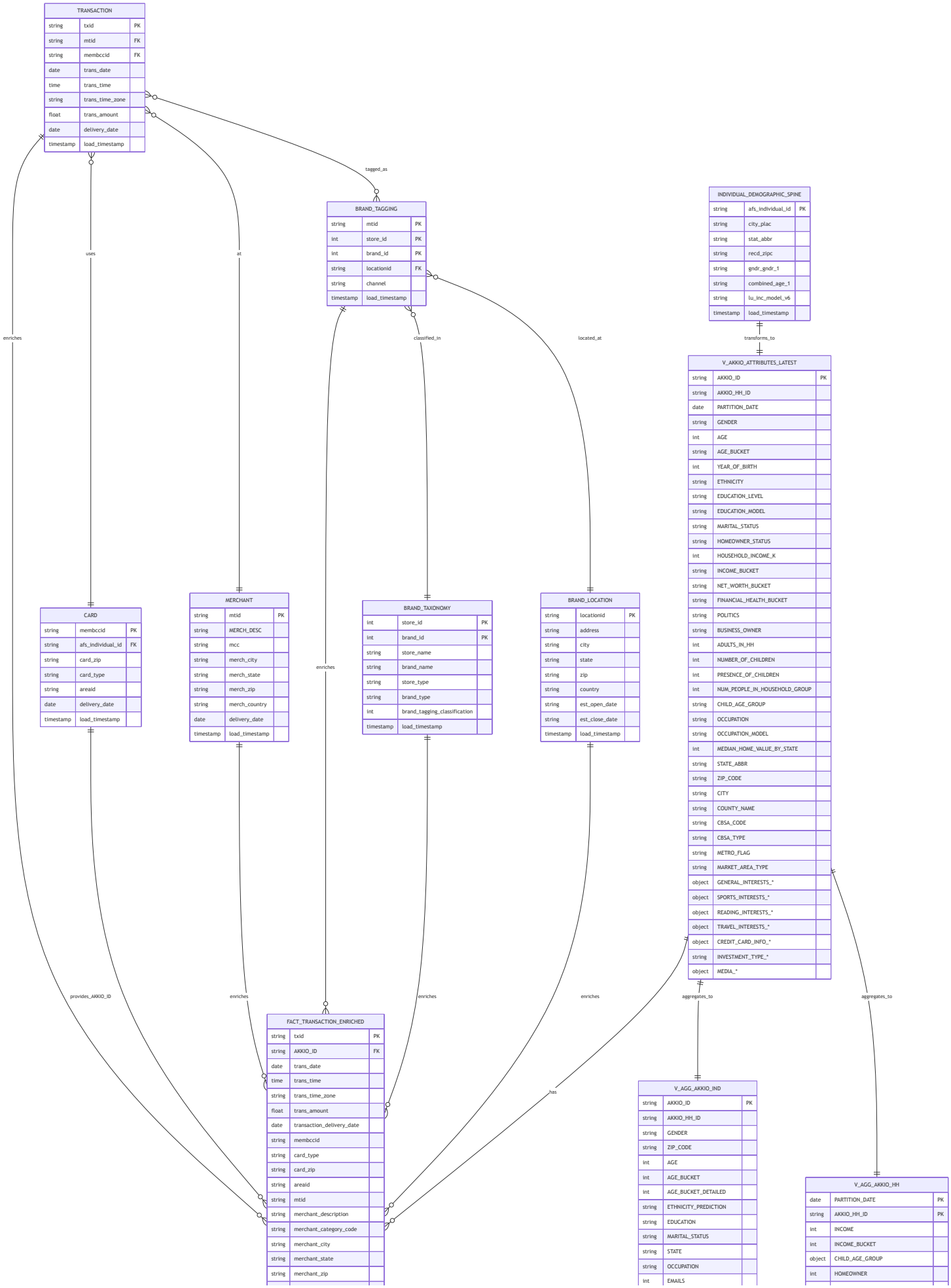
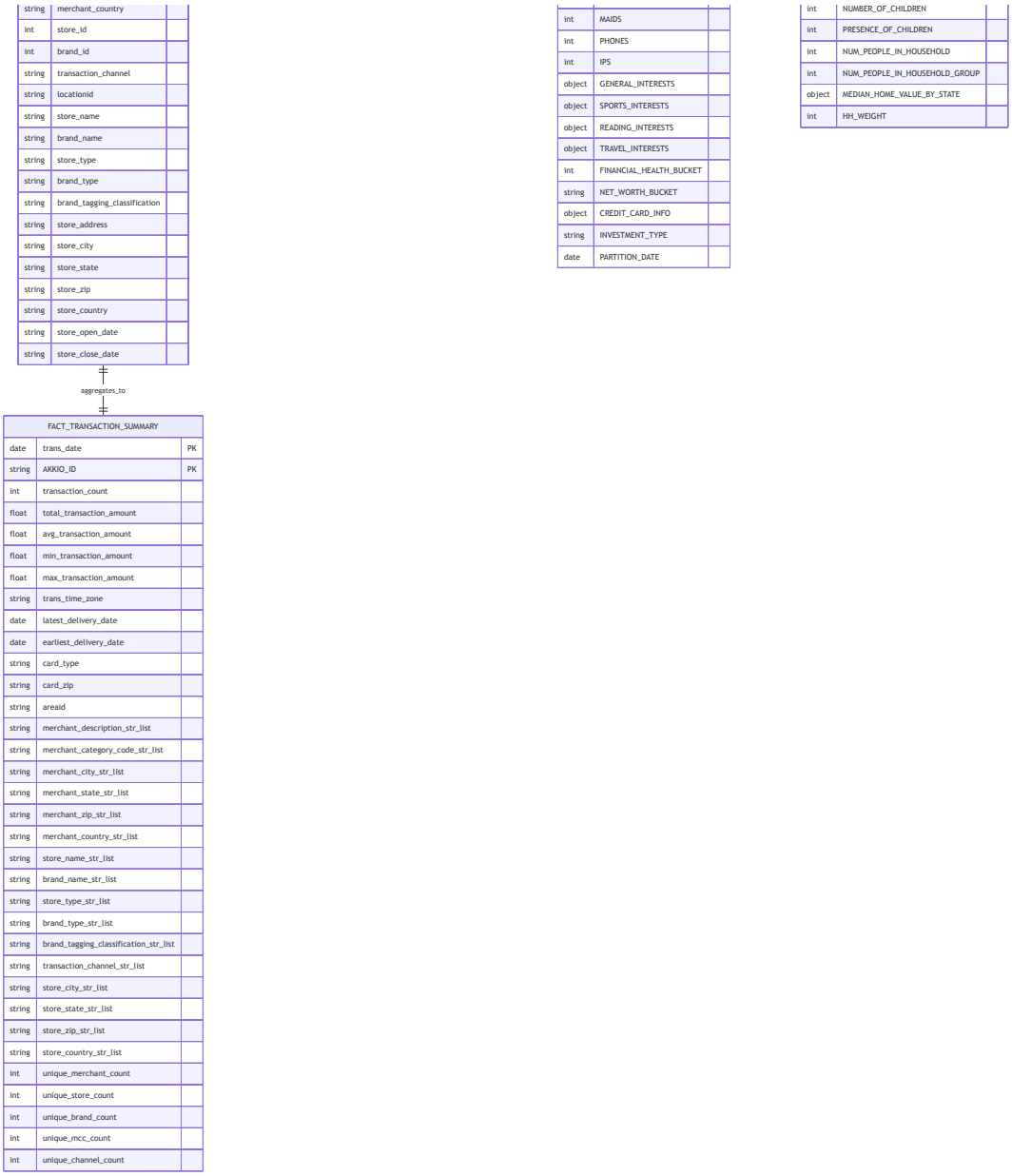
Data Model ERD

Entity Relationship Diagram


TRANSACTION
string txid PK
string mtid FK
string membccid FK
date trans_date
time trans_time
string trans_time_zone
float trans_amount
date delivery_date
timestamp load_timestamp
tagged_as
uses
at
enriches
BRAND_TAGGING
string mtid PK
int store_id PK
int brand_id PK
string locationid FK
string channel
timestamp load_timestamp
classified_in
located_at
INDIVIDUAL_DEMOGRAPHIC_SPINE
string afs_individual_id PK
string city_plac
string stat_abbr
string recd_zipc
string gndr_gndr_1
string combined_age_1
string lu_inc_model_v6
timestamp load_timestamp
transforms_to
V_AKKIO_ATTRIBUTES_LATEST
string AKKIO_ID PK
string AKKIO_HH_ID
date PARTITION_DATE
string GENDER
int AGE
string AGE_BUCKET
int YEAR_OF_BIRTH
string ETHNICITY
string EDUCATION_LEVEL
string EDUCATION_MODEL
string MARITAL_STATUS
string HOMEOWNER_STATUS
int HOUSEHOLD_INCOME_K
string INCOME_BUCKET
string NET_WORTH_BUCKET
string FINANCIAL_HEALTH_BUCKET
string POLITICS
string BUSINESS_OWNER
int ADULTS_IN_HH
int NUMBER_OF_CHILDREN
int PRESENCE_OF_CHILDREN
int NUM_PEOPLE_IN_HOUSEHOLD_GROUP
string CHILD_AGE_GROUP
string OCCUPATION
string OCCUPATION_MODEL
int MEDIAN_HOME_VALUE_BY_STATE
string STATE_ABBR
string ZIP_CODE
string CITY
string COUNTY_NAME
string CBSA_CODE
string CBSA_TYPE
string METRO_FLAG
string MARKET_AREA_TYPE
object GENERAL_INTERESTS_*
object SPORTS_INTERESTS_*
object READING_INTERESTS_*
object TRAVEL_INTERESTS_*
object CREDIT_CARD_INFO_*
string INVESTMENT_TYPE_*
object MEDIA_*
CARD
string membccid PK
string afs_individual_id FK
string card_zip
string card_type
string areaid
date delivery_date
timestamp load_timestamp
MERCHANT
string mtid PK
string MERCH_DESC
string mcc
string merch_city
string merch_state
string merch_zip
string merch_country
date delivery_date
timestamp load_timestamp
BRAND_TAXONOMY
int store_id PK
int brand_id PK
string store_name
string brand_name
string store_type
string brand_type
int brand_tagging_classification
timestamp load_timestamp
BRAND_LOCATION
string locationid PK
string address
string city
string state
string zip
string country
string est_open_date
string est_close_date
timestamp load_timestamp
provides_AKKIO_ID
enriches
enriches
enriches
enriches
aggregates_to
aggregates_to
has
FACT_TRANSACTION_ENRICHED
string txid PK
string AKKIO_ID FK
date trans_date
time trans_time
string trans_time_zone
float trans_amount
date transaction_delivery_date
string membccid
string card_type
string card_zip
string areaid
string mtid
string merchant_description
string merchant_category_code
string merchant_city
string merchant_state
string merchant_zip
V_AGG_AKKIO_IND
string AKKIO_ID PK
string AKKIO_HH_ID
string GENDER
string ZIP_CODE
int AGE
int AGE_BUCKET
int AGE_BUCKET_DETAILED
string ETHNICITY_PREDICTION
string EDUCATION
string MARITAL_STATUS
string STATE
string OCCUPATION
int EMAILS
V_AGG_AKKIO_HH
date PARTITION_DATE PK
string AKKIO_HH_ID PK
int INCOME
int INCOME_BUCKET
object CHILD_AGE_GROUP
int HOMEOWNER



Source Table Relationships (Data Lineage)

- **INDIVIDUAL_DEMOGRAPHIC_SPINE → V_AKKIO_ATTRIBUTES_LATEST**: Source table transformed into normalized individual attributes dimension
- **TRANSACTION → FACT_TRANSACTION_ENRICHED**: Core transaction facts enriched with additional attributes
- **CARD → FACT_TRANSACTION_ENRICHED**: Provides `AKKIO_ID` via `membccid` join (LEFT JOIN)
- **MERCHANT → FACT_TRANSACTION_ENRICHED**: Provides merchant details via `mtid` join (LEFT JOIN)
- **BRAND_TAGGING → FACT_TRANSACTION_ENRICHED**: Provides brand/store IDs and channel via `mtid` join (LEFT JOIN)
- **BRAND_TAXONOMY → FACT_TRANSACTION_ENRICHED**: Provides brand/store names and classifications via `store_id + brand_id` join from BRAND_TAGGING (LEFT JOIN)

- **BRAND_LOCATION** → **FACT_TRANSACTION_ENRICHED**: Provides store location details via `locationid` join from BRAND_TAGGING (LEFT JOIN)
- **TRANSACTION** (N) ————————> (1) **CARD**: Many transactions use one card
- **TRANSACTION** (N) ————————> (1) **MERCHANT**: Many transactions occur at one merchant
- **TRANSACTION** (N) ————————< (M) **BRAND_TAGGING**: Many transactions can be tagged with many brands
- **BRAND_TAGGING** (N) ————————> (1) **BRAND_TAXONOMY**: Many brand taggings reference one brand taxonomy
- **BRAND_TAGGING** (N) ————————> (1) **BRAND_LOCATION**: Many brand taggings reference one location

dbt Model Relationships

- **V_AKKIO_ATTRIBUTES_LATEST** (1) ————————< (N) **FACT_TRANSACTION_ENRICHED**: One individual can have many transactions
- **FACT_TRANSACTION_ENRICHED** (N) ————————> (1) **FACT_TRANSACTION_SUMMARY**: Many transaction detail rows aggregate to one daily summary row per individual
- **V_AKKIO_ATTRIBUTES_LATEST** (1) ————————> (1) **V_AGG_AKKIO_IND**: One individual per PARTITION_DATE aggregates to one individual aggregation row per PARTITION_DATE
- **V_AKKIO_ATTRIBUTES_LATEST** (1) ————————> (1) **V_AGG_AKKIO_HH**: One individual per PARTITION_DATE aggregates to one household aggregation row per PARTITION_DATE (currently 1:1, structured for future household scenarios)

Data Model Notes

dbt Models

- **V_AKKIO_ATTRIBUTES_LATEST**: Individual Attributes Dimension - One row per individual with all normalized demographic attributes. Primary key is `AKKIO_ID` (formerly `afs_individual_id`). Contains 800+ demographic attributes with normalized values for gender, ethnicity, politics, income, net worth, financial health, occupation, interests (general, sports, reading, travel), credit card info, investment types, and media consumption preferences. Generated from `INDIVIDUAL_DEMOGRAPHIC_SPINE` source table. Includes household composition fields (NUMBER_OF_CHILDREN, PRESENCE_OF_CHILDREN, CHILD_AGE_GROUP, etc.).
- **FACT_TRANSACTION_ENRICHED**: Detail Transaction Fact Table - Denormalized transaction table with `AKKIO_ID` for easy joining to attributes table. Contains granular detail about each individual transaction. Built by joining 6 source tables:

    - **TRANSACTION** (base table): Transaction facts (txid, trans_date, trans_time, trans_amount, etc.)
    - **CARD** (LEFT JOIN on `membccid`): Provides `AKKIO_ID` via `afs_individual_id`, plus card attributes
    - **MERCHANT** (LEFT JOIN on `mtid`): Provides merchant description, MCC, and location
    - **BRAND_TAGGING** (LEFT JOIN on `mtid`): Provides store_id, brand_id, channel, and locationid
    - **BRAND_TAXONOMY** (LEFT JOIN on `store_id + brand_id`): Provides store/brand names and classifications
    - **BRAND_LOCATION** (LEFT JOIN on `locationid`): Provides store address and location details
    - **Materialization**: Incremental table (clustered by trans_date, AKKIO_ID)
    - **Note**: Use `FACT_TRANSACTION_SUMMARY` for most queries unless transaction-level detail is required
- **FACT_TRANSACTION_SUMMARY**: Daily Transaction Summary Table - Aggregated transaction activity per day and individual (`trans_date`, `AKKIO_ID`). Optimized for RAG engine queries that need summary-level data. Contains transaction metrics (count, totals, averages), aggregated merchant/brand attributes as comma-separated lists, and unique counts. Source: `FACT_TRANSACTION_ENRICHED`.
    - **Grain**: One row per day per individual (trans_date, AKKIO_ID)
    - **Materialization**: Table (clustered by trans_date, AKKIO_ID)
    - **Use Case**: Preferred table for most analytics queries; use `FACT_TRANSACTION_ENRICHED` only when transaction-level detail is needed
- **V_AGG_AKKIO_IND**: Individual Aggregation Table - One row per individual (`AKKIO_ID`) per `PARTITION_DATE` with aggregated demographic attributes optimized for analytics. Generated from `V_AKKIO_ATTRIBUTES_LATEST`. Includes contact identifier placeholders (MAIDS, IPS, EMAILS, PHONES), interests as OBJECTs (GENERAL_INTERESTS, SPORTS_INTERESTS, READING_INTERESTS, TRAVEL_INTERESTS), credit card info as OBJECT, and financial attributes. Age buckets are encoded as integers (1-7 for AGE_BUCKET, 1-12 for AGE_BUCKET_DETAILED).
- **V_AGG_AKKIO_HH**: Household Aggregation Table - One row per household (`AKKIO_HH_ID`) per `PARTITION_DATE` with household-level attributes. Generated from `V_AKKIO_ATTRIBUTES_LATEST`. Includes household income, child age groups as OBJECT, homeowner status as integer (0/1), and household composition metrics. Currently 1:1 with individuals but structured for future scenarios where multiple individuals may share a household.

Design Principles

- Both `V_AKKIO_ATTRIBUTES_LATEST` and transaction tables use `AKKIO_ID` as the bridge for flexible querying
- Transactions are kept separate from individual attributes for optimal LLM query performance
- `FACT_TRANSACTION_SUMMARY` provides aggregated daily summaries optimized for RAG queries; use `FACT_TRANSACTION_ENRICHED` only when transaction-level detail is required
- All demographic fields are normalized (e.g., GENDER: MALE/FEMALE/UNKNOWN, ETHNICITY: HISPANIC/AFRICAN_AMERICAN/etc., POLITICS: DEMOCRAT_KNOWN/REPUBLICAN_INFERRED/etc.)
- All joins in `FACT_TRANSACTION_ENRICHED` are LEFT JOINs to preserve all transactions even if enrichment data is missing
- `FACT_TRANSACTION_ENRICHED` uses incremental materialization for efficient processing of new transactions