

Capstone Project 1: EDA - Inferential Statistics

This project focuses on determining the variables/factors that are most important in determining the possibility of securing a home loan.

The variables in the data includes 2 quantitative and 7 categorical. For the quantitative data a correlation will be evaluated using the linear regression model. While for the categorical data, a chi-square test will be applied and for those variables where visual inspection may lead to an incorrect assumption a proportion z-test will be completed.

What is the rate of loan approval?

The loans are approved at a rate of 0.7.



What's the correlation between loan amount and total income?

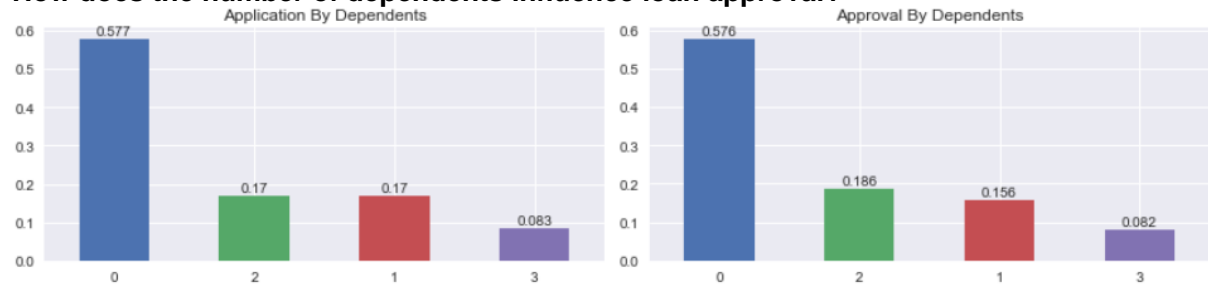
H_0 : There is no linear relationship between total income and loan amount.

H_A : There is a linear relationship between total income and loan amount.



The correlation coefficient of our observed data is $r = 0.657$ and the p-value is $4.7e-51$. With a p-value that is infinitesimally below 0.05 there is very strong evidence to reject H_0 for H_A .

How does the number of dependents influence loan approval?



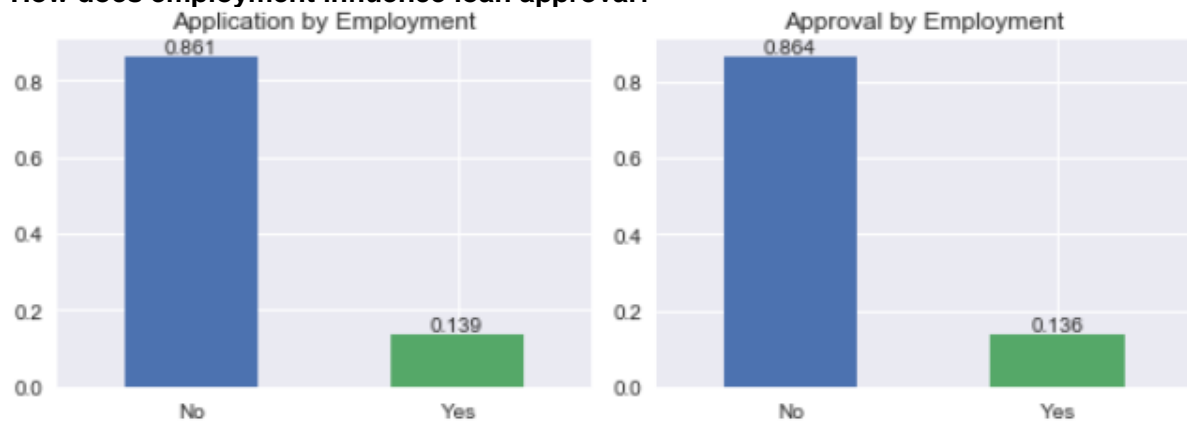
The rates of application and approval are indeed similar. Therefore, the number of dependents may not be taken into consideration when approving or denying a loan application.

H_0 : Loan approval is not related to the number of dependents.

H_A : Loan approval is related to the number of dependents.

The chi-square statistics is 3.55 and the p-value is 0.314. With a p-value (0.314) that is above 0.05 there isn't enough evidence to reject H_0 for H_A .

How does employment influence loan approval?



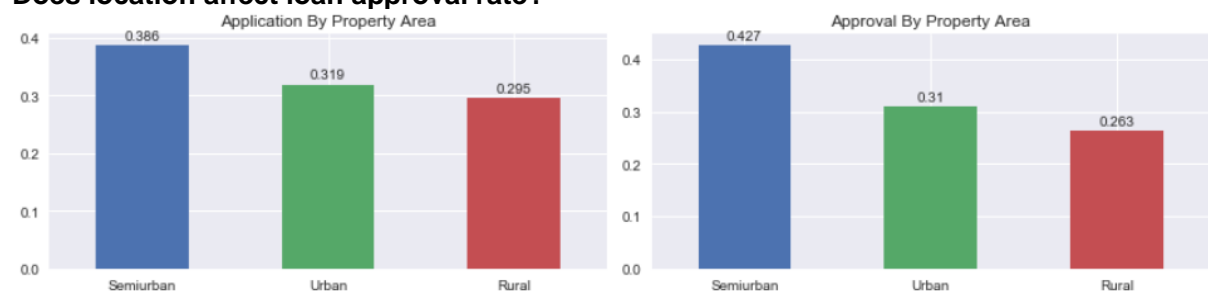
The rates of application and approval are indeed similar. Therefore, self-employment status may not be taken into consideration when approving or denying a loan application.

H_0 : There is no relationship between loan approval and employment type.

H_A : There is a relationship between loan approval and employment type.

The chi-square statistics is 0.01 and the p-value is 0.922. With a p-value (0.922) that is above 0.05 there isn't enough evidence to reject H_0 for H_A .

Does location affect loan approval rate?

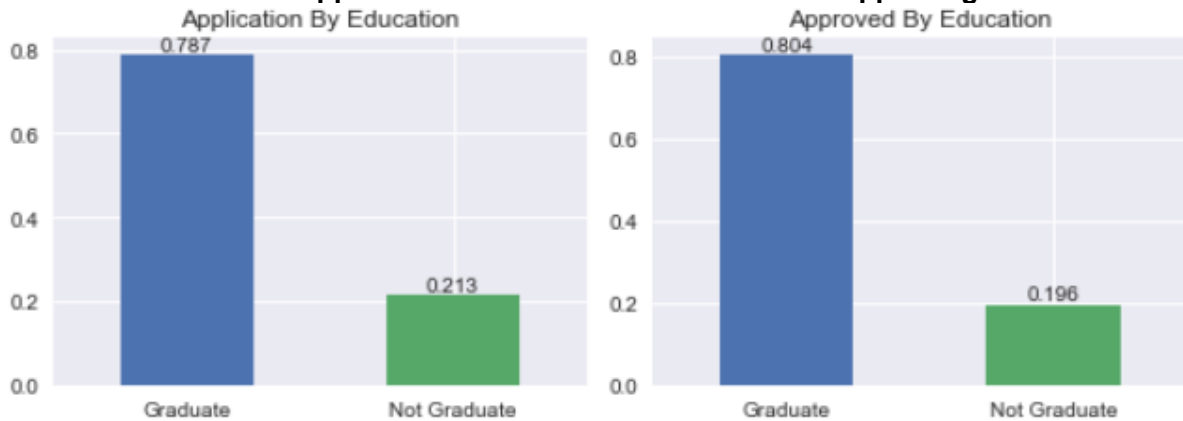


The ratio of approved (0.427) semiurban loans is greater than the ratio of semiurban applicants (0.386). This can lead to the conclusion that your location can influence your application.

H_0 : The variables loan approval and property location are independent.
 H_A : The variables loan approval and property location are not independent.

The chi-square statistics is 10.47 and the p-value is 0.0053. The p-value (0.0053) is below 0.05 there is enough evidence to reject H_0 for H_A .

Is the education of the applicant taken into consideration when approving a loan?



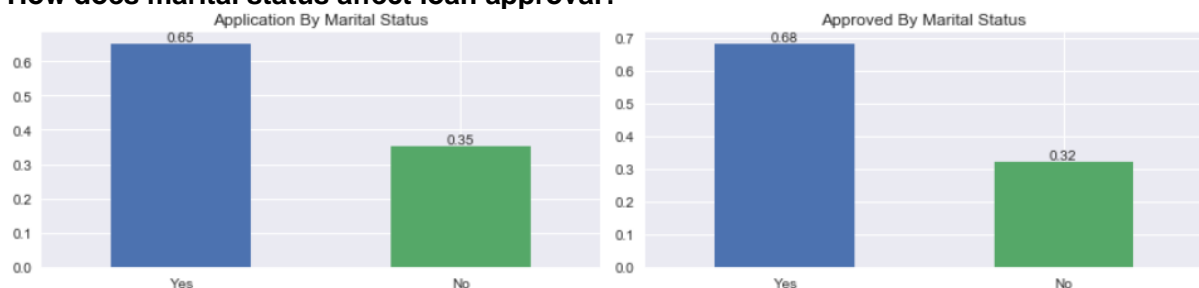
A greater portion of more educated persons loans were approved (0.804) compared to the proportion that applied (0.787). While a great proportion of less educated individuals were denied (0.253) than the proportion that applied (0.213). There's a higher likelihood that those who are less educated earn less and therefore this can affect their chances of receiving a loan.

H_0 : There is no statistically significant relationship between loan approval and educational status.

H_A : There is a statistically significant relationship between loan approval and educational status.

The chi-square statistics is 2.01 and the p-value is 0.156. The z-test statistics is -0.65 and the p-value is 0.514. Both a p-values are above 0.05, therefore there isn't enough evidence to reject H_0 for H_A .

How does marital status affect loan approval?



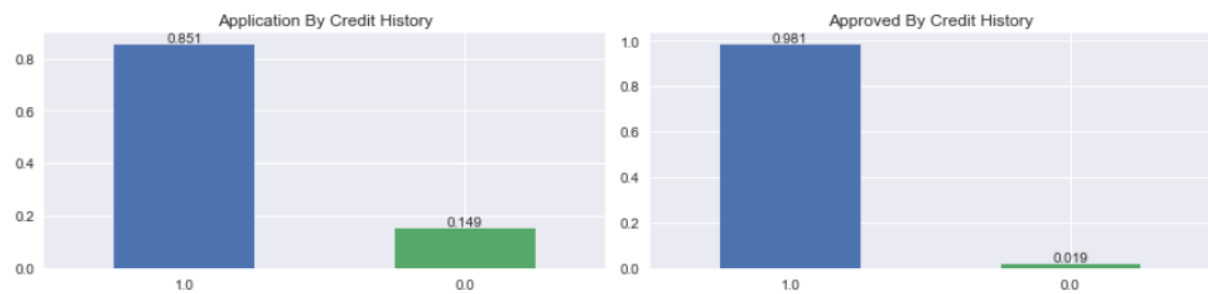
A larger portion of applicants are married and the proportion of approved (0.680) is greater than that which applied (0.651). Therefore, it is more likely that married applicants may have their loans approved. It is more likely that they are willing to apply jointly improving their chances due to a greater combined income.

H_0 : There is no relationship between loan approval and marital status.

H_A : There is a relationship between loan approval and marital status.

The chi-square statistics is 4.85 and the p-value is 0.028. The p-value (0.028) is below 0.05, there is enough evidence to reject H_0 for H_A .

Does having a credit history increase the chances of securing a loan?



Having a previous credit history (1.0) does increase the likelihood of receiving an approval. Creditors are more comfortable with someone who has already proven their credit worthiness, of the approved loans only 0.019 had no credit history.

H_0 : There isn't a statistical difference between loan approval and credit history.

H_A : There is a statistical difference between loan approval and credit history.

The chi-square statistics is 156.75 and the p-value is $5.81e-36$. With a p-value that is infinitesimally below 0.05 there is very strong evidence to reject H_0 for H_A .