**3.1 Project Plan: Due Week 10 (10 Pts)**

Deliver an approximately 2 page **plan**, in HTML or PDF format, addressing the following elements in order:

1. Project Title (?)
2. Group Members and their course – Thaina Gomez & Maxwell Miller-Golub
3. Topic - 1 sentence (Predicting Diabetes) <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset>
4. Question*s* of Interest: At least one for regression **and** one for classification.
   1. Regression:
      * Planned Approach: This includes framing the problem, the data, the methods, and the Intended Outcome.
      * Literature Review: Identify at least one article or publication *per group member* relevant to the topic. It does not have to be a scholarly article but it shall provide context for the topic and questions of interest to help in framing the analysis.
      * Data Assessment: this includes
        1. Identification of the data set and its source.
        2. Assessment of the data set and characteristics, e.g., number of observations and expected number and type of variables.
      * Planned Methods: this includes
        1. Identification of the expected methods for regression
        2. Identify planned approaches for assessing multicollinearity and model comparison as appropriate to the course. Any group with members from 627 must satisfy the 627-unique learning outcomes.
        3. Identification of potential ethical concerns.
        4. Risk Assessment and mitigation.
   2. Classification:
      * Planned Approach: This includes framing the problem, the data, the methods, and the Intended Outcome.
      * Literature Review: Identify at least one article or publication *per group member* relevant to the topic. It does not have to be a scholarly article but it shall provide context for the topic and questions of interest to help in framing the analysis.
      * Data Assessment: this includes
        1. Identification of the data set and its source.
        2. Assessment of the data set and characteristics, e.g., number of observations and expected number and type of variables.
      * Planned Methods: this includes
        1. Identification of the expected methods for both regression and classification.
        2. Identify planned approaches for assessing multicollinearity and model comparison as appropriate to the course. Any group with members from 627 must satisfy the 627-unique learning outcomes.
        3. Identification of potential ethical concerns.
        4. Risk Assessment and mitigation.
5. Deliverable: Option A (Poster and Presentation)
6. Schedule and Hours: A table for weeks 10-14 with the main activities, estimated hours, and expected outcomes for the week.
7. Group Member Responsibilities
8. Summary