LENDING CLUB CASE STUDY

Sudeep K S
Chetan D. Wargantiwar

AGENDA

- Case Study Objectives
- Methodology Used
- Data Analysis
- Data Cleaning
- Univariate Analysis
- Bivariate Analysis
- Recommendations

CASE STUDY OBJECTIVES

- We are a Consumer Finance Company which specializes in lending various types of loans to urban customers
- We must decide for loan approval based on the applicant's profile
  - If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
  - If the applicant is not likely to repay the loan, i.e., he/she is likely to default, then approving the loan may lead to a financial loss for the company
- We are provided with information about past loan applicants and whether they 'defaulted' or not
- The aim is to identify patterns to understand the driving factors (or driver variables) behind loan default to understand the driving factors (or driver variables) behind loan default

METHODOLOGY USED

DATA ANALYSIS

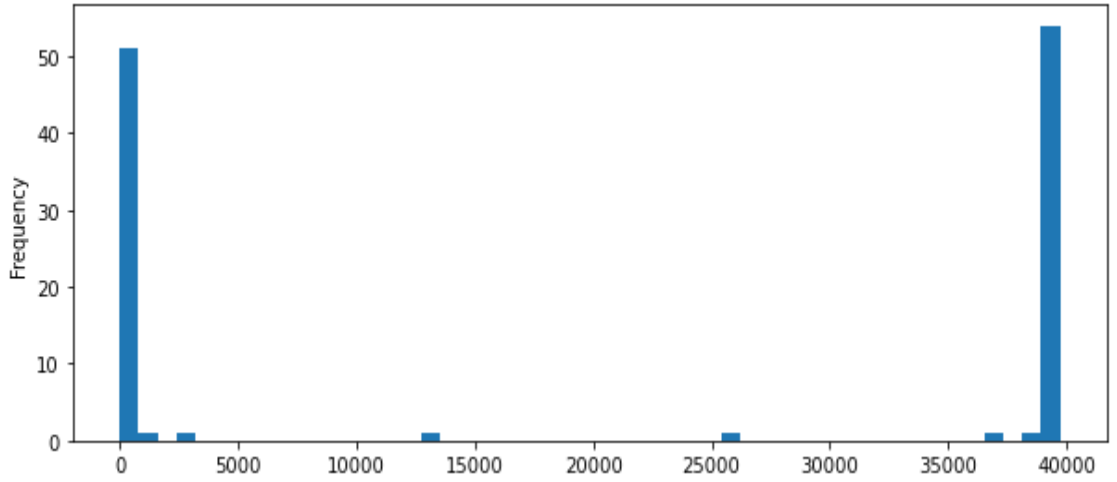
- We first need to analyze the data frame provided to us, that is the Loan.csv file
- For this purpose, we read the file and display the ***head*** , this gives me a glimpse into all the data I have in the Data frame
- I also check mode details using ***info*** function which gives us the total number of rows in the data frame
- Code from the python book :

```
loan= pd.read_csv(r'loan.csv')
loan.head()
loan.info()
```

DATA CLEANING

- Now that the data is analyzed, and we know the number of columns and entry details
- We must now understand all various columns from the dictionary provided
- Once we have been familiarized with the details of all columns, we must identify and remove unwanted data
- We must identify outliers or single value data and remove them from the data frame
  - Delete columns :
  - Remove Outliers/Unique values
  - Remove missing values
  - Remove duplicates if present
  - Filter and optimize entries/rows if required

DATA CLEANING



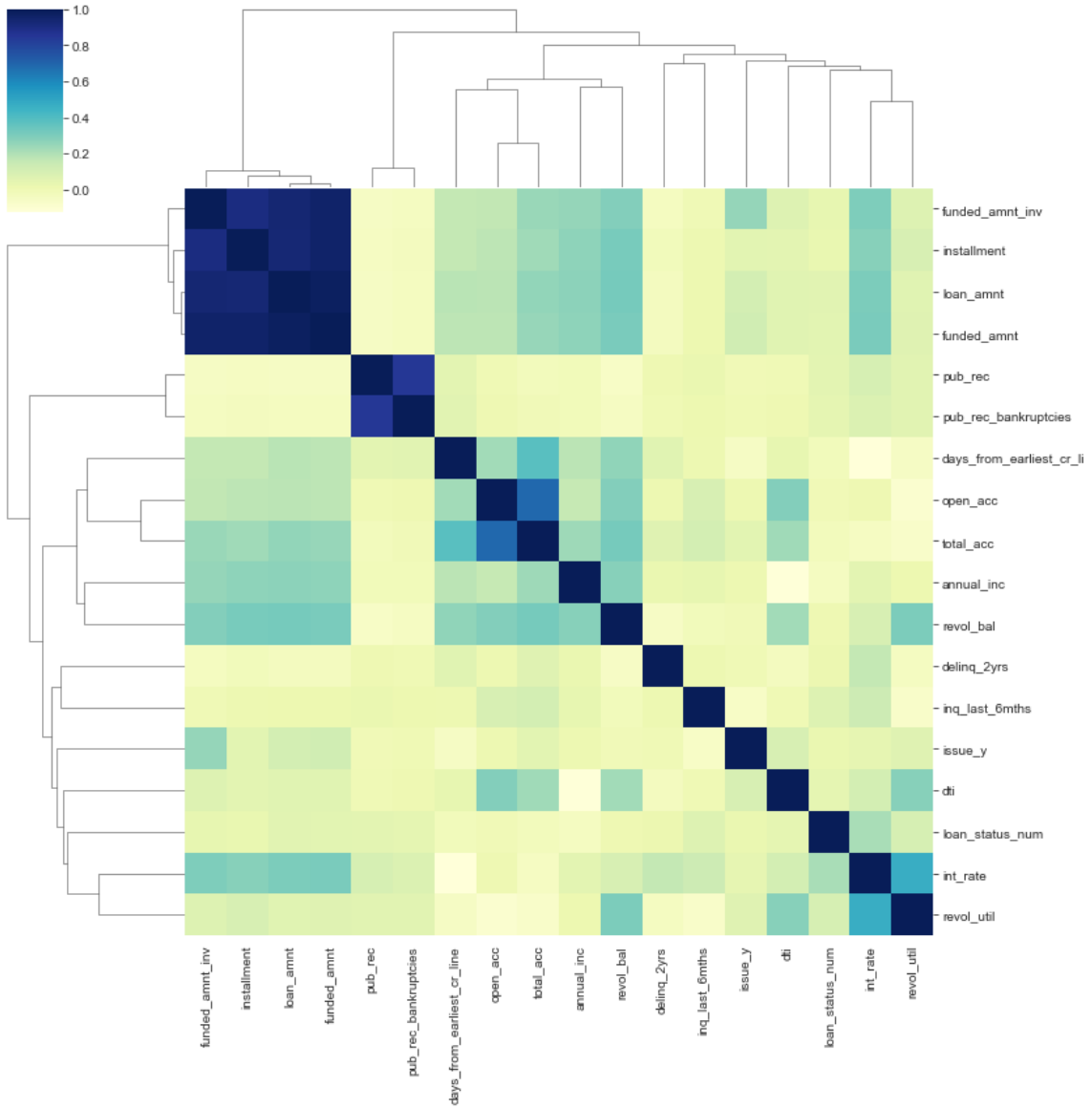
- In the above figure, X-axis contains number of NULL values and Y-axis contains number of columns.
- We can clearly see that there is a huge spike towards the 40K mark.
- The total records in our data is 39,717. We would obviously not need the columns that have all NULLs.
- Using the above information, we dropped columns with Null Data
- We also checked for unique values and dropped outliers
- From 111 columns , we came down to 44 with same number of rows
- The company wants to know which loan applications are risky.
  - The fields that are created after a loan application is approved not applicable to us
  - The fields like id, member_id & url are different for each application.
  - All the above can be dropped

DATA PREPARATION

- For Data preparation we now see the various parameters available and see if it can be useful to us or not for analysis
- We analyze the loan_status, int_rate, term, grade, sub_grade,issue_d and other parameters
- We prepare these data as per our needs an example we used:
  - *add new column like issued_year and issued_month which will be useful for our further analysis and drop this issue_d column*

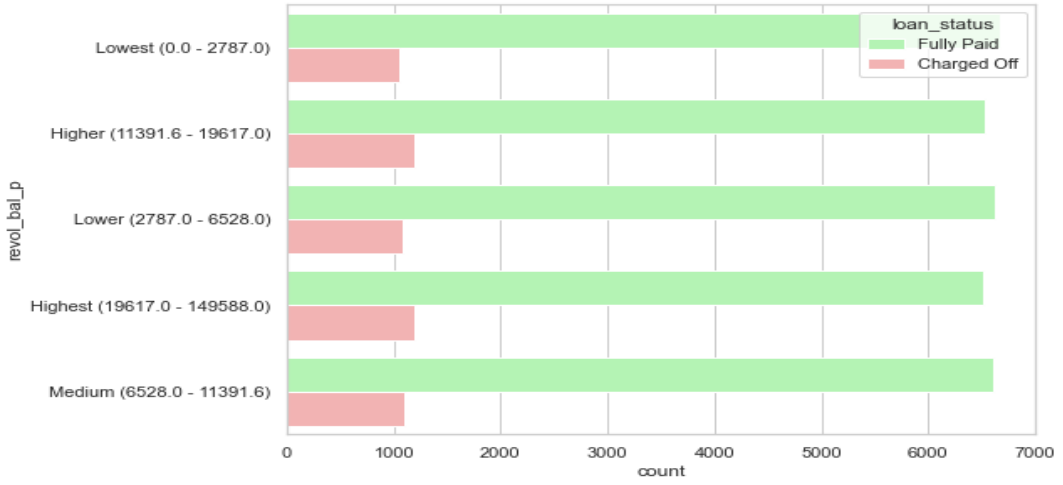
UNIVARIATE ANALYSIS

- As data is cleaned and prepared, we will now try to understand the correlation between the different numeric fields
- Since we know darker the value higher the correlation
  - loan_amnt, funded_amnt, funded_amnt_inv
  - installment have huge correlation
  - These fields are proportional to each other.
- Next, the public records related fields
  - pub_rec & pub_rec_bankruptcies
  - number of accounts related fields open_acc
  - total_acc are correlated.



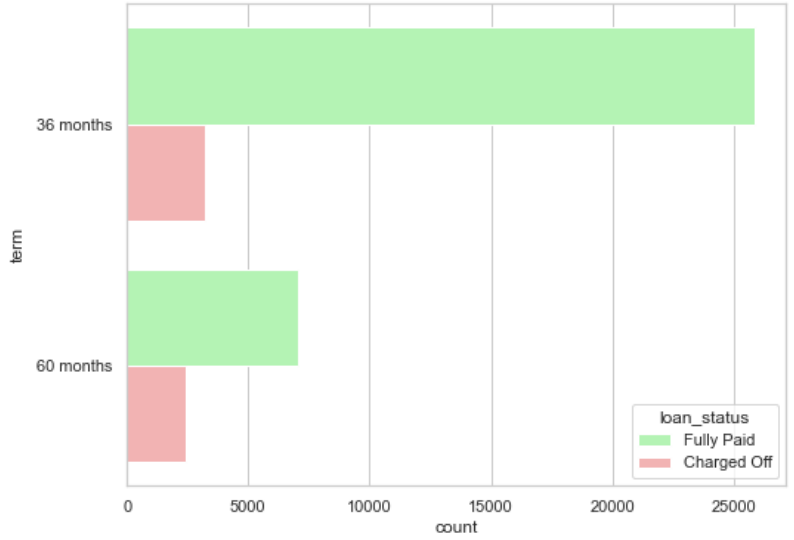
LOAN STATUS
(REVOLVING BALANCES)

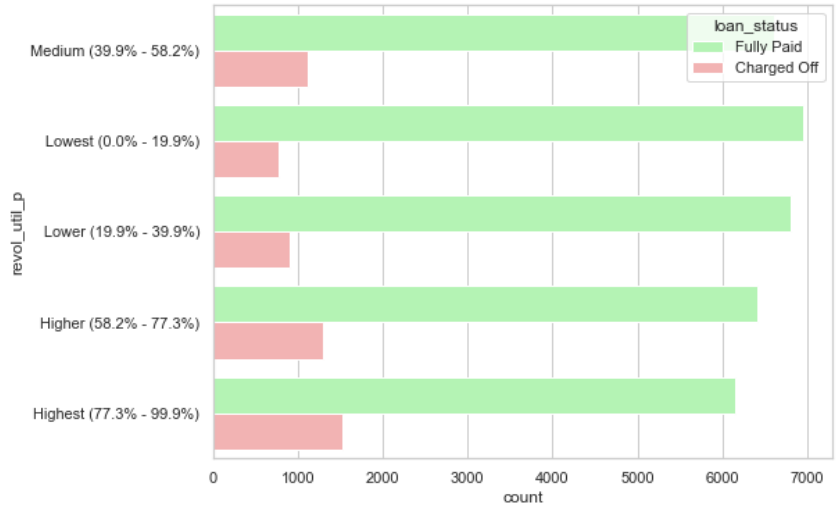
- from box plot we can conclude that "higher the amount" will tend to "Write off“
- The second plot shows the total credit revolving balances slightly influence the default percentage.
- Higher the revolving balance, bigger the chance of the loan getting defaulted.



LOAN STATUS
(TERM, REVOLVING UTILIZATION)

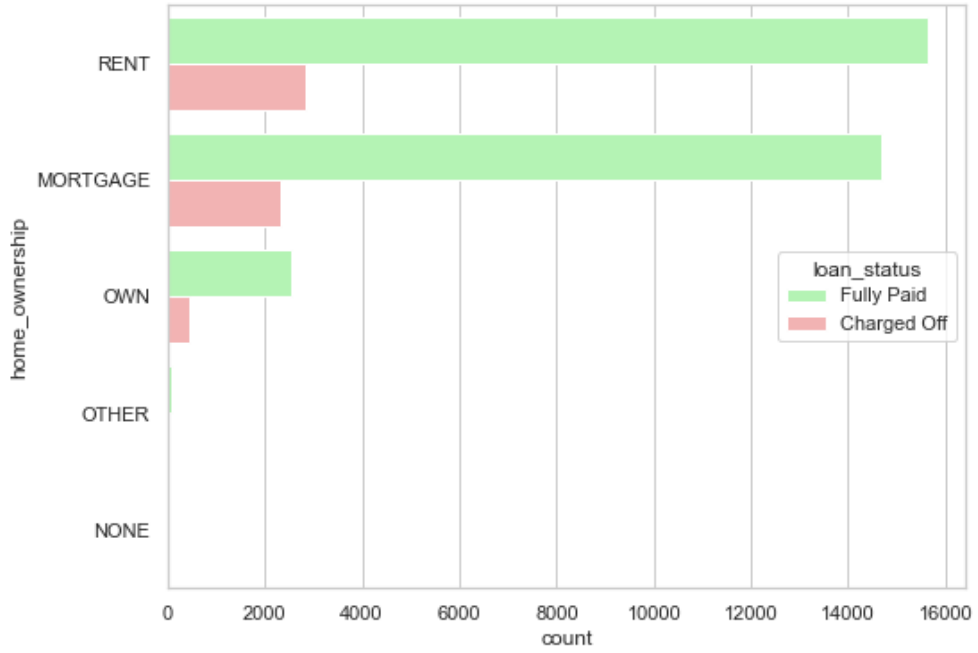
- The First Plot shows that loans with 5-year repayment term, the default percent is 25%
- for 3-year loan repayment term, the default is only for 11% of the cases
- Therefore, loan repayment term plays a factor in judging the default rate.
- The Second plot here shows that the revolving line utilization rate has a large impact to the default percentage. When this increases, the charged off percentage rises.

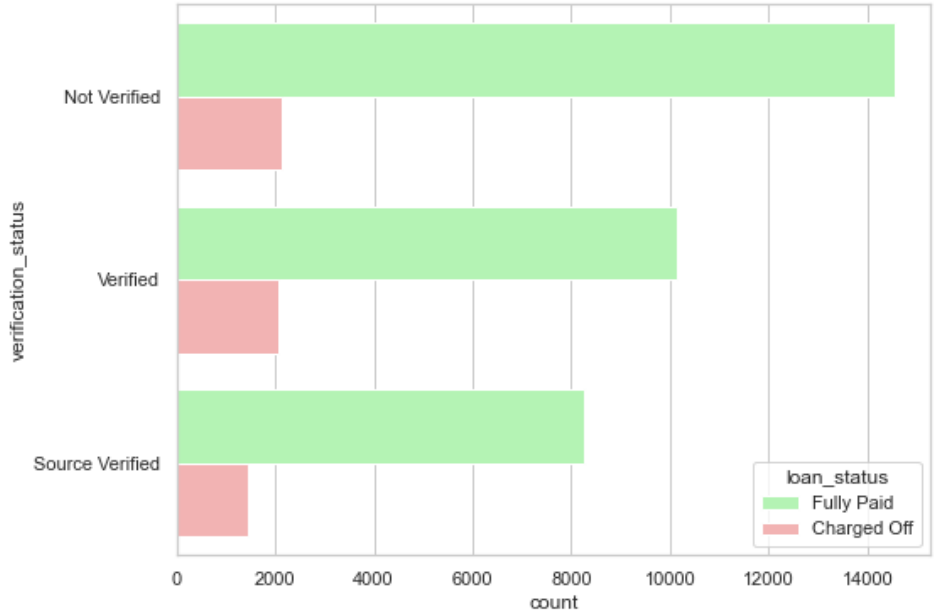




LOAN STATUS
(HOME OWNERSHIP, VERIFIED APPLICANTS)

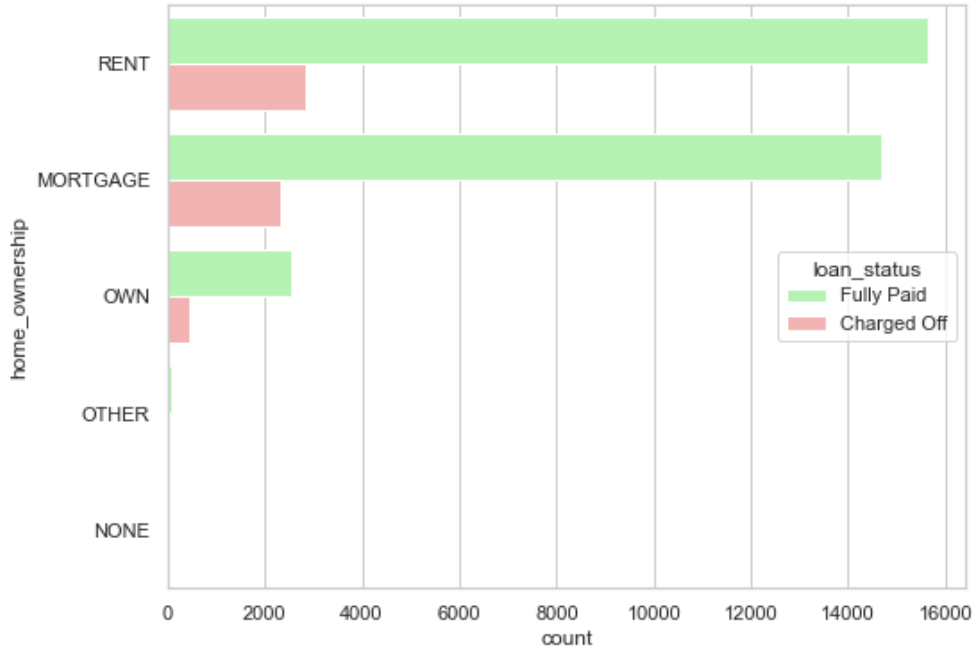
- The verified status for home ownership with the count is increasing in continues manner.
- The verification is a must for loan approval.
- Source verified applicants have the maximum chances of repaying loans

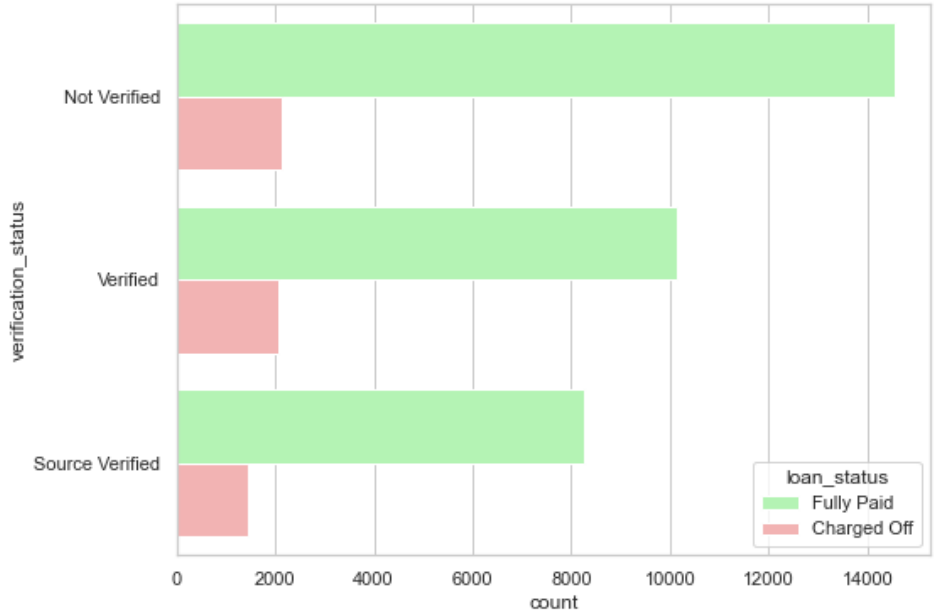




LOAN STATUS
(HOME OWNERSHIP, VERIFIED APPLICANTS)

- The verified status for home ownership with the count is increasing in continues manner.
- The verification is a must for loan approval.
- Source

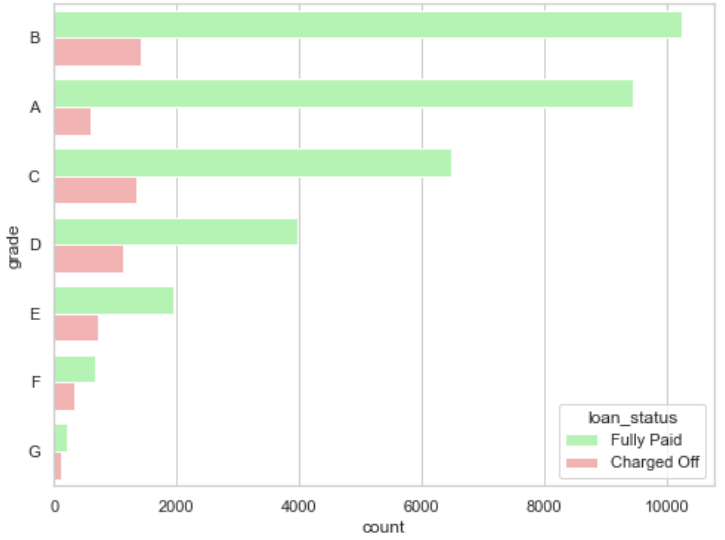


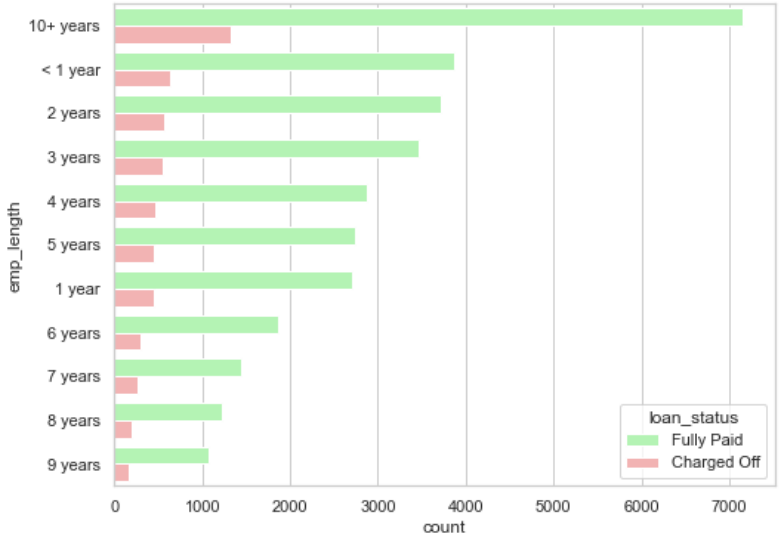


LOAN STATUS

(LOAN GRADES, EMPLOYMENT RECORDS)

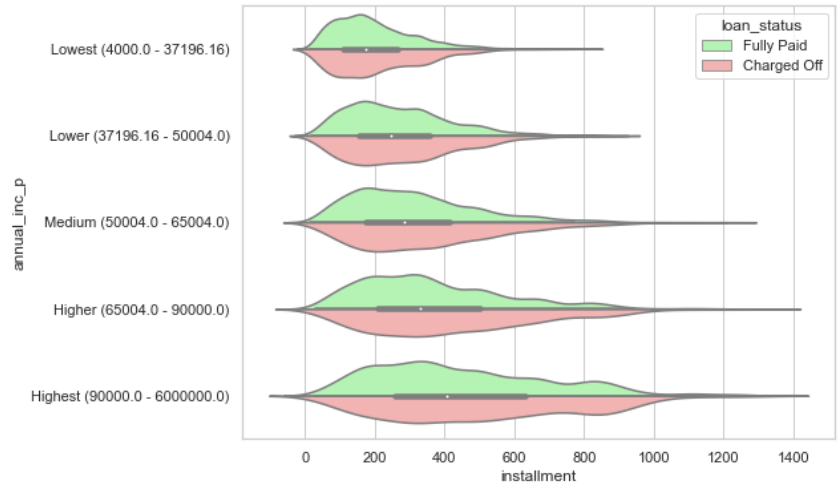
- The loan grades having highest default percentages. G, F, E and D form grades where default rate is much higher than others.
- The applicant having no employment will lead to Defaulted payment
- Applicants with more than (10+years) of experience pay the loan fully

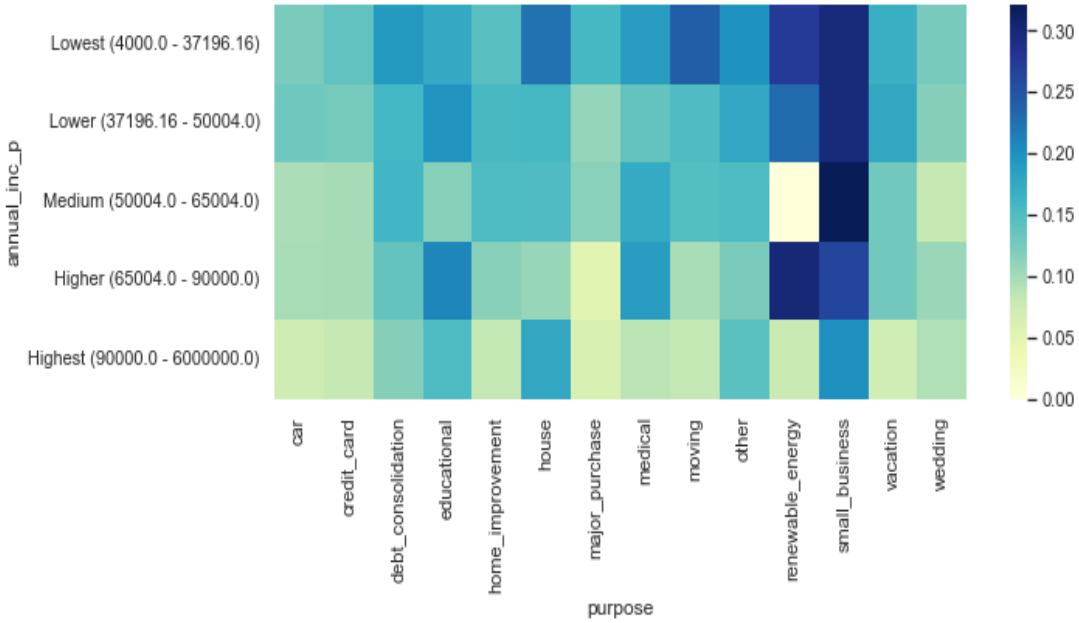




BIVARIATE ANALYSIS

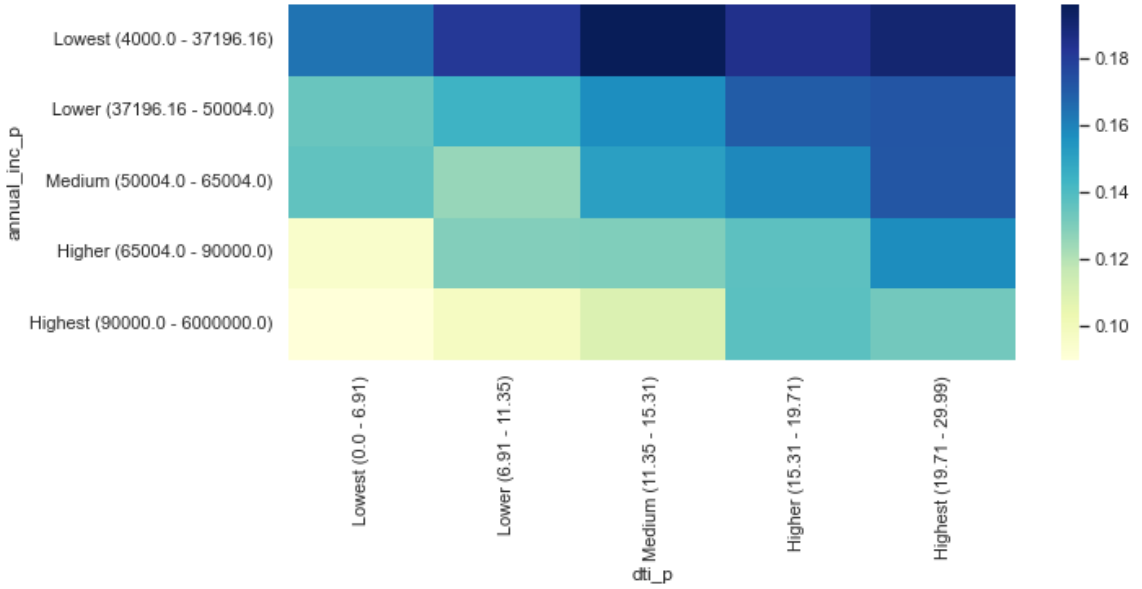
- The higher installments for any income group have a greater number of defaults.
- Plot of various income groups versus the risky purposes of loans for them.
  - small business loans for lowest and medium income group
  - renewable energy loans for higher income group





BIVARIATE ANALYSIS

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



RECOMMENDATIONS - 1

- Major 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)

RECOMMENDATIONS - 2

- Multifactor 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

RECOMMENDATIONS - 3

- 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)