# Lending Club Case Study

An Exploratory Data Analysis on the Loan Dataset

Analysis made by
**Vikas Soorya Krishnakumar**

# Problem Statement

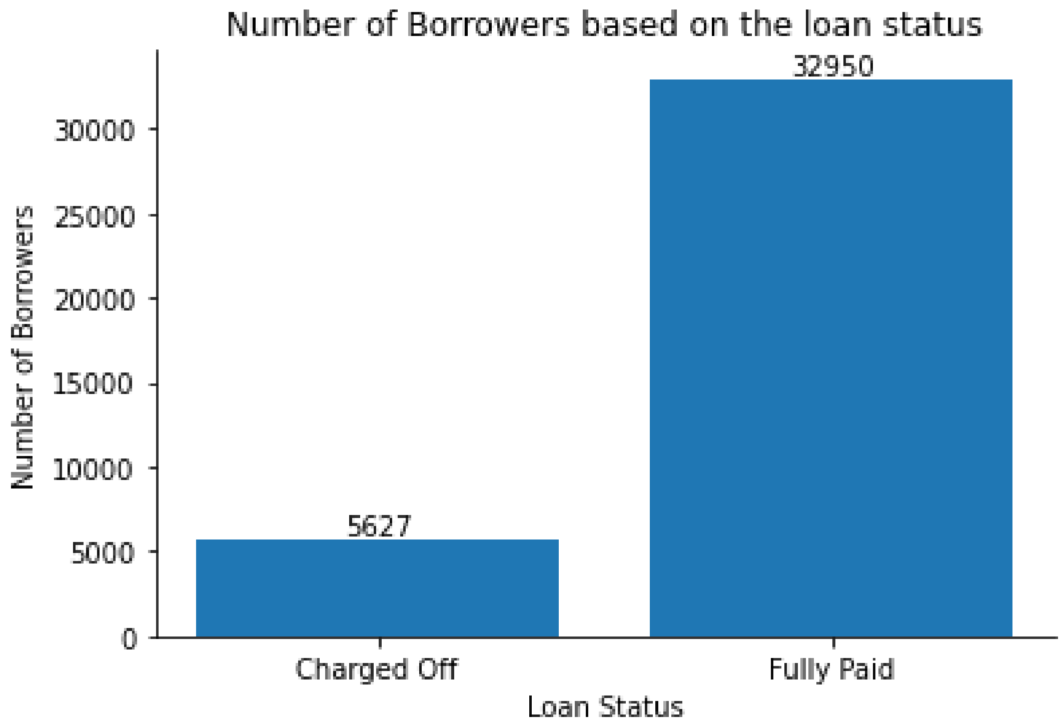
- A consumer finance company specializing in various types of loans to urban customers has provided a dataset of the past loans that were disbursed by the company.
- The dataset includes various consumer attributes and loan attributes.
- Consumer attributes include customer ID, location, annual income, credit inquiries, etc.,
- Loan attributes include the loan amount, interest on the loan, installment to be paid, date of disbursal, etc.,
- The loan attributes also includes the "loan status" which tells whether the loan is "Fully Paid" or "Charged off". Here, Charged off means the loan is defaulted by the borrower.

**Based on the past data, the company wants to identify the driving factors that impact the loan status (i.e. whether a loan would be fully paid or defaulted).**

# Analysis Approach

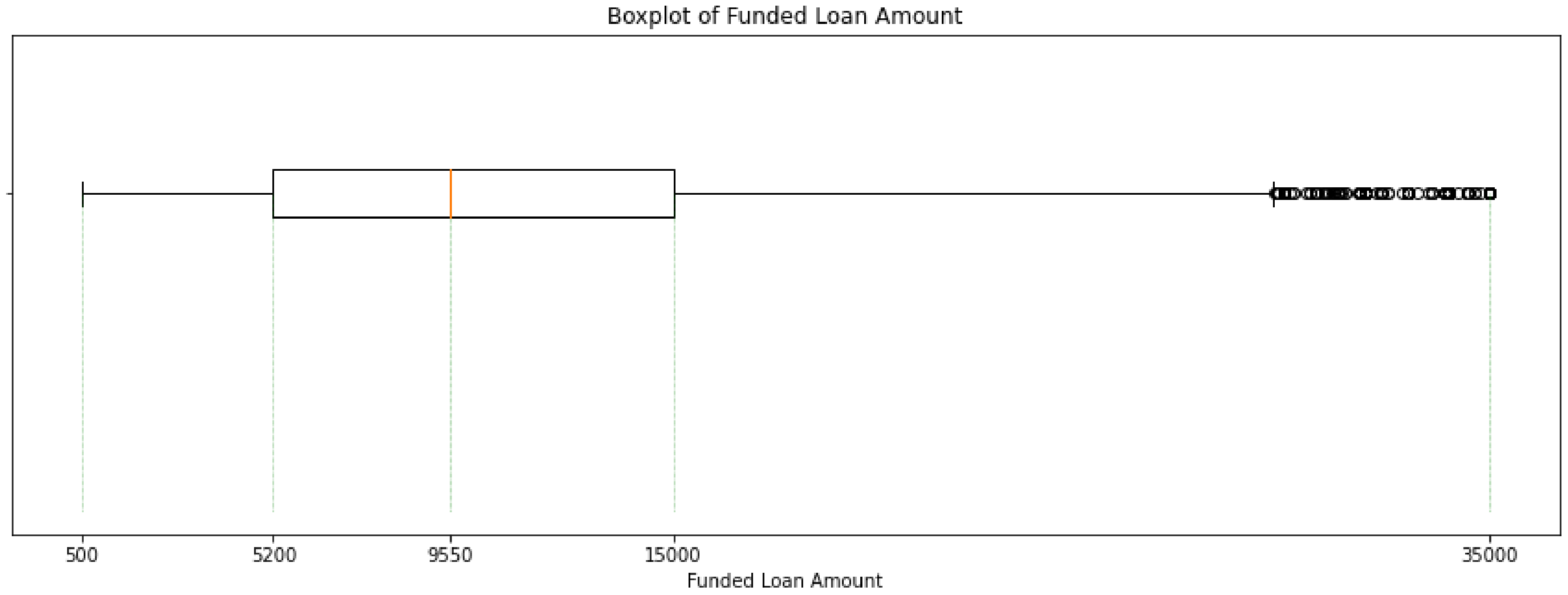
| Data Understanding | Data Cleaning | Univariate Analysis | Bivariate Analysis | Inference |
|---|---|---|---|---|
| Understand the dataset provided and make sense of what the different variables denote | Remove columns with invalid data or no data.<br>Convert data to appropriate datatypes like string to numbers, date-time, etc., | Analyze the distributions of the business relevant variables one at a time.<br>Create derived variables when required | Analyze the relationship of two different business relevant variables at a time<br>Create derived variables when required | Provide insights and suggestions from the data analysis. |

# Distribution of loan status



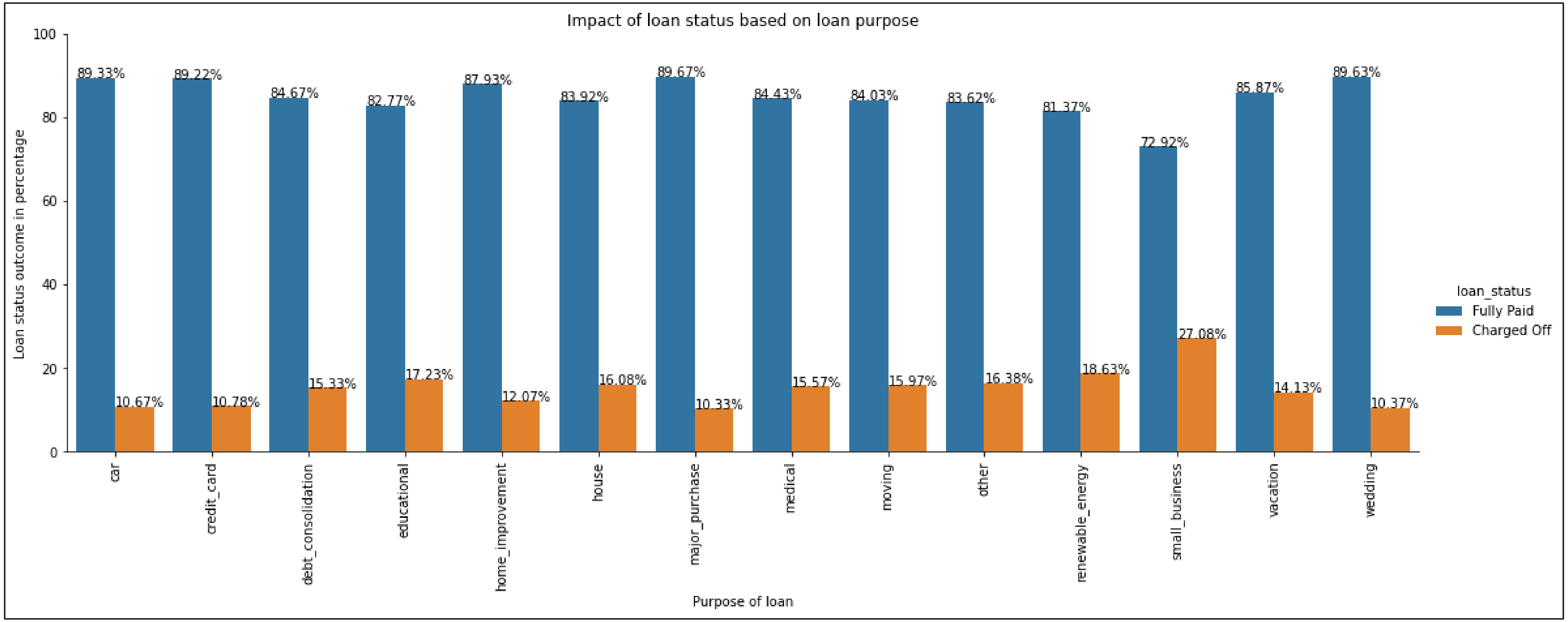

Out of the 38577 loans that were lent, 32950 (85.4%) loans were fully paid and 5627 (14.6%) loans were charged off.

# Distribution of funded loan amount




It can be observed that the minimum amount of loan funded is $500 and the maximum amount is $35000. It can also be seen that around 50% of the loan funded is between $5200 and $15000. The median amount of loan funded is $9550.

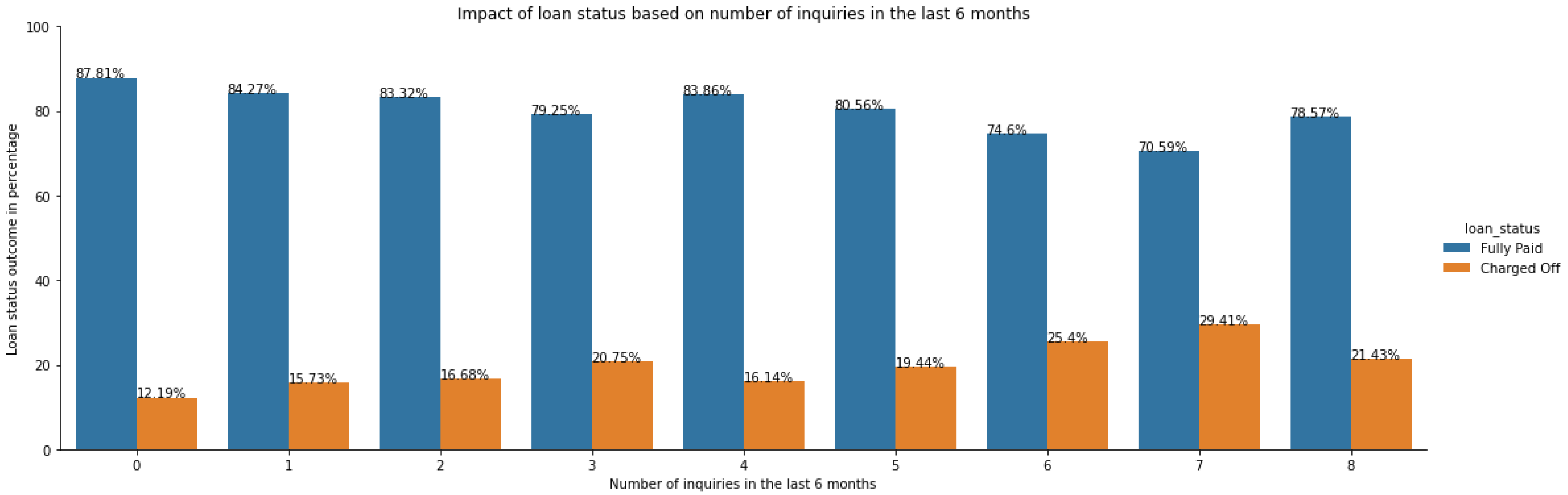
# Analysis: Loan Purpose impact on Loan Status




It can be noticed that loans provided for the purpose of "Small business" has the highest loan default percentage when compared to the other categories. Categories such as car, credit card, major purchase and wedding have the least loan default percentage.

Lending Club should avoid providing loans for the purpose of "Small business" unless there is a good collateral attached to it.

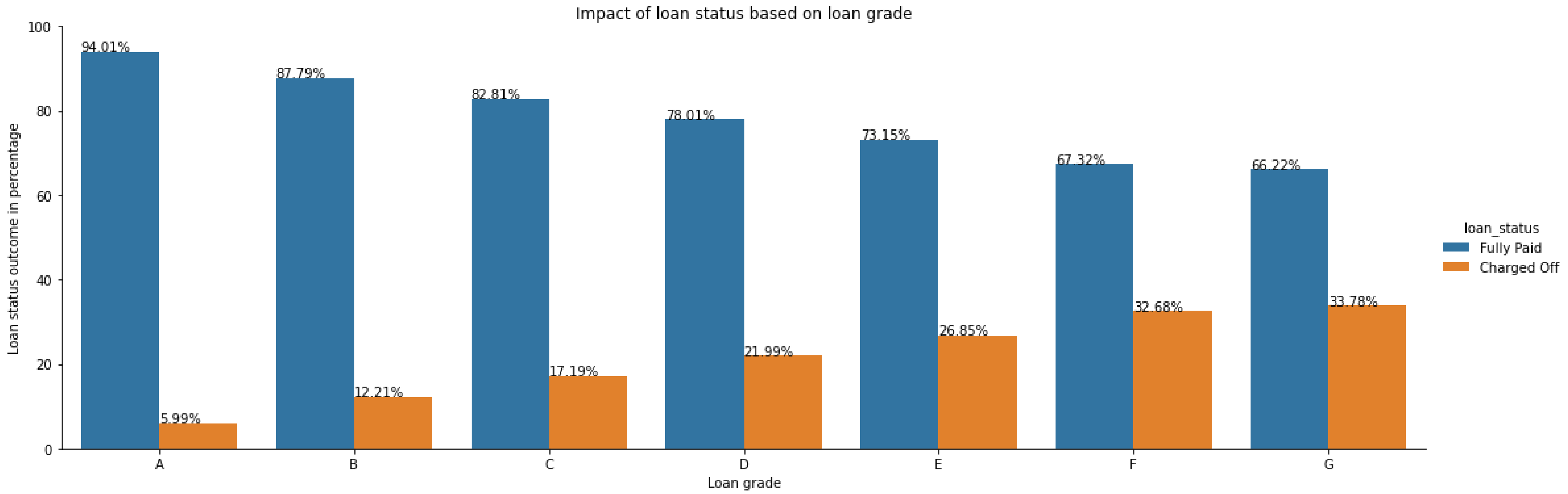
# Analysis: Credit inquiries impact on Loan Status




From the above plot, it can be inferred that borrowers with zero credit inquiries in the last 6 months are less likely to default. When the borrower has higher number of credit inquiries, the chances of loan getting defaulted increases.

Lending Club should reduce the number of loans provided to the borrowers with higher number of credit inquiries in the last 6 months.

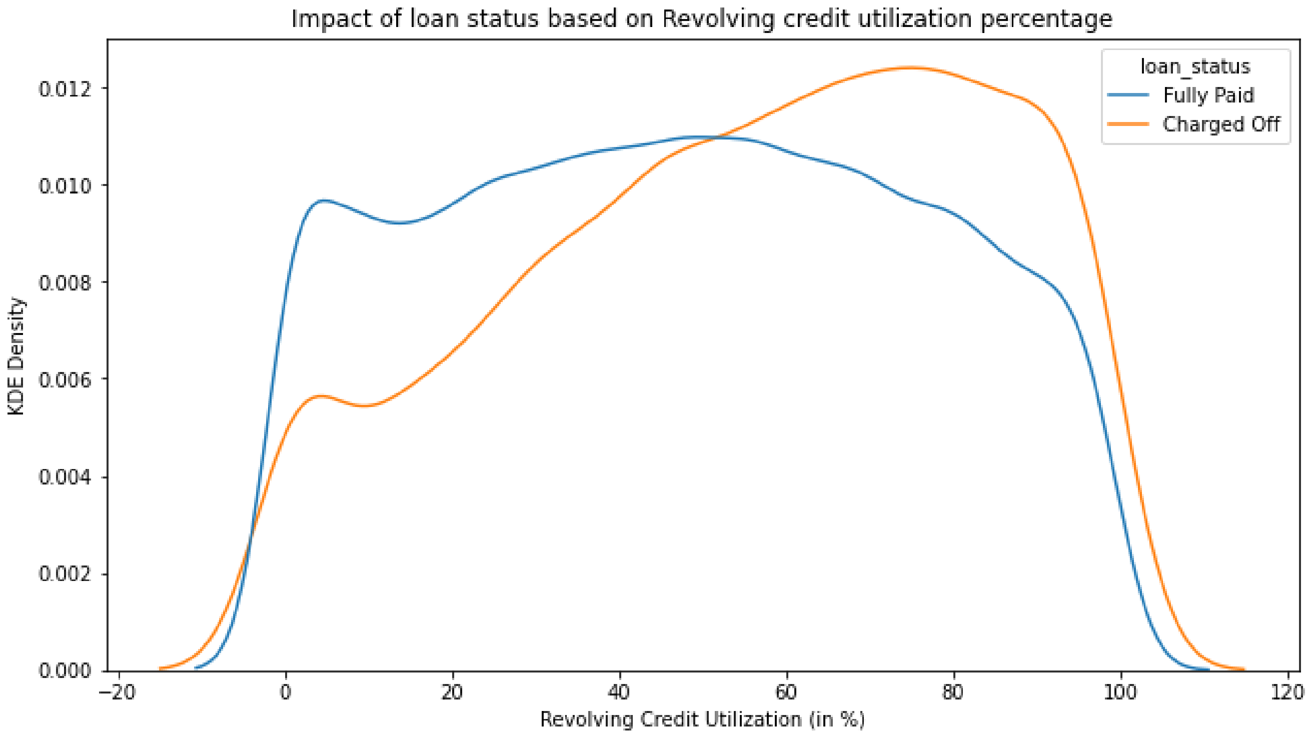
# Analysis: Loan Grade impact on Loan Status




It can be seen that the chances of a loan getting defaulted increases as the loan grade goes from A to G. Loan grade of A usually has low default chances and loan grade of G has the highest probability of getting defaulted.

Lending Club could focus providing loans only to the best grades to avoid credit loss.

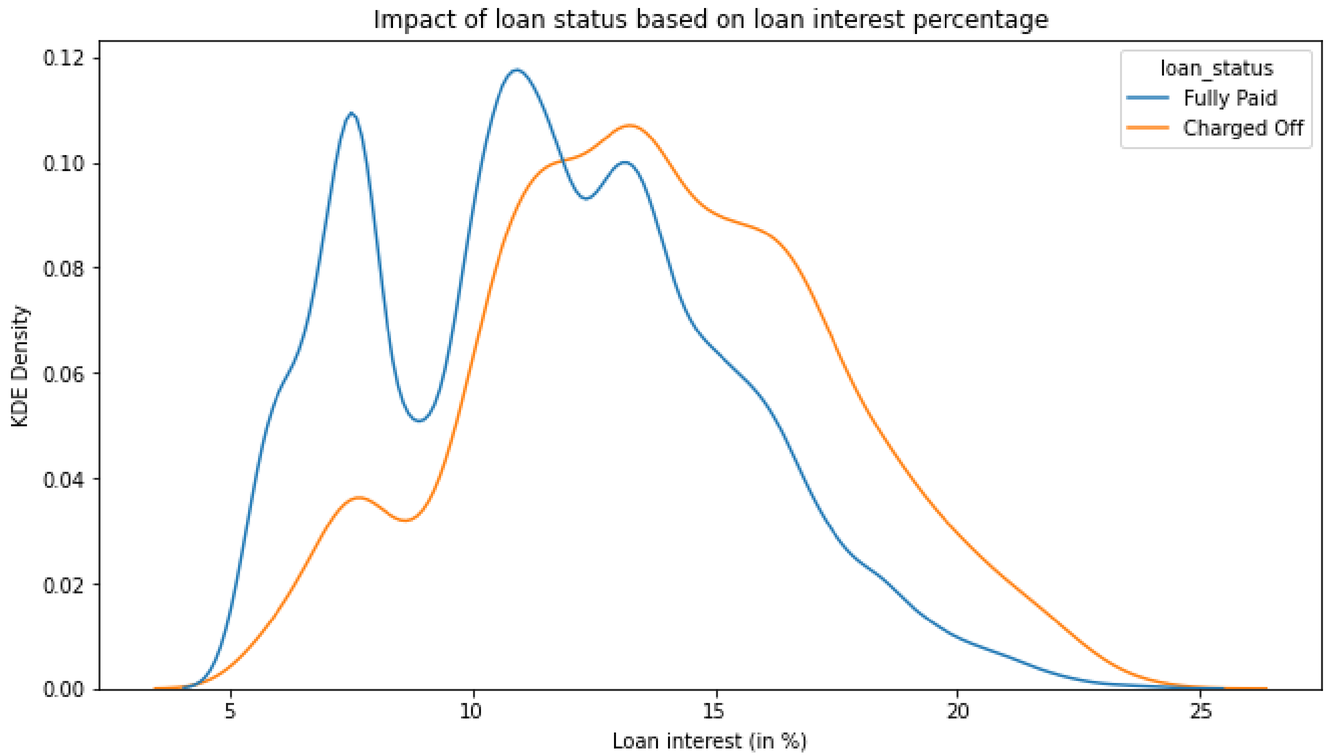
# Analysis: Credit Utilization impact on Loan Status




The above plot shows that borrowers who have a revolving credit utilization percentage above 40% tend to default on the loan more compared to the people who have lesser revolving credit utilization.

Lending Club should limit the loans offered to borrowers having high revolving credit utilization.

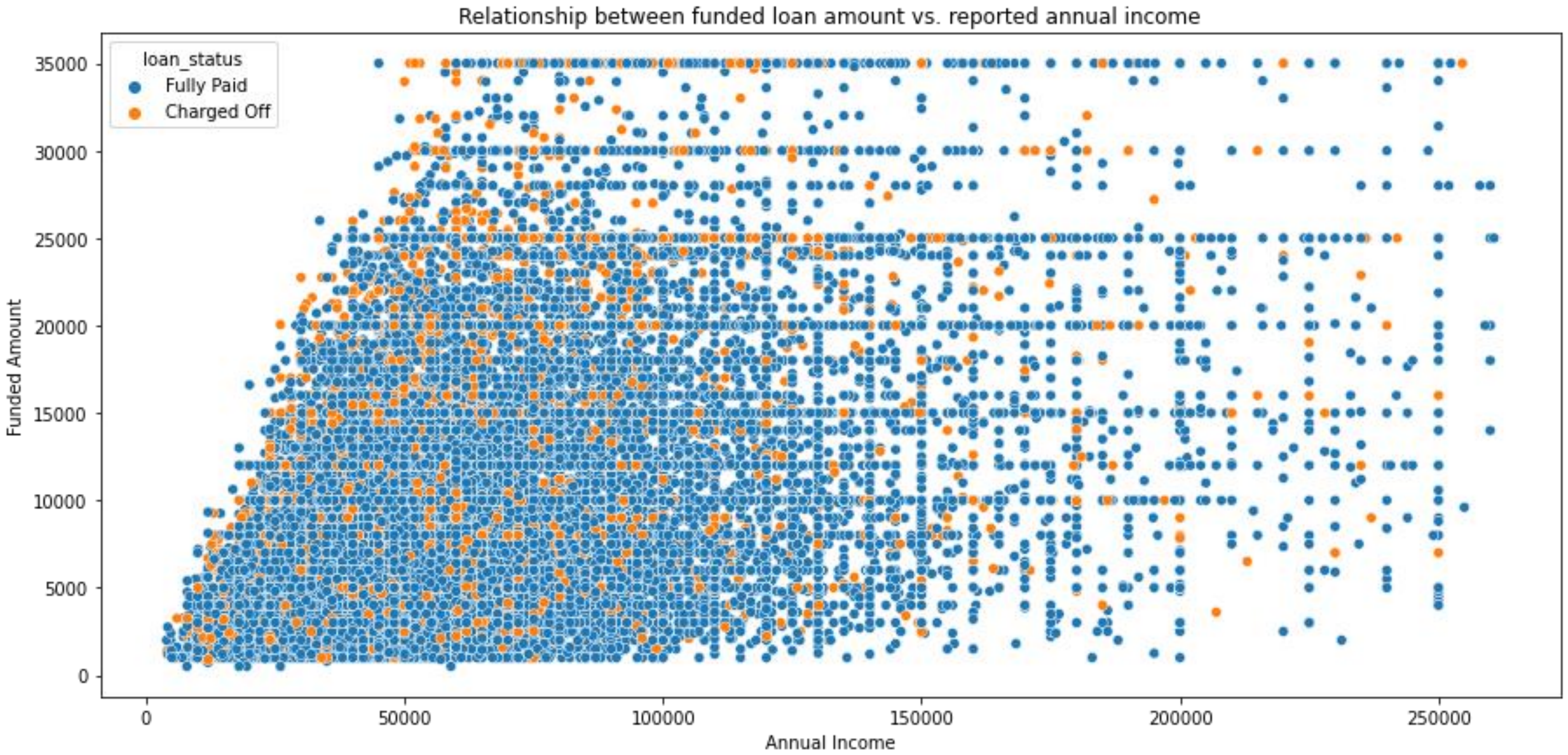
# Analysis: Loan interest percentage impact on Loan Status




From the above plot, it can be observed that the "Charged off" loans usually had a higher interest rate when compared to the "Fully paid" loans. The compounding effect of high interest could have made the borrowers not to pay the full amount back to Lending Club.

Lending Club could consider lowering the loan interest rates to make it easier for the borrowers to repay the money promptly.

# Analysis: Impact of funded loan amount and Reported annual income on the loan status

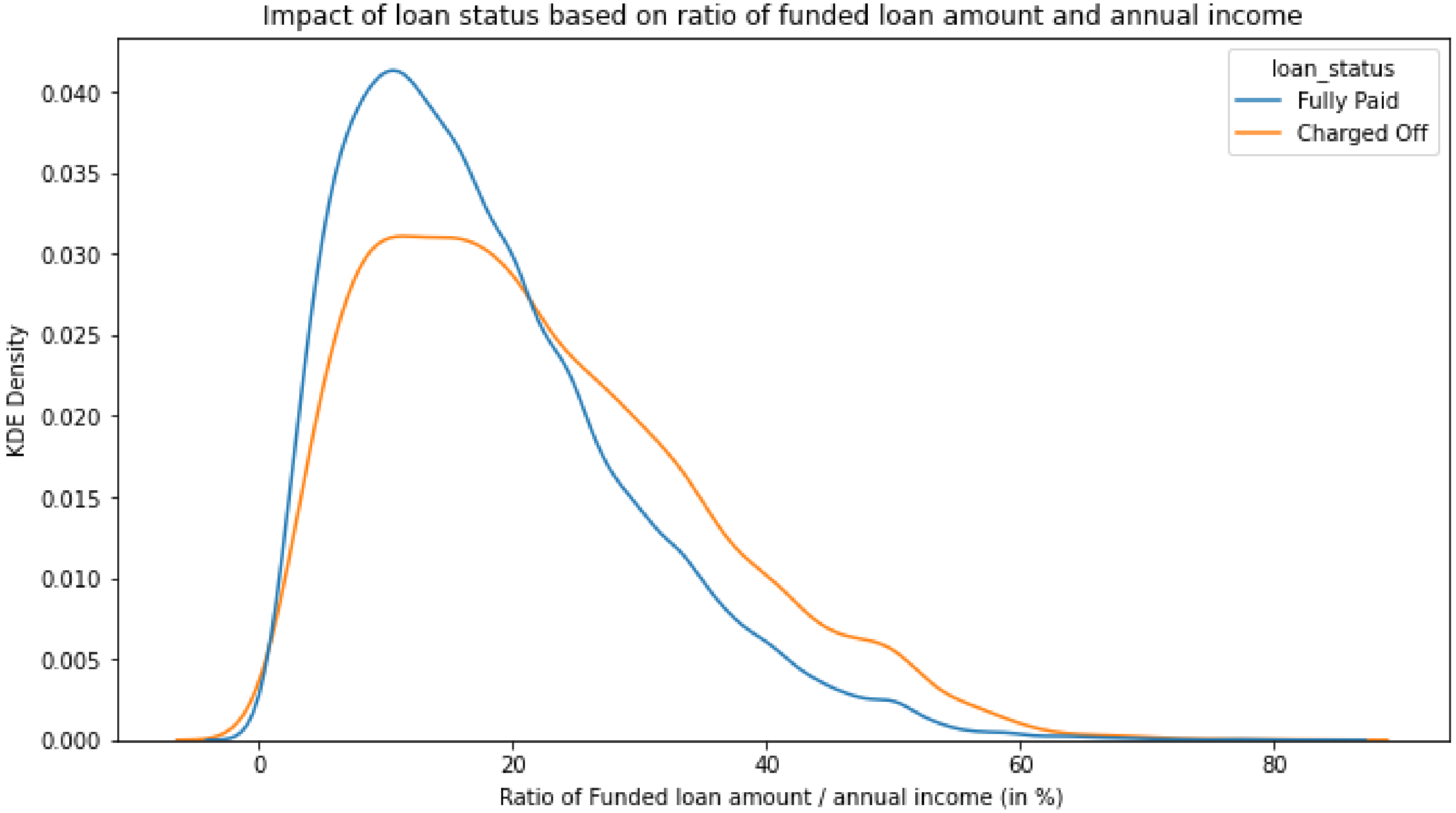



In the left side of the plot, we could see a slope indicating a small positive correlation between the annual income and the funded loan amount. However, this plot does not provide any insight when it comes to the loan status as the data is spread out almost evenly.

We have to derive a new data from the annual income and the funded loan amount data to get better insights.

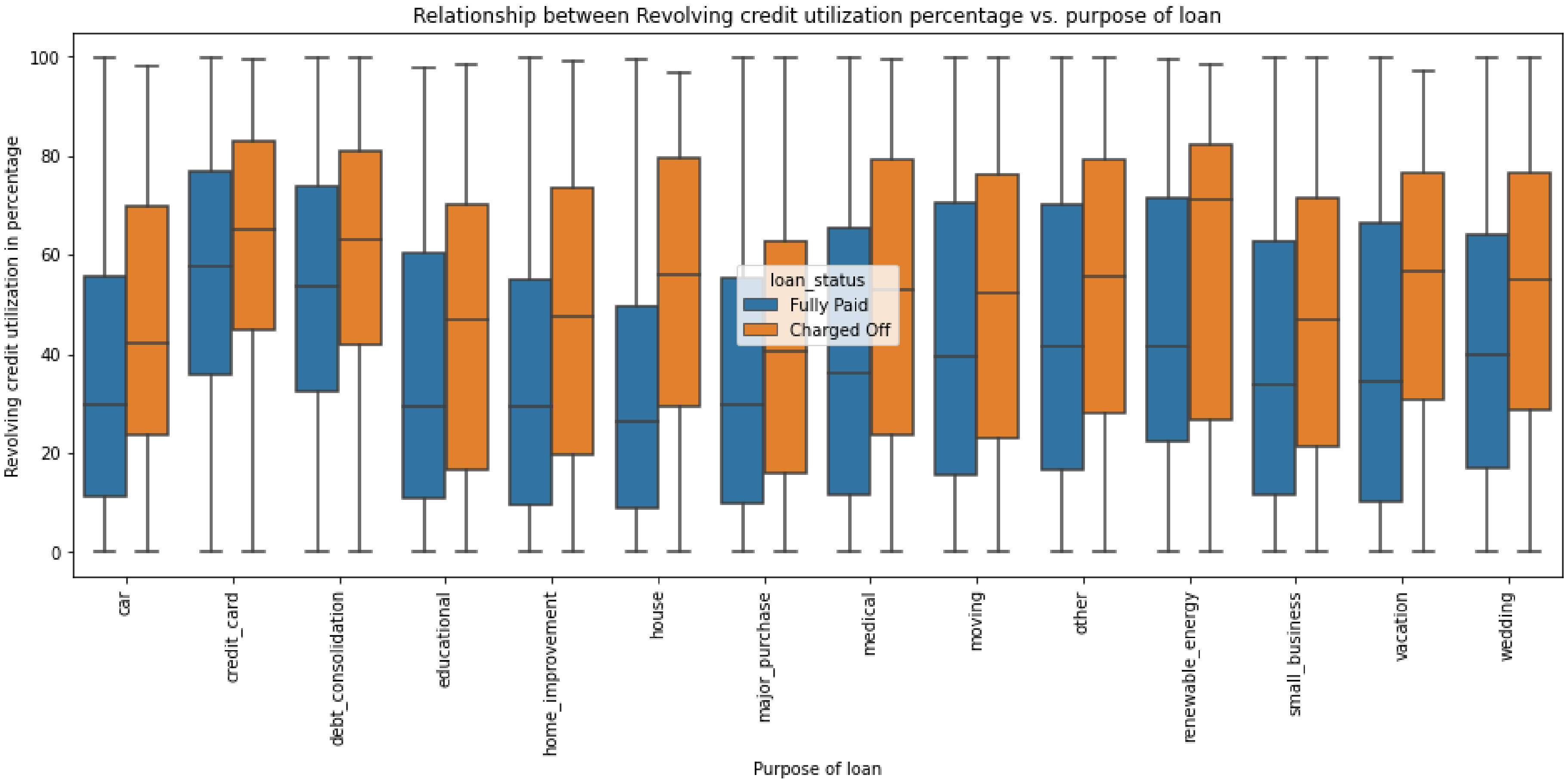
# Analysis: Impact of funded loan amount and Reported annual income on the loan status – Better insights

A ratio of funded loan amount and the reported annual income is derived and this is plotted as a distribution curve for the "Fully paid" and the "Charged off" loan status.



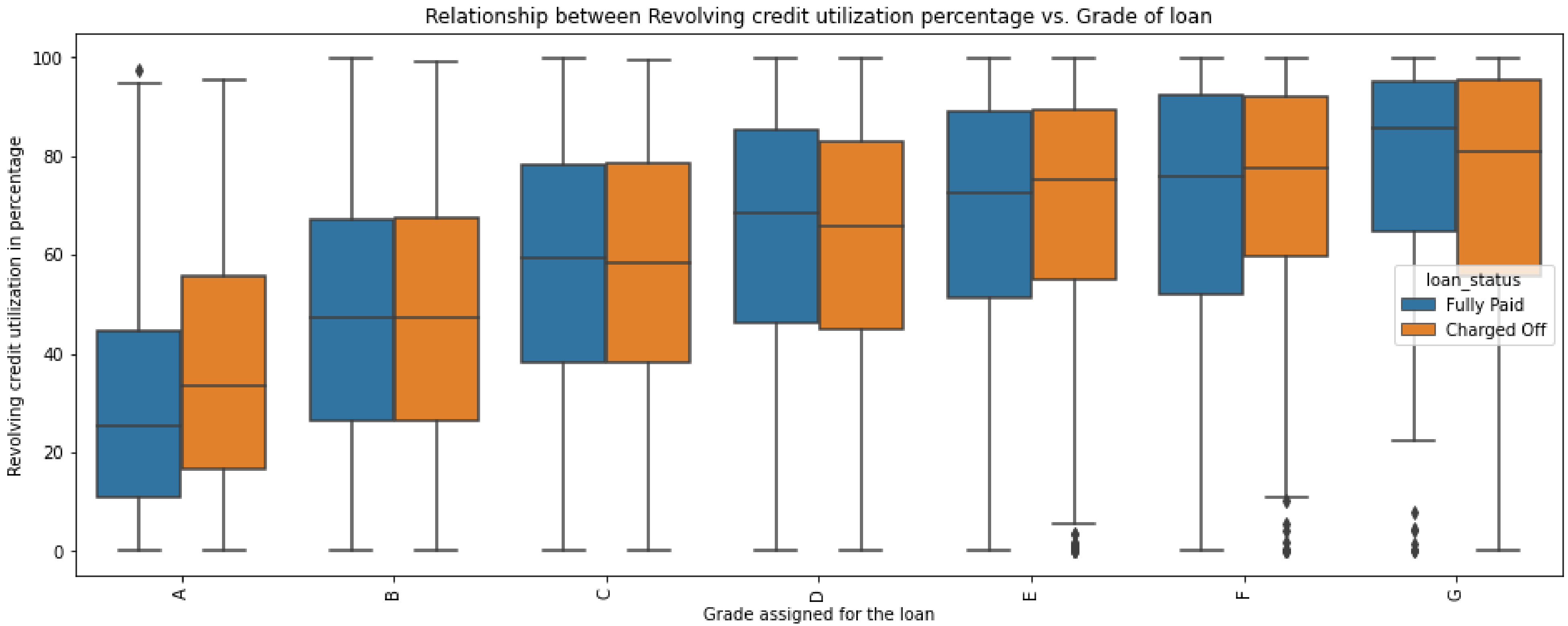

From the above plot, we can clearly see that, once the ratio of funded loan amount and annual income crosses 20%, the chances of loan getting defaulted becomes higher. Lending club should consider this ratio while disbursing loans.

# Analysis: Impact of Credit utilization percentage and Purpose of loan on the loan status




It can be seen from the above figure that the revolving credit utilization percentage is always higher for the people who default on the loan. There is a stark difference that can be seen in the "house" category where the majority of borrowers who had a higher revolving credit utilization percentage had defaulted on the loan. Lending Club should restrict loans provided with the purpose as "house".

# Analysis: Impact of Credit utilization percentage and Loan grade on the loan status




Though the distribution between "fully paid" and "charged off" are more or less similar for most grades, we could get a valuable insight from Grade A. We could observe that borrowers with the credit utilization percentage on the higher side tend to default on the loan even though the loan grade is A.

Lending club should prioritize credit utilization percentage above loan grade when approving loans.

# Inference

From the data analysis on the Lending Club's loan data, we could figure out some of the driving factors that affect whether the loan will be fully repaid or defaulted. These factors include:

- Purpose of the loan taken
- Loan interest percentage
- Ratio of loan amount and annual income
- Revolving credit utilization percentage
- Number of credit inquiries made in the last 6 months
- Grade of the loan

The lender should be wary of these facts to avoid credit loss. For high default risk cases, the lender could ask the borrower to pledge collaterals. This could in turn help avoid financial loss for the lending company.

# Suggestions

- Loans provided for the purpose of "Small Business" can be avoided. Instead focus on low default rate purposes such as car, credit card and wedding purposes.
- Avoid disbursing loans to borrowers with high revolving credit utilization percentage.
- Avoid lending loans to borrowers whose loan to annual income ratio is very high.
- Restrict loans to borrowers with high number of credit inquiries in the last 6 months.
- Try reducing the loan interest percentage to make it affordable for the borrowers. This may come at a slight financial loss but would help in lowering the credit loss.