Preparing for the Technical Interview

Lindsay Warrenburg
The Erdős Institute

Outline

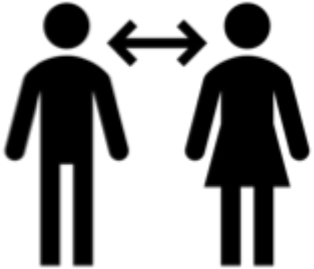
1. What is data science?
2. What types of data science are there?
3. What subject matter topics do I need to know?
4. What will my interview look like?
5. How do I prepare?

What is data science?

Data science is a buzzword

- “Data science” can mean very different things to different people

- **Analogy**: cognitive neuroscience and therapy are both types of psychology



Data science is a buzzword

- Technical interview prep will be different for different types of data science positions

Data science is a buzzword

- Technical interview prep will be different for different types of data science positions

- Before applying to a job position, you should have a general idea of which type of data science you want to go into

Data science is a buzzword

- Technical interview prep will be different for different types of data science positions

- Before applying to a job position, you should have a general idea of which type of data science you want to go into

- Job descriptions will give you some indication of the ‘flavor’ of data scientist, but sometimes you’ll learn what the company needs as you’re going through the interview process

Data science is a buzzword

The GOOD thing is that you do not need to know every topic from every data science flavor to do well in your interviews

What types of data science are there?

Data Engineering	Computer Vision	Data Product
Software Engineering	Genomics & Computational Biology	Decision Science
Computer Science	Human Factors Engineering	Business Intelligence
Machine Learning Research	Speech Processing & Audio Engineering	User Experience & Customer Insights
Machine Learning Engineering	Natural Language Processing (NLP)	Market Research

What subject matter topics do I need to know?

Machine Learning

- Supervised & Unsupervised Learning
- L1 vs. L2 regularization
- Sensitivity, Specificity, F1
- PCA, t-SNE, UMAP
- Random Forest vs. XGBoost vs. Catboost

Statistics & Probability

- Normal vs. Binomial vs. Poisson vs. Bernoulli distributions
- KL Divergence
- Q-Q plot
- OLS & ML Regression
- Power Analysis

Computer Science

- List comprehension
- Python environments (virtualenv, pip, conda)
- Arrays vs. Hashmaps vs. Queues vs. Linked Lists
- Time & Space Complexity / Big-O Notation
- Industry Terms / Services (AWS, GCP, Azure, Tableau, Power BI, Agile)

SQL

- WHERE
- DISTINCT
- LEFT, RIGHT, FULL OUTER, INNER JOIN
- SELECT
- COUNT

Business

- KPIs
- Churn
- LTV
- Funnel Analysis
- A/B Testing vs. Multi-armed Bandits

What will my interview look like?

Technical Interview Formats

- There are several formats for technical interviews

Technical Interview Formats

- There are several formats for technical interviews

- Each company's technical interview is (most likely) multiple steps, each with a different format

Technical Interview Formats

- There are several formats for technical interviews
- Each company's technical interview is (most likely) multiple steps, each with a different format
- Different companies use different combinations of interview formats

Demo Projects

- 5-min overview of a project you've done on your own

Demo Projects

- 5-min overview of a project you've done on your own
- Interviewers could be totally unfamiliar with data science techniques or could be a professional data scientist themselves

Demo Projects

- 5-min overview of a project you've done on your own
- Interviewers could be totally unfamiliar with data science techniques or could be a professional data scientist themselves
- You should have an annotated GitHub and PowerPoint for these projects

Demo Projects

- 5-min overview of a project you've done on your own
- Interviewers could be totally unfamiliar with data science techniques or could be a professional data scientist themselves
- You should have an annotated GitHub and PowerPoint for these projects

Hint: You can use your Erdős project!

White Boarding

- The interviewers will ask a question about a data science topic and you will explain it to them, using a whiteboard to draw pictures, write equations, write sample code

White Boarding

- The interviewers will ask a question about a data science topic and you will explain it to them, using a whiteboard to draw pictures, write equations, write sample code
- Very similar to how you'd respond if a student asked you a question about your lecture. They want to see how you think out loud.

White Boarding

- The interviewers will ask a question about a data science topic and you will explain it to them, using a whiteboard to draw pictures, write equations, write sample code
- Very similar to how you'd respond if a student asked you a question about your lecture. They want to see how you think out loud.
- Questions could be on anything, from **Please describe the relationship between the p-value and the assumptions of the Central Limit Theorem** to **Here is a list, L, of numbers: L = [9,3,37,1,110,36]. Please write a Python function that uses Timsort to sort this list from low to high?**

White Boarding

- The interviewers will ask a question about a data science topic and you will explain it to them, using a whiteboard to draw pictures, write equations, write sample code
- Very similar to how you'd respond if a student asked you a question about your lecture. They want to see how you think out loud.
- Questions could be on anything, from **Please describe the relationship between the p-value and the assumptions of the Central Limit Theorem** to **Here is a list, L, of numbers: L = [9,3,37,1,110,36]. Please write a Python function that uses Timsort to sort this list from low to high?**
- You can ask the interviewer clarifying instructions as you're answering. If you get stuck, they might give you a hint to see how you process new information

White Boarding

- The interviewers will ask a question about a data science topic and you will explain it to them, using a whiteboard to draw pictures, write equations, write sample code
- Very similar to how you'd respond if a student asked you a question about your lecture. They want to see how you think out loud.
- Questions could be on anything, from **Please describe the relationship between the p-value and the assumptions of the Central Limit Theorem** to **Here is a list, L, of numbers: L = [9,3,37,1,110,36]. Please write a Python function that uses Timsort to sort this list from low to high?**
- You can ask the interviewer clarifying instructions as you're answering. If you get stuck, they might give you a hint to see how you process new information
- The goal is for the interviewer to understand the boundaries of your knowledge -- if you get stuck, that's okay! They'll likely keep asking harder questions until you don't know anymore. No one interviewing will know all of the answers to these questions.

White Boarding

- The interviewers will ask a question about a data science topic and you will explain it to them, using a whiteboard to draw pictures, write equations, write sample code
- Very similar to how you'd respond if a student asked you a question about your lecture. They want to see how you think out loud.
- Questions could be on anything, from **Please describe the relationship between the p-value and the assumptions of the Central Limit Theorem** to **Here is a list, L, of numbers: L = [9,3,37,1,110,36]. Please write a Python function that uses Timsort to sort this list from low to high?**
- You can ask the interviewer clarifying instructions as you're answering. If you get stuck, they might give you a hint to see how you process new information
- The goal is for the interviewer to understand the boundaries of your knowledge -- if you get stuck, that's okay! They'll likely keep asking harder questions until you don't know anymore. No one interviewing will know all of the answers to these questions.

It helps to practice this, with other people, out loud, over a few weeks

Data Challenges

- Companies want to see how you handle data analysis for data you've never seen before. They will give you a data set, a question, and a time limit and you will present a complete analysis at the end of the time limit

Data Challenges

- Companies want to see how you handle data analysis for data you've never seen before. They will give you a data set, a question, and a time limit and you will present a complete analysis at the end of the time limit
- Dataset will typically be a csv but could be some other format

Data Challenges

- Companies want to see how you handle data analysis for data you've never seen before. They will give you a data set, a question, and a time limit and you will present a complete analysis at the end of the time limit
- Dataset will typically be a csv but could be some other format
- Time limit could be 2-3 hours or 2-3 days from when they give you the data to when you turn it in

Data Challenges

- Companies want to see how you handle data analysis for data you've never seen before. They will give you a data set, a question, and a time limit and you will present a complete analysis at the end of the time limit
- Dataset will typically be a csv but could be some other format
- Time limit could be 2-3 hours or 2-3 days from when they give you the data to when you turn it in
- The question prompt will be about a page long and include a problem statement, any relevant background information, and a description of the dataset (what each column name means, etc.)

Data Challenges

- Companies want to see how you handle data analysis for data you've never seen before. They will give you a data set, a question, and a time limit and you will present a complete analysis at the end of the time limit
- Dataset will typically be a csv but could be some other format
- Time limit could be 2-3 hours or 2-3 days from when they give you the data to when you turn it in
- The question prompt will be about a page long and include a problem statement, any relevant background information, and a description of the dataset (what each column name means, etc.)
- You should return (1) annotated Jupyter notebook and (2) 1-2 paragraph summary with takeaways and how your solution solves the problem statement

Data Challenges

- Companies want to see how you handle data analysis for data you've never seen before. They will give you a data set, a question, and a time limit and you will present a complete analysis at the end of the time limit
- Dataset will typically be a csv but could be some other format
- Time limit could be 2-3 hours or 2-3 days from when they give you the data to when you turn it in
- The question prompt will be about a page long and include a problem statement, any relevant background information, and a description of the dataset (what each column name means, etc.)
- You should return (1) annotated Jupyter notebook and (2) 1-2 paragraph summary with takeaways and how your solution solves the problem statement

Your Erdős project will be a close simulation to a Data Challenge (except you won't choose the prompt or data!)

Paired Coding

- Similar to a white board question except the interviewer watches you coding (or codes with you!)

Paired Coding

- Similar to a white board question except the interviewer watches you coding (or codes with you!)
- The interviewers can give you feedback and see how you handle const

Paired Coding

- Similar to a white board question except the interviewer watches you coding (or codes with you!)
- The interviewers can give you feedback and see how you handle constant feedback
- This type of interview is one of the easiest to conduct online so I think it is more likely during Covid than other types of interviews

Paired Coding

- Similar to a white board question except the interviewer watches you coding (or codes with you!)
- The interviewers can give you feedback and see how you handle const
- This type of interview is one of the easiest to conduct online so I think it is more likely during Covid than other types of interviews

Practicing with each other is the best way to prepare

Case Studies

- An extra-long version of a single whiteboard question

Case Studies

- An extra-long version of a single whiteboard question
- They will give you a problem with no perfect solution to see how you think out loud

Case Studies

- An extra-long version of a single whiteboard question
- They will give you a problem with no perfect solution to see how you think out loud
- You're expected to ask a lot of clarifying questions and engage with the interviewer. Without asking them questions, you won't be able to answer adequately

Case Studies

- An extra-long version of a single whiteboard question
- They will give you a problem with no perfect solution to see how you think out loud
- You're expected to ask a lot of clarifying questions and engage with the interviewer. Without asking them questions, you won't be able to answer adequately
- The question could be something like **Here's the background of our company in the last three years. Given this information, what can we do to reduce employee churn in departments X and Y.**

Case Studies

- An extra-long version of a single whiteboard question
- They will give you a problem with no perfect solution to see how you think out loud
- You're expected to ask a lot of clarifying questions and engage with the interviewer. Without asking them questions, you won't be able to answer adequately
- The question could be something like **Here's the background of our company in the last three years. Given this information, what can we do to reduce employee churn in departments X and Y.**
- Case studies can be about business problems, creating a data analysis plan, building an algorithm from scratch, etc.

Case Studies

- An extra-long version of a single whiteboard question
- They will give you a problem with no perfect solution to see how you think out loud
- You're expected to ask a lot of clarifying questions and engage with the interviewer. Without asking them questions, you won't be able to answer adequately
- The question could be something like **Here's the background of our company in the last three years. Given this information, what can we do to reduce employee churn in departments X and Y.**
- Case studies can be about business problems, creating a data analysis plan, building an algorithm from scratch, etc.

Reading published case studies and taking notes helps!

How do I prepare?

Preparation Tips

- Practice, practice, practice -- out loud and with others (not just silent reading by yourself!)
- Use our **tech interview materials on the Erdős website**
- Sign up for **practice groups**

Ace the Data Science Interview

DataLemur

Sign

Ace The SQL Interview

Practice SQL Interview questions asked by top tech companies **FOR FREE** on DataLemur. Made by Nick Singh, Best-Selling Author of Ace the Data Science Interview

COMPANY	TITLE	DIFFICULTY
LinkedIn	Data Science Skills	Easy
Facebook	Page With No Likes	Easy

Homework for this month

- Research the different types of data science jobs and figure out which type(s) best align with your interests, values, and career goals

- Go through tech interview folder and start prepping (yourself and with others)

Coming up next month

- White boarding & paired coding prep overview (with me)
- Dedicated group sessions for white boarding & paired coding
- Weekly problem sets starting after the next lecture

Contact me on our Slack channel :)