**Data Set:**

**Loan Data:**

Please refer to the dataset file below in order to work on the tasks:

1. **Loans.csv** : Information about the loans, including loan amount, interest rate, term, etc.
2. **Borrowers.csv** : Information about the borrowers, including personal details.
3. **Repayments.csv** - Information about loan repayments, including the amount repaid and the date.

**Tasks:**

1. **SQL Queries:**

Write SQL queries to retrieve the following information:

a. The total number of loans issued.

b. The average loan amount.

c. The number of loans issued per month.

d. The total amount repaid.

1. **Data Preparation:**

Preprocess the data to handle missing values, encode categorical variables, and create features for modeling.

1. **Data Analysis:**

a. Calculate the default rate for loans. Define "default" as loans that were not repaid within 30 days of the due date.

b. Identify any correlations between the borrower's age, employment status, and the likelihood of default.

1. **Problem Definition:**

Clearly define the problem you want to solve. In this case, it's predicting loan default risk.

1. **Feature Enginering:**Create new features that might be useful for predicting loan default.
2. **Exploratory Data Analysis:**

Conduct exploratory data analysis to understand the distribution of features, identify patterns, and visualize relationships between variables.

1. **Data Visualization:**

Create a visual representation (e.g., chart or graph) to show the distribution of loan amounts using the tools that you are comfortable with.

1. **Report:**Present your key findings in a brief report with supporting visualizations. Attach Python code if necessary and provide strategic recommendations based on your analysis.

**“*Demonstrate your proficiency in data analysis and understanding of P2P lending. Effectively communicate findings and recommendations. Strive for excellence independently.”***