Writing your BEST Data Science Application

Emily Kearney
PhD, Environmental Science, Policy, and Management
Data Science Lead - NYC

Kyle Frankovich
PhD, Cognitive Neuroscience
Data Science Lead - Remote

INSIGHT

Roadmap

1. What are my transferable skills?
2. The Application
3. Common Pitfalls
4. Questions

What are my transferable skills?

Who are Academic Scientists?

Job of an academic scientist:

- Plan and execute a study ~ years
- Collect and clean data
- Use programming and statistics/ML to discriminate between signal and noise
- Convey results to the scientific community

Who are Data Scientists?

Job of a data scientist in industry:

- Plan and execute a study ~ week-long sprints
- Collect and clean data
- Use programming and statistics/ML to discriminate between signal and noise
- Convey results to the team/company/investors
- Make data-informed decisions that directly impact a product and a business

The Modern Data Scientist

Math and Statistics

- Machine learning
- Statistical modeling
- Experimental design
- Bayesian inference
- Supervised learning
- Unsupervised learning
- Optimization

Programming and Database

- Computer science fundamentals
- Scripting language e.g., Python
- Statistical computing packages, e.g., R
- Databases; SQL
- Parallel databases and parallel query processing
- MapReduce concepts
- Hadoop, Hive, Spark
- Experience with AWS

Domain Knowledge and Professional Skills

- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic,proactive, creative, innovative, and collaborative

Communication and Visualization

- Able to engage with senior management
- Storytelling skills
- Translate data driven insights into decisions and actions
- Visual art design
- Knowledge of visualization tools

The Modern Data Scientist

Math and Statistics

- Machine learning
- Statistical modeling
- Experimental design
- Bayesian inference
- Supervised learning
- Unsupervised learning
- Optimization

Programming and Database

- Computer science fundamentals
- Scripting language e.g., Python
- Statistical computing packages, e.g., R
- Databases; SQL
- Parallel databases and parallel query processing
- MapReduce concepts
- Hadoop, Hive, Spark
- Experience with AWS

Things to highlight

Domain Knowledge and Professional Skills

- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic,proactive, creative, innovative, and collaborative

Communication and Visualization

- Able to engage with senior management
- Storytelling skills
- Translate data driven insights into decisions and actions
- Visual art design
- Knowledge of visualization tools

Outline of the Insight Application

- Describe your research to a non-expert.
- If you have done a side project, describe it.
- What coding languages do you know and how have you used them?
- What ML or statistical tools/methods are you familiar with?
- Why are you interested in data science?

Questions about your work:

Describe your research
to a non-expert.
&
If you have done a side project,
describe it.

What we are assessing:

- Communication of technical concepts to a **broad audience**
- Your use of/experience with **ML/stats outside of courses**
- Your ability to identify and communicate the **impact** of your work
- From the side project: what has been **your experience with data science**?
  - Did you push yourself into unfamiliar territory?
  - Did you produce a useful product that a real person would use?

Answering effectively

- Pretend that you are explaining your work to **a smart, but not technically inclined person**
- **Start big** and become more specific as you go
- Make sure you tell us the **WHY**
- Be **concise** and clear
- **Define** any jargon, or better yet, avoid it altogether
- **Do not list** publications in place of explanations, re-use an abstract, or copy straight from your resume

Example:

High-level Summary: Helping agriculture by supporting native pollinators

Some Details: Observational data from multiple studies and designed & implemented three independent experiments, used generalized linear mixed models and simulations to analyze the effect of different treatments on pollinators, ...

The Why: I gave concrete recommendations to farmers about how and when they should change their management of their fields in order to increase yields through natural pollination saving them $100-1000s

Questions about technical tools/methods:

What coding languages do you know?
&
What ML or statistical tools/methods are you familiar with?

What we are assessing:

- Your **breadth of experience** with technical tools/methods
- Your use of **ML/stats in research & other contexts**
- Your **depth** of programming knowledge
- Your **motivation** to transition to data science (if you do research in Matlab, are you learning Python?)

Answering effectively

- **Don’t list** everything that you have taken for a test drive
- Give us **context!**
- Tell us **where** you have used something = in “the wild” (i.e. research) or in a course
- Tell us **how long** you have worked with a language, tool, or method

Example - Languages:

R: 10+ years, from classes in undergrad, for research in internships & PhD; used for processing data, visualizations, modeling, and simulations

SQL: 1+ year, self-taught, created database for side project

Python: 1+ year, self-taught, used for side project; web scraping, simple ML models (linear regression & RF), & visualizations

Why are you interested in data science?

What we are assessing:

- Your **understanding** of data science as a career path
- Your **motivation** to transition to data science
- **Your interests!**

Answering effectively

- Tell us **why you are running towards data science! (**Not why you are running away from academia/your current job)
- **Share any explorations** you have done - talking to DSs, reading blogs, listening to podcasts, etc.
- Mention **specific examples** of DS in practice in industry that you are interested in
- Don’t tell us why you are a good fit for DS - the rest of the app will do that!

Common Pitfalls

Repeating Your Resume

Over/under-selling

Novelty or Academic Burnout

Applications are for context!

- Tell us what's NOT on your resume
- Focus on connecting the dots
- Explain the motivation behind your choices

Trust us to assess!

- Present yourself confidently
- Don’t assume that you are already “doing DS”
- Focus on interests
- Give us context and let us do the rest!

Most Data Science is not cutting edge

- “Interesting” is not always valuable
- Find the correct tool for the job
- If you want to research ML, academia might be the best place!

Your enthusiasm is critical

- Focus on moving forward
- Emphasize the positives
- Why data science? Be specific!

DEI : Insight Candidate Mentorship Program

What it is: a short-term commitment between applicants* who self-identified as belonging to at least one underrepresented group in tech and our alumni

Eligible historically underrepresented groups:

- Racial and ethnic groups: Black/African American, Hispanic, LatinX, American Indian, Native Alaskan, and Pacific Islander
- Gender minority groups: women, third-gender, non-binary, gender-fluid

What's offered:

- 30-60 minute call with an Insight alumnus
- Opportunity to discuss Insight interview, program experience, and workforce experience

*Candidates are eligible to receive mentorship regardless of admissions status

Questions?

Emily Kearney

Data Science Lead - NYC

linkedin.com/in/emily-kearney-phd

Kyle Frankovich

Data Science Lead - Remote

linkedin.com/in/kylefrankovich/

INSIGHT