

Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	739756
Project Title	Unveiling Baldness: Genetic And Environmental Dynamics
Maximum Marks	2 Marks

Data Quality Report:

The Data Quality Report 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 Quality Report:

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Missing values in the 'Age', 'Stress', 'Medical Condition', 'Smoking', 'Genetics', 'Weight Loss', 'Harmonal Changes', 'Medical Condition', and 'Hair Loss' columns.	Moderate	Use mean/median imputation.
Kaggle Dataset	Categorical data in the dataset	Moderate	Encoding has to be done in the data.