Pushpay Data Analysis Test

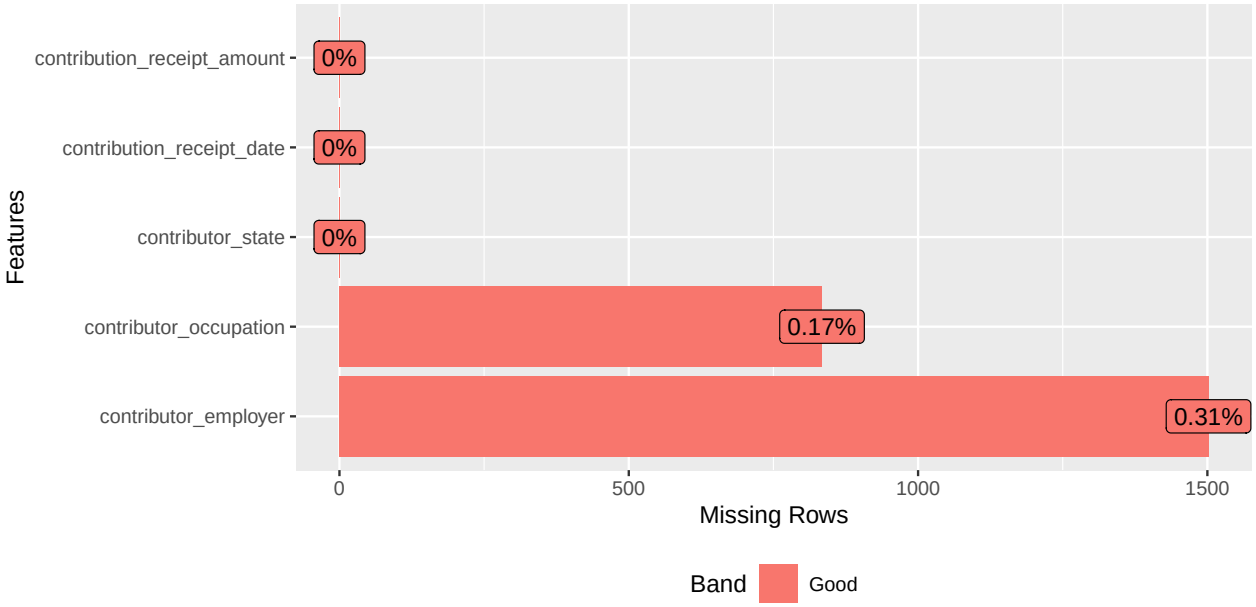
Tushar Sawhney

29 May 2021

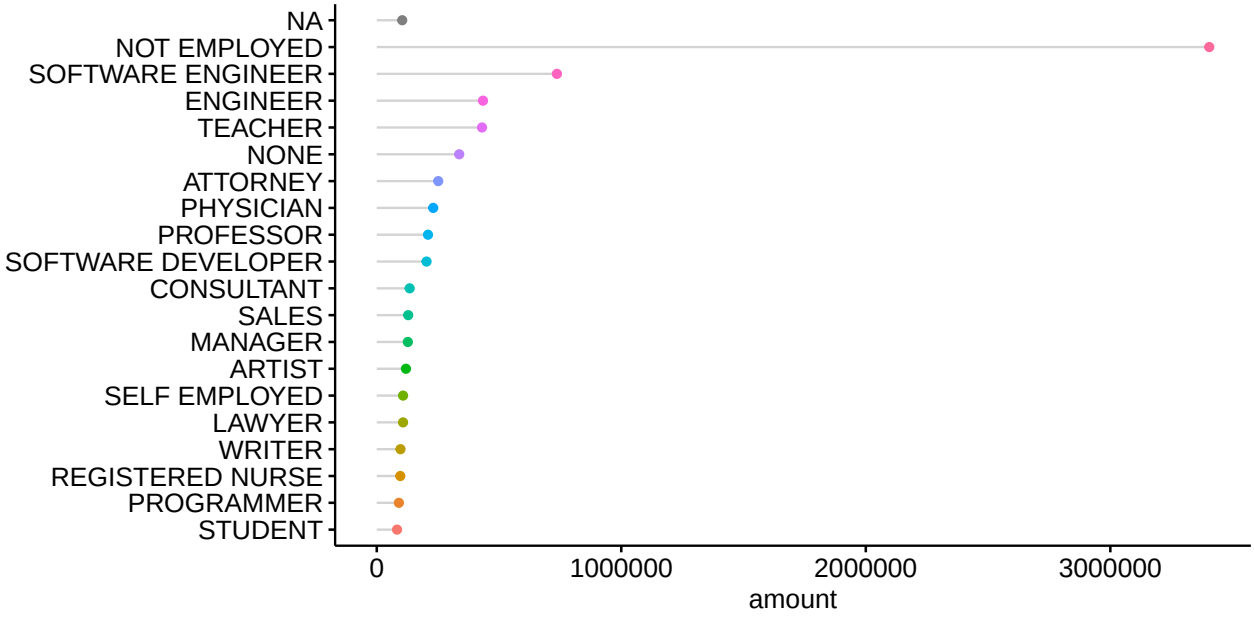
For the given task, and with the provided data, we have been asked specifically to check for the following criteria,

- Occupation
- Employer
- State
- Rate Over Time

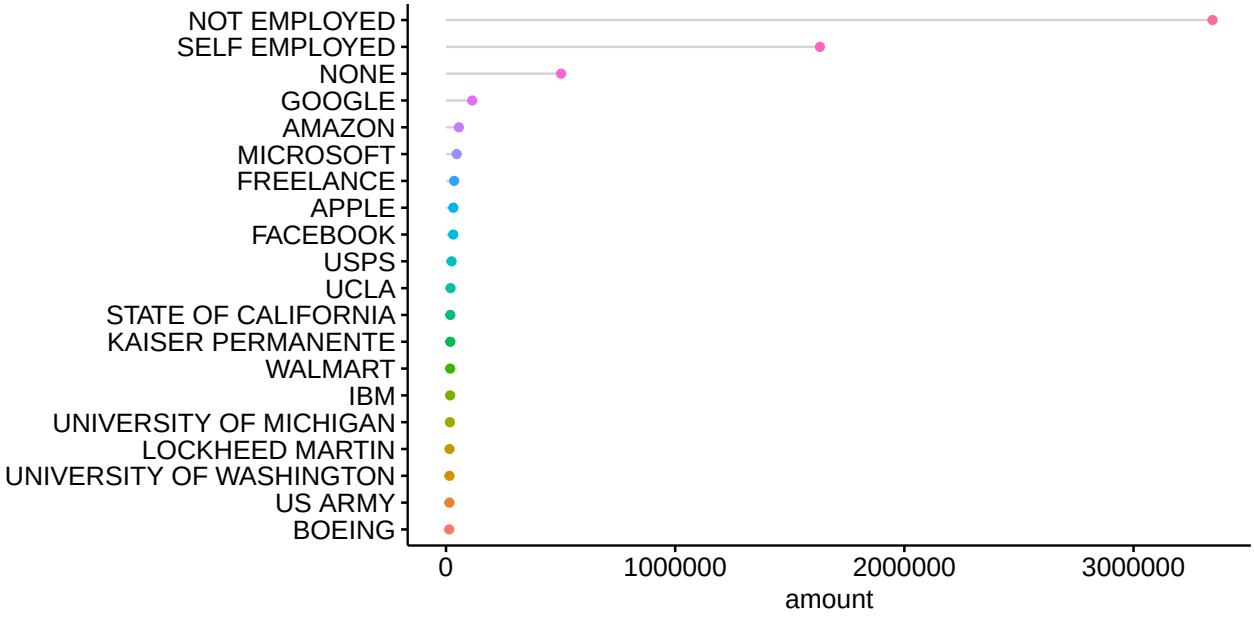
We can notice that from the derived data below, only 0.17% of Occupation and 0.31% of Employer data is missing. This gives us confidence to analyze further, as most of the data is NOT NULL.



From the following plot we can conclude that most of the contribution has been from the donors who were not employed then. They can either be unemployed, or retired people as well. It was followed by Software Engineers and Engineers. NONE here indicates that the donor didn't mention their employment. This maybe because they recognize the impact that policies such as Medicare for All and student loan debt relief could have on their well-being.

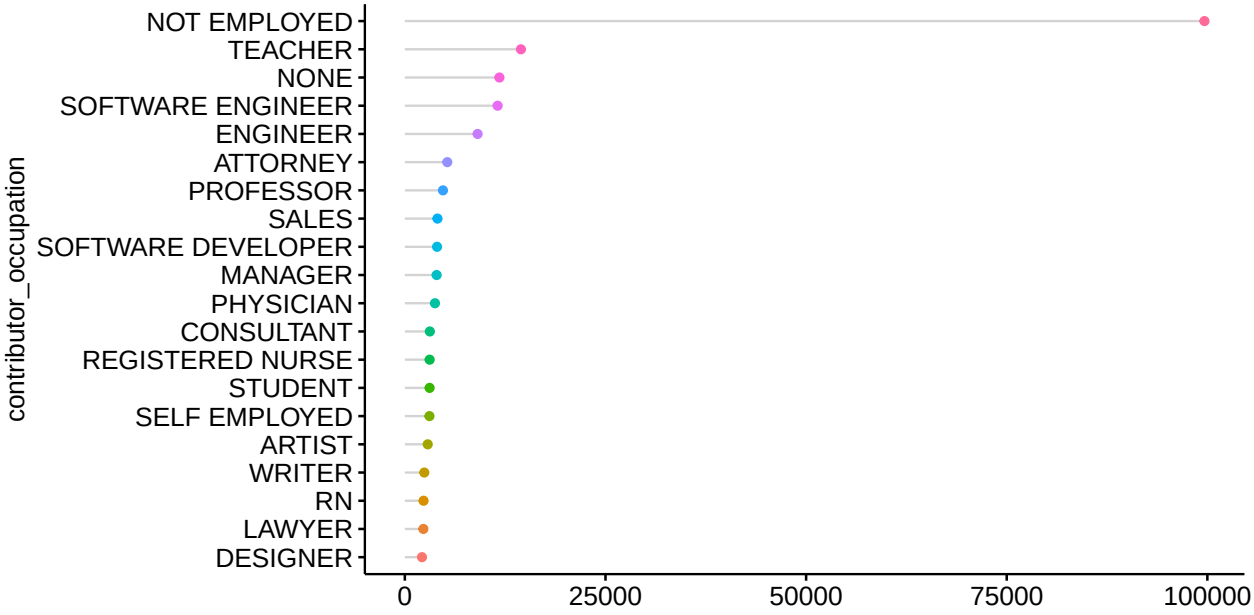


In terms of donors with employement, the ones who are Self Employed have donated the most, followed by those employed by Google, Amazon and Microsoft. NA here represents missing employer, and NONE represents donator didn't wish to disclose their employer.

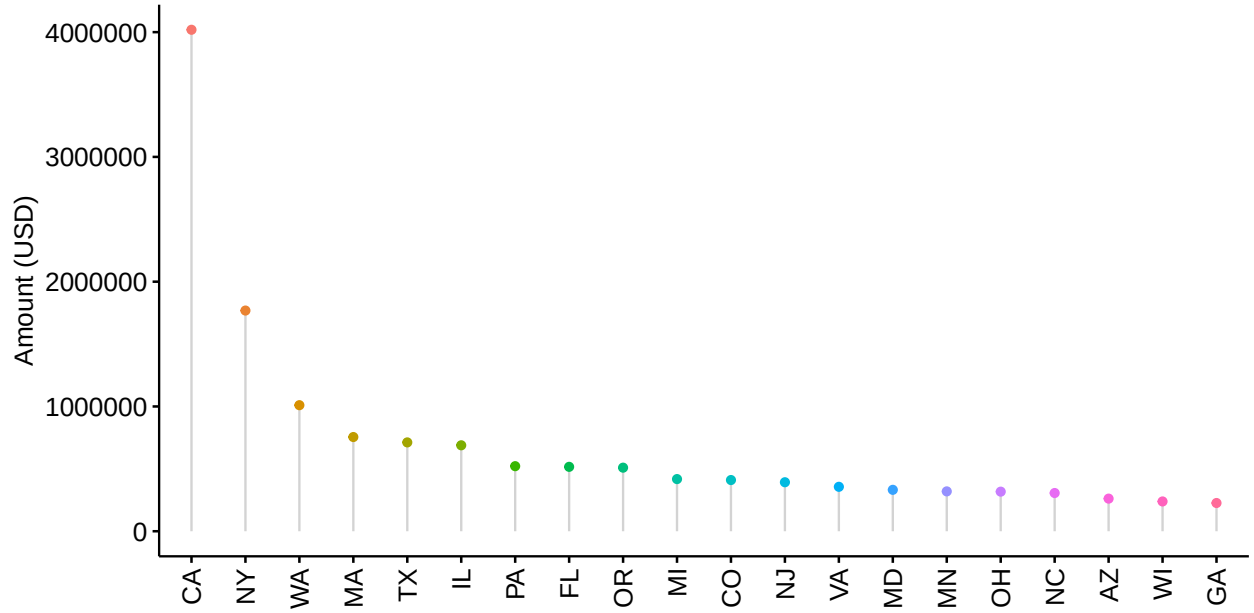


However, the number of contributions by donors by profession can be different from the sum of donations.

We can notice that even though the Software Engineers contributed to the greatest amount of donations, it was the number of Teachers made highest contributions in terms of the count. This means Donation/Profession was greatest for Software Engineers, even though more number Teachers donated.

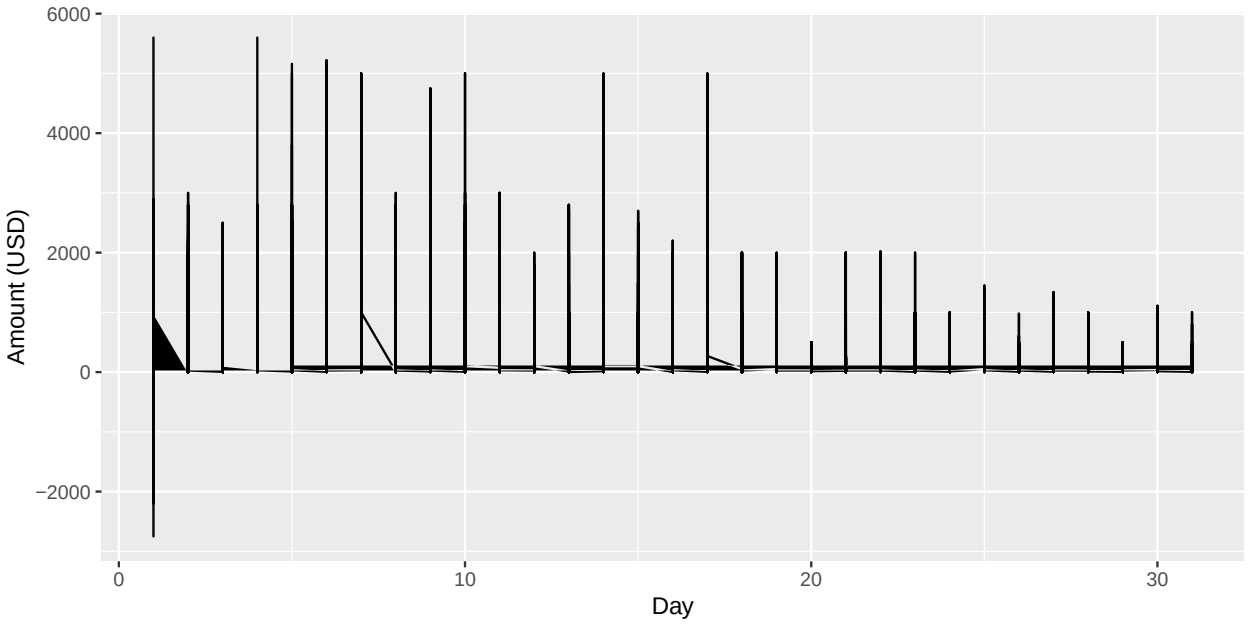


In terms of States, California witnessed the greatest contribution, followed by New York and Washington. This makes sense as presidential candidates generally raise the vast majority of their campaign dollars from donors in states with the largest concentrations of population and wealth.



The data available to us was from 2020-03-01 to 2020-03-31. From the plot above, we can notice that the maximum donation was made in the first week of March, followed by second.

There exists some negative contribution value as well, which I think is some error in the recorded data.


6000
4000
2000
0
−2000
Amount (USD)
0
10
20
30
Day