Lending Club Casestudy

Problem Statement:

- To identify patterns which indicates if the loan applicant is likely to default or not by using EDA.
- To identify the variables which are strong indicators of loan default.

Objectives:

- Understand the Lending Club Loan data and find insights.
- To identify whether applicant is default or not using EDA, to avoid financial loss.
- To understand driving factors which indicates loan to be default.

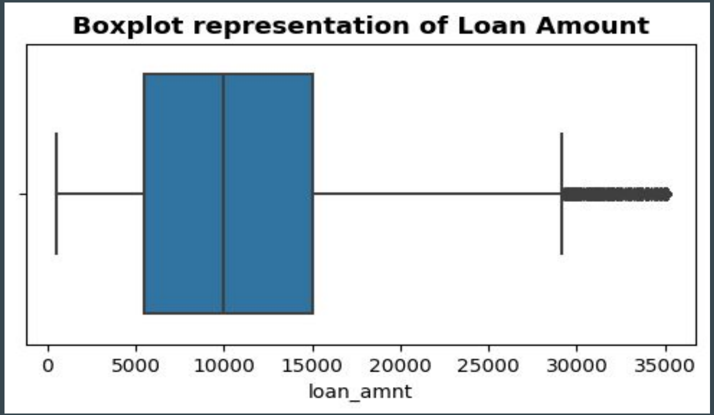
EDA Approach:

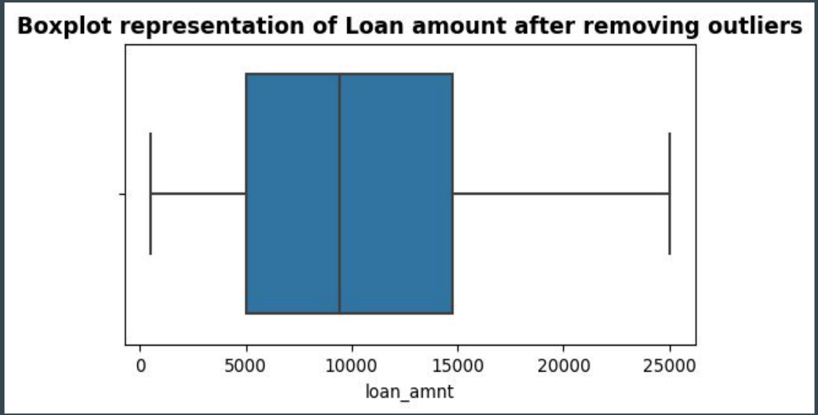
Data Cleaning and Manipulation:

- The dataset was having 57 null columns which were dropped off.
- Also the columns containing single value such as ‘0’ or ’INDIVIDUAL’, were also removed.
- The missing or null values in the column were imputed.
- The data redundancies were removed and columns such as term, int_rate and zip_code were manipulated and converted to their respective data types.

Outlier Treatment:

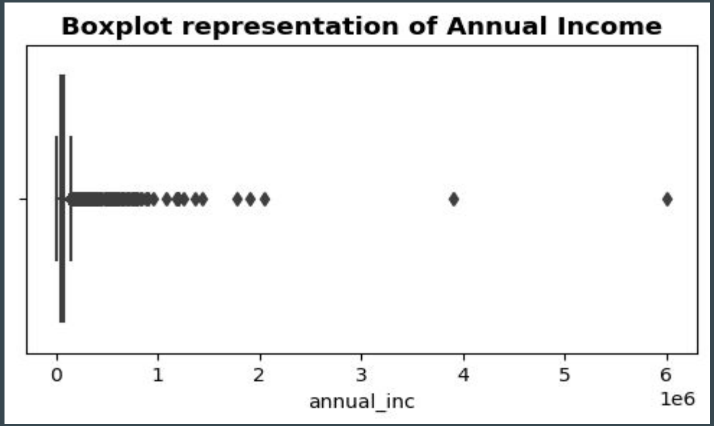
1. Loan Amount:

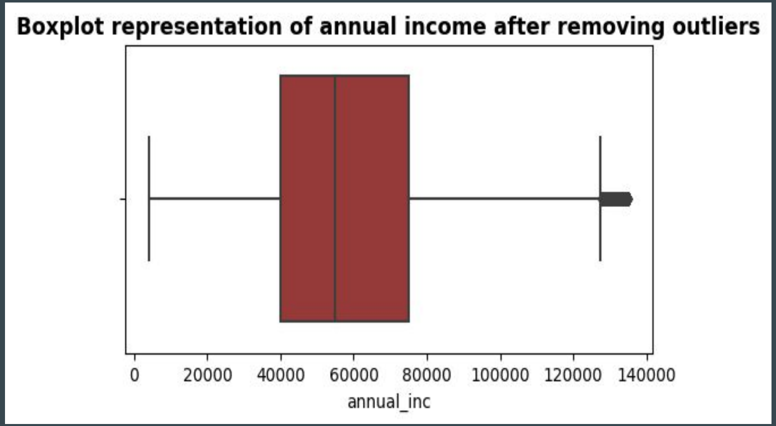




- Very few borrowers have loan amounts that exceed more than 25000, these data points are referred to as an outlier.
- These outliers are removed by applying outlier treatment, which improves the accuracy.

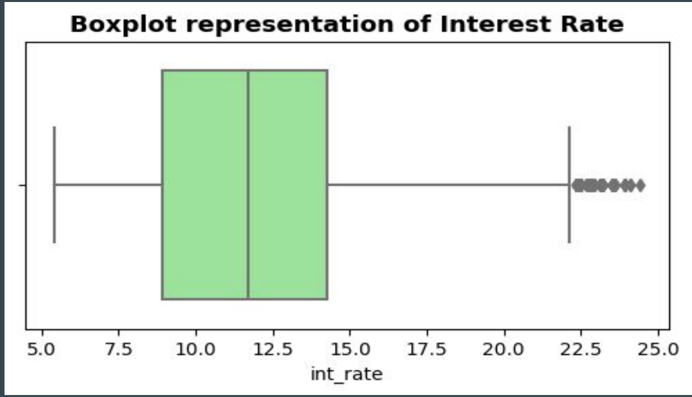
2. Annual Income:

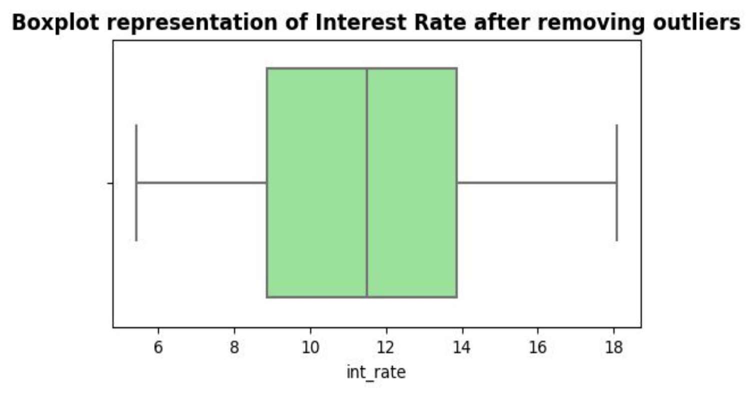




- Some borrowers have significantly high income amounts i.e. more than 100000.
- Due to their extreme values, the data points can reduce the accuracy and hence, they are removed by applying the outlier treatment.

3. Interest Rate:

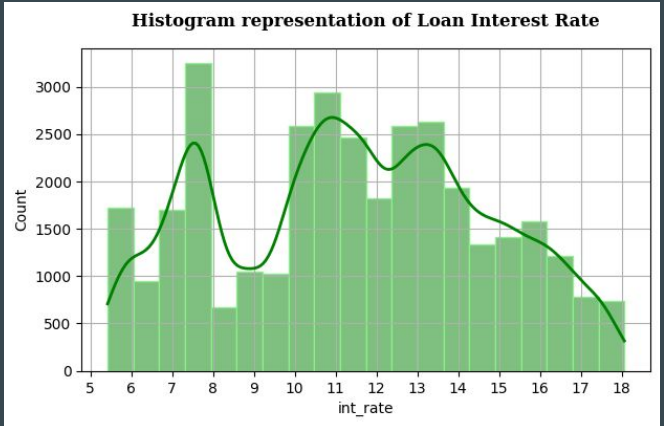




- Loans with significantly high interest rates certainly produce outliers.
- Applying outlier treatment will remove the outliers. This will improve the accuracy of the model.

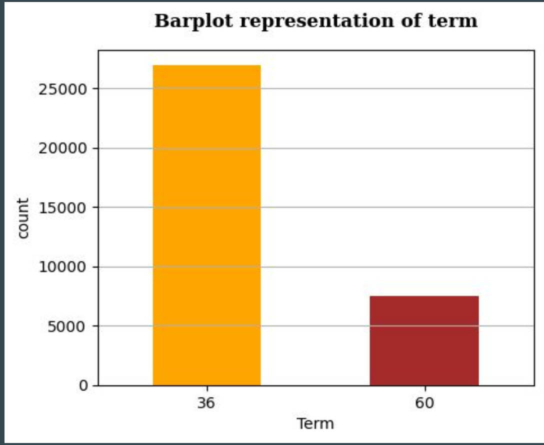
Univariate Analysis:

1. Interest Rate:



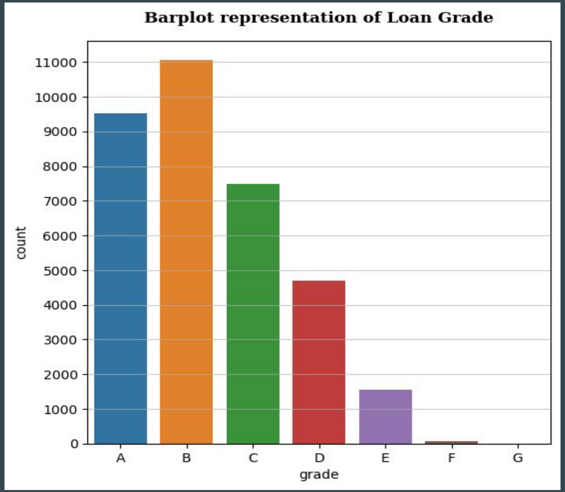
- The count of borrowers is high between 7%-8% and 10%-14%.
- The higher the interest rate more likely the borrower is to default.

2. Term:



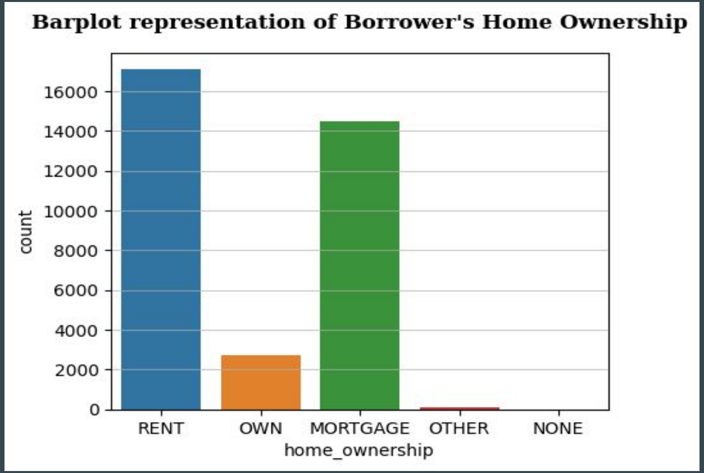
- The borrowers count who have borrowed loans with a term of 36 months is significantly high.

3. Grade:



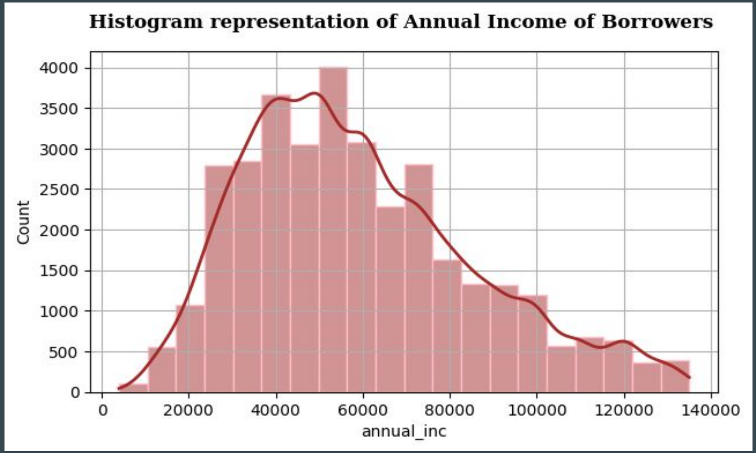
- Grade A loans have borrower count between 9000 -10000.
- Grade B loans have a higher count of borrowers, i.e. around 11000.
- Loan applicants with grade A and B are likely to be approved.

4. Home Ownership:



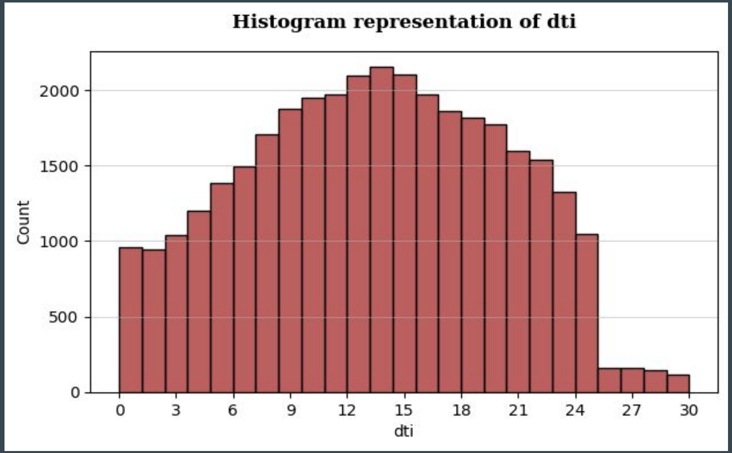
- Borrowers with Rent or Mortgage home have higher chances of loan approval.

5. Annual Income:



- The borrower count is high for annual income ranging between 35000 - 60000.
- So, applicant having annual income between 35k - 60k have high possibility of loan approval.

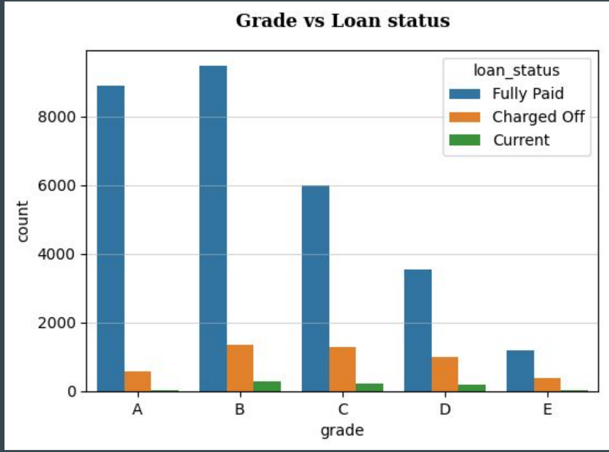
6. Debt to Income Ratio:



- The number of borrowers with an income range between 35000-65000 is quite high.

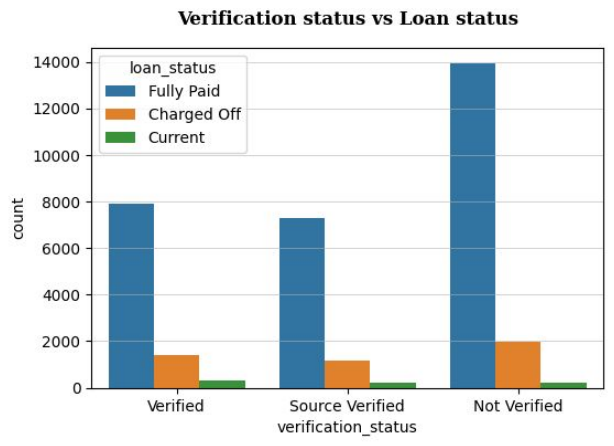
Bivariate Analysis:

1. Grade vs Loan Status:



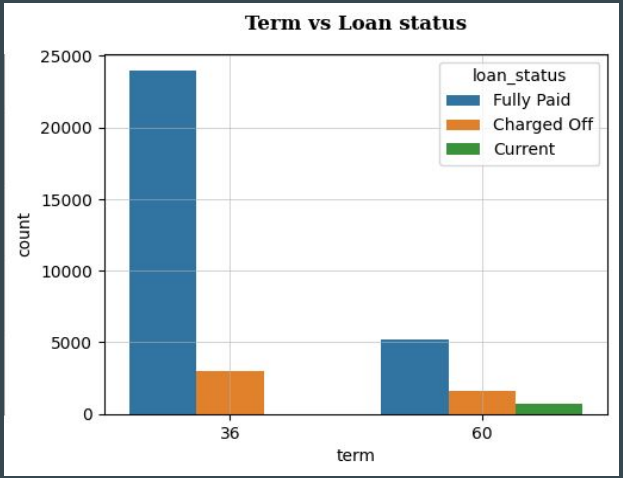
- Grade A and B have a higher count of 'Fully Paid' and 'Charged Off' borrowers .
- So, it can be inferred that borrowers with loan grades A and B are less likely to default.

2. Verification Status vs Loan Status:



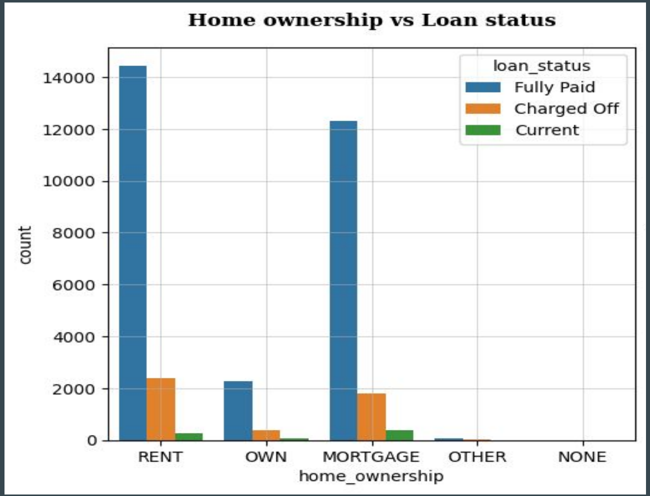
- Fully Paid and Charged Off borrowers count is higher for those with unverified income sources compared to verified or source verified borrowers.

3. Term vs Loan Status:



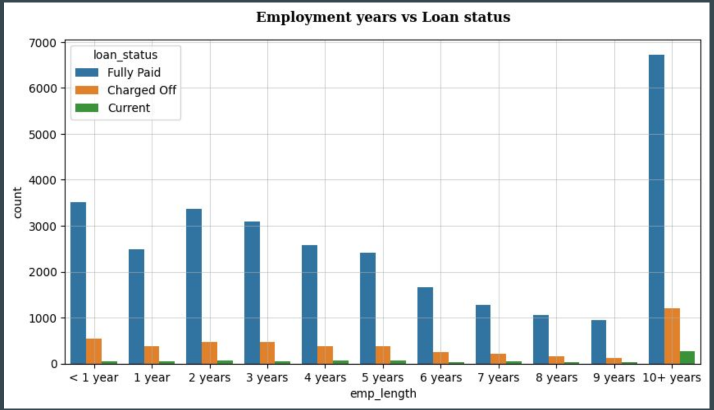
- There are high count of borrowers who have fully paid the loan amount over a period of 36 months.

4. Home Ownership vs Loan Status:



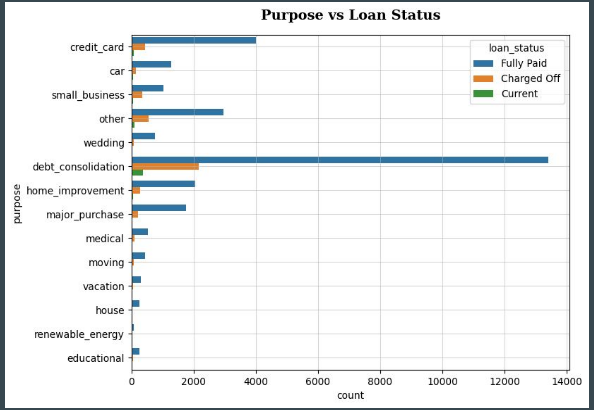
- The number of Fully paid borrowers for Rent and Mortgage are higher than Charged-Off.

5. Employment years vs Loan Status:



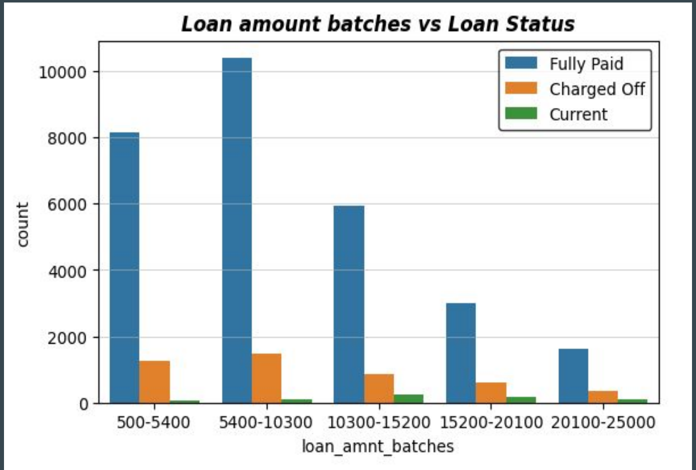
- Employment years of 10+ years has high count of borrowers. Also, it has a high count of Fully Paid borrowers.
- Hence, applicants with 10+ employment years have a higher chance of loan approval.

6. Purpose vs Loan Status:



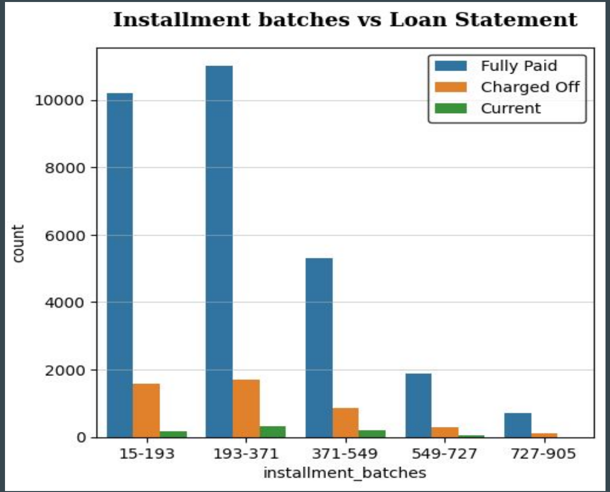
- Debt Consolidation has a high count of loan borrowers.
- Also, the count of Charged Off is significantly lower compared to Fully Paid borrowers.

7. Loan Amount vs Loan Status:



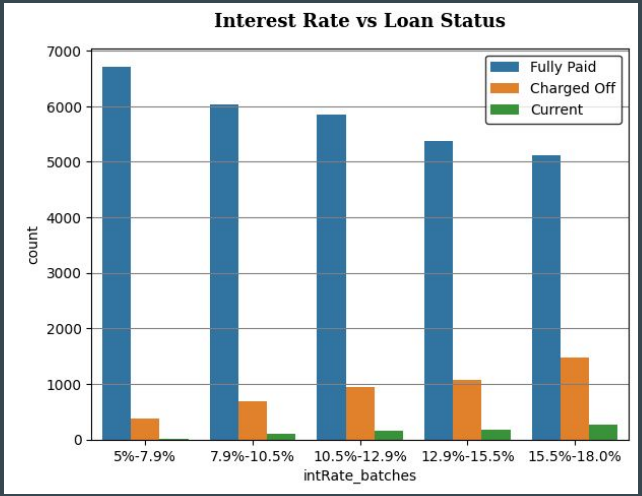
- Loans with loan amounts ranging from 500 - 10300 have a high count of borrowers.

8. Installment batches vs Loan Status:



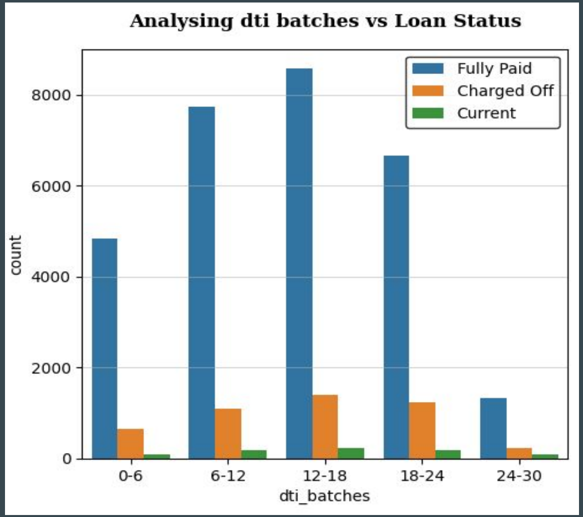
- The majority of borrowers have loan installments between 193 to 371 and the loan amount is fully paid.
- The count of Fully paid is comparatively lower than that of Defaulters.

9. Interest Rate vs Loan Status:



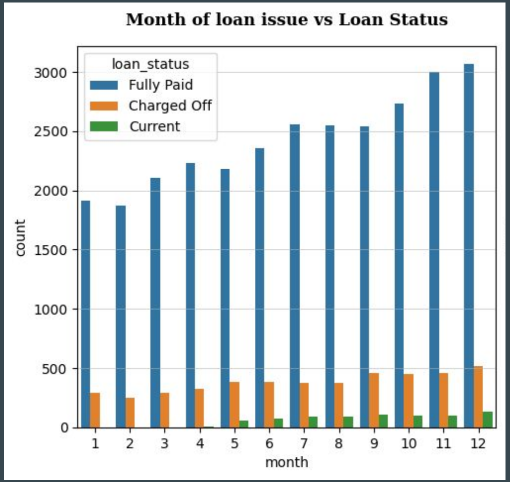
- While analysing Interest rate vs Loan status, as the interest rate increases the count of Fully paid decreases and Charged Off increases.

10. DTI batches vs Loan Status:



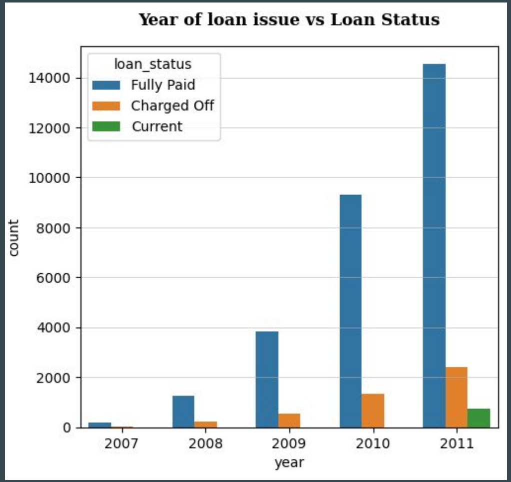
- The dti value between 6-12 has a significantly high count of borrowers with fully paid the loans.

11. Loan Issued Month vs Loan Status:



- The borrowers count is high for November and December.
- Also, the rate of both Fully Paid and Charged Off is high but the defaulters count is far less than the Fully paid count.

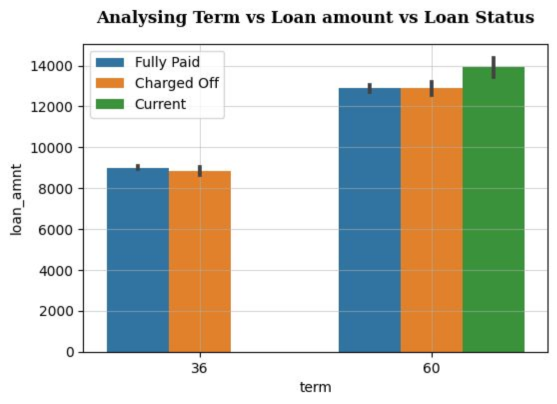
12. Loan Issued Year vs Loan Status:



- There is significant growth in the count of borrowers from 2007 - 2011.
- As the number of borrowers increases, both Fully Paid and Defaulters count increases, but the count of defaulters is much lower.

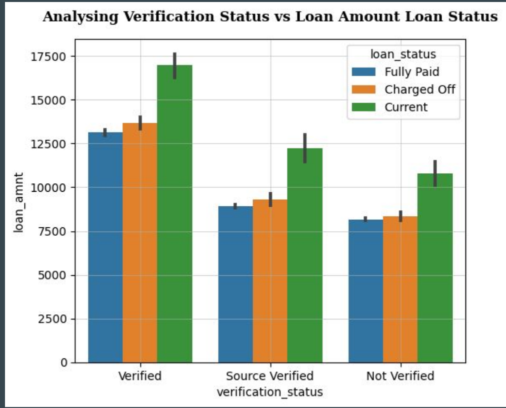
Segmented Univariate Analysis:

1. Term vs Loan Amount vs Loan Status:



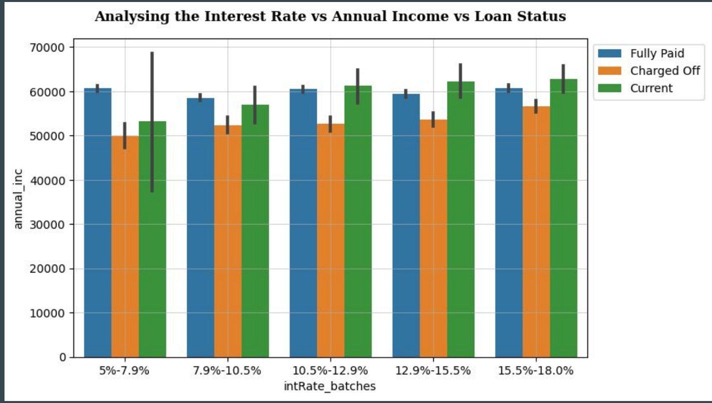
- For lower loan amounts the term is of 36 months.
- For higher loan amounts the term is of 60 months.
- The loan amount range for Fully Paid and Defaulters are same.

2. Verification Status vs Loan Amount vs Loan Status:



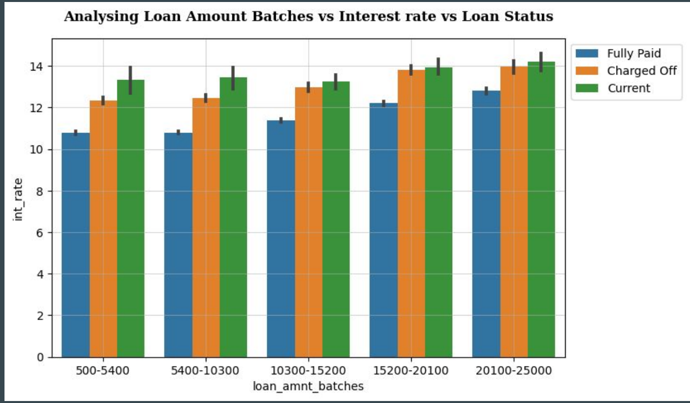
- The loan amount range for Fully Paid and Charged Off borrowers is similar for all categories.
- However, the Charged Off borrowers have a slightly higher loan amount.

3. Interest Rate vs Annual Income vs Loan Status:



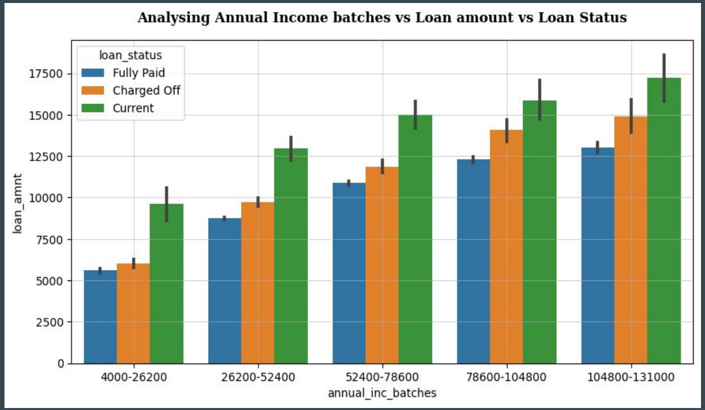
- Fully paid borrowers have an annual income range of 55000 to 65000 across all interest rate groups.
- Charged Off borrowers have annual income range between 50000 and 60000 across all interest rate groups.

4. Interest Rate vs Loan Amount vs Loan Status:



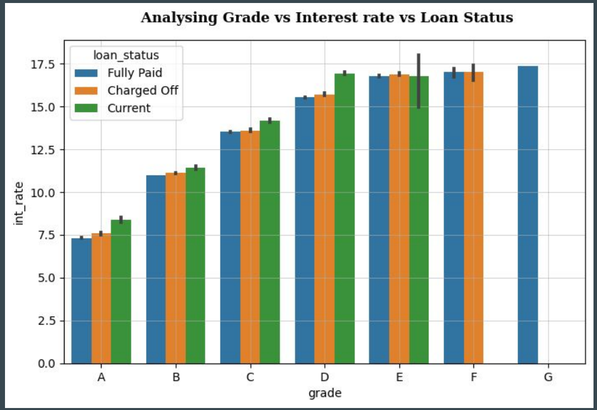
- Fully Paid borrowers have low interest rates compared to Charged Off borrowers.
- If the interest rate is too high compared to the borrower's annual income, there is a higher possibility of loan default.

5. Annual Income vs Loan Amount vs Loan Status:



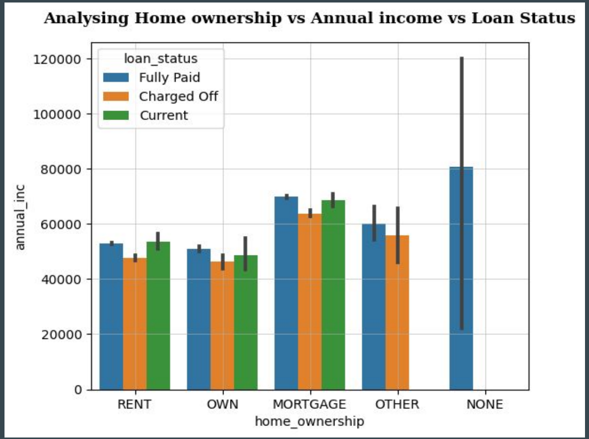
- The count of charged off loans is greater than fully paid loans.
- Loan amount for Charged Off is higher than that of Fully Paid.

6. Grade vs Interest Rate vs Loan Status:



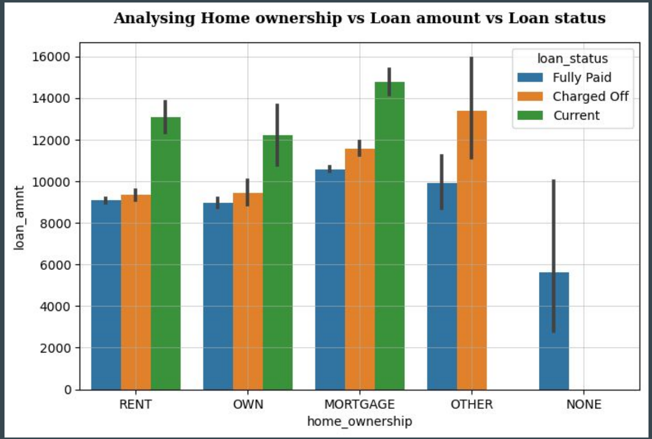
- As the Grade moves from A to G, the interest rate increases.
- The range of interest rate is the same for Fully Paid and Defaulters of the same grade

7. Home Ownership vs Annual Income vs Loan Status:



- Fully Paid, Current and Default borrowers fall into same annual income range.
- Fully Paid and Defaulted borrowers having home as Mortgage have high income i.e between 60000- 80000.

8. Home Ownership vs Loan Amount vs Loan Status:



- High loan amount increases the loan default risk.
- Borrowers having home ownership other than Rent, Owned or Mortgage and having high loan amount are likely to default.

Conclusion:

1. The count of borrowers for a loan with a term of 36 months is higher.
2. Loan applicants with grades A or B and with subgrades A4, A5, B3, B4, and B5 are likely to be approved.
3. Borrowers with 10+ years of employment have a high chance of getting loan approval.
4. Borrowers with a rental or mortgage home have higher chances of loan approval.
5. The number of borrowers who have taken out loans for debt consolidation is quite high.
6. The borrower count is high for those with open accounts between 6-9 i.e., above 3000.
7. Borrowers having loans with dti from 12 - 15 are higher in the count.
8. The total payment between 5000-7500 has a higher count of borrowers.
9. Applicants with 0 derogatory public records have a higher chance of loan approval.
10. Analyzing the interest rate shows that 5%–7.9% and 10.5%–12.9% have more loan borrowers.
11. The dti ranges from 12–18 and 6–12 have the highest count of fully paid borrowers.
12. The months of November and December had a higher number of loans issued, largely in 2011.

Recommendations:

- Loan of grade A or B with low interest rates and a term of 36 months will reduce the risk of defaulters.
- Applicants who have worked for 10 years or more are less likely to default, while those with 6-9 years of employment have a higher risk.
- The dti i.e. debt-to-income ratio between 16-18 can reduce the risk of default.
- Loans issued in October, November and December are more likely to default.
- Loans for Credit cards, Small businesses, Debt Consolidation and Houses have a low possibility to be default.
- As installments increase relative to annual income, the risk of default also increases.
- Loan applicants without home ownership, unverified loan source and with high loan amounts are more likely to default.
- Loan amount and interest rate were high for the borrower who has charged off. So, the loan amount and interest rate should be determined based on their annual income, loan grade and other aspects.

Driving Factors:

- Interest Rate
- Loan Grade
- Annual Income
- Debt to Income ratio
- Loan Amount
- Home Ownership

These features can help indicate if a person is likely to default.