

What's out there? Data Edition



Today's Agenda

- My story
- Tips for your job search
- Data and engineering roles explained

A little about me!

- **Hi!** I'm Ibby - most recently I ran the retention data team at Peloton, where I focused on user retention for their hardware business, and sometimes help screen technical applicants for roles across the company. I also helped teach part of Columbia's Data Analytics accreditation program, and have previously worked at Drift, Facebook, and a handful of startups. I graduated from Harvard in 2019 with a major in Statistics and a minor in Economics. I'm now working on a startup in the data space.

Check the boxes!

- **LinkedIn:** In today's world, having a really detailed LinkedIn profile is better than any cover letter or resume. Recruiters **will** contact you, make sure you've outlined technical experiences
- **Github:** Make sure you're committing often! Projects are always good to see (just commit the stuff you do in class tbh)
 - Note: **Kaggle** competitions are a great thing to participate in and the projects are awesome to populate your github with
- **Resume:** Your resume should include your technical experience, and skills should emphasize programming languages and tools

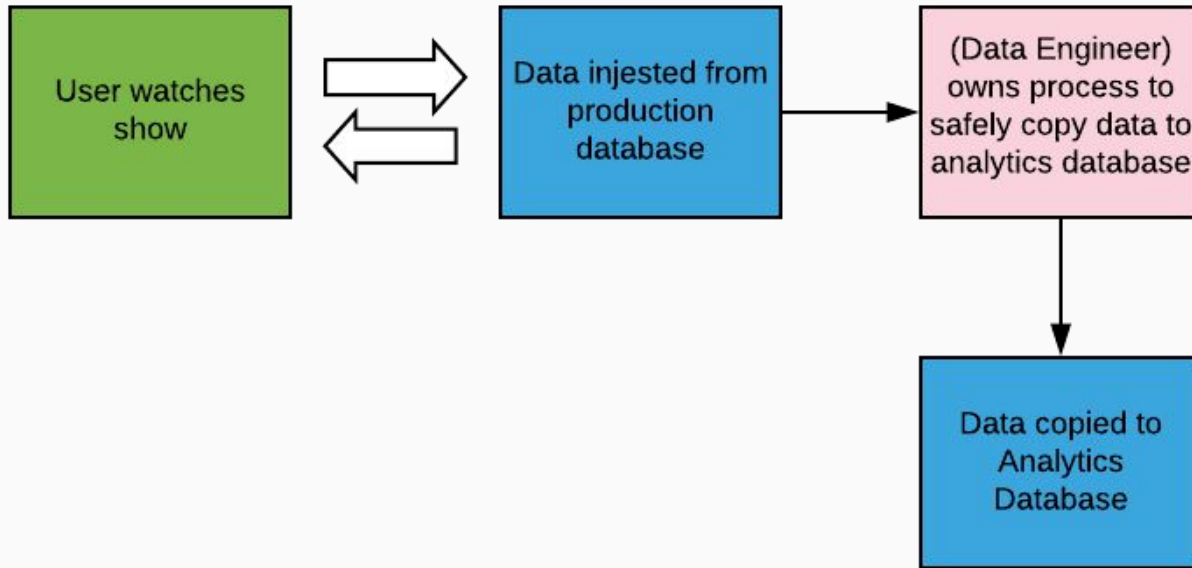
The myth

- **I *need* to work for a FAANG company.**
 - Nope, you don't. Technology is really abstracted at these large companies, and you're usually **maintaining existing products rather than working on new ones**. If you can handle being a bit more of a self-starter, **I'd suggest finding a smaller company** that's recently raised a late-stage venture round. It's structured enough to where you're working in one field, but early enough to where you get to wear different hats in said field (read: you're still learning!).

Data Engineer

- **Description:** Mainly in charge of moving data around - making sure that production data gets accurately moved to replica databases, and that pipelines exist so people can access said replica databases.
- **Stack:** Mainly dataflow-oriented (Python/SQL + tools like AWS Athena, Redshift, Snowflake, S3, Presto, DBT)
- **Using Netflix as an example:**

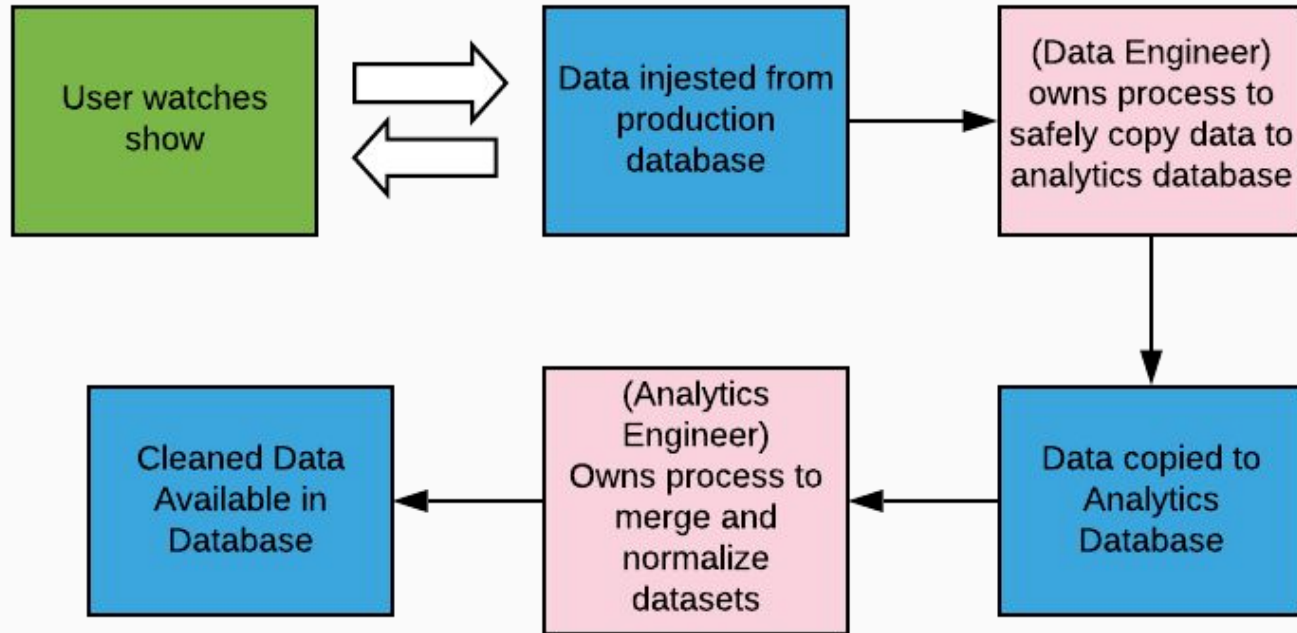
Data Engineer



Analytics Engineer

- **Description:** Sit between Data Engineers and other data consumers. In charge of doing data manipulation to regularly clean data and provide it to endpoints owned by data consumers, so they can glean insights/create dashboards
- .
- **Stack:** Mostly SQL data modeling, can be Python depending on how infrastructure is set up.

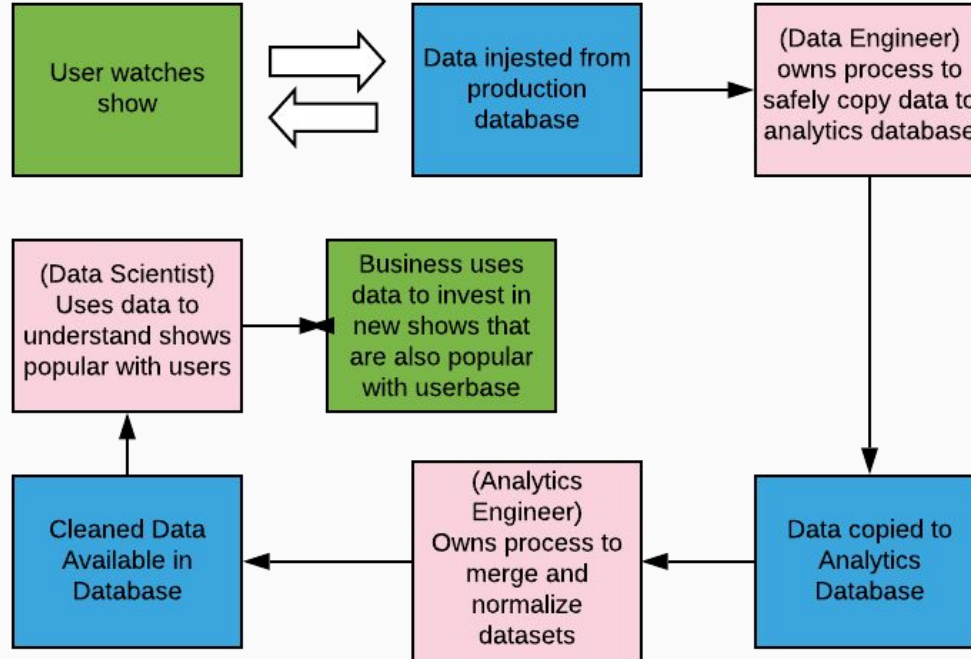
Analytics Engineer



Data Scientist

- **Description:** Use statistics and machine learning tools to understand trends/run observational studies and identify hidden patterns in data. Either on the business or product front, a DS's insights are used to drive various decisions. Depending on the technical background, their models are sometimes placed into some production capacity.
- **Stack:** Python (R can also be used, usually in more academic settings)

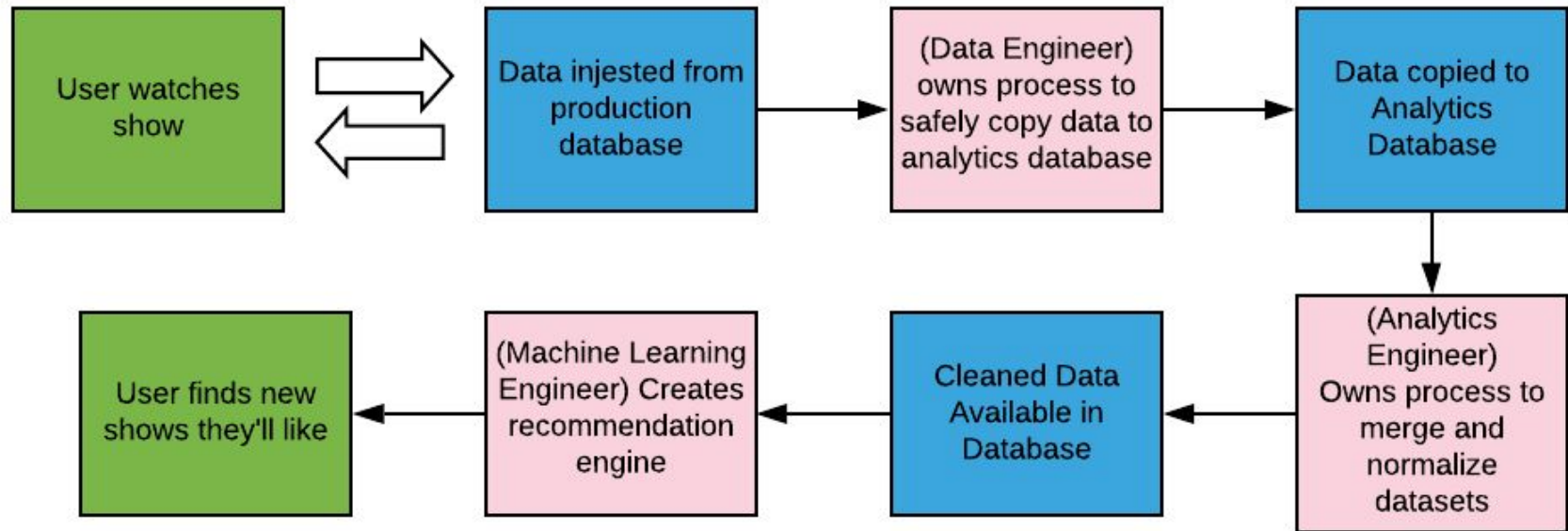
Data Scientist



Machine Learning Engineer

- **Description:** Usually work on product or marketing teams, ML engineers are usually a hybrid of data scientists and conventional software engineers. They train models to algorithmically make decisions and then launch said models into an automated production environment.
- **Stack:** Python (heavy reliance on libraries such as tensorflow or pytorch), C++

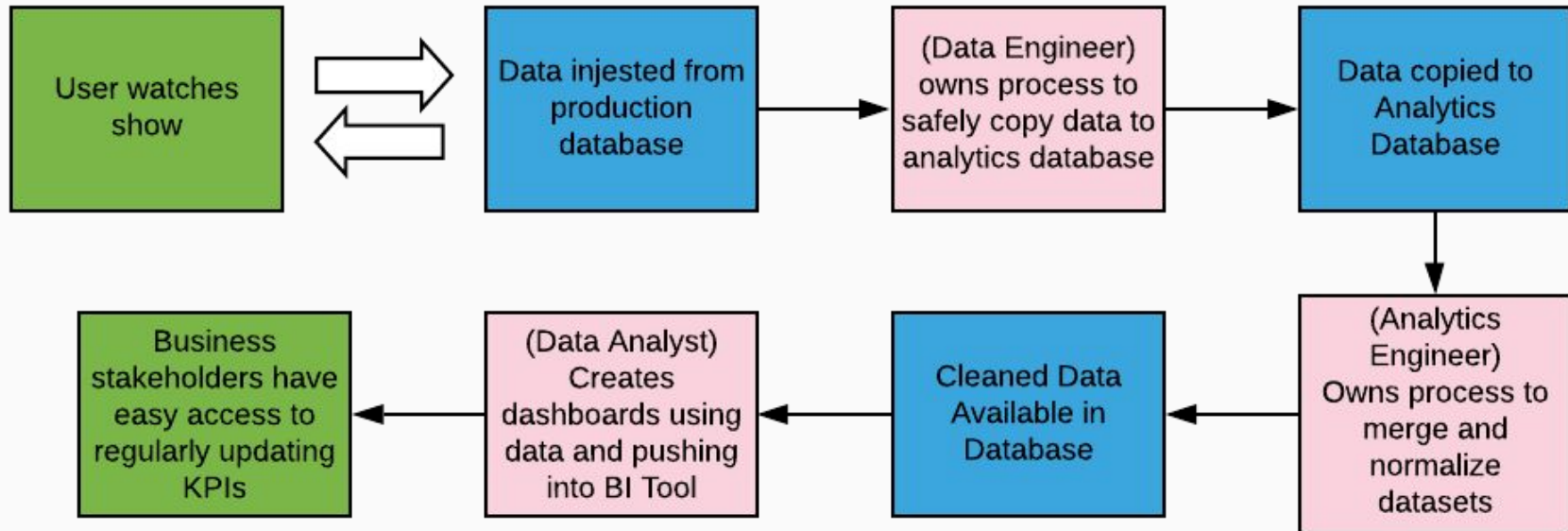
Machine Learning Engineer



Data Analyst

- **Description:** This one is the most variable, depending a lot on individual teams. In general, in charge of analyzing data from databases (more technical) or using BI-tools (less technical). In charge of understanding trends in KPI movements and relaying that to stakeholders. Many times are in charge of “reporting” in various business orgs.
- **Stack:** Some SQL, some templating depending on BI-tool used (Jinja/LookML), Tableau/PowerBI

Data Analyst



Visualization Engineer

- **Description:** A bit more of a rarity, Visualization engineers are closer to front-end developers and are in charge of creating custom visualizations where BI-tools won't do the trick.
- **Stack:** Javascript (Heavy use of D3.js and visualization libraries such as Leaflet.js), some Python

Product Manager (extra)

- **Description:** Usually a product “owner” and decides upcoming features, tests it with users, and works closely with designers and engineers.
- **Stack:** General design and system structure knowledge

Sales Engineer (extra)

- **Description:** For SaaS and enterprise products, Sales Engineers usually sit with account managers and either explain technical/and or security elements to potential customers.
- **Stack:** General design and system structure knowledge