

Department of Probability and Statistics  
University of California, Santa Barbara

Big Data Analytics, PSTAT 135  
Analyzing Voter Participation Disparities in the United States: A Big Data Approach

Ryan Yee, George Weale, Sarah Liang, and Ryan Lee  
Group 16  
Mar 14, 2024, - March 20th, 2024  
Submitted March 20th, 2024

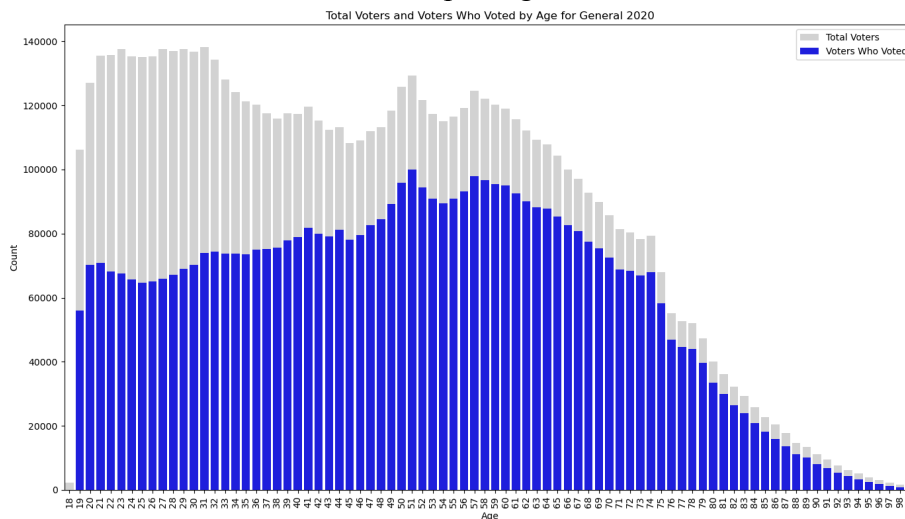
<b>Table Of Contents.....</b>	<b>2</b>
<b>I. Abstract.....</b>	<b>2</b>
<b>II. Preliminary Findings.....</b>	<b>2</b>
<b>III. Results.....</b>	<b>9</b>
<b>IV. Discussion.....</b>	<b>9</b>
Age and Voter Turnout.....	9
Gender Disparities.....	9
Ethnicity and Socioeconomic Factors.....	10
Lifestyle Interests.....	10
<b>V. Conclusion.....</b>	<b>10</b>

## I. Abstract

This PSTAT 135/235 project investigates racial disparities in voter turnout across the U.S., leveraging extensive voter files and big data tools with PySpark and Dataproc. Aimed at identifying the underlying causes of these disparities, the study examines demographic, socioeconomic, and geographic influences. Through data analysis, we hope to find insights for policies that could help our political campaign. In order to be able to complete necessary computing in an efficient manner, we decided to narrow our focus on Georgia. As a swing state, this will help prevent imbalances between democratic and republican voters and provide us with an insightful area to investigate. Our goal is to further understand voting patterns and behavior of U.S. residents, especially when considering differences between and effects of affluent versus non-affluent regions, voting by mail, and accessory lifestyle interests.

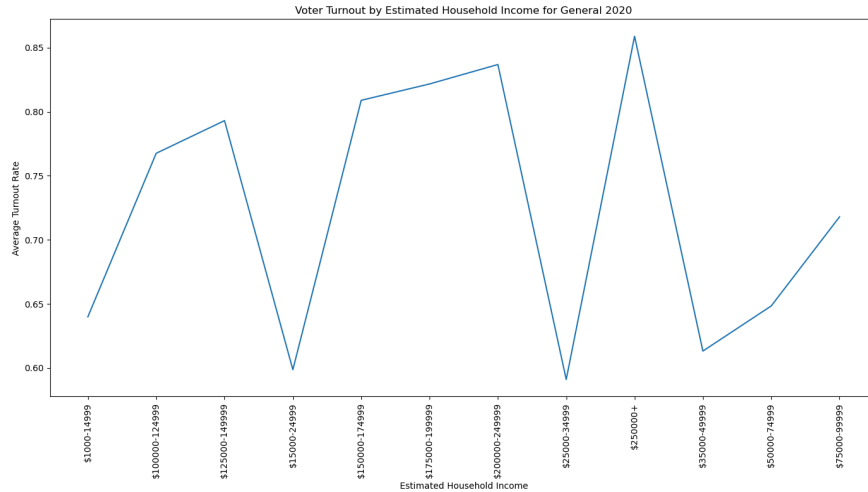
## II. Preliminary Findings

We decided to do our voter analysis on Georgia, a key swing state. In Georgia, voter turnout varies significantly with factors like age, gender, and income, showing lower participation among young people, lower-income individuals, and men. These disparities are linked to socioeconomic differences, reduced political engagement among marginalized groups, and barriers such as strict ID laws and limited polling sites.



**Figure 1:** Age vs Count of votes where blue is the amount that voted, and gray was the amount that could have voted. The younger population is larger and voting less.

The analysis of the provided data indicates a pattern of lower voter turnout among younger populations compared to older demographics shown in Figure 1. Recognizing that multiple factors likely contribute to this trend, we propose to undertake an exploratory analysis to identify the key determinants behind this phenomenon. Focusing on the demographic segment exhibiting the most significant disparity between actual voters and the eligible voting population will offer critical insights into strategies for enhancing voter participation. Notably, as individuals age, the gap between those who vote and those who abstain narrows, suggesting age-related shifts in voting behavior.



**Figure 2:** Estimated Household income vs average turnout rate and we see that the less money that you have the less you would vote.

Further investigation into trends revealed that younger voters, who also tend to fall into lower income brackets, show markedly lower turnout rates compared to wealthier counterparts. This disparity is hypothesized to stem from differing priorities and concerns among these groups, particularly regarding taxation policies shown in Figure 2. Wealthier individuals might be more motivated to vote due to direct financial implications, such as tax legislation, whereas those with lower incomes may prioritize immediate life challenges over political engagement. Identifying and addressing these underlying factors is crucial for developing targeted strategies to increase voter turnout, especially among demographics where the potential for impact is greatest.

```

Count and Ratio for each Interest Column:
Interest_Column Count_Yes Count_Null Ratio_Yes
0 Camping_Hiking 628 4249 0.128768
0 Cooking_General 786 4091 0.161165
0 Cooking_Gourmet 977 3900 0.200328
0 Crafts 1176 3701 0.241132
0 Current_Affairs_Politics 413 4464 0.084683
0 Education_Online 258 4619 0.052901
0 Electronic_Gaming 77 4800 0.015788
0 Exercise_Health 1552 3325 0.318228

```

**Table 1:** Table showing the amount of people that voted yes to different interests and the ratio of them vs the amount of total non responses

Based on Table 1 about the count and ratio for each interest column, we can analyze potential causes for disparity in voter turnout based on different interests/lifestyles. Here are some observations and insights:

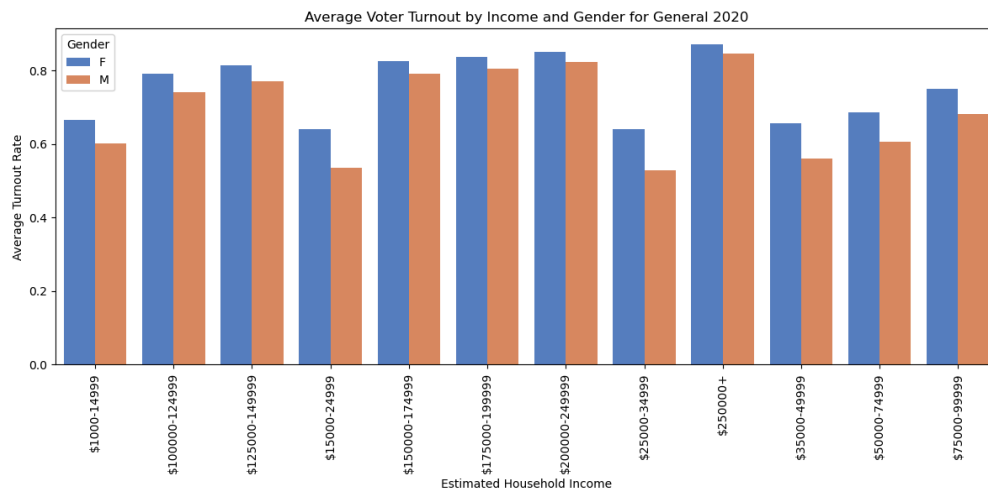
**Exercise\_Health:** This interest column has the highest ratio of Yes to Null values (0.318), indicating a significant interest among voters. It suggests that individuals who prioritize exercise and health-related activities may be more likely to participate in voting. This could be due to a sense of responsibility and awareness about health policies or community well-being.

**Cooking\_Gourmet:** Similarly, the interest in gourmet cooking shows a relatively high ratio (0.200). This might imply that individuals with a passion for culinary arts or gastronomy are engaged in civic activities like voting. It could be related to concerns about food policies, cultural representation, or community events related to food.

**Crafts and Camping\_Hiking:** These interests also show notable ratios (0.241 and 0.129, respectively). It suggests that individuals interested in outdoor activities, nature, and crafts are active in civic engagement. This could be linked to environmental concerns, outdoor recreational policies, or community initiatives related to arts and crafts.

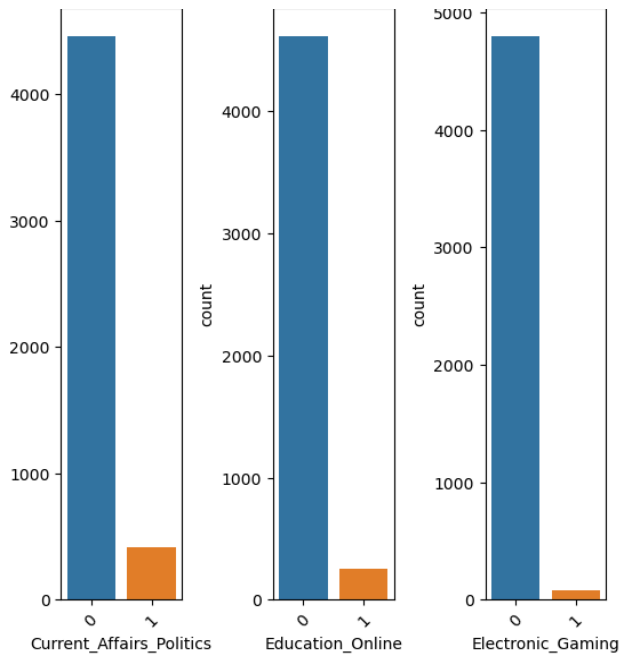
**Current\_Affairs\_Politics:** Surprisingly, the interest in current affairs and politics has a lower ratio (0.085). This might indicate that while individuals have an interest in these topics, it doesn't always translate directly into higher voter turnout. Other factors such as trust in the political system, accessibility to voting, or campaign effectiveness could influence this disparity.

**Education\_Online and Electronic\_Gaming:** These interests have the lowest ratios (0.053 and 0.016, respectively). It suggests that individuals interested in online education or electronic gaming may be less engaged in traditional civic activities like voting. This could be due to various reasons such as demographic differences, generational trends, or perceived priorities.



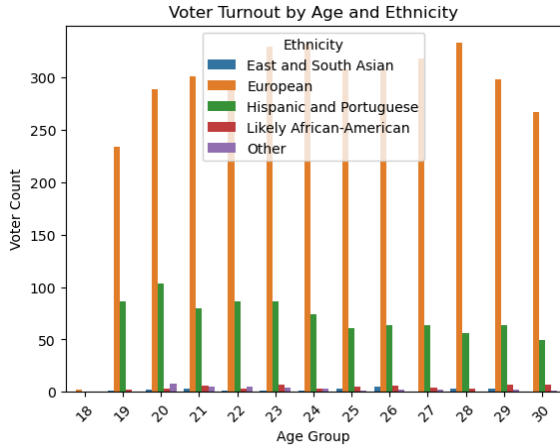
**Figure 3:** Estimated Household income vs voter turnout with two different bars for women and men in each of the groups showing that women vote more than men.

In addition, we decided to examine the differences in voter turnout by gender compared to income level. As shown in Figure 3, in every income cohort the female population has a higher voter turnout than the male population. We believe discovering the potential factors causing females to have a higher voter turnout ratio would allow us to uncover the necessary actions that must be taken in order to increase male voter turnout.



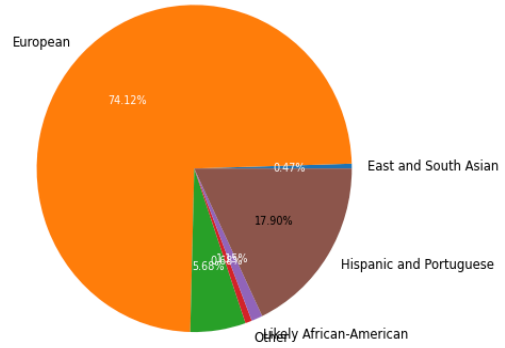
**Figure 4:** Count vs responses where 1 means yes and 0 means no or no response

We hypothesize that the reduced voter turnout among younger populations may be attributed to their concentration in college campuses, where a lack of awareness about regional laws or competing priorities diminishes their propensity to vote. This theory is supported by observations that younger, non-voting individuals often engage in online electronic gaming and show interest in current affairs, characteristics typical of college students with limited financial resources. This demographic profile aligns with our expectations and suggests that younger voters represent a significant opportunity for increasing electoral participation. Successfully engaging this group through targeted campaigning could unlock substantial gains in voter turnout, leveraging their potential to influence political outcomes significantly. Analyzing voter turnout data shown in Figure 4 among younger demographics reveals complexities beyond initial observations. While it appears that youth participation is low and white individuals vote in disproportionately high numbers relative to their population share, this analysis may not capture the full picture. Further investigation is warranted.



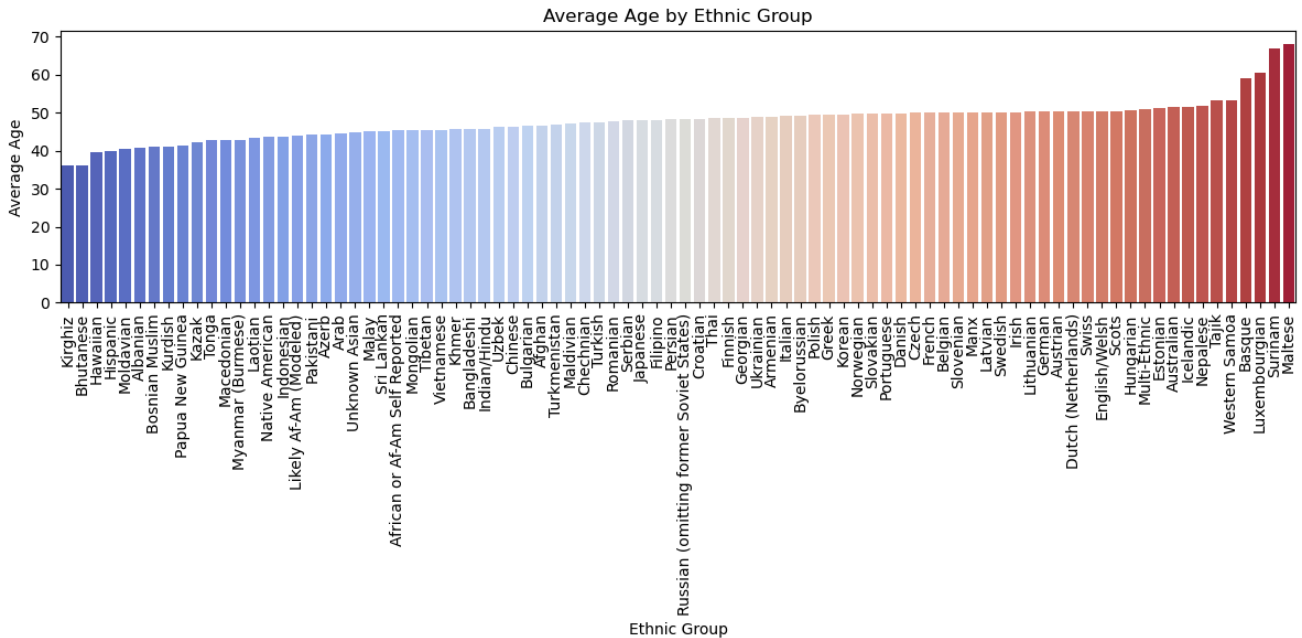
**Figure 5A:** Young age vs voter count with different ethnicity count for each age

**Ethnic Distribution of Georgia Voters who Specified Ethnicity**



**Figure 5B:** All voters in Georgia with different ethnicity groups shown

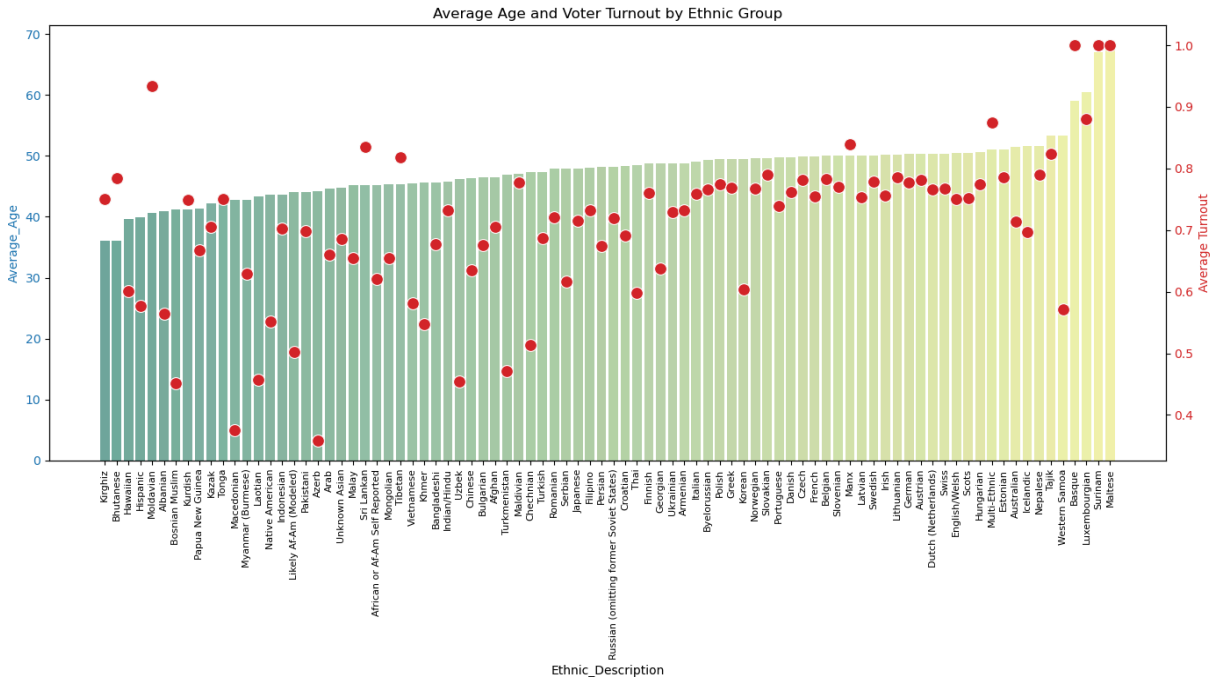
The data shown in Figures 5A and 5B show that Europeans represent approximately 74% of the voter turnout, suggesting a need to delve deeper into specific community behaviors and their correlation with age to understand disparities in voter engagement fully. Notably, older individuals, particularly of European descent, exhibit higher voting rates. This pattern prompts a closer look at how age and ethnic backgrounds intersect with electoral participation.



**Figure 6:** Segmented ethnic groups vs average age showing a specific set of ethnicities that have extreme low or high average ages

In Figure 6, we notice that some ethnic groups, like the Kirghiz and Bhutanese, have a relatively low average age of about 35 years. On the contrary, ethnic groups, like the Basque,

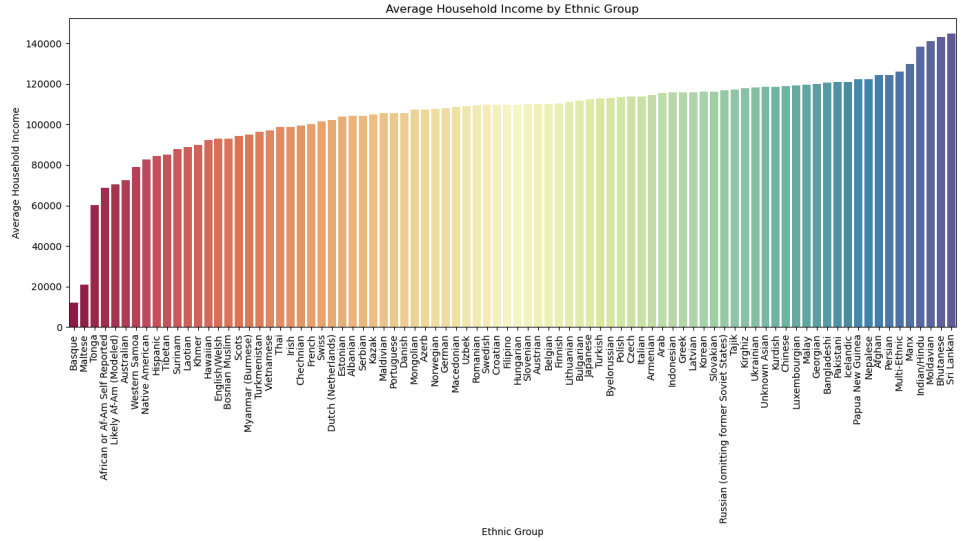
Luxembourgian, Surinam, and Maltese all have unusually high average ages of about 57-67 years. If we consider this in addition to voter turnout within each ethnicity, we can discover more about the possible nuances of ethnicity impacting voter turnout.



**Figure 7:** Specific ethnic group vs voter age and average turnout to show that there is a trend with older age ethnic groups

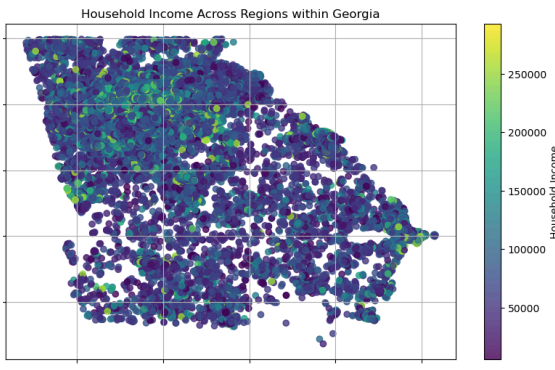
Upon expanding our analysis, we observed that certain ethnic communities, which are generally older, mirror the previously identified trend of higher voter turnout with age shown in Figure 7. However, this correlation diminishes among younger ethnic groups, hinting at potential influences of socioeconomic factors, such as average household income, on voting behavior.



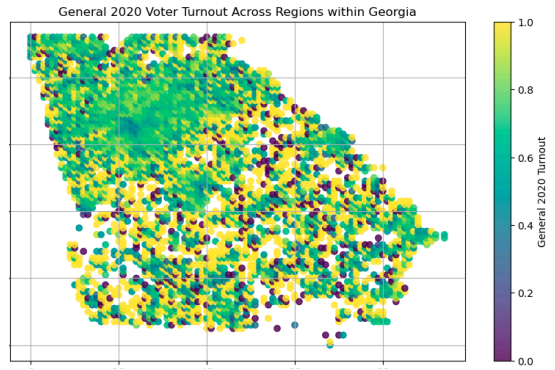


**Figure 8:** The same ethnic groups vs average household income showing a larger disparity in these groups.

When comparing ethnic groups to average household income, the data did not align as anticipated shown in Figure 8. Older groups with higher turnout rates did not necessarily correspond to those with the highest income levels. This discrepancy suggests that income alone may not be a definitive factor in predicting voter turnout across different communities. Given the diversity within smaller groups, this finding challenges the assumption that higher income correlates directly with higher electoral participation, indicating that further exploratory analysis is required to uncover more nuanced relationships within the data.



**Figure 9A:** Latitude vs Longitude with a heat map to show average household income



**Figure 9B:** Latitude vs Longitude with a heat map to show average voter turnout

Central Georgia, particularly the Atlanta area, exhibits higher household incomes compared to the more southern regions of the state shown in Figures 9A and 9B. The distribution of data reflects greater density and affluence, with more pronounced and consistent voter registration and activity in these wealthier areas. Notably, pockets of higher income and affluence are also identified in Columbus and Savannah, positioned along Georgia's coastal flanks. In these wealthier, more populous cities, the elevated cost of living and higher incomes correlate with a denser voter base.

Analyzing voter turnout, for the General 2020 election, reveals a nuanced picture. In Figure 9B, the binary representation of voter turnout, with 0 indicating no turnout and 1 indicating participation, provides insight into regional voting patterns. Intermediate values within this spectrum, calculated for specific areas in grid cells of 0.05 by 0.05 degrees, represent the average turnout. A value approaching 1 suggests high participation in the 2020 election, whereas a value near 0 indicates lower engagement. This analysis shows that affluent regions around Atlanta and Savannah consistently demonstrate higher average voter turnout compared to other areas, which exhibit greater variability in their data, alternating between extremes of 0 and 1. This variance points to a possible data scarcity in less affluent, less densely populated regions.

In summary, affluent areas tend to have higher voter turnout, likely influenced by both higher population density and greater affluence. This pattern underscores the interconnectedness of socioeconomic status, population distribution, and electoral participation in Georgia.

### **III. Results**

In this study, we used a logistic regression model to understand what affects whether people vote. Logistic regression is a way to see how different things (like someone's interests or how much money they make) might influence a yes/no outcome. We chose this model because it works well for predicting outcomes where there are two possible answers, such as voting yes or no.

First, we looked at how lifestyle interests (like exercise, cooking, or gaming) might predict voting behavior. The model's accuracy, measured by the AUC (Area Under the Curve) metric, was 57%. This means the model was somewhat able to distinguish between those who voted and those who didn't, based on their interests. However, this result wasn't very strong, which suggests that lifestyle interests alone don't tell us much about why someone would vote.

We tested a model using age, gender, and household income as predictors. This time, the AUC improved to 58.4%. This increase shows that these factors are more informative for predicting voting behavior than lifestyle interests alone.

When we combined lifestyle interests with age, gender, and household income in one model, the AUC reached 61.2%. This is the highest score among our tests, indicating that including a mix of personal interests and demographic factors gives a clearer picture of voting behavior.

### **IV. Discussion**

The analysis conducted in this study sheds light on several key factors influencing voter turnout and disparities across different demographic groups in Georgia. Our investigation revealed significant variations in voter participation based on age, gender, income level, ethnicity, and lifestyle interests. Here, we discuss the implications of our findings and their relevance to understanding voter behavior and designing effective strategies for increasing civic engagement.

One of the prominent findings of our study is the substantial difference in voter turnout

between younger and older demographics. Younger populations, particularly those in college or with lower incomes, exhibited lower participation rates compared to older individuals. This trend highlights the importance of targeting outreach and educational campaigns towards younger voters to increase their awareness of civic responsibilities and the impact of political decisions on their lives. Strategies such as providing voter education in educational institutions and leveraging digital platforms for engagement could be effective in mobilizing younger voters.

Our analysis also uncovered gender disparities in voter turnout, with females consistently showing higher participation rates across different income levels. Understanding the factors driving this disparity is crucial for developing initiatives to enhance male voter turnout. Possible strategies include addressing barriers to voting faced by men, such as accessibility issues or lack of awareness about voting procedures. Additionally, targeted messaging and outreach campaigns tailored to specific gender demographics could help bridge the gap in voter turnout between males and females.

Ethnicity and socioeconomic status were found to be significant determinants of voter turnout. While older age groups, particularly those of European descent, exhibited higher participation rates, disparities emerged among younger ethnic groups. These findings suggest that factors beyond income, such as cultural influences and historical contexts, play a role in shaping voting behavior. Addressing barriers to voting among marginalized communities and promoting inclusivity in the electoral process are essential steps towards achieving equitable representation and participation.

Analyzing lifestyle interests provided additional insights into voter behavior. While certain interests, such as exercise and health-related activities, showed a positive correlation with voter turnout, others like online education and electronic gaming displayed lower engagement in civic activities. These findings emphasize the need to tailor outreach efforts based on individuals' interests and priorities, ensuring that diverse segments of the population are actively engaged in the democratic process.

## **V. Conclusion**

Based on our findings we can conclude that compared to lifestyle variables, factors such as age, gender, and household income were the most impactful in determining voter turnout. In addition to these factors, we found that ethnicity could play a role in voter turnout as a few of the plots we created showed relationships between ethnicity and age as well as gender. In terms of a political campaign to encourage more voting, the young and lower income population should be targeted to increase voter turnout.

Our study underscores the multifaceted nature of voter participation disparities in Georgia. By leveraging big data analytics and advanced tools like PySpark, we were able to uncover patterns and trends that contribute to these disparities. Our findings highlight the importance of targeted interventions, educational initiatives, and policy changes to promote inclusive and equitable voter engagement.

Moving forward, future research should focus on refining predictive models, exploring

additional variables that influence voter behavior, and evaluating the effectiveness of intervention strategies. By addressing the underlying causes of voter turnout disparities and fostering a culture of civic engagement, we can work towards a more representative and participatory democracy in Georgia and beyond.