Homework 1 Donors

Kyle Walter

5/9/2021

## Introduction

For this assigment, I have been provide a data set that with various data about people who have made donations to an organzation.

These variables include the gender of the donor, the number of children that they have, the average give, time since last donation, income band and wealth band for their respective areas and whether not they own their repective homes.

I want to see if the data can help us answer three questions 1. Do Men or Women donate more money? 2. Does the wealth or homeowner status of a donor increase their donation amount? 3. Lastly does the size of the person family tend towards more donation?

## Data Exploration

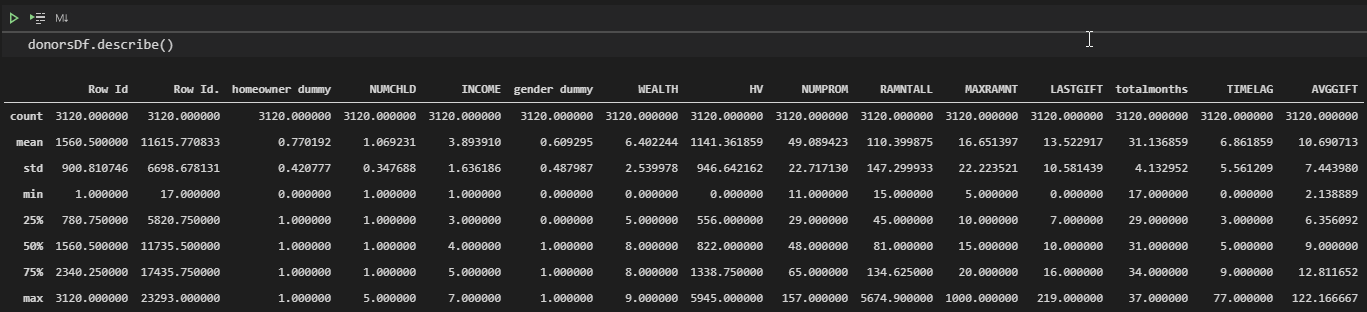
First thing I did was call the data into Python as Pandas Data Frame. This is a rectanglar shapped format similar to table in a database or a worksheet in excel with which to store an manipulate the data for the analysis.

The second step I did after reviewing the data dictionary that came with the file was removed the columns that will not provide information relative to this analysis.These focus on regional based information in the data such as which of the 5 regions the donor lives in, the median income, average income, and poverty line information for their region. While this data set is small, since python stores the object in memory while working with it, removing unneeded columns frees up memory that may be later on.

The next step I thought about looking at was one of the questions around who is more likely to be a 1x time donor. To do this I checked if the LastGift and lifetime giving were equal and saved them into a separate data frame. Then when looking at the statistics the “TIMELAG” in this new data frame, I noticed all the values were 5. Per the dictionary it looks like this field represents the time in months between the first and second gifts. There is definitely a data quality issues here, however; as I cannot ask the person who created the data, marked it as such and moved on to other questions I may be able to answer using this data.

## Question 1 Do women or men donate more money?

Since this question can likely be answered through continuing the Data Exploration, I’ve decided to add it here. For the next step, I returned to the full data set and took a look at the summary statistics using the describe method in pandas.

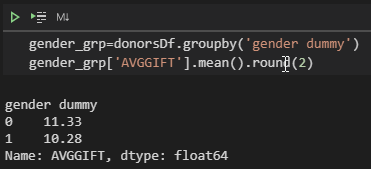


Summary Stats

As we can see from the summary statistics the data itself has slightly more women donors than men. The “gender dummy” field in the data dictionary says that the value of 1 equals female and 0 equals male. With the average around .6 means we have about 60-40 split of women to men in our data set. So women are more like to be converted to donating, but we do not know yet if they also donate more.

Other quick items to point out are the wealth bands of our donor population are an average of 6.4 so the average donor is in the top 3rd of the wealth bands. What is a bit more telling is the lower quantile of donors 25% are in wealth band 5 or less while by the halfway point of all donors we’ve reached wealth band 8. The population of our donors are relatively wealthy, however; when looking at income 50% of them fall between income of 3 and 5 so our population is relatively middle income.

Now returning the question at hand, let’s see if women or men from out dataset donate on average more money.



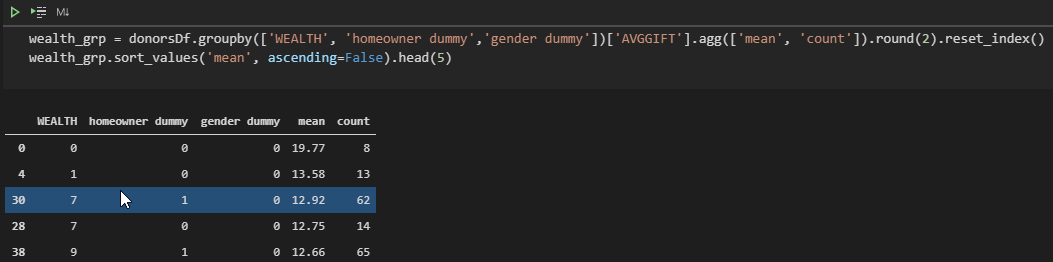
Average Amount Donated By Gender

As we can see from the output, men in the data on average are donating about 1 dollar more than then women donors despite the fact there are more women in the data set.

## Question 2 Does the wealth or homeowner status of a donor increase their donation amount?

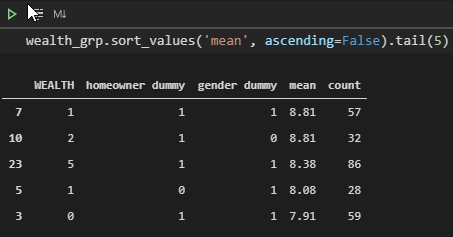
So as we noticed earlier it seems like the majority of our donors were home owners and in the higher wealth brackets. Do these happen to also be our highest donor population?

I used python to create a group data from by the wealth category, home ownership, and gender, and used this to report the top 5 highest donation group and lowest 5 donations groups. I did however; find a few surprising things, such as men in the lowest wealth category tend to donate the most with an overage donation of 19.77. However to act as an offset, I then added in a count of how many times this group is seen. 8 in total out of a data set of 3120 records. So it seems like this is likely 1 larger donation that that is dragging the group average way up. As seen below most of the others are in the 12 to 13 dollar range. And the third option Men who own their homes in wealth band 7 have 62 records and are likely the highest donor category.



5 Highest Donnor Groups

In the groups with the smallest donation women homeowners in the lowest wealth group are provide the smallest donations. This we have 59 records making up this row, and feels intuitively correct as they would like have lowest means by which to provide a donation.

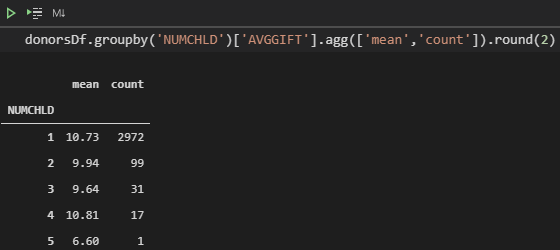


5 smallest donor groups

We also notice that with the exception of the middle record these seem to follow the the donors wealth band is most likely to tell us who will donate the least.

## Question 3 Does the size of the person family tend towards more donation?

Lastly, I wanted to know if the size of the household tended to drive the size of the donation



Number of Children

The data as seen above does not really lend itself well to the question at hand. As we see the largest donation is by households with 4 children, but there are only 17 records while households with 1 child are the next largest but make up almost the entire data set.

This could point to one of two things: 1. People tend to stop donating as their families grow. 2. The average family size can no longer be figured using their children since the number of children has dropped.

I tend to lean towards this data set representing the first, mainly because while I wouldn’t expect see a few hundred records for people with 4 kids, my own observations are that most people still have 2 children and would have expected a near even split between 1 and 2 child households.

The other issue that may come up with regular donors is we collected information from them at the time they started donating and the data itself has not been update to reflect changes in their “static data”.

Lastly, the numbers for 2,3, and 5 children do tend to follow the pattern as the number of children grows, the donation size will drop as resources become more scarce. There really are not enough observations to claim this is significant.

## Conclusion

Based on the smaller data set at hand, it appears that the gender and wealth group make a small difference in donation size. While those with large donations tend to be homeowners, the data itself says about 70% were homeowners. In Addition the lack of variable sized families in the data make it hard to tell if household size is impacting the donors decision making process in their size of the their gifts.