

Project Overview:

This project explores U.S. presidential election donations from December 2019 to October 2020, specifically for Joe Biden and Donald Trump. The goal is to understand the relationship between political donations and election outcomes. By analyzing how contributions were distributed across different states, professions, and donor demographics, the project sheds light on the broader implications of political donations on election results.

The objective was not only to assess the monetary support for the 2020 U.S. presidential candidates but also to understand the broader impact of contributions on the final election outcome. This analysis required a thorough, step-by-step process of data collection, cleaning, transformation, and visualization using SQL in BigQuery and Tableau.

Data Collection Challenges:

The dataset for this project was massive, spanning millions of rows from December 2019 to October 2020. As a result of its size, I had to download the data in **48 separate files**, each under 300 MB, as the platform did not allow larger downloads in one go. This approach introduced additional complexity, as it required consolidating all files into a single dataset for analysis.

The data, sourced from the Federal Election Commission (FEC), included individual contributions for both Biden and Trump, along with details such as contributor city, state, occupation, and the amount donated.

Data Preparation and Cleaning:

Once all 48 files were successfully downloaded, I began the process of data cleaning and preparation in Python and **BigQuery**. Here's a breakdown of the key steps:

All of the datasets initially had **78 columns** each, but many of them were not necessary for the analysis. Using **Python**, I reduced the number of columns to just **16 important columns** that were relevant to the project. These included key fields like `contributor_city`, `contribution_amount`, `contributor_state`, and `transaction_date`, among others.

Although I could have performed this column selection step in **SQL**, I chose to do it in Python to further involve Python in the project. Python provided flexibility in filtering and managing the datasets efficiently. By doing this, I ensured that only the most important data was retained, making the analysis both cleaner and faster.

This step was crucial for streamlining the large dataset and focusing only on the essential information for further analysis and visualization. Then I started more cleaning process in BigQuery and Here's a breakdown of the key steps:

1. Merging Datasets:

- The first challenge was to append all the data tables. Using SQL, I merged all 48 datasets into a single master table, named “**all_month_2019_2020**,” which contained over **14.4 million rows**.
- This step alone was critical in ensuring that the data was ready for analysis.

2. Handling Missing Values:

- Several columns, including **contributor_city**, **contributor_state**, **contributor_employer**, and **contributor_occupation**, had missing data.
- For example, there were **57,300 missing values** in the **contributor_city** column and **99,108 missing values** in the **contributor_employer** column. To maintain the integrity of the analysis, I filled these missing values with either “Unknown” or “Out_of_US,” depending on the context.
- I also noticed significant missing values for **contributor_id** (over 7 million), where I replaced null values with “Unknown.”

3. Primary Key Creation:

- To ensure uniqueness and remove any duplicates, I created a **composite primary key** by merging the **transaction_id** and **contribution_receipt_date** columns.
- This step allowed me to identify and remove **3,140 duplicate rows**, ensuring accurate results for further analysis.

4. Cleaning Up Remaining Null Values:

- After primary cleaning, I handled the remaining null values in columns such as **contribution_receipt_amount** and **fec_election_type_desc**. By addressing these issues, I ensured the dataset was clean, accurate, and ready for further exploration.

5. Exporting Data:

- Once the dataset was prepared, it was exported as a CSV file for use in **Tableau**, where deeper insights would be visualized.

Visualization (Tableau):

In Tableau, the cleaned and prepared data was visualized to bring the story to life:

1. Top Committees:

A bar chart analysis of the top 10 committees receiving the most donations, revealing the largest financial contributors to political campaigns and highlighting the disproportionate influence of certain committees.

2. Geographic Insights:

A map-based visualization showed the total donation collection by state. This included:

- **Top 10 states** with the highest donation amounts.
- **Bottom 10 states** with the lowest donation amounts. This geographic breakdown provided a clear visual representation of financial engagement across the U.S.

3. Biden vs. Trump:

A detailed analysis of Biden and Trump's donations over time:

- **Monthly donation trends** revealed when each candidate gained the most financial support.
- A comparison of **average donation amounts** showed how much, on average, supporters contributed to each candidate.
- Analysis of **donation sources** (individuals vs. organizations) helped illustrate the different financial backers behind each campaign.
- The impact of donations on **election results** by state was also visualized, demonstrating the relationship between financial contributions and voting outcomes.

4. Occupational Analysis:

A bar chart displaying the top professions contributing to Biden and Trump. This analysis highlighted which industries were more inclined to support each candidate, showing key trends and biases in professional financial support for the 2020 election.

Insights and Findings:

1. Relationship Between Donations and Election Results

- **By State:**
 - The analysis revealed a **strong correlation** between the **percentage of donations collected** and the **actual election results** in various states.
 - States where candidates received more donations often reflected higher support in the actual election, showing how financial contributions potentially influenced voter behavior and outcomes.
- **Key Observation:**
 - Higher donation percentages in key states often aligned with the final election results, indicating that financial backing was a strong predictor of success.

2. Total Contributions

The total contribution amounts reveal a significant financial disparity between Joe Biden and Donald Trump. Biden raised **\$763.9 million** in total, while Trump secured **\$474.9 million**. This large gap in fundraising may reflect differing levels of financial support from their respective donor bases. Additionally,

the average donation per transaction for Biden was **\$124**, compared to Trump's **\$89**. This suggests that Biden's contributors were willing to make larger individual donations, indicating stronger or more affluent financial backing.

This difference in the average transaction size could be attributed to several factors, such as differing campaign strategies, the socio-economic profiles of the donor bases, or the enthusiasm levels of supporters at key moments in the campaign. The financial advantage Biden had may have played a role in his ability to fund more campaign ads, outreach efforts, and mobilization efforts, potentially influencing the election's outcome.

3. State-wise Donation Distribution

When breaking down donations **by state**, California stands out as the top contributor, with a total of **\$226 million** donated to the election campaigns. Washington, D.C., followed as the second-highest contributor with **\$119 million**. Notably, no other state surpassed the \$100 million mark, showcasing a stark financial concentration in just two regions.

These figures indicate that the financial dynamics of elections are concentrated in a few key regions, which may have a disproportionate influence on political campaigns. The correlation between donations and election outcomes in these states also suggests that financial support could reflect voter behavior, making campaign funding a potential indicator of electoral success in certain states.

4. Contributions by Donor Demographics

A surprising finding in the analysis was the large volume of contributions coming from **retired individuals** (\$233 million) and **unemployed persons** (\$227 million). This might seem counterintuitive at first, but it can be understood when considering the strong political engagement among these groups. Retired individuals, in particular, often have more time to focus on political issues and may feel a greater sense of responsibility to shape the future through financial contributions.

Unemployed individuals contributing such a large amount might be reflective of a particularly engaged or motivated demographic, possibly due to the economic and political issues at stake in the 2020 election.

5. Donations by Occupation and Profession

- **Top Professions:**
 - Individuals from top companies and various professional sectors contributed heavily, with a noticeable bias towards certain candidates.

- **Individual contributions** made up **89.82%** of total donations, while **organizations** contributed **7.10%**.
- **Key Observation:**
 - The overwhelming majority of donations came from individuals, which underscores the importance of personal contributions in U.S. elections, as opposed to corporate or organizational funding.

6. Geographic Donation Trends

At the city level, Washington, D.C. contributed \$119 million, making it the top city for donations in the country. This isn't surprising since it's the political center of the U.S., with many political organizations, lobbyists, and government workers living there. Because it's the hub of federal government activity, the city is home to many people and groups who are heavily involved in politics, leading to a high amount of donations.

Arlington, VA, which is just across the river from Washington, D.C., contributed \$49 million, making it the second-highest city for donations, though much lower than Washington, D.C. Arlington's close location to the capital explains why so many people who are politically active, like professionals, consultants, and lobbyists, live and work there, which boosts the city's involvement in the election through donations.

7. Monthly Donation Trends

- **Biden vs. Trump:**
 - Biden saw significant spikes in contributions at key points in his campaign, showing a **sudden surge in contributions** during crucial months.
 - Trump's donation trends were more consistent but lower in comparison to Biden's.
- **Key Observation:**
 - Biden's campaign benefited from well-timed spikes in contributions, possibly reflecting campaign strategies or pivotal moments during the election that galvanized donor support.