2020 US Presidential Elections Campaigns Donations Analysis

It uses two datasets. A partial dataset for developing and the full dataset. A zipped version of the full dataset expands to a 3GB file. The full dataset can be accessed on AWS at s3://rw-cs696-data/P00000001-ALL.csv. The format of the files is described at the end of the document. The partial dataset has two differences from the original. The differences are described at the end of this document.

The data contains information about each donation made to presidential candidates in 2020. The source of the data is:

https://www.fec.gov/data/candidates/president/presidential-map/ Questions

1. How many donations did each candidate have?

2. What was the total amount donated to each candidate?

3. How many unique contributors did each candidate have?

4. What mean and standard deviation of the donations for each candidate.

5. What percentage of each campaign’s donations was done by small contributors, that is donations under $50?

6. Produce a histogram of the donations for the Trump and Biden campaign? The x-axis the amount of the donation and the y-axis the number of donors that gave that amount.

I used AWS Spark to answer the first 5 questions. For the 6th question, I processed the data on AWS and download the result using Python plotting tools to produce the histograms.

Data Format

Note that some(all?) of the rows in the data set contain an extra column.

Columns

cmte_id - ID of the committee that received the donation. Example: C00285254 cand_id - Candidate ID. Example: "P00013649"

cand_nm - Candidate name. Name is quoted as it contains a coma. Example: "Sanford, Mar- shall"

2020 US Presidential Elections Campaigns Donations Analysis

contbr_nm - Contributor name. Name is quoted as it contains a coma. Example: "KEITHLEY, BRAD"

contbr_city - Contributor city. Example: "ORANGE BEACH"
contbr_st - Contributor state. Example: "AK"
contbr_zip - Contributor zip. Example: "99501"
contbr_employer - Contributor employer. Example: "BONAPARTE FILMS LLC"
contbr_occupation - Contributor occupation. Example: "CONSULTANT"
contb_receipt_amt - Contributed amount. Example: 1000

contb_receipt_dt - Date contributed. Example: 09-SEP-19

receipt_desc - Often blank. Example: "" memo_cd - Often blank. Example: "" memo_text - Often blank. Example: "" form_tp - Example: "SA17A"

file_num - a unique number assigned to a report and all its associated transactions. Example: "1376946"

tran_id - Example: "AFBC1B0EF531D4CDCBE8"

election_tp - This code indicates the election for which the contribution was made. EYYYY (election plus election year). Options are: (P)rimary, (G)eneral, (O)ther, (C)onvention, (R)unoff, (S)pecial, or (R)ecount. Example: "P2020"

A slightly longer description of the columns can be found at: http://www2.stat.duke.edu/~cr173/ Sta102_Sp16/Lab/lab9.html

Partial Dataset Difference

In the original data all text data entires are quoted. In the partial dataset only the text data en- tires that contain a comma character (,) are quoted.