Hi everyone,

**Project Info**

You must complete a final project at the end of this course. The point of the project is for you to **show what you have learned** in the course and have you create something you can put in your new Data Science work portfolio and that you can brag about on Twitter. You are given a lot of freedom in choosing how you want to execute it, all I want is for you to (1) find a dataset that you like, either by downloading something that is relatively clean or by scraping it similarly to how we got our hands on the Wikipedia dataset, (2) analyze the dataset with methods that you have learned here or even something that goes beyond the course if you need it, and (3) write a blog post about it that your smart non-technical friends will be able to understand.

The project is split into **two parts**:

* **Part A:** Craft a **short presentation** (~5 minutes or less) that proposes your project idea. In it you should (1) present a motivation statement [think: *"Have you also wondered how come...?"*] (2) specify what it is you want to do/find out, (3) how you want to do it and (4) why it is interesting. You can go into as much depth as you want, and if you can explain it in one minute that's fine too. Just make sure that the viewer understands what your project is about.
* **Part B:**
  + ***1:*** Write a **blog post** that explains your research journey. Do not err on the too technical side; try instead to explain it in a language that the average inquisitive reader would understand. The blog post must link to your **code on Github**. The figures you show in your blog post MUST be somewhere in there. It's the technical counterpart to your non-technical blog post and it explains all the stuff behind the scenes. There also needs to be a link to **the dataset** you scraped/used, so that the reader can reproduce the results.
  + ***2:*** With your group, give a 10 minute **oral presentation** of your project. You don't have to make an overtly flashy presentation here, just stand up and help us understand what you have done.

**Yardstick**

My ambition for your project is that you produce something you can add to your project portfolio and use as leverage when applying for work. I will, therefore, judge your project according to how likely I think it is to impress a potential employer. When grading, I will put myself in the shoes of a tech recruiter with an open mind and good domain knowledge.

Also, have a look at the projects from previous years to have a better sense of what's expected of you. Use them to understand the space in which you could situate your own explorations. They demonstrate a good range of possibilities in terms of which data and models to work with and how to use the media (blogposts + code reps) to showcase your findings.

List of previous final projects and proposal videos that turned out great:

Dashboard

Links

<https://covid19datahub.io/articles/python.html>

https://worldwater.io/

<https://www.researchgate.net/publication/345360214_An_Ideal_Big_Data_Architectural_Analysis_for_Medical_Image_Data_Classification_or_Clustering_Using_the_Map-Reduce_Frame_Work>