Hi ,

I am Rishabh. I am an Analytics Engineer at Fetch and wanted to discuss some data quality issues I encountered during my work. I was working with three files containing information about users,brands and receipts of a retail or e-commerce platform.

These files had data issues namely missing data, duplicated information, nested data structure, and data inconsistencies. These issues if not addressed can hinder our decision-making leading to sub-optimal results and operational issues down the line. I would like to request some information from you to combat the above issues.

Firstly, I would like to inquire information about the data checks and validation steps currently implemented if any in the data pipeline. This information can help identify and correct errors and can speed up and improve the data analysis later on.

Secondly, I would like to request information about the business objectives as it can help formulate the analysis better and give a general direction for the analysis to go forward in. This information will also guide me in easily creating data assets that can assist me in my analysis.

Finally, I would like to request metadata about each of the files. This will aid me in resolving the above data issues appropriately. This can help me allocate importance to files accordingly. I have also developed a strategy to combat them and would like to discuss the same with you to gain a better perspective. I conducted my work using Python and SQL and can share my coding notebook if required.

I believe fixing these issues would complement the organization's performance and our decision-making process. I would highly appreciate the opportunity to discuss this further at your convenience.

Looking forward to your response.

Best,
Rishabh