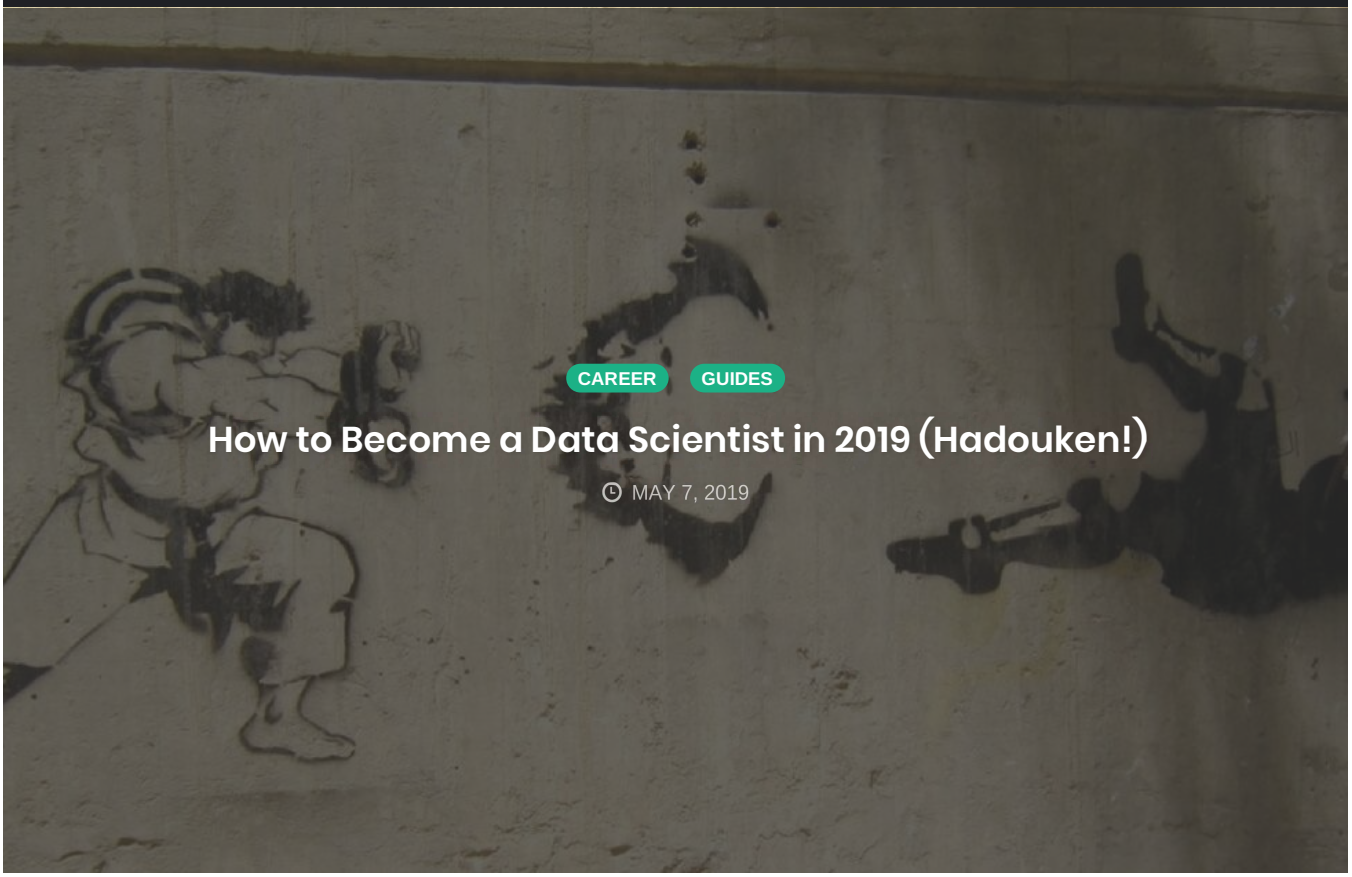
EliteDataScience

<<

Shorten the Learning Curve!
Download 10+ Bonus Resources for Data Science & Machine Learning

GET INSTANT ACCESS!



CAREER GUIDES

How to Become a Data Scientist in 2019 (Hadouken!)

MAY 7, 2019

586
SHARES

Follow me on linkedIn for more:
Steve Nouri
https://www.linkedin.com/in/stevenouri/

So you want to become a data scientist… that's fantastic! But as you may already know (or may soon find out), it's not quite that simple.

In fact, you'll most likely face some challenges that are unique to data science…

Challenge #1: WTF is a "data scientist?"

You could ask 10 data scientists and get 15 descriptions of what they do. The term "data scientist" is nowhere near as well defined as… say… "accountant" or "web developer."

It's a relatively young discipline, so different employers disagree on what a data scientist should be doing. And that brings us to the next challenge…

Challenge #2: Do I really need all that?

Many “how to become a data scientist” articles begin by listing a huge collection of skills, software, and concepts you’ll supposedly need to master…

Spark! Hadoop! Hive! Python! R! Storm! SQL! NoSQL! MaybeSQL?

…You’d think they were shouting moves from a street fighter game… *Hadouken!*

As a result, candidates often feel overwhelmed by too many things to learn, not knowing where to start. Then, after they do start, they’ll feel very busy but often won’t know if they’re making real progress.

In reality, most positions only expect you to have a handful of key skills, but those key skills differ from industry to industry and from employer to employer. For example, some data scientists never touch the Big Data tech stack while others use it every day. Candidates feel overwhelmed when they try to prepare for everything.

Therefore, rather than providing you a static list of skills and saying “go learn this and come back when you’re done,” we’d like to present a systematic approach to designing your own personalized roadmap.

Solution: Flip the order of operations.

The conventional order of operations is (1) start studying and learning skills, (2) write your resume, and then (3) search for jobs. This leads to the challenges described above.

Instead, let’s flip that process on its head. We’ll start with the job search itself in order to obtain a concrete target and a sense of direction.

1. Choose ONE industry.

In his book *The ONE Thing: The Surprisingly Simple Truth Behind Extraordinary Results*, Gary Keller attributes his success in building one of the world’s largest real estate companies to his habit of prioritizing a single task at a time. He focuses on that task until completion instead of attempting to multitask.

We can adopt the same mindset here. Data science is never done in a vacuum, so each industry requires different skills, programming languages, and qualifications.

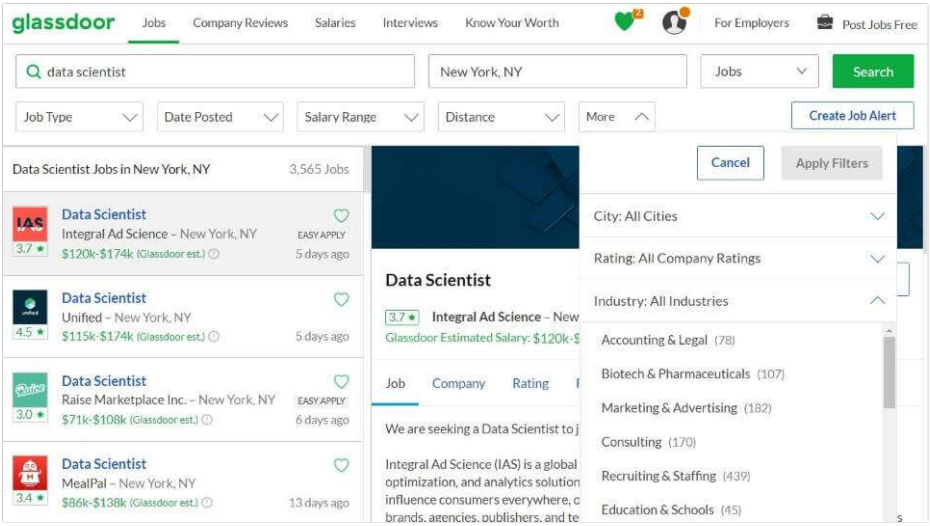
Limiting your initial search to ONE industry has many benefits. Not only will this **reduce the number of topics to study**, but it will also allow you to start building invaluable **domain knowledge** and adding **relevant portfolio projects.** These will give you a huge edge during the interview process.

Based on Glassdoor.com’s classifications, some popular industries for data scientists include:

- Biotech & Pharmaceuticals
- Marketing & Advertising
- Banking & Financial Services
- Internet & Tech
- Media & Publishing
- And more…

Note that the number of opportunities in each industry will vary by city. For example, at the time of this writing, San Francisco had 269 *Internet & Tech* listings and only 116 *Banking & Financial Services* listings while New York had 166 *Internet & Tech* listings and a whopping 595 *Banking & Financial Services* listings.

Therefore, we recommend going to **Glassdoor.com > Jobs** and then searching for data scientist positions in several cities you'd like to work in, then clicking the **More > Industry** dropdown to find a list of industries in those cities.



Narrowing by industry on Glassdoor.com

2. Pinpoint 5 target positions.

Next, go to a job board such as Glassdoor, LinkedIn Jobs, or Indeed and search for data science positions in your chosen industry. Don't just limit your search to "data scientist." Try other terms such as *data analyst*, *machine learning engineer*, or *quantitative analyst*.

As you'll discover, the problem is that there are *too many* options rather than too few, so we're going to eliminate many of them. Start reading through listings and try to get a qualitative feel of the work. Which software would you be using? What types of analyses would you be performing? Who would you be working with?

As you read through them, eliminate ones that:

- **(A) Aren't positions you'd be excited about.** This sounds obvious, but many people fall into the trap of "that job sounds OK enough." The "OK enough" mindset won't provide the necessary motivation when it really comes time to grind it out during preparation phase. Cut out that noise and tune into the signal from the positions you'd be really thrilled to have.
- **(B) Have requirements that are unrealistic to obtain within your target time frame.** For example, if a position requires a PhD and you don't have one, it's probably unrealistic to target that position without going back to school first. Of course going back to school is an option, but there are also plenty of excellent data scientist positions that don't require an advanced degree.

Now, if you're still months away from applying, you may be thinking that searching for target positions right now will be a waste of time. After all, won't these positions be filled by then?

Well, yes, these specific positions will most likely be filled by then, but that's not the point of this step. The point of this step is to set us up to define **concrete targets.** This step will help you identify the requirements for *your own* ideal data scientist position.

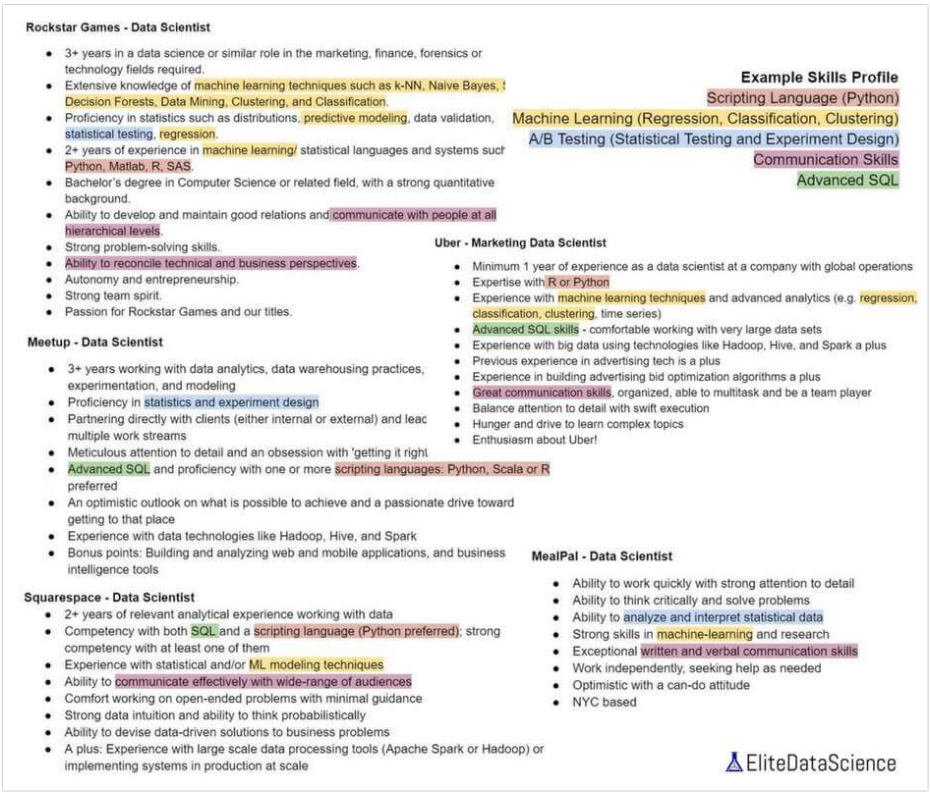
Once you've found your 5 target positions, download and save their complete job descriptions. We'll need them in the next step.

3. Create a "skills profile."

In the previous step, we gained a qualitative understanding of our target positions. Now, we'll distill the useful and actionable information into a "skills profile."

Look at the responsibilities and requirements for each position and try to pick out the ones that appear repeatedly. A good rule of thumb is to write down any skills that appear in at least 3 of the 5 target job descriptions.

Here's an example skills profile that we compiled from 5 data scientist positions in tech (the screenshot only shows the requirements portions of the job descriptions):



Example skills profile

As you can see, the skills that show up in at least 3 of the 5 target positions include:

- **Scripting Language** (Python)
- **Machine Learning** (Regression, Classification, Clustering)
- **A/B Testing** (Statistical Testing and Experiment Design)
- **Communication Skills**
- **Advanced SQL**

Now we're talking! We went from potentially dozens of skills, topics, and software down to just 5. Some of these skills have sub-skills and sub-concepts, but for study and preparation purposes, we can treat each of these as a single skill bundle.

Sure, there were a few “nice-to-haves” from the job descriptions that we left out, but that’s fine for now. The skills profile allows us to focus on bolstering the skills that will give us the biggest bang for our buck.

Just remember that successful candidates are rarely *fully* qualified. Most of the time, you only need to have ~60-80% of the qualifications in order to have a realistic shot of landing a job. Employers understand that most will still need to learn more on the job.

4. Write your "future resume."

In the 5th inning of Game 3 of the 1932 World Series, baseball legend Babe Ruth walked up to the mound at Wrigley Field and pointed his finger at the center-field bleachers. *I will hit a home run there…*

The next pitch comes in and… CRACK! The bat connects, sending the ball soaring exactly where he was pointing moments ago. That “called shot” was one of the most famous home runs in history.

Babe Ruth's Called Shot

This is basically what we’re going to do next. Based on the skills profile from the previous step, we will write our future resume, and it’s going to be impressive yet realistic.

Pretend you’re applying to those 5 target positions tomorrow, but you can write your resume as it would look 3-6 months later. How would the best version of yourself as a data scientist candidate look?

Here are a few ideas for what to include:

- If your resume has a ‘Skills’ section, go ahead and list the ones from the skills profile.
- If you’re currently working, are there any projects you can join that would give you relevant skills or experience? If so, include them.
- If you’re currently in school, are there any classes you can take that would give you relevant skills or experience? If so, include them.
- Are there any side projects you’d like to tackle and add to your portfolio? If so, include them. Bonus points for projects related to the industry you’d like to join.

Tip: When writing your future resume, you can use a different font color for clarity.

If you ever start feeling overwhelmed or pulled in too many directions, return to your future resume and the skills profile to re-center yourself. Your future resume will give you a concrete goal, and it can become a self-fulfilling prophecy.

Resources:

- **How to Write the Perfect Data Scientist Resume**

5. Start studying and practicing.

Now that we have our skills profile and future resume, we're finally ready to start learning, studying, and filling in any gaps. In other words, it's time to achieve that future resume.

Other guides often start with this step, which is like hopping into a car and just taking off in the general direction of your destination. Instead, we've opted to **install a GPS first**, so we'll get there faster and more reliably.

The process here is simple and iterative:

1. **Pick one of the skills/concepts you're missing.** If programming is one of them, then we recommend starting with that. The ability to program, especially in Python or R, will allow you learn other concepts faster because you'll be able to actually implement them and learn by doing.
2. **Block out X weeks to absorb everything you can about that skill/topic.** X is a number that you set, depending on your personal situation. The key is that you must set X in advance, which will give you an actual deadline… pretend you have a test coming up in X weeks. Parkinson's Law states that *"work expands so as to fill the time available for its completion,"* which we've found to be especially true during self-study. Self-imposed deadlines can reduce stress and overwhelm because they provide concrete milestones to aim for.
3. **Mix in plenty of hands-on practice.** For example, if you're learning SQL for this block, then grab a dataset, import it into a database server, and practice writing queries as you learn about JOIN, GROUP BY, and so on.
4. **Rinse and repeat (1) – (3).** Make your way through the skills profile you created in Step 3. Of course, you can always circle back if you feel like you haven't learned enough about a particular topic after your X week block.

Some candidates may need 6 months or more to learn everything, while others may only need to brush up for a week or so. It all depends on how well your existing skillset transfers to data science.

Treat this as a **long-term investment in yourself**, and don't try to rush because you're afraid of "missing out" on opportunities. Take the time you need, as there will always be more opportunities once you are ready for them.

Tip: We recommend setting consistent study blocks every week. Treat those blocks as a class you can't skip.

Resources:

- **Free Data Science Study Resources for Beginners**
- **Overview of Modern Machine Learning Algorithms**

6. Weave in end-to-end projects.

While you study, attempt an end-to-end project every other weekend. Start with a real-world dataset, pose an interesting question, and then try to answer it on your own.

That may include:

- Cleaning the data
- Wrangling it into a new format
- Engineering features
- Training a model with machine learning
- Creating visualizations
- And/or running hypothesis tests

Allow these projects to serve as **barometers for your progress.** For example, for your first attempt, you may discover that you don't even know how to structure a project or where to find data. That's OK! These projects are meant to teach you what you don't know.

As former U.S. Secretary of Defense Donald Rumsfeld once said:

> *"There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also **unknown unknowns.** These are things we don't know we don't know."*

These **end-to-end projects reveal those unknown unknowns**, turning them into known unknowns (thus allowing you to address them and make them known knowns).

Plus, projects keep you motivated, help you solidify knowledge, and look impressive on your resume. They will also give you great talking points for interviews down the road.

As you can tell, we absolutely love projects as a learning tool, and we firmly believe that they are the best way to prepare for a job in data science.

Resources:

- **Datasets for Data Science and Machine Learning**
- **Fun Machine Learning Projects for Beginners**

7. Hit your milestones, and then just apply!

The limbo between study mode and the formal job hunt causes many candidates to stagnate. How do you know you've prepared enough? Are you ready? Have you missed anything?

These concerns are common and reasonable, but there's an easy way around them: pick concrete milestones *beforehand*… and once you hit them, just start applying. We recommend **project-based milestones.**

For example, as soon as you complete **5 end-to-end projects** that you're happy with, polish them up, update your resume (i.e. modify or confirm the "future resume" from Step 4), and then fire off a few applications.

You may not feel fully ready yet, but that's fine! It's rare for any candidate to ever feel 100% ready, so the key is to maintain momentum and keep learning on-the-go.

The interview process will be a new challenge unto itself. By starting to apply as soon as you've hit your milestones, you can ease into the job search and get more opportunities to practice.

8. Prepare for the interview.

With enough prep time, Batman can beat anyone despite not having any super powers. Let's "batman" the interview process by starting early.

Many top companies have at least 3 rounds of interviews:

Round #1 – Phone Screen

This is typically an interview with HR, but you could be asked concept questions to screen your understanding of data science and machine learning.

To prepare for the phone screen, practice (but don't memorize) your responses to common interview questions. In addition, review key concepts in machine learning, A/B testing, or whichever other core skills you'll need for the position.

Round #2 – Take-Home Challenge

These are analytical challenges with datasets where you'll have ~24-48 hours to answer multiple objectives.

The best way to prepare for take-home challenges is to continue completing end-to-end projects, as they ensure you cover all the bases. After finishing a project, also practice revising/refactoring your code to make it clean, concise, and well commented.

To complete analytical take-home challenges, you'll need to be prepared for *Project Scoping & Planning*, *Exploratory Analysis*, *Data Cleaning*, *Feature Engineering*, *Modeling (Regression, Classification, Clustering)*, *A/B Testing*, and *Communication, Visualization, & Writing.*

Round #3 – Onsite "Super Day"

This is usually a day filled with analytical case questions, SQL coding challenges, technical interviews, and behavioral interviews.

At the least, you can expect more of the previous rounds' challenges, both in quantity and in difficulty. In addition, you might see new interview formats depending on the employer, such as consulting-like case questions or SQL exercises.

The good news is that the work you've done up to this point should've already given you a huge head start in terms of preparation. It's now just a matter of pushing through the final stretch.

Resources:

- **9 Mistakes to Avoid When Starting Your Career in Data Science**
- **Example Machine Learning Interview Concept Questions**

9. Full-court press!

The term "full-court press" comes from basketball, and it refers to when the defending team badgers their opponents throughout the whole court, instead of just near their own basket. This tactic drains energy quickly, so it's reserved only for critical moments.

Full-Court Press

Once you're the thick of the job search, we recommend entering a full-court press. Cut out as many distractions as possible and really ramp up effort.

Study. Apply. Interview. Study. Apply. Interview. Study. Apply…

At the end of the day, data scientist positions are competitive, but it's still just a numbers game. The best way to maintain momentum, especially through setbacks or rejections, is to keep your pipeline full of opportunities you're excited about.

It won't necessarily be easy, but you'll get there if you persist.

"Ambition is the path to success. Persistence is the vehicle you arrive in." ~ Bill Bradley

586
SHARES

« Previous Post
15 Ways Machine Learning Will Impact Your Everyday Life

Next Post »
How to Become a Data Scientist in 2019 (Hadouken!)

Free: Data Science Downloads

10+ Bonus Resources for Data Science & Machine Learning

GET INSTANT ACCESS!

L I N K S

Start Here
Homepage
Login

A R T I C L E S

Guides

Follow me on linkedIn for more:
Steve Nouri
https://www.linkedin.com/in/stevenouri/