UDACITY
(http://blog.udacity.com)

8 Skills You Need to Be a Data Scientist

By *Dave Holtz*
November 7, 2014

(https://twitter.com/intent/tweet?url=http://blog.udacity.com/2014/11/data-science-job-skills.html&via=udacity&text=8 Skills You Need to Be a Data Scientist)

Demystifying Data Science: 4 Kinds of Data Science Jobs and 8 Skills that Will Get You Hired

Interested in landing a job as a data scientist? You're in good company – a recent article (http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1) by Thomas Davenport and D.J. Patil in the Harvard Business Review calls 'data scientist' the sexiest job of the 21st century.

But how can you get your foot in the door? Many resources out there may lead you to believe that becoming a data scientist requires comprehensive mastery of a number of fields, such as software development, data munging, databases, statistics, machine learning and data visualization.

Don't worry. In my experience as a data scientist, that's not the case. You don't need to learn a lifetime's worth of data-related information and skills as quickly as possible. Instead, learn to read data science job descriptions closely. This will enable you to apply to jobs for which you already have necessary skills, or develop specific data skill sets to match the jobs you want.

4 Types of Data Science Jobs

"Data scientist" is often used as a blanket title to describe jobs that are drastically different. Here are four types of data science jobs:

A Data Scientist is a Data Analyst Who Lives in San Francisco: All joking aside, there are in fact some companies where being a data scientist is synonymous with being a data analyst. Your job might consist of tasks like pulling data out of MySQL databases, becoming a master at Excel pivot tables, and producing basic data visualizations (e.g., line and bar charts). You may on occasion analyze the results of

an A/B test or take the lead on your company's Google Analytics account. A company like this is a great place for an aspiring data scientist to learn the ropes. Once you have a handle on your day-to-day responsibilities, a company like this can be a great environment to try new things and expand your skillset.

Please Wrangle Our Data!: It seems like a number of companies get to the point where they have a lot of traffic (and an increasingly large amount of data), and they're looking for someone to set up a lot of the data infrastructure that the company will need moving forward. They're also looking for someone to provide analysis. You'll see job postings listed under both "Data Scientist" and "Data Engineer" for this type of position. Since you'd be (one of) the first data hires, there are likely many low-hanging fruit, making it less important that you're a statistics or machine learning expert. A data scientist with a software engineering background might excel at a company like this, where it's more important that a data scientist make meaningful data-like contributions to the production code and provide basic insights and analyses. Mentorship opportunities for junior data scientists may be less plentiful at a company like this. As a result, you'll have great opportunities to shine and grow via trial by fire, but there will be less guidance and you may face a greater risk of flopping or stagnating.

We Are Data. Data Is Us: There are a number of companies for whom their data (or their data analysis platform) is their product. In this case, the data analysis or machine learning going on can be pretty intense. This is probably the ideal situation for someone who has a formal mathematics, statistics, or physics background and is hoping to continue down a more academic path. Data Scientists in this setting likely focus more on producing great data-driven products than they do answering operational questions for the company. Companies that fall into this group could be consumer-facing companies with massive amounts of data or companies that are offering a data-based service.

Reasonably Sized Non-Data Companies Who Are Data-Driven: A lot of companies fall into this bucket. In this type of role, you're joining an established team of other data scientists. The company you're interviewing for cares about data but probably isn't a data company. It's equally important that you can perform analysis, touch production code, visualize data, etc. Generally, these companies are either looking for generalists or they're looking to fill a specific niche where they feel their team is lacking, such as data visualization or machine learning. Some of the more important skills when interviewing at these firms are familiarity with tools designed for 'big data' (e.g., Hive or Pig) and experience with messy, 'real-life' datasets.
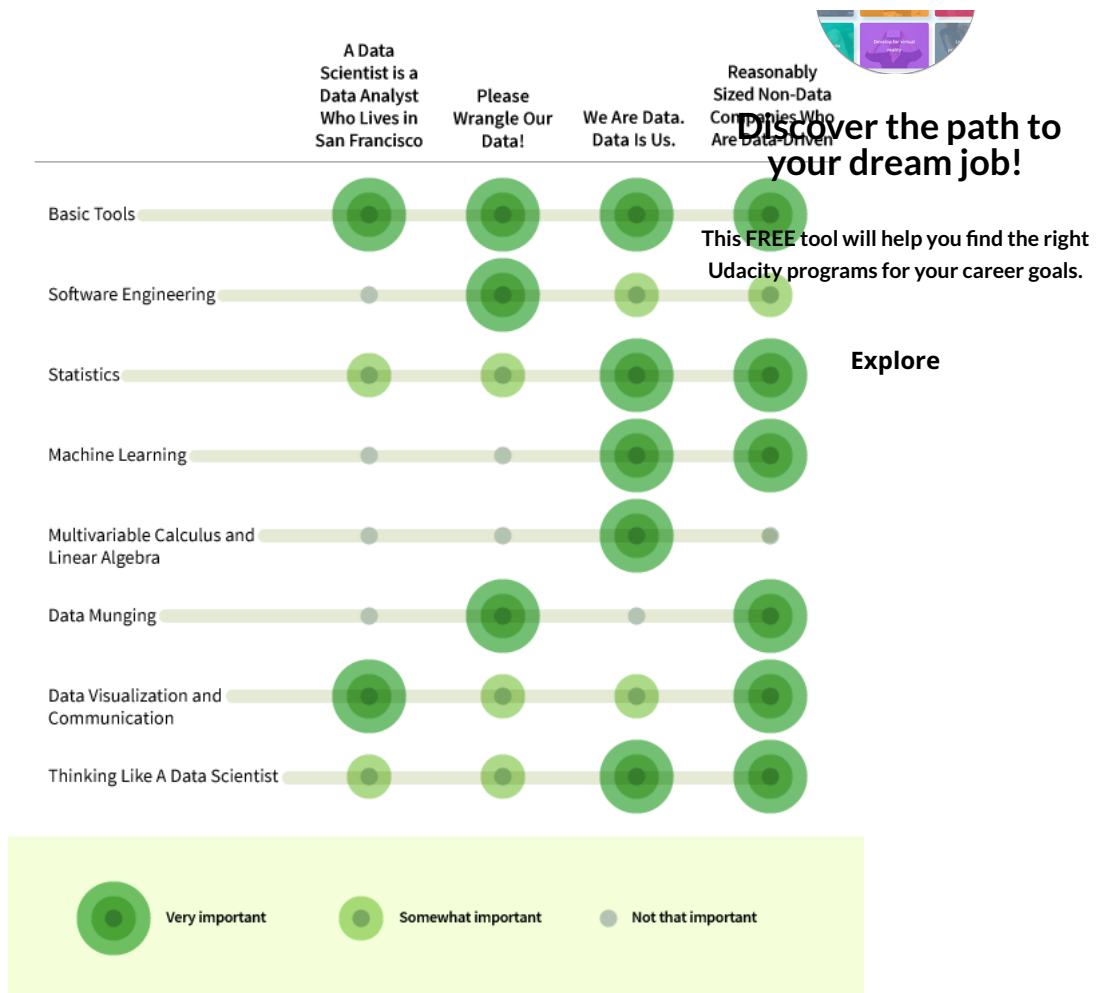
Hopefully this gives you a sense of just how broad the title 'data scientist' is. Each of the four company 'personalities' above are seeking different skillsets, expertise, and experience levels. Despite that, all of these job postings would likely say "Data Scientist," so look closely at the job description for a sense of what kind of team you'll join and what skills to develop.

	A Data Scientist is a Data Analyst Who Lives in San Francisco	Please Wrangle Our Data!	We Are Data. Data Is Us.	Reasonably Sized Non-Data Companies Who Are Data-Driven
Basic Tools	Very important	Very important	Very important	Very important
Software Engineering	Not that important	Very important	Somewhat important	Somewhat important
Statistics	Somewhat important	Somewhat important	Very important	Very important
Machine Learning	Not that important	Not that important	Very important	Very important
Multivariable Calculus and Linear Algebra	Not that important	Not that important	Very important	Not that important
Data Munging	Not that important	Very important	Not that important	Very important
Data Visualization and Communication	Very important	Somewhat important	Somewhat important	Very important
Thinking Like A Data Scientist	Somewhat important	Somewhat important	Very important	Very important

Legend: ● Very important ● Somewhat important ● Not that important

(http://1onjea25cyhx3uvxgs4vu325.wpengine.netdna-cdn.com/wp-content/uploads/2014/11/blog_dataChart_white.png)

4 types of data science jobs with a breakdown of the 8 skills you need to get the job

8 Data Skills to Get You Hired

This is the core set of 8 data science competencies you should develop:

Basic Tools: No matter what type of company you're interviewing for, you're likely going to be expected to know how to use the tools of the trade. This means a statistical programming language, like R or Python, and a database querying language like SQL.

Basic Statistics: At least a basic understanding of statistics is vital as a data scientist. An interviewer once told me that many of the people he interviewed couldn't even provide the correct definition of a p-value. You should be familiar with statistical tests, distributions, maximum likelihood estimators, etc. Think back to your basic stats class! This will also be the case for machine learning, but one of the more important aspects of your statistics knowledge will be understanding when different techniques are (or aren't) a valid approach. Statistics is important at all company types, but especially data-driven companies where the product is not data-focused and product stakeholders will depend on your help to make decisions and design / evaluate experiments.

Machine Learning: If you're at a large company with huge amounts of data, or working at a company where the product itself is especially data-driven, it may be the case that you'll want to be familiar with machine learning methods. This can mean things like k-nearest neighbors, random forests, ensemble methods – all of the machine learning buzzwords. It's true that a lot of these techniques can be implemented using R or Python libraries – because of this, it's not necessarily a dealbreaker if you're not the world's leading expert on how the algorithms work. More important is to understand the broadstrokes and really understand when it is appropriate to use different techniques.

Multivariable Calculus and Linear Algebra: You may in fact be asked to derive some of the machine learning or statistics results you employ elsewhere in your interview. Even if you're not, your interviewer may ask you some basic multivariable calculus or linear algebra questions, since they form the basis of a lot of these techniques. You may wonder why a data scientist would need to understand this

stuff if there are a bunch of out of the box implementations in sklearn or R. The answer is that at a certain point it become worth it for a data science team to build out their own implementations in house. Understanding these concepts is most important at companies where the product is defined by the data and small improvements in predictive performance or algorithm optimization can lead to huge wins for the company.

"Data scientist" is often used as a blanket title to describe jobs that are drastically different.

(hp://🐦.com/nn/🐦?ul=hp://blog.udacy.com/2014/11/daa-scnc-job-

sklls.hml&x=%E2%80%9CDaa%20scns%E2%80%9D%20s%20ofn%20usd%20as%20a%20blank%20l%20o%20d

Data Munging: Often times, the data you're analyzing is going to be messy and difficult to work with. Because of this, it's really important to know how to deal with imperfections in data. Some examples of data imperfections include missing values, inconsistent string formatting (e.g., 'New York' versus 'new york' versus 'ny'), and date formatting ('2014-01-01' vs. '01/01/2014', unix time vs. timestamps, etc.). This will be most important at small companies where you're an early data hire, or data-driven companies where the product is not data-related (particularly because the latter has often grown quickly with not much attention to data cleanliness), but this skill is important for everyone to have.

Data Visualization & Communication: Visualizing and communicating data is incredibly important, especially at young companies who are making data-driven decisions for the first time or companies where data scientists are viewed as people who help others make data-driven decisions. When it comes to communicating, this means describing your findings or the way techniques work to audiences, both technical and non-technical. Visualization wise, it can be immensely helpful to be familiar with data visualization tools like ggplot and d3.js. It is important to not just be familiar with the tools necessary to visualize data, but also the principles behind visually encoding data and communicating information.

Software Engineering: If you're interviewing at a smaller company and are one of the first data science hires, it can be important to have a strong software engineering background. You'll be responsible for handling a lot of data logging, and potentially the development of data-driven products.

Thinking Like A Data Scientist: Companies want to see that you're a (data-driven) problem solver. That is, at some point during your interview process, you'll probably be asked about some high level problem – for example, about a test the company may want to run or a data-driven product it may want to develop. It's important to think about what things are important, and what things aren't. How should you, as the data scientist, interact with the engineers and product managers? What methods should you use? When do approximations make sense?

Data science is still nascent and ill-defined as a field. Getting a job is as much about finding a company whose needs match your skills as it is developing those skills. This writing is based on my own firsthand experiences – I'd love to hear if you've had similar (or contrasting) experiences during your own process.

Note: Dave (https://twitter.com/daveholtz) is a data scientist at Airbnb. If you're curious about how the data science interview process works at Airbnb, Riley Newman (head of data science at Airbnb) has written a great post (http://www.quora.com/Airbnb/How-does-Airbnb-hire-data-scientists) on Airbnb's data science hiring process on Quora.

Discover the path to your dream job!

(http://blog.udacity.com/data-

This FREE tool will help you find the right Udacity programs for your career goals.

Explore

Ultimate Data Analyst Skills Checklist

Curious about getting started as a Data Analyst?

Get the Checklist

DOWNLOAD (HTTP://1ONJEA25CYHX3UVXGS4VU325.WPENGINE.NETDNA-CDN.COM/WP-CONTENT/UPLOADS/2014/12/UDACITYULTIMATESKILLCHECKLISTFORYOURFIRSTDATAANALYSTJOB.PDF)

ENROLL NOW

(https://www.udacity.com/nanodegree?utm_source=blog&utm_medium=referral&utm_campaign=sb_widget_enroll)

Popular Nanodegree Programs

Data Analyst (https://www.udacity.com/course/data-analyst-nanodegree--nd002)

Android Basics Nanodegree by Google (https://www.udacity.com/course/android-basics-nanodegree-by-google--nd803)

Front-End Web Developer Nanodegree (https://www.udacity.com/course/front-end-web-developer-nanodegree--nd001)

iOS Developer (https://www.udacity.com/course/ios-developer-nanodegree--nd003)

Machine Learning Engineer (https://www.udacity.com/course/machine-learning-engineer-nanodegree--nd009)

Android Developer (https://www.udacity.com/course/android-developer-nanodegree--nd801)

Senior Web Developer (https://www.udacity.com/course/senior-web-developer--nd802)

Latest Posts

Introducing Siraj Raval's Deep Learning Nanodegree Foundation Program! (http://blog.udacity.com/2017/01/siraj-raval-deep-learning-nanodegree-foundation-program.html)

Exploring the Jobs of Today, Tomorrow, and Beyond at Udacity's Intersect 2017 Conference (http://blog.udacity.com/2017/01/exploring-the-jobs-of-today-tomorrow-and-beyond-at-udacitys-intersect-2017-conference.html)

Udacity 2016: The Year in Review (http://blog.udacity.com/2016/12/year-in-review-2016.html)

A Wall Of Student Success (http://blog.udacity.com/2016/12/wall-student-success.html)

Switching Careers: Advice from a Nanodegree Graduate (http://blog.udacity.com/2016/12/switching-careers-advice-nanodegree-graduate.html)

SUBSCRIBE NOW

Dave Holtz

POSTS BY DAVE

Discover the path to your dream job!

This FREE tool will help you find the right Udacity programs for your career goals.

Explore

Share This Post

TWITTER

(HTTPS://TWITTER.COM/INTENT/TWEET?URL=HTTP://BLOG.UDACITY.COM/2014/11/DATA-SCIENCE-JOB-SKILLS.HTML&VIA=UDACITY&TEXT=8 SKILLS YOU NEED TO BE A DATA SCIENTIST)

FACEBOOK

(HTTP://WWW.FACEBOOK.COM/SHARER/SHARER.PHP?U=HTTP://BLOG.UDACITY.COM/2014/11/DATA-SCIENCE-JOB-SKILLS.HTML)

GOOGLE+

(HTTPS://PLUS.GOOGLE.COM/SHARE?URL=HTTP://BLOG.UDACITY.COM/2014/11/DATA-SCIENCE-JOB-SKILLS.HTML)

LINKEDIN

(HTTPS://WWW.LINKEDIN.COM/SHAREARTICLE?MINI=TRUE&URL=HTTP://BLOG.UDACITY.COM/2014/11/DATA-SCIENCE-JOB-SKILLS.HTML&TITLE=8 SKILLS YOU NEED TO BE A DATA SCIENTIST)

PREVIOUS POST

← **Course Announcement for iOS Developer Nanodegree: Intro to iOS App Dev with Swift (http://blog.udacity.com/2014/11/course-announcement-for-ios-developer.html)**

NEXT POST

Announcing the Platypus... Data Analyst Nanodegree (http://blog.udacity.com/2014/11/announcing-platypus-data-analyst.html) →

POPULAR NANODEGREE PROGRAMS   +

STUDENT RESOURCES   +

UDACITY   +

INQUIRIES   +

✖