

Third: communicate with stakeholders

Construct an email or slack message that is understandable to a product or business leader who is not familiar with your day-to-day work. Summarize the results of your investigation. Include:

- Key data quality issues and outstanding questions about the data
- One interesting trend in the data
 - Use a finding from part 2 or come up with a new insight
- Request for action: explain what additional help, info, etc. you need to make sense of the data and resolve any outstanding issues

Response:

Email Template –

Hi XXX,
Hope you are doing well!

As an Analyst, it is my responsibility to ensure that the team makes informed business decisions using our historical trends. I was analyzing our product catalog, users, and transaction data. During my analysis, I came across a couple of questions, can you help me resolve these?

Below is the list,

1. I noticed that the Product Catalogue dataset has multiple categories. Can you share how the segmentation works here? What is the difference between Category 3 and Category 4? Category 4 is also unavailable for many products, how important is this field?
2. We are missing barcodes for ~4K products, how can we retrieve them?
3. We have multiple data entry errors on the birthday field in the users table, 291 rows exceed the average human life expectancy age, and we need to check further with the average Fetch users age as well.
4. The transaction table seems to only have 24K receipt ID, while the table count is 50K. A general assumption would be every row is unique and records one transaction per receipt_id, can you clarify how the transaction record is perceived by the team?

FYI, I also discovered some interesting facts:

1. Walmart is the most popular store, it comprises 41% of the total transactions.
2. More than 65% of the users are Female.
3. About ~41% of the transactions included a product from the 'Health & Wellness' category.

I've attached some data samples from each of these tables below.

Let's discuss the inconsistencies in detail and I'd like to learn more about the dataset from a business/product perspective. Let me know how you'd like to take this further both email and meeting works for me (meeting is preferred)...

Best,
Ahmed

Slack Template –

Hi XXX,

As an Analyst, I work on ensuring the team makes informed business decisions using our historical trends. I was analyzing our product catalog, users, and transaction data. During my analysis, I came across a couple of questions, can you help me resolve these?

Below is the list,

1. I noticed that the Product Catalogue dataset has multiple categories. Can you share how the segmentation works here? What is the difference between Category 3 and Category 4? Category 4 is also unavailable for many products, how important is this field?
2. We are missing barcodes for ~4K products, how can we retrieve them?
3. We have multiple data entry errors on the birthday field in the users table, 291 rows exceed the average human life expectancy age, and we need to check further with the average Fetch users age as well.
4. The transaction table seems to only have 24K receipt ID, while the table count is 50K. A general assumption would be every row is unique and records one transaction per receipt_id, can you clarify how the transaction record is perceived by the team?

I also discovered some interesting facts:

1. Walmart is the most popular store, it comprises 41% of the total transactions.
2. More than 65% of the users are Female.
3. About ~41% of the transactions included a product from the 'Health & Wellness' category.

Happy to schedule some time this week or the next to deep dive into this. Let me know a time that works for you...

Thanks!