Data Science Career Track Student Interview Guide (Coding)

Student Version
April 2021

Springboard

Table of Contents

Mock Interview Overview

The **coding interview** will contain hands-on problems related to data management and basic data structures leveraging Python and SQL. This interview is critical for data scientists to demonstrate their coding competencies. You will code working solutions to problems.

The coding interview is relatively objective compared to the other interviews. Coding exercises in this interview **consist of two levels: Level 1 and Level 2**.

- **Level 1**: An exercise *directly related* to data transformation and analytics; Level 1 is typically a data management problem consisting of two parts:
    - A challenge that involves the use of Python
    - A challenge that involves the use of SQL
- **Level 2**: An exercise that will typically focus on more advanced questions in Python OR SQL, depending on your specialization and job focus:
    - Students taking the **Business Insider track (Specialization 2)** and focusing on analyst roles will receive a challenge on advanced SQL.
    - Students taking the **Generalist or Advanced ML tracks (Specialization 1 or 3)** and focusing on data science roles will receive a challenge on data structures and algorithms with Python.

You will interview through an online coding platform (coderpad.io), which will give you a sense of how an actual coding interview plays out. This objective interview process contains three criteria that the candidate will be evaluated against:

1. Working (and optimal solution) for Level 1 — Python
2. Working (and optimal solution) for Level 1 — SQL
3. Working (and optimal solution) for Level 2 — Python \ SQL

The interview grade will be on a **4-level scale** split across these key criteria. The coding exercises should meet the following guidelines:

- **Level 1** — The Python challenge will require you to work on sample data — transforming and analyzing it to extrapolate basic insights. Key focus should be on a problem related to data science (see appendix for sample problems).
- **Level 1** — The Python challenge should be coded using the Python standard set of libraries (base python). Third-party packages like numpy, pandas, etc. ***are not allowed***. This requirement's sole purpose is to ensure you are prepared for real-world interviews and ***not*** to make it more challenging for you on purpose.
- **Level 1** — The SQL challenge should focus on problems that are relevant to gather basic insights for a data science problem (e.g., aggregations, filtering, joins).
- **Level 2** — Given the 60-minute time frame, exercises at this level are not a required completion for passing the interview. You, however, can attempt these exercises to strengthen your skills.
- **Level 2** — The Python challenge will focus on a data structures and algorithms problem (see appendix for sample problems).
- **Level 2** — The SQL challenge will focus on advanced SQL constructs (self-join, window functions, CTEs, etc.)

Mock Interview Format

The coding interview will be a 60-minute interview with a focus on both Python and SQL challenges. This timed format exposes you to both coding against the clock and working and optimal solutions to given problems.

You should code working solutions to Level 1 challenges as a minimal requirement to pass the interview. A bonus will be awarded if you can also complete the Level 2 challenge in the 60-minute span.

Coding interviews will use the site: http://coderpad.io.

In general, it is ok to enable code execution for the Python problems to give you a way to debug your code, though parts of it may entail whiteboarding.

SQL questions will have you write SELECT statements. These questions will primarily require whiteboarding — no database will be created in advance to run queries live.

Interview Format:

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

Topic	Time Allocated
Introduction & Interview Expectations	5 minutes
Level 1: Python & SQL	35 minutes
Level 2: Python or SQL	15 minutes
Feedback and Conclusion	5 Minutes

Introductions & Interview Expectations: 5 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 coding interview. This will set the expectations of the interview and there will be time for any questions you have before starting the interview.

Basic expectations for passing the interview include:

- Minimal requirements:
  - The ability to code correct working solutions.
  - Level 1 Python and SQL.
- ***No third-party libraries, like numpy and pandas, are allowed in Python challenges.***
- In line with actual coding interview practices, Google search or any web search or reference material lookup ***is not allowed.***

Before the interview begins, you should clearly state your specialization and the roles you will be targeting. This information determines if the Level 2 challenge will be Python- or SQL-focused. (You should preferably mention this when booking or scheduling the mock interview so that the interviewer can prepare adequately in advance).

Once you start your call, you will be asked to join a Coderpad.io session.

Level 1 Python & SQL: 35 mins

The minimal requirement for passing this round requires you to code a working solution to both the Python **and** SQL challenges. In case you have not coded an optimal solution, you may be asked to work towards it (*if* and *only if* you have already completed coding a working solution).

Python challenges will involve using python standard libraries only as discussed earlier.

SQL will primarily involve whiteboarding since there won't be any existing database.

To view sample problems and topics to focus on, please refer to Appendix – A.

Level 2 Python OR SQL: 15 mins

This segment involves solving a more complex challenge based on your area of focus:

- **Data structures and algorithms** using Python if you have focused on the generalist or advanced ML specialization and are targeting data science jobs
- **Advanced SQL** if you have focused on the business insider specialization and are targeting analyst jobs

The expectations are very similar to Level 1, except this level is not required to pass the interview given the time constraints. However, you *should* attempt to solve the questions.

To view sample problems and topics to focus on, please refer to Appendix – A.

Feedback and Conclusion: 5 mins

Typically the last **5 - 10 minutes** will be spent by the interviewer giving you necessary feedback on your solutions.

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 are:

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

Coding 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
Level 1: Python Challenge	1	2	3	4
Level 1: SQL Challenge	1	2	3	4
Level 2: Python OR SQL Challenge	1	2	3	4

Total Score:

Fail (score<=6)		Pass (Score>6)	
Level 1: Not yet meeting expectations	Level 2: Approaching expectations	Level 3: Meeting expectations	Level 4: Exceeding expectations
3-4	5-6	7-10	11-12

Special Case: Any candidate who scores “Not yet meeting expectations” (score = 1) in ANY of the sections EXCEPT the Level 2 Challenge automatically fails the coding interview. This score would indicate that you are unable to code working solutions to basic python and/or SQL problems.

Appendix — A: Sample Coding Problems & Topics to Focus

This section covers places where you can practice problems and also topics you should focus on.

Problem Sources:

Both Python and SQL questions are available at:

- leetcode.com
- https://www.hackerrank.com
- careercup.com

For Level 1 Python, focus on the following aspects (using only base python):

- how to open a file in python, read in data, read or load json data, csv data etc
- how to work with strings, lists, dictionaries, list of dicts
- how to access keys and values of dicts, strings, list values etc.
- how to use loops \ comprehensions \ conditionals
- packages like collections and itertools might be useful
- functions like Counter, defaultdict etc.

For Level 1 SQL, focus on the following aspects:

- focus on select statement
- accessing columns of tables
- how to join tables
- filter data - where, having etc
- group by - sum, count etc

For Level 2 Python, focus on the following:

- essentials in data structures and algorithms
- basic searching \ sorting
- examples from leetcode or hackerrank
- finding duplicates or remove duplicates in a list
- finding max element or 2nd max element
- fibonacci series etc.

For Level 2 SQL, focus on the following:

- Advanced SQL constructs
- CASE WHEN
- Window functions
- CTEs
- Subqueries
- Multiple joins