DATA DICTIONARY

Data Dictionary

Column-Level Reference for Both Projects

Reporting Period	Current as of: Q1 2026
Document Version	1.0
Date Prepared	April 2026
Classification	Internal - Reference Document
Prepared By	Data & Analytics Team

DISTRIBUTION

Data Engineering - Analytics - Data Quality - New-Hire Onboarding

PREPARED BY	REVIEWED BY	APPROVED BY
_____ *Signature / Date*	_____ *Signature / Date*	_____ *Signature / Date*

1. Purpose and Conventions

This document is the column-level reference for both projects in the bundle. It documents every column in every table - source mapping, type, nullability, business meaning, valid values, and example data. It is the document an analyst opens to answer the question "what does this column actually mean?"

Convention	Detail
Case	snake_case for everything. No camelCase. No mixed case.
Surrogate keys	End in _sk (customer_sk, branch_sk, product_sk, channel_sk, txn_type_sk, date_sk).
Natural keys	Preserved from source with original name (customer_id, branch_id, product_id, order_id).
Booleans	Use is_ prefix (is_current, is_weekend, is_holiday). True/false values, never 1/0.
Currency	Bank warehouse: end in _ngn (amount_ngn, balance_after_ngn). Distributor: numeric without suffix (consistent with USD throughout).
Timestamps	End in _at for system timestamps (loaded_at). End in _timestamp for business timestamps (txn_timestamp).
Dates	End in _date for date-only fields (order_date, shipped_date, effective_from).
Counts	End in _count (transaction_count, total_orders, active_customer_count).

Table 1. Project-wide naming conventions.[1]

2. Project A - Distributor Source Tables

Project A reads directly from CSV source files without a warehouse intermediate. Each table below documents the source-file schema as consumed by the Tableau dashboard.

2.1 orders

Column	Type	Nullable	Description
order_id	INT	No	Primary key. Sequential order number assigned at order creation.
customer_id	VARCHAR(5)	No	Foreign key to customers. 5-character uppercase code (e.g., 'ALFKI').
employee_id	INT	No	Foreign key to employees. The sales rep who took the order.
order_date	DATE	No	Date the order was placed. Used for trend analysis and date filters.
required_date	DATE	Yes	Customer-requested delivery date. Not used in current dashboard.
shipped_date	DATE	Yes	Date the order left the warehouse. NULL for orders not yet shipped. Used for fulfillment-days calculation.
shipper_id	INT	No	Foreign key to shippers.
freight	DECIMAL(10,2)	No	Freight cost in USD. Used in Freight % of Revenue calculation.
ship_country	VARCHAR(50)	No	Destination country. Used in regional analysis.

Table 2. orders source schema.

2.2 order_details

Column	Type	Nullable	Description
order_id	INT	No	Foreign key to orders. Part of composite key.
product_id	INT	No	Foreign key to products. Part of composite key.
unit_price	DECIMAL(10,2)	No	Price per unit at time of order. May differ from current product price.
quantity	INT	No	Units ordered. Always positive.
discount	DECIMAL(5,4)	No	Decimal discount (0.10 means 10%). Range 0.00 to 0.25 observed.

Table 3. order_details source schema. Grain: one row per product per order.

> **WHY DISCOUNT IS DECIMAL, NOT PERCENTAGE**
> The discount is stored as a decimal (0.10) rather than a percentage (10). The revenue formula multiplies by (1 - discount) to compute the price actually paid. Storing as a percentage would require dividing by 100 in the formula, an extra computation per row.

2.3 products

Column	Type	Nullable	Description
product_id	INT	No	Primary key.
product_name	VARCHAR(80)	No	Display name. Used in product rankings.
category_id	INT	No	Foreign key to categories.
unit_price	DECIMAL(10,2)	No	List price. May differ from sale price in order_details.unit_price.
units_in_stock	INT	Yes	Current inventory. Not used in dashboard.
discontinued	BOOLEAN	No	True if the product is no longer sold. 8 products have this set to true.

Table 4. products source schema.

2.4 categories

Column	Type	Nullable	Description
category_id	INT	No	Primary key.
category_name	VARCHAR(50)	No	8 distinct values: Beverages, Condiments, Confections, Dairy Products, Grains & Cereals, Meat & Poultry, Produce, Seafood.
description	TEXT	Yes	Category description. Not used in dashboard.

Table 5. categories source schema.

2.5 customers

Column	Type	Nullable	Description
customer_id	VARCHAR(5)	No	Primary key. 5-character uppercase code.
company_name	VARCHAR(100)	No	Display name of the customer organization.
country	VARCHAR(50)	No	21 distinct values in dataset.
city	VARCHAR(50)	Yes	City of the customer's primary address.
contact_name	VARCHAR(50)	Yes	Individual contact at the customer org. Genericized in portfolio.
contact_title	VARCHAR(50)	Yes	Contact role. Not used in dashboard.

Table 6. customers source schema.

2.6 employees

Column	Type	Nullable	Description
employee_id	INT	No	Primary key.
first_name, last_name	VARCHAR(50)	No	Genericized for portfolio submission (Rep A, Rep B, etc.).
hire_date	DATE	No	Used for tenure analysis (out of scope).
title	VARCHAR(50)	Yes	Job title - Sales Representative, Sales Manager, etc.

Table 7. employees source schema.

2.7 shippers

Column	Type	Nullable	Description
shipper_id	INT	No	Primary key.
company_name	VARCHAR(50)	No	Genericized for portfolio (Carrier A, Carrier B, Carrier C).
phone	VARCHAR(20)	Yes	Contact phone. Not used in dashboard.

Table 8. shippers source schema.

3. Project B - Bank Warehouse Tables

Project B has a full warehouse schema. The tables below document every column in every warehouse table.

3.1 stg_transactions (staging)

Column	Type	Nullable	Description
txn_id	VARCHAR(30)	Yes	Transaction ID from source. NULL allowed in staging to catch missing IDs in data quality checks.
txn_datetime	TIMESTAMP	Yes	Transaction timestamp.
customer_id	VARCHAR(30)	Yes	Source customer ID.
customer_name	VARCHAR(150)	Yes	Source customer name.
tier	VARCHAR(30)	Yes	Customer tier at transaction time.
branch_id	VARCHAR(30)	Yes	Source branch ID.
branch_name	VARCHAR(150)	Yes	Source branch name.
state	VARCHAR(80)	Yes	Branch state.
product_id	VARCHAR(30)	Yes	Product ID.
product_name	VARCHAR(150)	Yes	Product display name.
product_type	VARCHAR(80)	Yes	Product category.
txn_type	VARCHAR(50)	Yes	Transaction type (Deposit, Withdrawal, Transfer, etc.).
channel	VARCHAR(50)	Yes	Channel (Branch, ATM, Mobile, POS, Internet Banking).
amount_ngn	NUMERIC(18,2)	Yes	Transaction amount in Nigerian Naira.
balance_after_ngn	NUMERIC(18,2)	Yes	Account balance after the transaction.
load_batch_id	VARCHAR(50)	Yes	ETL batch identifier for the load run.
loaded_at	TIMESTAMP	No	System timestamp when row was inserted into staging.

Table 9. stg_transactions schema. All business columns nullable to allow data quality validation post-load.

3.2 dim_date

Column	Type	Nullable	Description
date_sk	INT	No	Surrogate key. YYYYMMDD format (e.g., 20240315 for March 15, 2024).
full_date	DATE	No	Calendar date. UNIQUE constraint.
day_of_month	INT	No	1-31.
day_of_week	INT	No	1-7 (1=Monday, 7=Sunday).
day_name	VARCHAR(10)	No	Monday, Tuesday, etc.
week_of_year	INT	No	1-53. ISO week numbering.
month_num	INT	No	1-12.
month_name	VARCHAR(15)	No	January, February, etc.
quarter_num	INT	No	1-4.
quarter_name	VARCHAR(2)	No	Q1, Q2, Q3, Q4.
year_num	INT	No	Four-digit year.
is_weekend	BOOLEAN	No	True for Saturday and Sunday.
is_holiday	BOOLEAN	No	True for recognized public holidays. Calendar maintained separately.

Table 10. dim_date schema. Pre-generated; loaded once and extended as the date range grows.

3.3 dim_customer

Column	Type	Nullable	Description
customer_sk	BIGINT	No	Surrogate primary key. IDENTITY-generated.
customer_id	VARCHAR(30)	No	Natural key from source. NOT unique here - multiple versions per customer under SCD Type 2.
customer_name	VARCHAR(150)	No	Display name. SCD Type 1 (overwrite).
tier	VARCHAR(30)	No	Customer tier. SCD Type 2 (versioned). Valid values: Silver, Gold, Platinum.
effective_from	DATE	No	First date this version is active.
effective_to	DATE	No	Last date this version is active. '9999-12-31' for current.

Column	Type	Nullable	Description
is_current	BOOLEAN	No	True for the active version of each customer.
row_hash	TEXT	Yes	MD5 hash of versioned attributes. Used for change detection during incremental load.

Table 11. dim_customer schema.

3.4 dim_branch

Column	Type	Nullable	Description
branch_sk	BIGINT	No	Surrogate primary key.
branch_id	VARCHAR(30)	No	Natural key from source. Multiple versions allowed.
branch_name	VARCHAR(150)	No	Branch display name. SCD Type 2.
state	VARCHAR(80)	No	State the branch is located in. SCD Type 1.
effective_from	DATE	No	First date this version is active.
effective_to	DATE	No	Last date this version is active.
is_current	BOOLEAN	No	True for the active branch version.
row_hash	TEXT	Yes	Change-detection hash.

Table 12. dim_branch schema.

3.5 dim_product, dim_channel, dim_transaction_type

Simple Type 1 dimensions. No SCD versioning.

Column	Type	Nullable	Description
product_sk / channel_sk / txn_type_sk	BIGINT	No	Surrogate primary key in each respective table.
product_id (dim_product only)	VARCHAR(30)	No	Natural key. UNIQUE in dim_product.
product_name (dim_product)	VARCHAR(150)	No	Product display name.
product_type (dim_product)	VARCHAR(80)	No	Indexed for product-type filtering.

Column	Type	Nullable	Description
channel_name (dim_channel)	VARCHAR(50)	No	Valid values: Branch, ATM, POS, Mobile App, Internet Banking. UNIQUE.
txn_type_name (dim_transaction_type)	VARCHAR(50)	No	Valid values: Deposit, Withdrawal, Transfer, Fee, Bill Payment, etc. UNIQUE.

Table 13. Type 1 dimension schemas (compact view).

3.6 fact_transactions

Column	Type	Nullable	Description
txn_sk	BIGINT	No	Surrogate primary key.
txn_id	VARCHAR(30)	No	Degenerate dimension - transaction ID from source. UNIQUE.
txn_timestamp	TIMESTAMP	No	Business timestamp.
date_sk	INT	No	FK to dim_date.
customer_sk	BIGINT	No	FK to dim_customer. Resolved to the version active at txn_timestamp.
branch_sk	BIGINT	No	FK to dim_branch. Resolved to the version active at txn_timestamp.
product_sk	BIGINT	No	FK to dim_product.
channel_sk	BIGINT	No	FK to dim_channel.
txn_type_sk	BIGINT	No	FK to dim_transaction_type.
amount_ngn	NUMERIC(18,2)	No	Transaction amount. Fully additive.
balance_after_ngn	NUMERIC(18,2)	Yes	Account balance after transaction. Semi-additive (do not sum across time).
fee_amount_ngn	NUMERIC(18,2)	Yes	Fee charged. NULL until fee rules supplied; column reserved.
deposit_amount_ngn	NUMERIC(18,2)	No	Deposit amount (0 for non-deposit transactions). Default 0. Denormalized for query speed.
withdrawal_amount_ngn	NUMERIC(18,2)	No	Withdrawal amount (0 for non-withdrawal). Default 0.
transaction_count	INT	No	Constant value 1 (CHECK enforced). Sums to give transaction count without COUNT(*).

Table 14. fact_transactions schema. Grain: one row per banking transaction.

3.7 agg_monthly_branch_activity

Column	Type	Nullable	Description
month_key	INT	No	YYYYMM (e.g., 202403 for March 2024). Part of composite primary key.
branch_sk	BIGINT	No	FK to dim_branch. Part of composite PK.

Column	Type	Nullable	Description
product_type	VARCHAR(80)	No	Product type from dim_product, denormalized for grouping. Part of composite PK.
total_transaction_value	NUMERIC(18,2)	No	SUM(amount_ngn) for this month/branch/product_type.
deposit_value_ngn	NUMERIC(18,2)	No	SUM(deposit_amount_ngn) for this slice.
withdrawal_value_ngn	NUMERIC(18,2)	No	SUM(withdrawal_amount_ngn) for this slice.
fee_income_ngn	NUMERIC(18,2)	Yes	SUM(fee_amount_ngn). NULL until fee data populated.
transaction_count	INT	No	COUNT(*) for the slice.
active_customer_count	INT	No	COUNT(DISTINCT customer_sk) for the slice.

Table 15. agg_monthly_branch_activity schema. Refreshed nightly after fact load.

4. Calculated Fields and Derived Metrics

This section documents calculated fields not directly stored in any table - either calculated by Tableau (Project A) or computed at query time (Project B).

4.1 Project A - Tableau calculated fields

Calculated field	Definition
Revenue	[unit_price] * [quantity] * (1 - [discount]). Computed at order_details grain.
Average Order Value (AOV)	SUM([Revenue]) / COUNTD([order_id]). Distinct count of orders, not lines.
Fulfillment Days	DATEDIFF('day', [order_date], [shipped_date]). NULL if shipped_date is NULL.
Freight % of Revenue	SUM([freight]) / SUM([Revenue]) * 100. Use SUM on both for correct aggregation.
Discount Band	Bucketing on [discount]: 0%, 1-10%, 11-20%, >20%. CASE expression.
Product Status	IF [discontinued] = 1 THEN 'Discontinued' ELSE 'Active' END.

Table 16. Dashboard calculated fields with definitions.

4.2 Project B - Common warehouse derivations

Derivation	How computed
Transaction count	SUM(transaction_count) - the column is constant 1 per row. Equivalent to COUNT(*) but faster on aggregates.
Net flow	SUM(deposit_amount_ngn) - SUM(withdrawal_amount_ngn). Both columns are denormalized on the fact.
Active customer count	COUNT(DISTINCT customer_sk). NOT customer_id - the surrogate key correctly handles SCD Type 2.
Customer at transaction time	Join fact_transactions to dim_customer on customer_sk. The SCD Type 2 resolution is implicit in the join.
Customer (current state)	Join fact_transactions to dim_customer on customer_id, with filter is_current = TRUE.
Latest balance per customer	MAX(balance_after_ngn) within a customer/time window. Do not SUM.
Quarter-over-quarter growth	Self-join dim_date on quarter offset. (this_quarter_value - prev_quarter_value) / prev_quarter_value.

Table 17. Common derived metrics. Each documents the correct expression to avoid double-counting or incorrect aggregation.

5. Reference Values and Valid Domains

5.1 Project A - Distributor

Field	Valid values
categories.category_name	Beverages, Condiments, Confections, Dairy Products, Grains & Cereals, Meat & Poultry, Produce, Seafood.
customers.country	21 distinct values: Germany, USA, France, Brazil, UK, Austria, Canada, Sweden, Mexico, Spain, Venezuela, Italy, Belgium, Switzerland, Argentina, Denmark, Finland, Ireland, Norway, Poland, Portugal.
order_details.discount	0.00, 0.05, 0.10, 0.15, 0.20, 0.25. Six discrete values observed.
products.discontinued	TRUE for 8 products, FALSE for 69 products.
shippers.company_name (genericized)	Carrier A, Carrier B, Carrier C. (Three shippers.)

Table 18. Distributor reference values.

5.2 Project B - Bank

Field	Valid values
dim_customer.tier	Silver, Gold, Platinum. Three tiers.
dim_channel.channel_name	Branch, ATM, POS, Mobile App, Internet Banking.
dim_transaction_type.txn_type_name	Deposit, Withdrawal, Transfer, Bill Payment, Fee, Reversal, Adjustment.
dim_branch.state	Nigerian states. The reference list is maintained centrally; new states only added when a new branch opens.
dim_date.is_weekend	TRUE for Saturday and Sunday rows.
dim_date.is_holiday	TRUE for recognized public holidays. Maintained from a public-holidays calendar.
fact_transactions.amount_ngn	Positive numeric. Zero allowed only for reversal transactions (rare).
fact_transactions.transaction_count	Always 1 (CHECK constraint enforced).

Table 19. Bank warehouse reference values.

6. Cross-Reference - How to Find What

Quick reference for analysts looking for specific information.

Looking for	Table - Column
Project A - revenue	Computed: order_details.unit_price * .quantity * (1 - .discount)
Project A - total revenue by country	Join: order_details -> orders -> customers; SUM Revenue group by customers.country
Project A - shipper performance	orders.shipped_date - orders.order_date, group by orders.shipper_id
Project B - transaction value	fact_transactions.amount_ngn
Project B - customer at transaction time	fact_transactions.customer_sk -> dim_customer (SCD Type 2 resolved)
Project B - current customer state	dim_customer where is_current = TRUE
Project B - monthly branch reports	agg_monthly_branch_activity (or fact_transactions with date filter)
Project B - balance over time	fact_transactions.balance_after_ngn (use MAX not SUM for time aggregation)
Project B - deposit vs withdrawal	fact_transactions.deposit_amount_ngn and .withdrawal_amount_ngn (denormalized)
Both projects - row counts	See Tables 2-9 (Project A) and Tables 9-15 (Project B)

Table 20. Cross-reference index.

Document Control

Version	Date	Change Summary	Author
0.1	2026-02-18	Project A tables documented	Data Engineering
0.5	2026-03-01	Project B tables documented; conventions section added	Data Engineering
0.9	2026-03-12	Calculated fields and cross-reference index added	Data Engineering
1.0	2026-03-15	Approved as reference document	Data Engineering

Table 21. Revision history.