> User provided file path: C:\Users\mdudu\Documents\Data Explorer Project\Datasets\retail_sales_messy.csv

Dataset loaded successfully! It has 2200 rows and 15 columns.

--- Detecting Issues ---

Missing Values Per Column:

order_id 0

customer_id 0

customer_name 0

order_date 0

product_id 0

product_name 0

quantity 0

unit_price 0

discount% 219

total_price 0

payment_method 0

store_location 0

customer_email 381

phone_number 0

loyalty_status 877

Duplicate Rows: 180

Column Data Types:

order_id int64

customer_id object

customer_name object

order_date object

product_id object

product_name object

quantity int64

unit_price object

discount% object

total_price object

payment_method object

store_location object

customer_email   object

phone_number   object

loyalty_status   object

No negative values detected.

--- Cleaning Data ---

Detected boolean-like columns: []

Detected date-like column: 'order_date'

Attempted to parse 'order_date' with multiple formats. Invalid entries replaced with NaT.

The column 'order_date' has 361 missing or invalid entries.

> User chose: 1

Rows with missing or invalid dates in 'order_date' have been dropped.

'order_date' standardized to '%Y-%m-%d' format.

'order_date' cleaned and sorted chronologically.

Cleaning text-based column: 'customer_id'

Cleaning text-based column: 'customer_name'

Cleaning text-based column: 'order_date'

Cleaning text-based column: 'product_id'

Cleaning text-based column: 'product_name'

Cleaning text-based column: 'unit_price'

Cleaning text-based column: 'discount%'

Cleaning text-based column: 'total_price'

Cleaning text-based column: 'payment_method'

Cleaning text-based column: 'store_location'

Cleaning text-based column: 'customer_email'

Cleaning text-based column: 'phone_number'

Cleaning text-based column: 'loyalty_status'

Cleaning price column: 'unit_price'

Cleaning price column: 'total_price'

Cleaning discount column: 'discount%'

Cleaning phone number column: 'phone_number'

Analyzing column 'order_id' for outliers...

Column 'order_id' contains 0 outliers (below 26.25 or above 3964.25).

Analyzing column 'quantity' for outliers...

Column 'quantity' contains 82 outliers (below -1.0 or above 7.0).

> User chose option 3.

Skipping outlier handling for 'quantity'.

Analyzing column 'unit_price' for outliers...

Column 'unit_price' contains 0 outliers (below -5751.0 or above 13049.0).

Analyzing column 'discount%' for outliers...

Column 'discount%' contains 0 outliers (below -14.5 or above 45.5).

Analyzing column 'total_price' for outliers...

Column 'total_price' contains 151 outliers (below -15259.5 or above 34752.5).

> User chose option 3.

Skipping outlier handling for 'total_price'.

Column 'discount%' has 175 missing values.

> User chose option 1.

Rows with missing values in 'discount%' have been dropped.

The dataset has 153 duplicate rows.

> User chose: yes

Duplicate rows have been removed.

Column 'customer_id' contains non-numeric values.

> User chose: no

Column 'customer_name' contains non-numeric values.

> User chose: no

Column 'order_date' contains non-numeric values.

> User chose: no

Column 'product_id' contains non-numeric values.

> User chose: no

Column 'product_name' contains non-numeric values.

> User chose: no

Column 'payment_method' contains non-numeric values.

> User chose: no

Column 'store_location' contains non-numeric values.

> User chose: no

Column 'customer_email' contains non-numeric values.

> User chose: no

Column 'phone_number' contains non-numeric values.

> User chose: yes

Converted 'phone_number' to numeric. Invalid entries replaced with NaN.

Column 'loyalty_status' contains non-numeric values.

> User chose: no

--- Final Cleaned Dataset ---

	order_id	customer_id	customer_name	order_date	product_id	product_name	quantity	unit_price	discount%	total_price	payment_method	store_location	customer_email	phone_number	loyalty_status
0	1932	cust244	charles goodwin	20230101	prod134	screen protector	4	999	27.0	2917	cash	san jose ca	daniel28examplecom	7233369270	nan
1	1210	cust251	richard cook	20230101	prod102	usbccable	12	1299	4.0	14964	credit card	chicago il	lbaldwinexampleorg	5156491433	silver
2	2093	cust396	rebecca delgado	20230101	prod112	smart watch	2	12999	13.0	25998	credit card	san antonio tx	nan	8142235688	new
4	2624	cust497	jones kiara	20230101	prod112	smart watch	5	12999	29.0	46146	cash	philadelphia pa	stacey54exampleorg	3662059452	nan
5	2808	cust496	smith stephanie	20230102	prod145	portable charger	16	2999	20.0	38387	cash	dallas tx	pvaughnexamplenet	5883831697	nan