

POLLING AND SOCIAL MEDIA SENTIMENT U.S. PRESIDENTIAL ELECTIONS

A COMPARATIVE OBSERVATIONAL ANALYSIS

OUR TEAM

August
Völlbrecht



DATA SOURCING
& COLLECTION

Swathi
Murali



POLLING DATA &
BASELINE MODELING

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 accuracy 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?

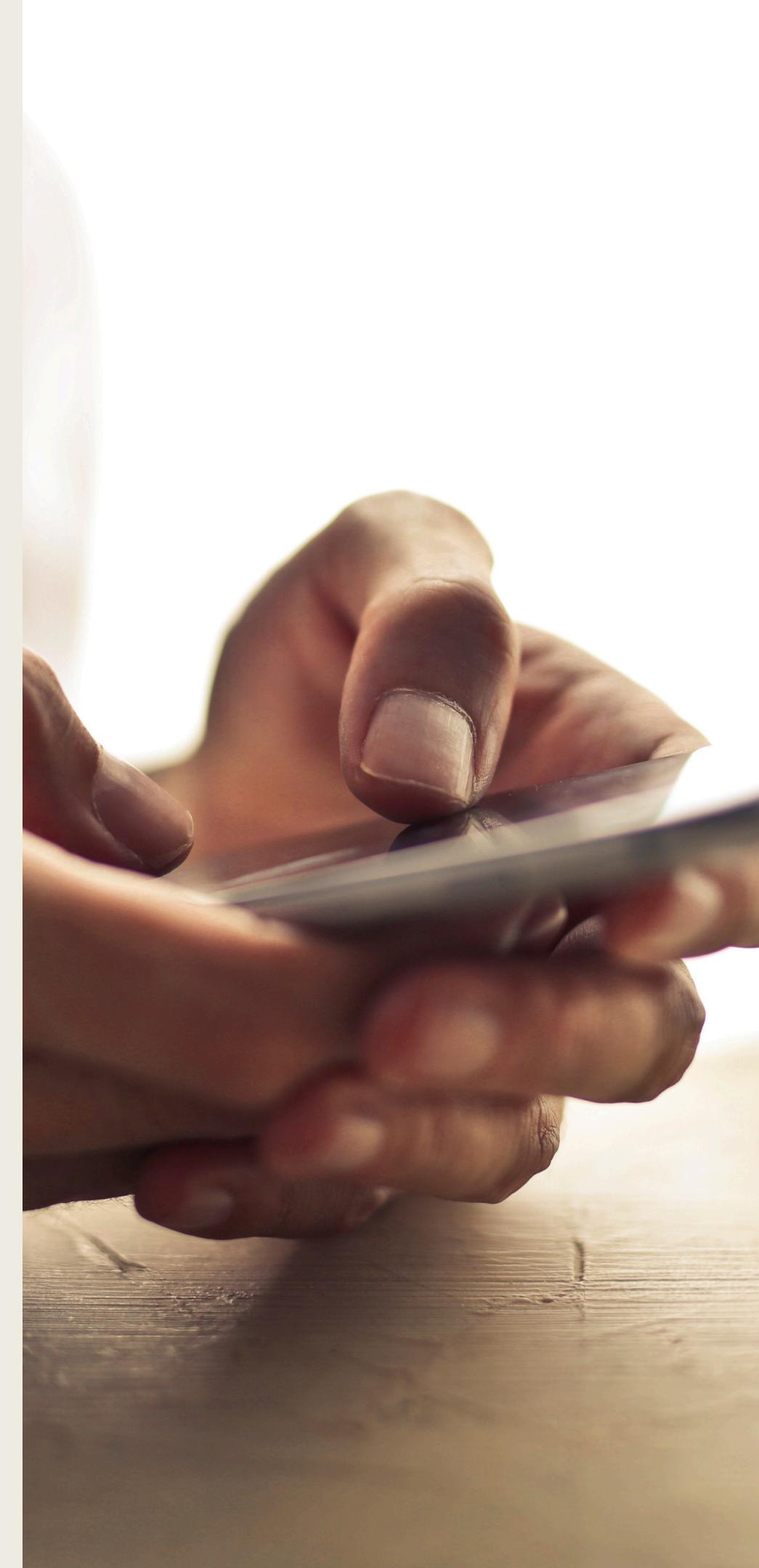
Our project puts polls and social media head to head to find out which data source can call it better.



THE BIG WHY?

Trusted Polls vrs Real-Time Sentiment

Historically 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 out perform traditional polling methodology.



PROJECT OVERVIEW

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)

Andrew Cuomo  @andrewcuomo · 1d
In it to win it.

1:29

7.7K 4.1K 5.1K 16M

Zohran Kwame Mamdani  @ZohranKMamdani

Contribute
contribute.nycvotes.org

2:51PM · 7/14/25 · 2.4M Views
1.3K 10K 164K 2.6K

PROJECT OVERVIEW

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



PROJECT OVERVIEW

THE BIG WHY?

Why This Matters for Everyone

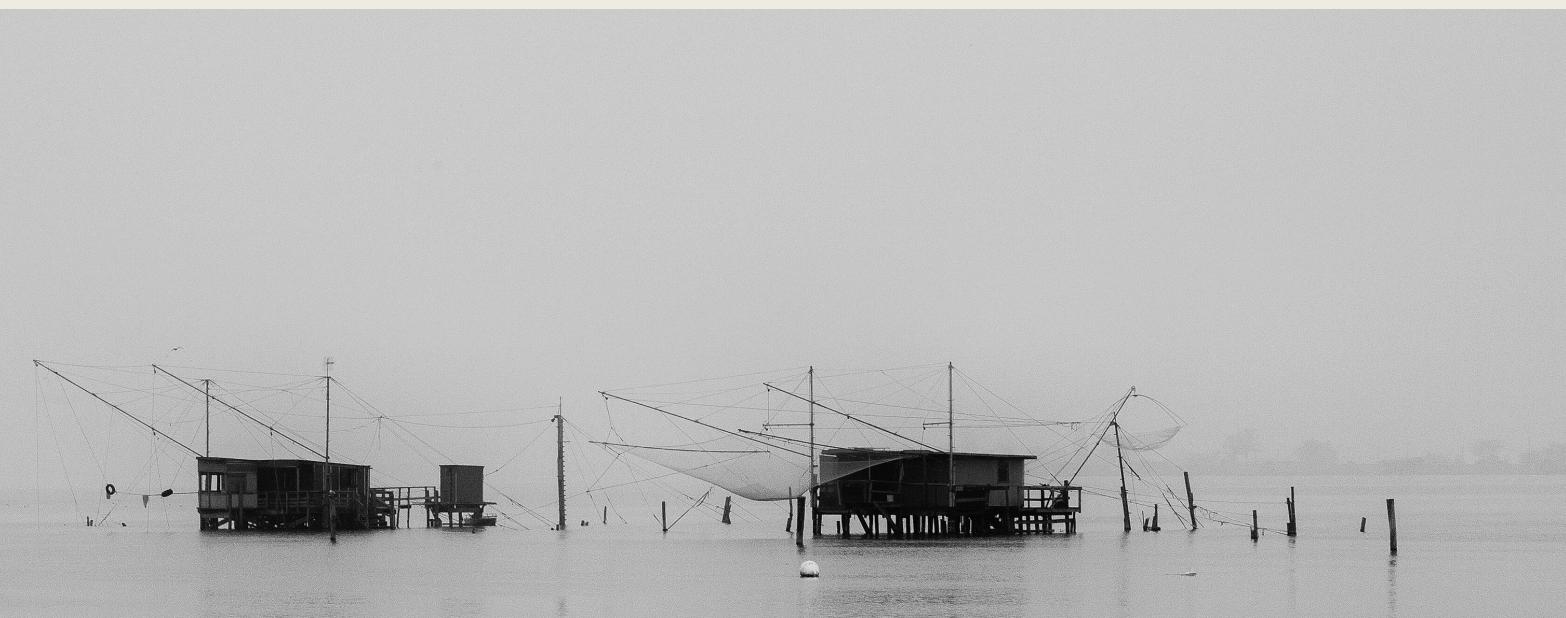
Understanding which data source is most predictive helps campaigns, journalists, and voters see past the hype and make smarter decisions. Whether polls or social sentiment turn out to be more accurate, our project sheds light on what tools will shape the future of U.S. presidential campaigns and how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



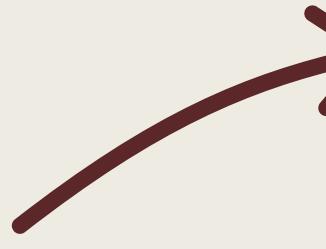
OUR DATA SOURCES & WHY THEY MATTER



Where:

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

SCRAPING AND BUILDING THE POLLING DATASET



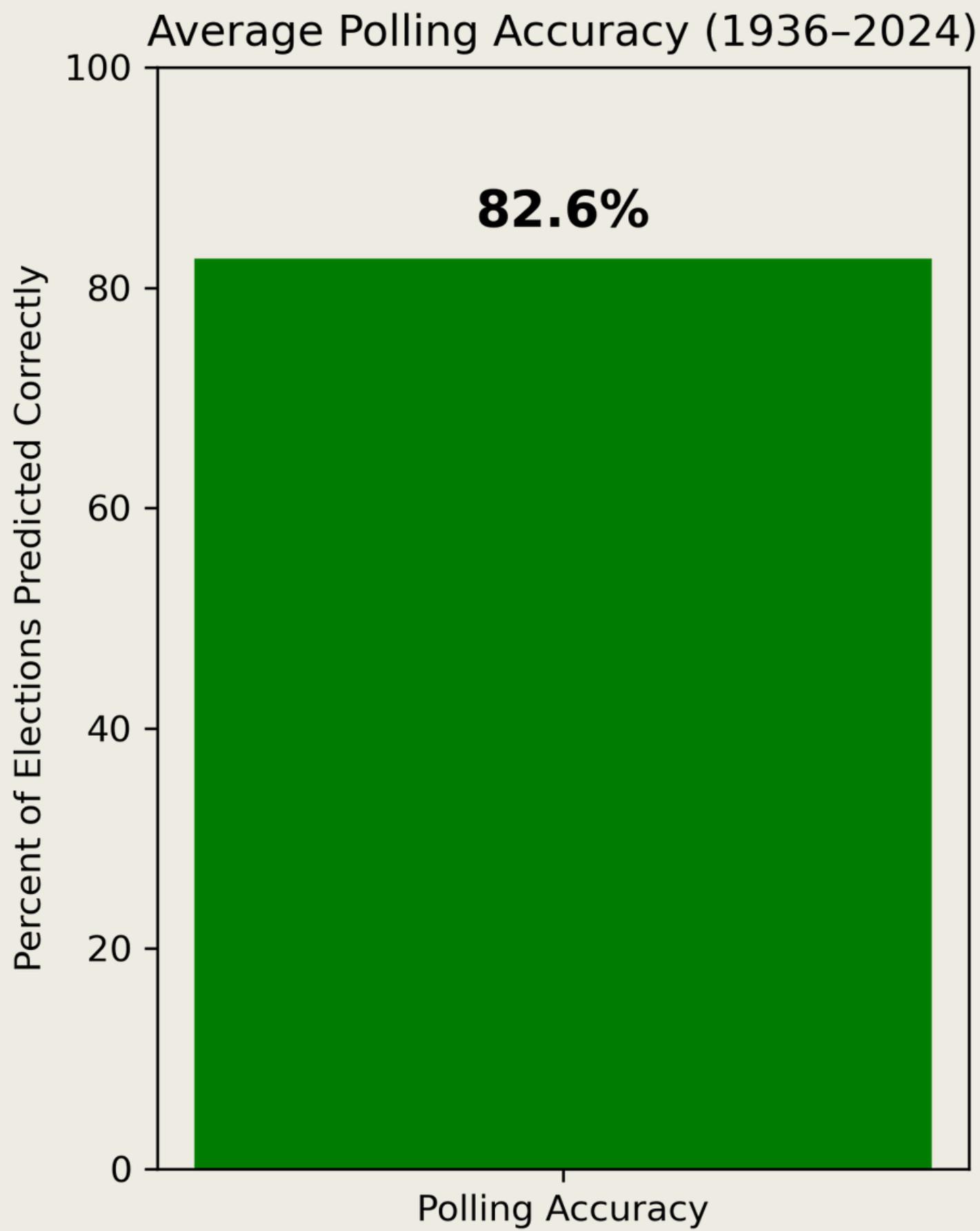
| | year | Poll_Leading_Margin | Poll_Leader | EC_election_winner | PC_election_winner | EC_Poll_Accurate | PC_Poll_Accurate |
|----|------|---------------------|-------------|--------------------|--------------------|------------------|------------------|
| 0 | 1936 | 5.89 | 0 | 0 | 0 | 1 | 1 |
| 1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

| | year | Poll_Leader | EC_election_winner |
|----|------|-------------|--------------------|
| 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 |

CAN POLLS STILL PREDICT THE WINNER?

BASELINE MODELING & ANALYSIS

At first glance...

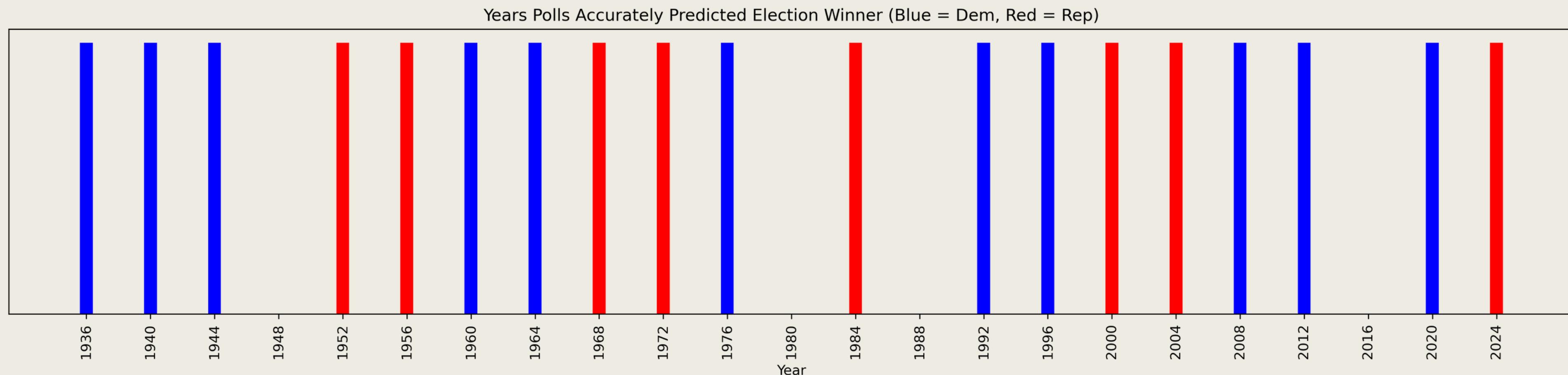


(09)

CAN POLLS STILL PREDICT THE WINNER?

BASELINE MODELING & ANALYSIS

When you start to look deeper ...

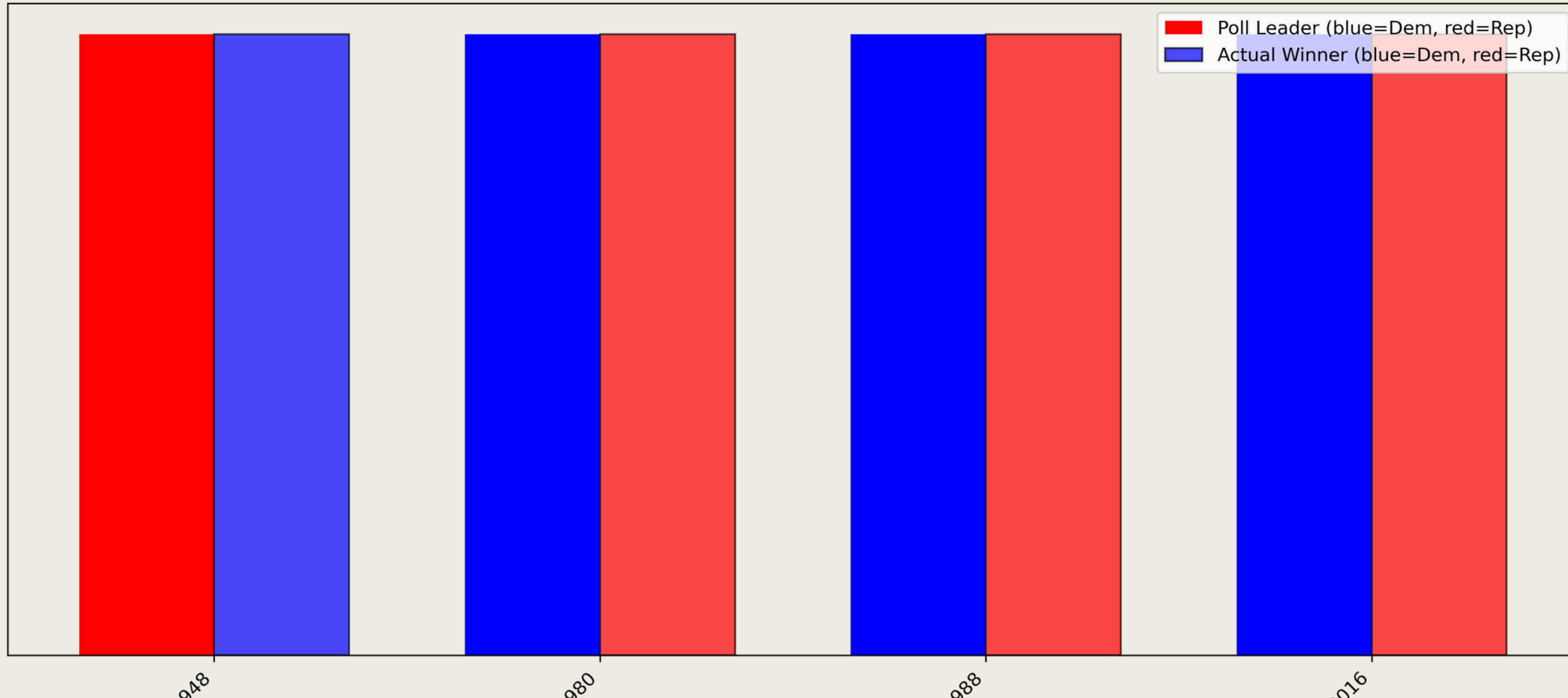


(10)

CAN POLLS STILL PREDICT THE WINNER?

What happened during the upset years?

Upset Years: Poll Prediction vs. Actual Winner



1948

1980

1988

2016

HISTORICAL
CONTEXT:
(CLICK THE
LINK)

[CLICK LINK: 1948](#)
[UPSET REASON](#)

[CLICK LINK: 1980](#)
[UPSET REASON](#)
[1980 UPSET REASON](#)

Upset Year

[CLICK LINK: 1988](#)
[UPSET REASON](#)

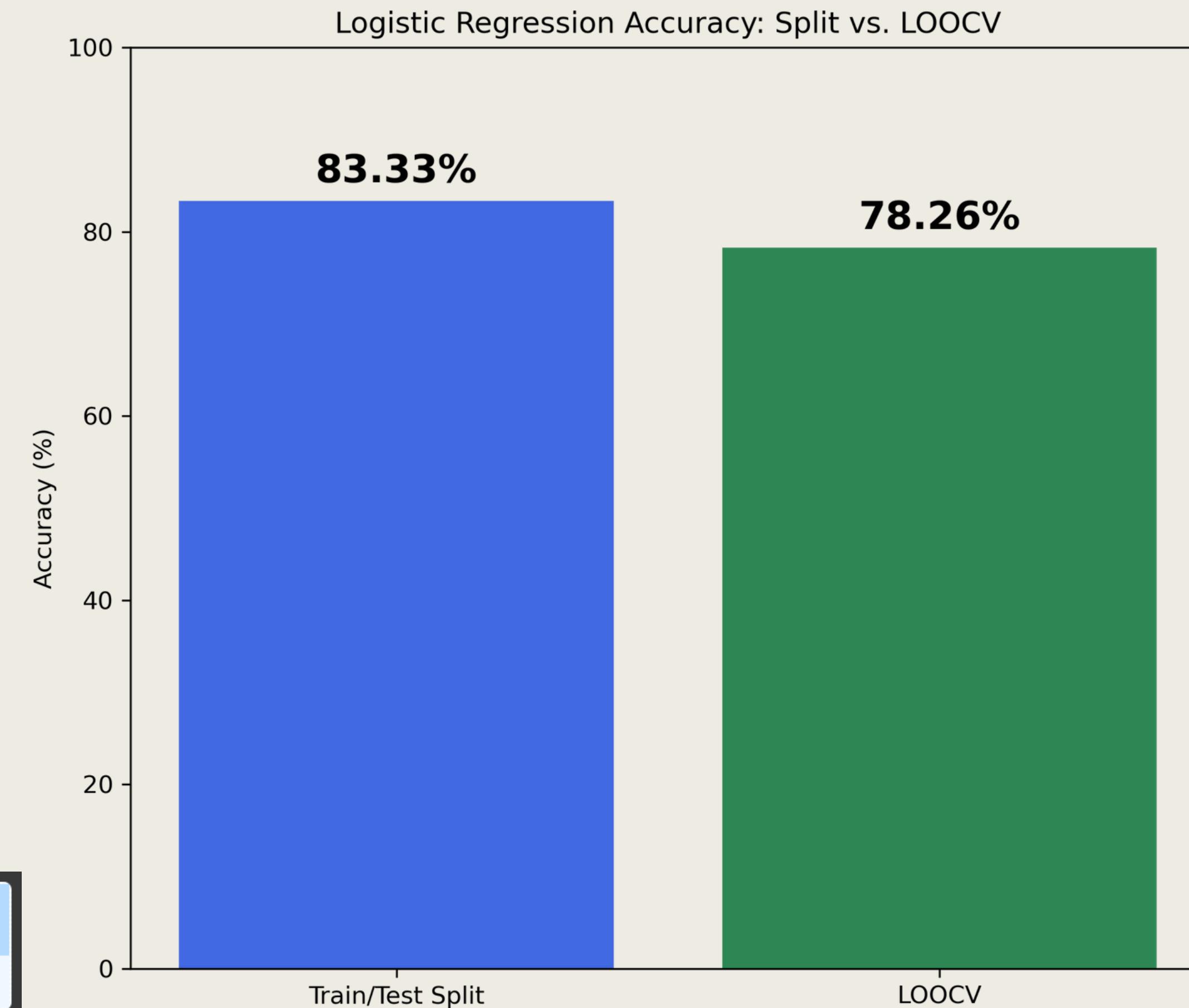
[CLICK LINK: 2016](#)
[UPSET REASON](#)
[2016 UPSET REASON](#)

(11)

CAN POLLS STILL PREDICT THE WINNER?

BASELINE MODELING & ANALYSIS

What did the Baseline Models Say?



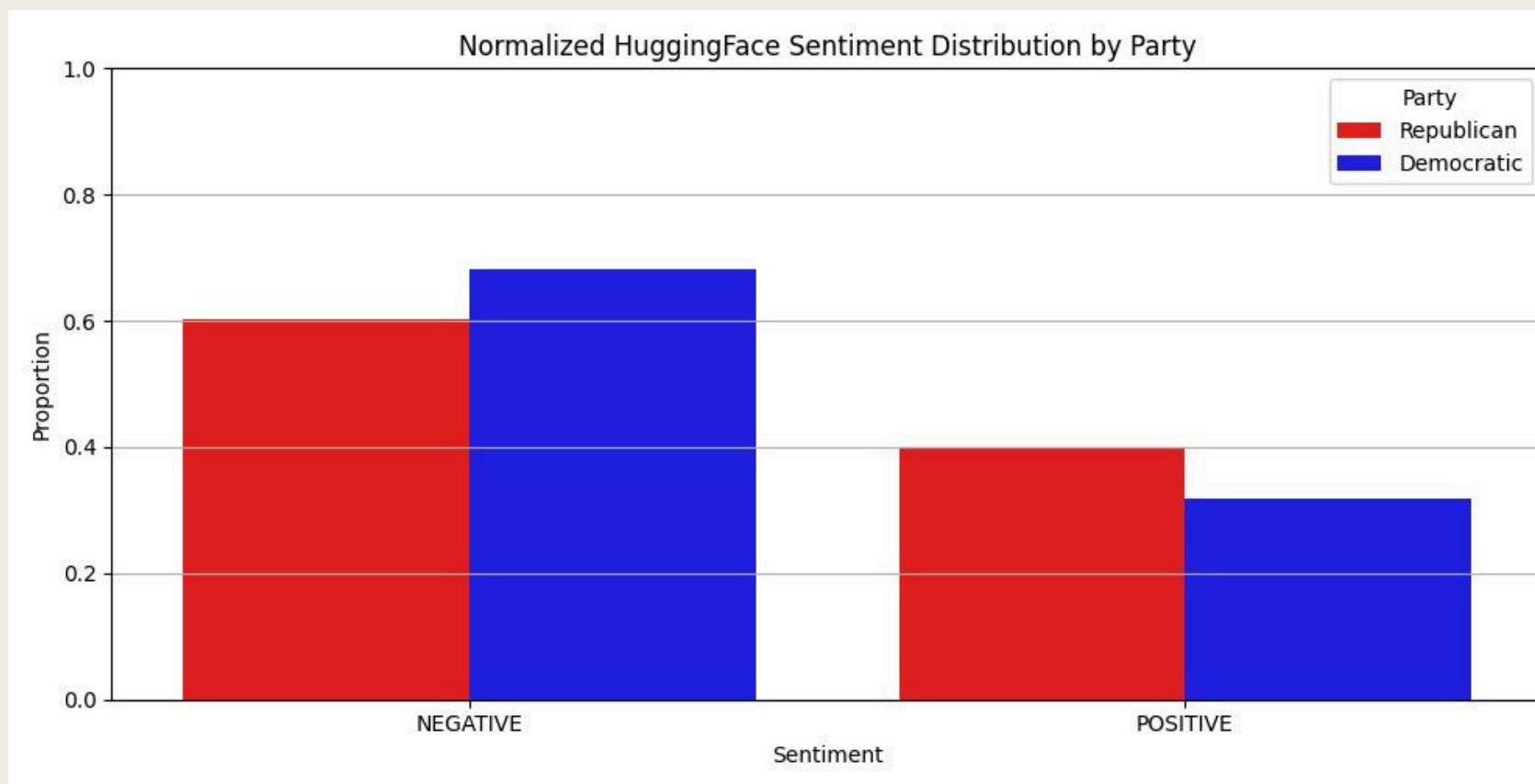
▼ LogisticRegression ⓘ ⓘ
LogisticRegression()

(12)

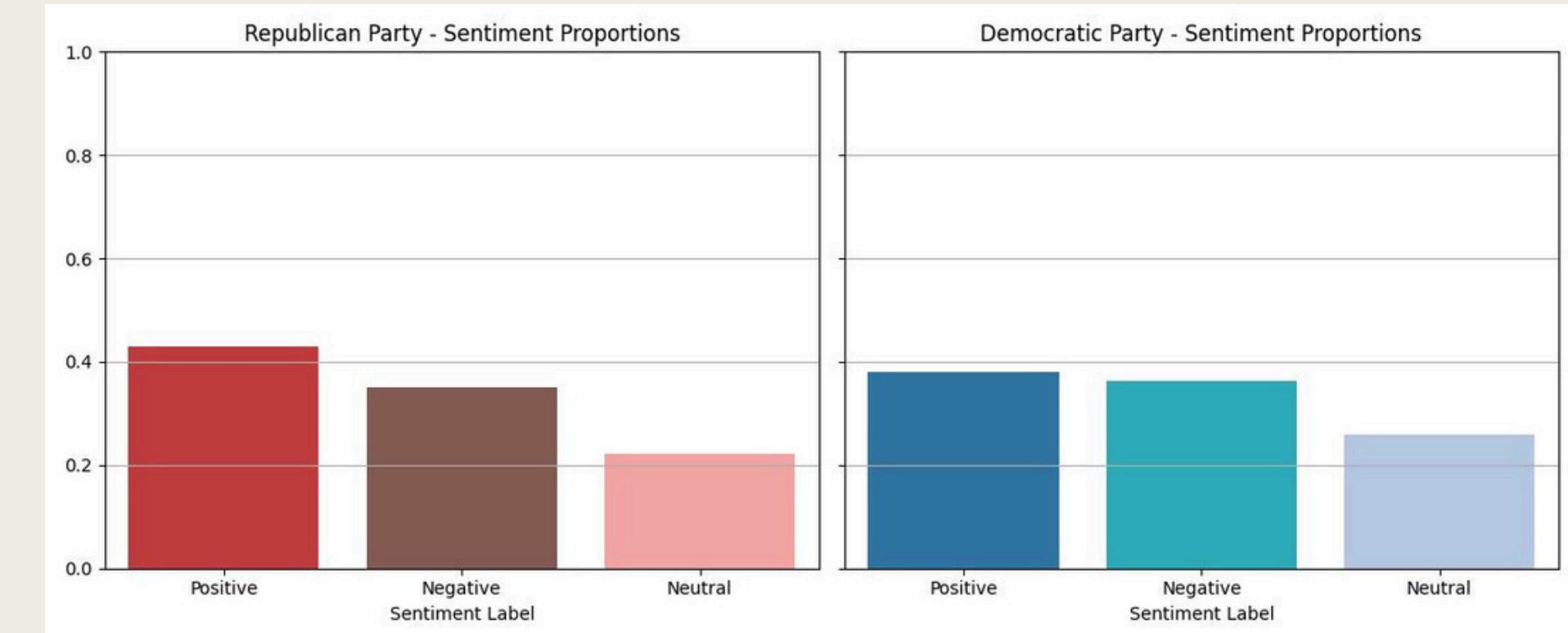
DATA COLLECTION: YOUTUBE COMMENTS

- search
- scrape
- refine
- repeat

MEASURING SENTIMENT: HOW WE TRACKED THE MOOD

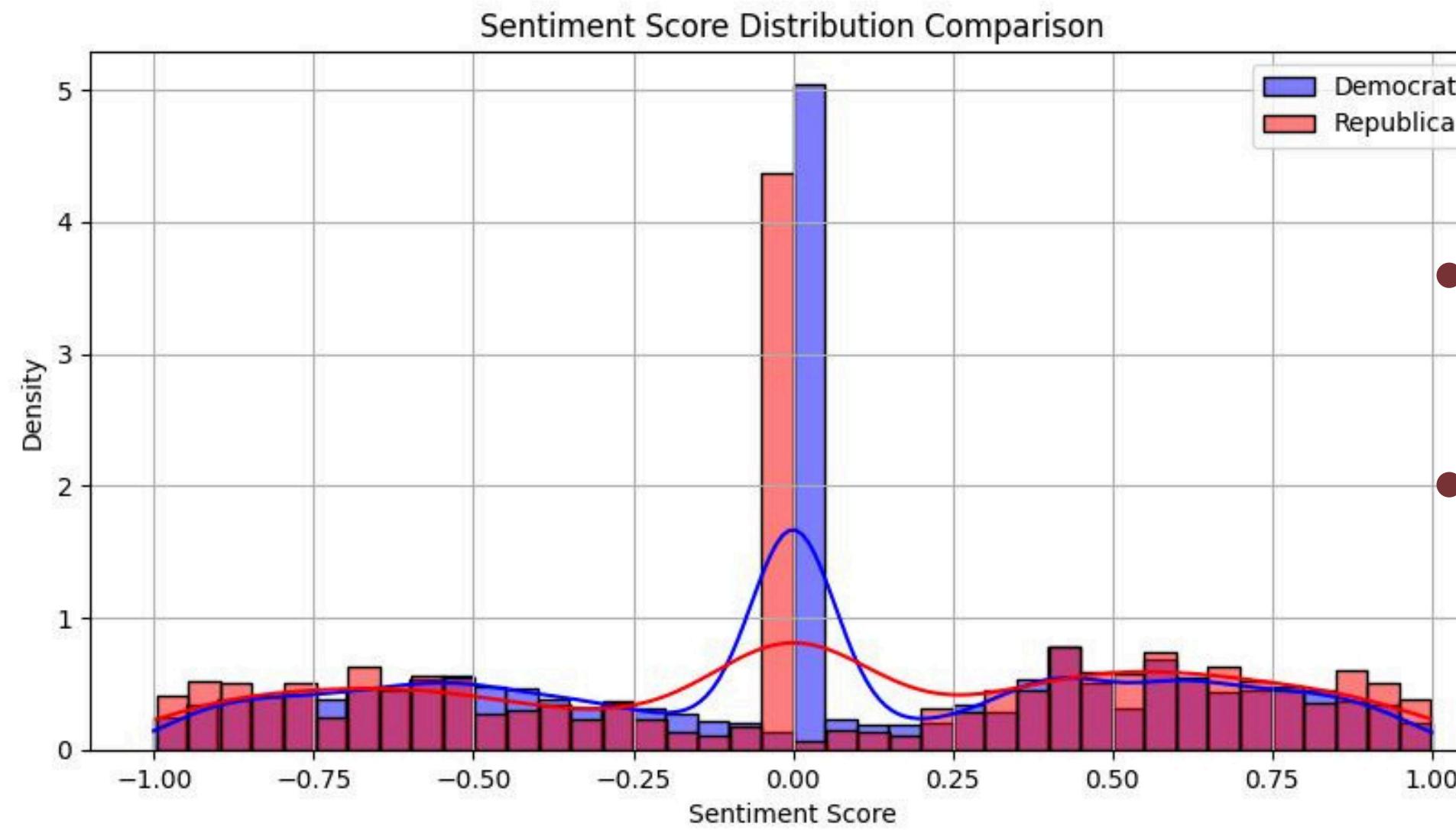


- Sentiment analysis models:
 - Vader NLP
 - Hugging Face



SENTIMENT MODEL RESULTS: DOES THE MOOD MATCH THE OUTCOME?

SENTIMENT ANALYSIS



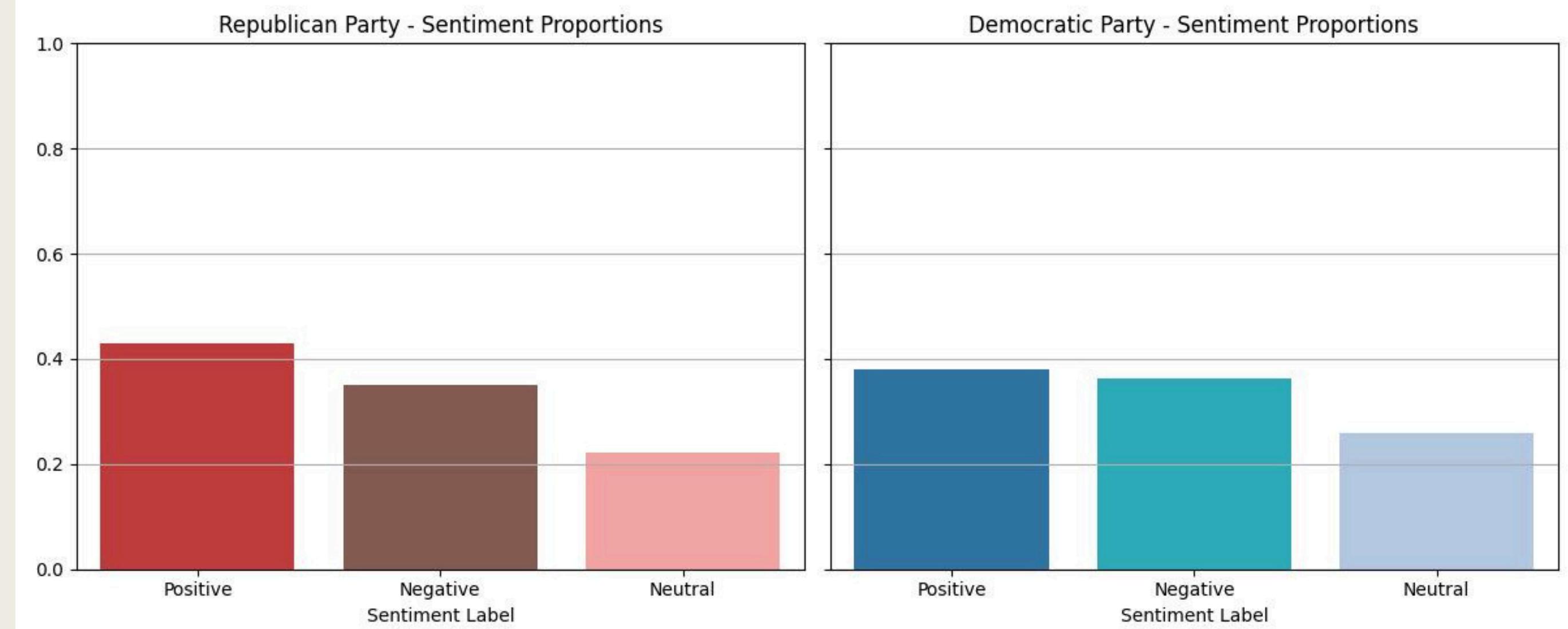
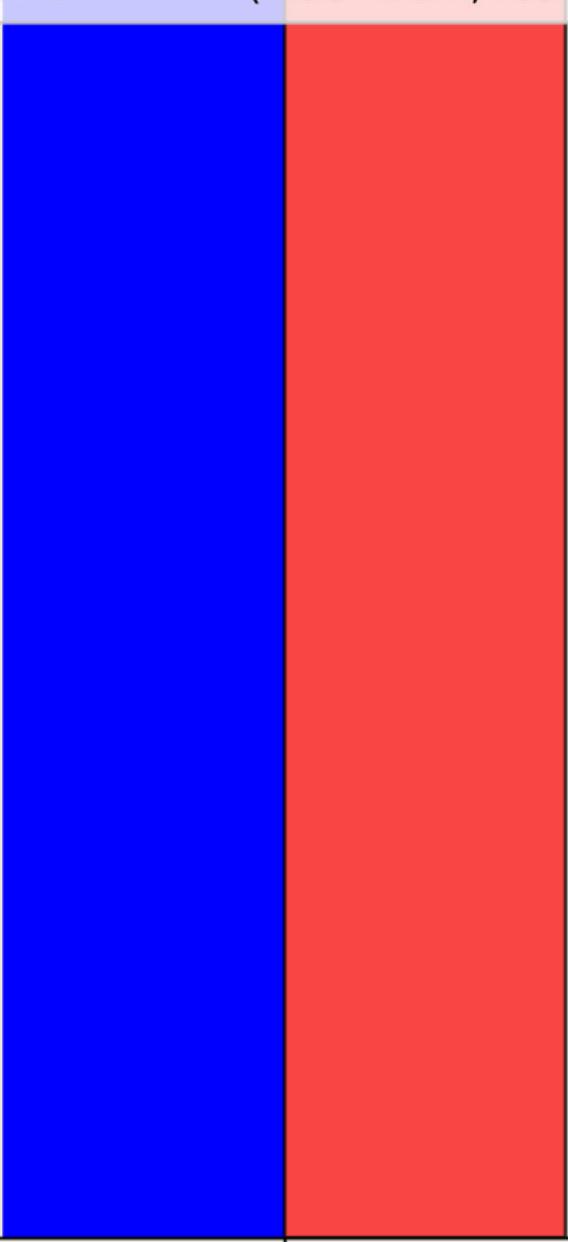
- How well did sentiment analysis predict election winners?
- Any interesting patterns, upsets, or trends from the sentiment data?

KEY TAKEAWAYS:

WHAT DID WE LEARN?

Upset Years: Poll Prediction vs. Actual Winner

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



LIMITATIONS & NEXT STEPS

- 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