Gmail

Jeffrey Katz <jeffreyerickatz@gmail.com>

Data Science soft skills

2 messages

Yu Chen <yoyothesheep@gmail.com> Sun, Mar 29, 2020 at 2:57 PM
To: Jeff Katz <jeffreyerickatz@gmail.com>

[I reworked this a bit, feel free to send now]

Context:

- As PM lead I've worked with 4 data scientists (2 from bootcamps), 1 data director, and 1 data intern at Flatiron School
- At Google I've worked w/ data scientists, data engineers, and strategy & operations analysts/principals from the top-3 consulting firms

General observations:

- Everybody knows how to crunch data
- Fewer people know how to draw insights from that data (e.g. spot interesting patterns or tell a story about real life from that data)
- Even fewer people know how to effect change with data (e.g. ask the right questions, pull the right data and analyze it, understand people/org in your company, present the data in a way that people care about it or would take action on it, continue to work with people to iterate on analyze <-> take action)

Traits / characteristics that make a data scientist a great addition to a team:

- **Teach others in the organization to be data-driven (exec-level).**
  - Example - let's say the Flatiron School's admissions team has cobbled together a process for recruiting and interviewing students. It includes emails, salesforce, calendar, technical interviews. Their process looks at top/low admissions reps every month and tries to learn best-practices from top performers. They've been driving towards a monthly closed goal for the last few years. A ninja-level data exec would understand the company and realize that the admissions team is locally optimizing. They would then develop strategies to improve the team & the company's performance. One step is talk to CEO and other high level execs to develop company goals & metrics (e.g. graduation rate) that unite depts, understand how each dept contributes to company goals, then discuss strategies to optimize globally by getting departments to work together. The 2nd step is to then figure out how to work with each individual department to help them change their processes and goal setting and data analysis. E.g. they would realize that because it's hard for admissions to tie their criteria to a graduation rate that's 15weeks out, they could develop early indicators. They would also know that the team needs an embedded data engineer to clean-up data from their various tools and also need an embedded data analyst to work with the VP and data engineer to develop early indicators plus dashboards that reps & managers can use.
- **Own a project (senior level).**
  - **Take a simple project request/description, figure out the needs, then follow through.** In the above example the data director tells the data scientist - I'd like you to work with a PM, engineer and the admissions VP to develop a new set of goals for the sales team. The data scientist would first interview the various stakeholders, understand what are the sales constraints and goals, understand what data is where, then develop a plan with the PM and engineer to define the data, get it, vet it, then finally train the team on using it.
    - **Ask the right questions** - they'd know enough about interviewing to know how to get the constraints and goals from various teams.
    - **Work with a team** - Work with PM & sales if data isn't even available and need a change in product or sales process. Work with engineers if data is partly available but need processing. Present regular progress updates to the team - ask for help if needed.
    - **Pull and analyze data** - understand data repositories, setup data pipelines (be clear about pipeline SLA). Know
    - **Help people use the data** - present the data, train users on dashboards/reports, make sure the data is clearly labeled and there's documentation on what stuff means (this is a common mistake). Follow up regularly to catch things that change.

    - **Another example** - let's say I have access to a ton of YouTube watchtime data, and we have an idea for a new product that will increase watchtime. But first - we need a data scientists' advice to 1) validate our assumption with existing logged data and 2) help us craft experiments to run to further test out our hypothesis. A really good data scientist would meet with the team, understand the need, then work with eng/pm/other data folks to pull the analysis and present the data back. The presentation should be in a story format that confirms/disproves our original hypothesis with caveats and next steps. If our hypothesis is confirmed, the data scientist then works with team to craft an experiment - the experiment parameters should be well designed so that the DS person can hand it off and not be on-call. I.e. the experiment parameters should be in the format of: collect <x> amount of data over <y> days <z> population, either through A/B arm or longitudinal experiment. This is the sensitivity analysis (what are the diff confidence intervals) and things to watch out for (don't p-hack).
- **Support a project (junior DS)**
    - A junior person can pull & analyze data, and sometimes train folks to use the output (not all of them can).

Examples of specific traits sometimes seen in data analysts/scientists that's difficult to work with:

- Don't apply the 80/20 rule - e.g. go down a rabbit hole and work for days without asking for any feedback
- Don't know how to draw insights from the data - e.g. somebody asks "so why are fewer students enrolling?" and the data analyst replies with a long story about how they pulled their data, caveats to their data and what they've observed, but not the actual answer.
- Present data in ways that are not useful - sales team wants to know which leads to prioritize, but the data dashboard is a huge pivot with many columns that don't actually sort or highlight "likely" candidates. Or - in a presentation the slides are full of dense numbers and not takeaways.
- Assume data is a certain way, or get frustrated if data isn't clean/perfect.

--
LinkedIn | Calendar

Jeffrey Katz <jeffreyerickatz@gmail.com> Sun, Mar 29, 2020 at 11:32 PM
To: Yu Chen <yoyothesheep@gmail.com>

Thanks so much. I'll share this with them on Tuesday
[Quoted text hidden]