Products Data Quality Report

Detected Data Issues found by LLM

- Inconsistent casing in 'category' and 'product_category' columns (e.g., 'CLOTHING', 'clothing', 'Sports').

- The 'brand' and 'manufacturer' columns have inconsistent casing and formatting (e.g., 'Brand D', 'BrandA', 'brandE', 'brand_b', 'BRAND-C'). Also, both columns contain empty strings, representing missing values, and they both have a high percentage of nulls.

- The 'description' column has a high percentage of null values (28.0%).

- The 'weight' column has a significant percentage of missing values (44.0%).

- The 'dimensions' column contains empty strings, representing missing values.

- The 'color' and 'size' columns contain empty strings, representing missing values, and a high percentage of nulls.

- The 'supplier_id' column has a high percentage of null values (26.0%).

- The 'created_date' column has a high percentage of null values (19.0%).

- The 'last_updated' column contains empty strings, representing missing values.

- The 'is_active' column has mixed data types (str, int, bool) and inconsistent string representations (e.g., 'yes', 'no', 'true', 'false', 1, 0).

- The 'rating' column has a high percentage of null values (30.0%).

- Potential redundancy between 'product_name' and 'item_name', as well as 'product_id' and 'item_id'.

- The columns 'price', 'list_price', and 'cost' are strings instead of numeric types.

- The 'category' and 'product_category' columns seem to have a many-to-many relationship which might not be correct. For example, a product can have category 'clothing' but product_category 'Electronics'.

Column-wise Summary generated by utility

Column   : product_id
Types   : str
Sample Values : ['PROD_001', 'PROD_002', 'PROD_003', 'PROD_004', 'PROD_005']
Unique Count : 200
Null %   : 0.0%
Notes   : None

Column   : item_id
Types   : int
Sample Values : [1, 2, 3, 4, 5]
Unique Count : 200
Null %   : 0.0%
Notes   : None

Column   : product_name

Types   : str

Sample Values : ['Product 1', 'Product 2', 'Product 3', 'Product 4', 'Product 5']

Unique Count : 200

Null %   : 0.0%

Notes   : None

Column   : item_name

Types   : str

Sample Values : ['Item 1 Name', 'Item 2 Name', 'Item 3 Name', 'Item 4 Name', 'Item 5 Name']

Unique Count : 200

Null %   : 0.0%

Notes   : None

Column   : description

Types   : str

Sample Values : ['Description for product 2', 'Description for product 3', 'Description for product 4', 'Description for product 6', 'Description for product 7']

Unique Count : 144

Null %   : 28.0%

Notes   : None

Column   : category

Types   : str

Sample Values : ['CLOTHING', 'clothing', 'Sports', 'Toys', 'electronics']

Unique Count : 8

Null %   : 0.0%

Notes   : None

Column   : product_category

Types   : str

Sample Values : ['Toys', 'clothing', 'Sports', 'Books', 'CLOTHING']

Unique Count : 8

Null %   : 0.0%

Notes   : None

Column   : brand

Types   : str

Sample Values : ['Brand D', 'BrandA', 'brandE', 'brand_b', '']

Unique Count : 6

Null % : 15.5%

Notes : None

Column : manufacturer

Types : str

Sample Values : ['brandE', 'BRAND-C', 'BrandA', '', 'brand_b']

Unique Count : 6

Null % : 18.5%

Notes : None

Column : price

Types : float

Sample Values : [485.55, 119.25, 436.02, 11.97, 267.75]

Unique Count : 199

Null % : 0.0%

Notes : None

Column : list_price

Types : float

Sample Values : [270.33, 582.58, 188.79, 383.12, 389.6]

Unique Count : 200

Null % : 0.0%

Notes : None

Column : cost

Types : float

Sample Values : [168.81, 167.96, 285.86, 120.18, 101.7]

Unique Count : 200

Null % : 0.0%

Notes : None

Column : weight

Types : float

Sample Values : [5.27, 10.09, 1.19, 4.67, 7.31]

Unique Count : 106

Null % : 44.0%

Notes : None

Column   : dimensions

Types   : str

Sample Values : ['13x42x43', '13x4x27', '', '37x19x11', '31x44x3']

Unique Count : 107

Null %   : 0.0%

Notes   : None

Column   : color

Types   : str

Sample Values : ['Black', 'Green', 'Red', 'Blue', '']

Unique Count : 6

Null %   : 12.5%

Notes   : None

Column   : size

Types   : str

Sample Values : ['S', 'XL', '', 'One Size', 'L']

Unique Count : 6

Null %   : 11.0%

Notes   : None

Column   : stock_quantity

Types   : int

Sample Values : [376, 119, 180, 347, 582]

Unique Count : 183

Null %   : 0.0%

Notes   : None

Column   : stock_level

Types   : int

Sample Values : [426, 453, 340, 196, 352]

Unique Count : 164

Null %   : 0.0%

Notes   : None

Column   : reorder_level

Types   : int

Sample Values : [36, 46, 25, 13, 35]

Unique Count : 48

Null % : 0.0%

Notes : None

Column : supplier_id

Types : str

Sample Values : ['SUP_04', 'SUP_02', 'SUP_15', 'SUP_03', 'SUP_17']

Unique Count : 20

Null % : 26.0%

Notes : None

Column : created_date

Types : str

Sample Values : ['2020-05-17', '2021-09-23', '2022-04-30', '2022-06-25', '2022-05-22']

Unique Count : 153

Null % : 19.0%

Notes : None

Column : last_updated

Types : str

Sample Values : ['2023-12-24T18:30:00.000Z', '2023-07-08T18:30:00.000Z', '2023-08-17T18:30:00.000Z', '2023-05-25T18:30:00.000Z', '']

Unique Count : 134

Null % : 0.0%

Notes : None

Column : is_active

Types : str, int, bool

Sample Values : ['yes', 'no', False, 'true', 'false']

Unique Count : 6

Null % : 0.0%

Notes : Multiple types found: ['str', 'int', 'bool'].

Column : rating

Types : float

Sample Values : [4.6, 2.6, 3.5, 2.1, 0.9]

Unique Count : 48

Null %   : 30.0%

Notes   : None