

**Data Collection and Preprocessing Phase**

Date	20 March 2025
Team ID	LTVIP2025TMID27193
Project Name	Cosmetic insights: Navigating Cosmetics Trends and costumer Insights with Tableau
Maximum Marks	3 Marks

**Data Quality Report Template:**

Data Source	Data Quality issue	Severity	Resolution Plan
Cosmetics datasets	Inconsistent Data Types	Moderate	<ul style="list-style-type: none"><li>• <b>Price:</b> Should ideally be numeric (e.g., float or integer). While it appears to be numeric, it's worth confirming that there are no non-numeric characters present.</li><li>• <b>Rank:</b> Should be numeric (float or decimal) to represent rating values. Verify that all values are within a reasonable range (e.g., 1-5 or 0-5).</li><li>• <b>Combination, Dry, Normal, Oily, Sensitive:</b> These are likely intended to be boolean (True/False or 1/0). The data appears consistent with 1/0, but it should be explicitly verified.</li></ul>
Cosmetics datasets	Missing Values and Incomplete Data	Low	<ul style="list-style-type: none"><li>• <b>Ingredients:</b> The "Ingredients" column is truncated in many entries (e.g., "Algae (Sea...", "Galactomy...", "Water, Dic..."). This makes it difficult to fully understand the product's composition.</li><li>• <b>Brand Names:</b> Some brand names are abbreviated (e.g., "DRUNK EL" instead of "Drunk Elephant"). Consistency is needed.</li></ul>
Cosmetics datasets	Data Inconsistency and Types	High	<ul style="list-style-type: none"><li>• <b>Inconsistent Capitalization:</b> Brand and product names show inconsistent capitalization (e.g., "LA MER" vs. "lamer" or "Truth Seru" vs. "Truth Serum")</li></ul>
Cosmetics datasets	Data Cleaning	High	<ul style="list-style-type: none"><li>• <b>Standardize Brand and Product Names:</b> Ensure consistent capitalization and full names.</li><li>• <b>Correct Typos:</b> Fix errors like "Bl" and ensure consistent spelling.</li><li>• <b>Complete Ingredient Lists:</b> If possible, obtain the full ingredient lists for each product.</li><li>• <b>Remove Special Characters:</b> Replace or remove special characters as needed.</li></ul>

Cosmetics datasets	Data Validation	High	<ul style="list-style-type: none"><li>• <b>Verify Rank Values:</b> Ensure all ranks are within the expected range.</li><li>• <b>Confirm Data Types:</b> Explicitly define the data types for each column (numeric, boolean, string).</li></ul>