Jeremy Yeung, Amy Jung, Joyce Li, and Noor Gill
Project 2 Design Document
29th July 2021
MIDS W200

Project 2:

Github Repository: Project2_Li_Jung_Gill_Yeung
Link: https://github.com/UC-Berkeley-I-School/Project2_Li_Jung_Gill_Yeung

Primary Dataset: Political TV Ad Archive AIRINGS ON TV
Link: http://politicaladarchive.org/data/

1. **Initial Plots, Figures, or Tables:**
   ○ Distribution of ads in swing states; ad topics; sponsor types (bar chart)
   ○ Percentage of candidate ads in swing states; whether ads were from the news or non-news channels (pie chart)
   ○ Start times/end times - What was most popular among candidates? (histogram)
   ○ Correlations between variables (scatterplot)
   ○ Number of ads relative to the population (heatmap)
2. **Variables and Insights:**
   ○ Network (TV channel on which the ad aired), Location (city and state aired), Program-Type ("News" or "not news," representing the type of TV programming), Start-Time (Date/time ad aired, start), End-Time (Date/time ad aired, end), Sponsor-Types (Candidate committee), Subjects (Subjects covered in the ad), Candidates (Candidate(s) named in the ad), Type (Campaign ad, issue ad, unknown, input by Internet Archive researchers - most are campaign), Message (Pro, con, mixed)
      ■ These can help us complete analysis on swing states and split out different types of ads for different candidates and analyze them
3. **Supplemental Datasets:**
   ○ US Census Data: Voting and Registration in the Election of November 2016 → to determine the efficacy of ads by analyzing voter and voter registration data that corresponds to each region/district
   → Link: https://www.census.gov/data/tables/time-series/demo/voting-and-registration/p20-580.html
   ○ Table 4a: Reported Voting and Registration, for States: November 2016 → Join by state name
   → Link:

- Potentially look into other voter data such as age, race, and sex (Tables 4b-4c) for further analysis.

4. **Final Report Organization:**
   - Research Question: Are political ads effective in swing states? (Is there an association between the number of ads displayed in a swing state and the percentage of voter registration?)
   →Sub-Question: Are political ads in swing states different (longer run times, more ads, more pro/con)?
   →Sub-question: In regions where political ads are effective, are pro ads for the candidate the individual voted for more or less effective compared to the cons ads for the candidate the individual did not vote for?
   - In terms of the project itself, start by data cleaning, EDA, deriving subsets of tables to reveal patterns, and visualizing different attributes.
   - In terms of the final report, we will present the data cleaning and sanity checks first. After, we will present each sub-question and then the analysis and visualizations that answer the question. At the end, we will combine the results and answer the main research question.