN OTROS MEDIOS             56986 non-null  int64
28  PERSONAS EN PROCESO DE EVACUACION    56986 non-null  int64
29  PERSONAS QUE RESISTEN A LA EVACUACION 56986 non-null  int64
30  VIVIENDAS AFECTADAS                 56985 non-null  float64
31  VIVIENDAS DESTRUIDAS                56986 non-null  int64
32  ESTABLECIMIENTOS EDUCATIVOS AFECTADOS 56986 non-null  int64
33  ESTABLECIMIENTOS EDUCATIVOS DESTRUIDOS 56986 non-null  int64
34  CENTROS DE SALUD AFECTADOS          56986 non-null  int64
35  CENTROS DE SALUD DESTRUIDOS         56986 non-null  int64
36  PUENTES AFECTADOS                   56986 non-null  int64
37  PUENTES DESTRUIDOS                  56986 non-null  int64
38  BIENES PUBLICOS AFECTADOS           56986 non-null  int64
39  BIENES PUBLICOS DESTRUIDOS          56986 non-null  int64
40  BIENES PRIVADOS AFECTADOS           56986 non-null  int64
41  BIENES PRIVADOS DESTRUIDOS          56986 non-null  int64
42  METROS LINEALES DE VÍAS AFECTADAS    56985 non-null  float64
43  CATEGORÍA DE VÍA                    52698 non-null  object
44  Ha. CULTIVO AFECTADAS                56964 non-null  float64
45  Ha. CULTIVO PERDIDAS                 56981 non-null  float64
46  Ha. DE COBERTURA VEGETAL QUEMADA    56986 non-null  float64
47  ANIMALES AFECTADOS                  56986 non-null  int64
48  ANIMALES MUERTOS                    56986 non-null  int64
49  DESCRIPCIÓN GENERAL DEL EVENTO      54144 non-null  object
50  AÑO                                  56986 non-null  int64
51  NIVEL DEL INF                        49638 non-null  object
52  Macroevento                          2122 non-null  object
53  calificación del evento peligroso    56986 non-null  object
54  Estado actual de la vía o del INF    15096 non-null  object
55  Vía Alternativa                      841 non-null  object
dtypes: datetime64[ns](1), float64(10), int64(30), object(15)
memory usage: 24.2+ MB

```

Figura 1. Visualización General Dataset.

Una vez que los datos están limpios en cuanto a valores faltantes se separó las variables cualitativas y cuantitativas para proceder a analizarlas por separado con gráficos individuales y observar cómo se encuentran distribuidas.

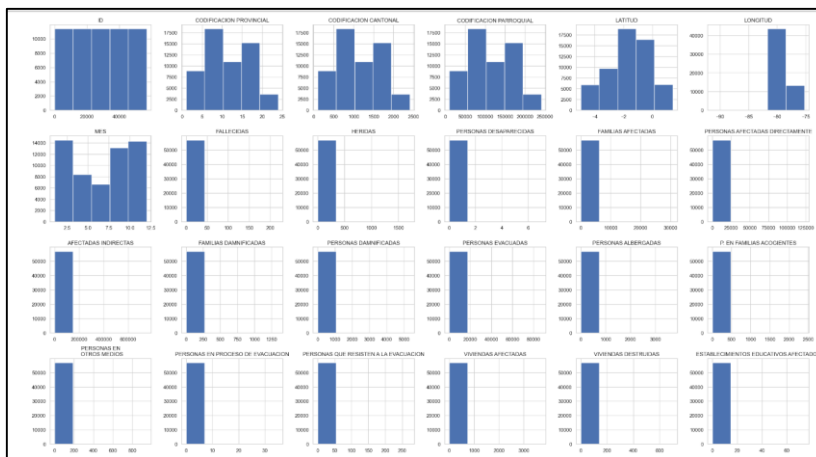


Figura 2. Distribuciones variables cuantitativas.

Con los gráficos se observó que los campos de catón y parroquia no aportarían al análisis ya que la data se dispersaría, por lo tanto, los eliminamos.

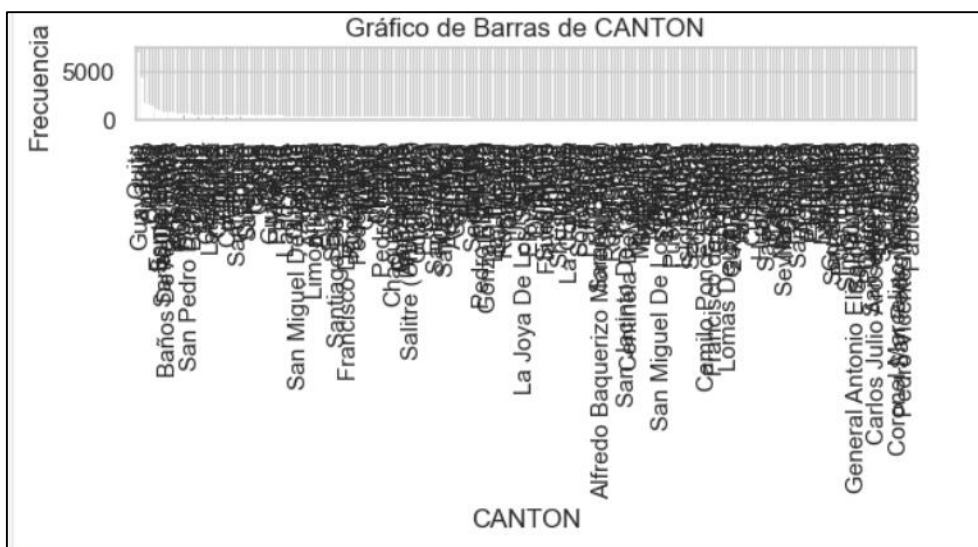


Figura 3. Distribución por cantón.

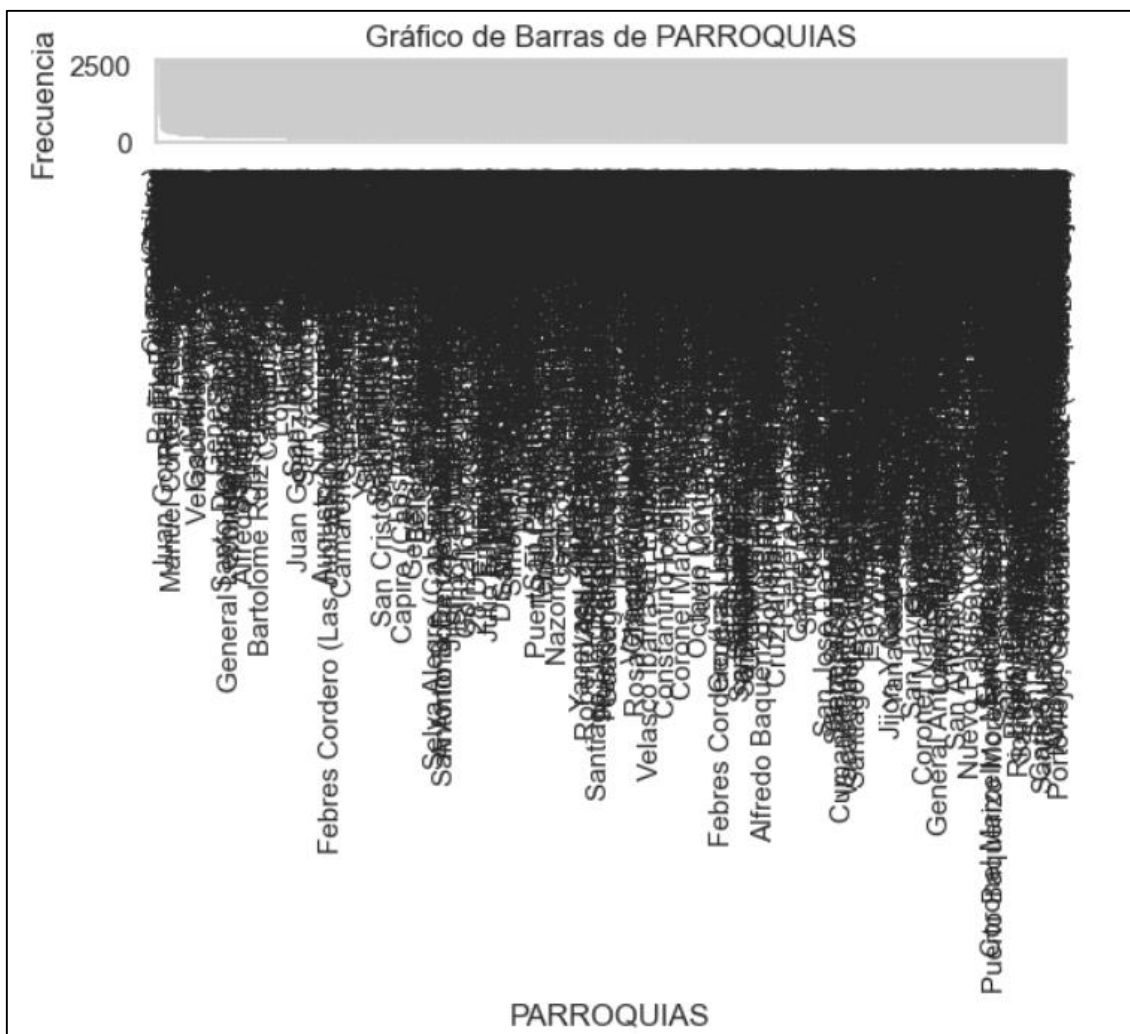


Figura 4. Distribución por parroquia.

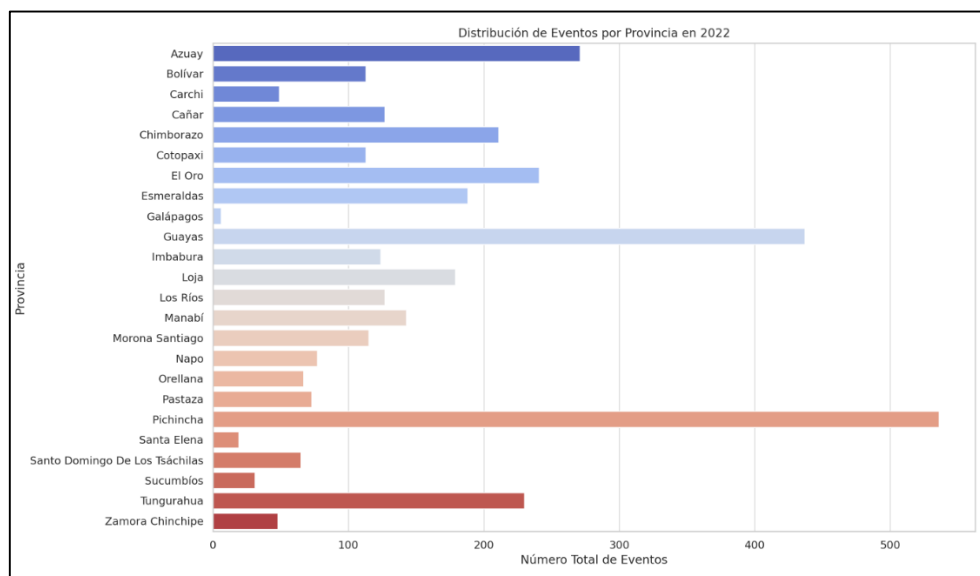
También se identificó que las columnas como: **EVENTO**, **CAUSA** y **CATEGORIA DEL EVENTO** contenían registros que no se encontraban homologados por lo tanto se realizó la debida gestión de los mismos, cambiando a todos los registros a mayúsculas y creando mapeos para su homologación, el proceso de codificación se lo puede revisar en la siguiente referencia (Crisanto, GitHub: PROYECTO CAPSTONE Eventos Peligrosos.ipynb, 2024).

Una vez que el dataset se encuentra limpio se procedió con la selección de características lo cual fue importante ya que permitió determinar cuáles son las variables relevantes para predecir el número de eventos. Esto puede incluir la fecha (año, mes), la provincia, y posiblemente derivar nuevas características que podrían ser útiles.

Debido a la naturaleza de los datos y del modelo que se empleó el cual fue el ARIMA se procedió con la selección de las variables **AÑO, MES, PROVINCIA y FECHA DEL EVENTO**, lo cual nos servirá para implementar el modelo.

Como el objetivo es cuantificar los eventos futuros por región o provincia el modelo ARIMA nos impide realizarlo de manera general usando el campo de provincia como una variable más, por lo cual se decidió filtrar el dataset por provincia y desarrollar el modelo ARIMA para las provincias con mayor cantidad de eventos.

Para ello se realizó un análisis de la distribución de eventos por región y año lo cual ayudó a entender cómo varían los eventos en diferentes provincias y periodos de tiempo, se realizó un gráfico de distribución regional para comparar el número total de eventos entre provincias en un período específico.



*Figura 5. Distribución de eventos peligrosos por provincia 2022.*

Este gráfico muestra la distribución de eventos por provincia para el año más reciente disponible en los datos. Lo cual permitió identificar rápidamente qué en las provincias de Azuay, Guayas, Pichincha y el Oro experimentaron la mayor cantidad de eventos en un período determinado, proporcionando una visión clara de las diferencias regionales en la frecuencia de eventos.

La distribución inicial de eventos por provincia y año muestra, por ejemplo, en Azuay, el número de eventos ha variado cada año desde 2010 hasta 2022, con

un aumento notorio hacia el 2022. Este tipo de análisis nos ayuda a entender las tendencias temporales y regionales de los eventos, lo cual es crucial para cualquier modelado predictivo.

Se graficó las tendencias temporales de las provincias seleccionadas, para ver cómo varía el número de eventos a lo largo del tiempo.

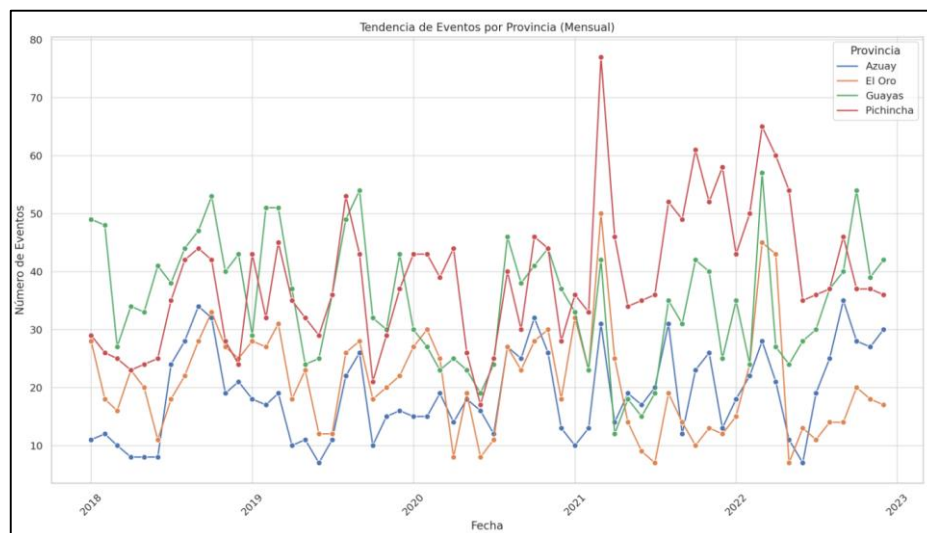


Figura 6. Tendencia mensual de eventos peligrosos por provincia.

Este gráfico muestra la tendencia mensual del número de eventos en las provincias seleccionadas (Azuay, Guayas, Pichincha, El Oro) a lo largo del tiempo, en donde podemos observar cómo varía el número de eventos mes a mes en estas regiones, lo cual es útil para identificar patrones temporales, como periodos de aumento o disminución en la frecuencia de eventos.

Estos análisis ofrecen una base sólida para entender las tendencias temporales y regionales de los eventos en el conjunto de datos. Este entendimiento es crucial para desarrollar modelos predictivos más precisos.

Una vez que se identificó las provincias con las que se va a desarrollar el modelo ARIMA, se procedió a preparar la información para que sea el input del modelo por provincia; PICHINCHA, GUAYAS, AZUAY y EL ORO, aquí se generó un campo nuevo AÑO\_MES en base a la FECHA DEL EVENTO que consta dentro del dataset y se agrupó por este campo nuevo para tener el número de eventos por cada mes.



### 7.2.1. Guayas

Cantidad de eventos por año y mes.

AÑO_MES	
2010-02	1
2010-03	4
2010-04	9
2010-05	4
2010-06	3
	..
2022-08	37
2022-09	40
2022-10	54
2022-11	39
2022-12	42

Figura 7. Cantidad de eventos peligrosos por periodo.

Una vez que tenemos la cantidad de eventos procedemos a implementar el modelo, se empezó identificando si la serie es estacionaria.

```

ADF Statistic: -2.881040698783497
p-value: 0.047597544271609865
Critical Values 1%: -3.4769274060112707
Critical Values 5%: -2.8819726324025625
Critical Values 10%: -2.577665408088415

```

Figura 8. Estadísticas de estacionariedad.

Estadístico ADF (-2.881040698783497). Este valor, obtenido del resultado de la prueba estadística, es crucial para determinar la estacionariedad de la serie de tiempo. La comparación de este estadístico con los valores críticos correspondientes a diferentes niveles de significancia permite evaluar la presencia de raíces unitarias. Si el estadístico ADF es más negativo que el valor crítico en un nivel de significancia dado, se puede rechazar la hipótesis nula de que existe una raíz unitaria en la serie, indicando que la serie es estacionaria.

Valor p (0.047597544271609865). El valor p indica la probabilidad de observar los datos actuales si la hipótesis nula fuera cierta. Un valor generalmente igual o menor a 0.05, sugiere que es posible rechazar la hipótesis nula, concluyendo así que la serie es estacionaria. En este caso, el valor p es aproximadamente 0.048, lo que sugiere una evidencia débil contra la hipótesis nula y apoya la conclusión de estacionariedad en el nivel de significancia del 5%.

Valores Críticos:

1%: -3.4769274060112707. Para rechazar la hipótesis nula con un nivel de significancia del 1%, el estadístico ADF debería ser menor que -3.4769. Dado que esto no ocurre en este caso, la hipótesis nula no puede ser rechazada en este nivel de confianza.

5%: -2.8819726324025625: Para rechazar la hipótesis nula con un nivel de significancia del 5%, el estadístico ADF debería ser menor que -2.8819. Dado que el estadístico ADF es marginalmente menor, esto proporciona una evidencia débil contra la hipótesis nula a este nivel de significancia.

10%: -2.577665408088415: Para rechazar la hipótesis nula con un nivel de significancia del 10%, el estadístico ADF debería ser menor que -2.5777. En este escenario, el estadístico ADF es menor, lo que permite rechazar la hipótesis nula con un nivel de confianza del 90%, sugiriendo que la serie de tiempo es estacionaria.

Basándose en que el estadístico ADF es ligeramente menor que el valor crítico en el nivel de significancia del 5% y el valor  $p$  es cercano a 0.05, se sugiere que hay una evidencia débil para rechazar la hipótesis nula de no estacionariedad, apoyando la conclusión de que la serie de tiempo es estacionaria. No obstante, debido a la cercanía del valor  $p$  al umbral de significancia y al leve cruce del estadístico ADF sobre el valor crítico, esta conclusión debe ser considerada con cautela.

Luego procedemos a identificar los parámetros óptimos para el modelo ARIMA, necesitamos observar las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) de la serie temporal. Estas funciones nos ayudan a determinar los valores iniciales para  $p$  y  $q$ , que corresponden a los componentes autoregresivos y de media móvil del modelo, respectivamente.

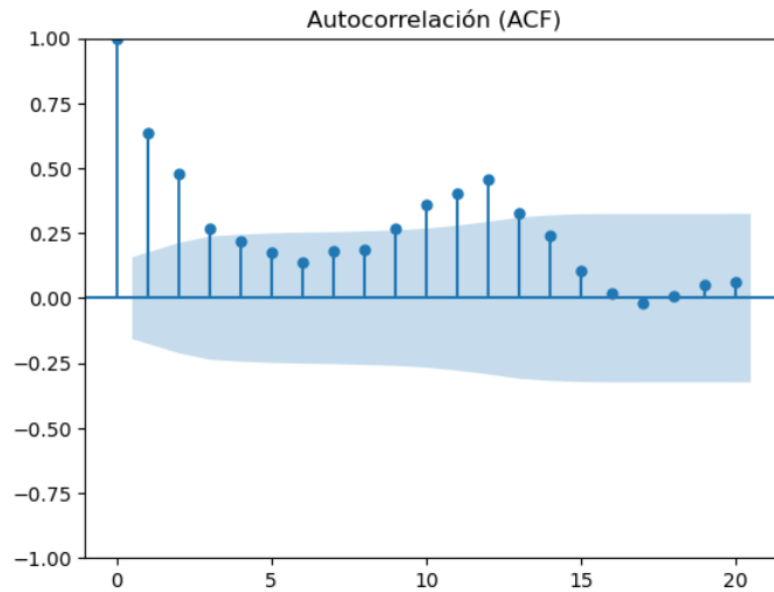


Figura 9. Autocorrelación ACF.

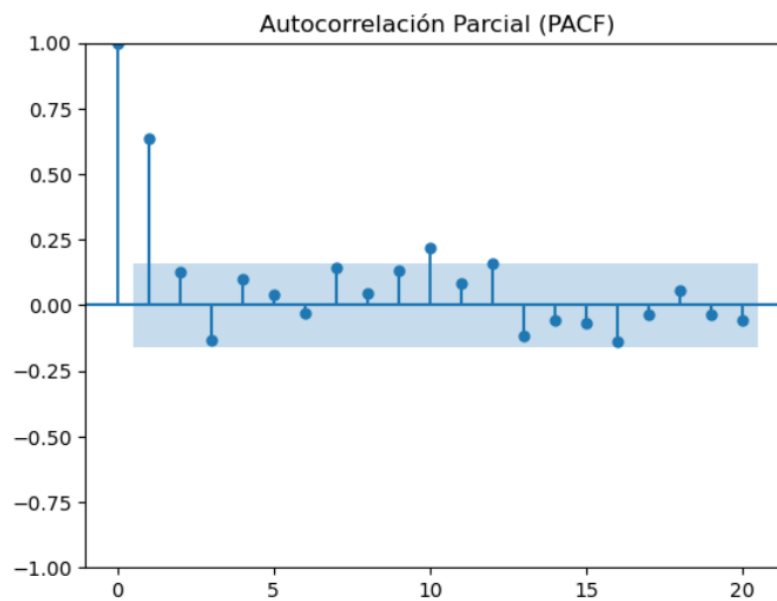


Figura 10. Autocorrelación parcial PACF.

ACF muestra una disminución gradual después del primer rezago, lo que sugiere que podríamos tener un componente de media móvil (MA) en nuestro modelo.

PACF muestra un corte brusco después del primer rezago, lo que indica la presencia de un componente autoregresivo (AR).

Basándonos en estas observaciones, un punto de partida razonable para nuestro modelo ARIMA fue  $p=1$ ,  $d=0$  (ya que la serie es estacionaria y no necesita



significativo y se procedió a evaluar de nuevo el modelo, obteniendo los siguientes resultados.

SARIMAX Results						
Dep. Variable:	y	No. Observations:	155			
Model:	ARIMA(1, 0, 0)	Log Likelihood	-587.861			
Date:	Sat, 24 Feb 2024	AIC	1181.722			
Time:	01:26:43	BIC	1190.852			
Sample:	02-01-2010	HQIC	1185.430			
	- 12-01-2022					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
const	28.4515	2.637	10.790	0.000	23.284	33.620
ar.L1	0.6515	0.056	11.559	0.000	0.541	0.762
sigma2	114.8767	11.048	10.398	0.000	93.222	136.531
Ljung-Box (L1) (Q):	1.54	Jarque-Bera (JB):	16.17			
Prob(Q):	0.22	Prob(JB):	0.00			
Heteroskedasticity (H):	0.97	Skew:	0.43			
Prob(H) (two-sided):	0.92	Kurtosis:	4.33			

Figura 12. Resumen evaluación modelo 2.

Se realizó una tabla comparativa con los valores resultantes de la evaluación del modelo con diferentes parámetros  $p$ ,  $d$ ,  $q$ .

	ARIMA(1,0,1)	ARIMA(1,0,0)
Parámetros	Modelo_1	Modelo_2
AIC	1181.55	1181.72
BIC	1193.72	1190.85

Figura 13. Comparación parámetros AIC y BIC.

Donde se puede observar que el valor de AIC del modelo\_2 es mínimamente superior al valor del modelo\_1 con una diferencia de 0.17, mientras que el valor de BIC del modelo\_2 es ligeramente inferior al valor del modelo\_1 con una diferencia de 2.87, por lo cual tomamos como modelo final el modelo\_2.

Una vez identificado el modelo con que vamos a trabajar se procedió a evaluar el modelo, es importante realizar un diagnóstico de los residuos del modelo para asegurarse de que no contienen patrones no capturados por el modelo. Esto generalmente implica verificar que los residuos se comporten como ruido blanco,

es decir, que sean independientes y estén distribuidos normalmente con media cero.

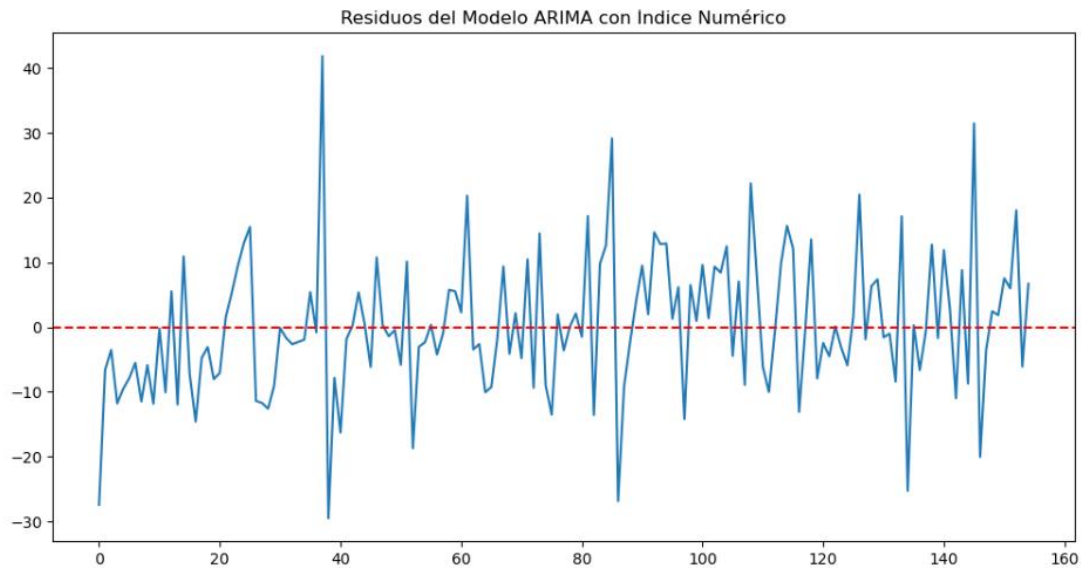


Figura 14. Gráfico de residuos del modelo ARIMA.

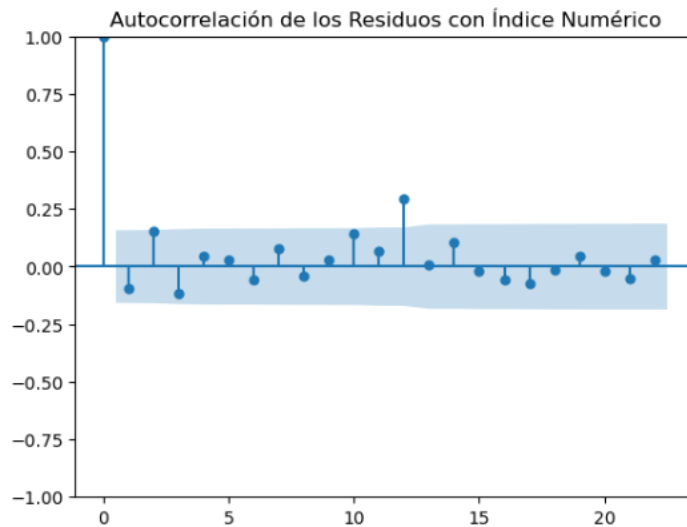


Figura 15. Autocorrelación de residuos del modelo ARIMA.

El gráfico de residuos muestra las desviaciones de los valores predichos respecto a los observados a lo largo del tiempo. La línea roja representa el valor 0, indicando donde los residuos serían ideales si el modelo capturara perfectamente la dinámica de la serie temporal. Observamos cierta variabilidad en los residuos, lo cual es esperado, pero idealmente estos no deberían mostrar patrones claros o tendencias a lo largo del tiempo.

La función de autocorrelación (ACF) de los residuos ayuda a identificar cualquier autocorrelación remanente en los residuos que no fue capturada por el modelo. Idealmente, pretendemos que todas las barras se mantengan dentro del área sombreada, indicando que no hay autocorrelaciones significativas a diferentes retrasos. En este gráfico, se observa que la mayoría de las autocorrelaciones de los residuos están dentro del intervalo de confianza, sugiriendo que el modelo ha capturado adecuadamente la estructura de dependencia en los datos.

A continuación, generó predicciones dentro de la muestra y las comparamos visualmente con los datos reales para evaluar el ajuste del modelo.

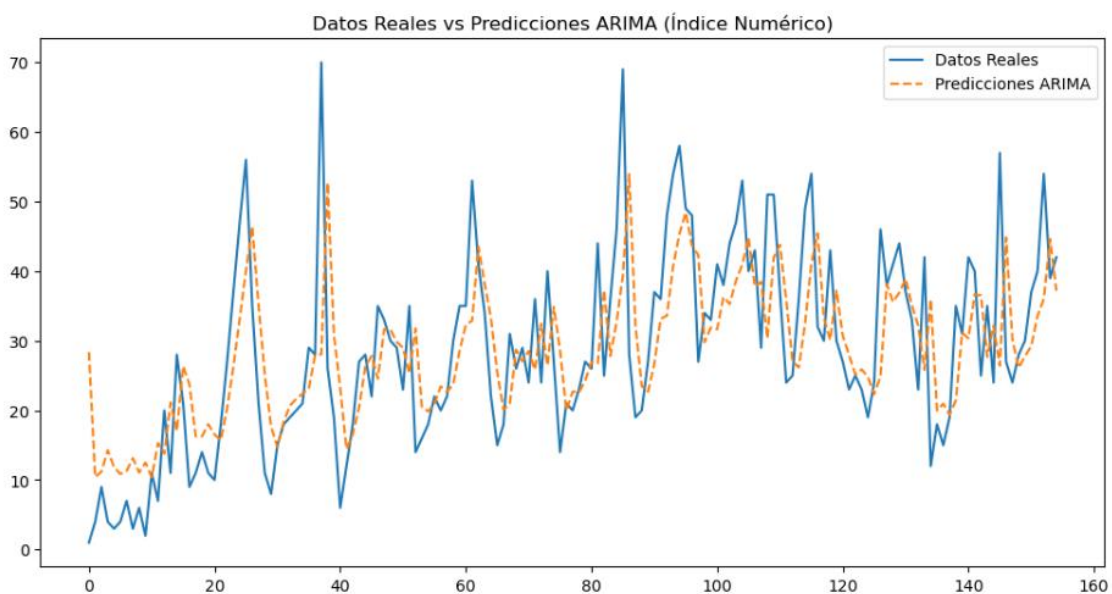


Figura 16. Datos reales vs predicciones de Guayas.

Como se puede observar en el gráfico, los Datos Reales se muestran como una línea continua azul, mientras que las Predicciones ARIMA están representadas por una línea discontinua naranja. Este tipo de visualización nos permite evaluar de manera intuitiva qué tan bien el modelo ARIMA captura la tendencia y la variabilidad de la serie temporal de eventos peligrosos, con lo cual podemos considerar cómo las predicciones se alinean con los datos reales en términos de tendencia general y variaciones, además de la existencia de patrones o períodos específicos donde el modelo parece desviarse significativamente de los datos reales.

Posteriormente realizamos las predicciones por mes junto con los límites superior e inferior por mes, no se debe olvidar que estas predicciones son únicamente de la provincia del Guayas.

Fecha	Predicción	Límite Inferior	Límite Superior
2023-01-31	37.769400	16.911256	58.627545
2023-02-28	35.381693	11.205560	59.557827
2023-03-31	33.601323	7.764108	59.438538
2023-04-30	32.273807	5.557700	58.989914
2023-05-31	31.283957	4.091489	58.476425
2023-06-30	30.545885	3.092145	57.999626
2023-07-31	29.995549	2.397616	57.593482
2023-08-31	29.585196	1.907421	57.262972
2023-09-30	29.279221	1.557154	57.001288
2023-10-31	29.051073	1.304411	56.797735
2023-11-30	28.880957	1.120630	56.641283
2023-12-31	28.754111	0.986190	56.522032

Figura 17. Predicciones por periodo de Guayas.

El proceso detallado del desarrollo del modelo se lo puede observar en la siguiente referencia (Crisanto, GitHub: PC Eventos Peligrosos Guayas.ipynb, 2024).

El proceso de ejecución del modelo para el resto de provincias seleccionadas (Pichincha, Azuay y El Oro) es muy similar, para el caso de estudio únicamente se presentará los resultados finales del desarrollo del modelo.

### 7.2.2. Pichincha

Para el caso de pichincha la serie no fue estacionaria de un inicio por lo que se tuvo que realizar transformaciones para hacerla estacionaria y continuar con el proceso, para más información sobre el modelo realizado de la provincia de Pichincha puede dirigirse a la siguiente referencia (Crisanto, GitHub: PC Eventos Peligrosos Pichincha.ipynb, 2024).



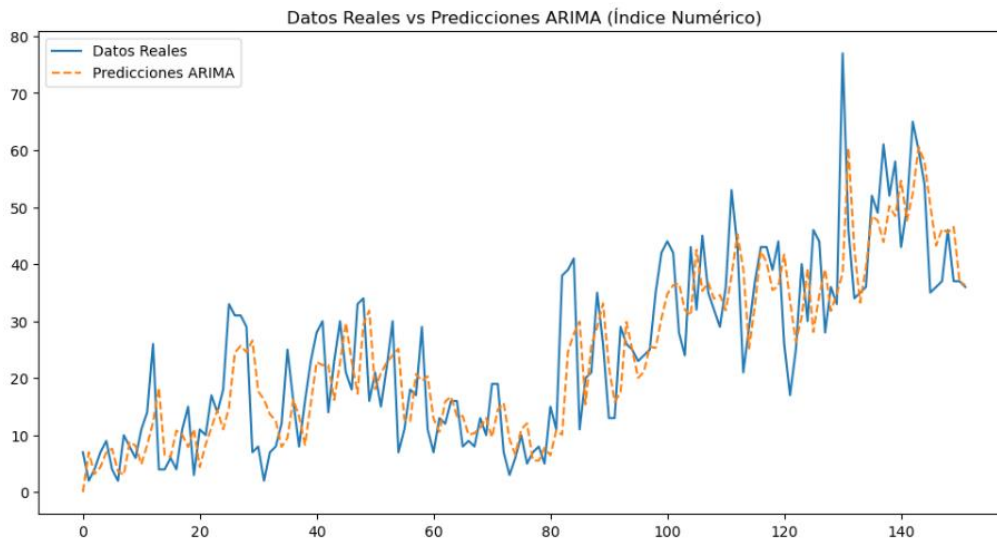


Figura 18. Datos reales vs predicciones de Pichincha.

Fecha	Predicción	Límite Inferior	Límite Superior
2023-01-31	33.478311	14.430903	52.525720
2023-02-28	31.503935	7.312802	55.695067
2023-03-31	29.958081	3.096369	56.819794
2023-04-30	28.747743	0.372902	57.122584
2023-05-31	27.800099	-1.463669	57.063867
2023-06-30	27.058133	-2.737458	56.853724
2023-07-31	26.477206	-3.639763	56.594174
2023-08-31	26.022364	-4.289932	56.334659
2023-09-30	25.666241	-4.765175	56.097657
2023-10-31	25.387413	-5.116797	55.891622
2023-11-30	25.169102	-5.379646	55.717849
2023-12-31	24.998173	-5.577846	55.574192

Figura 19. Predicciones por periodo de Pichincha.

### 7.2.3. Azuay

Para el caso de Azuay la serie fue estacionaria de un inicio por lo que se continuo con el proceso, para más información sobre el modelo realizado de la provincia de Azuay puede dirigirse a la siguiente referencia (Crisanto, GitHub: PC Eventos Peligrosos Azuay.ipynb, 2024).

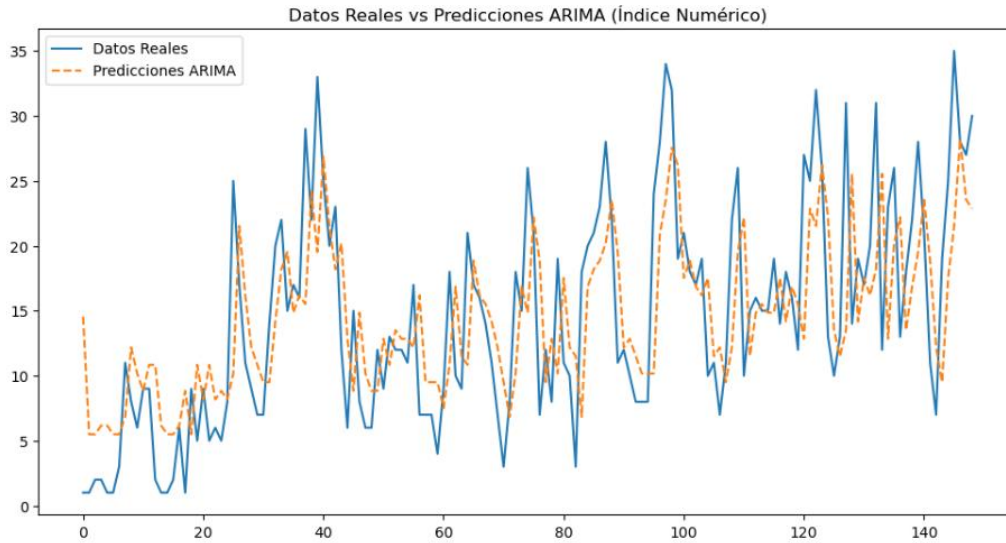


Figura 20. Datos reales vs predicciones de Azuay.

Fecha	Predicción	Límite Inferior	Límite Superior
2023-01-31	26.316524	14.214313	38.418736
2023-02-28	24.432277	10.413114	38.451441
2023-03-31	23.468409	8.743107	38.193711
2023-04-30	22.975351	7.908454	38.042249
2023-05-31	22.723133	7.447383	37.998882
2023-06-30	22.594112	7.164550	38.023675
2023-07-31	22.528113	6.969831	38.086395
2023-08-31	22.494352	6.819619	38.169086
2023-09-30	22.477082	6.692253	38.261911
2023-10-31	22.468248	6.576828	38.359667
2023-11-30	22.463728	6.467809	38.459648
2023-12-31	22.461417	6.362373	38.560461

Figura 21. Predicciones por periodo de Azuay.

#### 7.2.4. El Oro

Para el caso de El Oro la serie no fue estacionaria de un inicio por lo que se tuvo que realizar transformaciones para hacerla estacionaria y continuar con el proceso del modelo, para más información sobre el modelo realizado de la provincia de El Oro puede dirigirse a la siguiente referencia (Crisanto, GitHub: PC Eventos Peligrosos El Oro.ipynb, 2024).

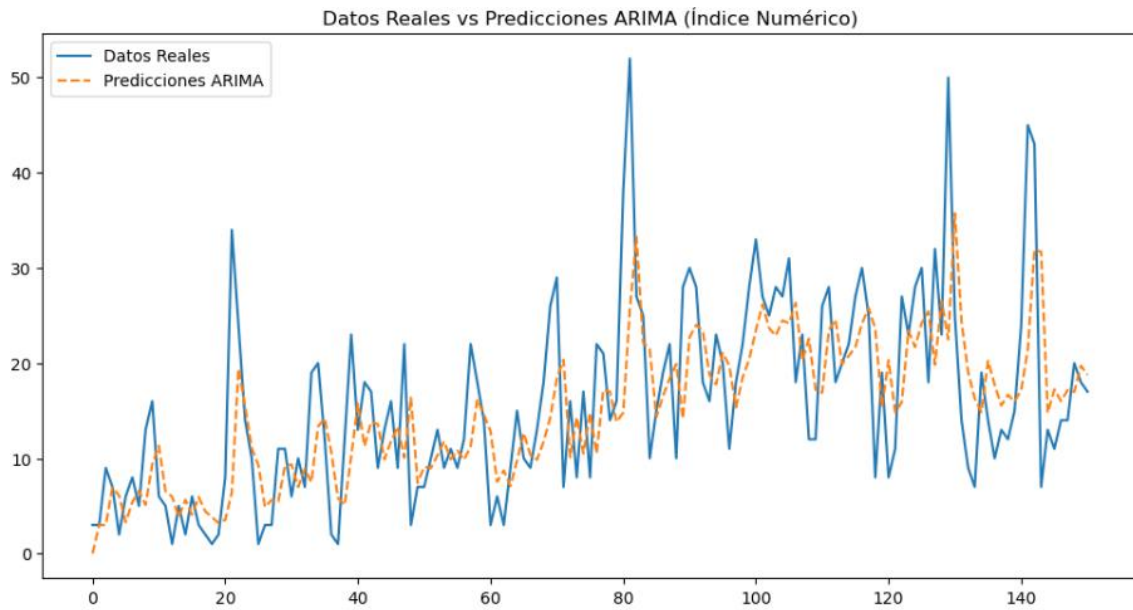


Figura 22. Datos reales vs predicciones de El Oro.

Fecha	Predicción	Límite Inferior	Límite Superior
2023-01-31	18.231524	2.705413	33.757634
2023-02-28	18.759042	1.474412	36.043671
2023-03-31	18.985002	1.197626	36.772378
2023-04-30	19.081791	1.083997	37.079585
2023-05-31	19.123250	1.000932	37.245568
2023-06-30	19.141009	0.924118	37.357900
2023-07-31	19.148616	0.848865	37.448367
2023-08-31	19.151874	0.774207	37.529542
2023-09-30	19.153270	0.699925	37.606615
2023-10-31	19.153868	0.625962	37.681773
2023-11-30	19.154124	0.552300	37.755948
2023-12-31	19.154234	0.478930	37.829538

Figura 23. Predicciones por periodo de El Oro.

## 8. RESULTADOS

A continuación, se presentan los resultados del modelo desarrollado para cada una de las provincias seleccionadas Guayas, Pichincha, Azuay y El Oro, en los cuales se visualizará tanto las predicciones los limites inferior y superior

En el siguiente gráfico se puede visualizar los valores de las predicciones por mes para el 2023 de las 4 provincias.

Provincia Fecha	AZUAY			EL ORO			GUAYAS			PICHINCHA		
	Predicción	Límite Inferior	Límite Superior	Predicción	Límite Inferior	Límite Superior	Predicción	Límite Inferior	Límite Superior	Predicción	Límite Inferior	Límite Superior
2023-01	26	14	38	18	3	34	38	17	59	33	14	53
2023-02	24	10	38	19	1	36	35	11	60	32	7	56
2023-03	23	9	38	19	1	37	34	8	59	30	3	57
2023-04	23	8	38	19	1	37	32	6	59	29	0	57
2023-05	23	7	38	19	1	37	31	4	58	28	-1	57
2023-06	23	7	38	19	1	37	31	3	58	27	-3	57
2023-07	23	7	38	19	1	37	30	2	58	26	-4	57
2023-08	22	7	38	19	1	38	30	2	57	26	-4	56
2023-09	22	7	38	19	1	38	29	2	57	26	-5	56
2023-10	22	7	38	19	1	38	29	1	57	25	-5	56
2023-11	22	6	38	19	1	38	29	1	57	25	-5	56
2023-12	22	6	39	19	0	38	29	1	57	25	-6	56
Total	275	95	457	227	13	445	377	58	696	332	-9	674

Figura 24. Vista general de predicciones por periodo y provincia.

Los límites superior e inferior en el contexto de las predicciones de un modelo ARIMA, o en general en modelos estadísticos y de machine learning, forman lo que se conoce como intervalos de confianza para las predicciones. Estos intervalos proporcionan un rango estimado dentro del cual se espera que caigan los verdaderos valores futuros con un cierto nivel de confianza, típicamente el 95%.

Cuando se dice que un intervalo de confianza tiene un nivel de confianza del 95%, significa que si repetimos el proceso de muestreo y cálculo del intervalo muchas veces (bajo las mismas condiciones), esperaríamos que el intervalo incluya el verdadero valor futuro el 95% de esas veces, por ende, estos intervalos de confianza son una medida de la incertidumbre asociada a una predicción. Cuanto más amplio es el intervalo, mayor es la incertidumbre. Factores como la variabilidad de los datos y el tamaño de la muestra pueden afectar a esta incertidumbre.

Cuando se aplique o se analice las predicciones estos intervalos ayudaran a los tomadores de decisiones a entender el rango posible de los resultados futuros y a planificar teniendo en cuenta tanto el mejor como el peor escenario.

Los valores de las predicciones que se obtuvieron, incluyendo las predicciones puntuales y los límites inferior y superior de los intervalos de confianza, son típicos de lo que se podría esperar de un modelo ARIMA con las condiciones mencionadas.

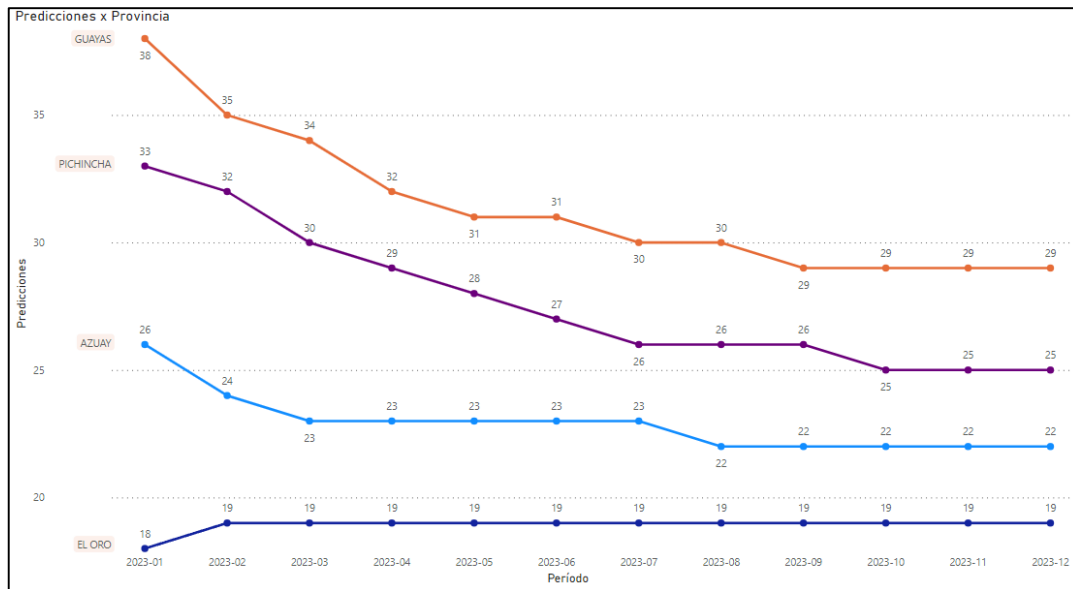


Figura 25. Comparación de predicciones por provincia.

Las predicciones puntuales que convergen hacia un valor constante como por ejemplo en las provincias de El Oro y Azuay sugieren que el modelo prevé que la serie temporal alcanzará una especie de equilibrio o nivel medio a largo plazo. Esto es común en series temporales que no tienen una tendencia clara o estacionalidad que el modelo pueda seguir hacia el futuro.

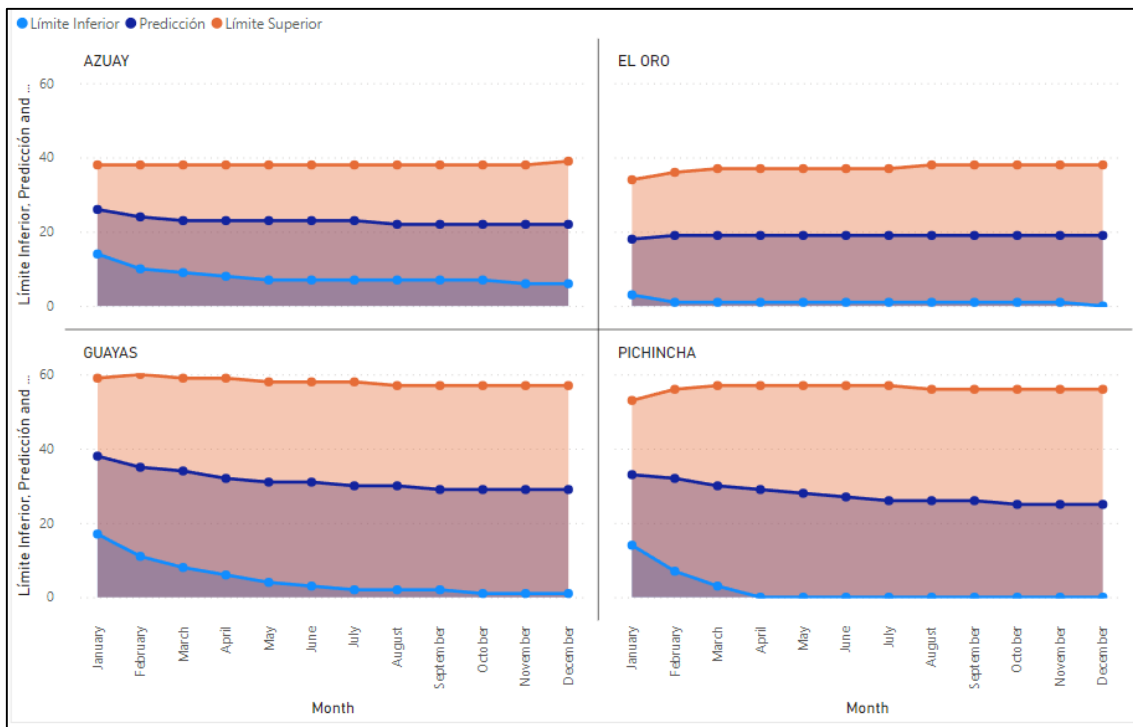


Figura 26. Comportamiento de los límites y predicciones.

De la misma manera como se puede observar los intervalos de confianza se amplían a medida que nos movemos hacia el futuro, esto refleja la incertidumbre inherente a las predicciones a largo plazo. Esto es esperado, ya que cuanto más intentamos predecir hacia el futuro, mayor es la incertidumbre sobre estos valores predichos. Sin embargo, el grado en que se amplían puede ser una señal de la confianza del modelo en sus propias predicciones.

## 9. DISCUSIÓN DE LOS RESULTADOS Y PROPUESTA DE SOLUCIÓN

Para poder interpretar y discutir los resultados es necesario realizar varios análisis como, por ejemplo:

Análisis Temporal, el cual examinará las predicciones mensuales a nivel agregado para detectar patrones o picos estacionales.

Análisis por Provincia, lo que nos permitirá comparar las predicciones medias de eventos peligrosos entre provincias, identificando las provincias con las predicciones más altas y las más bajas.

Y por último la evaluación de la Incertidumbre, esto analizará la amplitud de los intervalos de confianza para identificar predicciones con mayor incertidumbre.

A partir de los análisis de las predicciones de eventos peligrosos por provincia y mes, se han identificado varias tendencias y patrones clave. Estos hallazgos pueden informar recomendaciones específicas para autoridades provinciales y otros interesados en la gestión de eventos peligrosos. A continuación, se presentan los resultados del análisis:

### 9.1. Análisis de Tendencias Temporales

El análisis temporal muestra una disminución gradual en la media de las predicciones de eventos peligrosos a lo largo del año 2023. En enero, la predicción promedio de eventos peligrosos es la más alta, con un valor de aproximadamente 29 eventos, y disminuye consistentemente hasta diciembre, donde el promedio es de 24 eventos. Esta tendencia sugiere una variabilidad estacional en la probabilidad de eventos peligrosos, lo cual puede ser crucial para la planificación de recursos y estrategias de prevención a lo largo del año.

### 9.2. Comparación por Provincia

Cuando se examinan las predicciones por provincia, Guayas muestra el promedio más alto de eventos peligrosos predichos (aproximadamente 31),

seguido por Pichincha (aproximadamente 28), Azuay (aproximadamente 23), y finalmente El Oro (aproximadamente 19). Esta información es vital para dirigir los esfuerzos y recursos de manera efectiva, priorizando aquellas provincias con mayores niveles de riesgo predicho.

### 9.3. Evaluación de la Incertidumbre

La amplitud del intervalo de confianza proporciona una medida de la incertidumbre asociada con las predicciones. Pichincha presenta la mayor amplitud promedio de intervalos de confianza (aproximadamente 57), seguida por Guayas (aproximadamente 53), lo que indica una mayor incertidumbre en estas predicciones. En contraste, Azuay y El Oro muestran amplitudes más pequeñas (aproximadamente 30 y 36, respectivamente), lo que sugiere predicciones con menor incertidumbre.

En base a los análisis anteriormente presentados se plantean las siguientes propuestas de solución:

**Planificación de Recursos y Estrategias de Prevención.** Las autoridades deberían considerar la variabilidad estacional y aumentar la preparación durante los meses de mayor riesgo, especialmente al inicio del año.

**Priorización de Esfuerzos por Provincia.** Las provincias con mayores niveles de riesgo, como Guayas y Pichincha, deberían recibir atención prioritaria en términos de asignación de recursos y desarrollo de infraestructura de gestión de emergencias.

Estas propuestas buscan apoyar a las autoridades provinciales y otros stakeholders en la gestión eficaz de eventos peligrosos, basándose en un entendimiento profundo de las predicciones y su incertidumbre asociada.

### 9.4. IMPLICACIONES SOBRE INNOVACION EMPRESARIAL

Las estrategias propuestas tienden a ser innovadoras en su enfoque holístico y adaptativo hacia la gestión de eventos peligrosos. Al integrar análisis predictivo, personalización según eventos provinciales y tecnología de vanguardia, estas



estrategias representan un cambio significativo hacia una gestión de riesgos más proactiva, eficiente y basada en la evidencia. Contar con un enfoque innovador no solo mejorará la capacidad de respuesta ante eventos peligrosos, sino que también contribuirá a la resiliencia a largo plazo de las comunidades frente a futuras amenazas.

El análisis de tendencias temporales y la comparación por provincia juegan un papel crucial en la influencia de las estrategias de gestión de eventos peligrosos propuestas, permitiendo una asignación de recursos más informada y estratégica, así como una planificación anticipada para períodos de mayor riesgo. A continuación, se detalla cómo este análisis impactaría en la formulación de estrategias efectivas:

La identificación de una tendencia decreciente en el promedio de eventos peligrosos a lo largo del año sugiere una variabilidad estacional importante. Este conocimiento capacitará a las autoridades para:

Prepararse para períodos de alto riesgo, conociendo que el inicio del año presenta un pico en la predicción de eventos peligrosos, se pueden anticipar y asignar recursos adicionales, como personal de respuesta a emergencias, financiamiento y equipamiento, asegurando una preparación adecuada durante estos meses críticos.

Optimizar el uso de recursos en los meses con menores predicciones de eventos peligrosos, es posible reajustar o reducir los recursos asignados, permitiendo una gestión más eficiente del presupuesto y las operaciones.

La capacidad para identificar los meses con mayor y menor riesgo de eventos peligrosos posibilita el desarrollo y la implementación de estrategias de mitigación y preparación temporales, tales como:

Incrementar la frecuencia de simulacros y capacitaciones previo a los períodos de mayor riesgo, así también la implementación de campañas de comunicación dirigidas a educar a la población sobre precauciones y planes de evacuación antes de la llegada de los meses más riesgosos (Wannous & Velasquez, 2017).

El conocimiento anticipado de las tendencias permitirá a las autoridades y organizaciones transicionar de una respuesta reactiva a una proactiva frente a los eventos peligrosos, lo que conllevaría a mejorar la preparación general, las estrategias podrían no solo enfocarse en la respuesta sino también en la mejora de la infraestructura crítica y la capacidad de respuesta antes de la ocurrencia de los eventos, disminuyendo así su impacto potencial (Lundgren & McMakin, 2018). Además de invertir en tecnología de monitoreo y alerta temprana, esto es crucial para detectar y comunicar rápidamente los riesgos en los meses de mayor incidencia, mejorando la eficacia de las medidas de mitigación.

El análisis de incertidumbre, especialmente a través de la observación de los intervalos de confianza en las predicciones de eventos peligrosos, juega un papel fundamental en la influencia sobre las estrategias de gestión propuestas para estos eventos. Este enfoque no solo profundiza la comprensión sobre los niveles de riesgo, sino que también facilita una planificación más detallada y una respuesta adaptable, como, por ejemplo:

La priorización basada en el riesgo, en la cual la variabilidad en la incertidumbre de las predicciones entre diferentes regiones o periodos permite a las autoridades identificar aquellas áreas con mayor incertidumbre, señaladas por intervalos de confianza amplios. Esto lleva a una priorización en la inversión para la recopilación de datos adicionales, análisis más profundos, y medidas de mitigación enfocadas principalmente en estas regiones. La asignación eficiente de recursos hacia áreas donde las predicciones son menos precisas mejora la capacidad para evaluar y manejar el riesgo potencial (Hyndman & Athanasopoulos, 2018).

El desarrollo de planes de respuesta flexibles en donde la presencia de incertidumbre en las predicciones subraya la importancia de crear planes de respuesta que sean flexibles y capaces de adaptarse a distintos escenarios de eventos peligrosos. Estrategias que incluyen planes contingentes, que se activarían al superarse ciertos umbrales dentro de los intervalos de confianza, permitirían una actuación más dinámica y efectiva ante la materialización de estos eventos.

## 10. CONCLUSIONES Y RECOMENDACIONES

Se implementó un riguroso procedimiento para el procesamiento de los datos, asegurando su calidad a través de la limpieza de datos, el manejo efectivo de valores faltantes y la normalización de variables. Este proceso ha sido fundamental para establecer una base sólida que garantice la fiabilidad de los análisis posteriores. El esfuerzo en esta etapa ha permitido minimizar los errores y maximizar la utilidad de los datos para el modelado predictivo.

El modelo desarrollado ha sido validado contra datos reales, demostrando una precisión y fiabilidad aceptable en diferentes provincias. Esta validación nos ha permitido confirmar la robustez del modelo y su capacidad para cuantificar con eficacia los eventos peligrosos.

Basándonos en los insights obtenidos del modelo, hemos formulado una serie de estrategias dirigidas a las autoridades provinciales y otros stakeholders interesados en la optimización de la gestión de eventos. Estas estrategias están diseñadas para aprovechar las predicciones del modelo, mejorar la planificación, la asignación de recursos y la respuesta a las dinámicas cambiantes de los eventos, donde se asegura que su implementación conducirá a una gestión más eficaz y eficiente de los eventos peligrosos.

Dada la variabilidad en la incertidumbre de las predicciones, se recomienda una revisión continua de los modelos predictivos, especialmente para aquellas provincias con alta incertidumbre en sus predicciones, como Pichincha y Guayas. Esto puede incluir la recopilación de más datos o la revisión otros métodos de modelado.

Explorar y, cuando sea posible, integrar nuevas fuentes de datos que puedan enriquecer el modelo y mejorar la calidad de las predicciones. Esto puede incluir datos sociales, económicos, climáticos, entre otros.

Asegurar que todos los procesos de recolección, procesamiento y análisis de datos cumplan con las regulaciones vigentes sobre protección de datos y privacidad.

Fomentar una cultura organizacional que priorice la toma de decisiones basada en datos, utilizando las predicciones y recomendaciones del modelo para guiar la planificación y gestión de eventos.

## 11. REFERENCIAS

- Amin, M., & Ahn, H. (2021). Earthquake disaster avoidance learning system using deep learning. *Cognitive Systems Research*, 66, 221-235. doi:10.1016/j.cogsys.2020.11.002
- Ataollah, S., Dieu, T. B., Binh, T. P., Karim, S., & Kamran, C. (2017). Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environ Earth Sci*, 76, 1-18. doi:10.1007/s12665-016-6374-y
- Cheema, M., Saqib, M., & Iqbal, S. (2023). Rainfall Prediction using Big Data Analytics: A Systematic Literature Review. *International Journal of Computations, Information and Manufacturing (IJCIM)*, 3(2), 28-36. doi:10.54489/ijcim
- Crisanto, S. (25 de Febrero de 2024). *GitHub: PC Eventos Peligrosos Azuay.ipynb*. Obtenido de GitHub: [https://github.com/wcris/Proyecto\\_Capstone.git](https://github.com/wcris/Proyecto_Capstone.git)
- Crisanto, S. (25 de Febrero de 2024). *GitHub: PC Eventos Peligrosos El Oro.ipynb*. Obtenido de GitHub: [https://github.com/wcris/Proyecto\\_Capstone.git](https://github.com/wcris/Proyecto_Capstone.git)
- Crisanto, S. (25 de Febrero de 2024). *GitHub: PC Eventos Peligrosos Guayas.ipynb*. Obtenido de GitHub: [https://github.com/wcris/Proyecto\\_Capstone.git](https://github.com/wcris/Proyecto_Capstone.git)
- Crisanto, S. (25 de Febrero de 2024). *GitHub: PC Eventos Peligrosos Pichincha.ipynb*. Obtenido de GitHub: [https://github.com/wcris/Proyecto\\_Capstone.git](https://github.com/wcris/Proyecto_Capstone.git)
- Crisanto, S. (25 de Febrero de 2024). *GitHub: PROYECTO CAPSTONE Eventos Peligrosos.ipynb*. Obtenido de GitHub: [https://github.com/wcris/Proyecto\\_Capstone.git](https://github.com/wcris/Proyecto_Capstone.git)
- Deparday, V., Gevaert, C., Molinario, G., Soden, R., & Balog-Way, S. (2019). Machine Learning for Disaster Risk Management. *World Bank*, 51.
- Fan, C., Wu, F., & Mostafani, A. (2020). A Hybrid Machine Learning Pipeline for Automated Mapping of Events and Locations From Social Media in Disasters. *IEEE Access*, 8, 10478-10490. doi:10.1109/ACCESS.2020.2965550
- Huang, X., Song, J., & Jin, H. (2020). The casualty prediction of earthquake disaster based on Extreme Learning Machine method. *Natural Hazards*, 102, 873-866. doi:10.1007/s11069-020-03937-6
- Huang, Y., Jin, L., Sheng Zhao, H., & Yan Huang, X. (2018). Fuzzy neural network and LLE Algorithm for forecasting precipitation in tropical cyclones: comparisons with interpolation method by ECMWF and stepwise regression method. *Natural Hazards*, 91, 201-220. doi:10.1007/s11069-017-3122-x
- Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting Principles and Practice*. OTexts.
- Iparraquirre, O., Melgarejo, M., Castro, G., & Olaya, S. (2023). Classification of Tweets Related to Natural Disasters Using Machine Learning Algorithms. *International Journal of Interactive Mobile Technologies (IJIM)*. doi:10.3991/ijim.v17i14.39907

- Kameswari, L. (2023). A Review on Application of Machine Learning Techniques in Climate Change and Disaster Management. *Technoarete Transactions on Climate Change and Disaster Management Research*, 2(1).
- Kuglitsch, M., Pelivan, I., Ceola, S., Menon, M., & Xoplaki, E. (2022). Facilitating adoption of AI in natural disaster management through collaboration. *Nature Communications*, 13, 1579. doi:<https://doi.org/10.1038/s41467-022-29285-6>
- Kurlander, J., Waljee, A., Menees, S., Lipson, Rachel, & Kokaly, A. (2022). Regression and Random Forest Machine Learning Have Limited Performance in Predicting Bowel Preparation in Veteran Population. *Digestive Diseases and Sciences*, 67(7), 2827-2841. doi:10.1007/s10620-021-07113-z
- Linardos, V., Drakaki, M., Tzionas, P., & Karnavas, Y. (2022). Machine Learning in Disaster Management: Recent Developments in Methods and Applications. *Machine Learning & Knowledge Extraction*, 4(2), 446-473. doi:10.3390/make4020020
- Lu, S., Christie, G., Nguyen, T., Freeman, J., & Hsu, E. (2022). Applications of Artificial Intelligence and Machine Learning in Disasters and Public Health Emergencies. *Disaster Medicine and Public Health Preparedness*, 16(4), 1674-1681. doi:10.1017/dmp.2021.125
- Lundgren, R., & McMakin, A. (2018). *A Handbook for communicating environmental, safety, and Health risks*. WILEY.
- Ospina, R., Gondim, J., Leiva, V., & Castro, C. (2023). An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic: Methodology and Case Study in Brazil. *Mathematics*, 11(14), 3069. doi:10.3390/math11143069
- Sankaranarayanan, S., Prabhakar, M., Satish, S., & Jain, P. (2020). Flood prediction based on weather parameters using deep learning. *Journal of Water and Climate Change*, 11(4), 1766-1783. doi:10.2166/wcc.2019.321
- Secretaria de Gestion de Riesgos. (12 de 2023). *Secretaria de Gestion de Riesgos*. Obtenido de <https://www.gestionderiesgos.gob.ec/>
- Shaiba, H., Alaashoub, N., & Alzahrani, A. (2018). Applying Machine Learning Methods for Predicting Sand Storms. *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)*, 1-5. doi:10.1109/CAIS.2018.8441998
- Sit, M. A., Koylu, C., & Demir, I. (2019). Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma. *International Journal of Digital Earth*, 12(11), 1205-1229. doi:10.1080/17538947.2018.1563219
- Wannous, C., & Velasquez, G. (2017). United Nations Office for Disaster Risk Reduction (UNISDR)—UNISDR's Contribution to Science and Technology for Disaster Risk Reduction and the Role of the International Consortium on Landslides (ICL). *Advancing Culture of Living with Landslides*, 1, 109-115. doi:10.1007/978-3-319-59469-9\_6

Yasir, A., Hussain, F., & Haque, M. (2024). Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review. *Accident Analysis & Prevention*, 194. doi:<https://doi.org/10.1016/j.aap.2023.107378>