**Summary and Findings: Loan Approval Analysis**

This study aimed to identify the key factors influencing loan approval decisions. The primary variables analyzed were **income, loan amount, credit score, and employment experience** to determine their impact on approval rates.

For this project, a **dataset from Kaggle containing 44,972 loan applications** was used. The data was pre-processed, with extreme outliers removed to improve the accuracy of the analysis. The following key findings were identified:

**Descriptive Statistics & Data Distribution**

* + **All variables were positively skewed,** except for credit scores, which followed an approximately normal distribution.
  + **Age:** Most applicants were between **24-35 years old**.
  + **Income:** Right-skewed distribution, with most earning **below $100,000**.
  + **Loan Amount:** Most loans ranged from **$5,000-$15,000,** with a few high-value outliers.
  + **Credit Score:** The majority of applicants had scores **above 600**.
* **PMF & CDF Analysis**
  + **Probability Mass Function (PMF)** showeda **higher approval rate for lower-income applicants,** which was contrary to expectations.
  + **Cumulative Distribution Function (CDF)** confirmed that most borrowers fell within common income and loan amount ranges.
* **Correlation & Hypothesis Testing** 
  + **Loan Amount vs. Applicant Income:** Moderate positive correlation **(r = 0.36, p < 0.0001),** indicating that higher-income applicants tend to receive larger loans.
  + **Loan-to-Income Ratio & Loan Status**: A strong negative relationship—**higher loan percentages relative to income were associated with increased rejection probability (p < 0.0001)**.
  + **Employment Experience & Loan Approval: Extremely weak correlation (r = -0.02),** despite statistical significance.
  + **Credit Score & Interest Rate: No meaningful relationship (r = 0.01, p = 0.0176),** suggesting that other factors influence interest rates.

**Challenges & Limitations**

* **Most variables exhibited positive skewness**, affecting statistical assumptions.
* **Correlation does not imply causation -** factors such as lender policies, credit history details, and economic conditions may also influence loan approval.
* **Hypothesis testing and regression assume linearity**, which may not fully apply to all variables.
* **Additional relevant factors (**e.g., **debt-to-income ratio, loan purpose)** were not included in the dataset.

**Conclusions & Implications**

* **Loan-to-income ratio influences loan approval, but there is no strict cutoff rate where loans begin to be rejected.** Instead, **rejection rates gradually decrease after 20-30%, and approval rates increase beyond 30-40%.** This suggests that loan-to-income ratio alone is **not the primary factor** in approval decisions.
* **Higher-income applicants tend to receive higher loan amounts, but income alone is not the most significant approval factor.**
* **Lenders can use these insights to refine approval models, while borrowers should focus on maintaining a balanced loan-to-income ratio and other creditworthiness factors to improve their approval chances.**