

# WHICH INDUSTRIES DONATE TO DEMOCRATS

*By Rohit Dilip and Samuel Arnesen*

With over 20 candidates vying for the Democratic nomination, donors have plenty of options in deciding whom to support. Like single payer? Give Bernie Sanders \$2.70. Believe in a universal basic income? Donate to Andrew Yang.

(And of course, if you believe that love will defeat Donald Trump, [here's](#) Marianne Williamson's donation page. We'd also like to inform you that you're our lucky one millionth visitor, and you've just won a free iPad).

Although a clear [top tier](#) of candidates has emerged, if [history is any guide](#), we'll likely see at least a few major shakeups before the nominee is chosen. Already, six different candidates have spent some time in a polling average top three; this churn will likely increase as candidates commit gaffes, have surprising debate performances, or drop out of the race entirely. Political campaigns thus have a vested interest in which candidates benefit from changes in their competitors' fortunes.

Answering this question isn't as simple as just comparing the candidates' policy positions, since many of the Democrats hold similar policy views and it's difficult to come up with a rigorous way to differentiate between them. It's also unclear the extent to which voters make their choices based on policy preferences, as opposed to [other factors](#) like electability, gender, and age.

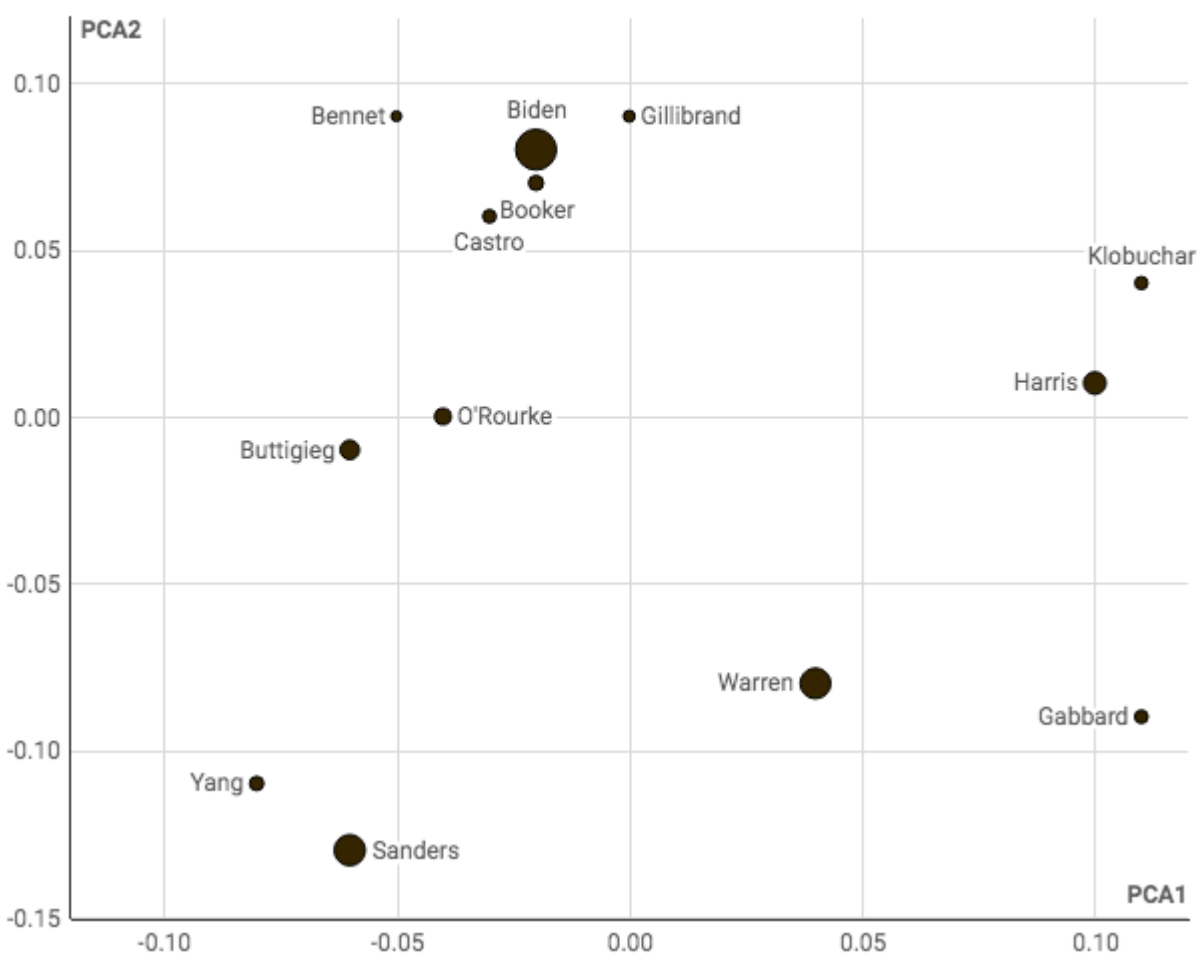
To answer this question, we looked at [quarterly donation data](#) released by the FEC, which provides the [name and employer](#) of all donors who have contributed more than \$200. To understand the flow of donations, we categorized every employer into one of seventeen industries [1], then counted the number of donors the candidates received from each of those industries [2]. While our industries are fairly broad, categorizing donors in this way can still provide us with information that categorization by education, income level, or gender cannot. For instance, a wealthy, white, educated finance executive probably might have different political preferences than a wealthy, white, educated college professor.

To visualize the results, we projected our data onto two dimensions using a technique called [principal component analysis](#) [3]. We adjusted for the fact that some candidates, like Sanders, have substantially more donors than others by describing each contender in terms of their reliance on different industries for donors. Take Sen. Michael Bennet as an example. Even though six times as many lawyers donated to Mayor Pete Buttigieg than to Bennet, the Colorado senator is more than twice as reliant on lawyers because lawyers make up a quarter of all his donors compared to roughly 12% for Buttigieg as Buttigieg has more donors overall.

Our results confirm several widely-held beliefs about the Democratic primary: Sanders relies most on blue-collar and service-sector workers, former professor and current Massachusetts Sen. Elizabeth Warren is the favorite among teachers and professors, and former Vice President Joe Biden is the frontrunner with businessmen, lawyers, and financiers. But the results also reveal some surprising parallels between many of the candidates and suggest that some of the main beneficiaries of missteps by the frontrunners may not be the candidates who are the most ideologically similar.

## Candidate Similarity Based on Industry of Donations

● candidate



### The Unpopular Populists

As a well-known Senator with one of the most liberal records in the [Senate](#), Sen. Kirsten Gillibrand appeared primed for a serious run for the Democratic nomination. Throughout the first six months of the campaign, she carved out a clear brand as an outspoken [advocate for women](#)

during a time when women's issues are becoming [increasingly important](#) to the Democratic party.

Like Gillibrand, Sen. Cory Booker also had a clear case for how he might win the nomination. Citing his background, political progressivism, and oratorical skills, Michelle Goldberg of the New York Times [proclaimed](#) that “if you were cooking up a candidate in a lab to take on Donald Trump, you might come up with someone a lot like Senator Cory Booker of New Jersey.”

And yet both Gillibrand and Booker have struggled. Gillibrand is stuck at 0.6% in the RealClearPolitics polling [average](#) and Booker is performing only slightly better at 2.2%. Our donor data suggests a possible reason why: they are competing for the same voters as Biden.

Of all the pairs of candidates, Biden and Booker have the most similar voter bases while Biden and Gillibrand are third. All three candidates heavily rely on businesspeople (Booker is 1<sup>st</sup>, Gillibrand is 2<sup>nd</sup>, and Biden is 4<sup>th</sup>), lawyers (Biden is 2<sup>nd</sup>, Gillibrand is 3<sup>rd</sup>, and Booker is 6<sup>th</sup>), and financiers (Booker is 2<sup>nd</sup>, Gillibrand is 3<sup>rd</sup>, and Biden is 5<sup>th</sup>). Individuals from those industries make up almost 60% of the three candidates' donors, compared to only 21% for Sanders, 28% for Warren, and 42% for Sen. Kamala Harris.

On one level, it is unsurprising that senators from New York and New Jersey rely on businesspeople and executives for their donations, nor is it surprising that a former Vice President might possess [connections](#) to wealthy donors. But it is still noteworthy that in industry after industry, Biden, Booker, and Gillibrand receive a constant share of the donor pool. After all, both Booker and Gillibrand differ strongly from Biden in their backgrounds and ideologies. Both [Booker](#) and [Gillibrand](#) attacked Biden in the second debate for past stances he has taken; the issues they have discussed the most -- [criminal justice issues](#) for Booker and [women's](#) issues for Gillibrand -- are widely considered to be [weak points](#) for Biden. Moreover, on issues from [single-payer health care](#) to [free trade](#), Booker and Gillibrand have consistently staked out more populist positions than Biden.

But if donor data is any guide, voters apparently still view Booker and Gillibrand in the same light as Biden despite the major differences in policies and background. This information takes a step towards resolving the question as to why two of 2020's most promising candidates have [failed](#) to gather momentum.

### **The Young White Guys**

On March 14<sup>th</sup>, former Texas representative Beto O'Rourke [declared](#) his candidacy for President of the United States. By the start of April, he had surged into 3<sup>rd</sup> place in the [polls](#) with his average support just shy of double digits. South Bend, Indiana Mayor Pete Buttigieg, meanwhile, was languishing at just [2%](#).

One month later, however, their fortunes reversed. By mid-May, Buttigieg was tied for third with Warren while O'Rourke had lost nearly 40% of his supporters. Though Buttigieg's stardom has faded somewhat since its peak in early spring, the two young, white, male candidates' relative positions in the polls have remained constant. As the New York Times [observed](#) back in April, O'Rourke and Buttigieg appeared to be competing for the same block of voters.

The donation data confirms this hypothesis. Both candidates rely more strongly on medical sector employees and less so on retirees than their competitors for the Democratic Party nomination. Overall, the O'Rourke and Buttigieg's donor bases are more similar than any other pair outside of Booker and Biden.

### **The Retirees Puzzle**

As two of the nation's most famous progressive icons, Elizabeth Warren and Bernie Sanders have fairly similar donor bases. Both are equally reliant on researchers (Warren is 2<sup>nd</sup> and Sanders is 3<sup>rd</sup>), teachers (Sanders is 1<sup>st</sup> and Warren is 2<sup>nd</sup>), and, perhaps surprisingly given their political stances, tech sector workers (Sanders is 2<sup>nd</sup> and Warren is 3<sup>rd</sup>). They are also similarly spurned by business executives and financiers (they're the bottom two for both categories).

But there is one sector of donors that strongly prefers Warren to Sanders: retirees. While Warren relies on retirees for a robust 13% of her donor base, Sanders' fellow older Americans make up only a paltry 2% of his donor base. A perusal of the five candidates most reliant on retirees for their donations hints at the distinguishing factor between Warren and Sanders in the eyes of retirees: in order, the top five are Rep. Tulsi Gabbard, Sen. Amy Klobuchar, Harris, Warren, and Gillibrand.

Why do retired donors so strongly prefer female candidates? It isn't immediately obvious whether this phenomenon can be explained by more than demographics or the ideological leanings of retired donors. However, even without an explanation, this finding is still noteworthy. [Polls](#) have consistently found that older Americans prioritize electability much more so than do younger Americans. At the same time, [research](#) also suggests that gender clouds voter perceptions of electability, with white, male candidates being perceived as [more electable](#) than women and people of color.

The fact that women running for the Democratic nomination disproportionately rely on retirees for their donor base suggests that the way in which age and gender affect voter preferences is more complicated than conventional polling wisdom would have us believe.

### **Who Benefits from a Misstep**

If current frontrunner Biden were to stumble, who would see the greatest gains in the polls? The possibility that his political parallels -- the other, moderate white men like Michael Bennet or Steve Bullock -- would benefit seems increasingly unlikely by the day, especially now that neither candidate is likely to make the [third debate stage](#). Another school of thought argues that

Sanders, as another older, well-known, white man with much more progressive beliefs than the former vice president, would pick up the majority Biden's supporters. [As the story goes](#), Sanders is the second choice among 29% of Biden supporters, outpacing Harris (20%), Warren (18%), and the remainder of the field.

Our analysis suggests another possibility: Booker and (to a lesser extent) HUD Secretary Julian Castro would instead be the main beneficiaries. Although their policy platforms are more liberal than Biden's, they rely on a similar mix of white-collar professionals for their donor bases. As such, a decline in Biden's fortunes might be exactly what they need to revitalize their flagging campaigns.

Declines by the other frontrunners have less clear effects in our analysis. If their first choice is unavailable, Sanders supporters might flock to his closest ideological competitor, Warren, but many may also move to the other two candidates that appeal to young, white men: Buttigieg and O'Rourke. Similarly, some Warren supporters might default to Sanders should her campaign become unviable, but others may also turn to Harris.

Ultimately, whether examining the professional choices of each candidate's donor base adds any information beyond that which can be gleaned from existing polls will become clearer as the campaign progresses and more candidates drop out. At the very least however, this data still offers the possibility of explaining why some of 2020's most promising candidates have faltered, and potentially, what might lie ahead should some of the current frontrunners unexpectedly stumble.

### **Notes on Methodology**

[1] The industries we used were: Retired, Law, Finance/Accounting, Low Level White-Collar, Government/Nonprofit, Medical, Entrepreneur/Manager/Executive, Real Estate, Researcher/Professor/Engineer, Agriculture, Teacher/Writer/Librarian, Technology, Religious, Service/Retail, Arts, Military/Police.

When we couldn't identify an industry from the employer/occupation data, we labelled them as N/A. We generally tried to sort by answering the question "what type of work does this person do?" which meant that, for example, a secretary at a law firm would go in "Low Level White Collar" and not "Law".

[2] Given that there were ~300,000 donors in our dataset, we could not label every single one by hand. Instead, we hand classified about 30,000 unique professions. Then, for the remaining donors, if their job description exactly matched a profession we hard-coded, they would be labelled as being in the same industry. If their profession did not exactly match a profession in our database, we used a simple multinomial naive Bayes classifier with normalization trained on our hand-labelled data.

[3] Broadly, Principal Component Analysis is a form of feature extraction where we convert a list of variables – in this case, our 17 industries – to an equivalent list composed of combinations of the original variables. These new variables (called principal components) are ordered by how well they describe the dataset. In many cases, only a few of the new variables contribute substantially to the data, and we can ignore the remaining new variables. This type of dimensionality reduction makes it much easier to visualize complex datasets (for a more detailed discussion of PCA, see [2]).