# Week 2.2 Project 1: Project Check- In/ Milestone 2

Aritzi Piedras Silva
Catherine Williams

- Any surprises from your domain from these data?
  After conducting some EDA – I was bit puzzled simply because I had to look up what each column abbreviation meant as well as understanding the meaning behind the values recorded such as:
  Age
  Sex – Value 1: Male
  Value 0 : Female
  CP – Chest pain
  Value1: Typical Chest Pain
  Value 2: Atypical Chest Pain
  Value 3: Non- Chest Pain
  Value 4: Asymptomatic
  TRESTBPS – Resting Blood pressure
  Chol – cholesterol levels in mg/dl
  FBS- Fasting Blood Sugar
  Value 1 = > 120 mg/dl = TRUE
  Value 0 = FALSE
  RESTECG – resting electrocardiographic results
  Value 0: Showing Probable or hypertrophy
  Value 1: normal
  Value 2: ST-T wave abnormality
  THALACH – max heart rate achieved
  EXANG- exercise induced angina/chest pain
  Value 1 = Yes
  Value 0 = No
  OldPeak – ST depression induced by exercise relative to rest

  Slope – slope of the peak exercise ST segment
  Value 1: Upsloping
  Value 2: Flat
  Value 3: Downsloping
  CA – Number of major Vessels (0-3)
  THAL – A blood disorder called thalassemia
  Value 3 = Normal
  Value 6 = Fixed Defect
  Value 7 = Reversable Defect
  Target – Hear Disease
  Value 0 = No
  Value 1 = Yes

  This was not a HUGE surprise, but it was enough to keep my busy and research how these attributes may influence heart disease prevention and awareness.

- The dataset is what you thought it was?

This dataset is something that I feel comfortable playing with – all data is Int64 except for 'oldpeak' which is a float64. Additionally, there were no missing values in the set, so I do not have to worry

about dropping certain data points. (Yet at least). I do plan on renaming my columns to make things easier for me, I also ran Pandas Profiling to make EDA a bit simpler – its my second time using pandas profiling so seeing it execute so fast with visualizations was beyond impressive.

- Have you had to adjust your approach or research questions?

I actually want more questions answered now that I have a general idea of what lives in the dataset as well as heart disease. I want to figure out which gender is most likely to have heart disease and what are the common traits for this gender to develop heart disease. I wanted to figure out which attributes habits can leader to an unhealthy lifestyle, but the dataset does not contain diets or factors that may contribute to high blood sugar, cholesterol, and heart rates.

- Is your method working?

After splitting my dataset into x and y variables for modeling – I kept on grabbing errors – I think these errors are reflected more on my expertise with dealing coding and machine learning. I'm going to do more research on medium and stack overflow and see how I can get the 3 algorithms to run without issues to compare their performance.

- What challenges are you having?

The main challenge I always face is really my lack of experience, I'm hoping that by project number two that majority of my errors as far as coding such as commas, wrong variable names, spaces, etc. will decrease as I get into the rhythm and flow of things.