**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 July 2024 |
| Team ID | 740713 |
| Project Title | Genetic Classification Of An Individual By Using Machine Learning |
| Maximum Marks | 6 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 | Basic statistics, dimensions, and structure of the data. |
| Univariate Analysis | Exploration of individual variables (mean, median, mode, etc.). |
| Bivariate Analysis | Relationships between two variables (correlation, scatter plots). |
| Multivariate Analysis | Patterns and relationships involving multiple variables. |
| Outliers and Anomalies | Identification and treatment of outliers. |
| **Data Preprocessing Code Screenshots** | |
| Loading Data |  |
| Handling Missing Data |  |
| Data Transformation |  |
| Feature Engineering |  |
| Save Processed Data |  |