**PROJECT 1:** *Purpose*: Exploratory Data Analysis of a chosen data set/industry. Data set must be at least 100 rows, and of decent quality. TWO main parts: Research Questions & Predictive/Statistical Modeling

**Part 1: Research Questions:** 2-4 research questions. What are your hypotheses, i.e., what do you think will be/expect the answer to be? Create 1-2 visualizations to answer that research question. Was your hypothesis correct? In total, I am expecting 6-8 professional visualizations (but you can do way more). Further, you must have at least 3 different types/styles of visualizations, such as bar, line, scatter, violin, map, box, donut, Etc.

**Part 2: Predictive/Statistical Modeling:** Build at least one regression (linear or logistic) with at least one feature to predict one target. Feel free to try multiple features. Look at correlations, look at the shape of data, histograms, etc. Do you need to normalize (z-scores) your data? What's the pattern you're noticing? Is your model any good? Why or why not? Did you expect your model to be any good? *OPTIONAL*: Use an API as a supplemental data set. Use an ANOVA or Statistical Hypothesis Test to add additional rigor to the statistical modelling.\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  
You will need to follow our Data Analysis Recipe (below in thread)\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***Deliverables**

1. Project Proposal
2. What is your data? Why did you choose this data set?
3. Inspiration - what other public analysis has been done on this data set?
4. What are some possible research questions?
5. What is something you might be able to predict with a regression?
6. What will be your color palette?
7. Roles and Responsibilities.
8. Complete sentences, grammar, PDF, Images as needed

2. Presentation Slide Deck

1. Introduce the project - why did you choose it
2. Consistent color scheme - think about design
3. How did you clean/prepare the data
4. Go over your research questions, hypotheses, and visualizations
5. Go over the regression - is it any good?
6. Live demo/use of the predictive model (optional)
7. Go over the statistical hypothesis testing (optional but recommended)
8. Conclusions - call to actions
9. Limitations/bias
10. Future work
11. Work Cited (non-official, just links)

3. Analytics Write Up

1. Same deal as the slide deck, but a more formal whitepaper
2. Provide more details about things such as the correlations, more images, more in-depth analysis
3. Due 1 week after project presentations into your GitHub
4. Introduce the Project
5. How did you clean the data? Be specific. this is more TECHNICAL writeup than your more business-oriented presentation
6. Go over your research questions, include images. How do the images support your conclusions?
7. Go over the regression - include all images, correlations, predicted vs actual plots. Is it any good? How do you know? How can you make it better?
8. Conclusions/Calls to Action
9. Limitations & Bias
10. Future Work
11. Works Cited (non-official)

 4. All Code in an Organized fashion in your GitHub. Neat, commented, etc.

1. You can use screenshots of code in your presentation and writeup, just don't make it overwhelming
2. I should be able to read through your notebooks and understand its purpose, what it does, and why

 5. Presentation Day

1. Everyone must speak
2. Total presentation around 15-20minutes
3. Business Casual appearance, video on
4. One person shares screen and walks through the slide deck
5. Leave time for questions
6. First group starts at 7pm sharp