**Sports Analytics Hackathon**

***Participants:***

The event will be open to all members of Columbia Statistics Club.

***Speakers:***

James Curley (Professor of Psychology)

Zaheer Benjamin (Head of Business Planning & Analytics, Chelsea Football Club) – will speak remotely

***Judges:***

James Curley (Professor of Psychology)

Greg Wawro (Professor of Political Science)

***Coaches:***

Zhuxi Cai (Recent Graduate of the MA in Statistics at Columbia University)

Xianyun Wang (Recent Graduate of the MA in Statistics at Columbia University)—both days

Guangyu Wang (Recent Graduate of the MA in Statistics at Columbia University)—moring of 1st day

***Location:***

SSW Room 903 (1255 Amsterdam Ave, New York, NY 10027)

***Schedules:***

16th April, Saturday, 9:00-17:00

|  |  |
| --- | --- |
| 9:00-10:00 | Registration, breakfast, ice-breaking |
| 10:00-10:30 | Announcement on rules, logistics, topics |
| 10:30-11:00 | Team-building |
| 11:00 | Hackathon starts |
| 12:00 | Lunch |
| 13:00-13:45 | Lecture 1-- James |
| 13:45-14:00 | Lecture 1 Q&A-- James |
| 14:00 | Continue hacking |
| 15:00-15:45 | Lecture 2—Zaheer  Data Mining & Analytics in Professional Sports |
| 15:45-16:00 | Lecture 2 Q&A-- Zaheer |
| 16:00-17:00 | Continue hacking |

17th April, Sunday, 9:00-17:00

|  |  |
| --- | --- |
| 9:00-9:30 | Breakfast |
| 9:30-12:00 | Continue hacking |
| 12:00 | Lunch |
| 15:00 | Stop hacking |
| 15:00-16:00 | Presentation |
| 16:00-16:30 | Judgment |
| 16:30-17:00 | Winner announced, prizes awarded |

***Data Description:***

The projects will be using PUBLIC DATA in sports field.

***Submitting Projects:***

Final submission consists of: Title/Name of project; a short description of project (~300 words); name of team & team members; data source; code (linked through Github); screen-cast or screenshots describing the project. A link will be provided for project submissions on Sunday.

***Presentations:***

Presentations will be judged during an electronic poster session on Sunday where judges will go around and ask teams to present their projects projected on a display or screen. Teams will present using their own computers. Time allotted for each presentation is 7 minutes (strictly enforced).

***Judging criteria:***

- Technical Difficulty

- Impressiveness

- Creativity

- Usefulness/Practicality

***Awards:***

To be determined

***Sponsors:***

Department of Statistics | Columbia University

Center for foundations of data science, Data Science Institute

***Task Description***

The theme of this hackathon is uncovering actionable insights from publicly available NBA datasets. You should use the provided list of questions to motivate your team’s project. You are not expected to answer every question or sub-question in the list - it is better to have a great answer to a narrow question than having a so-so answer to a broad question. Visualization or modeling approaches are both relevant, so long as your work illuminates some aspect of the datasets.

Data

o   NBA Play by Play Data. These are game logs that detail specific actions in specific games. The second link has structured datasets but older data. The third link contains a Python scraping code for ESPN data, which we haven’t tested.

o   <http://www.basketball-reference.com/boxscores/pbp/201402090LAC.html>

o   <http://basketballvalue.com/downloads.php>

o   <https://gist.github.com/drewdresser/8153408>

o   NBA Player Box Score statistics. These pages have statistics organized by player.

o   <http://www.basketball-reference.com/players/c/curryst01.html?lid=header_players>

o   NBA Team Box Score statistics. These pages have statistics organized by team.

o   <http://www.basketball-reference.com/teams/GSW/2016.html>

o   NBA Optical Tracking statistics. These pages have structured data based on optical tracking of games. Analysts have found it challenging to find meaningful insights from this data so use at your own risk.

o   [http://stats.nba.com/tracking/#!/player/](http://stats.nba.com/tracking/#%21/player/)

Questions

o   Time-out usage. How effective are NBA teams in using time-outs? What is the expected excess team wins for optimizing time-out usage? Does using time-outs to stop opponent team momentum (e.g. 8-0 scoring run) work? Does using time-outs during end-of-game situations work?

o   End-of-game fouling.  What is the optimal strategy for when to foul in end of game situations? How close to optimal had teams been operated? What is the expected excess team wins for optimizing end-of-game fouling?

o   Hack-a-Shaq. Read about this strategy here: <https://en.wikipedia.org/wiki/Hack-a-Shaq>. Does this strategy work? When does it work?

***Bios of Speakers and Judges***

**Kaiser Fung**

Professor Kaiser is a well known author and blogger on the topic of visualization. He provides training and advisory services in business analytics and data visualization. For over a decade, he held leadership positions on data teams at Vimeo, SiriusXM Radio, [X+1], and American Express. He directs the Applied Analytics program at Columbia University, and write the Statbusters column in The Daily Beast, with Columbia professor Andrew Gelman.

He teaches business analytics and data visualization at Columbia University, and runs workshops for corporations and organizations. He has also taught at New York University for over 10 years. He holds an MBA from Harvard Business School, and degrees in engineering and statistics from Princeton and Cambridge Universities.

**James Curley**

He has conducted and published research at molecular, systems, organismal and evolutionary levels of analysis in both animals and humans. The focus of his lab at Columbia is on the development of social behavior. Briefly, he is interested in how both inherited genetic variability and social experiences during development can shift individual differences in various aspects of social behavior and what the neuroendocrinological basis of these differences may be. He is also interested in the reliability and validity of social behavioral tests conducted in the laboratory and whether it is possible to utilize alternative statistical and methodological approaches to more appropriately assess social behavior. He also believes that it is critical to understand how the 'social brains' of humans and other animals have been differentially shaped by evolution and to acknowledge how this should better inform translational research.

**Zaheer Benjamin**

Zaheer Benjamin is Head of Business Planning & Analytics for Chelsea Football Club. In this role he is responsible for leading research and strategic analytics across all areas of Chelsea’s commercial operations. Prior to Chelsea Zaheer spent four years with the Phoenix Suns as VP of Business Planning & Basketball Analytics where he was broadly responsible for improving both commercial and basketball performance through the use of data driven decision making.

Zaheer holds an MBA from the MIT Sloan School of Management, as well as a Bachelor of Science & Engineering (BSE) in Computer Science from Princeton University. In addition to his experience at Chelsea FC and the Suns, Zaheer has held Strategy and Operations positions at Pfizer, Schering-Plough, Orlando Magic, and Sirius XM Satellite Radio.

**Tian Zheng**

Tian Zheng is associate professor of Statistics at Columbia University. She obtained her PhD from Columbia in 2002. Her research is to develop novel methods and improve existing methods for exploring and analyzing interesting patterns in complex data from different application domains. Her current projects are in the fields of statistical genetics, bioinformatics and computational biology, feature selection and classification for high dimensional data, and network analysis. Especially, Dr. Zheng have been developing statistical and computational tools for high dimensional data, searching for genetic interactions associated with complex human disorders, quantifying social structure and studying hard-to-reach populations using survey questions, with more than 40 peer-reviewed publications in journals including JASA, AOAS and PNAS. Her work was recognized with the 2008 Outstanding Statistical Application Award from the American Statistical Association, The Mitchell Prize from ISBA and a Google research award. She is on the editorial board of Statistical Analysis and Data Mining and Frontier in Genetics. She was Associate Editor for JASA from 2007 to 2013.

**Gregory Wawro**

Gregory Wawro (Ph.D., Cornell, 1997) specializes in American politics (including Congress, elections, campaign finance, judicial politics, and political economy) and political methodology. He is the author of Legislative Entrepreneurship in the U.S. House of Representatives and co-author (with Eric Schickler) of Filibuster: Obstruction and Lawmaking in the United States Senate, which is an historical analysis of the causes and consequences of filibusters. He has published articles in The American Journal of Political Science, The Annual Review of Political Science, Critical Review, Legislative Studies Quarterly, the Journal of Law Economics and Organization, and Political Analysis. His academic awards include the Richard J. Fenno Prize for best book in legislative studies in 2006, the E.E. Schattschneider Award, the Milton J. Esman Award, the CQ Prize for best paper presented in the Legislative Studies section at the 2002 APSA meeting, a Mellon Foundation Graduate Fellowship, and a John M. Olin Faculty Fellowship. He has been a visiting scholar at the Center for Basic Research in the Social Sciences at Harvard University.