# Introduction

In this project, we aim to explore the dynamics of electoral results by using various data sources (listed in subsequent sections) to predict outcomes in congressional elections and presidential elections within NY-02 congressional district. Our primary focus will be on identifying winners at different levels, analyzing the number and percentage of votes received by each candidate, and evaluating voter turnout, which reflects the percentage of eligible voters who participate in the electoral process.

By examining historical data alongside current trends, we hope to discover new patterns that influence voter behavior and decision-making. This analysis will allow us to see how factors such as demographics, previous election results, and campaign strategies impact electoral outcomes.

In addition to identifying winners, we will investigate voter turnout rates. High turnout often indicates strong democratic engagement, suggesting that people feel motivated and empowered to vote. Conversely, lower turnout may highlight issues such as disenfranchisement or apathy, raising questions about the effectiveness of outreach and education efforts aimed at voters.

Our project will use a variety of data sources, including past election results, demographic information, and voter registration statistics, to build a comprehensive understanding of the electoral process. We will employ data preparation and preprocessing techniques which will be crucial in ensuring the accuracy and reliability of our analyses. This will involve cleaning the data to remove inconsistencies and errors, transforming raw data into suitable formats for analysis, exploration and visualization.

As we progress, we will discuss our methodologies and findings in depth, allowing for a richer understanding of the complex factors influencing electoral outcomes. Through this project, we hope to shed light on the importance of informed voting and the critical role it plays in shaping our democracy.

# Project Approach

1. **Data Collection**:  
   We will begin by gathering historical data related to electoral outcomes, demographic information, and voter behavior within the NY-02 congressional district.
2. **Data Selection**:  
   From the collected data, we will select the most relevant datasets that will provide insights into election outcomes and voter patterns.
3. **Data Cleaning**:  
   The selected datasets will undergo a cleaning process to address inconsistencies, remove duplicates, and fill in any missing values. This step is crucial to ensure the accuracy and reliability of our analysis.
4. **Data Exploration**:  
   Once the data is clean, we will explore it to identify various patterns and trends related to election outcomes. This will involve visualizations and statistical analyses to gain insights into voter behavior and demographic influences.
5. **Pattern Identification**:  
   Through exploration, we will attempt to uncover relationships and patterns that may affect election results, such as correlations between demographic factors and voting behavior.
6. **Data Drilling**:  
   After identifying initial patterns, we will drill down into the data for a more granular analysis. This may involve examining specific precincts or demographic segments within the NY-02 district.
7. **Spatial Analysis**:  
   In the spatial analysis phase, we will union different datasets to identify relevant patterns in election outcomes. By integrating various sources of data, we can uncover spatial relationships and geographic trends that influence voter behavior and electoral results. This will help us visualize how location affects voting patterns within the district.

This structured approach will enable us to systematically analyze electoral dynamics in NY-02, ultimately leading to more informed predictions and insights regarding future elections.

# Data Sources

We have utilized the following data sources to gather relevant information for our project.

**Redistricting Data Hub**: This resource provides comprehensive information about congressional district boundaries and demographics. **Link**: [Redistricting Data Hub - New York](https://redistrictingdatahub.org/state/new-york/)

**NY Race 2022 (Shapefile)**:  
The NY Race 2022 dataset is a shapefile containing detailed electoral results from the 2022. It includes race-wise population data, providing a comprehensive view of voter demographics across different regions. This dataset is instrumental for analyzing voting patterns and candidate performance, particularly in the NY-02 district. Its geographical format allows for in-depth spatial analysis, making it easier to identify trends and correlations in electoral outcomes. **Dataset**: ny\_race\_2022\_t\_shp

**NY 2022 General Precinct Data**:  
This dataset provides detailed precinct-level election results, enabling us to assess voter preferences and the distribution of votes across different areas within NY-02. Analyzing precinct-level data will allow us to identify trends and variations in voting behavior within the district. **Dataset**: ny\_2022\_gen\_prec

**New York State GIS Data**:  
This resource offers geographic information about congressional districts, allowing for effective visualization of demographic and electoral data. It will help us analyze how geographical factors influence voting patterns in NY-02.**Link**: [New York State GIS Data - Congressional Districts](https://data.gis.ny.gov/datasets/nys-congressional-districts/explore?filters=eyJEaXN0cmljdCI6WzIsMl19)

**U.S. Census Bureau - My Congressional District**:  
This tool provides detailed demographic and socioeconomic data specific to congressional districts, including NY-02. It will help us understand the population characteristics, income levels, and other relevant metrics that may influence voting behavior in the district. **Link**: [Census Bureau - NY-02](https://www.census.gov/mycd/?st=36&cd=02)

**Congressman Andrew Garbarino's Official Website**:  
This site offers insights into the NY-02 district, including its demographics, community resources, and key issues affecting residents. Understanding these local factors will be crucial in analyzing voter behavior and election outcomes in the district. **Link**: [About NY-02 District](https://garbarino.house.gov/about/about-district)

**Data USA - Congressional District 2, NY**:  
This platform provides a comprehensive overview of the NY-02 congressional district, including demographic data, economic indicators, and education statistics. This information will be valuable in understanding the socio-economic context that may influence voter behavior in the district. **Link**: [Data USA - Congressional District 2, NY](https://datausa.io/profile/geo/congressional-district-2-ny)

By leveraging these data sources, we aim to create a comprehensive understanding of the electoral landscape in NY-02, facilitating informed predictions about future elections and enhancing discussions on civic engagement and representation in our democracy.

# EDA

# Data Selection

* In our analysis of the “***ny\_2022\_gen\_prec***” dataset, we focused specifically on the GCON02 columns, which detail the election results for the U.S. House in Congressional District 2. The selected columns for this analysis include **GCON02DGOR**, **GCON02OWRI**, and **GCON02RGAR**. We used this dataset for 2 main reasons-

**Geographical Context**: The analysis of this dataset allows for mapping precinct boundaries effectively, showing election results at precinct level. Precinct-level results in relation to their geographical boundaries will help us in identifying and analyzing electoral trends and patterns.

**Precinct-Level Election Results**:

**GCON02DGOR (Jackie Gordon - Democrat)**: This column provides the vote totals for Jackie Gordon, the Democratic candidate.

**GCON02OWRI (Write-In Votes)**: This column captures any write-in votes cast in the election for Congressional District 2.

**GCON02RGAR (Andrew R. Garbarino - Republican)**: This column details the vote totals for Andrew R. Garbarino, the Republican candidate.

* Similarly, from our “***New York State GIS Data***” dataset we have –

Utilized it to create accurate visualizations of **Congressional District boundaries**, enhancing the understanding of geographical contexts in electoral analysis.

We have used it to **identify the winners in each district**, allowing for a clearer picture of electoral outcomes across New York.

* From our “***ny\_race\_2022\_t\_shp***” dataset

We have extracted demographic information which we used to map the data with the GEOID of the precinct data. This integration allows for a clearer understanding of the racial composition within specific precincts.

By utilizing the race data, we identified race-wise population distributions in Congressional District 2. This analysis helps reveal the demographic diversity of the district, which is essential for understanding voter behavior and electoral dynamics.