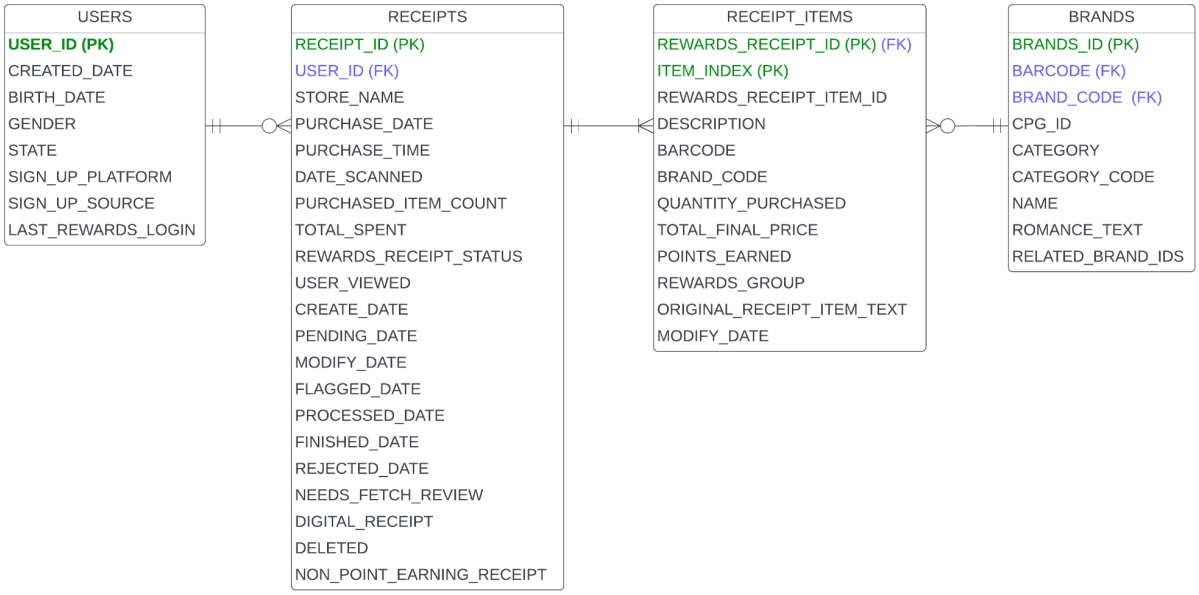
Data Analytics Take Home Test | Fetch

By Rushali Bhudeka

New Structured Relational Data Model



The ERD diagram above represents the relationship between USERS, RECEIPTS, RECEIPT_ITEM, and BRANDS.

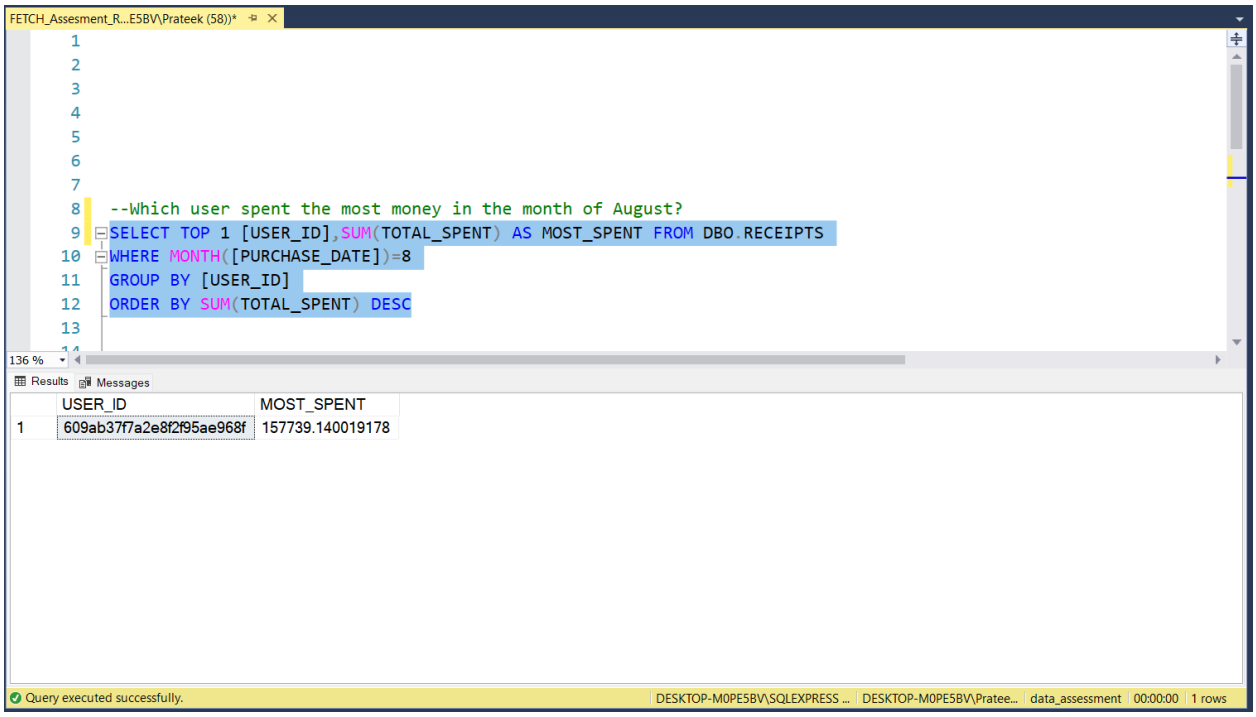
Table Name	**Primary KEY**
USERS	Users_ID
RECEIPTS	Recepit_ID
RECEIPT_ITEM	Rewards_Receipt_ID + Item_Index (CK)
BRANDS	Brands_ID

Write a query that directly answers the question(s).

- Which user spent the most money in August?

  Since the data for August is available for multiple years (2020 - 2022), the most money spent by the users was evaluated based on all years.

  The data was extracted from the Receipts table, and the attributes used to find the users with the most spent amount are - Purchase Date, Total_Spent, and User_ID.

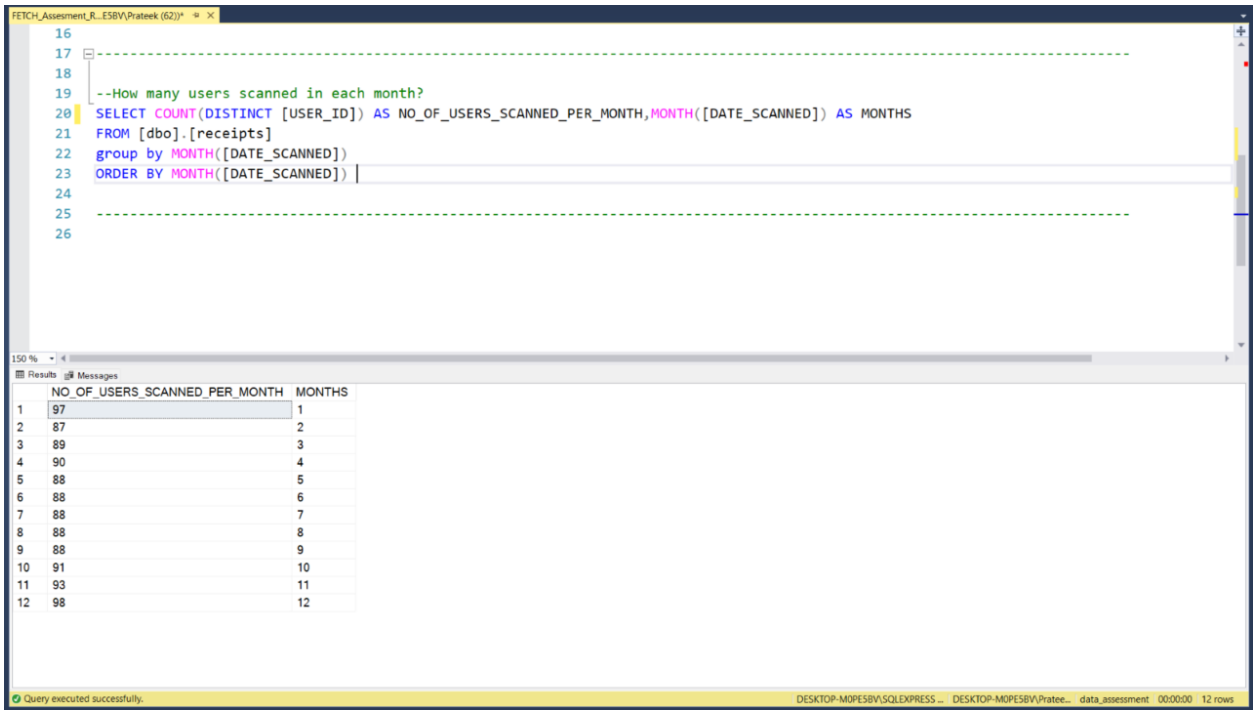


  Therefore, the user who spent the most in **August** (2020 - 2022) is **609ab37f7a2e8f2f95ae968f,** with total spending of 157739.14.

- How many users scanned in each month?

  The data for users from the Receipt table is from multiple years - 2021, 2022, and 2023. The data was analyzed based on all the years together. Therefore the result shows the number of users scanned monthly from 2021 to 2023.

  The data was extracted from the Receipts table, and the attributes used are - Date_Scanned and User_ID.



  The number of users that scanned each month:

Number of Users Scanned/ Month (2021 - 2023)	**Months**
97	January
87	February
89	March
90	April
88	May
88	June
88	July

88	August
88	September
91	October
93	November
98	December

Choose something noteworthy about the data and share it with a non-technical stakeholder.

One critical insight is the missing correlation between the Receipt_Items table and the Brands table. Even though a relation can be formed using the two attributes - Barcode and Brand Code, there are a lot of missing data from the Receipt_Items table.

As a result, the null values made it difficult to find the answers to the following questions from the stakeholders.

- Which brand saw the most dollars spent in June?
- What user bought the most expensive item?
- What is the name of the most expensive item purchased?

BRANDS Table:

BRANDS	Datatype	Null Values per Column
BRANDS_ID (PK)	Varchar	0
BARCODE	Integer	0
BRAND_CODE	Varchar	25
CPG_ID	Varchar	0
CATEGORY	Char	27
CATEGORY_CODE	Varchar	31
NAME	Varchar	0
ROMANCE_TEXT	Varchar	103
RELATED_BRAND_IDS	Varchar	243

Receipt_Items Table:

RECEIPT_ITEMS	Datatype	Null values per Column
REWARDS_RECEIPT_ID/ RECEIPT_ID (FK)	Varchar	0
ITEM_INDEX	Integer	0
REWARDS_RECEIPT_ITEM_ID	Varchar	0
DESCRIPTION	Varchar	1091
BARCODE	Integer	135369
BRAND_CODE	Varchar	205490
QUANTITY_PURCHASED	Integer	7756
TOTAL_FINAL_PRICE	Decimal	692
POINTS_EARNED	Decimal	341425
REWARDS_GROUP	Varchar	29844
ORIGINAL_RECEIPT_ITEM_TEXT	Varchar	1681
MODIFY_DATE	Timestamp	0

I choose a PowerPoint presentation to communicate my analysis with a non-technical team or stakeholders.

The link to the Powepoint is here - Fetch PPT - RB.pdf