1. Preliminary Data Inspection

Data Types and Missing Values: Check for consistency in data types (e.g., Gender, Married, BankCustomer are integers that might represent binary or categorical features). Identify and handle missing or null values if they are encoded in a non-standard format.

Summary Statistics: Compute summary statistics for numerical features (Age, Debt, YearsEmployed, CreditScore, Income) to understand their distribution, scale, and presence of outliers.

Categorical Features Analysis: For categorical features (Industry, Ethnicity, Citizen), examine the distribution of categories and consider encoding techniques for converting them into numerical values suitable for SVM.

Missing Values:

Duplicates:

Outliers:

2. Exploratory Data Analysis

i. Univariate Analysis:

Plot histograms for numerical features to understand their distributions.

Use bar plots for categorical features to inspect the frequency of each category.

ii. Bivariate Analysis:

Investigate the relationship between each feature and the target variable (Approved) using appropriate plots (e.g., box plots for numerical features, stacked bar charts for categorical features) to identify patterns or correlations.

iii. Correlation Analysis: Compute correlation matrix for numerical features to identify potential multicollinearity.

3. Feature Engineering:

Feature Transformation: Apply scaling on numerical features to standardize their scale for SVM.

Feature Creation: Consider creating new features that might be relevant for the prediction (e.g., combining Age and YearsEmployed into a new feature if relevant).

Feature Selection: Use statistical tests and domain knowledge to select a subset of features that are most relevant to the target variable.

Encoding Categorical Variables: Apply one-hot encoding or label encoding to transform categorical variables into a machine-readable format.

Scaling: Since SVM is sensitive to the scale of the input features, apply feature scaling (standardization or normalization) to ensure that all features contribute equally to the model performance.