Palak Gupta

23116071

**DAI-101 Assignment 1**

**Dataset**

The dataset contains data regarding loan applicants (Education, Applicant Income, Gender, Loan Amount, Term, Location etc.) and their credit history.

**Initial Dataset Structure**

The Dataset contains 13 columns and 367 rows with null values in some columns.

**Data Cleaning**

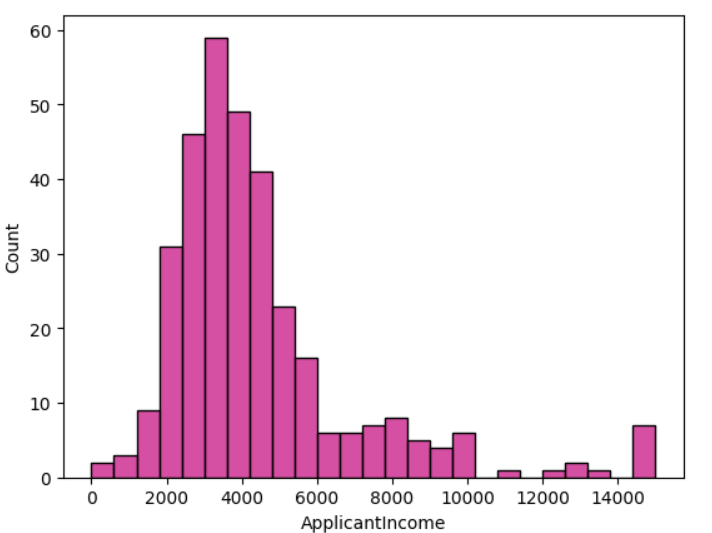
1. We first deal with null values by either replacing them with median/mean values in some columns or completely removing them in others.
2. Next, we try to visualize the outliers using boxplots and try to reduce them.
3. We also check for any duplicates, however this dataset doesn’t contain any.

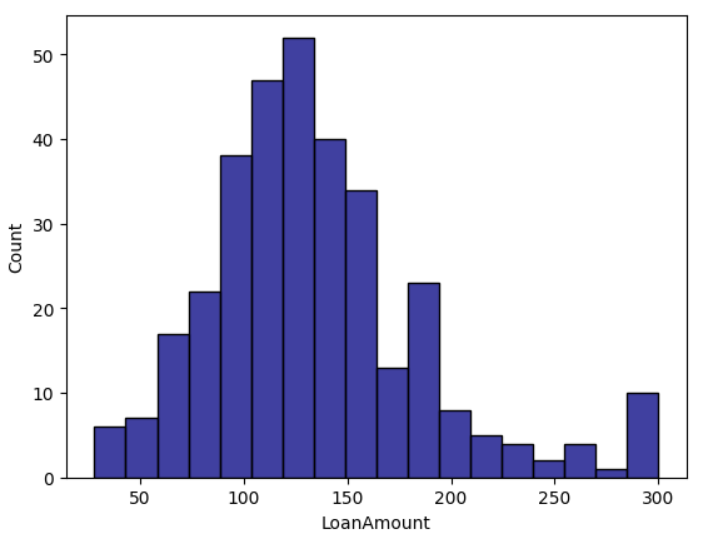
**Univariate Analysis**

We can see from the results that around 80% of total loan applicants were males and 20% were females.85% loans were approved. Applicants were almost equally distributed in urban , semi-urban and rural regions. Around 78% of the applicants were graduate (educated).

We can also observe from the statistics that Applicants have an average income of 4475 and median income of 3791. The loan amounts have an average of 133 and median of 126 , with max amount being 300.

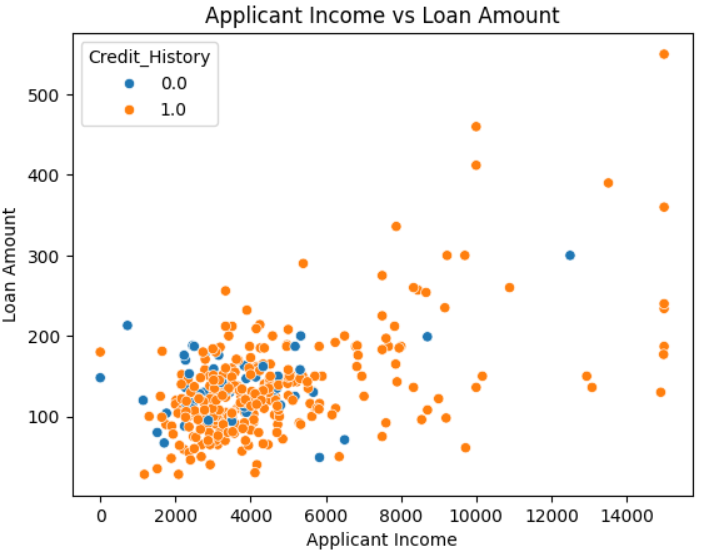
Plots:

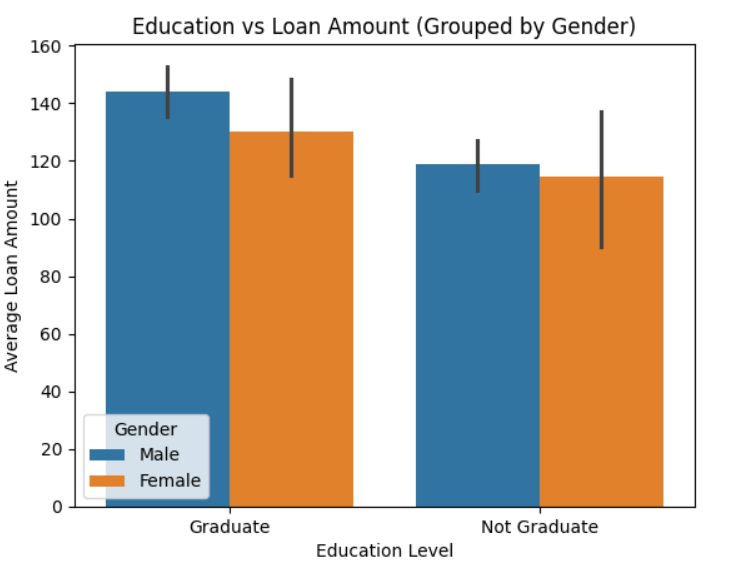


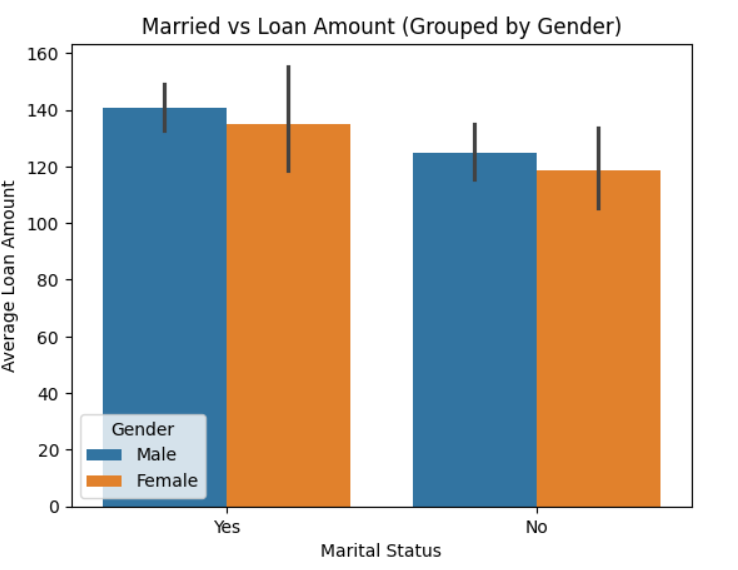


**Bivariate Analysis**

Scatter Plot

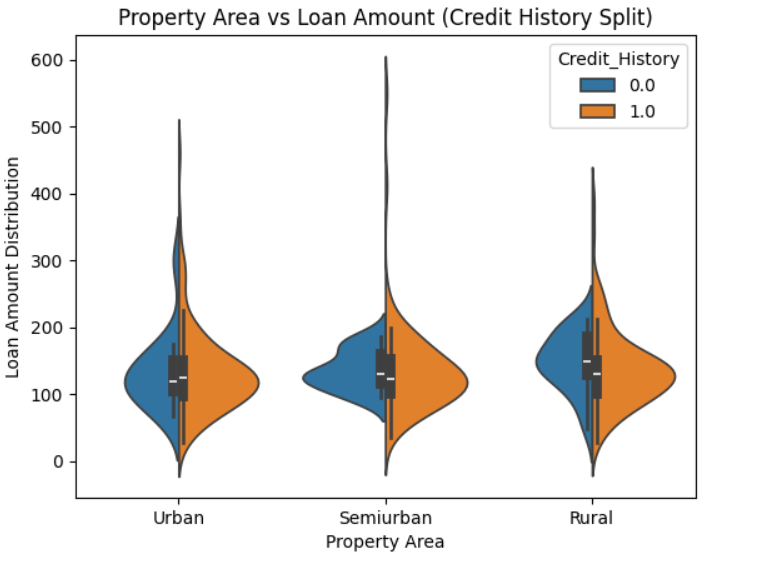




****

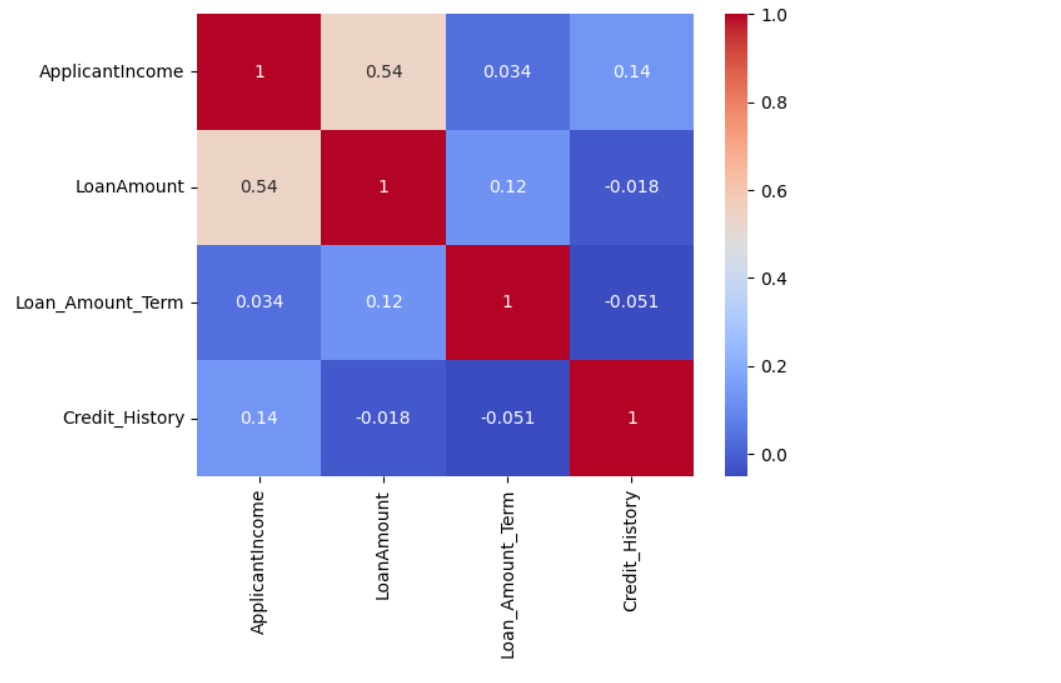
We can observe from the plot that Graduates and married people take more loans than non-graduates and unmarried. Moreover, Males take more loans than females.

Violin Plot

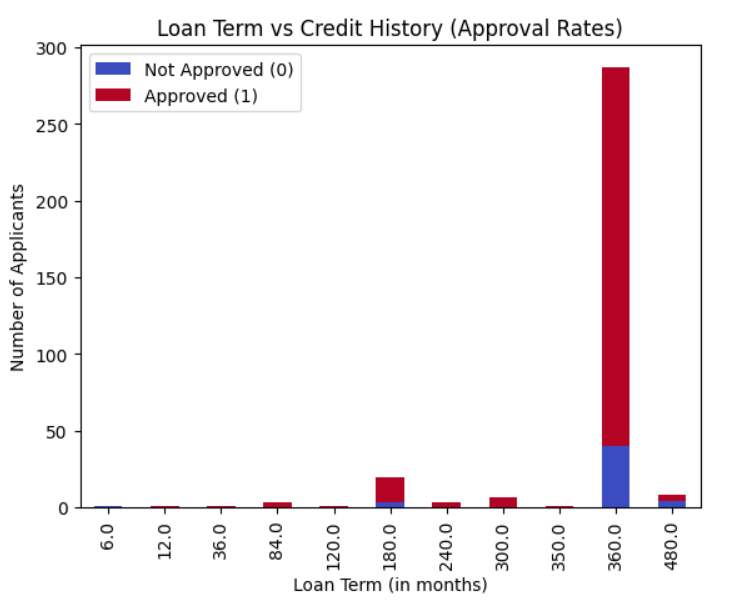


**Multivariate Analysis**

Heatmap



We can see correlation between Applicant Income and Loan Amount , which is expected. However interestingly , there is little correlation between Applicant Income and Credit History which means high income does not guarantee loan success.



This plot shows that most people apply for loans for a term of 360 days.