Orders Data Quality Report

Detected Data Issues found by LLM

- Redundant columns: 'order_id' (str) and 'ord_id' (int) likely represent the same information. Consider choosing one and removing the other.

- Redundant columns: 'customer_id' (int) and 'cust_id' (str) likely represent the same information. Consider choosing one and removing the other.

- Inconsistent date format in 'order_date' column: Mixed formats like '5/9/2023', '2023-11-30', and '1/13/2023' are present. Standardize to a single format (e.g., YYYY-MM-DD).

- Missing values in 'order_date' column: 9.0% null values.

- Missing values in 'order_datetime' column: 19.3% null values.

- Redundant columns: 'quantity' and 'qty' both represent quantity, but have different scales. Investigate the difference and consolidate or clarify their meaning.

- Inconsistent casing in 'status' and 'order_status' columns: Values like 'delivered', 'processing', 'CANCELLED', 'pending', 'SHIPPED' exist. Standardize the casing (e.g., all lowercase or title case).

- Redundant columns: 'status' and 'order_status' seem to represent the same information, but have slight differences and missing values. Investigate and consolidate.

- Missing values in 'status' column: 15.1% null values.

- Missing values in 'order_status' column: 13.3% null values.

- High percentage of missing values in 'notes' column: 78.6% null values. Consider if this column is necessary or if the missing values can be addressed.

- High percentage of missing values in 'tracking_number' column: 37.6% null values. Investigate the reason for missing tracking numbers.

- Inconsistencies between order_date and order_datetime: In some records, the dates are different (e.g., order_id ORD_00821). Investigate if this is an error.

- Possible calculation errors: Values in 'price', 'unit_price', 'quantity', 'discount', 'tax', 'shipping_cost', 'total_amount', and 'order_total' may not be consistent. Verify the calculations.

- Single Unique Value: 'notes' column has only one unique value ('Special delivery instructions') apart from null. This column is not providing much information.

Column-wise Summary generated using utility

Column   : order_id

Types   : str

Sample Values : ['ORD_00001', 'ORD_00002', 'ORD_00003', 'ORD_00004', 'ORD_00005']

Unique Count : 1000

Null %   : 0.0%

Notes   : None

```
Column       : ord_id
Types        : int
Sample Values : [1, 2, 3, 4, 5]
Unique Count  : 1000
Null %       : 0.0%
Notes        : None
----------------------------------------
Column       : customer_id
Types        : int
Sample Values : [180, 337, 384, 22, 168]
Unique Count  : 432
Null %       : 0.0%
Notes        : None
----------------------------------------
Column       : cust_id
Types        : str
Sample Values : ['CUST_0180', 'CUST_0337', 'CUST_0384', 'CUST_0022', 'CUST_0168']
Unique Count  : 432
Null %       : 0.0%
Notes        : None
----------------------------------------
Column       : order_date
Types        : str
Sample Values : ['5/9/2023', '2023-11-30', '2023-01-17', '2023-08-06', '1/13/2023']
Unique Count  : 480
Null %       : 9.0%
Notes        : None
----------------------------------------
Column       : order_datetime
Types        : str
Sample   Values   :   ['2023-01-03T17:48:00.000Z',   '2023-01-20T08:04:00.000Z',   '2023-02-20T13:37:00.000Z',
'2023-09-02T02:34:00.000Z', '2023-02-15T22:46:00.000Z']
Unique Count  : 806
Null %       : 19.3%
Notes        : None
----------------------------------------
Column       : product_id
Types        : str
```

Sample Values : ['PROD_198', 'PROD_044', 'PROD_001', 'PROD_129', 'PROD_074']

Unique Count : 199

Null %   : 0.0%

Notes   : None

Column   : item_id

Types   : int

Sample Values : [146, 41, 100, 153, 157]

Unique Count : 200

Null %   : 0.0%

Notes   : None

Column   : quantity

Types   : int

Sample Values : [8, 1, 9, 2, 7]

Unique Count : 10

Null %   : 0.0%

Notes   : None

Column   : qty

Types   : int

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

Unique Count : 5

Null %   : 0.0%

Notes   : None

Column   : unit_price

Types   : float

Sample Values : [130.62, 36.21, 146.71, 109.64, 413.82]

Unique Count : 990

Null %   : 0.0%

Notes   : None

Column   : price

Types   : float

Sample Values : [404.94, 324.83, 145.07, 232.03, 69.9]

Unique Count : 986

Null %   : 0.0%

Notes   : None

Column   : total_amount

Types   : float

Sample Values : [140.51, 887.54, 352.45, 674.66, 813.13]

Unique Count : 994

Null %   : 0.0%

Notes   : None

Column   : order_total

Types   : float

Sample Values : [570.3, 700.84, 834.33, 203.85, 178.07]

Unique Count : 993

Null %   : 0.0%

Notes   : None

Column   : shipping_cost

Types   : float

Sample Values : [0.0, 12.26, 17.65, 7.0, 13.77]

Unique Count : 591

Null %   : 0.0%

Notes   : None

Column   : tax

Types   : float

Sample Values : [9.98, 13.2, 14.0, 10.51, 50.25]

Unique Count : 894

Null %   : 0.0%

Notes   : None

Column   : discount

Types   : float

Sample Values : [0.0, 18.36, 30.96, 14.69, 38.16]

Unique Count : 279

Null %   : 0.0%

Notes   : None

Column   : status

Types   : str

Sample Values : ['delivered', 'processing', 'CANCELLED', 'pending', 'SHIPPED']

Unique Count : 6

Null %   : 15.1%

Notes   : None

Column   : order_status

Types   : str

Sample Values : ['SHIPPED', 'CANCELLED', 'processing', 'pending', 'delivered']

Unique Count : 6

Null %   : 13.3%

Notes   : None

Column   : payment_method

Types   : str

Sample Values : ['credit_card', 'paypal', 'bank_transfer', 'cash', 'debit_card']

Unique Count : 5

Null %   : 0.0%

Notes   : None

Column   : shipping_address

Types   : str

Sample Values : ['8984 Oak Ave', '6820 Oak Ave', '751 Oak Ave', '9781 Oak Ave', '6747 Pine Rd']

Unique Count : 984

Null %   : 0.0%

Notes   : None

Column   : notes

Types   : str

Sample Values : ['Special delivery instructions']

Unique Count : 1

Null %   : 78.6%

Notes   : None

Column   : tracking_number

Types   : str

Sample Values : ['TRK510443', 'TRK118744', 'TRK745174', 'TRK919582', 'TRK429287']

Unique Count : 624

Null %   : 37.6%

Notes   : None