Software Engineer Assignment

This is an assignment for you to complete as part of your application to SimPPL's Software Engineer (Web and ML Infrastructure) position. We expect a team player with experience building and scaling prototypes into software that is tested, documented, and prepared for client deployments. Part of this is to test your ability to work with existing codebases that students and junior developers have created with likely limited documentation, and inefficiencies–that still works, but just isn't the best version it could be. You are the individual that comes in and delivers a significant improvement to such a product. This is your chance to do exactly that, in an assignment we offer to you.

First, here is the assignment we offered to junior devs for internship positions:
📄 SimPPL: Incoming Applicant Assignments

We want you to take one of the applicant-submitted assignments and improve it. We offer two options to you at varying levels of difficulty explained below.

Why are we building this?

Asking interesting questions and getting the answers to them in a partially or fully automated way can unlock tremendous value for organizations and individuals interested in cross-platform analysis. You can automate the reports that are listed on this page, and this one. So read these reports and pick a question you can ask and generate insights about in a meaningful way for your end-users. We are looking to evaluate not just your programming abilities but also your thought process. And at the end of the day, regardless of whether you join SimPPL, we want you to show off this assignment as an example of what makes you an excellent candidate to build 'information integrity' tools.

Goals

- ☐ Add, refactor, or rewrite code to update one of two real intern-submitted assignments to turn it into a real-time, scalable system to query social media data starting from a proof-of-concept system / working prototype manually put together with scraped data.
- ☐ Include the same or additional data sources that are continuously streaming data and/or can search a platform to collect data on an automated basis (you are free to modify the choice of platform that data is being collected from).
- ☐ Demonstrate how the plots you create from the data can tell a meaningful story to a user of the system.
- ☐ Explain what is the question that a user might ask, how the system takes it as an input and share how plots answer that question (more plots/data don't necessarily mean better insights).
- ☐ Share with us a PDF report. This will focus on a small case study you created as a 1-page or 2-page report as an example of how your system can be used to generate the above linked types of reports.

Remember, we are **not** looking for the most number of features you can build or maximum no. of platforms you can club datasets from, into your system. This assignment is intended for you to share *how* you can think through a small-but-important problem and create a user-friendly system to operationalize your process. What are some important events worth tracking and developing a case study about? Which platforms might be good to study in case you need to switch away from Bitchute and Instagram? You are free to decide and accordingly build a solution.

Difficulty Level

Here are two options to pick as your starter assignment.

1. Easier: https://github.com/ShlokMandloi/bitchute_analytics_dashboard
    a. Shlok has built a really neat system to collect and query data from a platform called Bitchute. The UI is functional and plots look good but the goal of such a system should be to guide a user to ask a meaningful research question and get an answer to it. Show us how you evaluated what question(s) you wanted the system to answer and how you built on top of it to create a user-friendly tool.
2. Harder: https://github.com/rachit901109/simppl_task
    a. Rachit has built out an instagram data analysis pipeline that relies on a somewhat manual data collection using headless browsers / scraping. We want you to identify if you could partially or fully automate this. If you want to go ahead and switch this platform (Instagram) out for another where you are confident of getting continuous streaming data + the ability to search it for more data, you are free to do so. Look up the list of suggested platforms in the intern assignment.

You will be evaluated as per the quality of your submissions including the repository you share (it can be public) and based on the report you generate (we would prefer this be private until we complete the evaluation, after which you are free to release it as you wish). We really value people who can solve hard problems, but we also want to know if you are able to assess your own skills and deliver a meaningful report in a reasonable timeline. You are free to pick your choice of starter assignment.

Similar Systems
This is not a novel task and there are other systems out there that use smart and meaningful visualizations. Here is one example:
https://css.seas.upenn.edu/radicalization-at-a-glance-penn-media-accountability-project-launches-interactive-data-dashboard

Others like https://openmeasures.io/research or https://graphika.com/reports may offer additional insights into what kinds of questions make sense to ask. We want to see how you think, not just whether you can code.

Pay Structure

The proposed pay for this position ranges from USD 9,000 - USD 15,000 annually subject to experience and performance in the role. We offer a probationary period of 1 month in order to evaluate your performance, initiative, and ability to fit into the team before you join us on an annual contract. The contract is possible to extend subject to mutual agreement following consistently positive quarterly performance evaluations.There may be perks including international travel subject to performance and opportunities available during the contract.