**Data Collection and Preprocessing Phase**

| Date | 19 March 2025 |
| --- | --- |
| Team ID | LTVIP2025TMID19942 |
| Project Title | Cosmetic Insights Navigating Cosmetics Trends and Consumer Insights with Tableau |
| Maximum Marks | 10 Marks |

**Data Exploration and Preprocessing Template**

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

| **Section** | **Description** |
| --- | --- |
| Data Overview | * **Dataset Name:** Cosmetics Dataset * **Total Rows:** 1,472 * **Total Columns:** 11 * **Column Names:** Label, Brand, Name, Price, Rank, Ingredients, Combination, Dry, Normal, Oily, Sensitive * **Missing Values:** None detected * **Duplicate Rows:** None detected * **Data Types:**   + Categorical: Label, Brand, Name, Ingredients   + Numerical: Price, Rank, Combination, Dry, Normal, Oily, Sensitive |
| Data Cleaning | * No missing values detected, so no imputation needed. * No duplicate rows found. * Checked for inconsistencies in categorical fields. |
| Data Transformation | * Sorted dataset based on **Price** and **Rank** for better visualization. * Applied filtering based on **Skin Type Suitability** (Combination, Dry, Normal, Oily, Sensitive). * Created calculated fields for **average ranking by brand**. |
| Data Type Conversion | * No incorrect data types detected. * Ensured numeric fields (Price, Rank) were in appropriate formats. |
| Column Splitting and Merging | * No column splitting required. * Merged **skin suitability columns** into a single categorical field for easier analysis.. |
| Data Modeling | * Defined relationships between product attributes and rankings. * Ensured consistency in numerical and categorical attributes. |
| Save Processed Data | Saved cleaned dataset in CSV format for Tableau visualization. |