**Date: Mostly 2014.**

**Dataiku - Interview with Data Scientist Matthieu Scordia**

**Q. Out of all the requirements listed, which are the most critical for a candidate to possess? Why?**  
  
As a data scientist you need to have a **real solid knowledge of machine learning and be able to code in Python or R fluently**. And if you also spend time building **compelling stories around data**, or spend all **your free time on Kaggle** for instance, you’re probably a great candidate for us. You have to like playing with data, have a bit of a hacker mindset. – Dataiku interview

***Q. How do you typically read a data science resume? (i.e., Are there sections you skim, places you focus more time etc.)***  
  
Well before we even start reading the resume, we reply to candidates by sending them a data science exercise, like a small Kaggle competition. Usually, the candidate has to make predictions on a small dataset, using the language of his choice.  
  
Only about half of the candidate send us the exercise back. Then, we especially look at how the candidate handles difficulties, what hacks he finds to avoid them. This is a great way to gouge a data scientist’s real value, before even looking at his studies or his background. After that we have him come in for a job interview to see how he could fit in with the team, as well as interact with clients.

***Q. What should candidates avoid doing on their resume?***  
  
Tell us you graduated from Stanford instead of admitting you just completed Andrew NG’s Coursera MOOC. True story, candidates have done this.

**Q. What's one thing that's really compelling about the role that doesn't necessarily come through in the job description?**  
  
We are looking for passionate people and candidates who are proactive and have lots of ideas to contribute. It’s a major part of the job to love the tool Data Science Studio and want to contribute to making it better. Our boss always says that he only recruits people that are smarter than him, from all kinds of different backgrounds, and that makes going to working every day so much fun!

***Q. What are the biggest areas of opportunity/questions you want to tackle at Dataiku?***  
  
The main opportunity for us today is to democratise data science in companies. We want to tackle problems like:

* The fact that data scientist are in such demand that they’re hard to find, and companies often don’t know how to hire these types of profile yet.
* Even once data scientists are hired, data preparation takes up SO MUCH of their time and it definitely isn’t the most fun part, or the part they bring the most value to.
* There are a lots of incredible Big data technologies that are being developed, but they’re hard to assemble and often don’t work well together. (I suggest the great technoslavia blogpost)
* Because of communication problems between different teams and different profiles (engineers vs data scientists vs marketers of managers) models are rarely deployed into production in the **end and resources go to waste**.

**Q. What tools, techniques, programming languages etc do you use on a regular basis? Has this changed at all recently?**  
  
As you can imagine, I only work with Data Science Studio, the platform we developed at Dataiku. I get to test it and suggest new features to make the jobs of data scientists like me easier. DSS is great because I get to work with all the technologies and tools I like, integrated in the software: Python, SQL, R (if you want to), as well as Hadoop and Spark. I’m a python guy myself so I love to code and analyze data in Jupyter notebooks. Recently I started to learn javascript and d3js to make cool data visualization. I love building webapps to share results and have fun with them!

**Senior Data Scientist:**

*Leveraging your experience with real world data to...*

* Derive a set of new features that will help us better understand the interplay between geography and audience features to improve our model performance.
* Discover and explore third party data sources to determine their value for improving our model performance.
* Build new data-driven products and bring them to market.

*Provide technical leadership to...*

* Mentor other data scientists in algorithms, models, tools, and products that make the team more efficient.
* Participate in planning, roadmap, and architecture discussions to help evolve our data science into revenue-generating products.
* Engage in code and model reviews to continually raise the bar on our work.
* Draw data flows and architecture designs on the white board to encourage understanding and cohesive development towards your solution.
* Meet with customers and help map business needs into product requirements.

*Using your deep knowledge of numerical and statistical packages (Pandas, Numpy, Sklearn, R) to...*

* Implement a gradient-boosting classifier to predict whether a person is likely to visit a car dealership based on the advertising signals they’ve received.
* Use a Bayesian dynamic time series model to estimate the causal impact of an advertising campaign on sales at your neighborhood grocery.
* Model the complex interactions between system architecture components to refactor and rethink key components and models in an advertising system.
* Develop algorithms to optimize the setting of every lever in our advertising infrastructure.
* Analyze data to better understand how a neighborhood’s consumption of web pages correlates with visits to a local big box store.
* Build a time series model to forecast future sales of diapers for one of our clients.
* Model the effects of environmental changes on promotion effectiveness with multiple regression.

Looking at the very first bullet point "Implement a gradient-boosting classifier" all you need to do is to add the words "how to" in front of the key technical term and search for it on Google. The first result for me was this article -> [http://tullo.ch/articles/gradient-boosted-decision-trees-primer/](https://web.archive.org/web/20170505160407/http:/tullo.ch/articles/gradient-boosted-decision-trees-primer/). If you can read the article, know every word, be able to derive all the terms, and code it up in your favorite programming language, then you will have gone beyond the "basics" and be on your way.

**Your next action to level up from a junior data scientist to a senior data scientist...**

Your next action is to first look at three senior data scientist job postings you are interested in and find the knowledge gaps you have from where you are now to where you'll need to be. This way you'll learn something new that will help you get one step closer to getting leveling up to a senior data science role.

**Resume:**

**3. Format:**  
There are several mistakes to avoid here, but at a minimum make sure you

* Keep it to one page - if you can't be succinct / convey what you need to in one page you're sending a signal (and not a positive one!) about your communication style
* Use past tense
* Keep any descriptions succinct
* Avoid color coding - it is very off-putting and often makes it harder to read
* Send it as a pdf (the best way to ensure there are no scaling issues)

**5. "Proof Points"**  
This was alluded to earlier, but as part of ensuring the relevancy of the resume - and making it very easy for a Hiring Manger to digest your value proposition - is critically important to include as many "proof points" as possible that back up what you're writing / implying in terms of your ability to do the job. Without these, it is very hard to assess the strength and credibility of your candidacy as well as your commitment level - both of which will harm your chances of landing an interview. There are many different forms / ways to do this, with the right answer very dependent on your background. However, some common "proof points" are

* Relevant thesis
* Relevant course(s) taken - and grades if available
* Independent projects completed (e.g., Kaggle, self-driven work)
* Github profile - especially with the code from the independent projects
* Technical blog

**How to take action now!**  
Pull up your latest resume iteration and take a deep breath! **This is one iteration that \*\*will\*\* help!**Review the 5 areas highlighted in this post and think about each in relation to your resume. Which apply to you? Where are you falling into a common trap? For each of the 5 areas discussed here, list out the things that could be improved in your resume. Literally write them down in a notepad (paper or digital). You now have your checklist of what to work on and can focus your valuable time accordingly!

**2. Experience:**  
Don't list everything you've ever done! Your resume is not a work history document, it is a chance to showcase relevant job and/or project experiences … so, the Starbucks gig you had in college or the sales role you had after graduation should not be on there - stick to what matters for the position you're applying to.

**1. Principles:**  
An overarching rule to keep in mind at all time is that your resume should be tailored to each position you're applying to. Just churning out the same generic resume to hundreds of positions is not going to deliver the results you want. Some people may tell you "its a numbers game", which to some extent is true, but consider the Hiring Manager's perspective - they have to read hundreds of resumes / cover letters - for them, success is quickly being able to assess a candidate's suitability versus spend a lot of time wading through the details. As such, the more you can tailor your content to what they are looking for and make "proof points" stand out, the more chance you have of winning their numbers game (which is frankly the one that matters!).