**Subject**: Questions regarding data quality issues.

Hi Product/Business leader,

I am Jinal Sapariya, I am working as a Data Analyst at Fetch Rewards, Data & Analytics team. I am reaching out to ask you about a few data related questions that came up during the data exploration and modeling process. I would be grateful if you could provide your business perspective to these questions. They are as follows:

* What are the business use cases as it will help us to identify critical attributes from the data?
* What are the target KPIs we have to monitor with respect to this data?
* Do we have any other additional resources which we can use along with the provided data?
* Who is the targeted audience for this analysis report?

Moreover, during the initial analysis process, I found few data quality issues and I want to share some resolutions through which we can mitigate the impact.

|  |  |
| --- | --- |
| ISSUES | RESOLUTION |
| Many features have null values/ missing data. | We can use either Data Imputation techniques like fill null values with 0 or avg values based on binning some categories or if null counts are less and they do not impact the analysis, we can drop the values. |
| In Users table, user\_id had duplicate values | Due to duplicates we cannot consider this field as the primary key. We need to keep Unique key Constraint during early stage of data modeling |
| According to the Users Data Schema, the role should have a constant value set to "CONSUMER" but in data, I discovered that role had different values like "Fetch-staff". | Reach out to SME to clarify about the attribute. |
| DATE FORMAT not consistent | Define the Date format standards from the beginning of data modeling. Also, we can write a script which will convert into a single date format. |
| BonusPoints awarded but no receipt item list present in the receipt table | Easier way is to define a boolean field which help us to differentiate scenarios of having points with/without item list present |
| There is no clear common column to join the rewards item list in a receipt and the brands table. | The column that we have found useful is a brandCode but we can specify a particular reference column from the brands table into the rewardsReceiptItemList. |

In the early stage of data modeling, database security is a major factor. Thus, I want to know what are the security and privacy compliances? Through this we can enforce encryption techniques for secured storage of data and will help in optimizing our data assets.

To handle large volumes of data in production, we are planning to use indexing, caching techniques for faster retrieval of response data. This can also reduce the cost of complex operations. Maintaining consistent, clean, and normalized data, we can achieve high performance and scalability.

Please do let me know if you have any questions regarding any of the details described above. It will be my pleasure to even schedule a meeting to discuss further.

Thanks & Regards,

Jinal Sapariya