**Project 4 Proposal | Group 5 | 5.2.2024**

***GROUP PARTICIPANTS***

* Amanda Dunaway
* Charity Ovando
* Sergio Villa

***BACKGROUND & OBJECTIVES***

The primary objective is to utilize machine learning models on a dataset of patient information to predict whether a person has heart disease.

***APPROACH***

* Data:
  + UC Irvine Machine Learning Repository Dataset:
    - https://archive.ics.uci.edu/dataset/45/heart+disease
    - We will import the dataset into Python using the ucimlrepo package.
    - Next, we will perform any data cleaning and manipulation with Pandas.
  + The target variable will be a binary variable for whether an individual has heart disease.
  + The feature variables will consist of individual data for each patient, including integers such as age and cholesterol value, and categorical data, such as sex and meeting fasting blood sugar threshold.
* Model:
  + To start, we will use the Scikit Learn library to split our data into training and testing data.
  + For the first model, we will apply logistic regression.
  + For further optimization (if needed), we will implement some combination of the following.
    - Logistic regression with altered features.
    - Neural network model
    - Random forest model
* Visualization
  + We will develop a Tableau story to illustrate variables and results of interest.

***DELIVERABLES***

* 15-20 minute PowerPoint presentation outlining process and findings
* Git repository with all code and analysis