

TRABAJO DE GRADO
Opción Seminario-Diplomado.

Análisis de Datos sobre la Correlación entre Educación y la Delincuencia en Colombia

Corporación Universitaria Remington
Facultad de Ingenierías
Ingeniería de Sistemas

Cristian Camilo Gonzalez Cano¹
Ivonne Castaño²
John Edisson Amortegui Granada³
Opción de Trabajo de Grado Seminario
2024

¹ Estudiante de Noveno Semestre de Ingeniería de Sistemas UniRemington Pereira. E-mail:
cristian.gonzalez.5846@uniremington.edu.co

² Asesora Temática Seminario en Big Data y Ciencia de Datos. UniRemington Pereira. E-mail:
ivonne.castano@uniremington.edu.co

³ Asesor Metodológico Seminario en Big Data y Ciencia de Datos. UniRemington Pereira. E-mail:
ivonne.castano@uniremington.edu.co

Dedicatoria

Dedico este trabajo a todas las personas que de alguna forma me han ayudado y a todas las personas que a pesar de tener muchas situaciones que dificultan su aprendizaje y motivación siguen adelante, enfocados en el resultado y las metas que se proponen.

Agradecimientos

Quiero agradecer a mis familiares que están y los que ya no están, por el apoyo y la atención durante todos mis estudios, motivándome a continuar mejorando cada día, a los profesores por la creatividad, dedicación y entrega en cada clase, inspirándonos a tener nuevos logros y metas.

Tabla de contenidos

Resumen.....	5
Palabra clave.....	5
Pregunta orientadora de la búsqueda.....	6
Metodología de búsqueda de la información.....	7
Sustentación teórica de la pregunta.....	18
Conclusiones.....	19
Referencias.....	20

Resumen

En el presente ejercicio de profundización, aplica una variación de la metodología CRISP-DM de IBM para el análisis de datos y entender el impacto de la educación en la delincuencia en Colombia, extrayendo datos de valor a partir de unos data sets obtenidos de fuentes libres, que fueron ordenados y cruzados con la cantidad de hurtos presentados en las distintas ciudades de Colombia y la cantidad de estudiantes de educación preescolar, básica y media.

Palabras clave: Ciencia de Datos, Análisis, Delincuencia, Educación.

Pregunta orientadora de la búsqueda

Diversos estudios de la relación entre el nivel educativo y la criminalidad, específicamente en los casos de hurto, de acuerdo a varios estudios han sugerido que existe una correlación significativa, argumentando que una mayor educación puede reducir el interés de las personas a cometer delitos, a lo cual surge la pregunta sobre ¿cuál es el impacto de la educación en la delincuencia en Colombia mediante el Análisis de Datos?

Propósito

El propósito de este estudio es explorar en profundidad cómo la educación puede influir en el comportamiento delictivo en contextos específicos como el colombiano. Se analizarán diversas perspectivas teóricas y estudios empíricos para proporcionar una visión comprehensiva de este fenómeno social.

Metodología de búsqueda de la información

1. Entendimiento del negocio

El alcance de la educación en Colombia, ha incrementado significativamente con los distintos incentivos presentados por distintos gobiernos, organizaciones y empresas que requieren personal capacitado, por ello se ha presentado una gran disminución de analfabetismo, pero contrario a ello, los casos de delincuencia común como el hurto continúan presentándose en distintas ciudades del país, así como un aumento en algunas regiones, es por ello que se requiere analizar las distintas fuentes de información abierta, para intentar encontrar una relación directa entre la educación y los hurtos registrados en el 2018 y el 2023.

2. Enfoque analítico

De acuerdo a la información que se requiere analizar, el enfoque analítico será enfoque descriptivo, para lograr categorizar por niveles de educación y su relación con los hurtos presentados en los departamentos de Colombia.

3. Requisitos de datos

Para realizar este estudio se requieren los datos de la población total en los distintos departamentos de Colombia, los totales de población según su nivel de educación y el total de hurtos del año 2018 y 2023.

4. Recopilación de datos

Se recopilaron data sets de fuentes abiertas y libres basados en la necesidad del análisis como la plataforma nacional de datos abiertos de Colombia y se analizaron los estudios, sobre temas relacionados en el repositorio institucional de la universidad

nacional de Colombia, Sistema de Información Científica Redalyc y el Repositorio Institucional Universidad EAFIT. Para abordar la pregunta de investigación sobre el impacto de la educación en la delincuencia en el DANE “PERSONAS_DEMOGRAFICO_Cuadros CNPV_2018.XLSX” y en Datos con el nombre “Reporte_Hurto_por_Modalidades_Polic_a_Nacional_20240611”.

5. Comprensión de datos

Se identificaron las variables de los data set, los cuales revelan información acorde al análisis a realizar como totales de población según su nivel académico, dividido por departamento, nivel educativo, así como también variables irrelevantes para el actual estudio que no aporta correlación entre los mismos.

6. Preparación de los datos

Luego de analizar los datos previamente obtenidos, se determinó que se eliminarían datos irrelevantes y solo se utilizó una de las distintas hojas con las que contaba el data set “PERSONAS_DEMOGRAFICO_Cuadros CNPV_2018.XLS” dejando solo la hoja con nombre “17PD” y se agruparon los distintos grupos de la población que estaba segmentada por etnia, código de departamento, código de municipio, área y sexo, así como también el arma o medio de hurto, genero, grupo etario, y tipo de hurto, así como también se proyectó el crecimiento de la población al 2023 de acuerdo con los datos actuales del DANE, debido a que el último censo fue realizado en el 2018 y no se cuenta con un censo reciente, comparando las cifras presentadas en el 2018 con respecto a las del 2023 se realizó la proyección incrementando en cada departamento de acuerdo al porcentaje de población de cada departamento en el 2018.

7. Modelado

Tabla 1.

Resumen de la población por departamento y hurtos registrados en Colombia en 2018 y 2023.

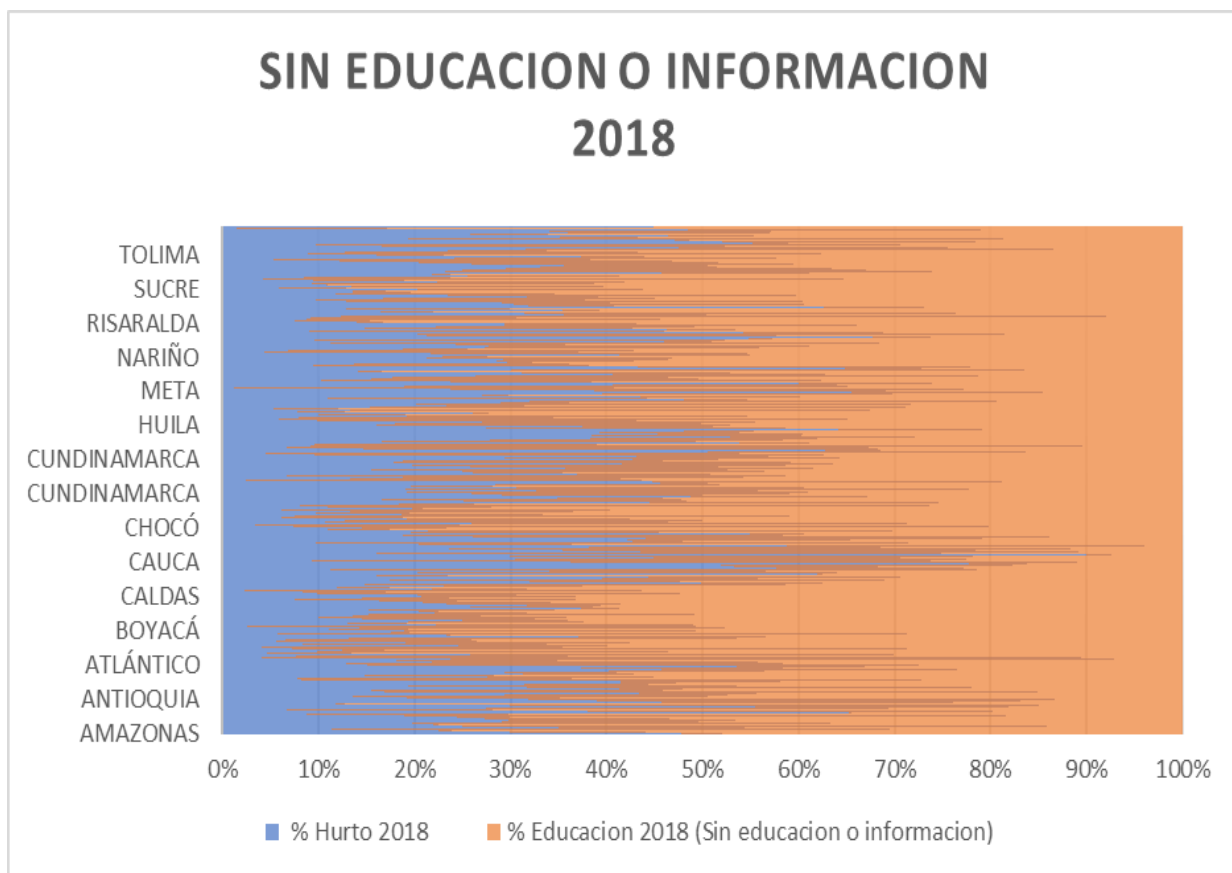
DEPARTAMENTO	HURTOS VEHÍCULOS 2018	TOTAL, SIN EDUCACIÓN SIN INFORMACIÓN 2018	POBLACIÓN 2018	HURTOS VEHÍCULOS 2023	TOTAL, SIN EDUCACIÓN O INFORMACIÓN 2023	POBLACIÓN 2023
AMAZONAS	29	2.673	38.451	46	3.394	48.818
ANTIOQUIA	9.189	285.731	5.495.595	14.938	401.512	6.977.358
ARAUCA	302	17.250	218.565	99	21.901	277.496
ATLÁNTICO	1.673	87.517	2.051.825	4.229	121.766	2.605.053
BOLÍVAR	1.277	114.949	1.642.540	1.250	141.473	2.085.414
BOYACÁ	149	47.019	886.574	454	76.801	1.125.619
CALDAS	151	47.911	816.064	205	50.027	1.036.097
CAQUETÁ	343	25.264	301.588	328	56.166	382.904
CASANARE	416	19.959	339.387	515	18.970	430.895
CAUCA	2.593	62.688	1.186.125	3.858	3.266.740	1.505.937
CESAR	1.422	84.829	1.049.644	1.884	75.401	1.332.657
CHOCÓ	401	23.736	198.270	644	31.449	251.729
CÓRDOBA	854	141.395	1.410.212	782	169.111	1.790.444
CUNDINAMARCA	8.704	371.902	9.215.612	14.915	481.932	11.700.393
GUAINÍA	3	2.818	27.304	6	3.578	34.666
GUAJIRA	1.034	112.474	704.826	1.191	95.801	894.866

DEPARTAMENTO	HURTOS VEHÍCULOS 2018	TOTAL, SIN EDUCACIÓN SIN INFORMACIÓN 2018	POBLACIÓN 2018	HURTOS VEHÍCULOS 2023	TOTAL, SIN EDUCACIÓN O INFORMACIÓN 2023	POBLACIÓN 2023
GUAVIARE	34	6.952	83.284	56	10.154	105.740
HUILA	771	51.631	901.866	2.854	99.466	1.145.034
MAGDALENA	857	91.060	1.114.372	1.087	110.750	1.414.837
META	969	55.616	817.503	1.576	36.517	1.037.924
NARIÑO	1.288	84.077	1.155.286	2.221	104.398	1.466.783
NORTE DE SANTANDER	1.092	83.414	1.177.459	1.773	101.644	1.494.934
PUTUMAYO	258	13.830	236.807	326	16.857	300.657
QUINDÍO	215	26.645	224.157	159	19.962	284.596
RISARALDA	232	39.409	799.211	605	50.035	1.014.700
SAN ANDRÉS	70	799	44.949	45	1.649	57.068
SANTANDER	1.403	82.811	1.806.845	2.384	109.358	2.294.020
SUCRE	517	83.814	772.189	438	101.069	980.392
TOLIMA	650	71.304	1.115.969	1.264	91.566	1.416.865
VALLE	5.656	160.168	3.504.622	9.333	207.886	4.423.323
VAUPÉS	2	2.108	25.477	7	2.676	32.346
VICHADA	25	10.220	63.434	34	12.976	80.538
Total, general	42.579	2.311.973	39.426.012	69.506	6.092.982	50.030.104

Fuente: (Autoría Propia, 2024)

Gráfico 1.

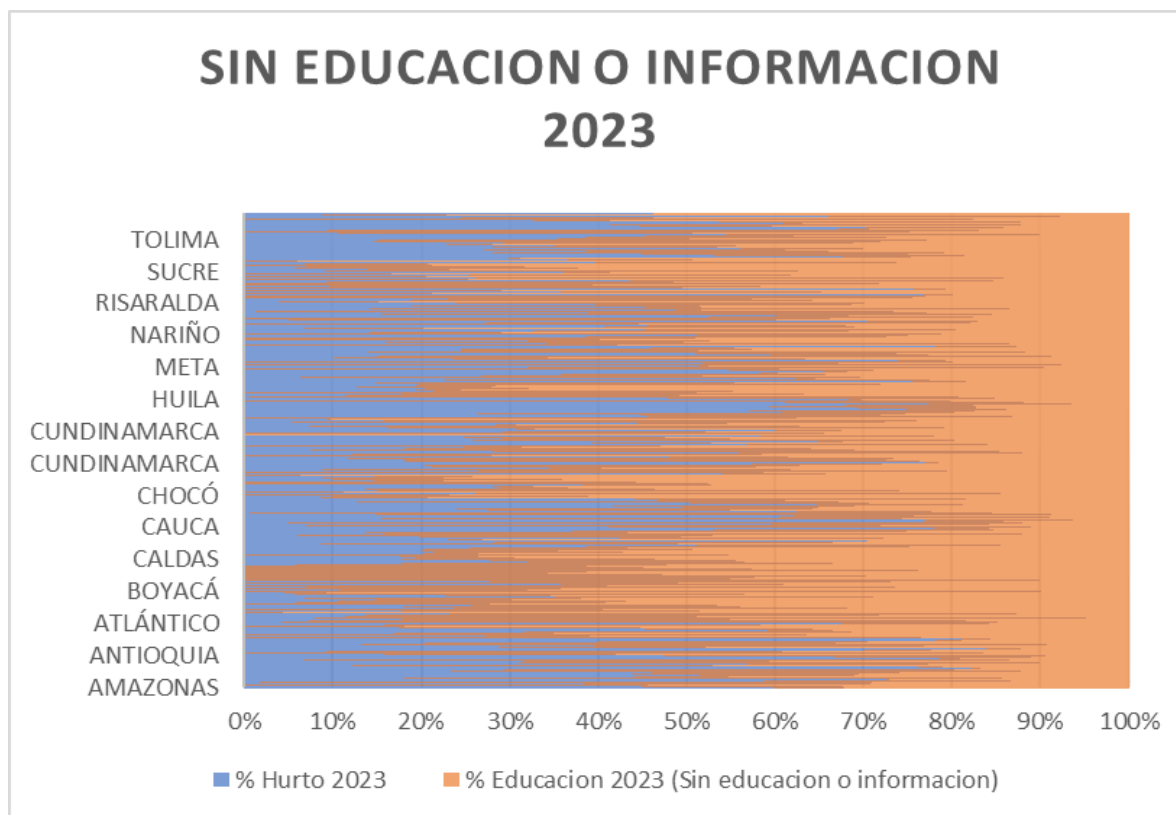
Comparativa de la población sin educación o sin información y hurtos reportados en el 2018.



Nota. El gráfico representa la población de Colombia reportada sin educación o sin información, comparándose con la cantidad de hurtos reportados en los departamentos en el año 2018.

Gráfico 2.

Comparativa de la población sin educación o sin información, comparado con la cantidad de hurtos reportados en el 2023.



Nota El grafico representa la población de Colombia reportada sin educación o sin información, comparándose con la cantidad de hurtos reportados en los departamentos en el año 2023.

Tabla 2.

Departamentos y código de departamento para el análisis

COD_DEPARTAMENTO	DEPARTAMENTO
1	AMAZONAS
2	ANTIOQUIA
3	ARAUCA
4	ATLÁNTICO
5	BOLÍVAR
6	BOYACÁ
7	CALDAS
8	CAQUETÁ
9	CASANARE
10	CAUCA
11	CESAR
12	CHOCÓ
13	CÓRDOBA
14	CUNDINAMARCA
15	GUAINÍA
16	GUAJIRA
17	GUAVIARE
18	HUILA
19	MAGDALENA
20	META
21	NARIÑO
22	NORTE DE SANTANDER
23	PUTUMAYO
24	QUINDÍO
25	RISARALDA
26	SAN ANDRÉS
27	SANTANDER
28	SUCRE
29	TOLIMA
30	VALLE
31	VAUPÉS
32	VICHADA
33	Total, general

Fuente: (Autoría Propia, 2024)

Tabla 3.

Covarianza de la población por departamento y hurtos registrados en Colombia en 2018 y 2023

	COD DEPARTAME NTO	HURTOS VEHÍCUL OS 2018	TOTAL, SIN EDUCACIÓN SIN INFORMACIÓ N 2018	POBLACIÓ N 2018	HURTOS VEHÍCUL OS 2023	TOTAL, SIN EDUCACIÓ N O INFORMACI ÓN 2023	POBLACIÓ N 2023
Media	17	2.581	140.120	2.389.455	4.212	369.272	3.032.127
Error típico	2	1.309	69.247	1.199.103	2.142	203.900	1.521.624
Mediana	17	650	55.616	817.503	782	76.801	1.037.924
Moda	#N/D	#N/D	#N/D	#N/D	#N/D	#N/D	#N/D
Desviación estándar	10	7.519	397.792	6.888.323	12.307	1.171.314	8.741.062
Varianza de la muestra	94	56.528.688	158.238.672.888	47.448.986.98 7.203	151.469.570	1.371.976.077. 774	76.406.169.604. 524
Curtosis	- 1,20	26,94	30,21	28,25	26,62	19,61	28,25
Coefficiente de asimetría	- 0,00	5,03	5,41	5,19	4,99	4,37	5,19
Rango	32	42.577	2.311.174	39.400.535	69.500	6.091.333	49.997.758
Mínimo	1	2	799	25.477	6	1.649	32.346
Máximo	33	42.579	2.311.973	39.426.012	69.506	6.092.982	50.030.104
Suma	561	85.158	4.623.946	78.852.024	139.012	12.185.967	100.060.207
Cuenta	33	33	33	33	33	33	33
Mayor (1)	33	42.579	2.311.973	39.426.012	69.506	6.092.982	50.030.104
Menor(1)	1	2	799	25.477	6	1.649	32.346
Nivel de confianza (95,0%)	3,43	2.665,96	141.051,00	2.442.492,96	4.363,98	415.329,82	3.099.445,91

Fuente: (Autoría Propia, 2024)

Tabla 4.

Coeficiente de correlación de la población por departamento y hurtos registrados en Colombia en 2018 y 2023.

	<i>COD DEPARTAMEN TO</i>	<i>HURTOS VEHÍCULO S 2018</i>	<i>TOTAL, SIN EDUCACIÓN SIN INFORMACI ÓN 2018</i>	<i>POBLACIÓ N 2018</i>	<i>HURTOS VEHÍCULO S 2023</i>	<i>TOTAL, SIN EDUCACIÓN O INFORMACI ÓN 2023</i>	<i>POBLACI ÓN 2023</i>
COD DEPARTAMENTO	1						
HURTOS VEHÍCULOS 2018	0,23892173	1					
TOTAL, SIN EDUCACIÓN SIN INFORMACIÓN 2018	0,26218563	09	1				
		0,9941721	0,99563535				
POBLACIÓN 2018	0,25243419	39	1	1			
		0,9992111	0,98797999	0,9941727			
HURTOS VEHÍCULOS 2023	0,23969058	87	2	9	1		
TOTAL, SIN EDUCACIÓN O INFORMACIÓN 2023	0,18864173	99	7	79	76	1	
		0,9941511	0,99563562	0,9999998	0,9941524	0,86866442	
POBLACIÓN 2023	0,25228396	2	3	7	92	1	1

Fuente: (Autoría Propia, 2024)

Tabla 5.

Regresión de la población por departamento y hurtos registrados en Colombia en 2023

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	2	2.719.808.095.609.450,00	1.359.904.047.804.720,00	1.474,81	0,00
Residuos	31	28.584.635.519.659,10	922.085.016.763,20		
Total	33	2.748.392.731.129.100,00			

Fuente: (Autoría Propia, 2024)

Tabla 6.

Regresión de la población por departamento y hurtos registrados en Colombia en 2023

	<i>Coefficient es</i>	<i>Error típico</i>	<i>Estadístic o t</i>	<i>Probabilid ad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>	<i>Inferior 97,0%</i>	<i>Superior 97,0%</i>
Intercepción	0	#N/D	#N/D	#N/D	#N/D	#N/D	#N/D	#N/D
	703,5413	27,96354	25,15923	3,42871E-	646,50935	760,5734	639,93511	767,147660
HURTOS VEHÍCULOS 2023	858	85	134	22	26	19	07	9
					-		-	
TOTAL, SIN EDUCACIÓN O INFORMACIÓN 2023	0,048106	0,296250	0,162383	0,872057	0,5561013	0,652314	0,6257501	0,72196288
	383	943	899	409	98	164	19	5

Fuente: (Autoría Propia, 2024)

8. Evaluación del modelo

Tabla 7.

Regresión de la población por departamento y hurtos registrados en Colombia en 2023

<i>Estadísticas de la regresión</i>	
Coeficiente de correlación múltiple	0,994786162
Coeficiente de determinación R ²	0,989599508
R ² ajustado	0,957005944
Error típico	960252,5797
Observaciones	33

Fuente: (Autoría Propia, 2024)

De acuerdo al coeficiente de correlación múltiple del modelo, si la cantidad de población sin educación o información aumenta tiene una correlación alta en la cantidad de hurtos, y las variables entre si tienen un alto índice de correlación representado por el coeficiente de determinación R² entre más cercano a 1 sea, es más preciso, por lo cual las variables son importantes para el análisis, confirmando la correlación entre las variables de población, nivel educativo y el nivel de hurtos.

9. Implementación

No vamos a hacer un uso práctico de los modelos que se encontraron, debido a que estamos realizando un modelo de análisis descriptivo, esta fase solo se debe usar en un modelo predictivo.

10. Retroalimentación

Si bien los datos son relevantes, Cano (2020) sugiere que la “Si bien fue un trabajo que explica de manera general las dinámicas poblacionales, la evidencia es consistente con la hipótesis planteada de que la educación tiene un efecto inverso sobre el crimen para Colombia.”(p.14), se debe ahondar en el tema con respecto a otras variables que

impactan, como grupo étnico, religión, sexo, entre otras que no se tuvieron en cuenta para un análisis mas exacto, que si bien los resultados son muy relacionados, no podemos excluir otras situaciones que impactan o influyen en el estudio, así como contar con datos mas actualizados para tener una mayor probabilidad de acercarnos a datos reales, pero que para este estudio fueron cumplidos en su totalidad.

Sustentación teórica de la pregunta

La fundamentación teórica que respalda la pregunta orientadora sobre el impacto de la educación en la delincuencia en Colombia se desarrollará, tomando como base la revisión de literatura relacionada. Para esto se ha utilizado una metodología de búsqueda que incluye recopilación y análisis de DataSets obtenidos desde la plataforma nacional de datos abiertos de Colombia, así como también una revisión a estudios disponibles en el repositorio institucional de la Universidad Nacional de Colombia, el Sistema de Información Científica Redalyc y el Repositorio Institucional de la Universidad EAFIT.

El enfoque descriptivo se basa en mediciones cuantitativas utilizadas para explorar las relaciones entre variables educacionales y tasas delictivas dentro del contexto colombiano. Este análisis tiene como finalidad señalar algunas tendencias, patrones y descubrimientos fundamentales que podrían establecer unas bases teóricas sólidas para futuras investigaciones empíricas. Diversos puntos de vista teóricos serán considerados, así como también investigaciones previas como: cómo el acceso a la educación, calidad educativa y otros factores socioeconómicos influyen en prevenir o aumentar delincuencia. El marco teórico estará nutrido por investigaciones significativas que han indagado cómo diferentes niveles de enseñanza pueden influir tanto en las decisiones criminales como los actos ofensivos cometidos.

Conclusiones

Al llevar a cabo el análisis de los datasets disponibles, utilizando el enfoque descriptivo, los modelados y al proyectar estos datos hasta el año 2023, se logra establecer una relación directa y clara entre el nivel educativo y los casos de hurto registrados en diferentes regiones de Colombia. Se observan similitudes en los porcentajes de hurto y los niveles de educación en varios departamentos, este análisis no considera adecuadamente otros factores determinantes que tienen un impacto significativo en los índices de criminalidad. Entre estos factores se incluyen la presencia y actividad de bandas criminales, las condiciones socioeconómicas locales como el desempleo y otros aspectos contextuales relevantes que influyen en las tendencias delictivas.

Es crucial reconocer que la complejidad de la criminalidad va más allá de la educación como único factor. Los entornos sociales, económicos y políticos juegan un papel crucial en la configuración de las dinámicas delictivas en cada región. Por lo tanto, cualquier análisis que busque comprender estas relaciones debe ser integral y considerar una gama más amplia de variables para proporcionar una evaluación precisa y completa del fenómeno delictivo en el contexto colombiano, que, si bien en los resultados expuestos en este análisis tienen una alta correlación entre el nivel educativo y la cantidad de hurtos, no son definitivos y se debe ampliar los datos para un mejor análisis.

Lista de referencias

Normas APA, (2019). *Manual 7a edición*. <https://normas-apa.org/>.

Castaño, I (2024). Metodología de la Ciencia de Datos.

https://drive.google.com/file/d/1A7yTe2tkbO_yovcKU_0b1_D_i97TWMK7/view

DANE, (2019). Población 2018.

http://systema59.dane.gov.co/redcol/CNPV2018/PERSONAS_DEMOGRAFICO_Cuadros%20CNPV_2018.XLSX

EAFIT, (2020). Efectos de la educación sobre la delincuencia: un análisis para la implementación de políticas públicas.

<https://repository.eafit.edu.co/server/api/core/bitstreams/72691b02-bdd9-4d52-ba57-25920e66003b/content>