I. INTRODUCTION

Voting is an important civic responsibility that may be less accessible to people depending on factors such as location and income. An important step towards addressing this inequity is recognizing trends of voter registration across municipalities. To this end, we will model and analyze voter registration data across Pennsylvania counties, comparing it to population, average income, age, and racial demographics. Our goal is to understand which demographics are less likely to be registered to vote. This project has significant, wide implications in policy and civic engagement. Citizens who vote are more likely to be informed on current events, and they play an active role in determining who represents them in office. This project aims to model and analyze human behavior, attempting to determine how different factors of an individual's identity come together to influence voter registration trends in their counties as a whole.

II. MOTIVATION

In recent years, there have been many efforts to increase voter registration around the nation. You can even see this on college campuses, as ambassadors aim to get more college students to vote - I myself registered to vote because of someone who helped me through the process when I stumbled into them on campus. Thus, by understanding the underlying factors that may influence voter registration numbers, such as population, income, age, and race, we can craft programs and initiatives that are better suited to engage specific populations, particularly groups which are underrepresented in the general population of voters.

III. LITERATURE REVIEW

One example of relevant literature that aligns with our project is the measurement of voter turnout during the elections. "Voter turnout is a measure of civic participation that many people believe best gauges the health of the electoral process" ("Voter turnout"). We are planning to do a similar measurement to find the specific demographics that are not showing up in the overall votes.

Additionally, previous scholars have sought to understand the variables that impact voter registration. Zhu (2021) found that some of the largest impacts on voter registration among young people included personal demographics (e.g., race and gender), income and socioeconomic status, and education. By focusing our analysis on the state of Pennsylvania, we hope to more closely understand the picture of voter registration and turnout in our home state.

"Voter turnout." MIT Election Lab, 28 April 2021, https://electionlab.mit.edu/research/voter-turnout. Accessed 20 October 2023.

Zhu, Claire. "A Study of Key Factors Influencing Youth Voter Turnout." *American Government and Politics*, 16 Aug. 2021, preprints.apsanet.org/engage/apsa/article-details/6116fcb64cb47934aa2e932d. Accessed 22 Oct. 2023.

IV. SCOPE OF WORK

The scope of our project is to analyze voter registration trends throughout Pennsylvania, with the hope of gaining an understanding of which communities are less likely to be registered. We intend to use this data to gather a full picture of voting demographics around the state. Though we do not plan to perform outreach to the under-registered demographics, this data could potentially be useful for campaigners and advocacy groups who want to most effectively target non-voters. Finally, this project gives us a clear picture of the intersections between voting status, income level, and location that could be extended to other states beyond Pennsylvania.

V. DATASET DESCRIPTION

One dataset we will be using is the United States Census Bureau demographic data for each county in Pennsylvania. Data for Allegheny County such as population, median income, age, racial makeup is available at https://www.census.gov/quickfacts/fact/table/alleghenycountypennsylvania/PST045222. Another one is the Pennsylvania Department of State's Voting & Election Statistics which gives information about voter registration by county. It includes the number of registered voters in each county and a breakdown of each county's voter registration by age under "All by Age." These spreadsheets can be found at https://www.dos.pa.gov/VotingElections/OtherServicesEvents/VotingElectionStatistics/Pages/VotingElectionStatistics.aspx.

Finally, we have used a dataset containing per-capita income information for all Pennsylvania counties as of 2019, available at https://apps.bea.gov/iTab. These are all obtained from publicly available government websites.

VI. EXPLORATORY DATA ANALYSIS

One piece of exploratory analysis from the dataset is the percentage of all voters of eligible age that actually did vote was highest in 1960 with 70%. It slowly decreased over the following elections until it was where only half of people old enough to vote were voting in 1988. Since then it has slowly made its way back up to 66.87% in 2020, and has been increasing since then putting it on a track to go up again in the next election.

VII. METHODOLOGY

In testing our hypothesis of the correlations between voter registration, per capita income, and population, we performed multivariable regression to understand the connection between these factors. By using the R language and R-Studio for statistical analysis, we were able to note trends between county income and voter registration trends (we will explore this in the following section.) Finally, our model and testing regressions confirmed our discoveries and provided a convincing connection between the variables.

VIII. RESULTS & CONCLUSION

In the regression of income vs votes, a scattered result showed us a positive relation between income and voter registration. The

slope was lower, indicating that income does not have as large of an impact on voter registration than population. Looking at the training model, we would say that it is a good fit, neither underfitting nor overfitting. However, the case is less clear for the test result. The test result underfits the original income vs votes regression, because it does not demonstrate as strong a correlation in its slope. This is likely due to there being less data points available on this model. The point at income just above 60,000 is a noticeable outlier, likely caused by insufficient training. We did this in order to see what factors influence voter registration and to what extent. Given what we found, we can now implement voter registration initiatives that target lower income populations to ensure that these individuals are also registering to vote at the same rate as higher income communities in order to ensure that everyone's voice is being heard equally in our democracy.