Asana Project Report - Job Site Safety

Group Members: Jason Raimondi, Jeremy Cryer, Maimuna Bashir
Asana Board Link: https://app.asana.com/0/1206390072539638/1206390072539638

- **What is the current health of your project? (1-2 paragraphs)**

  The current health of our project is good. We have made steady progress and have completed several key tasks such as creating the team Asana board, drafting the design document, and setting up the data lake and feature store notebooks. We are on track with our timeline and are actively working towards the successful completion of the project.

- **What have you completed so far? (bullet points)**

  - Create a team Asana board.
  - Draft the design document.
  - Create a GitHub repository.
  - Draft the Asana project report.
  - Complete sections of the design document.
  - Set up one notebook for the data lake
  - Set up one notebook for the feature store.

- **Is your project ahead of schedule or behind schedule? What is working or not working? (1-2 paragraphs)**

  Our project is currently on schedule. We have been able to complete tasks as planned and have not encountered any major issues or delays. The team's collaboration and communication have been effective, allowing us to work efficiently and stay on track. We will continue to monitor our progress and make adjustments if necessary.

- **Is the overall completion of tasks on track? (1 paragraph)**

  Yes, the overall completion of tasks is on track. With several tasks already completed and active work on ongoing items, we are confident that we will meet our project goals within the given timeline. We will continue to prioritize tasks and allocate resources

accordingly to ensure timely completion.

- **Have any tasks taken longer than expected? (1 paragraph)**

  No, so far, all tasks have been completed within the expected timeframe. The team has been proactive in managing tasks and addressing any potential challenges that arise. Regular communication and planning meetings have helped us identify and address any potential delays promptly.

- **How have you decided what tasks to put on the board? Have you been leveraging planning meetings? (1 paragraph)**

  Tasks are decided through collaborative discussions and planning meetings. During these meetings, team members share their insights and expertise to break down the project requirements into smaller, actionable tasks. We assign deadlines and prioritize tasks based on their dependencies and overall project timeline.

- **How have you been collaborating as a team? (1 paragraph)**

  As a team, we have been utilizing various collaboration tools to work together effectively. We use Asana to track and manage our tasks, ensuring transparency and accountability. Slack is our primary communication channel, facilitating quick and efficient discussions. Additionally, we leverage Google Sheets for document design and hold regular meetings on Zoom to discuss project progress, address any challenges, and make decisions collectively.

Job-Site-Safety---AAI-540

Printed from Asana

To Do

Module 7 - Individual Deliverable

Feb 26

Low On track

Module 7 - Individual Deliverable

Ma Feb 26

Low On track

Module 7 - Individual Deliverable

Je Feb 26

Low On track

Module 7 Deliverable - Codebase GitHub Repository

Feb 26

Low On track

Module 7 Deliverable - ML System Operation Validation

Feb 26

Low On track

Module 7 Deliverable - ML Design Document

Feb 26

Low On track

Module 6 Recommended Tasks

Feb 19

Low On track

Module 5 Recommended Tasks

Feb 12

Low On track

In Progress

Module 4 Deliverable - GitHub Repo Link(s)

Monday

Medium On track

Review

Done

- Module 4 Deliverable - Partial Design Document
   Monday
  Medium On track
- Module 4 Deliverable - Asana Project Report
   Monday
  Medium On track
- Module 2 - Team Asana Board
   Jan 22
  Medium On track
- Module 2 - GitHub Repository
  Je Jan 22
  Medium On track
- Module 2 - Draft Design Document
  Je Jan 22
  Medium On track

Job-Site-Safety---AAI-540

Printed from Asana

To Do

- [ ] **Jason E. Raimondi:** Module 5 Recommended Tasks due Feb 12, 2024

  Priority: Low
  Status: On track

  Topic: Monitoring

  - [ ] Implement model monitors on your ML system due Feb 10, 2024
  - [ ] Implement data monitors on your ML system due Feb 10, 2024
  - [ ] Implement infrastructure monitors on your ML system due Feb 10, 2024
  - [ ] Create a monitoring dashboard for your ML endpoint/job on Cloudwatch due Feb 10, 2024
  - [ ] Generate model and data reports on SageMaker due Feb 10, 2024

- [ ] **Jason E. Raimondi:** Module 6 Recommended Tasks due Feb 19, 2024

  Priority: Low
  Status: On track

  Topic: CI/CD

  - [ ] Implement CI/CD Pipeline to automate training, evaluation, and deployment due Feb 17, 2024

    - CI/CD pipeline should have checkpoints to evaluate model performance, model code and system integration.
    - Your CI/CD pipeline should train with training data, and evaluate with testing data.
    - You should try to improve your initial model and run it through your CI/CD pipeline.

- [ ] **Jason E. Raimondi:** Module 7 - Individual Deliverable due Feb 26, 2024

  Priority: Low
  Status: On track

  Presentation Video – Interview Responses

- [ ] **Maimuna:** Module 7 - Individual Deliverable due Feb 26, 2024

  Priority: Low
  Status: On track

  Presentation Video – Interview Responses

- [ ] **Jeremy:** Module 7 - Individual Deliverable due Feb 26, 2024

  Priority: Low
  Status: On track

  Presentation Video – Interview Responses

- [ ] **Jason E. Raimondi:** Module 7 Deliverable - Codebase GitHub Repository due Feb 26, 2024

  Priority: Low
  Status: On track

  Team Code - GitHub repository with the complete codebase for your ML system

  #3 Code

  Codebase GitHub Repository

The final presentation and submission of your codebase should support your ML Design Document and reflect teamwork.

1. Include a link to your team's GitHub repository in your design document.
2. Your GitHub repository should reflect the following:
    - Method
        - All of your code should be stored in GitHub in a clean and professional manner. Notebooks should be stored in .ipynb format.
        - Your code should be clean, have useful comments, and only include code that builds towards the project goal.
        - Your data should be stored in S3 and documented in your GitHub repository
        - Any graphics, such as charts/graphs that help explain your data, should be included in your .ipynb files.
    - ML Design
        - Codebase should be comprehensive and complete as an ML system codebase.
        - Codebase and design document should be mutually reinforcing, reflecting the parallel effort and scope of the ML system.
    - Teamwork
        - All team members should contribute to the GitHub repository.
        - Commit history will be available to the instructor for review.

☐ **Jason E. Raimondi:** Module 7 Deliverable - ML System Operation Validation due Feb 26, 2024

Priority: Low
Status: On track

Team Video - Screencast demonstration "ML System Operation Validation"

#2 Demonstration

ML System Operation Validation

1. Implement the project using AWS SageMaker.
2. Validate ML system operation.
3. Create either a video or document that demonstrates the following:
    - feature store and feature groups
    - infrastructure monitoring dashboards for your system
    - any model or data monitoring reports
    - CI/CD DAG in a successful and failed state
    - model registry
    - outputs of batch inference job or endpoint invocation

Format requirements:

- Video screencast with narration of required items. Record/screencast your screen and provide voice narration of the above content requirements. Consider using a recording software, such as Screencast-O-Matic or Zoom. Ensure that the sound quality of your video is good and each member presents an equal portion of the presentation. Export the video file to an **mp4 format**. View the Recording Video Presentation and Submission Guidelines in the **Course Resources.**
- Document with screenshots or project exports with appropriate descriptors. Screenshot demonstrations of the required items and assemble in a document; and/or export required PDF/HTML with code, comments,

and results within the notebook and include in the document. Export the final document as a PDF for submission.

☐ **Jason E. Raimondi:** Module 7 Deliverable - ML Design Document due Feb 26, 2024

Priority: Low
Status: On track

Team Document - Written report "ML System Design Document"

#1 Design Document

ML System Design

1. Finalize the ML System Design Document that you have drafted throughout the course for final submission to project stakeholders.
2. In your final document, make sure you include the following:
    - Include a clearly defined problem statement. This should be one to two paragraphs.
    - Include a clear description of how you will measure the impact of your work; this should tie directly to the goals. This should be one to two paragraphs.
    - Complete the security checklist and describe any risks surrounding sensitive data, bias, and ethical concerns.
    - Solution Overview. Describe your implementation of the solution. Each section should be two to three paragraphs and should describe your findings from the code in your GitHub repository. Your answers should be detailed and explain your rationale for the decisions you have made.
        - Data Sources
        - Data Engineering
        - Feature Engineering
        - Model Training & Evaluation
        - Model Deployment
        - Model Monitoring
        - CI/CD

The ML Design Document is for a technical audience and must be written in a clear, organized fashion.

In Progress

☐ **Jason E. Raimondi:** Module 4 Deliverable - GitHub Repo Link(s) due Feb 5, 2024

Priority: Medium
Status: On track

Deliverable #3: Share GitHub Repo Link

- 1 Notebook for setting up data lake
- 1 Notebook for setting up feature store

☑ ~~Collect a raw data set and store in an S3 Datalake~~ due Jan 29, 2024

☑ ~~Set up Athena tables to enable cataloging and querying of your data~~ due Jan 29, 2024

☑ ~~Perform exploratory data analysis on your data in a SageMaker notebook~~ due Jan 29, 2024

☑ ~~Initialize a feature store~~ due Jan 29, 2024

☑ ~~Design the feature groups needed for your system~~ due Jan 29, 2024

- [x] ~~Perform feature engineering on raw data and store features in feature groups~~ due Jan 29, 2024
- [x] ~~Split your feature data into training (~40%), test (~10%) validation (~10%) datasets~~ due Jan 29, 2024
- [x] ~~Reserve some data for "production data" (~40%)~~ due Jan 29, 2024
- [ ] Set up a benchmark model in SageMaker due Feb 3, 2024
  - The benchmark model should be simple, it can be a simple heuristic or a model with just a couple of features.
  - The idea here is to have a baseline for further improvements, and to get a minimum viable. product out before you start improving your system.
- [ ] Build, train and debug your ML model in a SageMaker model due Feb 3, 2024
  - This is your first real iteration on your model, it doesn't need to be perfect, you may want to revisit model development again once we implement CI/CD in module 6
- [ ] Evaluate your model and compare against your simple benchmark model due Feb 3, 2024
- [ ] Deploy your model to SageMaker (Batch process or Real Time Endpoint) due Feb 3, 2024
- [ ] Link your initial findings and codebase in your ML Design Document due Feb 3, 2024

Review

Done

- [x] **Jason E. Raimondi:** ~~Module 2 - Team Asana Board~~ due Jan 22, 2024

  Priority: Medium
  Status: On track

  Team to set up / collaborate on:

  - Asana View
  - Team Workflow(s)
  - Project Set Up

  Remaining:
  Add all module tasks

- [x] **Jeremy:** ~~Module 2 - GitHub Repository~~ due Jan 22, 2024

  Priority: Medium
  Status: On track

  Set up GitHub repository
  - Initialize Repo
  - Set up for sharing/permissions
  - Make sure teammates join for collaboration

- [x] **Jeremy:** ~~Module 2 - Draft Design Document~~ due Jan 22, 2024

  Priority: Medium
  Status: On track

  Complete the following sections and submit by due date.

  - Authors: Students must be in groups of two or three

- Business Name
- Publication Date
- GitHub Project Link:
- Asana Board Link:
- Project Background:
- Technical Background:
- Goals vs. Non-Goals:
- Data Sources:

Remaining:
Jason to submit - Completed 1/20
Share out .docx and .pdf files afterwards - Completed 1/20

1/21 (JER):
Remaining items were completed.
Moving task to "Done".

☑ **Jason E. Raimondi:** ~~Module 4 Deliverable - Partial Design Document~~ due Feb 5, 2024

Priority: Medium
Status: On track

Deliverable #1: Partial ML System Design Document
You and your team will use and build upon the ML Design Document that you started in Module 2and complete the following sections of their ML System Design Document:

- Data Sources:
- Data Engineering:
- Training Data:
- Feature Engineering:
- Model Training and Evaluation:
- Model Deployment:

☑ **Jason E. Raimondi:** ~~Module 4 Deliverable - Asana Project Report~~ due Feb 5, 2024

Priority: Medium
Status: On track

Deliverable #2: Asana Project Report

Export your Asana Project Report and respond to the following questions:

- What is the current health of your project? (1-2 paragraphs)
- What have you completed so far? (bullet points)
- Is your project ahead of schedule or behind schedule? What is working or not working? (1-2 paragraphs)
- Is the overall completion of tasks on track? (1 paragraph)
- Have any tasks taken longer than expected? (1 paragraph)
- How have you decided what tasks to put on the board? Have you been leveraging planning meetings? (1 paragraph)
- How have you been collaborating as a team? (1 paragraph)