Caitlin Ruble
Capstone 3 Project Ideas

1. Create a submission for this Kaggle competition, Image Classification of Stoke Blood Clot Origin: https://www.kaggle.com/competitions/mayo-clinic-strip-ai/overview/description

Description: The goal of this competition is to classify the blood clot origins in ischemic stroke. Using whole slide digital pathology images, you'll build a model that differentiates between the two major acute ischemic stroke (AIS) etiology subtypes: cardiac and large artery atherosclerosis.

Note: the data provided takes up 395.36 GB of space! I don't have anywhere near that kind of space on my machine. How can I work with the data without downloading it locally? I've seen stuff about loading it into CoLab, which is all well-and-good, but can I use that much space in CoLab? Or maybe simply running in through Kaggle's integrated notebooks?

2. Create a Skin Cancer Image Classifier:
    a. ultimate goal would be to launch a site where a user could submit an image of a mole/mark they're concerned about and get a percentage likelihood that it's cancerous (or abnormal)
    b. Looking for data….
        i. Compilation of Data Sources: https://www.thelancet.com/journals/landig/article/PIIS2589-7500%2821%2900252-1/fulltext
        ii. https://challenge.isic-archive.com/data/#2020
        iii. https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000
    c. reference paper for some work done at Stanford: https://cs.stanford.edu/people/esteva/nature/
    d. lots of code & project examples online from folks who have approached this challenge

3. A fish-keeping image classifier: a program that could take an image of a fish (or plant?) taken at a fish-store and give feedback on what the species is, what the needs of the species are, and potentially even whether it's compatible with their current tank setup at home.
    a. web-scrape fish data
    b. find labeled images via web search (?)
    c. maybe starting with "freshwater tropical" since that's what I'm most familiar with
    d. could start by scraping this Wikipedia page to get a framework/list of fish to build out the dataset and model on: https://en.wikipedia.org/wiki/List_of_freshwater_aquarium_fish_species