Dear Product Manager,

I have gone through the Fetch Rewards new batch of data .Could you please clarify my questions on the data as mentioned below

* Are the userID, receipt ID, brandID encrypted? Because I can see the values of ID’s as 5ff1e194b6a9d73a3a9f1052
* Why are values in date not displayed in the correct date format, is there any way I can get the dates in correct format. Without proper dates, it is difficult to analyze data related to dates

I performed some Exploratory Data Analysis on the data to study, observe the data and following are the data quality issues I found

* There are many null values in the dataset which contributes around 33% to 52% of total data. For dataset of small size, it is alarming to have null values in majority and these issues has to be addressed, decisions has to be taken to deal with the missing values 2.The datatype are not correct(i. For dates the datatypes is int but it has to be Date ii. Boolean columns has datatypes as int)
* Values in Date columns are incorrect, the data should be in format YYYY-MM-DD or MM-DD-YYYY
* Some columns have "" as values which are to be considered as null values

I would need the following to resolve data quality issues

* How the data is stored in the system and extracted
* Why are some values empty
* How should I deal with null values(Should I remove the fields with null values or replace them using methods available like filling the values with mean,ffill,bfill and interpolate

I need the answers for below mentioned to optimize the data assets

* The fields which are not required for our analysis
* Exact datatypes for each field

As our users grow in number we are likely to face performance issues when we work on the same database for our analysis. I suggest we move our data into Data warehouse using ETL tools so that we can perform OLAP without interfering with OLTP . This method reduces load on our database and makes data available for our analysis, we can also restrict access to data