Weekly Project SCRUM Report

Date: 4/8/2021

Project Name: Deviance or Deservingness? Opioids, Morality, and Economic Precarity

Team Members: Jayden Font, Della Lin, Nikita Jakkam, Michalina Jadick, Divya Gowravaram

Tasks Completed This Week:

- Early insight presentation!!
- Collected data from all of the campaign URLs and stored the data in excel sheets similar to our client's format
- Wrote a script to drop all blank and international campaigns and convert campaign dates to datetime objects to help with plotting and analysis in python-- this file can be expanded for all filtering tasks needed
- Developed figures for certain aspects of data:
- We created data visualizations to explore the full database such as:
    - Plotting the number of campaigns by year to visualize trends
    - Plotting the number of campaigns by keywords to see which keywords have more campaigns
    - Plot the average donation goal reached for each keyword
    - Generating a pairplot with Seaborn to check for correlations in data
    - Generated correlation matrix to further see correlations
    - Number of campaigns per tag
    - Campaigns by state
    - Line plot of shares and donors (normalized and unnormalized)
    - A map of the number of campaigns scraped pere state
    - A map of the locations of all campaigns in the United States (by city)
    - Percent of campaign money raised by keyword
- Sorted URL dates for data analysis on trends

Plans for Next Week

- start analyzing 2017 campaigns
    - Look into natural language processing tools to help build a model determining different types of sentiments in campaign titles/descriptions
        - Potentially using NLTK VADER for sentiment analysis
    - Test prediction capabilities of different models like logistic and multivariate linear regression using our collected attributes (measure success) based on Heather's ranking of keywords
    - Use PCA to reduce dimensions and group attributes that correlate closely into general metrics that measure a campaign like marketing success (shares, comments, etc.)
    - Cluster data and then reverse-engineer keyword weights/sentiments by which keywords appear in each cluster
- https://medium.com/ds3ucsd/gofundme-predictive-analysis-3cdc3218b1ba

Obstacles and Questions:

- Classes end in roughly two weeks, so we are limited in what we are able to complete

in terms of analysis by time
- There's a lot of subjectivity surrounding how we weigh different key words compared to others, which we would like to do in order to further our analysis of how a campaign is framed. We are going to try to implement strategies to reverse engineer this process but we're not yet sure how successful that will be. Heather is going to send us her rankings of the terms based on her expertise and the literature, but since that may not be well-validated in sociology right now, it's tough to classify sentiment when we don't already have predetermined labels for what is positive and negative.

Met with the client recently? When is the next meeting with the client?:
- Met with client on April 7th
- Meetings are biweekly so next meeting with client is April 14th