**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 20 June 2024 |
| Team ID | 740018 |
| Project Title | Determine: Loan from KIVA crowdfunding data |
| Maximum Marks | 6 Marks |

**Data Exploration and Preprocessing Template**.

Data Exploration and Preprocessing Template for KIVA Crowdfunding: Load data, handle missing values, explore basic statistics, visualize distributions, encode categorical variables, normalize/scale features, identify outliers, and prepare for modeling.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | Summary of the dataset, including number of rows and columns, data types of each column, and brief descriptions of each column. |
| Univariate Analysis | Distribution analysis of individual variables using histograms, bar charts, and descriptive statistics (mean, median, mode, standard deviation). |
| Bivariate Analysis | Examination of relationships between pairs of variables using scatter plots, correlation matrices, and pairwise plots to identify patterns and trends. |
| Multivariate Analysis | Investigation of interactions between multiple variables using heatmaps, PCA (Principal Component Analysis), and clustering to understand data structure. |
| Outliers and Anomalies | |  | | --- | | Identification and description of outliers and anomalies,  summarized in a table with details on detection method,  number of outliers, description, and potential impact. |   Top of Form  Bottom of Form |
| **Data Preprocessing Code Screenshots** | |
| Loading Data | Import the dataset into the environment using libraries such as pandas. Ensure correct data types and initial data review. |
| Handling Missing Data | Identify and address missing values by using methods like deletion, imputation, or filling with default values. |
| Data Transformation | Perform necessary data transformations like normalization, scaling, encoding categorical variables, and more. |
| Feature Engineering | Create new features from existing ones to improve model performance. Includes techniques like feature extraction and selection. |
| Save Processed Data | Create new features from existing ones to improve model performance. Includes techniques like feature extraction and selection. |