

Dirección de Desarrollo Digital

Unidad de Científicos
de Datos



**DEPARTAMENTO
NACIONAL DE PLANEACIÓN**



ANÁLISIS DE LA ENCUESTA TODOS POR LA AMAZONÍA

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• Consejería Presidencial para las Regiones
Sector	Territorial
Tecnologías utilizadas	Python
Fuentes de datos	Encuesta Todos por la Amazonía

Contenido

1. Presentación	2
2. Objetivos del proyecto	2
3. Metodología	2
4. Resultados	3
5. Conclusiones y recomendaciones	12
Contacto	12
6. ANEXOS	13



1. Presentación

El siguiente documento tiene como fin exponer los resultados encontrados en el análisis descriptivo y de texto en la encuesta “Todos por la Amazonía”, recolectada por el equipo de Consejería Presidencial para las Regiones. La encuesta se realizó a distintos grupos sociales en los diferentes departamentos del país; donde se encontraron tres tipos de preguntas: abiertas, selección única y selección múltiple, que tienen como objetivo ver qué tipo de soluciones encuentran los encuestados para ayudar a la reforestación de la Amazonía y cuál, según ellos, debe ser el acompañamiento de las organizaciones tanto públicas como privadas. Con este propósito se busca visibilizar los resultados de la encuesta para así realizar un análisis del texto enfocado principalmente en las preguntas abiertas y, a su vez, en el documento completo.

Abstract

The following document aims to expose the results found in the survey: "Todos por la Amazonía", developed by the Presidential Advisor's team for the Regions. The survey was conducted to different social groups in the different departments of the country; where three types of questions were found: open, single selection and multiple selection, which aim to see what kind of solutions the respondents find to help the reforestation of the Amazon and what, according to them, should be the accompaniment of both public and private organizations. With this purpose, we seek to make the results of the survey visible to carry out an analysis of the text focused mainly on the open questions and, in turn, on the complete document.

2. Objetivos del proyecto

2.1. General

Analizar la encuesta “Todos por la Amazonía” mediante el procesamiento del texto, con el propósito de obtener conclusiones que ayuden al equipo de Consejería Presidencial a entender qué soluciones ven los encuestados al problema de la deforestación en la Amazonía.

2.2. Específicos

1. Hacer un análisis exploratorio de la información sobre las preguntas cerradas (selección única y selección múltiple) que se plantearon en la encuesta.
2. Realizar análisis descriptivo de texto sobre las preguntas abiertas de la encuesta que ayuden a encontrar las generalidades de las soluciones propuestas por los ciudadanos
3. Sacar las principales conclusiones desde el punto de vista cuantitativo de la encuesta.

3. Metodología

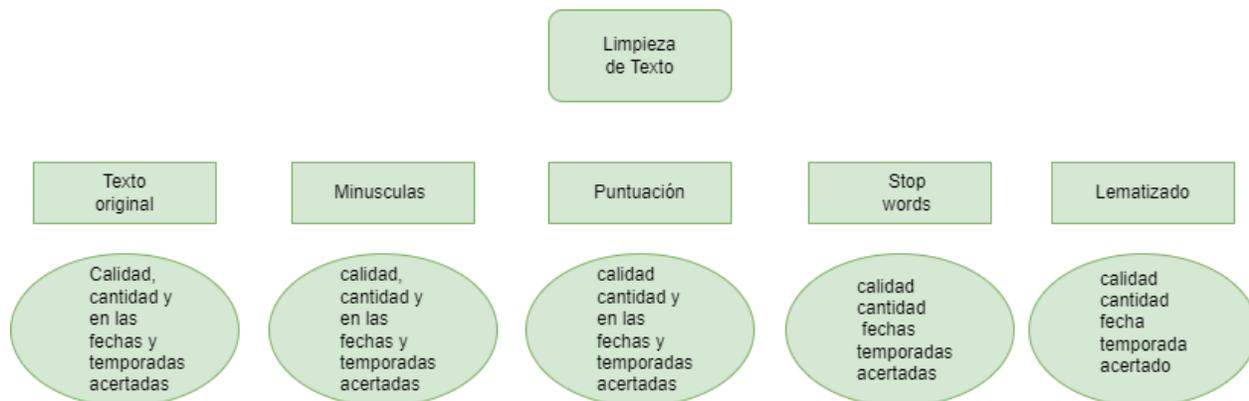
Para el análisis de la encuesta se realizaron procesos de limpieza de la base de datos de “Respuestas” junto con técnicas de procesamiento de texto. Para la limpieza inicial se utilizó la base de “Preguntas” para obtener el nombre completo de estas y así arreglarlas en la base de Respuestas dado que por lo amplio de las preguntas que se realizaron estas no salían completas en la base de respuestas lo cual dificultaba el entendimiento de la misma, también se unificó todo lo referente a la información de municipio de los encuestados en una sola variable con el fin de tener menos columnas y así ayudar al procesamiento de la información debido a que en la base original había una columna con el nombre de cada uno de los departamentos donde el encuestado debía colocar en que municipio de ese departamento vivía, esto generó bastantes valores faltantes y aumentaba la cantidad de columnas. El procesamiento de texto para las preguntas abiertas tuvo las siguientes fases:

3.1. Limpieza Básica

La limpieza básica del texto consistió en pasar el texto a minúsculas, eliminar signos de puntuación y *stop words*, las cuales son palabras vacías o que solo aportan ruido al análisis; generalmente son adverbios, conjunciones, preposiciones y artículos. En la Ilustración 1 se observa un diagrama de flujo del proceso con uno de los textos de la base de datos



Ilustración 1: Limpieza de texto



Fuente: Elaboración propia.

Como se observa en la Ilustración 1 en la fase de stopwords o palabras vacías se retiraron del texto los términos y, en, las ya que estos términos no entregan ninguna información adicional

3.2. Lematización

La segunda parte de procesamiento de las preguntas abiertas consistió en lematizar el texto. Este proceso consiste en transformar una palabra dada a la raíz, con el objetivo disminuir la variabilidad que pueda generar distintas versiones morfológicas de una misma palabra sin perder el significado original de lo que se quería transmitir. Adicionalmente la lematización ayuda a generar conteos de palabras más compactos y mejora la calidad de análisis exploratorios de texto. En la Tabla 1 se incluyen ejemplos de lematización junto con uno ya aplicado en la base de datos de las respuestas.

Tabla 1: Ejemplos de lematización sobre distintos textos.

Ejemplo	Ejemplo Lematizado
Soy, es, somos	ser
Mesita, mesota	mesa
Si sigue como vamos totalmente deforestada y sus aguas cristalinas contaminadas y una gran pérdida de fauna silvestre y conocimiento ancestral	seguir ir totalmente deforestar agua cristalino contaminado gran pérdida fauna silvestre conocimiento ancestral ¹

Fuente: Elaboración propia.

4. Resultados

En esta sección se presentan los hallazgos más importantes del análisis de la información estructurada y de texto

4.1. Estadísticas Descriptivas

Con ayuda de la metodología descrita en la sección 3 se consiguieron los siguientes resultados.

¹ En este ejemplo se aplicó limpieza de texto seguido de la lematización de palabras sobre la primera respuesta a la pregunta, ¿Cómo ve la Amazonía en el año 2030?



Tabla 2: Respuestas a la encuesta por departamento.

Departamento	Total, de Respuestas
Bogotá D.C.	292
Cundinamarca	112
Meta	108
Antioquia	100
Caquetá	97
Sucre	78
Guaviare	59
Córdoba	59
Valle Del Cauca	58
Norte De Santander	54
Quindío	50
Atlántico	49
Putumayo	44
Magdalena	44
Boyacá	34
Bolívar	33
Amazonas	31
Santander	31
Cauca	21
Arauca	20
Nariño	18
Tolima	17
Risaralda	16
Huila	15
Caldas	14
Cesar	12
Guainía	11
Casanare	11
Vaupés	10
La Guajira	6
Vichada	2
Chocó	1
San Andrés Y Providencia	1

Fuente: Elaboración propia.

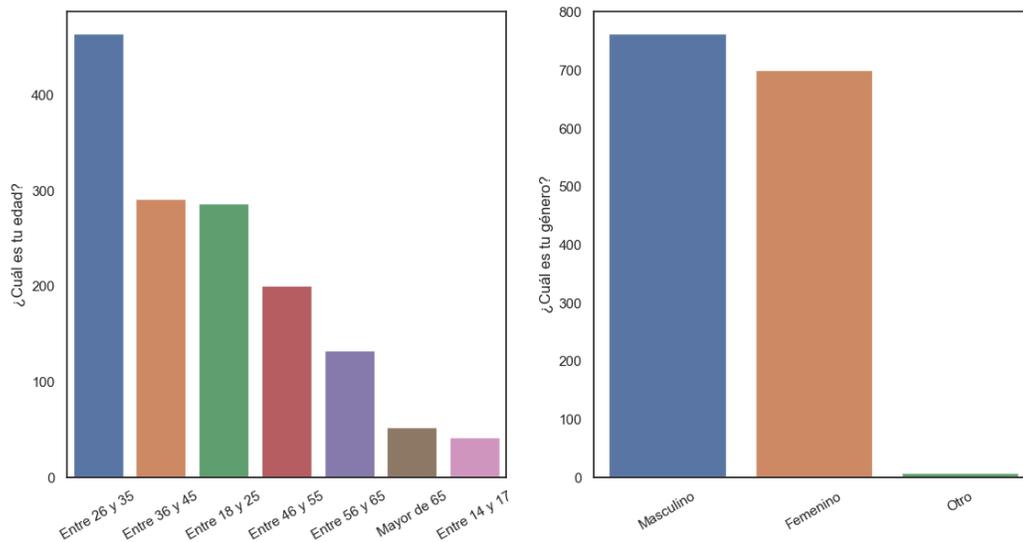
En la Tabla 2 se observa cuantas respuestas vienen de cada departamento, donde se observa que Bogotá se lleva el 20 % de las respuestas siendo el departamento que más tiene participación en la encuesta, seguido de Cundinamarca



con el 7.4%. En los últimos lugares se encuentran el archipiélago de San Andrés, Chocó y Vichada con un 0.6% de participación, Amazonas tiene un 2.05%.

En la Ilustración 2 se tienen las estadísticas descriptivas de la base de preguntas referentes a la edad y género de los encuestados.

Ilustración 2: Estadísticas descriptivas



Fuente: elaboración propia.

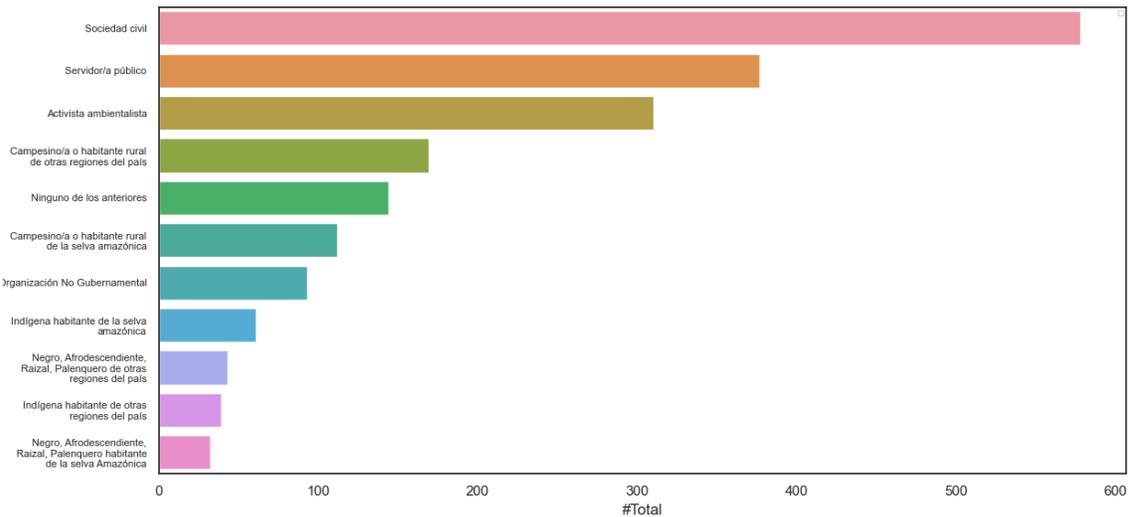
Se observa para la gráfica de Edad que aproximadamente el 70% de los encuestados se encuentran entre los 18 y 35 años siendo aquellos que tienen entre 26 y 35 el 31 % de los encuestados, seguido de los que están entre 36 y 45 años que representan el 19 %, infiriendo que los adultos jóvenes son los que más participación presentan en la encuesta. Los encuestados que están en edades muy jóvenes o viejos representan un 15% de participación siendo el grupo que esta entre los 14 y 17 años los que tienen un menor porcentaje de participación con un 2.7 % antecedido por los mayores de 65 con un 3.5%. En cuanto al género de los encuestados la situación esta más balanceada en



cuanto a participación femenina y masculina con 46% y 51% respectivamente, el grupo con una menor participación son todos aquellos que contestaron con otro género teniendo un 0.5% de participación.

En Ilustración 3 se observa la composición de grupos sociales en la base de datos de respuestas.

Ilustración 3: Grupos Sociales



Fuente: Elaboración propia

Para la creación de la gráfica se contó cuántas veces fue escogida la opción del grupo social ya que en esta pregunta se le permitía a los encuestados escoger más de una opción. Con la gráfica se observa que la mayoría de los encuestados se consideran que pertenecen a una sociedad civil, son servidores públicos y activistas ambientales, los grupos sociales con los cuales los encuestados se identificaron menos fueron: negro, afrodescendiente, raizal, palenquero; a su vez, de otras regiones del país (43), indígena habitante de otras regiones del país (39), negro, afrodescendiente, raizal, palenquero habitante de la selva Amazónica (32).

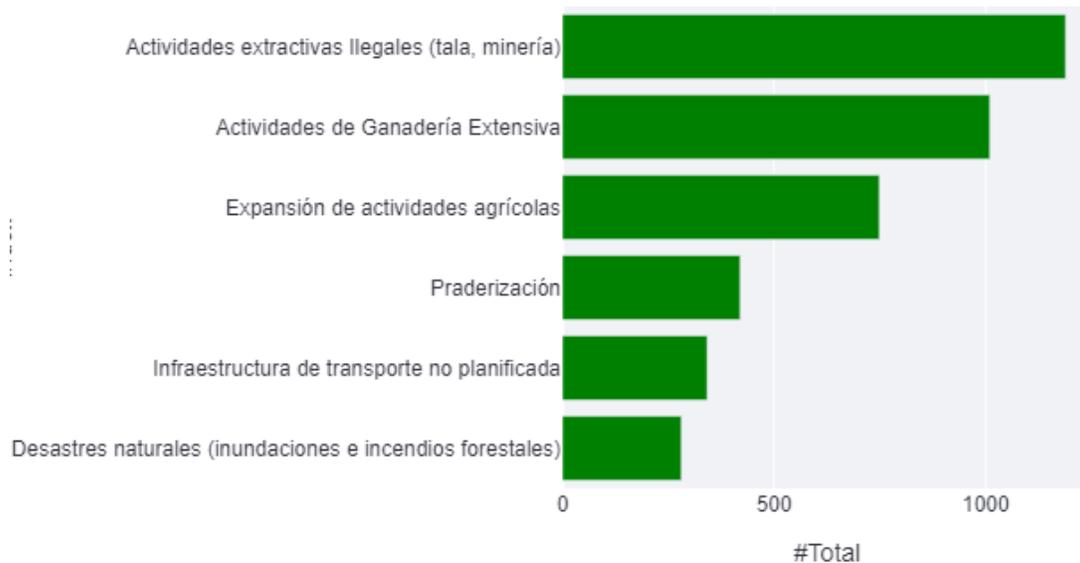
4.2. Análisis de Preguntas de Selección múltiple y única

En esta sección se presenta el análisis de las preguntas de selección tanto única como múltiple sin aplicar ningún filtro sobre la base de respuestas.

En la Ilustración 4 se observa el comportamiento de las respuestas a la pregunta de selección múltiple: “¿Cuáles crees que son las principales causas de la deforestación?” Para analizar la pregunta se contó cuantas veces se seleccionó cada una de las opciones. Se observa que el grupo más seleccionado con un 40% son "actividades extractivas ilegales", seguido con un 25% de "actividades de ganadería extensiva" y las que menos tuvieron participación fueron las opciones de "infraestructura de transporte" y desastres naturales con 7% y 8% respectivamente. Con ello se tiene que el 84% de las opciones seleccionadas son referentes a causas humanas para la deforestación.



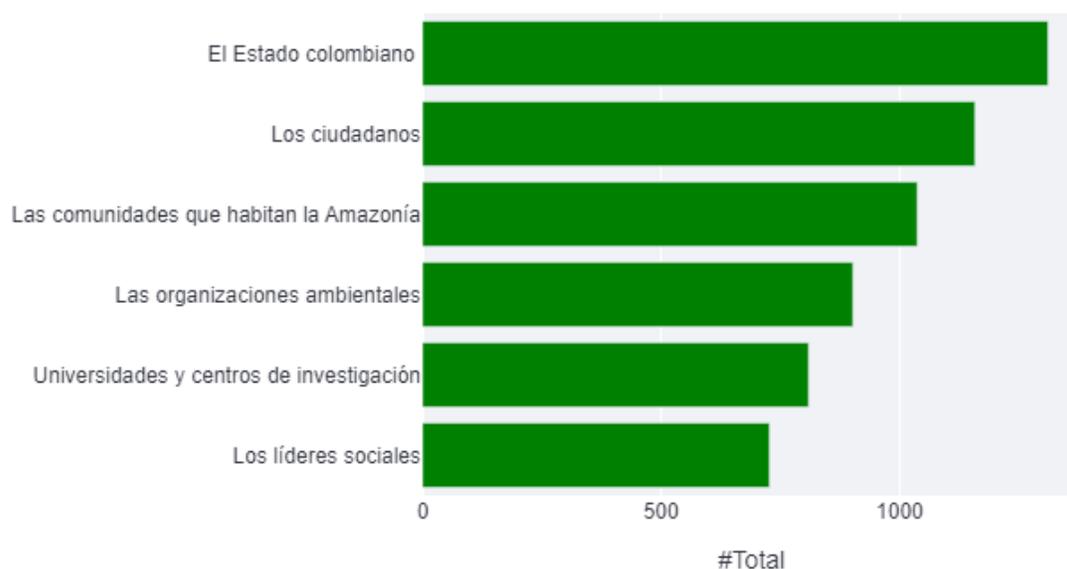
Ilustración 4: Distribución de respuestas a la pregunta ¿Cuáles crees que son las principales causas de la deforestación?



Fuente: Elaboración propia.

En la Ilustración 5 se observa el comportamiento de la pregunta “¿Quién crees que es el responsable de la conservación de la Amazonía colombiana?” Esta pregunta es de selección múltiple y para su análisis se contó cuantas veces cada una de las opciones era seleccionada. El Estado colombiano tiene una participación del 22% seguido de los ciudadanos con 19% y comunidades de la Amazonía con un 17%, y en las últimas posiciones están organizaciones ambientales, universidades y líderes sociales con 15%,13% y 12% respectivamente. Como se observa no hay diferencias muy grandes entre la participación de las distintas opciones para los encuestados resaltando que 41% pertenecen al estado y a los ciudadanos como los principales responsables de la conservación de la Amazonía.

Ilustración 5: “¿Quién crees que es el responsable de la conservación de la Amazonía colombiana?”

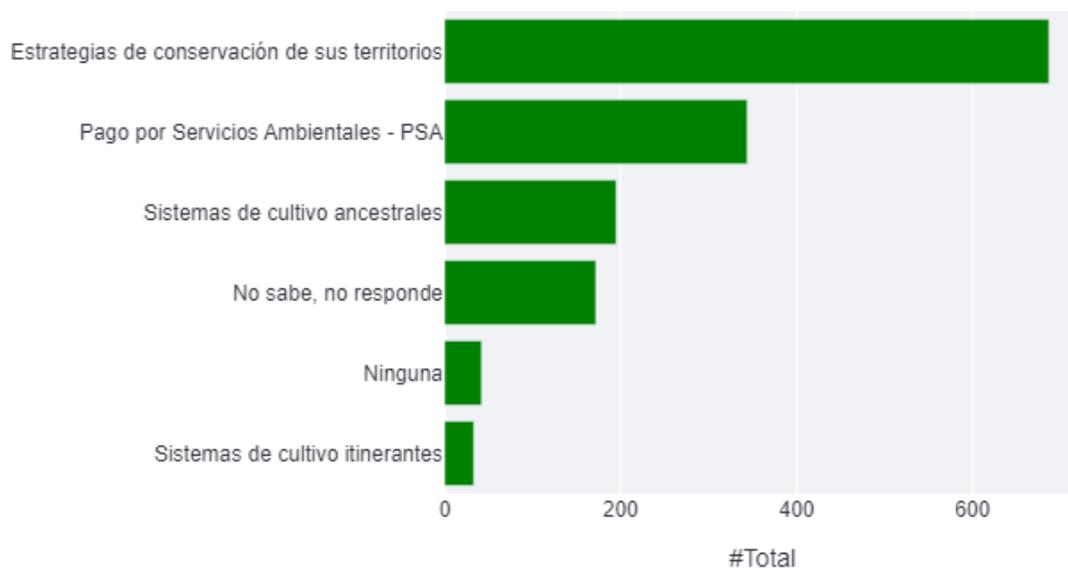


Fuente: Elaboración Propia.



En la Ilustración 6 se muestra los resultados de la pregunta: “El Estado colombiano ha reconocido la importancia de articular los conocimientos de poblaciones étnicas y campesinas en relación con el uso y manejo de la biodiversidad, a través de diferentes instrumentos y estrategias. ¿Cuál de esas consideras que genera un mayor impacto en la conservación de los ecosistemas Amazónicos y la protección de los saberes de las comunidades que habitan el territorio?”. Esta pregunta a diferencia de las anteriores es de selección única donde el 2% son datos faltantes. Para los encuestados las estrategias que más ayudan a la protección del medio ambiente son estrategias de conservación de sus territorios (45.5%), pago por Servicios Ambientales-PSA (23%), y Sistemas de cultivo ancestrales: (12%). El 11% no sabe cuáles de las opciones que se le presentaron puede ser una buena estrategia y el 2% piensa que Ninguna de las presentadas es una buena opción. Los sistemas de cultivo itinerantes son considerados como una buena estrategia por un 2% de los encuestados.

Ilustración 6: ¿Cuál de esas consideras que genera un mayor impacto en la conservación de los ecosistemas Amazónicos y la protección de los saberes de las comunidades que habitan el territorio?



Fuente: Elaboración Propia.

4.3. Análisis de Preguntas abiertas

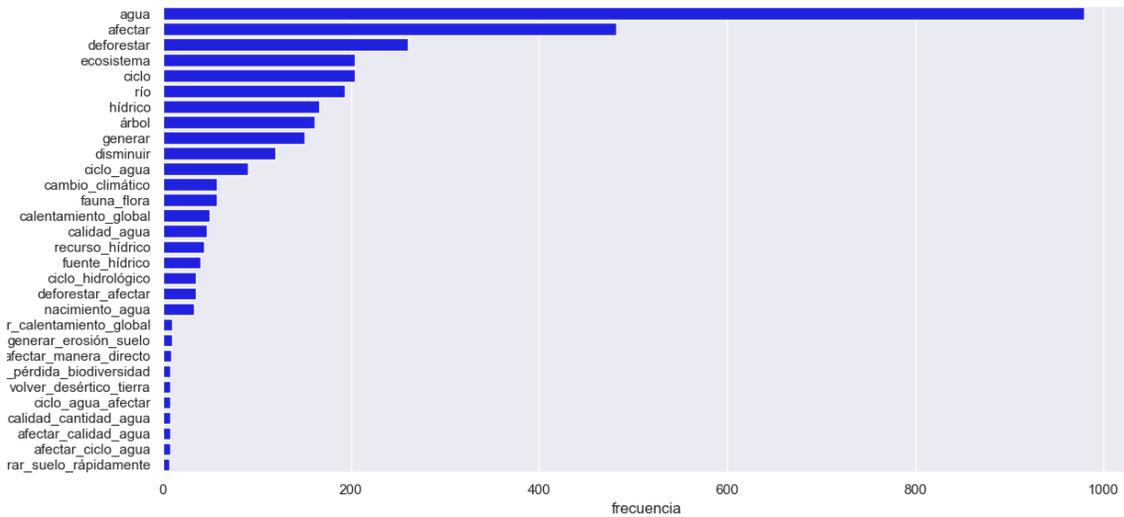
Para el análisis de las preguntas abiertas se siguió la metodología previamente mencionada, se trabajó con el texto lematizado como se mencionó en la Sección 3 y además de esto se realizó un análisis de n-gramas para la frecuencia de los distintos términos en las respuestas que los encuestados dieron. Los n-gramas consisten en secuencias de tamaño de n de palabras o *tokens* donde una de sus utilidades es la detección de palabras claves. Para las preguntas abiertas de la base de datos se usaron unigramas, bigramas y trigramas.

Para la pregunta: “La macro cuenca amazónica se refiere a todo el complejo acuático que alimenta al gran río Amazonas, incluyendo algunos nacimientos de agua en los páramos andinos. ¿De qué manera crees que la deforestación en la Amazonía afecta a todo este complejo de agua incluyendo la que llega hasta tu hogar?” se realizó un análisis de trigramas, tomando grupos de palabras de tres, para ver cuáles eran los términos más frecuentes y observar que palabras lo componen, se graficó el top diez. En la Ilustración 12 se evidencia esta gráfica donde observamos el término “agua” aparece tanto en los trigramas más frecuentes como en los menos frecuentes, el top tres de trigramas más frecuentes son “agua ver afectado”, “contribuir calentamiento global” y “generar erosión suelo”. Con esto se evidencia que los encuestados tienen muy presente que la deforestación afecta directamente el agua y también al cuidado del suelo que se observa en los trigramas “generar erosión suelo” y “volver desértico tierra”.



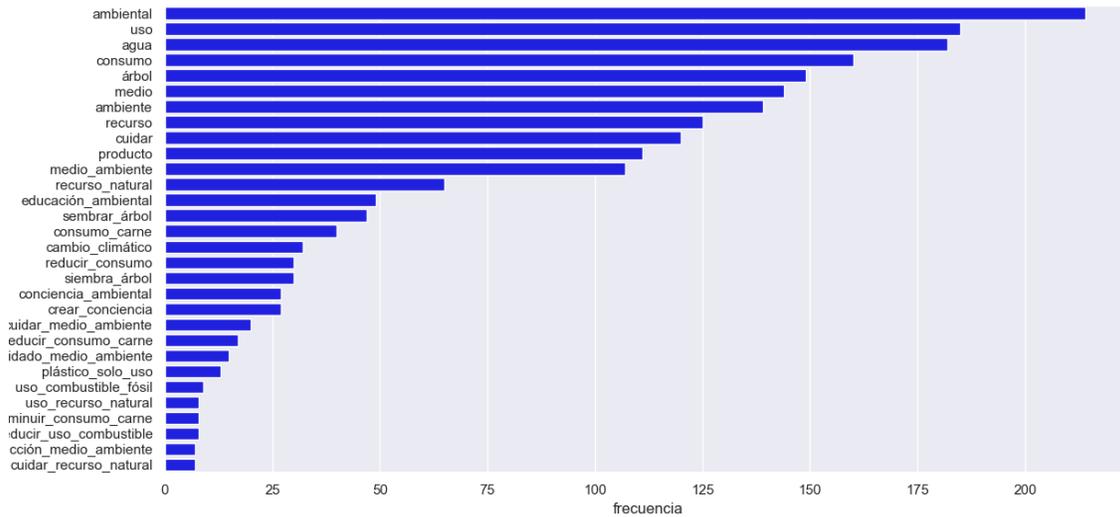
6. ANEXOS

Ilustración 12: ¿De qué manera crees que la deforestación en la Amazonía afecta a todo este complejo de agua incluyendo la que llega hasta tu hogar?



Fuente: Elaboración propia

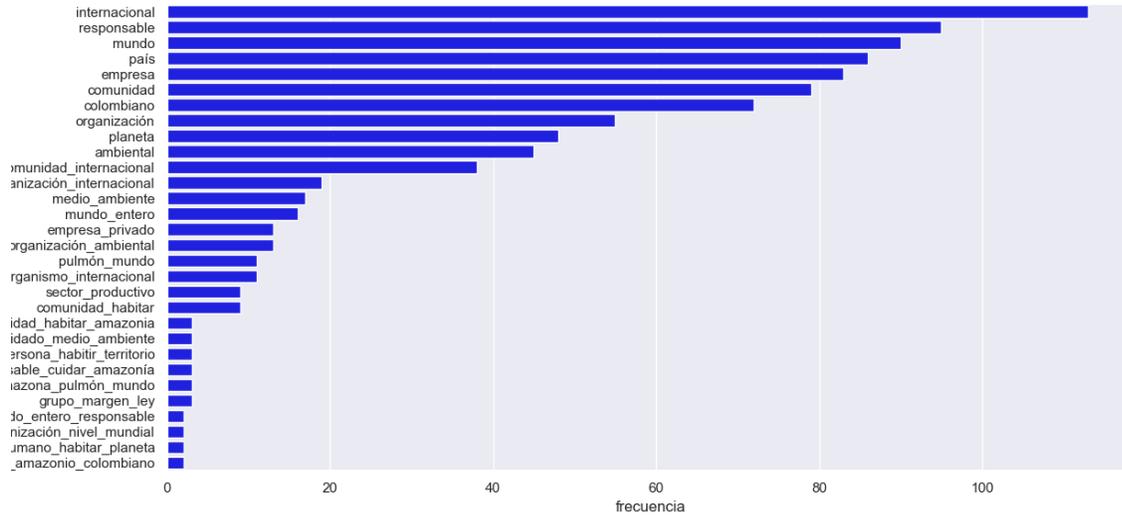
Ilustración 13: ¿Cuáles son tus compromisos ciudadanos para proteger a la Amazonia contra la deforestación y hacer frente a las causas del cambio climático?



Fuente: Elaboración propia

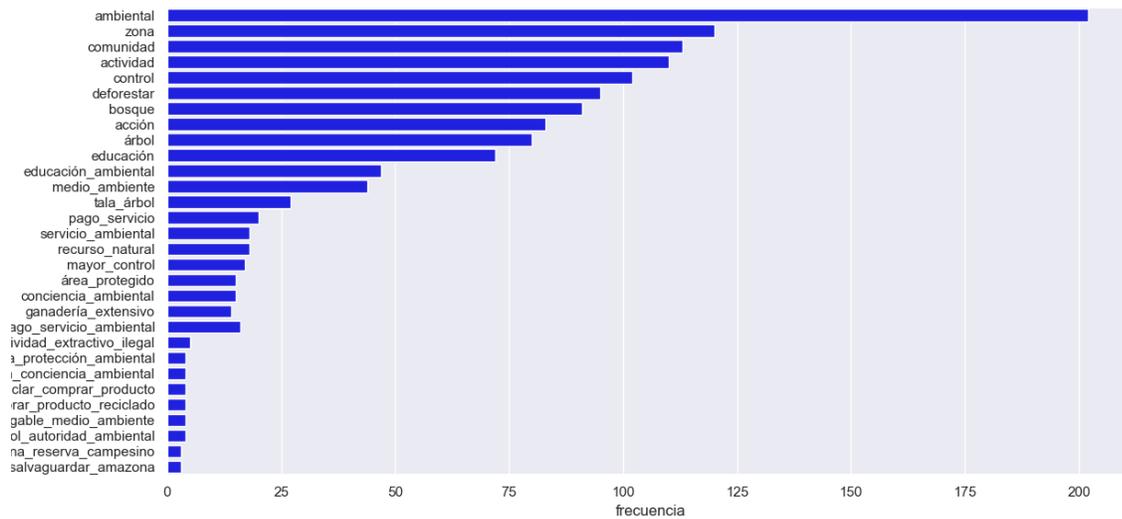


Ilustración 14: ¿Crees que hay otro responsable de la conservación de la Amazonía colombiana? ¿Quién?



Fuente: Elaboración Propia

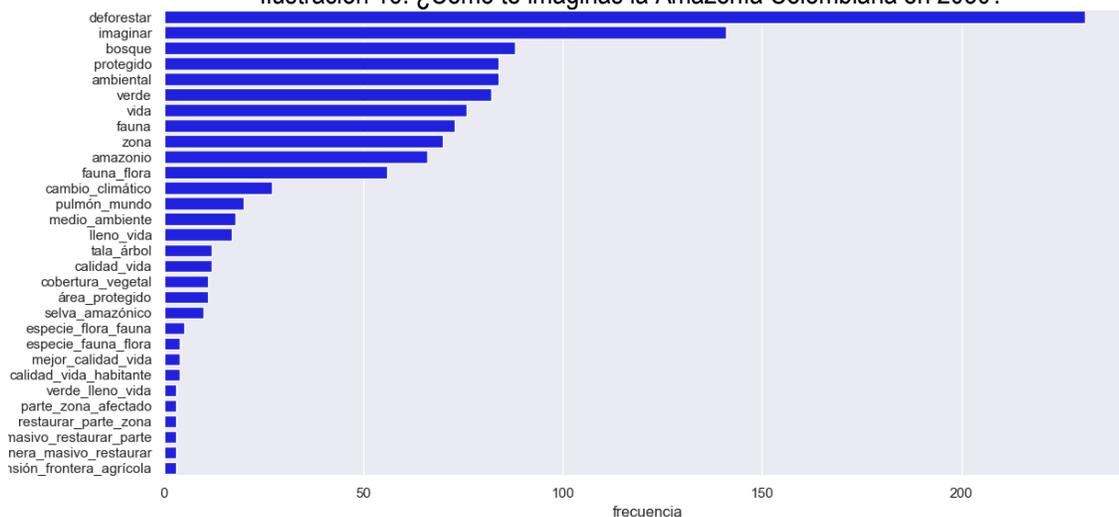
Ilustración 15: ¿Se te ocurre otra acción para frenar o disminuir la deforestación en la Amazonía colombiana? ¿Cuál?



Fuente: Elaboración Propia



Ilustración 16: ¿Cómo te imaginas la Amazonía Colombiana en 2030?



Fuente: Elaboración Propia.

Tabla 4: N-gramas para los compromisos

Compromisos	N-gramas
Cuidado del medio ambiente y sus recursos	Medio ambiente Recursos Recurso Natural Basura
Fortalecer educación ambiental	Crear conciencia Educación
Reducción de plásticos	Consumo carne Reducir consumo Disminución consumo
Compromiso de siembra de árboles	Sembrar árbol Plantar árbol Sembrado árbol arbolito
Reducción de plásticos	Plástico Plastico Solo uso
Reducción de minerías y actividades extractivas	Minería Mineria talar
Reducir uso de combustibles	Combustible fósil carbón
Aumentar uso de bicicletas	bicicleta

Fuente: Elaboración Propia