Data Scientist, Analytics, Intern
Screen Interview Guide

Welcome to your prep guide for your screen interview at Meta. Our data scientists and recruiters put together this guide so you know what to expect and how to prepare.

Our data scientists are embedded in product teams. This means they are expected to be co-owners of their product with the product and engineering partners, and are accountable to drive product outcomes.

As you're preparing, consider how this impacts your thinking as a data scientist. It might be helpful to keep in mind the core work of our data scientists:

- Use quantitative tools to uncover opportunities, set team goals and work with cross-functional partners to guide the product roadmap.
- Explore, analyze and aggregate large data sets to provide actionable information, and create intuitive visualizations to convey those results to a broad audience.
- Design robust and informative experiments, considering statistical significance, sources of bias, target populations and potential for positive results.
- Collaborate with engineers on logging, product health monitoring, and experiment design & analysis.
- Partner with data engineers on data infrastructure - tables, dashboards, metrics, and goals.
- Drive product decisions via actionable insights.

Interview Overview:

The structure of your interview

The interview will be 45 minutes:

- Introductions (5 minutes): The interviewer will first introduce themselves, explain what they do at Meta, and explain the internship opportunity. They will also ask you to introduce yourself and questions about your prior projects or internship experience.
- Analytical Case Study (20 minutes)
- Technical Analysis (15 minutes)
- Ask Us Anything (5 minutes): The last 5 minutes are your time to ask questions. This is a great opportunity to express your interest in Meta as well as ask questions about the internship, the broader team, and Meta in general.

You will be asked behavioral questions that draw from your previous experiences/coursework, as well as hypothetical questions on things you might encounter at Meta.

During the interview, you'll share your solutions on your computer via CoderPad, an online code-sharing interface. Prior to your interview, please familiarize yourself with the CoderPad interface; you can check out CoderPad's tips and tricks as you're preparing. Please also make sure that you have strong phone and internet connectivity in a quiet zone.

You will be with a Data Scientist one-on-one. Discussion with the Data Scientists is encouraged throughout the interview! If you have specific questions related to next steps or our decision timeline, please direct those to your recruiter.

How to Prepare:

During this interview you will be asked a series of questions that assess your technical skills, analytical skills, and product sense. **Strong Programming Skills, Sharp Analytical Thinking and Deep Product Sense are the three important ingredients for a Data Scientist to be successful at Meta.**

Analysis Case:

For the analytical questions, the interviewer is trying to understand how you solve business questions and problems, as well as how creative and articulate you are at thinking through these problems while solving them. It's not about arriving at the perfect or correct answer, but how you engage with the problem.

Spend some time engaging with Meta's products less as a user and more as someone who is tasked with improving or developing these products. This link outlines what we consider a "Meta product" (Facebook, IG, Messenger, Ads, Mobile, Timeline, News Feed, etc.). Note: It is not a complete list, and the analysis case may or may not refer to one of the products listed.

Put yourself in the shoes of the product team who built the product / feature:

- Why do you think they made certain decisions about how it works?
- What could be done to improve the product?
- What kinds of metrics you'd want to consider when solving for questions around health, growth, or the engagement of a product?
  How would you measure the success of different parts of the product?
  What metrics would you assess when trying to solve business problems related to our products?

How would you tell if a product is performing well or not?
How would you set up an experiment to evaluate any new products or improvements?

Aspects of these questions include:

- Thinking about which data sets are best suited to answer a product question
- Drawing inferences from a data set
- Combining multiple signals into a data-informed statement
- Mapping analytical insights back to product impact

Technical Analysis:

During the technical component of your interview, your interviewer will be assessing your ability to translate a high-level question into an execution strategy, and explain how the result is relevant and what aspects may still be lacking. The analytical case study will focus on questions to gauge your product sense and analytical skills.

- You will be given 1-2 data processing questions during this portion
- Answer the questions in SQL, Python, or R, whichever language you feel most comfortable with; we do not have a preference!
    - *Note that several of our data processing questions are designed to be solved in SQL (with tables and columns as input), but we accept any programming language that can get to the solution.*
- This will be a hands-on technical investigation of data problems.
- We are looking not only for coding skills, but also for the ability to design an operational approach to figure out a concrete answer to a specific question using data.

Sample Questions:

- Given timestamps of logins, figure out how many people on Facebook were active all seven days of a week on a mobile phone.
- How do you determine what product in Facebook was used most by the non-employee users for the last quarter? [Required parameters will be given]

Note, the actual data presented in the technical question will vary based on the question being asked, you can expect to see event-level data, dimension-level data, or both. This data set is designed for interview purposes and is not representative of the large data sets we work with at FB. Below is a mock data set and some questions. This example doesn't pertain to Meta but is representative of the data and questions you may see.

Mock Data and Questions:

Here's a couple of examples that don't pertain specifically to Meta but are representative of the kind of data and format with which you'll be presented:

Event-level data: an attendance log for every student in a school district date |

student_id | attendance

Dimension-level data: a summary table with demographics for each student in the district
student_id | school_id | grade_level | date_of_birth | hometown

Using this data, you could answer questions like the following:

- What was the overall attendance rate for the school district yesterday? ● Which grade level currently has the most students in this school district?
- Which school had the highest attendance rate? The lowest?

You will be expected to write code that would answer the data processing questions given based on a schema or set of schemas that will be provided to you. Whether you choose SQL, Python, or R, please follow standard coding style and best practices for easy readability.

Using SQL

- You may work in whatever dialect you like, but you'll be able to answer all questions with ANSI-standard functions (think PostgreSQL). If you use a dialect-specific syntax, you may need to explain it to your interviewer.
- Try to maintain a consistent capitalization / indentation style for readability.

Using Python or R

- Given the heavy focus on data manipulation, most people choose to use libraries, such as Pandas / NumPy in Python or dplyr in R. It's possible to solve the questions in pure Python/ R (or any Turing-complete language), but doing so will likely be much slower and more difficult.
- The interview will either be on a whiteboard or in a plain text environment, so there'll be no access to function autocomplete or help documentation.
- A few small mistakes in syntax won't automatically disqualify you, but pseudocode or a general explanation isn't acceptable. Be sure that you know the function names, input arguments, etc., to implement the core skills listed above.
- Interviewers generally know the most widely used libraries, but there's no guarantee that any individual interviewer will be familiar with the libraries you're mentioning. That's OK, but you may have to provide more guidance / context around what you're writing.
- If you don't have much experience performing joins and aggregations, you may wish to review that functionality for Python or R

In general, you can expect be assessed on on some subset of the following:

- Work with aggregate functions (count, sum) and be able to compute percentages with multiple aggregate functions in a single SELECT statement
- Utilize different types of Joins (IE: Left, Inner, Outer, etc.) including joining datasets that may not be on the same grain
- Utilize Union and Union All
- Work with concepts including Distinct, Random Sampling, De-Duplication, Optimization

7 Tips for Your First Round Interview:

1. **Have a clear picture first.** Ask clarifying questions about the product or situation if needed, before jumping straight into an answer. Not having a clear idea on what you want to do, or changing your mind halfway through, could slow you down.

2. **Think out loud.** It helps your interviewer follow along and learn about your problem-solving skills. They'll want to understand why you're making certain decisions.

3. **Don't worry about things like typos, syntax, etc.** Some people have the right code sitting there, but they take another minute or two going over everything manually trying to check if it's all correct. Feel free to hit "run" to see what error messages come up.

4. **If you aren't sure about syntax.** If you aren't sure about syntax, just ask the interviewer: "What's the syntax to check if an element is in a list again?"

5. **Be open.** Interviewers are asking questions that are not meant to stump you, but rather they're meant to get to know your skills and gauge how in-depth your knowledge is. It's OK not to know something. Communicate openly with the interviewer, approaching questions as a chance to collaborate with them.

6. **Listen for hints.** Interviewers might ask questions such as, "Are you sure you want to use a loop there?" This gives you an opportunity to consider alternatives to how to solve the problem, and to demonstrate that you can learn on the spot and implement feedback.

7. **Questions.** You'll most likely have some time at the end for questions for your interviewer. Some people find it easier to come up with a few questions in advance rather than thinking of them on the spot.

Appendix / Resources:

While we recognize that people interviewing with Meta may utilize outside sites / resources with posted interview questions and answers, we encourage caution, as we've found many of the proposed solutions to be incorrect. Our data scientists and recruiters have compiled these trusted links for you to review while you prep

Technical prep

- SQL Course
- Programmer Interview SQL Practice Database

- Mode Analytics SQL Tutorials
- HackerRank

Analytical prep

- How Experimentation Informs Product Development: LinkedIn
- The Pitfalls of A/B Testing in Social Networks
- Summary of Udacity A/B Testing Course
- Khan Academy Statistics & Probability Course
- Cracking the PM Interview by Gayle Laakmann McDowell

About Meta and Data Analytics

- Meta News
- Meta Data and Analytics Team
- Meta Career Stories
- Meta Analytics Blog
- Medium: I'm a Data Scientist - Here's Why I Work at FB: Part 1 + Part 2
- VP of Analytics Alex Schultz's Talk at Stanford on Growth at Facebook
- How We Shipped Reactions

Additional Resources:

- Experimentation in Data Science