Lending Club Case Study

Group Members:

- Viswanadha Venkat Prabhat
- Veda Charitha Bandikallu

Business Objective

- Lending Club, the largest online loan marketplace, facilitates personal loans, business loans, and medical procedure financing. This project's primary business objective is to enhance the company's decision-making process in loan approvals by identifying and mitigating credit risks associated with loan applicants.

- The company can reduce potential financial losses and optimise its loan portfolio by identifying these "risky" applicants who are likely to default. In other words, the company wants to understand the driving factors (or driver variables) behind loan default.

- Ultimately, this will strengthen the company's position as the leading online loan marketplace by improving its ability to manage risk and maintain profitability.

- Specifically, the goal is to leverage Exploratory Data Analysis (EDA) to uncover patterns and key factors that predict loan defaults.

Steps Taken

Data Cleaning

- Drop columns/Rows:
  - All NA
  - Single values (more than 90%)
  - more than 50% of values are NA
  - removing records with missing data (<1%), and so on
- Outlier Treatment

Data Manipulation

- Data type conversions,
- Dates Extraction (month, year, …),
- Removing unnecessary characters ('%', '+', …) , and so on

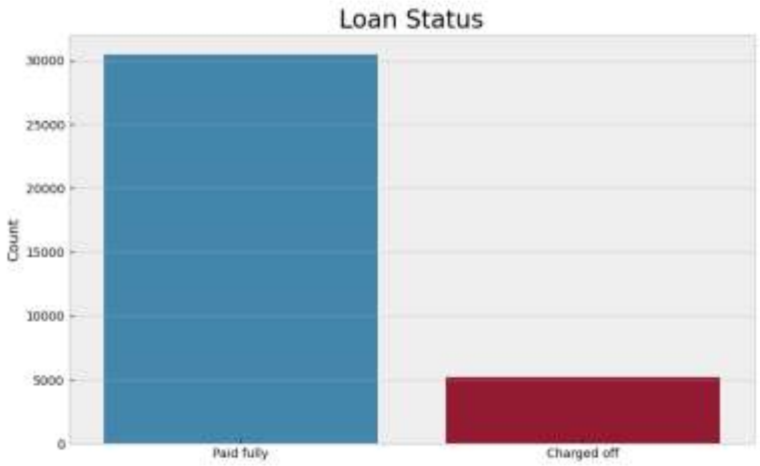
Data Analysis

- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

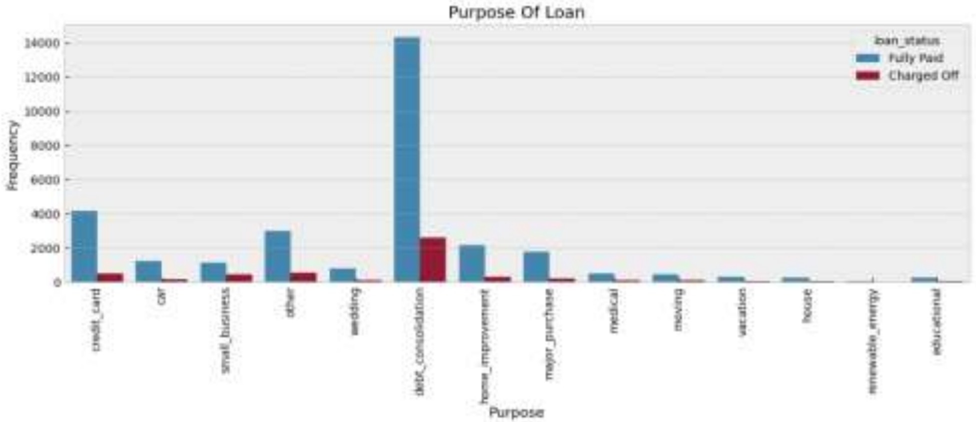
Recommendations

Analysis

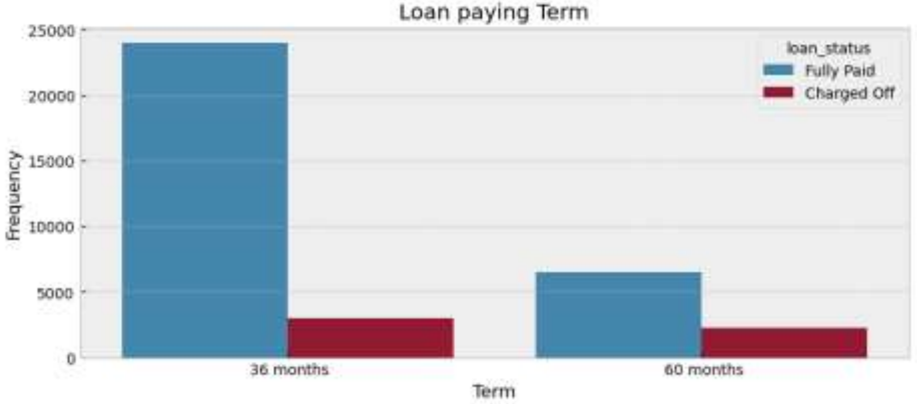
Univariate Analysis



In the given dataset, 85% of the borrowers have ‘Paid fully’, whereas 15% of borrowers have ‘charged off’ their loans.

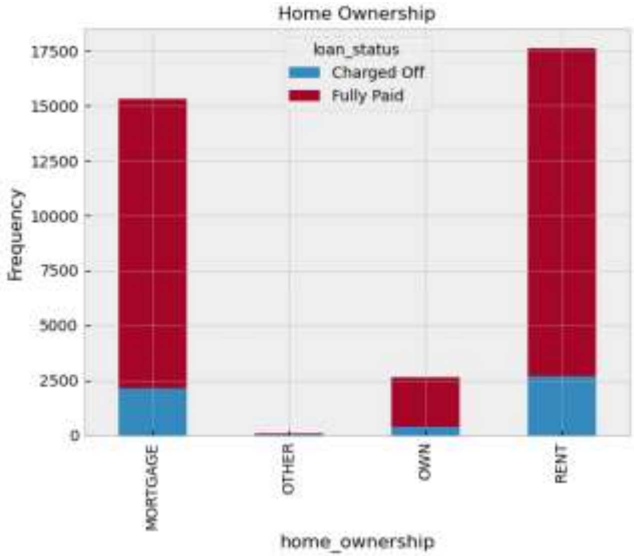


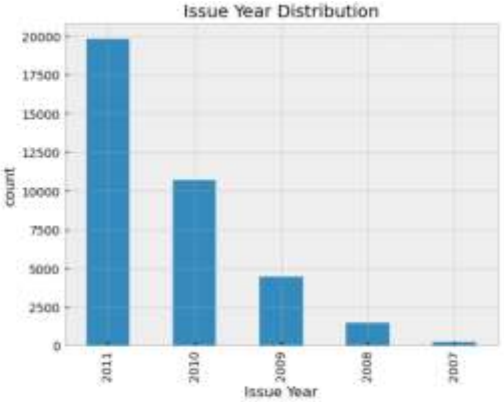
This graph displays the purpose of loan taken and the loan status.



The dataset contains 2 terms.
Ratio of ‘Fully Paid’:’Charged Off’ is more in the 60-month term.

The number of loans issued decreases steadily from December through February.
The highest number of loans issued occurs at the end of the year i.e., the holiday season

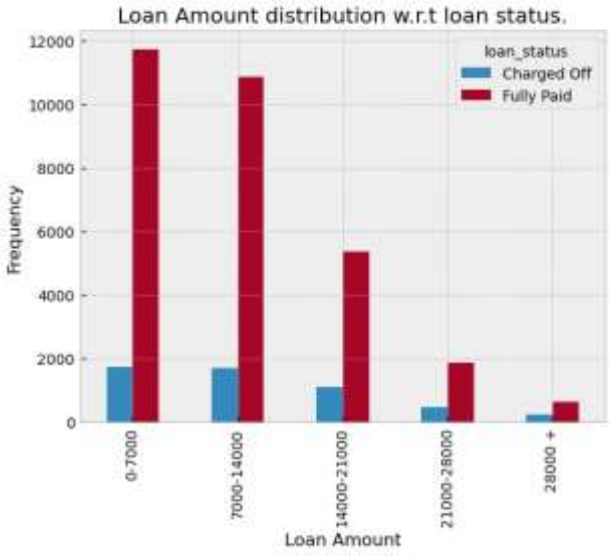




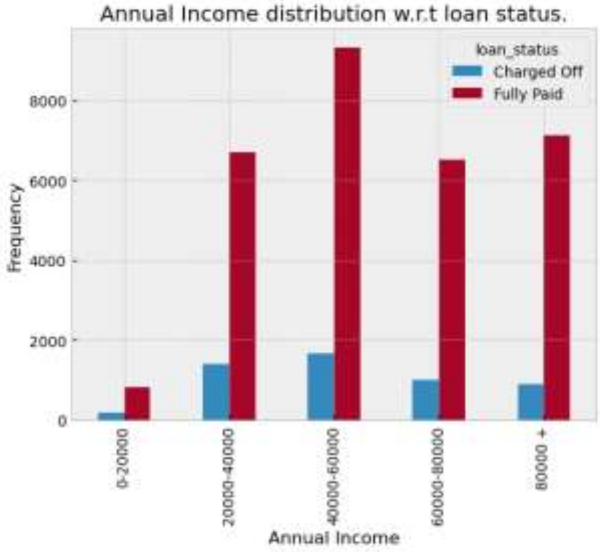
No. of loans issued by the lending club has increased over the years.

Home Ownership status of the borrower.
The relatively higher frequency of Charged Off loans in MORTGAGE and RENT categories may suggest a higher risk profile for these home ownership types.

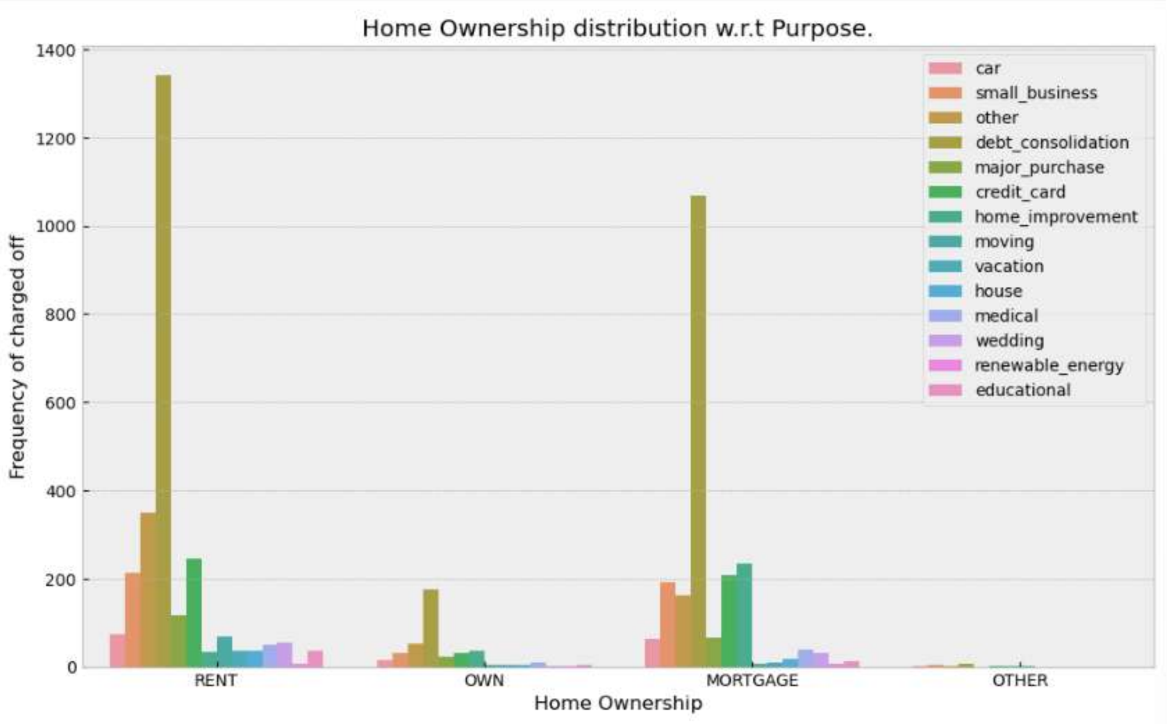
Segmented Univariate Analysis



This chart indicates that lower loan amounts tend to be fully paid more often, while higher loan amounts see a decrease in fully paid loans and an increase in charged-off and current loans.
Loan Amount can be a used as a driving factor

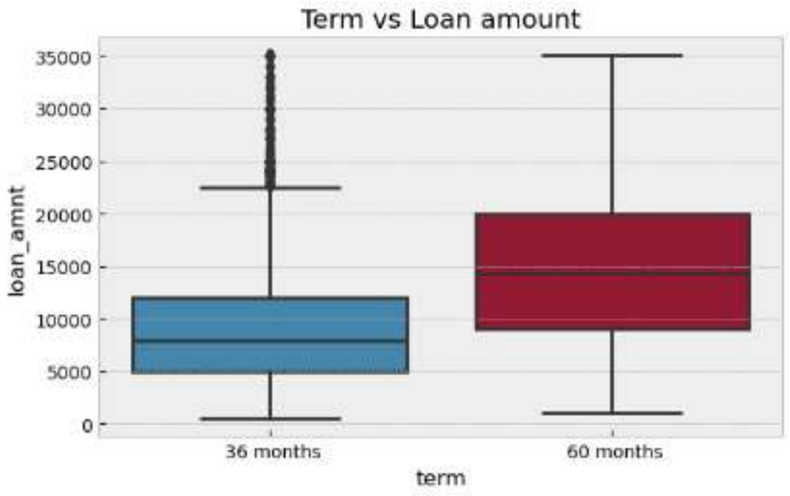


Charged Off ratio is getting reduced as annual income increases.
Annual Income is a driving factor

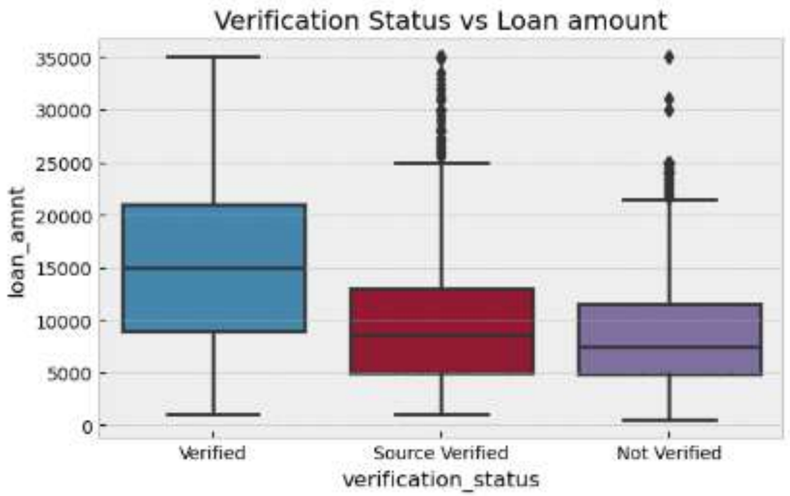


‘RENT’ and ‘MORTGAGE’ home ownership statuses are associated with higher frequencies of charged-off loans, especially for debt consolidation purposes.
The purpose of loan can be a driving factor.

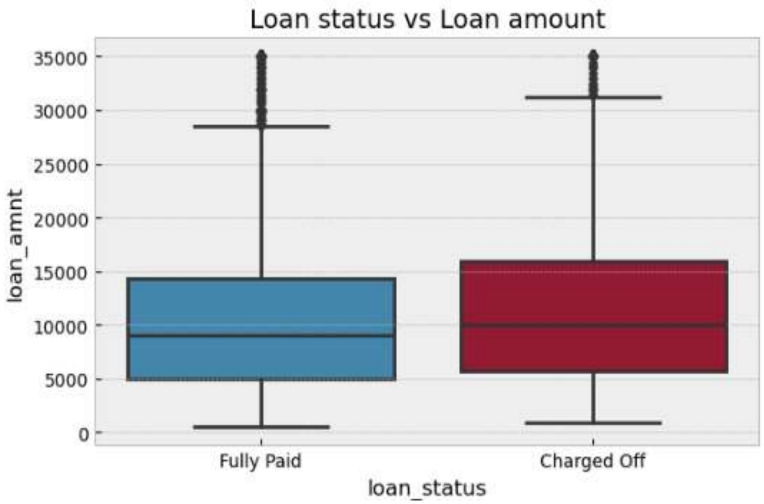
Bivariate Analysis

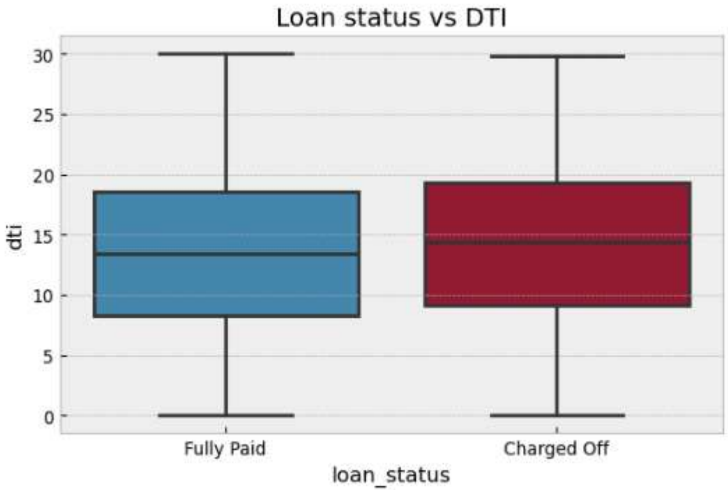


Loans with a 60-month term generally have higher amounts and more variability compared with a 36-month term.
Outliers in the 36-month term indicate that while most loans are within a certain range, there are exceptions where the loan amounts are significantly higher.



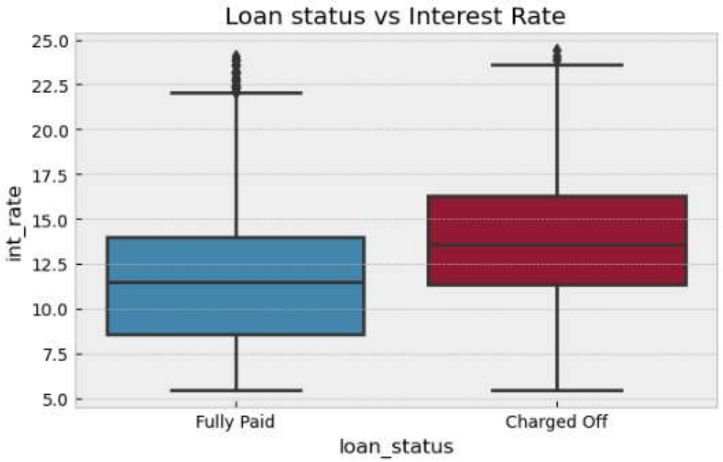
Compared to Source Verified and Not Verified borrowers, Verified Borrowers receive higher loan amounts as they are trusted.



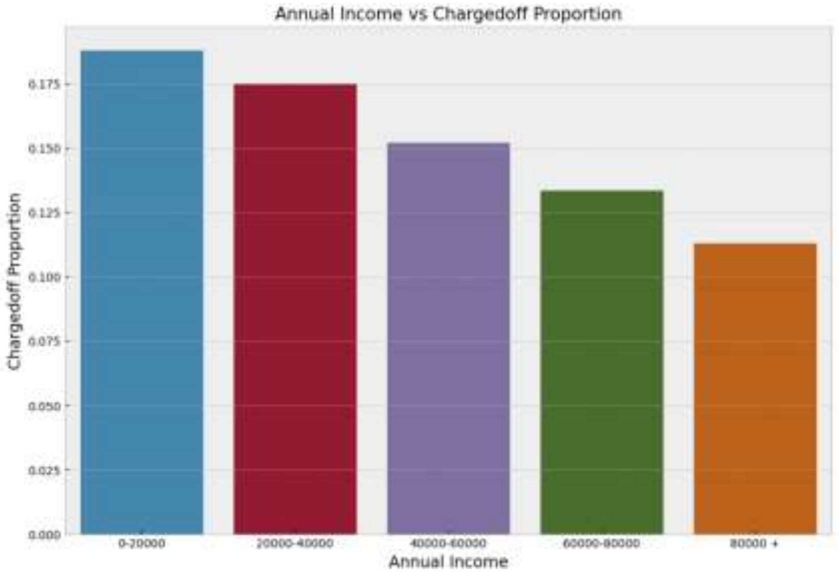


Borrowers with higher loan amounts might be at a slightly higher risk of defaulting, as evidenced by the larger median and wider IQR for "Charged Off" loans. However, the difference is not very large, suggesting that loan amount alone might not be a significant predictor.

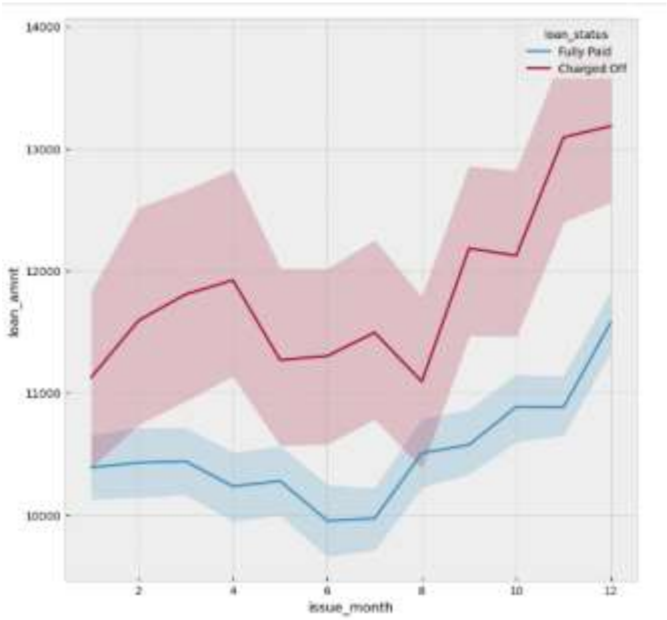
Borrowers with higher DTI are more likely to have their loans charged off, while those with lower DTI are more likely to pay off their loans or remain current.
DTI can a driving factor.



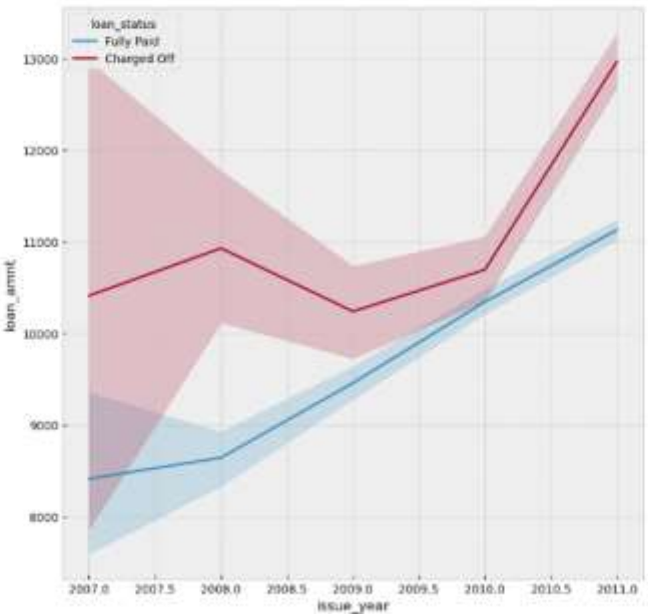
Borrowers with lower interest rates are more likely to repay their loans fully, while those with higher rates are more prone to default.
Current loans have a broader range of interest rates, which could indicate a mix of risk levels among current borrowers.
This can be driving factor



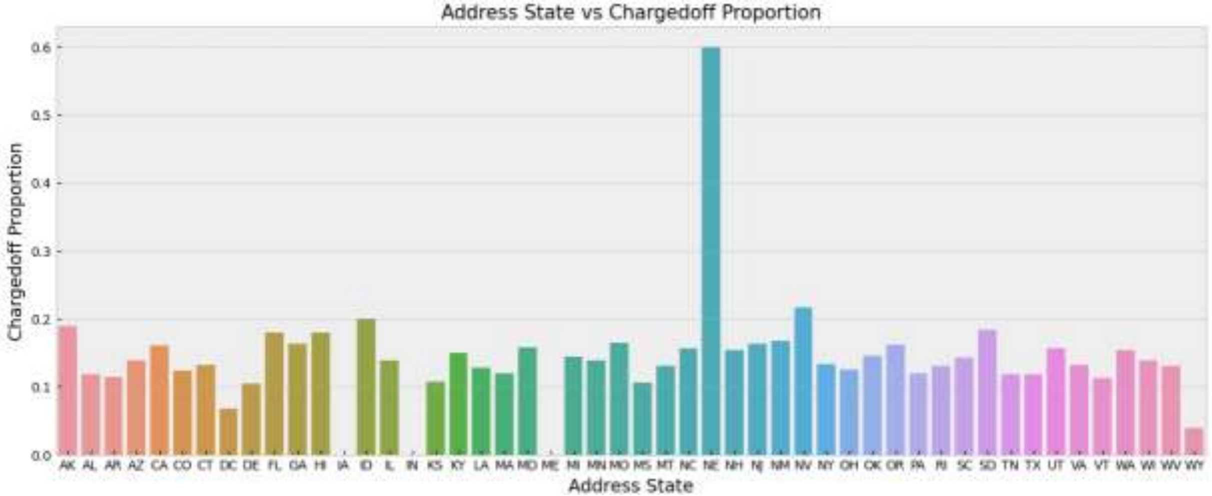
As the annual Income increases the chances to default reduces.
This can be considered as driving factor.



Monthly trends indicate that larger and more unpredictable loan amounts are more prone to being charged off, while loans with smaller, more consistent amounts are more likely to be fully paid.

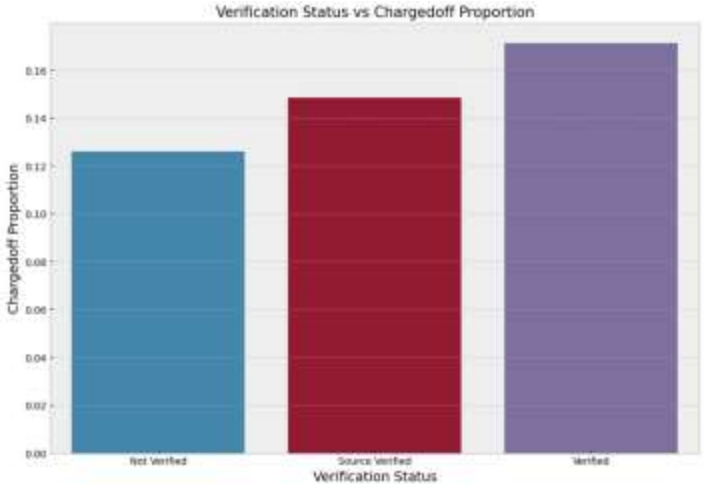


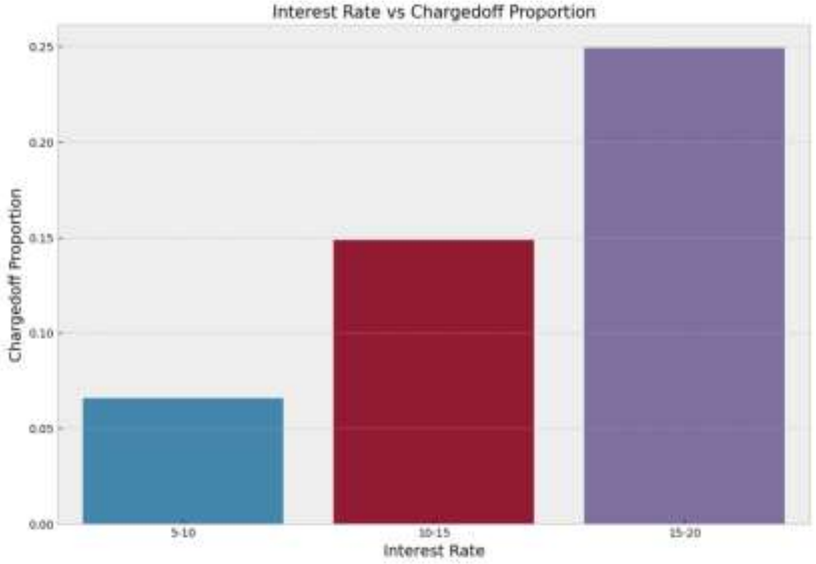
Yearly trends reveal that larger loan amounts are increasingly associated with a higher likelihood of being charged off, with a significant rise in this pattern after 2008, possibly due to economic conditions.

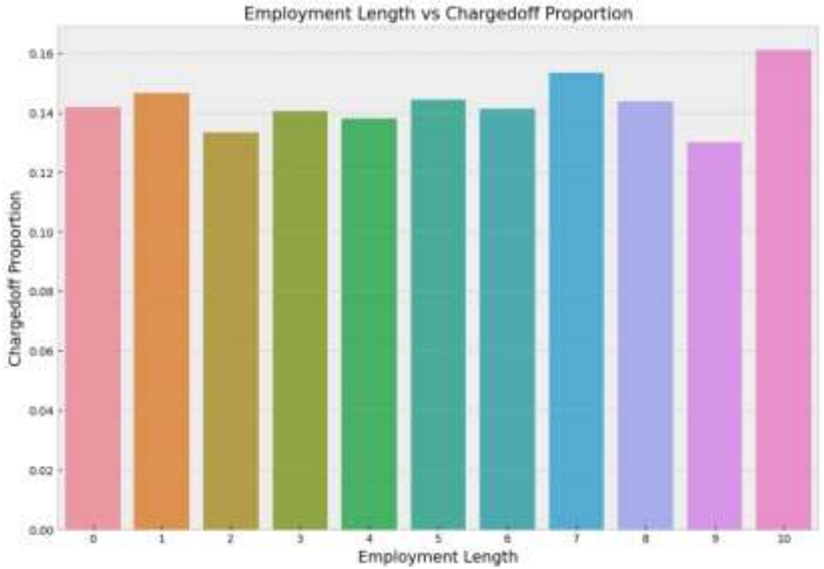


Nebraska (NE) has an exceptionally high charged-off proportion, far above any other state.
Most other states have charged-off proportions that are relatively close to each other, generally falling between 0.1 and 0.2.
States like South Carolina (SC) and Utah (UT) also have slightly higher charged-off proportions compared to others.
This can be considered as a driving factor as it's segregating borrowers as per state and helping give analysis on it.

Loans which are verified have slightly higher chances of being charged off.
Losses incurred is more even when the sources are verified.
This cannot be a driving factor

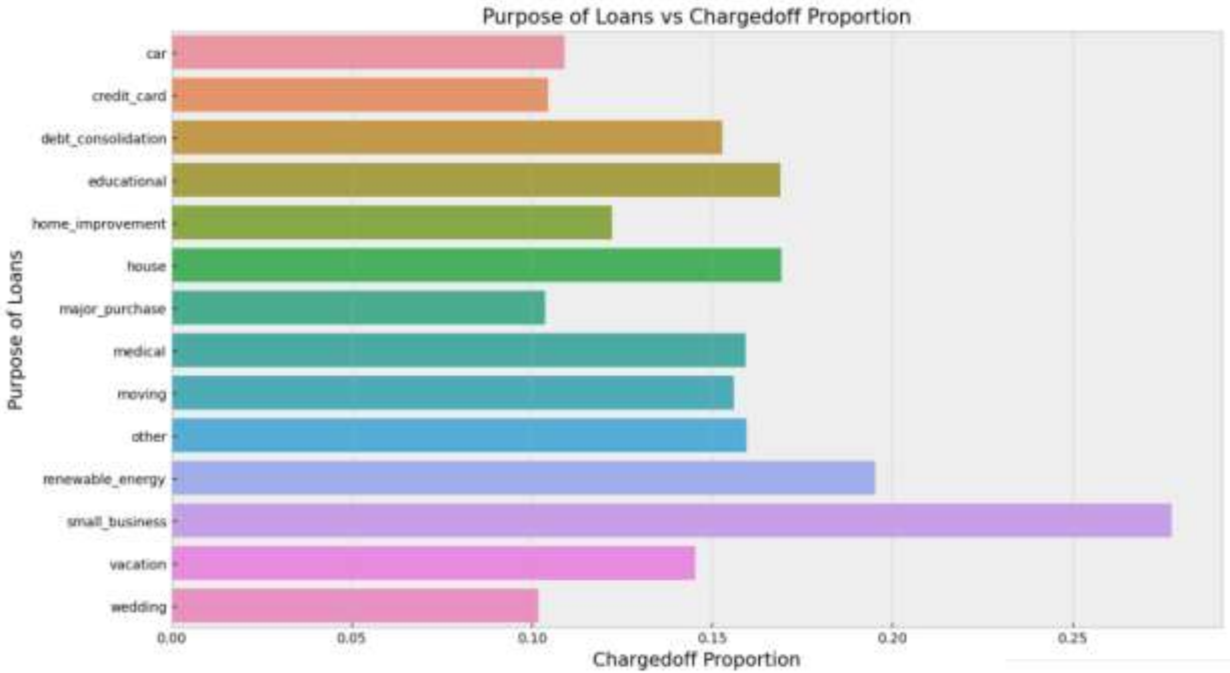






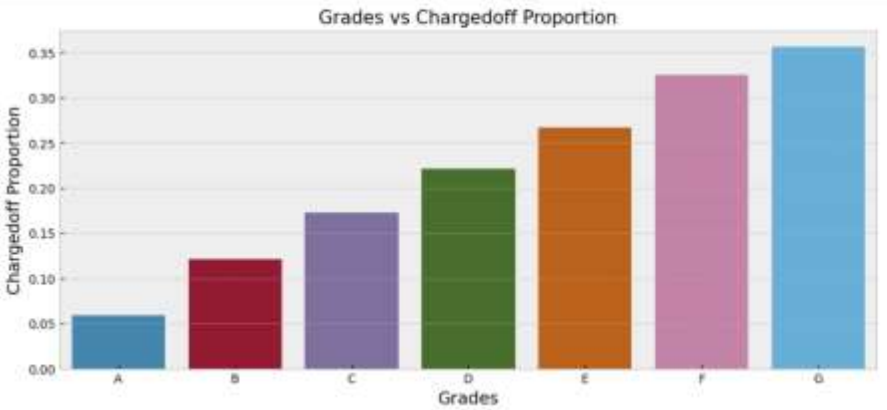
Loans with higher interest rate (15 – 20%) are more likely to be defaulted that the loans with lower interest rates.
Interest Rate is a driving factor.

Borrowers with 9 years experience defaulted least as per the graph, but the variance is low between the various employment lengths.
Hence, this is not a driving factor.

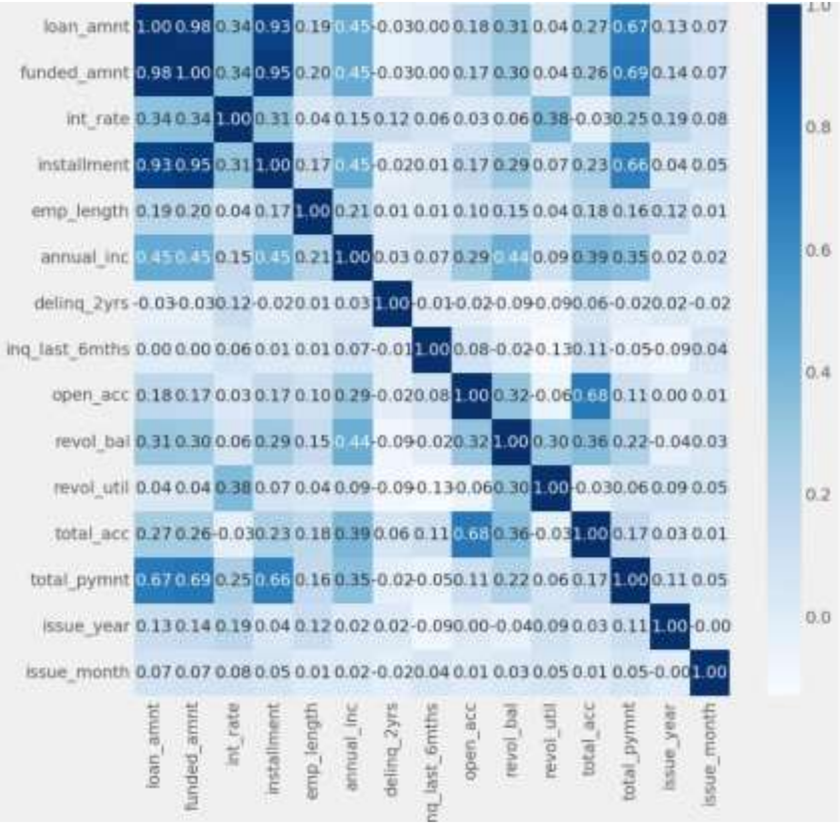


Small Business loans have the highest charged-off proportion.
Car loans have the lowest charged-off proportion.
Other categories like renewable energy, vacation, and medical loans also have relatively high charged-off proportions.
As there is a high variance between the purpose of loans and charged-off proportion, the purpose of loans is a driving factor.

As Grades are increasing (alphabetically), charged off proportion also increases.
A means lowest risk of loan default and G means higher risk of loan default.
Indicating that there are high chances of borrowers defaulting with lower grades (such as G).
This can be considered as driving factor.



Multivariate Analysis



Strong Positive Correlations:
- Loan Amount & Funded Amount
- Loan Amount & Installment
- Funded Amount & Installment
- Total Payment & Loan Amount
- Open Accounts & Total Accounts

Moderate Correlations
- Annual Income & Loan Amount
- Total Payment & Funded Amount

Weak or No Correlation
- Employment Length & Other Variables
- Inquiries in the Last 6 Months (inq_last_6mths)
- Delinquency in the Last 2 Years (delinq_2yrs)

Negative Correlations
- Revolving Balance & Interest Rate

Conclusion

The heatmap highlights strong relationships between key financial variables like loan amount, funded amount, and instalment while indicating that factors like employment length and recent credit inquiries have minimal impact on loan characteristics.

Recommendations

Based on the analysis done, the probability of defaulting is higher for borrowers who:

• Take loans with a term of 60 months.

• Have a loan status of "Verified," as they tend to take out larger loans with 60-month terms.

• Receive loans with an interest rate between 15% and 20%.

• Take loans for the purpose of starting a small business.

• Have lower credit grades, particularly F and G.

• Have home ownership status as "Rent" and take loans for debt consolidation.

• Have a low annual income (like between 0 and 20,000).

• Take loan amounts in the range of 0 to 14,000.

• are verified loan applicants, but it’s just a marginal loss compared to non-verified ones. Though its not a key factor, but needs to check the loan screening sources.