I just went through a job interview at Facebook for a data science position. Wanted to write here a full review. Online there are TONS of resources about job interviews, but they are all vague at best. Sometimes they even look fake! I never got asked to fit balls in a bus!

1st step -> How to get the interview. Good luck applying online...I think the only way to get an interview at places like FB is via internal referral.  At least that was my experience.

2nd step -> HR interview. Was just an informal chat about why I wanted to work at Facebook. I watched several videos on Youtube of FB executives talking about FB mission. I picked from the videos a couple of things that resonated to me and told HR that I really loved FB had that mission and wanted to be part of that. I focused on making that personal. Am an Indian so I told her about the importance that connecting people online can have in my own country giving examples.

3rd step -> They sent me a data challenge. It was about advertising. How to evaluate them and find the best performing ones. It was kinda hard! There were so many ways to look at it. Some ads were great for growth cause they got lots of clicks, but not profitable. Some profitable, but few clicks. Eventually, I created a very simple combination of growth and profitability and ranked ads based on that. It helped me that before I had done quite a few challenges from the book the collection of data science takehome challenges. One is similar to the FB challenge. Also, practice made me a bit more confident and I didn't panic under time pressure. At least I was less scared!

4th step-> Shared screen. SQL coding. Before this, I was terrorized. I had never coded in front of anyone in my life and SQL is definitely not my #1 language. I felt I didn't do very well. Pretty much none of my queries would have worked. At some point, I panicked so much I switched to Python! I told him: let me solve this in Python and then I do it in SQL! But somehow I passed it! Below I am putting one of the two queries I was asked to solve:

*1.  Consider a table lifetime\_music\_actions that for any dt value has all the user-song pairs that have ever existed and the count of listens for each pair. Write a query to update that table for today given that you have it for yesterday and music\_actions for today.*

*(For a given dt partition, the table has ALL user-song pairs ever recorded up through that day.)*

*The schema will be something like:*

*"lifetime\_music\_actions"*

*dt STRING #date*

*user\_id BIGINT*

*song\_id BIGINT*

*cnt INT*

*music\_actions logs when a user listens to a song. Its schema is:*

*dt STRING # date*

*timestamp BIGINT*

*user\_id BIGINT*

*song\_id BIGINT*

5th step -> On-site interview. I had  5 meetings: 2 Product managers, 2 data scientists and 1 HR. HR was just chatting. Didn't feel like an interview. For the others, this is the most important question each asked me. Some were pretty vague. I think it was my duty to better define them or ask them questions to narrow down the problem.

1. 1st PM: How would you estimate the life time value of a click on ad? Focus on how you would communicate this in simple terms to an advertiser.
2. 2nd PM: We have developed a new ad model. How would you test if this is better than the current model? Not just regarding revenue, but also thinking about user experience (I loved this question cause it showed how FB always thinks about user experience first!)
3. 1st DS: Build a model to classify if a click on FB comes from a real user or a bot (first time I heard about bots in my life!).
4. 2nd DS: Pick your favorite model and walk me through it. How it works, how it's built, why is your favorite. I chose a random forest, but I did bad here. I couldn't explain well how each tree was built. The splitting criteria.