



# FraudLens - dbt Documentation

**Project:** fraudlens **Version:** 1.0.1 **Database:** Snowflake (FRAUDLENS\_DB)

## Overview

This document describes the dbt data models for the FraudLens Healthcare Fraud Detection Pipeline. The models follow a medallion architecture with three layers: Bronze (raw), Silver (cleansed), and Gold (business-ready).

## Data Sources

### Bronze Layer Sources

Source	Database	Schema	Description
bronze	FRAUDLENS_DB	BRONZE	Raw CMS/HHS public datasets
nppes	AFFINE_NPPES_PROVIDER_DATA	REF_DW	Snowflake Marketplace NPPES dataset

### Bronze Tables

Table	Description	Key Columns
LEIE	Excluded Individuals/Entities	NPI, EXCLTYPE, EXCLDATE
MEDICARE_HOSPITAL_SPENDING	Hospital spending per beneficiary	FACILITY_ID, CLAIM_TYPE
PROVIDER_INFORMATION	Nursing home providers	FEDERAL_PROVIDER_NUMBER
LONGTERM_CARE_HOSPITAL	LTCH quality measures	CMS_CERTIFICATION_NUMBER
HOSPICE	Hospice provider info	CMS_CERTIFICATION_NUMBER
HOME_HEALTH_CARE	Home health agencies	CMS_CERTIFICATION_NUMBER
MEDICARE_PART_D_PRESCRIBERS	Prescription drug claims	PRSCRBR_NPI, TOT_DRUG_CST
OPEN_PAYMENTS_GENERAL	Pharma payments to physicians	Covered_Recipient_NPI
OPEN_PAYMENTS_RESEARCH	Research payments	Covered_Recipient_NPI
OPEN_PAYMENTS_OWNERSHIP	Ownership interests	Physician_NPI

### NPPES Tables (Snowflake Marketplace)

Table	Description	Key Columns
DIM_PROVIDER	Provider master data	NPI, ENTITY_TYPE_CODE, IS_ACTIVE
DIM_PROVIDER_ADDRESS	Provider addresses	NPI, ADDRESS_TYPE_CODE
DIM_PROVIDER_TAXONOMY	Provider specialties	NPI, TAXONOMY_CODE
REF_TAXONOMY_CODE	Taxonomy reference	TAXONOMY_CODE

---

## Staging Layer

---

All staging models are materialized as **views** in the STAGING schema.

Model	Description
stg_leie	LEIE exclusions with cleaned dates
stg_medicare_hospital_spending	Hospital spending with standardized columns
stg_provider_information	Nursing home providers
stg_longterm_care_hospital	LTCH facilities
stg_hospice	Hospice providers
stg_home_health_care	Home health agencies
stg_medicare_part_d_prescribers	Prescription claims
stg_open_payments_general	General pharma payments
stg_open_payments_research	Research payments
stg_open_payments_ownership	Ownership interests
stg_nppes_provider	NPPES provider master (active only)
stg_nppes_provider_address	Provider addresses
stg_nppes_provider_taxonomy	Provider specialties
stg_nppes_taxonomy_code	Taxonomy reference (active only)

---

## Silver Layer Models

---

All Silver models are materialized as **tables** in the SILVER schema.

| **provider**

Provider master table enriched with NPPES data and exclusion status

Column	Description	Tests
NPI	National Provider Identifier - Primary key	unique, not_null
ENTITY_TYPE	Individual or Organization	accepted_values
IS_EXCLUDED	True if provider is on LEIE exclusion list	not_null

## excluded\_providers

LEIE excluded providers enriched with NPPES data

Column	Description	Tests
EXCLUSION_ID	Unique identifier for the exclusion record	unique, not_null
EXCLUSION_DATE	Date of exclusion from Medicare/Medicaid	not_null
IS_CURRENTLY_EXCLUDED	True if still excluded (not reinstated)	not_null
IS_HIGH_RISK	True if excluded but still active in NPPES	not_null

## payments

Consolidated Open Payments (General, Research, Ownership)

Column	Description	Tests
PAYMENT_ID	Unique payment identifier	unique, not_null
PAYMENT_CATEGORY	Type: GENERAL, RESEARCH, or OWNERSHIP	not_null, accepted_values
PAYMENT_RISK_TIER	Risk tier based on payment amount	accepted_values
IS_RECIPIENT_EXCLUDED	True if recipient is on LEIE	not_null

## prescriptions

Medicare Part D prescriptions with fraud indicators

Column	Description	Tests
PRESCRIPTION_ID	Unique prescription record identifier	unique, not_null
NPI	Prescriber National Provider Identifier	not_null
RISK_TIER	Fraud risk tier	accepted_values
IS_PRESCRIBER_EXCLUDED	True if prescriber is on LEIE	not_null

## facilities

Unified healthcare facilities (Hospice, Home Health, LTCH, Nursing Homes)

Column	Description	Tests
FACILITY_ID	Unique facility identifier	unique, not_null
CCN	CMS Certification Number	not_null
FACILITY_TYPE	HOSPICE, HOME_HEALTH, LTCH, NURSING_HOME	not_null, accepted_values
OWNERSHIP_CATEGORY	FOR_PROFIT, NON_PROFIT, GOVERNMENT, OTHER	accepted_values

## hospital\_spending

Medicare hospital spending with benchmarks

Column	Description	Tests
SPENDING_ID	Unique spending record identifier	unique, not_null
CCN	CMS Certification Number	not_null
SPENDING_RISK_TIER	Spending risk classification	accepted_values
IS_HIGH_SPENDER	True if spending 50%+ above national avg	not_null

## Gold Layer Models

All Gold models are materialized as **tables** in the GOLD schema.

## provider\_360

Complete provider profile combining all data sources. Single source of truth for provider analysis.

**Joins:** provider (base) + payments\_summary + prescriptions\_summary + excluded\_providers

**Key Columns:** NPI, ENTITY\_TYPE, FULL\_NAME, SPECIALTY, IS\_EXCLUDED, TOTAL\_PAYMENT\_AMOUNT, TOTAL\_PRESCRIPTION\_COST, TOTAL\_FINANCIAL\_EXPOSURE

## fraud\_risk\_score

Composite fraud risk scoring for each provider. Score range: 0-100.

Risk Category	Max Points	Factors
Exclusion Risks	40	IS_EXCLUDED (+25), IS_EXCLUDED_HIGH_RISK (+15)
Payment Risks	20	MEGA/MAJOR_RECIPIENT (+5-10), PCT_HIGH_RISK_PAYMENTS
Prescription Risks	20	PCT_BRAND_CLAIMS, TOTAL_HIGH_RISK_DRUGS
Activity Anomalies	20	Both payments+prescriptions (+10), Inactive NPI (+10)

**Risk Tiers:** CRITICAL ( $\geq 70$ ) | HIGH ( $\geq 50$ ) | MEDIUM ( $\geq 30$ ) | LOW ( $\geq 10$ ) | MINIMAL ( $< 10$ )

## high\_risk\_alerts

Actionable alerts for fraud investigation. Each row = one alert requiring review.

Alert Type	Priority	Description
EXCLUDED_STILL_ACTIVE	1	Provider excluded but still active in NPPES
PRESCRIPTION_BY_EXCLUDED	2	Excluded provider prescribed medications
PAYMENT_TO_EXCLUDED	3	Excluded provider received pharma payments
HIGH_RISK_SCORE	4	Provider has critical fraud risk score
HIGH_BRAND_PRESCRIBER	5	High brand prescription rate (>80%)

## payments\_summary / prescriptions\_summary

Aggregated metrics per provider for payments and prescriptions respectively.

## Tests Summary

Model	Tests
provider	3
excluded_providers	4
payments	6
prescriptions	5
facilities	5
hospital_spending	4
<b>Total</b>	<b>31 tests</b>

# Configuration

## Materialization Strategy

Layer	Materialization	Schema
Staging	view	STAGING
Silver	table	SILVER
Gold	table	GOLD

## Running dbt

```
dbt deps          # Install dependencies
dbt run           # Run all models
dbt test          # Run tests
dbt docs generate && dbt docs serve # Documentation
dbt run --select staging.*      # Run specific layer
dbt run --select silver.* 
dbt run --select gold.*
```