

LENDING CLUB CASE STUDY

An analysis into customer and loan attributes influencing Loan Default

Group Members:

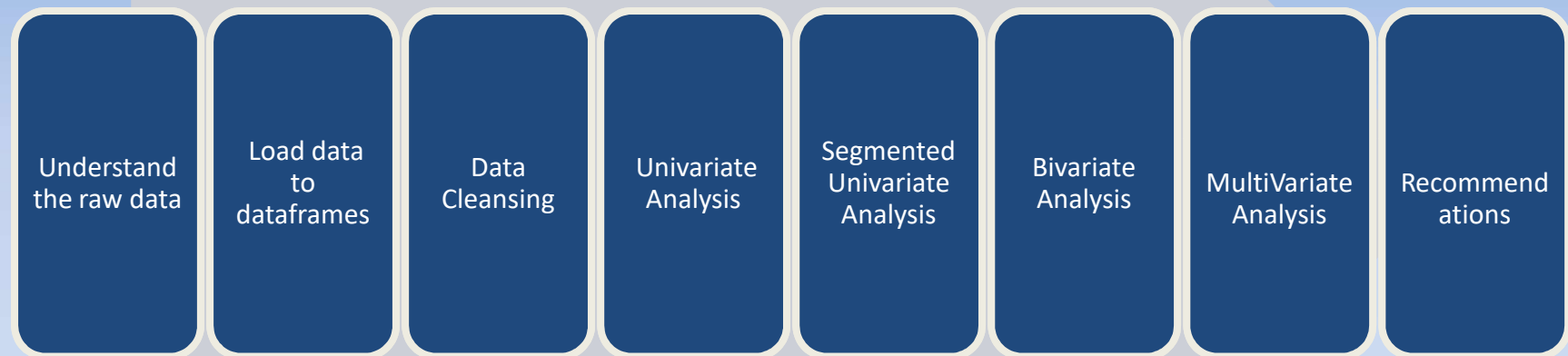
Ramanujam Srinivasan

Vysakh P S

PROBLEM STATEMENT

- Identify patterns which indicate the borrower is likely default.
- Provide the company with actionable insights that help them reduce the chance of a borrower defaulting.

Approach



Assumptions

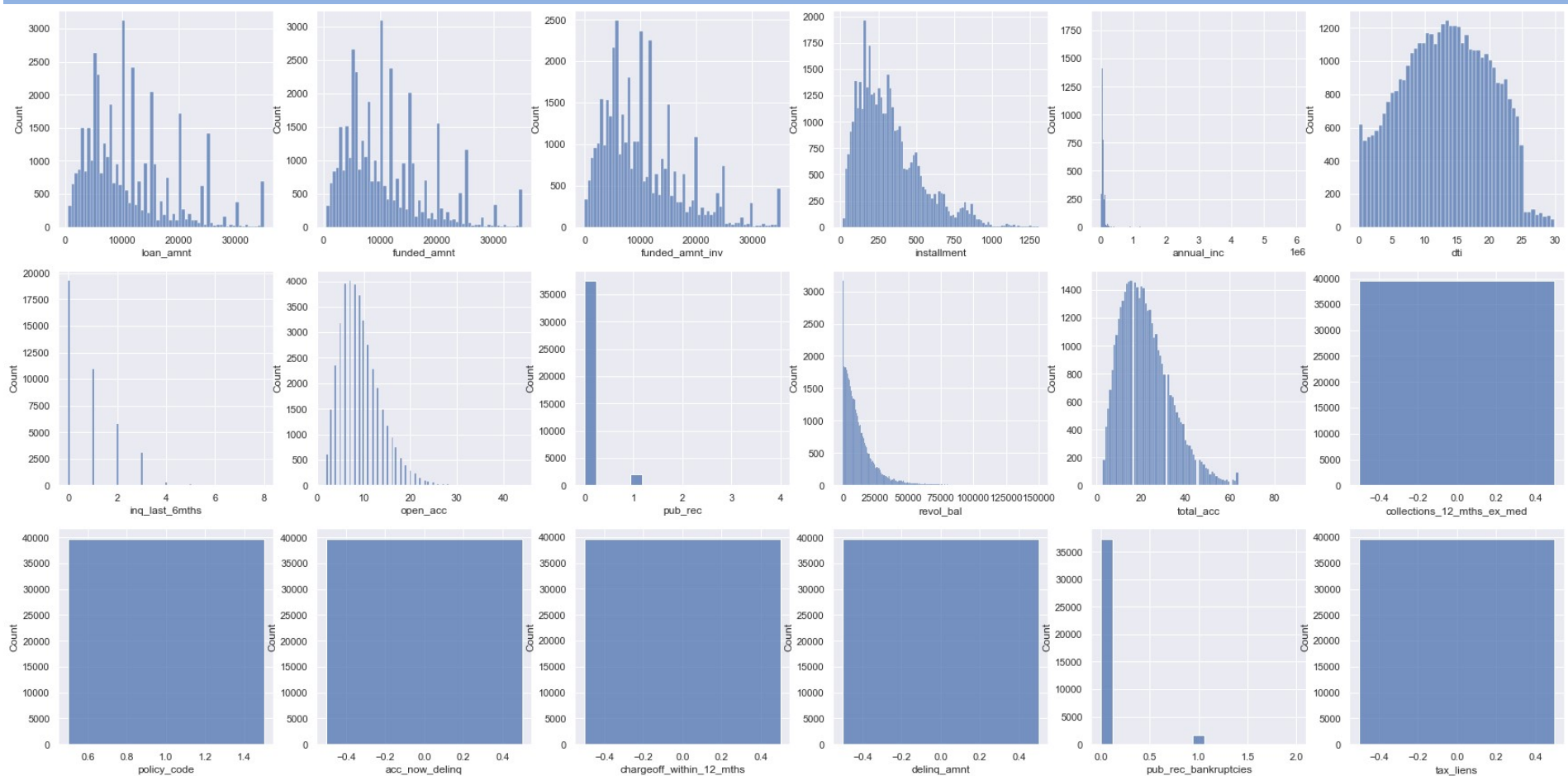
- Customer demographic and Behavior are not taken into consideration

Data Cleaning & Preparation

- Deleted all null value columns (Removes 54 columns)
- Removed columns with only one unique values (Removes 9 columns)
- Validated internal rules (open_acc should be \leq total_acc)
- Data Checks for Funded amt invested $>$ Funded Amt / LoanAmount was done to ensure validity of data
- Set appropriate Data Types
- Filtered Data (Removed entries with '**Current**' loan status).
- Deriving new Columns from existing columns (Closed Acct. Year etc)

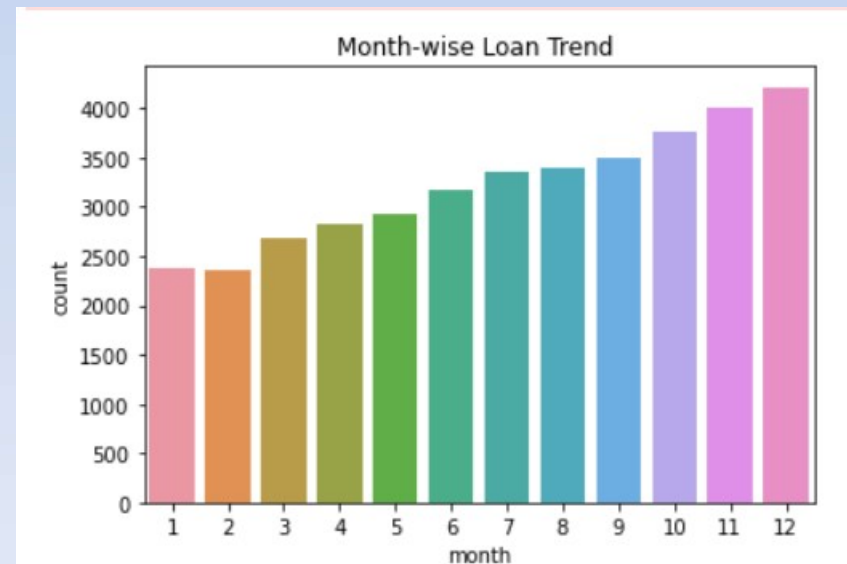
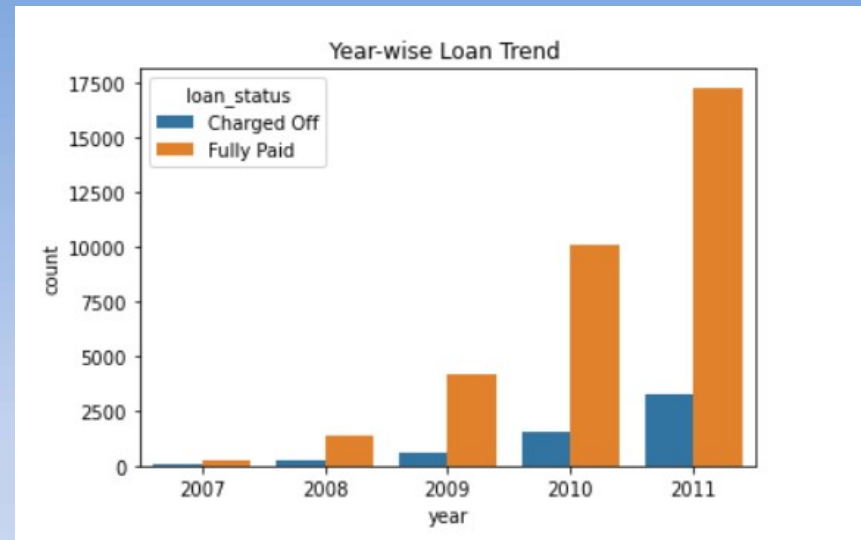
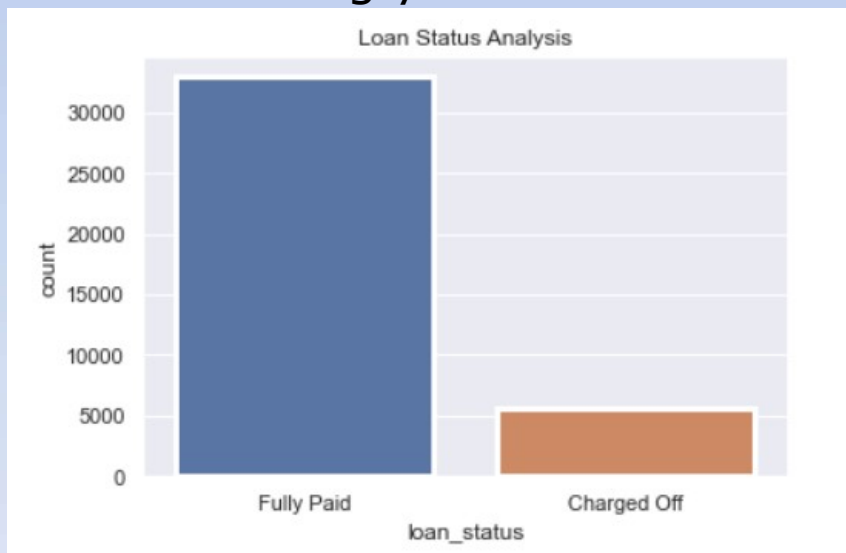
Univariant Analysis

Created histogram and box plot to understand the behavior of each variable.

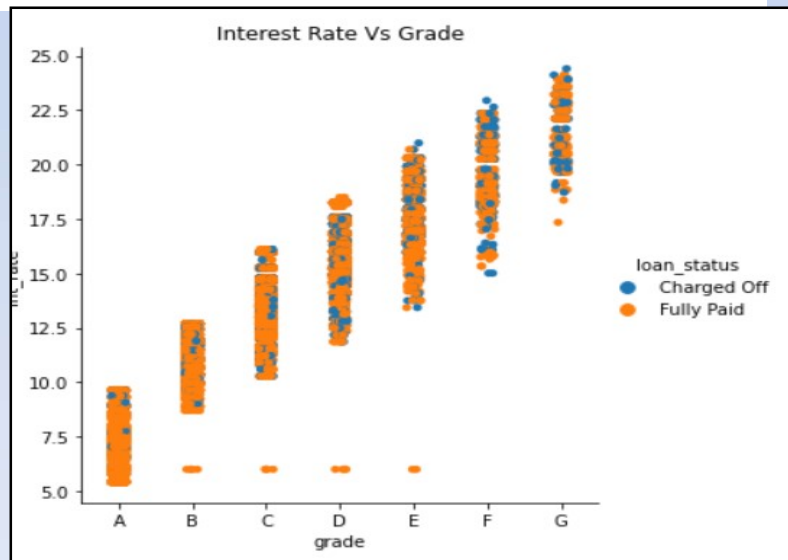
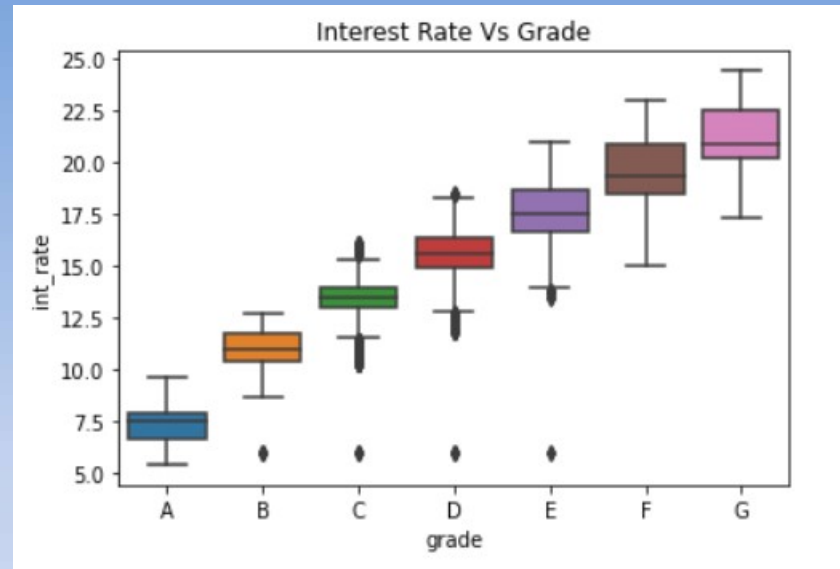
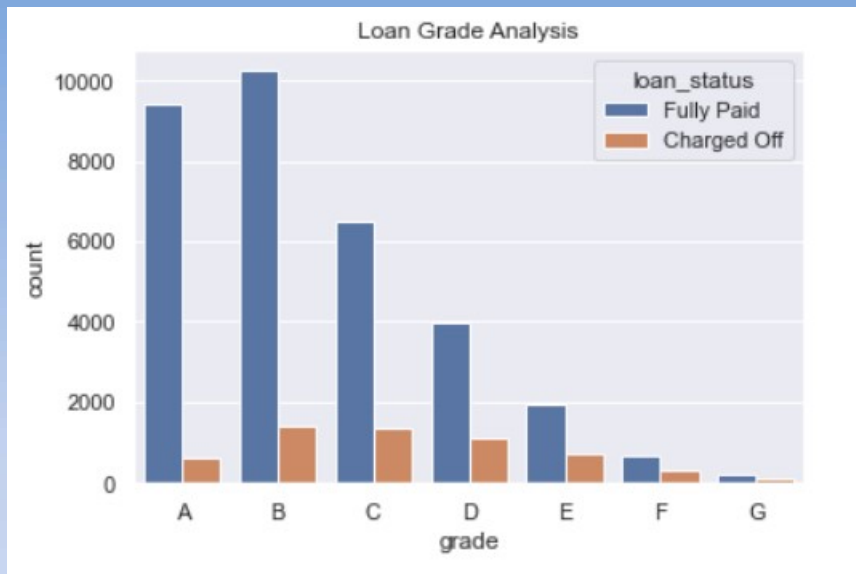


Loan Status Analysis

- Most of the applicants have Fully Paid the Loan
- Loan applicants increases steadily across year
- We see a trend of loans being taken during year end.

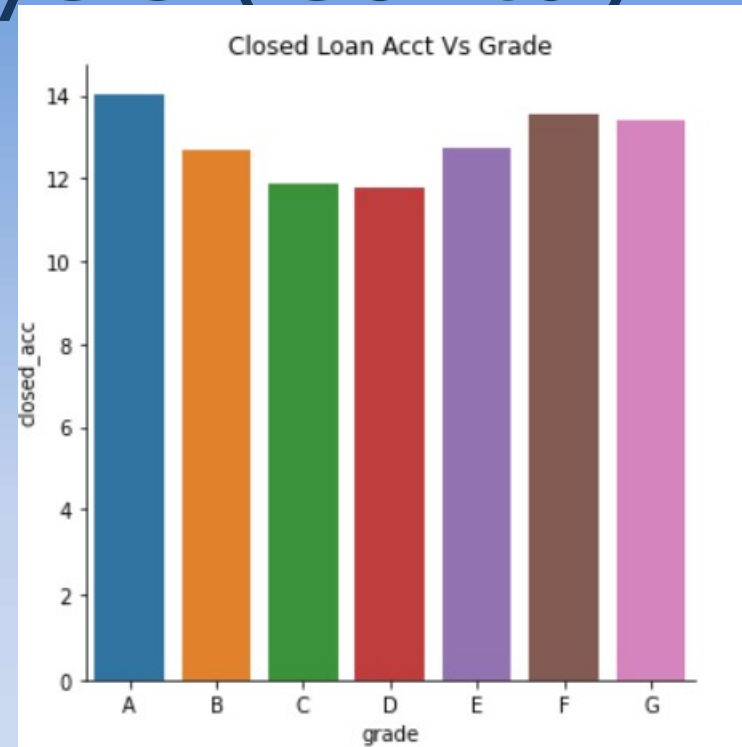
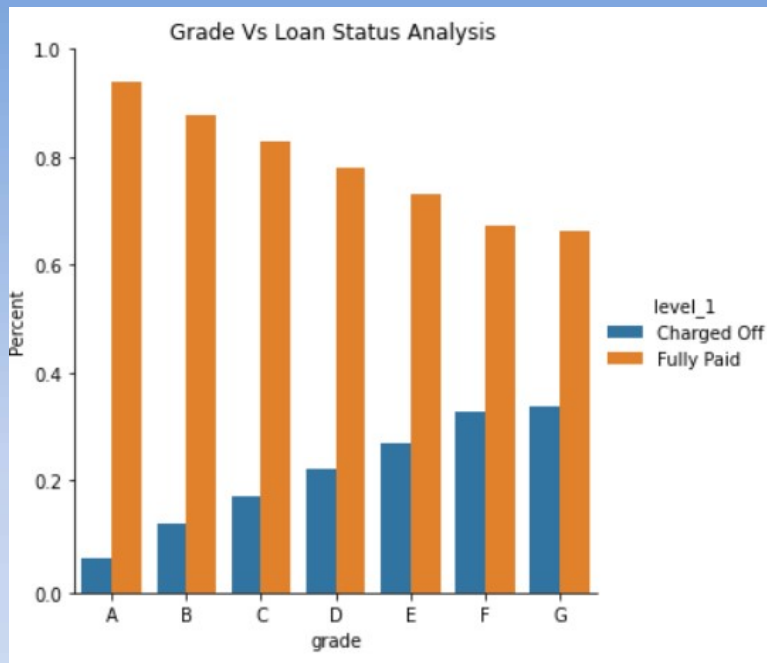


Loan Grade Analysis



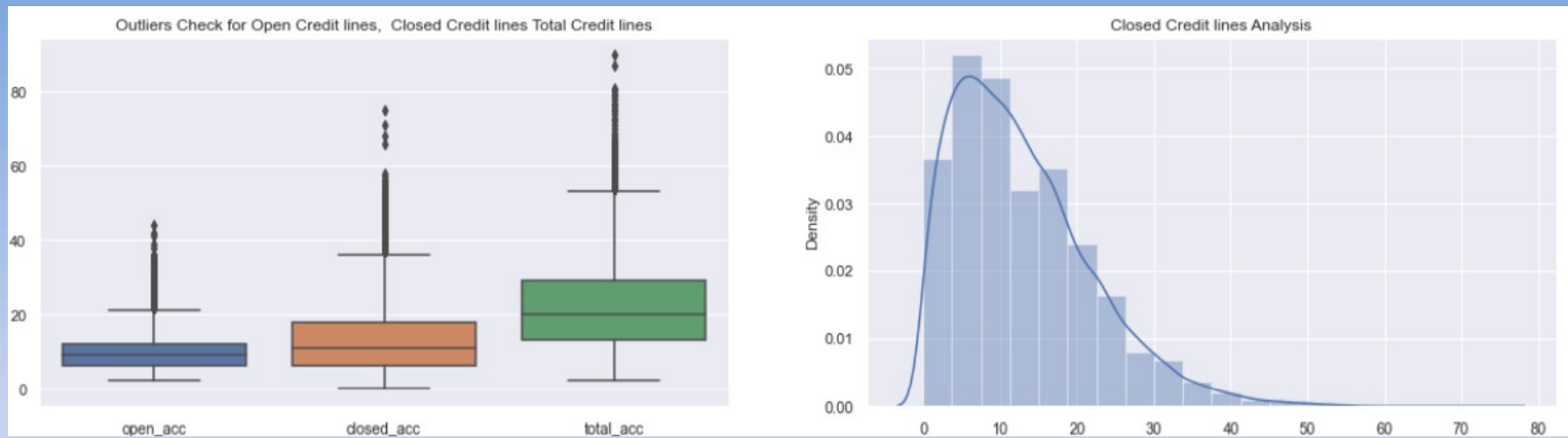
- ✓ Most of the loans are from grade A and B.
- ✓ Even though high proportion of Grade A loans, Defaulters are relatively less
- ✓ High Quality loans (A,B,C) have low interest rate
- ✓ Higher interest rate have higher tendency to default the loan.

Loan Grade Analysis (Contd)



- ✓ Defaulters increases from high grades to low grades, which indicates loans to Grade E, F, G needs to be approved with caution
- ✓ Vice Versa, Fully paid observation indicates loan applicants with higher grade (A,B) can be provided loans
- ✓ Derived Variable (Closed Acct) indicates Grade D & C are delinquent in closing the loans

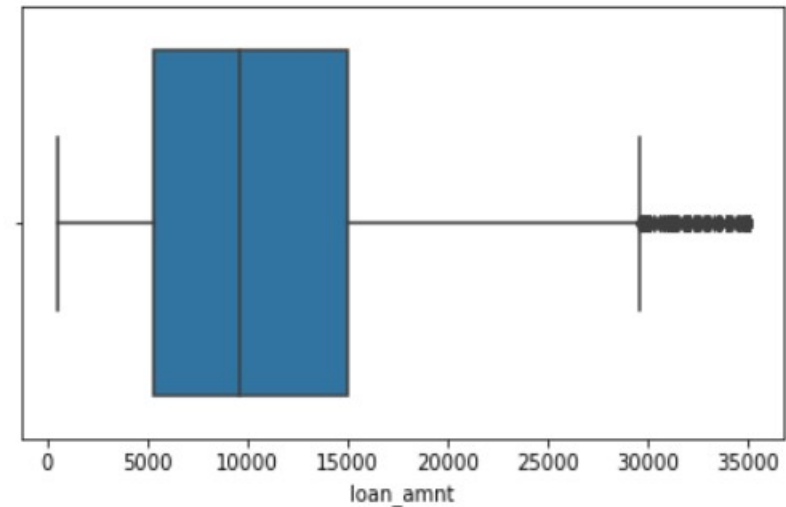
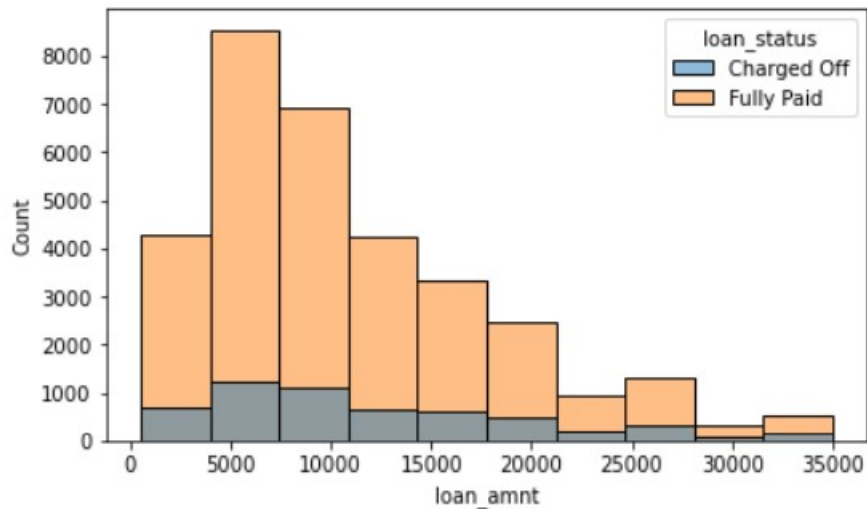
Closed Credit Line Analysis



- ✓ Closed Credit Line = Total Credit lines - Open Credit lines
- ✓ Box Plot indicates Closed Credit Line are higher, which is a good indicative that many loans are closed
- ✓ Closed Credit Line are left skewed.

Loan Amount Analysis

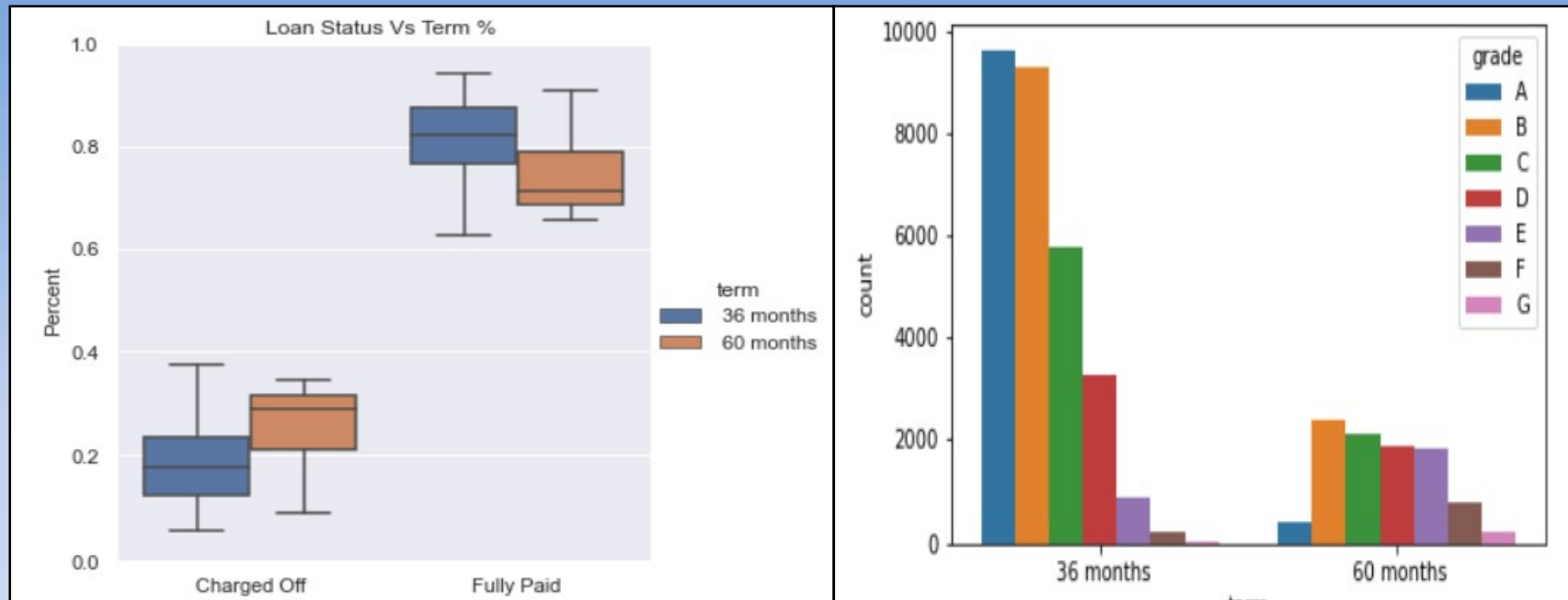
Loan Amount distribution



✓ Funded amount is left skewed. Most loan needs are in the range of 5000 – 15000

✓ Defaulters are more in range of 5000 – 15000

Loan Term Analysis



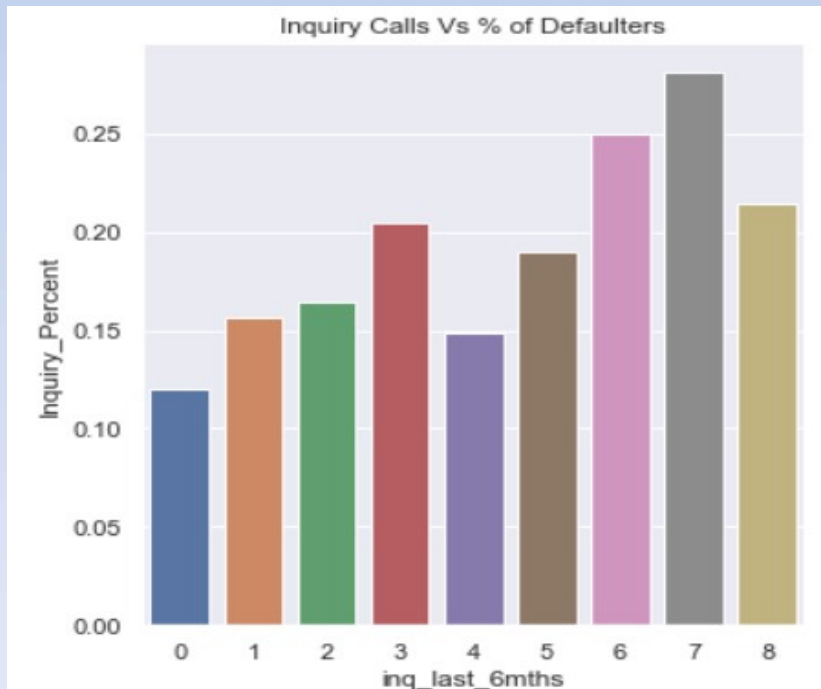
- ✓ Defaulters (Charged Off) are comparatively high in long term loans
- ✓ Short term loans are common in Less Risky Grades (A,B,C)

Other Variable Analysis



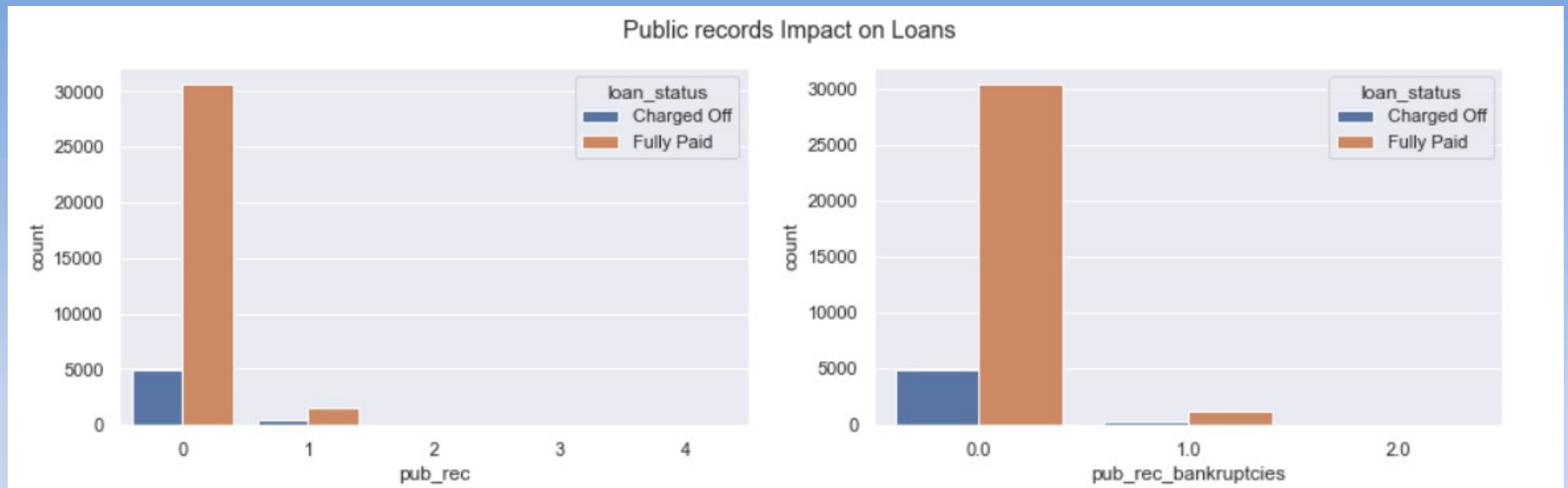
✓ As Employee Tenure increases, annual income increases

✓ Loans Payment are better when employee tenure is more



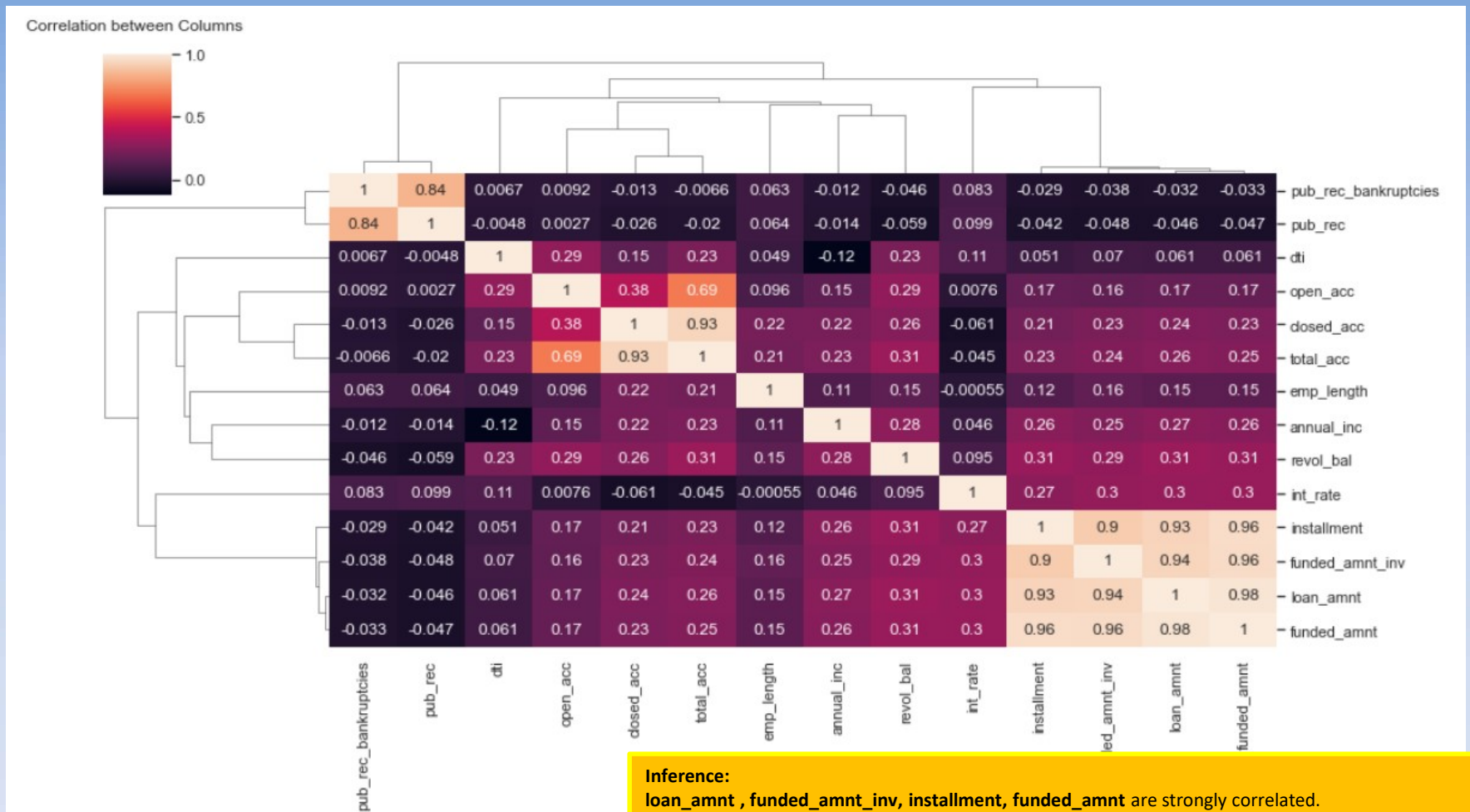
✓ Customer with more number of recent credit inquiries have higher chances being default

Other Variable Analysis



- ✓ Having No Public records is good sign for Loan
- ✓ Applicants with no public records have paid loans fully

Multivariate Analysis



Inference:

loan_amnt , funded_amnt_inv, installment, funded_amnt are strongly correlated.

•total_acct, closed_acct, open_acct are positively correlated.

•int_rate is negatively correlated with emp_length, Credit lines (total_acct,closed_acct, open_acct)

•annual_inc and dti is negatively correlated.

•pub_rec_bankruptcies & pub_rec are negatively correlated with loan_amnt, funded_amnt, funded_amnt_inv. annual_inc,total_acct

Conclusion

- ✓ The percentage of defaulting tends to increase from higher grade (A) to lower grades (G).
- ✓ Chances of defaults tend to increase with the number of recent credit inquiries by customer.
- ✓ Interest rate seems to be a contributing factor with higher interest resulting in higher chances of default.
- ✓ Clients with Low Income Group are more likely to default.
- ✓ Clients having long terms loans are risky
- ✓ Avoid Clients with Public Records
- ✓ Prefer Clients with employment history of 10+ years