# Progress Report

The progress report for the final project is due **Thursday** May 18th at **11:59** Pacific time on gradescope.

There should be only one submission per group: use the group option to tag your teammates.

The progress report will likely be around **3-5 pages** (excluding the appendix). Feel free to draw as much as you would like from your project proposal.

The progress report should include the following. The point breakdown is given by [**X points**] indicators, which serves as a checklist for writing your update. The assignment is graded for completion, as it is intended to get you started working on your project early.

## 1 Introduction

Your introduction should address:

- what is your research question? [**1 point**]
- why is it important? [**1 point**]
- what methods and data will you use to answer it? [**1 point**]
- how will you **validate** your work? I.e. how can you show that you've done a good job (that your method is better than other methods or is quantifying something real). [**1 point**]

If you plan to do any statistical hypothesis testing (i.e. reporting p-values that have meaning), **you should state the hypothesis you are testing** in this section. You should formulate the statistical hypothesis you are testing **before you do any examination of the data.** Otherwise you are "double-dipping" and your p-values don't have any meaning. If you engage in double-dipping, that's okay too, just make sure to discuss exactly what you did in a manner that's reproducible.

The point is, if you "double dip," your p-values don't mean anything, but there are still interesting questions you can answer without looking at p-values.

## 2 Literature Review

Include a discussion of any relevant literature to your work. Relevant literature can include

- papers in statistics and machine learning that provide methods of use to you
- papers in sports analytics (such as from JQAS) that are relevant to your project
- blog posts related to your project.

For at least **two** pieces of literature, you should include a more in-depth, paragraph long discussion in which you

- summarize the main point of the paper [**1 point each**]
- discuss its relevance to your project [**1 point each**]
- explain how you will be extending or applying the paper in the project, or how your work is different from the paper. [**1 point each**]

## 3 Exploratory Data Analysis

Exploratory data analysis (EDA) involves playing around with you data and getting to know it better. You should include **at least one plot** (i.e. a nice graphic) [**2 points**] that demonstrates that you have done some exploratory analysis. Be sure to discuss the plot and what conclusions you have drawn from it.

## 4 Your Progress

Summarize the progress you have been on your project. You should include **at least one plot** (i.e. a nice graphic) [**2 points**] that demonstrates that you made some progress towards answering your research question (this plot should be different from the plot you used in the EDA section). Discuss the plot, the conclusions you have drawn from it, and why it represents progress on your project. [**2 points**]

## 5 Next steps

Discuss your vision for what your finished project will look like, and what steps you will take to get there. [**2 points**] Discuss any challenges you anticipate, and how you will surmount them. [**2 points**]

## 6 Appendix

The appendix does NOT count for the page-count. Your appendix should contain all data and code (as long as you are not using any private data; obviously private data should not be included). The best way to link to data and code is through a link to a github repo, but any method is fine as long as your work is reproducible.