Team 8: Michelle Xie, Elbek Gaybullaev​, Theophilus Ghartey​, Jingsi Hu​

Dr. Fogel

BAN5573

30 April 2023

**Interview Case**

**Data Quality Issue (at least 1)**

Assuming we test the data quality issues of Users Data Schema were included with columns: "\_id," "state," "created date," "lastLogin", "role," and "active," one possible data quality issue could be missing or null values in the "lastLogin" column.

To identify this issue, we can use the following SQL query:

SELECT COUNT(\*) FROM users WHERE lastLogin IS NULL;

This query will count the rows in the "users" table where the "last login" column is null. If the result is more significant than zero, then we have missing values in the "lastLogin" column, which could indicate a data quality issue. Another possible issue could be inconsistent or invalid values in the "state" column. To identify this issue, we can use the following SQL query:

SELECT DISTINCT state FROM users;

This query will return a list of all unique values in the "state" column. If there are any unexpected values or inconsistencies in the list, we may have a data quality issue with the "state" column. For example, if we see deals like "NY", "New York", and "New York" in the list, then we may have inconsistent formatting or spelling of state names.

select \* from public.users u order by u.userid;

delete from public.users a using public.users b where a.userid = b.userid and a.lastlogin < b.lastlogin;

🡪 keeps the row with the most recent lastlogin value

delete from public.users a using public.users b where a.userid = b.userid and a.id > b.id;

🡪 keeps the row with the older ID

🡪 then you can make userid a primary key after cleanup

The two queries above were used. From the first query, we noticed that the user id was not unique. There were number rows with the same user id and information. To resolve this situation, the second query was used to delete the duplicates. Doing so, allows the user id to be unique and keeps the one with the most recent last login (assuming that it has the most up to date data).