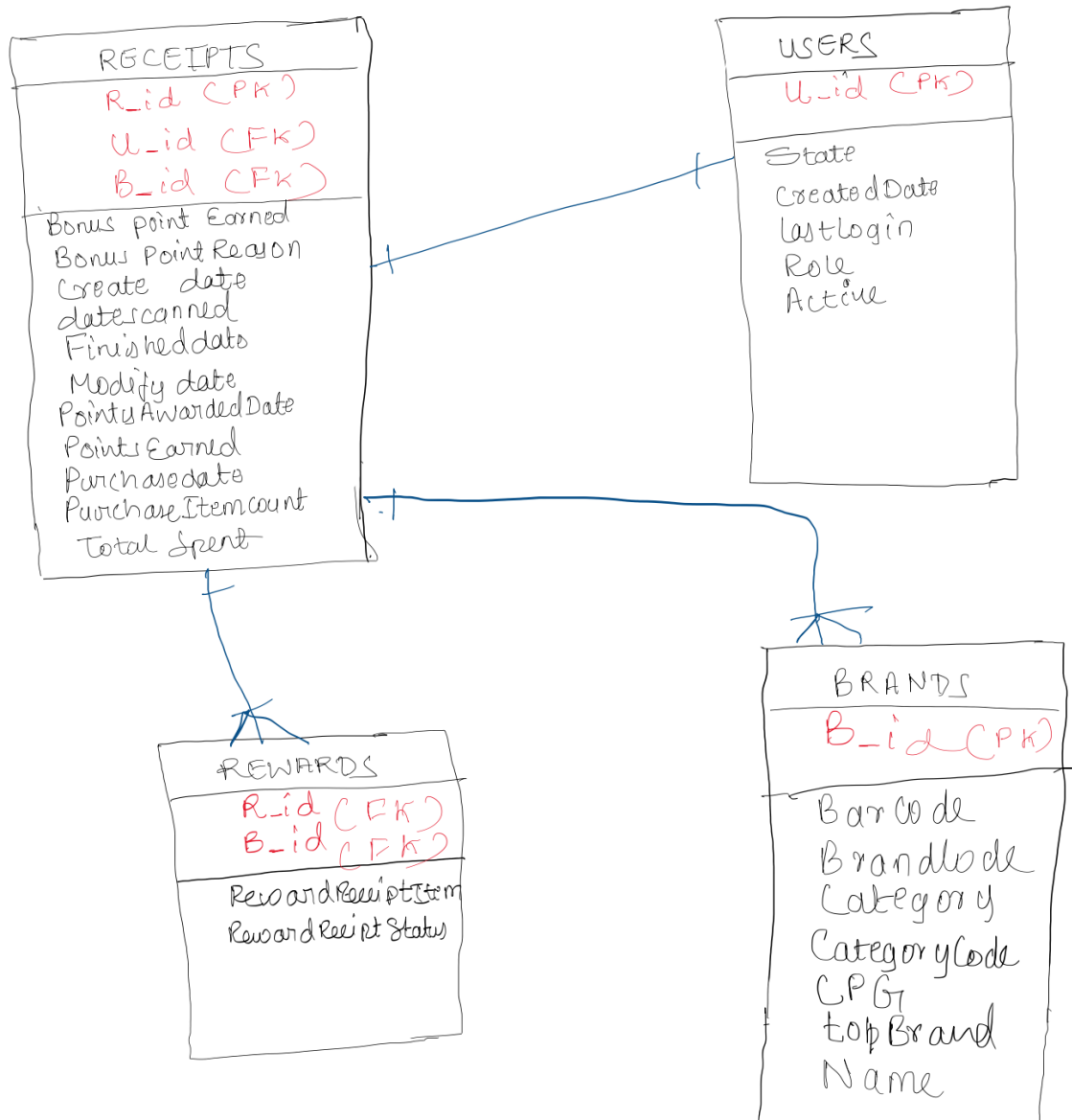


ER DIAGRAM:

Erd

Monday, June 6, 2022 3:24 PM



Data Quality Issues:

After checking for data quality issues in the Receipts.json I found that date has incorrect format (**createDate,DateScanned,finishedDate,ModifyDatePoints AwardedDate,PurchaseDate**) and it has to be in the Date format for Receipts table.

Another issue is that **rewardsReceiptItemList** contains multiple item list which can be sorted and mapped to each brand separately.

SQL QUERY

Query to give top 5 brands

with cte as(

Select B_id, count(R_id) as count,MONTH(datescanned) as Month
rank() over(partition by month order by count(R_id) desc) rk

from Receipt

group by B_id,MONTH(datescanned)

Where month(datescanned)= Max(month(datescanned))

Select C.B_id, C.count,B.Name

from cte C

join Brand B

on C.B_id = B.B_id

where C.rk < 5

Stakeholder Communication:

Hi All,

I had the chance to review the data provided in the previous email. I had few questions regarding the data:

1. Can you please tell me about the source of the data?
2. Is this data set related to the last 6 months?

I used python to understand the data and found multiple issues with the format of the data that has to change as per the final discussion.

I would also request to connect me to the Data Source team to find out the better ways to get the data and get the correct date format in the given datasets.