Project Deliverable 1

Michalina Jadick, Della Lin, Jayden Font, Divya Gowravaram, Nikita Jakkam

----

Sufficient data should have been collected to perform a preliminary analysis of the data and attempt to answer one question relevant to your project proposal which you will submit as a pull request. If data has already been collected for your project you must answer two questions.

Checklist-- Overview

Collect and pre-process a preliminary batch of data

- We are in the process of extracting urls/data from GoFundMe and building our web scraper using https://github.com/lmeninato/GoFundMe/tree/500b2beb0121124551b47f5bf02a6f62f5be6858 as a template.

Perform a preliminary analysis of the data

- Because we do not have a dataset to work through from the start of this project and are building it ourselves, our preliminary analysis of the data focuses on our takeaways from the process of building our data scraper.

Answer one key question

- First question to address: What do we use to define a successful campaign and why? (see below)

Refine project scope and list of limitations with data and potential risks of achieving project goal

- See highlights in limitations section in updated project proposal below.

Submit a PR with the above report and modifications to original proposal

Key Question we are working on :

- What do we use to define a successful campaign and why?

The main metrics we are using to measure campaign success in the preliminary stages are basic quantitative data like the ratio of how much was raised to the campaign goal. Based on the spread of the ratios we obtain, we can then set some thresholds that demonstrate success (for example, 80%). We can also measure the number of donors on the page and compare on average how much each individual donates to that particular campaign. Another straight-forward metric to obtain is number of shares, to gauge success in marketing the campaign. These last two metrics give us a sense of campaign popularity, and we can delve in deeper from there to determine what links there are between that popularity and other attributes like socioeconomic factors, geographic location, etc.

Steps to implementing a successful web scraper for GoFundMe:

1) Getting the URLS based on key terms from our list — we have a version of this built and operational
   a) First step: filter for "opiate"
2) Filter out duplicate urls or empty urls (other forms of data cleaning we can do to remove urls we don't need)
3) Extract data from our final list of urls
   a) Determine location of each features in HTML of GFM webpages
   b) To filter out international campaigns, the first piece of data we check could be the location of the organizer, and if they are not in the US we could skip
   c) Determine which features can be extracted using bs4 and which require Selenium
      i) Bs4: title, name, number of shares, followers, donors, money raised, location, year, links to images
      ii) Selenium (more challenging): updates to the page (how we can track multiple photos added to a page), comment section (difficult because you can't control for how many there are),
4) Storage of extracted data
   a) We will most likely store the URLs from step 1 in a separate .csv file for easier storage and organization
      i) We may save the keyword we used along with each URL so that we can categorize them later.
   b) While running the scraper, we can implement it so that we update a .csv file as it executes, so only data that is successfully scraped will be saved
   c) May need some way to merge data if we store in different files
   d) Should we merge directly with the client's current dataset?

Proposal up to date with latest decisions:

Deviance or Deservingness? Opioids, Morality, and Economic Precarity	
Contact	Heather Mooney hmooney@bu.edu
Organization	Boston University - Sociology Department
Organization Description	Sociology, broadly, looks at how macro-level systems, institutions, and ideologies shape life outcomes for individuals and groups. In this project, I explore how the ongoing opioid crisis, which has killed over 760,000 people since 1999, impacts support people and care workers.
Project Type	Data Science
Project Description	For this portion of the research, I analyze crowdfunded campaigns hosted on GoFundMe posted from 2010-2020. Using a variety of keywords, I explore campaigns related to drug-use and overdose to explore how competing frames of drug use and addiction change over time. In addition to exploring how race and gender impact framing and campaign success, I also explore the different relational, moral, and affective appeals that are made to potential donors online. An estimated 10.3 million people aged 12 or older misused opioids in 2017. 9.9 million people misused prescription pain relievers and 808,000 people used heroin. I hope this will be available to them and their networks, amplifying the impact significantly. This project has both policy implications and theoretical promise. Given the long-reaching effects of COVID-19 and the ongoing opioid crisis (which has been overshadowed by and accelerating since COVID-19 began), it will be important to understand how death, loss, and need are constructed by supporting people in times of (layered) crisis. This research represents a case to explore how morality and deservingness change over time and across populations. More broadly, my dissertation explores the "intersections" of social control and "rehabilitative poverty governance." This project provides concrete benefits by centralizing the experiences of care workers and support people--who are often on the "front lines" of service delivery--in order to further improve existing recommendations, policy, and programming.
Data Sets	N/A - I have been hand coding so I can share that, but not sure how useful it will be. GoFundMe data - campaign keyword search. Examples of keywords include: **Opiate** **Opioid** **Addiction** **Addict** **Heroin** Pain medication Pain medicine Pain killer(s)

	Drugs **Overdose** **Dependency** **Demon** **Recovery** **Rehabilitation** **Rehab** **Fentanyl** Unexpectedly Suddenly Epidemic Battle War ***focusing on the words in bold first, to narrow our project scope**
Suggested Steps	1. Scrape data from GoFundMe from 2010 - 2020 using a variety of keywords **with a strategy to optimize the number of relevant campaigns we extract data for** - There may be limitations to getting all data from 2010 to 2020, so a potential alternative we thought of is to break down the years into beginning (2010-2012) - middle (2015-17) - present (2019-2020) 2. Max 300 campaigns per year, all United States postings 3. Include photo data, campaign information, wall posting, and photo information 4. Include social media tagging and relevant pages 5. Clean data (filter out international & repeated campaign) 6. Devise a data visualization tactic to illustrate patterns & findings
Questions to be answered in Analysis	I'm trying to understand how a contested social phenomenon - drug use - is framed as deviant (moral failing) or deserving (medical condition) to a wide audience, and how that **stigma changes over time**. Particularly, I am interested in how this paradoxical problem is framed in relation to need in times of economic precarity and minimal financial/institutional support. Is drug use a criminal act in need of control, or is it a medicalized condition in need of care? What is construed as deviancy versus deserving of support? How is financial need for stigmatized conditions (ranging from rehabilitation services to memorial/funeral costs) framed, and how does that vary across time and by population? ● How is success of a campaign determined by humanizing descriptions/components, race/status of the victim and who is writing the campaign, social media interaction, etc.? ● What do we use to define a successful campaign and why?
Additional Information	**Tools and Methods** For scraping - Scrapy and Selenium webdriver (for searching GoFundMe and finding relevant campaigns), and Beautiful Soup (for scraping actual campaign data). ● Potential reference: https://github.com/lmeninato/GoFundMe ○ We will use this reference as a template, but it makes more sense to build a scraper from scratch so we can make it fit our needs For cleaning and preprocessing use Pandas to organize the dataset into dataframes for faster computation. Data visualization libraries such as Matplotlib, Seaborn, and Bokeh (interactive web-integratable visualizations).

Limitations / Potential Risks	- Since this a new project, we would have to scrape data ourselves before data analysis could be done which means that data analysis would have to take place later in the semester than expected - Because Selenium takes over the user's computer, it may require a lot of time to gather data which could prevent the owner of the computer from using their device for prolonged periods of time. In this case, we may need to look into using another device or remote access - GoFundMe does not appear to have a way of filtering results by year, so it may prove difficult to get data from GoFundMe campaigns that go all the way back to 2010 (possible solution: we could potentially get all the campaigns and then sort by year after) - How will we be handling images? Even downloading all the images would be a very heavy task. Will we focus on numerical data like the number of photos on a page? - Fetching comments may be difficult as it requires Selenium, and the number of comments varies for every page (this makes it hard to write a script to gather this information). Number of comments is much easier to get (could tie into engagement as a metric for success). - Some search results that appear in later pages return "Campaign Not Found", most likely meaning they were deleted but are still showing in searches. We would need some way to remove these posts? - Campaigns which are not currently accepting donations may not have the campaign goal listed — We could use a Try/Except Block to handle this without running into major issues - Later search results may lack data fields, meaning we may have empty fields for some campaigns
Questions for Client	- How much data is sufficient to start (i.e., for each time period how much should we collect at first?)<ol type="a"> - Since we cannot filter results by year directly on GoFundMe, we will have to determine this retroactively after scraping key term by key term. - Differences between content analysis & selective discourse analysis? - Since many of the GoFundMe campaigns are not explicit when it comes to disclosing opioid addiction, how would we know that campaigns being scraped are associated with opioid/drug addiction and are relevant to the study? - Is it possible to reach out to GoFundMe to inquire about gaining access to data related to our project?