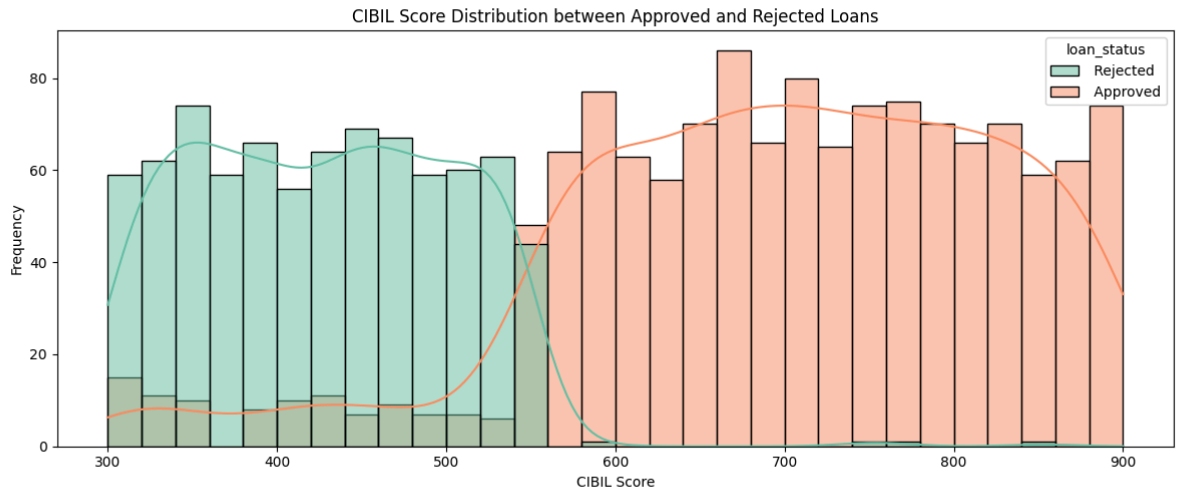
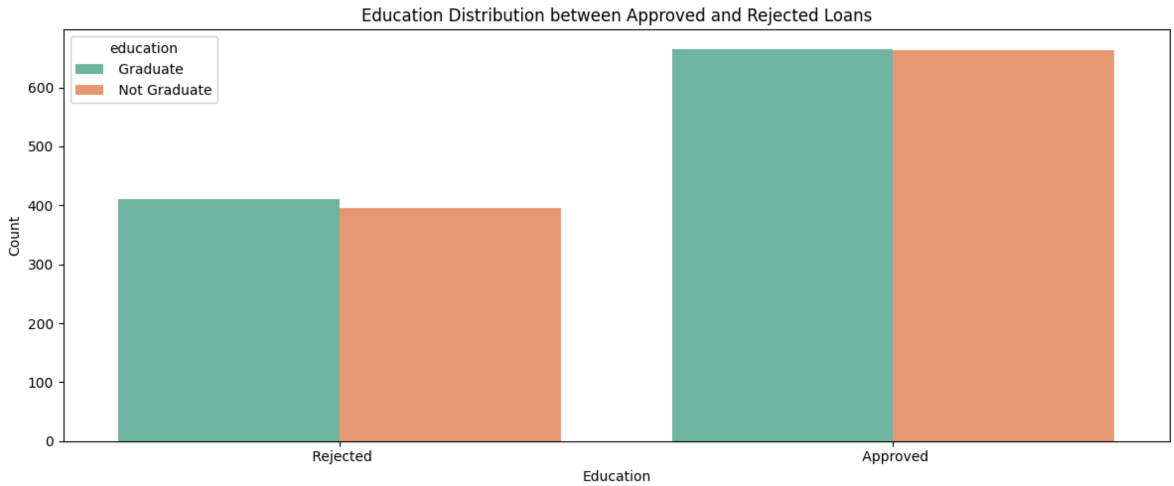
Question 1: Visualize the following and give description (maximum 100 words for all three factors).

a) CIBIL Score distribution between approved and rejected loans



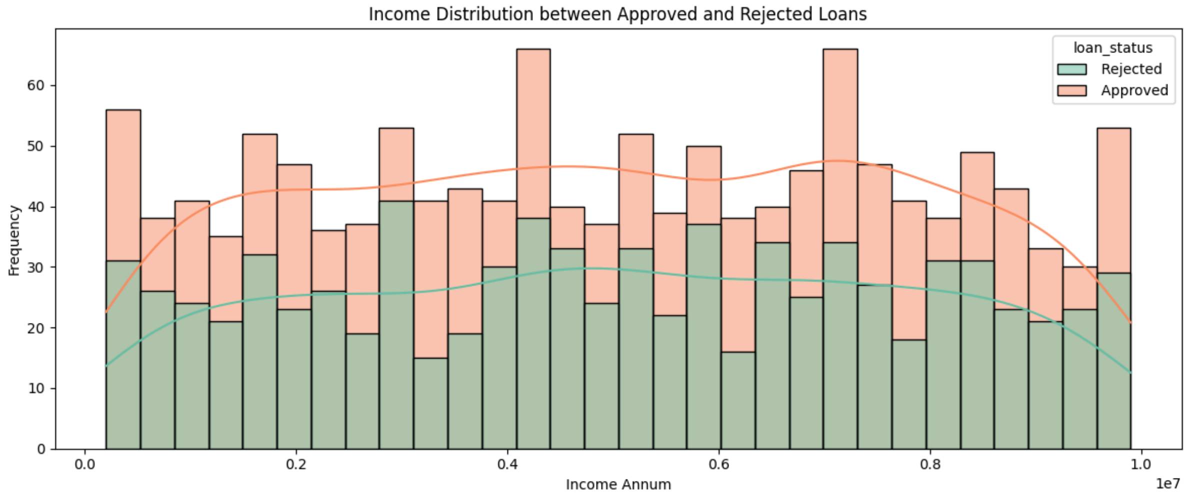
=> Approved loans generally have higher CIBIL scores (above 600), whereas rejected loans are concentrated in the lower score range (below 600), indicating that higher scores improve approval chances.

b) Education distribution between approved and rejected loans



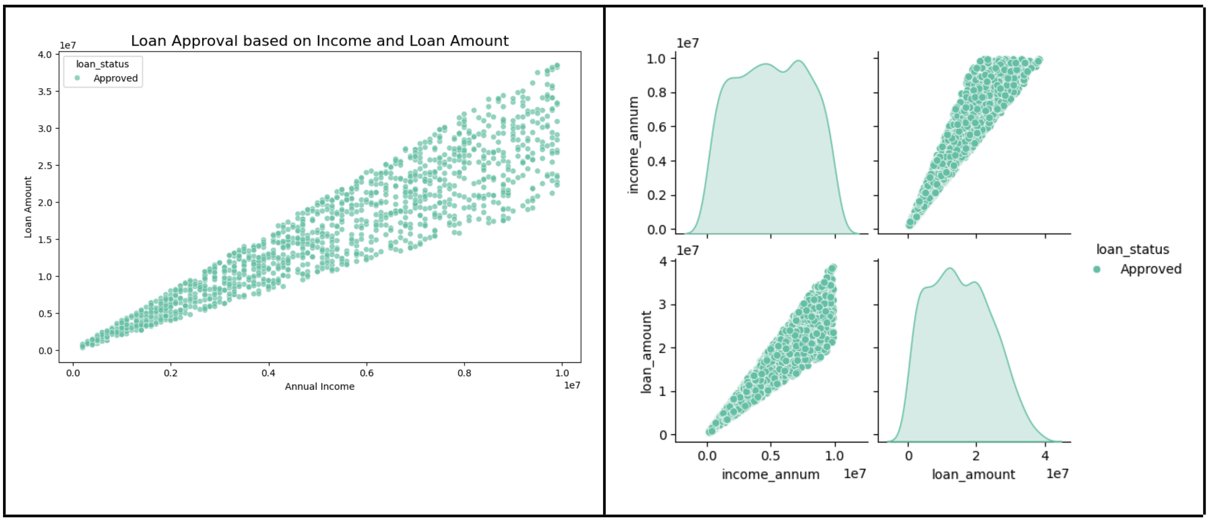
=> Education level does not significantly impact loan approval. Both graduates and non-graduates have similar counts in both approved and rejected categories.

c) Income distribution between approved and rejected loans



=> Income distribution shows a similar spread for both approved and rejected loans, with no strong relationship between higher income and loan approval.

Question 2: Visualize the relationship between income and loan amount for approved loans. Do you think the loan amount is directly proportional to the income level? Why or why not? (maximum 150 words)



=> The relationship between income and loan amount for approved loans appears to be positively correlated, as shown in both scatter plots. The loan amount generally increases as the annual income increases, suggesting a direct proportionality. This trend indicates that applicants with higher incomes are eligible for, or tend to request, larger loan amounts. However, the spread around the line of best fit suggests that while income is an influencing factor, it's not the sole determinant for the loan amount; other factors, such as credit score and asset values, may also play significant roles.

Question 3: If you were to include one more factor to predict loan outcomes, what specific factor would you choose to add, and what is the rationale behind this choice? (maximum 100 words)

Based on the performance of 5 best models, in addition to the "cibil_score" feature, you can consider the new feature "loan_to_income_ratio", calculated as the ratio of loan_amount to income_annum.

- A high loan_to_income_ratio suggests that a large part of the applicant’s income goes towards the loan, indicating a higher debt burden and potentially greater financial strain.
- Lenders often use this ratio to assess an applicant’s ability to manage loan payments relative to their income. Borrowers with a lower loan_to_income_ratio typically have more disposable income, making them less likely to default, while a higher ratio may indicate higher risk.