Q.Dose Dataset - Data Dictionary

Dataset Overview

- **Dataset Name:** Q.Dose Pharmaceutical Analytics Data

- **Purpose:** Detailed pharmaceutical prescription and claims analytics for specialized medications

- **Data Period:** 2017-2019 monthly prescription data

- **Record Type:** Individual prescription/claim records with detailed patient demographics and financial breakdowns

- **Total Records:** 200 rows (sample dataset)

- **Total Columns:** 51 columns

Column Definitions

Patient & Entity Identification

Column Name	Data Type	Description	Sample Values	Business Rules
ENTITY_NO	String	Encrypted patient/member identifier	"csdWMUuVPDFA6P/zQVEsl7DAb9y6hSX0A4nSHu8GgMA="	Base64 encoded hash for privacy protection

Temporal Dimensions

Column Name	Data Type	Description	Sample Values	Business Rules
DATE_KEY	String	Transaction date	"2017-03-01", "2019-09-01"	Format: YYYY-MM-DD
YEAR	Integer	Calendar year	2017, 2018, 2019	Range: 2017-2019
MONTH_KEY	Integer	Year-month combination	201703, 201909, 201812	Format: YYYYMM

Product Identification

Column Name	Data Type	Description	Sample Values	Business Rules
NAPPI9	Integer	9-digit National Pharmaceutical Product Interface code	720446001, 708286001	Unique pharmaceutical product identifier
PRODUCT_NAME	String	Brand/commercial product name	"Tuvigin", "Copaxone pre-filled syringe 1ml"	Official registered product name
NAPPI_MANUFACTURER	String	Manufacturing company	"Sandoz SA (PTY) LTD", "Pharmachemie"	Pharmaceutical manufacturer

Patient Demographics

Column Name	Data Type	Description	Sample Values	Business Rules
AGE_BUCKET	String	Broad age category	"Above 18 Yrs"	Currently only adult patients
AGE_GROUPS	String	Detailed age ranges	"Between 35-50", "Between 50-65", "Greater than 65"	5 distinct age brackets
GENDER	String	Patient gender	"F", "M"	Binary gender classification

Geographic Dimensions

Column Name	Data Type	Description	Sample Values	Business Rules
PROVINCE	String	Patient's province of residence	"GAUTENG", "WESTERN CAPE", "FREE STATE"	South African provinces (uppercase)
REGION_OF_RESIDENCE	String	Patient's specific residential area	"PRETORIA", "JOHANNESBURG", "RANDBURG"	City/town level granularity

Medical Scheme Information

Column Name	Data Type	Description	Sample Values	Business Rules
PLAN_GRP	String	Medical plan category	"Comprehensive", "Executive", "Saver"	11 unique plan types
PLAN_SCHEME	String	Medical scheme name	"Discovery Health Medical Scheme", "Anglo Medical Scheme"	9 different medical schemes
DEG_DESCR	String	Designated Service Provider (DSP) condition	"NEU040 - Multiple sclerosis"	Clinical condition code and description (sample data only)

Clinical Information

Column Name	Data Type	Description	Sample Values	Business Rules
IN_OUT_HOSPITAL_IND	Integer	Hospital vs outpatient indicator	0, 1	0=Outpatient, 1=Hospital
TREATING_DR	String	Prescribing physician specialty	"Neurologist", "General Practitioner"	Medical specialist category

Product Specifications

Column Name	Data Type	Description	Sample Values	Business Rules
STRENGTH	String	Medication strength/concentration	"0.5 - MG", "20 - MG/1ML", "30 - MCG/0\,5ML"	Dosage strength with units
SCHEDULE	Integer	Medicine scheduling category	4	Currently only Schedule 4 medications
PACK_SIZE	Integer	Units per package	28, 4, 15, 12, 1	Range: 1-28 units
DOSAGE_FORM	String	Medication form	"CAP", "INJ", "TAB", "INF"	Capsule, Injection, Tablet, Infusion

ATC Classification System (5-Level Hierarchy)

Column Name	Data Type	Description	Sample Values	Business Rules
ATC_DESCRIPTION	String	Active ingredient name	"FINGOLIMOD", "GLATIRAMER ACETATE"	WHO ATC active substance
ATC_LEVEL_DESC_1	String	ATC Level 1 - Anatomical main group	"L - Antineoplastic And Immunomodulating Agents"	Broadest therapeutic category
ATC_LEVEL_DESC_2	String	ATC Level 2 - Therapeutic subgroup	"L04 - Immunosuppressants", "L03 - Immunostimulants"	Therapeutic classification
ATC_LEVEL_DESC_3	String	ATC Level 3 - Pharmacological subgroup	"L04A - Immunosuppressants", "L03A - Immunostimulants"	Pharmacological classification
ATC_LEVEL_DESC_4	String	ATC Level 4 - Chemical subgroup	"L04AA - Selective Immunosuppressants"	Chemical classification
ATC_LEVEL_DESC_5	String	ATC Level 5 - Chemical substance	"L04AA27 - Fingolimod", "L03AX13 - Glatiramer Acetate"	Specific chemical substance

Provider Information

Column Name	Data Type	Description	Sample Values	Business Rules
PROVIDER_TYPE	String	Type of healthcare provider	"Pharmacy", "Hospitals"	Dispensing facility type
PROVIDER_GROUP	String	Provider network/group	"Independent Pharmacy", "Dischem", "Clicks"	7 unique provider groups
PROVIDER	String	Specific provider name	"Script Wise Pharmacy", "Southern Rx Pharmacy"	Individual facility name
PROVIDER_REGION	String	Provider's geographic region	"CARLETONVILLE", "EDENVALE", "PRETORIA"	Provider location
PROVIDER_PROVINCE	String	Provider's province	"GAUTENG", "WESTERN CAPE", "NORTH WEST"	South African provinces

Procedure Information

Column Name	Data Type	Description	Sample Values	Business Rules
BUCKET	String	Procedure bucket/category	"NA"	Currently not applicable
TR_PROCEDURE_CODE_DESCRIPTION	String	Associated procedure description	"NA", "Medical Per Diem", "Intravenous infusion..."	Related medical procedures

Financial Breakdown (Multiple Payment Sources)

Column Name	Data Type	Description	Sample Values	Business Rules
AMT_PAID_ATB	Float	Amount paid by Above Threshold Benefit	0	Currently zero across all records
AMT_PAID_CEB	Float	Amount paid by Catastrophic Emergency Benefit	0, 22282.31, 7671.57	Emergency benefit payments
AMT_PAID_GPN	Float	Amount paid by General Practitioner Network	0	Currently zero across all records
AMT_PAID_HCC	Float	Amount paid by Hospital Cash Cover	14000, 7529.9, 8898.67	Hospital coverage payments
AMT_PAID_HCC_ADMIN	Float	Amount paid by HCC Administration	0	Administrative fees (currently zero)
AMT_PAID_MEM	Float	Amount paid by Member	0	Member co-payments (currently zero)
AMT_PAID_MOB	Float	Amount paid by Medical Out of Benefits	0	Out-of-benefit payments (currently zero)
AMT_PAID_MSA	Float	Amount paid by Medical Savings Account	0	MSA deductions (currently zero)
AMT_PAID_PFR	Float	Amount paid by Private Fee for service	0	Private fee payments (currently zero)
AMT_PAID_PMB	Float	Amount paid by Prescribed Minimum Benefits	14000, 7529.9, 8898.67	PMB coverage payments
AMT_PAID_PMB_CHRONIC	Float	Amount paid by PMB Chronic benefits	14000, 7529.9, 8898.67	Chronic condition PMB payments
AMT_PAID_PROV	Float	Amount paid to Provider	14000, 7529.9, 8898.67	Total provider reimbursement
AMT_PAID_TP	Float	Amount paid by Third Party	0	Third-party payments (currently zero)

Summary Financial Metrics

Column Name	Data Type	Description	Sample Values	Business Rules
AMT_PAID	Float	Total amount paid	14000, 7529.9, 8898.67	Sum of all payment sources
AMT_CLAIMED	Float	Total amount claimed	22863.97, 7529.9, 8898.67	Original claim amount

Volume Metrics

Column Name	Data Type	Description	Sample Values	Business Rules
QTY	Float	Quantity dispensed	28, 4, 15, 12, 1	Number of units dispensed
CLAIMS	Integer	Number of claims	1, 2	Count of individual claims

Data Quality Notes

- **Patient Privacy:** ENTITY_NO is encrypted/hashed for patient confidentiality

- **Completeness:** Minimal null values (15 missing DEG_DESCR, 2 missing TREATING_DR)

- **Temporal Range:** Data spans 36 months (2017-2019)

- **Clinical Focus:** Dataset specifically targets Multiple Sclerosis medications

- **Financial Granularity:** Detailed breakdown of payment sources across 13 different benefit types

Business Context

Clinical Domain

- **Therapeutic Area:** Specialized pharmaceutical claims across various medical conditions

- **Medication Types:** High-cost, specialized pharmaceuticals requiring specialist supervision

- **Patient Population:** Adult patients (18+ years) with various medical conditions

Financial Structure

- **Payment Model:** Complex multi-source reimbursement system

- **Key Benefit Types:**

    - PMB (Prescribed Minimum Benefits) - primary coverage for chronic conditions

    - HCC (Hospital Cash Cover) - hospital-related expenses

    - CEB (Catastrophic Emergency Benefit) - high-cost emergency coverage

- **Cost Management:** Significant difference between claimed and paid amounts indicates cost control measures

Provider Network

- **Dispensing Channels:** Primarily specialized pharmacies (independent and chain)

- **Geographic Distribution:** Concentrated in major urban centers (Gauteng, Western Cape)

- **Specialty Focus:** Specialist-prescribed medications requiring specialized handling

Key Relationships

- Each record represents a unique prescription/claim event

- ENTITY_NO links to individual patients (encrypted for privacy)

- ATC hierarchy provides 5-level therapeutic classification

- Financial breakdown shows how claims are processed across different benefit pools
- Geographic dimensions track both patient residence and provider location

Usage Recommendations for Snowflake Implementation

Performance Optimization

- **Partitioning:** Partition by YEAR and MONTH_KEY for temporal queries
- **Clustering:** Cluster on NAPPI9 and PROVINCE for product and geographic analysis
- **Indexing:** Create indexes on ENTITY_NO, ATC_DESCRIPTION, and PROVIDER_GROUP

Data Types

- **Encrypted Fields:** Use VARCHAR for ENTITY_NO (encrypted patient IDs)
- **Financial Fields:** Use NUMBER(10,2) for monetary amounts
- **Categorical Fields:** Use VARCHAR with appropriate constraints
- **Temporal Fields:** Use DATE type for DATE_KEY, INTEGER for YEAR/MONTH_KEY

Security Considerations

- **PII Protection:** ENTITY_NO is pre-encrypted but should be treated as sensitive
- **Access Controls:** Implement role-based access for patient-level data
- **Audit Logging:** Track access to individual patient records

Analytics Opportunities

- **Utilization Analysis:** Track medication usage patterns by demographics
- **Cost Analysis:** Compare claimed vs. paid amounts across benefit types
- **Geographic Trends:** Analyze prescription patterns by province/region
- **Provider Performance:** Evaluate dispensing patterns across provider networks
- **Therapeutic Monitoring:** Track treatment adherence and switching patterns