**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 12 JULY 2024 |
| Team ID | SWTID1720438678 |
| Project Title | Early prediction of chronic kidney disease. |
| 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 | **Structure:** The data is organized in rows and columns, with column headers like "normal," "notpresent," "present," and seemingly unique patient identifiers in the first column.  **Dimensions:** It's difficult to determine the exact number of rows and columns from the limited view, but there are likely many rows representing individual patients and potentially dozens of columns containing various medical data points.  **Content:** The data includes a mix of numbers (potentially representing values for medical tests or measurements) and letters indicating presence or absence of certain conditions. |
| Univariate Analysis | Rbc,pc,pcc,ba,htn,dm,cad,appet,pe,ane |
| Bivariate Analysis |  |
| 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 |  |