# Data Collection and Preprocessing Phase

| | |
|---|---|
| Date | 17 March 2025 |
| Team ID | LTVIP2025TMID25168 |
| Project Title | Cosmetic Insights |
| Maximum Marks | 3 Marks |

## Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

| Data Source | Data Quality Issue | Severity | Resolution Plan |
|---|---|---|---|
| Customer Reviews Dataset | Missing values in product reviews (e.g., incomplete feedback) | Moderate | Use imputation techniques such as mode-based filling or predictive modeling to estimate missing values. |
| Sales Transactions Dataset | Duplicate records of sales transactions | High | Apply deduplication techniques using unique transaction IDs and timestamps. |
| Social Media Data | Presence of irrelevant or spam data | Moderate | Implement NLP-based filtering to remove spam and irrelevant content. |
| Product Attributes Dataset | Inconsistent product category names | High | Standardize category names using a predefined taxonomy and regex matching. |

| Data Source | Data Quality Issue | Severity | Resolution Plan |
|---|---|---|---|
| Customer Demographics Dataset | Outdated or incorrect customer details | Low | Perform data validation checks and use third-party enrichment services to update records. |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |