Introduction:

My name is Sha Zhao. I am currently a master's student in computer science at ASU and will be graduating in Fall 2024. I have a background in Environmental Engineering with both a bachelor's and master's degree. I am enthusiastic about merging my expertise in environmental studies and computer science to contribute towards addressing climate issues. My passion lies in working with environmental data, aiming to make a meaningful impact on improving our world.

As a computer science student, I had two internship experiences . In one of my internships, I have completed an internship focused on AI/data analytics at Office of Palo Alto City, where I utilized Python for data processing and generated email responses with a provided prompt by using Open AI and Llama Index. The efficiency of email responses has been significantly improved by 70%. And in the other, I worked as a full-stack developer, using react for frontend and node.js for backend to build a cooking website, where users can search recipe based on the ingredients they have.

With the knowledge gained from my M.S. program in Computer Science, and experience with SQL, NumPy, Pandas, Sklearn, MatPlotLib, Seaborn, as well as projects involving data mining and data visualizations, I am confident that I am a good fit for the internship.

Relevant projects

1, Data analysis - Improving College Educational Enrollment:

UVW College plans to improve their enrollment by developing a new marketing campaign based on the profiles of individules. To achieve this objective, I created various visualizations and analyzed the effect of different factors on income, including age, education, marriage-status, capital loss, hours-per-week, etc. I Utilized Pandas, Python, NumPy, Tableau, Power BI, Seaborn, Matplotlib to accomplish the development of marketing profiles for UVW College.

Github link: UVW College Data Analysis

2, Data analysis - Analyzing voting behavior and improving voter turnout

I identified patterns and correlations that can provide insights into the voting behavior of the public in the upcoming election to encourage voter turnout. I created various visualizations and analyzed the effect of different factors on age, income, degree, race, etc. and utilized Pandas, Python, NumPy, Seaborn, Matplotlib to finish it.

Colab link: Voting Data Analysis

What I have learned from the projects:

* Understanding the Data: I learned that the foundation of effective data visualization is a thorough understanding of the dataset. Given the huge amount of data, I spent a lot time to understand each factor.

* Creating Meaningful Visualizations: I learned how to transform data into Visualizations that tell a story. This involved selecting the right type of visualization for the data at hand, employing color, size, and layout effectively, and focusing on clarity and simplicity to communicate complex information easily.

*Critical Thinking and Analytical Skills: Through these projects, I also honed my critical thinking and analytical abilities. I learned to question assumptions, interpret patterns, and draw insights from visual data representations. This has been invaluable in making data-driven decisions and recommendations.

*Collaboration and Feedback: Engaging with colleagues for feedback on my visualizations has taught me the importance of collaboration in the data visualization process.

* Document: Recording everything is also important, it can help me and other people understand my code and my report.

The understanding of the project

I spend a lot of time to understand the Land Sector datasets and learn FLINTcloud templates . I have a basic understanding now.

1, Land Sector datasets

Land Sector datasets provide comprehensive information about land use, land cover, and various environmental factors that impact the land surface. It has Land Cover and Land Use, Soil

Properties, Climate data, and Bioclimatic Ecological zones. It provides boundaries and environmental factors by county and even by state.

2, FLINT.Cloud

FLINT.Cloud is like a special online space where we use the FLINT program. FLINT helps us keep track of forests and land. With FLINT.Cloud, we can use FLINT on the internet, which makes it easier and faster to see what's happening with the land and forests all over the world. The most amazing feature of FLINT is Real-time Monitoring. FLINT.Cloud can quickly get the latest information about the land and forests. This means we can see changes in the land, like new tree growth or areas that have been cut down, almost as soon as they happen.

3, GCBM

The Generalized Comprehensive Biogeochemical Model (GCBM) is like a computer program that helps us understand how forests grow and how they store carbon from the air. Imagine forests as big, green machines that can take in carbon dioxide and store it in trees and soil. The GCBM helps us see how much carbon these forests can hold and how things like cutting down trees or forest fires can change that.

4, The project

It is about making a special tool that uses big sets of land data to find small patches of forests all over the world. This tool will work with FLINT.Cloud to set up a GCBM for these forests. This system helps us simulate how forests grow and how they capture carbon from the air.

To do this, we will first study the land data closely to understand what it tells us about forests. We will look for specific details in the data that are important for keeping an eye on forests. Then, we will create a step-by-step process that uses this data to spot forest patches. We will use Python to build this process.

Our tool will be made to set up the GCBM system quickly and start simulations using FLINT.Cloud. To make sure everything works smoothly and keeps getting better, we will use CI/CD. This method helps us automatically update our results in a MongoDB database.

Finally we will have a working FLINT Forest Monitoring tool. This tool will be able to monitor forests and study carbon in an automated and flexible way. This is important for helping take care of forests and fight climate change.

Timeline

Week 1: Project Planning and Setup
Understand the project scope and objectives and set up development environment and tools.

Week 2-3: Study Land Sector Datasets
Understand the Land Sector datasets. Think a data workflow for identifying forest squares.

Week 4-5: Develop Data Workflow
Implement data preprocessing. Test the workflow on sample datasets to ensure accuracy.

Week 6-7: Develop Python-based Tool
Design the tool architecture for producing GCBM configurations.
Implement the tool using Python, integrating FLINT.Cloud templates.
Test the tool with sample data to ensure correct configuration generation.

Week 8-9: Dispatch Simulations
Integrate the tool with FLINT.Cloud to dispatch simulations for the identified forest squares.
Test the simulation results. And optimize the tool for efficient simulation dispatching.

Week 10: Integrate CI/CD Workflow
Set up a CI/CD pipeline for the project. Automate the process of pushing generated results to a remote MongoDB database.

Week 11: Documentation
Document the entire process, including data workflow, tool usage, and simulation dispatching.

Week 12: Testing and Refinement
Getting feedback from team members and mentors. Fix any issues.
Refine the tool and workflow based on test results and feedback.

Week 13: Finalization and Release
Finalize the documentation with any last-minute updates.
Prepare a template for users looking to try the tool.