



LENDING CLUB CASE STUDY

CONTENT

- PROBLEM STATEMENT
- ANALYSIS APPROACH
- ANALYSIS RESULTS
- KEY TAKE AWAYS

Abhay Pratap Singh(Group Facilitator)

Mansi Upadhyay

PROBLEM STATEMENT

1. Lending loans to '**risky**' applicants is the largest source of financial loss (called credit loss). To reduce this risk bank check the portfolio of an applicant.
2. In this case study we will use a landing club dataset to find out those information(**consumer attributes** and **loan attributes**) which helps the bank to know which variables which will work as key attributes(**Drivers**) in doing the risk assessment of the candidates applying for a loan.

ANALYSIS APPROACH

We have followed the following steps:

Step1: Data Loading.

Step2: Data Cleaning: Removing missing values, Dropping irrelevant variables, Data standardization, Removing Outlier.

Step3: Data Manipulation: Changing categorical or numeric variables, Driving new matrixes.

Step4: Data Analysis: Univariate Analysis, Univariate Segmented Analysis, Bivariate Analysis and Correlation.

ANALYSIS EXPLANATION & RESULTS

1. Univariate Analysis and Segmented Univariate Analysis.

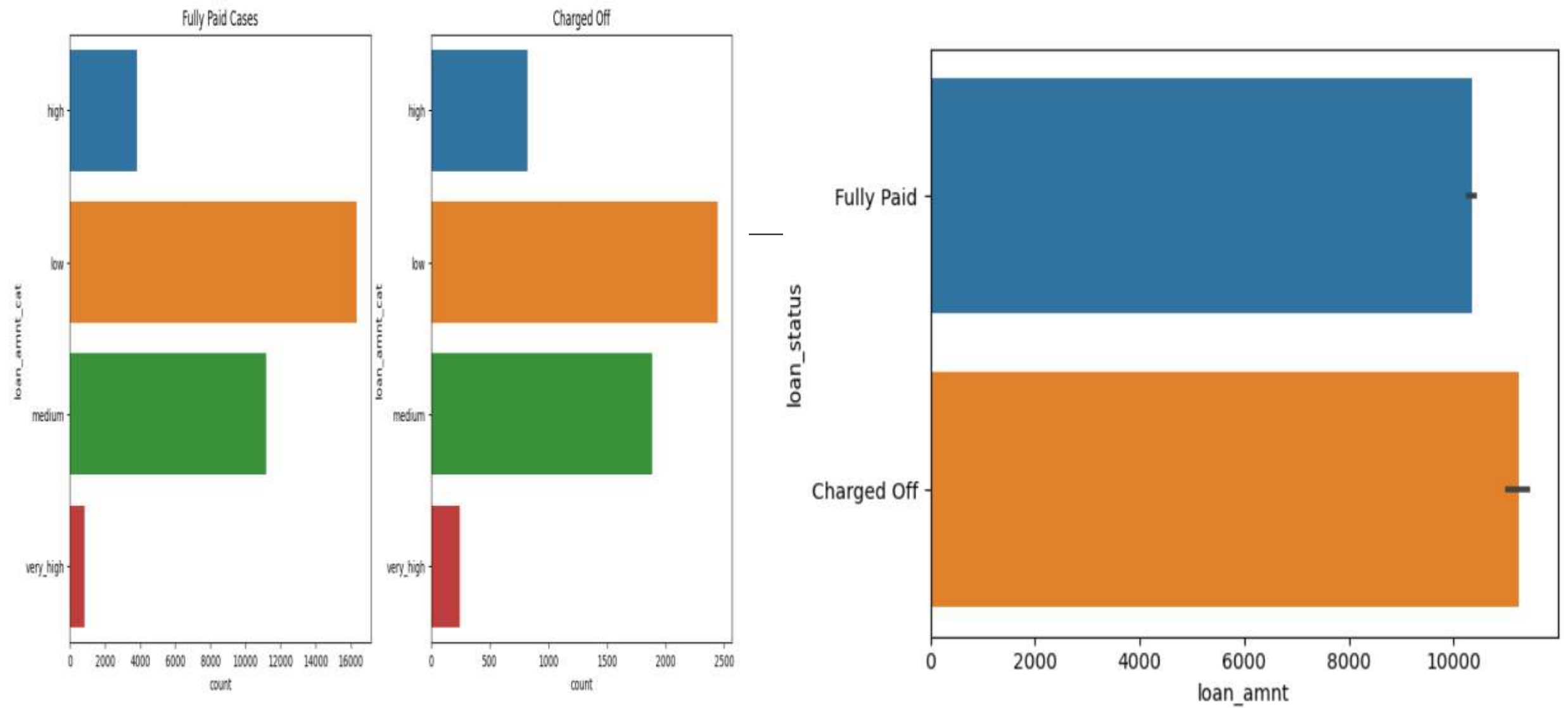
This is done to check the impact of single variable. Similarly segmented univariate is used to segment one single variables and check the trend.

2. Bivariate Analysis

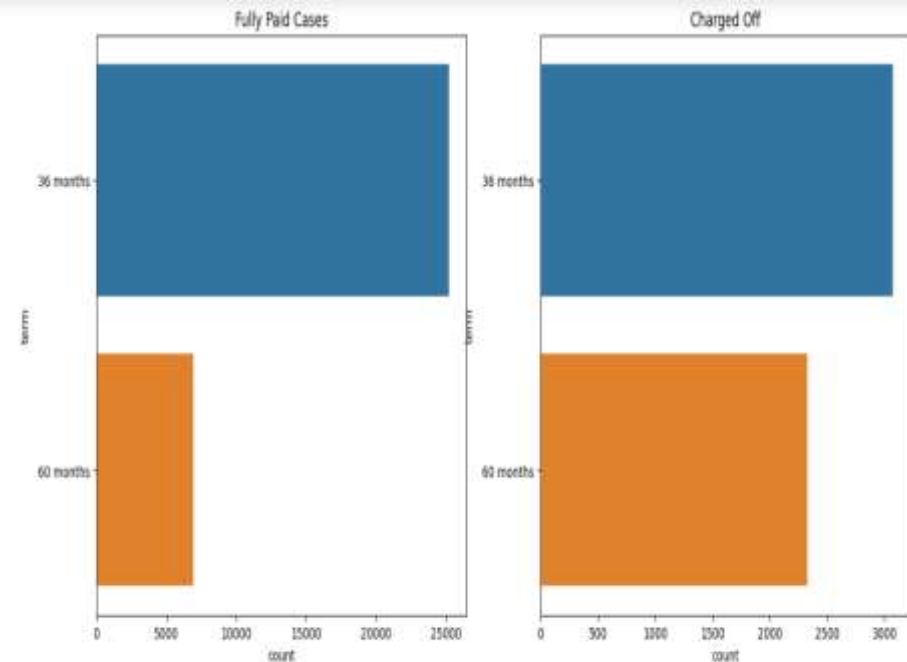
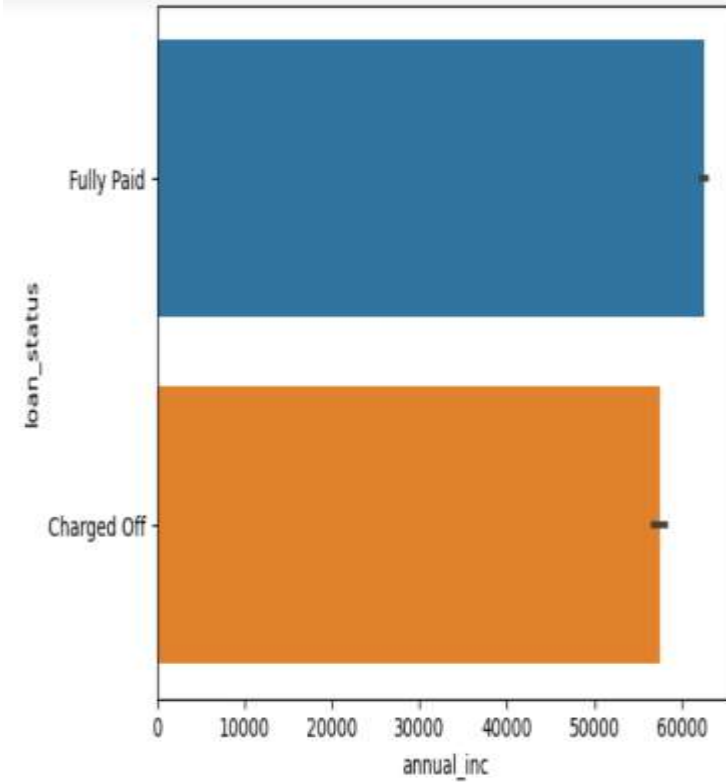
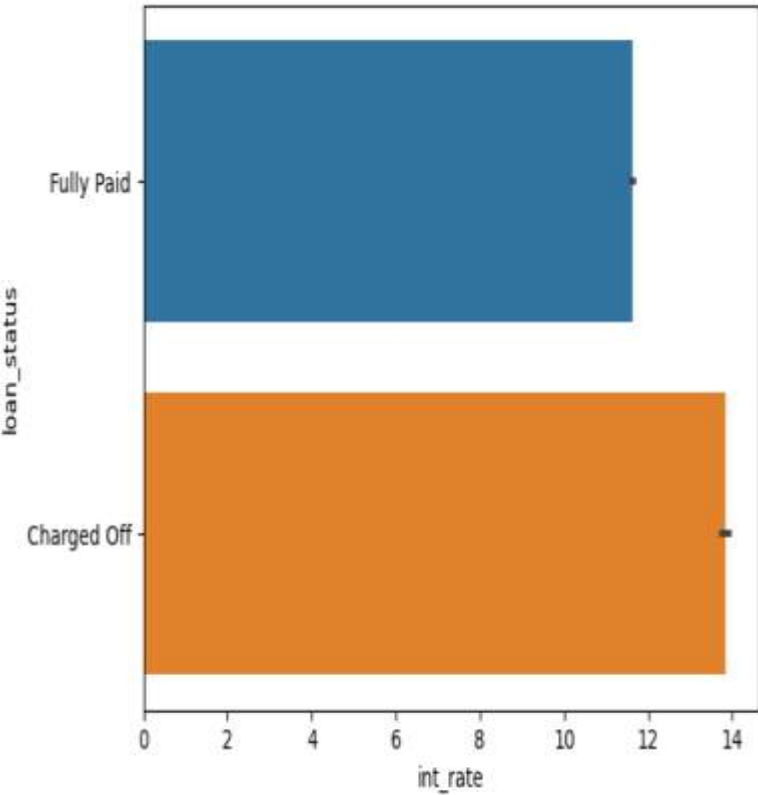
This is done to check the pairs of 2 variables and there impact.

3. Multivariate Analysis(Correlation)

This is done to check the correlation between numeric data in the dataset. To remove variables that are highly correlated and amplifying the outcome of analysis.

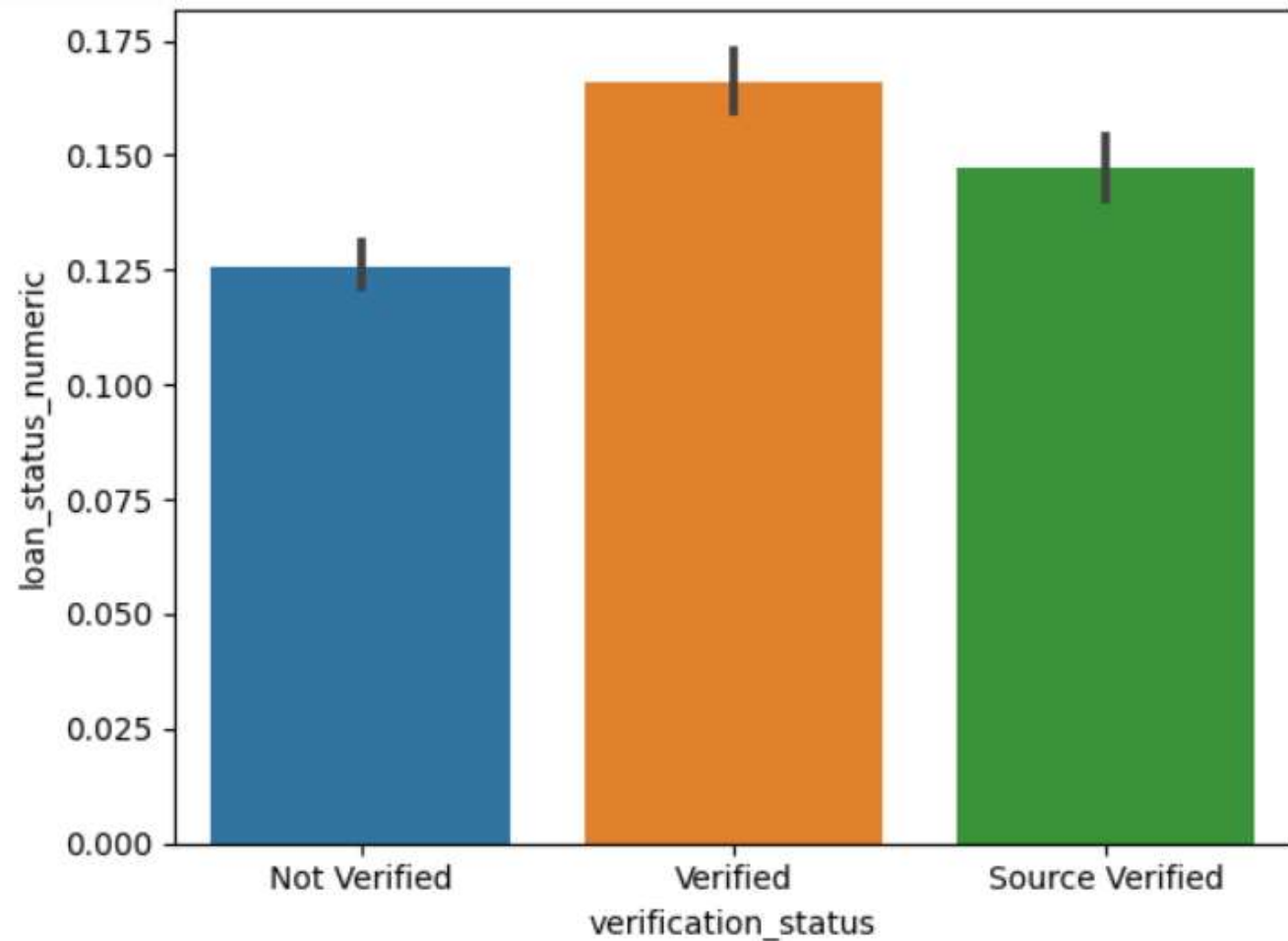
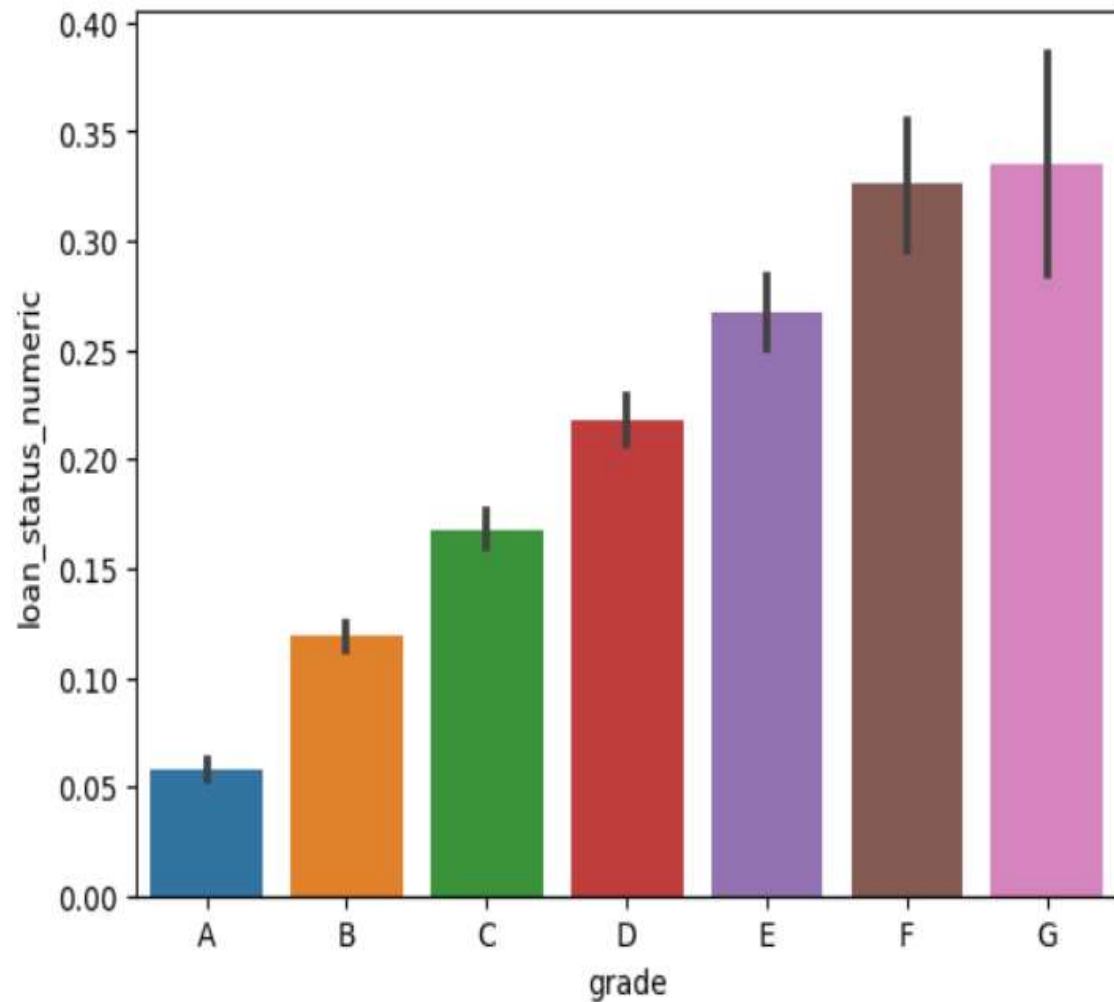


•The higher the loan amount the riskier it will become towards default status.

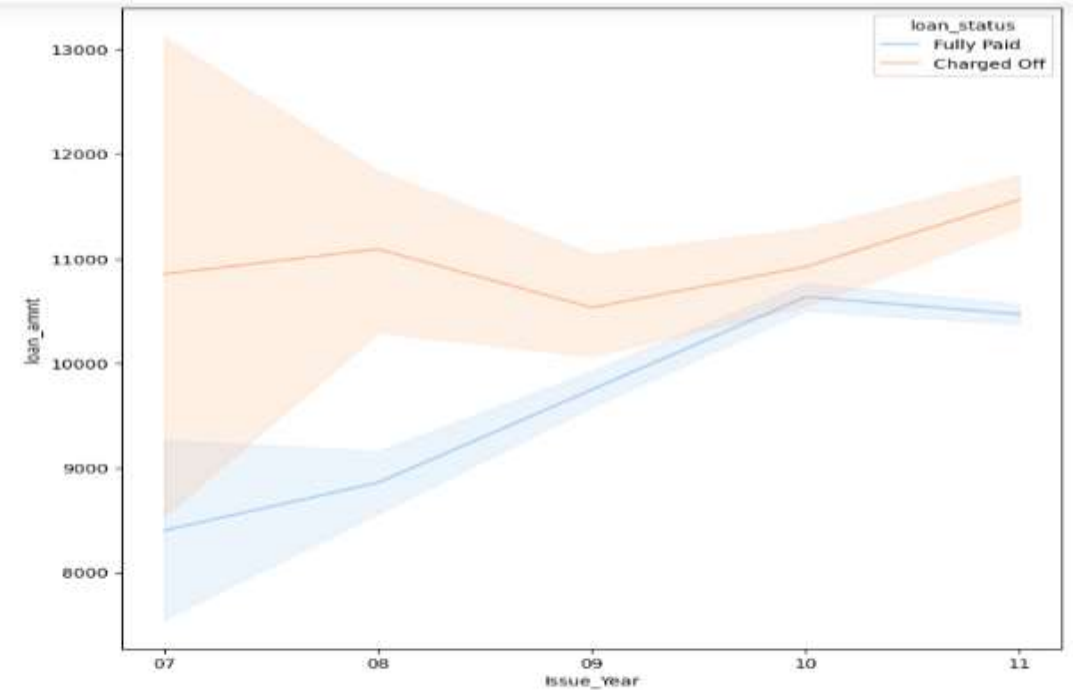
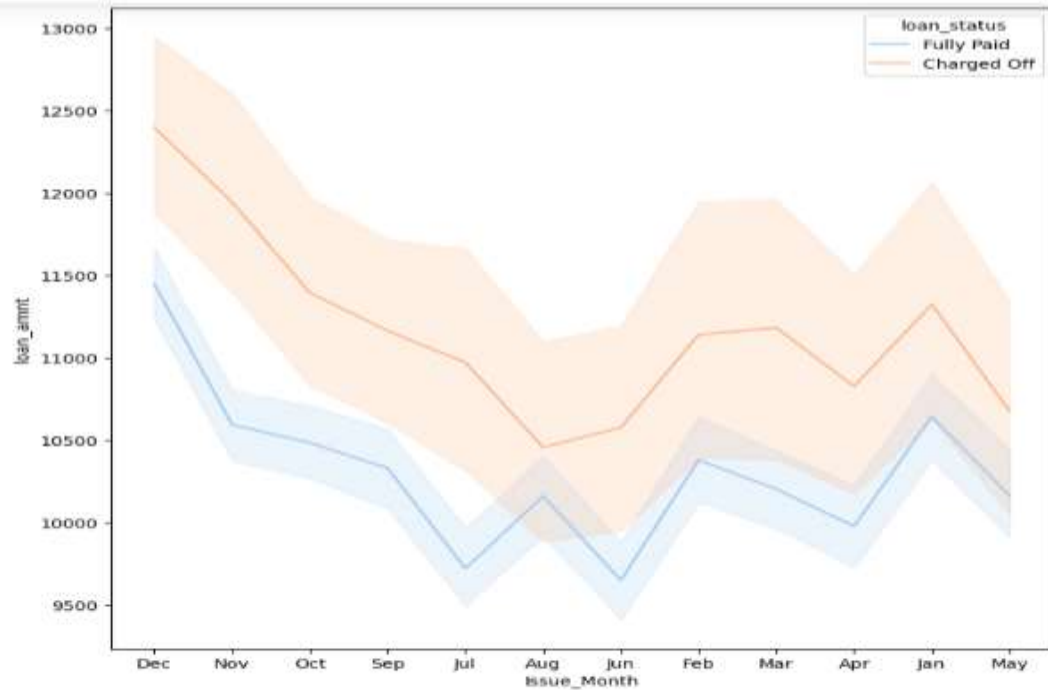


Conclusion for 1.3 : Long term(60months) seems to be more prevalent in Charged off or Defaulter category

1. High interest rate seems to be risk factor as most defaulter has high interest rate trend as compare to fully paid cases.
2. Annual income is higher in Fully Paid as compare to charged Off.
3. Longer term loan seems to get more defaulter than short term.



- Verification is directly affecting the loan status outcome. i.e the more verified a person details is the lesser the chances of him to become a defaulter.
- Values of loan status will vary from paid to defaulter if laon grades goes from A to G. Hence higher the grade the safer the outcome.



Conclusion for 1.6 : Effect of Issue_Month is almost same for type of loan status. In from 2010-2011 we see a drastic number of increase in defaulter v/s paid

Higher loan applications are seen in December month for every year in the dataset. Hence it can be used as a opportunity by bank to grow their customer by giving schemes or doing marketing campaigns.

Note: Correlation matrix created for numeric features

	loan_amnt	funded_amnt	int_rate	installment	annual_inc	dti	inq_last_6mths
loan_amnt	1.000000	0.980762	0.248718	0.923744	0.368576	0.069361	-0.004171
funded_amnt	0.980762	1.000000	0.264847	0.952578	0.368153	0.068027	-0.003867
int_rate	0.248718	0.264847	1.000000	0.230569	0.045820	0.111466	0.156679
installment	0.923744	0.952578	0.230569	1.000000	0.356785	0.055091	-0.004140
annual_inc	0.368576	0.368153	0.045820	0.356785	1.000000	-0.079448	0.016279
dti	0.069361	0.068027	0.111466	0.055091	-0.079448	1.000000	0.016399
inq_last_6mths	-0.004171	-0.003867	0.156679	-0.004140	0.016279	0.016399	1.000000
open_acc	0.180251	0.176676	-0.041129	0.167215	0.266597	0.294638	0.080729

loan_amnt is highly correlated with "installment","funded_amnt". Hence only consider of loan_amnt is enough which will not over amplify the effect of same kind of variables.

KEY TAKE AWAYS

A. Drivers for portfolio assessment/risk assessment.

1. Loan Grade

Findings: People with good loan grades have more tendency to repay the loan.

2. Term

Findings: Long term(60months) seems to be more prevalent in Charged off or Defaulter category than paid ones. Hence if bank provide loans for shorter term than that would be more safe in terms of business perspective.

3. Loan Amount

Findings: Loan amount is more likely to be high or very high for default cases as compare with fully paid cases. Hence lower the amount, lower the risk.

4. Verification

Findings: Verification is directly affecting the loan status outcome. i.e the verified a person details has lesser chance of becoming a defaulter.

5. Interest Rate

Findings: High interest rate seems to be risk factor as most defaulter has high interest rate trend as compare to fully paid cases. Hence if someone is going for high interest other drivers like verification status, Loan Grade, loan amount, Loan term should be considered very carefully.

6. Annual Income

Finding: The higher the annual income the safer the candidate is.

KEY TAKE AWAYS

B. Driver for Business Expansion

1. Issue_Month

Findings: Higher loan applications are seen in December month for every year in the dataset. Hence it can be used as a opportunity by bank to grow their customer by giving schemes or doing marketing campaigns.

THANK YOU !
