Six qualities of a great data scientist

What does it take to be a great data scientist? We thought long and hard about this while we were designing the [data science bootcamp for Metis](). Some of these you undoubtedly know without looking, but others may surprise you. These six hard and soft skills are not just things to look for in a data science hire, but are also things to cultivate in yourself to prepare for a successful and fulfilling career in data science. We made this list to help identify ideal students for our data science bootcamp, people who will make great students and great hires.

1. Statistical thinking

Data scientists are professionals who turn data into information, so statistical know-how is at the forefront of our toolkit. Knowing your algorithms and how and when to apply them is arguably the central task to a data scientist's work. However, to do this well can be an art and a science.

A good data scientist can model any data he is given and implement a toolbox full of algorithms to make statistically-informed predictions and recommendations. A great data scientist can smell something 'fishy' in the results she gets, senses that he needs to ask the client or stakeholder a few more questions before retreating to the code cave, and can make the difference between a game-changing insight and an expensive blind hunch.

Cultivating statistical thinking: Quantitative thinking sinks in the fastest when it is relevant to you. At Datascope, we keep our statistical thinking sharp by calling each other out on our bold claims, sometimes making friendly bets, and figuring out how to resolve it with statistics. For example, here's one you can try on your friends: how many wikipedia articles contain the word "the"? All of them? Nearly all of them? We put the over-under at 90%, wrote a script to calculate the wikipedia document frequency of the word "the", and shocked ourselves to find out the word "the" only appears in 85% of wikipedia articles.

2. Technical acumen

Data scientists write code and work with teams to produce tools, pipelines, packages, modules, features, dashboards, websites, and more. We write code on the back end and the front end. We do structured and unstructured. We sift through unfamiliar formats and legacy code, and "roll our own" tools when we can't find the solution we need.

A great data scientist has a hacker's spirit. Technical flexibility is as important as experience, because in this field the gold standards change with an alarming rate. Data scientists work together, love open source, and share our knowledge and experience to make sure that we can move at the speed of demand. If your data scientist is a quick study, you've made a sound investment beyond the current trend cycle.

Cultivating technical acumen: Write code, every day if possible. Learn about the tools you want to use, but don't just read about them, try them. Follow a tutorial. Change it and see what happens. Break it.

Check out someone's projects that are written in a language you don't know. If you read about a new tool or service that interests you, start by just making a "hello, world". Work in small bites.

3. Multi-modal communication skills

When the analysis is finished running, most of the time the results aren't pretty. That's not to say they are unhelpful, but they are often trapped in opaque readouts, or in plots that are sensical to the expert's eye but hieroglyphics to the rest of the team and stakeholders. Algorithmic output has to be interpreted and communicated to make the leap out of the data science team and into the hands of the rest of the company to be put to service in alignment with their usefulness.

A great data scientist can contextualize and translate a problem and its solution to interested parties of wildly varying backgrounds using common ground, metaphor, skillful listening, and storytelling. This includes the written communication that goes into a statement of work or a report, visual communication for clear and intuitive plots and visualization, and spoken communication for presentations, project specifications, check-in meetings, and iterative design. If your data scientist can stop a meeting when it's clear that not everyone is on the same page, draw a sketch on the whiteboard and elicit consensus from a diverse team, you have a deeply valuable person on your team.

Cultivating multi-modal communication: Practice writing and talking about your technical projects to "normal" people. Edit yourself down to the important parts. Learning to edit yourself is key. Practice visual communication with "crayon wireframes"-- we use these all the time, and they help us to think visually and iterate quickly. Sketches are also a great way to gut check if everyone is on the same page. If the words you're saying seem to match each other, but the pictures don't jive, you may have nipped some future problems in the bud.

4. Curiosity

So your data scientist is a technical and statistical whiz who can explain a Markov chain to a supermarket checker. What else separates the elite? The first of our trio fo invaluable soft skills is curiosity. Many who are drawn to data science find most alluring the opportunity to work on a constant stream of new and challenging puzzles. They are people who have been asking "why" and "how" since their mouths could form the words.

A good data scientist will take a request, implement it, and deliver the prediction or analysis with confidence. A great data scientist will come back asking for access to more data, or to interview users, or to try something new in the next iteration, because something he did triggered that curious itch. Curious data scientists might have a disdain for machine learning competitions because they can't access all of the levers and choice points to ask questions and dig deeper. Masters of curiosity are quick to question their own assumptions.

Cultivating curiosity: Are there stupid questions? Actually, probably… yes. But who cares? In our experience they are largely indistinguishable from really great questions that no one has bothered to ask yet. So we don't discriminate, and we ask many many questions that may or may not be stupid. Ask enough stupid questions and eventually you'll land on something brilliant. Seek out people different than you to talk through ideas. Let your mind wander. Raise your hand and ask questions. And never assume that the experts see everything that you do -- don't ignore that inner voice when it's telling you something isn't adding up. You might ask a stupid question (and learn), or you might ask a brilliant one (and discover).

5. Creativity

Creativity goes far beyond the obvious applications in communication and project design. Of course, a data scientist who can create an attractive and easy-to-grasp report or visual out of results that would take a couple of master's degrees to fully understand is a skill with enormous returns. Creativity provides the fuel for skillful communication, and that is not a hard sell.

Beyond aesthetics and communication, however, the best data scientists are creative problem solvers and have a peculiar relationship with the word "no." Your data scientist really wants to include those user-level data sets in the algorithm, but they are locked down in another silo within the company? She figures out a way to model their effects from the population statistics, or she generates a simulated report using dummy data to convince the c-suite that building a bridge between departments is well worth the risk or effort. The client wants to know how much foot traffic their potential new outlets will see, next Friday, and those data don't exist or aren't available? He uses publicly available transit data to build informed estimates and proposes a small and inexpensive sample-gathering project to build a turnstile-count-to-total-pedestrians conversion heuristic.

Great data scientists get annoyed by "no," so annoyed that they find a way to get around it, over it, or through it, or they back up and take another path altogether. Design constraints are infuriating and intoxicating. A data scientist who says "no," followed quickly by "wait, hold on… let me think," just might be a great creative thinker.

Cultivating creativity: When you have an idea, do it, and let go of judgment. Foster an environment that invites questioning, thinking outside the box, and a "yes, but" conversational atmosphere. Creativity is ingrained in culture: give people around you the safety to "fail," without judgement or criticism, and they'll likely grant you the same. When you're surrounded by people thinking and working creatively, you'll find your own creativity flourishing, too.

6. Grit

In all of the characteristics and skills above, we've talked about technologies that appear, demand attention, and fade faster than songs on the top forty, messy data that should but don't quite fit into the model the client had in mind, dead ends, wrong turns, roadblocks, and red tape, teams full of mixed

agendas and personalities; budgets, deadlines, clients, teams, and contractors; and the unicorn-magician-programmer-statistician who is supposed to bring it all together. Those who survive have a healthy inner store of grit.

All of these challenges and roadblocks could drive a reasonable person to an unplanned leave of absence. Grit is that inner drive that pulls us over obstacles, recasts setbacks as design constraints, propels us through fear of failure, keeps us walking through actual failure, helps us to resist the impulse to take things personally, and brushes that dirt off our shoulders. When grit is working, we're less competitive so we can encourage and learn from each other. We get the taste for tackling the new and the unknown.

Cultivating grit: Let go of perfectionism. We encourage ourselves to think of quick and dirty ways we can resolve problems or test solutions. We share our early efforts and listen to the feedback we get. Keep in mind that every expert has to start as a beginner, and a beginner trying to masquerade as an expert fools no one but himself. Don't be afraid to let your warts show -- what seems like a weakness, handled properly, is a powerful tool for leadership and growth.