



**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación



El futuro  
es de todos

DNP  
Departamento  
Nacional de Planeación

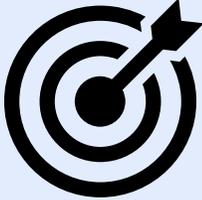
# Análisis PDTs de Departamentos y Municipios costeros 2020-2023

**Unidad de Científicos de Datos**  
Dirección de Desarrollo Digital

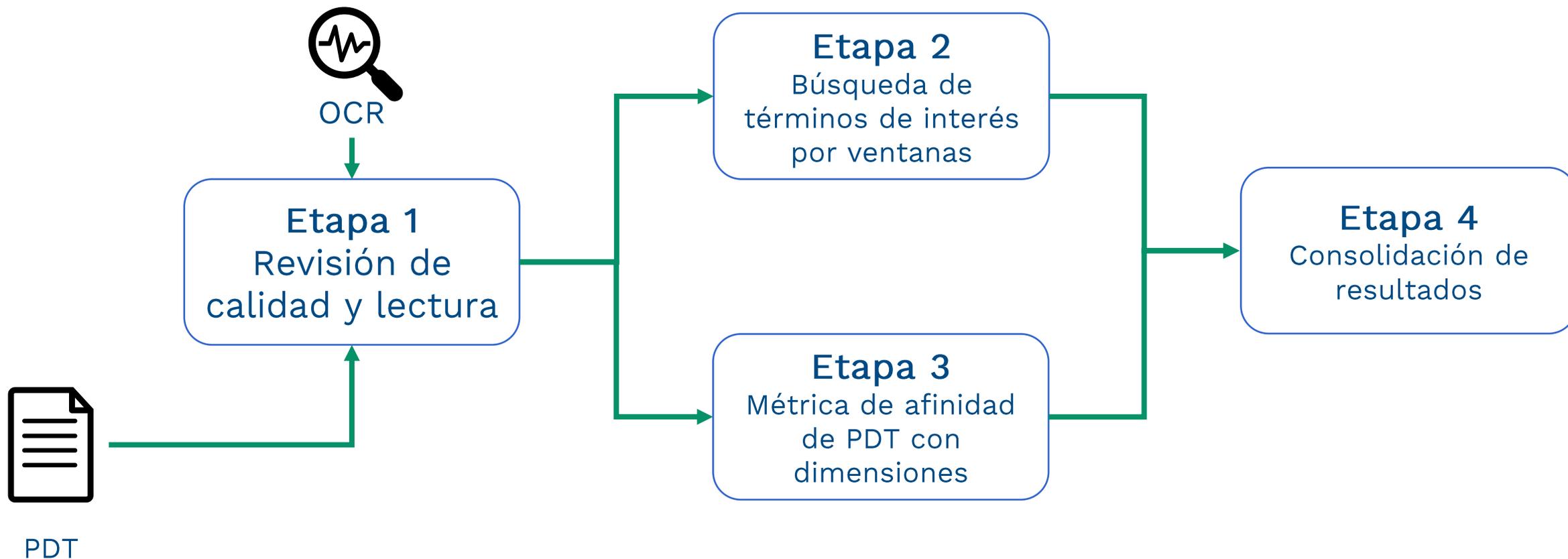
2021



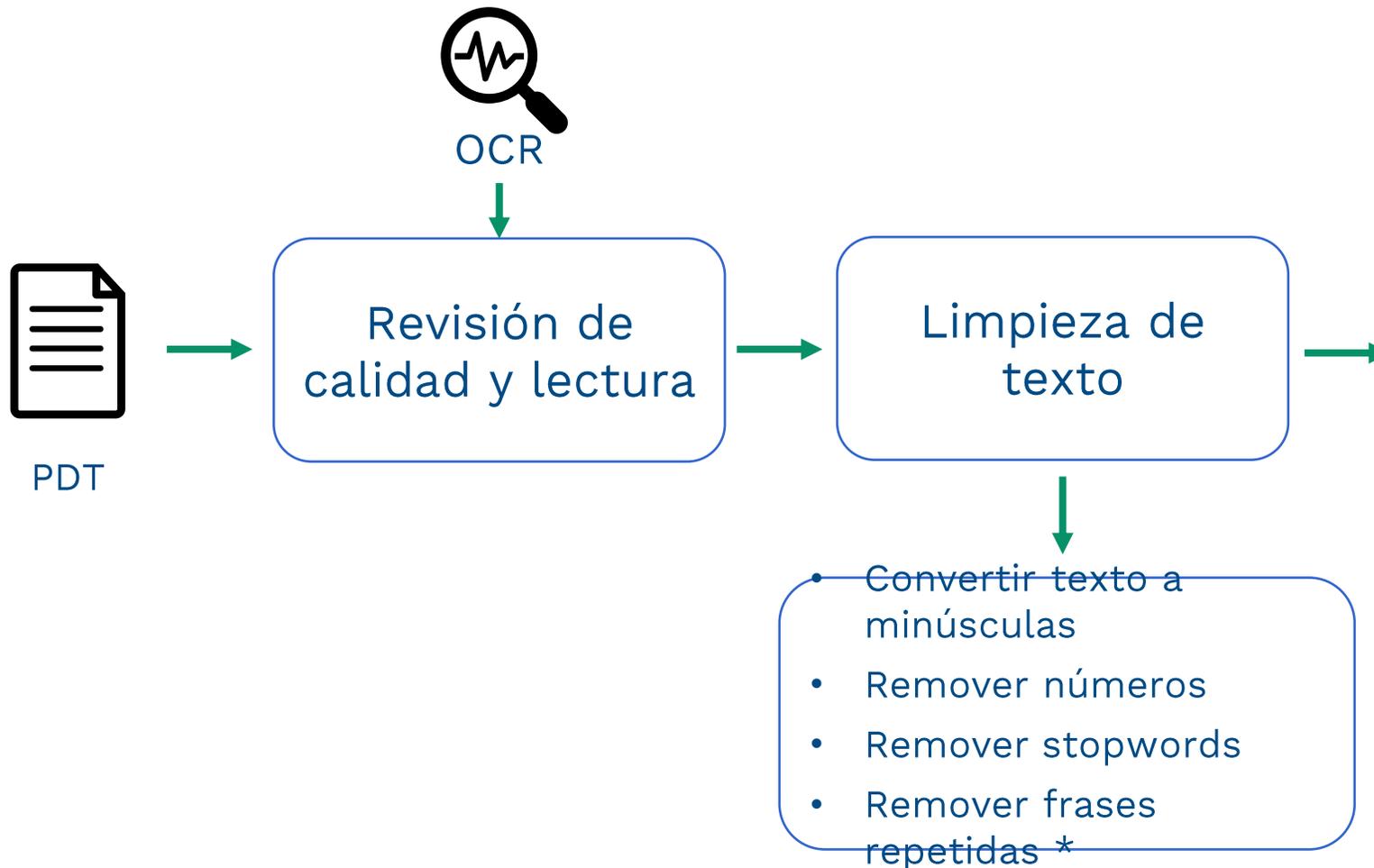
# Análisis PDTs de Departamentos y Municipios costeros

<h2>¿Qué?</h2>	<p>Realizar una comparación de reiteración de palabras en los PDT 2020-2023 de los departamentos y municipios costeros del país, con base a unos términos claves resultantes del diseño del Índice de Desarrollo.</p>	
<h2>¿Para qué?</h2>	<p>Cuantificar que tan alineados están los PDT costeros 2020-2023 con 7 dimensiones del Índice de Desarrollo</p>	
<h2>Logros</h2>	<ol style="list-style-type: none"> <li>1. Metodología de búsqueda flexible de términos.</li> <li>2. Indicador para cuantificar el número de términos dentro de cada PDT.</li> <li>3. Resultados consolidados para Departamentos y Municipios.</li> </ol>	

# Metodología general



# Metodología: Lectura - Limpieza



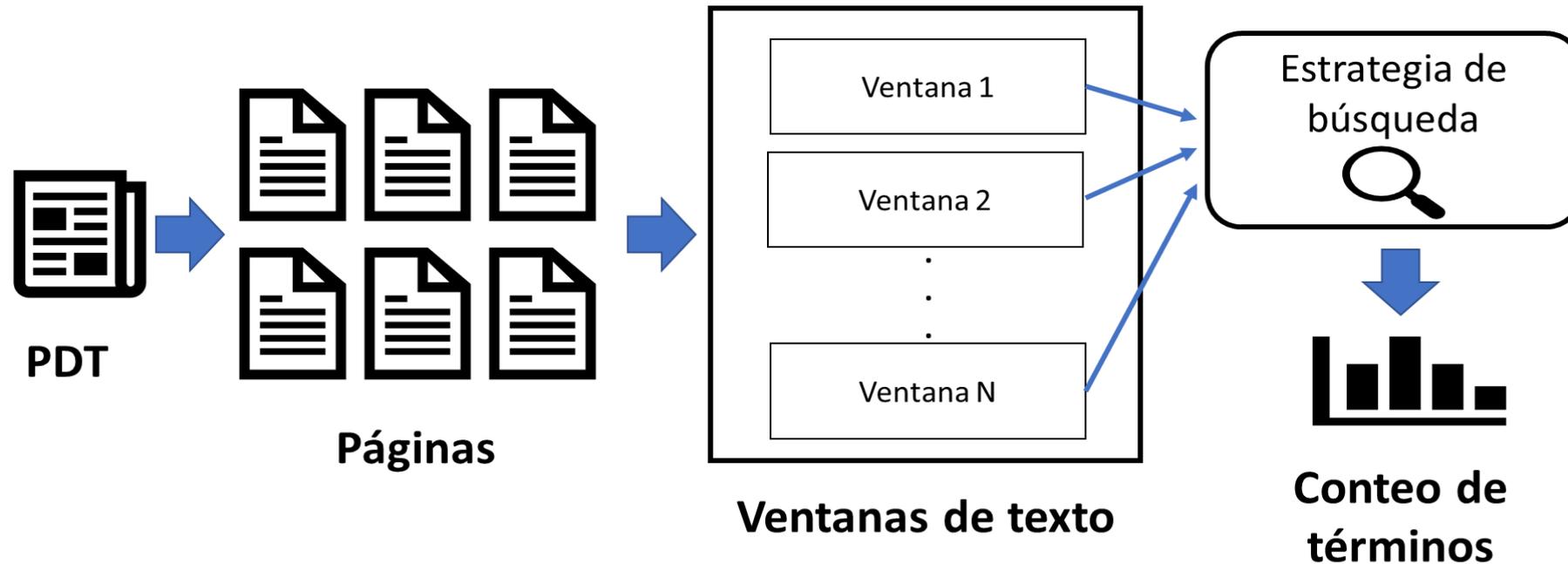
## Texto Original

Unidos por la vida. Unidos todos por la vida en Antioquia, es nuestro mayor propósito. Es el momento para forjar juntos una agenda que consolide a nuestro Departamento como un territorio saludable, seguro, en paz, con equidad y prosperidad, que nos ofrezca oportunidades para acordar una agenda colectiva, aprovechar así nuestras riquezas y forjar progreso con desarrollo humano sostenible. Desde el primer momento del gobierno, emprendimos la formulación colectiva de del Plan de Desarrollo UNIDOS POR LA VIDA, que representa la convergencia y el compromiso de toda Antioquia para trabajar por nuestro futuro común.

## Texto Limpio

unidos vida antioquia mayor propósito momento forjar juntos agenda consolide departamento territorio saludable seguro paz equidad prosperidad ofrezca oportunidades acordar colectiva aprovechar riquezas progreso desarrollo humano sostenible primer gobierno emprendimos formulación plan representa convergencia compromiso toda trabajar futuro común

# Metodología: Estrategia de búsqueda



- Tamaño ventana: 200 palabras
- 49 términos clasificados en 7 dimensiones
- Búsqueda flexible

# Metodología: búsqueda flexible

Expresión regular : **brechas\***|**mujer(?:es)\***|**hombres\***

las **brechas** que existen entre **hombres** y **mujeres** es igual a la **brecha** entre **hombre** y **mujer** o **brecha** entre **mujer** y **hombre**

Texto de prueba

**Brechas entre hombres y mujeres: 3 ocurrencias**

# Metodología: Cuantificación de frases en PDT

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

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

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

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

# Metodología: ejemplo

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

Sumatoria sobre todas las ventanas y términos (9 ocurrencias) y se divide entre el total de páginas (10 páginas) para un puntaje de

$$I_i = 0,9$$

Se multiplica el número de términos de cada ventana por el número de ocurrencias, esto es, (3x4) para la primera ventana, (1x1) para la segunda, (2x4) para la tercera. Se hace la suma de estos valores (21) y se divide sobre el número de páginas.  $IP_i = 2,1$

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

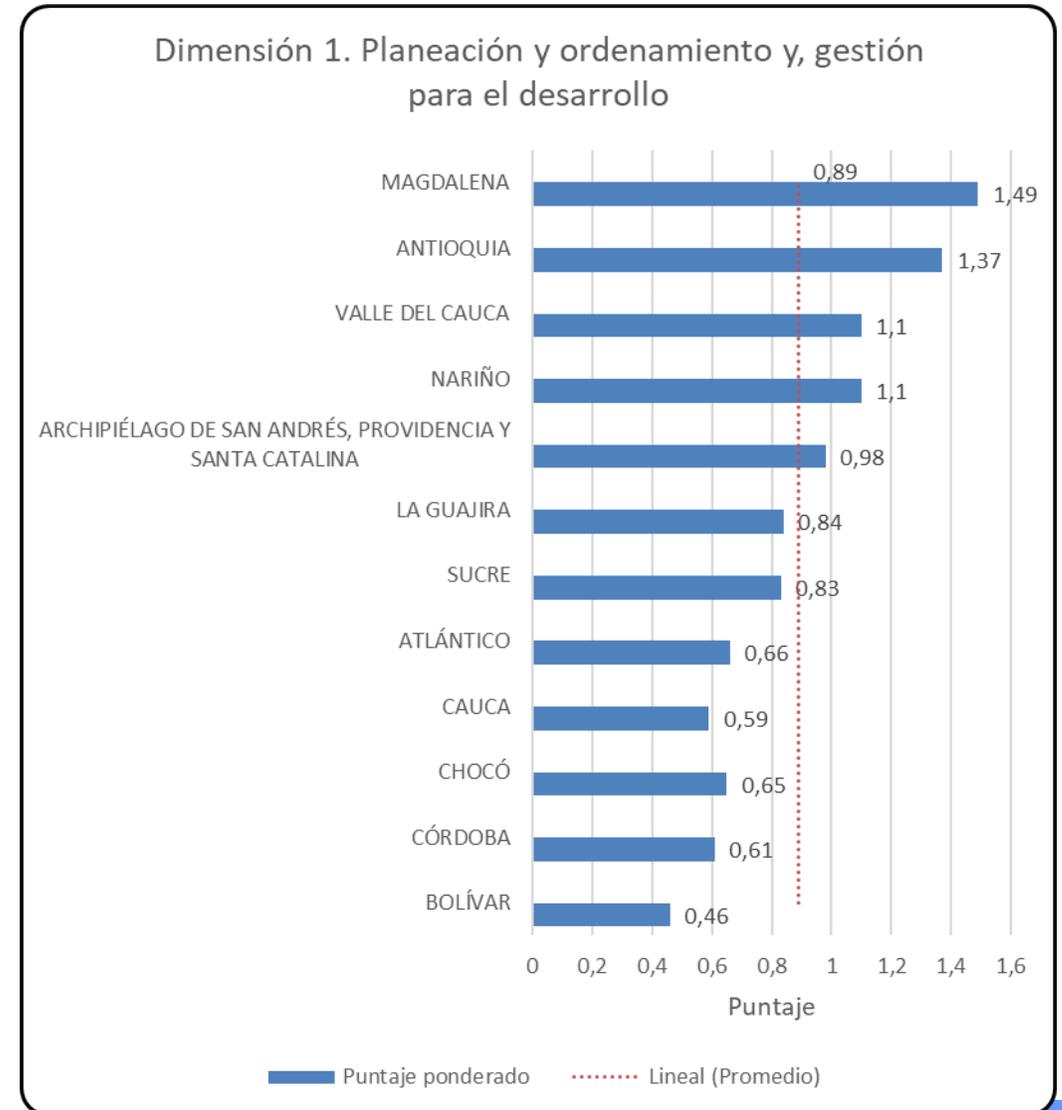
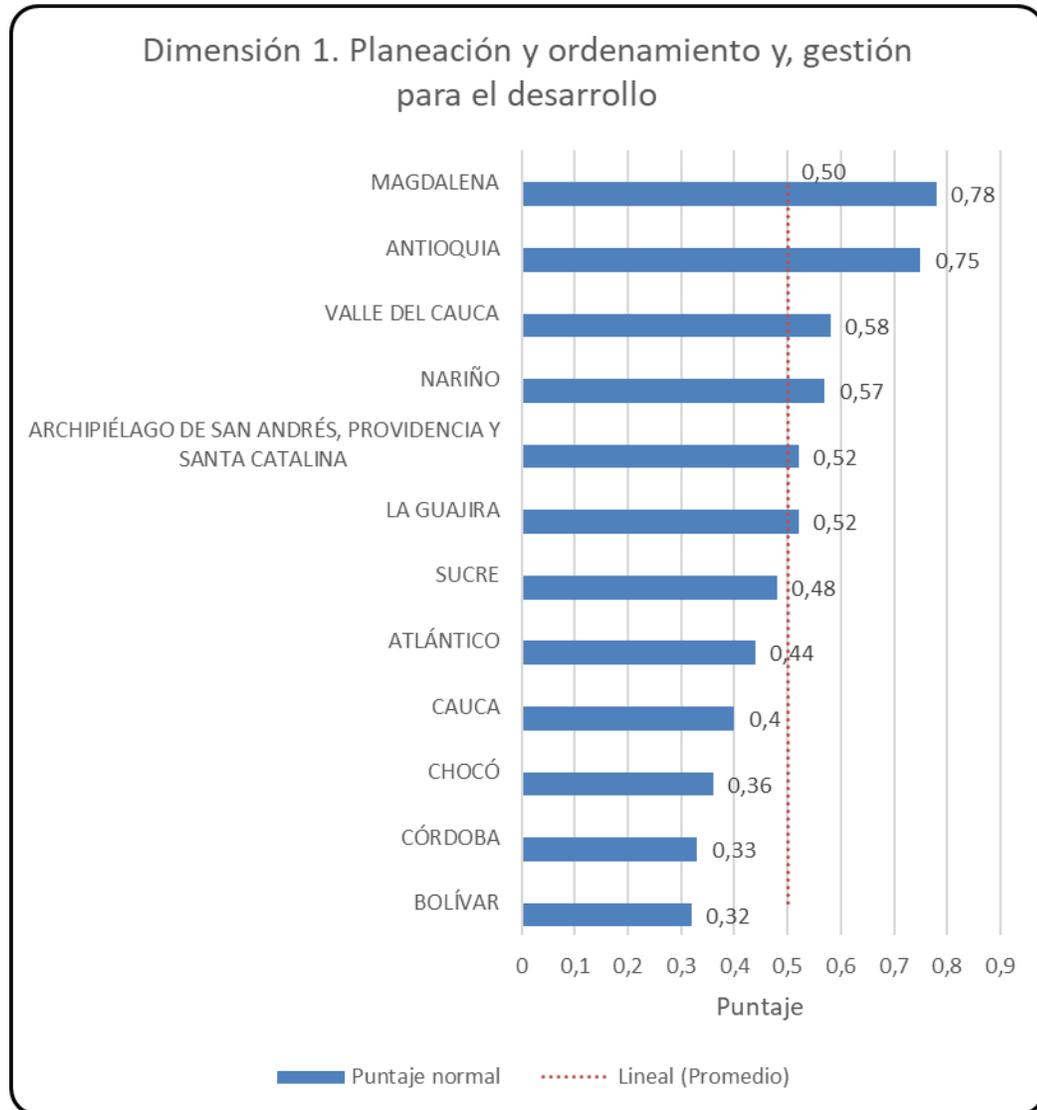
Sumatoria sobre todas las ventanas y términos (9 ocurrencias) y se divide entre el total de páginas (10 páginas) para un puntaje de

$$I_i = 0,9$$

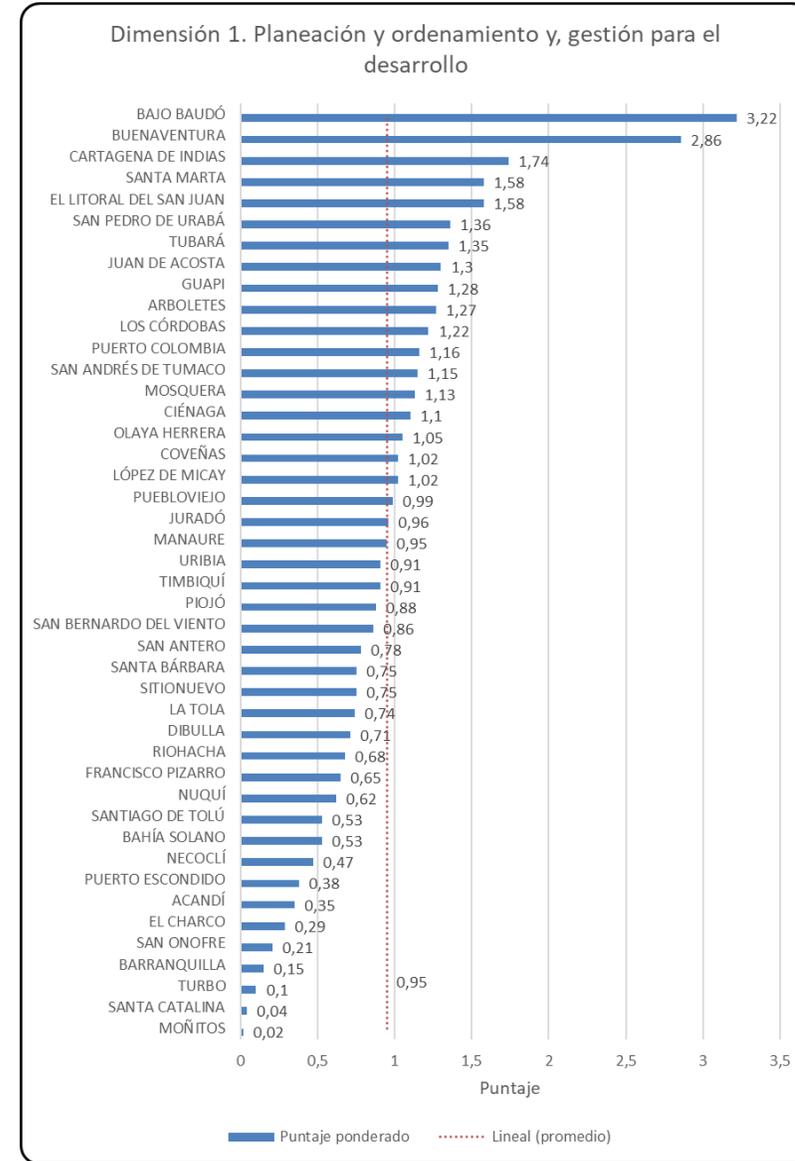
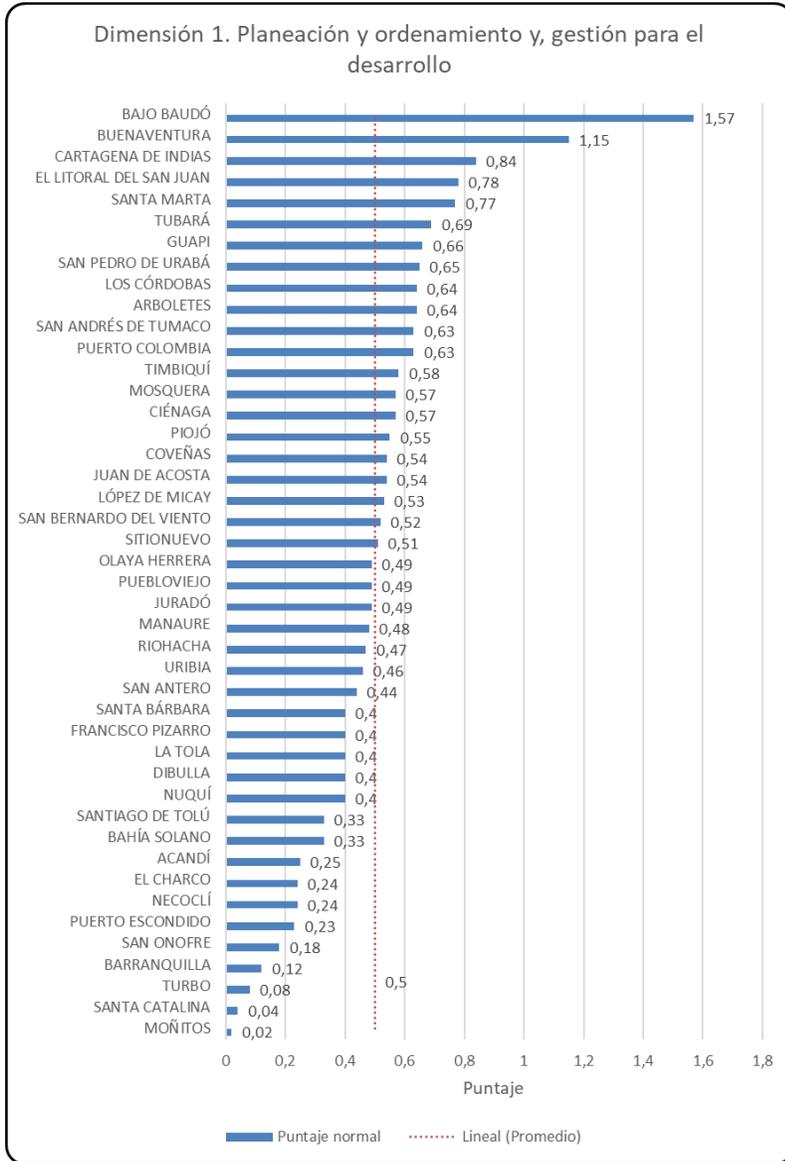
Se multiplica el número de términos de cada ventana por el número de ocurrencias, esto es, (1x4) para la primera ventana, (1x1) para la segunda, (2x4) para la tercera. Se hace la suma de estos valores (13) y se divide sobre el número de páginas.  $IP_i = 1,3$



# Resultados: Departamentos

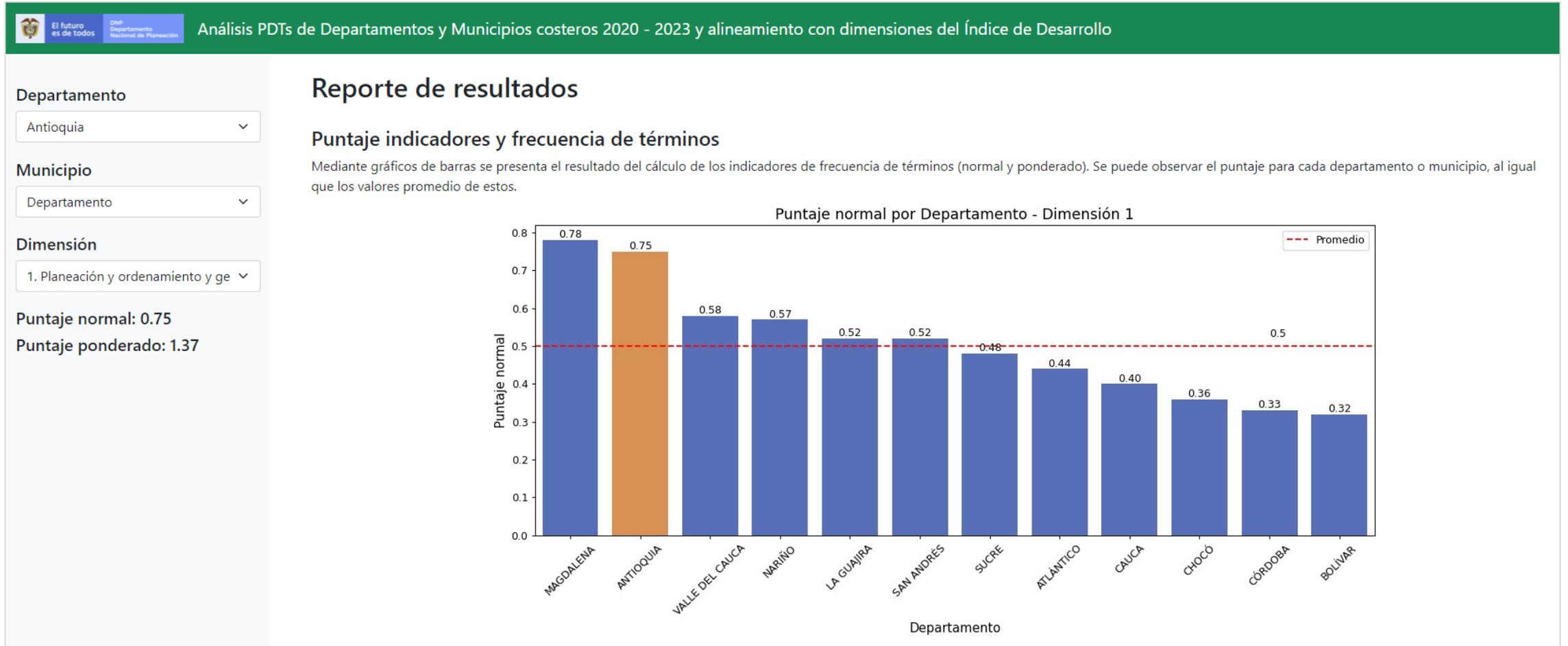


# Resultados: Municipios



# Reporte de resultados

Se tiene un reporte (archivo HTML) por cada dimensión.



# Conclusiones y recomendaciones

1

Se logró implementar una metodología de búsqueda flexible de términos, la cual permitió encontrar múltiples formas de los términos de interés.

2

Se recomienda que los resultados obtenidos se analicen desde la visión de un experto temático, dado a que este proyecto fue un análisis automatizado sobre los textos de los PDT.





**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación