

Lending Club Case Study

Group Facilitator – Purnachander Gajjala

Group Member – Pramod Khandare

Data Understanding :

- Number of loans: **39717**
- Number of attributes considered for each application: **111**
- Number of numerical attributes: **27**
- Number of categorical attributes: **84**
- Identified target column - Variable ***loan_status*** gives require information about status of loan.
- Identified columns with garbage columns - **68**
- Identified irrelevant columns - **18**
- Identified columns with mixed data types

Data Cleaning:

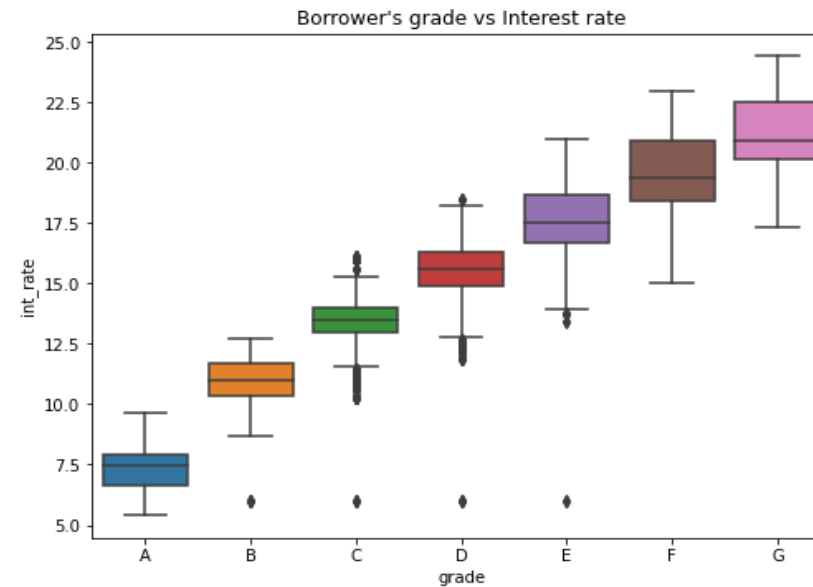
- Filtered out currently active loans
- Dropped columns with high percentage of garbage values
 - Dropped irrelevant columns
 - Handled columns with mixed data types
 - Handled columns wrongly mapped to object data type – int_rate, revol_util, earliest_cr_line, pymnt_plan, application_type
 - Removed percentage symbols
 - Converted ***term*** column to numeric

Data Cleaning :

- Converted earliest_cr_line column into datetime format
- Dropped columns with just one value
- Handled columns wrongly mapped to numeric data type – member_id, policy_code, acc_now_delinq, delinq_amnt
- Converted few numeric columns to object type
- Dropped columns with zero value: **31**
- Handled columns with high number of categories – addr_state
- Dropped columns with higher categories - **26**
- Set **id** as index, Dropped **policy_code** column
- Handled columns with low garbage values
- Dropping columns with zero value and low percentage of garbage values: **22**
- Replaced multiple garbage values with null
- Interpreted missing values in home_ownership column
- Interpreted missing values in emp_length column
- Interpreted missing values in revol_util column
- Interpreted missing values in pub_rec_bankruptcies column

Data Analysis :

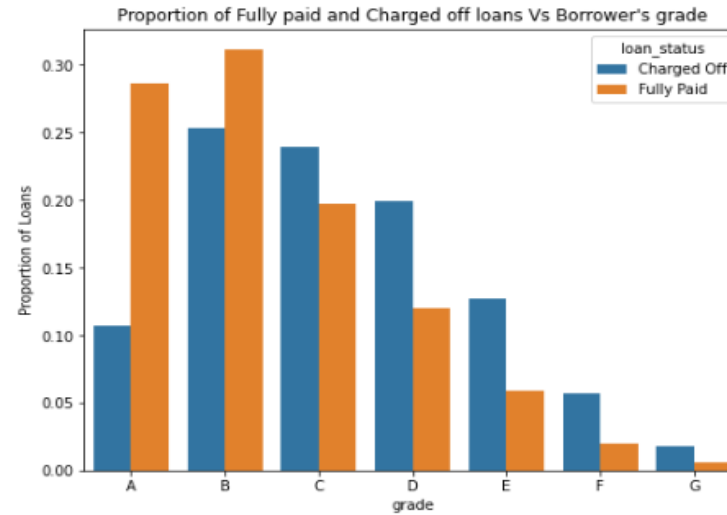
- Relation between **Grade** and **Interest rate** -
 - The **grade** of a borrower is based on past history of loans.
 - It generally indicates worthiness of a borrower.
 - Usually a borrower with high grade gets low interest rate offered and vice versa.



- Insights
 - Clearly above figure shows the relationship between borrower's grade and the interest rate at which loan is borrowed.
 - Borrowers with high grades like 'A', 'B' are likely to get loans at cheaper interest rates and vice versa

Data Analysis :

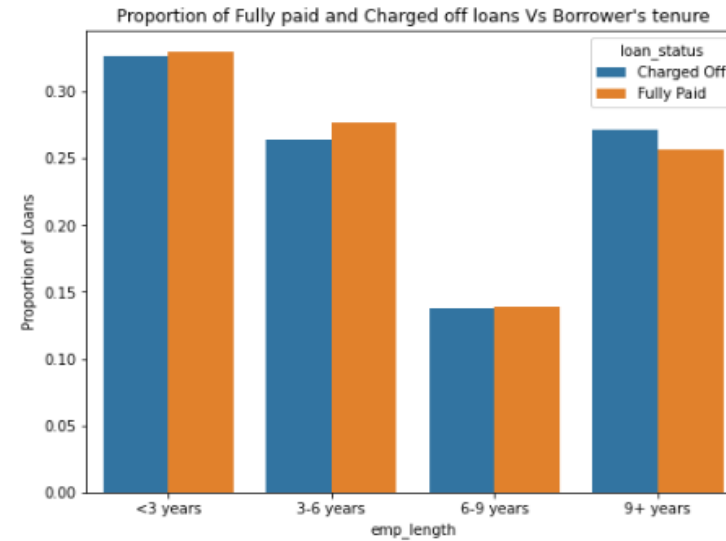
- Relation between **Grade** and **Loan Status**



- Insights
 - The proportion of defaulted loans increased with decrease in the borrower's grade.
 - Meanwhile, the proportion of fully paid loans decreased with decrease in the borrower's grade.
- Recommendation
 - Borrowers with grade 'A' are highly unlikely to default. Targeting these borrowers will reduce the risk.
 - Borrowers with grade 'C' and above are more likely to default.

Data Analysis :

- Relation between ***Employee Tenure*** and ***Loan Status***

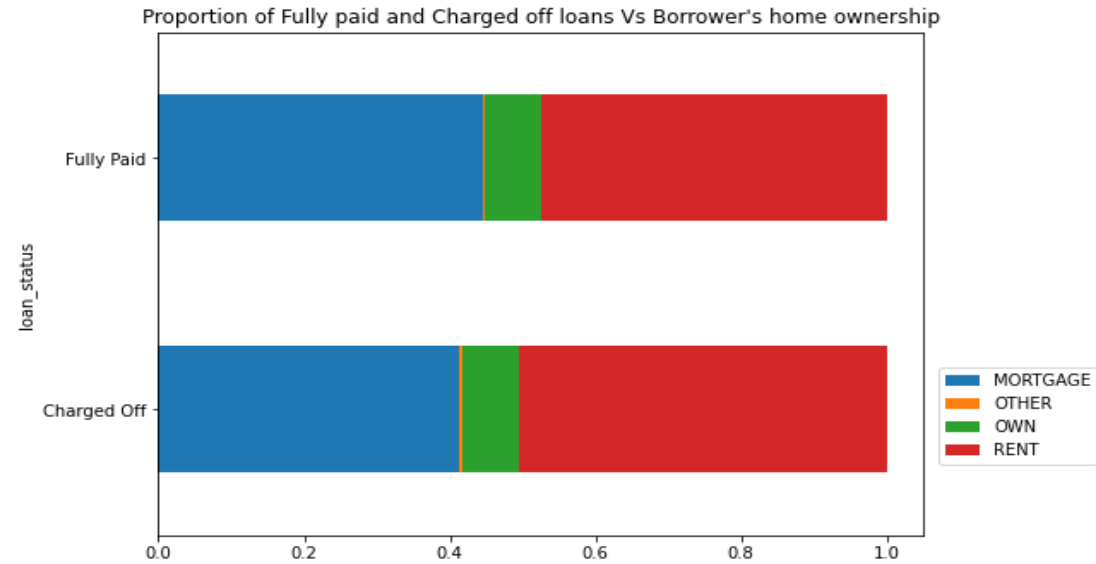


- Insights

- There is no significant difference in proportions of fully paid and charged off loans across different ranges of borrower's employment tenure.

Data Analysis :

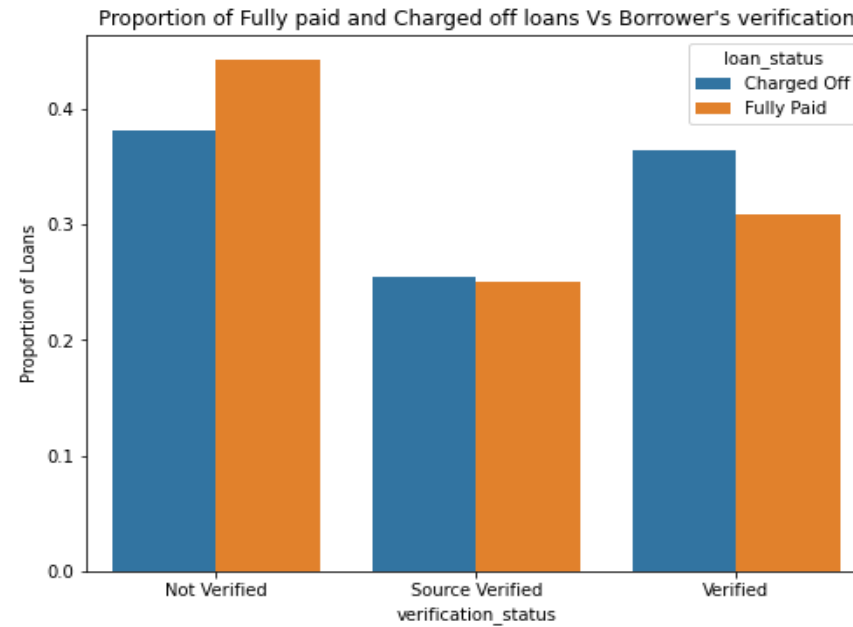
- Relation between ***Home Ownership*** and ***Loan Status***



- Insights
 - Borrowers staying in rented homes are slightly more likely to default.

Data Analysis :

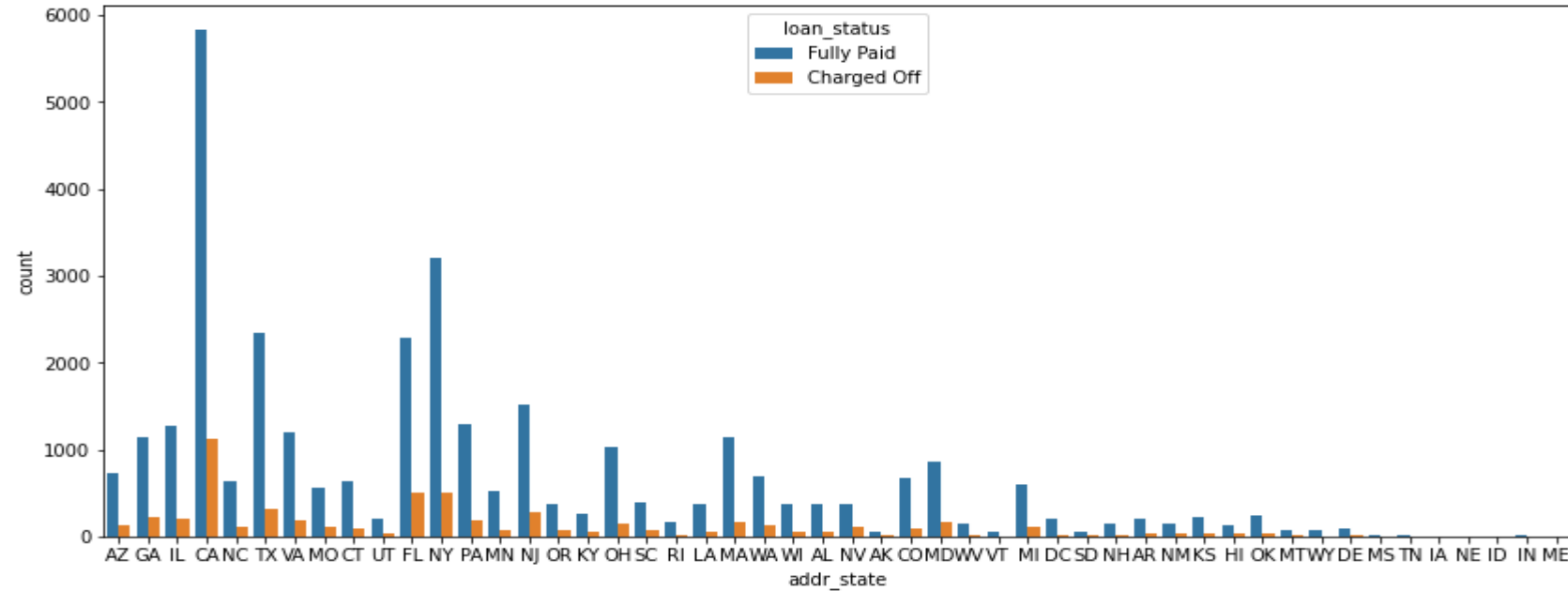
- Relation between ***Verification Status*** and ***Loan Status***



- Insights
 - Verifying the income source details of borrowers reduces the risk of default.

Data Analysis :

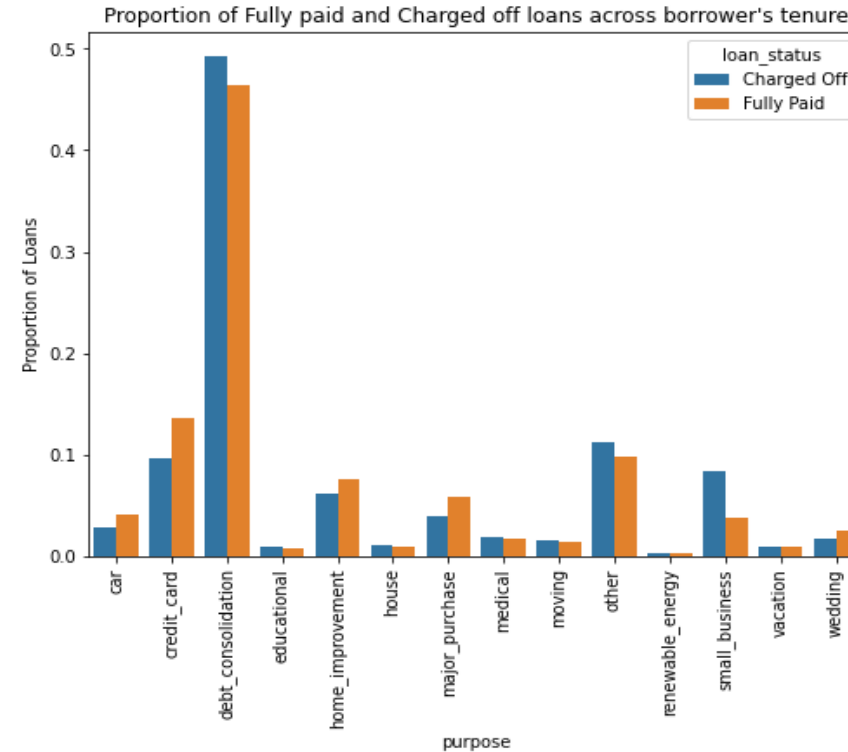
- Relation between ***Borrower's State*** and ***Loan Status***



- Insights
 - More number of charged off loans are associated with high cost of living states like California, Texas, Florida, New York, and New Jersey

Data Analysis :

- Relation between ***Loan Purpose*** and ***Loan Status***

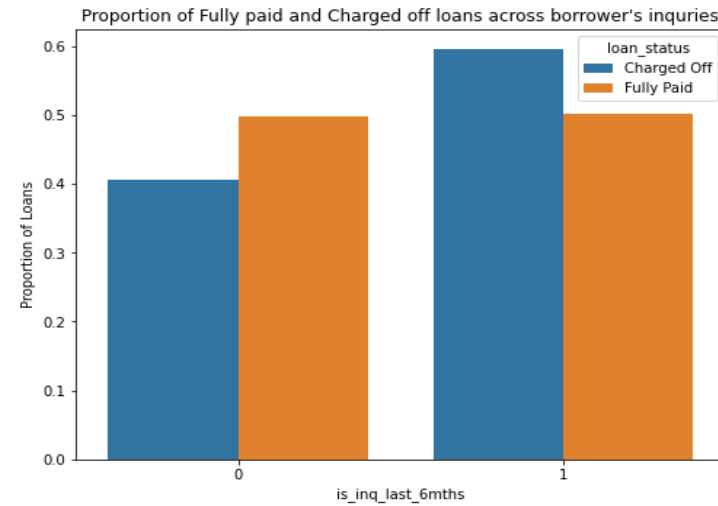
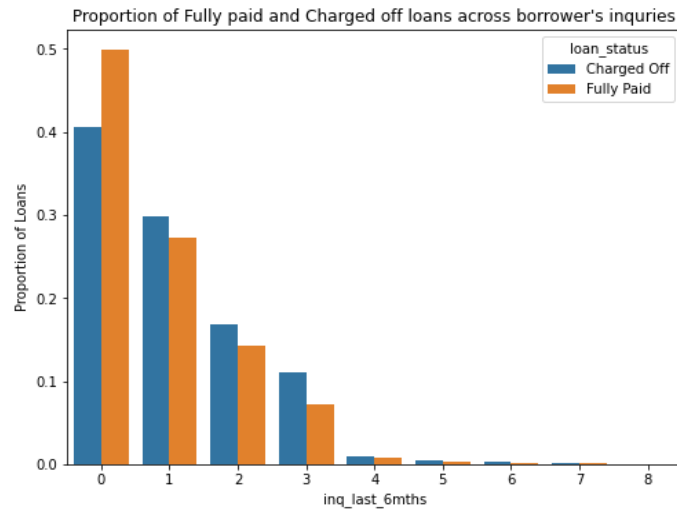


- Insights

- Borrowers taking loans for purpose of debt consolidation, small business and other are more likely to default.
- On other hand, borrowers taking loans for purpose of car, credit card, home improvement, major purchase and wedding are more unlikely to default.

Data Analysis :

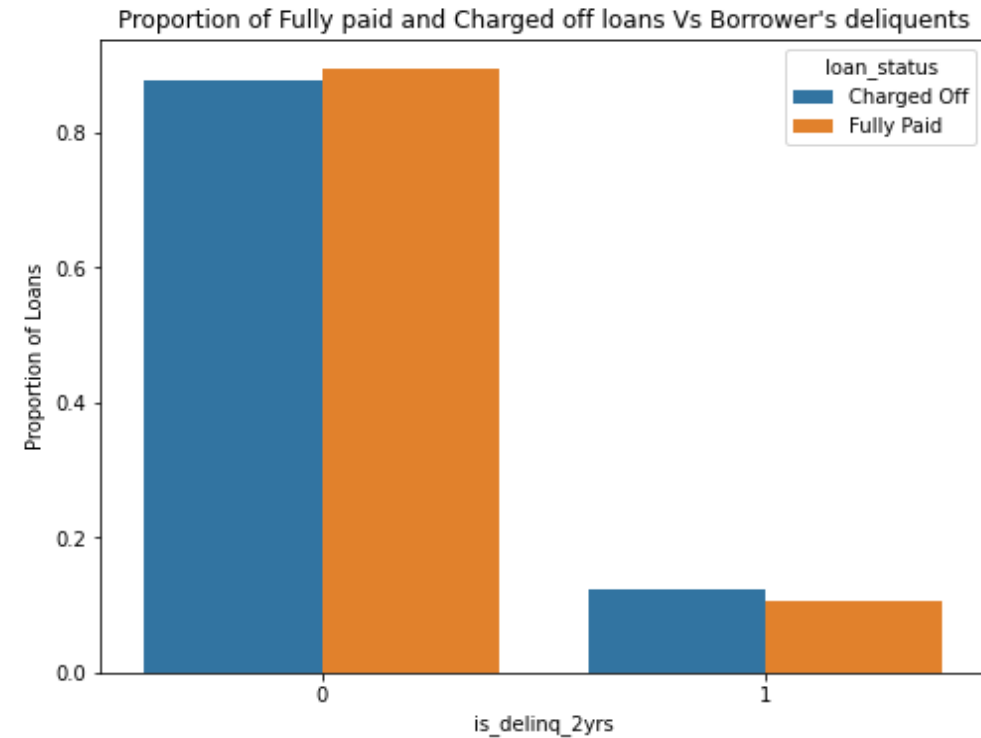
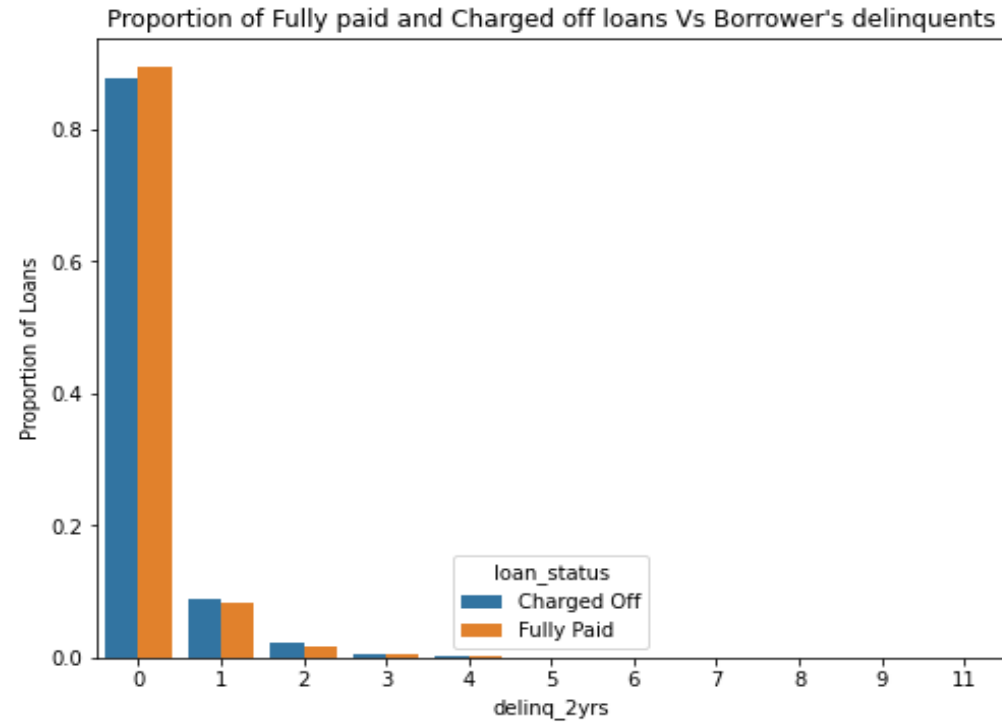
- Relation between ***Borrower's Inquiry*** and ***Loan Status***



- Insights
 - The borrower's who have done any inquiries in last 6 months, about their credit history, are more likely to default.
- Recommendation
 - Look for borrowers who have not done any inquiries in last six months for reducing the risk of default

Data Analysis :

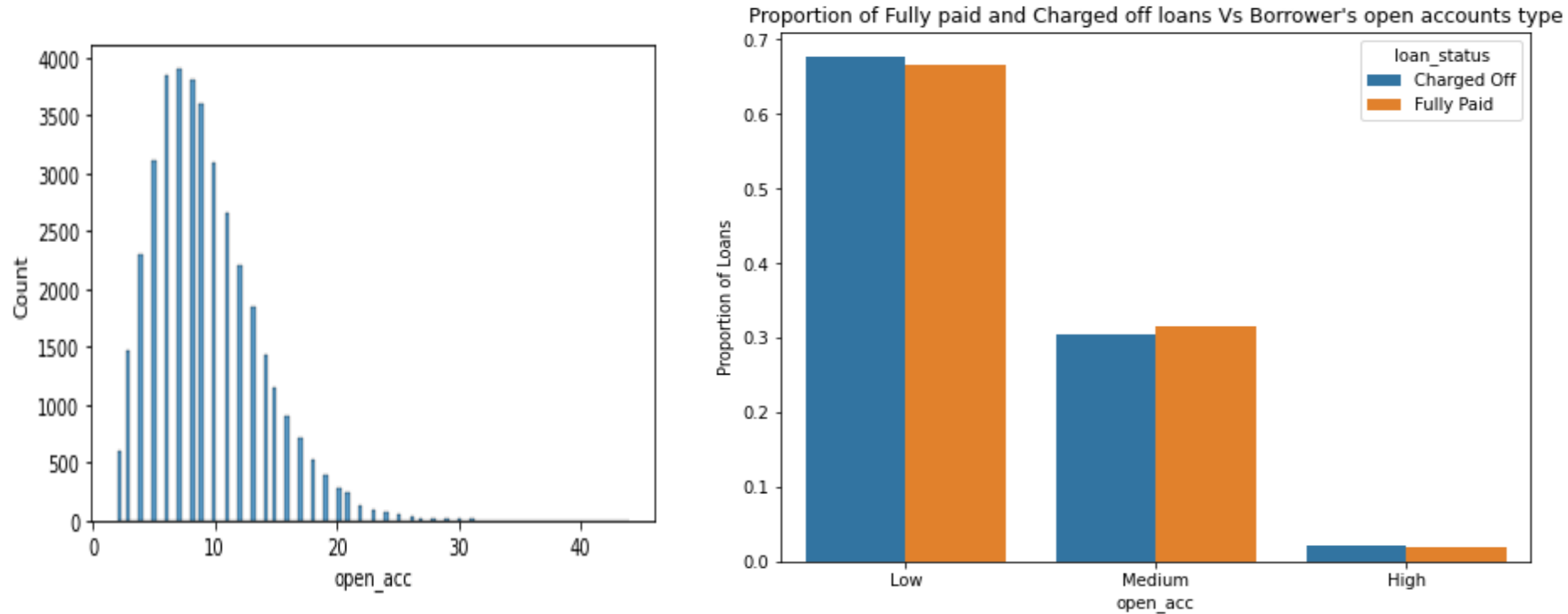
- Relation between ***Borrower's Delinq*** and ***Loan Status***



- Insights
 - The borrower's who were delinquent in last 2 years likely to default. However the difference from non delinquent users is not subtle

Data Analysis :

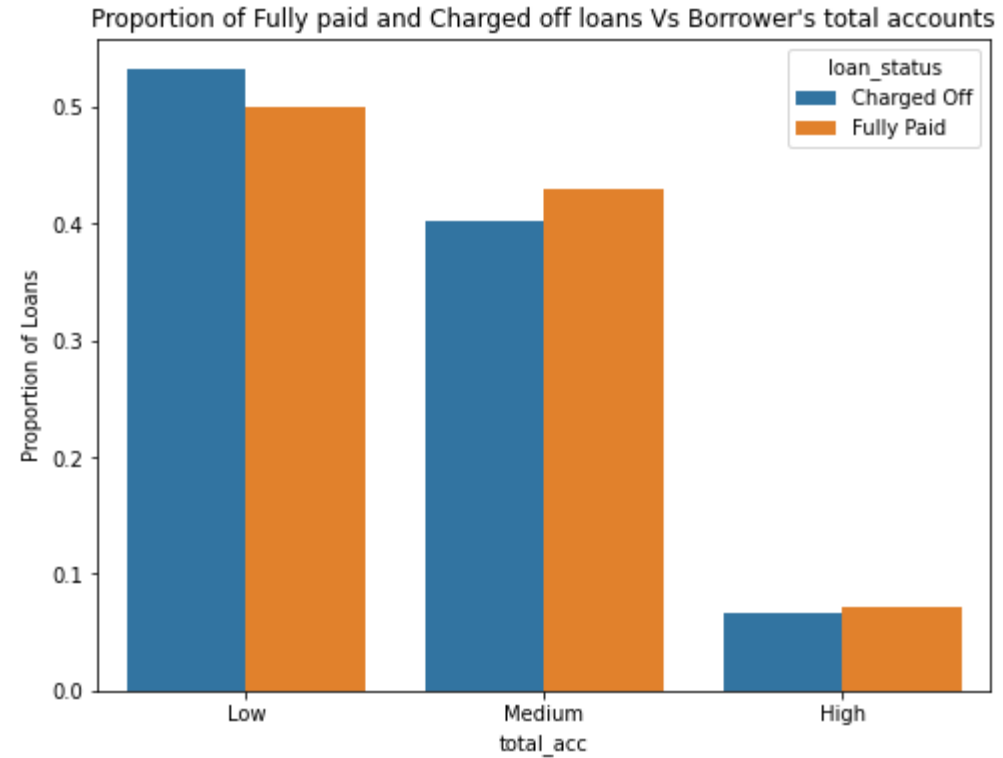
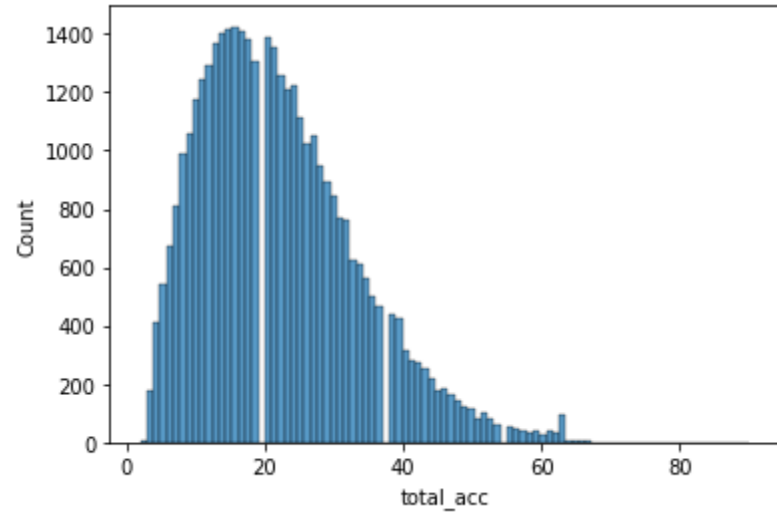
- Relation between ***Borrower's Open Credit Accounts*** and ***Loan Status***



- Insights
 - The proportion of charged off and fully paid loans across three different ranges of open accounts is almost same

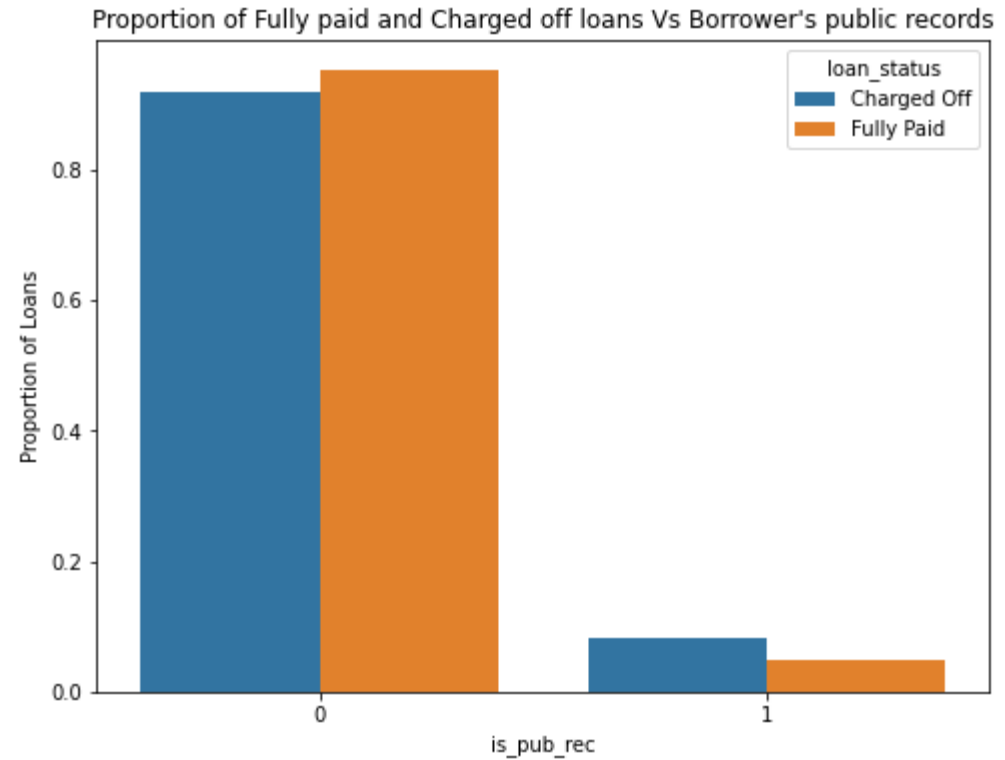
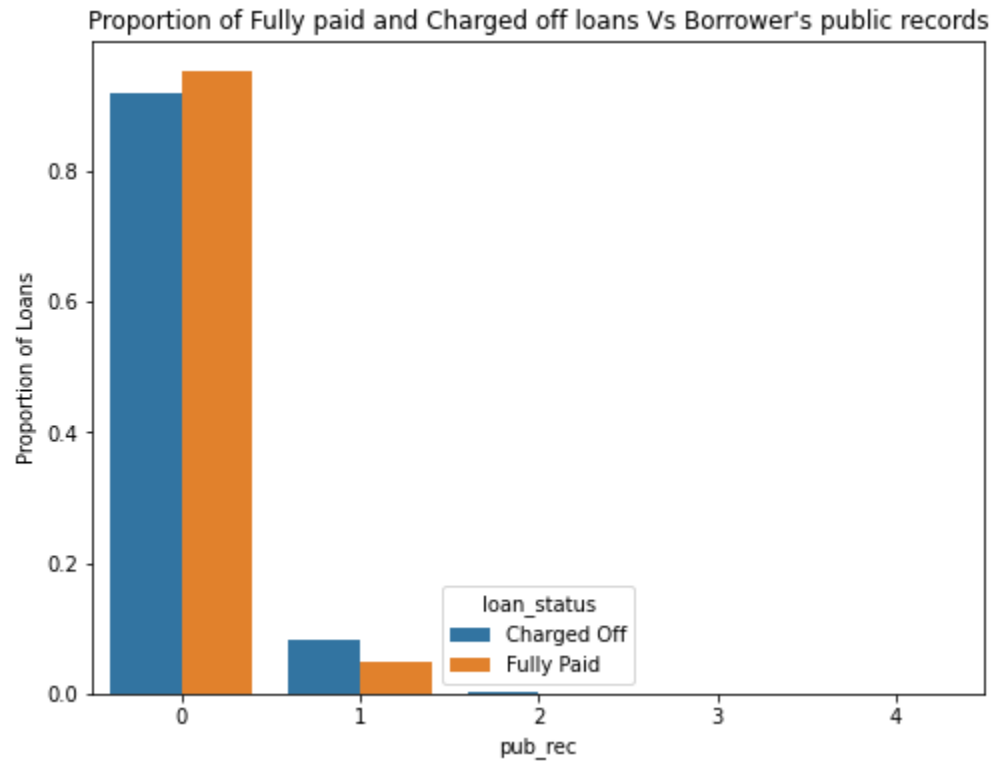
Data Analysis :

- Relation between ***Borrower's Total Records*** and ***Loan Status***



Data Analysis :

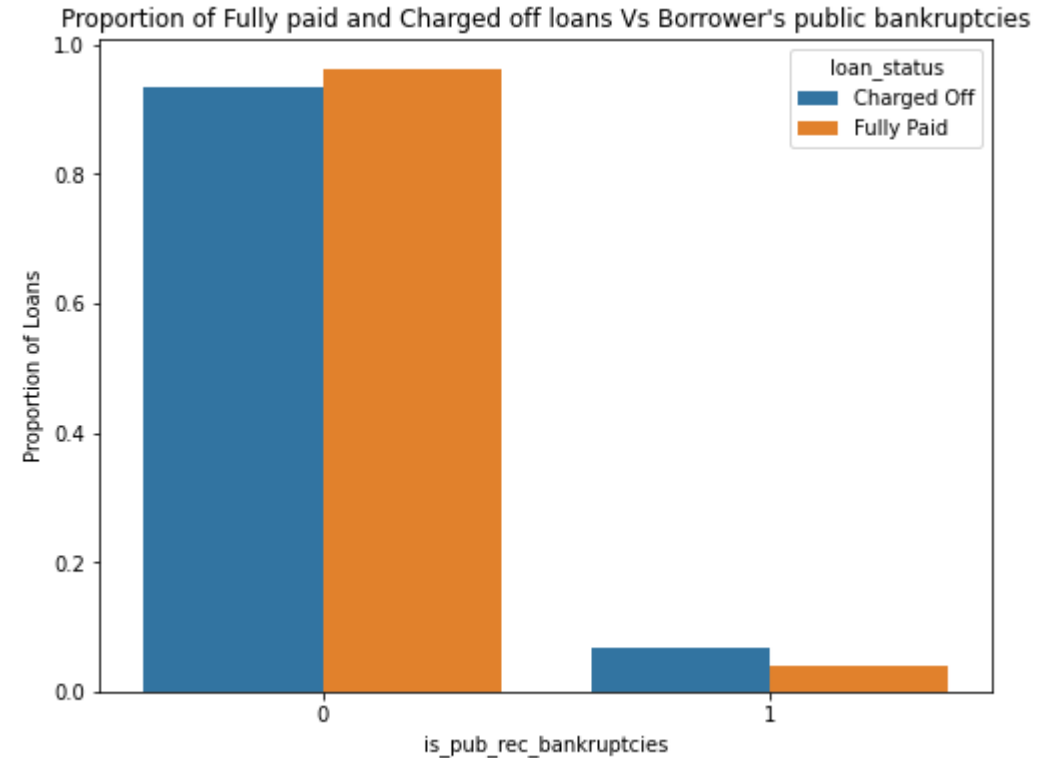
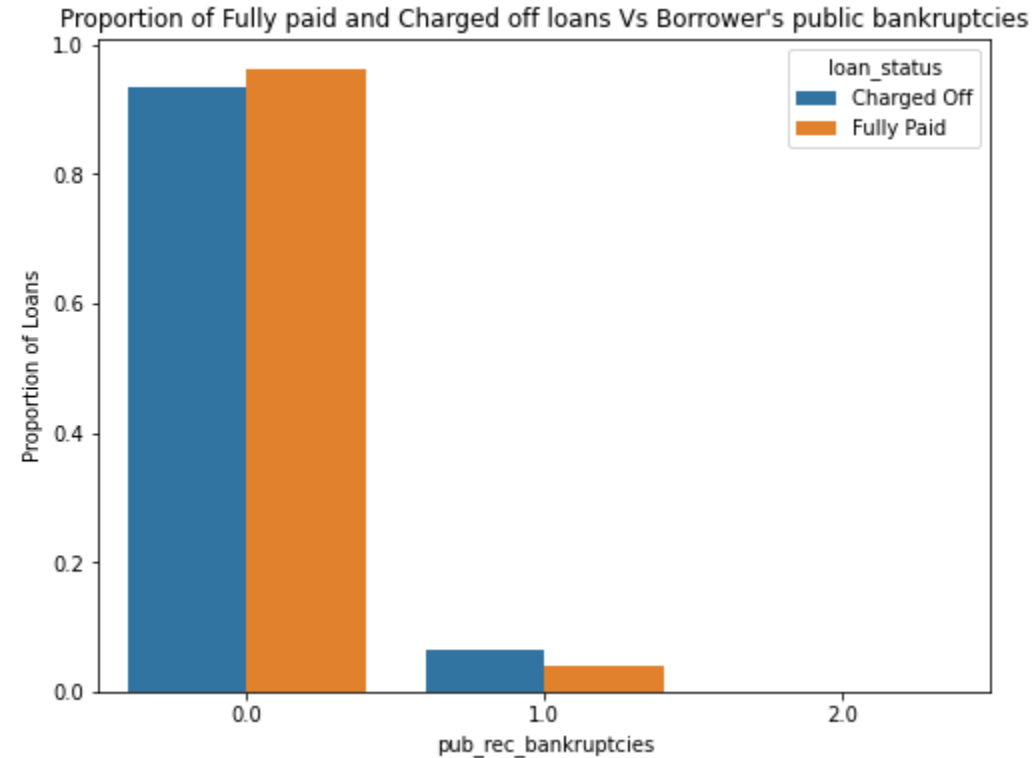
- Relation between ***Borrower's Public Records*** and ***Loan Status***



- Insights
 - Loan is charged off if there is single public record is present
- Recommendations
 - Check for any public record present before approving loan as the chances of loan getting charged off are very high

Data Analysis :

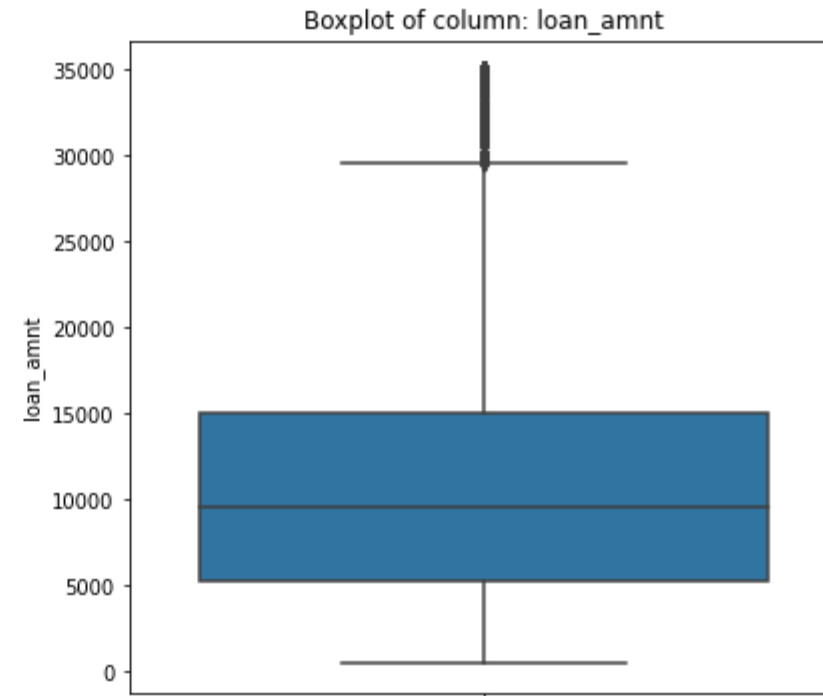
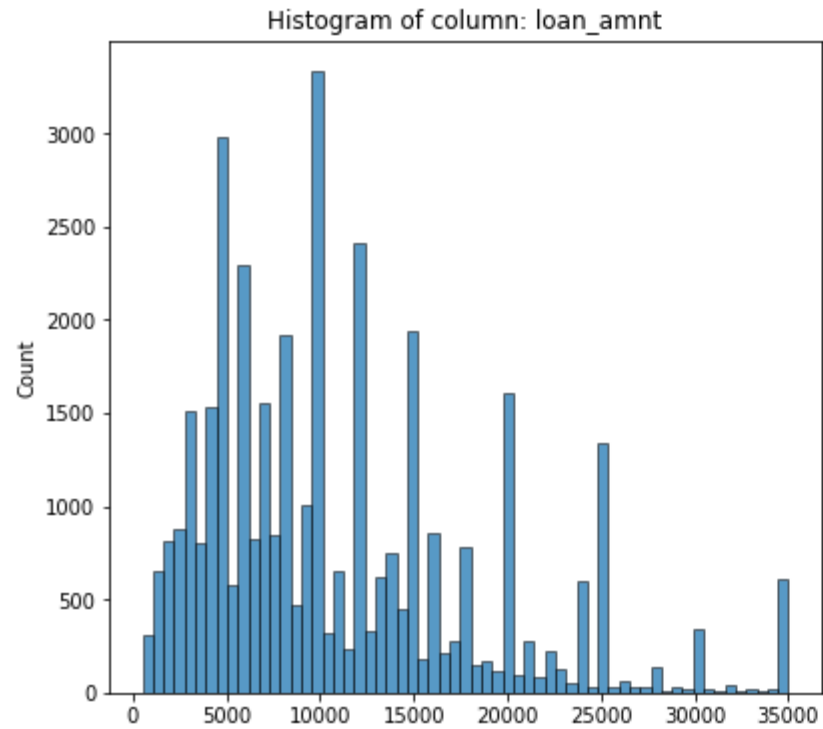
- Relation between ***Borrower's Public Bankruptcies*** and ***Loan Status***



- Insights
 - Fully paid are more when the bankruptcies value is Zero
- Recommendation
 - Don't approve loan for the user having record for bankruptcies

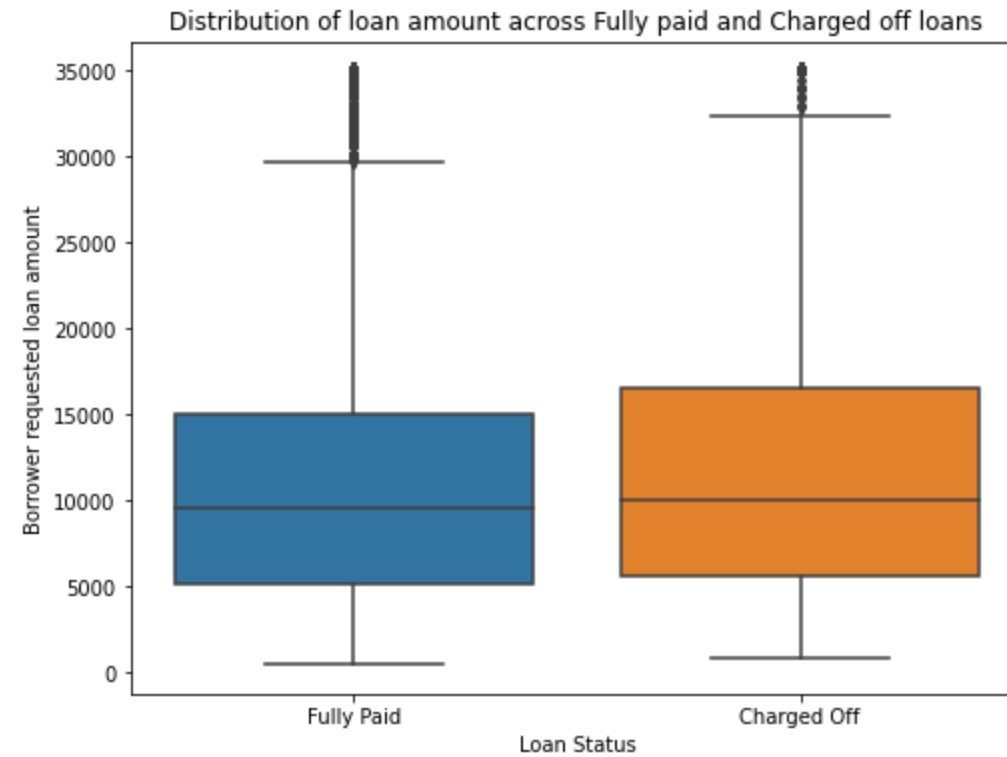
Data Analysis :

- Relation between ***Loan Amount*** and ***Loan Status***



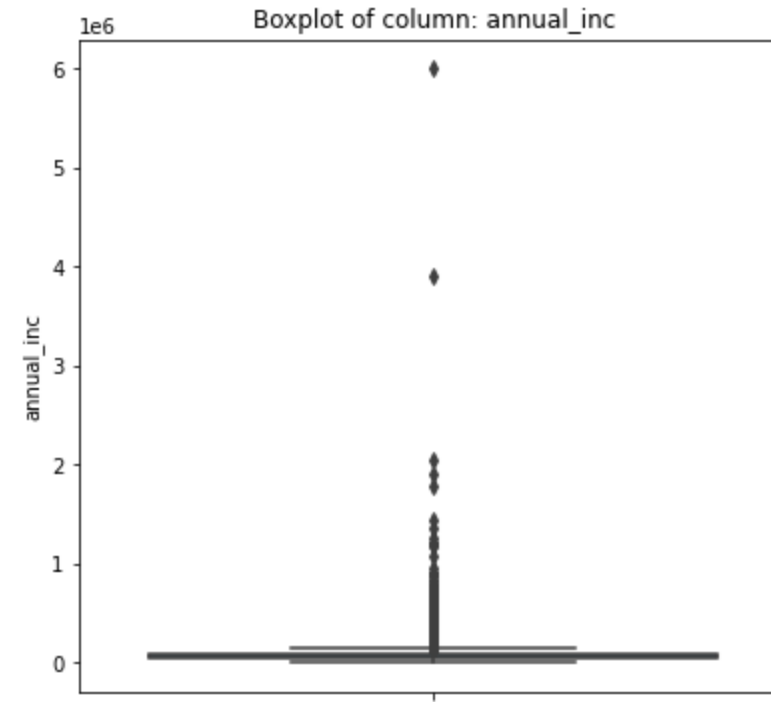
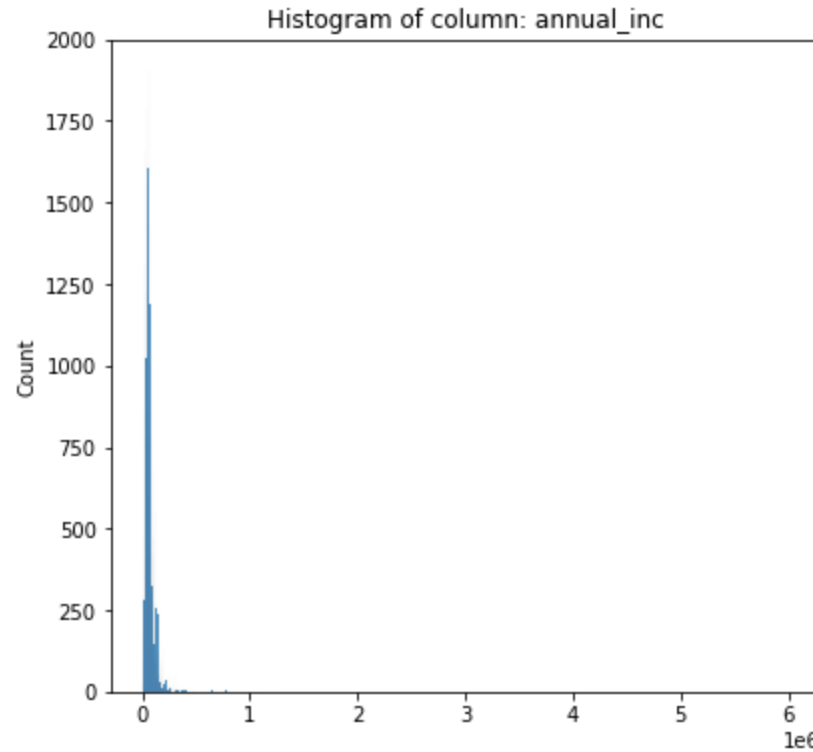
Data Analysis :

- Relation between ***Loan Amount*** and ***Loan Status***



Data Analysis :

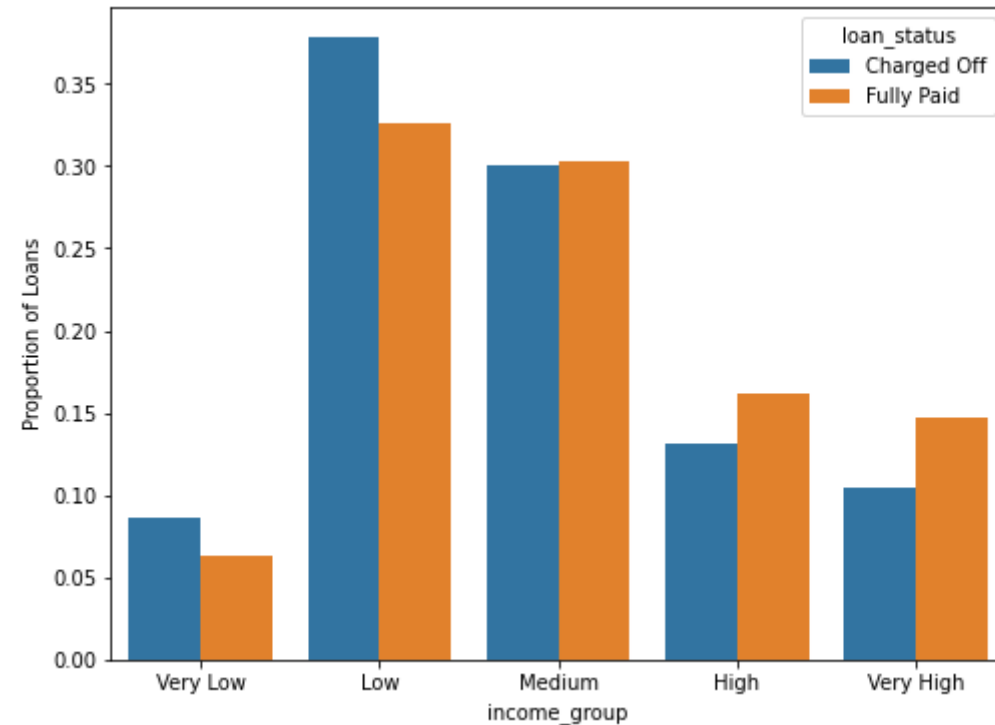
- Relation between ***Annual Income*** and ***Loan Status***



- Insights
 - Annual income of borrowers is highly skewed towards lower income values.
- Recommendation
 - Bin annual incomes of borrowers into various income groups and further analyze the distribution of loans within each income group.

Data Analysis :

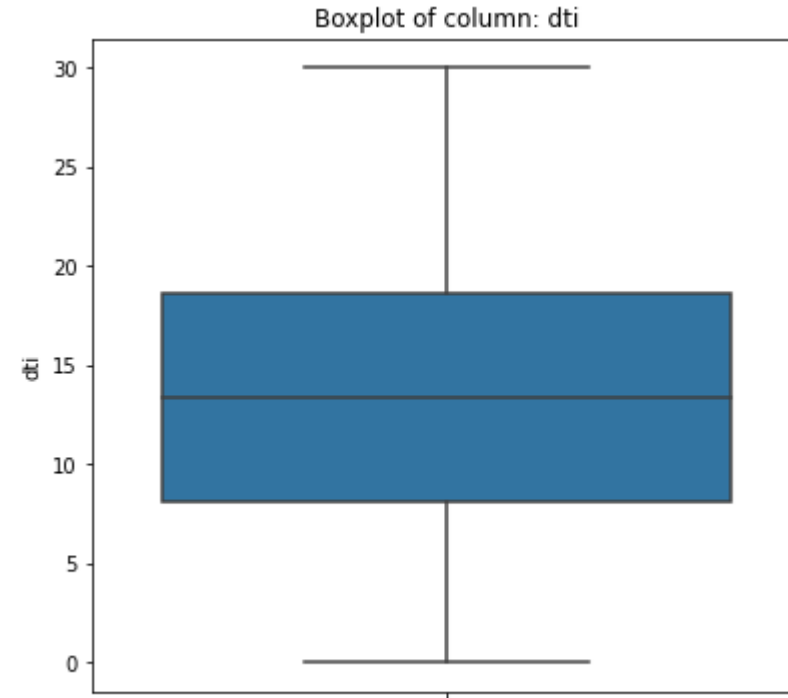
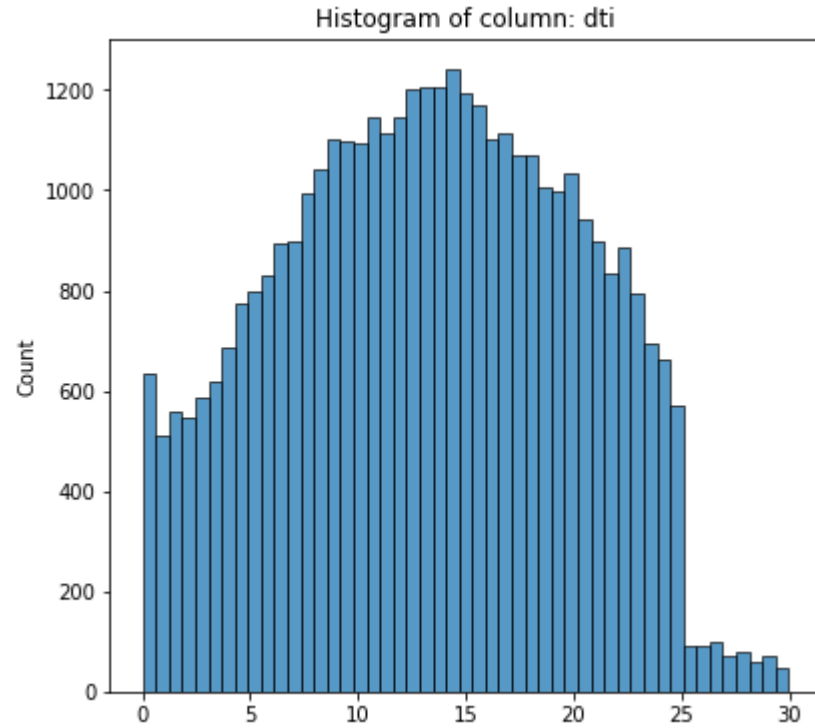
- Relation between **Annual Income** and **Loan Status**



- Insights
 - Borrowers belonging to **Very Low** and **Low** income groups ($\text{annual_inc} \leq 50000$) are more likely to get default.
 - On other hand, borrowers of **High** and **Very High** income groups ($\text{annual_inc} > 75000$) are less likely to default.
- Recommendations
 - More defaulters from the low annual income group

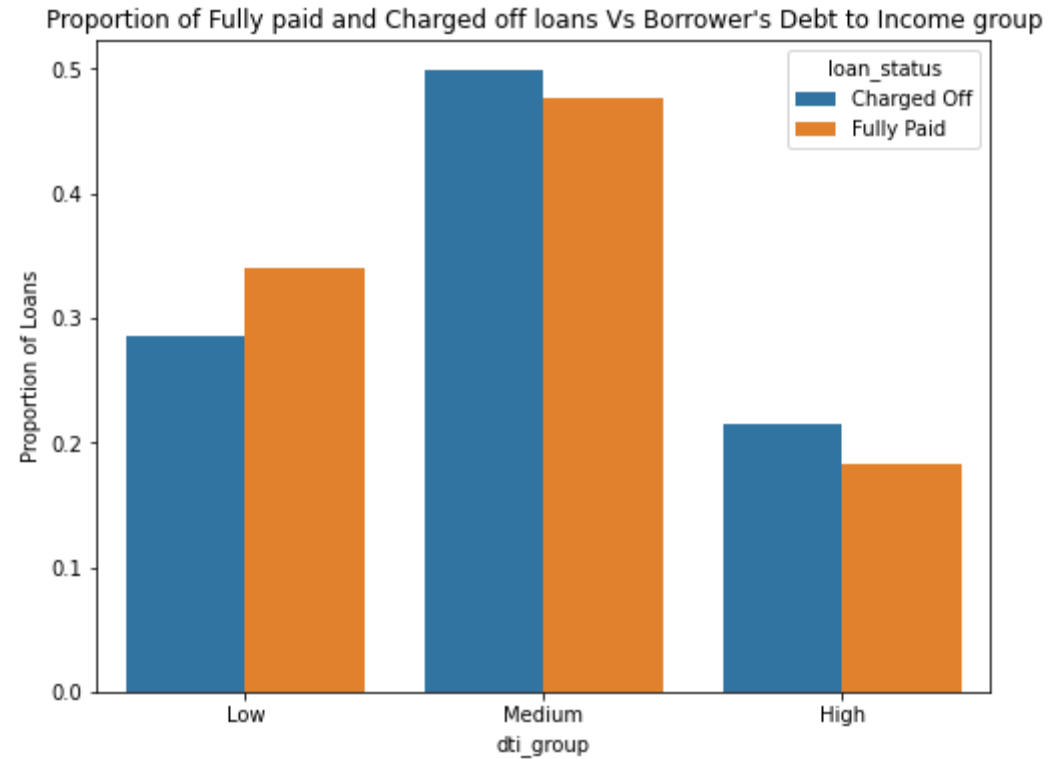
Data Analysis :

- Relation between ***Debt to Income*** and ***Loan Status***



Data Analysis :

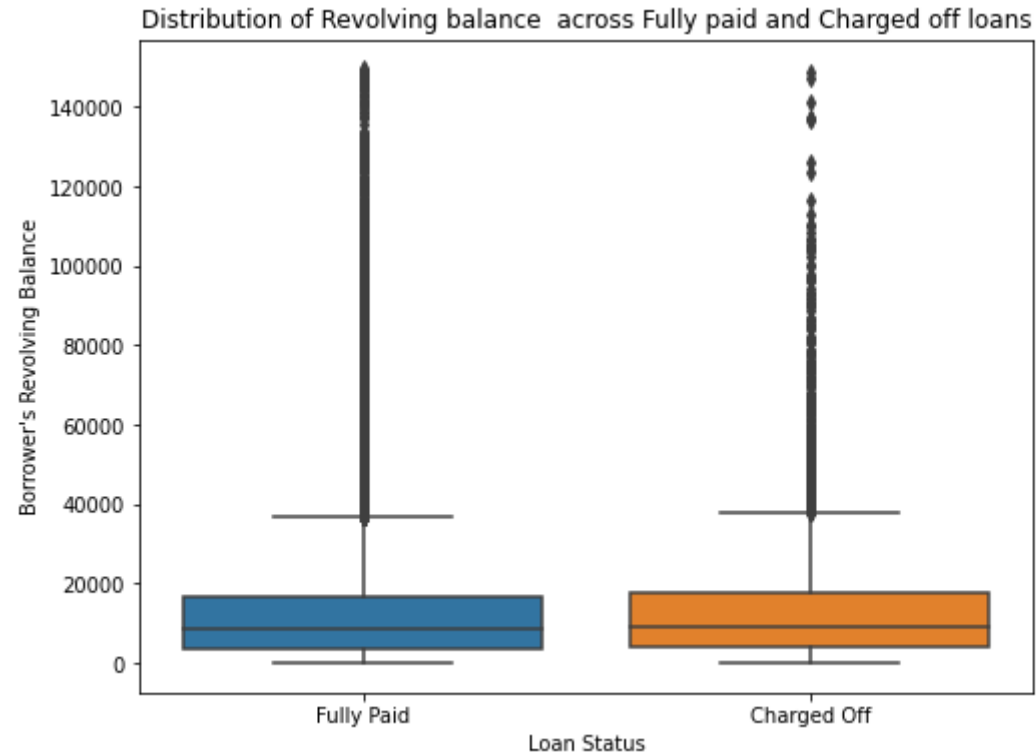
- Relation between **Debt to Income** and **Loan Status**



- Insights
 - Borrowers having **Low** debt to income group ($dti \leq 10$) are less likely to get default.
 - On other hand, borrowers having **High** debt to income group ($dti > 20$) are more likely to default.
- Recommendation

Data Analysis :

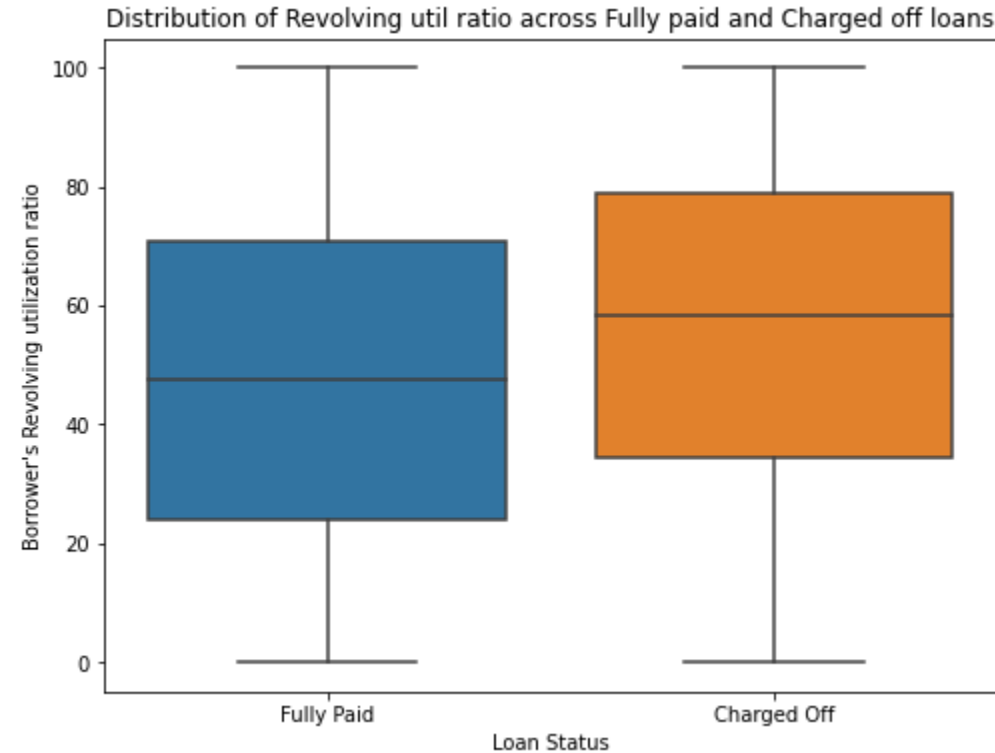
- Relation between ***Revolving Balance*** and ***Loan Status***



- Insights
 - Borrowers revolving balance is also highly skewed towards lower income values
- Recommendation
 - Bin revolving balance of borrowers into various groups and further analyze the distribution of loans within each group.

Data Analysis :

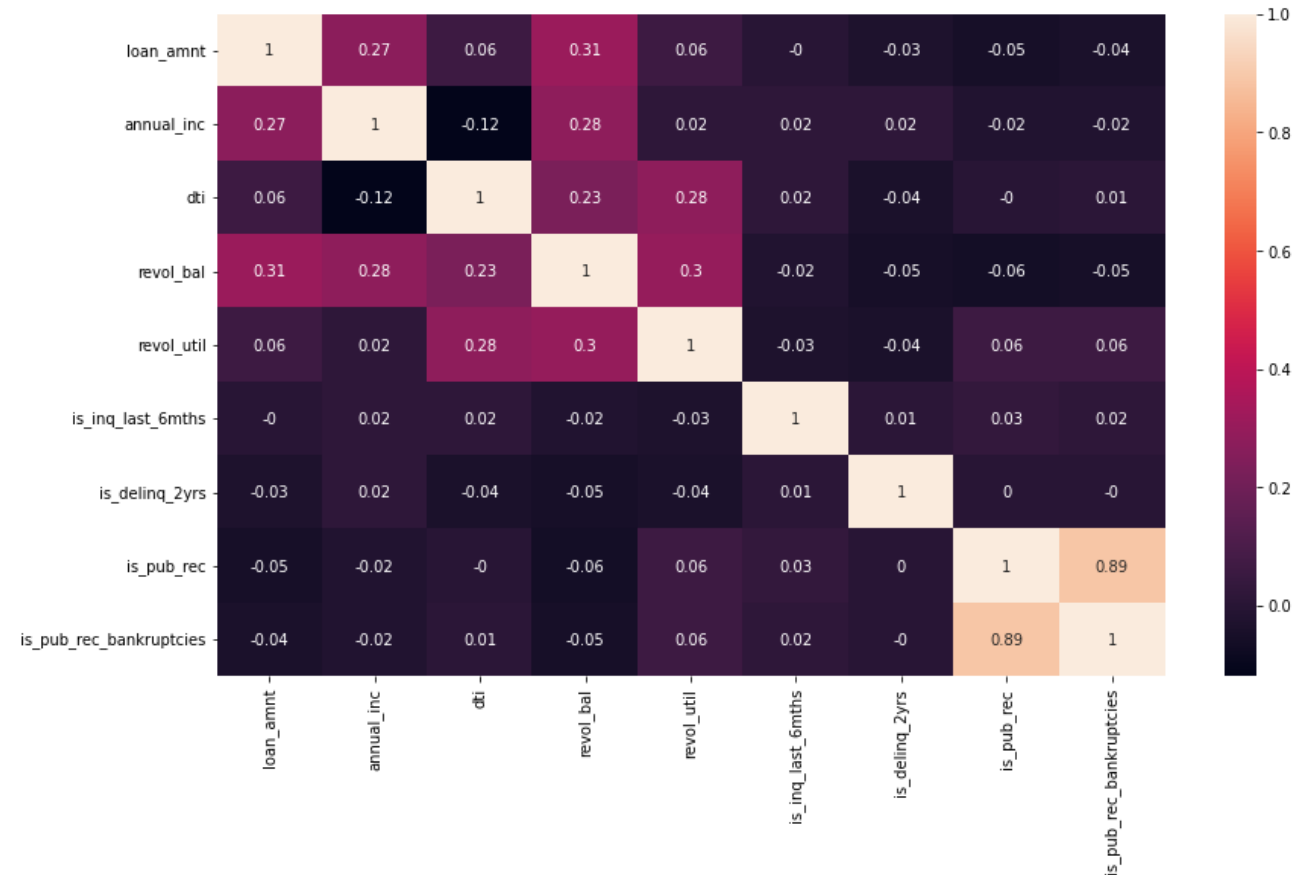
- Relation between ***Revolving Utilization*** and ***Loan Status***



- Insights
 - There is a significant difference in distributions of borrower's revolving utilization ratio.
 - A higher revolving ratio ($> 60\%$) indicates higher risk of default.
 - A lower revolving ratio ($< 40\%$) reduces risk of default.
- Recommendation
 - Target borrowers with lower revolving ratio for reducing risk of default

Data Analysis :

- Relation correlation across various numerical variables:



- Insights
 - The variables representing number of public records and number of public bankruptcies of a borrower are highly correlated (correlation = 0.89)
- Recommendation
 - Due to presence of high correlation we can either drop one of the variables or derive a ratio of these two variables.

Recommendations :

- Strong indicators
 - Grade
 - Borrowers with grade 'A' are highly unlikely to default. Targeting these borrowers will reduce the risk.
 - Borrowers with grade 'C' and above are more likely to default.
 - Revolving utilization ratio
 - Target borrowers with lower revolving ratio for reducing risk of default
 - is_inq_last_6mths
 - Look for borrowers who have not done any inquiries in last six months for reducing the risk of default
 - Debt to income group
 - Borrowers having Low debt to income group ($dti \leq 10$) are less likely to get default.
 - On other hand, borrowers having High debt to income group ($dti > 20$) are more likely to default.
 - Annual income group
 - More defaulters from the low annual income group

Recommendations :

- Weak indicators
 - home ownership
 - Borrowers staying in rented homes are slightly more likely to default.
 - verification status
 - Verifying the income source details of borrowers reduces the risk of default.
 - loan purpose
 - Borrowers taking loans for purpose of debt consolidation, small business and other are more likely to default.
 - On other hand, borrowers taking loans for purpose of car, credit card, home improvement, major purchase and wedding are more unlikely to default.
 - borrower's address state
 - More number of charged off loans are associated with high cost of living states like California, Texas, Florida, New York, and New Jersey
 - Revolving balance
 - Bin revolving balance of borrowers into various groups and further analyze the distribution of loans within each group.
 - is_delinq_2yrs
 - The borrower's who were delinquent in last 2 years likely to default. However the difference from non delinquent users is not subtle
 - is_pub_rec
 - Check for any public record present before approving loan as the chances of loan getting charged off are very high

Thank You