Individual campaign contributions for multiple years

*(this program is written for Insight Data Engineering - Coding Challenge 2018, If you want to use it for any other purpose please contact* [*diyar82@gmail.com*](mailto:diyar82@gmail.com)*)*

1. How to run and modify
   1. Th program is written in linux/bash for running it make sure it is still in executable mode (chmod a+x run.sh)
   2. The program counts donations made within same calendar year and previous years as repeated donations. You may modify it by changing statement “if [ Year2 -le Year1]”. Given Year2 is the donation listed former to donation Year1
   3. The program may take a while if your data is bigger than 1M lines. In that case please uncomment the progress counter ( ##echo “The Zip code” … 🡺 “The Zip code”). So you can track the progress.
2. General approach
   1. First the data is cleaned: In this step any unwanted, redundant and invalid data is removed.
   2. Label the data: this step is necessary so the chronical order of the data is preserved through the process
   3. Sorting the data: In this step, to decrease the computational cost data are sorted according to their Zip code. Instead of comparing data with the entire document, the data is only compared to the same zip code.
   4. The actual calculation. It uses to for loops that are iterating opposite to each other to catch first donor reoccurrence easily
   5. After the calculation the data is sorted back according to its chronological order and the label is removed to get the last desired output format
3. More Advanced computation

I have also created a more advanced code in which percentile for each individual recipient is calculated also, yet the “run\_tests.sh”, Provided by you, did not like the format. I am very happy to share that also

1. IMPORTANT

Since I applied for data science and data engineering program I got two emails from you. I uploaded the very same program for both of you