Loan Lending Club Case Analysis

By Darshil Ajay Parekh

Contents

- Problem Statement

- Business Understanding

- Risk

- Analysis Approach

- Univariate Analysis

- Bivariate Analysis

- Multivariate Analysis

- Factor whether an applicant will be Defaulter

- Decisive Factor whether an applicant will be Defaulter

Problem Statement

- Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.
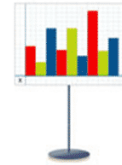

How Lending Club Works

Borrowers apply for loans.
Investors open an account.

Borrowers get funded.
Investors build a portfolio.

**Borrowers** repay automatically.
Investors earn & reinvest.

The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

- When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

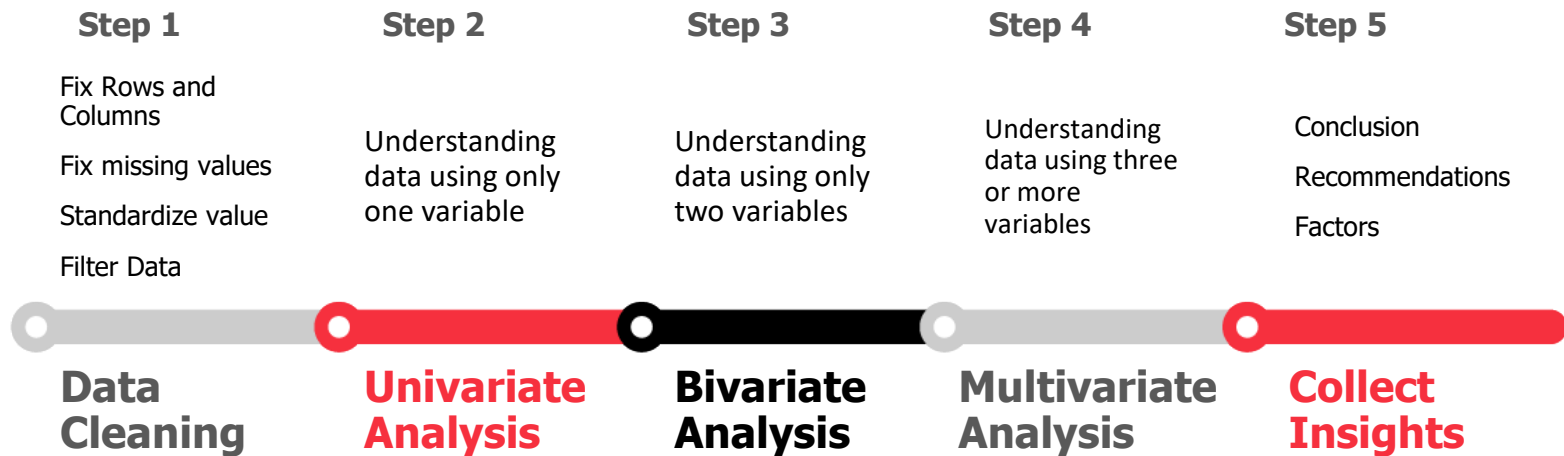1. **Loan accepted:** If the company approve the loan, there are 3 possible scenarios described below:

   - **Fully paid**:
   - **Current**:
   - **Charged-off**:

2. **Loan rejected**: The company had rejected the loan (because the candidate does not meet their requirements etc.).
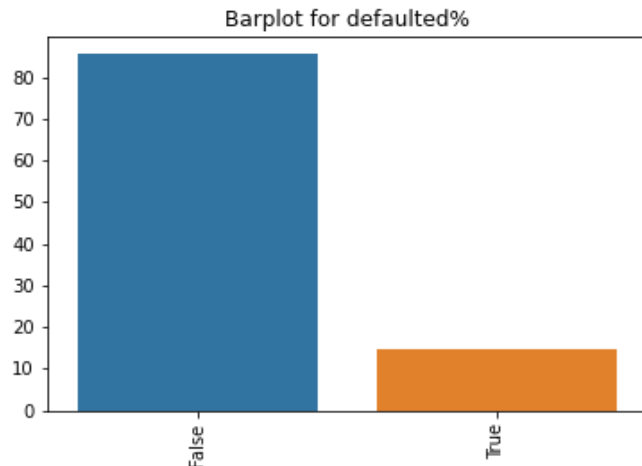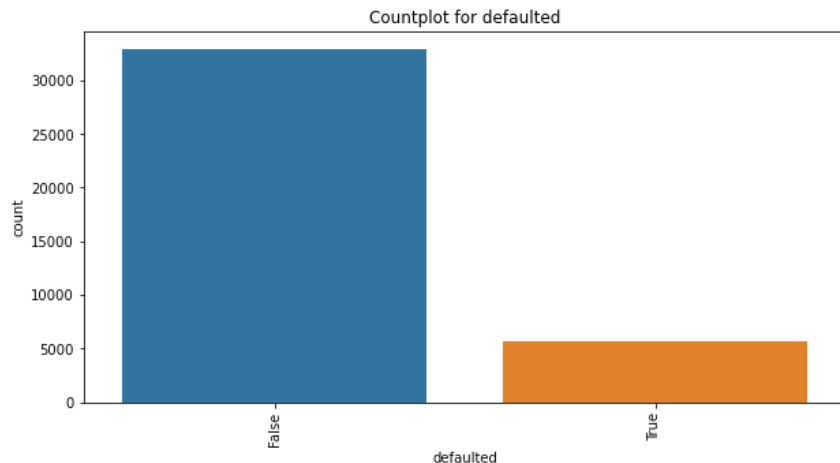
LOAN DATASET
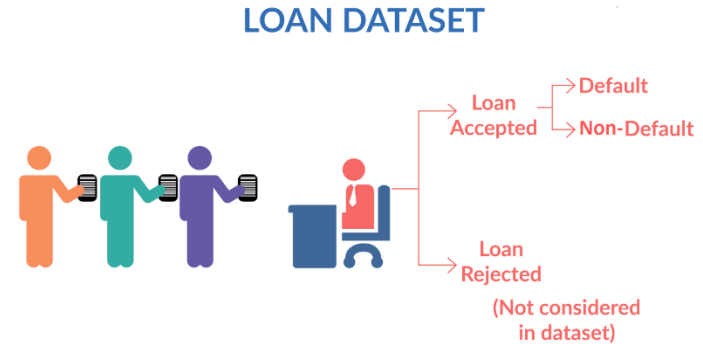
- When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's decision:

1. If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company

2. If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company
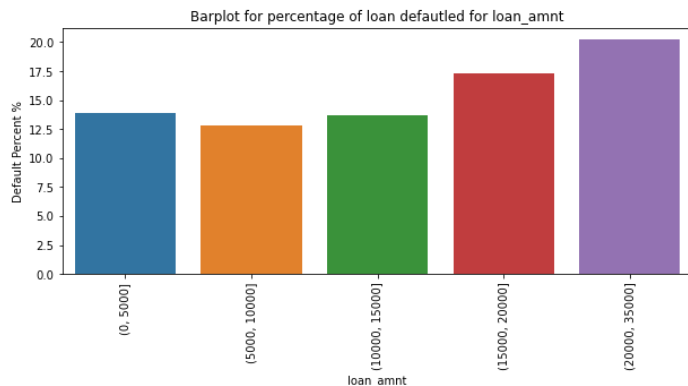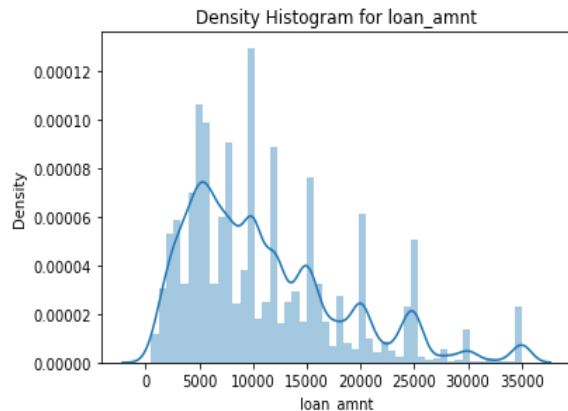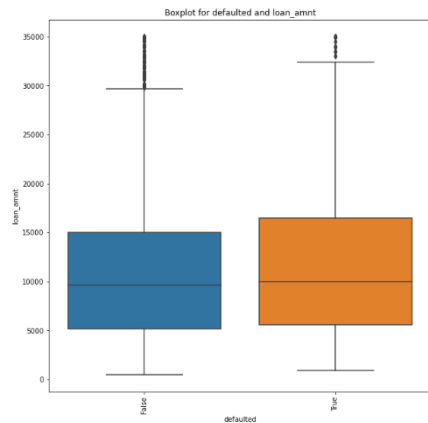
upGrad

Step 1	Step 2	Step 3	Step 4	Step 5
Fix Rows and Columns	Understanding data using only one variable	Understanding data using only two variables	Understanding data using three or more variables	Conclusion
Fix missing values				Recommendations
Standardize value				Factors
Filter Data				

Data Cleaning **Univariate Analysis** **Bivariate Analysis** **Multivariate Analysis** **Collect Insights**

- Defaulted Loan Percent - 14.58%

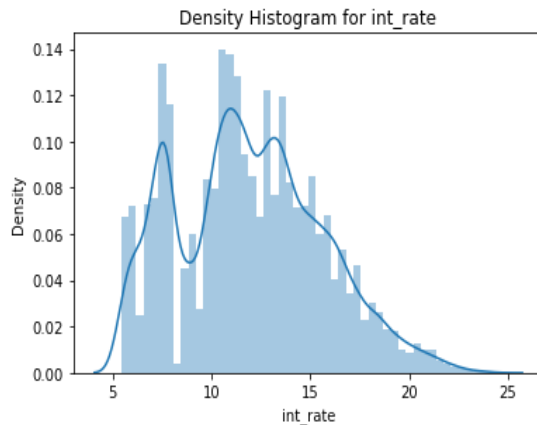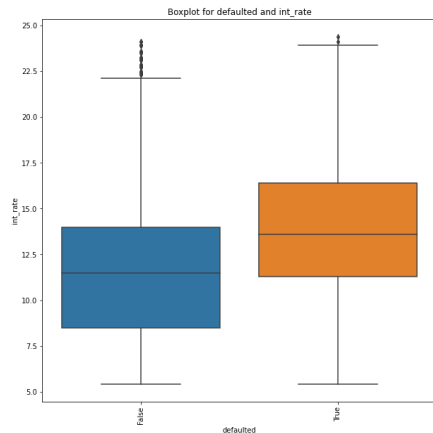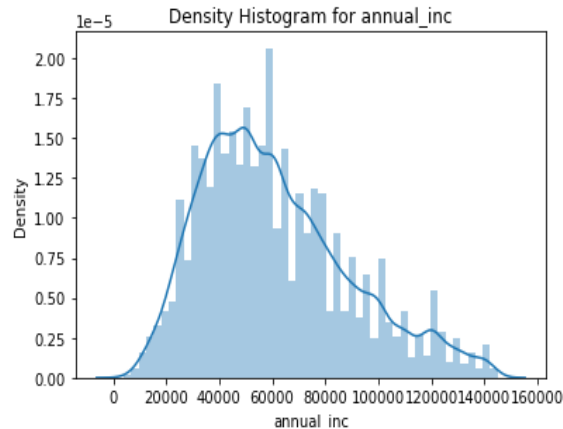- Non Defaulted Loan Percent - 85.41%
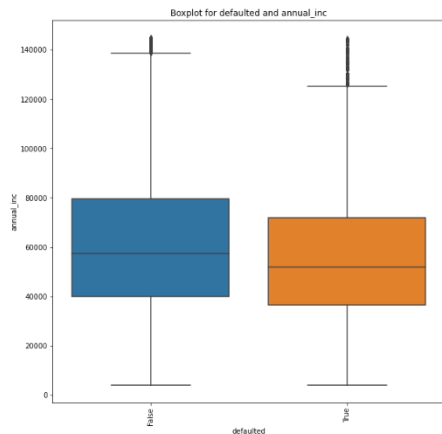
Analysis on Loan Amount

- There are unusual spike of loans for 10k,15k,20k,25k,30k and 35k.

- The loan default percent for loan amount 15k to 35k is higher than 0-15k and higher than overall defaulted Loan Percent 14.58%

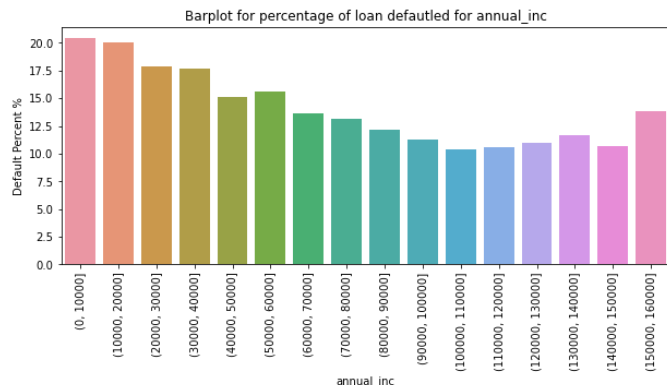- Similar trends have been noted for the funded amount and funded amount invested
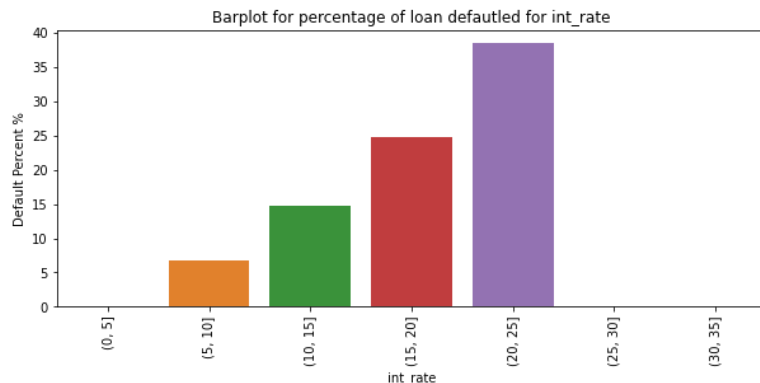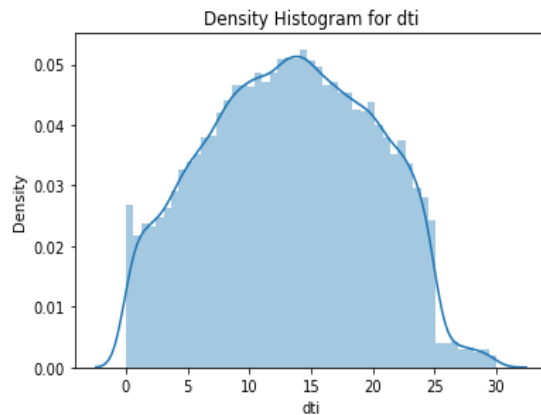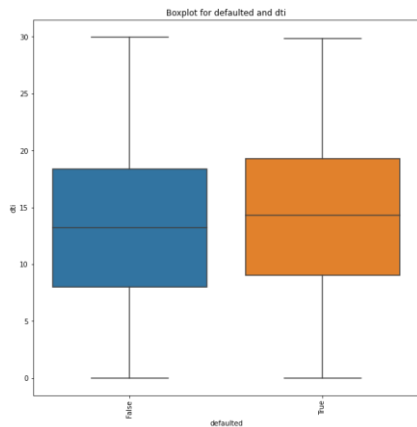
- There are unusual drop of loans for 8-10 Interest rate.

- The loan defaulted percent increase steeply with interest rate.

- The loan defaulted percent for 15-25% interest rate is higher than overall defaulted Loan Percent 14.58%

Boxplot for defaulted and annual_inc

Density Histogram for annual_inc

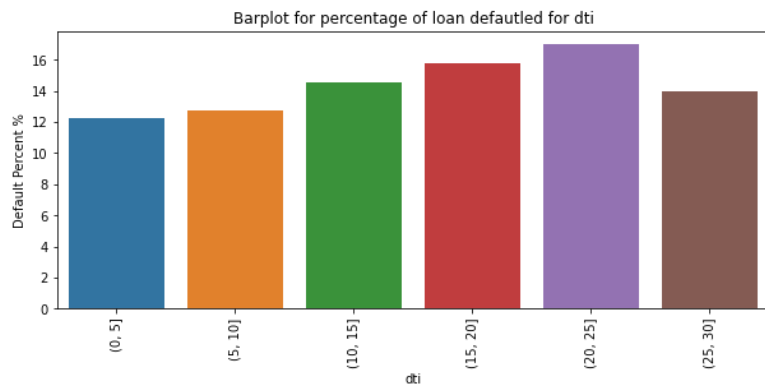Barplot for percentage of loan defautled for annual_inc

- The loan defaulted percent decrease gradually with Annual Income.

- The loan defaulted percent for 0-40000 annual income is higher than overall defaulted Loan Percent 14.58%
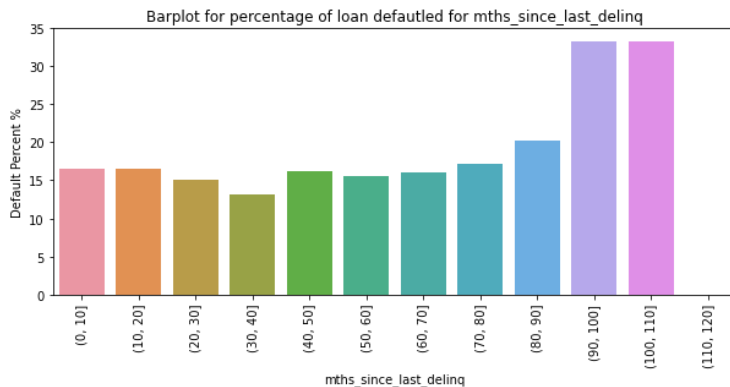
Boxplot for defaulted and dti

Density Histogram for dti

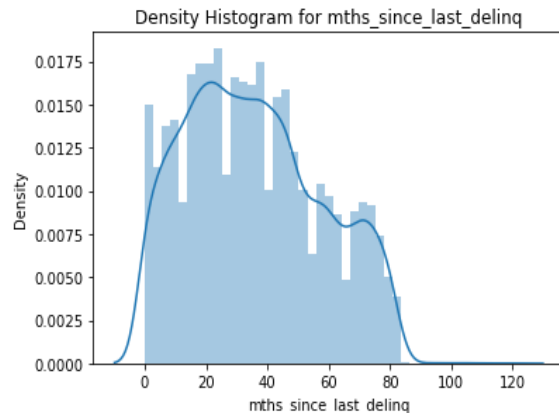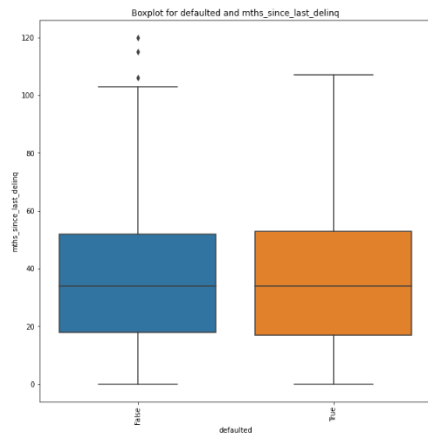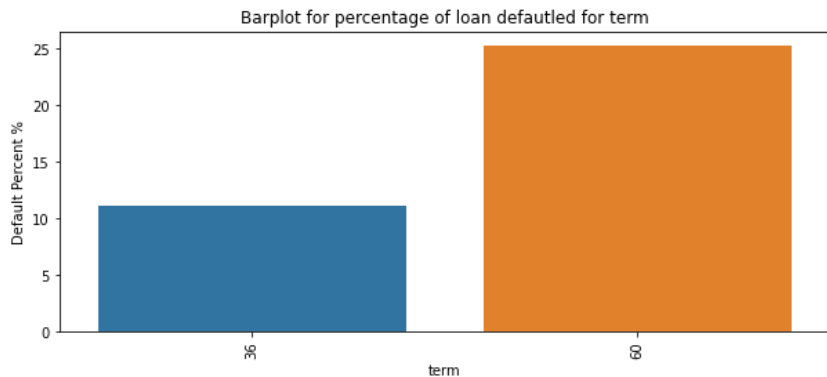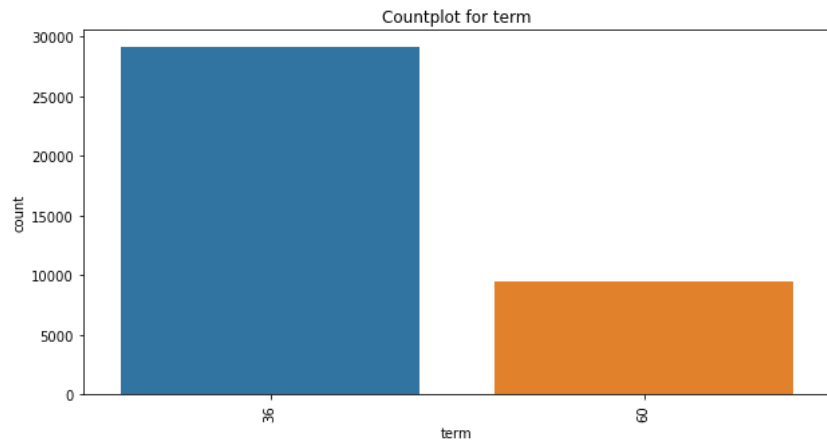Barplot for percentage of loan defautled for dti

- The loan defaulted percent increase gradually with DTI.

- The loan defaulted percent for 15-25 DTI is higher than overall defaulted Loan Percent 14.58%

Boxplot for defaulted and mths_since_last_delinq

Density Histogram for mths_since_last_delinq

Barplot for percentage of loan defautled for mths_since_last_delinq
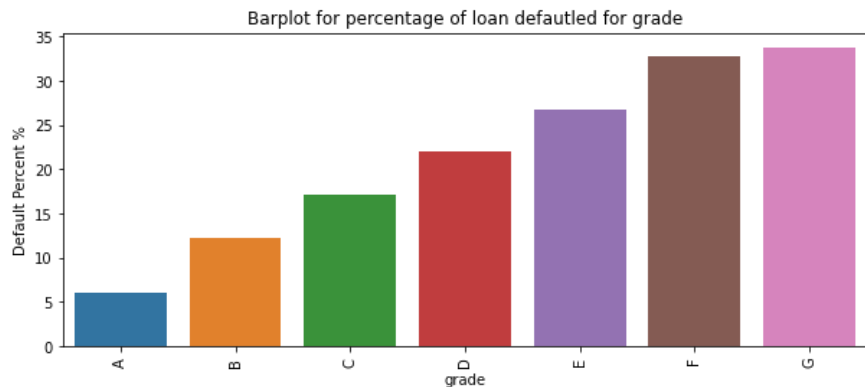
- The loan defaulted percent remains constant till 80 months.

- The loan defaulted percent for 80-110 months of borrowers crime is higher than overall defaulted Loan Percent 14.58%.

- The borrowers have taken more loans for 36 months term verses 60 months

- The loan defaulted percent for 60 months term is twice that of 36 months

- The loan defaulted percent for 60 months term is higher than overall defaulted Loan Percent 14.58%.
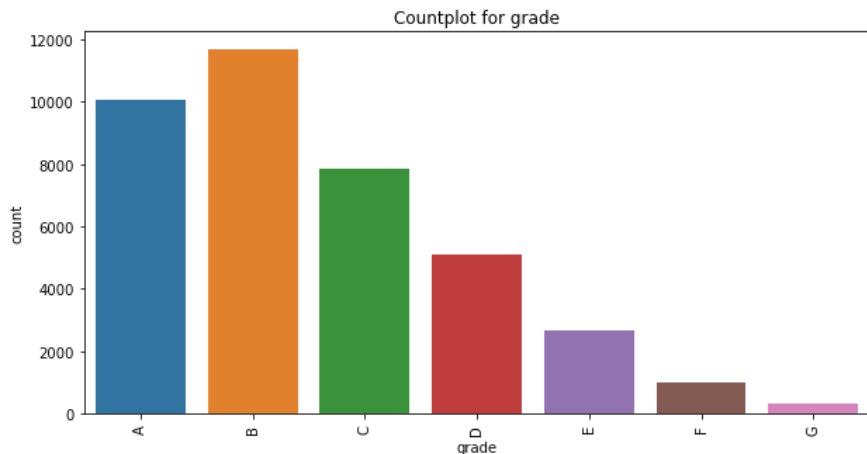
- The grade (A,B,C) have more number of loans.

- The loan defaulted percent increase with Grade from A - G

- The loan defaulted percent for (D,E,F,G) grade is higher than overall defaulted Loan Percent 14.58%.

Countplot for sub_grade
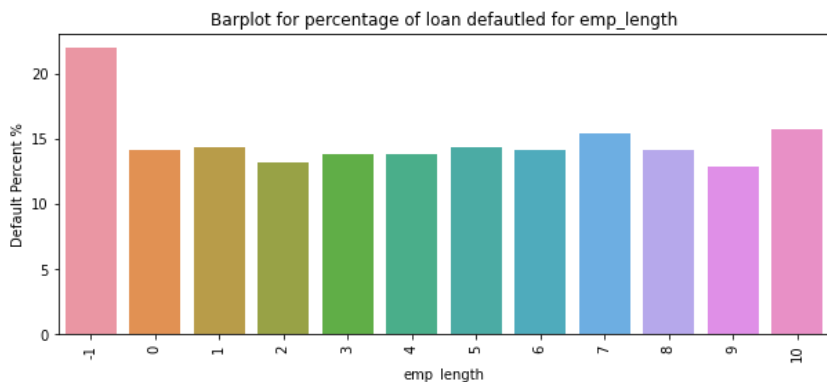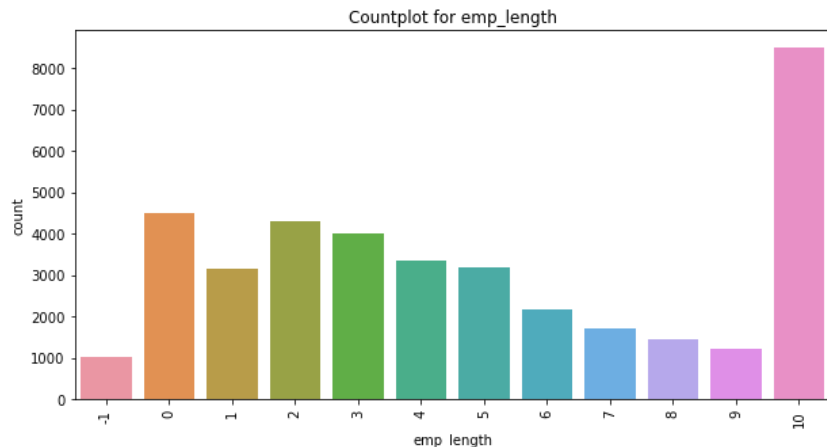
Barplot for percentage of loan defautled for sub_grade
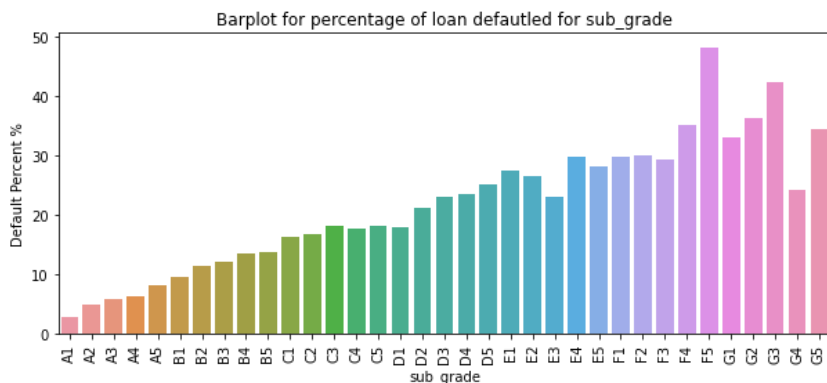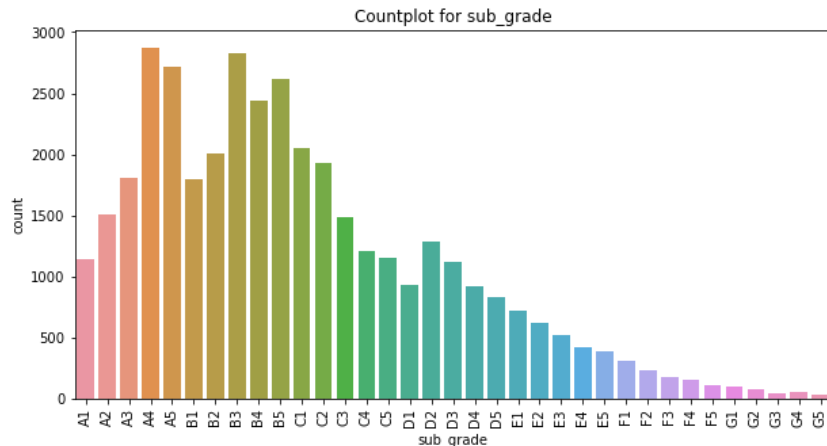
- The grade (A1 – C2) have more number of loans.

- The loan defaulted percent increase with Grade from A1 – G5

- The loan defaulted percent for (C2-G5) grade is higher than overall defaulted Loan Percent 14.58%.

Countplot for emp_length
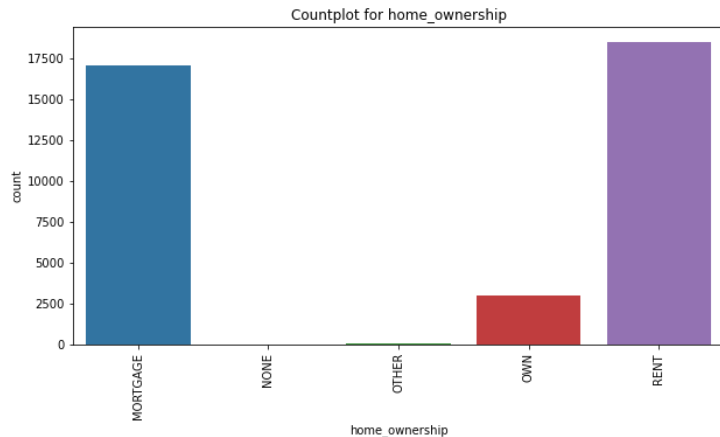

Barplot for percentage of loan defautled for emp_length

- The 10+ years of employment have the most number of loans.

- The loan defaulted percent remains constant throughout the employment years.

- Note : -1 denotes the missing values.

Countplot for home_ownership

Barplot for percentage of loan defautled for home_ownership

- Borrowers on rent and mortgage have most number of loans

- The loan defaulted percent remains constant for mortgage, own and rent

- For higher amount of loan the borrowers on mortgage the loan default spread is more.

Boxplot for loan_amnt and home_ownership segmented by defaulted
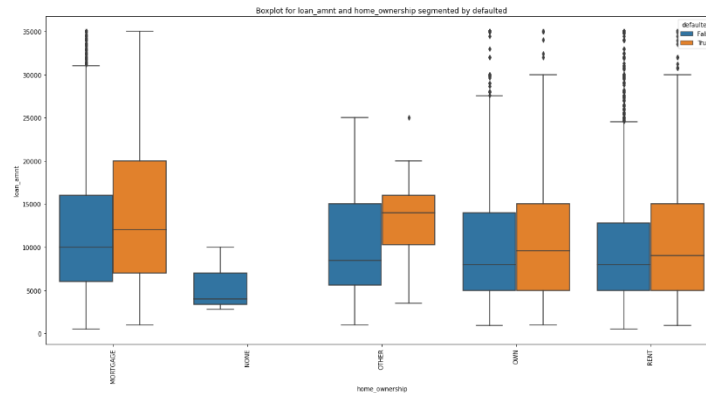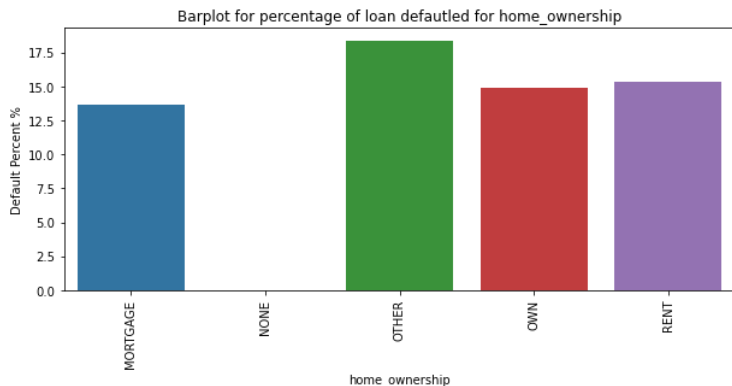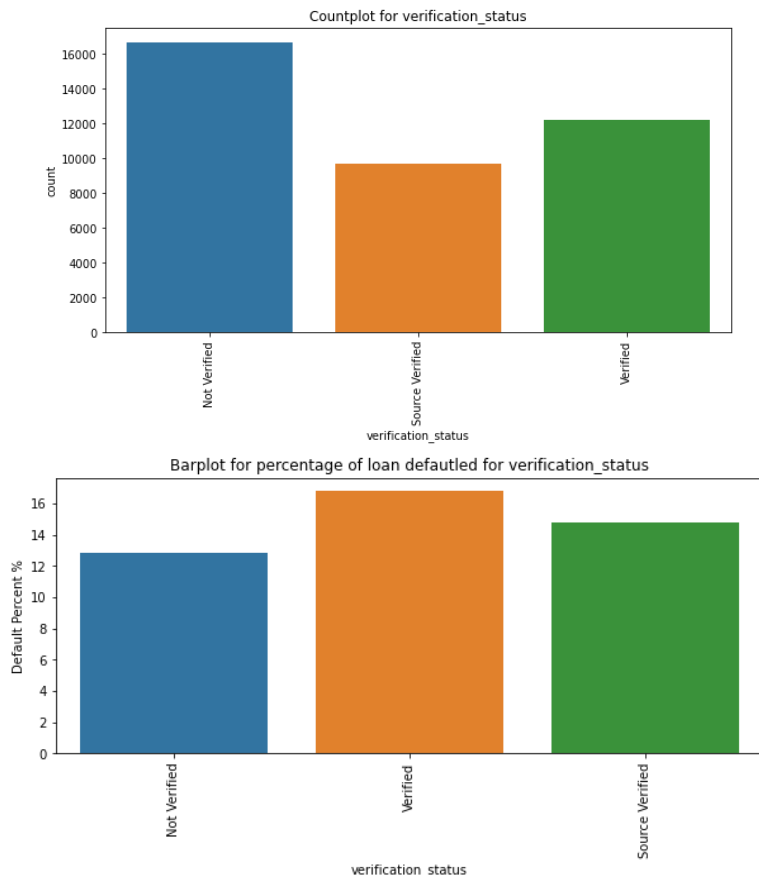
- The not verified borrowers have take more loans than verified and source verified.

- The loan defaulted percent for verified borrower is higher than overall defaulted Loan Percent 14.58%.

Countplot for purpose

Barplot for percentage of loan defautled for purpose

- The Debt consolidation and credit card borrowers have take more loans than others.

- The loan defaulted percent for small business and renewable energy borrowers is higher than overall defaulted Loan Percent 14.58%.

Countplot for pub_rec_bankruptcies
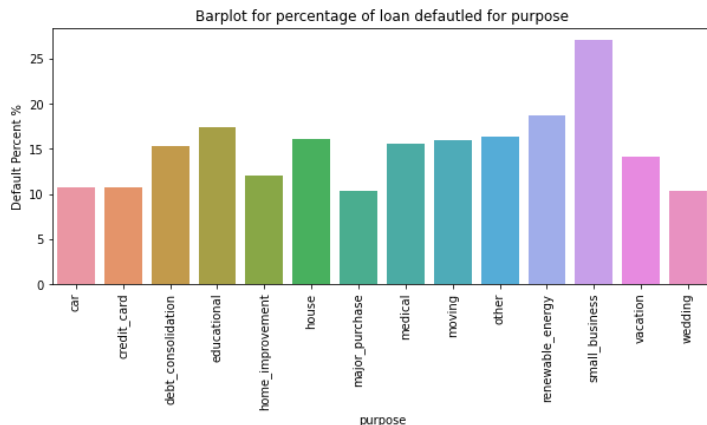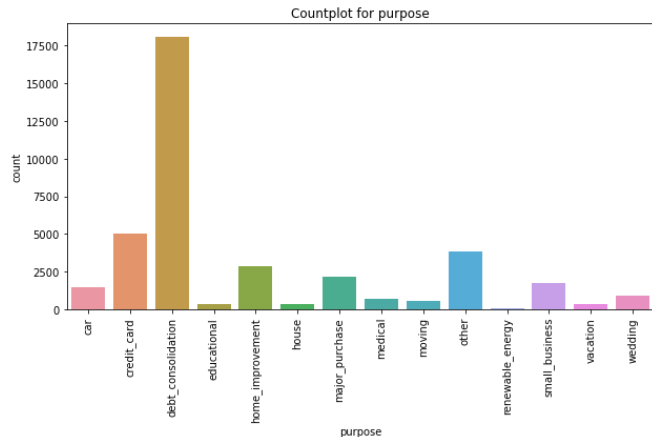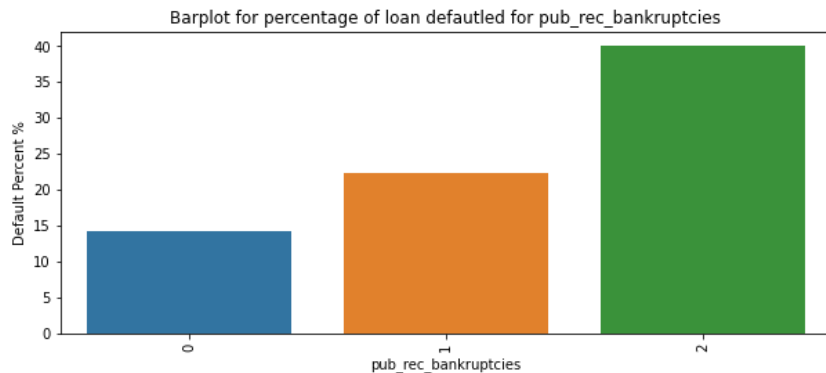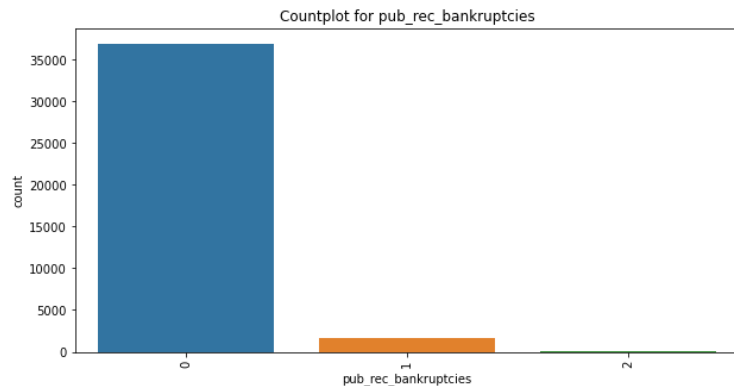
- Borrowers zero public bankruptcies have most number of loans

- The loan defaulted percent increases for one or more public bankruptcies record

- The loan defaulted percent for one or more bankruptcies is higher than overall defaulted Loan Percent 14.58%..

Barplot for percentage of loan defautled for pub_rec_bankruptcies

loglog plot for pub_rec_bankruptcies

Boxplot for int_rate and pub_rec_bankruptcies segmented by defaulted
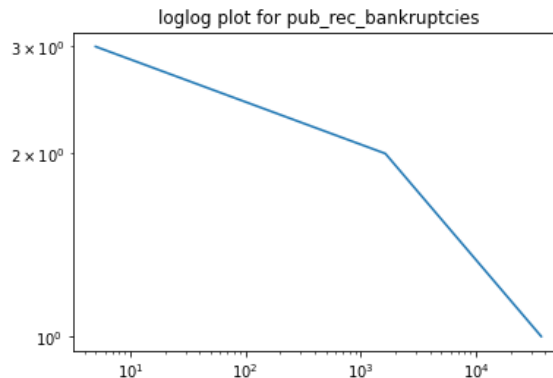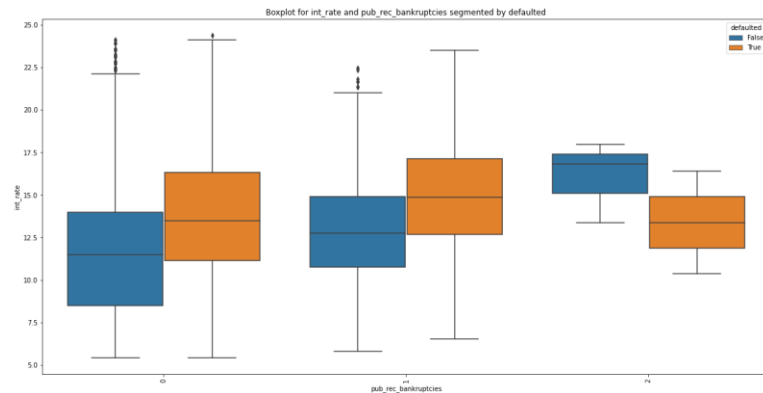
- INTEREST RATE AND PUBLIC BANRUPTIES RECORD : Borrowers with lower interest rate and 2 public bankruptcies have defaulted
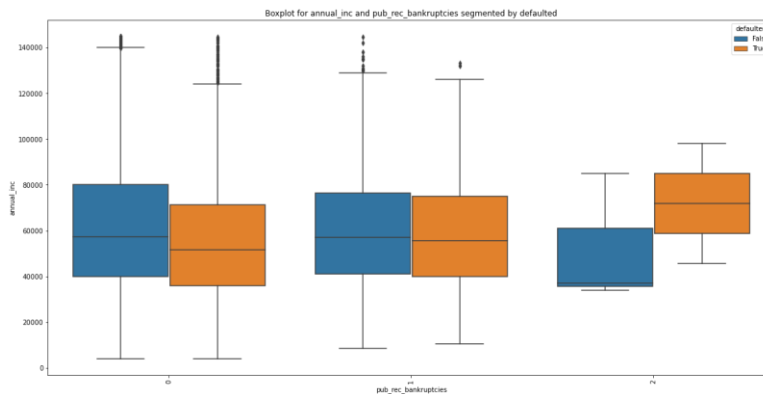
- INSTALLMENT AND PUBLIC BANRUPTIES RECORD : Borrowers with higher installment and 2 public bankruptcies have defaulted

- ANNUAL INCONE AND PUBLIC BANRUPTIES RECORD : Borrowers with higher Income and 2 public bankruptcies have defaulted

Boxplot for annual_inc and pub_rec_bankruptcies segmented by defaulted

Boxplot for loan_amnt and pub_rec_bankruptcies segmented by defaulted

Countplot for addr_state

Barplot for percentage of loan defautled for addr_state
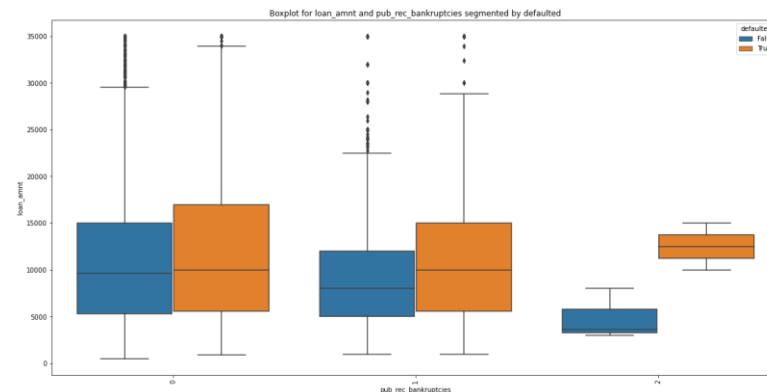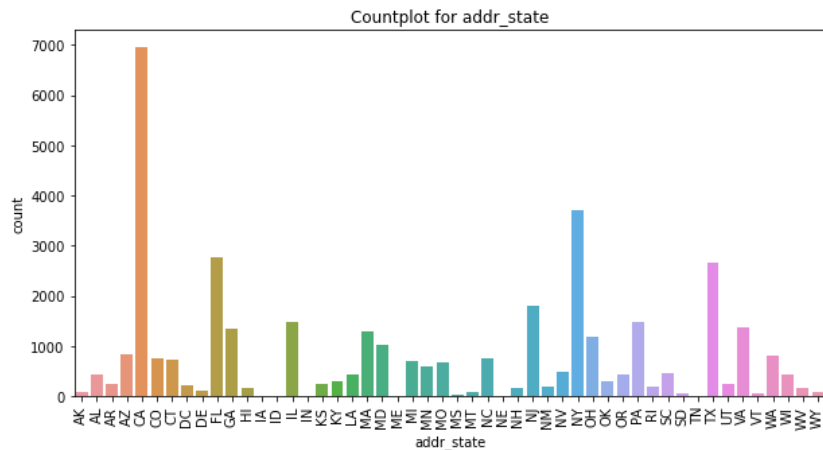
- We can see more number of loans from certain state like CA,FL,NY and NI

- Percent of loan defaulted is very high for state NE and high for NV,AK and SD.
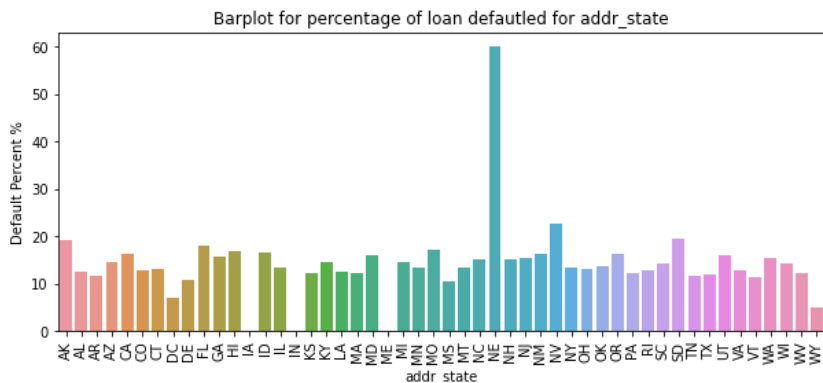
- Term and interest rate have the highest correlation with the defaulted loan with 0.17 and 0.21 coefficient

upGrad

- **Continuous Variable:**

1. LOAN_AMOUNT : Loan amount greater than 15000 dollars have higher default rate

2. FUNDED_AMOUNT : Funded amount greater than 15000 dollars have higher default rate
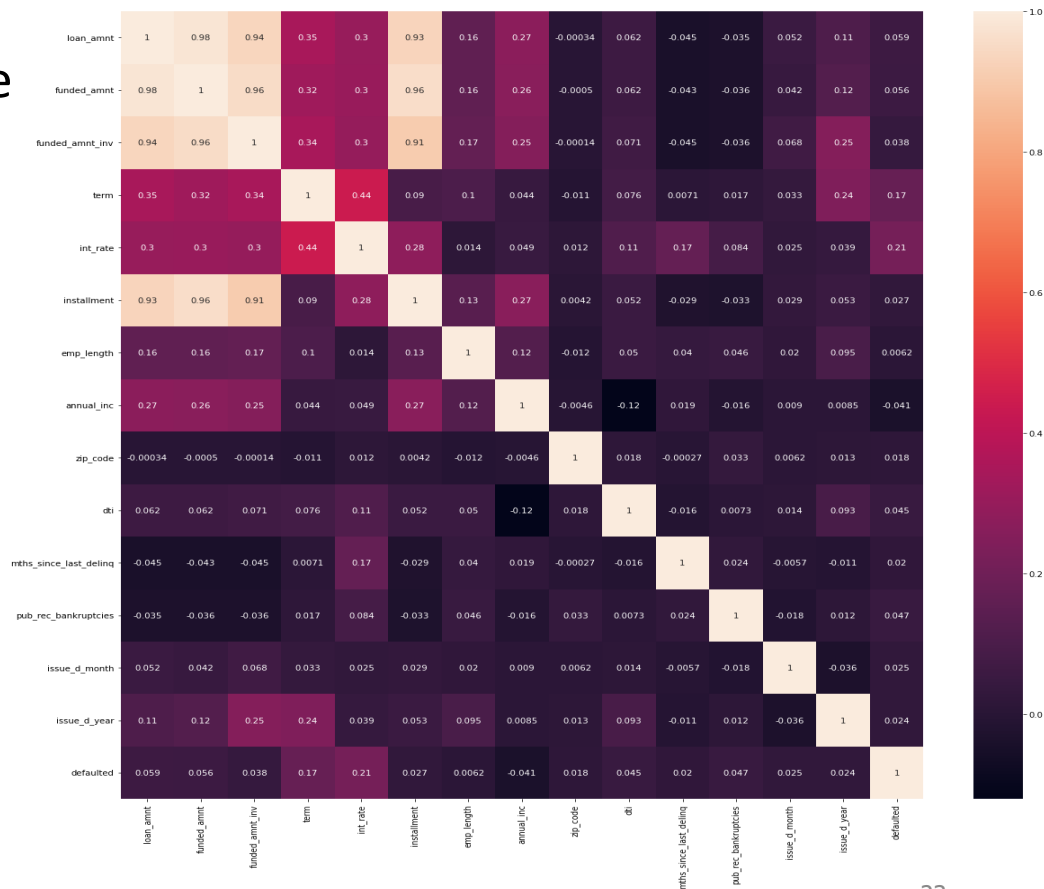
3. FUNDED_AMOUNT_INVESTED : Funded amount invested greater than 15000 dollars have higher default rate

4. INTEREST_RATE : As Interest rate increases the default rate increases steeply
    1. (5, 10]-> 6.739748%
    2. (10, 15]-> 14.820089%
    3. (15, 20]-> 24.826918%
    4. (20, 25]-> 38.441558%

5. ANNUAL_INCOME : As the annual income increase the default rate decreases

6. DTI : As dti increase the default rate increases

7. MONTHS_SINCE_LAST_DELINQ : Crime committed between 90 to 110 days have higher default percent

Factor whether an applicant will be Defaulter

- ## **Categorical Variable:**

1. TERM : 60 months term have a higher default rate than 36 months term

2. GRADE : As the Grade decreases (A B C D E F G) default rate increases

3. SUB_GRADE : As the Sub Grade decreases (A1 A2 B1 B2.....) default rate increases

4. VERIFICATION STATUS : Percent of loan defaulted is higher for verified borrowers

5. PURPOSE : Small business borrowers have high default rate

6. PUBLIC_BAKRUPTIES_RECORD : One or more public bankruptcies have higher default rate

7. STATE : Percent of loan defaulted is very high for state NE and high for NV and SD

8. EMPLOYEE TITLE : The following have the highest default rate among the top 20 title by frequency:
    1. walmart -> 25.242718%
    2. united parcel service -> 22.448980%

9. ZIP CODE : The following have the highest default rate among the top 20 zip code by frequency:
    1. 917xx -> 20.882353%
    2. 331xx -> 20.771513%

- FUNDED_AMOUNT_INVESTED

- INTEREST_RATE

- ANNUAL_INCOME

- DTI

- TERM

- GRADE

- SUB_GRADE

- PURPOSE

- PUBLIC_BAKRUPTIES_RECORD

Thank You

Contact Details :
LinkedIn : https://www.linkedin.com/in/darshil2848/
GitHub : https://github.com/darshil2848
Email : darshil2848@gmail.com