



Credit Exploratory Data Analysis Case Study

Rudranil Ganguly

SaiKumar



Purpose

- Credit Risk Analysis helps a organisation make decisions for loan approval of the lending entity based on their profile.
- Credit Risk Analysis leads to reduction of business and financial loss by avoiding bad or high risk applicants for loans

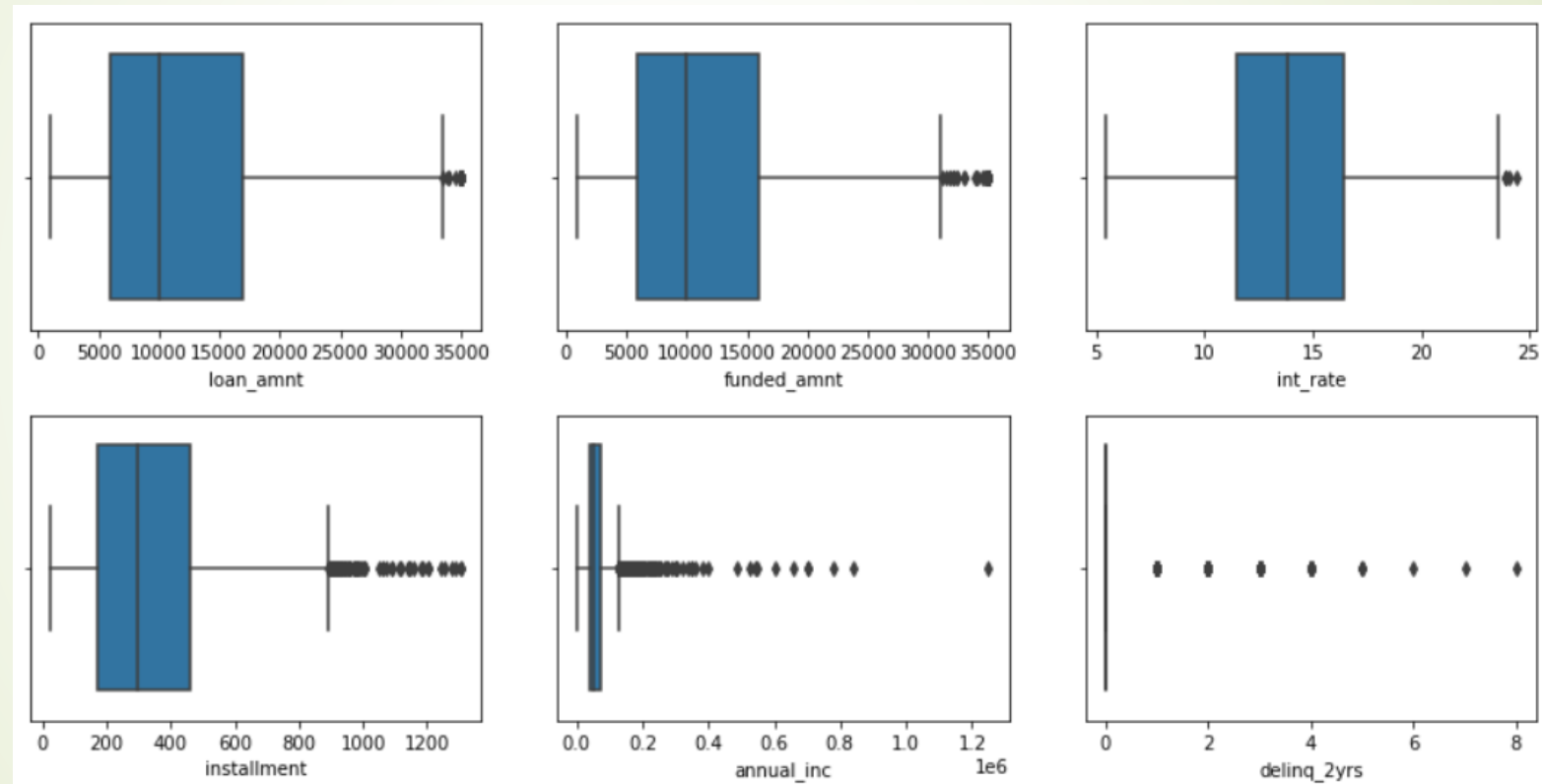


Steps

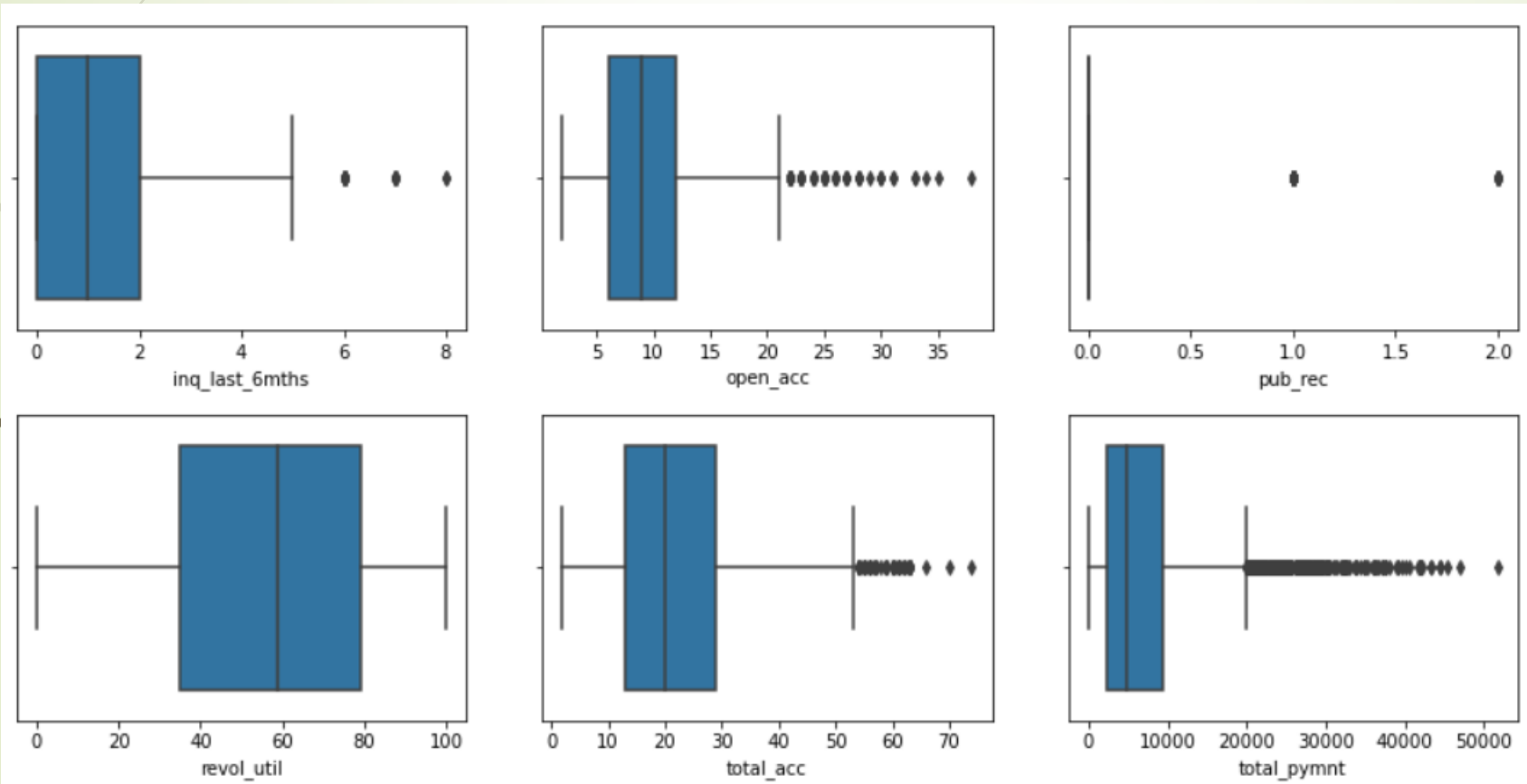


1. Data Understanding & Sourcing
2. Improve data quality by removing data quality issues
3. Introduce data transformations eg binning
4. Check for Data imbalances and perform Univariate and Bivariate analysis
5. Merge the data from current and past
6. Recommendation and Risk

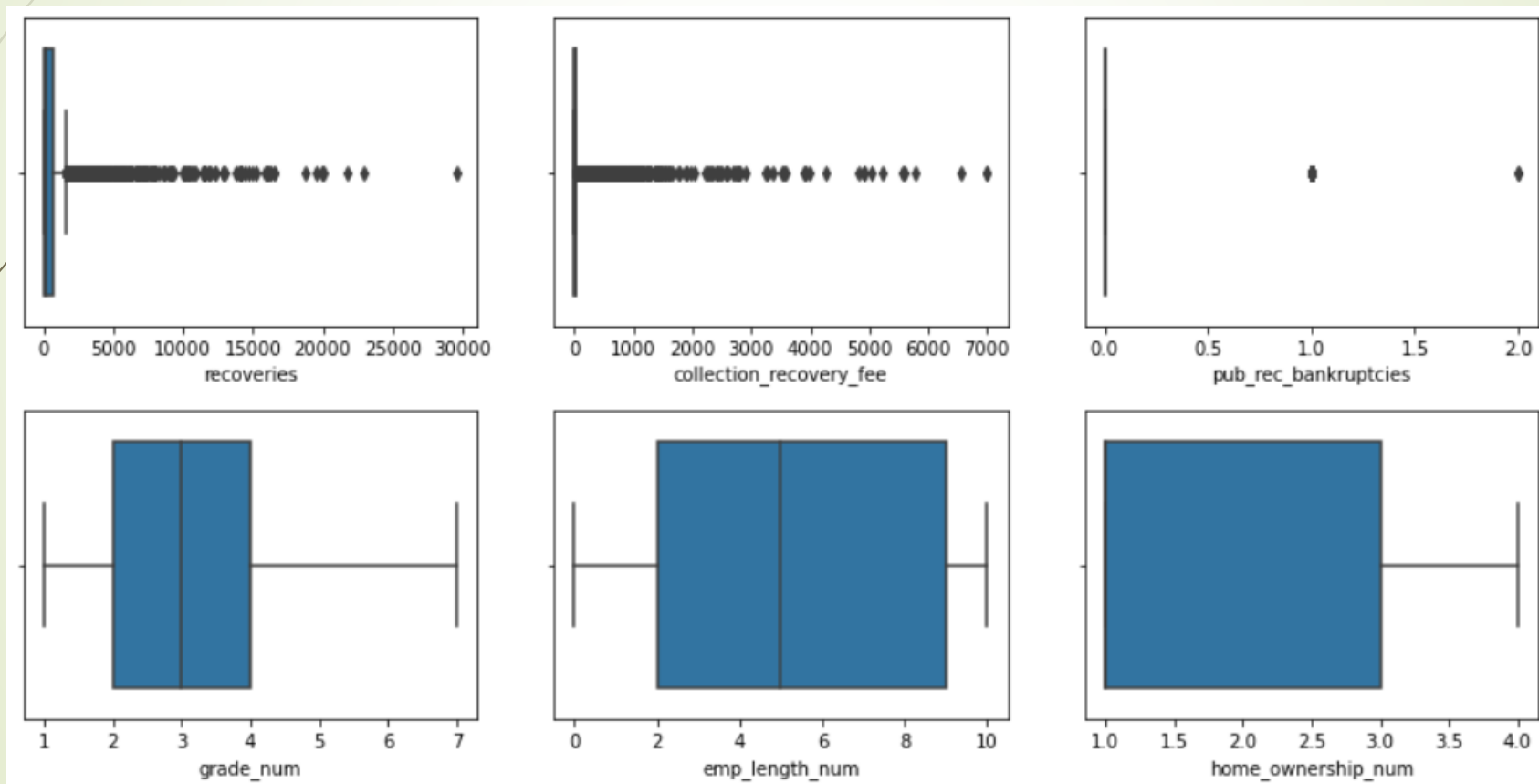
Univariate Analysis - Numerical Variables(Box Plot)



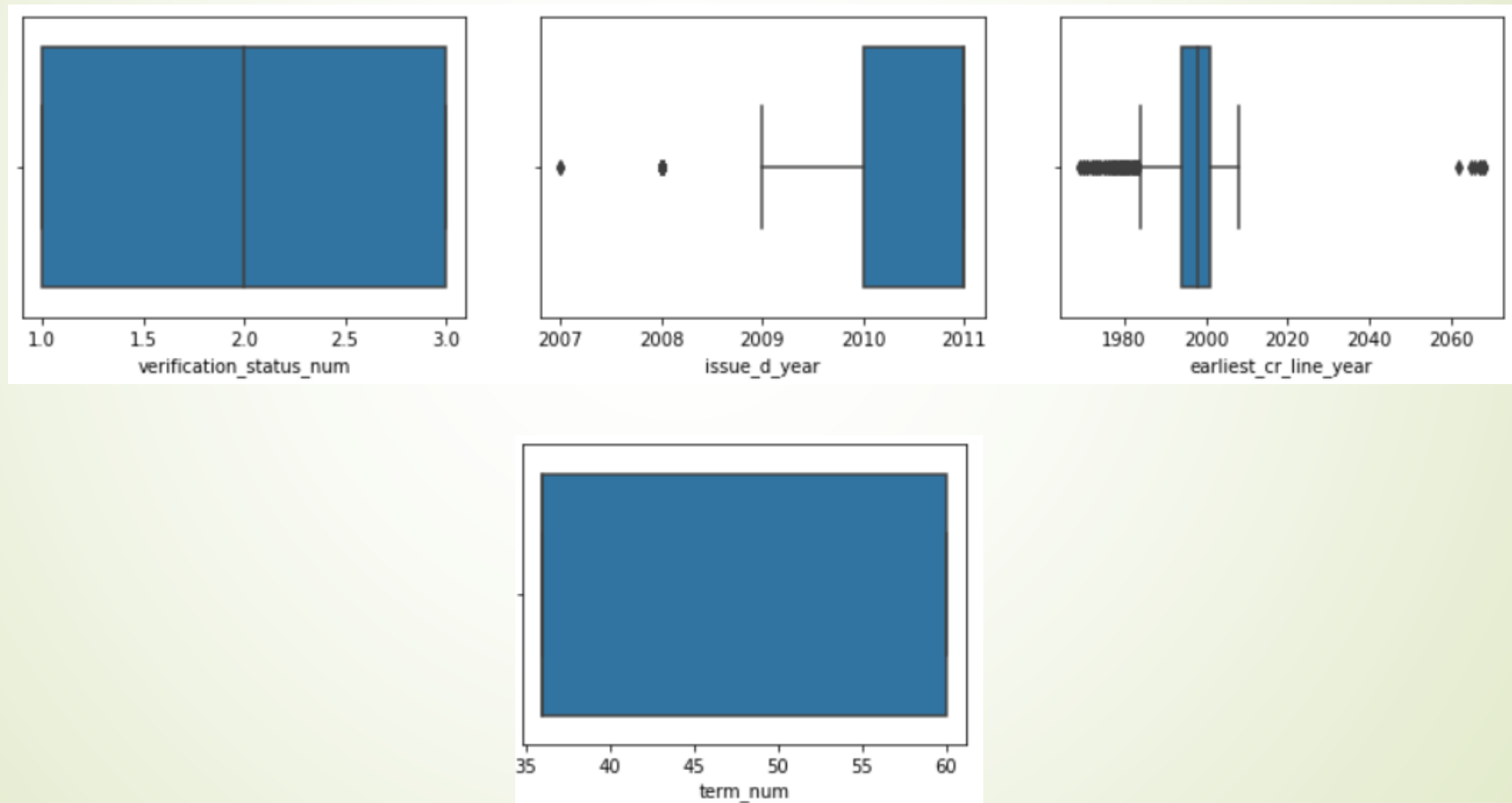
Cont..



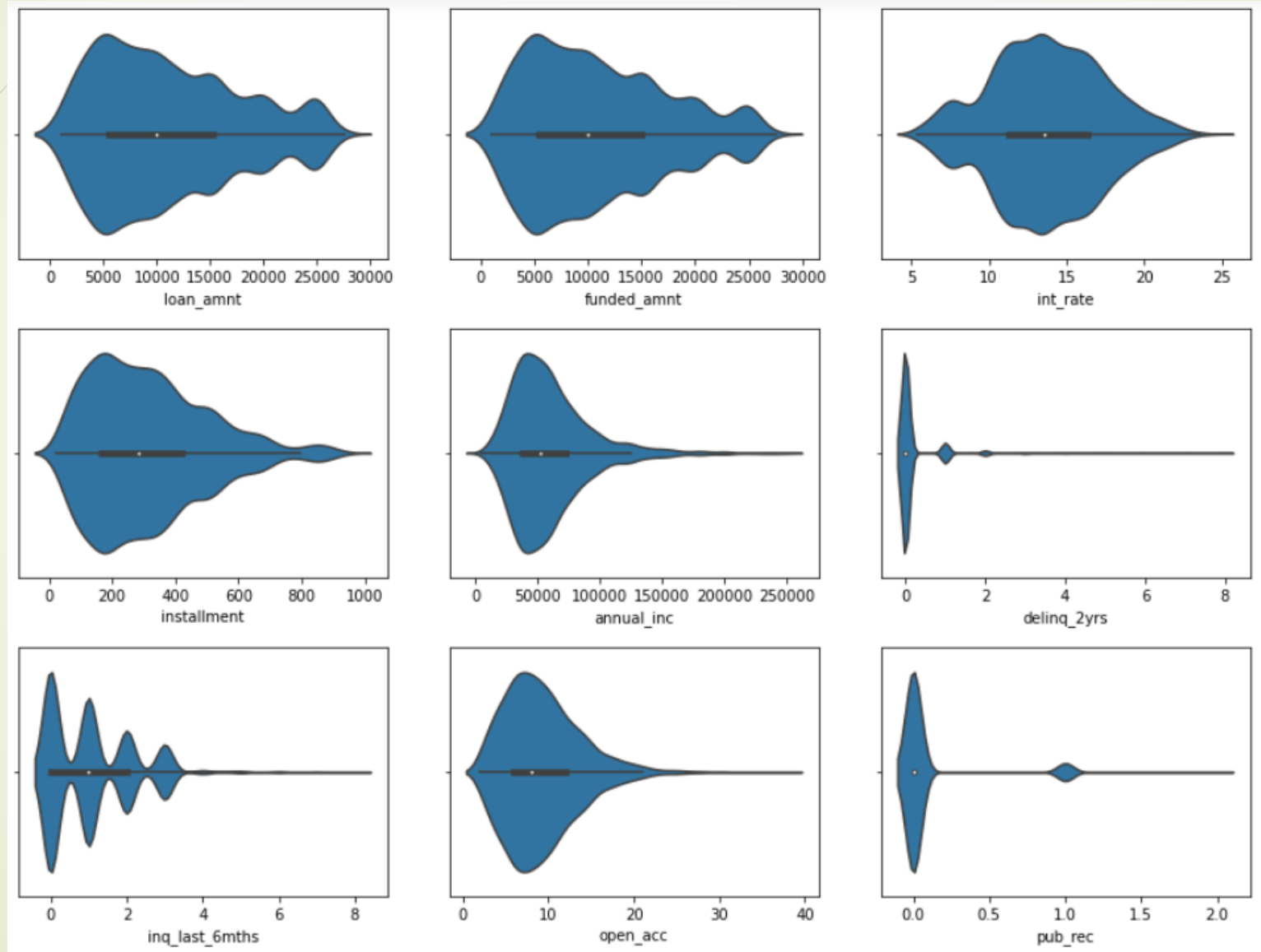
Cont..



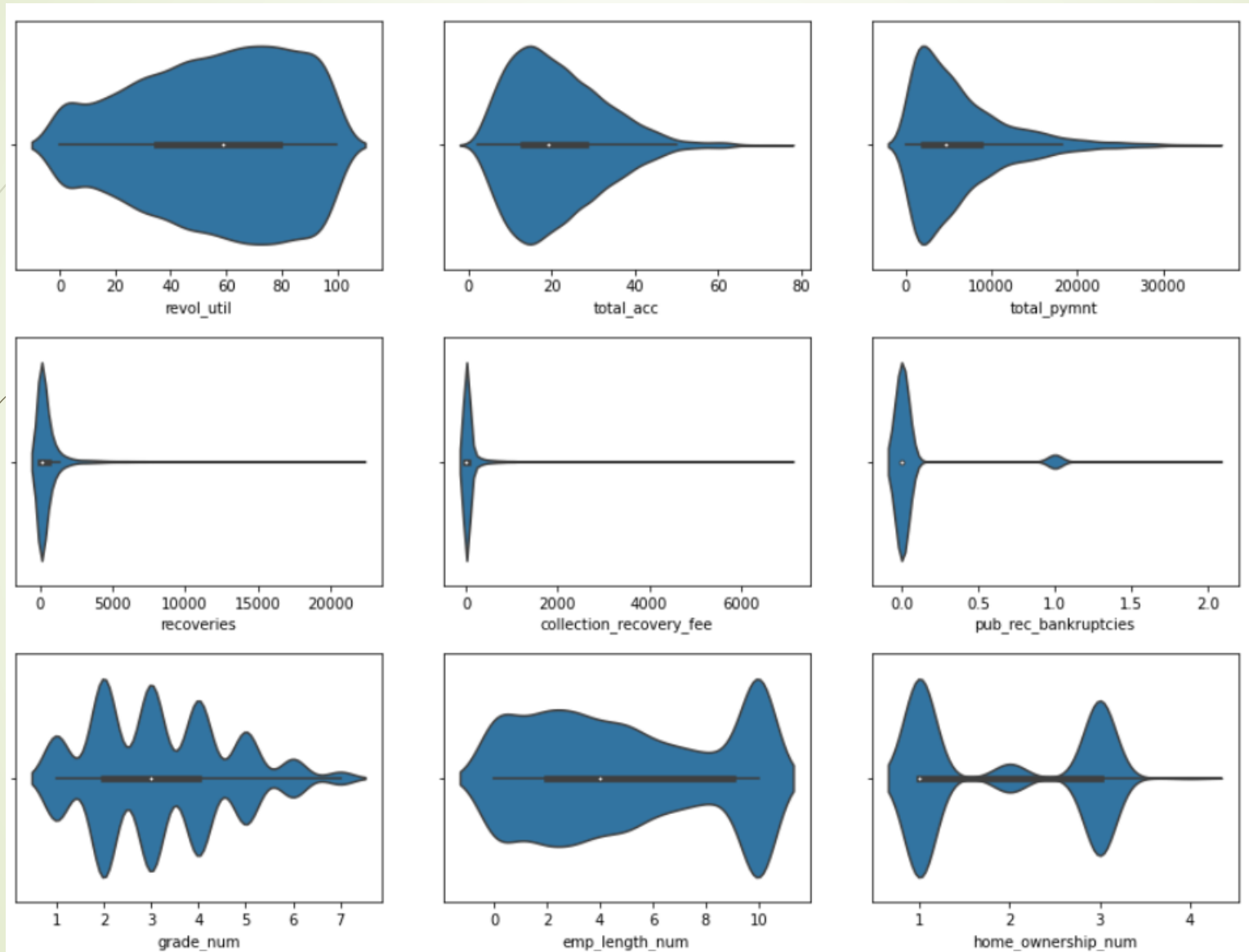
Cont..



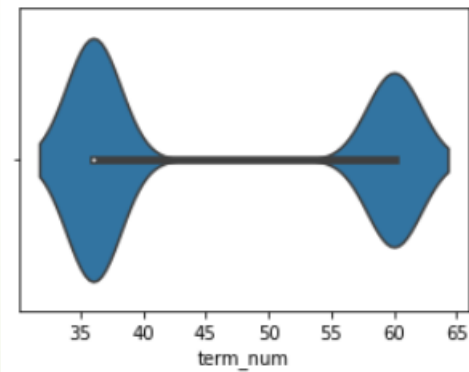
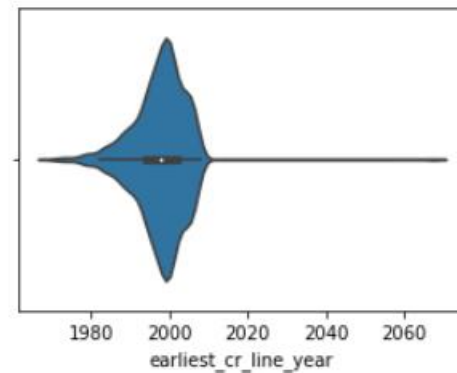
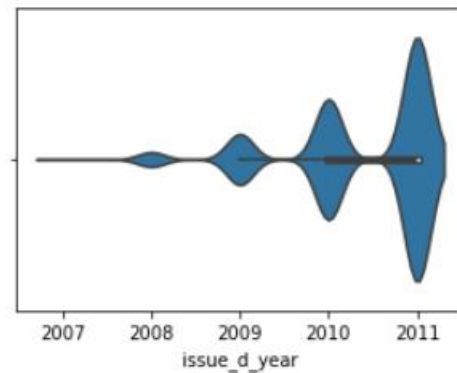
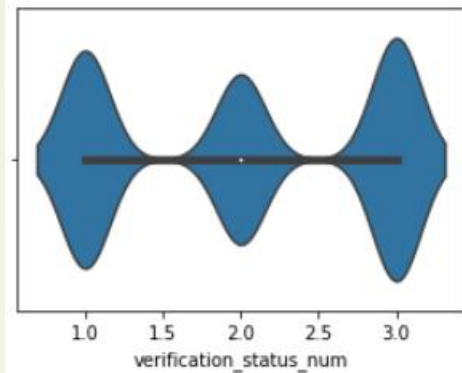
Univariate Analysis - Numerical Variables(Violin Plot) Post Outlier Removal



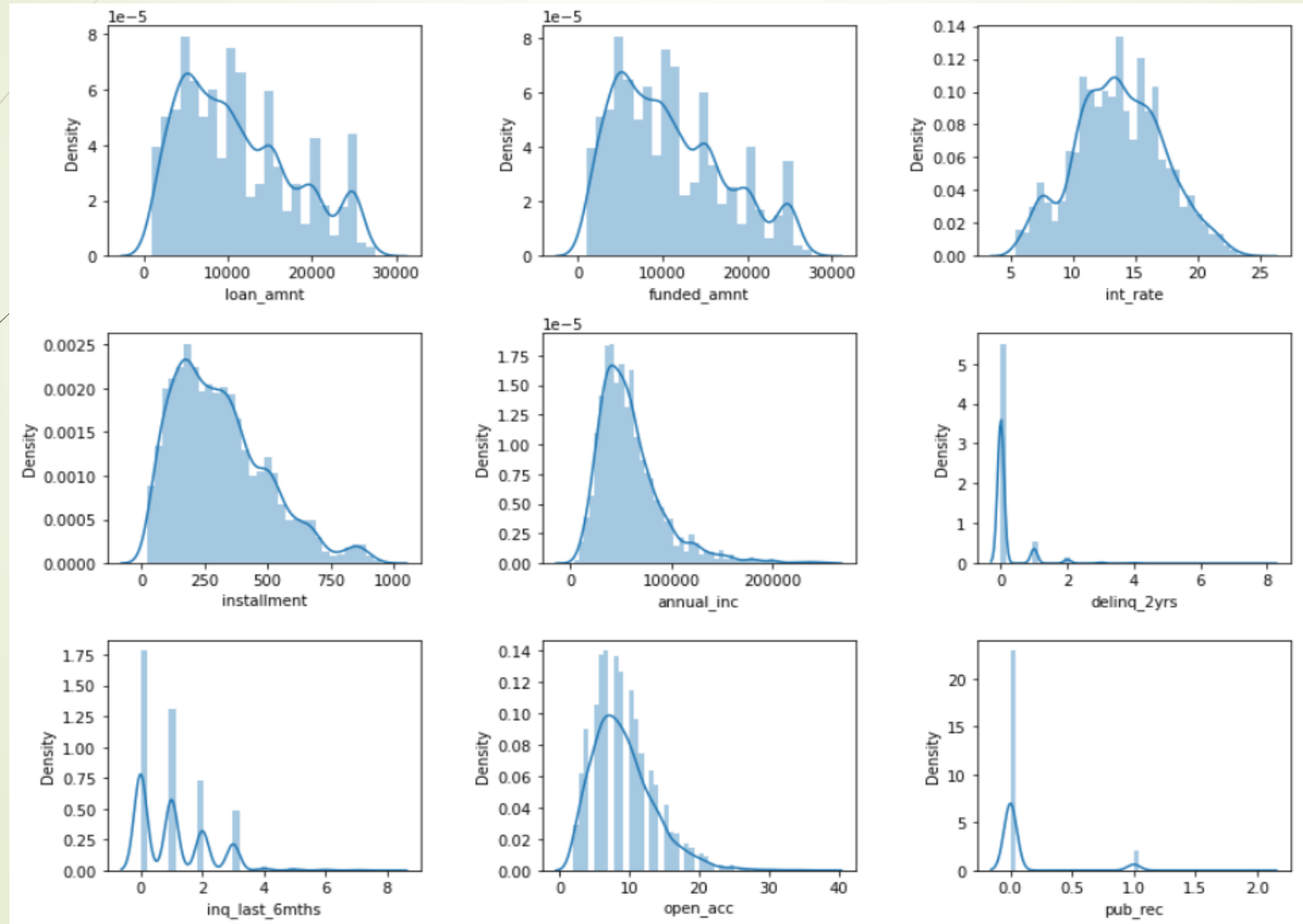
Cont..



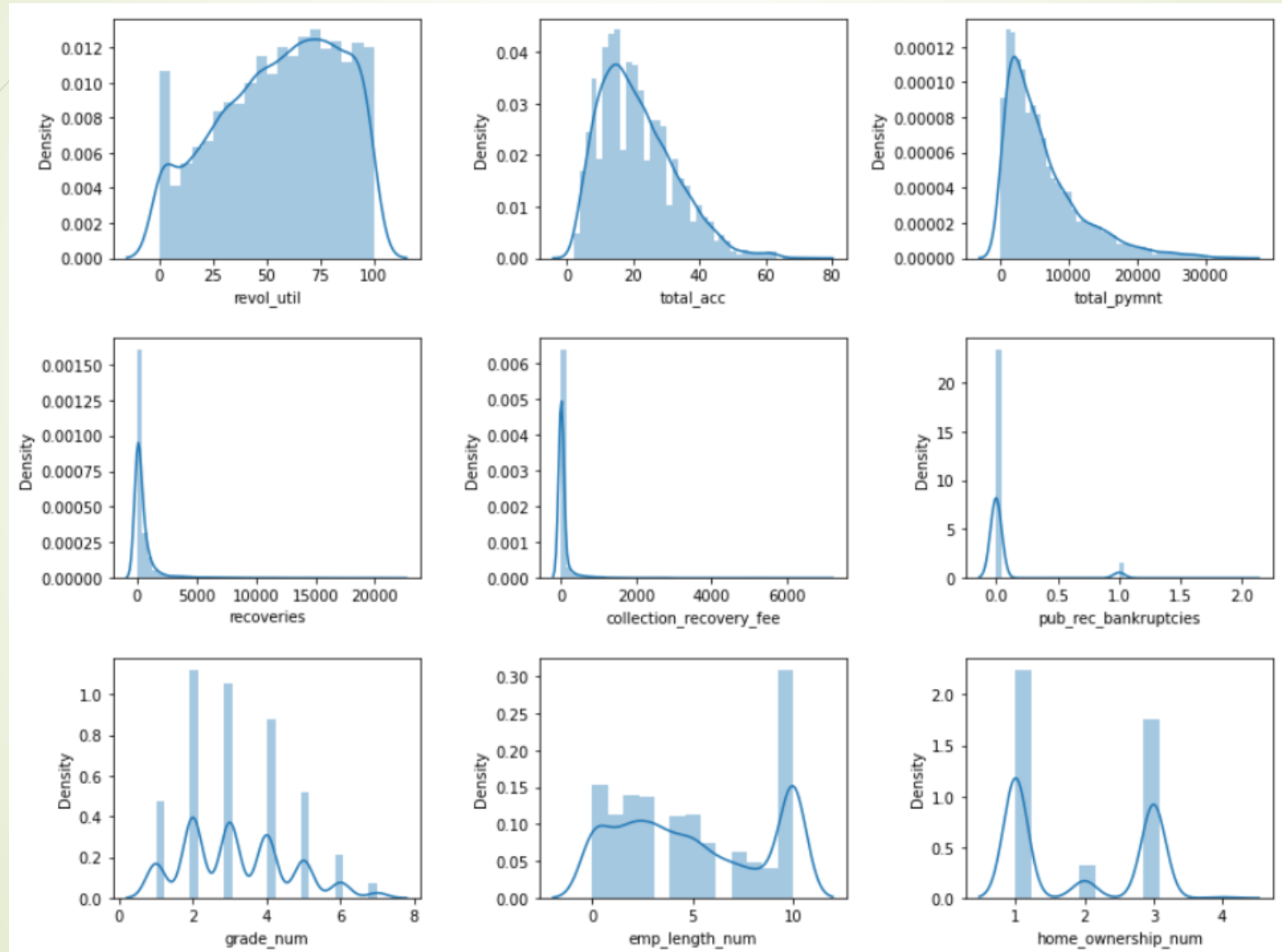
Cont..



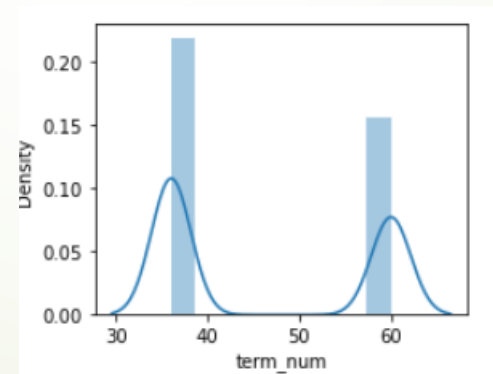
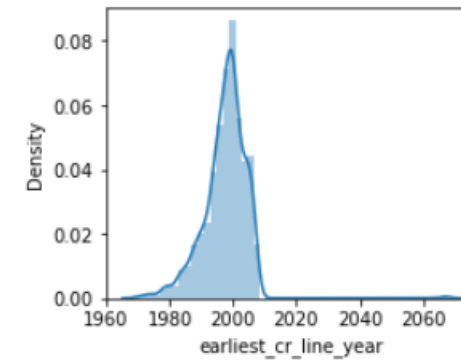
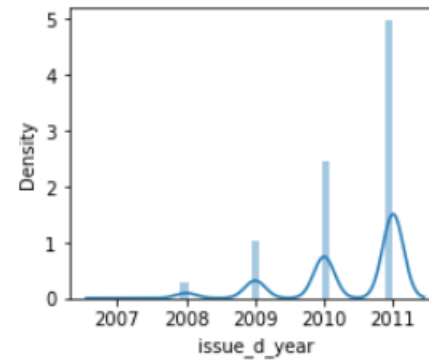
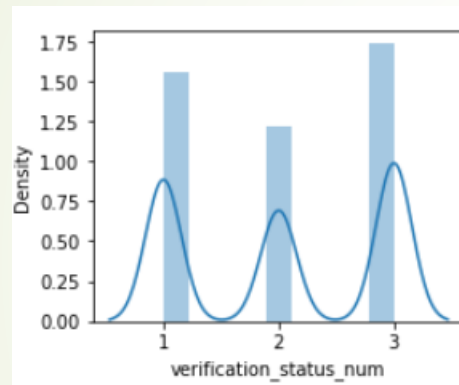
Univariate Analysis - Numerical Variables (Distribution Plot) Post Outlier Removal



Cont..



Cont..

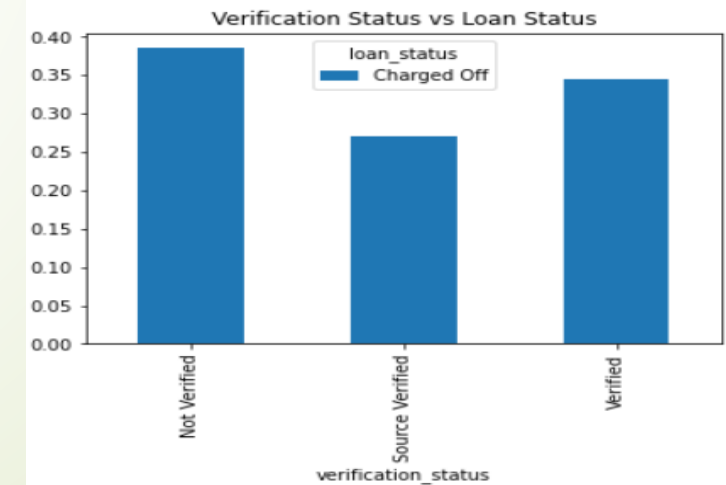
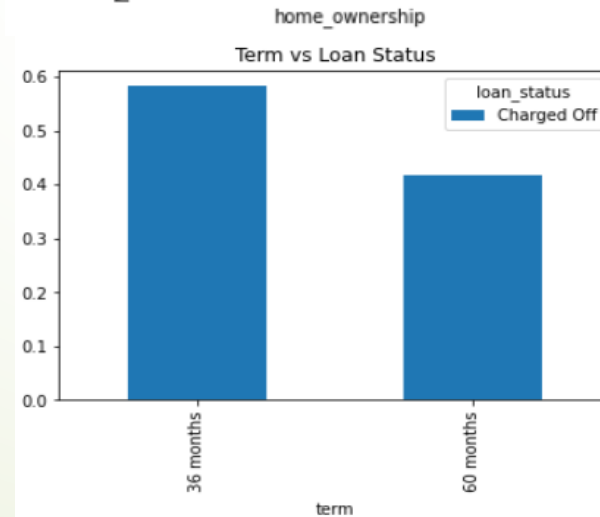
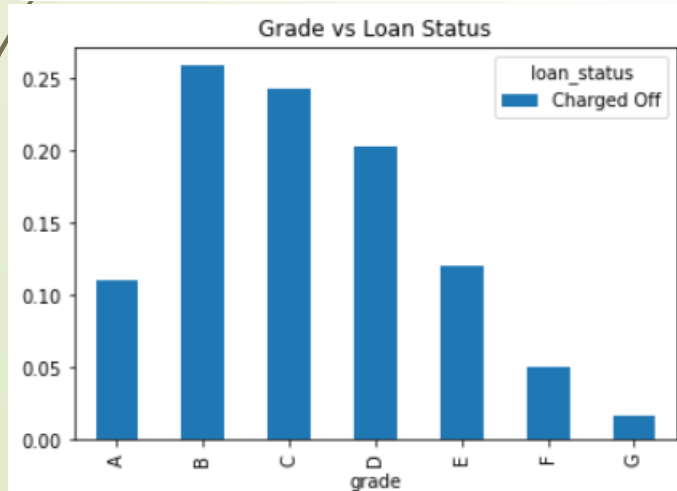
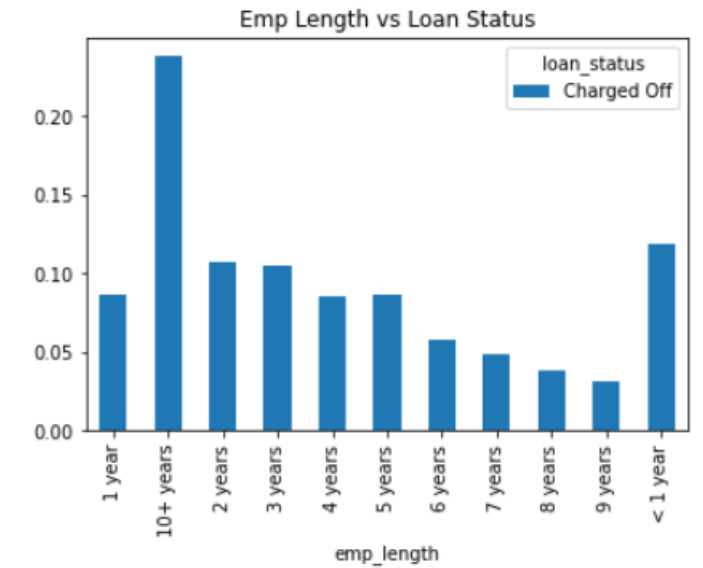
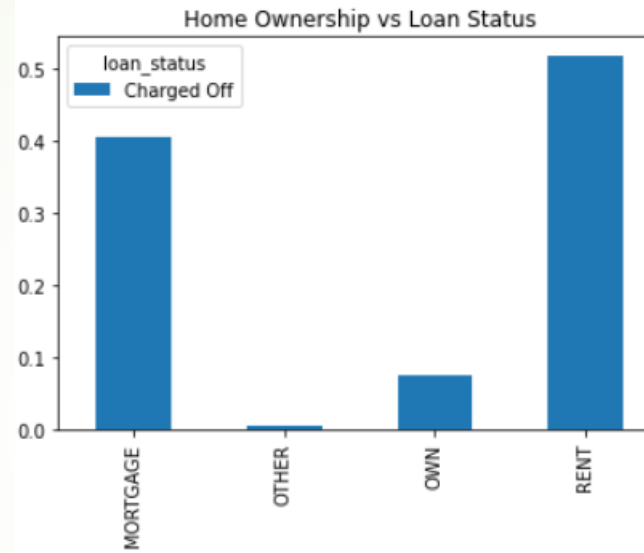
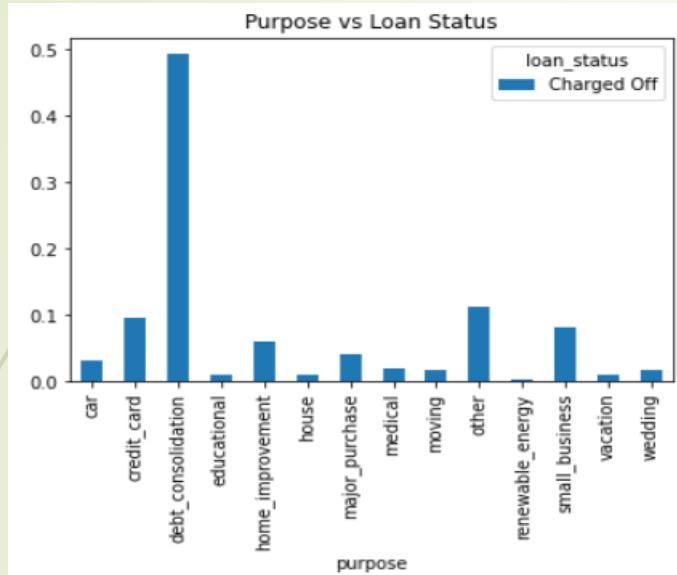




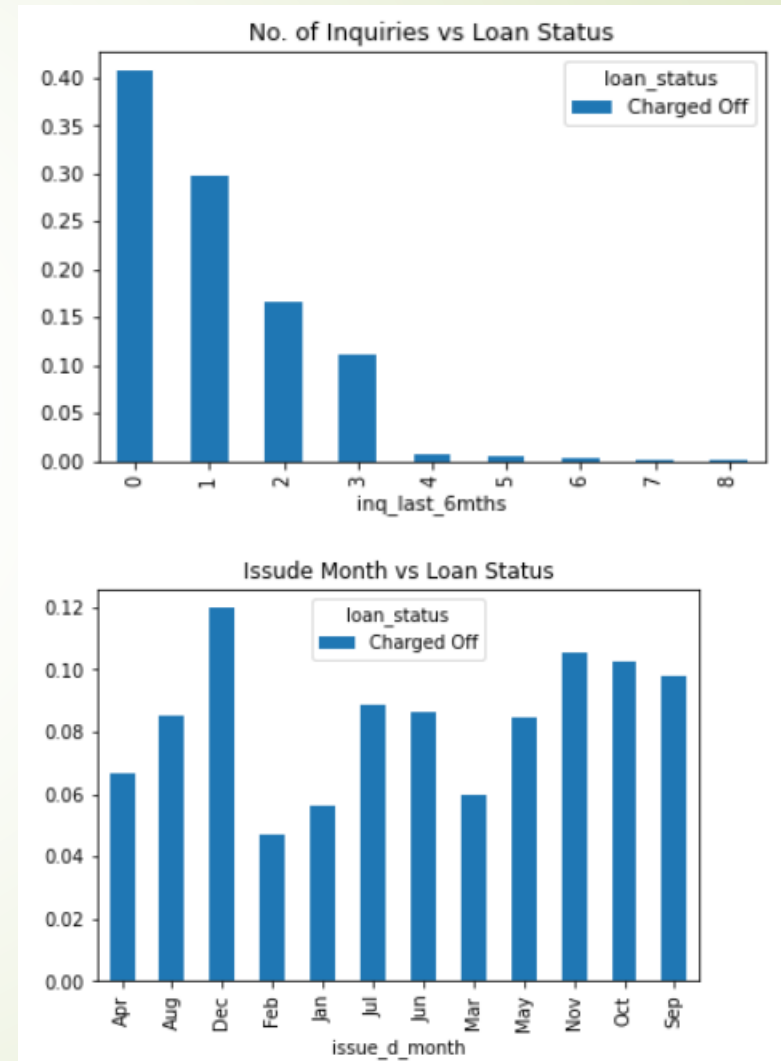
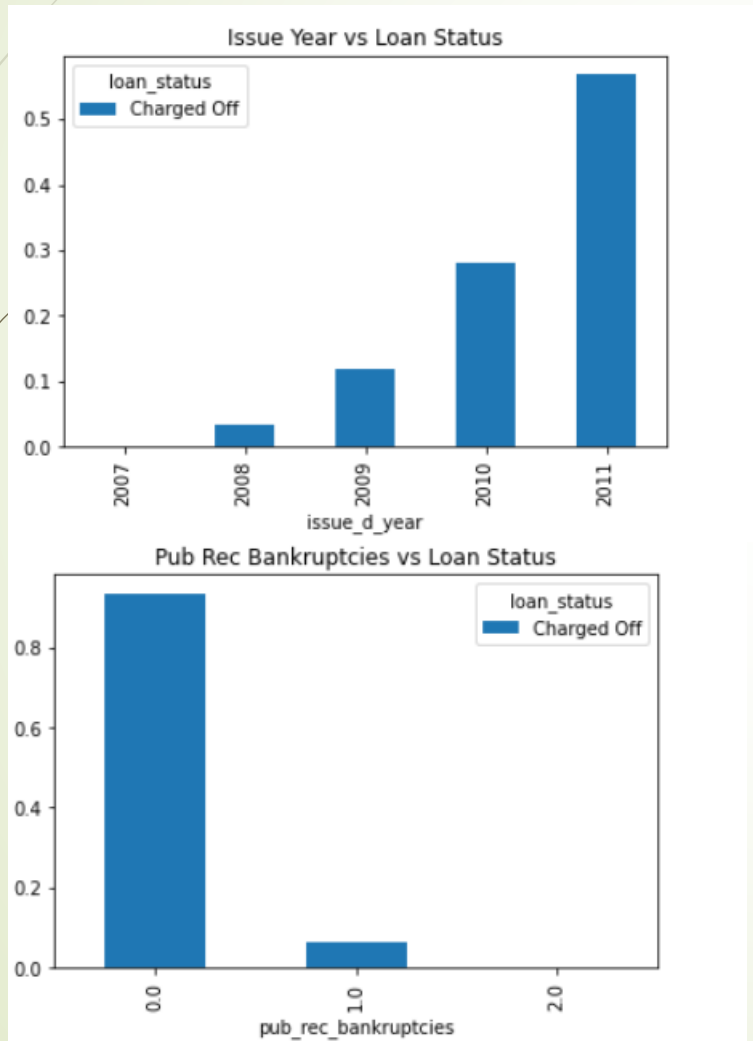
Univariate Analysis Continuous Variables Observations

- Most of the loan amounts are in the range of 5500 to 15250 dollars
- Most of the annual incomes are in the range of 37000 to 72000 dollars
- Most of the interest rates are in the range of 11.28% to 16.29%
- Most of the instalment amounts are in the range of 167 to 417 dollars
- Most of the revolving credit utilization is in the range of 34% to 99%
- Most of the loan grades are in the range B to D
- Most of the employee length are in the range of 2 years to 9 years

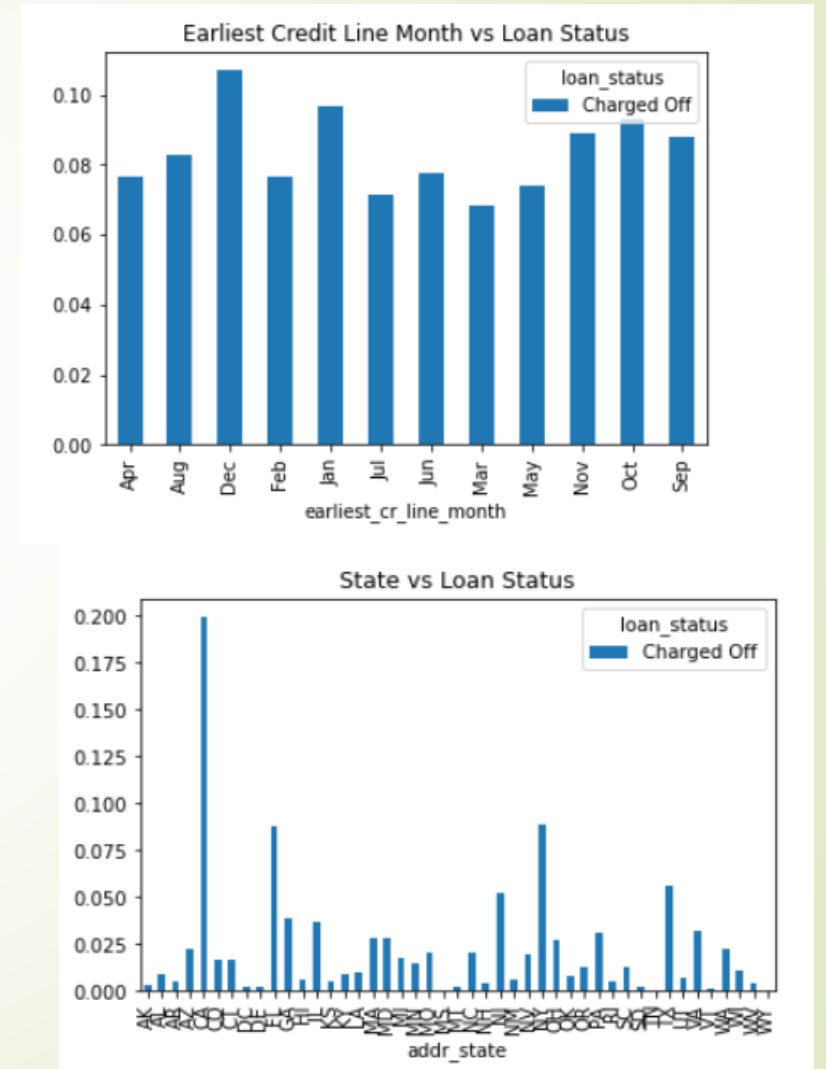
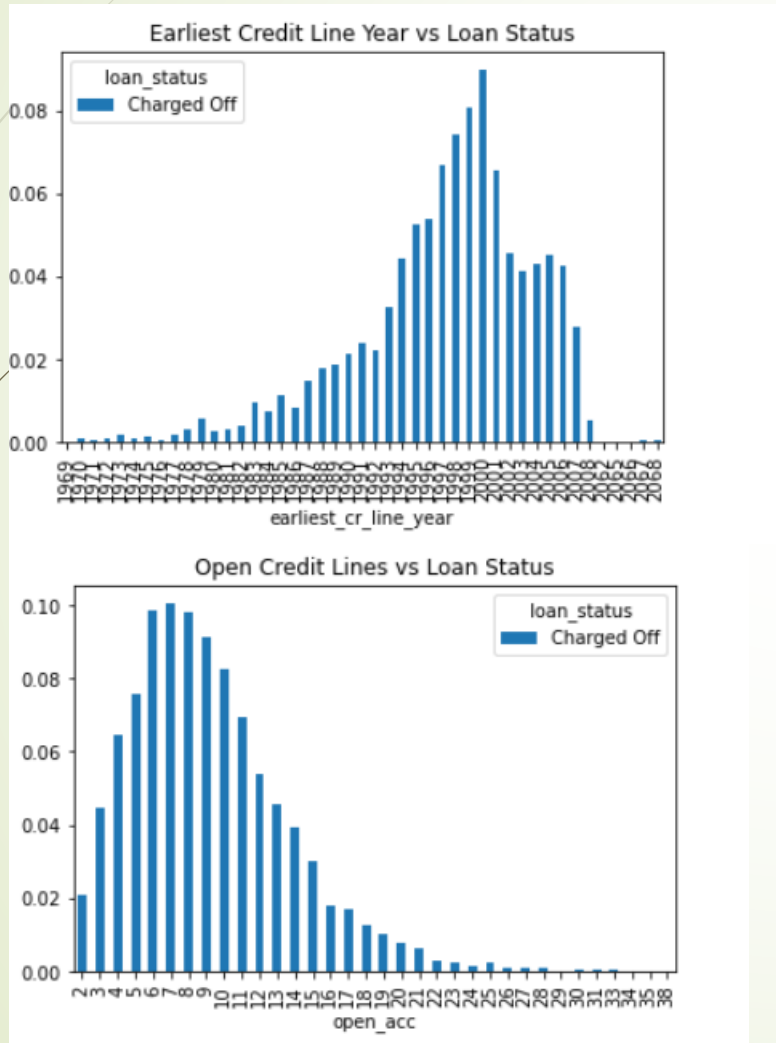
Univariate Analysis Categorical Variables



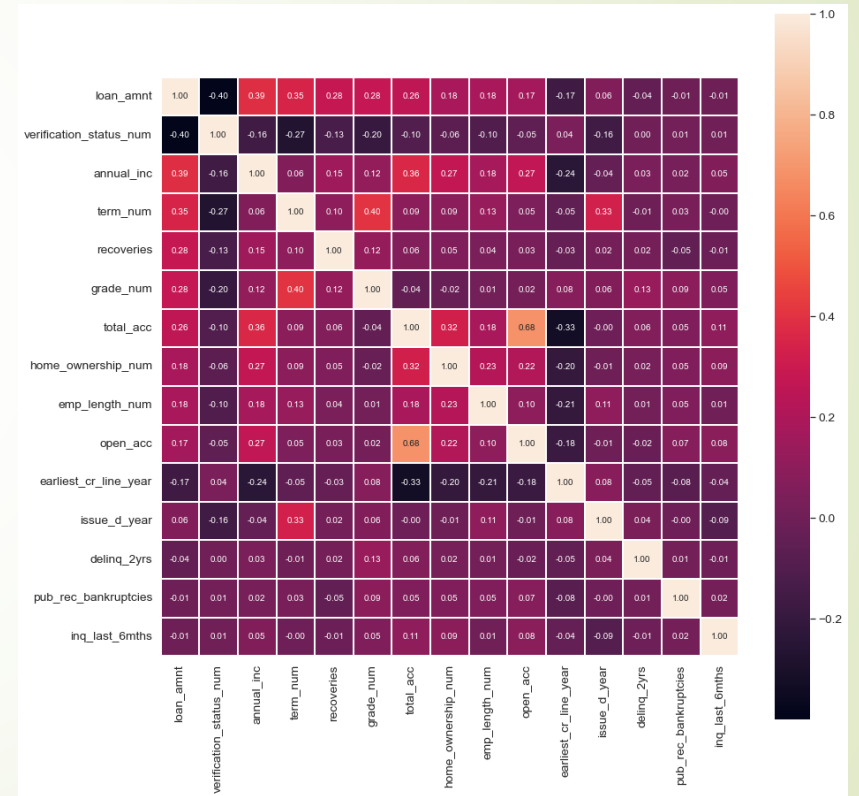
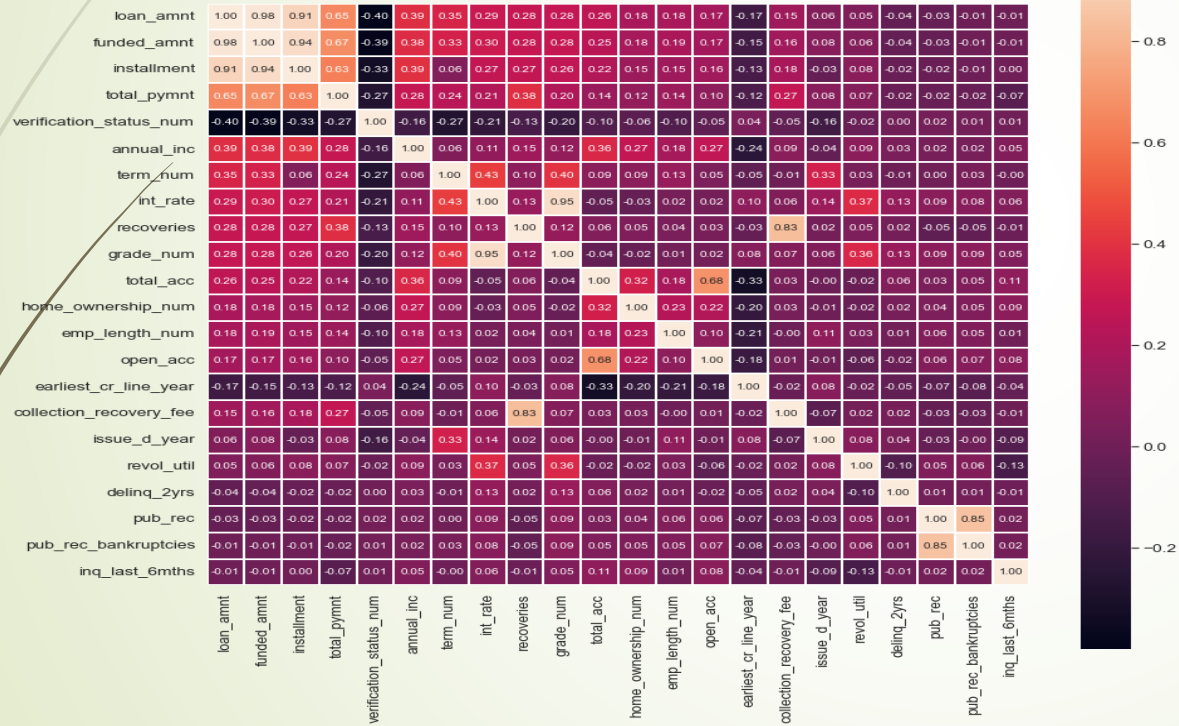
Cont..



Cont..



Bivariate Analysis (Heat Map)

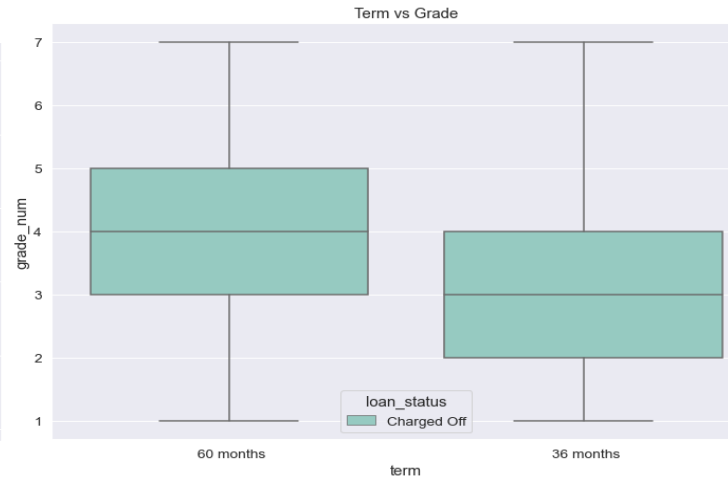
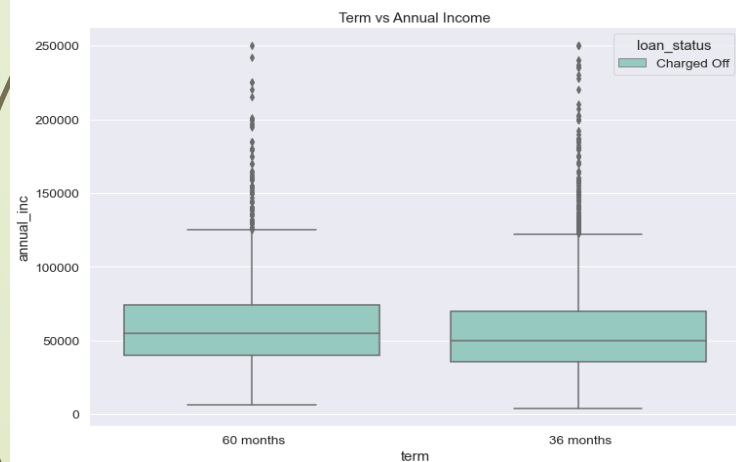
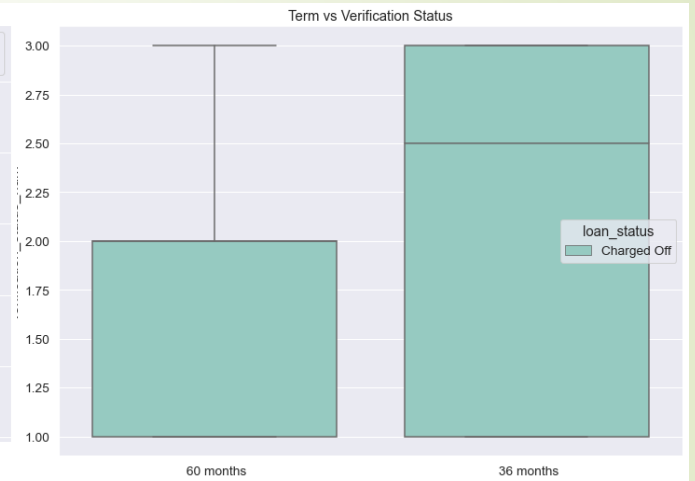
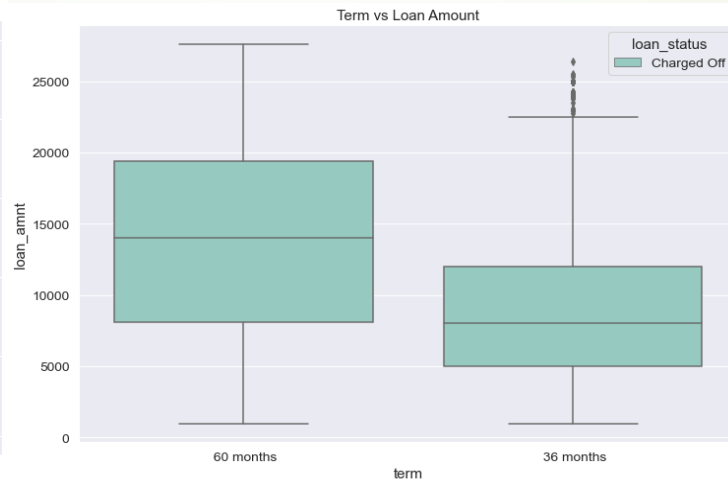
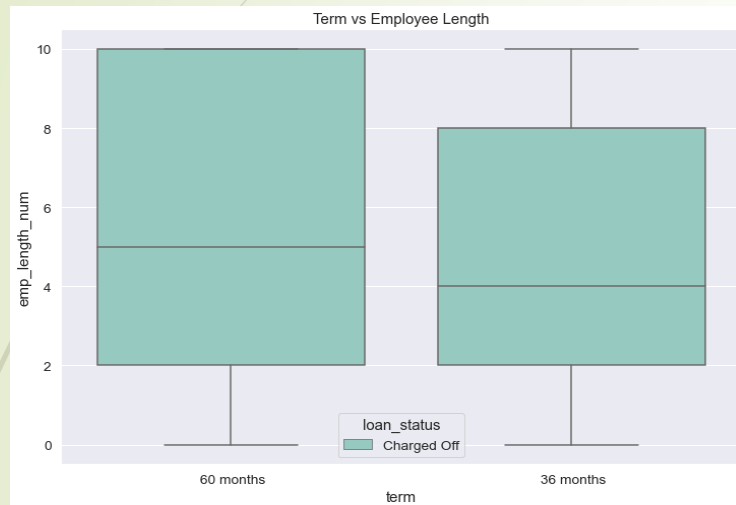




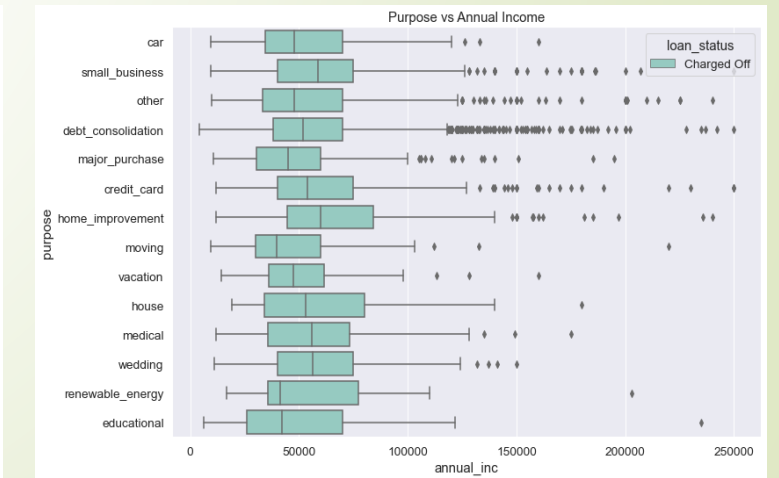
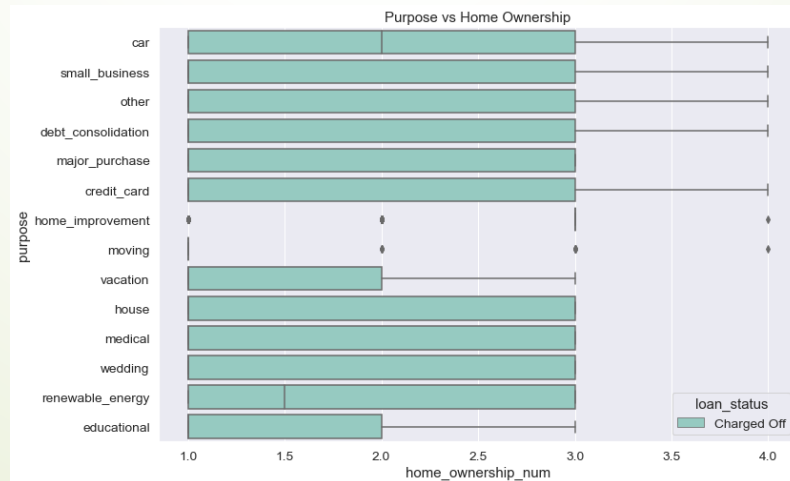
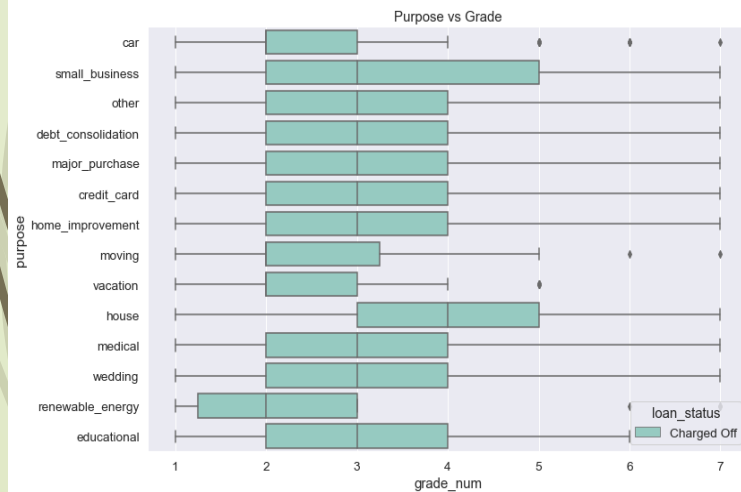
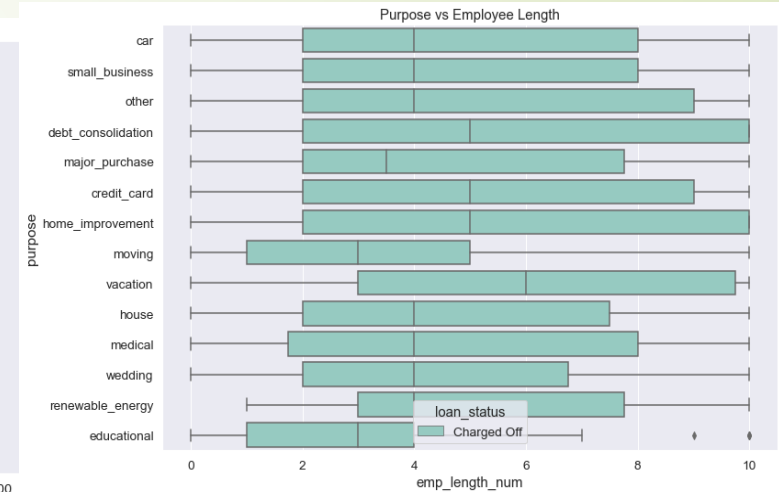
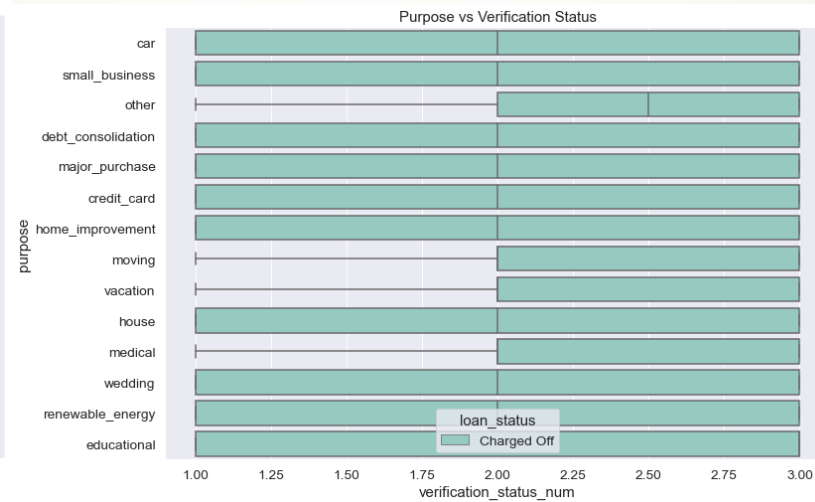
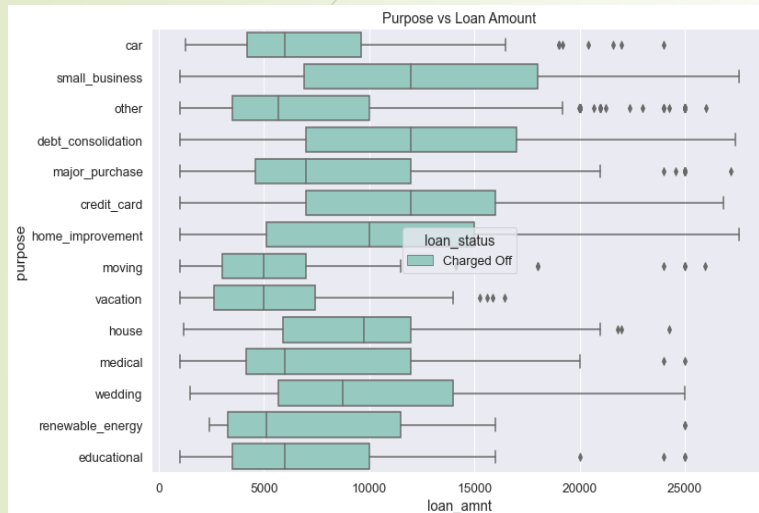
Bivariate Analysis Observations

- ▶ loan_amnt, funded_amnt, installment, total_pymnt are highly correlated and only one of them can be used.
- ▶ int_rate and grade are highly correlated and only one of them can be used.
- ▶ recoveries and collection_recovery_fee are highly correlated and hence, only one of them can be used.
- ▶ pub_rec and pub_rec_bankruptcies are highly correlated and hence, only one of them can be used.
- ▶ revol_util and int_rate are slightly correlated which means that higher revol_util will have higher interest rate too.
- ▶ emp_length and int_rate are not at all correlated - which means that int_rate will be higher for people with lesser emp experience.
- ▶ open_acc and verification_status has no clear correlation with any of the columns

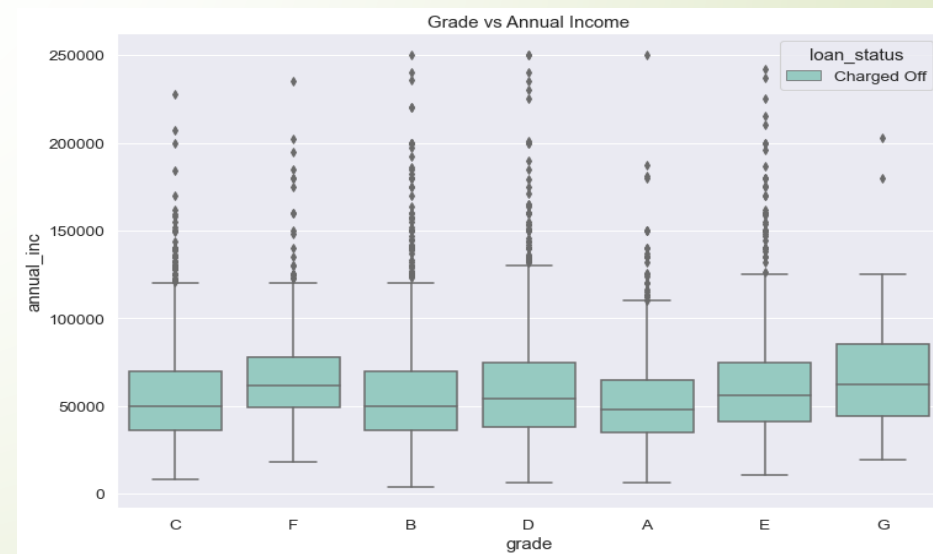
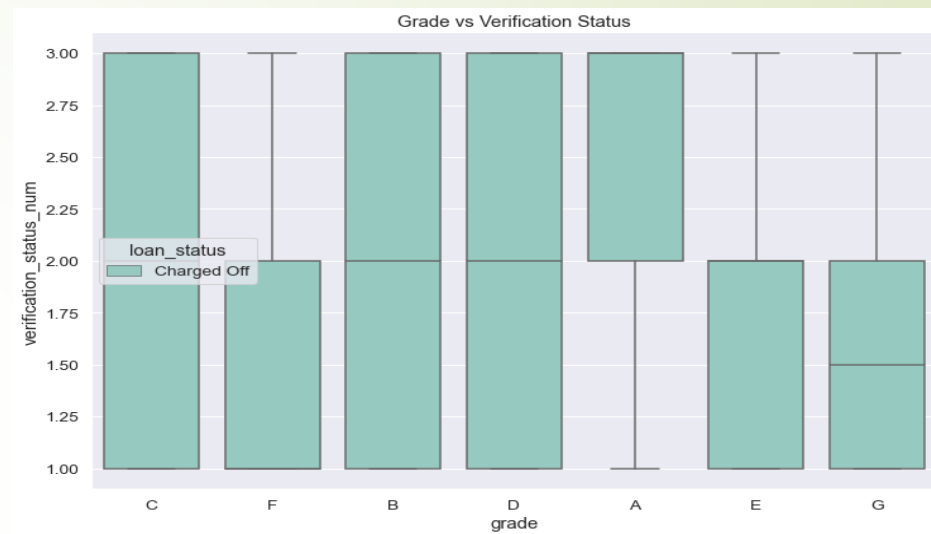
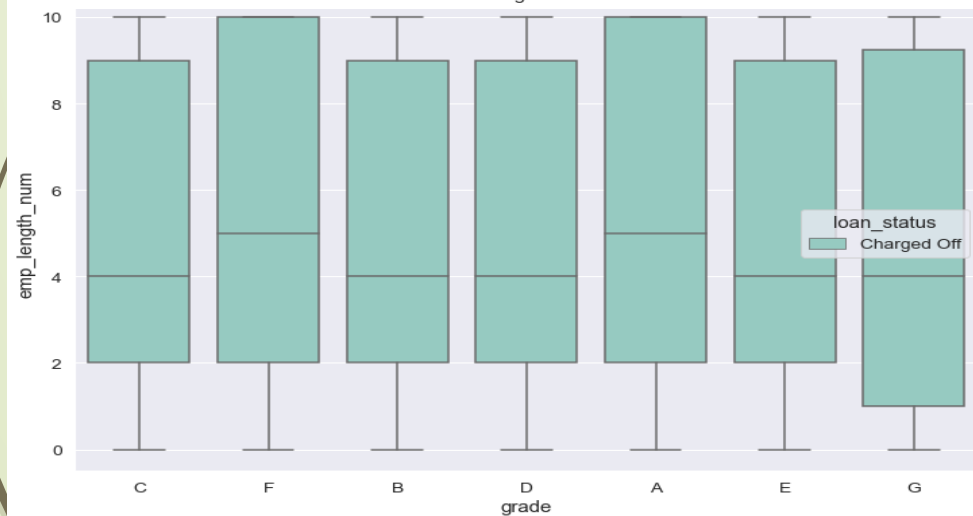
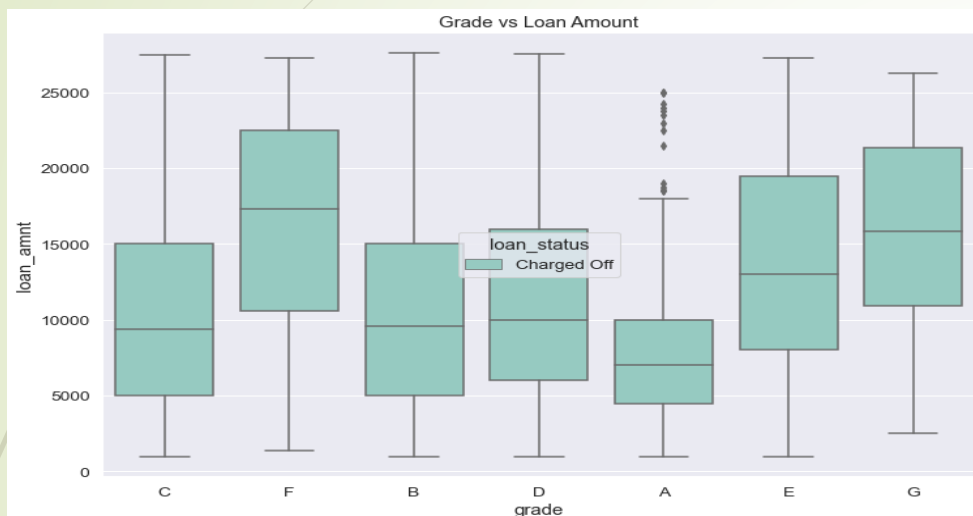
Bivariate Analysis on Loan Status



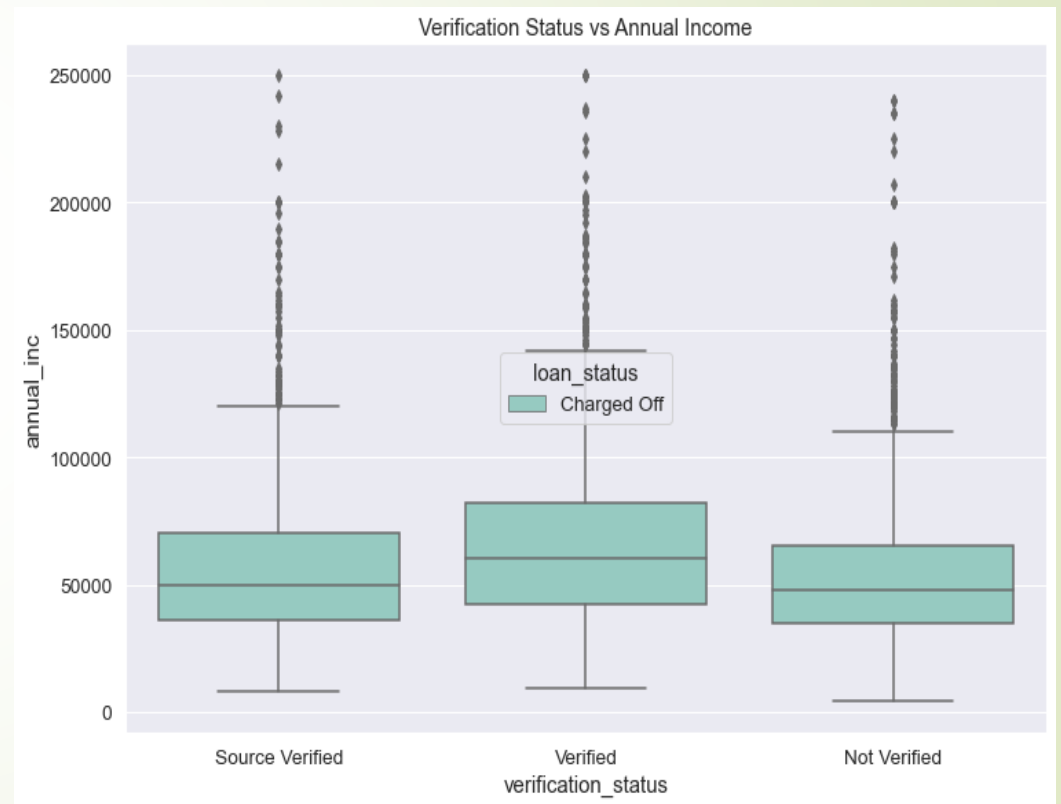
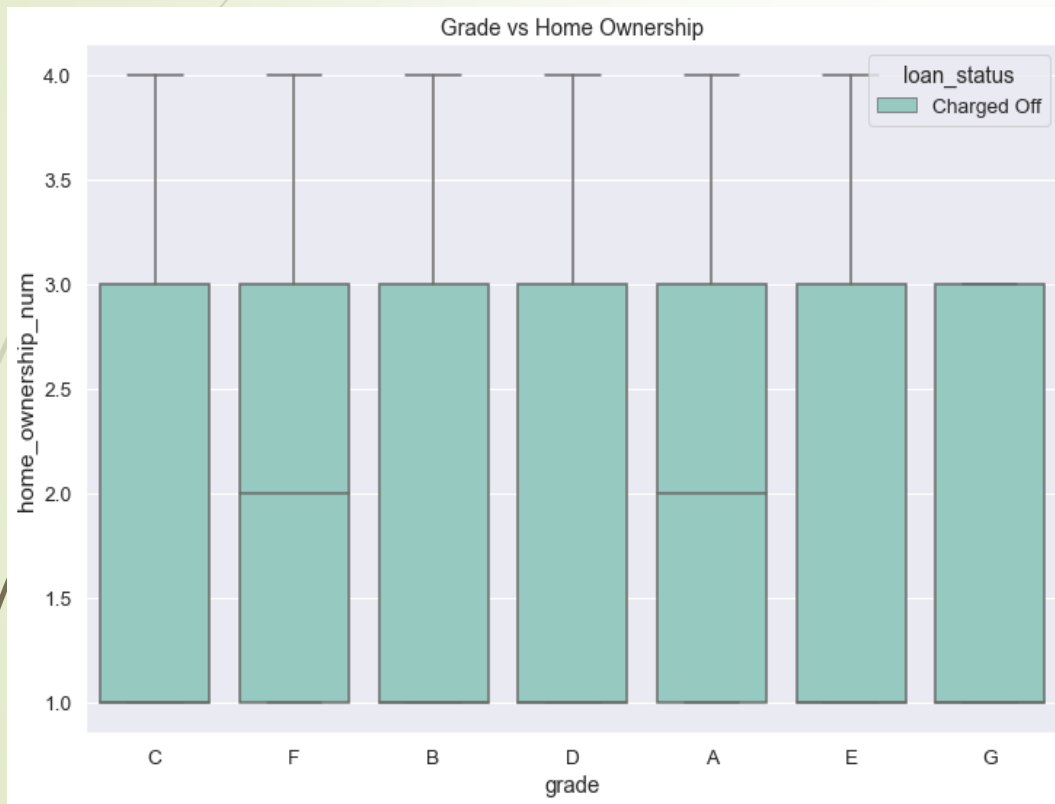
Cont..



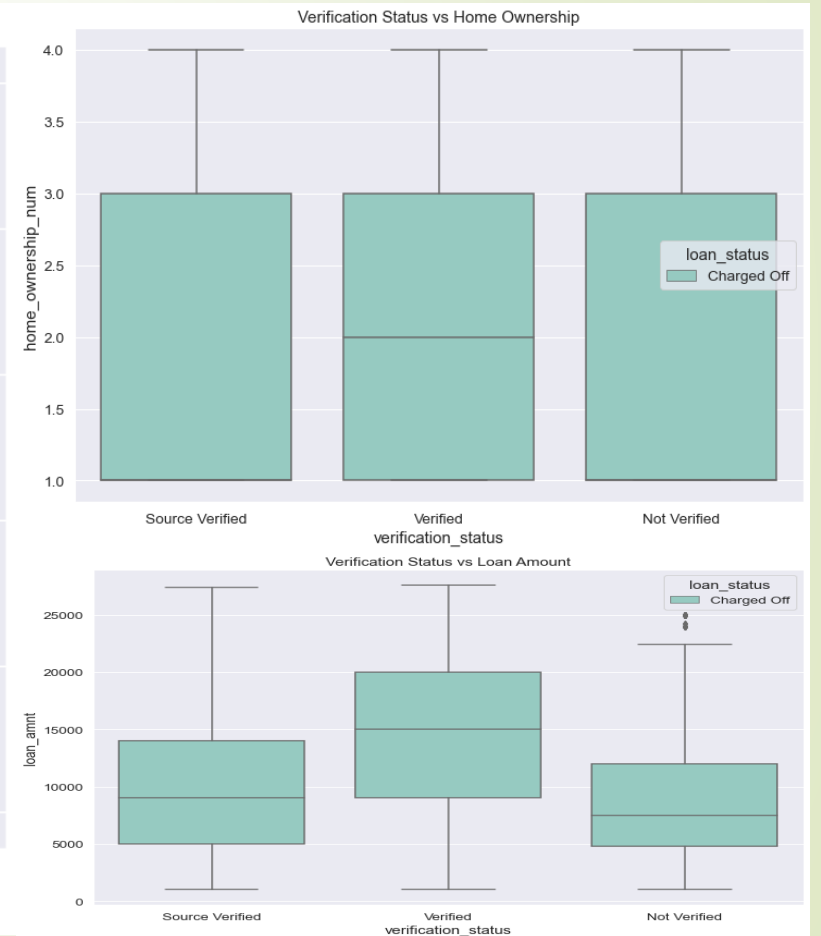
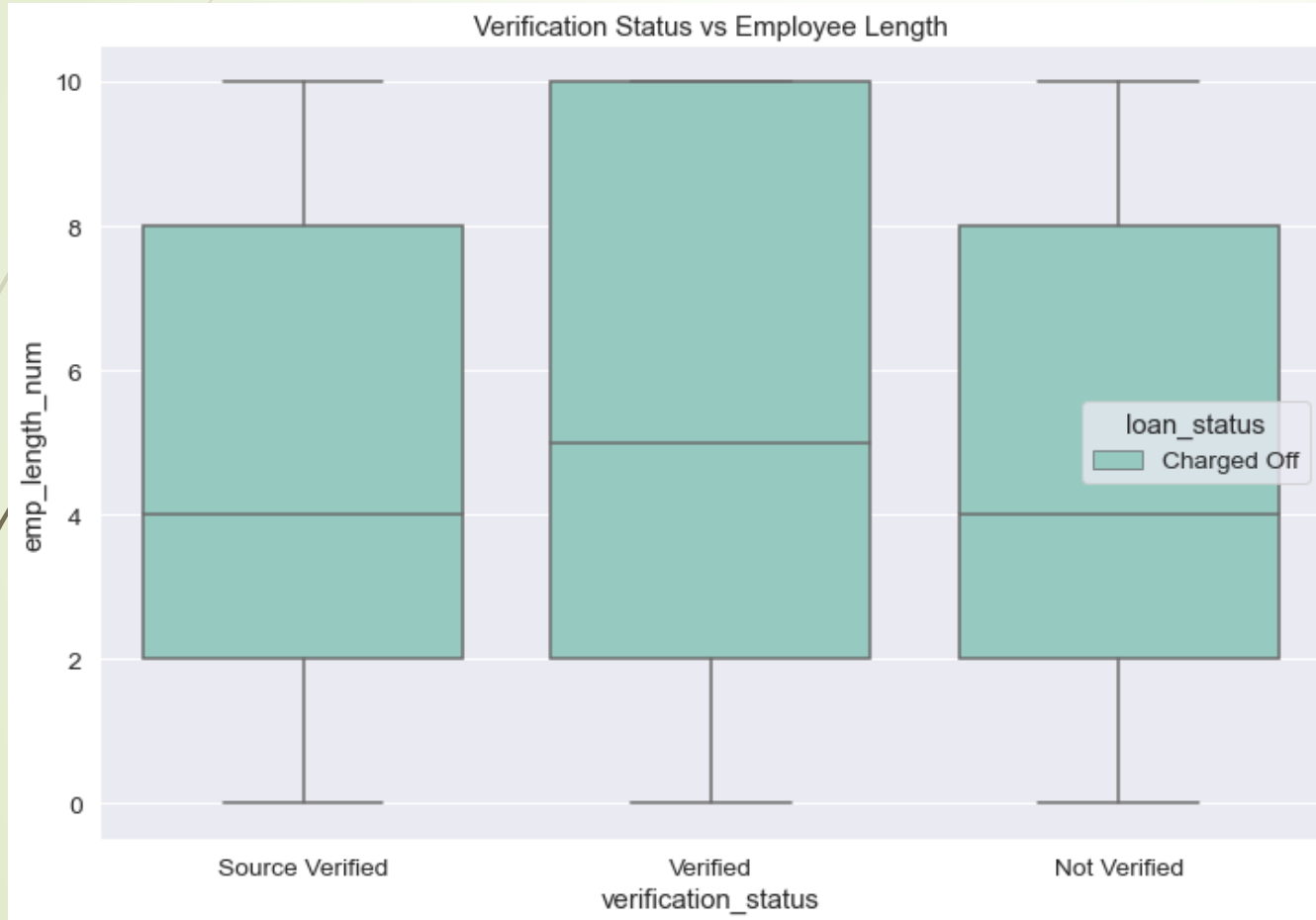
Cont..



Cont..



Cont..





Conclusions

- ▶ The top driving factors that can identify potential loan defaulters are:
 - **Grade** - E, F, G loan grades with respect to loan amount and B, C, D loan grades with respect to loan purpose
 - **Term** - Higher the term, higher the risk of a loan becoming default except with respect to verification status
 - **Annual Income** - All defaulted loan applicants have annual income in the range of 30000 - 80000 dollars
 - **Loan Amount** - All defaulted loan applicants have sanctioned loan amount in the range of 5000 - 22500 dollars
 - **Home Ownership** - Most of the defaulted loan applicants are living on "Rent"
 - **Verification Status** - Most of the defaulted loan applicants have "Verified" and "Source Verified" status



Recommendations

- Since Verification Status is one of the driving factors, **ensure** that the loan applicant's **income and the source of income are thoroughly verified**
- Since loan applicants living on "Rent" are already paying significant amount of their income towards their rent, try **adding "Rent" as a debt while calculating the debt-to-income ratio**
- Since lower loan amounts are at high risk of being defaulted, be a little **stringent on the "purpose" of loan**. Avoid giving loans for Debt Consolidation and Small Business. For all purposes, **give loans if loan applicant's annual income is higher than 80000 dollars**
- Try to have **multiple and flexible loan payment term** options - in the range of 24 months to 60 months
- Even if a loan has good grades assigned to it, **bad purpose of the loan might push a loan to default**.