



Data Science Career Track

Student Interview Guide

(General Data Science)

Student Version
April 2021

This document is the property of Springboard and is provided to you as a Springboard mock interviewer. By accessing, downloading or reading through this document you agree to not use or permit anyone else to use this document or any materials in it for any purpose other than your personal learning.



Table of Contents

Mock Interview Overview	2
Mock Interview Format & Grading Criteria	3
Introductions & Interview Expectations: 2 mins	3
Problem Introduction and Understanding: 5-8 mins	4
Solution Approach Explanation: 35 mins	4
Follow-up Questions: 10 mins	5
Feedback and Conclusion: 10 mins	5
Appendix – A: Sample Open-ended Problems	8

Mock Interview Overview

The **general data science interview** is a scenario-based interview that consists of technical and business components. This is a critical interview for data scientists where candidates must explain in detail how to solve a real-world business problem. They must draw in elements from business context, understanding, relevance, technical solutioning, and system design.

The **general data science interview (also known as the ML problem solving)** typically consists of the interviewer asking you to present a solution to a given data science problem related to a real-world business problem.

The interviewer will give you an open-ended problem based on a real-world scenario that would require you to discuss in detail your solution approach to the problem with a key focus on the following areas:

- Business understanding & problem formulation
- Solution approach & workflow
- Technical depth & breadth
- Scenario handling & system design

The general data science interview is a 60-minute interview with scope for you to understand a given open-ended business problem, articulate your thoughts, formulate and explain your solution approach, and also answer necessary questions asked by the interviewer.

Mock Interview Format & Grading Criteria

The 60-minute general data science interview focuses on both business problem understanding and technical depth and breadth of the data science solution workflow explained by you for a given open-ended data science problem.

The given problem will most likely incorporate elements of data cleaning, exploratory analysis, feature engineering/selection, and machine learning.

If the given problem has elements of natural language processing, computer vision, or deep learning, the focus will be more on the overall data science workflow. Remember: When working for an organization in the real world, you can encounter any problem as a data scientist!

Interview Format:

The structure of the 60-minute Interview has the following breakdown:

Topic	Time Allocated
Introduction & Interview Expectations	2 minutes
Problem Introduction & Understanding	8 minutes
Solution Approach Explanation	30 minutes
Follow-up Questions	10 minutes
Feedback and Conclusion	10 Minutes

Introductions & Interview Expectations: 2 mins

This will be a short introduction where the Mock Interviewer introduces himself/herself to you and provides an overview of the structure of the general data science interview. This will set the expectations of the interview, and there will be time for any questions you may have before starting the interview.

Problem Introduction and Understanding: 5-8 mins

For this segment of the interview, the interviewer will either verbally explain the open-ended problem scenario or send a one-pager that covers the details of the problem.

The problem will be defined along with some necessary background information and assumptions. For the next 5-8 minutes, you can process the given information and think about your solution approach. You can also ask the interviewer any questions relevant to the problem. These questions should help you solidify your understanding of the problem. This phase is akin to the business problem understanding step that typically happens between the data scientist (you) and the client/ domain experts (interviewer).

After this step, you should start explaining your solution approach. You are still free to ask some relevant questions when explaining a particular step in your approach – in case you feel you might be lacking additional information about the problem.

For a few sample problem examples, please refer to Appendix – A

Solution Approach Explanation: 35 mins

For this segment of the interview, you should explain your solution approach and workflow to design and build a system to solve the given business problem. You are not expected to code anything but verbally explain how you would approach the problem and design a solution. You are free to use a virtual whiteboard if you desire but the key focus remains on your ability to explain your approach clearly.

The workflow described for your solution should closely align with the Springboard Data Science Method taught in the curriculum. If necessary, there may be a few minor additions and changes depending on the open-ended problem as depicted in the following figure.



These following make up the four major phases of the workflow that you would also be evaluated on, i.e.:

- *Problem Identification & Formulation*
- *Data Cleaning & Understanding and EDA*
- *Modeling & Evaluation*
- *Findings & Recommendations*

You need to ensure your approach and workflow closely align with the Data Science Method and, in turn, the CRISP-DM workflow, which is an industry standard.

You might be asked technical or business-relevant questions at each step to test your technical breadth and depth.

Follow-up Questions: 10 mins

For this segment of the interview, the interviewer could choose to go in-depth to question the technical decisions taken by you for their solution workflow and ask you about how you would design a system so that the business can start using their proposed model technique. You may also be asked alternate scenarios in the form of “what if” questions.

Feedback and Conclusion: 10 mins

Typically, the interviewer will spend the last **5-10 minutes** giving you necessary feedback on your solution approach.

Grading Criteria and Rubric:

The mock interview will be graded on a **4-level scale** split across key criteria that check for professional data scientist competencies in line with business hireability and market expectations. The grading criteria is:

- **Level 1:** Not yet meeting expectations
- **Level 2:** Approaching expectations
- **Level 3:** Meeting expectations
- **Level 4:** Exceeding expectations

General Data Science Rubric:

The final decision for each technical interview is an accumulation of your performance for each of the following topics:

Grading Criteria	Not yet meeting expectations	Approaching expectations	Meeting expectations	Exceeding expectations
Business Understanding & Problem Formulation	1	2	3	4
Solution Approach & Workflow	1	2	3	4
Technical Depth & Breadth	1	2	3	4
Scenario Handling & System Design	1	2	3	4

Total Score:

Fail (score<=10)		Pass (Score>10)	
Level 1: Not yet meeting expectations	Level 2: Approaching expectations	Level 3: Meeting expectations	Level 4: Exceeding expectations
4-7	8-9	10-14	15-16

Special Case: Any candidate who scores “Not yet meeting expectations” (score = 1) in ANY of the sections EXCEPT the *Scenario Handling & System Design* section, automatically fails the general data science interview.

This would indicate that you:

- Are unable to understand the business problem and formulate it as a data science problem **OR**
- Showcase below-par understanding of how to approach the problem concerning the major steps in the data science lifecycle **OR**
- Lack the necessary technical rigor and depth concerning approaches and methodologies used in your solution approach.

Appendix – A: Sample Open-ended Problems

This section covers a couple of examples of open-ended problems that can be given to candidates during the interview.

Verbal Example: This is an example problem that is typically spoken out verbally to the candidate. Then, they can ask relevant questions and proceed with the interview.

“Every day, we process hundreds of millions of emails. First, we need to know whether each email is SPAM or NOT SPAM. Your task is to develop a method that determines whether or not an email is spam or not. You are free to assume that we already have labeled data.”

While this task involves natural language processing, you do not need extensive knowledge in NLP or need to be an NLP expert. Focus on the end-to-end process of building such a system (classification model)

Other Examples:

- <https://www.kaggle.com/blastchar/telco-customer-churn> – Given a dataset for telecom customers, design a system to predict which customers might churn.
- <https://www.kaggle.com/c/bike-sharing-demand/data> – Given a bike-sharing dataset, design a system to predict bike demand.