



Spring 2023

## Sports Data Science Job Requirements

Cam E. Morse  
*Gettysburg College*

Follow this and additional works at: [https://cupola.gettysburg.edu/student\\_scholarship](https://cupola.gettysburg.edu/student_scholarship)



Part of the [Data Science Commons](#), and the [Sports Studies Commons](#)

**Share feedback** about the accessibility of this item.

---

### Recommended Citation

Morse, Cam E., "Sports Data Science Job Requirements" (2023). *Student Publications*. 1067.  
[https://cupola.gettysburg.edu/student\\_scholarship/1067](https://cupola.gettysburg.edu/student_scholarship/1067)

This open access student research paper is brought to you by The Cupola: Scholarship at Gettysburg College. It has been accepted for inclusion by an authorized administrator of The Cupola. For more information, please contact [cupola@gettysburg.edu](mailto:cupola@gettysburg.edu).

---

## Sports Data Science Job Requirements

### Abstract

Data science is an extremely fast growing field in which job opportunities are opening in every industry related to data science. Within the data science field is the sports data science industry which has its own requirements and specificities that may not be present in other industries. In this paper, research is done using multiple job posting websites such as LinkedIn, Indeed, Sportstek jobs, and TeamworkOnline to explore the job descriptions of many different sports data science jobs. These job descriptions are then examined using Python coding to find the frequencies of specific data science skills in the various job postings to see which skills are most important for aspiring sports data scientists to learn.

### Keywords

Data science, Data analytics, Sports, Job Postings

### Disciplines

Data Science | Sports Studies

### Comments

Written for MGT 460: Individualized Study-Research.

### Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

## Sports Data Science Job Requirements

For this project I was interested in looking at the types of skills and competencies that were desired/required by data science job postings in the sports industry. This idea came from a content analysis that looked at the same topic but for data science postings in the healthcare industry ([Meyer](#)). From my prior research I hadn't found a similar study for the sports data science industry, so I was interested in seeing what skills/competencies would be desired, and whether sports data science jobs would require more knowledge of the specific industry of sports rather than just broad data science knowledge. To collect the data for this project, I used various job posting websites such as LinkedIn, Indeed, Sportstek jobs, and TeamworkOnline to gather job descriptions of data science/analyst positions for MLS, Premier League, NWSL, NBA, MLB, NHL, and NFL teams, as well as sports data companies, which I then put into an Excel file. Once I had this Excel file with all of the different job postings, I then converted it into a CSV file which allowed me to read it into Python and put it into a DataFrame. The data being in a DataFrame then made it so I could count the occurrences of a "string" or a word/phrase in a column of the DataFrame. I was interested in counting different skills/competencies that were present in the "Job Description" column, so I used code to count how many times my selected skills such as Python, SQL, and many other things occurred within that column. For the "Sports Knowledge/Experience" row in my results table I was unable to have Python count this automatically because the way that the job postings word requirements about this area can differ greatly from one posting to the next. For example, one job posting may say they require 3+ years working in the sports industry while another job posting may say they require experience working with a specific sport's data, such as soccer.

As shown in the table, Python was the skill that was most required/desired in sports data science job postings with SQL and R just behind it. Tableau was the most highly sought-after data visualization tool for data scientists to have experience with. Additionally, over half the job postings wanted applicants to have had some sort of experience with, or knowledge of, the sport that the team or company works with. As mentioned previously, there were some limitations on what skills/competencies could be automatically gathered using Python due to the different ways that these skills/competencies can be listed in job postings. For example, some jobs may require experience with data pipelines or storytelling, but the phrasing of these things can differ from job to job, so it is hard to capture all of them just using a singular term which is what the code that was used requires.

In conclusion, there are some skills that are very important for aspiring sports data scientists to learn such as Python, SQL, R, and even Excel. These skills are likely to be important in any data science industry, however, if someone is interested in going into the sports data science industry it is extremely helpful to have knowledge of the particular sport that you are applying to a job for. Not all data science job industries require prior knowledge or experience in the specific industry, such as in the healthcare industry, where many of the job postings only focus on your data science experience rather than your healthcare experience.

Required/Desired Skill	Count	Percent of Job Postings: n = 39
Python	29	74%
SQL	27	69%
R	27	69%
Excel	21	54%
Sports Knowledge/Experience	21	54%
Tableau	13	33%
Power BI	8	21%
PowerPoint	6	15%
Machine Learning	6	15%
AWS	6	15%
Azure	4	10%
Shiny	4	10%
Use of Opta Data	3	8%
SAS	2	5%
Hadoop	1	3%