

Data Privacy Impact in Political Advertising

Google Ads

Jose Lugo Gonzalez

Brief Plan

Week 1: Case Study Introduction

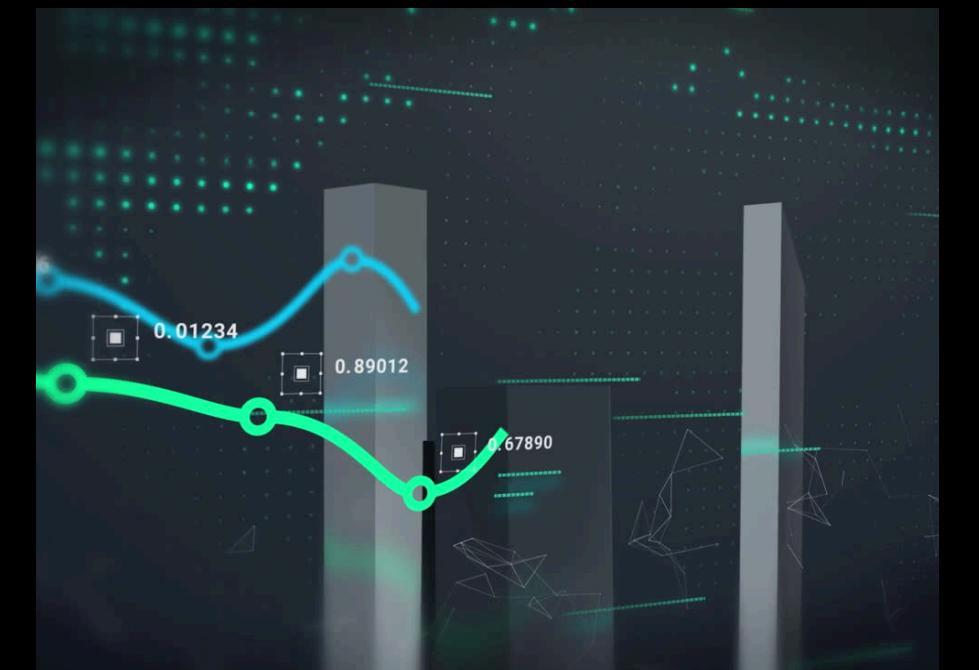
Week 2: Group Formation and Responsibilities

Week 3: Presentation of Case Study Analysis

Week 4-7: Development

Week 8: Final Deliverables

Case Study Intro



Case Study

My study highlights the need for stronger regulations to protect user privacy and ensure balanced information from Google Ads.

Ethical Issue

This study examines how Google Ads' targeting and personalization create filter bubbles, reinforcing biases, compromising data privacy, and influencing voter perceptions.

Goal

The goal is to explore the ethical impact of these practices and propose solutions for more transparent and fair political advertising.

Research

Since I will be working independently, I will conduct thorough research, gather relevant references, and source a suitable dataset online. Additionally, I will explore the source scripts Google uses for data collection to gain a deeper understanding of their methods.

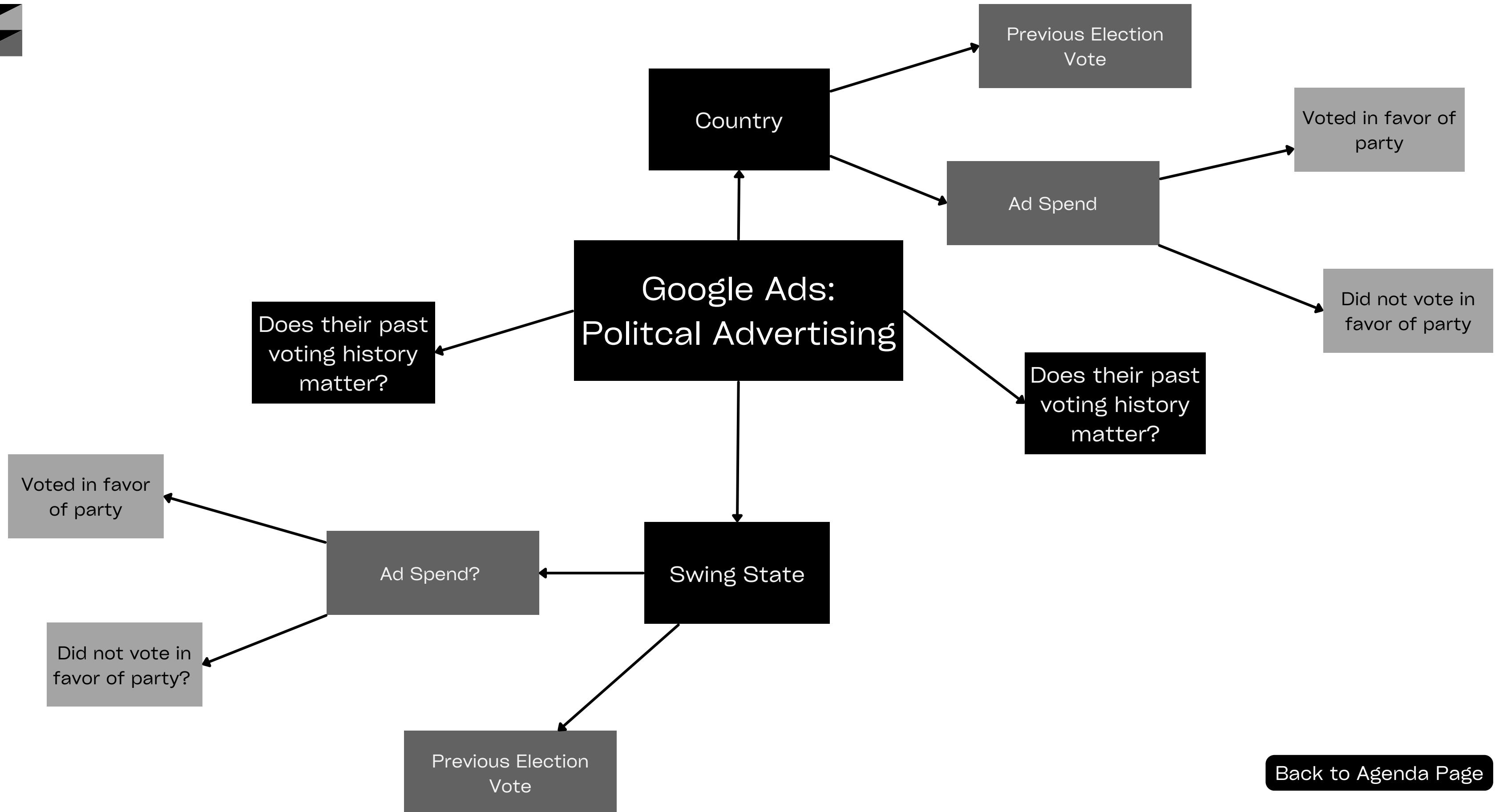
CRISP-DM & The Data Science Project Lifecycle

I will apply the CRISP-DM framework to guide this project through the entire Data Science Project Lifecycle. I'll also leverage the skills and knowledge gained from MSDS 600 to enhance my approach.

Case Study

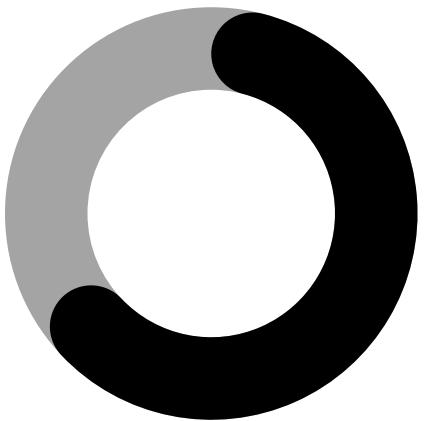
I will compile all my findings and data into a comprehensive final paper. If it aligns well with my personal portfolio, I may also upload it to GitHub to showcase my work.

My Responsibilities



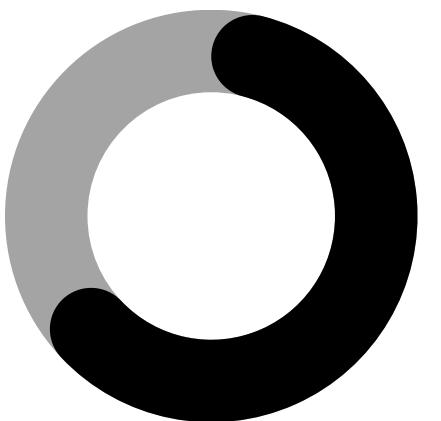
[Back to Agenda Page](#)

Key Performance Indicators



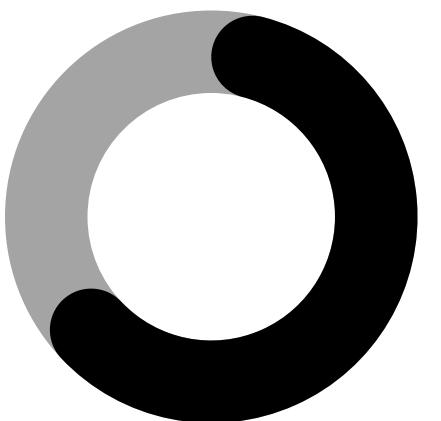
Voted for Ad They
Interacted with

Did they come across Ad and vote for
the party affiliated



Ad Spend in State / City

What was the ad spend in certain area



Cost/Winner

Does more spend often resonate in
winning the area?

Data Collection and EDA Analysis

[Back to Agenda Page](#)

Sources and Data Collection

- <https://www.opensecrets.org/online-ads/map/google>
- <https://www.kaggle.com/datasets/dankawaguchi/google-ads>
- <https://www.statista.com/statistics/910251/largest-political-advertisers-on-google/>
- <https://www.ppchero.com/unpacking-the-data-behind-trump-and-harris-presidential-election-ad-spending/>

Data Cleaning

Currently my plan is to break it down by state and see how many times and ad was fed here in here and what style of ad was most effective and if they won the state or not.

Predictive Analysis

After cleaning up my Data and doing a an extensive EDA I plan to work with Dr. Ghulam on guiding me towards how I should create a predictive analysis towards predicting the influence of political advertising.



Project Overview: Political Ad Spend & Voter Behavior

This project explores the real-world impact of political advertising on voter behavior across the United States, with a special focus on swing states. In an era of high-volume media influence, understanding whether increased ad exposure truly shifts voter decisions is both politically and socially relevant—especially as campaign spending continues to break records.

The analysis integrates multiple datasets:

- Political ad spend by candidate and region
- County-level election results
- Demographic and socioeconomic data
- Nielsen DMA (Designated Market Area) reach information
-

By combining these sources, I aim to uncover patterns between media exposure and election outcomes.

During the Exploratory Data Analysis (EDA) phase, I'll visualize trends using:

- Choropleth maps to highlight ad spend and voting margins across states
- Correlation heatmaps to explore relationships between variables like income, education, and turnout
- Scatter plots and bar charts to compare ad types and their localized effects
-

The core hypothesis is that higher political ad spend in a region correlates with an increased likelihood of voting for the advertised candidate. To evaluate this, I will build a predictive model—starting with logistic regression—to classify which candidate a region is likely to vote for based on ad spend and key demographic indicators.

Ultimately, this project seeks to test the power of exposure and influence, and whether more money really does mean more votes.

	A	B
1	Column Name	Description
2	state	U.S. state
3	county	County name
4	dma	Nielsen Designated Market Area (for ad reach)
5	population	Total population
6	median_income	Median household income
7	education_level	% with Bachelor's or higher
8	party_registration_dem	% registered Democrats
9	party_registration_rep	% registered Republicans
10	total_ad_spend	Total \$ spent on political ads in the area
11	candidate_a_spend	\$ spent by Candidate A
12	candidate_b_spend	\$ spent by Candidate B
13	winner	Who won the vote in this area
14	vote_margin	Vote difference in % between top 2 candidates
15	voter_turnout	% of eligible voters who voted
16	ad_type_tv, ad_type_radio, ad_type_digital	Breakdown of ad types (optional but cool!)

Data Collection

This is the columns I am currently working on completing I am fairly close to where I need the data to be. I anticipate once this is done it will take me 1-2 days to complete the EDA and start to get somewhere with where I am seeing.



First Party Data Collection Pt 1

Class Involvement on Ad Influence

Who should I vote for? 

Sponsored

 Jordan Lumen Campaign
www.jordanlumen.com

[Digital Detox Weekends | Connect To What Matters](#)

We're more connected than ever—and more distracted than ever. Vote for Jordan Lumen. Join us in restoring balance with nationwide Digital Detox Weekends.

Sponsored

 Jordan Lumen Campaign
www.jordanlumen.com

[Digital Detox Weekends | Connect To What Matters](#)

We're more connected than ever—and more distracted than ever. Vote for Jordan Lumen. Join us in restoring balance with nationwide Digital Detox Weekends.

Sponsored

 Jordan Lumen Campaign
www.jordanlumen.com

[Digital Detox Weekends | Connect To What Matters](#)

We're more connected than ever—and more distracted than ever. Vote for Jordan Lumen. Join us in restoring balance with nationwide Digital Detox Weekends.

JORDAN LUMEN

★ FOR PRESIDENT ★



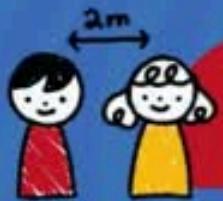
*Civic
Education
Reform*

01

VOTE WISELY AND SAFELY



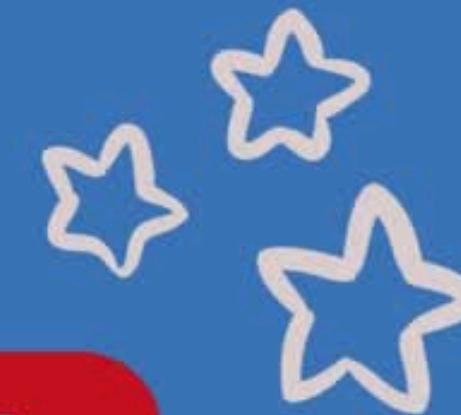
Read the instructions at the precinct entrance.



Follow the signages for smooth foot traffic.



Leave after voting to avoid crowding.



Send From: jlugogonzalez@regis.edu

To info@opensecrets.org

Cc

Bcc

Draft saved at 10:22 PM

2020 Election Data: Graduate Student Project

Hi OpenSecrets Team,

My name is Lugo Gonzalez, and I am a graduate student at Regis University studying Data Science. As part of a semester-long project, I am analyzing and developing a predictive model focused on advertising spending by presidential candidates.

I came across your site and was wondering if there is an option to purchase data on the total ad spend for each candidate in each state during the 2020 election. I would greatly appreciate any guidance or information you can provide on how to access this data.

Please feel free to respond at your earliest convenience.

Thank you for your time and assistance.

Best regards,
Lugo Gonzalez

(Ctrl) v

	A	B	C	D	E	F	G	H	I	J
1	state	country	population	median_income	education_level	party_registration_dem	party_registration_rep	total_ad_spend	candidate_a_spend	candidate_b_sp
2	Alabama	United States	5,024,279	56,802	40%	40%	50%			
3	Alaska	United States	733,391	75,455	30%	30%	45%			
4	Arizona	United States	7,151,502	65,330	33%	33%	35%			
5	Arkansas	United States	3,011,524	59,490	29%	35%	45%			
6	California	United States	39,538,223	86,846	36%	46%	24%			
7	Colorado	United States	5,773,714	83,379	39%	40%	33%			
8	Connecticut	United States	3,605,944	94,363	41.60%	48%	24%			
9	Delaware	United States	989,948	69,380	33.80%	47%	23%			
10	Florida	United States	21,477,737	71,739	32.30%	37%	35%			
11	Georgia	United States	10,711,908	62,660	34.10%	45%	38%			
12	Hawaii	United States	1,455,271	69,646	39%	55%	18%			
13	Idaho	United States	1,839,106	61,872	29.90%	20%	50%			
14	Illinois	United States	12,812,508	75,434	37.80%	47%	28%			
15	Indiana	United States	6,785,528	64,410	33.30%	37%	41%			
16	Iowa	United States	3,190,369	64,357	33%	35%	37%			
17	Kansas	United States	2,937,880	68,638	33.40%	30%	40%			
18	Kentucky	United States	4,505,836	58,240	30.80%	30%	60%			
19	Louisiana	United States	4,661,468	61,929	26.50%	30%	55%			
20	Maine	United States	1,362,359	68,164	37.60%	45%	38%			
21	Maryland	United States	6,177,224	79,240	39.70%	50%	23%			
22	Massachusetts	United States	7,029,917	95,331	42.20%	60%	31%			
23	Michigan	United States	10,077,331	64,188	33.40%	47%	39%			
24	Minnesota	United States	5,706,494	75,251	39.20%	48%	34%			
25	Mississippi	United States	2,961,279	52,022	24.20%	30%	60%			
26	Missouri	United States	6,154,913	65,382	32.50%	35%	45%			
27	Montana	United States	1,084,225	67,360	29.90%	30%	45%			
28	Nebraska	United States	1,981,504	72,892	32.50%	30%	40%			
29	Nevada	United States	3,138,259	69,829	27.30%	40%	35%			
30	New Hampshire	United States	1,377,529	83,396	37.50%	45%	25%			
31	New Jersey	United States	9,288,994	85,894	40.60%	45%	30%			
32	New Mexico	United States	2,117,522	57,991	28.50%	45%	30%			
33	New York	United States	20,201,249	87,153	39.80%	60%	26%			
34	North Carolina	United States	10,439,388	65,039	32.20%	40%	33%			
35	North Dakota	United States	779,094	72,019	31.60%	20%	40%			
36	Ohio	United States	11,799,448	64,592	32.70%	35%	45%			

Update On Data Collection

There has been many challenging things I have discovered upon gathering the data and attempting my first few EDA's. I am now working with Open Secrets a major company known for collecting data around elections.

Numbers for the most recent election are still being populated as we speak and those numbers are not final.

1

Getting this much data with limited time is very difficult and working with a team would have been useful.

2

3

Only 30 of the 50 states in my study have approved data such as this be open. There are 20 states that do not show ad spend info

Survey

I have created a survey to test the influence of ads in our classroom and then use the findings to see how we stand in comparison to the population as a whole. I will be updating this as everyone answers.

I please ask that you take the time to take the 3 minute survey!



Please Take
the time now
to complete
my survey

Questions Responses 1 Settings

President Election

B I U ↲ ✖

Class

President Vote

- Jordan Lumen
- Diego Ricardo
- Chase Conway Smith III
- Elise Johnson

Did seeing the Ads Last Week leave you remembering the candidates name?

- Yes
- No

What Ad Resonated with you the most?

- Google Search
- Voter Registration Video

+

✖

Tr

▲

▶

■

Plan

To explore the relationship between political ad spend and candidate success in the 2020 election, I plan to use a mix of regression and classification models. I'll begin with linear regression to establish a baseline understanding of how ad spend correlates with vote share. From there, I'll introduce more flexible models like Random Forest and XGBoost regressors, which are well-suited for capturing complex, non-linear relationships and can account for interactions between different variables such as spend timing, geography, and platform type.

If the focus shifts from predicting vote share to predicting win/loss outcomes (particularly on a state level), I will also experiment with classification models like logistic regression and tree-based classifiers. The training process will involve either a standard 80/20 train-test split or K-Fold cross-validation to ensure the models generalize well.

To evaluate the effectiveness of the regression models, I'll rely on metrics like the R-squared score (R^2) to determine how much variance in vote share can be explained by ad spend and related features. Additionally, I'll use RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) to measure the average prediction error, giving a sense of how close the model's predictions are to the actual outcomes.

Findings From Survey

01

**Jordan Lumen
Won The
election with
37.5% of the
total voters**

02

**The
effectiveness of
the ads helping
remember the
candidates
name was 50%
Effective**

03

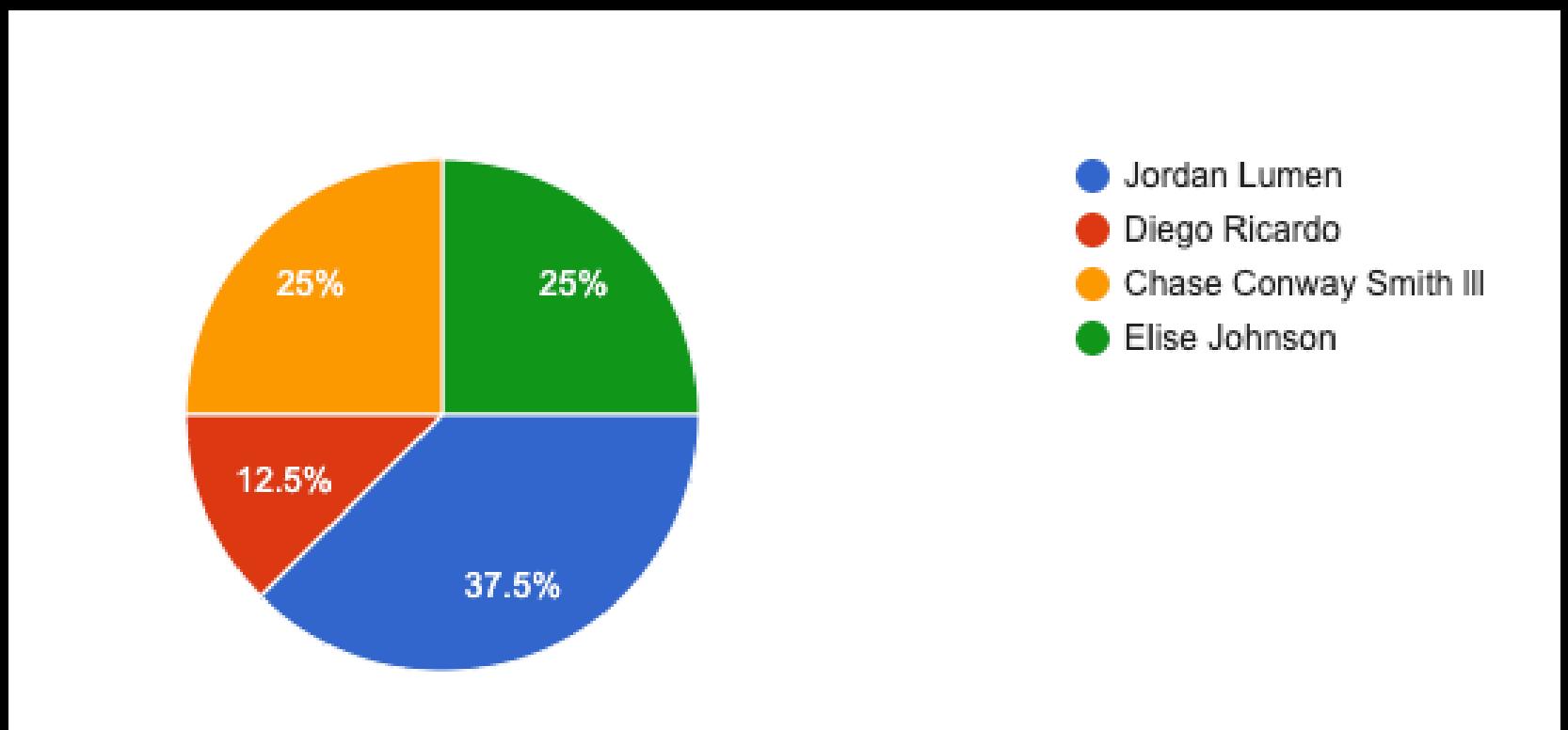
**72.5% of
respondents say
they voted for
candidite
because they
recognized the
name**

04

**The ads that
resonated the
most with our
class was flyer
and voter
registration
video**

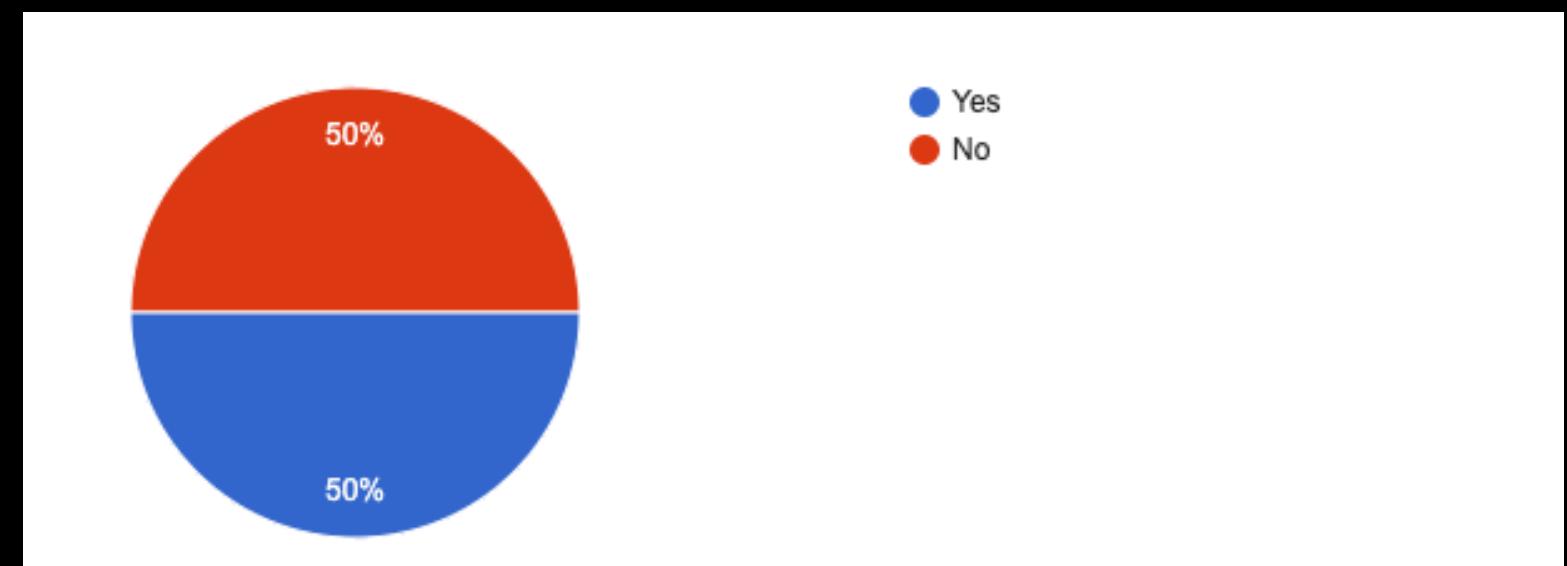
President Vote

What this chart will do for my overall project is showcase that from the 14 respondants in the class that Jordan Lumen won the election. And also the only candidite that exposed our class to ads. showcasing the insight that there is a 37.5% chance they were influenced by the ads.



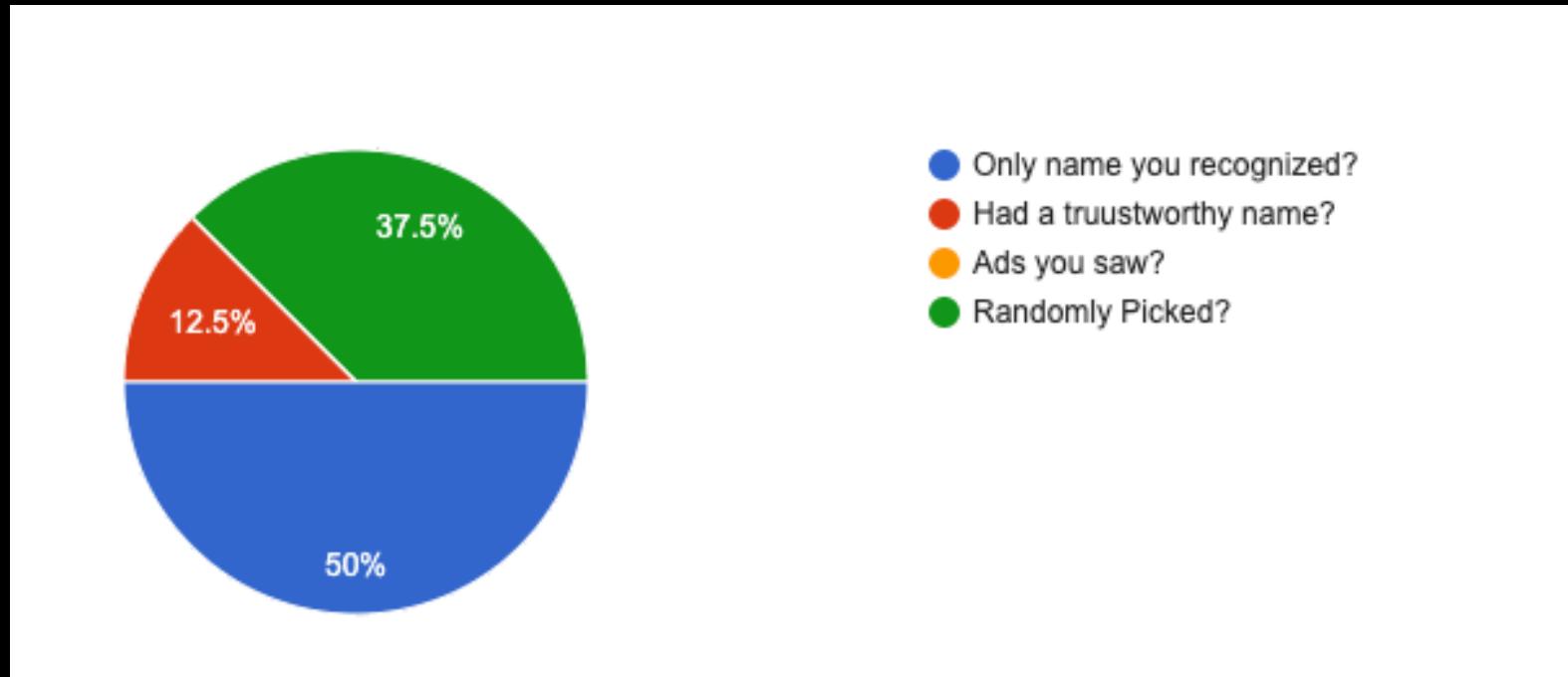
Did Ads help?

The class states that the ads had a 50/50 effect on the 14 participants, somehow a perfect split between yes and no. This helps showcase that there is an element of unpredictability here.



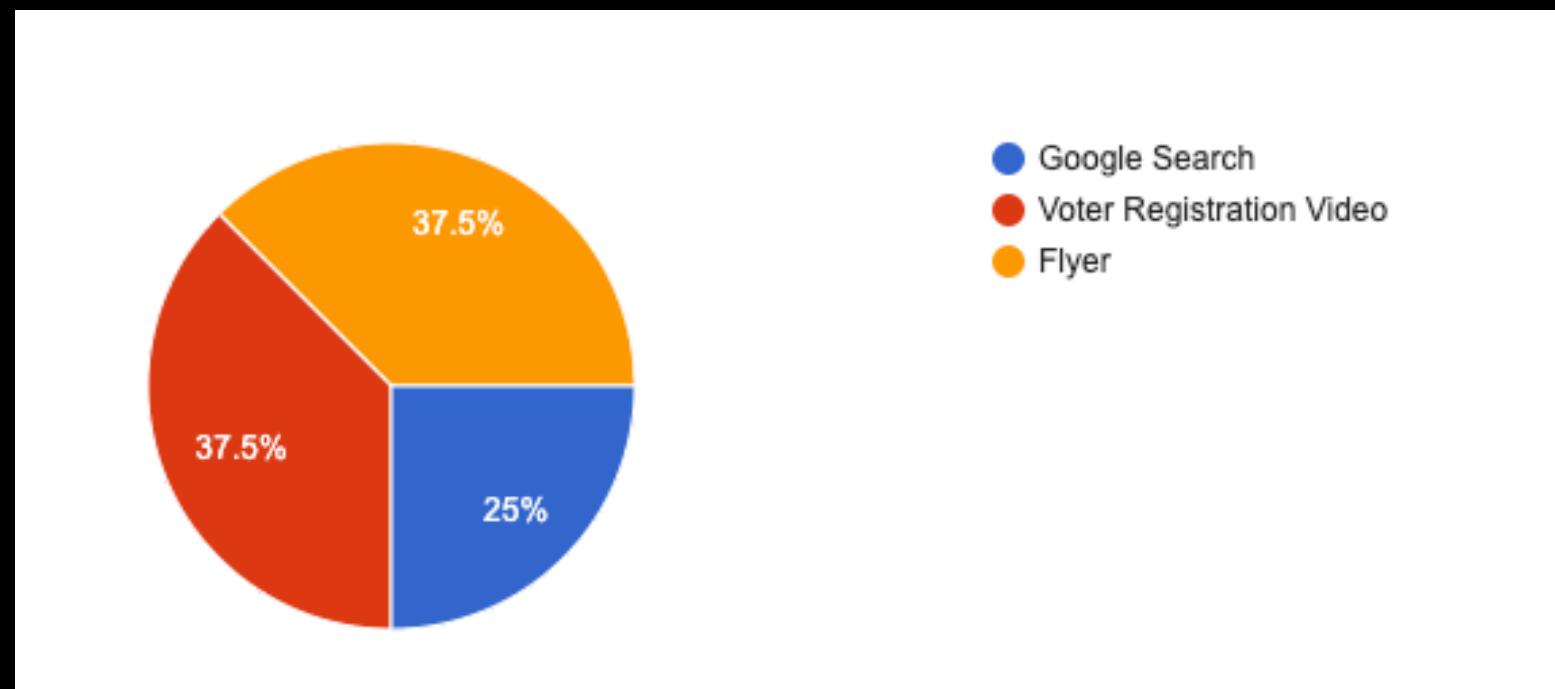
Why did you choose your candidate

I was shocked to see nobody picked the “Ads you saw?” option, however that one was thrown in there as an easy win. But really the answers I was looking for was blue and red. Upon creating this, I found a survey that did something similar and they said that people are hesitant to accredit ads for their choices in important and at time moral decisions. But the first two were put in there to be alternatives as to the effect of ads.



Ad Format

This one was just an interesting one in case down the line if I decide to develop this project I might consider the effectiveness of ad formats in my project particular for elections. As I can see that some of these may be portrayed as more reliable which could be a great additive to this project.



Update on EDA and Analysis

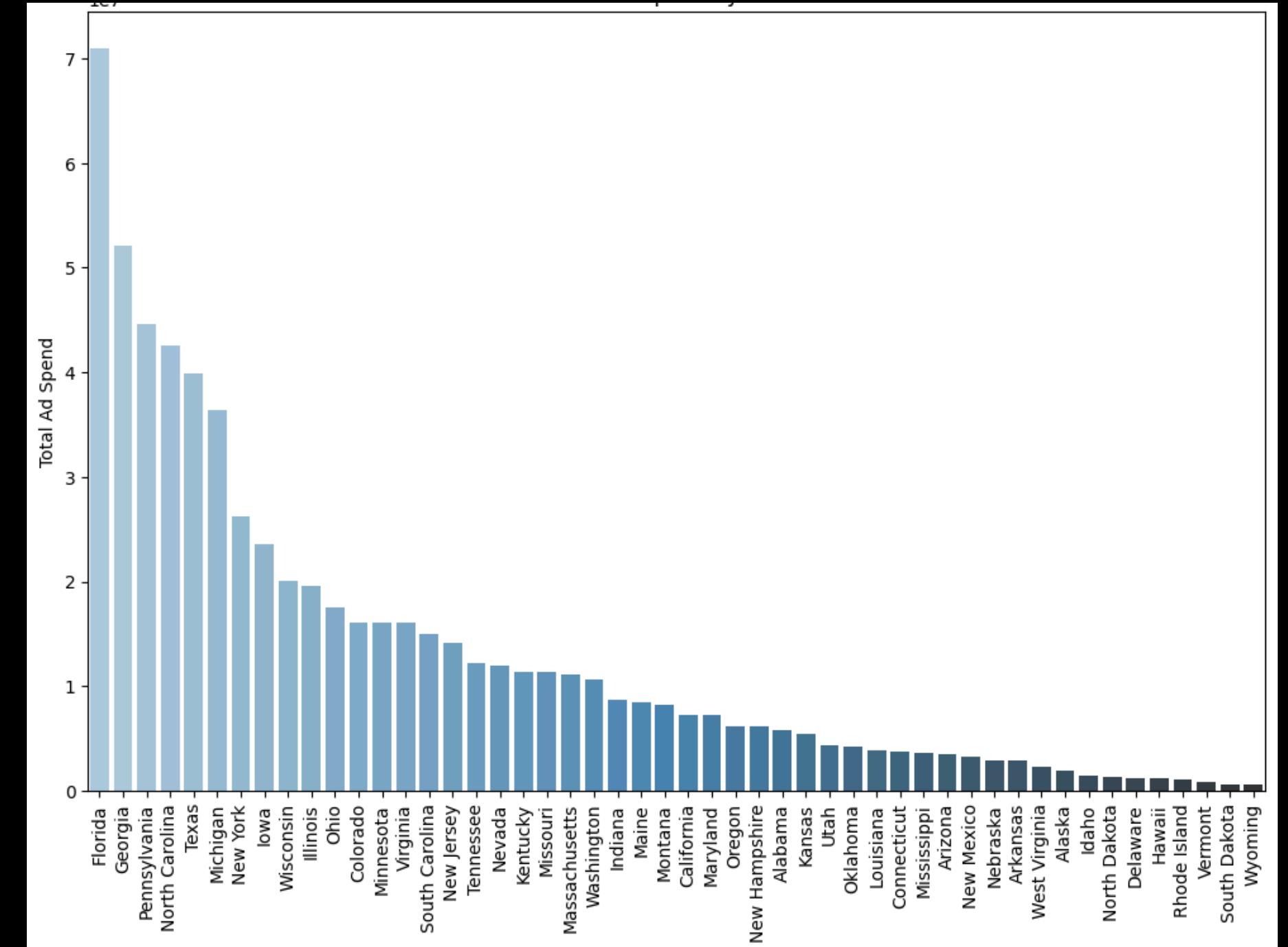
Made Great process this week and plan to be done with this project by Next Sunday! Finishing the data collection was by far the most difficult the people ghosting me at the data center so I decided to look up every state individually and cross reference which was not fun at all.

Update: Heard back from the data company, they had me sign up for a student account in order to get the data so will compare theirs to mine.

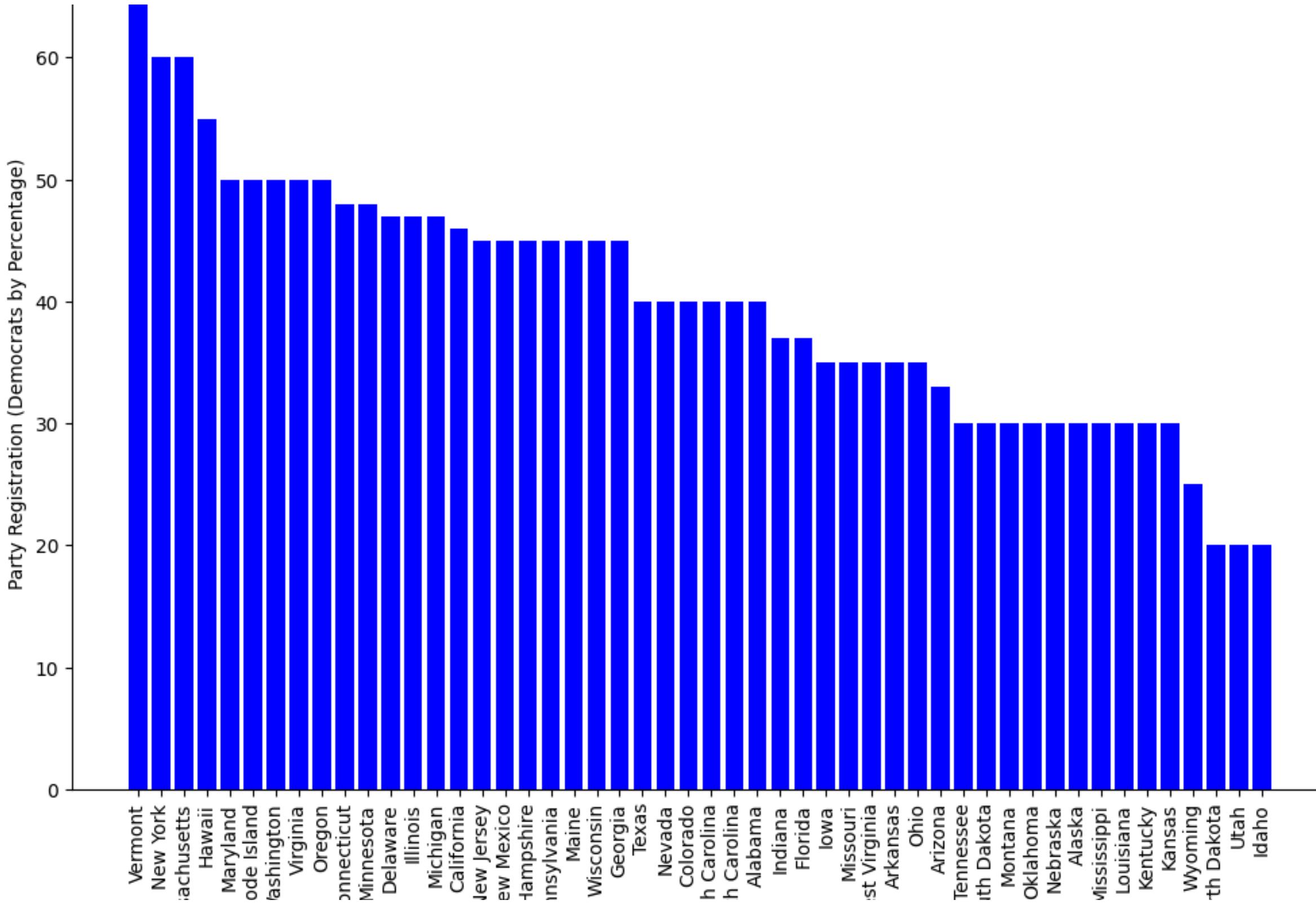
A	B	C	D	E	F	G	H	I	J	K	L
state	country	population	median_income	education_level	party_registration_dem	party_registration_rep	total_ad_spend	candidate_a_spend	candidate_b_spend	winner	voter_turnout
Alabama	United States	5,024,279	56,802	40%	40%	50%	5850000	3,100,000	2,750,000	Trump	61.46%
Alaska	United States	733,391	75,455	30%	30%	45%	1940000	1,200,000	740,000	Trump	68.23%
Arizona	United States	7,151,502	65,330	33%	33%	35%	3480000	2,000,000	1,480,000	Biden	66.63%
Arkansas	United States	3,011,524	59,490	29%	35%	45%	2930000	1,700,000	1,230,000	Trump	56.17%
California	United States	39,538,223	86,846	36%	46%	24%	7296000	4,000,000	3,296,000	Biden	68.83%
Colorado	United States	5,773,714	83,379	39%	40%	33%	16,093,300	9,000,000	7,093,300	Biden	76.69%
Connecticut	United States	3,805,944	94,363	41.60%	48%	24%	3,749,000	2,100,000	1,649,000	Biden	70.09%
Delaware	United States	989,948	69,380	33.80%	47%	23%	1,245,800	700,000	545,800	Biden	68.34%
Florida	United States	21,477,737	71,739	32.30%	37%	35%	70,934,900	40,000,000	30,934,900	Trump	68.87%
Georgia	United States	10,711,908	62,660	34.10%	45%	38%	52,114,000	30,000,000	22,114,000	Biden	70.79%
Hawaii	United States	1,455,271	69,646	39%	55%	18%	1,165,000	650,000	515,000	Biden	67.08%
Idaho	United States	1,839,106	61,872	29.90%	20%	50%	1,395,800	800,000	595,800	Trump	55.35%
Illinois	United States	12,812,508	75,434	37.80%	47%	28%	19,588,900	11,000,000	8,588,900	Biden	64.94%
Indiana	United States	6,785,528	64,410	33.30%	37%	41%	8,757,200	5,000,000	3,757,200	Trump	66.52%
Iowa	United States	3,190,369	64,357	33%	35%	37%	23,547,800	13,000,000	10,547,800	Trump	61.06%
Kansas	United States	2,937,880	68,638	33.40%	30%	40%	5,410,200	3,000,000	2,410,200	Trump	72.04%
Kentucky	United States	4,505,836	58,240	30.80%	30%	60%	11,383,900	6,000,000	5,383,900	Trump	65.10%
Louisiana	United States	4,661,468	61,929	26.50%	30%	55%	3,891,100	2,200,000	1,691,100	Trump	64.41%
Maine	United States	1,362,359	68,164	37.60%	45%	38%	8,490,200	4,800,000	3,690,200	Biden	64.14%
Maryland	United States	6,177,224	79,240	39.70%	50%	23%	7,255,000	4,100,000	3,155,000	Biden	74.87%
Massachusetts	United States	7,029,917	95,331	42.20%	60%	31%	11,087,100	6,200,000	4,887,100	Biden	69.45%
Michigan	United States	10,077,331	64,188	33.40%	47%	39%	36,431,500	20,000,000	16,431,500	Biden	71.31%
Minnesota	United States	5,706,494	75,251	39.20%	48%	34%	16,060,800	9,100,000	6,960,800	Biden	73.27%
Mississippi	United States	2,961,279	52,022	24.20%	30%	60%	3,586,200	2,000,000	1,586,200	Trump	79.21%
Missouri	United States	6,154,913	65,382	32.50%	35%	45%	11,304,900	6,500,000	4,804,900	Trump	60.63%
Montana	United States	1,084,225	67,360	29.90%	30%	45%	8,203,800	4,600,000	3,603,800	Trump	65.49%
Nebraska	United States	1,961,504	72,892	32.50%	30%	40%	2,943,600	1,700,000	1,243,600	Trump	71.34%
Nevada	United States	3,138,259	69,629	27.30%	40%	35%	12010400	7,000,000	5,010,400	Biden	69.06%
New Hampshire	United States	1,377,529	83,396	37.50%	45%	25%	6,113,700	3,500,000	2,613,700	Biden	64.25%
New Jersey	United States	9,288,994	85,894	40.60%	45%	30%	14,138,700	8,000,000	6,138,700	Biden	73.96%
New Mexico	United States	2,117,522	57,991	28.50%	45%	30%	3,245,200	1,900,000	1,345,200	Biden	72.04%
New York	United States	20,201,249	87,153	39.80%	60%	26%	26,248,800	15,000,000	11,248,800	Trump	60.49%
North Carolina	United States	10,439,388	65,039	32.20%	40%	33%	42,583,400	24,000,000	18,583,400	Trump	64.94%

► Total Ad Spend Distribution

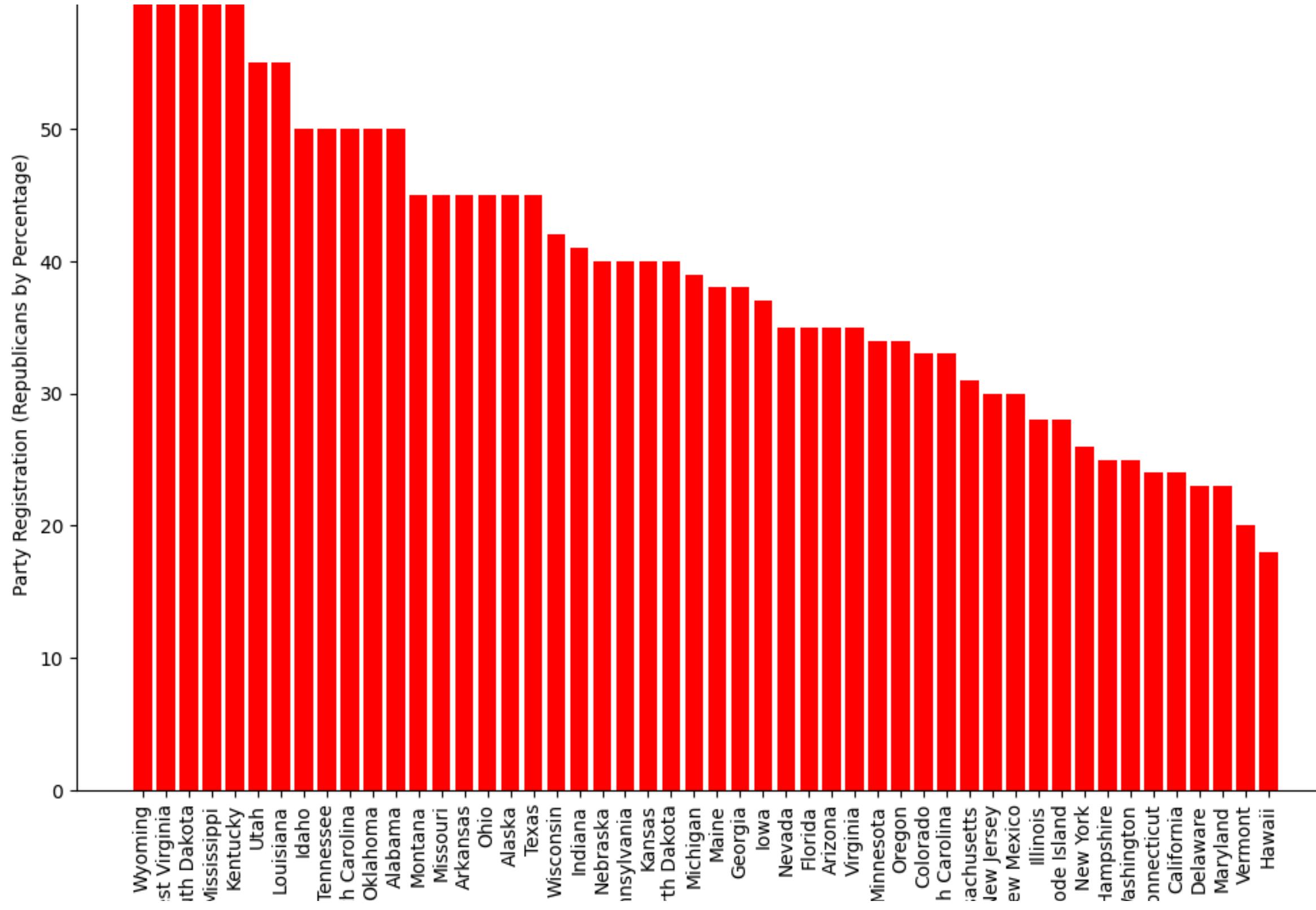
Bar Chart showcasing the most ad spend on an individual state.



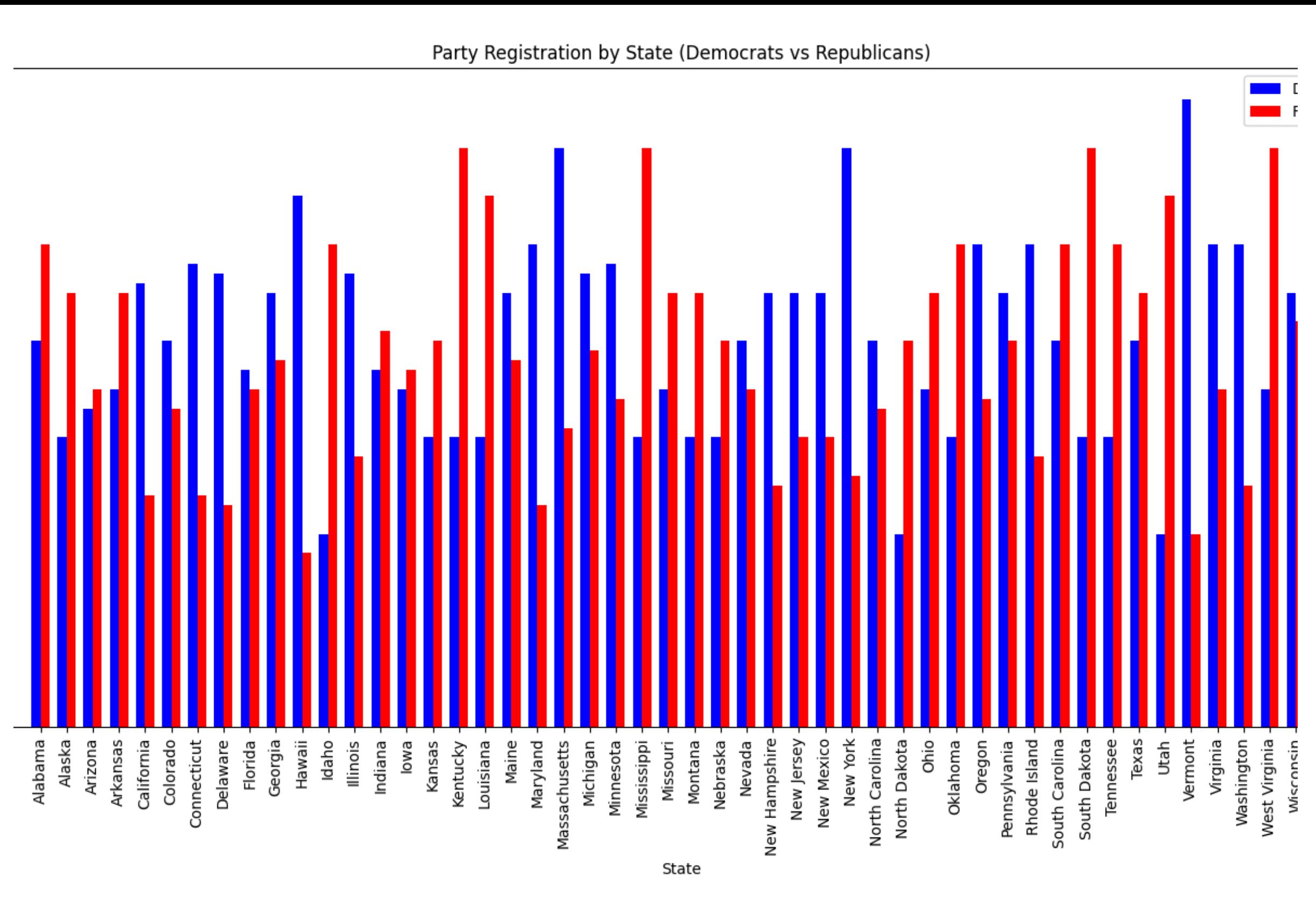
Democrats Registered



Republican Registration

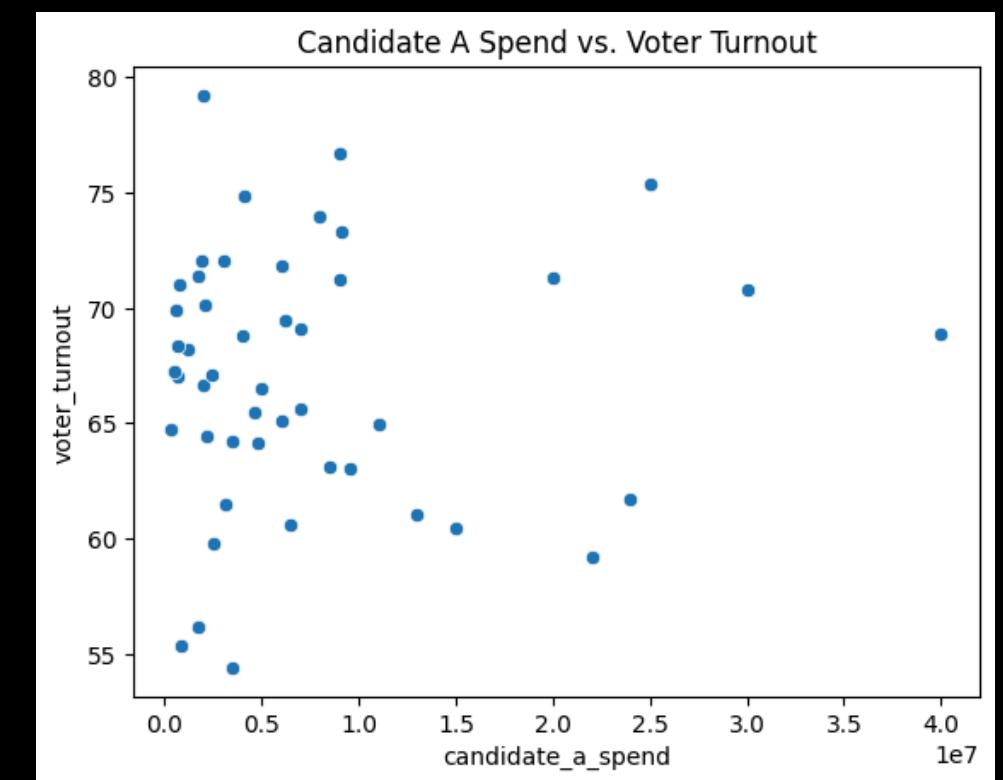
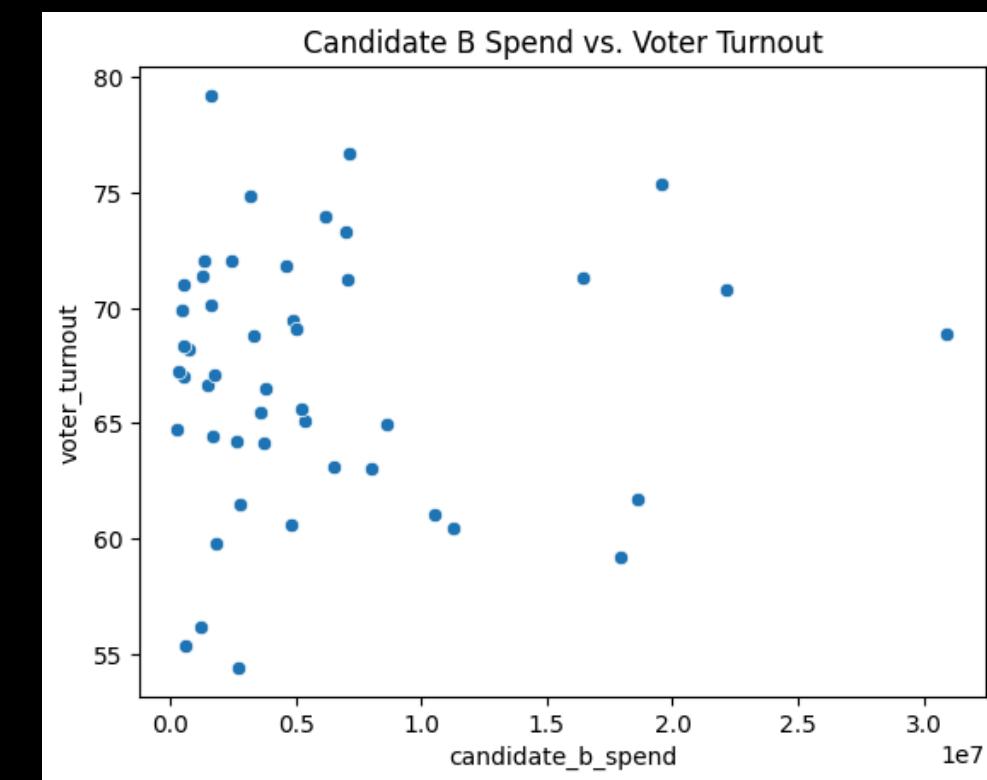


Comparison

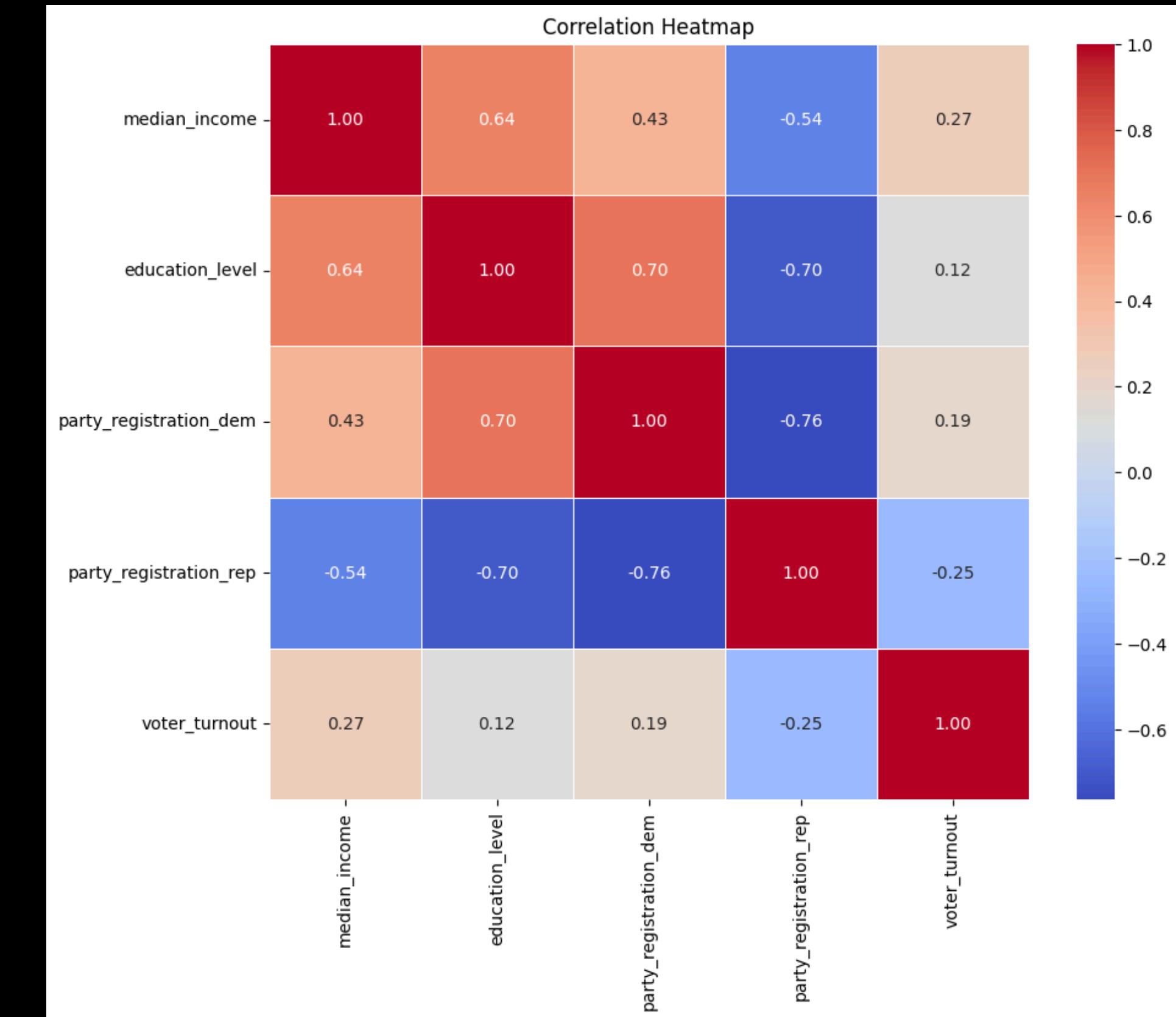


Scatter Plots

Scatter plots showed moderate correlation between ad spend and voter turnout, suggesting exposure may drive engagement —especially for Candidate A in certain counties

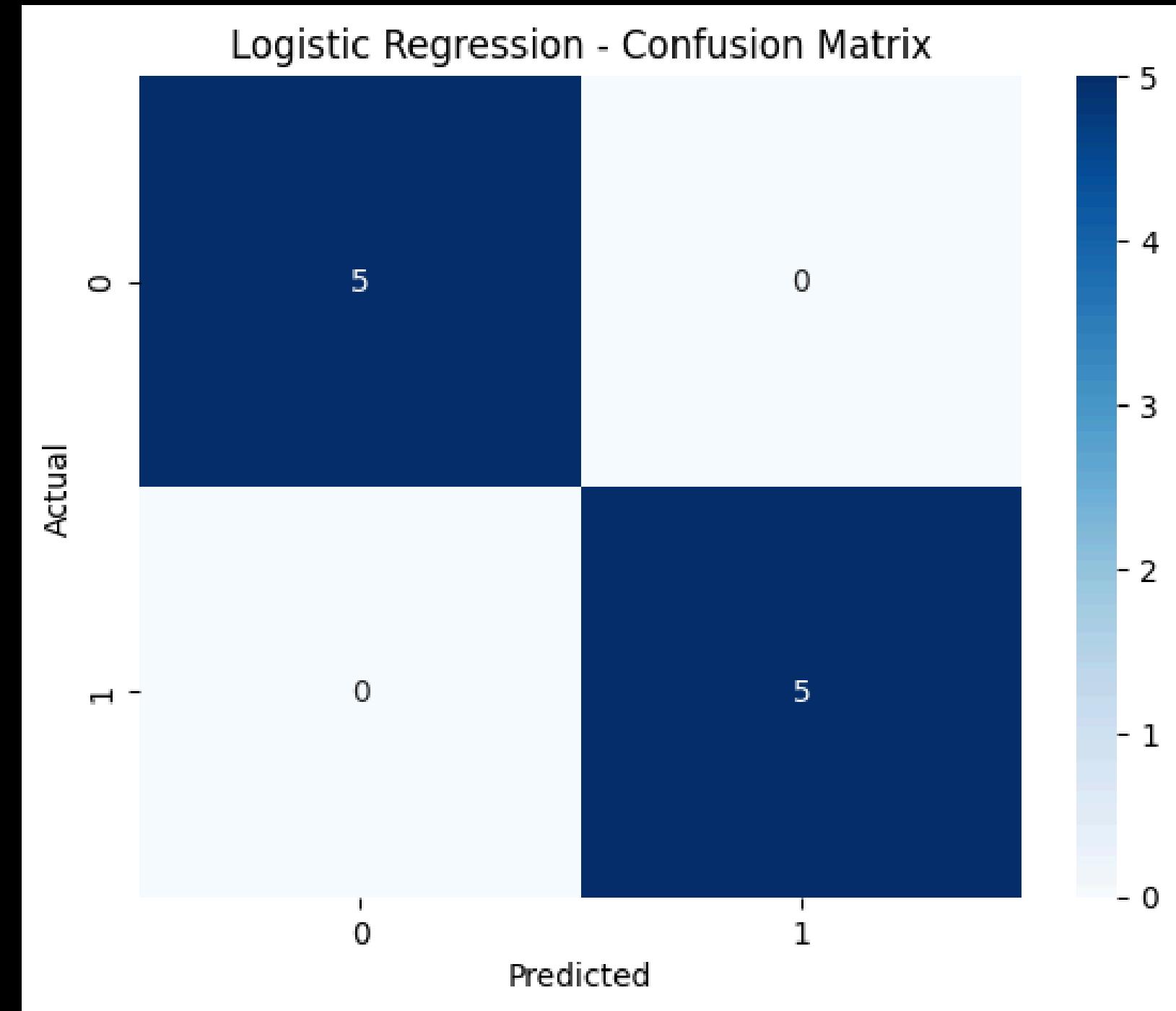


- Positive correlation between education level, income, and voter turnout.
- Party registration metrics also had directional influence, validating their inclusion in the predictive model.



Logistic Regression

--- Logistic Regression ---
Accuracy: 1.00 Precision: 1.00
Recall: 1.00 F1-Score: 1.00



Key Thoughts



Outcome

My logistic regression achieved a predictive accuracy of 100%, showing that ad spend and key demographic indicators can moderately predict election outcomes.

Thought 1

YMy model must be overfitting (memorizing the data instead of generalizing).

Thought 2

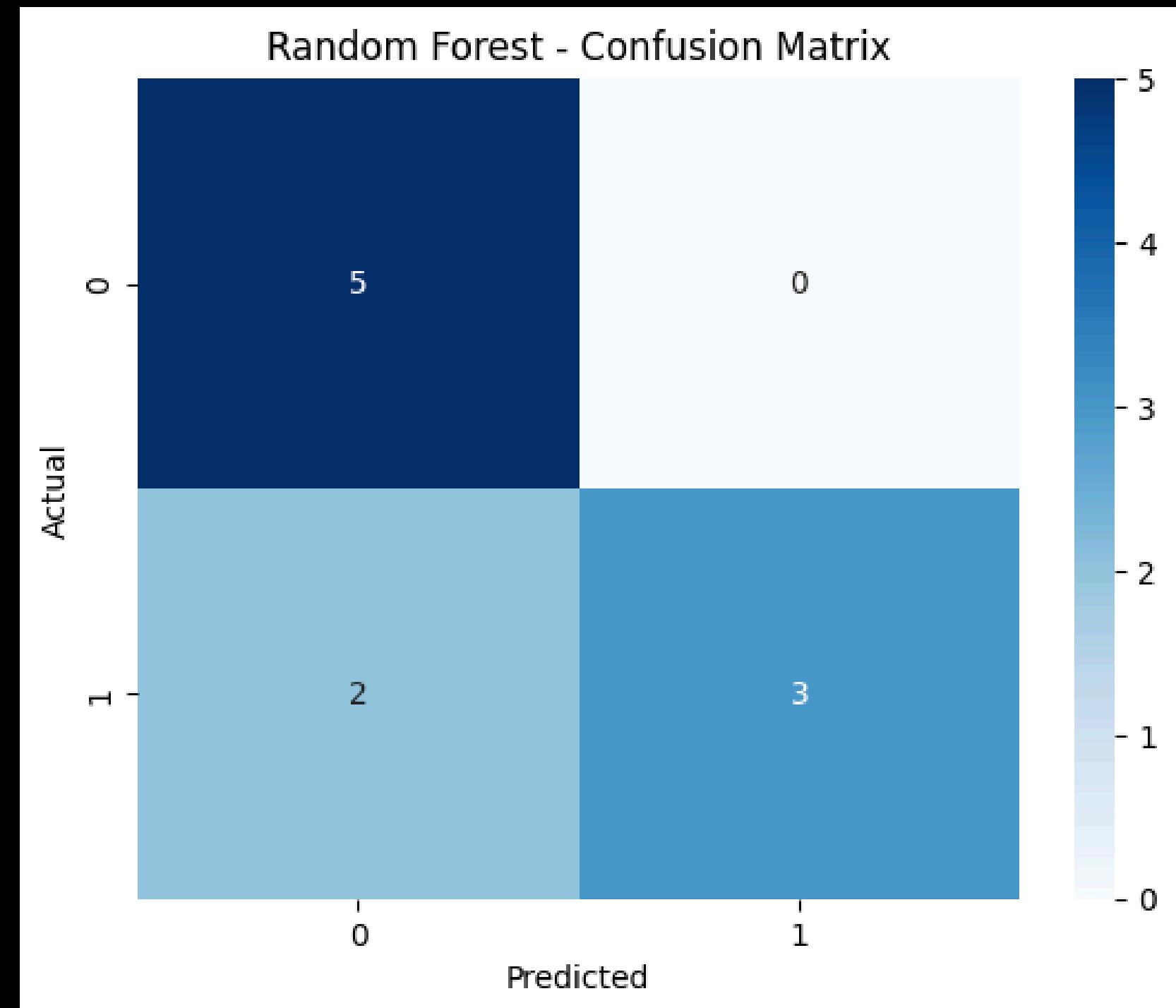
data might be too clean or too small. Which is likely since I ahed to result in one campaign year.

Thought 3

Perfect scores like this are extremely rare in real-world political modeling. So this can't be accurate.

Random Forest

--- Random Forest --- Accuracy:
0.80 Precision: 0.86 Recall: 0.80
F1-Score: 0.79



Key Thoughts



Outcome

My Random Forest is predicting correctly predicted the winner in 80% of the test cases.

Thought 1

Unlike the logistic regression model, the Random Forest model shows more realistic, less suspicious results.

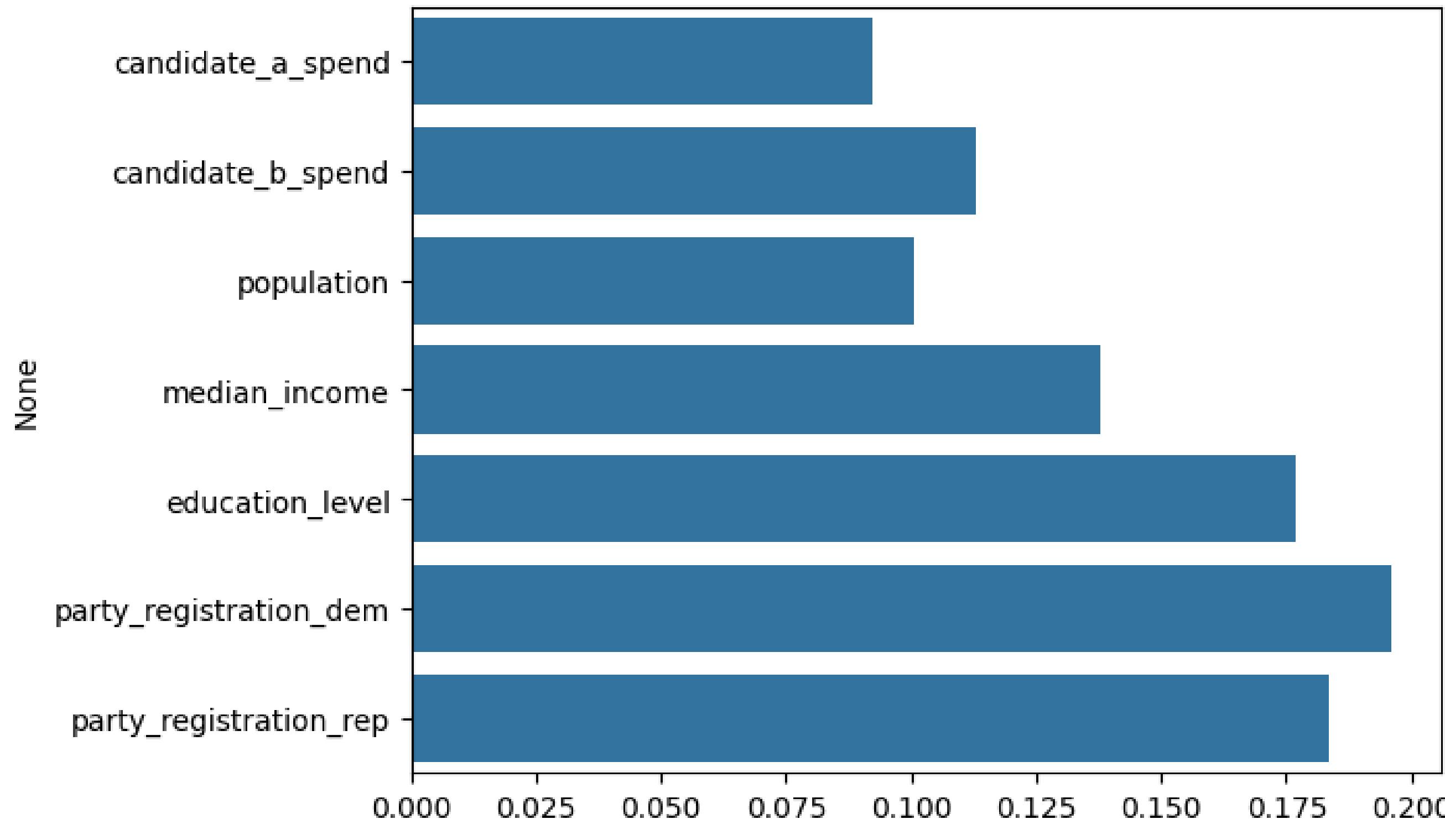
Thought 2

80% accuracy in a political outcome prediction task is respectable, especially with multiple variables.

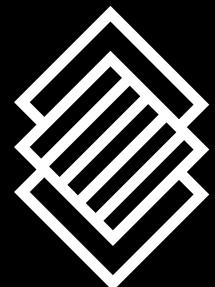
Thought 3

A slightly higher precision than recall suggests your model errs on the side of caution when predicting winners.

Random Forest - Feature Importance



The Ethical Cost of Prediction



While this project showed that political ad spend, combined with a bit of demographics, can accurately predict election outcomes, it raises critical ethical concerns.



If models can anticipate how a state will vote based on ad saturation, it reveals just how persuasive and invasive political advertising can be. The ability to micro-target individuals based on income, education, and party affiliation edges into manipulation, blurring the line between persuasion and control.



Voters must be aware that the ads they see are not neutral and in fact are designed to influence, often based on deeply personal data.



This reinforces the ethical responsibility to research the messages we receive, question their motives, and advocate for greater transparency in political media. In a world of predictive power, awareness is one of the last defenses against manipulation.