Hey Mark/Katie,

I've had a chance to dig through the data sets you sent me on Wednesday and wanted to share what I've found so far. Here's a summary of the data quality issues I encountered, some interesting trends I noticed, and a couple of questions I still have.

- **Missing Data**: When I compare the 3 tables to one another I notice there are a lot of records missing from both the Users and the Product Tables. This is probably my biggest concern at the moment. For products, a little less than half of the barcodes in the transaction log have corresponding products. For users, the story is actually a lot worse. Currently we show less than 1% of users in the user log. I'm thinking/hoping that I have just a subset of data, and would like to request access to the full data set. Can you help with that?
- **Null Values in Transaction Data**: The transactions log has numerous records where either the final_quantity or final_sale columns are null. While I've been able to merge some of these records by matching barcode, receipt, and user/store/date, there are still many that need cleaning. Based on what I'm seeing, I think that the problem is potentially that the records are being split for some reason, but I need to talk to someone in Engineering to make sure they aren't pulling in any bad data from source and to make sure that the issue isn't just related to the exports I've received.
- **Other issues**: I'm seeing some other smaller problems within the data like duplicate records within the products log and some odd birthdays in the users log. These aren't quite as big of deal, but still need to understand them more and hopefully get this data fixed.
- **Some good news**: So, it's not all bad news. Based on some initial findings with the current user table it seems like we've had some good user growth in the 12 months leading up to September. We have over 18% growth during that time period, and I'm thinking that it might look even better once we have the full user logs to analyze.

In terms of next steps, I'd like to propose we schedule a meeting with someone from the DBA team and the Engineering team as soon as possible. It would be helpful to confirm record counts on each of the datasets, and I'd like to better understand if there's any additional data being logged that could support our analysis. I'll go ahead and put something on the calendar in the next couple of days and will send out invites shortly.

In the meantime, could you let me know if there are any Confluence pages, documentation, or internal resources related to the data pipelines or point me to the right contact person? This will help me get a better picture of the data flow and help expedite resolving some of the issues we've encountered.

Please let me know if you have any questions or if there's anything specific you'd like me to focus on during these meetings.

Thanks for your help!

Dustin