



Facultad de Ciencias Económicas y Empresariales

# **Voces de la Diáspora: Analizando la Percepción Pública sobre la Migración Venezolana a través de las redes sociales**

Autor: Julia Jenaro Barrio

Director: Jenny Alexandra Cifuentes Quintero

MADRID | Junio 2025



# Resumen

La diáspora venezolana constituye uno de los procesos migratorios más complejos y de mayor impacto del siglo XXI, caracterizado por su magnitud y por una serie de causas estructurales interrelacionadas. Este fenómeno se enmarca en una crisis prolongada de naturaleza económica, política e institucional que ha deteriorado gravemente las condiciones de vida en el país. El análisis socioeconómico realizado en este trabajo revela que la histórica dependencia del petróleo, junto con una gestión ineficaz, sanciones internacionales y un entorno autoritario, ha dado lugar a un colapso sistémico que ha impulsado la salida de más de siete millones de personas, conformando una diáspora de alcance regional y global.

En este contexto, la percepción pública representa una dimensión analítica de gran relevancia para comprender las construcciones simbólicas y las actitudes sociales en torno al éxodo venezolano. La literatura especializada ha documentado el valor de las redes sociales como espacios dinámicos de producción discursiva, donde es posible capturar en tiempo real narrativas, emociones y posicionamientos colectivos. Entre las metodologías más empleadas destacan el modelado de tópicos mediante Latent Dirichlet Allocation (LDA) y el análisis de sentimientos a través de herramientas como VADER, que permiten identificar tanto los ejes temáticos predominantes como la polaridad afectiva de las publicaciones. A partir de este enfoque, el presente trabajo propone un análisis empírico de las percepciones ciudadanas sobre la migración venezolana en Reddit, integrando dichas técnicas con el fin de ofrecer evidencia complementaria para el estudio de la diáspora. Los resultados del modelado de tópicos y el análisis de sentimientos convergen en una representación predominantemente crítica del fenómeno migratorio venezolano en Reddit. A nivel discursivo, las publicaciones se articulan en torno a cuatro grandes ejes, con una atención destacada a las cuestiones legales y estructurales, mientras que los aspectos administrativos reciben una menor cobertura. Desde el plano emocional, el análisis de sentimientos destaca una orientación mayoritariamente negativa en las narrativas digitales, especialmente en lo relativo al control institucional y la gestión federal de la migración. No obstante, el uso de registros satíricos en el tópico sobre liderazgo populista introduce una dimensión humorística que matiza la carga afectiva, sugiriendo que la crítica política se canaliza también mediante recursos expresivos no convencionales.

**Palabras Clave:** migración venezolana, *Reddit*, percepción ciudadana, modelado de tópicos, análisis de sentimiento.

# Abstract

The Venezuelan diaspora represents one of the most complex and impactful migration processes of the 21st century, characterized by its magnitude and a set of interrelated structural causes. This phenomenon is framed within a prolonged crisis of economic, political, and institutional nature that has severely deteriorated living conditions in the country. The socioeconomic analysis carried out in this work reveals that the historic dependence on oil, combined with ineffective governance, international sanctions, and an authoritarian environment, has led to a systemic collapse that has driven the departure of over seven million people, forming a diaspora with both regional and global reach.

In this context, public perception becomes an analytically relevant dimension for understanding the symbolic constructions and social attitudes surrounding the Venezuelan exodus. Specialized literature has documented the value of social media as dynamic spaces for discursive production, where it is possible to capture in real time the narratives, emotions, and collective stances. Among the most widely used methodologies are topic modeling using Latent Dirichlet Allocation (LDA) and sentiment analysis through tools like VADER, which allow for the identification of both the dominant thematic axes and the emotional polarity of the posts.

From this perspective, the present study proposes an empirical analysis of citizen perceptions of Venezuelan migration on Reddit, integrating these techniques to provide complementary evidence for the study of the diaspora. The results of the topic modeling and sentiment analysis converge in a predominantly critical representation of the Venezuelan migratory phenomenon on Reddit. Discursively, the posts are structured around four main axes, with a notable focus on legal and structural issues, while administrative aspects receive less attention. Emotionally, the sentiment analysis highlights a predominantly negative tone in digital narratives, especially regarding institutional control and federal migration management. However, the use of satirical registers in the topic on populist leadership introduces a humorous dimension that softens the emotional charge, suggesting that political criticism is also expressed through unconventional expressive means.

**Keywords:** Venezuelan migration, *Reddit*, perception, topic modeling, sentiment analysis.

# Agradecimientos

Quiero expresar mi más sincero agradecimiento a mi tutora, Jenny Alexandra Cifuentes Quintero, por su constante apoyo, orientación y por toda la comunicación a pesar de las 6 horas de diferencia horaria que nos separaban, al yo estar parte de este proceso viviendo en Estados Unidos.

Su acompañamiento y constancia han sido fundamentales para el desarrollo y la finalización de este trabajo.

Quiero agradecer también a todas las personas que conocí durante mi estancia en la frontera entre México y Estados Unidos. Gracias por enseñarme el valor de la gratitud, por su luz, sus historias y sus enseñanzas. Gracias por migrar, por buscar un futuro mejor para sus familias. Ojalá llegue el día en que nadie se vea obligado a dejar su hogar. Ustedes me inspiraron a llevar a cabo este trabajo.

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Objetivos . . . . .	4
1.2.1. Objetivo General . . . . .	4
1.2.2. Objetivos Específicos . . . . .	4
1.3. Estructura del documento . . . . .	5
<b>2. El Contexto Migratorio de Venezuela y el Uso del Machine Learning para su Análisis en Redes Sociales</b>	<b>6</b>
2.1. El contexto socioeconómico de Venezuela como origen de la migración masiva	6
2.2. Flujos migratorios en Venezuela: patrones y destinos . . . . .	15
2.2.1. Diáspora venezolana en Colombia: puerta de entrada y refugio . . . . .	16
2.2.2. Diáspora venezolana en Brasil: refugio en la frontera amazónica . . . . .	17
2.2.3. Diáspora venezolana en México: país de paso hacia el norte . . . . .	20
2.2.4. Diáspora venezolana en Estados Unidos: el sueño económico de los venezolanos . . . . .	23
2.2.5. Diáspora venezolana en España: el hogar culturalmente familiar cruzando el Atlántico . . . . .	25
2.3. Técnicas de Machine Learning para el análisis de redes sociales en estudios migratorios . . . . .	26
<b>3. Diseño Metodológico para el Análisis de Percepción Pública en Redes Sociales</b>	<b>33</b>
3.1. Recolección y preparación de los datos . . . . .	34
3.2. Pre-procesamiento del texto . . . . .	35
3.3. Análisis de N-Gramas . . . . .	36
3.4. Modelado de tópicos . . . . .	37
3.5. Análisis de sentimientos . . . . .	40
<b>4. Resultados</b>	<b>42</b>
4.1. Adquisición y preparación de los datos . . . . .	42
4.2. Pre-procesamiento de los datos . . . . .	48

4.3. Análisis de <i>N-Gramas</i> . . . . .	49
4.4. Modelado de tópicos . . . . .	52
4.5. Análisis de sentimientos . . . . .	60
<b>5. Conclusiones</b>	<b>64</b>
<b>Bibliografía</b>	<b>70</b>

# Índice de figuras

1.1. Diáspora Venezolana por país . . . . .	2
2.1. PIB per cápita . . . . .	9
2.2. Índice de pobreza en Venezuela por Estado (como porcentaje de la población)	10
2.3. Índice de desarrollo humano (0-1) . . . . .	11
2.4. Causas de muerte en Venezuela (2021) . . . . .	12
2.5. Índice de democracia (0-10) . . . . .	13
2.6. Cambio del Bolívar Venezolano frente al Dólar . . . . .	13
2.7. Fuentes de energía en Venezuela . . . . .	15
2.8. Población Venezolana en Colombia . . . . .	17
2.9. Venezolanos en Brasil en 2022 y 2023 . . . . .	18
2.10. Venezolanos que tienen la tarjeta de residencia en México (2024) . . . . .	21
2.11. Inmigrantes venezolanos en México, 2024 . . . . .	22
2.12. Venezolanos en Estados Unidos (2000-2021) . . . . .	23
2.13. Migración Venezolana en Estados Unidos (2023) . . . . .	24
3.1. Metodología de análisis . . . . .	33
3.2. Representación gráfica del modelo LDA . . . . .	38
4.1. Comentarios diarios: Septiembre 2024-Febrero 2025 . . . . .	44
4.2. Número de comentarios por idioma . . . . .	45
4.3. Distribución de la longitud de los comentarios . . . . .	45
4.4. Top 10 Subreddits con más comentarios . . . . .	46
4.5. Top 10 Hashtags más utilizados . . . . .	47
4.6. Nube de palabras . . . . .	48
4.7. Unigramas con Mayor TF-IDF . . . . .	50
4.8. Bigramas con Mayor TF-IDF . . . . .	51
4.9. Trigramas con Mayor TF-IDF . . . . .	52
4.10. Coherencia por Número de Tópicos . . . . .	53
4.11. Distancia intertópica, modelo K=4 . . . . .	54
4.12. Distribución de los tópicos en el corpus textual . . . . .	59
4.13. Distribución de la puntuación compuesta . . . . .	60

4.14. Número de publicaciones por puntuación . . . . .	61
4.15. Distribución del sentimiento por tópico . . . . .	63

# Índice de tablas

2.1. Artículos representativos en el análisis de percepción pública sobre inmigrantes usando técnicas de analítica de textos. . . . .	32
4.1. Tópicos, palabras clave, bigramas y trigramas más relevantes de los tópicos seleccionados. . . . .	55
4.2. Estadísticos descriptivos por tema dominante . . . . .	62

# Acrónimos

<i>API</i>	Application Program Interface
<i>AWS</i>	Amazon Web Services
<i>BCV</i>	Banco Central de Venezuela
<i>BD</i>	Big Data
<i>BERT</i>	Bidirectional Encoder Representations from Transformers
<i>BoW</i>	Bag of Words
<i>CBP One</i>	Customs and Border Protection One
<i>CEPAL</i>	Comisión Económica para América Latina y el Caribe
<i>CENDES</i>	Centro de Estudios de Desarrollo
<i>ChatGPT</i>	Chat Generative Pre-trained Transformer
<i>Conare</i>	Comité Nacional para los Refugiados
<i>CRISPR-</i>	Clustered Regularly Interspaced Short Palindromic Repeats
<i>Cas9</i>	
<i>DACA</i>	Deferred Action for Childhood Arrivals
<i>DDHH</i>	Derechos Humanos
<i>DHS</i>	Departamento de Seguridad Nacional
<i>EEUU</i>	Estados Unidos
<i>EmoLex</i>	National Research Council Canada Word-Emotion Association Lexicon
<i>EIU</i>	Economist Intelligence Unit
<i>ELMICT</i>	Ensemble Learning for Multilingual
<i>ENCOVI</i>	Encuesta Nacional de Condiciones de Vida
<i>ETPV</i>	Estatuto Temporal de Protección para Migrantes Venezolanos
<i>FMI</i>	Fondo Monetario Internacional
<i>GA</i>	Google Assistant
<i>GDELT</i>	Global Database of Events, Language, and Tone
<i>IA</i>	Inteligencia Artificial
<i>IAG</i>	Inteligencia Artificial Generativa
<i>IDF</i>	Inverse Document Frequency
<i>IDH</i>	Índice de Desarrollo Humano
<i>IS</i>	Ingeniería del Software
<i>KDT</i>	Knowledge Discory in Text

<i>LDA</i>	Latent Dirichlet Allocation
<i>LLM</i>	Large Language Models
<i>LSA</i>	Latent Semantic Analysis
<i>MALLET</i>	Machine Learning for Language Toolkit
<i>ML</i>	Machine Learning
<i>NB</i>	Naïve Bayes
<i>NFT</i>	Activos Digitales no Fungibles
<i>NLP</i>	Procesamiento del Lenguaje Natural
<i>NPMI</i>	Información Mutua Normalizada Puntual
<i>NRC</i>	National Research Council
<i>OAE</i>	Organización de los Estados Americanos
<i>OIM</i>	Organización Internacional para las Migraciones
<i>OPI</i>	Observatorio Permanente de la Inmigración
<i>OPS</i>	Organización Panamericana de la Salud
<i>OTAN</i>	Organización del Tratado del Atlántico Norte
<i>PCA</i>	Principal Component Analysis
<i>PDVSA</i>	Petróleos de Venezuela Sociedad Anónima
<i>PIB</i>	Producto Interior Bruto
<i>PLN</i>	Procesamiento del Lenguaje Natural
<i>PP</i>	Partido Popular
<i>RA</i>	Realidad Aumentada
<i>RRSS</i>	Redes Sociales
<i>RSC</i>	Responsabilidad Social Corporativa
<i>RV</i>	Realidad Virtual
<i>SMOTE</i>	Synthetic Minority Over-Sampling Technique
<i>SVM</i>	Support Vector Machine
<i>TF</i>	Term Frequency
<i>TF-IDF</i>	Term Frequency - Inverse Document Frequency
<i>TFG</i>	Trabajo de Fin de Grado
<i>TPS</i>	Temporary Protected Status
<i>t-SNE</i>	Stochastic Neighbor Embedding

*TS-NMF* Topic-Supervised Non-Negative Matrix  
*TVRH* Tarjeta de Visitante por Razones Humanitarias  
*UCAB* Universidad Católica Andrés Bello  
*UCV* Universidad Central de Venezuela  
*UNICEF* Fondo de las Naciones Unidas para la Infancia  
*VADER* Valence Aware Dictionary for Sentiment Reasoning

# Capítulo 1

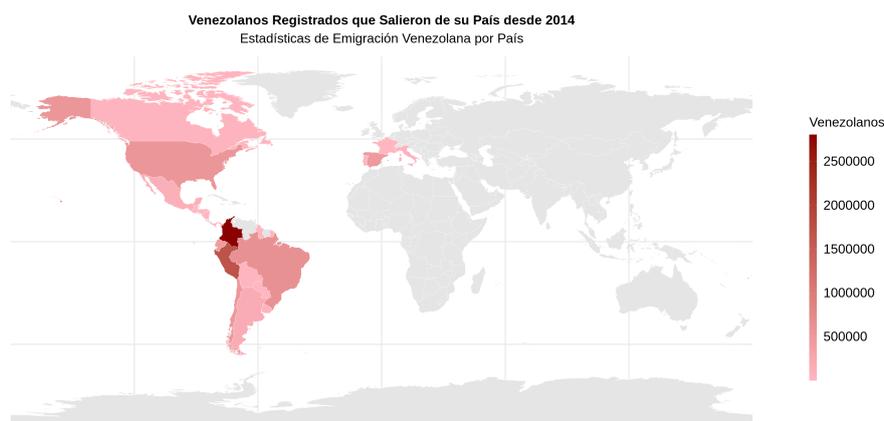
## Introducción

### 1.1. Motivación

La diáspora venezolana trasciende la dimensión demográfica y representa una historia de vidas que han sido forzadas a abandonar sus hogares en busca de un futuro más seguro y de dignidad humana. Cada migrante no solo lleva consigo el peso de la experiencia traumática de la emigración, sino también sus aspiraciones y el anhelo de ser tratado con el respeto que merece. En un contexto global donde la migración es frecuentemente abordada desde una perspectiva política, es relevante recordar que detrás de cada cifra hay una historia y una persona que desea ser comprendida con empatía. Humanizar el fenómeno migratorio implica considerar aspectos más allá de los desafíos sociales y económicos que enfrentan los países receptores y reconocer los procesos personales y emocionales de adaptación que atraviesan los migrantes. Este fenómeno no solo impacta a quienes dejan su país de origen, sino también a las generaciones futuras que nacen y crecen en naciones extranjeras. Estas nuevas generaciones enfrentan el reto de integrar las identidades culturales de sus países de residencia con las que heredan de sus progenitores, planteando así complejas dinámicas sociales que requieren atención en los estudios de movilidad humana.

La migración venezolana, además, se ha convertido en uno de los fenómenos económicos y sociales más significativos de las últimas décadas en América Latina. Este éxodo masivo representa el mayor desafío en términos de movilidad humana para la comunidad iberoamericana, sin precedentes en la historia reciente de la región (Eguren and Estrada, 2019). Según datos de la Agencia de Naciones Unidas para los Refugiados (ACNUR) más de 7,7 millones de venezolanos han dejado su país, de los cuales más de 1,2 millones han solicitado asilo y alrededor de 347.000 han sido reconocidos oficialmente como refugiados (ACNUR España, 2023). Este movimiento migratorio ha generado desafíos profundos para los países receptores, no solo en términos de integración laboral, social y cultural, sino también con respecto a la presión que la inmigración ejerce sobre los sistemas de salud, educación y empleo.

Como se observa en la Figura 1.1, Colombia es el principal país receptor de migrantes



**Figura 1.1: Diáspora Venezolana por país**  
Fuente de datos: (Statista, 2023). Elaboración propia.

venezolanos, acogiendo a aproximadamente 2.875.743 personas, lo que representa cerca del 37 % de la diáspora venezolana. En respuesta a esta situación, Colombia ha implementado políticas como el ‘Estatuto Temporal de Protección para Migrantes Venezolanos’ (ETPV), que busca garantizar a los migrantes acceso a derechos fundamentales regularizados, como la educación, el empleo y la atención sanitaria (Gobierno de Colombia, 2024). Perú, por su parte, es el segundo país en volumen de inmigrantes venezolanos, con más de 1.542.004 personas acogidas. El gobierno peruano estableció el ‘Permiso Temporal de Permanencia’ (PTP) para regularizar la situación de los migrantes; sin embargo, enfrenta grandes desafíos debido al aumento del desempleo y la limitada inversión en infraestructura, lo cual también ha impulsado la emigración de ciudadanos peruanos (Gobierno de Perú, 2024).

En tercer lugar, Estados Unidos ha recibido a aproximadamente 545.200 inmigrantes venezolanos. Su política migratoria incluye el ‘Estatus de Protección Temporal’ (TPS), implementado en 2023 para otorgar protección temporal a los venezolanos y evitar su deportación; sin embargo, los trámites burocráticos asociados a este programa resultan lentos y dificultosos para los migrantes de esta nacionalidad, limitando su acceso rápido a derechos laborales y sociales fundamentales (U.S. Citizenship and Immigration Services, 2024). Actualmente, desde el inicio de la presidencia de Donald Trump esta política ha sido prorrogada, con la intención de suspenderla exclusivamente a migrantes venezolanos. En América Latina, otros países destacados como Ecuador, Chile y Brasil también han visto un incremento considerable en la llegada de migrantes venezolanos, quienes enfrentan tanto oportunidades como desafíos de integración en sus nuevas comunidades. En Europa, España se ha convertido en el principal destino para los venezolanos, en gran parte debido a la proximidad cultural e idiomática. Aunque un número significativo de migrantes solicita asilo en España, las tasas de aprobación son bajas, y muchos reciben permisos de residencia en lugar del estatus de refugiado, lo cual limita el acceso a ciertos beneficios sociales. Por ejemplo, en 2023, de las

más de 60.500 solicitudes de venezolanos en España, 40.674 obtuvieron la residencia temporal como protección humanitaria, pero muy pocos lograron el reconocimiento de asilo formal (Diario, 2024).

La migración venezolana plantea complejos desafíos en los países de acogida, no solo en términos de capacidad logística y administrativa, sino también en lo referente a la cohesión social. En este sentido, comprender la percepción pública sobre los migrantes resulta de gran relevancia para promover políticas de integración eficaces y sostenibles que mejoren las condiciones de los migrantes y la convivencia en las comunidades receptoras. La opinión pública y las actitudes de la sociedad hacia los migrantes venezolanos juegan un papel muy importante, ya que estas percepciones influyen en la aceptación y disposición de los ciudadanos locales para facilitar o dificultar el proceso de integración. La integración no depende exclusivamente de la disposición de los migrantes para adaptarse, sino también de la manera en que son percibidos y aceptados en sus nuevos entornos, lo que impacta directamente su capacidad para participar activamente en la vida social, económica y cultural del país receptor (González-Rábago, 2014).

En este contexto, las redes sociales, al ser espacios de libre expresión y participación, se han convertido en una fuente valiosa para analizar en tiempo real las percepciones sobre fenómenos migratorios. Estas plataformas reflejan actitudes predominantes y opiniones diversas de múltiples grupos sociales y geográficos, ofreciendo una perspectiva integral sobre la opinión pública (Eguren and Estrada, 2019). A través de los comentarios y debates generados en redes sociales, es posible observar tanto expresiones de apoyo y empatía hacia los migrantes como manifestaciones de rechazo y discriminación. Estos discursos reflejan el clima social en torno a la inmigración y permiten identificar tanto factores que facilitan la integración como aquellos que representan barreras. Además, los comentarios en redes no solo expresan las opiniones populares; también pueden llegar a influir en la creación de políticas públicas, ya que los debates en línea a menudo impactan el discurso político y las respuestas gubernamentales ante la migración (Freire-Vidal, 2019).

Dentro de las diversas redes sociales, Reddit representa una plataforma particularmente útil para el análisis de estas percepciones debido a su estructura organizada en comunidades temáticas (subreddits), donde los usuarios pueden debatir abiertamente sobre temas específicos, incluida la migración (Reddit, 2024). A diferencia de otras redes sociales que tienden a favorecer interacciones rápidas y breves, Reddit permite el desarrollo de discusiones detalladas, lo cual facilita la obtención de opiniones más elaboradas y contextuales. La diáspora venezolana es un tema recurrente en estas comunidades, lo que convierte a Reddit en un recurso valioso para explorar las percepciones y dinámicas sociales que se desarrollan en torno a este fenómeno migratorio (Eguren and Estrada, 2019). El análisis de estas discusiones podría revelar actitudes polarizadas hacia la migración, reflejando tanto la empatía como los prejuicios hacia los inmigrantes venezolanos y, en última instancia, contribuyendo a entender la influencia del sentimiento público en las políticas y prácticas de integración.

Con este fin, se han propuesto diversas técnicas de procesamiento del lenguaje natural (NLP) para analizar las percepciones públicas sobre la migración en redes sociales. Entre estas técnicas, destacan el análisis de sentimiento, el modelado de tópicos y el análisis de redes semánticas (Gualda and Rebollo, 2020; Mantilla Muñoz and Velandia Ochoa, 2023). El análisis de sentimiento permite identificar las emociones y actitudes predominantes en los textos, clasificándolos como positivos, negativos o neutros, lo que facilita una comprensión rápida del tono general de la conversación pública. Por otro lado, el modelado de tópicos es útil para descubrir los temas principales que emergen en las discusiones, mostrando los aspectos de la migración que generan mayor interés o controversia. Además, el análisis de redes semánticas permite mapear las relaciones entre conceptos importantes en el discurso, mostrando la conexión de ciertos términos o ideas en el debate sobre la migración. Estas técnicas ofrecen una forma eficiente de procesar grandes volúmenes de datos textuales, extrayendo patrones y tendencias que serían difíciles de detectar manualmente, y resultan actualmente necesarios para entender mejor las dinámicas sociales y las percepciones colectivas en torno a la inmigración. Dado el impacto que las percepciones públicas tienen en el proceso de integración de los migrantes y la relevancia de las redes sociales como espacios donde estas actitudes se expresan, el presente Trabajo de Fin de Grado se enmarca en la aplicación de técnicas de NLP para analizar las discusiones sobre la migración venezolana en Reddit. Con un enfoque en el modelado de tópicos y el análisis de sentimientos, este estudio tiene como objetivo fundamental identificar patrones, temas recurrentes y actitudes predominantes en las conversaciones sobre la diáspora venezolana, proporcionando información relevante de las dinámicas sociales que rodean este fenómeno migratorio.

## **1.2. Objetivos**

### **1.2.1. Objetivo General**

El presente Trabajo de Fin de Grado en Business Analytics tiene como objetivo principal analizar las percepciones y actitudes públicas sobre la inmigración venezolana en el contexto de la diáspora, a través del análisis de contenido de blogs y discusiones publicadas en redes sociales, principalmente en Reddit, utilizando técnicas de procesamiento del lenguaje natural (NLP), como el modelado de tópicos y el análisis de sentimiento para identificar tendencias, temas recurrentes y opiniones predominantes.

### **1.2.2. Objetivos Específicos**

El Objetivo General se puede desglosar en objetivos específicos que permiten una visión más concreta del estudio a realizar y, que se presentan a continuación:

- Contextualizar la problemática social de la migración venezolana en el contexto de su diáspora, mediante el análisis descriptivo y estadístico de fuentes de datos disponibles, tales como informes gubernamentales, estudios de organizaciones internacionales y bases de datos públicas, para ofrecer un marco de referencia que complemente el análisis de la percepción pública en redes sociales.
- Identificar y analizar las técnicas más utilizadas para el análisis automático de textos en redes sociales (principalmente en Reddit), con un enfoque en estrategias de NLP, mediante una revisión de la literatura especializada en este campo.
- Realizar un análisis de la percepción pública sobre la inmigración venezolana a través de Reddit, utilizando técnicas de NLP como el modelado de tópicos y el análisis de sentimiento, con el fin de proporcionar información integral y relevante sobre la opinión pública en torno a la diáspora venezolana.

### **1.3. Estructura del documento**

El presente Trabajo de Fin de Grado se organiza en cinco capítulos principales. En el capítulo 1, se presenta una introducción que expone la motivación para el análisis de la percepción pública de la migración venezolana y detalla los objetivos fundamentales del estudio. El capítulo 2 contextualiza la situación sociopolítica y económica de Venezuela en la última década, destacando los factores que han impulsado la migración masiva de ciudadanos venezolanos hacia otros países de la región y del mundo. Asimismo, se examina el fenómeno de la diáspora venezolana con especial atención a los vínculos que los migrantes mantienen con su país de origen y las implicaciones de esta migración para las comunidades receptoras. En este capítulo también se revisa la bibliografía relevante sobre el análisis de la diáspora venezolana en estudios previos, así como estudios que aplican técnicas de análisis de redes sociales, minería de datos y métodos de NLP para evaluar la percepción pública de la migración en plataformas digitales.

En el capítulo 3, por su parte, se describe detalladamente la metodología del estudio, incluyendo los procesos de minería de datos y la obtención y preprocesamiento de los datos extraídos de Reddit. Además, se explican las técnicas de NLP empleadas, como el modelado de tópicos y el análisis de sentimiento, que permiten identificar temas recurrentes y evaluar las actitudes expresadas en torno a la migración venezolana. El capítulo 4 presenta los resultados obtenidos a partir de estas técnicas. Finalmente, el capítulo 5 ofrece un resumen de los aspectos más relevantes del análisis y expone las conclusiones alcanzadas.

## Capítulo 2

# El Contexto Migratorio de Venezuela y el Uso del Machine Learning para su Análisis en Redes Sociales

### 2.1. El contexto socioeconómico de Venezuela como origen de la migración masiva

Venezuela, situada en el norte de Sudamérica y limitando con Colombia, Brasil y Guyana, se distingue por su extraordinaria riqueza natural y biodiversidad. Aproximadamente el 15 % de su territorio está constituido por parques nacionales protegidos, los cuales albergan algunos de los paisajes más emblemáticos del continente. Entre ellos destaca el Salto Ángel, la catarata más alta del mundo, y el Lago de Maracaibo, conocido por ser el escenario del ‘Relámpago del Catatumbo’, un fenómeno meteorológico único que genera cerca de 1,6 millones de rayos al año. Este extraordinario espectáculo, además de su relevancia cultural, constituye un importante atractivo turístico para la región. Adicionalmente, Venezuela es reconocida por su cacao de alta calidad, considerado uno de los mejores a nivel mundial gracias a la combinación de sus condiciones climáticas tropicales y la fertilidad de sus suelos (AZERTAC, 2024).

El territorio venezolano también destaca por su excepcional riqueza en recursos minerales. Entre los metales industriales más relevantes se encuentran el hierro, oro, níquel, bauxita y cobre, además de minerales estratégicos como el coltán y el titanio. Estos recursos desempeñan un papel crucial en sectores clave como la manufactura y la construcción, apoyados por la disponibilidad de minerales no metálicos como arcillas, azufre y yeso. La actividad minera del país se concentra en el Arco Minero del Orinoco, una vasta región ubicada en el estado Bolívar que, aunque representa un importante potencial económico, ha generado controversias debido a sus impactos ambientales y sociales (SOS Orinoco, 2024). Además,

Venezuela cuenta con yacimientos significativos de piedras preciosas, como diamantes y cuarzo. No obstante, el mayor recurso económico del país es el petróleo. Con reservas probadas de 300.900 millones de barriles, Venezuela se posiciona como el mayor productor mundial, consolidando al crudo como el eje central de su economía (Agüero, 2012).

Desde mediados del siglo XX, el petróleo ha sido el pilar fundamental de la economía venezolana, estableciéndose como la principal fuente de ingresos del país. Durante décadas, la economía dependió casi exclusivamente de las exportaciones petroleras, lo que permitió un crecimiento sostenido en periodos de altos precios internacionales. Sin embargo, esta dependencia también evidenció la vulnerabilidad del modelo económico frente a fluctuaciones en el mercado global. En los últimos años, este panorama se ha agravado debido a la desinversión en infraestructura, altos niveles de corrupción administrativa y las sanciones internacionales impuestas, principalmente por Estados Unidos desde 2018 (Agüero, 2012). La producción petrolera, que alcanzó un máximo de más de 3 millones de barriles diarios en 2008, ha sufrido una drástica disminución como resultado del deterioro de la infraestructura energética y las restricciones impuestas por dichas sanciones (Lugo, 2024).

En 2019, las sanciones intensificaron su impacto al limitar severamente las exportaciones de crudo y congelar los activos de la estatal Petróleos de Venezuela S.A. (PDVSA), profundizando aún más la crisis económica. En 2022, la administración de Joe Biden permitió a empresas como Chevron retomar operaciones bajo licencias específicas, condicionando estas a avances políticos dentro de Venezuela (Abrams, 2024). Este renovado interés estadounidense en el petróleo venezolano refleja no solo la importancia estratégica de este recurso, sino también la necesidad de diversificar fuentes energéticas en un contexto de cambios en las dinámicas geopolíticas globales (Harper and Koh, 2024). Sin embargo, la falta de transparencia en los recientes procesos electorales en Venezuela y las tensiones políticas en Estados Unidos, especialmente con la reelección de Donald Trump, podrían conducir a la implementación de nuevas sanciones más estrictas, perpetuando las dificultades económicas del país (Harper and Koh, 2024)

Aunque las sanciones han tenido un impacto devastador, la flexibilización parcial en 2022 permitió una relativa estabilización del mercado petrolero venezolano. Esto contribuyó a reducir la presión inflacionaria, aunque los niveles de inflación permanecieron elevados, con una tasa interanual del 155,8%, según datos del Fondo Monetario Internacional (FMI) y el Banco Mundial (Harper and Koh, 2024). Este nivel, sin embargo, representa una mejora significativa en comparación con 2018, cuando la hiperinflación superó el 1.000.000%. Factores como la dolarización parcial, el menor crecimiento de los agregados monetarios y una disminución en la depreciación del bolívar (92% en 2022 frente al 315% en 2021) ayudaron a mitigar la crisis económica. Además, los ingresos fiscales provenientes del petróleo aumentaron gracias al incremento en los volúmenes de exportación y a los precios internacionales más altos. Sin embargo, la recaudación tributaria no petrolera disminuyó un 31,5% en términos reales respecto al año anterior, reflejando la fragilidad estructural de las finanzas públicas

(Melvin et al., 2024). El impacto positivo de estas medidas se reflejó en un crecimiento del Producto Interno Bruto (PIB) de Venezuela del 12 % en 2022, impulsado principalmente por la recuperación parcial de la producción petrolera y los precios internacionales favorables. No obstante, este crecimiento económico no se tradujo en una mejora significativa en el PIB per cápita, que sigue siendo extremadamente bajo en comparación con los niveles previos a la crisis. Las exportaciones petroleras, aunque incrementadas respecto a años anteriores, apenas alcanzaron el 30 % de los niveles registrados en 2014, lo que evidencia la magnitud de la contracción económica acumulada en la última década (Kitroeff and Kurmanaev, 2022).

El crecimiento económico del 12 % del PIB en 2022, asociado a la flexibilización de sanciones y la recuperación parcial de la producción petrolera, no se refleja en el PIB per cápita, que continúa siendo extremadamente bajo en comparación con los niveles previos a la crisis (Observatorio Gasto Público Cedice, 2022). La Figura 2.1 ilustra esta evolución, mostrando una caída significativa en el PIB per cápita desde 2013, tras la entrada de Maduro al poder que conllevó la intensificación de la crisis económica. Aunque ha habido una leve estabilización en los últimos años, los niveles actuales son significativamente inferiores a los registrados antes de 2014, año en el que Maduro asumió la presidencia y se agudizó la recesión económica. Esta caída en el PIB per cápita, un indicador clave de prosperidad, refleja un deterioro profundo en el nivel de vida de la población, que sigue enfrentando limitaciones severas para acceder a bienes esenciales como alimentos, salud y educación (Naciones Unidas, 2022). El declive económico también ha impactado gravemente el mercado laboral, con una disminución de los salarios reales, un aumento del desempleo y un incremento del subempleo. Según datos oficiales, en el primer semestre de 2022 se registraron 3.892 protestas relacionadas con salarios y condiciones laborales (Bermúdez, 2022). Este contexto, marcado por la pérdida generalizada del poder adquisitivo y el colapso de servicios básicos, ha sido un factor determinante en la migración masiva de venezolanos, quienes, desde 2013, han abandonado el país en busca de mejores oportunidades de vida.

El impacto del deterioro económico en Venezuela se manifiesta de manera especialmente crítica en el incremento sostenido de los niveles de pobreza general y extrema desde el año 2013. Tal como se presenta en la Figura 2.2, en 2021 un total de 15 estados del país registraron tasas de pobreza extrema superiores al 75 %, lo que evidencia la magnitud y extensión territorial de esta problemática. Entre las regiones más afectadas destacan el estado Amazonas, con un 87 % de su población en situación de pobreza extrema, seguido de Monagas (86 %), Falcón (84 %) y Sucre (83 %). Estos elevados porcentajes ponen de relieve una crisis estructural que impacta tanto a las zonas rurales como a las urbanas, limitando la capacidad de amplios sectores de la población para acceder a condiciones de vida mínimamente adecuadas. Según datos del Centro de Estudios de Desarrollo (CENDES) de la Universidad Central de Venezuela (UCV), en 2014 el 48 % de la población venezolana se encontraba por debajo de la línea de pobreza; no obstante, para 2022 esta proporción superó el 90 %, reflejando un deterioro social sin precedentes en la historia reciente del país (Naciones Unidas, 2022).

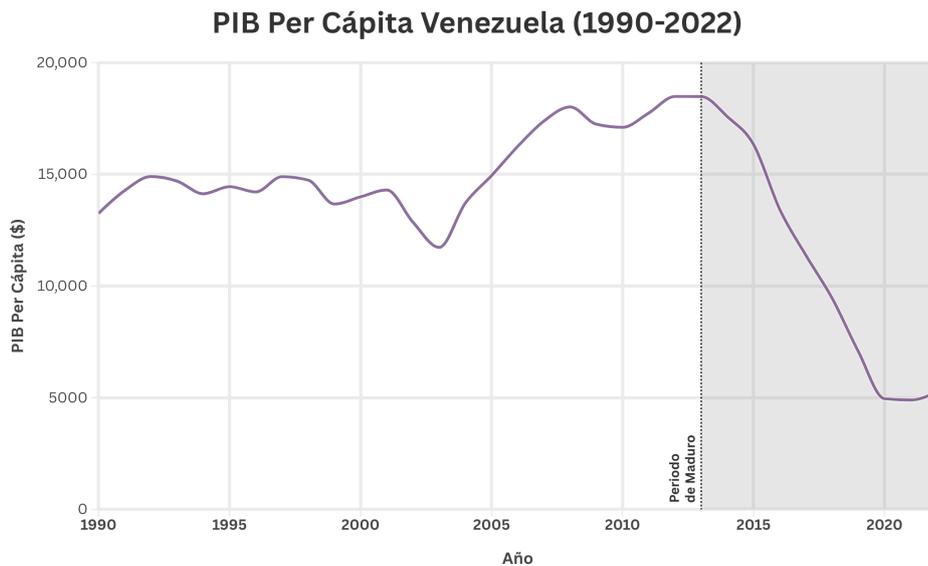


Figura 2.1: PIB per cápita  
Elaboración propia. Fuente de Datos: (Our World in Data, 2024c)

De esta población, aproximadamente el 60 % se encuentra en condiciones de pobreza extrema, definida como aquellos hogares cuyos ingresos diarios son inferiores a un dólar, de acuerdo con las estimaciones del Observatorio de la Venezuela Analítica. Esta situación ha provocado una escasez significativa de recursos básicos, afectando a más del 30 % de los hogares venezolanos, muchos de los cuales carecen de los recursos necesarios para satisfacer sus necesidades básicas de alimentación, salud y vivienda de manera adecuada (Guyana, 2024). El colapso económico también ha tenido un efecto devastador sobre los sistemas de salud y educación, que no han logrado mantenerse operativos frente a la magnitud de la crisis. Según la Encuesta Nacional de Condiciones de Vida (ENCOVI), elaborada por la Universidad Católica Andrés Bello (UCAB), el 45 % de la población venezolana en situación de pobreza que presenta problemas de salud no accede a servicios médicos debido a la escasez de medicamentos, la falta de equipamiento y la carencia de personal sanitario capacitado. Esta crisis se ha visto profundizada por la migración masiva de profesionales del sector salud y del ámbito educativo, fenómeno que ha debilitado aún más la prestación de estos servicios esenciales. A su vez, la disminución de la tasa de fertilidad registrada en los últimos años refleja las severas restricciones económicas que enfrentan las familias, limitando su capacidad para garantizar el bienestar de un mayor número de hijos (Guyana, 2024).

Por otro lado, el Índice de Desarrollo Humano (IDH) de Venezuela ha mostrado un retroceso significativo desde 2013, como se observa en la Figura 2.3. Este indicador, que mide el desarrollo de un país en función de la esperanza de vida, la educación y el ingreso per cápita, presentó una tendencia ascendente sostenida hasta 2012, momento en el cual Venezuela se encontraba en la categoría de desarrollo humano medio-alto. Sin embargo, a partir de ese año, el IDH comenzó a disminuir de forma pronunciada, alcanzando en 2020 niveles

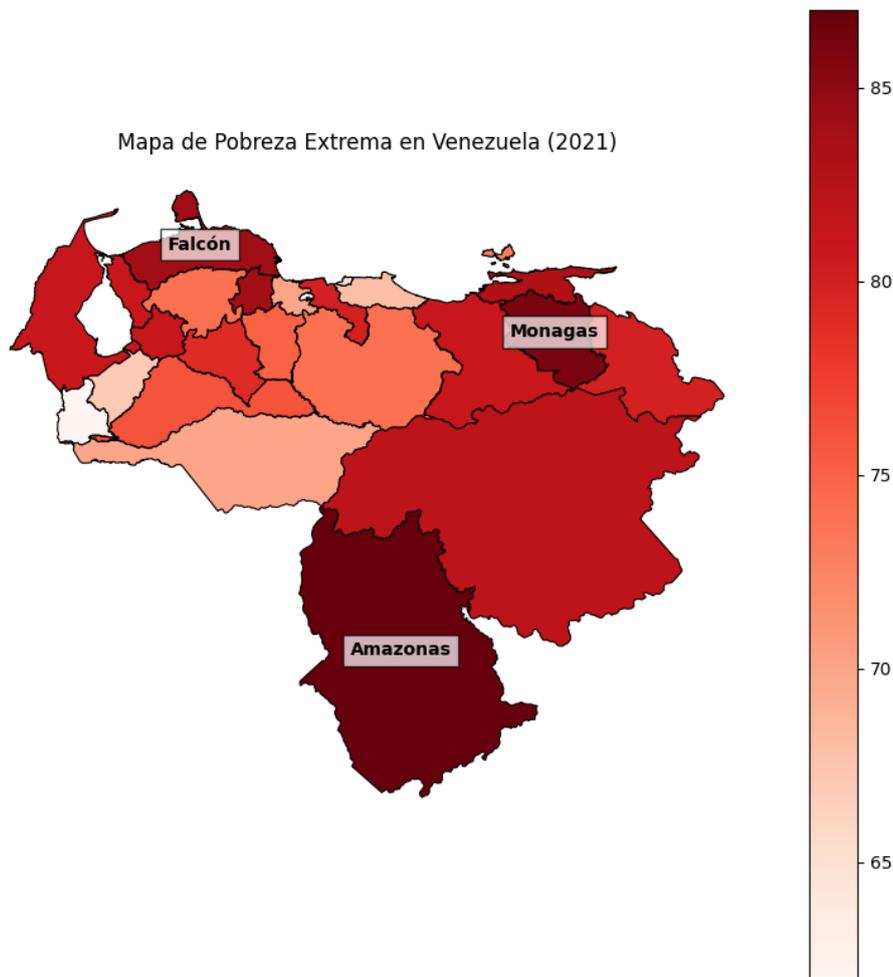


Figura 2.2: Índice de pobreza en Venezuela por Estado (como porcentaje de la población)  
Elaboración propia. Fuente de Datos: (ENCOVI, 2021)

propios de países con desarrollo humano bajo. Esta caída evidencia el impacto acumulado de factores como el deterioro económico, el colapso de los servicios básicos, y la precarización de las condiciones de vida (Expansión, 2021). El descenso en el IDH también refleja las profundas desigualdades sociales que se han intensificado desde 2013, un fenómeno característico de la crisis venezolana. Según datos de la Encuesta Nacional de Condiciones de Vida (ENCOVI), las áreas rurales y los estados más alejados de Caracas, como Táchira, son los más afectados. Estas desigualdades territoriales han profundizado la vulnerabilidad de amplios sectores de la población, especialmente en regiones con menor acceso a servicios básicos y oportunidades económicas (Encovi, 2023). Asimismo, la reducción de la actividad económica femenina ha sido otro factor crítico, con solo un 34,3 % de participación laboral femenina. Según ENCOVI, 1,8 millones de mujeres no trabajan debido a responsabilidades de cuidado, aunque manifiestan interés en participar en el mercado laboral, lo que evidencia un importante potencial económico desaprovechado en el contexto de la crisis (Encovi, 2023).

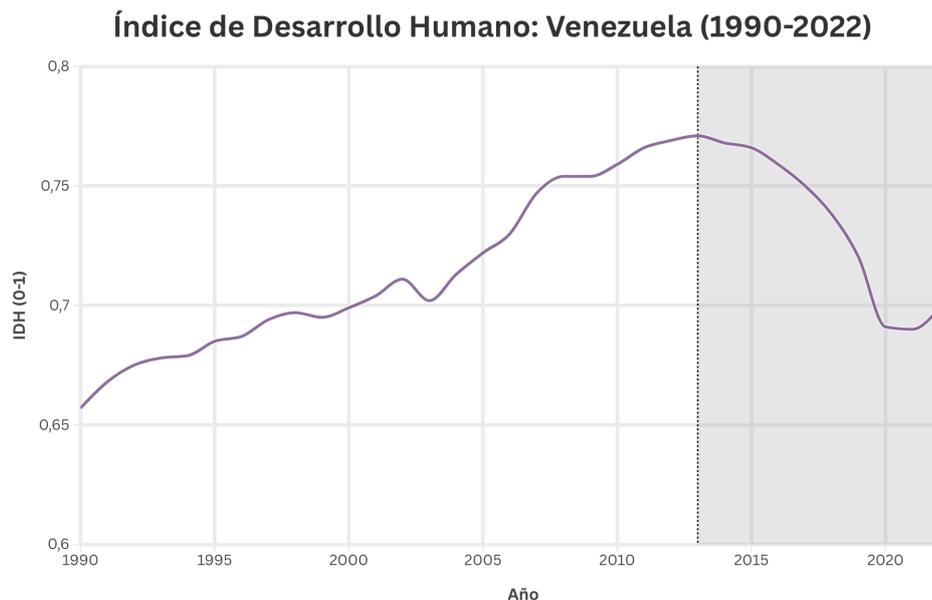


Figura 2.3: Índice de desarrollo humano (0-1)  
Fuente de datos: (Our World in Data, 2023). Elaboración propia.

En línea con el deterioro del desarrollo humano, las causas de mortalidad en Venezuela evidencian los efectos más devastadores de la crisis. En 2021, la principal causa de muerte fue la malnutrición infantil, una consecuencia directa de la inseguridad alimentaria que afecta al 20 % de los hogares debido a la pérdida de poder adquisitivo (Guyana, 2024). La Figura 2.4 muestra esta situación, seguida por el consumo de drogas como la segunda causa más relevante, agravada por la falta de recursos médicos adecuados para tratar las adicciones. Según CEPAL y OPS, este problema está estrechamente vinculado al colapso del sistema sanitario nacional (Superintendencia Nacional Antidrogas, 2023). Además, UNICEF señala que desde 2015 una parte significativa de la población ha dependido de fuentes de agua contaminadas, lo que ha provocado un aumento en enfermedades parasitarias y gastrointestinales, especialmente en los sectores más vulnerables. La carencia de servicios básicos sanitarios ha exacerbado esta situación, contribuyendo al deterioro de la salud pública y reflejando el impacto más profundo de la crisis humanitaria.

En el contexto de la crisis humanitaria y económica que atraviesa Venezuela, las dinámicas políticas han sido determinantes en el deterioro general del país. Según *The Economist Intelligence Unit* (EIU), Venezuela dejó de ser considerada una democracia en 2014, pasando a ser clasificada como un “autoritarismo híbrido”. Este cambio estuvo motivado por el aumento de la corrupción, la falta de transparencia y el progresivo debilitamiento de las instituciones desde la llegada de Nicolás Maduro al poder (Economist Intelligence, 2023). Las elecciones presidenciales de 2013, que marcaron el inicio de su mandato, estuvieron ya marcadas por controversias debido a la ausencia de transparencia. Como se observa en la Figura 2.5, el índice democrático del país, relativamente estable hasta 2014, sufrió una caída

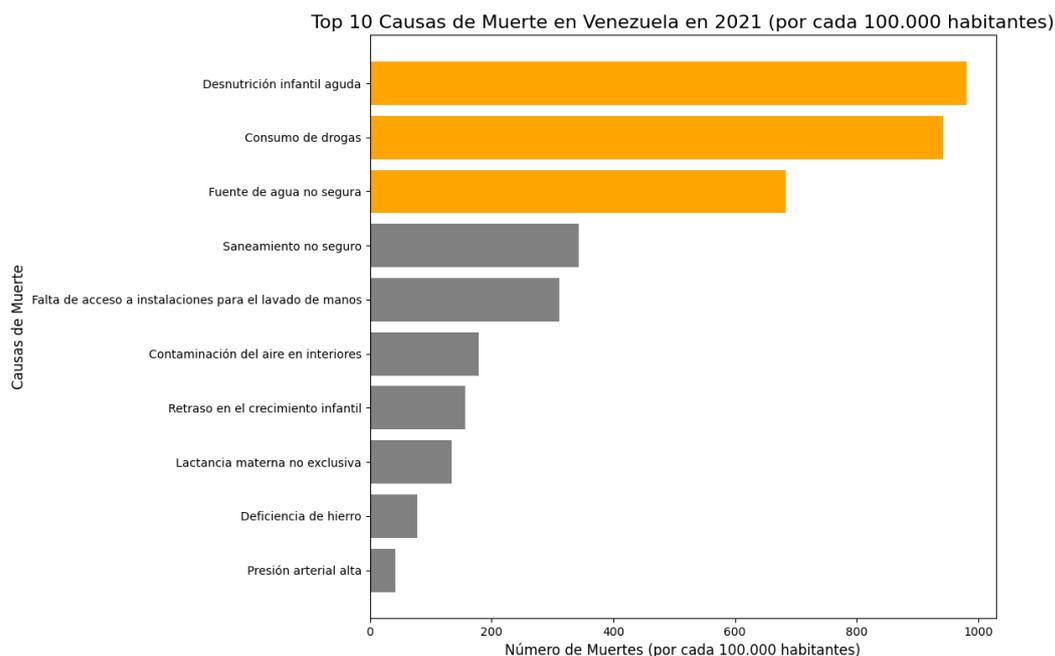


Figura 2.4: Causas de muerte en Venezuela (2021)  
Fuente de datos: (Our World in Data, 2021a). Elaboración propia.

drástica a partir de 2015, consolidándose como un régimen autoritario. Durante el mandato de Maduro, las restricciones a la libertad de expresión y la represión de la oposición se han intensificado. Su controvertida reelección en 2018, ampliamente rechazada por la comunidad internacional, no impidió que mantuviera el poder (Economist Intelligence, 2023). Este deterioro democrático ha exacerbado las ya profundas desigualdades sociales y económicas, alimentando la crisis humanitaria y acelerando el éxodo masivo de venezolanos.

En paralelo, la devaluación de la moneda venezolana, el bolívar, se ha convertido en una de las expresiones más visibles de la crisis económica. Como muestra la Figura 2.6, la tasa de cambio del bolívar frente al dólar ha incrementado exponencialmente desde 2013 debido a la hiperinflación descontrolada mencionada previamente (Zambrano Sequín, 2022). En 2018, el gobierno introdujo el bolívar soberano, eliminando cinco ceros de la moneda en un intento por contener la inflación y restaurar el poder adquisitivo. Sin embargo, esta medida no logró estabilizar la economía, ya que 2018 registró un récord hiperinflacionario del 1.698.488 %, según datos del Banco Central de Venezuela (BCV) y el FMI (BBC News Mundo, 2018). La continua depreciación del bolívar refleja la falta de confianza en la gestión económica del gobierno y las graves distorsiones estructurales que enfrenta el país.

En el marco del deterioro económico y social de Venezuela, las políticas monetarias y las condiciones del sistema eléctrico han profundizado las desigualdades existentes. En 2021, se introdujo el bolívar digital, eliminando seis ceros de la moneda en un intento por frenar los efectos de la hiperinflación, pero sin éxito. Aunque el 60 % de las transacciones en el país se realizan en dólares, lo que ha permitido estabilizar ciertos precios, esta dolarización parcial

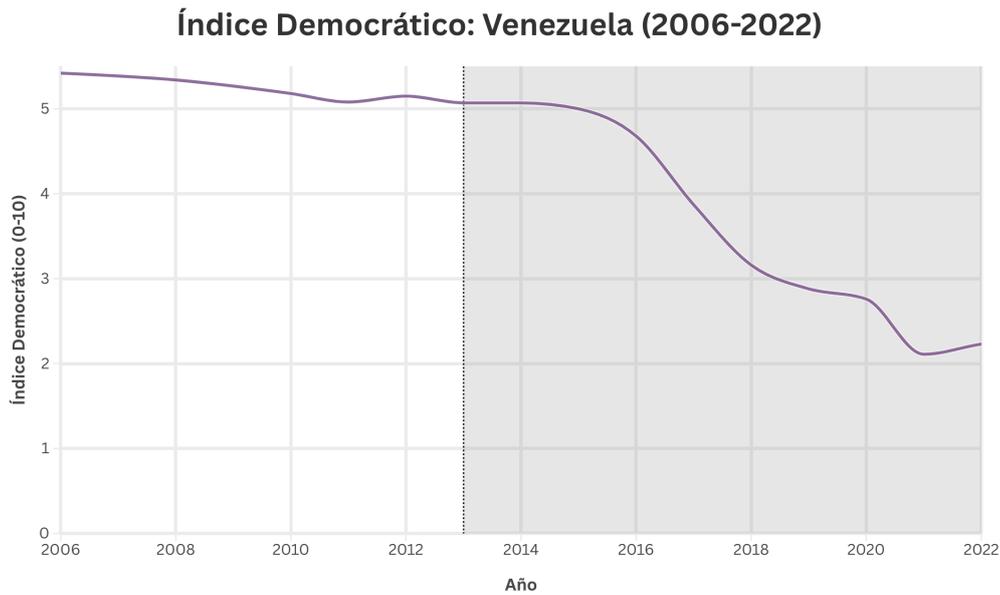


Figura 2.5: Índice de democracia (0-10)  
Fuente de datos: (Our World in Data, 2024a). Elaboración propia.

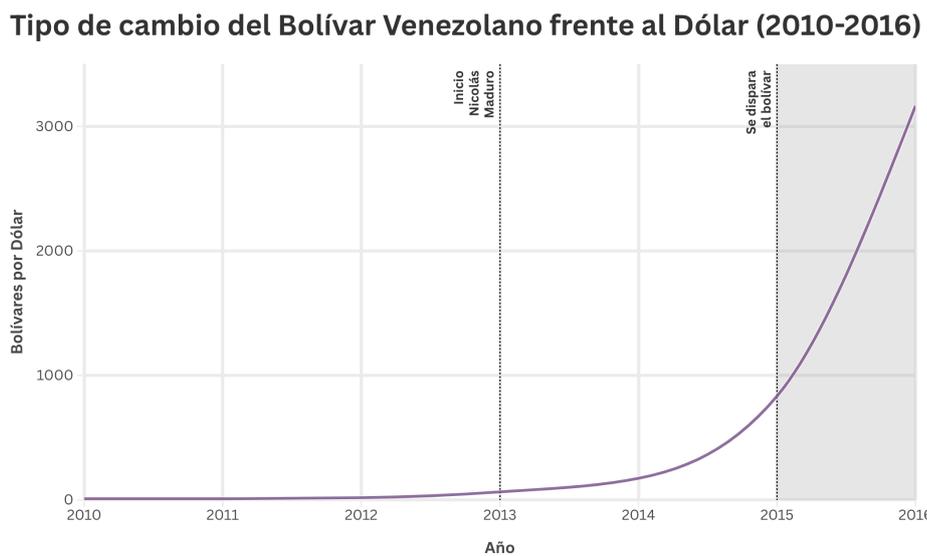


Figura 2.6: Cambio del Bolívar Venezolano frente al Dólar  
Fuente de datos: (Kaggle, 2025). Elaboración propia.

ha incrementado las disparidades, ya que una gran parte de la población no tiene acceso a divisas (Zambrano Sequín, 2022). Como muestra la Figura 2.6, desde 2018 el tipo de cambio del bolívar frente al dólar ha aumentado de forma exponencial, reflejo de la depreciación continua de la moneda local. En 2022, el salario mínimo mensual era inferior a 5 dólares, y en 2024, esta cifra disminuyó aún más a 3,60 dólares (130 bolívares). A pesar de las medidas gubernamentales compensatorias, como el Bono de Guerra Económica (60 dólares) y el Cestaticket Alimentación (40 dólares), estos subsidios no han sido suficientes para mitigar las carencias estructurales, beneficiando principalmente a trabajadores del sector público,

jubilados y pensionados (Durán, 2021).

Si bien estas medidas buscan ofrecer cierto alivio económico, el acceso desigual a servicios básicos, como la electricidad, sigue siendo una de las principales fuentes de descontento social. Los apagones constantes, tanto en áreas urbanas como rurales, evidencian las fallas estructurales del sistema eléctrico nacional, cuya gestión prioriza el control político sobre las inversiones técnicas necesarias. Estas deficiencias han agravado la percepción de ineficacia institucional y han intensificado las disparidades regionales (González Oquendo, 2019). La crisis de la industria petrolera, exacerbada por las sanciones internacionales y la falta de inversión en infraestructura, ha reducido significativamente la producción y refinación de petróleo y gas, disminuyendo su participación en la generación de electricidad desde la llegada de Maduro al poder. Además, la falta de mantenimiento y modernización de las plantas termoeléctricas ha evidenciado la ausencia de una planificación energética sostenible, dejando al país dependiente de fuentes limitadas (Doval, 2023).

En la Figura 2.7, se observa que una proporción significativa de la electricidad en Venezuela proviene de fuentes hidroeléctricas, representadas en azul. La central hidroeléctrica más importante, la Represa de Guri, es una de las más grandes del mundo y constituye la base de la generación eléctrica nacional. No obstante, la dependencia casi total de esta fuente ha convertido al país vulnerable a eventos climáticos extremos, como la sequía de 2016, que tuvo un impacto severo en la capacidad de generación hidroeléctrica. Esta vulnerabilidad ha generado frecuentes cortes eléctricos que afectan no solo a los hogares, sino también a infraestructuras críticas, como hospitales, escuelas y universidades. La interrupción del suministro eléctrico limita el acceso a servicios esenciales de salud y educación, agravando las desigualdades sociales y comprometiendo el desarrollo humano, esto se apreció especialmente a partir del año 2015 (Doval, 2023). Esta situación refleja que las fallas en la diversificación de las fuentes energéticas y en la gestión inadecuada de los recursos existentes han contribuido al agravamiento de la crisis humanitaria en Venezuela, con efectos que trascienden los indicadores económicos e impactan directamente en la calidad de vida de millones de personas.

En resumen, la crisis en Venezuela representa un colapso multidimensional que abarca los ámbitos económico, social y político desde 2013. La dependencia histórica del petróleo como eje central de la economía nacional, sumada a las sanciones internacionales y a una gestión gubernamental caracterizada por la corrupción y la falta de planificación, ha provocado una contracción masiva del PIB y del PIB per cápita. Esto ha reducido drásticamente el poder adquisitivo de la población, alimentando una hiperinflación sin precedentes y situando a más del 90 % de los venezolanos bajo la línea de pobreza, de los cuales más del 60 % vive en condiciones de pobreza extrema (Naciones Unidas, 2022). Los sistemas de salud y educación han colapsado debido a la falta de inversión y al éxodo masivo de profesionales, lo que ha resultado en una creciente mortalidad infantil e inseguridad alimentaria en numerosos hogares (Guyana, 2024). A este panorama se añade la progresiva desintegración y pérdida de credibilidad de las instituciones, junto con el deterioro de la infraestructura eléctrica, ca-

### Fuentes de energía para producción de electricidad Venezuela (2000-2022)

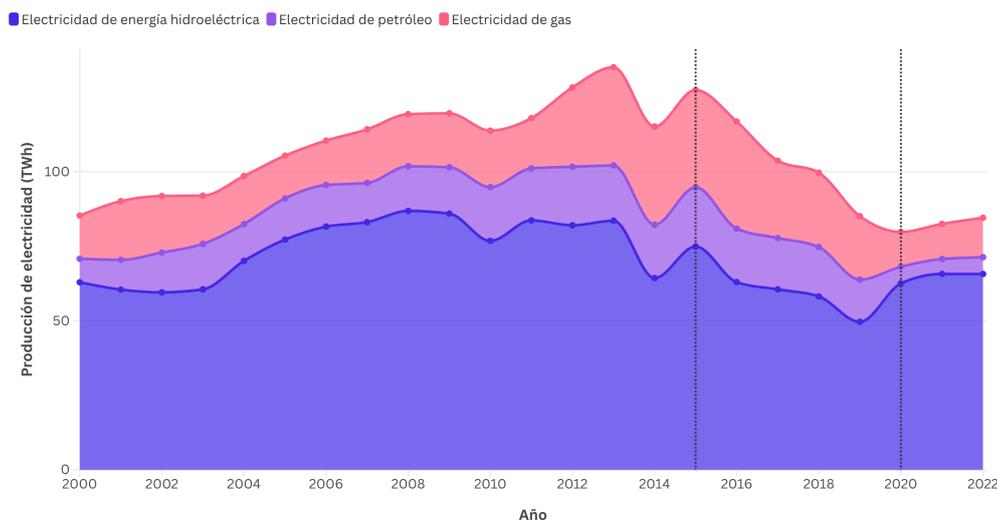


Figura 2.7: Fuentes de energía en Venezuela  
Fuente de datos: (Our World in Data, 2024b). Elaboración propia.

racterizado por apagones frecuentes y una dependencia excesiva de fuentes hidroeléctricas altamente vulnerables a eventos climáticos. Estas deficiencias han restringido aún más el acceso a servicios esenciales, agravando las desigualdades sociales. Estos factores combinados han dado lugar a una crisis humanitaria de enormes proporciones, impulsando a más de 7 millones de venezolanos a abandonar su país en busca de condiciones de vida dignas. La migración masiva, más que una respuesta al deterioro interno, refleja el impacto acumulado de un modelo económico no sostenible, la descomposición institucional y la falta de políticas públicas efectivas para garantizar derechos básicos.

## 2.2. Flujos migratorios en Venezuela: patrones y destinos

El éxodo masivo de ciudadanos venezolanos ha dado lugar a una diáspora amplia y diversa, distribuida a lo largo de América Latina, América del Norte y Europa. Si bien todos estos destinos comparten la recepción de migrantes que huyen de una crisis humanitaria compleja, cada país presenta particularidades en cuanto a su proximidad geográfica, vínculos históricos, marcos jurídicos migratorios, niveles de acogida institucional y percepciones sociales hacia los recién llegados. Las siguientes subsecciones examinan en detalle cinco destinos representativos de la diáspora venezolana, Colombia, Brasil, México, Estados Unidos y España, con el fin de identificar tanto los patrones de llegada y asentamiento como los desafíos específicos de integración que enfrentan los migrantes en cada contexto.

### **2.2.1. Diáspora venezolana en Colombia: puerta de entrada y refugio**

Desde el año 2014, Colombia se ha consolidado como el principal país receptor de migrantes y refugiados venezolanos en América Latina y el Caribe, desempeñando simultáneamente el rol de país de tránsito y de acogida permanente. La proximidad geográfica, los vínculos históricos y sociales, así como la relativa permeabilidad de su frontera compartida, han convertido a Colombia en el primer destino en el marco del éxodo venezolano. Esta situación ha generado profundas transformaciones sociales, institucionales y económicas tanto en las regiones fronterizas como en las principales ciudades del país (Proyecto Migración: Venezuela, 2024). Desde 2015, Colombia ha mantenido un compromiso sostenido con una política migratoria abierta, que ha sido ampliamente reconocida como ejemplar en la región. Basada en la premisa de que una migración bien gestionada representa una oportunidad para el desarrollo, la estrategia colombiana ha priorizado la integración social y económica de los migrantes, enfocándose especialmente en las poblaciones más vulnerables.

Entre las medidas adoptadas se incluyen la emisión de permisos de trabajo, tránsito y estadía; el acceso a subsidios habitacionales; la inclusión en programas sociales, educativos y sanitarios; y la protección de la niñez, la reunificación familiar y la lucha contra la trata de personas (Grupo Banco Mundial, 2021). Este enfoque ha contado con el respaldo de organismos multilaterales, como el Banco Mundial, que desde 2017 colabora estrechamente con el gobierno colombiano en la implementación de una respuesta migratoria que promueva simultáneamente el desarrollo humano y el crecimiento económico del país. En este contexto, el Estatuto Temporal de Protección para Migrantes Venezolanos (ETPV), lanzado en 2021, constituye un pilar central de la estrategia colombiana. Esta política permite regularizar a los migrantes durante un periodo de diez años, facilitando su acceso a educación, salud y empleo formal, lo que a su vez contribuye a reducir la informalidad y fortalecer la cohesión social (Grupo Banco Mundial, 2021). Según datos recientes, Colombia concentra el mayor número de tarjetas de residencia emitidas para migrantes venezolanos en 2024. Asimismo, la Resolución 5477, expedida por el Ministerio de Relaciones Exteriores el 22 de julio de 2022, estableció la Visa de Residente Permanente, dirigida a aquellos ciudadanos que deseen establecerse de forma indefinida en el país, fortaleciendo aún más el marco jurídico de protección e integración (Affirma Legal, 2024).

Por otro lado, los aportes económicos de la población venezolana en Colombia también han sido significativos. Según estimaciones de la Organización Internacional para las Migraciones (OIM), los migrantes y refugiados venezolanos generaron un impacto económico superior a los 529,1 millones de dólares. El Estudio de Impacto Fiscal de la Migración Venezolana en Colombia (2022) reveló que aproximadamente el 2 % de los ingresos fiscales del país provino de esta población, con una tasa de participación laboral cercana al 90 % y un 20 % con educación formal (OIM: ONU Migración, 2024). Estos datos destacan el papel positivo de la migración en la economía nacional, más aún en el caso colombiano, que

representa el núcleo más amplio de la diáspora venezolana.

La Figura 2.8 ilustra de manera clara el crecimiento exponencial de la población venezolana en Colombia entre 1990 y 2020. Mientras que en los primeros años del periodo la migración era marginal, se observa un incremento sustancial a partir de 2015, con un aumento particularmente pronunciado en 2020, cuando la población supera los 1,7 millones de personas. Este ascenso está directamente vinculado con la profundización de la crisis humanitaria en Venezuela, pero también con la apertura de políticas migratorias en Colombia, que facilitaron el asentamiento regular de esta población. Durante la pandemia de COVID-19, a pesar del cierre inicial de fronteras, Colombia mantuvo una postura de cooperación binacional que permitió la repatriación de más de 119.000 venezolanos por motivos de reunificación familiar y emergencia sanitaria. Sin embargo, 2020 también marcó un punto de inflexión en los flujos migratorios hacia Colombia, ya que miles de personas cruzaron la frontera en busca de condiciones mínimas de subsistencia, dada la precariedad del sistema sanitario venezolano.



Figura 2.8: Población Venezolana en Colombia  
Fuente de datos: (Our World in Data, 2021b). Elaboración propia.

En síntesis, la diáspora venezolana en Colombia constituye un caso paradigmático dentro del panorama migratorio regional. No solo por su magnitud, sino por el enfoque integral adoptado por el Estado colombiano en materia de regularización e integración.

### **2.2.2. Diáspora venezolana en Brasil: refugio en la frontera amazónica**

Brasil, país vecino al sur de Venezuela, ha desempeñado un papel significativo como receptor de la diáspora venezolana, especialmente a partir de 2018. Desde entonces, más de 800.000 ciudadanos venezolanos han cruzado la frontera en busca de protección y mejores

condiciones de vida. Si bien el idioma ha representado una barrera para una parte de los migrantes, el Gobierno Federal ha implementado políticas de acogida y regularización que han facilitado su asentamiento, transformando a Brasil en el quinto país del mundo con mayor número de venezolanos acogidos, según datos de la Organización de los Estados Americanos (OEA) (Jiménez, 2020). En respuesta a este fenómeno, en 2018 se puso en marcha la “Operación Acogida”, una estrategia humanitaria coordinada por el Gobierno Federal basada en tres pilares fundamentales: la gestión de fronteras (que incluye recepción, documentación y atención médica), el alojamiento temporal en ciudades fronterizas, y la reubicación voluntaria hacia otras regiones del país con mejores oportunidades económicas (ONU, 2023). Esta última ha sido de gran relevancia para descongestionar las zonas de entrada, particularmente el estado de Roraima, cuya capital, Boa Vista, se ha convertido en un centro de recepción masiva desde la apertura del paso fronterizo por Pacaraima.

Las Figuras 2.9a y 2.9b permiten visualizar la evolución en la distribución geográfica de los migrantes venezolanos en Brasil. En 2022, la gran mayoría de la población migrante se concentraba en el estado de Roraima, al norte del país, donde se localiza el principal punto de ingreso. Sin embargo, en 2023 se aprecia una redistribución significativa hacia estados del sur, como Paraná y Santa Catarina. Este desplazamiento geográfico es resultado directo del componente de reubicación voluntaria impulsado por la Operación Acogida, que busca facilitar la integración socioeconómica en regiones con mayor capacidad de absorción laboral y acceso a servicios básicos (ONU, 2023).

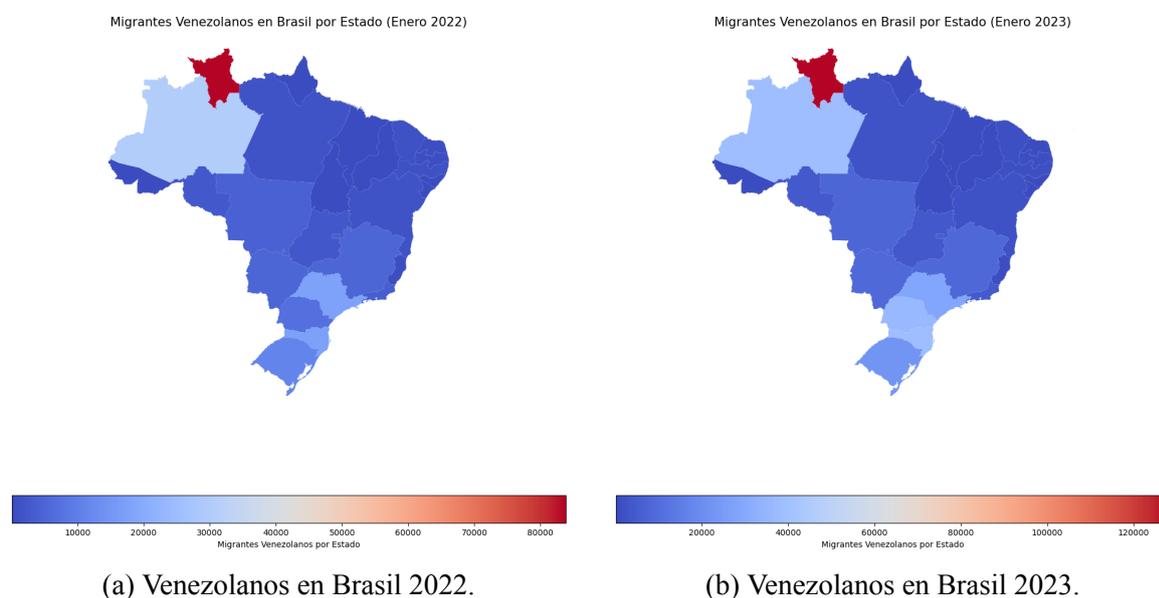


Figura 2.9: Venezolanos en Brasil en 2022 y 2023  
Elaboración propia. Fuente de datos: (Kaggle, 2023)

El marco normativo brasileño ha garantizado el acceso de la población venezolana a múltiples derechos, entre ellos la documentación legal, atención sanitaria, educación y movilidad

territorial. Los refugiados reconocidos pueden obtener un documento de identificación fiscal, acceder a la sanidad pública y matricular a sus hijos en el sistema educativo. Asimismo, se encuentran habilitados para participar en programas sociales como Bolsa Família, una política pública orientada a la reducción de la pobreza y al aumento del poder adquisitivo de los hogares. Durante la pandemia de COVID-19, también se les otorgó acceso a la renta básica de emergencia, de la que se beneficiaron más de 42.500 personas venezolanas, según datos de la Caixa Econômica Federal (Jiménez, 2020).

Desde 2022, el Gobierno Federal ha mantenido una política de fronteras abiertas para refugiados y migrantes venezolanos, aunque se reconoce que persisten desafíos administrativos en la frontera que han dificultado la tramitación eficiente de los procesos de regularización. Para ser reconocidas como refugiadas, las personas venezolanas deben cumplir con cinco criterios establecidos por el Comité Nacional para los Refugiados (Conare): presentar documento de identidad (vigente o vencido), tener el último ingreso al país registrado, no contar con antecedentes penales, ser mayores de edad y no haber solicitado otro estatus migratorio en Brasil (ACNUR, 2024). Una innovación importante ha sido la adopción del procedimiento *prima facie* por parte del Conare, mediante el cual se reconoce automáticamente a los ciudadanos venezolanos como refugiados sin necesidad de entrevistas extensas. Esta medida fue justificada por la constatación oficial de que en Venezuela se cometen violaciones graves y generalizadas de derechos humanos. Como resultado, se ha agilizado el proceso de otorgamiento del estatus de refugiado, del que actualmente se benefician más de 49.000 personas en Brasil, de las cuales el 95 % son de nacionalidad venezolana (Jiménez, 2020).

No obstante, la implementación de estas políticas tiene lugar en un contexto socioeconómico particularmente desafiante. Brasil enfrenta elevados niveles de pobreza y desigualdad, así como una marcada heterogeneidad en la cobertura y calidad de los servicios públicos entre regiones. Esta situación no solo afecta a la población migrante, sino también a la ciudadanía brasileña, generando flujos internos hacia estados con mejores condiciones de vida. Tanto nacionales como extranjeros se enfrentan a problemáticas compartidas, como la violencia, el desempleo y la insuficiencia de servicios esenciales en diversas zonas del país. En este contexto, los principales estados de acogida de la población venezolana son Roraima, Amazonas, São Paulo y Santa Catarina, tal como se observa en la Figura 2.9b. De acuerdo con un informe de la OIM publicado en marzo de 2022, el 95 % de los migrantes venezolanos entrevistados manifestó su intención de permanecer en Brasil. Esta voluntad de permanencia persiste a pesar de los desafíos de integración: el 42 % de los encuestados reportó dificultades para acceder al empleo, y un 16 % mencionó obstáculos relacionados con el acceso a servicios básicos como educación, salud y alimentación (Paredes, 2023).

A pesar de estas limitaciones y de la presión que la migración ejerce sobre los servicios públicos en regiones fronterizas, como el estado de Roraima, donde el gobernador Antonio Denarium señaló que el 40 % de las camas hospitalarias están ocupadas por ciudadanos venezolanos, Brasil ha sostenido su compromiso humanitario (Jiménez, 2020). A través de la

iniciativa “Operación Acogida”, el país ha procurado mejorar la gestión y recepción de personas migrantes provenientes de Venezuela, en un esfuerzo por armonizar los principios de solidaridad internacional con los desafíos estructurales internos de un país receptor. Aun con recursos limitados y tensiones sociales latentes, esta política ha sido reconocida como una experiencia destacada de cooperación internacional para enfrentar una de las crisis migratorias más relevantes de América Latina en las últimas décadas.

### **2.2.3. Diáspora venezolana en México: país de paso hacia el norte**

México ha emergido como un territorio clave dentro de los flujos migratorios procedentes de Venezuela, desempeñando una doble función como país de tránsito hacia Estados Unidos y, cada vez con mayor frecuencia, como lugar de asentamiento temporal o permanente. Esta posición geopolítica, como punto de conexión entre América Latina y América del Norte, convierte a México en un espacio de alta presión migratoria, en particular por su frontera sur, relativamente permeable con Guatemala, y su frontera norte, fuertemente controlada por los Estados Unidos. Durante el año 2023, se registraron niveles sin precedentes de entradas irregulares y solicitudes de asilo por parte de ciudadanos venezolanos, lo que situó a México entre los principales países receptores de la región en términos relativos (Delgado Linero, 2024).

En respuesta, el gobierno mexicano ha adoptado medidas orientadas a la regularización temporal, como la Tarjeta de Visitante por Razones Humanitarias (TVRH), que permite a los migrantes acceder a servicios básicos como salud y educación. Además, México ha reforzado su cooperación con organismos internacionales como la OIM y ACNUR, con el objetivo de garantizar la protección de los derechos fundamentales de las personas migrantes (Delgado Linero, 2022). No obstante, persisten importantes desafíos en términos de seguridad y derechos humanos, especialmente en las zonas fronterizas, donde se han reportado casos sistemáticos de extorsión, secuestros y abusos. Esta realidad complejiza la labor institucional y pone a prueba los compromisos adquiridos por el Estado mexicano en el marco del Pacto Mundial para la Migración Segura, Ordenada y Regular, el cual promueve políticas basadas en la protección, la integración y la justicia migratoria (Venemex, 2018).

La Figura 2.10 ilustra la evolución mensual de las renovaciones de tarjetas de residencia por parte de ciudadanos venezolanos durante el año 2024. Se observa una tendencia ascendente en los primeros siete meses del año, con aumentos especialmente pronunciados en los meses de abril y julio. Esta dinámica puede interpretarse como un reflejo del creciente número de personas que optan por establecerse legalmente en el país, ya sea por la dificultad de cruzar hacia Estados Unidos o por una mayor percepción de estabilidad y acceso a derechos dentro del territorio mexicano.

En el contexto de la frontera norte, estados como Tamaulipas y Chihuahua (particularmente Ciudad Juárez) han adquirido una importancia estratégica como puntos de tránsito ha-

**Renovación de tarjetas de residencia para venezolanos en México (2024)**

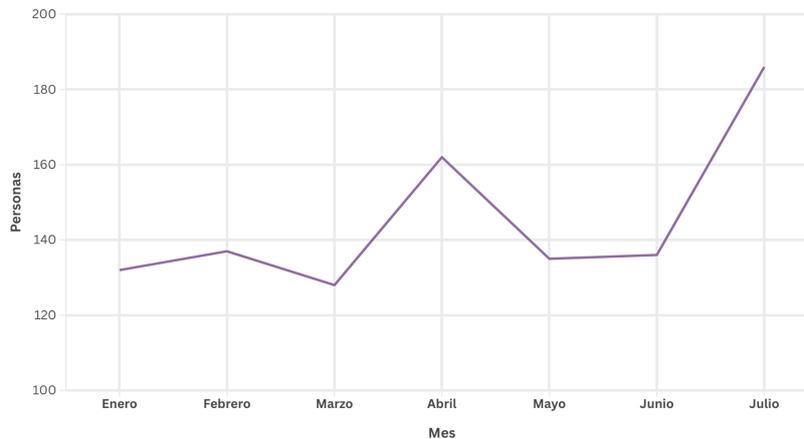


Figura 2.10: Venezolanos que tienen la tarjeta de residencia en México (2024)  
Fuente de datos: (Secretaría de Gobernación (SEGOB), 2024a). Elaboración propia.

cia los Estados Unidos. La derogación del Título 42 en 2023, una medida instaurada durante la pandemia para restringir el acceso por razones sanitarias, marcó un punto de inflexión en la dinámica migratoria. Aunque el gobierno de Joe Biden impulsó acuerdos bilaterales para ordenar y canalizar los flujos, el volumen de devoluciones desde Estados Unidos ha generado presión sobre las autoridades mexicanas, provocando tensiones sociales en comunidades fronterizas y desafíos logísticos en la gestión de personas en tránsito (ONU, 2024).

Esta presión en las regiones fronterizas del norte se refleja también en la distribución espacial de la población migrante venezolana dentro del territorio mexicano. La Figura 2.11 permite observar que los estados con mayor número de inmigrantes venezolanos en 2024 fueron Chiapas, en la frontera sur con Guatemala, y Sonora, en el norte, limítrofe con Estados Unidos, seguidos por Tabasco. Este patrón confirma el carácter eminentemente transitorio de México en la ruta migratoria venezolana, con una concentración notable en las zonas de entrada y salida del país, mientras que la presencia en estados del centro permanece reducida. La elección de Sonora como principal punto de cruce durante la segunda mitad de 2023 por encima de otras rutas como Chihuahua o Baja California, pone de manifiesto la continua transformación de los corredores migratorios ante el reforzamiento de los controles y las dinámicas de devolución.

Aunque muchos migrantes venezolanos aspiran a continuar hacia Estados Unidos, una parte considerable permanece en territorio mexicano, ya sea de forma voluntaria o forzada por las circunstancias, integrándose paulatinamente en las economías locales. Esta inserción ocurre principalmente en sectores como el comercio y los servicios, donde los migrantes comienzan a contribuir al desarrollo económico de los estados receptores. Sin embargo, enfrentan importantes desafíos derivados de la irregularidad migratoria, la falta de reconocimiento de sus credenciales laborales y episodios de discriminación. Frente a esta situación, el Estado

mexicano ha comenzado a fortalecer sus programas de integración, con el objetivo de reducir las barreras de acceso a derechos y promover la inclusión efectiva de la población migrante en el tejido social y económico nacional (Delgado Linero, 2024).



Figura 2.11: Inmigrantes venezolanos en México, 2024

Fuente de datos: (Secretaría de Gobernación (SEGOB), 2024b). Elaboración propia.

La complejidad del tránsito migratorio por el norte de México se ha visto agravada en los últimos años por la implementación de mecanismos de entrada regulada hacia Estados Unidos, como el sistema digital *CBP One*. Esta aplicación, desarrollada por el gobierno estadounidense, obliga a las personas migrantes a solicitar una cita para su ingreso por los puertos de entrada autorizados. Sin embargo, el sistema ha sido ampliamente criticado por su carácter aleatorio y por los requerimientos tecnológicos que impone, excluyendo de facto a quienes llegan sin dispositivos móviles funcionales o sin conectividad, tras haber atravesado múltiples países en condiciones de vulnerabilidad y extorsión. El tiempo medio de espera para obtener una cita mediante *CBP One* puede superar los dos meses, lo que obliga a miles de personas a permanecer en ciudades del norte de México, como Nogales, Ciudad Juárez o Tijuana, en condiciones precarias, esperando su turno de ingreso.

Un requisito adicional del sistema es que la localización GPS del dispositivo debe coincidir con zonas cercanas a la frontera, lo que genera una concentración desproporcionada de personas en esas regiones. En 2022, las posibilidades de ingreso regular eran aún más limitadas: mientras que familias con menores o menores no acompañados podían acceder mediante listas gestionadas en los puntos de entrada, los adultos solos eran frecuentemente deportados de forma inmediata, en ocasiones tras la incautación de documentos y pertenencias personales. Organizaciones de la sociedad civil, como *Kino Border Initiative*, desempeñan un papel relevante en este contexto, proporcionando alojamiento temporal, servicios básicos y asesoría jurídica a las personas migrantes. Estas iniciativas intentan paliar las carencias estructurales

de atención humanitaria en la frontera, así como empoderar a los migrantes en el conocimiento y ejercicio de sus derechos.

#### 2.2.4. Diáspora venezolana en Estados Unidos: el sueño económico de los venezolanos

Estados Unidos se ha convertido en uno de los principales destinos de la diáspora venezolana, atraída por las oportunidades económicas, la estabilidad institucional y la garantía de derechos humanos, elementos que posicionan al país como un referente global de seguridad jurídica y democracia liberal. De acuerdo con el *Migration Policy Institute*, más de 545.000 venezolanos residen actualmente en Estados Unidos, cifra que ha aumentado de forma sostenida en la última década en respuesta al agravamiento de la crisis política, económica y humanitaria en Venezuela. Este crecimiento se encuentra documentado en la Figura 2.12, la cual muestra una tendencia ascendente particularmente acelerada a partir del año 2015. Si bien se observa un aumento también en el número de venezolanos nacidos en territorio estadounidense, la población migrante recién llegada representa la mayoría, lo que refleja la magnitud y urgencia del fenómeno.

**Población Venezolana en Estados Unidos (2000-2021)**

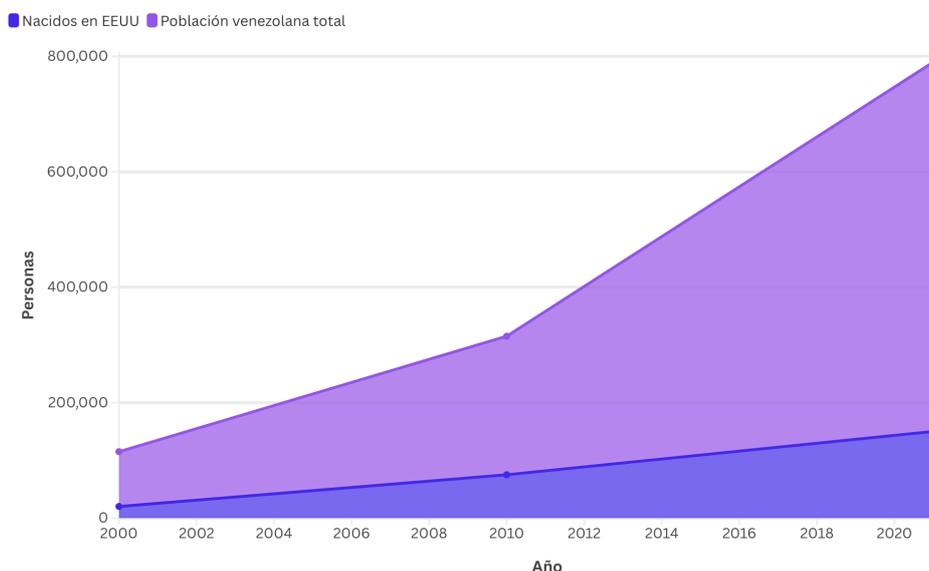


Figura 2.12: Venezolanos en Estados Unidos (2000-2021)  
Fuente de datos: (Pew Research Center, 2021). Elaboración propia.

La distribución geográfica de los migrantes venezolanos dentro del país se muestra en la Figura 2.13. Florida destaca con diferencia como el estado con mayor concentración de esta población, especialmente en las áreas metropolitanas de Miami y Orlando. Esta localización responde a una combinación de factores como las redes familiares preexistentes, la cercanía cultural, el idioma y la infraestructura comunitaria de apoyo. Otros estados relevantes en

términos de recepción son Texas, Nueva York y Georgia, donde también se han establecido comunidades significativas. Este patrón de asentamiento sugiere que, más allá de la migración como respuesta a una crisis, existe una lógica de inserción territorial estratégica que facilita la integración progresiva de los migrantes.

### Distribución de inmigrantes venezolanos en EE. UU.

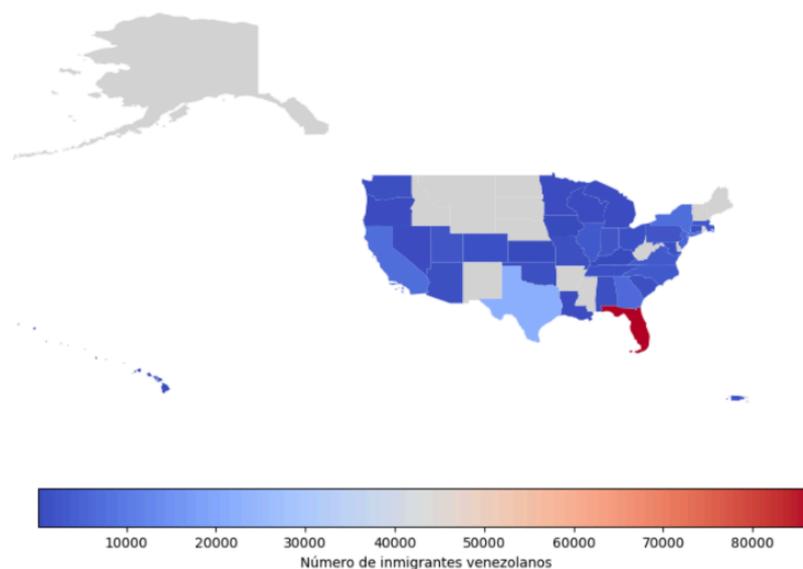


Figura 2.13: Migración Venezolana en Estados Unidos (2023)  
Fuente de datos: (U.S. Census Bureau, 2023). Elaboración propia.

Frente a este flujo migratorio creciente, el gobierno de Estados Unidos ha adoptado distintas medidas para gestionar la presencia de ciudadanos venezolanos en su territorio. En 2021, la administración de Joe Biden activó el Estatus de Protección Temporal (TPS), que otorga permisos de residencia y trabajo a personas de nacionalidades afectadas por crisis humanitarias. En el caso venezolano, la medida fue renovada en 2023, beneficiando a más de 470.000 personas (Moslimani, 2024). No obstante, el TPS solo es aplicable a quienes se encontraban en el país antes de la fecha de corte establecida, lo cual ha dejado fuera a miles de migrantes que ingresaron posteriormente, generando una situación de inseguridad jurídica y desigualdad de acceso a derechos (Center, 2023). Otra política migratoria de relevancia ha sido el Programa de Parole Humanitario, diseñado para otorgar entrada legal temporal a ciudadanos de países con condiciones extraordinarias, como Venezuela, Cuba, Haití y Nicaragua. Aunque esta herramienta ha permitido el ingreso regulado de ciertos grupos, su implementación ha sido criticada por los altos costes asociados y por el acceso limitado de personas en situación de vulnerabilidad extrema (Citizenship and Services, 2023). Paralelamente, la derogación del Título 42 en 2023, una medida introducida durante la pandemia para justificar la expulsión inmediata de migrantes por motivos sanitarios, fue reemplazada por la aplica-

ción del Título 8, que establece sanciones más severas para aquellos que crucen la frontera de forma irregular, incluyendo deportaciones inmediatas y restricciones futuras de entrada.

El panorama migratorio se torna aún más incierto ante el regreso de Donald Trump a la presidencia en 2025. Entre las propuestas que ha anunciado se encuentra la ejecución de una campaña masiva de deportaciones, el uso de recursos militares para reforzar la seguridad fronteriza y la reimplementación de vetos migratorios dirigidos a países específicos. Aunque ha esbozado medidas para retener talento extranjero, como otorgar tarjetas de residencia automática a graduados de universidades estadounidenses, su agenda migratoria general se basa en un enfoque restrictivo que podría generar consecuencias humanitarias significativas (Español, 2024). De concretarse estas políticas, la comunidad venezolana en Estados Unidos podría enfrentarse a nuevos obstáculos para su regularización, permanencia e integración, en un contexto ya marcado por la incertidumbre migratoria.

### **2.2.5. Diáspora venezolana en España: el hogar culturalmente familiar cruzando el Atlántico**

España se ha consolidado como el principal destino europeo de la diáspora venezolana, albergando la comunidad más numerosa de venezolanos en Europa. A finales de 2023, se estimaba que más de 500.000 ciudadanos venezolanos residían en territorio español, según datos de Eurostat y del Observatorio Permanente de la Inmigración (OPI) (Ministerio de Inclusión, Seguridad Social y Migraciones, 2024). Esta preferencia migratoria se explica por una confluencia de factores históricos, culturales y administrativos: el idioma compartido, las similitudes socioculturales, la existencia de acuerdos bilaterales en materia de reconocimiento de títulos académicos y las relaciones diplomáticas fluidas entre ambos países facilitan la elección de España como destino. Además, para los migrantes venezolanos, el proceso de regularización y obtención de permisos de residencia resulta más accesible que en otros países de la Unión Europea.

En términos de políticas migratorias, España ha implementado mecanismos específicos que benefician particularmente a esta comunidad. Desde 2019, un número creciente de venezolanos ha accedido a la protección internacional en su modalidad subsidiaria, la cual otorga residencia y permiso de trabajo, aun cuando no se conceda el estatus formal de asilo. Según el Ministerio del Interior, los venezolanos constituyen el colectivo que más solicitudes de protección internacional presenta anualmente (Ministerio de Inclusión, Seguridad Social y Migraciones, 2024). Además, la legislación española contempla otras vías de regularización, como el arraigo social, aplicable tras tres años de permanencia continuada en el país. Este marco normativo, junto con el efecto de la llamada “migración en cadena”, es decir, la atracción ejercida por comunidades ya asentadas en ciudades como Madrid, Barcelona y Valencia, ha contribuido a reforzar la posición de España como principal receptor europeo de esta diáspora (CEMVE, 2024).

El análisis comparado de los destinos principales de la migración venezolana refleja la multiplicidad de trayectorias y las condiciones diferenciadas de acogida. En Estados Unidos y México, la búsqueda del llamado “sueño americano” se enfrenta a crecientes obstáculos normativos y restricciones fronterizas, especialmente en el actual contexto político marcado por el retorno de discursos antiinmigrantes. En cambio, Colombia y Brasil, por su proximidad geográfica, actúan como receptores naturales, aunque con capacidades institucionales limitadas y desafíos notables en cuanto a infraestructura de acogida e integración, como se evidencia particularmente en el caso del estado brasileño de Roraima. España, por su parte, destaca por ofrecer un entorno institucional más favorable, reforzado por afinidades culturales y una estructura legal que facilita la inserción social y económica.

En este contexto, la diáspora venezolana pone de manifiesto no solo las consecuencias devastadoras de una crisis estructural prolongada en su país de origen, sino también la necesidad urgente de respuestas coordinadas a nivel internacional. Dichas respuestas deben contemplar tanto la asistencia humanitaria inmediata como el diseño de políticas sostenibles de integración, capaces de fomentar la autonomía, dignidad y participación plena de las personas migrantes en las sociedades receptoras. La experiencia venezolana evidencia, asimismo, la capacidad de resiliencia de millones de personas que, ante la ausencia de horizontes viables en su territorio, emprenden procesos de movilidad transnacional como única vía para reconstruir sus vidas.

### **2.3. Técnicas de Machine Learning para el análisis de redes sociales en estudios migratorios**

La integración de personas migrantes constituye un proceso multidimensional que trasciende el acceso a derechos básicos, abarcando dimensiones económicas, psicológicas, sociales, lingüísticas y cívicas (Organización Internacional para las Migraciones, 2024). Esta complejidad exige no solo políticas públicas inclusivas, sino también el empoderamiento de las comunidades de acogida para facilitar la participación activa de los migrantes en la vida social. En este contexto, las opiniones y actitudes de la población local resultan de gran relevancia, ya que influyen directamente en la construcción de sociedades multiculturales cohesionadas (Luzes, 2024). De esta manera, comprender dichas percepciones es importante para identificar barreras simbólicas, detectar focos de tensión y diseñar intervenciones eficaces que favorezcan la convivencia. En el caso específico de la diáspora venezolana, la percepción pública varía significativamente según el país de acogida. En Colombia, por ejemplo, la implementación de políticas públicas orientadas a la integración ha contribuido a una institucionalización más sólida del fenómeno migratorio (UNHCR ACNUR, OIM, 2023). Sin embargo, en otros contextos, la narrativa pública ha estado marcada por la polarización y el rechazo. Ante esta diversidad de escenarios, el análisis de redes sociales emerge como una

herramienta eficaz para explorar en tiempo real las actitudes sociales hacia los migrantes. Estas plataformas permiten captar el discurso público de forma espontánea y descentralizada, ofreciendo evidencia empírica sobre el clima social que rodea los procesos migratorios y facilitando el diseño de políticas basadas en datos.

En este sentido, el uso de técnicas de análisis automático de textos adquiere especial relevancia para el estudio de la percepción pública sobre la migración. La minería de textos, o *text mining*, permite procesar y analizar grandes volúmenes de datos generados en redes sociales, convirtiéndose en una herramienta necesaria para el análisis de fenómenos sociales complejos. Su aplicación en plataformas digitales como Reddit o X facilita la identificación de patrones discursivos, emociones y temas recurrentes en torno a la migración, proporcionando evidencia empírica sobre las actitudes colectivas hacia los migrantes. A través de técnicas avanzadas de procesamiento del lenguaje natural (PLN) y algoritmos de *machine learning* (ML), es posible detectar tanto el tono emocional de los mensajes, por medio del análisis de sentimientos, como las narrativas más frecuentes, a través del modelado de tópicos, permitiendo así una comprensión más matizada y dinámica de los discursos que circulan en el espacio digital. Esta aproximación resulta especialmente útil para analizar dinámicas de polarización, cambios en el clima social y diferencias entre contextos geográficos, ofreciendo insumos valiosos para la formulación de políticas públicas basadas en evidencia.

Este tipo de análisis ha sido implementado en diversos contextos nacionales con el fin de examinar la construcción discursiva de actitudes hacia la población migrante en redes sociales. Un ejemplo representativo es el estudio de (Freire-Vidal, Y. & Graells-Garrido, E., 2019), que analiza el posicionamiento ideológico de los usuarios de *Twitter* (ahora X) en Chile frente al fenómeno migratorio durante el año 2017. A partir de una muestra de 206.353 publicaciones emitidas por 49.346 usuarios distintos, los autores exploraron dos dimensiones clave: la “empatía” y la “amenaza”, como formas contrastantes de representar la inmigración. Para capturar estas actitudes, se empleó el algoritmo *Topic-Supervised Non-Negative Matrix Factorization* (TS-NMF), una técnica de modelado temático que permite incorporar supervisión parcial a partir de ejemplos previamente definidos por el investigador. Esta metodología fue complementada con el análisis de *Term Frequency-Inverse Document Frequency* (TF-IDF), que permitió identificar los términos más característicos en función de su relevancia relativa dentro del corpus. Los resultados revelaron diferencias notables entre ambos grupos: mientras que los usuarios del grupo “empatía” tendían a utilizar términos como “integración”, “salud” o “educación”, y preferían la denominación “migrantes”, los del grupo “amenaza” recurrían con mayor frecuencia a palabras como “delincuentes”, “control” o “ilegales”, asociadas a representaciones negativas vinculadas al gasto público y la competencia laboral. Estas divergencias léxicas no solo reflejan marcos interpretativos opuestos, sino también tensiones sociales más profundas en torno a la convivencia y la inclusión.

En un contexto político particularmente polarizado, el estudio de (Calderón et al., 2020) profundiza en la forma en que ciertas actitudes negativas hacia la inmigración se radicalizan

y cristalizan en discursos de odio explícitos. Este trabajo se centró en el análisis de 244.095 publicaciones en español que contenían la palabra “Vox”, publicados entre noviembre de 2018 y abril de 2019, coincidiendo con el ascenso electoral del partido en España. Tras una fase inicial de codificación manual, se identificaron 1.977 publicaciones (aproximadamente el 0,8 %) como expresiones de odio dirigidas contra personas migrantes. Para su clasificación, se utilizó la taxonomía propuesta por (Miró-Llinares, 2016), que distingue entre tres niveles de peligrosidad: (1) ofensas contra valores individuales o colectivos (por ejemplo, insultos o estereotipos humillantes); (2) restricciones a derechos y libertades (como llamados a la expulsión o exclusión); y (3) incitación directa a la violencia física. La mayoría de las publicaciones identificadas se concentraron en los dos primeros niveles, con un 56 % de mensajes ofensivos y un 42 % que promovían restricciones, mientras que solo un 2 % incitaban a la violencia directa. Para explorar la estructura temática del discurso de odio, se aplicó el algoritmo *Latent Dirichlet Allocation* (LDA), configurado con cinco tópicos. Aunque el modelo presentó un bajo índice de coherencia, permitió identificar patrones relevantes en el lenguaje utilizado. Entre los tópicos extraídos destacaron temas como el control institucional (“PP”, “ilegales”), la criminalidad asociada a la inmigración (“delincuentes”), el acceso a ayudas públicas (“millones”, “expulsar”), el terrorismo (“negro”) y la representación de España como frontera africana (“invasión”).

Ampliando la perspectiva analítica hacia una dimensión comparativa transnacional, el estudio de (Rowe et al., 2021) constituye una contribución significativa al explorar las variaciones en las actitudes hacia la migración en el marco de crisis globales, como la pandemia de COVID-19. A diferencia de trabajos centrados exclusivamente en contextos nacionales, esta investigación analiza la evolución del discurso migratorio en Twitter durante la primera fase de la pandemia en cinco países: España, Alemania, Italia, Reino Unido y Estados Unidos. El análisis se basó en una muestra estratificada de 30,39 millones de tuits recolectados a través de la *Twitter Premium API*, aplicando filtros geográficos, lingüísticos y un proceso de traducción automática al inglés para estandarizar el corpus multilingüe. La metodología combinó el análisis de sentimientos, empleando el algoritmo VADER, con modelado de tópicos mediante LDA, lo que permitió identificar doce temas centrales del debate digital. Entre los más destacados se encontraron tópicos relacionados con abusos de derechos humanos, políticas migratorias en Estados Unidos y la situación de los migrantes vulnerables en Grecia y Turquía. En el caso particular de España, la narrativa predominante giró en torno al racismo y la xenofobia, concentrando el 33 % de los tuits analizados. Los resultados también mostraron un alto grado de polarización emocional, con un 37,6 % de mensajes marcadamente negativos frente a un 20 % de tono positivo, y escasa representación de discursos neutrales. No obstante, el estudio no identificó un incremento sostenido del sentimiento antiinmigrante, sino más bien picos emocionales vinculados a eventos coyunturales.

En esta misma línea de estudios que combinan análisis de sentimientos y modelado de tópicos, la investigación de (Tun-Mendicuti et al., 2024) se focaliza en el tratamiento discursivo

de las caravanas migrantes centroamericanas entre 2018 y 2021, contrastando dos contextos nacionales: México, como país de tránsito, y Estados Unidos, como país de destino. El estudio explora cómo variables mediáticas y geográficas influyen en la intensidad emocional del discurso expresado en redes sociales. Para ello, se analizaron 14.932 tuits geolocalizados, 6.740 procedentes de EE.UU. y 3.802 de México, complementados con datos de GDELT (*Global Database of Events, Language, and Tone*) sobre cobertura mediática diaria y estadísticas censales relativas a la distribución de migrantes. Al igual que el trabajo de (Rowe et al., 2021), esta investigación empleó una estrategia metodológica dual que integra el algoritmo VADER para el análisis de sentimientos con el modelado de tópicos mediante LDA. En el caso de Estados Unidos, los tópicos predominantes giraron en torno a la militarización de la frontera, discursos políticos y declaraciones del entonces presidente Donald Trump. Por su parte, en México sobresalieron temas vinculados a la asistencia humanitaria y los derechos humanos. En cuanto al análisis de polaridad emocional, se observaron diferencias significativas entre ambos países: mientras que en EE.UU. predominaban los mensajes negativos (media de  $-0.533$ ), en México las puntuaciones positivas y negativas mostraron una mayor simetría ( $-0.490$  y  $0.486$ , respectivamente). Un resultado destacado fue que, tras aplicar una codificación manual a una muestra aleatoria, se evidenció que la intensidad del sentimiento era un indicador más robusto que la polaridad para interpretar la postura del emisor. Finalmente, el análisis de variables explicativas mostró que, en EE.UU., ni la cobertura mediática ni los factores geográficos fueron estadísticamente significativos, mientras que en México sí se detectó una relación entre mayor exposición mediática y niveles más elevados de carga emocional, así como una menor polarización en estados fronterizos o con alta concentración de población migrante.

A diferencia de los estudios previamente revisados, que combinan enfoques no supervisados como LDA y análisis léxico con VADER para explorar patrones temáticos y emocionales en los discursos sobre migración, el trabajo de (Romero Vega, 2021) introduce un cambio metodológico relevante al adoptar un enfoque estrictamente supervisado. Este estudio se centra en la detección automática de contenido xenófobo en tuits relacionados con la migración venezolana en Ecuador, aplicando técnicas de clasificación dentro del marco de KDT (*Knowledge Discovery in Text*). Para ello, se recopilaron 9.888 publicaciones mediante la herramienta Twitter Scraper con un filtro geográfico específico, seguidos de un preprocesamiento que incluyó la eliminación de stopwords. Dado el desequilibrio de clases, con una proporción reducida de publicaciones xenófobas frente a mensajes ofensivos o neutros, se empleó la técnica SMOTE (*Synthetic Minority Over-sampling Technique*) para generar ejemplos sintéticos de la clase minoritaria. Además, se utilizó un conjunto de datos previamente etiquetado a través de *crowdsourcing*, siguiendo una taxonomía tripartita: contenido xenófobo, ofensivo y otros. El algoritmo que arrojó los mejores resultados fue SVM (*Support Vector Machine*), con una puntuación F1 del 94 %, superando a modelos como la regresión logística (93 %) y *Naive Bayes* (89 %). Este estudio destaca por su precisión en distinguir

discursos de odio específicos, más allá de la mera detección de lenguaje ofensivo,

Continuando con el análisis de representaciones discursivas sobre la migración venezolana en contextos latinoamericanos, el estudio de (Silva Aguilar and Ferré Pavia, 2024) se enfoca en la plataforma *Twitter* (ahora X) para examinar el debate político en Colombia. A través de un enfoque cualitativo basado en codificación manual, se analizaron 15.703 publicaciones originales extraídas de un corpus inicial de 160.095 mensajes que incluían hashtags relacionados con la temática migratoria. La clasificación se articuló mediante una tipología de encuadres que incluyó tres niveles: genéricos (ejes temáticos amplios como política o derechos), específicos (formas de representación del colectivo migrante) y narrativos (estructura discursiva). Entre los encuadres genéricos predominó el enfoque de “factores e implicaciones políticas” (26 %), seguido de “prescripción y evaluación de políticas públicas” (14,3 %) y “sentimiento público” (13,8 %), lo que evidencia una alta politización del discurso, enmarcado en el contexto electoral colombiano. En cuanto a los encuadres específicos, los migrantes venezolanos fueron representados principalmente como “víctimas”, ya sea de conflictos armados (25 %) o crisis humanitarias (17,3 %), aunque también emergieron representaciones negativas, como “amenaza al orden público” (15,7 %) o “amenaza fiscal” (11 %). Resulta significativo que el encuadre positivo de “héroe” solo aparezca en un 9,3 % de los casos, indicando una limitada visibilidad de narrativas empoderadoras. Desde la perspectiva narrativa, la mayoría de los mensajes se construyen en tercera persona y adoptan un punto de vista externo al sujeto migrante. Además, se constató que los tuits eran emitidos mayoritariamente por usuarios comunes, con escasa participación de actores institucionales, medios de comunicación y mujeres.

Por último, desde un enfoque más lingüístico y con un aporte metodológico innovador, el estudio de (Vitiugin, 2024) introduce la dimensión del *code-mixing* en las conversaciones digitales sobre migración. En este trabajo, se analizan publicaciones en Reddit procedentes de comunidades como r/Finland, r/Korea y r/GoingToSpain, con el objetivo de identificar mensajes que combinan dos idiomas, por ejemplo, inglés y finlandés, dentro de un mismo texto. Esta práctica, común en contextos migratorios, se asocia a menudo con la expresión de identidades híbridas y la adaptación cultural en entornos digitales. Para abordar esta tarea, los autores desarrollaron ELMICT (*Ensemble Learning for Multilingual Identification of Code-mixed Texts*), un modelo de clasificación que combina salidas de múltiples tokenizadores con un modelo preentrenado multilingüe (XLM-RoBERTa), ajustado específicamente para detectar patrones de *code-mixing*. Este enfoque superó en desempeño a soluciones tradicionales basadas en árboles de decisión y al modelo ChatGPT-3.5, alcanzando un F1-score del 97,6 %. Desde el punto de vista temático, los hilos con mayor incidencia de *code-mixing* abordaban asuntos de alta sensibilidad como defensa nacional, empleo, autenticación bancaria y geopolítica, especialmente vinculados a la guerra en Ucrania y la OTAN, aunque también se detectaron con frecuencia en temas cotidianos como vivienda, salud, energía o reciclaje. Los autores subrayan que la integración de modelos como ELMICT en servicios públicos digi-

tales, como chatbots gubernamentales, podría mejorar significativamente la accesibilidad y la confianza de los migrantes, al adaptarse mejor a sus patrones reales de comunicación en línea.

El análisis de estos estudios revela patrones claros, como se puede observar en la siguiente Tabla 2.1. El análisis conjunto de estos estudios permite identificar una serie de patrones comunes en la forma en que la migración es representada y debatida en las plataformas digitales. En primer lugar, el modelado de tópicos mediante LDA se ha consolidado como una herramienta relevante para estructurar y sintetizar los ejes discursivos predominantes, permitiendo detectar temáticas recurrentes vinculadas tanto a aspectos estructurales como a narrativas simbólicas. Paralelamente, el análisis de sentimientos, especialmente mediante el uso del algoritmo VADER, ha facilitado la evaluación de la carga emocional asociada a los discursos, revelando altos niveles de polarización y una convivencia constante entre expresiones de rechazo y solidaridad. Los enfoques supervisados, como los modelos SVM y *Naive Bayes*, han mostrado una alta eficacia en la identificación de contenido discriminatorio, particularmente en contextos caracterizados por tensiones sociopolíticas. Además, las investigaciones basadas en encuadres discursivos han destacado la oscilación entre la representación de los migrantes como víctimas y su caracterización como amenaza, dependiendo de las coyunturas políticas o del tratamiento mediático. Por último, los estudios recientes que incorporan dimensiones lingüísticas como el *code-mixing* han abierto nuevas líneas de análisis sobre las formas reales de expresión y adaptación comunicativa de las comunidades migrantes.

En este marco de referencia se sitúa la presente investigación, que busca aportar una comprensión empírica de la percepción pública sobre la inmigración venezolana en países de habla inglesa, tomando como fuente los discursos generados en *Reddit*. A partir de los resultados de la literatura existente, este trabajo se propone combinar técnicas de análisis de sentimientos y modelado de tópicos para examinar tanto la carga emocional como las principales narrativas que estructuran el debate digital en torno a la diáspora venezolana. Si bien estudios previos han explorado fenómenos migratorios específicos mediante enfoques cuantitativos y cualitativos, la mayoría se ha centrado en plataformas como *Twitter* (ahora X) o en contextos geográficos aislados. En contraste, este estudio adopta una perspectiva más amplia, al considerar múltiples comunidades y al emplear una plataforma menos explorada como *Reddit*, caracterizada por permitir una expresión discursiva más extensa y menos condicionada por el formato. De este modo, se busca contribuir al campo con una aproximación complementaria que permita identificar patrones emocionales y temáticos emergentes, así como ofrecer evidencia empírica útil para el diseño de intervenciones comunicativas y políticas públicas informadas.

Referencia	Tamaño del Data Set	Método de Adquisición	Objetivo	Algoritmo	Resultados
(Freire-Vidal, Y. & Graells-Garrido, E., 2019)	206,353 publicaciones de 49,346 usuarios	Filtrado por palabras clave y agrupación temática, recolección en 2017 en Chile	Analizar la percepción ciudadana (empatía vs amenaza) hacia inmigrantes	Modelado de tópicos: TS-NMF y TF-IDF	Identificaron dos grupos de usuarios: “empatía” con términos como “migrantes”, “salud” y “educación”; y “amenaza” con “inmigrantes”, “delincuentes” e “ilegales”.
(Calderón et al., 2020)	244,095 publicaciones con la palabra “Vox”	Filtrado por palabra clave “Vox”, codificación manual para clasificar discurso de odio	Identificar el discurso de odio contra inmigrantes en contexto político	Modelado de tópicos: LDA (5 tópicos)	Ofensas (56%), restricción de derechos (42%), violencia directa (2%). Tópicos: control institucional, criminalidad, ayudas públicas, terrorismo, invasión africana.
(Rowe et al., 2021)	30,39 millones de publicaciones en 5 países	Twitter (ahora X) Premium API (1.25M publicaciones mensuales), muestreo aleatorio estratificado, traducción automática	Comparar actitudes hacia migración durante COVID-19 en 5 países	Modelado de tópicos: LDA (12 tópicos) y análisis de sentimiento: VADER	Fuerte polarización emocional: España con 37.6% tuits negativos y 20% positivos. Tópicos: abusos DD.HH., Trump e inmigración, migrantes vulnerables.
(Tun-Mendicuti et al., 2024)	14,932 tuits geolocalizados (6,740 EE.UU. y 3,802 México)	Twitter Scraper, complementado con GDELT y censos nacionales	Examinar impacto geográfico y mediático en sentimientos sobre caravanas migrantes	Modelado de tópicos: LDA (4 tópicos por país) y análisis de sentimiento: VADER	EE.UU.: publicaciones más negativos (-0.533) que positivos (0.470). México: equilibrio. Cobertura mediática asociada a mayor carga emocional en México.
(Romero Vega, 2021)	9,888 publicaciones	Filtro geográfico en Twitter (ahora X), limpieza de datos y balanceo de clases (SMOTE)	Detectar contenido xenófobo contra venezolanos	Clasificación supervisada: SVM, Regresión logística, Naive Bayes	SVM alcanzó 94% F1-score. Detección precisa de tuits xenófobos, diferenciados de ofensivos/neutros.
(Silva Aguilar and Ferré Pavia, 2024)	15,703 publicaciones de un corpus de 160,095	Filtrado por hashtags migratorios, codificación manual	Analizar encuadres temáticos, específicos y narrativos sobre migración venezolana	Codificación manual (no ML)	Predominó el encuadre de “factores políticos” (26%), “prescripción de políticas” (14,3%) y “sentimiento público” (13,8%). Migrantes vistos como víctimas (25%) o amenaza (15,7%).
(Vitiugin, 2024)	No especifica número exacto	Análisis de publicaciones en Reddit (r/Finland, r/korea, r/GoingToSpain)	Detectar code-mixing en conversaciones migratorias	Modelo supervisado: ELMICT con XLM-RoBERTa	Alta precisión (F1 97,6%). Code-mixing frecuente en empleo, defensa y temas cotidianos. Potencial para uso en servicios públicos digitales.

Tabla 2.1: Artículos representativos en el análisis de percepción pública sobre inmigrantes usando técnicas de analítica de textos. Elaboración propia.

## Capítulo 3

# Diseño Metodológico para el Análisis de Percepción Pública en Redes Sociales

Este capítulo presenta la metodología empleada en el desarrollo del presente TFG, centrado en el análisis automatizado de publicaciones en la plataforma digital *Reddit*. Tal como se ha expuesto previamente, el objetivo principal de esta investigación es examinar la percepción pública sobre la inmigración venezolana, a partir de la recopilación y procesamiento de comentarios generados por usuarios de dicha plataforma. Los comentarios analizados son en inglés, provenientes principalmente de usuarios en Estados Unidos. Para alcanzar este propósito, se ha diseñado un procedimiento estructurado en cinco fases principales, representadas esquemáticamente en la Figura 3.1, que se describen en detalle a lo largo del capítulo: (1) recolección y preparación de los datos, (2) pre-procesamiento textual, (3) análisis descriptivo mediante N-gramas, (4) modelado de tópicos y (5) análisis de sentimiento.

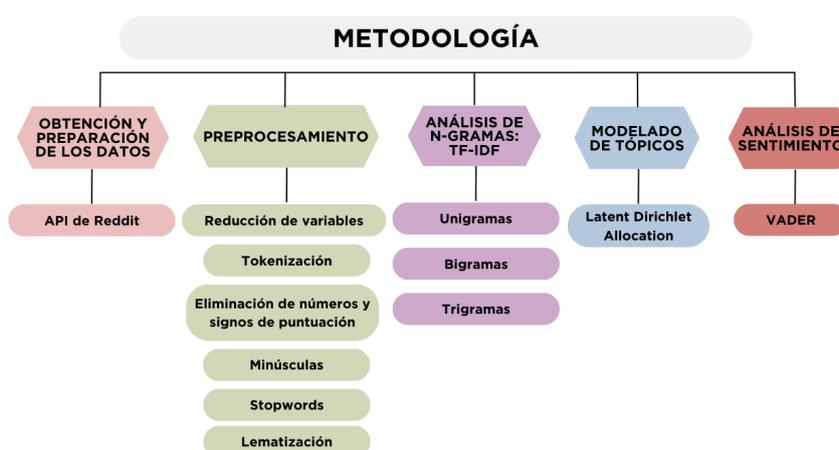


Figura 3.1: Metodología de análisis  
Elaboración propia.

### 3.1. Recolección y preparación de los datos

Para llevar a cabo el análisis de la percepción pública sobre la migración venezolana en la plataforma Reddit, se diseñó un proceso sistemático de recolección y preparación de datos, basado en el acceso a la interfaz de programación de aplicaciones (API) de Reddit mediante la biblioteca `asyncpraw` en el lenguaje de programación Python. Esta herramienta permite realizar solicitudes de manera eficiente y asíncrona, optimizando el tiempo de extracción de grandes volúmenes de información. La configuración inicial incluyó las credenciales necesarias para garantizar una autenticación segura y respetar las políticas de uso de la plataforma.

La recolección de datos abarcó un periodo de cinco meses desde octubre del 2024 hasta febrero del 2025 incluidos, con el objetivo de capturar una muestra representativa de las discusiones generadas en torno a este fenómeno migratorio. Este intervalo temporal se seleccionó debido a la alta intensidad de los flujos migratorios venezolanos durante dicho periodo, así como a la creciente atención pública e institucional que ha recibido esta problemática. Asimismo, el contexto político actual, como la exclusión de los migrantes venezolanos del programa de TPS en Estados Unidos y las recientes políticas de deportación tras el retorno de Donald Trump al poder, destaca la pertinencia de analizar las expresiones espontáneas de la ciudadanía en redes sociales, dado su valor como termómetro de opinión y herramienta para comprender la evolución de las actitudes públicas frente a la migración. Estos grandes cambios que viven los venezolanos concretamente en Estados Unidos se ven completamente reflejados en el análisis ya que captura la situación antes y después de las elecciones estadounidenses. Es importante recalcar que el análisis es de comentarios sobre migrantes venezolanos en países de habla inglesa, pero muchos reflejan la situación en Estados Unidos.

Desde un punto de vista operativo, la recolección se realizó mediante la búsqueda de publicaciones y comentarios que incluyeran términos clave como “migración” y “Venezuela”. Para limitar la búsqueda al marco temporal definido, se utilizó un sistema de filtrado por marcas de tiempo (*timestamps*), ejecutando el proceso de manera iterativa día a día en sentido descendente desde la fecha actual. Además, se estableció un número máximo de publicaciones por día, tanto para evitar sobrecargar la API como para cumplir con sus restricciones técnicas.

Posteriormente, se procedió a la depuración estructural del conjunto de datos, conservando únicamente aquellas columnas relevantes para el análisis, tales como el contenido textual del comentario, la fecha de publicación o el nombre del subreddit. Esta reducción tiene como finalidad optimizar la eficiencia computacional en las fases posteriores del procesamiento, eliminando información redundante o irrelevante para los objetivos del estudio. Asimismo, a partir de la variable de fecha, se generaron variables derivadas como el año, mes y día de publicación. Esta transformación permite estructurar el análisis temporal con mayor granularidad, facilitando la identificación de tendencias, fluctuaciones o posibles eventos coyunturales.

Complementariamente, se llevó a cabo una exploración descriptiva de la distribución temporal y estructural de los datos. Se analizó la frecuencia de publicaciones a lo largo del tiempo para detectar picos de actividad asociados a eventos relevantes, y se examinó la longitud de los comentarios como indicador de la extensión y complejidad de las intervenciones discursivas. Además, se exploró el uso de *hashtags* en el corpus, generando un ranking de las etiquetas más frecuentes. Esta información permitió identificar los temas más recurrentes y las entidades que concentran la atención de los usuarios. Finalmente, se identificaron los *subreddits* más activos en términos de volumen de participación, permitiendo reconocer las comunidades virtuales con mayor dinamismo en el tratamiento del fenómeno migratorio.

Con el fin de garantizar la coherencia lingüística del análisis y la adecuada aplicación de las técnicas de PLN, se procederá a identificar el idioma original de cada publicación contenida en el corpus. Esta etapa es necesaria, dado que muchas herramientas de análisis semántico y sintáctico, así como diversos algoritmos de modelado, presentan un desempeño diferencial según el idioma de entrada. La selección del idioma de trabajo se basará en criterios cuantitativos, priorizando aquella lengua que concentre el mayor volumen de comentarios. En función de la naturaleza del corpus y del carácter internacional de la plataforma Reddit, es previsible que el inglés represente la proporción dominante.

En síntesis, las tareas de exploración preliminar permiten establecer una base para el análisis posterior, al proporcionar una caracterización inicial del comportamiento del corpus tanto en términos cuantitativos como contextuales. La identificación de patrones temporales, la evaluación de la extensión de los comentarios, el reconocimiento de temáticas recurrentes y la detección de comunidades digitales relevantes, constituyen elementos importantes para orientar las decisiones metodológicas en etapas subsiguientes.

## 3.2. Pre-procesamiento del texto

Tras la etapa de análisis exploratorio, se procedió al pre-procesamiento textual del corpus con el fin de garantizar la calidad de los resultados en las fases analíticas posteriores. El primer aspecto considerado fue el tratamiento de posibles duplicados. En el contexto de plataformas digitales como *Reddit*, la repetición de publicaciones, frases o expresiones no siempre responde a errores técnicos o redundancias indeseadas, sino que puede constituir un reflejo genuino de dinámicas comunicativas colectivas. Fenómenos como el copy-paste, el eco discursivo, las campañas coordinadas o el uso reiterado de ciertas fórmulas retóricas representan mecanismos mediante los cuales las comunidades digitales refuerzan mensajes, enfatizan posturas o consolidan narrativas compartidas.

Desde esta perspectiva, la eliminación sistemática de textos duplicados podría suponer una pérdida significativa de información, distorsionando la manera en que se expresan y refuerzan determinados tópicos dentro del discurso colectivo. Lejos de considerarse como

ruido, la presencia de duplicados puede ser interpretada como un indicador de temática o resonancia social. Por este motivo, y tras comprobar que dichas repeticiones no obedecían a errores durante el proceso de extracción ni a artefactos introducidos en fases anteriores, se decidió conservar estos registros.

Posteriormente, se implementó una función de limpieza para la depuración lingüística del texto. Esta función incluyó la eliminación de elementos estructuralmente irrelevantes como enlaces (URLs), menciones a usuarios, hashtags, cifras numéricas y palabras excesivamente breves (de uno o dos caracteres), así como la corrección de posibles errores de codificación y la normalización de espacios en blanco. Asimismo, se amplió el conjunto estándar de stop-words en inglés incorporando términos adicionales cuya presencia frecuente podría interferir en la detección de patrones discursivos significativos (por ejemplo, “like”, “also”, “really”, “thing”).

A continuación, se procedió a la lematización del corpus con el propósito de reducir cada término a su forma base, lo cual facilita la unificación de expresiones semánticamente equivalentes que podrían presentarse con variaciones morfológicas. Esta etapa es particularmente relevante en el análisis de discursos sobre migración, ya que permite agrupar bajo una misma raíz conceptos clave que, de otro modo, podrían fragmentarse. Por ejemplo, formas como “migrants”, “migrating” y “migration” fueron normalizadas a “migrate”, mientras que términos como “arrivals”, “arrived” y “arriving” se redujeron a “arrive”. Esta transformación permite una representación más precisa del contenido temático y mejora la consistencia de los resultados en las etapas analíticas posteriores. De esta manera, la versión lematizada del texto resultante es la entrada estandarizada para las siguientes fases de análisis, como la extracción de n-gramas, el modelado de tópicos y el análisis de sentimientos.

### 3.3. Análisis de N-Gramas

Una vez completadas las fases de adquisición y preprocesamiento del corpus textual, se procede a realizar un análisis exploratorio basado en la extracción de secuencias léxicas recurrentes. Para ello, se implementa una técnica fundamental en minería de texto: el análisis de *N-gramas*, entendido como la identificación de secuencias consecutivas de  $n$  palabras dentro del texto. Esta estrategia permite detectar patrones lingüísticos recurrentes, relaciones sintácticas frecuentes y unidades de significado compuestas, que resultan útiles para explorar la estructura semántica de los discursos en redes sociales. El análisis se desarrollará en distintos niveles de granularidad: los *unigramas* permiten identificar los términos más frecuentes de manera aislada; los *bigramas* exploran coocurrencias de dos palabras consecutivas; y los *trigramas* permiten detectar expresiones más complejas. Estas combinaciones son especialmente relevantes para caracterizar temas de discusión, detectar frases recurrentes o incluso identificar construcciones idiomáticas propias del lenguaje.

A partir de la extracción de N-gramas, se calculará una métrica de relevancia que no se limita al conteo bruto de frecuencias. Si bien la frecuencia absoluta proporciona una primera aproximación al peso de un término o expresión, se recurre a una medida más robusta y ampliamente utilizada: la métrica *Term Frequency - Inverse Document Frequency* (TF-IDF). Esta técnica estadística permite ponderar la importancia relativa de una palabra o secuencia dentro de un conjunto de documentos, penalizando aquellas que aparecen frecuentemente y realzando aquellas más distintivas. La métrica TF-IDF se construye a partir del producto entre dos componentes. El primero, la Frecuencia del Término (TF), cuantifica las veces que aparece un término específico  $t$  en un documento  $d$ , ajustado por la longitud del mismo:

$$TF(t, d) = f(t, d) \quad (3.1)$$

El segundo componente, la Frecuencia Inversa del Documento (IDF), busca reducir el peso de aquellos términos que aparecen de forma recurrente en la mayoría de los documentos. Se calcula como el logaritmo del cociente entre el número total de documentos  $N$  y el número de documentos que contienen el término  $t$ :

$$IDF(t, N) = \log \left( \frac{N}{df_t} \right) \quad (3.2)$$

Finalmente, la métrica TF-IDF se obtiene al multiplicar ambas medidas:

$$TF-IDF = TF(t, d) \times IDF(t, N) \quad (3.3)$$

Este cálculo proporciona una puntuación que refleja la relevancia relativa de un término dentro de un documento, en función de su frecuencia local y su distribución global en el corpus. En términos interpretativos, un valor elevado de TF-IDF indica que el término es característico del documento en el que aparece, ya que se presenta con frecuencia en ese texto pero con escasa presencia en el resto de los documentos analizados. Por el contrario, una puntuación baja sugiere que la palabra es común en múltiples documentos y, por tanto, aporta poca información diferenciadora. En el contexto de este estudio, esta métrica resulta útil para destacar expresiones significativas que surgen en publicaciones individuales sobre migración venezolana, pero que no están presentes de manera sistemática en el conjunto general de comentarios.

### 3.4. Modelado de tópicos

El modelado de tópicos constituye una metodología clave en el análisis automatizado de textos, cuyo objetivo es identificar estructuras latentes de significado a partir de grandes volúmenes de datos no estructurados. Este enfoque permite descomponer un corpus en temáticas recurrentes, facilitando una comprensión más profunda del contenido discursivo y

de las narrativas implícitas en el lenguaje utilizado por los usuarios. En este estudio, se ha implementado con el fin de detectar los principales ejes temáticos asociados a la percepción pública sobre la migración venezolana en publicaciones extraídas de la plataforma Reddit.

Una de las técnicas más ampliamente utilizadas para esta tarea es la Latent Dirichlet Allocation (LDA), un modelo generativo de tipo probabilístico no supervisado, propuesto por (Blei et al., 2003). LDA asume que cada documento es una combinación de varios tópicos, y que cada tópico está representado por una distribución de probabilidad sobre palabras. Este modelo es especialmente valioso en contextos en los que no se dispone de anotaciones manuales, ya que infiere automáticamente patrones temáticos a partir de los datos sin necesidad de etiquetas previas (Kibe, 2024; Murel and Kavlakoglu, 2024).

En el marco de este trabajo, se implementa el modelo LDA para extraer tópicos a partir de comentarios en inglés de usuarios de Reddit relacionados con la migración venezolana. La Figura 3.2 muestra una representación gráfica del funcionamiento interno del modelo LDA. El modelo parte de un corpus de documentos  $D$ , en este caso, publicaciones de Reddit, donde cada documento está formado por una secuencia de  $N$  palabras ( $w$ ). El modelo introduce dos parámetros de Dirichlet:  $\alpha$ , que controla la distribución de tópicos por documento ( $\theta$ ), y  $\beta$ , que regula la distribución de palabras por tópico ( $\varphi$ ). Un valor alto de  $\alpha$  implica que cada documento incluirá una mezcla más diversa de tópicos, mientras que un valor elevado de  $\beta$  sugiere que los tópicos estarán compuestos por un mayor número de palabras con probabilidad significativa. El parámetro  $K$  representa el número total de tópicos definidos a priori, y es una decisión crítica para la calibración del modelo, pues influye directamente en la granularidad de los resultados obtenidos. A lo largo del proceso de generación, cada palabra  $w$  se asocia a un tópico  $z$  mediante una asignación estocástica, y a su vez, cada tópico se genera en función de la distribución  $\varphi$ . La variable latente  $\theta$ , por su parte, describe la proporción de tópicos en cada documento.

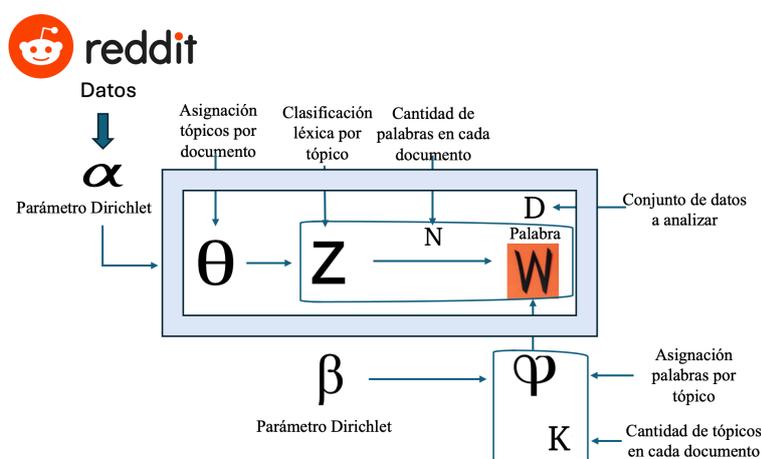


Figura 3.2: Representación gráfica del modelo LDA  
Elaboración propia.

El corpus final se transforma en un conjunto de documentos representados como bag-of-words, y se ajustan diversos modelos LDA variando sistemáticamente el parámetro  $K$  para identificar el número óptimo de tópicos. La elección del valor de  $K$  constituye una etapa fundamental en la configuración del modelo, ya que define el número de tópicos que se buscarán en el corpus. Para determinar el valor óptimo, se emplea el índice de coherencia, una métrica diseñada para evaluar la cohesión semántica de los tópicos generados. Este índice mide la similitud entre las palabras más representativas de cada tópico, basándose en su co-ocurrencia dentro del corpus. Las variantes más utilizadas del índice de coherencia incluyen la Normalized Pointwise Mutual Information (NPMI) y la similitud del coseno, las cuales permiten estimar la afinidad semántica entre términos clave (Mimno, D., McCallum, A. and others, 2011). Una mayor puntuación en esta métrica indica una mayor calidad semántica del modelo, ya que refleja tópicos internamente consistentes y conceptualmente significativos.

Para determinar el número óptimo de tópicos  $K$ , se procederá a ajustar múltiples modelos LDA variando sistemáticamente este parámetro. El objetivo es identificar el valor de  $K$  que proporcione el mejor equilibrio entre calidad interpretativa y simplicidad del modelo. En otras palabras, se busca una configuración que capture adecuadamente la diversidad temática del corpus sin introducir un número excesivo de tópicos que dificulte la interpretación o genere redundancias semánticas. Esta elección se basará en la maximización del índice de coherencia, complementada con un análisis cualitativo de los resultados obtenidos para cada valor de  $K$ . Al seleccionar el modelo que ofrezca la mayor coherencia temática con el menor número de tópicos posible, se garantiza un balance adecuado entre desempeño y claridad interpretativa, lo cual es necesario para los objetivos exploratorios de este estudio.

Complementariamente, se calcula la distancia intertópica lo que permite visualizar la disimilitud entre los distintos tópicos generados y comprobar que estén bien separados temáticamente. Esta medida resulta útil para evitar la redundancia temática dentro del modelo y asegurar que los tópicos generados representen dimensiones diferenciadas del discurso analizado. Dicha distancia se visualiza frecuentemente mediante técnicas de reducción de dimensionalidad como *t-Distributed Stochastic Neighbor Embedding* (t-SNE) o *Principal Component Analysis* (PCA), que permiten proyectar el espacio temático en dos dimensiones. En estas representaciones, cada tópico es ilustrado como un círculo cuya superficie es proporcional a su peso en el corpus, y cuya proximidad respecto a otros círculos indica su grado de solapamiento semántico (Blei and Lafferty, 2007). Como criterio adicional de validación, se comprobará que los círculos estén lo más separados posible, ya que una mayor separación sugiere una menor superposición semántica entre los tópicos, lo cual garantiza una representación temática más diferenciada y, por tanto, una segmentación más fiable del contenido discursivo.

### 3.5. Análisis de sentimientos

El análisis de sentimientos es una técnica central en el procesamiento del lenguaje natural que permite identificar la polaridad emocional de un texto, positiva, negativa o neutral, y, con ello, aproximarse a las actitudes implícitas en los discursos. Su utilidad radica en la capacidad de procesar de forma automatizada grandes volúmenes de contenido generado por usuarios, facilitando la evaluación del clima afectivo en espacios digitales. En el contexto del presente estudio, esta metodología se aplica a comentarios en inglés publicados en *Reddit* sobre la migración venezolana en países de habla inglesa, con el fin de detectar patrones discursivos, estimar la carga afectiva de los mensajes y explorar las narrativas sociales asociadas a este fenómeno.

Como se detalla en la sección 2.3, una de las metodologías más extendidas para el análisis automatizado de sentimientos es el uso de diccionarios léxicos, los cuales contienen conjuntos de palabras asociadas a etiquetas emocionales predefinidas. Estas herramientas permiten asignar una valoración subjetiva al contenido textual sin necesidad de datos previamente etiquetados. En el presente trabajo se ha seleccionado el diccionario VADER (*Valence Aware Dictionary and Sentiment Reasoner*) como principal recurso para la evaluación de los sentimientos expresados en los comentarios recopilados en *Reddit*. VADER es un sistema basado en reglas y orientado a textos informales, diseñado específicamente para la interpretación de opiniones en redes sociales. Su fortaleza radica en su capacidad para incorporar elementos del lenguaje digital contemporáneo, como emoticonos, puntuación enfática, uso de mayúsculas y modificadores léxicos, permitiendo una detección más precisa de la intención emocional del mensaje (Pano, 2020). Además, a diferencia de otros enfoques supervisados, VADER no requiere entrenamiento previo sobre corpus etiquetados, lo que lo convierte en una herramienta idónea para contextos de análisis exploratorio y no supervisado.

El sistema asigna a cada palabra del texto un valor de sentimiento en una escala que oscila entre -4 (extremadamente negativo) y +4 (extremadamente positivo), permitiendo estimar tanto la polaridad (positiva, negativa o neutral) como la intensidad emocional. Para ajustar esta valoración al contexto del mensaje, VADER aplica cinco reglas fundamentales (Hutto, 2014):

1. Uso de signos de puntuación: Los signos de exclamación intensifican la emoción sin alterar su orientación. Por ejemplo, la frase “This coffee is amazing!” transmite una intensidad emocional mayor que “This coffee is amazing”.
2. Mayúsculas: El uso de palabras completamente en mayúsculas enfatiza el sentimiento. Así, “I am VERY happy with the results” comunica una emoción más intensa que “I am very happy with the results.”
3. Modificadores de grado: Palabras que amplifican o atenúan el sentimiento, como “ex-

tremely”, “slightly”, “barely”, afectan la fuerza emocional del mensaje. Por ejemplo, “This movie was absolutely brilliant” tiene una carga afectiva más fuerte que “This movie was brilliant”, mientras que “This movie was somewhat interesting” reduce dicha carga.

4. Conjunción adversativa “but”: Esta conjunción actúa como un modificador estructural que desplaza el peso semántico hacia la segunda parte de la oración. En la frase “The hotel was nice, but the staff was rude”, el sentimiento negativo sobre el personal domina la percepción general.
5. Negaciones: Las negaciones invierten o mitigan la polaridad de los términos adyacentes. Así, “The movie was not bad” tiene un significado más positivo que “The movie was bad”.

Finalmente, VADER calcula una puntuación compuesta a partir de la suma de las valoraciones individuales ajustadas por las reglas anteriores. Esta puntuación se normaliza en un rango continuo entre -1 y 1, donde los valores próximos a -1 indican una fuerte carga negativa, los valores cercanos a +1 reflejan una carga positiva significativa, y los valores en torno a 0 se interpretan como neutralidad afectiva.

Por último, es importante destacar que, con el fin de optimizar la precisión del análisis de sentimientos mediante VADER, se han considerado ajustes específicos en la etapa de preprocesamiento de los datos textuales. A diferencia de otras técnicas de PLN que requieren una limpieza más detallada, VADER es particularmente sensible a características del lenguaje informal presentes en redes sociales, como la puntuación enfática, los emoticonos, el uso de mayúsculas y ciertos modificadores léxicos. Por esta razón, en este estudio se conservarán intencionadamente elementos como signos de exclamación, emojis y *stop words*, ya que son componentes necesarios para la correcta interpretación del tono emocional del mensaje. Asimismo, se evitará la conversión del texto a minúsculas, preservando la capitalización original que VADER utiliza como un indicio de énfasis emocional. No obstante, se eliminarán elementos considerados irrelevantes para el análisis, tales como *hashtags*, enlaces web, menciones a otros usuarios y cadenas numéricas aisladas, con el propósito de reducir el ruido semántico en el cálculo de la puntuación compuesta. Esta estrategia de preprocesamiento, alineada con las recomendaciones metodológicas del desarrollador de VADER, busca garantizar una evaluación más fiel de los sentimientos expresados en comentarios informales publicados en plataformas sociales.

# Capítulo 4

## Resultados

El presente capítulo expone los resultados derivados de la implementación de la metodología de análisis detallada en el Capítulo 3. A través de un enfoque basado en técnicas de PLN, se analizaron publicaciones de la plataforma *Reddit* relativas al fenómeno de la migración venezolana. Todo el código empleado para el preprocesamiento textual, la aplicación de modelos y la visualización de resultados ha sido desarrollado en Python y se encuentra disponible en un repositorio público para su consulta <sup>1</sup>.

### 4.1. Adquisición y preparación de los datos

Para el presente estudio se recopilieron todos los comentarios publicados en *Reddit* entre septiembre de 2024 y febrero de 2025, conformando un corpus con una delimitación temporal acotada pero representativa para los objetivos analíticos planteados. La elección de este periodo responde a su alta relevancia sociopolítica, al coincidir con la campaña presidencial de Donald Trump en Estados Unidos y con el debate público originado por el anuncio de la cancelación del TPS para ciudadanos venezolanos, hecho público en febrero de 2025 (Citizenship and Services, 2025). Ambos acontecimientos generaron una intensa cobertura mediática y un aumento significativo en las discusiones digitales sobre migración, identidad nacional y seguridad, lo que convierte este marco temporal en especialmente propicio para el análisis de las narrativas sociales compartidas en espacios digitales. La recopilación de los datos se llevó a cabo mediante la *Reddit API*, utilizando una serie de palabras clave relacionadas con la temática migratoria, tales como “migrantes y venezuela”. Esta estrategia permitió extraer un volumen significativo de contenido relevante para el estudio. El corpus final está compuesto por un total de 30.904 comentarios, que constituyen la base empírica tanto del análisis descriptivo como del modelado temático y del análisis de sentimientos.

A continuación, se analizó la evolución temporal de la conversación digital mediante un análisis del volumen mensual de comentarios recogidos durante el periodo comprendido en-

---

<sup>1</sup><https://github.com/juliajenarobarrio/Julia-Jenaro-TFG>

tre septiembre de 2024 y febrero de 2025. Tal como se aprecia en la Figura 4.1, la actividad discursiva en la plataforma experimentó un crecimiento sostenido a lo largo de estos meses. Durante septiembre, octubre y noviembre de 2024, la conversación digital muestra un comportamiento fluctuante con picos moderados, sin superar en la mayoría de los días los 500 comentarios. Esta fase inicial refleja una participación aún limitada, posiblemente vinculada a eventos políticos preliminares y debates en redes incipientes. No obstante, a partir de enero de 2025 y con mayor intensidad durante febrero de 2025, se observa un repunte significativo en la frecuencia de publicaciones, alcanzando su punto máximo con más de 3.000 comentarios en un solo día. En particular, el 1 de febrero se documentaron 3.125 comentarios, seguido de 1.000 el 4 de febrero y 1.875 el 11 del mismo mes. Un análisis cualitativo exploratorio sobre el contenido de los comentarios publicados en estos días revela un notable aumento en las referencias a temas asociados con deportaciones masivas, noticias de *Fox News* sobre migrantes venezolanos, el anuncio por parte del presidente Trump de deportar a más de 10.000 personas de dicha nacionalidad, así como menciones a supuestos vuelos de repatriación y criminalización de la población venezolana en EE.UU.

Este pico discursivo coincide con un acontecimiento de gran relevancia en la agenda migratoria estadounidense: el anuncio oficial, realizado a comienzos de febrero de 2025, de la finalización del TPS para ciudadanos venezolanos que habían sido acogidos bajo la designación de 2023, tras la eliminación del Título 42. La medida fue promulgada el 5 de febrero, estableciendo como fecha de expiración el 7 de abril de 2025, momento en el cual miles de beneficiarios perdieron su estatus legal y su derecho a trabajar, salvo que tuvieran acceso a otras vías de regularización. Si bien se interpusieron recursos legales para retrasar o suspender la medida, la fecha final no fue modificada. Este contexto explica el aumento exponencial de la conversación digital en torno a la migración venezolana durante dicho mes.

Además del análisis temporal, se realizó una exploración de las características lingüísticas del corpus para evaluar la diversidad idiomática presente en las publicaciones. La Figura 4.2 muestra la distribución de los comentarios según el idioma detectado automáticamente. Como se observa, el inglés constituye el idioma predominante, con más de 13.000 publicaciones, lo que confirma su centralidad en las discusiones sobre migración venezolana dentro de la plataforma Reddit. Le siguen, en volumen, el *igbo* y el español, con cifras superiores a los 10.000 y 5.000 comentarios respectivamente. Asimismo, se identificaron contribuciones marginales en otros idiomas, como luxemburgués y alemán, aunque su frecuencia fue escasa. No obstante, cabe señalar que una inspección manual de los comentarios catalogados como escritos en *igbo* mostró que muchos de ellos estaban vacíos o carecían de contenido textual significativo. Este resultado sugiere la existencia de errores en los procesos automáticos de detección del idioma, posiblemente debido a entradas generadas por spam, plantillas repetitivas o estructuras textuales no reconocibles por los modelos de clasificación lingüística. En contraste, el español representa una proporción significativa del corpus, pero no alcanza el volumen ni la consistencia observada en las publicaciones en inglés.

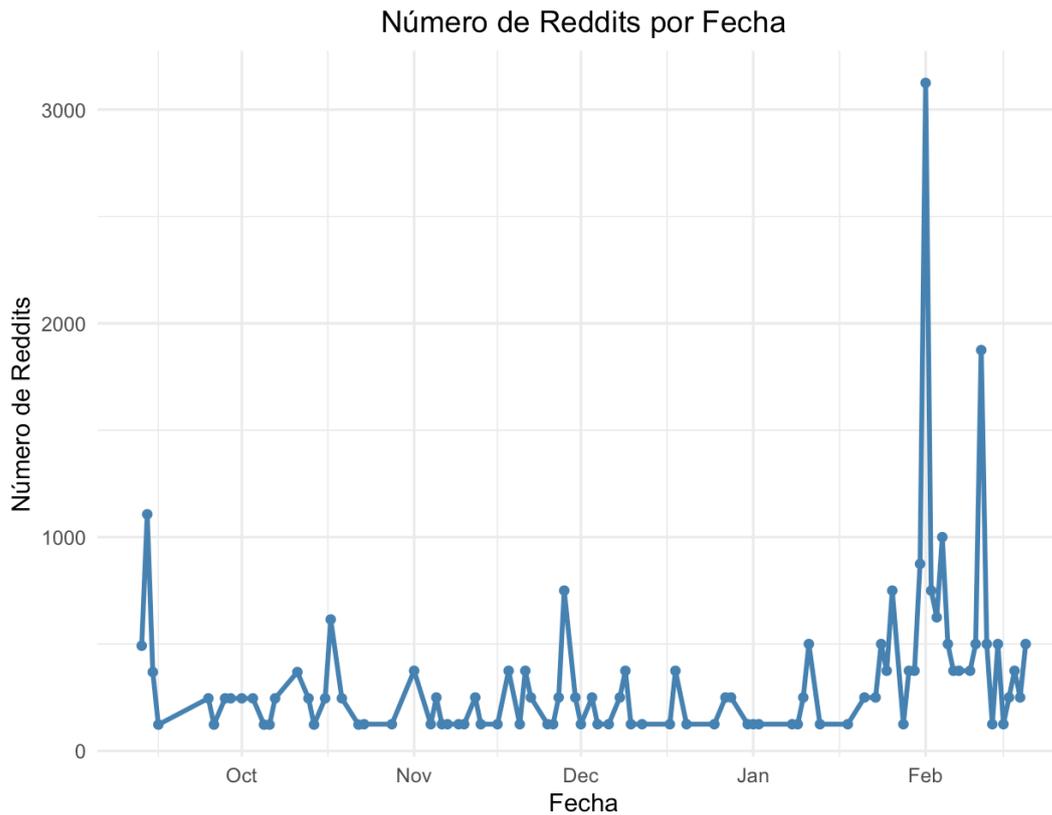


Figura 4.1: Comentarios diarios: Septiembre 2024-Febrero 2025  
Fuente de datos: (Reddit, 2024). Elaboración propia.

Por estos motivos, y considerando la necesidad de garantizar la coherencia semántica y metodológica de los análisis posteriores, se ha decidido acotar el estudio al conjunto de comentarios redactados en inglés. Esta decisión no solo responde a criterios de representatividad, dado que el inglés es la lengua más frecuente en el corpus, sino también a que trabajar con un solo idioma permite aplicar técnicas de PLN de forma más precisa, evitando errores derivados de ambigüedades multilingües o de inconsistencias sintácticas entre lenguas. Además, facilita la interpretación comparativa de resultados en análisis como el modelado de tópicos o el análisis de sentimientos, los cuales se sustentan en vocabularios sensibles a la estructura y al matiz lingüístico.

Una vez delimitado el corpus, se procedió a una caracterización adicional de las propiedades estructurales del texto. Entre ellas, la longitud de los comentarios constituye una métrica relevante, ya que permite explorar la extensión típica de las intervenciones y anticipar el grado de elaboración discursiva de los usuarios. La Figura 4.14 muestra la distribución de la longitud de los comentarios, medida en número de caracteres. Se observa que la mediana se sitúa en torno a los 2.000 caracteres, lo cual sugiere una predominancia de mensajes breves o de extensión moderada, consistentes con el formato habitual de interacción en *Reddit*. La mayor concentración de observaciones se ubica por debajo de los 4.000 caracteres,

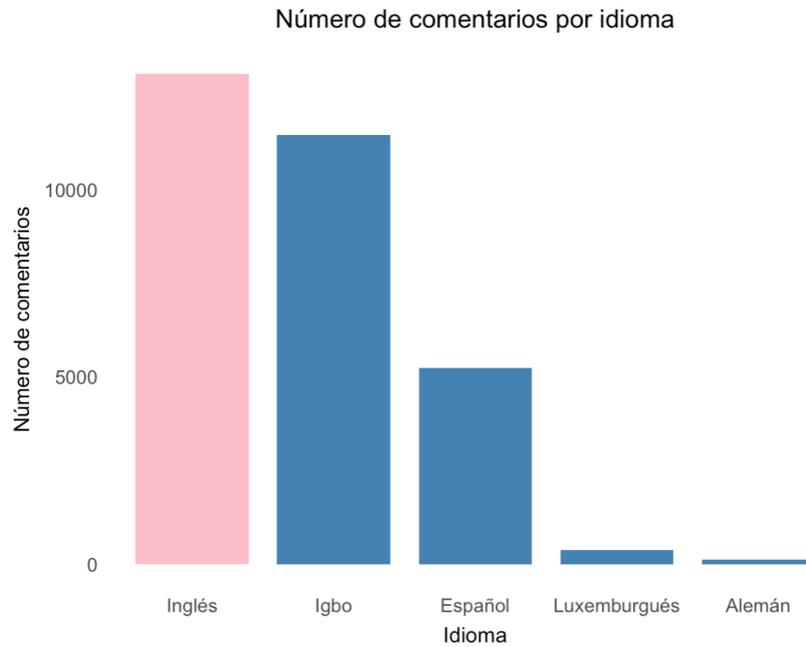


Figura 4.2: Número de comentarios por idioma  
Elaboración propia.

mientras que un número reducido de comentarios presenta longitudes superiores a los 10.000 caracteres. Estos casos atípicos podrían estar vinculados a publicaciones narrativas extensas, testimonios personales detallados o discusiones argumentativas más desarrolladas.

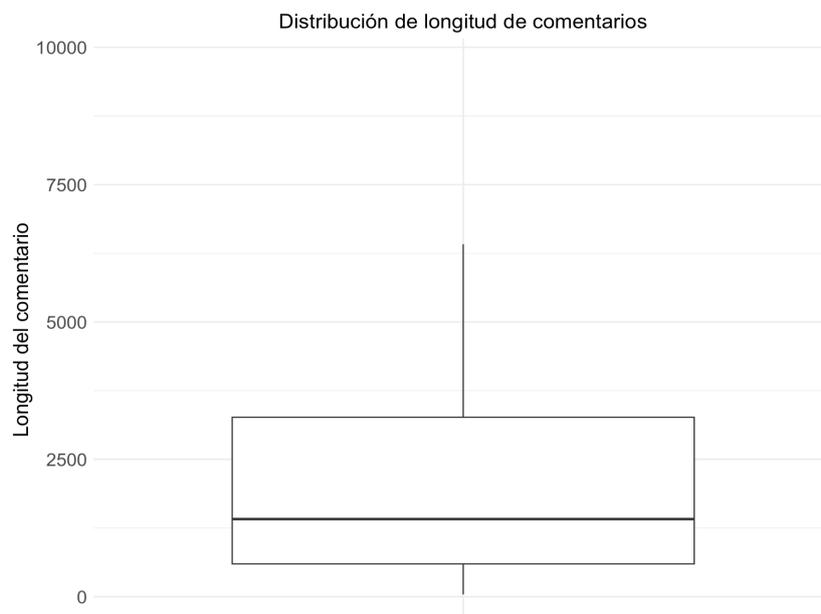


Figura 4.3: Distribución de la longitud de los comentarios  
Elaboración propia.

Complementariamente, se evaluó la distribución temática de las discusiones a través de

la identificación de los *subreddits* con mayor volumen de participación durante el periodo de estudio. La Figura 4.4 muestra los veinte *subreddits* que concentraron el mayor número de comentarios. Destaca de forma prominente el subreddit *vzla*, que aglutina más de 2.500 publicaciones, posicionándose como el principal espacio de conversación en torno a este fenómeno. Le siguen *AutoNewspaper* y *neoliberal*, con aproximadamente 1.000 comentarios cada uno, así como otros subreddits relevantes como *WhatTrumpHasDone*, *ABCWorldNews*, *Conservative* y *Miami*, que reúnen entre 400 y 600 comentarios respectivamente.

La presencia de estos espacios refleja la diversidad temática desde la cual se aborda la cuestión migratoria en *Reddit*. Mientras que subreddits como *vzla* y *Miami* tienden a reunir testimonios personales, noticias de interés comunitario o debates sobre políticas migratorias con impacto directo en la población venezolana, otros como *neoliberal* y *Conservative* articulan discusiones ideológicas más amplias sobre economía, gobernanza y seguridad, reflejando distintas posturas políticas ante la migración. De manera similar, *WhatTrumpHasDone* y *ABCWorldNews* concentran atención mediática sobre eventos coyunturales, aportando un enfoque más informativo o noticioso.

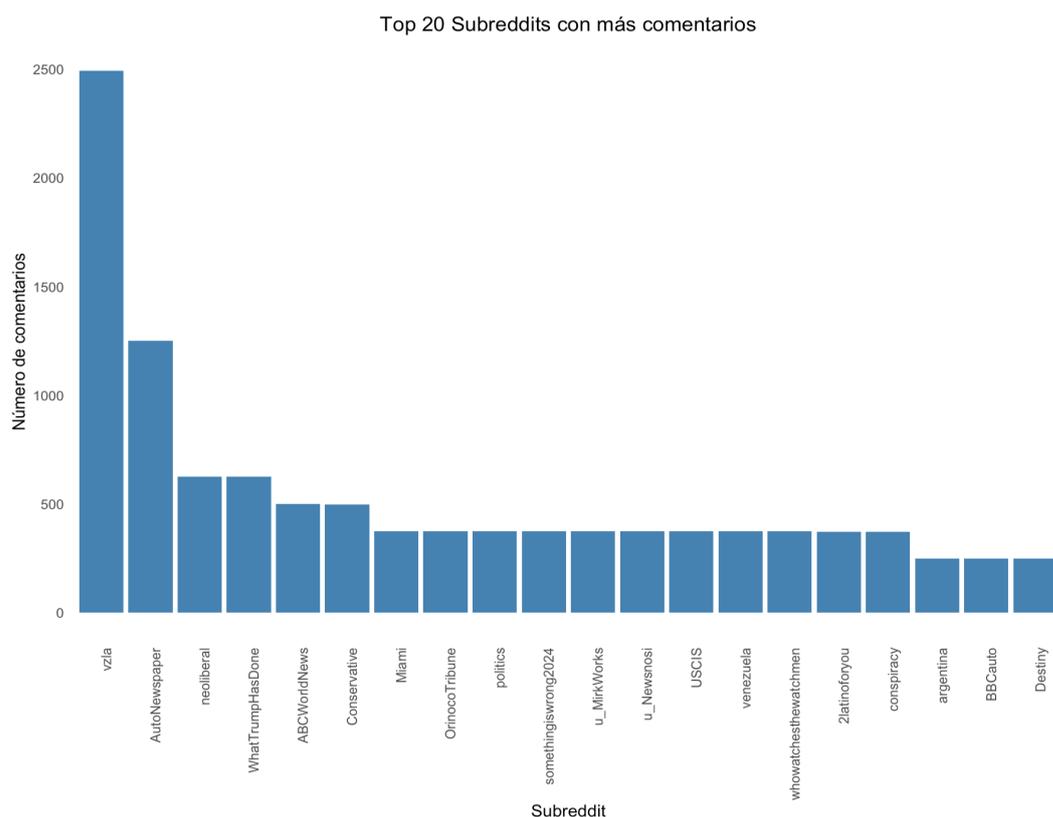


Figura 4.4: Top 10 Subreddits con más comentarios  
Elaboración propia.

Con el fin de profundizar en la identificación de tópicos recurrentes y entidades centrales dentro del discurso digital, se realizó un análisis de frecuencia de los *hashtags* empleados por

los usuarios en los comentarios. Esta estrategia permite detectar términos clave que sintetizan temas de conversación, expresan posturas ideológicas o refieren a eventos mediáticos de alta relevancia. En esta línea, se construyó un *ranking* con los veinte *hashtags* más utilizados a lo largo del periodo analizado, cuyo resultado se presenta en la Figura 4.5.

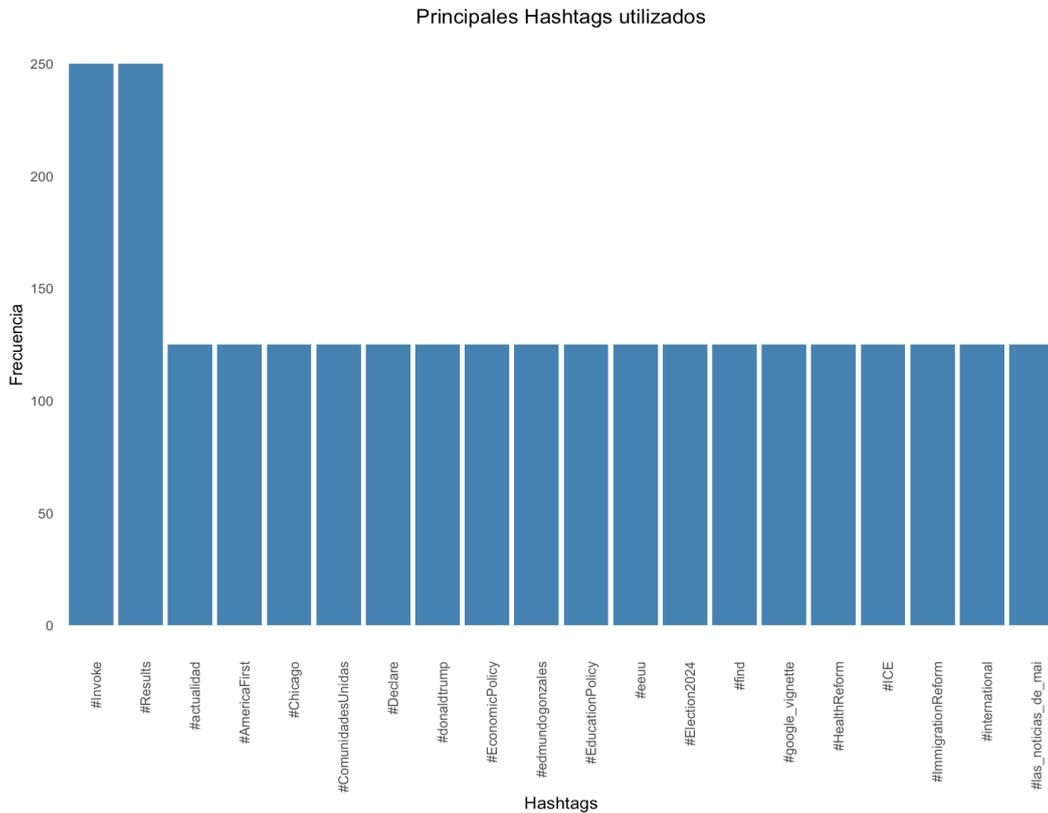


Figura 4.5: Top 10 Hashtags más utilizados  
Elaboración propia.

El análisis de los *hashtags* más frecuentes muestra una notable concentración temática en torno a cuestiones políticas y migratorias. Entre las etiquetas más recurrentes se encuentran #Invoke y #Results, ambas con más de 250 menciones, seguidas por #actualidad, #AmericanFirst y #Chicago, que superan las 100 apariciones. Esta distribución sugiere una articulación discursiva que oscila entre lo local y lo global: por un lado, *hashtags* como #Chicago o #ComunidadesUnidas destacan dinámicas territoriales específicas y formas de organización comunitaria; por otro, términos como #ICE, #ImmigrationReform o #AmericanFirst reflejan debates estructurales sobre política migratoria, seguridad nacional y soberanía.

La prominencia de etiquetas asociadas a la administración Trump, especialmente #AmericanFirst, indica que una parte significativa de las conversaciones está asociada a posicionamientos ideológicos vinculados al nacionalismo conservador y al endurecimiento de las políticas migratorias. Este resultado coincide con estudios previos que destacan la capacidad de las redes sociales para amplificar discursos polarizados en contextos electorales o de



El análisis de la figura muestra la prominencia de conceptos vinculados al aparato institucional y la política migratoria, con términos como “trump”, “state”, “government”, “country” y “united”, que reflejan el contexto político y la centralidad de la figura de Trump en los discursos analizados. Asimismo, aparecen términos relacionados directamente con la migración venezolana (“migrant”, “venezuela”) y la seguridad nacional (“border”, “security”, “law”, “agency”), lo cual destaca la idea de que la migración se enmarca dentro de narrativas securitarias y de control en Estados Unidos. Resaltan también referencias a fenómenos económicos y laborales (“job”, “worker”, “economic”), así como a la dimensión geopolítica y transnacional (“mexico”, “china”, “russia”). Estas menciones sugieren que los discursos sobre la migración venezolana no son un tema aislado, sino que se insertan en un marco más amplio de disputas ideológicas y geopolíticas que pueden abarcar desde cuestiones de comercio internacional hasta debates sobre seguridad y política exterior.

Además, la presencia de términos vinculados a las instituciones del Estado y su respuesta frente a la inmigración (“agency”, “law”, “order”) muestra también una preocupación constante por los mecanismos de control y las políticas públicas implementadas, indicando que la conversación en torno a la migración venezolana puede articularse también con la percepción de la eficacia o legitimidad de estas medidas. Finalmente, la coexistencia de términos como “venezuela”, “migrant”, “trump” y “security” evidencia la tensión entre discursos humanitarios, que podrían apelar a la solidaridad y los derechos humanos, y visiones que entienden la migración como una amenaza para la seguridad y la estabilidad nacional.

### 4.3. Análisis de *N-Gramas*

El análisis de n-gramas permite identificar las unidades léxicas más representativas dentro de un corpus textual. En lugar de basarse exclusivamente en la frecuencia absoluta, en este estudio se ha empleado la métrica TF-IDF (*Term Frequency-Inverse Document Frequency*), que permite ponderar los términos en función de su relevancia relativa dentro del conjunto de documentos, atenuando el peso de las palabras comunes y destacando aquellas que aportan un mayor valor informativo. Esta estrategia es útil para detectar los ejes temáticos con mayor capacidad discriminativa en discusiones digitalmente distribuidas como las observadas en Reddit. En el caso de los unigramas, representados en la Figura 4.7, se identifican como más relevantes términos como “trump”, “country”, “migrant”, “venezuela” y “war”, lo cual sugiere que el discurso gira en torno a temáticas geopolíticas, identitarias y político-institucionales. A su vez, aparecen vocablos relacionados con el aparato estatal y el contexto político estadounidense, como “government”, “order”, “president” y “election”. Esta selección de términos, ponderada por su especificidad contextual, muestra que los comentarios analizados no solo hacen referencia a la situación migratoria venezolana, sino que se puede articular dentro de marcos discursivos más amplios que incluyen la gobernanza, la seguridad

nacional y la política exterior.

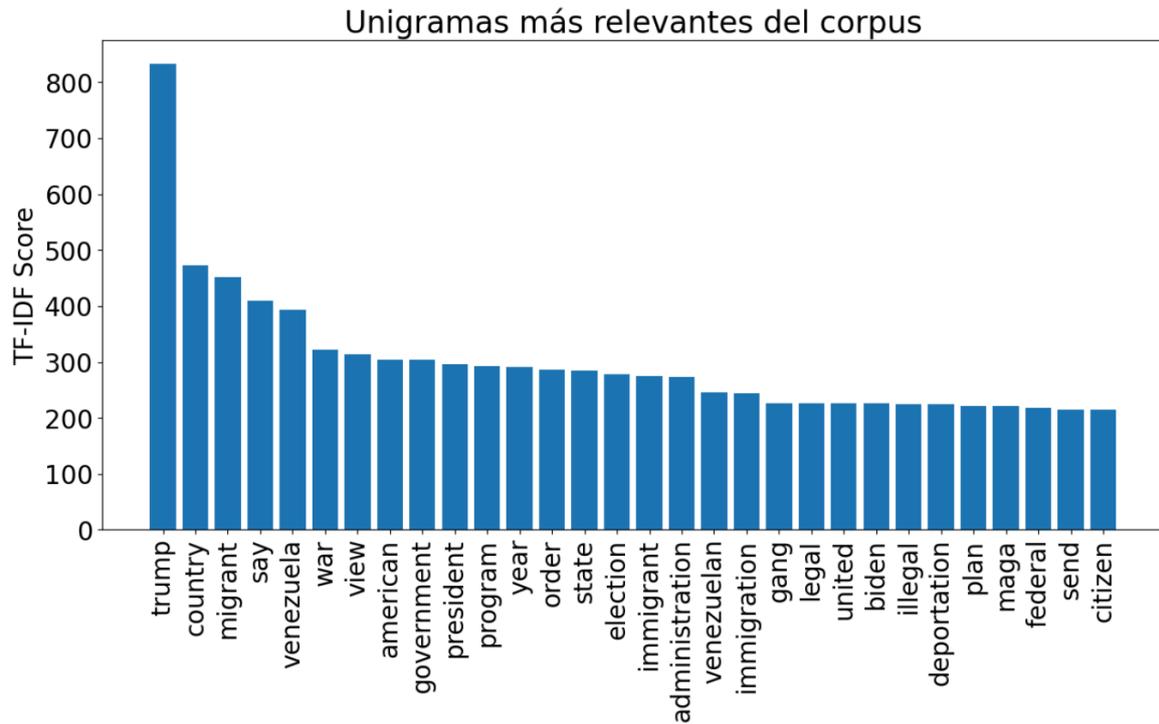


Figura 4.7: Unigramas con Mayor TF-IDF  
Elaboración propia.

En cuanto al análisis de bigramas (pares de palabras consecutivas), que podemos observar en la Figura 4.8, destacan combinaciones como “united states”, “tiny hands”, “trump administration” y “biden administration”, que muestran la importancia de los actores políticos estadounidenses y la figura de Trump en los discursos sobre la migración venezolana. La expresión “tiny hands” es una referencia sarcástica y humorística hacia Donald Trump, frecuentemente utilizada en redes sociales para burlarse de su figura y liderazgo. También destacan “cuba venezuela” y “latin america”, lo que sugiere la relevancia de espacios geográficos específicos y de la comunidad migrante venezolana en estas regiones. La aparición de “legal status”, “asylum seeker” y “executive order” indica la preocupación por los marcos legales y administrativos que rigen la movilidad humana, mientras que menciones como “fox news” y “rubio say” reflejan la influencia mediática y de actores políticos concretos en la construcción de estas narrativas. “Rubio say” alude a declaraciones públicas de Marco Rubio, senador estadounidense, quien ha tenido un papel destacado en la política migratoria y en la relación con Venezuela, especialmente por sus posturas críticas hacia el régimen venezolano y su defensa de políticas más restrictivas de inmigración.

Por su parte, la Figura 4.9 presenta los trigramas con mayor peso según la métrica TF-IDF, los cuales refuerzan las tendencias discursivas previamente identificadas, al tiempo que

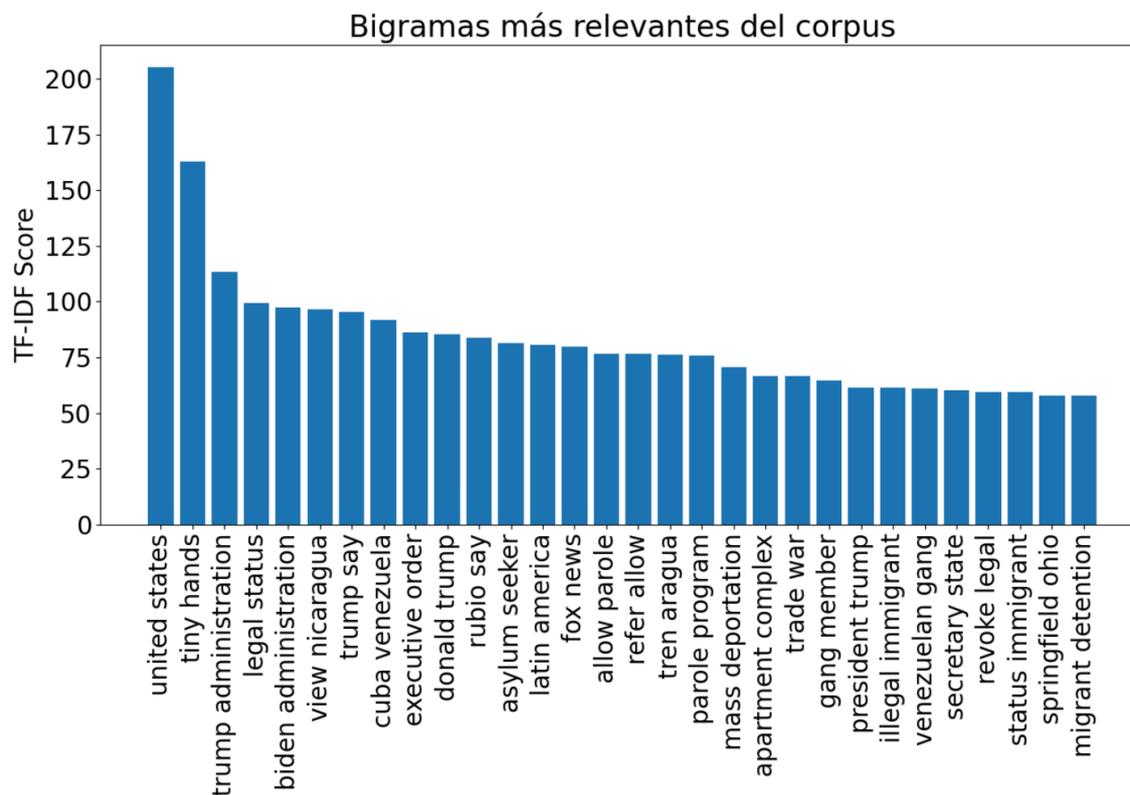


Figura 4.8: Bigramas con Mayor TF-IDF  
Elaboración propia.

añaden un nivel de especificidad léxica que permite capturar con mayor precisión las construcciones semánticas más representativas del corpus. Expresiones como “allow parole program”, “refer allow parole” y “legal status immigrant” destacan la relevancia del programa de *parole* humanitario y del estatus legal de los migrantes como ejes temáticos clave. Estos términos reflejan la preocupación y la importancia del *parole* humanitario como política migratoria en Estados Unidos, especialmente relevante para los migrantes venezolanos que buscan regularizar su situación. Otros trigramas, como “rick rojas atlanta” y “atlanta bureau chief”, muestran la importancia de actores mediáticos concretos en la circulación de estos relatos. “Rick Rojas Atlanta” se refiere a un periodista del *New York Times* que ha documentado la situación migratoria desde Atlanta, contribuyendo a la difusión y legitimación de estos discursos. Finalmente, combinaciones como “tiny hands administration” y “venezuela wave migrant” reflejan la convergencia entre figuras políticas, dinámicas migratorias y representaciones irónicas, lo que indica que el debate en torno a la migración también podría estructurarse mediante recursos humorísticos y formas discursivas marcadas por la polarización.

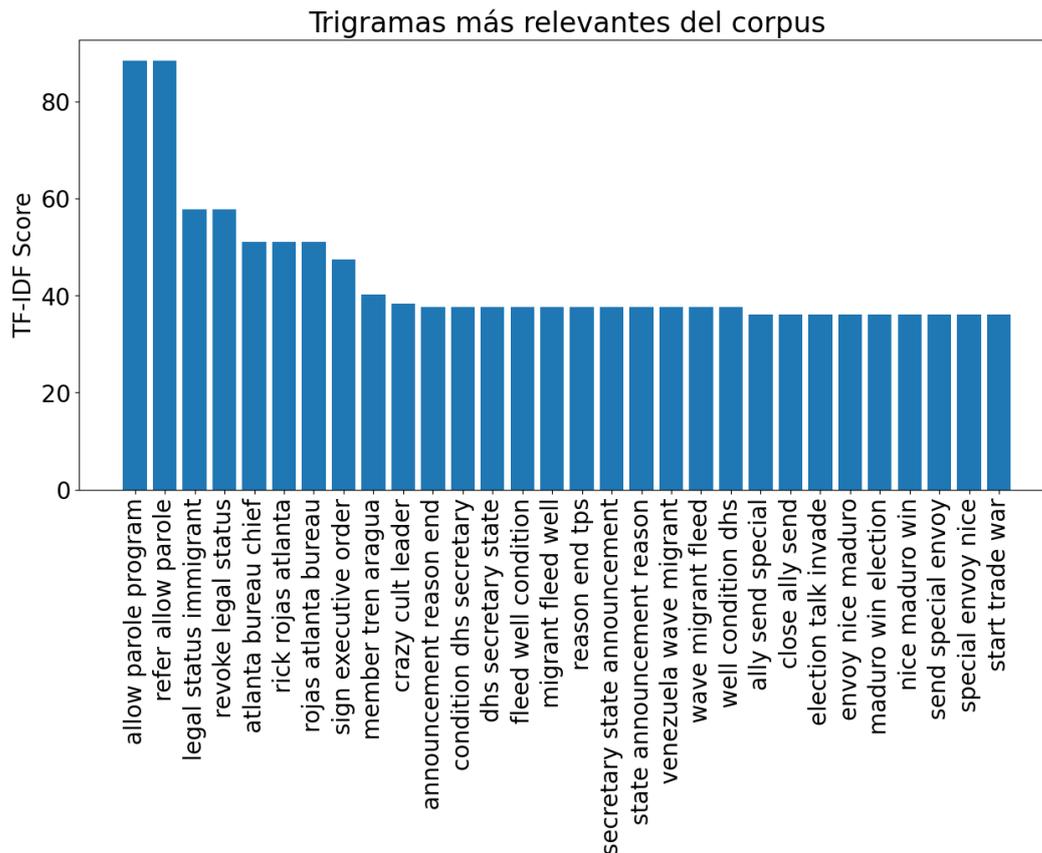


Figura 4.9: Trigramas con Mayor TF-IDF  
Elaboración propia.

## 4.4. Modelado de tópicos

Siguiendo el procedimiento descrito en la Sección 3.4, el modelado temático se llevó a cabo mediante Latent Dirichlet Allocation (LDA). Conforme a las recomendaciones metodológicas, la determinación del número de tópicos  $K$  se basó en un procedimiento iterativo que combina la métrica de coherencia temática con la inspección de la distribución intertópica. En concreto, se estimaron modelos para valores de  $K$  comprendidos entre 2 y 15; para cada configuración se calculó el índice de coherencia, una medida que evalúa la cohesión semántica interna de los tópicos, y se visualizó la separación entre ellos en un espacio de proyección bidimensional. La Figura 4.10 muestra la evolución del índice de coherencia a lo largo de dicho rango de valores, lo que permitió identificar el punto en el que se alcanza el mejor compromiso entre granularidad temática y diferenciación semántica.

Tal como se muestra en la Figura, los valores más altos del índice de coherencia se concentran en los modelos con  $K = 4, 6, 8$  y  $10$ . Entre estas opciones, el modelo con  $K = 4$  ofrece el compromiso más equilibrado entre calidad temática e interpretabilidad. En primer lugar, presenta un nivel de coherencia comparable al de configuraciones con mayor número

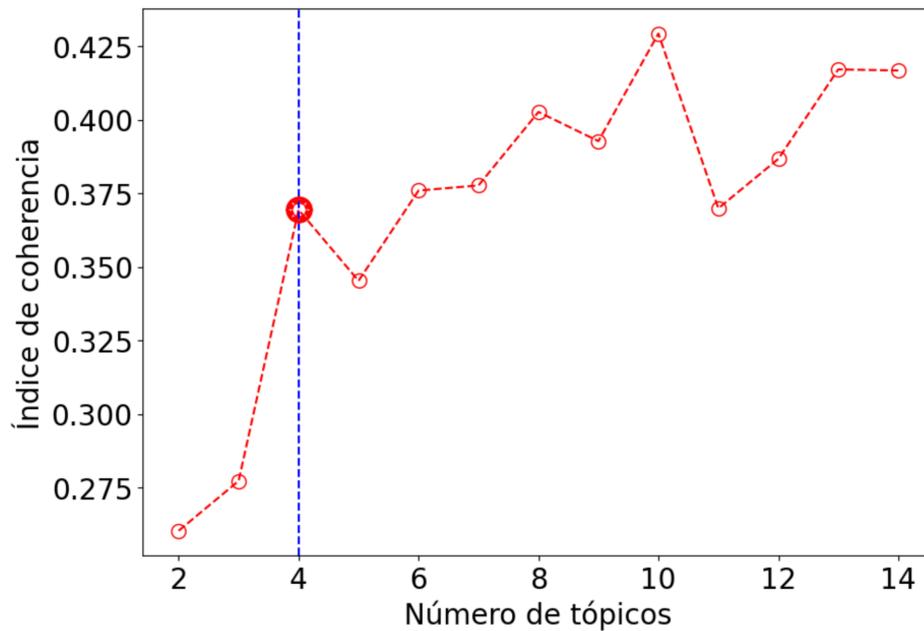


Figura 4.10: Coherencia por Número de Tópicos  
Elaboración propia.

de tópicos, evitando al mismo tiempo la fragmentación temática que tiende a producirse cuando el número de componentes se incrementa. En segundo lugar, favorece una comprensión más clara y sistemática de los resultados, al limitar el volumen de información que debe ser analizada e interpretada por el investigador. Para validar esta elección, se estimaron las distancias intertópicas correspondientes a los modelos con  $K = 4, 6, 8, 10$ , mediante el mapa de distancia intertópica. Los resultados mostraron que, a partir de  $K = 6$ , los tópicos comienzan a mostrar solapamientos significativos en el espacio temático, situación que se agrava con  $K = 8$  y alcanza su punto más crítico en  $K = 10$ , donde al menos seis componentes presentan intersecciones parciales relevantes. Esta superposición semántica dificulta la asignación clara de significados diferenciados a cada tópico, comprometiendo tanto la estabilidad como la utilidad analítica del modelo. En contraste, el modelo con  $K = 4$  mantiene una separación adecuada entre los tópicos, asegurando su delimitación temática.

La Figura 4.11 visualiza el mapa de distancia intertópica correspondiente a la configuración con  $K = 4$ , generado a partir de escalado multidimensional. En esta representación, los círculos indican los tópicos y su tamaño es proporcional a la prevalencia relativa en el corpus. Se observa una disposición claramente diferenciada, sin evidencias de solapamiento sustantivo, lo que indica que cada tópico constituye una dimensión temática distinta y no redundante del discurso analizado. Desde una perspectiva metodológica, esta distribución confirma la solidez estructural del modelo, al tiempo que facilita su interpretación cualitativa y su integración en etapas analíticas posteriores. La separación semántica entre componentes observada en la Figura 4.11 responde a los estándares teóricos recomendados para la eva-

luación de modelos de tópicos, particularmente en lo que respecta a la coherencia interna y la independencia temática. En conclusión, estos resultados respaldan la elección de  $K = 4$  como la configuración óptima para representar las principales líneas discursivas del corpus estudiado.

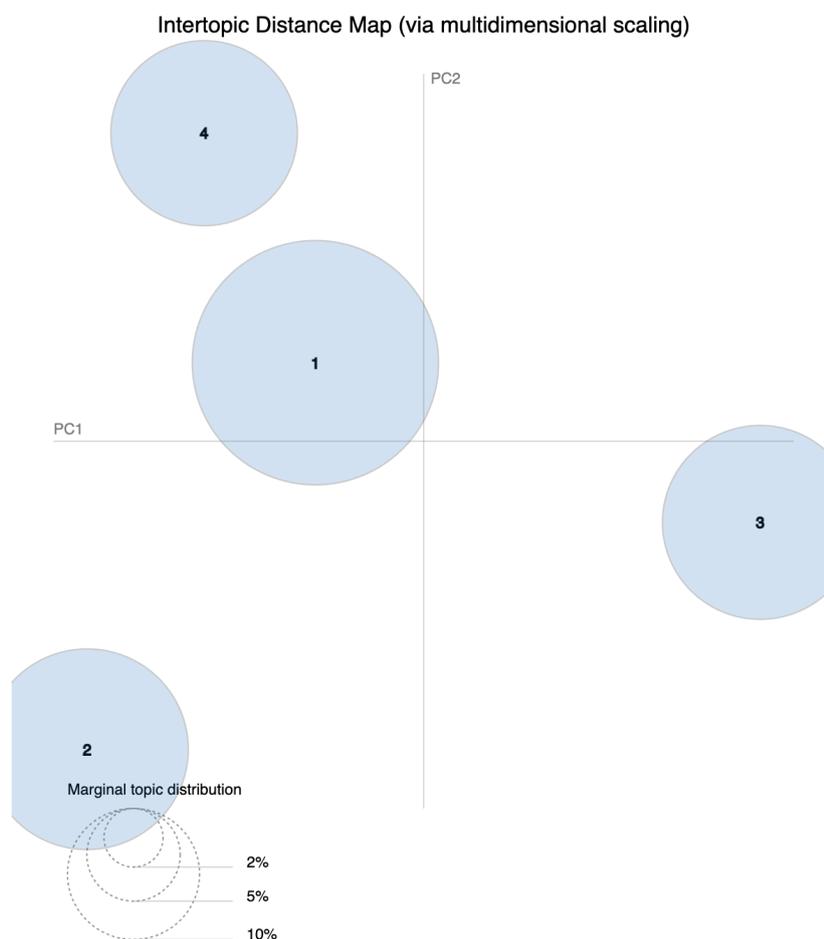


Figura 4.11: Distancia intertópica, modelo  $K=4$   
Elaboración propia.

Posteriormente, se procedió a la categorización temática de los tópicos generados por el modelo LDA, con base en el análisis cualitativo de las palabras clave, bigramas y trigramas con mayor peso en cada uno de los tópicos. Esta fase permitió asignar una etiqueta representativa a cada tópico a partir de los términos con mayor probabilidad dentro de su distribución léxica. La Tabla 4.1 resume las etiquetas asignadas a cada tópico, caracterizando su contenido principal y facilitando una interpretación más precisa de los temas presentes en el corpus.

<b>Tópico</b>	<b>Categoría</b>	<b>Palabras clave</b>	<b>Bigramas</b>	<b>Trigramas</b>
1	Discurso jurídico y legalidad	“migrant”, “country”, “trump”, “citizen”, “election”, “government”, “say”, “administration”, “tiny” y “american”	“tiny hands”, “united states”, “trump administration”, “rubio say” y “allow parole”	“allow parole program”, “refer allow parole”, “member tren aragua”, “election tiny hands” y “tiny hands administration”
2	Administración federal y control institucional migratorio	“order”, “federal”, “trump”, “program”, “fire”, “department”, “freeze”, “funding”, “employee” y “year”	“executive order”, “biden administration”, “sign executive”, “federal worker” y “funding freeze”	“sign executive order”, “agency international develop”, “ordered federal agency”, “federal funding freeze” y “office personnel management”
3	Causas estructurales y reconocimiento selectivo del refugiado	“trump”, “view”, “president”, “war”, “migrant”, “say”, “government”, “country”, “united” y “venezuela”	“view nicaragua”, “united states”, “fox news”, “trump say” y “legal status”	“legal status immigrant”, “revoke legal status”, “atlanta bureau chief”, “rick rojas atlanta” y “rojas atlanta bureau”
4	Liderazgo populista y exclusión migratoria	“trump”, “state”, “war”, “political”, “fix”, “maga”, “american”, “world”, “country” y “vote”	“united states”, “cult leader”, “bashar assad”, “donald trump” and “niagara falls”	“crazy cult leader”, “able issue deportation”, “anymore lmao trump”, “deportation long sick” y “early wow hide”

Tabla 4.1: Tópicos, palabras clave, bigramas y trigramas más relevantes de los tópicos seleccionados.

Elaboración propia.

En el marco del modelo de tópicos generado, el **Tópico 1** se interpreta como una representación del discurso institucional y jurídico asociado a la migración, con énfasis en el debate sobre el estatus legal, el derecho al asilo y las medidas excepcionales implementadas según criterios de nacionalidad o contexto geopolítico. La categoría asignada, **“Discurso jurídico migratorio y legalidad”**, sintetiza el eje central de este clúster, caracterizado por la preocupación sobre quién puede permanecer legalmente en Estados Unidos y bajo qué condiciones normativas. El conjunto de términos más relevantes, entre ellos “migrant”, “country”, “trump”, “citizen”, “election”, “government”, “administration” y “american”, sugiere que el núcleo semántico del tópico gira en torno a la intersección entre migración, ciudadanía y decisiones políticas. El bigrama “trump administration” destaca esta orientación, aludiendo a la centralidad de las políticas migratorias aplicadas durante ese mandato, mientras que expresiones como “rubio say” están asociadas a declaraciones de figuras políticas relevantes, como el senador Marco Rubio, cuyo posicionamiento ha sido especialmente influyente en el tratamiento del caso venezolano. Durante este periodo, se han impulsado medidas que han restringido drásticamente las vías legales de entrada y permanencia en el país, tales como la disminución de la cuota de refugiados, la suspensión temporal del programa DACA (*Deferred Action for Childhood Arrivals*), o la implementación de la política de “Tolerancia Cero”, que derivó en la separación forzada de familias migrantes.

Asimismo, varias combinaciones léxicas de alto peso contribuyen a matizar el alcance de este tópico. El bigrama “allow parole” y trigramas vinculados, como “refer allow parole”, aluden a la figura del parole humanitario, mecanismo que autoriza la entrada temporal de migrantes por razones humanitarias pero no garantiza estabilidad jurídica a largo plazo. Su recurrencia evidencia la percepción de inestabilidad legal y la sensación de hallarse en un “limbo” migratorio, sentimientos especialmente acentuados en la población venezolana que depende de este recurso y teme la deportación una vez vencido el permiso. En segundo término, la aparición del trigramas “member tren aragua” introduce una dimensión securitaria al conectar la migración con la criminalidad organizada. El Tren de Aragua, banda delictiva originaria de Venezuela, se destaca así en el discurso digital, reforzando narrativas que asocian la movilidad humana con amenazas a la seguridad pública. Por último, expresiones como el bigrama “tiny hands” y trigramas tales como “election tiny hands” o “tiny hands administration” aportan un registro satírico al tópico. Estas referencias, popularizadas como meme político desde la campaña presidencial de 2016, ridiculizan el liderazgo de Donald Trump y revelan el uso estratégico del humor como herramienta de deslegitimación en el debate migratorio.

El **Tópico 2** se centra en el papel del Estado y sus instituciones en la regulación de los flujos migratorios, estructurado bajo la categoría **“Administración federal y control institucional migratorio”**. Este clúster muestra una narrativa orientada a la gestión operativa del fenómeno migratorio, destacando el rol de las agencias gubernamentales, los programas específicos y los recursos presupuestarios asignados para su implementación. Entre las palabras

clave más representativas se encuentran “order”, “federal”, “trump”, “program”, “department”, “funding”, “freeze” y “employee”, lo que refleja un discurso centrado en decisiones ejecutivas, asignaciones de fondos y dinámicas administrativas. Un elemento particularmente relevante en este marco es el incremento del presupuesto destinado al control fronterizo durante la campaña de Trump, que permitió ampliar significativamente el número de agentes desplegados en zonas estratégicas del país, como parte de una política migratoria más estricta. La presencia destacada del bigrama “executive order” y de trigramas como “sign executive order” o “biden administration” refuerza esta orientación institucional. Estas expresiones se asocian al uso de órdenes ejecutivas como mecanismo decisivo en la formulación e implementación de políticas migratorias, tanto durante el mandato de Donald Trump como en la administración de Joe Biden. Más allá de su efecto normativo inmediato, estas órdenes también cumplen una función simbólica, al comunicar de forma directa las prioridades del poder ejecutivo. De igual modo, se hace referencia a la implicación de entidades específicas como la *Office of Personnel Management* o la *Agency for International Development*, cuyas funciones pueden ser ampliadas, restringidas o reorientadas en función de los decretos presidenciales vigentes.

Asimismo, la co-ocurrencia de trigramas como “federal funding freeze”, “federal worker” y “ordered federal agency” destacan el impacto directo que las decisiones administrativas ejercen sobre los recursos humanos y financieros de las dependencias gubernamentales involucradas en la atención migratoria. La suspensión temporal de partidas presupuestarias destinadas a programas de reasentamiento o asistencia legal repercute de forma inmediata en los migrantes recién llegados, quienes dependen de tales servicios para avanzar en sus procesos de integración y regularización. Expresiones aisladas como “fire”, “freeze” y “year” enfatizan, además, la dimensión laboral de estas medidas, aludiendo a despidos, congelación de contrataciones y bloqueos presupuestarios que introducen una incertidumbre sostenida en el personal federal y en la continuidad de las iniciativas. En conjunto, estos elementos podrían interpretarse como indicios de que la gestión administrativa estaría siendo utilizada, al menos en ciertos contextos, como una herramienta adicional de control migratorio.

En paralelo a las dimensiones institucionales previamente descritas, el **Tópico 3** orienta el análisis hacia los factores exógenos y estructurales que configuran las dinámicas contemporáneas de movilidad migratoria forzada. Bajo la categoría analítica “**Causas estructurales y reconocimiento selectivo del refugiado**”, este tópico reúne un conjunto de términos que remiten a contextos de conflicto, crisis humanitaria e inestabilidad política. Palabras clave como “war”, “migrant”, “country”, “venezuela”, “president”, “government” o “trump” permiten identificar una narrativa en la que la migración aparece vinculada a violaciones de derechos humanos, colapsos institucionales o persecuciones ideológicas, reconociendo su carácter forzoso en muchos casos. En este marco, el bigrama “view Nicaragua” actúa como punto de referencia comparativa, mostrando que los usuarios tienden a establecer paralelismos discursivos entre distintas situaciones geopolíticas, particularmente entre Nicaragua y

Venezuela, para evidenciar patrones regionales de desplazamiento.

Los trigramas “atlanta bureau chief”, “rick rojas atlanta” y “rojas atlanta bureau” incorporan un matiz mediático-institucional que sugiere la presencia de actores periodísticos en la configuración del discurso. La frecuente alusión a Rick Rojas, corresponsal del *New York Times* en Atlanta, podría indicar que la cobertura de este periodista actúa como lente a través de la cual los usuarios interpretan los efectos locales de las decisiones migratorias federales. Su labor informativa, orientada a documentar las repercusiones de dichas políticas en comunidades concretas, entre ellas la venezolana residente en Georgia, parece servir de referencia para discutir hipotéticos escenarios de deportación, exclusión o asistencia institucional en función del ciclo político. De manera análoga, la aparición de trigramas como “trump say” o “legal status immigrant” podría apuntar a la influencia que las declaraciones de figuras políticas ejercen sobre las percepciones de legalidad y pertenencia. En este contexto, términos como “revoke legal status” parecen reflejar la inquietud, no exenta de incertidumbre, de que protecciones provisionales, tales como el TPS o el *parole* humanitario, puedan ser revertidas en función de vaivenes políticos.

Por último, el **Tópico 4** parece centrarse en la dimensión simbólica y afectiva que rodea la figura de Donald Trump dentro del debate migratorio digital, categoría que puede describirse provisionalmente como **“Liderazgo populista y discursos de deportación”**. El conjunto de palabras clave, “trump”, “maga”, “state”, “war”, “political”, “american”, “country”, “vote”, sugiere un discurso fuertemente ideologizado, en el que la migración, más que tratarse como fenómeno social o humanitario, se instrumentaliza con fines de movilización política. Ciertos trigramas, tales como “crazy cult leader”, junto con expresiones del tipo “anymore lmao trump”, apuntan a una representación satírica del expresidente, retratado hipotéticamente como líder carismático cuyos seguidores asumirían de forma acrítica su retórica.

En este mismo registro aparece la mención de líderes como *Bashar al-Assad*, presidente de Siria, cuya presencia en los trigramas sugiere que algunos usuarios establecen paralelismos entre distintos conflictos internacionales y los procesos de desplazamiento forzado asociados a regímenes autoritarios. Del mismo modo, combinaciones léxicas tales como “able issue deportation” o “deportation long sick” denotan un tono crítico y cargado de frustración respecto a los protocolos de deportación, percibidos por determinados participantes del debate como medidas desproporcionadas o de aplicación arbitraria. La recurrencia de expresiones humorísticas, por ejemplo “crazy cult leader” o “early wow hide”, refuerza la hipótesis de que el discurso digital articula la crítica política mediante recursos satíricos que caricaturizan tanto a las figuras de poder como a las políticas de expulsión.

En la Figura 4.12 se muestra la frecuencia absoluta de cada uno de los tópicos identificados en el corpus textual. El tópico 1, categorizado como **“Discurso jurídico y legalidad”**, es el más mencionado, con un total de **4.863 publicaciones**. Esta alta presencia sugiere que la comunidad de *Reddit* presta especial atención a temas legales como el *parole* humanitario, el estatus migratorio y los mecanismos de control asociados al perfil de los migrantes

venezolanos, como su vinculación con organizaciones criminales o su elegibilidad para permanecer en EE.UU. de forma temporal o permanente. Le sigue, el tópico 3, definido como **“Causas estructurales y reconocimiento selectivo de refugiados”**, el cual agrupa **4.484 publicaciones**. Este tópico aporta una mirada contextual e internacional al debate migratorio, vinculando el caso venezolano con otras crisis geopolíticas como Nicaragua o Siria y destaca la contradicción entre las causas objetivas de migración forzada y la falta de reconocimiento legal como refugiados.

El Tópico 4, asignado a la categoría **“Liderazgo populista y exclusión migratoria”**, registra una frecuencia relativa de 2.246 publicaciones. Aunque su volumen es inferior al de los dos primeros clústeres, resulta relevante por la carga simbólica y afectiva que incorpora: la aparición de recursos humorísticos y satíricos, particularmente en torno a la figura de Donald Trump, sugiere que parte del debate se articula a través de narrativas meméticas que, de forma indirecta, pueden contribuir a moldear la percepción pública sobre la población venezolana. Por último, el Tópico 2, vinculado con la **“Administración federal y control institucional migratorio”**, concentra 1.494 publicaciones, lo que lo convierte en el menos representado del conjunto. Esta menor presencia sugiere que los aspectos estrictamente burocráticos, órdenes ejecutivos, reasignaciones presupuestarias o directrices administrativas emitidas tanto por la administración Trump como por la de Biden, generan un interés comparativamente reducido dentro de las discusiones analizadas. Sin embargo, su inclusión en el modelo confirma la existencia de un segmento discursivo que, aunque minoritario, se centra en la dimensión procedimental de la política migratoria.

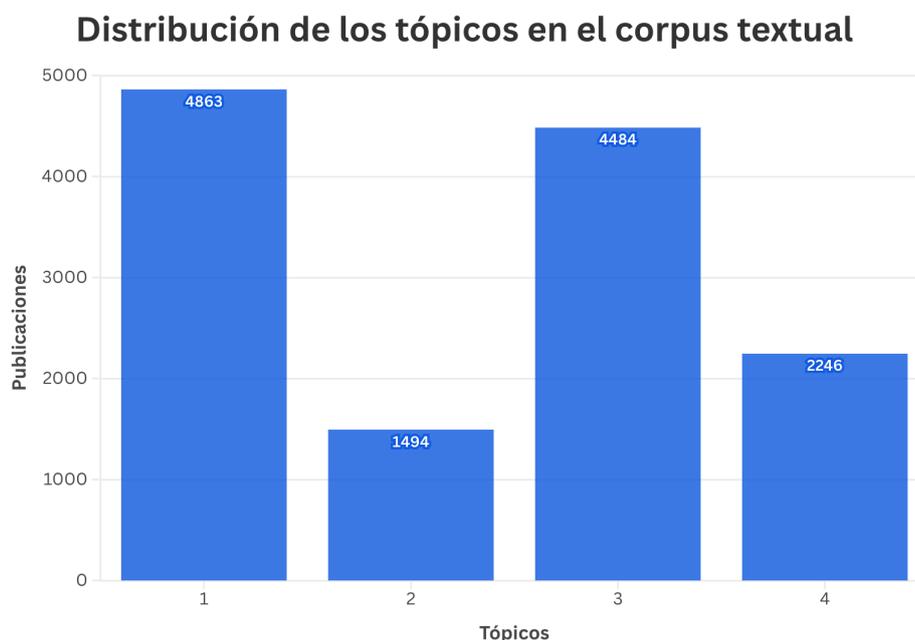


Figura 4.12: Distribución de los tópicos en el corpus textual  
Elaboración propia.

## 4.5. Análisis de sentimientos

En esta etapa del análisis, tal como se mencionó en la Sección 3.5, se analizan las emociones expresadas por los usuarios de *Reddit* en comunidades digitales de habla inglesa respecto a la migración venezolana. Para ello, se ha aplicado el modelo VADER, que detecta la carga emocional de los mensajes mediante la asignación de una puntuación de polaridad en un rango de  $-1$  (negativo) a  $+1$  (positivo), clasificando los mensajes como positivos, negativos o neutros. A fin de preservar los elementos relevantes para el algoritmo, como el uso de mayúsculas, signos de puntuación y emoticonos, se implementó un proceso de limpieza específico que difiere del aplicado en fases anteriores. Esta estrategia permite captar con mayor precisión los matices afectivos del discurso digital.

Tras este procesamiento y la aplicación del modelo VADER, se elaboró un gráfico de cajas para representar la distribución de la puntuación compuesta de los mensajes analizados, como se muestra en la Figura 4.13. El diagrama revela una clara tendencia hacia la negatividad: la mediana se sitúa por debajo de cero (aproximadamente en  $-0.5$ ), lo que indica que más del 50% de los comentarios presentan una carga emocional predominantemente negativa. Además, el rango intercuartílico abarca desde valores cercanos a  $-1$  hasta alrededor de  $0.5$ , lo que evidencia una amplia dispersión en los sentimientos expresados. La asimetría observada, con una mayor concentración de valores negativos, confirma la presencia de un clima afectivo desfavorable en torno a la migración venezolana dentro de las comunidades digitales de habla inglesa.

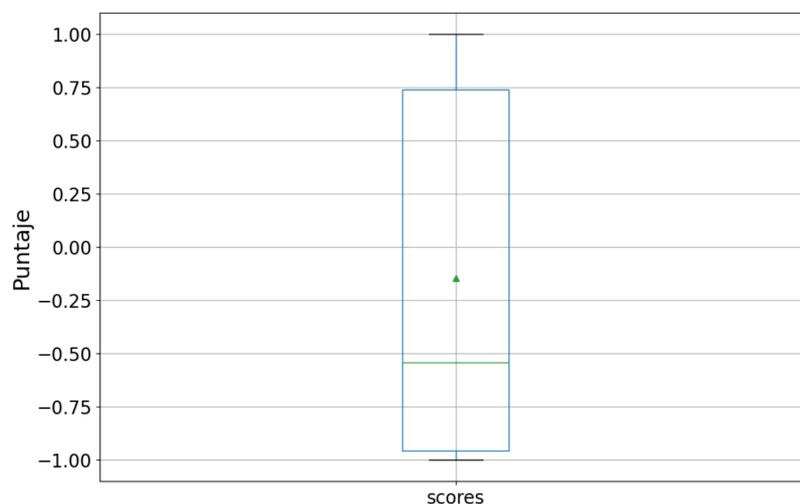


Figura 4.13: Distribución de la puntuación compuesta  
Elaboración propia.

Adicionalmente, para complementar el análisis de sentimiento, se presenta en la Figura 4.14 un histograma que representa la distribución de la puntuación compuesta de las publicaciones analizadas. La gráfica muestra una distribución claramente bimodal, con concen-

traciones elevadas en los extremos de la escala: especialmente en el valor mínimo (-1.0) y, en menor medida, en el máximo (1.0). Esta forma revela una tendencia general hacia la negatividad, con una densidad destacada de publicaciones que expresan emociones negativas intensas respecto a la migración, muchas de las cuales pueden estar relacionadas con discursos de criminalización, rechazo o preocupación por la llegada de migrantes venezolanos. A pesar de ello, se observa también una presencia notable de publicaciones con puntuaciones positivas, particularmente entre los valores de 0.5 y 1.0, lo que indica que existe también un espacio discursivo más favorable o empático sobre la población migrante. El pico en torno a la puntuación 0.0 señala una proporción importante de comentarios neutrales, sin carga emocional explícita.

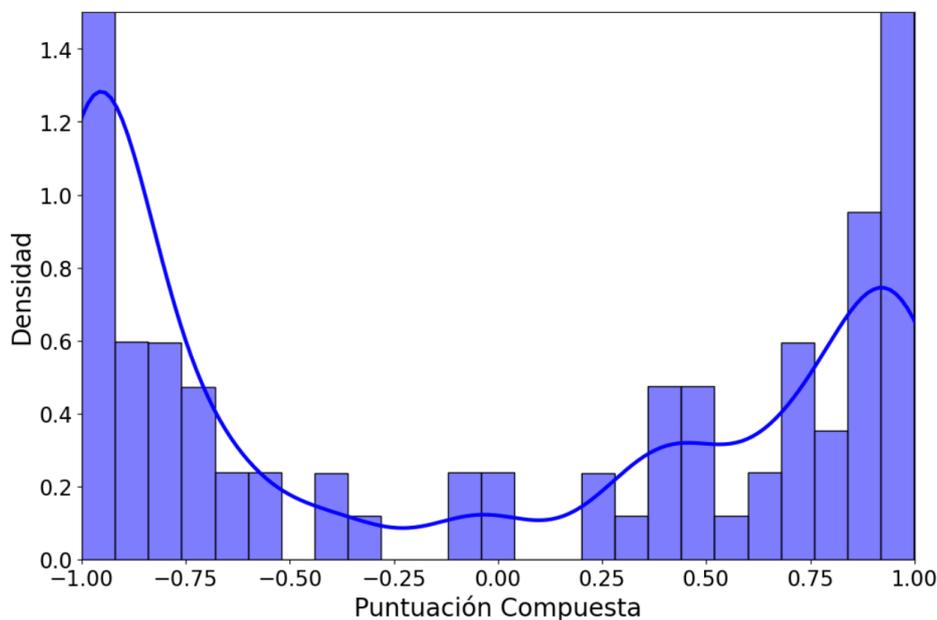


Figura 4.14: Número de publicaciones por puntuación  
Elaboración propia.

Posteriormente a esta inspección global, se evaluó la distribución del sentimiento dentro de cada uno de los cuatro tópicos identificados, a fin de discernir aquellos grupos temáticos que concentran la orientación afectiva más positiva y aquellos que presentan la evaluación más negativa, como se puede observar en la Figura 4.15. Para cada tópico se calcularon, además, los estadísticos descriptivos básicos, valor máximo, valor mínimo, media, desviación estándar y coeficiente de asimetría, cuyos resultados se resumen en la Tabla 4.2.

En relación con el tópico 1, “*Discurso jurídico y legalidad*”, se observa que presenta una mayor concentración de publicaciones con sentimiento negativo asociado, con una media de -0.149712, y a pesar de tener una media negativa es el segundo tópico con más sentimiento negativo, presentando una puntuación mínima de -0.9994, algo que ya nos indica que en general las conversaciones sobre migración venezolana en países de habla inglesa, principal-

mente Estados Unidos, tienen un sentimiento en general negativo. Este tópico gira en torno a políticas migratorias restrictivas, la criminalización del migrante (por ejemplo, con referencias al “Tren de Aragua”) y situaciones de vulnerabilidad legal que afectan directamente a las personas migrantes.

El tópico 2, “*Administración federal y control institucional migratorio*” presenta el mayor sentimiento negativo de los cuatro tópicos, con una puntuación mínima de -1.000, y una media de -0.335582. El coeficiente de asimetría de 0.864267 indica una distribución sesgada hacia la derecha, lo que significa que hay una mayor concentración de puntuaciones por debajo de la media, pero también una cola larga hacia los valores más altos o positivas. Esto sugiere que aunque muchas personas presentan un sentimiento muy negativo hacia la administración federal migratoria, hay ciertas publicaciones que tienen un sentimiento positivo. Este es el tópico con menos comentarios, como se ha referido anteriormente, ya que se refiere a temas como decisiones administrativas que afectan directamente a los recursos operativos de los programas clave y la regularización de las personas.

En cuanto al tópico 3, respecto los debates sobre las “*Causas estructurales y el reconocimiento selectivo del refugiado*”, muestra también una tendencia negativa significativa (media de -0.154), aunque menos acentuada. El coeficiente de asimetría es de 0.748437, lo cual indica que aunque la tendencia de los comentarios sea negativa, hay una concentración menor de comentarios con sentimiento positivo. Esto sugiere que, si bien el discurso está marcado por la crítica, probablemente hacia las causas estructurales del desplazamiento forzado o las políticas restrictivas de asilo, también existen publicaciones que visibilizan con empatía las situaciones que enfrentan los migrantes.

Por último, el tópico 4 categorizado como “*Liderazgo populista y exclusión migratoria*” es el único que presenta una media levemente positiva (0.005), lo cual puede explicarse por la presencia de registros discursivos irónicos o satíricos. El uso frecuente de expresiones humorísticas en este tópico parece suavizar la carga emocional negativa que normalmente acompaña a las discusiones sobre exclusión migratoria, sugiriendo que parte del discurso se articula desde una crítica lúdica o paródica al liderazgo populista y sus políticas.

<b>Tópico</b>	<b>Punt. Máxima</b>	<b>Punt. Mínima</b>	<b>Media</b>	<b>Desv. Típica</b>	<b>Coef. Asimetría</b>
1	0.9991	-0.9993	-0.149712	0.797199	0.283591
2	0.9905	-1.0000	-0.335582	0.864267	0.716988
3	0.9897	-0.9984	-0.154153	0.748437	0.151077
4	0.9993	-0.9999	0.005756	0.941026	-0.014935

Tabla 4.2: Estadísticos descriptivos por tema dominante  
Elaboración propia.

En conclusión, el análisis de sentimientos realizado sobre los distintos tópicos identificados revela una carga emocional predominantemente negativa en el discurso digital en torno

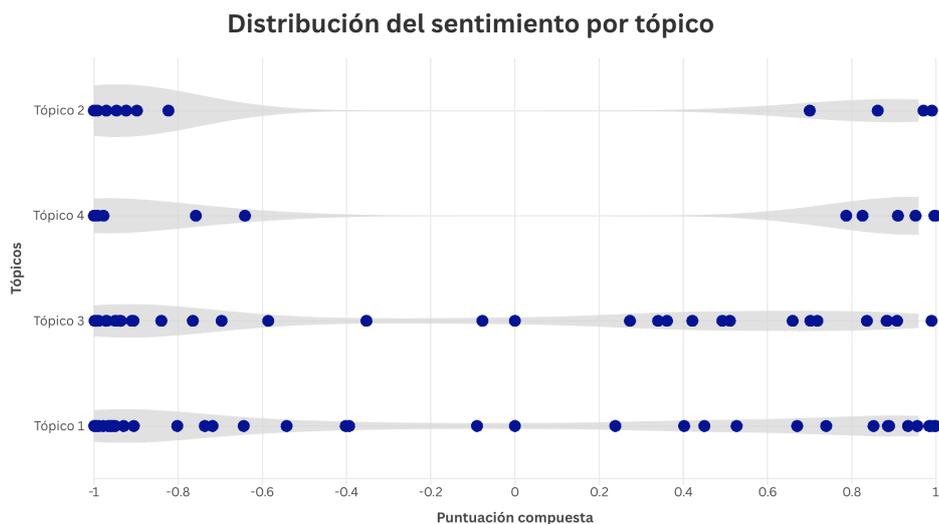


Figura 4.15: Distribución del sentimiento por tópico  
Elaboración propia.

a la migración venezolana en comunidades de habla inglesa. Este sesgo afectivo refleja no solo el impacto emocional que suscita el fenómeno migratorio en el imaginario colectivo, sino también las tensiones asociadas a las políticas de control, las condiciones de regularización y las causas estructurales del desplazamiento. La única excepción relativa se encuentra en el tópico vinculado al liderazgo populista, donde el uso de recursos satíricos atenúa la negatividad general del discurso.

# Capítulo 5

## Conclusiones

La migración venezolana constituye uno de los fenómenos más significativos de desplazamiento humano en el siglo XXI, tanto por su magnitud como por sus causas profundamente estructurales. A diferencia de otras olas migratorias motivadas por conflictos bélicos o desastres naturales, el éxodo venezolano responde a una crisis prolongada de carácter económico, político y social, que ha deteriorado de forma generalizada las condiciones de vida de la población. Esta situación ha impulsado a más de siete millones de personas a abandonar el país en busca de seguridad, estabilidad y oportunidades, configurando una diáspora de alcance regional y global que plantea retos sustantivos en materia de derechos humanos, gobernanza e integración.

El presente estudio incorpora un análisis socioeconómico preliminar que permite concluir que la crisis venezolana no puede ser comprendida adecuadamente sin considerar la histórica dependencia del petróleo como eje central de su estructura económica. Esta vulnerabilidad estructural, agravada por una gestión ineficiente, la imposición de sanciones internacionales y un entorno institucional de carácter autoritario, ha derivado en una contracción severa tanto del PIB como del PIB per cápita. Esta caída se ha traducido en un empobrecimiento masivo de la población: más del 90 % vive actualmente por debajo de la línea de pobreza, y alrededor del 60 % en pobreza extrema. A ello se suma el colapso de los sistemas de salud, educación y energía, la devaluación constante de la moneda nacional, y un deterioro crítico en indicadores clave como el Índice de Desarrollo Humano, la mortalidad infantil y la inseguridad alimentaria. La combinación de estos factores ha deteriorado no solo el poder adquisitivo y la calidad de vida, sino también la cohesión social y la capacidad del Estado para garantizar derechos básicos. Además, el análisis presentado muestra que la progresiva pérdida de calidad democrática, reflejada en la caída del índice de democracia y en la consolidación de un régimen autoritario, ha profundizado la desconfianza institucional y acelerado la migración. Las decisiones políticas adoptadas en los últimos años han contribuido a restringir libertades, debilitar el tejido institucional y obstaculizar la inversión en infraestructuras esenciales, como el sistema eléctrico nacional. Esta crisis multidimensional ha afectado de manera desigual a

la población, exacerbando las brechas territoriales, de género y de clase social.

La percepción pública constituye un indicador de gran relevancia para captar la dimensión social y simbólica del éxodo venezolano, ya que refleja temores, actitudes de acogida o rechazo y narrativas sobre sus causas y efectos, elementos decisivos para el diseño de políticas migratorias y para los procesos de integración de la población desplazada. El análisis de conversaciones en redes sociales, como por ejemplo X o Reddit, ofrece un observatorio en tiempo real de esas narrativas. Estos datos permiten identificar tópicos emergentes, medir el sentimiento colectivo y detectar variaciones temáticas o territoriales ausentes en las fuentes oficiales, brindando así evidencia complementaria para entender la complejidad de la diáspora. Para extraer este tipo de información sobre las percepciones ciudadanas en redes sociales, diversos estudios han propuesto estrategias metodológicas que combinan técnicas de minería de texto, análisis de sentimientos y modelado de tópicos. La literatura especializada ha utilizado enfoques como TF-IDF, LDA y NMF para identificar patrones discursivos, mientras que el análisis de sentimientos se ha apoyado en herramientas como VADER para estimar la polaridad emocional de los mensajes. En contextos de alta tensión política, también se han implementado modelos supervisados como SVM y regresión logística para detectar contenido discriminatorio, complementados con técnicas de balanceo como SMOTE. Algunos trabajos recientes han incorporado dimensiones más complejas, como el *code-mixing*, mediante modelos multilingües basados en transformadores que permiten identificar adaptaciones lingüísticas propias de comunidades migrantes.

En este contexto, el presente trabajo tiene como objetivo principal analizar la percepción pública sobre la inmigración venezolana en comunidades digitales de habla inglesa, utilizando como fuente de información las publicaciones generadas en la plataforma Reddit. A través de la aplicación de técnicas de PLN, como el análisis de sentimientos y el modelado de tópicos, se busca identificar las principales narrativas y emociones asociadas al fenómeno migratorio, así como explorar la diversidad de representaciones y actitudes que circulan en el espacio digital. Esta aproximación permite no solo complementar la comprensión de la diáspora venezolana desde una perspectiva cuantitativa y discursiva, sino también ofrecer insumos empíricos que contribuyan al diseño de políticas públicas basadas en evidencia. En términos metodológicos, el estudio se sustenta en un flujo de trabajo reproducible que abarca cuatro fases principales. Primero, se extrajeron mediante la API de Reddit  $\approx 150.000$  publicaciones y comentarios publicados entre octubre de 2024 y febrero de 2025 que incluían términos relacionados con “Venezuela” y “migración”, preservando metadatos temporales y de comunidad. Segundo, se realizó un preprocesamiento específico: se conservaron duplicados significativos, se eliminaron enlaces, menciones y hashtags irrelevantes, se amplió la lista de *stopwords* y se aplicó lematización para unificar variantes morfológicas. Tercero, el corpus limpio se sometió a un análisis exploratorio de N-gramas ponderados mediante TF-IDF, seguido de modelado de tópicos con LDA. Para la selección del número óptimo de tópicos  $K$  se basó en la maximización del índice de coherencia y en la inspección de la distancia inter-

tópica proyectada con t-SNE. Finalmente, se evaluó la carga emocional de cada mensaje con VADER, manteniendo puntuación, emoticonos y capitalización para preservar información relevante para el algoritmo, lo que permitió triangular temáticas y sentimientos, y así ofrecer un análisis más completo de las narrativas sobre la diáspora venezolana.

Los resultados obtenidos a partir del análisis léxico indican que las construcciones semánticas más representativas del corpus giran en torno a expresiones vinculadas al *parole* humanitario, el estatus legal de los migrantes y la importancia de actores mediáticos concretos. Combinaciones como “allow parole program” y “legal status immigrant” reflejan una preocupación central por la legalidad y estabilidad migratoria, mientras que referencias como “rick rojas atlanta” apuntan a la influencia de determinados periodistas en la articulación del discurso público sobre la migración venezolana. Asimismo, la presencia de expresiones irónicas o satíricas, por ejemplo, “tiny hands administration”, sugiere que parte del debate adopta un registro discursivo humorístico, mediante el cual se cuestionan y critican figuras políticas y medidas de exclusión, configurando un espacio narrativo marcado por la polarización. El modelo de tópicos con  $K = 4$ , validado mediante métricas de coherencia y análisis de distancia intertópica, permitió identificar cuatro líneas discursivas diferenciadas: (1) el discurso jurídico y legalidad, centrado en mecanismos de regularización migratoria y percepción de legitimidad institucional; (2) la administración federal y control institucional, que aborda medidas ejecutivas y restricciones presupuestarias desde un enfoque procedimental; (3) las causas estructurales y el reconocimiento selectivo del refugiado, que vinculan la migración forzada con crisis internacionales y la inconsistencia de los marcos legales de protección; y (4) el liderazgo populista y exclusión migratoria, donde predomina una narrativa ideologizada, afectiva y a menudo satírica. La mayor frecuencia de los tópicos 1 y 3 denota un interés predominante en aspectos legales y estructurales del fenómeno migratorio, mientras que la menor presencia del tópico 2 sugiere un grado limitado de atención hacia los componentes administrativos.

Finalmente, se realizó un análisis de sentimientos que complementa las conclusiones sobre el modelado de tópicos y muestra un predominio de valoraciones desfavorables en las comunidades de Reddit estudiadas. El análisis de la puntuación compuesta obtenida mediante VADER revela un sesgo general hacia la negatividad, con una intensidad particularmente acusada en el Tópico 2 (“Administración federal y control institucional migratorio”), cuyos mensajes expresan frustración ante los recortes presupuestarios, las decisiones ejecutivas y la incertidumbre laboral que afectan la gestión de la migración. Los Tópicos 1 (“Discurso jurídico y legalidad”) y 3 (“Causas estructurales y reconocimiento selectivo del refugiado”) también muestran una orientación emocional negativa, aunque de menor intensidad, lo que sugiere una preocupación sostenida por la seguridad jurídica y las causas estructurales del éxodo. En contraste, el Tópico 4 (“Liderazgo populista y discursos de deportación”) presenta la media de sentimiento más próxima a la neutralidad e incluso levemente positiva, atribuible a la introducción de recursos satíricos, como «crazy cult leader» o «anymore lmao

trump», que, al ridiculizar figuras políticas, atenúan la carga afectiva negativa. Estos resultados destacan la existencia de un discurso digital predominantemente crítico, matizado por registros humorísticos que actúan como mecanismos de contestación simbólica frente a las políticas de exclusión migratoria.

En conclusión, los resultados derivados del modelado de tópicos y el análisis de sentimientos del discurso digital constituyen una fuente valiosa de información para la identificación de áreas susceptibles de mejora en el diseño y la implementación de políticas migratorias. La comprensión de las inquietudes predominantes en la esfera pública, relativas al estatus legal, la seguridad, o la legitimidad del reconocimiento como refugiados, permite afinar los marcos normativos y comunicacionales, fortaleciendo así la capacidad de respuesta institucional. Asimismo, la monitorización continua de estas narrativas ofrece un insumo empírico relevante para la formulación de estrategias de integración más sensibles al contexto social y discursivo, contribuyendo al desarrollo de políticas públicas basadas en evidencia y orientadas a la cohesión social.

Desde una perspectiva prospectiva, futuros trabajos podrían orientarse a la ampliación y diversificación del corpus analizado, incorporando publicaciones procedentes de otras plataformas digitales como Instagram o TikTok, cuyos formatos multimodales y audiencias diferenciadas ofrecen nuevas posibilidades para el estudio de los discursos migratorios en entornos comunicativos más visuales y performativos. Asimismo, la inclusión de variables geográficas, a través de técnicas de geolocalización aplicadas a redes sociales con mayor densidad de metadatos espaciales, permitiría analizar las variaciones territoriales en la percepción pública, identificando regiones de mayor polarización o aceptación hacia la comunidad venezolana en el contexto estadounidense. Adicionalmente, la implementación de enfoques longitudinales posibilitaría analizar la evolución de las narrativas migratorias a lo largo del tiempo, estableciendo vínculos entre eventos coyunturales, como procesos electorales o crisis regionales, y cambios en la opinión pública digital. Otro eje relevante de investigación futura radica en la comparación entre comunidades migrantes, lo que permitiría contrastar la percepción social hacia la diáspora venezolana con la de otros colectivos, como los migrantes cubanos o mexicanos, a fin de identificar patrones discursivos comunes y divergencias significativas. Finalmente, una línea de desarrollo metodológico consistiría en incorporar modelos de aprendizaje profundo para el análisis de sentimiento, con el objetivo de mejorar la precisión en la detección de matices emocionales complejos y elevar el grado de granularidad analítica del estudio.

## **Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado**

**ADVERTENCIA:** Desde la Universidad consideramos que *ChatGPT* u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Julia Jenaro Barrio, estudiante de Doble Grado en Business Analytics y Relaciones Internacionales (E-6 + Analytics) de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado “Voces de la Diáspora: Analizando la Percepción Pública sobre la Migración Venezolana a través de las redes sociales”, declaro que he utilizado la herramienta de Inteligencia Artificial Generativa *ChatGPT* u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. Referencias: Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
2. Metodólogo: Para descubrir métodos aplicables a problemas específicos de investigación.
3. Interpretador de código: Para realizar análisis de datos preliminares.
4. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.
5. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja.
6. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
7. Traductor: Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: Junio 2025

Firma: Julia Jenaro Barrio

# Referencias

- Abrams, E. (2024). Council on foreign relations: Chevron, the biden administration, and the maduro regime. <https://www.cfr.org/blog/chevron-biden-administration-and-maduro-regime>. Accedido: Noviembre 15, 2024.
- ACNUR (2024). Entrada y regularización documental de personas venezolanas en brasil. <https://help.unhcr.org/brazil/es/informacion-para-la-poblacion-venezolana/16493-2/>. Accedido: noviembre 17, 2024.
- ACNUR España (2023). Llamamiento de emergencia: Situación de venezuela. <https://www.acnur.org/es-es/emergencias/situacion-de-venezuela>. Accedido: Octubre 7, 2024.
- Affirma Legal (2024). Visa de residente en colombia - visa tipo r. <https://www.affirmalegal.com/blog/visa-de-residente-en-colombia-tipo-r/>. Accedido: noviembre 17, 2024.
- Agüero, A. (2012). La dimensión geopolítica y económica del petróleo venezolano en las primeras décadas del siglo xx y su repercusión en el estado cojedes. Master's thesis, *Tiempo y Espacio*, vol. 22, no. 58, Caracas, dic. 2012.
- AZERTAC (2024). Agencia estatal de noticias de azerbaiyán: Datos curiosos sobre venezuela que debería saber. [https://azertag.az/es/xeber/datos\\_curiosos\\_sobre\\_venezuela\\_que\\_debera\\_saber-2466851](https://azertag.az/es/xeber/datos_curiosos_sobre_venezuela_que_debera_saber-2466851). Accedido: Noviembre 16, 2024.
- BBC News Mundo (2018). Venezuela: Nicolás maduro anuncia que le quitará 5 ceros al bolívar en medio de una galopante inflación que el fmi prevé que llegue al 1.000.000 % este año. <https://www.bbc.com/mundo/noticias-america-latina-44962277>. Accedido: Noviembre 16, 2024.
- Bermúdez, A. (2022). "los salarios de hambre": cómo el gobierno de maduro redujo los ingresos de los funcionarios públicos y desató una nueva ola de protestas en venezuela. <https://www.bbc.com/mundo/noticias-america-latina-62613543>. Accedido: Noviembre 16, 2024.
- Blei, D. M. and Lafferty, J. D. (2007). A correlated topic model of science. <https://www.cs.columbia.edu/~blei/papers/BleiLafferty2007.pdf>. Accedido: 2 de febrero, 2025.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of*

- machine Learning research*, 3(Jan):993–1022.
- Calderón, C. A., de la Vega, G., and Herrero, D. B. (2020). Topic modeling and characterization of hate speech against immigrants on twitter around the emergence of a far-right party in spain. *Social Sciences*, 9(11):188.
- CEMVE (2024). I estudio del cemve sobre integración de la migración venezolana en españa. <https://www.ongredreto.org/i-estudio-del-cemve-sobre-integracion-de-la-migracion-venezolana-en-espana/>. Accedido: noviembre 17, 2024.
- Center, N. I. J. (2023). Initial temporary protected status (tps) for venezuelan nationals (form i-821). [https://immigrantjustice.org/sites/default/files/content-type/services/documents/2023-10/I-821\\_Initial\\_TPS\\_Venezuela\\_English\\_2023-10-05.pdf](https://immigrantjustice.org/sites/default/files/content-type/services/documents/2023-10/I-821_Initial_TPS_Venezuela_English_2023-10-05.pdf). Accedido: noviembre 17, 2024.
- Citizenship, U. and Services, I. (2023). Processes for cubans, haitians, nicaraguans, and venezuelans. <https://www.uscis.gov/CHNV>. Accedido: noviembre 17, 2024.
- Citizenship, U. and Services, I. (2025). Estatus de protección temporal. <https://www.uscis.gov/es/programas-humanitarios/estatus-de-proteccion-temporal#:~:text=AVISO%3A%20El%205%20de%20febrero,designaci%C3%B3n%20del%202023%20para%20Venezuela>.
- Delgado Linero, M. (2022). La complejidad del fenómeno migratorio en méxico y sus desafíos. <https://www.undp.org/sites/g/files/zskgke326/files/2022-10/PNUDLAC-working-paper-30%20Mexico-ES.pdf>. Accedido: noviembre 17, 2024.
- Delgado Linero, M. (2024). Sobre los venezolanos y la política migratoria en méxico. <https://www.observatoriovenezolanodemigracion.org/post/sobre-los-venezolanos-y-la-politica-migratoria-en-mexico>. Accedido: noviembre 17, 2024.
- Diario, E. (2024). Venezolanos encabezan las solicitudes de asilo en españa. <https://eldiario.com/2024/09/09/venezolanos-encabezan-las-solicitudes-de-asilo-en-espana/>. Accedido: Noviembre 2, 2024.
- Doval, O. (2023). Sector eléctrico en venezuela. <https://efectococuyo.com/opinion/sector-electrico-en-venezuela/>. Accedido: Noviembre 15, 2024.
- Durán, M. A. (2021). Venezuela: estrategias para sobrevivir en el límite. <https://www.nuso.org/articulo/Venezuela-Maduro-crisis/>. Accedido: Noviembre 15, 2024.
- Economist Intelligence (2023). Democracy index 2023. <https://www.eiu.com/n/campaigns/democracy-index-2023/>. Accedido: Noviembre 16, 2024.
- Eguren, J. and Estrada, C. (2019). Análisis comparativo y transversal de la migración venezolana en iberoamérica. Master's thesis, Universidad Pontificia Comillas.
- ENCOVI (2021). Pobreza extrema población 2021. <https://insoencovi.ucab.edu.ve/indicador-pobreza/>. Accedido: febrero 4, 2025.
- Encovi (2023). Encovi 2023: 6 de cada 10 hogares están en pobreza extre-

- ma. <https://finanzasdigital.com/encovi-2023-6-de-cada-10-hogares-están-en-pobreza-extrema/>. Accedido: Noviembre 16, 2024.
- Español, C. (2024). ¿qué propone donald trump en materia de inmigración para su segundo mandato en ee.uu.? <https://cnnespanol.cnn.com/2024/11/06/que-propone-donald-trump-inmigracion-segundo-mandato-ee-uu-trax/>. Accedido: noviembre 17, 2024.
- Expansión (2021). Venezuela-Índice de desarrollo humano - idh. <https://datosmacro.expansion.com/idh/venezuela>. Accedido: Noviembre 16, 2024.
- Freire-Vidal, Y. (2019). Characterization of local attitudes toward immigration using social media. Master's thesis, Universidad del Desarrollo Santiago, Chile.
- Freire-Vidal, Y. & Graells-Garrido, E. (2019). Characterization of local attitudes toward immigration using social media. Master's thesis, Universidad del Desarrollo, Santiago, Chile.
- Gobierno de Colombia (2024). Cancillería. <https://www.cancilleria.gov.co/>. Accedido: Octubre 23, 2024.
- Gobierno de Perú (2024). Superintendencia nacional de migraciones. <https://www.gob.pe/migraciones>. Accedido: Octubre 23, 2024.
- González Oquendo, L. J. (2019). Colapso eléctrico y colapso gerencial en venezuela. Master's thesis, Universidad del Zulia.
- González-Rábago, Y. (2014). Los procesos de integración de personas inmigrantes: límites y nuevas aportaciones para un estudio más integral. Master's thesis, Universidad del País Vasco.
- Grupo Banco Mundial (2021). Integración social y económica de los migrantes venezolanos en colombia. <https://www.bancomundial.org/es/events/2021/11/11/integracion-social-y-economica-de-los-migrantes-venezolanos-en-colombia>. Accedido: noviembre 17, 2024.
- Gualda, E. and Rebollo, C. (2020). Big data y twitter para el estudio de procesos migratorios. Master's thesis, Universidad de Huelva.
- Guyana, C. U. (2024). Encovi 2023: los hallazgos más destacados sobre las condiciones de vida de los venezolanos. Master's thesis, Universidad Católica Andrés Bello.
- Harper, N. and Koh, F. (2024). Access to oil, geopolitical pressures: What is the us' motive behind easing venezuelan sanctions? <https://www.channelnewsasia.com/world/access-oil-geopolitical-pressures-what-us-motive-behind-easing-venezuela-sanctions-china-russia-3926136>. Accedido: Noviembre 15, 2024.
- Hutto, C. Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. Master's thesis, University of Michigan. Proceedings of the international aaai conference on web and social media (Vol. 8, pp. 216–225).
- Jiménez, C. (2020). La odisea de los venezolanos que se convirtieron en ciudadanos bra-

- sileños. <https://elpais.com/internacional/2020-12-19/la-odisea-de-los-venezolanos-que-se-convirtieron-en-ciudadanos-brasilenos.html>. Accedido: noviembre 17, 2024.
- Kaggle (2023). Immigration to brazil 2022-2023. <https://www.kaggle.com/datasets/danttis/immigration-to-brazil-2022-2023>. Accedido: 17 de noviembre, 2024.
- Kaggle (2025). Bolívar/dólar price 2010-2021. <https://www.kaggle.com/datasets/unestudiantemas/bolivardolar-price-20102021>. Accedido: enero 2, 2025.
- Kibe, K. (2024). Topic modeling using latent dirichlet allocation (lda). <https://www.analyticsvidhya.com/blog/2023/02/topic-modeling-using-latent-dirichlet-allocation-lda>. Accedido: 2 de febrero, 2025.
- Kitroeff, N. and Kurmanav, A. (2022). El gobierno de biden considera levantar algunas secciones a venezuela. <https://www.nytimes.com/es/2022/10/12/espanol/venezuela-sanciones-estados-unidos.html>. Accedido: Noviembre 16, 2024.
- Lugo, O. (2024). Los datos de opep que ratifican colapso petrolero de venezuela (y de pdvsa). <https://elestimulo.com/elinteres/energia/2023-07-18/los-datos-de-opep-que-ratifican-colapso-petrolero-de-venezuela-y-de-pdvsa/>. Accedido: Noviembre 15, 2024.
- Luzes, M. (2024). Evolución de la percepción pública hacia la migración en la región. <https://migdep.colmex.mx/destinomexico/marta-luzes.html>. Accedido: Enero 12, 2025.
- Mantilla Muñoz, W. A. and Velandia Ochoa, J. S. (2023). Detección de discursos polarizados y de odio en twitter, en el contexto de las elecciones presidenciales de colombia en 2022 usando natural language processing. Master's thesis, Universidad Unab.
- Melvin, J., Winston, K., Espejo, S., and Mogollon, M. (2024). Venezuela gets partial oil sanctions relief in exchange for election assurances. <https://www.spglobal.com/commodityinsights/en/market-insights/latest-news/oil/101923-venezuela-gets-partial-oil-sanctions-relief-in-exchange-for-election-assurances>. Accedido: Noviembre 15, 2024.
- Mimno, D., McCallum, A. and others (2011). Optimizing semantic coherence in topic models. <https://aclanthology.org/D11-1024.pdf>. Accedido: 2 de febrero, 2025.
- Ministerio de Inclusión, Seguridad Social y Migraciones (2024). Observatorio permanente de la inmigración (opi). [https://www.inclusion.gob.es/web/opi/estadisticas/observatorio\\_permanente\\_inmigracion](https://www.inclusion.gob.es/web/opi/estadisticas/observatorio_permanente_inmigracion). Accedido: noviembre 17, 2024.
- Miró-Llinares, F. (2016). Taxonomía de la comunicación violenta y el discurso del odio en internet. *IDP. Revista de Internet, Derecho y Política*, (22):1–16.
- Moslimani, M. (2024). How temporary protected status has expanded under the biden administration. <https://www.pewresearch.org/short-reads/2024/03/29/how-temporary-protected-status-has-expanded-under-the-biden-administration/>. Accedido: noviembre 17, 2024.

- Murel, J. and Kavlakoglu, E. (2024). What is latent dirichlet allocation? <https://www.ibm.com/think/topics/latent-dirichlet-allocation>. Accedido: 2 de febrero, 2025.
- Naciones Unidas (2022). “las alarmantes condiciones de vida en venezuela se agravan cada día”, alertan expertos en derechos humanos. <https://news.un.org/es/story/2018/02/1426751>. Accedido: Noviembre 16, 2024.
- Observatorio Gasto Público Cedice (2022). Evolución data macroeconómica. <https://cedice.org.ve/ogp/wp-content/uploads/2022/12/CVI-2022.pdf>. Accedido: Noviembre 16, 2024.
- OIM: ONU Migración (2024). Estudio de impacto fiscal de la migración venezolana en colombia: Realidad vs. potencial. <https://respuestavenezolanos.iom.int/sites/g/files/tmzbd1526/files/documents/2024-04/informe.pdf>. Accedido: noviembre 17, 2024.
- ONU (2023). Operación acogida brinda a los migrantes venezolanos un nuevo comienzo en brasil. <https://news.un.org/es/story/2023/04/1519997>. Accedido: noviembre 17, 2024.
- ONU (2024). Estadísticas migratorias para méxico. boletín anual 2023. <https://mexico.un.org/es/264927-estad%C3%ADsticas-migratorias-para-m%C3%A9xico-bolet%C3%ADn-anual-2023>. Accedido: noviembre 17, 2024.
- Organización Internacional para las Migraciones (2024). Integración y cohesión social de los migrantes. <https://www.iom.int/es/integracion-y-cohesion-social-de-los-migrantes>. Accedido: Enero 12, 2025.
- Our World in Data (2021a). Annual number of deaths by cause. <https://ourworldindata.org/grapher/annual-number-of-deaths-by-cause>. Accedido: 20 de diciembre, 2025.
- Our World in Data (2021b). Migration flows: Emigrants from venezuela. <https://ourworldindata.org/explorers/migration-flows?table&Select+a+country=Venezuela&Outflow+or+Inflow=Emigrants%3A+Where+people+born+in+the+selected+country+moved+to&country=USA~GBR~FRA~MEX~IND~CHN~DEU~AUS~BRA~ZAF>. Accedido: 17 de noviembre, 2024.
- Our World in Data (2023). Human development index. <https://ourworldindata.org/grapher/human-development-index?v=1&csvType=full&useColumnShortNames=false>. Accedido: 2 de diciembre, 2024.
- Our World in Data (2024a). Democracy index (eiu). <https://ourworldindata.org/grapher/democracy-index-eiu?v=1&csvType=full&useColumnShortNames=false>. Accedido: 2 de diciembre, 2024.
- Our World in Data (2024b). Electricity production by source. <https://ourworldindata.org/grapher/electricity-prod-source-stacked?v=>

- 1&csvType=full&useColumnShortNames=false. Accedido: 11 de diciembre, 2024.
- Our World in Data (2024c). Gdp per capita (maddison). <https://ourworldindata.org/grapher/gdp-per-capita-maddison?v=1&csvType=full&useColumnShortNames=false>. Accedido: 5 de diciembre, 2024.
- Pano, T. Kashef, R. (2020). A complete vader-based sentiment analysis of bitcoin (btc) tweets during the era of covid-19. Master's thesis, Big Data and Cognitive Computing, 4(4).
- Paredes, N. (2023). La operación con la que brasil está recibiendo a miles de migrantes venezolanos. <https://www.bbc.com/mundo/noticias-america-latina-65562186>. Accedido: noviembre 17, 2024.
- Pew Research Center (2021). U.s. hispanics: Venezuelan-origin population. <https://www.pewresearch.org/chart/us-hispanics-venezuelan-origin-population/>. Accedido: 2 de enero, 2025.
- Proyecto Migración: Venezuela (2024). Esta es la línea de tiempo de la migración venezolana a colombia. <https://migravenezuela.com/frontera/esta-es-la-linea-de-tiempo-de-la-migracion-venezolana-a-colombia/>. Accedido: noviembre 17, 2024.
- Reddit (2024). Reddit homepage. <https://www.reddit.com>. Accedido: Octubre 2, 2024.
- Romero Vega, R. (2021). Análisis de sentimientos en twitter para descubrir contenido xenófobo hacia los inmigrantes venezolanos en ecuador. Master's thesis, Universidad Nacional de Loja.
- Rowe, F., Mahony, M., Graells-Garrido, E., and Rango, M. (2021). Using twitter to track immigration sentiment during early stages of the covid-19 pandemic. *Data Policy*, 3.
- Secretaría de Gobernación (SEGOB) (2024a). Cuadros de boletín anual - política migratoria. <https://www.politicamigratoria.gob.mx/es/PoliticaMigratoria/CuadrosBOLETIN?Anual=2024&Secc=2>. Accedido: 15 de noviembre, 2024.
- Secretaría de Gobernación (SEGOB) (2024b). Cuadros de boletín anual - política migratoria. <https://portales.segob.gob.mx/es/PoliticaMigratoria/CuadrosBOLETIN?Anual=2024&Secc=3>. Accedido: 15 de noviembre, 2024.
- Silva Aguilar, A. and Ferré Pavia, C. (2024). El debate político en xtwitter sobre la inmigración venezolana en colombia: análisis desde el framing. Master's thesis, Universidad Complutense de Madrid.
- SOS Orinoco (2024). Sos orinoco: Salvando la amazonía venezolana. <https://sosorinoco.org/es/>. Accedido: Noviembre 16, 2024.
- Statista (2023). Ranking de países con mayor número de emigrantes procedentes de venezuela a nivel mundial en 2023. <https://es.statista.com/estadisticas/1261404/paises-con-mayor-numero-de-migrantes-venezolanos-en-el-mundo/>. Accedido: 29 de enero, 2025.
- Superintendencia Nacional Antidrogas (2023). Informe nacional antidrogas.

- <https://www.sunad.gob.ve/wp-content/uploads/2022/09/Portada-Informe-Pagina-web-01.png>. Accedido: Noviembre 16, 2024.
- Tun-Mendicuti, A., Kim, J., and Mulder, C. H. (2024). Understanding opinions towards migrants in transit: An analysis of tweets on migrant caravans in the us and mexico. In *Proceedings of the 2024 ACM Web Science Conference*. ACM.
- UNHCR ACNUR, OIM (2023). Recepción e integración de personas migrantes y refugiadas en ciudades de las américas. <https://www.acnur.org/sites/default/files/2023-06/recepcion-e-integracion-de-personas-migrantes-y-refugiadas.pdf>. Accedido: Enero 12, 2025.
- U.S. Census Bureau (2023). Venezuelans per state. <https://data.census.gov/table/ACSDT5YSPT2021.B25014?q=venezuelans%20per%20state>. Accedido: 2 de enero, 2025.
- U.S. Citizenship and Immigration Services (2024). U.s. citizenship and immigration services. <https://www.uscis.gov/>. Accedido: Octubre 23, 2024.
- Venemex (2018). Informe venemex: situación de venezolanos en méxico. <https://venemex.org/docs/20180713-informe-migracion-venemex.pdf>. Accedido: noviembre 17, 2024.
- Vitiugin, F., L. (2024). Unraveling code-mixing patterns in migration discourse: Automated detection and analysis of online conversations on reddit. Master's thesis, Aalto University Finland.
- Zambrano Sequín, L. (2022). Dolarización y desdolarización: ¿un dilema en venezuela? Master's thesis, Instituto de Investigaciones Económicas y Sociales, Universidad Católica Andrés Bello.