**Subject**: Data Quality and Management

Hello Sir/Mam,

This is Govind Maheshwari from Data Analytics team, and I was asked to analyze the file for data quality issues. I have gone through the file and came up with initial findings and they are as follows:

* Quite a lot of null values are there in columns of Brands table and Receipt Table
* User Id in Users table should be unique and not duplicated (wastage of memory)
* Misleading Table name i.e., user table doesn't have any detail about user
* Points are earned without buying any product
* The users in receipt table are not there in users table. (Data extract is not consistent)
* Brand table and receipt table has same column name BrandCode but both appear to be different and linked to each other
* Naming Convention of columns can be improved

In Addition to it I had some question about the data to resolve the data quality issues

* Do we have any metadata of tables and column to efficiently resolve some of the problems which I mentioned above.
* Are these 3 tables the only tables that we use most of the time or there are more tables.
* What is the purpose of the data, is it to store for the data warehouse or to find insights and improve the business and if latter is the case then what are the KPI’s involved.
* How much data do we have at present and what is the average increasing rate of the data.
* Need to understand the functioning of the App to understand how data is getting generated and can be improved at the source level i.e., my enforcing more checks and constraints.

Moreover, as company is gaining attraction from both companies and users, more will follow and the data will be increasing exponentially so we need to design the data warehouse such that it will not cause problem in near future regarding storage issues, performance and scalability etc. We can take different storage in cloud for more recent data and old data respectively so the performance will always be at par. The queries written to retrieve those data should be efficient. The data needs to be indexed for faster retrieval.

We can discuss all the points in more detail in a meeting and can collectively find out further problem and their solutions

Regards

Govind Maheshwari