Project

Problem:

I want to contribute to a charity that addresses disease X. What charity should I pick?

Input:

Disease: <text box>

State: <dropdown menu>

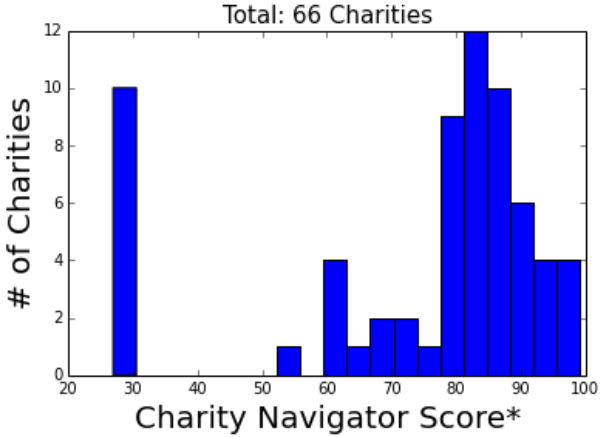
Accredited: <dropdown menu: yes or no>

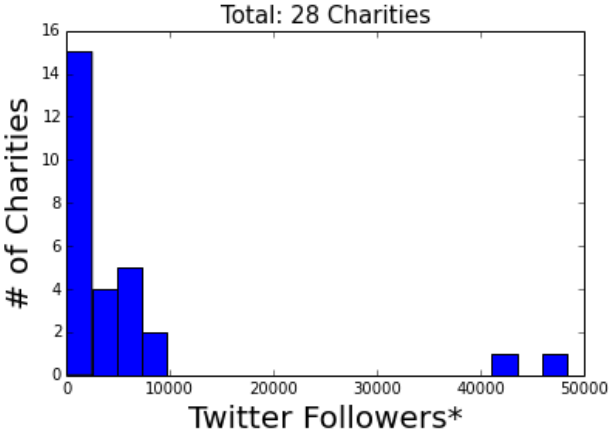
Focus: <dropdown menu: research, outreach, etc.

Output:

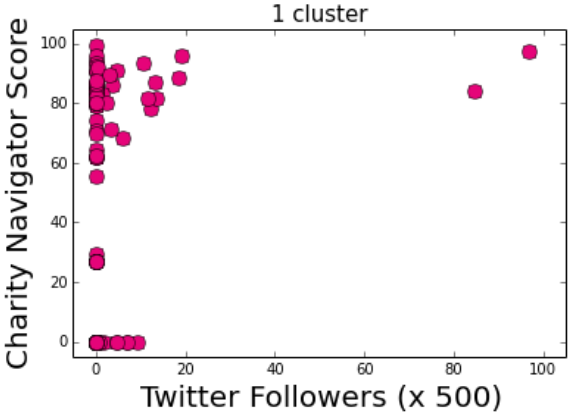
Name	Address	City	Charity Navigator Score	BBB Accreditation	Twitter Followers
National Multiple Sclerosis Society	733 Third Ave, 3rd Floor	New York, NY 10017	83.97	Yes	42,255
Multiple Sclerosis Foundation	6520 N Andrews Ave	Fort Lauderdale, FL 33309	86.86	No	6,557

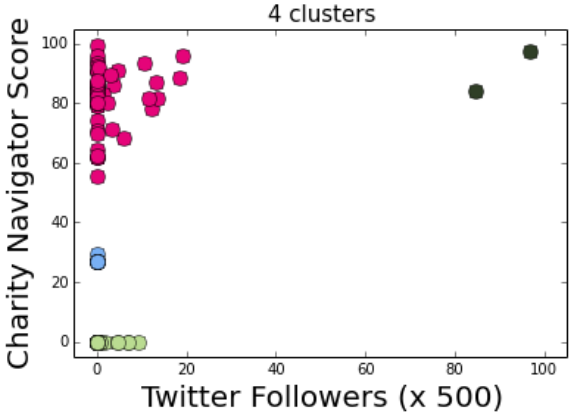
Data

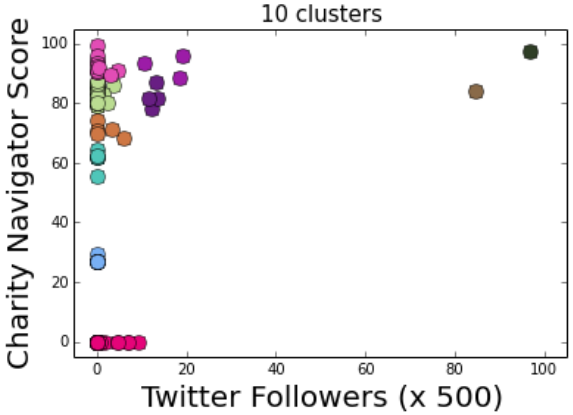




* Histograms do not include charities with no score or no followers.







Algorithm

1. Web scraping
    a. Basic setup working
    b. Need to *pull more info* and *handle irregular data better*
2. Data cleaning and merging
    a. Pipeline written and working
3. Clustering
    a. K-means implemented with two features
    b. Need to *add features*, *evaluate clustering*, and *choose best charities*
4. Output
    a. Basic output done
    b. Need to *allow sorting* and *give recommendations*

Algorithm

1. Scrape charity databases for specified disease.
    a. Preliminary search done.
    b. To do: Add more diseases.
    c. To do: Do more comprehensive search for each disease.
2. Search for additional info on each charity.
    a. Basic search implemented.
    b. To do: Improve cleaning.
    c. To do: Import more information.
3. Merge data from multiple sources.
    a. Pipeline written and working.
4. Cluster charities by quality metrics.
    a. Very basic clustering implemented.
    b. To do: Incorporate more features.
    c. To do: Implement clustering metric.
    d. To do: Choose best charities.
5. Print output.
    a. Basic output written.
    b. To do: Allow user to sort list.
    c. To do: Arrange by score based on clustering.