Hello,

I, Pratik Talreja have been working on unstructured data to diagram a new structured relational data model. Following that, I worked on queries to provide analytics on the performance of different brands based on different aspects.

While working on the exercise using SQL queries, I had a few questions and came across issues about data-

Questions-

1. What's the source of the data. How frequently the data was fetched from these sources?

2. I observed a lot of missing data. So it would be interesting to see that is it because the data was never recorded or there's some issue when the unstructured data is being migrated into the data source?

Issues-

1.
   - In the tables, I found multiple **duplicate** entries for each user. This can make the results of query inconsistent. Hence it would be better to consider just **one instance per user**.

2.
   - In brands table, in my opinion, Brand Code should also be **unique** for each brand.

3.
   - There are many **missing** values in different tables which makes the results of certain queries inconsistent.

Additional Details Required-

1. Data Source- More information about where the data is fetched and how frequently.

Performance issue I expect to face-

The python script which converts unstructured data to structured data is not yet production-ready. To make it production-ready, I intend to follow below steps-

- Creating specific python virtual environments for each separate deployment
- Installing static code analysis libraries in the virtual environment as dev dependencies
- Turn any hardcoded values like database credentials into parameters.
- Building a Docker fIle to containerize the code.

Please let me know if you have suggestions to make the process more efficient. Your time and effort is much appreciated.

Best Regards,
Pratik Sunil Talreja.