**Introduction**

While the presidential election season is in full swing, we decided to explore polling data sources that exist online. There are several individual sources that could be found online; however, the website RealClear Politics is a location that gathers, summarizes, and presents the results of the various polls in one location. It should be noted that, while this website is good of a summary view, the underlying polling data must be extracted from the various polling sources (if available) for further review and analysis. The polling sources include Emerson College, The Economist Magazine, The New York Times/Sienna College, CBS News, and many others. Some sources are free, while others incur a fee. It should be noted that the polls tend to discriminate between “Registered Voters” (RV) and “Likely Voters” (LV), and the common belief is the LV are better more indicative of election results. However, a Berkley Haas Study in 2020 reported that while the polls reached a 95% confidence level for statistical reporting, the actual election results only matched with the polls 60% of the time.

**Data Sources**

We are currently in discussion to identify the data sources for analysis, and the type of analysis we wish to discuss. The sources are varied and include tables on websites, attached PDF documents, and CSV files. Some will require us prepare the data through another platform before we are able to evaluate and analyze the data. This also needs to include a matching/pairing of questions and response on polls to insure equivalency of the questions. Data that has currently been identified include The New York Times/Sienna, Roanoke College, and Emerson College Polls.

**Collaboration Platform**

**Entity-Relationship (ER) Diagram**

1. Identify the Purpose
   1. Analysis of polling data
2. Identify entities - once you have identified these entities, add them in rectangles.
   1. People
      1. Tasks
   2. Sources
      1. Polling Data
   3. Platforms
      1. Relational database
   4. Functional Operations
      1. Data Identification
      2. Importing
      3. Tidy/Transformation
      4. Analytics
   5. Deliverables
      1. Presentation
      2. Analysis
3. Identify relationships - how are these entities related?
4. Identify attributes - what are the key attributes of the defined entities?
5. Database Design

Cited Works

1. https://www.realclearpolitics.com/epolls/2024/president/2024GeneralElection.html#!
2. Kotak, A., & Moore, D. A. (2020). Public Election Polls are 95% Confident but only 60% Accurate.