Project Title: Loan Approval Analysis

Problem Statement:

The loan approval process often involves evaluating multiple factors to determine whether an applicant qualifies for a loan. This project aims to analyze a dataset containing information on demographics, income, loan amounts, and credit history to understand the factors influencing loan approval decisions. The goal is to identify key patterns in the data and build a predictive model that can forecast loan approval outcomes, helping financial institutions make more informed and data-driven decisions.

Objectives:

1. **Missing Values Treatment:**

To investigate the impact of missing values in the dataset and the strategies to handle them (e.g., imputation, removal).

2. Demographic Insights:

To explore how factors like gender, marital status, number of dependents, education, and employment status influence the likelihood of loan approval.

3. **Income & Loan Amount Analysis:**

To analyze the impact of applicant and co-applicant income on loan approval.

To determine whether the loan amount requested has any correlation with income or loan approval.

4. **Credit and Loan Term Insights:**

To assess the influence of credit history and loan term on loan approval decisions.

5. **Property Area Analysis:**

To understand how the location of the property affects loan approval chances.

Tasks:

Task 1: Data Exploration

- Load and inspect the dataset.

- Perform initial descriptive statistics on numerical columns (ApplicantIncome, CoapplicantIncome, LoanAmount, Loan_Amount_Term, Credit_History).
- Analyze the distribution of categorical columns (Gender, Married, Dependents, Education, Self_Employed, Property_Area).
- Visualize and identify missing values in the dataset.

Task 2: Handling Missing Data

- Decide on strategies to handle missing data (e.g., Fill in missing numbers with estimates for numerical columns, use the most common category for missing values in categorical columns, or remove data that is missing too much information.").
- Implement the chosen methods for handling missing values.
- Check how missing data handling affects the dataset and model performance.

Task 3: Demographic Analysis

- Explore the relationship between loan approval and gender using bar plots.
- Analyze how marital status affects loan approval rates.
- Examine the influence of the number of dependents on loan approval.
- Investigate loan approval by education background.
- Assess the impact of self-employment status on loan approval.

Task 4: Income and Loan Amount Analysis

- Visualize the relationship between applicant income and loan approval using boxplots.
- Examine how co-applicant income influences loan approval.
- Perform correlation analysis between applicant income, co-applicant income, and loan amount.
- Compare loan amounts requested by different demographic groups (gender, marital status, education).

Task 5: Credit History and Loan Term Analysis

- Examine the loan approval rate for applicants with different credit histories.
- Analyze the relationship between loan term (Loan_Amount_Term) and loan approval rate.
- Investigate the interaction between credit history and loan term.

Task 6: Property Area and Loan Approval

- Analyze the distribution of loan approvals across different property areas (Urban, Semiurban, Rural).
- Investigate if property area has a significant impact on loan amounts requested.