

Get started

Open in app



Follow

572K Followers



Conquer the Python Coding Round in Data Science Interviews

My experience from interviewing as a Data Scientist in 2021 in Bangalore, India.

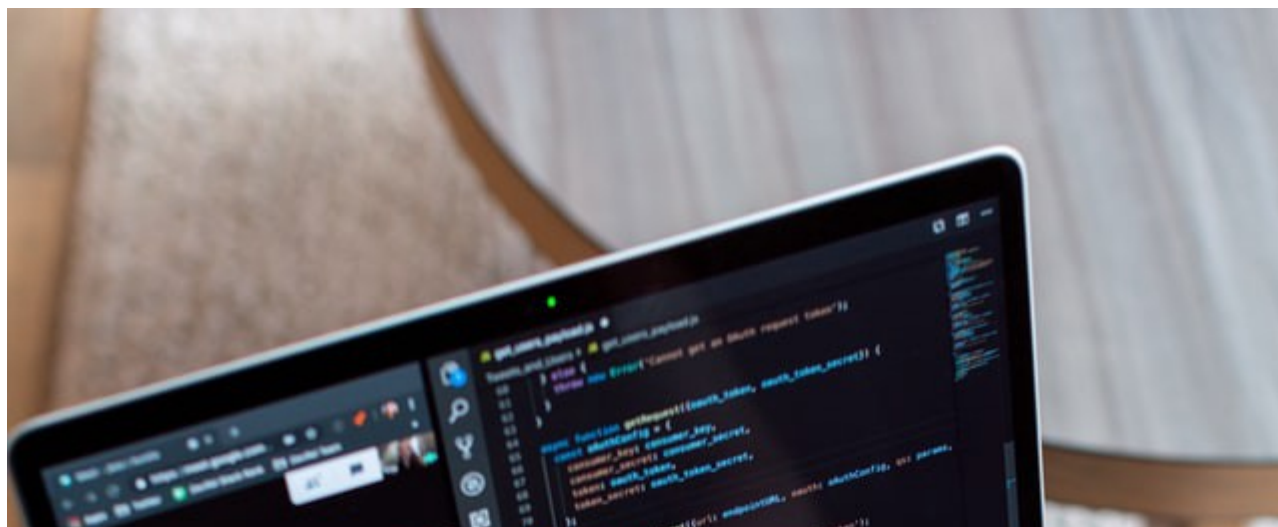


Divya Choudhary Jul 5 · 5 min read

The coding round has become an integral part of Data Science interviews. As ubiquitous it may be, it is also a dreaded round for many. With this post, I aim to fight fear with information, by sharing the different types of coding interviews and questions I encountered recently.

Let us look at the different Formats of execution and questions asked, and understand what concept is being tested by the questions.

Format1 — Live coding



[Get started](#)[Open in app](#)

Photo by [ThisisEngineering RAEng](#) on [Unsplash](#)

You are asked to open an editor (Jupyter notebook) and share your screen with the interviewer. It's appreciated when a candidate talks through their process to keep the interviewer on the same page. Often the interviewer nudges candidates in the expected direction. Most interviewers are considerate and allow minor Googling for syntax etc.

Some examples:

[Get started](#)[Open in app](#)

[https://www.kaggle.com/sudalairajkumar/covid19-in-india?](https://www.kaggle.com/sudalairajkumar/covid19-in-india?select=covid_19_india.csv)
[select=covid_19_india.csv](#)

Concepts Tested

1. *Exploratory Data Analysis*
2. *Data Cleaning*
3. *Handling missing values*
4. *Data Frame manipulations*
5. *Performing aggregate operations*
6. *Datetime manipulations*
7. *Data Visualization*

Q2. Create a dummy Dataset that has sensor values

Timestamp	Pressure (sensor 1)
12 – Dec 2019 00:11:03 am	10
21 – Dec 2019 11:10:03 pm	10
01 – Jan 2020 10:13:53 pm	13
02 – Jan 2020 08:11:07 am	21

Timestamp	Temperature (sensor 2)
14/12/2019 00:11:03:563	100
21/12/2019 21:10:03:098	130
02/01/2020 10:13:53:234	132

[Get started](#)[Open in app](#)

Sensor data, dummy dataset (image by author)

Find the highest ratio of Pressure to Temperature in this time range.

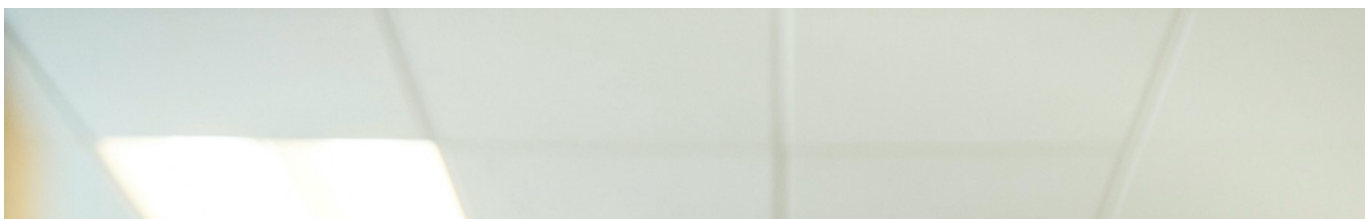
Concepts Tested –

1. *Exploratory Data Analysis*
2. *Data Cleaning*
3. *Data Frame manipulations*
4. *Performing aggregate operations*
5. *Datetime manipulations*
6. *Data Visualization*

Q3. Write a code that takes a number from the user and outputs all Fibonacci numbers less than the user input.

Concepts Tested –

1. *Basic Python Programming*
2. *Basic logical thinking*
3. *Awareness of Data Structure and Algorithms (This problem can be solved faster with Dynamic programming)*



[Get started](#)[Open in app](#)

Photo by [LinkedIn Sales Solutions](#) on [Unsplash](#)

[Get started](#)[Open in app](#)

Companies use this test as a way to check on various areas of Data Science. So, it's common to see a mix of objective type questions on Probability and statistics, Machine Learning, Deep Learning (like 10 questions of 2 points each) and a few coding questions on SQL, Machine Learning and Data Structures and Algorithms (rare but noticeable) (2–5 questions, their weightages based on complexity and concept)

Few examples:

Q1. Given the CDF of a distribution, find its mean.

Concept Tested — Basic Probability theory

Q2. Person A decides to go on a sky diving trip. Based on his research, the probability of a glitch resulting in death is 0.001. If A goes on 500 sky dives, what is the probability of death?

Solution Options:

- a) .50
- b) .29
- c) .39
- d) .01

Concept Tested — Knowledge of Distributions

Q3. Given a dataset of soft drinks sold in 4 stores in Europe, for a period of 1 year. Perform Data Wrangling and visualization. Can you predict future demand and identify how different features influence that? Please explain your findings effectively to technical and non-technical audiences using comments and visualizations, if appropriate.

[Get started](#)[Open in app](#)

Q4. Questions on writing SQL queries on sample datasets.

Concepts Tested — *Joins, Partition and Rank, Order by, Group by*

Q5. Given two numbers a, b ; $a < b$. Print output of

$$f(a,b) = g(a) + g(a+1) + g(a+2) + \dots + g(b-2) + g(b-1) + g(b)$$

where $g(x)$ is defined as all Fibonacci numbers less than x .

Concepts Tested —

1. *Python Programming*
2. *Logical thinking*
3. *Data Structure and Algorithms*

Q6. Given a number X, find the smallest sum of two factors (a, b) of X

Concepts Tested —

1. *Python Programming*
2. *Logical thinking*

Q7. Read the data present in this [link](#) using a get request. It contains the complete works of Shakespeare. Remove digits and stop words present in this [link](#). Now count the number of unique words present in the text.

Concepts Tested —

1. Using Get and Pull requests to read data
2. Basic NLP like stop word removal
3. Split a string into tokens

[Get started](#)[Open in app](#)

Photo by [BRUNO EMMANUELLE](#) on [Unsplash](#)

Format 3 — Onsite Case study/ Take home Assignment

This can be an Onsite (nowadays Online) short duration (1–3 hours) case study, or a take home assignment of 3–7 days where a candidate is given a sample dataset (which is fairly similar to a real dataset in size and complexity) and asked to solve for a business objective in 90 mins. Post that they need to walk the interview panel through their solution and thought process.

In the take home version, most companies understand that people have their day jobs as well and are considerate enough to extend the timeline within reason.

Pro Tip:

This format is a great place to showcase your breadth and depth of knowledge. Big points are given for going beyond the stated problem. For example, find literature (papers) similar

[Get started](#)[Open in app](#)

learning solution and compare and contrast their results. Maybe one model outperforms another in some segments of the data. Investigate those segments. Try building an ensemble model of the two.

This showcases to the panel that you go beyond the minimum requirement and you bring all your skills to the table.

Example: Retail transactions Case study

The company gave me a transactions dataset of a gift shop for a 2-year period.

Business problem: Predict whether a customer is going to buy a product next month.

They insisted I solve it using a deep learning technique.

This was the perfect place to display the range of my skills across the gamut of model building, and I fully used the tip mentioned above to convert the interview into an offer.

Conclusion

I hope with this post I have allayed the fear of coding round and levelled the playing field to some extent. It is to show the variety of ways companies test a candidate on their coding and problem-solving skills. Practice on the concepts mentioned above certainly goes a long way.

For Questions, you may leave a Comment or message me on [LinkedIn](#).

You may also like:

4 types of Data Science Interview Questions — Part1

Real questions with a strong ML focus

towardsdatascience.com

[Get started](#)[Open in app](#)

4 types of Data Science Interview Questions — Part2

Real questions with a strong business focus

towardsdatascience.com

Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. [Take a look.](#)

[Get this newsletter](#)[Coding Interviews](#)[Python](#)[Data Science](#)[Data Science Interview](#)[Python Interview Question](#)[About](#) [Write](#) [Help](#) [Legal](#)

Get the Medium app

