

Campaign Finance City Council And State Representative Scorecard

Aaron Jacob Varghese - aaronjv@bu.edu, U60942775

Raj Vipani - rvipani@bu.edu, U11275347

Sun Diwen - jack123@bu.edu, U37462503

Link to data -

<https://drive.google.com/drive/folders/1E0cY1iGYHMwfR30olp5J8CM-kXOnEg6F?usp=sharing>

All code is uploaded to github project repo

1. **Outline of the project:** In this project we aim to analyze the donations made to the city and state representatives of Massachusetts for the years 2016 - 2019. Out of the two teams working on the same project, we divided the work and our team is analyzing the state representatives data. Owned by "Progressive Mass", a state-wide member driven grassroots organization built by organizers and activists regardless of political affiliations from across Massachusetts to advocate for progressive policy, this project aims at answering questions and correlating specific industries for instance healthcare, Law enforcement and real-estate to the donations being contributed by people of these industries.

Our goal in this project was to continue the work done last semester, specifically focusing on the healthcare industry and the funds transferred between entities that operate in this industry. We were also focused on analyzing the funds received by state representatives.

Some of the questions we sought to answer were -

- a. How much each representative receives.
- b. How the amount they receive corresponds to their power i.e. are they on special committees where they wield a lot of power?
- c. Is there a time of year of years where the donated amount goes up?
- d. How much do the Political Action Committees and Unions donate and to whom?

Being a continuation of the same project from last semester, we started by analyzing and verifying the results done by the previous semester team. We continued by scraping the Secretary of State website (<http://corp.sec.state.ma.us/CorpWeb/CorpSearch/CorpSearch.aspx>), collecting data from the OCPF website (<https://www.ocpf.us/>) and the data from Boston Businesses. Furthermore, we cleaned the collected data and after the preliminary analysis we are working forward on a deeper analysis of the data acquired and producing visualizations of the consolidated results.

2. Completed Tasks:

a. Data Collection

- i. **OCPF** - This data was provided to us by the clients in the form of Microsoft Access databases.
- ii. **Secretary of State** - This website has records of every business entity operating within Massachusetts, which includes annual reports of all active entities. Our goal was to download the annual report PDF's which contained details like the type of business, executives and addresses.

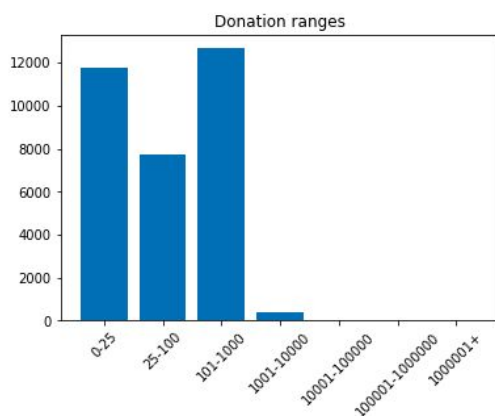
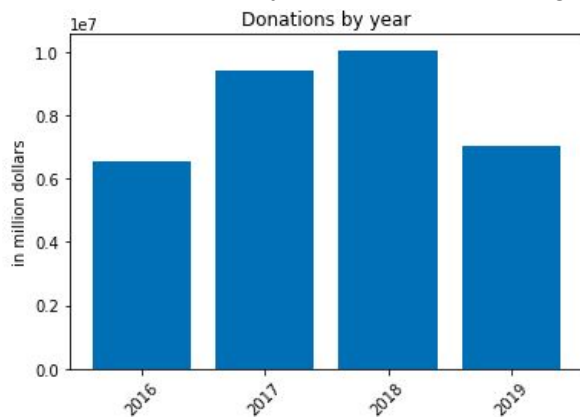
- iii. **Boston Businesses** - This was a website that contained the listings of businesses in the city of Boston, filtered by neighborhood. Our goal was to scrape the site for the details of every listed company.

b. Data cleaning and merging

- i. **Merging of Master and Receipts table in OCPF database** - We worked on the analysis of the OCPF database by merging the Master and Receipts tables in the OCPF database on the basis of the "Report ID", the id assigned to the individual donors after the donations made.
- ii. **Use Sec. State PDF's to get business types and use that to augment the filtering process in OCPF data** - We went through each of the downloaded PDF files to obtain the details of each active business entity. Most importantly, we needed the type of business, which could be used to filter through the OCPF data, as our team was focused on the healthcare industry.

c. Data analysis

- i. **Analysis of OCPF data** - For the preliminary analysis of the OCPF data, we created histogram charts of the overall contributions made to candidates per year and analysed the donation ranges.



Calculated ranges

```
{'0-25': 11789, '25-100': 7744, '101-1000': 12664, '1001-10000': 398, '10001-100000': 37, '100001-1000000': 20, '1000001+': 7}
```

Furthermore, we analysed the OCPF data on the basis of the PAC donations. For this step, we used the PAC names in the healthcare industry from the state of Massachusetts (provided by the Project Clients), and filtered the OCPF data to get a consolidated result of which PAC donated to which candidate and how much amount

during the period of interest of our analysis.

Donor_Name	Amount_Donated
Massachusetts Health & Hospital Association	25099250.00
Massachusetts Nurses Association	10543601.46
Republican Governors Association	6625000.00
Roman Catholic Archdiocese Of Boston	1700000.00
Sheldon Adelson	1000000.00
Massachusetts Nurses Association Region 5	750000.00
1199 Seiu Ma Pac	569050.00
Laura And John Arnold	510000.00
Steward Health Care System Llc	505100.00
Massachusetts Nurses Association Region 2	350000.00
Knights Of Columbus	300000.00
New England District Council Health & Welfare ...	279120.90
Strong Economy For Growth	228000.00
New England District Council Health & Welfare	203409.11
Partners Healthcare	200000.00
Wynn Resorts, Limited	200000.00
Massachusetts Nurses Association Region 3	200000.00
Gmmb, Inc	190737.42
Partners Healthcare System, Inc	190385.60
Massachusetts Hospital Association, Inc	185947.55

Candidate_Received	Amount_Donated
Coalition to Protect Patient Safety	2.580052e+07
Committee to Ensure Safe Patient Care	1.250142e+07
Commonwealth Future Independent Expenditure Pol...	6.625000e+06
The Campaign for a Safe and Healthy Massachusetts	4.249864e+06
Charles D. Baker	6.740445e+05
Ironworkers Union Local 7 Pol Action Comm	5.429560e+05
Patients for Affordable Drugs Action Independen...	5.102500e+05
1199 SEIU MA PAC	4.301240e+05
Committee to Protect Massachusetts Hospital Care	3.763332e+05
Karyn E. Polito	3.237270e+05
Freedom for All Massachusetts, Inc.	3.113760e+05
Massachusetts Dental Society Political Action C...	3.041890e+05
Maura T. Healey	2.376597e+05
Raise Up Massachusetts 2018	2.153900e+05
Robert A. DeLeo	2.107700e+05
Martin J. Walsh	1.997670e+05
Massachusetts Senior Coalition	1.683370e+05
Ronald Mariano	1.290400e+05
Jay Gonzalez	1.156177e+05
Baker-Polito 2019 Inaugural Committee	1.004687e+05

3. Pending Tasks

- a. **Further data analysis using Sec. State PDF data to filter OCPF** - For this step we aim to couple the data collected from the Secretary of State website and use the keywords such as the industry the businesses are registered in and the key personnel in these businesses to discover trends in the donations to the state representative candidates and see if there are any particular underlying trends.
- b. **Looking into specific committees that the recipients might be in** - For this step of analysis we are looking to identify key members from the Joint Committee on Healthcare Financing, Joint Committee on Mental Health, Substance Use and Recovery and Joint Committee on Public Health of the state of Massachusetts. We have already compiled the list

of all the members sitting on these committees along with the various committees they sit on and for the next step we plan to visualize the donations made to these members which would further help us analyse the donation behaviour.

- c. **Grouping data by recipients and donors to find patterns** - For this task we hope to find patterns such as focusing on the time when donations are made to candidates i.e. concentrating on the relation between the time when a legislation is introduced and seeing if it causes a change in the trend of donations.

4. Issues

- a. **Sec. State website slow and buggy** - For us, the part of collecting annual reports was divided into two phases. First we collected the links to the annual reports. The code we developed for this used to often crash and we'd have to change the cookie and viewstate parameters in our code. Second was the phase of downloading the PDFs, which was slow and it took longer than estimated because of several bugs in the website and an anti-scrapers that we had to bypass.
- b. **Data download process was delayed as the website was changed** - The process to download the data from the secretary of state website was further delayed because midway through the process the website was changed, which complicated our data collection code a lot, forcing us to change it and introduce some measures that made the overall process slightly slower.

For both the issues discussed above, we immediately reached out to our project PM and the clients from Progressive mass and had kept them in the loop as we worked to get through it.

- c. **Lots of different formats used in the PDF's, hence hard to collect data** - Another difficulty that we faced and had to overcome was the different formats of the annual reports on the secretary of state website. It took quite a bit of trial and error analysis to model the format of the PDF's as accurately as possible.