



Decoding Data Science Roles

A guide to the different jobs data scientists (really) do

Stephanie Kirmer, April 2024

OCAIR

www.stephaniekirmer.com

About Me

- Worked my way up in academia and industry
- Industry experience in: mechanical engineering, logistics, travel/hospitality, distributed computing, and data privacy/security
- University teaching experience: sociology, health policy, research literacy
- Chicago! 



Agenda

1. Quick intro
2. Meet the Archetypes
3. Advancement & Choosing Org Size
4. Practice Interpreting Job Posts

Introduction

The Essentials

- Most data scientist roles will cover more than one function, at least in some part
- None of the archetypes I am going to describe are necessarily the “Job Title” you’ll find on a posting
 - You have to read and interpret to figure out which one/s they really mean
 - Some of these archetypes won’t even be called “Data Scientist” depending on the industry
- The rates of pay for these archetypes can vary- our field doesn’t value all skills equally, even when they are incredibly important and difficult.

Analytics Guru

Help the company figure out whether they are meeting goals, and whether stuff is working the way it should.

Key Functions

- Analyzing success of other departments and functions
- Understanding metrics and measures of success very deeply

Technologies & Skills

- Languages: R, SQL (Python is very much optional)
- Tableau, Looker, maybe PowerBI
- Data visualization best practices
- A/B testing

People will ask you:

- can you make me a dashboard?
- I need “the numbers” for this important meeting!

Feature Builder

Make the product better, ideally, with some sort of DS/ML wizardry.

Key Functions

- Understand the product and come up with ways to make it better using DS/ML
- Know what the customer needs and don't get fooled by hype
- Push back against OTHER people who get fooled by hype

Technologies & Skills

- Languages: Python, SQL
- Need to have software engineering competencies (APIs, unit testing, etc)
- May be useful to learn languages like Java or Ruby to work with engineering teams

People will ask you:

- Does our product use AI?

Infrastructure Builder

aka, ML (Ops) Engineer. Make the pipes that move the data & model predictions around.

Key Functions

- Know where the data comes from that the DS models need to use
- Take the DS/ML tools other people build and figure out how to make them scale
- Understand your company's application/s and make sure the data science stuff doesn't break anything

People will ask you:

- It's 2 am, the model's not making the right predictions, get out of bed and fix it

Technologies & Skills

- Languages: Python, SQL, and at least one production language like Java or Ruby
- Need to have software engineering competencies (APIs, unit testing, etc)
- Docker; observability tools like DataDog; scaling technologies like Spark/Dask
- Cloud computing systems (AWS, Azure, GCP)

*Big Companies

Internal Only

Build ML tools for other people in your company, not the external customers.

Key Functions

- Understand other departments at your company and what they do
- Build DS/ML tools to help internal stakeholders work better (and convince them to use your solution)
- Maintain the tools you build and keep them up to date with needs

Technologies & Skills

- Languages: Python, SQL
- Need to have software engineering competencies (APIs, unit testing, etc)
- May be useful to learn languages like Java or Ruby to work with engineering teams

People will ask you:

- The boss wants me to do this really slow repetitive task, is there a way this can be automated?
- This tool you built for me stopped working right, can you fix it?

www.stephaniekirmer.com

*Big Companies

Researcher

Think and write about cutting edge ML science stuff.

Key Functions

- Read a lot of scholarly articles about latest advances in ML
- Advise other people in DS/ML about techniques they might find useful from academia
- Go out and give talks about how advanced your company's ML program is

Technologies & Skills

- Languages: Python. Maybe distributed computing like Spark or Dask
- Deep learning frameworks (Tensorflow or PyTorch)
- If you just do thinking about big ideas and don't build models, then you may not need other programming languages.

People will ask you:

- I heard about this crazy new technology, can we do something with that?

www.stephaniekirmer.com

Solutions Engineer

Build and integrate data science tools for your company's external customers.

Key Functions

- Customer engagements (consulting) or direct customer support pre/post sales
- Learn what systems the customer is using and figure out how to make that work with your company's product
- Teach customers how to use the DS/ML capabilities of your product, or how to integrate with their systems

Technologies & Skills

- Languages: Python, SQL, and a smattering of other stuff so you're able to deal with whatever your customers are using in-house.
- Presentation skills: public speaking as well as data visualization and analysis

People will ask you:

- I have a customer that wants to know what we do with DS/ML, can you jump on a call?

My Current Role

Senior Machine Learning Engineer

What I do

- Build models that will help our product work better for customers (Feature Builder) - 75%
- Get on calls to explain our product's ML stuff to customers (Solutions Engineer) - 15%
- Do analyses about how well our product is currently working (Analytics Guru) - 5%
- Think about how we might incorporate new cutting edge stuff (Researcher) - 5%

I am the only DS/ML employee in our company, so I wear a lot of hats, and it is very hard work- I highly advise against doing this if you are not at least 3-5 years into your career.

Small versus Big Company

Pros and Cons

Small Company

- - Less guidance and help to learn (few or no other data scientists)
- + More opportunities to take on responsibility, you can try new things
- + Easier to move between roles if you want to
- - Risk of burnout from wearing too many hats
- + Opportunities to do open source work/use OSS

Big Company

- + More structure and support
- + More experienced people in your specific area to learn from
- - You may get siloed into your specific role
- - More red tape to try new things/slower innovation
- + More budget for conferences and development

Seniority and Advancement

Moving up that ladder

How to Advance

- Soak up technical skills from everyone around you (no matter if they are in DS or not)
- ALSO soak up “soft” skills. A lot of these roles involve understanding people and groups, not just code
- The higher you advance, the more the soft skills matter

What Advancement Means for Your Role

- More diversification of responsibilities and expectations
- Strategizing and planning more, versus carrying out instructions
- You’re expected to have informed opinions about what to do based on experience

Reading Job Posts

www.stephaniekirmer.com

Reading a Job Posting

Analytics Guru

- Title: Data Scientist, Analytics at DoorDash
- “dive into large datasets to guide decision-making”
- “Identify and measure levers to help move essential metrics and make recommendations”
- R/SQL/Python, Looker/Tableau/Chartio
- A/B testing, experimentation, hypothesis testing

Data Scientist / Senior Data Scientist, Analytics

Job ID: 2730279

San Francisco, CA; Seattle, WA; Ne

About the Team

The Analytics team is looking for experienced Data Scientists and Senior Data Scientists to guide measurement, strategy, and t a variety of teams and levels. Data Scientists at DoorDash work to uncover insights and turn them into relevant recommendations. Analytics is integral to all operational areas at DoorDash.

Please apply here for all non-managerial levels within the following analytics teams:

- Consumer & Growth
- Business Operations
- Dasher & Logistics
- Customer Experience & Integrity
- Merchant, Ads & Sales
- New Verticals
- International Data Science

About the Role

<https://careers.doordash.com/jobs/5523494>

Reading a Job Posting

Feature Builder (with Analytics Guru)

- Title: Data Scientist at Aven
- “Partner across the company ... to shape product strategy using rigorous scientific solutions”
- “Design, implement and launch innovative data science solutions to empower data-driven decisions at scale”
- R/SQL/Python
- “Experience designing and implementing complex modeling solutions from ideation to production at scale”



Data Scientist

Aven · Campbell, CA · 1 month ago · Over 100 applicants

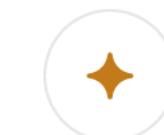
On-site · Full-time

11-50 employees

7 of 10 skills match your profile - you may be a good fit

Job poster joined LinkedIn in 2007 ?

See how you compare to over 100 other applicants. [Try Premium for \\$0](#)



Am I a good fit for this job?

How can I best position myself?

Easy Apply

Save

About the job

Aven Data Scientist

Quick Facts:

Focus: Asset-Backed Credit Cards

Funding: \$140+ million, Series C

Headquarters: Campbell, CA

Website: <https://www.aven.com/>

Are you ready to make a profound impact through data?

www.stephaniekirmer.com



Reading a Job Posting

Feature Builder

Data Scientist, Algorithms-Fulfillment Optimization

San Francisco, CA

Apply

- Title: Data Scientist, Algorithms at Lyft
- “Drive the Science roadmap of the Dispatch problem area...to direct creations and improvements of algorithms and models underpinning the Dispatch systems and products”
- “Perform exploratory data analysis to gain a deeper understanding of the problem”
- “Proficiency with Python and working in a production coding environment”
- “Proven experience with building and evaluating machine learning models or optimization models”

At Lyft, our mission is to improve people's lives with the world's best transportation. To do this, we start with our own community by creating an open, inclusive, and diverse organization.

Data Science is at the heart of Lyft's products and decision-making. As a Data Scientist on the cross-functional team, you will work in a dynamic environment, where we embrace moving quickly to build the world's best transportation. Data Scientists take on a variety of problems ranging from shaping critical business decisions to building algorithms that power our internal and external products. We're looking for passionate, driven Data Scientists to take on some of the most interesting and impactful problems in ridesharing.

https://app.careerpuck.com/job-board/lyft/job/7223971002?gh_jid=7223971002

www.stephaniekirmer.com

Reading a Job Posting

Infrastructure Builder

- Title: ML Infrastructure Engineer at Photon
- This example doesn't tell you much about the responsibilities, but you can get hints from the competencies they want
- "Proficiency in programming languages Python, Java"
- "Strong understanding of Containerization technologies (e.g., Docker, Kubernetes)."
- "Experience with cloud platforms such as AWS, GCP, or Azure."

ML Infrastructure Engineer | Onshore

MD, United States and 1 more

JOB DESCRIPTION

- Experience in Machine Learning engineer or Infrastructure roles, with a focus on Machine Learning infrastructure.
- Proficiency in programming languages Python, Java
- Experience with cloud platforms such as AWS, GCP, or Azure. GCP experience is strongly preferred.
- Familiarity with machine learning frameworks (e.g., TensorFlow, PyTorch) and libraries (e.g., scikit-learn).
- Knowledge of DevOps principles and tools (e.g., CI/CD pipelines, Terraform).
- Strong understanding of Containerization technologies (e.g., Docker, Kubernetes).
- Excellent problem-solving and communication skills.
- Experience with big data technologies such as Apache Spark or Hadoop Preferred.

APPLY NOW

JOB INFO

Job Identification

19102

Posting Date

03/15/2024, 03:24 PM

Locations

 Maryland, Maryland, US  Mountain View, California, US

<https://www.photon.com/careers>

www.stephaniekirmer.com

Reading a Job Posting

Internal Only (with Analytics Guru)

- Title: Staff Data Scientist at Devoted Health
- "The Staff Data Scientist will serve as a crucial partner to the Payment Integrity Operations team, which is responsible for managing claim audit and adjustment functions and driving projects to prevent repetitive errors."
- "Working with the PI team to build out data warehousing infrastructure in Snowflake to ... improve Devoted's internal operating platform"
- "Leveraging domain knowledge of healthcare billing and claims processing"
- "Equally comfortable building dashboards and reports for KPIs, and doing deep-dive statistical analysis; such as predictive modeling and experimental design to influence the business"

The image shows a job posting for a Staff Data Scientist position at Devoted Health. The top banner features the Devoted logo and the tagline "Every member is family." Below the banner, the job title "Staff Data Scientist" is displayed, followed by a blue "Apply" button. To the left of the job description, there is a location icon with "Remote USA" and "Waltham Massachusetts Office". To the right, there are icons for "Full time", "Posted 30+ Days Ago", and "R1818". The job description text is as follows:

At Devoted Health, we're on a mission to dramatically improve the health and well-being of older Americans by caring for every person like family. That's why we're gathering smart, diverse, and big-hearted people to create a new kind of all-in-one healthcare company – one that combines compassion, health insurance, clinical care, service, and technology - to deliver a complete and integrated healthcare solution that delivers high quality care that everyone would want for someone they love. Founded in 2017, we've grown fast and now serve members across the United States. And we've just started. So join us on this mission!

Job Description

[https://devoted.wd1.myworkdayjobs.com/
Devoted/job/Remote-USA/Staff-Data-
Scientist_R1818](https://devoted.wd1.myworkdayjobs.com/Devoted/job/Remote-USA/Staff-Data-Scientist_R1818)

www.stephaniekirmer.com

Reading a Job Posting

Researcher

- Title: Senior Applied ML Researcher at Yurts AI
- “Advanced Research”: “push the state of the art in natural language processing”
- “Rapid Prototyping”: “Develop rapid research prototypes or conduct simulations to further evaluate the impact of research”
- “Thought Leadership”: “Author research papers, technical blogs to share and generate impact of research results across the team and in the research community.”
- “Identify new and upcoming research areas by interacting with potential external and internal collaborators”
- Python, deep learning frameworks

Sr Applied ML Researcher

at Yurts ([View all jobs](#))

San Francisco (Hybrid)

Company Overview:

At Yurts, we are committed to pushing the boundaries of artificial intelligence and driving innovation in enterprise solutions. Our goal is to create cutting-edge applications that leverage the power of LLMs and Generative AI to transform businesses. We are seeking a highly talented and motivated Senior Research Scientist passionate about delivering value to the portfolio of products and experiences connected to Generative AI offered by Yurts. If you have experience designing new algorithms in the NLP/LLM space, pursuing rapid prototyping of the latest techniques from the academic world to deliver mission impact at enterprise scale and enjoy working alongside product teams to iteratively improve capabilities, we would love to hear from you!

Job Summary:

Are you a seasoned Machine Learning researcher with a deep passion for Natural Language Processing (NLP) and Large Language Models (LLM)? Do you thrive on tackling the most complex challenges in AI and possess a track record of groundbreaking research in these domains? If so, we invite you to join our cutting-edge team as a Senior Machine Learning Researcher.

<https://boards.greenhouse.io/yurtsai/jobs/4101739007>

www.stephaniekirmer.com

[Apply Now](#)

Reading a Job Posting

Solutions Engineer

- Title: Specialist Solutions Architect at Databricks
- “trusted technical ML expert to both Databricks customers and the Field Engineering organization”
- “guide customers in architecting production-grade ML applications on Databricks”
- “Collaborate with the product and engineering teams to represent the voice of the customer”
- Experience with NLP and generative AI



Why Databricks Product Solutions Resources About

DATA + AI SUMMIT

Overview Culture Benefits Diversity Engineering Research Students & new grads

[Back to search results](#)

Specialist Solutions Architect – Data Scientist/Machine Learning Engineer

United States

[Apply now](#)

FEQ425R84

This role can be remote.

As a Specialist Solutions Architect (SSA) – Data Scientist / ML Engineer, you will be the trusted technical ML expert to both Databricks customers and the Field Engineering organization. You will work with

<https://www.databricks.com/company/careers/field-engineering/specialist-solutions-architect---data-scientist-machine-learning-engineer-7266250002>

www.stephaniekirmer.com

Questions?

Thanks for having me!

- Women in Data Science job board: <https://www.widsworldwide.org/opportunities/>
- Networking opportunities: PyData meetups, R user groups, American Statistical Association