**Hiring Exceptional Data Scientists**

**The outliers: going beyond just technical skill**

With the waning of the initial excitement surrounding applied data science, it is pertinent to reflect on the factors that contributed to the success of some initiatives during the peak of the hype cycle. It is worth considering what was instrumental in surpassing the mostly unrealistic expectations set during this period. However, the question at hand may be better framed as “who?” rather than “what?”. In this post, I share my observations and thoughts on what I think makes an **exceptional** data scientist — those that kept impressing and demonstrating business impact with their all-round skill and endless passion.

**Active self-learner**

An exceptional data scientists will demonstrate a deep sense of curiosity and a keen drive for ongoing learning. In my experience, most data scientists are faced with problems that require novel solutions that are not the run-of-the-mill supervised or unsupervised methods. Solutions for such problems require active research on topics not always directly relevant to the problem at hand, exploring online resources, keeping up-to-date with the latest from the open-source communities, reading books, engaging in conversations with peers, or attending data science meetups where they get to externalise ideas. This type of individual may also contribute to the field by developing tools or algorithms (although not necessarily enrolled in a postgraduate degree). They will also engage with academia to keep up-to-date with research and substantiate their own learning, which is an invaluable resource in my opinion.

Self-learning data scientists can critically assess new information and associate it to their ongoing projects or those of their peers. Furthermore, they exhibit an open-minded attitude and remain receptive to new ideas and perspectives, and possess a willingness to evaluate and challenge their own assumptions as they continue to learn and expand their skills. Such individuals exhibit a high degree of self-efficacy, which enables them to be highly motivated to learn and persevere. I’ve spoken to exceptional data scientists that also practice meta-cognition. They will reflect on their own cognitive processes and learning methods, which is highly beneficial in regulating and directing their own thoughts and knowledge towards a specific purpose. In my view, this contributes towards enhancing their learning efficacy and their ability to problem-solve more effectively.

How I would recognise such individuals is by looking at their public repositories where they implemented difficult theoretical concepts or algorithms derived from scientific papers. How did they approach the problem at hand? What literature and resources did they refer to? How did they ultimately develop a viable solution? Such insights can provide a useful understanding of their thinking process and approach towards problem-solving. I would also look out for those data scientists that use the learning-by-teaching method where teaching others help reinforce their own understanding and knowledge on the subject matter. This type of candidate is highly beneficial for a team, since they are open to share their learnings! As part of my interview process, I ask candidates about the methods or algorithms that they would like to learn (or better understand) and why. What is on their wish-list?

**Effective communicator**

Exceptional data scientists have a vocabulary and communication style that allows them to clearly articulate both business and technical information, taking into account the needs and perspectives of their audience*.* They may use active listening skills such as restating or summarising what a speaker has said and ask follow-up questions. They are able to put themselves in their audience’s shoes demonstrating empathy. This is often necessary when data scientists need to explain difficult concepts to non-technical business stakeholders. They also have belief in their own ability, but also able to convey confidently where they have uncertainties. They can keep an audience engaged during presentations and use storytelling to make their message memorable and impactful. They are great collaborators with product owners, business stakeholders and peers and are aware of group dynamics, which is crucial in recognising potential obstacles and opportunities for group success. They exhibit trust in others and are adept at building robust working relationships, maintaining a high level of interdependence within the group and resolving conflict in a positive and constructive manner. This also enhances group cohesion and facilitates successful outcomes.

**Creative problem-solver**

Exceptional data scientists possess a high level of fluid intelligence, which enables them to solve novel problems, think abstractly, and adapt to new situations. This facilitates their inventiveness and grasp of new concepts, often leading to the generation of new business ideas (an entrepreneurial mindset if you will) that can be transformed into new practical products or solutions. Some may display divergent thinking, capable of generating a wide range of ideas in response to a challenge or problem, which involves thinking beyond conventional approaches, making novel connections and considering multiple perspectives. They are imaginative, curious, and open to new experiences.

To identify such individuals, I would pay attention to their cognitive flexibility when asking them to come up with different approaches to a problem. The focus here is not necessarily for the candidate to come up with correct solutions, but rather to assess their capacity for original thinking by connecting seemingly unrelated concepts to generate meaningful associations relevant to solving the problem.

**Domain professional**

Apart from their technical skill, some may possess considerable knowledge and expertise in a particular field or industry that is pertinent to your recruitment objectives. Distinguished data scientists may have a good understanding of business (e.g., business strategy, objectives, etc.), which help them identify and work autonomously on core problems related to business. The same goes for other fields (e.g., engineering, manufacturing, finance, etc.). These data scientists have an advantage in using their domain knowledge in understanding the nuances of the data they are working with and to develop models or insights that are relevant and actionable. They would also know what model assumptions represent the underlying dynamics of the domain best, which can be gauged during interviews. Hiring domain professionals can accelerate data science initiatives in a particular domain.

**Technical proficiency**

Here are some key technical areas that exceptional data scientists take seriously.

**Cogentmodel assumptions:** The ability to identify and explain underlying logic, principles and assumptions mathematically, using coherent arguments and evidence to support their position. The exceptional data scientist can explain their tools and identify the ideal scenarios to leverage them.

**Systems orientated reasoning:** They will approach problems by considering the larger end-to-end system in which they are embedding a solution. They meticulously analyse the inter-relationships among various data and model components of a system, taking into account how these components are being consumed or interacted with. They also consider the impact of changes in one part of the system on the system as a whole and are interested in understanding the underlying structure and feedback loops that give rise to system behaviour. I’ve also observed exceptional data scientists working closely with end-users of the system to leverage user-experience feedback.

**Metrics driven:** They place a strong emphasis on using data and quantitative measures (appropriate statistics) to drive decision-making and to evaluate performance. They are focused on tracking and analysing KPIs and using that information to make informed decisions to improve models or systems. For example, they will know exactly what metrics to use for specific models, how to interpret them, and what cognitive biases and statistical biases (e.g., Goodhart’s law, Simpson’s paradox, Berkson’s paradox, etc.) to be aware of.

**Data orientated:** They are highly focused on the collection, analysis, and interpretation of data in order to inform decision-making and gain insights into various phenomena. The individual is comfortable working with data, is skilled in data processing, data analysis, and data visualisation. They are detail-oriented in their data exploration process and will search for explanations for anomalous discoveries — never sweeping hidden knowledge under the carpet.

**Open-source projects:** They may have their own public open-source projects or repositories where they and other contributors frequently contribute. This individual actively contributes code towards a common goal and their contributions are valued and accepted by the community.

There are different ways in which interviews can be conducted. Personally, I do not prefer assigning time limit coding tasks to data science candidates. Doing so, undermines almost everything written in this post. Instead, I would review their public code repositories to assess their coding style, contributions, and problem-solving approaches. By examining the comments within their code and analysis, one can gauge their thought process and observations. Follow-up questions during interviews related to their public code can help clarify gaps. Alternatively, asking a data science candidate to talk through an interesting project they did (with the protection of confidential information), or providing them with a predetermined case study to solve conceptually (not with code) should provide ample opportunity for a hiring team to identify most of the characteristics mentioned in this post.

P.S. Please ensure that your job requirements are realistic and representative of the actual role; otherwise, exceptional candidates will lose interest in your company. Know what practices have been successful within your present team, what your company/team culture is like, what complimentary skills your team requires, **and how you will retain and grow the exceptional data scientists after you’ve hired them!**

To end on a humours note — an exceptional data scientist is like a Kalman filter ;)

* **Adaptable:** they should be able to adapt their models and analyses as new information or data becomes available.
* **Robust:** they should be robust to noisy and uncertain data (or people). They should be able to utilise uncertainty and filter out noise and other irrelevant information to focus on what matters.
* **Prediction:** they should be able to make accurate predictions based on past data (and current observations), identify trends and patterns in the data and use them to make informed predictions/recommendations.
* **Balance accuracy and complexity:** they should find the best model to fit a given set of data and extract the most useful information from it while keeping computation time reasonable by using simple (often linear) assumptions.