

Fetch Example Diagram Documentation

Link to schema: <https://app.quickdatabasediagrams.com/#/d/jCPv6V>

@skrzym ERD for Fetch Rewards Coding Exercise - Analytics Engineer

Major modifications from raw JSON files include the flattening of nested components in 'receipts.json'. Specifically details about line items and bonus point reasons.

Technical team question: Are logs available for slow changing dimension analysis?. Questions like 'Did 'top brand' status impact the brand's frequency on receipts by top users?' are not answerable with just this data.

Please see notes per table below for further considerations about the proposed model.

receipts

Only comment here would be potentially renaming any 'date' field to 'at' if its a specific timestamp vs just a date value for easy differentiation in queries and downstream models.

Field	Description	Type	Default	Other
receipt_id				PK
user_id				FK
created_date				
scanned_date				
finished_date				
modified_date				
points_awarded_date				
purchased_date				
rewards_receipt_status				
points_earned				
bonus_points_earned				
purchased_item_count				
total_spent				

receipt_bonus_point_reasons

Would love more details about bonus point reason breakdowns (Points earned by this code for this specific transaction, bonus point campaign associations). Potential future ask by marketing or customer success teams to know how effective certain bonus incentives were or to do A/B testing here.

Field	Description	Type	Default	Other
receipt_bonus_point_reason_id				PK
receipt_id				FK
bonus_points_earned_reason				

users

Role was indicated to be a constant field but there are more than one unique values populating it. Potential DQ concern.

Field	Description	Type	Default	Other
user_id				PK
last_login_at				
created_date	Potentially rename to 'created_at'			
is_active				
state				
role	Potential DQ concern			

brands

Unclear what barcode is supposed to be referring to here since the data is at the 'brand' grain. In addition, barcode looks to have a data quality issue as many of them start with the exact same string of 5-6 digits.

Field	Description	Type	Default	Other
brand_id				PK
brand_code	Known data quality issue that violates the UNIQUE constraint			FK, UNIQUE
name				
barcode				
category				
category_code				
cpg				
is_top_brand				

line_items

This table is completely manufactured from data within receipt.json elements. EDA shows core elements like barcode, price, quantity, description are split amongst 3 "classes" - Valid, Flagged, Missing(Original_Metabrite). This makes this table somewhat sparse and a potential candidate for consolidation.

Field	Description	Type	Default	Other
line_item_id				PK
receipt_id				FK
brand_code				FK
barcode				
user_flagged_barcode				
user_flagged_new_item				

Field	Description	Type	Default	Other
user_flagged_price				
user_flagged_quantity				
user_flagged_description				
needs_fetch_review_reason				
points_not_awarded_reason				
points_payer_id				
rewards_group				
rewards_product_partner_id				
partner_item_id				
description				
quantity_purchased				
item_price				
discounted_item_price				
price_after_coupon				
final_price				
needs_fetch_review				
prevent_target_gap_points				
metabrite_campaign_id				
original_metabrite_barcode				
original_metabrite_description				
original_metabrite_quantity_purchased				
original_final_price				
original_metabrite_item_price				
original_receipt_item_text				
competitor_rewards_group				
item_number				
points_earned				
target_price				
competitive_product				
deleted				