Social Media Data Science Pipelines Project 3: Answering Questions

November 28, 2022

1 Introduction

Now that you have a live data collection system running, a strong understanding of what the data looks like, and some ideas that you can use the data to answer, it's time to execute. In this project, you will thus be using your data to answer some questions as well as adding some interactivity to basic data characterization.

2 Project Description

Simply put, you will *answer at least one of the research questions* you proposed in Project 2 as well as provide some sort of tool (I suggest making it a Web based dasbboard) that will allow for *interactive querying and results from **two** analyses from Project 2* (you may also do two totally two analyses). The this tool is not expected to be super complicated. We are also open to other tools. E.g., a command line tool that allows us to perform some analysis (with different parameters) or a Web API, etc.

3 Project Deliverables

There are three deliverables for this project.

1. Project proposal.
    - GitHub Classroom: `https://classroom.github.com/a/9_zIE7JS`
    - Due 11:59PM December 5th
2. Project implementation.

- GitHub: Classroom: `https://classroom.github.com/a/il8hkAKq`
- Due 11:59PM December 16th

3. Project report.
    - GitHub Classroom: `https://classroom.github.com/a/fCIXPKGF`
    - Due 11:59PM December 16th

This project also includes a demo over Zoom. Slots will be opened and announced during finals week (or earlier by appointment if desired). There will be many slots opened *except* on Friday the 16th. This is to avoid everyone waiting until the last minute to work on things, as well as to ensure you have *ample opportunity* to make sure your report is the best quality you can make it.

3.1 Project Proposal

The proposal for this project is much simpler than for the other projects.

1. Specify which research question you are going to answer in your report.
2. Specify which two (or more) analyses from your Project 2 you intend to include in your interactive tool.
3. Specify which tools you intend to use (e.g., plotting libraries, Web app frameworks, etc.)

3.2 Project Implementation

You will be required to submit all the the code you created. You are also free to use essentially any library or tool that you want (if you have any concerns, reach out as soon as possible). Please note that this code *must* be executable by us.

Additionally, you must give a demo of a live implementation to us via Zoom during finals week. You must also make this demo reachable by us if we are connected to the university VPN!

NB: You are *not* allowed to use any hosted 3rd party services. E.g., no Tableu or anything that isn't hosted on your VM. *If you have any questions here, please email us!*

For those of you that use Python, you might want to check out Flask (`https://github.com/pallets/flask`) which is approximately equivalent to Sinatra.

You can also check out Django (`https://www.djangoproject.com/`) which is a more full featured Web application framework in Python, or Ruby on Rails (`https://rubyonrails.org/`) which is a more full featured Web application framework in Ruby.

GRADING: The implementation for this project has a more specific grading mechanism.

- a README file in the format specified for Project 1.
  - 10 points.
- There is some kind of reachable demo that shows two analyses from Project 2. Interactivity not required.
  - 60 points.
- At least one analysis allows varying at least one parameter.
  - 10 points.
- At least two analyses allows varying at least one parameter.
  - 10 points.
- *Style points.* Do something that we think is cool, or looks nice, etc.
  - ≥ 1 points.
  - **NB:** there is no cap on style points, only a floor.

3.3 Project Report

The report for this one has no specific requirements (beyond formatting) except that you attempt to answer one or more of the research questions from your proposal. The report will be graded with an eye looking for three things:

1. How well thought out, reasoned, and justified (in the text) the analyses performed are with respect to the question being answered.
2. How well executed the analyses are.
3. Presentation. E.g., plots looking nice, proper captions and text that refers to the figures, properly cited references, etc.

Although not explicitly required, it is *heavily* suggested that the report contain some plots and tables. The more you are able to successfully communicate what you did, the better your grade will be.

With respect to questions of length: there is no minimum or maximum length on the report. Use whatever space you think you need to sufficiently answer whatever question(s) you are asking.

4 Grading

- Proposal is worth 25 points.
- Implementation is worth 50 points.
- Final report is worth 25 points.