Analytics Engineer Assessment

This assessment focus on providing a mini case that we expect from the Analytics Engineer role, you need to showcase your skills on data exploration, cleaning, validating, and in the end transforming it into a usable data warehouse by using Kimball methodology.

This exercise is intended to test your hands-on skills on building a easy to use data warehouse using Kimball dimensional modelling. Think about this scenario: our marketeers want to study the relationship between different sales facts, e.g. which product style and category are more popular in city of Mumbai? What is the sales trend over seasons? What if I want to have some insights on the last period over sales and want to build a dashboard to report on this continuously? You can help them by building a Kimball style data warehouse. Please try to make fact and dimension tables tidy, clean, and easy to understand so that our marketeers can do analysis effortlessly.

The solution has to perform the following actions:

- Cleaning, validating, and exploring the raw dataset.
- Present your findings using a visualization tool but don't spend too much time on this (feel free to choose any open source tool).
- Create data warehouse structured in Kimball star schema that enables marketeers to do analysis (preferably written in SQL/DBT, you can also write in PySpark/Pandas).
- Adding some testing on your data models to ensure consistent data quality.
- Lead us through your data models building process and explain your choices of each step.

Notes and requirements:

- Optional: using DBT to construct this project
- Optional: deploy the data models on a public cloud provider
- You are free to choose any programming language, technology or service
- During the interview you will be requested to give a presentation of your solution to explain your choices in 1 hour.
- 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 how you thought of the architecture, approach and any future improvements you could make.
- Please share your results 24 hours before the interview through a public repository.