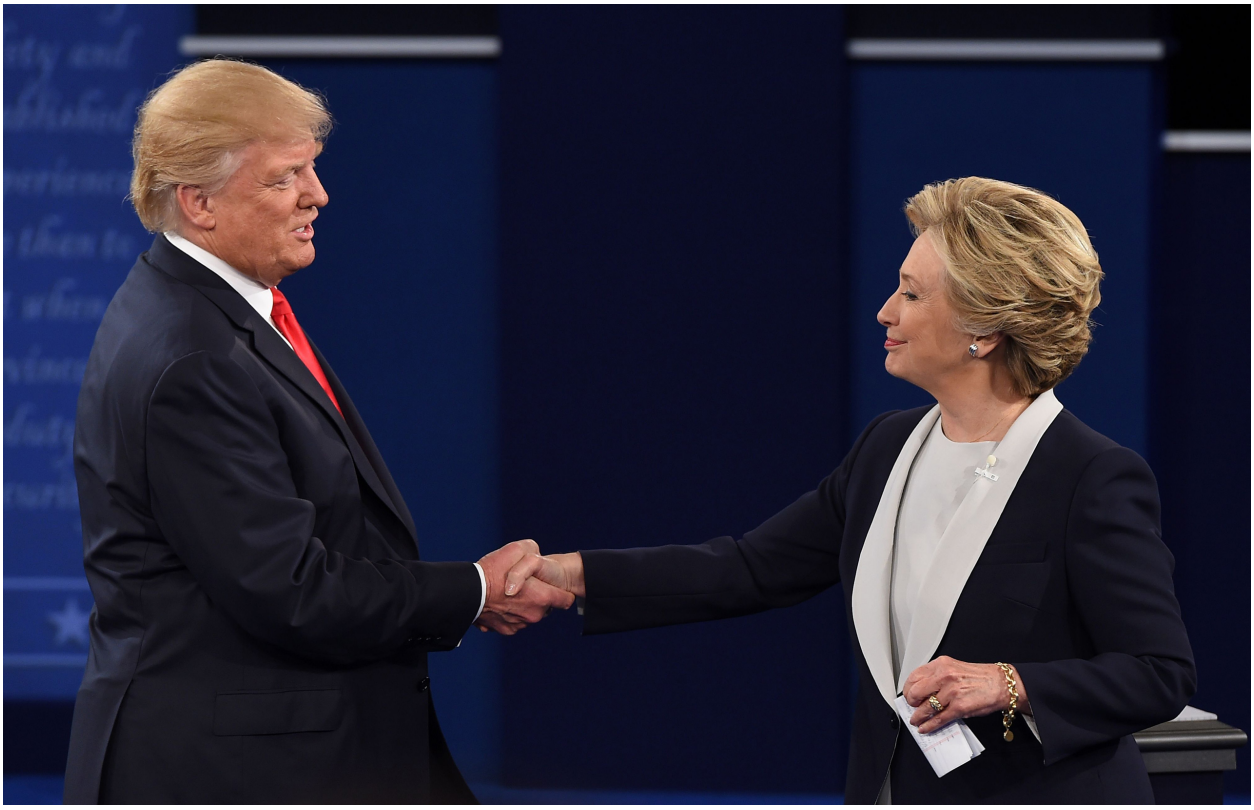
Data analysis on FEC report of presidential campaign finances in 2016



Course: MSBA 305

Team paper Project

Instructor: Dr. Mayank Johri

Team members: Sadegh Bamohabbat Chafjiri, Peiwen Li, Ramin Rafat

November 9th 2018

Introduction

In this project, we will perform the statistical analysis including descriptive, predictive, and prescriptive analysis on Federal Election Commission report. “The Federal Election Commission (FEC) is an independent regulatory agency established in 1975 to administer and enforce the Federal Election Campaign Act (FECA), which requires public disclosure of campaign finance information. The FEC publishes campaign finance reports for presidential and legislative election campaign candidates on the Campaign Finance Disclosure Portal." (Federal Election Commission, 2017). The FEC report includes a summary report of presidential campaign finances for election 2016. In this project, we perform financial analysis to measure the main attributes. Moreover, We do correlation analysis to find relationship between effective attributes in which finalist candidates have considerable proportion. Finally, we fit a regression model on 2 main attributes.

Data Collection and Description:

The Federal Election Commission (FEC) report had 2 appropriate datasets related to election campaign finances in 2016. We found our data set in csv format on Kaggle. The available data sets were expenditures and Summary. In “expenditures” file, data set included all of operating expenditures by campaign committee and cost of campaigning in different states including advertising, traveling, advertising, staffing and etc. (Federal Election Commission, 2017). In “summary” file, there were less measures such as total offsets, Cash on hand, total Loan, Net operating expenditures and etc.

By comparison the files, we found that summary file included the appropriate attributes for statistical analysis and as a suitable file satisfied the team project purposes for 2 two reason; Firstly, It had a desired file size (252 * 91) while expenditures file had a huge size in range of (65.5K* 16) out of the team paper scale. Secondly, summary file included the BI data which can make a good model to predict the relationship between different financial attributes. Summary file had a multivariate and Time-Series Characteristics. It had one record for each financial report by the presidential campaign committees during the 2016 primary and general election campaigns. It is valuable to add that we reduced the columns which were not related to our analysis in the summary file.

Descriptive Statistics:

In this section, we did the filtering process on all factors of presidential campaign finances. We looked for the main factors in which the finalist presidential candidates, Donald Trump and Hillary Clinton, had considerable proportions. In other words, We kept those financial attitudes helped finalist presidential candidates to have a better result.

Filtering the attributes: In the first step, by use of bar chart in Microstrategy tool, we performed the filtering process on all financial attributes and determined the main attributes in which Donald Trump's and Hillary Clinton's campaign had the major shares. Among those attributes, it seems Donald Tramp had considerable shares in *total loans, Committee payables, Candidate Contribution, net operating expenditures and Total offsets* which might positively impact on his result of election. We compared main candidates' campaign finances by bar-chart. As you see in fig. 1 to fig. 5, Donald Trump undeniably had a bigger share in *total loans, Committee payables, Candidate Contribution a*mong the all attributes *while* Hillary Clinton only relatively had a bigger proportion in *net operating expenditures and total offsets in comparison with Trump or Sanders*. Interestingly, among the main attributes, the total offset had a ratio of similarity with final result of election.

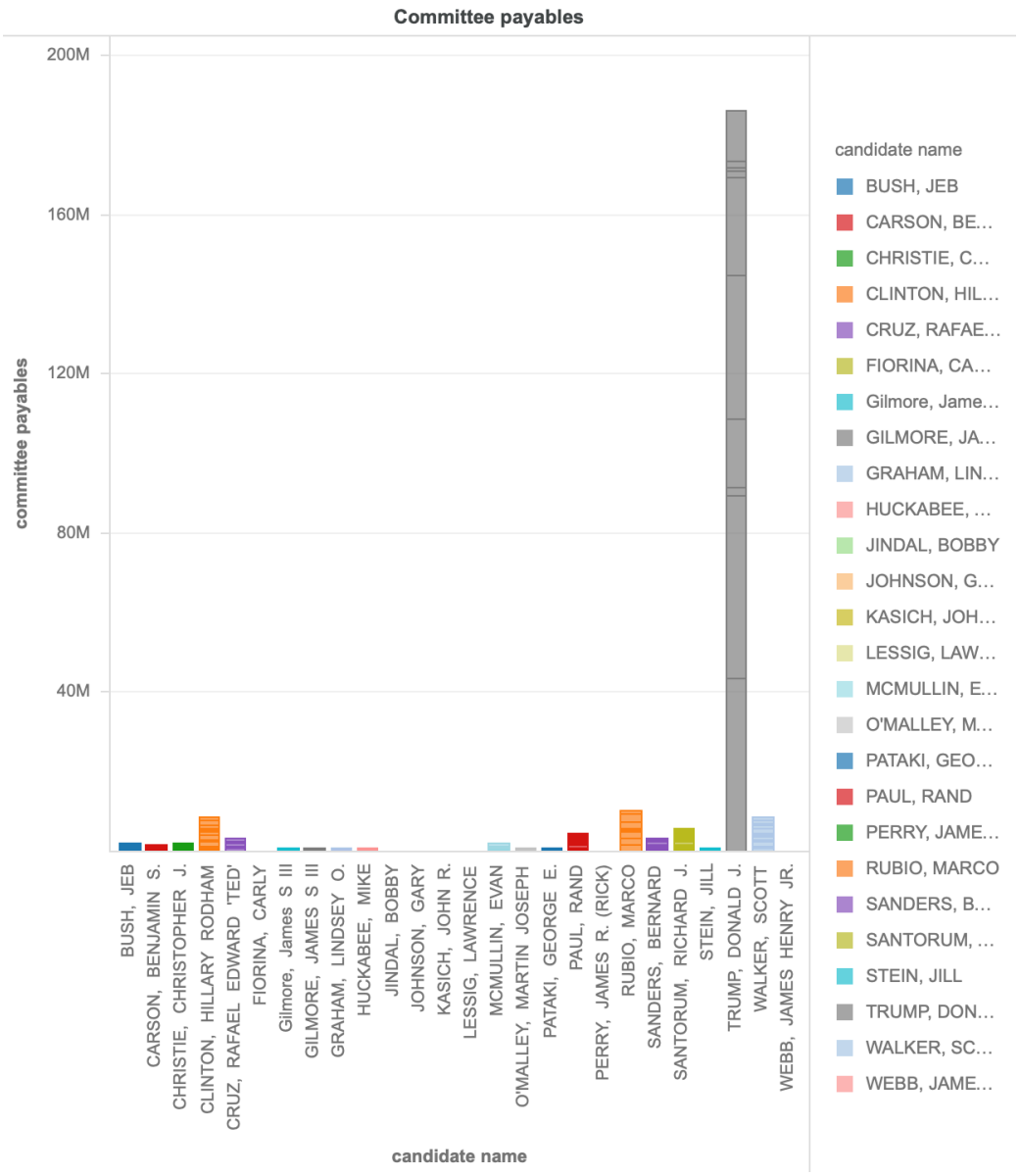


Fig 1. Total loans

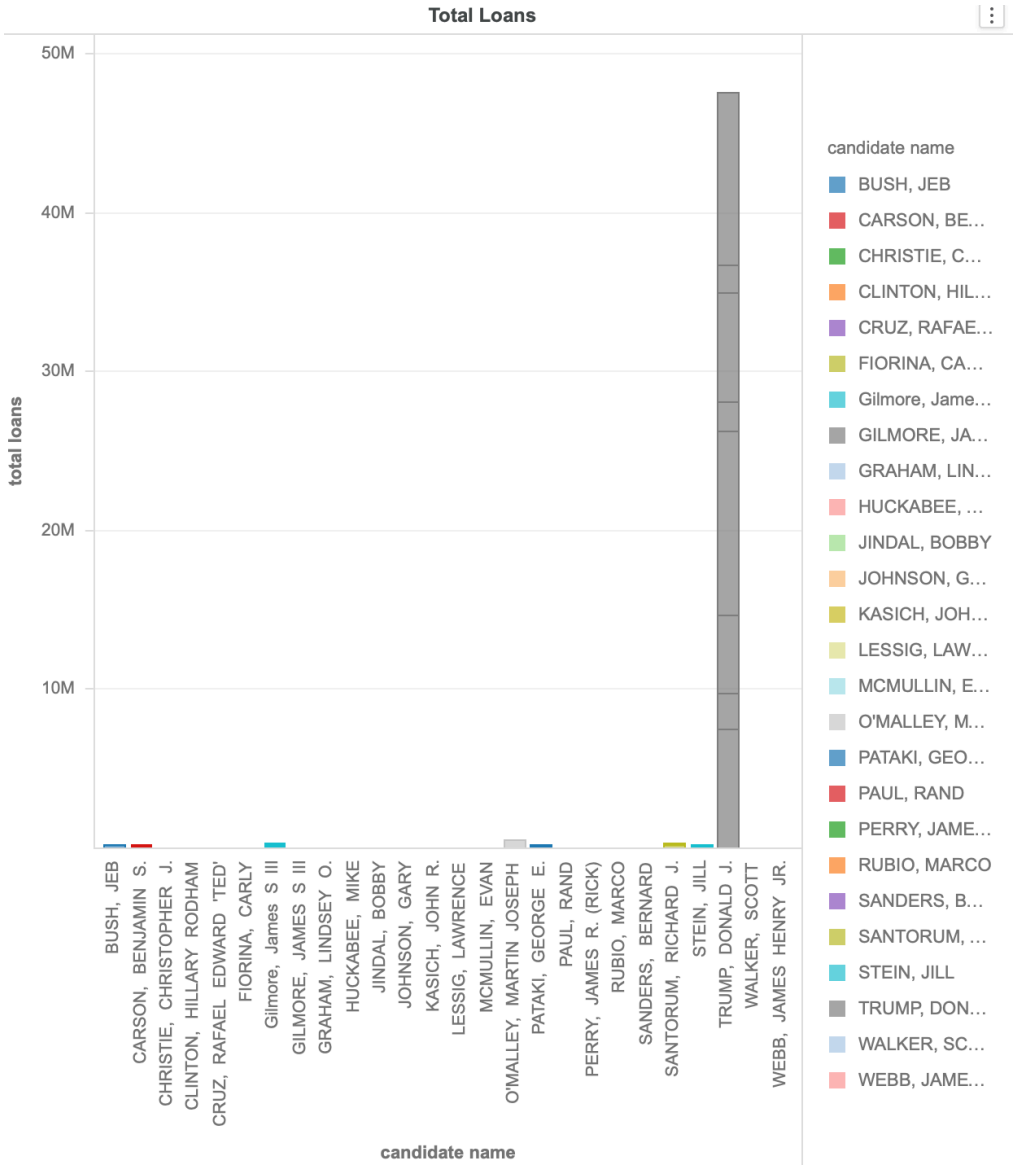


Fig 2. Committee payables

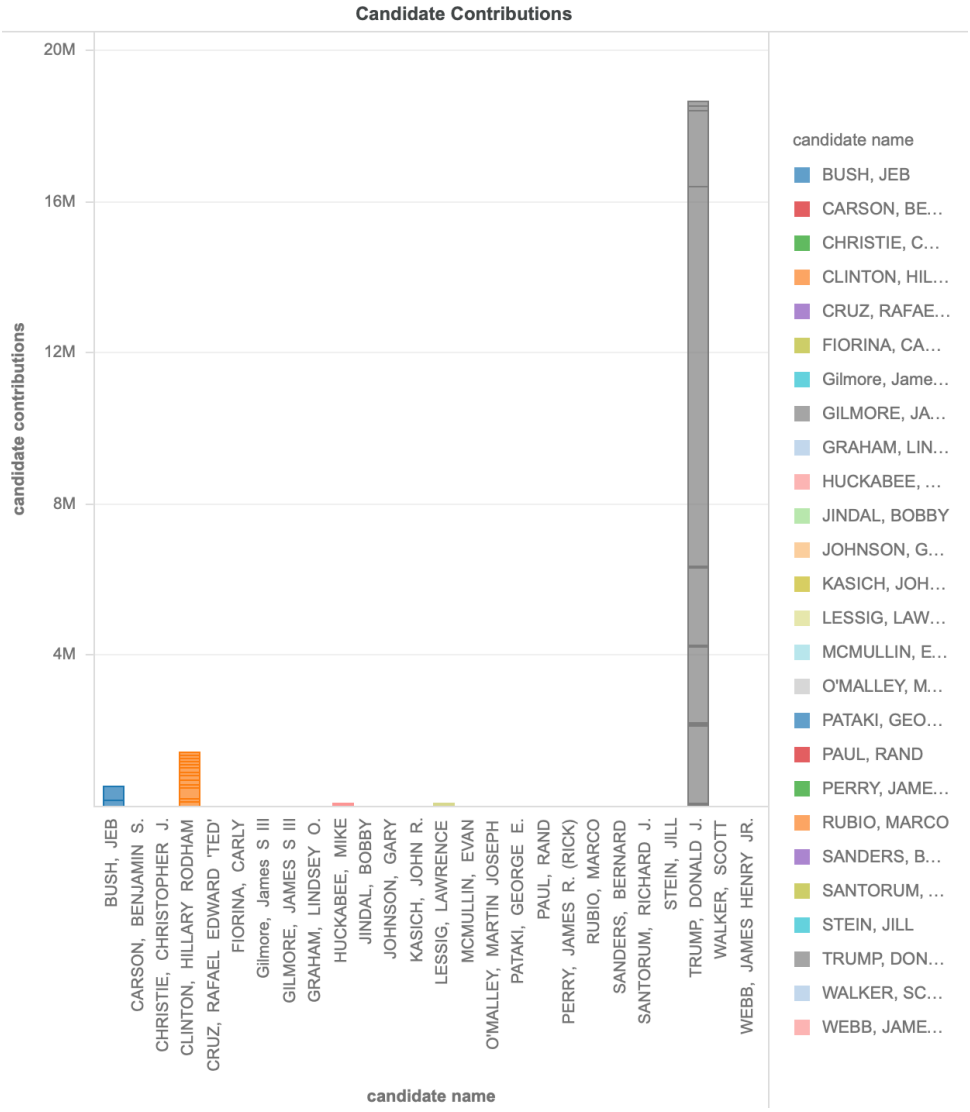


Fig 3. Candidate contributions

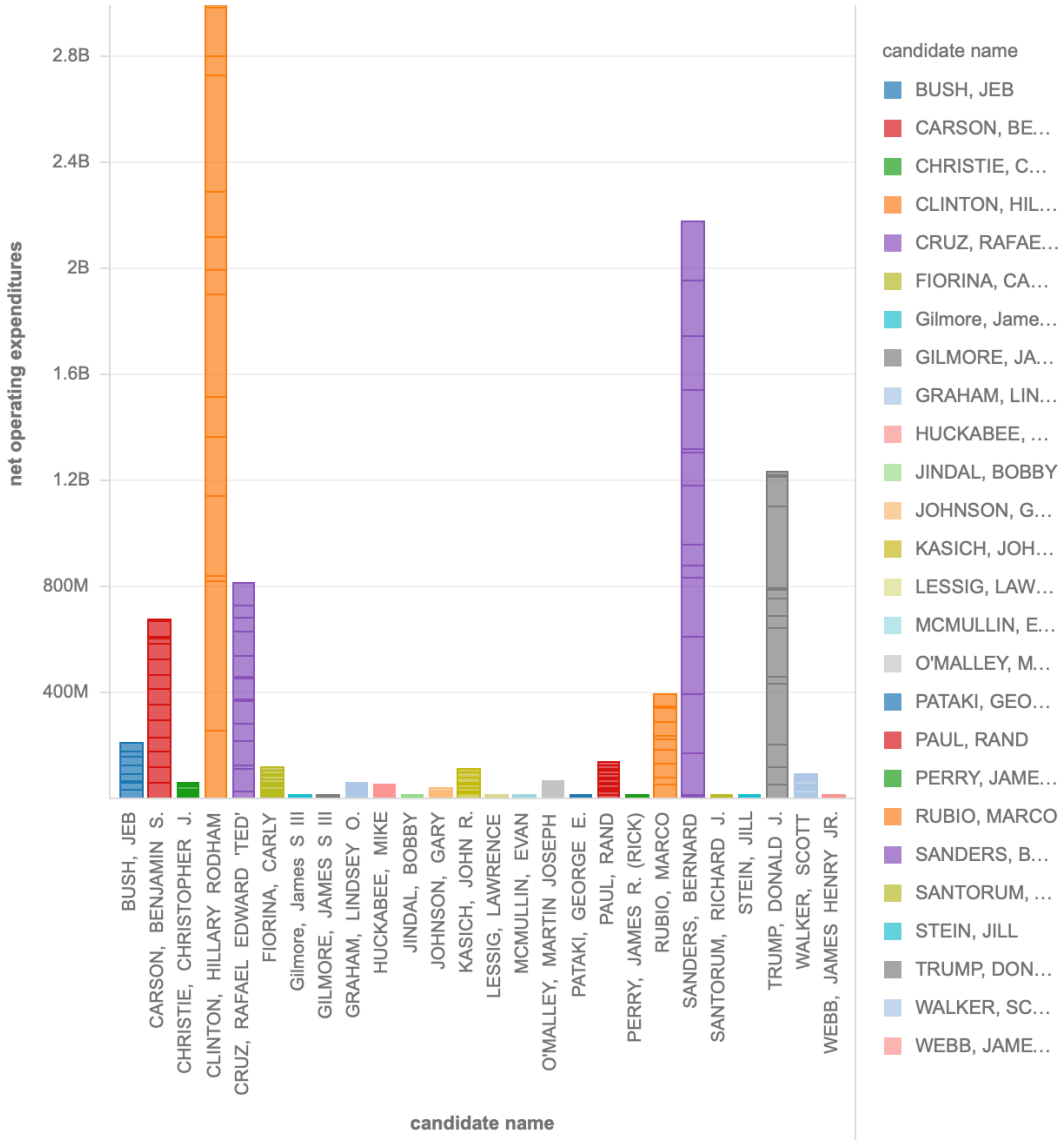


Fig 4. Net Operating expenditure

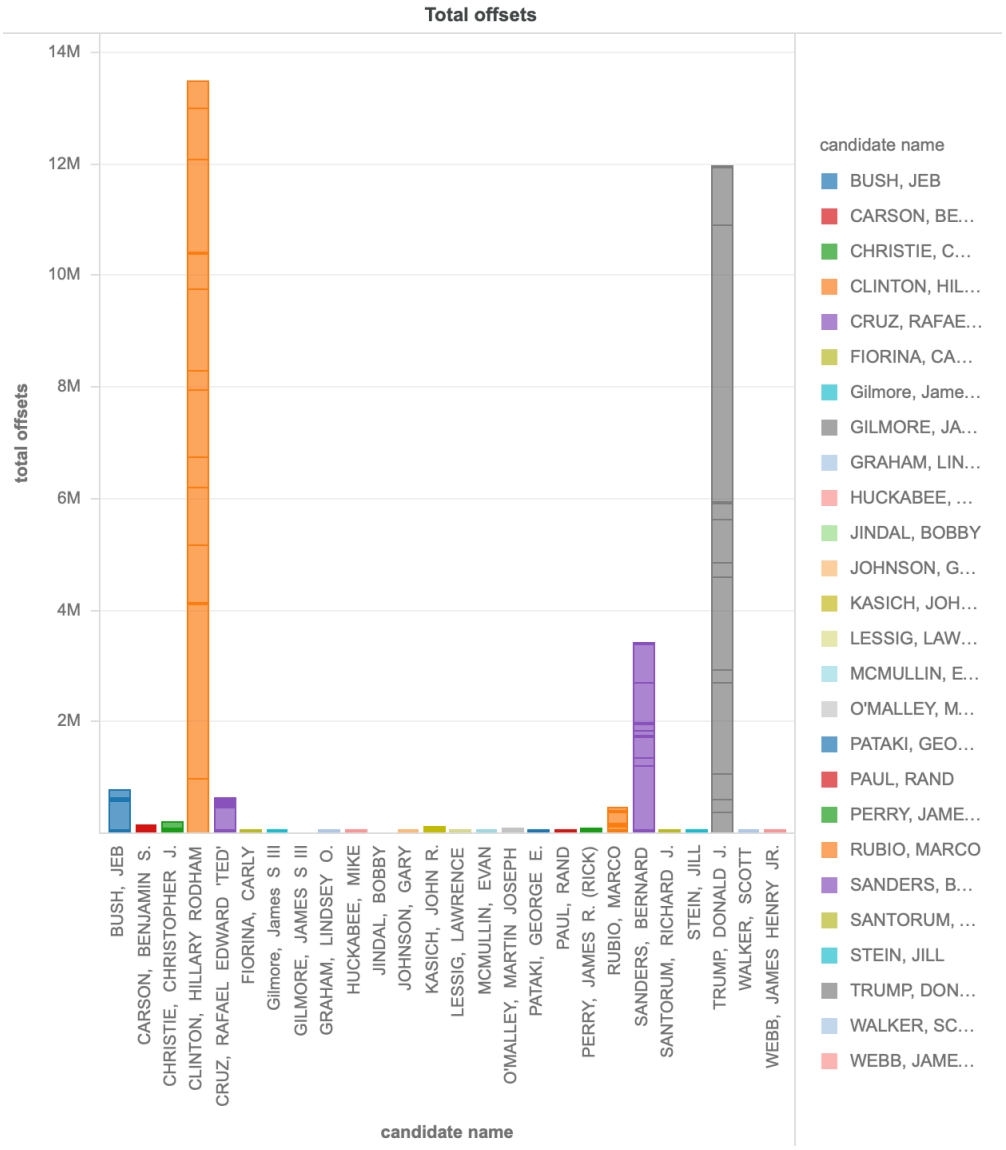


Fig 5. Total offsets

Moreover, in table 1, we showed a summary of statistical characteristics for 4 main attributes i.e. *Total loans, Committee payables, candidate contribution and net operating expenditure*. The table presented the characteristics of attributes including *Standard deviation, median, mean, minimum, maximum,1st, 2nd and 3rd quartile values*.

Moreover, it showed that net operating expenditures was the attribute with the widest range of finance from 9.7*10^3 to 5.6* 10^8 while the candidate contribution had the smallest interval between 0 to 10^7.

Table 1. Summary on main attributes

	total_loans	com_payables	CandContrib	net_oper_expen
count	2.520000e+02	2.520000e+02	2.520000e+02	2.520000e+02
mean	1.943118e+05	9.605061e+05	8.194392e+04	3.691930e+07
std	1.231151e+06	4.951185e+06	6.800684e+05	7.530228e+07
min	0.000000e+00	0.000000e+00	0.000000e+00	9.740090e+03
25%	0.000000e+00	0.000000e+00	0.000000e+00	1.434390e+06
50%	0.000000e+00	4.554070e+04	0.000000e+00	7.282073e+06
75%	0.000000e+00	4.235328e+05	0.000000e+00	3.403114e+07
max	1.154212e+07	4.570319e+07	1.003364e+07	5.629233e+08

In Fig 6, the box plot schematically depicted the mean, 1st, 2nd and 3rd quartile values for main attributes. The scale of graph is logarithmic in range of 10^8. The graph showed the mean, maximum and minimum values for each of attributes. Maximum, minimum and outlier Values for total loans, Committee payables, candidate contributions were very compressed in logarithmic scale (base 10^8) in comparison with net operating expenditures. As we mentioned in prior, box-plot confirmed that net operating

expenditure had the widest range of finance and outliers. Third-forth of Net operating expenditure values is less than 7.5 *10^7.

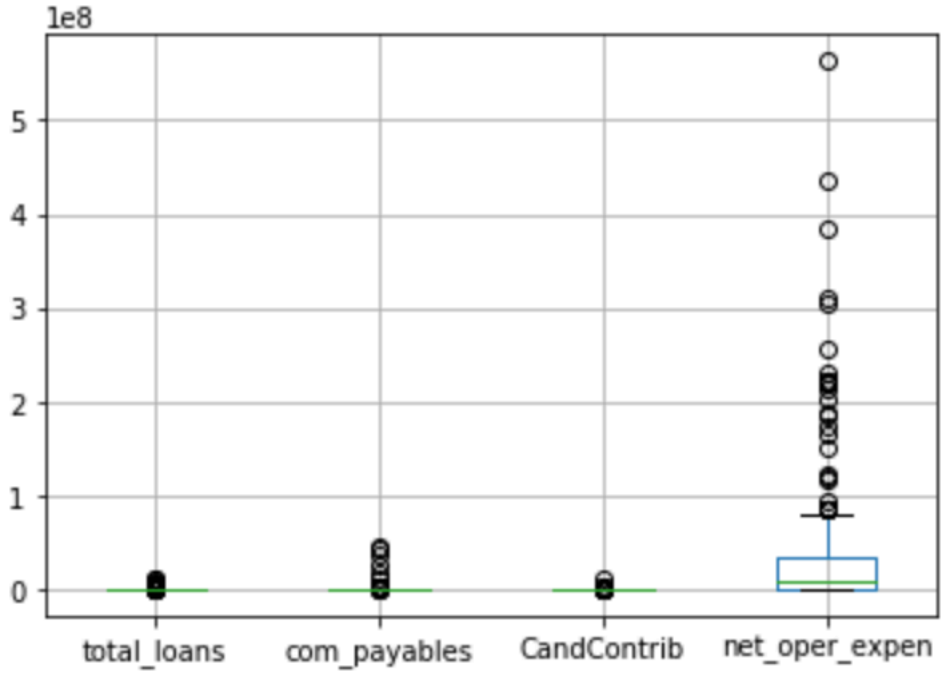


Fig 6. box plot for main attributes

Correlation Analysis

we performed the correlation analysis for all attributes. The table size was too big. Therefore, We could only show the main attributes in Fig 7. It depicted that the net operating expenditures, total offsets, Total loans and committee payables had a highly correlation with most of attributes whereas candidate contributions had a weakly correlation with most of attributes. In addition, in fig 8, we did the correlation analysis on main attributes. It can be seen that among the main attributes, the committee payable

and total loans were only attributes with high ratio of correlation. Other attributes were not highly correlated with each other.

	total_offsets	net_operating_expenditures	candidate_contributions	total_loans	committee_payables
total_offsets	1	0.58	0.73	-0.33	-0.29
net_operating_expenditures	0.58	1	0.17	-0.42	-0.37
candidate_contributions	0.73	0.17	1	-0.16	-0.17
total_loans	-0.33	-0.42	-0.16	1	0.73
committee_payables	-0.29	-0.37	-0.17	0.73	1
cashonhand_start	0.36	0.82	-0.077	-0.58	-0.53
total_receipts	0.72	0.79	0.5	-0.5	-0.52
subtotal	0.59	0.87	0.23	-0.58	-0.57
total_disbursements	0.8	0.87	0.38	-0.39	-0.37
cashonhand_end	-0.0074	0.43	-0.093	-0.55	-0.55
net_contributions	0.41	0.93	-0.029	-0.51	-0.48
individual_contributions	0.43	0.63	0.11	-0.62	-0.6
political_party_contributions	0.36	0.41	0.38	-0.24	-0.23
other_political_contributions	0.079	0.053	-0.1	-0.4	-0.37
total_contributions	0.52	0.64	0.24	-0.63	-0.61
transfer_from_committee	0.68	0.74	0.58	-0.41	-0.4
candidate_loans	-0.33	-0.42	-0.16	1	0.73
operating_exp_offsets	1	0.58	0.73	-0.33	-0.29
other_receipts	0.25	0.5	-0.23	-0.36	-0.34
total_receipts.1	0.72	0.79	0.5	-0.5	-0.52
operating_expenditures	0.8	0.87	0.38	-0.39	-0.37
transfer_to_committees	-0.2	-0.24	-0.069	0.48	0.052
individual_refunds	0.4	0.71	-0.085	-0.49	-0.46
other_political_refunds	0.56	0.34	0.74	-0.25	-0.23
total_contribution_refunds	0.41	0.71	-0.081	-0.49	-0.46
other_disbursements	0.85	0.36	0.88	-0.14	-0.13
total_disbursements.1	0.8	0.87	0.38	-0.39	-0.37
individual_contributions_todate	0.38	0.93	-0.06	-0.5	-0.47
political_party_contributions_todate	0.52	0.9	0.27	-0.35	-0.34
other_political_contributions_todate	0.19	0.74	-0.26	-0.48	-0.45
candidate_contributions_todate	0.76	0.29	0.89	-0.24	-0.23
total_contributions_todate	0.41	0.93	-0.031	-0.51	-0.48
transfer_from_committee_todate	0.65	0.94	0.35	-0.37	-0.36
candidate_loans_todate	0.2	-0.19	0.48	0.27	0.41
total_loans_todate	0.2	-0.19	0.48	0.27	0.41
operating_exp_offsets_todate	0.65	0.97	0.3	-0.48	-0.42
total_offsets_todate	0.65	0.97	0.3	-0.48	-0.42
other_receipts_todate	0.2	0.85	-0.24	-0.37	-0.36
total_receipts_todate	0.55	0.99	0.16	-0.47	-0.42
operating_expenditures_todate	0.58	1	0.18	-0.43	-0.37
transfer_to_committees_todate	0.076	-0.32	0.38	0.54	0.52
individual_refunds_todate	0.24	0.86	-0.21	-0.42	-0.4
other_political_refunds_todate	0.31	0.86	-0.14	-0.5	-0.47
total_contribution_refunds_todate	0.24	0.86	-0.21	-0.42	-0.4
other_disbursements_todate	0.85	0.63	0.75	-0.27	-0.25
total_disbursements_todate	0.58	1	0.17	-0.43	-0.37

Fig 7. correlation analysis between all attributes

	total_loans	com_payables	CandContrib	net_oper_expen
total_loans	1	0.78	0.0099	-0.0049
com_payables	0.78	1	-0.0038	0.032
CandContrib	0.0099	-0.0038	1	0.28
net_oper_expen	-0.0049	0.032	0.28	1

Fig 8. correlation analysis between main attributes

In table 2, we performed the correlation analysis on 4 main attributes of Summary report. The available tool of “correl” on data sheet of Google Drive presented the the ratio of correlation between two finalist presidential candidates. It showed that candidates were moderately correlated in total offsets, net operating expenditures and committee payable. Instead, they were weekly correlated in candidate contributes. In addition, they were independent in total loan because all values of total loan for Hillary Clinton were zero whereas Donald Trump had non-zero amounts.

Table 2. correlation analysis between Donald Trump and Hillary Clinton

fx total_receipts

	A	B	C	D	E
1	committee_name	total_offsets	net_operating_expenditures	committee_payables	candidate_contributions
2	HILLARY	982685.91	256850956.9	490632.09	121574.57
3	HILLARY	3146574.22	562923302.4	0	99703.2
4	HILLARY	150	18203538.26	574147.67	278821.09
5	HILLARY	1045858.64	304828517.2	214311.13	80340.98
6	HILLARY	1025126.72	219985867.8	144099.98	97735.8
7	HILLARY	529342.26	151181385.6	953048.57	128271.18
8	HILLARY	1208539.43	385715359.7	626094.77	101077.41
9	HILLARY	357537.34	93419002.61	1056361.42	99889.78
10	HILLARY	1462703.18	123305355	923628.74	92946.3
11	HILLARY	641411.9	174075255.7	585702.63	107308.25
12	HILLARY	14127.85	435335310.5	111237.9	50479.9
13	HILLARY	1649910.69	74037916.21	978863.24	89325.83
14	HILLARY	936217.79	186812338	1168389.38	102860.92
15	HILLARY	470644.13	43059983.57	647245.1	0
16	TRUMP	381890.57	55080299.12	43454544.39	32711.68
17	TRUMP	222223.58	61593101.69	45703185.39	45325.27
18	TRUMP	451752.44	86605214.58	0	2052825.34
19	TRUMP	1632121.7	231365922.3	2086572.41	30681.67
20	TRUMP	234581.56	23433901.79	17534058.41	31751.95
21	TRUMP	1662600.81	183805879.9	0	2036805.19
22	TRUMP	260725.84	46109417.28	35926174.39	35095.63
23	TRUMP	774231.7	68613970.5	0	2046629.75
24	TRUMP	303266.53	32587098.66	24384058.41	32056.68
25	TRUMP	1364.25	1413310.04	1804747.23	4049.64
26	TRUMP	4965180.88	310202386.3	766756.67	10033640.55
27	TRUMP	1065329.19	115430715.4	0	2037118.08
28	TRUMP	7695.47	5440690.92	1804747.23	100779.63
29	TRUMP	258.26	12171475.52	12620297.41	113737.74
30	Correlation (Hillary,Trump)	-0.3491613056	0.3173803122	-0.3746512731	-0.08340374591

Statistical tests:

In this section, we performed ANOVA (analysis of variance) to evaluate difference between the mean of total offsets and other attributes. Moreover, we performed ANOVA between net operating expenditures and other attributes too. As you know, ANOVA is applicable when we have more than two groups of data which t.test does not work properly and applying one way ANOVA is faster than t.test to compare data groups by means and find the statistical significance. In table 3, we performed ANOVA on the attributes.

Null Hypothesis: There was no difference between mean of total offsets and attributes in below:

```
federal_funds,individual_contributions,political_party_contributions,other_po
litical_contributions,candidate_contributions,total_contributions,transfer_fr
om_committee,candidate_loans,other_loans,total_loans,operating_exp_offsets,le
gal_acct_offsets,total_offsets,other_receipts,total_receipts,federal_funds_to
_date,individual_contributions_todate,political_party_contributions_todate,ot
her_political_contributions_todate,candidate_contributions_todate,total_contr
ibutions_todate,transfer_from_committee_todate,candidate_loans_todate,other_l
oans_todate,total_loans_todate,operating_exp_offsets_todate,total_offsets_tod
ate,other_receipts_todate,total_receipts_todate
```

As you know, 0.05 is a special value for statistical significance in data analytics. If the p-value is less than 0.05, null hypothesis is rejected and it means there is a statistical significance between assumed groups. In this case, there was no difference between mean of total offsets and aforementioned attributes because p-values were less than 0.05 and

null hypothesis were rejected. In other words, there was a statistical significance between total offsets and attributes(see table 3).

table 3: analysis of variance between total offsets and other attributes

	PR(>F)
federal_funds	0.000000
individual_contributions	0.000000
political_party_contributions	0.000000
other_political_contributions	0.000000
candidate_contributions	0.000000
total_contributions	0.000000
transfer_from_committee	0.000000
candidate_loans	0.000000
other_loans	0.000000
total_loans	0.000000
operating_exp_offsets	0.000000
legal_acct_offsets	0.000000
total_offsets	0.000000
other_receipts	0.000000
total_receipts	0.000000
federal_funds_to_date	0.000000
individual_contributions_todate	0.000000
political_party_contributions_todate	0.000000
other_political_contributions_todate	0.000000
candidate_contributions_todate	0.000000
total_contributions_todate	0.000000
transfer_from_committee_todate	0.000000
candidate_loans_todate	0.000000
other_loans_todate	0.000000
total_loans_todate	0.000000
operating_exp_offsets_todate	0.000000
total_offsets_todate	0.000000
other_receipts_todate	0.000000
total_receipts_todate	0.000000

Null Hypothesis: There is no difference between mean of net operating expenditures and attributes in below:

table 4: analysis of variance between net operating expenditures and other attributes

	PR(>F)
subtotal	4.109215e-03
total_disbursements	4.101375e-03
cashonhand_end	4.070104e-03
net_contributions	7.435768e-289
federal_funds_to_date	1.640971e-219
total_contributions_todate	3.060779e-289
transfer_from_committee_todate	4.928874e-281

```
candidate_loans_todate             1.066030e-210
other_loans_todate                 9.582391e-159
total_loans_todate                 1.452176e-258
total_offsets_todate               1.011706e-293
other_receipts_todate              6.877940e-275
total_receipts_todate              4.058168e-287
operating_expenditures_todate       3.374753e-95
transfer_to_committees_todate       7.708108e-87
fundraising_disbursements_todate    1.098939e-07
candidate_loan_repayments_todate    1.221378e-09
other_loan_repayments_todate        2.474082e-22
total_loan_repayments_todate        2.175427e-59
total_contribution_refunds_todate  3.835637e-234
other_disbursements_todate          2.151744e-86
total_disbursements_todate          5.496102e-87
```

In other word, because of being less than 0.05, p-valuses made null hypothesis rejected and there was a statistical significance between net operating expenditures and attributes in table 4.

Multiple Regression Analysis:

In this section, we performed the regression analysis to model total offsets and net operating expenditures. As we evaluate before, Donald Trump and Hillary Clinton had considerable shares in total offsets and net operating expenditures. We selected total offset and net operating expenditures for two reasons; Firstly, total offsets and net operating expenditures respectively behaved moderately and highly correlated with majority of attributes. Secondly, they had approximately a similar graph with result of election (specially total offset). It might help decision makers get an idea about effective attributes to model the result of election. Our attributes are discrete, therefore the only way to build a model is to perform multiple regression by factorizing technique. Applied technique in this project is backward selection method in Ordinary least squares (OLS). We removed the attributes having p-value more than 0.05 and the attributes having small p-value less than 0.05 were remained (see table 5). Moreover, We accepted the model if R-squared was more than 0.75. we made a regression model and checked the significancy of each attribute through the p-value.

Table5. Regression analysis for total offsets

OLS Regression Results

Dep. Variable:	total_offsets	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	1.443e+25
Date:	Mon, 12 Nov 2018	Prob (F-statistic):	0.00
Time:	19:30:41	Log-Likelihood:	3355.0
No. Observations:	252	AIC:	-6666.
Df Residuals:	230	BIC:	-6588.
Df Model:	21		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.065e-09	3.68e-08	0.029	0.977	-7.14e-08	7.35e-08
federal_funds	-0.1022	4.19e-12	-2.44e+10	0.000	-0.102	-0.102
individual_contributions	-0.0204	1.52e-11	-1.35e+09	0.000	-0.020	-0.020
political_party_contributions	-0.0204	5.82e-11	-3.51e+08	0.000	-0.020	-0.020
other_political_contributions	-0.0204	1.52e-11	-1.35e+09	0.000	-0.020	-0.020
candidate_contributions	-0.0204	1.52e-11	-1.35e+09	0.000	-0.020	-0.020
total_contributions	-0.0818	1.3e-11	-6.3e+09	0.000	-0.082	-0.082
transfer_from_committee	-0.1022	4.18e-12	-2.44e+10	0.000	-0.102	-0.102
candidate_loans	-0.0341	1.48e-12	-2.31e+10	0.000	-0.034	-0.034
other_loans	-0.0341	1.63e-12	-2.1e+10	0.000	-0.034	-0.034
total_loans	-0.0681	2.81e-12	-2.43e+10	0.000	-0.068	-0.068
operating_exp_offsets	0.3390	2.19e-11	1.55e+10	0.000	0.339	0.339
legal_acct_offsets	0.2198	4.03e-11	5.45e+09	0.000	0.220	0.220
total_offsets	0.5588	1.85e-11	3.02e+10	0.000	0.559	0.559
other_receipts	-0.1022	4.21e-12	-2.43e+10	0.000	-0.102	-0.102
total_receipts	0.1022	4.18e-12	2.45e+10	0.000	0.102	0.102
federal_funds_to_date	-0.0184	4.41e-12	-4.18e+09	0.000	-0.018	-0.018
individual_contributions_todate	-0.0037	1.76e-11	-2.1e+08	0.000	-0.004	-0.004
political_party_contributions_todate	-0.0037	6.74e-11	-5.46e+07	0.000	-0.004	-0.004
other_political_contributions_todate	-0.0037	1.76e-11	-2.1e+08	0.000	-0.004	-0.004
candidate_contributions_todate	-0.0037	1.75e-11	-2.1e+08	0.000	-0.004	-0.004
total_contributions_todate	-0.0147	1.5e-11	-9.82e+08	0.000	-0.015	-0.015
transfer_from_committee_todate	-0.0184	4.39e-12	-4.19e+09	0.000	-0.018	-0.018
candidate_loans_todate	-0.0061	1.5e-12	-4.09e+09	0.000	-0.006	-0.006
other_loans_todate	-0.0061	1.62e-12	-3.8e+09	0.000	-0.006	-0.006
total_loans_todate	-0.0123	2.95e-12	-4.17e+09	0.000	-0.012	-0.012
operating_exp_offsets_todate	-0.1191	2.19e-11	-5.45e+09	0.000	-0.119	-0.119
total_offsets_todate	0.1007	1.85e-11	5.43e+09	0.000	0.101	0.101
other_receipts_todate	-0.0184	4.38e-12	-4.21e+09	0.000	-0.018	-0.018
total_receipts_todate	0.0184	4.39e-12	4.2e+09	0.000	0.018	0.018

Omnibus:	156.293	Durbin-Watson:	0.332
Prob(Omnibus):	0.000	Jarque-Bera (JB):	859.815
Skew:	2.622	Prob(JB):	1.97e-187
Kurtosis:	10.375	Cond. No.	1.13e+16

R Squared Equation for total offsets:

total_offsets=

-0.1022*federal_funds-0.0204*individual_contributions-0.0204*political_party_contributions-0.0204*other_political_contributions-0.0204*candidate_contributions-0.0818*total_contributions-0.1022*transfer_from_committee-0.0341*candidate_loans-0.0341*other_loans-0.0681*total_loans+0.3390*operating_exp_offsets+0.2198*legal_acct_offsets+0.5588*total_offsets-0.1022*other_receipts+0.1022*total_receipts-0.0184*federal_funds_to_date-0.0037*individual_contributions_todate-0.0037*political_party_contributions_todate-0.0037*other_political_contributions_todate-0.0037*candidate_contributions_todate-0.0147*total_contributions_todate-0.0184*transfer_from_committee_todate-0.0061*candidate_loans_todate-0.0061*other_loans_todate-0.0123*total_loans_todate-0.1191*operating_exp_offsets_todate+0.1007*total_offsets_todate-0.0184*other_receipts_todate+0.0184*total_receipts_todate

Table 6. Regression analysis for net operating expenditures

OLS Regression Results

Dep. Variable:	net_operating_expenditures	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	7.352e+07
Date:	Mon, 12 Nov 2018	Prob (F-statistic):	0.00
Time:	20:26:34	Log-Likelihood:	-2968.0
No. Observations:	252	AIC:	5974.
Df Residuals:	233	BIC:	6041.
Df Model:	18		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	3673.8957	2935.748	1.251	0.212	-2110.107	9457.898
subtotal	8.1055	2.797	2.898	0.004	2.596	13.615
total_disbursements	-8.1076	2.797	-2.899	0.004	-13.618	-2.598
cashonhand_end	-8.1150	2.797	-2.901	0.004	-13.625	-2.605
net_contributions	-63.8803	0.246	-259.437	0.000	-64.365	-63.395
federal_funds_to_date	-9.4190	0.072	-130.079	0.000	-9.562	-9.276
total_contributions_todate	54.4477	0.209	260.430	0.000	54.036	54.860
transfer_from_committee_todate	-9.4315	0.039	-240.070	0.000	-9.509	-9.354
candidate_loans_todate	-3.1452	0.026	-119.075	0.000	-3.197	-3.093
other_loans_todate	-3.1418	0.045	-70.207	0.000	-3.230	-3.054
total_loans_todate	-6.2869	0.033	-192.050	0.000	-6.351	-6.222
total_offsets_todate	-10.4321	0.038	-272.261	0.000	-10.508	-10.357
other_receipts_todate	-9.4345	0.042	-225.867	0.000	-9.517	-9.352
total_receipts_todate	9.4437	0.037	255.006	0.000	9.371	9.517
operating_expenditures_todate	10.4998	0.298	35.184	0.000	9.912	11.088
transfer_to_committees_todate	9.5039	0.298	31.845	0.000	8.916	10.092
fundraising_disbursements_todate	8.9251	1.629	5.480	0.000	5.717	12.134
candidate_loan_repayments_todate	3.3955	0.536	6.334	0.000	2.339	4.452
other_loan_repayments_todate	3.0341	0.281	10.808	0.000	2.481	3.587
total_loan_repayments_todate	6.4296	0.290	22.191	0.000	5.859	7.000
total_contribution_refunds_todate	-54.3743	0.361	-150.574	0.000	-55.086	-53.663
other_disbursements_todate	9.5046	0.300	31.673	0.000	8.913	10.096
total_disbursements_todate	-9.5113	0.298	-31.902	0.000	-10.099	-8.924

Omnibus:	480.961	Durbin-Watson:	2.256
Prob(Omnibus):	0.000	Jarque-Bera (JB):	260498.327
Skew:	-11.016	Prob(JB):	0.00
Kurtosis:	158.962	Cond. No.	1.55e+19

R Squared Equation for net operating expenditures:

```
net_operating_expenditures=

8.1055*subtotal-8.1076*total_disbursements-8.1150*cashonhand_end-63.88
03*net_contributions-9.4190*federal_funds_to_date+54.4477*total_contri
butions_todate-9.4315*transfer_from_committee_todate-3.1452*candidate_
loans_todate-3.1418*other_loans_todate-6.2869*total_loans_todate-10.43
21*total_offsets_todate-9.4345*other_receipts_todate+9.4437*total_rece
ipts_todate+10.4998*operating_expenditures_todate+9.5039*transfer_to_c
ommittees_todate+8.9251*fundraising_disbursements_todate+3.3955*candid
ate_loan_repayments_todate+3.0341*other_loan_repayments_todate+6.4296*
total_loan_repayments_todate-54.3743*total_contribution_refunds_todate
+9.5046*other_disbursements_todate-9.5113*total_disbursements_todate
```

Conclusion:

In this project we evaluated descriptive, correlation and regression analysis on the report of FEC presidential campaign finance. In descriptive analysis, we performed a filtering process to distinguish the main attributes in which finalist candidates, Donald Trump and Hillary Clinton had the considerable shares. In the second step, we performed exploratory data analysis and box-plot on main attributes to depict more details about statistical characteristics such as Standard deviation, median, mean, minimum, maximum,1st, 2nd and 3rd quartile values. In addition, We did a full and partial correlation analysis on attributes. Furthermore, we did ANOVA on data set and determined with which attributes total offsets or net operating expenditures were statistically significant. Finally, we did multiple regression analysis to model total offsets or net operating expenditures with multiplying the coefficients to significant attributes by OLS method.

Reference:

Federal Election Commission. (2017, January 17). 2016 Presidential Campaign Finance. Retrieved from https://www.kaggle.com/fec/presidential-campaign-finance/home

Federal Election Commission. (2017, January 17). 2016 Presidential Campaign Finance. Retrieved from https://www.kaggle.com/fec/presidential-campaign-finance#summary.csv

Table of responsibility

	responsible for each chapter	Documentation
Introduction, Data Collection and Description	Sadegh Bamohabbat Chafjiri, Ramin Rafat, Peiwen Li	Sadegh Bamohabbat Chafjiri
Descriptive Statistics	Peiwen Li	
Correlation Analysis	Ramin Rafat	
Statistical tests Multiple Regression Analysis	Sadegh Bamohabbat Chafjiri	
Conclusion	Sadegh Bamohabbat Chafjiri, Ramin Rafat, Peiwen Li	