Lending Club Case Study

Teja Gunupudi

Shelly Goel

Data Identifications and Assumptions
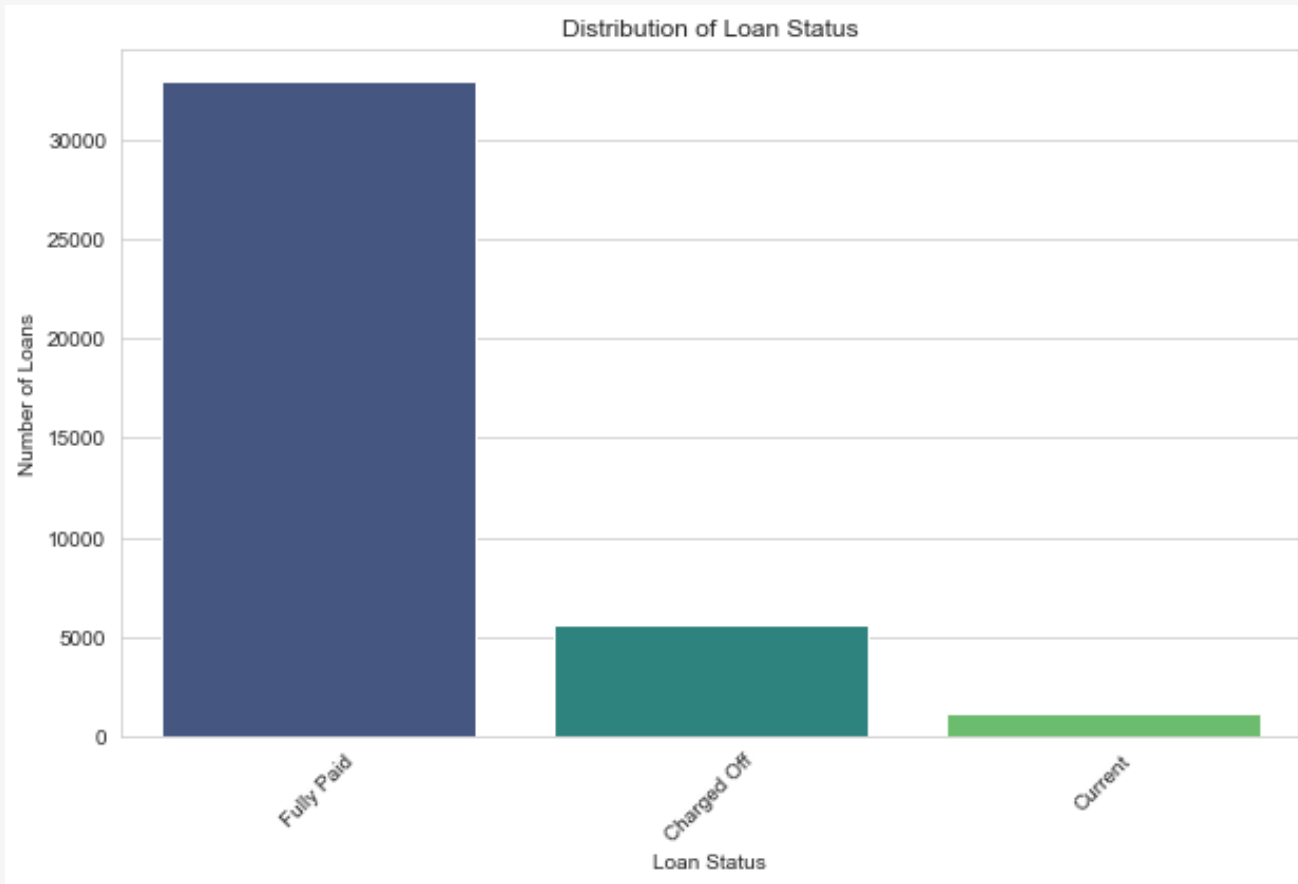
The loan data consists of multiple columns 111

Dropped columns with more than 90% missing data and finally landed with 55 columns

Additional columns were dropped because either they are user related (user id, etc.) OR post processing of loan. Columns remaining are 26

Univariate Analysis

Distribution of Loan Status

Loan Status Distribution

- A large majority of loans have been "Fully Paid"

- A significant number of loans are "Charged Off", indicating applicants who have defaulted

- A smaller portion is still "Current", meaning these loans are ongoing
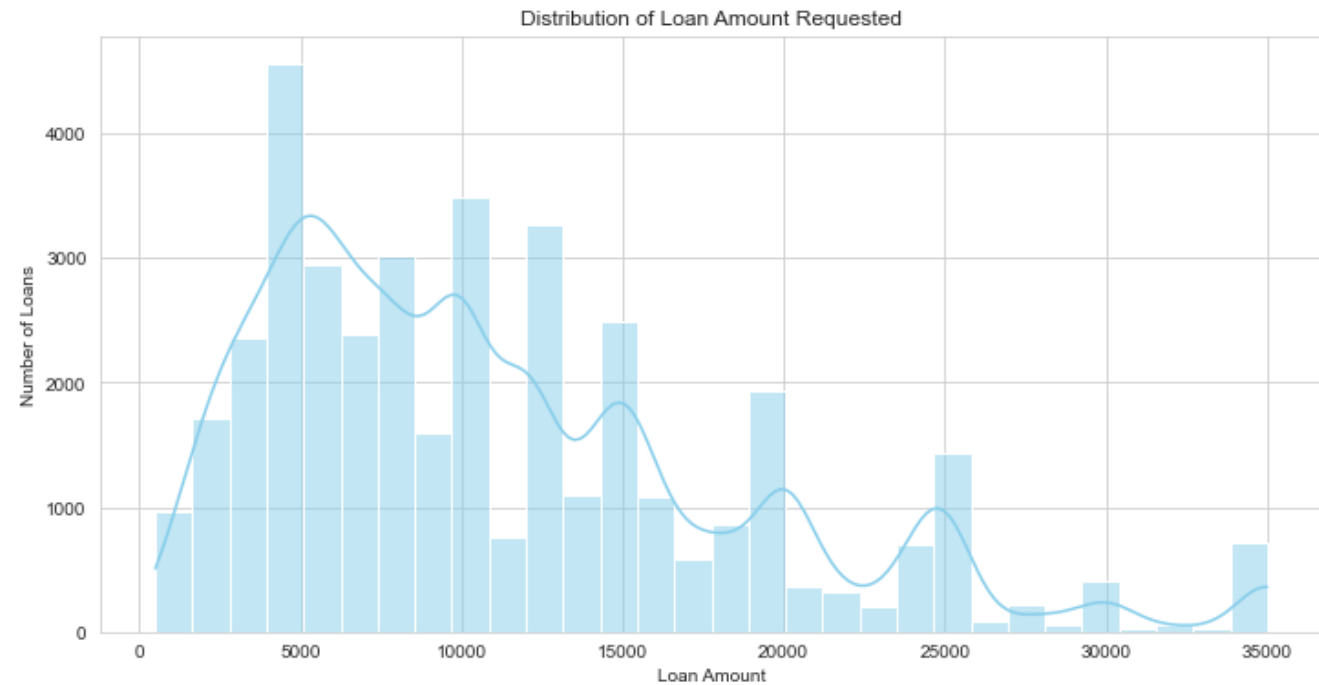
Loan Amount

- More defaults when 'loan_amnt' is between 0-7k

- As the loan amount increases, the number of requests tends to decrease, indicating fewer people request higher loan amounts
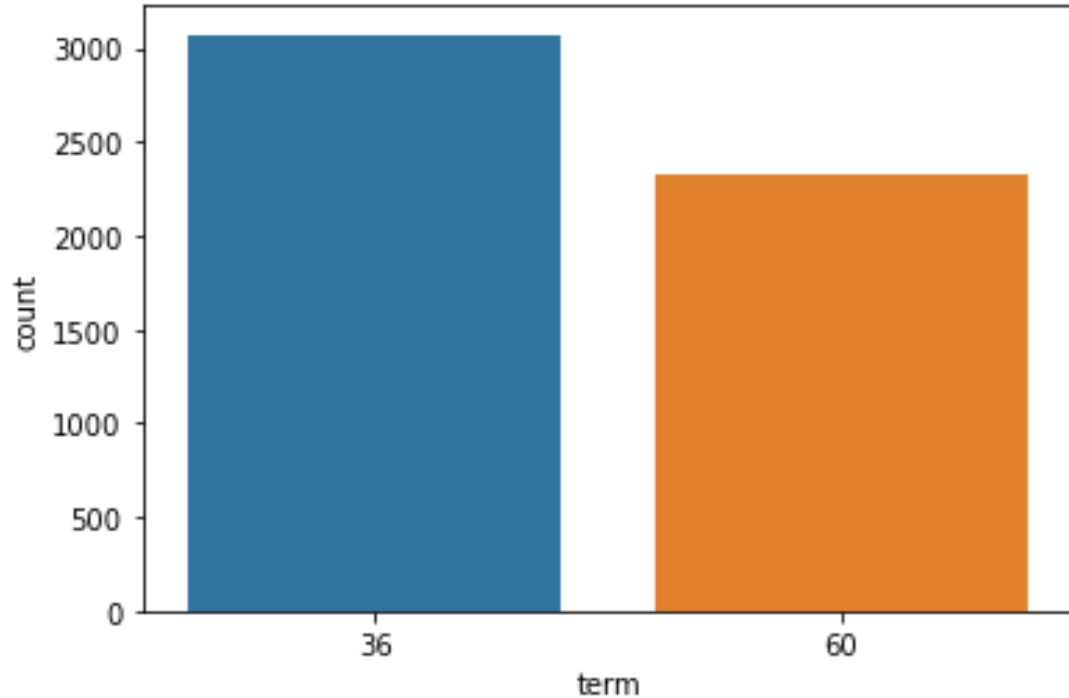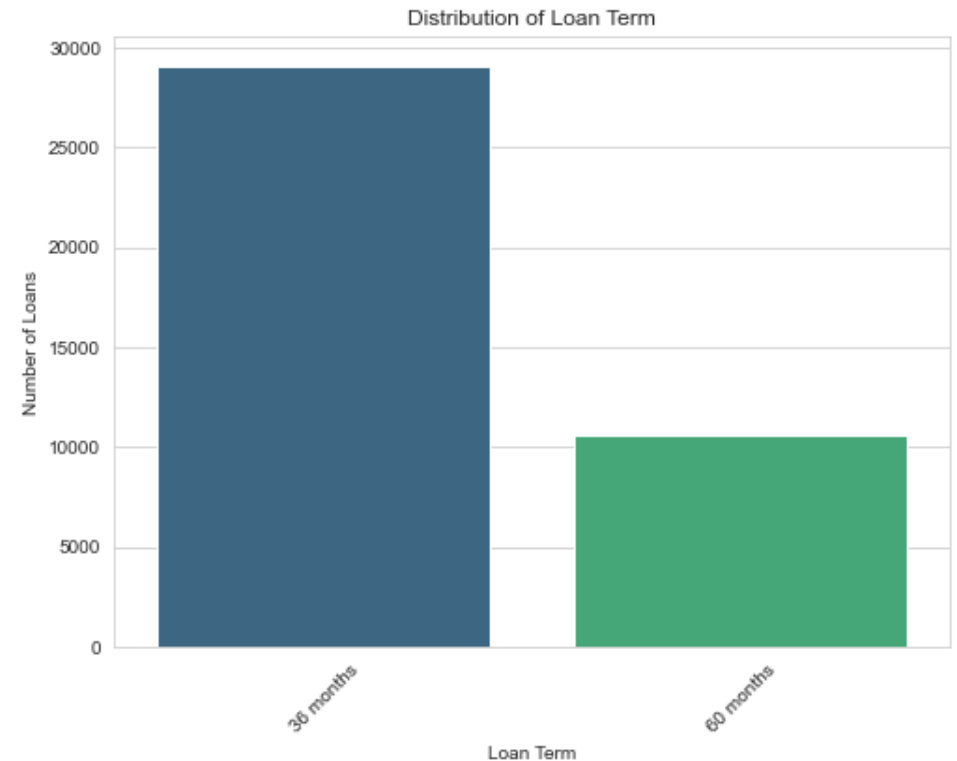
Charged off category
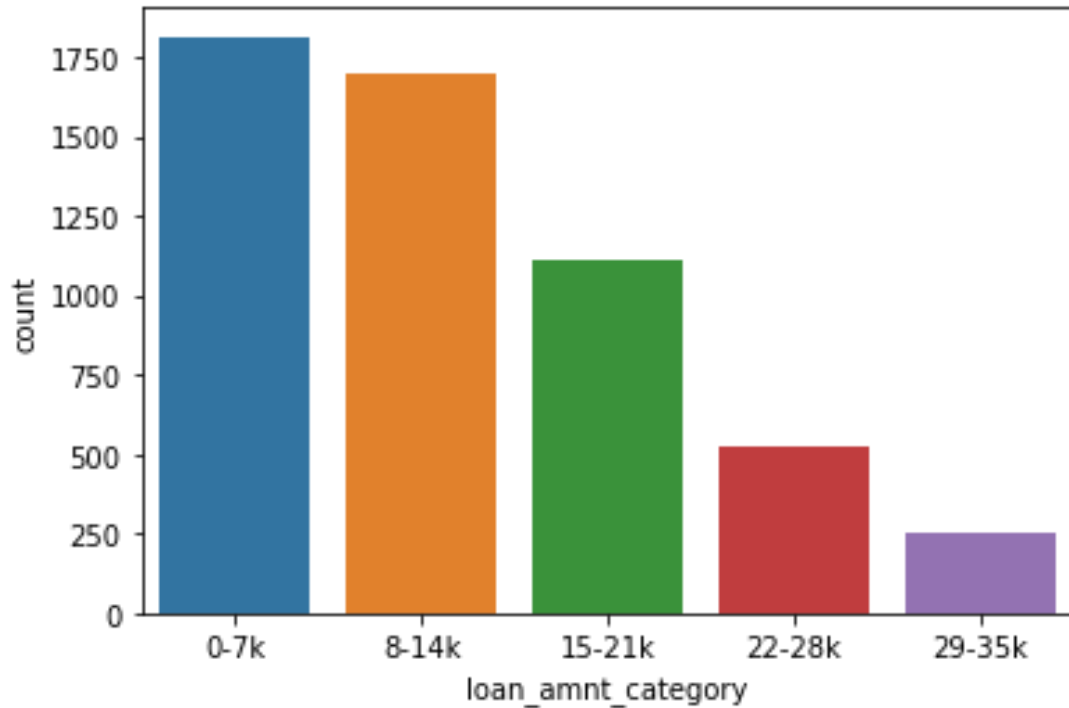
All Categories

Loan Term

- Majority of loans and default have a term of 36 months
- This indicates that borrowers generally prefer shorter-term loans
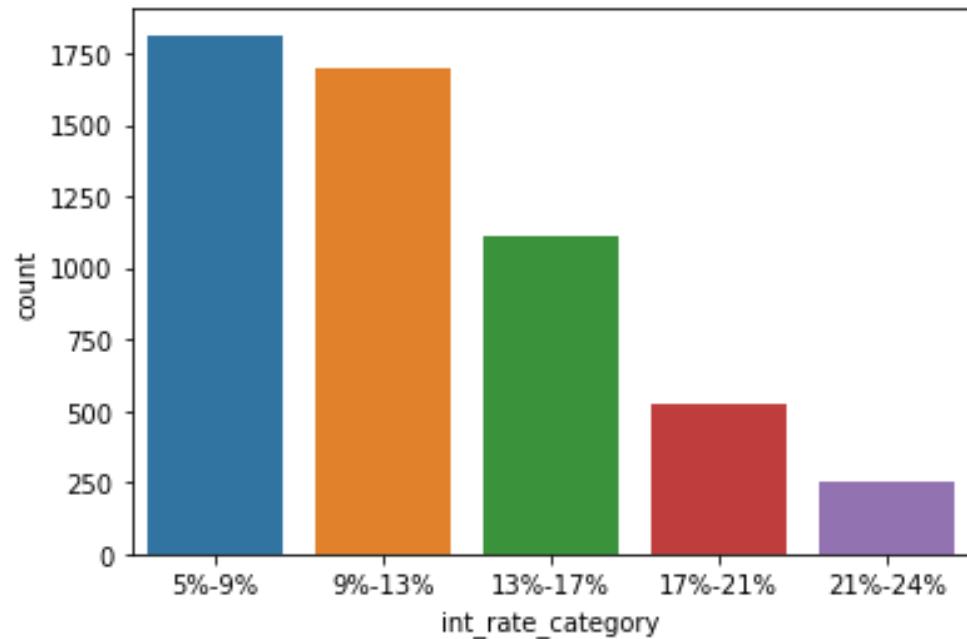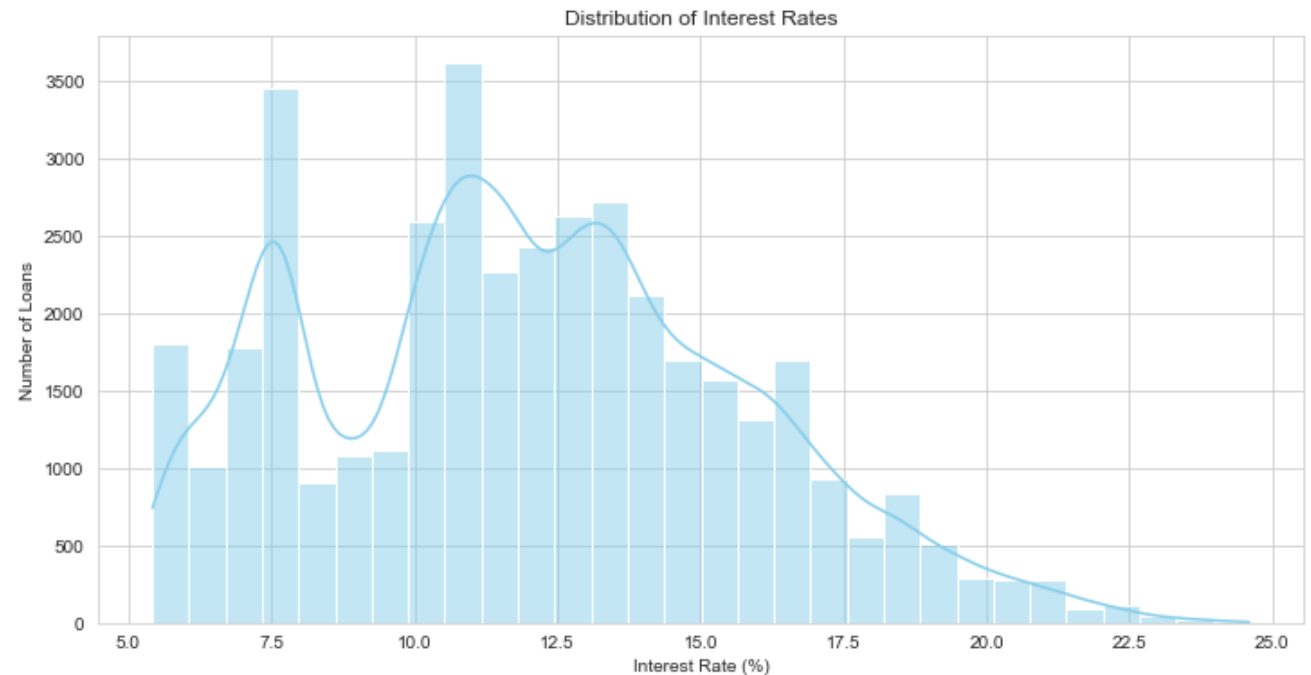
Charged off category

All Categories

Interest Rates

- More defaults when 'int_rate' falls in 5-9%
- The interest rates are mostly distributed between 5% and 25%
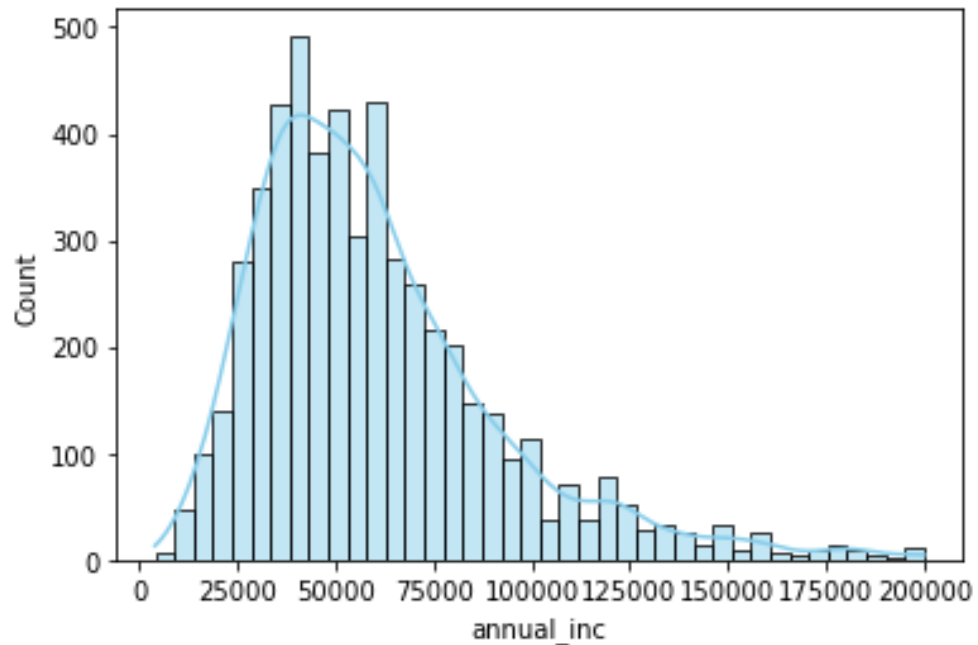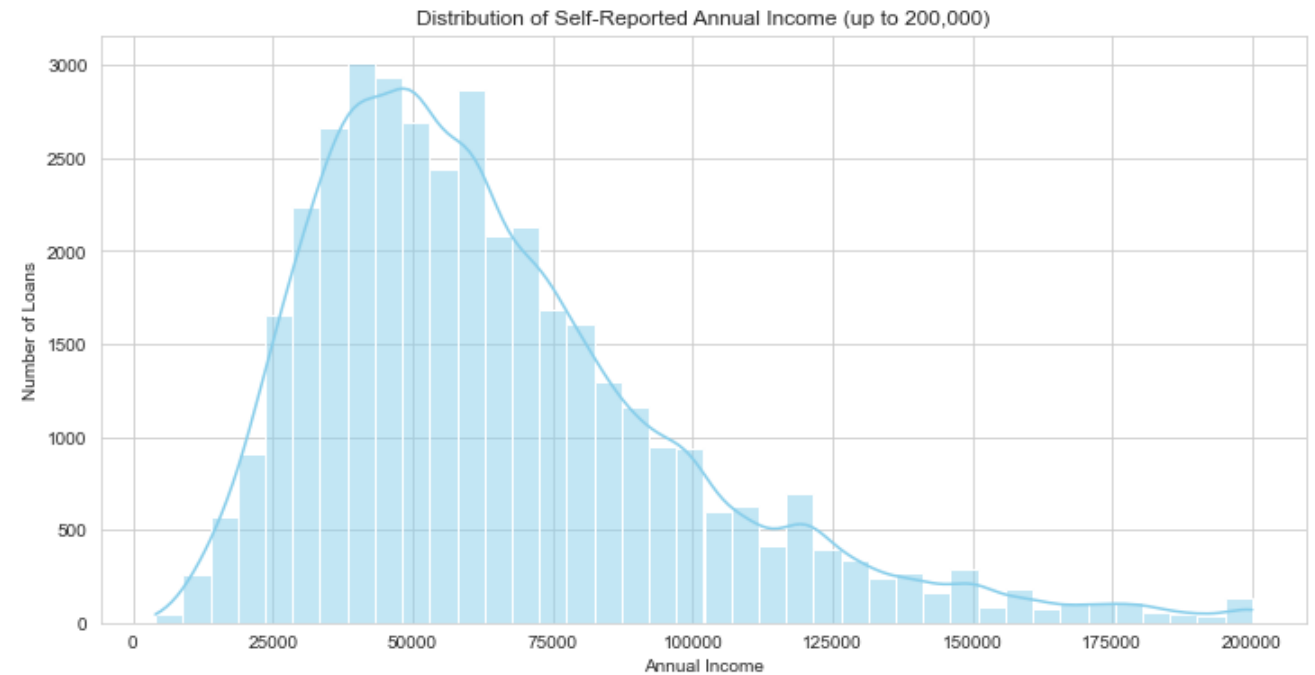
Charged off category

All Categories

Annual Income

- Annual income around 30k -50k has more default

- Most borrowers report an annual income in the lower range, up to 200,000, with a concentration below 100,000

Charged off category

All Categories

Distribution of Self-Reported Annual Income (up to 200,000)
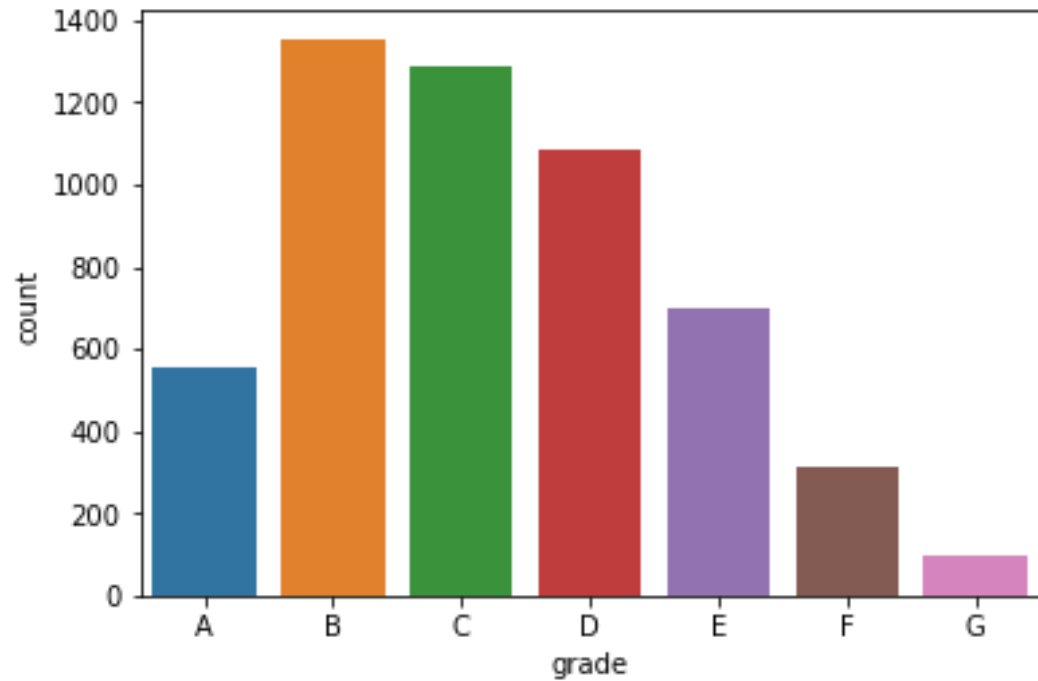
Grade

- More defaults for grade B
- Grades 'G' and 'F' are the least common, suggesting that very few loans are assigned these lower grades

Charged off category

All Categories

Distribution of Loan Grades
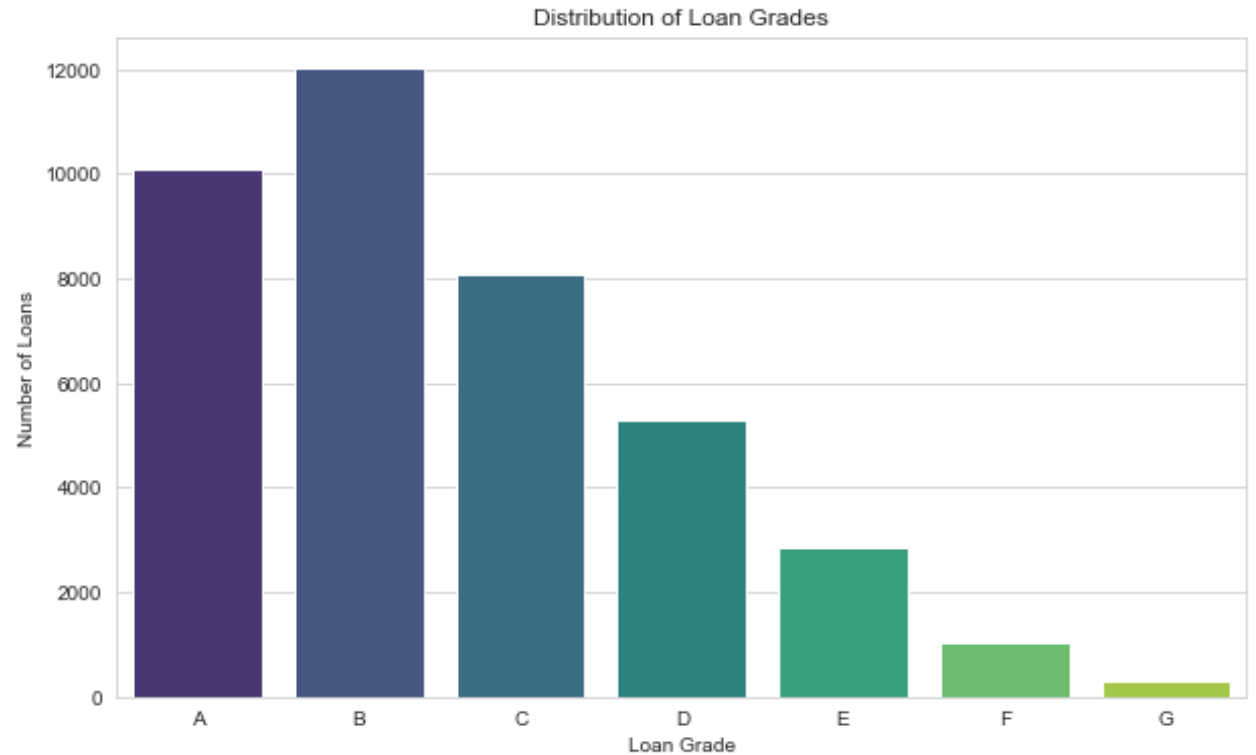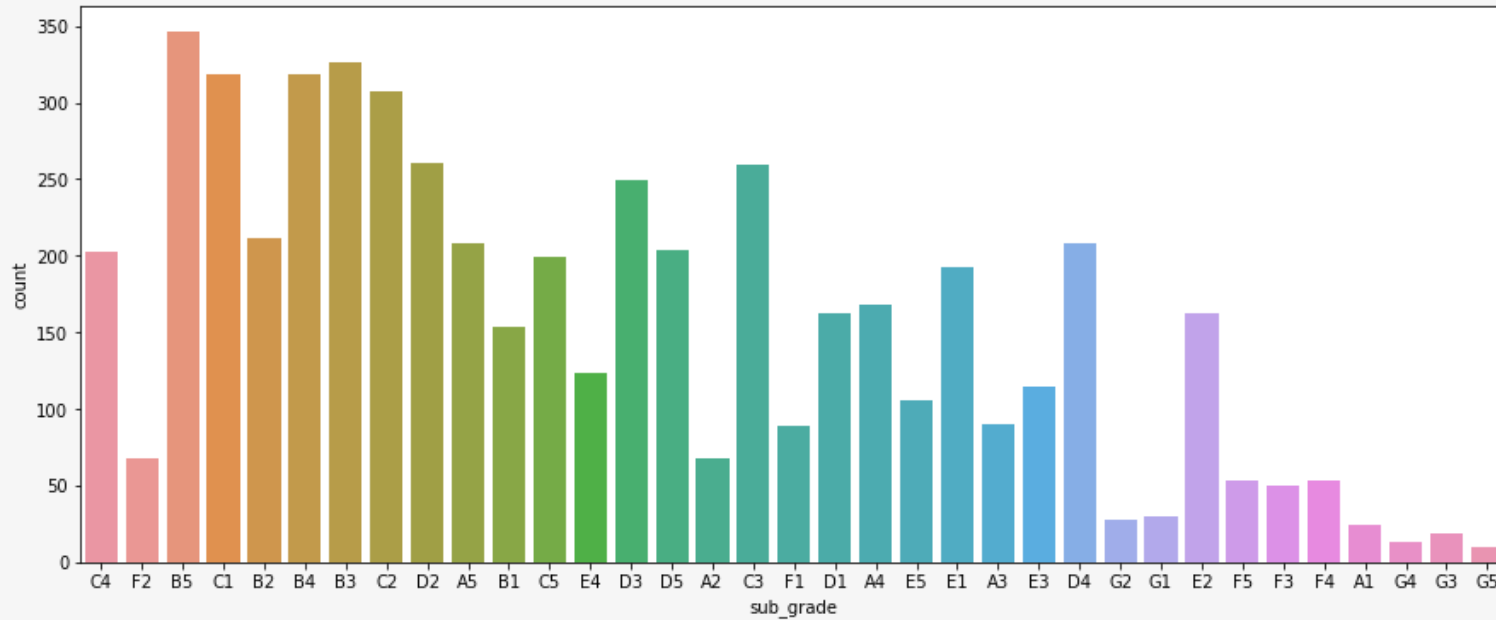
Sub grade for charged off Loans

- More defaults for subgrade B5
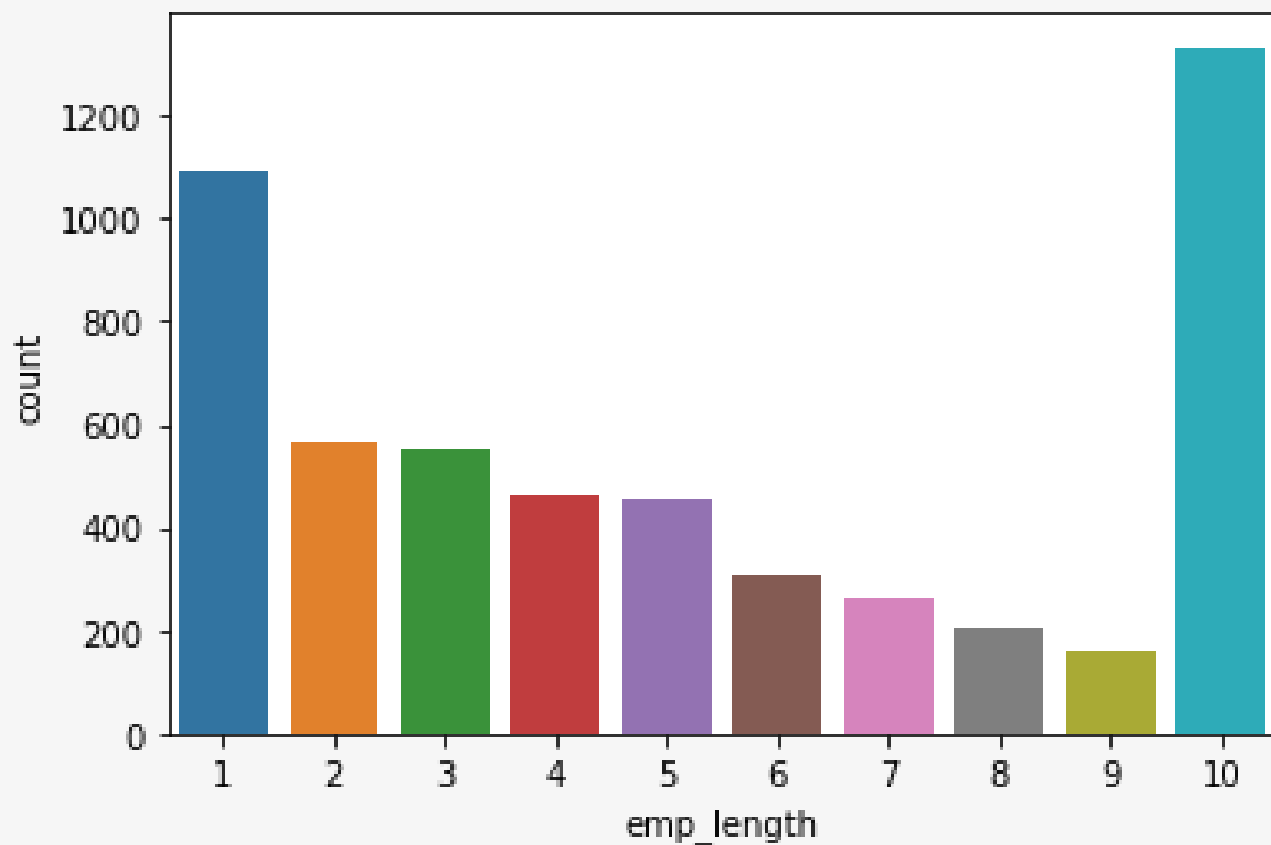
Employment Length for charged off Loans

- More than 10 years have high chance of default
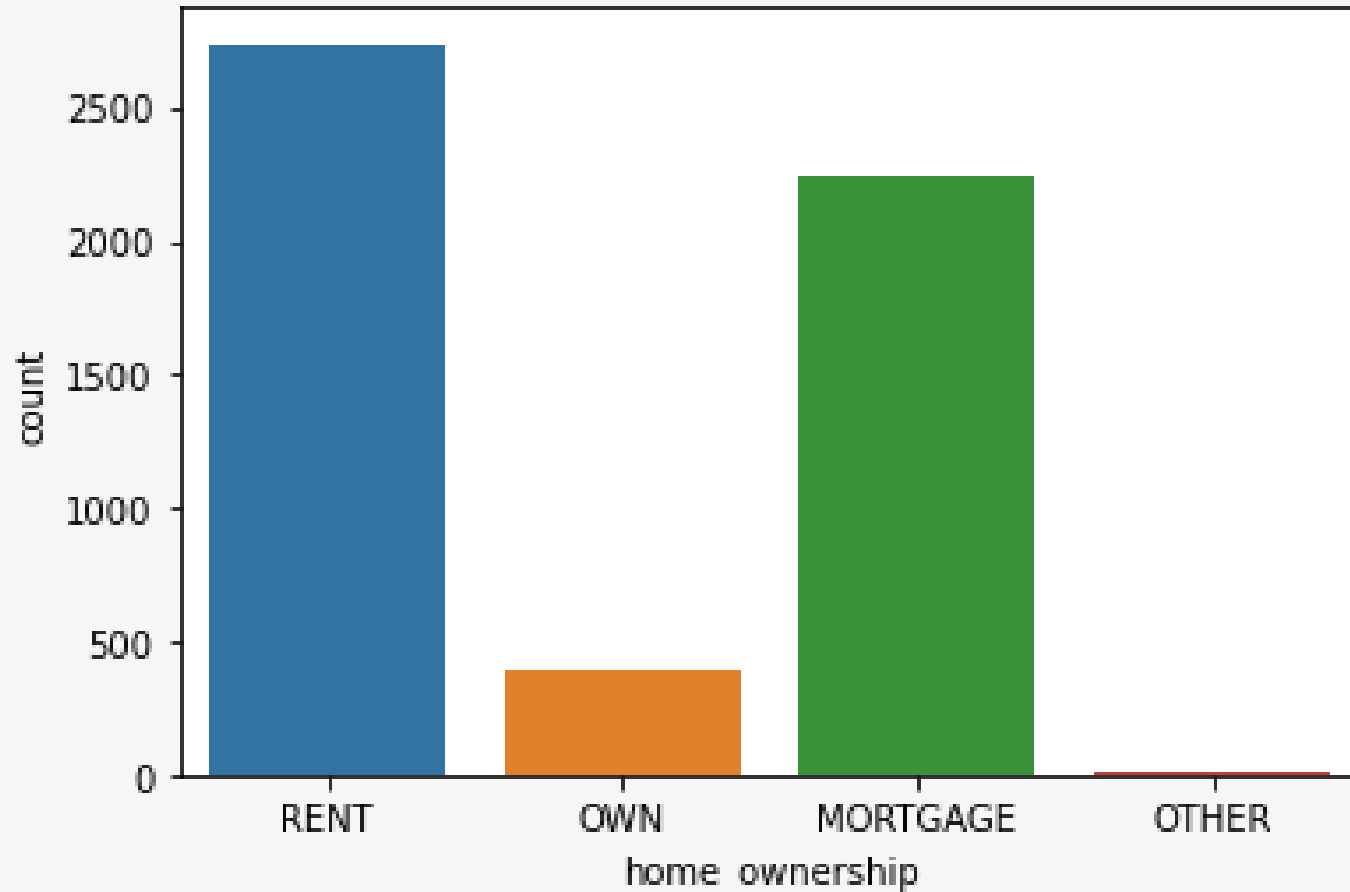
Home Ownership for charged off Loans

- More defaults when 'home_ownership' is 'RENT' and 'MORTGAGE'
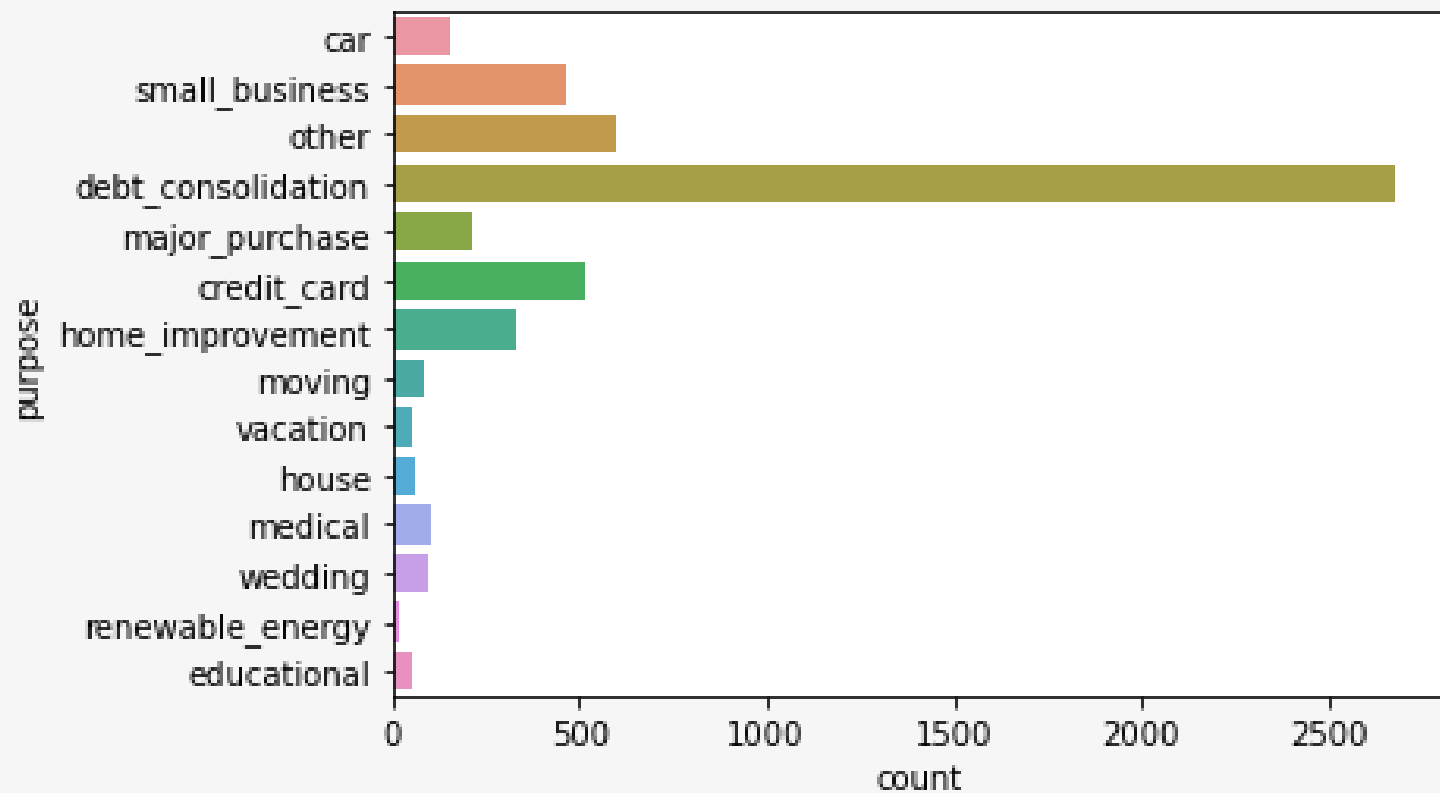
Purpose for charged off Loans

- More defaults when 'purpose' of loan is 'debt_consolidation'
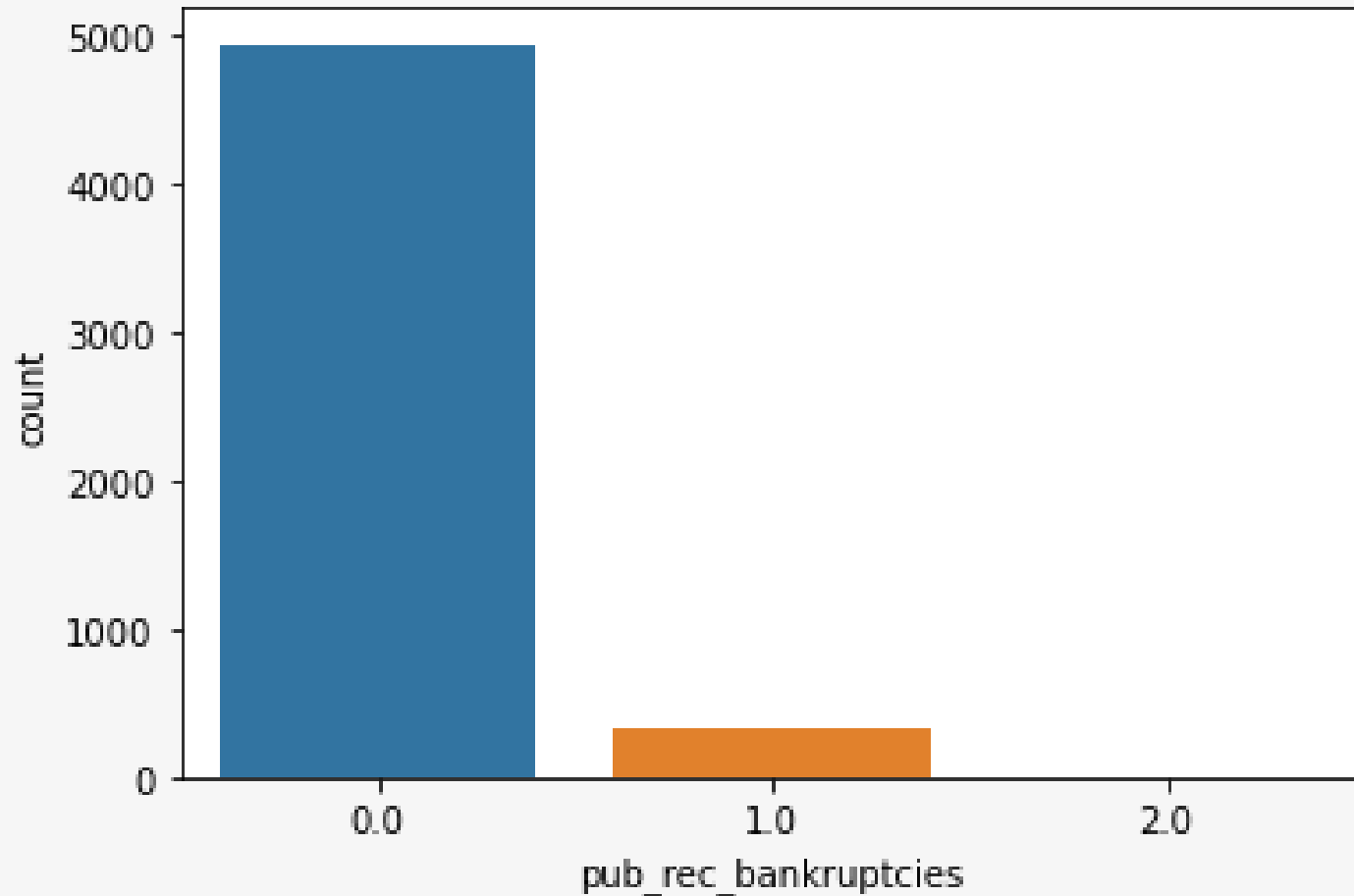
Public record for charged off Loans

- More defaults when 'public_rec_bankruptcies' is 0

- This represents the first default of the customers with the credit company
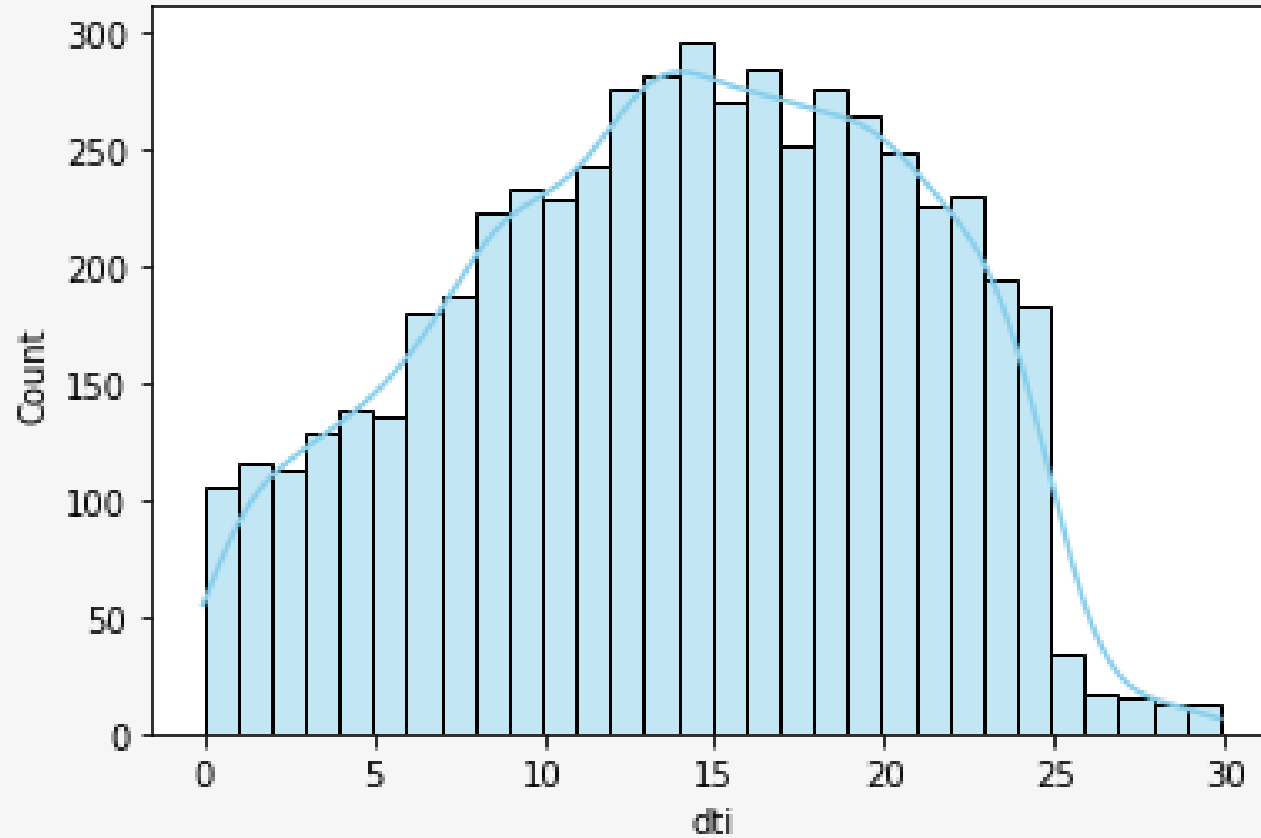
dti for charged off Loans

- More defaults when 'dti' is between 12-18
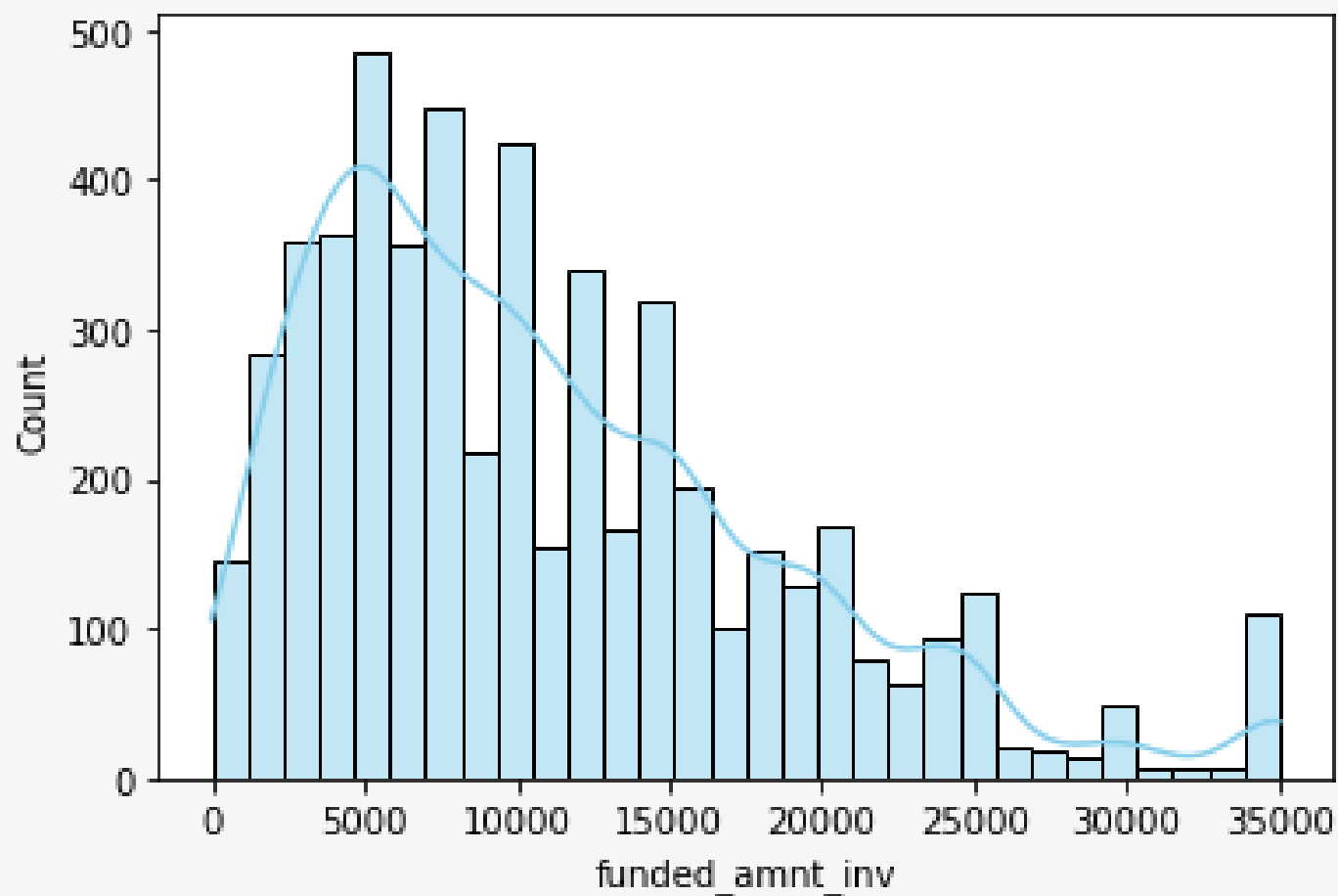
Funded amount inv for charged off Loans

- More defaults when 'loan_amnt' is between 5-10k
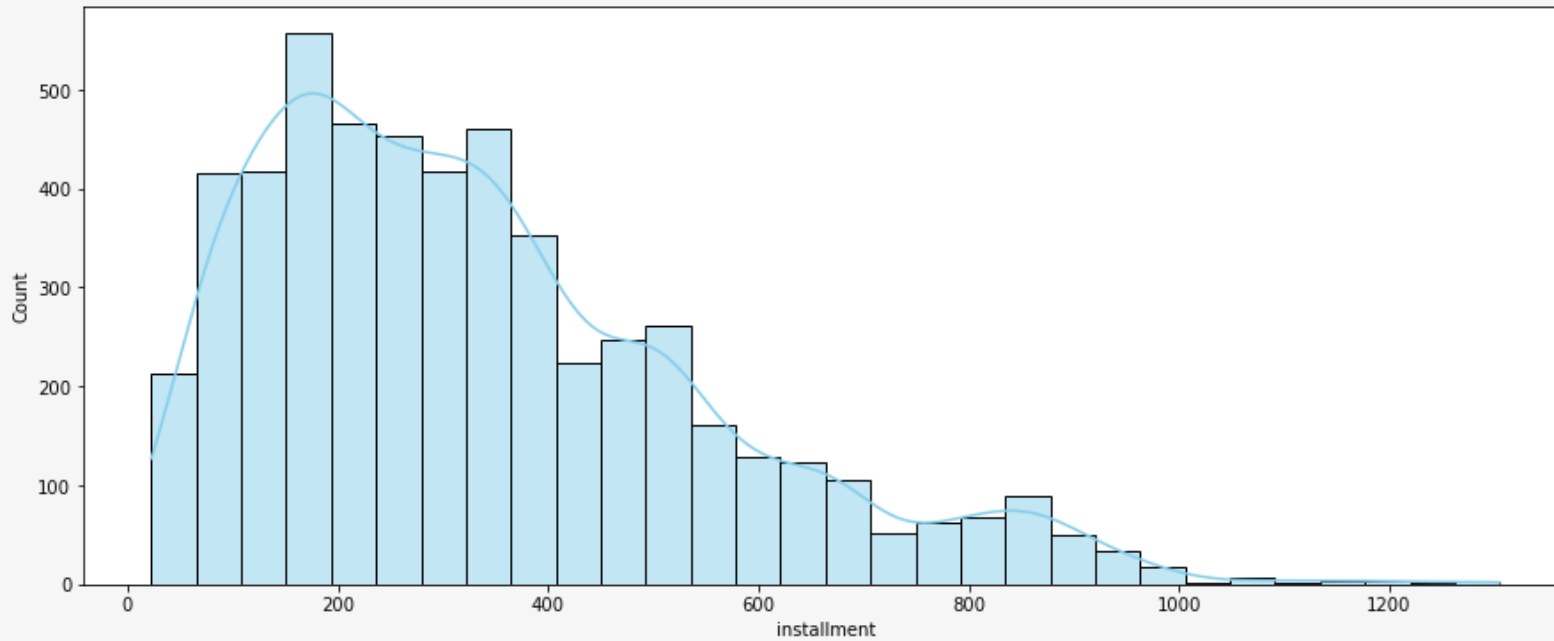
Installment for charged off Loans

- More defaults when 'installment' are between 150 to 300
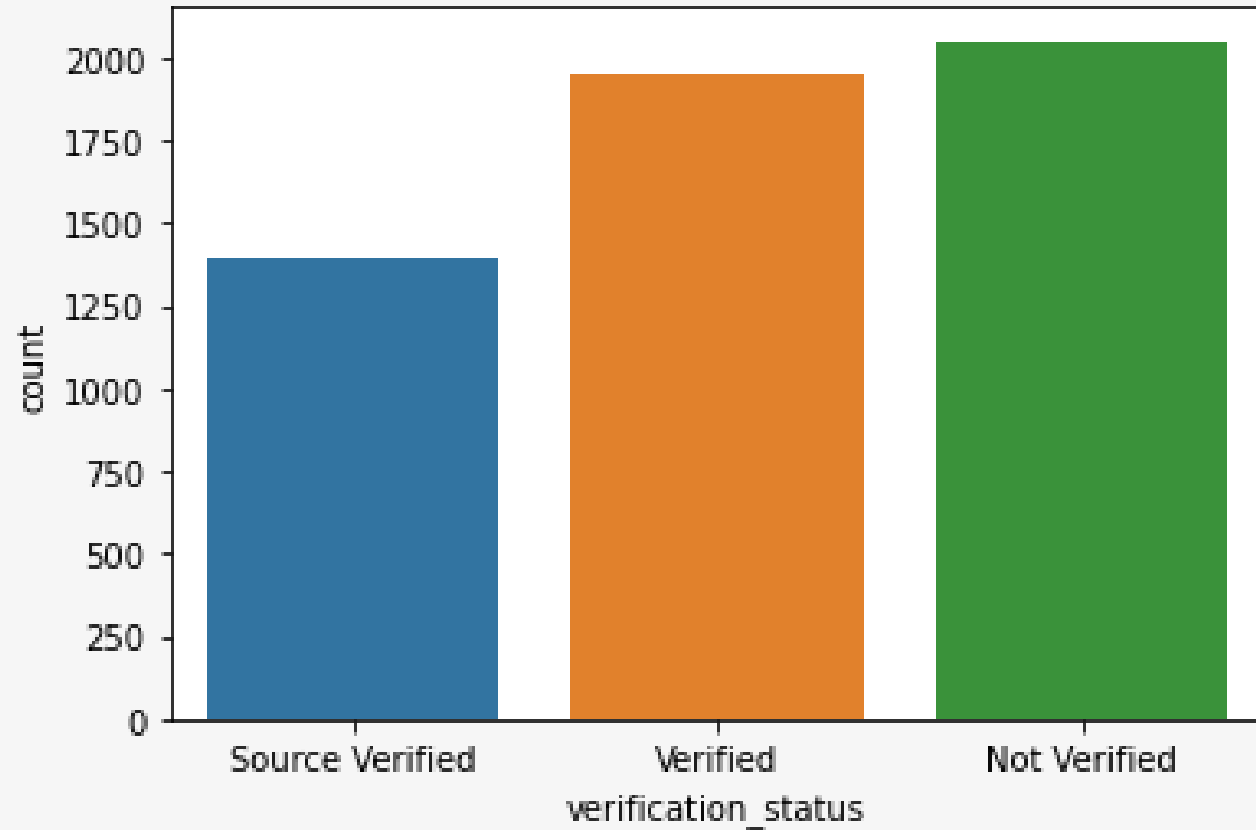
Verification status for charged off Loans

- More defaults when 'verification_status' is 'Not verified'

Univariate Analysis Summary-1

- **Loan Status:** A significant number of loans were defaulted on, with a smaller portion still ongoing.

- **Loan Amount:** Most defaults occurred for loan amounts ranging from 0-7k.

- **Loan Term:** The majority of defaults were for shorter-term loans (36 months).

- **Interest Rates:** Defaults were more common for interest rates in the 5-9% range.

- **Annual Income:** Borrowers with incomes around 30k-50k showed higher defaults.

- **Grade & Subgrade:** Grade 'B' and subgrade 'B5' experienced higher defaults.

- **Employment Length:** Borrowers employed for more than 10 years had a higher chance of defaulting.

Univariate Analysis Summary-2

- **Home Ownership:** Default rates were higher for those who rented or had a mortgage.

- **Purpose:** Debt consolidation was the primary purpose behind most defaulted loans.

- **Public Record:** Those without previous bankruptcies (public record = 0) had higher defaults.

- **DTI:** Borrowers with a DTI between 12-18 showed more defaults.

- **Funded Amount:** Defaults were higher for funded amounts between 5-10k.

- **Installment:** Defaults were more common for installment amounts between 150-300.

- **Verification Status:** Loans that weren't verified experienced more defaults.

Bivariate Analysis
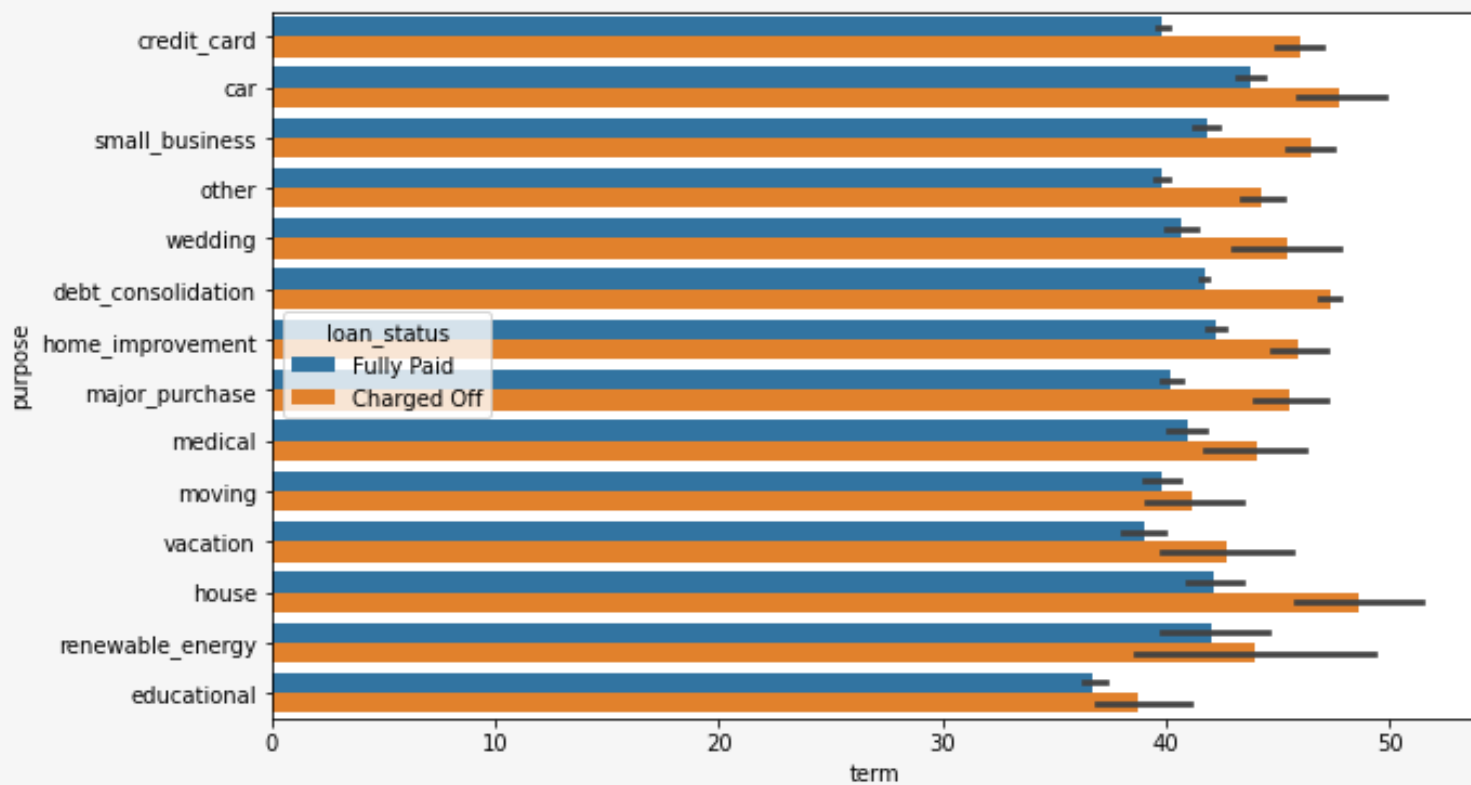
Term vs Purpose

- House, home_improvement, debt_consolidation, car with longer terms has higher default rate
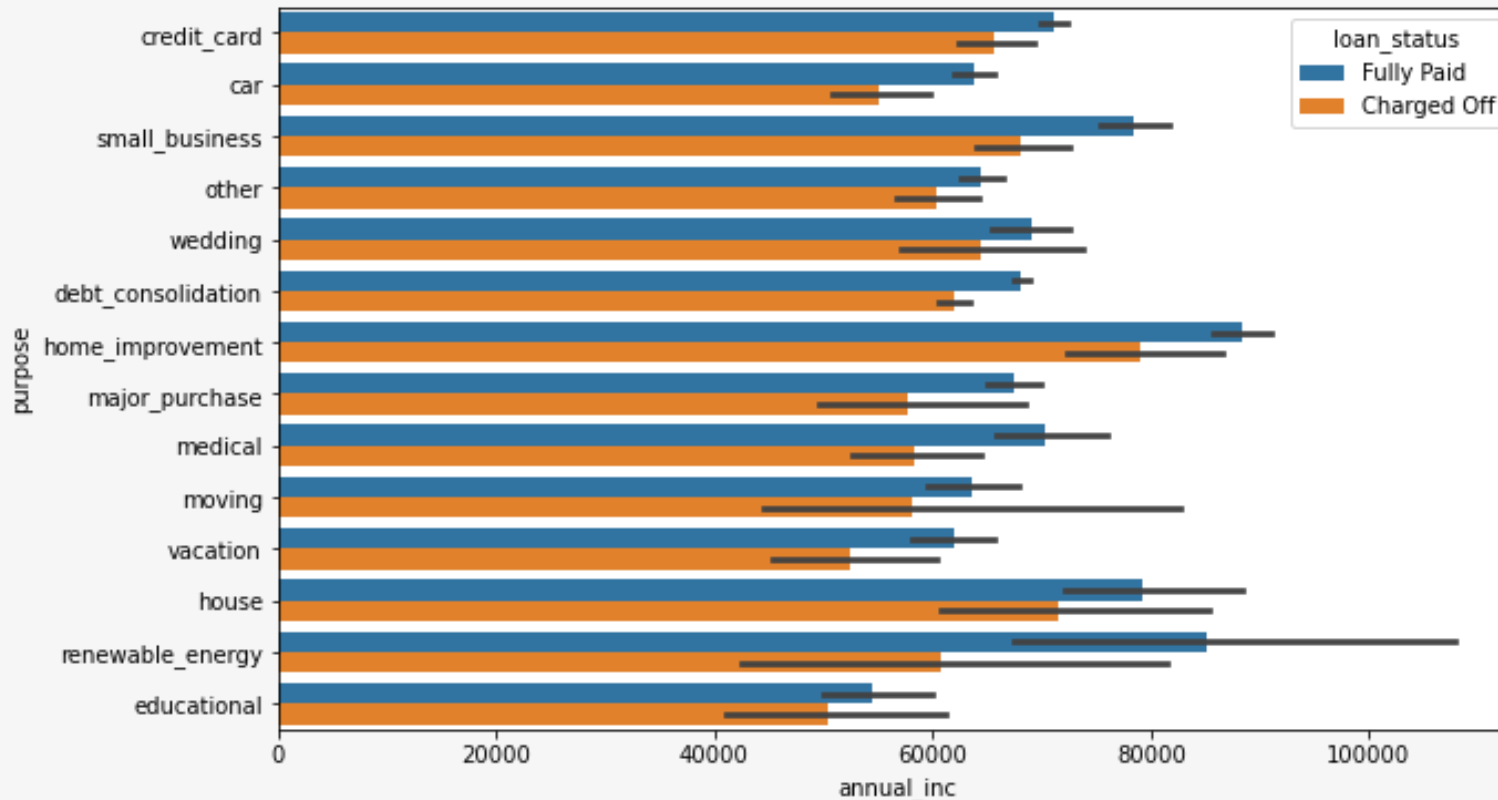
Purpose vs Annual Income

- Applicants with higher salary mostly applied loans for "home_improvment", "house", "renewable_energy" and "small_businesses" have higher default rate
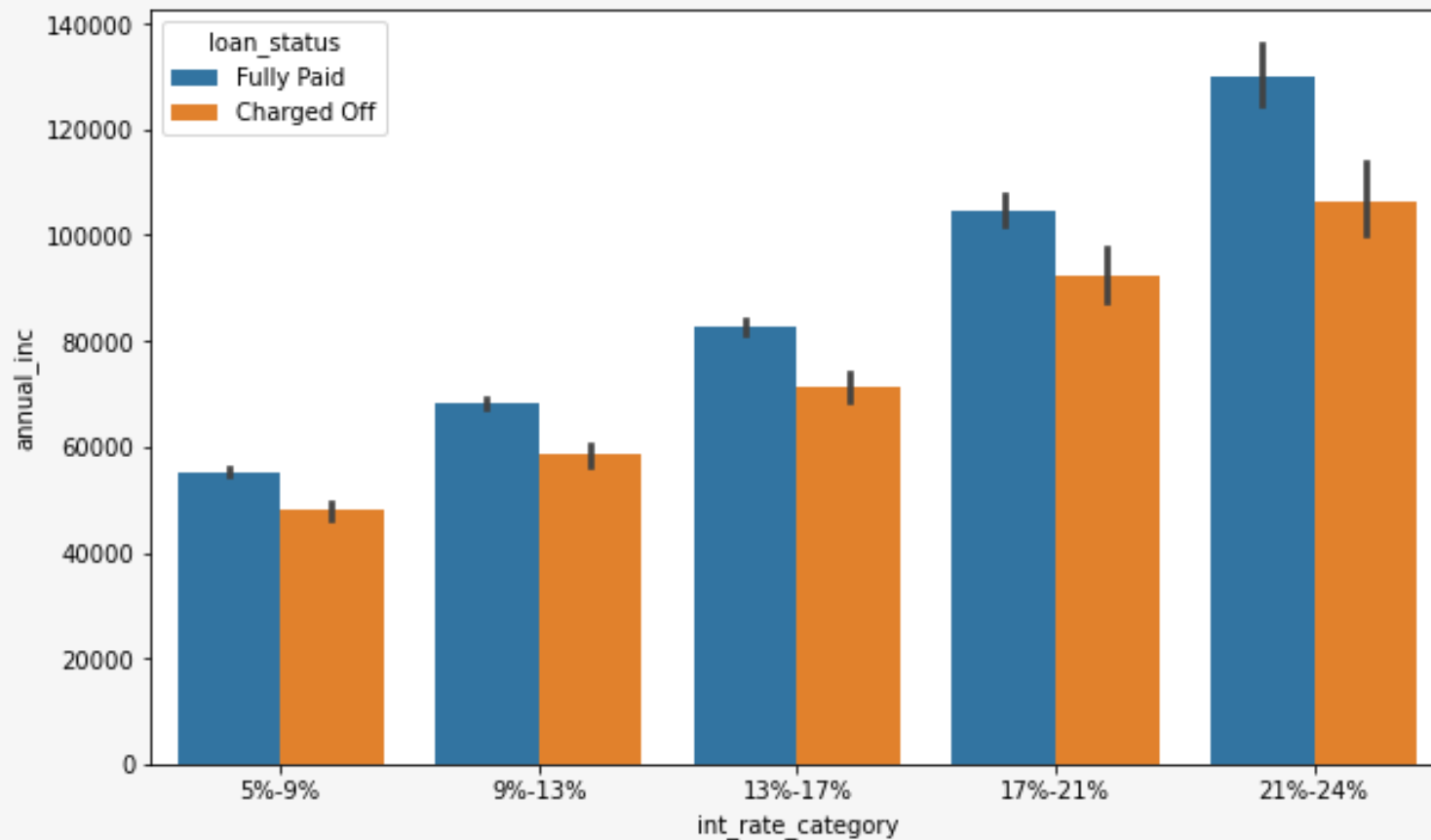
Interest rate vs Annual Income

- Higher salary range gets loan on higher interest and higher defaults

- Higher defaults when Applicants who receive interest at the rate of 21-24%
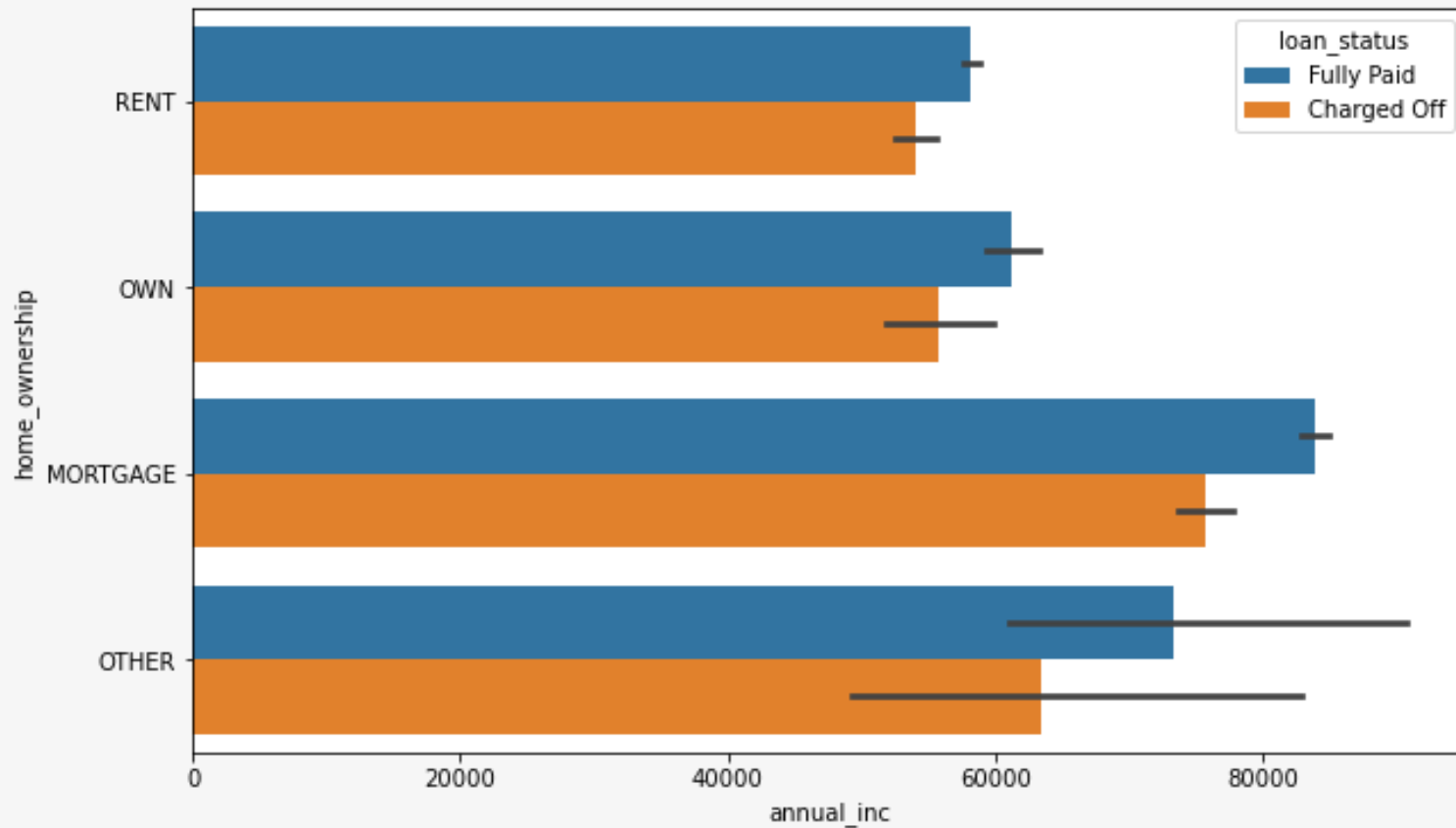
Home Ownership vs Annual Income

- Applicants whose home ownership is 'MORTGAGE' or 'OTHER' and have income of 60-70k have higher defaults
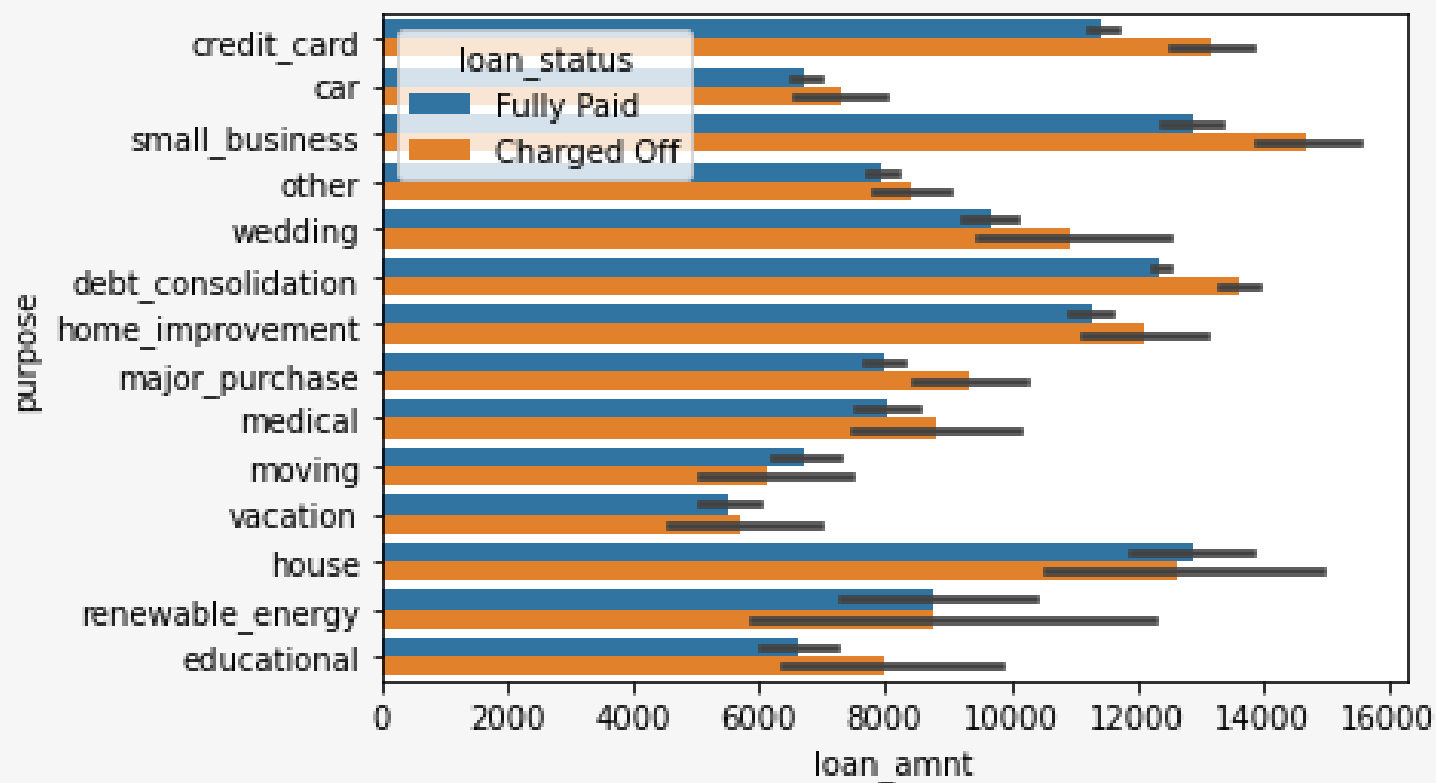
Purpose vs Loan Amount

- Applicants who have taken a loan for 'small business', 'credit_card', 'debt_consolidation' and the loan amount is greater than 12k have high default
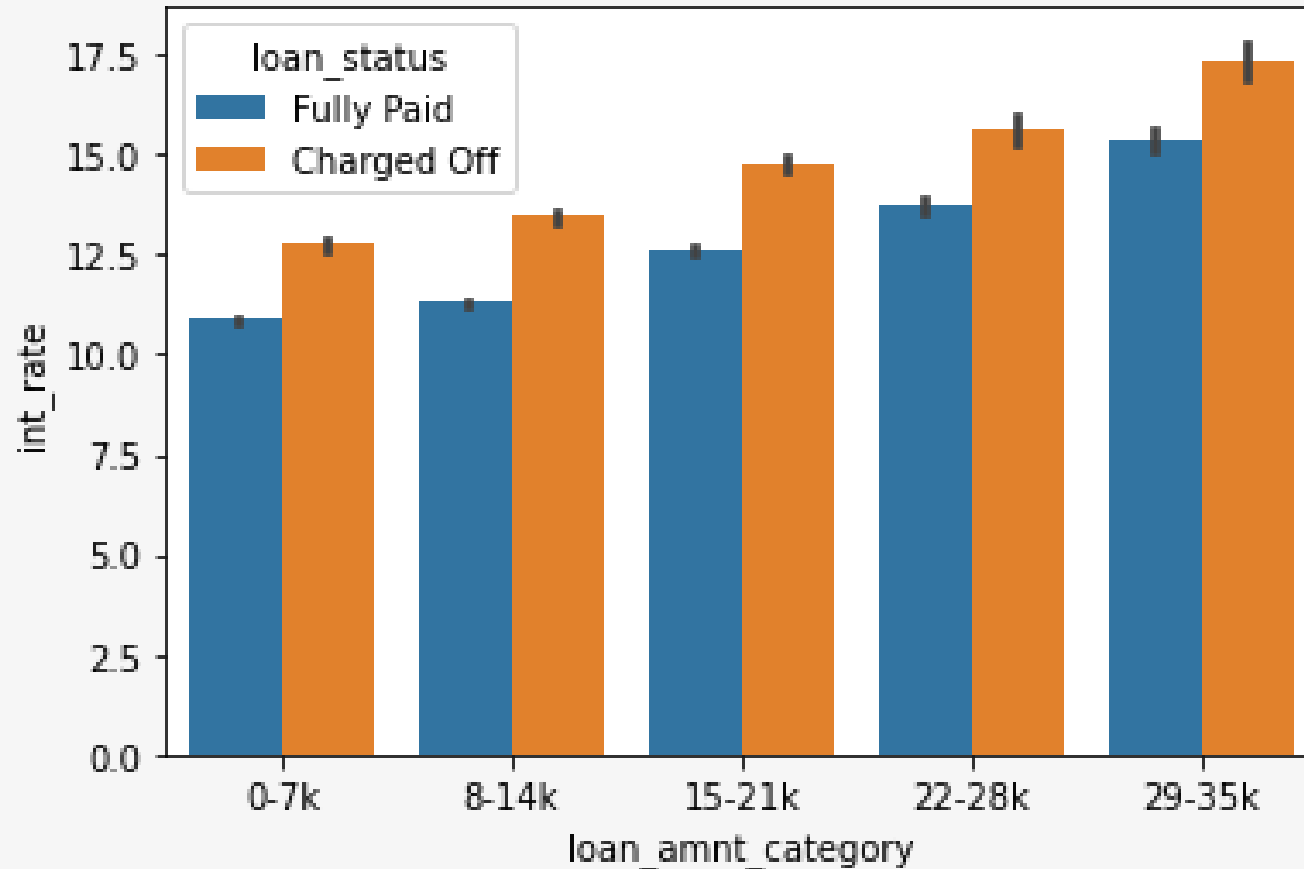
Interest rate vs Loan Amount Category

- The default rate seems to be similar for interest rate and loan amount categories combinations. This doesn't seem to impact the default rate

Grade vs Loan Amount

- When grade is F and loan amount is between 15k-20k have high default
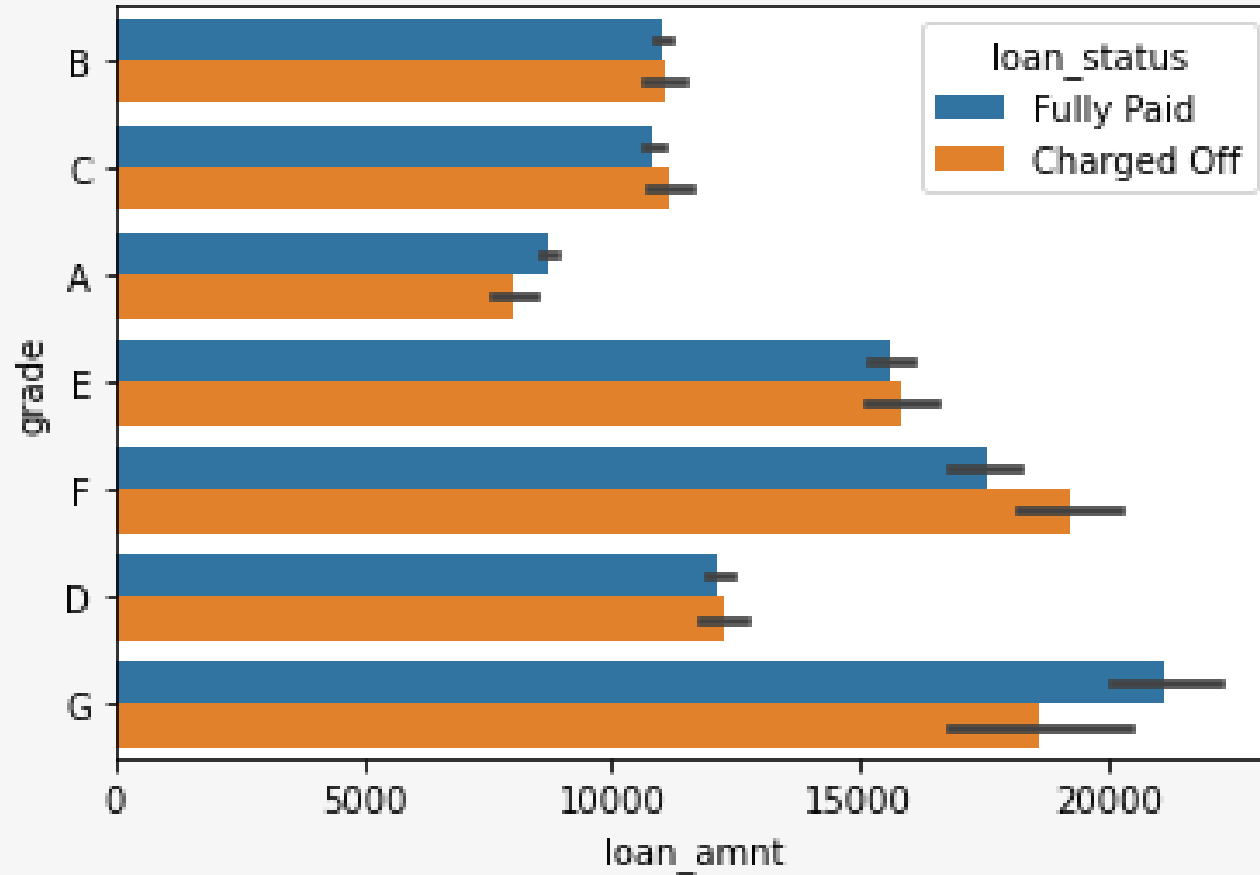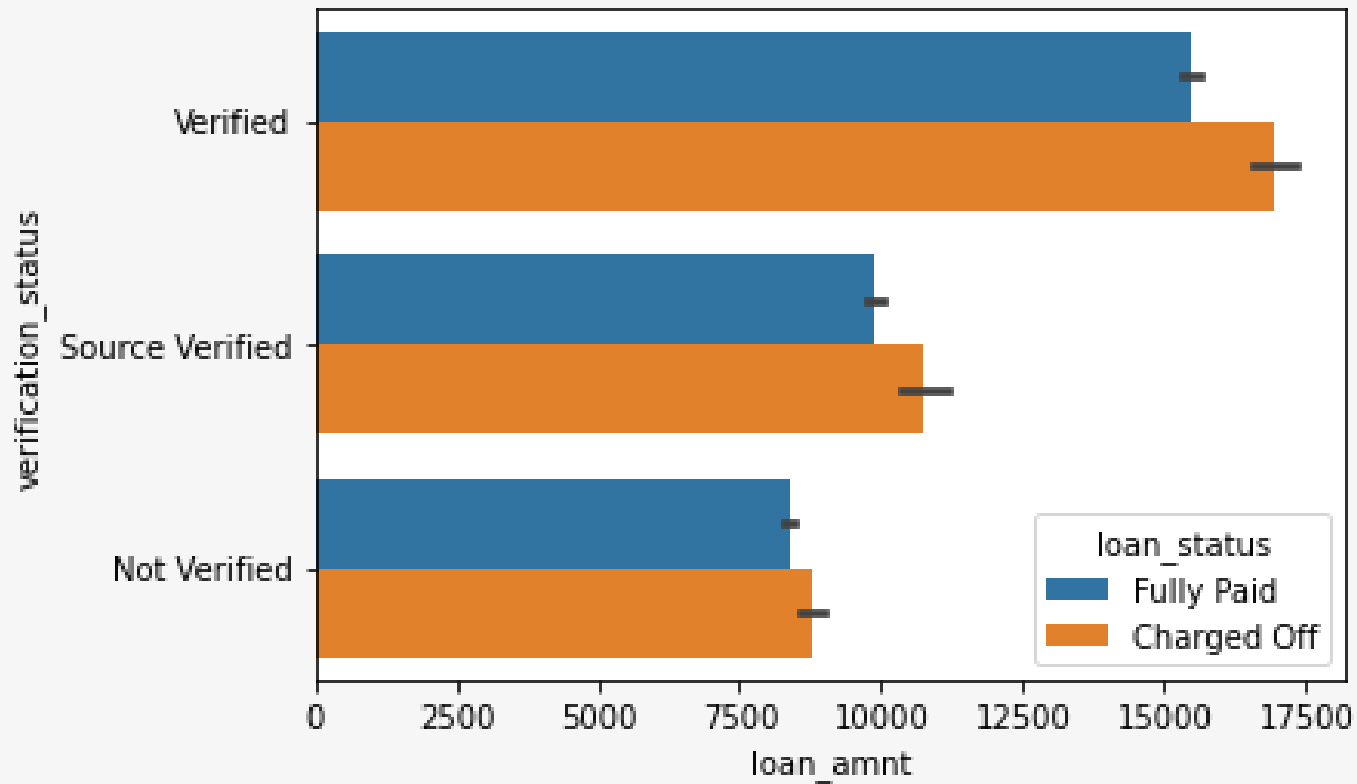
Verification status vs Loan amount

- Looking at the verification status data, verified loan applications tend to have higher loan amount

- Which might indicate that the firms are first verifying the loans with higher values

- When the loan is verified, and loan amount is above 16k

- Verified loans have high default rate
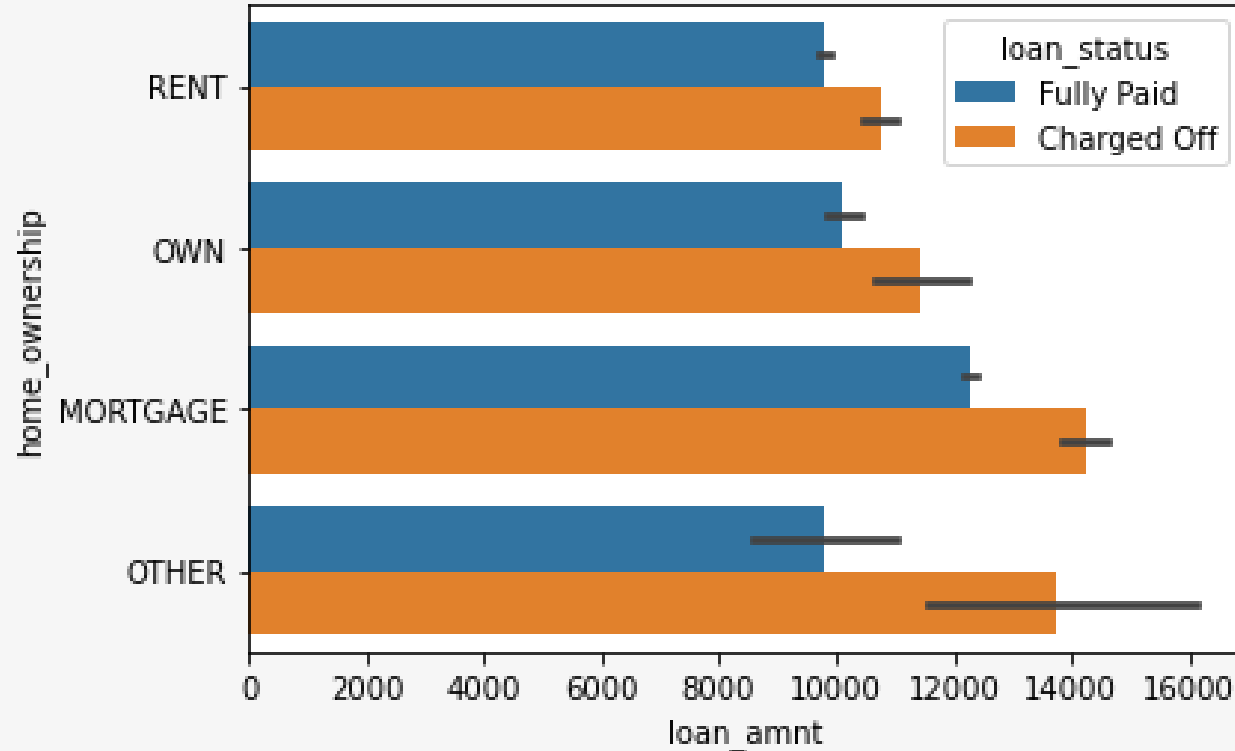
Home ownership vs Loan amount

- Applicants whose home ownership is 'OTHER' or 'MORTGAGE and have loan of 12-16k have high default
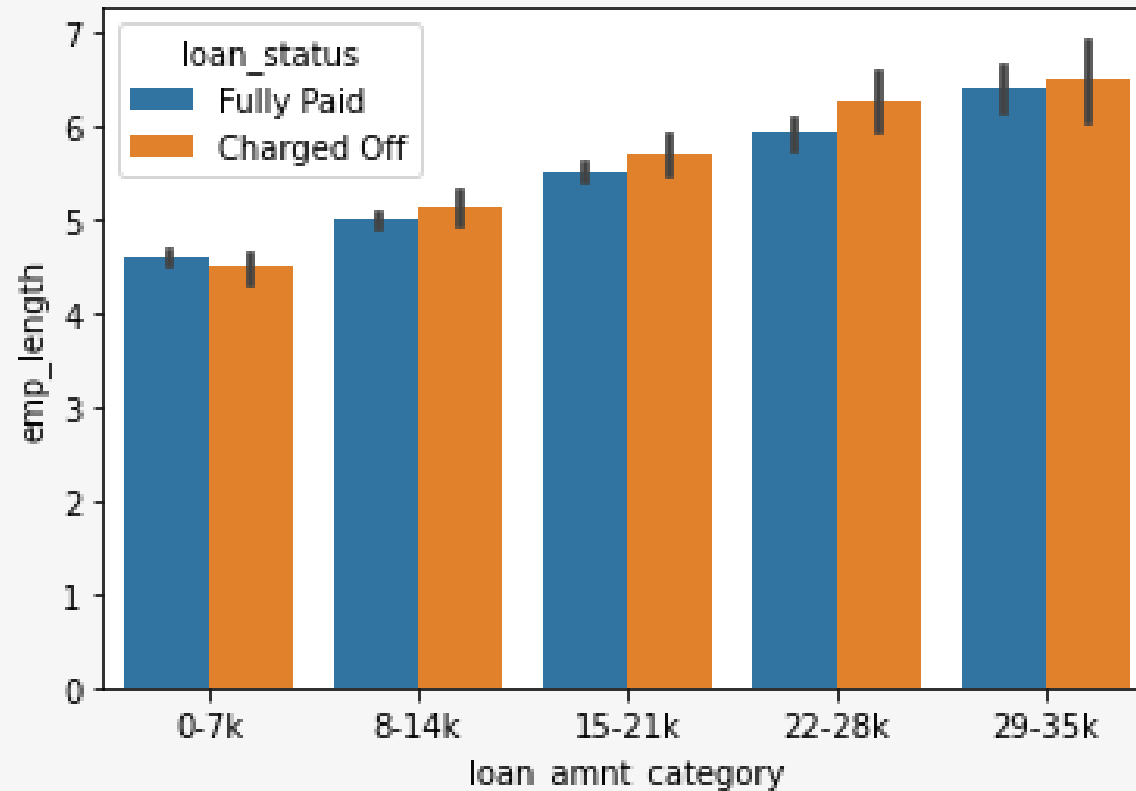
Emp length vs Loan amount category

- Employees with longer working history got the loan approved for a higher amount

- Employment length and loan amount combination does not seem to have much impact on default rate
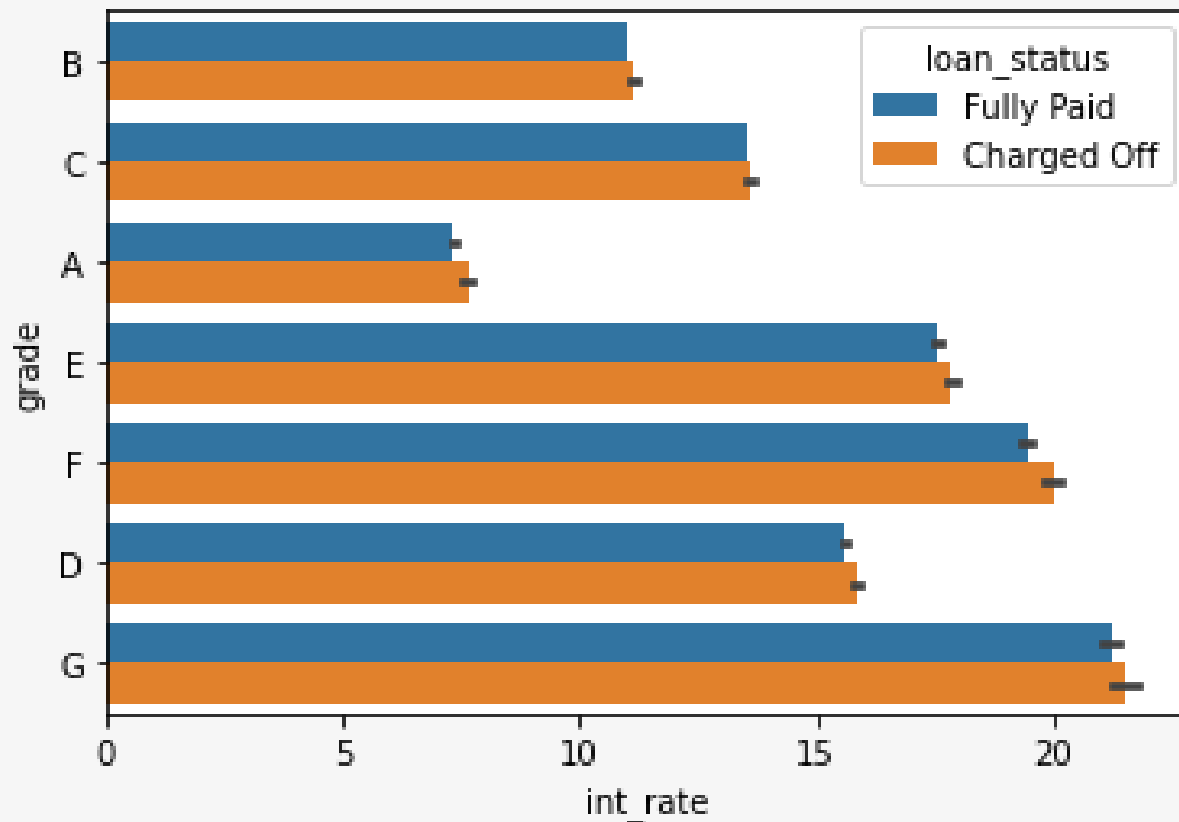
Grade vs Interest rate

- The interest rate for charged off loans is higher than that of fully paid loans in all the grades

- Looking deeply, there doesn't seem to be much cor-relation between grade and interest rate as all the grades seems to have the similar default rate as compared to paid loans

Bivariate Summary - 1

- **Term vs. Purpose:** Longer-term loans for purposes like housing, home improvement, debt consolidation, and car have a higher default rate

- **Purpose vs. Annual Income:** Applicants with higher incomes, especially those taking loans for home improvement, house, renewable energy, and small businesses, show a higher default rate

- **Interest Rate vs. Annual Income:** Higher salaried individuals tend to receive loans at higher interest rates and show higher defaults, especially when the interest rate is between 21-24%

- **Home Ownership vs. Annual Income:** Applicants with a mortgage as their home ownership type and an income range of 60-70k have higher defaults

- **Purpose vs. Loan Amount:** Loans taken for small business purposes and amounts greater than 14k show higher defaults

Bivariate Summary - 2

- **Interest Rate vs. Loan Amount Category:** Not much impact on the default rate as all the categories seems to have similar trend in default rate.

- **Grade vs. Loan Amount:** Loans with a grade of 'F' and amounts between 15k-20k show higher defaults.

- **Verification Status vs. Loan Amount:** Verified loans tend to have higher amounts, indicating a preference for verifying larger loans. Verified loans above 16k in amount show higher defaults.

- **Home Ownership vs. Loan Amount:** Applicants with a mortgage and loans between 14k-16k have a high default rate.

- **Employment Length vs. Loan Amount Category:** Not much impact on the default rate as all the categories seems to have similar trend in default rate.

- **Grade vs. Interest Rate:** Charged off loans generally have higher interest rates across all loan amount groups. Not much impact on the default rate as all the categories seems to have similar trend in default rate.

Other Data Insights
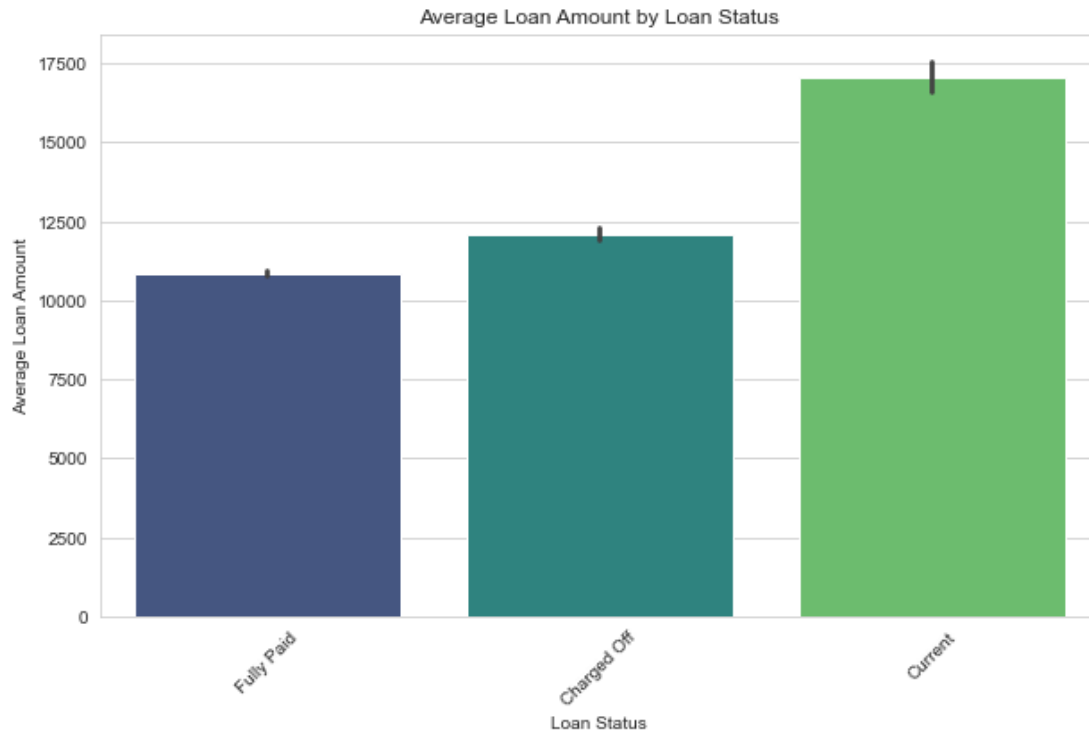
Average Loan Amount by Loan Status

- The bar plot depicts the average loan amount by loan status

- Loans that are "Charged Off" (defaulted) have a slightly higher average loan amount than those that are "Fully Paid"

- Loans that are "Current" have the highest average loan amount among the three statuses

- This suggests that larger loans might have a slightly higher risk of default, but this difference is not solidly taken
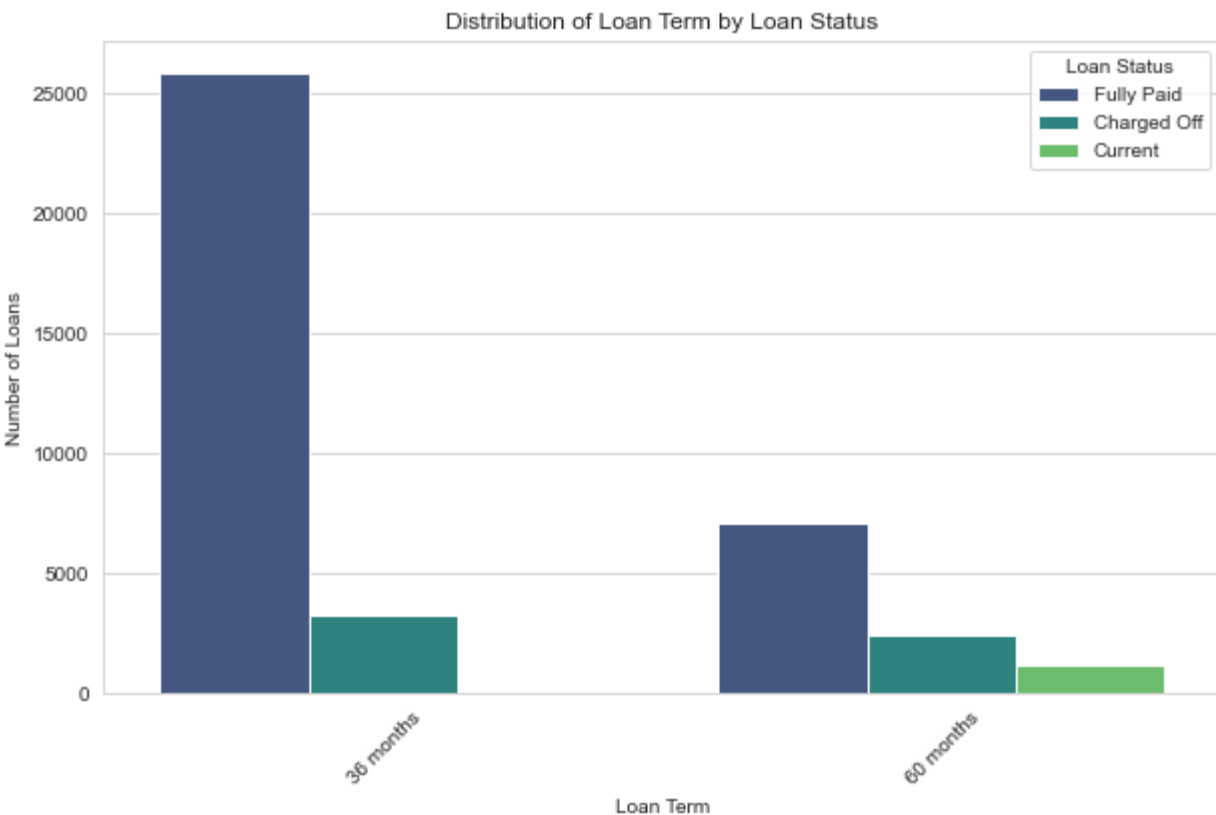
Loan Term by Loan Status

- The plot showcases the distribution of loan terms by loan status

- **For 36-month loans:**
  - A majority are "Fully Paid"
  - A smaller portion are "Charged Off", and an even smaller subset is "Current"

- **For 60-month loans:**
  - The number of "Charged Off" loans is relatively higher compared to the 36-month loans, suggesting a higher risk associated with longer-term loans
  - The ratio of "Fully Paid" to "Charged Off" is narrower for 60-month loans compared to 36-month loans

- The number of "Charged Off" loans is relatively higher compared to the 36-month loans, suggesting a higher

Average Interest Rate by Loan Status
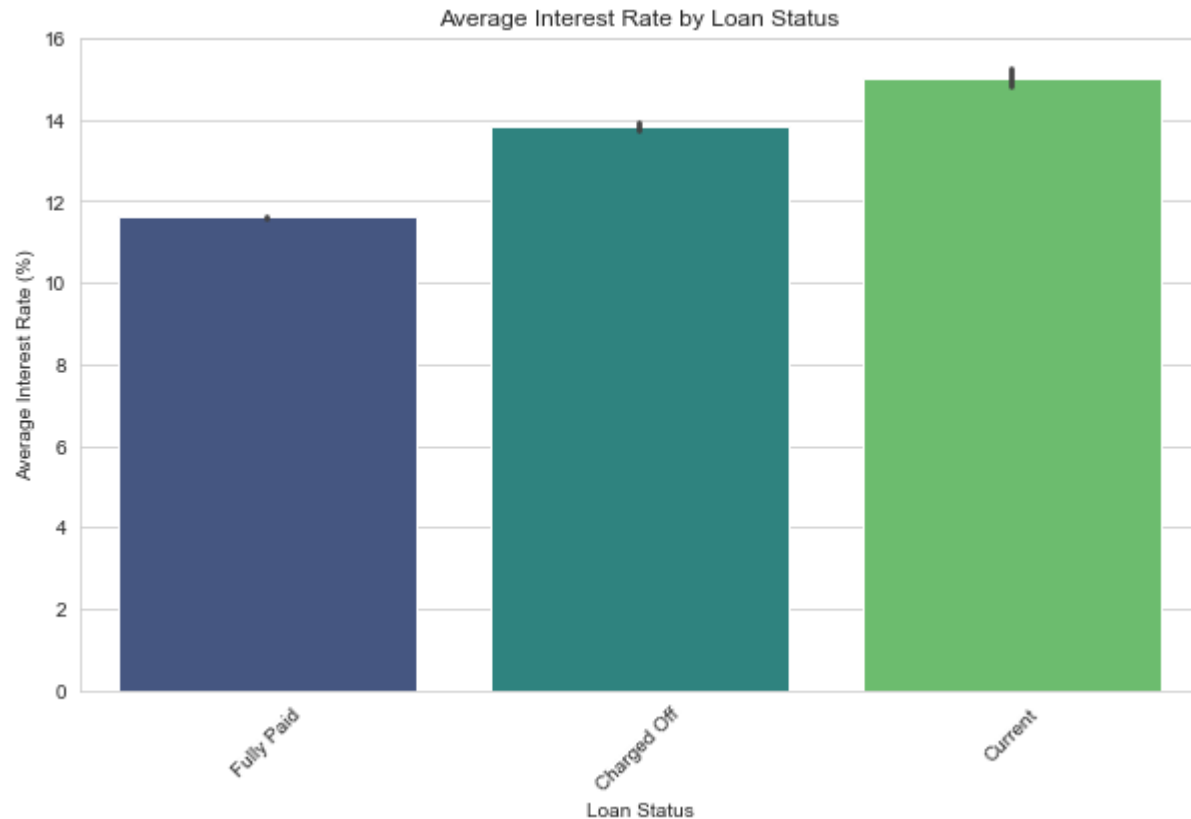
Interest Rate by Loan Status

- The bar plot showcases the average interest rate by loan status

- Loans that are "Charged Off" (defaulted) have a considerably higher average interest rate than those that are "Fully Paid"

- Loans that are "Current" also have a slightly higher average interest rate compared to "Fully Paid" loans, but less than "Charged Off" loans

- This suggests that higher interest rates are associated with a higher risk of default.

- Borrowers with higher interest rates might struggle more with repayments, leading to defaults
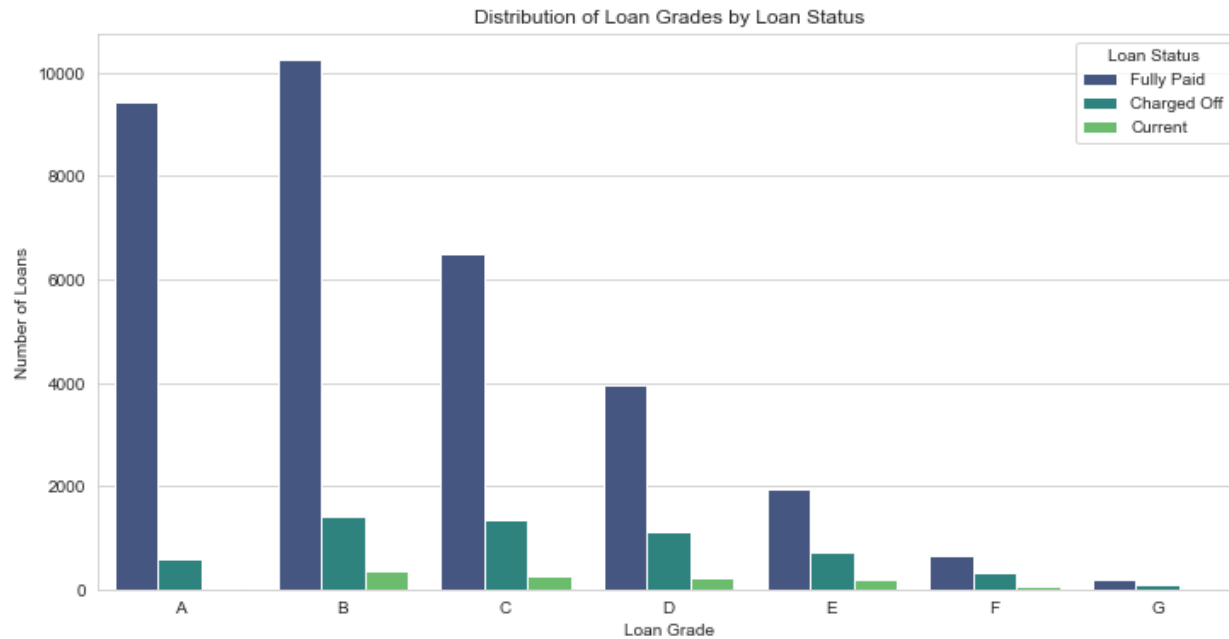
Loan Grades by Loan Status

- The plot showcases the distribution of loan grades by loan status

- For higher grades like 'A' and 'B':
  - A large majority are "Fully Paid", with relatively fewer "Charged Off" loans

- As we move to lower grades like 'C', 'D', and especially 'E', 'F', and 'G':
  - The number of "Charged Off" loans increases in proportion, making the ratio of "Fully Paid" to "Charged Off" narrower

- This indicates that lower-grade loans, which are typically associated with higher risk, have a greater likelihood of defaulting

Median Annual Income by Loan Status
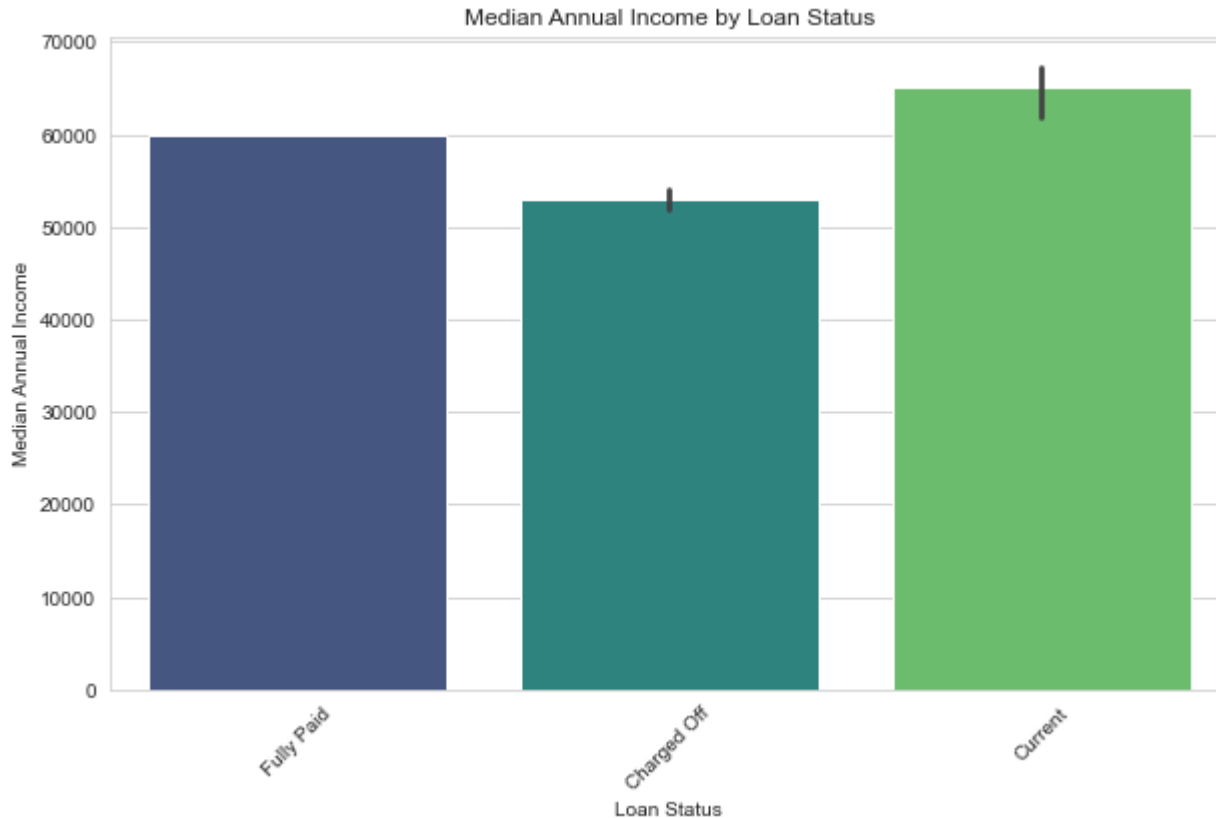
Annual Income by Loan Status

- The bar plot illustrates the median annual income by loan status

- Borrowers with "Charged Off" loans (defaulted) have a slightly lower median annual income compared to those with "Fully Paid" loans

- The median annual income for borrowers with "Current" loans is slightly higher than both "Fully Paid" and "Charged Off" loans

- This suggests that income levels might play a role in defaults, with borrowers reporting lower incomes being slightly more susceptible to defaulting

Trend Over Time

- In 2011 we can see more defaults
- In December month we can see more defaults

Yearly Trend
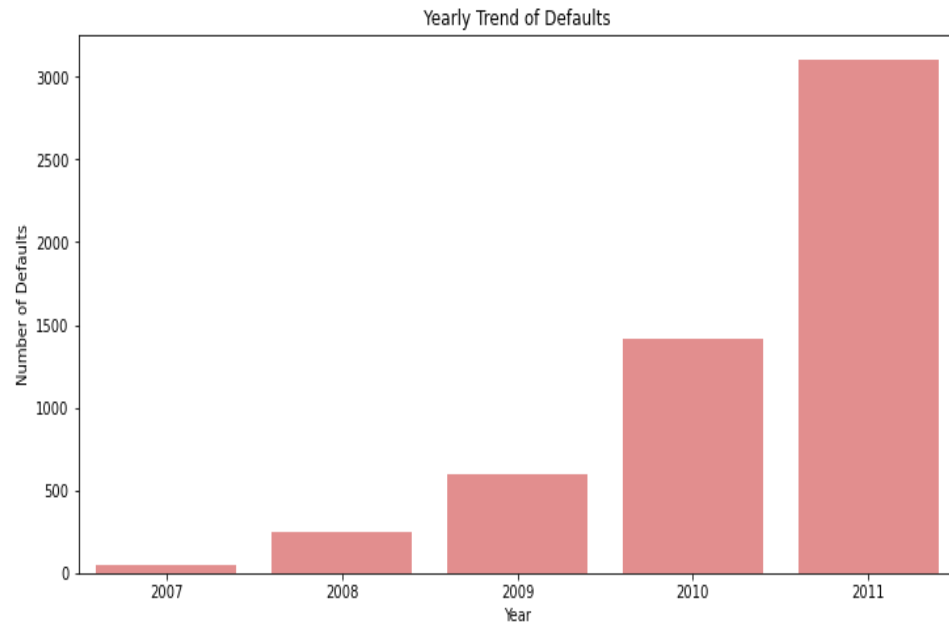
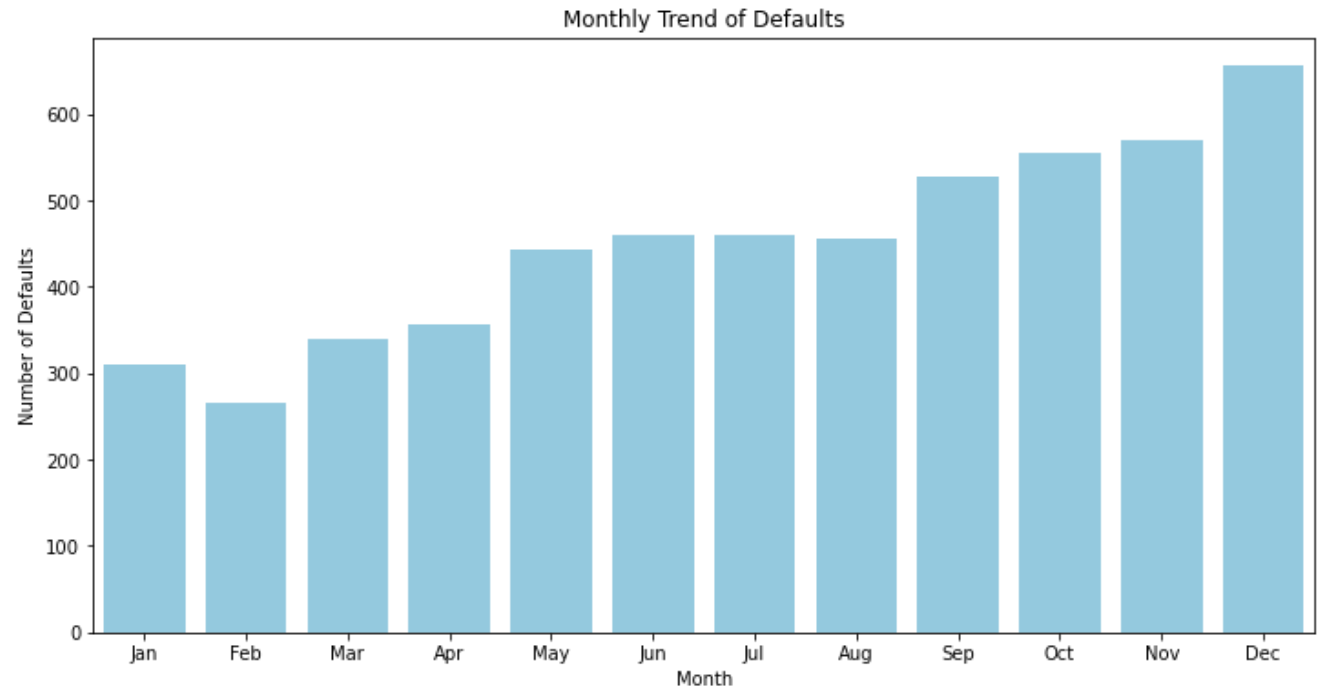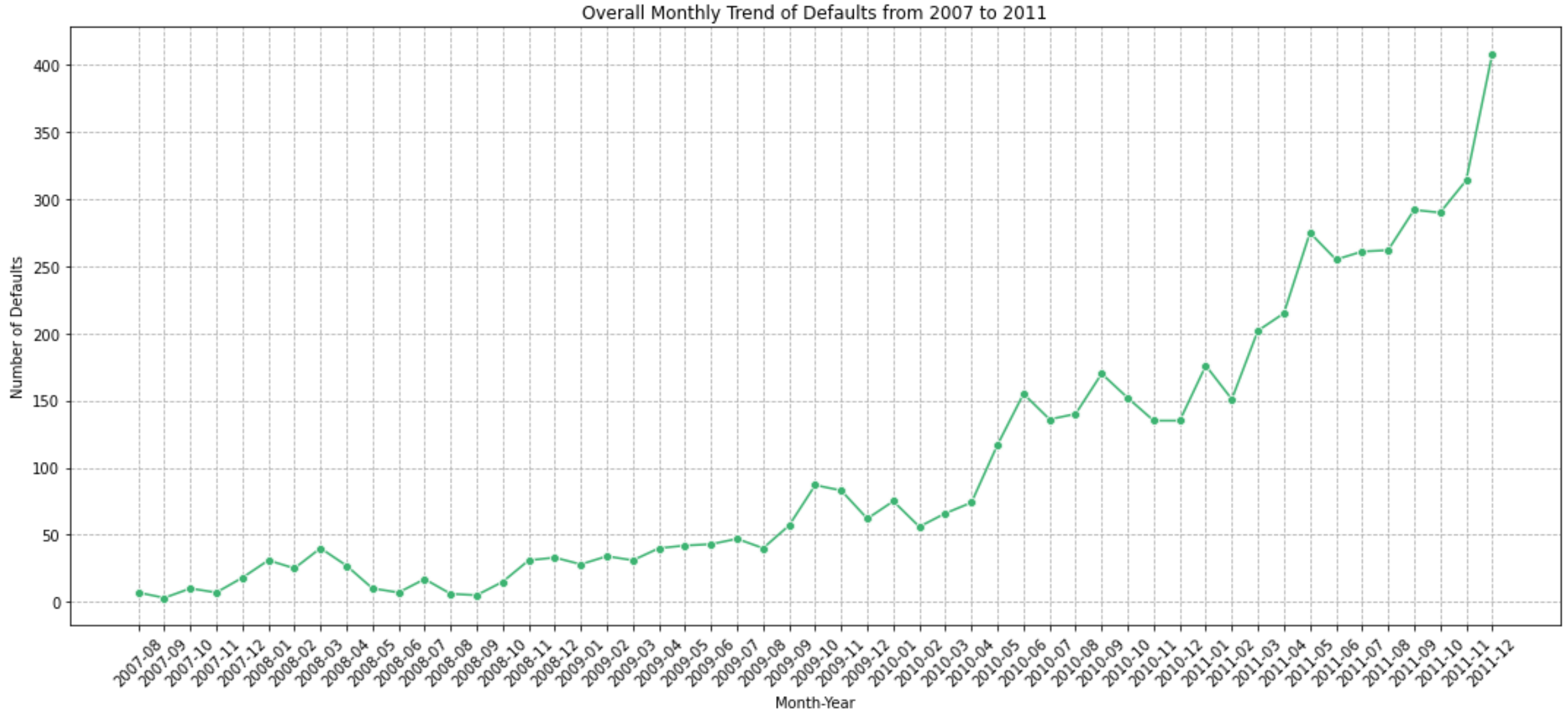## Monthly Trend

The default trend: 2011 December have too high defaults

Overall Monthly Trend of Defaults from 2007 to 2011

Other Analysis Summary

1. **Loan Amount vs Loan Status:**
   - Loans that defaulted ("Charged Off") have a slightly higher average loan amount than those that were "Fully Paid"

2. **Loan Term vs Loan Status**:
   - Longer-term loans (60 months) have a relatively higher proportion of defaults compared to shorter-term loans (36 months)

3. **Interest Rate vs Loan Status**:
   - Loans that defaulted have a significantly higher average interest rate compared to those that were fully paid

4. **Loan Grade vs Loan Status**:
   - Lower loan grades (like 'E', 'F', 'G') have a higher proportion of defaults compared to higher grades (like 'A' and 'B')

5. **Annual Income vs Loan Status**:
   - Borrowers who defaulted have a slightly lower median annual income than those who fully paid their loans

6. **Trend**:
   - 2011 and December have high default cases (too much high)

Summary of Loan Default Analysis

Loan Amount & Interest Rates: Smaller loans (0-7k) and interest rates between 5-9% show higher defaults.

Income & Purpose: Applicants with incomes around 30k-50k, especially those taking loans for debt consolidation, show a higher likelihood of default.

Loan Term: Shorter-term loans (36 months) are more likely to default.

Grade & Employment: Loans with grade 'B' and long-term employees (10 years) have higher defaults.

Home Ownership: Renters and those with mortgages show higher default rates.

Verification & Loan Amount: Larger loans tend to be verified more, and those above 16k show higher defaults.

Interest Rate & Annual Income: Higher salaried individuals receiving loans at higher interest rates (21-24%) are more likely to default.

Key Takeaway: To minimize credit loss, the lending company should be cautious with applicants showing the above attributes. Further verification, adjusted loan terms, or varied interest rates can be considered for these profiles.

THANK YOU