

Dirección de Desarrollo Digital

Unidad de Científicos
de Datos



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación



PROCESAMIENTO NATURAL DEL LENGUAJE EN PLANES DE DESARROLLO TERRITORIAL

INFORME FINAL

Dependencias y entidades involucradas	Departamento Nacional de Planeación <ul style="list-style-type: none">• Dirección de Desarrollo Digital - Unidad de Científicos de Datos• Dirección de Descentralización y Desarrollo Regional- Grupo de estudios territoriales
Sector	Planeación
Tecnologías utilizadas	Python
Fuentes de datos	PDT, PND, planes estratégicos y estudios de caso

Contenido

1. Presentación	2
2. Objetivos del proyecto	2
3. Metodología	2
4. Resultados	11
5. Conclusiones y recomendaciones	30
6. Socialización	30
Contacto	30
ANEXOS	31
Anexo 1	31
Anexo 2	36
Anexo 3	39
Anexo 4	47
Anexo 5	49



1. Presentación

El análisis de los Planes de Desarrollo Territorial del periodo de gobierno 2020-2023 y su alineación con el Plan Nacional de Desarrollo 2018-2022 y los planes estratégicos y estudios de caso de asociaciones municipales y departamentales permite identificar las apuestas estratégicas que registran los Planes de Desarrollo Territorial y evaluar su cumplimiento. Asimismo, se identifica si los territorios tienen apuestas de articulación subnacional y las categorías económicas que más se mencionan a través de la búsqueda de términos relevantes de cada una de estas.

Para llevar a cabo este análisis se hace uso de técnicas exploratorias de texto y se obtienen como resultados: gráficos de barras y nubes de palabras que resumen el contenido de todos los documentos, búsqueda de términos relevantes que permiten la asociación de cada documento con temáticas relevantes, identificación de similitud y disimilitud entre los PDT y planes estratégicos, estudios de caso y el PND. La disimilitud entre documentos se resume con la búsqueda de las palabras exclusivas de cada documento.

The analysis of the Territorial Development Plans for the 2020-2023 government period and their alignment with the 2018-2022 National Development Plan and the strategic plans and case studies of municipal and departmental associations allows to identify the strategic bets registered in the Plans of Territorial Development and evaluate its compliance. Likewise, it is identified if the territories have sub-national articulation bets and the economic categories that are mentioned the most through the search for relevant terms of each of these.

To carry out this analysis, exploratory text techniques are used, and the following results are obtained: bar graphs and word clouds that summarize the content of all documents, search for relevant terms that allow the association of each document with relevant topics, identification of similarity and dissimilarity between the PDT and strategic plans, case studies and the PND. The dissimilarity between documents is summarized with the search for the unique words of each document.

2. Objetivos del proyecto

2.1. General

Identificar y analizar las apuestas estratégicas registradas en los Planes de Desarrollo Territorial (PDT) 2020-2023 y la alineación de los PDT con el Plan Nacional de Desarrollo (PND) 2018-2022, mediante el uso de técnicas de análisis y minería de texto.

2.2. Específicos

1. Lectura y adecuación de los documentos PDT 2020-2023, PND 2018-2022 y documentos de texto relacionados mediante técnicas de OCR (donde sea necesario) y lectura automática de documentos.
2. Identificar las apuestas estratégicas registradas en los PDT mediante técnicas exploratorias de texto.
3. Establecer el nivel de alineación entre estos PDT y el PND mediante medidas de similitud de texto.
4. Identificar si los territorios contemplan apuestas de articulación del nivel subnacional en sus Planes de Desarrollo y analizar cuáles son los movilizadores de estas apuestas, a partir de los análisis sobre los textos.

3. Metodología

3.1. Lectura de los documentos

En esta primera etapa se realiza la lectura de todos los PDT disponibles, el PND, los planes estratégicos y estudios de caso necesarios para el estudio. Para algunos de ellos se utiliza el OCR (Reconocimiento óptico de caracteres) que facilita la extracción de texto de imágenes o documentos escaneados. Una vez se realiza este procedimiento se obtienen unos documentos en formato .txt que permiten un posterior análisis.



3.2. Limpieza de los documentos

La etapa de limpieza de texto se aplica tanto para los documentos PDT, PND, estudios de caso y planes estratégicos como a cada uno de los términos de interés que se buscarán en los documentos, este proceso incluye 4 pasos:

1. Remover signos de puntuación, números y conversión a minúsculas: se retiran todos los signos de puntuación del texto tales como comas, puntos, signos de exclamación así como los números. Después todo el texto es convertido a texto en letras minúsculas.
2. Remoción de *stopwords* o palabras vacías: estas palabras no tienen significado por sí solas, sino que modifican o acompañan a otras. Este conjunto se encuentra conformado por artículos, pronombres, preposiciones y adverbios.
3. Quitar palabras propias del contexto: Se remueve del texto palabras que pueden ser muy frecuentes dado la naturaleza del estudio, para ello se removieron nombres propios de personas, municipios y departamentos de Colombia y números.
4. Lematización de texto: se refiere a la transformación de todas las formas flexionadas de una palabra en su lema, por ejemplo, el plural, femenino y diminutivo de una palabra se transforma en su lema: mesas, mesón y mesitas se transforman todas en su lema, mesa. Este proceso reduce el tamaño del vocabulario (número de palabras que conforman todo el universo de los documentos) lo que facilita la búsqueda automática y coincidencia de los términos de interés dentro de los documentos.

Nota: dado que los términos relacionados con articulación subnacional contienen números es necesario que los textos de búsqueda los conserven, por esta razón cuando se realiza la búsqueda de estas palabras se omite la remoción de números en la limpieza de los textos.

3.3. Frecuencia de las palabras dentro de los documentos

Con los textos en formato plano se realiza la identificación de la frecuencia de las palabras y a través de gráficos como nubes de palabras y gráficos de barras se representan los términos con mayor aparición dentro de los documentos.

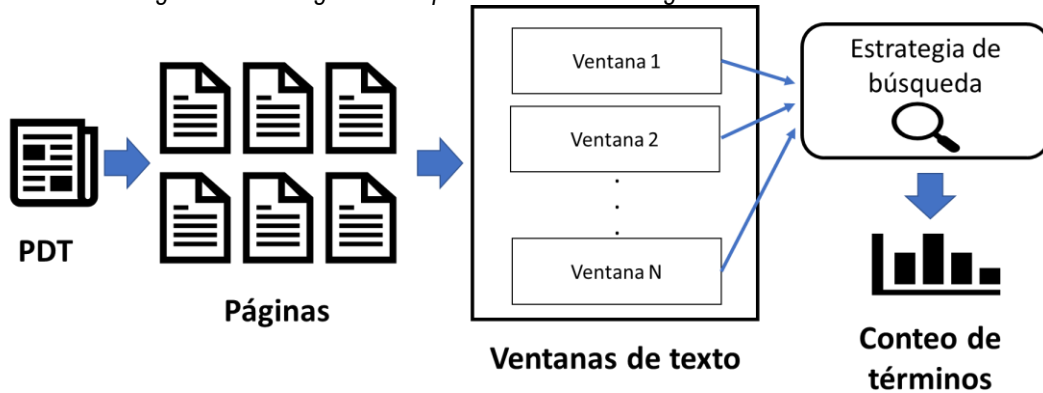
3.4. Búsqueda de frases y términos en los PDT

En la Figura 1 se muestra el esquema general de la búsqueda de términos relacionados con cada categoría predefinida en los documentos PDT. Cada documento PDT después de la etapa de limpieza de texto, pasa por los siguientes pasos:

1. El documento es dividido en páginas para el análisis de términos.
2. Cada página del PDT a su vez es dividido en fragmentos de texto más pequeños, denominados ventanas de análisis. Cada ventana de análisis está compuesta por 60 palabras. En algunos casos puede contener menos palabras debido al número de palabras presentes por página, es decir, una página con 353 palabras tendrá 5 ventanas de 60 palabras y una de 53 palabras.
3. Cada ventana entra a la estrategia de búsqueda de términos y al final del proceso se consolida el conteo de número de apariciones de los términos en las ventanas. La estrategia de búsqueda se explica a continuación:



Figura 1: Estrategia de búsqueda de términos de género en documentos PDT



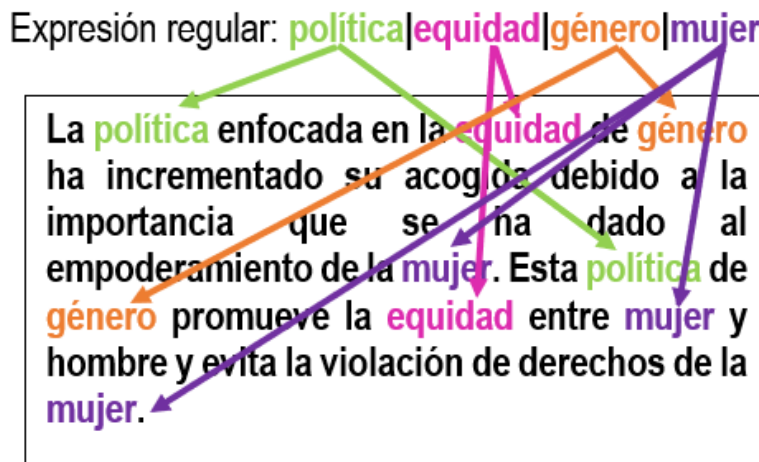
Fuente: elaboración propia

Teniendo en cuenta que un término o frase de interés se puede presentar de diferentes maneras en los documentos PDT, la estrategia de búsqueda permite hacer búsquedas flexibles de términos basado en expresiones regulares¹. Por ejemplo, la palabra *natural* aparece contenida en la palabra *naturaleza*, sin embargo, son dos palabras distintas que deben analizarse por separado, para dar solución a este problema se establece que la palabra *natural* sea contada solo si aparece idéntica, para ello la expresión regular es `\bnatural\b`.

Por otro lado, los términos compuestos por dos o más palabras, se buscan por separado en las ventanas de análisis, por ejemplo, la frase “Política de equidad de género para las mujeres”, es buscada como `política|equidad|género|mujer`, la raya vertical o pleca (|) actúa como una condición lógica “o” y permite que se busquen las cuatro expresiones al mismo tiempo. En la Figura 2 se ilustra su funcionamiento sobre un texto de prueba.

Después de identificar todas las palabras que cumplen el patrón de la expresión regular, se hace un conteo por cada condición separada por la pleca (|). En el ejemplo de la Figura 2, hay 4 condiciones y cada una tiene 2 ocurrencias a excepción de *mujer* que tiene 3. Para encontrar el número de apariciones de la frase “Política de equidad de género para las mujeres”, se halla el mínimo de ocurrencias entre todas las condiciones, en este caso 2, por lo que la estrategia de búsqueda da como resultado del conteo 2 ocurrencias para esa ventana de análisis.

Figura 2: Ejemplo del funcionamiento de la búsqueda de términos



¹ Las expresiones regulares son patrones que se utilizan para hacer coincidir combinaciones de caracteres en cadenas de texto.



Texto de prueba

Política de equidad de género para las mujeres: 2 ocurrencias

Este proceso se hace para cada término y cada ventana dentro del PDT. Al final se consolidan los conteos de los términos para todo el documento. Es importante tener en cuenta que algunas palabras pueden presentar conteos duplicados dado que aparecen solas y también al interior de una frase, por ejemplo, desastre natural y natural contienen la palabra “natural”. Todos los conteos de palabras se normalizan de tres formas: primero, con respecto al número de palabras que contiene cada documento, dividiendo el valor de la frecuencia de aparición en el número total de palabras que contiene, segundo, con respecto al número de páginas que tiene el documentos y tercero, respecto a la suma de la frecuencia de los 2000 términos más frecuentes dentro del documento.

3.5. Búsqueda de términos relacionados con las palabras identificadas

Después de tener los resultados de la búsqueda de términos se identifican las palabras que los rodean con el fin de establecer las temáticas que se relacionan con dichos términos. Como son más de mil términos, este proceso de búsqueda se agrupa por categorías para facilitar la interpretación de los resultados. El proceso se realiza con los 100 términos más cercanos al término de interés, sin embargo, este valor puede variar pues en el caso de algunas palabras no se alcanzan a encontrar 100 términos.

En la Figura 3 se observa un ejemplo de la búsqueda de términos relacionados con las palabras de interés.

Figura 3: Ejemplo del funcionamiento de la búsqueda de términos en una ventana de análisis.

Departamento requiere disponer
tierras aptas producción
agropecuaria número significativo
productores condiciones vida
adecuadas eviten migración jóvenes
permita dirección más técnica
procesos productivos.

Fuente: elaboración propia

Para el caso del ejemplo, la palabra de interés es productores, después de ser identificada con el proceso descrito en la Sección 3.4, dentro del texto se buscan las palabras que la rodean en la ventana que se encontró el término, en este caso son las que se encuentran subrayadas con color amarillo.

Este proceso se realiza para cada una de las palabras y después de esto las palabras identificadas como relacionadas con cada término de interés son agrupadas por la respectiva categoría a la que pertenecen.

Como resultado de este proceso se obtiene una tabla con la siguiente estructura:

Tabla 1: estructura resultados búsqueda de términos relacionados

Departamento	Término inicial	Categoría	Término relacionado	Cantidad
Antioquia	Sostenibilidad	Cambio climático	desarrollo	20



Antioquia	Ambiental	Cambio climático	desarrollo	20
-----------	-----------	------------------	------------	----

Fuente: elaboración propia

Con estos resultados se elimina la columna “Término inicial” y se agrupan los datos para cada categoría, como resultado de ese proceso se obtiene una estructura como la que se observa en la Tabla 2.

Tabla 2: términos relacionados, estructura final

Departamento	Categoría	Término relacionado	Cantidad
Antioquia	Cambio climático	desarrollo	40

Fuente: elaboración propia

3.6. Cálculo de similitud entre documentos

Con el fin de comparar el contenido de algunos PDT municipales y departamentales con los estudios de caso y planes estratégicos se lleva a cabo un cálculo de similitud entre textos. En la Tabla 3 se observan los municipios y departamentos cuyos PDT fueron comparados y los respectivos documentos correspondientes a esta comparación.

Tabla 3: documentos comparados

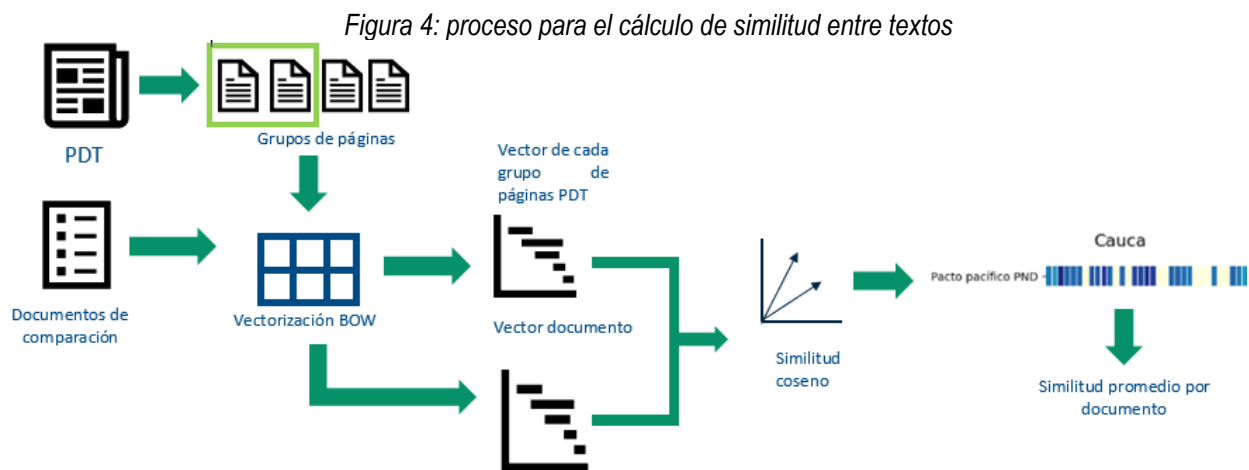
Documento	PDT
Estudio de caso RAPE Central Plan estratégico RAPE Central	Boyacá Cundinamarca Huila Meta Tolima
Estudio de caso AMCO Acuerdo metropolitano AMCO Plan estratégico AMCO	Pereira Dosquebradas La Virginia
Estudio de caso CARTAMA	Fredonia Jericó La Pintada Montebello Pueblorrico Támesis Tarso Valparaíso Venecia
Estudio de caso AMUNORCA	Buenos Aires Caldono Corinto Guachené Jamboló



	Miranda Puerto Tejada Santander de Quilichao Suárez Toribío Villa Rica
Capítulo regional pacto pacifico PND	Cauca Valle Nariño Chocó

Fuente: elaboración propia

En la Figura 4 se observa el proceso para el cálculo de similitud entre textos.



Fuente: elaboración propia

A continuación se explica el paso a paso del proceso para el cálculo de similitud entre los textos:

1. Dado que los documentos PDT en general tienen una mayor extensión que los documentos estudios de caso y planes estratégicos, no se pueden comparar directamente, pues daría como resultado valores de similitudes pequeñas inducidas por la diferencia de longitud del contenido de los textos. Por tal motivo, se dividieron los PDT por grupos de páginas, para poder comparar directamente cada grupo con cada documento. Esta estrategia permite resaltar las secciones de cada documento PDT en las cuales existe mayor similitud con los planes estratégicos y los estudios de caso, además de poder calcular un índice de similitud promedio más acertado que uno total por documento.
2. Una vez se tiene la representación granular de cada PDT, se realiza una vectorización (representación numérica) de los textos de cada PDT y cada estudio de caso, plan estratégico y el capítulo regional pacto pacifico del PND. Para ello, se utilizó el modelo de bolsa de palabras (Bag of Words), que consiste en construir una matriz con palabras que pertenecen al corpus de los documentos y son únicas por columna. En dicha matriz, las filas muestran el número de veces que aparece cada palabra en un determinado PDT o documento, según corresponda. Por ejemplo, si un PDT fuera únicamente “acabar pobreza extrema”, las columnas “acabar”, “pobreza” y “extrema” tendrían un valor de 1 en la fila de la matriz, mientras las demás columnas tendrían un valor de 0. Otro ejemplo se presenta en la Tabla 4, allí se tienen dos cadenas de texto; “Tasa de ocupación” y “Tasa de



formalidad laboral”, las columnas están compuestas por las palabras contenidas en los textos y cada texto corresponde a un vector fila.

Tabla 4: ejemplo modelo bolsa de palabras (BOW)

	de	ocupación	formalidad	tasa	laboral
Tasa de ocupación	1	1	0	1	0
Tasa de formalidad laboral	1	0	1	1	1

Fuente: elaboración propia

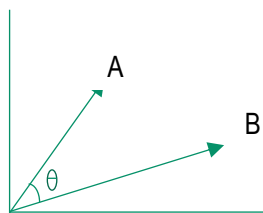
- Con base en el punto anterior se procedió a vectorizar cada grupo de páginas de los PDT con base en las palabras presentes en los textos de los estudios de caso, planes estratégicos y PND. Esto es esencial debido a que los vectores finales necesitan estar en el mismo espacio vectorial para ser comparables. En términos prácticos, la realización de este procedimiento sería equivalente a, por un lado, eliminar de los PDT las palabras que no estuvieran en los textos de referencia (estudios de caso, planes estratégicos y PND), para luego vectorizarlos fácilmente como nuevas filas en la misma matriz.
- Con la representación numérica de cada documento se realiza el cálculo de la similitud coseno entre el vector de cada grupo de páginas de cada PDT y los textos de comparación correspondientes, obteniendo una medida de similitud para cada grupo de páginas de cada PDT con su respectivo documento de comparación. La similitud coseno es una medida de la similitud existente entre dos vectores que se encuentran en un mismo espacio entre los que se evalúa el coseno del ángulo comprendido entre ellos. En la Ecuación 1 se observa la ecuación utilizada para encontrar esta similitud.

Ecuación 1: similitud coseno

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Para este caso, A y B son respectivamente la representación vectorial del documento de comparación y el grupo de páginas de cada PDT. En la Figura 5 se ve gráficamente este procedimiento el procedimiento gráficamente

Figura 5: ángulo entre dos vectores A y B



Fuente: elaboración propia

- Finalmente, los resultados son graficados en mapas de calor donde se observa la similitud entre cada página de los PDT y el documento con el cual es comparado. Antes de realizar estos gráficos, los valores de similitud menores a 0.2 para los PDT municipales y 0.3 para los PDT departamentales son reemplazados por 0, con el fin de eliminar el ruido de similitudes muy bajas dentro de los mapas de calor. Posteriormente los resultados que se pueden observar representados en los mapas de calor se utilizan como insumo para encontrar un promedio de la similitud y un promedio de similitud “ponderado” que excluye de la sumatoria los grupos de



páginas que tienen similitud menor al umbral de 0.2 o 0.3 según corresponda. Este procedimiento se encuentra explicado en la Ecuación 2 y Ecuación 3

Ecuación 2: promedio de similitud

$$Promedio = \frac{\sum_g S(V_c, g)}{G}$$

Ecuación 3: promedio ponderado de similitud

$$Promedio_ponderado = \frac{\sum_g S(V_c, g | S > t)}{G | S > t}$$

S = similitud

V_c = documento de comparación

g = grupo de páginas

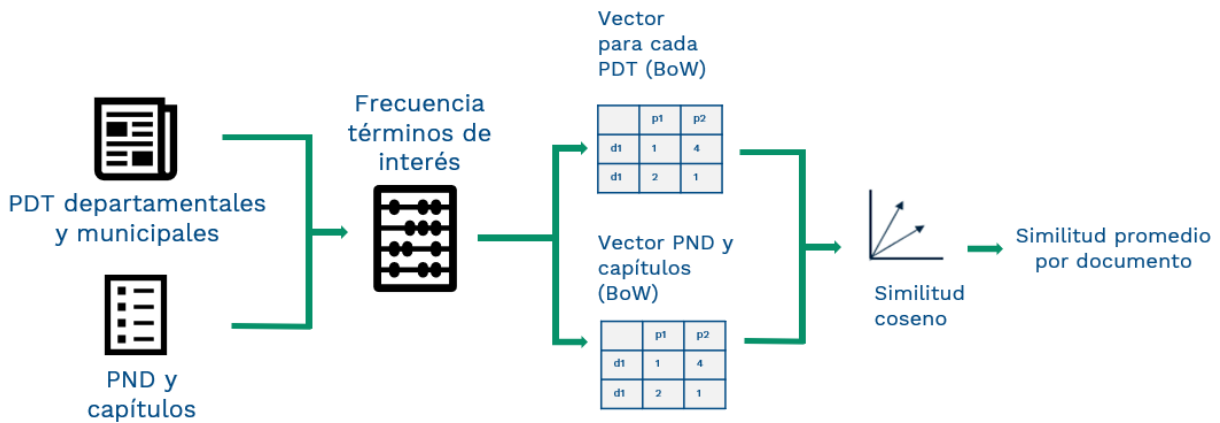
G = total de grupo de páginas por PDT

t = umbral de similitud

3.7. Similitud entre PDT y PND

Para el cálculo de similitud entre los PDT municipales y departamental, el PND y los capítulos de: equidad, legalidad, emprendimiento, pacto por la descentralización, pactos por la productividad, pacto por la construcción de paz, pacto por la equidad para las mujeres y pacto por la equidad de oportunidades para grupos indígenas, negros, afros, raizales, palenqueros y Room del PND, se llevó a cabo la metodología que se observa en la Figura 6.

Figura 6: metodología cálculo de similitud PDT-PND



Fuente: elaboración propia.

1. El primer paso es la identificación y extracción de los capítulos del PND que van a ser analizados. Posteriormente estos se limpian y lematizan según el proceso que está especificado en los entregables 1, 2 y 3.
2. Una vez se tienen los documentos adecuados, se realiza la búsqueda de los términos asociados a las 30 categorías analizadas en el Entregable 2 para los PDT, el PND y cada capítulo a analizar. Como resultado de este proceso se obtiene una matriz Bag of Words (BoW) con todas las palabras correspondientes a las diferentes categorías de análisis y la frecuencia de aparición de cada una dentro de cada documento, esto es la representación vectorial.



3. Con las representaciones vectoriales de cada documento se realiza el cálculo de similitud coseno entre el vector de:
 - Cada PDT departamental y el PND
 - Cada PDT departamental y cada capítulo del PND
 - Cada PDT departamental con los demás PDT departamentales
 - Cada PDT municipal y el PND
 - Cada PDT municipal y cada capítulo del PND
 - Cada PDT municipal con los demás PDT municipales
 -

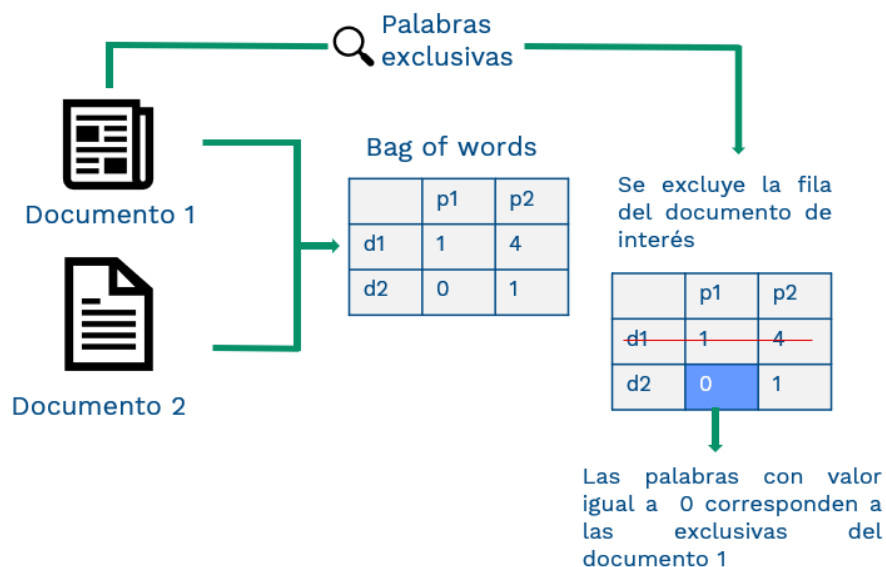
Este proceso da como resultado una medida de similitud entre cada par de documentos comparados.

La similitud coseno es una medida de la similitud existente entre dos vectores que se encuentran en un mismo espacio entre los que se evalúa el coseno del ángulo comprendido entre ellos.

3.8. Palabras exclusivas como indicador de disimilitud entre PDT y PND

Como indicador de disimilitud entre los documentos se toman las palabras exclusivas de cada uno. Esto quiere decir, aquellos términos que únicamente se encuentran dentro de un documento. El proceso llevado a cabo para encontrar estos términos se resume en la Figura 7.

Figura 7: búsqueda de palabras exclusivas



Fuente: elaboración propia

1. Se construye una matriz Bag of Words que contiene la totalidad de palabras que aparecen en todos los textos cuyas columnas corresponden a las palabras que aparecen en todos los documentos (el vocabulario del conjunto de documentos), cada fila corresponde a cada documento y el elemento (d_i, p_j) de la matriz corresponde a la frecuencia o número de veces que aparece la palabra j en el documento i .
2. Para determinar las palabras exclusivas de un documento, es decir las que aparecen en este y no en los demás, se elimina de la matriz Bag of Words total la fila correspondiente al documento de interés.
3. Con la nueva matriz Bag of Words se identifican los elementos (d_i, p_j) con valor igual a 0. Estos corresponden a las palabras que aparecen únicamente en el documento de interés.

Este proceso de comparación se llevó a cabo para tres grupos:



- Cada PDT departamental se compara con el PND.
- Cada PDT municipal se compara con el PND.
- Cada capítulo del PND se compara con los demás capítulos.

Al encontrar todas las palabras exclusivas de cada documento se ordenan por su frecuencia de aparición y dado que algunas de ellas son palabras sin contexto se eliminan las que tienen una frecuencia menor a cierto umbral. Para los capítulos del PND este umbral es 3, para los PDT departamentales 10 y para los PDT municipales 3. Estos valores se obtienen al realizar una observación de la calidad de los resultados iterando con distintos umbrales.

4. Resultados

A través del desarrollo metodológico descrito en la Sección 3, se obtuvieron los resultados que se presentan a continuación.

4.1. Nubes de palabras

Las nubes de palabras son una representación visual que muestra la frecuencia relativa de las palabras y bigramas dentro de un texto. Aquellos términos que son más frecuentes (es decir, que aparecen más dentro de los documentos) se ven más grandes y se encuentran en el centro de la nube. A medida que los términos son menos frecuentes, pierden relevancia y por consiguiente tamaño en la representación visual. Este proceso se realizó para los PDT de cada departamento y cada municipio del que se tenía disponibilidad. De manera ilustrativa, en la Figura 8 se puede observar la nube de palabras para los PDT del departamento del Amazonas y el municipio Leticia.

Figura 8: Nubes de palabras departamento Amazonas(izquierda) municipio Leticia(derecha)



Fuente: elaboración propia

De igual forma existen las nubes de bigramas cuya única diferencia con las nubes de palabras es que muestran los términos compuestos por dos palabras (bigramas) que son más frecuentes. En la Figura 9 se observa la nube de bigramas para el departamento del Huila y el municipio Neiva.



Figura 9: Nube de bigramas departamento Huila (izquierda) municipio Agrado (derecha)

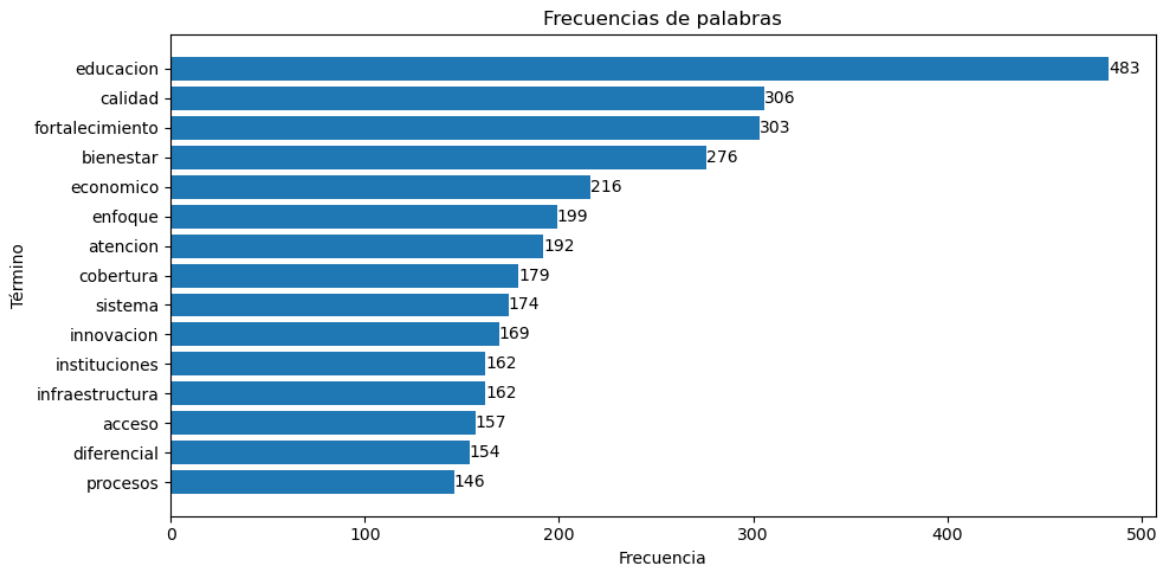


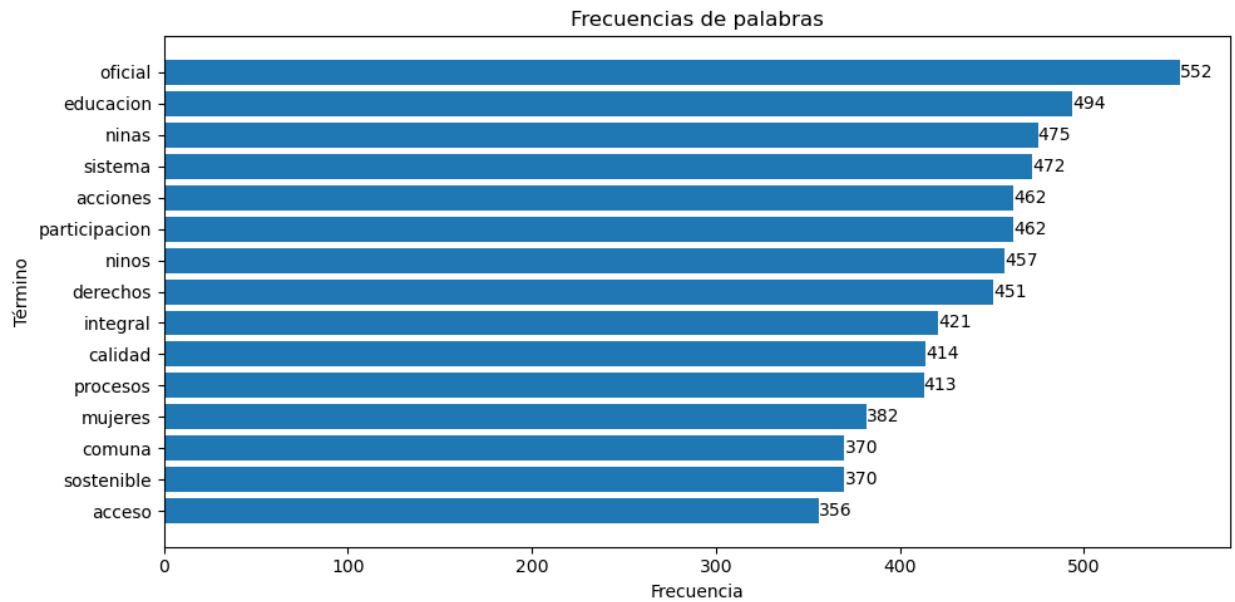
Fuente: elaboración propia

4.2. Gráficos de barras

Al igual que las nubes de palabras, los gráficos de barras muestran la frecuencia de los términos, en este caso los gráficos de barras de realizaron para los 15 términos más frecuentes de los documentos y se puede observar el número de veces que aparecen dentro de los documentos. En la Figura 10 se puede observar un ejemplo de los gráficos de barras para el PDT del departamento de Antioquia y municipio Medellín.

Figura 10: Gráficos de barras departamento Antioquia(arriba) municipio Medellín(abajo)

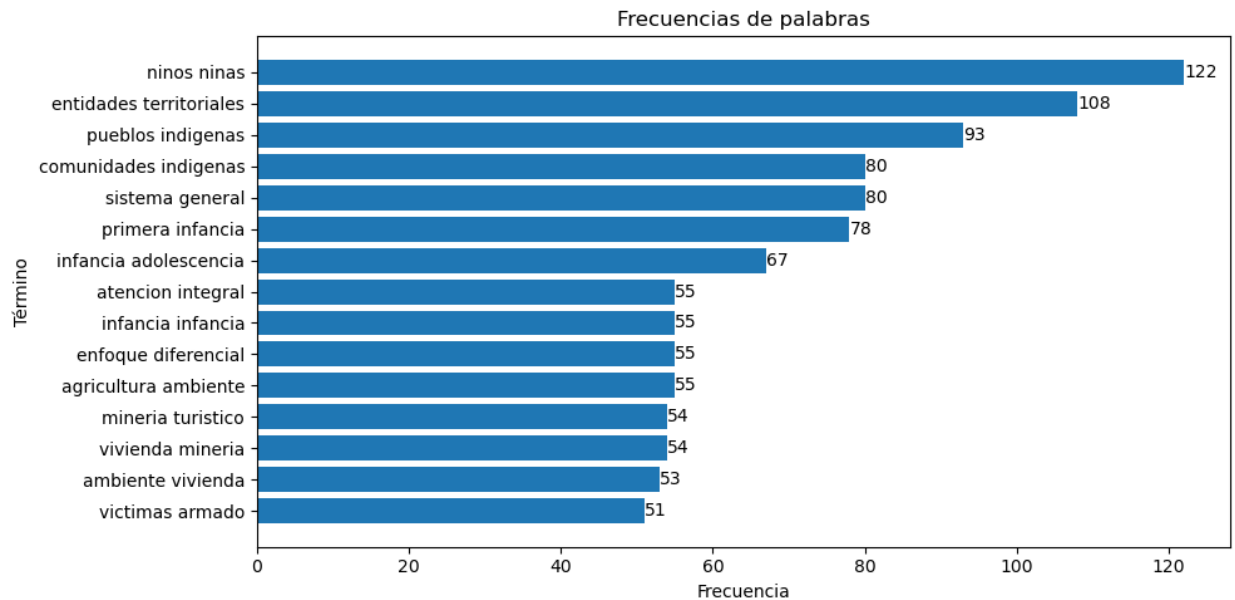


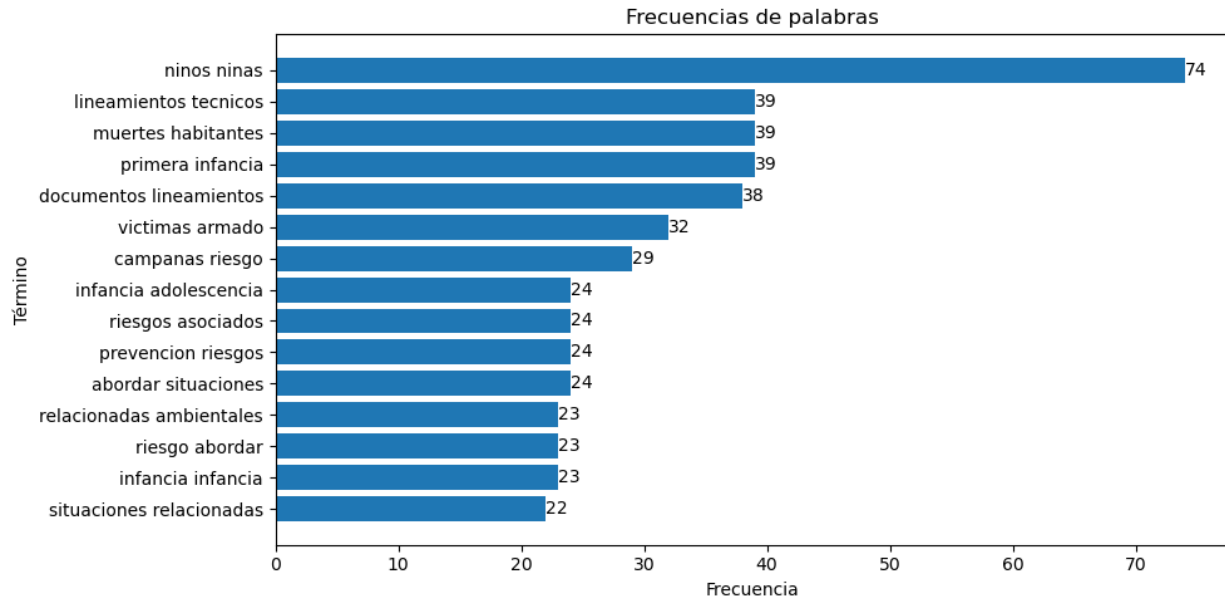


Fuente: elaboración propia

Asimismo, en la Figura 11 se observan los gráficos de barras con la frecuencia de los 15 bigramas con mayor aparición para los PDT del departamento de Vaupés y su capital Mitú.

Figura 11: Gráfico de barras para bigramas departamento Vaupés(arriba) y municipio Mitú(abajo)





Fuente: elaboración propia

4.3. Términos y frases en los PDT

Del proceso de búsqueda descrito en la Sección 3.4 de este documento se obtuvieron los resultados que se encuentran en los archivos anexos este documento. En este se encuentra la frecuencia de aparición de las palabras de interés que fueron buscadas en todos los documentos, a nivel municipal y departamental.

En la Tabla 5 se observan los departamentos cuyos PDT tienen mayor frecuencia de términos para cada una de las 30 categorías. Este es el resultado de la suma de la frecuencia de aparición de todos los términos preestablecidos para cada categoría.

Tabla 5: Departamentos que más mencionan cada categoría

Categoría	Departamento	Categoría	Departamento
Adulto mayor	Caldas	Ordenamiento territorial	Antioquia
Agricultura y desarrollo rural	Meta	Participación ciudadana	Caldas
Agua y saneamiento básico	Vaupés	Paz	Valle del cauca
Asuntos étnicos	Valle del cauca	Planeación y seguimiento	Córdoba
Cambio climático	Antioquia	Primera infancia y adolescencia	Santander
Cultura	Antioquia	Reactivación exclusivos	Antioquia
Derechos	Santander	Reactivación relevantes	Meta
Discapacidad	Santander	Recursos	Córdoba
Educación	Amazonas	Salud	Santander
Empleo y trabajo	Antioquia	Sector minero-energético	Chocó
Enfoque territorial	Valle del cauca	Seguridad	Antioquia
Equidad	Antioquia	Transparencia y lucha contra la corrupción	Caldas
Equidad de género	Santander	Transporte	Antioquia
Innovación y competitividad	Caldas	Turismo	Caldas



Justicia	Santander	Víctimas	Córdoba
----------	-----------	----------	---------

Fuente: elaboración propia

Estos resultados pueden dar un indicio de los temas más representativos para el plan de gobierno de cada departamento y las categorías que son más tratadas en sus documentos.

De igual forma se realiza este procedimiento para los términos de las categorías de articulación subnacional. En el Anexo 1 se observa la frecuencia total de aparición de los términos correspondientes a cada categoría en los PDT departamentales y municipales.

En la Tabla 6 se observan los 20 términos más frecuentes de las categorías de articulación subnacional encontrados en los PDT municipales y departamentales.



Tabla 6: términos más frecuentes categorías articulación subnacional

Departamentos		Municipios	
Palabra	Frecuencia	Palabra	Frecuencia
departamento	13235	departamento	67333
departamental	9714	plan estratégico	29556
región	2874	departamental	27058
regional	2281	regional	20162
plan estratégico	1672	región	16408
proyecto departamental	1241	asociación	9836
desarrollo regional	1066	desarrollo regional	9485
inversiones departamentales	780	esquema	8158
asociación	610	inversiones públicas	7937
inversiones públicas	596	coordinación	6863
objetivo de desarrollo departamental	575	am	6595
coordinación	550	cofinanciación	5699
proyecto regional	484	enfoque territorial	5065
enfoque territorial	475	relaciones municipales	5013
esquema	443	asociación de municipios	4114
gestión regional	387	gestión regional	4043
provincia	347	proyecto departamental	4012
cofinanciación	331	proyecto regional	3853
nivel regional	324	nivel regional	3488
gestión asociada	264	objetivo de desarrollo departamental	2852

Fuente: elaboración propia

En la Tabla 7 se observan los 20 términos más frecuentes relacionados con la categoría articulación subnacional- términos que implican articulación subnacional para los PDT departamentales y municipales. Asimismo, en el Anexo 2 se encuentra este análisis para las categorías restantes.

Tabla 7: Palabras relacionadas más frecuentes para la categoría articulación subnacional- términos que implican articulación subnacional

Departamentos		Municipios	
Palabra	Frecuencia	Palabra	Frecuencia
desarrollo	9612	plan	36542
plan	6909	municipio	25147
territorial	6440	desarrollo	22942
departamento	5962	municipal	20438



social	5832	programa	19887
público	5612	nacional	16034
servicio	4989	social	13578
programa	4847	territorial	13372
municipio	4823	público	12471
secretaría	4563	estratégico	11979
departamental	4404	sector	11235
nacional	4283	gestión	11106
base	3562	proyecto	10905
sector	3355	servicio	10405
recurso	3335	sostenible	10361
producto	3137	objetivo	10352
acción	2822	acción	9766
participación	2809	ambiental	9709
objetivo	2671	línea	8331
indicador	2624	producto	8103

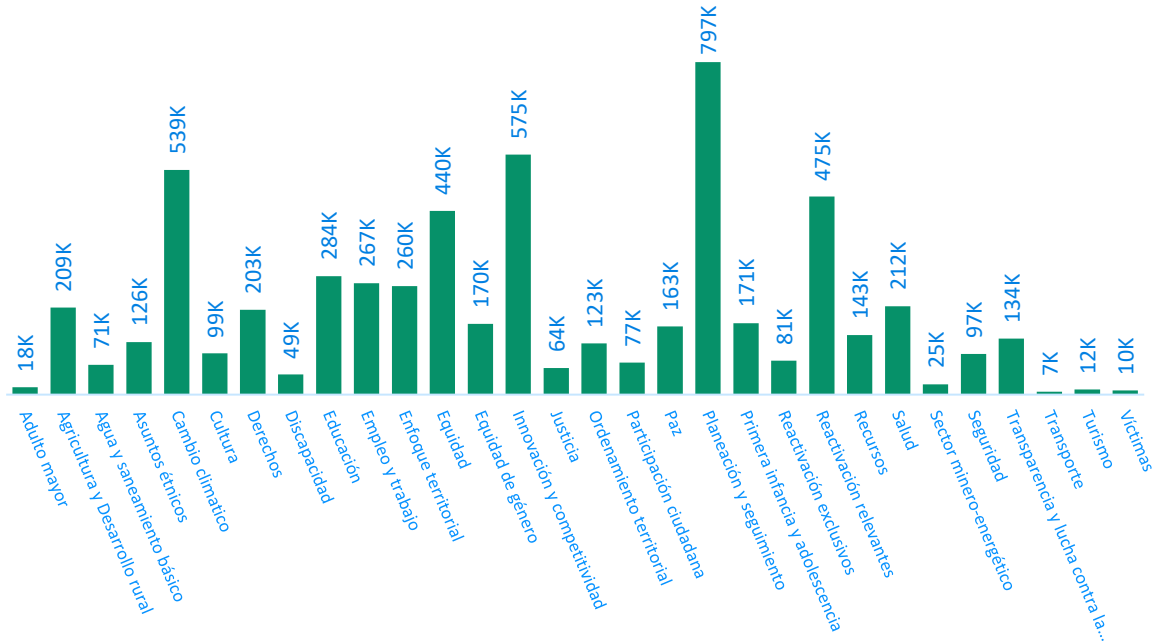
Fuente: elaboración propia

4.4. ***Términos relacionados con las palabras identificadas***

Los resultados obtenidos del proceso descrito en la Sección 3.5 se puede consultar en los archivos anexos este documento. La Figura 12 muestra qué tan frecuentes son todos los términos relacionados con cada categoría para el total de departamentos.



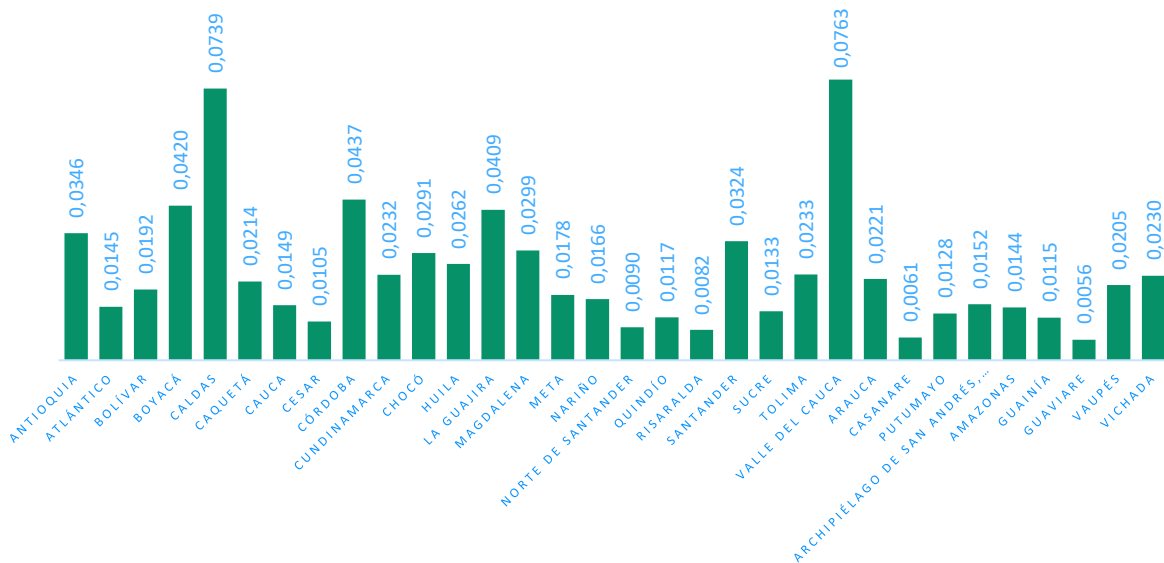
Figura 12: Frecuencia del total de términos por categoría para todos los departamentos



Fuente: elaboración propia

Las categorías Planeación y seguimiento, innovación y competitividad y cambio climático son los más mencionados en todos los PDT departamentales. En las Figuras 13, 14 y 15 se observa la distribución de la cantidad de términos para todos los departamentos en cada una de estas categorías.

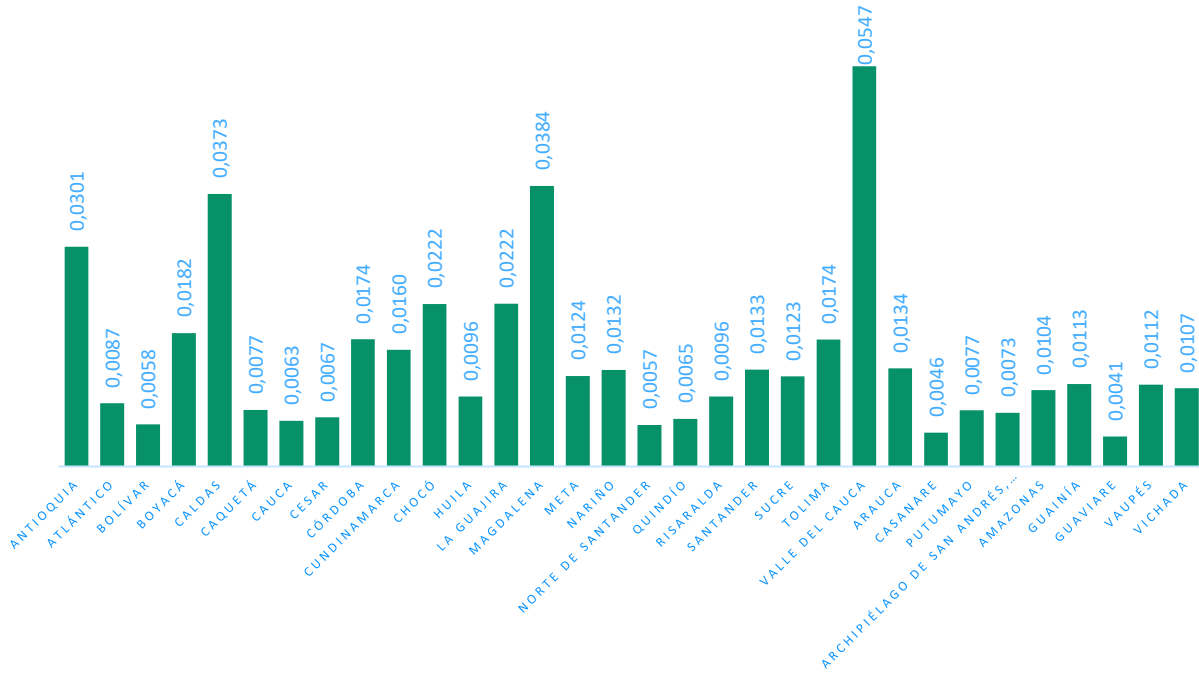
Figura 13: Cantidad de términos normalizados para todos los departamentos, categoría Planeación y seguimiento



Fuente: elaboración propia

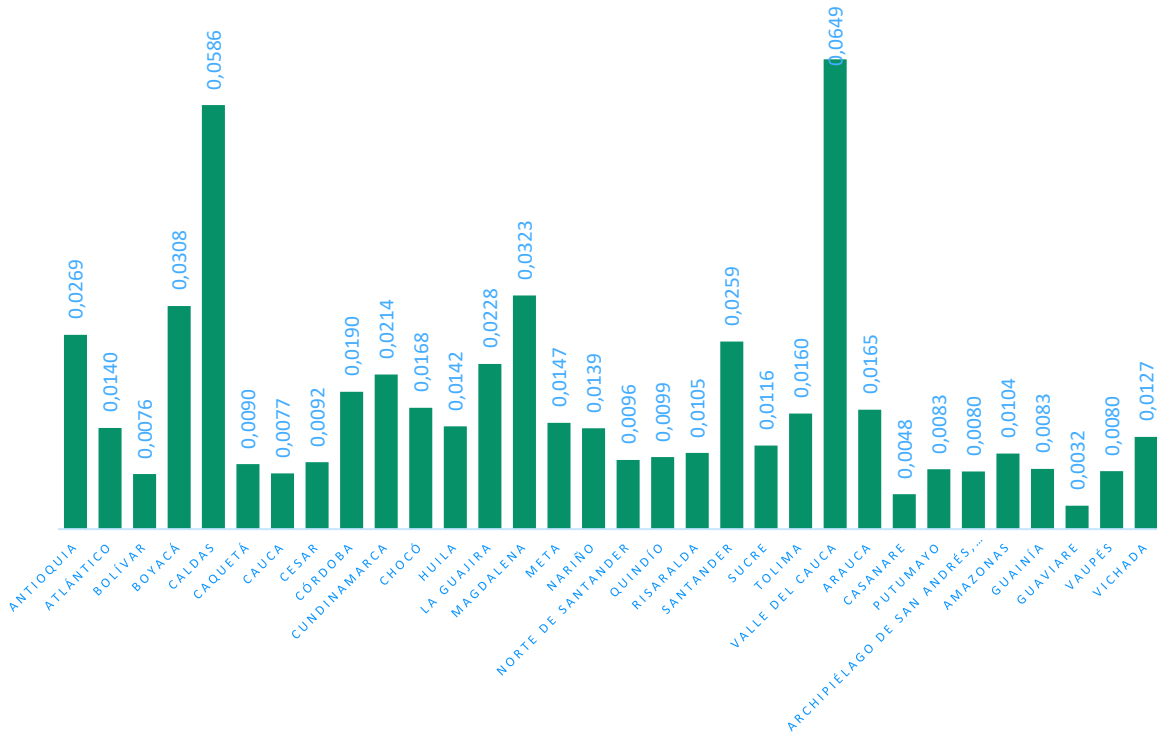


Figura 14: Cantidad de términos normalizados para todos los departamentos, categoría Cambio climático



Fuente: elaboración propia

Figura 15: Cantidad de términos normalizados para todos los departamentos, categoría Innovación y competitividad

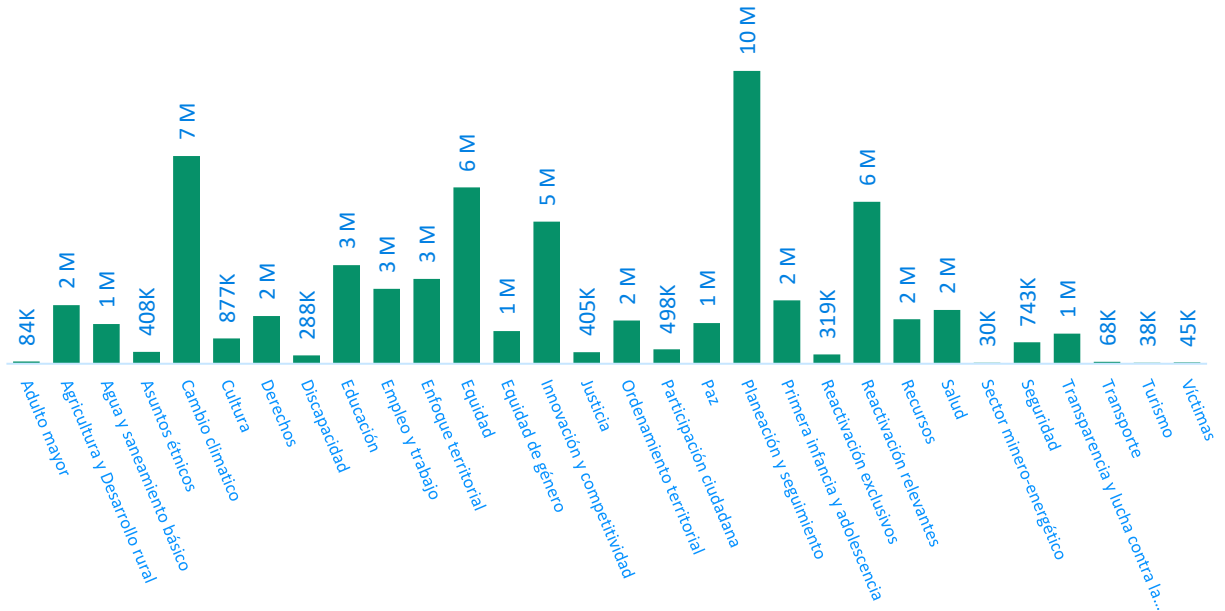


Fuente: elaboración propia



La Figura 16 muestra qué tan frecuentes son todos los términos relacionados con cada categoría para el total de municipios.

Figura 16: frecuencia del total de términos por categoría para todos los municipios.

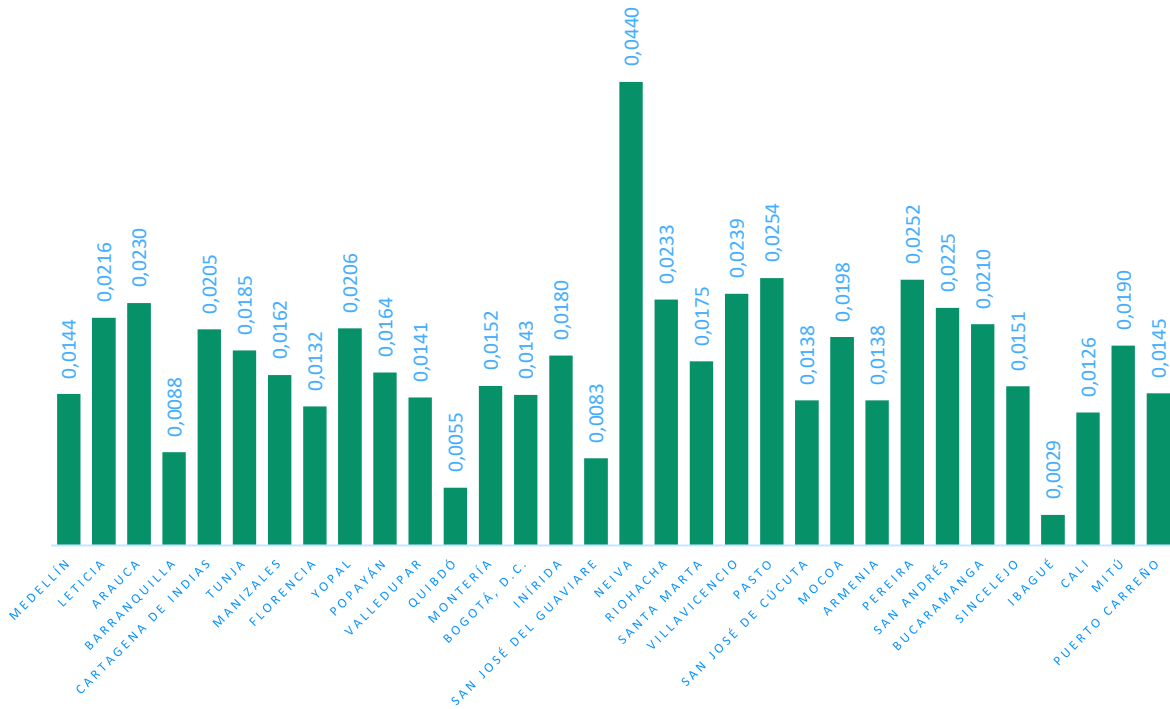


Fuente: elaboración propia

Las categorías Planeación y Seguimiento, Cambio Climático y los términos de Reactivación relevantes son los más mencionados en todos los PDT municipales. En las Figuras 17, 18 y 19 se observa la distribución de la cantidad de términos para todos los municipios capital en cada una de estas categorías.

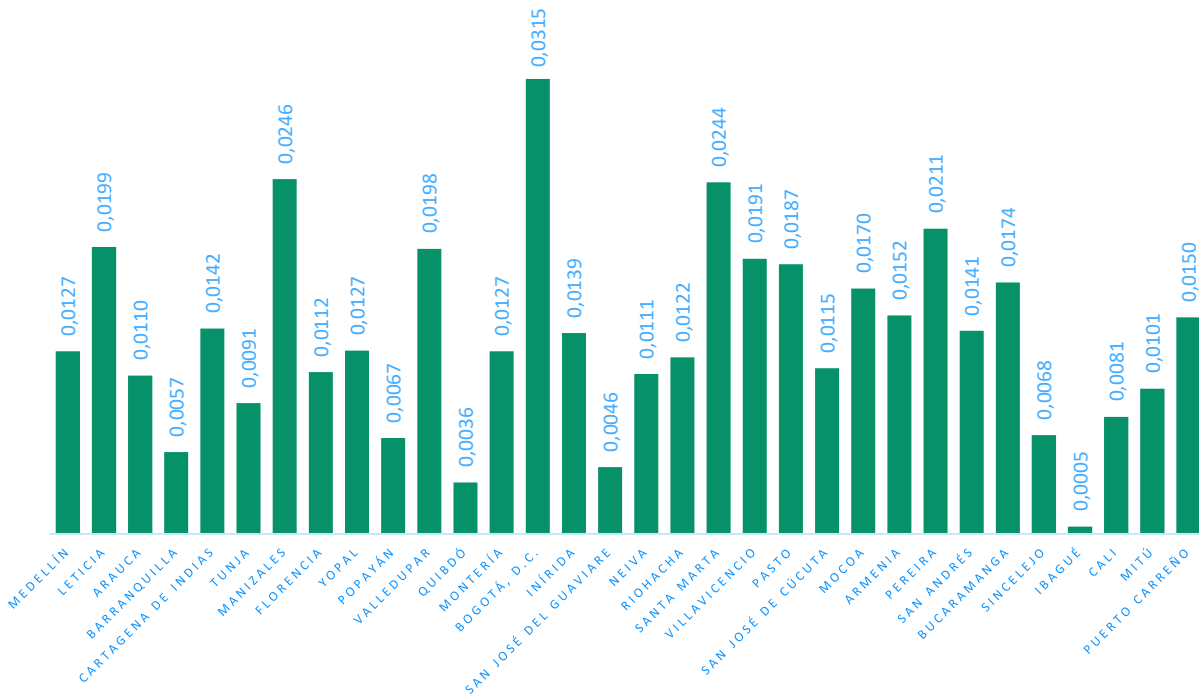


Figura 17: Cantidad de términos normalizados para los municipios capital, categoría Planeación y seguimiento



Fuente: elaboración propia

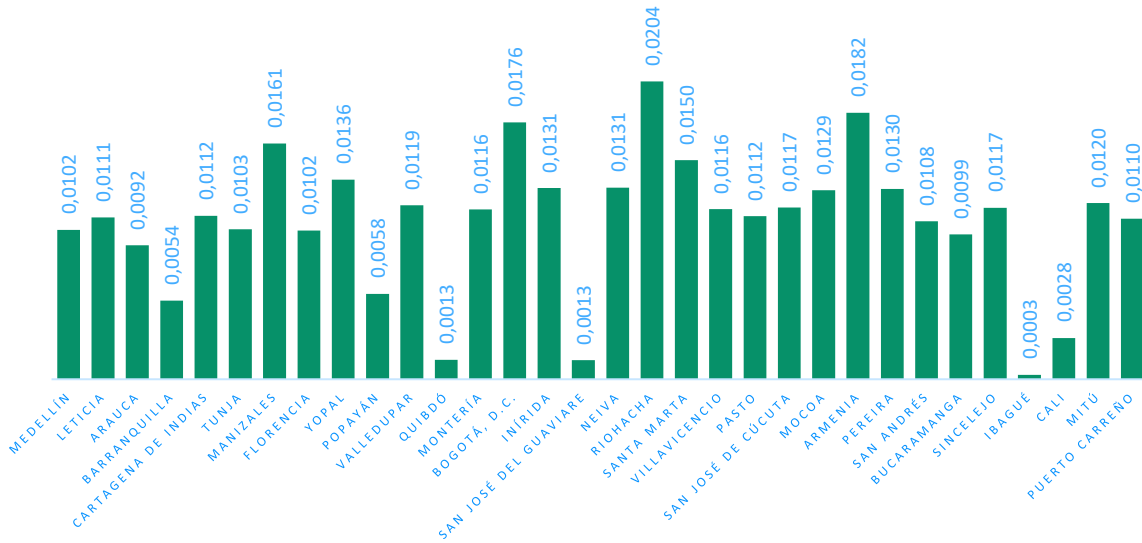
Figura 18: Cantidad de términos normalizados para los municipios capital, categoría Cambio climático



Fuente: elaboración propia



Figura 19: Cantidad de términos normalizados para los municipios capital, categoría Reactivación relevantes



En la Tabla 8 y Tabla 9 se observan las 20 palabras que más ocurren (de mayor a menor ocurrencia) alrededor de los términos relacionados con la categoría “Agricultura y desarrollo rural” y “Empleo y trabajo” para todos los departamentos y municipios respectivamente.

Tabla 8: palabras relacionadas con agricultura y desarrollo rural para todos los departamentos

Palabra	Cantidad	Palabra	Cantidad	Palabra	Cantidad	Palabra	Cantidad
1. seguro	2002	6. base	1496	11. atención	1303	16. territorial	1247
2. población	1773	7. política	1454	12. derecho	1289	17.indicador	1207
3. sector	1680	8. producto	1407	13.proyecto	1287	18. sistema	1184
4. nacional	1604	9. estrategia	1374	14. persona	1253	19.rural	1167
5. acción	1593	10. número	1366	15. nivel	1249	20. integral	1157

Fuente: elaboración propia

Tabla 9: palabras relacionadas con Empleo y trabajo para todos los municipios

Palabra	Cantidad	Palabra	Cantidad	Palabra	Cantidad	Palabra	Cantidad
1. desarrollo	123212	6.sector	44654	11.servicio	31963	16.rural	21695
2. población	59285	7.educación	43722	12.producto	28127	17.calidad	16963
3. plan	58991	8.gestión	43701	13.acceso	25634	18.participación	15221
4. social	52190	9.número	35416	14trabajo	24849	19.política	14057
5.programa	48741	10.persona	34228	15.indicador	23737	20.bienestar	13730

Fuente: elaboración propia

4.5. Similitud entre documentos

Teniendo en cuenta que la comparación por páginas en ocasiones arroja similitudes muy bajas porque el tamaño de los documentos de comparación es de más de una página, se realizan iteraciones agrupando los documentos



PDT por diferentes números de páginas hasta encontrar el que mayores similitudes genere. De este proceso se encuentra que las agrupaciones óptimas para cada grupo de comparación son las siguientes.

PDT, documentos RAPE central: 4 páginas

PDT, documentos AMCO: 2 páginas

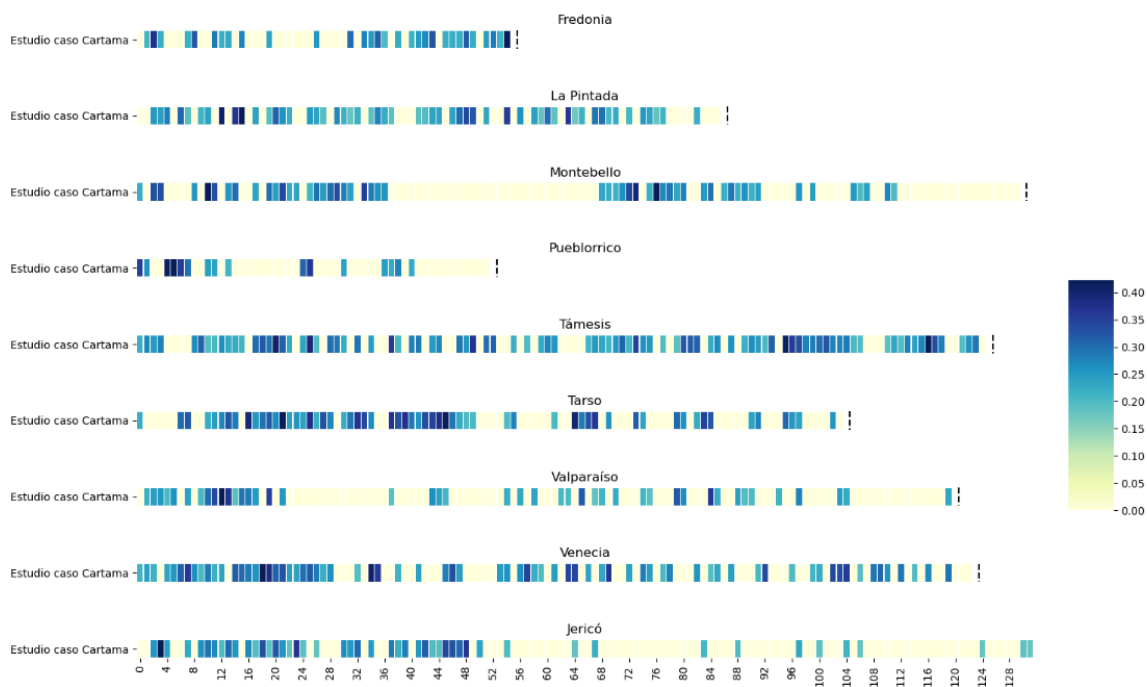
PDT, documentos CARTAMA: 2 páginas

PDT, documentos AMUNORCA: 2 páginas

PDT, pacto pacífico PND: 4 páginas

Los resultados de este proceso se resumen en los siguientes mapas de calor:

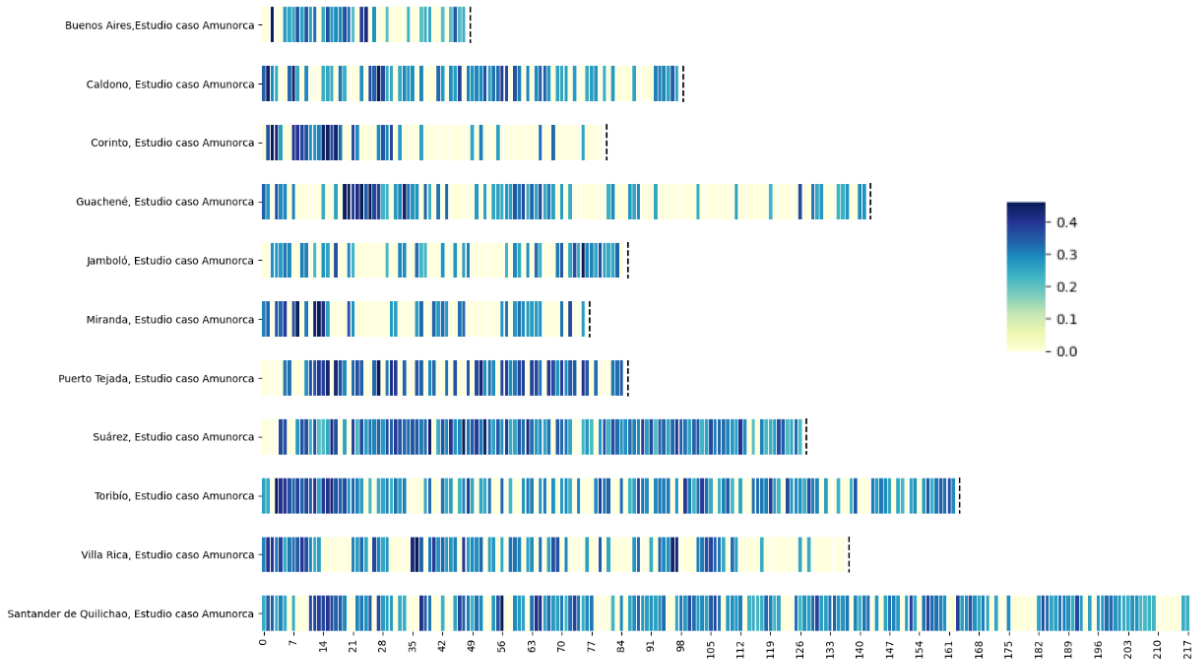
Figura 20: similitud representada en mapa de calor para los municipios CARTAMA



Fuente: elaboración propia

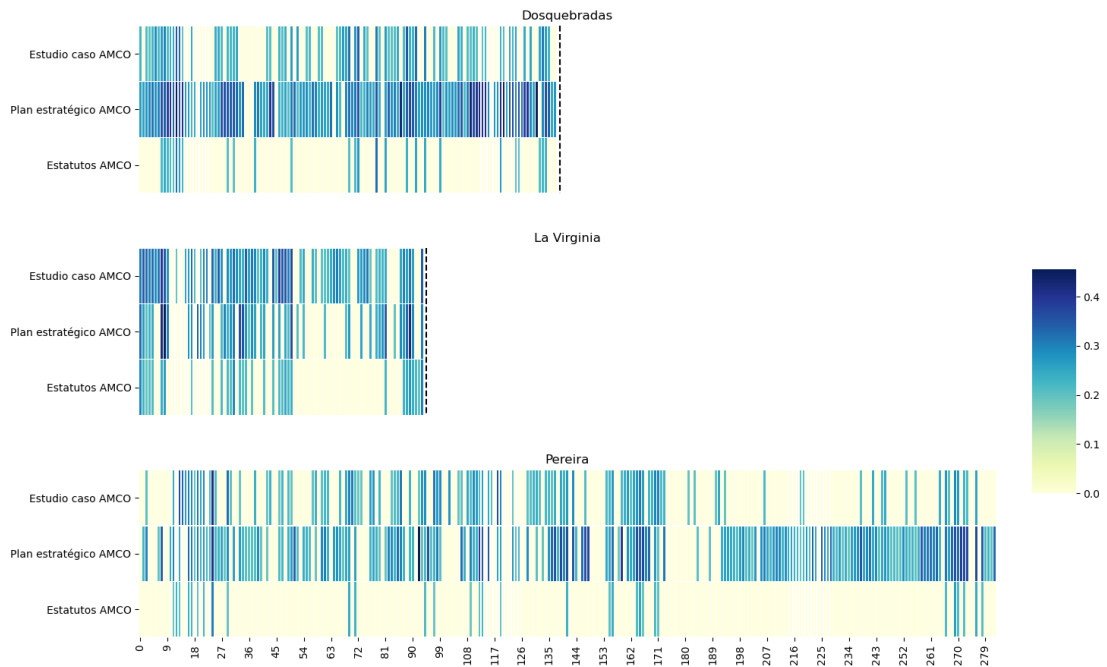


Figura 21: similitud representada en mapa de calor para los municipios AMUNORCA



Fuente: elaboración propia

Figura 22: similitud representada en mapa de calor para los municipios AMCO



Fuente: elaboración propia



Figura 23: similitud representada en mapa de calor para los departamentos RAPE y de la región pacífica



Fuente: elaboración propia

En los mapas de calor se observa la similitud entre el grupo de páginas y su respectivo documento de comparación. En el eje x aparece el número de páginas (grupos de 2 o 4 páginas) que tiene el documento más largo y con las líneas punteadas negras se señala la terminación de cada documento. Los tonos de azul más oscuros indican una mayor similitud entre los documentos comparados y en cuanto el color se acerca a los tonos amarillos la similitud es más baja, estas convenciones se pueden observar en la barra de color que se encuentra al lado derecho de las figuras

Además de los mapas de calor, y con el fin de tener un indicador total de la similitud entre cada documento de comparación con su respectivo PDT se pondera la similitud de todas las páginas por documento. En la Tabla 10 se observa esta información. En la columna “Promedio” está el promedio de la similitud para todos los grupos de páginas y en la columna “Promedio ponderado” se encuentra el promedio de similitud para los grupos de páginas que tienen un valor de similitud superior a cierto umbral definido para evitar la inclusión de similitudes muy bajas que no representan valor para el análisis, este umbral se definió para todos los municipios como 0.2 y para los departamentos como 0.4.

Tabla 10: similitud promedio entre los PDT y documentos de comparación

PDT	Documento de comparación	Promedio	Promedio Ponderado
Dosquebradas	Estudio caso AMCO	0,21	0,24
	Plan estratégico AMCO	0,30	0,30
	Estatutos AMCO	0,16	0,24
La Virginia	Estudio caso AMCO	0,25	0,28
	Plan estratégico AMCO	0,22	0,28
	Estatutos AMCO	0,17	0,23



Pereira	Estudio caso AMCO	0,19	0,26
	Plan estratégico AMCO	0,24	0,28
	Estatutos AMCO	0,13	0,23
Buenos Aires	Estudio caso AMUNORCA	0,24	0,39
Caldono	Estudio caso AMUNORCA	0,22	0,35
Corinto	Estudio caso AMUNORCA	0,19	0,51
Guachené	Estudio caso AMUNORCA	0,20	0,42
Jambolo	Estudio caso AMUNORCA	0,20	0,40
Miranda	Estudio caso AMUNORCA	0,19	0,41
Puerto Tejada	Estudio caso AMUNORCA	0,22	0,35
Suárez	Estudio caso AMUNORCA	0,29	0,33
Toribio	Estudio caso AMUNORCA	0,24	0,34
Villa Rica	Estudio caso AMUNORCA	0,21	0,39
Santander de Quilichao	Estudio caso AMUNORCA	0,24	0,32
Fredonia	Estudio caso CARTAMA	0,22	0,38
La Pintada	Estudio caso CARTAMA	0,24	0,37
Montebello	Estudio caso CARTAMA	0,18	0,34
Pueblorrico	Estudio caso CARTAMA	0,17	0,36
Támesis	Estudio caso CARTAMA	0,25	0,35
Tarso	Estudio caso CARTAMA	0,24	0,35
Valparaíso	Estudio caso CARTAMA	0,20	0,37
Venecia	Estudio caso CARTAMA	0,23	0,35
Jericó	Estudio caso CARTAMA	0,19	0,36
Boyacá	Estudio de caso RAPE	0,21	0,34
	Plan estratégico RAPE	0,29	0,38
Cauca	Pacto pacífico PND	0,26	0,39
Cundinamarca	Estudio de caso RAPE	0,26	0,36
	Plan estratégico RAPE	0,37	0,42
Valle	Pacto pacífico PND	0,25	0,39
Huila	Estudio de caso RAPE	0,22	0,33
	Plan estratégico RAPE	0,29	0,36
Meta	Estudio de caso RAPE	0,19	0,31
	Plan estratégico RAPE	0,28	0,39
Nariño	Pacto pacífico PND	0,27	0,39
Tolima	Estudio de caso RAPE	0,21	0,31
	Plan estratégico RAPE	0,29	0,37
Chocó	Pacto pacífico PND	0,20	0,39

Fuente: elaboración propia



4.6. Similitud entre PDT y PND

Al realizar el análisis de similitud se pueden identificar los PDT municipales y departamentales que mayor similitud tienen con el PND y cada capítulo.

En la Tabla 11 y Tabla 12 se pueden ver los PDT departamentales y municipales que tienen mayor similitud con el PND y cada capítulo analizado.

Tabla 11: PDT departamentales más similares al PND y sus capítulos

PND	PDT	Similitud
PND	Risaralda	0,87
Capítulo equidad	Cesar	0,78
Capítulo legalidad	Risaralda	0,65
Capítulo emprendimiento	Meta	0,66
Capítulo descentralización	Quindío	0,69
Capítulo productividad	Guainía	0,49
Capítulo construcción paz	Arauca	0,62
Capítulo equidad mujeres	Chocó	0,56
Capítulo equidad grupos	Chocó	0,55

Fuente: elaboración propia

Tabla 12: PDT municipales más similares al PND y sus capítulos

PND	PDT	Similitud
PND	Génova	0,87
Capítulo equidad	Montería	0,83
Capítulo legalidad	Copacabana	0,70
Capítulo emprendimiento	Dosquebradas	0,65
Capítulo descentralización	Sopó	0,72
Capítulo productividad	El Rosario	0,59
Capítulo construcción paz	Roberto Payán	0,64
Capítulo equidad mujeres	Casabianca	0,71
Capítulo equidad grupos	Guapi	0,59

Fuente: elaboración propia

Además, en la Tabla 13 se puede observar el promedio de similitud a nivel departamental y municipal con el PND y cada capítulo.

Tabla 13: promedio de similitud departamental y municipal

PND	Promedio departamental	Promedio municipal
PND	0,79	0,69
Capítulo equidad	0,70	0,62
Capítulo legalidad	0,55	0,47
Capítulo emprendimiento	0,54	0,46
Capítulo descentralización	0,52	0,46

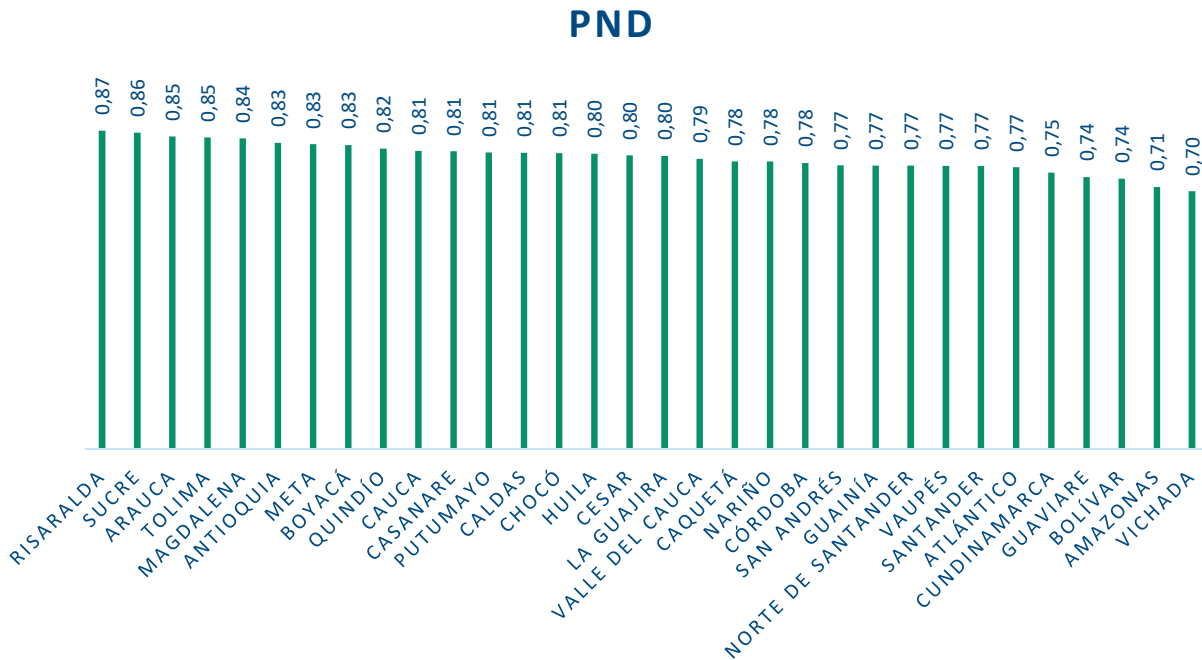


Capítulo productividad	0,37	0,31
Capítulo construcción paz	0,52	0,46
Capítulo equidad mujeres	0,39	0,33
Capítulo equidad grupos	0,44	0,38

Fuente: elaboración propia

En las siguientes figuras se puede observar el ranking de similitud entre los PDT departamentales y el PND. En el Anexo 3 **Error! No se encuentra el origen de la referencia.** se encuentra esta información para cada capítulo.

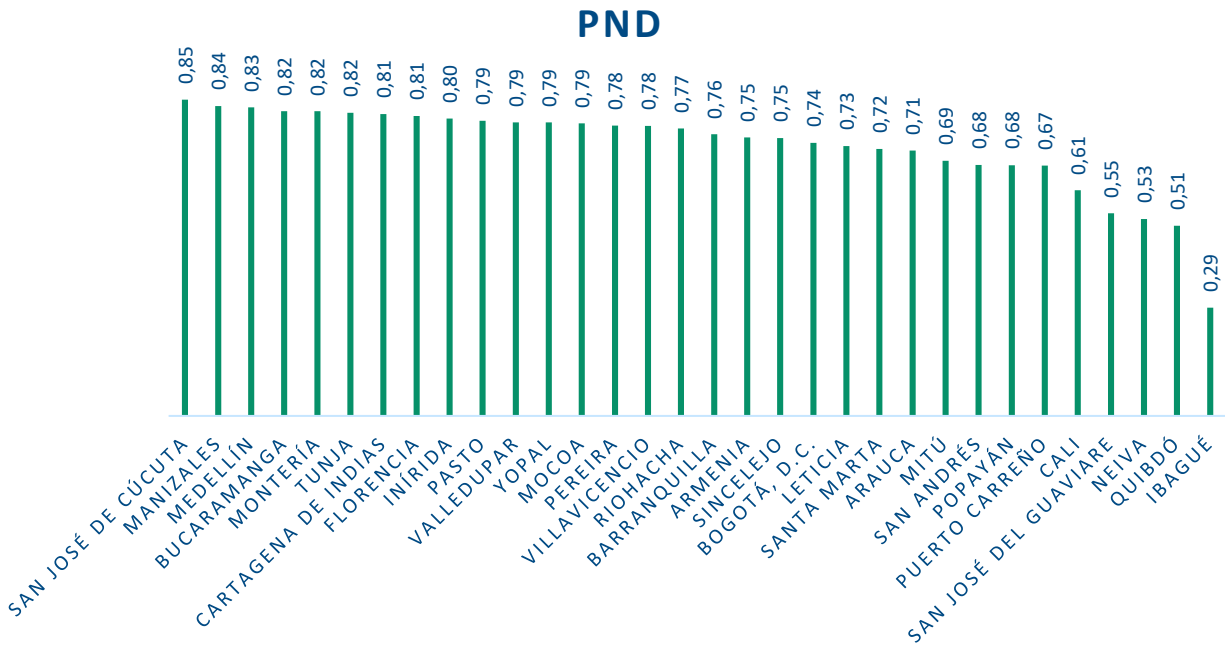
Figura 24: similitud entre PDT departamentales y el PND



Fuente: elaboración propia



Figura 25: similitud entre los PDT municipales y el PND



Fuente: elaboración propia

4.7. Palabras exclusivas

Del proceso descrito en la Sección 3.8 se obtienen los siguientes resultados:

En la Tabla 14 se pueden observar los 10 términos exclusivos más frecuentes para cada capítulo del PND y en el Anexo 4 se encuentra esta información para los PDT municipales y departamentales.

Tabla 14: top 10 términos exclusivos capítulos PND

Capítulo equidad	Capítulo legalidad	Capítulo emprendimiento	Capítulo descentralización	Capítulo productividad	Capítulo construcción paz	Capítulo equidad mujeres	Capítulo equidad grupos
vejez	consulado	mincomercio	inmueble	malformación	guardar	empareja	tentación
saludable	lavado	fitosanitario	macroproyecto	pirámide	antipersona	ingeniería	afrocolombiano
prejuicio	blando	mercantil	equilibrado	protesta		igualitario	indígenas
deportista	judicialización	intangible	suburbanización	traza		reproducción	indigno
barrio	terrorismo	bancario	aeródromo			anticonceptivo	nativo
rendimiento	anticorrupción	firmado	edificabilidad			marital	amazónica
recreativo	carcelario	patente	ecosistémico			feminidad	restauración
federación	dni	procolombia	aeropuerto			casada	sagrado
malnutrición	disfunción	sanidad	obsolescencia			sobrecarga	tradición
propietario	ciberespacio	microfinanza	multi propósito			popular	bélica

Fuente: elaboración propia



5. Conclusiones y recomendaciones

1. Algunos PDT municipales no fueron tenidos en cuenta dentro del análisis dado que del proceso de lectura no se obtuvieron resultados óptimos que permitieran realizar un análisis de calidad. En el Anexo 5 se encuentran los códigos divipola de los municipios que fueron excluidos.
2. Se desarrolló una metodología de búsqueda de términos relacionados con cada tema basado en expresiones regulares, lo que permite encontrar múltiples formas de una frase en los textos.
3. El PDT departamental que mayor similitud presenta con el PND respecto al uso de los términos de las 30 categorías es el de Risaralda, en cuanto al municipal es el PDT de Génova.
4. El promedio de similitud de todos los PDT departamentales con respecto al PND es 0,79 y a nivel municipal 0,69.
5. El capítulo de productividad del PND es el que presenta menor similitud en promedio con respecto a los PDT departamentales y municipales, por otra parte el que mayor similitud presenta es el capítulo de equidad.
6. Las palabras exclusivas son una herramienta para diferenciar el contenido de los diferentes documentos analizados.
7. La lectura y limpieza de los documentos no arroja resultados perfectos, esto ocasiona que los textos de análisis contengan palabras que no tienen contexto, con el fin de disminuir estas palabras se llevó a cabo una corrección ortográfica que incrementó el tiempo de desarrollo.
8. Se recomienda validar los resultados de las palabras exclusivas desde la vista de un experto dado que la corrección ortográfica puede ocasionar errores en los resultados finales.

6. Socialización

El proyecto ha sido socializado con el Grupo de Estudios Territoriales de la Dirección de Descentralización y Desarrollo Regional y con la Dirección de Desarrollo Digital del DNP.

Contacto

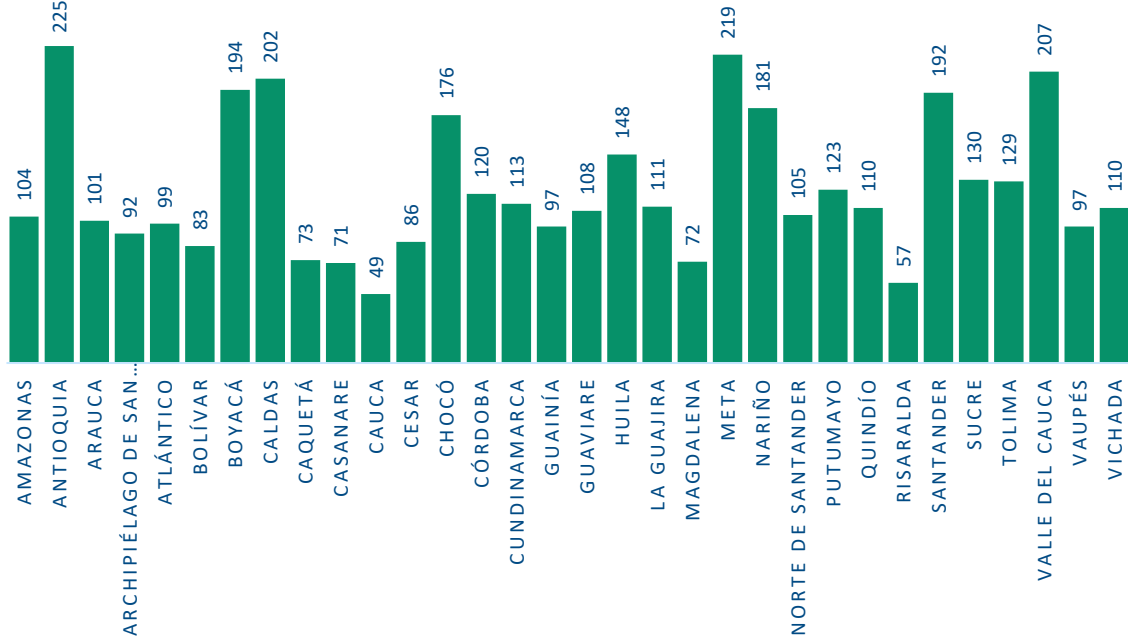
Si tiene alguna duda, comentario o sugerencia sobre este proyecto, o si le gustaría conversar con la Unidad de Científicos de Datos sobre la posibilidad de una nueva fase para el mismo, puede comunicarse con nosotros a través del correo electrónico ucd@dnpp.gov.co.



ANEXOS

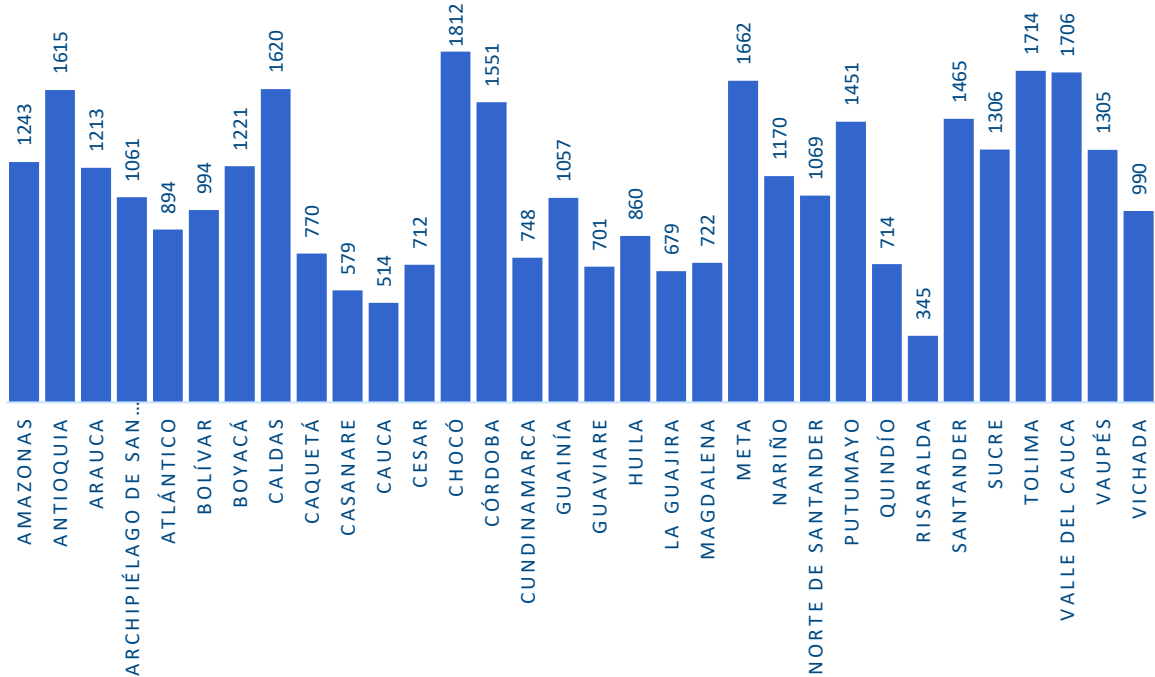
Anexo 1

Figura 26: frecuencia total de términos de articulación subnacional



Fuente: elaboración propia

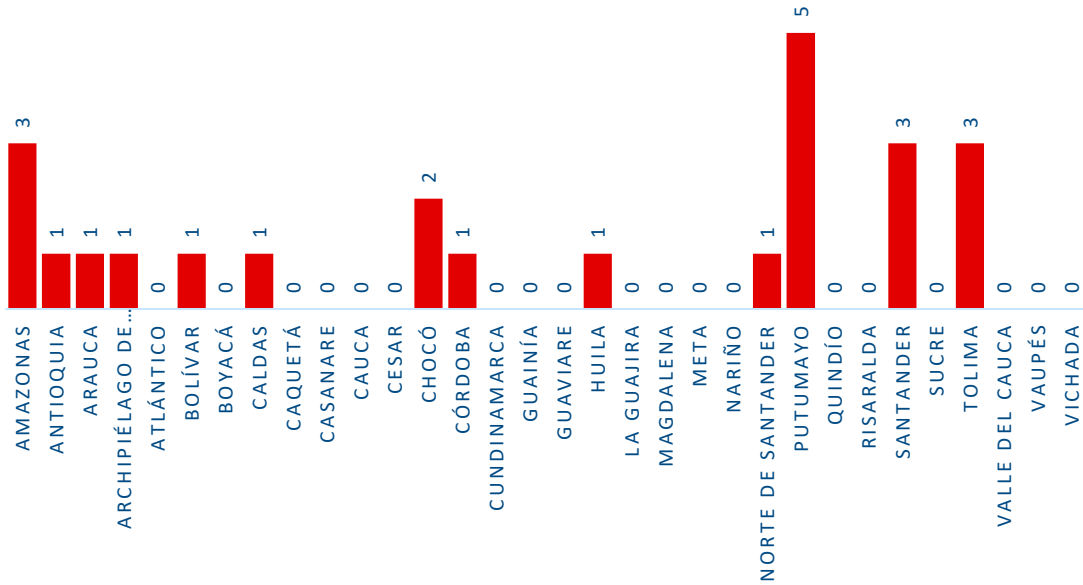
Figura 27: frecuencia total de términos de enfoque territorial





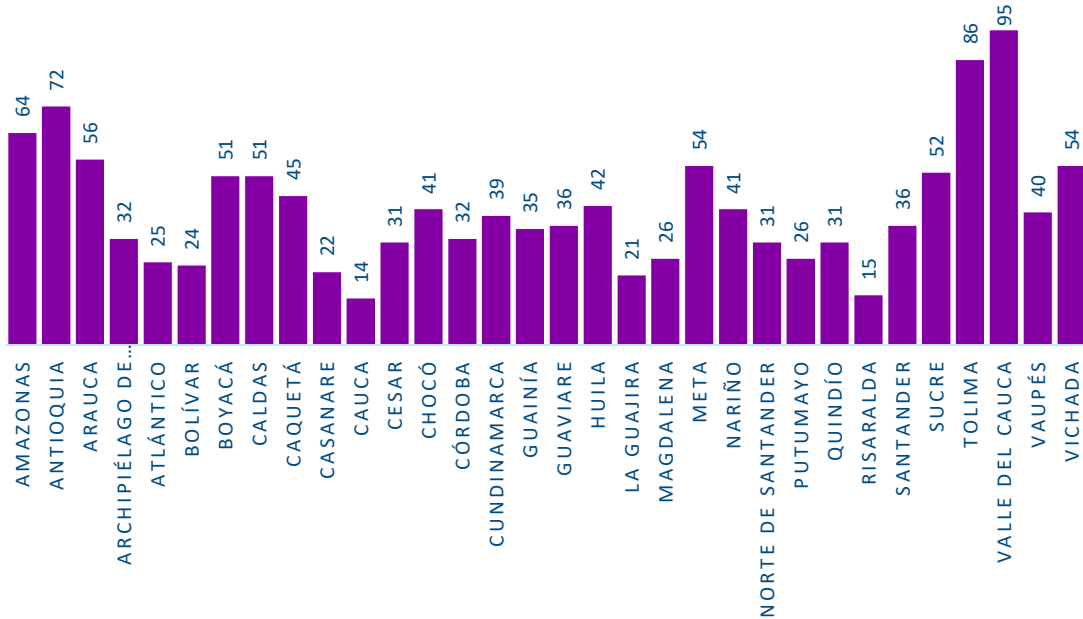
Fuente: elaboración propia

Figura 28: frecuencia total de términos de normatividad



Fuente: elaboración propia

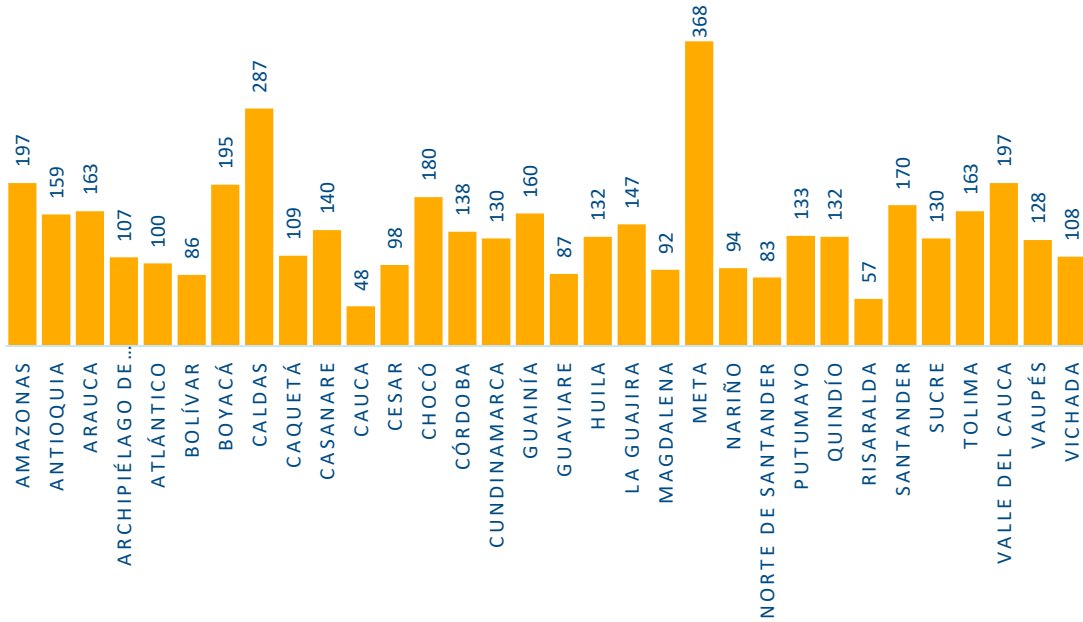
Figura 29: frecuencia total de términos de ordenamiento territorial



Fuente: elaboración propia

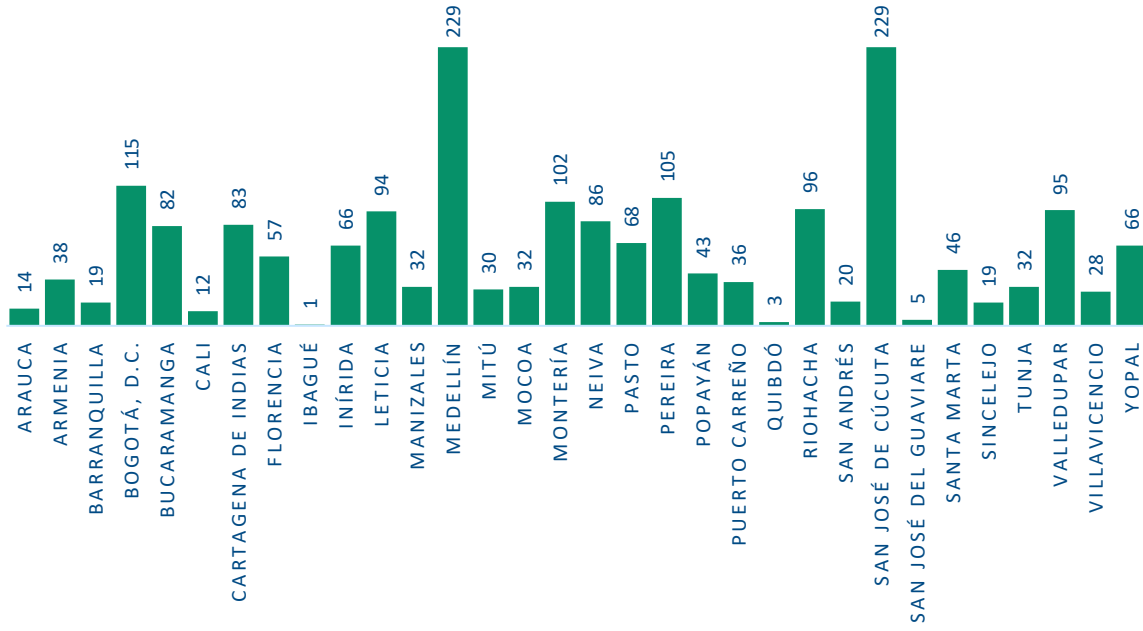


Figura 30: frecuencia total de términos de recursos



Fuente: elaboración propia

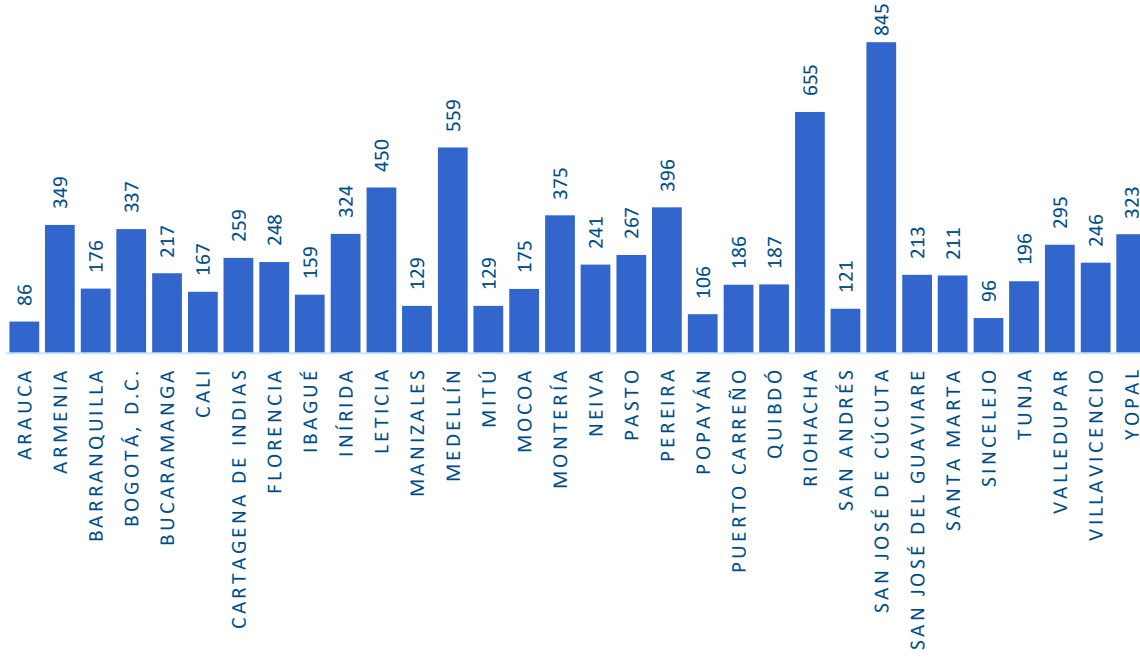
Figura 31: frecuencia total de términos de articulación subnacional



Fuente: elaboración propia

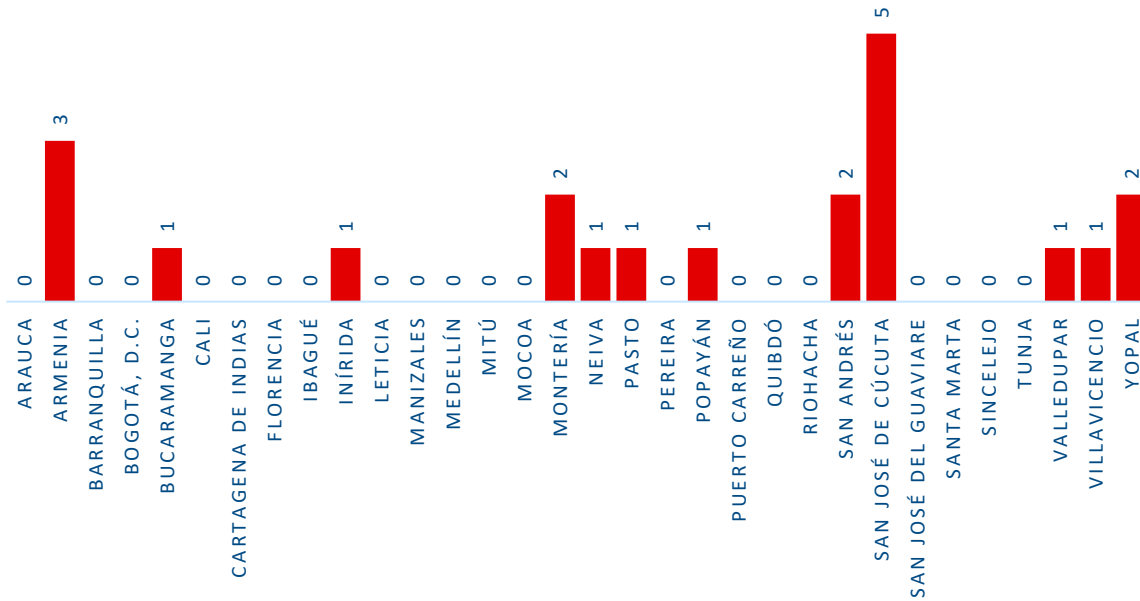


Figura 32: frecuencia total de términos de enfoque territorial



Fuente: elaboración propia

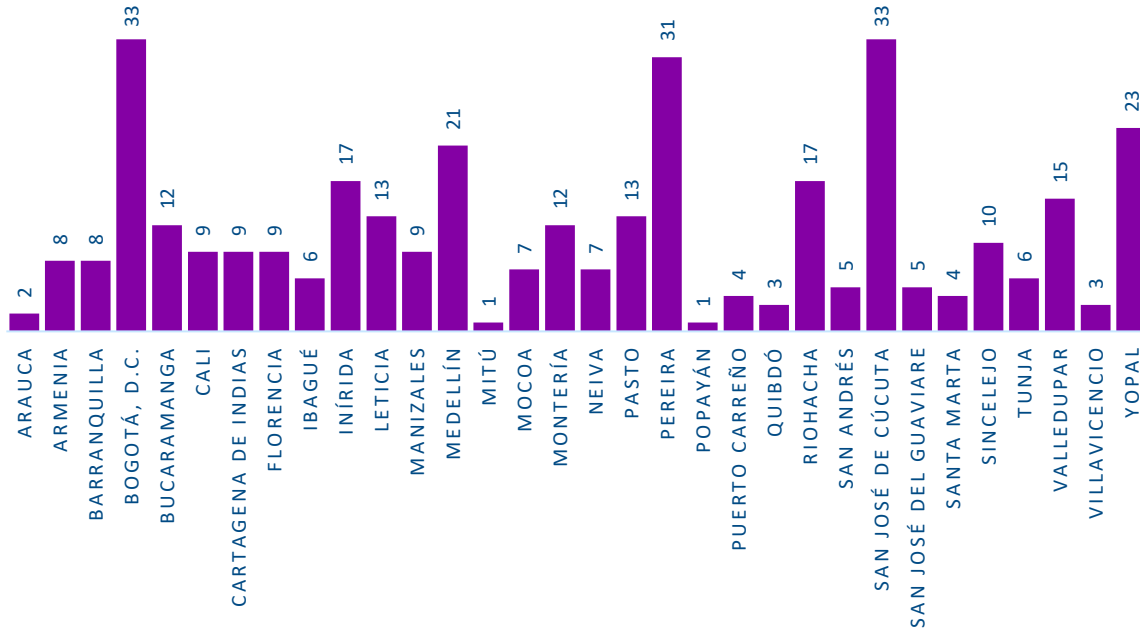
Figura 33: frecuencia total de términos de normatividad



Fuente: elaboración propia

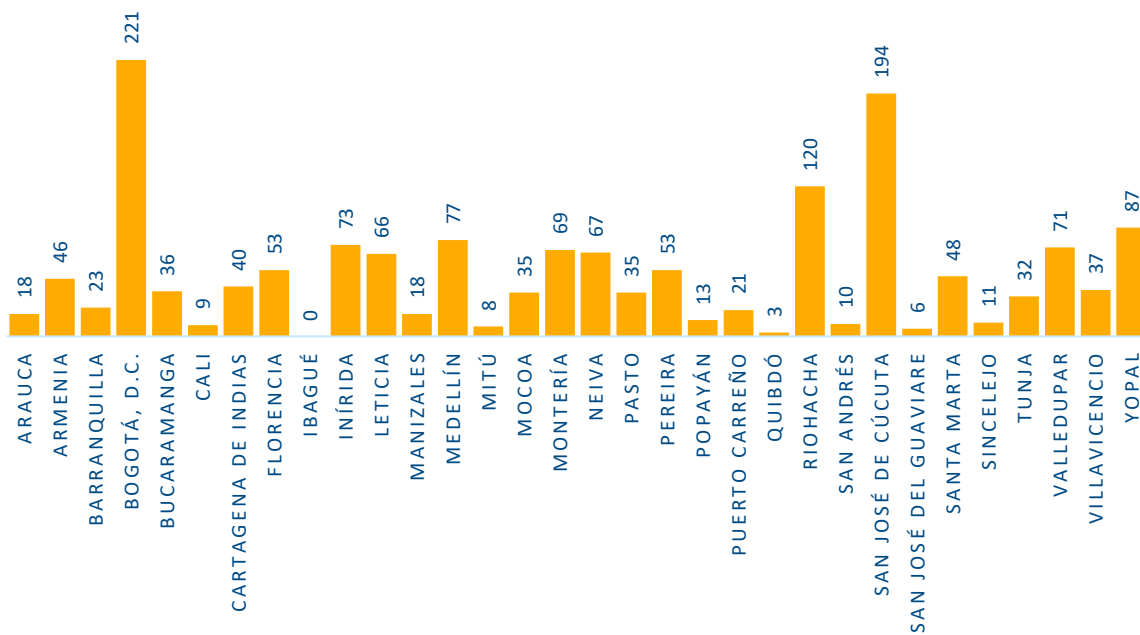


Figura 34: frecuencia total de términos de ordenamiento territorial



Fuente: elaboración propia

Figura 35: frecuencia total de términos de recursos



Fuente: elaboración propia



Anexo 2

Tabla 15: Palabras relacionadas más frecuentes para la categoría Enfoque territorial -términos que implican articulación subnacional

Departamentos		Municipios	
Palabra	Frecuencia	Palabra	Frecuencia
desarrollo	13335	municipio	208457
departamento	10831	desarrollo	190648
plan	8000	por	185916
servicio	5245	municipal	134068
departamental	5196	plan	117914
programa	4851	programa	78075
nacional	4682	nacional	72650
municipio	4429	territorial	58920
gestión	4135	sector	58658
proyecto	3523	social	54065
indicador	3507	servicio	53014
producto	3482	público	50592
público	3369	gestión	49166
estrategia	3328	población	46563
población	3192	línea	43448
nivel	3053	indicador	42470
recurso	2923	proyecto	41725
sistema	2856	rural	41107
línea	2816	concejo	38139
sector	2801	producto	36454

Fuente: elaboración propia

Tabla 16: Palabras relacionadas más frecuentes para la categoría Normatividad-términos que implican articulación subnacional

Departamentos		Municipios	
Palabra	Frecuencia	Palabra	Frecuencia
programa	642	ley	2044
departamento	614	por	960
nacional	560	plan	252
disposición	528	desarrollo	242
gestión	497	artículo	234
rural	467	orgánico	215
proyecto	334	dictar	204



mujer	325	municipio	188
documento	321	territorial	182
planeación	254	decreto	173
seguridad	219	área	158
tasa	205	norma	151
productivo	203	región	139
persona	201	régimen	135
estrategia	199	organización	127
infancia	192	funcionamiento	113
cultura	182	recurso	112
riesgo	170	metropolitano	110
sistema	161	ambiental	109
garantizar	150	distrito	106

Fuente: elaboración propia

Tabla 17: Palabras relacionadas más frecuentes para la categoría Ordenamiento territorial-términos que implican articulación subnacional

Departamentos		Municipios	
Palabra	Frecuencia	Palabra	Frecuencia
plan	7507	desarrollo	8890
municipio	4066	plan	8299
desarrollo	3996	por	7623
social	3506	municipio	6605
acción	3144	municipal	4114
programa	3057	territorial	3751
departamento	2797	programa	3603
población	2724	ambiental	3213
proyecto	2551	gestión	3134
público	2544	pacto	3083
línea	2462	nacional	2814
departamental	2416	servicio	2745
económico	2265	infraestructura	2590
asamblea	2229	vial	2560
información	2102	ordenamiento	2450
nacional	2091	sector	2376
territorio	1964	público	2201
sector	1780	estratégico	2163
gestión	1625	proyecto	2109



educación	1528	vía	2049
-----------	------	-----	------

Fuente: elaboración propia

Tabla 18: Palabras relacionadas más frecuentes para la categoría Recursos-términos que implican articulación subnacional

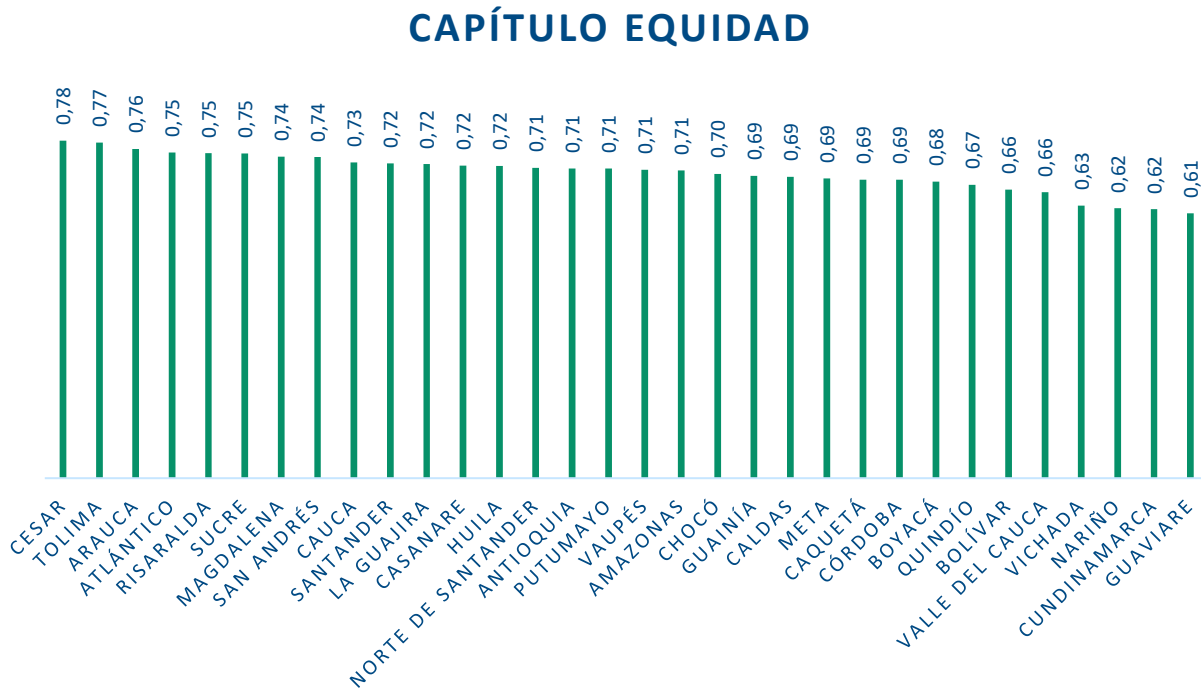
Departamentos		Municipios	
Palabra	Frecuencia	Palabra	Frecuencia
cid	18511	desarrollo	32173
desarrollo	6170	plan	30301
plan	5407	por	26442
departamental	4405	municipio	21557
departamento	4279	público	18218
sector	4224	inversión	17684
población	3660	municipal	17275
proyecto	3475	recurso	15688
nacional	3449	programa	15416
indicador	3139	nacional	15132
público	3000	gestión	10982
municipio	2988	sector	10158
programa	2736	proyecto	9213
sistema	2444	servicio	9177
gestión	2417	social	9140
gobierno	2405	territorial	8135
producto	2390	general	7534
territorio	2136	gobierno	7151
social	1861	sistema	7046
infraestructura	1794	sgp	6844

Fuente: elaboración propia



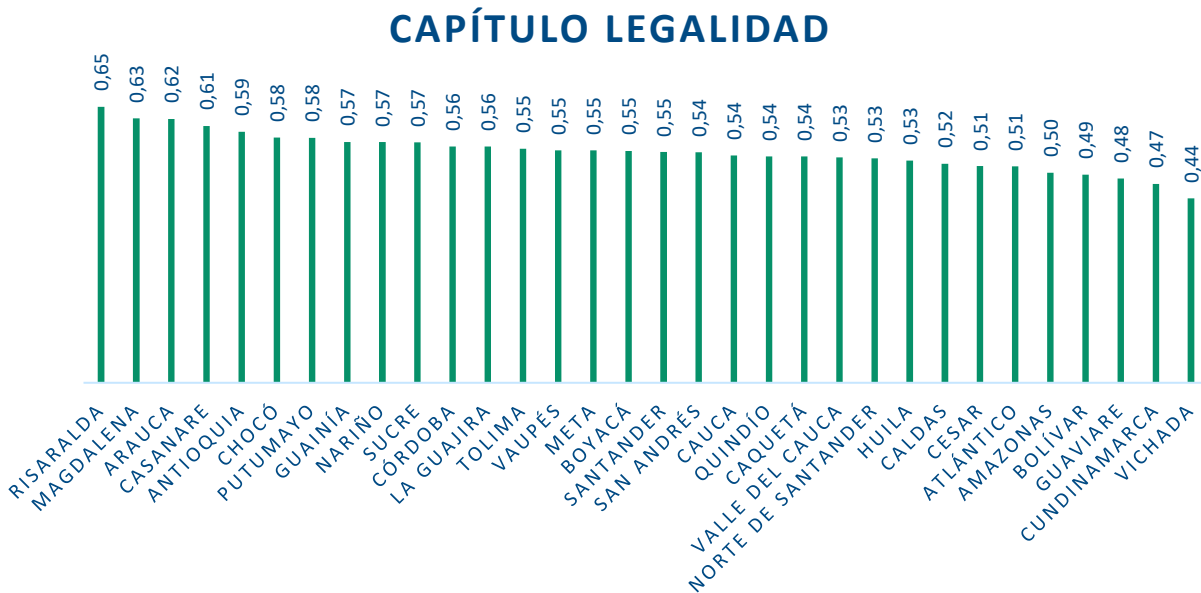
Anexo 3

Figura 36: similitud entre PDT departamentales y el capítulo equidad del PND



Fuente: elaboración propia

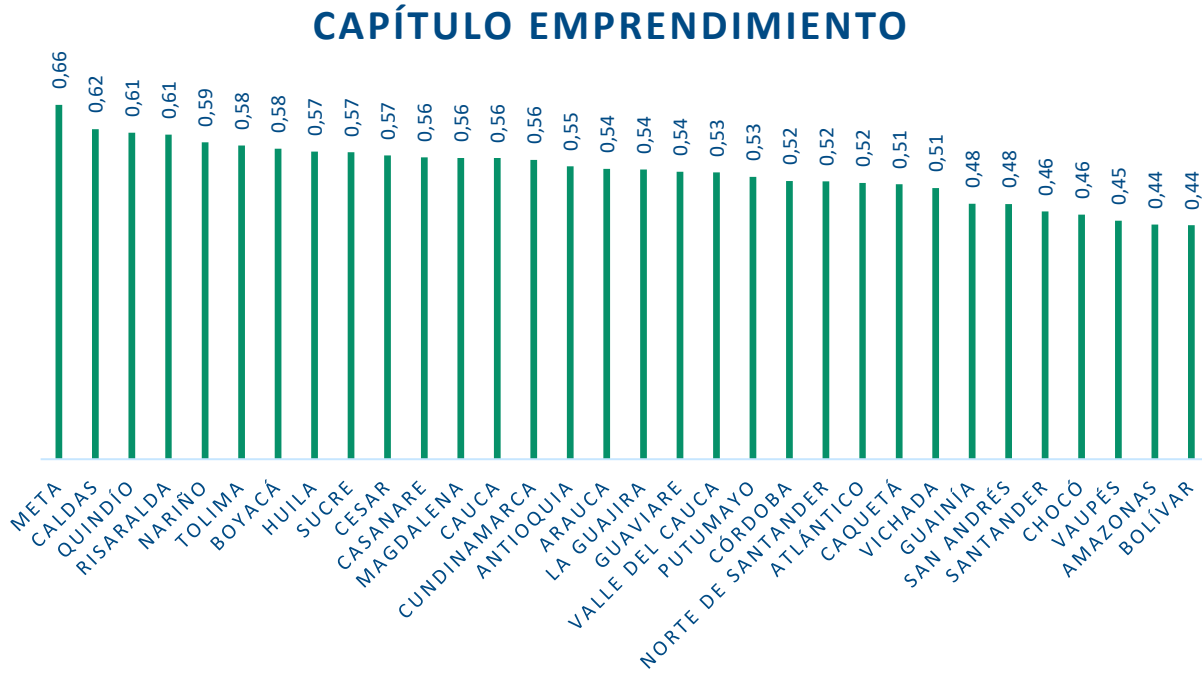
Figura 37: similitud entre PDT departamentales y el capítulo legalidad del PND



Fuente: elaboración propia

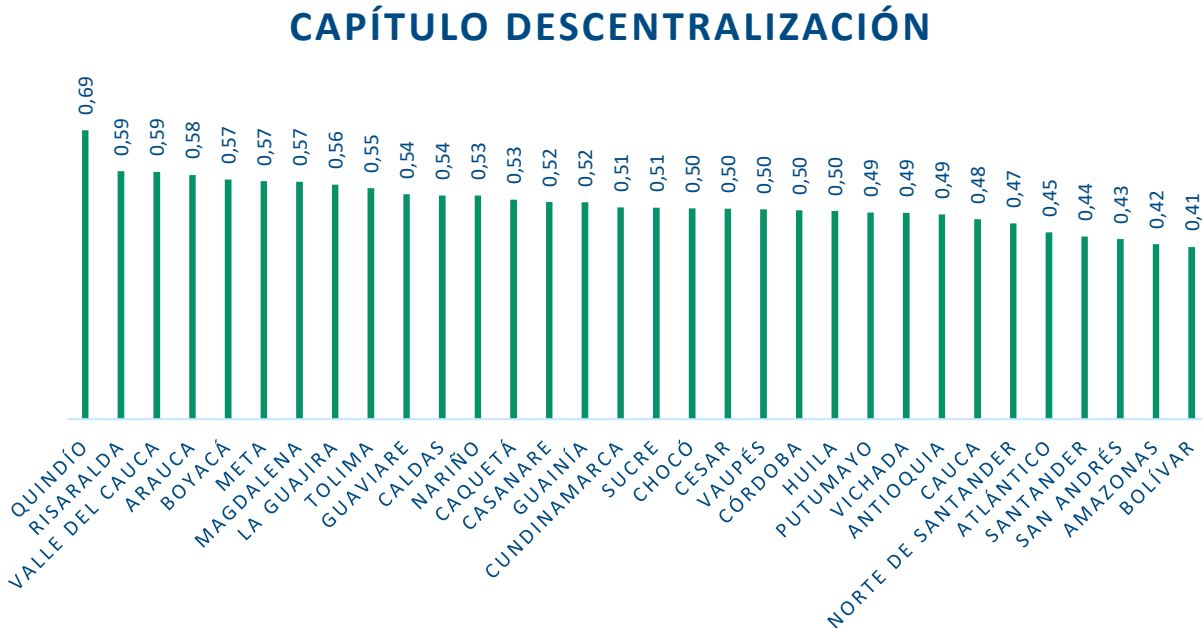


Figura 38: similitud entre PDT departamentales y el capítulo emprendimiento del PND



Fuente: elaboración propia

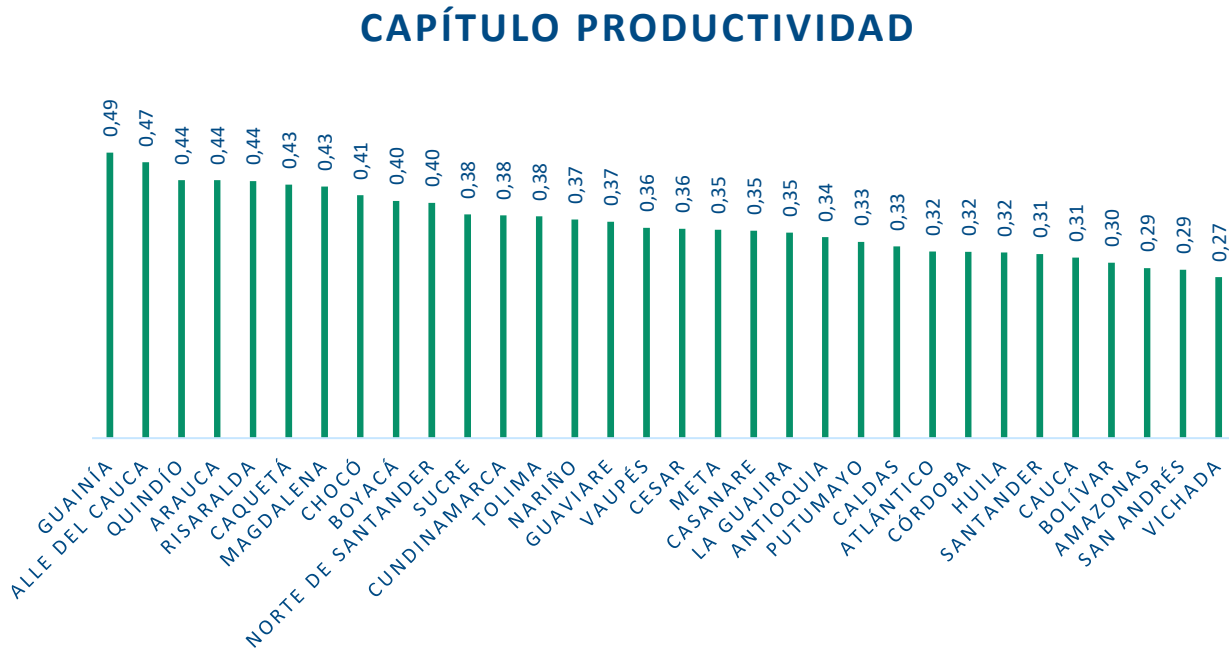
Figura 39: similitud entre PDT departamentales y el capítulo descentralización del PND



Fuente: elaboración propia

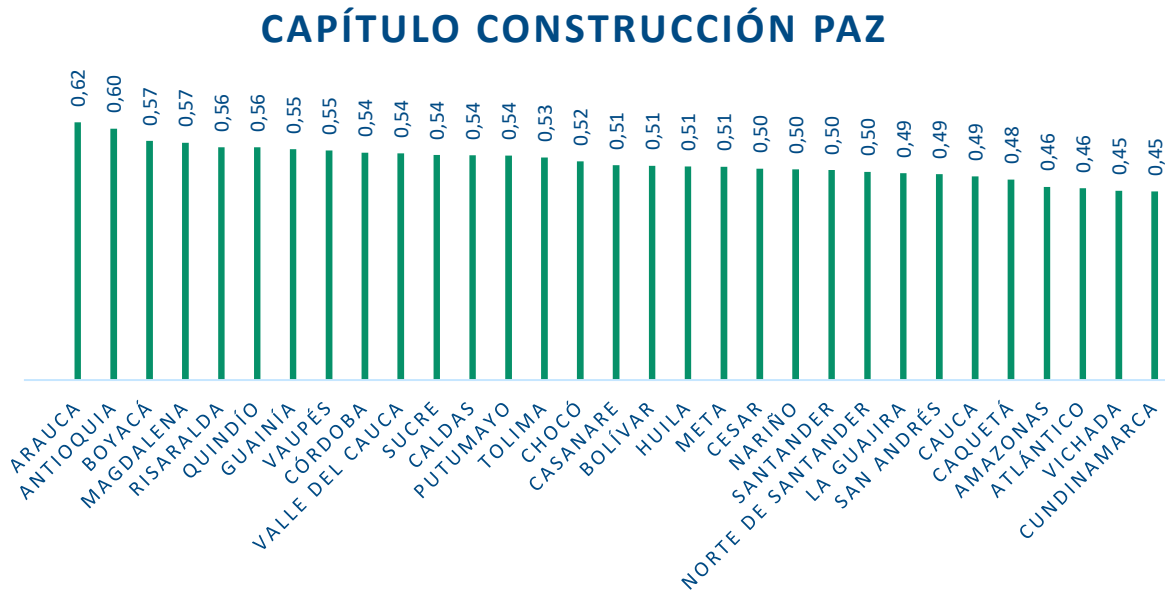


Figura 40: similitud entre PDT departamentales y el capítulo productividad del PND



Fuente: elaboración propia

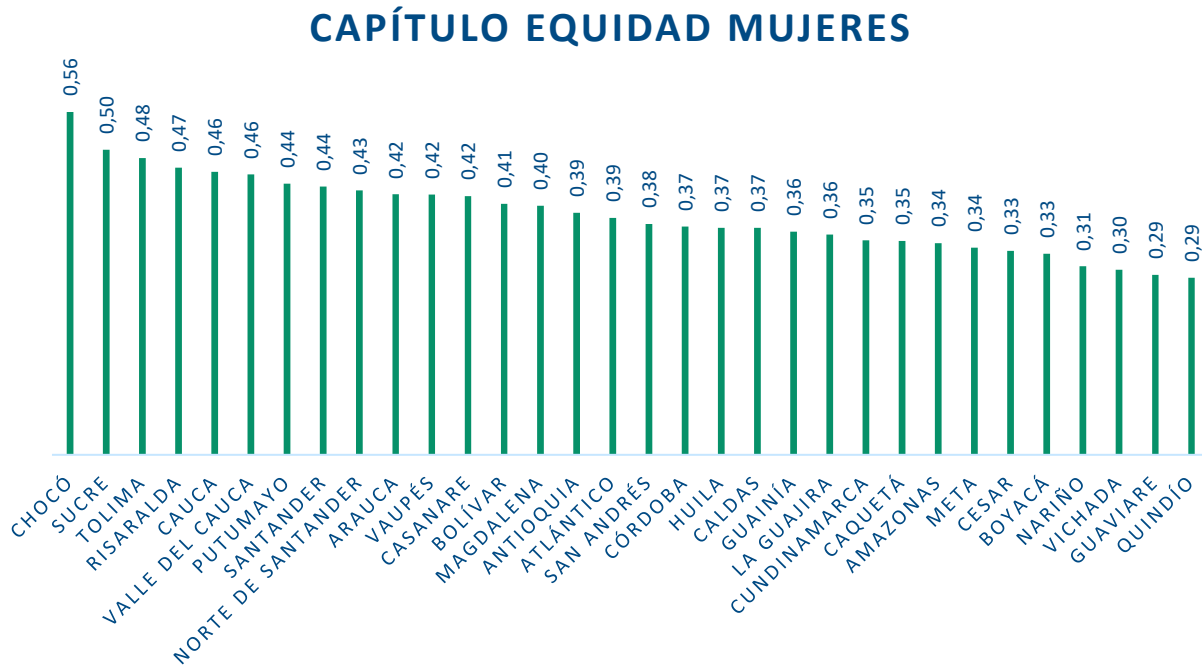
Figura 41: similitud entre PDT departamentales y el capítulo construcción paz del PND



Fuente: elaboración propia



Figura 42: similitud entre PDT departamentales y el capítulo equidad mujeres del PND



Fuente: elaboración propia

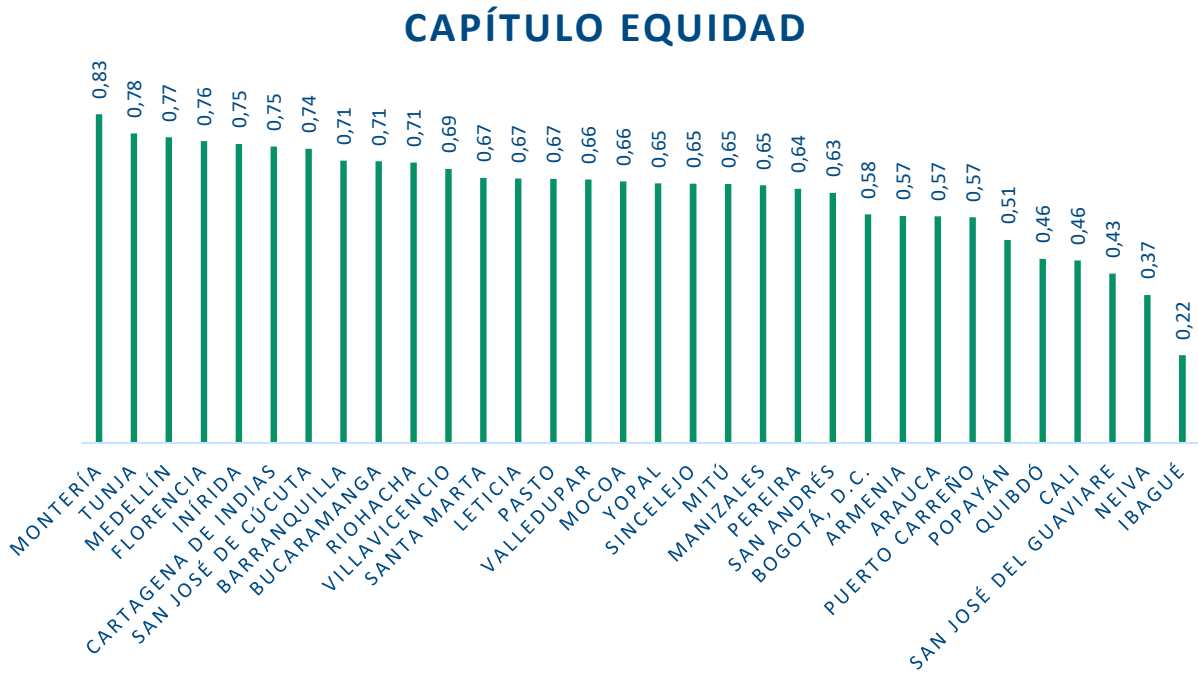
Figura 43: similitud entre PDT departamentales y el capítulo equidad grupos del PND



Fuente: elaboración propia

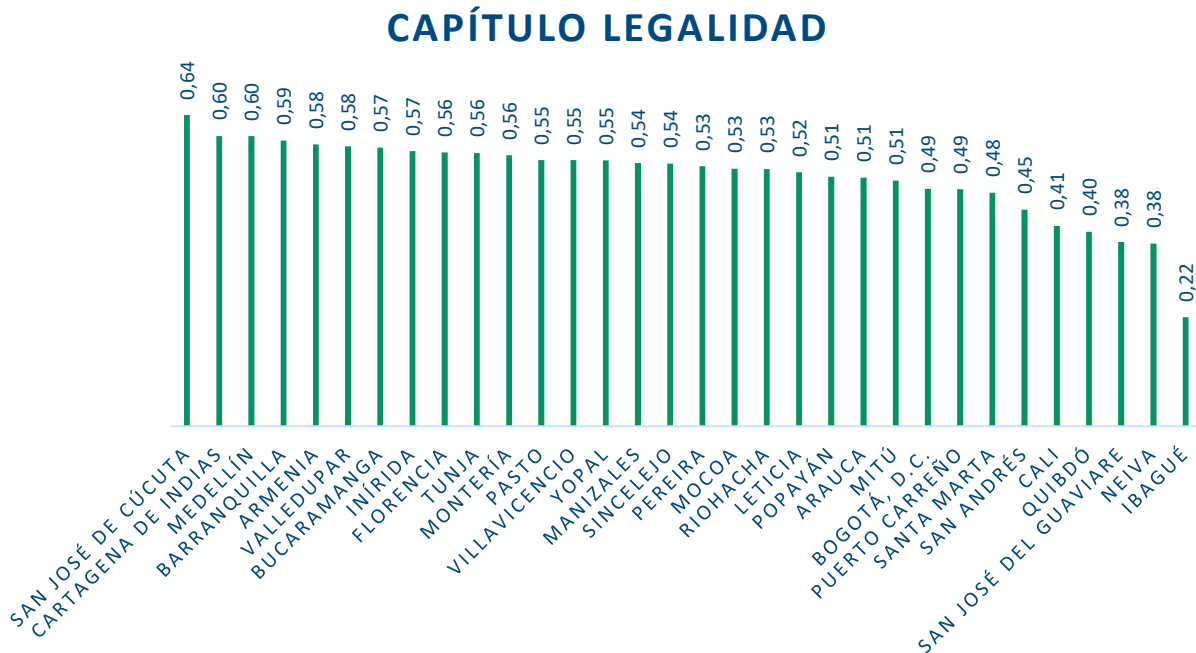


Figura 44: similitud entre los PDT municipales y el capítulo equidad del PND



Fuente: elaboración propia

Figura 45: similitud entre los PDT municipales y el capítulo legalidad del PND

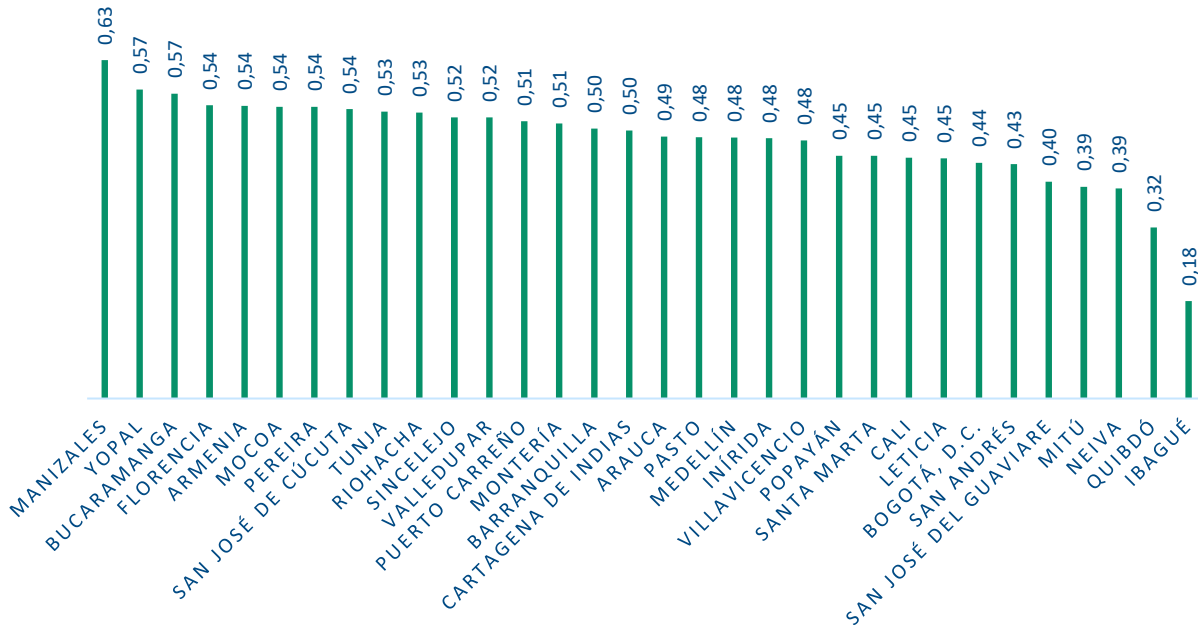


Fuente: elaboración propia



Figura 46: similitud entre los PDT municipales y el capítulo emprendimiento del PND

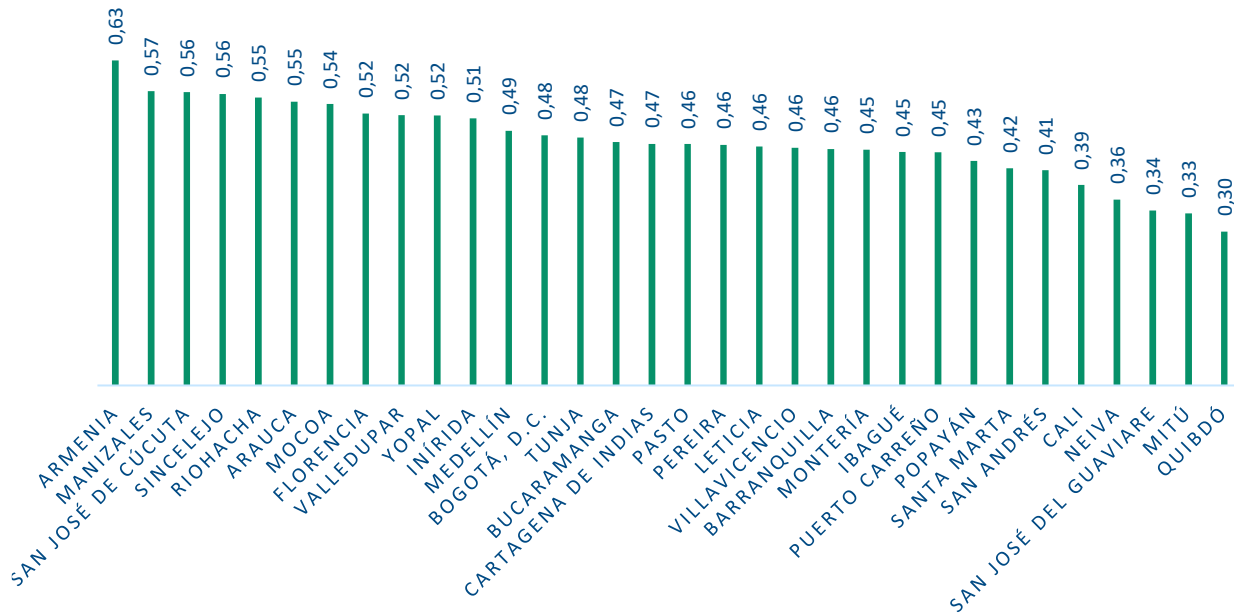
CAPÍTULO EMPRENDIMIENTO



Fuente: elaboración propia

Figura 47: similitud entre los PDT municipales y el capítulo descentralización del PND

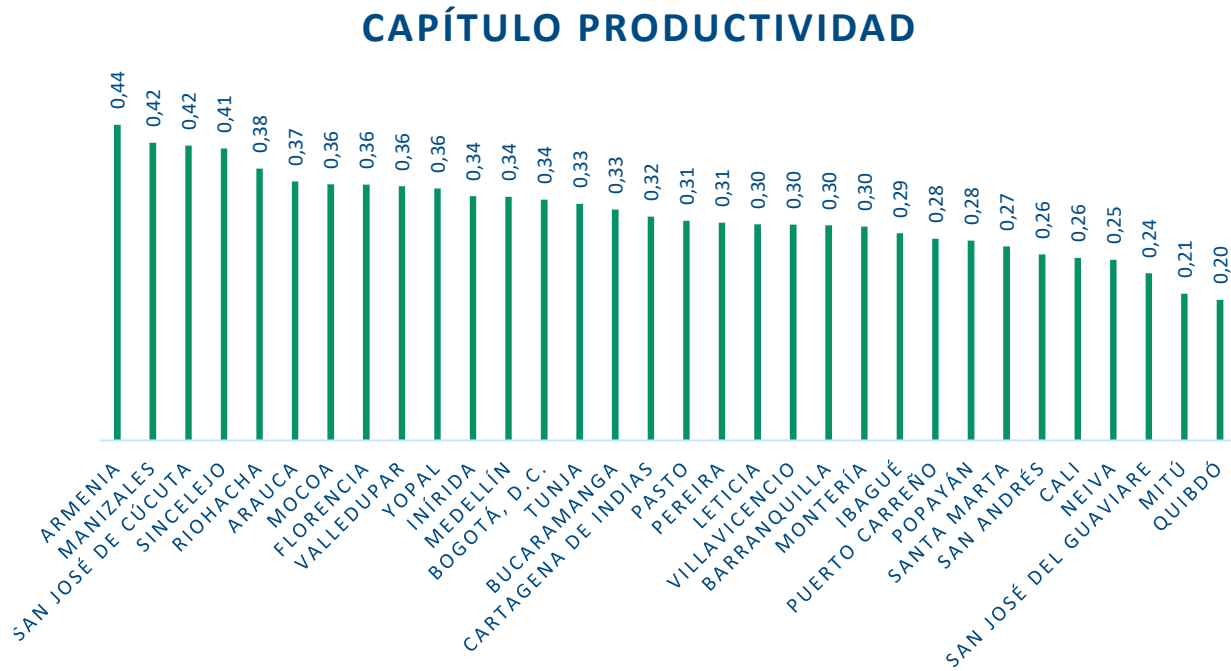
CAPÍTULO DESCENTRALIZACIÓN



Fuente: elaboración propia

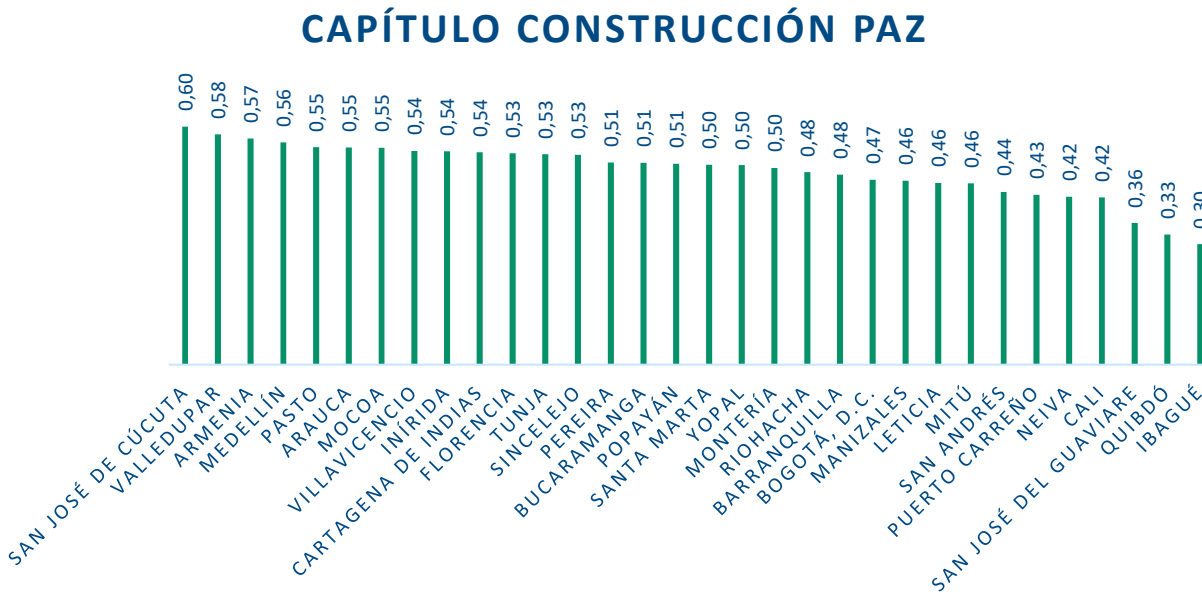


Figura 48: similitud entre los PDT municipales y el capítulo productividad del PND



Fuente: elaboración propia

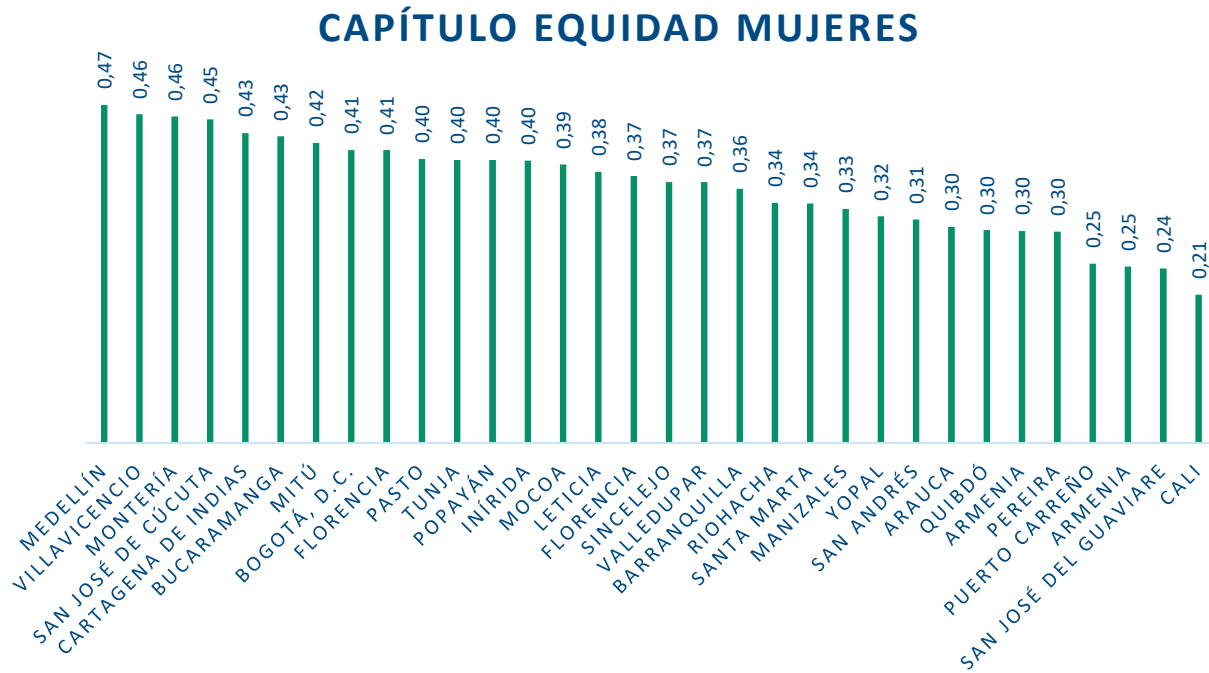
Figura 49: similitud entre los PDT municipales y el capítulo construcción paz del PND



Fuente: elaboración propia

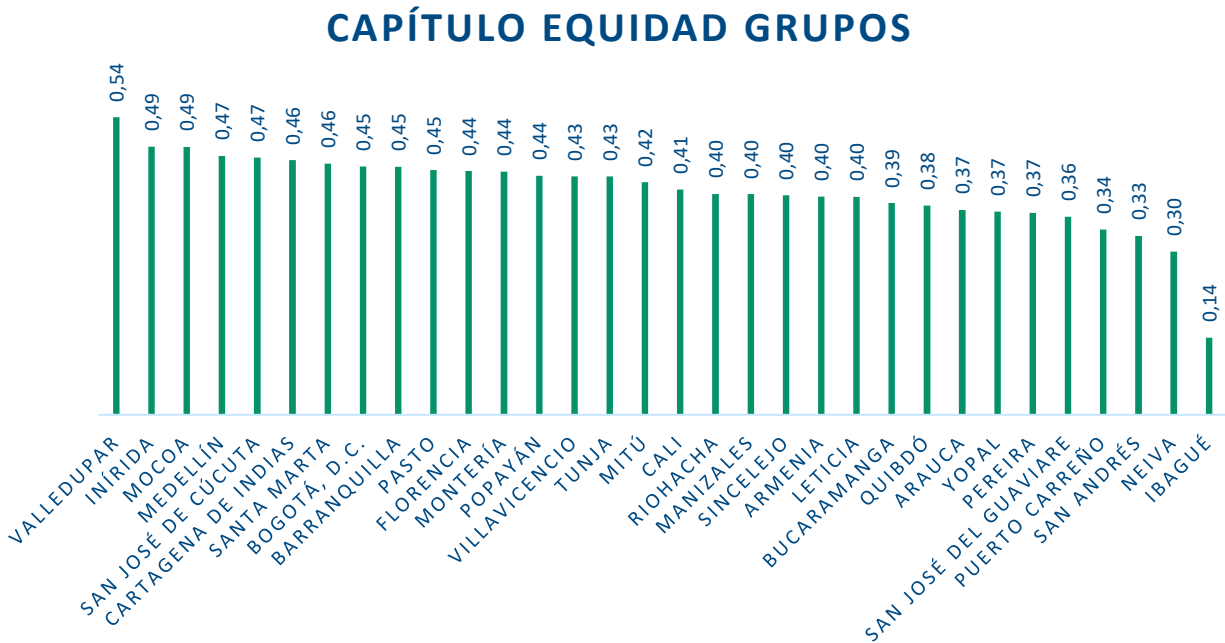


Figura 50: similitud entre los PDT municipales y el capítulo equidad mujeres del PND



Fuente: elaboración propia

Figura 51: similitud entre los PDT municipales y el capítulo equidad grupos del PND



Fuente: elaboración propia



Anexo 4

Tabla 19: top 8 palabras exclusivas PDT departamentales

Antioquia	subregión	poblacional	covid	seccional	corresponsable	anualizado	resiliencia	organizacional
Atlántico	estampilla	subregión	indeporte	matrícula	sectorial	tributación	covid	hídrico
Bolívar	bolivarense	raizal	palenquero	covid	poblacional	ordenanza	lineamiento	organizacional
Boyacá	provincia	sectorial	indeporte	covid	ordenanza	hídrico	poblacional	prestador
Caldas	cofinanciación	caldense	ordenanza	poblacional	subregión	cafetero	lineamiento	covid
Caquetá	presupuestal	caguán	pdet	moción	lineamiento	caqueteño	prescolar	fragua
Cauca	subregión	presupuestal	poblacional	lineamiento	precipitación	caucano	rabia	emcaservicio
Cesar	sectorial	simat	trazador	cesarense	idecesar	gei	fecundidad	kilómetro
Córdoba	cordobés	sinú	presupuestal	cerveza	regalía	subregión	cigarrillo	sectorial
Cundinamarca	cundinamarqués	provincia	consumido	referenciar	saneado	hídrico	provincial	covid
Chocó	subregión	chocoano	hídrico	morbilidad	botadero	priorizado	concertación	baudó
Huila	lineamiento	huilense	poblacional	palenquero	cigarrillo	raizal	equivaler	sectorial
La Guajira	presupuestal	eólico	guajira	wayuu	sectorial	poblacional	lineamiento	regalía
Magdalena	subregión	covid	macondo	magdalenense	poblacional	hídrico	cerro	pdet
Meta	ordenanza	gobernanza	programación	lineamiento	metense	prestador	sectorial	hídrico
Nariño	presupuestal	covid	nariñense	hídrico	pdet	lineamiento	poblacional	priorizado
Norte De Santander	subregión	migrante	matrícula	poblacional	pdet	adecuación	migratorio	suroriental
Quindío	quindiano	sectorial	covid	cafetero	lineamiento	ordenanza	clúster	estampilla
Risaralda	risaraldense	poblacional	lineamiento	migrante	sectorial	cafetero	hídrico	prestador
Santander	provincia	santandereano	lineamiento	poblacional	prestador	prevalencia	sectorial	ordenanza
Sucre	covid	subregión	migrante	morrosquillo	cuatrenio	playa	sucreño	reincorporado
Tolima	ordenanza	hídrico	tolimense	poblacional	covid	sectorial	evidenciar	estampilla
Valle Del Cauca	vallecaucano	subregión	palenquero	raizal	hídrico	poblacional	concertación	covid
Arauca	situacional	sectorial	covid	presupuestal	araucano	priorizado	evidenciar	entierro
Casanare	provivir	ordenanza	lineamiento	covid	propiciar	sectorial	kilómetro	mitigación
Putumayo	epidemiológico	notificar	evidenciar	hospitalización	quebrado	perinatal	morbilidad	poblacional
San Andrés	raizal	sai	poblacional	normativo	insular	ordenanza	prestador	coralino
Amazonas	covid	municipalizado	cofinanciación	subsidiado	amazónico	paraná	reemergente	cabildo
Guainía	lineamiento	covid	poblacional	ordenanza	amazonía	raizal	palenquero	mitigación
Guaviare	ordenanza	lineamiento	guaviarense	amazónico	concertación	estampilla	hídrico	letalidad
Vaupés	poblacional	municipalizado	sectorial	lineamiento	indígena	morbilidad	caracterizar	papunahuo
Vichada	planeacion	insumo	cigarrillo	poblacional	estampilla	sabana	caracterizar	fecundidad

Fuente: elaboración propia



Tabla 20: top 8 palabras exclusivas PDT municipales

Agrado	presupuestal	kilómetro	honorable	predial	lineamiento	petrolífero	destinación	facultado
Algeciras	pdet	inexistencia	vacuna	casual	dosis	hídrico	paraíso	insumo
Bahía Solano	quebrado	playa	honorable	árbol	baudó	aseguramiento	húmedo	hídrico
Barranquilla	covid	priorizado	localidad	secretar	migrante	poblacional	hídrico	palenquero
Chía	covid	aplazado	honorable	expedir	coronavirus	poblacional	indiferencia	vertimiento
Chinú	migrante	reincorporado	cualitativo	matrícula	cordobés	covid	palenquero	insumo
Cravo Norte	pdet	migrante	árbol	poblacional	adecuación	numérico	ramal	hídrico
El Cerrito	cafetero	hídrico	poblacional	sectorial	serranía	predial	vulneración	propiciar
Guacarí	presupuestal	poblacional	honorable	lineamiento	mantenido	resiliente	interrelación	habilitante
La Belleza	belleza	presupuestal	rubi	adecuación	bovino	simat	fedegan	playa
La Playa	playa	destinación	esparcimiento	adecuación	subsidiado	mipg	kilómetro	prescolar
La Unión	hídrico	priorizado	capacitar	poblador	honorable	estampilla	lineamiento	prestador
Lenguazaque	hídrico	cualitativo	enel	respectivo	adecuación	presupuestal	eot	predial
Manta	priorizado	presupuestal	enel	adecuación	respectivo	kilómetro	pertinencia	vertimiento
Moñitos	adecuación	subsidiado	comisaria	playa	poblacional	pecuario	resiliente	pertinencia
Olaya	presupuestado	morbilidad	quebrado	poblacional	playa	guayabo	olayense	maligno
Páez	poblacional	hídrico	lúdico	prescolar	mochilero	adecuación	quebrado	subsidiado
Puerto Guzmán	hospitalario	optimizado						
Recetor	fotografía	quebrado	hídrico	prestador	lámpara	poblacional	tecnificado	vertimiento
Riofrío	honorable	adecuación	predial	prevalencia	presupuestal	fecundidad	hídrico	viento
San Antonio Del Tequendama	adecuación	mitigación	covid	enel	lineamiento	vulnerar	poblacional	propender
San Francisco	cofinanciación	formalizado	sectorial	típico	festival	quebrado	poblacional	reactivado
Támara	miramar	priorizado	cementerio	mantenido	resiliente	evidenciar	festival	pecuario
Tame	priorizado	hídrico	inexistencia	trazador	presupuestal	migrante	pedemonte	destinación
Tibú	pdet	priorizado	subsidiado	aseguramiento	hídrico	núcleo	lineamiento	migrante
Tiquisio	lineamiento	quebrado	concertación	covid	honorable	poblacional	mitigación	sabana
Villa Caro	cociente	adecuación	liberal	quebrado	ramal	subsidiado	ordenanza	sisben
Villagarzón	villagarzón	árbol	núcleo	pdet	cabildo	guineo	amazónico	priorizado
Zambrano	poet	pdet	covid	subregión	coronavirus	sectorial	gestante	adecuación

Fuente: elaboración propia



Anexo 5

- 05042
- 05045
- 05129
- 05145
- 05237
- 05368
- 05628
- 05647
- 05658
- 05660
- 05664
- 05679
- 05792
- 05837
- 05895
- 08137
- 08685
- 13580
- 15104
- 15232
- 15696
- 15757
- 23350
- 25281
- 25324
- 25436
- 25489
- 25580
- 25645
- 25743
- 25781
- 25867
- 27800
- 47703
- 50689
- 52240
- 68179
- 68575
- 68720
- 70678
- 70717
- 73024
- 73152
- 73236
- 73275
- 73283
- 73504
- 76036
- 76041
- 76111
- 76126
- 76243
- 76400
- 76616
- 81220
- 85125
- 85400
- 86755
- 88001
- 91263
- 91405
- 91407
- 91430
- 91460
- 91530
- 91536
- 91669
- 91798
- 94343
- 94883
- 94884
- 94885
- 94886
- 94887
- 94888
- 97511
- 97777
- 97889