STATS 507 – Data Science Analytics using Python
Project Proposal Guideline (10 pts)

In the project proposal, you'll pick a project idea to work on early.

General Guideline
- Your project must integrate at least one Hugging Face **model** or one open **dataset**.
- Each project should be done individually and must be implemented in Python.
- You should scope your project to be about what you could complete within the time constraints.
- The final project should be submitted as a GitHub repository containing a well-documented executable codebase along with a (at least) 2-page summary report (summary.md or pdf) that concisely describes your problem, approach, and results. You can have preliminary results in your proposal but choose a problem that is non-trivial and not a Ph.D. thesis either.
- You are expected to spend at least 40 hours on your project.

Potential Paths
- Dataset-Centric Projects: choose an open dataset that we didn't cover in class for meaningful analysis. https://huggingface.co/docs/datasets/en/index
- Model-Centric Projects: choose a model/models from Hugging face hub to build innovative applications. https://huggingface.co/models
- Many more!

Proposal Requirements
- Format: at least 2 pages, at least 1 figure.
- Sections:
  Overview
  - Describe the background and motivation of the project.
  - Why would you want to do this project?
  - What kind of data/model do you want to work on?
  - What kind of insights are you expecting from this projects?

  Prior Work
  - Literature review: conduct a comprehensive review of existing literature, focusing on recent advancements and state-of-the-art.
  - Describe potential methods that can be used to achieve the project goal.

  Preliminary Results
  - Data understanding (dataset statistics, data quality assessment, initial insights, potential challenge)
  - Basic model (resource requirements, performance bottlenecks…)
  - What tools from class did you use, what tools will you explore?

Project Deliverables
- What will a successful project produce?
- What are your sub-goals?

Timeline
- Week 1-2
- Week 3-4
- Week 5