Ontario General Election Analysis

Motivation

We wanted to analyze the results of the 2018 Ontario general election. In particular we were interested in the role that financial donations play in predicting success in elections on a per-district level. By examining data sources that were available from Elections Ontario, it seemed that it would be possible to determine relationships between the success of candidates and parties running for election, and the financial donations that they recieved in the lead-up to the election.

Table of Contents

Gathering Data

The data that we were interested in analyzing were data related to the results of the 2018 Ontario general election. We were especially interested in finding data related to the financial contributions towards the political parties involved in the election and to the candidates running in different electoral districts. We were also interested in other data related to the electoral districts. The data for the financial contributions from the years 2014-2018 were used, because those were the years leading up to the 2018 election, and seemed the most applicable donations towards that particular election. The financial contribution data was downloaded from the Elections Ontario website in the form of a .csv file[1]. The results of the general elections in 2011, 2014 and 2018, were similarly retrieved from the Elections Ontario website[2]. Data for the population and area were scraped from the Elections Ontario website[3] since it was not available for download and the information on the page for each district needed to be combined. The results for the 2022 election were not available, and were instead retrieved from Wikipedia[4].

---

1. Elections Ontario. (n.d.). Retrieved August 14, 2022, from https://results.elections.on.ca/en/data-explorer?fromYear=1867&toYear=2022&electionId=504&levelOfDetail=candidate ↩

2. Elections Ontario. (n.d.). Retrieved August 14, 2022, from https://finances.elections.on.ca/en/contributions?fromYear=2014&toYear=2022 ↩
3. Elections Ontario. (n.d.). Retrieved August 14, 2022, from https://voterinformationservice.elections.on.ca/en/electoral-district/1 ↩
4. Wikimedia Foundation. (2022, July 27). 2022 Ontario general election. Wikipedia. Retrieved August 14, 2022, from https://en.wikipedia.org/wiki/2022_Ontario_general_election ↩

Preparing Data

After the data were obtained, we needed to prepare data frames for analysis, making sure that in each data frame the data is tidy, so that each row is one observation, and each column is one variable with one value.

Challenges

One significant difficulty that occured was due to the change of the Ontario electoral districts in 2018. The 107 districts that existed in the 2011 and 2014 elections were increased by 17 to 124, and many of the districts had their boundaries redrawn. This presented a significant hurdle since donations to many districts in the years leading up to the election could not be mapped directly onto the districts that existed during the election. It was decided to only use the districts that did not change. Fortunately, 76 of the original 107 districts remained the same and represent a significant sample.

Another issue is that many of the donation entries were not to a particular candidate in an electoral district, and were instead to the provincial party. Since these donations represented a majority of the money donated during the 2014-2018 period examined, they needed to be accounted for.

lastly, the existence of a large number of political parties that won no ridings, and received very little or no financial donations threatened to make the analysis unecessarily messy, so the decision was made to only include parties that won at least one riding, so the Liberal, Progressive Conservative, Green, and New Democratic parties were the only ones used.

New Attributes

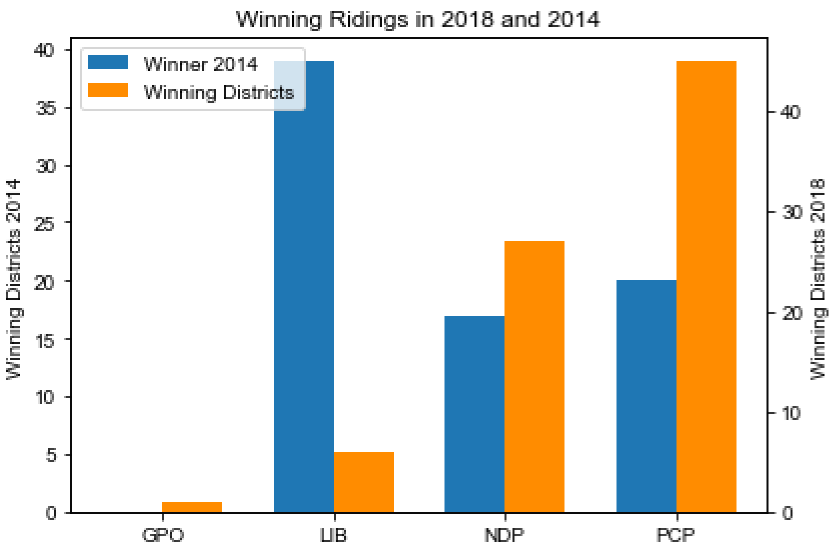
While there were many attributes included in the data downloaded, several others were created to help in the analysis. First, the total donations that were recieved by the party were calculated and added to each candidate. The population density in people per square kilometer was calculated and included, as well as the percentage of the population that turned out to vote. test

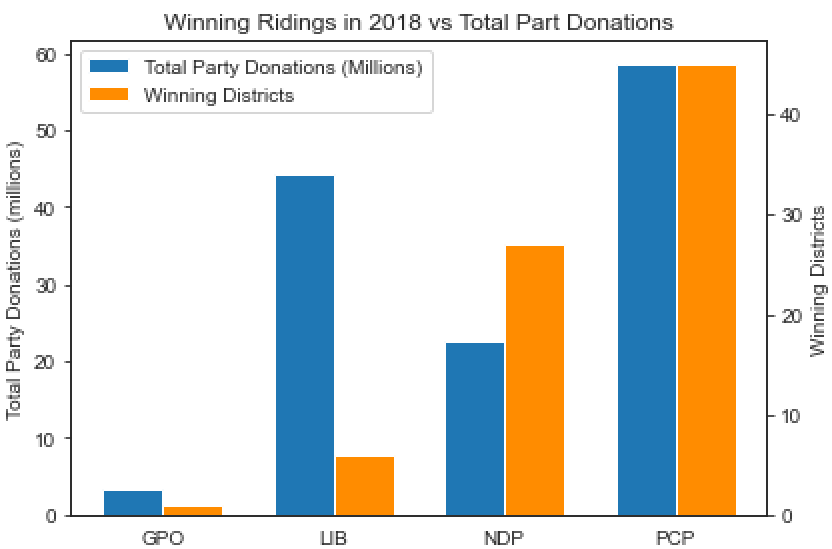
---

Initial Analysis

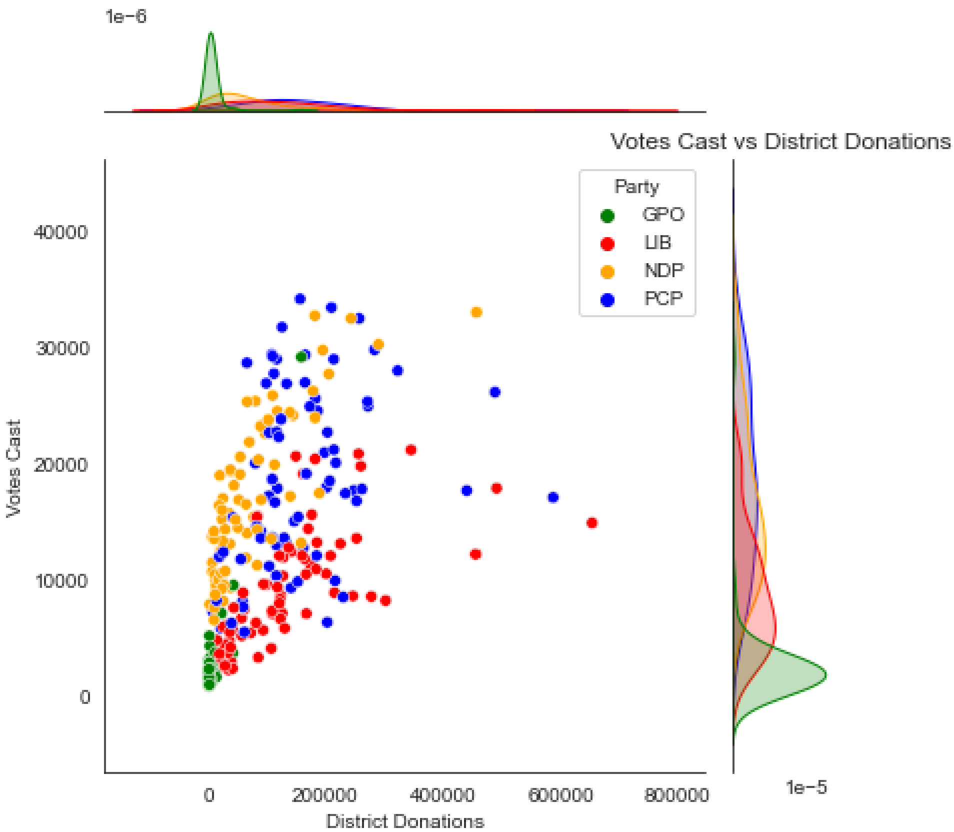
The first thing we can look at is the relationship between the amount of money that each party recieved from 2014-2018 and their relative performance in the 2018 elections, and while we can see a relationship between donations and election performance, we can see again that the Liberals did relatively poorly.

At first we can compare how each party did in the 2018 election relative to the 2014 election, and it is clear that the biggest change is that the Liberals did much worse in 2018, despite still recieving a significant number of donations.

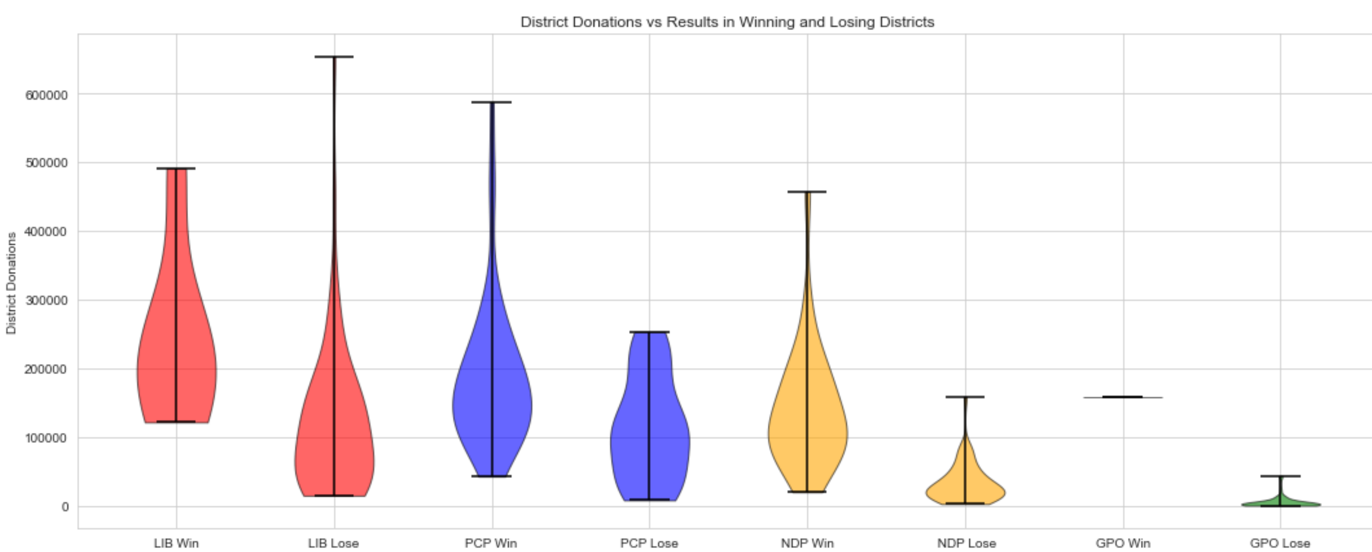




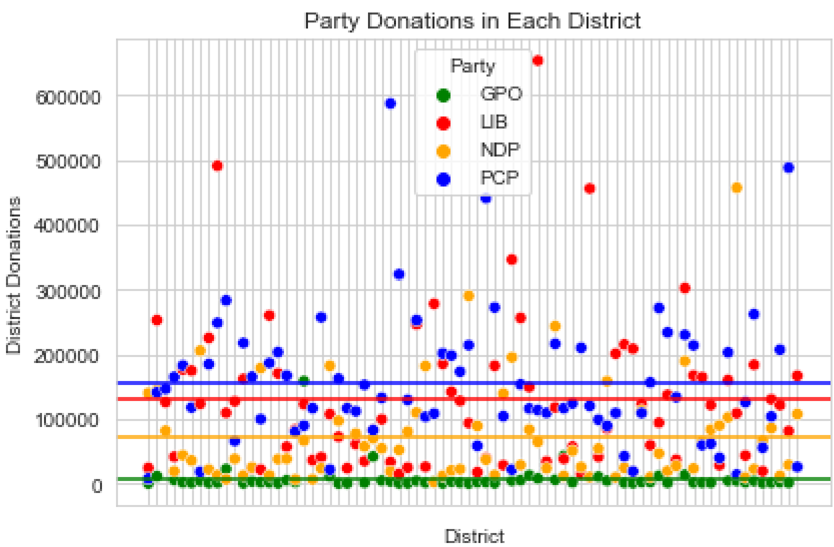
The next thing is to address the main question we are attempting to answer, what effect do donations have on election results? We can start with a graph of the relationship between the District Donations and the Votes Cast. Do donations to a district's candidate results in more votes for that candidate? If so we would expect to see a positive correlation between the two variables. Examining the plot reveals what appears to be a positive relationship between the two variables, which suggests that donations are correlated to the number of votes cast. It can also be seen that the relationship is not evenly spread out among the parties. The two smaller parties, the Green party and the NDP, seem to have the strongest relationship, with the Liberals having the weakest. This will be examined further in the Linear Regressions section.



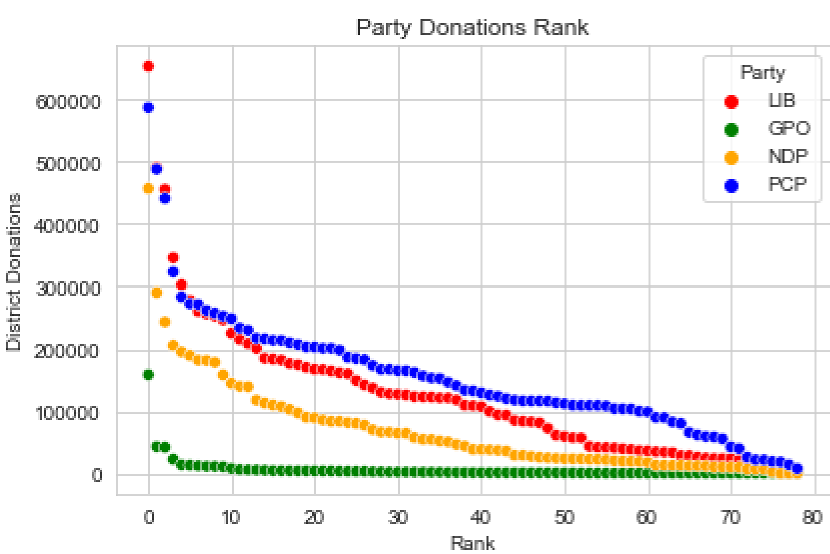
The next thing we can look at is a comparison of the distributions of donations in the different districts, seperated by whether the district won or not. We can see above that candidates that recieved more donations tended to recieve more votes, but did winning candidates for each party tend to recieve more donations? Below we can violin plots that show the distribution of donations in winning and losing districts for each party. We can see that for every party, the winning districts did tend to recieve more donations. We can also see that for the Green party, the Conservative party, and the NDP, the outlier districts with very higher donations were largely in the winning districts, whereas the Liberals had losing districts with very high donation amounts.



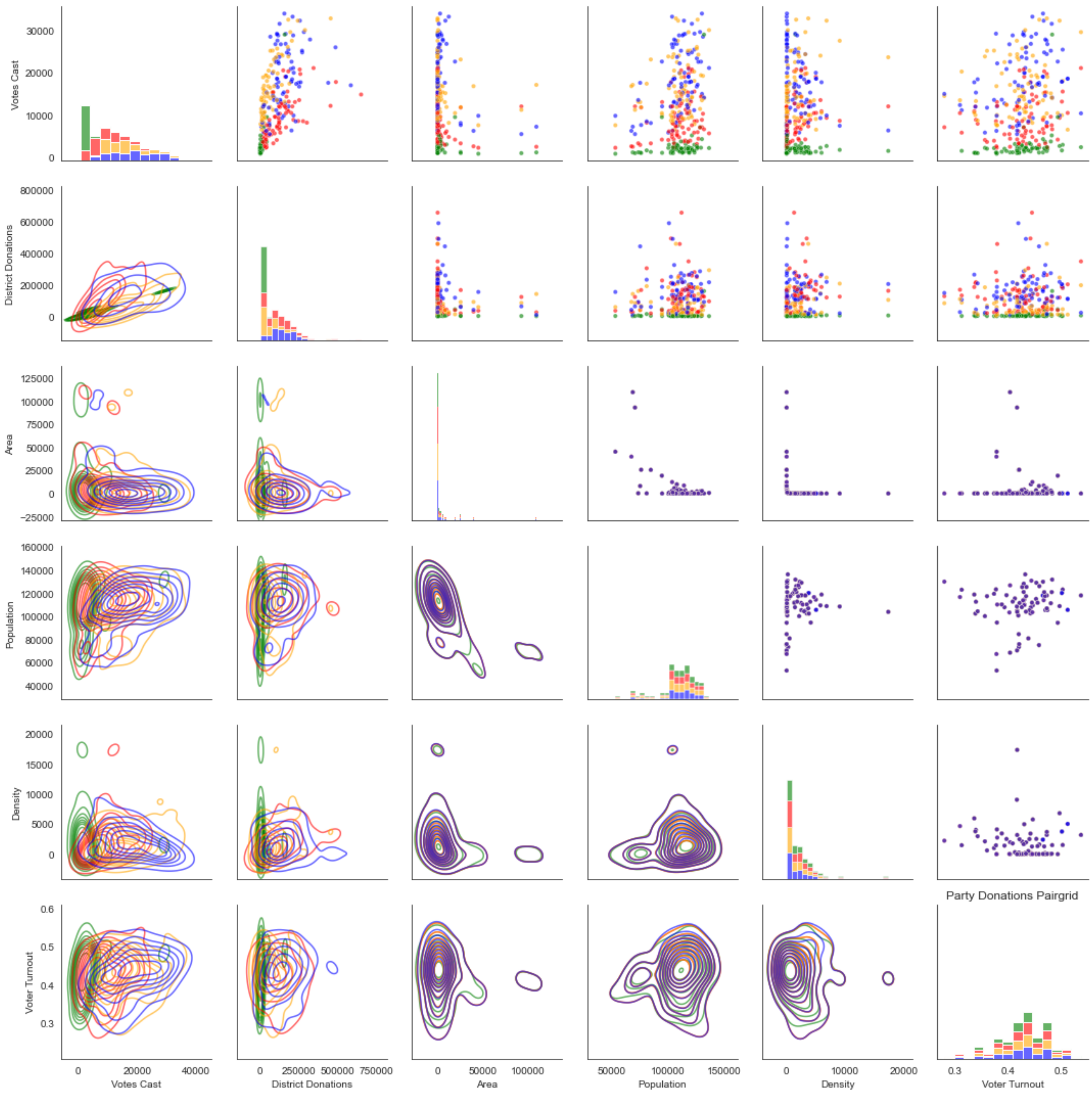
We can now produce a plot of the donations to each party, in each district, to see if there is any clustering. It could be possible that districts that recieve more donations from one party tend to recieve more from the others. However, there does not appear to be this sort of relationship, and the correlation between the donations to different parties in the same districts are very low.



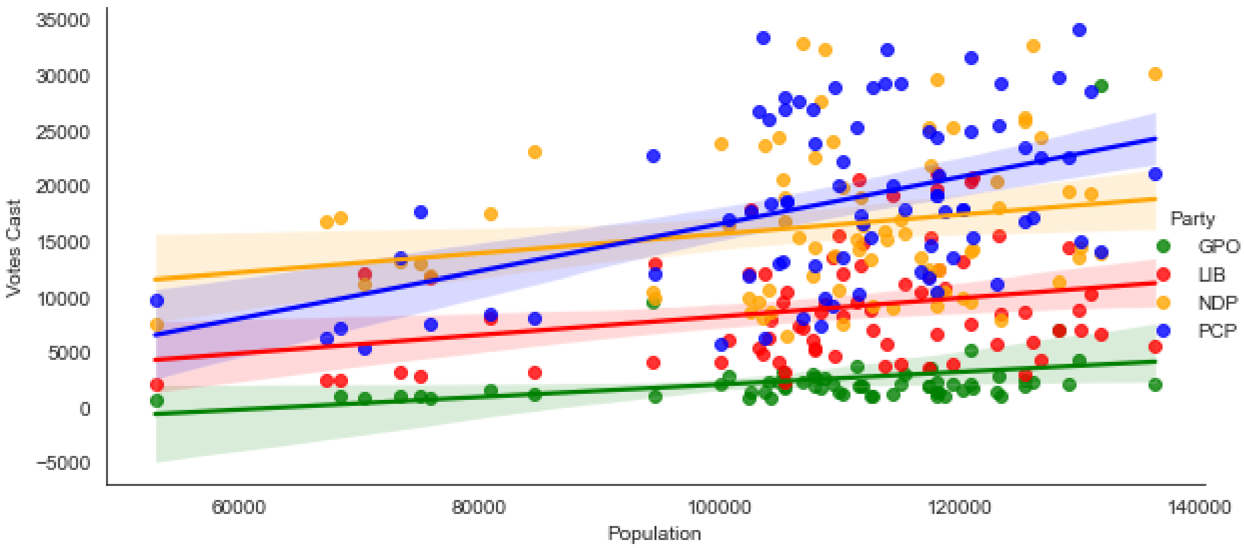
We can also look at the donations per district, sorted in descending order, and similarly, a histogram of the number of districts within certain bins of donations. This gives us an idea about how the donations are distributed to the districts for each party. It is clear that there is a clear difference between the parties in terms of how wealthy they are. The Conservative party recieves the most money, followed by the Liberals, the NDP and the Green party. There is also an interesting feature where the conservative donations take longer to drop, there are noticeably more districts with medium levels of donations to the Conservatives than to the other parties.



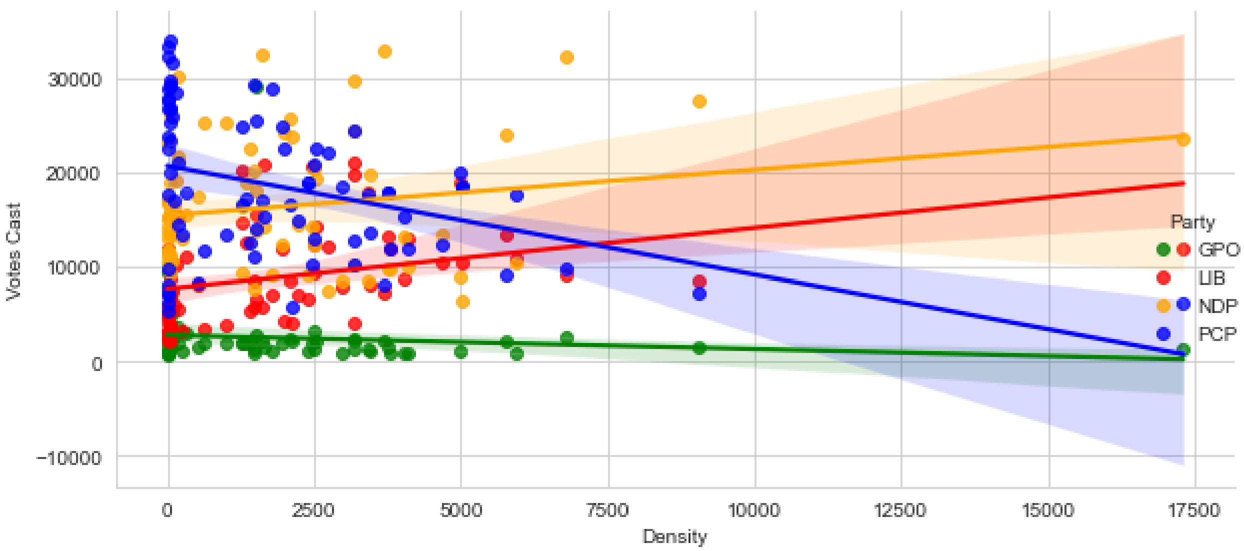
Before continuing, we can visualize all of the variables to see if there are any interesting relationships to examine more closely.



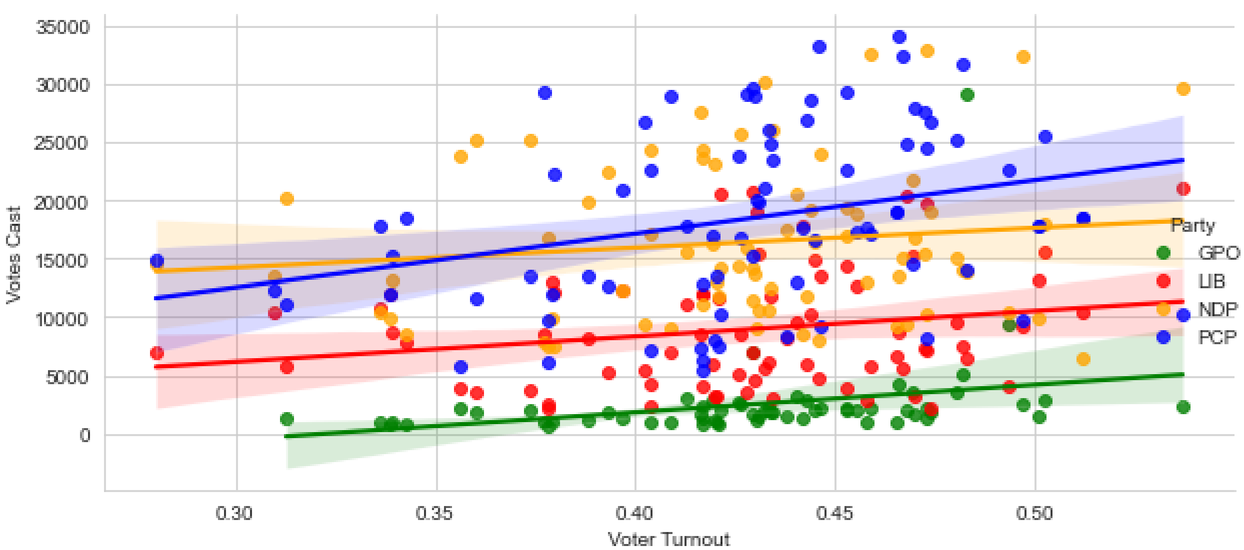
There are several relationships we can see above that we can examine in more detail. There appear to be some relationships relating Population, Density, and Voter Turnout, differing by party. If we look below, if the Votes Cast are compared to the Population, the population Density, and the Voter Turnout, we can see that the trends are similar for the NDP, the Liberals and the Green part, however the Conservatives seems to have recieved more votes for districts with higher populations, and higher voter turnouts, and much lower for denser districts. There is also an interesting feature shared by many of the graphs where the relationships seem to be bimodal.



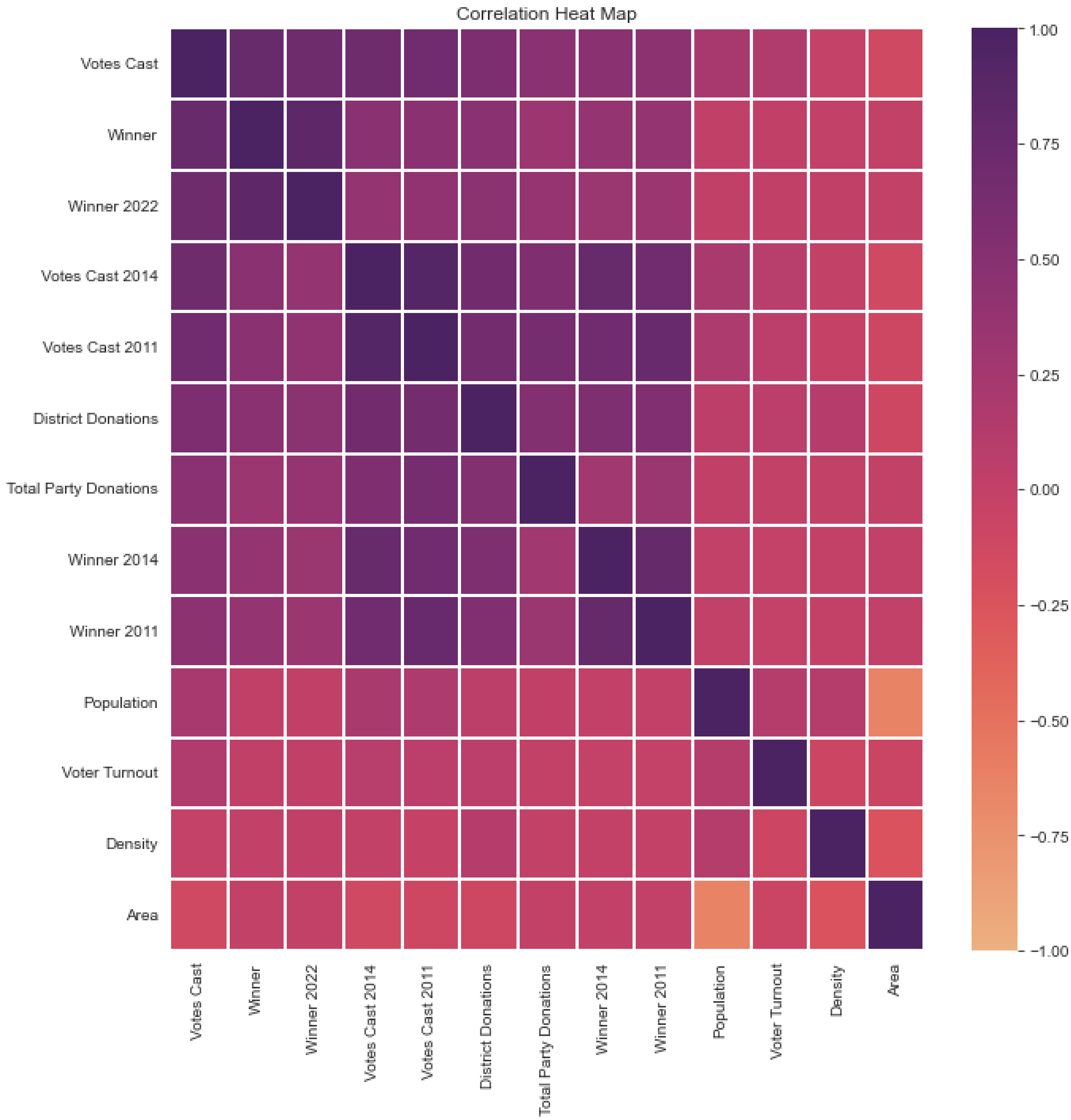
Votes Cast
Population
Party
GPO
LIB
NDP
PCP



Votes Cast
Density
Party
GPO
LIB
NDP
PCP



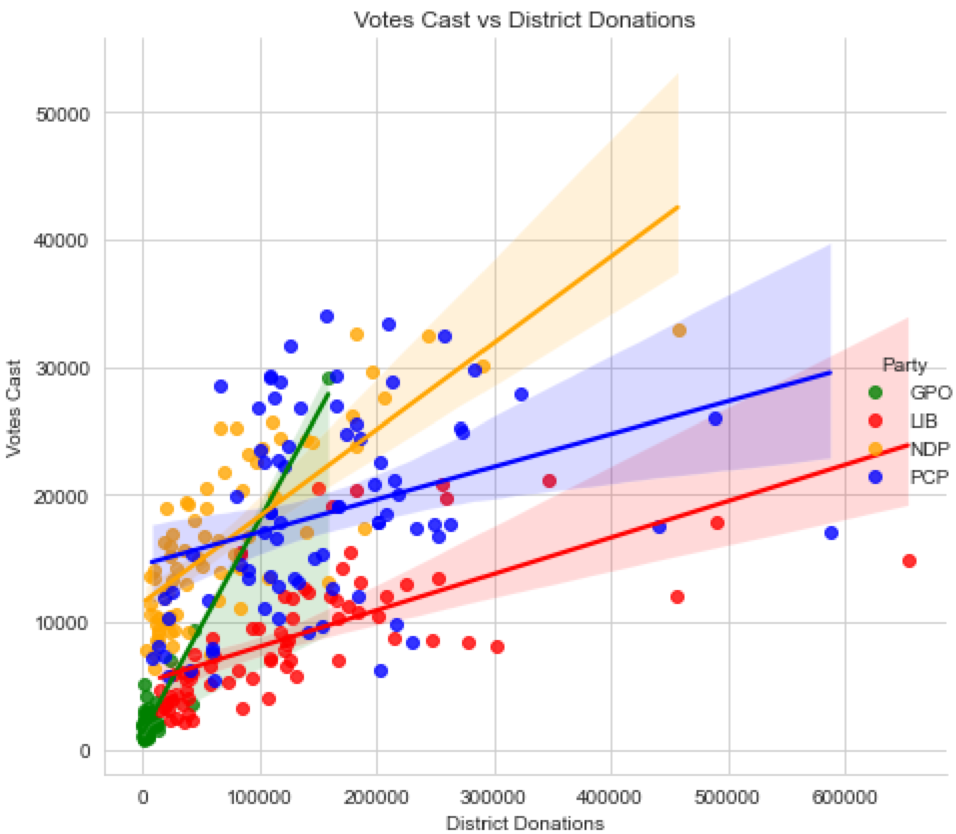
Votes Cast
Voter Turnout
Party
GPO
LIB
NDP
PCP

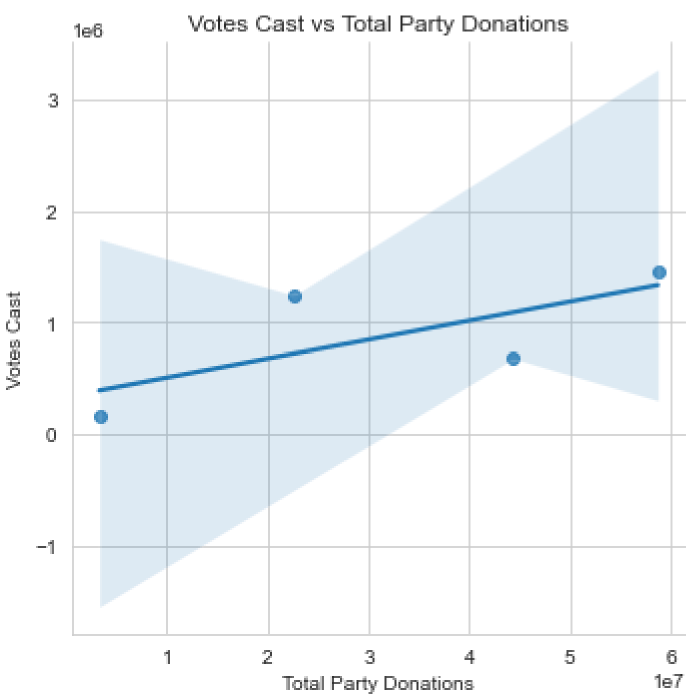


Next we can examine the correlations between variables, especially those most closely correlated with Votes Cast. The results of previous elections and the donations all have a positive corelation with the number of votes recieved by each candidate, whereas the variables related to the size, population, density and voter turnout have much weaker relationships.

Linear Regressions

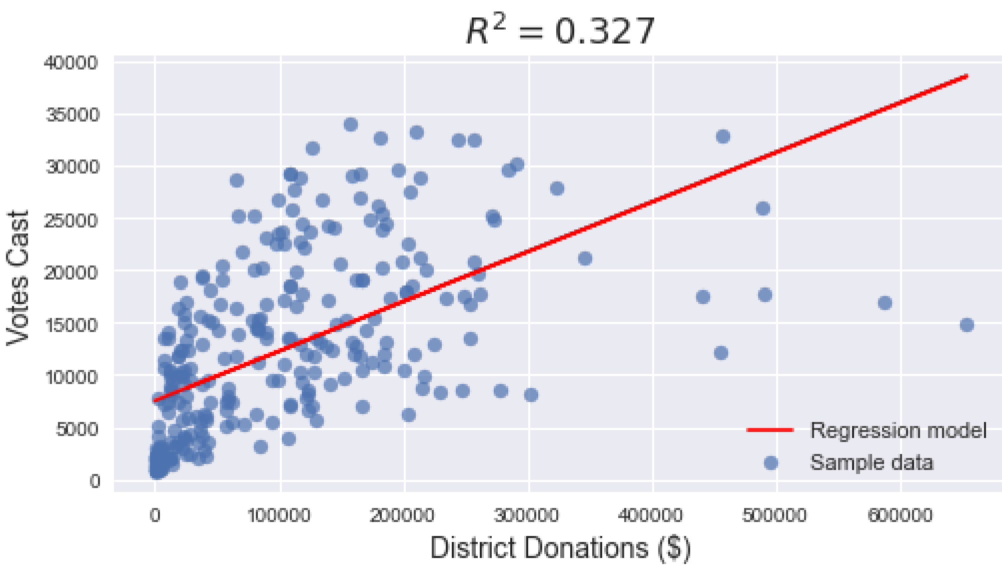
The next thing to do is see if we can make a predictive model for the number of Votes Cast based on the information available. Since the main motivation was to analyze the relationship that financial donations have with electoral success, it makes sense to being with a plot comparing the Votes Cast to the District Donations, and to the Total Party Donations. The plot with District Donations suggest there is a positive correlation between the two variables for all parties, and a plot of the Votes Cast vs the Total Party Donations also suggests a positive relationship, although the sample size is very small.

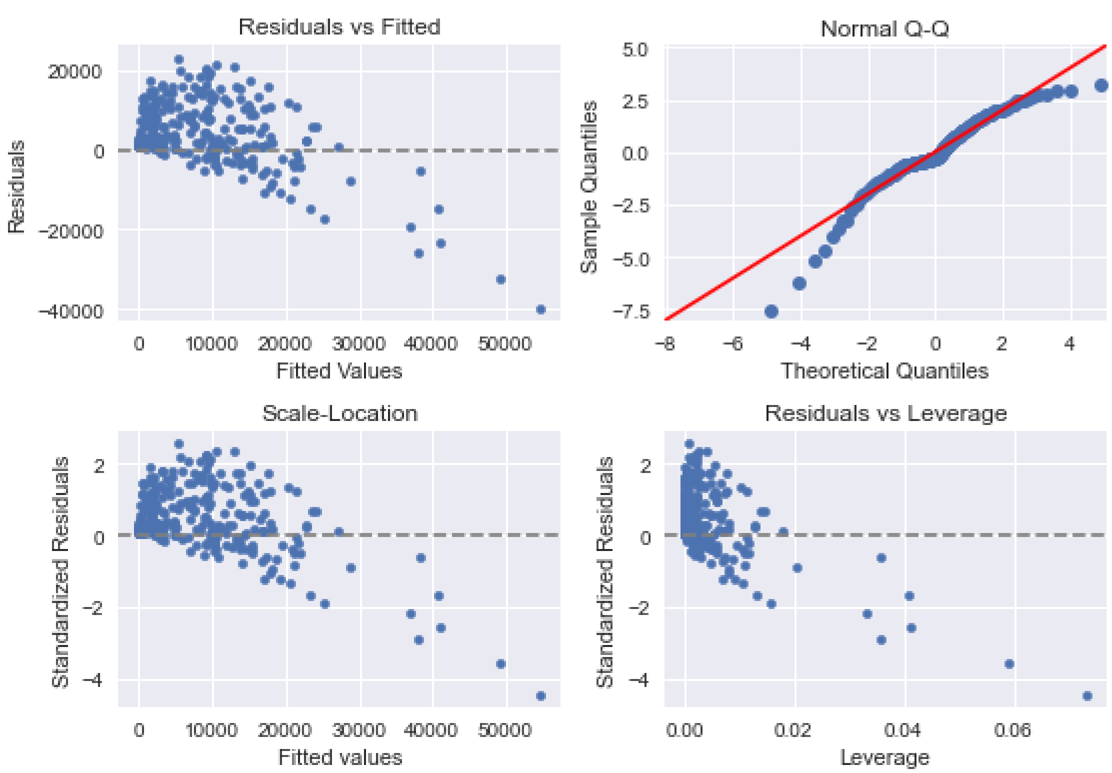




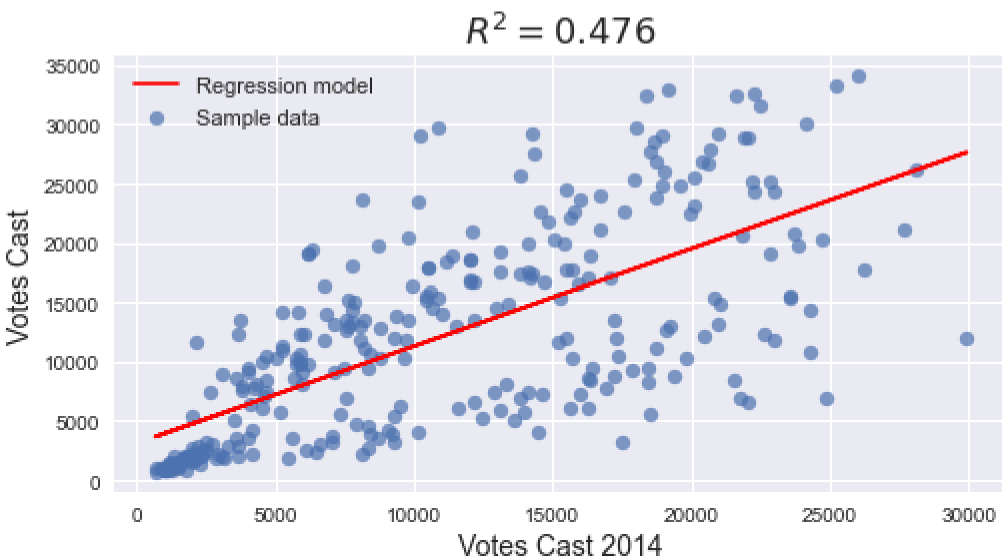
We can also saw in the Initial Analysis section that these appears to be a positive relationship between the votes in 2018 and in the previous years' elections, so it makes sense to examine those variables as well. First we can examine the correlations to see which variables may be suitable targets for a linear regression. It appears that the votes cast in previous elections, the donations, and the winning districts from the previous elections all have a moderate to strong correlation with the votes cast in 2018.

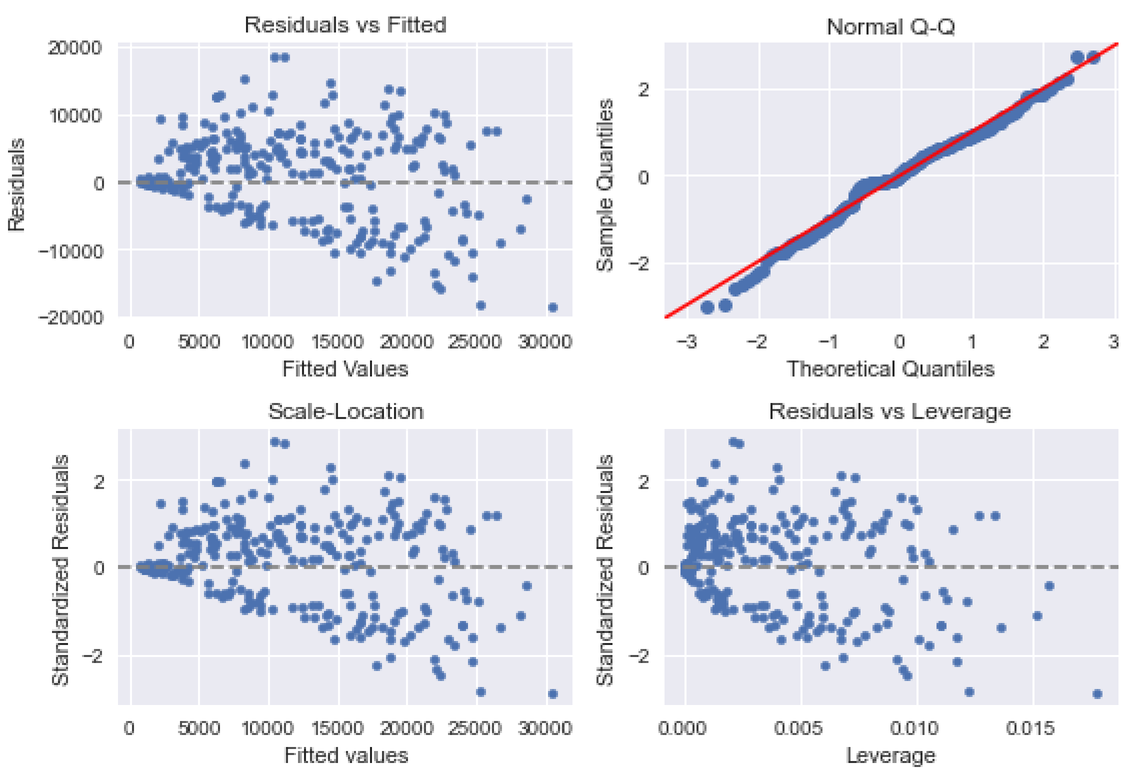
Now we can examine several linear regressions. If we use the donations, we obtain a significant model with a moderate R-squared, and small P for all coefficients, suggesting that donations do usefully account over a third of the observed variation in votes counts. The models using only District Donations or Total Party Donations alone do not provide as strong a relationship so they appear to be somewhat independent. If we closely examine a single variable regression using District Donations, we can see several issues with the regression. First, looking at the residuals vs the fitted values and the leverage, we can see that there is still a linear pattern. It is likely because their is another, unaccounted-for variable that is having an effect on the Votes Cast. We can also see from the QQ plot that the data is left-skewed.





Regressions that use votes from previous elections work even better, which isn't too surprising. In this case however there are significant multicollinearity issues when using the results from both previous elections, so using only the votes from 2014 provides the best model. In this case, the QQ plot hsows the data is fairly normal, however the residuals show clearly that the data is heteroscedastic, with higher variance at high donation levels.





Machine Learning Predictions

K-Nearest Neighbour

Below we use a nearest neighbour and random forest machine learning model to create a predictive model. We are trying to take each candidate, and based on the data available, create a model that will predict whether that candidate will win their riding or not. The first model used is a k-nearest neighbours model, and several different k values are used to determine if their is an ideal k to use to provide the most accurate model. In the end result, the best model produces an accuracy of approximately 85%.

	k	Accuracy
0	1	0.842027
1	3	0.859595
2	5	0.850676
3	7	0.851081
4	10	0.847838
5	20	0.846081
6	30	0.830676

Random Forest

The next model used is a random forest model. Several different number of estimators are used, the more estimators typically will return better results. This model performed better than the knn model with an accuracy closer to 89% with a larger number of estimators.

	n	Accuracy
0	10	0.879324
1	100	0.888784
2	250	0.890135

Using the better random forest model, we can do k-fold cross validation in order to test if the model might be effective on new data. The validation returned a similar accuracy around 89%

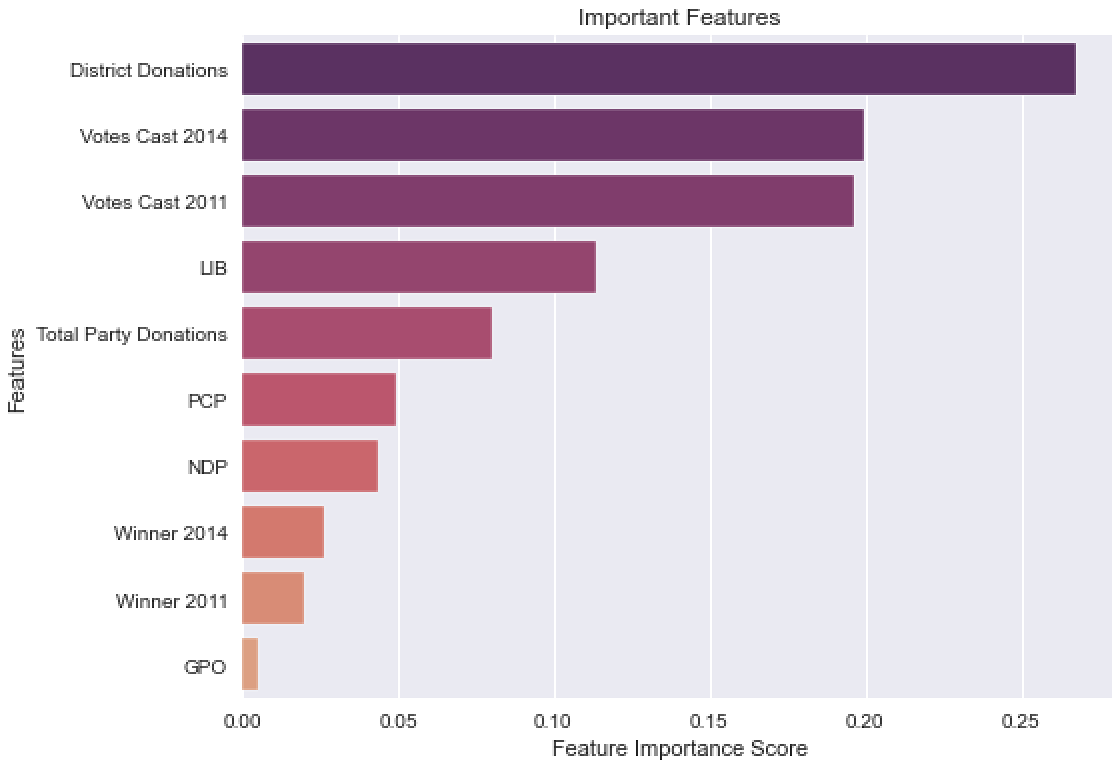
	Fold	Accuracy
0	1	0.954545
1	2	0.954545
2	3	0.863636
3	4	0.818182
4	5	0.818182
5	6	0.863636
6	7	0.863636
7	8	0.954545
8	9	0.909091
9	10	0.954545

	Cross-Validation Accuracy	Standard Deviation
0	0.895455	0.053974

Finally, the model can be tested on a completely new set of data: the results of the 2022 election. using the model on these data provided an even higher 91% accuracy, suggesting the model may be effective even with completely new data.

	n	Accuracy
0	10	0.905676
1	100	0.911351
2	250	0.913378

Lastly, we can examine the importances of the different variables according to the model. Most interestingly, the District Donations is the variable with the highest importance.



Conclusions

There does appear to be a relationship between donations and election outcomes. We have seen that parties that recieve a larger total amount of financial donations tend to win more seats in an election, and also that candidates that recieve more donations are more likely to win their riding. A linear regression to predict the number of votes that a candidate will recieve, based on the donations they recieve, accounts for slightly more than a third of the variation in the number of votes recieved. Similarly, in an effective machine learning model that predicts winning candidates with approximately 90% accuracy, the most important feature is the donations that the candidate recieved.

Appendix

Examples of the data frames used for analysis are shown below.

The first prepared data set has each row as one of the major political parties, and a variety of attributes related to that party, their financial donations and their electoral results.

Party	Votes Cast	Winner	Total Party Donations	Votes Cast 2014	Winner 2014	Votes Cast 2011	Winner 2011	Vote %	Winner 2022
GPO	158184.0	1.0	3296726.92	148313.0	0.0	76467.0	0.0	3.173885	1.0
LIB	677660.0	6.0	44351026.06	1225996.0	39.0	1099295.0	38.0	14.564344	7.0
NDP	1244682.0	27.0	22617023.03	812789.0	17.0	700620.0	13.0	27.404930	22.0
PCP	1458137.0	45.0	58720546.33	1010616.0	20.0	1012849.0	25.0	30.856840	48.0

In the second prepared data set, each row is an electoral district, and the columns are the donations for each major political party.

	District	GPO Donations	LIB Donations	NDP Donations	PCP Donations
0	Algoma Manitoulin	200.00	24228.00	140030.03	7863.00
1	Beaches East York	11456.09	252929.30	144667.45	141033.34
2	Bruce Grey Owen Sound	5618.00	41521.00	19125.00	165078.66
3	Burlington	1640.00	176591.40	43830.15	182884.54
4	Cambridge	1000.00	174796.01	35807.17	117852.38

The third data set is the one that is involved in most of the analysis. Each row represents one candidate for the election.

	District	Party	Votes Cast	Winner	District Donations	Votes Cast 2014	Winner 2014	Votes Cast 2011	Winner 2011	Vote %	Winner 2022	Total Pa Donati
0	Algoma Manitoulin	GPO	1025.0	0.0	200.00	828.0	0.0	684.0	0.0	0.037087	0.0	3296726
1	Algoma Manitoulin	LIB	2365.0	0.0	24228.00	6504.0	0.0	7397.0	0.0	0.085571	0.0	44351026
2	Algoma Manitoulin	NDP	17105.0	1.0	140030.03	14171.0	1.0	11585.0	1.0	0.618894	1.0	22617023
3	Algoma Manitoulin	PCP	7143.0	0.0	7863.00	4589.0	0.0	6141.0	0.0	0.258449	0.0	58720546
4	Beaches East York	GPO	2128.0	0.0	11456.09	2329.0	0.0	1025.0	0.0	0.043541	0.0	3296726