

Instituto Tecnológico y de Estudios Superiores de Occidente

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial
15018, publicado en el Diario Oficial de la Federación del 29 de noviembre de 1976.

Departamento de Matemáticas y Física
Maestría en Ciencia de Datos



Educación y Prevención del Delito: Un Análisis de la Influencia de los Centros Educativos en la Incidencia Delictiva de Jalisco

TRABAJO RECEPCIONAL que para obtener el GRADO de
Maestro en Ciencia de Datos

Presenta:
Oscar Zambrano Aguirre

Director:
Mtro. BYRON MICHAEL MOTTA BONILLA

Tlaquepaque, Jalisco, 27 de mayo de 2024

Educación y Prevención del Delito: Un Análisis de la Influencia de los Centros Educativos en la Incidencia Delictiva de Jalisco

Oscar Zambrano Aguirre

Resumen

En el presente trabajo se aborda el problema de la incidencia delictiva en el estado de Jalisco, México, considerando la educación como un poderoso instrumento de cambio. La investigación se centra en establecer una correlación entre el número y tipos de escuelas a nivel municipal y la incidencia y tipo de delitos en esos lugares.

El objetivo central es discernir las mejores estrategias para la creación de instituciones educativas que puedan contribuir a una disminución en los índices de violencia.

Inicialmente, se realiza un análisis exploratorio de datos para comprender la distribución actual de delitos y centros educativos. Posteriormente, se presentan modelos estadísticos que expliquen la relación entre las variables. Los resultados del estudio podrían proporcionar observaciones sobre qué municipios podrían beneficiarse más de intervenciones educativas.

Finalmente, se enfatiza la urgente necesidad de invertir en educación como mecanismo de prevención del delito y promoción de una sociedad más justa y segura.

Esta investigación proporciona un enfoque que invita a la reflexión al vincular directamente la educación con la prevención del delito, enfatizando la relevancia de políticas educativas estratégicas para el bienestar social.

Tabla de Contenidos

	Página
1 Introducción	15
1.1. Contexto	15
1.2. Justificación	19
1.3. Problema	20
1.4. Objetivos	20
1.4.1. Objetivo general	20
1.4.2. Objetivos específicos	20
2 Metodología	21
2.1. Descripción de los datos	21
2.2. Instituto de Información Estadística y Geográfica del Estado de Jalisco - IIEG	21
2.2.1. Índices de criminalidad del estado de Jalisco	22
2.3. INEGI	27
2.3.1. Geografía y Demografía - INEGI	27
2.4. Datos Jalisco	31
2.4.1. Centros educativos Activos Jalisco	32
2.5. Clasificador de delitos en Jalisco por nivel de violencia	34
2.5.1. Escala de medición	34
2.6. Análisis exploratorio	36
2.6.1. Construcción de DataSet unificado	36
2.6.2. Análisis Descriptivo Inicial	39
2.6.3. Distribución de las Variables	42
2.6.4. Análisis PCA	51
2.7. Descripción de los modelos	56
2.8. Descripción de las métricas	60
2.9. Descripción de los experimentos o simulaciones	63
3 Resultados y discusión.	71
3.1. Resultados y discusión	71
4 Conclusiones y trabajo futuro.	79
4.1. Conclusiones	79
4.2. Trabajo futuro	80
4.3. Trabajo relacionados	80

Índice de figuras

Página

1.1.	Esta gráfica ilustra la distribución de matrículas en diferentes niveles educativos. La mayoría de los estudiantes se encuentran en primaria (40.9%), seguido de secundaria (18.3%) y preescolar (14%). Los niveles de educación superior y medio superior representan un 25.8% combinado. La educación inicial y especial, aunque esenciales, abarcan sólo el 1.9% del total, reflejando áreas potencialmente menos atendidas en el sistema educativo.	16
2.1.	La serie de gráficos muestra la tendencia en la incidencia de diversos tipos de delitos en un periodo que abarca varios años(Diciembre 2017 a febrero 2024). Cada gráfico detalla un tipo específico de delito, desde robos en diferentes contextos hasta delitos más graves como homicidios y feminicidios.	23
2.2.	Esta serie de gráficos muestra la evolución mensual de delitos como violencia familiar, abuso sexual infantil, lesiones dolosas, violación y feminicidio desde 2017 hasta 2024. Las tendencias destacan un aumento en la mayoría de los delitos a partir de 2020, reflejando posibles cambios en los patrones de comportamiento social y efectividad de las políticas de prevención. Estos datos son cruciales para identificar periodos críticos y podrían orientar las estrategias de intervención y asignación de recursos hacia las áreas más afectadas.	25
2.3.	Estos gráficos muestran la tendencia mensual en los robos de autopartes y motocicletas desde 2017 hasta 2024. Ambas categorías presentan una considerable volatilidad en sus cifras, con picos que sugieren episodios de alta incidencia. La variabilidad en el robo de autopartes y motocicletas refleja los desafíos en la gestión y prevención de este tipo específico de crimen, lo cual podría ser indicativo de la necesidad de estrategias dirigidas y mejoradas para combatir eficazmente estos delitos.	26
2.4.	Estos gráficos representan la distribución de la densidad de población por km ² , destacando una concentración predominante en rangos bajos con algunos valores atípicos significativos que indican áreas de mayor densidad.	42

2.5. El histograma y el boxplot transformados proporcionan una perspectiva clara de la centralidad y dispersión de los datos, facilitando la interpretación estadística y la planificación basada en la densidad poblacional.	42
2.6. distribución y outliers, de la variable "Mayores de 15 años sin educación por población total	43
2.7. distribución después del preprocesamiento	43
2.8. Distribución y outliers, de las variables Para las variables: Sin acceso a servicios de salud, población mayor de edad, Mayores de edad con educación básica	44
2.9. Post-procesamiento. distribución y outliers, de las variables Para las variables: Sin acceso a servicios de salud, población mayor de edad y Mayores de edad con educación básica	44
2.10. Estos histogramas representan la distribución de docentes, alumnos y escuelas por población, destacando una concentración significativa en valores bajos para cada variable. Esta visualización refleja que, en su mayoría, las áreas tienen una baja densidad de docentes, alumnos y escuelas en relación a su población.	45
2.11. Post-procesamiento, las variables son mencionadas en cada gráfica. .	45
2.12. Estos histogramas muestran la frecuencia de turnos continuos, discontinuos y matutinos por población, evidenciando una predominancia de turnos matutinos en comparación con otros horarios.	46
2.13. Post-procesamiento, las variables son mencionadas en cada gráfica. .	46
2.14. Estos histogramas muestran la distribución de turnos nocturnos, vespertinos y escuelas privadas por población. Los datos indican que los turnos nocturnos y las escuelas privadas son poco comunes, mientras que los turnos vespertinos, aunque algo más frecuentes, siguen siendo limitados. Estos patrones pueden informar políticas educativas y la asignación de recursos.	46
2.15. Post-procesamiento, las variables son mencionadas en cada gráfica. .	46
2.16. Estos histogramas visualizan la distribución por población de instituciones públicas, educación básica y educación especial. La mayoría de las áreas tienen una baja proporción de instituciones públicas y educación básica por población, mientras que la educación especial es aún menos común, con una distribución que indica una disponibilidad muy limitada en la mayoría de las áreas.	47
2.17. Post-procesamiento, las variables son mencionadas en cada gráfica. .	47
2.18. Estos histogramas muestran la distribución per cápita de instituciones de educación inicial, media superior y superior. La mayoría de las áreas tienen una baja proporción de estas instituciones en relación con su población, destacando una disponibilidad particularmente limitada para la educación superior. Esto subraya la necesidad de ampliar el acceso a niveles educativos más avanzados en muchas regiones. . . .	47
2.19. Post-procesamiento, las variables son mencionadas en cada gráfica. .	47

2.20. Estos histogramas representan la frecuencia de abuso sexual infantil, feminicidios y homicidios dolosos por población. Todos los gráficos muestran una concentración alta de casos cerca del cero, indicando que, aunque los incidentes son relativamente raros en términos per cápita, su presencia es preocupante y varía ampliamente entre diferentes áreas. Esto subraya la importancia de políticas enfocadas en la prevención y la justicia para estos delitos graves.	48
2.21. Post-procesamiento, las variables son mencionadas en cada gráfica.	48
2.22. Estos histogramas ilustran la distribución per cápita de lesiones dolosas, robos a bancos y robos a carga pesada. Los datos muestran una concentración predominante de frecuencias cerca de cero para todos los delitos, indicando que estos eventos son infrecuentes en relación a la población total. Sin embargo, la presencia de estos delitos, aunque baja, subraya la necesidad de medidas de seguridad y prevención específicas para cada tipo de crimen.	48
2.23. Post-procesamiento, las variables son mencionadas en cada gráfica.	48
2.24. Estos histogramas representan la frecuencia de robos a cuentahabientes, robos a vehículos y robos a negocios por población. Todos los gráficos muestran que la mayoría de las áreas tienen una baja incidencia de estos delitos, con una concentración predominante de casos cercana a cero. Esta visualización es fundamental para identificar regiones con necesidades específicas de seguridad y para la planificación de recursos policiales y preventivos.	49
2.25. Post-procesamiento, las variables son mencionadas en cada gráfica.	49
2.26. Estos histogramas visualizan la distribución de robos a persona, robos a vehículos particulares y robos en casas habitación por población. En cada caso, la concentración de eventos es predominantemente baja, con la mayoría de las áreas mostrando valores cercanos a cero, lo que indica una baja frecuencia de estos delitos en relación a la población total. Este análisis subraya la variabilidad regional en la incidencia de robos y puede guiar las estrategias de prevención y seguridad local.	49
2.27. Post-procesamiento, las variables son mencionadas en cada gráfica.	49
2.28. Estos histogramas muestran la distribución de robos de autopartes, robos de motocicletas y violaciones por población. Cada histograma refleja una baja frecuencia general de estos delitos en la mayoría de las áreas, con la mayor parte de los datos agrupados cerca del cero. Esto indica que, aunque los incidentes ocurren, su prevalencia por población es relativamente baja en la mayoría de las regiones estudiadas.	50
2.29. Post-procesamiento, las variables son mencionadas en cada gráfica.	50
2.30. Este histograma muestra la distribución de casos de violencia familiar por población. La gran mayoría de los datos están concentrados cerca de cero, indicando que, en la mayoría de las áreas, la frecuencia de violencia familiar es relativamente baja en relación con la población total.	50
2.31. Post-procesamiento, las variables son mencionadas en cada gráfica.	50

2.32. Este mapa de calor de la matriz de correlación visualiza las relaciones entre diferentes indicadores socioeconómicos y de criminalidad por población. Los tonos rojos indican una correlación positiva fuerte, mientras que los tonos azules representan una correlación negativa. Este análisis ayuda a identificar patrones y posibles interdependencias entre variables como educación, salud, y tipos de delito, proporcionando una base para decisiones de política y priorización de recursos.	51
2.33. Tomando 6 variables podremos alcanzar más del 90 % de la varianza acumulada.	53
2.34. Mapa de calor que representa la matriz de correlación con la reducción de variables.	54
2.35. Tomando 3 variable podremos alcanzar mas del 95 % de la varianza acumulada.	55
2.36. Mapa de calor que representa la matriz de correlación con la reducción de variables.	55
2.37. Gráfica del método del codo mostrando la inercia por número de clusters para determinar el número óptimo. El punto de inflexión alrededor de 4 a 5 clusters sugiere el equilibrio ideal entre el número de clusters y la compacidad de cada uno.	61
2.38. Representación en componentes principales (PC ₁ y PC ₂) de la distribución de delitos, instituciones y población por cluster. Cada color representa un cluster diferente, destacando la variabilidad y agrupación de los datos en el espacio de las dos principales componentes.	63
2.39. Visualización 3D utilizando las tres principales componentes para mostrar la distribución de delitos, instituciones y población por cluster. Cada color identifica un cluster distinto, revelando cómo se diferencian estos grupos en un espacio tridimensional más complejo.	64
2.40. gráfica correspondiente a los indicadores de violencia	65
2.41. gráfica correspondiente a los indicadores de población y educación	66
2.42. Comparativa de la precisión en la predicción de casos de abuso sexual infantil utilizando tres modelos de regresión diferentes: Gradient Boosting Regressor, Regresión Lineal y MLPRegressor. Cada gráfico muestra la relación entre los valores predichos por el modelo y los valores reales, donde la línea discontinua representa la predicción perfecta.	67
2.43. Comparativa de la precisión en la predicción de casos de Homicidio doloso utilizando tres modelos de regresión diferentes: Gradient Boosting Regressor, Regresión Lineal y MLPRegressor.	68
2.44. Comparativa de la precisión en la predicción de casos de Femicidio utilizando tres modelos de regresión diferentes: Gradient Boosting Regressor, Regresión Lineal y MLPRegressor.	68
2.45. Comparativa de la precisión en la predicción de casos de abuso violacion utilizando tres modelos de regresión diferentes: Gradient Boosting Regressor, Regresión Lineal y MLPRegressor.	69

2.46. Comparativa de la precisión en la predicción de casos de violencia familiar utilizando tres modelos de regresión diferentes: Gradient Boosting Regressor, Regresión Lineal y MLPRegressor.	69
3.1. Femicidio - Frecuencia de los casos de femicidio.	76
3.2. Femicidio - Frecuencia de los casos de Violaciones.	77

Índice de tablas

	Página
1.1. Esta tabla detalla la distribución de alumnos, docentes y escuelas en diversos niveles educativos y modalidades dentro del sistema educativo. La información está segmentada entre educación pública y privada, así como entre modalidades escolarizadas y no escolarizadas. Los datos resaltan la predominancia de la educación pública en casi todos los niveles, particularmente en la educación primaria, y proporciona una comparativa de género en la matrícula, mostrando una presencia significativa tanto de mujeres como de hombres en las diferentes etapas educativas.	18
2.1. Descripción de los campos de la base de datos	23
2.2. Descripción de campos seleccionados para el análisis	31
2.3. Planteles Educativos, Descripción de campos seleccionados para el análisis . .	33
2.4. Esta tabla clasifica diversos delitos por niveles de gravedad del 1 al 5, proporcionando una justificación basada en el impacto emocional, físico y psicológico en las víctimas, así como el efecto en la comunidad. Los delitos de nivel 5, como el homicidio doloso, violación, abuso sexual infantil y feminicidio, son reconocidos por su severidad extrema y consecuencias profundas. Los niveles inferiores reflejan delitos que, aunque menos graves, siguen teniendo un impacto significativo en la seguridad y el bienestar general. Esta clasificación ayuda a entender mejor la urgencia y prioridad que deberían tener las políticas de prevención y respuesta legal.	35
2.5. Planteles Educativos, Descripción de campos seleccionados para el análisis . .	35
2.6. Descripción de campos seleccionados para el análisis	36
2.7. Información general del DataFrame.	39
2.8. Estadísticas iniciales de los campos.	40
3.1. Resultados de los modelos para abuso sexual infantil.	74
3.2. Resultados de los modelos para homicidio doloso.	75
3.3. Resultados de los modelos para feminicidio.	75
3.4. Resultados de los modelos para el delito de violación.	77
3.5. Resultados de los modelos para el delito de violencia familiar.	78

Este trabajo representa la culminación de un camino lleno de aprendizaje y enseñanzas, pero también es cierto que fue un camino marcado con diversos retos y sacrificios. Sacrificios que no solo me pertenecen a mí, también a la gente que me rodea, mis padres, amigos, mi novia. Quienes me han ofrecido su apoyo, son ellos quienes en los momentos donde me pude sentir mas abrumado me han brindado un poco de su energía y aliento para continuar.

Dedicado a...

*a **ASTRAZENECA, Gobierno de Zapopan e ITESO**, Gracias a su generoso apoyo para que pudiera cursar esta maestría, he podido alcanzar nuevos horizontes en mi educación y desarrollo profesional. Esta oportunidad ha sido un verdadero catalizador para mi crecimiento*

A mi lead Manager, Claudia Oliva

Quiero agradecerte de corazón por tu comprensión y flexibilidad durante mi maestría. Hubo momentos en los que las actividades parecían sobrepasarme, y tu apoyo fue fundamental para poder superarlos

***A mis amigos Ricardo de León y Lilivette Cruz**, Compañeros de esta travesía, definitivamente esta aventura no hubiera sido lo mismo sin la energía que trajeron ustedes dos. Gracias por ser parte de este equipo que formamos, donde el apoyo mutuo fue clave para hacer que este camino valiera la pena.*

***A mi novia, Ana Luisa García**, Gracias por ser mi apoyo y mi resistencia. Tus porras, palabras de aliento y amor fueron ingredientes indispensables que me impulsaron a dar ese esfuerzo adicional y lo mejor de mí.*

***A mi Madre, Hirma Aguirre**, Gracias por siempre creer que lo que sea que me proponga lo puedo cumplir. Y aunque quizá no sea del todo cierto, pero escucharlo de ti me hace sentir con super poderes, me llena de confianza y me hace sentir capaz de lograr lo que sea.*

***Y a mi Padre, Sergio Zambrano**, Quien con su ejemplo de vida me ha enseñado qué tan importantes son la constancia y la perseverancia, Que la vida tarde o temprano premia aquellos que están dispuestos a dar lo mejor de sí.*

Termino con estas palabras, que al escucharlas de el tomaron un nuevo significado y me han acompañado a lo largo de mi camino.

“Se como la luz de la aurora boreal, que va en aumento hasta que el día es perfecto.”

1 Introducción

1.1 Contexto

El estado de Jalisco, reconocido por su profunda cultura y tradiciones emblemáticas, cubre una extensión de 78,596 km² en el corazón de México [1].

Con una población que supera los 8.3 millones de habitantes [2], Jalisco representa un intrincado equilibrio entre zonas urbanas y rurales. Esta población se divide principalmente entre el Área Metropolitana de Guadalajara (AMG), compuesta por diez municipios, y el interior del estado que comprende los 115 municipios restantes. Pero al igual que muchas otras regiones de México, Jalisco enfrenta desafíos cruciales, principalmente en áreas de delincuencia y educación.

Para el ciclo escolar 2021-2022, Jalisco cuenta con 2 millones 279 mil 231 alumnos en 15 mil 015 escuelas estatales. De ellos, 373 mil 240 (16.4 %) están matriculados en instituciones particulares y 1 millón 905 mil 991 (83.6 %) en escuelas públicas[3].

Al desglosar la matrícula total se tiene que:

Primaria: 933,401 estudiantes

Secundaria: 416,478 estudiantes

Nivel medio superior 319,518 estudiantes

Nivel superior: 248,716 estudiantes

Preescolar: 281,773 estudiantes

Educación inicial: 9,681 estudiantes

Educación especial: 28,618 estudiantes

Para mayor detalle consulte la tabla (Tabla 1.1):

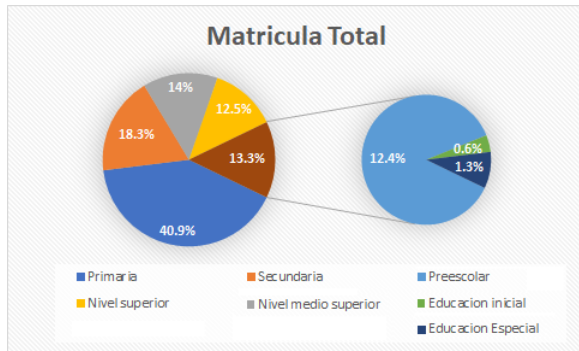


Figura 1.1: Esta gráfica ilustra la distribución de matrículas en diferentes niveles educativos. La mayoría de los estudiantes se encuentran en primaria (40.9%), seguido de secundaria (18.3%) y preescolar (14%). Los niveles de educación superior y medio superior representan un 25.8% combinado. La educación inicial y especial, aunque esenciales, abarcan sólo el 1.9% del total, reflejando áreas potencialmente menos atendidas en el sistema educativo.

A pesar de estas cifras, Jalisco ocupa el lugar 22 a nivel nacional en analfabetismo para personas mayores de 15 años, con un 4.4%, en contraste con la media nacional de 6.9%. Cuando se divide por género, el 4.1% de los hombres y el 4.6% de las mujeres en ese rango de edad son analfabetos[4].

La educación, más allá de ser un derecho, es un vehículo poderoso para el cambio social. En Jalisco, la relación entre la educación y la delincuencia es palpable. La falta de oportunidades educativas de calidad, junto con otros factores socioeconómicos, puede desencadenar descontento, desempleo y, en última instancia, incrementos en la delincuencia.

Aunque Jalisco ha logrado significativos avances en el acceso a la educación, enfrenta el reto constante de elevar su calidad. A pesar de que el sistema educativo estatal acoge a más de 2.2 millones de estudiantes, estas cifras, por impresionantes que sean, encubren notables desigualdades y retos. Mientras la educación básica prevalece, albergando al 74.3% de los estudiantes[4], las tasas de cobertura en los niveles medio superior y superior experimentan una marcada reducción. Las áreas de equidad, pertinencia y eficiencia se perfilan como áreas críticas de enfoque.

Esta insuficiencia educativa no solo influye en factores como la pobreza, sino que también tiene un impacto directo en la seguridad pública. De hecho, la insuficiencia y falta de calidad educativa podrían estar en la raíz de los incrementos en la violencia en el estado.

En años recientes, Jalisco ha experimentado un incremento en la incidencia delictiva, una tendencia que no solo representa cifras sino vidas afectadas, familias desplazadas y la desintegración social. Los 16 delitos de seguimiento prioritario, que abarcan desde el homicidio doloso hasta la violencia familiar, ilustran la magnitud de la crisis. Las alarmantes cifras publicadas por el Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública indican que, durante el primer trimestre de 2023, se registró una víctima de homicidio doloso cada 20 minutos en Jalisco, y sorprendentemente, el 70% de estos homicidios se cometieron usando armas de fuego[5].

Aunque algunos delitos, como los robos, han mostrado una tendencia decreciente, el panorama general sigue siendo sombrío. La administración federal actual ya se ha catalogado como una de las más violentas en la historia

reciente. Esto plantea una pregunta esencial: ¿De qué manera la educación, como derecho y herramienta transformadora, podría contribuir a modificar el panorama de violencia y delincuencia en Jalisco?[6]-[7].

(Ver Bibliografía 4.3)

Tablas

Tipo / modalidad / nivel	Alumnos			Docentes	Escuelas
	Total	Mujeres	Hombres		
Total del Sistema Educativo	2,252,329	1,135,058	1,117,271	135,259	14,734
Público	1,877,829	942,944	934,885	98,681	11,629
Privado	374,500	192,114	182,386	36,578	3,105
Modalidad escolarizada	2,209,567	1,111,374	1,098,193	130,045	14,531
Público	1,866,222	936,966	929,256	97,415	11,586
Privado	343,345	174,408	168,937	32,630	2,945
Educación básica	1,641,333	808,770	832,563	76,416	13,079
Público	1,427,850	703,676	724,174	63,076	10,835
Privado	213,483	105,094	108,389	13,340	2,244
Educación inicial	9,681	4,646	5,035	743	186
General	9,448	4,528	4,920	730	173
Indígena	233	118	115	13	13
Público	1,527	747	780	212	66
Privado	8,154	3,899	4,255	531	120
Educación preescolar	281,773	139,917	141,856	14,266	5,109
General ^I	272,436	135,364	137,072	13,242	3,980
Indígena	1,146	529	617	78	54
Cursos comunitarios	8,191	4,024	4,167	946	1,075
Público	233,909	116,127	117,782	11,122	4,022
Privado	47,864	23,790	24,074	3,144	1,087
Educación primaria	933,401	457,479	475,922	33,851	5,624
General	923,699	452,676	471,023	33,118	5,138
Indígena	6,324	3,119	3,205	322	104
Cursos comunitarios	3,378	1,684	1,694	411	382
Público	824,503	403,683	420,820	29,076	4,993
Privado	108,898	53,796	55,102	4,775	631
Educación secundaria	416,478	206,728	209,750	27,556	2,160
General ^{II}	252,806	125,355	127,451	18,180	1,198
Telesecundaria	39,866	19,626	20,240	2,249	646
Técnica	123,806	61,747	62,059	7,127	316
Público	367,911	183,119	184,792	22,666	1,754
Privado	48,567	23,609	24,958	4,890	406
Educación media superior	319,518	169,245	150,273	27,552	1,145
Bachillerato general	233,289	127,293	105,996	17,238	932
Bachillerato tecnológico	61,475	30,779	30,696	6,955	168
Profesional técnico bachiller	13,107	5,810	7,297	2,029	30
Profesional técnico	11,647	5,363	6,284	1,330	15
Público	268,449	143,230	125,219	20,651	679
Privado	51,069	26,015	25,054	6,901	466
Educación superior^{III}	248,716	133,359	115,357	26,077	307
Licenciatura	235,729	126,720	109,009	21,702	256
Normal	5,691	4,475	1,216	606	20
Universitaria y tecnológica	230,038	122,245	107,793	21,096	236
Posgrado	12,987	6,639	6,348	4,375	132
Público	169,923	90,060	79,863	13,688	72
Privado	78,793	43,299	35,494	12,389	235

Tabla 1.1: Esta tabla detalla la distribución de alumnos, docentes y escuelas en diversos niveles educativos y modalidades dentro del sistema educativo. La información está segmentada entre educación pública y privada, así como entre modalidades escolarizadas y no escolarizadas. Los datos resaltan la predominancia de la educación pública en casi todos los niveles, particularmente en la educación primaria, y proporciona una comparativa de género en la matrícula, mostrando una presencia significativa tanto de mujeres como de hombres en las diferentes etapas educativas.

1.2 *Justificación*

La delincuencia y la violencia son desafíos persistentes que enfrentan muchas regiones del mundo, incluido el estado de Jalisco en México. Estos problemas no solo afectan la calidad de vida y la seguridad de los residentes, sino que también tienen un impacto económico y social significativo, desalentando la inversión, el turismo y el desarrollo sostenible. A pesar de las múltiples intervenciones y estrategias implementadas para abordar estos desafíos, aún queda mucho por hacer.

Dentro de este panorama, la educación surge como una herramienta vital y transformadora. Numerosos estudios han demostrado que el acceso a una educación de calidad tiene el potencial de ofrecer oportunidades económicas, fomentar la situación social y, en última instancia, reducir la delincuencia. Al dar a las personas las habilidades y conocimientos necesarios para triunfar en la comunidad, la educación puede funcionar como un mecanismo potente de prevención delictiva, presentando opciones distintas a la criminalidad y promoviendo una estructura social más unida y robusta.

Sin embargo, en Jalisco, como en muchos lugares, la relación entre educación y delincuencia no se ha explorado en profundidad a nivel municipal inclusive a nivel de colonias. Los municipios, con sus características únicas y desafíos específicos, pueden beneficiarse enormemente de un análisis detallado que vincule la presencia y calidad de las instituciones educativas con las tasas de delincuencia.

Este proyecto, al utilizar técnicas de ciencia de datos para analizar y correlacionar la disponibilidad y niveles de instituciones educativas con la incidencia delictiva, tiene el potencial de proporcionar perspectivas valiosas que pueden guiar las políticas y estrategias de desarrollo. Al identificar áreas donde el fortalecimiento de la infraestructura educativa puede tener el mayor impacto en la reducción de la delincuencia, se pueden hacer recomendaciones basadas en evidencia para donde se deberían crear nuevos planteles educativos y de qué tipo.

En última instancia, al comprender mejor esta relación, Jalisco tiene la oportunidad de hacer inversiones estratégicas en educación que no sólo enriquecerán la vida de sus ciudadanos sino que también contribuirán a la creación de un estado más seguro y próspero.

1.3 *Problema*

En Jalisco, los índices de delincuencia continúan siendo un desafío. Si bien se han implementado estrategias de seguridad, la correlación entre la educación y la delincuencia en el ámbito municipal sigue siendo ambigua.

A pesar de que existen múltiples investigaciones sobre educación y delincuencia a nivel macro, el análisis detallado a nivel municipal y a un nivel más granular de colonia en Jalisco es escaso. La ciencia de datos, con su capacidad para procesar y analizar grandes conjuntos de datos, ofrece una oportunidad única para abordar esta brecha. El problema científico radica en determinar, utilizando técnicas de análisis, si existe una correlación significativa entre la presencia, cantidad y niveles de planteles educativos y las variaciones en la incidencia delictiva en los municipios de Jalisco. Además, se busca discernir qué tipos y niveles educativos tienen el mayor impacto preventivo en la delincuencia.

La pregunta principal es: ¿Influye la presencia de instituciones educativas en los índices de criminalidad de un municipio?

1.4 *Objetivos*

1.4.1 *Objetivo general*

Analizar y determinar la relación entre la presencia y nivel de planteles educativos y la incidencia delictiva en el estado de Jalisco, a nivel de colonia, para proponer estrategias educativas que contribuyan a la disminución de la delincuencia.

1.4.2 *Objetivos específicos*

1. Recolectar y depurar las bases de datos relacionadas con la incidencia delictiva y las instituciones educativas en Jalisco (INEGI, IIEG y SEP).
2. Establecer métricas y categorías para clasificar la incidencia delictiva y el nivel educativo de los planteles en cada municipio.
3. Utilizar técnicas de ciencia de datos para identificar patrones o tendencias entre la delincuencia y la educación en el estado.
4. Desarrollar recomendaciones basadas en los hallazgos para guiar futuras políticas públicas o inversiones en educación como estrategia de prevención delictiva.

2 Metodología

2.1 Descripción de los datos

A continuación se incluye la descripción de los conjuntos de datos utilizados en este análisis del estado de Jalisco a nivel de colonia, lo cual nos otorga una perspectiva diferente y más puntual. Estos datos abarcan índices de criminalidad, instituciones educativas, así como información geográfica y demográfica.

2.2 Instituto de Información Estadística y Geográfica del Estado de Jalisco - IIEG

El portal de internet IIEG es la plataforma oficial del Instituto de Información Estadística y Geográfica del Estado de Jalisco (IIEG), siendo una herramienta muy útil para aquellos que buscan datos confiables y actualizados sobre el estado de Jalisco. Es una excelente fuente para comprender la situación social, económica y geográfica de Jalisco.[8]

Datos e información: *Ofrece acceso a una amplia gama de datos sobre diversos temas relacionados con Jalisco, como:*

- Población (información demográfica)
- Índices de criminalidad
- Economía (indicadores socioeconómicos)
- Educación
- salud
- Medio ambiente
- Geografía (mapas)

Recursos: *El sitio web proporciona varios recursos para los usuarios, como:*

- Paneles interactivos y visualizaciones de datos
- Conjuntos de datos descargables en diferentes formatos
- Informes y análisis estadísticos
- Boletines económicos
- Información acerca de las áreas geográficas en Jalisco (municipios)

2.2.1 Índices de criminalidad del estado de Jalisco

Dentro del portal del IIEG, existe una sección denominada "Plataforma de seguridad". Esta sección contiene un apartado llamado "laboratorio de datos", que facilita el acceso a información detallada sobre incidencia delictiva en el estado de Jalisco, permitiendo incluso obtener datos específicos a nivel de colonia [9].

Laboratorio de Datos: incluye la definición de términos utilizada por la Legislatura Estatal, de acuerdo con la Fiscalía del Estado.

Descripción DataSet:

Este dataset es útil para el análisis y la visualización de la seguridad y el crimen en el estado de Jalisco, permitiendo una comprensión más profunda de las tendencias delictivas a nivel localizado.

Este dataset cuenta con 555,158 registros. Se analizaron 126 municipios y 3,188 colonias, considerando 16 tipos de delito tipificados:

- Robo a carga pesada
- Robo a vehículos particulares
- Homicidio doloso
- Robo casa habitación
- Robo de motocicleta
- Robo a negocio
- Robo a persona
- Lesiones dolosas
- Robo a int de vehículos
- Violación
- Violencia familiar
- Abuso sexual infantil
- Robo de autopartes
- Robo a cuentahabientes
- Femicidio
- Robo a bancos

El período de estudio abarca desde Enero de 2017 hasta febrero de 2024.

Descripción de Campos

- **Fecha:** Fecha en que se cometió el delito (formato dd/mm/aaaa)
- **Delito:** Tipo de delito cometido
- **Latitud:** Coordenada de latitud del lugar del delito
- **Longitud:** Coordenada de longitud del lugar del delito
- **Colonia:** Colonia donde se cometió el delito
- **Municipio:** Municipio donde se cometió el delito
- **Clave municipio:** Clave INEGI del municipio
- **Hora:** Hora en que se cometió el delito (formato hh:mm)
- **Bien afectado:** Tipo de bien afectado por el delito
- **Zona geográfica:** Zona geográfica donde se cometió el delito (AMG, Interior)

Campo	Tipo de Dato	Campo	Tipo de Dato
Fecha	Fecha (String)	Latitud	Coordenada (float)
Delito	Texto (String)	Longitud	Coordenada (float)
Colonia	Texto (String)	Clave municipio	Numérico (int)
Municipio	Texto (String)	Hora	Hora (time)
Bien afectado	Texto (String)	Zona geográfica	Texto (String)

Tabla 2.1: Descripción de los campos de la base de datos

Comportamiento de los delitos

Este conjunto de gráficas ilustra el comportamiento de los distintos tipos de delito registrados entre diciembre de 2017 y febrero de 2024. Cada gráfica muestra las variaciones en la frecuencia de los delitos a lo largo del tiempo, permitiendo identificar tendencias, incrementos o disminuciones significativas en la ocurrencia de cada tipo de delito en el periodo mencionado.

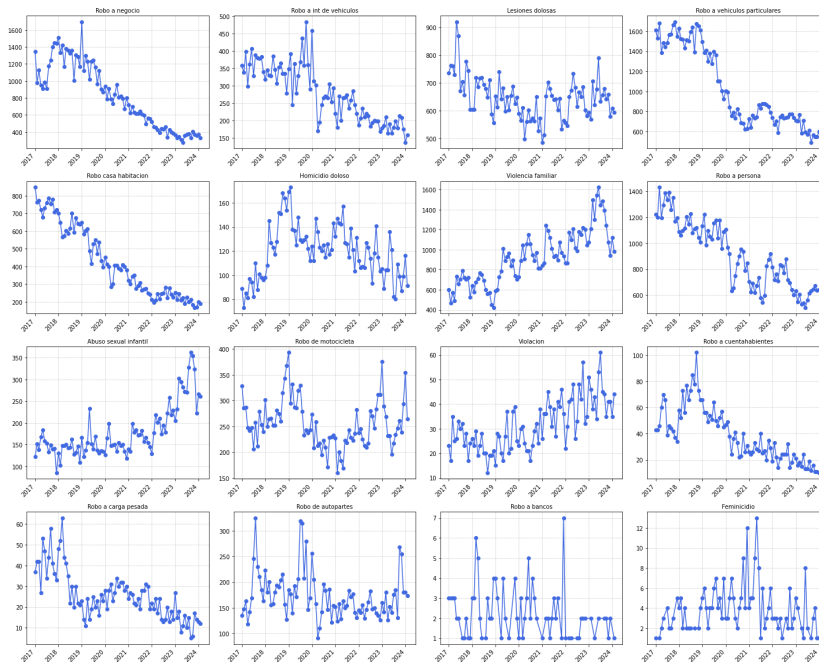


Figura 2.1: La serie de gráficos muestra la tendencia en la incidencia de diversos tipos de delitos en un periodo que abarca varios años(Diciembre 2017 a febrero 2024). Cada gráfico detalla un tipo específico de delito, desde robos en diferentes contextos hasta delitos más graves como homicidios y feminicidios.

Aquí algunas observaciones clave:

Robo a negocio, robo a vehículos particulares y robo a cuentahabientes muestran una tendencia general a la baja a lo largo del tiempo, lo cual podría indicar una mejora en las medidas de seguridad o en la eficacia de las estrategias de prevención en estos ámbitos.

Violencia familiar y abuso sexual infantil, por otro lado, presentan picos y variaciones significativas, lo que podría reflejar variaciones en la tasa de reportes más que una verdadera disminución en la incidencia.

Homicidio doloso y feminicidio exhiben fluctuaciones con periodos de aumento seguidos de disminuciones, sugiriendo que son influenciados

por factores complejos y posiblemente diferentes medidas de intervención y cambios en políticas públicas.

Delitos como **robo de motocicleta y robo a bancos** tienen menos incidencia en comparación con otros tipos, pero sus picos son abruptos y podrían estar vinculados a eventos específicos que alteran la tendencia.

Considerando que a lo largo de los años los gobiernos han implementado diversas estrategias para combatir la violencia, es posible notar que ciertos delitos han experimentado una disminución, señalando así la efectividad de algunas de estas tácticas. Sin embargo, es crucial reconocer que mientras algunos delitos han mostrado una tendencia a la baja, otros han permanecido estables o, incluso, han registrado un incremento. En vista de esto, el análisis subsiguiente se centrará exclusivamente en aquellos delitos que demandan una intervención urgente para su reducción.

Se considerarán los siguientes delitos aquellos que serán analizados debido a que muestran un aumento o se mantienen constantes en su frecuencia.

- Lesiones dolosas
- Violación
- Violencia familiar
- Abuso sexual infantil
- Femicidio

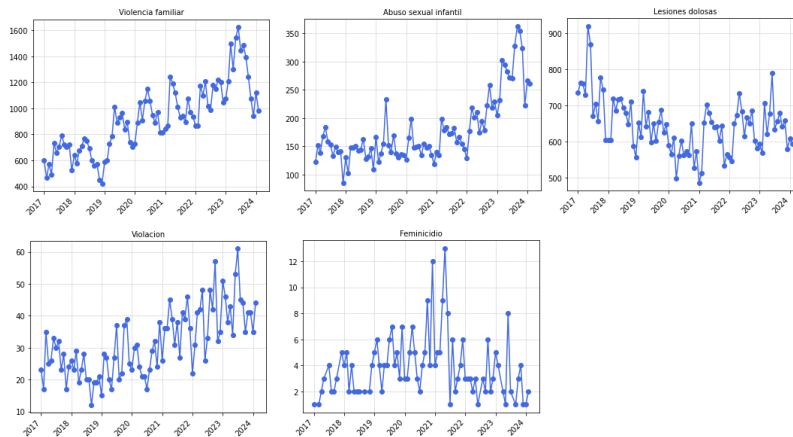


Figura 2.2: Esta serie de gráficos muestra la evolución mensual de delitos como violencia familiar, abuso sexual infantil, lesiones dolosas, violación y feminicidio desde 2017 hasta 2024. Las tendencias destacan un aumento en la mayoría de los delitos a partir de 2020, reflejando posibles cambios en los patrones de comportamiento social y efectividad de las políticas de prevención. Estos datos son cruciales para identificar periodos críticos y podrían orientar las estrategias de intervención y asignación de recursos hacia las áreas más afectadas.

Violencia familiar: Exhibe un notable aumento a partir de finales de 2020, alcanzando picos más altos en 2023. Esto podría sugerir un aumento en los casos reportados o un incremento real en la incidencia de este delito.

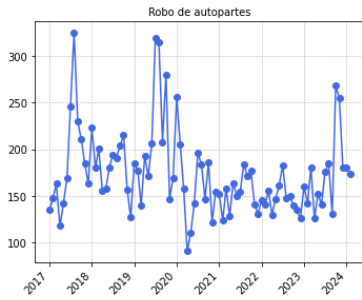
Abuso sexual infantil: Las cifras muestran fluctuaciones significativas con un aparente aumento en la frecuencia de los casos hacia 2022 y 2023. La variabilidad en la incidencia puede reflejar cambios en los mecanismos de reporte o sensibilización sobre este tema.

Lesiones dolosas: También muestra una tendencia general al alza desde 2021, con varias fluctuaciones a lo largo del tiempo, lo que podría indicar períodos de conflictos sociales o alteraciones de la paz pública más frecuentes.

Violación: Los casos de violación presentan un aumento progresivo desde 2021, con picos más notorios en los últimos años, lo que sugiere un creciente problema de seguridad y justicia en esta categoría.

Femicidio: A pesar de la volatilidad en los datos, hay un pico dramático observado en 2021, seguido de una disminución. Este tipo de delito, sin embargo, mantiene un nivel preocupante de incidencia a lo largo del tiempo.

- Robo de autopartes



- Robo de motocicleta

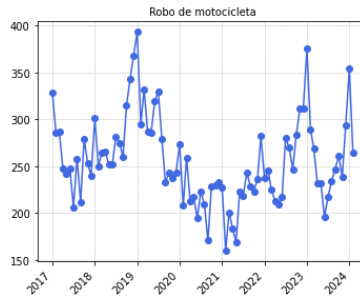


Figura 2.3: Estos gráficos muestran la tendencia mensual en los robos de autopartes y motocicletas desde 2017 hasta 2024. Ambas categorías presentan una considerable volatilidad en sus cifras, con picos que sugieren episodios de alta incidencia. La variabilidad en el robo de autopartes y motocicletas refleja los desafíos en la gestión y prevención de este tipo específico de crimen, lo cual podría ser indicativo de la necesidad de estrategias dirigidas y mejoradas para combatir eficazmente estos delitos.

En los gráficos de robos de autopartes y motocicletas, observamos una constancia en la frecuencia de los incidentes a lo largo del periodo analizado (2017-2024). Aunque hay picos que indican episodios de alta incidencia en momentos específicos, no se aprecia una tendencia clara ni hacia la mejora ni hacia el empeoramiento general de la situación.

2.3 INEGI

Es un organismo público autónomo responsable de normar y coordinar el Sistema Nacional de Información Estadística y Geográfica, así como de captar y difundir información de México en cuanto al territorio, los recursos, la población y economía, que permita dar a conocer las características de nuestro país y ayudar a la toma de decisiones.

2.3.1 Geografía y Demografía - INEGI

Los datos Geográficos y Demográficos son la clave para entender la composición y las necesidades de la población, dimensiones de las comunidades, educación y el empleo.

Censo población y vivienda 2020 INEGI información demográfica y social de colonias del estado Jalisco 2020.[10]

El Censo de Población y Vivienda se lleva a cabo cada 10 años en México. El más reciente fue el Censo 2020, que tuvo lugar del 2 al 27 de marzo.

Descripción DataSet:

El Censo de Población y Vivienda 2020 (Censo 2020) se realizó del 2 al 27 de marzo; en éste participaron poco más de 147 mil entrevistadoras y entrevistadores que recorrieron los cerca de dos millones de kilómetros cuadrados del territorio nacional, visitando cada una de las viviendas para obtener información sobre estas, contar a la población que vive en México e indagar sobre sus principales características demográficas, socioeconómicas y culturales.[10]

Este dataset cuenta con 5,567 registros. El dataset contiene 225 variables. Se analizaron 126 municipios y 3,729 colonias.

Descripción de Campos

- NUMCOL Número único de colonia
- MUNS Municipios en los que se encuentra
- NOMBRE Nombre de la colonia
- CP Código postal
- supm2 Superficie en kilómetros cuadrados
- POBTOT Población total
- POBFEM Población femenina
- POBMAS Población masculina
- P0A2 Población de 0 a 2 años
- P0A2F Población femenina de 0 a 2 años
- P0A2M Población masculina de 0 a 2 años
- P3YMAS Población de 3 años y más
- P3YMASF Población femenina de 3 años y más
- P3YMASM Población masculina de 3 años y más
- P5YMAS Población de 5 años y más
- P5YMASF Población femenina de 5 años y más
- P5YMASM Población masculina de 5 años y más
- P12YMAS Población de 12 años y más
- P12YMASF Población femenina de 12 años y más
- P12YMASM Población masculina de 12 años y más
- P15YMAS Población de 15 años y más
- P15YMASF Población femenina de 15 años y más
- P15YMASM Población masculina de 15 años y más
- P18YMAS Población de 18 años y más
- P18YMASF Población femenina de 18 años y más
- P18YMASM Población masculina de 18 años y más
- P3A5 Población de 3 a 5 años
- P3A5F Población femenina de 3 a 5 años
- P3A5M Población masculina de 3 a 5 años
- P6A11 Población de 6 a 11 años
- P6A11F Población femenina de 6 a 11 años
- P6A11M Población masculina de 6 a 11 años
- P8A14 Población de 8 a 14 años
- P8A14F Población femenina de 8 a 14 años
- P8A14M Población masculina de 8 a 14 años
- P12A14 Población de 12 a 14 años
- P12A14F Población femenina de 12 a 14 años
- P12A14M Población masculina de 12 a 14 años
- P15A17 Población de 15 a 17 años
- P15A17F Población femenina de 15 a 17 años
- P15A17M Población masculina de 15 a 17 años
- P18A24 Población de 18 a 24 años

- **P18A24F** Población femenina de 18 a 24 años
- **P18A24M** Población masculina de 18 a 24 años
- **P15A49F** Población femenina de 15 a 49 años
- **P60YMAS** Población de 60 años y más
- **P60YMASF** Población femenina de 60 años y más
- **P60YMASM** Población masculina de 60 años y más
- **RELHM** Relación hombres-mujeres
- **POB014** Población de 0 a 14 años
- **POB1564** Población de 15 a 64 años
- **POB65MAS** Población de 65 años y más
- **PROMHNV** Promedio de hijas e hijos nacidos vivos
- **PNACENT** Población nacida en la entidad
- **PNACENTF** Población femenina nacida en la entidad
- **PNACENTM** Población masculina nacida en la entidad
- **PNACOE** Población nacida en otra entidad
- **PNACOEF** Población femenina nacida en otra entidad
- **PNACOEM** Población masculina nacida en otra entidad
- **PRES2015** Población de 5 años y más residente en la entidad en marzo de 2015
- **PRES2015F** Población femenina de 5 años y más residente en la entidad en marzo de 2015
- **PRES2015M** Población masculina de 5 años y más residente en la entidad en marzo de 2015
- **PRESOE15** Población de 5 años y más residente en otra entidad en marzo de 2015
- **PRESOE15F** Población femenina de 5 años y más residente en otra entidad en marzo de 2015
- **PRESOE15M** Población masculina de 5 años y más residente en otra entidad en marzo de 2015
- **P3YMHLLI** Población de 3 años y más que habla alguna lengua indígena
- **P3YMHLLIF** Población femenina de 3 años y más que habla alguna lengua indígena
- **P3YMHLLIM** Población masculina de 3 años y más que habla alguna lengua indígena
- **P3HLINHE** Población de 3 años y más que habla alguna lengua indígena y no habla español
- **P3HLINHEF** Población femenina de 3 años y más que habla alguna lengua indígena y no habla español
- **P3HLINHEM** Población masculina de 3 años y más que habla alguna lengua indígena y no habla español
- **P3HLIHE** Población de 3 años y más que habla alguna lengua indígena y habla español
- **P3HLIHEF** Población femenina de 3 años y más que habla alguna lengua indígena y habla español
- **P3HLIHEM** Población masculina de 3 años y más que habla alguna lengua indígena y habla español
- **P5HLI** Población de 5 años y más que habla alguna lengua indígena
- **P5HLINHE** Población de 5 años y más que habla alguna lengua indígena y no habla español
- **P5HLIHE** Población de 5 años y más que habla alguna lengua indígena y habla español
- **PHOGIND** Población en hogares censales indígenas
- **POBAFRO** Población que se considera afroamericana o afrodescendiente
- **POBAFROF** Población femenina que se considera afroamericana o afrodescendiente
- **POBAFROM** Población masculina que se considera afroamericana o afrodescendiente
- **PCONDISC** Población con discapacidad
- **PCDISCMOT** Población con discapacidad para caminar, subir o bajar
- **PCDISCVIS** Población con discapacidad para ver, aun usando lentes
- **PCDISCLENG** Población con discapacidad para hablar o comunicarse
- **PCDISCAUD** Población con discapacidad para oír, aun usando aparato auditivo
- **PCDISCMOT2** Población con discapacidad para vestirse, bañarse o comer
- **PCDISCMEN** Población con discapacidad para recordar o concentrarse
- **PCONLIMI** Población con limitación
- **PCLIMCSB** Población con limitación para caminar, subir o bajar
- **PCLIMVIS** Población con limitación para ver, aun usando lentes
- **PCLIMHACO** Población con limitación para hablar o comunicarse
- **PCLIMOAUD** Población con limitación para oír, aun usando aparato auditivo
- **PCLIMMOT2** Población con limitación para vestirse, bañarse o comer
- **PCLIMRECO** Población con limitación para recordar o concentrarse
- **PCLIMPMEN** Población con algún problema o condición mental
- **PSINDLIM** Población sin discapacidad, limitación, problema o condición mental
- **P3A5NOA** Población de 3 a 5 años que no asiste a la escuela
- **P3A5NOAF** Población femenina de 3 a 5 años que no asiste a la escuela
- **P3A5NOAM** Población masculina de 3 a 5 años que no asiste a la escuela
- **P6A11NOA** Población de 6 a 11 años que no asiste a la escuela
- **P6A11NOAF** Población femenina de 6 a 11 años que no asiste a la escuela
- **P6A11NOAM** Población masculina de 6 a 11 años que no asiste a la escuela
- **P12A14NOA** Población de 12 a 14 años que no asiste a la escuela
- **P12A14NOAF** Población femenina de 12 a 14 años que no asiste a la escuela
- **P12A14NOAM** Población masculina de 12 a 14 años que no asiste a la escuela
- **P15A17A** Población de 15 a 17 años que asiste a la escuela
- **P15A17AF** Población femenina de 15 a 17 años que asiste a la escuela
- **P15A17AM** Población masculina de 15 a 17 años que asiste a la escuela
- **P18A24A** Población de 18 a 24 años que asiste a la escuela
- **P18A24AF** Población femenina de 18 a 24 años que asiste a la escuela
- **P18A24AM** Población masculina de 18 a 24 años que asiste a la escuela
- **P8A14AN** Población de 8 a 14 años que no sabe leer y escribir
- **P8A14ANF** Población femenina de 8 a 14 años que no sabe leer y escribir
- **P8A14ANM** Población masculina de 8 a 14 años que no sabe leer y escribir
- **P15YMAN** Población de 15 años y más analfabeta
- **P15YMANF** Población femenina de 15 años y más analfabeta
- **P15YMANM** Población masculina de 15 años y más analfabeta
- **P15YMSE** Población de 15 años y más sin escolaridad
- **P15YMSEF** Población femenina de 15 años y más sin escolaridad
- **P15YMSEM** Población masculina de 15 años y más sin escolaridad
- **P15PRIIN** Población de 15 años y más con primaria incompleta
- **P15PRIINF** Población femenina de 15 años y más con primaria incompleta
- **P15PRIINM** Población masculina de 15 años y más con primaria incompleta
- **P15PRICOF** Población de 15 años y más con primaria completa
- **P15PRICOF** Población femenina de 15 años y más con primaria completa
- **P15PRICOM** Población masculina de 15 años y más con primaria completa
- **P15SECIN** Población de 15 años y más con secundaria incompleta
- **P15SECINF** Población femenina de 15 años y más con secundaria incompleta
- **P15SECINM** Población masculina de 15 años y más con secundaria incompleta
- **P15SECCO** Población de 15 años y más con secundaria completa
- **P15SECCOF** Población femenina de 15 años y más con secundaria completa
- **P15SECCOM** Población masculina de 15 años y más con secundaria completa
- **P18YMPB** Población de 18 años y más con educación pos-básica
- **P18YMPBF** Población femenina de 18 años y más con educación pos-básica
- **P18YMPBM** Población masculina de 18 años y más con educación pos-básica
- **GRAPROES** Grado promedio de escolaridad
- **GRAPROESF** Grado promedio de escolaridad de la población femenina
- **GRAPROESM** Grado promedio de escolaridad de la población masculina
- **PEA** Población de 12 años y más económicamente activa
- **PEAF** Población femenina de 12 años y más económicamente activa

- PEAM Población masculina de 12 años y más económicamente activa
- PEINAC Población de 12 años y más no económicamente activa
- PEINACF Población femenina de 12 años y más no económicamente activa
- PEINACM Población masculina de 12 años y más no económicamente activa
- POCUPADA Población de 12 años y más ocupada
- POCUPADAF Población femenina de 12 años y más ocupada
- POCUPADAM Población masculina de 12 años y más ocupada
- PDESOCUP Población de 12 años y más desocupada
- PDESOCUPF Población femenina de 12 años y más desocupada
- PDESOCUPM Población masculina de 12 años y más desocupada
- PSINDER Población sin afiliación a servicios de salud
- PDERSS Población afiliada a servicios de salud
- PDERIMSS Población afiliada a servicios de salud en el IMSS
- PDERISTE Población afiliada a servicios de salud en el ISSSTE
- PDERISTEE Población afiliada a servicios de salud en el ISSSTE estatal
- PAFILPDOM Población afiliada a servicios de salud en PEMEX, Defensa o Marina
- PDERSEGP Población afiliada a servicios de salud en el Instituto de salud para el Bienestar
- PDERIMSSB Población afiliada a servicios de salud en el IMSS BIENESTAR
- PAFILPRIV Población afiliada a servicios de salud en una institución privada
- PAFILOTRAI Población afiliada a servicios de salud en otra institución
- P12YMSOLT Población de 12 años y más soltera o nunca unida
- P12YMCASA Población de 12 años y más casada o unida
- P12YMESEPA Población de 12 años y más que estuvo casada o unida
- PCATOLICA Población con religión católica
- PROCRIEVA Población con grupo religioso protestante/cristiano evangélico
- POTRASREL Población con otras religiones diferentes a las anteriores
- PSINRELIG Población sin religión o sin adscripción religiosa
- TOTHOG Total de hogares censales
- HOGJEFF Hogares censales con persona de referencia mujer
- HOGJEFM Hogares censales con persona de referencia hombre
- POBHOG Población en hogares censales
- PHOGJEFF Población en hogares censales con persona de referencia mujer
- PHOGJEFM Población en hogares censales con persona de referencia hombre
- VIVTOT Total de viviendas
- TVIVHAB Total de viviendas habitadas
- TVIVPAR Total de viviendas particulares
- VIVPARHAB Viviendas particulares habitadas
- TVIVPARHAB Total de viviendas particulares habitadas
- VIVPARDES Viviendas particulares deshabitadas
- VIVPARUT Viviendas particulares de uso temporal
- OCUPVIVPAR Ocupantes en viviendas particulares habitadas
- PROMOCUP Promedio de ocupantes en viviendas particulares habitadas
- PROOCUPC Promedio de ocupantes por cuarto en viviendas particulares habitadas
- VPHISODT Viviendas particulares habitadas con piso de material diferente de tierra
- VPHISOTI Viviendas particulares habitadas con piso de tierra
- VPHIDOR Viviendas particulares habitadas con un dormitorio
- VPH2YMASD Viviendas particulares habitadas con dos dormitorios y más
- VPH1CUART Viviendas particulares habitadas con sólo un cuarto
- VPH2CUART Viviendas particulares habitadas con dos cuartos
- VPH3YMASC Viviendas particulares habitadas con 3 cuartos y más
- VPHCELEC Viviendas particulares habitadas que disponen de energía eléctrica
- VPHSELEC Viviendas particulares habitadas que no disponen de energía eléctrica
- VPHAGUADV Viviendas particulares habitadas que disponen de agua entubada en el ámbito de la vivienda
- VPHAEASP Viviendas particulares habitadas que disponen de agua entubada y se abastecen del servicio público de agua
- VPHAGUAFV Viviendas particulares habitadas que no disponen de agua entubada en el ámbito de la vivienda
- VPH TINACO Viviendas particulares habitadas que disponen de tinaco
- VPHCISTER Viviendas particulares habitadas que disponen de cisterna o aljibe
- VPHXCSA Viviendas particulares habitadas que disponen de excusado o sanitario
- VPHLETR Viviendas particulares habitadas que disponen de letrina (pozo u hoyo)
- VPHDRENAJ Viviendas particulares habitadas que disponen de drenaje
- VPHNODREN Viviendas particulares habitadas que no disponen de drenaje
- VPHCSERV Viviendas particulares habitadas que disponen de energía eléctrica, agua entubada de la red pública y drenaje
- VPHNDEAED Viviendas particulares habitadas que no disponen de energía eléctrica, agua entubada, ni drenaje
- VPHDSADMA Viviendas particulares que disponen de drenaje y sanitario con admisión de agua
- VPHNDACMM Viviendas particulares habitadas que no disponen de automóvil o camioneta, ni de motocicleta o motoneta
- VPHSBIEN Viviendas particulares habitadas sin ningún bien
- VPHREFRI Viviendas particulares habitadas que disponen de refrigerador
- VPHLAVAD Viviendas particulares habitadas que disponen de lavadora
- VPHMICRO Viviendas particulares habitadas que disponen de horno de microondas
- VPHAUTOM Viviendas particulares habitadas que disponen de automóvil o camioneta
- VPHMOTO Viviendas particulares habitadas que disponen de motocicleta o motoneta
- VPHBICI Viviendas particulares habitadas que disponen de bicicleta como medio de transporte
- VPHRADIO Viviendas particulares habitadas que disponen de radio
- VPHTV Viviendas particulares habitadas que disponen de televisor
- VPHPC Viviendas particulares habitadas que disponen de computadora, tablet o laptop
- VPHTELEF Viviendas particulares habitadas que disponen de línea telefónica fija
- VPHCEL Viviendas particulares habitadas que disponen de teléfono celular
- VPHINTER Viviendas particulares habitadas que disponen de Internet
- VPHSTVP Viviendas particulares habitadas que disponen de servicio de televisión de paga
- VPHSPMVP Viviendas particulares habitadas que disponen de servicio de películas, música o videos de paga por Internet
- VPHCVJ Viviendas particulares habitadas que disponen de consola de videojuegos
- VPHSINRTV Viviendas particulares habitadas sin radio ni televisor
- VPHSINLTC Viviendas particulares habitadas sin línea telefónica ni teléfono celular
- VPHSINCINT Viviendas particulares habitadas sin computadora ni Internet
- VPHSINTIC Viviendas particulares habitadas sin tecnologías de la información y de la comunicación (TIC)

Aunque el conjunto de datos abarca 225 variables, se decidió focalizar el análisis en tan sólo 14 columnas. Esta selección se debe a la vastedad de información recabada en un censo, de la cual sólo una fracción se relaciona directamente con temas de educación, nuestro interés principal en este estudio. Cabe destacar que, aunque nuestro análisis se concentra en factores educativos, explorar la inseguridad desde diferentes ángulos utilizando este mismo conjunto de datos representa una oportunidad prometedora para futuros proyectos.

La variables a considerarse son las siguientes:

- Colonia
- Código postal
- Superficie en kilómetros cuadrados
- Población femenina
- Población masculina
- Población femenina de 15 años y más
- Población masculina de 15 años y más
- Población de 15 años y más sin escolaridad
- Población femenina de 15 años y más sin escolaridad
- Población masculina de 15 años y más sin escolaridad
- Población femenina de 18 años y más con educación posbásica
- Población masculina de 18 años y más con educación posbásica
- Población sin afiliación a servicios de salud

Campo	Tipo de dato
Colonia	Entero (Integer)
Código postal	Entero (Integer)
Superficie en kilómetros cuadrados	Entero (Integer)
Población femenina	Entero (Integer)
Población masculina	Entero (Integer)
Población femenina de 15 años y más	Entero (Integer)
Población masculina de 15 años y más	Entero (Integer)
Población de 15 años y más sin escolaridad	Entero (Integer)
Población femenina de 15 años y más sin escolaridad	Entero (Integer)
Población masculina de 15 años y más sin escolaridad	Entero (Integer)
Población femenina de 18 años y más con E. posbásica	Entero (Integer)
Población masculina de 18 años y más con E. posbásica	Entero (Integer)
Población sin afiliación a servicios de salud	Entero (Integer)

Tabla 2.2: Descripción de campos seleccionados para el análisis

2.4 Datos Jalisco

En esta plataforma ("Plataforma de seguridad del portal del IIEG") se encuentran los datos de carácter público generados por las dependencias y entidades gubernamentales, están disponibles en formatos digitales en línea, y pueden ser usados, reutilizados y redistribuidos por cualquier interesado[11].

Datos e información: Ofrece acceso a una amplia gama de datos sobre diversos temas relacionados con Jalisco, como:

- Ciencia y tecnología
- Desarrollo e integración social
- Desarrollo económico y comercio
- Educación
- Empleo
- Gobierno
- Información estadística y geográfica
- Infraestructura
- salud
- Turismo
- Movilidad
- Medio ambiente

2.4.1 Centros educativos Activos Jalisco

Registro de escuelas activas, desde nivel preescolar a profesional técnico, que contiene la información general del centro de trabajo, claves INEGI de localidad y municipio, nombre.

Descripción DataSet:

En este conjunto de datos se recopila información crucial sobre los planteles educativos ubicados en el estado de Jalisco, desglosada a nivel de colonia. Esta detallada segmentación es de vital importancia, ya que facilita el análisis a un nivel granular, permitiéndonos examinar la incidencia delictiva específica de cada colonia, tal como se propuso en nuestro análisis.

Este dataset cuenta con 16,082 registros.

Se analizaron 125 municipios y 155 colonias, Es importante destacar que, de las 3 mil colonias presentes en el estado de Jalisco, únicamente 155 cuentan con planteles educativos de algún nivel escolar.

Esta información es pertinente hasta el año 2023.

Descripción de Campos

- Clave del centro de trabajo
- Clave del turno
- Nombre del turno
- Nombre del centro de trabajo
- Tipo educativo
- Nivel educativo
- Servicio educativo
- Nombre del control (Público o Privado)
- Tipo de sostenimiento
- Clave de la entidad federativa
- Nombre de la entidad
- Municipio ID
- Nombre del municipio o delegación
- Clave de la localidad
- Nombre de localidad
- Domicilio
- Número exterior
- Entre la calle
- Y la calle
- Calle posterior
- ID Colonia
- Colonia
- Código postal

- Clave lada
- Teléfono
- Extensión del teléfono
- Página web
- Alumnos total (hombres)
- Alumnos total (mujeres)
- Alumnos total
- Docentes total (hombres)
- Docentes total (mujeres)
- Docentes total
- Aulas en uso
- Aulas existentes
- Tipo de localidad U - Urbano, R - Rural
- Ubicación de la escuela-localidad al Oeste del Meridiano de Greenwich, expresada en grados, minutos y segundos
- Ubicación de la escuela-localidad al norte del Ecuador, expresada en grados, minutos y segundos
- Ubicación de la escuela-localidad al Oeste del Meridiano de Greenwich, expresada en grados
- Ubicación de la escuela-localidad al norte del Ecuador, expresada en grados

La variables a considerarse son las siguientes:

- Colonia
- Tipo educativo (BÁSICA, MEDIA SUPERIOR, SUPERIOR, CAPACITACIÓN, INICIAL y ESPECIAL)
- Nivel educativo (PRIMARIA, BACHILLERATO, LICENCIATURA, PREES-COLAR, SECUNDARIA, INICIAL GENERAL, FORMACIÓN PARA EL TRABAJO y INICIAL)
- Alumnos total (hombres)
- Alumnos total (mujeres)
- Alumnos total
- Tipo de localidad U - Urbano, R - Rural

Campo	Tipo de dato
Colonia	Texto (<i>string</i>)
Tipo educativo	Texto (<i>string</i>)
Nivel educativo	Texto (<i>string</i>)
Alumnos total (hombres)	Entero (<i>Integer</i>)
Alumnos total (mujeres)	Entero (<i>Integer</i>)
Alumnos total	Entero (<i>Integer</i>)
Tipo de localidad	Texto (<i>Integer</i>)

Tabla 2.3: Planteles Educativos, Descripción de campos seleccionados para el análisis

2.5 Clasificador de delitos en Jalisco por nivel de violencia

En función de poder dotar de una forma de medición he creado este listado que clasifica los delitos en Jalisco según su nivel de violencia. Realmente esta lista está basada en mi propia interpretación. Para hacerlo, he tenido en cuenta lo que dicen las leyes, así como los informes de organizaciones que se dedican a la seguridad y los derechos humanos.

Nota: Es importante mencionar que la percepción de la gravedad de los delitos puede cambiar dadas las propias circunstancias de cada delito. Pero es importante poder dotar de una ponderación general, si quisiéramos ser más precisos debería ser necesario hablar directamente con las autoridades y se evalué cada delito para poder asignar un valor.

2.5.1 Escala de medición

Se han organizado los delitos en una escala del 1 al 5, donde el 1 representa los delitos que, desde mi punto de vista, son menos violentos. Estos suelen ser aquellos cuyas consecuencias son principalmente materiales y no ponen directamente en peligro la integridad física de las personas. En el extremo opuesto con valor de 5 se ha colocado aquellos delitos que considero los más graves y violentos. Estos son los que, por su naturaleza, causan gran daño profundo y directo a las víctimas.

Descripción - Escala de medición

- **5 (Alta violencia y daño causado):** Los delitos como el homicidio doloso, feminicidio, violación y abuso sexual infantil son clasificados con la máxima ponderación de violencia debido al daño físico y psicológico directo sobre las víctimas.
- **4 (Violencia y daño significativos):** Aquellos delitos donde implica una confrontación directa, la cual puede derivar en daños físicos, así como emocionales, así como aquellos delitos que implican un alto grado de organización y violencia.
- **3 (Daño moderado con impacto directo a individuos o propiedad):** Los robos a persona, negocio, y a cuentahabientes suelen involucrar confrontación o violencia directa hacia las víctimas, teniendo un impacto psicológico significativo además del daño material.
- **2 (Impacto material con menor violencia directa):** Robos de vehículos, motocicletas y al interior de vehículos generalmente involucran un daño económico significativo pero menos violencia física directa hacia las personas, aunque el riesgo de violencia está presente.
- **1 (Menor impacto directo en la seguridad física de las personas):** El robo de autopartes, aunque representa un delito económico, tiene el menor impacto directo en la seguridad y bienestar físico de las personas.

Delito	Nivel	Justificación
Robo de autopartes	1	Este delito afecta el patrimonio de las víctimas y genera un clima de inseguridad en la comunidad [12].
Robo a int de vehículos	2	Similar al robo de vehículos particulares, este delito genera un alto grado de intimidación y afectación emocional en las víctimas[13].
Robo a vehículos particulares	2	Si bien no siempre implica violencia física, este delito genera un alto grado de intimidación y afectación emocional en las víctimas[14].
Robo de motocicleta	2	Similar al robo de vehículos particulares, este delito genera un alto grado de intimidación y afectación emocional en las víctimas[15].
Lesiones dolosas	3	El nivel de violencia varía según la gravedad de las lesiones, pero en general este delito genera un daño físico y psicológico en las víctimas[16].
Robo a negocio	3	Este delito afecta el patrimonio de las víctimas y genera un clima de inseguridad en la comunidad[17].
Robo a persona	3	Este delito implica un contacto físico con la víctima, lo que genera un alto grado de intimidación y afectación emocional[18].
Robo a cuentahabientes	3	Este delito implica un alto grado de intimidación y afectación emocional en las víctimas, ya que se aprovecha un momento de vulnerabilidad[19].
Robo a carga pesada	4	Este delito implica un alto grado de organización y violencia, ya que suele realizarse con armas de fuego y con la participación de varios sujetos[20].
Robo casa habitación	4	Este delito viola la seguridad y privacidad del hogar, lo que genera un alto impacto emocional en las víctimas[21].
Violencia familiar	4	Este delito afecta a las víctimas en el ámbito privado del hogar, lo que genera un alto impacto emocional y dificultad para escapar de la violencia[22].
Robo a bancos	4	Este delito implica un alto grado de organización y violencia, ya que suele realizarse con armas de fuego y con la participación de varios sujetos[23].
Homicidio doloso	5	Es el delito más grave, ya que atenta contra el derecho fundamental a la vida[24].
Violación	5	Este delito es una grave forma de violencia sexual que genera un profundo daño físico y psicológico en las víctimas[25].
Abuso sexual infantil	5	Este delito es una grave forma de violencia sexual que genera un profundo daño físico y psicológico en las víctimas, especialmente por su vulnerabilidad[26].
Feminicidio	5	Es un asesinato por razones de género que atenta contra la vida y la dignidad de las mujeres[27].

Tabla 2.4: Esta tabla clasifica diversos delitos por niveles de gravedad del 1 al 5, proporcionando una justificación basada en el impacto emocional, físico y psicológico en las víctimas, así como el efecto en la comunidad. Los delitos de nivel 5, como el homicidio doloso, violación, abuso sexual infantil y feminicidio, son reconocidos por su severidad extrema y consecuencias profundas. Los niveles inferiores reflejan delitos que, aunque menos graves, siguen teniendo un impacto significativo en la seguridad y el bienestar general. Esta clasificación ayuda a entender mejor la urgencia y prioridad que deberían tener las políticas de prevención y respuesta legal.

La variables a considerarse son las siguientes:

- Delito
- Escala

Campo	Tipo de dato
Delito	Texto (<i>string</i>)
Escala	Numerico (<i>integer</i>)

Tabla 2.5: Planteles Educativos, Descripción de campos seleccionados para el análisis

2.6 Análisis exploratorio

2.6.1 Construcción de DataSet unificado

En esta sección se realizará el procedimiento necesario para formar un dataset final, el cual estará conformado por las tablas de Índices de criminalidad, Información geográfica y demográfica, Instituciones educativas y la ponderación de los delitos.

A continuación se enlistan los pasos a seguir para la construcción del dataset unificado

Los pasos se enlistan en el orden que fueron ejecutados, en la mayoría de los pasos no se tendrá que profundizar en una mayor explicación, salvo por aquellos que por su complejidad si lo requieran.

1. **Eliminación de columnas**, Se ha llevado a cabo una limpieza inicial de variables en cada conjunto de datos, eliminando aquellas columnas que no son necesarias para el análisis propuesto. Las columnas que se han eliminado se encuentran detalladas en el documento proporcionado. Dado que la lista de variables eliminadas es extensa, sólo se conservan aquellas que están explícitamente indicadas a continuación.

Criminalidad	Datos Demográficos	Instituciones educativas
Fecha	Colonia	Colonia
Delito	Código postal	Tipo educativo
Colonia	Superficie en kilómetros cuadrados	Nivel educativo
Clave municipio	Población femenina	Alumnos total (hombres)
Bien afectado	Población masculina	Alumnos total (mujeres)
	Población femenina de 15 años y más	Alumnos total
	Población masculina de 15 años y más	Tipo de localidad Texto
	Población de 15 años y más sin escolaridad	
	Pob fem de 15 años y más sin escolaridad	
	Pob masc de 15 años y más sin escolaridad	
	Pob fem de 18 años y más con E. posbásica	
	Pob masc de 18 años y más con E. posbásica	
	Pob sin afiliación a servicios de salud	

Tabla 2.6: Descripción de campos seleccionados para el análisis

2. **Asignación de ID de munición**, Cada dataset contaba con su propio ID de municipio, teniendo en común solamente el nombre, se realizó un merge de los dataset con el catálogo para obtener un ID de munición en común.
3. **Creación de ID único para los campos llave 'ID Municipio' y 'Colonia'**, Para facilitar las tareas de mapeo de datos se crea un ID único para cada municipio y colonia, esto debido a que muchos municipios tienen nombres de colonias iguales.
4. **Eliminación de registros sin 'ID único'**, Se realiza un proceso de limpieza de los datos con un 'ID único' erróneo, ya que en los dataSet se encontró

con colonias sin municipio asignado o con más de un municipio asignado, estos registros no serían de utilidad afectado al modelo.

5. Limpieza de dataset's

a) Criminalidad

- 1) Eliminación de valores N.D., 'NO DISPONIBLE' y los valores nulos en las columnas relacionadas con el delito cometido.
- 2) Conteo de apariciones por tipo de delito de cada colonia.
- 3) Unión con tabla de escala de delito, esto para poder asignar el valor descrito previamente a cada uno de los delitos.
- 4) Se multiplica la cantidad de delitos cometidos por su ponderación asignada.
- 5) Eliminación de columnas innecesarias.

b) Demografía

- 1) Eliminación de valores ND2020 (no disponible previo a 2020), Son datos que no contiene toda la información ya que en el sistema nuevo estos datos eran opcionales.
- 2) Limpieza de datos faltantes.
- 3) Sumarización de valores numéricos en función del 'ID único' esto debido a hay colonias que están divididas en 2 o 3 fracciones y cada una se registra de forma independiente, pero para fines prácticos este análisis las considerada como una misma.
- 4) Instituciones Educativas
- 5) Se eliminan datos faltantes (no se encontraron)
- 6) Se sumaliza la cantidad de instituciones educativas de cada nivel educativo de cada colonia.

6. Generación de variables Dummy Las variables dummy son variables binarias que se utilizan en modelos estadísticos para representar categorías cualitativas con números. Estas variables toman el valor de 1 para indicar la presencia de una categoría específica y 0 para indicar su ausencia, lo que permite incluir efectos categóricos en modelos que sólo aceptan variables numéricas.[28]

a) Criminalidad

- 1) Las variables dummy se construyen en base al campo 'Tipo de delito' esto nos genera una tabla donde cada columna corresponde a cada delito cometido.

b) Instituciones educativas

- 1) Turno, Se crean variables dummy para los valores 'Matutino' y 'Vespertino'
- 2) Tipo de institución, 'Pública' y 'Privada'
- 3) Nivel educativo, Especial, inicial, media superior y superior

7. Unión de los dataset's, Después de haber realizado el preprocesamiento de los datos, nuestro objetivo es consolidar toda la información en un único dataset. Para lograr esto, vamos a fusionar los distintos conjuntos de datos en uno sólo. Para llevar a cabo esta fusión, utilizaremos un identificador único compuesto por dos campos: el municipio y la colonia. Este identificador

único nos permitirá relacionar los registros de los diferentes conjuntos de datos de manera precisa.

Es importante destacar que durante el proceso de fusión, aquellos registros cuyo identificador único no tenga su correspondiente en otro conjunto de datos serán eliminados. Esto garantiza que sólo conservemos los registros que puedan ser completamente integrados y que evitemos la inclusión de datos incompletos o sin relación directa.

8. **Convertir variables en función de la población,** El objetivo es obtener valores comparables de diferentes indicadores, como la cantidad de escuelas, delitos cometidos, entre otros, en función de la distribución de la población. Para lograr esto, vamos a normalizar estos indicadores para que reflejen la cantidad de eventos por cada mil habitantes.

Importancia: La normalización por cada mil habitantes nos permite homogeneizar las poblaciones, teniendo en cuenta tanto aquellas con una extensión territorial mayor como aquellas con una densidad poblacional más alta. De esta manera, podemos comparar de manera más precisa la incidencia de los diferentes indicadores en diferentes áreas, independientemente de su tamaño o densidad poblacional.

2.6.2 Análisis Descriptivo Inicial

La tabla describe un conjunto de datos con 37 campos (columnas) y 4058 observaciones (filas).

#	Column	Non-Null Count	Dtype
0	ID único	0 nan de 4058	object
1	Superficie territorial	0 nan de 4058	float64
2	Mayor de 15 años sin educación_por_pobl	0 nan de 4058	float64
3	Sin acceso a servicios de salud_por_pobl	0 nan de 4058	float64
4	población_por_pobl	0 nan de 4058	float64
5	población mayor de edad_por_pobl	0 nan de 4058	float64
6	población mayor de edad con educación_por_pobl	0 nan de 4058	float64
7	Cantidad de docentes_por_pobl	0 nan de 4058	float64
8	Cantidad de alumnos_por_pobl	0 nan de 4058	float64
9	Cantidad de escuelas_por_pobl	0 nan de 4058	float64
13	Turno Nocturno_por_pobl	0 nan de 4058	float64
14	Turno Vespertino_por_pobl	0 nan de 4058	float64
15	control_PRIVADO_por_pobl	0 nan de 4058	float64
16	Institucion Publica_por_pobl	0 nan de 4058	float64
17	Tipo educativo_BÁSICA_por_pobl	0 nan de 4058	float64
18	educación Especial_por_pobl	0 nan de 4058	float64
19	educación Inicial_por_pobl	0 nan de 4058	float64
20	Media Media Superior_por_pobl	0 nan de 4058	float64
21	educación Superior_por_pobl	0 nan de 4058	float64
22	Abuso sexual infantil_por_pobl	0 nan de 4058	float64
23	Feminicidio_por_pobl	0 nan de 4058	float64
24	Homicidio doloso_por_pobl	0 nan de 4058	float64
25	Lesiones dolosas_por_pobl	0 nan de 4058	float64
26	Robo a bancos_por_pobl	0 nan de 4058	float64
27	Robo a carga pesada_por_pobl	0 nan de 4058	float64
28	Robo a cuentahabientes_por_pobl	0 nan de 4058	float64
29	Robo a int de vehiculos_por_pobl	0 nan de 4058	float64
30	Robo a negocio_por_pobl	0 nan de 4058	float64
31	Robo a persona_por_pobl	0 nan de 4058	float64
32	Robo a vehiculos particulares_por_pobl	4058 non-null	float64
33	Robo casa habitacion_por_pobl	0 nan de 4058	float64
34	Robo de autopartes_por_pobl	0 nan de 4058	float64
35	Robo de motocicleta_por_pobl	0 nan de 4058	float64
36	Violacion_por_pobl	0 nan de 4058	float64
37	Violencia familiar_por_pobl	0 nan de 4058	float64

Tabla 2.7: Información general del DataFrame.

Campo	std	min	mean	max	count	75 %	50 %	25 %
1	645.459	2.437	374.65	8049.114	4058	392.594	158.459	67.056
2	0.165	0	0.131	2.554	4058	0.168	0.079	0.028
3	2.023	0	1.86	54.18	4058	2.638	1.257	0.492
4	6.466	0	6.886	65.299	4058	9.97	4.904	2.086
5	4.893	0	5.101	65.117	4058	7.446	3.599	1.5
6	3.073	0	2.55	38.14	4058	3.58	1.497	0.515
7	0.511	0	0.066	21.481	4058	0.007	0	0
8	8.801	0	1.213	342.78	4058	0.169	0	0
9	0.034	0	0.005	1.263	4058	0.001	0	0
10	0	0	0	0.011	4058	0	0	0
11	0.005	0	0	0.252	4058	0	0	0
12	0.027	0	0.003	1.121	4058	0.001	0	0
13	0	0	0	0.025	4058	0	0	0
14	0.007	0	0.001	0.254	4058	0	0	0
15	0.023	0	0.002	1.01	4058	0	0	0
16	0.015	0	0.003	0.508	4058	0	0	0
17	0.026	0	0.004	1.009	4058	0	0	0
18	0.002	0	0	0.112	4058	0	0	0
19	0.003	0	0	0.168	4058	0	0	0
20	0.008	0	0	0.339	4058	0	0	0
21	0.003	0	0	0.189	4058	0	0	0
22	0.052	0	0.011	1.77	4058	0.011	0	0
23	0.002	0	0	0.147	4058	0	0	0
24	0.06	0	0.007	2.692	4058	0.004	0	0
25	0.191	0	0.037	8.679	4058	0.041	0.01	0
26	0.001	0	0	0.04	4058	0	0	0
27	0.015	0	0.001	0.63	4058	0	0	0
28	0.009	0	0.001	0.417	4058	0	0	0
29	0.059	0	0.01	2.137	4058	0.004	0	0
30	0.224	0	0.039	7.259	4058	0.02	0	0
31	0.238	0	0.039	9.067	4058	0.018	0	0
32	0.191	0	0.042	6.266	4058	0.023	0.001	0
33	0.102	0	0.027	3.749	4058	0.032	0.006	0
34	0.027	0	0.005	1.018	4058	0	0	0
35	0.093	0	0.014	3.895	4058	0.012	0	0
36	0.009	0	0.001	0.309	4058	0	0	0
37	0.222	0	0.057	6.973	4058	0.065	0.017	0

Tabla 2.8: Estadísticas iniciales de los campos.

Primeros hallazgos

Superficie territorial: Los valores de la superficie territorial varían significativamente, con un máximo muy por encima del tercer cuartil, lo que indica que hay áreas con una superficie territorial muy grande en comparación con la mayoría.

Variables de salud ("Sin acceso a servicios de salud por población"): Al igual que con la educación, hay un sesgo donde algunos lugares tienen acceso significativamente menor en comparación con otros.

Variables demográficas ("Población por km²"): Los valores sugieren variabilidad en la densidad de población, con algunas áreas mucho más densamente pobladas que otras.

Variables con mínimos de 0 y bajos promedios: Muchas de las variables tienen un mínimo de 0 y medias bajas, lo que sugiere que hay una gran cantidad de áreas sin incidentes reportados para esas variables o con valores muy bajos.

Distribución de los datos: Los valores de la mediana (%) más bajos que la media (mean) en la mayoría de las variables sugieren una distribución sesgada hacia la derecha para estas variables.

Diversidad de las métricas: La diversidad en las escalas de las métricas indica la necesidad de normalizar o estandarizar los datos antes de realizar análisis comparativos o de correlación.

2.6.3 Distribución de las Variables

En la siguiente sección, se llevará a cabo un análisis detallado de un par de variables seleccionadas. Este servirá como ejemplo ilustrativo del proceso sistemático de análisis que se ha aplicado individualmente a cada una de las variables del conjunto de datos

Varibale población por km²

Este muestra la distribución de la población por kilómetro cuadrado. Se observa una concentración de datos hacia el lado izquierdo del gráfico, indicando que muchas zonas tienen una baja densidad de población, con algunas excepciones que alcanzan valores más altos.

Con el gráfico de boxplot se puede observar la distribución de la población por kilómetro cuadrado. La caja central representa los cuartiles Q2 (mediana) y Q3, mientras que las líneas (bigotes) se extienden para mostrar el resto de la distribución, excepto por los valores atípicos, que son los puntos fuera de los bigotes. El boxplot confirma que hay una cantidad significativa de valores atípicos, lo que sugiere una variabilidad considerable en la densidad de población entre diferentes zonas.

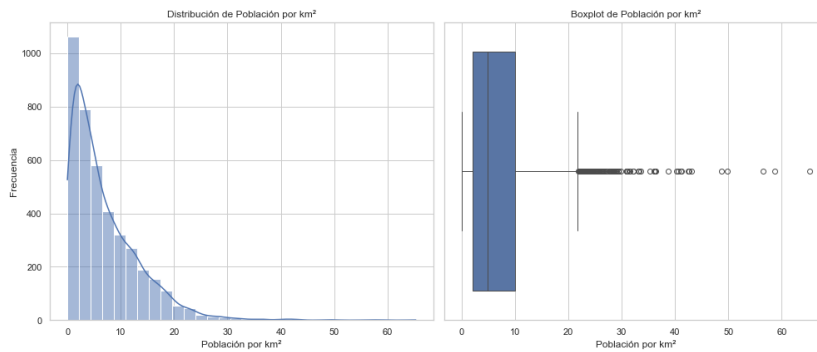


Figura 2.4: Estos gráficos representan la distribución de la densidad de población por km², destacando una concentración predominante en rangos bajos con algunos valores atípicos significativos que indican áreas de mayor densidad.

Una forma común de reducir el sesgo es aplicar una transformación logarítmica. Esta transformación es efectiva especialmente para datos que tienen una distribución sesgada hacia la derecha, como parece ser el caso de la variable "población por km²"

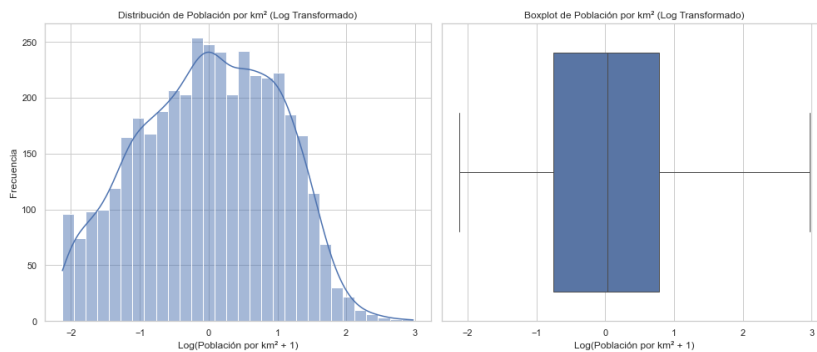


Figura 2.5: El histograma y el boxplot transformados proporcionan una perspectiva clara de la centralidad y dispersión de los datos, facilitando la interpretación estadística y la planificación basada en la densidad poblacional.

Mayor de 15 años sin educación en relación a la población total

Este muestra la distribución de la población sin educación. Se observa una concentración de datos hacia el lado izquierdo del gráfico, indicando que muchas zonas tienen una baja densidad de población, con algunas excepciones que alcanzan valores más altos.

Con el gráfico de boxplot se puede observar la distribución de la población sin educación. La caja central representa los cuartiles Q2 (mediana) y Q3, mientras que las líneas (bigotes) se extienden para mostrar el resto de la distribución, excepto por los valores atípicos, que son los puntos fuera de los bigotes. El boxplot confirma que hay una cantidad significativa de valores atípicos, lo que sugiere una variabilidad considerable en la densidad de población entre diferentes zonas.

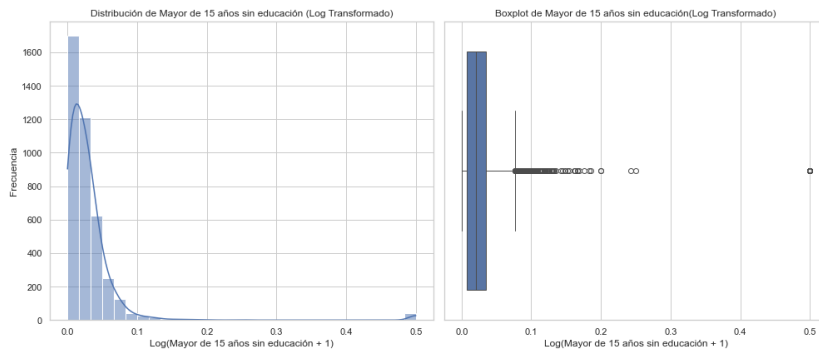


Figura 2.6: distribución y outliers, de la variable "Mayores de 15 años sin educación por población total"

Una forma común de reducir el sesgo es aplicar una transformación logarítmica. Esta transformación es efectiva especialmente para datos que tienen una distribución sesgada hacia la derecha, como parece ser el caso de la variable "población por km2"

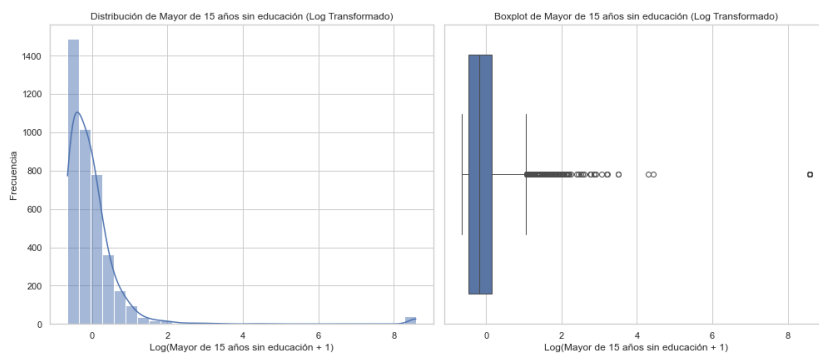


Figura 2.7: distribución después del preprocesamiento

Análisis visual de las variables:

Para las variables:

- "Sin acceso a servicios de salud"
- "población mayor de edad"
- "Mayores de edad con educación básica"

Se observa un sesgo significativo. Por lo tanto, se procede a aplicar la misma transformación a estas variables con el fin de estandarizar los valores y mitigar cualquier distorsión en el análisis subsiguiente

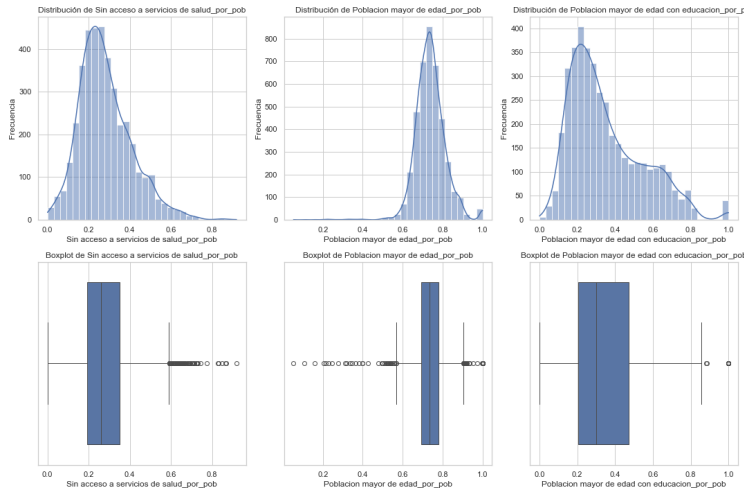


Figura 2.8: Distribución y outliers, de las variables Para las variables: Sin acceso a servicios de salud, población mayor de edad, Mayores de edad con educación básica

De nueva cuenta se aplica una transformación logarítmica, y también se aplica una estandarización de los datos.

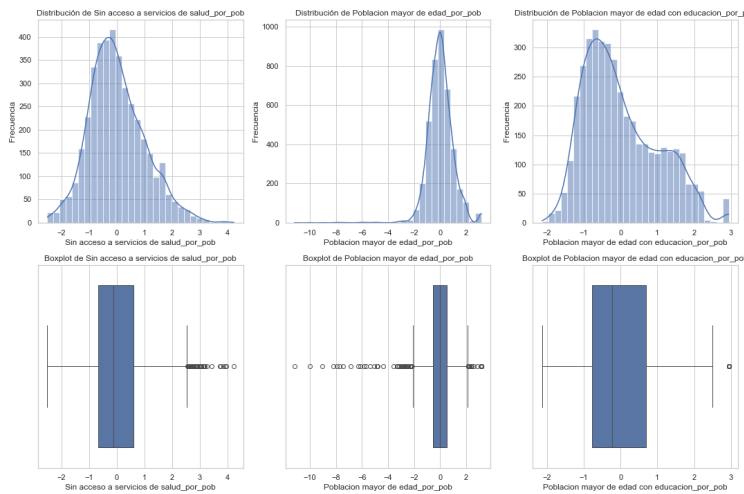


Figura 2.9: Post-procesamiento. distribución y outliers, de las variables Para las variables: Sin acceso a servicios de salud, población mayor de edad y Mayores de edad con educación básica

Análisis visual del resto de variables:

Para las variables:

- Cant. de docentes [2.11 - 2.11]
- Cant. de alumnos [2.11 - 2.11]
- Cant. de escuelas [2.11 - 2.11]
- Turno continuo [2.13 - 2.13]
- Turno discontinuo [2.13 - 2.13]
- Turno matutino [2.13 - 2.13]
- Turno nocturno [2.15 - 2.15]
- Turno vespertino [2.15 - 2.15]
- control privado [2.15 - 2.15]
- Institucion Publica [2.17 - 2.17]
- Tipo educativo básica [2.17 - 2.17]
- educación especial [2.17 - 2.17]
- educación Inicial [2.19 - 2.19]
- Media Media Superior [2.19 - 2.19]
- educación Superior [2.19 - 2.19]
- Abuso sexual infantil [2.21 - 2.21]
- Femicidio [2.21 - 2.21]
- Homicidio doloso [2.21 - 2.21]
- Lesiones dolosas [2.23 - 2.23]
- Robo a bancos [2.23 - 2.23]
- Robo a carga pesada [2.23 - 2.23]
- Robo a cuentahabientes [2.25 - 2.25]
- Robo a int de vehiculos [2.25 - 2.25]
- Robo a negocio [2.25 - 2.25]
- Robo a persona [2.27 - 2.27]
- Robo a vehiculos particulares [2.27 - 2.27]
- Robo casa habitacion [2.27 - 2.27]
- Robo de autopartes [2.29 - 2.29]
- Robo de motocicleta [2.29 - 2.29]
- Violacion [2.29 - 2.29]
- Violencia familiar [2.31 - 2.31]

Para el conjunto de datos analizado, se observa un sesgo en todas las variables, lo cual era anticipado. Esto se debe a que no se esperaba encontrar escuelas en todas las colonias, al igual que la presencia de delitos varía significativamente entre las distintas áreas.

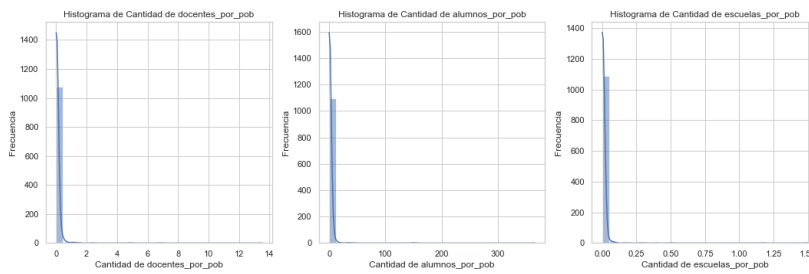


Figura 2.10: Estos histogramas representan la distribución de docentes, alumnos y escuelas por población, destacando una concentración significativa en valores bajos para cada variable. Esta visualización refleja que, en su mayoría, las áreas tienen una baja densidad de docentes, alumnos y escuelas en relación a su población.

Post transformación logarítmica y normalizar

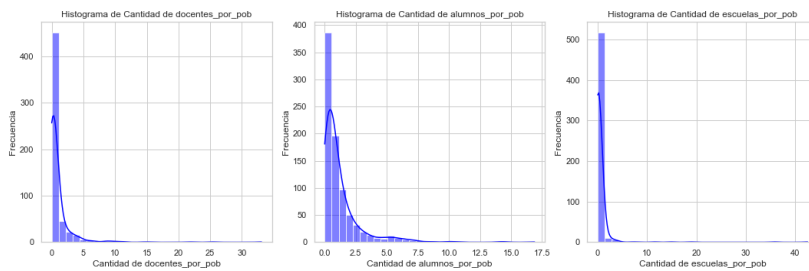


Figura 2.11: Post-procesamiento, las variables son mencionadas en cada gráfica.

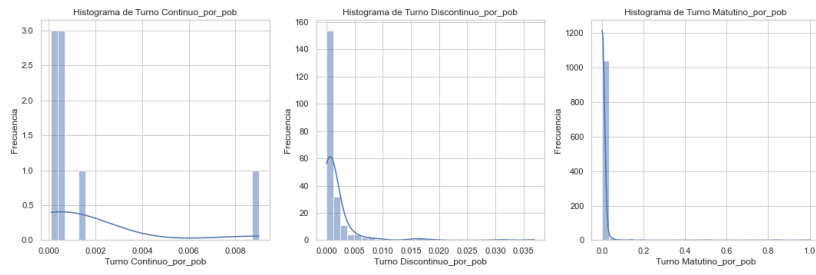


Figura 2.12: Estos histogramas muestran la frecuencia de turnos continuos, discontinuos y matutinos por población, evidenciando una predominancia de turnos matutinos en comparación con otros horarios.

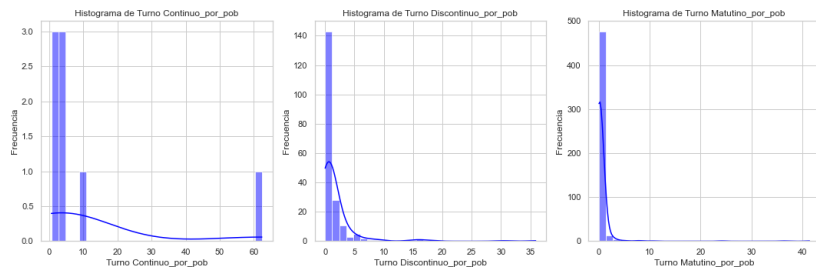


Figura 2.13: Post-procesamiento, las variables son mencionadas en cada gráfica.

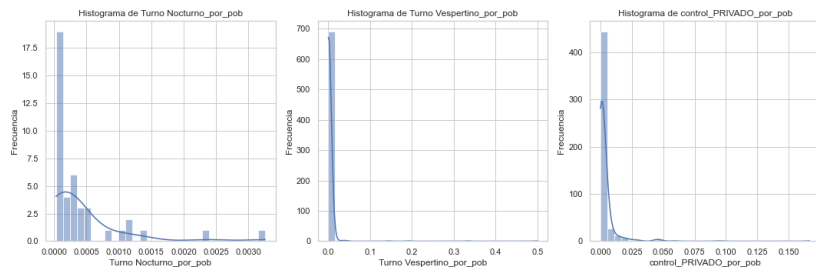


Figura 2.14: Estos histogramas muestran la distribución de turnos nocturnos, vespertinos y escuelas privadas por población. Los datos indican que los turnos nocturnos y las escuelas privadas son poco comunes, mientras que los turnos vespertinos, aunque algo más frecuentes, siguen siendo limitados. Estos patrones pueden informar políticas educativas y la asignación de recursos.

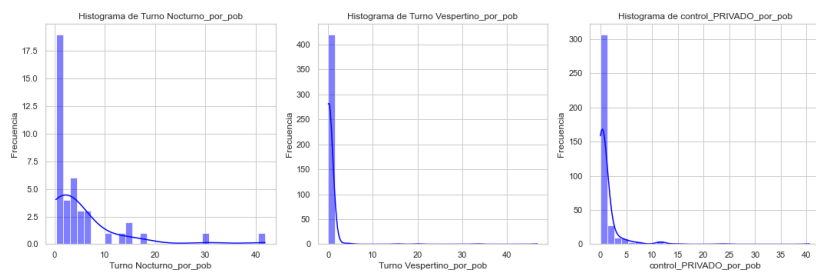


Figura 2.15: Post-procesamiento, las variables son mencionadas en cada gráfica.

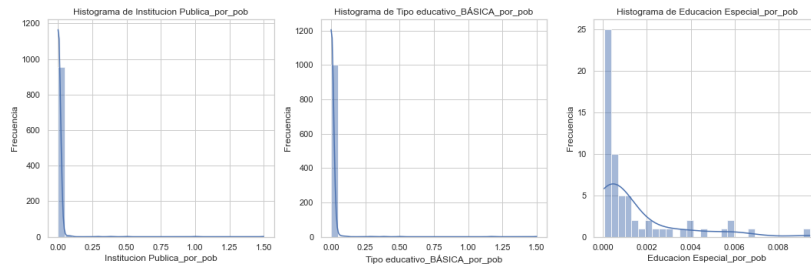


Figura 2.16: Estos histogramas visualizan la distribución por población de instituciones públicas, educación básica y educación especial. La mayoría de las áreas tienen una baja proporción de instituciones públicas y educación básica por población, mientras que la educación especial es aún menos común, con una distribución que indica una disponibilidad muy limitada en la mayoría de las áreas.

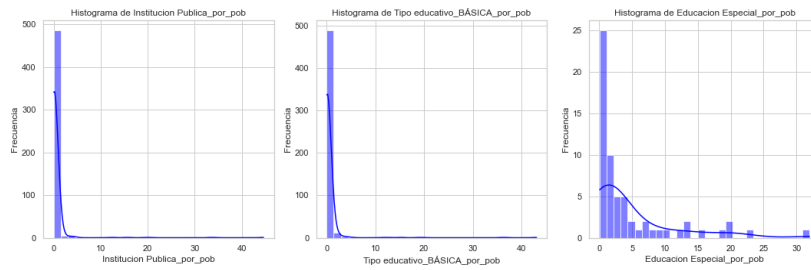


Figura 2.17: Post-procesamiento, las variables son mencionadas en cada gráfica.

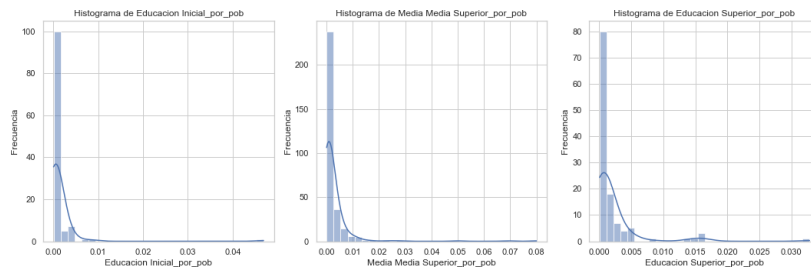


Figura 2.18: Estos histogramas muestran la distribución per cápita de instituciones de educación inicial, media superior y superior. La mayoría de las áreas tienen una baja proporción de estas instituciones en relación con su población, destacando una disponibilidad particularmente limitada para la educación superior. Esto subraya la necesidad de ampliar el acceso a niveles educativos más avanzados en muchas regiones.

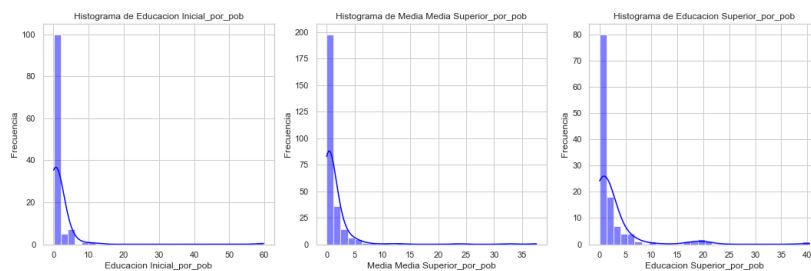


Figura 2.19: Post-procesamiento, las variables son mencionadas en cada gráfica.

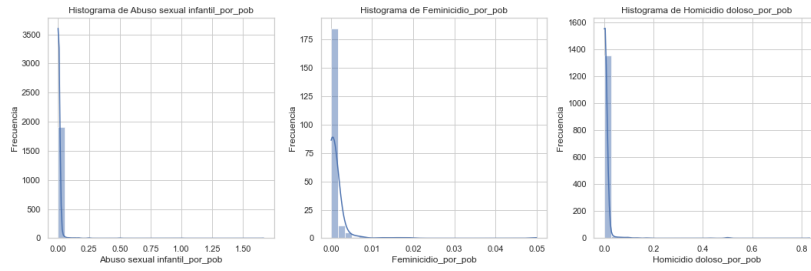


Figura 2.20: Estos histogramas representan la frecuencia de abuso sexual infantil, feminicidios y homicidios dolosos por población. Todos los gráficos muestran una concentración alta de casos cerca del cero, indicando que, aunque los incidentes son relativamente raros en términos per cápita, su presencia es preocupante y varía ampliamente entre diferentes áreas. Esto subraya la importancia de políticas enfocadas en la prevención y la justicia para estos delitos graves.

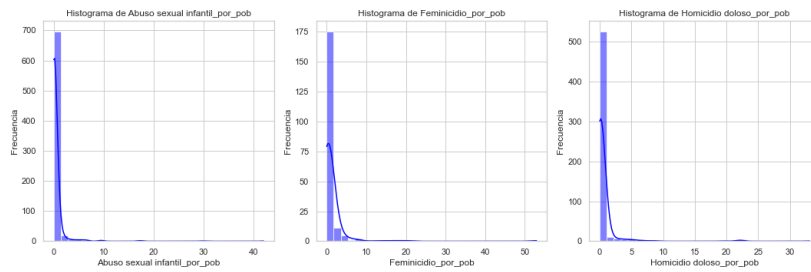


Figura 2.21: Post-procesamiento, las variables son mencionadas en cada gráfica.

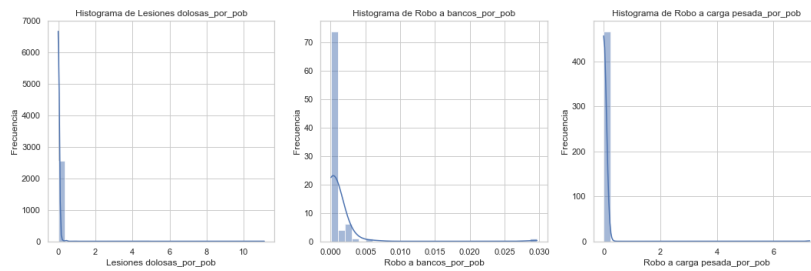


Figura 2.22: Estos histogramas ilustran la distribución per cápita de lesiones dolosas, robos a bancos y robos a carga pesada. Los datos muestran una concentración predominante de frecuencias cerca de cero para todos los delitos, indicando que estos eventos son infrecuentes en relación a la población total. Sin embargo, la presencia de estos delitos, aunque baja, subraya la necesidad de medidas de seguridad y prevención específicas para cada tipo de crimen.



Figura 2.23: Post-procesamiento, las variables son mencionadas en cada gráfica.



Figura 2.24: Estos histogramas representan la frecuencia de robos a cuentahabientes, robos a vehículos y robos a negocios por población. Todos los gráficos muestran que la mayoría de las áreas tienen una baja incidencia de estos delitos, con una concentración predominante de casos cercana a cero. Esta visualización es fundamental para identificar regiones con necesidades específicas de seguridad y para la planificación de recursos policiales y preventivos.

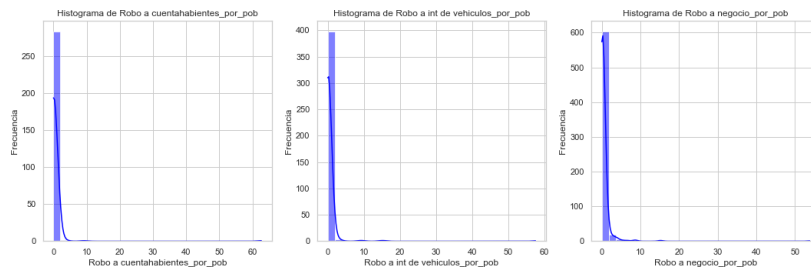


Figura 2.25: Post-procesamiento, las variables son mencionadas en cada gráfica.

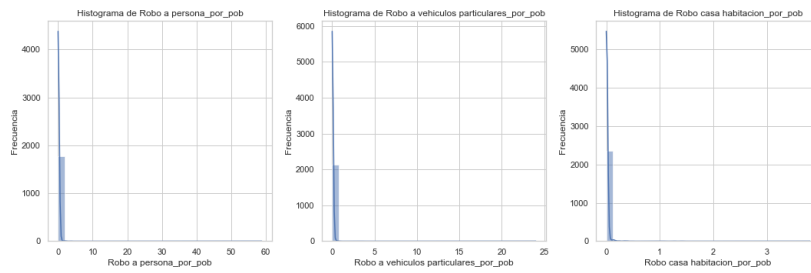


Figura 2.26: Estos histogramas visualizan la distribución de robos a persona, robos a vehículos particulares y robos en casas habitación por población. En cada caso, la concentración de eventos es predominantemente baja, con la mayoría de las áreas mostrando valores cercanos a cero, lo que indica una baja frecuencia de estos delitos en relación a la población total. Este análisis subraya la variabilidad regional en la incidencia de robos y puede guiar las estrategias de prevención y seguridad local.

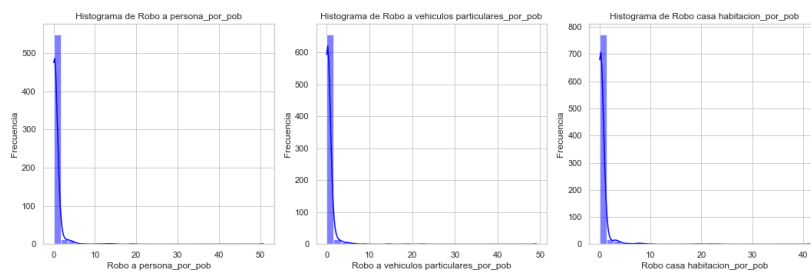


Figura 2.27: Post-procesamiento, las variables son mencionadas en cada gráfica.

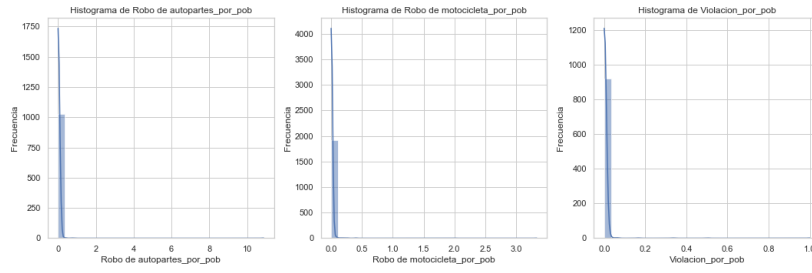


Figura 2.28: Estos histogramas muestran la distribución de robos de autopartes, robos de motocicletas y violaciones por población. Cada histograma refleja una baja frecuencia general de estos delitos en la mayoría de las áreas, con la mayor parte de los datos agrupados cerca del cero. Esto indica que, aunque los incidentes ocurren, su prevalencia por población es relativamente baja en la mayoría de las regiones estudiadas.

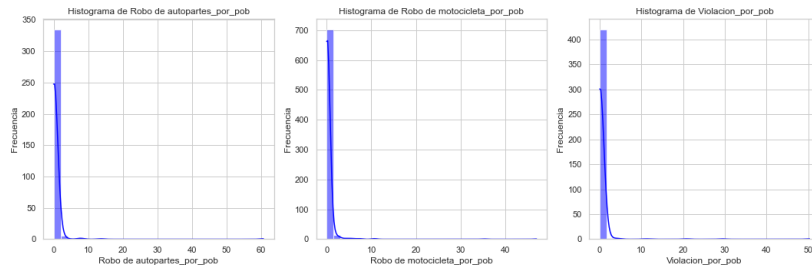


Figura 2.29: Post-procesamiento, las variables son mencionadas en cada gráfica.

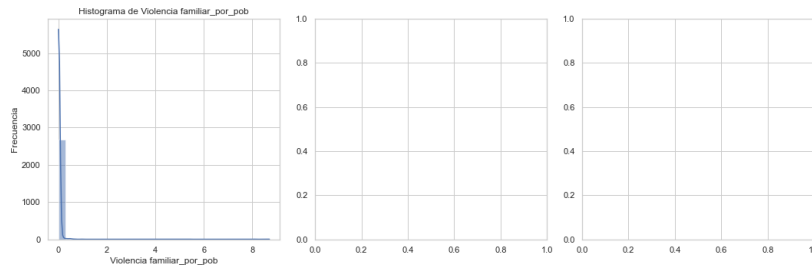


Figura 2.30: Este histograma muestra la distribución de casos de violencia familiar por población. La gran mayoría de los datos están concentrados cerca de cero, indicando que, en la mayoría de las áreas, la frecuencia de violencia familiar es relativamente baja en relación con la población total.

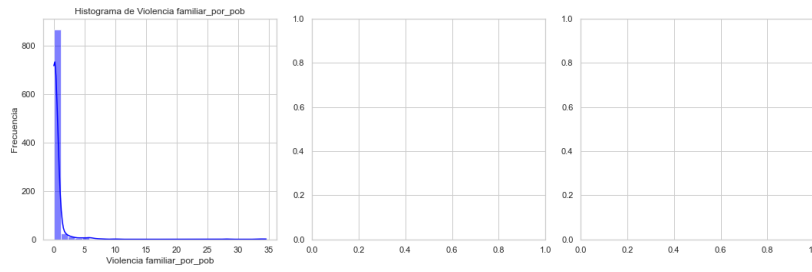


Figura 2.31: Post-procesamiento, las variables son mencionadas en cada gráfica.

2.6.4 *Análisis PCA*

El siguiente paso consiste en realizar un análisis de correlación. El objetivo principal de este procedimiento es identificar variables que contribuyan mínimamente a la varianza explicada de los datos. Además, se busca determinar la posibilidad de crear nuevas variables que permitan explicar el modelo de manera eficiente, pero con un menor número de variables. Este enfoque tiene como finalidad optimizar el modelo, manteniendo o mejorando su capacidad explicativa mediante la reducción de la complejidad y el número de variables involucradas."

A continuación, se presenta el mapa de calor resultante de la matriz de correlación, que incluye todas las variables involucradas en el análisis. Este mapa de calor revela varios hechos interesantes y patrones significativos, los cuales se describen detalladamente a continuación:

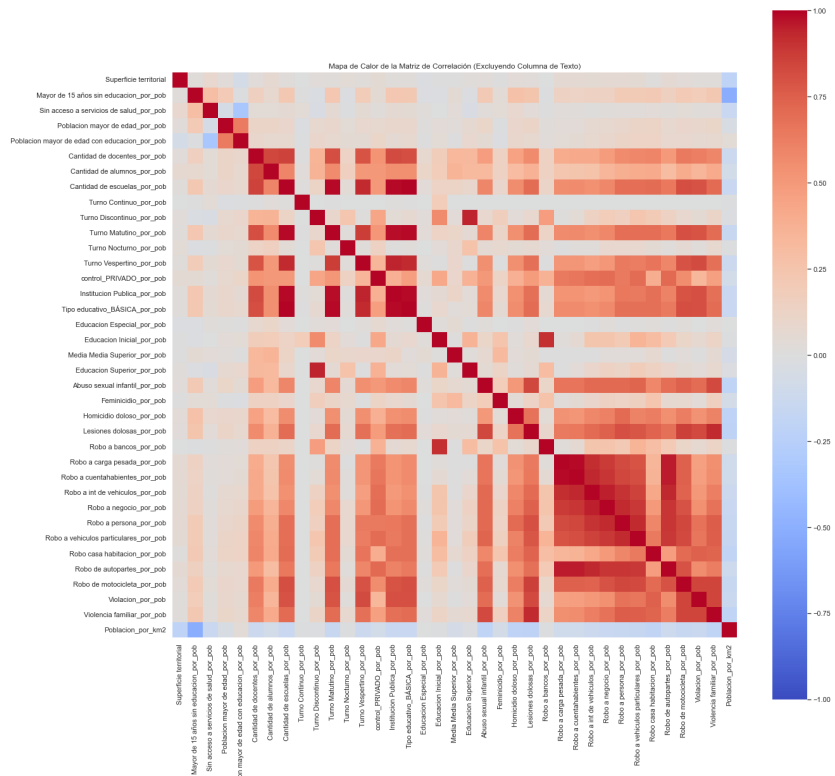


Figura 2.32: Este mapa de calor de la matriz de correlación visualiza las relaciones entre diferentes indicadores socioeconómicos y de criminalidad por población. Los tonos rojos indican una correlación positiva fuerte, mientras que los tonos azules representan una correlación negativa. Este análisis ayuda a identificar patrones y posibles interdependencias entre variables como educación, salud, y tipos de delito, proporcionando una base para decisiones de política y priorización de recursos.

Primeras observaciones:

Correlaciones Positivas Fuertes: Hay varias variables con alta correlación positiva entre sí, particularmente las que están relacionadas con la población y educación, como 'Población', 'Población mayor de edad' y 'Población mayor de edad con educación'. Esto podría indicar que las áreas con mayor densidad de población tienden a tener también

una mayor proporción de población adulta y población adulta con educación, lo cual es lógico desde un punto de vista demográfico y social.

Variables de Educación: Las variables relacionadas con instituciones educativas, como 'Cantidad de docentes', 'Cantidad de alumnos' y 'Cantidad de escuelas', también muestran correlaciones positivas, lo que sugiere que estas métricas tienden a aumentar juntas. Es probable que en áreas con más escuelas haya también más docentes y alumnos.

Variables de Delitos: Las variables relacionadas con delitos parecen tener correlaciones variables entre sí. Esto podría reflejar la complejidad y la multicausalidad del fenómeno del crimen, en el que diferentes tipos de delitos no necesariamente ocurren con la misma frecuencia en las mismas áreas.

Posible Redundancia de Datos: La existencia de correlaciones altas sugiere que algunas de estas variables pueden ser redundantes. Por ejemplo, si dos variables están altamente correlacionadas.

Pasos a seguir

Creación de Nuevas Variables: Dado que algunas variables están altamente correlacionadas, podrían aplicarse técnicas de reducción de dimensiones como el Análisis de Componentes Principales (PCA) para simplificar el modelo sin perder capacidad explicativa.

Ejecución PCA

En el análisis del dataset, se implementó una rutina para resaltar variables con correlaciones fuertes (mayores de 0.7) pero inferiores a la perfección (menores de 0.95). Este método nos ayuda a identificar relaciones intensas entre las variables sin caer en la redundancia de una correlación absoluta.

```

1 # Se obtienen booleanos para correlaciones fuertes
2 correlaciones_fuertes = matriz_correlacin.abs() > 0.7
3 # Se obtienen booleanos para correlaciones que no son perfectas
4 correlaciones_no_perfectas = matriz_correlacin.abs() < 0.95
5

```

Listing 2.1: Código para encontrar correlaciones fuertes y no perfectas

El análisis arrojó que las siguientes variables tenían una correlación muy fuerte con otras variables.

Nota: Ninguna se descartó con las justificaciones siguientes, pero si se realizara una reducción de dimensionalidad con PCA.

Si bien son pocos registros, pero es esperado ya que no hay muchas escuelas especiales en el estado, pero aún juegan un papel importante. Así mismo las escuelas de tiempo completo, si bien cada vez son menos estas también son relevantes para el estudio.

educación Especial

Turno Continuo

Los siguientes delitos son de los que menos eventos se tienen pero es importante considerar todos los delitos aunque estos parezcan poco relevantes.

Robo a bancos

Robo a cuentahabientes

PCA - Variables de violencia

Se tomará las 16 variables relacionadas con la violencia y posterior al análisis PCA podemos observar que mas del 90% de la varianza se puede explicar con 6 variables. figuras: (2.33, 2.34)

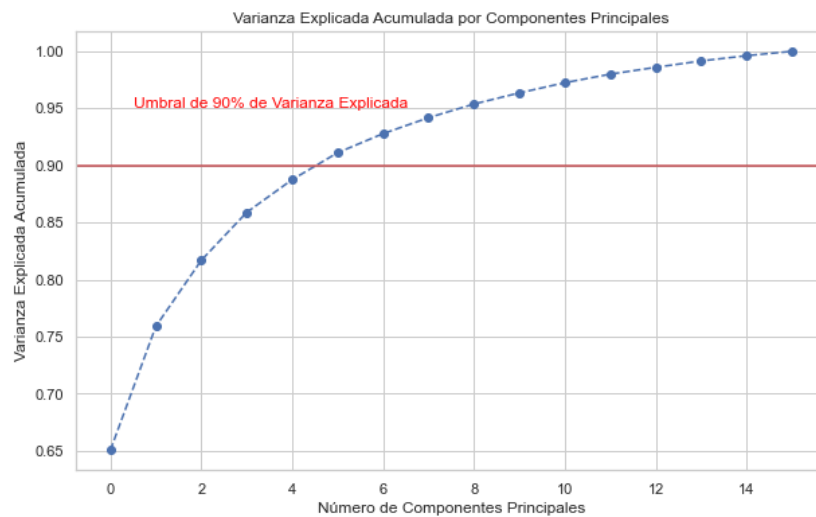


Figura 2.33: Tomando 6 variables podremos alcanzar más del 90% de la varianza acumulada.

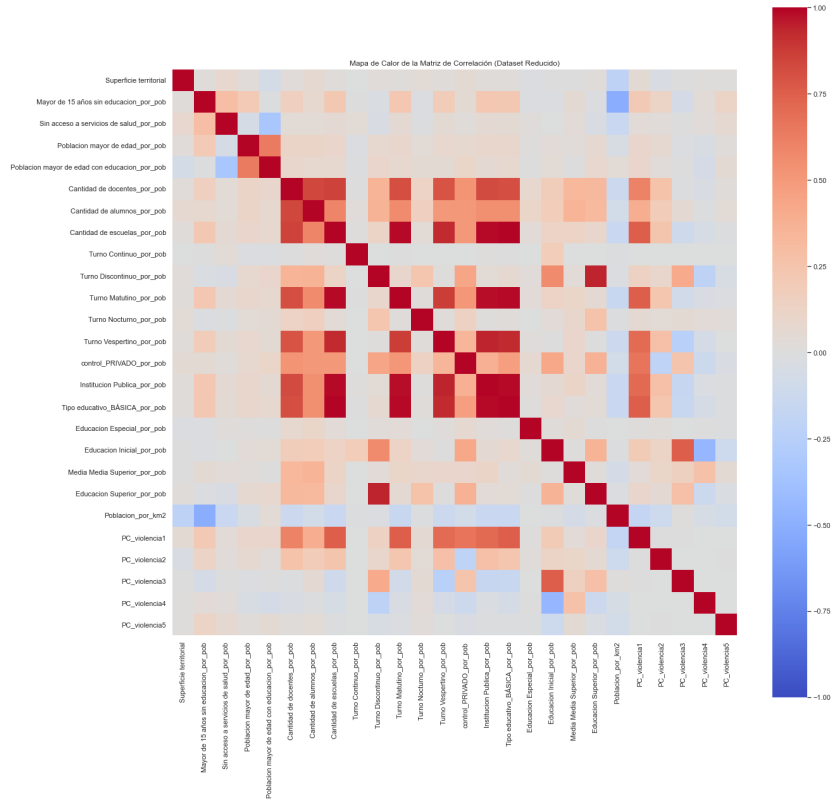


Figura 2.34: Mapa de calor que representa la matriz de correlación con la reducción de variables.

PCA - Variables de población

Se tomará las 4 variables relacionadas con la población y posterior al análisis PCA podemos observar que mas del 90% de la varianza se puede explicar con 3 variables. figuras: (2.35, 2.37)

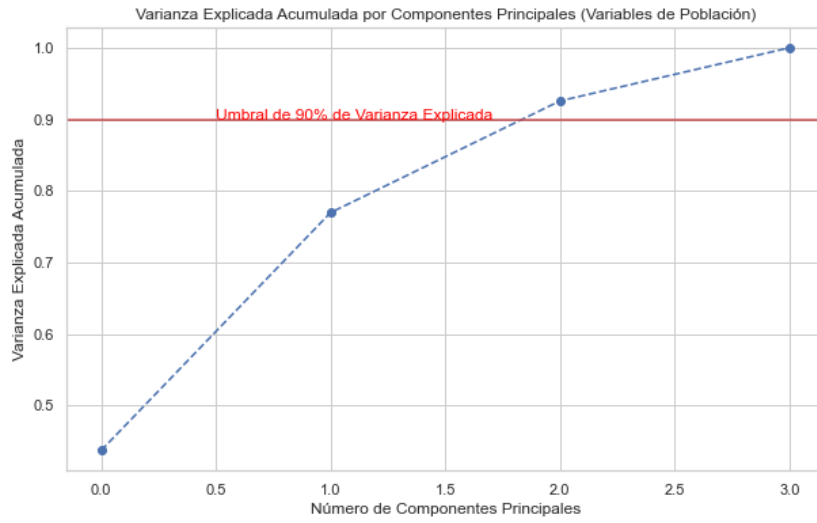


Figura 2.35: Tomando 3 variable podremos alcanzar mas del 95% de la varianza acumulada.

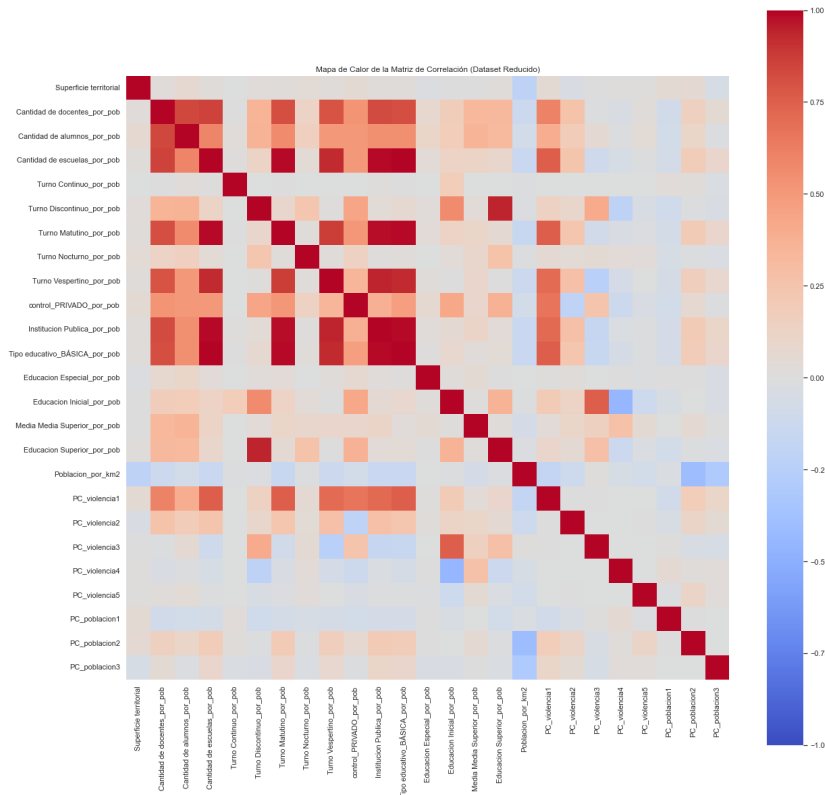


Figura 2.36: Mapa de calor que representa la matriz de correlación con la reducción de variables.

2.7 Descripción de los modelos

En esta sección, vamos a procesar la información que ya hemos preparado. Como primer paso, realizaremos una clasificación no supervisada. Nuestro objetivo es conseguir una clasificación que equilibre las variables de población, educación y violencia.

Nota: Es importante destacar que no estamos buscando crear un ranking de las zonas de más a menos violentas. Más bien, esperamos que el modelo nos ofrezca claridad sobre qué zonas necesitan atención prioritaria debido a sus características específicas

K-means

El algoritmo K-means es un método de agrupamiento que divide un conjunto de datos en k grupos o clusters. En este proceso, los datos se agrupan de tal manera que los puntos en el mismo clúster son más similares entre sí que los puntos en otros clusters.[29]

Un cluster, o agrupamiento, es un conjunto de objetos, puntos o datos que son similares entre sí y, al mismo tiempo, distintos de los objetos en otros grupos. En el contexto del análisis de datos, un cluster puede verse como una colección de datos que comparten propiedades comunes. El agrupamiento ayuda a identificar patrones subyacentes y estructuras en un conjunto de datos, facilitando la comprensión y la organización de la información de manera más eficiente. Los clusters son fundamentales en diversas aplicaciones como la segmentación de mercado, la organización de documentos, la clasificación de imágenes y muchas otras áreas donde se busca extraer información valiosa a partir de grandes volúmenes de datos.

¿Cómo funciona el algoritmo K-means?

Para utilizar el algoritmo K-means, primero se especifica el número de clusters deseados (k). Por ejemplo, al establecer k igual a 2, su conjunto de datos se agrupará en 2 grupos, mientras que si establece k igual a 4, agrupará los datos en 4 grupos.

Cada grupo está representado por su centro o centroide, que corresponde a la media aritmética de los puntos de datos asignados al grupo. De esta manera, el algoritmo funciona a través de un proceso iterativo hasta que cada punto de datos está más cerca del centroide de su propio grupo que de los centroides de otros grupos, minimizando la distancia dentro del grupo en cada paso.

regresión lineal

La regresión lineal es un método estadístico utilizado para modelar la relación entre una variable dependiente y una o más variables independientes. En su forma más simple, se aplica a una sola variable independiente y se llama regresión lineal simple. La idea central es encontrar una línea recta que mejor se ajuste a los datos observados, de modo que podamos predecir los valores futuros de la variable dependiente basándonos en los valores pasados de la variable independiente.[30] [31]

Tipos de Regresión Lineal:

Regresión Lineal Simple: Se utiliza cuando hay una sola variable independiente. El objetivo es examinar cómo influye esa variable en la variable dependiente.

Regresión Lineal Múltiple: Se emplea cuando hay varias variables independientes. Aquí se analiza la influencia de todas esas variables en la variable dependiente.

Modelo Matemático:

La ecuación de la regresión lineal simple es:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- y es la variable dependiente.
- x es la variable independiente.
- β_0 es el punto de intersección con el eje y .
- β_1 es la pendiente de la recta.
- ε representa el término de error (la variabilidad no explicada por la relación entre x e y).

Método de los Mínimos Cuadrados:

La regresión lineal busca la línea recta que minimiza la suma de los cuadrados de las diferencias entre los valores observados y los predichos. Visualmente, se representa como una línea en un diagrama de dispersión.

Interpretación:

Cuanto mayor sea la relación lineal entre las variables, más precisa será la predicción. La recta de regresión describe cómo cambia y con respecto a x .

GBM - Gradient Boosting Machine

Gradient Boosting Machines *GBMs* son una técnica de aprendizaje automático ampliamente utilizada y altamente efectiva. Estos modelos pertenecen a la familia de algoritmos de ensemble learning. [32][33]

Los GBMs combinan múltiples modelos de aprendizaje débiles, generalmente árboles de decisión, para crear un modelo predictivo más preciso y robusto.

Utilizan el método de boosting, que implica agregar árboles uno tras otro para corregir los errores de predicción de modelos anteriores.

Son ampliamente utilizados en problemas de clasificación y regresión en datos tabulares.

Como función?:

Inicialmente, se entrena un modelo débil en los datos, por ejemplo, un árbol de decisión.

Luego, se calculan los errores residuales entre las predicciones del modelo débil y los valores reales.

En cada iteración, se ajusta un nuevo modelo débil para predecir los errores residuales.

Los modelos débiles se combinan mediante ponderación para formar un modelo más fuerte.

GBM con GridSearch

Un grid search, es una técnica utilizada para encontrar los mejores parámetros para un modelo en el aprendizaje automático. Básicamente, consiste en probar sistemáticamente múltiples combinaciones de parámetros para determinar cuál de estas configuraciones da los mejores resultados según un criterio de evaluación específico. Esta técnica es muy útil cuando se trabaja con modelos que tienen varios hiperparámetros que pueden ser ajustados, como es el caso de las Máquinas de Boosting Gradiente (GBM). [34]

Como función?:

Definición del espacio de parámetros: Primero, se define un "grid" cuadrícula de parámetros. Esto incluye todas las combinaciones posibles de los valores de los parámetros que deseas probar. Por ejemplo, en un modelo GBM, podrías querer ajustar la tasa de aprendizaje, el número de árboles y la profundidad máxima de cada árbol.

Entrenamiento del modelo: Luego, se entrena un modelo para cada combinación de parámetros en el grid. Cada modelo se evalúa utilizando un método de validación, como la validación cruzada, para obtener una estimación de su rendimiento.

Evaluación del rendimiento: Se evalúa el rendimiento de cada modelo usando una métrica específica, como puede ser el error cuadrático medio (MSE) para tareas de regresión o la precisión para tareas de clasificación.

Selección del mejor modelo: Finalmente, se selecciona la combinación de parámetros que dio lugar al mejor rendimiento del modelo según la métrica de evaluación elegida.

MLP Regression - Redes Neuronales

La Regresión de MLP es una técnica de aprendizaje automático utilizada para predecir valores numéricos continuos. Se basa en una arquitectura de red neuronal artificial con múltiples capas interconectadas.[35]

Como función?:

Arquitectura: Un MLP consta de capas, incluyendo una capa de entrada, una o más capas ocultas y una capa de salida.

Neuronas: Cada neurona en las capas ocultas aplica una función de activación a la suma ponderada de sus entradas.

Aprendizaje: Durante el entrenamiento, el modelo ajusta los pesos para minimizar el error entre las predicciones y los valores reales.

evaluación Evalúa el modelo utilizando métricas como el error cuadrático medio *MSE* o el coeficiente de determinación *R-cuadrado*.

métricas para clasificación

Silhouette Score:

Para este estudio, emplearemos la métrica de Silhouette como indicador de la calidad de los clusters identificados. La elección de esta métrica se justifica porque ofrece una medida clara de hasta qué punto cada objeto del dataset se asemeja más a su propio cluster en comparación con otros clusters. Un alto puntaje de Silhouette sugiere que los objetos están bien emparejados a sus propios clusters y mal emparejados a los vecinos, lo cual es un indicador de clustering adecuadamente definido y separado. Además, esta métrica es particularmente útil cuando las etiquetas verdaderas no están disponibles, como es nuestro caso en un contexto de aprendizaje no supervisado.[36]

La métrica de Silhouette se calcula utilizando dos valores para cada objeto en el dataset: la cohesión y la separación.

La cohesión a es la distancia media entre un objeto y todos los otros objetos en el mismo cluster, indicando cuán cercanamente está agrupado el objeto a su propio cluster.

Por otro lado, la separación b es la distancia media mínima de un objeto a todos los objetos en el cluster más cercano al cual el objeto no pertenece, reflejando cuán bien separado está el objeto de los clusters vecinos.

El coeficiente de Silhouette para un objeto específico se calcula con la siguiente fórmula[37]:

$$s = \frac{b - a}{\max(a, b)}$$

Donde:

a es la media de las distancias entre el objeto y todos los demás objetos en su propio cluster.

b es la menor media de las distancias entre el objeto y todos los objetos en un cluster diferente (el más cercano).

El valor de s varía entre -1 y 1 , donde un valor alto indica que el objeto está bien emparejado a su propio cluster y diferenciado de otros clusters. Un valor de s cercano a 1 significa que el objeto está bien colocado dentro de su cluster y lejos de otros clusters, un valor de s cercano a 0 indica que el objeto está en la frontera entre dos clusters, y un valor de s cercano a -1 sugiere que el objeto estaría mejor situado en otro cluster.

Esta métrica permite evaluar la calidad de los clusters de una manera objetiva y es especialmente útil en escenarios de aprendizaje no supervisado donde las etiquetas verdaderas de los datos no están disponibles.[38]

Determinar el número óptimo de clusters:

Método del codo

El método del codo determina el número óptimo de clusters en un análisis de agrupamiento observando el punto donde aumentar el número de clusters deja de reducir significativamente la variación dentro de los clusters

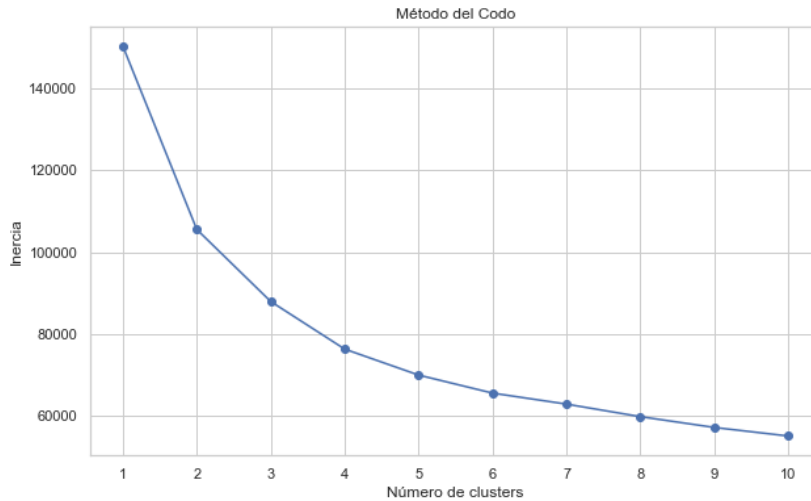


Figura 2.37: Gráfica del método del codo mostrando la inercia por número de clusters para determinar el número óptimo. El punto de inflexión alrededor de 4 a 5 clusters sugiere el equilibrio ideal entre el número de clusters y la compacidad de cada uno.

En esta gráfica, no se ve un punto donde la curva hace un doblez claro, pero parece que después del 4 o 5, la línea se va poniendo más plana. Eso significa que si seguimos poniendo más grupos, no estamos ganando mucho en términos de agrupar mejor los datos. Así que, quizás escoger 4 o 5 grupos estaría bien para este caso.

Prueba de silhouette score

Dado que la gráfica del codo no marca una flexión muy marcada, pero nos podemos inclinar por 4 o 5 clusters, se hará la prueba de silhouette score, con 4 y 5 clusters.

```

1 from sklearn.metrics import silhouette_score
2 # Aplicar K-Means con 4 y 5 clusters y calcular el Silhouette Score para
   cada uno
3 for k in [4, 5]:
4     kmeans = KMeans(n_clusters=k, random_state=42)
5     cluster_labels = kmeans.fit_predict(df_normalized)
6     # Calcular el Silhouette Score
7     silhouette_avg = silhouette_score(df_normalized, cluster_labels)
8     results[k] = {
9         'Inertia': kmeans.inertia_,
10        'Silhouette Score': silhouette_avg
11    }
12

```

Listing 2.2: evaluación Silhouette

Resultados:

Con 4 clusters: Silhouette Score de 0.7359.

Con 5 clusters: Silhouette Score de 0.7741.

Dado que el objetivo es maximizar la similitud dentro de los clusters y minimizar la similitud entre clusters diferentes, un silhouette score de 0.7741 con 5 clusters es preferible.

métricas para regresión

R-cuadrado (Coeficiente de determinación): Coeficiente de determinación, cuyo símbolo es R^2 (R cuadrado), es un estadístico que mide la bondad de ajuste de un modelo de regresión. El coeficiente de determinación muestra lo bien que se ajusta un modelo de regresión a un conjunto de datos, es decir, indica el porcentaje explicado por el modelo de regresión.[?].

Por lo tanto, cuanto mayor sea el coeficiente de determinación, mejor será el modelo de regresión. Aunque no siempre se cumple esta condición, en principio nos interesa un coeficiente de determinación lo más grande posible. Más abajo veremos cómo interpretar el coeficiente de determinación.

Así pues, la fórmula para calcular el coeficiente de determinación es la siguiente:

$$R^2 = 1 - \frac{\sigma_r^2}{\sigma^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Raíz del Error Cuadrático Medio (RMSE): La Raíz del Error Cuadrático Medio o RMSE (Root Mean Squared Error) es una medida de desempeño cuantitativa utilizada comúnmente para evaluar métodos de pronóstico de demanda. En este contexto RMSE consiste en la raíz cuadrada de la sumatoria de los errores cuadráticos. En comparación con la Desviación Media Absoluta o MAD, RMSE amplifica y penaliza con mayor fuerza aquellos errores de mayor magnitud.[?]

La fórmula de cálculo del RMSE se muestra a continuación:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (A_t - F_t)^2}$$

En conclusión, las dos métricas seleccionadas son clave para determinar

la eficiencia del modelo. El coeficiente de determinación R cuadrado nos proporcionará una medida de cuánta variabilidad de los datos es capturada por el modelo, mientras que el RMSE nos dará una estimación de la desviación promedio de las predicciones del modelo respecto a los valores reales. Juntas, estas métricas nos ofrecen una visión integral de la performance del modelo.

2.9 Descripción de los experimentos o simulaciones

Clasificación - K-means

La estrategia que se seguirá consiste en ejecutar el modelo de k-means, un método de clasificación no supervisada. Después de realizar un análisis mediante el coeficiente de Silhouette, se ha determinado que la configuración óptima del modelo incluye 5 clusters. El objetivo principal de aplicar k-means con esta cantidad de grupos es descubrir una clasificación que refleje y capture las posibles variaciones existentes entre las variables demográficas, las características de las instituciones educativas y los índices de violencia. Esto nos permitirá entender mejor la estructura subyacente de los datos y posiblemente revelar patrones significativos que de otro modo podrían pasar desapercibidos.

Post - Clasificación

gráfico de dispersión

Un gráfico de dispersión para las dos y tres primeras componentes principales después de aplicar PCA, con cada punto coloreado según su cluster para visualizar cómo se separan los clusters en un espacio de menor dimensión.

Nota: El PCA permite representar los datos en gráficas 2D o 3D. Esta representación conserva las estructuras esenciales del conjunto de datos original, facilitando así su interpretación visual

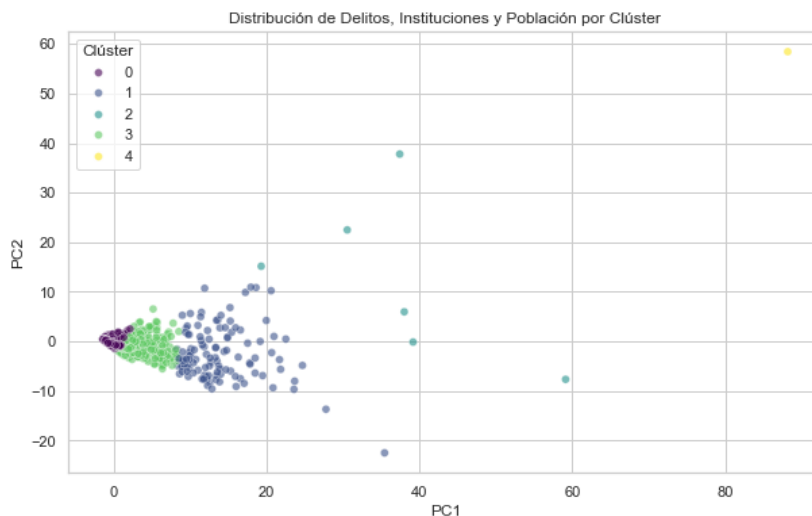


Figura 2.38: Representación en componentes principales (PC₁ y PC₂) de la distribución de delitos, instituciones y población por cluster. Cada color representa un cluster diferente, destacando la variabilidad y agrupación de los datos en el espacio de las dos principales componentes.

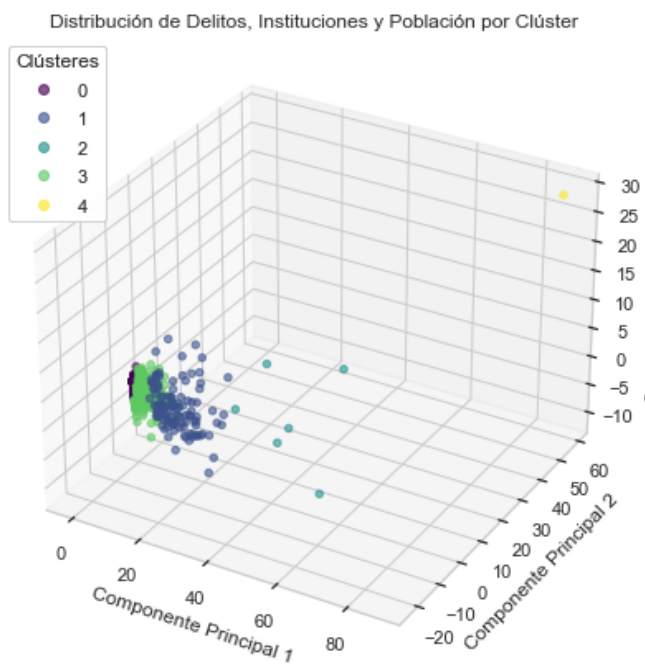


Figura 2.39: Visualización 3D utilizando las tres principales componentes para mostrar la distribución de delitos, instituciones y población por cluster. Cada color identifica un cluster distinto, revelando cómo se diferencian estos grupos en un espacio tridimensional más complejo.

Gráfica de barras Promedio de los distintos indicadores de cada cluster.

Indicadores Violencia

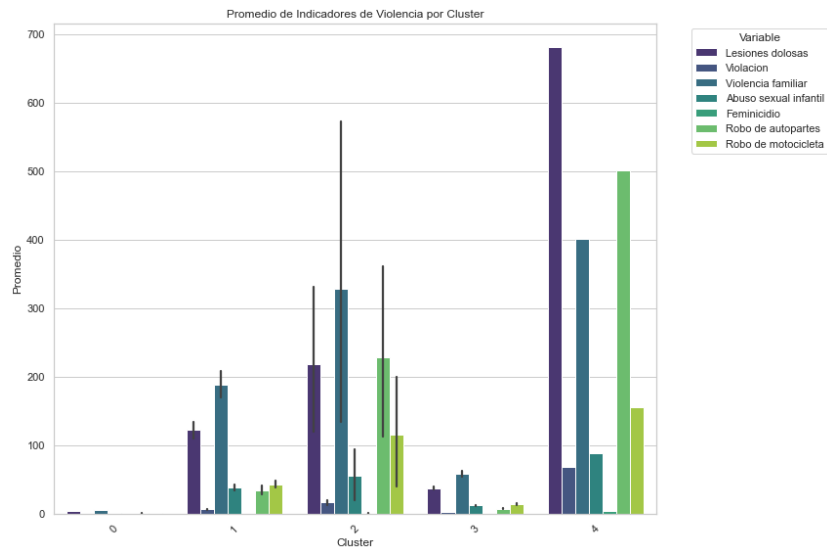


Figura 2.40: gráfica correspondiente a los indicadores de violencia

Cada barra corresponde a un cluster diferente, y los colores de las barras representan distintos tipos de indicadores de violencia, como lesiones dolosas, violación, violencia familiar, abuso sexual infantil, feminicidio, robo de autopartes y robo de motocicleta.

Nota: La interpretación de cada cluster se describe en la sección de resultados.

Indicadores educación y población

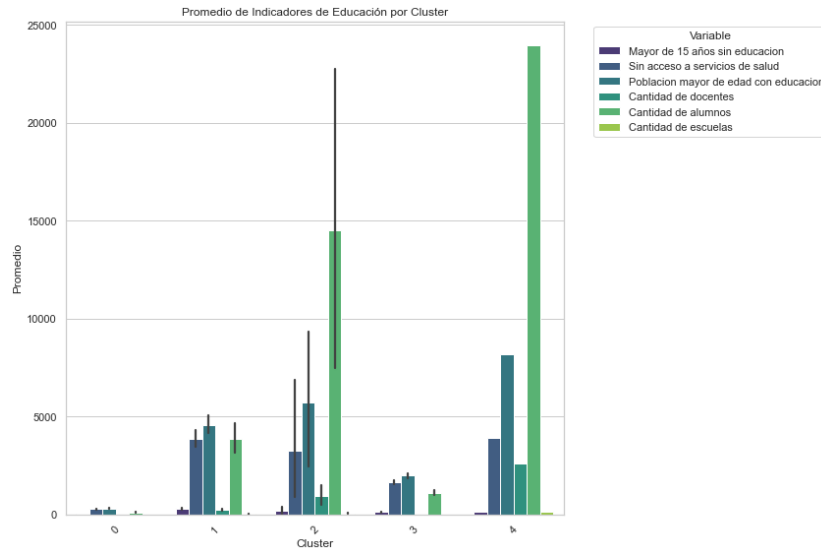


Figura 2.41: gráfica correspondiente a los indicadores de población y educación

Cada barra representa un indicador de educación distinto, como el número de personas mayores de 15 años sin educación, la población sin acceso a servicios de salud, la población mayor de edad con educación, la cantidad de docentes, alumnos y escuelas.

Regresión [Regresión lineal - GBM - MLP]

En esta sección, se llevarán a cabo experimentos utilizando tres modelos de regresión.

Como se indicó previamente en el documento, los delitos a analizar son los siguientes:

- Abuso sexual infantil
- Feminicidio
- Homicidio doloso
- Violación
- Violencia familiar
- Feminicidio

Recordando que con el paso del tiempo, se ha observado un incremento constante en la incidencia de estos delitos específicos en el estado de Jalisco.

A continuación, se presentará un conjunto de dos gráficas para cada delito, ilustrando su comportamiento en la predicción realizada por los modelos. Asimismo, se incluirán las métricas de R cuadrado (coeficiente de determinación) y RMSE (Error Cuadrático Medio Raíz) para proporcionar una evaluación cuantitativa de la precisión de las predicciones.

Abuso sexual infantil

gráficas

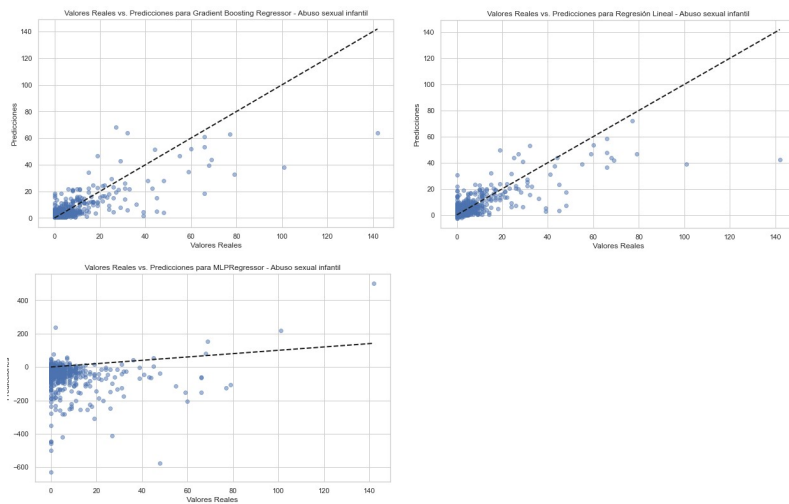


Figura 2.42: Comparativa de la precisión en la predicción de casos de abuso sexual infantil utilizando tres modelos de regresión diferentes: Gradient Boosting Regressor, Regresión Lineal y MLPRegressor. Cada gráfico muestra la relación entre los valores predichos por el modelo y los valores reales, donde la línea discontinua representa la predicción perfecta.

Homicidio Doloso

gráficas

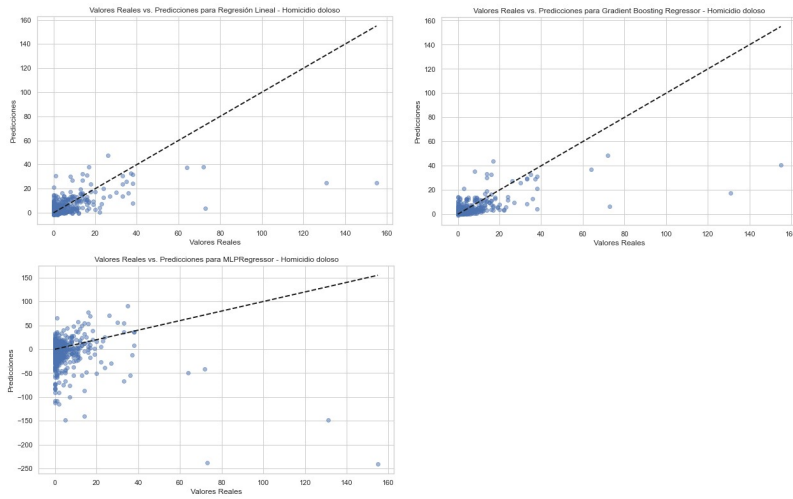


Figura 2.43: Comparativa de la precisión en la predicción de casos de Homicidio doloso utilizando tres modelos de regresión diferentes: Gradient Boosting Regressor, Regresión Lineal y MLPRegressor.

Feminicidio

gráficas

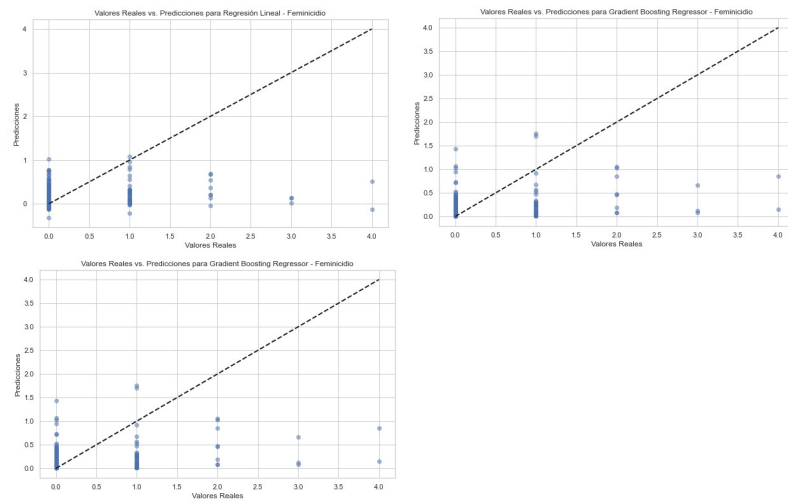


Figura 2.44: Comparativa de la precisión en la predicción de casos de Feminicidio utilizando tres modelos de regresión diferentes: Gradient Boosting Regressor, Regresión Lineal y MLPRegressor.

Violacion

gráficas

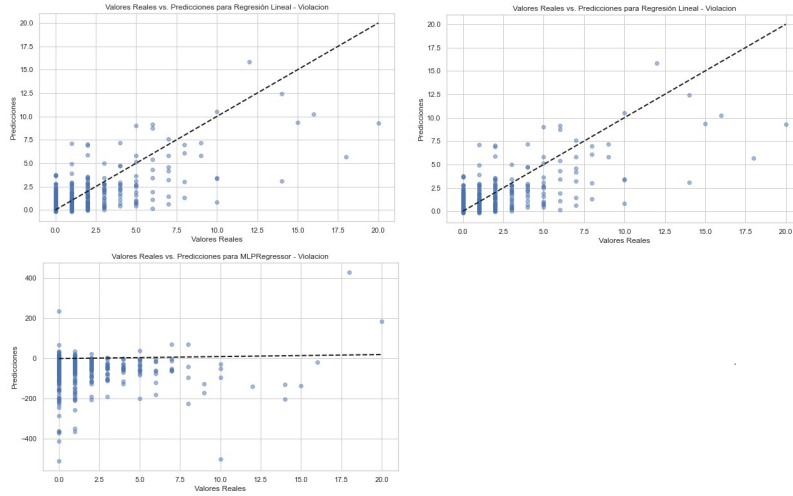


Figura 2.45: Comparativa de la precisión en la predicción de casos de abuso violacion utilizando tres modelos de regresión diferentes: Gradient Boosting Regressor, Regresión Lineal y MLPRegressor.

Violencia familiar

gráficas

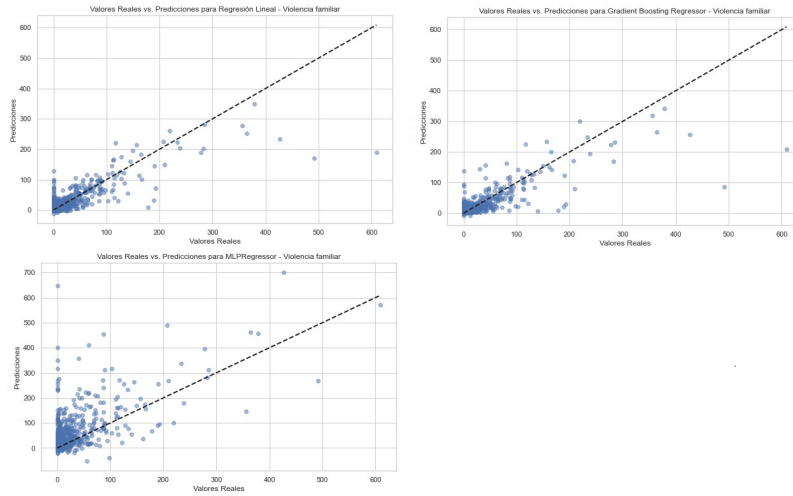


Figura 2.46: Comparativa de la precisión en la predicción de casos de violencia familiar utilizando tres modelos de regresión diferentes: Gradient Boosting Regressor, Regresión Lineal y MLPRegressor.

3 Resultados y discusión

En este capítulo se presentan los resultados obtenidos del desarrollo de este trabajo y una discusión sobre el objeto de estudio.

3.1 Resultados y discusión

Interpretación de cada uno de los clasificadores

Cluster 0: ver [2.40 - 3.2]

Delitos: Significativamente más bajos en el cluster 0, comparado con el promedio total.

Escuelas: También más bajo en el cluster 0, frente al promedio total.

Población: Menor en el cluster 0 que en el promedio general.

Población sin educación: En el cluster 0 es considerablemente más baja comparada con el promedio total.

Violaciones: Son más bajas en el cluster 0 en comparación con el promedio total.

Sin educación: Menor en el cluster 0 frente al promedio total.

Robo casa habitación: Muy bajo en el cluster 0.

En resumen, el cluster 0 se caracteriza por tener menores tasas en todas las categorías comparadas con los promedios generales del conjunto de datos. Esto podría indicar que las áreas agrupadas en el cluster 0 tienden a ser más seguras, con menos delitos y violaciones, y una población con mejor acceso a la educación.

Cluster 1: ver [2.40 - 3.2]

Educación: Muestra un aumento considerable en la cantidad de docentes y alumnos comparado con el Cluster 0, lo que sugiere una mayor inversión en educación o una mayor densidad poblacional. Sin embargo, también muestra un mayor número de personas mayores de 15 años sin educación.

Violencia: Los niveles de violencia, especialmente en términos de violencia familiar, lesiones dolosas y abuso sexual infantil, son significativamente más altos que en el Cluster 0. Esto podría indicar áreas con problemas sociales más complejos y una necesidad de políticas de intervención específicas.

Cluster 2 ver [2.40 - 3.2]

Educación: Similar al Cluster 1 en términos de indicadores educativos, aunque con algunas diferencias notables en población educada y acceso a servicios de salud, sugiriendo variaciones en el acceso a educación y servicios básicos.

Violencia: Este cluster muestra niveles aún más altos de ciertos tipos de violencia, como lesiones dolosas y violación, lo que sugiere una prevalencia significativa de crimen violento y posiblemente áreas urbanas con desafíos de seguridad más pronunciados.

Cluster 3 ver [2.40 - 3.2]

Educación: Exhibe un perfil educativo mixto, con una cantidad de docentes y alumnos que sugiere un nivel intermedio de recursos o accesibilidad educativa. La población mayor de edad con educación es relativamente alta, lo que indica un acceso razonable a la educación para adultos.

Violencia: Los niveles de violencia son moderados comparados con otros clusters, con una incidencia de violencia familiar y lesiones dolosas que sugiere preocupaciones de seguridad pero en una escala menos extrema que en los Clusters 1 y 2.

Cluster 4 ver [2.40 - 3.2]

Educación: Destaca por tener la mayor cantidad de docentes y alumnos, así como la mayor cantidad de escuelas, indicando áreas con una fuerte inversión en educación, posiblemente zonas urbanas con una densa red de instituciones educativas.

Violencia: Muestra los niveles más altos de violencia en varias categorías, incluyendo violencia familiar y feminicidio, lo que refleja áreas con graves problemas de seguridad. A pesar de los altos niveles de recursos educativos, estos problemas de violencia presentan desafíos significativos para la comunidad.

Interpretación General ver [2.38 - 2.39] Los clusters varían significativamente tanto en términos de indicadores de educación como de violencia. Los Clusters 0 y 3 podrían representar áreas con mejor seguridad relativa y diferentes niveles de acceso a la educación, mientras que los Clusters 1, 2 y 4 sugieren áreas con mayores desafíos tanto en seguridad como en educación, posiblemente reflejando diferencias entre áreas urbanas y rurales, o entre comunidades con distintos niveles de recursos y problemas sociales. Este análisis nos ayuda a comprender cómo las condiciones de educación y violencia pueden variar entre diferentes áreas, subrayando la importancia de enfoques específicos y dirigidos para abordar estos desafíos.

Resumen por Cluster(Municipios):

Cluster 0: 107 municipios representados.

Cluster 1: 13 municipios representados.

Cluster 2: 2 municipios representados.

Cluster 3: 44 municipios representados.

Cluster 4: 1 municipio representado.

Resumen por cluster(Colonias):

Cluster 0:

Zapopan: 470 colonias

Tlajomulco de Zúñiga: 236 colonias

Tonalá: 231 colonias

Puerto Vallarta: 196 colonias

Guadalajara: 164 colonias

Cluster 1:

Guadalajara: 48 colonias

Zapopan: 29 colonias

Tlajomulco de Zúñiga: 14 colonias

Tonalá: 9 colonias

Tlaquepaque: 7 colonias

Cluster 2:

Guadalajara: 5 colonias

Tlaquepaque: 1 colonia

Cluster 3:

Guadalajara: 120 colonias

Zapopan: 81 colonias

Tlaquepaque: 42 colonias

Tlajomulco de Zúñiga: 38 colonias

Tonalá: 34 colonias

Cluster 4:

Guadalajara: 1 colonia

Regresión [Regresión lineal - GBM - MLP]

En esta sección, se presentan los resultados de los experimentos utilizando dos modelos de regresión.

Abuso sexual infantil

Resultados

Regresión Lineal: Con un RMSE de 5.91 y un R^2 de 0.63, este modelo tiene un ajuste razonable, explicando alrededor del 63 % de la variabilidad de los datos.

Gradient Boosting Machine (GBM) con GridSearchCV: Muestra un desempeño similar a la regresión lineal con un RMSE ligeramente mejor y el mismo R^2 . Los parámetros optimizados indican que el modelo ha sido afinado para mejorar la precisión.

Red Neuronal MLP: Tiene un rendimiento muy pobre con un RMSE extremadamente alto y un R^2 negativo, indicando que las predicciones del modelo son imprecisas y peores que simplemente tomar el promedio de los datos.

Delito	Modelo	RMSE	R^2	Mejores Parámetros	Mejor Score (RMSE)
Abuso sexual infantil	Regresión Lineal	5.91	0.63	N/A	N/A
Abuso sexual infantil	GBM con GridSearchCV	5.85	0.63	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	5.78082
Abuso sexual infantil	Red Neuronal MLP	70.33	-51.85	N/A	N/A

Tabla 3.1: Resultados de los modelos para abuso sexual infantil.

Homicidio Doloso

Resultados

Regresión Lineal: Con un RMSE de 6.44 y un R^2 de 0.37, este modelo tiene un ajuste moderado, explicando cerca del 37% de la variabilidad de los datos.

Gradient Boosting Machine (GBM) con GridSearchCV: Presenta un RMSE de 6.11, que es ligeramente mejor que la regresión lineal, y un R^2 de 0.44, indicando que puede explicar alrededor del 44% de la variabilidad en los datos. Los parámetros optimizados muestran un esfuerzo por refinar el modelo.

Red Neuronal MLP: Tiene un RMSE de 25.90, el cual resulta un valor más alto comparado con el resto de modelos, y un R^2 de -9.11, lo cual es extremadamente bajo y sugiere que el modelo es inadecuado, dando predicciones muy imprecisas.

Delito	Modelo	RMSE	R^2	Mejores Parámetros	Mejor Score (RMSE)
Homicidio doloso	Regresión Lineal	6.44	0.37	N/A	N/A
Homicidio doloso	GBM con GridSearchCV	6.11	0.44	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	5.76797
Homicidio doloso	Red Neuronal MLP	25.90	-9.11	N/A	N/A

Tabla 3.2: Resultados de los modelos para homicidio doloso.

Feminicidio

Resultados

Regresión Lineal: Presenta un RMSE bajo (0.32) pero un R^2 muy bajo (0.06), indicando que solo explica el 6% de la variabilidad del delito de feminicidio, lo que puede considerarse un ajuste débil.

GBM con GridSearchCV: Tiene un RMSE ligeramente mejor que la regresión lineal (0.31) y un R^2 algo mayor (0.14), mostrando una mejora en la capacidad del modelo para explicar la variabilidad de los datos. Los parámetros optimizados proporcionan la mejor precisión alcanzada en la validación cruzada.

Red Neuronal MLP: Con un RMSE extremadamente alto (31.41) y un R^2 negativo grande (-9053.98), este modelo muestra un rendimiento muy pobre, sugiriendo que es completamente inadecuado para modelar este tipo de delito.

Delito	Modelo	RMSE	R^2	Mejores Parámetros	Mejor Score (RMSE)
Feminicidio	Regresión Lineal	0.32	0.06	N/A	N/A
Feminicidio	GBM con GridSearchCV	0.31	0.14	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.30779
Feminicidio	Red Neuronal MLP	31.41	-9053.98	N/A	N/A

Tabla 3.3: Resultados de los modelos para feminicidio.

Nota: Las gráficas sugieren que la mayoría de las entidades observadas tienen cero o muy pocos casos de feminicidio, con una marcada disminución

en la frecuencia a medida que aumenta el número de casos. En cuanto a la relación entre la población mayor de edad con educación y los feminicidios, no parece haber una correlación clara basada en el gráfico de dispersión; los feminicidios ocurren en distintos niveles de población educada mayor de edad sin mostrar un patrón evidente.

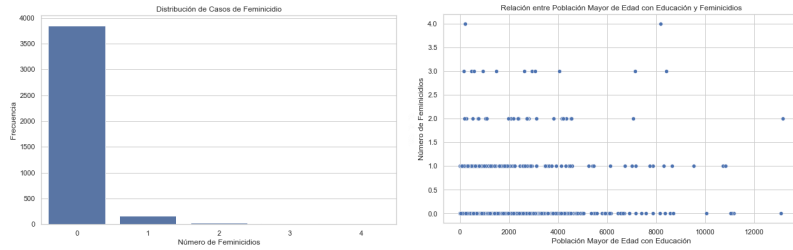


Figura 3.1: Femicidio - Frecuencia de los casos de feminicidio.

Violacion

Resultados

Regresión Lineal: Tiene un ajuste moderado con un RMSE de 1.19 y explica un 55% de la variabilidad del delito, lo cual es razonablemente bueno para un modelo sencillo.

GBM con GridSearchCV: Muestra un ligero mejoramiento en comparación con la regresión lineal, con un RMSE de 1.15 y un R² de 0.58, indicando que puede predecir los casos de violación con una precisión ligeramente superior.

Red Neuronal MLP: Exhibe un rendimiento muy deficiente con un RMSE extremadamente alto y un R² negativo muy grande, lo que sugiere que el modelo no es adecuado para predecir este tipo de delito o que necesita una reconfiguración significativa.

Delito	Modelo	RMSE	R ²	Mejores Parámetros	Mejor Score (RMSE)
Violación	Regresión Lineal	1.19	0.55	N/A	N/A
Violación	GBM con GridSearchCV	1.15	0.58	{'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100}	1.54587
Violación	Red Neuronal MLP	58.66	-1077.38	N/A	N/A

Tabla 3.4: Resultados de los modelos para el delito de violación.

Nota: Las gráficas sugieren que la mayoría de las entidades observadas tienen cero o muy pocos casos de Violaciones, no es que sea una cantidad insignificante pero no los casos suficientes para encontrar una relación lineal.

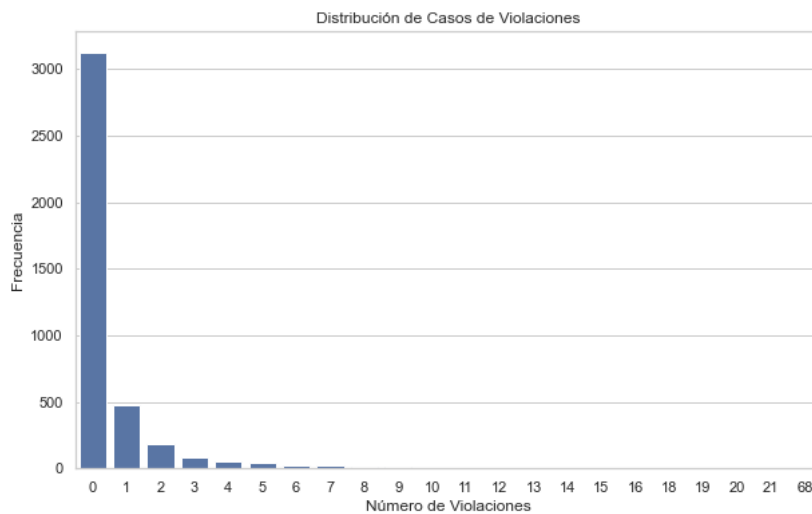


Figura 3.2: Femicidio - Frecuencia de los casos de Violaciones.

Violencia familiar

Resultados

Regresión Lineal: Con un RMSE de 24.36 y un R^2 de 0.70, este modelo muestra que puede explicar un 70% de la variabilidad en los datos, lo cual indica un buen ajuste.

GBM con GridSearchCV: A pesar de tener un RMSE ligeramente superior (25.48), el R^2 es algo menor (0.67), indicando un ajuste levemente inferior al modelo de regresión lineal. Los parámetros seleccionados muestran un intento de optimización del modelo.

Red Neuronal MLP: Con un RMSE muy alto (54.77) y un R^2 negativo (-0.53), este modelo no predice bien el delito de violencia familiar y probablemente necesite ser reevaluado o ajustado.

Delito	Modelo	RMSE	R^2	Mejores Parámetros	Mejor Score (RMSE)
Violencia familiar	Regresión Lineal	24.36	0.70	N/A	N/A
Violencia familiar	GBM con GridSearchCV	25.48	0.67	{'learning_rate': 0.2, 'max_depth': 3, 'n_estimators': 100}	24.59205
Violencia familiar	Red Neuronal MLP	54.77	-0.53	N/A	N/A

Tabla 3.5: Resultados de los modelos para el delito de violencia familiar.

4 Conclusiones y trabajo futuro

4.1 Conclusiones

Este estudio cumplió con su objetivo general de analizar y determinar la relación entre el número de instituciones educativas y la incidencia delictiva en Jalisco a nivel de colonia. A partir de los datos recolectados y depurados, se establecieron métricas claras y categorías que permitieron clasificar la incidencia delictiva y los planteles educativos, conforme a los objetivos específicos propuestos.

Las técnicas de ciencia de datos revelaron la dificultad de identificar patrones que apoyen la hipótesis planteada, pero esto puede no resultar tan sorprendente ya que se tiene la contra hipótesis de que los delincuentes tienden a no cometer delitos en sus propias comunidades, desafiando la suposición de que el incremento de instituciones educativas está correlacionado directamente con la disminución de la delincuencia.

Pero es importante rescatar que se identificaron zonas con densidades poblacionales similares, pero con diferencias significativas en la cantidad de planteles educativos. Esto sugiere que la presencia de instituciones educativas por sí sola no es suficiente para prevenir la delincuencia, y destaca áreas con potenciales carencias educativas que podrían estar generando delincuentes. Por otro lado, algunas regiones con alta presencia educativa y baja proporción de adultos sin educación mostraron altos índices de delitos. Esto indica que otros factores, además de la educación, contribuyen a la criminalidad, sugiriendo la necesidad de estrategias de seguridad más robustas que aborden las causas subyacentes de la delincuencia.

Es crucial resaltar que el estudio permitió identificar delitos de alta gravedad como violencia familiar, violaciones, feminicidios y abuso infantil, que no han visto una reducción a pesar de las estrategias implementadas en años pasados. Estos delitos, que tienen un gran impacto en la sociedad, requieren de un enfoque más especializado. La persistencia de estos delitos a altos niveles indica la necesidad de desarrollar estrategias más efectivas y específicas que aborden estas formas graves de criminalidad.

En consecuencia, se recomienda que las futuras políticas públicas y las inversiones educativas no solo se enfoquen en la cobertura sino también en la calidad y en programas de intervención directa que tengan en cuenta las complejidades de cada tipo de delito.

A pesar de que mejorar las políticas educativas es fundamental, es importante reconocer que el tema de la violencia es considerablemente más complejo y abarca múltiples dimensiones. Además de la educación, existen otros factores que pueden tener un impacto significativo en la regeneración del tejido social. La creación de más espacios como canchas deportivas y bibliotecas, así como centros de entrenamiento técnico y escuelas integrales, son cruciales. Asimismo, las actividades recreativas como talleres de arte, programas de mentoría, y clubes de ciencia y tecnología, pueden fomentar un ambiente comunitario más fuerte y seguro. La implementación de programas de desarrollo comunitario y espacios de encuentro social, como cafés comunitarios y jardines públicos, también pueden jugar un papel vital en este proceso de renovación. Todas estas iniciativas, al complementarse, tienen el potencial de fortalecer el tejido social y crear comunidades más resilientes y armónicas.

4.2 Trabajo futuro

Este estudio sienta las bases para explorar nuevas preguntas y desafíos futuros. Uno de los avances más significativos sería la inclusión de datos concretos sobre poblaciones delictivas, específicamente, información sobre criminales condenados. Cruzar este estudio con tal información es crucial, ya que podría proporcionarnos insights más precisos sobre qué comunidades abordar con intervenciones educativas. La tendencia observada en este trabajo sugiere que, a menudo, los delincuentes no cometen delitos dentro de sus propias comunidades. Por lo tanto, implementar estrategias educativas basadas solo en la densidad poblacional o en tasas de criminalidad podría no ser efectivo. Se hace imprescindible identificar aquellos barrios donde la falta de servicios educativos y de salud contribuyen a la delincuencia, para así desarrollar programas educativos que aborden el problema desde su origen.

Un segundo enfoque derivado de este trabajo es la necesidad de determinar los factores que causan la violencia en ciertas áreas. Aunque el estudio indica una alta presencia de instituciones educativas en zonas de elevada criminalidad, lo que sugiere que la educación por sí sola no es la solución, es imperativo identificar las características que hacen que un área sea propensa a la violencia. Con esta información, podríamos diseñar estrategias de seguridad más enfocadas y efectivas.

4.3 Trabajo relacionados

EL EFECTO DE LA EDUCACIÓN EN LA TASA DE CRIMINALIDAD: UN PANEL DE DATOS PARA LOS DEPARTAMENTOS DE LA SIERRA DEL PERÚ 2012 AL 2017 [39]

Resumen extraído del documento:

La presente investigación usa datos de paneles de la tasa de criminalidad para un grupo de departamentos del Perú por el periodo de 2012 al 2017, basados en los datos del Instituto Nacional de Estadística e Informática, los reportes anuales de criminalidad de la Policía Nacional del Perú y el Ministerio de Salud del Perú, para analizar la relación de la educación y la tasa de criminalidad tanto entre departamentos como a través del tiempo; así mismo, se utilizó a la victimización como tasa de criminalidad. La

metodología utilizada es Mínimos Cuadrados Generalizados Factibles para solucionar los problemas de heterocedasticidad. Los resultados muestran principalmente que el nivel de educación alcanzado influye negativamente en la tasa de criminalidad.

Modelo predictivo aplicado a las Estadísticas Policiales del Poder Judicial de Costa Rica para el análisis de la criminalidad por regiones [40]

Resumen extraído del documento:

Por medio del análisis descriptivo realizado a los Datos Abiertos del Poder Judicial de Costa Rica, se reflejan números alarmantes en cuanto a la cantidad de denuncias impuestas en el Organismo de Investigación Judicial, superando las cincuenta mil denuncias en el año 2019. Con estos números en mente, el objetivo de esta investigación es generar un modelo de análisis de datos que permita potencializar estas estadísticas y señalar anticipadamente las regiones con mayor propensión de sufrir delitos en el próximo quinquenio, con el fin de promover la proactividad, tanto del ciudadano como de la policía para estar alerta y evitar los próximos delitos. Se aprovechan los modelos de predicción estadística mediante métodos matemáticos comprobados y aplicables a los datos obtenidos y su comportamiento durante los años 2015-2019. El análisis realizado refleja la necesidad de aplicar el algoritmo de regresión lineal simple el cual es aplicado a la solución desarrollada que está disponible para todos los costarricenses en el sitio Tableau Public. Los resultados muestran predicciones negativas para el país, especialmente en el Gran Área Metropolitana, el comportamiento de los delitos impactará significativamente esta zona, lo cual indica la necesidad de establecer programas de fortalecimiento policial, mejoras en la educación y el empleo para contrarrestar los potenciales delitos proyectados. Palabras clave – predicción de delitos, patrullaje preventivo, estadísticas policiales, lucha contra el crimen.

Bibliografía

- [1] G. del Estado de Jalisco, "Jalisco: Cultura y tradiciones." <https://www.jalisco.gob.mx/es/jalisco/cultura-y-tradiciones>, 2023. [En línea; accedido el 26-marzo-2024].
- [2] I. N. de Estadística y Geografía, "Censo de población y vivienda 2020." https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/censos/poblacion_y_vivienda/2020/default.html, 2020. [En línea; accedido el 26-marzo-2024].
- [3] S. de Educación Pública, "Estadísticas del sistema educativo nacional 2021-2022." https://www.planeacion.sep.gob.mx/Doc/estadistica_e_indicadores/principales_cifras/principales_cifras_2021_2022_bolsillo.pdf, 2022. [En línea; accedido el 26-marzo-2024].
- [4] Instituto de información estadística y geográfica de Jalisco, "Información sociodemográfica por colonia, jalisco (colonias enteras), inegi2020-ine2021." https://iieg.gob.mx/ns/?page_id=882, 2021. [En línea; accedido el 26-octubre-2023].
- [5] Instituto de información estadística y geográfica de Jalisco, "Homicidio doloso, enero de 2021." https://iieg.gob.mx/ns/wp-content/uploads/2021/02/reporte_mensual_homicidio_enero_2021.html, 2021. [En línea; accedido el 14-Febrero-2024].
- [6] El País México, "Las cifras de homicidios en el sexenio de López Obrador y cómo "a la baja" no implica un país menos violento." <https://elpais.com/mexico/2023-09-05/las-cifras-de-homicidios-en-el-sexenio-de-lopez-obrador-y-como-a-la-baja-no-implica-un-pais-menos-violento.html>, 2023. [En línea; accedido el 14-Febrero-2024].
- [7] E. Política, "Cifra de 100,000 homicidios confirma tres años de violencia inédita con AMLO." <https://politica.expansion.mx/mexico/2021/10/21/cifra-de-100-000-homicidios-confirma-tres-anos-de-violencia-inedita-con-amlo>, 2021. Consultado el 26 de marzo de 2024.
- [8] IIEG, "Instituto de información estadística y geográfica del estado de Jalisco." <https://iieg.gob.mx/ns/>, 2024. Consultado el 26 de marzo de 2024.
- [9] IIEG, "Plataforma de seguridad." https://iieg.gob.mx/plataforma_seguridad/home, 2024. Consultado el 26 de marzo de 2024.

- [10] INEGI, "Censo población y vivienda 2020 inegi información demográfica y social de colonias del estado jalisco." <https://datos.jalisco.gob.mx/temas/informacion-estadistica-y-geografica>, 2020. Consultado el 26 de marzo de 2024.
- [11] D. A. Jalisco, "Datos abiertos jalisco." <https://datos.jalisco.gob.mx/>, 2023. Consultado el 26 de Septiembre de 2023.
- [12] F. Jalisco, "Robo autopartes." <https://fiscalia.jalisco.gob.mx/prevencion-social/robo-de-auto-partes/>, 2023. Consultado el 26 de diciembre de 2023.
- [13] REPUVE, "Consulta no oficial de repuve." <https://www.repuve-consultar.com/jalisco>, 2023. Consultado el 26 de diciembre de 2023.
- [14] INEGI, "Encuesta nacional de seguridad pública urbana 2021, datos al cuarto trimestre del año." https://www.inegi.org.mx/rnm/index.php/catalog/654/variable/F13/V944?name=BP1_6_2, 2023. Consultado el 26 de diciembre de 2023.
- [15] C. A. Arturo Sánchez, "¿sabes cuántos robos a casa habitación hay en el país?" <https://feherandfeher.com/sabes-cuantos-robos-a-casa-habitacion-hay-en-el-pais/>, 2022. Consultado el 26 de diciembre de 2023.
- [16] G. de México, "Incidencia delictiva." <https://www.gob.mx/sesnsp/articulos/incidencia-delictiva>, 2024. Consultado el 21 de Febrero de 2024.
- [17] D. . O. F. del Estado, "Dirección de investigación de robo a negocio." <https://info.jalisco.gob.mx/area/direccion-de-investigacion-de-robo-negocio>, 2023. Consultado el 26 de diciembre de 2023.
- [18] C. S. D. G. D. E. D. Jalisco, "Fe logra vinculación a proceso de un sujeto por su probable participación en despojar de sus pertenencias a una pareja." <https://www.jalisco.gob.mx/es/prensa/noticias/158086>, 2023. Consultado el 26 de diciembre de 2023.
- [19] C. S. D. G. D. E. D. Jalisco, "Fe desarticula banda dedicada al robo de cuentahabientes tipo conejero; logra la vinculación de cuatro posibles integrantes." <https://fiscalia.jalisco.gob.mx/comunicacion-social/boletin-de-prensa/boletin-1941---20230809>, 2023. Consultado el 26 de diciembre de 2023.
- [20] E. Informador, "Robo al transporte de carga crece en jalisco." <https://www.informador.mx/jalisco/Carreteras-Robo-al-transporte-decarga-crece-en-Jalisco-20230619-0018.html>, 2023. Consultado el 26 de diciembre de 2023.
- [21] E. Informador, "Repuntan robos a casa-habitación en jalisco." <https://www.informador.mx/jalisco/Seguridad-en-Jalisco-Repuntan-robos-a-casa-habitacion-en-el-estado-20230701-0096.html>, 2023. Consultado el 26 de diciembre de 2023.

- [22] D. Jalisco, "Unidades de atención a la violencia familiar." <https://difjalisco.gob.mx/ppna/unidades-atencion-violencia-familiar>, 2023. Consultado el 26 de diciembre de 2023.
- [23] Telediario, "Cómo fue el robo a camioneta de valores afuera de bbva en polanco, guadalajara." <https://www.telediario.mx/comunidad/asalto-banco-en-guadalajara-como-fue>, 2024. Consultado el 10 de Febrero de 2024.
- [24] C. P. Jalisco, "Artículo 63." https://leyes-mx.com/codigo_penal_jalisco/63.htm, 2023. Consultado el 26 de diciembre de 2023.
- [25] E. Informador, "Denuncian abuso sexual a custodias en puente grande; la sse investiga el caso." <https://www.informador.mx/jalisco/Puente-Grande-Denuncian-abuso-sexual-a-custodias-en-el-penal-la-SSE-investiga-el-caso-20230215-0069.html>, 2023. Consultado el 26 de diciembre de 2023.
- [26] F. del Estado FE, "Abuso sexual infantil." <https://fiscalia.jalisco.gob.mx/prevencion-social/abuso-sexual-infantil>, 2023. Consultado el 26 de diciembre de 2023.
- [27] M. Jalisco, "Feminicidios." <https://mide.jalisco.gob.mx/mide/panelCiudadano/detalleIndicador/1671>, 2023. Consultado el 26 de diciembre de 2023.
- [28] W. J.M, "Introductory econometrics: A modern approach. south-western cengage learning," 2012. Consultado el 23 de febrero de 2023.
- [29] IBES, "K-means, ¿como funciona?." <https://www.iebschool.com/blog/algorithm-k-means-que-es-y-como-funciona-big-data/>, 2023. Consultado el 21 de febrero de 2024.
- [30] U. de los Andes, "Introducción a la regresión lineal: definición y aplicaciones." <https://programas.uniandes.edu.co/blog/regresion-lineal>, 2023. Consultado el 21 de febrero de 2024.
- [31] DATAtab, "Regresion lineal." <https://datatab.es/tutorial/linear-regression>, 2023. Consultado el 21 de febrero de 2024.
- [32] P. J. Rousseeuw, "'silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, 1987.
- [33] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, pp. 1189–1232, 2001.
- [34] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, first ed., 2006.
- [35] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "'learning representations by back-propagating errors," *Nature*, vol. 323, 1986.
- [36] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

- [37] L. Kaufman and P. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. 1990.
- [38] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques, Third Edition*. third ed., 2011.
- [39] D. C. Leiva, F. M. Valverde, N. A. Fioravanti, and J. C. Retana, "Modelo predictivo aplicado a las estadísticas policiales del poder judicial de costa rica para el análisis de la criminalidad por regiones." <https://repositorio.ulacit.ac.cr/bitstream/handle/20.500.14230/10631/REF-1632841769-1.pdf?sequence=1>. Consultado el 01/05/2024.
- [40] H. R. C. C. . A. V. Ninahuanca, "El efecto de la educación en la tasa de criminalidad: Un panel de datos para los departamentos de la sierra del Perú 2012 al 2017." https://repositorio.ulima.edu.pe/bitstream/handle/20.500.12724/17796/T018_48105753_T.pdf?sequence=1&isAllowed=y. Consultado el 01/05/2024.