

Dirección de Desarrollo Digital

Unidad de Científicos
de Datos



**DEPARTAMENTO
NACIONAL DE PLANEACIÓN**



ÍNDICE PARA LA PRIORIZACIÓN DE ÁREAS PROTEGIDAS DEL RUNAP

INFORME FINAL

Dependencias y entidades involucradas	Departamento Nacional de Planeación <ul style="list-style-type: none">• Dirección de Economía Naranja y Desarrollo Digital - Unidad de Científicos de Datos•
Sector	Defensa
Tecnologías utilizadas	R, Power BI
Fuentes de datos	RUNAP

Contenido

1. Presentación	2
2. Objetivos del proyecto	3
3. Metodología	3
4. Resultados	19
5. Conclusiones y recomendaciones	22



1. Presentación

Actualmente el país cuenta con 170 áreas protegidas, en las cuales, se generan diferentes tipos de delitos ambientales, de acuerdo con sus recursos naturales propios; algunos ejemplos de estos delitos son: tala ilegal de árboles, caza de especies en vía de extinción o minería ilegal, lo cual genera daños e invasiones en estas áreas (Colombia, 2021). Por lo anterior, se manifiesta la necesidad de controlar estos delitos, en lo cual surge la idea de generar una herramienta de visualización en el cual se pueda detectar en cuales áreas existe un mayor riesgo de delitos ambientales, a través del desarrollo de un indicador que recoja la información de variables estratégicas que permita una aproximación acertada en el riesgo, y de esta forma prevenir nuevos delitos en estas áreas.

De acuerdo con lo anterior, se desarrolló un indicador con las variables INCAUTACIONES, CAPTURAS, DELITOS, HECTÁREAS, DRENAJES, dada la importancia de cada una de ellas en la toma de decisiones para el cuidado de cada una de las áreas. Se analizaron los datos para 170 áreas protegidas que contara con información.

Se aplicó un método de PCA para determinar el número de componentes en que se agrupan las variables y la dimensión de estas componentes para generar un peso a las variables. De acuerdo con esto, se desarrolla un modelo de regresión lineal para predecir el riesgo en cada una de las áreas.

Finalmente, se desarrolla un visualizador en la herramienta Power BI, el cual permite identificar en un mapa de área protegidas, cual es el riesgo para cada una de ellas; también permite visualizar el total de delitos, capturas, incautaciones o drenajes para cada una de las áreas, lo cual permite identificar o asociar los delitos, capturas, incautaciones o drenajes las áreas protegidas con su nivel de riesgo.

Presentation:

Currently the country has 170 protected areas, in which different types of environmental crimes are generated, according to their own natural resources, some examples of these crimes are: illegal felling of trees, hunting of endangered species or mining illegal, which generates damage and invasions in these areas (Colombia, 2021). Due to the above, the need to control these crimes is manifested, in which the idea of generating a visualization tool arises in which it can be detected in which areas there is a greater risk of environmental crimes, through the development of an indicator that collect the information of strategic variables that allow a correct approximation in the risk, and in this way prevent new crimes in these areas.

In accordance with the above, an indicator was developed with the variables SEIZURES, CATCHES, CRIMES, HECTARES, DRAINAGES, given the importance of each of them in decision-making for the care of each one of the areas. Data was analyzed for 170 protected areas that will have information.

A PCA method was applied to determine the number of components in which the variables are grouped and the dimension of these components to generate a weight for the variables. Accordingly, a linear regression model is developed to predict the risk in each of the areas.

Finally, a viewer is developed in the Power BI tool, which allows identifying on a map of protected areas, what is the risk for each of them; It also allows visualizing the total crimes, captures, seizures or drainages for each of the areas, which allows identifying or associating crimes, captures, seizures or drainages of the protected areas with their level of risk.



2. Objetivos del proyecto

2.1. General

Desarrollar una herramienta para identificar las amenazas ambientales y de seguridad en las áreas protegidas que contribuya a la toma de decisiones de la Fuerza Pública para su intervención y protección.

2.2. Específicos

1. Creación de un indicador que priorice el nivel de riesgo o amenaza para cada una de las áreas protegidas del RUNAP.
2. Desarrollar un visualizador que permita identificar las áreas por su nivel de riesgo, así como la información asociada a cada una de estas áreas.

3. Metodología

Inicialmente se realiza un análisis descriptivo de los datos, para que de esta forma se entienda como están compuestas las bases entregadas por el RUNAP: INCAUTACIONES, CAPTURAS, DELITOS, HECTÁREAS, DRENAJES

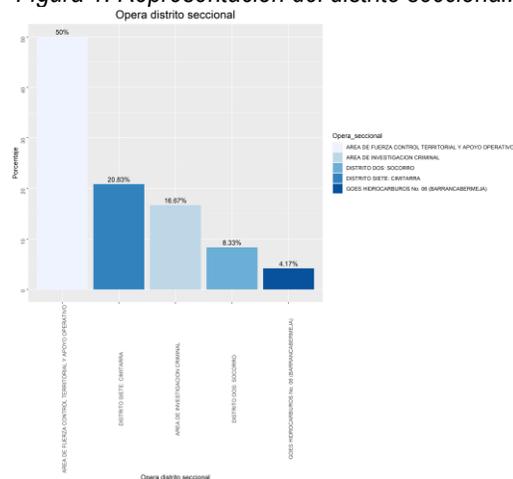
3.1. Análisis descriptivo de los datos:

Colombia suscribió el convenio de Diversidad Biológica a través de la Ley 165 de 1994, con base en la cual se formuló la Política Nacional de Biodiversidad y se adquirió el compromiso de conformar y consolidar un Sistema Nacional de Áreas Protegidas – SINAP. Colombia es uno de los cinco países con mayor diversidad biológica a nivel internacional y como parte del Convenio de Diversidad Biológica se encuentra comprometida en “establecer y mantener al año 2010 para las zonas terrestres y al año 2012 para las marinas, sistemas nacionales y regionales de áreas protegidas completos, eficazmente gestionados y ecológicamente representativos que contribuyan al logro de los objetivos del Convenio.”

3.1.1. Capturas

Dentro de las áreas protegidas se trabaja a diario para poder capturar a personas que no cumplen con los manuales de respeto a la fauna y flora y desde allí se generan capturas las cuáles se registran en el SINAP con todos los delitos que se presentan en estos lugares, a continuación, se hará una descripción de las capturas registradas.

Figura 1: Representación del distrito seccional:





La Serranía de los Yariguies, con 11 capturas lleva la delantera de el en listado de 24 áreas que han presentado capturas, donde también se puede evidenciar que el artículo 328: “ILÍCITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES” es el más vulnerado, y presentándose el mayor número de capturas en áreas protegidas con categoría Distritos Regionales de Manejo Integrado

Tabla 1: Cuenta del número de capturas por áreas

Área protegida	número de delitos en área	hectáreas por área
Serranía de los Yariguies	11	419,225
La Serranía de Chiribiquete	2	4,268,095
Kirpas Pinilla la Cuerera	2	294
Lago de Socha gota	2	8,150
Del Humedal San Silvestre	1	69,985
Campoalegre	1	21,132
Río Pure	1	999,880
Barbas Bremen	1	4,316
Serranía La Lindosa - Angosturas II	1	28,224
Palmar del Titi	1	2,622
Quebrada Honda y Canos Parrado y Buque	1	0
Total general	24	4,268,095

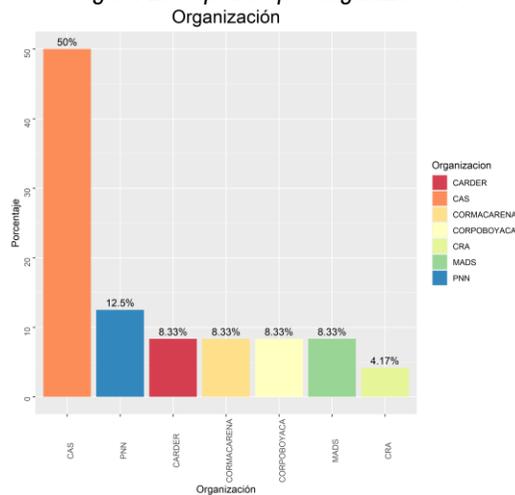
Tabla 2: Cuenta por número de conducta de captura

Conducta	Frecuencia de la conducta
ARTICULO 328. APROVECHAMIENTO ILICITO DE LOS RECURSOS NATURALES RENOVABLES	12
ARTÍCULO 332. EXPLOTACIÓN ILÍCITA DE YACIMIENTO MINERO Y OTROS MATERIALES	6
ARTICULO 333. DANOS EN LOS RECURSOS NATURALES Y ECOCIDIO	5
ARTICULO 339. DANOS EN LOS RECURSOS NATURALES Y ECOCIDIO (MODALIDAD CULPOSA)	1
Total general	24

La organización CAS cuenta con el 50% del manejo de estas capturas siendo esta la de mayor representatividad en el enlistado:

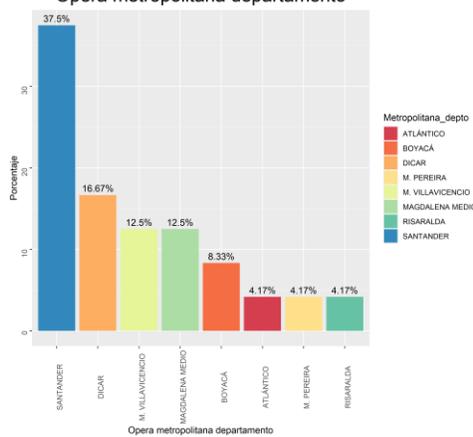


Figura 2: Capturas por organización.



El departamento de Santander es el más afectado con 12 capturas dentro del enlistado seguido de Villavicencio y magdalena medio a continuación en la siguiente grafica podemos ver cuáles son las organizaciones más representativas por área protegida:

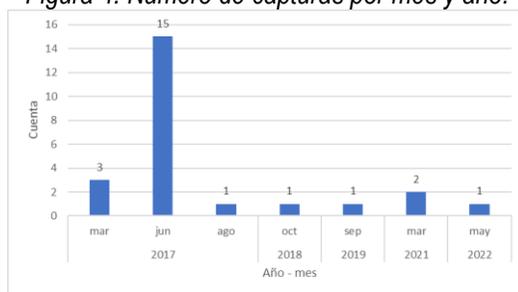
Figura 3: Capturas por departamento.



Según el reporte de Capturas el municipio con más casos es el de Landázuri con un 16,67% de las capturas, seguido de Barrancabermeja y Villavicencio, el año con más reportes fue el 2017.



Figura 4: Número de capturas por mes y año.



Haciendo un estudio de correlación entre las hectáreas y hechos cometidos en los diferentes lugares, se evidencia que no existe una correlación entre estos mismos, donde alas hectáreas más grandes presentan en promedio 2 hechos y la que más hechos presenta tiene una extensión mucho más pequeña como lo podemos ver en el siguiente Gráfico:

Figura 5: Número de capturas por mes y año.

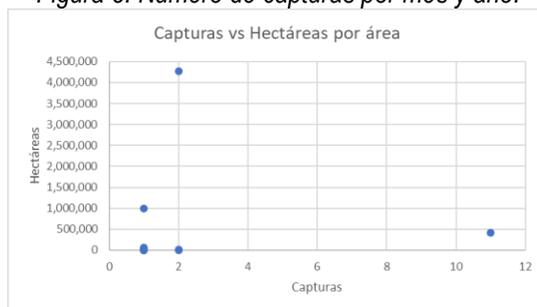


Tabla 3: Correlación Hectáreas – número de capturas

	<i>número de delitos en área</i>	<i>hectáreas por área</i>
número de delitos en área	1	
hectáreas por área	0.758013229506984	1

3.1.2. Delitos

Dentro de los delitos la Serranía de los Yariquies sigue siendo la de más hechos registrados llevándose un 24,66% de peso entre la totalidad de delitos reportados seguido del Humedal de san silvestre con 10,1% y Amacayacu con 3,9 donde el género masculino es el que representa el 60% de los delitos, junto con los distritos de manejo integrado que tratan el 60% de los casos.



Figura 6: Áreas con mayores delitos - top 20



Figura 7: Causas de los delitos
Conteo de crimen por causa

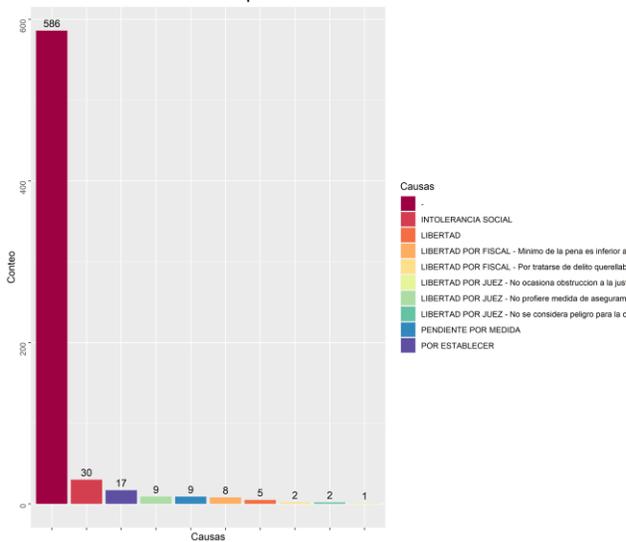
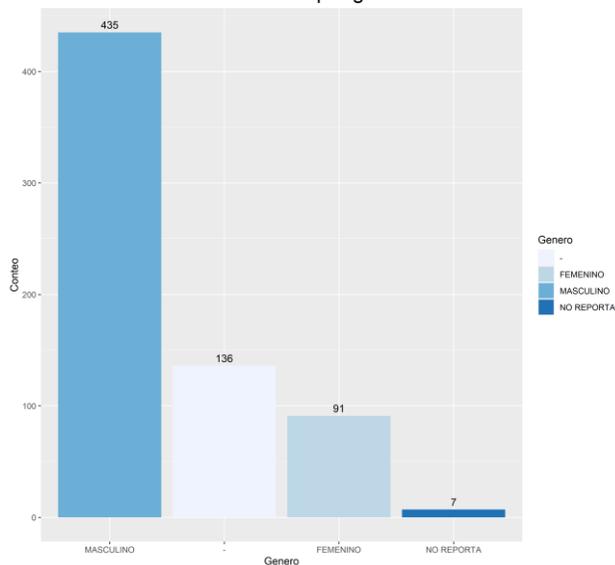


Figura 8: Conteo de crimen por sexo
Conteo de crimen por genero



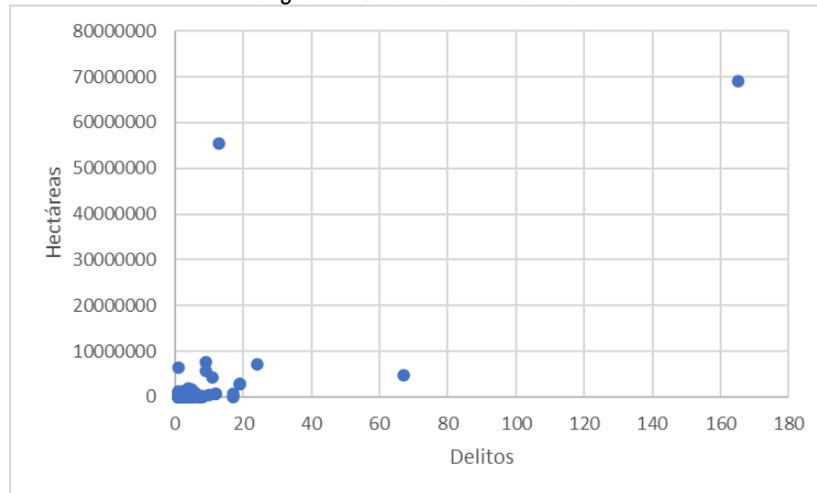
Los delitos por hectárea si están correlacionados positivamente por su extensión, es decir que hay más delitos en las áreas con mayor número de hectáreas:

Tabla 4: Matriz de correlación entre delitos y hectáreas por área.

	número de delitos en área	hectáreas por área
número de delitos en área	1	
hectáreas por área	0.758013229506984	1



Figura 9: Delitos vs Hectáreas



Los distritos regionales de manejo integrado son las categorías con más delitos son 403 seguidos de Parque Nacional Natural con 116 casos y en tercer lugar Reservas Forestales Protectoras Nacionales:

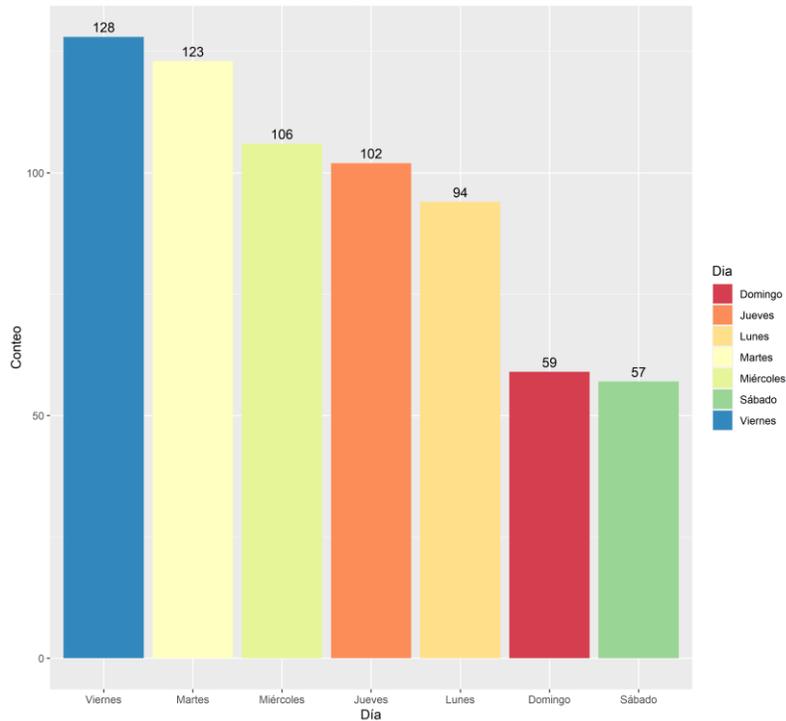
Tabla 5: Cuenta de delitos por categoría

Categoría	Cuenta de categoría
Distritos Regionales de Manejo Integrado	403
Parque Nacional Natural	116
Reservas Forestales Protectoras Nacionales	72
Distritos de Conservación de Suelos	23
Parques Naturales Regionales	16
Reservas Forestales Protectoras Regionales	11
Reserva Natural	10
Santuario de Fauna y Flora	6
Vía Parque	4
Reserva Natural de la Sociedad Civil	3
Áreas de Recreación	3
Distritos Nacionales de Manejo Integrado	1
Santuario de Flora	1
Total, general	669

La mayor cantidad de delitos se presentaron el viernes con un 19%, seguido de los martes con un 18,30% y los miércoles con un 15%:

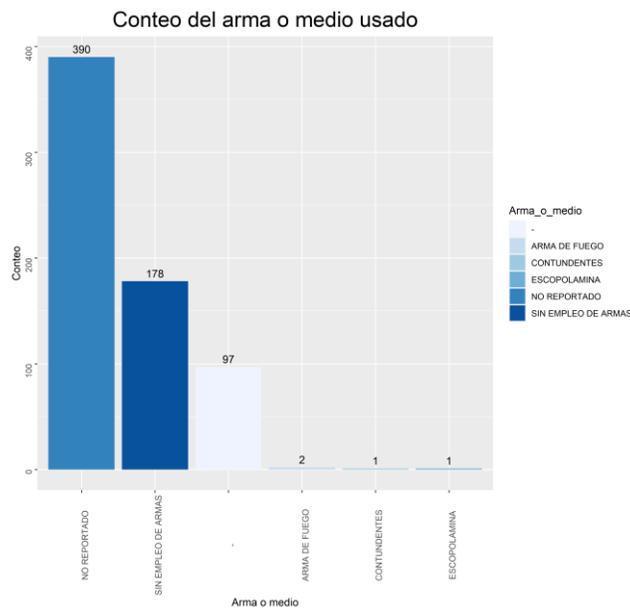


Figura 10: Delitos por día de la semana.
Conteo de delitos por día de la semana



Dentro de la revisión de delitos, otra de las variables que se encuentra despoblada es la de causas, donde aún se encuentra oportunidad para poder identificar por qué, estos hechos quedan impunes:

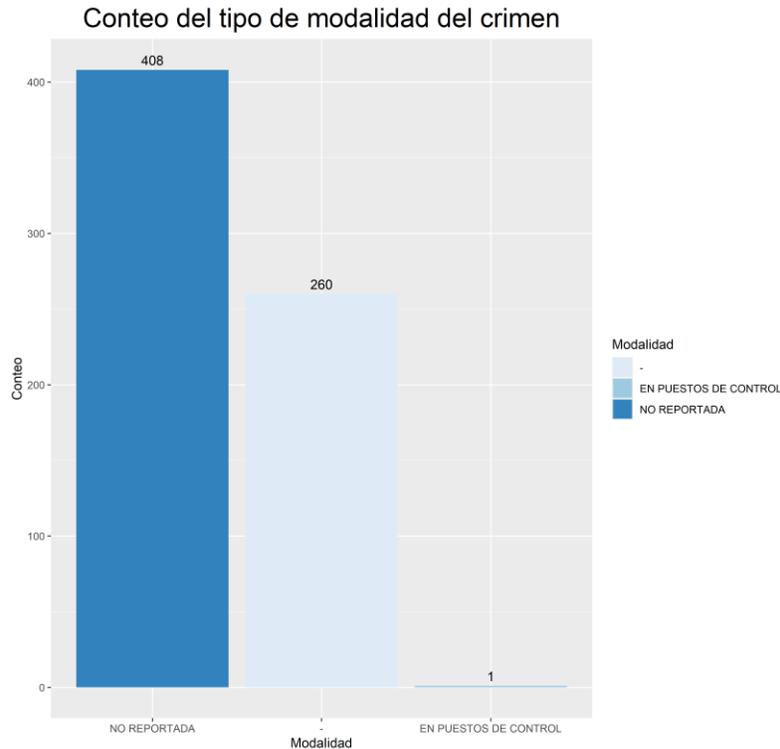
Figura 11: Medio de ejecución del delito.





Los hechos también tienen una característica importante que es la nacionalidad de la persona que afecta el área donde su gran mayoría son colombianos, pero también quedan algunos otros extranjeros los cuales son tienen un reporte exacto de nacionalidad, pero afectan las zonas protegidas. El artículo 328: “ILÍCITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES” es el más reiterativo con un 33.78% de casos de todo el en listado, es de resaltar que la gran mayoría de avisos se hacen por denuncias.

Figura 12: Medio de reporte del delito



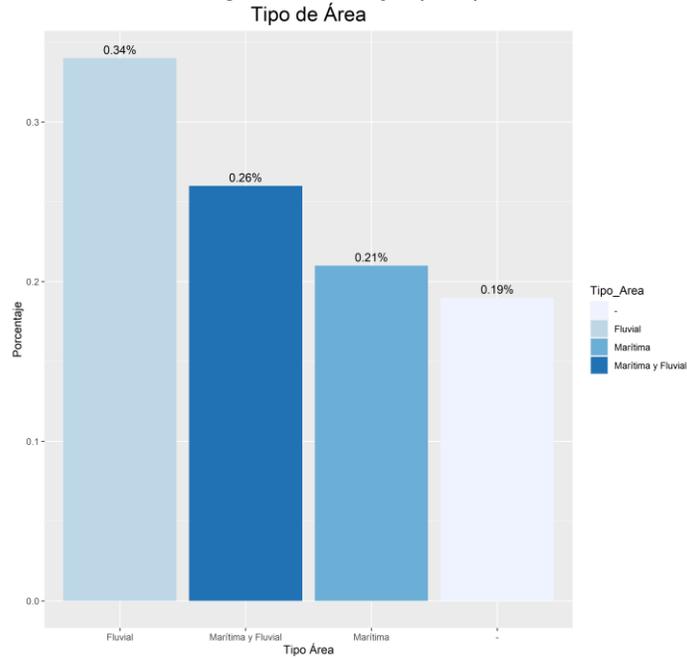
3.1.3. Drenajes

El paso de drenaje por las áreas protegidas es de gran importancia pues las afectaciones pueden generar un daño transversal e irreparable, es por lo que se hace necesaria una revisión de los drenajes presentes dentro de cada una de las áreas protegidas, donde Sanquianga ocupa el primer lugar seguido de encanto de los Manglares del bajo Baudó. Los hechos dentro de estas zonas son representados en gran medida por organizaciones como PNN y CODECHOCO, los registros más altos se incidencias se presentan en el año 2017.

Dentro de las áreas más afectadas se encuentran las Fluviales, luego las marítimas-fluviales y por último Marítimas:

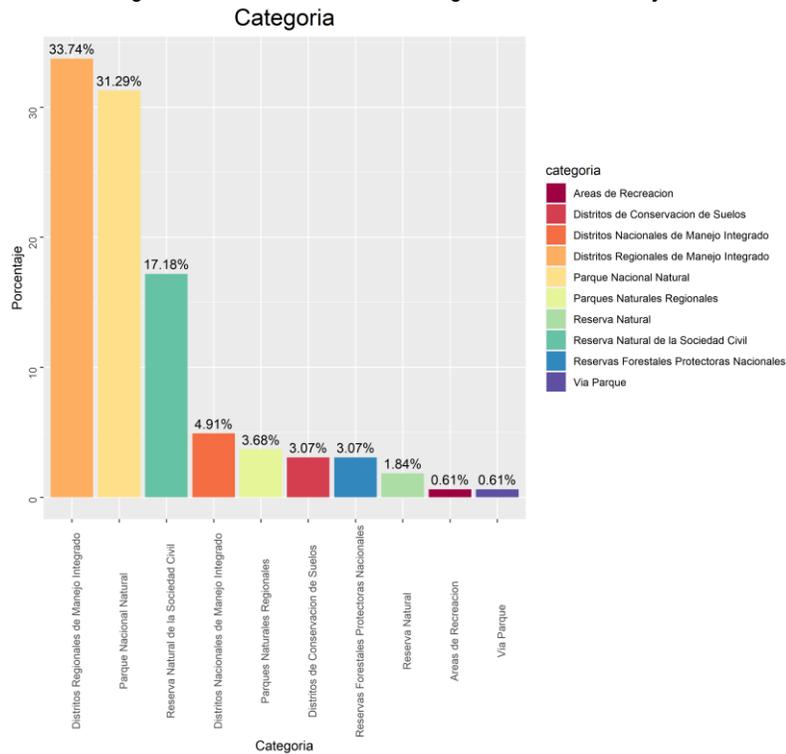


Figura 13: Drenajes por tipo de área.



Dentro de las categorías los Distritos Regionales de Manejo integrado y Parque Nacional Natural llevan la delantera con más oportunidades hídricas:

Figura 14: Recuento de las categorías de los drenajes:



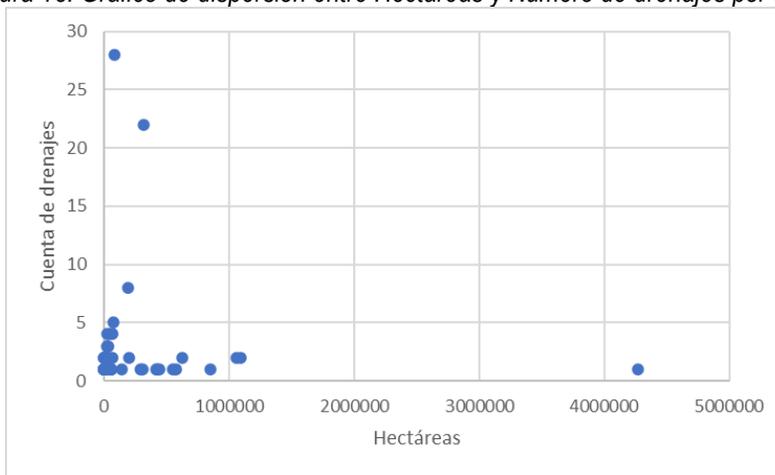


En la verificación de Hectáreas por número de drenajes cercanos a estas mismas se evidencia que no hay correlación entre los hechos reportados y la magnitud de estas áreas:

Tabla 5: Matriz de correlación entre drenajes y hectáreas por área.

	hectáreas_	Drenajes
hectáreas_	1	
Drenajes	0.00893351	1

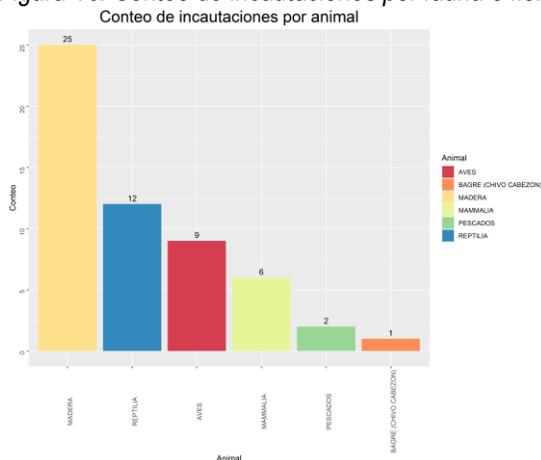
Figura 15: Gráfico de dispersión entre Hectáreas y Número de drenajes por área.



3.1.4. Incautaciones

Las incautaciones reportan la cantidad de Animales salvados en la zona, dentro de este enlistado se evidencia como el animal más afectado el Madera con 25 incautaciones seguido de reptiles y aves con un total de 55 registros de incautaciones.

Figura 16: Conteo de incautaciones por fauna o flora





Donde Lorica y Barrancabermeja tienen la mayor cantidad de reportes y el área más afectada sigue siendo la Serranía de los Yariquies donde se presenta un total de 12 casos, seguido del complejo Cenagoso de bajo Sinú. Los distritos regionales siguen siendo las categorías con mayor presencia teniendo un 67% de participación en los delitos

Figura 17: Conteo de incautaciones por categoría
Conteo de incautaciones por categoría

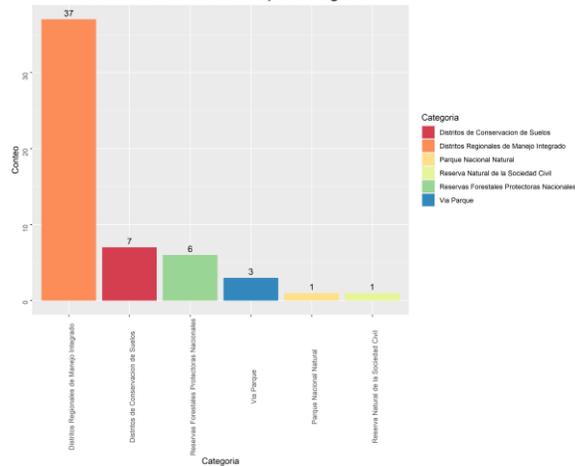
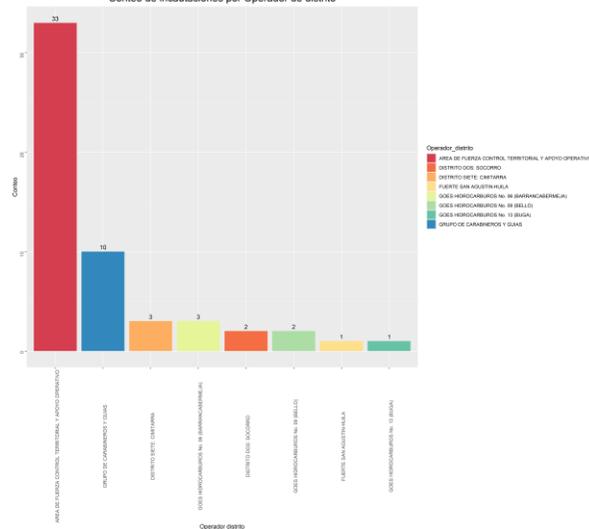


Figura 18: Incautaciones por operador de distrito
Conteo de incautaciones por Operador de distrito



En el gráfico anterior podemos evidenciar las unidades que más apoyan por área para la defensa y prevención de estos delitos.

Teniendo en cuenta la información anterior, es de resaltar que 37 de los delitos tienen una categoría de Distritos regionales de manejo integrado, y haciendo un análisis de correlación entre los hechos y las hectáreas de cada uno

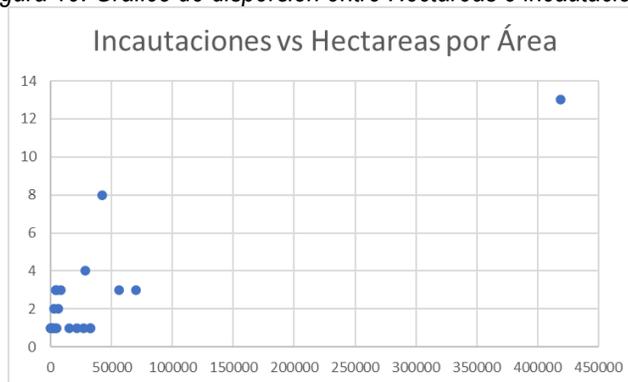


de estos delitos si se tiene una relación, mientras más grande es el territorio más son las incautaciones que se han realizado

Tabla 6: Matriz de correlación entre hectáreas e incautaciones

	Hectáreas	Número de Delitos
Hectáreas	1	
Número de Delitos	0.85123669	1

Figura 19: Gráfico de dispersión entre Hectáreas e incautaciones



3.2. Desarrollo del indicador:

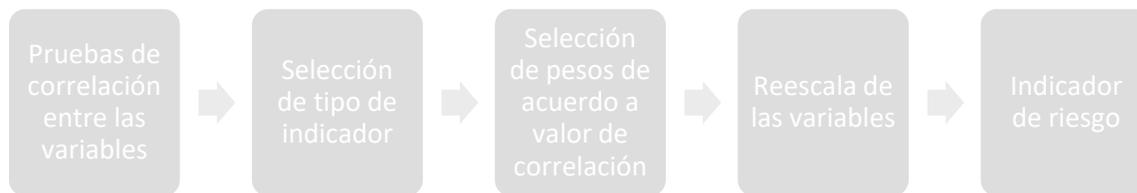
Se reemplazaron por 0 los valores no encontrados para cada una de las áreas, según sus variables:

- Incautaciones: 148 registros no encontrados.
- Delitos: 53 registros no encontrados.
- Drenaje: 96 registros no encontrados.
- Capturas: 159 registros no encontrados.

De acuerdo con la completitud de los datos debido al total de registros encontrados, se propone realizar un análisis de componentes principales (PCA) con el fin de determinar cuál es la representatividad de cada una de las variables (Rivas & Martínez Arias, 1991) con las que se cuenta para construir el índice de priorización de áreas protegidas; con esto, se pretende realizar un modelo de regresión lineal, a través de la posible correlación de las variables de las cuales suponemos, tendrán influencia (Granados, 2016) para calcular el índice de priorización para cada una de las áreas con las que se cuenta información.

Algunos de los beneficios del modelo PCA es que no requiere una suposición de normalidad multivariada para los datos puesto que dentro de las aplicaciones del PCA se encuentra la normalización de estos (Tripp, s.f.), además, el PCA es utilizado para cálculo de variables latentes (Camacho & Pico Marco, 2010), como es el caso de índice de priorización de áreas protegidas.

Figura 20: Flujo para la creación del indicador de riesgo



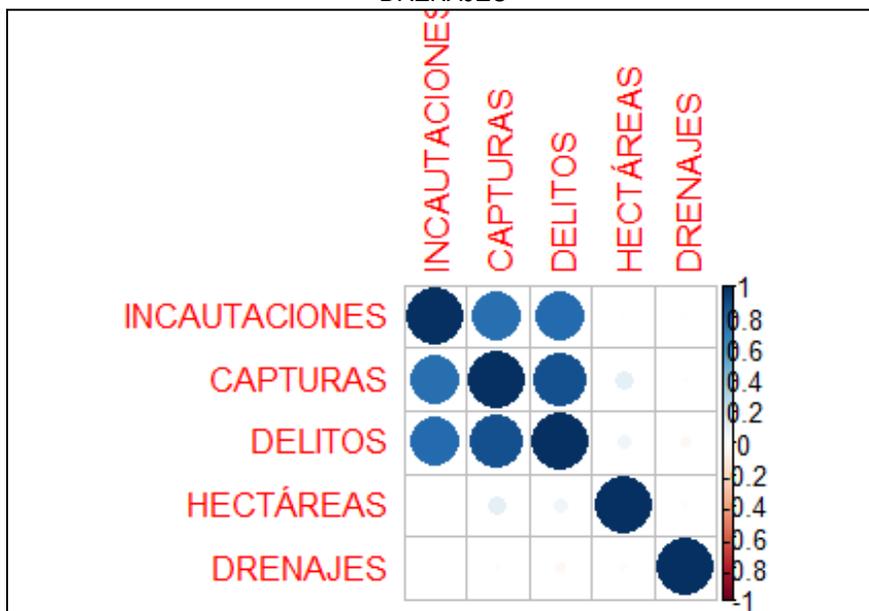
Fuente: Elaboración propia

Inicialmente, se realiza un conteo de las variables: INCAUTACIONES, CAPTURAS, DELITOS, HECTÁREAS, DRENAJES, para cada área protegida. (Base: "Variables para reescalar.xlsx")

Utilizamos el programa R para realizar el análisis de las variables:

Cargamos los datos, y realizamos un análisis de correlación entre todas las variables, en el cual podemos observar que hay correlaciones diferentes a cero:

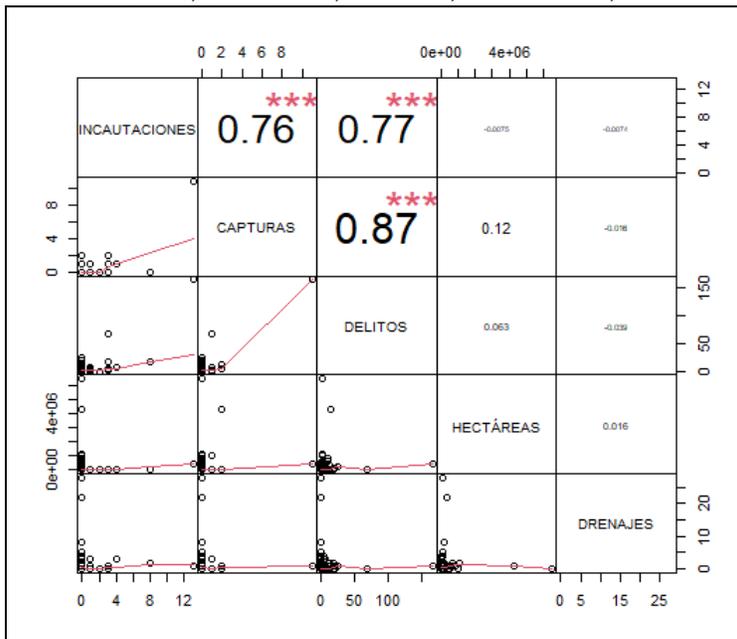
Figura 21: Gráfico de correlaciones entre las variables INCAUTACIONES, CAPTURAS, DELITOS, HECTÁREAS, DRENAJES



Fuente: Elaboración propia



Figura 22: Matriz de correlación y diagramas de dispersión bivariados con una línea ajustada para las variables INCAUTACIONES, CAPTURAS, DELITOS, HECTÁREAS, DRENAJES:



Fuente: Elaboración propia

Variables que presentan una correlación alta: #incautaciones + hectáreas + #delitos

Realizamos de igual manera la prueba de Kaiser_Meyer_Olkin (KMO) para calcular el valor de la Medida de adecuación de muestreo (MSA por sus siglas en inglés), el cual resulta en 0.73, lo cual significa que los datos serían idóneos para detectar estructuras o componentes:

Figura 23: Ejecución del código en R de la prueba Kaiser_Meyer_Olkin

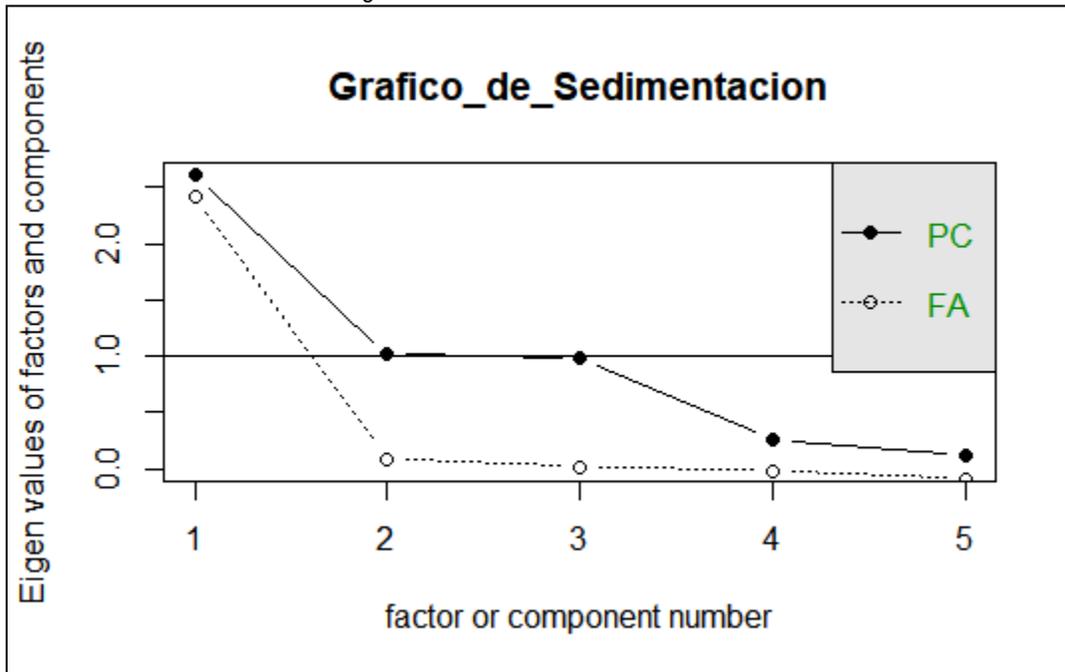
```
> kmo(datos1)
Kaiser-Meyer-Olkin factor adequacy
Call: kmo(r = datos1)
Overall MSA = 0.73
MSA for each item =
INCAUTACIONES    CAPTURAS    DELITOS    HECTÁREAS    DRENAJES
0.85            0.70        0.70        0.30         0.30
```

Fuente: Elaboración propia

Se realiza un gráfico de sedimentación para identificar cuantos posibles componentes podríamos utilizar para crear el indicador, que, de acuerdo con sus resultados, podríamos usar entre 2 o hasta 3 componentes, puesto que, hasta el tercer componente, está encima de la línea de rango aceptable:



Figura 24: Gráfico de Sedimentación.



Fuente: Elaboración propia

Se re escalan y centran los datos para realizar el análisis de componentes principales (PCA):

Figura 25: Ejecución del código en R para el escalamiento y centralización de los datos:

```
> componentes_cli <- prcomp(datos1, scale=TRUE, center = TRUE)
> componentes_cli
Standard deviations (1, .., p=5):
[1] 1.6154511 1.0096141 0.9920373 0.5116789 0.3536154

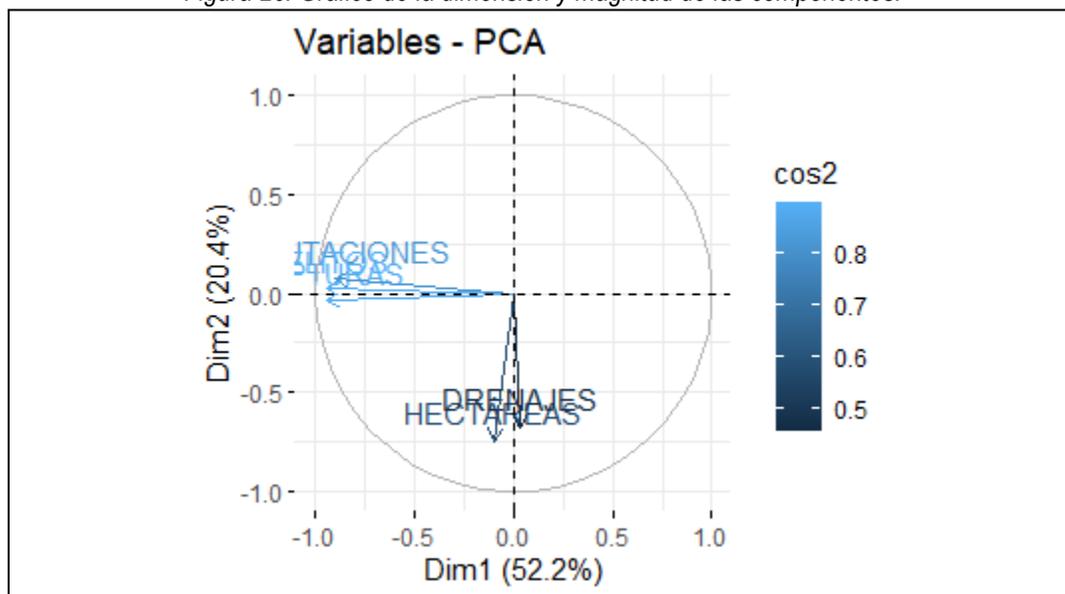
Rotation (n x k) = (5 x 5):
      PC1      PC2      PC3      PC4      PC5
INCAUTACIONES -0.55622408  0.07050765  0.10612887 -0.82070689  0.02864112
CAPTURAS      -0.58485954 -0.03781038 -0.01277294  0.41574751  0.69534204
DELITOS       -0.58658732  0.02532752  0.01106194  0.37614716 -0.71670411
HECTÁREAS     -0.06314888 -0.73858701 -0.66123830 -0.10759565 -0.04109174
DRENAJES      0.02194578 -0.66891464  0.74243859  0.02303723 -0.01805049
```

Fuente: Elaboración propia

Luego generamos una representación gráfica con vectores para los componentes:



Figura 26: Gráfico de la dimensión y magnitud de las componentes:



Fuente: Elaboración propia

De acuerdo con los resultados, tendríamos dos componentes:

Componente 1: Drenajes + Hectáreas.

Componente 2: Incautaciones + Capturas + Delitos.

El componente 1 representa 52.2% de la varianza de los datos y el componente 2 el 20.4%. En total, con los datos disponibles podríamos representar el 72.2% de la varianza de los datos.

Con el análisis inicial de PCA y sus resultados se propone construir el indicador de cero (0) a cinco (5) para la priorización de áreas protegidas (IPAP) donde cero (0) significaría ausencia de riesgo y cinco (5) el mayor riesgo; con la **metodología de porcentajes**.¹² de la siguiente forma:

$$IPAP = \frac{5}{1 + e^{-((INCAUTACIONES+CAPTURAS+DELITOS)*0.2 - (HECTÁREAS+DRENAJES)*0.5)}}$$

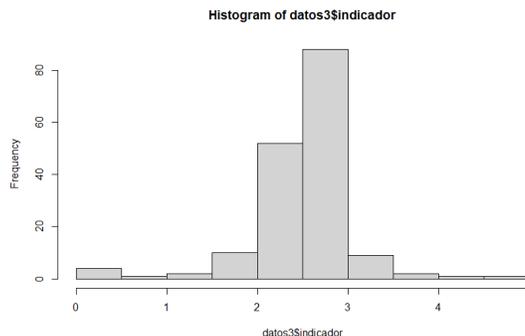
Generamos un histograma para visualizar la distribución del riesgo por áreas, en el cual podemos encontrar que, en su mayoría, las áreas tienen un riesgo entre 2 y 3.

¹ Consejo Nacional de Evaluación de la Política de Desarrollo Social. Manual para el Diseño y la Construcción de Indicadores. Instrumentos principales para el monitoreo de programas sociales de México. México, DF: CONEVAL, 2013.

² MARCANO, Arismar; CARTAYA, Scarlet; PACHECO, Henry y MENDEZ, Williams. Estimación de pesos ponderados de variables para generar mapas de susceptibilidad a movimientos en masa a través de la Evaluación Espacial Multicriterios. Terra [online]. 2015, vol.31, n.50, pp.55-80. ISSN 1012-7089.



Figura 7: Histograma de la distribución del indicador de riesgo



De acuerdo con estos resultados, se crea un modelo lineal para el indicador de riesgo por áreas:

Figura 8: Ejecución del código en R para la creación del modelo a través de una regresión lineal:

```
modelo <- lm(indicador$indicador ~ indicador$INCAUTACIONES +  
            indicador$CAPTURAS + indicador$DELITOS + indicador$HECTÁREAS +  
            indicador$DRENAJES)
```

Fuente: Elaboración propia

4. Resultados

A través del desarrollo metodológico descrito en la Sección 3, se obtuvieron los resultados que se presentan a continuación. Toda retroalimentación desde un punto de vista experto o de usuario por parte de Subdirección de seguridad y defensa, es bienvenida. Este insumo será de gran ayuda para mejorar la calidad y utilidad de los resultados obtenidos, de manera que agreguen mayor valor.

Se reemplazaron por 0 los valores no encontrados para cada una de las áreas, según sus variables:
notas

Incautaciones: 148 registros no encontrados.

Delitos: 53 registros no encontrados.

Drenaje: 96 registros no encontrados.

El resultado del modelo de regresión lineal es el siguiente:



Figura 9: Resultados del modelo de regresión lineal:

```

Residuals:
  Min       1Q   Median       3Q      Max
-0.62291 -0.05619  0.04169  0.06708  0.71801

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.630e+00  1.734e-02 151.690 < 2e-16 ***
indicador$INCAUTACIONES  1.899e-01  1.898e-02  10.003 < 2e-16 ***
indicador$CAPTURAS    -4.145e-02  3.656e-02  -1.134  0.25855
indicador$DELITOS     6.488e-03  2.389e-03   2.716  0.00731 **
indicador$HECTÁREAS  -4.863e-07  2.525e-08 -19.264 < 2e-16 ***
indicador$DRENAJES   -1.166e-01  5.326e-03 -21.897 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.198 on 164 degrees of freedom
Multiple R-squared:  0.8859,    Adjusted R-squared:  0.8824
F-statistic: 254.6 on 5 and 164 DF,  p-value: < 2.2e-16

```

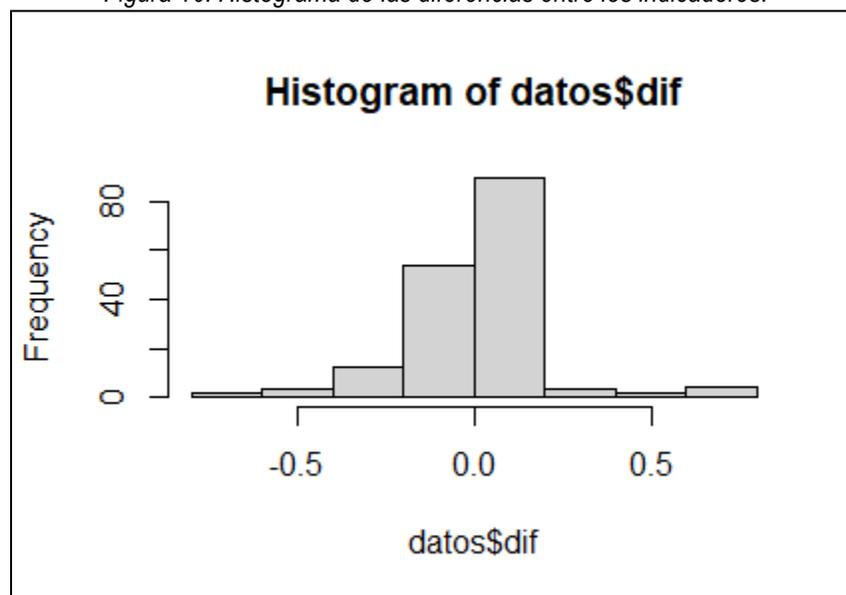
Fuente: Elaboración propia

$$\begin{aligned}
 \text{Indicador} = & 2.63 + 0.189(\text{INCAUTACIONES}) - 0.041(\text{CAPTURAS}) + 0.00648(\text{DELITOS}) \\
 & - 0.00000048(\text{HECTÁREAS}) - 0.116(\text{DRENAJES})
 \end{aligned}$$

De acuerdo con el primer coeficiente del modelo, un área que no tenga registros de delitos, capturas, incautaciones, drenajes o que su área sea 0, tendrá un riesgo dado de 2.6.

Con el modelo generado se realiza una predicción del indicador y se asigna como nueva variable a la data, se asigna la predicción a la base, se calculan las diferencias entre los dos indicadores y se genera un histograma para verificar la aproximación del modelo:

Figura 10: Histograma de las diferencias entre los indicadores:





Fuente: Elaboración propia

Desviación mínima: -0.62
Desviación máxima: 0.71
Desviación media: -0.62
Mediana: 0.041

Se propone clasificar los riesgos en tres categorías:

Tabla 5: Asignación del nivel de riesgo.

RIESGO	CLASIFICACIÓN
Mayor o igual a 4	Riesgo Alto
Mayor a 2 y menor a 4	Riesgo medio
Menor o igual a 2	Riesgo bajo

Fuente: Elaboración propia

Con la clasificación por categorías de riesgo las áreas se distribuyen de la siguiente forma:

Tabla 6: Recuento de áreas clasificadas por riesgo.

TOTAL DE ÁREAS	CLASIFICACIÓN
2	Riesgo Alto
151	Riesgo medio
17	Riesgo bajo

Fuente: Elaboración propia

En la siguiente tabla se muestran las áreas clasificadas con riesgo bajo y riesgo alto:

Tabla 7: Áreas clasificadas con riesgo bajo y riesgo alto.

ÁREAS RIESGO BAJO	ÁREAS CON RIESGO ALTO
Área Marina Protegida de la Reserva de Biosfera Seaflower	Serranía de los Yariguies
La Serranía de Chiribiquete	Del Humedal San Silvestre
Puinawai	
Yaigoje Apaporis	
Rio Pure	
Nukak	
Sierra de la Macarena	
El Tuparro	
Bosque Seco del Occidente Antioqueno	
Sanquianga	
Encanto de los manglares del Bajo Baudó	
Cabo Manglares Bajo Mira y Frontera	
Los Katios	
La Sierpe	
Uramba Bahía Málaga	



Ciénaga de Barbacoas	
Cahuinari	

Fuente: Elaboración propia

En general, en la unión de las bases del archivo “Consolidada 09082022.xlsx” se encontraron 169 áreas, lo cual, está alejado del total de áreas presentadas en la base “Calculo índice.xlsx” en el cual se presenta el indicador dado inicialmente por la Subdirección de seguridad y defensa.

Las variables se escalaron entre 0 y 5 se le dieron los pesos propuestos en el modelo, el archivo con el indicador y las variables escaladas se adjunta con este documento para que sea consultado, así como el código creado para este proceso en el programa R y los manuales de ejecución.

De acuerdo con el cálculo, el área con mayor riesgo es la Serranía de los Yariguies, seguido de Del Humedal San Silvestre.

Se adjunta a este documento el archivo “indicador&indicador_pred.csv”, en el cual están las áreas con el indicador generado para cada una de estas con la predicción del modelo, y el archivo “Escala de variables y cálculo de indicador.R” en cual se realizó el análisis de PCA el cálculo del indicador.

5. Conclusiones y recomendaciones

A partir de la metodología desarrollada y de los resultados obtenidos para cumplir los objetivos de este entregable, planteados en el plan de trabajo, se presentan a continuación las principales conclusiones obtenidas por el equipo de la UCD y las principales recomendaciones para la continuación del proyecto.

1. Se logró construir el índice de priorización de áreas protegidas a través del método PCA para asignar dimensiones y pesos a las variables, y posteriormente una regresión lineal para generar un modelo de predicción con las variables Incautaciones, Delitos, Capturas, Drenajes y Hectáreas.
2. El 88% de las áreas se encuentran en un riesgo medio, el 0.5% presentan un riesgo alto, y el 10% en riesgo bajo.
3. Se confirma que las variables Incautaciones, capturas, Delitos tienen correlaciones entre sí.
4. Se evidencia que no hay correlación entre el número de delitos/capturas/incautaciones con el tamaño del área protegida.
5. Las variables drenajes y Hectáreas son las que producen la mayor variabilidad al modelo, la cual se intenta capturar a través del método de PCA.
6. Se propone adicionar al primer cálculo del indicador una proporción de delitos por hectáreas para disminuir el efecto de la varianza de la variable HECTÁREAS, puesto que el rango de esta va desde cero (0) hasta seis millones quinientos un mil setecientos (6'501.700).
7. Se propone realizar un agrupamiento en el resultado de indicador para las áreas y de esta forma definirle su riesgo.
8. Se recomienda actualizar los datos de cada una de las variables con el RUNAP y hacerles recomendaciones sobre “data quality” para mejorar posibles análisis futuros e la información.
9. Se recomienda adquirir una licencia en Power BI Pro para publicar el visualizador del indicador en línea.