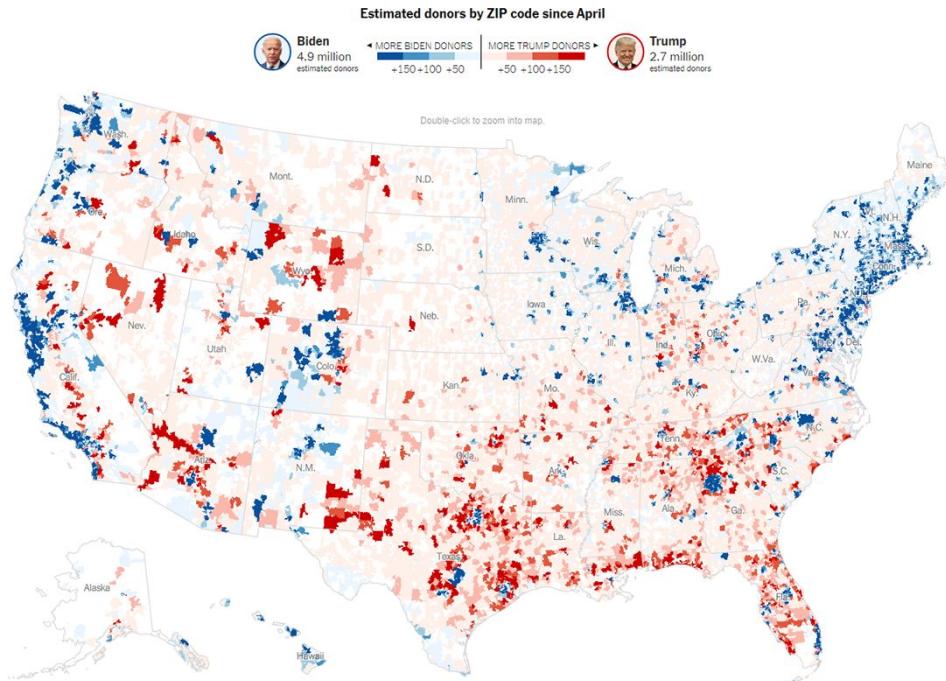


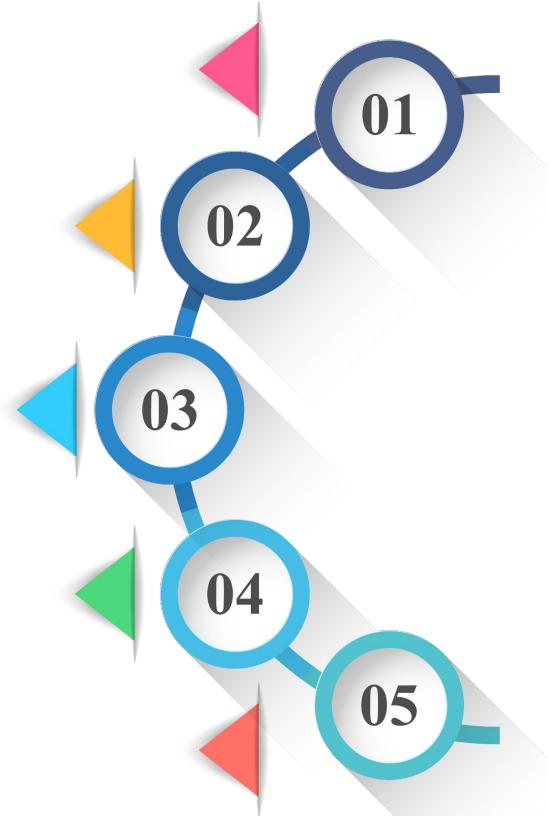


US Election Campaign Finance Analytics

Data 606 - Capstone : Ozgur Ozturk, PhD



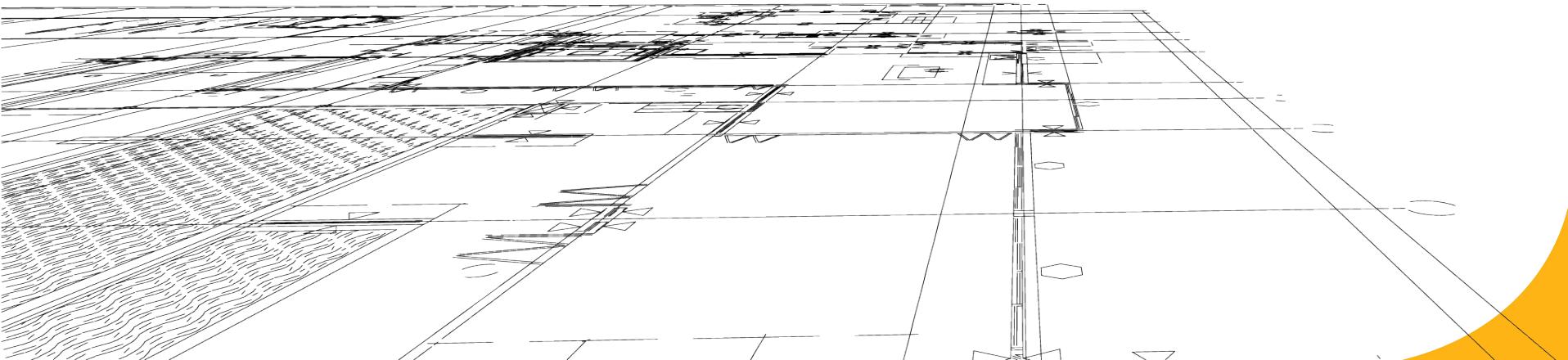
By
Harshitha Yentrapragada
Shivani Badinehal
Udveg Jukanti



1. Introduction and Abstract.
2. Data Source and Methodology.
3. Software Tools and Technologies
4. Analysis of Election Campaign using visualizations.
5. Model Prediction
6. Tableau Dashboards
7. Project Goal
8. Conclusion
9. Future Scope
10. References



Introduction and Abstract

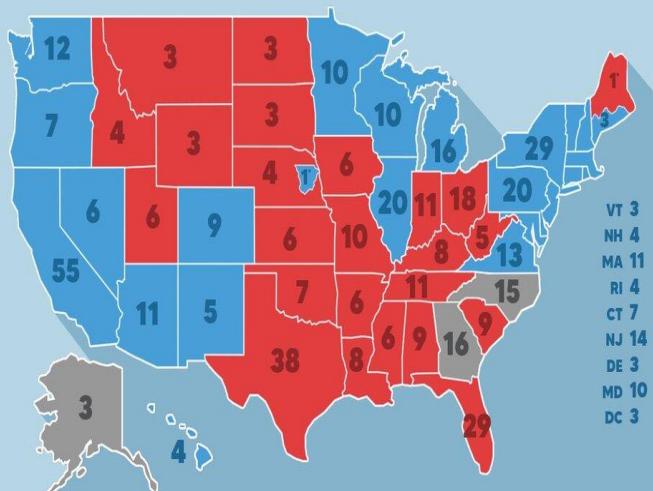


Introduction



According to conventional thinking, the federal government has a negligible impact on American political campaigns and elections. A closer study indicates that the federal government has also steadily grown its involvement in campaigns and elections over the past 50 years, even if states still hold most of the responsibility for election administration. One element that impacts the local market and the global economy is the US presidential election. Researchers need to focus more on political events and how they affect the growth of competitive local markets as well as how they affect the global economy. Given the importance of the US Presidential Election and its potential to have an impact on the world economy, a significant number of academics and politicians in the US have made efforts to forecast the results in order to develop policies based on the results.

Abstract



- Elections and campaigns are two distinct processes that are governed differently. With the exception of campaign funding, the federal government has comparatively limited control over political campaigns. The main election administration assistance provided by the federal government is to states. The main objectives of current election campaigns and policies are voter turnout and electioneering. The federal law that regulates campaigns and elections may not always apply to subjects like issue advocacy or general attempts to sway public opinion.
- The main objectives are permissible and prohibited sources of contributions and expenditures, Contribution limits, prohibition on fundraising, permissible and prohibited uses of campaign funds, public disclosure of contributions and expenditures.

Data Source And Methodology

Data Source

Retrieved from :

US Campaign Finance Data – Federal Election Commission(FEC)

URL:

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

Description:

- The size of the dataset is approximately 15 gb and has almost 28 million people information.
- This dataset includes info of candidates, committees, Political action committees, House and Senate campaigns, as well as transactions, individual and committee contributions, operations and independent expenditures for US elections.



Data Source - Files Used



Download bulk data

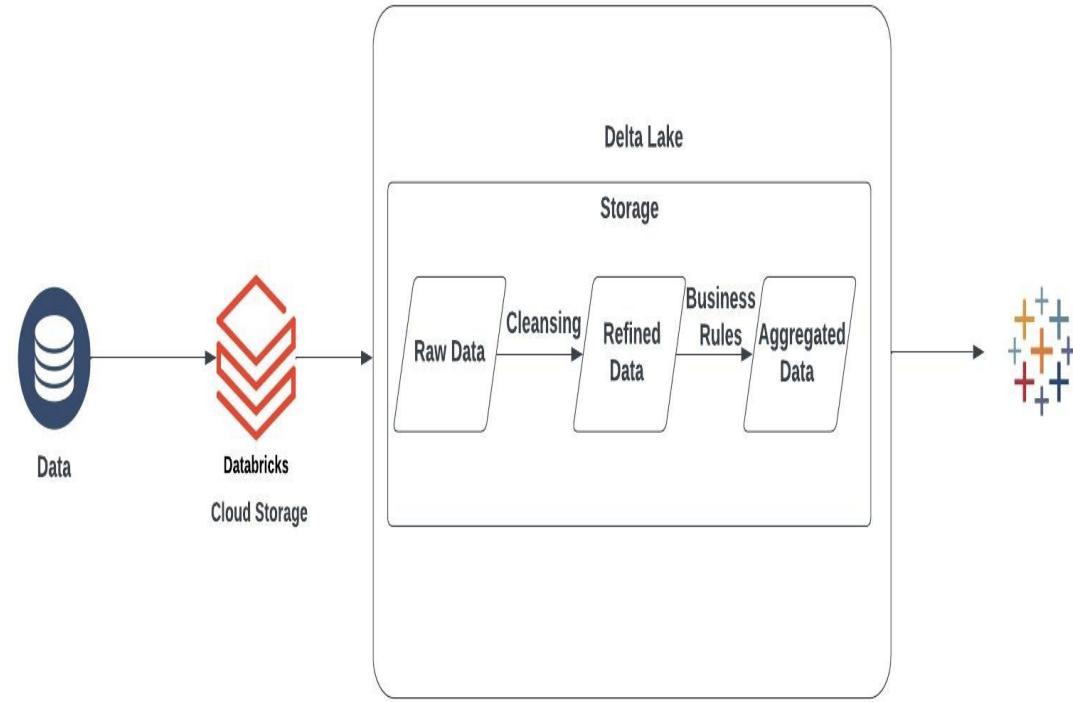
Downloadable bulk data files contain data from statements and reports filed with the Commission in a form that may be useful to users performing in-depth campaign finance research. The files, which were previously located on the Commission's file transfer protocol (FTP) server, can be very large because they contain transaction-level data. The update schedule of these files varies from daily to weekly. Expand each file's accordion to read more about what is contained in the file.

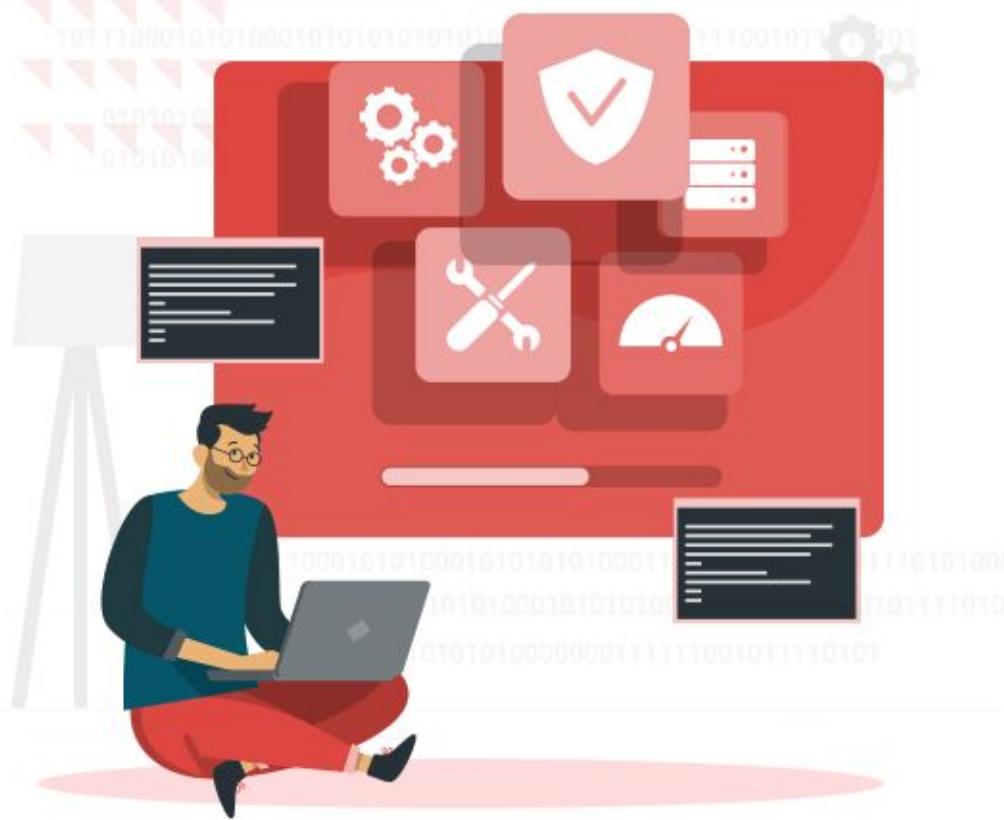
All candidates	+
Candidate master	+
Candidate-committee linkages	+
House/Senate current campaigns	+
Committee master	+
PAC summary	+
Contributions by individuals	+
Contributions from committees to candidates & independent expenditures	+
Any transaction from one committee to another	+
Operating expenditures	+
Electronically filed reports (.fec files)	+

Data Source - Fec Data Column Names and Abbreviations

<https://www.fec.gov/campaign-finance-data/committee-summary-file-description/>

```
'CAND_ID': 'Candidate identification',
'CAND_NAME': 'Candidate name',
'CAND_ICI': 'Incumbent challenger status',
'PTY_CD': 'Party code',
'CAND_PTY_AFFILIATION': 'Party affiliation',
'TTL_RECEIPTS': 'Total receipts',
'TRANS_FROM_AUTH': 'Transfers from authorized committees',
'TTL_DISB': 'Total disbursements',
'TRANS_TO_AUTH': 'Transfers to authorized committees',
'COH_BOP': 'Beginning cash',
'COH_COP': 'Ending cash',
'CAND CONTRIB': 'Contributions from candidate',
'CAND LOANS': 'Loans from candidate',
'OTHER LOANS': 'Other loans',
'CAND LOAN REPAY': 'Candidate loan repayments',
'OTHER LOAN REPAY': 'Other loan repayments',
'DEBTS_OWED_BY': 'Debts owed by',
'TTL_INDIV_CONTRIB': 'Total individual contributions',
'CAND OFFICE_ST': 'Candidate state',
'CAND OFFICE_DISTRICT': 'Candidate district',
'SPEC_ELECTION': 'Special election status',
'PRIM_ELECTION': 'Primary election status',
'RUN_ELECTION': 'Runoff election status',
'GEN_ELECTION': 'General election status',
'GEN_ELECTION_PERCENT': 'General election percentage',
'OTHER_POL_CMTE CONTRIB': 'Contributions from other political committees',
'POL_PTY CONTRIB': 'Contributions from party committees',
'CVG_END_DT': 'Coverage end date',
'INDIV_REFUNDS': 'Refunds to individuals',
'CMTE_REFUNDS': 'Refunds to committees'}
```





Software Tools and Technologies

Software Tools and Technologies

1. Data Bricks
2. Spark
3. Tableau



Data Bricks

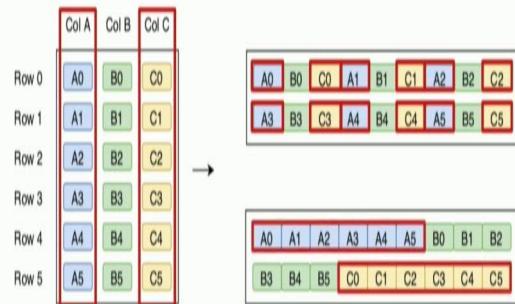


- The AWS, Azure, and Google clouds can all support the use of Databricks to handle enormous volumes of raw, unprocessed data.
- The computing resources and settings that make up a Databricks cluster are used to run data engineering, data science, and data analytics workloads such production ETL pipelines, streaming analytics, ad-hoc analytics, and machine learning.

Parquet

- The very best encoding and compression techniques are supported by Apache Parquet. When used with modern cloud technologies like Amazon Athena, Redshift Spectrum, Big Query, and Azure Data Lakes, Apache Parquet reduces the cost of data file storage and increases the efficiency of data queries.
- With the ability to read and write Parquet files, Spark SQL maintains the original data's schema automatically. All columns are automatically changed to be nullable when creating Parquet files for compatibility reasons.
- Reading and writing Parquet files is supported by Spark SQL, which automatically preserves the schema of the original data. All columns are automatically altered to be nullable when producing Parquet files due to compatibility concerns.

Row-wise vs Columnar



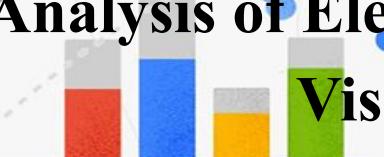
List Of Files After Loading

```
%fs ls "/FinalProject"
```

Table

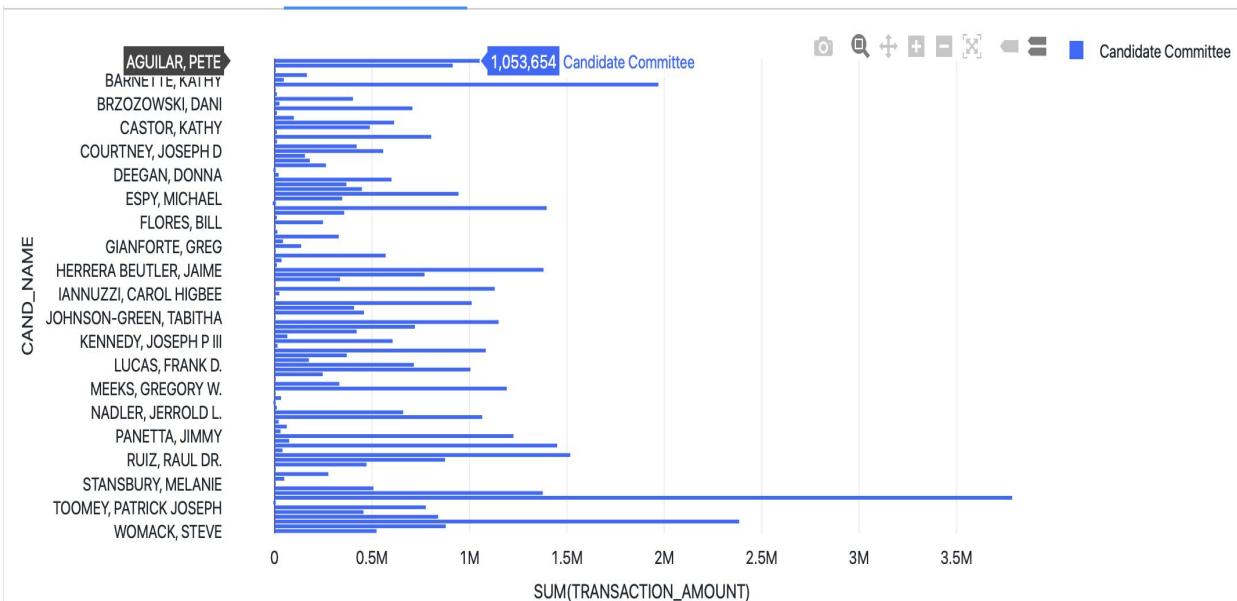
	path	name	size	modificationTime
1	dbfs:/FinalProject/All_Candidates_Cleaned/	All_Candidates_Cleaned/	0	0
2	dbfs:/FinalProject/Cont_by_indv/	Cont_by_indv/	0	0
3	dbfs:/FinalProject/PAC/	PAC/	0	0
4	dbfs:/FinalProject/all/	all/	0	0
5	dbfs:/FinalProject/cand_master/	cand_master/	0	0
6	dbfs:/FinalProject/ccl/	ccl/	0	0
7	dbfs:/FinalProject/cm/	cm/	0	0
Showing all 13 rows.				

Analysis of Election Campaign using Visualizations



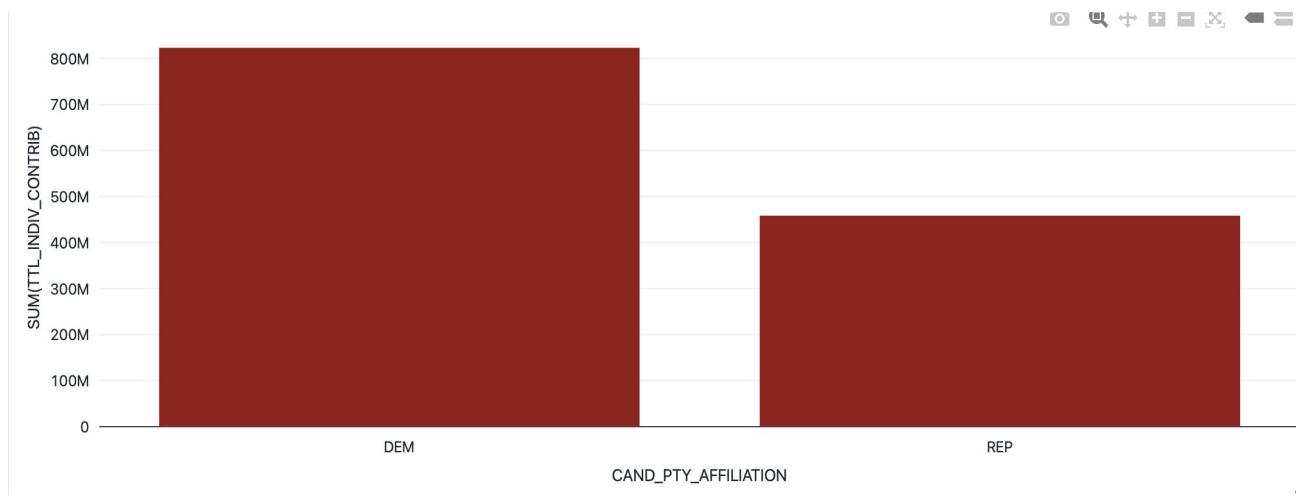
Transaction amount for each candidate contesting

Analysis of the transaction amounts made by each candidate in the campaign, where the candidate's name (CAND_NAME) and the sum of their transaction amounts (SUM(TRANSACTION_AMOUNT)) are shown. In other words, it involves examining the total amount of money spent by each candidate on their campaign and presenting this information alongside the candidate's name.



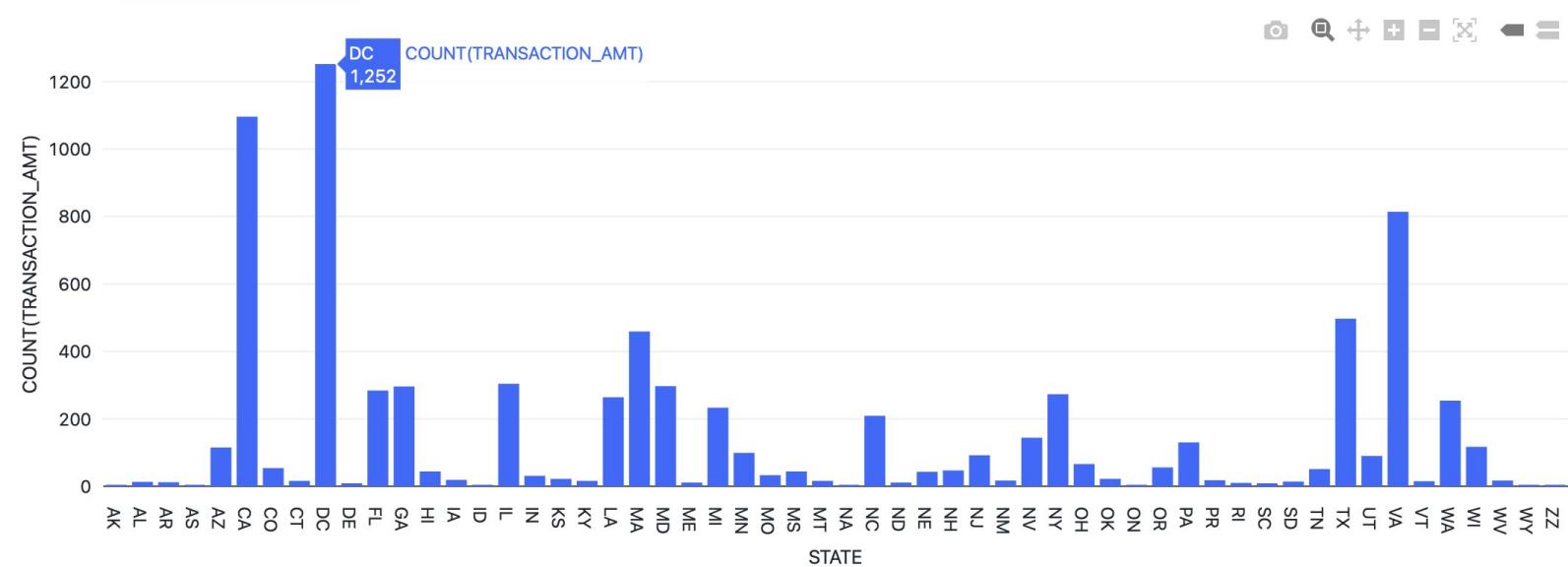
Total Individual Contributions made by candidates

Analysis of the individual contributions made to each candidate in the campaign, where the candidate's party affiliation (CAND_PTY_AFFILIATION) and the sum of their individual contributions (SUM(TTL_INDIV_CONTRIB)) are shown. In other words, it involves examining the total amount of money raised by each candidate through individual contributions and presenting this information alongside the candidate's party affiliation.

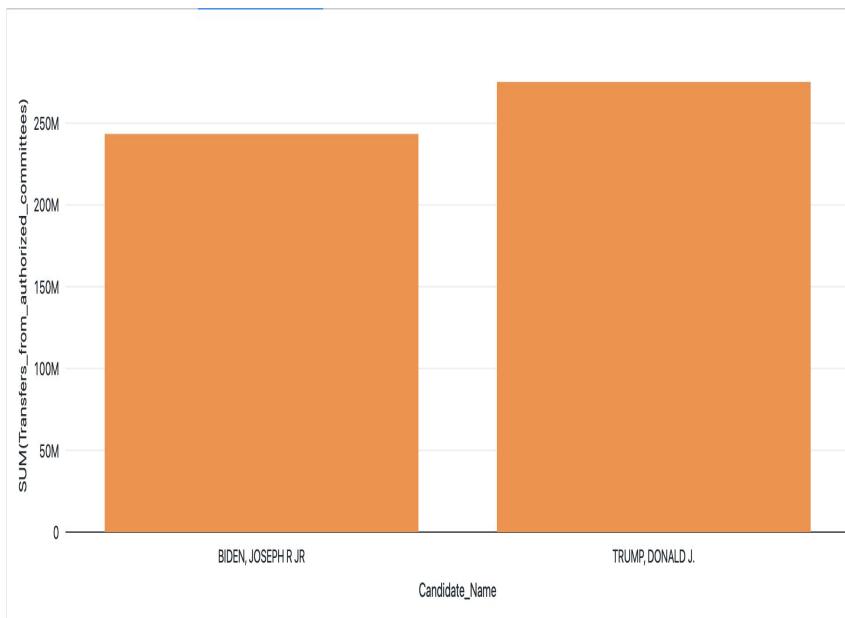


Transaction amount in each state

Analysis of each state's transaction amounts and present the result as a graph, one showing the STATE and the other displaying the count of transaction amounts using the COUNT(TRANSACTION_AMT) :



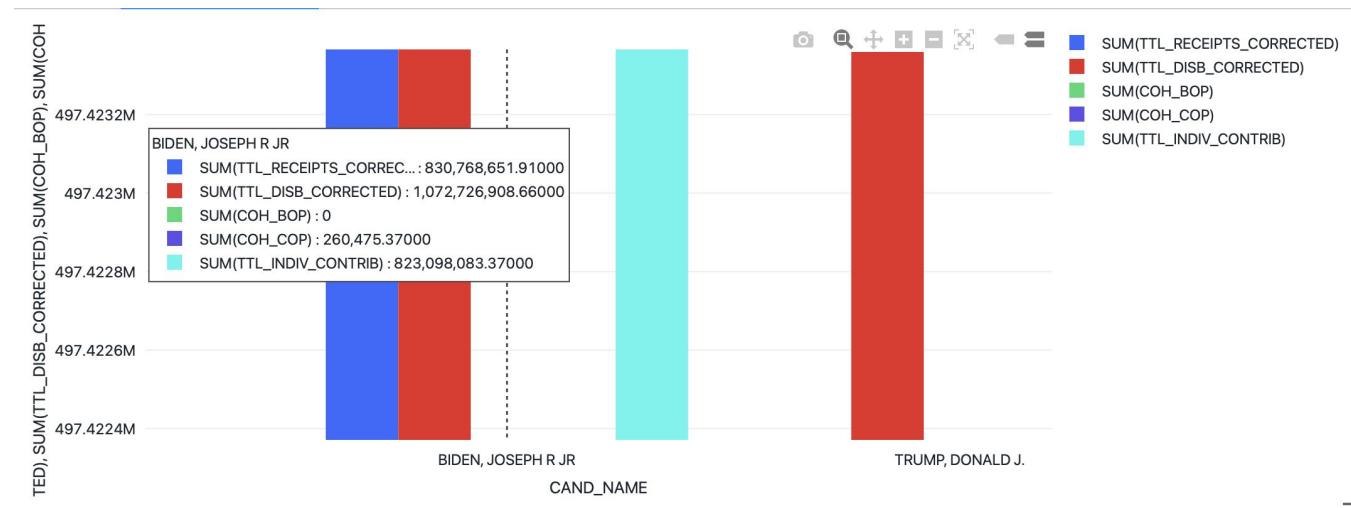
Visualization for transfers from authorized committees to the candidates:



In general, funds may be transferred between authorized committees of the same candidate (for example, from a previous campaign to a current campaign committee) without limit as long as the committee making the transfer has no net debts outstanding.

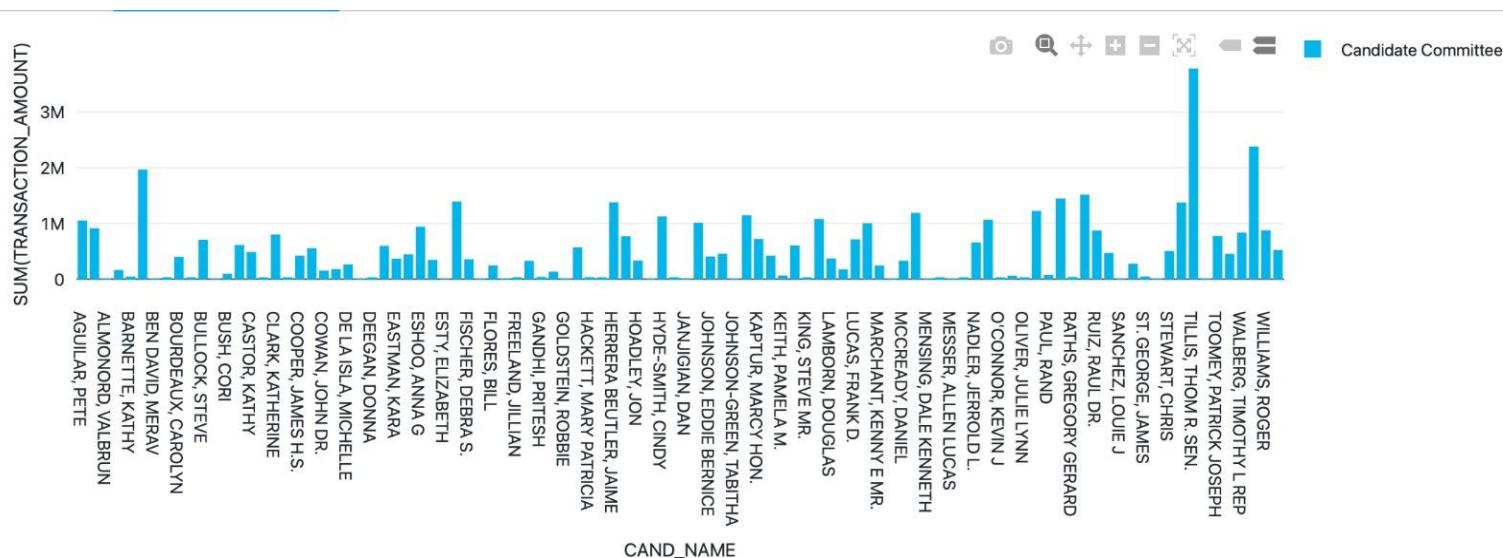
Visualization of COH_BOP and COH_COP for each party

As a result of the analysis , there may be inconsistencies between the figures displayed in this tracker and the totals claimed by the campaign, specifically in regards to COH_BOP (the beginning cash on hand balance) and COH_COP (the closing cash on hand balance) which is clearly been visualised for the understanding :

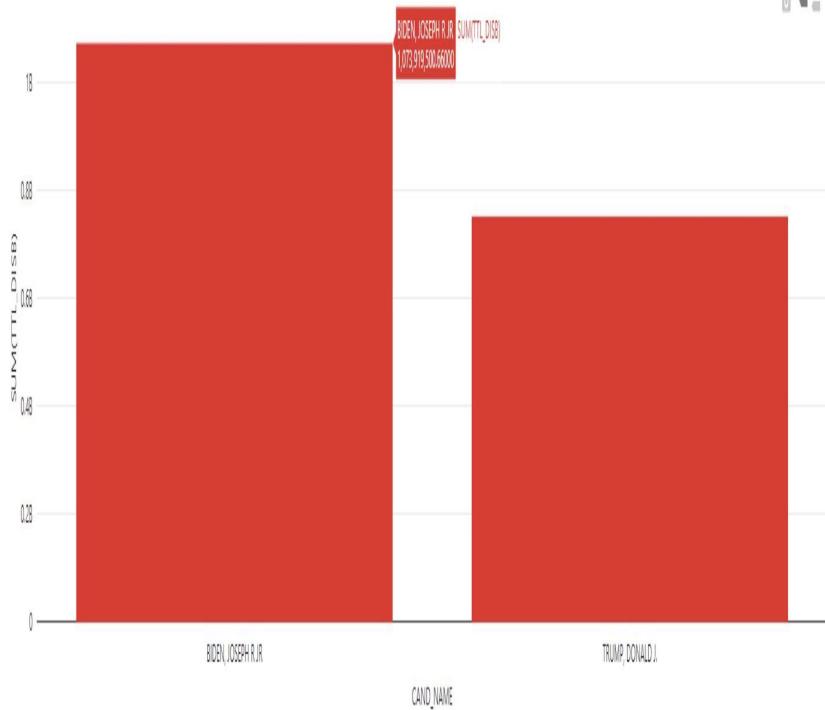


Transaction amount for each candidate contesting

Analysis of each contesting candidate and sum of their transaction amounts that reflects CAND_NAME AND SUM(TRANSACTION_AMOUNT) as shown below :



Total disbursements by Candidates



Disbursement is a border term that covers both expenditure and other kinds of payments(those not made to influence federal election)

The analysis presented involves calculating the total amount disbursed by individual candidates. This data is presented as a graph where the first bar displaying the candidate's name (under the heading CAND_NAME), and the second one showing the sum of their disbursements (under the heading SUM(TTL_DISB)) using the SUM function.

Visualization of Transaction amount by state for campaigning

The result of the analysis showing the total amount spent on campaigning by candidates in each state. This would provide insights into the distribution of campaign spending by state, and highlight any patterns or trends in the data.

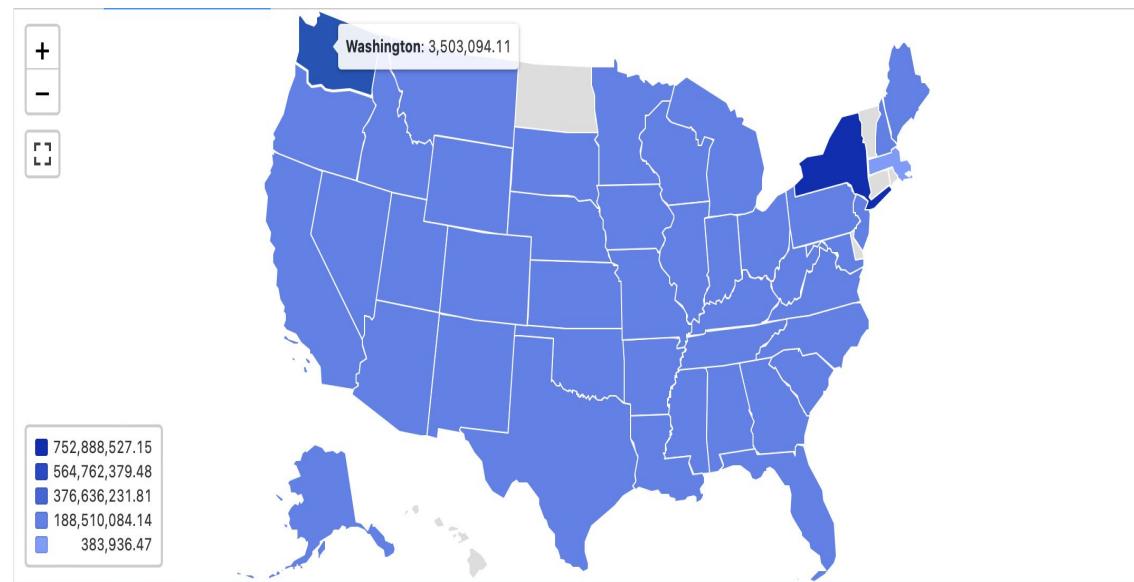
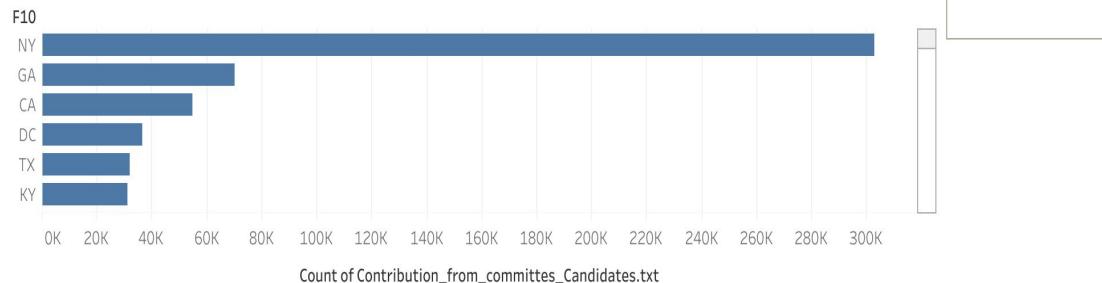


Tableau Dashboards

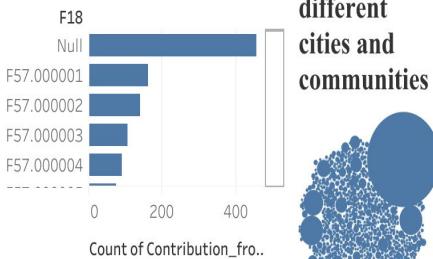
Contributions from each state



Number of Communities

ADRIENNE BELL	GANDHI FOR TEXAS			
CITIZENS FOR MAT				
CORY GARDNER	GINA ORTIZ		SYLVIA GARCIA FOR	

Contributions from each community



Contributions from different cities and communities



Using a Tableau dashboard to show the contribution from each community, city, and state in the United States of America is a powerful way to gain insights into where contributions are coming from and identify areas for potential outreach or engagement. It allows campaign managers and other stakeholders to make data-driven decisions and optimize their fundraising strategy.

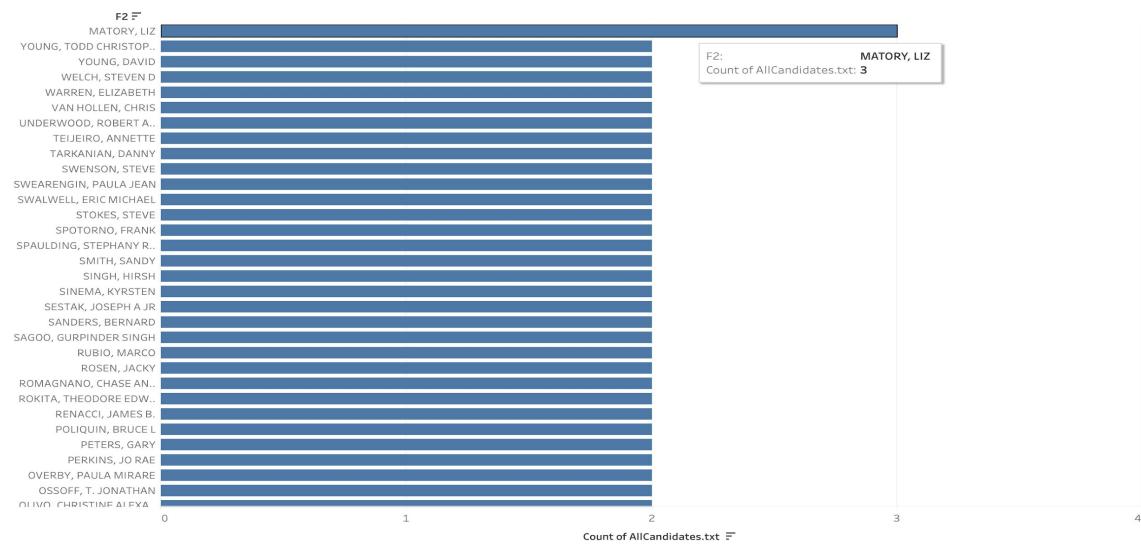
Tableau Dashboard link:

https://public.tableau.com/app/profile/harshitha.yentrapragada3999/viz/FE_CProject/Dashboard1

All candidates from different states for election campaign

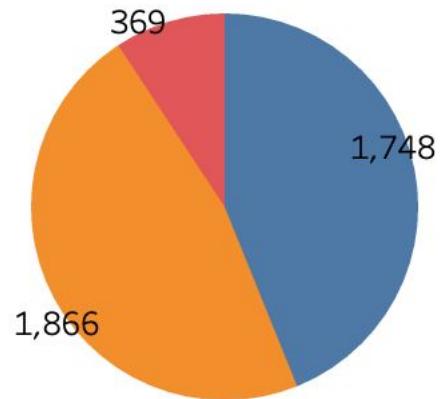
Using tableau dashboard is a useful method to illustrate the candidates from different states who are running in an election campaign. The graph can display the quantity of candidates in each state, and a distinct bar can represent each state.

Count of all candidates



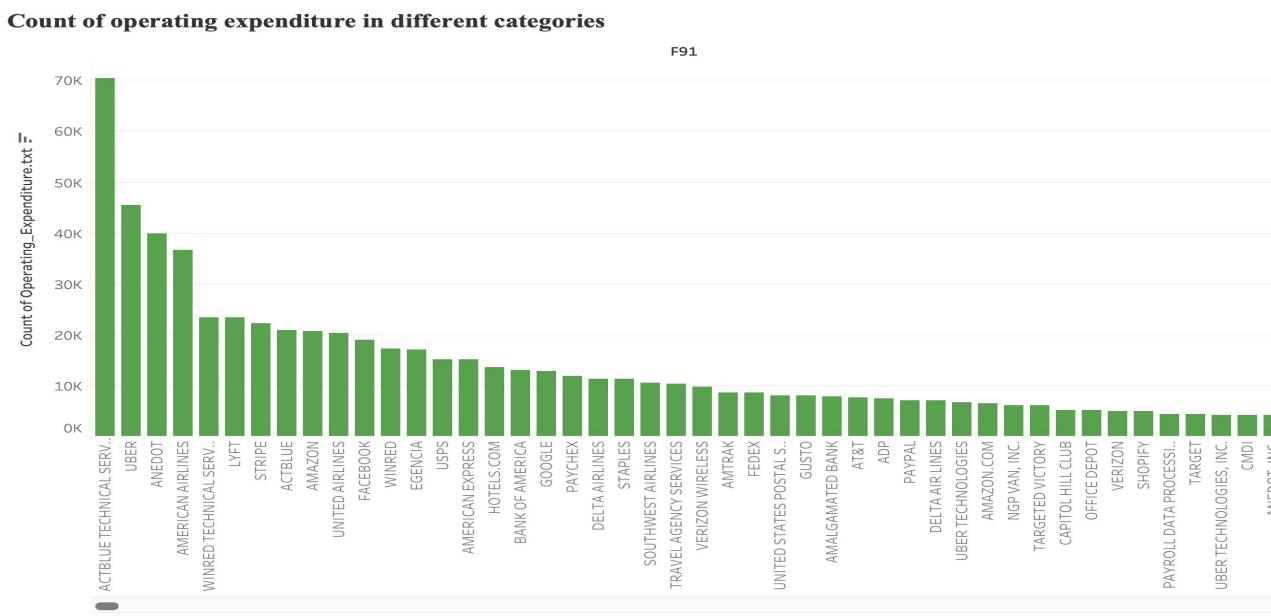
Pie Chart Representation

Total number of candidates in campaign

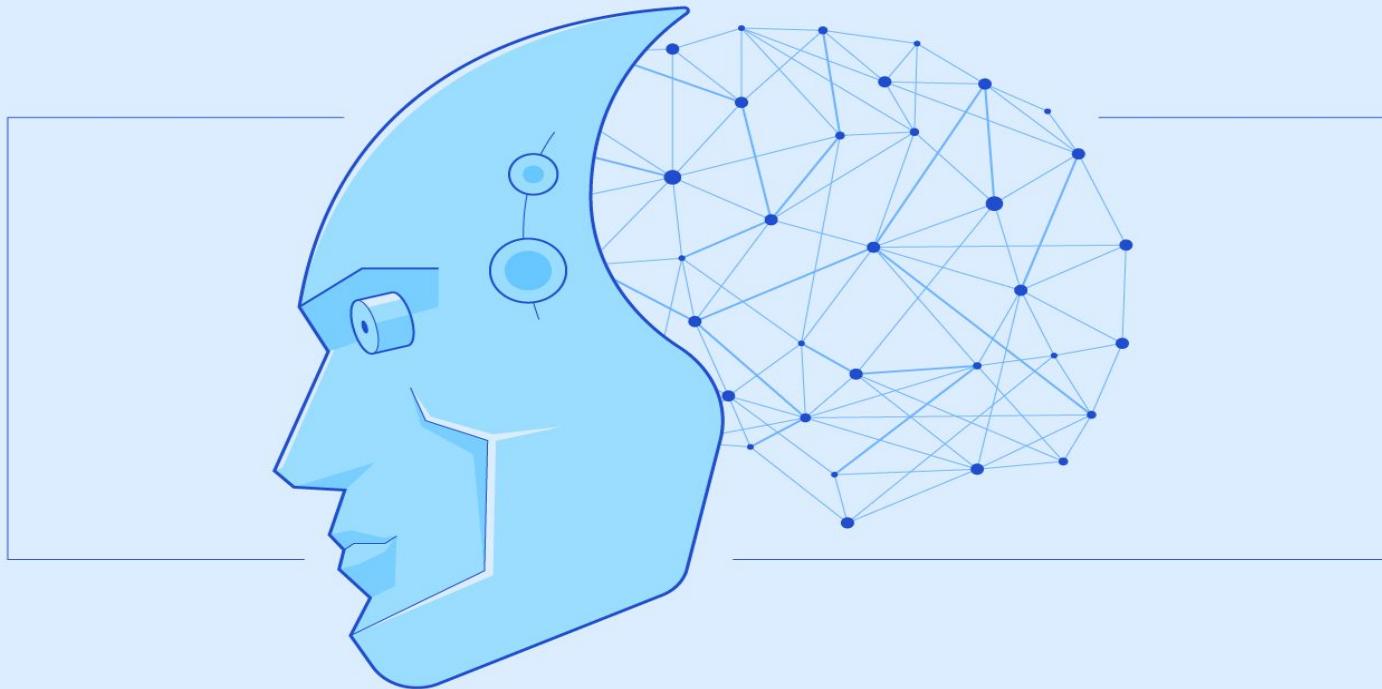


A pie chart is an ideal visual representation to display the total number of candidates in a campaign. It shows the proportional share of candidates belonging to each political party, with different colors or patterns assigned to each party. This makes it easy to understand the breakdown of candidates in the campaign at a glance.

Using a Tableau dashboard is an excellent way to visualize and analyze the count of operating expenditures in different categories. The dashboard can display a variety of information about the expenditures, such as the total amount spent, the count of expenditures in each category, the percentage of the budget allocated to each category, and the change in spending over time.



Machine Learning



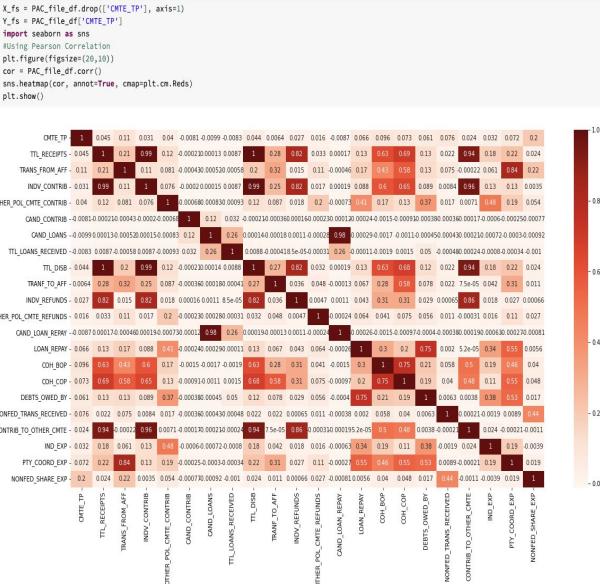
Using ML on Analysis

How Can We apply Machine Learning on US Campaign Data

Machine learning can be used on US campaign data to analyze and predict various aspects of the campaign, such as voter behavior, preferences, and turnout. Some of the ways in which machine learning can be applied to campaign data include:

1. Voter segmentation: Machine learning algorithms can analyze voter data to identify patterns and segment voters based on their demographics, voting history, and other factors. This can help campaigns to tailor their messaging and outreach to specific groups of voters.
2. Predictive modeling: Machine learning algorithms can be used to build predictive models that can forecast voter behavior, such as voter turnout or which candidate a voter is likely to support. This can help campaigns to allocate resources more effectively and target their outreach efforts to the voters who are most likely to be swayed.
3. Find out the relation between the characteristics of the population and the number of votes of Democratic and Republican.

- After performing EDA I had an idea about the features but to have a better understanding of features I have used Pearson correlation which is a method used for numerical variables to find the correlation.
- After using Pearson correlation I was able to extract features but there were non-numerical columns that had to be encoded for applying machine learning.
- I performed encoding then divided the dataset into training and testing data.



task took 2.84 seconds -- by nishanthreddy@gepal1.com at 5/9/2023, 7:04:28 PM on My Cluster

Linear Regression to predict the contribution

In the all candidates file there was very high correlation between two columns and when plotted they had graph as shown.

Using linear regression I was able to predict the total receipts that a committee spends based on receipts collection.

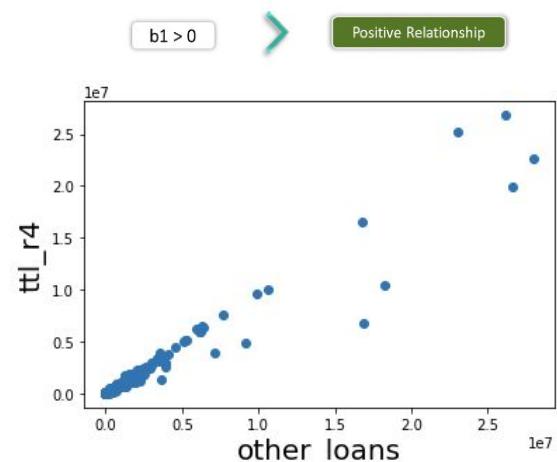
Use Case :-

The model will be able to predict committee's expenditure based on loans they are considering, which will help part to decide how much loan they have to take.

Accuracy :-

R2 Score = (0.9628120579557788, 2)

An R2 score near 1 means that the model is able to predict the data very well.



$$y = b_0 + b_1 x$$

Random Forest to predict the Committee Type :-

In the PAC file there were features which basically help to predict the Committee type.

Limitation :-

I wanted to use Random Forest classifier initially as this was basically a classification model and there were more than one feature. In the Random forest model to avoid correlation between models I have used bagging method.

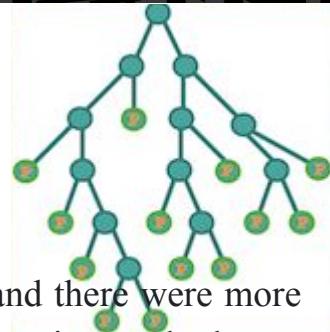
I didn't want to limit the model so I worked even on Decision Tree, however due to overfitting of data the learning was more inclined towards more trained data .

Use Case :-

There are 16 committee types which are basically decided based on the certain features, however with this model I was able to select two main features which affected it most hence helping us derive committee type.

Accuracy :-

R2 Score = (0.8974)



Decision Tree

All Candidates file and Candidates master file has close relation which will allow to find out whether a candidate has made it to elections or committee.

Limitation: -

Before selecting a model it is important to tune models so that we can find a better version of model in this particular case while using decision tree initially the data was overfitting and the tree was very complex so I decided to change the max depth which reduced the accuracy but it was more comprehensible for analysis.

Use Case -

It will allow to predict the threshold for all the candidates to know if they would be elected in committee or not based on various features.

Metrics -

Accuracy - (0.9212)

I didn't want to limit the model so I worked even on Decision Tree, however due to overfitting of data the learning was more inclined towards more trained data .

Project Goal

- The ultimate goal of US campaign finance analysis is to increase transparency and accountability in the political system and to inform the public about the role of money in politics. By providing accurate and accessible data, analysis, and reporting on campaign finance, this project seeks to empower voters, policymakers, and advocates to make informed decisions and take action to strengthen the democratic process.
- US campaign finance analysis is committed to the project goal of providing insights and understanding about the financing of political campaigns in the United States. The analysis aims to shed light on the sources and amounts of campaign contributions, how these contributions are spent, and the impact they have on the political process.
- We aimed the project to help candidates and committee understand the factors that have been affecting campaign and how they can use the models we have incorporated to decide expenditures, places to focus, decide committee type, overall trend analysis and many more interesting insights.

Conclusion

- The raw data from FEC has been transformed into meaningful insight where one can understand the campaigning expenditures made by candidates, individual committees to the parties in contesting election and also explains the states which have contributes highest campaigning funds.
- This will help us to identify the candidate behavior and overall factors which are affecting the US Campaign. We were able to identify expenditure based on state which will help identify the state wise analysis.
- The Machine Learning model helped us to identify some crucial information such as how a committee spends money based on other loans and receipts, committee type and also if the candidate had made it to committee or not.
- It becomes easier to identify the candidates who are spending the most in each category and state. This information can be used to gain insights into the campaign strategies of each candidate and their priorities. Additionally, it can help campaign managers and stakeholders to optimize their own fundraising and expenditure strategies based on the successes of their competitors.

Future Scope

- From the tableau dashboards we can see the number of contributions by individuals as well as committee groups. Naturally, there are many more candidates than there are funds raised by one contender. In order to elect a president, each party has its own committee, which can coordinate with the official campaign and raise money for the war chest.
- The number of financial information that must be received, processed, and made public has sharply increased as a result of the roughly doubling of the number of financial transactions that must be reported to the FEC each election cycle over the past three election cycles through the 2020 cycle. Additionally, campaigns frequently release their combined fundraising totals prior to filing their FEC reports, so there may be discrepancies between what this tracker has compared to what the campaign claims
- One potential area of future focus for US campaign finance analysis is the use of emerging technologies and data sources to improve the accuracy and timeliness of campaign finance reporting. For example, the use of machine learning algorithms and natural language processing techniques could help to automate the analysis of campaign finance data, making it faster and more efficient to identify trends and anomalies in the data.

References

- <https://www.fec.gov/>
- <https://www.nytimes.com/interactive/2020/10/25/us/politics/trump-biden-campaign-donations.html>
- <https://www.gao.gov/assets/gao-20-66r.pdf>
- <https://crsreports.congress.gov/product/pdf/R/R45302>
- <https://www.everycrsreport.com/reports/R44318.html>



Thank You!