4 Columbia Business School

Sports Analytics: Block Week Syllabus (Summer 2024)

Instructor

Professor Mark Broadie
Office: 1131 Kravis, eleventh floor
E-mail: `mnb2@gsb.columbia.edu`
Phone: 212-854-4103

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 main application area. These skills include critical thinking, mathematical modeling, statistical analysis, predictive analytics, game theory, optimization and simulation. Most of the application areas in the course will be to sports, but other business applications will be explored in homework assignments.

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) 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)? Are betting markets efficient? How to manage money in sports betting?

The main sports discussed in the course will be soccer, baseball, football, basketball, hockey and golf. Tennis, and other sports will be briefly discussed. Similar ideas developed in the course apply to many other sports, including cricket, rugby, darts, auto racing, horse racing, Australian rules football, skiing, curling, track and field, and even card games such as blackjack, poker, etc.

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.

Course work and grading

Grades will be based on the following weights:

- Class participation: 20%

- Concept check quizzes (4, individual): 20%

- Pre-session quizzes (group discussion, individual submission): 5%

- Homework (five total: 2 group, 3 individual): 20%

- Final exam (individual): 35%

Class participation will be based on: (i) nameplate (25%, based on the fraction of classes your attend with your nameplate clearly visible), (ii) attendance (25%, based on fraction of classes attended and arrived on time, unless previously excused), (iii) laptop closed (25%, except when laptops needed for class exercises), and (iv) contributions in class (25%, based on my impressions of participation in class, and other contributions including guest speaker questions, finding typos, bringing interesting articles to our attention).

Concept check quizzes: There will be daily concept check quizzes that are to be done individually (no discussion with other students).

Homework: Each homework will be based on a real-world application of the ideas and methods discussed in class. To account for varying backgrounds, homework will be graded on effort and partially on correct answers. Effort means that the work is clearly presented (e.g., the main formulas in spreadsheets are shown and explained) and the logic and steps are clearly shown and explained. The homework grades will range on a 1 to 5 scale, with meanings: 1 (little effort), 3 (work clearly explained though some wrong answers), 4 (all work clearly explained and almost all answers correct OR work not clearly explained even if all answers are correct), 5 (all work clearly detailed and explained and all answers correct).

Final: The final exam will have a maximum time limit of four hours. A practice final and solutions will be provided on Canvas. The final will include multiple choice questions, short "concept check"-type questions, and questions that are "homework level of difficulty." The final will be due within two weeks after the class ends.

Pre-Work

Before the class starts, you will be required to either: (i) install the Business Analytics Excel add-in (**this version of the add-in runs under Windows only, so Mac users will need to boot into Windows to install and use the add-in**), or (ii) take the Python level 1 class `https://www8.gsb.columbia.edu/courses/python` which will allow you to run python programs distributed in the course. Anyone who has not completed the pre-work at least three days before class begins will be removed from the class.

Prerequisites

Computer skill prerequisites: Good Excel skills are required. No programming experience is required. All examples in class will have both Excel and python equivalents and separate demon-

stration videos. You will have the choice of running the analyses yourself in Excel or python. For the Excel option, you'll need the Business Analytics Windows Excel add-in which requires **you to have a laptop computer that is either a Windows machine or a Mac machine that can boot into Windows in order to run Windows Excel.** For the python option, you'll need to have installed python and have taken the python level 1 basics course.

Detailed knowledge of sports is not required, but familiarity with the rules of baseball, American football, basketball, and golf will be assumed. For example, in baseball, you should understand the basic rules and what the terms bunting and base-stealing mean as described in `https://en.wikipedia.org/wiki/Baseball_rules`. For football, you should know about plays, downs, and scoring, as described in `https://en.wikipedia.org/wiki/American_football_rules`.

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, dynamic programming, game theory 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.

Guest speakers

There will be two or three guest speakers during the course. Past speakers have been professionals working in sports analytics groups for professional teams or sports analytics companies.

Textbook: There is no required textbook for the course. Recommended books include:

- *Mathletics, 2nd edition*, 2021, Wayne Winston, Princeton University Press
- *Scorecasting*, 2011, Moskowitz and Wertheim, Crown Archetype
- *The Success Equation*, 2012, Michael Mauboussin, Harvard Business Review Press
- *Every Shot Counts*, 2014, Mark Broadie, Avery Press
- *Trading Bases*, 2013, Joe Peta, Dutton

Course outline

The block week sessions will be numbered 1 for Monday morning, 2 for Monday afternoon, ..., 10 for Friday afternoon.

Sessions 1-4. Measuring and predicting the performance of players and teams

- Course overview
- Building an expected goals (xG) model in soccer
- Ratings teams and measuring the strength of schedule
- Predicting future performance of players and teams
- Assessing alternative models of performance
- Streaks, momentum, mean-reversion and hot hands in sports

Sessions 5-7. Decision making in sports

- Markov modeling of games: states, transition probabilities and state values
- Baseball: analysis of bunting, base-stealing and other strategies
- Football: analysis of run versus pass, punt or go-for-it
- Soccer: analysis of passes and "on ball" value (Markov chain analysis)
- Hockey: analysis of when to pull the goalie (dynamic programming)
- Assessing the value of a play using run value added and win probability added
- Decision making with strategic interactions (game theory)

Sessions 8-9. Sports betting

- Overview: betting markets, odds, setting betting lines
- Investigating the efficiency of betting markets
- Setting betting lines
- Money management and the Kelly criterion

Session 10. Golf analytics: performance measurement, decision making and sports betting