Ajay Venkatesh Pandravada

Fetch Rewards Coding Exercise - Data Analyst

Data Review and Data Model

Data Sources:

- Users
- Brands
- Receipts

Pre-Processing:

- The code reads the JSON files using pandas and converts them into DataFrames.
- The data is pre-processed to extract relevant information and convert date fields into appropriate date-time format and break nested JSON list to a new dataframe
- The code merges and cleans the data to create structured tables for further analysis.

Data Tables:

The provided data sources were not in a normalized form, which caused issues and increased time complexity while querying the data. To address these concerns, normalization techniques were applied to create data tables in the Third Normal Form (3NF). This normalization process removed partial and transitive dependencies, resulting in a more streamlined and efficient data structure.

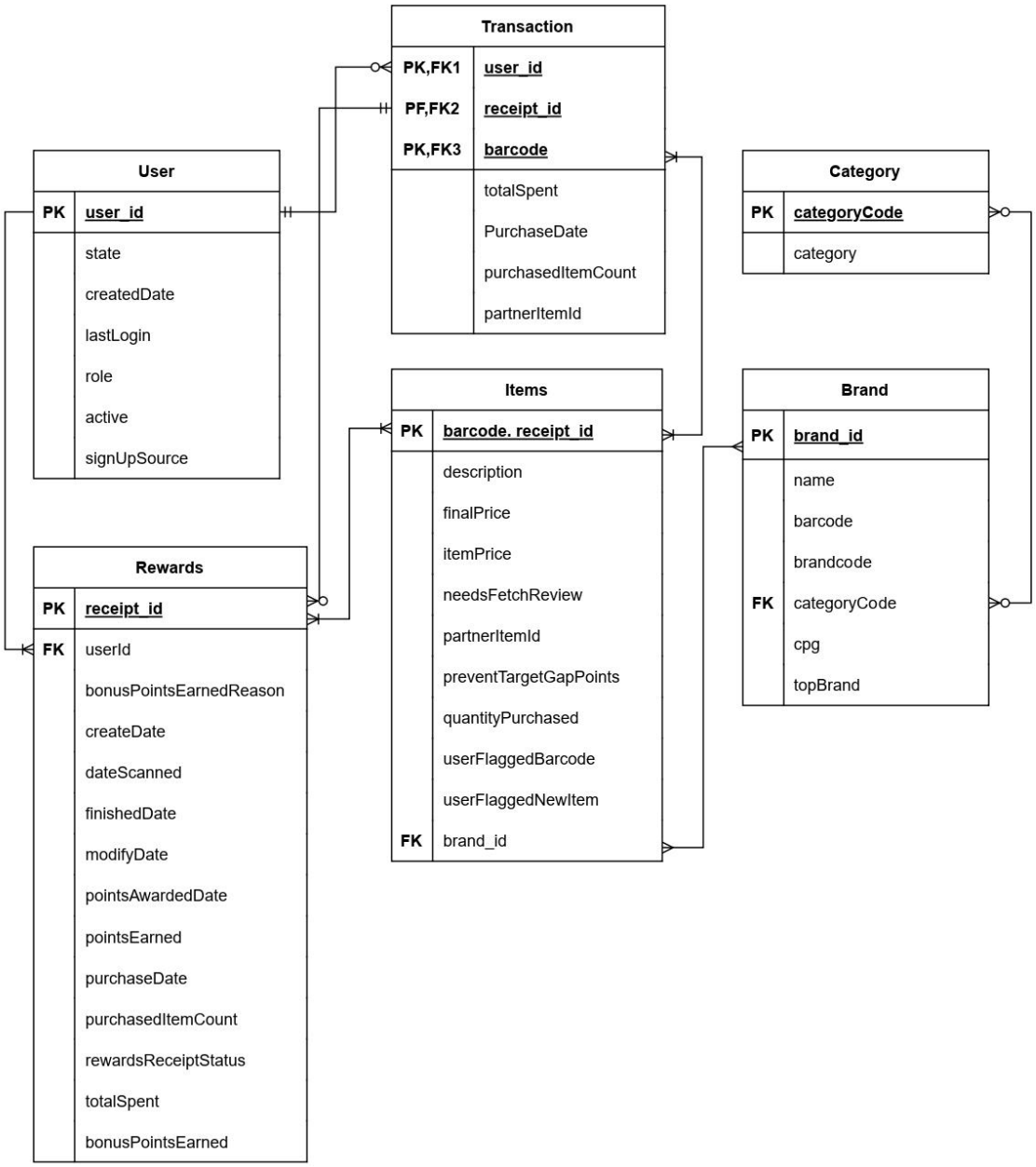
- user_table: Represents user-related information such as user ID, state, created date, and last login.
- brand_table: Represents brand-related information, including brand ID, barcode, brand code, and category.
- category_table: Contains the mapping between category and category code.
- rewards_table: Contains information about the rewards associated with each receipt.
- items_table: Represents individual items present in each receipt, including the receipt ID, barcode, description, and other relevant details.
- transaction_table: Represents transaction-related information in the star schema design, including the user ID, receipt ID, barcode, partner item ID, purchase date, and final price.

Structured Relational Data Model:

- The code transforms unstructured data into structured relational tables based on the entities present in the data.
- These tables are connected using primary keys and foreign keys to establish relationships and enable efficient querying and analysis.

Ajay Venkatesh Pandravada

Relational Schema:



The schema consists of multiple entities that represent different aspects of the data. Each entity has its own set of attributes, which provide specific details about the entity. In the schema, the primary key (PK) uniquely identifies each record within an entity, while the foreign key (FK) establishes relationships with other entities.