REPORT\_TEXT\_MINING (Trump Political campaign)

Table of Content

[**Introduction 3**](#_jnth06q2ak59)

[**Named Entities 4**](#_f291v7rtxxxv)

[Methodology 4](#_t8h4sj5lgo6t)

[Results 5](#_xatik0pfybi8)

[Relation to the business model 6](#_l5ukthuc3ry3)

[**Crowd Reaction Analysis 7**](#_i1ikoduz4d5z)

[Methodology 7](#_ind8natv9aua)

[Results 8](#_4gf03v482kcj)

[Relation to the business model 9](#_itl85f3o73ze)

[**Trump’s criticism on Biden 10**](#_2zzndoitkmuo)

[Methodology 10](#_hjdbw1ryuh6x)

[Results 12](#_b9endor05wqz)

[Relation to the business model 13](#_gmq19aj1yu0k)

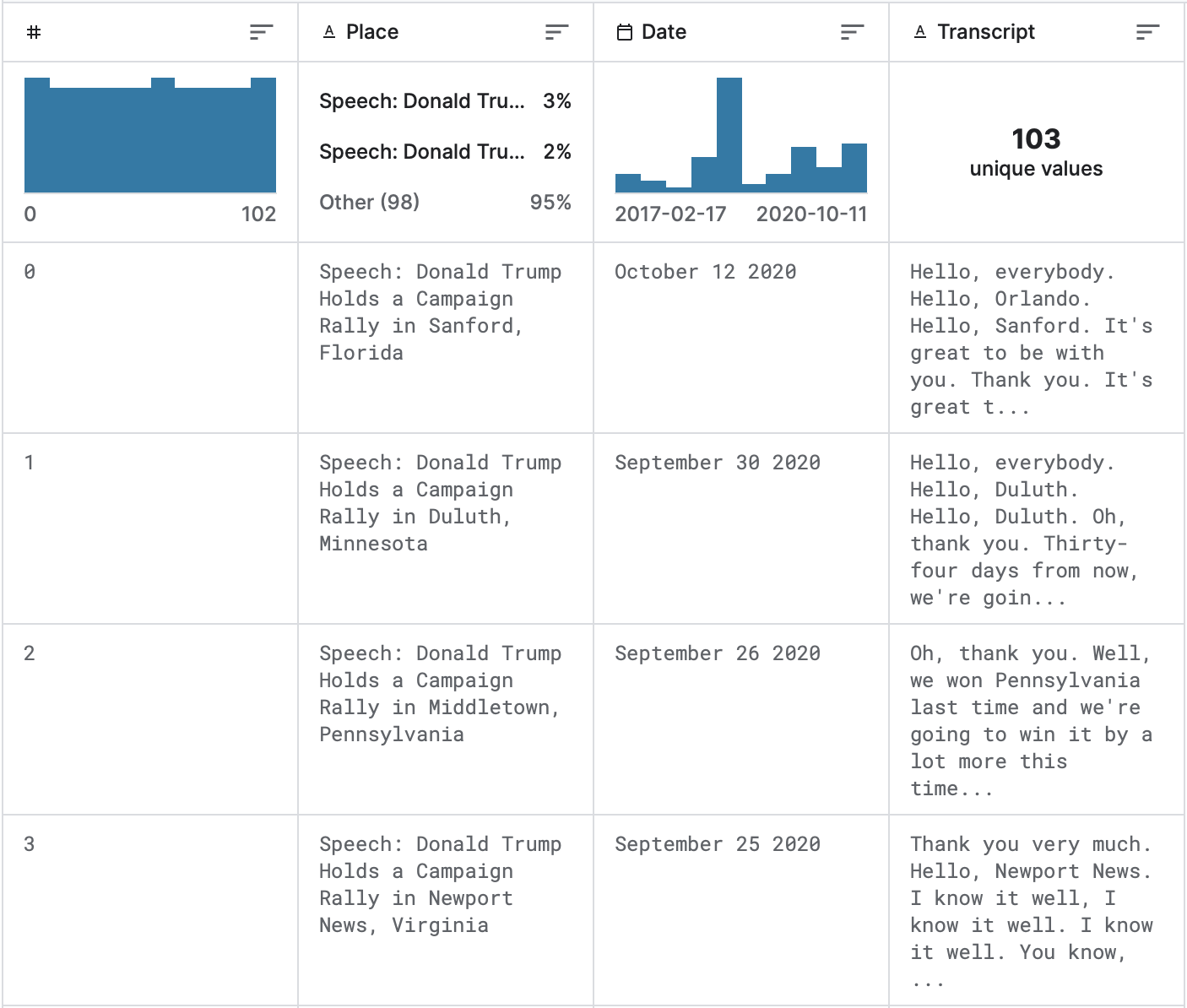
[**Conclusions 15**](#_r6e4ok4os5bm)

# Introduction

The last few years in U.S. history were marked by significant political polarization, with the presidency of Donald Trump serving as a focal point for numerous national debates and movements. His rallies, characterized by their direct and often confrontational style, were critical in mobilizing his supporters and influencing public opinion. These events not only served as platforms for political engagement but also reflected and influenced the broader dynamics within American politics. They highlighted persistent critiques against both the political establishment and the media, while frequently targeting the Democratic Party, thus shaping the political discourse of the era.

Our project's primary objective is to dissect these rally speeches to unearth the recurrent themes, claims, and strategies employed by Donald Trump to rally his base and criticize his opponents. By employing text mining techniques, we aim to systematically identify and analyze the patterns in his speech, revealing the underlying strategies that could be pivotal for the Democratic Party to understand and counteract effectively. This analysis is crucial not only for crafting pointed responses in political discourse but also for enhancing our strategic communication and campaign planning to resonate better with a broader electorate.

x

The dataset at our disposal is a comprehensive collection of transcripts from Donald Trump's political rallies, covering the period from his first rally in 2017 post-inauguration to his rally post-Covid-19 recovery in October 2020. It includes detailed information on the dates, locations, estimated attendance, and full transcripts of speeches. Additionally, parts of these transcripts capture audience reactions, such as cheering and chanting, which provide insights into the crowd dynamics and the impact of certain statements. Given the rich detail captured in these transcripts, the dataset serves as an invaluable resource for understanding the rhetoric employed and the reception it garnered across different states and cities.

This report is structured to provide a detailed analysis through several lenses. Firstly, we will delve into **Named Entities** to identify key figures, locations, and topics mentioned frequently, using a methodology rooted in natural language processing. Next, we focus on **Crowd Reaction Analysis** to gauge the emotional temperature during various segments of the rallies and its implications for the messages that resonated most. We will also specifically analyze **Trump's Criticism on Biden**. Each section will conclude with a discussion on the **Relation to the Business Model** of the Democratic Party, highlighting strategic insights and possible countermeasures. Lastly, we will reflect on **What Did Not Work**, considering the limitations and challenges encountered in our analysis, setting the stage for future research and strategy refinement.

# 

# Named Entities

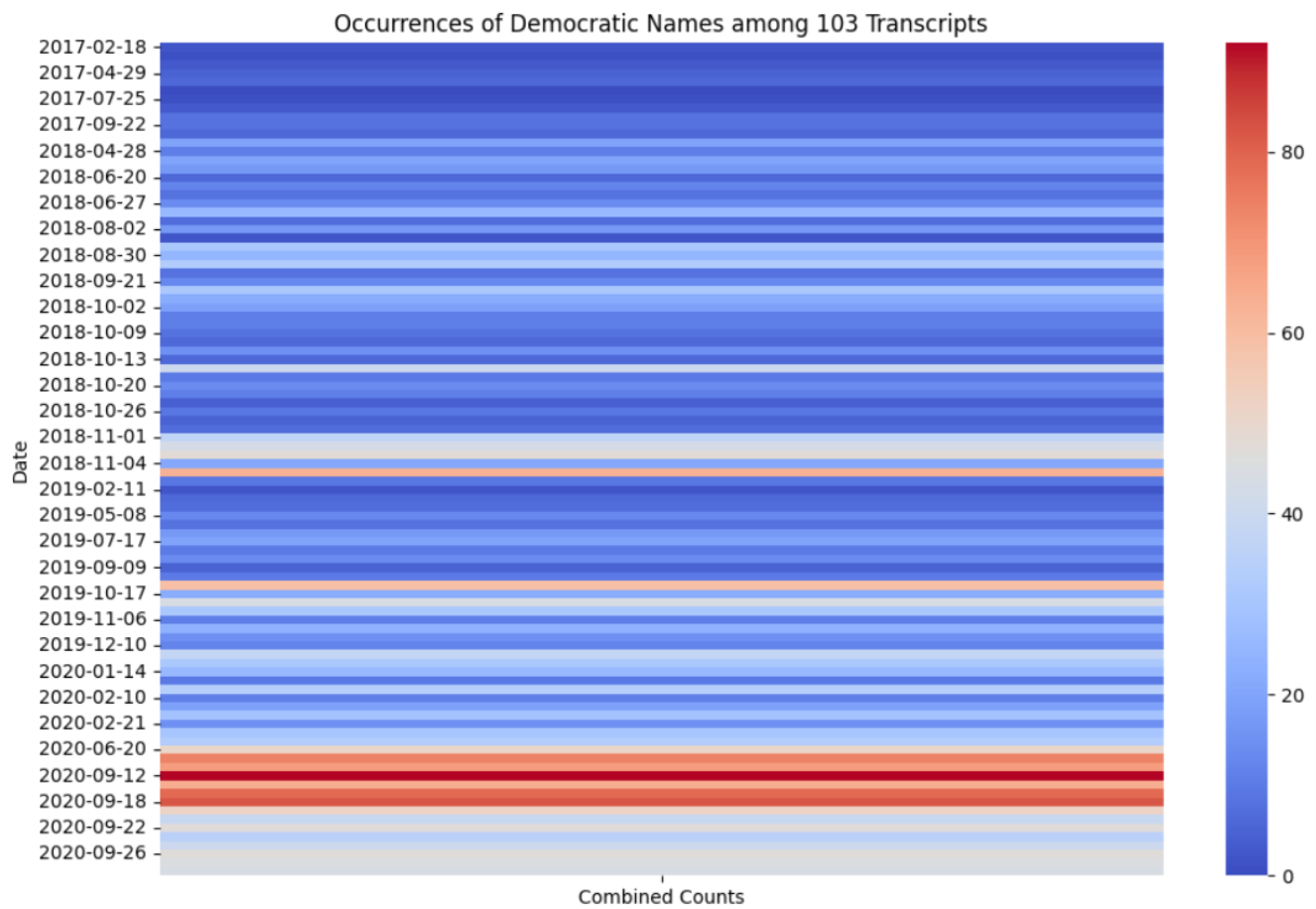
## Methodology

* **Cleaning text** : We removed all encodings and crowd reactions (which were between brackets) which were considered noise and not relevant for this step specifically , using regex patterns.
* **POS tagging and NER :** We did POS tagging first to identify the nouns within the text and thought it would help the NER as well because it emphasizes the nouns (although within python POS tagging is already integrated with NER). We retrieved the names of people ( using spacy’s “PERSON” tag) to see who are the main people targeted within Trump’s speech and among them the main democratic figures ( which concern our democratic party directly) who Trump is targeting. We had to combine some first names with corresponding family names as the results for each were separated at first.
* **Filtering dataset:** We filtered the dataset to keep all the sentences that contained named entities found from the previous step.
* **Lemmatization:** Standardizing words ensures consistency in entity recognition, as variations of the same word (e.g., "run," "running," "ran") are treated as the same entity. This improves the accuracy of the analysis and reduces dimensionality.
* **Filtering by frequency and Stopwords removal**: Filtering by frequency helps remove mainly very recurrent grammatical words. Removing stop words helps clean the data and reduce noise ( like auxiliary verbs etc ….), ensuring that the analysis is focused on relevant information. This step improves the quality of the results by eliminating distractions.
* **Cosine similarity:** We did a cosine similarity heatmap between all 103 transcripts ( filtered by the filters described previously to see if we found a pattern within Trump’s speech and detect if there’s an evolution or not. We used TF-IDF representation and cosine similarity because they were the best fitted for varying document sizes ( especially after processing steps). We found a pattern ( like a cluster within itself towards the end of 2020 transcripts (see picture below) which were nearing election time in November 2020).
* **Frequency analysis:** Trying to investigate the nature of the similarity of the clusters towards the end, we tried to see if there is any correlation between cosine similarity graph and number of democratic named entities mentioned (utilizing frequency analysis to quantify the occurrence of Democratic names, providing insights into Trump's targeting strategy). We obtained a similar pattern which drives us to conclude that Trump made a switch towards the end focusing more on attacking his opponents. Also, we can pick one transcript towards the end that will represent a pseudo-summary (or the most representative transcript concerning the democratic party among all other 103 transcripts.). We recommend reading one transcript among the last cluster to get more context and description of how the democratic figures were represented by Trump.

## Results

Cosine Similarity Heatmap of sentences containing named Entities among all 103 Transcripts over Time

## 



## Relation to the business model

We wanted to identify foremost the main democratic names targeted by Trump so we can advise our party on which people are concerned.

In the frenzied arena of political rallies, every word uttered carries weight, every name mentioned holds significance. Such is the case in the rallies orchestrated by Donald Trump, where the art of naming becomes a strategic tool in shaping narratives and influencing public opinion. As the curtains of the election draw closer, Trump's propensity to intensify the naming of Democratic political figures becomes increasingly evident, revealing a deliberate targeting strategy aimed at rallying his base and undermining his opponents.

Trump's rallies serve as a stage where the spotlight is often cast on individuals—be it allies or adversaries. Through meticulous analysis of named entities, it becomes apparent that Trump strategically amplifies the mention of Democratic leaders like Joe Biden, Hillary Clinton, and Bernie Sanders as the electoral clock ticks louder. Their names reverberate through the crowd, punctuating Trump's speeches with accusations, critiques, and rallying cries. This deliberate focus on Democratic figures underscores Trump's efforts to galvanize support among his base by framing the election as a battle against familiar opponents.

However, behind the veil of Trump's rhetoric lies an opportunity for the Democratic Party to wield named entity recognition as a potent tool in shaping their own narrative and countering his attacks. By dissecting the frequency and context of names mentioned in Trump's rallies, Democrats can glean valuable insights into his targeting strategy and vulnerabilities. Understanding which Democratic leaders and policies draw Trump's ire allows the party to tailor its responses, fortify its defenses, and pivot its messaging to resonate with voters.

Moreover, the prominence of certain names like Nancy Pelosi, Chuck Schumer, and Barack Obama underscores Trump's broader assault on Democratic institutions and policies. Recognizing this, the Democratic Party can leverage named entity analysis to not only defend its leaders but also to champion its agenda and rally public support. By strategically amplifying the voices of Democratic figures and highlighting the issues that resonate with voters, Democrats can counter Trump's narrative and seize control of the political discourse.

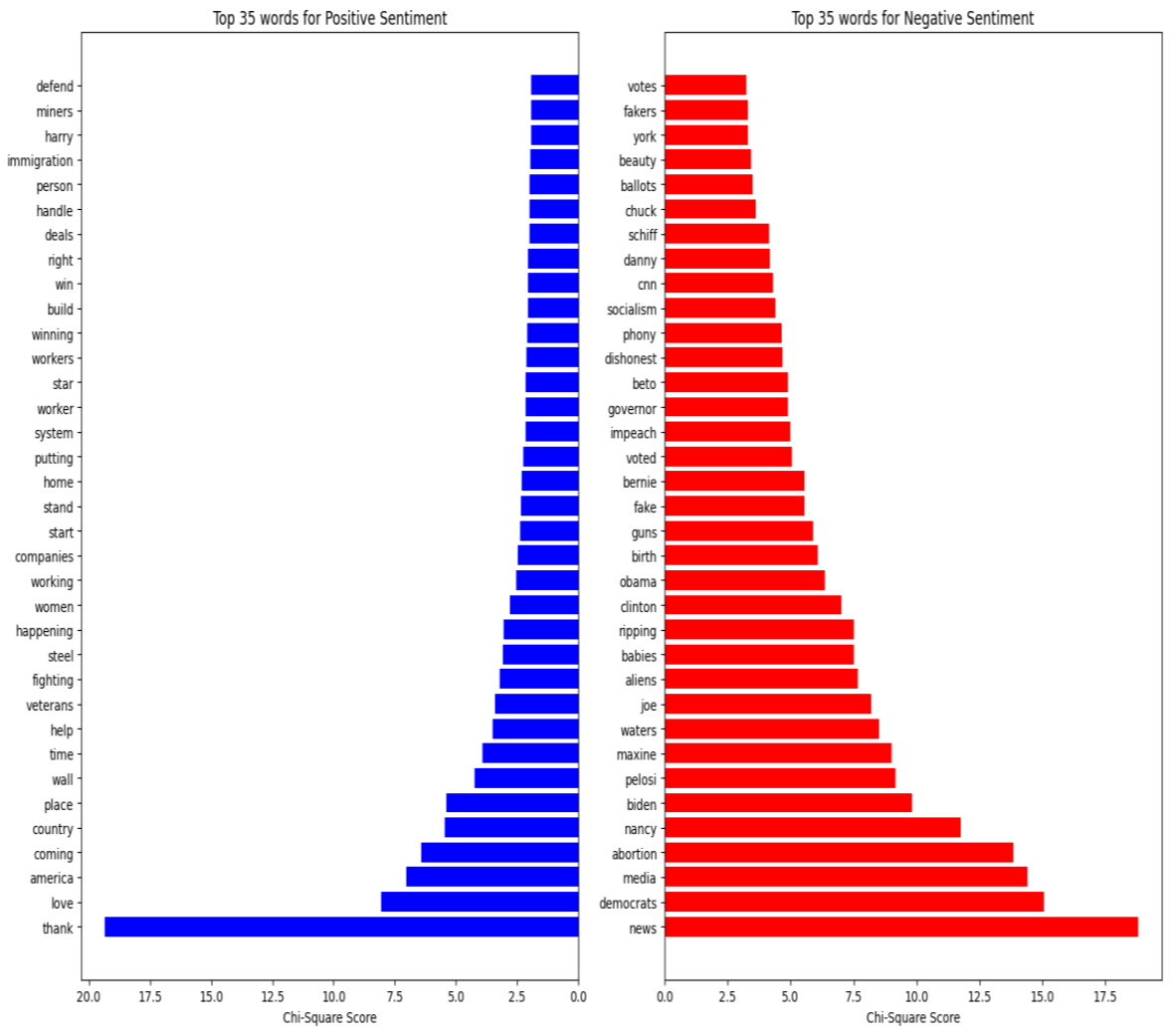
In essence, named entity recognition transcends mere linguistic analysis. It serves as a window into the intricacies of political strategy and discourse. Trump's relentless naming of Democratic figures in his rallies unveils his targeting tactics, laying bare the battleground upon which the electoral contest unfolds. Armed with this knowledge, the Democratic Party stands poised to harness the power of named entities as a shield against attacks and a sword to shape the narrative in their favor. As the election draws near, the significance of names uttered in Trump's rallies cannot be overstated. It is not merely a matter of rhetoric, but a battleground where the fate of the nation hangs in the balance.

# Crowd Reaction Analysis

## Methodology

* **Filtering dataset:** We filtered all the crowd reactions using regex pattern and with associated preceding phrase ( that triggered the reaction). By filtering the dataset to extract only phrases triggering crowd reactions, we narrow down the focus to the most impactful moments in Trump's rallies. This allows for a targeted analysis of the emotional responses elicited from the audience.
* **Lemmatization:** Lemmatization aids in standardizing words to their base or dictionary form, reducing variations and ensuring consistency in sentiment analysis. This helps in accurately capturing the emotional tone of crowd reactions.
* **Stop words:** Removing stopwords ( like repeated generic words like Audience, chants etc …)helps in cleaning the data and reducing noise, ensuring that the sentiment analysis is based on meaningful words that convey emotions. This step improves the accuracy of sentiment classification.
* **Sentiment Analysis using SentiWordNet:** SentiWordNet assigns sentiment scores to words based on their positivity or negativity, allowing for automated sentiment analysis. This enables the classification of crowd reactions into positive or negative categories, providing insights into the audience's emotional response.
* **Class creation:** Assigning labels based on sentiment scores to categorize reactions for further analysis.Creating classes based on sentiment scores facilitates the categorization of crowd reactions into meaningful groups, such as positive or negative. This classification simplifies the analysis and allows for targeted interpretation of sentiment trends and allows feature extraction to find most discriminative words between the 2 classes.
* **POS tagging:** Retaining only nouns, verbs, and adjectives ( of the preceding phrases) for more meaningful analysis. ( we also added a stop words list to remove meaningless auxiliary verbs).
* **TF-IDF and CHI square analysis:** TF-IDF (Term Frequency-Inverse Document Frequency) and Chi-square analysis highlight words that are statistically significant in distinguishing between positive and negative reactions. This information helps in understanding the factors driving sentiment and informs strategic responses to counter negative perceptions. The results showed some meaningless verbs that we added to the stop words list to get the next best chi-square score and the final results are in the figure below.

## Results



Interpretation of results (above table) : We did a bit of research with the words we got to better interpret the idea behind them. Mainly , the crowd reacted positively to hype statements like “ thank you for coming”, “we’re gonna make America great again”, “we’re gonna build a wall”, other positive reactions were triggered when Trump mentioned US veterans, US women, US workers etc …… On the other hand , the subjects that triggered negative reactions were mainly when Trump was mentioning the fake news and fake representation of the republican party in the news ( words in table were : news, media and CNN). Other topics were as follow ( I will give the word in table then the meaning behind it in parenthesis) :

* Abortion ( the republican crowd don’t agree with democrats to legalize abortion)
* Guns ( the republican crowd are against confiscating guns)
* Ballots , votes, dishonest ( Trump was referring to corruption in the voting process among democrats)
* Aliens ( Trump accuses democrats of welcoming illegal aliens ( foreigners) into the country

Also a lot of named entities appeared in results which demonstrates the key figures who are disliked by the republican crowd. ( Maxine Waters, Joe Biden, Nancy Pelosi, Adam Schiff, Danny O’connor,Chuck Schummer).

## Relation to the business model

The rallies gathered 20,000 people on average and so with over 103 rallies, the total number gathered was over 2 million people. It is crucial to analyze crowd reaction and discover which topics touches the heart of the people the most. Those topics will weigh more than others since the people represent the voters and winning them over is crucial to win the election.

As the Democratic Party, we recognize the power of crowd reactions as a barometer of public sentiment and a catalyst for change. In the face of divisive rhetoric and fear-mongering tactics used by our opponents, such as Trump's insistence on building a wall and demonizing immigrants as "aliens," crowd reactions serve as a rallying cry for unity and compassion. When we hear the voices of those who reject hate and embrace inclusivity, it reaffirms our commitment to building bridges, not walls and standing up for the rights of all Americans, regardless of background or origin.

During the 2017 to 2020 presidential run, the stark contrast between the Democratic and Republican positions on reproductive rights became abundantly clear. While Joe Biden and the Democratic Party championed a woman's right to make her own healthcare decisions, including access to safe and legal abortion, Donald Trump and the Republican Party sought to restrict these rights and roll back decades of progress. Trump's administration implemented policies aimed at defunding Planned Parenthood, appointing anti-choice judges to federal courts, and pushing for restrictive abortion legislation at the state level.

In the face of such attacks on reproductive freedom, crowd analysis becomes a vital tool for the Democratic Party to combat misinformation and mobilize support. By amplifying the voices of those who advocate for reproductive justice and bodily autonomy, we can push back against attempts to erode women's rights and control over their own bodies. Crowd reactions serve as a powerful testament to the urgency of protecting and expanding access to reproductive healthcare for all Americans, regardless of gender or socioeconomic status.

Moreover, crowd analysis enables us to address the risk of misrepresentation and distortion perpetuated by the likes of Trump and his allies. By providing factual information and engaging in honest dialogue with the public, we can counteract harmful narratives and ensure that our message is accurately conveyed. Through strategic communication and outreach efforts informed by crowd analysis, the Democratic Party can effectively advocate for policies that uphold the rights and dignity of all individuals, while exposing the divisive tactics employed by our opponents.

Rectifying misconceptions and communicating our truth is paramount for the Democratic Party in navigating the complex terrain of public sentiment. Beyond addressing negative perceptions surrounding the party's stance on abortion, it's imperative to tackle all subjects that elicited adverse reactions, including those highlighted by Trump's divisive rhetoric. By employing strategic communication, we can reclaim narratives surrounding issues such as immigration, media integrity, and political leadership. Equally important is amplifying the positive reactions, such as the support for aiding veterans. As a party that values all citizens of America and reveres the sacrifices made by veterans in service of their country, we must emphasize our unwavering commitment to supporting and uplifting those who have served. Through transparent dialogue, empathetic engagement, and a dedication to truth, the Democratic Party can forge meaningful connections with the public, dispel misinformation, and champion policies that reflect our shared values of compassion, equality, and unity

# 