UCLA Extension Data Science Intensive

Instructor: William Yu

Capstone Project

- The goal of the capstone project is that you will use the skills and models you learned in this course and apply them onto a real-world dataset. Do the data analytics and tell a coherent and good story from your results.
- In the project, it will be the best that you can demonstrate your **problem-solving** skills or strategies. Problem-solving could be in the following fields:
  - Finding and collecting a good-quality, meaningful, relevant, or interesting dataset.
  - Your dataset might be good but contain a lot of missing values, noises, etc. You can demonstrate your effort and skills on how to manage a messy dataset and turn it into a well-organized one.
  - Demonstrate your knowledge and skills on machine-learning models: such as and how to use them to analyze the data. Or you can extend a specific advanced machine learning model, such as Random Forest or deep learning to explore the details of it.
  - Find a problem that has yet to be discussed or analyzed.
  - Find a practical solution to the problem you are facing in your current/past work projects.
- There will be three deliveries:
  - PowerPoint slides.
  - R script, Markdown, Python script, or Jupyter.
  - The dataset.
- Similar to the spirit of GitHub, all your slides, codes, and datasets will be shared with your classmates after the end of the course, so you can learn from each other. Note: if your dataset cannot be shared, let me know. Then I will not share them.
- Either in your R Markdown or use a separate word file to clearly describe and explain your dataset and data analytics process and findings.
- Prepare your slides around **20 to 30** pages. Each of you will have **20 to 30 minutes** to present your project and results in the final week. And all the class and me will ask questions after your presentation **(5 to 10 minutes)**.
- To make a good presentation, there are three keys:
  - Really understand your data, your subject, and your mythology.
  - Data visualization is the best form in your presentation.
  - Practice.
- Your presentation date will be either on **May 30 (M), May 31 (Tu) or June 1 (W)**. Let me know which day you prefer to do presentation. *Note:* for those who don't inform me your preference day, you will usually be assigned to the first day.
- After the presentations, you might need to/want to revise your scripts and slides. The deadline to submit all the materials to Canvas is at 11:59pm on **June 10 (F)**.