

Lending Loan Case Study

Sanath P Kanchan

The Problem

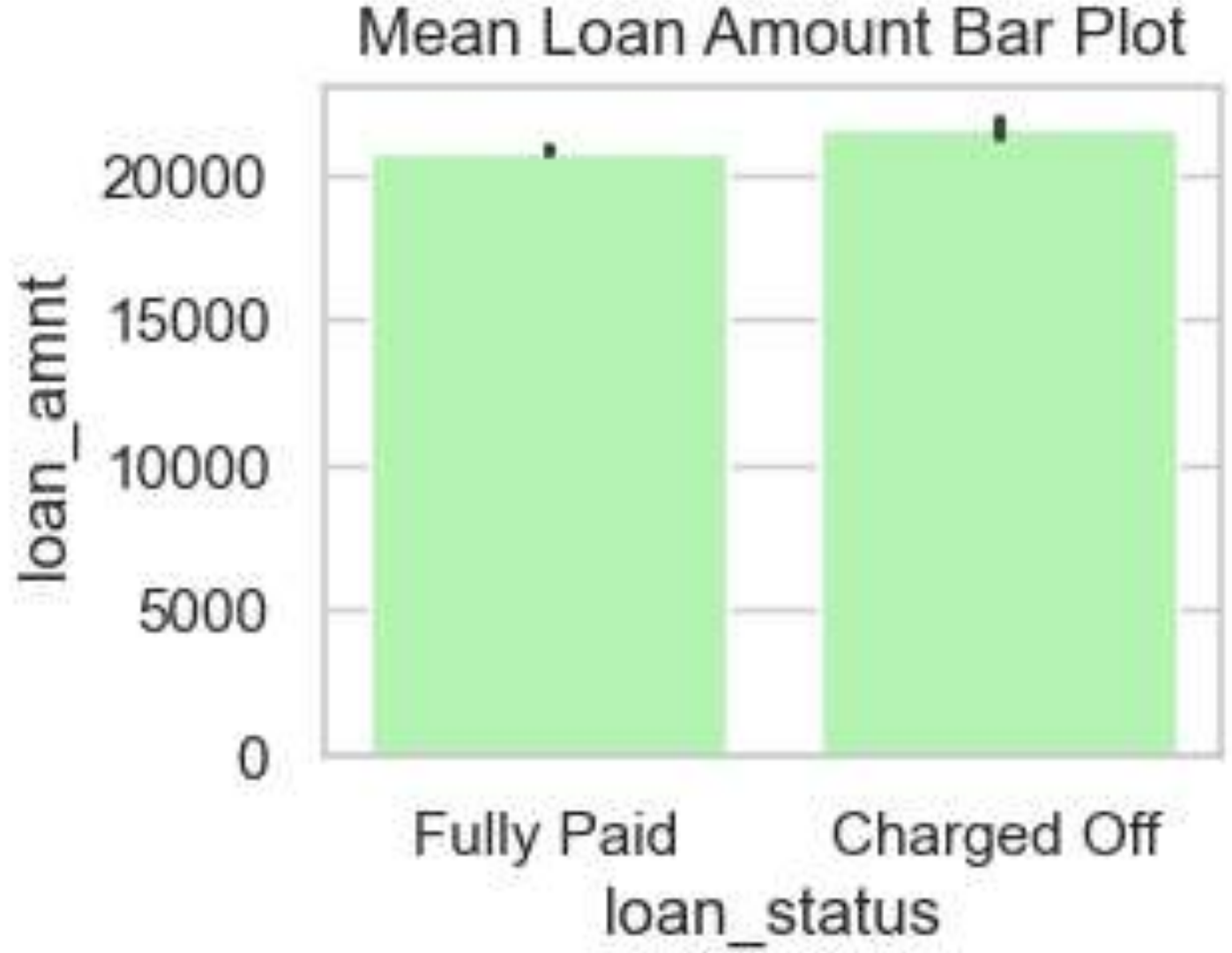
- **Company** - Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.
- **Context** - Lending Club wants to understand the **driving factors** behind loan default, i.e. the **driver variables** which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.
- **Problem Statement** - As a data scientist working for Lending Club analyze the dataset containing information about past loan applicants using EDA to understand how *consumer attributes* and *loan attributes* influence the tendency of default

Analysis Approach

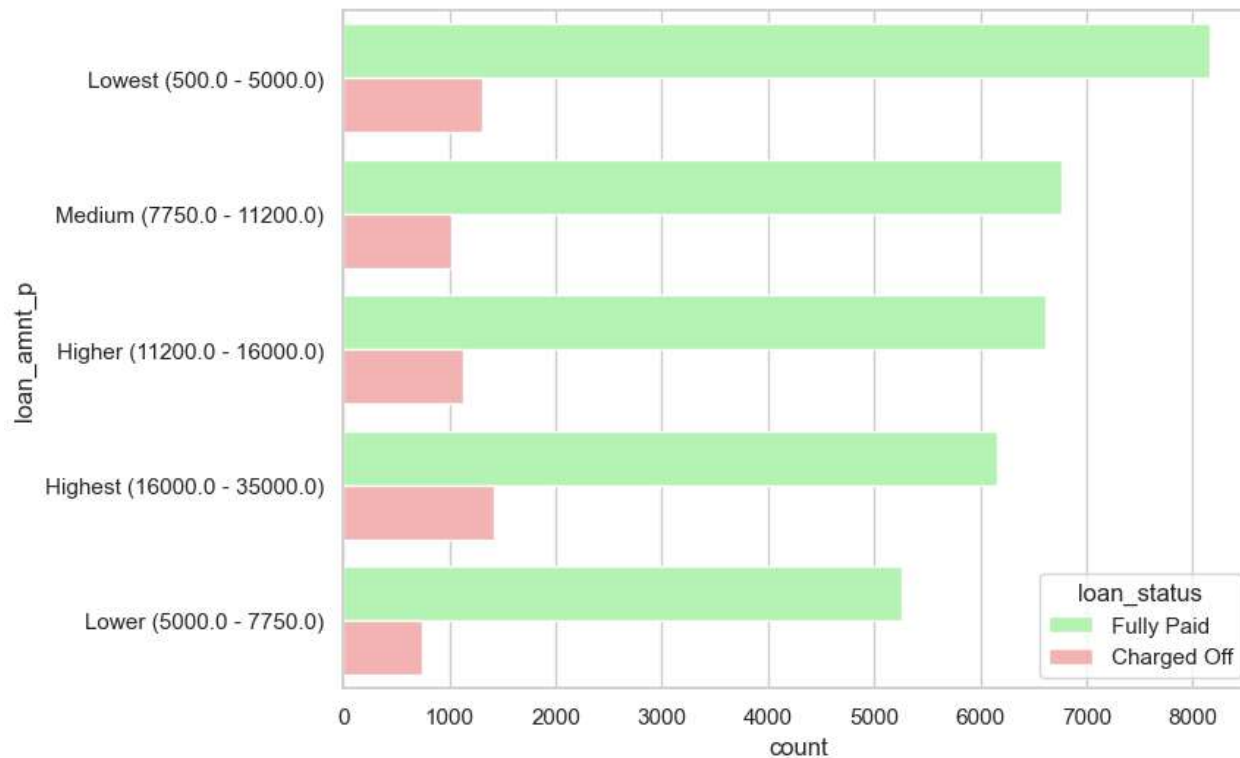
- Clean Data - > Drop columns with null values, all random values or single category value > Convert values to proper int, float, date representations
- Univariate Analysis - > Check distributions and frequencies of various numerical and categorical variables > Create derived variables.
- Segmented Univariate Analysis - > Analyze variables against segments of other variables > Create derived variables
- Bivariate Analysis - > Do correlation analysis Check how two variables affect each other or a third variable > Analyze joint distributions
- Summarize Results - > Publish insights and observations.

Analysis – Loan Status

- From this bar plot we can conclude that "higher the amount" will tend to "Write off".

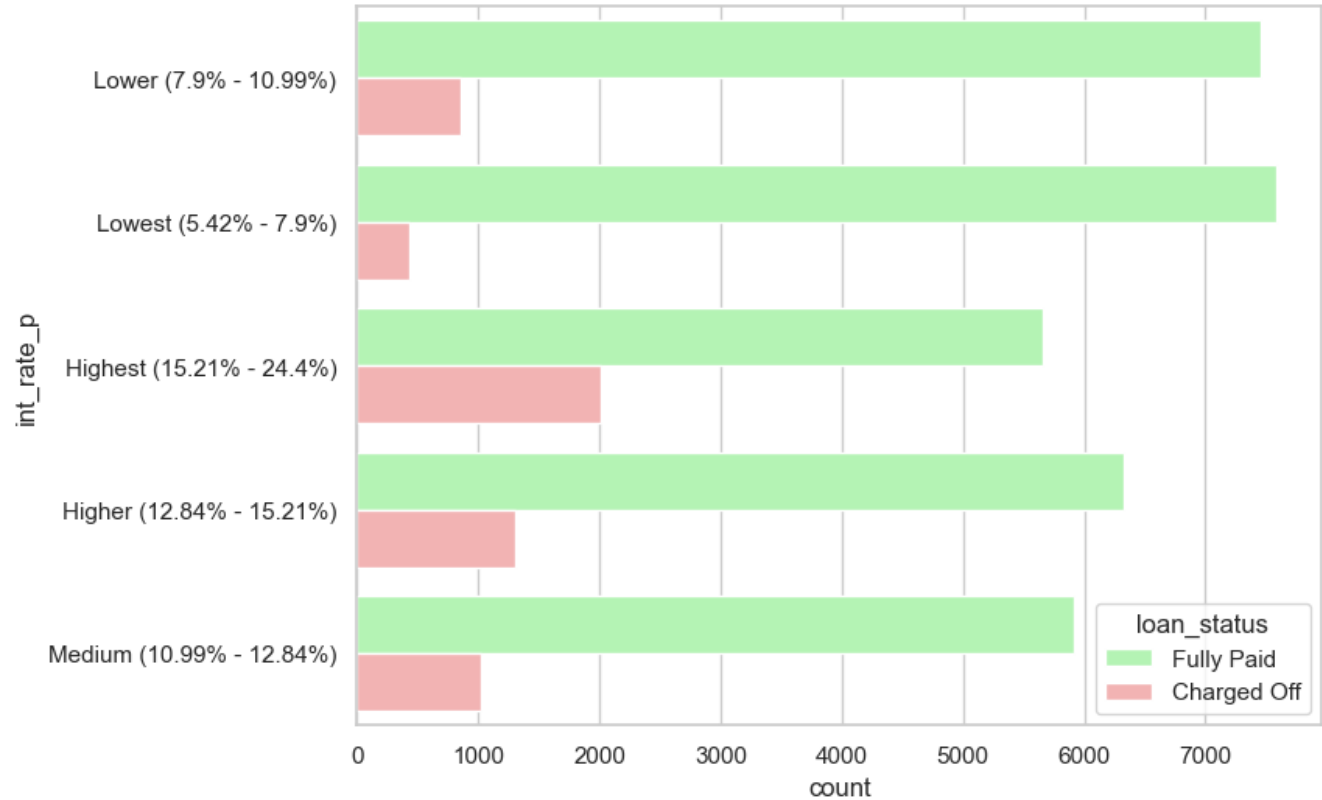


Analysis - Understanding Loans



- Higher the loan amount, greater the chance of the loan getting default.

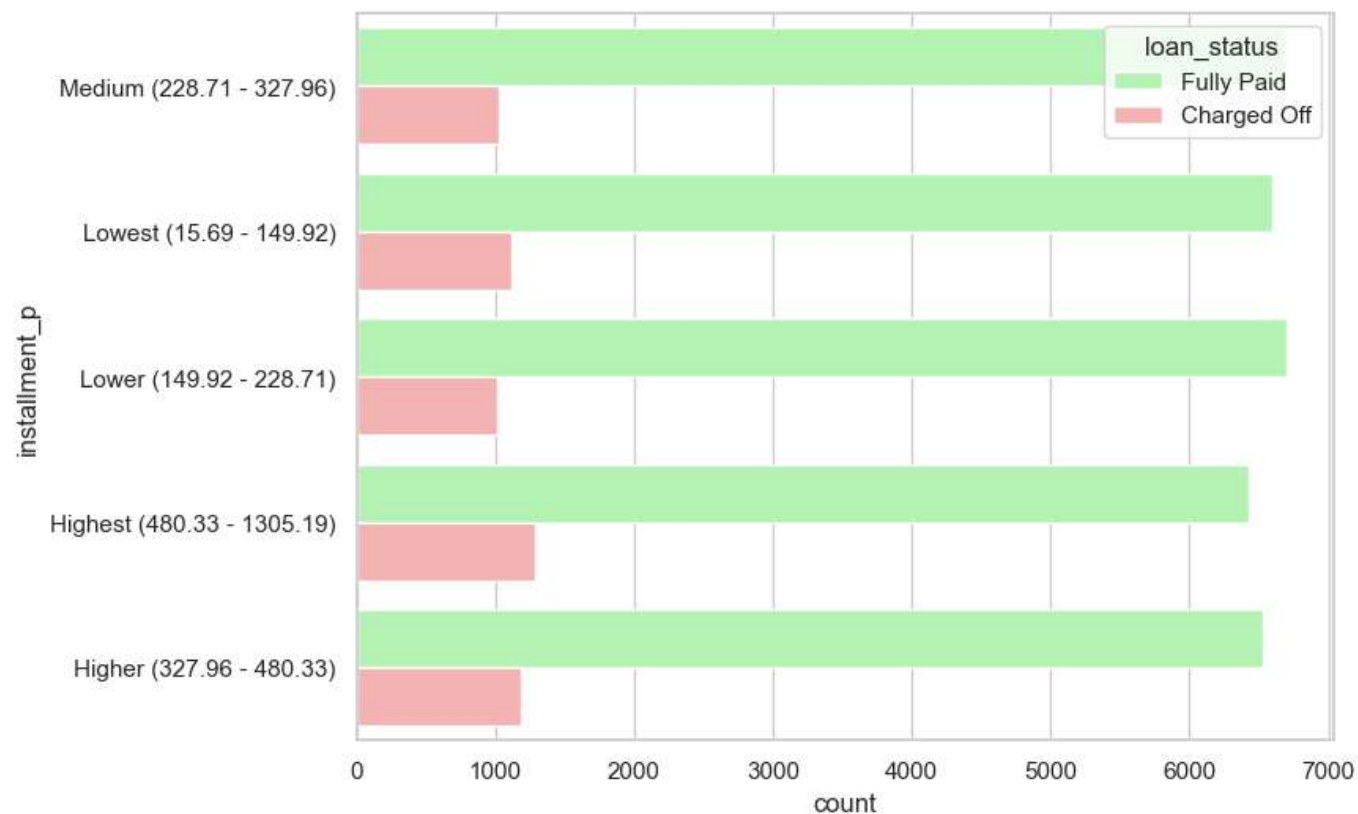
Analysis - Defaults by Interest Rate



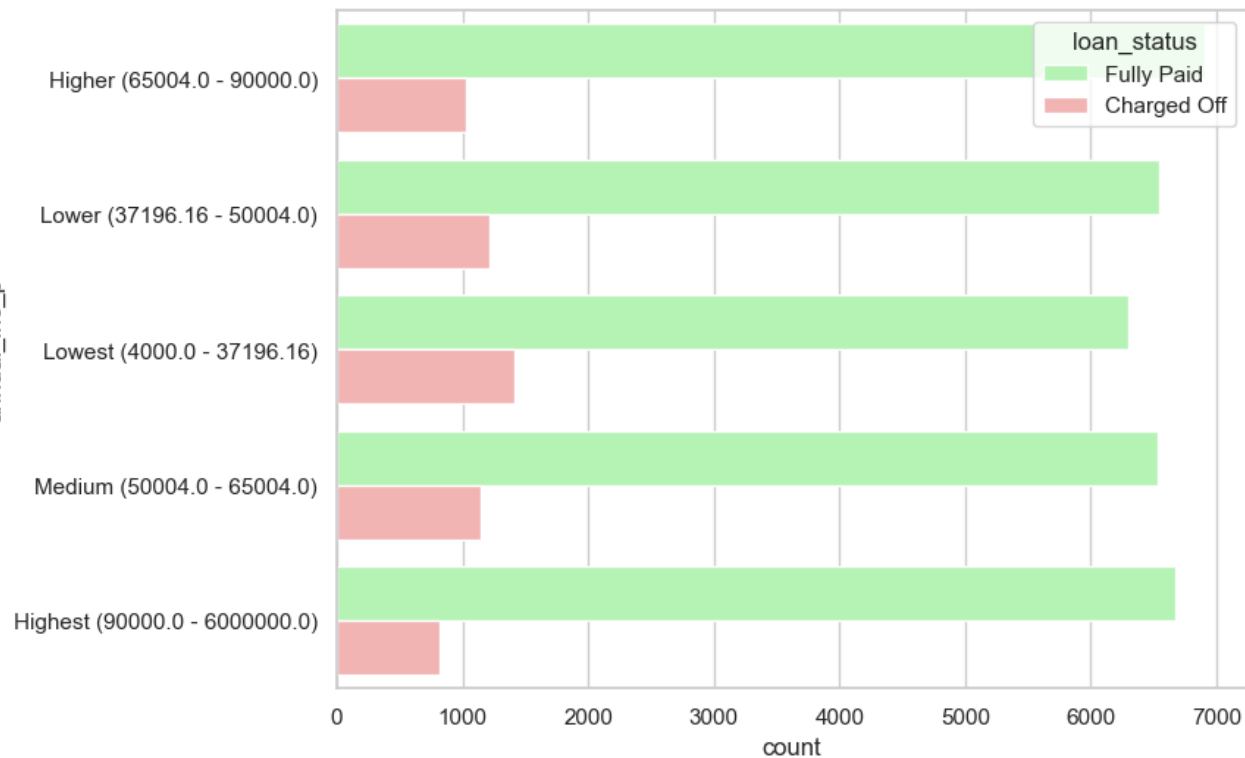
- Higher the interest rate leads to higher charged off%

Analysis - Defaults by Installment Amt

- Above data and plot shows that higher installment amounts shows higher default percentages.



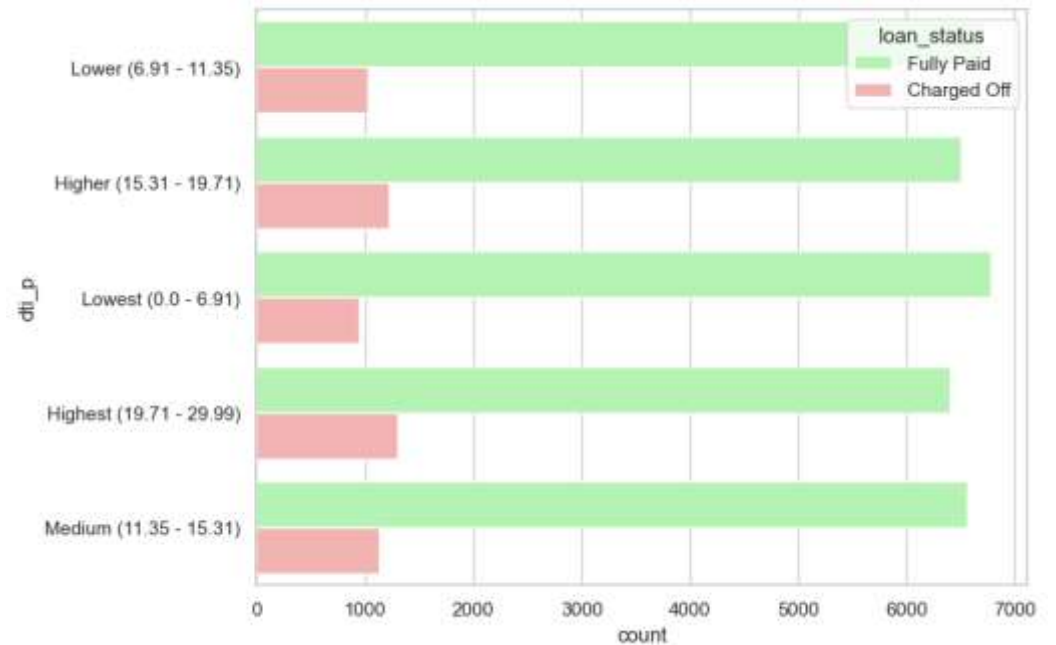
Analysis – Repayment consider Income

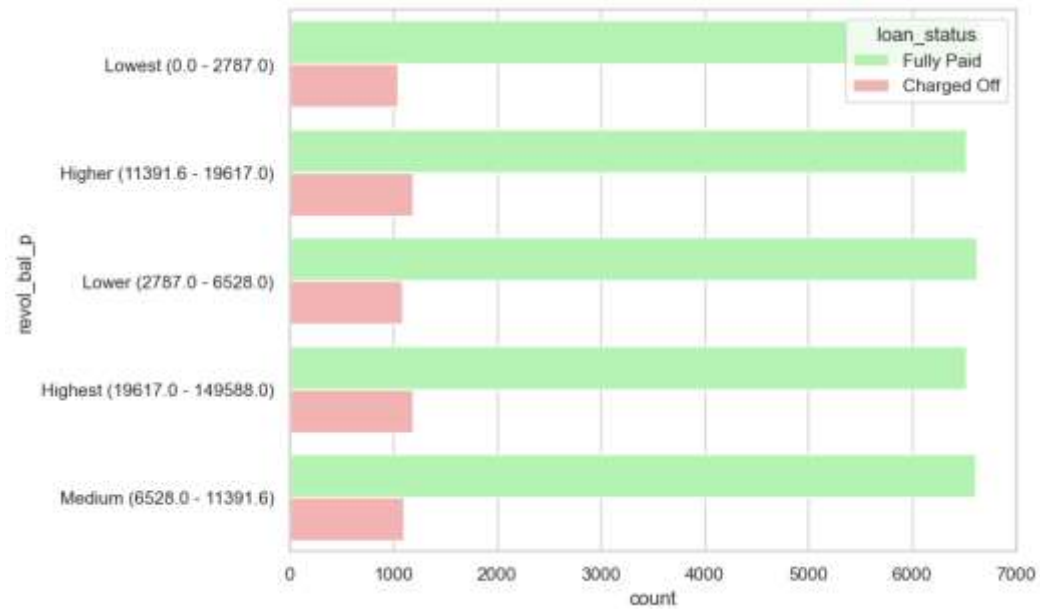


- Higher the income higher the repayment %

Analysis – DTI (Debt to income ratio)

- Higher DTI (debt to income ratio) will lead to higher charged off %

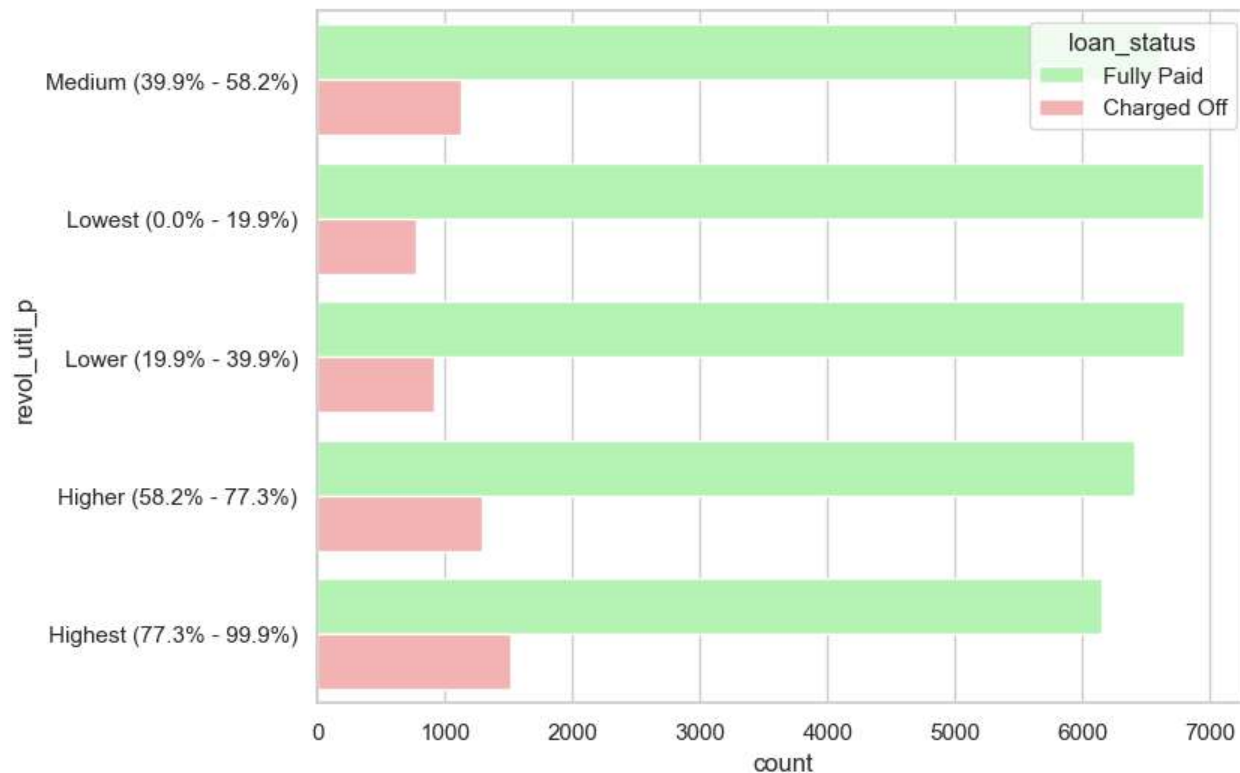




- This shows the total credit revolving balances slightly influence the default percentage. Higher the revolving balance, bigger the chance of the loan getting defaulted.

Analysis - Revolving balances

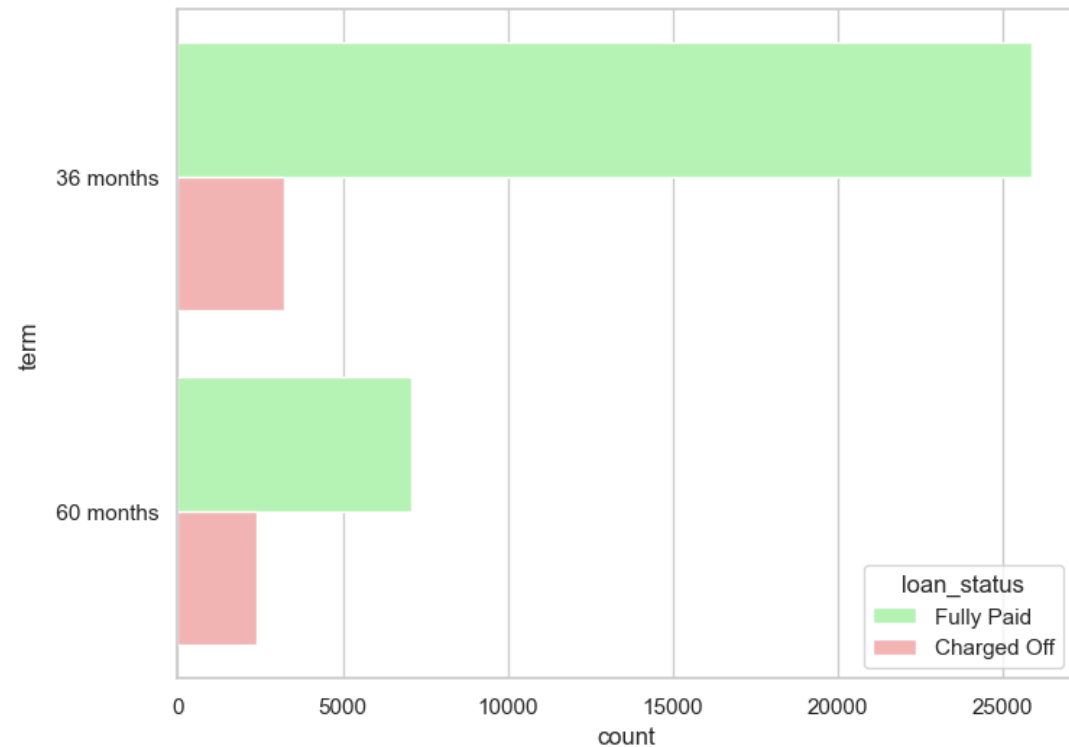
Analysis - Revolving line



- This data shows that the revolving line utilization rate has a large impact to the default percentage. When this increases, the charged off percentage rises.

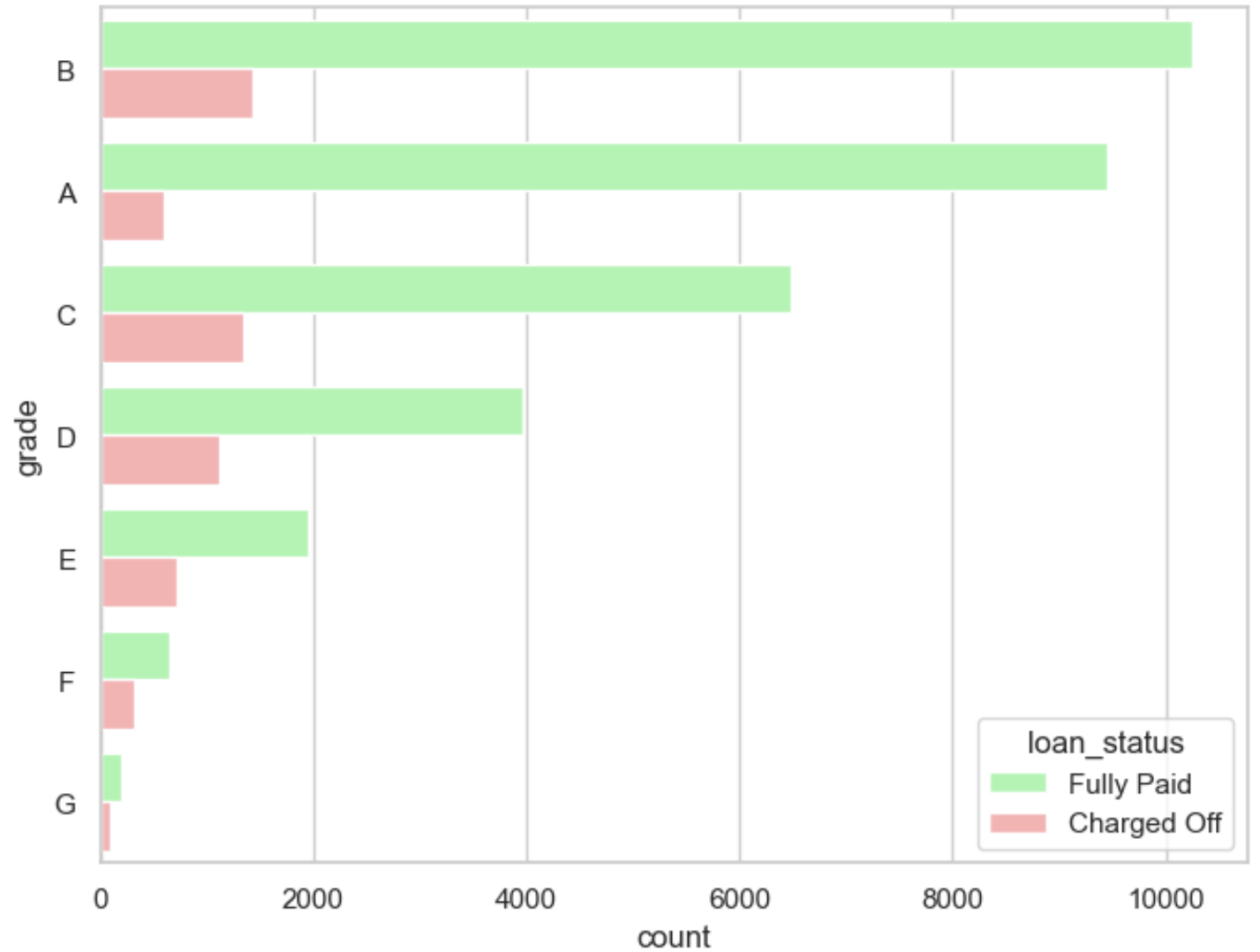
Analysis - Term

- For loans with 5-year repayment term, the default percent is 25%. And for 3-year loan repayment term, the default is only for 11% of the cases. Therefore, loan repayment term plays a factor in judging the default rate.



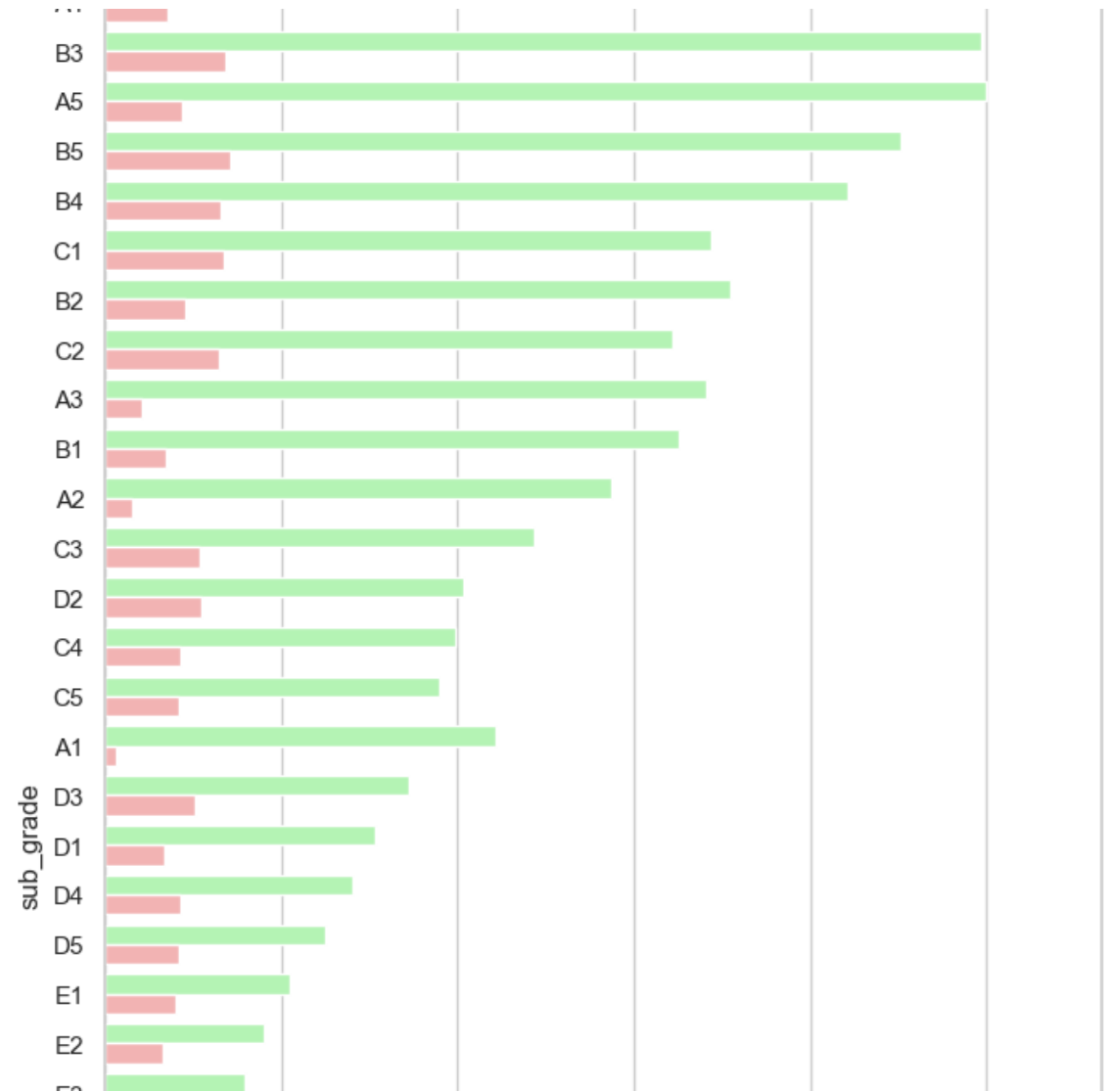
Analysis – Loan Grades

- We can clearly see that loan grades having highest default percentages. G, F, E and D form grades where default rate is much higher than others.

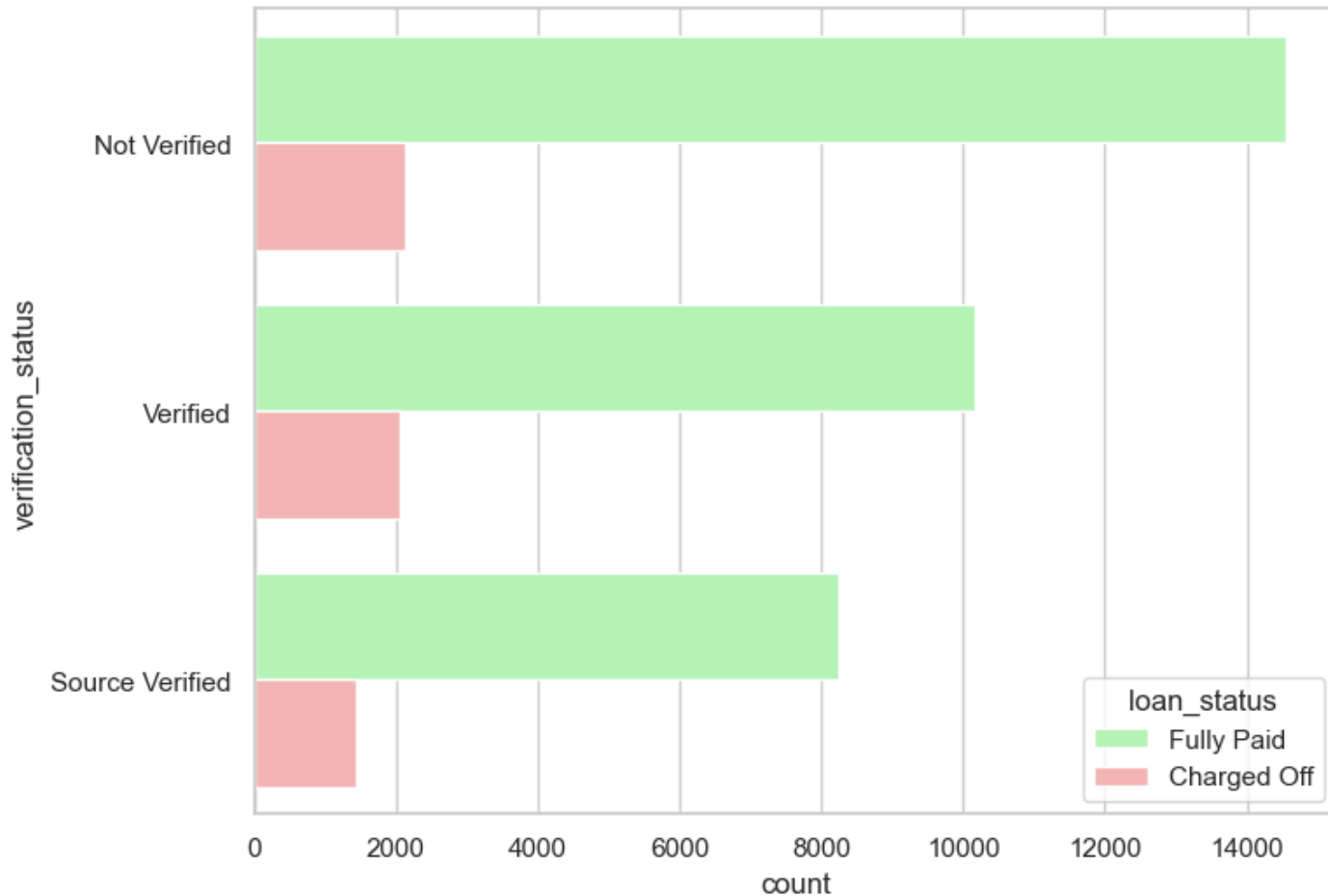


Analysis – Loan Sub-grade

- The above table shows the loan sub-grade versus the default percentage. The G3 and F5 sub-categories have above 40% default rate. This field is a clear indicator of the default percent.

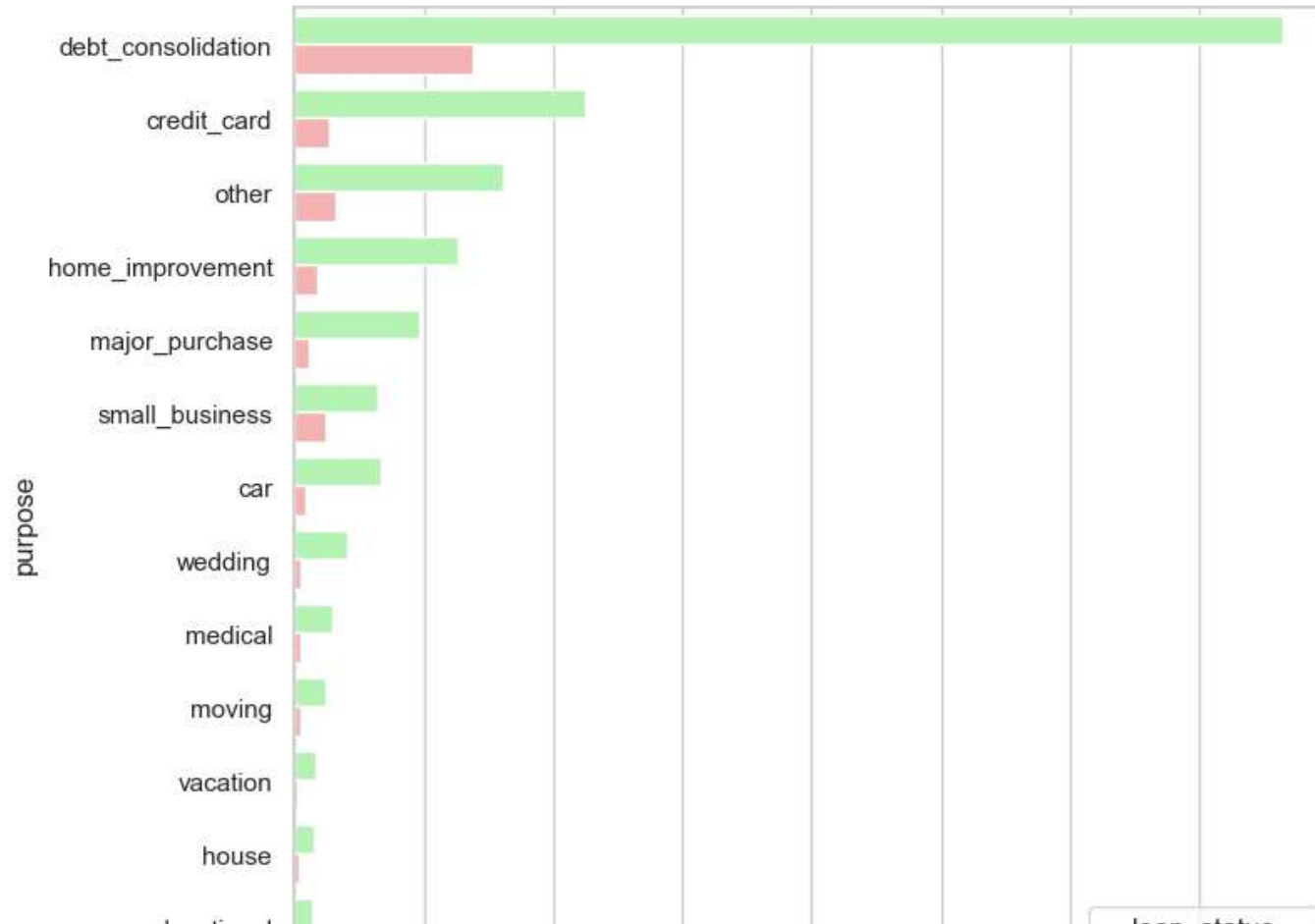


Analysis – Verified Applicants

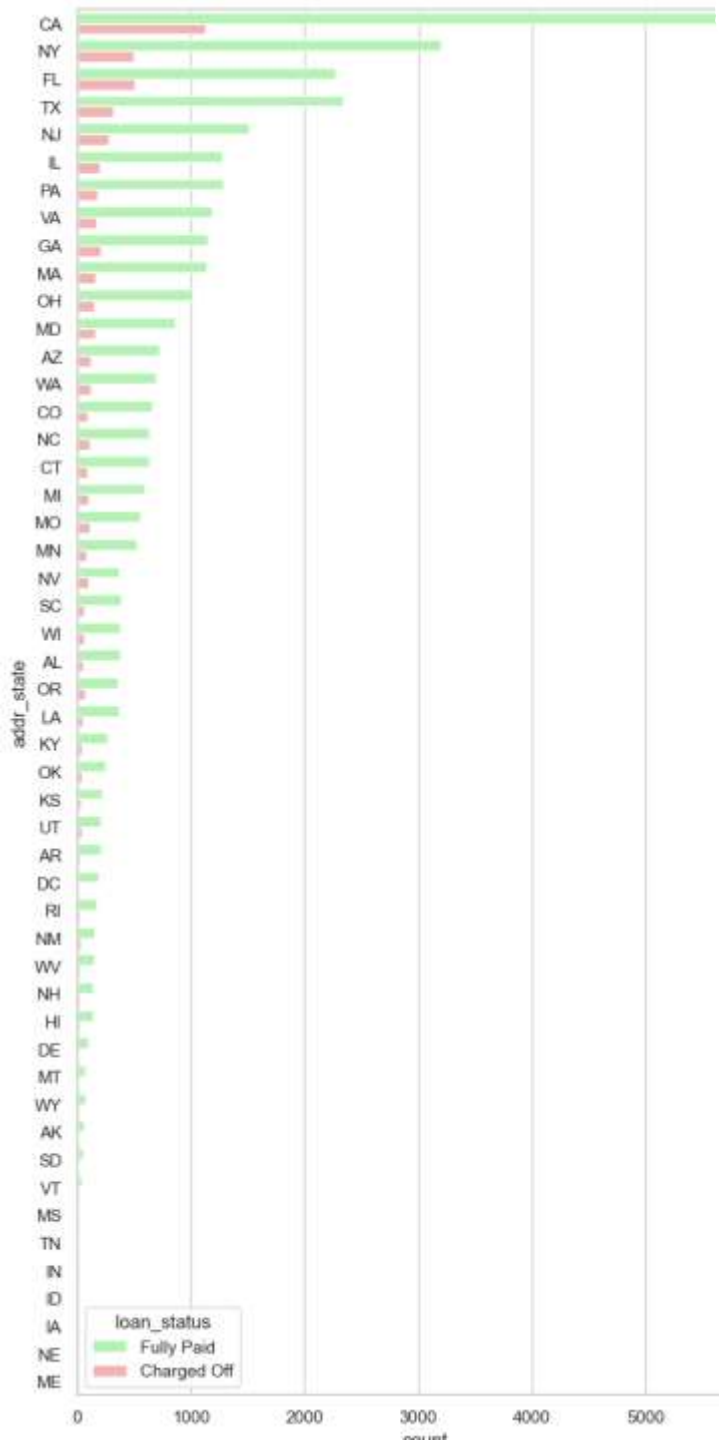


- interestingly verified applicants have more defaulted % which means the verification process is not correct.

Analysis – Purpose of Loans



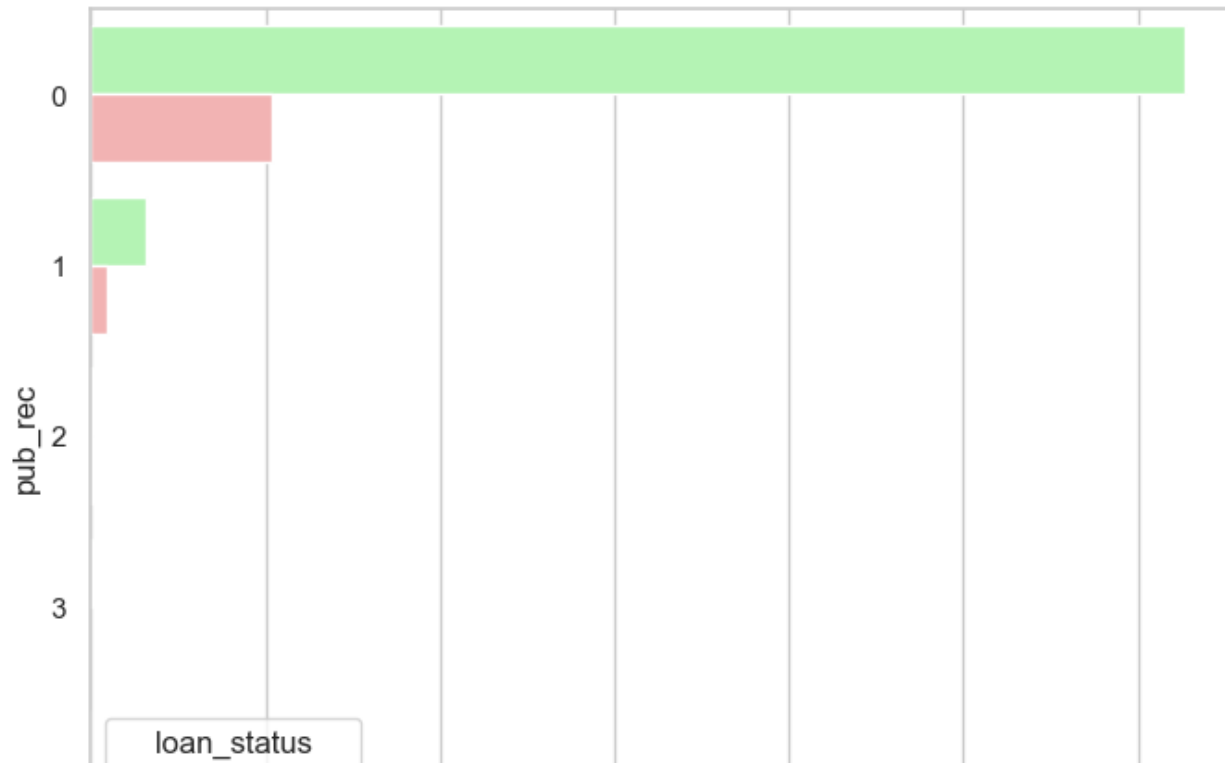
- From the above analysis it is evident that the loans taken for small business, renewable energy and educational are the riskier ones.



Analysis - States

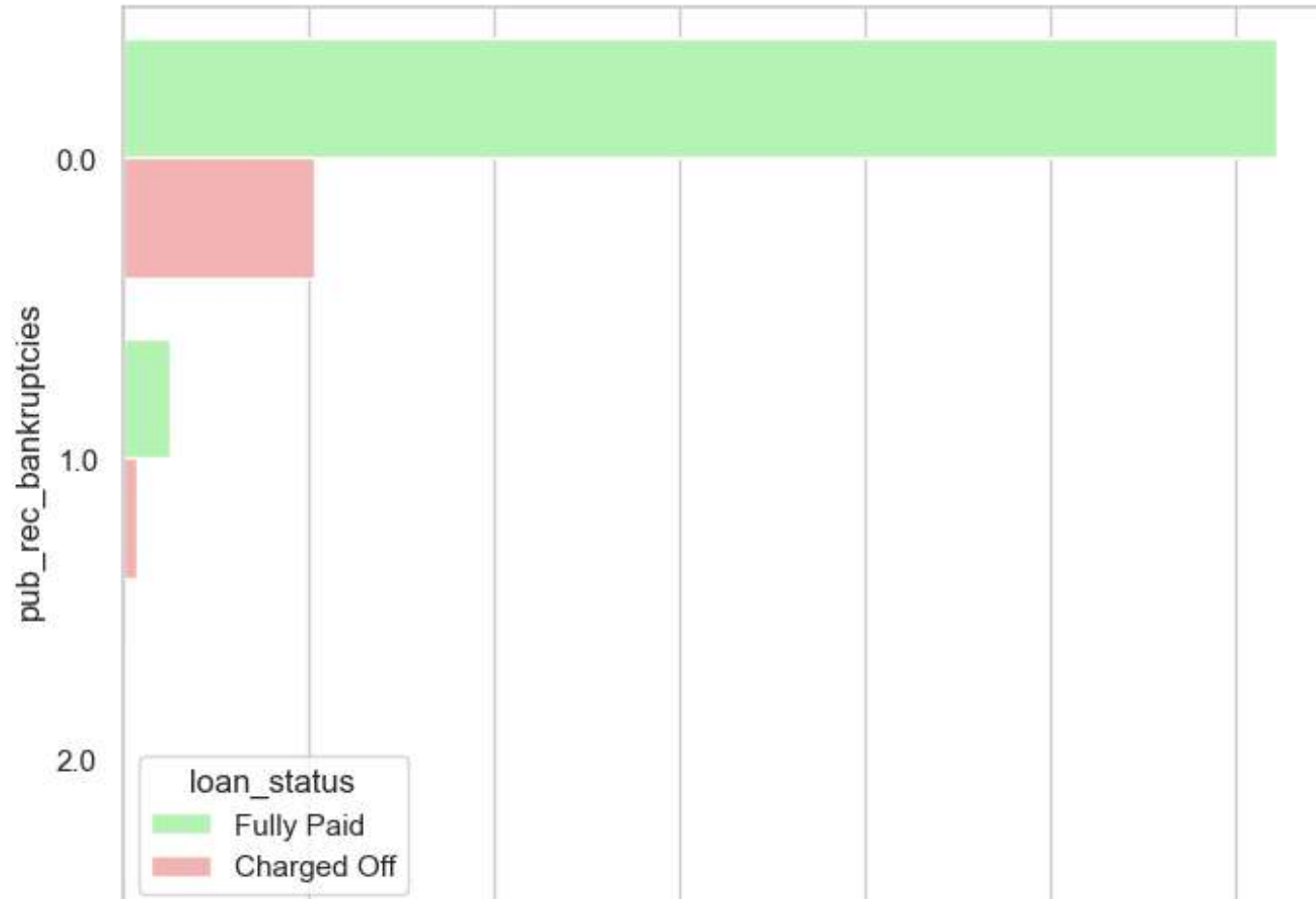
- From the above table we can see that the some of the risky states are NE, NV, SD, AK, FL, MO

Analysis – Public Records



- Above data and graph clearly shows that individuals with non-zero derogatory public records have higher chances of charged off

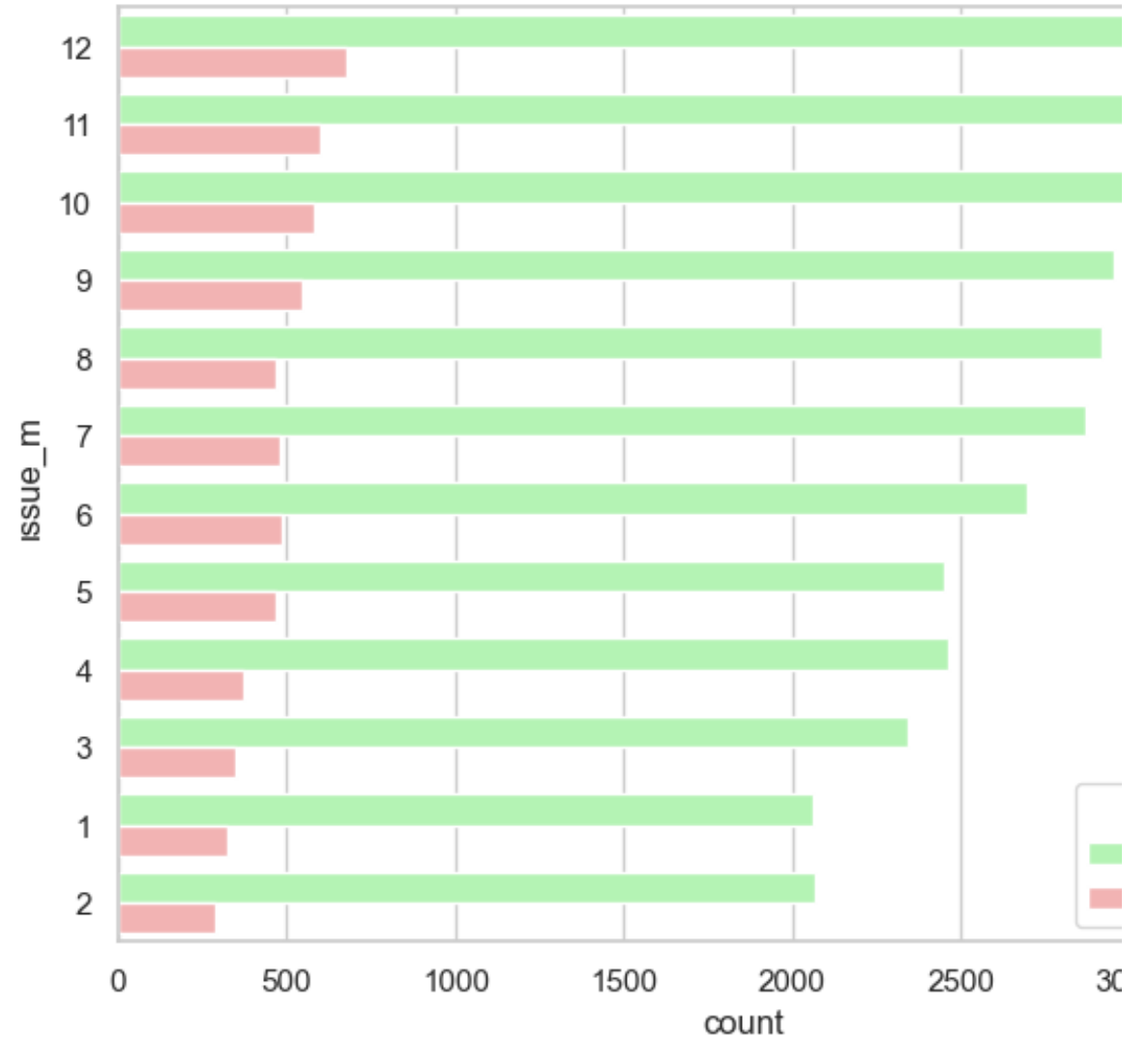
Analysis - Public bankruptcy records



- Higher the number of public bankruptcy records, bigger the chance of defaulting the loan.

Analysis – Month

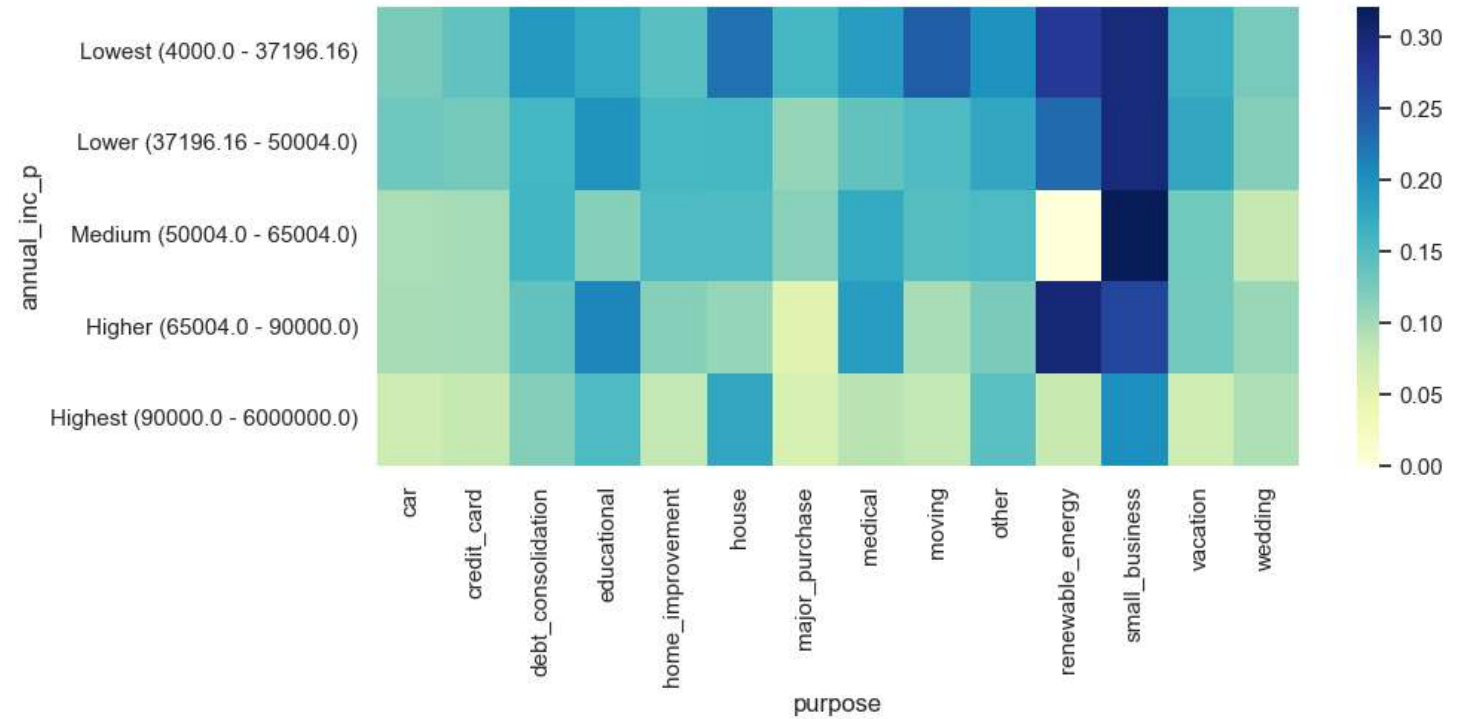
- we can clearly see December is the month which has the highest number of loan applications per year and also have the biggest default ratio. May be people take loan for travel/party during Christmas and then are not able to pay back. Month of May is also another one, which is during the summer break and right before the Memorial day and Independence day breaks in US where people love to travel.



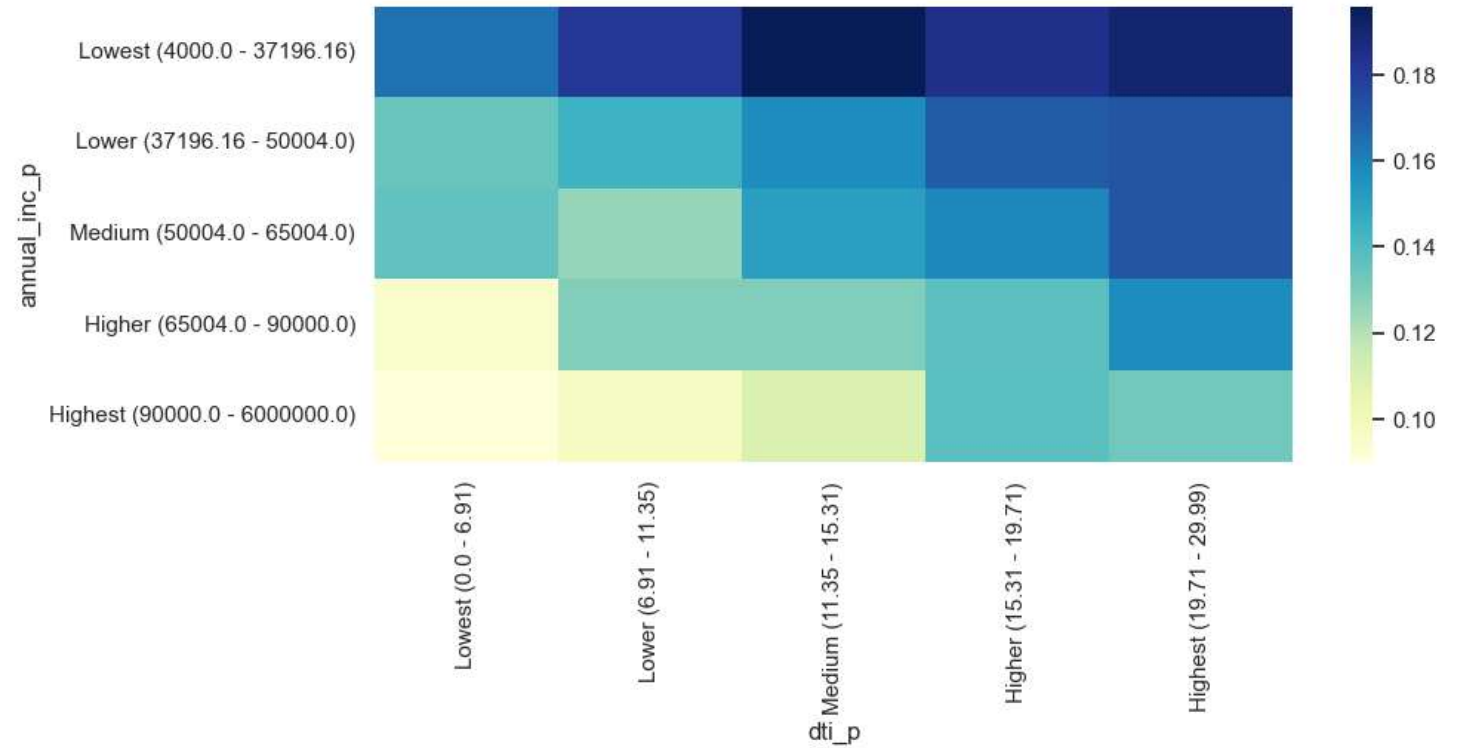
Analysis - Income groups vs Risky purposes of loans

Plot of various income groups versus the risky purposes of loans for them. Some examples are:

- Small business loans for lowest and medium income groups
- Renewable energy loans for higher income group



Analysis - Debt-to- income group vs Income range vs Repayment



- Medium debt-to-income group in the lowest income range is the most risky when it comes to loan repayment.

Summarize Results

Hereby we come to an end of the EDA of the loan data set and finding some of the drivers for loan default. Apart from the ones highlighted below, I am sure there will be multiple others too; however, according to me, these are the most impactful ones.

Minor Impact

- Higher loan amount (above 16K)
- Higher installment amount (above 327)
- Lower annual income (below 37K)
- Higher debt to income ratio (above 15%)
- Applicant's address state (NV, SD, AK, FL, etc.)
- Loan issue month (Dec, May, Sep)

Heavy impact

- Higher interest rate (above 13%)
- Higher revolving line utilization rate (above 58%)
- Repayment term (5 years)
- Loan grade & sub-grade (D to G)
- Missing employment record
- Loan purpose (small business, renewable energy, educational)
- Derogatory public records (1 or 2)
- Public bankruptcy records (1 or 2)

Combined impact

- High loan amount & interest rate for lower income group
- High installment and longer repayment term
- Home ownership (other) and loan purpose (car, moving or small business)
- Residential state and loan purpose
- Income group and loan purpose

THANK YOU