Exploratory Data Analysis for Loan Prediction

In the world of machine learning, **data preprocessing** plays a crucial role in ensuring that the raw data is transformed into a format that can be efficiently and effectively used by machine learning algorithms. The primary goal is to clean the data, handle missing or incorrect values, deal with outliers, and transform variables to enhance the model's predictive power.

The dataset in question pertains to loans, with columns such as client_id, loan_type, loan_amount, repaid, loan_id, loan_start, loan_end, and rate. This dataset needs to undergo several preprocessing steps before it can be fed into a machine learning model, including handling missing values, outliers, and encoding categorical variables.

Problem Statement:

The task is to preprocess the given **loan data** to prepare it for machine learning. This includes:

- **Importing necessary libraries** and loading the dataset.

- **Identifying and handling missing values** (if any).

- **Detecting and treating outliers** to ensure data consistency.

- **Transforming variables** to bring all data into a usable format for the model.

- **Scaling numerical features** to standardize their values.

- **Encoding categorical variables** (like loan_type) into numerical formats.

- **Creating new variables** if necessary, to better capture features that may assist in predictive modelling.

- **Splitting the data** into a training and test set for model validation.

Dataset: loans_EDA