So you think you want to become a CDO

One of the dynamics of data science is constant change. So you have to continually invest in your own professional development. As you prepare for the role of Chief Data Officer, test your readiness in the areas described below. I'm not suggesting you have to be an expert in these areas, just literate. If you're literate, it will significantly demystify the world of data science. As the song says, "You can't lead where you don't go. You can't teach what you don't know."

Data wrangling

Data Science teams spend 80% of their time, energy and money on data wrangling...getting data ready to analyze by transforming and mapping that data from one "raw" data form into another with the intent of making it more appropriate and valuable for a variety of downstream purposes. I'm not talking about ETLs or getting data into spreadsheet. How up-to-date are your skills at the following?

- ☐ Data cleaning – removing erroneous data in a corpus of data.
- ☐ Data curation – selecting, managing and documenting the provenance of data
- ☐ Data editing – correcting errors in a corpus of data.
- ☐ Data fusion / data integration / semantic mapping – data integration across multiple sources
- ☐ Data pre-processing –cleaning data in data mining for analysis purposes
- ☐ Data scraping – extracting parts of a corpus of data with automated tools.

Statistics

Statistics are the heart of data science. What's your statistics literacy? Can you explain each of the following concepts to an executive audience, using recent examples?

- ☐ Chi^2
- ☐ decile
- ☐ Interquartile range (IQR)
- ☐ mean, mode, median,
- ☐ min & max
- ☐ quartile
- ☐ R and R^2
- ☐ Range
- ☐ Standard Deviation

Yes they're basics, but they matter.

Visualizations

A picture is worth a thousand words, and today's analytics relies on visualizations and infographics, not on tabular reports and spreadsheets. What's your visualization literacy? Can you explain each of the following to an executive audience, using recent examples?

- ☐ geo-spatial maps
- ☐ graph database relationship maps
- ☐ funnel
- ☐ heat
- ☐ zoom line
- ☐ multi-axis
- ☐ treemap
- ☐ box-and-whisker

When was the last time you looked at online examples of the best infographics and visualizations? If it's more than 6-9 months, schedule some time for it.

Algorithms

Machine learning relies on statistical algorithms. What's your statistical literacy? Can you explain each of the following to an executive audience, using recent examples? Do you know what it says about the shape of the data when each algorithm is found to be accurate for that data?

- ☐ boosted decision tree regression
- ☐ decision tree
- ☐ K-means, K-means clustering
- ☐ K-nearest neighbors
- ☐ linear regression
- ☐ logistic regression
- ☐ naive Bayes & Bayesian methods
- ☐ random forest

But wait, there's more.

- ☐ Can you explain the differences between accuracy and recall?
- ☐ Can you explain what deep provenance is?
- ☐ Can you explain the differences between supervised, unsupervised and reinforcement learning

There are a whole lot more concepts to learn, but this is a pretty good starter kit for a CDO.

What's your computer science literacy?

Data Science requires computer science knowledge and skills. How up-to-date are your skills? And a word of caution. I've taken a shortcut in some of the examples that follow, by naming products rather than functions. Don't take the product names literally; they're just proxies for the functions. If your organization uses different products for the same functions then adopt those products.

- ☐ Have you used open source SDKs like Anaconda & Jupiter Notebooks?
- ☐ Have you learned to code Python or R?
- ☐ Have you deployed, trained and tested a few machine learning algorithms?
- ☐ Have you done code pulls, pushes & commits on GitHub (or similar tool)?
- ☐ Have you written data stories / user stories and stored them in JIRA (or similar tool)?
- ☐ Have you edited scripts and saved them in Ansible (or similar tool)? Have you used those scripts to stand-up and tear down an infrastructure stack in one of the leading general purpose clouds?
- ☐ Have you edited digital recipes and stored them in Jenkins (or similar tool)?
- ☐ Have you spent any time working with data wrangling tools like StreamSets, cloudera SDX, gluu, or similar tools?
- ☐ Have you spent any time working with data visualization tools like Qlik, Tableau or PowerBI?
- ☐ Have you worked with modern collaboration tools like SLACK, CrowdChat, StormBoard, Teams?
- ☐ When was the last time you got a briefing from one of the leading cloud vendors on their technical direction?
- ☐ What do you know about thin edge, thick edge, AR & VR technologies?

And as long as we're on the topic, when was the last time you were hands-on with at least one of the tools? If it's more than 12 months, schedule some time for it.

What's your Artificial Intelligence (AI) literacy?

Look, we both know that AI is 99.999999% artificial and 0.000001% pretending to be intelligent. The reality is that AI tools are still very much prone to percolating the biases and assumptions made by their designers and the biases in the data that trains the algorithms. These oversights can

lead to serious problems that amplify existing prejudices and can undermine the goals of better decision making. All the resources you and your team apply to AI & machine learning aren't going to deliver value unless and until you install new guardrails and develop a culture of assessing conclusions with appropriate skepticism. All the same AI is part of the game.

- ☐ How many of the CB Insights AI.100 list are you familiar with?
- ☐ When was the last time you explored the offerings that might be useful for your team's value proposition?
- ☐ How up-to-date is your knowledge of the offerings that are already at General Availability?
- ☐ Are you familiar with the native AI services offered by each of the leading cloud vendors?

What's your legislation / regulation literacy?

Your organization operates in an environment shaped by legislation and regulation. There are rules and guidelines, especially when it comes to data. There are standards and best practices. Meet with your organization's Legal Affairs team and get an update. Identify the industry bodies your organization participates in. Talk to your (new) peers in competing organizations and across your supply chain; learn what they're paying attention to or concerned about.

What's your data science ethics literacy?

The world of data science is ripe with instances of well-meaning professionals deploying AI/ML solutions that created serious harm. The space is littered with examples unintended and seriously damaging consequences. Why is that?

An AI algorithm is code that writes code. Algorithms are not products, they're processes. We will never be sure what an AI process does until we run it. Here's the rub, that code uses your data to do the detailed work (assigning weights in tables). Since AI algorithms are incentive-seeking machines; the biggest problem in AI is always misaligned incentives. This means the old adage "garbage-in → garbage-out" applies to AI and ML. This isn't about the quality of the data, so much as it's about the suitability of the data. If your organization is like every other organization, your data is primarily historical. It represents the past you're trying to move away from and **not** the future you want to create.

Don't be the next headline. Examine the examples that have been in the headlines and understand the root causes behind them. Diversity is critical: people, models, data, mindsets, backgrounds. Think through how you'd build a data science team accordingly. Data science is a team sport after all.

How are your soft skills?

No, I'm not kidding. Soft skills matter. You and your team are going to have to persuade a lot of people to believe in you enough to accept your findings and then your recommendations. I started this paper with a sports, metaphor, but what you're actually going to build is a boutique consulting organization. You do that from the top down. That means you need people with good public speaking skills...facilitation skills...consulting skills...writing skills...story telling skills...business savvy...executive presence. How good are your skills today? What do you need to do to improve them? How good are the skills of your existing people? What do you need to do to improve them?

Move to action

This is where you have to make the decision to prioritize your own professional development for peak performance as a CDO. Given your answers to all the questions above, what is your professional development plan? If you don't have a format you prefer, try this format.

Learning Agenda
5 things I want to learn in the next 12 months: 1. 2. 3. 4. 5.
5 skills I want to develop in the next 12 months: 1. 2. 3. 4. 5.
5 experiences I want to have in the next 12 months: 1. 2. 3. 4. 5.
5 strengths I want to increase in the next 12 months: 1. 2. 3. 4. 5.
5 weaknesses I want to improve in the next 12 months: 1. 2. 3. 4. 5.

Now schedule the time for these activities in your calendar. You're going to be really busy and if you don't schedule the time now, you're less likely to invest the time later. Look, sports metaphors aren't always useful, but I'm going to use several here. Data Science is a team sport. As CDO you're in a player-coach role. That means you've got to suit up and show up ready to play. The foundational act of leadership is to lead by example.