



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas

Identificación de puntos calientes mediante análisis multivariante y geoespacial de
extorsiones en la ciudad Guayaquil empleando aprendizaje automático

PROYECTO INTEGRADOR

Previo a la obtención del Título de:

Matemático

Presentado por:

Pedro Joel Luzuriaga Cedeño

Oscar Antonio Scott Macías

GUAYAQUIL - ECUADOR

Año: 2025

DEDICATORIA

A la memoria de Pedro, mi padre, y Gabriel, mi hijo, cuya presencia resuena en cada latido de mi corazón. A mi madre, Estrella, y mis hermanas, Katherine, Luisa y Angie, quienes, con su amor y apoyo constante, fueron mi escudo y mi espada a lo largo de este camino.

Pedro Luzuriaga C.

DEDICATORIA

A Dios, por darme la fortaleza y el coraje para culminar esta etapa. A mi amada familia: mi madre Kariely, mi padre Oscar, mi querida hermana Adamaris y a mi abuelo Jorge, cuya presencia vive siempre en mi corazón. Y a todos aquellos que saben amar, esto es para todos ustedes.

Oscar Scott M.

AGRADECIMIENTOS

Agradezco a los profesores, Johny Pambabay y Sergio Bauz, por su guía y aporte a mi formación académica a través de este trabajo. A la profesora Elimar Marchan, por ser un gran ejemplo en mi carrera. A mi compañero de tesis, Oscar, quien sin su apoyo este proyecto no habría sido posible. A Dayana, mi novia, por su amor y apoyo incondicional, que me dieron la fuerza para continuar y culminar mis estudios, incluso en los momentos más difíciles.

Pedro Luzuriaga C.

AGRADECIMIENTOS

Expreso mi más profundo agradecimiento a los profesores Johny Pambabay y Sergio Bauz por su paciencia, confianza y dedicación al impartir sus conocimientos a este proyecto. A la profesora Elimar Marchan, por su constante apoyo. A todos mis docentes de la carrera, en especial al profesor Jesús Aponte, a quien admiro profundamente por su entrega y paciencia en la enseñanza. Finalmente, mi gratitud a mi compañero de tesis, Pedro, cuya colaboración fue fundamental en este proceso.

Oscar Scott M.

DECLARACIÓN EXPRESA

“Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; *Pedro Joel Luzuriaga Cedeño, Oscar Antonio Scott Macías,* y damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual”



Pedro Luzuriaga C.



Oscar Scott M.

EVALUADORES

Luz Marchán Mendoza, Ph.D.

PROFESOR DE LA MATERIA

Jhony Pambababy, Ph.D.

DIRECTOR

RESUMEN

La extorsión en Guayaquil constituye un fenómeno delictivo que impacta gravemente la seguridad y estabilidad social. El objetivo principal de esta investigación es analizar la distribución espacial y temporal de los delitos de extorsión en la ciudad, integrando técnicas estadísticas y modelos geoespaciales para identificar zonas críticas, planteando la hipótesis de que este delito no se distribuye de manera aleatoria, sino que se concentra en áreas específicas debido a factores socioeconómicos. Se analizó, mediante herramientas como RStudio y GeoDa, la base de datos de llamadas de emergencia al 911 Guayaquil desde el 2021 hasta 2024 proporcionada por la Empresa Pública Municipal para la Gestión de Riesgos y Control de Seguridad de Guayaquil en colaboración con el Centro de Estudios e Investigaciones Estadística. Se utilizaron índices como la Razón Estandarizada de Mortalidad y Riesgo Relativo, análisis multivariante mediante el Análisis de Componentes Principales y métodos geoespaciales como Getis-Ord G_i^* y el Índice Local de Moran. Los resultados mostraron un incremento significativo de la tasa de extorsión por año en todos los distritos y circuitos de la ciudad de Guayaquil. A través de los métodos geoespaciales y el análisis multivariante se obtuvieron mapas de significancia y agrupamiento para identificar puntos calientes en circuitos. Se concluye que existen agrupaciones significativas de circuitos considerados puntos calientes desde la zona noroeste hasta la zona sur de Guayaquil, mientras que los puntos fríos están concentrados en la zona noreste de la ciudad, evidenciando una distribución no aleatoria de este delito.

Palabras Clave: Extorsión, Mapas de Criminalidad, Análisis Espacial, Riesgo Relativo.

ABSTRACT

Extortion in Guayaquil constitutes a criminal phenomenon that seriously affects security and social stability. The main objective of this research is to analyze the spatial and temporal distribution of extortion crimes in the city, integrating statistical techniques and geospatial models to identify critical areas, positing the hypothesis that this crime is not randomly distributed but rather concentrated in specific areas due to socioeconomic factors. Using tools such as RStudio and GeoDa, the database of emergency calls to 911 Guayaquil from 2021 to 2024 provided by the Empresa Pública Municipal para la Gestión de Riesgos y Control de Seguridad de Guayaquil in collaboration with the Centro de Estudios e Investigaciones Estadística was analyzed. Indices such as the Standardized Mortality Ratio and Relative Risk were used, as well as multivariate analysis through Principal Component Analysis and geospatial methods such as Getis-Ord G_i^ and the Local Moran's Index. The results showed a significant increase in the extortion rate per year in all districts and circuits of the city of Guayaquil. Through geospatial methods and multivariate analysis, significance and clustering maps were obtained to identify hot spots in circuits. It is concluded that there are significant clusters of circuits considered hot spots from the northwest to the south of Guayaquil, while cold spots are concentrated in the northeast of the city, evidencing a non-random distribution of this crime.*

Keywords: *Extortion, Crime Maps, Spatial Analysis, Relative Risk.*

ÍNDICE GENERAL

RESUMEN	I
ABSTRACT	II
ABREVIATURAS	V
ÍNDICE DE FIGURAS	VII
ÍNDICE DE TABLAS	VIII
CAPÍTULO 1	1
1. INTRODUCCIÓN	1
1.1 Descripción del problema	2
1.2 Justificación del problema	3
1.3 Objetivos	6
1.3.1 Objetivo General	6
1.3.2 Objetivos Específicos	6
1.4 Marco teórico	6
1.4.1 Antecedentes	6
1.4.2 Palabras clave	9
1.4.3 Base teórica	11
1.4.4 Modelos matemáticos	12
CAPÍTULO 2	24
2. METODOLOGÍA	24

2.1	Recopilación de datos	24
2.2	Análisis descriptivo	26
2.3	Análisis Jerárquico Bayesiano	27
2.4	Análisis Multivariante	31
2.5	Análisis Geoespacial	31
CAPÍTULO 3		35
3. RESULTADOS Y ANÁLISIS		35
3.1	Análisis estadístico	35
3.1.1	Análisis descriptivo	35
3.1.2	Análisis multivariante	50
3.2	Análisis geoespacial	54
3.2.1	Getis-Ord G_i^*	55
3.2.2	Índice local de Moran	62
CAPÍTULO 4		72
4. CONCLUSIONES Y RECOMENDACIONES		72
BIBLIOGRAFÍA		
APÉNDICES		

ABREVIATURAS

KDE: Kernel Density Estimation

PCA: Principal Component Analysis

t-SNE: t-distributed Stochastic Neighbor Embedding

MDS: Multidimensional Scaling

SMR: Standardized Mortality Ratio

RR: Relative Risk

ÍNDICE DE FIGURAS

Figura 3.1	Evolución de extorsiones en Guayaquil desde el año 2021 al 2024 según semanas.	36
Figura 3.2	Comparación de Forest Plots para RR y SMR	41
Figura 3.3	Tasas de delitos por circuitos en los distritos del Norte.	42
Figura 3.4	Tasas de delitos por circuitos en los distritos del Centro.	42
Figura 3.5	Tasas de delitos por circuitos en los distritos del Sur.	43
Figura 3.6	Tasas de delitos por circuitos en los distritos del Oeste.	43
Figura 3.7	Dispersión de las dos componentes por medio de MDS	52
Figura 3.8	Dispersión de las dos componentes por medio de PCA	52
Figura 3.9	Dispersión de las dos componentes por medio de t-SNE	53
Figura 3.10	Evolución temporal de los mapas de llamadas de auxilio, distribuidos por distritos y coordenadas geográficas.	54
Figura 3.11	Contigüidad de la Torre en sus distintos órdenes	56
Figura 3.12	Contigüidad de la Reina en sus distintos órdenes.	57
Figura 3.13	Resultados de SMR con orden 1 con Getis-Ord G_i^*	58
Figura 3.14	Resultados de SMR con orden 1 y 2 incluido con Getis-Ord G_i^*	59
Figura 3.15	Resultados de RR con orden 1 con Getis-Ord G_i^*	60
Figura 3.16	Resultados de RR con orden 1 y 2 incluido con Getis-Ord G_i^*	61
Figura 3.17	Resultados de RR con orden 1 por I de Moran Local Univariante	63

Figura 3.18 Diagrama de dispersión con coeficiente de correlación de RR de orden 1 por I de Moran Local Univariante	63
Figura 3.19 Resultados de RR con orden 1 y 2 incluido por I de Moran Local Univariante	64
Figura 3.20 Diagrama de dispersión con coeficiente de correlación de RR de orden 2 por I de Moran Local Univariante	64
Figura 3.21 Resultados de PC1svd con orden 1 por I de Moran Local Univariante	66
Figura 3.22 Diagrama de dispersión con coeficiente de correlación de PC1svd de orden 1 por I de Moran Local Univariante	66
Figura 3.23 Resultados de PC1svd con orden 1 y 2 incluido por I de Moran Local Univariante	67
Figura 3.24 Diagrama de dispersión con coeficiente de correlación de PC1svd de orden 1 y 2 incluido por I de Moran Local Univariante	67
Figura 3.25 Resultados de PC2svd con orden 1 por I de Moran Local Univariante	69
Figura 3.26 Diagrama de dispersión con coeficiente de correlación de PC2svd de orden 1 por I de Moran Local Univariante	69
Figura 3.27 Resultados de PC2svd con orden 1 y 2 incluido por I de Moran Local Univariante	70
Figura 3.28 Diagrama de dispersión con coeficiente de correlación de PC2svd de orden 1 y 2 incluido por I de Moran Local Univariante	70

ÍNDICE DE TABLAS

Tabla 1.1	Distritos administrativos de Guayaquil	8
Tabla 3.1	Tasas por cada 100000 habitantes por distritos para los años 2021 al 2024	37
Tabla 3.2	Valores de SMR por Distrito (Ordenados de Mayor a Menor)	38
Tabla 3.3	Valores de RR por Distrito (Ordenados de Mayor a Menor)	40
Tabla 3.4	Tasas por año en los Circuitos más significativos, los 5 Circuitos más altos	44
Tabla 3.5	Tasas por año en los Circuitos más significativos, los 5 Circuitos más bajos	44
Tabla 3.6	Valores de SMR y RR en los circuitos del Distrito Florida	45
Tabla 3.7	Valores de SMR y RR en los circuitos del Distrito Nueva Prosperina	45
Tabla 3.8	Valores de SMR y RR en los circuitos del Distrito Pascuales	46
Tabla 3.9	Valores de SMR y RR en los circuitos del Distrito Portete	46
Tabla 3.10	Valores de SMR y RR en los circuitos del Distrito Sur	47
Tabla 3.11	Valores de SMR y RR en los circuitos del Distrito Progreso	47
Tabla 3.12	Valores de SMR y RR en los circuitos del Distrito Ceibos	47
Tabla 3.13	Valores de SMR y RR en los circuitos del Distrito Esteros	48
Tabla 3.14	Valores de SMR y RR en los circuitos del Distrito 9 de Octubre	48
Tabla 3.15	Valores de SMR y RR en los circuitos del Distrito Modelo	49
Tabla 3.16	Circuitos con RR más alto	50
Tabla 3.17	Circuitos RR más bajo	50

CAPÍTULO 1

1. INTRODUCCIÓN

La extorsión es un fenómeno delictivo que afecta gravemente la seguridad y la estabilidad social en ciudades como Guayaquil. Su impacto altera las dinámicas económicas y sociales, reforzando un clima de inseguridad. Este proyecto tiene como objetivo principal analizar cómo se distribuyen espacial y temporalmente los delitos de extorsión en la ciudad, empleando herramientas estadísticas avanzadas y métodos de análisis geoespacial.

En este trabajo se busca analizar la distribución geográfica de los delitos de extorsión en la ciudad, identificando las zonas con mayor incidencia y explorando posibles relaciones con variables como pobreza, desempleo y densidad poblacional. Al integrar datos geoespaciales y estadísticas multivariantes, se pretende descubrir patrones que permitan comprender mejor este fenómeno y aportar información clave para la toma de decisiones.

Se propone que la extorsión en Guayaquil no se distribuye de forma aleatoria, sino que se concentra en zonas específicas debido a factores socioeconómicos y características del entorno urbano.

Mediante la identificación de las zonas en donde es más frecuente y propensa la presencia de este delito se podrá generar mapas de riesgo y reconocer patrones relevantes que sirvan como base para el diseño de estrategias de prevención e intervención más efectivas en los sectores más afectados de la ciudad.

1.1 Descripción del problema

El delito de extorsión en América Latina constituye un fenómeno social complejo profundamente vinculado al crimen organizado, la corrupción política y los factores socioeconómicos de la región. La extorsión, caracterizada por el uso de amenazas para exigir pagos bajo coacción, adopta diversas formas y se ve intensificada por la debilidad de los marcos jurídicos y la prevalencia de la corrupción (Solis and Rojas, 2008).

Los altos índices de extorsión que se han suscitado en el Ecuador generan un clima de miedo, afectando no solo a la inversión y al crecimiento económico, sino también a la estabilidad social y a la confianza en las instituciones. El sector empresarial y los ciudadanos enfrentan constantes amenazas de los grupos delictivos, lo que impacta negativamente al desarrollo integral del país (Velez, 2024).

La relación entre pobreza y delincuencia resalta las raíces socioeconómicas que alimentan varias actividades delictivas, como la extorsión. Este vínculo destaca cómo las condiciones de desigualdad económica favorecen la proliferación de este tipo de delitos, complicando aún más su abordaje y prevención (Bourguignon, 2009).

Aunque la extorsión sigue siendo un desafío importante, surge la idea que las reformas legales y el refuerzo de la aplicación de la ley son relevantes para desarticular estas redes delictivas. Sin embargo, la naturaleza profundamente arraigada de la corrupción complica estos esfuerzos, lo que indica que un enfoque polifacético es pertinente para combatir la extorsión.

El estudio de la delincuencia, en particular la extorsión, en entornos urbanos como Guayaquil, ha llamado la atención debido al aumento de la violencia y sus implicaciones para la seguridad pública y la planificación urbana de ciudades densamente pobladas.

El fenómeno de la extorsión no es solo una violación de los derechos individuales, sino también un factor que contribuye a una inestabilidad social más amplia. Según Bertelli (2023), los delitos relacionados con la extorsión han alcanzado niveles alarmantes, afectando principalmente a familias y negocios, que suelen ser blanco de ataques tras negarse a pagar tarifas exorbitantes. Este contexto evidencia el impacto devastador que la extorsión ha tenido en la seguridad y el tejido social de las comunidades.

La dinámica espacial de los delitos revela la existencia de puntos calientes o zonas de concentración, es decir, las áreas donde el delito se manifiesta de forma significativa. Estas zonas suelen caracterizarse por altos niveles de violencia e intimidación, exacerbando la inseguridad y la desconfianza en las instituciones. Además, la falta de estrategias específicas para abordar las particularidades de la extorsión en diferentes contextos geográficos agrava su proliferación y dificulta la implementación de soluciones efectivas (He et al., 2022).

Este estudio, además de explorar la necesidad de utilizar técnicas estadísticas avanzadas y modelos geoespaciales para detectar focos de extorsión en Guayaquil, también proporciona ideas prácticas para los responsables políticos y los organismos encargados de hacer cumplir la ley.

Al centrarse en la cartografía de los puntos críticos de extorsión, se pretende proporcionar una comprensión más profunda del problema, lo que resulta esencial para diseñar estrategias efectivas que promuevan la seguridad y la sostenibilidad social en la comunidad.

1.2 Justificación del problema

Un estudio estadístico geoespacial de los delitos de extorsión en Guayaquil se justifica por varios factores críticos que ponen de relieve desafíos únicos y la necesidad de intervenciones

específicas en la ciudad.

En América Latina, en particular, en Guayaquil-Ecuador se ha producido un aumento significativo de la extorsión y la delincuencia organizada, problemas que se han arraigado profundamente en muchas comunidades. La diversificación de los grupos criminales ha llevado a un aumento de las actividades de extorsión, que a menudo están vinculadas a patrones más amplios de violencia e inseguridad en la región. Comprender la dinámica espacial de estos delitos puede ayudar a identificar los puntos conflictivos y fundamentar las estrategias de aplicación de la ley (Solis and Rojas, 2008).

Existe una considerable variación espacial y temporal en los índices de criminalidad en Guayaquil y América Latina. Por ejemplo, las tasas de homicidio fluctúan significativamente entre distintos países y regiones, influidas por factores como la competencia entre bandas y las condiciones socioeconómicas. Un análisis geoespacial es fundamental para comprender estos patrones, ya que permite identificar cómo se distribuyen geográfica y temporalmente los delitos de extorsión. Este enfoque ofrece información clave para abordar las dinámicas delictivas de manera más efectiva y fundamentar la formulación de políticas públicas dirigidas a mejorar la seguridad (Briceño-León et al., 2008).

La extorsión está estrechamente relacionada con factores socioeconómicos como la pobreza, el desempleo y la falta de acceso a la educación. Diversos estudios han demostrado que las zonas con mayores índices de pobreza suelen presentar una mayor incidencia de actividades delictivas graves, incluida la extorsión. El análisis geoespacial y estadístico permite explorar de manera más precisa la relación entre variables socioeconómicas y las tasas de extorsión. Esto proporciona una base sólida para identificar patrones y desarrollar estrategias

específicas que aborden las causas estructurales de la delincuencia (Solis and Rojas, 2008).

La extorsión suele ser un delito poco denunciado debido al temor a represalias y a la falta de confianza en las instituciones encargadas de la seguridad pública. Estas “cifras negras” dificultan una evaluación precisa de su impacto en las comunidades afectadas, creando un vacío crítico en la comprensión de su alcance real.

Un enfoque estadístico geoespacial contribuye a estimar la magnitud del problema, analizando los patrones de incidentes denunciados en combinación con datos sociodemográficos y económicos. Esta metodología ofrece una visión más completa y permite superar parcialmente las limitaciones asociadas a los eventos no denunciados (Gaviria and Pagés, 2002).

La información obtenida de un estudio estadístico geoespacial tiene el potencial de fundamentar las respuestas políticas a múltiples niveles en áreas identificadas como vulnerables. Al identificar las zonas de alto riesgo de extorsión, los gobiernos pueden asignar recursos de manera más eficaz, poner en marcha programas de prevención teniendo en cuenta las necesidades de la comunidad y mejorar las estrategias de aplicación de la ley fortaleciendo la capacidad de respuesta, aumentando así la seguridad de los ciudadanos (Ratcliffe, 2010).

En síntesis, un estudio estadístico geoespacial sobre los delitos de extorsión en Guayaquil es esencial para captar las complejidades de la delincuencia en la ciudad. Este enfoque proporciona un marco integral para analizar la interacción espacial y temporal de diversos factores, generando insumos críticos para el desarrollo de estrategias más efectivas de prevención e intervención frente a la delincuencia.

1.3 Objetivos

1.3.1 Objetivo General

Adaptar un modelo de análisis estadístico, geoespacial y automático que clasifique los mecanismos utilizados por los extorsionadores, la demografía de las víctimas y los factores que contribuyen a la extorsión, para la identificación de los sitios con mayor ocurrencia de este delito.

1.3.2 Objetivos Específicos

- Clasificar las distintas formas de extorsión, como la intimidación física, el chantaje, la ciberextorsión y los chantajes de protección, para la identificación patrones comunes.
- Examinar las características demográficas y geográficas de las víctimas de extorsión, para la determinación de perfiles y delimitación de zonas de mayor vulnerabilidad.

1.4 Marco teórico

1.4.1 Antecedentes

El crecimiento de delitos en los últimos años en América Latina es alarmante, especialmente en Ecuador, considerado un país muy inseguro. Desde el 2020 en adelante las denuncias por extorsión aumentaron de manera acelerada. Particularmente, en Guayaquil el delito de extorsión tuvo un incremento significativo. Estudios indican que delitos como la extorsión atentan de manera directa al crecimiento económico, inversión y flujo del turismo. La región, en general, sufre un desbalance en relación con sus años anteriores, teniendo que cancelar festividades patronales y eventos públicos por causa de dicho delito. En el trabajo de He et al. (2022), se emplearon los modelos Moran's I y el Getis-Ord Gi para estudiar la

distribución espacial de diversos delitos, como robo o homicidio pero no se estudió la extorsión como un caso de delito. Esta es una oportunidad para desarrollar una investigación que permita identificar los puntos calientes de extorsión en la ciudad de Guayaquil y, además, diseñar un modelo que caracterice los factores socioeconómicos y geográficos asociados. La implementación de este proyecto de investigación contribuye a la criminología y, además, proporciona información útil para la toma de decisiones del Gobierno Municipal y Nacional, generando estrategias para la prevención del delito.

En este contexto, resulta relevante considerar cómo la ciudad de Guayaquil se encuentra organizada territorialmente, ya que esto permite la implementación de políticas y estrategias en función de las necesidades locales. La ciudad está situada en la zona 8 de planificación, junto con Samborondón y Durán, y está dividida en 10 distritos administrativos. Cada distrito, a su vez, se subdivide en circuitos que facilitan la gestión territorial y la provisión de servicios. Esta estructura territorial, establecida por la Secretaría Nacional de Planificación y Desarrollo (SENPLADES) en su acuerdo ministerial 557-2012 página 5, tiene como propósito principal desconcentrar los servicios públicos y atender de manera eficiente las necesidades de la ciudadanía. La desconcentración de servicios se lleva a cabo a través de seis ministerios, siendo particularmente relevante para este estudio el Ministerio del Interior, en coordinación con la Policía Nacional del Ecuador, quienes lideran la estrategia para la prevención y el control del delito. Cada distrito fue renombrado por la Empresa Pública Municipal para la Gestión de Riesgos y Control de Seguridad de Guayaquil (SEGURA EP), en la recopilación de información para su base de datos. (Véase Tabla 1.1)

Tabla 1.1.
Distritos administrativos de Guayaquil

Código-Distrito	Nombre del Distrito	Nombre del Distrito según SEGURA EP
09D01	Ximena 1	Sur
09D02	Ximena 2	Esteros
09D03	García Moreno, Letamendi, Ayacucho, Olmedo, Sucre, 9 de Octubre, Urdaneta, Pedro Carbo, Roca	9 de Octubre
09D04	Febres Cordero	Portete
09D05	Tarqui 1 - Tenguel	Modelo
09D06	Tarqui 2	Florida
09D07	Pascuales 1	Pascuales
09D08	Pascuales 2	Nueva Prosperina
09D09	Tarqui 3	Ceibos
09D10	Progreso, El Morro, Posorja, Zona de expansión de Guayaquil	Progreso

La subdivisión de los distritos por circuitos es la siguiente:

- Distrito 9 de Octubre: comprende los circuitos 9 de Octubre, Antepara, Chile, Cristo del Consuelo, Garay, Venezuela y Victoria.
- Distrito Ceibos: comprende los circuitos Chongón, Los Ceibos y Puerto Hondo.
- Distrito Esteros: comprende los circuitos Esteros, Malvinas Norte, Malvinas Sur, Trinitaria Norte y Trinitaria Sur.
- Distrito Florida: comprende los circuitos Bellavista, Florida, Juan Montalvo, Mapasingue,

Martha de Roldós y Samanes.

- Distrito Modelo: comprende los circuitos Alborada, Atarazana, Guayacanes, Kennedy, Sauces y Urdesa.
- Distrito Nueva Prosperina: comprende los circuitos Flor de Bastión, Fortín, Monte Sinai, Nueva Prosperina, Nuevo Guayaquil y Paraíso de la Flor.
- Distrito Pascuales: comprende los circuitos Bastión Popular Norte, Bastión Popular Sur, Monte Bello, Mucho Lote, Orquídeas, Pascuales, Puente Lucía y San Francisco.
- Distrito Portete: comprende los circuitos Batallón del Suburbio, Chala, Cisne, Portete, Salinas y Suburbio.
- Distrito Progreso: comprende los circuitos Morro, Posorja y Progreso
- Distrito Sur: comprende los circuitos 7 Lagos, Centenario, Floresta, Guasmo, Puná y Unión de Bananeros.

1.4.2 Palabras clave

Según el artículo 185 del Código Orgánico Integral Penal del Ecuador (Asamblea Nacional del Ecuador, 2014), “la persona que, con el propósito de obtener provecho personal o para un tercero, obligue a otro, con violencia o intimidación, a realizar u omitir un acto o negocio jurídico en perjuicio de su patrimonio o el de un tercero” comete el delito de extorsión.

En el trabajo realizado por Echeverri et al. (2018) en Medellín, se identificaron nueve modalidades de extorsión: contacto directo, rifas ficticias, tarjetas personales de presentación, recibos de parqueo, compra forzada de artículos, llamadas telefónicas, cuentas bancarias para

consignación, cartas de circulación pública y secuestro extorsivo. Según el Observatorio Ecuatoriano de Crimen Organizado (2023) estas mismas modalidades pueden ser adaptadas al contexto ecuatoriano.

En el marco de este trabajo, un punto caliente (término traducido del inglés hot spot) es un área geográfica caracterizada por una incidencia criminal significativamente mayor al promedio. Esto implica un riesgo de victimización más elevado para quienes la habitan o la transitan.

Además, los puntos calientes pueden diferir en intensidad, reflejando variaciones en cuánto exceden los índices delictivos respecto al promedio (Eck et al., 2005).

En el análisis multivariante se utilizan técnicas de agrupación, de aglomeración y jerárquicas, así mismo, reducción de la dimensión para analizar conjuntos de datos complejos. Estos métodos permiten identificar patrones y relaciones entre múltiples variables relacionadas con los casos de extorsión (Hair et al., 2014).

El análisis geoespacial en los Sistemas de Información Geográfica (SIG) desempeña un papel fundamental en la visualización y el análisis de datos espaciales. Los métodos de detección de puntos calientes, como Getis-Ord G_i^* , pueden identificar zonas con niveles significativamente altos de incidentes de extorsión (Getis and Ord, 1992).

En términos de modelización estadística, algoritmos avanzados como HOTTOR se emplean para detectar puntos calientes dentro de grandes conjuntos de datos. Este algoritmo se centra en la identificación de cambios significativos en las variables a través de diferentes condiciones, superando a los métodos tradicionales como ANOVA en conjuntos de datos complejos (Anderson et al., 2016).

Para La Integración de fuentes de datos se destaca la importancia de combinar diversos

tipos de datos (por ejemplo, informes sobre delitos, factores socioeconómicos, etc.) para crear una visión global de la dinámica de la extorsión, facilita una comprensión más matizada de los factores que contribuyen a los patrones delictivos (Aguilar et al., 2014).

1.4.3 Base teórica

La investigación sobre el delito de extorsión en Guayaquil se apoya en teorías criminológicas, geográficas y socioeconómicas que ayudan a comprender la dinámica de este tipo de crimen y sus implicaciones para la seguridad pública y el desarrollo de la sociedad. Estas teorías ofrecen una base para la utilización de modelos estadísticos y geospaciales en la detección y caracterización de puntos calientes de extorsión, a la vez que justifican la relevancia de los factores socioeconómicos en el análisis delictivo.

La teoría de la desorganización social sostiene que ciertos factores, como la pobreza, la densidad de población y la heterogeneidad social, pueden generar condiciones de desorganización en ciertas áreas geográficas, lo cual incrementa la probabilidad de actividades delictivas. Esta teoría es fundamental para comprender cómo el contexto socioeconómico y el deterioro de las condiciones de vida en determinadas zonas urbanas pueden contribuir a la manifestación del delito de extorsión, al crear un ambiente donde el control social es insuficiente y las redes criminales encuentran condiciones favorables para operar (Shaw and McKay, 1942).

La teoría de las actividades rutinarias establece que el delito tiene mayor probabilidad de ocurrencia cuando existen tres elementos: un delincuente motivado, una víctima potencial y la ausencia de un guardián capaz de prevenir el crimen. Esta teoría es relevante para el análisis de la extorsión, ya que permite entender cómo ciertos factores, como la actividad económica de una zona y la densidad comercial pueden incrementar el riesgo de este delito. Los patrones

de movilidad y actividades económicas en áreas urbanas de Guayaquil pueden ser claves para identificar puntos calientes de extorsión, justificando así el enfoque geoespacial de este estudio (Cohen and Felson, 1979).

La teoría de la ventaja acumulativa, postula que una ventaja inicial en algún aspecto tiende a acumularse con el tiempo, mientras que las desventajas también se mantienen. En el contexto de la criminalidad, esta teoría ayuda a explicar cómo ciertos lugares pueden volverse más vulnerables a crímenes recurrentes, como la extorsión, debido a factores históricos, socioeconómicos y geográficos. La identificación de puntos calientes mediante modelos geoespaciales como Moran's I y Getis-Ord Gi permite visualizar la persistencia y concentración de eventos delictivos en zonas específicas de Guayaquil (Sampson, 1993).

1.4.4 Modelos matemáticos

La identificación de áreas de alta concentración de delitos de extorsión se basa en diversos métodos geoespaciales que pueden aplicarse mediante herramientas como QGIS y RStudio. A continuación, se describen detalladamente los métodos principales, acompañados de sus notaciones matemáticas.

Estimación de Densidad mediante Núcleos

La Estimación de Densidad mediante Núcleos (KDE, por sus siglas en inglés) es una técnica estadística empleada para estimar la función de densidad de probabilidad de una variable aleatoria. En el ámbito del análisis espacial, la KDE es particularmente eficaz para identificar áreas con alta concentración de delitos, como la extorsión, al generar una representación continua de la densidad de eventos en una región específica. Esto facilita la

visualización y el análisis de patrones espaciales en los datos delictivos.

La KDE consta de los siguientes componentes:

- Kernel: La función kernel se emplea para asignar pesos a los puntos según su distancia a una ubicación específica. Entre las funciones de kernel más utilizadas se encuentran el kernel gaussiano, el cual se expresa matemáticamente como:

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} \quad (1.1)$$

donde u representa la distancia al punto que se desea estimar.

- Ancho de banda (Bandwith): El parámetro de ancho de banda h regula el nivel de suavidad en la estimación de densidad. Un valor pequeño de h puede generar un ajuste excesivo, mostrando demasiados detalles, mientras que un valor grande puede suavizar en exceso, perdiendo información relevante. La selección del ancho de banda es fundamental y puede realizarse mediante métodos como la validación cruzada.
- Estimación de la densidad KDE: esta estimación en un punto x se calcula de la siguiente forma:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (1.2)$$

donde n representa al número de puntos de datos, x_i los puntos de datos observados y K la función Kernel.

En el ámbito del análisis de la delincuencia, el método KDE es una herramienta útil para identificar áreas con altas concentraciones de incidentes, como los relacionados con la extorsión.

El proceso suele implicar los siguientes pasos:

1. Recopilación de datos: Reunir información espacial sobre los incidentes de extorsión registrados.
2. Cálculo de la densidad mediante KDE: Generar una superficie de densidad que permita resaltar las zonas con mayores tasas de criminalidad.
3. Visualización: Representar los niveles de densidad mediante gradientes de color, lo que facilita la detección de las áreas denominadas puntos calientes.

Como ejemplo de aplicación, en un estudio realizado en la ciudad de San Francisco, se utilizó KDE para visualizar puntos calientes de delincuencia aplicando anchos de intervalos variables (por ejemplo, 750 *m*, 1.000 *m* y 1.500 *m*). Los mapas de densidad resultantes ayudaron a las fuerzas del orden a asignar los recursos de forma eficaz identificando las zonas con una agrupación significativa de incidentes de extorsión (Kalinic and Krisp, 2018).

Estadísticas de Escaneo Espacial

El método de Estadísticas de Escaneo Espacial (SSS, por sus siglas en inglés) es una técnica estadística ampliamente utilizada para identificar puntos calientes, en diversos contextos, incluido el estudio de la criminalidad. Este método resulta particularmente eficaz para localizar regiones con altos niveles de incidentes, como casos de extorsión, mediante el análisis detallado de datos espaciales.

El SSS opera mediante la creación de múltiples ventanas de barrido superpuestas, generalmente de forma circular, en una región geográfica específica, comparando la cantidad de eventos observados con los esperados bajo una hipótesis nula. Cada ventana se analiza para determinar su significancia estadística al comparar la cantidad de eventos ocurridos en su

interior, como incidentes de extorsión, con el número esperado de sucesos según una hipótesis nula que asume una distribución aleatoria.

Notación matemática

Sea:

- n : el número total de regiones del área de estudio.
- x_i : el número observado de sucesos en la región i .
- E_i : el número esperado de sucesos en la región i , calculado a partir de la tasa global de sucesos.

El ratio (cociente) de likelihood (likelihood ratio) para un conglomerado determinado, definido por una ventana de exploración, puede expresarse como:

$$L = \frac{\prod_{i \in C} x_i!}{\prod_{i \in C} E_i!} \quad (1.3)$$

Donde:

- C indica el conjunto de regiones dentro de la ventana de exploración.
- El producto recorre todas las regiones i incluidas en el conjunto C .

Para determinar si una agrupación identificada es significativa desde el punto de vista estadístico, el SSS emplea simulaciones de Monte Carlo. Este proceso genera una distribución de cocientes de verosimilitud bajo la hipótesis nula, la cual se compara con el cociente observado para obtener un valor p que indique la significancia estadística.

Recientes desarrollos han incorporado un enfoque flexible en las estadísticas de escaneo espacial, que permite el uso de ventanas de exploración con formas no circulares. Esta

característica es clave para identificar conglomerados irregulares en datos, como aquellos relacionados con actividades delictivas que se ajustan a trazados urbanos o accidentes geográficos.

El SSS ha demostrado ser una herramienta efectiva para identificar áreas con alta incidencia delictiva, incluyendo delitos como la extorsión. Diversos estudios han mostrado que este enfoque permite localizar conglomerados de alta criminalidad y apoyar a las autoridades en la distribución de recursos y la planificación estratégica (He et al. (2022), Runadi and Widyaningsih (2017), Tango and Takahashi (2005)).

Estadístico Getis-Ord G

El estadístico Getis-Ord G, en particular su variante Getis-Ord G_i^* , es una herramienta ampliamente utilizada para identificar patrones espaciales, especialmente en estudios relacionados con el análisis del crimen, como los delitos de extorsión. Este método permite detectar agrupaciones relevantes de valores altos (puntos calientes) y valores bajos (puntos fríos) en conjuntos de datos geoespaciales.

El estadístico Getis-Ord G_i^* se formula matemáticamente de la siguiente manera (Getis and Ord, 1992):

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{\sqrt{\frac{(n \sum_{j=1}^n w_{i,j}^2) - (\sum_{j=1}^n w_{i,j})^2}{n-1}}} \quad (1.4)$$

Donde:

- x_j representa el valor del atributo en la característica j .
- $w_{i,j}$ es el peso espacial que vincula las características i y j .
- n es el número total de características.

- $\bar{X} = \frac{\sum_{j=1}^n x_j}{n}$ es la media de los valores del atributo.

El valor calculado de G_i^* se interpreta como un z -score que indica el nivel de importancia de la agrupación.

- Un z -score positivo alto señala una agrupación significativa de valores altos (puntos calientes).
- Un z -score negativo bajo indica una agrupación significativa de valores bajos (puntos fríos).
- Un z -score cercano a cero sugiere que no hay evidencia de agrupación significativa.

Para evaluar la significancia estadística, se utilizan los valores p , donde los niveles de confianza más comunes son:

- $p < 0.01$: nivel de confianza del 99%.
- $p < 0.05$: nivel de confianza del 95%.
- $p < 0.10$: nivel de confianza del 90%.

En el análisis de los delitos de extorsión, el estadístico Getis-Ord G_i^* puede utilizarse para detectar áreas con una incidencia significativamente alta, en comparación con lo que se esperaría bajo una distribución aleatoria (Jana and Sar, 2016). El proceso de análisis consta de varios pasos clave:

- Preparación de datos: Recopilar y preprocesar datos espaciales relacionados con los incidentes de extorsión, asegurando su calidad y coherencia.

- Esquema de ponderación: Definir una matriz de pesos espaciales que represente las relaciones entre las diferentes unidades geográficas (por ejemplo, barrios, distritos o regiones).
- Cálculo: Aplicar la fórmula del estadístico G_i^* a cada unidad geográfica para determinar su nivel de agrupación espacial.
- Interpretación: Analizar los z -score y los valores p obtenidos para identificar puntos calientes (agrupaciones significativas de valores altos) y puntos fríos (agrupaciones significativas de valores bajos).

Local Moran's I

Local Moran's I es un método estadístico utilizado para detectar conglomerados espaciales y valores atípicos, especialmente útil para analizar fenómenos como los delitos de extorsión, mediante la identificación de áreas con asociaciones espaciales significativas. Este método ayuda a identificar áreas en las que la ocurrencia de sucesos, como los delitos, es significativamente mayor o menor de lo esperado en función del contexto circundante.

El estadístico Local Moran's I se define como:

$$I_i = \frac{n}{W} \cdot \frac{\sum_{j=1}^n w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sigma^2} \quad (1.5)$$

donde:

- I_i es Local Moran's I para la característica i .
- n es el número total de características.
- W es la suma de las ponderaciones de la característica i .

- w_{ij} es el peso espacial entre las características i y j .
- x_i y x_j son los valores de la variable de interés en los lugares i y j .
- \bar{x} es la media de la variable en todas las características.
- σ^2 es la varianza de la variable en todas las características.

Los valores de Local Moran's I pueden interpretarse del siguiente modo:

- Los valores positivos indican una agrupación de valores similares (alto-alto o bajo-bajo).
- Los valores negativos indican que el valor de una característica es un valor atípico en relación con sus vecinos (alto-bajo o bajo-alto).

Para aplicar este método se emplean los siguientes pasos:

- Preparación de los datos: Agregar datos sobre delincuencia en una estructura espacial (por ejemplo, celdas de cuadrícula o unidades geográficas).
- Creación de la matriz de pesos: Definir una matriz de pesos espaciales que cuantifique la relación entre unidades vecinas. Puede basarse en la contigüidad o en la distancia.
- Cálculo de Local Moran's I: Aplicar la fórmula para calcular el Local Moran's I para cada unidad espacial.
- Prueba de significancia estadística: Utilizar permutaciones para generar un valor p para cada estadístico Local Moran's I para determinar la significancia. Un valor p bajo indica una agrupación significativa.

- Mapeo de resultados: Visualizar los resultados en un mapa para identificar los puntos calientes (áreas con altos índices de criminalidad) y los puntos fríos (áreas con bajos índices de criminalidad).

Local Moran's I no sólo identifica los conglomerados, sino que también detecta los valores atípicos, proporcionando una comprensión más matizada de los patrones espaciales en comparación con métodos globales como Global Moran's I o la estadística Getis-Ord G_i^* (Vidanapathirana, 2021).

Para abordar herramientas como la KDE, el SSS, el estadístico Getis-Ord G y el Local Moran's I, es importante que exista una visión de los métodos de reducción de la dimensión que permiten la caracterización multivariante de los delitos de extorsión. Las técnicas de reducción de dimensionalidad son fundamentales para analizar conjuntos de datos complejos, especialmente en el contexto del análisis multivariante de los delitos de extorsión. Estos métodos permiten simplificar los datos mientras se conservan sus características esenciales, lo que facilita su interpretación y modelado. En este caso, nos centraremos específicamente en una técnica particular, explorando su formulación matemática y revisando la bibliografía relevante que respalda su aplicación.

Embebido Estocástico de Vecinos Distribuidos en t

El Embebido Estocástico de Vecinos Distribuidos en t (t-SNE), es una técnica altamente efectiva para visualizar datos de alta dimensionalidad, al proyectarlos en un espacio de menor dimensión, generalmente de dos o tres dimensiones. Es particularmente útil para identificar la estructura en conjuntos de datos complejos, como aquellos relacionados con delitos de extorsión, al agrupar datos similares y separar aquellos que son diferentes.

El algoritmo t-SNE consta de dos etapas principales:

1. Distribución de probabilidad de alta dimensión: Para un conjunto de n objetos de alta dimensión x_1, x_2, \dots, x_n , t-SNE calcula primero las probabilidades condicionales $p_{j|i}$ que representan la similitud entre los puntos x_i y x_j . Se define como:

$$p_{j|i} = \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}\right)} \quad (1.6)$$

Aquí, σ_i es el bandwidth del kernel gaussiano, que puede determinarse mediante un parámetro llamado *perplejidad* (*perplexity*):

$$H(P_i) = -\sum_j p_{j|i} \log_2 p_{j|i}$$

La perplejidad es una medida del número efectivo de vecinos y suele fijarse entre 5 y 50.

2. Distribución de probabilidad de baja dimensión: En el espacio de baja dimensión, t-SNE define una distribución de probabilidad $q_{i|j}$, similar a una distribución t de Student con un grado de libertad (que tiene colas más pesadas que una gaussiana):

$$q_{i|j} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_i - y_k\|^2)^{-1}} \quad (1.7)$$

El objetivo de t-SNE es minimizar la divergencia de Kullback-Leibler entre estas dos distribuciones:

$$C = KL(P\|Q) = \sum_i \sum_j p_{j|i} \log\left(\frac{p_{j|i}}{q_{i|j}}\right) \quad (1.8)$$

Aplicaciones de t-SNE en delitos de extorsión

Diversos estudios recientes han empleado t-SNE, para analizar conjuntos de datos complejos, permitiendo identificar patrones y relaciones relevantes. Algunos ejemplos destacados incluyen:

- Detección de la violencia basada en el contexto: Una investigación utilizó t-SNE para estudiar diálogos relacionados con la extorsión y el chantaje, identificando patrones específicos en el uso del lenguaje según el tipo de delito. Este enfoque permitió comprender cómo se agrupan los diálogos en función de sus características semánticas (Kim et al., 2024).
- Constructos psicológicos en la delincuencia: Otro estudio aplicó t-SNE para visualizar motivaciones psicológicas subyacentes en el comportamiento delictivo, incluida la extorsión. Al proyectar datos psicológicos de alta dimensionalidad en un espacio reducido, los investigadores detectaron patrones de agrupamiento que reflejan motivaciones vinculadas con estos delitos (Walsh et al., 2020).
- Técnicas de visualización comparativa: Se evaluó la efectividad de t-SNE frente a otros métodos de reducción de dimensionalidad, como PCA y UMAP. Los resultados demostraron que t-SNE supera a estas técnicas al revelar estructuras subyacentes en datos que involucran diversas clases de delitos, incluyendo la extorsión (Van der Maaten and Hinton, 2008).

Aplicando este marco teórico, en la presente investigación se pretende desarrollar un modelo sólido que no sólo identifique los puntos calientes, sino que también caracterice los factores subyacentes que contribuyen a los incidentes de extorsión. Es decir, la integración de técnicas estadísticas, geoespaciales y de modelización automática presenta una vía para mejorar la comprensión de los focos de extorsión en Guayaquil. Mediante el análisis de ciertos factores, este estudio pretende aportar valiosos conocimientos sobre la dinámica de la delincuencia urbana e informar sobre respuestas políticas eficaces.

Finalmente, se espera que este enfoque aporte información útil para los responsables políticos y las fuerzas del orden.

CAPÍTULO 2

2. METODOLOGÍA

2.1 Recopilación de datos

Teniendo en cuenta la división de Guayaquil, la Empresa Pública Municipal para la Gestión de Riesgos y Control de Seguridad de Guayaquil, (Segura EP) recolectó las llamadas de auxilio al 911 desde el 1 de enero del 2021 al 23 de septiembre del 2024. En conjunto con la Coordinación Técnica de Estadística y Datos, se elaboró una base de datos en Excel donde se recopilan los llamados de auxilio relacionados con posibles casos de extorsión. En esta base de datos se incluyeron las siguientes variables:

- Año: año de la emergencia (desde el 2021 al 2024).
- Mes: mes de la emergencia (desde enero a diciembre).
- Semana: semana en la que se produjo la emergencia (se etiquetaron a las semanas en orden cronológico desde la semana 1 hasta la semana 53).
- Día: día de la emergencia (de lunes a domingo).
- Fecha: fecha de la emergencia (día/mes/año).
- Fecha hora: fecha y hora de la emergencia.

- Latitud: latitud de la emergencia.
- Longitud: longitud de la emergencia.
- Cantón: cantón de la emergencia (Guayaquil).
- Distrito: distrito de la emergencia.
- Circuito: circuito de la emergencia.
- Subcircuito: subcircuito de la emergencia.
- Descripción: descripción de la emergencia (desde abril de 2024 se tiene una descripción completa y ya no se incluyen comentarios).
- Tipología: tipología de la emergencia (todas fueron encasilladas como extorsión).
- idComentario: identificador del comentario de la emergencia.
- Comentarios: comentarios de la emergencia (solo hasta marzo de 2024).

Posteriormente, se incorporó la variable Modalidad, la cual, siguiendo el trabajo realizado por Echeverri et al. (2018), permitió clasificar cada llamado de emergencia según el tipo de modalidad de extorsión correspondiente, realizando un análisis de los comentarios asociados a cada caso.

Con el objetivo de obtener indicios que contribuyan al desarrollo de la investigación, se realizó un análisis descriptivo mediante el programa RStudio, de los datos recolectados. Este análisis busca explorar patrones y tendencias que puedan ser clave para una comprensión más profunda de la extorsión en Guayaquil. A continuación, se detalla la metodología empleada para este análisis.

2.2 Análisis descriptivo

Se calculó la tasa de incidentes por cada 100,000 habitantes, utilizando las poblaciones proporcionadas por el Centro de Estudios e Investigaciones Estadísticas (CEIE) de la ESPOL. Estas tasas se obtuvieron dividiendo el número de incidentes registrados en cada estrato (distrito y circuito) entre su población respectiva, y multiplicando el resultado por 100,000 (Véase ecuación (2.1)). Este cálculo se realizó en cada uno de los años 2021, 2022, 2023 y 2024 para identificar variaciones en las tasas a lo largo del tiempo.

$$Tasa = \frac{\text{Cantidad de incidentes}}{\text{Población por estrato}} \cdot 100000 \quad (2.1)$$

El cálculo de los casos esperados para cada distrito en el año 2024 se realizó tomando como referencia los incidentes registrados durante los años 2021, 2022 y 2023. Estos casos esperados representan el número de incidentes de extorsión que se anticiparían si el comportamiento de cada distrito siguiera el patrón poblacional de Guayaquil. Para determinar el número esperado de casos E_i en un distrito i , se aplicó la técnica de estandarización indirecta mediante la ecuación (2.2).

$$E_i = \sum_j r_j^{(s)} n_j^{(i)} \quad (2.2)$$

Aquí, $r_j^{(s)}$ es la tasa de incidencia en el estrato j (calculada como el número de casos dividido entre la población), y $n_j^{(i)}$ corresponde a la población del estrato j en el distrito i .

El cálculo del número de casos esperados se realizó por medio de RStudio, utilizando el paquete `SpatialEpi()` y su función `expected()`. Para ello, se proporcionaron como argumentos el vector que contiene la población de cada estrato en cada área, el vector que incluye el número de casos correspondientes a cada estrato en cada área, y la cantidad total de estratos considerados.

Un caso de criminalidad puede abordarse como un caso de epidemiología forense, ya que esta disciplina utiliza métodos de salud pública para analizar y valorar hipótesis relacionadas con problemas delictivos en una comunidad, facilitando la evaluación de intervenciones y estrategias preventivas (Zúñiga, 2018).

Para clasificar las modalidades de extorsión y evaluar el nivel de riesgo en diferentes distritos, se utilizaron dos medidas comúnmente empleadas en el área de la salud:

El SMR (Standardized Mortality Ratio) compara el número de casos observados con los casos esperados, calculados utilizando la función `expected` y considerando los datos de 2021, 2022 y 2023 junto con sus respectivas poblaciones. A partir de estos estratos, se estimó el valor esperado. Para cada distrito o circuito i el SMR viene dada por

$$SMR_i = \frac{Y_i}{E_i} \quad (2.3)$$

Donde Y_i representa el número de casos observados en 2024 y E_i el número de casos esperados.

2.3 Análisis Jerárquico Bayesiano

El Riesgo Relativo (RR), es una medida que expresa, de manera relativa, la relación entre la probabilidad de que un evento ocurra en un grupo expuesto y la probabilidad de que ocurra en un grupo no expuesto. (Mendivelso and Rodríguez, 2019).

Esta medida evalúa la probabilidad de que un incidente, ocurra en un estrato específico en comparación con otro, proporcionando una clasificación de los distritos según su nivel de peligro.

Se calcula de la siguiente manera:

$$RR = \frac{\text{Incidencia en el grupo expuesto}}{\text{Incidencia en el grupo no expuesto}} \quad (2.4)$$

Donde,

- $RR > 1$: El evento es más probable en el grupo expuesto.
- $RR = 1$: No hay diferencia en la probabilidad entre los grupos.
- $RR < 1$: El evento es menos probable en el grupo expuesto.

Teniendo en cuenta la naturaleza y el contexto de este trabajo, se adaptó un modelo basado en la estimación del riesgo de cáncer de pulmón en Pensilvania, el cual provee un enfoque estadístico bayesiano para estimar el riesgo relativo ajustado por factores estructurales, no estructurales donde se incluyen covariables espaciales.

Los pasos necesarios para ajustar el modelo y visualizar los resultados utilizando el paquete INLA en R se describen a continuación.

Sean Y_i y E_i el número de casos observados y esperados de extorsión, respectivamente, y sea RR_i el riesgo relativo asociado al distrito i , con $i = 1, 2, \dots, n$. El modelo se define como:

$$Y_i | RR_i \sim \text{Poisson}(E_i \times RR_i), i = 1, 2, \dots, n \quad (2.5)$$

Además, el riesgo relativo RR_i se modela en la escala logarítmica mediante la expresión:

$$\log(RR_i) = \beta_0 + \sum_{i=1}^n (\beta_i \times cov_i) + u_i + v_i \quad (2.6)$$

Donde

- β_0 es un intercepto
- β_i es el coeficiente de la covariable
- u_i es el efecto espacial estructurado,

$$u_i | U_{-i} \sim N \left(\bar{u}_{\delta_i}, \frac{1}{\tau_u n_{\delta_i}} \right) \quad (2.7)$$

- v_i es el efecto no estructurado,

$$v_i \sim N\left(0, \frac{1}{\tau_v}\right) \quad (2.8)$$

Para definir el efecto aleatorio espacial, se generó una matriz de vecindad utilizando las funciones `poly2nb()` y `nb2INLA()` del paquete `spdep`. Inicialmente, con la función `poly2nb()`, se creó una lista de vecinos, donde las áreas vecinas de una región específica son aquellas con las que comparte bordes. Cada elemento de la lista `nb` corresponde a una región y contiene los índices de sus vecinos. Por ejemplo, `n[[2]]` almacena los vecinos asociados al área 2.

Se mostraron los vecinos de distritos específicos de Guayaquil utilizando un mapa. Por ejemplo, se puede mostrar los vecinos de los distritos 6, 9 y 10. En primer lugar, se creó un objeto `SpatialPolygonsDataFrame` con el mapa de Distritos, y los datos que contiene una variable llamada `distrito` con los nombres de los distritos, y una variable ficticia llamada `neigh` que indica los vecinos de los distritos 6, 9 y 10. `neigh` es igual a 1 para los distritos que son vecinos de los distritos 6, 9 y 10, y 0 en caso contrario.

A continuación, se añadieron variables llamadas `long` y `lat` con las coordenadas de cada variable, y una variable `ID` con el `id` de los distritos.

Se creó el mapa con la función `ggplot()` de `ggplot2`. En primer lugar, se convirtió el mapa, que es un objeto espacial de la clase `SpatialPolygonsDataFrame`, en un simple objeto `feature` de la clase `sf` con la función `st_as_sf()` del paquete `sf`.

Se creó un mapa mostrando la variable `neigh`, y añadiendo etiquetas con los ids de las zonas.

Posteriormente, se utilizó la función `nb2INLA()` para convertir la lista `nb` en un archivo denominado `map.adj`, que contiene la representación de la matriz de vecindad requerida para

trabajar con INLA. Este archivo `map.adj` se guarda automáticamente en el directorio de trabajo, el cual puede verificarse mediante la función `getwd()`. Luego, se cargó el archivo `map.adj` empleando la función `inla.read.graph()` del paquete INLA, y el resultado se asignó al objeto `g`, que se utilizará más adelante para definir el modelo espacial en INLA.

El modelo considera dos efectos aleatorios: u_i , que representa la variabilidad espacial estructurada residual, y v_i , que captura el ruido no estructurado. Para ello, se incluyeron dos vectores en el *data frame* para indexar estos efectos aleatorios. El vector de índices asociado a u_i se denominó `re_u`, mientras que el vector correspondiente a v_i se denominó `re_v`. Ambos vectores, `re_u` y `re_v`, se definieron como $1, \dots, n$, donde n es la cantidad de condados.

La fórmula del modelo se definió incluyendo la variable respuesta en el lado izquierdo y los efectos, tanto fijos como aleatorios, en el lado derecho. Los efectos aleatorios se especificaron mediante la función `f()`, cuyos argumentos incluyen el nombre de la variable y el modelo seleccionado. Para u_i , se utilizó `model="besag"` con la matriz de vecindad representada por `g`, mientras que para v_i , se optó por `model="iid"`.

Se realizó la estimación del modelo utilizando la función `inla()` con las distribuciones previas por defecto de INLA. Para ello, se definieron la fórmula, la familia, el conjunto de datos y el número esperado de casos. Además, se configuró el parámetro `control.predictor` con `list(compute=TRUE)` para calcular las distribuciones posteriores de las predicciones.

Este análisis permitió identificar los distritos más conflictivos y mostrar indicios de los patrones temporales y geográficos de las extorsiones, sentando las bases para un entendimiento más profundo y una aplicación oportuna de los métodos matemáticos en la detección geográfica de los puntos calientes.

2.4 Análisis Multivariante

Para analizar la relación entre las tasas de extorsión de 2021 a 2024, junto con los índices SMR y RR, se aplicaron técnicas de reducción de dimensionalidad mediante el programa GeoDa. Este análisis permitió la reducción de dimensiones.

Inicialmente, se utilizó un archivo en formato Shapefile (.shp), proporcionado por el CEIE, al cual se le añadieron las variables de interés. Este archivo fue cargado en GeoDa, donde se seleccionó la opción "Cluster", que permite realizar análisis multivariantes con métodos como PCA, MDS y t-SNE.

Para cada técnica, se seleccionaron las seis variables (tasas de extorsión 2021-2024, SMR y RR) y para cada caso se obtuvieron dos variables. En el caso de PCA, se aplicó el criterio de Kaiser, el cual sugiere retener únicamente los componentes con valores propios mayores a 1.

Finalmente, los resultados obtenidos con PCA, MDS y t-SNE fueron comparados para evaluar cuál de las técnicas proporcionaba una representación más efectiva de la estructura de los datos.

2.5 Análisis Geoespacial

En la aplicación de los modelos matemáticos se utilizó GeoDa, el cual es un software encasillado dentro de los Sistemas de Información Geográfica (SIG), especializado en el análisis espacial y estadístico, diseñado para facilitar la visualización, exploración y modelado de patrones espaciales.

Se inició el método de Local Moran's I univariante, desde el icono de la barra de herramientas Mapas de clústeres de GeoDa. Se incluyó como opción de nivel superior en la lista

desplegable resultante, de esta forma hace que aparezca el cuadro de diálogo Configuración de variables, que enumera las variables disponibles, así como el archivo de pesos que en este caso, se utilizaron los valores de SMR y RR calculados anteriormente.

Dando clic en Aceptar aparece un cuadro de diálogo para seleccionar el número y los tipos de gráficos que sirvieron para el análisis posterior. Se seleccionó el mapa de conglomerados, el mapa de significación y el gráfico de dispersión de Moran.

El mapa de significancia destaca las ubicaciones donde el estadístico local resulta significativo, representando el nivel de significancia con tonos de verde progresivamente más oscuros. Este mapa inicia en $p < 0.05$ e incluye todas las categorías de significancia relevantes según el número de permutaciones realizado.

Se elaboraron mapas de conglomerados espaciales, que identificaron áreas con asociaciones espaciales significativas. Estas áreas se clasificaron en las siguientes categorías:

- Alto-Alto: Zonas con altas tasas de extorsión rodeadas de zonas similares.
- Bajo-Bajo: Áreas con bajas tasas rodeadas de zonas similares.
- Alto-Bajo y Bajo-Alto: Valores atípicos espaciales.

Se mantuvo la configuración de 99999 permutaciones y un valor p de 0.05, variando el valor de p para 0.01 y 0.001 para la visualización del mapa de conglomerados y el mapa de significación.

Se utilizó la opción de Guardar resultados en el menú de opciones, para su análisis posterior.

En la búsqueda del método más eficiente para la identificación de los puntos calientes de los casos de extorsiones en Guayaquil, se utilizaron adicionalmente los estadísticos Getis-Ord G y

su variante Getis-Ord G_i^* en el programa GeoDa. Se seleccionaron dentro del menú desplegable generado por el icono de la barra de herramientas Cluster Maps.

A continuación, se eligieron las opciones use row-standardized weights y generar sólo el Cluster Map marcando las casillas correspondientes.

Se seleccionó la configuración predeterminada de 999 permutaciones, con $p = 0.05$, donde se obtuvo el mapa de significación para ambos estadísticos.

En los mapas mostrados se identificaron también las áreas con asociaciones espaciales significativas mencionadas anteriormente (Alto-Alto, Bajo-Bajo, Alto-Bajo y Bajo-Alto) con la salvedad de que los valores espaciales atípicos se clasifican ahora como parte de los conglomerados (uno en el grupo Alto-Alto y otro en el grupo Bajo-Bajo).

Para el contraste de resultados en esta investigación, se utilizó QGIS, el cual es otro sistema de información geográfica (SIG) que permite la visualización, análisis y edición de datos espaciales. QGIS se utiliza en diversas disciplinas para la gestión y análisis de datos geoespaciales, permitiendo realizar tareas como la geolocalización, la modelización espacial y la representación gráfica de información geográfica.

Para la visualización de mapas y puntos con mayor aglomeración por cada año se utilizaron los plugins Density Análisis y Fast Density Analisis para la aplicación de la Estimación de Densidad mediante Núcleos (KDE), donde se seleccionó un tamaño del pixel de 0.001 y un radio estimado de 4.5 kilómetros.

Se agregaron archivos en formato Shapefile con la división de Guayaquil por circuitos que provienen del Centro de Estudios e Investigaciones Estadísticas, en adición, se convirtió la base de datos de Excel en el formato CSV donde se leyeron las coordenadas x como Longitud y y como

Latitud como parámetros espaciales

Finalmente se ejecutó la KDE en QGis para su posterior presentación de los mapas por cada uno de los años 2021, 2022, 2023 y 2024, donde se encuentran pintados los sitios con mayor aglomeración en los casos de extorsión color rojo más intenso y con un color rojo menos intenso donde hay poca aglomeración. De esta forma, se puede analizar la evolución de los casos de extorsiones reportados y determinar cuáles son los focos de mayor propagación.

CAPÍTULO 3

3. RESULTADOS Y ANÁLISIS

Los primeros resultados se obtuvieron a partir de un análisis estadístico, mostrando la evolución del delito, por medio de cantidades absolutas y tasas de riesgo. Con estos resultados se realizó un análisis multivariante que dió paso a realizar un análisis geoespacial, para la detección de puntos calientes.

3.1 Análisis estadístico

El análisis descriptivo fue realizado con Rstudio y el análisis multivariante fue realizado con GeoDa.

3.1.1 *Análisis descriptivo*

Con los datos de las llamadas de emergencia del 911-Guayaquil, se elaboró un análisis inicial para visualizar el comportamiento de la evolución de las extorsiones en Guayaquil. Con esto se pudo observar el incremento considerable que tuvo este delito en el Cantón, con lo cual era necesario realizar un análisis a profundidad y comprender con mayor precisión los patrones delictivos. (Véase Figura 3.1)

Figura 3.1.

Evolución de extorsiones en Guayaquil desde el año 2021 al 2024 según semanas.



En la figura 3.1 se observa un incremento a partir del tercer trimestre del año 2022, y al prevalecer un índice tan alto, se muestran indicios de delincuencia organizada, por lo tanto fue necesario entender el comportamiento en cada uno de los distritos de la ciudad.

Análisis por distritos

Como fue mencionado anteriormente, el Cantón Guayaquil cuenta con 10 distritos, 9 urbanos y 1 rural según SEMPLADES. En cada distrito se encontró las tasas por cada 100000 habitantes, en la tabla 3.1 se muestra los resultados obtenidos.

Tabla 3.1.*Tasas por cada 100000 habitantes por distritos para los años 2021 al 2024*

Distrito	2021	2022	2023	2024
9 De Octubre	3.17	12.31	40.83	60.55
Ceibos	1.81	12.64	42.99	65.72
Esteros	0.67	5.70	24.96	49.49
Florida	1.93	9.63	44.49	65.49
Modelo	3.77	22.30	92.23	96.07
Nueva Prosperina	0.25	9.68	47.23	97.66
Pascuales	1.91	13.02	93.59	88.10
Portete	1.62	8.83	29.16	58.28
Progreso	3.49	1.75	25.99	63.55
Sur	1.27	4.08	61.94	75.20

En la tabla 3.1 se puede observar un aumento del delito por cada uno de los distritos de la ciudad, teniendo un crecimiento más alto en Nueva Prosperina, pasando de una tasa de 0.25 en 2021 a una tasa de 97.66 en el 2024. En el distrito Modelo también se observa un cambio significativo, pasando de una tasa de 3.77 a una de 96.07 desde el 2021 al 2024.

Aunque esto es un análisis bastante acertado, se realizó el análisis por medio del SMR. A partir de los datos del 2021 al 2023, se encontró un valor esperado a partir de la función `expected()` en Rstudio.

(Véase el código A.1, que muestra el cálculo del valor esperado y el SMR).

En el código, se multiplicó el valor esperado, denotado como v , por la fracción $\frac{270}{366}$. Este ajuste se realizó porque el valor esperado proyecta el total para el año 2024, pero los datos

disponibles solo cubren hasta el 25 de septiembre de 2024, lo que corresponde a 270 días del año. De este modo, se multiplicó por este factor para asegurar que ambos valores representen el mismo período de tiempo, alineando la proyección con los datos disponibles hasta esa fecha.

Con los valores esperados y la cantidad de delitos observados en el 2024, se calculó el SMR al dividir los casos observados para los esperados.

Dado que se trabaja en un dashboard con Shiny en RStudio, los valores ajustados se almacenaron en una variable reactiva para su posterior uso en el dashboard.

En la tabla 3.2 se muestran los valores obtenidos.

Tabla 3.2.
Valores de SMR por Distrito (Ordenados de Mayor a Menor)

Distrito	SMR
Progreso	2.79
Nueva Prosperina	2.34
Esteros	2.16
Portete	2.02
Florida	2.15
Ceibos	1.57
Sur	1.53
9 De Octubre	1.47
Modelo	1.54
Pascuales	1.11

Un valor de $SMR < 1$ indica que el distrito tiene un factor protector, mientras que un $SMR > 1$ sugiere la presencia de un factor contaminante, conceptos tomados del ámbito de la

salud. Sin embargo, al analizar los datos, se observa que todos los distritos presentan un factor contaminante. El distrito con el menor factor contaminante es Pascuales, mientras que el más afectado es Progreso

El análisis mediante el SMR resulta eficaz para estudiar el comportamiento de cada circuito, al utilizar los datos de años anteriores para calcular el valor esperado de un año y compararlo con los casos observados. Sin embargo, esta metodología puede verse influenciada por la población de cada distrito. Por ello, también se calculó el índice RR para complementar el análisis

(Véase el código A.2, que muestra el cálculo del RR).

El RR incluye un análisis espacial, teniendo en cuenta el archivo ShapeFile del Cantón Guayaquil para evaluar las relaciones de vecindad. Primero, se carga el archivo ShapeFile, y posteriormente, se asigna un nombre a cada polígono según corresponda.

En la variable `datos` se encuentra la base de datos analizada, la cual se utiliza para contar la cantidad de delitos por cada distrito. Se realizó un reordenamiento para utilizar los datos de manera pertinente y se asignaron nombres según corresponde a `tabla_tasa_reordenada`.

Se creó la variable espacial `mapa_exto` con el ShapeFile cargado y la `tabla_tasa_reordenada`, se creó la vecindad de cada distrito por guardado en la variable `nb`, también se creó el grafo `g` para posteriormente crear los vectores `re_u` y `re_v` que son el efecto espacial estructurado y no estructurado respectivamente.

Se guardó en la variable `formula` toda la formula necesaria para agregarla dentro de la función `inla()`. La fórmula contiene los valores observados en 2024 y como covariables contiene las tasas del 2021 al 2023 y los efectos estructurados y no estructurados con sus modelizaciones correspondientes.

Con esto se obtuvieron tres resultados importantes, el RR, el LL y el UL, donde LL representa el límite inferior y UL representa el límite superior. Dichos resultados se pueden observar en la tabla 3.3.

Tabla 3.3.
Valores de RR por Distrito (Ordenados de Mayor a Menor)

Distrito	RR
Esteros	2.28
Nueva Prosperina	2.16
Portete	2.04
Pascuales	1.77
Ceibos	1.83
Florida	1.71
Progreso	1.63
9 De Octubre	1.56
Sur	1.50
Modelo	1.06

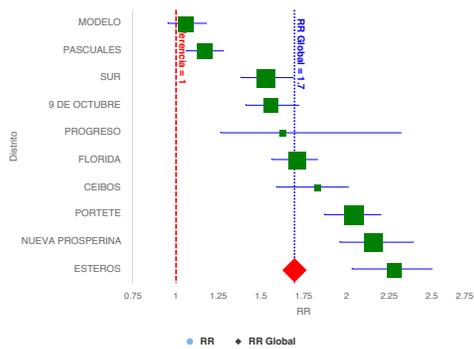
En la tabla 3.3 podemos observar que los distritos con mayor factor contaminante son Esteros, Nueva Prosperina y Portete y los que cuentan con un menor factor contaminante son Sur, 9 De Octubre y Pascuales, donde se evidenció que el distrito Sur bajó una posición respecto al SMR y Esteros siendo el de la primera posición, contrario al SMR que estaba en la tercera posición.

Para facilitar la visualización de la información del SMR y RR, en la figura 3.2 se presentan los Forest Plots de cada medida, donde el tamaño del cuadrado representa la población de cada

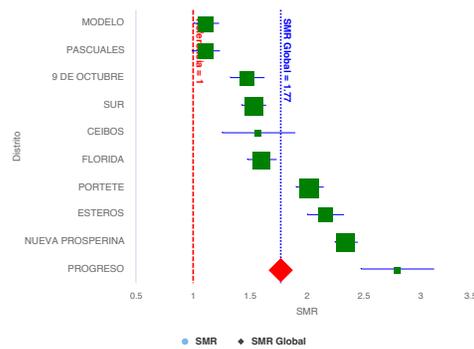
distrito.

Figura 3.2.
Comparación de Forest Plots para RR y SMR

(a)
Forest Plot de RR



(b)
Forest Plot de SMR

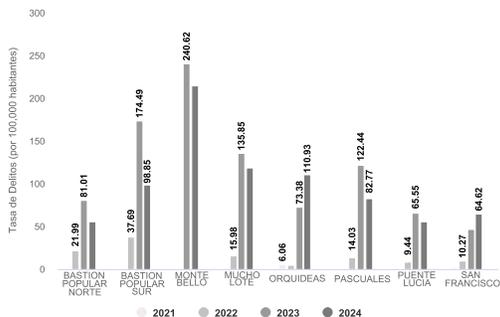


Como se puede observar, ninguno de los distritos presenta un factor protector, ya sea según el SMR o el RR. Por lo tanto, el análisis a nivel de distrito no resulta tan efectivo para comprender la distribución del delito. Es necesario realizar un análisis más detallado a nivel de circuito dentro de cada distrito, lo que permite identificar estratos específicos y obtener información más precisa y relevante.

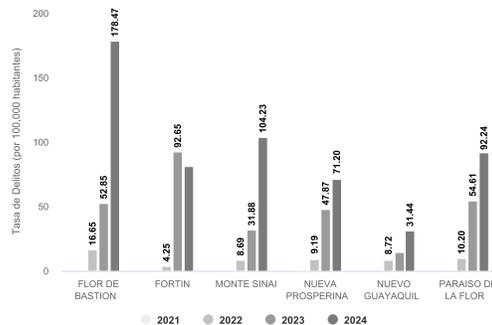
De forma similar, se elaboró primero un análisis por cada distrito, analizando sus circuitos y presentando sus resultados. En las figuras 3.3, 3.4, 3.5 y 3.6 se muestran los resultados por cada uno de los distritos.

Figura 3.3.
Tasas de delitos por circuitos en los distritos del Norte.

(a)
Tasas por Circuitos del Distrito Pascuales



(b)
Tasas por Circuitos del Distrito Nueva Prosperina



(c)
Tasas por Circuitos del Distrito Florida

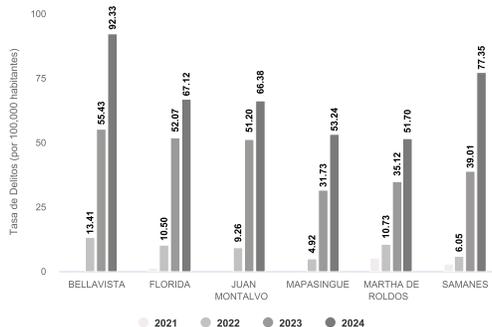
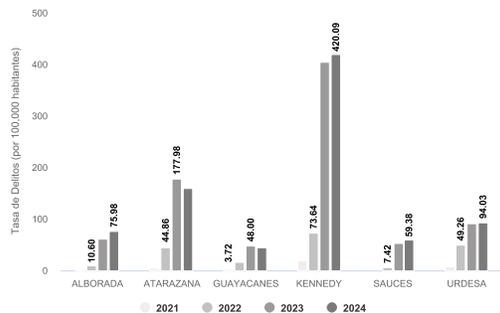


Figura 3.4.
Tasas de delitos por circuitos en los distritos del Centro.

(a)
Tasas por Circuitos del Distrito Modelo



(b)
Tasas por Circuitos del Distrito 9 de Octubre

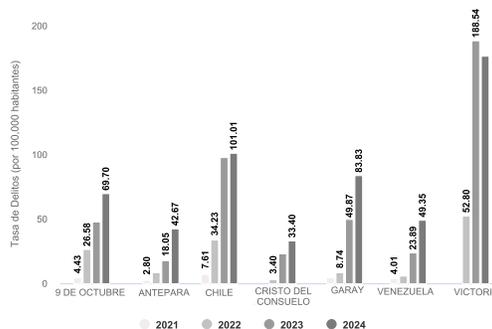
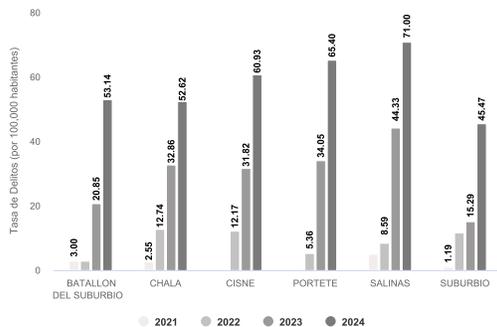
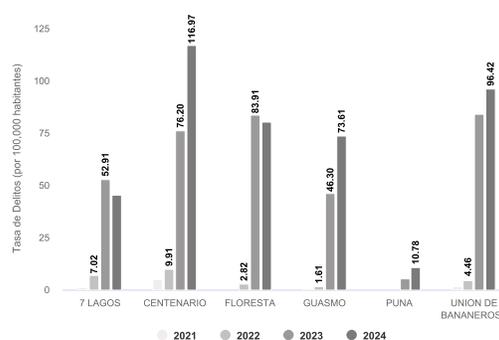


Figura 3.5.
Tasas de delitos por circuitos en los distritos del Sur.

(a)
Tasas por Circuitos del Distrito Portete



(b)
Tasas por Circuitos del Distrito Sur



(c)
Tasas por Circuitos del Distrito Esteros

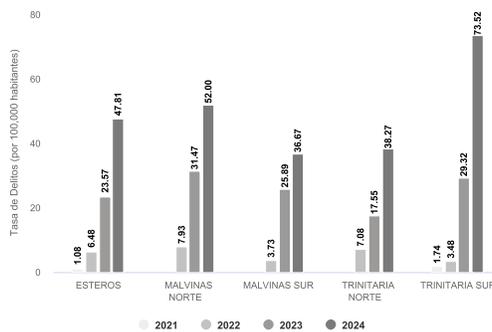
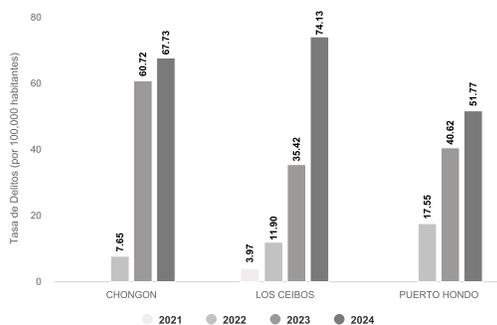
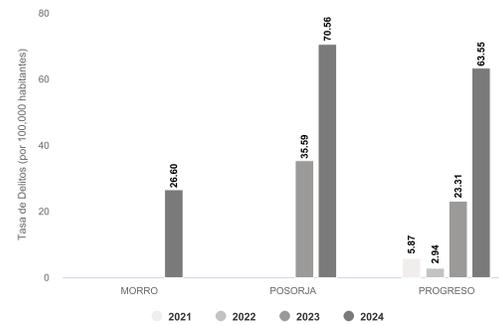


Figura 3.6.
Tasas de delitos por circuitos en los distritos del Oeste.

(a)
Tasas por Circuitos del Distrito Ceibos



(b)
Tasas por Circuitos del Distrito Progreso



En general, se observa un incremento significativo en todos los distritos, entre 2021 y 2024. Los circuitos que más destacan en este aumento, se presentan en la tabla 3.4, donde los primeros cinco circuitos con mayores tasas en los últimos dos años son Kennedy, Monte Bello, Flor de Bastión, Victoria y Atarazana. Por otro lado, en la tabla 3.5 se encuentran los cinco circuitos con las tasas más bajas: Puna, Morro, Nuevo Guayaquil, Malvinas Sur y Trinitaria Norte.

Tabla 3.4.

Tasas por año en los Circuitos más significativos, los 5 Circuitos más altos

Distrito	Circuito	2021	2022	2023	2024
Modelo	Kennedy	19.64	73.64	404.23	420.09
Pascuales	Monte Bello	0.00	0.00	240.62	214.69
Nueva Prosperina	Flor de Bastion	0.00	16.65	52.85	178.47
9 de Octubre	Victoria	0.00	52.80	188.54	176.55
Modelo	Atarazana	5.61	44.86	177.98	159.91

Tabla 3.5.

Tasas por año en los Circuitos más significativos, los 5 Circuitos más bajos

Distrito	Circuito	2021	2022	2023	2024
Esteros	Trinitaria Norte	1.74	7.08	17.55	38.27
Esteros	Malvinas Sur	0.00	3.73	25.89	36.67
Nueva Prosperina	Nuevo Guayaquil	0.00	8.72	14.41	31.44
Progreso	Morro	0.00	0.00	35.59	26.60
Sur	Puna	0.00	0.00	5.43	10.78

De manera similar, se calcularon tanto el SMR, como el RR para cada circuito de los distintos distritos. Los valores obtenidos para ambos índices se muestran en las tablas 3.6 a

3.15.

Tabla 3.6.*Valores de SMR y RR en los circuitos del Distrito Florida*

Circuito	SMR	RR
Samanes	2.20	1.90
Bellavista	1.84	1.74
Mapasingue	1.99	1.85
Juan Montalvo	1.50	1.41
Florida	1.44	1.46
Martha de Roldos	1.38	1.42

Tabla 3.7.*Valores de SMR y RR en los circuitos del Distrito Nueva Prosperina*

Circuito	SMR	RR
Flor de Bastion	3.52	3.21
Monte Sinai	3.52	3.37
Paraiso de la Flor	1.95	1.92
Nuevo Guayaquil	1.86	2.03
Nueva Prosperina	1.67	1.73
Fortin	1.15	1.28

Tabla 3.8.
Valores de SMR y RR en los circuitos del Distrito Pascuales

Circuito	SMR	RR
Orquideas	1.80	1.71
San Francisco	1.50	1.49
Monte Bello	1.22	1.30
Puente Lucia	1.02	1.35
Mucho Lote	1.07	1.08
Pascuales	0.83	0.95
Bastion Popular Norte	0.74	0.88
Bastion Popular Sur	0.64	0.75

Tabla 3.9.
Valores de SMR y RR en los circuitos del Distrito Portete

Circuito	SMR	RR
Portete	2.27	2.19
Suburbio	2.20	2.15
Batallon del Suburbio	2.71	2.47
Cisne	1.90	1.90
Salinas	1.67	1.71
Chala	1.50	1.69

Tabla 3.10.*Valores de SMR y RR en los circuitos del Distrito Sur*

Circuito	SMR	RR
Puna	2.71	1.89
Guasmo	2.07	2.01
Centenario	1.76	1.71
Union de Bananeros	1.47	1.49
Floresta	1.27	1.34
7 Lagos	1.02	1.13

Tabla 3.11.*Valores de SMR y RR en los circuitos del Distrito Progreso*

Circuito	SMR	RR
Progreso	2.71	2.19
Posorja	2.71	2.58
Morro	1.00	2.05

Tabla 3.12.*Valores de SMR y RR en los circuitos del Distrito Ceibos*

Circuito	SMR	RR
Los Ceibos	1.98	1.83
Chongon	1.36	1.56
Puerto Hondo	1.22	1.65

Tabla 3.13.
Valores de SMR y RR en los circuitos del Distrito Esteros

Circuito	SMR	RR
Trinitaria Sur	2.91	2.64
Trinitaria Norte	2.13	2.09
Esteros	2.10	1.98
Malvinas Norte	1.81	1.83
Malvinas Sur	1.69	1.83

Tabla 3.14.
Valores de SMR y RR en los circuitos del Distrito 9 de Octubre

Circuito	SMR	RR
Antepara	2.00	1.92
Venezuela	1.99	1.82
Garay	1.82	1.64
Cristo del Consuelo	1.69	1.75
9 de Octubre	1.20	1.21
Victoria	1.00	1.11
Chile	0.99	1.13

Tabla 3.15.
Valores de SMR y RR en los circuitos del Distrito Modelo

Circuito	SMR	RR
Alborada	1.41	1.36
Sauces	1.36	1.32
Kennedy	1.16	1.18
Atarazana	0.96	1.01
Guayacanes	0.92	1.10
Urdesa	0.86	0.98

En las tablas 3.16 y 3.17 se muestran los circuitos que más destacan según los índices SMR y RR. Entre los que presentan el mayor factor contaminante, se encuentran Monte Sinaí, Flor de Bastión, Posorja, Trinitaria Sur y Batallón del Suburbio. De estos, Monte Sinaí y Flor de Bastión destacan por contar con un índice particularmente alto, lo cual es alarmante. Por otro lado, los circuitos con los índices más bajos, incluyen Bastión Popular Sur y Norte, Pascuales, Urdesa y Atarazana. En estos casos, todos presentan un factor protector según el SMR, y aunque Atarazana no tiene un factor protector significativo, su índice RR de 1.01 indica que no posee un factor contaminante relevante.

Tabla 3.16.
Circuitos con RR más alto

Distrito	Circuito	SMR	RR
Nueva Prosperina	Monte Sinai	3.52	3.37
Nueva Prosperina	Flor de Bastion	3.52	3.21
Progreso	Posorja	2.71	2.58
Esteros	Trinitaria Sur	2.91	2.64
Portete	Batallon del Suburbio	2.71	2.47

Tabla 3.17.
Circuitos RR más bajo

Distrito	Circuito	SMR	RR
Pascuales	Bastion Popular Sur	0.64	0.75
Pascuales	Bastion Popular Norte	0.74	0.88
Pascuales	Pascuales	0.83	0.95
Modelo	Urdesa	0.86	0.98
Modelo	Atarazana	0.96	1.01

En el análisis geoespacial posterior se muestran los puntos calientes y fríos, entre los cuales algunos de los expuestos en la tabla estarán dentro del análisis.

3.1.2 Análisis multivariante

Con las variables de tasas de 2021 a 2024, SMR y RR, se realizó un análisis mediante Análisis de Componentes Principales (PCA), Escalamiento Multidimensional (MDS) y t-Incrustación de Vecinos Estocásticos (t-SNE). Estas técnicas fueron implementadas en GeoDa,

obteniendo el siguiente comportamiento.

PCA

Se realizó el PCA bajo el método de Descomposición de Valores Singulares (SVD), y la transformación estandarizada Z . (Véase el Test A.3).

Al realizar el análisis por medio del método Eigen, los resultados son iguales al método anterior, por lo tanto los resultados serán trabajados por medio del método SVD.

Teniendo en cuenta el criterio de Kaiser, en este método se consideraron dos componentes principales guardadas en el programa como $PC1_{svd}$ y $PC2_{svd}$.

MDS

Se elaboró un análisis de Escalamiento Multidimensional, utilizando las mismas 6 variables, con un número de dimensiones igual a 2. Se empleó el método `classic metric` y la transformación estandarizada (Z). El resultado de este análisis se muestra en el diagrama de dispersión de la figura 3.7.

En este caso, no se cuenta con un test de MDS, pero es posible compararlo con los resultados de dispersión del PCA, con lo que se observa una gran similitud en los resultados.

Figura 3.7.
Dispersión de las dos componentes por medio de MDS

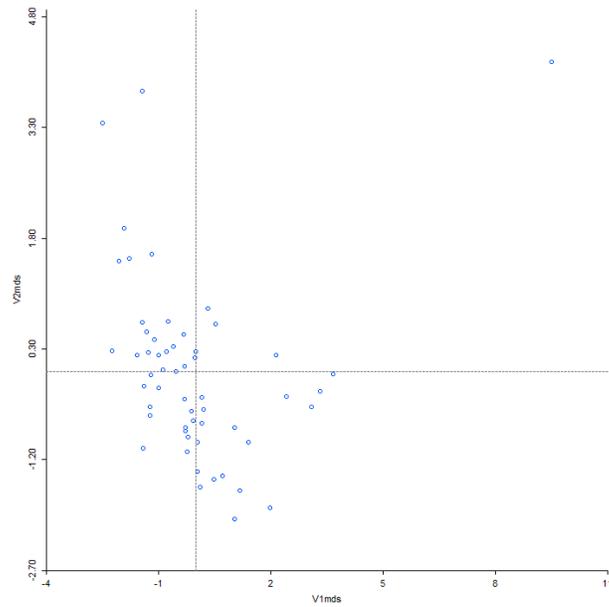
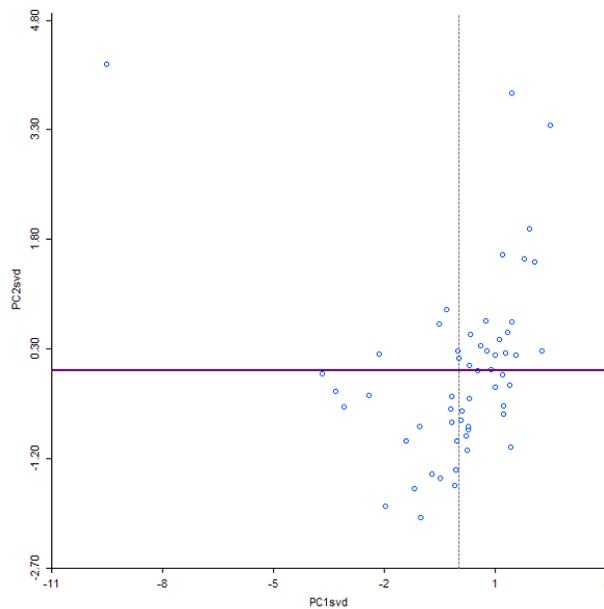


Figura 3.8.
Dispersión de las dos componentes por medio de PCA



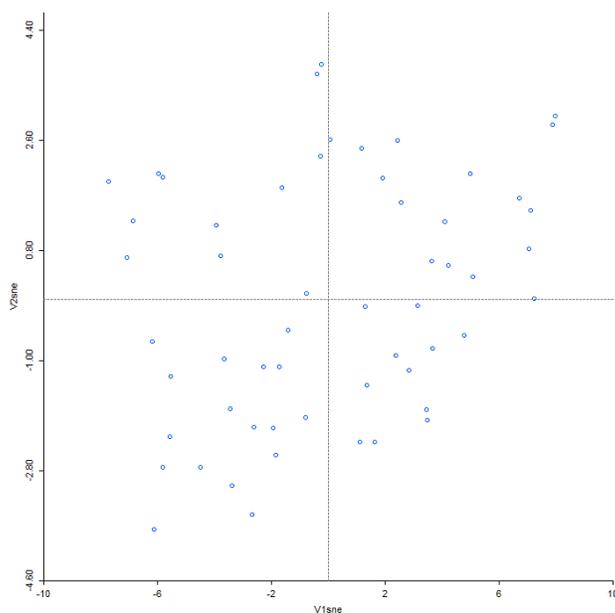
Debido a la naturaleza del programa utilizado, los valores del MDS muestran un

comportamiento similar, pero con valores negativos en la variable, es decir, se cumple que $PC1_{svd} = -V1_{mfs}$. Por lo tanto, se concluye que ambos métodos presentan el mismo comportamiento, y, en consecuencia, se opta por utilizar el PCA en lugar del MDS.

t-SNE

El t-SNE es una herramienta más sofisticada que emplea métodos diferentes, pero al aplicar este método con las variables disponibles, se obtuvo el siguiente resultado.

Figura 3.9.
Dispersión de las dos componentes por medio de t-SNE



Este método muestra resultados muy diferentes, los cuales no tienen una similitud a los resultados anteriores vistos en PCA y MDS. Luego, al realizar el análisis geoespacial, se observa que dicho método muestra resultados poco realistas, alejados de los encontrados por los método tradicional. Se sugiere que este problema es causado por la correlación de las variables y eso puede llevar al método a obtener resultados fuera del análisis que se espera realizar.

3.2 Análisis geoespacial

Para el análisis Geoespacial, se trabajó por medio de dos métodos, el método Getis-Ord G_i^* y el Local Moran I. A continuación en las figuras 3.10a, 3.10b, 3.10c y 3.10d se presentan mapas que revelan la evolución de las extorsiones en el Cantón Guayaquil de manera espacial.

Figura 3.10.

Evolución temporal de los mapas de llamadas de auxilio, distribuidos por distritos y coordenadas geográficas.

(a)

Mapa de llamadas de auxilio en 2021.



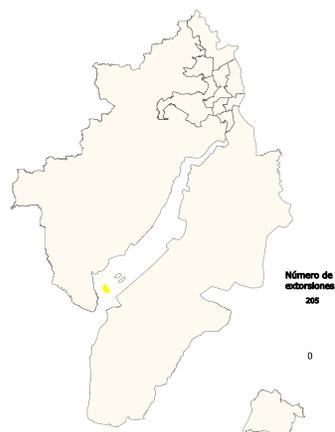
(b)

Mapa de llamadas de auxilio en 2022.



(c)

Mapa de llamadas de auxilio en 2023.



(d)

Mapa de llamadas de auxilio en 2024.



Dado que este delito se considera, en el contexto de este trabajo, un problema de salud pública, es importante reconocer que existe un patrón de contaminación que comienza en la cabecera cantonal y se extiende hacia las áreas rurales del cantón. Con base en los análisis previos, se observa un patrón de relación de este delito con la delincuencia organizada. Comprender estos patrones es fundamental para entender la estructura y organización de los grupos delictivos involucrados.

3.2.1 *Getis-Ord Gi**

Los resultados obtenidos, desde la parte descriptiva, espacial y multivariante se utilizaron para identificar patrones comunes, debido a esto se utilizó el RR, SMR, PC1svd y PC2svd. Se aplicó el *Getis-Ord Gi** mediante GeoDa.

Para aplicar los métodos, es necesario definir los pesos espaciales. En GeoDa, existen dos formas de definirlos: la opción "Torre" y la opción "Reina".

A continuación se presentan las distintas formas de los pesos espaciales en GeoDa.

En la figura 3.11 se muestra la contigüidad de la torre en sus distintos órdenes.

Figura 3.11.*Contigüidad de la Torre en sus distintos órdenes***(a)***Contigüidad de la Torre orden 1*

1	2	3
4	5	6
7	8	9

(b)*Contigüidad de la Torre orden 2*

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

(c)*Contigüidad de la Torre orden 1 y 2**incluido*

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

La Contigüidad de la Torre toma como vecinos aquellos elementos cercanos en el norte, sur, este y oeste (Véase figura 3.11a), no toma los elementos diagonales, aquellos que están en noroeste o noreste. Luego en Orden 2, considera como vecinos aquellos que son vecinos de sus vecinos (Véase figura 3.11b).

Para este tipo de análisis, al tratarse de un municipio, es necesario no considerar este tipo de peso espacial, así que se utilizó la contigüidad de la Reina, como se muestra en la figura 3.12.

Figura 3.12.
Contigüidad de la Reina en sus distintos órdenes.

(a)
Contigüidad de la Reina orden 1

1	2	3
4	5	6
7	8	9

(b)
Contigüidad de la Reina orden 2

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

(c)
Contigüidad de la Reina orden 1 y 2

incluido

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

En contigüidad de la Reina mostrada en la figura 3.12a se consideran vecinos a todos los elementos que están inmediatamente al lado de ellos (Véase figura 3.12a), en orden 2 (Véase figura 3.12b) se considera que los vecinos son los vecinos de sus vecinos.

Para el análisis de este trabajo se consideró pesos de la Contigüidad de la Reina de orden 1 y de orden 1 y 2 incluido (Véase figura 3.12c).

A continuación se presentan los resultados por medio de cada uno de los métodos, pesos y variables a analizar.

Para todos los casos se utilizó una permutación de 99999 y un $p < 0.05$.

Por SMR

Considerando contigüidad de la Reina y orden 1, se muestran los siguientes resultados.

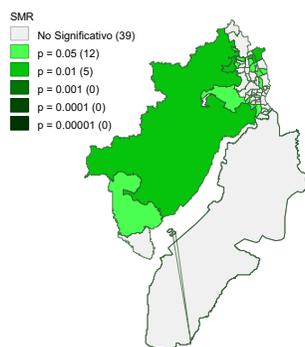
Figura 3.13.

Resultados de SMR con orden 1 con Getis-Ord G_i^*

(a)

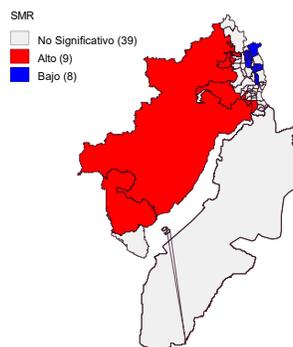
Mapa de significancia de SMR con orden 1 con

Getis-Ord G_i^*

**(b)**

Mapa de clusters de SMR con orden 1 con Getis-Ord

G_i^*



Los circuitos que se encuentran como puntos calientes (Véase figura 3.13) a través de esta metodología son El Morro, Progreso, Puerto Hondo, Monte Sinaí, Nuevo Guayaquil, Paraíso la Flor, Suburbio, Trinitaria Sur y Esteros. Como puntos fríos se identifican a Orquídeas, Mucho Lote, Bastion Popular Norte y Sur, Juan Montalvo, Alborada, Sauces y Kennedy. El SMR no es una medida confiable en este método, debido a que es altamente afectada por la población de cada circuito. Sin embargo, los circuitos identificados llegan a ser altamente conflictivos, lo que refleja la necesidad de una atención inmediata por parte de las autoridades.

Ahora se muestran los resultados con el orden 1 y 2 incluidos.

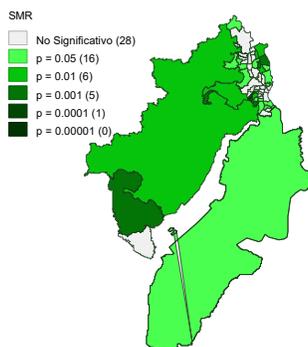
Figura 3.14.

Resultados de SMR con orden 1 y 2 incluido con Getis-Ord G_i^*

(a)

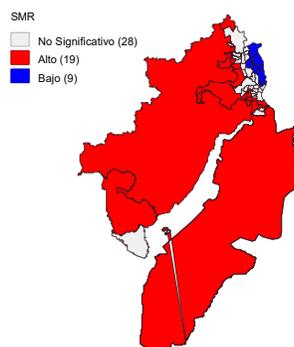
Mapa de significancia de SMR con orden 1 y 2

incluido con Getis-Ord G_i^*

**(b)**

Mapa de clusters de SMR con orden 1 y 2 incluido

con Getis-Ord G_i^*



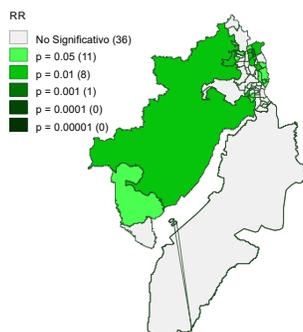
Considerando pesos espaciales de orden 2, incluyendo el orden 1 por medio del SMR, se observan 19 puntos calientes (Véase figura 3.15), que son los mismos observados en orden 1 excepto Paraíso la Flor, además resaltan nuevas zonas de riesgo como Chongon, Fortin, Puente Lucía, Los Ceibos, Suburbio, Chala, Cisne, Trinitaria Norte, Malvinas Sur, Floresta, 7 Lagos y Puna, contando en total 11 nuevos puntos calientes. Entre los puntos fríos se observan Orquideas, Mucho Lote, Juan Montalvo, Samanes, Guayacanes, Saucos, Alborada, Atarazana y 9 de Octubre, de los cuales 5 de estos circuitos coinciden con los mostrados en el orden 1. Se observa una tendencia delictiva sectorizada, que sugiere una ruta de este delito, aunque no todos los circuitos coincidan, a rasgos generales se evidencia un camino de contaminación en la extorsión.

Por RR

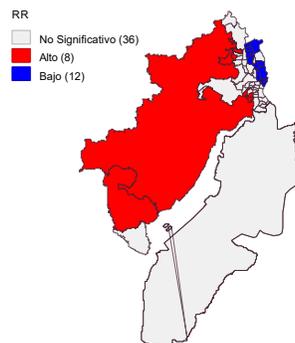
Considerando contigüidad de la Reina y orden 1, se muestran los siguientes resultados.

Figura 3.15.Resultados de RR con orden 1 con Getis-Ord G_i^* **(a)**

Mapa de significancia de RR con orden 1 con

Getis-Ord G_i^* **(b)**

Mapa de clusters de RR con orden 1 con Getis-Ord

 G_i^* 

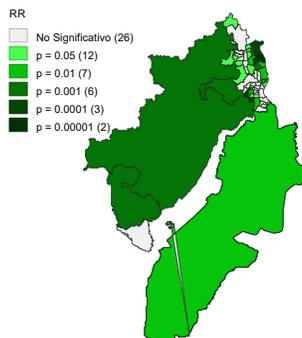
Ahora, considerando el RR con pesos espaciales de orden 1, se observa que el cantón Guayaquil cuenta con 8 puntos calientes (Véase la figura 3.15). Los puntos calientes que se identifican son El Morro, Progreso, Paraíso la Flor, Nuevo Guayaquil, Monte Sinaí, Suburbio, Trinitaria Sur y Norte. El comportamiento de los puntos calientes es muy similar al detectado con el SMR, en el RR no se está considerando como punto caliente Puerto Hondo ni Esteros, pero se agrega como punto caliente Trinitaria Norte. Esto deja ver que ambas medidas tienen un objetivo inicial, pero que el RR tiene un origen distinto. Entre los puntos fríos se identifican a Orquídeas, Mucho Lote, Pascuales, Bastion Popular Norte y Sur, Juan Montalvo, Saucés, Alborada, Kennedy, Atarazana, 9 de Octubre y Victoria. Se observa que todos los que fueron considerados puntos fríos según SMR de orden 1, ahora son considerados puntos fríos según RR de orden 1.

A continuación se muestran los resultados con el Orden 1 y 2 incluidos.

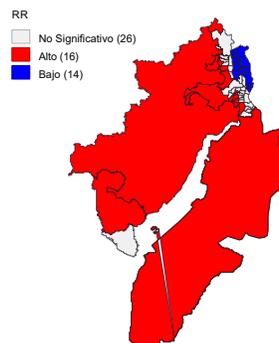
Figura 3.16.

Resultados de RR con orden 1 y 2 incluido con Getis-Ord G_i^*

(a)
Mapa de significancia de RR con orden 1 y 2
incluido con Getis-Ord G_i^*



(b)
Mapa de clusters de RR con orden 1 y 2 incluido
con Getis-Ord G_i^*



Se evidencian 16 circuitos como puntos calientes (Véase figura 3.16), entre los cuales destacan los circuitos de El Morro, Progreso, Puna, Puente Lucía, Monte Sinaí, Nuevo Guayaquil, Nueva Prosperina, Los Ceibos, Puerto Hondo, Chongon, Suburbio, Chala, Cisne, Trinitaria Norte y Sur y Esteros. Entre los puntos fríos se identifican a los circuitos Orquídeas, Mucho Lote, Pascuales, Bastion Popular Norte y Sur, Juan Montalvo, Samanes, Guayacanes, Sauces, Alborada, Martha de Roldos, Atarazana, 9 de Octubre y Victoria. De igual forma que en el SMR con orden 2, muchos de los puntos calientes y fríos son migrados considerando el RR como variable, manteniendo la sectorización que se viene mostrando en los mapas anteriores.

Al realizar el análisis y considerar que una de las medidas más confiable es el RR, ya que sus resultados son a partir de una transformación logit, y al comparar con los resultados obtenidos por medio del método del índice local de Moran, se continuó con el análisis del siguiente método para corroborar sus resultados con lo analizado en este método.

A diferencia del G_i^* , el índice Local Moran analiza los estratos por 4 especies,

considerándolos Alto-Alto a los puntos calientes y Bajo-Bajo los puntos fríos, contando también con categorizaciones de valores atípicos en las categorías de Bajo-Alto y Alto-Bajo. Por lo tanto, se continuó con el método para corroborar los resultados.

Además que el índice Local Moran proporciona cuatro clasificación de los estratos, también proporciona un índice de correlación espacial, el cual permite analizar el ajuste de correlacion de la variable que se está considerando para cada caso. Por lo tanto, se continuó con el análisis del índice local de Moran

3.2.2 Índice local de Moran

Se analizó el RR y ambas componentes principales llamadas PC1svd y PC2svd que fueron extraídas a partir del PCA. Se inició con el análisis de RR, para observar el comportamiento y compararlo con la metodología anterior.

Por RR

Considerando contigüidad de la Reina y orden 1, se muestran los siguientes resultados.

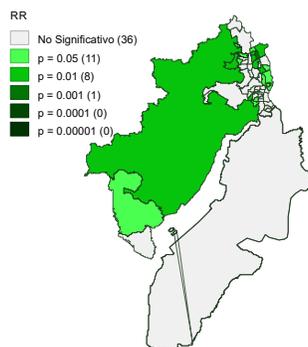
Figura 3.17.

Resultados de RR con orden 1 por I de Moran Local Univariante

(a)

Mapa de significancia de RR con orden 1 por I de

Moran Local Univariante

**(b)**

Mapa de clusters de RR con orden 1 por I de Moran

Local Univariante

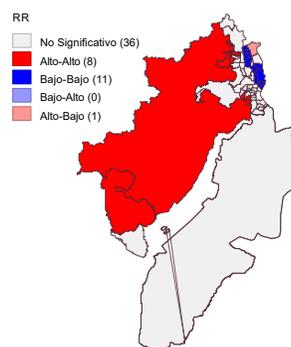
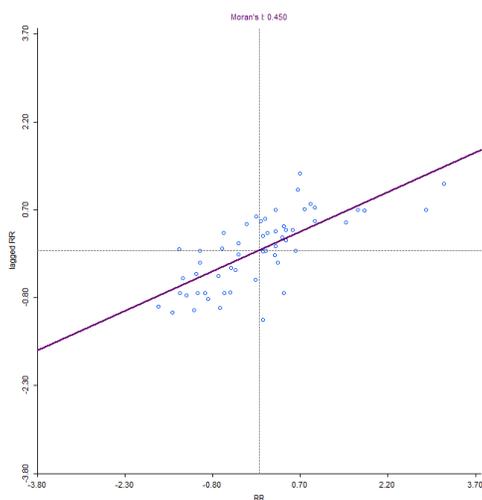
**Figura 3.18.**

Diagrama de dispersión con coeficiente de correlación de RR de orden 1 por I de Moran Local Univariante



Se observó que los clusters son iguales a la metodología anterior, pero en este caso se considera que el circuito Orquideas es considerado como atípico, en la metodología anterior era considerado como punto frío. Los puntos calientes y puntos fríos son exactamente iguales que la metodología anterior con RR como variable, tomando orden 1.

Para esta variable existe un índice de correlación espacial I de Moran de 0.45, siendo esta la correlación más alta entre todos los análisis realizados.

Veáse los resultados con el Orden 1 y 2 incluidos en la figura 3.19.

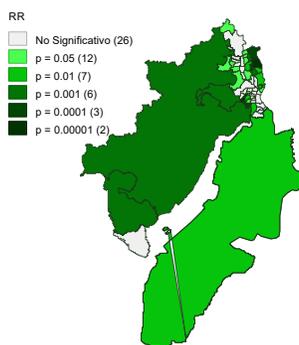
Figura 3.19.

Resultados de RR con orden 1 y 2 incluido por I de Moran Local Univariante

(a)

Mapa de significancia de RR con orden 1 y 2 por I

de Moran Local Univariante



(b)

Mapa de clusters de RR con orden 1 y 2 por I de

Moran Local Univariante

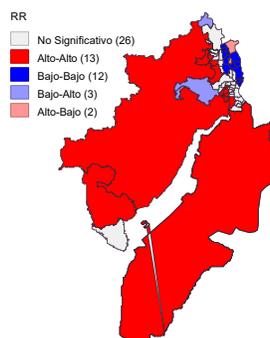
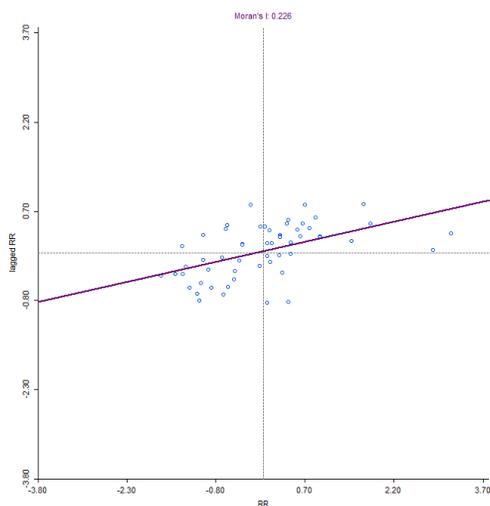


Figura 3.20.

Diagrama de dispersión con coeficiente de correlación de RR de orden 2 por I de Moran Local Univariante



De manera similar que en el orden 1, los clusters son iguales que en el RR con la Getis-Ord

G_i^* , pero ahora se considera Chongon, Puerto Hondo y Puente Lucía como atípicos de la categoría Bajo-Alto, siendo estos atípicos. Antes estos circuitos fueron considerados como puntos calientes. Dentro de los puntos fríos se encuentran los mismos que la metodología anterior, pero los circuitos Orquideas y Samanes ahora entran en la categoría Alto-Bajo, siendo estos atípicos.

El índice de correlación I de Moran es de 0.226, siendo esta correlación menos de la mitad que la correlación anterior.

Se observa en las figuras 3.17 y 3.19, los resultados son los mismos que con la metodología anterior, con diferencia que algunos de los circuitos que eran fríos o calientes en el Geits-Ord G_i^* , en la nueva metodología son considerados como atípicos de la categoría Alto-Bajo o Bajo-Alto.

Por lo tanto, se consideró utilizar para el análisis los resultados la metodología del índice local de Moran.

Por PC1svd

Considerando contigüidad de la Reina con orden 1, se obtuvieron los resultados observados en la figura 3.21. Esta variable proviene de un proceso de refinamiento multivariante, la cual contiene una varianza acumulada del 58.73% de la varianza total. Por lo que esta variable posee un comportamiento con mayor relevancia. Las variables que cuentan con mayor peso en esta variable son las tasas, teniendo mayor scores que el RR y el SMR. Por lo que se interpreta que esta variable provee mayor información de las tasas del 2021 al 2024.

Figura 3.21.

Resultados de PC1svd con orden 1 por I de Moran Local Univariante

(a)

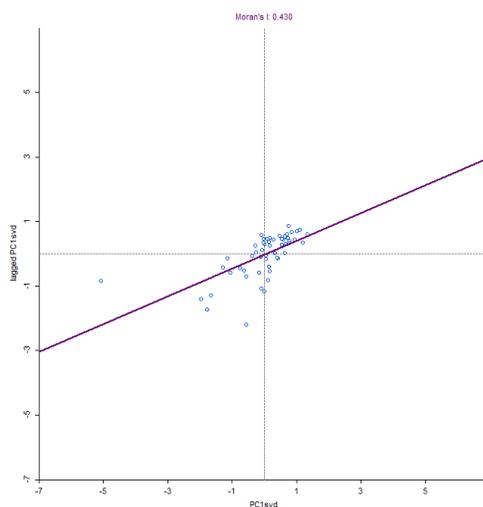
Mapa de significancia de PC1svd con orden 1 por I de Moran Local Univariante

(b)

Mapa de clusters de PC1svd con orden 1 por I de Moran Local Univariante

**Figura 3.22.**

Diagrama de dispersión con coeficiente de correlación de PC1svd de orden 1 por I de Moran Local Univariante



Se observa que en el mapa según esta variable tiene 9 puntos calientes, entre los cuales se muestran a El Morro, Progreso, Monte Sinaí, Suburbio, Cisne, Trinitaria Norte y Sur, Esteros y Malvinas Sur. Como puntos fríos se muestran a los circuitos Alborada, Atarazana, Urdesa,

Kennedy, 9 de Octubre, Victoria y Garay.

El índice de correlación de I de Moran para esta variable es de 0.43, siendo este el segundo índice de correlación más alto de todo el análisis de esta investigación.

A continuación se muestra con el Orden 1 y 2 incluidos en la figura 3.23.

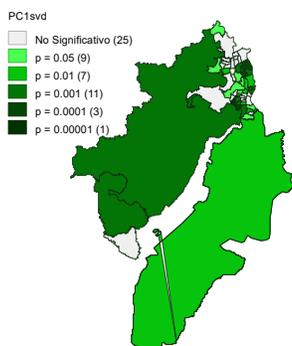
Figura 3.23.

Resultados de PC1svd con orden 1 y 2 incluido por I de Moran Local Univariante

(a)

Mapa de significancia de PC1svd con orden 1 y 2

por I de Moran Local Univariante



(b)

Mapa de clusters de PC1svd con orden 1 y 2 por I

de Moran Local Univariante

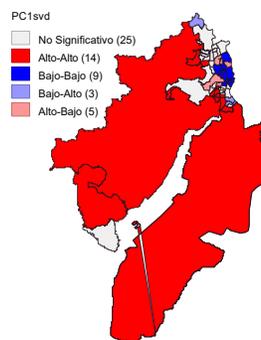
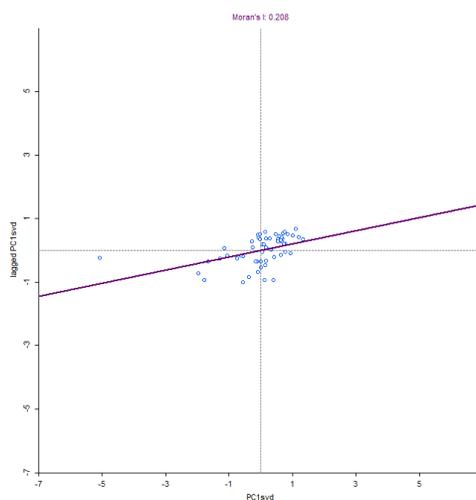


Figura 3.24.

Diagrama de dispersión con coeficiente de correlación de PC1svd de orden 1 y 2 incluido por I de Moran

Local Univariante



Los circuitos calificados como puntos calientes son El Morro, Progreso, Puna, Monte Sinaí, Nueva Prosperina, Chongon, Suburbio, Cisne, Chala, Trinitaria Norte y Sur, Malvinas Sur, Floresta y Guasmo. Entre los puntos fríos se observan a Sauces, Alborada, Martha de Roldos, Urdesa, Kennedy, Atarazana, 9 de Octubre, Victoria y Garay. Entre los puntos atípicos en la clasificación de Alto-Bajo destacan Sauces, Juan Montalvo, Samanes, Bellavista y Salinas. En la clasificación de Bajo-Alto se encuentran a Puente Lucía, 7 Lagos y Floresta.

Se entiende que los atípicos en la clasificación Alto-Bajo eran puntos fríos que empezaron a contaminarse y ahora ya no son fríos, entre los atípicos de la clasificación Bajo-Alto se interpreta que eran puntos calientes que empezaron a mejorar su situación y ya no son considerados como puntos calientes.

Se observa que el índice de correlación I de Moran es de 0.208.

Por PC2svd

Considerando contigüidad de la Reina y orden 1, se muestran los resultados observados en la figura 3.25.

Esta variable también proviene del PCA, la cual contiene un refinamiento adicional al RR o al SMR, teniendo una varianza acumulada del 25.98% de la varianza total. Por lo que esta variable posee un comportamiento más importante. Las variables que cuentan con mayor peso en esta variable son el RR y el SMR, siendo la variable que recoge mayor información del RR y el SMR.

Figura 3.25.

Resultados de PC2svd con orden 1 por I de Moran Local Univariante

(a) Mapa de significancia de PC2svd con orden 1 por I de Moran Local Univariante

(b) Mapa de clusters de PC2svd con orden 1 por I de Moran Local Univariante

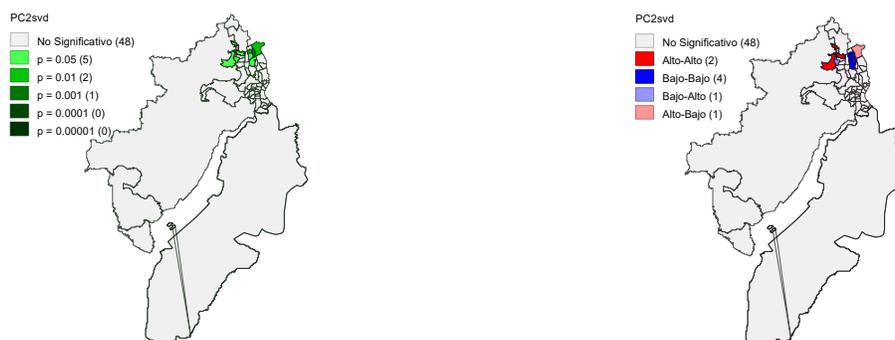
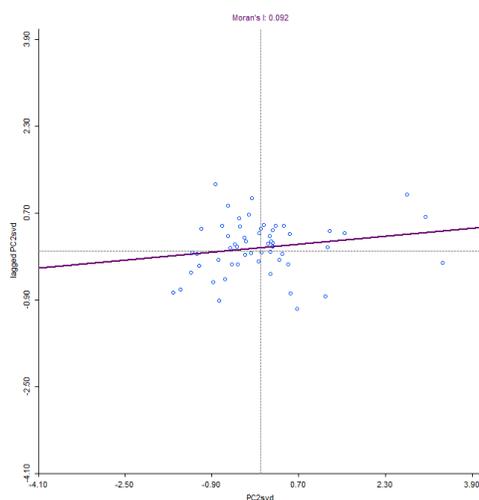
**Figura 3.26.**

Diagrama de dispersión con coeficiente de correlación de PC2svd de orden 1 por I de Moran Local Univariante



Con este orden, se identifican 2 puntos calientes que son los circuitos Paraíso la Flor y Monte Sinaí. Entre los puntos fríos se identifican los circuitos Mucho Lote, Bastión Popular Norte y Sur y Juan Montalvo. En los atípicos de categoría Alto-Bajo, se encuentran Nuevo Guayaquil y

en la categoría Bajo-Alto, a Orquideas.

En este caso se cuenta con un índice de correlación I de Moran es de 0.208.

Los resultados con el Orden 1 y 2 incluidos se observan en la figura 3.23.

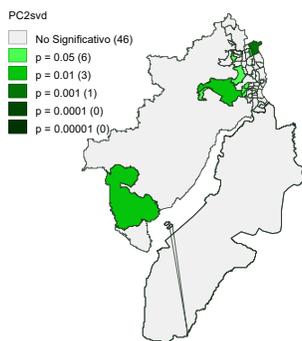
Figura 3.27.

Resultados de PC2svd con orden 1 y 2 incluido por I de Moran Local Univariante

(a)

Mapa de significancia de PC2svd con orden 1 y 2

por I de Moran Local Univariante



(b)

Mapa de clusters de PC2svd con orden 1 y 2 por I

de Moran Local Univariante

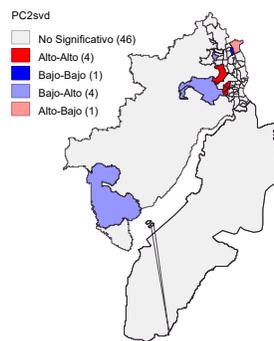
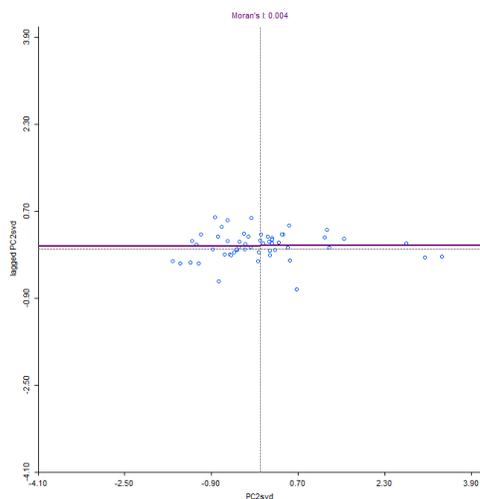


Figura 3.28.

Diagrama de dispersión con coeficiente de correlación de PC2svd de orden 1 y 2 incluido por I de Moran

Local Univariante



Para este caso se identifican 4 puntos calientes, Ceibos, Suburbio, Batallón del Suburbio y Portete. Se observa un solo punto frío, el circuito Mucho Lote. Entre los puntos atípicos de la categoría Alto-Bajo se encuentran el circuito Orquideas y en la categoría Bajo-Alto a los circuitos Puerto Hondo, Chongon, Nuevo Guayaquil y el Morro.

En este caso se cuenta con un índice de correlación espacial I de Moran de 0.004, siendo esta correlación la más baja de todas las anteriores.

CAPÍTULO 4

4. CONCLUSIONES Y RECOMENDACIONES

Esta investigación demuestra una forma efectiva de identificar puntos críticos de extorsión en Guayaquil, combinando análisis estadísticos y geoespaciales. A diferencia de otros trabajos, no solo se señalan las zonas más afectadas, sino que también se considera cómo el delito se propaga en el tiempo y el espacio. Aunque los resultados son sólidos, integrar factores socioeconómicos ayudaría a entender mejor las causas del problema. Esta metodología puede ser clave para mejorar la seguridad y optimizar recursos en la ciudad. A futuro, sería ideal incluir datos en tiempo real y aplicar este enfoque a otros delitos. A continuación, se presentan las principales conclusiones y recomendaciones.

Conclusiones

- El modelo geoespacial Índice Local de Moran univariante demostró ser efectivo para identificar puntos calientes de extorsión en Guayaquil, con altos índices de correlación espacial. Esto sugiere que la distribución de los delitos no es aleatoria, sino que se concentra en áreas específicas, lo que es crucial para la planificación de intervenciones en salud pública.
- Los resultados indican que los puntos calientes están relacionados con circuitos con bajos índices de pobreza, lo que sugiere que los grupos organizados operan en áreas con mayor

actividad económica. Este hallazgo es relevante para la salud pública, ya que la extorsión puede generar estrés y ansiedad en la población afectada, impactando negativamente en su bienestar.

- El análisis de Riesgo Relativo (RR) y la Razón Estandarizada de Mortalidad (SMR) revelaron que la propagación de la extorsión sigue patrones similares a los de una enfermedad contagiosa, lo que refuerza la necesidad de abordar este problema desde un enfoque epidemiológico.

Recomendaciones

- Se recomienda incorporar variables como el índice de pobreza, desempleo y densidad poblacional en futuros análisis para comprender mejor los factores que contribuyen a la extorsión. Esto permitiría diseñar intervenciones más efectivas desde un enfoque de salud pública.
- Extender el análisis a nivel de subcircuitos policiales permitiría una comprensión más detallada del comportamiento del delito dentro de cada circuito, lo que es esencial para la implementación de estrategias de prevención y control.
- Con la finalidad de extender este estudio a nivel Provincial, Regional o a todo el País, se recomienda medir el costo computacional de los modelos presentados, para así conocer si es estable para conjuntos de datos muy grandes.

BIBLIOGRAFÍA

Aguilar, E., Manzano, M., and Ramírez, J. (2014). Integración de fuentes de datos en el análisis delictivo urbano. *Ciudad y Territorio Estudios Territoriales*, pages 79–102.

Anderson, D., Sweeney, D., and Williams, T. (2016). *Estadística para administración y economía (13.a ed.)*. Cengage Learning.

Asamblea Nacional del Ecuador (2014). *Código Orgánico Integral Penal (COIP)*.

Bertelli, M. (2023). Aljazeera. Accedido: 8 de noviembre de 2024.

Bourguignon, F. (2009). *Crime as a Social Cost of Poverty and Inequality: A Review Focusing on Developing countries*. Revista Desarrollo y Sociedad, Universidad de los Andes, Facultad de Economía, CEDE.

Briceño-León, R., Villaveces, A., and Concha-Eastman, A. (2008). Understanding the uneven distribution of the incidence of homicide in latin america. *International Journal of Epidemiology*, 37:751–757.

Cohen, L. and Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, pages 588–608.

Echeverri, C., Zuluaga, L., and Maya, N. (2018). *La extorsión en Medellín como fenómeno del orden social, poder político y control territorial*. Alcaldía de Medellín, Secretaría de Seguridad y Convivencia y Observatorio de Seguridad Humana de Medellín, Medellín, Colombia.

- Eck, J., Chainey, S., Cameron, J., Leitner, M., and Wilson, R. (2005). *Mapping crime: Understanding hot spots*. National Institute of Justice.
- Gaviria, A. and Pagés, C. (2002). Patterns of crime victimization in latin american cities. *Journal of Development Economics*, 67:181–203.
- Getis, A. and Ord, J. (1992). *The analysis of spatial association by use of distance statistics*.
- Hair, J., Black, W., Babin, B., and Anderson, R. (2014). *Multivariate Data Analysis (7.a ed.)*. Pearson.
- He, Z., Lai, R., Wang, Z., Liu, H., and Deng, M. (2022). Comparative study of approaches for detecting crime hotspots with considering concentration and shape characteristics. *International journal of environmental research and public health*.
- Jana, M. and Sar, N. (2016). Modeling of hotspot detection using cluster outlier analysis and getis-ord g_i^* statistic of educational development in upper-primary level, india. *Modeling Earth Systems and Environment*, 2:1–10.
- Kalinic, M. and Krisp, J. (2018). Kernel density estimation (kde) vs. hot-spot analysis—detecting criminal hot spots in the city of san francisco. In *AGILE 2018 Conference Papers*, Lund, Sweden.
- Kim, M., Yeen, H., and Koo, M. (2024). Towards context-based violence detection: A korean crime dialogue dataset. In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 603–623.
- Mendivelso, F. and Rodríguez, M. (2019). Riesgo relativo. *Revista Médica Sanitas*, 22(2):72–75.

Observatorio Ecuatoriano de Crimen Organizado (2023). Radiografía de la extorsión: Tipologías y resultado de la encuesta nacional de victimización de casos de extorsión empresarial. Último acceso: 28 de noviembre de 2024.

Ratcliffe, J. (2010). *Crime Mapping: Spatial and Temporal Challenges*, pages 5–24. Springer New York.

Runadi, T. and Widyaningsih, Y. (2017). Application of hotspot detection using spatial scan statistic: Study of criminality in indonesia. In *AIP Conference Proceedings*, volume 1827.

Sampson, R. (1993). *The community context of violent crime*. Sociology and the Public Agenda.

Shaw, C. and McKay, H. (1942). Juvenile delinquency and urban areas. *University of Chicago Press*.

Solis, L. G. and Rojas, F. (2008). *Crimen organizado en américa latina y el caribe*. Catalonia.

Tango, T. and Takahashi, K. (2005). A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics*, 4:1–15.

Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11).

Velez, M. S. (2024). Análisis teórico de la evolución del delito de extorsión como fenómeno criminal en el ecuador. *Polo del Conocimiento*, pages 97–111.

Vidanapathirana, N. (2021). A comparison of spatial clustering assessment methods.

Walsh, J., Heazlewood, I., DeBeliso, M., and Climstein, M. (2020). Application of t-distributed stochastic neighbor embedding (t-sne) to clustering of social affiliation and recognition

psychological motivations in masters athletes. *International Journal of Sport, Exercise and Health Research*, 4(1):1–6.

Zúñiga, I. (2018). Epidemiología forense: una disciplina al servicio de la justicia. *Instituto Mexicano del Seguro Social*.

APÉNDICES

APÉNDICE A

```
1 # Calcular el valor esperado (V) para cada distrito en 2021-2023
2 VE_list <- lapply(distritos, function(distrito) {
3   V_distrito <- historico %>%
4     filter(Distrito == distrito, Año %in% c(2021, 2022, 2023))
5   if(nrow(V_distrito) == 0) {
6     return(data.frame(Distrito = distrito, V = NA))
7   }
8   V <- expected(V_distrito$Poblacion,
9     V_distrito$Delitos, nrow(V_distrito))
10  data.frame(Distrito = distrito, V = V*270/366)
11 })
12
13 VE_df <- do.call(rbind, VE_list)
14 valores$VE_df <- VE_df
15
16 # Observados en 2024 por distrito
17 P_2024 <- delitos_por_distrito_ano %>%
18   filter(Año == 2024) %>%
19   left_join(poblacion_distrito_ano, by=c("Distrito", "Año")) %>%
20   rename(Observed = Delitos, poblacion = Poblacion)
21
```

```

22 # Calcular SMR = Observed_2024 / V
23 SMR_df <- VE_df %>%
24   left_join(P_2024, by = "Distrito") %>%
25   mutate(SMR = Observed / V)
26
27 # Unir SMR a tabla_tasa_base
28 tabla_tasa <- tabla_tasa_base %>%
29   left_join(SMR_df %>% select(Distrito, SMR,
30     Observed, poblacion), by = "Distrito")
31
32 tabla_tasa$SMR<-round(tabla_tasa$SMR,2)
33
34 valores$tabla_tasa <- tabla_tasa

```

Código A.1. código en R para el cálculo del valor esperado y el SMR

```

1 # Shapefile de distritos
2 setwd("C:\\Tesis_graduación\\Distritos_Policia_Validos")
3 shp_my <- my_shp <- st_read("Guayaquil0901.shp", promote_to_multi = TRUE)
4 st_crs(my_shp)
5 my_shp <- my_shp %>% st_transform(4326)
6 my_shp <- sf::as_Spatial(my_shp$geometry)
7 my_shp@polygons[[1]]@ID <- "SUR"
8 my_shp@polygons[[2]]@ID <- "ESTEROS"
9 my_shp@polygons[[3]]@ID <- "9 DE OCTUBRE"
10 my_shp@polygons[[4]]@ID <- "PORTETE"
11 my_shp@polygons[[5]]@ID <- "MODELO"
12 my_shp@polygons[[6]]@ID <- "FLORIDA"
13 my_shp@polygons[[7]]@ID <- "PASCUALES"
14 my_shp@polygons[[8]]@ID <- "NUEVA PROSPERINA"
15 my_shp@polygons[[9]]@ID <- "CEIBOS"
16 my_shp@polygons[[10]]@ID <- "PROGRESO"
17 distritos_shp <- sapply(my_shp@polygons, function(x) x@ID)
18
19 # Datos totales por distrito (para etiqueta)
20 data <- datos %>% group_by(Distrito) %>% count()
21 tabla_tasa <- tabla_tasa %>%
22   left_join(data %>% select(Distrito, n), by="Distrito")
23
24 idx <- match(distritos_shp, tabla_tasa$Distrito)
25 tabla_tasa_reordenada <- tabla_tasa[idx, ]

```

```

26 tabla_tasa_reordenada <- as.data.frame(tabla_tasa_reordenada)
27 rownames(tabla_tasa_reordenada) <- distritos_shp
28
29 # Crear SpatialPolygonsDataFrame para INLA
30 mapa_exto <- SpatialPolygonsDataFrame(my_shp,
31                                     tabla_tasa_reordenada,
32                                     match.ID = TRUE)
33 nb <- poly2nb(shp_my)
34 nb2INLA("map.adj", nb)
35 g <- inla.read.graph(filename = "map.adj")
36 mapa_exto$re_u <- 1:nrow(mapa_exto@data)
37 mapa_exto$re_v <- 1:nrow(mapa_exto@data)
38
39 # Agregar V (esperado) al dataset
40 mapa_exto@data <- mapa_exto@data %>%
41   left_join(VE_df %>% select(Distrito, V), by = "Distrito")
42
43 formula <- round(mapa_exto@data$Observed, 0) ~
44   mapa_exto@data$"2021" + mapa_exto@data$"2022" + mapa_exto@data$"2023" +
45   f(re_u, model = "besag", graph = g, scale.model = TRUE) +
46   f(re_v, model = "iid")
47
48 res <- inla(formula, family = "poisson", data = mapa_exto@data,
49             E = mapa_exto@data$V,
50             control.predictor = list(compute = TRUE))

```

```
51
52 mapa_exto$RR <- res$summary.fitted.values[, "mean"]
53 mapa_exto$LL <- res$summary.fitted.values[, "0.025quant"]
54 mapa_exto$UL <- res$summary.fitted.values[, "0.975quant"]
55
56 tabla_tasa <- tabla_tasa %>%
57   left_join(mapa_exto@data %>% select(Distrito, RR, LL, UL),
58             by = "Distrito")
59
60 tabla_tasa$RR <- round(tabla_tasa$RR, 2)
61
62 valores$tabla_tasa <- tabla_tasa
```

Código A.2. código en R para el cálculo del RR

```
1 PCA method: svd
2
3 Standard deviation:
4 1,877208 1,248563 0,727879 0,553094 0,247511 0,142143
5
6 Proportion of variance:
7 0,587318 0,259818 0,088301 0,050985 0,010210 0,003367
8
9 Cumulative proportion:
10 0,587318 0,847136 0,935437 0,986422 0,996633 1,000000
11
12 Kaiser criterion: 2,000000
13
14 95% threshold criterion: 3,000000
15
16 Eigenvalues:
17 3.52391
18 1.55891
19 0.529807
20 0.305913
21 0.0612617
22 0.0202046
23
24 Variable Loadings:
25
```

	PC1	PC2	PC3	PC4	PC5	PC6
--	-----	-----	-----	-----	-----	-----

26	RR	0.35518	0.575062	-0.0424226	0.17944	0.677281
		0.224594				
27	SMR	0.33422	0.606669	-0.00637974	0.0566101	-0.718074
		0.0370831				
28	Tasa2021	-0.367505	0.32259	0.763438	-0.415302	0.0639485
		0.0383763				
29	Tasa2022	-0.458919	0.0868744	0.208587	0.853037	-0.0709296
		0.0750664				
30	Tasa2023	-0.484465	0.154968	-0.453622	-0.232307	-0.0732118
		0.690057				
31	Tasa2024	-0.426346	0.406977	-0.407481	-0.102383	0.105736
		-0.681833				

32

33

34 Squared correlations:

35		PC1	PC2	PC3	PC4	PC5	PC6
36	RR	0,44455	0,51552	0,0009534	0,009850	0,0281018	0,00101914
37	SMR	0,39363	0,57375	2,1564e-05	0,000980	0,0315878	2,77811e-05
38	Tasa2021	0,47593	0,16222	0,308791	0,052762	0,0002505	2,97716e-05
39	Tasa2022	0,74215	0,01176	0,0230511	0,222604	0,0003082	0,00011385
40	Tasa2023	0,82708	0,03743	0,10902	0,016509	0,0003283	0,00962075
41	Tasa2024	0,64054	0,25820	0,0879699	0,003206	0,0006849	0,00939329

Código A.3. Resultados del PCA