

# ER diagram

fact table  
(or)  
(central table)

dimension

receipts	
receipt_id (PK)	varchar
user_id (FK)	varchar
create_date	Timestamp
date_scanned	timestamp
finished_date	timestamp
points_earned_date	timestamp
points_earned	float
purchase_date	timestamp
purchased_item_count	integer
rewards_receipt_status	text
total_spent	float

receipt_items	
receipt_id (FK)	varchar
barcode	varchar
description	text
final_price	float
item_price	float
quantity_purchased	integer
points_earned	float
points_not_awarded_reason	text
rewards_group	text
rewards_product_partner_id	varchar

dimension table

users	
user_id (PK)	varchar
active	bool
role	varchar
signup_source	varchar
state	varchar
created_date	timestamp
last_login	timestamp

PK → Primary key  
FK → Foreign key

1 — \*  
↳ one → many  
relation

dimension table

brands	
barcode	varchar
category	varchar
category_code	varchar
brand_name (PK)	varchar
top_brand	bool
brand_id (PK)	varchar
cpg_id	varchar
cpgref	varchar
brand_code	varchar

## Summary:-

→ The above ER diagram is the only possible relational model to have a clean & efficient model

→ Decisions made based on analysis & info given

1<sup>st</sup>:- Converted Receipts (Central table) into two normalized table to keep them clean

2<sup>nd</sup>:- Avoided possible Star Schema to avoid having possible null values

3<sup>rd</sup>:- Added 100+ users into Users table to create relationship model

4<sup>th</sup>:- Brands table

↳ Inspected barcodes, brand names, brand-id columns

→ found 7 brands have two brand-Ids and barcodes

→ After inspection found that they were wrongly populated

→ Ended up creating (brand-Id, brand-name) as primary key

→ Making { brand-Id, brand-name, bar-code } as primary key is possible, but it's possible that a brand has multiple barcodes  
↓

→ Made a question for business stakeholders about barcodes column  
↓

The validation made due to data provided