

# **VoteAmerica Project - Team 2**

## **Final Report**

### **CS506-A1**

#### **Team members:**

Rania Hakimi  
Saurabh Johri  
Kaiyan Xu  
Temi Lajumoke

#### **Introduction:**

VoteAmerica is a non profit organization that was formed by elections and technology experts with the goal of increasing voter turnout during the 2020 elections. The organization sent over 100 million messages across the united states among them were 8 million reminders. Our project consists of the audience analysis of the people targeted by their campaign. We aim to analyze the audience demographics, understand who was reached, analyze the voting rates and contactability of the voters and present our results in an informative manner.

#### **Results:**

##### **Early Analysis:**

We first started our analysis by answering key questions provided in the proposal such as how many people were targeted and from them how many people actually voted.

After removing duplicates of voter id we found that the number of people targeted is 2,804,686. We then looked at the vote method and assumed that the people with a “nan” value as the vote method did not vote and from that we extracted the number of targeted people that voted which is 2,006,224.

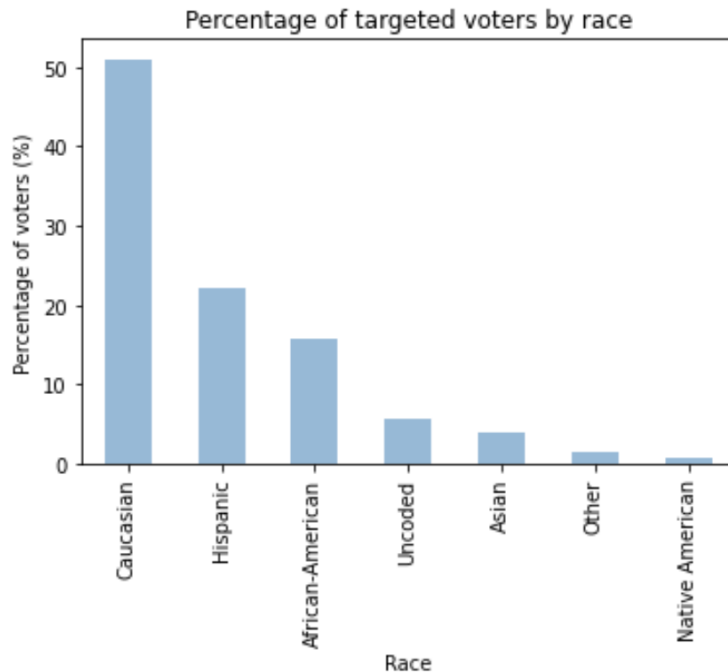
We found that 2.98% of the voters have different “state” information in the target voter dataset and messages dataset. This is because they are from the Virgin Islands, and they are likely to receive messages to another state. On average, each voter id has 1.11 cell phone ids. Since Vote America sent messages to all the cell phone ids a voter can have, we would only analyze the message information that is sent to the cell phone id that is most likely to be used by the voter.

To compare our previous results with the United States we extracted from the American Community Survey data, the population of the United States on a 5 year age from 2015 to 2019 which is 324,697,795.

## Demographic breakdown of targeted voters:

### By race:

The graph below shows the percentage of targeted voters based on race:



When looking at the race breakdown of the targeted voters we see that caucasian represents % 51 of the targeted voters,

### Demographic breakdown of targeted voters based on race in states where Caucasian was not the majority:

The following table shows the demographic breakdown of race for states where ‘Caucasian’ wasn’t the majority race unlike the overall statistic depicted by the graph above. For example, **Florida, Texas, North Carolina, and Georgia** have majority voters that are not Caucasian.

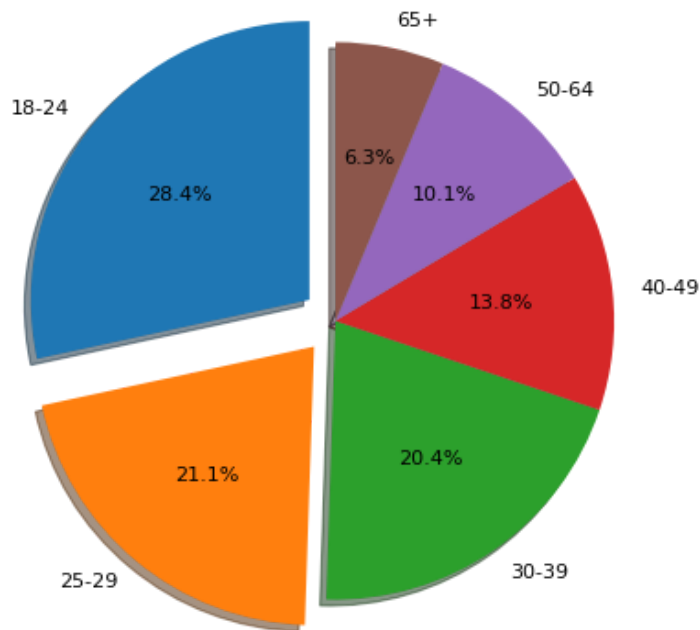
Race\States	Texas	Georgia	North Carolina	Florida
African-American	16.70%	53.26%	46.91%	22.34%
Asian	9.72%	10.12%	7.52%	5.39%
Caucasian	21.01%	23.73%	27.57%	30.68%
Hispanic	43.72%	7.39%	9.30%	35.74%
Native American	0.12%	0.36%	1.26%	0.46%
Uncoded	8.46%	1.06%	1.47%	0.50%

Other	0.27%	3.86%	5.96%	4.88%
-------	-------	-------	-------	-------

Highest % in state

### By age:

The pie chart below shows the breakdown of targeted voters given the age buckets:  
[18-24, 25-29, 30-39, 40-49, 50-64, 65+]



The majority of voters are young, with the most belonging to the age group of 18-24 years old followed by 25-29 years old and decreasing percentage of total voters for increasing age groups.

### By gender:

The table below shows the percentage of voters based on gender:

Gender	Men	Women	Unknown
Percentage	35.83 %	55.05%	9.12 %

Where we see that women have been the most reached.

### By marital status:

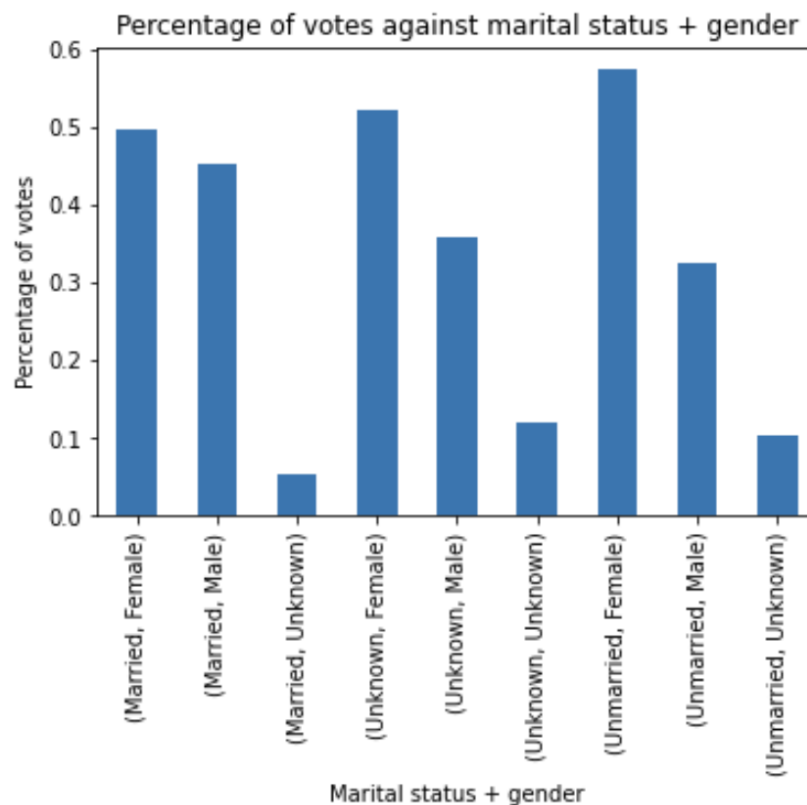
The table below shows the percentage of voters based on marital status:

Marital Status	Married	Unmarried	Unknown
Percentage	25.53 %	68.90 %	5.57 %

The most votes came from unmarried people followed by married people and people with an unknown marital status. People in all three categories most commonly voted through absentee ballots.

### By marital status + gender:

When combining marital status and gender we found that the most targeted groups were both unmarried women and men, as we can see in the graph below:



We added below a table with the actual number of voters.

Marital status\Gender	Men	Women	Unknown
Married	355,812	390,413	41,682
Unmarried	688,490	1,219,217	218,920
Unknown	218,920	89,600	20,887

## Demographic breakdown of New American Majority Group:

We started here our analysis of the New American Majority Group

After removing the duplicates of voter ids we computed the percentage of people in the group vs not in the group:

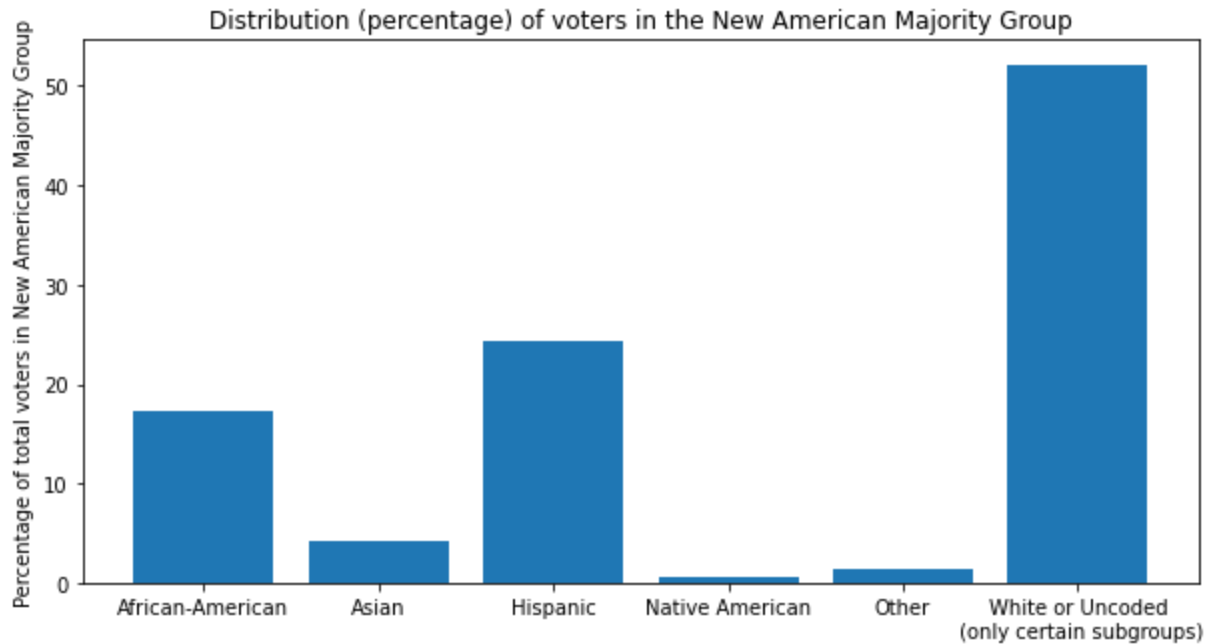
Group	New American Majority Group	Non-New American Majority Group
Percentage %	90.28%	9.72%

## Percentage Breakdown of new american majority subgroups:

New American Majority Group		Percentage of People Texted	Percentage of people who Voted
African-American		15.58	16.57
Asian		3.78	4.26
Hispanic		22.01	20.91
Native American		0.57	0.49
Other		1.35	1.59
White or Race Unknown	Unmarried Women 30-49	12.64	12.01
	Young Men (18-29)	13.58	13.43
	Young Women (18-29)	16.70	17.72
	Young, Gender Expansive Person (18-29)	4.02	3.65
	Young, Gender Unknown Person (18-29)		
Not Part of New American Majority		9.72	9.32

From these results we can see that VoteAmerica targeted effectively the community intended with a percentage of 90.28% from the whole targeted voters.

## Breakdown based on race in the group:

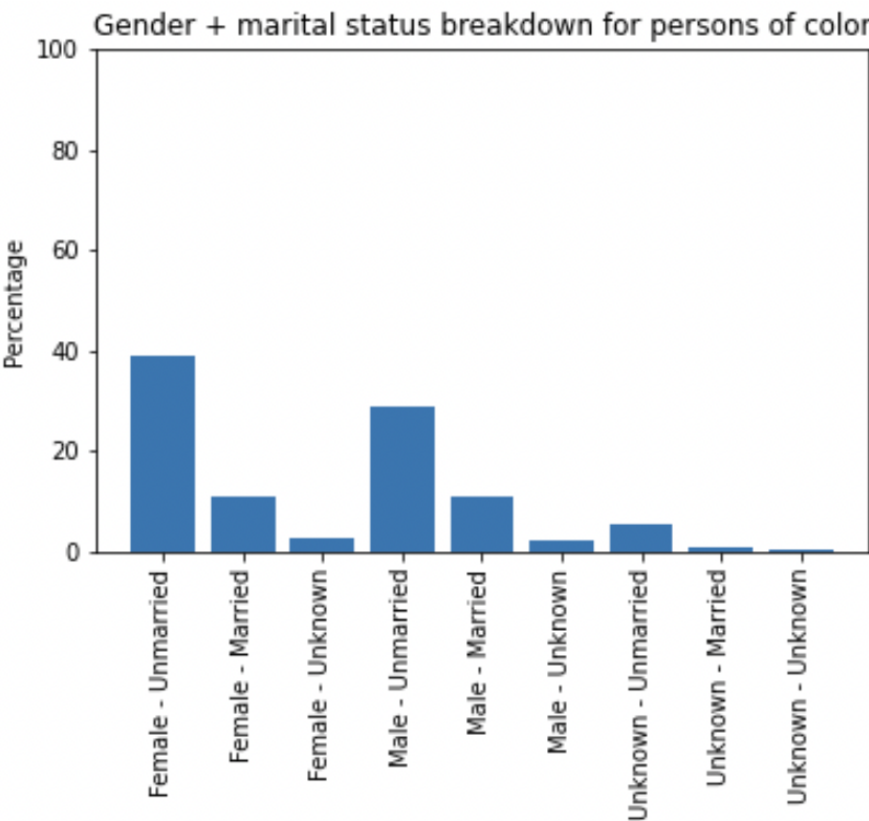


The above graph shows the breakdown of the targeted people that belong to the New American Majority Group. The graph shows a percentage of the total in the New American Majority Group for each group. The White or Uncoded race group only accounts for certain subgroups which are: Unmarried Women from age 30-49, Young Men from age 18-29, Young Women from age 18-29, Young Gender Expansive Person from age 18-29 and Young Gender Unknown Person from age 18-29. In the provided data, we only had three values for gender (male, female and unknown) so we combined the Gender Expansive and Gender Unknown categories into Gender Unknown.

Gender\Marital Status	Married	Unmarried	Unknown
Men	10.86%	28.99%	1.92%
Women	11.01%	38.71%	2.45%
Unknown	0.49%	5.22%	0.36%

The above table shows the breakdown for POC by gender + marital status where we found that unmarried women and unmarried men have the highest percentage in POC (NAMG).

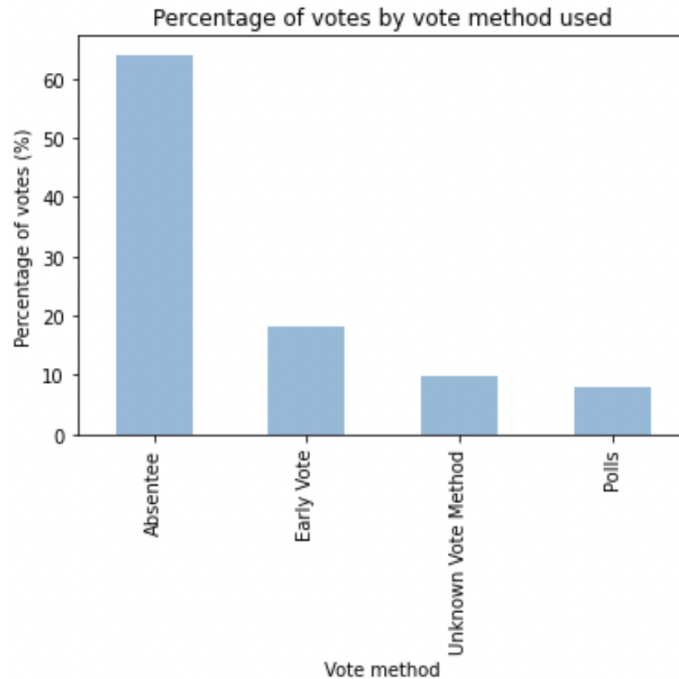
Below is a graph for better visualization:



## Analysis of vote method:

Looking at the different voting methods on the targeted voters, “absentee” vote method was the highest voting method used, amounting to over 60% of the overall voting methods.

The graph below represents the percentage of votes for each voting method used by the voters.



Since VoteAmerica targeted people who received vote-by-mail ballots that they requested or were sent to them by the government, these results make sense.



## Vote method by state:

After comparing the vote method for each state, we found that most of them have the absentee voting method as the highest except KS (vote method unknown), NE (vote method unknown), OH (vote method unknown) and Arizona (early vote) as shown in graph below.

Vote Methods Distribution in Targeted States				
	Absentee (%)	Early Vote (%)	Polls (%)	Unknown Vote Method (%)
AZ	1.09	90.18	8.40	0.33
CO	86.15	1.33	11.16	1.35
FL	78.90	13.69	7.20	0.20
GA	91.99	0.18	0.33	7.50
IA	93.19	0.07	6.59	0.15
KS	0.70	0.14	0.07	99.10
ME	96.01	0.09	3.73	0.17
MI	88.65	0.11	11.09	0.14
MN	84.83	0.22	14.85	0.09
MT	97.32	0.52	1.62	0.54
NC	63.16	28.46	8.00	0.38
NE	0.88	0.12	0.15	98.84
NV	52.88	27.16	18.64	1.32
OH	0.41	0.09	0.05	99.45
PA	84.73	0.10	15.02	0.15
TX	80.35	15.54	3.18	0.93
UT	97.73	0.90	0.46	0.91
WI	86.26	0.35	13.13	0.26
% in the State				
100%	75%	50%	25%	0%

For the states of KS, NE and OH, having an unknown vote method does not necessarily mean that it's not absentee.

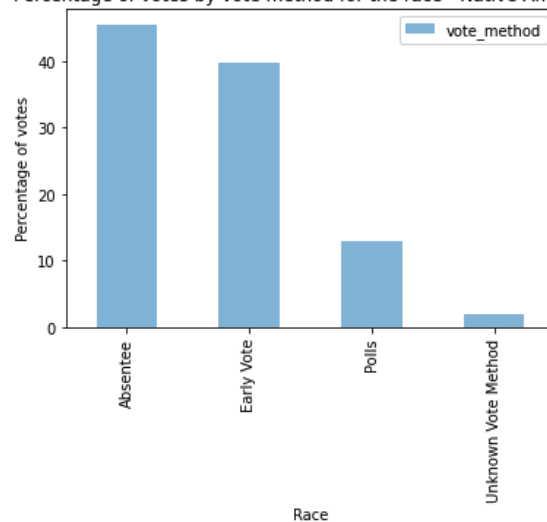
For these states we know that the vote method has not been accurately collected which explains the very low percentages for Absentee, Early vote and Polls vote methods. Therefore the breakdown of vote method by states is not necessarily meaningful.

## Comparing vote method in each race:

Race \ Vote method	Absentee	Early Vote	Polls	Unknown vote method
African-American	72.61%	8.96%	6.73%	11.70%
Asian	77.30%	12.00%	5.40%	5.30%
Caucasian	62.45%	16.82%	8.03%	12.70%
Hispanic	59.25%	29.18%	9.72%	1.85%
Native American	45.54%	39.68%	13.01%	1.76%

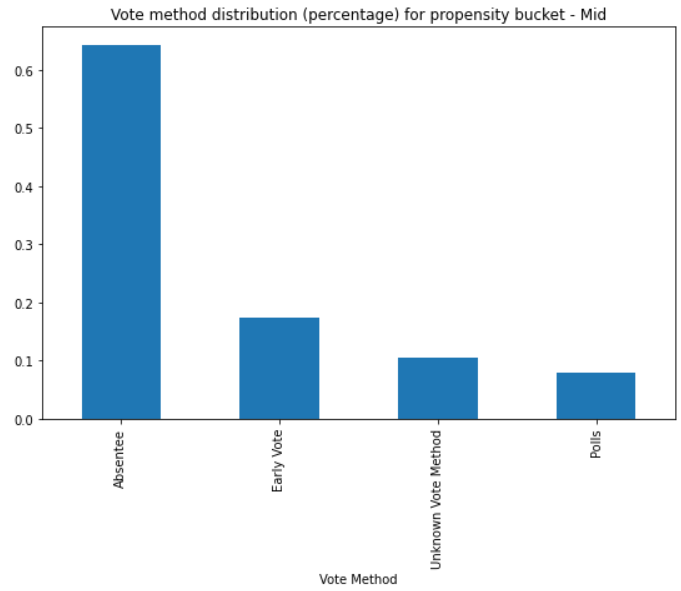
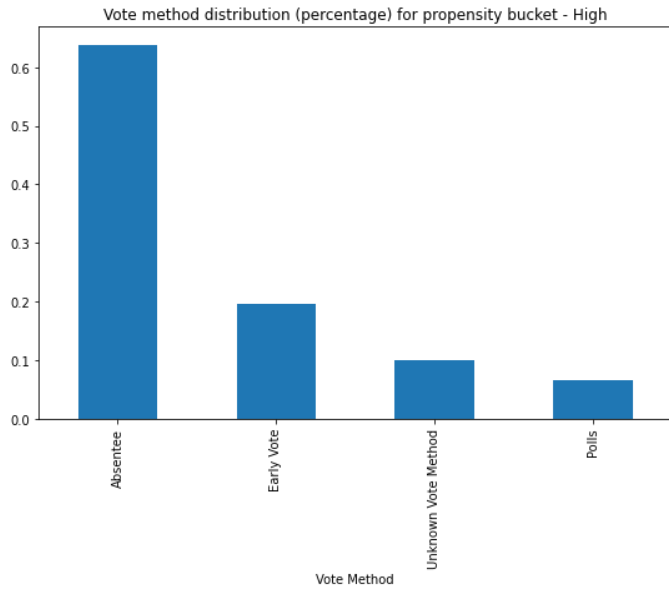
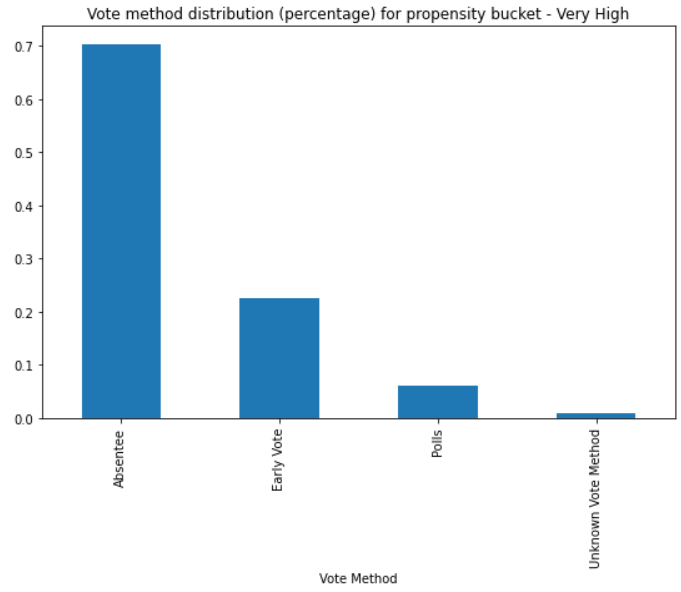
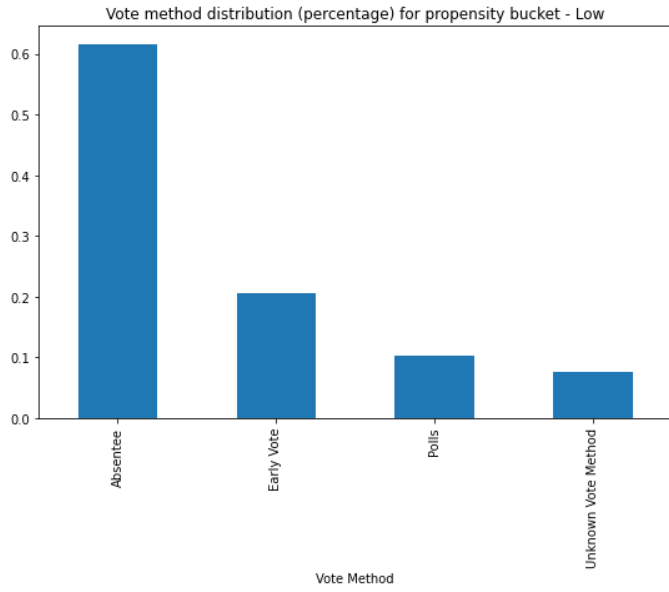
Similarity Highlighted

Percentage of votes by vote method for the race - Native American



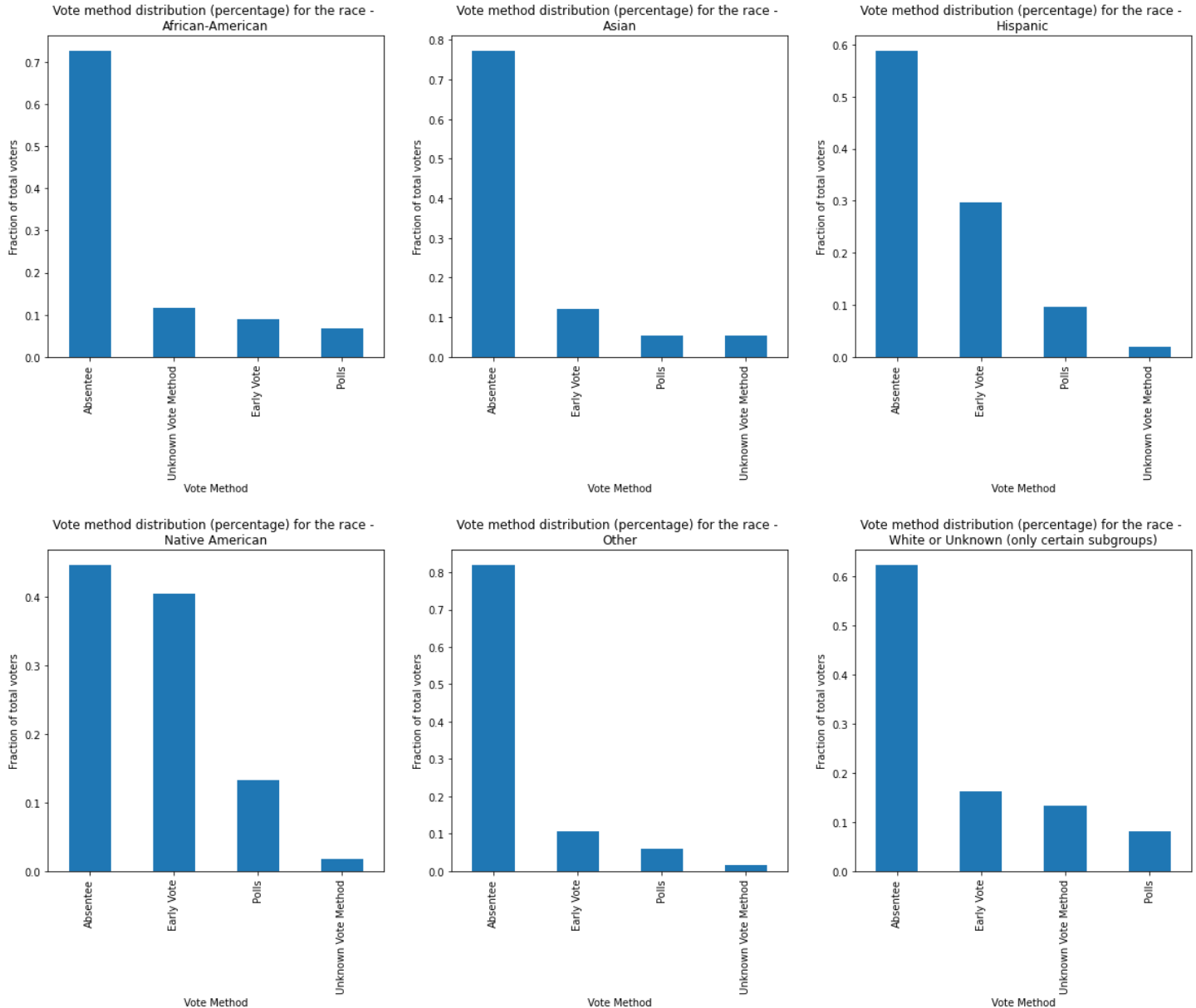
The highest number of votes came from the Caucasian demographic, while the lowest number from Native Americans. After comparing the voting method for each race, we noticed similar patterns in terms of voting methods. Similar to the overall vote method trend, people from different races voted by absentee ballot most commonly. However, Hispanic and Native American voters also had early voting numbers comparable to absentee ballots.

## Vote method analysis against Vote propensity:



All vote propensities show a similar pattern with Absentee method having the highest representation followed by Early Vote and then Unknown Vote Method. Polls had the least representation among all voting methods across all propensity buckets.

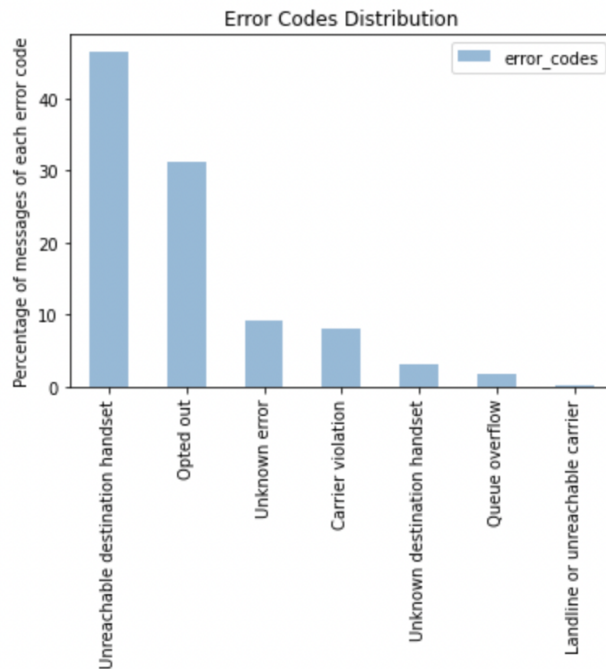
## Vote method distribution for the New American Majority Group:



All groups have similar patterns in vote method used as well with most people using absentee ballots.. However, Native Americans had a comparable turnout of early voters among all the groups followed by Hispanic people.

## Contactability Analysis:

### Overall contactability:



In this part, only messages of the cell phone that is most likely to be the cell phone used by voters are considered. After dropping duplicates that share both the same voter\_uuid and error\_code, there are 3,105,013 messages in total, and the number of messages that went through (with a nan value in error\_code) is 2,568,263, so the overall contactability is 82.7%. Since duplicates are dropped, each error code only counts once if it happens multiple times on a voter, every type of error code is counted and avoids repetitive counts as one user may constantly face the same issue when not receiving the messages.

From the graph, the most usual case for a message not to go through is “unreachable destination handset,” which includes that the device is likely powered down, out of the service area, or may not accept the messages. The second common case is that voters chose to opt-out.

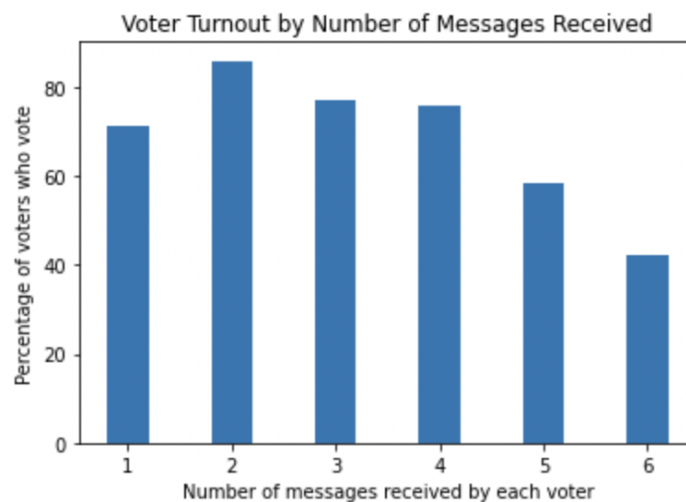
### Patterns:

Focusing on the voters who choose to opt-out, we filtered the above messages and got 167,335 messages from those voters who would choose to opt-out. By calculating the average number of messages sent to this part of voters, we found that they decided to opt out after receiving 1.54 messages on average.

Error Codes Distribution in Targeted States								
	Unreachable destination handset	Opted out	Unknown error	Carrier violation	Landline or unreachable carrier	Unknown destination handset	Queue overflow	Total error codes in the State
AZ	56.19%	20.25%	13.68%	6.06%	0.05%	3.75%	0.03%	67235
CO	33.90%	39.86%	9.05%	14.12%	0.06%	2.94%	0.06%	120169
FL	45.73%	28.86%	8.69%	6.81%	0.07%	2.63%	7.21%	104669
GA	55.09%	22.91%	7.71%	11.51%	0.10%	2.18%	0.50%	34418
IA	48.18%	41.03%	6.59%	2.67%	0.12%	0.97%	0.44%	5172
KS	49.19%	32.04%	7.63%	5.36%	0.07%	3.69%	2.00%	5387
ME	37.97%	49.45%	7.55%	4.76%	0.09%	0.09%	0.09%	2291
MI	54.44%	26.64%	10.06%	3.60%	0.11%	5.01%	0.14%	41532
MN	39.45%	46.41%	6.59%	3.91%	0.02%	3.60%	0.02%	4887
MT	49.38%	38.06%	7.41%	4.56%	0.09%	0.50%	0.00%	7924
NC	57.55%	24.22%	9.56%	6.42%	0.10%	1.18%	0.98%	9126
NE	49.47%	33.48%	6.59%	4.05%	0.08%	6.21%	0.13%	3976
NV	67.57%	23.89%	0.54%	0.99%	3.54%	2.28%	1.19%	12134
OH	52.62%	30.76%	9.42%	6.19%	0.05%	0.92%	0.04%	33079
PA	48.40%	29.00%	9.60%	6.35%	0.04%	6.46%	0.14%	36678
TX	48.75%	26.80%	9.77%	11.24%	0.26%	2.40%	0.77%	5292
UT	37.98%	44.64%	7.52%	4.85%	0.53%	4.39%	0.08%	27141
WI	43.35%	33.07%	8.09%	4.34%	0.03%	1.90%	9.24%	3735
% in the State				100%	75%	50%	25%	0%

After comparing the error codes in our targeted states, the pattern of error codes are generally consistent among all the states. The differences occur in CO, ME, MN, UT, where opt out messages occur more than unreachable destination handset. In NV, there is a higher proportion of voters who face landline or unreachable carriers than other states.

### Voting rate based on number of messages received:



In general, as the number of messages that a voter receives increases, their likelihood of not voting increases. However, it is surprising to find that the voting rate of voters receiving 2 messages is higher than those receiving 1 message. This is probably because of the timeframe and time gap voters receive messages. Other than that, voters receiving more messages have lower voter turnout, basically due to their original low voting propensity.

## Error code analysis:

The below table illustrates the distribution of messages and their error codes over each cell accuracy score, with 1 being the lowest and 4 being the highest. A cell accuracy of 0 means there was no data available. As expected, the cell accuracy of 4 has the highest percentage of messages that went through. Percentage of opt-out is also highest when cell accuracy is 4.

Error Codes Distribution in Each Cell Accuracy Score					
Cell Accuracy Score	0	1	2	3	4
Unreachable destination handset	48.61%	47.31%	50.84%	48.45%	28.13%
Opted out	28.77%	29.36%	26.33%	30.94%	50.10%
Unknown error	10.27%	9.64%	9.38%	9.84%	5.46%
Carrier violation	9.31%	8.66%	7.04%	8.52%	6.73%
Landline or unreachable carrier	0.07%	0.19%	0.30%	0.05%	0.20%
Unknown destination handset	1.05%	4.38%	4.20%	0.41%	6.41%
Queue overflow	1.91%	0.46%	1.91%	1.78%	2.96%
Total Error Codes	88393	127321	144202	115017	61817
Went through	476573	530770	616098	511232	433590
% went through	84.35%	80.65%	81.03%	81.63%	87.52%
% opt out	4.50%	5.68%	4.99%	5.68%	6.25%

% in each cell accuracy score

100%

75%

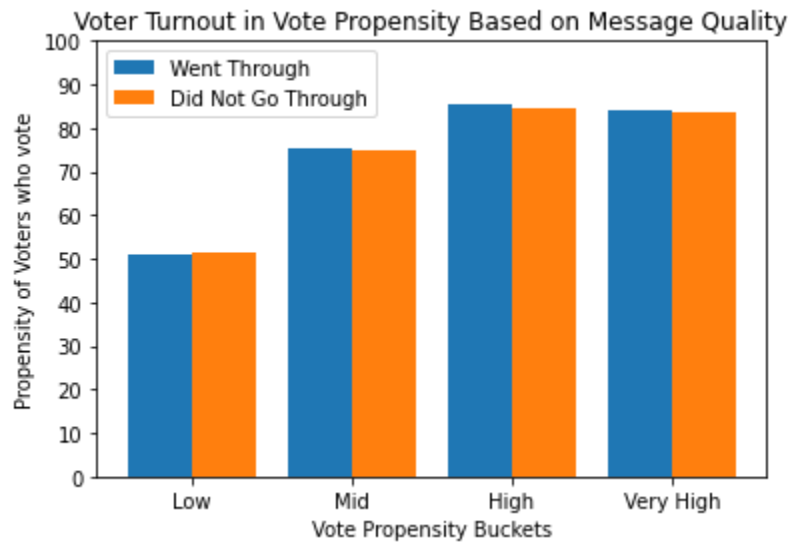
50%

25%

0%

## Voting rate based on whether the messages went through in each voting propensity buckets:

Assuming voters with similar voting propensity have the same likelihood of voting originally, comparing voter turnout based on whether the messages went through to the voters can prove whether messages sent by Vote America stimulate voters to vote.

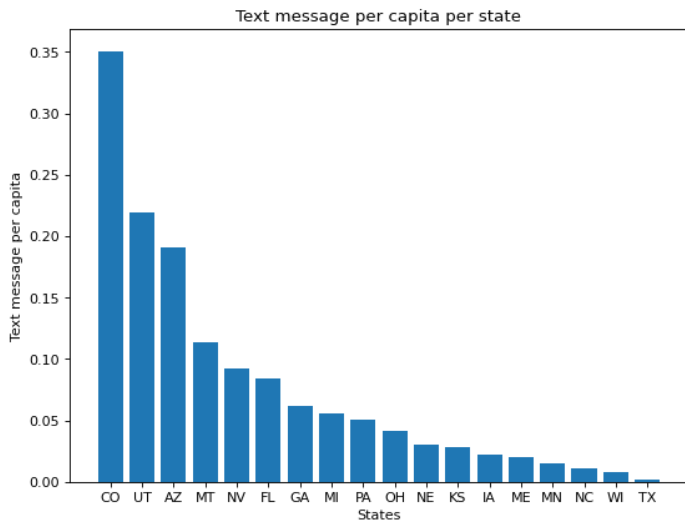


In Mid, High and Very High propensity buckets, the voting rate of voters whose messages went through successfully are slightly higher than voters whose messages did not go through. However, in low propensity, the voting rate of voters whose messages not went through is slightly. However, since the main goal is to stimulate voters having mid propensity to vote, sending messages to voters is still effective.



## Message per capita per state:

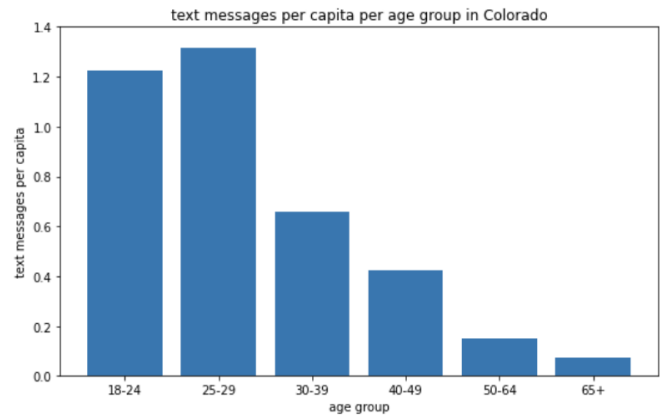
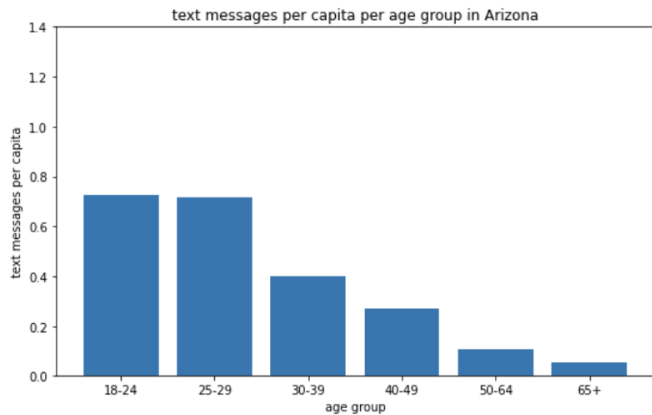
We defined message per capita per state as a measure of the number of text messages received by a voter in each of the states observed. We computed this by generating the total number of messages sent in each state and then extracting the total number of residents from the 2019 5-year acs census data. Finally, we divided the total number of messages sent in each state by the total number of residents in the respective state.



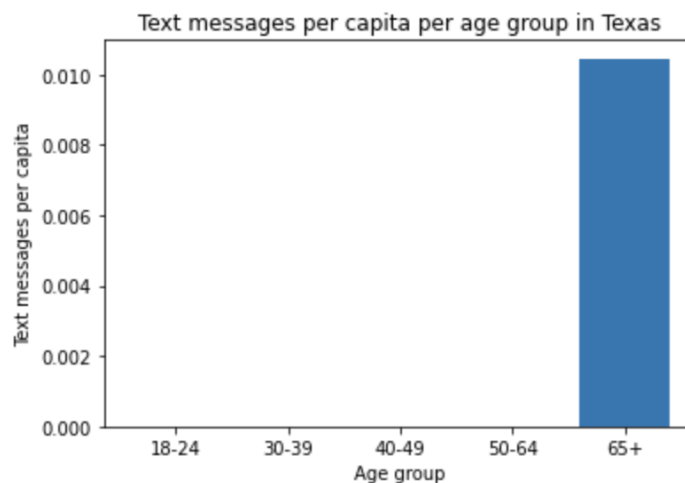
From our results, we can observe that Colorado had the highest text message per capita per state of any state. This was closely followed by Utah and then Arizona. While states like Texas and Wisconsin had the lowest text message per capita per state. Interestingly, the top 3 states with the highest text messages per capita lie along the same/similar geographical longitude. These states had a cumulative text message per capita greater than those of all other observed states combined.

## Message per capita per age group per state:

We defined message per capita per age group per state as a measure of the number of text messages received by a voter in each age group in each of the states observed. We computed this by generating the total number of messages sent to voters in each age group in every state and then extracting the total number of residents in the said age group in that state from the 2019 5-year acs census data. Finally, we divided the total number of messages sent to each age group in every state by the total number of residents in that age group in the respective state.

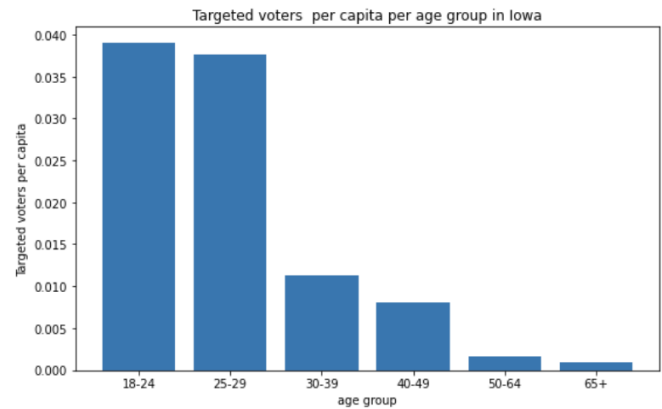
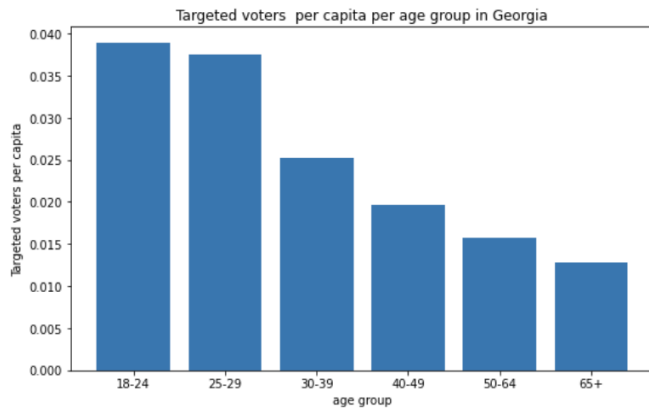


We extracted results for all 18 states observed. Highlighting similar and contrasting results in three interesting states, we see that in Colorado and Arizona, voters between the ages of 18 and 29 were targeted fairly similarly with respect to the total population in each age group. In Arizona, for every 10 individuals from both age groups (18-24 and 25-29) about 7 text messages were sent. While In Colorado, 12 to 14 text messages were sent for every 10 individuals in both age groups. In both states, we also see a steady decline in text messages per capita received for older age groups. In contrast, for Texas, our results show that people above the age of 65 received the most text messages per capita, while the other age groups barely received any with respect to the total population of their age groups.

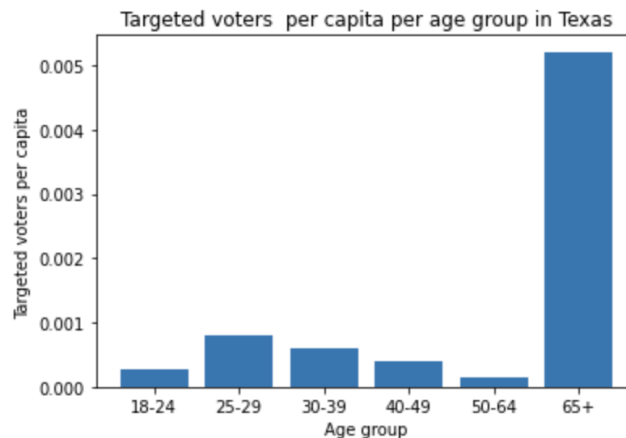


## Targeted voters per age group per capita per state:

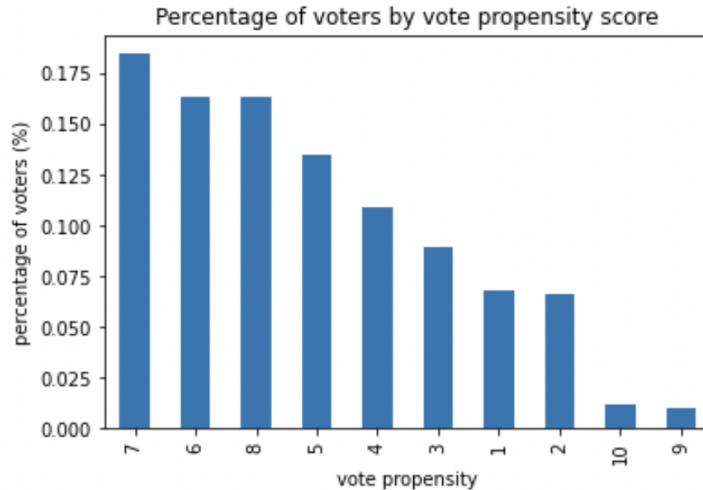
We defined targeted voters per capita per age group per state as a measure of the number of voters targeted in each age group in each of the states observed. We computed this by extracting the total voters targeted in each age group in every state and then extracting the total number of residents in the said age group in that state from the 2019 5-year acs census data. Finally, we divided the total number of voters targeted in each age group in every state by the total number of residents in that age group in the respective state.



We extracted results for all 18 states observed. Highlighting similar and contrasting results from a few interesting states, we see that in Georgia and Iowa, voters between the ages of 18 and 29 and voters between the ages of 25 and 29 were targeted fairly similarly with respect to the total population in the respective age groups. We also see a somewhat steep decline in targeted voters per capita as the ages in the age groups increase. With the aged population being targeted up to 8 times less. In contrast, for Texas, our results show that people above the age of 65 were targeted the most relative to their total population, while voters in the other age groups barely received about 4 times less (on average).



## Vote Propensity Analysis:



The following graph represents the distribution of vote propensity or the likelihood that a person will vote. Most voters received a score of 7, which meant they were on the higher end of the scale and most people targeted are likely to vote in the elections.

### Vote propensity distribution for New American Majority Group:

By applying the following breakdown of voting propensity: Low (1-3), Mid (4-7), High (8-9), Very High (10), we generated graphs of vote propensity distribution (percentage) for each race in the New American Majority Group.

Vote propensity distribution for New American Majority Group				
	Mid	Low	High	Very High
African-American	60.30%	23.53%	15.78%	0.39%
Asian	61.41%	19.09%	18.32%	1.17%
Hispanic	58.97%	23.97%	16.11%	0.95%
Native American	51.31%	31.60%	16.06%	1.02%
White or Unknown	60.62%	20.71%	17.36%	1.31%
percentage in race	100%	60%	30%	0%

All groups have similar patterns in vote propensity as most people belong to the mid propensity (4, 5, 6 and 7) group. However, Native Americans had the highest percentage of low propensity voters among all the groups followed by African-American and Hispanic people.