**Final Project Progress Report**

**1. Project info**

* Team Name

Runaway

* Team Members (first and last name, email)

Agens XU jingzhi@uchicago.edu

Anny Liu hanyunl@uchicago.edu

Huanrui CHEN hchen0628@uchicago.edu

Tianle YE tye1@uchicago.edu

* GitHub Repository Link

https://github.com/macs30122-winter24/final-project-runaway

**2. Project Description**

Update and expand the description you submitted for the proposal. Your updated summary should include more details about the project. At the minimum, please include:

* Research question(s) and/or description of the issue/topic you tackle with the project
* Brief description of the significance/relevance of the project from a social science perspective. You can include literature if you would like, but that’s not required! I am looking at a persuasive narrative of why we care about it as (computational) social scientists
* For the main concepts that your project studies:
  + List them and provide a working definition of them
  + Explain how you plan to measure each of them in the context of your project (e.g. how you operationalize them)
* Formulate at least one hypothesis and explain what you expect to find with the data (hypotheses can change but oftentimes it is very useful to formulate them explicitly to orient your data collection and analysis). If your final product is an interface, define what information the interface will provide and what the users will gain from it.

**3. Data Sources**

Collect a sample of data from each source before completing this part; only list sources if you are sure you can gather data from them. For each data source you plan to use, include the following (organize them as you prefer: written form, table, list, etc.):

* type of data collection (scraping or downloading)
* time frame (e.g. all year 2021, month of January 2024, etc.)
* how much data you have collected so far OR you plan to collect (size of the dataset, or expected results from scraping)
* possible data reliability/validity issues and how you plan to mitigate them
* provide additional information that could be useful to share (e.g., unit of analysis, search terms if you collect textual data, etc.)

For example:

“Data source #1. Reddit: we plan to collect all posts and comments from the subreddit named x from y to z period; so far, we have collected… this sample is the most accurate to answer your project question because… limits of this sample include…”

**4. Data cleaning/wrangling**

Describe data cleaning and/or data wrangling/merging  tasks that you plan to perform to get the data into the right shape. This can be a general description of data cleaning procedures or a detailed list for each data source (this part will greatly vary according to the project).

For example:

“We plan to merge area zip codes with Airbnb data, aggregate all sales on a specific item, perform x and y data cleaning/processing which is required for textual analysis on x data” etc.

**5. Data analysis and visualization**

Describe the data analyses and visualizations you plan to perform on your data. Make sure to include the following:

* List your dependent (what you want to explain) and your independent variables; if applicable, how you plan to construct these variables (e.g., “our x variable is at the individual level but we need it at the state level, thus we aggregate the data”, etc.)
* Explain which data analysis approach/es you plan to take (e.g., “we plan to conduct sentiment analysis using x data source, we use labeled data developed by x source which are applicable to our data for this and this reason, the training dataset consists of, etc.; we plan to conduct logistic regression using x data source” etc.); each project with a data analysis section should include descriptive statistics and visualizations of the data
* Describe what you expect to visualize/display

If your final deliverable is not a traditional data analysis, but more like a product (an interactive interface, etc.), then you won’t have a traditional data analysis, but please describe what you plan to generate by providing as many details as possible.

**6. Responsibilities**

List the responsibilities of each team member. Remember to account for the in-class presentation, short video, final report, and README. See Canvas “Final Project Overview.”

**7. Workplan (optional)**

If useful for the group, create a plan with a list of tasks and a timeline that specifies when each task should be completed between now and the in-class presentations and final submission.

This current report will serve as the basis for the final report, which will be an updated version of this report (e.g., it will include the last version/updated version of the same info you reported here + a summary of the project’s takeaways). This and the final report lengths do not matter as long as all required information is included (typically groups write a few pages).