Author: Jeff P.
Fall 2018, CIT-129, CCAC

Access link (github):

https://github.com/milominderbinder22/cit129/tree/master/finalProject

Description:

This project is a jupyter notebook that reads-in a .csv (provided in git repo) of anonymized donor data for a local non-profit (gift amounts by year, appeals to the donor to give by year, zip code of the donor), and provides some baseline tools for analyzing the data by year and zip code.

Instructions for use are provided in each Jupyter notebook cell (along with code documentation and notes on challenges and uncompleted but desirable tasks). The primary functionality from the user's perspective consists in the ability to select a donation year (such as '08-09'), and create a dataframe of the top-giving zip codes in that year. The default is an aggregate donation amount of over $1000, zip code-wide. In addition (and which was the impetus for this project), columns in the dataframe display metrics regarding the number of appeals for donations made to that zip code, along with derived or calculated columns regarding the efficacy of those appeals (both in dollar, and percentage terms).

Running further jupyter nb cells gives a statistical summary of the dataframe, followed by a colorful bar chart of the top-giving zip codes, and their "Response Rate" to appeals.

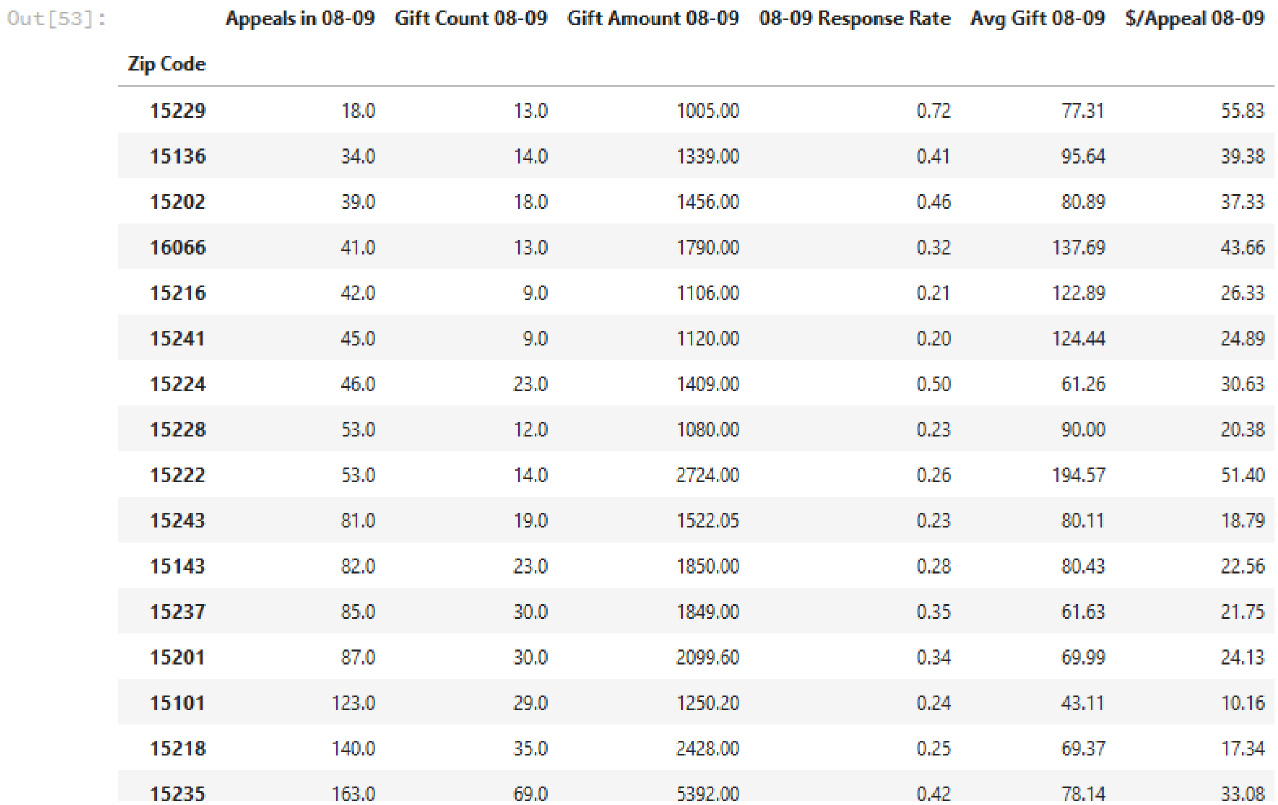
Notes on the workflow:

The data was cleaned with OpenRefine. (Notes pertaining to this process are in the jupyter nb's, themselves). Much of this cleaning could have been done programatically, but it may have been fortuitous that it was not. Eyeballing the data in an OpenRefine facet revealed that 1286 zip codes of donors and potential donors were listed as "UNKNOWN." Investigating the magnitude of the potential error this may cause may be a worthwhile endeavor. Additionally, examining the data in OpenRefine facets revealed the presence of substantial outliers in donation amounts. For instance, one congregation donated a sum of $500,000 as a single gift. The zip code of that particular congregation happened to be 15147. It is a mathematical certainty that the average gift size (just for instance) is skewed for that zip code during that giving period. Inasmuch as the impetus of this analysis project was to examine the efficacy of individual giving, being aware of the sheer magnitude of such outliers is a good starting point, and omitting them and reanalyzing the data may be an appropriate next step. For that matter, one might well omit any donations from congregations or family foundations, as well. While they are important sources of development, one imagines that their choice to give or not give is not made on the basis of a mailer (an appeal).

Difficulties/Challenges:

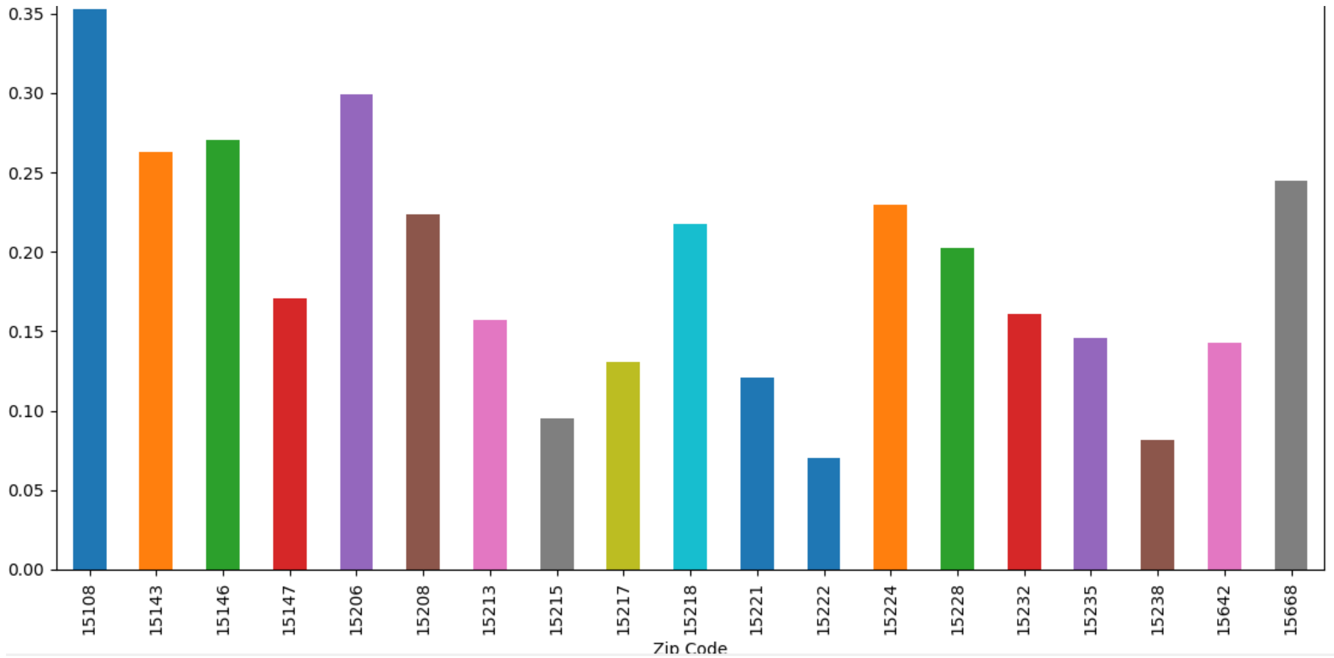
The chart proved to be more difficult than I had anticipated. Much of the discussion about axis values centers around continuous values, not categorical variables like zip codes. Also, a design choice had to be made about whether to calculate and add all of the derived data to a single data frame, or whether to generate it dynamically, per the user's requested year, on each pass. I went with the latter option, because I felt I was telling a story in the jupyter notebook. However, analyzing data across years would become much easier had I chosen the former design.

A few screenshots:



Out[53]:

Zip Code	Appeals in 08-09	Gift Count 08-09	Gift Amount 08-09	08-09 Response Rate	Avg Gift 08-09	$/Appeal 08-09
15229	18.0	13.0	1005.00	0.72	77.31	55.83
15136	34.0	14.0	1339.00	0.41	95.64	39.38
15202	39.0	18.0	1456.00	0.46	80.89	37.33
16066	41.0	13.0	1790.00	0.32	137.69	43.66
15216	42.0	9.0	1106.00	0.21	122.89	26.33
15241	45.0	9.0	1120.00	0.20	124.44	24.89
15224	46.0	23.0	1409.00	0.50	61.26	30.63
15228	53.0	12.0	1080.00	0.23	90.00	20.38
15222	53.0	14.0	2724.00	0.26	194.57	51.40
15243	81.0	19.0	1522.05	0.23	80.11	18.79
15143	82.0	23.0	1850.00	0.28	80.43	22.56
15237	85.0	30.0	1849.00	0.35	61.63	21.75
15201	87.0	30.0	2099.60	0.34	69.99	24.13
15101	123.0	29.0	1250.20	0.24	43.11	10.16
15218	140.0	35.0	2428.00	0.25	69.37	17.34
15235	163.0	69.0	5392.00	0.42	78.14	33.08



Zip Code