Analytics Engineer – Technical Assessment

This assessment focuses on testing your level of knowledge and skills to assess your Analytics Engineering capabilities. You will get some raw data that needs to be transformed into a usable data warehouse by using the Kimball dimensional modeling technique.

The case

You are working for an online retail company that sells products on Amazon. Our marketeers will have many questions about the performance of our sales on Amazon to better understand the customer and take the appropriate actions. To better facilitate our marketeers, you are asked to build a simple data warehouse using Kimball dimensional modeling that will make it easy for marketeers to perform analysis on.

Example questions that our marketeers might have:

- Which product styles and categories are the most popular in the city of Mumbai?
- What is the sales trend for different seasons?

The assignment

- The data is given as a .csv (Amazon Sale Report.csv) and is included in the invitation. You are free to use any database (e.g. Postgres, DuckDB or any Cloud database for example) as a source for DBT.
- Design a data warehouse using the Kimball star schema model, aimed at facilitating marketers' analytical tasks. Implement it in SQL and utilise DBT.
- Make simple visualizations which demonstrates you understand how data will be used in BI tools and how to structure data for use in BI tools. You can use any tool to visualize. Don’t spend too much time on this.
- The expected preparation time is approximately 4-6 hours. Should you not be able to finish in this time, we value also the thought process of how you thought of the architecture, approach and any future improvements you could make.
- Store all the code and other relevant output in a Github repository.

The presentation

- The interview will take 1 hour in which you will present your solution and we can ask questions.
- You must present your findings and way of working using a presentation. Try to limit the presentation to 40 minutes so there is room for questions. Topics to include:
    - How did you explore the data?
    - How did you clean the data?
    - How did you create the facts and dimensions and what was the reasoning behind it?
    - How did you set up DBT?
    - How did you prepare your data for use in BI?
    - How do you guarantee data quality?

- Please share your Git repository containing the solution with the people participating in the assessment 24 hours in advance.