Alaska's Voter Turnout
PSTAT 135: Big Data - Final Project

by Allester Ramayrat, Kyle Felder, Kyle Kim, & Spencer Zeng

PSTAT 135 - UCSB

Professor Oh Sang Yun

March 22, 2023

Table of Contents:

1. Introduction

Alaska, the 49th state that entered the Union in January 1959, participated in its first presidential election in 1960. It had three electoral votes that year and has maintained that number ever since. Alaska has voted Republican since, only voting Democratic for Lyndon Johnson in 1964.

After seeing its lowest percentage in state voter turnout ever in 2018, Alaska began its new initiative in 2020- the Ranked Choice Voting General Election system - to counteract the decrease in voters. The Ranked Choice Voting General Election system allows Alaskan voters to rank the potential candidates, permitting them to have an opinion on other candidates if their top choice is not elected. In comparison, Alaska is different from other U.S. states in terms of geography, population and demographic, and economy. Let's explore some of these factors:

With approximately 586,000 square miles of land, Alaska is considered the largest state in the union, with approximately 0.91 square miles per individual (State of Alaska 2020). From 2002-2013, Alaska saw a gradual increase in population, but its trend became stagnant and decreased from 2016 to the present day. According to the Alaska Department of Labor and Workforce Development, more individuals are moving out of Alaska compared to those moving in. Though the birth rate is higher than the death rate and emigration rates, Alaska has yet to see a net positive increase in population (Zak 2023). Alaska's demographic profile shows that it is a majority White state at 66.7%, followed by American Indian or Alaskan Native at 14.8%, Asian at 5.4%, and Black at 3.6% (Wiki 2023). In addition, Alaska has had a long history of voter suppression targeted at the Native American community, which will be explored later. Alaska's economic approach has stayed consistent throughout many decades, with oil production, fishing, and tourism being the state's main sources. Oil revenues supply nearly 85% of the state's budget. The economy is also based on other factors, such as federal and state expenditures and research and development due to its unique geographical location, and a small portion is based on agricultural production which is lower compared to other states (Britannica 2021).

We will be taking a look into Alaska's general election in 2018 to analyze voter turnout and the importance of numerous factors such as demographic and commercial data that led to various voting implications and patterns.

2. Problems of Interest

Our problem of interest is which factors influence voter turnout in Alaska. Specifically, we will be looking at the 2018 voter turnout in the general election. We chose this year because it was Alaska's lowest voter turnout in the state's history with only ~49% of registered voters casting a vote. Thus, we would like to look at what is causing this decline by investigating the most important predictors of a person voting. Through our literature review, we've identified two categories of interest that may help us in solving our problem: Demographic Data and Commercial Data.

2.1 Demographic Data

Since it has been found that age and ethnicity have had strong correlations with voter turnout in the past, we decided to investigate this more with the ample demographic data provided. Specifically, we looked at ethnicity, county, gender, and age. As stated above, Alaska's population has plateaued with many people starting to leave the state, thus studying voter registration data can also provide insights into demographic shifts in Alaska. For example, changes in the age, ethnicity, or gender composition of registered voters could have important implications for voter turnout. In Alaska, there has been a history of voter suppression when it comes to Native Alaskans. Redistricting and the lack of voting materials in native languages have diminished the effect of voters of that ethnicity. According to acluak.org (2022), "Of the ballots rejected, about 3 percent were rejected because of ballot defects that could have been corrected if voters had been notified of defects, such as missing witness signatures, voter identifying numbers, or voter signatures. Rejection rates were significantly higher in rural areas, where the percentage of Alaska Native voters is substantially higher." Due to all of this, we think that demographic data could have a strong influence on voter turnout.

2.2 Commercial Data

Secondly, Commercial Data can give us insight into how likely a person is to vote. A person's interests such as their hobbies or commercial products that they purchase as well as a person's contributions made to political, religious, environmental, and many more groups may be useful to determine voter turnout. In addition, information on education level, income, home size, number of children, etc. are also predictors that we are interested in looking into. According to the National Library of Medicine (2021), "The strong correlation between education and voting is among the most robust findings in social science." There is also known to be a correlation between income and voter turnout, so we believe this data regarding one's socioeconomic status and family could also provide valuable insight.

3. Data Descriptions & Methods

The dataset used in the following analysis contains 726 columns and 548,259 rows of information on registered voters in the state of Alaska. The data was collected and put together by the Center for Anti-Racist Research (CAR) at BU and contains geographic, demographic, and household information, voting history, and much more for each registered voter.

Our first step was to figure out which of the 726 columns would be informative predictors of voter turnout. We began our EDA with a general understanding of the columns as well as what information is available from the data set. Of the 726 columns we found that only 80 columns contained 0% *null* values while 296 contained 100% *null* values. Thus, there is a lot of missing data within the 726 columns as a result we should take into consideration the columns that have a small ratio of missing values while dropping columns that give us no information.

3.1 Data Grouping

Individually looking into each of the 726 columns of the data set is expected to require a substantial amount of time; therefore, as a way to more efficiently understand each column of the data, we categorized the data based on the naming convention of the columns. From this we subsetted the data into 8 groups to be analyzed. The groups were:

- Voter Information: The first 65 columns contained data that we classified as Voter Information. These columns contained information that served as a person's identification (i.e. Voter ID, Name, Phone Number, Residence Address, and Mailing Address). These 65 columns made up most of the 80 columns that contained 0% *null* values.
- Voter Description: The next 18 columns contained descriptive information about the voter such as their Age, Gender, Ethnicity, and Party.
- District Data: The 253 columns that followed were related to which district they were a part of. The District appeared to be based on the voter's address as people with similar addresses had the same districts. Some districts included were school, water, fire, and medical districts.
- Commercial Data: The following 45 columns were a voter's commercial data. These columns contained the prefix 'CommercialData_' followed by the name of the column. This included information on the types of magazines they purchase, their property information, estimated household and area income, education, and occupation.
- Commercial Data LL: The next 90 column headers contained the prefix 'CommercialDataLL_' followed by the name of the column. This subgroup of data contained information on what a voter likes to do. The columns contained a boolean value of whether or not a voter donates to certain organizations, is a gun owner, is interested in certain hobbies, or is a pet owner of a certain kind.

- Election Returns: The following 168 columns contained the county and precinct turnout percentages of each political party and candidate per general and primary election from 2008 to 2018.
- Donations: The next 7 columns were voters' information on their donations over the last four election cycles. This included the amount they donated, the number of times they donated, and to who they primarily donated.
- Election Participation: The last 80 columns contained a boolean of whether or not a voter voted in a particular election. It contained the general, primary, and other elections from 2000 to 2018. A column for the 2022 election was present; however, there was no data present in the column.

3.2 Data Assessment

Upon further inspection of each group, we found that many of the columns in the *Voter Information* group may not be useful for the purposes of our analysis. This is because these group columns consisted of strings of names, phone numbers, and addresses; as a result, the ratio of distinct values to the number of rows in most of the columns was relatively high. This means that there would not be a sufficient number of observations per feature; resulting in no meaningful relationships toward the outcome. Since these columns made up the majority of the complete columns, we must note that the columns that are most useful to our analysis are not always the ones with the most observations.

The *District Data* group provided the districts for each resource (school, water, medical, etc.) each voter was a part of. Nearly all of the columns within this group contained all *null*. Despite most of the data being *null* in this group, we were able to find several columns that may be of interest to our analysis. The *County* and *Precinct* column contained the county and precinct that the voter resided in. This may be helpful in our analysis as the values of the Election Returns group are aggregated upon the voter's county and precinct; thus, these columns are a useful identifier for retrieving the election returns per county and precinct. In addition, we can make use of additional data sources such as county population to determine the voter turnout between rural and urban areas. We will use the *County* column as an aggregate for the groups we created as well.

In a similar way to the *District Data* group, the columns of the Commercial *Data LL* group had a large ratio of *null* values, and on average each column within the group contained ~3% of *non-null* values. Since there were no columns of particular interest within this, we opted to drop the *Commercial Data LL* group of columns from the dataset.

The *Donations* group was an interesting set of columns that can be used in a variety of ways. We think this group of columns contains some valuable information that can identify how certain states donate to politicians or organizations as a result of political and marketing

campaigns. However, these columns were not included in our analysis and will be covered further in **Section 6: Future Work**.

At the end of our Data Assessment stage, we identified several relevant groups and specific columns that warranted further analysis. These groups and columns were the *Voter Description, Commercial Data,* and *Election Participation* groups as well as the *County* column from the *District Data* group. We will cover these groups in-depth as well as aggregations of the groups in the next section.

4. Findings

After assessing and pre-processing the data, we conducted exploratory data analysis (EDA) on the selected groups and columns. We split our analysis by Ethnicity, Age, and Income & County Turnout to identify potential patterns in the data.

The *Voter Description* group contained the demographic data information that we were interested in as mentioned in **Section 2.1**. The *Commercial Data* group and the *County* column contained commercial data information that we mentioned in **Section 2.2**. The *Election Participation* group contains the necessary information related to voter turnout. Afterward, we select, fit, and compared model performances. The methodology for preparing our data for model fitting is included at the beginning of **Section 4.4**.

4.1 Ethnicity

One of the features of interest is a voter's ethnic group. The *Voter Description* group contained information on a voter's ethnic description which contained over 100 different ethnicities classifying 90% of the voters. In addition, each of the ethnicities were grouped into 1 of 5 categories which provide an aggregation that provides better readability and representation of the groups.

Our data represented in **Figure 4.1** below indicates the ethnicity makeup of the registered voters in Alaska is 85.91% European (White), 6.64% Hispanic or Portuguese, 2.87% East and South Asian, and 1% African-American. The ethnic makeup of Alaska's registered voters to Alaska's population is not similar

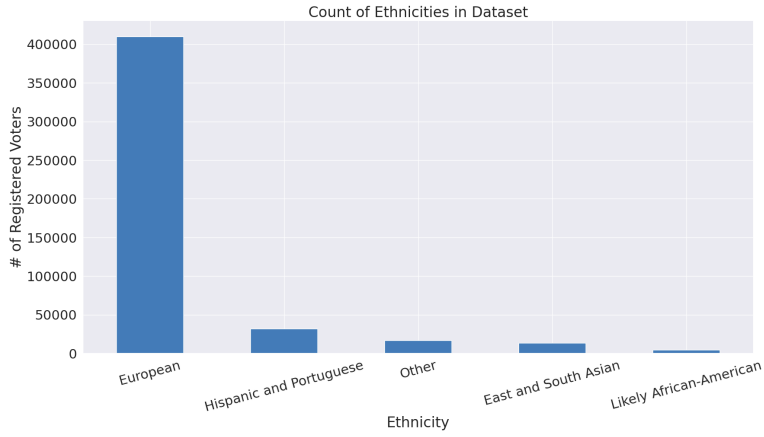

Figure 4.1: # of Voters by Ethnicity

at all. As the Europeans (White) make up 66.7% of Alaska's population as mentioned, Alaska has had a history of voter suppression towards Native Alaskans. Our data indicate that this is not exclusive to Native Alaskans, but other minority groups as well. Thus, we can further investigate this issue by plotting the voter turnout of each ethnic group.

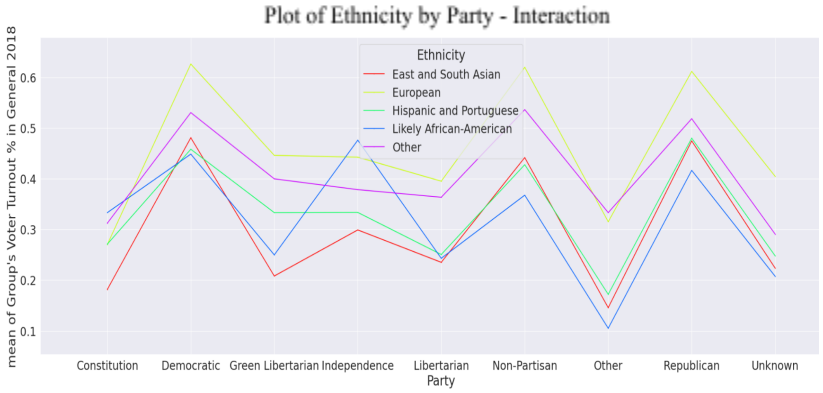

Figure 4.2 shows the voter turnout % of the 5 groups of different ethnicities and their associated parties. Parties with the highest turnouts for each ethnic group typically are Democratic, Non-Partisan, and Republican.

Figure 4.2: Barplot of Political Party aggregated by Ethnicity

According to **Figure 4.3**, Europeans have a voter turnout rate of 51.89% during the 2018 election, while other ethnicities have 39.97%, African Americans have 36.56%, Hispanics and Portuguese have 34.64%, and East and South Asians have 32.79%. The data shows that Europeans show up to the polls 1.3~1.5 times more than minorities.

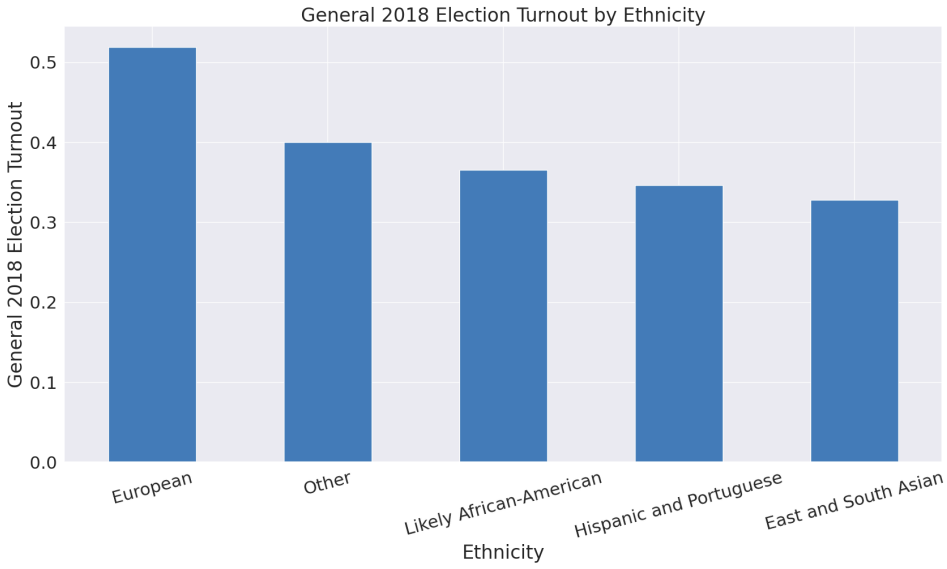

Figure 4.3: Barplot of the 2018 Voter Turnout by Ethnicity

4.2 Age

In the next section, we took a look at the age component of the dataset in relation to voter turnout. **Figure 4.4** shows a histogram of the distribution of ages for those who did vote (blue) and for those who did not vote (orange) in 2018.

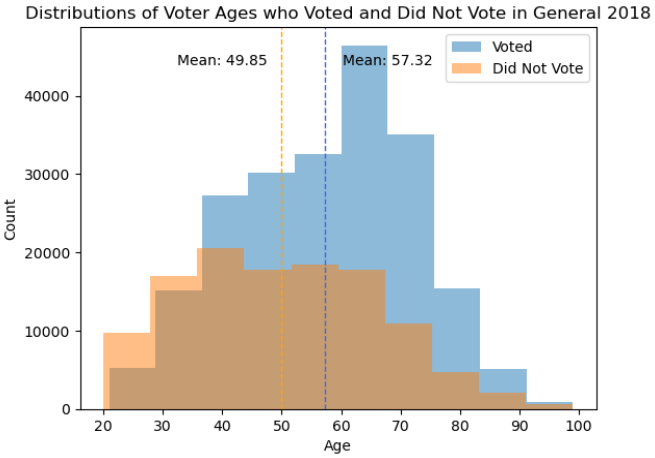

Figure 4.4: Histogram of Number of Voters & Non-Voters by Age

There are more non-voters from the approximate ages of 20 to 35, but ages 40 onward have exponentially more engaged voters than non-voters. The average age of voters who voted in 2018 is roughly 57 years, while the average age of voters who didn't vote is roughly 49 years. As a result, we find that younger people in Alaska are not voting as much as the older population of Alaska.

4.3 County Turnout and Income

We believe that a person's purchases and purchasing power influence whether or not they will turn out to the voting polls. As a result, we are particularly interested in the variables of the *Commercial Data* group as it aligns with the goals of our analysis. In order to understand some of the features of this group, we plotted the distributions and created aggregations to find correlations.

From the *Election Participation* group, we created **Figure 4.5** below that shows the Voter Turnout %'s in each Alaskan county by year - as we can see, there are numerous counties that hit a record low percentage in voter turnout, as many counties seeing a decrease from the previous biannual time interval. This connects us back to our initial research, where we found that Alaska hit a record low of ~49% of voters in 2018.

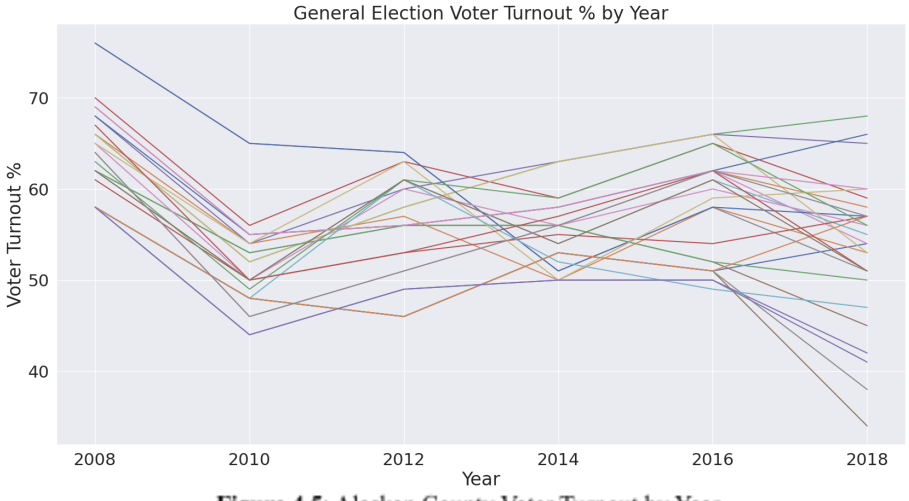

Figure 4.5: Alaskan County Voter Turnout by Year

Inspecting the data visually in **Figure 4.6,** it is clear there is a positive correlation between average household income and voter turnout, however, there may be stronger indicators of voter turnout out there.

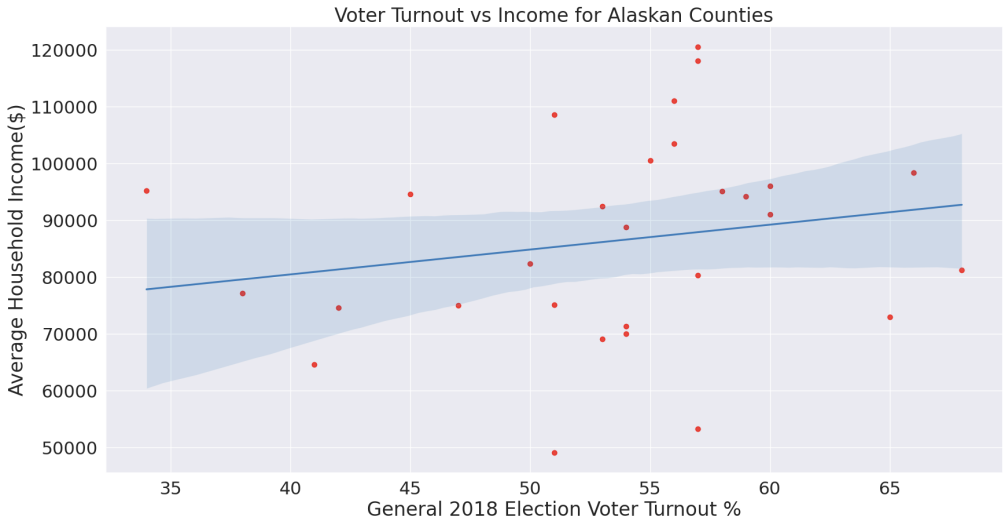

Figure 4.6: Voter Turnout & Income Correlation

Figure 4.7 shows a histogram that tells us that the average household income is ~$90k, and is distributed normally. We then wanted to investigate the correlation between household income level and voter turnout. We opted to do this by looking specifically at the mean income for each county and the percentage of each county that voted.

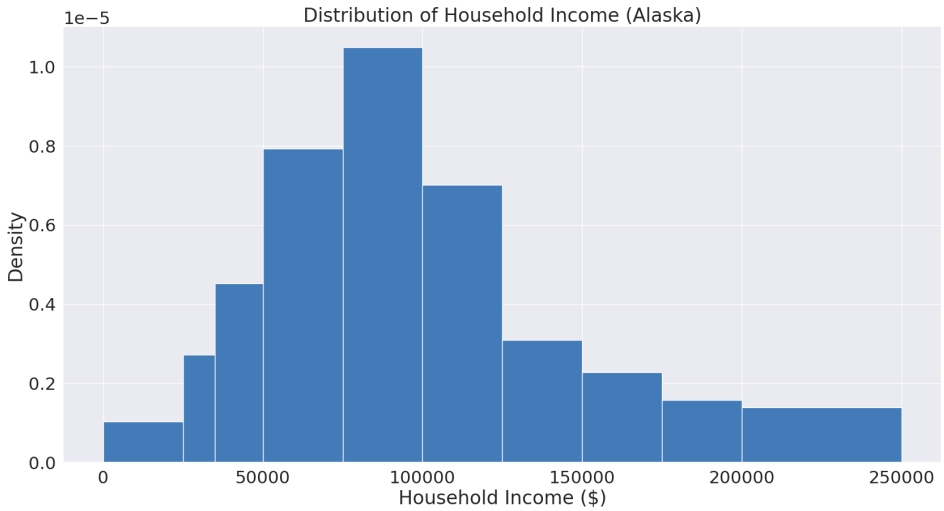

Figure 4.7: Histogram of the Household Income in Alaska

4.4 Models

To begin building a model to predict voter turnout, we started with a Decision Tree model which used Commercial Data and Demographic Data (County, Age, Gender, and Ethnicity). We then fit a Logistic Regression model on the same features for comparison. Then, we fit another Logistic Regression model but only on the Commercial Data. Finally, we found our best result with a Lasso Regression model on just the Demographic Data (County, Age, Gender, Income, and Ethnicity).

To prepare the data for model fitting, the appropriate columns from Commercial Data, Demographic Data, or both are selected. Demographic Data includes County, Age, Gender, and Ethnicity. Duplicate income and house value columns that were found unreliable are dropped. For Commercial Data columns in % or $, the sign is dropped and the data type is changed to integer. For the many Commercial Data columns where null represents a 0 or 'N', null values are replaced with 'N' if the column was of type string, and 0 if the column was of type integer. The meaning of null in each column was verified in metadata prior to transformations. Since we dropped all rows with null values as the final step before model fitting, we found that dropping three columns that had more than 90% null data provided us with much more data to test and train on.

4.4.1 Decision Tree Classifier (Commercial and Demographic Data)

A decision tree classification machine learning model is a predictive algorithm that utilizes a tree-like structure to classify data into different categories based on a set of rules. The algorithm makes decisions based on the features of the input data and follows a path down the decision tree until a leaf node is reached, which represents the final classification. The decision tree is constructed by recursively splitting the data based on the most informative features until the desired level of accuracy is achieved.

```
+----------+-----+--------------------+
|prediction|label|            features|
+----------+-----+--------------------+
|       0.0|  0.0|(6321,[2,6,10,14,...|
|       0.0|  0.0|(6321,[2,6,10,14,...|
|       0.0|  0.0|(6321,[2,6,10,14,...|
|       0.0|  1.0|(6321,[2,6,10,14,...|
|       0.0|  0.0|(6321,[2,6,10,14,...|
+----------+-----+--------------------+
only showing top 5 rows
```

Figure 4.8: Prediction Table from Decision Tree

After the transformations stated above, we have a training dataset with 159,428 rows and a testing dataset with 67,925 rows each with 44 features after performing a 70% and 30% train/test split. Producing a table to evaluate our model's predictions, we found that it correctly classified if a person voted or not 62% of the time. Given that the voter turnout percentage across Alaska was 49% in the general 2018 election, our model is better than a coin flip but we would like to see higher accuracy.

4.4.2 Logistic Regression Model (Commercial and Demographic Data)

A logistic regression machine learning model is a type of algorithm that models the probability of an event occurring based on a set of input features. It is commonly used for binary classification problems where the output is either Yes or No. The algorithm estimates the probability using a logistic function and learns the optimal values of the parameters through an optimization process that minimizes the error between the predicted probability and the actual outcomes of the training data.

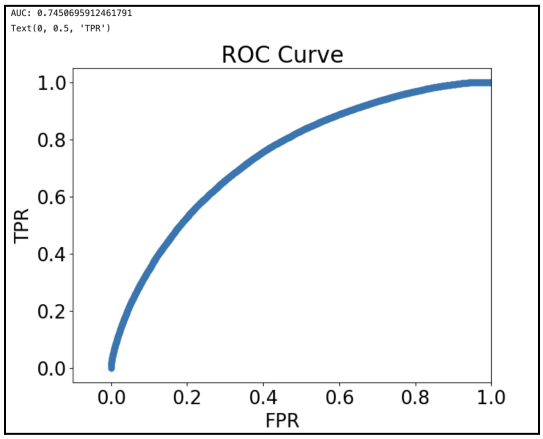

Figure 4.9: ROC Curve on Training Set

Using the same dataset as the decision tree model, we have a training dataset with 159,428 rows and a testing dataset with 67,925 rows each with 44 features after performing a 70% and 30% train/test split. Using the same features as the decision tree model, we found that our logistic regression model had a test AUC of 0.70. While we cannot directly compare accuracy to AUC, the two metrics are positively correlated and indicate a similar level of success. Below, we created a training data ROC curve plot to give us an idea of the true positive rate (TPR) vs false positive rate (FPR). We also created a plot of the beta coefficients and this showed us that out of the 6,000 features created, roughly 5,000 were very close to 0, while 1,000 had absolute values greater than 1 and up to 10.

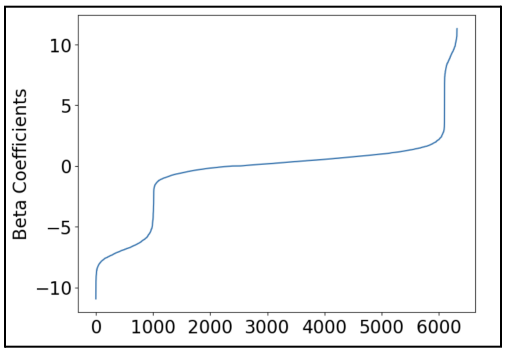

Figure 4.10: Beta Coefficients Plot

4.4.3 Logistic Regression Model (Commercial Data)

While we saw positive results in our models with Commercial and Demographic Data, we thought a large number of features could be causing the model to overfit and thus be less effective. To combat this, we fit another logistic regression model with just the Commercial Data columns. Here we have a training dataset with 281,887 rows and a testing dataset with 120,548 rows each with 40 features after performing a 70% and 30% train/test split.

This model gave us a test AUC of 0.73, outperforming our prior logistic regression

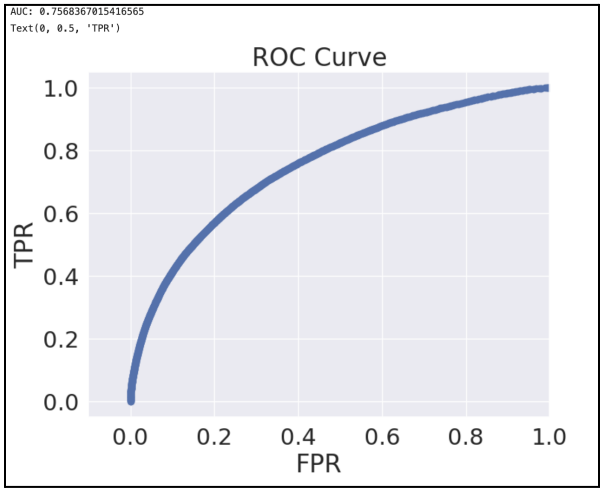

Figure 4.11: ROC Curve on Training Set

model with more features. While this model is our most successful yet, the AUC is very close to the prior model and we see a very similar training ROC curve and beta coefficient plot.

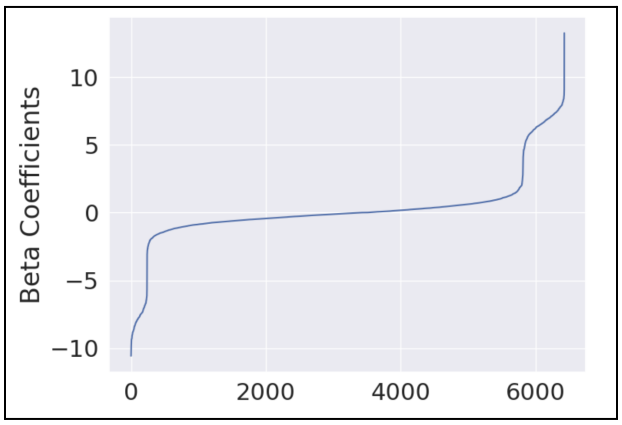

Figure 4.12: Beta Coefficients Plot

4.4.4 Lasso Regression Model (Demographic Data)

The prior three models showed some effectiveness but still produced the wrong prediction ~30% of the time. Given that there are around 40-44 features in those models, it is possible there's some overfitting. To avoid overfitting and improve model accuracy, we decided to do feature selection manually. Based on our exploratory analysis (see **Section 4.1-4.3**), we reduced our predictors to only five, including gender, age, ethnicity, income, and county. To get a better sense of feature importance, we fitted a Lasso regression of the selected predictors on '*General_2018*':

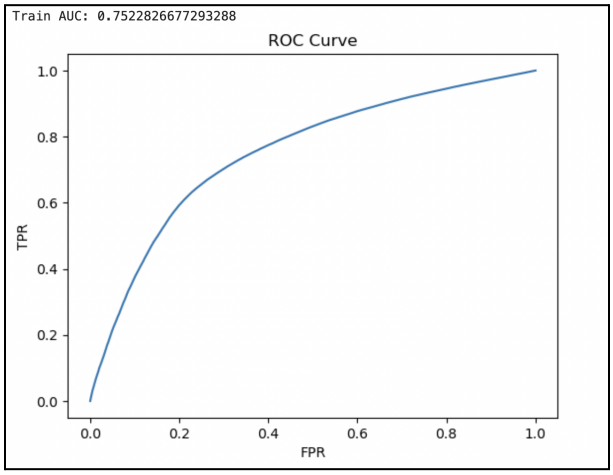

Figure 4.13: ROC curve on Training set

The training AUC and testing AUC for this model are both approximately 0.753, suggesting that our model is not overfitting. In addition, we see improvement in the testing AUC compared with previous models.

	coefs	name
29	0.369955	County_ALEUTIANS WEST
26	0.372337	County_WRANGELL
21	0.380401	County_NORTH SLOPE
17	0.393563	County_PRINCE OF WALES HYDER
22	0.406198	County_NORTHWEST ARCTIC
11	0.406457	County_KETCHIKAN GATEWAY
7	0.411757	County_MATANUSKA SUSITNA
12	0.424746	County_KODIAK ISLAND
25	-0.429618	County_HAINES
8	0.465906	County_FAIRBANKS NORTH STAR

Figure 4.14: LASSO coefficients of county features

We have 188 distinct features for the county column, so we decided to examine them separately from other features. **Figure 4.14** shows the ten highest Lasso coefficients for the county column. We can see County Haines has a relatively high negative association with people not voting.

When examining Lasso coefficients for the rest of the features, Europeans have a relatively high negative association with people not voting, while East and South Asians are positively associated with people not voting. In addition, the effect of income on the response variable is small compared to other selected features (age, ethnicity, gender, and county).

	coefs	name
34	-0.000004	Income
4	0.019620	EthnicGroups_EthnicGroup1Desc_Other
1	-0.031512	Voters_Age
3	0.100564	EthnicGroups_EthnicGroup1Desc_Hispanic and Por...
0	0.189138	Voters_Gender_M
5	0.321383	EthnicGroups_EthnicGroup1Desc_East and South A...
2	-0.364184	EthnicGroups_EthnicGroup1Desc_European

Figure 4.15: LASSO coefficients of features other than counties

5. Discussion

The Alaska dataset offers ample demographic and commercial data that could provide valuable insights into voter turnout in Alaska. Demographic factors such as age, ethnicity, gender, county, and socioeconomic status have been found to have strong correlations with voter turnout in the past. Furthermore, Alaska's population plateau and history of voter suppression towards Native Alaskans make it crucial to investigate how demographic shifts and voting patterns correlate. On the other hand, the commercial data on personal interests, magazine subscriptions, and contributions to various groups could also give valuable insights into how different interests and socioeconomic factors correlate with voter turnout. Overall, the data mentioned above could provide useful information on how to improve voter turnout and address voter suppression issues in Alaska.

In **Section 4.1** - Ethnicity, we found that the voter registration demographics do not properly represent Alaska's population. Voter suppression is not unique to Native Alaskans but affects all minority groups within Alaska as minority groups are severely underrepresented. In addition to being underrepresented in the voter registration, but also have significantly lower voter turnout as the European ethnicity group has a voter turnout 1.3~1.5x larger than minority ethnic groups.

As a result of our exploratory data analysis, we found correlations between voter turnout and many columns. Specifically, looking at ethnicity, Figure 4.3 illustrates how people of European ethnicity had a voter turnout of ~55% while all other ethnicities were below 40%. This indicated that ethnicity may be a very informative feature for our models. We also found (**Figure 4.4**) that the mean age of voters was ~57 years old while the mean age of non-voters was ~49 years old. This is a large difference and could also be informative when utilized in models. Additionally, we also found a positive correlation between voter turnout and county (**Figure 4.6**). Intuitively, people in population-dense areas will likely have an easier time voting so we believed this would also be a strong predictor. The last column we investigated was income which was also shown to have a small positive correlation with voter turnout when

grouped by county (**Figure 4.7**). As a result of our exploratory analysis, we felt as if we had identified key predictors of voter turnout.

We first fitted a decision tree model of all the commercial and demographic factors on the response variable ('*General_2018*') and obtained 62% accuracy on the testing set. In addition, we fit a Logistic Regression model on the same predictors and response which yielded an AUC of 0.70 on the testing dataset. This result is not ideal, so we decided to improve our model by shrinking the number of predictors to avoid overfitting and improve model accuracy. Among all the columns of Commercial Data, we only examined income in our exploratory analysis. Hence, We decided to fit a Logistic Regression Model on just the Commercial Data to investigate those columns' strengths as predictors. The testing AUC on this model is around 0.73, slightly higher than the previous two. The fact that a smaller number of predictors led to an improved AUC prompted us to try fitting a model with manually selected features. In particular, gender, age, ethnicity, income, and county were selected because our exploratory analysis suggested that these features were closely related to voter turnout, and then Lasso regression was used to better understand feature importance. The testing AUC of this model is around 0.75, which is the highest one among all the models. Looking at the Lasso coefficients, we were able to grasp which features were most and least effective in practice.

To determine which features were most effective in predicting voter turnout we primarily analyzed the Lasso Regression model (**Section 4.4.4**). Using this model, we were able to obtain our highest AUC yet of 0.75 and investigate the coefficient sizes for each feature. Surprisingly, the coefficient for income was reduced to essentially zero, indicating that it was not useful in terms of predicting voter turnout. While it initially seemed like it would be a powerful predictor, it, along with all the other Commercial Data columns was found to not be very helpful in our models relative to other predictors. The coefficient for age was also reduced to near 0 and is not useful in our model. Gender was another predictor we looked at and when used in our Lasso Regression model, it was found to be more influential than income and age, but not as powerful as some of the other predictors we looked at. The Lasso Regression model coefficients reinforce our initial beliefs about ethnicity being a very strong predictor of voter turnout, especially when a person was European or East/South Asian. Due to the variation in voter turnout from each ethnic group, it was able to provide extremely valuable insight into if a person voted or not. Similarly, the coefficients for *County* were the highest of all in our Lasso Regression model. Specifically, being from the county Kodiak Island and North Star proved to be the strongest predictor that someone would vote while being from Haines was the strongest predictor someone would not vote.

In summary, by using different features in different models we were able to find which features were the most effective. Once narrowed down to a small list with just county, age, gender, income, and ethnicity, we were able to create a Lasso Regression model that produced

coefficients indicating how strong each predictor was. We found that *Age* and *Income* were relatively useless, *Gender* had some influence, and *Ethnicity* and *County* were the most effective predictors of voter turnout.

6. Future Work

While we have looked at hundreds of thousands of rows of data with hundreds of different columns, there is still a lot of potential for future work.

As mentioned in **Section 3.2** the *Donations* group of columns was interesting and would offer great insight if given the chance to be looked into further. It would be helpful to create aggregations of donation data among the county, income levels, and ethnic groups as it would inform us of the participation of voters outside of the polls. Participation outside of the polls may be a useful predictor in determining political alignment and voter turnout. It would be interesting to explore how voters' contributions toward certain politicians and organizations affect how and if they vote. We expect that we will find a positive correlation between donating and voting in the election as well as the politician or organization that they primarily contribute to will share similar political alignments as the voter.

Our decision tree performance can be enhanced by utilizing the powerful technique of pruning, which reduces our overfitting and enhances their interpretability and reliability. Through the process of pruning, we can simplify the tree structure and prevent irrelevant training data patterns. This improves the generalization of the tree and results in higher accuracy when the pruning is well-balanced to avoid overfitting using cross-validation.

In general, for all of our models, we've had low accuracy with the Decision Tree, Logistic, and Lasso Regression. This may be due to the response variable '*General_2018*' that we have selected. As identified in **Section 4.3** the '*General_2018*' column had the record lowest voter turnout rate. This anomaly may not be able to be properly explained by the data set alone and may require supplemental data to explain the reason for the occurrence of a low voter turnout rate. If we were to perform our regression on the other Election Years where the voter turnout was balanced, we may experience a higher accuracy in our models.

In addition, Alaska is a relatively unique state in terms of its size, population, and location in the world. While this is part of the reason we chose to look into Alaska in the first place, investigating if these results are similar across other states would also give us a good idea of whether our findings are unique to Alaska or hold across the rest of the USA. Due to the huge amount of election data available, the number of ways to model the data, and the importance of findings regarding voter turnout, there is so much potential future work that can be done.

7. Bibliography

Alaska Department of Education & Early Development. "About Alaska's Geography." Alaska.gov, State of Alaska, 2020, https://alaska.gov/Kids/learn/aboutgeography.htm.

American Civil Liberties Union of Alaska. "After State of Alaska Fails to Address Voter Disenfranchisement Issues, Civil Rights Groups File Suit." ACLU of Alaska, 23 Feb. 2022, https://www.acluak.org/en/news/after-state-alaska-fails-address-voter-disenfranchisement-issues-civil-rights-groups-file-suit.

Britannica, The Editors of Encyclopaedia. "Alaska - Economy." Encyclopædia Britannica, Encyclopædia Britannica, Inc., 18 Feb. 2021, https://www.britannica.com/place/Alaska/Economy.

National Center for Biotechnology Information. "The Correlation between Education and Voter Turnout in Alaska." Journal of Public Health Management and Practice, vol. 28, no. 3, 2021, pp. E1-E8, doi: 10.1097/PHH.0000000000001344, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8685665/#:~:text=The%20strong%20correlation%20between%20education,associated%20with%20higher%20voter%20turnout.

Wikipedia contributors. "Demographics of Alaska." Wikipedia, Wikimedia Foundation, 19 Mar. 2023, https://en.wikipedia.org/wiki/Demographics_of_Alaska#:~:text=White%3A%2066.7%25.

Zak, Kyle. "Alaska's Population Rose Slightly in 2022, but More People Continue to Leave than Arrive." Anchorage Daily News, 16 Jan. 2023, https://www.adn.com/alaska-news/2023/01/16/alaskas-population-rose-slightly-in-2022-but-more-people-continue-to-leave-than-arrive/.