Lending Club Case Study

Group Members:
Lakshay Saxena
Nishtha Choubey

Business Understanding

Lending club a finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

•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

Objective

The company wants to understand the driving factors (or driver variables) behind loan default (loan_status = 'Charged Off'), i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Analysis Approach

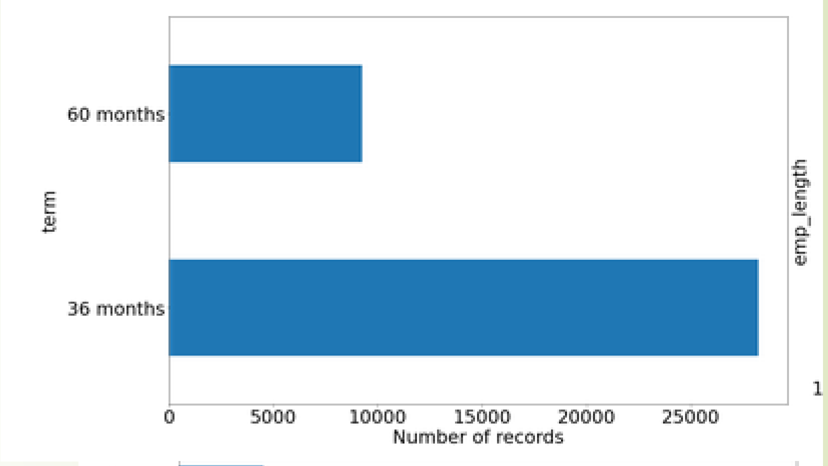
1. Data Cleaning.
2. Data preparation.
3. Derived Metrics.
4. Univariate Analysis.
    - Categorical Variable
    - Continuous Variable
5. Univariate Segmented Analysis.
    - Categorical Variable
    - Continuous Variable
6. Bivariate Analysis.
    - Categorical Variable
    - Continuous Variable
7. Summary

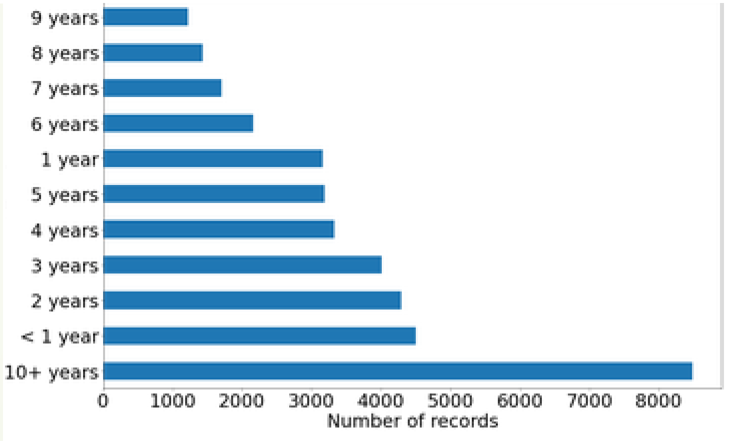
Steps for Data Cleaning

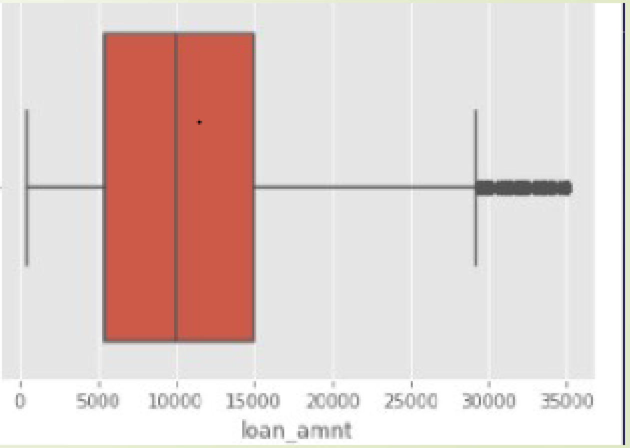
1. Checking for null values and there percentage in dataset.
2. Removing columns and rows containing more than 30% null values.
3. Removing unnecessary variables and their respective rows.
4. Remove duplicate variables and rows if any existed.
5. Converting column in proper format. Ex- removing % sing and converting required column to float type
6. Deriving required month, year and no of days column from date column

Univariate Analysis

- Charged off accounts constitutes of around 14% of loan data
- Customer prefer to take loans for shorter duration of time
- Customer with high employment duration(10+) tend to take more loans
- Loan Amount is in range of 5-15k and has lot of outliers

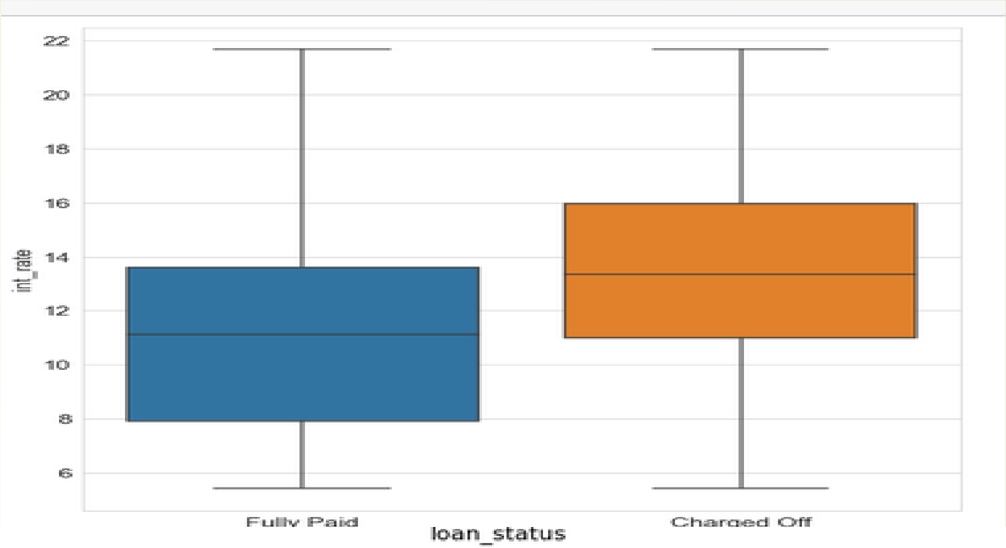


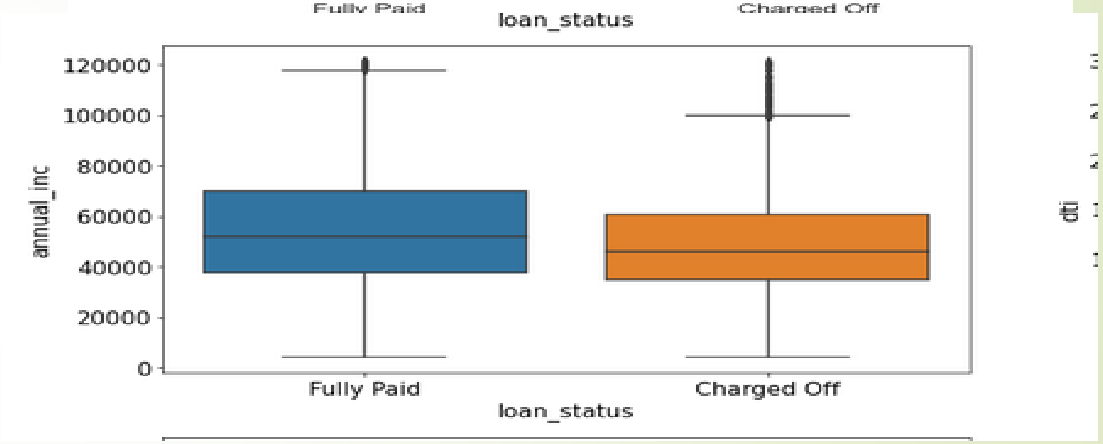


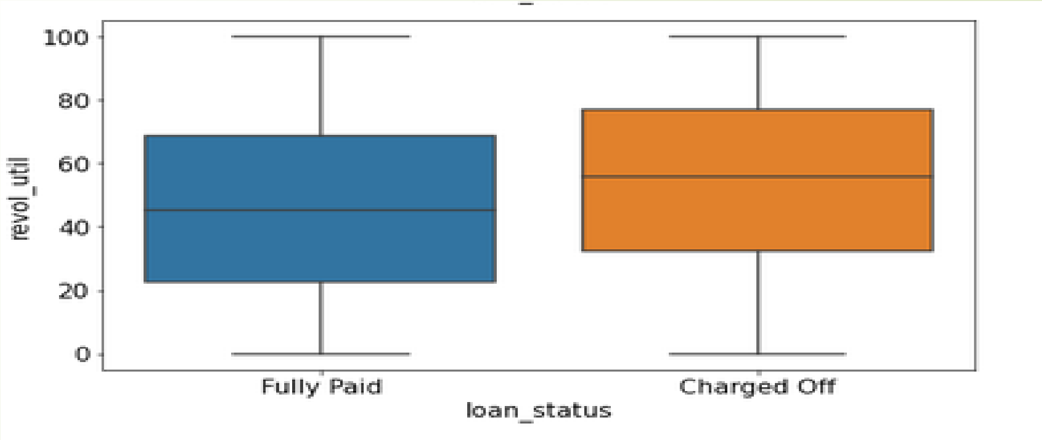


Segmented Univariate Analysis (Continuous Variable)

- higher interest rate tends to higher chances of default. there is huge difference in between charged off vs fully paid category
- higher annual_inc tends to lower chances of default. there is huge difference in 75percentile for both category
- higher revol_util rate tends to higher chances of default. there is huge difference in for both category
- higher dti value tends to higher chances of default. there is difference in 75percentile for both category



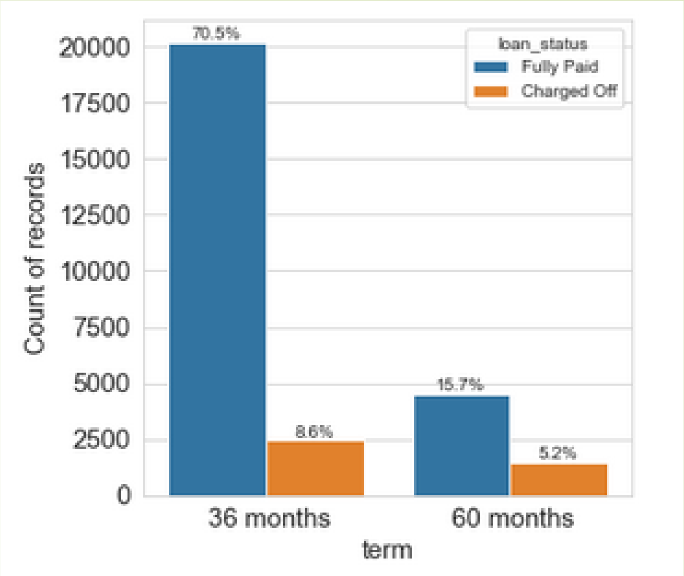


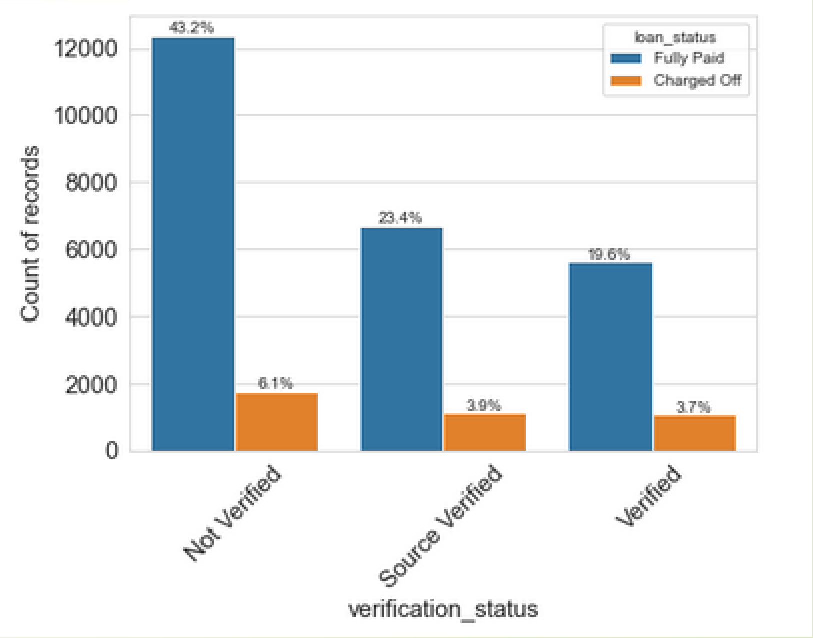


Segmented Univariate Analysis (Categorical)

- For loans with 5 year repayment term have high charged off ration compared to loan with 3year payment term (33% vs12%). Therefore, loan repayment term plays a important factor in predicting the default rate. Customer prefer to take loans for shorter duration of time

- The ratios of charged off per category are Not verified - 14% Source Verified - 16% and Verified - 19% Surprisingly the trend is unexpected verified applicants have more defaulted % which means the verification process has some issue.

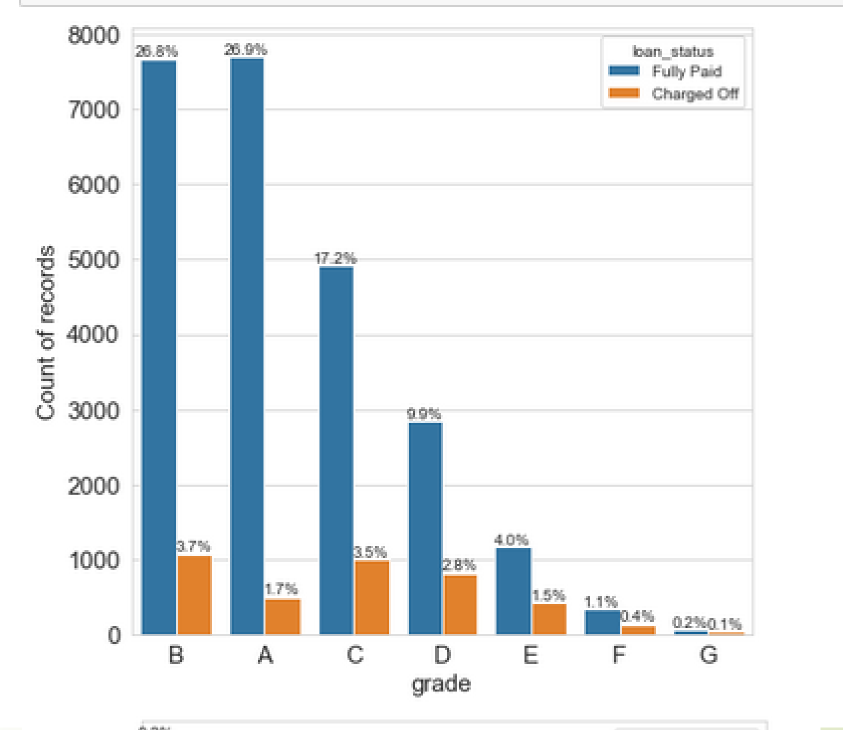


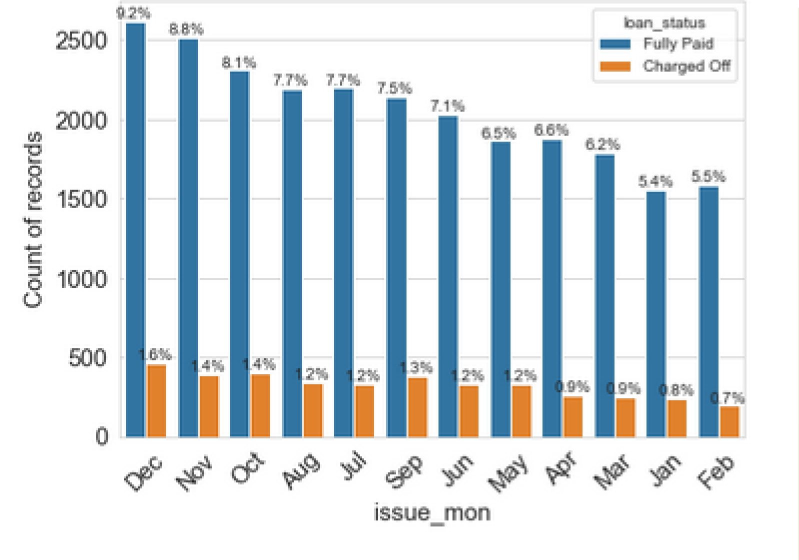


Segmented Univariate Analysis (Categorical)

- The ratios of charged off to fully charged are B-13% A-6% C-20% D-28% E-37% F- 37% , G-50% 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. In graph we can see the lower grade have high proportion of charged off account

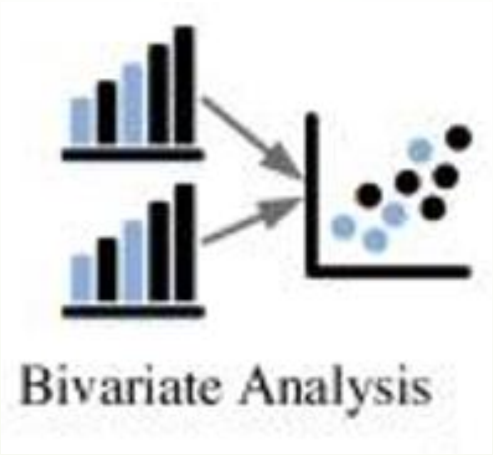
- Charge of % ratio was higher in month of Dec and May around 18% where as for other months its in range of 15-12%. Dec also sees high no of loan application may be because of holiday season people apply for loan for holiday and vacation. Also previously we have observed that vaacation purpose had high charge off value which supports our theory.





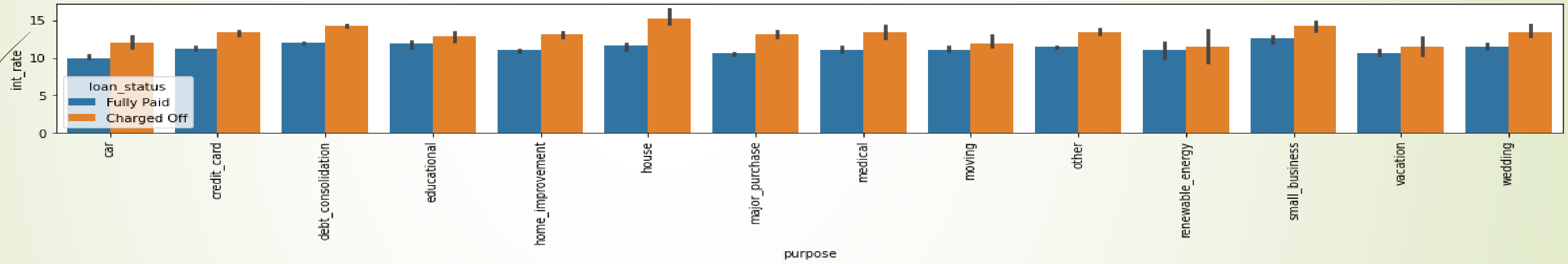
Bivariate Analysis

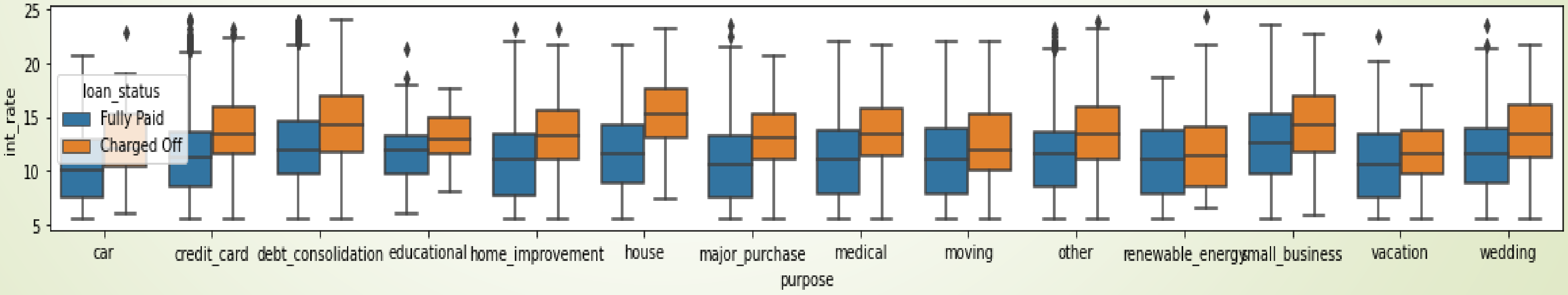
- Analyzing two variables behavior with respect to a third variable.
- Example: Analyzing Grade and Loan Amount with respect to Loan Status



Purpose vs Interest Rate Analysis

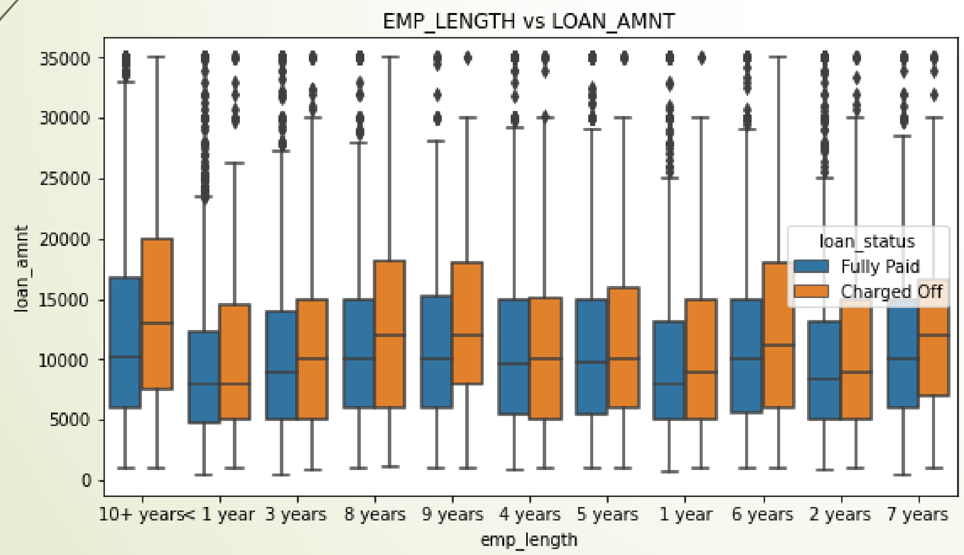
- From the plots it is clear that people taking loans for houses are the highest defaulters if interest rate is high.

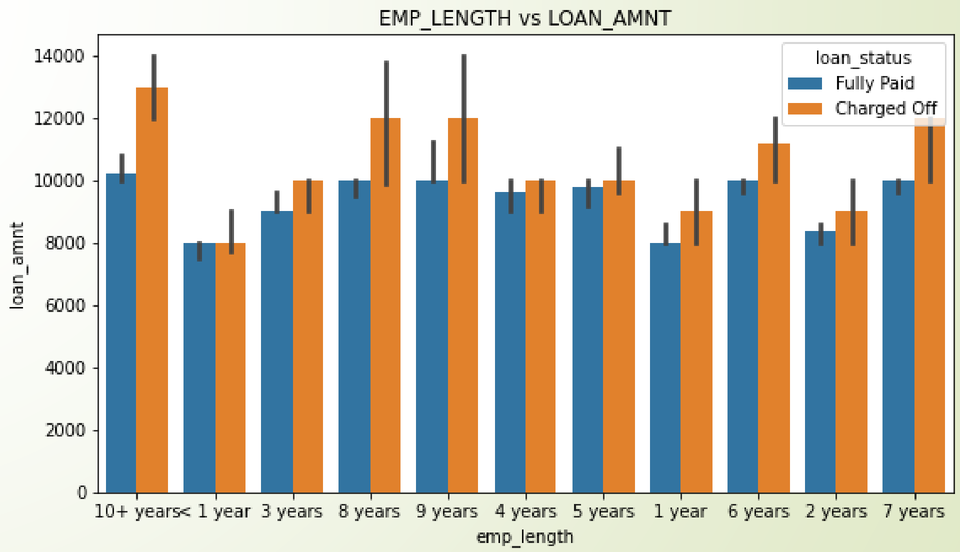




Employment Length vs Loan Amount Analysis

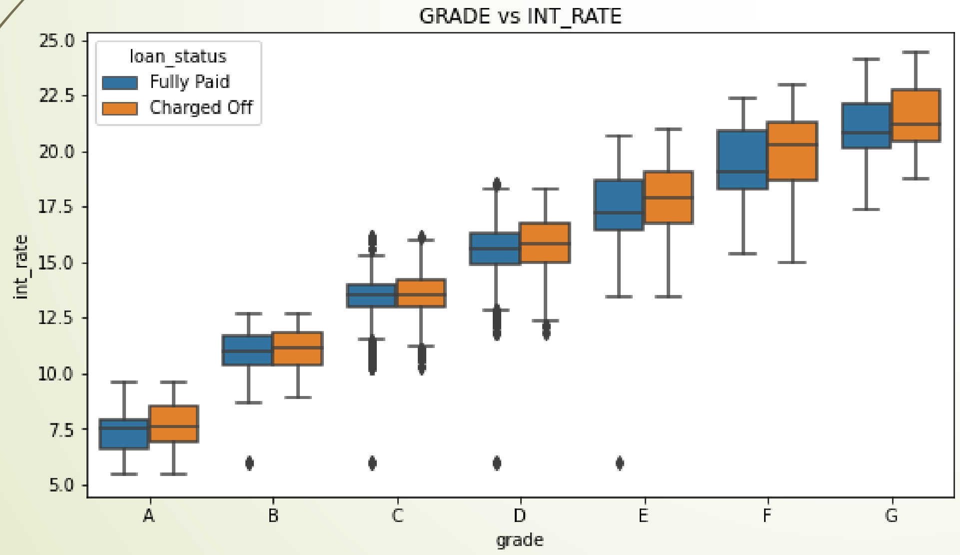
- From the plots it is clear that borrowers with higher employment lengths and taking more loan amounts are more prone to default the loan.

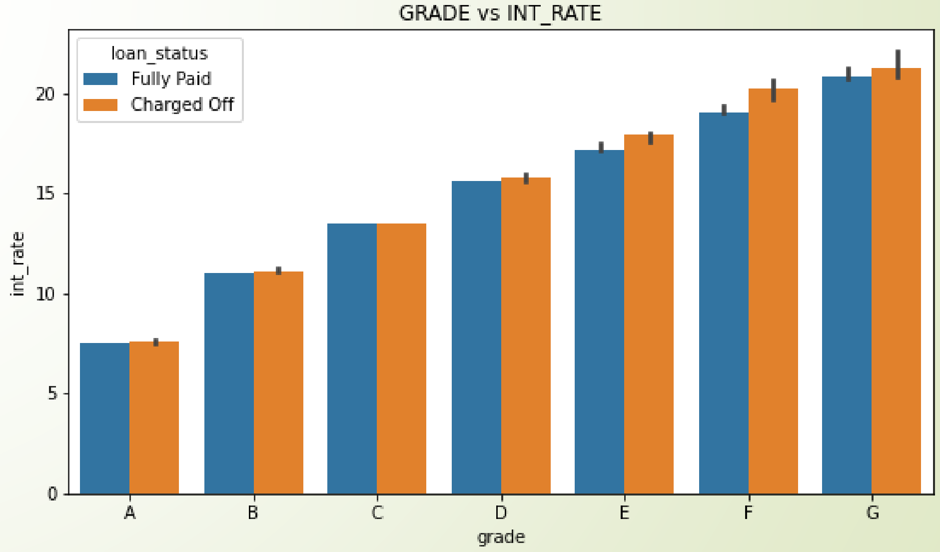




Grade vs Interest Rate Analysis

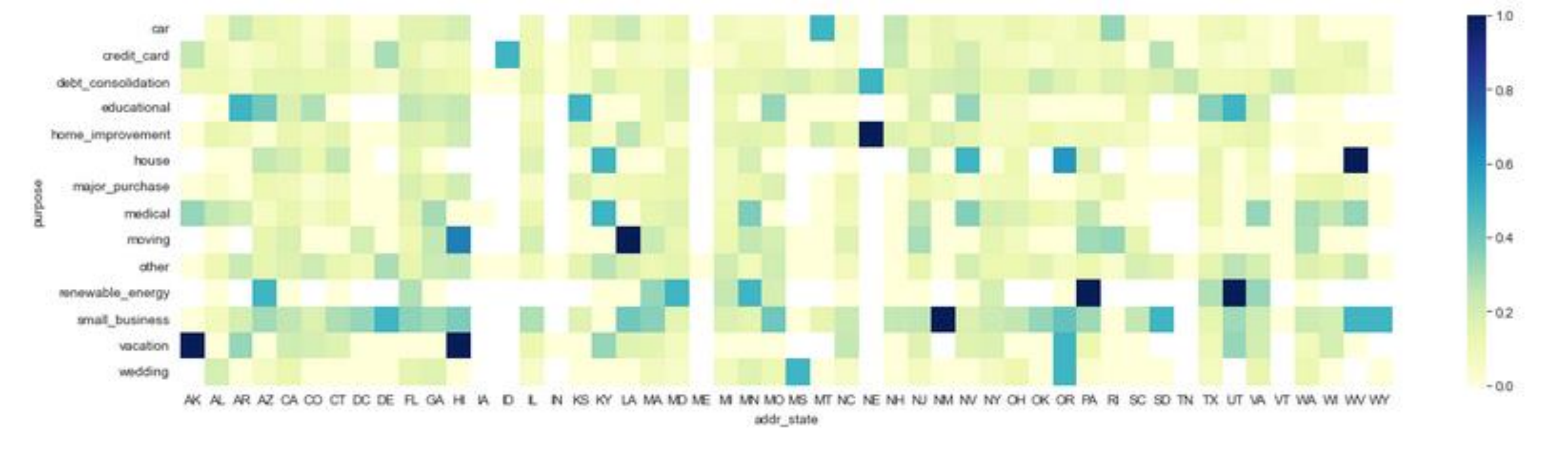
- From the plots it is clear as the grade and interest rate are inversely proportional, that is, if grade decreases, interest rate increases. That increase the rate of default the loan.





Purpose vs State

- Observation: vacation loans in AK, HI
- education loans in AR, KS, UT
- small business loans in DE, NM, WV, wY
- wedding loans in MS, OR all these combination of loans are risky



Recommendations

- Higher interest rate loans are more prone to be charged off
- Higher revolving line utilization rate loans are more prone to be charged off.
- Loan taken with Repayment term (5 years) are more prone to be charged off.
- Loan give to grade (D to G) are more prone to be charged off
- Loan purpose (small business, renewable energy, educational)
- Loan given to customer with Derogatory public records (1 or 2) are more prone to be charged off
- Lending company should not provide loans to the borrowers taking loans for house with high interest rate, they are more prone to default the loan.
- Lending company should not provide loan to the borrower who has a 10+ years of employment length and demanding a huge amount of loan as they are more prone to default the loan.
- Lending company should not provide loans to a higher grade borrowers as for higher grade borrowers interest rate increase which increase the chance to default the loan.