

# Dirección de Desarrollo Digital

Unidad de Científicos  
de Datos



**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación



### Análisis de presencia en diarios de los últimos 5 gobiernos

#### Entidad

Departamento Nacional de Planeación

- Dirección de Desarrollo Digital.
- Dirección General

#### Sector

Planeación

#### Lenguaje

Python

#### Fuente de datos

Base con información de portadas de los dos diarios principales del país (El Tiempo y El Espectador) durante el primer año de gobierno de los últimos 5 gobiernos

#### Presentación

Los medios de comunicación tienen una influencia bastante importante sobre la opinión de la ciudadanía acerca de diversos temas. La manera en la que se presenta la información, junto con lo que se decide presentar, pueden incidir directamente en la relevancia y la recepción de diferentes asuntos. En este proyecto se estudia la prensa escrita, por medio de un análisis a los dos periódicos principales de Colombia, El Tiempo y El Espectador. En particular, se desea observar cómo la prensa ha informado, desde sus portadas, sobre los últimos 5 periodos presidenciales, y sus respectivos presidentes y gobiernos. Para cada periodo presidencial se estudiaron las portadas de ambos diarios durante el primer año de gobierno, para estudiar temas como presencia e impacto de los gobiernos en los medios, así como un análisis exploratorio de los principales temas abordados por cada periódico.

*The media have a fairly important influence on the opinion of the public on various issues. The way in which the information is presented, together with what they decide to present, can directly affect the relevance and reception of different matters. This project studies the written press, through an analysis of the two main Colombian newspapers, El Tiempo and El Espectador. In particular, it is desired to observe how the press has reported, from its covers, on the last 5 presidential terms and their presidents and governments. For each presidential term, the covers of both newspapers were studied during the first year of government, to study issues such as the presence and impact of governments in the media, as well as an exploratory analysis of the main topics addressed by each newspaper.*

#### Objetivo general

Determinar cómo ha evolucionado la presencia del presidente y el gobierno en prensa durante el primer año de gobierno de los últimos 5 gobiernos.

#### Objetivos específicos

1. Identificar la cantidad de veces que cada presidente es mencionado en alguna parte de las portadas de periódicos.
2. Cuantificar y analizar la evolución del impacto (positivo, neutro o negativo) que tiene cada presidente y su gobierno a lo largo de su primer año de gobierno.
3. Estudiar qué entidades públicas fueron mencionadas durante cada año de mandato, y con qué frecuencia fueron mencionadas.
4. Realizar análisis exploratorio y no supervisado sobre los textos de las portadas, para conocer los términos y temas más relevantes reportados durante cada año de gobierno



### Metodología

El desarrollo de este proyecto se puede dividir en cuatro etapas, como se muestra en la Figura 1. A continuación se describe el trabajo realizado en cada una de estas etapas.

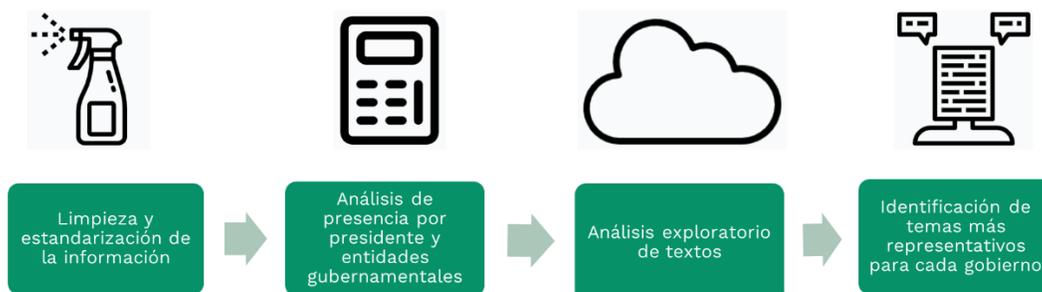


Figura 1: Diagrama de desarrollo del proyecto

### **Limpieza y estandarización de la información**

La base de portadas de periódicos, recolectada por asesores de Dirección General del DNP, contiene bastante información de utilidad, incluyendo:

- Fecha de publicación de la portada
- Título principal de la portada
- Título y contexto de noticias sobre el presidente y/o el gobierno (si las hay)
- Impacto de noticias sobre el presidente y/o el gobierno, marcado como “positivo”, “negativo” o “neutro” (si las hay)
- Lista de entidades públicas mencionadas en cada portada (si las hay)

Sobre estas variables de interés hubo necesidad de realizar preprocesamiento, para que fueran aptas para el análisis. En primer lugar, para las variables de texto (títulos y contextos), se realizó una limpieza estándar, en la que se pasó todo el texto a minúsculas, se removieron signos de puntuación y palabras que no aportan mucho significado (*stopwords*); también se reemplazaron los valores faltantes por cadenas de texto vacías, para facilitar el procesamiento de las variables. Luego, la lista de entidades públicas fue estandarizada (una misma entidad podía ser mencionada o escrita de diferentes maneras) y separada, de manera que se pudiera contar la frecuencia de aparición de cada entidad por separado. Finalmente, las dos variables de impacto (del presidente y del gobierno) fueron convertidas a columnas numéricas, dando un peso de +1 a los impactos positivos, de -1 a los impactos negativos y de 0 cuando el impacto era neutro o no había (valores faltantes).

Adicionalmente, se crearon algunas variables nuevas, para complementar el análisis. Dos de estas variables consisten en uniones de todo el texto disponible en cada portada (juntando todas las columnas de texto pre - procesadas) y una versión ligeramente distinta, en la que se unen solo los textos correspondientes al presidente o al gobierno. También se utilizó la fecha de cada portada para construir la variable “mes de gobierno”, que identifica, para cada portada, en qué mes (del 0 al 12) del primer año de gobierno fue publicada. De igual forma se construyó la variable “periodo” para identificar cada periodo presidencial, diferenciándolo de otros periodos, incluso si tuvieron el mismo presidente (como fue el caso con Juan Manuel Santos o Álvaro Uribe). Finalmente, se construyeron varias variables para determinar si cada presidente es mencionado en diferentes partes de la portada. Para construir estas variables se tuvo en cuenta el periodo de gobierno en el que cada portada fue



# El futuro es de todos

DNP  
Departamento  
Nacional de Planeación

publicada, buscando solamente el nombre del presidente de turno. Además del nombre de cada presidente (en distintas versiones) también se buscaron palabras como “presidente” y “mandatario”.

## **Análisis de presencia e impacto**

Una vez los datos estuvieron adecuados para los análisis, el siguiente paso fue analizar la presencia en portadas de los presidentes en cada año de gobierno, así como el impacto agregado (es decir, la sumatoria) para el gobierno y el presidente de turno. Estos análisis se hicieron en general, para todo el año de gobierno, y por meses, para estudiar cómo evolucionaba el impacto y la presencia en medios a lo largo de los diferentes meses del primer año de gobierno. El análisis de presencia de cada presidente (entendido como si fue mencionado o no) se hizo tanto de manera diferenciada para distintas partes de las portadas (el título principal, el título de gobierno, el título de presidente y los contextos) como de manera agregada (si eran mencionados en cualquier título, o en cualquier parte de la portada).

Adicionalmente, se hizo un conteo de frecuencias de mención o aparición de cada entidad pública en las portadas. Este análisis permite estudiar cuáles entidades han tenido más relevancia mediática para cada gobierno durante su primer año, lo que puede a su vez dar alguna idea de los proyectos o prioridades de cada gobierno en ese periodo de tiempo.

## **Análisis exploratorio de textos**

Una vez hechos los análisis de presencia, impacto y entidades públicas mencionadas, el siguiente paso es realizar un análisis exploratorio a los textos (títulos y contextos) de las portadas. Una herramienta que permite hacer exploración visual de textos, de manera sencilla, es la nube de palabras. Las nubes de palabras son herramientas gráficas que presentan los términos o expresiones más frecuentes de un texto e indican la frecuencia de aparición de cada uno de estos, mediante su color y/o tamaño. Este tipo de análisis permite tener una idea de cuáles eran los temas más relevantes mencionados en prensa durante cada año de gobierno.

## **Identificación de temas y términos más representativos**

Finalmente, se llevó el análisis de los textos un paso más allá, para identificar los temas más relevantes publicados en prensa por el medio de técnicas de modelamiento de temas. Para realizar esto, se realizaron dos operaciones sobre los textos.

En primer lugar, se utilizó la técnica de TF-IDF (siglas en inglés para *Term frequency – Inverse document frequency*) para obtener una representación vectorial de los textos. Esta técnica permite representar un texto como un vector de números que dependen de dos factores: la frecuencia con la que aparece cada término en dicho texto y la frecuencia con la que aparece el término en el conjunto total de documentos. Esta transformación vectorial fue ajustada para tener en cuenta términos de 1 y 2 palabras, y se acotó para tener en cuenta solamente los 20.000 términos más frecuentes en el conjunto de portadas. Esta acotación se hizo para evitar trabajar con vectores de dimensionalidad demasiado alta, debido a la inclusión de términos muy poco frecuentes.

Una vez se tienen las representaciones vectoriales de los textos, se utilizó una técnica llamada LDA (siglas en inglés para *Latent Dirichlet Allocation*), que permite categorizar un conjunto de documentos (en este caso, los textos de cada portada) dentro de una cantidad determinada de grupos debido a la similitud de algunos de sus componentes. La técnica de LDA permite encontrar temas comunes entre los documentos, lo que a su vez permite determinar los términos (palabras o bi-gramas) más representativos de cada tema, y la medida en la que un texto pertenece a uno o más temas. Con esta información es posible entonces determinar los términos que representan cada tema, en orden de peso o relevancia.



Para realizar este modelamiento de temas no se consideró el título principal de las portadas, pues este título hablaba en muchas ocasiones de temas ajenos al gobierno, que podrían añadir mucho ruido al análisis. Esto fue particularmente cierto en el caso del periódico El Espectador, cuyos títulos principales trataban principalmente temas ajenos al gobierno de turno.

### Resultados

Los resultados de los análisis realizados en este proyecto fueron compilados y presentados en dos reportes interactivos, uno por cada periódico estudiado. A continuación, se muestran algunas de las visualizaciones y resultados obtenidos.

### Análisis de presencia e impacto

La Figura 2 muestra las frecuencias de mención o aparición de cada presidente en el periódico El Tiempo durante sus primeros años de gobierno. Se puede apreciar que Álvaro Uribe tuvo mucha más presencia mediática en sus dos gobiernos, mientras que Juan Manuel Santos e Iván Duque presentan unas cifras similares.

periodo	Iván Duque	Juan Manuel Santos 1	Juan Manuel Santos 2	Álvaro Uribe 1	Álvaro Uribe 2
Aparece Título Principal	42	29	23	26	50
Aparece Título Presidente	62	61	44	56	127
Aparece Título Gobierno	0	1	0	0	0
Aparece Algún Título	65	64	44	56	127
Aparece Contexto Presidente	86	109	105	137	152
Aparece Contexto Gobierno	3	6	4	6	3
Aparece Alguno	120	124	119	151	186

Figura 2: Frecuencia de mención de los presidentes en distintas partes de las portadas (El Tiempo)

Por otro lado, la Figura 3 muestra una tabla con el impacto acumulado de los presidentes y gobiernos durante su primer año de gobierno, de acuerdo con las portadas de El Tiempo. Se puede observar que en general los impactos son positivos, con excepción de Iván Duque y el primer periodo de Álvaro Uribe.

periodo	Iván Duque	Juan Manuel Santos 1	Juan Manuel Santos 2	Álvaro Uribe 1	Álvaro Uribe 2
Impacto del gobierno	-4	9	25	-7	1
Impacto del presidente	30	65	39	50	11

Figura 3: Impacto acumulado de presidentes y gobiernos para cada periodo (El Tiempo)

### Evolución de presencia e impacto en el tiempo

De manera similar, estos análisis se pueden hacer por unidades de tiempo más pequeñas. En este caso se calculó el impacto y las menciones consolidadas de manera mensual, y se estudió su evolución durante el primer año de cada gobierno. La Figura 4 muestra la comparación de esta evolución para dos periodos presidenciales distintos. El informe generado permite interactuar con estas gráficas, permitiendo elegir qué gobiernos mostrar, y hacer *zoom* en partes específicas de la gráfica.



Cantidad de veces que cada presidente es mencionado en la portada

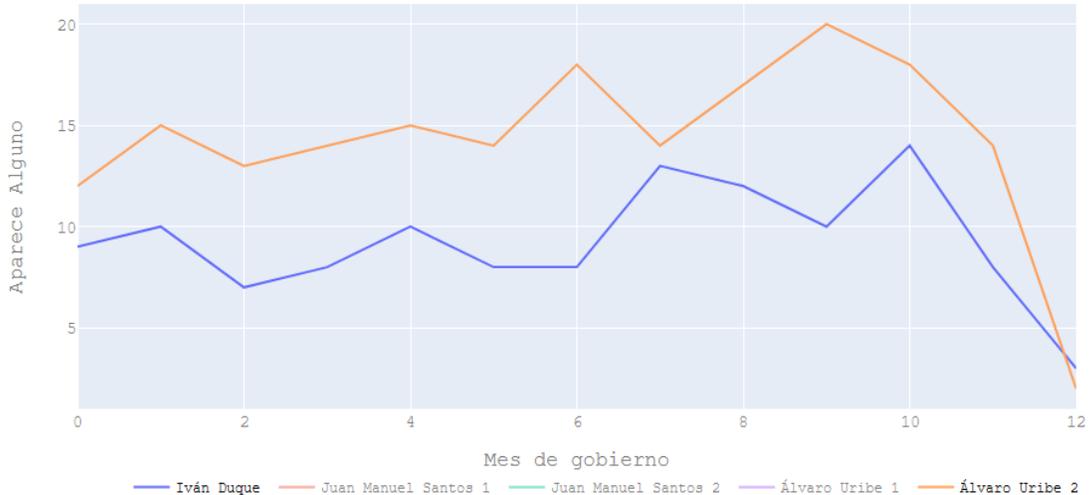


Figura 4: Comparación de presencia en portadas mes a mes entre Iván Duque y Álvaro Uribe, segundo periodo (El Espectador)

De manera similar, se pueden observar y comparar el impacto consolidado mensual. La Figura 5 muestra la comparación de impactos entre el gobierno de Iván Duque y el segundo gobierno de Juan Manuel Santos según el periódico El Espectador, mostrando un panorama más favorable para el segundo.

Impacto del gobierno

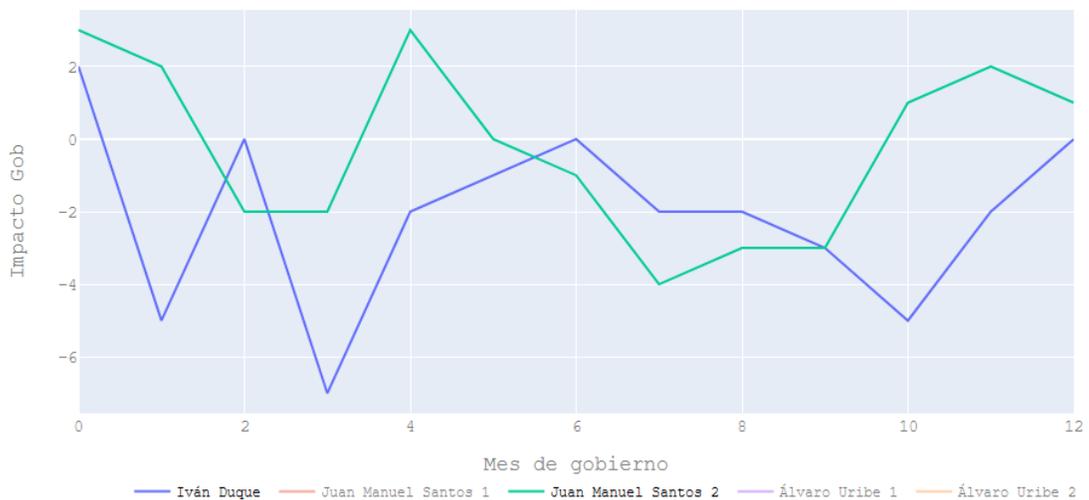


Figura 5: Comparación de impacto acumulado del gobierno mes a mes entre Iván Duque y Juan Manuel Santos, segundo periodo (El Espectador)

Finalmente, también se pueden ver las entidades públicas mencionadas por los periódicos durante cada año de gobierno. La Figura 6 muestra las entidades mencionadas por el periódico El Tiempo durante el primer año



# El futuro es de todos

DNP  
Departamento  
Nacional de Planeación

de gobierno de Juan Manuel Santos. Se puede observar que la Cancillería tuvo bastante protagonismo, junto a otras entidades más esperadas entre las más mencionadas, como el Ministerio de Hacienda y el Ministerio de Defensa.

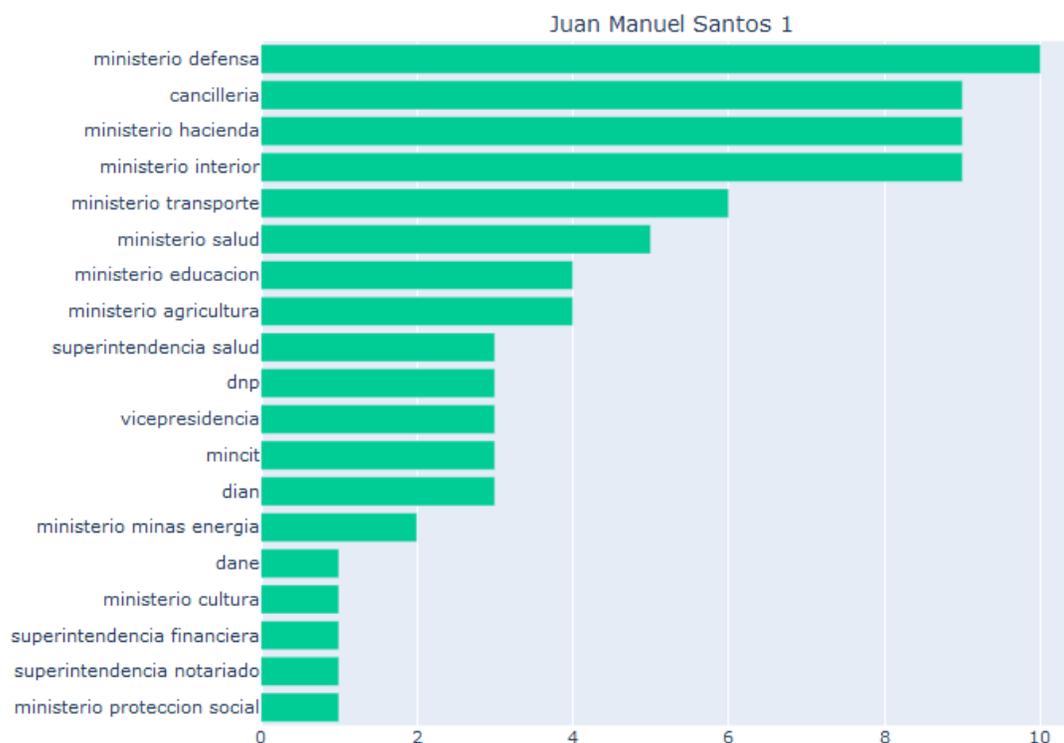


Figura 6: Entidades públicas mencionadas durante el primer gobierno de Juan Manuel Santos (El Tiempo)

## Análisis de textos

En cuanto al análisis exploratorio, la Figura 7 muestra dos nubes de palabras para las portadas durante el primer año del segundo gobierno de Juan Manuel Santos. La nube de palabras de la izquierda muestra los términos más frecuentes, mientras que la nube de la derecha muestra los bi-gramas (conjuntos de dos palabras). Estas gráficas permiten obtener una idea general sobre los temas más tratados por la prensa en ese periodo de tiempo.



# El futuro es de todos

DNP  
Departamento  
Nacional de Planeación

## Términos más frecuentes: Juan Manuel Santos 2



Figura 7: Nubes de palabras y bi-gramas para el primer año del segundo gobierno de Juan Manuel Santos (El Tiempo)

Finalmente, la Figura 8 muestra los términos que, según la técnica de LDA, son más representativos de los 6 temas (un número predeterminado). Aunque algunos temas no son del todo claros a partir de sus términos más representativos, otros pueden ser determinados con mayor claridad. La cantidad de temas es un parámetro que puede ser modificado (junto a otros parámetros del algoritmo LDA) para conseguir un conjunto de temas que se diferencien entre sí.

### Iván Duque

	Tema_1	Tema_2	Tema_3	Tema_4	Tema_5	Tema_6
0	paz	insiste	paro	acuerdo	días	ley
1	congreso	odebrecht	seguridad	corte	paz	nuevo
2	debate	salud	nacional	insolvencia	político	objeciones
3	objeciones	tan grave	anticorrupción	billones	donald trump	dane
4	jep	seis	billones	consumo	donald	ministro
5	pacto	minga	maicao	gran	trump	política
6	reforma	ley	acuerdo	meses	acuerdo	comercio
7	digna	grave	ayer	posesión	onu	cambios

Figura 8: Términos más representativos para temas – Iván Duque (El Espectador)

### Conclusiones y recomendaciones

1. Los estudios de presencia e impacto mostraron ciertas diferencias en las líneas editoriales de los periódicos analizados. Al comparar los resultados de ambos diarios, se puede observar que El Tiempo



# El futuro es de todos

DNP  
Departamento  
Nacional de Planeación

es más propenso a mencionar temas relacionados al presidente o el gobierno de turno, mientras que El Espectador hace esto con menos frecuencia, y usualmente con un impacto más negativo.

2. En cuanto a las menciones de los presidentes en las portadas, se pudo observar que este comportamiento ha sido relativamente estable en los últimos 3 gobiernos para El Tiempo, mientras que en El Espectador se aprecia una tendencia creciente, en la que desde el 2006 el presidente de cada gobierno ha sido mencionado más veces en portadas que su par del gobierno inmediatamente anterior.
3. El estudio de entidades públicas mencionadas en cada gobierno, junto con sus respectivas frecuencias, permite dar una buena idea de cuáles fueron los temas prioritarios de cada gobierno durante su primer año. Estos resultados son consistentes entre ambos diarios.
4. Finalmente, el análisis (exploratorio y de temas) de los textos disponibles también fue bastante útil para tener una mejor idea de los temas y términos más relevantes para los periódicos durante cada año de gobierno.

## **Socialización**

Este proyecto y sus resultados fueron presentados a asesores de Dirección General del DNP.