

# Q.CheckUp Lite Dataset - Data Dictionary

## Dataset Overview

- **Dataset Name:** Q.CheckUp Lite Healthcare Analytics Data
- **Purpose:** Healthcare claims and utilization analytics for medical products and services
- **Data Period:** Monthly healthcare transaction data
- **Record Type:** Transaction-level records aggregated by product, provider, and location
- **Total Records:** 200 rows (sample dataset)
- **Total Columns:** 24 columns

## Column Definitions

### Temporal Dimensions

| Column Name | Data Type | Description                      | Sample Values | Business Rules |
|-------------|-----------|----------------------------------|---------------|----------------|
| YEAR        | Integer   | Calendar year of the transaction | 2021          | Format: YYYY   |
| MONTH_NO    | Float     | Calendar month number            | 1.0           | Range: 1-12    |

### Product Identification

| Column Name  | Data Type | Description  | Sample Values   | Business Rules  |
|--------------|-----------|--|---|---|
| NAPPI9       | Integer   | 9-digit National Pharmaceutical Product Interface code | 227044001, 454018003                                      | Unique product identifier in South African healthcare |
| MANUFACTURER | String    | Name of the product manufacturer                       | "Systagenix Wound Management (Pty) Ltd",<br>"Apex Vision" | Company name as registered                            |

### Provider Information

| Column Name       | Data Type | Description                           | Sample Values   | Business Rules                 |
|-------------------|-----------|---------------------------------------|---|--------------------------------|
| PRACTICE_NO_DESCR | String    | Practice or facility name             | "Care At Midstream Sub Acute Facility", "Dermatologist" | Healthcare provider identifier |
| CATEGORY_DESCR    | String    | Provider category/type                | "Hospitals", "Dermatologist", "Pharmacy"                | 13 unique categories           |
| PROVIDER_GROUP    | String    | Provider group or network affiliation | "NHN", "Life Healthcare", "Dis-Chem"                    | 24 unique provider groups      |

Geographic Dimensions

| Column Name    | Data Type | Description            | Sample Values                             | Business Rules               |
|----------------|-----------|------------------------|---|------------------------------|
| P_PROVINCE     | String    | Provider province code | "Gauteng", "North West", "Kwa Zulu Natal" | South African province codes |
| PROVINCE_DESCR | String    | Full province name     | "Gauteng", "Western Cape", "Free State"   | Full province descriptions   |

Product Hierarchy (4-Level Classification)

| Column Name  | Data Type | Description                          | Sample Values                              | Business Rules                                |
|--------------|-----------|--------------------------------------|--|---|
| HIGH_LEVEL_1 | String    | Top-level product category with code | "WouM: Wound Management", "Sut: Sutures"   | Format: "CODE: Description"                   |
| HIGH_LEVEL_2 | String    | Second-level product subcategory     | "WouM-SynD: Synthetic Dressing"            | Format: "PARENT_CODE-CHILD_CODE: Description" |
| HIGH_LEVEL_3 | String    | Third-level product specification    | "WouM-SynD-SiliD: Silicone Dressing"       | Further refinement of product type            |
| HIGH_LEVEL_4 | String    | Fourth-level detailed specification  | "WouM-SynD-SiliD-SiliD: Silicone Dressing" | Most granular product classification          |

Simplified Product Hierarchy

| Column Name | Data Type | Description                   | Sample Values                             | Business Rules              |
|-------------|-----------|-------------------------------|---|-----------------------------|
| TR_LEVEL_1  | String    | Simplified top-level category | "Wound Management", "Sutures", "Syringes" | Clean version without codes |

Concatenated Classifications

| Column Name          | Data Type | Description                                 | Sample Values   | Business Rules                     |
|----------------------|-----------|---|---|------------------------------------|
| LEVELS_CONCAT_TILL_2 | String    | Levels 1-2 concatenated with pipe separator | "Wound Management   Synthetic Dressing"   | Format: "Level1   Level2"          |
| ALL_LEVELS_CONCAT    | String    | Levels 1-3 concatenated                     | "Wound Management   Synthetic Dressing   Silicone Dressing"                     | Format: "Level1   Level2   Level3" |
| ALL_LEVELS_CONCAT_4  | String    | All 4 levels concatenated                   | "Wound Management   Synthetic Dressing   Silicone Dressing   Silicone Dressing" | Complete hierarchy path            |

Financial Metrics

| Column Name    | Data Type | Description                              | Sample Values        | Business Rules                                      |
|----------------|-----------|--|----------------------|---|
| AMT_CLAIMED_TY | Float     | Amount claimed this year (current year)  | 0.00, 12.42, 36.57   | Currency in South African Rand (ZAR)                |
| AMT_CLAIMED_LY | Float     | Amount claimed last year (previous year) | 324.72, 180.00, 0.00 | Currency in South African Rand (ZAR)                |
| AMT_PAID_TY    | Float     | Amount paid this year (current year)     | 0.00, 12.42, 36.57   | May be less than claimed due to deductibles/co-pays |
| AMT_PAID_LY    | Float     | Amount paid last year (previous year)    | 324.72, 143.70, 0.00 | Actual reimbursement amount                         |

Volume Metrics

| Column Name | Data Type | Description                         | Sample Values | Business Rules                  |
|-------------|-----------|-------------------------------------|---------------|---------------------------------|
| UNITS_TY    | Float     | Number of units dispensed this year | 0, 23, 2, 1   | Physical quantity of products   |
| UNITS_LY    | Float     | Number of units dispensed last year | 3, 1, 0, 20   | Previous year comparison volume |

Technical Fields

| Column Name | Data Type | Description                 | Sample Values | Business Rules                          |
|-------------|-----------|-----------------------------|---------------|---|
| WF          | Integer   | Workflow or processing flag | 1, 2, 3, 4, 5 | Range: 1-11, likely processing sequence |

## Data Quality Notes

- **Completeness:** No null values present in sample dataset
- **Uniqueness:** NAPPI9 codes are mostly unique (185 unique values across 200 records)
- **Referential Integrity:** Geographic and provider codes appear consistent
- **Temporal Coverage:** Sample data represents January 2021 (YEAR=2021, MONTH\_NO=1)

## Business Context

- **TY/LY Comparison:** "This Year" vs "Last Year" allows for year-over-year analysis
- **Hierarchical Classification:** 4-level product taxonomy enables drill-down analysis
- **Geographic Analysis:** Province-level geographic breakdown for regional insights
- **Provider Analysis:** Multiple provider dimensions (category, group, individual practice)
- **Financial vs Volume:** Separate tracking of monetary and unit volume metrics

## Key Relationships

- Each record represents a unique combination of: Product (NAPPI9) + Provider + Time Period
- Product hierarchy levels are nested (Level 1 → Level 2 → Level 3 → Level 4)
- Concatenated fields are derived from individual hierarchy levels
- Financial amounts (claimed vs paid) show the relationship between requests and actual reimbursements

## Usage Recommendations for Snowflake Implementation

- **Partitioning:** Consider partitioning by YEAR and MONTH\_NO for performance
- **Indexing:** Create indexes on NAPPI9, P\_PROVINCE, and CATEGORY\_DESCR for common queries
- **Data Types:** Use appropriate Snowflake data types (NUMBER for financials, VARCHAR for text)
- **Hierarchies:** Consider implementing as separate dimension tables for better normalization