The project summary for Individual Project-Intro to Data Analytics

1. **A clear statement of the topic/purpose of the project, and for whom the results of the project are intended.**

Topic: US 2020 Election Campaign Fundraising Data Analysis

Objective: Collect US Election 2020 funds data to summarize various parties and candidates' fund collection amounts and expenditure. The tasks I have performed are:

- Determine the 10 candidates who collected the most amount of funds.
- Summaries and comparisons between two major political party candidates, Joe Biden's and Donald Trump's finances.
- Determine and plot the 10-election committees lists who raised most from the individual contributor.
- Comparison of 50 States fundraised amount from August 30th, 2020, to September 8th, 2020, and analyze the trend.

Targeted audience:

- Individuals working in the field of political finance and analysis
- Various political fundraisers, committees, and affiliations
- Major political parties and candidates planning for future campaign
- News organization and finance analyst
- Students who are studying political science or who would like to be in a field with politics.

2. **A description of the data used for the project**
    a. Who collected the data?

The data is collected by The US Federal Election Commission (FEC). I have found similar data from Kaggle but I couldn't use the data as it was collected in the

month of May 2020. I wanted to use a more recent time data set, so I have downloaded the bulk raw data provided in the FEC and processed it for exploration.

b. How the data was collected

Campaign committees, candidates, political parties, committees, political action committees, and other filters submit election-related data to the Federal Election Commission in paper and electronic filing. FEC website provides access for the raw data download to everyone. Besides they have an API for data collection for the users.

Data set Links:

https://www.fec.gov/data/browse-data/?tab=bulk-data

Data set 1: All candidates

Data set 2: Contributions by Individuals (itcont_2020_20200830_20200908)

Data set 3: Committee master

c. What data was collected (what variables were recorded, who/what was the data collected from – people? places? etc.)

FEC has collected data across all the states, counties, cities, and towns. It collects data from various political parties, candidates, affiliates, business organizations, and individual contributors. The data set contains transaction information about the fundraising and spending of election campaigns. A few of the variables are:

Candidate Information: Candidate name, Incumbent status, Location, Party affiliation, Total amount raised, Donor information, Type, Expenditure, Transaction ID, etc.

Contributor Information: Committee name, ID, Type of committee, Donation amount, Party affiliation, Location details, Date of transaction, etc.

d. Why the data was collected

The Federal Election Commission is the independent regulatory agency deals with administering and enforcing the federal campaign finance law to protect the integrity of the federal campaign finance process by providing transparency.

e. When the data was collected

FEC was established in 1975.FEC website has data available from 1978 to up to these days. The data sets I have downloaded are dated from 2019 to 2020. I have downloaded my datasets in November 2020.

3. Methods

a. How was the data prepared for analysis?

As the downloaded data set was a .text file without the column's name, the first step was to read raw data and integrate them with the column's name to convert it into a data frame. One of the data sets, Individual Contribution contains day-to-day transaction data of individual donation amounts from January 2020 to up to date. As it's a very big data set, I had to load only 10 days of transaction data containing almost 2 million rows of data.

i. Were data sets joined together?

I have used three data sets. I had to join them a few times to create a master file that contains all the necessary variables required. I have used the left join for the data set merging.

ii. What variables were included in the analysis

Candidate Information: Candidate name, Incumbent status, Location, Party affiliation, Total amount raised, Donor information, Type, Expenditure, Transaction ID, Date of transactions, etc.

Individual Contributor Information: Name, ID, Type of committee, or affiliation they belong to, Donation amount, Party affiliation, Location details, Date of transaction, Occupation, Employer, etc.

Committee Information: Committee ID, Name, Types, Location, Transaction amount, Party Affiliation, Candidate affiliation, etc.

b. An outline of how the data was analyzed

i. **Dropping unnecessary Variables:** All the data sets contained 20 to 30 variables. For my analysis, all the variables weren't necessary, so I have selected only the required data variable column for further process.

ii. **Drop missing (NA) values and White spaces:** Data sets contain lots of missing values (NA), White spaces, and negative transaction amounts. Using filtering, drop_na, and order to handle negative, missing values and white spaces accordingly.

iii. **Remove duplicate values:** There were many duplicated values that existed like many transactions from the same donor. Finding the unique donor ID and adding all transactions from the same ID was used to handle the duplication issue.

iv. **Datatype Correction:** Most of the variables were character types that needed to change into factor types. Besides the Transaction Date, I convert the character type and integer type of data variable into Date type.

v. **Removing special character in a String:** There is a comma (,) in between the candidate's last name and last name. Using text manipulation, I removed the comma and fixed the name order properly.

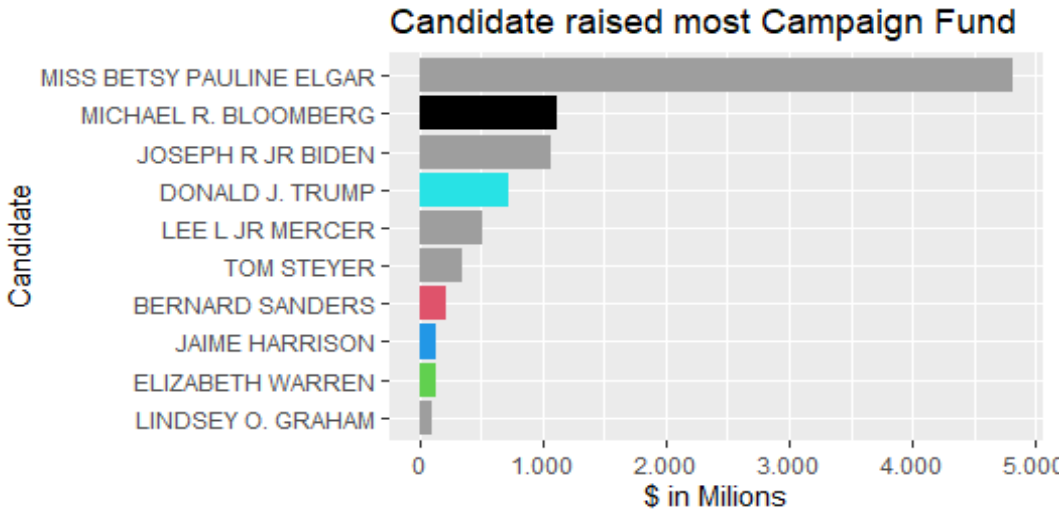
vi. **Formatting Dates:** One of the major issues was dealing with Date values. The date value was in integer type containing only integer numbers for example "8192020". As it's only a 7-digit number, r commands couldn't read the data in Date format. I had to use "str_split_fixed" and "gsub" to introduce commas (,) in-between month, days, and year values so the reader could read data like the date format.

vii. **Filtering:** Most of the data modeling methods involved Filtering in accordance with "group_by" to get the results.

viii. **Plotting:** For plotting, I have used a Bar chart and an Interactive US state map. have saved the Interactive US state Maps in HTML format for submission.

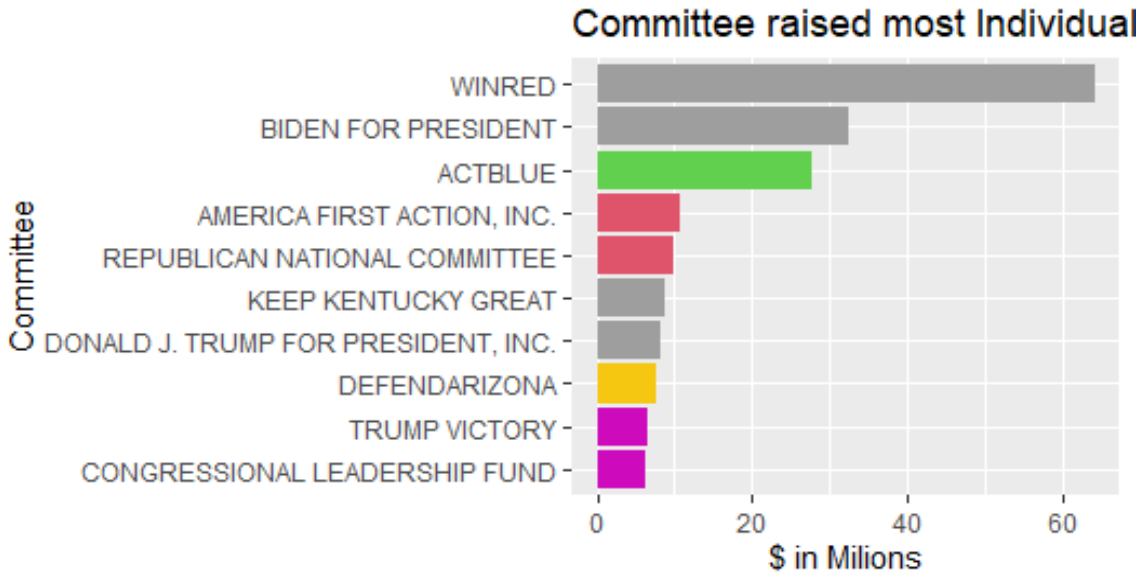
4. **Results of the analysis – include any relevant tables/charts**

1. **<u>10 Candidates who raised most Campaign Fund in 2020 Election</u>**



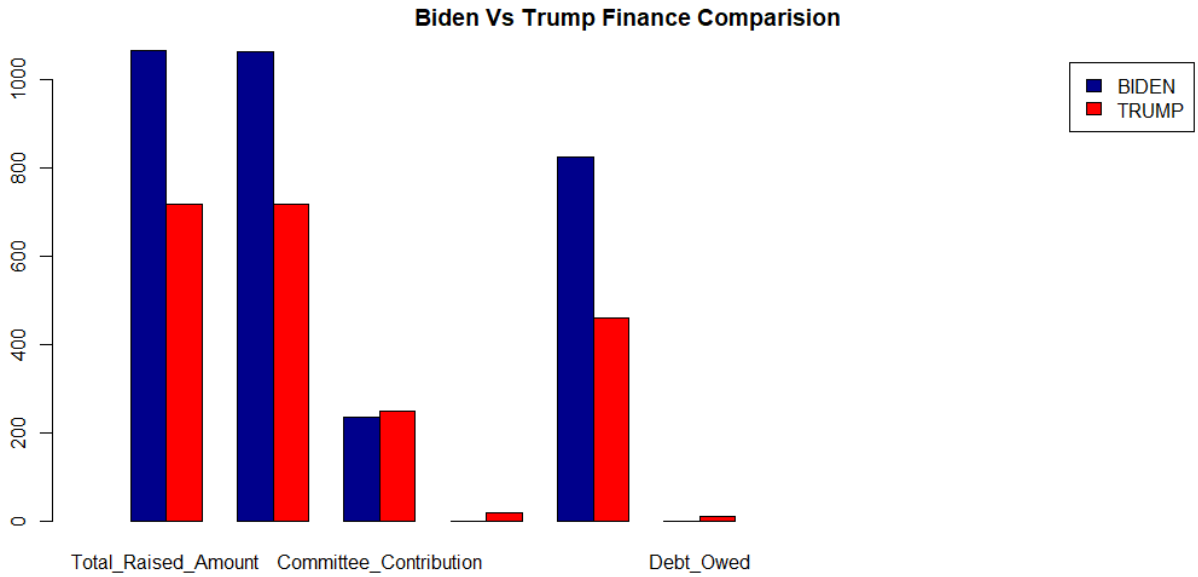
Insight: According to the plot, Miss Betsy Elgar had raised the most funding amount of more than $4 Billion in the 2020 election which I think is incorrect. Besides Michael Bloomberg, Joe Biden and President Trump had raised around a billion-dollar that seems reasonable.

2. **<u>10 Committees That Raised Most Fund from Individual Contribution:</u>**



Insights: WINRED is an independent committee that raised the most fund in the time period August 30th to September 8th this year. Besides, President-elect Joe Biden's Campaign raised more than 40 Million which was more than the Republican National Committee and President Trump campaign fund together.

3. Biden Vs Trump Finance Comparisons for 2020 Election



Candidate	Total Raised Amount (m)	Total Spent Amount (m)	Committee Contribution (m)	Cash on Hand left (m)	Individual Contribution (m)	Debt Owed By (m)
Joe Biden	1064.61	1063.05	235.5	1.56	823.1	0
Donald Trump	7177.3	718.17	249.72	18.41	459.41	11.33

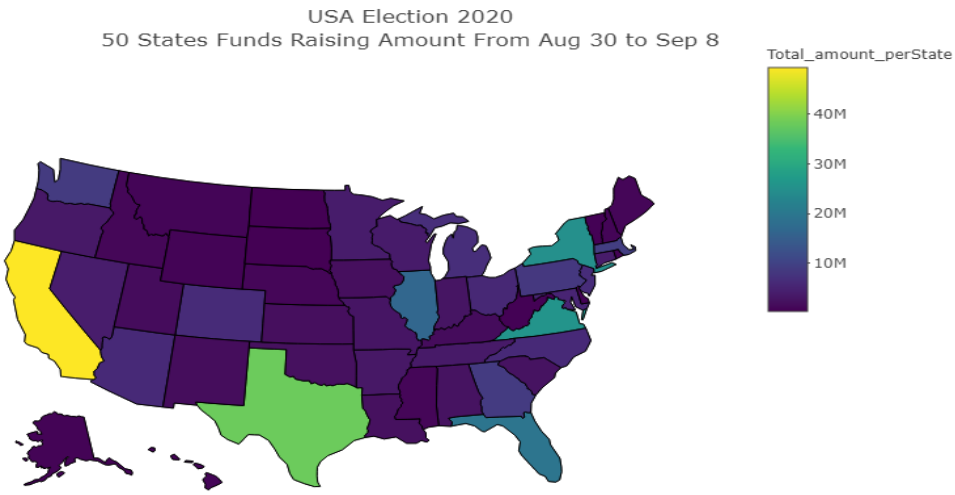
Insights: Joe Biden raised and spent more campaign funds in the year 2020 than president Trump. Besides Joe Biden raised more funds from Individual

Contribution than the Political Committee contribution that is true as historically we saw the Democratic party raise more money from small grass root donors. On the other hand, President Trump and the Republican party did better in the committee donors. The Donald Trump Campaign spent more than the amount they raised. Part of that extra money came from the loan as the campaign debt was 11.3 million.

4. **USA Election 2020 States-wise Funds Raising Amount from Aug 30th,2020 to Sep 8th,2020**

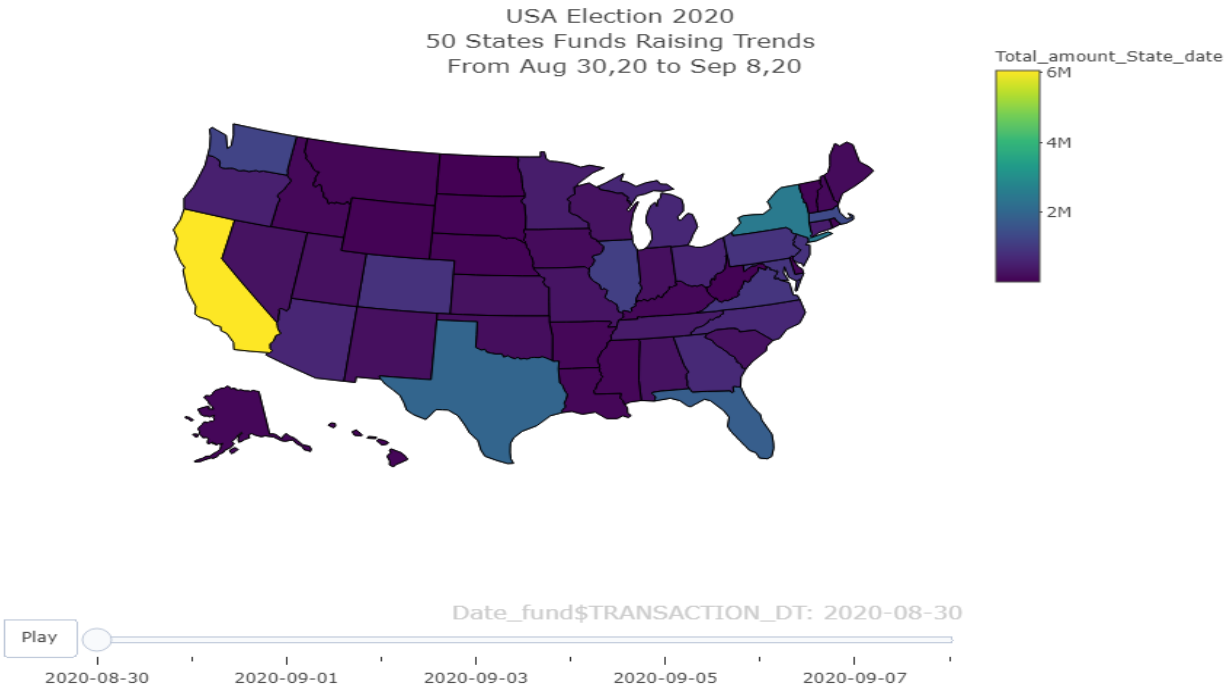
States	Fund Raised Amount (Million)
California - CA	49.33
Texas-TX	38.04
Virginia- VA	25.65
New York-NY	24.91
Florida-FL	19.36
Illinois -IL	16.53
Massachusetts-MA	9.17
Georgia-GA	8.67
Washington-WA	8.62
Pennsylvania-PA	8.16

Insights: California and Texas raised more amounts than other States. Considering both states' size, population size, and economy it seems very reasonable and correct.



file:///C:/Users/user/Desktop/3372/Ind%20project/US_Statewise_fund.html

5. **<u>USA Election 2020 States-wise Funds Raising Trends from Aug 30th,2020 to Sep 8th,2020</u>**



file:///C:/Users/user/Desktop/3372/Ind%20project/Day_Day_states_funding_plot.html

Insights:

a. From August 30th, 2020 to September 8th, 2020, most of the days California (CA) was the top state that raised the most funds. On the day of august 31st, CA raised most funding of 15.1M where on September 7th 1.86M which was the lowest amount but still top in comparison with other states.
b. Texas was the 2nd position in raised Individual funds. It overtook CA September 3rd and tied with Ca on August 31st
c. Most states raised the most amount of donation on August 31st and the least amount on September 7th. August 31st was the last day of the month and also Monday, the 1st day of the week when most people have their Payroll money in hand. With money in hand, individual donors most likely spent more than other days.
d. On September 2nd, Virginia raised the most amount of money 17.02 M among all other states for the given 10 days data period. I recognize it as an exceptional occurrence or outlier as among the 10 days period,8 days VA even couldn't raise a million.

5. Discussion/Conclusion

a. Discuss the results – what do the results reveal about the selected topic, how might the results be used by the intended audience?

According to my observation, major credible organizations do a similar kind of analysis and chart representation for Election Campaign Finance Analysis. Considering this fact, I think my task selections and results for this project are consistent with my topic and intentions.

This analysis could be used by future campaigns to get insights on the Campaigning method, find out the target audience and committee, and make a plan ahead. Besides, the research could use similar approaches to determine the trends and changes among demographics. For example, now social media has become a very popular medium to reach the target audience than 10 years ago. In 2020, both major political parties raised a record number of the amount compared

to their previous generation despite the pandemic. More analytical research is necessary to discover the trends and changes.

b. Strengths/weaknesses of the data/analysis

Strength: I enjoyed working on this project as I have learned many new functions to deal with data, especially the Interactive US-State Plot mapping. As I haven't done the Data Visualization courses yet, it was quite hard for me to figure out the process. I am motivated to work on various other projects to learn new and fun content.

Weakness: I think my code is not efficient as I lack knowledge of many functions that could make me write down fewer syntaxes. Due to this issue, many times I have to create more variables. Besides, my knowledge of US politics is limited, so my insights might be wrong.

c. Future analyses based on the project

i. K-nn clustering to segment data
ii. Prediction on US 2022 Election results based on Fundraising and Expenditure data.