

Political Project Report

Introduction

The success of a democratic system hinges on active voter engagement. Robust democracies thrive when voter turnout is elevated. Policymakers are increasingly scrutinizing the role of different factors within the democratic framework. Although disparities in voter participation among groups are acknowledged, the precise underlying factors remain somewhat elusive. Conducting a thorough analysis of pertinent data is crucial in uncovering insights to address these inquiries.

Using the Apache Spark framework, I used SparkSession to manipulate the data provided. I cleaned the data to exclude missing data and assessed the variables in comparison to participation based on party, marital status, ethnicity, gun, and home ownership in the state of Pennsylvania. Since Pennsylvania is a famous swing state, the voter result would be quite important for both parties. This resulted in a dataset of over 5 million observations, removing a bulk of the original data. We also divided these based on party affiliation, to explore correlations between political issues and voter affiliation.

Exploratory Data Analysis

In the project, I aimed to evaluate and answer the following questions given the results:

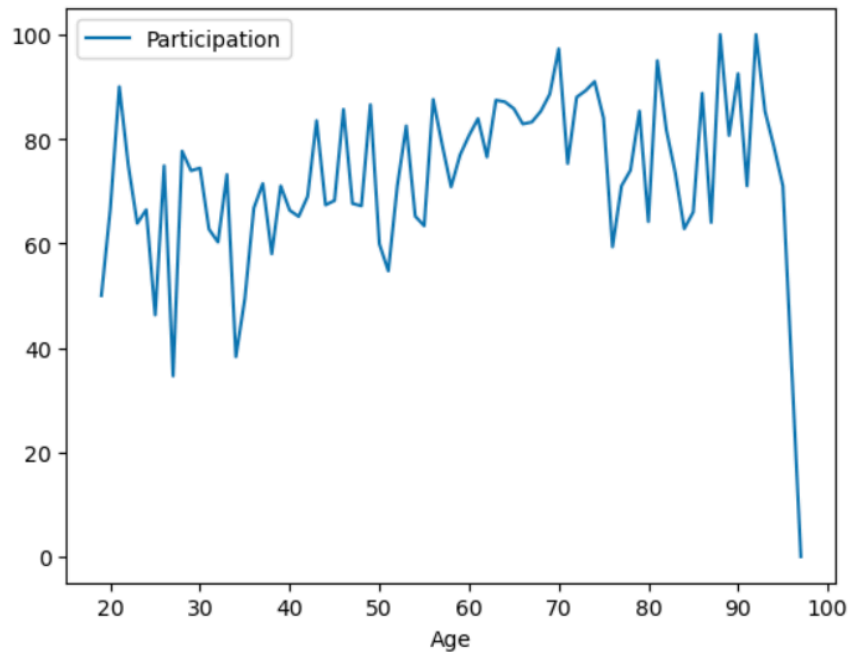
- Is there a disparity in voter turnout among households based on income level?
How does house ownership and/or housing cost impact voting turnout? Do counties with higher homeownership have higher voter turnout?
 - Yes. Higher-income households, on average, have significantly higher voter turnout than lower-income households.
- Is there a connection between the quality of school districts and voter turnout?
 - Not directly; we can infer from the data that higher-income people live in higher-income areas, and since the school districts in those areas are funded by property taxes, they are of higher quality. Thus, we can conclude that higher-quality school districts have higher voter turnout.
- What are the leading causes for the disparity in turnout for voters of a different ethnicity?
 - The lowest voting ethnicity in our findings were Hispanic or Portuguese descent individuals. This is because of one of two reasons; they are undocumented and unable to vote, or because of the language barrier that suppresses their ability to vote.

- Are the gaps in opinions between the races somehow related to the level of participation in the electoral system?
 - Republican African-American and Republican White individuals are the groups with the lowest and highest voter turnout, respectively. African Americans are historically Democratic and liberal, so the Republican counterpart of this demographic is incredibly stigmatized when it comes to voting for this party. Conversely, White Americans are historically conservative, so it makes sense that they, with their nationalist pride and ability to proficiently read and write in English, have the highest voter turnout.

Additionally, here are the general findings of the entire dataset:

- Republicans had higher voter participation than Democrats
- Married households had higher voter turnout than non-traditional households
- Single-person households voted significantly less than any other household above 2. It can be inferred that the biggest driving factor in this case is that people in households with over 2 people care more about the people around them and will be motivated to vote. This is the case for parents voting for issues based on their children, or there being a higher probability of each individual in a voter-eligible household being willing to vote.
- There is a trend of increasing voter turnout with age. As individuals grow older, they tend to exhibit greater participation in the electoral process. This aligns with

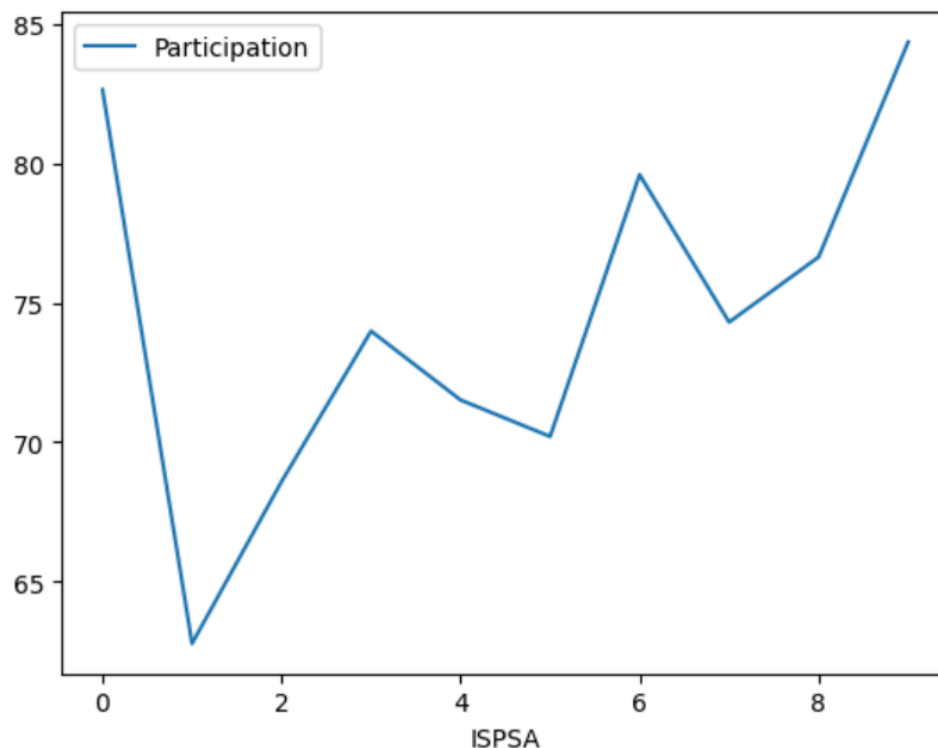
the common perception that older individuals are more inclined to engage in political activities compared to their younger counterparts.



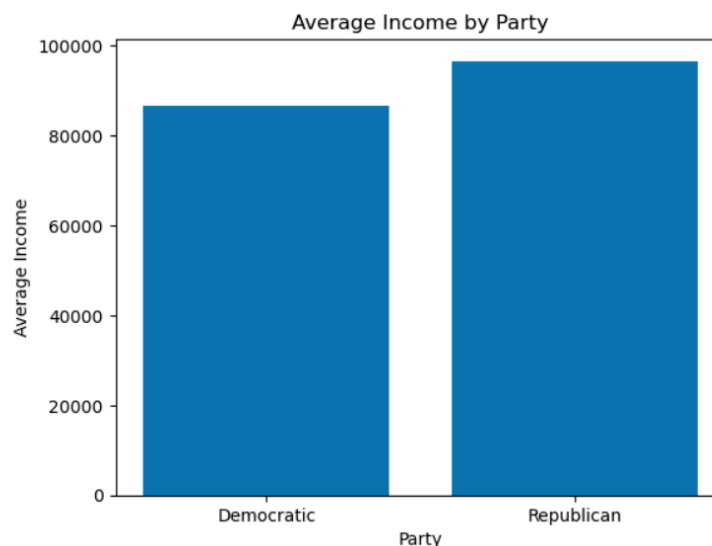
- Hispanic and Portuguese descent individuals have the lowest participation rate of any ethnicity. This is likely due to the lack of voter eligibility within these communities.
- Individuals who own guns are more likely to vote. The 2nd Amendment in the United States is an issue tackled every election year, and these individuals find the retention of their right to own a firearm as incredibly important. 83% (on average) are participants, compared to the 73% of people who do not own a firearm.
- Homeowners are more likely to vote than renters. This is because of issues such as property tax rates, NIMBYs (Colloquialism for “Not in my backyard” - opposition by residents to infrastructure developments in their residential areas),

and zoning and regulatory laws that allow homeowners to develop their homes as they deem appropriate.

- ISPSA(Index of Social Position for Small Areas) is a mix of education and income information to suggest where a voter lies on a scale of 0 to 9 on the "social-economic ladder". There is also an increasing trend of voter turnout with ISPSA. Individuals with higher socioeconomic status tend to have a better understanding of the political process, the importance of voting, and the potential impact of their participation. They also typically have more resources to overcome logistical barriers to voting. Higher socio-economic status is often associated with greater levels of civic engagement and participation in community activities. Individuals who are actively involved in their communities are more likely to perceive voting as a civic duty and to prioritize participation in the electoral process.



- Democrats, on average, are younger than Republicans. Considering our investigation of this dataset doesn't include the median for these groups, the extreme cases (younger individuals tend to be more liberal, and older generations tend to be more conservative) aren't included, swaying the ages.
- Republicans make more (about \$10,000, or 11.25% more) than Democrats. It makes sense that higher-income individuals are more likely to be Republican, as their voting stances rely heavily on the government encroaching on their tax dollars.
- Home values for both groups are similar, within \$30,000 of each other. Republicans pay more; this isn't an indicator of their political affiliation, but rather that home prices around the United States are approximately \$245,000.



- We can see the strongest positive correlation with participation is age and then home value. But overall, the correlation is not that significant, meaning the factors are not that decisive in determining the voter turnout.

	Participation	Age	Family_Head_Count	Income	ISPSA	Home_value
Participation	1.000000	0.201746	0.131241	0.134801	0.102892	0.160212
Age	0.201746	1.000000	-0.172232	-0.102586	0.063990	0.009132
Family_Head_Count	0.131241	-0.172232	1.000000	0.286762	0.127943	0.256046
Income	0.134801	-0.102586	0.286762	1.000000	0.438773	0.612768
ISPSA	0.102892	0.063990	0.127943	0.438773	1.000000	0.526119
Home_value	0.160212	0.009132	0.256046	0.612768	0.526119	1.000000

Regression analysis

Utilizing linear regression models for predicting participation based on multiple factors, several insights have emerged. The Random Forest Regression model performed the best across all evaluation measures, like RMSE, R-squared, and MSE. Surprisingly, the Linear Regression model emerged as the second-best performer, showing similar performance to the Random Forest model.

Examination of feature importance values derived from both the Random Forest and Decision Tree models indicates that features 1 and 9 exhibit relatively high importance, with feature 9 attaining the highest importance score. Feature 1 corresponds to the political party affiliation (Democrat or Republic), a crucial determinant in participation prediction for the reasons stated above. Feature 9 pertains to the ethnicity of the participating individuals, which makes sense considering ethnicity factors have an effect on lifestyles and community which in turn affects voting. Additionally, most voters in the dataset are European, which heavily influences the model's accuracy.

Conclusion

Given these findings, we can answer the following question: What concrete actions can the political campaign take? We can answer this question based on political parties:

Democrat or Republican:

- Democratic
 - Democrats are characterized by lower average income, composed of marginalized communities, and younger individuals. There is no clear connection between their voting rate and any single issue. Still, a vast majority of people who vote are single-issue voters (as signaled by the Republican voter turnout). Campaigns should focus on which issue drives the most Democrats to vote, and use this as the frontrunner of the campaign. The real-life equivalent of this is found in the 2016 election, where Donald Trump sought the Republican candidacy by infamously emphasizing his heightened border restriction.
- Republicans
 - Republicans are characterized by older, mostly White, gun-owning individuals. A whopping 83% of gun owners vote, so emphasizing the importance of retaining this right to bear arms to minimize encroaching on their freedoms would be the best option to both retain and increase the number of these voters. Guns are extremely challenged in Democratic states such as California, so this would be a difficult place to win votes.

In conclusion, the analysis of voter participation in Pennsylvania reveals significant insights into the dynamics shaping electoral engagement. I found that voter turnout varies significantly across demographic factors such as income level, ethnicity, homeownership, and political affiliation. Higher-income households, older individuals, and homeowners exhibit higher rates of voter participation, while disparities persist among ethnic groups, with Hispanic and Portuguese descent individuals showing the lowest participation rates.

Furthermore, the predictive modeling highlights the importance of political party affiliation and ethnicity in determining voter turnout, with features related to these factors demonstrating high importance in participation prediction models.

Moving forward, political campaigns can leverage these findings to tailor their strategies effectively. For Democratic campaigns, identifying and emphasizing key issues that resonate with marginalized communities and younger individuals could bolster voter turnout. Meanwhile, Republican campaigns may find success in mobilizing older, predominantly White, gun-owning individuals by emphasizing the importance of preserving Second Amendment rights.

By understanding the nuanced drivers of voter participation, policymakers and political actors can work towards fostering greater inclusivity and engagement within the democratic process, thereby strengthening the foundations of democracy in Pennsylvania and beyond.