Q2_report_Tego

TegoChang

10/22/2021

Summary

This analysis is to investigate the potential factors, especially the demographic ones, that could affect the turnout rate of the US 2020 election in North Carolina. During the process, data cleaning, aggregation, and merging have been performed to combine the registered and history data in the same variables. Exploratory data analysis has then proceeded and several signs were found as a reference for the following modeling process. The finalized hierarchical model helps us dive into how age is impacting the turnout rate in the election, and sort of explains why the Republicans outperform the Democrats in certain counties.

Introduction

It is widely acknowledged that the voting behaviors in the election are related to several factors, e.g., demographic ones, locations, or the party you tend to vote for. Our purpose is to construct a hierarchical model to answer the following questions:

- How did different demographic subgroups vote in the 2020 general elections?
- Did the odds of voting differ by county? Which counties differ the most?
- How did the turnout rate differ between males and females for different parties?
- How did the turnout rate differ among age groups for different parties?

Conclusion

The answers to our questions of interest are:

- 1) The odds of turnout for males is considered 1.06 times compared with that for females, which is six percent higher. However, it is not statistically significant. 2) The odds of turnout for age group 41 to 65 is considered the highest, which is 2.7 times compared with that for age group 18 to 25. 3) On race, when Asian is set as the baseline, the odds of turnout for the pacific islander, undesignated, and white have the higher odds of turnout. However, there are just too few observations for the pacific islanders. 4) On ethnicity, when the hispan and latino category is set as the baseline, the non-hispan and non-latino category has a higher odds of turnout.
- The odds of voting seem to differ by county in 2020 from our EDA, so we added varying intercept by county in our model. Further, we also observed an obvious trend change when evaluating if there is an interaction between party and county. Thus, varying slope by county has been added for the party predictors. In this case, the varying intercept in our model is 0, but if we select a party, say Democrat, Wake county seems to have the highest odds of turnout, while Randolph and Mcdowell have the lowest ones.

- Among the interaction between sex and party, males who vote for Democrats and unaffiliated are considered statistically significant. The odds of turnout for males to vote for Democrats is 1.47 times, increasing by forty-seven percent, compared with females to vote for CST.
- Among the interaction between age and party, there're many combinations that are statistically significant. One of the interpretations for these combinations could be that the odds of turnout for age group 26 to 40 who vote for Democrats is 2.03 times, increasing by one hundred and 3 percent, compared with the age group 18 to 25 who vote for CST.

Besides the above questions of interest, we further dive into the voter's age and investigate which age group has a higher impact on the turnout between Democrats and republicans. In the interaction between age and party, the 26 to 40 age group is the only statistically significant one compared with the other two age groups, with age group 18 to 25 as a baseline. Then, we further calculate the odds of turnout between the Democrats and Republicans at this age group, the result turns out that Republicans seems higher than Democrats, with odds of turn out is 2.41 for the former and 2.05 for the latter when the baseline is the age group of 18 to 25 voting for CST.

Limitations

During EDA, we find out that race could have an interaction with party. Further, applying a varying slope by county to race could also be a feasible option. However, due to too few observations for part of the categories in race, e.g., Pacific Islander, we did not put both into action. This could be considered as future work to further explore the analysis.