Venkatasubramanian Narasimman
MS in Information Management
University of Illinois at Urbana Champaign

Fetch Rewards Data Analytics Internship Take Home Assessment
Exercise 1 – Data review and Relational Data Model

Tables:

users
brands
receipts
receipt_items

Keys:

Primary Key of users – ID – referenced in receipts as USER_ID
Primary Key of brands – ID (not a foreign key in any relation)
Primary Key of receipts – ID – referenced as REWARDS_RECEIPT_ID in receipt_items
Primary Key of receipt_items – REWARDS_RECEIPT_ITEM_ID (not a foreign key)

Another meaningful key of receipt_items can be a composite key with the combination of REWARDS_RECEIPT_ID and ITEM_INDEX.
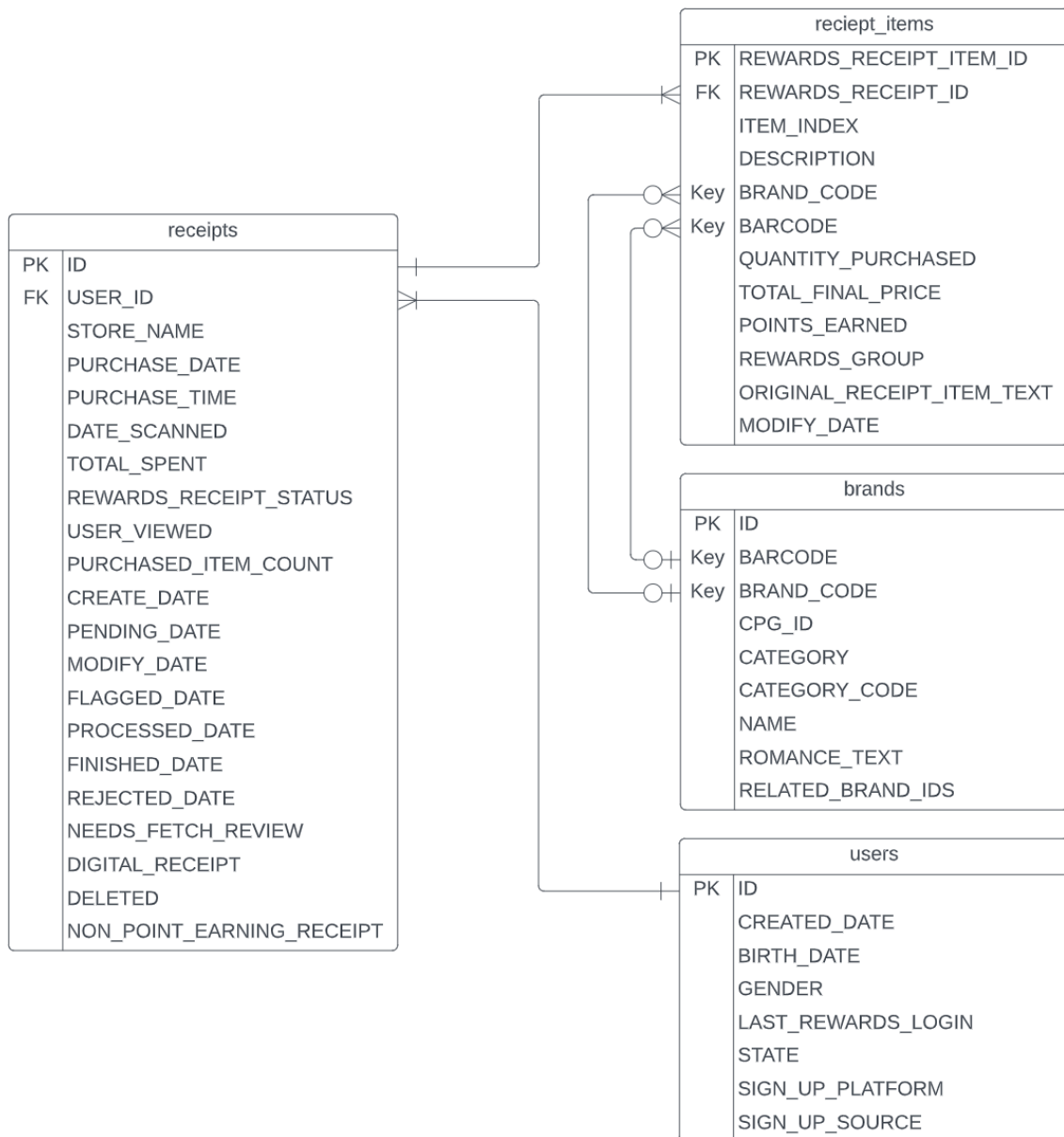
BRAND_CODE and BARCODE are found in both receipt_items and brands. However it is not necessarily common to both the relations.

Cardinalities:

1. users to receipts – one to [one or many] (one or many receipts can be made by one user)
2. receipts to receipt_item – one to [one or many] (one receipt ID is referenced atleast once or multiple times, that is for each of the items in that particular receipt in the receipt_items relation)
3. brands to receipt_items – 0 or one to 0 or many (any of BRAND_CODE and BARCODE appearing in either of brands or receipt_items may or may not be present in the other relation.)

Relational Schema

Each box denotes an entity (relation) along with the attributes of the entity. The primary key of the entity is denoted by PK and the foreign key is denoted by FK.
The cardinalities are also marked using the crow feet diagram (as mentioned above)

Category depends on the category code. The brands relation can be decomposed where we have a separate relation for category with just the category and the category_code.

Using the relational data model, a schema was created.

```
create database fetch_rewards;
use fetch_rewards;

SET innodb_strict_mode = 0;
SET FOREIGN_KEY_CHECKS = 0;

CREATE TABLE IF NOT EXISTS `brands` (
 `ID` VARCHAR(30) NOT NULL PRIMARY KEY,
 `BARCODE` VARCHAR(40),
 `BRAND_CODE` VARCHAR(1024),
 `CPG_ID` VARCHAR(1024),
 `CATEGORY` VARCHAR(1024),
 `CATEGORY_CODE` VARCHAR(1024),
 `NAME` VARCHAR(1024),
 `ROMANCE_TEXT` VARCHAR(1024),
 `RELATED_BRAND_IDS` VARCHAR(1024)
);

CREATE TABLE IF NOT EXISTS `users` (
 `CREATED_DATE` VARCHAR(1024),
 `BIRTH_DATE` VARCHAR(1024),
 `GENDER` VARCHAR(1024),
 `LAST_REWARDS_LOGIN` VARCHAR(1024),
 `STATE` VARCHAR(1024),
 `SIGN_UP_PLATFORM` VARCHAR(1024),
 `SIGN_UP_SOURCE` VARCHAR(1024),
 `ID` VARCHAR(24) NOT NULL PRIMARY KEY
);


CREATE TABLE IF NOT EXISTS `receipts` (
 `ID` VARCHAR(36) NOT NULL PRIMARY KEY,
 `STORE_NAME` VARCHAR(86),
 `PURCHASE_DATE` VARCHAR(20),
 `PURCHASE_TIME` VARCHAR(11),
 `DATE_SCANNED` VARCHAR(24),
 `TOTAL_SPENT` VARCHAR(25),
 `REWARDS_RECEIPT_STATUS` VARCHAR(1024),
 `USER_ID` VARCHAR(24),
 `USER_VIEWED` VARCHAR(1024),
 `PURCHASED_ITEM_COUNT` VARCHAR(20),
 `CREATE_DATE` VARCHAR(24),
 `PENDING_DATE` VARCHAR(24),
 `MODIFY_DATE` VARCHAR(24),
```

```sql
 `FLAGGED_DATE` VARCHAR(24),
 `PROCESSED_DATE` VARCHAR(24),
 `FINISHED_DATE` VARCHAR(24),
 `REJECTED_DATE` VARCHAR(24),
 `NEEDS_FETCH_REVIEW` VARCHAR(4),
 `DIGITAL_RECEIPT` VARCHAR(10),
 `DELETED` VARCHAR(10),
 `NON_POINT_EARNING_RECEIPT` VARCHAR(10),
  FOREIGN KEY (USER_ID) REFERENCES users (ID)
);


CREATE TABLE IF NOT EXISTS `receipt_items` (
 `REWARDS_RECEIPT_ID` VARCHAR(36),
 `ITEM_INDEX` VARCHAR(20),
 `REWARDS_RECEIPT_ITEM_ID` VARCHAR(32) NOT NULL PRIMARY KEY,
 `DESCRIPTION` VARCHAR(400),
 `BARCODE` VARCHAR(40),
 `BRAND_CODE` VARCHAR(150),
 `QUANTITY_PURCHASED` VARCHAR(20),
 `TOTAL_FINAL_PRICE` VARCHAR(20),
 `POINTS_EARNED` VARCHAR(30),
 `REWARDS_GROUP` VARCHAR(70),
 `ORIGINAL_RECEIPT_ITEM_TEXT` VARCHAR(410),
 `MODIFY_DATE` VARCHAR(24),
  FOREIGN KEY (REWARDS_RECEIPT_ID) REFERENCES receipts (ID)
);
```

The create.sql file has been added to the repository.

In order to insert records into the relations, a python script was created to convert CSV files into SQL insert statements.

The python script has been added to the repository.

Insert files were too large to be added to the repository. Snapshots of a file from the MySQL Workbench has been attached.

```sql
use fetch_rewards;

INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5a8c35dde4b0ccf165fac9e6", "511111904175", "PEPPERIDGE FARM", "5a734034e4bbd58f376be874", "Grocery", "GROCERY", "Pepperidge Farm", "Pepperidge Farm has b
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("6234af8f4e09b6067c237adb", "511111212997", "CHEX MIX", "62339666a8942a67af4934aa1", "Snacks", "SNACKS", "CHEX MIX", "", "");
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5332f7d3e4b03c9a25efd14e", "511111803393", "", "5332f52e4b03c9a25efd0aa", "Snacks", "SNACKS", "Cheez-It", "", "");
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5d0412e9a3a0185149904fd26", "511111304982", "BETTER HOMES & GARDENS", "53e10d6368abd3c7065097cc", "Magazines", "MAGAZINES", "Better Homes & Gardens Magazi
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("621e777eacedc865cefa99a7", "511111912859", "TRUVIA", "621e7754d759b10969cbcc80", "Baking", "BAKING", "Truvia®", "", "");
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5887a051e4b02187f85cdad0", "511111401209", "BAKER'S", "53e10d6368abd3c7065097cc", "Baking", "BAKING", "Baker's", "Baker's premium ingredients bring your
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("592466bee41d61fcea3d135", "511111700746", "PG TIPS", "5332f5f6e4b03c9a25efd0b4", "Beverages", "BEVERAGES", "PG Tips", "PG Tips strong, distinct blend co
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5887a0fde4b02187f85cdad1", "511111801191", "CAPRI SUN", "53e10d6368abd3c7065097cc", "Beverages", "BEVERAGES", "Capri Sun", "The go-to drink for real, act
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5bd2011f90fa074576779a17", "511111804652", "HUGGIES", "550b2565e4b001d5e9e41467", "Baby", "BABY", "Huggies", "The first moment you hug your baby, the res
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5332f6ede4b03c9a25efd0e5", "511111603870", "BLUE BUNNY", "53e10d6368abd3c7065097cc", "Frozen", "FROZEN", "Blue Bunny", "", "");
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("592466bde41d61fcea3d120", "511111509940", "AXE", "5332f5f6e4b03c9a25efd0b4", "Personal Care", "PERSONAL_CARE", "Axe", "Explore Axe's universe of men's g
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("57ebc026e4b0ac389136a338", "511111582098", "KNUDSEN", "53e10d6368abd3c7065097cc", "Dairy & Refrigerated", "DAIRY_AND_REFRIGERATED", "Knudsen", "Delicious
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("68abf7099b6e385567eee851", "511111709848", "LESTOIL", "53e10d6368abd3c7065097cc", "Cleaning & Home Improvement", "CLEANING_AND_HOME_IMPROVEMENT", "LESTOI
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5e7537244f392912df7c86d0", "511111386672", "VIZZY", "5332f709e4b03c9a25efd0f3", "Beer, Hard Cider & Seltzer", "BEER_HARD_CIDER_AND_SELTZER", "Vizzy", "In
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("604a430bac99f04b86c39a4a", "511111586707", "BABY DOVE", "5332f5f6e4b03c9a25efd0b4", "Personal Care", "PERSONAL_CARE", "Baby Dove", "From baby wash to bab
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5332f756e4b03c9a25efd113", "511111587343", "MURRAY", "53e10d6368abd3c7065097cc", "Baking", "BAKING", "Murray", "", "");
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("61b25afd2e105238e19ad07b", "511111011048", "MUS KITCHEN", "61b25af3a6516866771696da", "Candy & Sweets", "CANDY_AND_SWEETS", "MUS KITCHEN", "", "");
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5f4e516eac99f03f103c8e12", "511111907169", "FRONTERA FROZEN", "5e825d64f221c312e698a62a", "Frozen", "FROZEN", "Fronteram", "Mexican foods inspired by reg
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5332fa73e4b03c9a25efd21d", "511111203049", "W", "53e10d6368abd3c7065097cc", "Beverages", "BEVERAGES", "Caffeine-free Diet Coke", "", "");
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("57ebbd44e4b0ac389136a32d", "511111002200", "BULL'S-EYE", "53e10d6368abd3c7065097cc", "Condiments & Sauces", "CONDIMENTS_AND_SAUCES", "Bull's Eye", "The b
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("61dc5755f04efd3f1949fe8f", "511111512462", "", "610c4a3b5806c5140572e820", "", "", "Glaceau smartwater", "", "");
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("592466bde41d61fcea3d123", "511111600910", "BRUMMEL AND BROWN", "5e5ea9cc9c50da1291a6d074", "Dairy & Refrigerated", "DAIRY_AND_REFRIGERATED", "BRUMMEL AN
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("58861c7e4e8d0d20bc42c4e9", "511111901372", "PERONI", "63a0b3fdd1e6539543fdc2a", "Beer, Hard Cider & Seltzer", "BEER_HARD_CIDER_AND_SELTZER", "Peroni Ita
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("62b3fe9fcbf3d978d9ed32f1", "511111614357", "GOPUFF", "62b3fddacbf3d978d9ed3270", "Retailers", "RETAILERS", "Gopuff", "From snacks and drinks to cleaning
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5da60932a60b07376833e356", "511111405810", "ONE A DAY® KIDS AND TEENS", "53e10d6368abd3c7065097cc", "Health & Wellness", "HEALTHY_AND_WELLNESS", "ONE A D
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5d66d38ca3a010093ab34724", "511111605263", "CLEAN EATING", "5d5d4fd16d573b23d1bc7909", "Magazines", "MAGAZINES", "Clean Eating Magazine", "Clean Eating p
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("62699290d5bf52abce40aaa", "511111013402", "DENTASTIX", "61ba20c32064d21d1548c32e", "Pet", "PET", "DENTASTIX", "", "");
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("58b598dbe4b0857c2ddb7253", "511111101017", "MICKEYS", "63a0b3fdd1e6539543fdc2a", "Beer, Hard Cider & Seltzer", "BEER_HARD_CIDER_AND_SELTZER", "Mickey's
INSERT INTO `brands` (`ID`, `BARCODE`, `BRAND_CODE`, `CPG_ID`, `CATEGORY`, `CATEGORY_CODE`, `NAME`, `ROMANCE_TEXT`, `RELATED_BRAND_IDS`) VALUES ("5f1aea63ac99f054aeb9e7ee", "511111906797", "LIQUID PLUMR", "53e10d6368abd3c7065097cc", "Cleaning & Home Improvement", "CLEANING_AND_HOME_IMPROVEMENT", "L
```