

FSU SPORTS ANALYTICS SPEAKER SERIES

FSU SPORTS ANALYTICS SPEAKER SERIES

Slide Link: fsu2025.sportsdataverse.org

Saiem Gilani

OVERVIEW

The topic of our conversation will center around how to get access to sports data and why you should know your way around the data generated by the sports in which you are interested.

ABOUT ME

Saiem Gilani - Lead engineer and founder of the SportsDataverse

X @SAIEMGILANI

@SAIEMGILANI

292

@SAIEMGILANI.BSKY.SOCIAL

604

BACKGROUND

Born and raised a Seminole and a proud Tallahassee native. I am an FSU alumnus in mathematics and went to graduate school at Georgia Tech for analytics. My general domain of work is machine learning and data science with a current focus on sports.

WHERE ARE DATA SKILLS USED IN SPORTS?

- Data-Driven Decision Making
 - Performance Optimization: player performance, tracking fitness, etc
 - Tactical Insights: analyzing opponent tendencies, optimizing game plans, etc
- Player Evaluation
 - Scouting: evaluating player performance, identifying potential recruits, etc
 - Drafting: evaluating draft prospects, making informed decisions on draft picks, etc

WHERE ARE DATA SKILLS USED IN SPORTS? (CONT'D)

- Fan Engagement
 - Social Media: analyzing fan sentiment, creating engaging content, etc
 - Marketing: targeted advertising, personalized experiences, etc
- Revenue Generation
 - Sponsorship Insights: data helps teams demonstrate ROI to sponsors by quantifying fan engagement and brand exposure
 - Dynamic Pricing: engineers create models for ticket pricing that adjust based on demand, maximizing revenue.

WHERE ARE DATA SKILLS USED IN SPORTS? (CONT'D)

- Injury Prevention
 - Wearable Technology: monitoring player health data, identifying injury risks, etc
 - Load Management: Teams can track training loads and recovery times to minimize overtraining and optimize performance
- Operational Efficiency
 - Automated Workflows: streamlining processes like video analysis, data collection, and reporting, saving time for coaches and analysts.
 - Integration of Systems: integrating disparate data sources, creating a cohesive ecosystem for decision-making.

HOW DID YOU GET STARTED?

I attended the FSU Sports Analytics Summit 2020!

I wrote up some of my thoughts and observations on incoming coach Mike Norvell's presentation on a handful of analytics-related topics. This became my first article for [Tomahawk Nation](#), the SBNation blog covering the Seminoles.



Further reading

- Gilani, Saiem (2020). [Thoughts on FSU head coach Mike Norvell's sports analytics presentation](#). Tomahawk Nation.

STARTED MEETING GREAT FOLKS ONLINE

Simultaneously, I started working with the `{cfbscrapR}` package (now archived) to help write analytics driven articles.

I would not be here without my collaborators from the `cfbscrapR` team:

- Meyappan Subbiah
- Parker Fleming

I quickly became involved with contributing to my first open-source package on [GitHub](#), eventually becoming a co-author. I then developed the successor to the package, `{cfbfastR}`.



WENT TO ANOTHER CONFERENCE

- I used my experience from going to the FSU Sports Analytics Symposium to sharpen my networking and communication skills
- A couple weeks later, I went to the 2020 MIT Sloan Sports Analytics Conference
- Got to meet and see some sports analytics celebrities like Seth Partnow, John Hollinger, and [Alok Pattani](#)
- Competed in the Hackathon, an exceptional opportunity to work with and chat with very talented individuals about shared research ideas and further steps we could take with our projects

EVERYTHING CAME TO A SCREECHING HALT



THEN, I HAD AN IDEA

I had a thought I am sure many of the long-standing members of the sports analytics community has had.

- *what if getting sports data for analysis was easy?*
- *what if we worked together to build the data infrastructure for research?*
- *how much further would we get?*

THE SPORTSDATAVERSE

- An organization trying to make the sports data and analytics industry more diverse, inclusive, and accessible by providing high-quality resources for end-users and opportunities for practical code skill development for those that join the effort



- A set of packages for loading and scraping sports data in R, Python, and Node.js with focus placed on play-by-play data



THE STRENGTH OF THE SPORTSDATAVERSE

- A community of developers committed to developing and maintaining open-source sports data packages and pipelines as on-going public utilities



- A set of corresponding data repositories which allow fast loading of the data for users and collectively form one of the largest open-source sports data resources with **over 250Gb of data produced** from the packages I contribute to



- Our organization helps establish the bench of developers from diverse backgrounds to spearhead projects and make contributions

OUR PROGRESS SO FAR

20+ R packages with over a dozen sports leagues covered.

Pro Leagues

- NBA
- WNBA
- NBA G-League
- MLB
- NHL
- PWHL
- NWSL
- A boatload of soccer leagues

Collegiate Leagues

- College Football
- Men's College Basketball
- Women's College Basketball
- College Baseball
- College Softball
- College Football Recruiting
- College Basketball Recruiting

OUR PROGRESS SO FAR +

Access to loadable SDV-provided data and functions in the `sportsdataverse` python module and access to ESPN endpoints. Additional modules include: `sportypy`, `collegebaseball`, `nws1py`, and `recruitR-py`

Access to ESPN endpoints (among other websites) via the `sportsdataverse` node.js module for easy web application development.

WHY USE THE SPORTSDATAVERSE?

The first public conversation on the SportsDataverse projects happened at the Carnegie Mellon Sports Analytics Conference. The [paper](#) I wrote for the conference was selected as the winner for the Data and Software contribution, Open Track for their reproducible research competition.

- It was built for you enthusiasts and soon-to-be entrants into the field
- Allows users to quickly access seasons worth of datasets (which are updated nightly via automated [GitHub Actions](#)) via loading function calls, taking the burden off users to maintain their own web scraping scripts
- This in-turn provides significantly easier opportunities for reproducible research and reporting

Further reading

- Gilani, Saiem (2021). [The SportsDataverse: An Open Sports Data Initiative](#). 2021 CMU Sports Analytics Conference.

GREAT... BUT HOW DOES THIS HELP ME?

WELL, THERE IS A FAIRLY DIRECT PIPELINE FROM...

- Contributing to open-source projects
- Using open-source resources to create your own open-source sports analytics projects and portfolio
- Being an active member of the open-source sports analytics community
- Attending conferences and meetups to meet other sports analytics enthusiasts and professionals (I recommend the [CMU Sports Analytics Conference](#))
- Competing in collaborative hackathons and data challenges

...TO GETTING A JOB IN SPORTS ANALYTICS

More on this topic

- Ventura, Sam (29 October 2022). [Open-Sourcing the Sports Analytics Hiring Process](#). 2022 CMU Sports Analytics Conference.

HOW DO I GET STARTED?

- Start with one of the many open-source sports data packages available from the [SportsDataverse](#) and their tutorials at their respective documentation sites
- Use the data to create your own projects and analyses to publish in blogs or on social media
- Work through introductory sports analytics texts like [Analyzing Baseball Data with R, 3rd Ed.](#) which uses the `baseballr` package to get you started with baseball data analysis

FIND OPPORTUNITIES TO BUILD YOUR PORTFOLIO

- Contests are a great way to get involved with the community and meet other sports analytics enthusiasts
- Start sharing your code and projects online **so that people may see them**
- You may get opportunities to share projects created during the interview process on GitHub and in your portfolio
- Listen to thought-provoking conversations and let your curiosity lead you to new projects which address the questions using the data you have access to

THEN SHARE ON SOCIAL MEDIA!

- There are countless examples of people getting hired straight off their analysis on Twitter
- Build a following to increase your network reach
- Embrace non-traditional media and practice adapting communication of your analyses to different formats and audiences
- Be prepared to have not nice things said about your work
- Take feedback constructively and incrementally improve your projects

CONTINUE YOUR LEARNING JOURNEY

AthlyticZ offers courses in Python, R, Shiny Applications, Stan programming and beyond!

PLATFORM & INFRASTRUCTURE

- Fully functional Learning Management System built with Node.js and React + Google Cloud Platform for seamless coding experience.

PARTNERSHIPS & TEAM

- Curricula designed by industry and academic veterans from the NBA, MLB, ESPN, Notable authors from recognized textbooks in R/Python and members at Columbia University + UPenn.

ATHLYTICZ ACADEMY

STATE-OF-THE-ART OFFERINGS

- Customized curricula to meet the market needs for a blend of data science with sports analytics applications

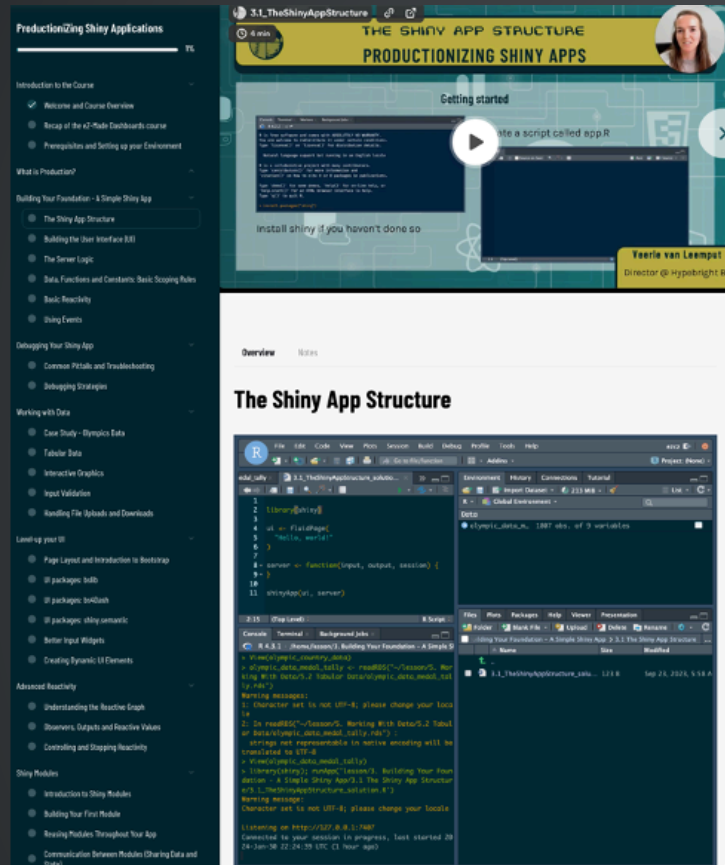
PER STUDENT IDE'S/CPU'S

- All coding modules will appear with preinstalled packages and data through a familiar IDE

CUSTOM CMS

- Instead of an off the shelf LMS/CMS – we invested in the development of a customized environment to ensure scalability, bespoke content management, code ownership and more.

ATHLYTICZ PREVIEW



Sign up for preview of our Bayesian Modeling course using real NBA and soccer data

FREE TRIAL: <https://athlyticz.com/stan-i-preview>

SOME INSPIRATIONS AND HEROS

MY BEAUTIFUL AND BRILLIANT WIFE, MADIHA, AND MY FAMILY

MY COLLABORATORS FROM THE **cfbfastR** TEAM:

- Akshay Easwaran
- Jared Lee
- Eric Hess



THE CREATOR OF **COLLEGEFOOTBALLDATA.COM**:

- Bill Radjewski



THE **nflverse** TEAM:

- Sebastian Carl
- Ben Baldwin
- Tan Ho





THANK YOU

- FSU Sports Analytics Club for creating a wonderful speaker series
- The seriously awesome community of developers that helps build and maintain resources
- All y'all for listening in

LEARN MORE

- SportsDataverse.org

 @SPORTSDATAVERSE  @SPORTSDATAVERSE.ORG 122

- [cfbfastR docs](#)

 @CFBFASTR 

- [Game on Paper](#) - for a look at the [sportsdataverse](#) python package serving live CFB expected points and win probability metrics

 @GAMEONPAPER  @GAMEONPAPER.COM 4479

- [AthlyticZ.com](#) - for state of the art data science training from sports industry veterans

 @ATHLYTICZ  @ATHLYTICZ.BSKY.SOCIAL 324

- [guideR](#) - compilation of sports analytics resources

- Connect with me:

 @SAIEMGILANI  @SAIEMGILANI 292  @SAIEMGILANI.BSKY.SOCIAL 604

QUESTIONS?