

Stat 143: Final Project

Issued during class on 4/2. Final project report & code due on 4/30 (submit to Canvas, or by email to laurieshaw@g.harvard.edu). Project presentations will take place during class time on the 21st, 23rd and 28th April.

The final project is your opportunity to apply some of ideas, methods and data that we have covered in class to a question or topic of your choice in sports analytics. You are free to choose your own topic, but we have also provided a list of project ideas on the following page.

You may make use of any data or code provided throughout this course, but note that one of the grading categories is *innovation*: the degree to which you go beyond some of the examples that we covered in class. Examples of innovation might include:

- Exploring a question or topic not covered in class [*the history of home advantage in a sport, when should you withdraw the goaltender at the end of a hockey game?*]
- An application of the methods introduced in class [*how does RAPM correlate from one season to the next? Cluster pitch types and study the pitching of your favourite pitcher.*]
- Implementing a method discussed but not coded in class [*calibrate and apply an Elo (or Glicko) system for a sport of your choice*].
- Make use of a new dataset (i.e., unrelated to the data provided on canvas).
- Creativity in presenting your analysis (e.g., makes use of interactive visualizations).

The TFs and I are happy to discuss your project ideas (including the availability of data) during office hours and sections. Furthermore, the classes on the 14th and 16th April have been allocated to project work: you may work on your project in one of the break-out rooms or drop-in to ask a question.

This is the final assignment for Stat143. You should therefore dedicate to it the time equivalent to four weeks' worth of problem sets plus study for an exam.

Presentation & Report

Projects will be graded based on a short, 5-minute presentation (+ 3 minutes of questions) **given in class during 21st, 23rd or 28th April** (10% of final grade) and a final project report (40%). The presentation is an opportunity to get some feedback on your work so far. You will be expected to present some initial results, but it does not have to be your final analysis. A sign-up sheet will be circulated so that you can choose the data & time of your presentation. The final report should be in the style of a research paper, up to 4 pages in length (including figures) and is **due on Friday 30th April**.

Grading

The presentation is worth 10% of your final mark; the project report is worth 40%. We will grade project reports based on the following criteria:

1. Methodology (techniques used to analyze data to address your research topic)
2. Evaluation (interpretation of results)
3. Presentation (including clarity of text and visualizations)
4. Innovation (as described above)

Group working

You may work in groups of up to three people (although you do not have to), however you must submit your own report. We will expect group projects to be more ambitious in scope than a single-person project.

Project Ideas

- Study the history of home advantage in soccer, baseball, or football – how has it changed over time?
- Has COVID affected home advantage?
- How does each pitch change the game state (expected runs scored over the remainder of the inning)?
- Cluster pitch types and study the pitching of your favourite pitcher.
- Is there a significant difference in expected goals models between leagues in soccer?
- Do some players significantly exceed the total of their xG?
- How does regularized, adjusted plus-minus (RAPM) correlate from one season to the next?
- Build an expected-goals model for basketball shots.
- Can you measure aging curves for NBA players from their RAPM score?
- Can you implement a simple pitch control model in soccer (see lectures 19 & 20)?
- When should a team run or pass the ball in football?
- When should you withdraw the goalie in hockey?
- Is there any evidence for momentum in team results (i.e. outcomes on a team level, rather than a player level)?
- Explore the statistics behind Elo system: how does the standard deviation in observed ratings relate to the k factor (see lecture 4)
- Calibrate and apply Elo system for a sport of your choice
- Compare your match predictions with those from a publicly available source.