Sports Analytics: Course Description and Syllabus

Time and room

The class meets on Wednesdays from 9:00 am–12:15 pm in Uris 333
Dates: First class January 27, 2016; last class April 27, 2016

Instructor

Professor Mark Broadie
Office: 415 Uris Hall, fourth floor
E-mail: `mnb2@gsb.columbia.edu`
Phone: 212-854-4103
Office hours: Wednesdays, 12:30 pm–2:00 pm (or by appointment)

FIRST CLASS MANDATORY

Students will not be allowed to add the course unless they attend the first class.

Course work and grading

Grades will be based on the following weights:

- Final project and presentation: 40%
- Class participation (including attendance): 10%
- Homework (lowest score dropped): 40%
- Concept check quizzes: 10%

There will be a homework assignment for almost every class. These may be done individually or in a group of up to three people (which can change from assignment to assignment). If you would like help finding a team, please e-mail the TA. Assignments will be a mix of data analysis, thought questions, and reading. In computing your homework assignment grade, the lowest score will be dropped. There will be one concept check quiz per week. There are to be done individually

The final project can be on a sports topic of your choosing and can be done in groups of three or less. A good level to shoot for would be an article that could be published on

`fivethirtyeight.com/sports` or `sportsanalyticsblog.com`.

The project could also be on a topic from the *Mathletics* book not covered in class or from the academic literature. You could investigate an interesting result from one sport and apply it to another sport. The write-up should be approximately five pages, not including figures, charts, or supporting material in an appendix. Project presentations will be given on the last two class sessions.

Core culture

Core culture (present and on time, prepared and participating) is expected of everyone in the course. It is important for learning the material and for us learning from each other—and is more fun for all of us.

Connection to the core

This course builds on the Business Analytics and Statistics core courses. The course will use the tools of statistical analysis, predictive analytics, optimization and simulation. Issues of risk and return from Corporate Finance will be used in the course, as will game theory concepts from Managerial Economics and Strategy.

Course description

Sports analytics refers to the use of data and quantitative methods to measure performance and make decisions to gain advantage in the competitive sports arena. This course builds on the Business Analytics core course and is designed to help students to develop and apply analytical skills that are useful in business, using sports as the application area. These skills include critical thinking, mathematical modeling, statistical analysis, predictive analytics, game theory, optimization and simulation. These skills will be applied to sports in this course, but are equally useful in many areas of business.

There will be three main topics in the course: (1) measuring and predicting player and team performance, (2) decision-making and strategy in sports, and (3) fantasy sports and sports betting. Typical questions addressed in sports analytics include: How to rank players or teams? How to predict future performance of players or teams? How much is a player on a team worth? How likely are extreme performances, i.e., streaks? Are there hot-hands in sports performances? Which decision is more likely to lead to a win (e.g., attempt a stolen base or not in baseball, punt or go for it on fourth down in football, dump and chase or not in hockey, pull the goalie or not in hockey)? How to form lineups in daily fantasy sports? How to manage money in sports betting? How to analyze various "prop" bets?

The main sports discussed in the course will be baseball, football, basketball, hockey, and golf. Soccer, tennis, and other sports will be briefly discussed. Students are welcome to pursue any sport in more detail (e.g., cricket, rugby, auto racing, horse racing, Australian rules football, skiiing, track and field, or even card games such as blackjack, poker, etc.) in a project.

Class sessions will involve a mixture of current events, lecture, discussion, and hands-on analysis with computers in class. Each session will typically address a question from a sport using an important analytical idea (e.g., mean reversion) together with a mathematical technique (e.g., regression). Because of the "laboratory" nature of part of the sessions, students should bring their laptops to each class.

Guest speakers

There will be several guest speakers during the course. Potential speakers include:

- Jason Rosenfeld, Director of Basketball Analytics for the NBA (and formerly with the Charlotte Hornets)
- James Morrow, Fidelity Investments, hockey analytics and investing
- Jared Lander, Lander Analytics, football analytics and other applications of predictive analytics
- Joey Levy, Columbia undergrad, Founder and CEO of Draftpot (daily fantasy sports website)
- Dan Singer, senior partner at McKinsey, head of Sports and Gaming Practice
- Michael Lopez, Professor Skidmore College, NCAA March madness expert, freelance writer for Sports Illustrated/SI.com and fivethiryeight.com

Prerequisites

The prerequisites for the course are the *Managerial Statistics* and *Business Analytics* core courses. Good Excel skills are required and we will use the Business Analytics Excel add-in for analyzing data. Detailed knowledge of sports is not required, but familiarity with the rules of baseball, football, basketball, hockey and golf will be assumed. There are many websites and popular books where you can learn about these sports, if needed. For example, in baseball, you should understand the basic rules and what the terms bunting and base-stealing mean, e.g., as described in `https://en.wikipedia.org/wiki/Baseball_rules`.

Recommended textbook: *Mathletics*, 2009, Wayne Winston, Princeton University Press.

Additional recommended books

- *Every Shot Counts*, 2014, Mark Broadie, Gotham Books
- *The Success Equation*, 2012, Michael Mauboussin, Harvard Business Review Press
- *Big Data Baseball*, 2015, Travis Sawchik, Flatiron Books
- *Trading Bases*, 2013, Joe Peta, Dutton
- *Scorecasting*, 2011, Moskowitz and Wertheim, Crown Archetype
- *Analyzing Baseball Data with R*, 2013, Chapman & Hall/CRC

Software

All assignments are expected to be done in Excel, with some assignments requiring the use of the Business Analytics Excel add-in. There is no requirement to use other software for the assignments (though we might use Visual Basic in class). However, if you are familiar with another statistical package or programming language, you are welcome to use it for the project (e.g., you might want to use R if your project involves analysis of pitch/fx data).

Course outline (subject to change)

Sessions 1-4. Rating teams and players and predicting future performance

- Course overview
- Getting data from the web
- Which team is the best? Which team has the best offense? How to measure the strength of schedule?
- How to rate predictive models?
- Using ratings models to simulate future performance
- Bayes: using small data to predict long-run performance
- Streaks, momentum, mean-reversion and hot hands in sports
- Reading: *Mathletics*, chapters 40 and 43

Sessions 5-6. Decision-making in sports

- Baseball: analysis of bunting and base-stealing strategies
- Football: analysis of run versus pass, punt or go-for-it
- Run value added and win probability added
- Applications of state space / Markov chain and game theoretic methods
- Reading: *Mathletics*, chapters 6, 20–24

Sessions 7-9. Fantasy sports and sports betting

- Overview: betting markets, odds, "prop" bets, daily fantasy sports and fantasy leagues
- Contests and strategies: head-to-head, double up and guaranteed prize pool
- Performance prediction: exponential smoothing, mean reversion
- Forming lineups (volatility and correlation effects) and portfolios of lineups (integer and nonlinear optimization and simulation)
- Predict ownership fractions: is there a recency bias?
- Money management, Kelly criterion
- Analysis of prop bets
- Reading: *Mathletics*, chapters 40 and 43

Session 10. Golf analytics

- Measuring golf performance: What is the key to Tiger's success?
- Handicapping
- Official World Golf Rankings
- Predicting golf performance
- Improving golf performance: optimal risk-reward tradeoffs
- Optional reading: *Every Shot Counts*, chapters 5 and 8

Sessions 11-12. Project presentations
