

# Dirección de Desarrollo Digital

Unidad de Científicos  
de Datos



**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación



## Análisis de los PDT de municipios y departamentos costeros y su similitud con dimensiones del diseño del índice de desarrollo

### INFORME FINAL

<b>Dependencias y entidades involucradas</b>	Departamento Nacional de Planeación <ul style="list-style-type: none"><li>• Dirección de Desarrollo Digital - Unidad de Científicos de Datos</li><li>• Dirección de Descentralización y Desarrollo Regional</li></ul>
<b>Sector</b>	Planeación
<b>Tecnologías utilizadas</b>	Python, Jinja
<b>Fuentes de datos</b>	Documentos PDT de departamentos costeros Documentos PDT de municipios costeros

## Contenido

1. Presentación .....	2
2. Objetivos del proyecto .....	2
3. Metodología .....	2
4. Resultados .....	9
5. Conclusiones y recomendaciones .....	10
6. Socialización .....	10
7. Contacto .....	11
ANEXOS .....	11
Anexo 1 Expresiones regulares utilizadas en la estrategia de búsqueda de términos .....	11



## 1. Presentación

Los Planes de Desarrollo Territorial (PDT) constituyen el instrumento de política más importante para orientar las acciones y apuestas de los entes territoriales en un período de gobierno. El contenido de estos documentos se convierte en la carta de navegación, y por ello resulta fundamental incluir en este instrumento lineamientos específicos que permitan adelantar las decisiones y acciones en diferentes pilares temáticos, como son planeación, economía, seguridad, ambiente y otros. Por lo anterior, es vital realizar análisis completos de todo lo que se consigna en los PDT con el objetivo de identificar cuáles son las estrategias y falencias territoriales para la formulación de políticas públicas relacionadas con estos pilares.

Uno de los grandes obstáculos para realizar estos análisis es la cantidad de información textual presente en estos documentos, en el marco de los departamentos y municipios costeros se tienen 56 documentos PDT en total. Para lo cual analizar estos documentos de forma manual es una tarea tediosa que implica un gasto alto en recurso humano y tiempo. Dado lo anterior y en aras de facilitar esta tarea, la Unidad de Científicos de Datos desarrolló una metodología basada en técnicas de análisis y minería de texto para buscar términos de interés con el fin de dar un indicador para medir que tan alineados están los PDT departamentales y municipales costeros con las diferentes dimensiones que componen el índice de desarrollo.

*The Territorial Development Plans (PDT) are the most important policy instrument to guide the actions and challenges of the territorial entities in a government period. The content of these documents becomes the navigation chart, and therefore it is essential to include in this instrument, specific guidelines that allow the advancement of decisions and actions in different thematic pillars, such as planning, economy, security, environment and others. For this reason, it is vital to carry out a complete analysis of everything included in the PDTs in order to identify the strategies and territorial shortcomings for the formulation of public policies related to these pillars.*

*One of the major obstacles to carry out these analyses is the amount of textual information present in these documents; within the framework of the coastal departments and municipalities, there are 56 PDT documents in total. Analyzing these documents manually is a tedious task that implies a high expenditure of human resources and time. Given the above and to facilitate this task, the Data Scientists Unit developed a methodology based on analysis and text mining techniques to search for terms of interest in order to provide an indicator to measure how aligned the departmental and municipal coastal PDTs are with the different dimensions that make up the development index.*

## 2. Objetivos del proyecto

### 2.1. General

Analizar los Planes de Desarrollo Territoriales (PDT) de 44 Municipios Costeros (MC) y 12 Planes de Desarrollo Departamentales (PDD) del periodo 2020-2023, con el propósito de cuantificar la presencia de diversos términos de interés clasificados en 7 grupos temáticos (o dimensiones) y definir que tan alineados se encuentran.

### 2.2. Específicos

1. Realizar un análisis comparativo de cada documento de plan de desarrollo (PDT o PDD) con cada una de las dimensiones y sus respectivas palabras claves asociadas resultantes del diseño del Índice de Desarrollo de los MC, para identificar el nivel de similitud entre estos.
2. Realizar la implementación de gráficos descriptivos de los textos de interés.
3. Generar un reporte web consolidando los resultados obtenidos del proyecto.

## 3. Metodología

### 3.1. Lectura de los documentos PDT

Para el desarrollo de este proyecto se trabajó con un total de 12 Planes de Desarrollo Departamentales y 44 Planes de Desarrollo Municipales en formato PDF. En las Tabla 1 y Tabla 2 se presenta el listado de documentos utilizados.



Tabla 1. Listado de PDD

Código DIVIPOLA	Departamento
05	Antioquia
08	Atlántico
13	Bolívar
19	Cauca
23	Córdoba
27	Chocó
44	La Guajira
47	Magdalena
52	Nariño
70	Sucre
76	Valle del Cauca
88	San Andrés, Providencia Y Santa Catalina Islas

Fuente: Elaboración propia

Tabla 2. Listado de PDM

Código DIVIPOLA	Departamento	Municipio
05051	Antioquia	Arboletes
05490	Antioquia	Necoclí
05665	Antioquia	San Pedro de Urabá
05837	Antioquia	Turbo
08001	Atlántico	Barranquilla
08372	Atlántico	Juan de Acosta
08549	Atlántico	Piojó
08573	Atlántico	Puerto Colombia
08832	Atlántico	Tubará
13001	Bolívar	Cartagena
13673	Bolívar	Santa Catalina
19318	Cauca	Guapi
19418	Cauca	López de Micay
19809	Cauca	Timbiquí
23419	Córdoba	Los Córdoba
23500	Córdoba	Moñitos
23574	Córdoba	Puerto Escondido
23672	Córdoba	San Antero
23675	Córdoba	San Bernardo del Viento
27006	Chocó	Acandí
27075	Chocó	Bahía Solano
27077	Chocó	Bajo Baudó
27250	Chocó	Litoral del San Juan
27372	Chocó	Juradó
27495	Chocó	Nuquí
44001	La Guajira	Riohacha
44090	La Guajira	Dibulla
44560	La Guajira	Manaure
44847	La Guajira	Uribia



Código DIVIPOLA	Departamento	Municipio
47001	Magdalena	Santa Marta
47189	Magdalena	Ciénaga
47570	Magdalena	Puebloviejo
47745	Magdalena	Sitionuevo
52250	Nariño	El Charco
52390	Nariño	La Tola
52473	Nariño	Mosquera
52490	Nariño	Olaya Herrera
52520	Nariño	Francisco Pizarro
52696	Nariño	Santa Bárbara de Iscuande
52835	Nariño	Tumaco
70221	Sucre	Coveñas
70713	Sucre	San Onofre
70820	Sucre	Tolú
76109	Valle del Cauca	Buenaventura

Fuente: Elaboración propia

Dado que los documentos se encuentran en formato PDF, el primer paso corresponde a la lectura de los documentos y realizar la transformación de estos a texto plano, para así poder tener la información en un formato que permita realizar los procesos de identificación de términos de interés. La calidad del texto resultante de la lectura y extracción de información de los PDF dependerá en gran medida de la calidad de estos. Para proceder con la lectura de los documentos se utilizó la función *leer\_texto* de la librería *ConTexto*, la cual fue implementada por la UCD y facilita diversas tareas de procesamiento de texto. La función *leer\_texto* permite leer y extraer el texto de documentos en diferentes formatos y en caso de requerirse utiliza un OCR o Reconocimiento Óptico de Caracteres para leer información de imágenes.

### 3.2. Limpieza de texto

Una vez se finaliza la etapa de lectura de los documentos se debe realizar una limpieza de estos, con el propósito de estandarizar los textos y remover información que pueda generar errores en el análisis. En el desarrollo de este proyecto la etapa de limpieza consistió principalmente en 3 pasos:

1. Se pasa el texto a minúscula. Lo anterior se debe a que el texto en los computadores tiene una representación numérica para cada carácter (o letra) en una palabra, por lo que la palabra "Planeación" no es equivalente a "planeación", al pasar el texto a minúscula estandarizamos el texto y reducimos los posibles errores al momento de analizar los documentos.
2. Se eliminan signos de puntuación como comas, puntos, punto y coma; y adicionalmente se remueven números y espacios múltiples. Lo anterior se realiza nuevamente con el objetivo de estandarizar el texto y remover caracteres que no aportan valor al momento de analizar los documentos.
3. Por último, se removieron "palabras repetidas". Estas hacen referencia a frases que se repiten consecutivamente dentro de un texto. Este paso se requiere principalmente por la manera en que el computador "lee" las tablas dentro de un documento, tomemos como ejemplo la Figura 1.

Figura 1. Recorte tabla de metas contenida en el PDT del departamento de Atlántico.

## Metas

Indicadores de resultado				
Responsable	Indicador	Línea Base	Meta	ODS
Secretaría de Salud	Tasa de mortalidad en menores de 5 años	16,26	13,76	3. Salud y Bienestar
Secretaría de Salud	Tasa de mortalidad por infecciones respiratorias agudas (IRA)	16,26	15	3. Salud y Bienestar
Secretaría de Salud	Tasa de mortalidad por enfermedades isquémicas del corazón	99,45	95,11	3. Salud y Bienestar
Secretaría de Salud	Tasa de mortalidad por tumor maligno de la próstata	19,65	18,56	3. Salud y Bienestar
Secretaría de Salud	Tasa de mortalidad por tumor maligno de la mama	19	18,53	3. Salud y Bienestar
Secretaría de Salud	Tasa de mortalidad por tumor maligno del cuello uterino	11,01	11	3. Salud y Bienestar
Secretaría de Salud	Tasa de mortalidad por enfermedades crónicas de vías respiratorias inferiores	18	17,99	3. Salud y Bienestar
Secretaría de Salud	Tasa de incidencia de violencia de género e intrafamiliar	135,79	135,78	3. Salud y Bienestar
Secretaría de Salud	Tasa de intento de suicidio	54,56	52,12	3. Salud y Bienestar
Secretaría de Salud	Tasa de mortalidad por desnutrición en menores de 5 años	8,37	8,36	3. Salud y Bienestar
Secretaría de Salud	Nacidos vivos con bajo peso al nacer	8,74	8,17	3. Salud y Bienestar

Fuente: PDT de Atlántico 2020 – 2023

La lectura de esta tabla el computador la realiza por columnas, para la columna “Responsable” se tiene como resultado la frase “Secretaría de Salud” repetida múltiples veces en un mismo texto (teniendo en cuenta que es el mismo texto para cada fila). Igualmente sucede con la columna ODS, en donde se obtiene la frase “3. Salud y Bienestar”. Dado lo anterior, se procede a utilizar una función que elimina las frases repetidas, para este ejemplo dejando una sola vez el texto “Secretaría de Salud” y “3. Salud y Bienestar” respectivamente.

### 3.3. Dimensiones de búsqueda y términos de interés en los PDT

Con el objetivo de medir qué tan alineados están los PDT con el índice de desarrollo, la Subdirección de Ordenamiento y Desarrollo Territorial (SODT) propuso una estrategia de búsqueda basada en 7 dimensiones de análisis. En donde cada dimensión está compuesta por diferentes términos de interés, en la Tabla 3 se listan las dimensiones de análisis y sus términos asociados.

Tabla 3. Listado de dimensiones y términos asociados

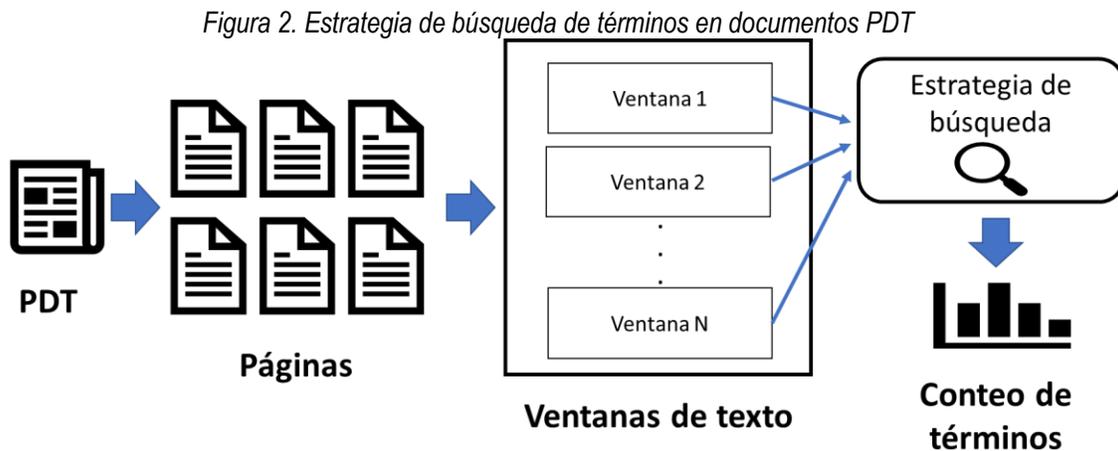
Dimensión	Términos asociados
1. Planeación y ordenamiento y, gestión para el desarrollo	Planeación, ordenamiento, gestión para el desarrollo, ordenamiento territorial, catastro multipropósito, suelo, naturales.
2. Conectividad, infraestructura y transporte	Conectividad, infraestructura, transporte, energía, alcantarillado, acueducto, internet, vivienda.
3. Capital humano	Capital humano, demografía, mortalidad, salud, educa, Pruebas Saber.
4. Actividades económicas	Economía, productividad, personas ocupadas, tasa de ocupación, desempleo, banca, financiación.
5. Gobernabilidad, desempeño y transparencia local	Gobernanza, transparencia, medición de desempeño, medición del desempeño, desempeño fiscal, desempeño institucional, regalías.
6. Seguridad y convivencia	Seguridad, convivencia, homicidio, hurto, delito, violencia, minas, desplazamiento, coca.
7. Ambiente	Medio ambiente, ecosistema, foresta, basura, residuo, área protegida, áreas protegidas.

Fuente: Elaboración propia

### 3.4. Estrategia de búsqueda de frases y términos sobre los PDT

En la Figura 2, se presenta el esquema general de búsqueda de términos en los documentos PDT. Este esquema se aplica a las siete dimensiones de búsqueda mencionadas en la sección anterior. Cada documento PDT después de la etapa de limpieza de texto, pasa por los siguientes pasos:

1. El documento es dividido en páginas para su análisis de términos.
2. Cada página del documento a su vez es dividida en fragmentos de texto más pequeños, denominados *ventanas* de análisis. Cada ventana de análisis está compuesta por 200 palabras (tamaño promedio de un párrafo), debido a que es difícil identificar correctamente párrafos en el texto. En algunos casos puede contener menos palabras debido al número de palabras presentes por página. Por ejemplo, una página con 453 palabras tendrá entonces 2 ventanas de 200 palabras y una de 53 palabras.
3. Cada ventana entra a la estrategia de búsqueda de términos relevantes 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:



Fuente: Elaboración propia

Debido a que los términos de interés pueden aparecer expresados de diversas formas (diferentes conjugaciones, en singular/plural u otras variaciones), la estrategia de búsqueda utiliza una metodología basada en expresiones regulares<sup>1</sup> para cada frase de las dimensiones de términos de interés. Por ejemplo, la frase “Gestión para el desarrollo” puede aparecer como “Gestión en el desarrollo”; “Gestión para garantizar el desarrollo” o “Gestión enfocada en el desarrollo”, entre muchas otras combinaciones. Tratar de armar todas las combinaciones posibles sobre el cómo podría aparecer la frase, puede dar como resultado cientos de combinaciones. Por tal motivo, las expresiones regulares ofrecen una alternativa para hacer esa búsqueda más lapsa pudiendo identificar múltiples formas de la frase. Para dar solución al problema anterior se utiliza la expresión regular “*gesti[oó]n|desarrollo*”.

En la Figura 3 se ilustra un ejemplo de uso de expresiones regulares sobre un texto de prueba; la expresión “*gesti[oó]n*” es capaz de encontrar las palabras *gestion* o *gestión*. La raya vertical o pleca (|) actúa como una condición lógica “o” y permite que se busquen las dos expresiones al mismo tiempo.

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

Figura 3: Ejemplo del funcionamiento de las expresiones regulares

Expresión regular: **gesti[oó]n** | **desarrollo**

El enfoque principal de esta política es asegurar la correcta **gestión** para el **desarrollo** del plan propuesto y lograr obtener las metas planteadas. Igualmente se debe contemplar la **gestión** que se venido realizando en periodos anteriores frente al **desarrollo** del PND y la **gestión** de recursos necesarios.

Fuente: Elaboración propia

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 anterior, hay 2 condiciones y la palabra gestión tiene 3 ocurrencias, mientras el término desarrollo tiene 2. Para encontrar el número de apariciones de la frase “Gestión para el desarrollo”, se halla el mínimo de ocurrencias entre todas las condiciones, en este caso 2, por lo que la estrategia de búsqueda da como resultado 2 para esa ventana de análisis.

Cada una de las frases y términos de interés de las 7 dimensiones de análisis tienen asociada una expresión regular para la estrategia de búsqueda (**¡Error! No se encuentra el origen de la referencia. Anexo 1 ¡Error! No se encuentra el origen de la referencia. ¡Error! No se encuentra el origen de la referencia.**). Este proceso se hace para cada término y para cada ventana dentro del PDT. Al final se consolidan los conteos de los términos para todo el documento.

### 3.5. Indicador para medir la alineación de los PDT con el índice de desarrollo

Para medir qué tan alineados están los documentos PDT con el índice de desarrollo se proponen dos puntajes, *promedio de términos* y un *promedio ponderado de términos*.

- **Puntaje de frecuencias “normal” de términos**

El promedio de términos es una sumatoria sobre el número de ocurrencias de los términos de interés en un PDT, normalizado por el número de páginas que contenga el documento. Este indicador es la forma natural de medir la aparición de términos de género en el documento. Matemáticamente, el indicador  $I_i$  sobre el  $i$ -ésimo documento PDT se define como:

$$I_i = \frac{\sum_v \sum_{j=1}^J t_v(j)}{P_i}$$

Donde el subíndice  $v$  representa la ventana de análisis sobre el texto,  $t(j)$  representa el número de ocurrencias de la dimensión  $j$  en la ventana  $v$  y  $P_i$  es el número de páginas con información del  $i$ -ésimo PDT.

- **Puntaje de frecuencias ponderado de términos**

Esta medida se plantea con el objetivo de conceder más importancia a las ventanas que contienen más de un término de la misma dimensión durante la búsqueda, bajo el supuesto que los documentos PDT que tengan capítulos dedicados a una temática asociada a la dimensión de análisis obtengan una mayor puntuación en comparación a aquellos que solo mencionan los términos de forma aislada. Matemáticamente, el indicador promedio ponderado  $IP_i$  sobre el  $i$ -ésimo documento PDT se define como:



$$IP = \frac{\sum_v \sum_{j=1}^J T_v d_v(j)}{P_i}$$

Donde el subíndice  $v$  representa la ventana de análisis sobre el texto,  $d(j)$  representa el número de ocurrencias de la dimensión  $j$  en la ventana  $v$ ,  $T_v$  representa el número de dimensiones diferentes que aparecen sobre la ventana  $v$  y  $P_i$  es el número de páginas con información del  $i$ -ésimo PDT.

Para dar un entendimiento más profundo de la diferencia entre estos dos indicadores se da el siguiente ejemplo:

Suponga que se tiene un documento PDT con 10 páginas, en el cual se analizan 3 ventanas, como se consigna en la Tabla 4

Tabla 4. Resultados para ejemplificar los indicadores

Páginas en el documento: 10		
Ventanas	Términos encontrados	Conteo
1	Planeación	1
	Gestión para el desarrollo	2
	Ordenamiento territorial	1
2	Catastro multipropósito	1
3	Planeación	3
	Catastro multipropósito	1

Fuente: Elaboración propia

El indicador de promedio de términos hace la sumatoria sobre todas las ventanas y términos (9 ocurrencias) y lo divide entre el total de páginas (10 páginas) para un puntaje de 0,9.

El indicador ponderado, primero hace la suma de ocurrencias sobre cada una de las ventanas (4 en la primera, 1 en la segunda y 4 en la tercera), después se fija cuantos términos aparecen en cada ventana (3 términos en la primera, 1 término en la segunda, 2 términos en la tercera). Por último, multiplica el número de términos de cada ventana por el número de ocurrencias de estos, esto es,  $3*4$  para la primera ventana,  $1*1$  para la segunda y  $2*4$  para la tercera. Se hace la suma de estos valores y se divide sobre el número de páginas. El resultado de este indicador será  $21/10 = 2,1$ .

En la Tabla 5 se presenta un ejemplo similar con el propósito de entender cómo cambian los valores de los indicadores.

Tabla 5. Resultados para ejemplificar los indicadores

Páginas en el documento: 10		
Ventanas	Términos encontrados	Conteo
1	Ordenamiento territorial	4
2	Catastro multipropósito	1
3	Planeación	3
	Catastro multipropósito	1

Fuente: Elaboración propia



El indicador de promedio de términos hace la sumatoria sobre todas las ventanas y términos (9 ocurrencias) y lo divide entre el total de páginas (10 páginas) para un puntaje de 0,9.

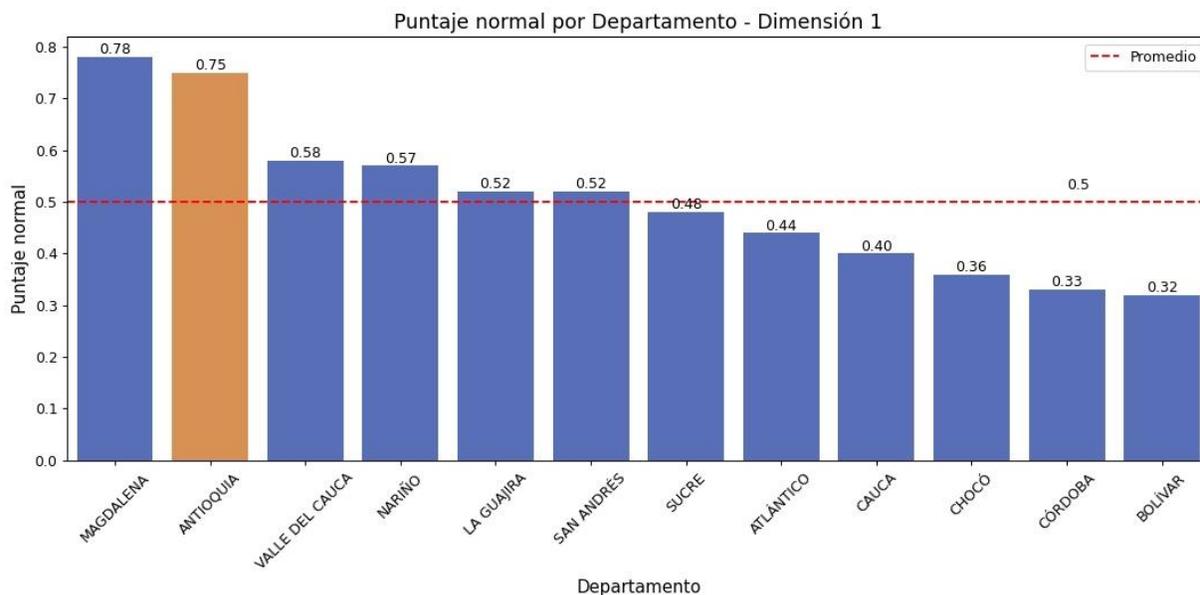
El indicador ponderado, primero hace la suma de ocurrencias sobre cada una de las ventanas (4 en la primera, 1 en la segunda y 4 en la tercera), después se fija cuantos términos aparecen en cada ventana (1 término en la primera, 1 término en la segunda, 2 términos en la tercera). Por último, multiplica el número de términos de cada ventana por el número de ocurrencias de estos, esto es,  $1*4$  para la primera ventana,  $1*1$  para la segunda y  $2*4$  para la tercera. Se hace la suma de estos valores y se divide sobre el número de páginas. El resultado de este indicador será  $13/10 = 1,3$ .

La interpretación de los indicadores propuestos se debe realizar en un contexto global, pues el valor por si solo no provee información relevante. Visto en conjunto, expresan de manera general cuántas frases o términos de interés fueron encontradas en los documentos analizados.

#### 4. Resultados

La metodología propuesta para la búsqueda de términos de interés se aplicó de forma automática a los 12 documentos PDT 2020 - 2023 de los entes departamentales costeros, al igual que a 44 documentos PDT de 2020 – 2023 de entes municipales costeros, para cada una de las 7 dimensiones de análisis. Como entregable del proyecto se entregaron los resultados en archivos de Microsoft Excel, al igual que se implementó un reporte web para cada una de las dimensiones de análisis. Estos corresponden a archivos HTML los cuales contienen gráficas descriptivas de los resultados obtenidos. La Figura 4 y la Figura 5 ilustran los resultados obtenidos para el puntaje de frecuencia de términos a nivel departamental y municipal de la Dimensión 1, destacando el departamento de Antioquia y el municipio de Arboletes respectivamente.

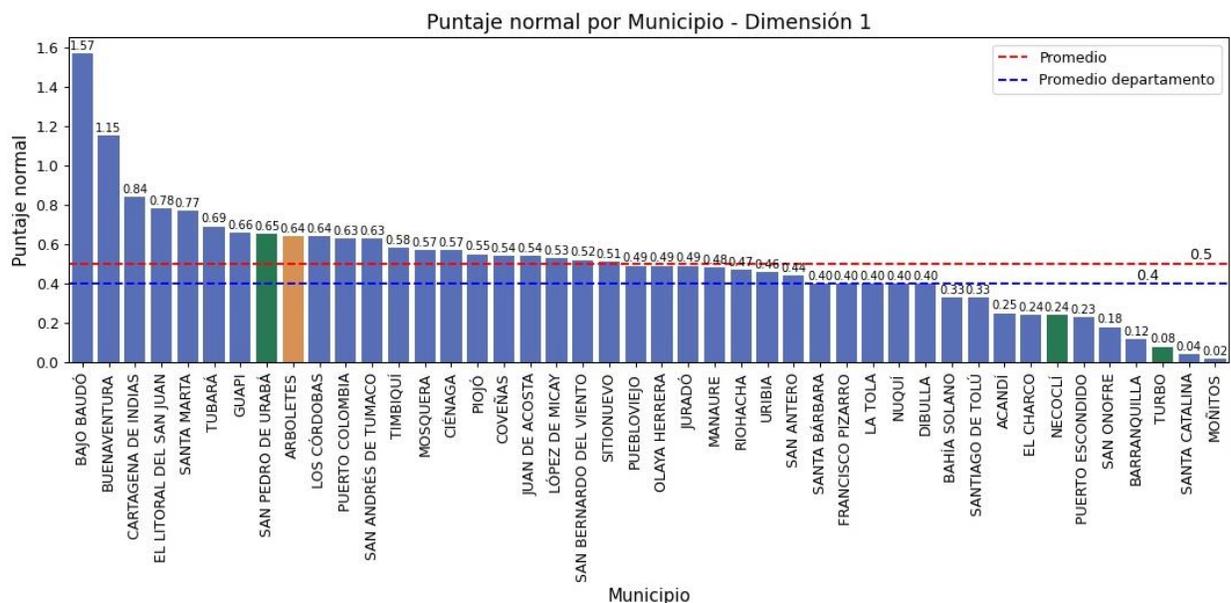
Figura 4. Puntaje de frecuencias normal por departamentos



Fuente: Reporte de resultados – dimensión 1



Figura 5. Puntaje de frecuencias normal por municipios



Fuente: Reporte de resultados – dimensión 1

Adicionalmente los reportes contienen gráficos de frecuencia de cada uno de los términos de interés, al igual que una tabla resumen de resultados y gráfico de dispersión léxica del documento analizado.

### 5. Conclusiones y recomendaciones

A partir de la metodología implementada y de los resultados obtenidos para cumplir los objetivos de este proyecto, 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 un mejor uso y aprovechamiento del proyecto.

1. Se desarrolló una metodología de búsqueda de términos de interés basado en expresiones regulares, lo que permite encontrar múltiples formas de una frase en los textos.
2. Se entregaron dos hojas de cálculo con los resultados consolidados para departamentos y municipios, donde se puede consultar los indicadores promedio y ponderado, y el conteo de frases encontradas en cada documento PDT.
3. Teniendo en cuenta que este fue un análisis automatizado sobre los textos de los PDT, se aconseja que los resultados se analicen desde la visión de un experto temático.
4. Como trabajo futuro se recomienda trabajar en conjunto con los expertos temáticos para proponer un indicador más robusto y que brinde más información, debido a que los actuales, simplemente cuantifican la cantidad de frases presentes en los documentos PDT.
5. Para mayor información sobre el uso de los reportes web de resultados, se recomienda consultar el documento "Manual de usuario.pdf" entregado junto a este documento.

### 6. Socialización

La metodología y resultados de este proyecto fueron socializados con la Subdirección de Ordenamiento y Desarrollo Territorial (SODT).



## 7. 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](mailto:ucd@dnpp.gov.co).

## ANEXOS

### Anexo 1 Expresiones regulares utilizadas en la estrategia de búsqueda de términos

Dimensión	Términos	Expresión regular
1. Planeación y ordenamiento y, gestión para el desarrollo	Planeación	planeaci[oó]n
1. Planeación y ordenamiento y, gestión para el desarrollo	Ordenamiento	ordenamiento
1. Planeación y ordenamiento y, gestión para el desarrollo	Gestión para el desarrollo	gesti[oó]n desarrollo
1. Planeación y ordenamiento y, gestión para el desarrollo	Ordenamiento territorial	ordenamiento territorial
1. Planeación y ordenamiento y, gestión para el desarrollo	Catastro multipropósito	catastro multiprop[oó]sito
1. Planeación y ordenamiento y, gestión para el desarrollo	Suelo	suelo
1. Planeación y ordenamiento y, gestión para el desarrollo	Naturales	natural(?:es)
2. Conectividad, infraestructura y transporte	Conectividad	conectividad
2. Conectividad, infraestructura y transporte	Infraestructura	infraestructura
2. Conectividad, infraestructura y transporte	Transporte	transporte
2. Conectividad, infraestructura y transporte	Energía	energ[ii]a
2. Conectividad, infraestructura y transporte	Alcantarillado	alcantarillado
2. Conectividad, infraestructura y transporte	Acueducto	acueducto
2. Conectividad, infraestructura y transporte	Internet	internet
2. Conectividad, infraestructura y transporte	Vivienda	vivienda
3. Capital humano	Capital humano	capital humano
3. Capital humano	demografía	demograf[ii]a
3. Capital humano	Mortalidad	mortal(?:idad)
3. Capital humano	Salud	salud
3. Capital humano	Educación	educaci[oó]n
3. Capital humano	Pruebas saber	prueba(?:s) saber
4. Actividades económicas	Economía	econom[ii]a
4. Actividades económicas	Productivo	productiv
4. Actividades económicas	Personas ocupadas	persona(?:s) ocupada(?:s)
4. Actividades económicas	Tasa de ocupación	tasa ocupaci[oó]n
4. Actividades económicas	Desempleo	desempleo
4. Actividades económicas	Banca	banca



Dimensión	Términos	Expresión regular
4. Actividades económicas	Financiera	financiera
5. Gobernabilidad, desempeño y transparencia local	Gobernabilidad	gobernabilidad
5. Gobernabilidad, desempeño y transparencia local	Transparencia	transparencia
5. Gobernabilidad, desempeño y transparencia local	Medición de desempeño	medici[oó]n desempeño
5. Gobernabilidad, desempeño y transparencia local	Desempeño fiscal	desempeño fiscal
5. Gobernabilidad, desempeño y transparencia local	Desempeño institucional	desempeño instituci
5. Gobernabilidad, desempeño y transparencia local	Regalías	regal[ii]as
6. Seguridad y convivencia	Seguridad	seguridad
6. Seguridad y convivencia	Convivencia	convivencia
6. Seguridad y convivencia	Homicidio	homicidio
6. Seguridad y convivencia	Hurto	hurto
6. Seguridad y convivencia	Delito	delito
6. Seguridad y convivencia	Violencia	violencia
6. Seguridad y convivencia	Minas	minas
6. Seguridad y convivencia	Desplazamiento	desplazamiento
6. Seguridad y convivencia	Coca	coca
7. Ambiente	Medio ambiente	medio ambiente
7. Ambiente	Ecosistema	ecosistema
7. Ambiente	Forestal	forestal
7. Ambiente	Basura	basura
7. Ambiente	Residuo	residuo
7. Ambiente	Área protegida	[aá]rea(?:s) protegida(?:s)

Fuente: Elaboración propia