

Data Collection and Preprocessing Phase

Date	20 March 2025
Team ID	LTVIP2025TMID26724
Project Title	Cosmetic Insights: Navigating Cosmetics Trends and Consumer 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">Price: 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.Rank: 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).Combination, Dry, Normal, Oily, Sensitive: 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.
Cosmetics datasets	Missing Values and Incomplete Data	Low	<ul style="list-style-type: none">Ingredients: 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.Brand Names: Some brand names are abbreviated (e.g., "DRUNK EL" instead of "Drunk Elephant"). Consistency is needed.
Cosmetics datasets	Data Inconsistency and Typos	High	<ul style="list-style-type: none">Inconsistent Capitalization: Brand and product names show inconsistent capitalization (e.g., "LA MER" vs. "la mer" or "Truth Seru" vs. "Truth Serum").
Cosmetics datasets	Data Cleaning	High	<ul style="list-style-type: none">Standardize Brand and Product Names: Ensure consistent capitalization and full names.Correct Typos: Fix errors like "Bi" and ensure consistent spelling.Complete Ingredient Lists: If possible, obtain the full ingredient lists for each product.Remove Special Characters: Replace or remove special characters as needed.

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