Lending Club Case Study

Lending Club Case Study

- Author: Manas Ranjan Das
- Author: Palak Kour

Business Problem Statement

A consumer finance company, is one of 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. When the company receives a loan application, the company must decide for loan approval based on the applicant's profile.

- If the applicant is likely to repay the loan, then 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

Business Objective

The data for past multiple year's information about loan applicants and whether they ‘defaulted’ or not has been shared. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Solution Approach

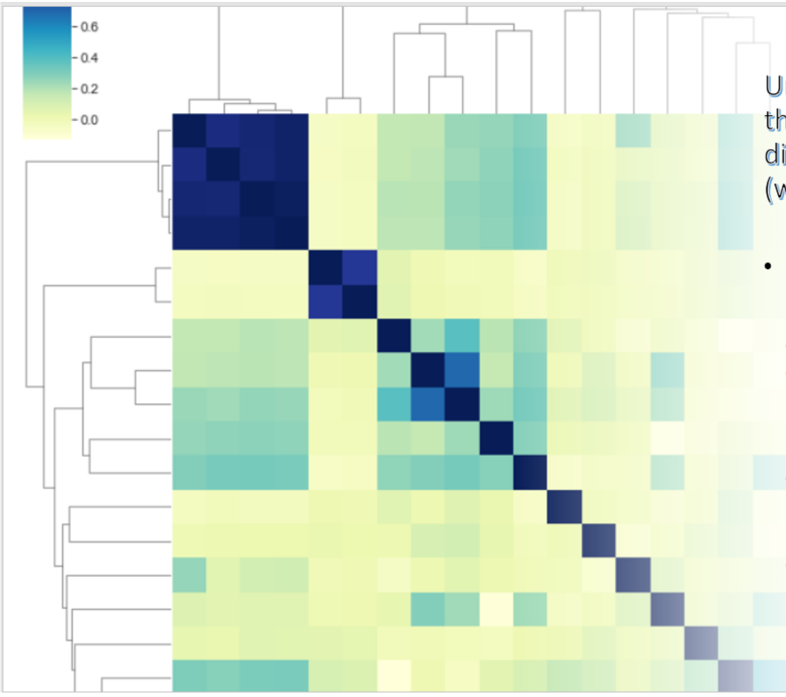
We will use EDA(Exploratory Data Analysis) to understand how consumer attributes and loan attributes influence the tendency of default.

Univariate Analysis (Single Attribute impact) Key Observations

- We have validated and can see the trend as higher the loan amount, the chances of loan default is more.
- We have also observed that higher is the interest, the greater is the chance of loan default.
- We have observed higher default percentages increase along with higher instalment amounts.
- The trend with higher annual income show lesser default percentages in our charts.
- We have observed that higher debt to income ratio impacts default percentages, and the trend is higher.
- We have also observed that total credit revolving balances slightly influence the loan default percentage. Higher the revolving balance, greater the chance of the loan getting defaulted in this data set.
- We have also found that revolving line utilization rate has high impact on the loan default percentage. When this increases, the charged off percentage also increases.

Univariate Analysis (Single Attribute impact) Key Observations

- We have observed that the loan repayment term plays a factor in predicting the default rate
- We have observed that for G, F, E and D form grades, the default rate is much higher than others
- We also observed that the G3 and F5 sub grades have default rate greater than 40%. This field is a clear indicator of loan default percentage.
- We have observed that Verified applicants have higher loan default percentage. Hence, we infer that there is gap in the verification process
- We have observed that loans was taken for small_business, renewable_energy and educational purpose have higher risks associated with it in terms of loan default cases.
- We have observed that NE, NV, SD, AK, FL, MO states from all the list show higher risks towards loan default.
- We found that applicants with non-zero derogatory public records have greater chances of loan default percentage.
- We have observed that higher is the number of public bankruptcy records, greater is the chance of loan default percentage.
- We have observed that the December month(**As this specific period coincides with festive season**) has the highest number of loan applications per year and the biggest default ratio. Similarly, the Month of May has similar trend as oy coincides with summer break and right before the Memorial day and Independence day breaks in US.

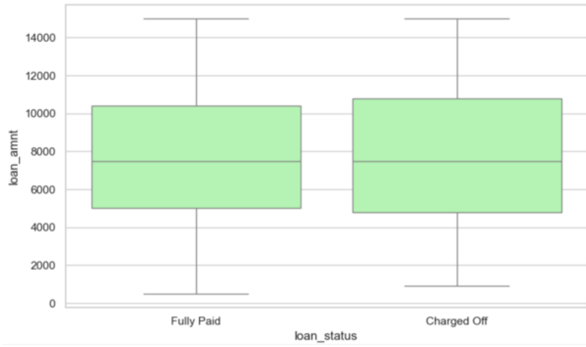


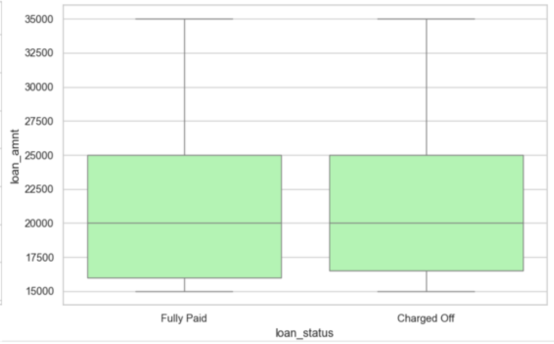
Univariate Analysis: validate the corelation between the different numeric attributes (with high correlation value)

- **Observation: As darker shades indicate heigher correlation , we infer that loan_amnt, funded_amnt, funded_amnt_inv and installment have huge correlation.**

  The public records related fields pub_rec, pub_rec_bankrupcies and number of accounts related fields open_acc & total_acc are correlated.

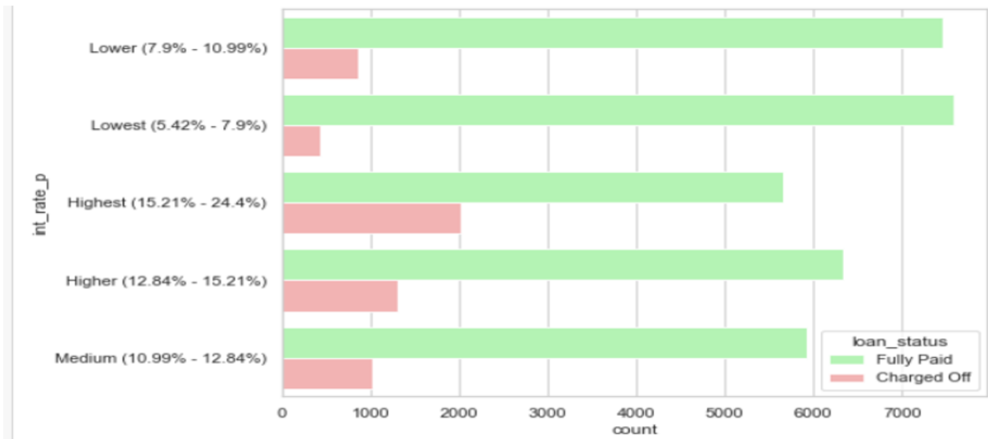
Boxplot findings on the loan_status against loan_amnt





Observation: As there is difference between mean and std, we are considering 75% as limit. The first one is less than 75% and 2nd one is above 75%
if we compare both the trends that higher the loan amount, the chances of Charged Off is more.

Univariate Analysis: Int_rate

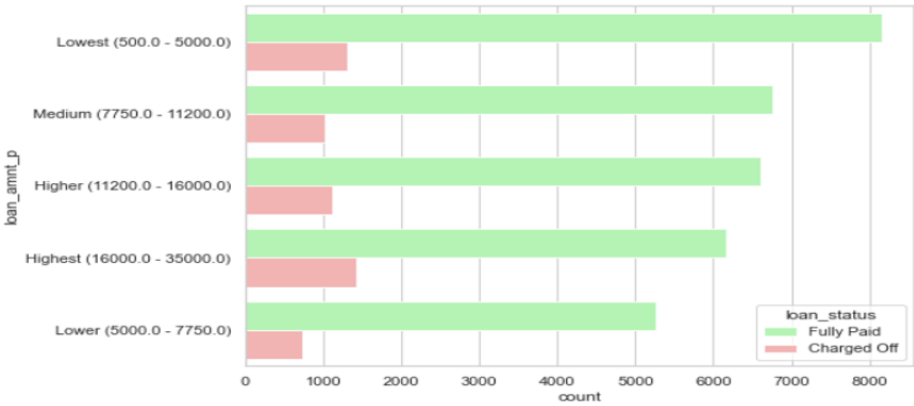


Observation: We have observed that higher is the interest, the greater is the chance of loan default.

Univariate Analysis: Installment

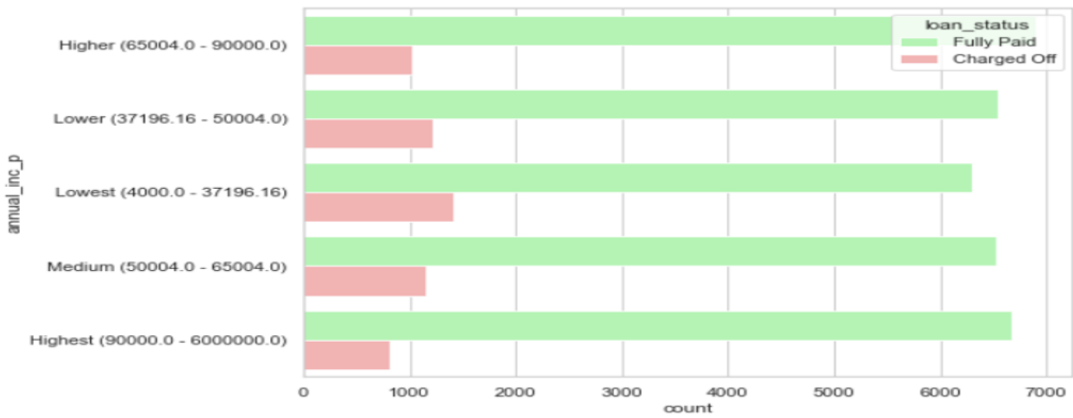
Observation: We have observed that higher instalment amounts show higher default percentages.

Univariate Analysis: loan_amount



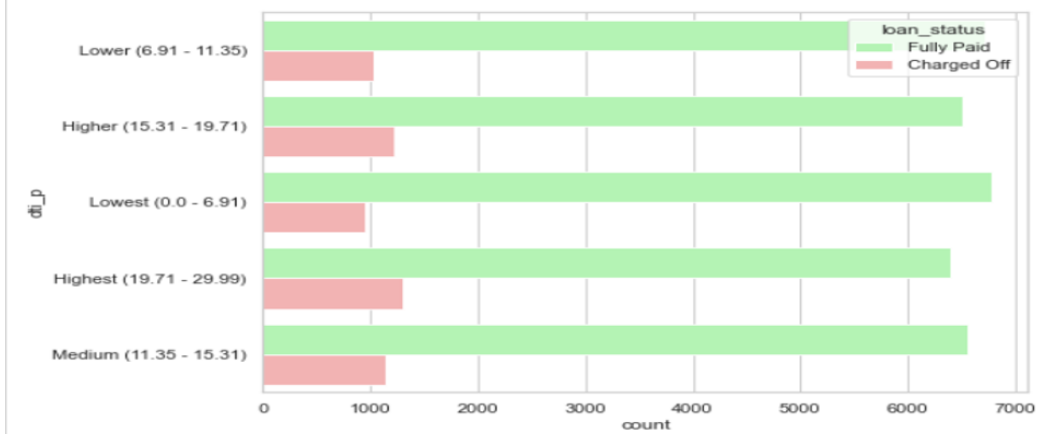
Observation: We have observed that higher is the loan amount, the greater is the chance of loan default.

Univariate Analysis: annual_inc



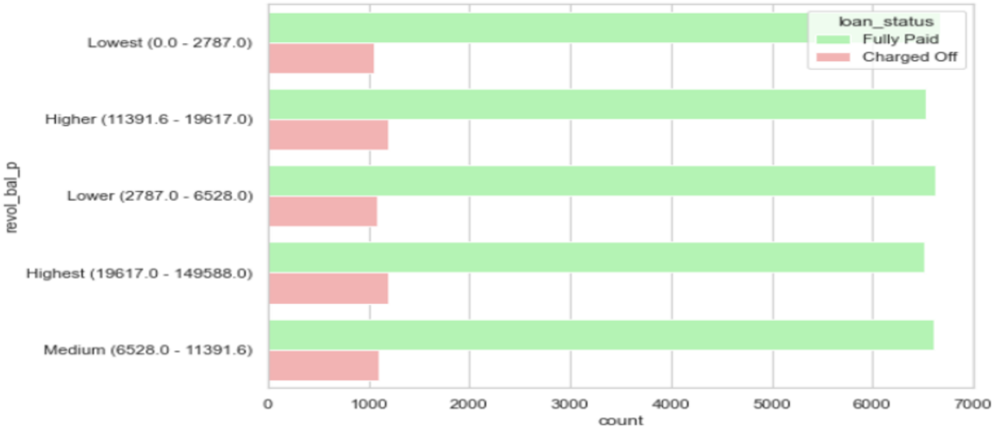
Observation: We have found that higher annual income show lesser default percentages.

Univariate Analysis: dti



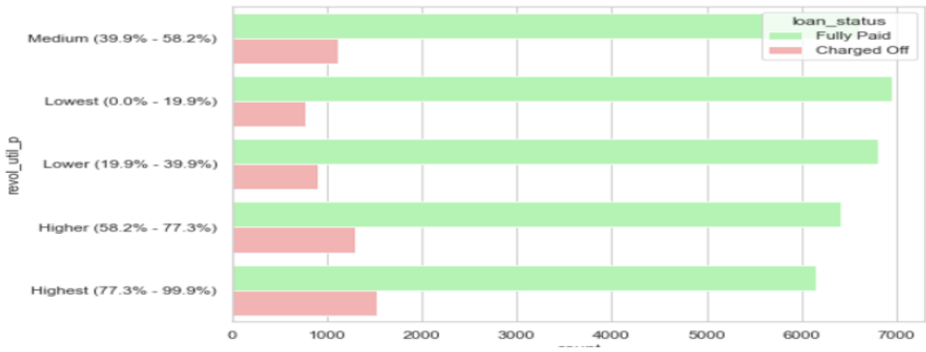
Observation: We have found that higher debt to income ratio show higher default percentages.

Univariate Analysis: revol_bal



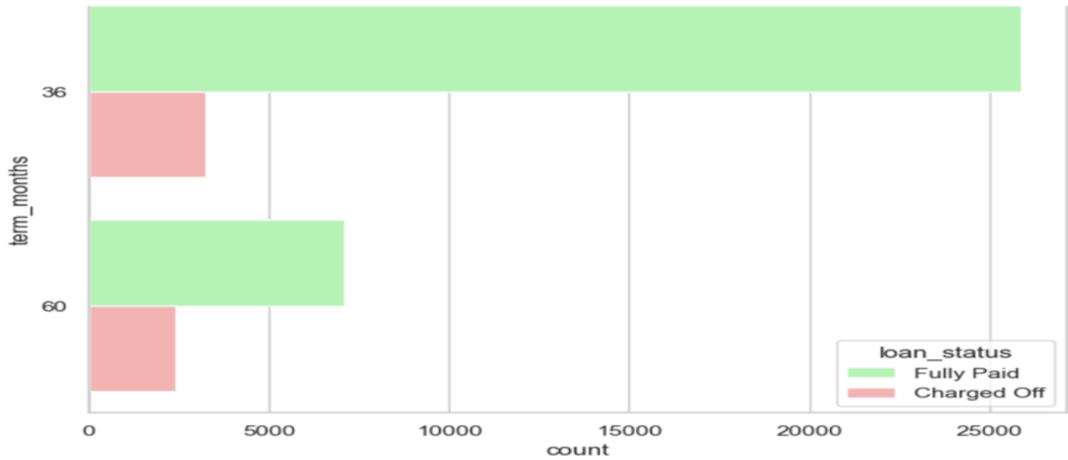
Observation: We have found that total credit revolving balances slightly influence the default percentage.

Univariate Analysis: revol_util



Observation: We have observed that revolving line utilization rate has high impact to the loan default percentage.

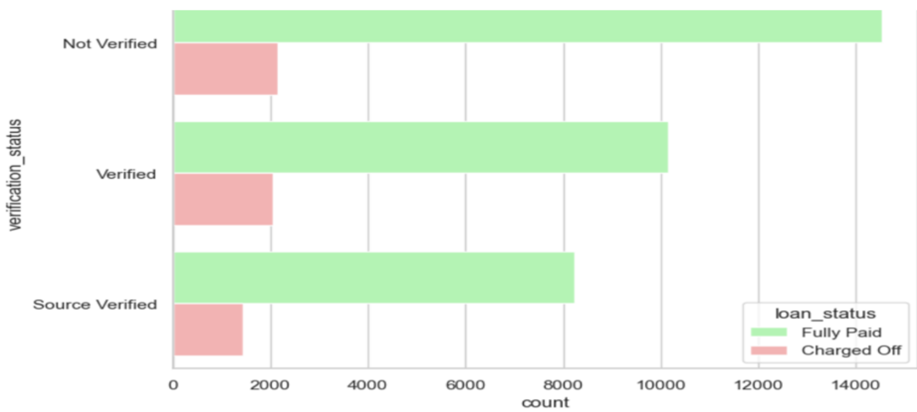
Univariate Analysis: term



Observations:

- For loans with 5 years repayment term, the default percent is 25% of the cases.
- For 3 years loan repayment term, the default is only for 11% of the cases.
- Hence loan repayment term plays a factor in predicting the default rate.

Univariate Analysis: verification_status

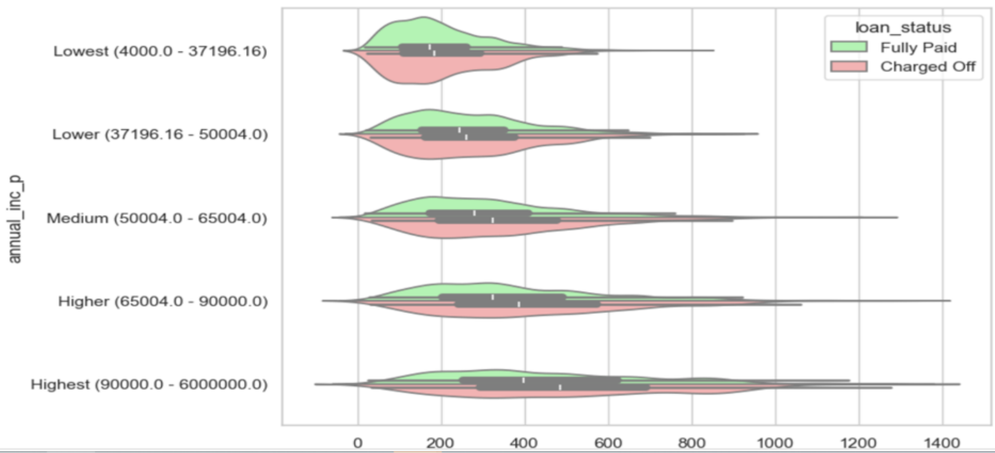


Observation: We have observed that the Verified applicants have higher loan default percentage. Hence, there is gap in the verification process, and it needs special attention.

Bivariate Analysis (Combined attribute impact) Key Observations

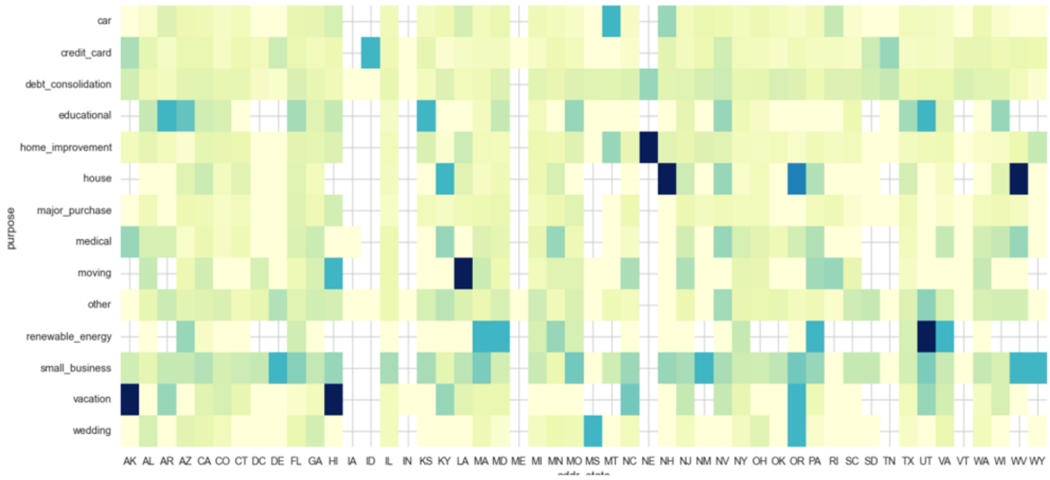
- We have observed the **loan repayment term** plays a factor in predicting the default rate percentage parameter.
- We have observed that higher **instalments vs. income group** have a greater number of loan defaults percentage.
- We have also observed that greater is the correlation of **addr_state** with **Purpose** of the loan, the greater is the chance of loan defaults percentage.
- The comparison of **annual income vs. loan purpose**,
    - we observed that small business loans for lowest and medium income groups have high chances of loan default.
    - The Renewable energy loans for higher income group has similar trend to have high chances of loan default.
- We have also compared **annual income with debt-to-income group**, we have observed that medium debt-to-income group in the lowest income range shows high trend of loan default

Bivariate Analysis:'installment','annual_inc'



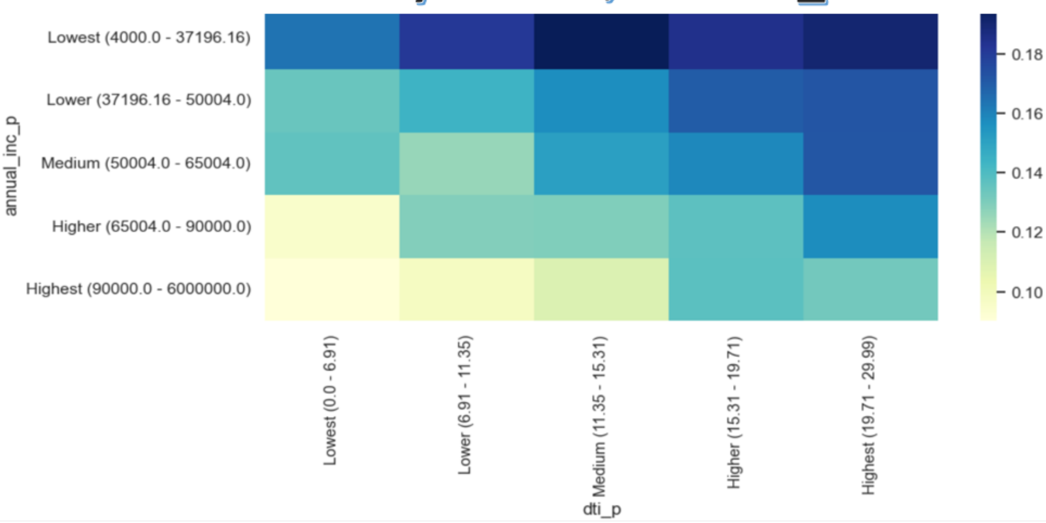
Observation: We could see the higher instalments for any income group have a greater number of loan defaults.

Bivariate Analysis:'addr_state', 'purpose'



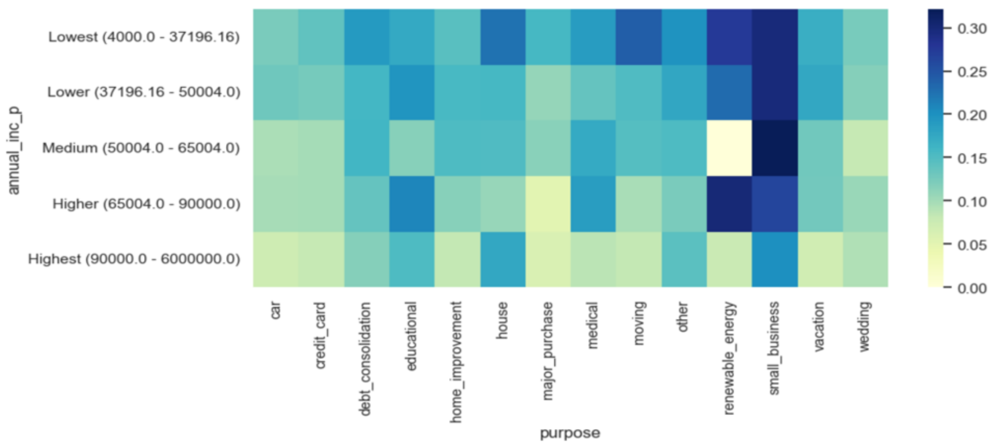
Observation: We have observed that darker the intersection of addr_state has with the purpose of the loan, the greater is the chance of loan default percentage.

Bivariate Analysis:'dti','annual_inc'



Observation: We have also observed that medium debt-to-income group in the lowest income range shows high trend of loan default.

Bivariate Analysis:’purpose','annual_inc'



Observation: By comparing the annual income with loan purpose we observe that small business loans for lowest and medium income groups have high chances of loan default.
Renewable energy loans for higher income group show the same trend

Trend from Analysis

We have categorized our results under 2 categories

The Attributes showing lesser impact to the trend of loan default

The Attributes showing higher impact to the trend of loan default

Lesser impact

Lower annual income (below around 37000)

Higher loan amount (above around 16000)

Higher debt to income ratio (above 15%)

Higher instalment amount (above 327)

Loan issue month (Dec, May)

Applicant’s address state (NV, SD, AK, FL)

Higher impact

Higher interest rate (above 13%)

Repayment term (5 years)

Loan purpose (small business, renewable energy, educational)

Higher revolving line utilization rate (above 58%)

Loan grade & sub-grade (D to G)

Public bankruptcy records (1 or 2)

Derogatory public records (1 or 2)

Missing employment record

Conclusion

- Based on our analysis and validation towards effects of combined attributes on the tendency of loan default, please find the details as below:
  - High instalment and longer repayment term
  - Residential state and loan purpose
  - Income group and loan purpose
  - High loan amount and interest rate for lower income group
  - Home ownership (other) and loan purpose (car, moving or small business etc.)
- **Recommendations**
  - The Organization should be very cautious in approving the loans with high amount and interest rates to lower income group applicants.
  - The Organization can structure the loans in a manner to avoid high instalment and longer repayment terms.
  - The Organization should be careful in approving small business loans to low- and medium-income groups. The same situation with renewable energy loans for high income group. The verification process should be in place stringent the process.
  - The Organization should validate employment records and should not approve loans for applicants with no proper details.
  - The Organization should try to avoid approving loans to applicants having Derogatory public records and Public bankruptcy records (1 or 2).
  - The Organization should revisit applicant verification process as we observed that verified applications have more tendency on loan default.

THANK YOU