

# DATASET UNDERSTANDING

## 1. Transaction Table

### Purpose:

Captures **every payment transaction** made by clients using their cards. This is the **fact table** for analysis.

### Key Points:

- Grain: **One row = one transaction**
- Timezone: Stored as UTC (**TIMESTAMP**), convertible to **Asia/Jakarta**
- **amount** is already numeric (**FLOAT**)
- Contains merchant info, MCC (Merchant Category Code), and transaction method (Chip vs Swipe)

### Schema:

Field name	Type	Description
<b>id</b>	INTEGER	Unique transaction ID.
<b>date</b>	TIMESTAMP	Date & time of transaction (UTC).
<b>client_id</b>	INTEGER	ID of the user making the transaction (joins to <b>Users.id</b> ).
<b>card_id</b>	INTEGER	ID of the card used (joins to <b>Cards.id</b> ).
<b>amount</b>	FLOAT	Transaction value; negative values = refunds.
<b>use_chip</b>	STRING	Method: "Chip Transaction" or "Swipe Transaction".
<b>merchant_id</b>	INTEGER	Unique merchant identifier.
<b>merchant_city</b>	STRING	City where merchant is located.
<b>merchant_state</b>	STRING	State where merchant is located.
<b>zip</b>	FLOAT	Postal code of merchant location.

<code>mcc</code>	INTEGER	Merchant Category Code — defines merchant type/category.
<code>errors</code>	STRING	Error message/code if transaction failed (mostly null).

#### Business Relevance:

- Main source for **spending patterns**, **merchant preferences**, and **channel usage**.
- `mcc` allows category-level aggregation (e.g., restaurants, retail, travel).
- `use_chip` can indicate adoption of secure payment methods.
- `amount` patterns are crucial for **revenue estimation** and **fraud detection**.

## 2. Cards Table

#### Purpose:

Holds details about each card, including brand, type, credit limit, and security features. This is a **dimension table** that enriches transactions with product-level attributes.

#### Key Points:

- Grain: **One row = one card**
- Links to both the transaction (`card_id`) and the card owner (`client_id`)
- Contains boolean fields for security risk checks (`has_chip`, `card_on_dark_web`)

#### Schema:

Field name	Type	Description
<code>id</code>	INTEGER	Unique card ID (joins to <code>Transaction.card_id</code> ).
<code>client_id</code>	INTEGER	Owner's user ID (joins to <code>Users.id</code> ).
<code>card_brand</code>	STRING	Brand (e.g., Visa, MasterCard, Amex).
<code>card_type</code>	STRING	Type (Credit, Debit, Prepaid).

card_number	INTEGER	Full card number (sensitive, not used directly in analysis).
expires	STRING	Expiry date (MM/YYYY).
cvv	INTEGER	CVV code (sensitive, not used directly in analysis).
has_chip	BOOLEAN	Whether card has an EMV chip.
num_cards_issued	INTEGER	Number of cards issued for this account.
credit_limit	INTEGER	Credit limit in currency units.
acct_open_date	STRING	Account opening date (MM/YYYY).
year_pin_last_changed	INTEGER	Year PIN was last updated.
card_on_dark_web	BOOLEAN	Whether card appears in dark web leaks.

#### Business Relevance:

- card\_type and card\_brand help segment transaction patterns by product.
- has\_chip + use\_chip combination can reveal risky behavior (chip-capable but swiped).
- card\_on\_dark\_web can be used for **fraud/risk flags**.
- credit\_limit can be compared to spending patterns for credit utilization analysis.

### 3. Users Table

#### Purpose:

Stores demographic, geographic, and financial attributes for each client. This is a **dimension table** that enriches transactions with user-level context.

#### Key Points:

- Grain: **One row = one user**
- Contains **age, income, credit score**, and **location**.

- Includes derived segments like **retirement\_age** and financial indicators (**total\_debt**, **num\_credit\_cards**).

#### Schema:

Field name	Type	Description
<b>id</b>	INTEGER	Unique user ID (joins to <b>Transaction.client_id</b> ).
<b>current_age</b>	INTEGER	User's current age.
<b>retirement_age</b>	INTEGER	Expected retirement age.
<b>birth_year</b>	INTEGER	Birth year.
<b>birth_month</b>	INTEGER	Birth month.
<b>gender</b>	STRING	Gender of the user.
<b>address</b>	STRING	Street address.
<b>latitude</b>	FLOAT	Geographic latitude.
<b>longitude</b>	FLOAT	Geographic longitude.
<b>per_capita_income</b>	INTEGER	Area's per capita income.
<b>yearly_income</b>	INTEGER	User's yearly income.
<b>total_debt</b>	INTEGER	Total debt value.
<b>credit_score</b>	INTEGER	Credit score (higher = better).
<b>num_credit_cards</b>	INTEGER	Number of credit cards owned.

#### Business Relevance:

- Demographics (**age**, **gender**) enable segmentation analysis.
- yearly\_income** and **credit\_score** can be linked to spending or risk profiles.
- Location fields can be mapped to merchant states for geo-analysis.
- Debt and credit card count may indicate financial health and potential risk.

