page contents

KRISTEN KEHRER

HOME   ABOUT   CONTACT   DISCLOSURES & PRIVACY

How to Ace the In Person Data Science Interview

I've written previously about my recent data science job hunt, but this article is solely devoted to the in-person interview. That full-day, try to razzle-dazzle em', cross your fingers and hope you're well prepared for what gets thrown at you. After attending a ton of these interviews,

I've found that they tend to follow some pretty standard schedules.

You may meet with 3–7 different people, and throughout the span of meeting with these different people, you'll probably cover:

- Tell me about yourself
- Behavioral interview questions
- "White boarding" SQL
- "White boarding" code (technical interview)
- Talking about items on your resume
- Simple analysis interview questions
- Asking questions of your own

Tell me about yourself

I've mentioned this before when talking about phone screens. The way I approach this never changes. People just want to hear that you can speak to who you are and what you're doing. Mine was some variation of:

I am a Data Scientist with 8 years of experience using

statistical methods and analysis to solve business problems across various industries. I'm skilled in SQL, model building in R, and I'm currently learning Python.

Behavioral Questions

Almost every company I spoke with asked interview questions that should be answered in the STAR format. The most prevalent STAR questions I've seen in Data Science interviews are:

- Tell me about a time you explained technical results to a non-technical person
- Tell me about a time you improved a process
- Tell me about a time with a difficult stakeholder, and how was it resolved

The goal here is to concisely and clearly explain the *Situation, Task, Action and Result.*

My response to the "technical results" questions would go something like this:

Vistaprint is a company that sells marketing materials for small businesses online (always give context, the interviewer may not be familiar with the company). I had the opportunity to do a customer behavioral segmentation using k-means. This involved creating 54 variables, standardizing the data, plenty of analysis, etc. When it was time to share my results with stakeholders, I had really taken this information up a level and built out the story. Instead of talking about the methodology, I spoke to who the customer segments were and how their behaviors were different. I also stressed that this segmentation was actionable! We could identify these customers in our database, develop campaigns to target them, and I gave examples of specific campaigns we might try. This is an example of when I explained technical results to non-technical stakeholders. (always restate the question afterwards).

For me, these questions required some preparation time. I gave some real thought to my best examples from my experience, and practiced saying the answer. This time paid-off. I was asked these same questions over and over throughout my interviewing.

White Boarding

White Boarding SQL

This is when the interviewer has you stand at the whiteboard an answer some SQL questions. In most scenarios, they'll tape a couple pieces of paper up on the whiteboard.

One will be a table with (for example) ids and names (lets call this NamesTable), the other paper might include ids,

dates, and purchases (lets call this PurchasesTable). You get the idea, you're about to write SQL queries to answer their questions.

ID	NAME
1	Tim
2	Kristen
3	Jim
4	Paul
5	Elaine
6	Cliff
7	Tina
8	Susie
9	Harold
10	Stanley

ID	PRODUCT	DT
1	chair	2017-05-01 00:..
1	table	2018-02-05 00:..
2	chair	2017-05-07 00:..
3	candle	2018-01-05 00:..
3	book	2017-05-01 00:..
3	table	2018-02-12 00:..
4	candle	2018-01-16 00:..
5	book	2017-07-26 00:..
6	book	2017-05-03 00:..
7	computer	2018-01-15 00:..
7	cup	2017-05-22 00:..
7	cup	2017-09-11 00:..
8	computer	2018-01-14 00:..
9	cup	2017-10-22 00:..
10	cup	2017-09-10 00:..

They'll ask a series of questions such as:

Write a query to get all of the names -

select names from NamesTable

NAME
Harold
Tina
Paul
Elaine
Tim
Kristen
Jim
Cliff
Susie
Stanley

Write a query to get the names and purchases -

select names, purchases

from NamesTable as n

join PurchasesTable as p

on n.id = p.id

NAME	PRODUCT
Tim	chair
Tim	table
Kristen	chair
Jim	candle
Jim	book

Jim	table
Paul	candle
Elaine	book
Cliff	book
Tina	computer
Tina	cup
Tina	cup
Susie	computer
Harold	cup
Stanley	cup

Write a query to get the names and purchases for purchases made after some date (that they randomly pick, here is December 2017) -

select names, purchases

from NamesTable as n

join PurchasesTable as p

on n.id = p.id

where p.dates > '2017–12–31'

NAME	PRODUCT
Tim	table
Jim	candle
Jim	table
Paul	candle

| Tina | computer |
| Susie | computer |

Write a query to get the names and count of purchases for those having at least 2 purchases -
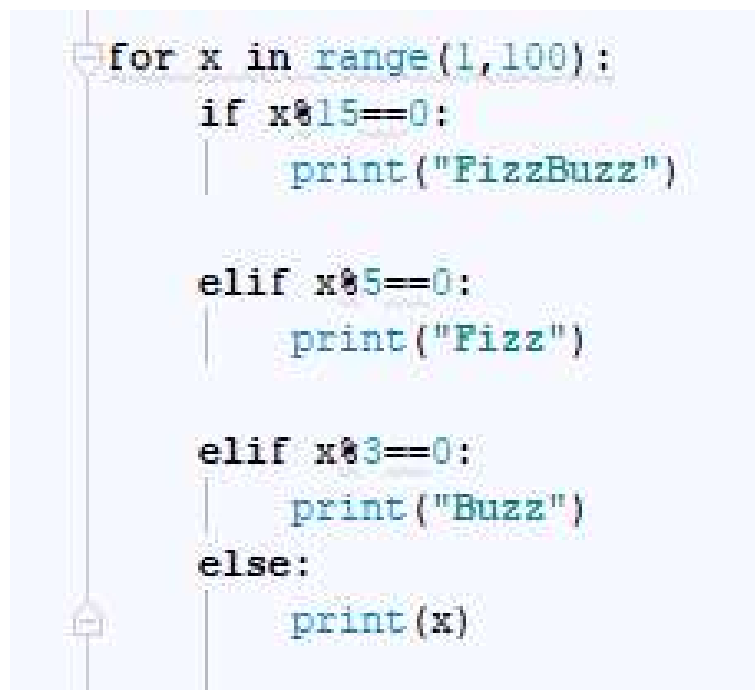
select names, count(purchases) as cnt

from NamesTable as n

join PurchasesTable as p

on n.id = p.id

group by names

having count(purchases) ≥ 2

NAME	COUNT
Tina	3
Tim	2
Jim	3

You get the idea. This shouldn't be too complex, an interviewer once asked me a query that required joining back to the same table, but it shouldn't get much more complex than this.

White Boarding Code

As mentioned in my previous article. I was asked FizzBuzz two days in a row by two different companies. A possible way to write the solution (just took a screenshot of my computer) is below:

```python
for x in range(1,100):
    if x%15==0:
        print("FizzBuzz")

    elif x%5==0:
        print("Fizz")

    elif x%3==0:
        print("Buzz")
    else:
        print(x)
```

The coding problem will most likely involve some loops, logic statements and may have you define a function. The hiring manager just wants to be sure that when you say

you can code, you at least have some basic programming

you can code, you at least have some basic programming knowledge.

Items on Your Resume

I've been asked about all the methods I mention on my resume at one point or another (regression, classification, time-series analysis, MVT testing, etc). I don't mention my thesis from my Master's Degree on my resume, but casually referenced it when asked if I had previously had experience with Bayesian methods.

The interviewer followed up with a question on the prior distributions used in my thesis.
I had finished my thesis 9 years ago, couldn't remember the priors and told him I'd need to follow up.

I did follow up and send him the answer to his question, they did offer me a job, but it's not a scenario you want to find yourself in. If you are going to reference something,

be able to speak to it. Even if it means refreshing your memory by looking at wikipedia ahead of the interview. Things on your resume and projects you mention should be a home run.

Simple Analysis Questions

Some basic questions will be asked to make sure that you have an understanding of how numbers work. The question may require you to draw a graph or use some algebra to get at an answer, and it'll show that you have some business context and can explain what is going on. Questions around changes in conversion, average sale price, why is revenue down in this scenario? What model would you choose in this scenario? Typically I'm asked two or three questions of this type.

I was asked a probability question at one interview. They asked what the expected value was of rolling a fair die. I was then asked if the die was weighted in a certain way, what would the expected value of that die be. I wasn't allowed to use a calculator.

Questions I asked:

Tell me about the behaviors of a person that you would consider a high-performing/high-potential employee.

Honestly, I used the question above to try and get at whether you needed to work 60 hours a week and work on the weekends to be someone who stood out. I pretty

frequently work on the weekends because I enjoy what I do, I wouldn't enjoy it if it was expected.

What software are you using?

Really, I like to get this question out of the way during the phone screen. I'm not personally interested in working for a SAS shop, so I'd want to know that upfront. My favorite response to this question is "you can use whatever open source tools you'd like as long as it's appropriate for the problem."

Is there anything else I can tell you about my skills and qualifications to let you know that I am a good fit for this job?

This is your opportunity to let them tell you if there is anything that you haven't covered yet, or that they might be concerned about. You don't want to leave an interview with them feeling like they didn't get EVERYTHING they needed to make a decision on whether or not to hire you.

When can I expect to hear from you?

I also ask about the reporting structure, and I certainly ask about what type of projects I'd be working on soon after starting (if that is not already clear).

Summary

I wish you so much success in your data science interviews. Hopefully you meet a lot of great people, and have a positive experience. After each interview, remember to send your thank you notes! If you do not receive an offer, or do not accept an offer from a given company, still go on LinkedIn and send them connection requests. You never know when timing might be better in the future and your paths might cross.

To read about my job hunt from the first application until I accepted an offer, click here.

If you liked this article, sign up for my email list! Here

© **2018 Kristen Kehrer**