

Screening Test/Interview Questions with Answers for Data Analyst (Remote) Position

Candidate's Name: Jayasurya Rathinagiri

1. Are you currently employed?

No, I recently graduated and am currently on an F1 student visa, awaiting my OPT.

2. Why did you go into data analysis?

The reason I went into Data Analysis is that, In the initial stages of my career, I was focused on development. However, working with large volumes of data and realizing the potential of what can be achieved with that data completely changed my perspective. Initially, data might seem like mere values, but the endless options for visualizations, identifying trends, and recognizing patterns reveal its incomparable potential. From that point on, I was fully committed to data analysis.

3. Which data analysis software are you well-versed in?

I am proficient in MySQL, Excel, Tableau, Power BI (Business Intelligence), and Python (with libraries like Pandas, NumPy, SciPy, and scikit-learn).

4. Do you prefer to work independently or as part of a team?

I am comfortable working both independently and as part of a team. I have experience collaborating with teams to deliver successful projects.

5. When dealing with multiple projects, how do you prioritize your tasks?

The first thing I do is establish clear milestones for each project and begin working early. I start with the project that has the nearest deadline and then move on to the one that requires the most time and effort.

6. What is your process for cleaning data?

- The first step is to import the dataset and then create a working copy of the dataset to avoid making accidental changes to the original data.
- Next, I remove duplicates. For example, in MySQL, I can use the **ROW_NUMBER** function to identify duplicates based on all columns. In Excel, I can use conditional formatting to flag duplicates or use the remove duplicates feature directly.
- The next step is standardizing the data, which includes trimming whitespace and standardizing the variation of the values.
- After that, I handle null and blank values by identifying them and either eliminating them or populating them using existing values.
- Next, I modify the data types as needed. For instance, sometimes a date might be stored as text, so I use the **STR_TO_DATE** function to convert it to the **date** type.
- Finally, I remove any unnecessary rows and columns that are not needed for analysis.

7. How do you explain technical concepts to a non-technical audience?

I use simple language and analogies to explain technical concepts. I also use visual aids like charts and graphs to make the information more accessible.

8. What's the largest data set you've worked with?

I worked with a data set of over 23 million entries for an Airbnb market analysis project.

9. What's your favorite project that you've worked on?

My favorite project was developing a Tableau dashboard for Airbnb market analysis, where I provided key insights for investors.

10. How did you make sure you hit the deadline?

I ensured timely completion by setting clear milestones, regularly tracking progress, and adjusting plans as needed to stay on track.

11. Tell me about your experience working remotely.

I have experience working remotely during my internship at JP Morgan Chase & Co. It was a virtual internship where I effectively collaborated with my team using various online tools, including Slack and Forage.

12. What is your process when you start a new project?

My process involves deep understanding of the project requirements, gathering, and analyzing relevant data, and developing a clear plan with milestones and deadlines.

13. What was your most difficult data analyst project?

The most difficult data analyst project I worked on was the Tableau Dashboard Development for Airbnb Market Analysis. For this project, I used three different datasets: one containing all the details of typical Airbnb listings, another with calendar details such as listing dates and listing IDs, and the last one with reviews containing review IDs and review texts.

I joined these three datasets, resulting in over 25 million rows of data with hundreds of columns. After cleaning the data, I reduced it to 23 million rows. However, I faced a challenge because Tableau Public allows only 12 million rows at a time. To overcome this, I optimized the joins and condensed the data to bring it below the threshold.

14. Your employer wants you to find all sales quota data from the past three months to see if recent incentives enhanced employee productivity. How does data cleaning factor into your analysis?

Data cleaning is essential for accurate analysis of sales quota data. It involves removing errors, standardizing formats, handling missing values, and detecting outliers. Clean data ensures reliable comparisons of pre- and post-incentive performance and ensure there is no false information, leading to valid conclusions about the impact of incentives on employee productivity.

15. What makes you stand out from other data analyst candidates for this position?

My strong technical skills, combined with my ability to effectively communicate with both technical and non-technical audiences, make me a standout candidate. Additionally, my experience in handling large datasets and delivering actionable insights sets me apart.