

# Mastering the Data Science Interviews Loop

Documenting summaries, and lessons learning from personal experience interviewing with both small startups, and several top-tier companies including: Microsoft, Twitter, Airbnb, Branch, Apple, and NVIDIA: the most efficient ways to ace the data science interviews.

We’ve also gathered knowledge from various data scientists who have interviewed hundreds of candidates at those companies.

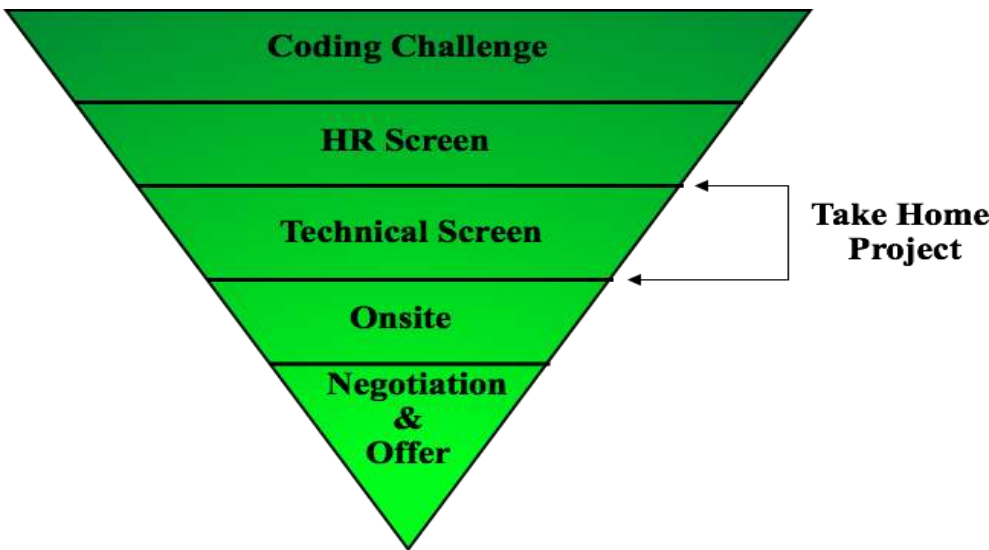
We seek to provide a full picture, and a comprehensive summary of the questions, and what to expect/ encounter during the rounds of data science interviews in a small company or a FAANG.

*Every battle is won or lost before its even fought: Sun Tzu*

## Table of Contents

- 1. The General Data Science Interview Process
- 2. Data Science Interview Tracks
- 3. Preparation for the Specific Data Science Interviews Subjects
- 4. Onsite Data Science Interviews
- 5. What Interviewers check, and evaluate during the interviews?
- 6. Tips to Ace the Data Science Interviews
- 7. Mistakes data scientists make during the Data Science interviews
- 8. Negotiation the job offer

### 1.0 The General Data Science Interview Process



- 1. A recruiter initial phone call
- 2. 1 or 2 rounds of technical phone screening (TPS)

Kennedy Wangari || Data Scientist || AI Community Lead

3. Case study project/ take-home data assignment
4. A 4 ~ 5-hour onsite interview, typically includes 3 ~ 4 rounds of technical interviews.
5. A behavioral interview with the hiring managers.
6. Technical Discussion/ interview with the tech team lead/VP of Data Science/ executives/ data scientists.
7. Wait forever and get paranoid
8. Negotiation. Always negotiate.

## 2.0 Data Science Interview Tracks

### a) Product Analytics (~70% on the market)

- Requirements: practical experience launching products; strong business acumen; advanced SQL skills
- Examples: Data Scientist, Analytics at Airbnb; Data Scientist at Lyft; Data Scientist at Facebook; Product Analyst at Google

### b) Modeling (~20% on the market)

- Requirements: knowledge of machine learning (not only how to use it but also the underlying math and theory); strong coding ability
- Examples: Data Scientist, Algorithms at Lyft; Data Scientist, Algorithms at Airbnb; Applied Scientist at Amazon; Research Scientist at Facebook

### c) Data Engineering (~10% on the market)

- Requirements: end to end data scientists with data engineering skills; knowledge of distributed systems; MapReduce and Spark; practical experience working with Spark; strong coding ability
- Examples: Data Scientist, Foundation at Airbnb; Data Scientist at some startups

## Airbnb Data Science tracks of work

Many companies consider three tracks of data science work to meet the needs of the business — **Analytics**, **Inference**, and **Algorithms**.

## Data Scientist – Analytics

Defines and monitors metrics, creates data narratives, builds tools

## Data Scientist – Algorithms

Builds and interprets algorithms that power data products

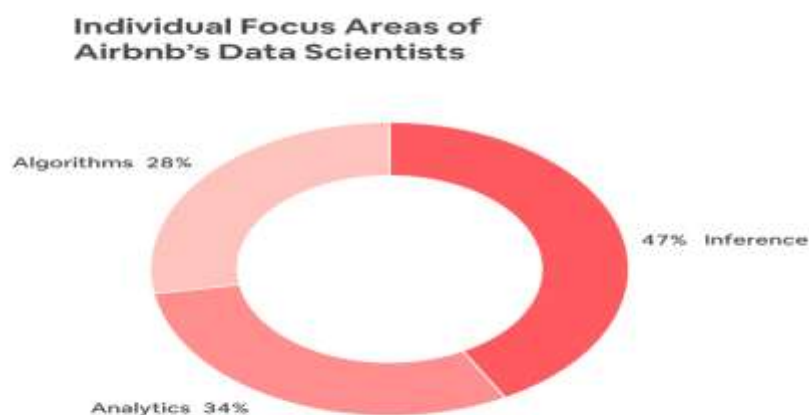
## Data Scientist – Inference

Establishes causal relationships with statistics

- Analytics - Defines and monitors metrics, creates data narratives, and builds tools to drive decisions
- Algorithms - Builds and interprets algorithms that power data products
- Inference - Employs statistics to establish causal relationships
- Foundation - Demonstrates ownership and accountability for data quality and code (*expected for all tracks*)

### Business (*expected for all tracks*)

- Ownership - Able to drive projects to success, enables others, owns impact
- Influence - Communicates clearly, demonstrates teamwork, and builds relationships
- Enrichment - Contributes to team-building through mentorship, culture, recruiting, and diversity efforts



So before starting to search for a role, it's important to determine what flavor of data science appeals to you. Based on your response to that, what you study and what questions you'll be asked will vary.

Despite the differences in the types, generally speaking, they'll follow a similar interview loop although the particular questions asked may vary.

***In light of my own interviewing experience, the rest of this presentation is strongly tailored towards the Product Analytics Data Science Interviews.***

### 3.0 Preparation for the Specific Subjects

Data science interviews, like other technical interviews, require plenty of preparation. There are a number of subjects that need to be covered in order to ensure you are ready for back-to-back questions on statistics, programming, and machine learning.

How do I prepare for the specific content that comes up in the interviews for the product, and metric driven track?

These interviews focus more on asking product questions like what kind of metrics you would use to show what you should improve in a product. These are often paired with SQL and some Python questions.

#### Sample Companies Data Science Interviews:

**Airbnb** — Product heavy, metrics diagnostics, metrics creation, A/B testing, tons of behavioral questions, and take-home material.

**Netflix** — Product-sense questions, A/B testing, experimental design, metric design

**Microsoft** — Programming heavy, binary tree traversal, SQL, machine learning

**Expedia** — Product, programming, SQL, product sense, machine learning questions about SVM, regression and decision tree

### 1. Product Sense, Experiment Design, and A/B Testing Questions

Learn, and practice how to answer the hardest product, metric, and A/B Testing questions such as 'Should we add the love button on Facebook? Which metric would you choose for this new feature? Etc.

Knowing how to design experiments and carry out A/B tests for different business use cases is an essential skill to have as a data scientist. Companies want every team member to have the capabilities to understand basic stats concepts (e.g., hypothesis testing, mean/median, variance, probability distributions, sample size calculation, power calculation, etc.), design and analyze experiments, and apply them in a business setting.

The expectation is not only that the candidate should have the theoretical knowledge for these questions, but also that they must know how to proactively use this knowledge to guide product development in a scientific way. This includes instances like how to design success metrics, set up an experiment plan, and provide timely insights to guide ramping up a test from a small pilot to 100% member base, which often also requires iterations to get the product features right for member satisfaction and desired business outcomes.

For junior-level data scientists, most companies focus more on the basic understanding of key stats concepts and the business sense candidates show when

**Kennedy Wangari || Data Scientist || AI Community Lead**

applying those concepts to real use cases.

For senior-level candidates, we'd expect them to have relevant industry experience and in-depth statistics knowledge to not only answer questions, but also drive towards solutions that create the optimized level of member experience and business impact.

*Develop product sense, create metrics, and design robust A/B tests.*

- Define, measure, and report the key performance indicators/ metrics for the product:
- What are some of the key metrics that the product might want to optimize?
- How to design robust A/B tests
- Understand what experimental design is.
- What is Applied A/B testing, and how to interpret important results statistically?
- How to validate the new feature by current users' behaviors?
- An important metric goes down, how would you dig into the causes?
- What metrics would you use to quantify the success of Youtube ads (this could also be extended to other products like Snapchat filters, Twitter live-streaming, Fortnite new features, etc)
- How do you measure the success or failure of a product/product feature
- Google has released a new version of its search algorithm, for which they used A/B testing. During the testing process, engineers realized that the new algorithm was not implemented correctly and returned less relevant results. Two things happened during testing:
  - People in the treatment group performed more queries than the control group.
  - Advertising revenue was higher in the treatment group as well.
  - What may be the cause of people in the treatment group performing more searches than the control group? There are different possible answers here.
  - Given there are no metrics being tracked for Google Docs, a product manager comes to you and asks what are the top five metrics you would implement?
  - In addition, let's say there's a dip in the engagement metric of Google Docs. What would you investigate?
  - Let's say we want to implement a notification system for reminding nurses to discharge patients at a hospital. How would you implement it?
  - Let's say at LinkedIn we want to implement a green dot for an "active user" on the new messaging platform. How would you analyze the effectiveness of it for roll out?

## **2. Programming, and Coding Questions**

A coding interview can appear during the technical phone screen (TPS), onsite, or both. There could even be multiple rounds of coding interviews during the onsite interviews, depending on the coding proficiency expected. In general, you should expect coding interviews in at least one stage of an overall Data Science interview loop.

During the TPS, the delivery of the coding interview will typically be through online integrated development environments (IDEs) such as CoderPad, HackerRank, CodeByte, CodeSignal, and even internal company solutions.

During onsite sessions, either an online IDE or a whiteboard can be used. In the current remote interview environment, the former is used by default.

**Kennedy Wangari || Data Scientist || AI Community Lead**

A company could send you a link/ URL which can be co-edited, where you are supposed to answer the questions, write simple codes in R, Python, SAS, C++, and submit online. The length of a coding session ranges from 45 minutes to 1 hour and it usually involves one or more to a maximum of 5 questions with tricky coding programs. The choice of language is typically flexible, but most candidates will choose Python for its simplicity.

When you're doing a coding challenge, it's important to keep in mind that companies aren't always looking for the 'correct' solution. They may also be looking for code readability, good design, or even a specific optimal solution.

### Example:

- Find out what's wrong in the code
- What will be the output of this code?
- How will you change this program to perform better?
- Will this code execute?
- Python Scripting

1. Fizzbuzz
2. Given a list of timestamps in sequential order, return a list of lists grouped by weekly aggregation.
3. Given a list of characters, a list of prior of probabilities for each character, and a matrix of probabilities for each character combination, return the optimal sequence for the highest probability.
4. Given a log file with rows featuring a date, a number, and then a string of names, parse the log file and return the count of unique names aggregated by month.

### • Data Manipulation Coding Questions

This type of question is not as common as the other types. They ask candidates to carry out data processing and transformations without using SQL or any data analysis library such as pandas. Instead, candidates are only allowed to use a programming language of choice to solve the problems.

In essence, this type of interview question, data manipulation, expects candidates to have a high-level understanding of logic.

- It comes down to fundamental programming in Python or another language.
- Come up with a solution and then try to find a more efficient algorithm

The most challenging part of data manipulation is to understand the logic behind the question. Even the most sophisticated programming comes down to Python fundamentals, e.g., string access and manipulation, different data types, and for and while loops, etc.

### Some common examples include:

- Representing two datasets as dictionaries, and joining them together on some given key values.
- Given a dictionary of dictionaries representing a JSON blob, doing some basic parsing to extract particular entries.

**Kennedy Wangari || Data Scientist || AI Community Lead**

- Writing a function that is similar to the “spread” or “gather” functions in R’s tidy package, and testing it using a dataset.
- Calculating a 30-day rolling profit.
- Parsing event logs and returning the count of unique strings by day/month/year.

### 3. Algorithms, and Data Structures Questions

This type of coding question aims at evaluating candidates’ proficiency in introductory Computer Science fundamentals. These fundamental topics can include, but are not limited to:

Expect questions on data structures, and algorithms. Prioritize, and learn data structures irrespective of you getting interviewed in any work domains. This applies mostly for the top companies.

- **Data Structures:** Arrays, Hashmaps/Dictionary, Heaps, Sets, Stack/Queues, Strings, and Tree/Binary Tree.
- **Algorithms:** Binary Search, Recursion, Sorting, and Dynamic Programming. Sorts, searches, insertion sort vs. merge sort, trees.

Some additional topics such as Linked Lists and Graphs (Depth First Search or Breadth-First Search) are less likely to occur during this type of interview.

Typically, multiple questions will be asked about a single scenario, ranging from simple to hard. Each question may cover a unique data structure or algorithm. Here is an example of a classic problem that revolves around finding the median of a list of numbers:

- Part 1: Find the median using any method. Candidates can use a built-in sorting function and simply return the median after sorting.
- Part 2: The interviewer now asks for a more optimized version of finding the median. In this setting, knowledge of common algorithms, such as quickselect, will come in handy.
- Part 3: Finally, the question is changed to a “streaming” version of computing medians, meaning that the data comes in an online fashion rather than as a fixed list of numbers. In this case, the candidate would likely resort to the use of heaps (slightly more challenging).

This type of question may also appear as an **applied business problem**. For such questions, the candidate is expected to code up a solution to a hypothetical applied problem, which is usually related to the company’s business model. These questions are easy to medium in the level of difficulty (based on the categorization of Leetcode). The key here is to understand the business scenario and exact requirements before coding.



- **Fibonacci.** Return the n-th Fibonacci number, which is computed using this formula:
- **Most frequent outcome.** We have two dice of different sizes (D1 and D2). We roll them and sum their face values. What are the most probable outcomes?
- **Reverse a linked list.** Write a function for reversing a linked list.
- **Binary search.** Return the index of a given number in a sorted array or -1 if it's not there.
- **Deduplication.** Remove duplicates from a sorted array

***Communication is still the King! Ask for clarification questions before proceeding to the coding part.***

## 4. Statistical Modelling/ Machine Learning Questions

Statistical modeling or machine learning skills are required for a data scientist to perform their job well. The aspects most companies look for is the candidate's ability to formalize a business problem into a machine learning problem, select the proper modeling algorithms, and build out the models following the right process of training, testing, and validation.

Typically, a data scientist on the Analytics team doesn't focus the majority of their time on modeling work. However, we have found that having knowledge of common machine learning algorithms and knowing how to apply them to specific business contexts are key to the success of a data scientist's career.

Without the proper understanding in this area, it's easy to cause incorrect interpretation of data, which may lead to imperfect decision-making or worse outcomes. The constant debates around "correlation versus causation" and "wrong data is worse than no data" are all good examples.

What is also really important is how the candidate picks the right machine learning algorithms (knowing the pros and cons from each, e.g., logistic regression, linear regression, decision-tree, deep learning, etc.) for the type of business problems he or she is solving for.

The modelling questions are more frequent nowadays, the interview questions for a product data scientist are mainly geared toward how to apply those models rather than the underlying math and theories.

Machine learning questions are all over the board: from a practical problem (how would you go about setting up the data, cross validation, modelling, performance monitoring for a given scenario) to generic questions (how do you deal with a categorical variable with high cardinality) to more theoretical.

### Expect questions around:

- Concepts on Data understanding, wrangling, exploration, transformation, visualizations, EDA, Data Cleaning, feature engineering and other fundamental skills.
- Concepts on each of the machine learning algorithms in detail.

**Kennedy Wangari || Data Scientist || AI Community Lead**



- Accuracy measurement, and practical machine learning deployment and implementation.
- Regression, and general modelling problems.
- Supervised Learning: Decision Tree, Linear Regression, Logistic Regression (using stochastic gradient descent), and K-nearest Neighbors.
- Unsupervised Learning: K-means Clustering
- Bootstrapping
- Confidence Intervals and their significance
- how over sampling works
- significance of the ROC curve and how to interpret a ROC curve
- How Random forest works
- Practical experience about Overfitting and Underfitting
- Practical experience about variable selection
- Basics of Logistic Regression

Example: How does the linear regression algorithm determine what the best coefficient values are?

The point is to see how deeply you understand linear regression, which is critical because in many data science roles you won't just work with algorithms in a black box; you'll actually put them into action. This category of question tests how much you know about what's actually happening beneath the surface.

So this is one of those "show your work" moments. Trace out every step of your thinking and write down the equations. As you're writing out the solution, describe your thought process so the interviewer can see your mathematical logic at work.

#### **Mathematics, and Statistics most asked concepts include:**

- Simulation: Monte Carlo simulations, weighted sampling, simulating Markov chains, etc.
- Prime Numbers / Divisibility: Calculations involving divisibility of natural numbers, Euclidean algorithm for computing the greatest common divisor of two natural numbers, etc.

Some common questions include:

- Estimating the value of Pi using simulation.
- Enumerating all prime numbers up to a given natural number N.
- Simulating a multinomial distribution using uniform random numbers.

### **5. Data Wrangling/ Munging/ Manipulation: the SQL Questions**

Generally, there will be at least one interview focused on SQL. In addition, the interviewers may take you through the entire process of developing a product, choosing metrics to track and then querying to measure the effectiveness of that metric.

For most companies, typically, you'll be asked to perform a series of data manipulations,

including aggregation, distribution, ordering, etc., with programming languages such as SQL, R, or Python, for a particular dataset to demonstrate their capabilities in this area. The goal is not to test the exact syntax, but rather test the right approach and thought process, and how well the candidate can make reasonable judgments based on the business context.

### Practice SQL questions:

- Leetcode database problems:
- HackerRank SQL Problems
- How would you handle NULLs when querying a dataset?
- How will you explain JOIN Function in SQL, in the simplest possible way?
- Select all customers who purchased at least two items on two separate days from Amazon.
- What is the difference between DDL, DML, and DCL?
- Given a payment transactions table and a customer's table, return the customer's name and the first transaction that the customer made.
- Given a payments transactions table, return a frequency distribution of the number of payments each customer made. (I.E. 1 transaction — 100 customers, 2 transactions — 50 customers, etc...)
- Given the same payments table, return the cumulative distribution. (At least one transaction, at least two transactions, etc...)
- Given a table of — friend1|friend2. Return the number of mutual friends between two friends.

## 6. Probability and Statistics Questions

It goes without saying that a strong grasp of statistics is important for solving different data science problems. Chances are you'll be tested on your ability to reason statistically and your knowledge of statistical theory.

At large tech companies, do expect to receive an occasional probability or statistics question. While the questions won't necessarily require complex math if you haven't thought about independent and dependent probabilities in a while, then it is good to review setting up the basic formulas

There can be questions on elementary statistics concepts like probability, probability distribution, standard statistics, hypothesis testing, mean/median, variance, probability distributions, sample size calculation, power calculation, etc.), design and analyze experiments, and how to apply them in a business setting.

Example: What is the difference between Type I error and Type II error?

Proving your mettle requires showing you understand the fundamentals of statistics. But more than that, interviewers also want to see whether you're capable of using the technical language and logic of statistics to grapple with ideas you may not often approach that way—and still communicate them clearly. So be no-nonsense in your response. Use the relevant statistical knowledge to arrive at your answer, but be as direct as possible about whatever you're asked to define.

## 7. Presentation/ Use Case/ Problem-Solving Scenario Study Project/ Questions

Being a data scientist is not an easy job, especially when it comes to the requirement of understanding business use cases really well in order to solve problems in a data-driven way. This requires good business domain knowledge, critical analytical thinking, familiarity with carrying out root cause analysis, and the ability to communicate results effectively to influence business decision-making.

These questions are meant to see how you envision your work delivering products or services from end to end. Scenario questions don't test for knowledge in every field; they're meant to explore a product's life cycle from beginning to delivery and see what limits the candidate might have at each stage of that process. But these questions also evaluate holistic knowledge—for instance, what it takes to manage a team to deliver a final product—to determine how candidates perform in team situations.

The capabilities assessed here include the ability to solve a business case with the right analytical approach and reasonable data intuition, as well as the ability to make relevant and actionable recommendations based on data insights.

The case studies could be from business domains like products, marketing, or sales, which are all based on what you would experience on a daily basis at work.

The scenario-based questions are designed to test your experience and knowledge in different fields of data science, to find out the practical limits of your abilities.

Demonstrate your applied knowledge as thoroughly as you can, and you'll come off well in any case analysis.

**Example:** If you were a data scientist at a web company that sells shoes, how would you build a system that recommends shoes to visitors?

Get to fully demonstrate the business sense knowledge you've acquired. This helps the company to identify the best sub-teams within Analytics that candidates could be allocated to.

Be honest about where you can add a lot of value, but don't be shy about where you expect to get a little bit of help from your teammates. Try to relate how your technical knowledge can help with business outcomes, and always explain the thought process behind your choices and the assumptions that guide them.

- **Presentation**

Some companies require candidates to either present the take-home assignment or a project of which they are most proud. Still, other companies would ask you about your most impactful project during behavioral interviews. However, no matter what the form the key is to make your presentation interesting and challenging.

That sounds great, but how do you do that? My main recommendation is to **think through all the details**, such as high-level goals and success metrics to ETL to modeling

**Kennedy Wangari || Data Scientist || AI Community Lead**

implementation details, to deployment, monitoring, and improvement. The little things add up to make a great presentation rather than one big idea. Here are a few questions worth rethinking to help reach your ideal presentation:

- What were the goal and the success metric of the project?
- How do you decide to launch the project?
- How do you know whether customers are benefiting from this project? By how much?
- How do you test it out? How to design your A/B test?
- What was the biggest challenge?

When presenting a project, you want to engage the audience. To make my presentations interesting, I often share **interesting findings and the biggest challenges** of the project. But the best way to make sure you are engaging is practice. Practice and practice out loud.

## 8. Behavioral Questions: Human Resource, and Competency Based Interviews

These questions are meant to test your soft skills, and see if you fit in culturally within the company.

Mainly conducted by the HR. Luckily, the questions are almost always the same, so you can prepare for them well ahead of time. One company even sent me a booklet about their “company values” and told me I’d be interviewed about how I reflect those values in my daily life.

The intent here is to identify whether the role you’re interviewing for suits your personality and temperament, and to identify why you’re moving on from a previous position.

Don't overthink it or imagine that the key here is really any different from any other type of interview: Just understand the role well, avoid talking about issues you've had in the past with specific people, and be professional when describing what you disliked and why. A data science role may call for an analytical mind, but hiring managers still want to hear what makes you passionate.

Get to understand the company’s mission, and core values. These greatly helps to answer questions.

- Why us? / What do you value most in a job?
- What have you liked and disliked about your previous position?
- Introduce yourself / Why are you leaving your current job?
- The biggest success/failure/challenge in your career. Other versions: Tell me about a time you resolved a conflict or you’ve had to convince your manager or a PM on something.

This part of the interview is just as, if not more, important than the others. It could be the person conducting this side of the interview has the last say. Have a look over some example Competency based questions such as ‘what’s your biggest weakness’, ‘what are your own standards of success’ or one of my favourites — ‘name a time when things haven’t gone your way and how did you overcome it’. The interviewers will be looking for specific examples describing exactly what you did in certain situations, not what the team’s role as a whole was, or what you would do in a hypothetical situation.

You can choose to use relevant examples from your current job, a previous role or a situation outside of work altogether. You will be asked to discuss the example in some detail. It is likely that the interviewers will then follow with some probing questions, possibly to clarify a particular point.

They will be interested in the outcome of the situation, whether there was anything you learned from the experience etc. The interviewers want a lot more than one word answers or statements. This is a real opportunity to display character and grit and what you're made of as a human being—it may not be your favourite part of the interview though do embrace it and treat it with as much respect as your Technical tests.

At the end of the day—the business has seen your profile, they know you've got a Degree in Comp Science / Stats and they know you can use Python or R—**now they need to know who you are.**

- **Soft Skills Evaluation:**

Most companies closely evaluate candidates' communication, project management, and influence skills, all of which are considered equally important.

The art of being a data scientist includes how you effectively influence others based on what you've found from the data, which oftentimes can be the hardest part in driving a data-driven decision-making culture. The types of questions we ask can include: how do you summarize your findings in a clear and succinct way, how do you handle the situation if the stakeholders are not convinced based on the analysis results, how do you respond to questions about the algorithms/methodology from people who are not technical, and how do you manage a project that isn't going as planned and turn it around? Ultimately, the goal is to take the insights generated from the analysis and effectively influence critical decision-making, which drives business impact. The "hard skills" and "soft skills" need to work together for the success of a data scientist.

## **Tips to Ace your Data Science Interviews**

Documenting best practices, actionable, and game changing tips on how to prepare effectively for the various rounds of data science interviews.

### **1. Prepare adequately for the "Why Us" and "What feature or product would you add" questions**

Most companies will ask you why you're interested in them. Get to know what features you could add to their offerings. What you'd add. Sometimes they ask for your favorite feature of their products. These are to test your knowledge and interest in the company. Come with prepared answers to both before every interview. Use the company's product before the interview if you can.

### **2. Use the company product/ services**

Go deep into the product beforehand to help yourself stand out. You can show up with a

list of ideas that could be added to their offerings. Go the extra mile. Do this for the companies that you are excited about.

Before applying to Slack, I built a Slackbot, I stayed at an Airbnb the night before the final round interview at Airbnb and interviewed my host about his experience on the platform. This will help you to express that you are really committed to the company.

### **3. Ask high-quality meaningful, and intelligent questions to the interviewer**

Have a list of well-prepared questions ready to ask.

- ✓ Do you have any concerns about my skillset or background that I can address for you?
- ✓ If you could go back in time to the day you started working at this company, and give yourself a piece of advice about the job, what would tell yourself?
- ✓ If you had to give your company's data infrastructure/cleanliness a grade (A-F) what would it be and why?

This question allows you to perceive 2 things.

(1) To what extent might my work devolve into data engineering tasks and

(2) How mature is the company? Is there a lot of technical debt? I've heard that private equity firms often ask this question in a circuitous way by requesting some kind of analysis of the company they're investigating; if it takes the company longer than a day or two to respond to the request, they infer that the company's data infrastructure is lacking."

The first allows you to play offense a bit. It shows you're confident in yourself. The second tends to be one they don't have a prepared answer for, and usually teaches you something valuable about the company and the job. Even if you don't end up working at that company, you tend to get some good advice for wherever you do end up working!

## **How You Are Evaluated During the Data Science Interviews?**

There are **4** major qualities you want to convey during your interview.

Logical Reasoning:

The interviewer wishes to see candidates make logical connections between the information provided and the ultimate answer. You should therefore describe clearly what is needed for the computation and how you would write the code to solve the problem, before diving into the actual coding.

Communication:

The effectiveness of your communication matters significantly. Before coding, clearly communicate your thought process. If the interviewer asks questions at any point during the interview, you need to be able to explain the reasoning of your assumptions and choices.

Code Quality and Best Practices:

The interviewer will also evaluate your overall code quality. While the standard expectations in a DS interview would not be as high as those in a software engineering interview, candidates should still focus on several aspects:

- Whether the code is executable without any syntax error.
- Cleanliness and conciseness.
- Whether the solution is optimized in terms of run-time/storage efficiency.
- General coding best practice, e.g. modularity, handling of edge cases, naming conventions, etc.

Proficiency:

Just as with software engineering coding interviews, for DS coding interviews, it is reasonable to expect multi-part questions and sometimes multiple questions. In other words, speed is also important. Being able to solve more questions within a limited amount of time is a signal of overall proficiency.

## Top Mistakes Data Scientists Make in Interviews

You've already done the hard weightlifting work to get to this step, so now it's time to finish strong.

This list contains top mistakes that are stumbling blocks during the various rounds of interviews.

1. Being unprepared to discuss your projects during interviews.
2. Underestimating the value of business domain knowledge.
3. Neglecting communication skills.
4. Not asking enough/ smart questions to the interviewers.
5. Not thinking in a structured manner
6. Discussing the same past project similarly for different questions, and situations.
7. Assuming interviewers know my past experiences
8. Not considering the business impact of your models.
9. Avoid jargons or concepts you're unsure of.
10. Not overviewing the whole Data Science process

- **Being unprepared to discuss projects.**

Having projects in your portfolio serves as a major safety net for "how would you" type interview questions. Instead of speaking in hypotheticals, you'll be able to point to concrete examples of how you handled certain situations.

In addition, many hiring managers will specifically look for your ability to be self-sufficient because data science roles naturally include elements of project management. That means you should understand the entire data science workflow and know how to piece everything together.

**To avoid this mistake:**

- Complete end-to-end projects that allow you to practice every major step (i.e. Data Cleaning, Model Training, etc.).
- Organize your methodology. Data science should be deliberate, not haphazard.

**Kennedy Wangari || Data Scientist || AI Community Lead**



- Review and practice describing past projects from any internships, jobs, or classes you've taken.

- **Underestimating the value of domain knowledge.**

Technical skills and machine learning knowledge are the basic prerequisites for landing a data science position. However, to truly stand out above the competition, you should learn more about the specific industry you'll be applying your skills to.

Remember, data science never exists in a vacuum.

**To avoid this mistake:**

- If you're interviewing for a position at a bank, brush up on some basic finance concepts.
- If you're interviewing for a strategy position at a Fortune 500, practice a few case interviews and learn about drivers of profitability.
- If you're interviewing for a startup, learn about its market and try to discern how it will gain a competitive edge.
- In short, taking a little bit of extra initiative here can pay big dividends!

- **Neglecting communication skills.**

Currently, in most organizations, data science teams are still very small compared to developer teams or analyst teams. So while an entry-level software engineer will often be managed a senior engineer, data scientists tend to work in more cross-functional settings.

Interviewers will look for your ability to communicate with colleagues of various technical and mathematical backgrounds.

**To avoid this mistake:**

- Practice explaining technical concepts to non-technical audiences. For example, try explaining your favorite algorithm to a friend.
- Prepare bullet point responses to common interview questions and practice delivering your answers.
- Practice analyzing various datasets, extracting key insights, and presenting your findings.

- **Not Thinking in a Structured Manner**

When people come across case studies, guesstimates and puzzles during their data science interview, the first instinct is to jump the answer. There's not much thought behind how to structure your thoughts – a big no-go for an interviewer.

**Example:** Suppose you have recently joined a transport company as the CEO. The company has been posting heavy losses recently. How would you go about turning around the situation?

Most interviewees start by listing off ideas, like "analyze the pricing", "look at the overall costs", "look at the route planning", etc. That is the absolute wrong way to go about things! Consider this a rejection for sure.

Lay down a framework using pen and paper (or a whiteboard). Putting a structure to your thoughts showcases your thinking ability – a must-have skill in data science.

- **Not Communicating your Thought Process with the Interviewer**

Focusing too much on the answer and not on the process is a sure shot way to failing your data science interview.

Think about it for a moment – the aim behind an interview is to help the interviewer understand the thought and reasoning behind your problem solving skills. Right? The interviewer does not care much about if your answer is precise to the decimal point.

There might not even be a right or wrong answer in the first place! So make sure you communicate your thought process to the interviewer, including the assumptions you are making. It's a win for both sides.

- **Blind use of machine learning libraries**

When you are given a case study, you often have an advantage you can capitalize on: you choose the model(s) to use. That means that you can anticipate some of the questions interviewers might ask you!

For example, if you end up using an `XGBClassifier` for your task, try to understand how it works, as deeply as you can. Everyone knows it's based on decision trees, but which other "ingredients" do you need for it? Do you know how XGBoost handles missing values? Could you explain Bagging and Boosting in layman's terms?

Even if you end up using linear regression, you should have a clear idea about what is happening under the hood, and the meaning behind the parameters you set. If you say "I set the learning rate to X", and somebody follows with "What's a learning rate?", it's quite bad if you cannot at least spend a few words on it.

- **Not considering the business impact**

You must always make sure to consider the business impact of your model. I believe this mistake is specifically popular for Data Scientists because they focus solely on the model and its performance, but fail to mention how the business was impacted.

You want to highlight your results in a way that is accurate but more importantly, *impactful*. You most likely have impactful results but you may have failed to let the interviewers know this point. You can phrase your answers like this:

- *"I worked on the Decision Tree model that automated a manual process, saving the process 50% time, and 50% money, creating time and money for bettering that product."*

If you refer to your 99% accuracy constantly but fail to mention its impact, you can expect that the interviewers will think you do not know how to work within a business and are more educational oriented. Sometimes even lower accuracy is better if the overall process is faster and more impactful in some way.

Pretend like you were hiring someone—you would want to know they can help your business.

- **Not Overlooking the whole Data Science Process**

Strive to include a discussion around the Data Engineering and Machine Learning components that happen before and after a main Data Science project.

***The interviews want to know:***

- How you got the data
- how you preprocessed it
- How the model was changed to object-oriented programming format
- How tests were made
- How it was deployed
- How it was integrated into your product

These points may not be something you have performed yourself, but the interviewers know that every Data Scientist is not also a Data Engineer, Machine Learning Engineer, or Software Engineer for that matter. What they are testing you on is if you were aware of the whole process from start to finish and who worked on what. If you answer this question correctly, then the company will see that you fall into a more specialized Data Science role, and could possibly learn the beginning or end parts of the Data Science process.

- **Not asking enough questions to the interviewers**

Not asking enough questions shows a few things:

- You are not interested in the company
- You did not pay attention well enough to come up with a question
- Could show you are overconfident
- Could show you are hard to work with

No one wants to work with a know-it-all that is overly confident. Similarly, no one wants to work with someone who is not interested in the company or its goals and respective Data Science projects. Most importantly, when you do not ask enough questions, it sounds like you did not care enough to listen. Ultimately, what you are portraying as you fail to ask questions to the interviewers, is that you would not be a good candidate to collaborate with.

- **Assuming interviewers know my past experiences/ describing your projects work**

What you are doing with interviewers is just that — storytelling. A common mistake I have made was assuming the interviewers knew some information about the background of my project.

*They most likely will know nothing — some will not even read your resume.*

You'll need to set the scene when answering a question by providing basic information for the Data Science project you performed. This type of explanation will show the interviewers that you can work with stakeholders and other non-Data Science people.

**Some key points to bring up are:**

- What was the business problem?
- Or why did you want to do the project in the first place?
- Who was involved? (Product Managers, Software Engineers, etc.)
- What was the process?
- How did you do it?
- Where in the grand scheme of the business did this project fit in?
- What were the results?
- How were they perceived?
- How many people did you help/money did you save/time to you save?

Once you outline your past projects in this format, it will better paint the picture of your answer.

- **Don't Try to Fumble Out Answers for a Topic you haven't Studied Before**

This is a very common mistake people have. There will be certain questions you won't know the answer to. That's ok – it's human to not know everything. But candidates still try to answer these questions by making up answers on the spot. This isn't a great look.

Let the interviewer know that you aren't an expert on the topic. Highlight a way to solve the problem which you already know. For example, if you don't know how a boosting algorithm works, you could solve the same problem using a technique you do know. And later, learn boosting and get back to the interviewer.

This shows two important things to the interviewer:

- Your ability to solve a problem by looking at different methods
- Your willingness to learn and apply yourself

- **Expecting only Technical Interviews as part of the Interview Process**

There are several layers to a data science interview. The process isn't one-dimensional! Just preparing for questions around tools and techniques will not land you the role. Of course these technical skills matter, but there are other equally important topics you will be judged on.

There will be an interaction with the project team, a case study related to the domain, role plays, and much more. You should prepare for all these formats.

## Example Questions I Got in Interviews

- Here's experiment data. Analyze it. What do you think?
- Here's a bunch of data. Build a model to predict this metric. Why'd you do it that way?
- Here's a scenario. What dependent variable would you pick? What experiment would you run to measure impact?

- Here are some schemas for a database. Write SQL to answer this question with this data.
- Multiple choice test which required calculating answers from data quickly
- Basic algorithms + data structures questions, big-O notation, etc.
- A management consulting / business school style business case
- Here's a Data Science scenario. What questions would you want to ask of data?
- "Here is information about our users. Find the features that lead to increased viewership, write it out in code, and present it to us."
- **My favorite interview question:** "Pick a Data Science project you did, and let's walk through it in detail"
- Fit interviews, talking about the product, my interests and my working style
- Didn't get any probability questions or code white boarding, but many peers did

## Interviewer: Do you have any Questions for Me???

### Sample Questions to the interviewers

Connect, Culture, Challenges and Close

- Ask valuable, intelligent and insightful business questions.
- Get to understand the company and what drives them. Why do they need a data scientist, and how can you add value to their functions and operations in those terms.
- What areas do they see where you can drive efficiency, augment people's decisions with machine learning and data science, improve pricing or revenue then show them how your skills apply to those problems.
- Demonstrate how your skills and value will apply to those business problems
- What does the future of this department look like?
- The best value comes from the questions you ask the business as it shows you are understanding the problem
- What's the company's biggest challenge this year and how this job will help overcome it?
- How will I measure performance so I know I'm having a positive impact?
- If there were some skills or experience that you would wish I had, what would they be?
- Who is the most successful data scientist hire and why?
- Who didn't succeed as a new hire and why did they fail in their role?
- What's the company's biggest challenge this year and how this job will help overcome it?
- How will I measure performance so I know I'm having a positive impact?
- If there were some skills or experience that you would wish I had, what would they be?
- What are the next steps in the process?
- How can my skills solve your problems?
- Tell me about yourself?

## The Technical Screening Interview

### Sample Interview Outline

Kennedy Wangari || Data Scientist || AI Community Lead

- Make me understand why you are fit for this job and the company. Ensure you have your details ready, not just general statements that apply to any company.
  - Showcase how you reason through a problem. A perfect solution isn't possible, but I care more on your thoughtful process.
  - Do ensure to demonstrate your ability and focus on communicating your work to technical and non-technical audiences. It's great to be deeply technical but how do you translate that to others.
  - Ensure you know your interviewers before-hand. Their profiles, names beforehand and ask them questions about themselves and what they do. ***It all goes a long way.***
  - Illustrate that you know your strengths and weaknesses. We aren't good at everything.
- 
- **Outline of the interview:** What happens in the next one hour
  - **Introduction:** I tell a bit about what I do-and then ask the candidate to do the same
  - **Ask briefly to describe the projects** the candidate is dealing with at the moment
  - **Then I ask to pick one of the projects** and dig deeper
  - The candidate's role in the project
  - How it started and why, the problem it solves and who drives it-the candidate or someone else
  - Why certain models were picked
  - How one of these models work, I cherry-pick some of the technical details
  - Model deployment, how the tech stack looks like, who takes care of the deployment
  - How communication with the team happens, also communication with PM and other teams.
  - How much hands-on work vs non-hands on work
  - Next I move to give a simple coding exercise. More practical than a typical LeetCode problem. The level of LeetCode problem is easy.

## General Preparations for the rounds of Data Science Interview

While preparing for the set of interviews, I focus on 2 things:

- The product offering of the company
- What I would want to know about them

- Check Glassdoor Interviews for the specific company's data scientist interviews, and the coding questions.
- Conduct a thoroughly research about the company, read, and inquire internally on ongoing data science projects, research resources, articles, and ongoing work related to the role you are was interviewing for.
- Go through the Cracking the Coding Interview for coding questions, tackle mock up data science interviews, and practice how to answer the behavioral questions using the CARL technique. (Context, Action, Result, Learning)
- Prepare questions to ask the interviewing team with special focus on their business operations, functions and the role. It shows you understand the problem.
- Practice how to answer the behavioral questions using the CARL technique. (Context, Action, Result, Learning)
- Conduct research, and more profiling to the backgrounds of the interviewing team (their profiles, any public works shared, articles, interviews, and AI Podcasts. This helped me greatly to know what perspective they have on the job/ tasks, what they do, and to prepare adequately before-hand questions to ask about themselves and what they do.
- Conduct a thorough, intensive deep research on the products, role, and the company, its goals, challenges, and initiatives, and ongoing projects for the next 6-12 months. This provides insights on how you could plug in, and potential data analytic use cases, and challenges from the perspective of the teams and people involved.
- If interviewing for a product data scientist role, do greatly interact, and utilize the company products/ service. Gather, and share insightful, valuable feedback from your experiences during the interviews. This provided a great baseline for our engagements, and discussions with the recruiters.
- Get to prepare intelligent, and valuable business questions to ask the interviewing team. The best value comes from the questions you ask the business as it shows you are understanding the problem.
- Get to know more info about the workplace in advance (review Glassdoor)
- Get to know your **resume inside out** as you may be asked to take them through it. Walk the interviewers through your resume.
- Know what your expectation of salary, benefits, bonus and commission
- Try to find out what fair market value is, or ask your recruiter.
- Be ready to explain your past work and projects and their impact.
- Which techniques(relevant to data science) did you use in your scientific work
- What languages are you familiar with(Python, C++, Flask for Web Development)

### **Learn as much as you can about the specific position**

Look over the description of the position and try your best to find out what you would be doing. The type of position will heavily influence what kind of questions you would be getting during the interview.

#### **Will you be:**

- **Designing and interpreting experiments to test variants of the product?**

Expect some questions regarding A/B Testing, questions regarding which metrics would be best to optimize, and questions about how to best evaluate your experimental results.

- **Doing deep dives to understand more about how users use your product?**

Expect questions that test your ability to carry a data project from end to end and to effectively and faithfully communicate my findings. Expect to discuss projects from previous experiences or your education and communicate what you were able to find and what you did.

- **Doing applied research on inference, prediction, or optimization problems?**



These positions are a lot more custom and may require a PhD. Recommended: do read through the job description to see what they might be looking for, and studying up on academic techniques to solve some problems that the team you're interviewing for maybe facing.

- **Developing algorithms for a data product?**

For example Uber's Surge Pricing feature or LinkedIn's People You May Know Feature. Depending on your specific role, you may get a traditional software engineering interview with a focus on processing large amounts of data, or be asked about your previous experience solving large-scale, difficult, and custom data problems.

There are many more roles of a data scientist, do your research on both the product and the role before you set foot in the interview room.

***The key question to be asking yourself is: within my role at the company, what is the best way to understand and improve the product and the business using data?***

**Sample:**

***I want to work for you long term.....*** sell yourself answering this question. Give reasons of how your passions and skills align with the company's business

Data Science is growing. So am I: Be prepared to explain how you are keeping up with the latest/ greatest insights on Data Science, where do you think the area will be in 3 years?

Read into the backgrounds of the people that will be interviewing you so that you can know what perspective they have on the job/ task.

"You don't just find and get a great job. You find and win a great job against a pool of very competitive candidates who may want that job as much, if not more, than you do. Finding and winning a great job is a competitive sport that requires as much career athleticism and perseverance as making it to the Olympics. You must be in the finest career shape possible in order to win."