**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 June 2024 |
| Team ID | 740003 |
| Project Title | Disease prediction using Machine Learning |
| Maximum Marks | 6 Marks |

**Data Exploration and Preprocessing Template**

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | Dimension:  8 rows x 131 columns  Descriptive stastistics: |
| Univariate Analysis |  |
|  |  |
| Bivariate Analysis |  |
|  |  |
| Multivariate Analysis |  |
| Outliers and Anomalies | - |
| **Data Preprocessing Code Screenshots** | |
| Loading train Data |  |
| Loading test Data |  |
| Handling Missing Data  In train and test |  |
| Data Transformation |  |
| Feature Engineering | - |
| Save Processed Data | - |