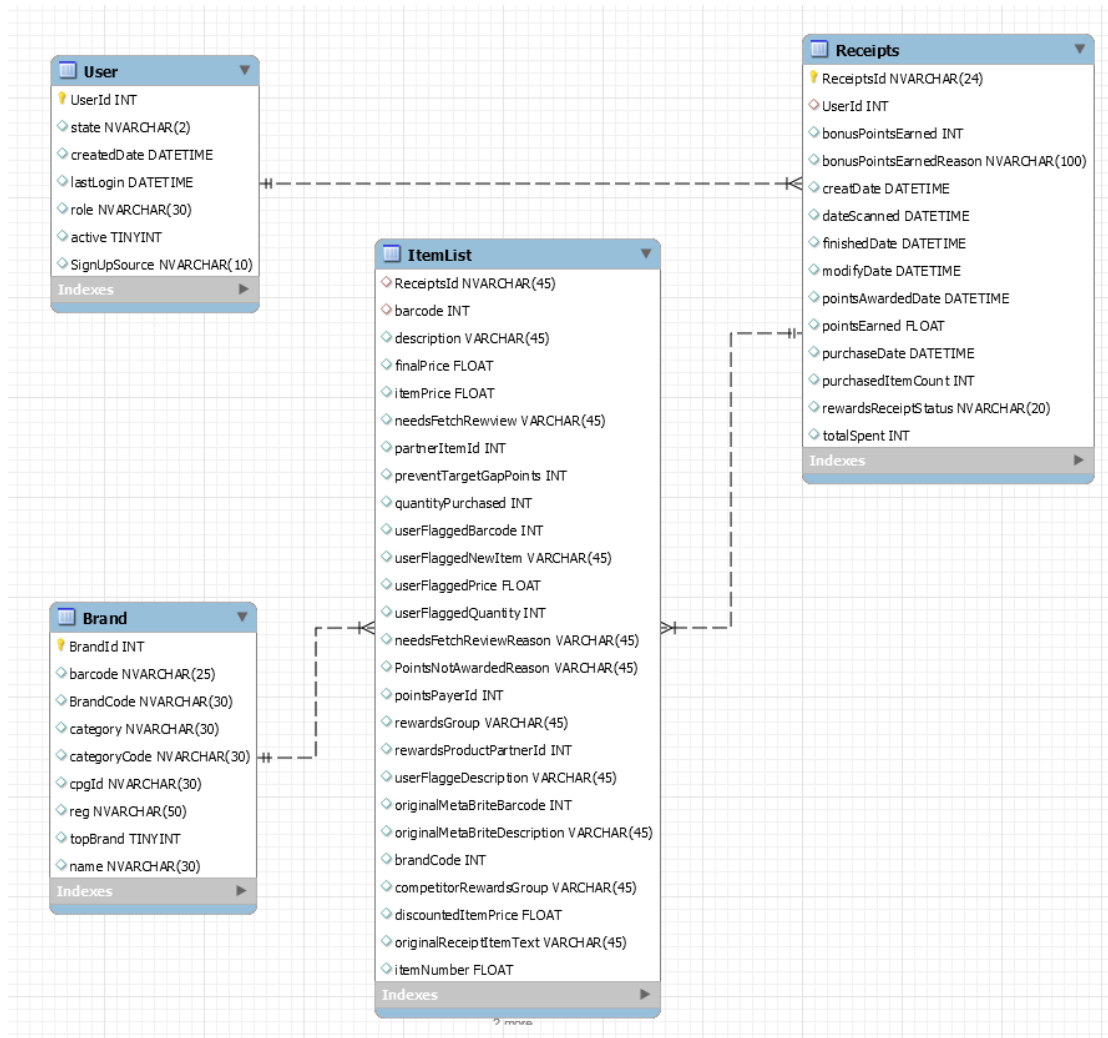# Fetch Rewards Coding Exercise - Data Analyst

Jianxiang Zhai

1. ER diagram

2. SQL scenarios
   Please refer to the SQL files.

3. Evaluate Data quality
   The overall missing data value is high, most of duplicated data concentrates on Users table.
   My data exploration mainly focused on two parts:
     1). Relationship of barcode
     2). Relationship of userId
   The boxplot is to check outliers for points earned.
   Please refer to the py file for the detailed codes.

4. Communicate with stakeholders

Dear Sir/Madam,

Hope you are doing well!

I'm writing this regarding of the recent data analysis task, after converting json data into csv file and finishing the normalization. I build an ER diagram and have some further analyze about the entire data that be mentioned during the meeting. The whole data issue I found are as follows:

- The entire data quality of these three tables are relatively low, there are missing values present among all the tables, including some duplicated rows regarding some important information such as User Id.
- After analyzing the barcode between brands and ItemList, I found that the barcode is totally ununiformed, and only 15.64% of barcode in brands table can be found matched with the barcode in ItemList table.
- Do you have any independent data set regarding item information? The low-quality data of ItemList may because the entire information is curved into one column of receipts table.
- What KPI do you concern the most? Can I have more details about this KPI?

We can have a further discuss regarding to these issues if you have any questions.

Best regards,
Jianxiang Zhai