

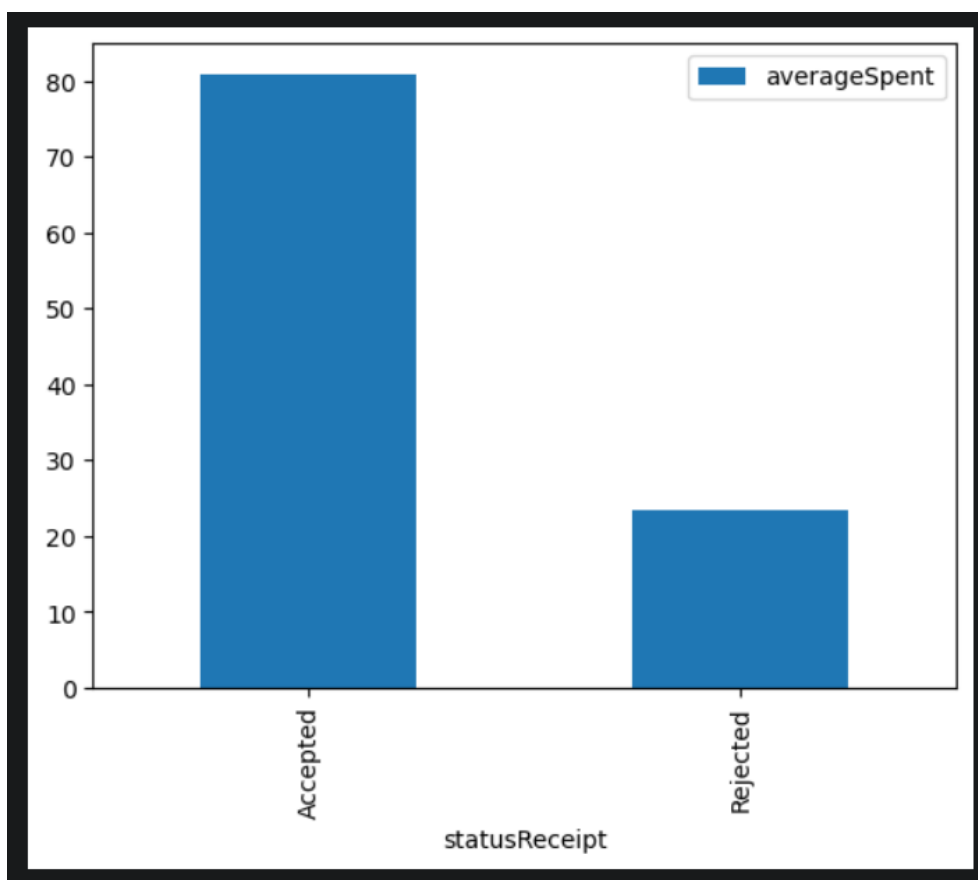
Construct an email or slack message that is understandable to a product or business leader who isn't familiar with your day-to-day work. This part of the exercise should show off how you communicate and reason about data with others. Commit your answers to the git repository along with the rest of your exercise.

- What questions do you have about the data?
- How did you discover the data quality issues?
- What do you need to know to resolve the data quality issues?
- What other information would you need to help you optimize the data assets you're trying to create?
- What performance and scaling concerns do you anticipate in production and how do you plan to address them?

Slack message to a business stakeholder

Hi Rob!

When a user scans a receipt it, we assign a status to the receipt. The status could be submitted, finished, flagged, rejected, pending, or rejected. Once the status is finished we can consider the status as the user accepted his rewards. If it's pending it might take some time to get accepted or rejected. Currently, in the data I have received the following can be observed. As per data average spent total spent on a scanned receipt which is accepted is around 80 dollars and Rejected is around 23 dollars.



Slack message to a Technical Lead

More information on barcode and brand code is helpful in tackling the issue of joining the receipts with the brand table would be great.

I created for loop to catch the exception on the if either barcode or brandcode is empty. I also created a separate dataframe with receiptlist items to identify how many of them are empty.

The main scaling concern is regarding the rewardsreceiptItemListcolumn this column might be heavy. Moving this to a separate table can be helpful. My ER diagram shows how this can be addressed.