**Association rule mining:**

We ran Apriori association rule mining on the political subset of our data, consisting of political contributions to candidates by industry, as well as election results. For this analysis, a number of new variables were created; the winner variable, to denote which candidates actually won their elections; indrank, a ranking of the industries that contributed to each candidate in each year by the amount contributed; and the total contributions and votes each candidate received in each election cycle was binned into categorical variables (candtotallevel and voteslevel, respectively) with four levels (very low, mid-low, mid-high, and high). Outliers, defined as candidate/year/industry level observations where either the number of votes received or the total amount contributed was more than one interquartile range less than/greater than the 25th/75th percentile values, were removed from the dataset for this analysis. Apriori rule mining was run to find rules with a minimum confidence level of .2, under three different support levels (.4, .2, .05). Eclat rule mining also was utilized but, as it yielded nearly identical results to the Apriori algorithm, the analysis focuses on the rules generated by the Apriori algorithm.

A variety of characteristics occurred frequently with the incumbent variable, a binary variable which represents whether or not a candidate was an incumbent. Incumbent and election winner occurred together in 50.5% of the observations in the dataset; incumbent candidates won 96.1% of reelection opportunities, and 86% of all election winners were incumbents (rules 1, 2). This is in line with the historical average proportion of incumbents who win reelection[[1]](#footnote-1), providing evidence that our dataset is representative of the real world. Another way in which the rule mining analysis mirrored popular opinions about American politics is in the relationship it revealed between a candidate’s fundraising and election success. 81% of candidates who raised a very low amount of funding lost their election bids, and 76% of election losers raised a very low amount of funding, while 80.9% of those who raised a high amount of funding won their election bids (rules 3, 4, 5).

Challenger (non-incumbent) and election loser also frequently occurred together, in 39.2% of the observations. 82.6% of challengers lost, and 95.1% of losing candidates were challengers (rules 6, 7). The incumbent variable also had a strong relationship with candidate funding levels; in general, incumbents received more funding than challengers. Challenger was frequently associated with very low levels of campaign contributions, occurring together in 32.3% of observations; 68% of challenger candidates received a very low amount of funding, and 83.4% of all candidates who received a very low amount of funding were challengers (rules 8,9). On the other end of the funding distribution, 75.1% of candidates who received a high amount of funding were incumbents, and 72.1% of candidates who received a mid-high amount of funding were incumbents (rules 10, 11).

Apriori association rule mining revealed some interesting frequent itemsets featuring the political party variable. Over the time period examined (2004-2014), belonging to the Republican party was frequently associated with being elected. 62.7% of Republican candidates won their elections, and 56% of election winners were Republicans (rules 14,15). Interestingly, 57.7% of Democratic candidates also won election over the time period (rule 16). There weren’t enough independent candidates in the dataset to show up in the association rules even at support level .05, but one takeaway from the Democratic and Republican results is that Independent candidates generally do not do very well in American congressional politics; out of the 1259 Independent candidates in the dataset, only 16 were elected.

By ranking every industry that contributed to each candidate by contribution amount, we hoped to determine if the breakdown of a candidate’s contributions affects the candidate’s performance in elections. However, even with a minimum support level of .05, there were only two combinations of the industry rank variable, an industry, and another variable that occurred frequently enough to get picked up by the association rule generation. The “industry” not for profit occurred with an indrank rank of 1 (meaning nonprofits were the candidate’s primary source of funding) and a very low level of funding in 5.4% of the observations; 48.4% of candidates whose primary industry was not for profit received a very low level of funding (rule 18). Considering that overall, 43.3% of candidates received a very low level of funding, this is not a very significant result; candidates whose main contributor was the not for profit industry were slightly more likely to receive a very low level of funding than the average candidate. Not for profit also occurred with an indrank of 1 and the election loser indicator in 6.1% of cases; 54.6% of candidates whose primary source of funding was nonprofits lost their elections (rule 19). This is a fairly significant result; only 40.8% of candidates in the dataset lost their elections, so candidates whose largest contributor was nonprofits were much more likely than average to lose.

Apriori rule mining revealed many associations which were in line with our expectations; while the rules we found strengthened our conviction that our dataset is representative of the real world, they failed to bring much new information to light. The main takeaway from the frequent itemset mining is that in American congressional politics, life is hard for challengers; incumbents have a large fundraising advantage and win an incredibly high proportion of elections in which they participate. Deeper analysis is necessary to determine if *where* a candidate’s funding comes actually impacts election results.

Table 1 - Interesting Frequent Itemsets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Rule** | **Support** | **Confidence** | **Note** |
| 1 | {INCUMBENT=1} => {WINNER=1} | 0.505272256 | 0.961323423 | 96.1% of incumbents won reelection |
| 2 | {WINNER=1} => {INCUMBENT=1} | 0.505272256 | 0.859453102 | 85.9% of winners were incumbents |
| 3 | {CANDTOTALLEVEL=Very Low} => {WINNER=0} | 0.313362143 | 0.810226155 | 81% of candidates who raised very low amounts lost |
| 4 | {WINNER=0} => {CANDTOTALLEVEL=Very Low} | 0.313362143 | 0.760402685 | 76% of losing candidates raised very low amounts of money |
| 5 | {CANDTOTALLEVEL=High} => {WINNER=1} | 0.074641314 | 0.80922039 | 80.9% of candidates who raised a high amount won |
| 6 | {INCUMBENT=0} => {WINNER=0} | 0.391771824 | 0.825827139 | 82.6% of challengers lost |
| 7 | {WINNER=0} => {INCUMBENT=0} | 0.391771824 | 0.950671141 | 95.1% of losers were not incumbents |
| 8 | {CANDTOTALLEVEL=Very Low} => {INCUMBENT=0} | 0.322731201 | 0.834450702 | 83.4% of candidates who raised a very low amount were challengers |
| 9 | {INCUMBENT=0} => {CANDTOTALLEVEL=Very Low} | 0.322731201 | 0.680294418 | 68% of challengers raised a very low amount of funding |
| 10 | {CANDTOTALLEVEL=High} => {INCUMBENT=1} | 0.069282627 | 0.751124438 | 75.1% of candidates who raised a high amount of funding were incumbents |
| 11 | {CANDTOTALLEVEL=Mid-High} => {INCUMBENT=1} | 0.138738116 | 0.720855039 | 72.1% of candidates who raised a mid-high amount were incumbents |
| 12 | {VOTESLEVEL=Very Low} => {INCUMBENT=0} | 0.125808124 | 0.87602311 | 87.6% of the candidates who received very low vote counts were challengers |
| 13 | {VOTESLEVEL=High} => {INCUMBENT=1} | 0.137942956 | 0.89301701 | 89.3% of candidates who received a high vote count were incumbents |
| 14 | {PARTY=R} => {WINNER=1} | 0.329230769 | 0.627090741 | 62.7% of Republican candidates won |
| 15 | {WINNER=1} => {PARTY=R} | 0.329230769 | 0.560011761 | 56% of election winners were Republicans |
| 16 | {PARTY=D} => {WINNER=1} | 0.258461538 | 0.577252722 | 57.7% of Democratic candidates won |
| 17 | {WINNER=1} => {PARTY=D} | 0.258461538 | 0.439635401 | 44% of election winners were Democrats |
| 18 | {PRIMARY.INDUSTRY=Not for profit,indrank=1} => {CANDTOTALLEVEL=Very Low} | 0.053759723 | 0.484272812 | 48.4% of candidates whose primary industry was not for profit raised a very low amount of funding |
| 19 | {PRIMARY.INDUSTRY=Not for profit,indrank=1} => {WINNER=0} | 0.060674157 | 0.546558704 | 54.7% of candidates whose primary industry was not for profit lost |

1. https://www.opensecrets.org/bigpicture/reelect.php [↑](#footnote-ref-1)