
Campaign Finance City Council and State Representative Scorecard

Yuang Liu (U99473611), Jun Xiao(U85900288), Qi Yin (U31787103), Yiming Miao(U76158067)

Department of Computer Science

Boston University

yaliu@bu.edu



Contents

1 Reproduce Results	3
1.1 Scraper for the Sec of State part	3
1.2 Download city of boston data	3
1.3 Analyzing OCPF data	3
1.4 Instructions of APIs	3
2 Database	4
2.1 OCPF data	4
2.2 Annual report for companies from N-Z	4
2.3 Business type for companies from N-Z	4
2.4 All info for companies from N-Z	4
2.5 Info from City of Boston	4
3 Results and Visualizations	5
3.1 City Committees	5
3.2 Results	8
3.2.1 Distribution of Donations to City Committees	8
3.2.2 A Deeper Look into the Real Estate Field	8
4 Answers to the Questions	12

1 Reproduce Results

1.1 Scraper for the Sec of State part

- i. Run `get_all_links.py` script to extract all the companies links for a specific letter, which you should specify in the python script. The script will save all the links to a txt file, where you should specify the name of the txt file such as `links_for_z.txt`
- ii. Run `get_all_pdf.py` script to download all the annual reports for the companies. You have to specify the txt file storing the links for companies. The script runs very slow, it will take you one hour to download only 1,200 to 1,500 links. Sometimes, you have to use VPN to change your IP, for that the Sec of State website has a very strict anti-scraper management. (Actually, it takes us 8 whole days to download all pdfs for companies from N-Z on three computers).
- iii. Run `analyze_pdf.py` script to get the companies name and corresponding business type. And run `analyze_pdf_extra_info.py` to get some extra information such as Street Address, Registered Agent, Street Address of Corporation Principal Office, Names and Addresses of Board Members or Executives.

Up to here, we finish downloading and analyzing all companies for one specific letter from the Sec of State website. If you want to download pdfs for other letters. Just go through all the above steps again.

1.2 Download city of boston data

Just run the `scrapy_for_city_of_boston.py` script, it will collect all the companies relating to the real estate field based on the keywords. Actually, this scraper is only a backup solution for fear that we are not able to handle the Sec of State website.

1.3 Analyzing OCPF data

- i. Use `filter_ocpf.py` script to filter the ocpf data to get all the raw data from 01-01-2016 to 12-31-2019. The intermediate data will be stored in the file you specify in the script.
- ii. Use `get_ocpf_by_pcas.py` to get ocpf data only relating to PCA. The intermediate data will be stored in the file you specify in the script.

1.4 Instructions of APIs

After the two steps above, we've already got the raw data to analyze, then run the `analyze_ocpf_data_with_keywords.py` script. There are many APIs in this script, here are some instructions over them:

- (1) `get_filtered_data_xxx`: extract data relating with a specific field 'xxx' based on keywords
- (2) `compute_amount_by_cpfid`: count the amount of money a receiver received and sort by the amount from lowest to highest
- (3) `compute_amount_by_season`: count the amount of donations based on season
- (4) `compute_amount_by_year`: count the amount of donations based on the year
- (5) `computer_amount_by_size`: count the amount of donations based on the size of the donations (under \$25, from \$25-\$50 etc)
- (6) `computer_amount_by_city`: count the amount of donations based on city
- (7) `computer_amount_by_state`: count the amount of donations based on state
- (8) `computer_amount_by_contributor_type`: count the amount of donations based on contributor's type (individual or committees)
- (9) `the_most`: get the top 10 companies donator regarding with total amount, average amount per donation and total numbers of donation
- (10) `companies_insight`: see which companies donate the most number of different receivers and which companies donate the most time to one receiver

All steps above are the way to generate the result data. And the visualization jobs are mostly done with the help of Microsoft Excel.

2 Database

All the datasets are shown as follows:

2.1 OCPF data

<https://drive.google.com/drive/folders/19Ne8561l4XdHYnpyJpWHziV0G74dKp2n?usp=sharing>

2.2 Annual report for companies from N-Z

https://drive.google.com/drive/folders/1kMET_SQKGrKh81KeOt1xhzUta8-mpucM?usp=sharing

2.3 Business type for companies from N-Z

https://drive.google.com/drive/folders/12DA-eT82_1iDsWZdHoLKOa6Y2l1ty_Ub?usp=sharing

2.4 All info for companies from N-Z

<https://drive.google.com/drive/folders/1SYcDUuOsn0aOQ-lFodipP0vDwmGNwcIF?usp=sharing>

2.5 Info from City of Boston

https://drive.google.com/drive/folders/1CjNdmvsTaEkKzN_BylggM-eJoXcp1D0F?usp=sharing

3 Results and Visualizations

3.1 City Committees



ANDREA CAMPBELL(15931)

Real Estate

WAYS AND MEANS (MEMBER)

Law Enforcement

PUBLIC SAFETY AND CRIMINAL JUSTICE (COMMITTEE CHAIR)

Higher Education

EDUCATION (VICE CHAIR)



ANNISSA ESSAIBI GEORGE(15618)

Real Estate

WAYS AND MEANS (VICE CHAIR)

Law Enforcement

PUBLIC SAFETY AND CRIMINAL JUSTICE (MEMBER)

CIVIL RIGHTS (MEMBER)

Healthcare

STRONG WOMEN, FAMILIES, AND COMMUNITIES (MEMBER)

Higher Education

EDUCATION (COMMITTEE CHAIR)



ED FLYNN(14391)

Real Estate

WAYS AND MEANS (MEMBER)

HOUSING AND COMMUNITY DEVELOPMENT (MEMBER)

Law Enforcement

PUBLIC SAFETY AND CRIMINAL JUSTICE (MEMBER)

CIVIL RIGHTS (VICE CHAIR)



FRANK BAKER(15333)

Real Estate

PLANNING, DEVELOPMENT, AND TRANSPORTATION (VICE CHAIR)

WAYS AND MEANS (MEMBER)

HOUSING AND COMMUNITY DEVELOPMENT (MEMBER)



JULIA MEJIA(17092)

Real Estate

WAYS AND MEANS (MEMBER)

Law Enforcement

CIVIL RIGHTS (COMMITTEE CHAIR)

Healthcare

STRONG WOMEN, FAMILIES, AND COMMUNITIES
(MEMBER)

KENZIE BOK(17164)

Real Estate

WAYS AND MEANS (COMMITTEE CHAIR)

PLANNING, DEVELOPMENT, AND TRANSPORTATION
(MEMBER)

HOUSING AND COMMUNITY DEVELOPMENT
(VICE CHAIR)

Energy

ENVIRONMENT, RESILIENCY, AND PARKS (MEMBER)



KIM JANEY(16537)

Higher Education

EDUCATION (MEMBER)



LIZ BREADON(17111)

Real Estate

PLANNING, DEVELOPMENT, AND TRANSPORTATION
(MEMBER)

HOUSING AND COMMUNITY DEVELOPMENT
(MEMBER)

Energy

ENVIRONMENT, RESILIENCY, AND PARKS (MEMBER)

Healthcare

STRONG WOMEN, FAMILIES, AND COMMUNITIES
(COMMITTEE CHAIR)

LYDIA EDWARDS(16314)

Real Estate



PLANNING, DEVELOPMENT, AND TRANSPORTATION (MEMBER)

HOUSING AND COMMUNITY DEVELOPMENT (COMMITTEE CHAIR)

Energy

ENVIRONMENT, RESILIENCY, AND PARKS (MEMBER)

Higher Education

EDUCATION (MEMBER)

MATT O'MALLEY(14092)

Law Enforcement

PUBLIC SAFETY AND CRIMINAL JUSTICE (MEMBER)

CIVIL RIGHTS (MEMBER)

Energy

ENVIRONMENT, RESILIENCY, AND PARKS (COMMITTEE CHAIR)

Healthcare

STRONG WOMEN, FAMILIES, AND COMMUNITIES (VICE CHAIR)

Higher Education

EDUCATION (MEMBER)

MICHAEL FLAHERTY(12892)

Real Estate

WAYS AND MEANS (MEMBER)

Law Enforcement

PUBLIC SAFETY AND CRIMINAL JUSTICE (VICE CHAIR)

Higher Education

EDUCATION (MEMBER)

MICHELLE WU(15563)

Real Estate

PLANNING, DEVELOPMENT, AND TRANSPORTATION (COMMITTEE CHAIR)

Law Enforcement

CIVIL RIGHTS (MEMBER)

Energy

ENVIRONMENT, RESILIENCY, AND PARKS (VICE CHAIR)

Healthcare

STRONG WOMEN, FAMILIES, AND COMMUNITIES (MEMBER)

RICARDO ARROYO(17105)



Real Estate

PLANNING, DEVELOPMENT, AND TRANSPORTATION (MEMBER)

Law Enforcement

PUBLIC SAFETY AND CRIMINAL JUSTICE (MEMBER)

CIVIL RIGHTS (MEMBER)

Higher Education

EDUCATION (MEMBER)

3.2 Results

3.2.1 Distribution of Donations to City Committees

These pie charts show the distribution of donations from different industry for each city committees listed above, the total amount is the sum of individual donations and PAC donations for that some industries are lack of PAC data.

The pie charts are shown in Figure 1.

According to the charts, the higher the committee's position in the corresponding field, the more inclined it is for him/her to receive more donations in this field. Generally speaking, the committees are more likely to receive donations from real estate, higher education, and healthcare. Besides, the amount of donation is not closely related to the committees' positions as we imagined before. It may be only related to the committee himself/herself, since the committee who received more donation from one category may also receive more from other categories.

3.2.2 A Deeper Look into the Real Estate Field

1. The distribution of donations from different PACs for each city committees are shown in Figure 2.

2. Compute by Season

The result is shown in Table 1 and Figure 3.

3. Compute by Year

The result is shown in Table 2 and Figure 4.

4. Compute by Size

The result is shown in Table 3 and Figure 5.

Size	Number	Amount
< \$25	422	\$4304.76
\$25 – \$99	1884	\$86518.04
\$100 – \$249	4903	\$669819.92
\$250 – \$499	3714	\$976471.49
\$500 – \$1000	3833	\$1942196.61
> \$1000	2530	\$3210500.00

Table 3: Compute by Size

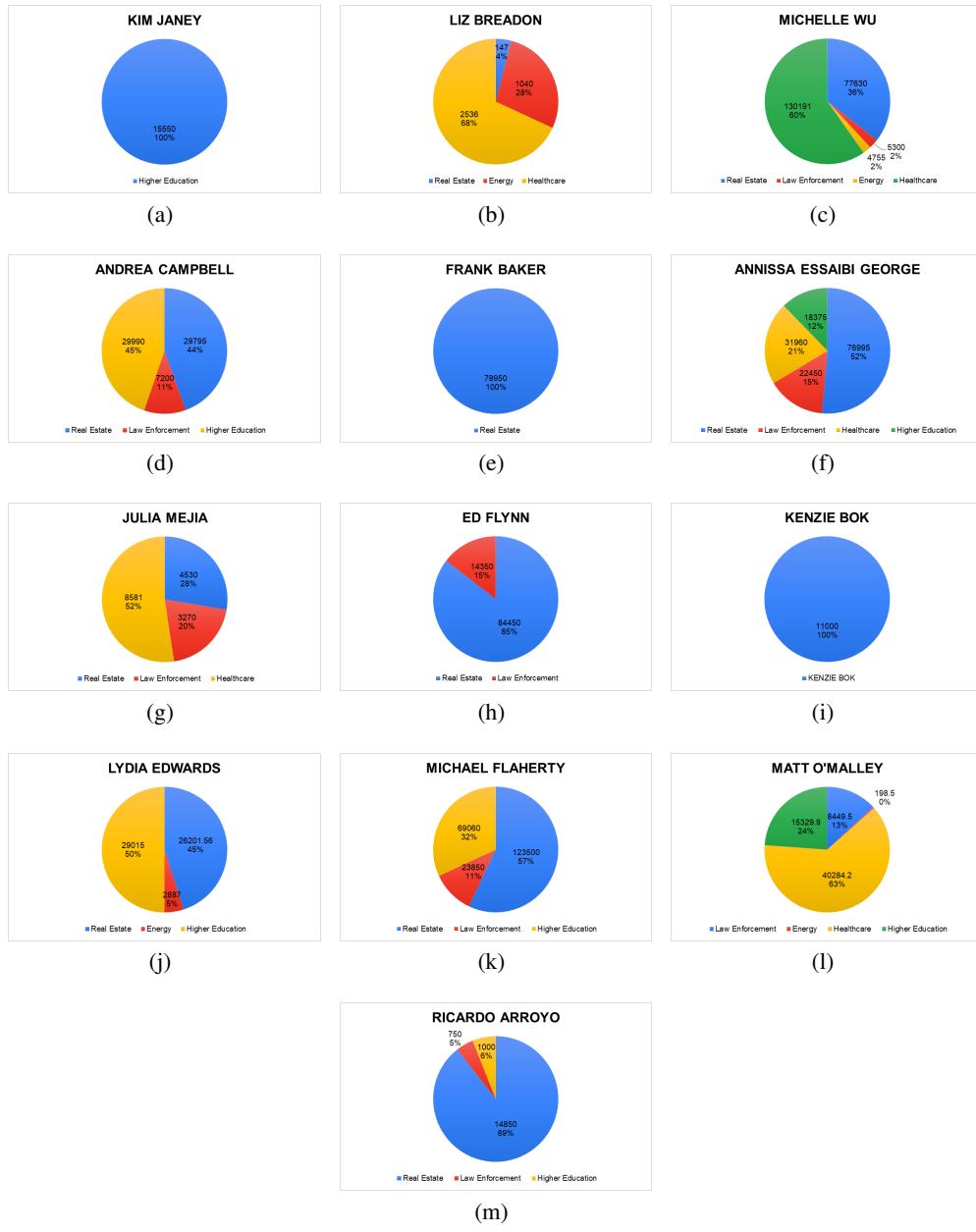


Figure 1: Distribution of Donations to City Committees

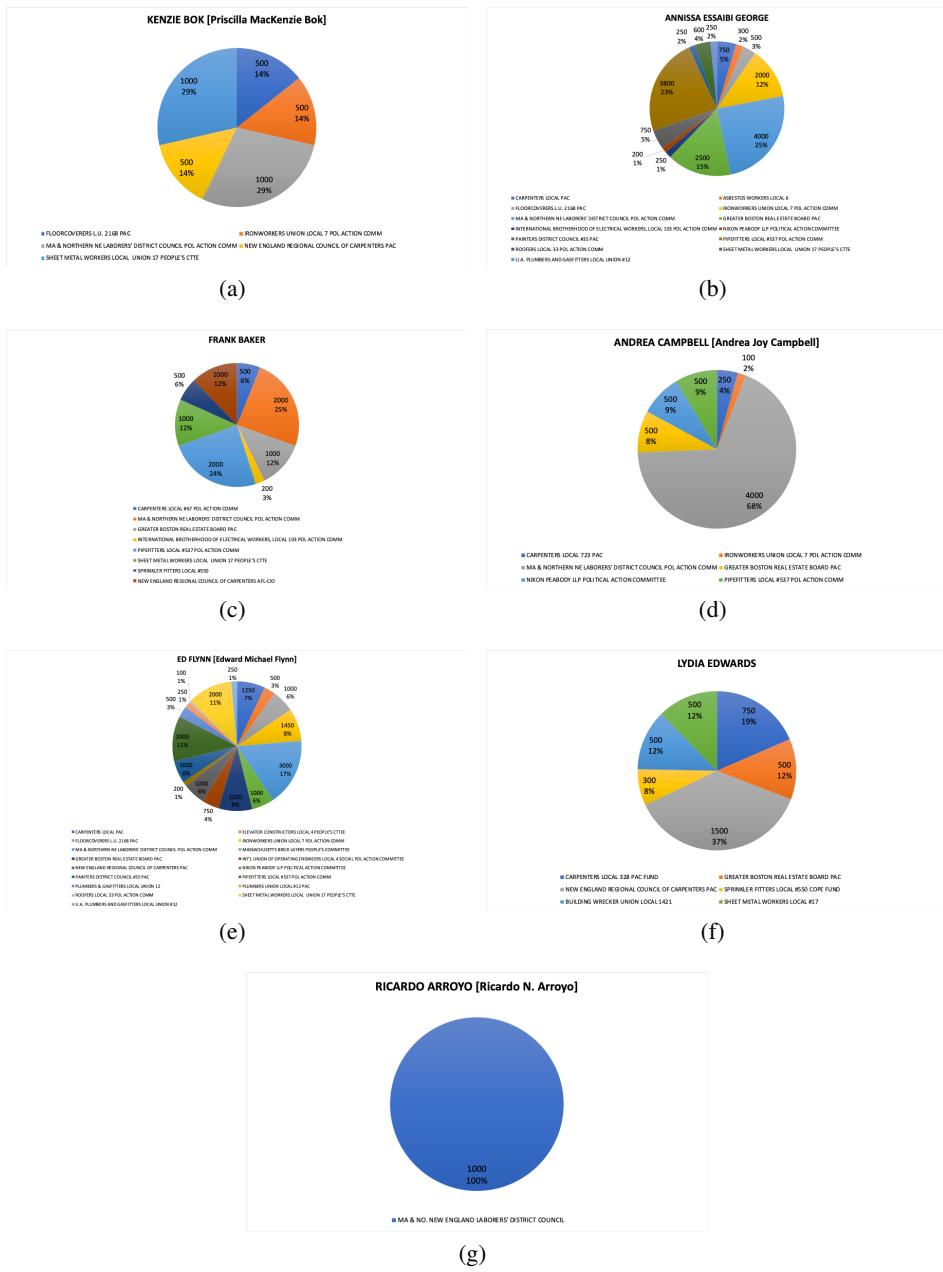


Figure 2: Distribution of donations from different PACs for each city committees

Season	Amount
1	\$1134122.81
2	\$1901463.20
3	\$1347196.58
4	\$2507028.23

Table 1: Compute by Season

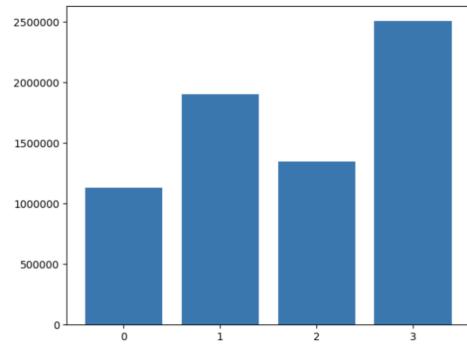


Figure 3: Compute by Season

Year	Amount
2016	\$1942829.81
2017	\$1736838.23
2018	\$1739569.30
2019	\$1470573.48

Table 2: Compute by Year

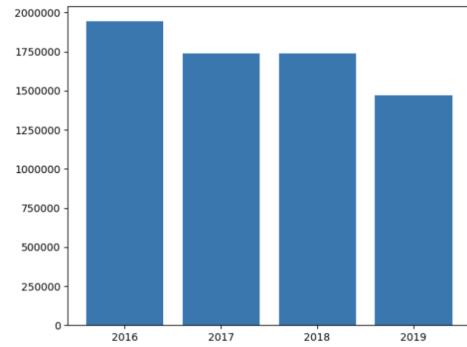
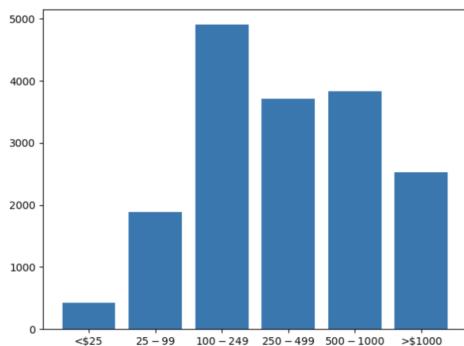
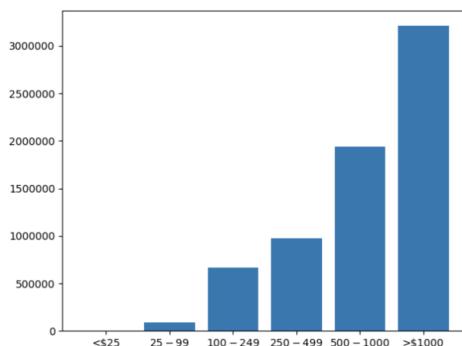


Figure 4: Compute by Year



(a) Number by Size



(b) Amount by Size

Figure 5: Compute by Size

4 Answers to the Questions

1. What is the average donation (\$) to all city committees by industry?
The results are shown in Table 4.

	Total	Average
Real estate	528048.6	356.1
Law Enforcement	85609.5	268.1
Higher Education	178319.9	303.5
Energy	8880.5	231.7
Healthcare	213552.2	195.6

Table 4: The average donation (\$) to all city committees by industry.

2. Which city committees are the outliers i.e. larger volume of donations received? How much larger?

The results are shown in Table 5.

	People	Amount
Real estate	MICHAEL FLAHERTY	123500.0
Law Enforcement	ANNISSA ESSAIBI GEORGE	22450.0
Higher Education	MICHAEL FLAHERTY	69060.0
Energy	MICHELLE WU	4755.0
Healthcare	MICHELLE WU	130191.0

Table 5: The outliers of city committees.

3. Is there a pattern or insights of these individuals? E.g. time of year, donating to multiple candidates from the same companies, etc.)

- (a) Here list (companies, donations receivers, how many times, the amount of donations), and shows the amount the companies donated and how many times they donated to one specific receiver. We only list donation times more than 2,000, and the results are shown in Table 6.
- (b) Here list (companies name, number of different receivers), and shows each company donated to how many receivers. Here only list companies that donate to more than 200 different receivers and the results are shown in Table 7.

4. Do members of the committees most relevant to a given industry receive a significantly larger volume of donations? The analysis and results are shown in Part 3.2.1.

5. How much are PACs and Unions giving (look at this alone as well as by industry)?

The results are shown in Table 8.

6. Who are the largest industry donors (include PACs, Unions, Companies, Individuals)

- (a) Top 10 companies in total amount are shown in Table 9
- (b) Top 10 companies in average amount are shown in Table 10.
- (c) Top 10 companies in total times are shown in Table 11.

7. What percent of Reps overall receive donations from the same firm by industry. (eg, 60% of reps receive from RE firms who make donations, while only 25% receive from Law Enf donating firms)

The results are shown in Table 12.

Company	Donation Receiver	Time	Amount
VERIZON NEW ENGLAND INC.	Int'l Brotherhood of Electrical Workers Local Union 2222 Pol Action Comm	157478	313080
MASS PCA - CP MASS	1199 SEIU MA PAC	16540	138808.1
VERIZON SERVICES CORP.	Int'l Brotherhood of Electrical Workers Local Union 2222 Pol Action Comm	14149	20423.5
MASS PCA - STAVROS	1199 SEIU MA PAC	12673	97343.7
STATE OF MA	Local 509 Service Employees Int'l Union Comm on Pol Ed MA Workers' Pol Action Comm.	12284	33302.6
MASS PCA - NORTH SHORE	1199 SEIU MA PAC	7301	56961.8
STATE OF MASSACHUSETTS	Local 509 Service Employees Int'l Union Comm on Pol Ed MA Workers' Pol Action Comm.	6947	18864.8
RETIRED	Charles D. Baker	4085	1098682.2
VERIZON CORPORATE SVCS. CORP	Int'l Brotherhood of Electrical Workers Local Union 2222 Pol Action Comm	3858	5055
MCOFU	MA Correction Officers Federated Union PAC M.C.O.F.U PAC	3298	14611.8
STOP AND SHOP	Chapter 25 Associated the Nat'l DRIVE PAC of the Int'l Brotherhood of Teamsters	2620	21483.8
CITY OF BOSTON	Martin J. Walsh	2194	463308.2

Table 6: The amount the companies donated and how many times they donated to one specific receiver.

Company	Number of Different Receivers
SELF-EMPLOYED	1221
RETIRED	1029
COMMONWEALTH OF MASSACHUSETTS	473
NOT EMPLOYED	393
N/A	256
HARVARD UNIVERSITY	218
UNEMPLOYED	210
CITY OF BOSTON	210

Table 7: Each company donated to how many receivers.

	PAC	Individual
Real estate	103350.0	424698.6
Law Enforcement	12650.0	72969.5
Higher Education	N/A	178319.9
Energy	N/A	8880.5
Healthcare	200.0	213352.2

Table 8: The amount PACs and Unions give.

Name	Amount
SELF-EMPLOYED	9905121.9
RETIRED	8104532.1
THE BAUPOST GROUP LLC	3979550.0
HIGHFIELDS CAPITAL MANAGEMENT LP	2865350.0
NOT EMPLOYED	2537793.7
BAIN CAPITAL	1429300.0
ARVEST BANK GROUP	1125000.0
LAS VEGAS SANDS CORPORATION	1000000.0
CITY OF BOSTON	859949.6
BLOOOMBURG LP	740000.0

Table 9: Top 10 companies in total amount.

Name	Amount
ARVEST BANK GROUP	1125000.0
LAS VEGAS SANDS CORPORATION	1000000.0
ACTION NOW INITIATIVE LLC	510000.0
HIGHFIELDS CAPITAL MANAGEMENT LP	260486.3
DE SHAW & CO	250000.0
CENTAURUS ADVISORS	250000.0
PAR CAPITAL MANAGEMENT INC.	250000.0
BLOOOMBURG LP	246666.7
THE BAUPOST GROUP LLC	234091.2
REGENT ABLE ASSOCIATE	200000.0

Table 10: Top 10 companies in average amount.

Name	Times
VERIZON NEW ENGLAND INC.	157495
RETIRED	32545
SELF-EMPLOYED	25486
MASS PCA - CP MASS	216540
VERIZON SERVICES CORP.	14149
NOT EMPLOYED	13963
MASS PCA - STAVROS	12674
STATE OF MA	12366
MASS PCA - NORTH SHORE	7301
STATE OF MASSACHUSETTS	7076

Table 11: Top 10 companies in total times.

	Total	Committees Received	Portion
Real estate	6889810.8	528048.6	7.66%
Law Enforcement	2077230.3	85609.5	4.12%
Higher Education	11340422.3	178319.9	1.57%
Energy	763041.1	8880.5	1.16%
Healthcare	16801269.5	213352.2	1.57%

Table 12: Percent of Reps overall receive donations from the same firm by industry.