



ANÁLISIS DE DISCURSOS ELECTORALES

Discursos de candidatos de Estados Unidos en 2020

Josefina Delgado, Mateo Bentancur
Introducción a la Ciencia de Datos, 2025

1.	Introducción.....	2
2.	Metodología.....	2
2.1.	Datos de partida	2
2.2.	Detección de problemas en la calidad de los datos	2
2.3.	Discursos por candidatos	2
2.4.	Normalización del texto y conteo de palabras.....	2
2.5.	Generación de matriz de menciones entre candidatos	3
3.	Resultados	3
3.1.	Datos de partida	3
3.2.	Detección de problemas de calidad de los datos	3
3.3.	Discursos por candidatos	4
3.4.	Normalización del texto y conteo de palabras.....	6
3.5.	Generación de matriz de menciones entre candidatos	8
4.	Reflexiones finales	9
5.	Bibliografía.....	10

1. Introducción

En las elecciones presidenciales de Estados Unidos en 2020, celebradas el 3 de noviembre, el candidato Joe Biden resultó electo, con 306 votos electorales frente a 232 que consiguió Donald Trump.

Según artículos de The New York Times, The Wall Street Journal y NPR, los temas de la campaña estuvieron centrados en el COVID-19, la muerte de dos afroamericanos durante manifestaciones, la Ley del Cuidado de Salud a Bajo Precio, el cambio climático, entre otros. [1] [2] [3]

Este trabajo práctico busca realizar un análisis acerca de los discursos de los candidatos estadounidenses en 2020, usando herramientas de ciencia de datos.

2. Metodología

2.1. Datos de partida

Se utilizó el *data frame* de los discursos dados por los candidatos durante la campaña electoral, este contiene el nombre del candidato, la fecha del discurso, el título, la transcripción completa, la ubicación donde se llevó a cabo y el tipo de evento o discurso.

El trabajo fue realizado en Google Colab, usando Python como lenguaje de programación. Finalmente, el trabajo se cargó en GitLab.

2.2. Detección de problemas en la calidad de los datos

Usando la función 'isna' se obtuvieron los datos faltantes. Utilizando la función 'unique' se obtienen datos únicos para realizar una inspección visual para encontrar inconsistencias.

2.3. Discursos por candidatos

Se listaron todos los candidatos que aparecen al menos una vez en la columna *speaker*. Para aquellos con múltiples candidatos, se utilizó la función 'split' para separar los nombres utilizando la coma como separador. Se creó una nueva fila para cada orador manteniendo el resto de la información.

Con el data frame actualizado se contó la cantidad de veces que aparece el orador en la columna *speaker*. Ordenando y obteniendo el top 5.

Para estos candidatos, se realizó un gráfico de barras con los discursos por mes a lo largo de la campaña, agregando ceros cuando no hay información de discursos del candidato en un mes. En esta instancia se evaluaron los momentos claves de la campaña.

2.4. Normalización del texto y conteo de palabras

Para contar las palabras, primero se normalizó el texto: se pasaron todas las letras a minúsculas, se eliminaron las marcas de tiempo, los nombres de los candidatos al comienzo de cada línea, los signos de puntuación y los espacios duplicados. Luego, se listaron las palabras más usadas y se hizo un gráfico de barras adaptado.

Para analizar los discursos, se identificaron los temas principales (como economía, salud, inmigración, etc.) y se listaron palabras clave para cada uno. Posteriormente, se procesaron los textos y se contó cuántas veces aparecía cada palabra clave en cada intervención.

Cuando un discurso era compartido por varios candidatos, se lo contó por separado para cada uno. Con estos datos, se calculó cuántas veces mencionó cada tema cada candidato. Para poder comparar los resultados, esos valores se pasaron a porcentajes sobre el total de menciones de cada candidato, y se mostraron en un mapa de calor. Así se puede ver fácilmente qué temas fueron más importantes para cada figura política.

2.5. Generación de matriz de menciones entre candidatos

Este análisis busca identificar con qué frecuencia los candidatos mencionan a sus oponentes en sus discursos. Para ello, se partió de una lista con los nombres de cinco figuras políticas clave, tanto en su forma completa como en la forma más común con la que suelen ser mencionados (por ejemplo, "Joe Biden" y "Biden").

Se recorrieron los discursos de cada candidato y, en cada uno, se contabilizó cuántas veces aparecía el nombre de otro candidato. No se contabilizaron las menciones a sí mismos.

Los resultados se almacenaron en una matriz en la que las filas representan a quien habla y las columnas a quien es mencionado. Cada celda muestra la cantidad de veces que un candidato hizo referencia a otro.

3. Resultados

3.1. Datos de partida

A continuación, se muestra el primer dato del data frame crudo (Tabla 1).

index	speaker	title	text	date	location	type
0	David Perdue	Georgia Sen. David Perdue Speech Transcript at Trump Rally. Mocks & Mispronounces Kamala Harris' Name	David Perdue: (00:01) How great is it to be back in Macon, Georgia with Donald Trump coming to Macon again? I want to thank Governor Kemp because he laid it out just now, but let me remind you of something that happened in 2002. Sunny just mentioned it. Because of you, we turned this thing around, and for the last 18 years, we have built a state that's now the number one state in the country in which to do business. David Perdue: (00:25) That did not happen by accident. Brian Kemp is the third governor in a row that understands how to do that. And thanks to you, that's going to continue. Today, we have an opportunity. Every generation has an opportunity like this. They have their moment of crisis. With Donald Trump, it's all about promises made and promises delivered. After eight years of Joe Biden, you hear Joe lately? Joe said the other day, he said, "Oh yeah, you guys were enjoying that economic Renaissance because of what we did with Barack Obama." David Perdue: (01:06) Now guys, I'm just a dumb business guy for right over that hill, right there, and I can tell you, that is a lie right out of the pit of hell. Donald Trump, after eight years of Barack Obama and Joe Biden giving us the lowest economic output in U.S. history, and we felt that right here in Georgia, Donald Trump gave us the best economic turnaround in U.S. history, y'all. David Perdue: (01:36) Seven and a half million new jobs, highest middle class income ever measured, the lowest unemployment for African Americans, Asian Americans, and Hispanics ever measured. After 800,000 people, under Barack Obama, fell into poverty, under Donald Trump, six and a half million people have lifted themselves out of poverty, y'all. Now that's America. David Perdue: (01:57) But it took Donald Trump, not the bureaucrats in Washington, not some career politician, but it took Donald Trump to build that wall. It took Donald Trump to move that embassy from Tel Aviv to Jerusalem. It took Donald Trump to name 200 new federal judges, including Neil Gorsuch and Brett Kavanaugh. And next week, we will confirm Amy Coney Barrett. Mark my words. David Perdue: (02:43) It's one thing to stand up here and cheer. It's one thing to stand up and remind everybody. This guy is presidential. He didn't happen by accident. How in the world, in our political system, could Donald J. Trump come on the scene in 2016 and do what he did? Tell me. But God's watching. He's watching what we're going to do right now. What's at stake is this, and I mean this: It's our time to stand up to the socialist movement that for the last 100 years has been trying to take over our beloved America. David Perdue: (03:27) It didn't just start in the '30s. It started long ago, but they believe they have it in their grasp now. They believe these CNN polls and CNBC polls. Oh, let me tell you, they don't know the big red poll right here. But this is what I believe is a stake. Mark my words. They believe this. That if they get the majority of the Senate and they win the White House, Joe Biden has already said, your taxes are going up, guys. David Perdue: (03:56) He's already said. You know those regulations that those mean old demonic Republicans, all those regulations you pulled back, we're going to give it back to you. You know that healthcare that we screwed up? We're going to do it again. But most importantly, guys, they want to slack the Supreme Court. Did you see that? They want to slack the Supreme Court. Now what President Trump has done is brought balance back to the Supreme Court. These aren't activists. These are constitutionalists. They're up there to protect you and me. David Perdue: (04:32) But the most insidious thing that Chuck Schumer and Joe Biden are trying to perpetrate, and Bernie and Elizabeth and Kamala, or what? Kamala or Kamala or Kamala mala mala, I don't know. Whatever. They're trying to perpetrate a lie, and that is that socialism, this radical socialism that they're trying to sell us, is better than our beloved democracy. We're not going to let that stand. David Perdue: (04:57) And here's how they're going to do it. You slack the courts, you change the voting rules in the Senate, and pretty soon you start adding democratic states. You get rid of the electoral college or change it like they can do, and pretty soon you're talking about a one-party state, guys. That failed in Russia in 1917, it failed in Cuba in 1959, Venezuela today. You can't find a country in the world where that has ever worked. David Perdue: (05:19) Now why in the world would we ever go down that road, particularly after the last 75 years in this country where God has blessed us so greatly that we can come out today and say what we want independent of what that media says? I need you to do one thing. Stand with me, stand with our President, and let's make damn sure that the road to socialism never runs through the state of Georgia. God bless you guys.	Oct 16, 2020	Macon, Georgia	Campaign Speech

Tabla 1: Primer dato del data frame.

3.2. Detección de problemas de calidad de los datos

Existen datos faltantes en algunas columnas (Tabla 2). Esto puede afectar el análisis, en particular cuando la cantidad de datos faltantes es significativa.

Columna	Cantidad de datos faltantes
speaker	3

title	0
text	0
date	0
location	18
type	21

Tabla 2: Datos faltantes

De la inspección visual se detectaron datos incompletos en la columna *speakers*, como ‘???’. Asimismo, aparecen datos inconsistentes, donde el código de agrupación de oradores no permite determinar exactamente los candidatos que participaron en el discurso, como ‘Multiple Speakers’, o ‘Democratic Candidates’. Estos errores en la calidad de los datos pueden afectar el conteo de discursos por candidatos realizado más adelante.

En la columna *location*, aparecen nuevamente inconsistencias en los datos. Algunos datos corresponden a cadenas de TV, como ‘Fox News’ o ‘ABC’, pero no queda claro si es un discurso realizado en el canal de forma presencial o virtual. Asimismo, existen locaciones más o menos específicas, como ‘Wisconsin’ y ‘Milwaukee, Wisconsin’, o ‘Darby, Pennsylvania’ y ‘Pennsylvania’. Estas inconsistencias podrían modificar los análisis sobre la campaña, como por ejemplo, el impacto de los discursos presenciales o el interés de realizar un discurso en algún sitio particular por el suceso de algún evento.

3.3. Discursos por candidatos

Los candidatos que lideran la cantidad de discursos son los más populares y los más votados (Tabla 3).

Candidatos	Cantidad de discursos
Joe Biden	81
Donald Trump	54
Mike Pence	20
Bernie Sanders	17
Kamala Harris	16

Tabla 3: Cantidad de discursos por candidato

Por otro lado, la campaña se intensifica en los 3 meses previos al 3 de noviembre (Figura 1).

El pico de discursos de Trump y Biden en septiembre coincide con el debate presidencial. Los dos principales candidatos a la vicepresidencia Harris y Pence debatieron en octubre, como se observa en la figura 1. [4]

Joe Biden tiene una presencia más constante en su campaña, manteniendo un compromiso constante. Donald Trump, por otro lado, tiene una estrategia que se concentra sobre el final de la campaña.

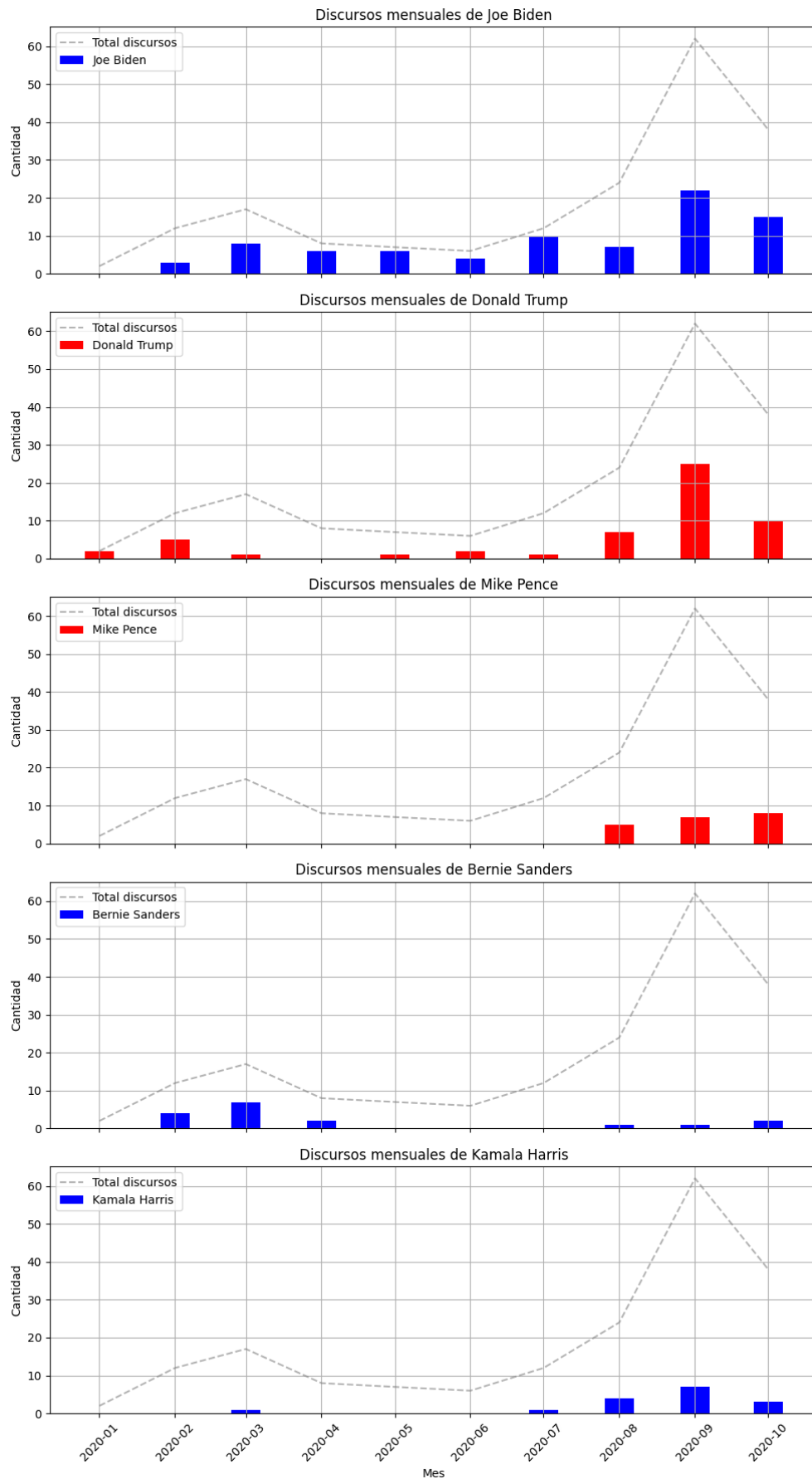


Figura 1: Discursos por candidatos a lo largo del tiempo

3.4. Normalización del texto y conteo de palabras

Al tratarse de una transcripción de un discurso hablado, es de esperarse que las palabras más frecuentes sean conjunciones y conectores. Esta tendencia es común entre los cinco candidatos analizados. (Figura 2)

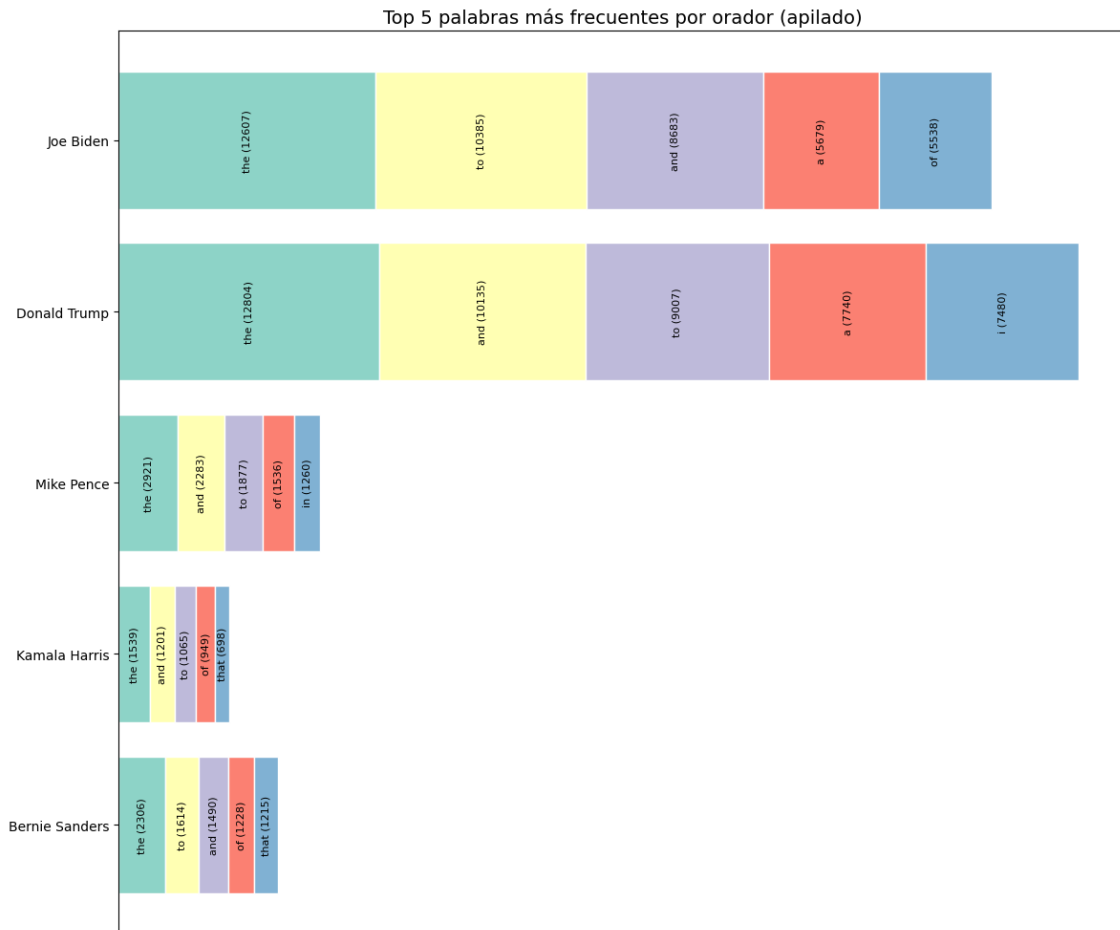


Figura 2: Cantidad de palabras por candidato

Dado que Joe Biden y Donald Trump cuentan con una mayor cantidad de discursos registrados, es esperable que también presenten un mayor volumen total de palabras. Sin embargo, se observa que, aunque Biden dio 27 discursos más que Trump, este último usó una mayor cantidad de palabras. Esta diferencia se explica porque, en promedio, los discursos de Trump son casi el doble de extensos, lo que naturalmente implica un uso más frecuente de conjunciones y conectores. En cuanto al resto de los candidatos, cuentan con una cantidad considerablemente menor de discursos, y presentan un volumen total de palabras significativamente más bajo. (Tabla 4)

Candidatos	Promedio de palabras por discurso
Donald Trump	6345
Joe Biden	3373
Kamala Harris	1962

Bernie Sanders	2820
Mike Pence	2809

Tabla 4: Características de los discursos por candidatos

Para realizar un análisis más enriquecedor acerca del enfoque que cada candidato da a sus discursos, se seleccionaron distintos ejes temáticos con palabras claves relacionadas a los mismos. Los ejes temáticos y sus palabras claves se muestran a continuación:

COVID: covid, coronavirus, pandemic, vaccine, vaccination, mask, quarantine, lockdown, health crisis.

Economía: economy, jobs, employment, tax, unemployment, business, market.

Justicia racial: race, racism, racial, justice, equality, black lives matter, blm, discrimination, civil rights, protest.

Salud: healthcare, insurance, obamacare, medicare, hospital, doctor, patient, public option.

Cambio climático: climate, environment, global warming, pollution, green energy, renewable, sustainability.

Inmigración: immigration, border, wall, deportation, visa, migrant, undocumented.

Seguridad pública: police, crime, violence, gun, shooting, safety.

Política exterior: china, russia, iran, north korea, allies, nato, trade war, treaty.

Corte Suprema: supreme court, justice, judge, constitution.

Derechos civiles: equal rights, lgbtq, women rights, freedom.

Educación: education, school, student, college, university, tuition, remote learning.

En la distribución porcentual de los temas mencionados por los candidatos se observa que la economía es el tema central de la campaña, lo cual tiene sentido considerando el peso que tiene Estados Unidos en el mercado global y el impacto que tienen las decisiones económicas del país. (Figura 3)

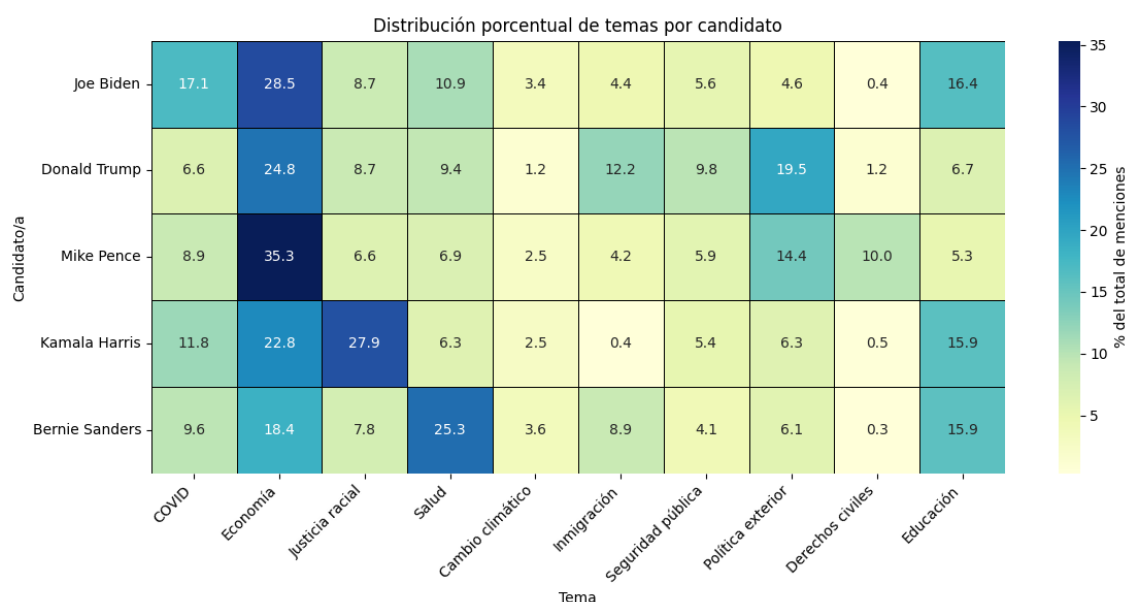


Figura 3: Distribución de temas por candidato

Los discursos de Joe Biden se centraron en la economía, la pandemia de COVID-19 y la educación. Kamala Harris abordó temas como la justicia social, la economía y la educación. Por su parte, Bernie Sanders hizo énfasis en la salud, la economía y la educación.

En cuanto a los republicanos, Donald Trump enfocó sus intervenciones en la economía, la política exterior y la inmigración, mientras que Mike Pence destacó la economía, la política exterior y los derechos civiles.

Los candidatos demócratas coinciden en su enfoque en la economía y la educación, temas alineados con su énfasis en la equidad social, mientras que los republicanos muestran un interés común en la economía y la política exterior, en línea con su visión de libre mercado y seguridad nacional.

A pesar de tener una gran industria, el cambio climático y la agenda ambientalista no parece ser de gran importancia para la sociedad estadounidense.

Es importante destacar que esta representación está basada en el conteo de palabras clave asociadas a cada eje temático. Esto introduce un sesgo en el análisis: la selección de términos puede dejar fuera sinónimos o formas indirectas de mencionar los temas. Además, no se considera el tono con el que se habla de cada asunto (si es positivo, negativo o neutro), ni la duración total de los discursos.

3.5. Generación de matriz de menciones entre candidatos

En la Tabla 5 se muestra el conteo de menciones de un candidato a otro. Las filas representan el candidato y las columnas a quien es mencionado. Como primera observación, se destaca que los candidatos presidenciales son las figuras más mencionadas en los discursos, siendo Donald Trump —quien ocupaba la presidencia en ese momento— el más nombrado. Por otro lado, se observa que los candidatos republicanos hacen referencia a los candidatos presidenciales del partido demócrata aproximadamente el doble de veces que lo contrario.

	Joe Biden	Donald Trump	Bernie Sanders	Kamala Harris	Mike Pence
--	-----------	--------------	----------------	---------------	------------

Joe Biden	0	492	10	25	4
Donald Trump	679	0	36	33	39
Bernie Sanders	62	139	0	1	0
Kamala Harris	78	74	0	0	7
Mike Pence	297	367	6	65	0

Tabla 5: Matriz de menciones entre candidatos

4. Reflexiones finales

Este trabajo muestra cómo la ciencia de datos puede ayudarnos a entender mejor qué están diciendo los políticos y hacia dónde apuntan sus discursos. Con herramientas simples, como contar palabras clave y visualizar los resultados, se puede tener una idea bastante clara de los temas más mencionados y las prioridades que se repiten.

Pero también es importante tener en cuenta que los datos no hablan por sí solos: todo análisis está influido por las decisiones que tomamos, como qué palabras elegir o cómo agrupar los temas. Eso quiere decir que, según cómo armemos el análisis, podemos llegar a conclusiones distintas.

A partir del análisis realizado, se podrían intentar responder algunas preguntas:

1. ¿Qué tono tienen las menciones de los ejes temáticos?

Siguiendo con el análisis realizado de los diversos ejes temáticos, sería muy interesante aplicar una técnica procesamiento de lenguaje natural a los fragmentos donde se mencionan las palabras claves, para conocer si el tono es positivo, negativo o neutro. Existen librerías de Python como *TextBlob* para puntuar el sentimiento del texto.

2. ¿Existe correlación entre la cantidad de discursos y la cantidad de electores en los diferentes estados?

Sería relevante explorar si existe una correlación entre la cantidad de discursos realizados en los distintos estados y la cantidad de electores que aporta cada uno en el Colegio Electoral. Para ello, se debería contar la cantidad de discursos por estado utilizando la columna *location*, y luego contrastarlo con los votos electorales asignados a cada estado

3. ¿Qué tipo de intervenciones utilizan los candidatos en su campaña?

Resulta interesante analizar el tipo de intervenciones realizadas por los candidatos, por ejemplo, discursos de campañas, entrevistas, entre otros. De modo de obtener información acerca de la estrategia de campaña. Esto puede lograrse contabilizando los registros según la columna *type*.

4. ¿Cuántas veces se nombran presidentes anteriores en los discursos?

Otro aspecto relevante es evaluar cuántas veces los oradores mencionan a expresidentes. Para esto, se puede construir una lista de nombres de expresidentes (por ejemplo, “Obama”, “Bush”, “Reagan”) y luego contar la cantidad de veces que aparecen en la transcripción de los discursos. Esto puede servir como un indicador de referencias a gestiones anteriores como estrategia discursiva.

5. Bibliografía

- [1] Armour, S., & Siddiqui, S. (2020, September 25). *Supreme Court vacancy makes health care leading issue in 2020 elections*. The Wall Street Journal. <https://www.wsj.com/politics/policy/supreme-court-vacancy-makes-health-care-leading-issue-in-2020-elections-11600948807>
- [2] Bouie, J. (2020, June 3). *Trump or Biden? George Floyd has shown us the answer*. The New York Times. <https://www.nytimes.com/2020/06/03/opinion/george-floyd-trump-biden.html>
- [3] Hersher, R. (2020, October 22). *There's a lot at stake for the climate in the 2020 election*. NPR. <https://www.npr.org/2020/10/22/925468716/theres-a-lot-at-stake-for-the-climate-in-the-2020-election>
- [4] Wikipedia contributors. (s.f.). *Debates presidenciales de Estados Unidos de 2020*. Wikipedia. https://es.wikipedia.org/wiki/Debates_presidenciales_de_Estados_Unidos_de_2020