

Preliminary Data Analysis of Reciept, Users & Brands dataset

Aniruddha Tambe <tambe.ani@northeastern.edu>

Tue 2/1/2022 12:00 AM

To: Aniruddha Tambe <tambe.ani@northeastern.edu>

Hi,

I have performed the preliminary data analysis of Receipts, Brands & Users dataset using SQL queries on Azure Data Studio.

Below is an overview the analysis:

1. Duplication found in Users
2. Numerous null values found in all 3 datasets
3. Inconsistency found in Receipts and Brands & Users
 - a. Users found on Receipts not present in Users dataset

A detailed report can be found in my shared github repository.

Before performing a complete data analysis some information to be confirmed,

1. Information contained with the Rewards Reciept Item List needs to clarified
2. Data Imputation logic, if needed, needs to clarified for the entire dataset
3. Business logic behind reward generation needs clarification

Please help arrange for the knowledge transfer sessions for above mentioned points.

Prior to migrating to production, all the data inconsistencies must be fixed for data analysis queries to work seamlessly.

Thanks,
Aniruddha Tambe