

# Course Syllabus

[Edit](#)

## Sports Analytics: A Capstone Course












### Professor Abraham Wyner

The class will meet every Thursday at 3:30 PM. Technically, we end at 6:30 but I will strive to finish closer to 6:00 PM (if possible).

You grade will be based on 4 assignments. The first is an individual assignment and the second is a group assignment as are the last two. You can switch up your groups, after each group assignment. Groups can be up to four students, ideally 3 in each group.

My office is 309 WARB, let me if you want to talk. We have two PhDs for the course. Hua Wang ([wanghua@wharton.upenn.edu](mailto:wanghua@wharton.upenn.edu) (<mailto:wanghua@wharton.upenn.edu>)) and Ryan Brill ([ryguy123@sas.upenn.edu](mailto:ryguy123@sas.upenn.edu) (<mailto:ryguy123@sas.upenn.edu>)). Hua will be managing the kaggle submissions and can help with coding and statistics. Ryan is an expert on sports analytics and he is the go to person for sports related questions (as well as statistics and coding). In addition, Justin Lipitz and Sarah Hu, students, will also be assisting me with the course and will be available to help you if needed.

## Course Summary:

Date	Details	Due
Thu Jan 12, 2023	 <a href="https://canvas.upenn.edu/courses/1704320/assignments/10817093">Background</a> ( <a href="https://canvas.upenn.edu/courses/1704320/assignments/10817093">https://canvas.upenn.edu/courses/1704320/assignments/10817093</a> )	due by 11:59pm
Thu Jan 19, 2023	 <a href="https://canvas.upenn.edu/courses/1704320/assignments/10862553">Assignment 1</a> ( <a href="https://canvas.upenn.edu/courses/1704320/assignments/10862553">https://canvas.upenn.edu/courses/1704320/assignments/10862553</a> )	due by 11:59pm
Tue Jan 24, 2023	 <a href="https://canvas.upenn.edu/calendar?event_id=3844588&amp;include_contexts=course_1704320">Ryan Brill's Office Hours</a> ( <a href="https://canvas.upenn.edu/calendar?event_id=3844588&amp;include_contexts=course_1704320">https://canvas.upenn.edu/calendar?event_id=3844588&amp;include_contexts=course_1704320</a> )	4pm to 5pm
Thu Jan 26, 2023	 <a href="https://canvas.upenn.edu/courses/1704320/assignments/10864139">Assignment 2</a> ( <a href="https://canvas.upenn.edu/courses/1704320/assignments/10864139">https://canvas.upenn.edu/courses/1704320/assignments/10864139</a> )	due by 3:30pm
Tue Jan 31, 2023	 <a href="https://canvas.upenn.edu/calendar?event_id=3844589&amp;include_contexts=course_1704320">Ryan Brill's Office Hours</a> ( <a href="https://canvas.upenn.edu/calendar?event_id=3844589&amp;include_contexts=course_1704320">https://canvas.upenn.edu/calendar?event_id=3844589&amp;include_contexts=course_1704320</a> )	4pm to 5pm
Tue Feb 7, 2023	 <a href="https://canvas.upenn.edu/calendar?event_id=3844590&amp;include_contexts=course_1704320">Ryan Brill's Office Hours</a> ( <a href="https://canvas.upenn.edu/calendar?event_id=3844590&amp;include_contexts=course_1704320">https://canvas.upenn.edu/calendar?event_id=3844590&amp;include_contexts=course_1704320</a> )	4pm to 5pm
Thu Feb 9, 2023	 <a href="https://canvas.upenn.edu/courses/1704320/assignments/10817092">Assignment 3</a> ( <a href="https://canvas.upenn.edu/courses/1704320/assignments/10817092">https://canvas.upenn.edu/courses/1704320/assignments/10817092</a> )	due by 3:30pm
Tue Feb 14, 2023	 <a href="https://canvas.upenn.edu/calendar?event_id=3844591&amp;include_contexts=course_1704320">Ryan Brill's Office Hours</a> ( <a href="https://canvas.upenn.edu/calendar?event_id=3844591&amp;include_contexts=course_1704320">https://canvas.upenn.edu/calendar?event_id=3844591&amp;include_contexts=course_1704320</a> )	4pm to 5pm
Tue Feb 21, 2023	 <a href="https://canvas.upenn.edu/calendar?event_id=3844592&amp;include_contexts=course_1704320">Ryan Brill's Office Hours</a> ( <a href="https://canvas.upenn.edu/calendar?event_id=3844592&amp;include_contexts=course_1704320">https://canvas.upenn.edu/calendar?event_id=3844592&amp;include_contexts=course_1704320</a> )	4pm to 5pm
Thu Feb 23, 2023	 <a href="https://canvas.upenn.edu/courses/1704320/assignments/10817094">Final Project</a> ( <a href="https://canvas.upenn.edu/courses/1704320/assignments/10817094">https://canvas.upenn.edu/courses/1704320/assignments/10817094</a> )	due by 3:30pm
Tue Mar 7, 2023	 <a href="https://canvas.upenn.edu/calendar?event_id=3844594&amp;include_contexts=course_1704320">Ryan Brill's Office Hours</a> ( <a href="https://canvas.upenn.edu/calendar?event_id=3844594&amp;include_contexts=course_1704320">https://canvas.upenn.edu/calendar?event_id=3844594&amp;include_contexts=course_1704320</a> )	4pm to 5pm

Sports Analytics  
A Capstone Course Proposal and Syllabus  
Professor Abraham Wyner  
February 5<sup>th</sup> , 2020  
Revision

## **Overview**

The “Moneyball revolution” in sports and inspired great interest in the transformative potential of statistics. Since classroom instruction is often highly theoretical and applications are usually simulation based, this “capstone” course will not only introduce students to the growing field of sports analytics but it will also allow our student to implement and integrate their extensive knowledge base by means of a deep dive into real sports data sets to solve real problems. While the context will be sports related and the expectation of students is that they are interested and knowledgeable about most major sports, the skills and techniques gained will be widely applicable and generalizable with applications in diverse areas.

The course will be open to students who have completed most of their course requirements for the concentration in Statistics or Business Analytics and have demonstrated mastery of the fundamental methods of data analysis. In the course, they will learn to apply their skills, receive feedback and learn new techniques.

## **Pre-requisites:**

Stat 471 (or 474), Stat 430 and knowledge of R or Python or an equivalent statistical coding language.

## **Format**

The tentative plan is to meet once, weekly in the second quarter for 3 hours. Students will turn in one short midterm assignment and will also be graded based on classroom participation. The first assignment will be a simple analysis and is expected to be a replication of a published analysis. The second, more extensive, final project is to be entirely original and will address a specific sports related managerial decision on or off the field.. For example, an investigation of the hot hand in baseball could be used to formulate a decision rule to determine if players should be removed based on “slumps” or elevated in the lineup when “hot”. Another example, could be to determine optimal strategies for using the various types of post-touchdown conversions being introduced into the NFL and the XFL.

It is hoped (but not required) that projects will involve collaborations among sports teams and related companies who will provide problems, data, and guidance. Student projects, in both parts, will be collaborative involving teams of 3-4 students.

## Topics

1. Introduction to Sports Analytics: The use of data for evaluation, prediction, strategy and training. We will discuss the general ways that analytics have impacted sports.
2. The Fundamental Mathematical Models of expected points and win probability: Markov-State-Space (Baseball), Binomial (Basketball), Poisson (Soccer and Hockey) and Regression (Football).
3. The homefield advantage: How big is it? Has it changed? What causes it? We will read several papers on home field advantage and study the difference between predictive and interpretive modeling.
4. Nationality Bias in Olympic Diving and Figure Skating Competition, Referee bias in Basketball and European Football. Non-parametric statistical methods of estimating p-values.
5. Hot Hand Fallacy! Is there a hot-hand in sports? Can it be measured?
6. Tracking Data and Machine Learning. Lessons from the NFL Data Bowl.

## Reading List:

1. **Moneyball**, Michael Lewis.
2. Selections from “**Handbook of Statistical Methods and Analyses in Sports**, 1st Edition Jim Albert, Mark E. Glickman, Tim B. Swartz, Ruud H. Koning.
3. Silberzahn, Raphael, et al. “**Many analysts, one data set: Making transparent how variations in analytic choices affect results.**” *Advances in Methods and Practices in Psychological Science* 1.3 (2018): 337-356.
4. **Scorecasting: The Hidden Influences Behind How Sports Are Played and Games Are Won** [L. Jon Wertheim](#), [Tobias Moskowitz](#)
5. **Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football** 2012 by [Wayne L. Winston](#) (Author).
6. **Sprawlball: A Visual Tour of the New ERA of the NBA**, by Kirk Goldsberry.
7. Emerson, John W., et al. “**Assessing Judging Bias: An Example From the 2000 Olympic Games.**” *The American Statistician*, vol. 63, no. 2, 2009, pp. 124–131. *JSTOR*, [www.jstor.org/stable/25652240](http://www.jstor.org/stable/25652240).
8. TVERSKY, A. and GILOVICH T. (1989b), “**The Hot Hand: Statistical Reality or Cognitive Illusion?**” *Chance*, 2, 31–34.