From my analysis of the data given, Receipts, Users and Brands, I found that there are some issues regarding the data. Firstly, I noticed that there is not a clear way to connect Brands with either of the other two while it is simple to connect Receipts and Users after looking more at the data, I think we could connect Brands to Receipts from rewardsReceiptItemList and brandCode, but there are quite a few rows where rewardsReceiptItemList don’t have brandCode.

Another issue is the number of Null values, duplicated data, and values which are unclear. For unclear data, it is mainly the testbrand in Brands; I am unsure if it is an error or if it is just used as a place holder. For Null values there are quite a few of them. In Users, there are around 52 Null values from its 495 rows, 650 Null values in Brands from its 1167 rows, and 585 Null values in Receipts from its 1119 rows. For Users there aren’t a lot of Null values, but out of those 495 rows, 283 of them are duplicate data. For Brands and Receipts around 50% of the data contains Null values and majority of these are in categories or dates, so it would be hard to impute a value for them.

A way to clean up those Null Values would to simply drop them, but then we would be losing a lot of the data, so I am unsure if I should drop those rows. For duplicated data I think it is best to just simply drop them because they are not going to be too helpful.