Final Project

FINAL PROJECT: CONTEXT

A larger project compared to datathon
Designed for the trainees to apply and showcase their data science skills learned through this program
Unlike Datathon, teams should find their problems and datasets.
We will give you a set of links to find interesting datasets; you may find datasets from other sources as well.
The details are on the course website.

FINAL PROJECT: DELIVERABLES

Deliverables: A production level application, involving a dashboard (using the DASH package in python), and a report.

Interactive Front-End: A non-technical user should be able to use this to get meaningful outputs and visualizations. The target audience is a government official or someone in the private sector - it should be clear to them how to get value out of this and what that value is.

AWS-hosted Database: This should persist all relevant data, and supply it to the front-end.

AWS-hosted Data Analysis & Computation: These should perform all computation on the data itself that is relevant to the application. They can be structured in a manner of the team's choice - microservices, scripts, etc. However, they must live in and run off of the AWS compute engine (NOT a local machine)

SUPPORT AND JOURNALING

As with the Datathon project, **your team must clearly indicate what the percentage split in effort was among the team members.**

As with the Datathon project, team are expected to keep a Google Doc journal of their progress on the final project, and write 1 - 2 paragraphs at the end of each day in this journal (a single journal document should contain both the Datathon project and Final project notes)

SCOPE OF THE PROJECT

Just as for Datathon, Scoping your project is key to success. It the involves the following:

Having a clear mental map of what you want to achieve

Clearly estimating of how long each step will take you to accomplish

Having a time table so you can deliver this on time

Important: You should get approval from your respective TA about the scope of your project and the end of week 3 before going deeper into the analysis.

SCOPING FOR THE FINAL PROJECT

Unlike the Datathon project, this scoping document will be more involved, and should include:

- Your team's problem you set out to solve (multiple versions, starting from V1 (minimal viable products), and work your way up to more ambitious versions)
- High-level overview of how your application plans on solving this problem
- Outline of who the primary users would be and how they would interact with the app
- Descriptions of app features
- The datasets your team is planning on using
- How you plan on using each of these datasets/how they contribute to your app

PROJECT PRESENTATIONS

The last week of classes has no instruction. All of the time will be dedicated to finishing up the projects.

The last day will involve the final project presentations.

Each team will have a booth where they will set up a laptop which contains the interactive dashboard.

We will have dignitaries visiting from both the private and public sector, and you will walk them through your project.

MILESTONES

WEEK 3

Multiple problem versions chosen

All versions should naturally build on top of each other

V1 should be pretty easy and reasonable

The last version can be moonshot - the idea is that you must implement for V1 first, then V2, etc. so as to guarantee some sort of finished build by the end of Week 10. If you can get to V2, V3, etc. by the end, great, if not, at the very least you have something done that you can present)

Datasets sourced

Project scoping plan/proposal written

Introduction section of final report written

MILESTONES (CONTINUED)

WEEK 5

Front-end design finalized

Information channel to and from AWS components designed

WEEK 6

Datasets wrangled & cleaned

WEEK 7

Database tables designed

Finalized datasets loaded into AWS-hosted database

MILESTONES (CONTINUED)

WEEK 8

Front-end build finalized

Link from front-end to AWS-hosted database established

WEEK 9

Analysis & modeling of datasets finished

Relevant code from this embedded into AWS compute engine

AWS compute engine integrated with database and front-end

WEEK 10

App final touches and quality assurance checks

Report finished & Presentation prepared

CERTIFICATION REQUIREMENTS

For receiving a certificate, your team must have achieved all of the milestones.

At the end of the program, we will conduct a survey whereby we ask each student to rate how each of their teammates contributed to the Datathon and final project; this information will be used in determining certification.

Timing for Final Project vs. Datathon

Students should be working on the final projects and the datathon project during the class time.

You are expected to work 80% of your time on the final project and 20% of your time on the Datathon.