**Project Prototype**

**CS216**

**Understanding Voter Turnout for Different Geographic Areas in National Elections**

Aileen Niu (an280), David Boham (dob13), Matthew Lu (ml689), Jerry Liu (jml208), Mich Lulo (mpl28)

Date: November 4th, 2024

**Part 1: Introduction and Research Questions (15 points)**

The election is tomorrow! Many questions surrounding voting and its related statistics have come up in discussion. Elections are a crucial topic and affect the lives of all citizens, so we deemed this topic more than sufficiently important to explore.

Due to the importance of elections and the recent rise of concern for bias, we hope to analyze the larger causes of voter turnout differences. We will evaluate the following questions:

1. How does voter turnout differ per geographical area on the national level depending on a variety of socioeconomic factors such as race, salary, age, and the type of election? ← we will use voter registration data to identify how many elections someone participated in, then identify which socioeconomic factors might be good predictors for turnout
2. How have these factors become more or less relevant in the past several years, and what are their long-term implications? ←identify relationships and changes between
3. What strategies can we implement to increase voter turnout in areas where there is little voter action?

Our research is **substantial** because it addresses a serious question about making voting accessible to the entire country. By analyzing specific factors and trends across geographic regions, we can predict the likelihood of different factors to affect turnout. In addition, elections are a very important topic to the government, making voting data available through many censuses. For this reason, our project is **feasible**. As we enter the upcoming election, it’s good to help make problems related to voting transparent to the public. The more people are aware, the more we can combat issues and make elections ‘fair’ across all states. The January 6th insurrection of the Capitol combined with growing concern over bias in votes makes this research **relevant.**

Changes from proposal: Our topic introduction and research questions have not changed much from our proposal. We still plan to look at elections

**Part 2: Data Sources (15 points)**

Part 3: What Modules Are You Using? (15 points)

1. **Module 3: Visualization**

Looking for trends and patterns in our data will be essential for answering our research question. To do this, we must make data visualizations. We will create different visualizations such as line plots to visualize how certain demographic factors affect voter turnout. Histograms and barplots will also be used to compare demographic factors and voter turnout rates.

1. **Module 4: Data Wrangling**

To prepare our data for visualizations and analysis, we will use key ideas from the data wrangling module. This process starts with data cleaning, which includes handling missing values by either imputing them based on logical estimates or removing rows that have excessive gaps, depending on the context. Next, standardizing data formats is essential, such as ensuring consistent labels for demographic categories like age groups or race. Additionally, it is crucial to detect and correct data entry errors, such as inconsistent spellings or numerical values entered incorrectly. After cleaning, data aggregation might be needed to group data by categories, such as by geographic region or socioeconomic status, allowing for the examination of patterns and trends in turnout rates. Filtering the data to focus on relevant subsets, such as specific election years or voter age brackets, is another key step that must be done.

1. **Module 6: Combining Data**

Once the data is cleaned and properly formatted, merging it with supplementary datasets—such as census data—can provide further context and enable more granular insights into the relationship between demographic variables and voter turnout. This approach to data combining ensures that the dataset is ready for deeper statistical analysis, visualizations, and interpretation. We may find related data such as voter registration data for further data analysis.

1. **Module 7: Statistical Inference**

Through statistical inference methods such as t-tests or we can test for differences in mean turnout rates between groups, while regression analysis can help determine the strength and direction of the relationship between demographic factors (e.g., income, education level) and turnout. Additionally, confidence intervals can be used to quantify the uncertainty around estimated statistics, providing a range that likely captures the true population parameter. Through these methods, statistical inference allows us to not only detect patterns but also quantify their significance, ensuring that the insights derived from the data are robust and generalizable.

1. **Module 8: Prediction & Supervised Machine Learning**

This module equips us with the tools of linear regression, model prediction, and basic dataset training operations. Predictive tools can be very helpful for both extrapolation and interpo

Part 4: Preliminary Results and Methods (15 points)

Part 5: Reflection and Next Steps( 10 points)