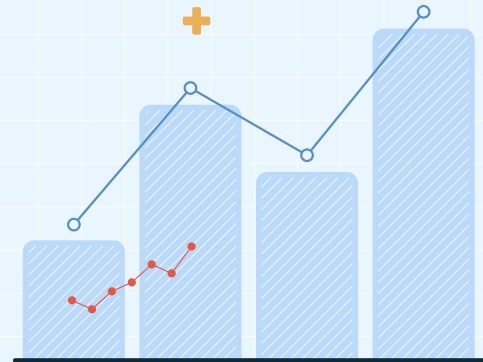
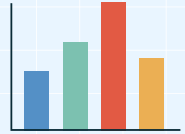
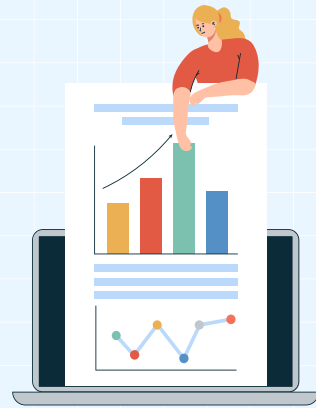
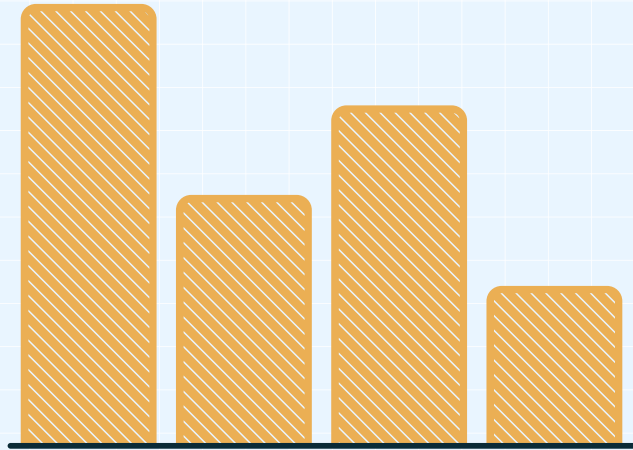


Lending Club Case Study

- Om Modi & Maneesh Darisi





“Quick to Borrow,
Is Always Slow to Pay .”

—German Proverb



Table of contents



01

Problem Statement

This tells about the problem statement for case study

02

Approach

This tells about the data driven approach taken

03

Univariate Analysis

This describes the unique characteristics of every individual feature



04

Bivariate Analysis

This describes the impact of one feature over other

05

Multivariate Analysis

This describes the influence of multiple features on other

06

Conclusions

This describes the inferences of analysis with recommendations



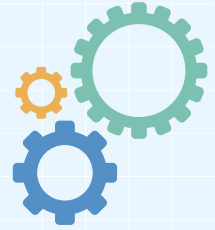
01

Problem Statement





Statement

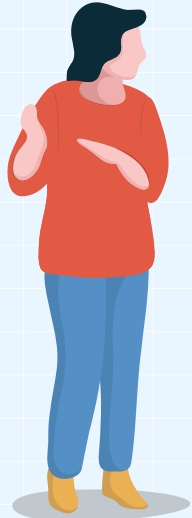


When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile

Objective –

The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed causing huge loss to the lenders.

Objective is to **identify the risky loan applicants** at the time of loan application so that such loans can be reduced thereby cutting down the amount of credit loss To the lending organization.



Approach

Data Cleaning

- Null Value Removal
- Outlier treatment
- Unique Value Removal
- Single Value Columns
- Domain Based Removal

Transformation

- Data Type Transformation
- Derived Columns
- Null Value impute using median/mode

Univariate Analysis

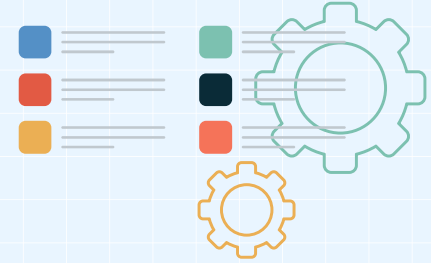
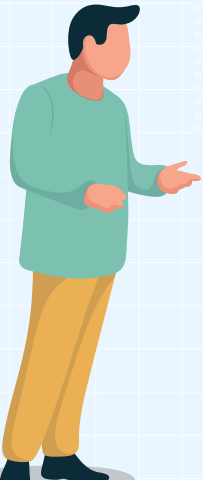
- Histogram + KDE
- Box Plot
- Count Plot
- Pie Plot

Bivariate Analysis

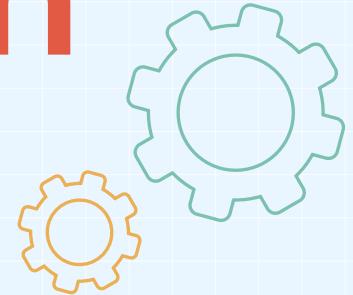
- Bar Plot (hue)
- Box Plot (hue)

Multivariate Analysis

Pair Plot
Heat Map



Data Cleaning And Transformation

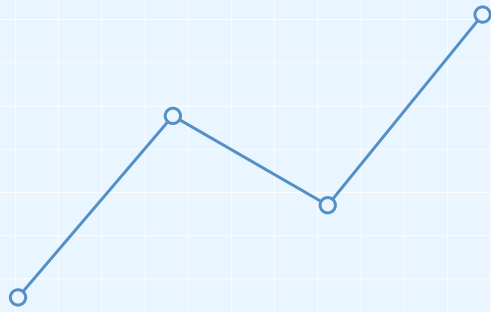




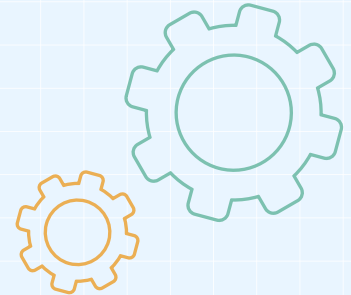
Data Cleaning & Transformation

- **Threshold for Null Values:** Rows with more than 70% null values were dropped, resulting in the elimination of approximately 55 columns.
- **Threshold for NA Values:** Columns with more than 50% NA values were dropped, leading to the removal of one column.
- **Unique Values:** Columns with only one unique value or all unique values were dropped, resulting in the removal of 9 and 3 columns, respectively.
- **Elimination of Loan Completion Attributes:** 21 columns related to attributes describing loans after completion were removed.
- **Removal of Loan and Customer Titles:** Columns related to loan and customer titles were removed.
- **Elimination of Zip Code:** Due to a large number of unique values causing potential skewness, the zip code column was removed.
- **Imputation of Missing Values:** Columns with at least 50% values were imputed using median and mode.
- **Outlier Detection:** Outliers were detected using box plots and eliminated using a threshold of mean plus three times the standard deviation for numerical columns.
- **Removal of Current Loan Status Rows:** Rows with current loan status were removed as they were not relevant to the analysis.
- **Data Type Standardization:** Data types of columns were adjusted, trimming additional characters and standardizing them to integers and floats.



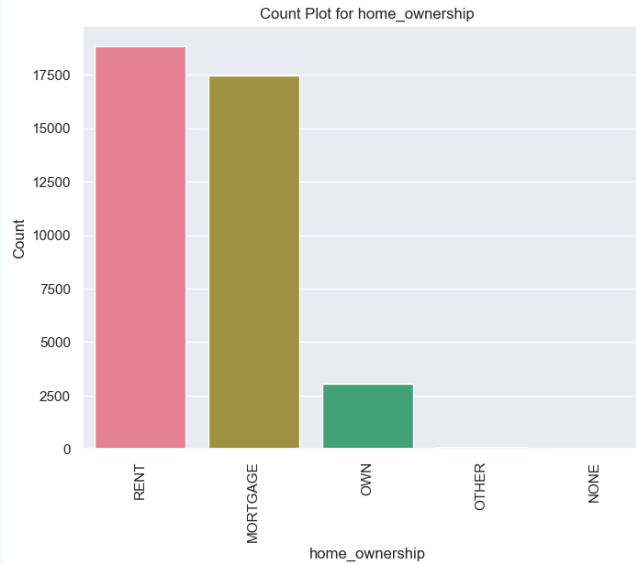


Univariate Analysis



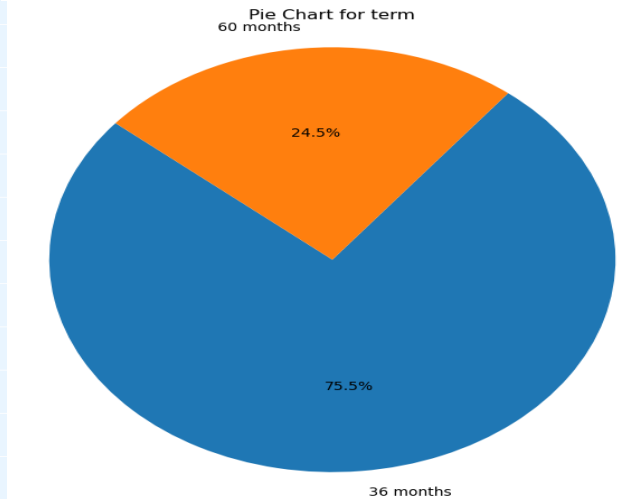
Home Ownership

- Most of the loan applicants are either living on Rent or on Mortgage, Very few have their own house



Term

- Majority of loans are for 36 months (75.5%) and 60 months (24.5%)



State

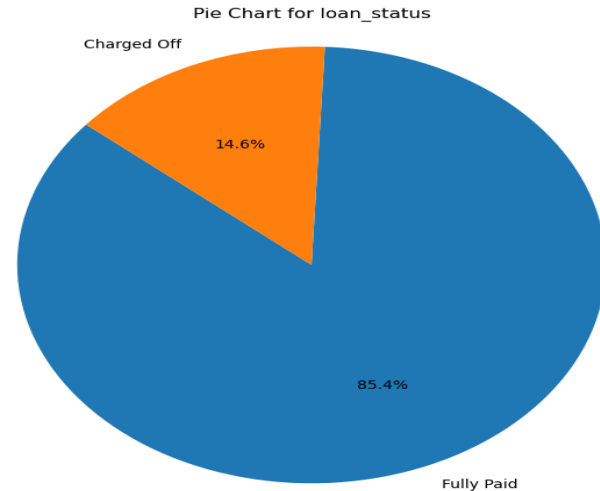
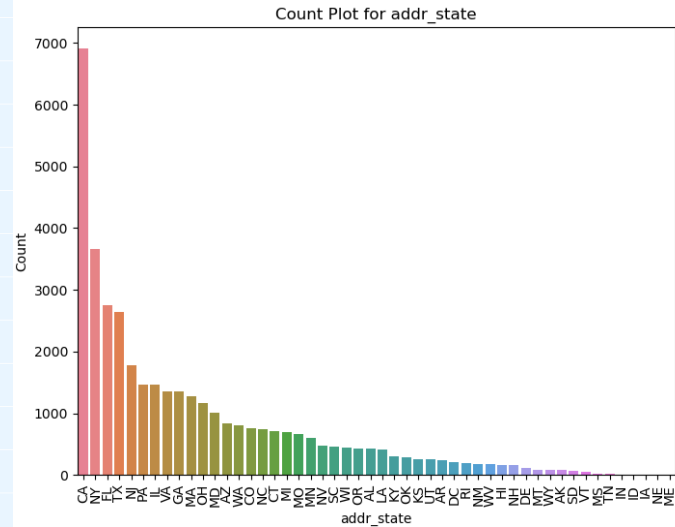


- CA and NY account for most of the loan applications

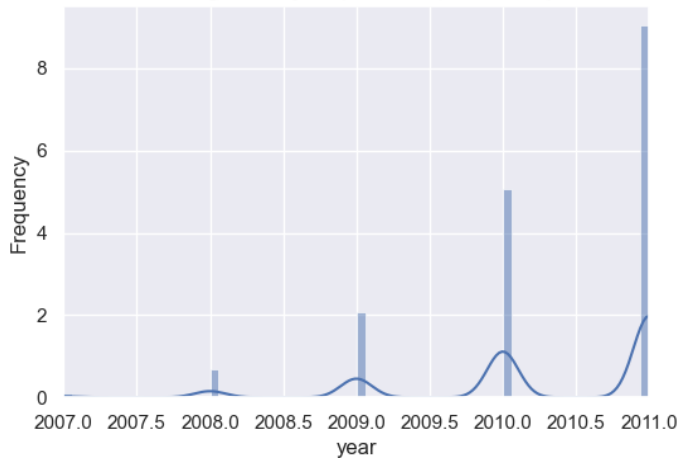


Loan Status

- Majority of loans are for 36 months(75.5%) and 60 months(24.5%)



Histogram for year (Min: 2007, Max: 2011)



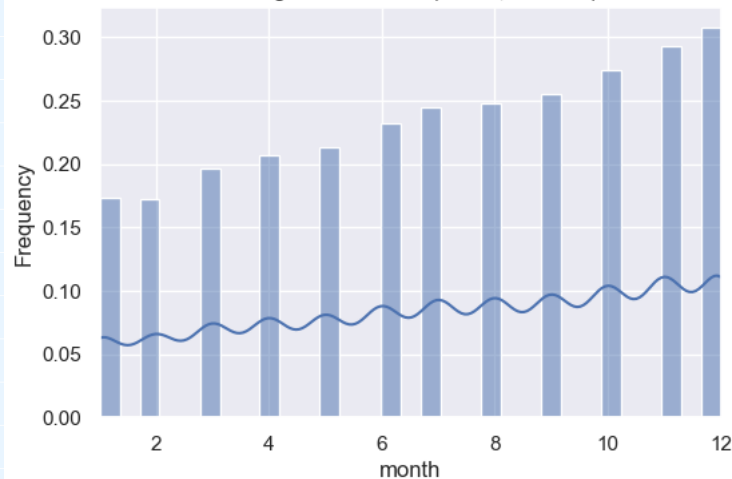
Year

- We can notice year on year growth in number of loans

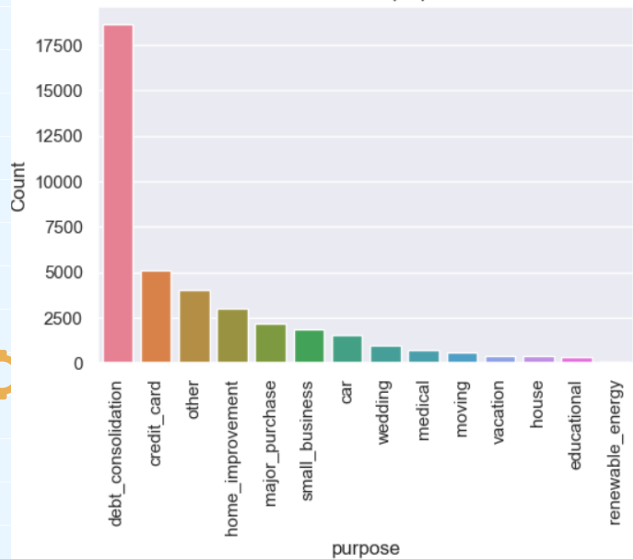
Month

- Based on the month there is month on month growth in the loan approvals.
- Q4 has the highest number of loan application.

Histogram for month (Min: 1, Max: 12)

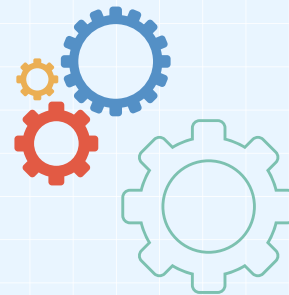


Count Plot for purpose



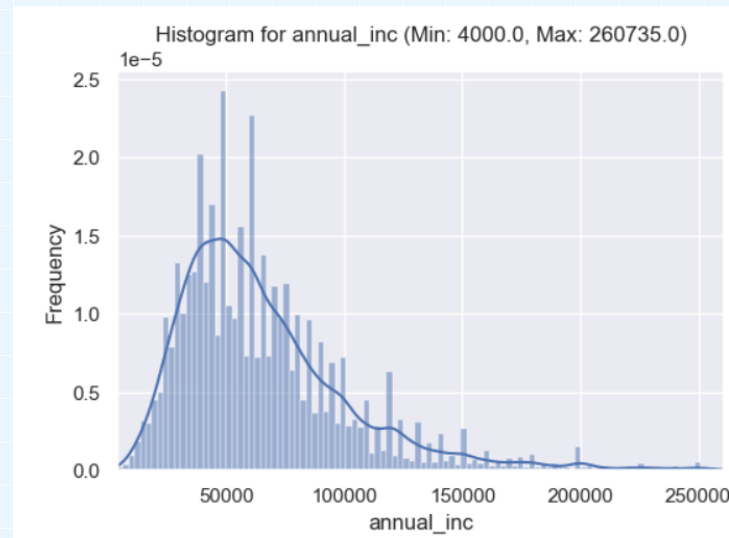
Purpose

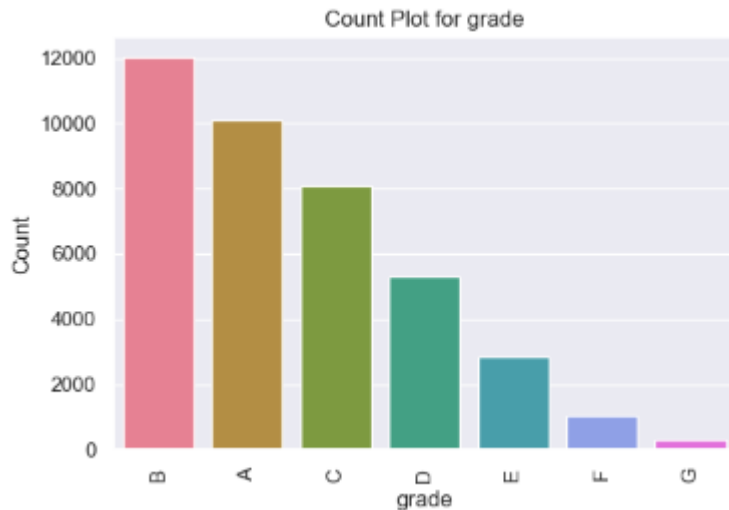
- Majority of Applicants have taken loan for debt consolidation and credit card



Annual Income

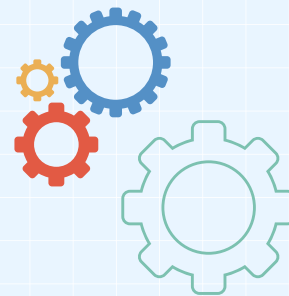
- Majority of annual income is between 50k-150k.





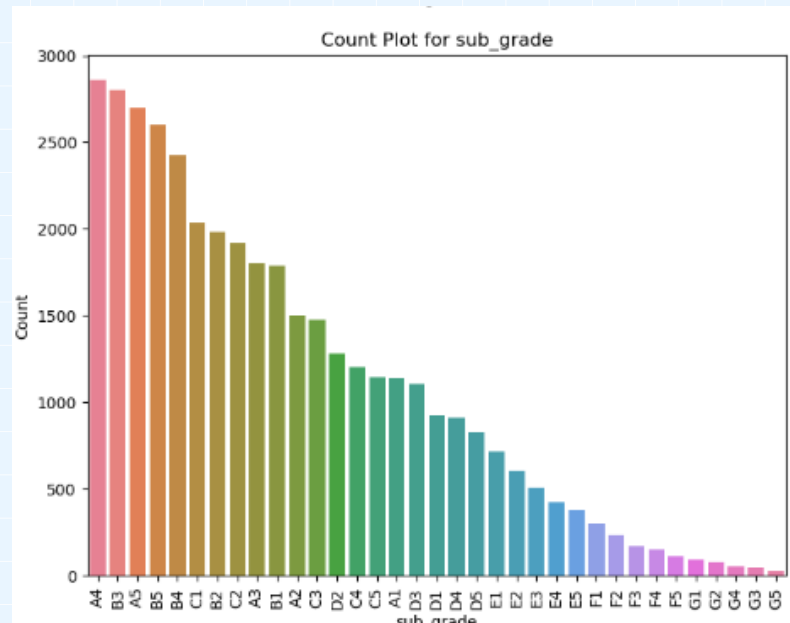
Grade

- A,B,C,D accounts for 90% of the loan applicants compared to E,F,G

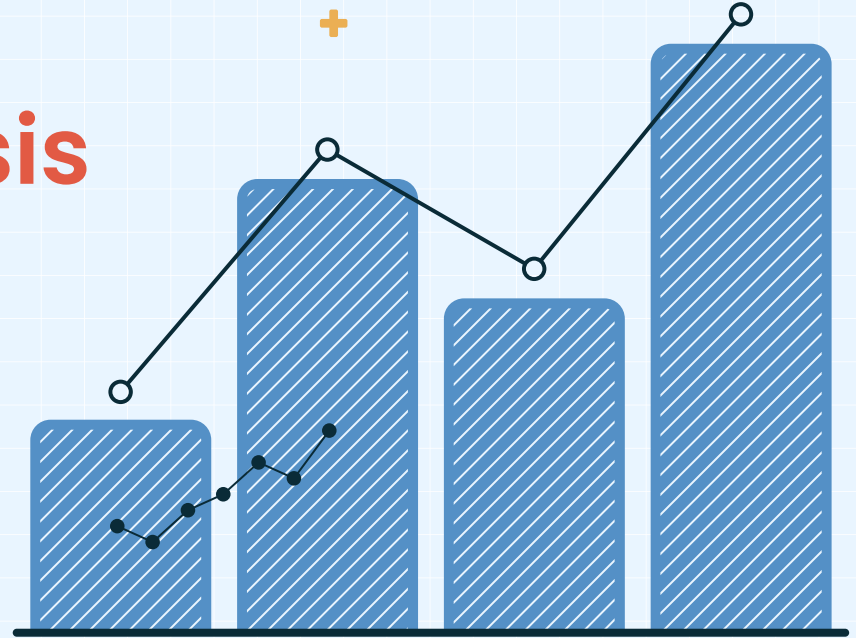


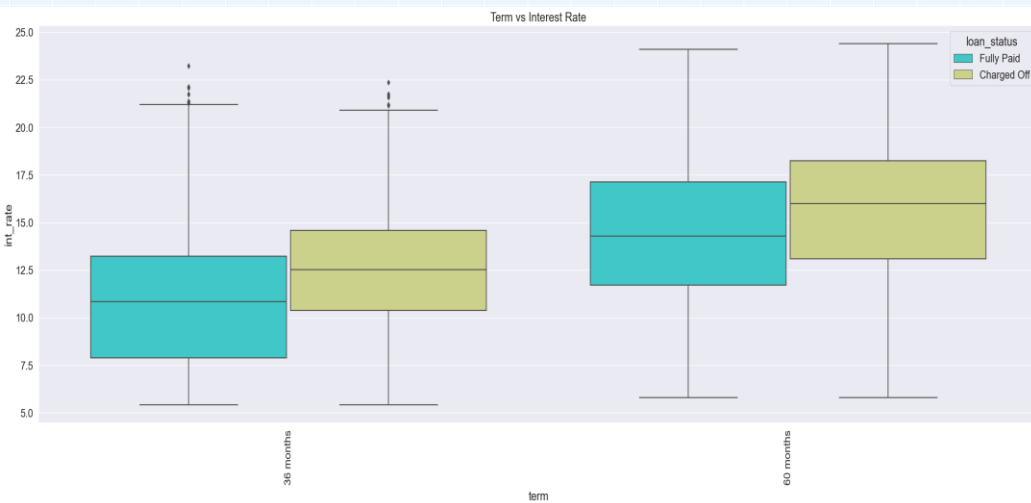
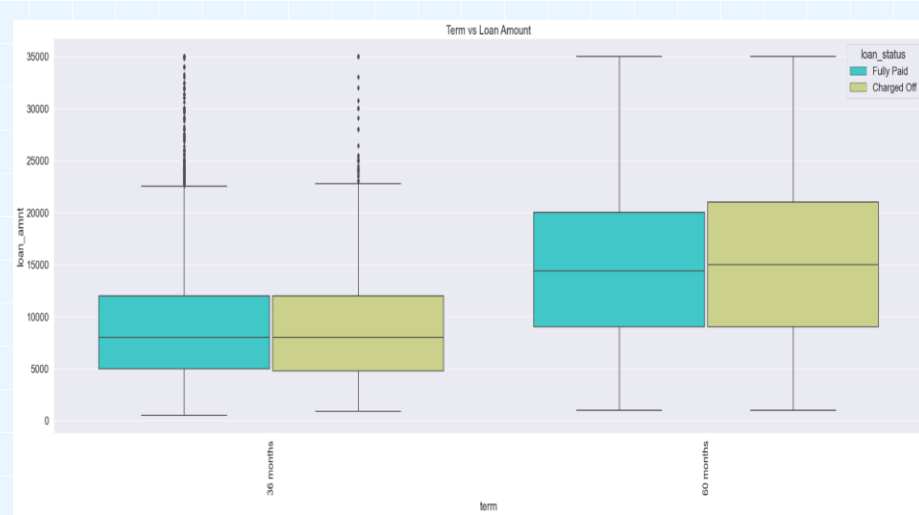
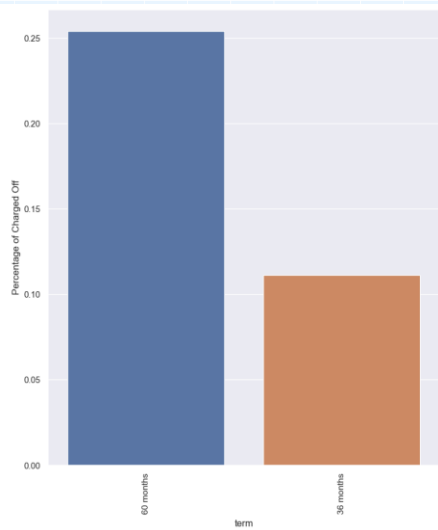
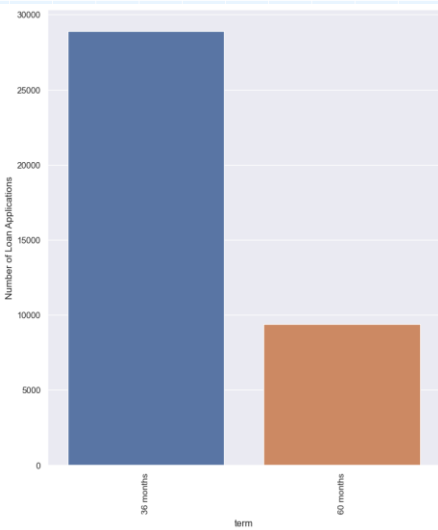
Sub Grade

- Similar trend to grade is also noticed in the subgrade. We see that as we move from A-G Subgrades the loan applicants decrease.



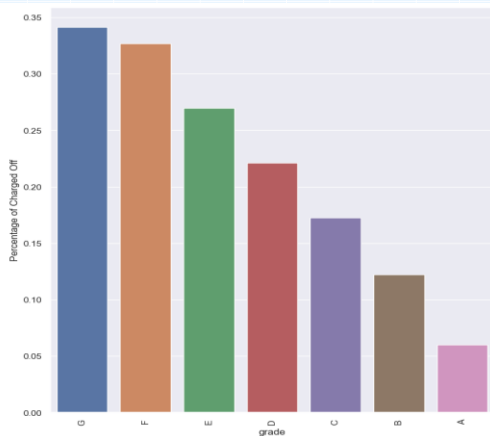
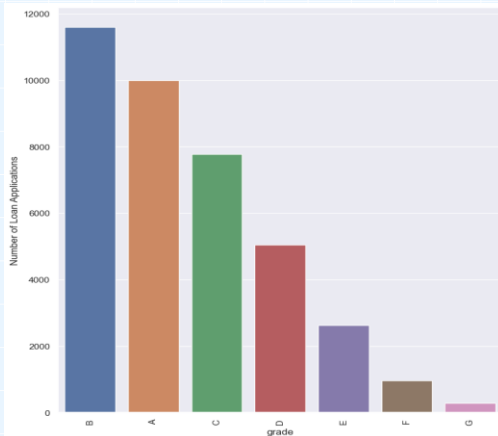
Bivariate Analysis





- As the term increase the risk of default also increases
- Irrespective of term as rate of Interest increases ($> 12.5\%$) the default chance is high.

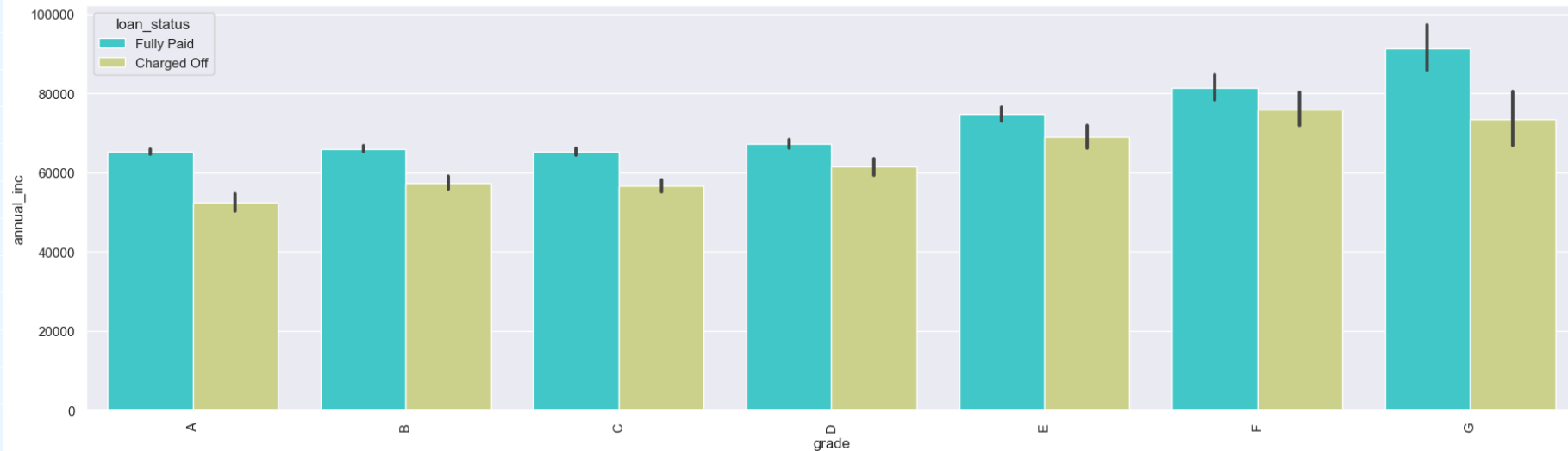


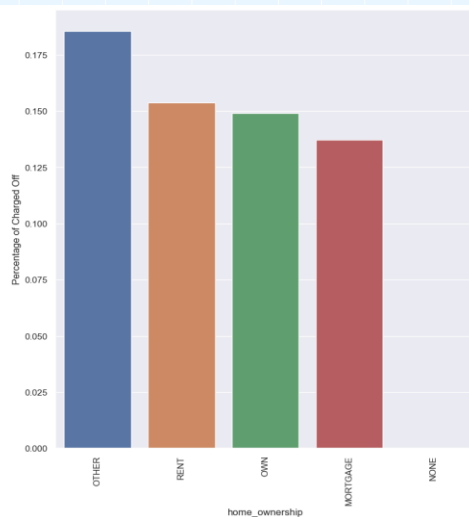
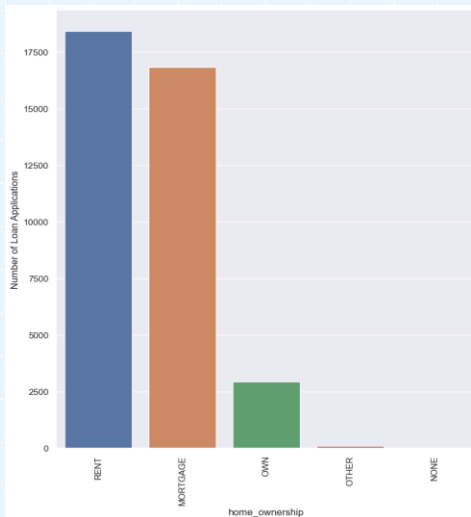


- As the loan grades increases the Charged Off percentage decreases
- As the grade moves from A to G the annual income increase so does the risk of default

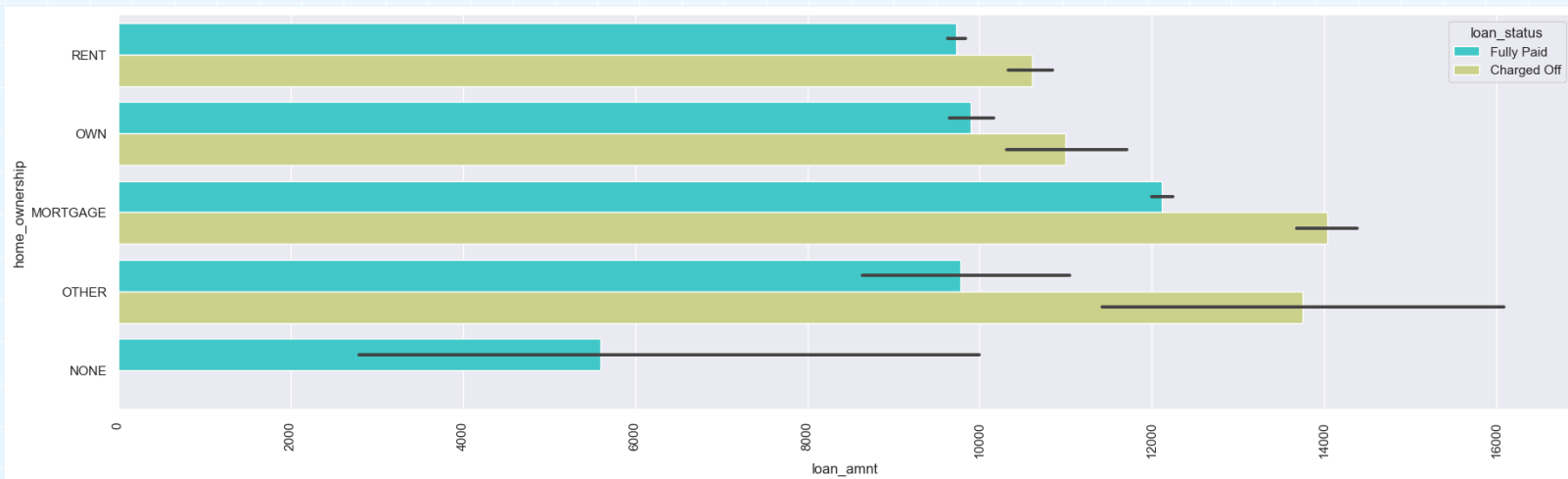


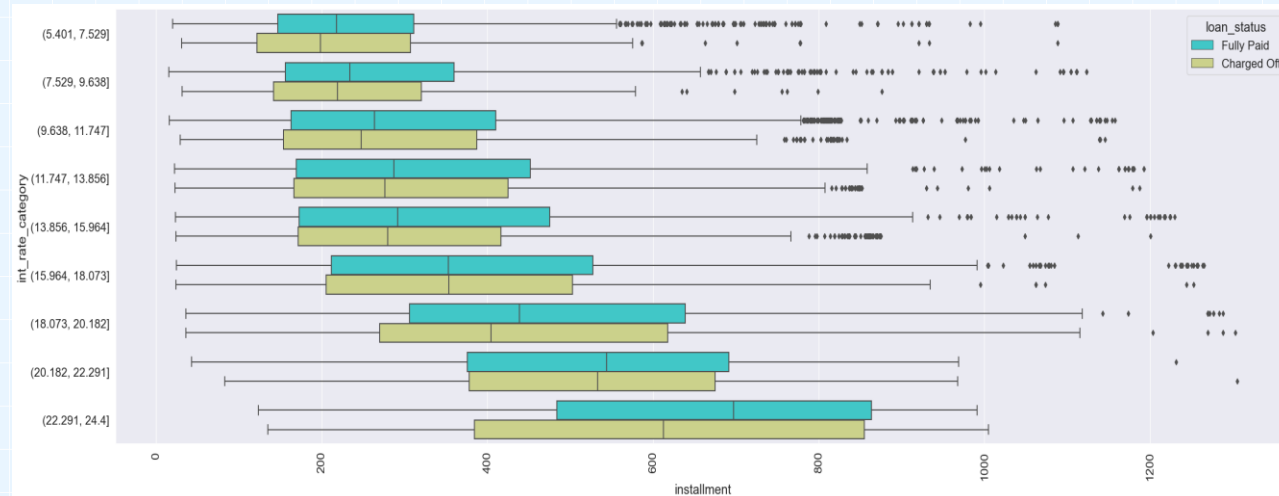
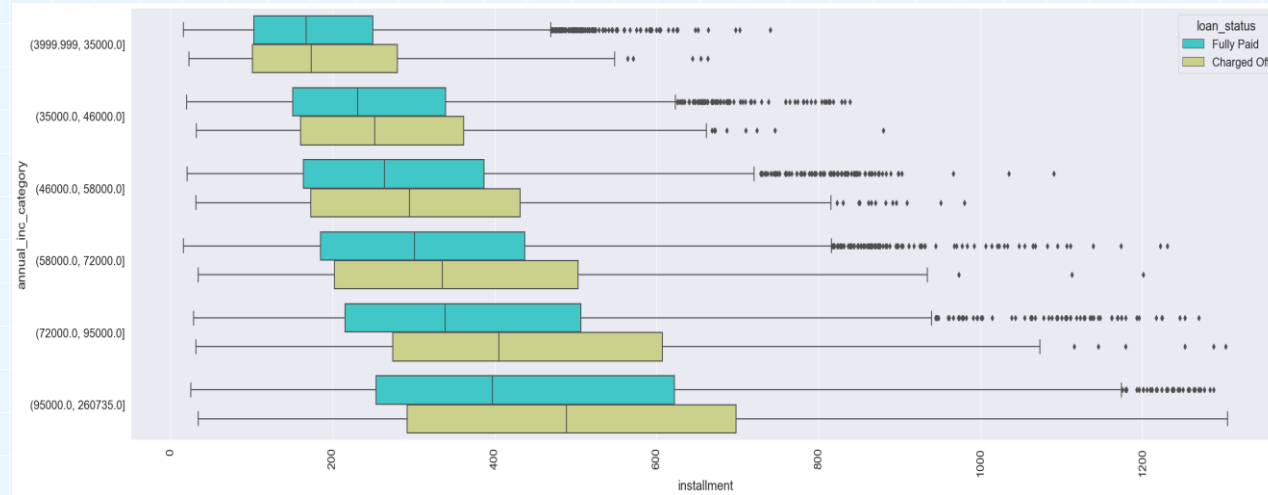
Grades vs Annual Income



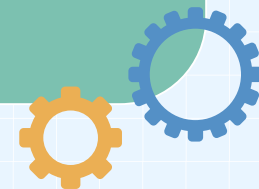


- Loan amount greater than 10K in Owned and Other category is highly likely to get charged off
- Loan amount greater than 12K in Mortgage category is highly likely to get charged off

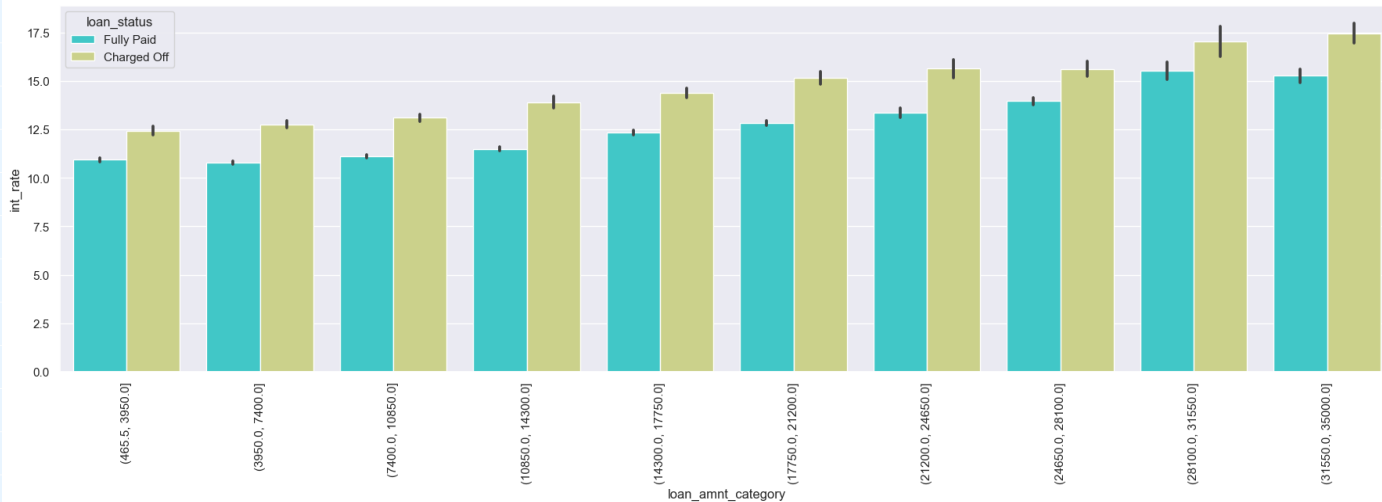




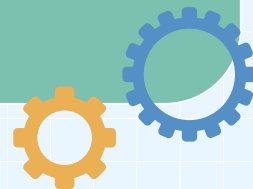
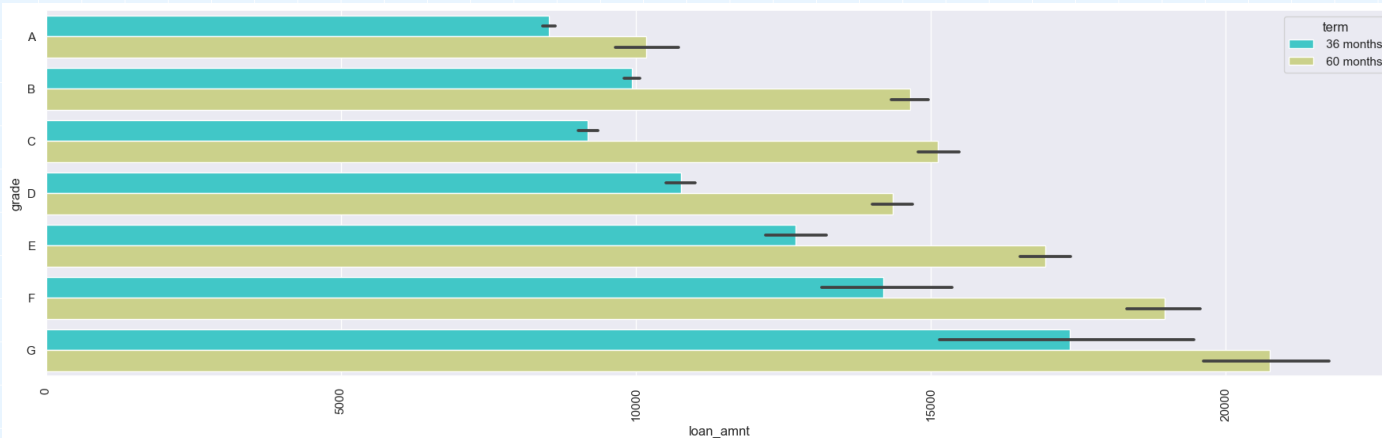
- For all the income category, if the installment increases then the charged off percent also increases
- If Installment Amount is higher than 300 the default rate is high.
- Above 15% interest rate there is drastic increase in charged off percent along with increase in installment.

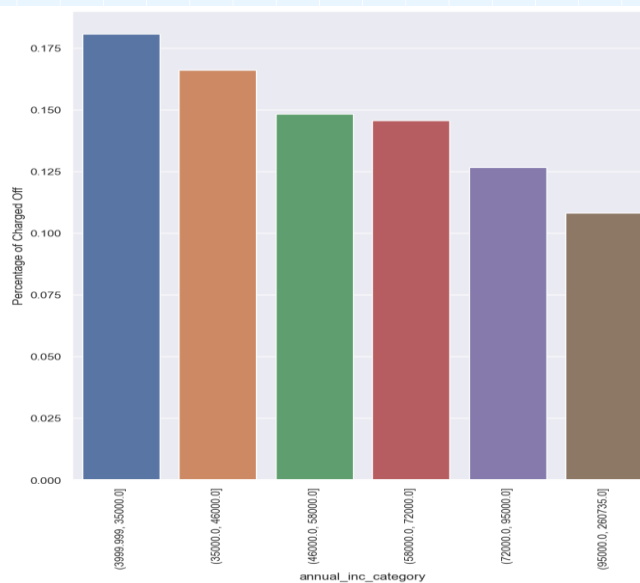
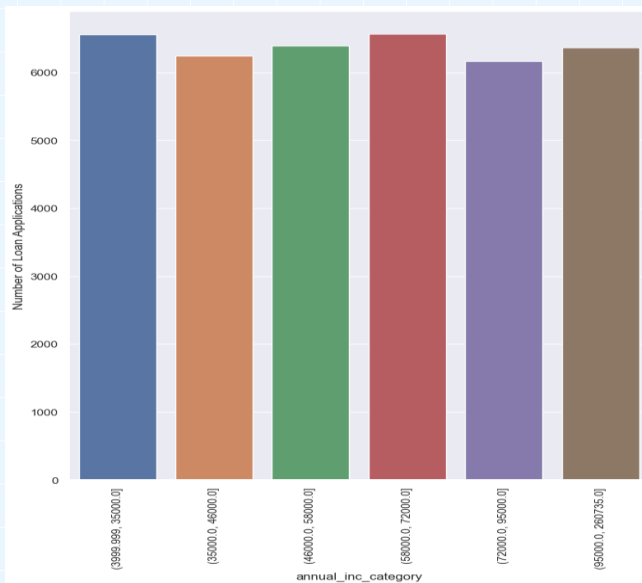
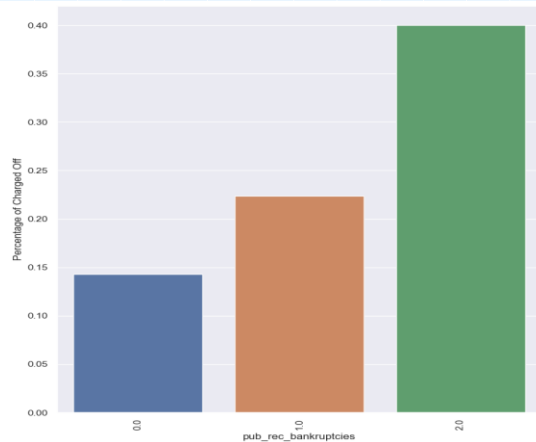
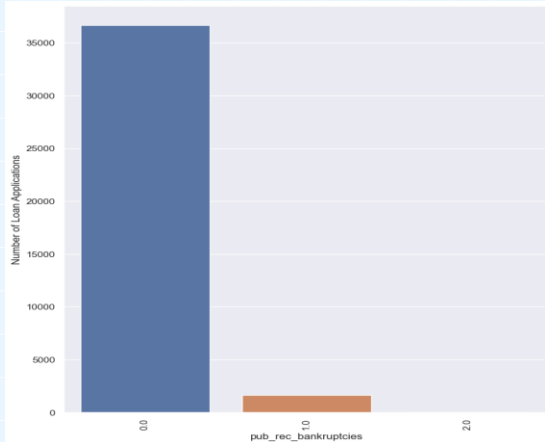


Interest Rate vs Loan Amount



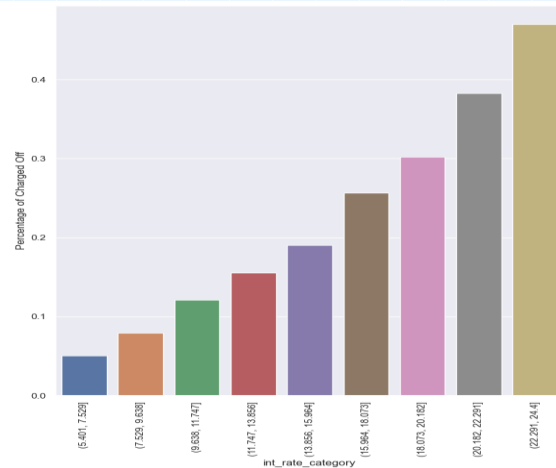
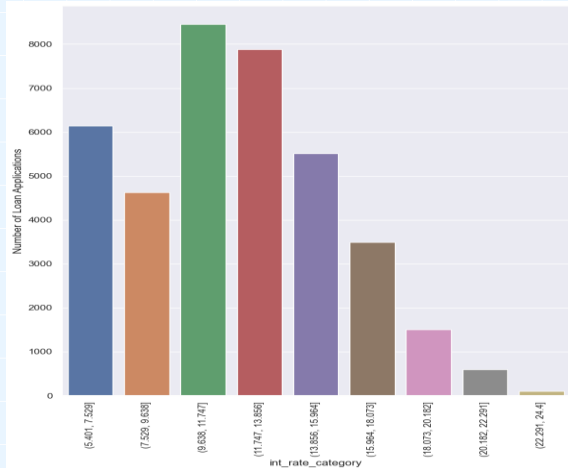
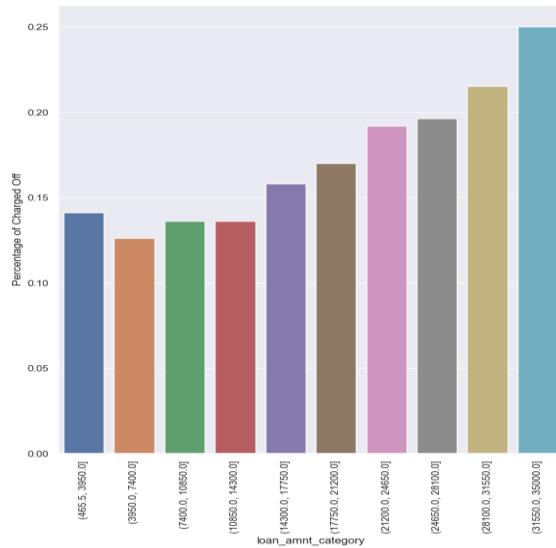
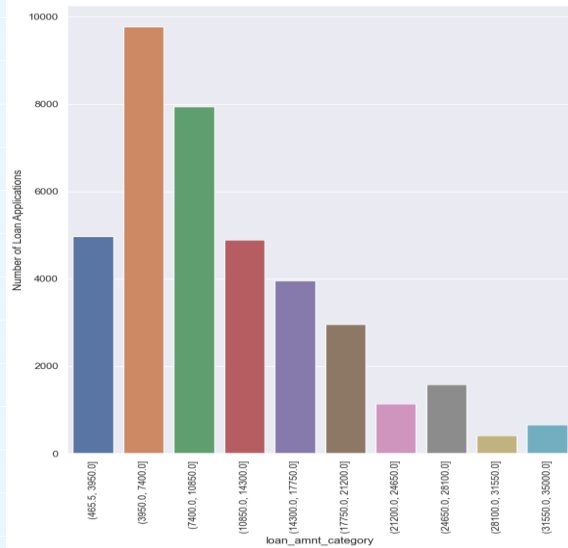
- If Loan amount grows along with growth in the interest rates, the risk of default also grows.
- Median loan amounts of Grades E, F, G are much higher than A, B, C, D which also is seen from earlier plot that more defaults are seen in E, F, G grades in loan amounts > 15K range



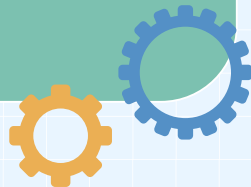


- Applications having bankruptcies records (1, 2) prior to applying are highly likely to default. More default risk where loan amount > 15k for public bankruptcies 1 and 2.
- Highest Charged off Percent is for applicant below 35k annual income and highest loan is taken by them,. The bar plot shows that the as the income increases the chances of defaulting reduces

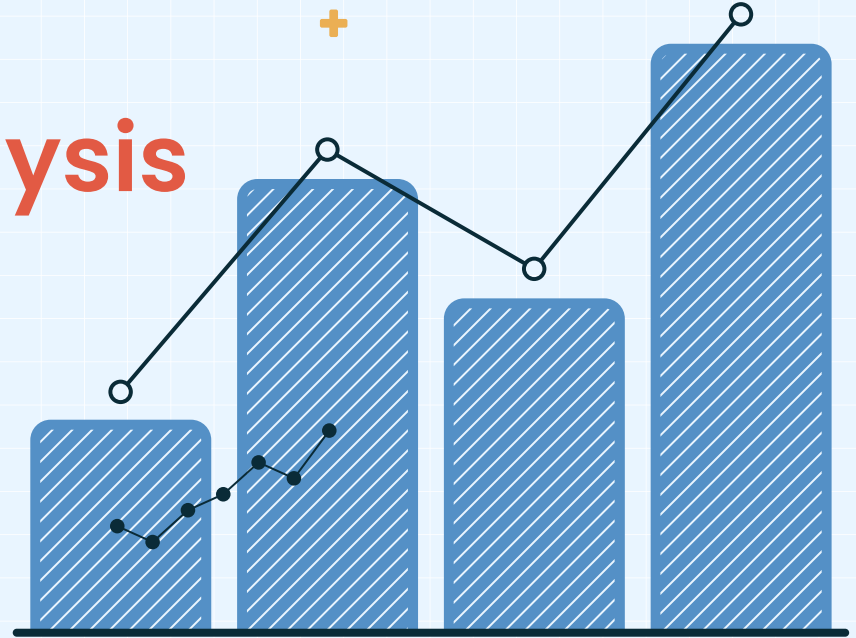


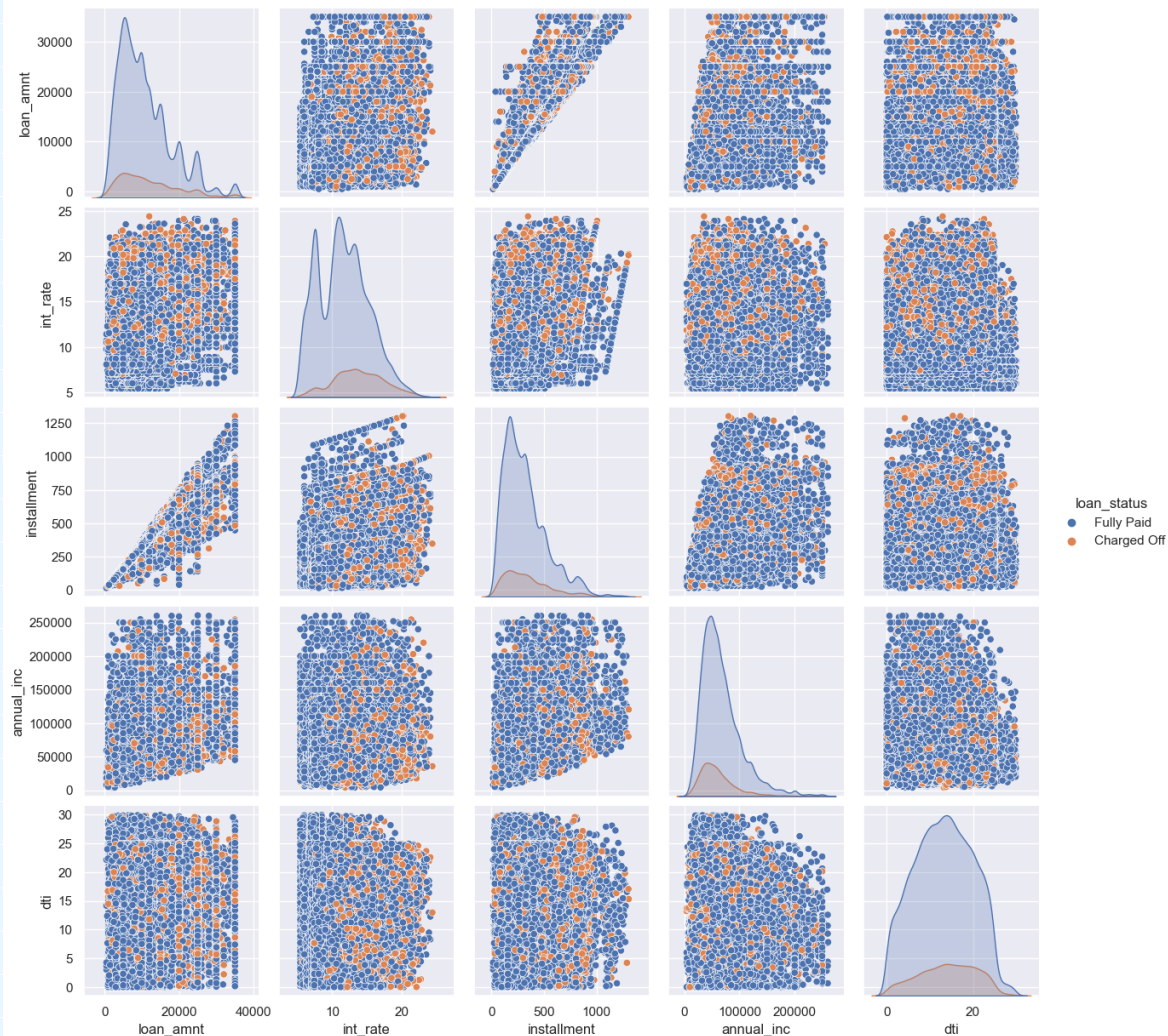


- The default rate increases with increase in loan amount, the increase in default percent is more on amounts > 15K. So limiting the loan amount to less than 15K might help in reducing the default.
- As interest rate grows the chances of default rate also is high.

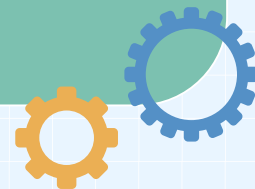


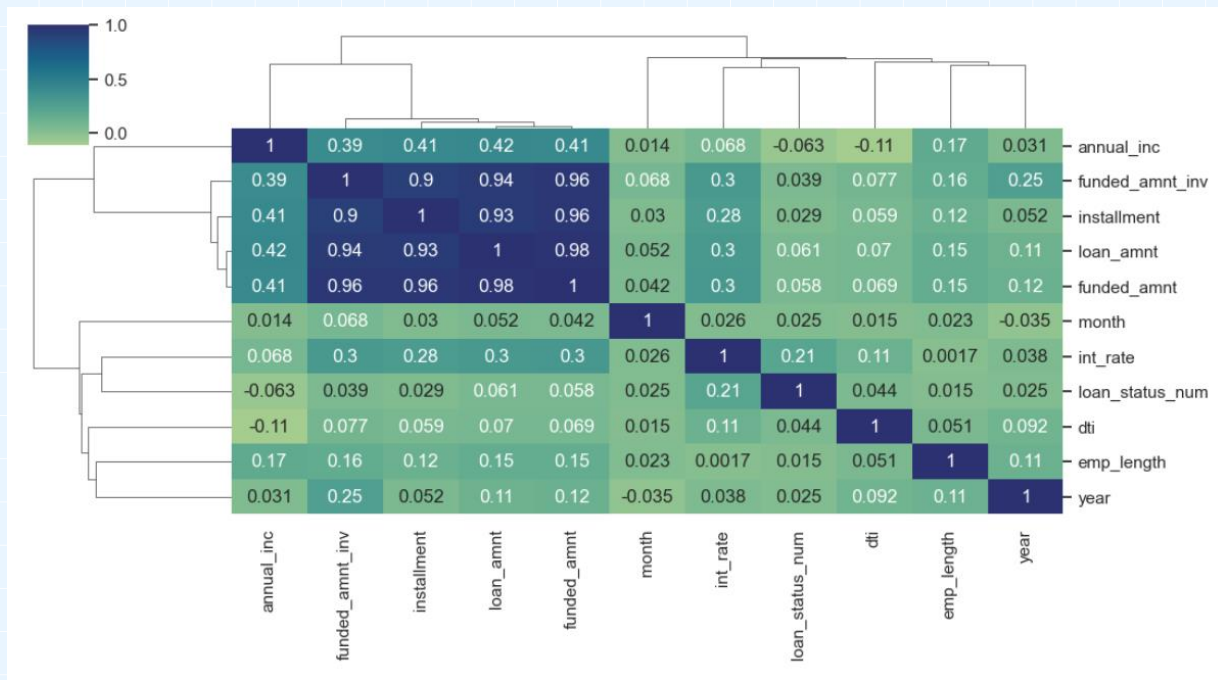
Multivariate Analysis





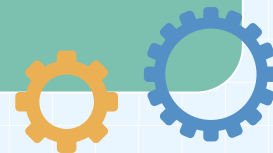
- **Lower Interest Rates with lower Annual Incomes have lower Charged Off loan status**
- **Lower DTI with lower interest rates have lower charged off Loan Status**
- **Installments increase with increase in Loan Amount and number of Charged off as are less in high installments**

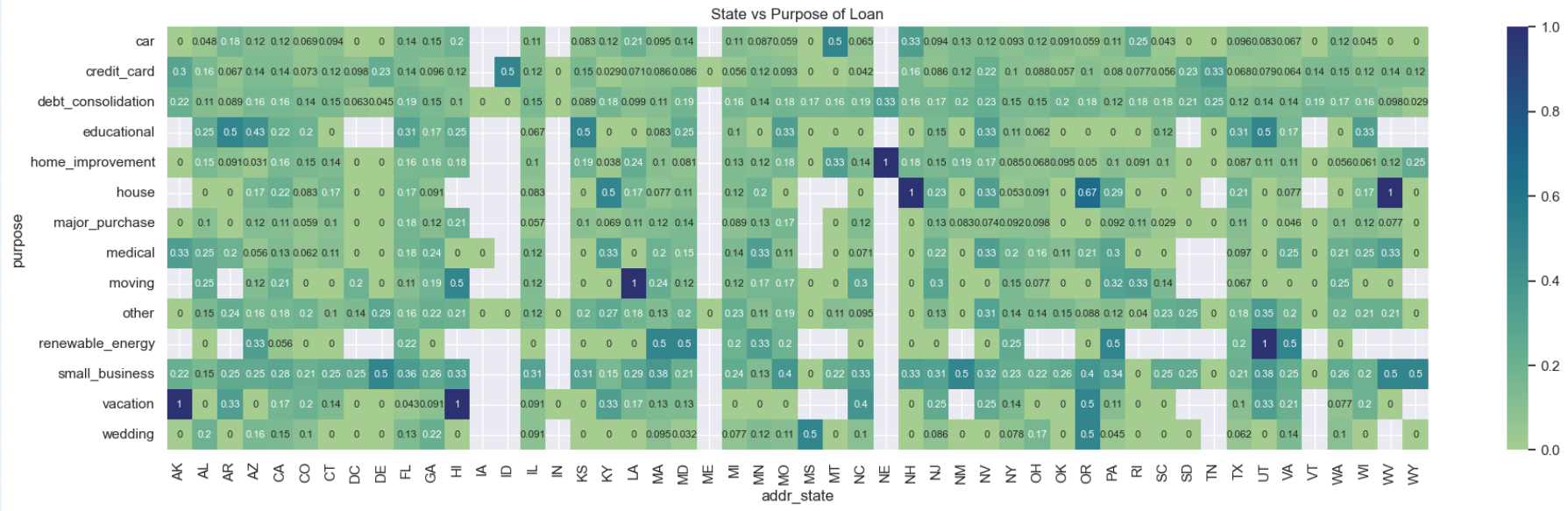




Correlation Matrix with Dendrogram –

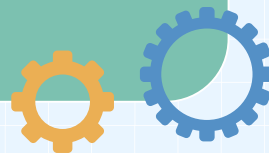
- Loan Amount , Amount Funding, Investor Amount Funding , Installments are highly correlated
- DTI is negatively correlated to Annual Income , which makes sense as since lower the salary more is the ration of salary going towards debt payment

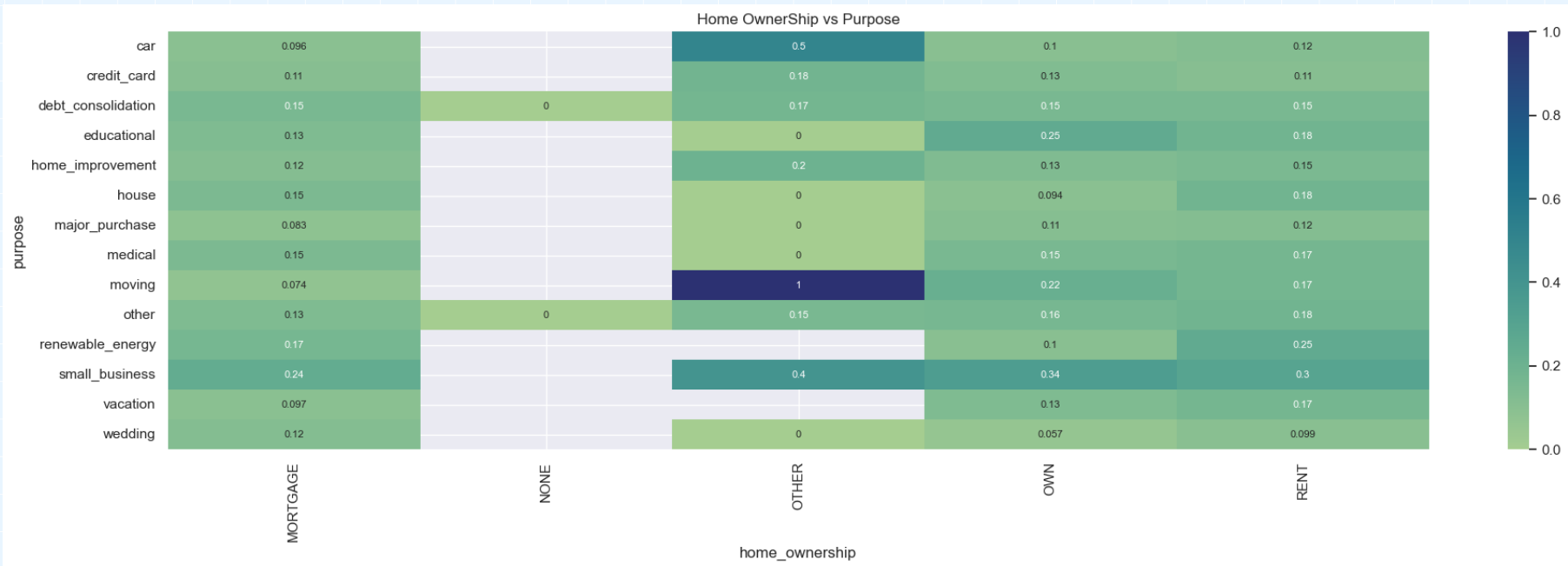




The loans defaulting chances for particular purpose in a specific State are higher for below-

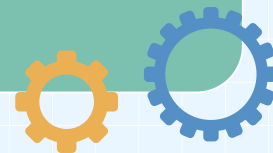
- Moving Loans in LA, HI
- Small Business Loans in DE, NM, WV, WY
- Vacation Loans in AK, HI, OR, NC
- Renewable Energy Loans in MA, MD, PA, UT, VA
- House Loans in NH, WV, OR, KY
- Educational Loans in AR, KS, UT
- Wedding Loans in MS, OR



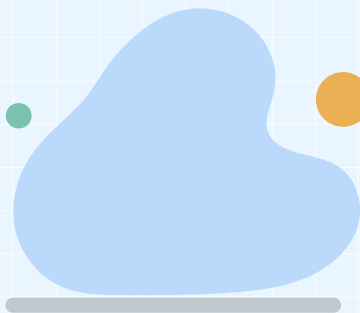
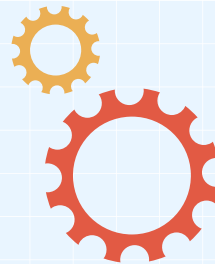
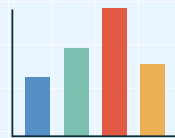
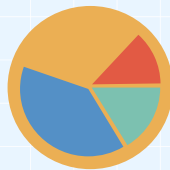


The loans defaulting chances for Home ownership in Purpose category are higher for below–

- Home Ownership as Other for Moving, Small Business, Car.



Conclusion and Recommendations

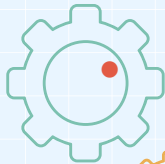




Conclusions



- High Interest rates above 12.5% have higher chances of default.
- Public Bankruptcy 1, 2 have high chances of default
- Loan Grade and Subgrade E, F, G have more default rate
- Loans having 60 months term have higher risk of default, if combined with interest rates risk is higher for 36 months term as well.
- Loan applications for purposes – Small Business, Renewable energy, Education have high risk of getting charged off
- Higher loan amount (above 15k) has more chance of default.
- Higher Installment amount above 300 have more risk.
- Higher DTI ratio above 15% have more risk of default





Recommendations



- Better to **reduce the interest rates for small businesses** to reduce the default rates less than 12%.
- Based on the analysis the loan company should **expand on various states apart from NY and CA only which have earning potential** or more equivalent to these states.
- Should for **target employee length groups more between 3- 9 years by providing lower interest** poses like home ownership.
- **Risky purposes such as vacation ,moving etc the rate of interest should be increased** as the risk of default exposure will be covered . If not to **avoid certain states with certain purposes which have a high default** record should be avoided.
- **Lending less than 10 K** which mitigate the risk of default.
- Should avoid giving low income group high amount loans above 10 k especially states having low income.
- **Offer Reduced Term of loan repayment like 4 yrs option** as well, since longer terms tend to default more.
- **Loan applications should be verified properly** before approving.



Thank You

