



# FROM POLLS TO POSTS: WILL THE REAL WINNER PLEASE STAND UP?

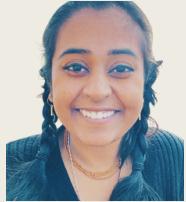
COMPARING THE PREDICTIVE POWER  
OF POLLING AND SOCIAL MEDIA  
SENTIMENT FOR U.S.  
PRESIDENTIAL ELECTIONS 



| July 2025

# OUR TEAM

Swathi  
Murali



POLLING DATA &  
BASELINE MODELING

August  
Vollbrecht



YOUTUBE DATA  
COLLECTION

Waleed  
Abedin



SENTIMENT ANALYSIS  
MODELING

Our team brings together skills in data scraping, machine learning, and analytics to tackle real world questions about elections and predictive modeling. Swathi focused on polling and baseline models, August led YouTube data collection, and Waleed handled sentiment analysis. Together, we explored what it takes to predict a U.S. presidential election winner.

# PROBLEM STATEMENT

How accurate is polling historically?

Can social media sentiment out perform or enhance traditional polling methodology?

Our project puts polls and social media head to head to find out which data source can call it better.



# THE BIG WHY?

## From Trusted Polls to Real-Time Signals

U.S. presidential elections. Polls gave campaigns, analysts, and voters a sense of where things stood. But now, with new technology and social media, we have the ability to analyze millions of opinions in real time. The question is whether these new tools are more accurate or if polls are still the best bet.

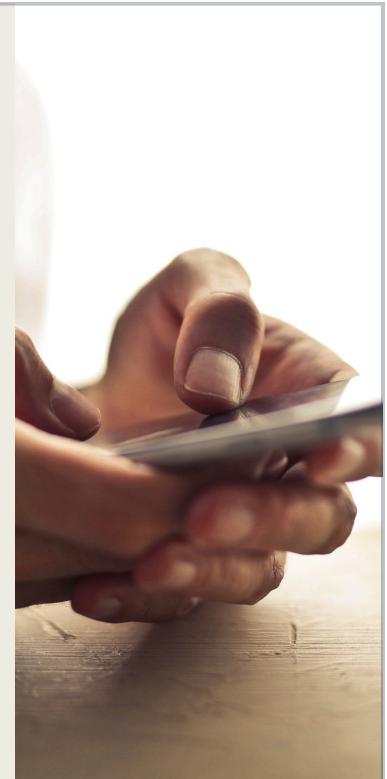
### FROM TRUSTED POLLS TO REAL-TIME SIGNALS

DRIVING CAMPAIGN DECISIONS

WHO SHOULD ANALYSTS AND PUNDITS TRUST?

WHY THIS MATTERS FOR EVERYONE

(03)



## Driving Campaign Decisions

Campaigns rely on predictive tools to decide which voters to target, historically this was heavy reliance on polling: what issues to emphasize, and how to position their candidate. If social media sentiment better captures the public mood and trending topics, it could reshape campaign strategy, messaging, and ground game. On the other hand, if polling remains the most reliable method, traditional tactics stay center stage.

(04)



I'm gonna read you AI's funny obv Cuomo tweaked desc of a case of social media being a powerful tool in politics:

Based on available information, Zohran Mamdani did not make money directly from commenting on Andrew Cuomo's Instagram post. Instead, his reply to Andrew Cuomo's relaunch video included a fundraising link, leading to a significant influx of donations to his campaign.

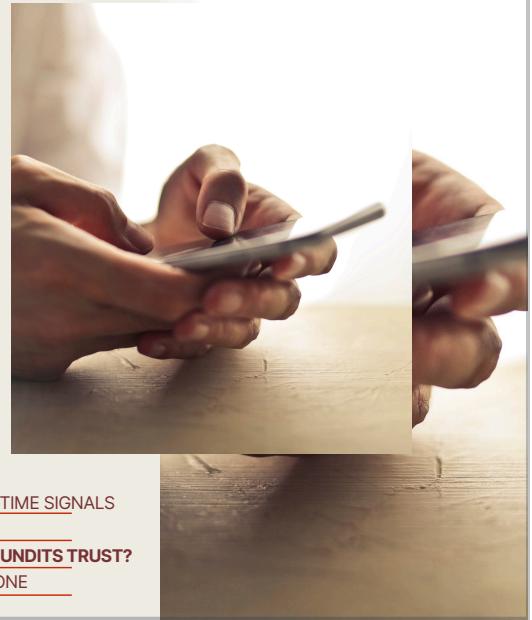
# THE BIG WHY?

## Who Should Analysts and Pundits Trust?

Media outlets and political analysts have long used polls to frame election narratives, allocate coverage, and forecast outcomes. As social media analytics become more sophisticated, there is a real question about which signals are worth following and how to avoid missing big shifts in voter sentiment, like in 2016.

(05)

FROM TRUSTED POLLS TO REAL-TIME SIGNALS  
DRIVING CAMPAIGN DECISIONS  
**WHO SHOULD ANALYSTS AND PUNDITS TRUST?**  
WHY THIS MATTERS FOR EVERYONE



Election forecasts are not just for pundits—they shape campaign strategy, media narratives, and even voter turnout.

Traditional polling has been trusted for decades, but big surprises like 2016 made everyone question if it is still enough.

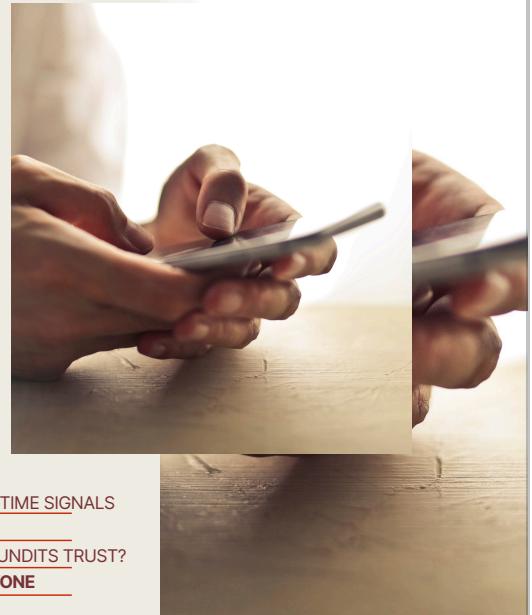
With social media everywhere, analysts have to decide: stick with classic polls, follow online sentiment, or try both?

Understanding these signals is key to not missing major shifts in public mood.

# THE BIG WHY?

## Why This Matters for Everyone

Understanding which data source is most predictive helps campaigns, journalists, and voters see past the hype and make smarter decisions. Whether polls or social sentiment turn out to be more accurate, our project sheds light on what tools will shape the future of U.S. presidential campaigns and how Americans experience and influence the election process.



(06)

FROM TRUSTED POLLS TO REAL-TIME SIGNALS  
DRIVING CAMPAIGN DECISIONS  
WHO SHOULD ANALYSTS AND PUNDITS TRUST?  
WHY THIS MATTERS FOR EVERYONE

For campaign strategists, the stakes are high: knowing which signal to trust can mean the difference between focusing resources where it matters and missing crucial battlegrounds.

For voters, being able to see beyond the headlines and hype gives them more power to make informed choices, instead of just following the narrative set by polls or viral trends.

Ultimately, the way we measure and interpret public opinion is shaping not just who wins, but how Americans engage with democracy itself.

# OUR DATA SOURCES & WHY THEY MATTER

PROJECT OVERVIEW



(07)

Where:

- National opinion polls (Wikipedia)
- Official election results (270towin)
- Social Media Sentiment (YouTube)

I'd drop the & why they matter because each of us will talk about that when we introduce the data we were scraping. less is more on slides :)

I'd also drop the "why did we choose these sources for our comparison" each of us will cover that when they talk about the data they collected...

# SCRAPING AND BUILDING THE POLLING DATASET

year	Poll_Leading_Margin	Poll_Leader	EC_election_winner	PC_election_winner	EC_Poll_Accurate	PC_Poll_Accurate
0	1936	5.89	0	0	0	1
1	1940	5.38	0	0	1	1
2	1944	4.89	0	0	1	1
3	1948	8.60	1	0	0	0
4	1952	12.11	1	1	1	1
5	1956	20.83	1	1	1	1
6	1960	2.93	0	0	1	1
7	1964	38.22	0	0	1	1
8	1968	8.08	1	1	1	1
9	1972	23.89	1	1	1	1
10	1976	10.80	0	0	1	1
11	1980	9.24	0	1	1	0
12	1984	11.95	1	1	1	1
13	1988	9.50	0	1	1	0
14	1992	11.76	0	0	0	1
15	1996	15.95	0	0	1	1
16	2000	5.87	1	1	0	1
17	2004	3.57	1	1	1	1
18	2008	5.04	0	0	1	1
19	2012	2.59	0	0	1	1
20	2016	3.85	0	1	0	1
21	2020	7.15	0	0	1	1
22	2024	1.27	1	1	1	1

year	Poll_Leader	EC_election_winner	ALYSIS
0	1936	Democrat	Democrat
1	1940	Democrat	Democrat
2	1944	Democrat	Democrat
3	1948	Republican	Democrat
4	1952	Republican	Republican
5	1956	Republican	Republican
6	1960	Democrat	Democrat
7	1964	Democrat	Democrat
8	1968	Republican	Republican
9	1972	Republican	Republican
10	1976	Democrat	Democrat
11	1980	Democrat	Republican
12	1984	Republican	Republican
13	1988	Democrat	Republican
14	1992	Democrat	Democrat
15	1996	Democrat	Democrat
16	2000	Republican	Republican
17	2004	Republican	Republican
18	2008	Democrat	Democrat
19	2012	Democrat	Democrat
20	2016	Democrat	Republican
21	2020	Democrat	Democrat
22	2024	Republican	Republican

(08)

Collecting & Cleaning  
Features Engineering  
Key challenges and surprises during this process

This is how I built the core dataset for our analysis of how polling has historically done in predicting election outcomes as well as in our model, how accurate it could potentially be in predicting future outcomes.

Started by scraping opinion polling data from Wikipedia for every U.S. presidential election cycle since 1936.

Matched this data with official election results from 270toWin, including both Electoral College and popular vote outcomes for each year.

Calculated the average poll percentage for both Democratic and Republican candidates for each election year.

Determined the overall poll leader for each year, not just based on one last poll, but as an average across the whole cycle.

Created a “Poll Leading Margin” feature by subtracting the trailing party’s polling percentage from the leader’s, so we could measure not just who led, but by how much.

Added new columns to show:

Electoral College vote share per party

Popular vote share per party

Who actually won the Electoral College

Engineered an “EC\_Poll\_Accurate” column to flag if the poll leader matched the actual winner (1 if yes, 0 if not).

Then cleaned the data set for use for my baseline model with the columns shown here.

Then I

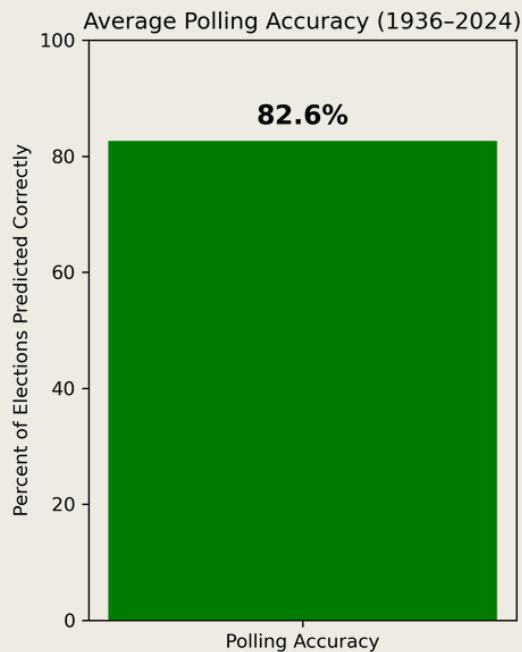
Renamed columns with party names for a clearer visual, so we could easily see which years the polls matched the outcome and which years they missed.

All of these steps make it possible to dig into not just whether polling was right or wrong, but also to understand why—based on lead size, vote share, and trends over time.

# CAN POLLS STILL PREDICT THE WINNER?

BASELINE MODELING & ANALYSIS

At first glance...



(09)

At first glance, it looks like polls do a pretty solid job at predicting who will win. If you just average the number of times the polls picked the right winner across all election years, you get an accuracy score of about 82.6 percent.

This is the basic historical “batting average” for polls—how often the polling leader actually ended up winning the Electoral College.

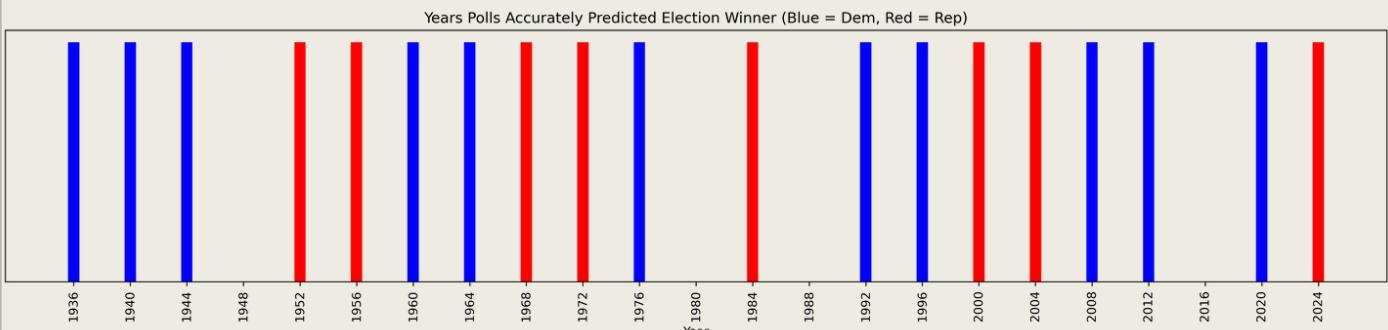
But this score doesn’t tell the whole story. It does not show us which years the polls got wrong, because it was still inaccurate a good chunk of the time, so even how big the misses were, or whether certain types of elections are easier or harder to predict.

To really test if polls have true predictive power, we have to dig deeper than just the average. We need to look at upsets, trends over time, and where polling consistently fails.

# CAN POLLS STILL PREDICT THE WINNER?

BASELINE MODELING & ANALYSIS

When you start to look deeper ...



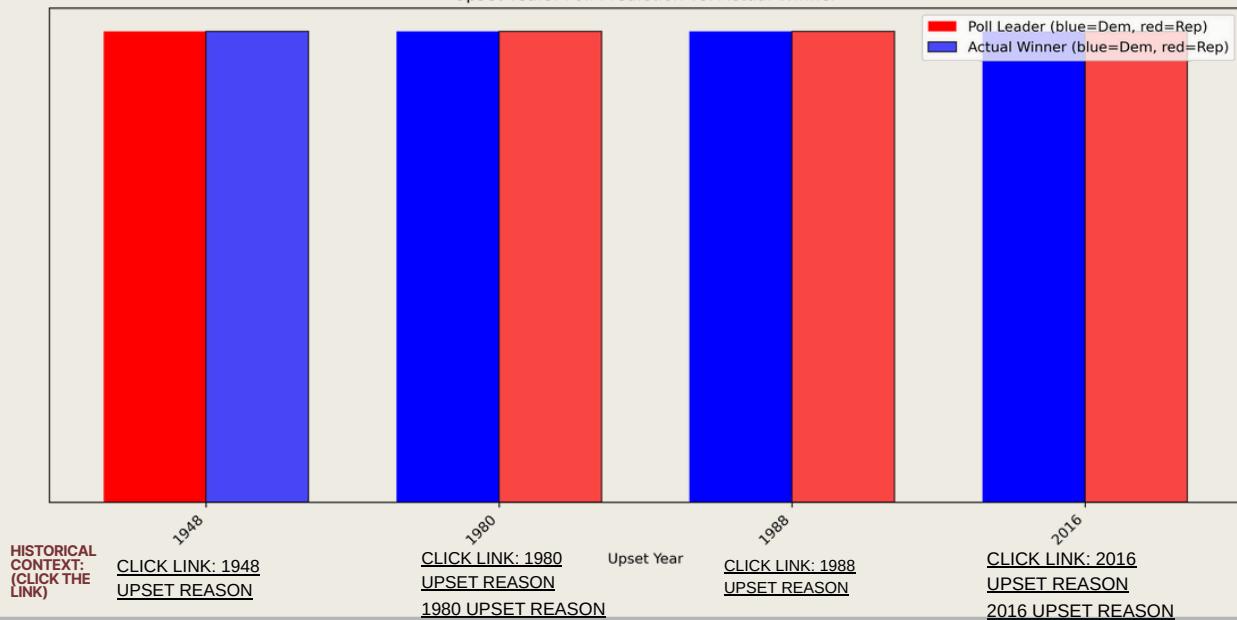
(10)

I built this chart to show, year by year, whether polls predicted the winner. Bars mean polls were right—blue for Democrats, red for Republicans. Years without a bar are the upset years, and those are the ones I wanted to understand.

# CAN POLLS STILL PREDICT THE WINNER?

What happened during the upset years?

Upset Years: Poll Prediction vs. Actual Winner

[CLICK LINK: HISTORIC CONTEXT FOR POLLING](#)


Digging into the history, I saw that these upsets almost always followed big, late shifts in public opinion, usually right after a major debate or campaign event—things that polls often missed.

In 1948 (Truman vs. Dewey), everyone assumed Dewey would win. Pollsters stopped polling early, headlines declared Dewey the winner, but a late surge gave it to Truman.

In 1980 (Carter vs. Reagan), polls showed a close race, but Reagan's debate performance and positive message in the final week won over undecided voters, leading to a landslide.

In 1988 (Dukakis vs. Bush), negative ads and Bush's steady image turned the tide at the last minute, and polls could not keep up.

In 2016 (Clinton vs. Trump), national polls showed Clinton ahead, but state polls missed crucial shifts and underrepresented key voter groups. Trump swung states like Michigan and Pennsylvania at the end.

The common thread: In these upsets, there was always a late rush of sentiment that polls failed to track—often sparked by debates or last-minute campaign moments. This happened both in the “old school” polling years and in the age of social media.

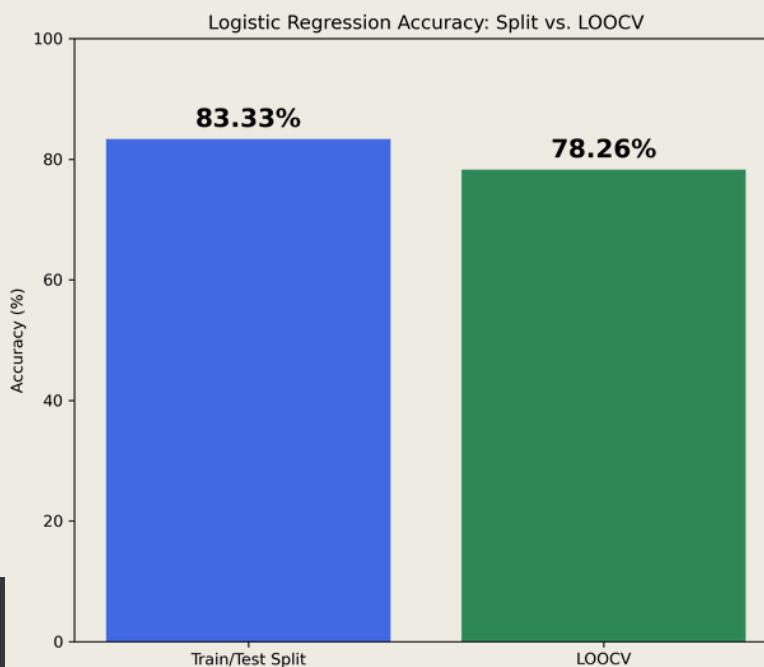
It really makes me wonder: If we had better ways to track real-time public mood—like sentiment analysis from YouTube or Twitter—could we have seen these upsets coming? With attitudes shifting so quickly online, this is why our project is looking at both polling and social media to see

what works best.

# CAN POLLS STILL PREDICT THE WINNER?

BASELINE MODELING & ANALYSIS

What did the Baseline Models Say?



For my baseline, I went with logistic regression because it's what pollsters typically use and it works best when you have a small dataset with a simple binary outcome like "Democrat or Republican wins."

When I ran a basic train/test split, my model predicted about 83 percent of the elections correctly. That's almost exactly the same as our historical polling average.

But the thing is, with only 23 elections, one lucky or unlucky split can swing the results a lot. That's why I wanted a more stable and fair measure.

So I used something called Leave-One-Out Cross-Validation, or LOOCV. This method trains the model on all the years except the last one and tests the model on every single year, one at a time, so it gives you a true average of how well the model would do on any possible scenario, not just one random split.

When I did that, the mean accuracy dropped a little to 78 percent. That makes sense, because LOOCV is less likely to be thrown off by outliers and gives a more realistic estimate of the model's performance. Because, again, the generalizability.

The main takeaway here is that LOOCV gives me a stronger sense of how well the model can actually generalize to new elections, not just the data it has already seen. It reminds me to be careful not to over-interpret any single accuracy score, especially with a small dataset.

So for the rest of my analysis, I'm sticking with LOOCV to keep my results grounded and as reliable as possible. Overall, polling models using logistic regression are pretty strong—but as we've seen, they're not perfect, and those upset years matter.

# DATA COLLECTION: YOUTUBE COMMENTS

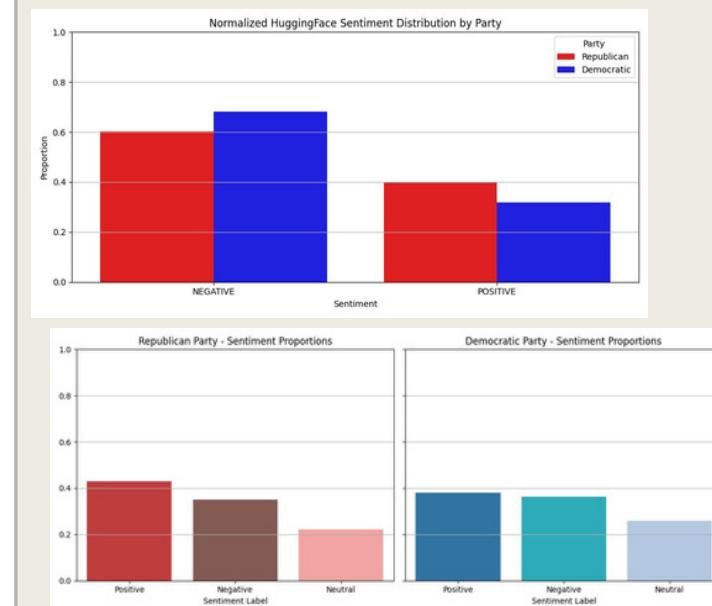
- search
- scrape
- refine
- repeat

(13)

- started by thinking we could use facebook found some very threatening videos urging me to not even think about it, apparently i was committing a lot of thought crimes
- scrapes around 700,000 comments
- 73 individually selected videos base don media outlet
- this was the basis of modeling 3 test case campaigns.

# MEASURING SENTIMENT: HOW WE TRACKED THE MOOD

SENTIMENT ANALYSIS

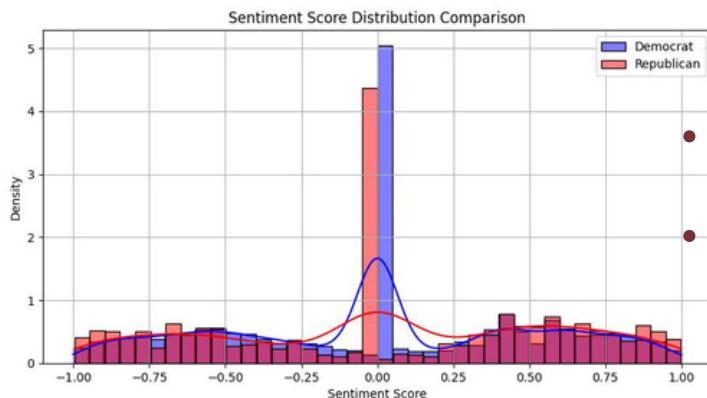


- Sentiment analysis models:
  - Vader NLP
  - Hugging Face

(14)

# SENTIMENT MODEL RESULTS: DOES THE MOOD MATCH THE OUTCOME?

SENTIMENT ANALYSIS



- How well did sentiment analysis predict election winners?
- Any interesting patterns, upsets, or trends from the sentiment data?

(15)

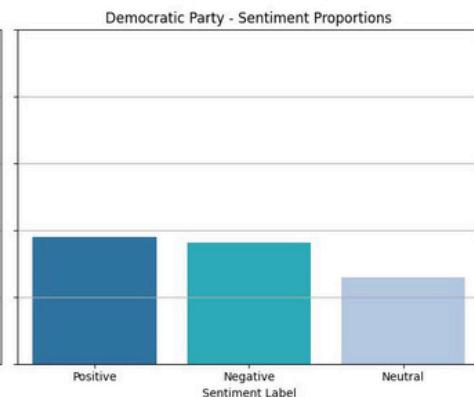
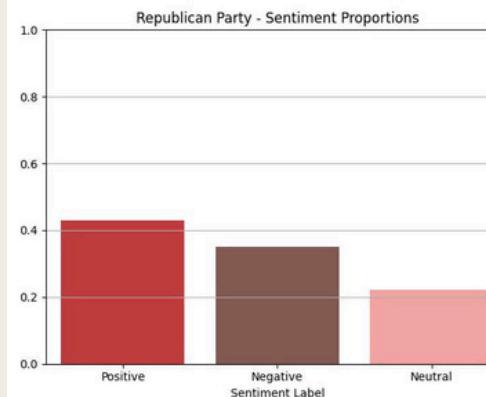
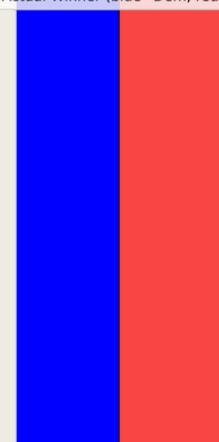
# KEY TAKEAWAYS:

RECOMMENDATIONS

## WHAT DID WE LEARN?

Upset Years: Poll Prediction vs. Actual Winner

Poll Leader (blue=Dem, red=Rep)  
Actual Winner (blue=Dem, red=Rep)



(16)

### Key Takeaways: What Did We Learn?

Looking at 2016, you can really see how sentiment analysis picked up on something that polling just did not catch. The YouTube comment sentiment for that year gave us a clue about why the polls were so off—there was a whole segment of voters whose attitudes and enthusiasm never really showed up in the polling data.

After 2016, there was a lot of talk about whether polls could really be trusted and what we were missing by relying only on those traditional methods. If campaigns, analysts, and even the media had been tracking sentiment across platforms like YouTube in real time, it might have been possible to spot these changes in mood and enthusiasm before Election Day.

As technology keeps moving forward, I think we will see more people turning to sentiment analysis to get a fuller, more realistic picture of what is going on. Polling is still useful for drilling down into specific questions or targeting messaging, but tracking overall sentiment could become just as important—maybe even more important—for actually predicting what is going to happen.

For analysts and strategists, this means being ahead of the curve and giving campaigns data they can actually use to adjust strategy on the fly. For voters, it means having a more honest view of the national mood and what is actually resonating. And for everyone, it opens up new ways to understand and respond to the changes that are happening in real time during an election season.

It is important to mention that our sentiment analysis is tracking things post-election, so there are some real limitations that come with looking at the data after the fact rather than in the moment.

We will get into those challenges and what we can do moving forward on the next slide.

# LIMITATIONS & NEXT STEPS

CONCLUSION

- Different sentiment analysis models offer varying strengths and limitations
- Similarly, different social media platforms tend to reflect the views of distinct segments of the population
- Next Steps:
  - Conduct a more in-depth evaluation of the performance and suitability of various sentiment analysis models
  - Expand data collection to include more diverse sources

(17)