Data Analysis of Loan Approvals: Key Insights for Decision-Making

Name : **Lekkala Sai Jayanth** **Student Id :** **23073007**

Email Id : sl24abn@herts.ac.uk **Link :** Git Hub

The leading rows involve the most essential features, such as income, the amount of the loan, and the status of the loan. Knowing these variables is necessary for exploring relationships within the data, allowing for a deep study of loan approval factors with potential trends for further analysis. The dataset has 4169 entries with all values available and 13 columns. This represents that it involves numeric features like income and loan amount along with categorical variables such as education and loan status, providing a complete and balanced view of all factors that might influence the approvals and possible relationships among them.

The summary statistics in general overview statistics for some key variables in the loan data, showcasing central tendencies and variability. The average loan amount is highly significant and indicates substantial borrowing behaviour. These statistics are crucial in understanding better decision-making and targeted strategies for loan approvals, which eventually improve financial assessments and customer profiling.

Figure1: Loan amount distribution

The histogram in ***Figure 1*** provides a graphical view of the distribution of loan amounts to help find patterns and outliers in data. This graph selection provides an evident view of frequency distribution, and it shows that most loans cluster around higher values, but a few lower loans point out other types of borrowing behaviours among applicants, which is left-skewed.

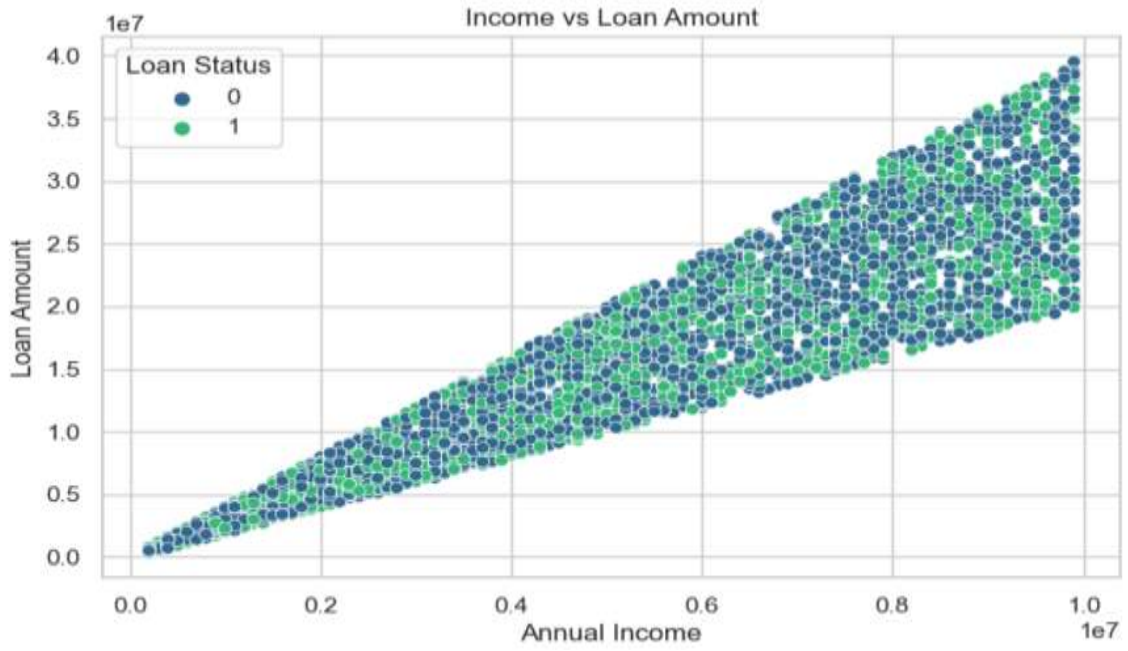

Figure 2: Income vs loan amount

The scatter plot in ***Figure 2*** shows the relationship between annual income and loan amounts, categorised by loan status. This visualisation effectively shows how income levels influence loan approvals and the tendency for higher-income applicants to receive larger loans.

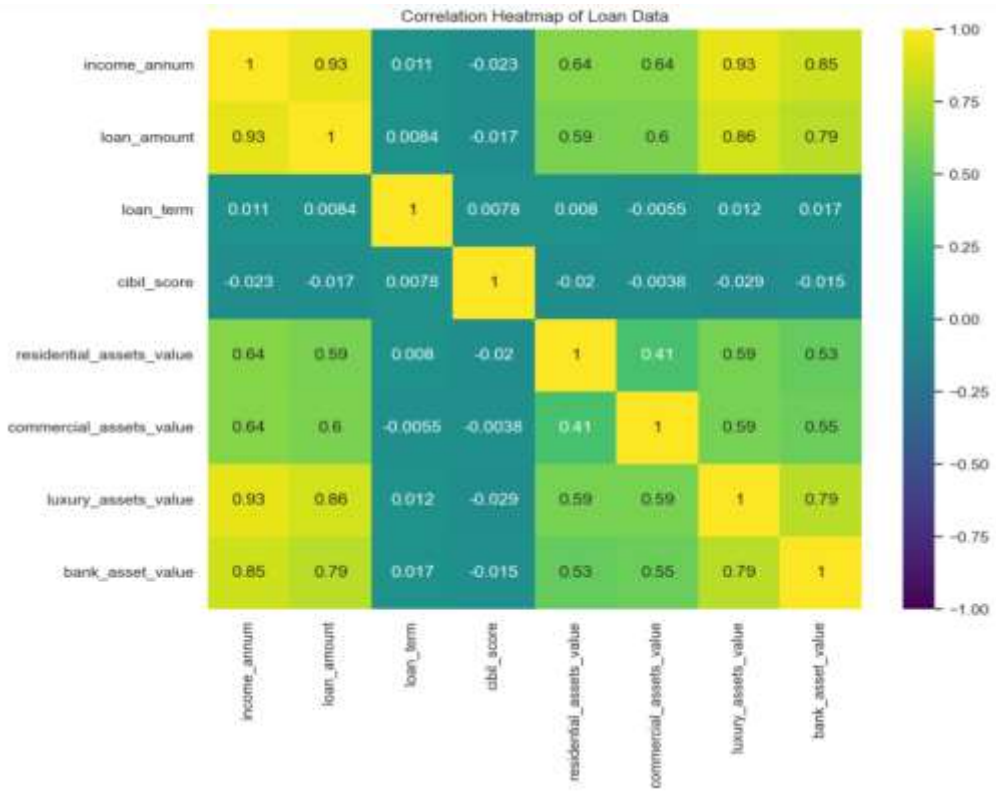

Figure 3: Correlation heatmap

The heatmap in ***Figure 3*** depicts the strength and direction of correlations of the critical numerical variables within the loan data. Annotated with correlation coefficients, it shows income is highly and positively correlated (0.93) with the loan amount, indicating that people with higher incomes tend to get loans easily due to their capability to repay them.