

Boston University
CS506
Bay State Banner
Deliverable 2 & 3

Zilin Zhang (U87038789)
Mingyan Yang (U30215865)
Yujing Chen (U70567267)
Yicheng Li (U29503597)

1. Introduction

The Bay State Banner is an independent newspaper primarily geared toward the readership interests of the African-American community in Boston, Massachusetts. It owns news weekly that reports on the political, economic, social and cultural issues that are of interest to African American and English speaking Latinos in Boston and throughout New England. Started in 1965 by Melvin B. Miller, a graduate of Harvard College and Columbia Law School, the Banner provides its estimated 120,000 weekly print readers and 300,000 online readers with reports on a mix of local, national, international news and cultural events. The Banner is the newspaper of record for this community and has chronicled the struggles and successes of the last 50 years.

To carry out the political plans, politicians need a large amount of money to support the activities. And they need donors. Actually, the real estate industry is the largest donor to Massachusetts politicians. It covers brokers, developers, architects, and anyone with a vested investment in building. They acquire a large number of profits from the real estate industry and become the power of the politicians to support the politicians' activities.

The real estate industry covers many people who have a vested investment in building, including brokers, developers, architects and ect.. Our client Yawu Miller, a journalist writing for the Bay State banner would like to examine the lobbying power of the real estate industry in Massachusetts.

2. Data Collection & Processing

2.1 Data Source

Our data source are The Secretary of Commonwealth's website:

<http://corp.sec.state.ma.us/corpweb/CorpSearch/CorpSearch.aspx>

Our task is to collect all business filings for the past 10 years that specify "real estate" as the industry. We need to type "a" or "b" or "c", etc...then view filings, annual report, extract PDFs from all links; and extract names of bulist "real estate" as industry.

2.2 Scraping & Data Downloading

We use Python code to scrape the Secretary of Commonwealth's website.

Since we were not familiar with the scraping, we use selenium IDE to create the basic skeleton of the code. Then we modified the code to the one that works for our scraping. And we choose the Chrome driver as the driver scraping the data, so we download the Chrome driver to help the scraping.

You can find our Python code in the "code" folder. Just run the *Scraping.py* and we can download all the PDFs automatically.

2.3 Changing File Format

Just run the *Modify.py* in the "code" folder and we can change all the files format from ".aspx" into ".pdf".

2.4 Extracting

We have two kinds of dataset extracted from the downloaded PDFs. One only contains name and description and the other contains many other detailed information like address and member information. Our project manager also provided us OCPF data to do the filtering. All of them are saved in excel files.

2.5 Filtering & Data Preparation

We use the following keywords related to real estate to filter the two kinds of data from the secretary website using the column "description":

Realtor	Real Estate
Contractor	Construction
Builder	Building
Carpenter	Property
Painter	Property Management
Architect (ambiguous)	Surveyor
Developer (ambiguous)	Civil Engineer/ing
Planner	Realtor
Environmental	Realty
Reviewer	
Surveyor	

Also, we use the following name of PACS to filter the OCPF data using the name of PACS (This is only a part of it. We don't list all the names of PACS here.). We also use the keywords showing above to filter the OCPF data using columns "Occupation" and "Employer":

Greater Boston Real
Estate Board PAC
NAIOP MA Commercial
Real Estate
Development Political
Action Committee

Real Estate Bar
Association for MA PAC

We will show the filtered data in the next part.


3. Data Description & Visualization

You can view our filtered data here:

<https://drive.google.com/drive/folders/0AONGMu03aGUKUk9PVA>

3.1 PDFs from the Secretary Website

According to the PDFs we have downloaded, most PDFs have the same format like this image:

		The Commonwealth of Massachusetts		Minimum Fee: \$100.00	
		William Francis Galvin			
		Secretary of the Commonwealth, Corporations Division One Ashburton Place, 17th floor Boston, MA 02108-1512 Telephone: (617) 727-9640			
Annual Report (General Laws, Chapter 156D, Section 16.22; 950 CMR 113.57)					
Identification Number: <u>463158640</u>					
1. Exact name of the corporation: <u>A&B AUTO INC</u>					
2. Jurisdiction of Incorporation: State: <u>MA</u> Country:					
3,4. Street address of the corporation registered office in the commonwealth and the name of the registered agent at that office: Name: <u>WILLIAM J. PLASKO JR.</u> No. and Street: <u>10 GUILD ST</u> City or Town: <u>NORWOOD</u> State: <u>MA</u> Zip: <u>02062</u> Country: <u>USA</u>					
5. Street address of the corporation's principal office: No. and Street: <u>10 GUILD ST</u> City or Town: <u>NORWOOD</u> State: <u>MA</u> Zip: <u>02062</u> Country: <u>USA</u>					
6. Provide the name and addresses of the corporation's board of directors and its president, treasurer, secretary, and if different, its chief executive officer and chief financial officer.					
Title		Individual Name First, Middle, Last, Suffix		Address (no PO Box) Address, City or Town, State, Zip Code	
PRESIDENT		WILLIAM J PLASKO JR		12 ROSEDALE RD NORWOOD, MA 02062 USA	
PRESIDENT		WILLIAM J PLASKO JR		12 ROSEDALE RD NORWOOD, MA 02062 USA	
TREASURER		WILLIAM J PLASKO JR		12 ROSEDALE RD NORWOOD, MA 02062 USA	
SECRETARY		WILLIAM J PLASKO JR		12 ROSEDALE RD NORWOOD, MA 02062 USA	
VICE PRESIDENT		WILLIAM J PLASKO JR		12 ROSEDALE RD NORWOOD, MA 02062 USA	
DIRECTOR		WILLIAM J PLASKO JR		12 ROSEDALE RD NORWOOD, MA 02062 USA	
7. Briefly describe the business of the corporation: <u>OWN AND OPERATE A DRIVING BUSINESS AND FACILITIES</u>					
8. Capital stock of each class and series:					
Class of Stock	Par Value Per Share Enter 0 if no Par	Total Authorized by Articles of Organization or Amendments		Total Issued and Outstanding	

CNP	\$0.00000	Num of Shares 10,000	Total Par Value \$0.00	Num of Shares 10,000
-----	-----------	-------------------------	---------------------------	-------------------------

9. Check here if the stock of the corporation is publicly traded: ☐

10. Report is filed for fiscal year ending: 12/31/ 2016

Signed by WILLIAM J PLASKO JR, its PRESIDENT
on this 27 Day of April, 2017

© 2001 - 2017 Commonwealth of Massachusetts
All Rights Reserved

3.2 Extracted Data from PDFs

The two kinds of dataset collected are shown below.

3.2.1 The file contains name and description

Name	Description
JML HOLDING COMPANY I	REAL ESTATE
THE WALNUT HILL GROUP, LLC	PROVIDE REAL ESTATE BROKERAGE SERVICES AS WELL AS NOTARY SERVICES.
THE WARBY GROUP LLC	SCIENCE EDUCATION, AND ENVIRONMENTAL CONSULTING; AND ANYTHING ELSE LEGAL
THE WARREN GROUP, INC.	PUBLICATION OF REAL ESTATE & BANKING NEWS & DATA
THE WATER CONNECTION, LLC	REAL ESTATE HOLDING COMPANY
THE WATER PLANET COMPANY	ENVIRONMENTAL CONSULTING AND OPERATIONS
THE WATERMAN REALTY MANAGEMENT	COMMERCIAL REAL ESTATE MANAGEMENT
TWC LIQUIDATION, INC.	SPECIALTY SUBCONTRACTOR
THE WATERPROOFING COMPANY, LLC	SPECIALTY SUBCONTRACTOR WATERPROOFING, DAMPROOFING, CAULKING AND MASO
THE WAY CONSTRUCTION, INC	CONSTRUCTION
THE WEATHERIZE GUYZ, LLC	CONSTRUCTION
THE WECO GROUP, LTD.	REAL ESTATE
THE WEEPING BEECH, INC.	LANDSCAPE DESIGN & CONSTRUCTION
THE WELL FOR INTEGRATIVE MEDICINE, LLC	THE PURPOSES OF THE LIMITED LIABILITY COMPANY ARE TO RENT REAL PROPERTY AND
UL INFORMATION AND INSIGHTS INC.	OCCUPATIONAL AND ENVIRONMENTAL SAFETY & COMPLIANCE
THE WESTFORD LLC	SINGLE FAMILY HOME PURCHASE, RENOVATION, AND SALE. INVESTMENT PROPERTY.
THE WESTWOOD GROUP PROPERTY	PROPERTY OWNERSHIP, DESIGN AND CONSTRUCTION AND OTHER PURPOSES PERMITTED
THE WHALE LLC	REAL ESTATE
THE WHARF PROPERTIES LLC	REAL ESTATE
THE WHEELRIGHT, LLC	PURCHASING, MANAGING, AND SELLING REAL ESTATE, AND ANY OR ALL OTHER LAWFUL
THE WHITING-TURNER CONTRACTING COMPANY	CONSTRUCTION MANAGER/GENERAL CONTRACTOR
THE WILBUR GROUP, INC.	CONSTRUCTION/REMODELING
THE WILDER COMPANIES, LTD.	REAL ESTATE MANAGEMENT & LEASING
WILDER & CHAVEZ LLC	REAL ESTATE DEVELOPMENT
THE WILEMPAT, L.L.C.	PURCHASE, REPAIR, RENOVATE, RENT AND SELL ALL KINDS OF REAL ESTATE FOR INVEST
THE WILLI MANAGEMENT CORPORATION	MANAGEMENT OF REAL ESTATE
THE WILLIAM FALL GROUP, INC.	REAL ESTATE VALUATION
THE WILLIAM GERALD GROUP LLC	TO OWN, OPERATE, SELL AND DEVELOP REAL ESTATE
THE WILLIAMSVILLE CENTER, LLC	OWNERSHIP AND INVESTMENT IN REAL ESTATE AND OWNERSHIP AND OPERATION OF AD
THE WILMAT COMPANY, INC.	RESIDENTIAL/COMMERCIAL CONSTRUCTION AND REMODELING

3.2.2 The file contains many other detailed information

Name of Corporation	Street Address	Street Address of Corporation Principal Office	Registered Agent	Description	Member Title	Member Individual Name	Member Address
NETCOV MANAGED IT SERVICES LLC	3 ELECTRONICS	3 ELECTRONICS AVE UNIT 102, DANVERS, MA, 0192	TERRANCE L LEDGER	MANAGED SERVICE PROVIDER FOR INFORMATION TECHNOLOGY SERVICES	MANAGER	TERRANCE L LEDGER	19A BAYNS HILL ROAD BOXFORD, MA 01921 USA
NETCOV MANAGED IT SERVICES LLC	3 ELECTRONICS	3 ELECTRONICS AVE UNIT 102, DANVERS, MA, 0192	TERRANCE L LEDGER	MANAGED SERVICE PROVIDER FOR INFORMATION TECHNOLOGY SERVICES			
NETCOV MANAGED IT SERVICES LLC	3 ELECTRONICS	3 ELECTRONICS AVE UNIT 102, DANVERS, MA, 0192	TERRANCE L LEDGER	MANAGED SERVICE PROVIDER FOR INFORMATION TECHNOLOGY SERVICES			
NAIL STYLUS CORPORATION	55 OAKWOOD AVE 2, REVERE, MA, 02151, USA			NAIL CARE SERVICE	PRESIDENT	BILLY K HUYNH	55 OAKWOOD AVE, 2 REVERE, MA 02151 USA
NAIL STYLUS CORPORATION	55 OAKWOOD AVE 2, REVERE, MA, 02151, USA			NAIL CARE SERVICE	PRESIDENT	BILLY KIM HUYNH	55 OAKWOOD AVE #2 REVERE, MA 02151 USA
NAIL STYLUS CORPORATION	55 OAKWOOD AVE 2, REVERE, MA, 02151, USA			NAIL CARE SERVICE	PRESIDENT	BILLY KIM HUYNH	55 OAKWOOD AVE #2 REVERE, MA 02151 USA
NAIL STYLUS CORPORATION	55 OAKWOOD AVE 2, REVERE, MA, 02151, USA			NAIL CARE SERVICE	PRESIDENT	BILLY KIM HUYNH	55 OAKWOOD AVE #2 REVERE, MA 02151 USA
NAIL STYLUS CORPORATION	55 OAKWOOD AVE 2, REVERE, MA, 02151, USA			NAIL CARE SERVICE	TREASURER	BILLY KIM HUYNH	55 OAKWOOD AVE #2 REVERE, MA 02151 USA
NAIL STYLUS CORPORATION	55 OAKWOOD AVE 2, REVERE, MA, 02151, USA			NAIL CARE SERVICE	SECRETARY	BILLY KIM HUYNH	55 OAKWOOD AVE #2 REVERE, MA 02151 USA
NAIL STYLUS CORPORATION	55 OAKWOOD AVE 2, REVERE, MA, 02151, USA			NAIL CARE SERVICE	DIRECTOR	BILLY KIM HUYNH	55 OAKWOOD AVE #2 REVERE, MA 02151 USA
NATWELLS TRADING, INC.	8 MIKAILA WAY, WOBURN, MA, 01801, USA			ENERGY TRADING	PRESIDENT	HONGYANG LIU	8 MIKAILA WAYWOBURN, MA 01801 USA
NATWELLS TRADING, INC.	8 MIKAILA WAY, WOBURN, MA, 01801, USA			ENERGY TRADING	TREASURER	BECK CHEN	8 MIKAILA WAYWOBURN, MA 01801 USA
NATWELLS TRADING, INC.	8 MIKAILA WAY, WOBURN, MA, 01801, USA			ENERGY TRADING	SECRETARY	BECK CHEN	8 MIKAILA WAYWOBURN, MA 01801 USA
NATWELLS TRADING, INC.	8 MIKAILA WAY, WOBURN, MA, 01801, USA			ENERGY TRADING	DIRECTOR	HONGYANG LIU	8 MIKAILA WAYWOBURN, MA 01801 USA
NATWELLS TRADING, INC.	8 MIKAILA WAY, WOBURN, MA, 01801, USA			ENERGY TRADING	DIRECTOR	BECK CHEN	8 MIKAILA WAYWOBURN, MA 01801 USA
NORTHEAST FLEX HEATERS, INC.	251 FORGE RD., WESTPORT, MA, 02790, USA			MANUFACTURER	PRESIDENT	CHARLES MICHAUD	251 FORGE RD. WESTPORT, MA 02790 USA
NORTHEAST FLEX HEATERS, INC.	251 FORGE RD., WESTPORT, MA, 02790, USA			MANUFACTURER	PRESIDENT	CHARLES MICHAUD	251 FORGE RD. WESTPORT, MA 02790 USA

3.3 OCPF Data

After filtering OCPF data in the “Last_Name” column using the name of PACS, we can get the information under the “Committee” type as below.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Amount	First Name	Last Name	Exp1003	Address	City	State	Zip	Occupation	Employer	Contributor	Date	myDate	Report_ID	Full Name	District	District_Cod	Category	Report_Canv	Report_Canv	Report_Offic	Report_Comm	Name	
250.00		Sheet Metal Workers	2015-01-22	1157 Adams	Dorchester	MA	02124			Committee	2015-01-22	2015/1/22	215677	Timothy J T Toomey Jr., 26th Middle	283 D			Timothy J. Toomey Jr.		Cambridge I Toomey Committee			
500.00		New England Regional	2015-03-30	750 Dorches	Boston	MA	02125-1132			Committee	2015-03-30	2015/3/30	217668	Jason Palitsch Palitsch, Jas 11th Worces	354 N			Jason Palitsch		House 11th Palitsch Committee			
500.00		Assoc Builders & Cent	2015-05-19	100 Uniform	Woburn	MA	01801			Committee	2015-05-19	2015/5/19	203915	Hannah Kari Kane, Hann 11th Worces	354 N			Hannah Kane		House 11th Kane Committee			
250.00		MA & Northern NE La	2015-04-06	7 Labors	Hopkinton	MA	01748			Committee	2015-04-06	2015/4/6	537661	Susannah M Lee, Susan 2nd Franklin	241 N			Susannah M Lee		House 2nd Lee Committee			
500.00		MA & Northern NE La	2015-05-18	7 Labors	Hopkinton	MA	01748			Committee	2015-05-18	2015/5/18	538302	Denise Provost Provost, Den 27th Middle	284 N			Denise Provost		House 27th Provost Committee			
500.00		MA Labors District	2015-04-30	7 Labors	Hopkinton	MA	01748			Committee	2015-04-30	2015/4/30	538777	Dennis Rosa Rosa, Denn 4th Worces	347 N			Dennis Rosa		House 4th V Rosa Committee			
500.00		MA & Northern NE La	2015-05-22	7 Labors	Hopkinton	MA	01748			Committee	2015-05-22	2015/5/22	538882	Ellen Story Story, Ellen 3rd Hampsh	257 N			Ellen Story		House 3rd Story Committee			
250.00		MA Labors District	2015-04-21	7 Labors	Hopkinton	MA	01748			Committee	2015-04-21	2015/4/21	538945	Fred J. Barnc Barrows, Fre 1st Bristol	209 N			Fred J. Barrows		House 1st B Barrows Committee			
500.00		MA & Northern NE La	2015-04-06	7 Labors	Hopkinton	MA	01748			Committee	2015-04-06	2015/4/6	538961	Josh Outler Outler, Josh 6th Plymouth	317 N			Josh Outler		House 6th P Outler Committee			
125.00		Carpenters Local Uno	2015-06-03	21 Mazzo	Randolph	MA	02368			Committee	2015-06-03	2015/6/3	538961	Josh Outler Outler, Josh 6th Plymouth	317 N			Josh Outler		House 6th P Outler Committee			
250.00		Sheet Metal Workers	2015-05-15	1157 Adams	Dorchester	MA	02124			Committee	2015-05-15	2015/5/15	538961	Josh Outler Outler, Josh 6th Plymouth	317 N			Josh Outler		House 6th P Outler Committee			
400.00		Ma Assoc. of Realtors	2015-03-21	1400 Main S	Waltham	MA	02451			Committee	2015-03-21	2015/3/21	538961	Josh Outler Outler, Josh 6th Plymouth	317 N			Josh Outler		House 6th P Outler Committee			
100.00		Int'l Union of Operati	2015-06-15	16 Trotter D	Medway	MA	02053			Committee	2015-06-15	2015/6/15	538961	Josh Outler Outler, Josh 6th Plymouth	317 N			Josh Outler		House 6th P Outler Committee			
250.00		Plumbers Union Local	2015-06-17	1240 Messa	Boston	MA	02125			Committee	2015-06-17	2015/6/17	538961	Josh Outler Outler, Josh 6th Plymouth	317 N			Josh Outler		House 6th P Outler Committee			
500.00		MA & Northern NE La	2015-05-01	7 Labors	Hopkinton	MA	01748			Committee	2015-05-01	2015/5/1	538972	Ruth B. Balser Balser, Ruth 12th Middle	266 N			Ruth B. Balser		House 12th Balser Committee			
300.00		MA Assoc. of Realtors	2015-04-06	1400 Main S	Waltham	MA	02451			Committee	2015-04-06	2015/4/6	538978	Stephen Kulik Kulik, Steph 1st Franklin	240 N			Stephen Kulik		House 1st F Kulik Committee			
500.00		MA & Northern NE La	2015-04-18	7 Labors	Hopkinton	MA	01748			Committee	2015-04-18	2015/4/18	538978	Stephen Kulik Kulik, Steph 1st Franklin	240 N			Stephen Kulik		House 1st F Kulik Committee			
200.00		Real Estate Bar Assoc	2015-05-19	50 Congress	Boston	MA	02101			Committee	2015-05-19	2015/5/19	538980	Louis L. Kaf Kafka, Louis 8th Norfolk	304 N			Louis L. Kafka		House 8th N Kafka Committee			
200.00		Greater Boston Real E	2015-05-19	One Center	Boston	MA	02108			Committee	2015-05-19	2015/5/19	538980	Louis L. Kaf Kafka, Louis 8th Norfolk	304 N			Louis L. Kafka		House 8th N Kafka Committee			
500.00		MA & Northern NE La	2015-05-19	7 Labors	Hopkinton	MA	01748			Committee	2015-05-19	2015/5/19	538980	Louis L. Kaf Kafka, Louis 8th Norfolk	304 N			Louis L. Kafka		House 8th N Kafka Committee			
500.00		MA Labors District	2015-03-22	7 Labors	Hopkinton	MA	01748			Committee	2015-05-22	2015/3/22	538986	Angelo J. P. Puppulo Jr., 12th Hampc	253 N			Angelo J. Puppulo Jr.		House 12th Puppulo Jr. Committee			
100.00		Carpenters Local #10	2015-06-19	29 Oakland	Springfield	MA	01108			Committee	2015-06-19	2015/6/19	538986	Angelo J. P. Puppulo Jr., 12th Hampc	253 N			Angelo J. Puppulo Jr.		House 12th Puppulo Jr. Committee			
200.00		Pipefitters Local #53	2015-06-20	35 Travis Str	Allston	MA	02134			Committee	2015-06-20	2015/6/20	538986	Angelo J. P. Puppulo Jr., 12th Hampc	253 N			Angelo J. Puppulo Jr.		House 12th Puppulo Jr. Committee			
500.00		MA & Northern NE La	2015-05-26	7 Labors	Hopkinton	MA	01748			Committee	2015-05-26	2015/5/26	538987	Todd M. Smola Smola, Todd 1st Hampde	242 N			Todd M. Smola		House 1st H Smola Committee			
500.00		Massachusetts & Nort	2015-04-01	7 Labors	Hopkinton	MA	01748-2694			Committee	2015-04-01	2015/4/1	539023	Paul J. Donato Donato, Pas 36th Middle	292 N			Paul J. Donato		Friends of Paul J. Donato			
250.00		Sheet Metal Workers	2015-04-03	1157 Adams	Dorchester	MA	02124-5710			Committee	2015-04-03	2015/4/3	539023	Paul J. Donato Donato, Pas 36th Middle	292 N			Paul J. Donato		Friends of Paul J. Donato			
500.00		MA & Northern NE La	2015-03-18	7 Labors	Hopkinton	MA	01748			Committee	2015-03-18	2015/3/18	539040	Jennifer Ber-Benson, Jen 37th Middle	294 N			Jennifer Benson		House 37th Benson Committee			
125.00		Nixon Peabody Lip P	2015-06-01	1300 Clinton	Dorchester	MA	02124			Committee	2015-06-01	2015/6/1	539040	Jennifer Ber-Benson, Jen 37th Middle	294 N			Jennifer Benson		House 37th Benson Committee			
500.00		International Brotherh	2015-06-02	256 Freepor	Boston	MA	02122			Committee	2015-06-02	2015/6/2	539040	Jennifer Ber-Benson, Jen 37th Middle	294 N			Jennifer Benson		House 37th Benson Committee			
400.00		Ma Assoc. of Realtors	2015-06-09	1400 Main S	Waltham	MA	02451			Committee	2015-06-09	2015/6/9	539040	Jennifer Ber-Benson, Jen 37th Middle	294 N			Jennifer Benson		House 37th Benson Committee			
500.00		Pipefitters Local #53	2015-06-09	35 Travis Str	Allston	MA	02134			Committee	2015-06-09	2015/6/9	539040	Jennifer Ber-Benson, Jen 37th Middle	294 N			Jennifer Benson		House 37th Benson Committee			
500.00		Assoc Builders & Cent	2015-06-09	100 Uniform	Woburn	MA	01801			Committee	2015-06-09	2015/6/9	539040	Jennifer Ber-Benson, Jen 37th Middle	294 N			Jennifer Benson		House 37th Benson Committee			
200.00		Greater Boston Real E	2015-06-04	One Center	Boston	MA	02108			Committee	2015-06-04	2015/6/4	539040	Jennifer Ber-Benson, Jen 37th Middle	294 N			Jennifer Benson		House 37th Benson Committee			
200.00		MA Systems Contract	2015-06-06	1 Merchant	Sharon	MA	02067			Committee	2015-06-06	2015/6/6	539040	Jennifer Ber-Benson, Jen 37th Middle	294 N			Jennifer Benson		House 37th Benson Committee			
200.00		MA Building Trades	2015-06-04	256 Freepor	Dorchester	MA	02122			Committee	2015-06-04	2015/6/4	539147	Denise Garlick Garlick, Den 13th Norfolk	309 N			Denise Garlick		House 13th Garlick Committee			
500.00		MA Labors District	2015-03-25	7 Labors	Hopkinton	MA	01748			Committee	2015-03-25	2015/3/25	539147	Denise Garlick Garlick, Den 13th Norfolk	309 N			Denise Garlick		House 13th Garlick Committee			
500.00		MA & Northern NE La	2015-06-11	7 Labors	Hopkinton	MA	01748			Committee	2015-06-11	2015/6/11	539142	Cory Adkins Adkins, Cory 14th Middle	271 N			Cory Adkins		House 14th Adkins Committee			
500.00		MA & Northern NE La	2015-04-03	7 Labors	Hopkinton	MA	01748			Committee	2015-04-03	2015/4/3	539183	Tom Sannicandri Sannicandri, 7th Middle	264 N			Tom Sannicandri		House 7th N Sannicandri Committee			
200.00		Associated Subcontr	2015-06-27	31 State Str	Boston	MA	02109			Committee	2015-06-27	2015/6/27	539183	Tom Sannicandri Sannicandri, 7th Middle	264 N			Tom Sannicandri		House 7th N Sannicandri Committee			
100.00		Pipefitters Local #53	2015-02-27	35 Travis Str	Allston	MA	02134			Committee	2015-02-27	2015/2/27	539249	James Murphy Murphy, Jan 4th Norfolk	300 N			James Murphy		House 4th N Murphy Committee			
100.00		Ironworkers Union Lo	2015-03-05	195 Old Col	South Boston	MA	02127			Committee	2015-03-05	2015/3/5	539249	James Murphy Murphy, Jan 4th Norfolk	300 N			James Murphy		House 4th N Murphy Committee			
250.00		International Brotherh	2015-03-09	256 Freepor	Boston	MA	02122			Committee	2015-03-09	2015/3/9	539249	James Murphy Murphy, Jan 4th Norfolk	300 N			James Murphy		House 4th N Murphy Committee			
250.00		U.A. Plumbers and Ga	2015-03-16	1240 Messa	Boston	MA	02125			Committee	2015-03-16	2015/3/16	539249	James Murphy Murphy, Jan 4th Norfolk	300 N			James Murphy		House 4th N Murphy Committee			
100.00		Labors Local 133 Pro	2015-04-24	P.O. Box 69	Quincy	MA	02269			Committee	2015-04-24	2015/4/24	539249	James Murphy Murphy, Jan 4th Norfolk	300 N			James Murphy		House 4th N Murphy Committee			
500.00		MA & Northern NE La	2015-04-24	7 Labors	Hopkinton	MA	01748			Committee	2015-04-24	2015/4/24	539249	James Murphy Murphy, Jan 4th Norfolk	300 N			James Murphy		House 4th N Murphy Committee			

4. Analysis

Currently we are waiting for Team 1 to merge OCPF data and their data so that we could move to the next step.

When analyzing, we want to figure out which real estate company donated most. To do this, we would try to build a model to estimate the relationship between the influence and the donations, address, descriptions and so on. So when the model is built, we can use it to determine which corporation or individual is the most influential. Also, we still have to compare the donations from PACS/unions and individuals.

5. Data Presentation

To clearly see the influence of the real estate corporation of the whole state, we will use charts like histogram or line chart to list our corporations. However, we are still waiting for team 1 to merge OCPF datasets to a fully datasets.

6. Project Complexity and Team Creativity

6.1 Challenges and obstacles

- 1) Because of the huge amount of data, we cannot go to the next page when searching for some letters who contain too many links. We solved this problem by choosing to show all links when searching the letters.
- 2) When we were scraping the website, at first we were using Octoparse to scrape the website since it is very convenient. However, we were rejected by the website due to the high frequency of visiting. Then we wrote our own Python code to do the scraping and added a random *sleep()* in the code to reduce the frequency of visiting. Also, we tried to restart the chrome driver every 100 times downloading so that we wouldn't be rejected. Then we could download PDFs without rejecting.
- 3) Another problem is that there are millions of PDFs to download. We have been downloading for a whole week and we got only about 100,000 PDFs. It may take more than a month to download all the PDFs. Then our project manager provided us the data from Campaign Finance Team since we are using the same website and data.

6.2 Team Creativity

When solving the rejecting problem, we find our own way to avoid being rejected. Adding *sleep()* and restarting chrome driver every 100 downloads can solve the problem well.

Up to now, our clients do not have any dissatisfaction.