What informs charitable giving

Faryal Usman

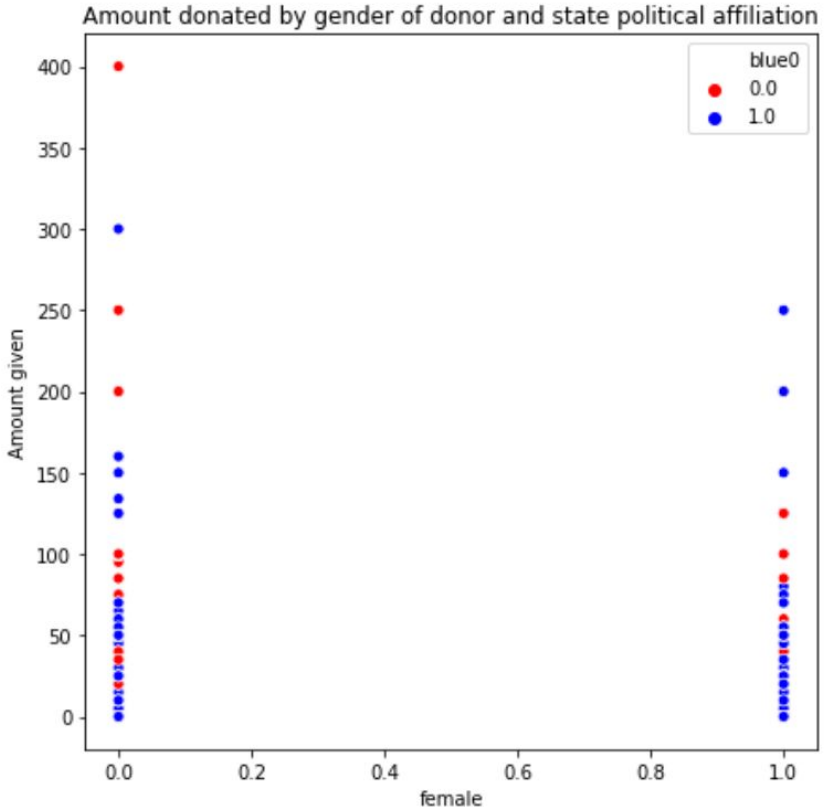
Paper and Data Background

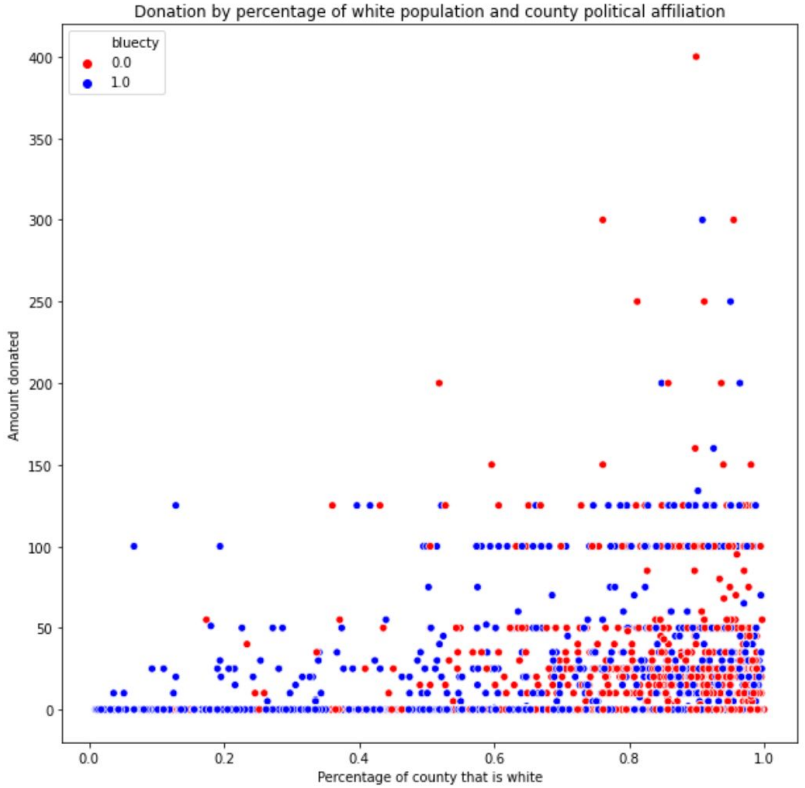
- The authors use a field experiment to see what treatment conditions, assigned to donors of a not-for-profit contribute to donation amount.
- The data is comprised of
  - State and county level demographic variables that are also economic and politcal (e.g. state or county level political affiliation, racial composition of area, proportion of urban population of area, median household income etc.)
  - Specific demographic variables of donors (gender of donor, whether donor is a couple, the time since last donation, the highest previous donation amount, the frequency of donation)
  - Various treatment conditions
- Treatment conditions are comprised of
  - General treatment condition
  - Matching ratio
  - Suggested donation amount

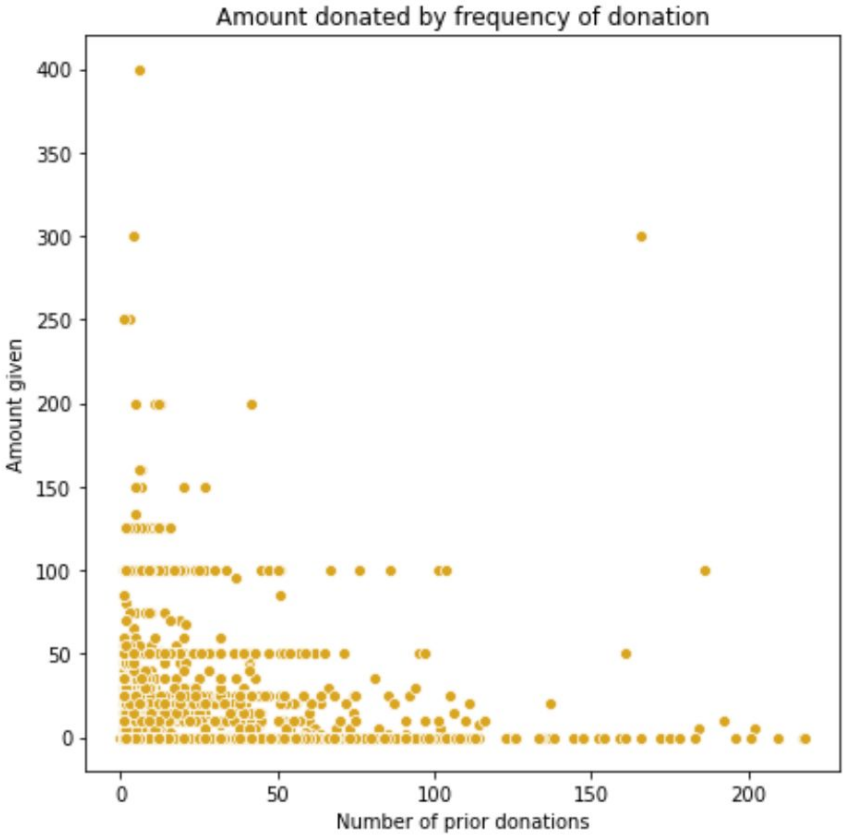
Research questions

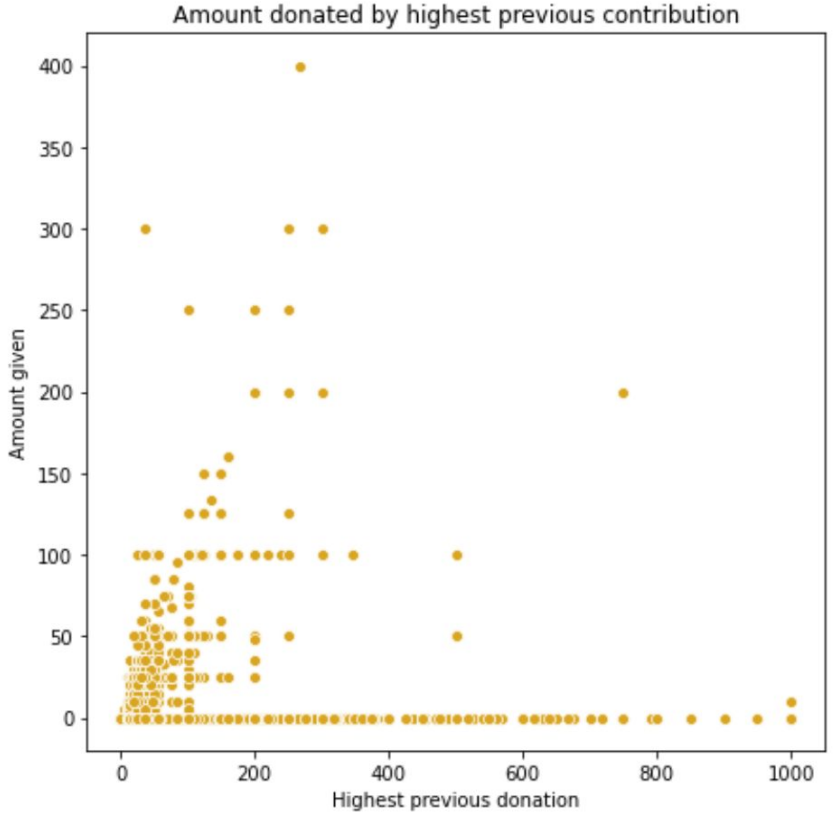
1. For the variables gender of donor and racial compostion of where the donor comes from, is there a relationship between these variables to the dependent variable of amount donation in a signficant way? If so, what is the relationship?
2. What variables in general are the most important when it comes to informing amount donated?
3. For the various treatment conditions (ratio of matching donation, donation amount suggested), is there an association with amount donated?

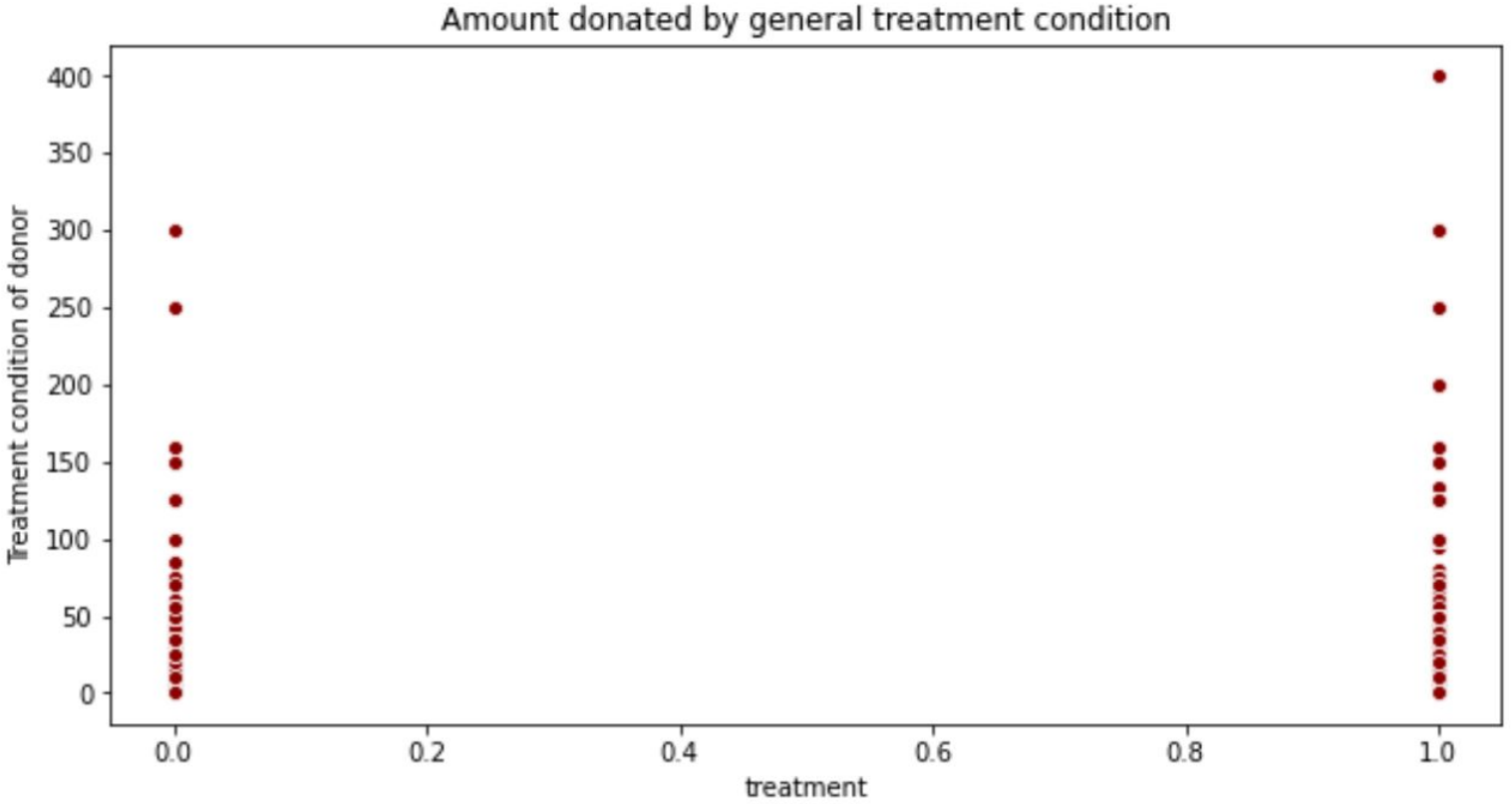
Analysis Part 1: EDA



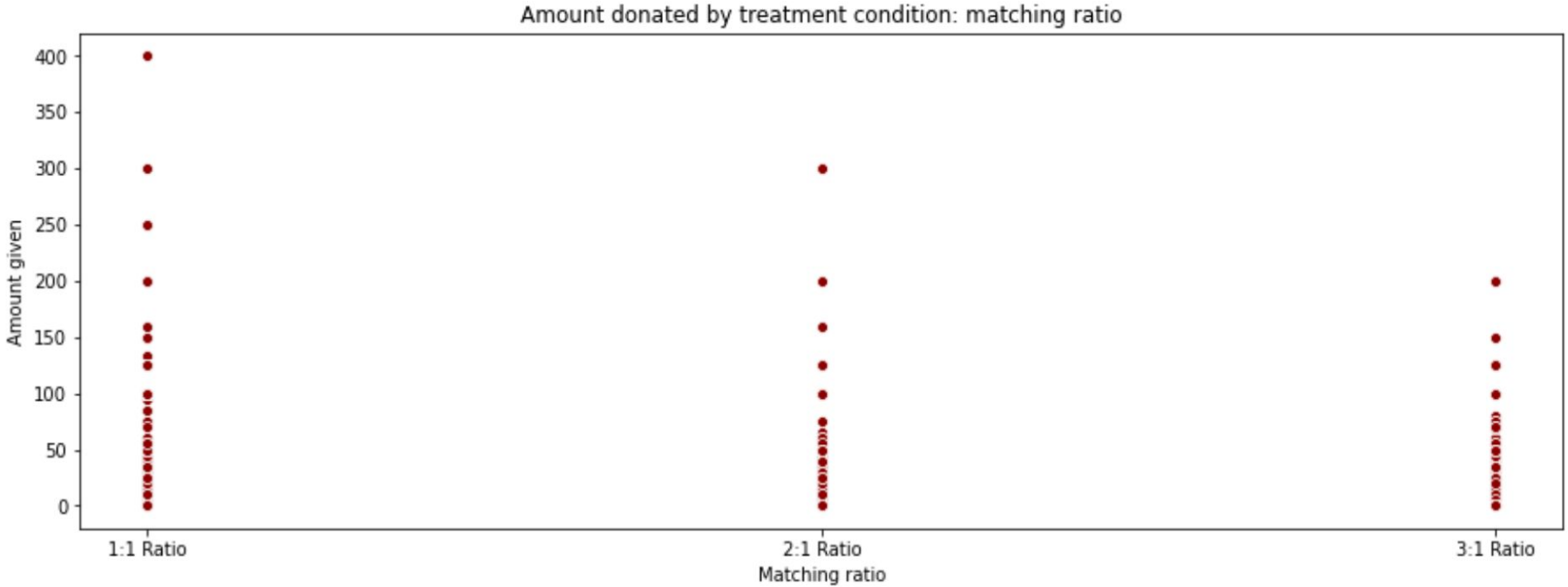



Amount donated by frequency of donation
Amount given
Number of prior donations
0
50
100
150
200
0
50
100
150
200
250
300
350
400

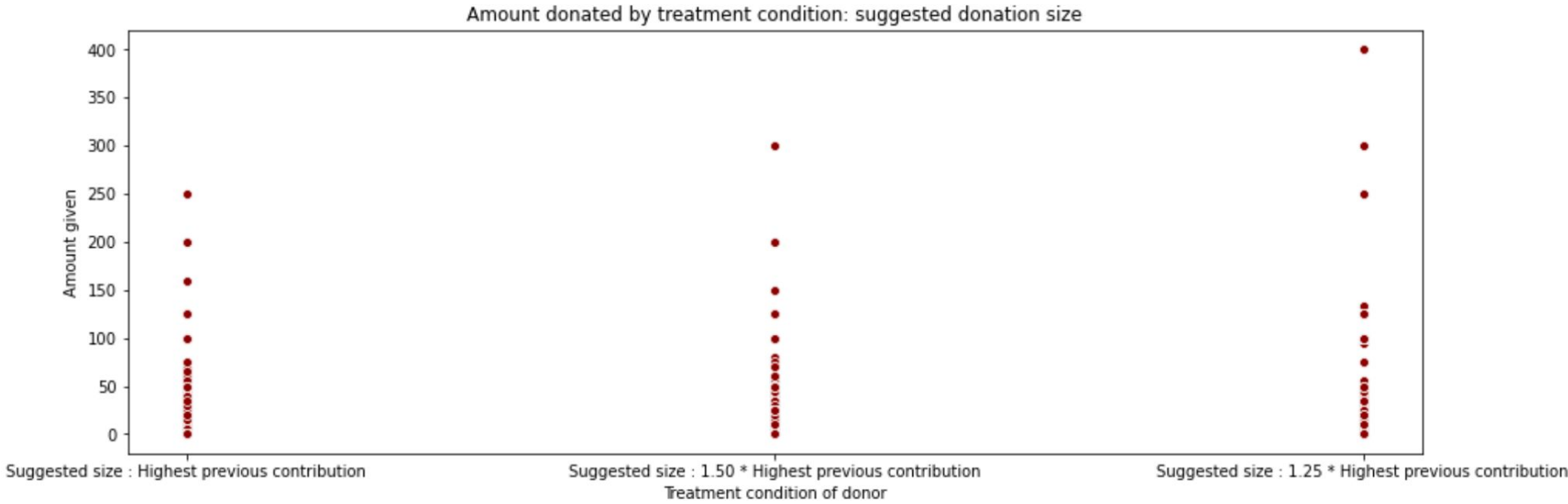

Amount donated by highest previous contribution
Amount given
Highest previous donation
0
200
400
600
800
1000
0
50
100
150
200
250
300
350
400



Treatment condition seems to have higher donation



1:1 Matching Ratio seems to have the highest donations



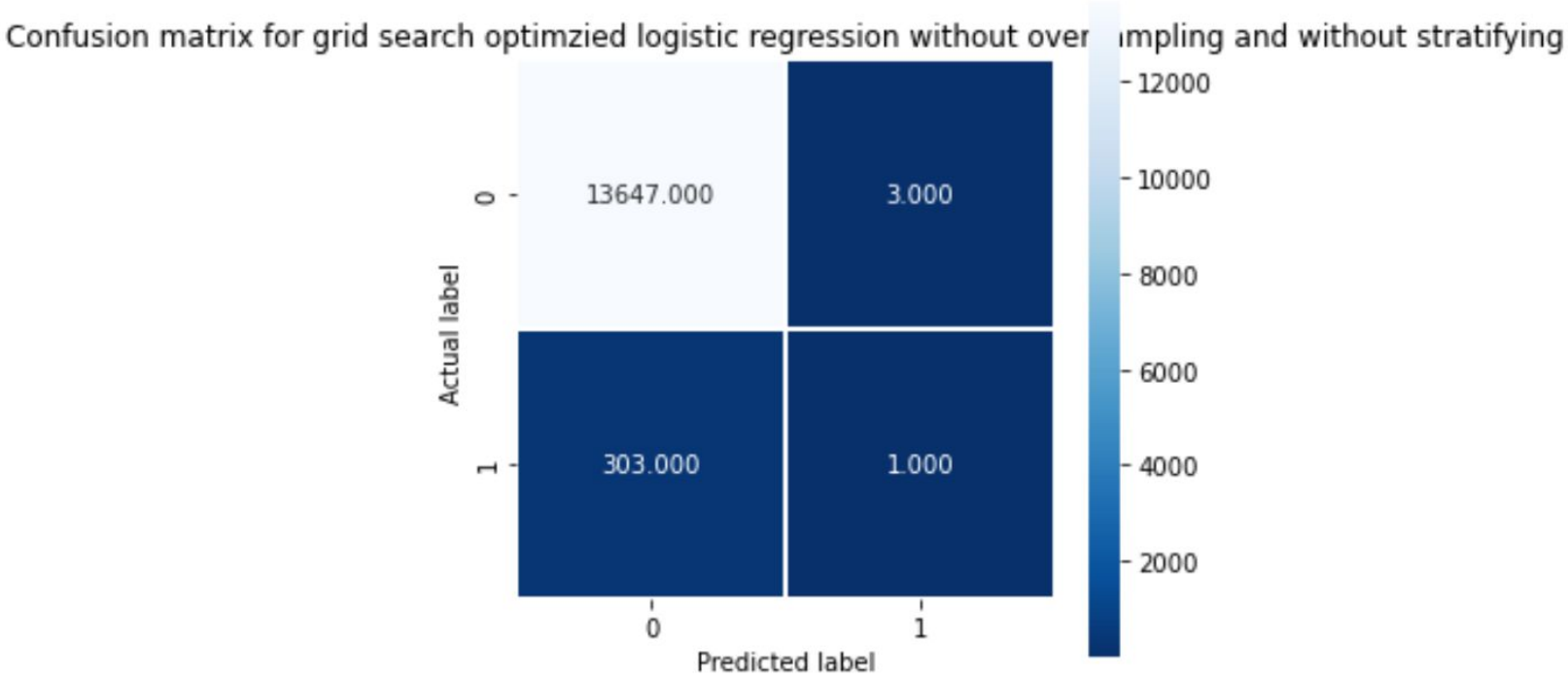
1.25 * Highest previous contribution suggestion has the highest donations

Analysis Part 2 .1 : Machine Learning Linear Regression

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.8069	1.055	0.765	0.445	-1.262	2.876
ave_hh_sz	-0.3036	0.194	-1.565	0.118	-0.684	0.077
blue0	0.3650	0.524	0.696	0.486	-0.663	1.393
bluecty	-0.4976	1.282	-0.388	0.698	-3.009	2.014
cases	0.0298	0.049	0.602	0.547	-0.067	0.127
close25	-0.1099	0.129	-0.854	0.393	-0.362	0.142
couple	-0.3414	0.144	-2.377	0.017	-0.623	-0.060
dormant	-0.1993	0.118	-1.683	0.092	-0.431	0.033
female	-0.1385	0.093	-1.495	0.135	-0.320	0.043
freq	0.0622	0.005	12.965	0.000	0.053	0.072
hpa	0.0055	0.001	8.681	0.000	0.004	0.007
ltmedmra	-0.5763	0.087	-6.616	0.000	-0.747	-0.406

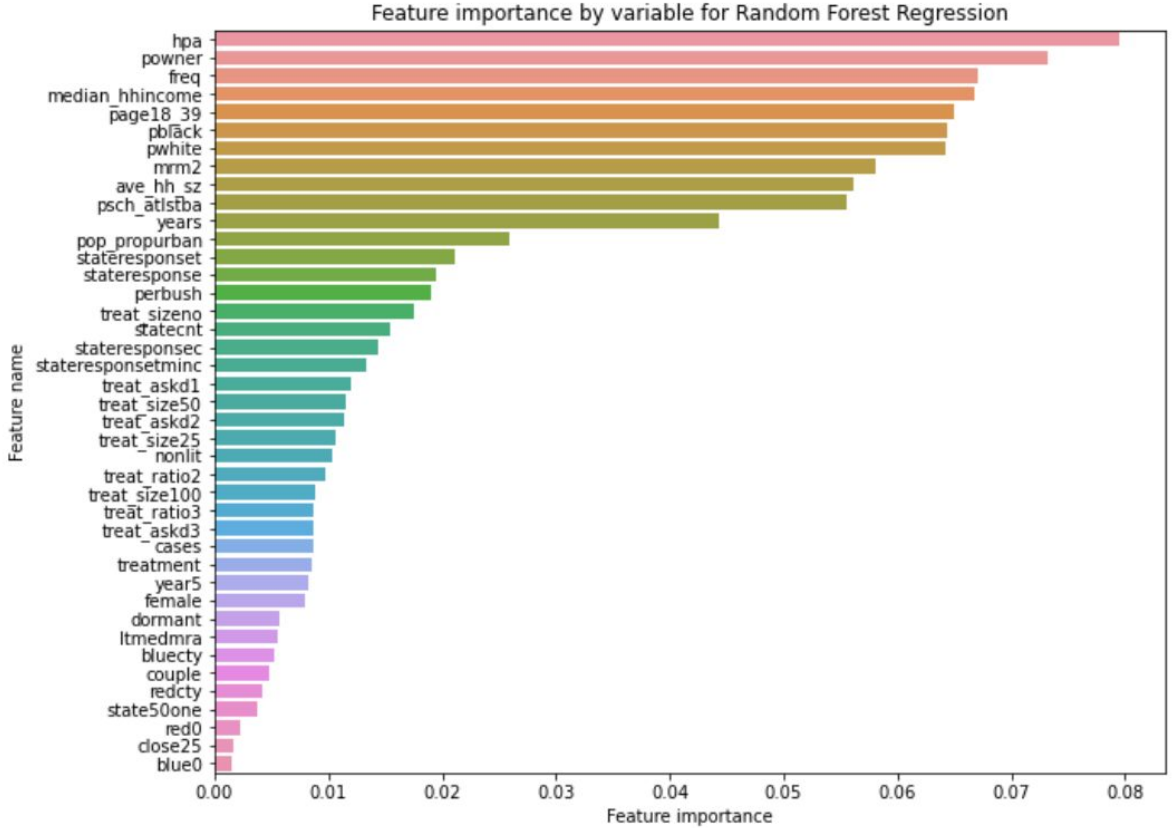
A majority of the results variables that were demographic based were not significant. On the other hand variables like freq (frequency of donation) and hpa (highest previous contribution) were significant . This is because they were specific to the donor.

Analysis Part 2 .2 : Logistic Regression classification report

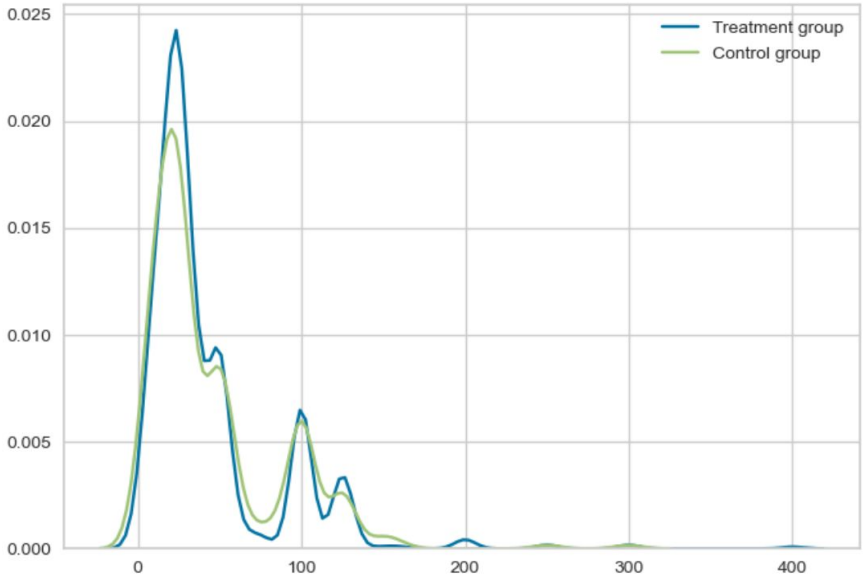


Logistic regression has type 2 error (bottom left corner). It is misidentifying donors as non-donors

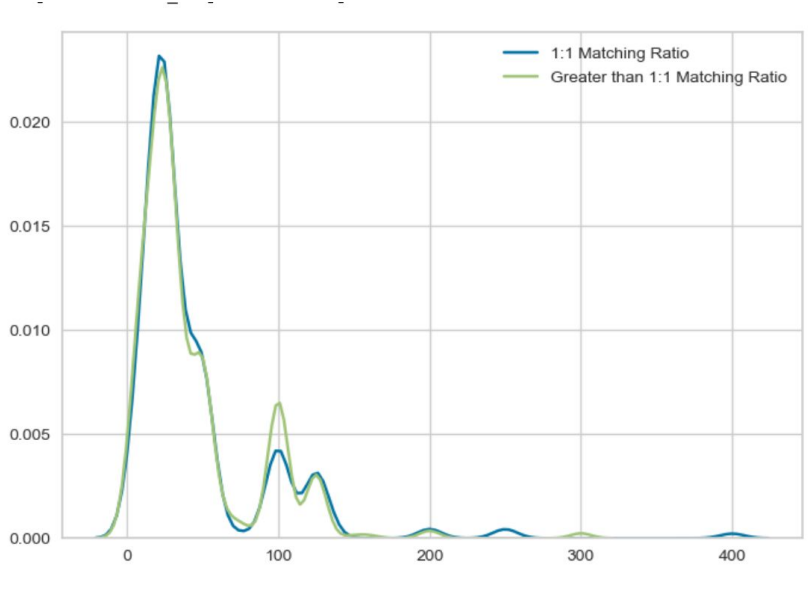
Analysis Part 2 .3 : Feature selection Random Forest



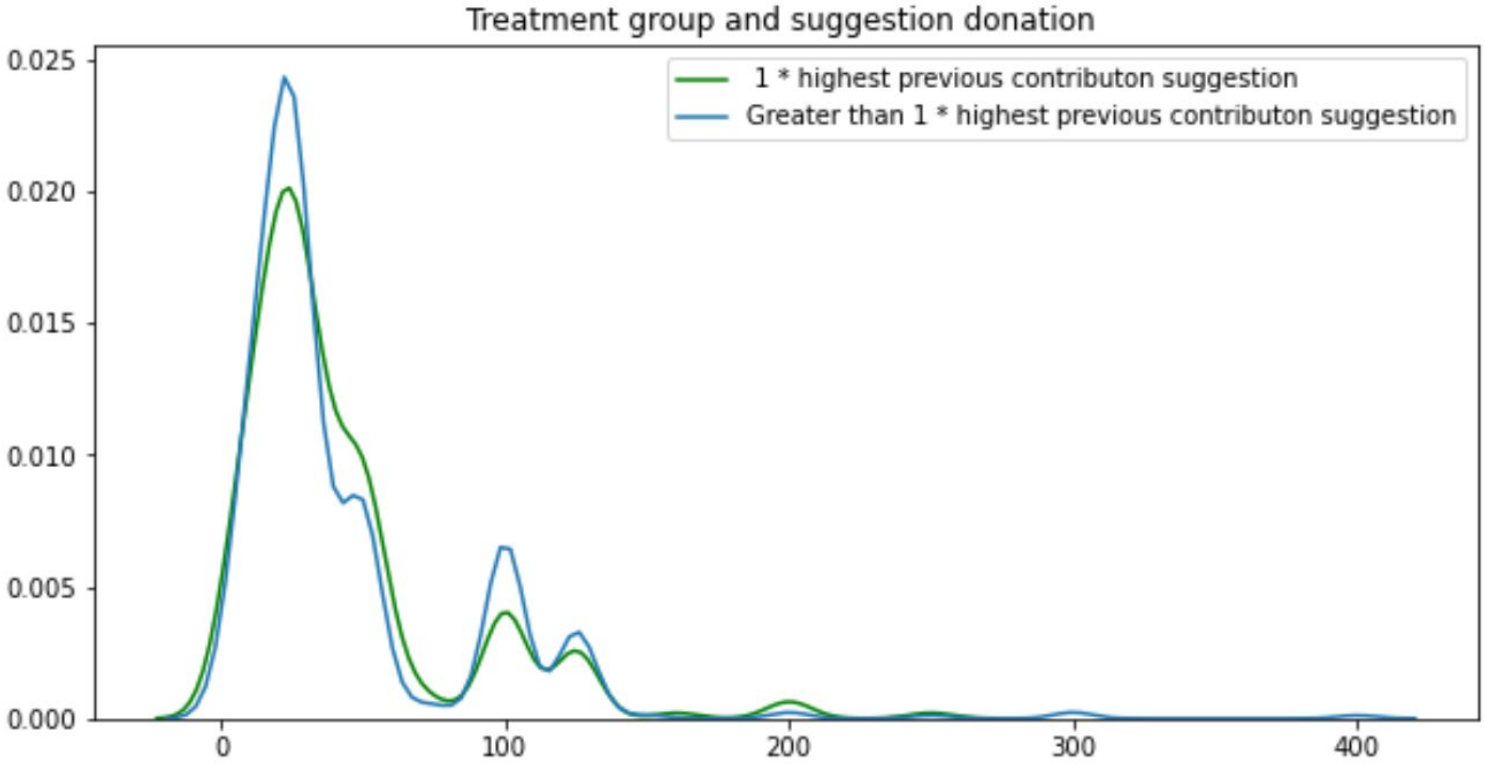
Analysis Part 3: Statistical testing of treatment conditions



P value: 0.0555.
Null hypothesis rejected, no difference in the mean donations of two groups



P value: 0.669
Null hypothesis rejected, no difference in the mean donations of two groups



P value: 0.429.
Null hypothesis rejected, no difference in the mean donations of two groups

Conclusion

- Between the treatment conditions tested, no significant difference in the treatment conditions.
  - Visual difference in the treatment conditions
- Demographic variables that are non-specific are not a good indicator of donation amount
- Donor specific variables (not gender) such as highest previous donation, frequency of donation and time since previous donation are signficant.

References

Data:

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910/DVN/27853

Paper:

https://www.nber.org/papers/w12338