

Informe sobre discursos de políticos en las elecciones de EEUU de 2020

Tarea 1 - Introducción a la Ciencia de Datos 2025

Introducción:

Este informe presenta un análisis de datos relacionados a discursos de políticos en el marco de las elecciones de EEUU, desde enero hasta octubre de 2020.

Vale la pena recordar, para brindar algo de contexto, que en esas elecciones se enfrentaron, la candidatura demócrata, compuesta por el exvicepresidente Joe Biden y la senadora californiana Kamala Harris, al entonces presidente republicano Donald Trump y al vicepresidente Mike Pence. Bernie Sanders suspendió su campaña presidencial en abril de 2020 y pasó en ese momento a respaldar a Biden.

Calidad de los datos:

En primer lugar, se analiza la calidad de los datos. El data frame contiene 269 filas y 6 columnas vinculadas a las variables: speaker, title, text, date, location, type.

Al revisar los elementos del data frame se determinó que hay 42 campos con datos faltantes, que representan un 2,6%. La mitad de los datos faltantes, 21, se encuentran en la variable "type", seguida por 18 campos nulos en la variable "location" y por 3 campos nulos en el dato del orador del discurso, "speaker". El total de filas con datos faltantes es de 32, representando 11,89% del total de filas. Hay filas como las 151 y 156 que tienen hasta tres elementos con valores nulos: speaker, location y type, mientras que en el resto de las filas con datos nulos el número de campos nulos oscilan entre uno y dos elementos.

En relación al registro de los datos se observan algunas particularidades que pueden derivar en dificultades para hacer los análisis. Por ejemplo, en la columna "speaker", la mayoría de los discursos se corresponde con un solo candidato pero también hay discursos que se asocian simultáneamente con varios candidatos (por ejemplo, "Joe Biden, Barack Obama" o "democratic candidates"). También se observa un registro donde el orador no aparece identificado ('???').

En relación a las locaciones, la mayoría incluyen el nombre de la ciudad y el estado pero hay otras que no siguen la lógica como por ej 'ABC', 'NBC', 'white house', 'CNN'. También en la forma de escribir la ciudad y el estado, la mayoría escriben con palabras el nombre completo del estado pero por ejemplo hay otra que dice 'Freeland, MI'. Estos aspectos serían obstáculos para un eventual análisis geográfico de los discursos.

Cantidad y cronología de los discursos

Analizando propiamente los datos, en relación a la cantidad de discursos por candidato/a, la tabla N°1 muestra la cantidad de discursos realizados por los políticos que tuvieron más presentaciones. Aquí se puede ver una mayor presencia de Joe Biden. Posiblemente esta presencia es más fuerte si se consideran los 8 discursos que se identificaron bajo el nombre de candidatos demócratas.

Tabla N° 1: Cantidad de discursos por candidato/a¹

Candidato	Cantidad de discursos
Joe Biden	81
Donald Trump	55
Mike Pence	20
Bernie Sanders	17
Kamala Harris	16
Democratic Candidates	8

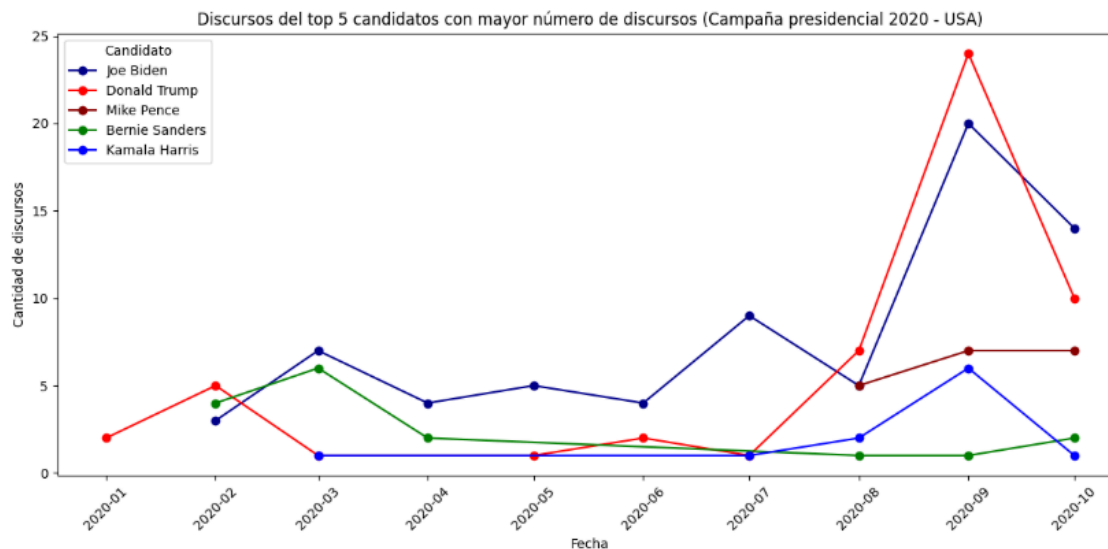
A partir de este punto, para el análisis de los discursos, se considerarán estos 5 políticos y en particular, únicamente los discursos que se identificaron con un solo orador. La tabla que muestra la cantidad de discursos considerando oradores únicos se muestra en Anexos y como se puede ver, se siguen las mismas tendencias de la tabla presentada anteriormente.

Uno de los aspectos relevantes a destacar es el momento en que se dan los discursos. En este sentido, el gráfico de la figura 1 muestra que a partir del mes julio la cantidad de discursos de Biden se fue incrementando y para el resto de los candidatos esto ocurrió en agosto, a excepción de Sanders, donde mayoritariamente sus discursos se ubican en marzo, época donde era candidato presidencial. Se observa que el mes de septiembre fue el que registró mayor número de discursos, esto se puede relacionar con el inicio de la fase decisiva de la campaña y el interés de movilizar votantes. Igualmente, resulta importante destacar la fecha del debate presidencial, el cual se realizó el 29 de septiembre y generó mayor movilización y repercusión mediática.

Otro elemento a considerar son las locaciones de los discursos, en el mes de septiembre la mayoría fueron realizados en estados claves para la definición de la campaña como Florida, Pensilvania, Ohio y Wisconsin. En la tabla del anexo 3 se pueden ver las cantidades correspondientes para cada estado. En el caso de Florida, Pensilvania y Wisconsin en 2016 tuvieron un margen de 3 puntos o menos en las elecciones presidenciales, y Ohio, aunque no en 2016, sí en 3 elecciones entre 1992 y 2020 tuvo un margen de 3 puntos o menos. Estos son algunos de los estados altamente competitivos que históricamente han oscilado entre el voto a uno o a otro partido en las elecciones presidenciales, denominados “swing states”. Muchas veces las campañas políticas en EEUU tienden a enfocarse en estos estados.

¹ Aquí se consideran los discursos donde estuvieron presentes los políticos, solos o acompañados por otros políticos.

Fig.1 : Cantidad de discursos por top 5 candidatos con más discursos



Análisis de palabras de los discursos:

A partir de acá, tanto para el análisis del conteo de palabras como para las palabras más frecuentes de los discursos de los políticos, se tomó la decisión de trabajar únicamente los discursos del tipo “campaign speech” de los 5 políticos que tenían mayor cantidad de discursos. Esto se debe a que en otro tipo de discursos más interactivos o de debate las características del discurso son generalmente algo distintas. Por ejemplo, en los discursos de “town hall” los aspirantes responden preguntas e inquietudes planteadas por la audiencia, algo similar ocurre en las “interview” a partir de preguntas de periodistas, por lo que los temas o las palabras más frecuentes surgen del intercambio, así como también la duración del discurso. Es decir, estos formatos pueden tener características distintas a los discursos de campaña donde cada político y su equipo decide donde se pone el foco y de qué manera.

También en los debates las menciones a otros candidatos pueden tener mayor frecuencia y dependen fuertemente de los involucrados en el debate. Por este motivo, el análisis se centrará en los discursos de campaña, que a su vez, representan el tipo de discurso más frecuente de los datos que se tienen (180 discursos de campaña de un total de 248², algo más del 72% de los discursos).

Cantidad de palabras

Tanto para el conteo de palabras como para el análisis de las palabras más frecuentes se realizó la limpieza del texto: se quitaron datos de los momentos en que se capturaba cada parte del discurso así como el nombre del orador en esos lugares específicos del texto; se normalizó el texto convirtiendo todas las palabras a minúsculas; a la función se le añadió eliminar paréntesis, espacios múltiples y signos de puntuación, ya que al momento de contabilizar el número de palabras pueden influir en el resultado y dar datos erróneos. Sin este proceso de limpieza y normalización, los resultados habrían sido imprecisos.

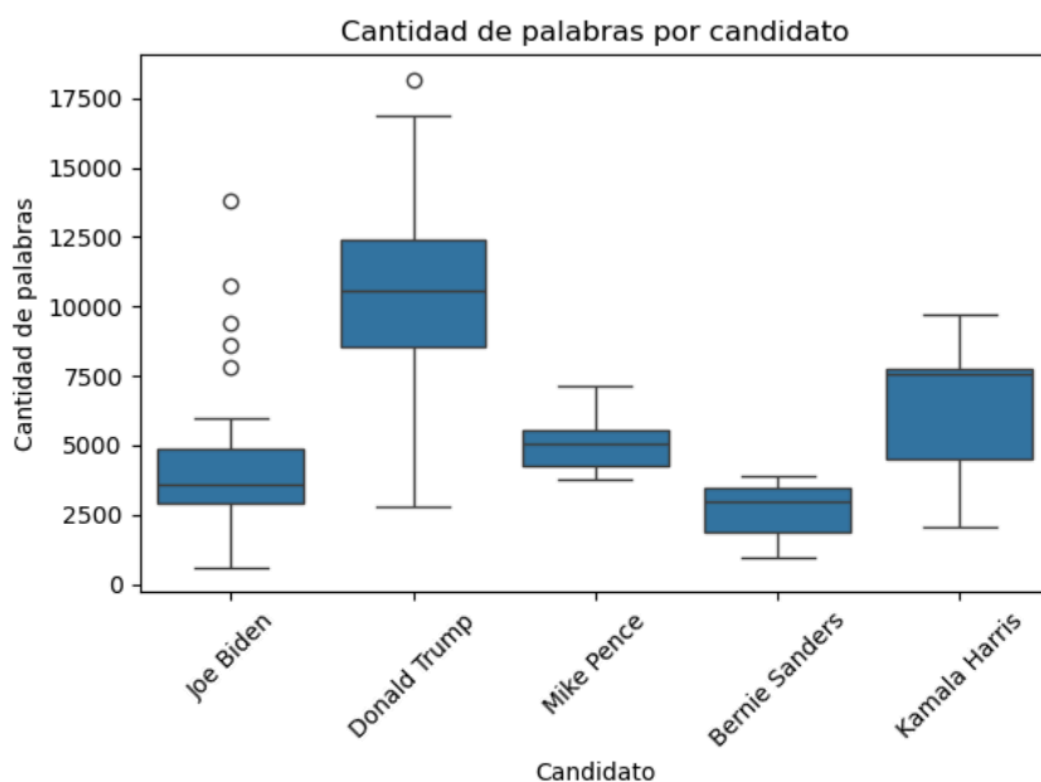
En relación a la cantidad de palabras por candidato se observa que Donald Trump tiene un total de palabras muy superior a los demás aún cuando no fue el candidato que realizó más cantidad de discursos. Trump pronunció un total de 429.174 palabras en 41 discursos, seguido por Biden con un

² Esta cifra representa el total de discursos no vacíos en el campo “type”.

total de 182.773 palabras en 43 discursos de campaña. En la tabla del anexo se puede ver la cantidad de palabras de los cinco candidatos analizados.

El gráfico de la fig 2. ilustra algunos aspectos más al respecto. Si bien los discursos de Trump presentan mayor variabilidad en relación a la cantidad de palabras, se puede apreciar que mayoritariamente la cantidad de palabras es superior al resto de los candidatos, donde la mediana de palabras se sitúa por encima de las 10000 palabras, valor únicamente superado en 2 discursos atípicos de Biden. Por otro lado, los discursos de Biden, su principal opositor, rara vez superan las 7000 palabras y son relativamente similares, en extensión, a los de Mike Pence y Bernie Sanders, que pueden considerarse generalmente cortos y consistentes, sin grandes variaciones en su duración. En el caso de los discursos de Kamala Harris, la mediana se ubica cercana a las 7500 palabras, por lo que sus discursos pueden tender a ubicarse en una posición intermedia en cuanto a duración.

Fig.2: Cantidad de palabras por candidato en los discursos de campaña.



Palabras más frecuentes

Para el conteo de palabras más frecuentes se aplicó una función en el campo "WordList" (texto limpio separado en palabras, usado previamente para el conteo) para identificar las tres palabras que más se repetía según los discursos de cada mandatario. Para ello se importó el módulo "re" que permite identificar los elementos coincidentes en el campo "WordList".

El resultado obtenido se muestra en la figura 3.

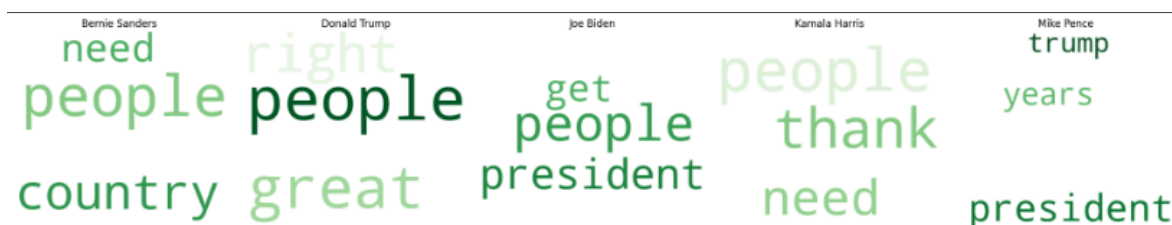
Fig. 3 Palabra más frecuente por candidato presidencial



En la figura n°3 se puede observar que la palabra más frecuente es “the” el cual es un artículo definido, seguidamente los conectores “and” y “to”, por ello el resultado no es el esperado ya que no son palabras que aporten contenido por sí solas, su función es conectar y estructurar oraciones. Si lo que se quiere es dar un contexto del mensaje que cada mandatario quería dar en sus discursos se debe crear una lista con estas palabras denominadas “stopwords”, eliminar: artículos, preposiciones, conjunciones, pronombres y verbos comunes, esto con la finalidad de establecer un criterio al momento de eliminar palabras mitigando el sesgo que podría generar eliminar palabras de manera aleatoria.

Al aplicar la función de eliminar palabras el resultado fue que la palabra más usada por los candidatos Biden, Trump, Sanders y Harris fue "people", mientras que para Pence fue “president”. Poniendo el foco en los cuatro candidatos a presidente y vicepresidente, se puede inferir que tanto Trump como los candidatos demócratas apelan a la conexión social y emocional con los votantes, mientras que Pence al tener las palabras más frecuentes como “president” y “trump” su discurso parece estar orientado a reforzar la figura de Trump al ser parte de la fórmula del partido republicano. Parte del eslogan de Trump es “Make America Great Again” y la palabra “great” fue una de las que más repitió, lo cual sigue con su narrativa durante la campaña.

Fig.4 Palabras más frecuente por candidato presidencial luego de aplicar función de eliminar palabras



En relación a las menciones a candidatos en los discursos, como se puede apreciar en la tabla N°2, el candidato con mayor número de automenciones fue Donald Trump con 481, seguido de Joe Biden con 163. De manera general los candidatos hicieron un número considerable de automenciones debido a que el tipo de discurso considerado es de “Campaign Speech”, lo que determina que la modalidad sean discursos referentes a sí mismos y planes de campaña de los candidatos.

Donald Trump junto con Joe Biden concentran el mayor número de referencias cruzadas respecto al resto de los candidatos, es decir, el candidato que más menciona Trump es Biden (806 menciones) y recíprocamente, el candidato que más menciona Biden es Trump (305 menciones). Esto se debe a una estrategia de confrontación directa entre los principales candidatos, siendo estos dos los candidatos a la presidencia. Si bien Trump parece centrarse más en su contrincante, la cantidad de menciones analizadas en relación a la cantidad total de palabras de estos candidatos es similar, ya

que Trump menciona a Biden aproximadamente 1 vez cada 532 palabras³ y Biden menciona a Trump 1 vez cada 599 palabras⁴.

Por otra parte, Mike Pence hizo mayor referencia a Trump con 490 menciones, seguido de Biden con 376. Esto se relaciona directamente con su rol como candidato a vicepresidente de la fórmula del partido republicano. Mientras que de forma similar Kamala Harris la mayor cantidad de menciones las realizó a su compañero de fórmula presidencial del partido demócrata. En el anexo 4, se incluye el grafo que permite visualizar las menciones.

Tabla N° 2: Cantidad de menciones por candidato

	Trump	Harris	Pence	Biden	Sanders
Trump	481	26	46	806	48
Harris	47	59	1	67	0
Pence	490	52	20	376	4
Biden	305	16	1	163	8
Sanders	86	1	0	37	6

Consideraciones finales:

En este informe se realizaron análisis preliminares a partir de los datos disponibles. Aún es posible profundizar en ciertos aspectos y continuar investigando algunas cuestiones.

En primer lugar, si bien se hizo un breve análisis de las palabras más frecuentes, quedan algunas preguntas interesantes a responder: ¿cómo varió el contenido de los discursos según la ubicación? ¿Hubo temas específicos priorizados en la campaña en las distintas localidades y estados? Para intentar responder estas preguntas se haría el análisis de la frecuencia de ciertas palabras vinculadas a temas importantes (seguridad, empleo, etc..) teniendo en cuenta la ubicación geográfica de los discursos. Siguiendo esa línea, y mirando el resultado de las elecciones, se podría focalizar en los estados que cambiaron y se inclinaron hacia Biden (Arizona, Georgia, Pensilvania, Wisconsin y Michigan) y analizar el contenido de los discursos de Biden en esos estados.

En segundo lugar, teniendo en cuenta que la campaña se realizó en 2020, enmarcada en el auge del COVID, cabe plantearse alguna pregunta relacionada a este tema, por ejemplo, ¿qué tanto se mencionó ese tema en los discursos? ¿alguno de los candidatos mencionó el tema más que otros?. Para responder estas preguntas se realizaría un conteo por palabras relacionadas a la temática tales como: covid, vacunas, salud, pandemia, virus etc., que permitan tener un panorama de la postura de cada candidato y su evolución en el tiempo (incremento o disminución del número de palabras).

Finalmente, otra pregunta relevante puede ser ahondar en otro tipo de discurso más interactivo, por ejemplo los town hall e investigar cuáles son los temas principales que surgen en los planteos de los

³ 806 menciones en un total de 429174 palabras.

⁴ 305 menciones en un total de 182773 palabras.

ciudadanos. Esto requeriría analizar el contenido de las palabras separando los discursos por las intervenciones.

ANEXOS

Anexo 1:

Orador	Cantidad de Discursos
Joe Biden	71
Donald Trump	53
Mike Pence	19
Bernie Sanders	16
Kamala Harris	11
Democratic Candidates	8
Multiple Speakers	5
Joe Biden, Kamala Harris	4
Elizabeth Warren	4
Pete Buttigieg	4

Anexo 2:

Cantidad total de palabras

Speaker	Total_Words
Donald Trump	429.174
Joe Biden	182.773
Mike Pence	84.924
Kamala Harris	43.840
Bernie Sanders	24.129

Anexo 3

Total discursos por Estado (Septiembre 2020)

Estado	Total Discursos
Pennsylvania	8

Ohio	5
Wisconsin	5
Florida	4
Minnesota	3
North Carolina	3
Arizona	2
Delaware	2
Michigan	2
Nevada	2
Georgia	1
MI	1
New Hampshire	1
Virginia	1

Anexo 4. Grafo de menciones

Grafo de menciones entre mandatarios en discursos de campaña - USA 2020

