DATA SCIENCE EAST AFRICA

DATA SCIENCE CASE STUDY QUESTIONS.

Day 12 /20

Introduction :

The data science case study is often the most difficult part of the hiring process. After sending in a resume and passing the recruiter screening along with the initial interview, this final stage often makes or breaks an applicant’s hiring potential.

Designed to simulate a company’s current and past projects, case study problems rigorously examine how a candidate approaches prompts, communicates their findings, and works through roadblocks.

In order to understand how to pass the case study section, it's important to first understand what interviewers are looking for when applicants work through these prompts. Often, at this point in the process, prospects have already demonstrated sufficient technical understanding and skills for the position, so this is no longer a question of whether or not they can perform job duties.

Instead, case studies look to *understand the interviewee's thought process* – the ability to think on their feet through problems that don't have a singular solution. Real life cases aren't binary– there is no black-and-white-yes-or-no answer. Rather, due to all the ambiguities, candidates will need to *demonstrate decisiveness* in their investigations, as well as a capacity to consider impacts and topics from a variety of angles.

Perhaps even more importantly, the ability to *effectively communicate* conclusions will be *heavily* highlighted in data science case study problems. Real working conditions require a great deal of information exchange across teams and divisions, so part of the interviewer's focus will be on the system through which a candidate processes and explains their answer, and consequently, exactly what details are falling through the cracks.

Types of Data Science Case Studies

There are three main types of data science case studies :

1). Product questions.

2). Modeling and machine learning questions.

3). Business case questions.

Product Case Study Questions.

This type of case study tackles a specific product or feature, often tied to the interviewing company. As such, it is extremely beneficial to research current projects and research developments across different divisions, as it might end up as the case study topic!

In this type of data science case study, interviewers are generally looking for a sense of business intuition revolving around product mechanics. The most important part is to identify which metrics should be proposed to understand a product.

Example of a product case study question:

Suppose you're working as a data scientist at Facebook. How would you measure the success of private stories on Instagram, where only certain chosen friends can see the story?

Modeling and ML Case Questions

Modeling case studies are more varied and designed around developing some sort of insight into building models around business problems. These questions can range from applying machine learning to solve a specific case scenario to assessing the validity of a hypothetical existing model.

The modeling case study requires a candidate to evaluate and explain any certain part of the model building process.

A common case study problem would be for a candidate to explain how they would build a model for a product that exists at the company or another company.

For example:

Describe how you would build a model to predict Uber ETAs after a rider requests a ride.

Many times this can be scoped down into specific portion of the model building process. For example taking the example above, we could break it up to:

How would you evaluate the predictions of an Uber ETA model?

Or

What features would you use to predict the Uber ETA for ride requests?

Business Case Questions

Similar to product questions, business case problems are tackling a problem specific to the business. Common topics are often tied around having candidates assess the best option for certain business plans, and formulating a process for solving a specific problem. Other examples could include estimation and calculation, as well as applying problem solving to a larger case.

As with the product variant, it is helpful to read up on the interviewing company's products and ventures beforehand to have some exposure to possible topics.

Example business case question:

You work as a data scientist for a ride-sharing company. An executive asks how you would evaluate whether a 50% rider discount promotion is a good or bad idea. How would you implement it? What metrics would you track?

Framework for Data Science Case Studies.

There are four main steps to tackling every data science case study problem, regardless of the type:

- clarify.
- make assumptions.
- gather context.
- provide data points and analysis.

Clarify

The first step is to *gather more information.* More often than not, these case studies are designed to be confusing! There will be unorganized data intentionally supplemented with extraneous or omitted information, so it is the candidate's job in this step to even out this inherent disadvantage.

Make Assumptions

The next step is where the thought process really starts to be outlined. With all the data provided, it's important to *start investigating and discarding possible hypotheses*. Developing insights here is complementary to the ability to fine tune and glean information from the previous step, and the understanding gained there is paramount to forming a successful hypothesis.

The goal of this is to reduce scope of the problem at hand and ask the interviewer questions upfront that allow you to tackle the meat of the problem instead of focusing on random edge cases.

Hypothesize and Propose a Solution.

Now that a hypothesis is formed, gathering context is the next step towards fleshing out an answer. This is where the problem should be reframed given the new information gathered in the last two steps. Remember that there isn't an expected singular solution, and as such, there is a certain freedom here to determine the exact path for investigation. Consider how to define different metrics in the context of the problem.

Provide Data Points and Analysis.

Finally, providing data points and analysis involves choosing and prioritizing a main metric. As with all prior factors, this step must be tied back to the hypothesis and the main goal of the problem. From there, it's important to trace through and analyze different examples– from the main metric–in order to validate the hypothesis.

Regards,
Data Science East Africa.