TASK 1: Bank Loan Classification Report

- **Submitted by: Jonathan Rai**

Approach:

- In this analysis, I performed data preprocessing (checking null values, dropping irrelevant columns which does not help in making predictions), exploratory data analysis (EDA), performed log transformation on skewed variables, feature engineering, and logistic regression to predict whether a customer will accept a personal loan or not.

Key Findings:

- After preprocessing, missing values were imputed for “Home Ownership” with mode and “Income” with mean value as they were necessary features for making predictions.
- Dropped irrelevant columns such as gender, customer id, experience, zip code.
- Reasons of dropping these columns:
  - Although 90.4% data is missing in 'Gender' column, but this column does not play significant role is loan approval or not, so it’s better to drop.
  - The customer id does not have relation with loan so dropping it too.
  - Since, income is usually given a priority for loan approval rather than work experience so, we remove experience column too.
  - zip code also does not directly contribute for loan approval so we remove it.
- After that, performed feature engineering to prepare dataset for modeling.
- EDA revealed insights into the distribution of variables and their relationships.

- Log transformation was applied to skewed variables to improve model performance.
- A Logistic Regression model was trained to predict personal loan acceptance with an accuracy of 97%.

Insights and Observations

- The log transformation helped reduce skewness to variables, leading to improve model performance.
- EDA showed various insights as influence of income and education on personal loan, High correlation between CCAvg (Average monthly spending with the credit card in thousands) and Income etc.
- The Logistic Regression Model demonstrated good precision and accuracy.