LENDING CLUB CASE STUDY

Exploratory Data Analysis (EDA)

Submitted by:

Ankit Pateriya

Shishir Kumar

DATA OVERVIEW

- The data has 39717 records with 111 columns.
- The data has no duplicate rows.
- The Data has no empty rows.
- The data has 54 empty columns.

DATA CLEANING & DERIVING COLUMNS

- The data requires cleaning and accordingly it was done.
- Null values were removed
- Outliers were there hence removed.
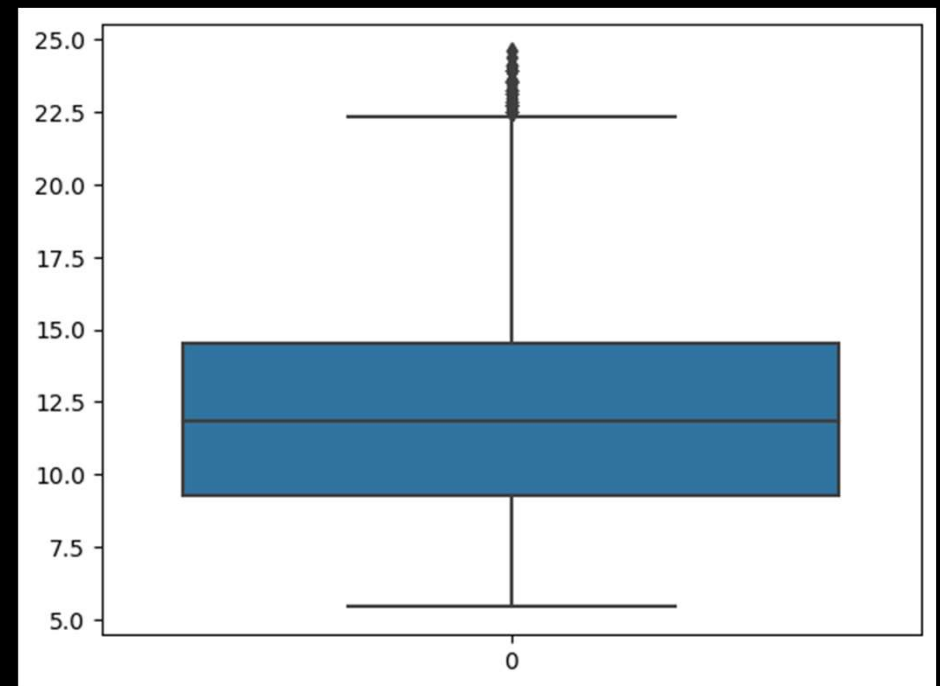- Certain columns were required to be derived for better analysys.
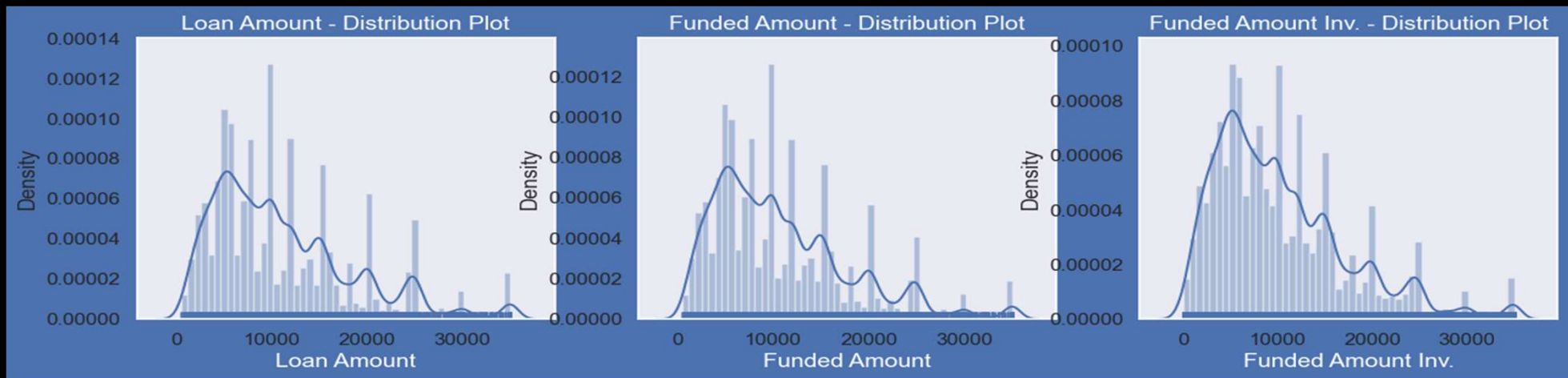
DATA ANALYSIS

- Following types of data analysis was done:-
- 1. Univariate Analysis.
- 2. Segmented Univariate Analysis.
- 3. Bivariate Analysis.
- 4. Multivariate Analysis.

UNIVARIATE ANALYSIS

- **Analysis of interest rate.**

- Average interest rate is 12 %

- After 75% percentile interest rate zoomed to 25% from 15%.
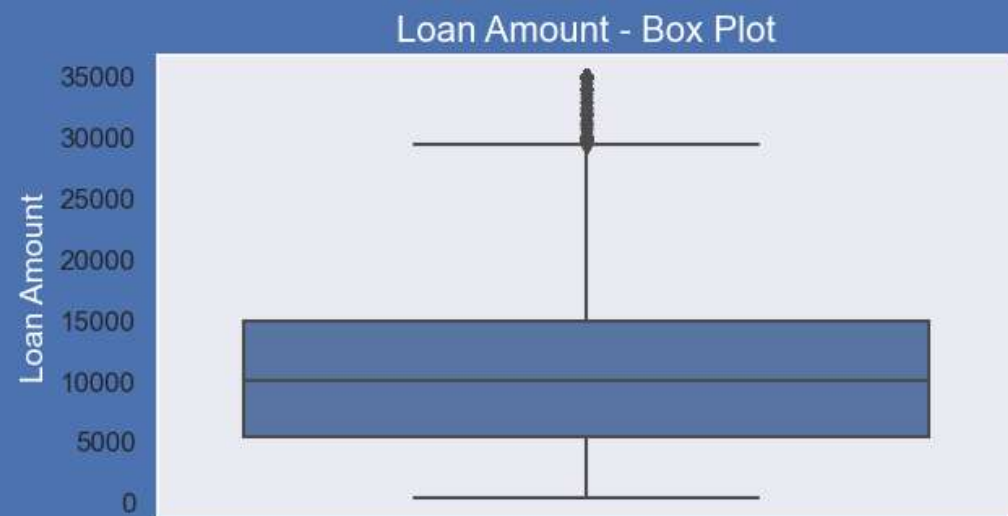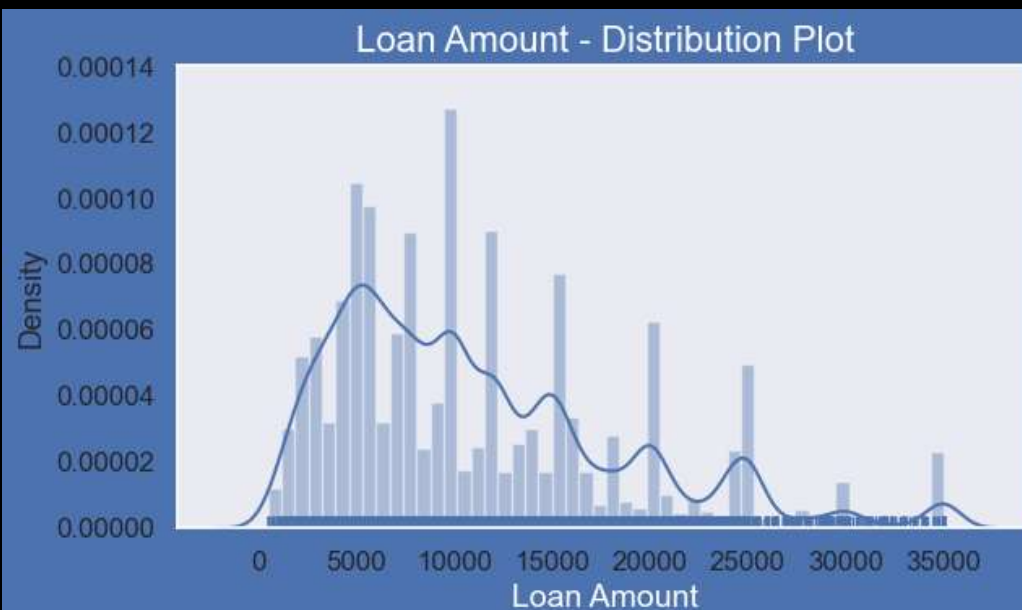
UNIVARIATE ANALYSIS

UNIVARIATE ANALYSIS

- Observation:
- Distribution of amounts for all three looks very much similar.
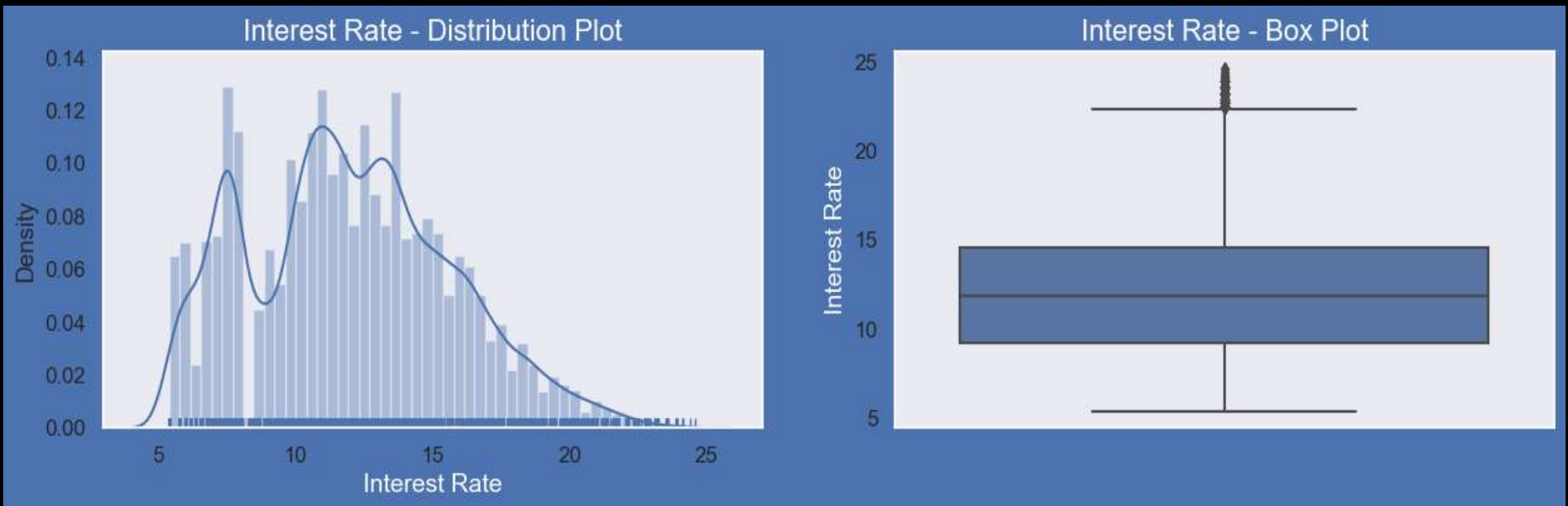- We will work with only loan amount column for rest of our analysis

UNIVARIATE ANALYSIS

- **Observations** :
- Below plots show that most of the Loan amounts are in range of 5000 - 15000
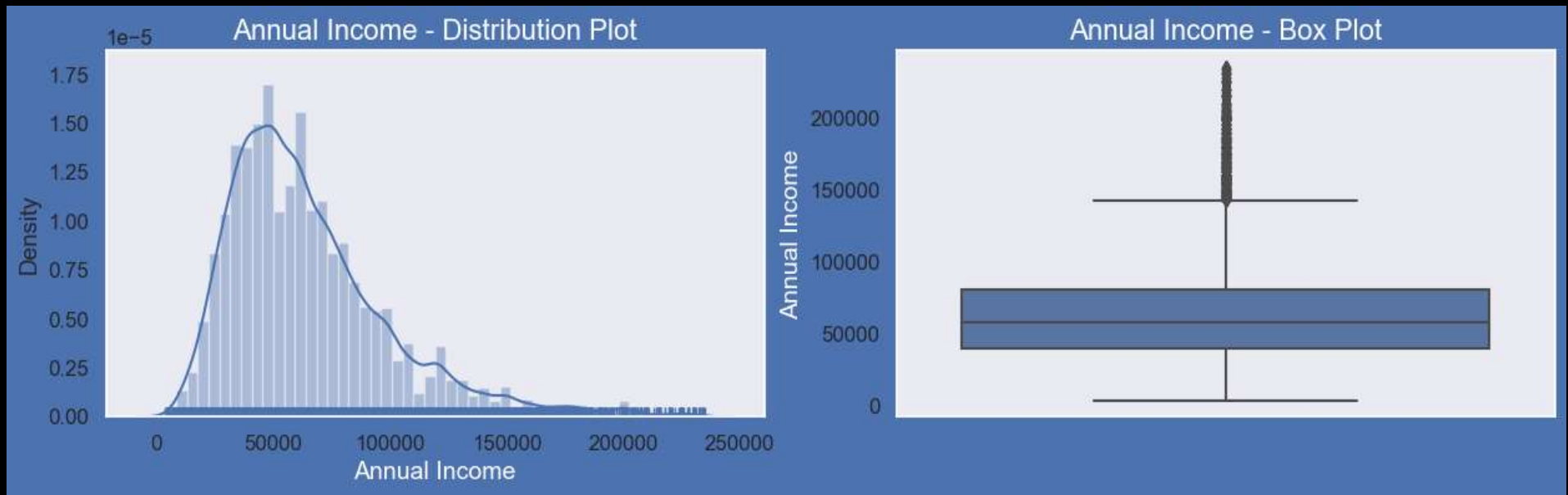
UNIVARIATE ANALYSIS

- Observations :

- Below plots show that most of the Interest Rates on loans are in range of 10% - 15%
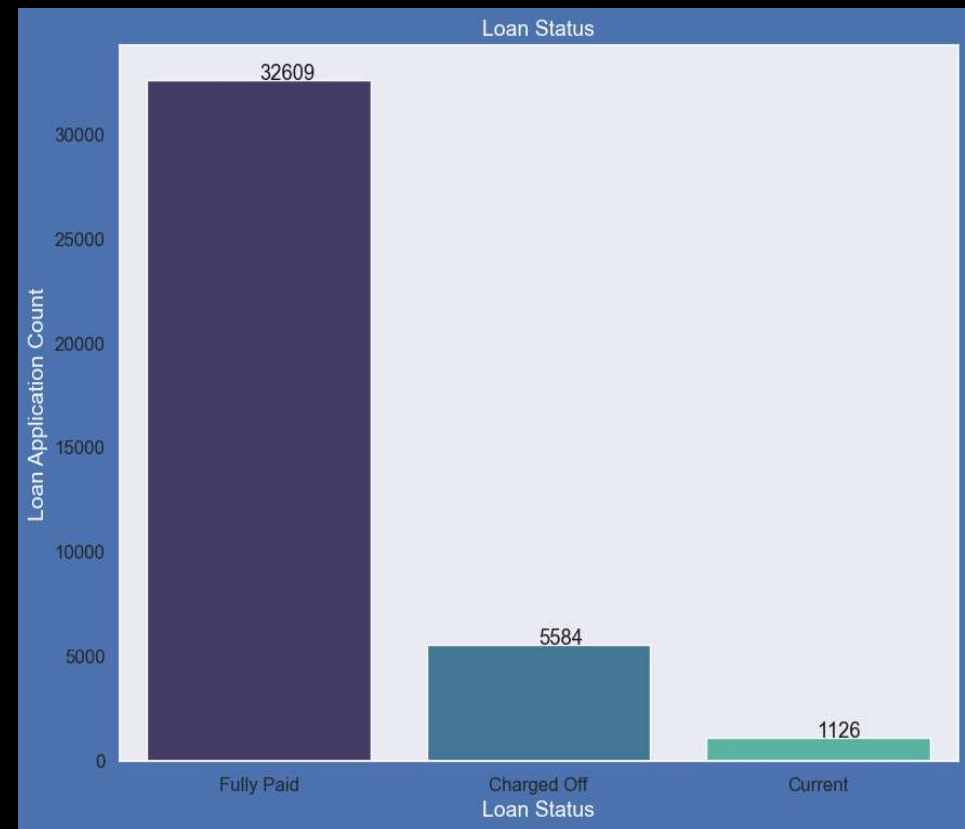
UNIVARIATE ANALYSIS

- Observations :
- Below plots show that most of the borrower's Annual incomes are in range of 40000- 80000
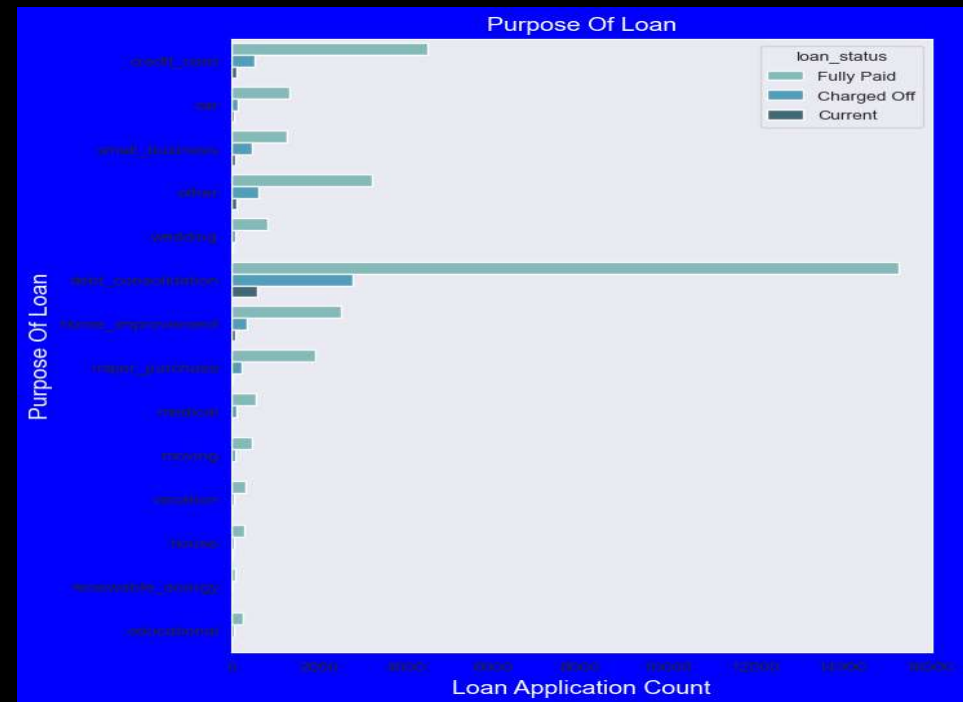
UNIVARIATE ANALYSIS

- **Observations** :

- Adjoining plot shows that close to 14% loans were charged off out of total loan issued.

UNIVARIATE ANALYSIS

- **Observations** :
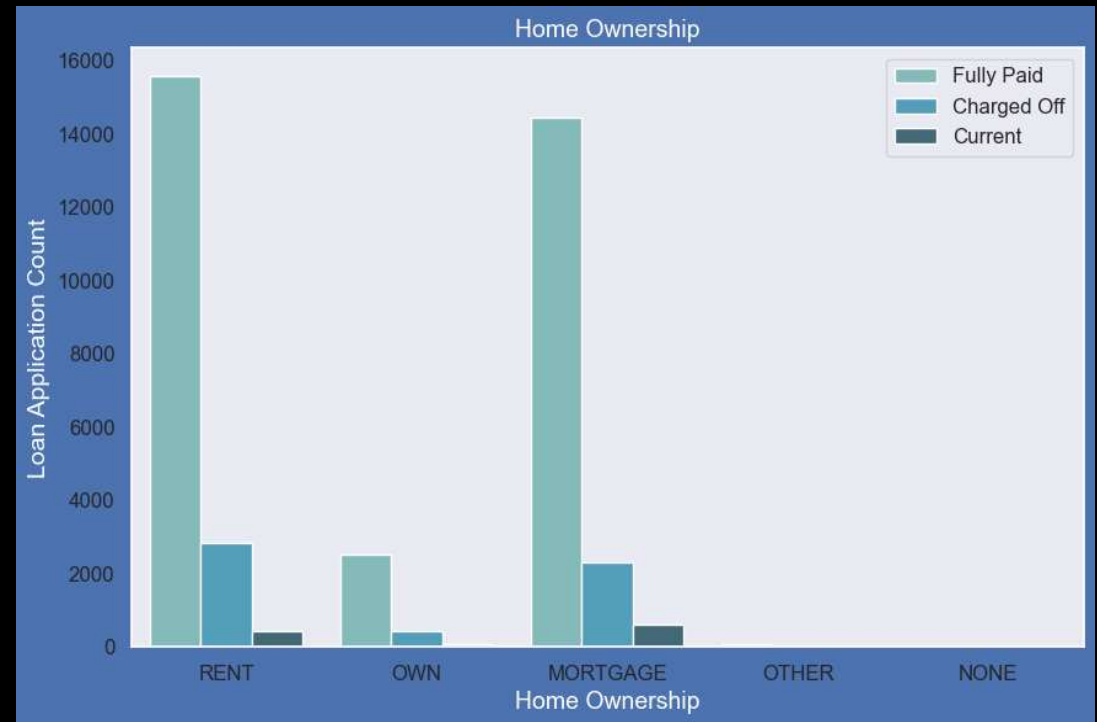- Below plot shows that most of the loans were taken for the purpose of debt consolidation & paying credit card bill.
- Number of charged off count also high too for these loans
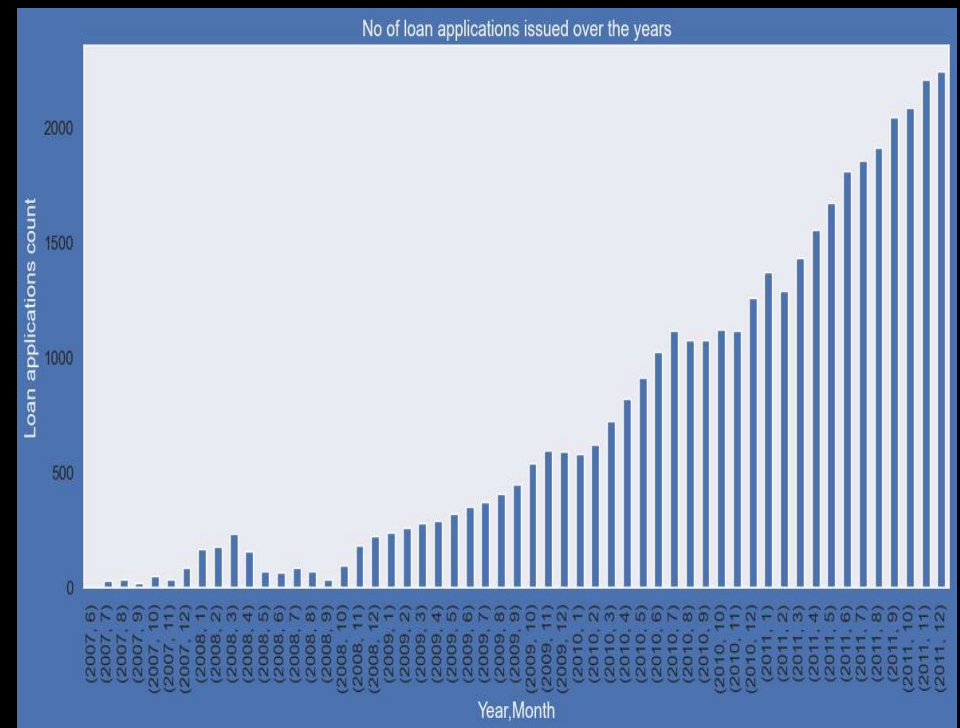
UNIVARIATE ANALYSIS

- **Observations** :

- Below plot shows that most of them living in rented home or mortgaged their home.

- Applicant numbers are high from these categories so charged off is high too.
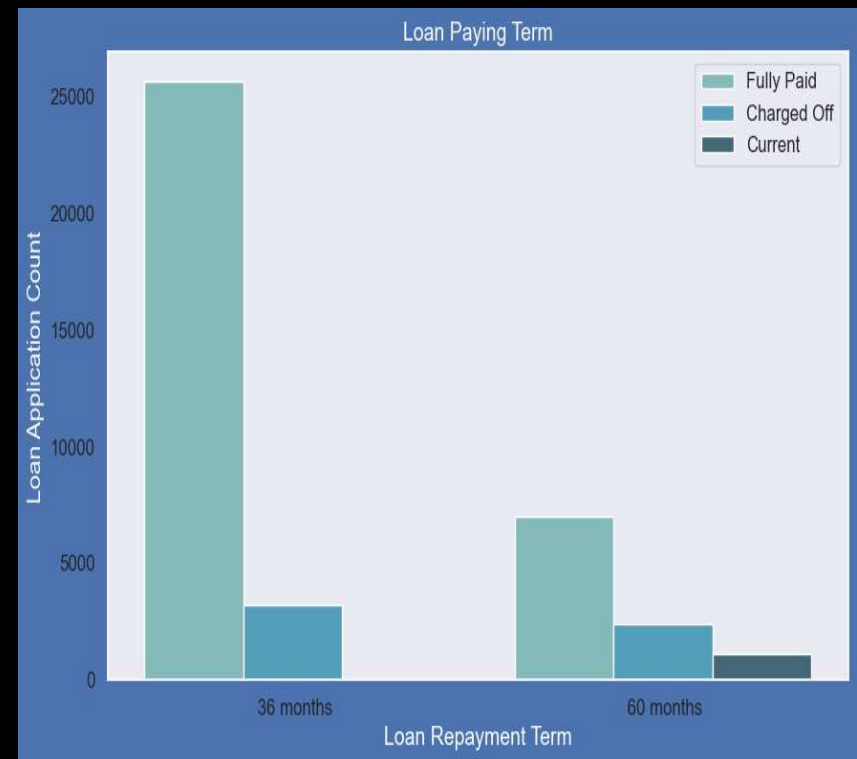
UNIVARIATE ANALYSIS

- Observation is that count of loan application is increasing every passing year.

- so increase in number of loan applications are adding more to number of charged off applications.

- Number of loans issued in 2008(May-October) got dipped, may be due to Recession.

No of loan applications issued over the years

- **Observations** :

- Below plot shows that those who had taken loan to repay in 60 months had more % of number of applicants getting charged off as compared to applicants who had taken loan for 36 months.

BIVARIATE ANALYSIS.

- Observation is that Loan amount, investor amount, funding amount are strongly correlated.

- Annual income with DTI(Debt-to-income ratio) is negatively correlated.

- Debt income ratio is the percentage of a consumer's monthly gross income that goes toward paying debts.

- That means when annual income is low DTI is high & vice versa.

- Positive correlation between annual income and employment years.

- That means income increases with work experience.
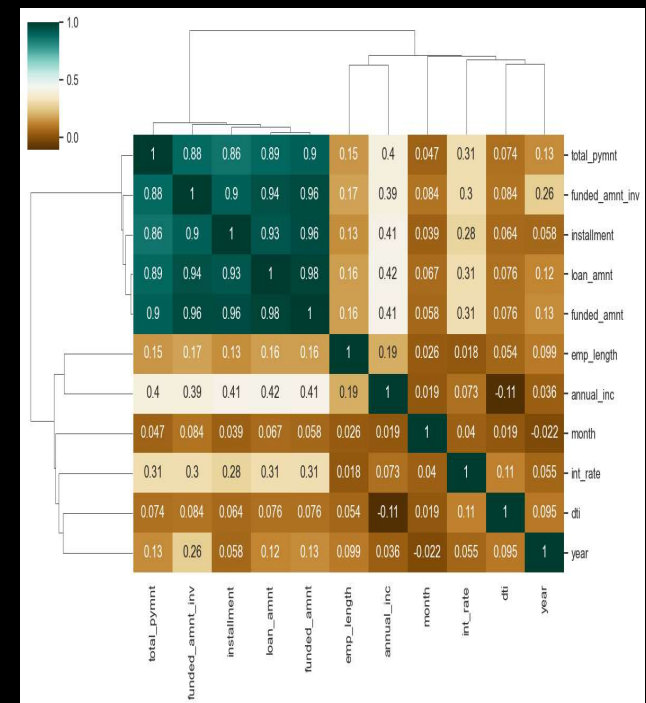
BIVARIATE ANALYSIS

- Observations:

- Income range 80000+ has less chances of charged off.

- Income range 0-20000 has high chances of charged off.

- Notice that with increase in annual income charged off proportion got decreased.
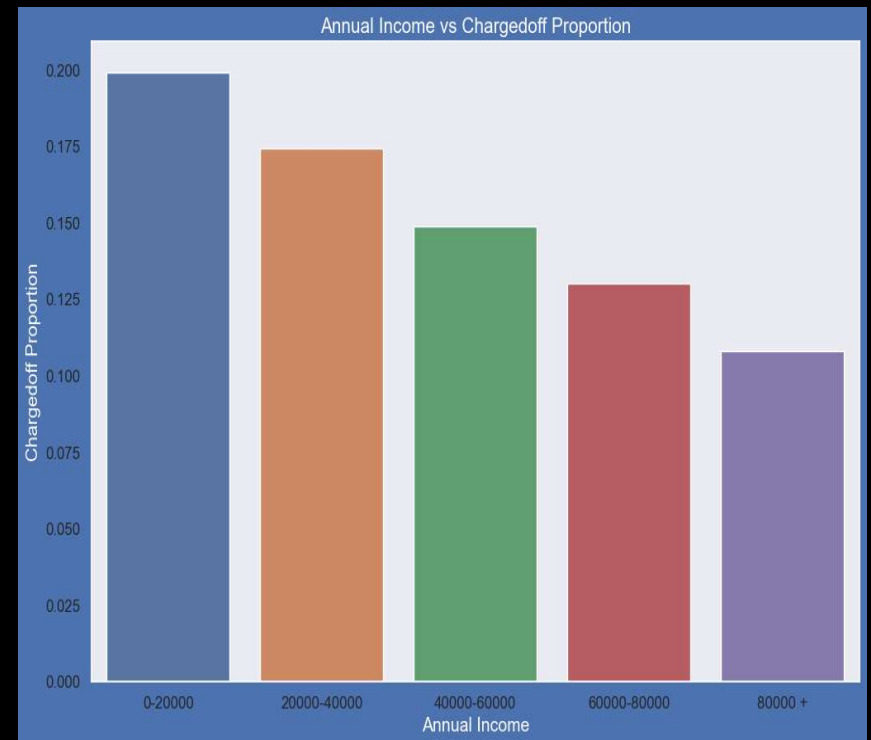
BIVARIATE ANALYSIS

- Analysis on purpose of Loan against Chargedoff_Proportion.

- Small Business applicants have high chances of getting charged off.

- Rrenewable_energy where charged off proportion is better as compare to other categories
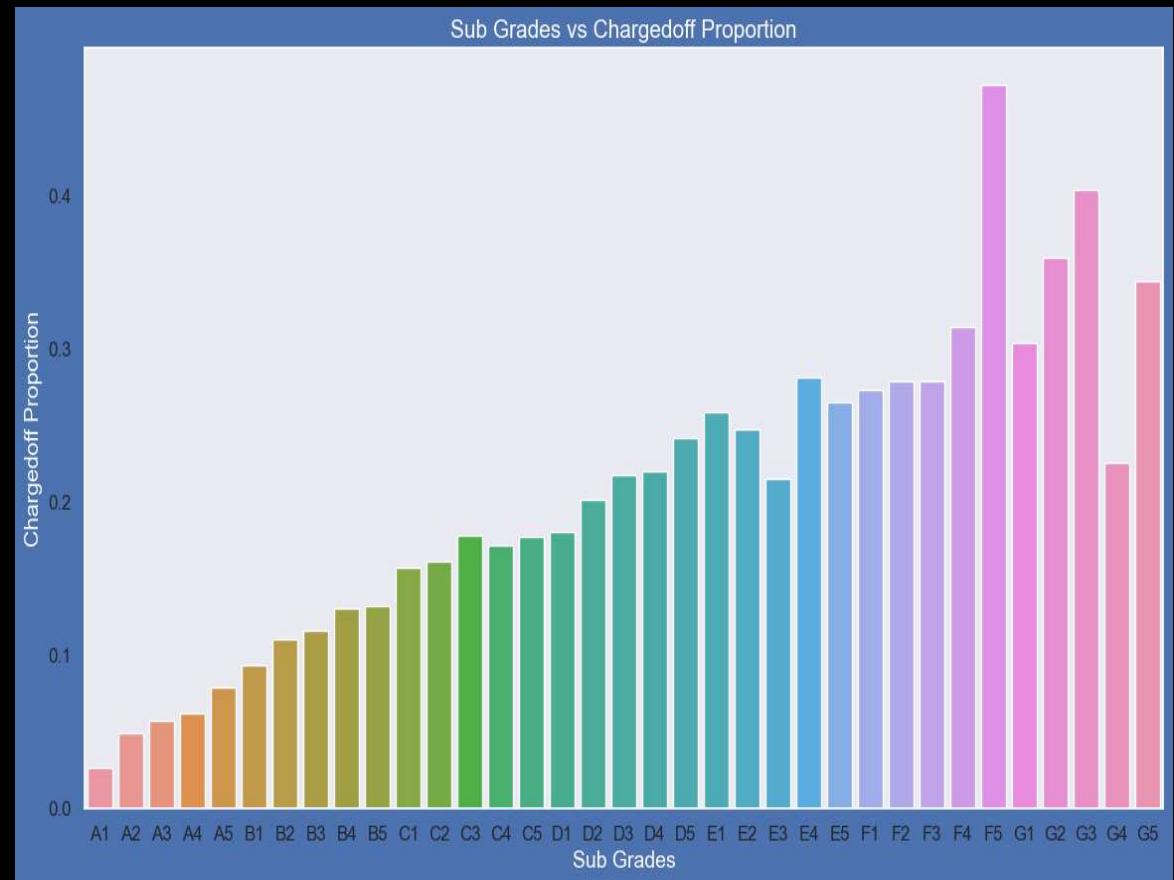

Purpose of Loans vs Chargedoff Proportion
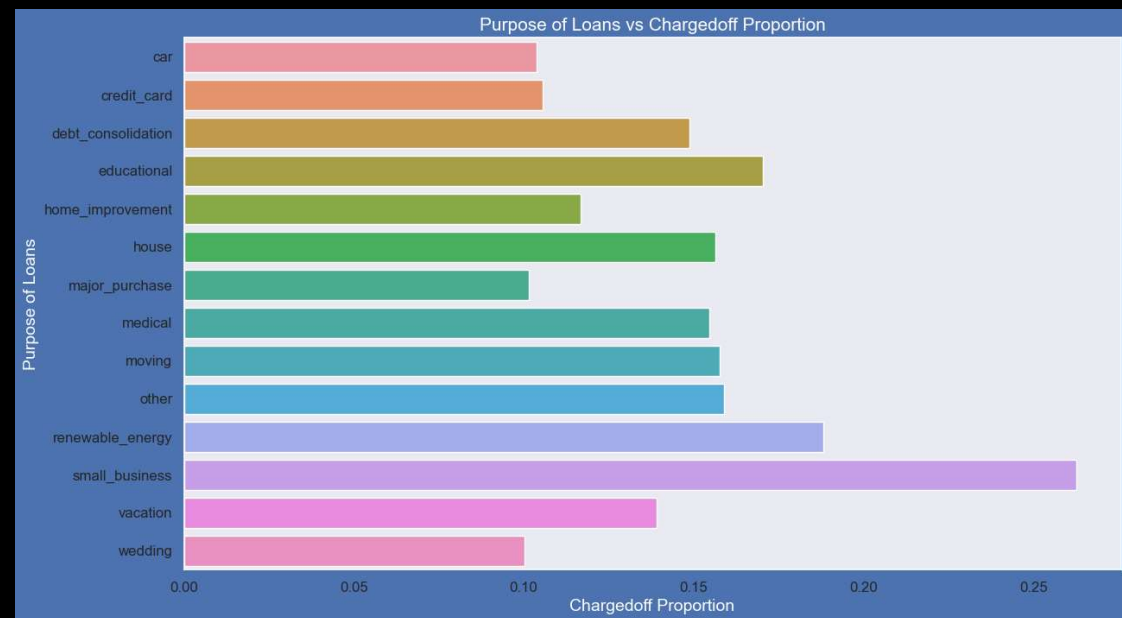
BIVARIATE ANALYSIS

- Observations:
- Grade "A" has very less chances of charged off.
- Grade "F" and "G" have very high chances of charged off.
- Chances of charged off is increasing with grade moving from "A" towards "G"

Grades vs Chargedoff Proportion
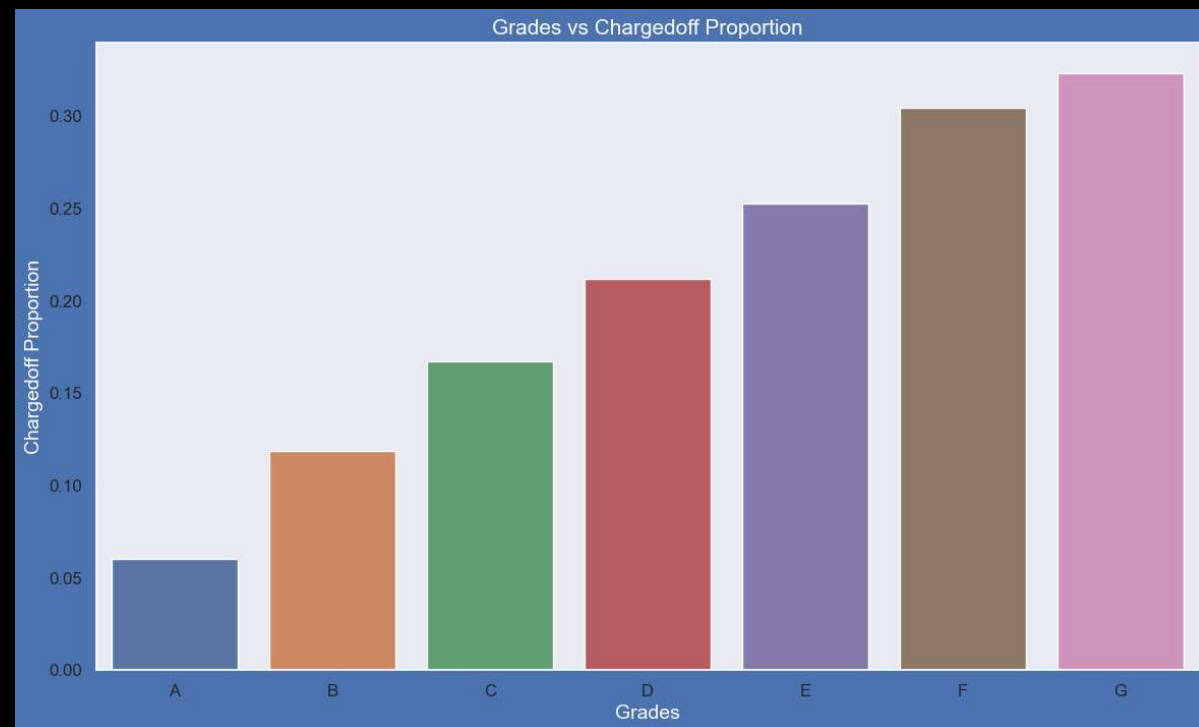
BIVARIATE ANALYSIS

- Sub Grades of "A" has very less chances of charged off.

- # sub Grades of "F" and "G" have very high chances of charged off.

- # proportion of charged off is increasing with sub grades moving from sub grades of "A" towards sub grades of "G"


Sub Grades vs Chargedoff Proportion

BIVARIATE ANALYSIS

- Interest rate less than 10% has very less chances of charged off. Interest rates are starting from minimum 5 %.

- Iinterest rate more than 16% has good chances of charged off as compared to other category interest rates.

- Charged off proportion is increasing with higher interest rates.

Interest Rate vs Chargedoff Proportion
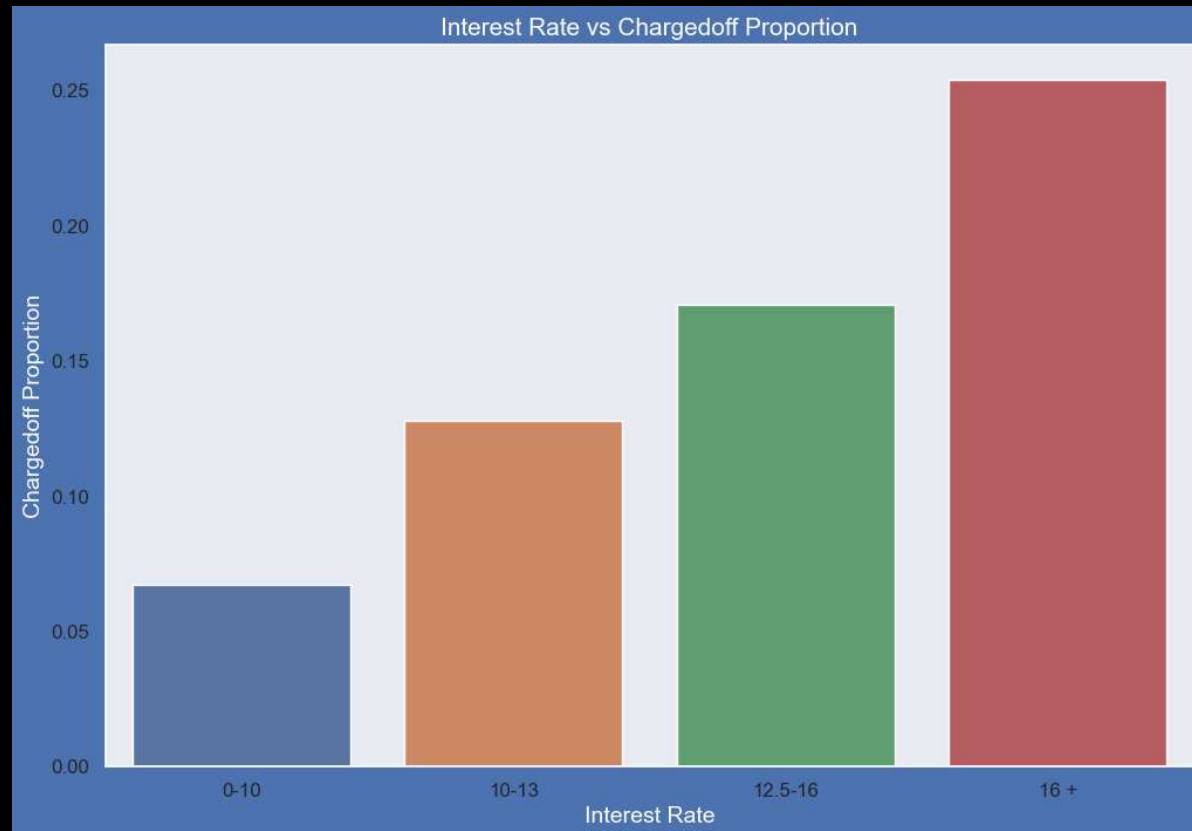
BIVARIATE ANALYSIS

- Those who are not working or have less than 1 year of work experience have high chances of getting charged off.

- It makes sense as with less or no experience they don't have source of income to repay loan.

- Rest of the applicants have more or less same chances of getting charged off.

Employment Length vs Chargedoff Proportion
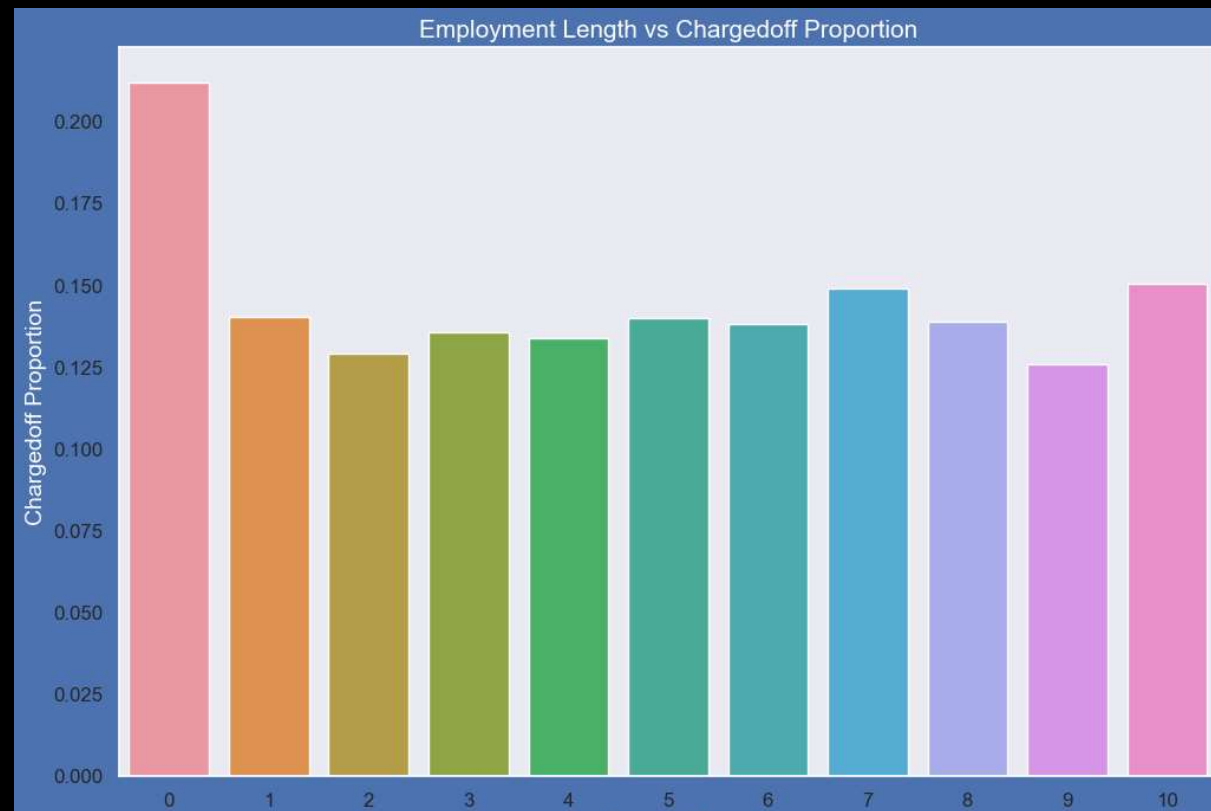
BIVARIATE ANALYSIS

- Observations:
- States NE has very high chances of charged off but number of applications are too low to make any decisions.
- NV,CA and FL states shows good number of charged offs in good number of applications.

Address State vs Chargedoff Proportion
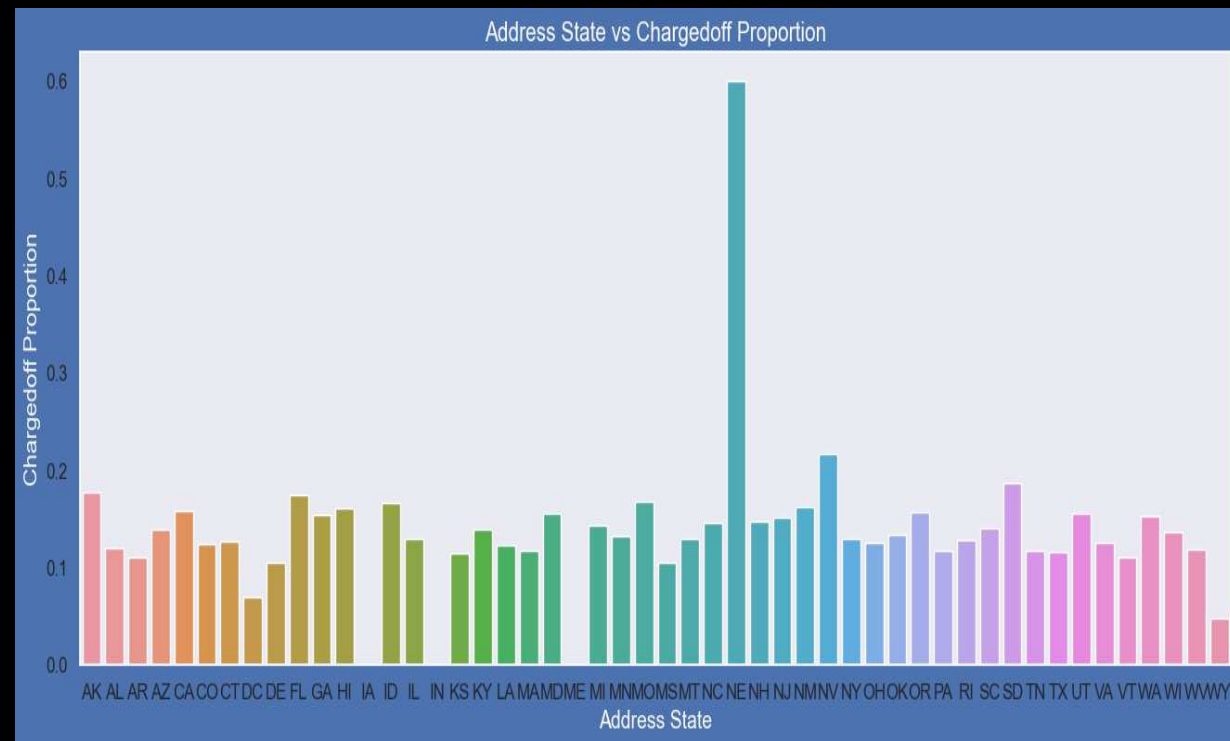
BIVARIATE ANALYSIS

- Observations:
- There is not much difference in charged off proportion.
- This variable doesn't provide any insights for charged off.
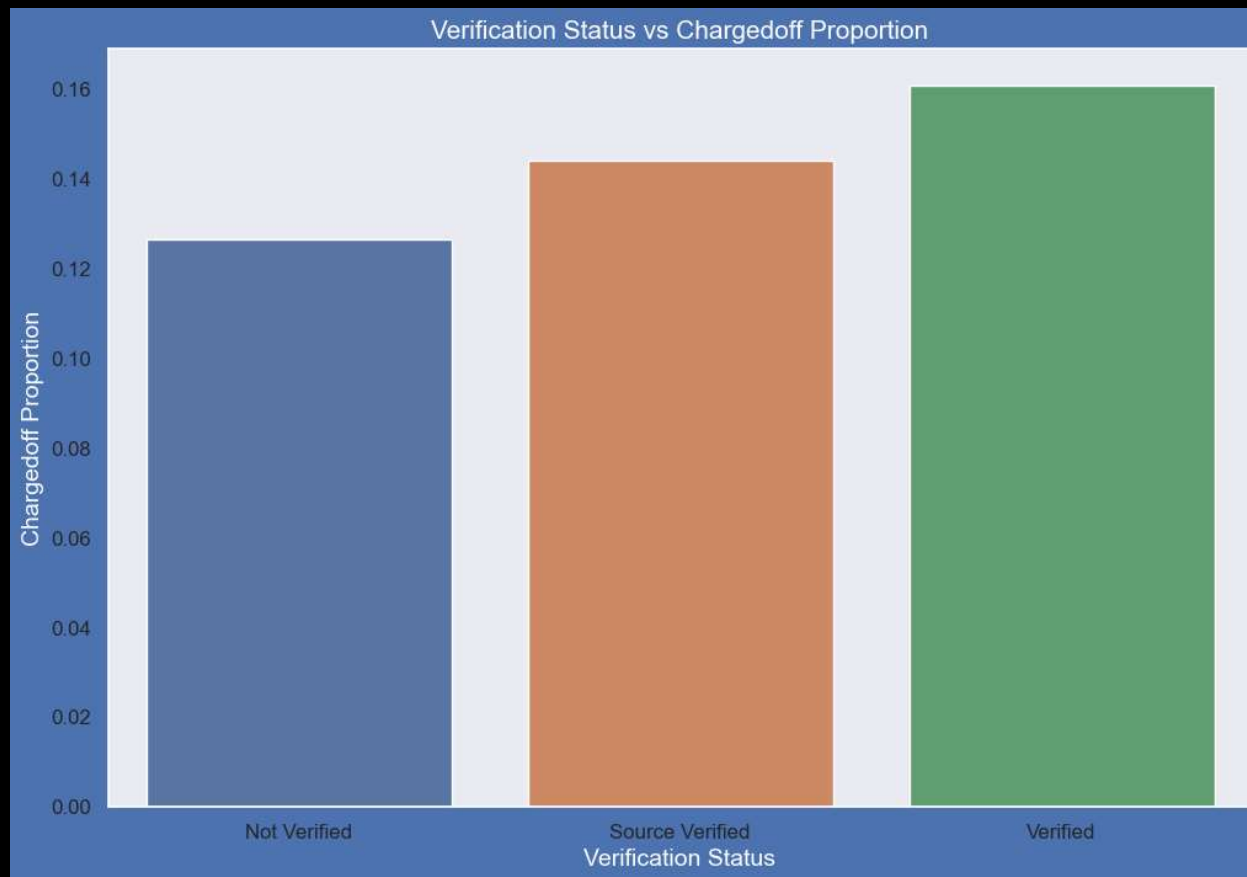
BIVARIATE ANALYSIS

- Observations:
- Those who already have pub_rec_bankruptcies value 1, have charged off proportion higher than who have no pub_rec_bankruptcies.
- pub_rec_bankruptcies count 2 has even higher charged off proportion but those numbers are not significant to decide.
- Not known is the column for which we don't have any information about borrower.
- This also makes sense that who has defaulted before has more chances of defaulting in future as well.
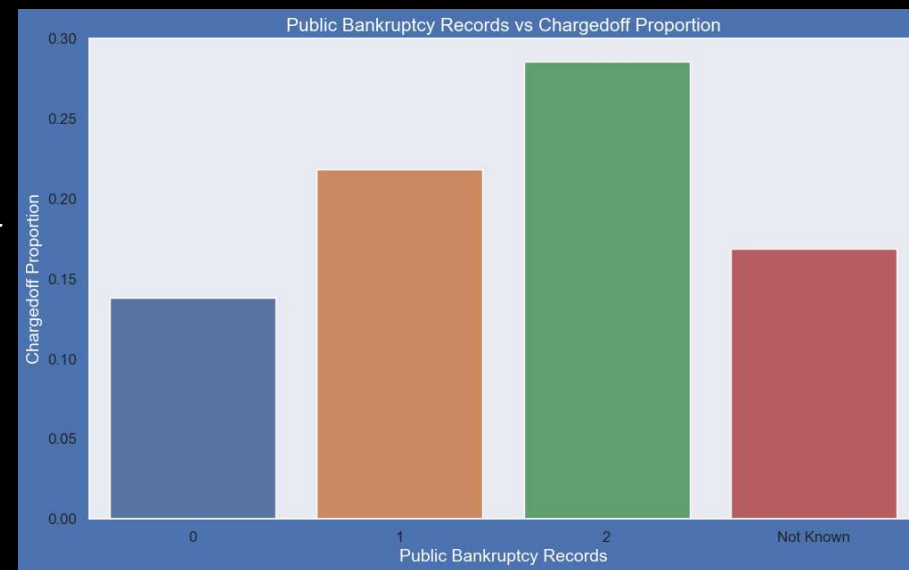
BIVARIATE ANALYSIS

- Observations:
- A derogatory item is an entry that may be considered negative by lenders because it indicates risk and hurts
- your ability to qualify for credit or other services. Public records and collections are derogatory items because they reflect financial obligations that were not paid as agreed.
- Those who already have pub_rec value 1 or 2 have charged off chances higher than who have no Derogatory Public Record.
- pub_rec count 3-4 has less numbers so cannot reach on any conclusions.


Derogatory Public Record vs Chargedoff Proportion
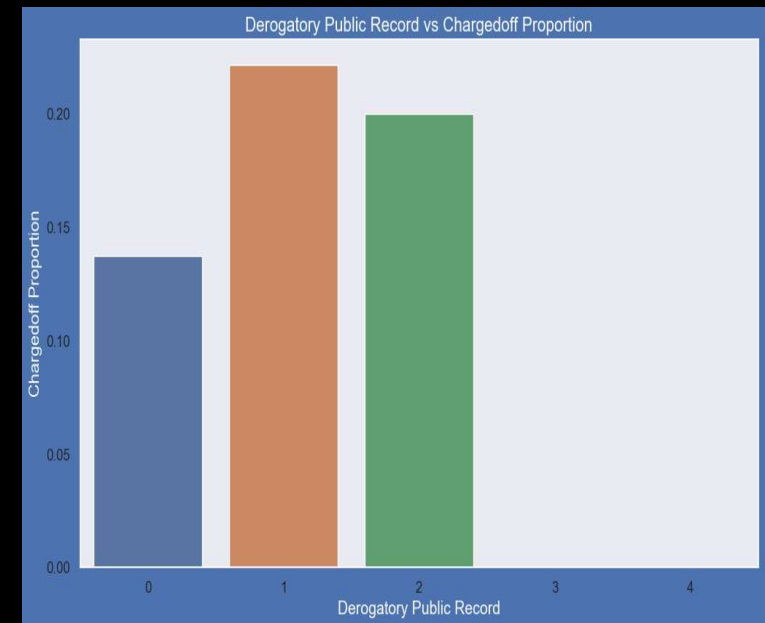
BIVARIATE ANALYSIS

- Observations:
- Median,95th percentile,75th percentile of loan amount is highest for loan taken for small business purpose among all purposes.
- Debt consolidation is second and Credit card comes 3rd.


Purpose of Loan Vs. Loan Amount
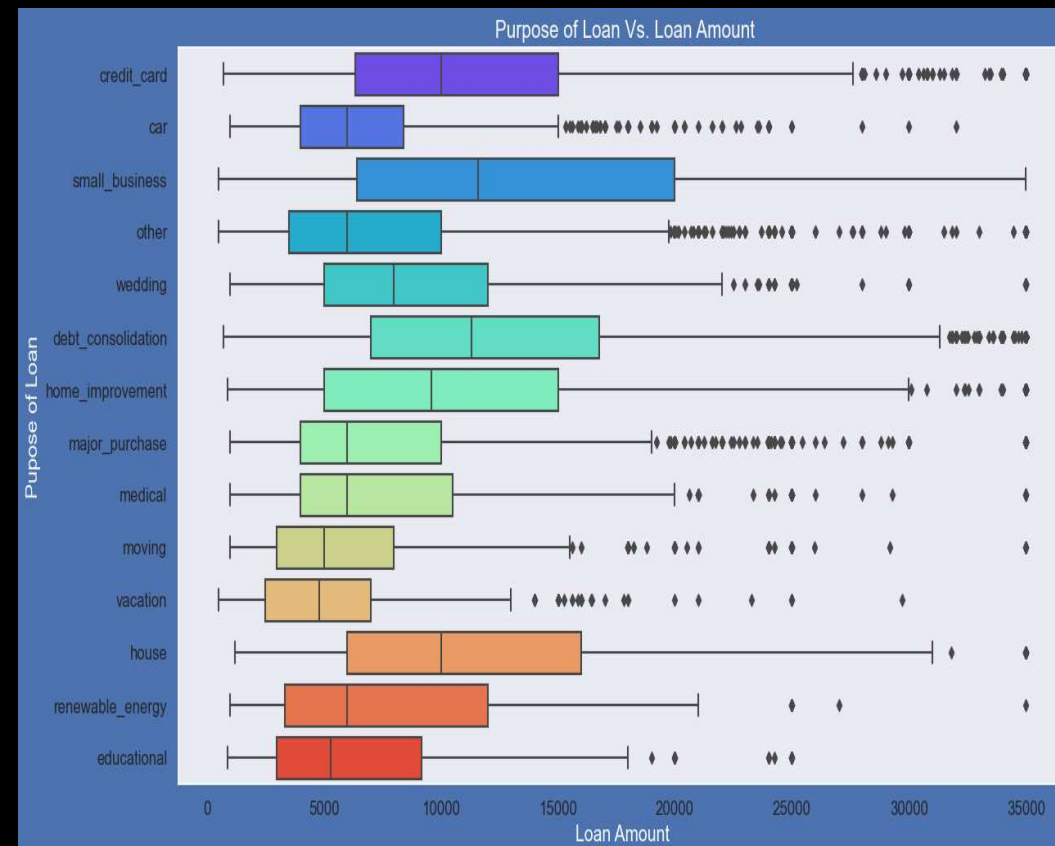
BIVARIATE ANALYSIS

- Observations:

- Loan taken for small business purpose, Debt consolidation and Credit card are somewhat evenly distributed as compare to loan taken for other purposes


Purpose of Loan vs Loan Amount
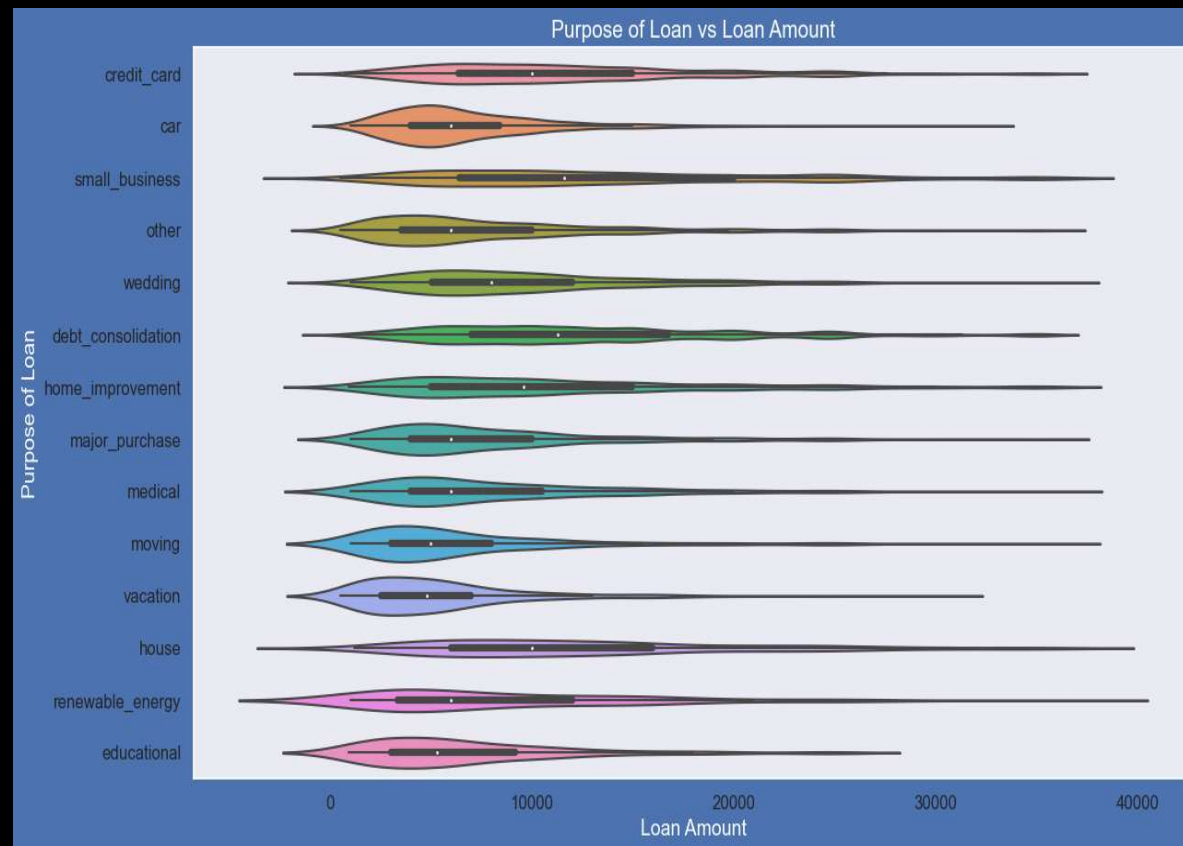
BIVARIATE ANALYSIS

- Observations:

- It is clear that average interest rate is higher for 60 months loan term.

- Most of the loans issued for longer term had higher interest rates for repayment.


Term of Loan vs Interest Rate
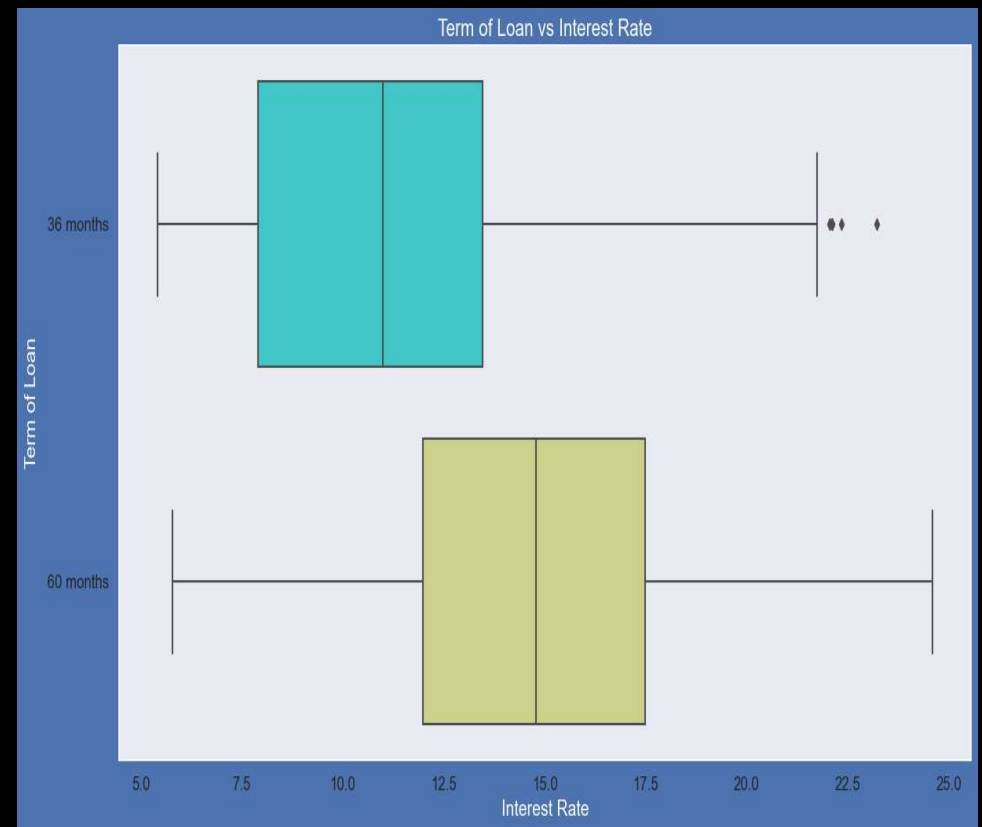
BIVARIATE ANALYSIS

- Observations:
- It is clear that average interest rate is highest for small business purpose.
- Loans taken for small business purposes had to repay the loan with more interest rate as compared to other.
- Debt consolidation is 2nd where borrowers had to pay more interest rate.
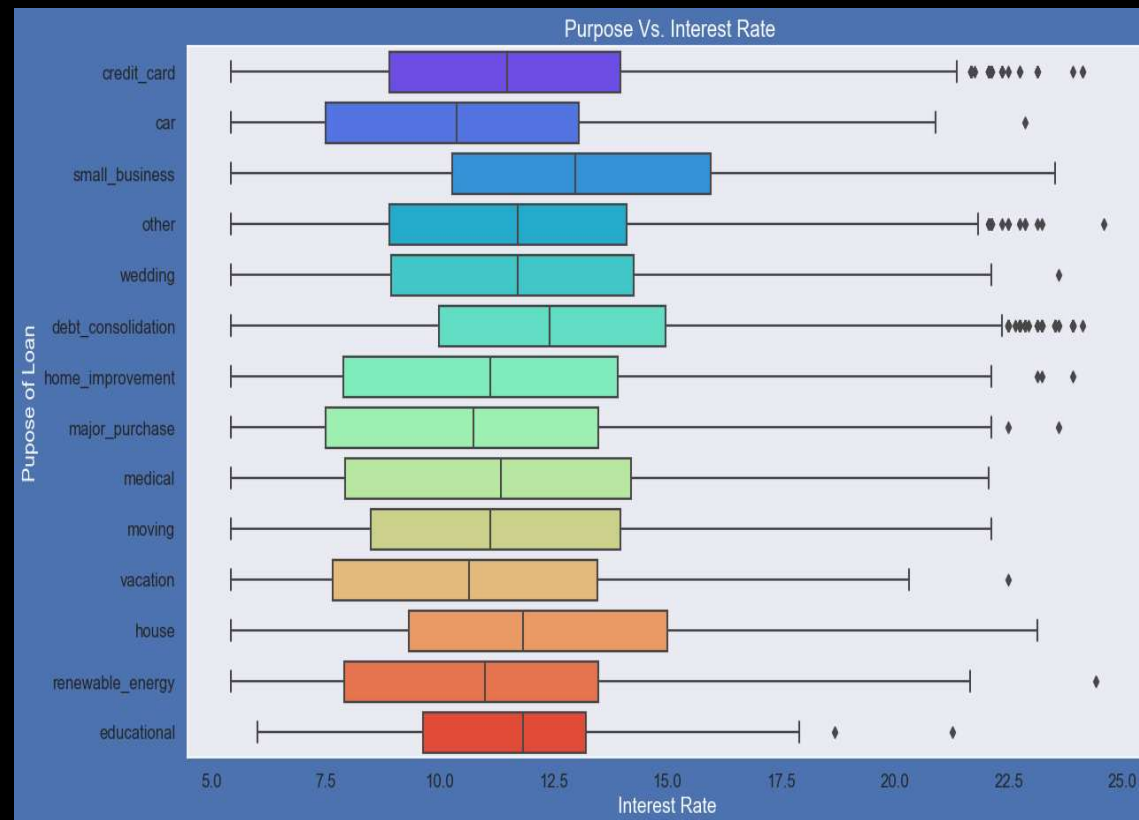
BIVARIATE ANALYSIS

- Observations:
- A-grade is a top letter grade for a lender to assign to a borrower.
- The higher the borrower's credit grade, the lower the interest rate offered to that borrower on a loan.
- It is clear that interest rate is increasing with grades moving from A to G.


Grade Vs. Interest Rate
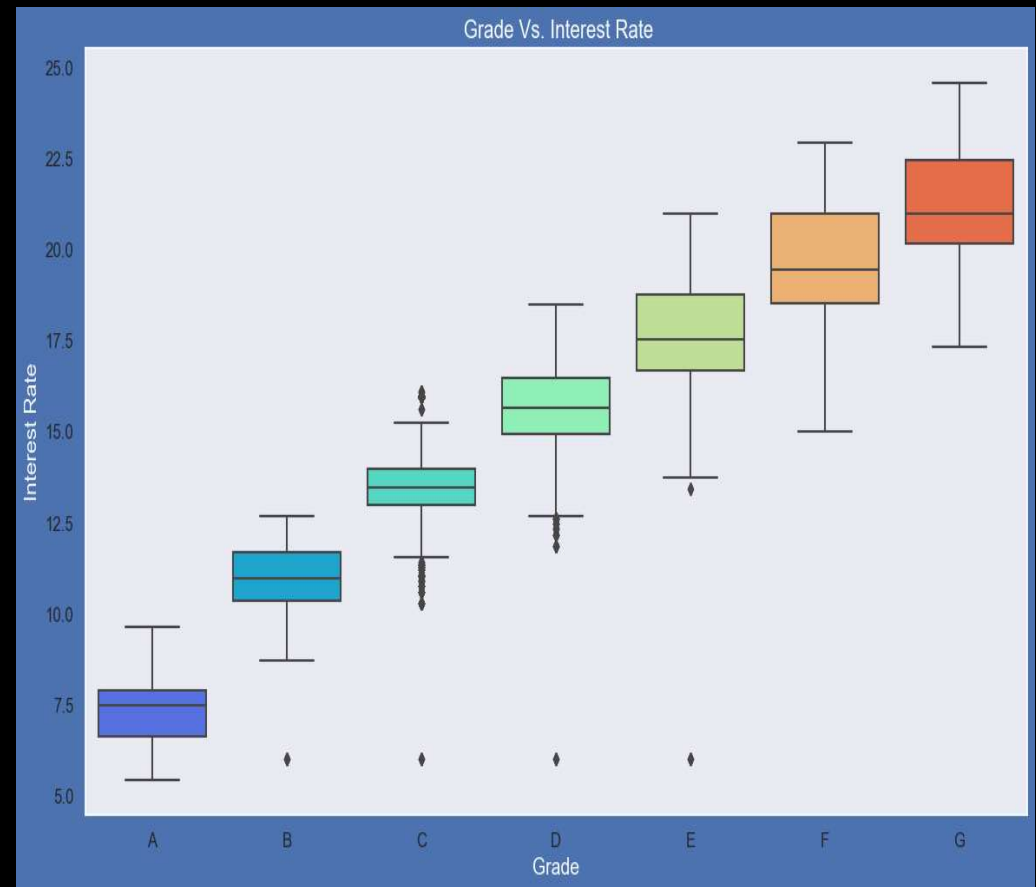
BIVARIATE ANALYSIS

- Observations:
- It is clear that interest rate is increasing with loan amount increase.
- probably when loan amount is more it is taken for longer loan term, we saw earlier that longer the loan term more the interest rate.

Loan Amount Vs. Interest Rate
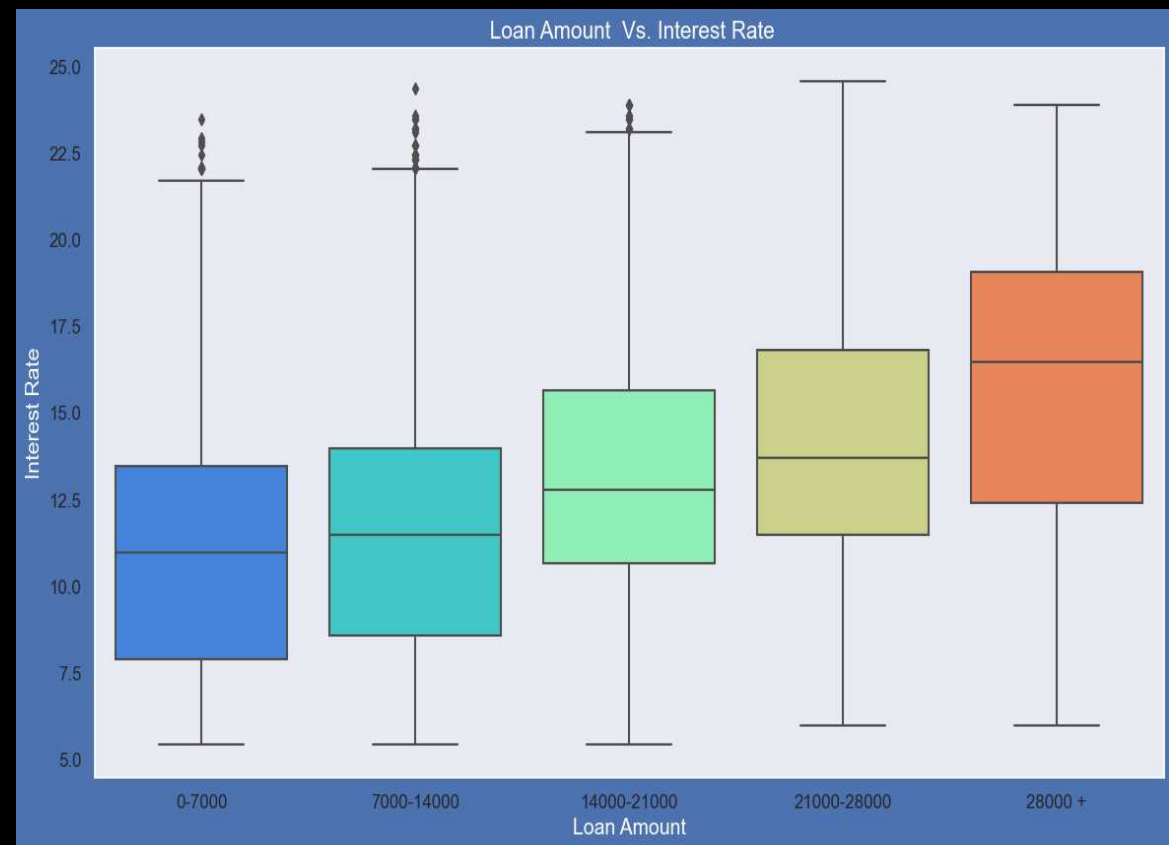
BIVARIATE ANALYSIS

- Observations:

- If your DTI is low enough you may get a lower interest rate.

- Plot shows no significant variation but there is slight increase in interest rate with increase in DTI.


Debt To Income Ratio Vs. Interest Rate
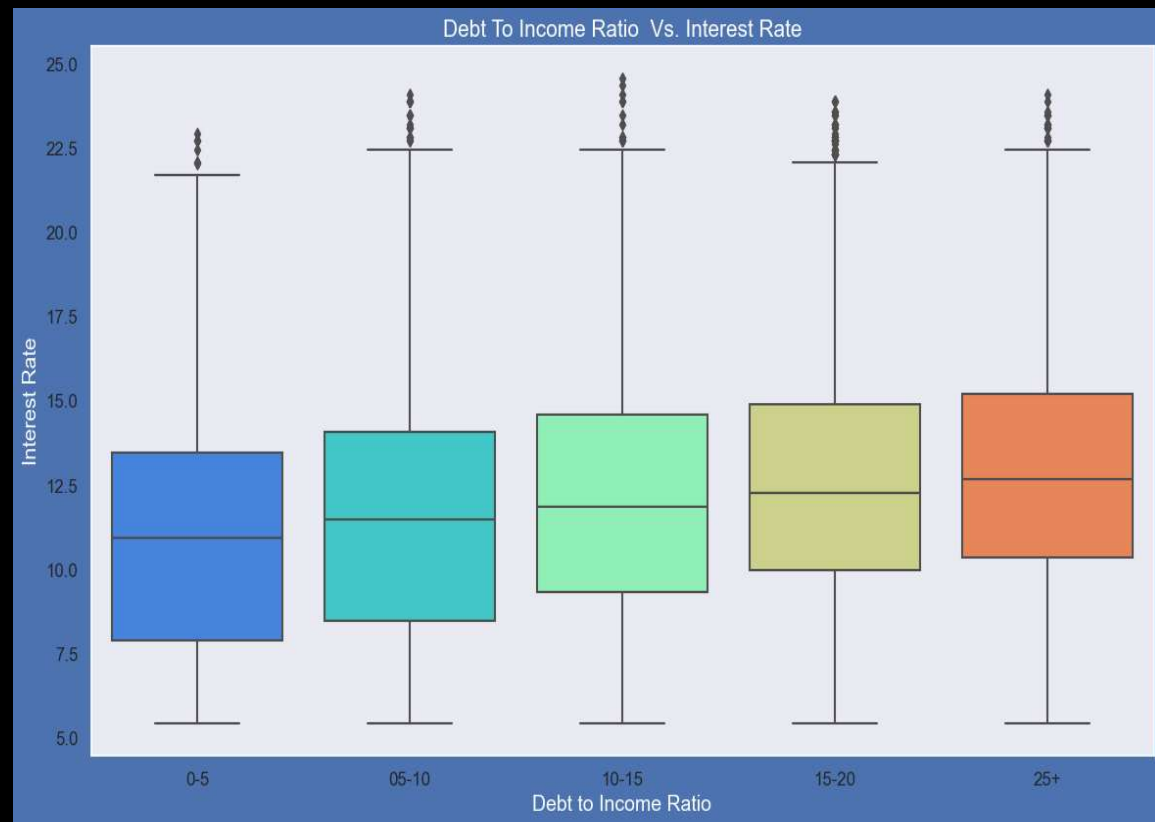
BIVARIATE ANALYSIS

- Observations:
- Higher percentage of loan amount is recovered when annual income is high.
- Plot shows no significant variation but there is slight increase in recovery percentage with increase in annual income.

Percentage of Loan amount Recovered   Vs. Annual Income
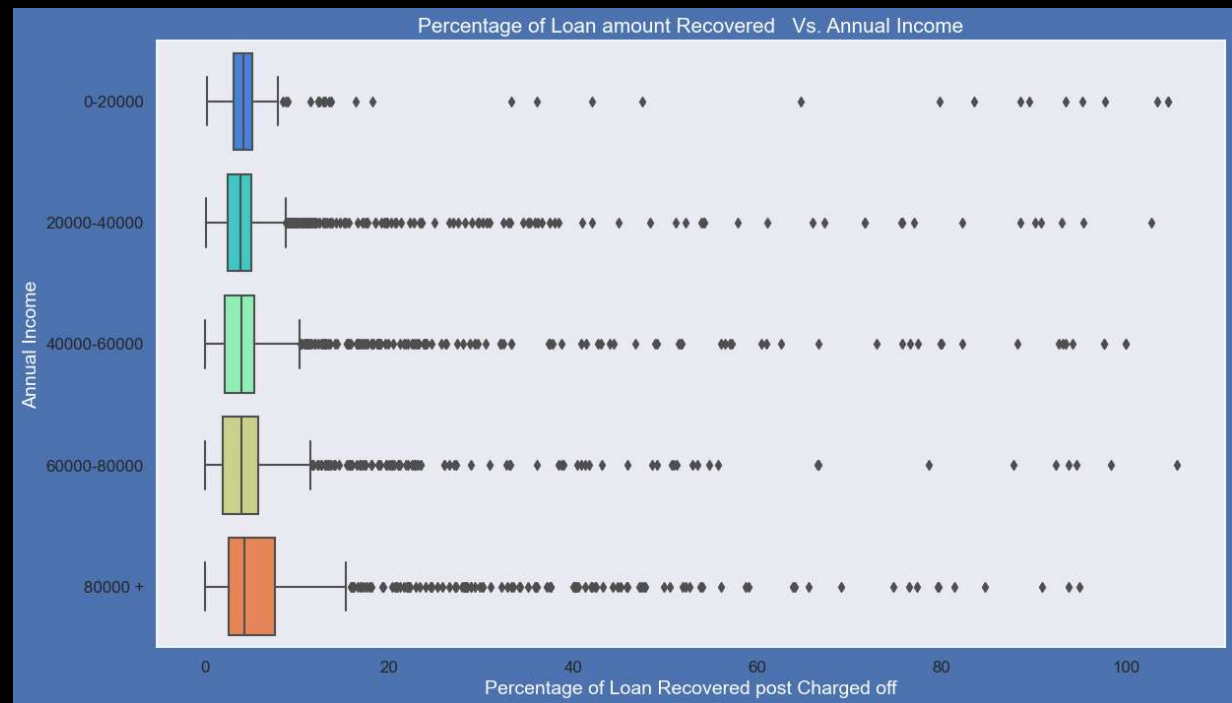
BIVARIATE ANALYSIS

- Observations:

- From this we can conclude that the ones getting 'charged off' have lower annual incomes than the ones who are paid fully' for each and every grade (i.e. at same interest range)
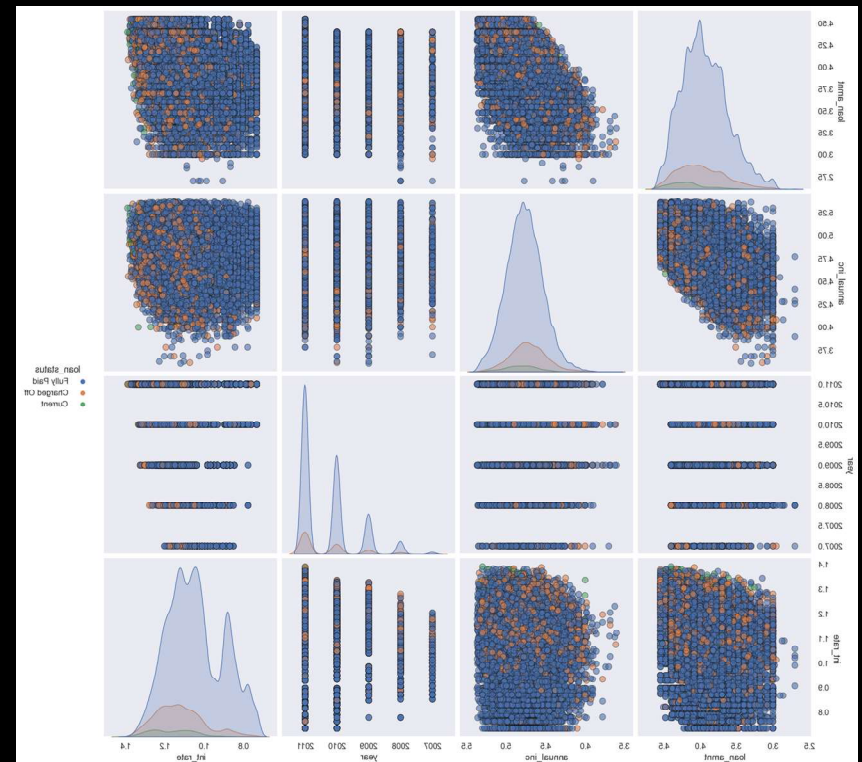

Grade vs Annual income
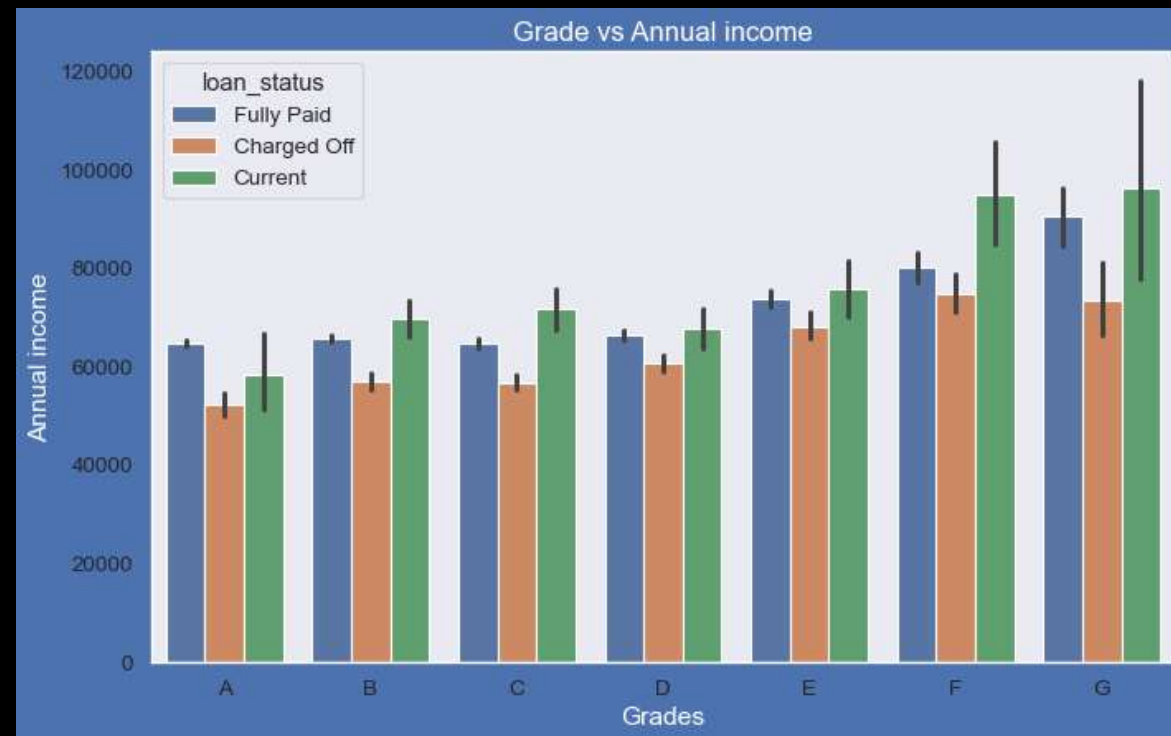
BIVARIATE ANALYSIS

- Observations:
- Plot shows interest rate is increasing slowly with increase in year.
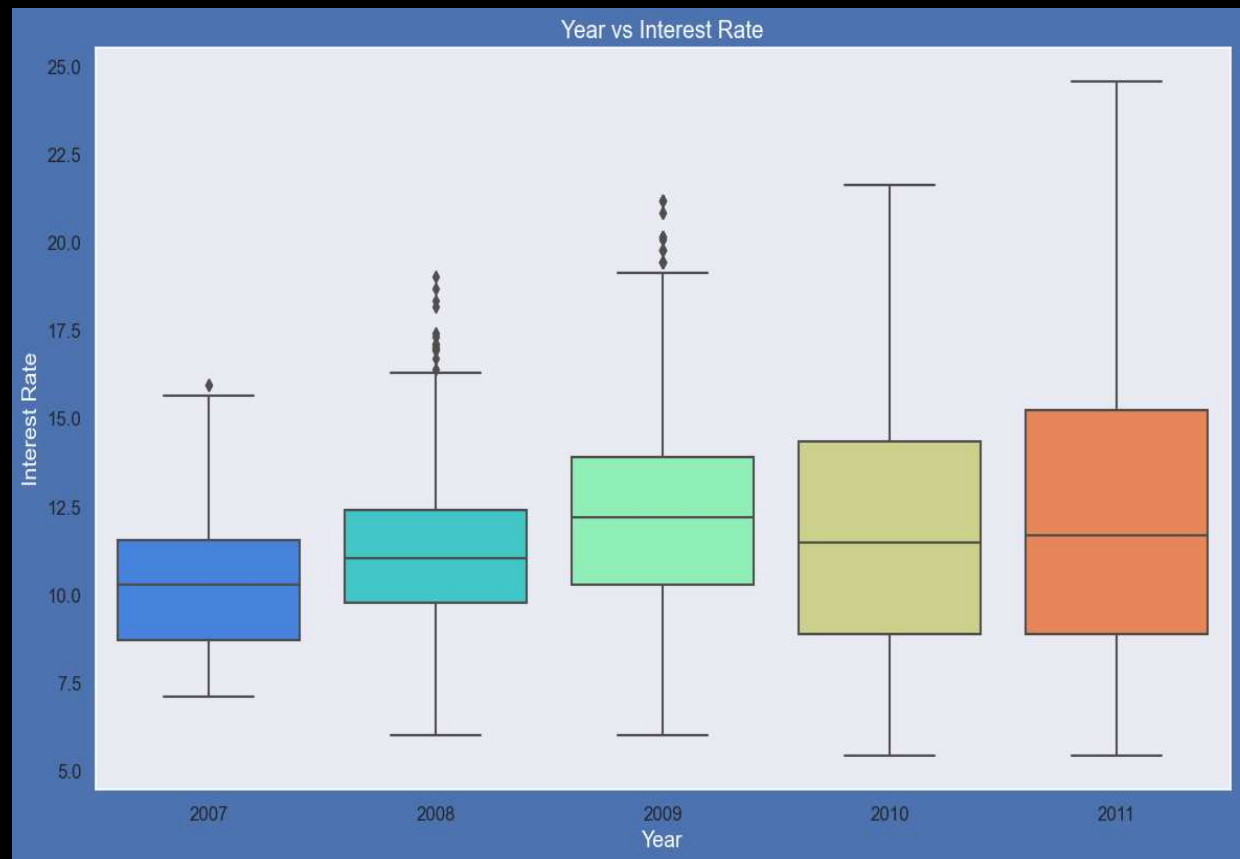
Year vs Interest Rate

MULTIVARIATE ANALYSIS

- Observation:
- Higher the interest rate higher charged off ratio
- Higher the annual income higher the loan amount slightly.
- increase in number of charged off with increase in year.
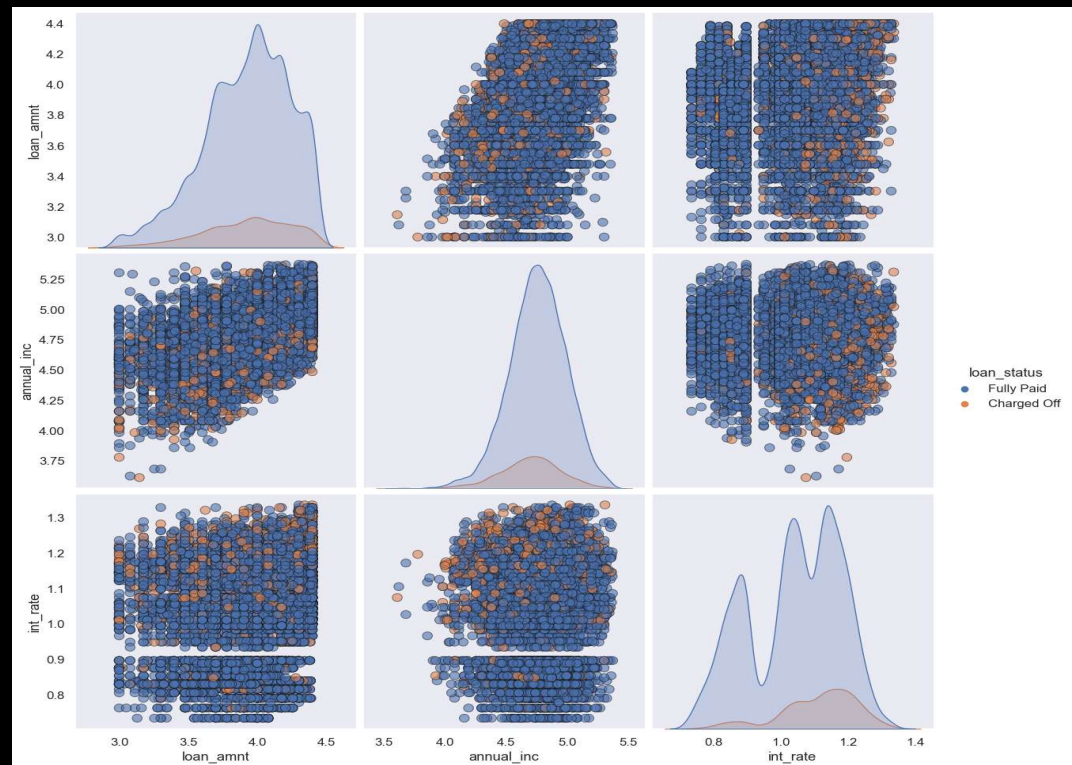- interest rate is increasing with loan amount increase

MULTIVARIATE ANALYSIS

- Observation:
- Higher the interest rate higher charged off ratio
- Higher the annual income higher the loan amount slightly
- Interest rate is increasing with loan amount increase this results in high charged off.

CONCLUSION

- Based on the analysis of the data, some possible recommendations are:

- Lower interest rates for higher loan amounts: The data shows that most of the loan amounts are in the range of 5000 - 15000, and the interest rates increase sharply after the 75th percentile. This may discourage borrowers from taking larger loans or make them default on their payments. Lowering the interest rates for higher loan amounts may attract more customers and reduce the risk of default.

- Focus on credit grade and employment status: The data shows that the credit grade and the employment status are important factors that affect the loan status. Borrowers with higher credit grades and full-time employment are more likely to pay back their loans. Focusing on these segments may improve the quality and profitability of the loans.

- Expand the loan purpose categories: The data shows that the loan purpose is a categorical variable with 14 values, but most of the loans are for debt consolidation, credit card, or other purposes. Expanding the loan purpose categories may help to understand the needs and preferences of the borrowers better and offer more tailored products and services.