**DATABASE MANAGEMENT & SQL**

***INTERVIEW ASSIGNMENT***

**TEAM # 2**

**Caleb Gibson**

**Dhruvi Agarwal**

**Sejal Bhargava**

**Shambhavi Dhakal**

**1) ERD**

**Table

Description automatically generated**

**2) Queries**

Out of the 4 queries, we were able to successfully solve 2 of them in SQL. The queries that we were able to solve were query 3 and query 4. Despite the challenges faced with the remaining queries, we gained valuable experience and insights. We were unable to run the queries that were associated with brands and the trans table because none of them had a common column.

**Q3.**

SELECT "rewardsReceiptStatus", COUNT(\*) AS count\_receipts, ROUND(AVG("totalSpent"), 2) AS avg\_total\_spent

FROM trans

WHERE "rewardsReceiptStatus" IN ('FINISHED', 'REJECTED')

GROUP BY "rewardsReceiptStatus"

ORDER BY "rewardsReceiptStatus" ASC;

"FINISHED" 518 80.85

"REJECTED" 71 23.33

**Q4.**

SELECT "rewardsReceiptStatus", SUM("purchasedItemCount") AS total\_items\_purchased

FROM trans

WHERE "rewardsReceiptStatus" IN ('FINISHED', 'REJECTED')

GROUP BY "rewardsReceiptStatus"

ORDER BY total\_items\_purchased DESC;

"FINISHED" 8184

"REJECTED" 173

**Graphical user interface, application, table

Description automatically generated**

SELECT "rewardsReceiptStatus", COUNT(\*) AS count\_receipts, SUM("purchasedItemCount") AS total\_items\_purchased

FROM trans

WHERE "rewardsReceiptStatus" IN ('FINISHED', 'REJECTED')

GROUP BY "rewardsReceiptStatus"

ORDER BY total\_items\_purchased DESC;

**Graphical user interface, application, table

Description automatically generated**

**3) Data Quality**

The major data quality issue is that across the tables, many have inconsistent column names and inconsistent data types when it comes to the data within the rows.

**4) Next Steps**

The given data was run on SQL to assess the overall quality; however, multiple issues were identified for the data.

* Initially assessed the data by looking at Excel files and then saw quality issues when importing data.
* The data was highly inconsistent. Multiple tables had missing and null values.
* Multiple duplicate entries existed for all tables, leading to issues while creating a primary key for the SQL database.
  + Needed to create a new column in the user’s table to input it.
* Names for the tables and records were unsystematic with some names in all caps while others followed a weird pattern of lower and uppercase letters.
  + The query kept returning an error stating that a particular column did not exist due to the case-sensitive names.
  + Slow down query creation
* Record names had anomalies with names of brands also including referencing numbers within the brand title.
  + Showcasing an overall issue of multiple data types of variables within columns
* The ID columns can create performance issues as they are long making it harder to type and needing more data storage space.
* Inefficient join conditions were found in different tables.
  + No ability to join brands table to trans.
  + Unable to run all the queries.

**Questions needed answered for further analysis of the tables to know:**

* What connection could be made to connect the brand and receipts table?
* When assessing the brand’s table and the trans table, there are no columns that match each other. Thus, when trying to run a query to assess brands and their transactions, there is no way to do so. So, a column needs to be added to the brand’s table.
* Are there supposed to be ‘Accepted’ results in the “rewardsReceiptStatus” column of the receipts table?
  + If not, there are non but there is a ‘Finished’ result that could be referring to ‘Accepted.’
* Why all barcodes are the same in the brands table?