Statistics and Trends

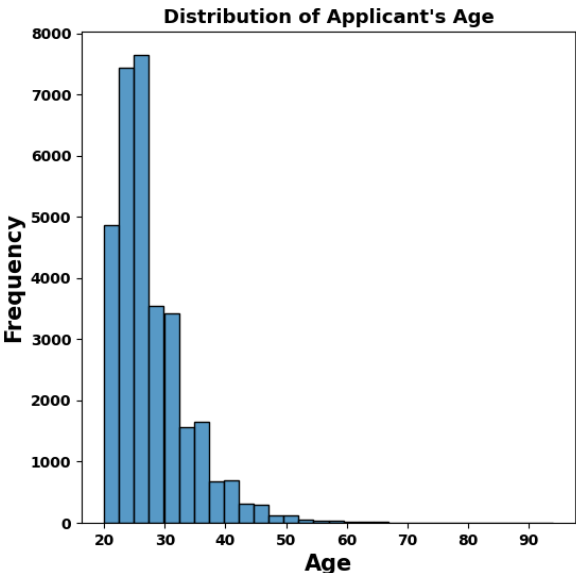**Student Name :** Vishal Kumar Senthil Kumar

Student ID : 23076841
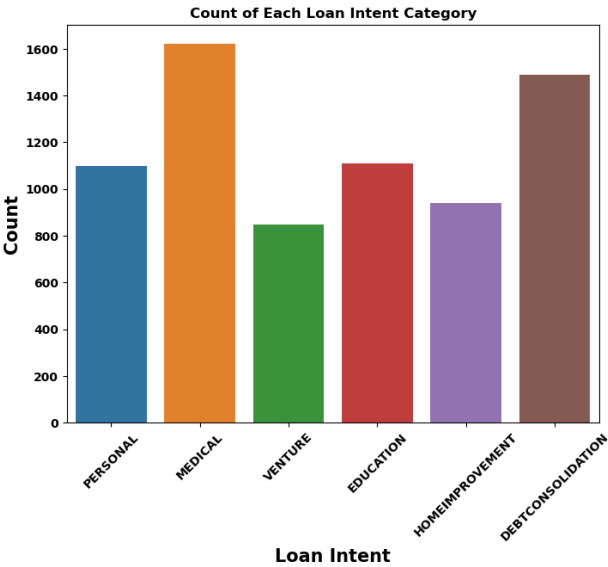
Github link : https://github.com/vishalkumar041298/statistics_and_trends

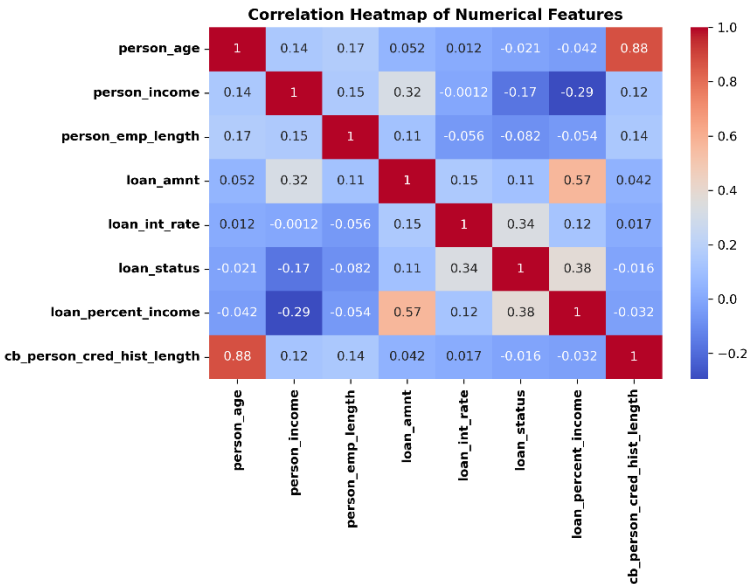# LOAN DEFAULTERS ANALYSIS REPORT

This report analyzes a credit risk dataset, focusing on characteristics that may influence loan approval and default rates. By examining demographic, financial, and loan-specific attributes, we aim to uncover insights that help understand applicant behavior and potential risk factors. Through visual exploration, we can interpret the relationships between variables such as age, income, home ownership, and loan intent. This approach supports data-driven lending decisions, providing valuable guidance for financial institutions to manage risk more effectively.

Count of Each Loan Intent Category

Distribution of Applicant's Age

The second chart displays the distribution of defaulters across different loan purposes. Categories like "MEDICAL" and "DEBT CONSOLIDATION" show the highest frequencies, indicating a high demand for loans in these areas. In contrast, "Venture" and "Home Improvement" have fewer applicants, reflecting a lower demand. This trend, when examined alongside age distribution, suggests that **younger adults are more inclined to take out loans for medical expenses**, likely because they are less likely to have medical insurance.
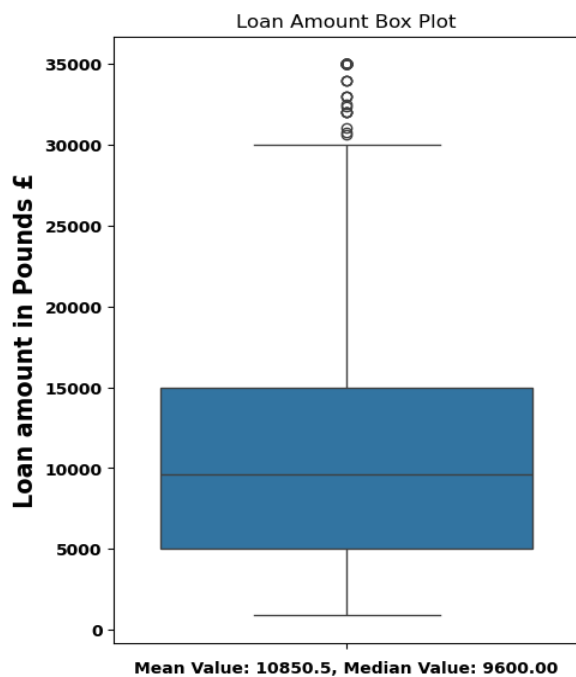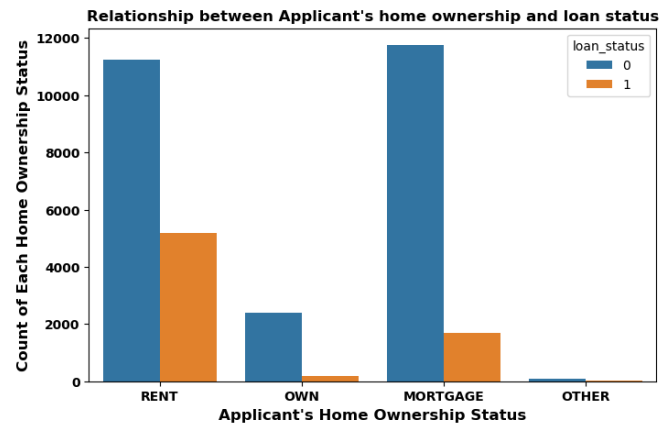
The first chart shows the histogram of age distribution that most loan applicants are between 20 and 40 years old, with fewer applications from those above 50. This trend could indicate that **younger individuals are more likely to seek loans**, potentially for ventures or education. Additionally, age may correlate with factors like employment length and income, as younger applicants often have shorter job histories and lower incomes. Age data helps financial institutions consider generational trends in borrowing, informing policies that balance access and risk management.

Correlation Heatmap of Numerical Features

The heatmap visualizes the correlations between key numerical features, such as age, income, employment length, loan amount, and interest rate. A high positive correlation exists between credit history length and age, suggesting that older applicants generally have lengthier credit histories. Moderate correlation between loan amount and percent income indicates that larger loans often represent a higher percentage of the applicant's income. Low correlation with loan status suggests that individual features may not strongly predict default, emphasizing the need for a combined approach in assessing credit risk.

The bar chart compares loan status (default vs. non-default) across different home ownership categories. Renters and mortgage holders make up the majority, with renters showing a higher tendency to default. This relationship suggests that **applicants who own their homes or have mortgages might be more financially stable compared to renters**.

The box plot illustrates the distribution of loan amounts, **with a median value near £9,600 and an average of around £10,850**. Outliers indicate that some applicants have received loans above £30,000, reflecting a range in borrowing needs. This variability can guide lenders in segmenting their customer base, potentially offering different loan products to applicants with varying financial needs. Outlier loans may carry higher risk, so institutions might adjust lending strategies for high-value loans.

The scatter plot shows the correspondence between income per annum and loan amount, with loan status (default or non-default) color-coded. Higher incomes generally correspond with larger loans, but defaults (highlighted in orange) appear across all income levels, indicating that higher income does not necessarily eliminate the risk of default. This visualization underscores the importance of a holistic approach in evaluating creditworthiness.