Risk analytics

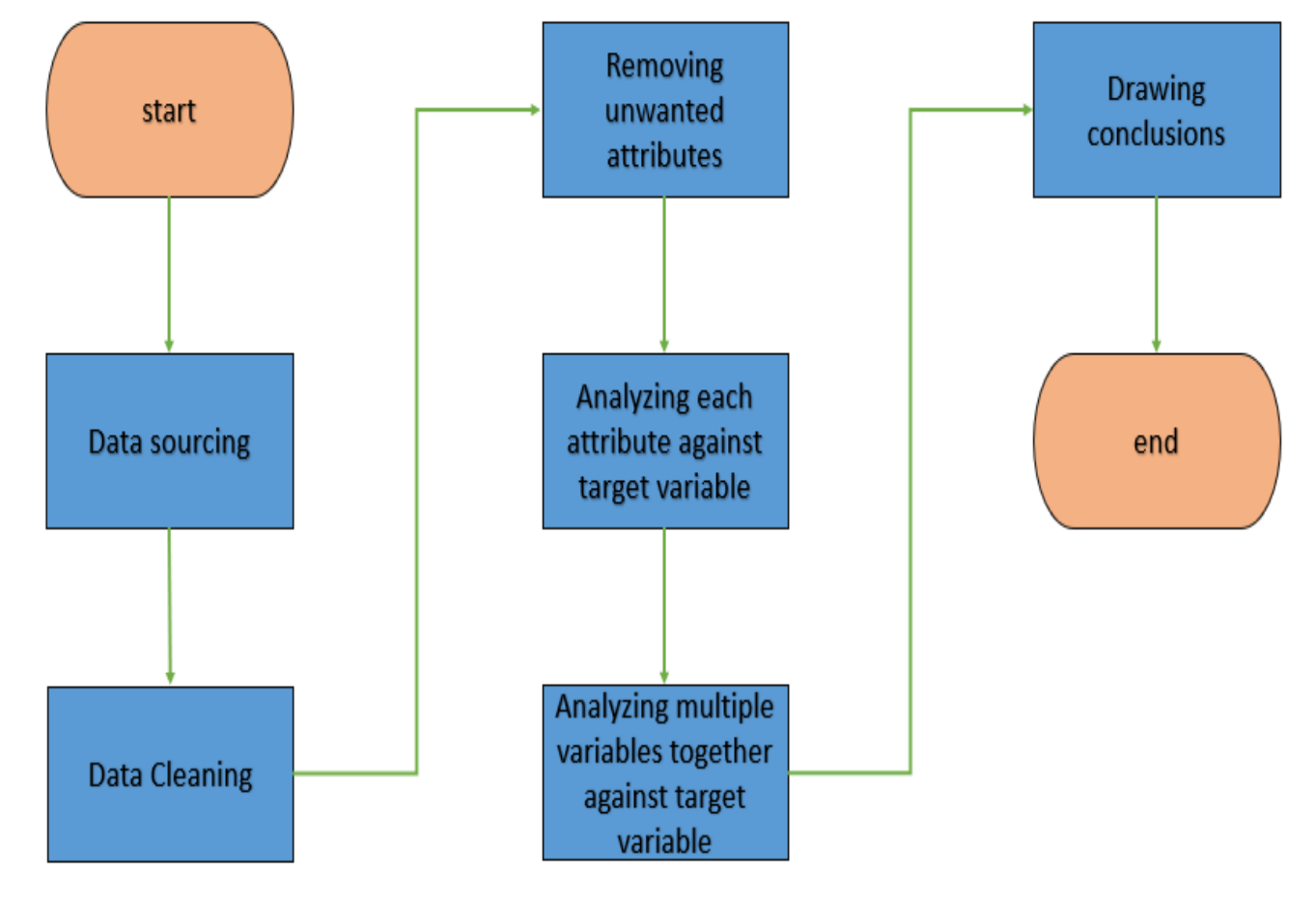
- Business Objective: To identify the driving factors behind the loan default (i.e. the variables which are strong indicators of default.

- Description:

  - Lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Our goal is to find such patterns which are indicators of loan default.
  - To achieve this goal, the methods of EDA i.e. exploratory data analysis has been followed. Even though EDA has no exact process / set of rules that commonly applies to all the problems, it does have general steps that can be followed and it needs the corresponding domain knowledge.

- Dataset contains complete loan data for all loans issued through the time period 2007 to 2011. There are no records of the applicants where loan was rejected.

Flow of analysis



- Data is sourced from lending club.
- Data is treated for handling missing values, removing duplicity, standardizing the values and units.
- As per business judgement, we will be filtering out customers which are still in the process of loan repayment and are not yet considered defaulters.
- Then we will carry out univariate and bivariate analysis using visualization to get the insights.
- At last we will be making an inference from the analysis.

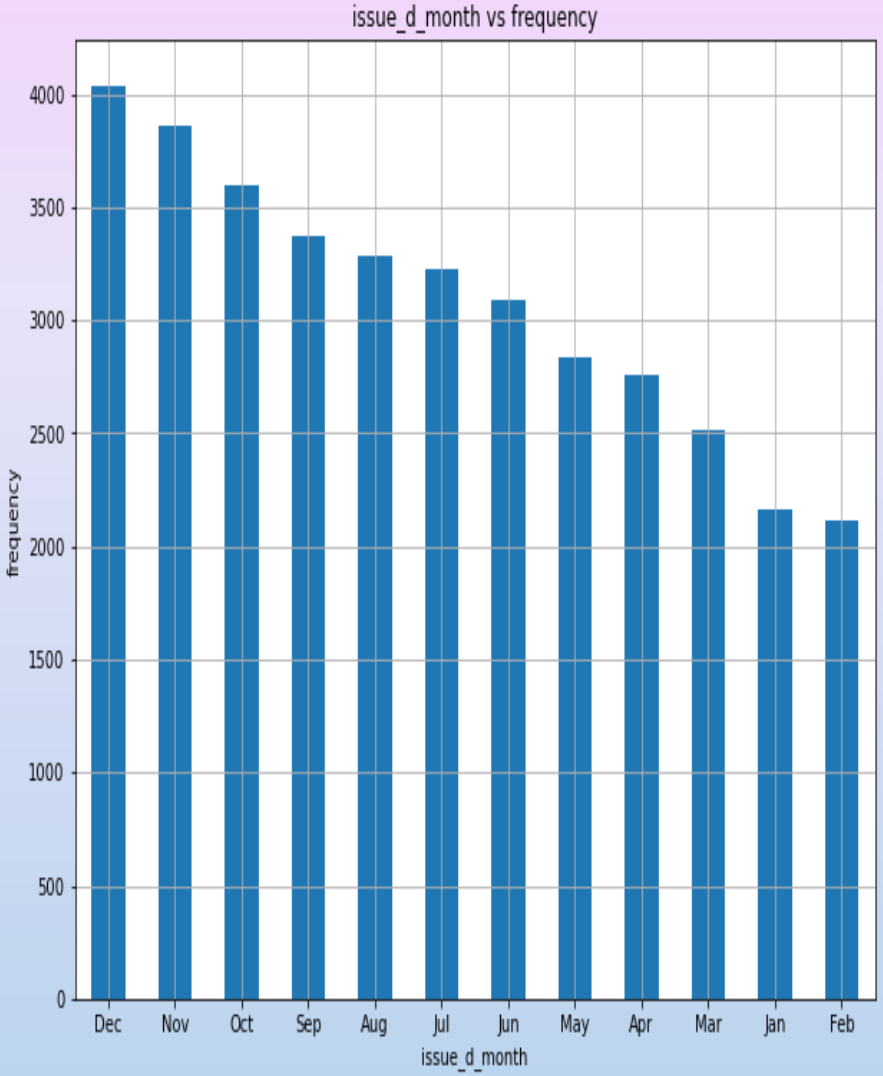
Data Cleaning

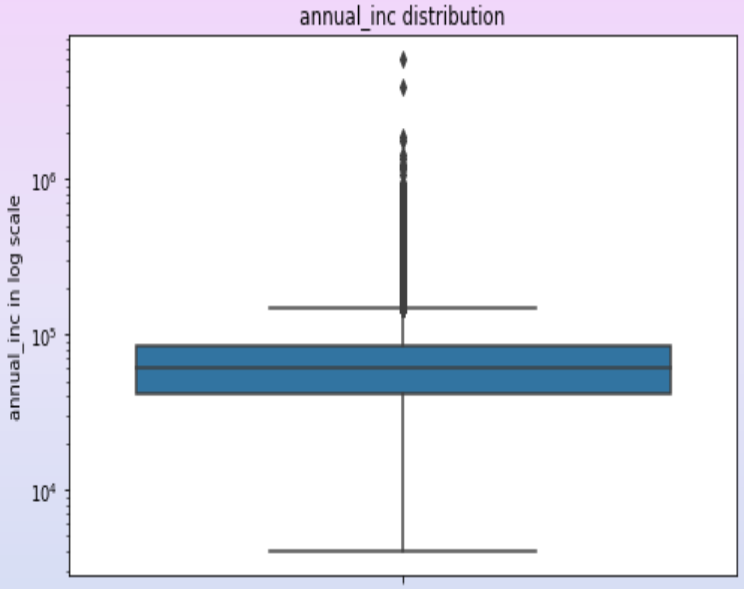
- After removing columns with all missing values, we are still left with large number of attributes. We will further narrow them down so we can prioritize on strongest driving factors of default. For that we will divide dataset into two types: financial attributes and customer behavioral attributes. Since at the time of loan approval, latter may not be available with us to help us make decision, so we will be removing behavioral attributes.

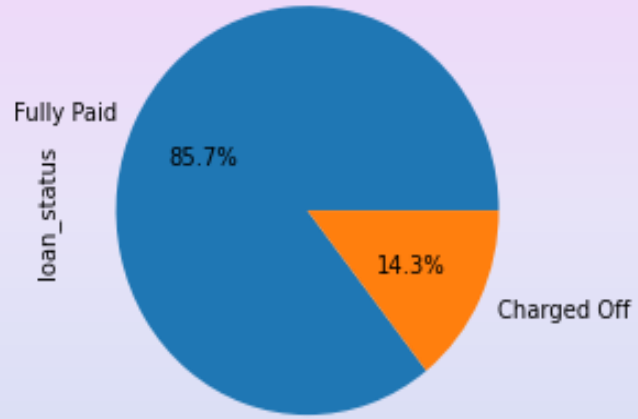
Financial attributes	Customer behavioral attributes
Annual income, DTI, loan term, employee serving experience, interest rate, loan amount, Installment, grade, sub-grade, public records, public record bankruptcies, purpose, loan status etc.	Last payment date, last payment amount,revolving balance, revolving balance utilization, total payment, total payment investment, total received principle, total received interest, total received late fee, recoveries, collection recovery fee, last credit pull date etc.

- We will be removing the missing values instead of imputing them to avoid introduction of bias since very low percentage of missing values are present in attributes that interest us.

Understanding the data with univariate analysis

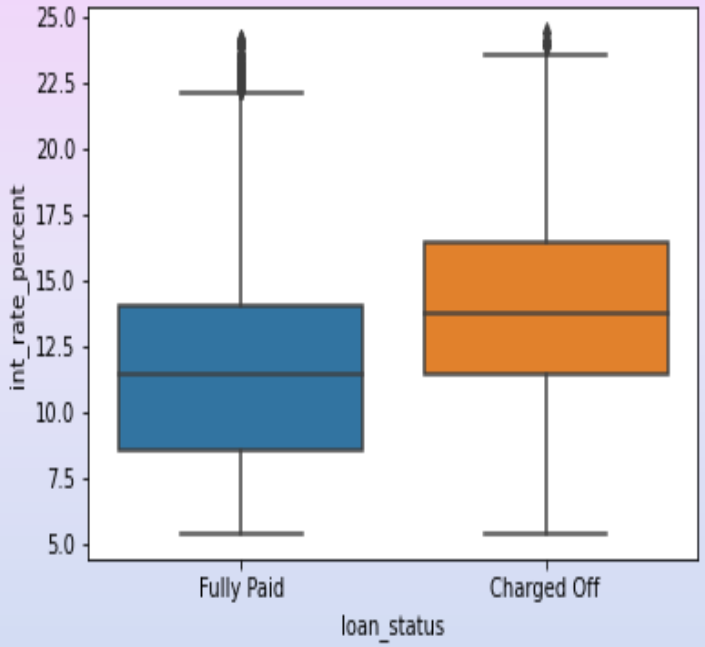


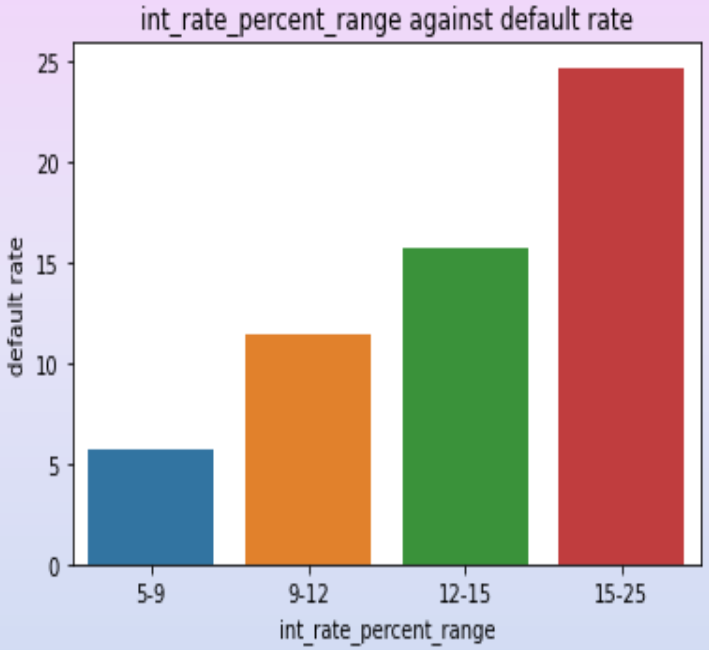


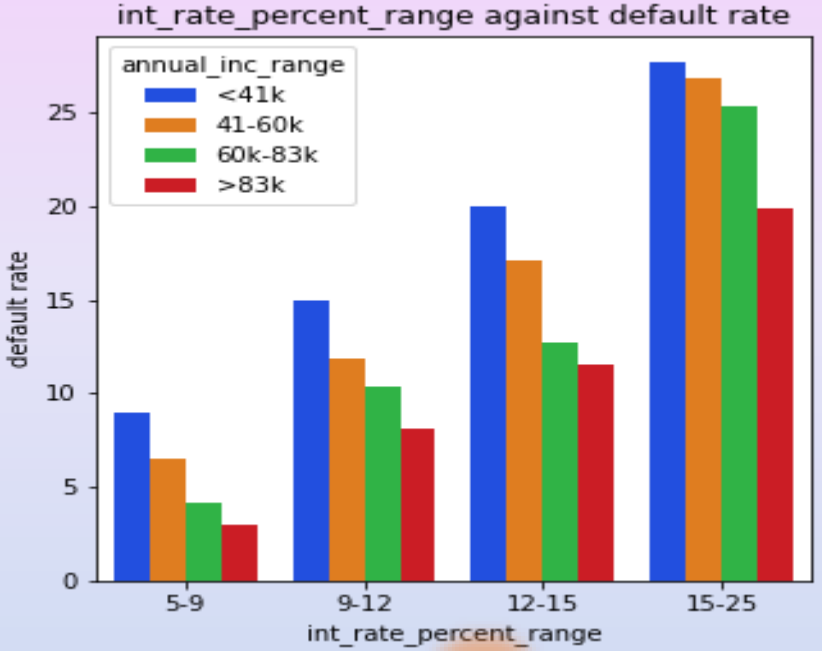


Even though univariate analysis will not help us much to determine driving variables, it will be definitely helpful for understanding the data. E.g. It can be seen that maximum loans are issued in December. There is large deviation in annual income. We have 14.3% of defaulters in our dataset.

Getting hidden patterns with bivariate analysis

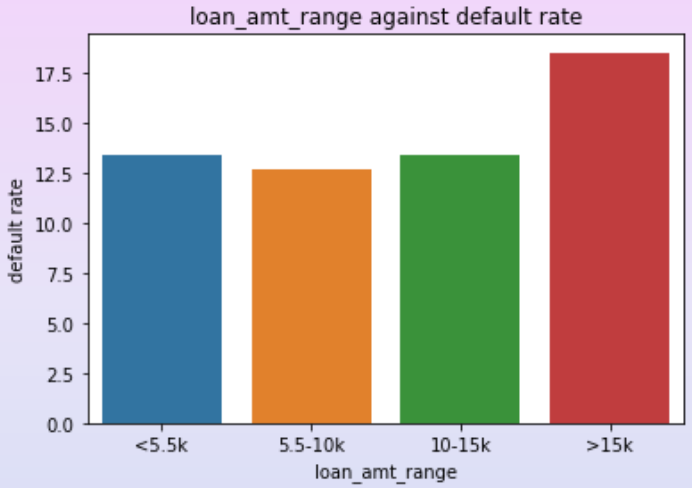


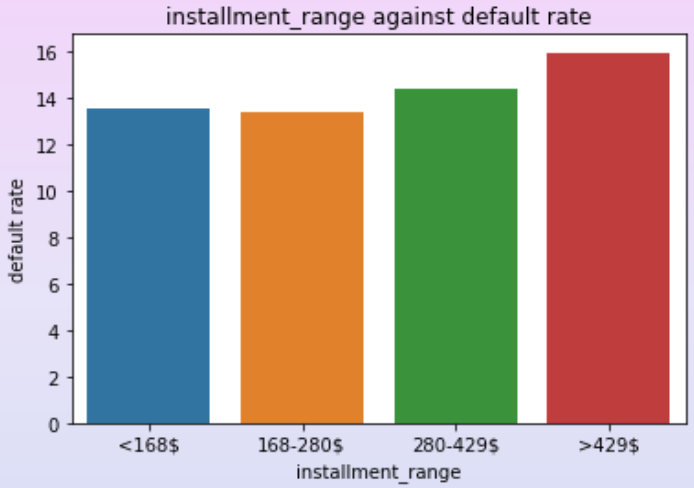


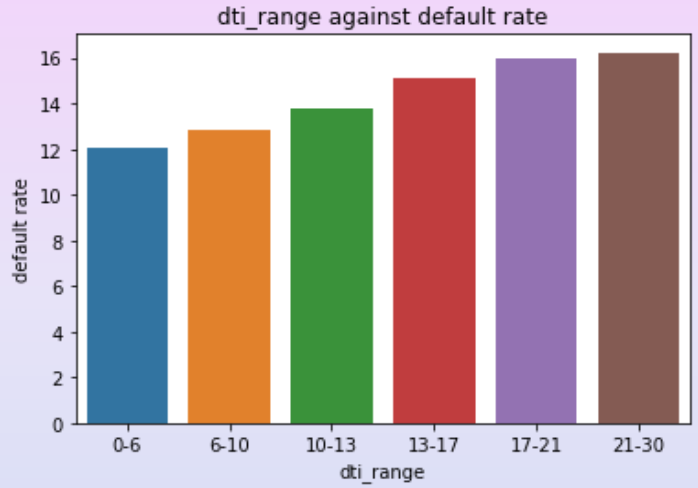


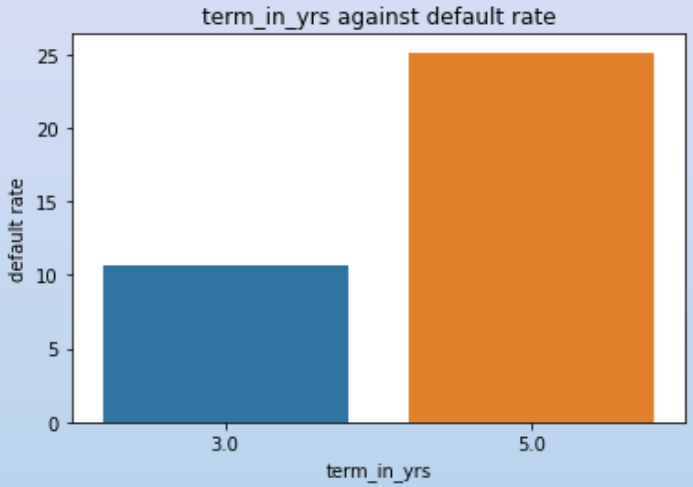
- Median value of interest rate is higher for ‘charged off’ than ‘fully paid’ loan status.
- As interest rate is increasing, rate of default is also increasing.
- For each range of interest rate, default rate is maximum for customers with lowest income range.

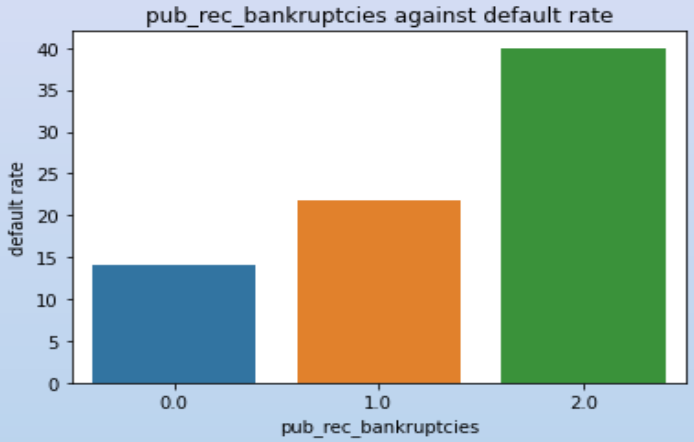
➢ Charging customers with very high interest rate may reduce the damage that will cause incase of default, but it will also increases chances of default. So interest rate should be fine tuned.



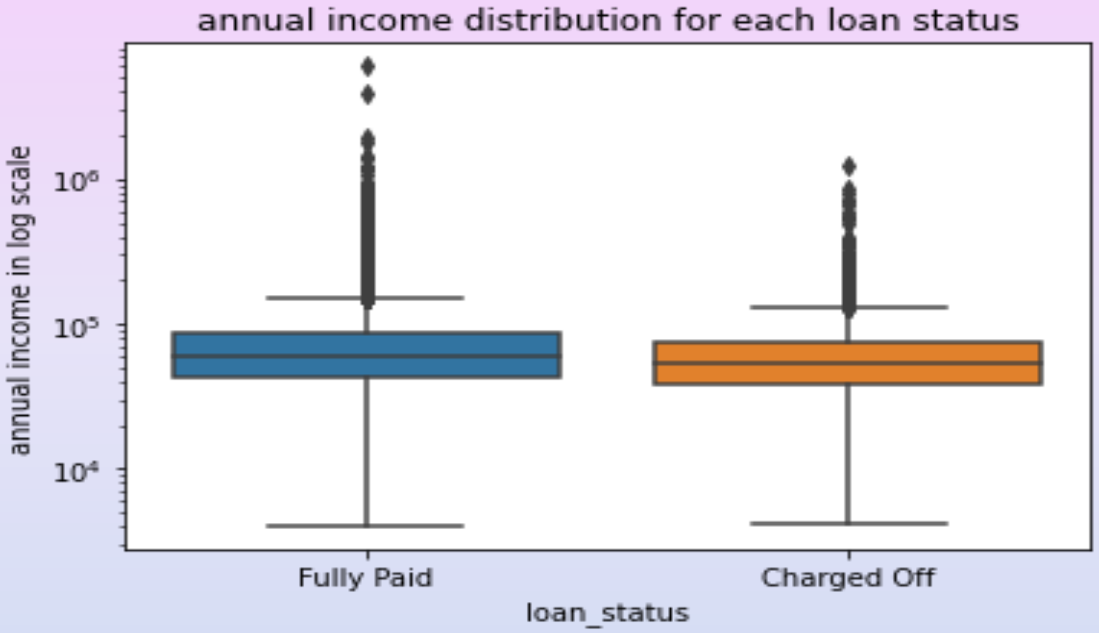


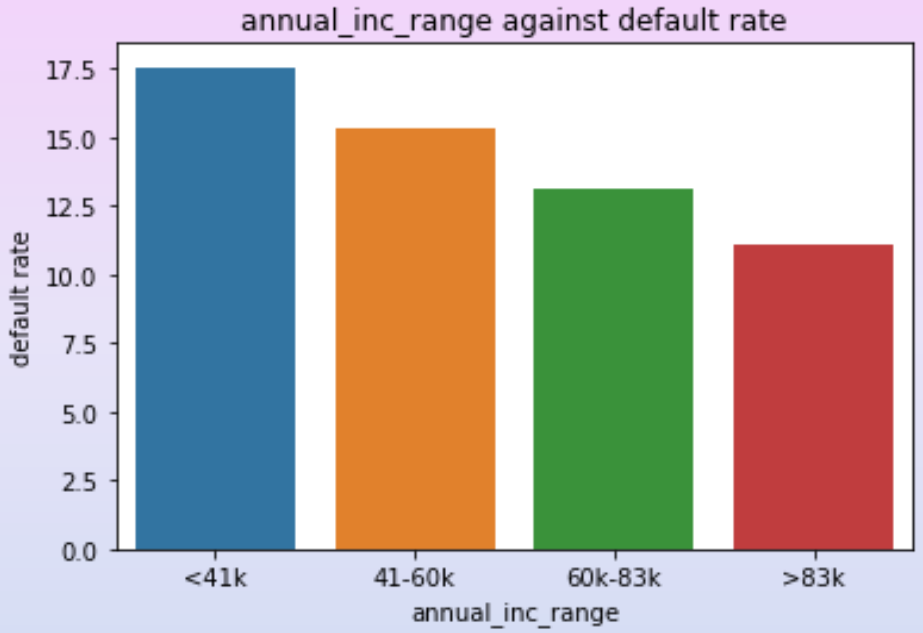






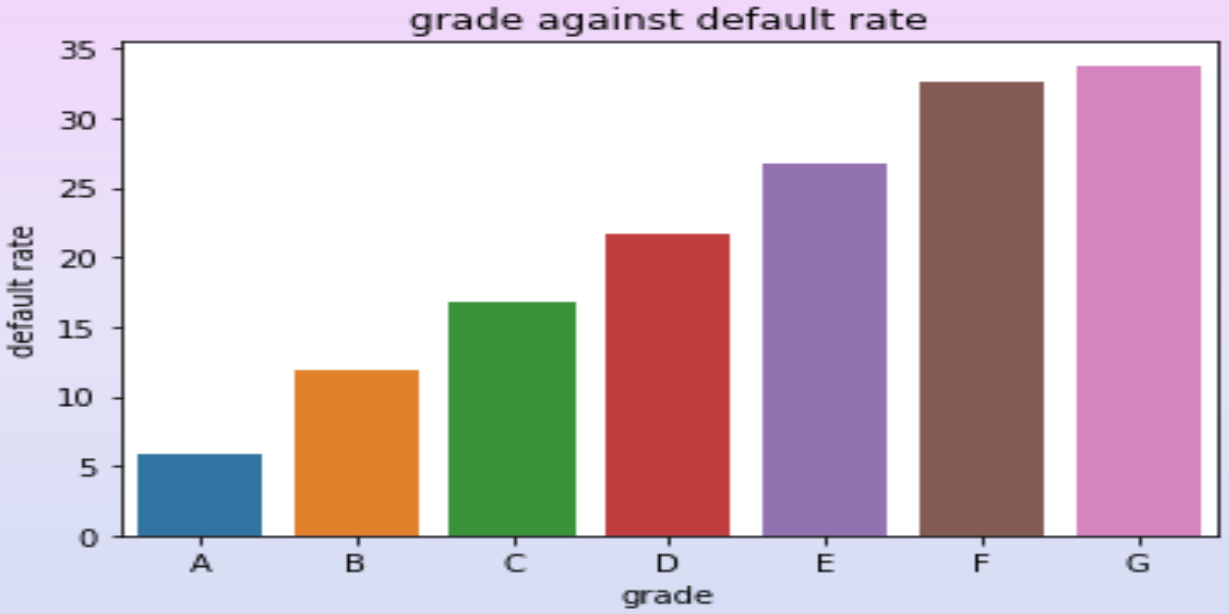
We can clearly see that default rate is increasing with increase in term, dti range, installment range loan amount and pub rec bankruptcies. Thus these can be clearly used as default indicators.

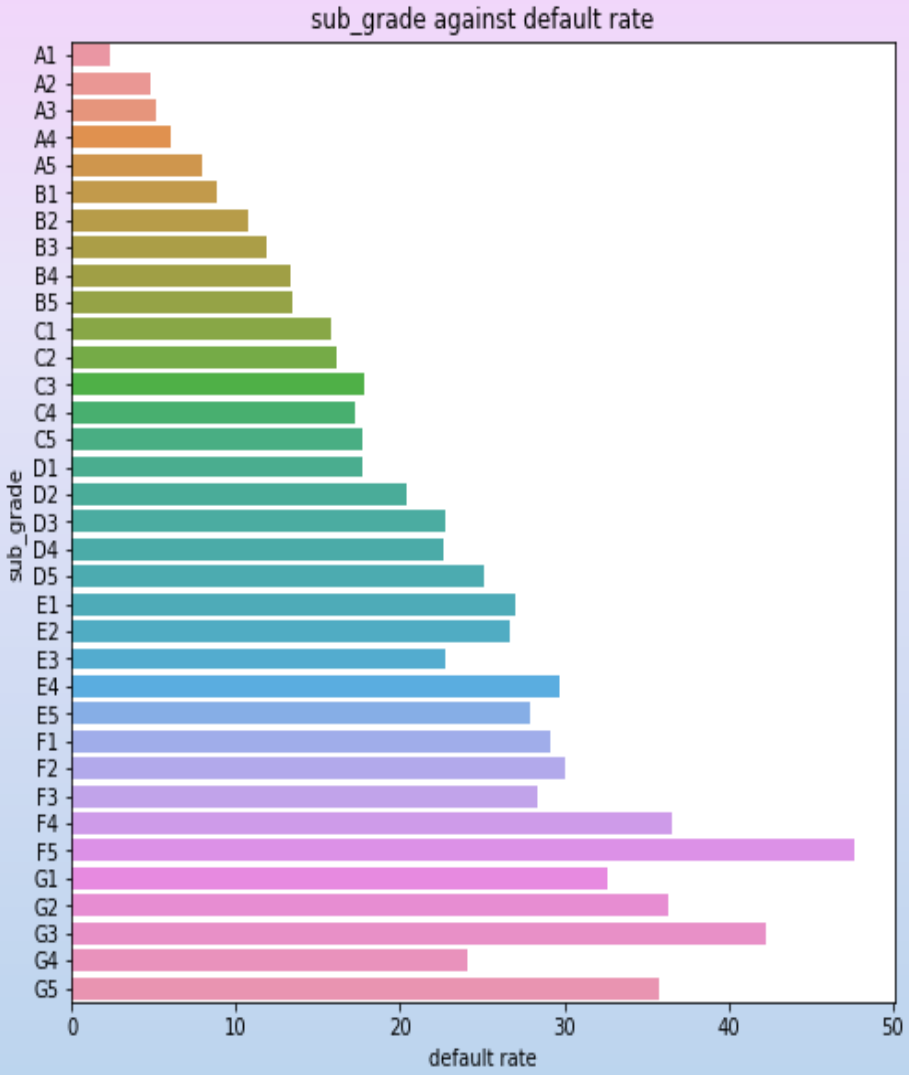




- Approx. same median value of annual income for each loan status type.
- Default rate is increasing as income range is decreasing.

➢ Lower income range appliers should be denied of loan or charged higher but appropriate interest rate.

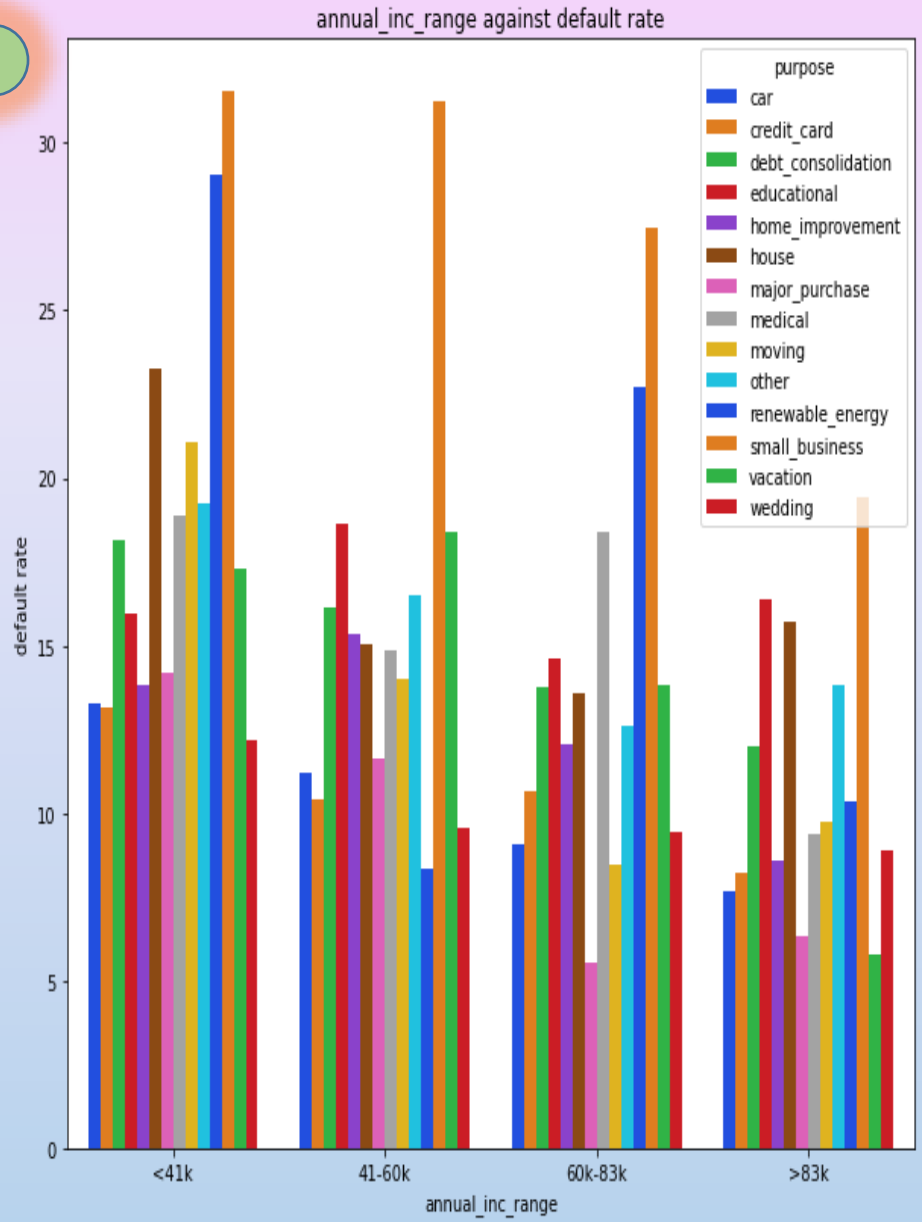


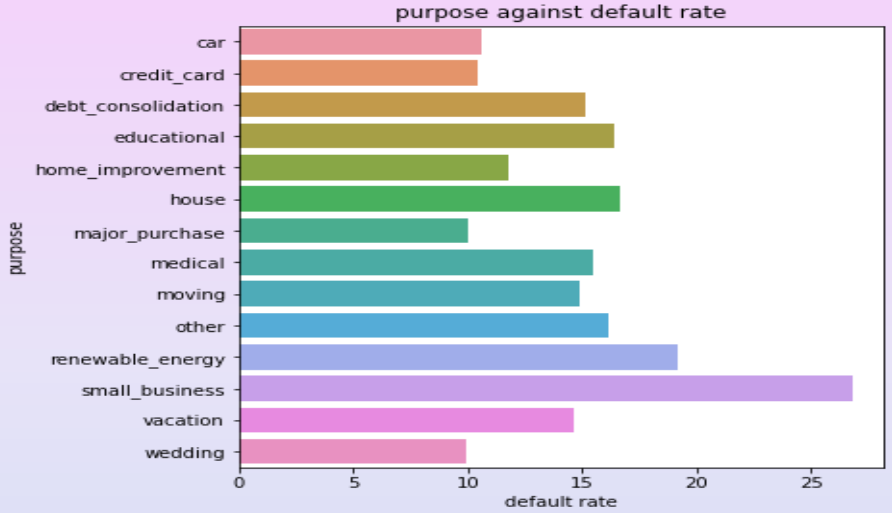


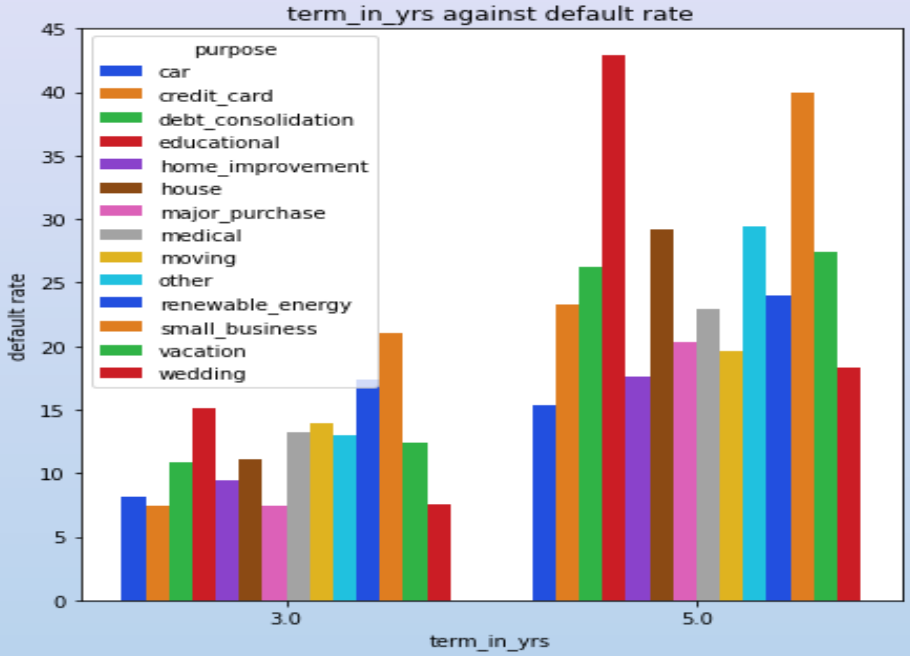
- As grade increases from A to G, default rete is also increasing.
- Till C5 sub-grade, default rate is around 20%.

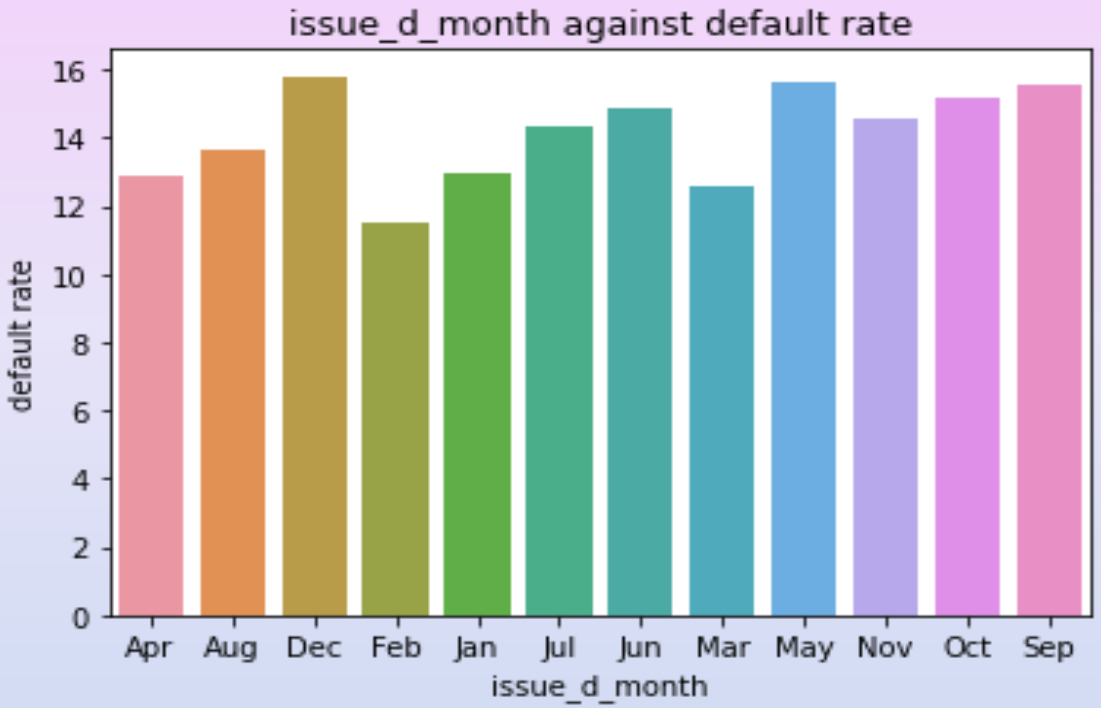
➢ Granting loan for grade above c should be done cautiously.

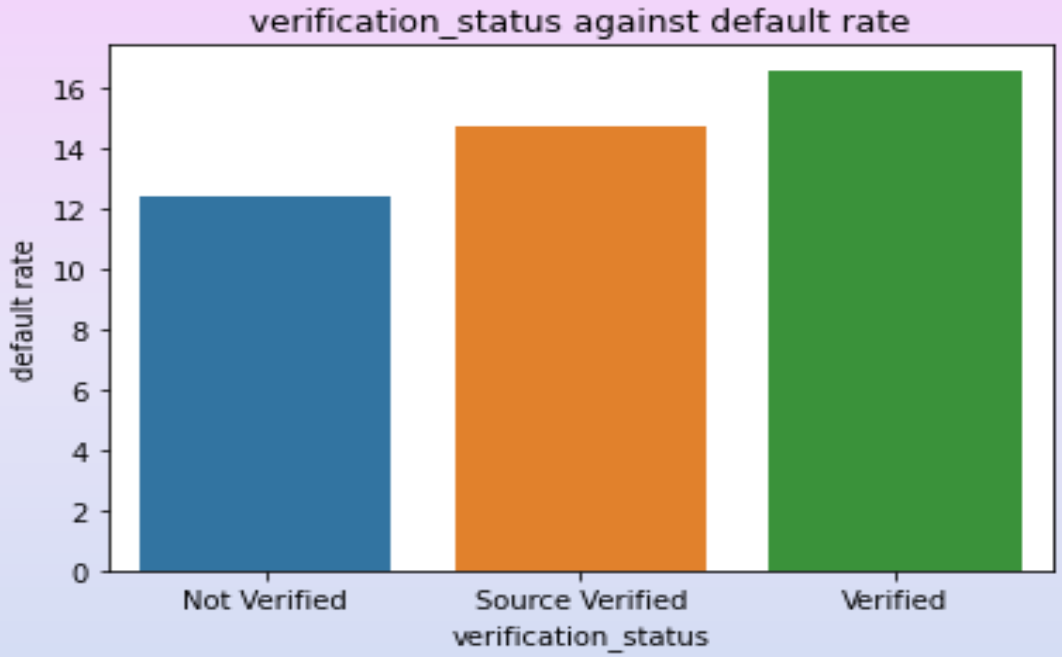
- Default rate is maximum for purpose category ‘small business’ which is followed by ‘renewable energy’ and educational
- For every annual income range, Default rate is maximum for small business type. it is followed by renewable energy category for the income range '<41k' and '60-83k'











- default rate is maximum for month of December, May. It may be due to the fact that people spend more on festive season such as Christmas and memorial day which takes place in month of December and may.
- Default rate is observed to be maximum for verified category and least for not verified. Which is a bit contradicting to general thought of clients with 'not verified' income would be more likely default and the ones with verified have more credibility and thus less likely to default. Perhaps verification criterion/process is flawed and need to be examined.

Inference

- Default rate is maximum for purpose category ‘small business’. Which is followed by ‘renewable energy’ and educational category.
- As sub grade moves from A to G (A1 to G5), default rate is also generally increasing. As per business criterion threshold grade can be set such that for applicant with grade above this loan is rejected.
- As term duration increases default rate also goes up. We can reduce the damage by charging suitable higher interest rate for longer terms.
- Loan amount is also an indicator of default rate. Default rate is higher for higher loan amount. It is maximum for loan amount >15k.
- Default rate increases with decrease in annual income. Loans should be sanctioned very cautiously for those with annual income <41k.
- DTI is also a very strong indicator of default. By setting threshold of dti value according to business objective, we can avoid credit loss.
- Installment amount, interest rate are also indictors of defaulters as they are positively correlated.
- To avoid business loss, if we are approving risky loans with much higher interest rate in order to reduce credit loss incase of default, then also we are increasing chances of default. So interest rate should be finely tuned.
- Applicants with two or even one public record bankruptcies should be denied of loan.