



Joseph Dunst &lt;joseph.dunst@gmail.com&gt;

---

## Data Quality Questions

---

Joseph Dunst <joseph.dunst@gmail.com>  
Draft

12 April 2021 at 18:28

Good afternoon Susan!

My name is Joey and I'm on the analytics/engineering side of things. I wanted to reach out about a few questions that the engineering team needed greater clarity on with ensuring that our team is in line with what the business is expecting and we have long term, big picture goals in mind with our current initiatives.

If you have time to chat on one one, let me know and we can set up a call!

The primary question I have is about our users data. We have a lot of stored receipts that reference users that are not in our system. This discrepancy was found by searching a recently pulled set of receipts against a recently pulled set of users in our system and seeing where the users in receipts did not exist in our users database.

What this means is, unless we have a purpose for keeping info about people not using our app, we are unnecessarily using up space and reducing performance since we have to process information that doesn't relate to a current user before finding data that does relate to a current user.

So, should we store data related to deactivated users? Or should we retain it, and create an identifier that would let us reference and apply it to a user who may come back in the future? Once I have a decision, I can relate it back to my Product Owner to get the ball rolling on addressing it ASAP.

One other thing - over time, our dataset is likely going to grow very large, which will inevitably lead to slower performance. We may want to consider storing data past a certain date in a separate place and creating a different means of accessing historical vs. current user data (think last year vs 2 years or older). Do you have any customer use data you could provide for us to assess and analyze to help drive a decision about the above?

Best,  
Joey