

Estudio de Incautaciones de Fauna Silvestre mediante el Análisis de Datos

Corporación Universitaria Remington

Diego Alexander Gómez Taborda¹

Ivonne Castaño Osorio²

Jhon Edison Amortegui Granada³

Opción de Trabajo de grado Seminario

2024

¹ Estudiante de Noveno Semestre de Ingeniería de Sistemas Uniremington sede Pereira. Email:

diego.gomez.0444@miremington.edu.co

² Tutor temático seminario Big Data y Ciencia de Datos. Email: ivonne.castano@uniremington.edu.co

³ Tutor metodológico seminario Big Data y Ciencia de Datos. Email:

john.amortegui@uniremington.edu.co

Dedicatoria

Con inmensa gratitud y profundo amor, dedico este trabajo de grado a Dios, fuente de sabiduría y fortaleza, por iluminar mi camino durante toda la carrera y por darme la perseverancia para alcanzar esta meta, a mis padres pilar fundamental en mi vida, por su amor incondicional, apoyo inquebrantable y por inculcarme los valores que me han permitido ser la persona que soy hoy en día. Gracias por creer en mí siempre, por motivarme a perseguir mis sueños y por ser mi inspiración constante. A mi hermana, mis sobrinos y mi cuñado, gracias por su cariño, apoyo y por estar siempre presentes en los momentos más importantes de mi vida. Agradezco profundamente su comprensión y por compartir mi alegría en este logro tan significativo. A mis compañeros de clase, gracias por compartir conmigo esta experiencia tan enriquecedora, por la colaboración y por los buenos momentos que compartimos. Gracias por ser parte de este logro que también es de ustedes.

Agradecimientos

Quiero agradecer profundamente a todos los docentes que han marcado mi camino durante la carrera, gracias por su dedicación, su conocimiento y su pasión por la enseñanza, todo esto ha sido fundamental en mi formación como ingeniero de sistemas. A todas las personas que han creído en mí, a todos ellos agradezco por estar presentes en mi vida y por haber contribuido de alguna manera a mi formación personal y profesional. Su apoyo ha sido invaluable para alcanzar esta meta.

Contenido

Resumen.....	5
Pregunta orientadora de la búsqueda	6
Comprensión del negocio.....	7
Enfoque analítico	7
Requisitos de datos	8
Recopilación de los datos.....	8
Comprensión de los datos	8
Preparación de datos.....	8
Modelado	9
Evaluación.....	10
Implementación.....	10
Retroalimentación.....	10
Sustentación teórica de la pregunta.....	11
Figuras y tablas.....	13
Conclusiones.....	19
Referencias	20

Resumen

En este proyecto se busca llevar a cabo un análisis integral y exhaustivo de un conjunto de datos publicado en la plataforma de Datos Abiertos por la Corporación Autónoma Regional de Risaralda CARDER, referente a las incautaciones de fauna silvestre. El principal objetivo es utilizar herramientas avanzadas de ciencia de datos para determinar que especies son confiscadas con mayor frecuencia, que municipios del departamento de Risaralda tienen los niveles más altos en cuanto al comercio ilegal de vida silvestre y así poder identificar que entidades gubernamentales tienen la mayor cantidad de incautaciones. Este análisis proporciona una comprensión más profunda de los patrones y tendencias del tráfico de fauna silvestre en la región, buscando proporcionar información valiosa a las autoridades y entidades correspondientes. De esta manera, las entidades territoriales llevarán a cabo estrategias más efectivas, esto con el fin de mitigar el impacto en el tráfico e ilegal de fauna silvestre. Este estudio de análisis de datos se realiza haciendo uso de la metodología para la ciencia de datos CRISP-DM.

CRISP-DM incluye un modelo y una guía, estructurados en seis fases, algunas de las cuales son bidireccionales, es decir que de una fase en concreto se puede volver a una fase anterior para poder revisarla, por lo que la sucesión de fases no tiene por qué ser ordenada desde la primera hasta la última. En la figura 4 se puede observar las fases en las que se divide CRISP-DM y las posibles secuencias a seguir entre ellas. (Galán, 2015, p. 21)

Palabras clave: Análisis de Datos, Autoridad Ambiental, CARDER, Datos, Especies, Fauna Silvestre.

Pregunta orientadora de la búsqueda

El análisis de datos sobre las incautaciones de fauna silvestre es fundamental en la toma de decisiones para la conservación y gestión de la biodiversidad. La autoridad ambiental de Risaralda en este caso la Corporación Autónoma Regional de Risaralda CARDER y de la mano de otras entidades a nivel regional, se enfrentan a un enorme desafío, el de proteger la biodiversidad de la región mientras aborda las amenazas que enfrentan las especies silvestres. En este contexto, realizar una investigación sobre el impacto que tiene el tráfico de fauna silvestre en la región, lo anterior lleva a plantear la siguiente pregunta:

¿Cuáles son los patrones y tendencias en las incautaciones de fauna silvestre en el departamento de Risaralda durante los últimos años, en términos de las especies afectadas, los municipios con mayor incidencia, y las autoridades involucradas?

Metodología de búsqueda de la información

El presente trabajo tendrá como objetivo el uso de la metodología CRISP-DM, la cual permite la ejecución avanzada de análisis de ciencia de datos.

La metodología CRISP-DM es una de las más empleadas actualmente para el desarrollo de proyectos de minería de datos. En 1997, se puso en marcha bajo el financiamiento del Programa de Investigación y Desarrollo en Tecnologías de Información de la Unión Europea. (Espinosa, 2020, p. 2).

Esta metodología consta de diez fases principales, las cuales se describen a continuación.

Comprensión del negocio

Esta primera fase es probablemente la más importante y aglutina las tareas de comprensión de los objetivos y requisitos del proyecto desde una perspectiva de negocio, con el fin de convertirlos en objetivos técnicos y en un plan de proyecto. Sin lograr comprender dichos objetivos, ningún algoritmo por muy sofisticado que sea permitirá obtener resultados fiables. (Galán, 2015, p. 22)

Con ello se busca definir cuáles serán los objetivos principales del proyecto, a donde se quiere llegar con el análisis realizado, buscando resolver una pregunta central con los datos suministrados.

Enfoque analítico

Una vez definido el problema a resolver, se elige el enfoque apropiado, ya sea realizando un análisis de los datos obtenidos de manera predictiva o tomando estos datos y dando un enfoque descriptivo de los mismos. El enfoque analítico también busca determinar de una manera concisa como plantear y mostrar estos datos en ocasiones de forma estadística buscando dar la respuesta adecuada.

Requisitos de datos

En esta etapa de la metodología CRISP-DM se busca identificar los datos requeridos, analizando de manera adecuada el contenido, los formatos y las fuentes para realizar un estudio detallado de la información que se va a analizar.

Recopilación de los datos

Se debe realizar una recopilación de datos inicial, esta recopilación permite evaluar si los datos suministrados son óptimos para efectuar el análisis, de lo contrario el científico de datos deberá indagar a profundidad buscando obtener información sólida y veraz. Una vez que estos datos son recopilados el científico de datos tendrá una buena percepción de los datos con lo que trabajará.

Comprensión de los datos

La fase de comprensión de datos de CRISP-DM implica estudiar más de cerca los datos disponibles de minería. Este paso es esencial para evitar problemas inesperados durante la siguiente fase (preparación de datos) que suele ser la fase más larga de un proyecto. (IBM, 2021)

En esta etapa de la metodología CRISP-DM el científico de datos interpreta si los datos recopilados son útiles para resolver el problema expuesto, adicionalmente se busca identificar variables relevantes logrando con ello detectar posibles problemas en la calidad de los datos, para finalmente realizar la documentación.

Preparación de datos

La fase de preparación de datos cubre todas las actividades necesarias para construir el conjunto de datos final [datos que se introducirán en la(s) herramienta(s) de modelado] a partir de los datos brutos iniciales. Es probable que las tareas de preparación de datos se realicen varias veces y no en ningún orden prescrito. Las tareas incluyen la selección

de tablas, registros y atributos, así como la transformación y limpieza de datos para herramientas de modelado. (Chapman et al., 2000, p. 20).

En esta fase el científico de datos se encarga de limpiar la información suministrada, transformándola y preparándola para el modelado, esta fase se construyen nuevas variables o se transforman las existentes para facilitar el análisis de los datos proporcionados. Los datos obtenidos en formato CSV, se cargaron en Excel, y se validaron con Power Query, posteriormente se validaron si las columnas presentaban errores, estos errores se corrigieron, se eliminaron columnas con datos innecesarios y que no fueran relevantes para el análisis. Se ajustaron los formatos de fecha, corrigiendo las celdas con formatos incorrectos, se convirtieron las celdas que se necesitaban en formato número que estaban en formato texto. Esto con el fin de dar una limpieza profunda a la información que va a ser analizada.

Modelado

Durante esta fase, se desarrollan y se evalúan diferentes modelos para realizar el análisis de los datos suministrados, se selecciona que tipo de modelo se presentara la información, ya sea un enfoque descriptivo, analítico o de correlaciones; todo esto con el fin de mostrar una información veraz y de óptima calidad.

Se utilizaron técnicas para el análisis de los datos como regresión lineal la cual permite predecir el comportamiento de una variable (dependiente) a partir de otra independiente. (Dagnino, 2014).

Se utilizaron métodos de análisis descriptivos para el estudio de los datasets. Este método busca hacer generalizaciones o predicciones sobre una población a partir de una muestra. (Monroy)

Evaluación

La evaluación se realiza durante el desarrollo del modelo y antes de que se implemente, esto permite evaluar la calidad del modelo y así poder analizar si este cumple con la solicitud inicial. Adicionalmente se busca interpretar los resultados obtenidos buscando patrones, tendencias y relaciones entre cada una de las variables.

Implementación

En esta fase, y una vez que el modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio, esto puede hacerse por ejemplo cuando el analista recomienda acciones basadas en la observación del modelo y sus resultados, o por ejemplo aplicando el modelo a diferentes conjuntos de datos. (Galán, 2015, p. 32)

Retroalimentación

Esta fase de la metodología CRISP-DM no se implementó durante el análisis de los datos.

Sustentación teórica de la pregunta

La conservación de los ecosistemas se ha convertido en una preocupación mundial debido a las alarmantes cifras que demuestran la extinción de fauna silvestre y la pérdida de sus hábitats naturales. Colombia es uno de los países con mayor biodiversidad del mundo, por tal motivo se ha visto enfrentado a grandes desafíos en busca de la protección de su fauna silvestre. El departamento de Risaralda es un punto importante debido a su rica biodiversidad y al actuar de los seres humanos sobre sus ecosistemas.

Las leyes colombianas respecto a la conservación y protección de la fauna silvestre son robustas y explícitas e indican lo siguiente: “La biodiversidad del país, por ser patrimonio nacional y de interés de la humanidad, deberá ser protegida prioritariamente y aprovechada en forma sostenible.” (Ministerio de Ambiente y Desarrollo Sostenible, 1993).

El decreto reglamentario (Decreto 1608 de 1978) indica “asegurar el cumplimiento de las normas relacionadas con la protección, aprovechamiento y conservación de la fauna silvestre” (Ministerio de Ambiente y Desarrollo Sostenible, 1978).

Las autoridades encargadas de hacer vigilancia y control a estas actividades incluyen a la Policía Nacional, las Corporaciones Autónomas Regionales y otras entidades gubernamentales.

Realizar análisis de información haciendo uso de herramientas metodológicas tales como la metodología para la ciencia de datos CRISP-DM, ayuda a entender las incautaciones que se han hecho, examinando si estas incautaciones han sido efectivas. Estos análisis permiten indicar que municipios del departamento de Risaralda tienen mayor incidencia en el tráfico de fauna silvestre, permitiendo evidenciar áreas críticas donde se necesita fortalecer la vigilancia y la educación ambiental.

Estas herramientas de ciencia de datos permiten también identificar cuáles son las especies que más han sido incautadas durante los últimos años, permitiendo así priorizar esfuerzos para su conservación; realizar este tipo de análisis de datos ayuda también a reconocer que entidades gubernamentales están teniendo un papel más relevante, quienes está realizando las incautaciones y así revelar las necesidades de capacitación y recursos.

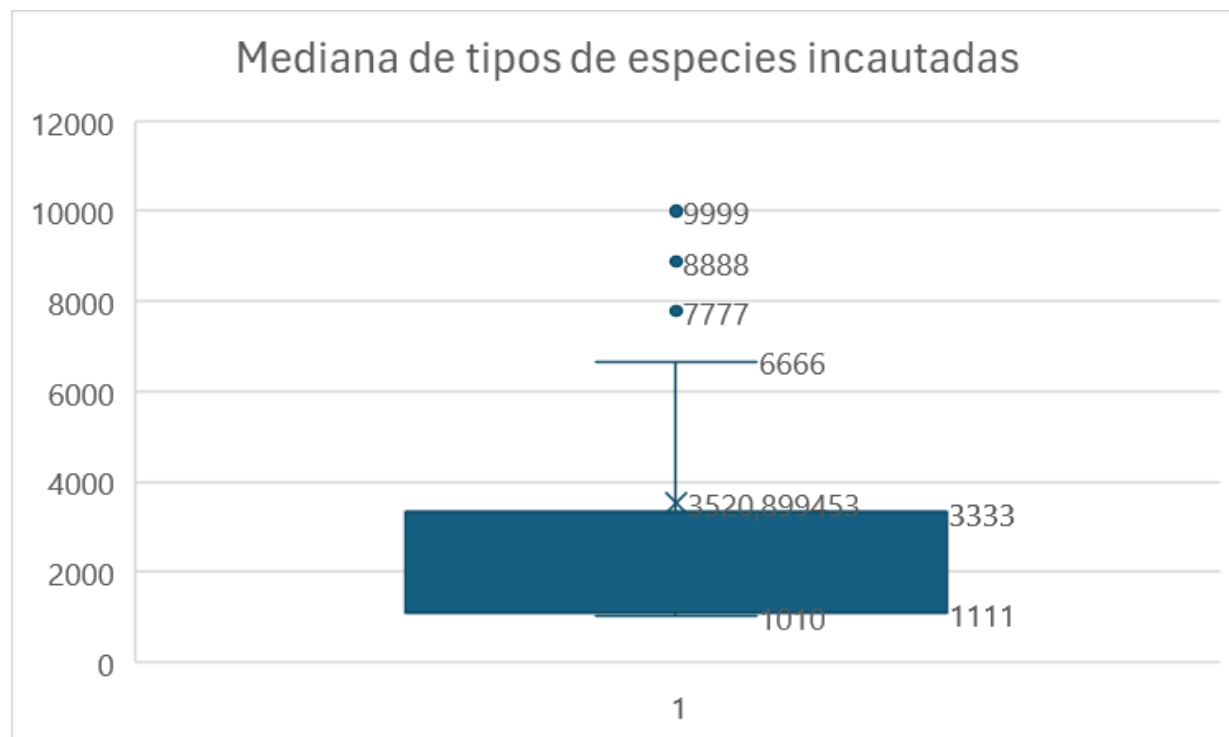
Las incautaciones de fauna silvestre pueden ocurrir en diferentes sitios, ya sean barrios, veredas, mercados, aeropuertos o incluso residencias privadas. Analizar y comprender estos datos utilizados mediante la ciencia de datos permite diseñar estrategias sólidas para así realizar intervenciones más efectivas y así enfocarse en realizar campañas de educación ambiental y concientización.

El análisis de los datos permite evidenciar las cantidades de animales y especies incautadas mostrando la magnitud del problema que presenta el tráfico ilegal de fauna silvestre, comparar estos datos a lo largo de los años ayuda a evaluar la efectividad de las políticas y estrategias implementadas y así poder mejorarlas según sea necesario. Este análisis riguroso mediante metodologías de ciencia de datos como CRISP-DM permitiría evidenciar cuales son los picos estacionales, fechas o eventos específicos que disparan el tráfico de fauna silvestre.

Los resultados obtenidos con este análisis no solo contribuirán con la conservación local, sino que también ayudarán a proporcionar datos valiosos para ahondar en esfuerzos de conservación tanto en el departamento como a nivel nacional.

Gráfico 2

Tipos de Especies Incautadas

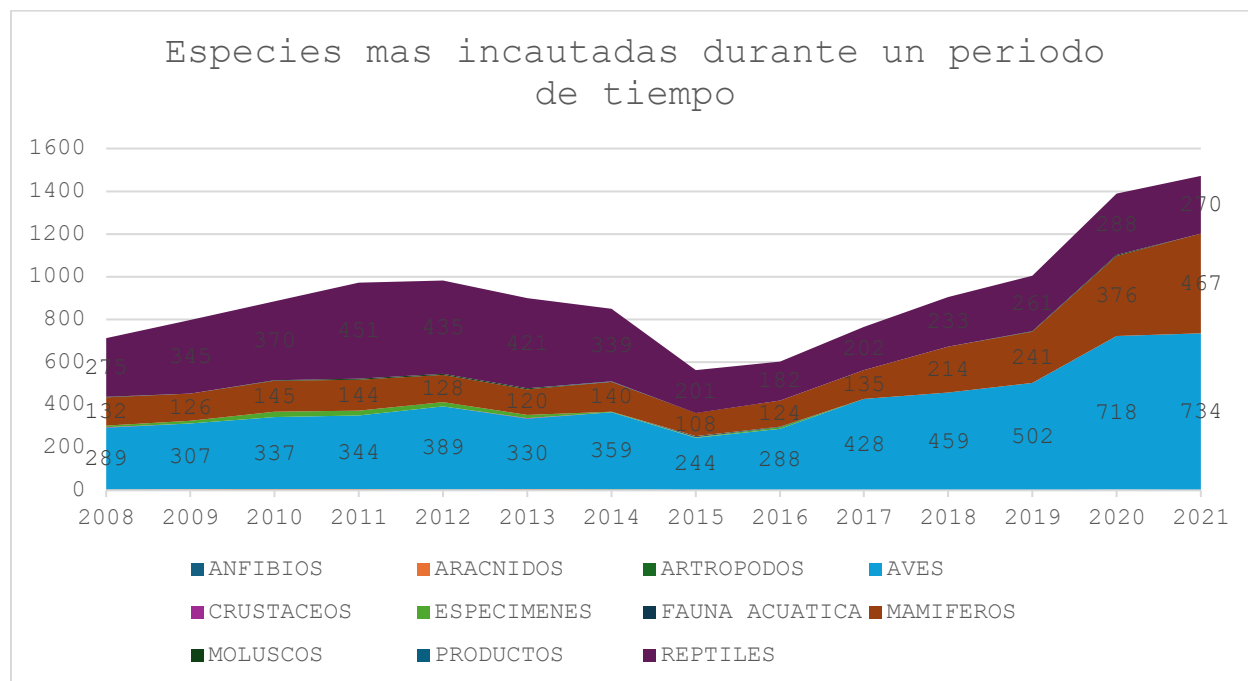


Fuente: (Autoría Propia)

Nota. El gráfico de cajas y bigotes permite mostrar la variación desde el mínimo hasta el máximo, en este caso podemos evidenciar cual es la mediana de la especie más incautada, este análisis sobre las incautaciones de fauna silvestre en Risaralda permite identificar patrones, tendencias sobre cuales son las especies más incautadas y así tomar medidas más focalizadas para ayudar a proteger la biodiversidad en Risaralda.

Gráfico 3

Especies más Incautadas durante el 2008 y el 2021



Fuente: (Autoría Propia)

Nota. El gráfico de áreas apiladas permite mostrar los tipos de especies más incautadas y cómo van aumentando o disminuyendo sus incautaciones durante un periodo de tiempo, el cual está comprendido entre los años 2008 y 2021.

Tabla 1*Análisis Sobre Incautaciones de Fauna Silvestre*

Estadísticas de la regresión	
Coefficiente de correlación múltiple	0,045782222
Coefficiente de determinación R ²	0,002096012
R ² ajustado	0,001627952
Error típico	2,745582856
Observaciones	12799

ANÁLISIS DE VARIANZA							
	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F		
Regresión	6	202,5408041	33,75680069	4,478083331	0,000154848		
Residuos	12792	96428,97697	7,538225217				
Total	12798	96631,51777					

	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 97,0%	Superior 97,0%
Intercepción	6,155120317	7,780412385	0,791104637	0,428897601	-9,095650751	21,40589138	-10,73096187	23,0412025
66	0	0	65535	#¡NUM!	0	0	0	0
66001000	-8,33694E-08	1,17751E-07	-0,708011936	#¡NUM!	-3,1418E-07	1,47441E-07	-3,38929E-07	1,7219E-07
1	0,118566466	0,035286261	3,360131209	0,000781333	0,049400121	0,187732811	0,041983546	0,195149386
3333	3,36061E-05	1,03521E-05	3,246297113	0,001172176	1,33144E-05	5,38978E-05	1,11385E-05	5,60736E-05
983	0,000143181	0,000167153	0,856582521	0,391691695	-0,000184465	0,000470826	-0,000219598	0,000505959
55	0,00343236	0,001870853	1,834649749	0,066580812	-0,000234792	0,007099512	-0,000628013	0,007492734

Fuente: (Autoría Propia)

Nota. Se realizó un análisis de regresión lineal para predecir la cantidad de incautaciones de fauna silvestre en Risaralda basado en las variables año, código municipio, código de tipo de especie y código de autoridad que incauta.

- La variable dependiente es la cantidad de incautaciones.

Las variables independientes incluyen el año, el código de municipio, el código de tipo de especie y el código de la autoridad que incauta.

Tabla 2

Análisis de Incautaciones de Fauna Silvestre en Risaralda: Tendencias y Concentración Geográfica (2008-2021)

Año		Cantidad		Codigo de departamento		Codigo municipio	
Media	2015,07578	Media	1,23773438	Media	66	Media	66107216,17
Error típico	0,03747806	Error típico	0,02428659	Error típico	0	Error típico	1826,393197
Mediana	2015	Mediana	1	Mediana	66	Mediana	66001000
Moda	2021	Moda	1	Moda	66	Moda	66001000
Desviación estándar	4,24015838	Desviación estándar	2,74771388	Desviación estándar	0	Desviación estándar	206632,8023
Varianza de la muestra	17,9789431	Varianza de la muestra	7,54993158	Varianza de la muestra	0	Varianza de la muestra	42697114997
Curtosis	-1,36152457	Curtosis	1945,37947	Curtosis	#¡DIV/0!	Curtosis	2,385577419
Coefficiente de asimetría	-0,09138968	Coefficiente de asimetría	40,5505533	Coefficiente de asimetría	#¡DIV/0!	Coefficiente de asimetría	1,9405293
Rango	13	Rango	149	Rango	0	Rango	686000
Mínimo	2008	Mínimo	1	Mínimo	66	Mínimo	66001000
Máximo	2021	Máximo	150	Máximo	66	Máximo	66687000
Suma	25792970	Suma	15843	Suma	844800	Suma	8,46172E+11
Cuenta	12800	Cuenta	12800	Cuenta	12800	Cuenta	12800
Mayor (1)	2021	Mayor (1)	150	Mayor (1)	66	Mayor (1)	66687000
Menor(1)	2008	Menor(1)	1	Menor(1)	66	Menor(1)	66001000
Nivel de confianza(97,0'	0,08133985	Nivel de confianza(97,0'	0,05270997	Nivel de confianza(97,0%)	0	Nivel de confianza(97,0'	3963,880346

Fuente: (Autoría Propia)

Nota. El análisis estadístico muestra que las incautaciones de fauna silvestre en Risaralda se realizaron principalmente entre el 2008 y el 2021, con una media alrededor del año 2015, la mayoría de las incautaciones involucraron una sola especie y que las actividades se concentraron en algunos municipios del departamento.

Tabla 3*Análisis de Correlación de Incautaciones de Fauna Silvestre en Risaralda*

	Cantidad	Año	Codigo de departamento	Codigo municipio	Codigo situacion	Codigo tipo especie	Codigo nombre comun	Codigo autoridad incauta
Cantidad	1							
Año	-0,02525923	1						
Codigo de departamento	#jDIV/0!	#jDIV/0!	1					
Codigo municipio	-0,00315292	0,079649976	#jDIV/0!	1				
Codigo situacion	0,03222405	0,368707805	#jDIV/0!	0,03897107	1			
Codigo tipo especie	0,02990066	0,171422211	#jDIV/0!	0,05940002	0,11475388	1		
Codigo nombre comun	0,00293258	-0,083555107	#jDIV/0!	-0,01069532	-0,02592522	-0,1605069	1	
Codigo autoridad incauta	0,01305641	-0,284420821	#jDIV/0!	0,01761924	-0,03760288	-0,07850728	0,042418978	1

Fuente: (Autoría Propia)

Nota. El análisis de correlación de los datos de incautaciones de fauna silvestre en Risaralda revela que no hay correlaciones significativas entre la mayoría de las variables. Esto significa que la cantidad de incautaciones de fauna silvestre no parece estar relacionada con ningún factor en particular. Sin embargo, se observó que hay una correlación negativa débil entre el nombre común de la especie incautada y la autoridad que incauto la fauna silvestre. Esta correlación sugiere que la cantidad de incautaciones de fauna silvestre de algunas especies comunes pueden ser mayor por parte de una autoridad que por otra.

Conclusiones.

En conclusión, el estudio sobre las incautaciones de fauna silvestre en el departamento de Risaralda es fundamental para analizar, comprender y así poder abordar el tráfico ilegal de especies. Este análisis de datos mediante a metodologías de ciencia de datos como CRISP-DM permitirá identificar que especies se han incautado más durante los últimos años, que entidades gubernamentales han influido en dichas incautaciones, que municipios tienen los más altos índices de trata de especies protegidas y los principales desafíos y oportunidades para así poder mejorar la protección de la biodiversidad en la región, “fortaleciendo las capacidades de las autoridades y fomentando una mayor conciencia pública sobre la importancia de conservar nuestra fauna silvestre”. (Chavarro, 2022)

Referencias

- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinnartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0.
- Chavarro Cadesa, J. E. (2022). *Medio ambiente: licencias y protección de los recursos naturales (3a. ed.)*. Grupo Editorial Nueva Legislación SAS.
<https://elibro.net/es/lc/remington/titulos/68882>
- Dagnino, J. (2014). *Regresión Lineal*. Rev Chil Anest.
- Espinosa Zuñiga, J. J. (03 de agosto de 2020). *Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública*. Scielo:
https://www.scielo.org.mx/scielo.php?pid=S1405-77432020000100008&script=sci_arttext#B6
- Galán Cortina, V. (octubre de 2015). *Aplicación de la Metodología CRISP-DM a un Proyecto de Minería de Datos en el Entorno Universitario*. Universidad Carlos III de Madrid:
https://www.academia.edu/download/49071533/PFC_Victor_Galan_Cortina_DM.pdf
- IBM. (17 de 08 de 2021). *Guía de CRISP-DM de IBM SPSS Modeler*. IBM:
<https://www.ibm.com/docs/es/spss-modeler/saas?topic=understanding-data-overview>
- Ministerio de Ambiente y Desarrollo Sostenible. (31 de julio de 1978). *Ministerio de Ambiente*.
<https://www.minambiente.gov.co/documento-entidad/decreto-1608-de-1978/>
- Ministerio de Ambiente y Desarrollo Sostenible. (22 de diciembre de 1993). *Ministerio de Ambiente*. <https://www.minambiente.gov.co/documento-normativa/ley-99-de-1993/>
- Monroy Saldívar, S. (2008). *Estadística descriptiva*. Insituto Politécnico Nacional:
<https://elibro.net/es/ereader/remington/74722?page=1>