Data Science Project Structure (100% total)

1. Critical Task: Data Analysis and Dashboard Creation (60% of total grade)

a) Data Collection and Preprocessing (12%)

- Gather data from relevant sources - more than 10k rows, minimum 5k.
- Clean and preprocess the data.
- Handle missing values and outliers.
- Perform feature engineering if necessary.

b) Exploratory Data Analysis (EDA) (15%)

- Conduct descriptive statistics.
- Visualize data distributions and relationships.
- Identify patterns and trends.
- Formulate hypotheses based on observations.

c) Statistical Analysis (10%)

- Perform relevant statistical tests.
- Validate or reject hypotheses.
- Interpret results in the context of the problem.

d) Insights and Recommendations (8%)

- Draw meaningful conclusions from the analysis.
- Provide actionable recommendations based on findings.
- Discuss limitations and potential biases in the analysis.

e) Streamlit Dashboard Creation (15%)

- Develop an interactive dashboard using Streamlit.
- Visualize key findings from your analysis.
- Implement interactive elements (e.g., filters, dropdowns).
- Present insights clearly and effectively.
- Ensure the dashboard is user-friendly and visually appealing.

2. Extra Tasks: Machine Learning and Advanced Techniques (40% of total grade)

a) Machine Learning Model Development (20%)

- Select appropriate algorithm(s) for the problem.
- Split data into training and testing sets.
- Train and validate the model.

b) Model Evaluation and Interpretation (10%)

- Assess model performance using relevant metrics.
- Interpret model results and feature importance.
- Discuss model limitations and potential improvements.

c) Advanced Techniques (choose at least one) (10%)

- Implement deep learning techniques.
- Develop a recommendation system.
- Perform natural language processing.
- Apply time series analysis or forecasting.
- Develop a web scraping and data collection pipeline.
- Tune hyperparameters for optimal performance.
- Use `make_pipeline` to streamline preprocessing and modeling steps.

Project Deliverables

1. Jupyter Notebook or Python scripts containing all code and analysis.
2. Streamlit dashboard application (Python script and any necessary assets).
3. Written report (PDF) explaining the project, methodology, findings, and conclusions.
4. Presentation slides summarizing the project (5-10 slides).

Evaluation Criteria

- Clarity and organization of code and analysis.
- Depth and appropriateness of analytical techniques used.
- Quality of insights and recommendations.
- Creativity in approach and problem-solving.
- Functionality, design, and user experience of the Streamlit dashboard.
- Presentation quality and ability to communicate results effectively.

Timeline

- Project proposal presentation: 5-7th August 2024.
- Project proposal due: 9th August 2024 - Submission.
- 1st Midpoint check-in review: 19th - 23rd August 2024.
- 2nd Midpoint check-in review: 2nd - 6th September 2024.
- Final project submission: 16th September 2024 (11:59 pm).
- Project presentations: 17th September 2024.

Streamlit Dashboard Requirements

- Interactivity: Include at least two interactive elements (e.g., dropdowns, sliders, buttons).
- Visualizations: Minimum of three different types of charts or graphs.
- Data Insights: Clearly display key findings and insights from your analysis.
- User Experience: Ensure the dashboard is intuitive and easy to navigate.
- Code Quality: Well-organized, commented code for the dashboard.
- Integration: Seamlessly integrate the dashboard with your data analysis workflow.