**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 09 July 2024 |
| Team ID | 739734 |
| Project Title | Evolving efficient classification patterns in Lymphography |
| 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: 614 rows × 13 columns Descriptive statistics |
| Univariate Analysis |  |

|  |  |
| --- | --- |
| Bivariate Analysis |  |
| Multivariate Analysis |  |
| Outliers and Anomalies | Outliers in lymphography data can be data points that fall outside the expected range for lymph node size, flow patterns, or tracer distribution. These outliers may indicate errors during imaging, unusual anatomical features, or potential diseases requiring further investigation. Anomalies in lymphography could be specific patterns, like unexpected blockages or leakage, that deviate from the norm and warrant specialist attention. |
| **Data Preprocessing Code Screenshots** | |
| Loading Data |  |
| Handling Missing Data | No missing attributes |

|  |  |
| --- | --- |
| Data Transformation |  |
| Feature Engineering | Attached the codes in final submission. |
| Save Processed Data | Done |