upGrad

LendingClub

Lending Club Case Study

Group Members:

1. Gaurav Tripathi
2. Varun Priyadarshi

Abstract

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

When reviewing loan applications, company faces 2 risks as follows:

    A. Turning down credit worthy loan application, and losing business.

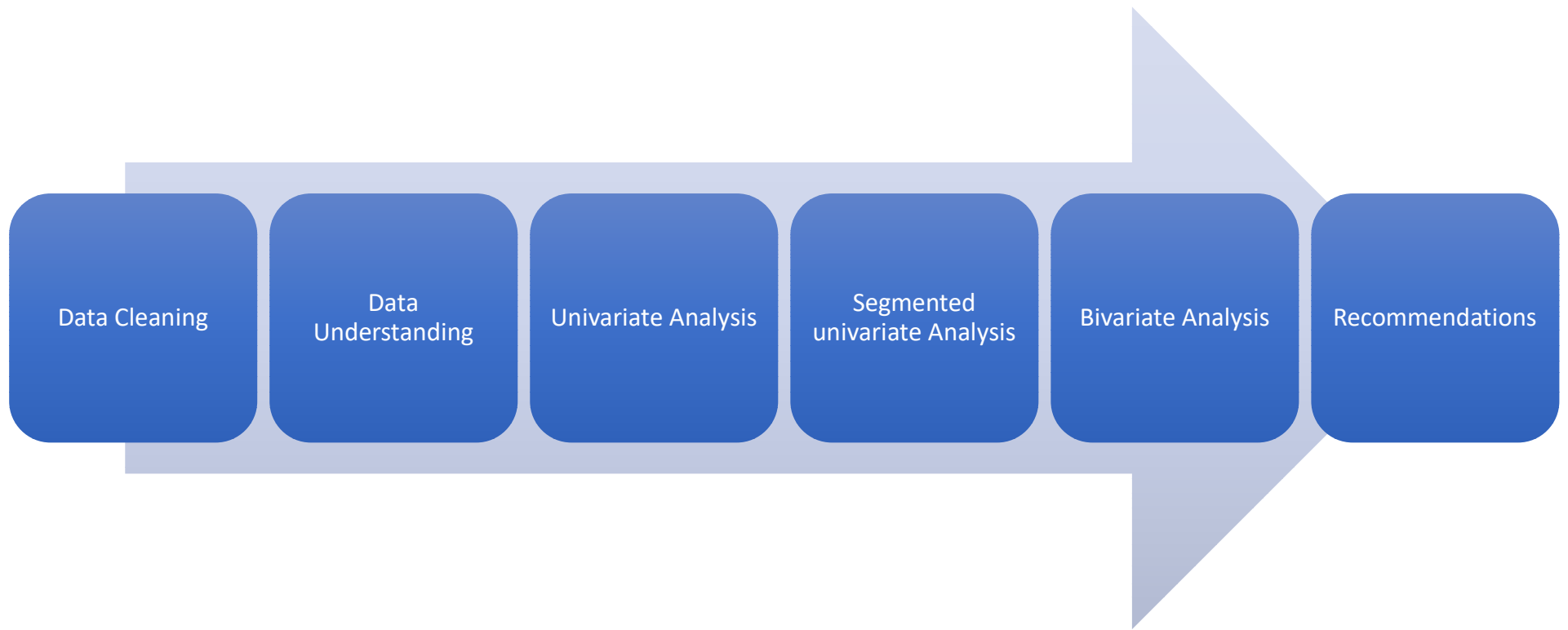   B. Suffering financial losses by approving loan applicants who are more likely to default.

Objective

- The objective of this analysis is to evaluate applicants' creditworthiness using variables like credit history, income, property, and employment stability to reduce these risks.

- *"In other words, 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. ---[UpGrad] "*
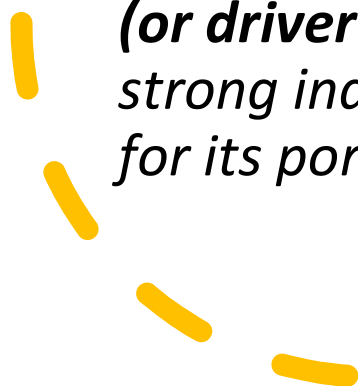
Exploratory Data Analysis Process

Data Cleaning → Data Understanding → Univariate Analysis → Segmented univariate Analysis → Bivariate Analysis → Recommendations

Detailed description for each Task

1. Data Cleaning - Handling Missing value , outlier, duplicates and removing unnecessary variable

2. Data Understanding- Working with the Data dictionary, checking all the variable and domain specific use. Explore the data set by summarizing the main character of the variables.

3. Univariate Analysis- Statistical Technique to summarize the characteristic of a single variable in a data set.

4. Segmented Univariate Analysis- Segmented univariate analysis makes it possible to examine a variable in greater detail by considering the differences between segments. It aids in finding patterns, trends, or anomalies that are unique to a subgroup but may not be seen when analysing the variable as a whole.

5. Bivariate Analysis- Examining the relationship between two variables in a data set and helps in understanding how changes in one variable relate to changes in another.

Basic Information About the data

- No of Columns: 111
- No of Rows: 39717
- No of Missing Values: 2263364
- No of Unique Values: 410681
- No Of Duplicates:0
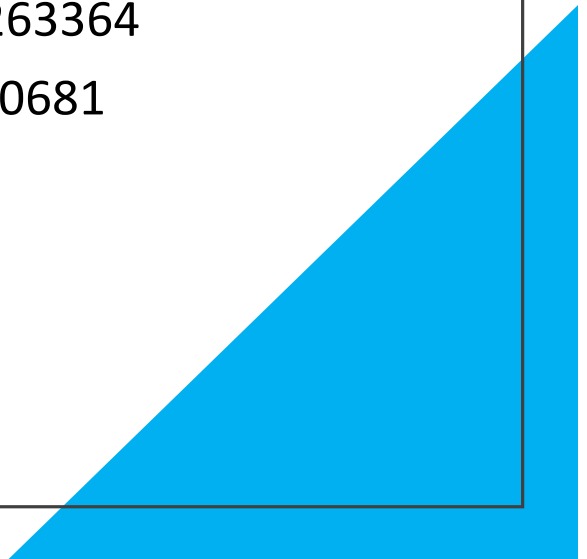
Derived Variables

1. issue_d_year
2. issue_d_month
3. issue_d_weekday
4. issue_d_weekday_name
5. earliest_cr_line_year
6. earliest_cr_line_month
7. approved_loan_amnt_ratio

Univariate Analysis

- From the summary statistics, we can understand the mean, standard deviation, min, max and 25%, 50% and 75% percentile for all the numerical columns.

- Next, we will see the univariate analysis for all these variables in following slides.

|	count	mean	std	min	25%	50%	75%	max
id	38577.00	676378.71	209263.90	54734.00	512033.00	656423.00	829146.00	1077501.00
loan_amnt	38577.00	11047.03	7348.44	500.00	5300.00	9600.00	15000.00	35000.00
funded_amnt	38577.00	10784.06	7090.31	500.00	5200.00	9550.00	15000.00	35000.00
funded_amnt_inv	38577.00	10222.48	7022.72	0.00	5000.00	8733.44	14000.00	35000.00
int_rate	38577.00	11.93	3.69	5.42	8.94	11.71	14.38	24.40
installment	38577.00	322.47	208.64	15.69	165.74	277.86	425.55	1305.19
annual_inc	38577.00	68777.97	64218.68	4000.00	40000.00	58868.00	82000.00	6000000.00
dti	38577.00	13.27	6.67	0.00	8.13	13.37	18.56	29.99
delinq_2yrs	38577.00	0.15	0.49	0.00	0.00	0.00	0.00	11.00
inq_last_6mths	38577.00	0.87	1.07	0.00	0.00	1.00	1.00	8.00
open_acc	38577.00	9.28	4.40	2.00	6.00	9.00	12.00	44.00
pub_rec	38577.00	0.06	0.24	0.00	0.00	0.00	0.00	4.00
revol_bal	38577.00	13289.49	15866.49	0.00	3650.00	8762.00	16912.00	149588.00
revol_util	38527.00	48.70	28.36	0.00	25.20	49.10	72.30	99.90
total_acc	38577.00	22.05	11.43	2.00	13.00	20.00	29.00	90.00
total_pymnt	38577.00	11866.97	8809.86	0.00	5513.50	9674.05	16136.95	58563.68
total_pymnt_inv	38577.00	11271.20	8700.85	0.00	5019.59	9066.66	15329.78	58563.68
last_pymnt_amnt	38577.00	2746.24	4494.65	0.00	217.42	568.26	3447.25	36115.20
pub_rec_bankruptcies	37880.00	0.04	0.20	0.00	0.00	0.00	0.00	2.00
approved_loan_amnt_ratio	38577.00	93.79	17.33	0.00	98.25	99.93	100.00	100.00

Loan Amount

The listed amount of the loan applied for by the borrower.

- **Observations**:
  From the plots, more number of people took loan amount of 9600, and also median of distribution is 9600. And very few people took more than 30000 loan amount.
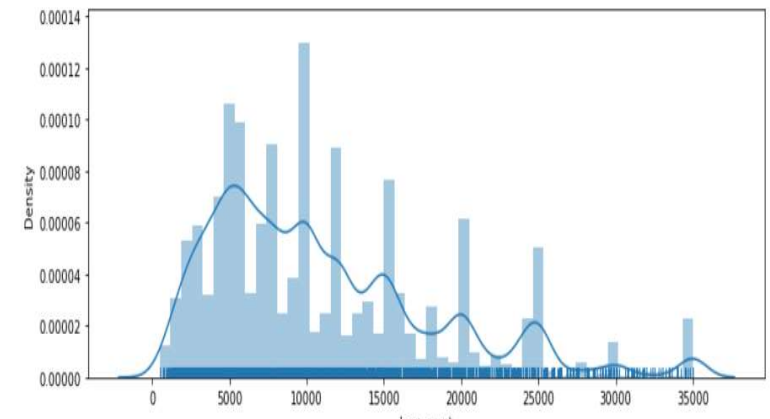
```
count    38577.00
mean     11047.03
std       7348.44
min        500.00
5%        2400.00
10%       3090.00
25%       5300.00
50%       9600.00
75%      15000.00
90%      21600.00
95%      25000.00
99%      35000.00
max      35000.00
```

Loan Status

Observations:

85% of borrowers have paid the loan fully where as 15% have defaulted the loan but if we observe the Loan status distribution where loan amount is greater than 30K , 23% have defaulted loan .

loan_status Distribution

loan_status Distribution_data_loan_amount_GreaterThan30K
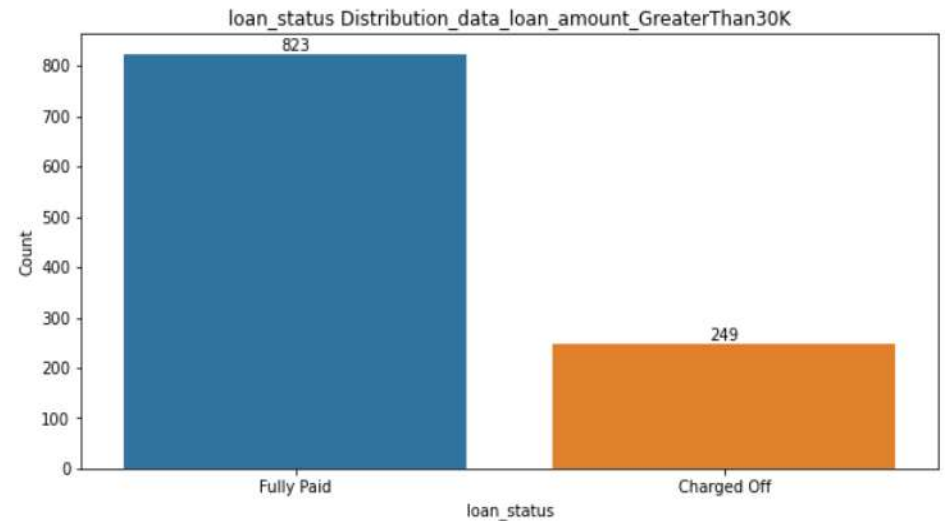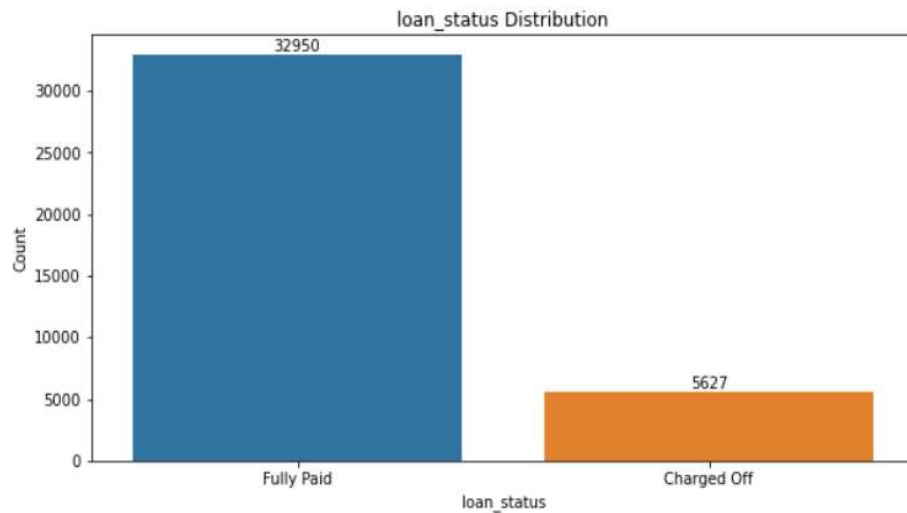
DTI - Debt To Income ratio

- **Observations:** Looks like there are no outliers and the distribution is very much similar to normal distribution. This is a good sign that all the loans were given to borrowers who have Debt To Income ratio less than 30.

Debt to Income distribution
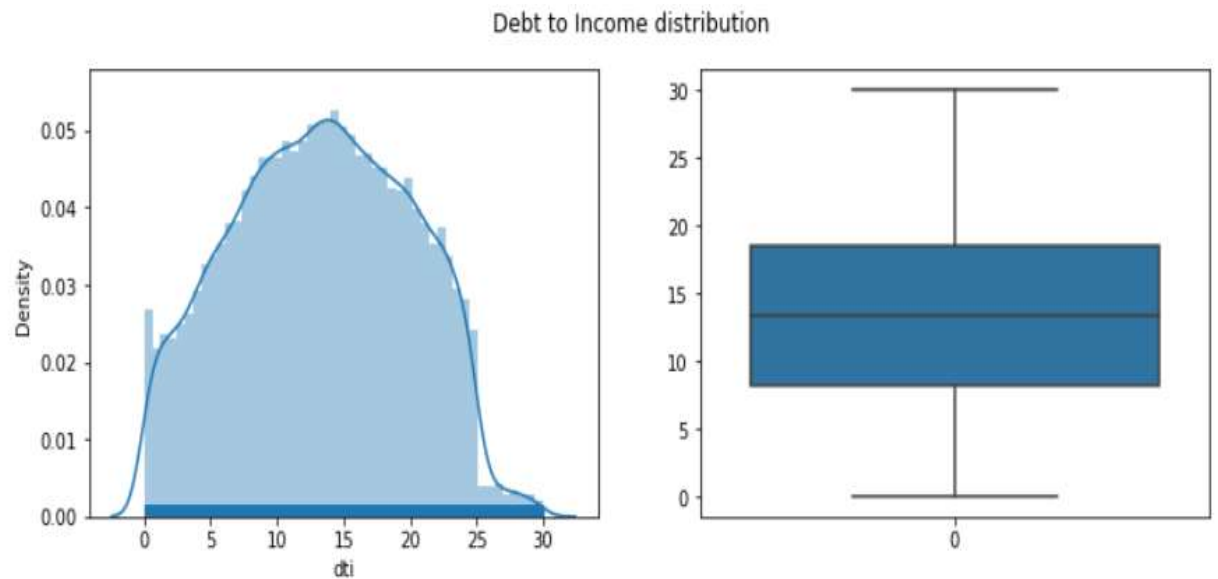
Loan payment Term

- **Observations:**

- if we see the overall data, borrowers have taken 36 months tenure more than 60 months but if we see the loan amount greater than 30K, that subset of data, borrowers took 60 months tenure mostly.

term Distribution

term Distribution_data_loan_amount_GreaterThan30K
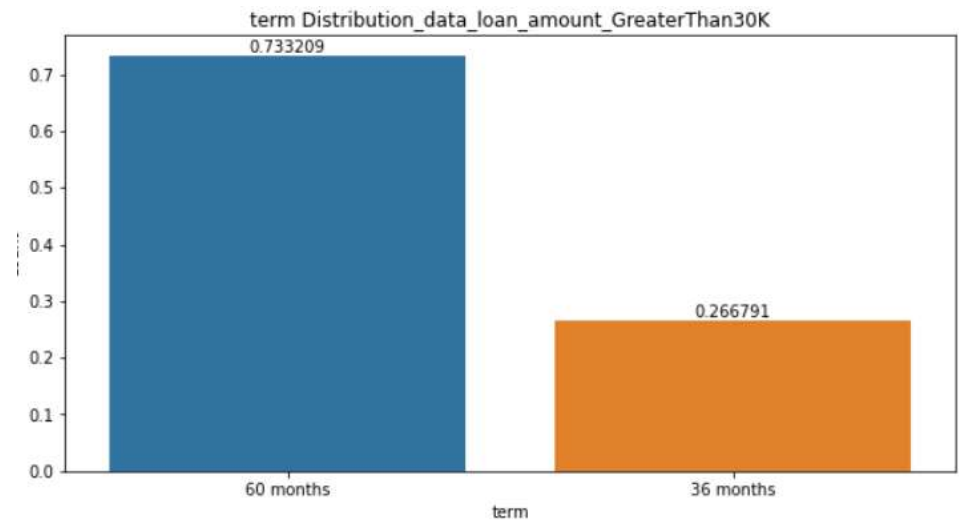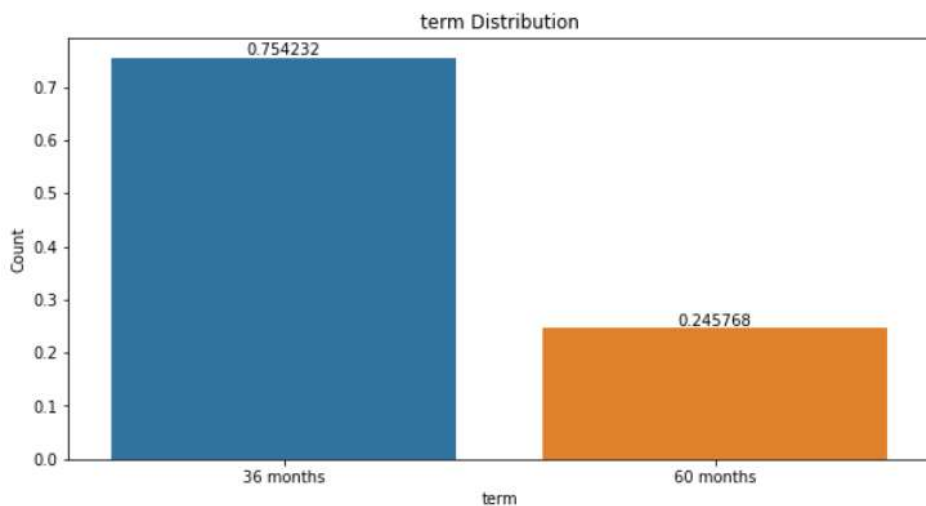
Employment length

- **Observation:**
- Borrowers are mostly 10+ years employment length and for higher loan amount most of the borrowers are having higher employment length

emp_length Distribution_data_loan_amount_GreaterThan30K

emp_length Distribution
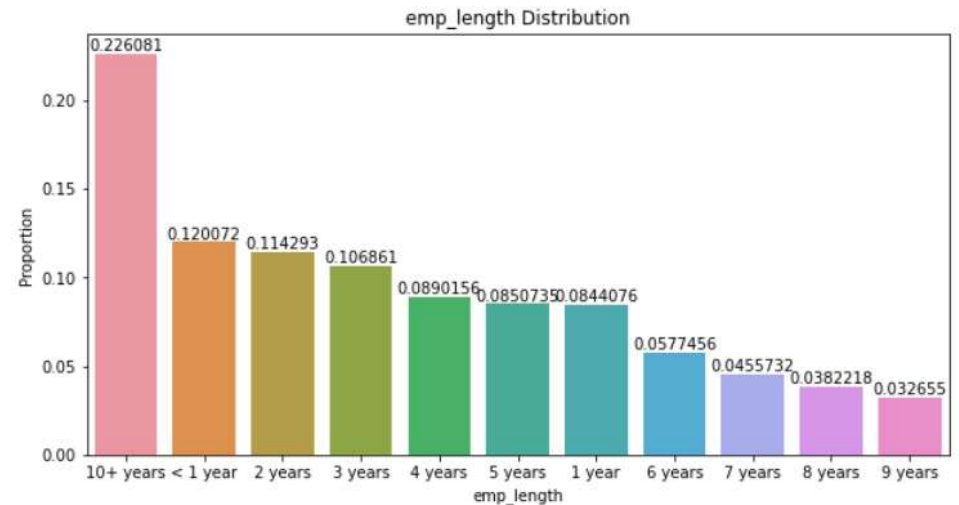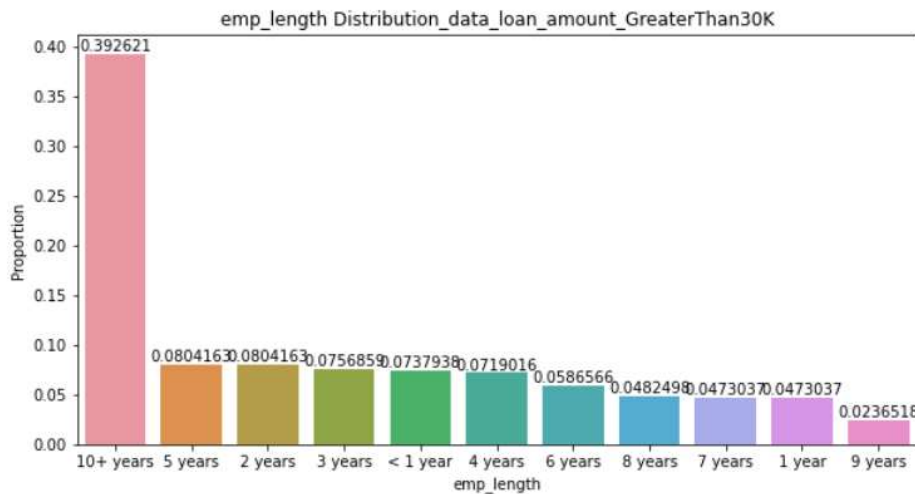
Home Ownership

- **Observations:**
- The loan borrowers are mostly having rented and mortgage houses.

home_ownership Distribution
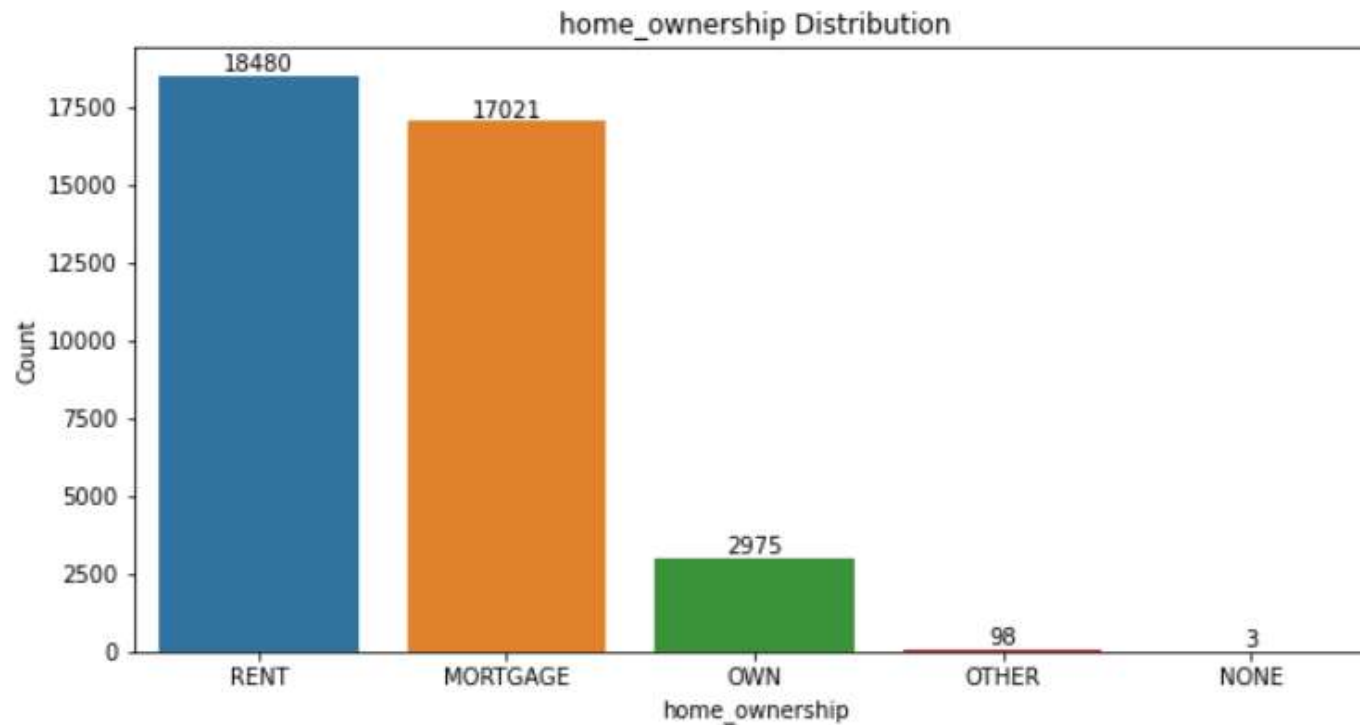
Loan Application Verification

- **Observation:**
- Most of the cases, loans are not verified but in case of higher loan amount, applications are mostly verified.

verification_status Distribution

verification_status Distribution_data_loan_amount_GreaterThan30K
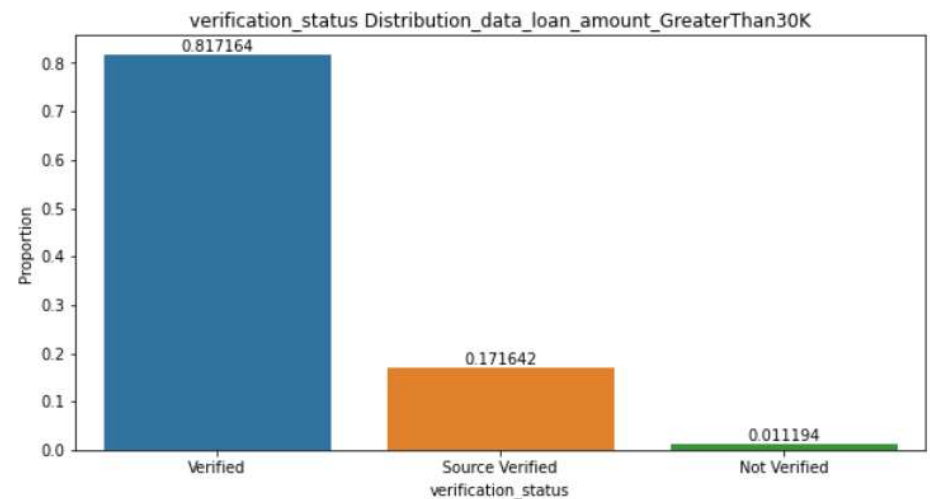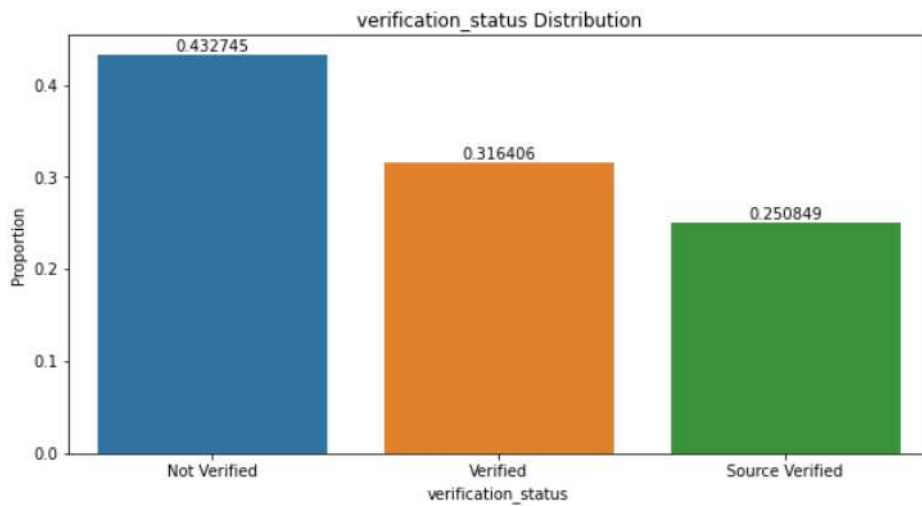
Applicants Grade

- **Observations:**
- Most borrowers fall under A and B grades then other grades but if the loan amount greater than 30K, then most of the borrowers fall in Grade E
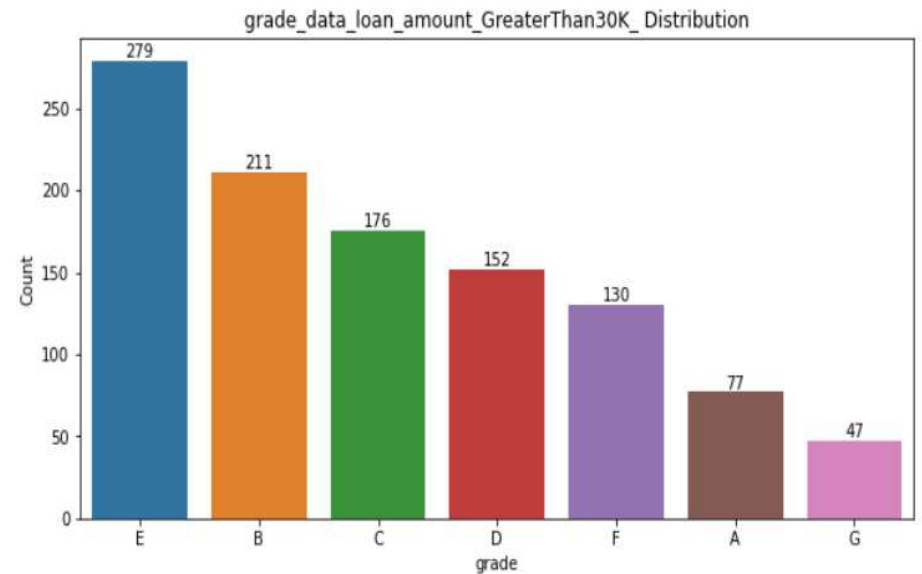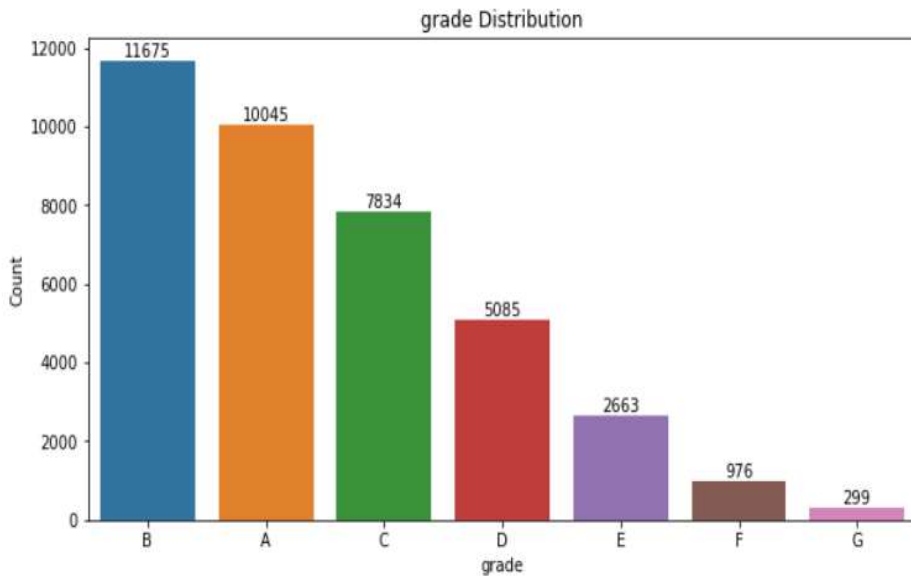
Loan Applicant's State

- **Observation:**¶
- For overall data set, top, Applicants are mostly from CA, NY and FL , whereas for loan amount greater than 30K, Top 3 locations are CA, TX and NY respectively

addr_state Distribution_Top 10

addr_state _data_loan_amount_GreaterThan30K_Distribution_Top 10
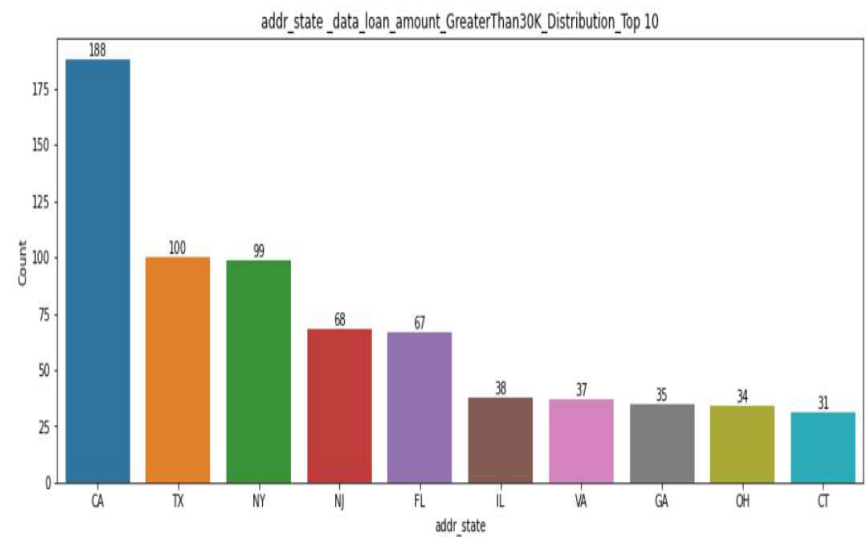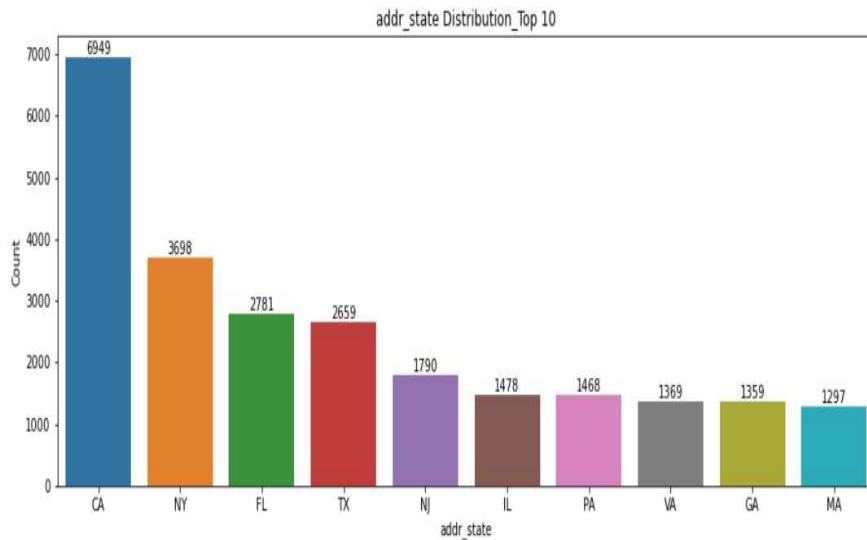
- **Loan Issue Year-** The lending club(LC) has doubling loan issues every year.
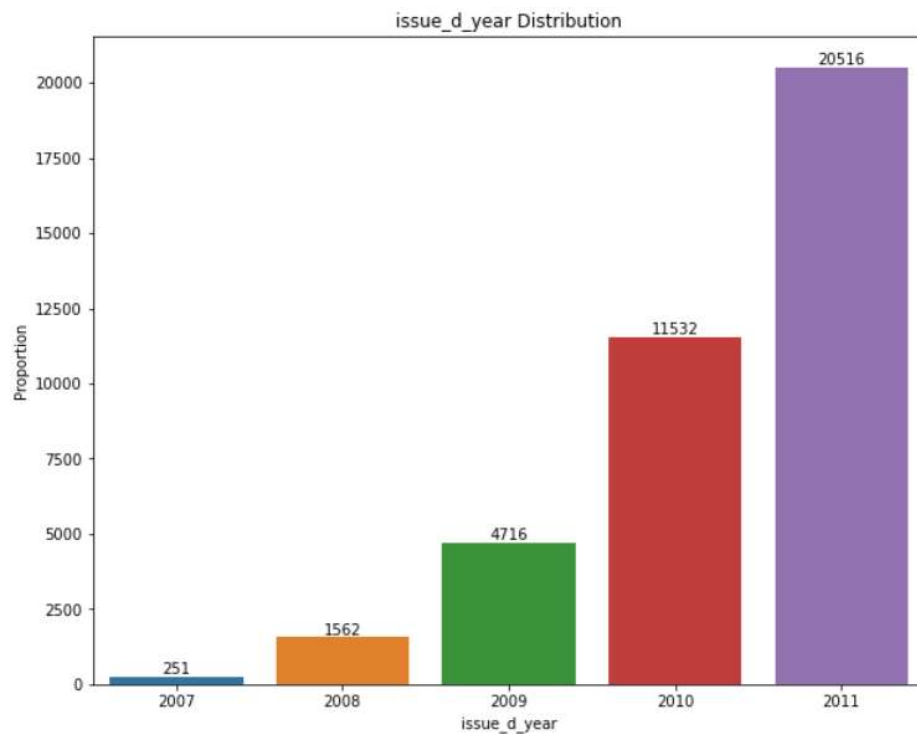
- **Loan Issue Month-**There are more issues of loan in last 3 months every end of the year i.e., Oct, Nov and Dec.

issue_d_year Distribution

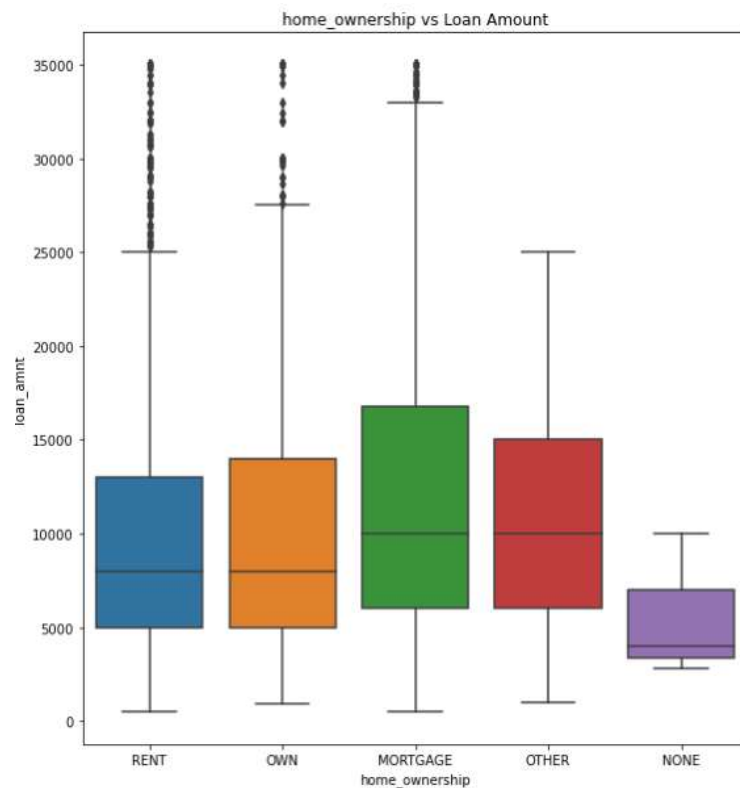issue_d_month Distribution
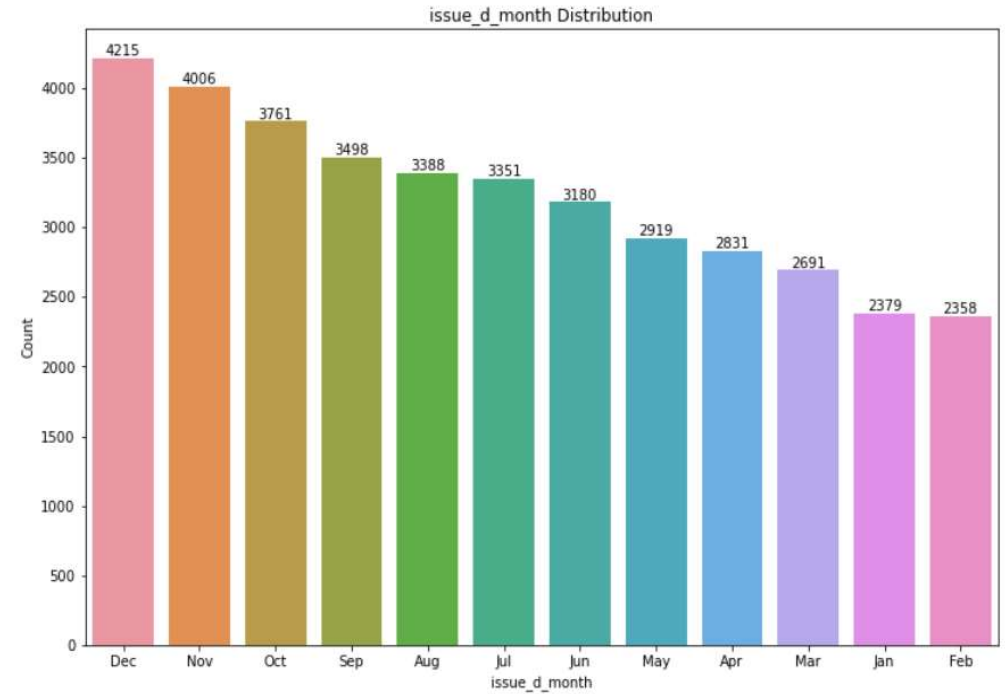
SEGMENTED UNIVARIATE ANALYSIS

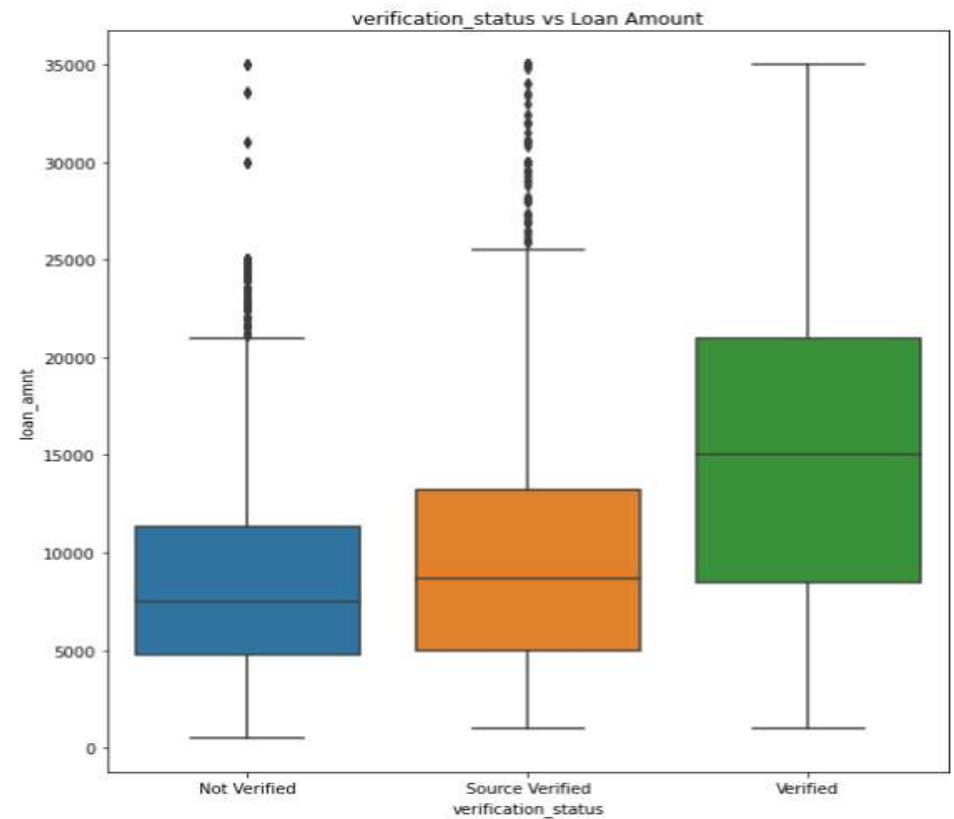- **Terms vs Loan Amount**-Higher amount loans have high tenure i.e, 60 months.
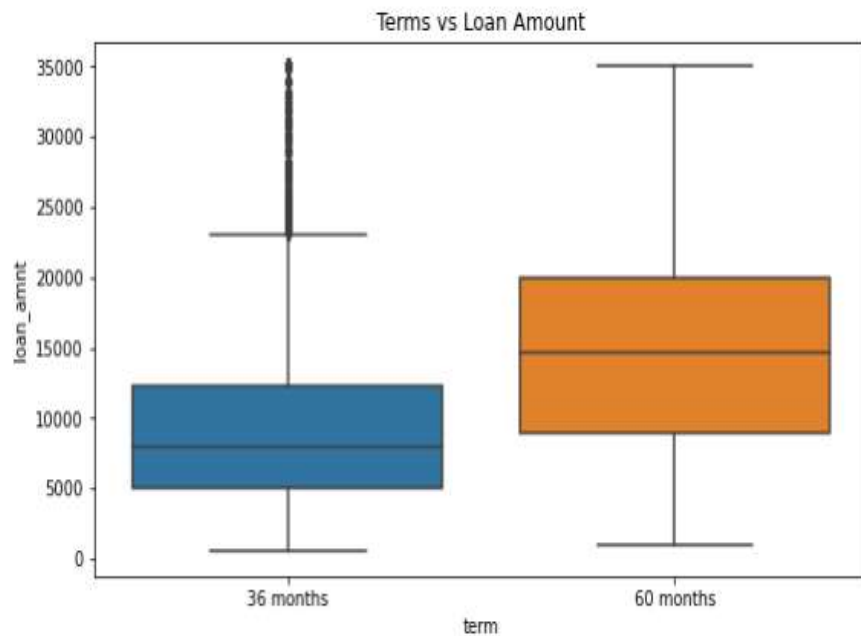
- **Grade vs Loan Amount**-Grade 'G' have taken max loan amount. From A to G, the loan amount is increasing.
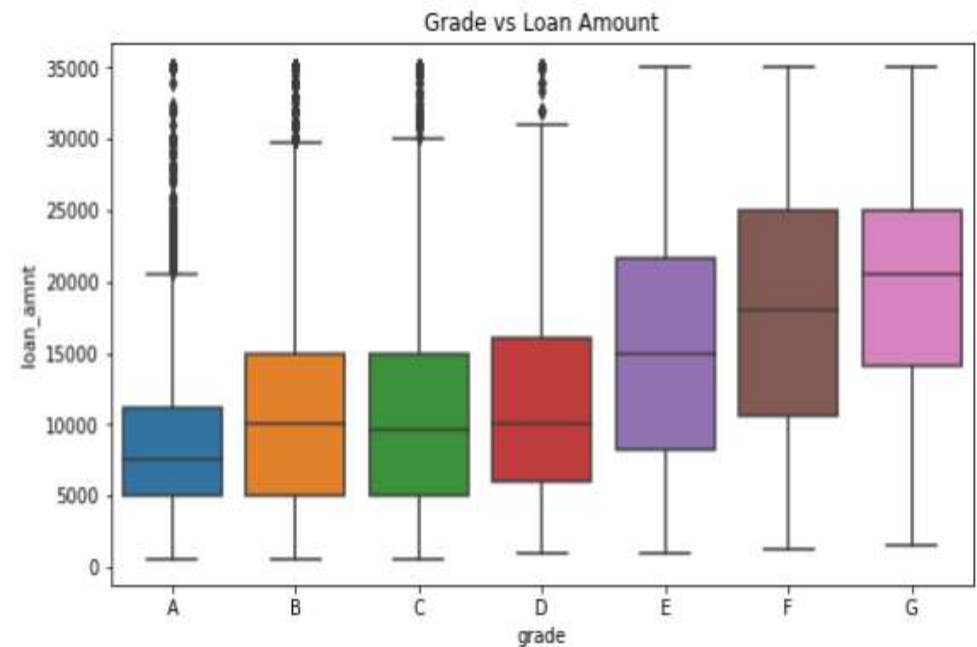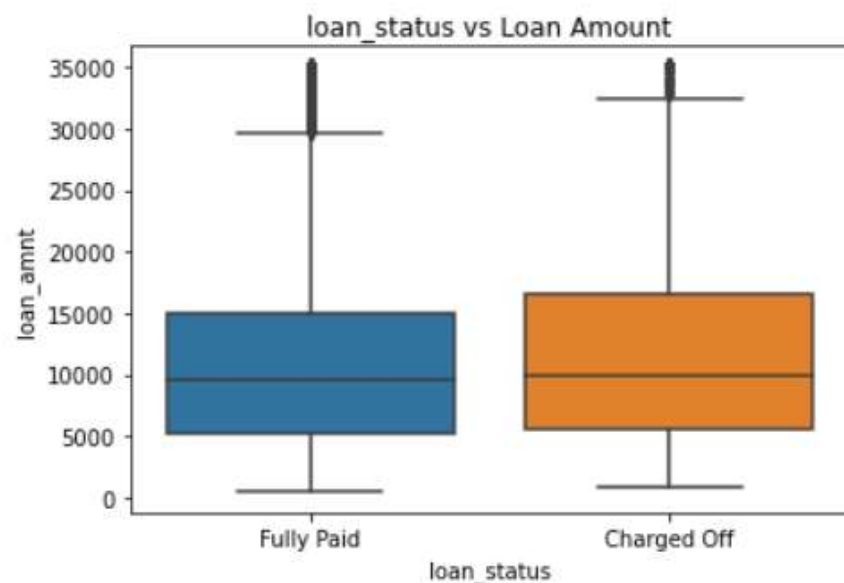
- **Home_Ownership vs Loan Amount-** More borrowers are from MORTGAGE and also the median loan amount is high for MORTGAGE owned borrowers.

- **Verification_status vs Loan Amount**-Median Loan amount for verified borrowers are the highest.

- **Loan Status vs Loan Amount-** Charged Off loans have higher amounts than Fully Paid ones.

- **Purpose vs Loan Amount-** Highest Mean value for Loan taken is for small_business followed by house and debt_reconsolidation.


loan_status vs Loan Amount

purpose	loan_amnt
small_business	13365.11
house	12753.41
debt_consolidation	12503.46
credit_card	11571.41
home_improvement	11308.00
wedding	9755.86
renewable_energy	8537.99
medical	8122.76
major_purchase	8072.60
other	7936.42
educational	6810.46
car	6746.73
moving	6568.01
vacation	5417.60

Emp_Length vs Loan Amount

1. With higher employment length borrowers have high tendency to take high loan amount.
2. In total data, 23% borrowers are having employment length more than 10 years.

Borrower % as employment length group

```
10+ years    0.23
< 1 year     0.12
2 years      0.11
3 years      0.11
4 years      0.09
5 years      0.09
1 year       0.08
6 years      0.06
7 years      0.05
8 years      0.04
9 years      0.03
```


emp_length vs Loan Amount
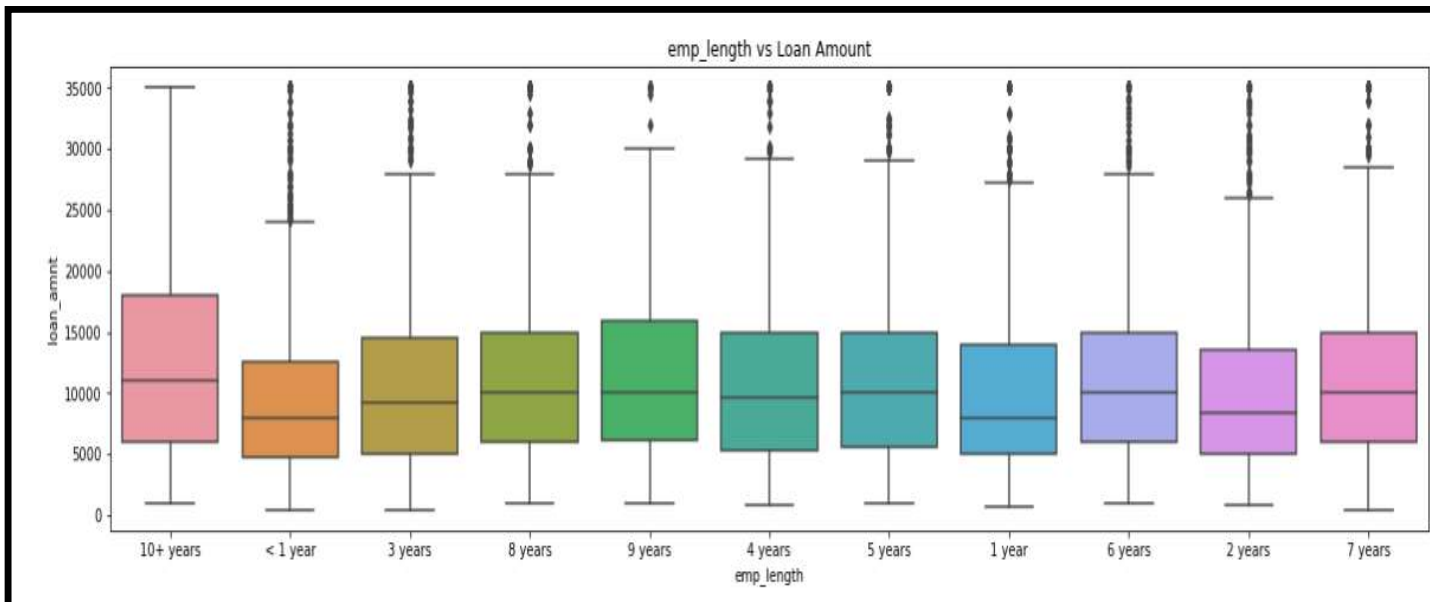
```
                loan_amnt
emp_length
10+ years       12808.10
9 years         11903.30
8 years         11741.55
7 years         11565.11
6 years         11372.50
5 years         11114.08
4 years         10839.33
3 years         10636.75
2 years         10124.56
1 year          10056.82
< 1 year         9571.94
```
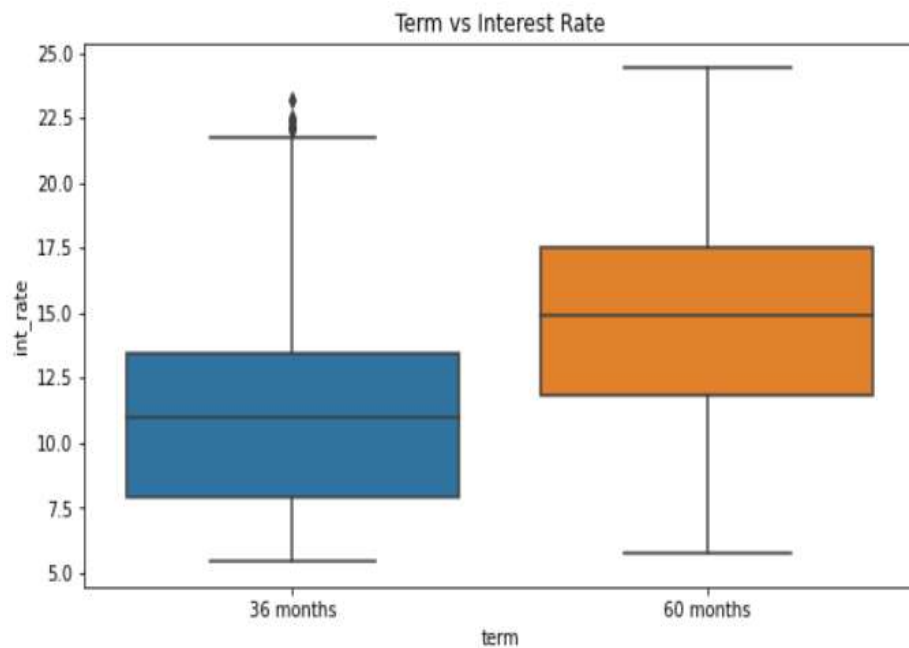
Terms vs funded amount

For 36 months terms, there are a few borrowers who received high loan amount which are shown as outliers.
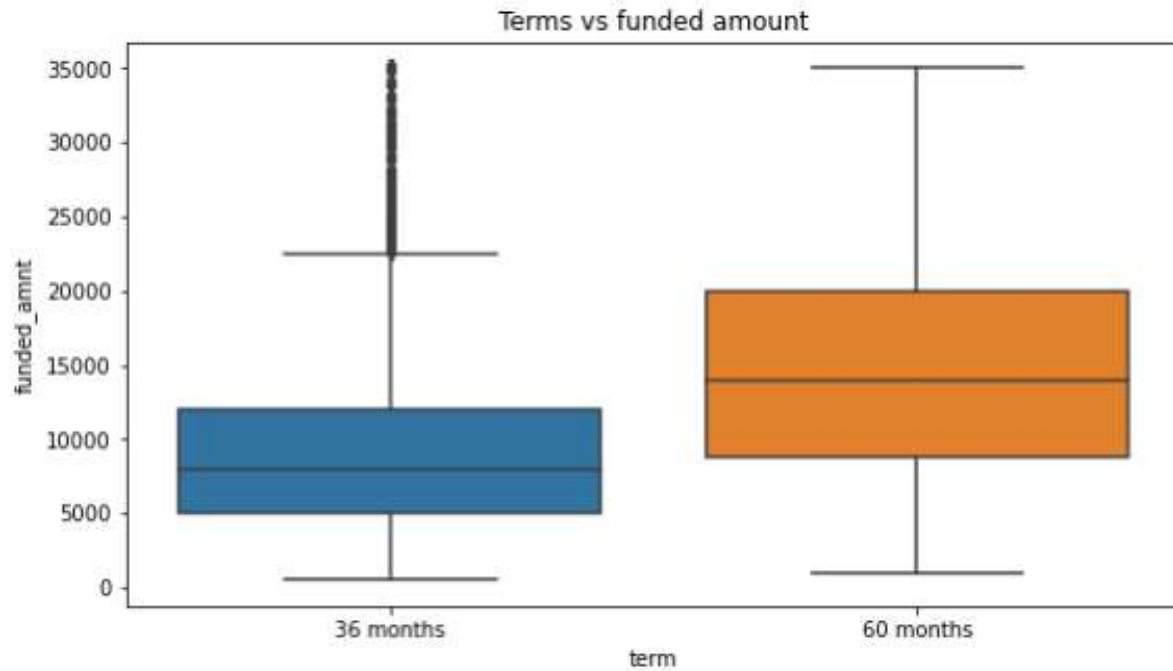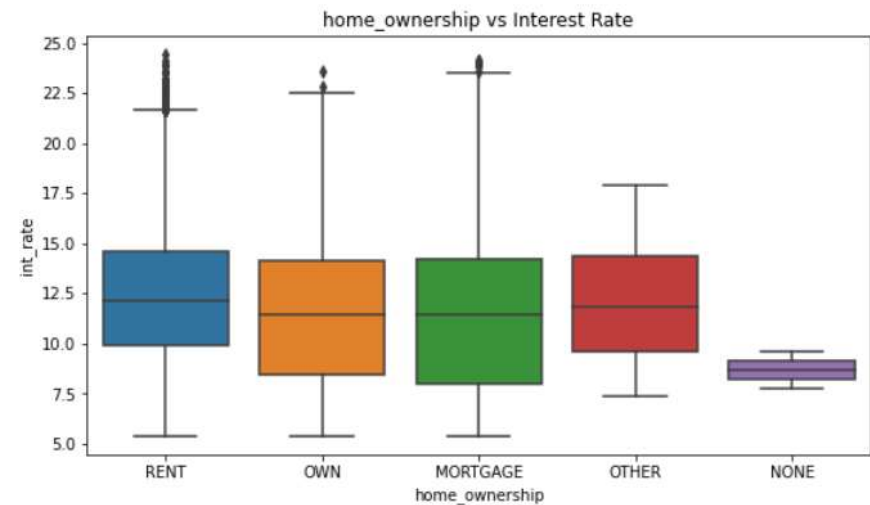
- **Term vs Interest Rate-** The interest rates are higher for higher tenure loans.

- **Home Ownership vs Interest Rate-** Borrowers having own house or kept mortgage for loan, they got less interest rate since Banks have the security.
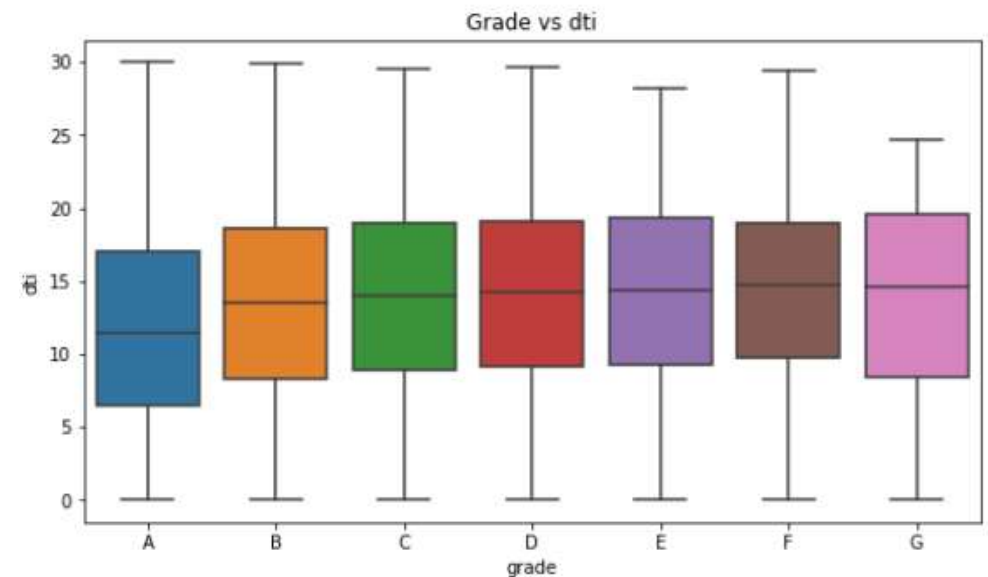

Term vs Interest Rate

home_ownership vs Interest Rate

```
                       int_rate
home_ownership
RENT                     12.22
OTHER                    12.04
OWN                      11.69
MORTGAGE                 11.66
NONE                      8.70
```

- **Verification Status vs Interest Rate-** Not verified loan applications get less interest which may be risky for the Bank
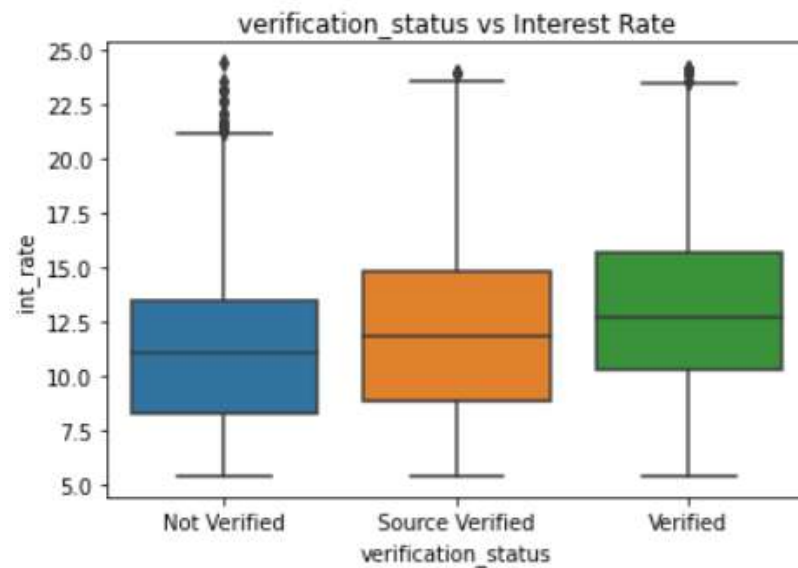
- **Grades vs DTI** -A Grade borrowers are having low DTI than Other grades. DTI should be low for having high repayment percentage.
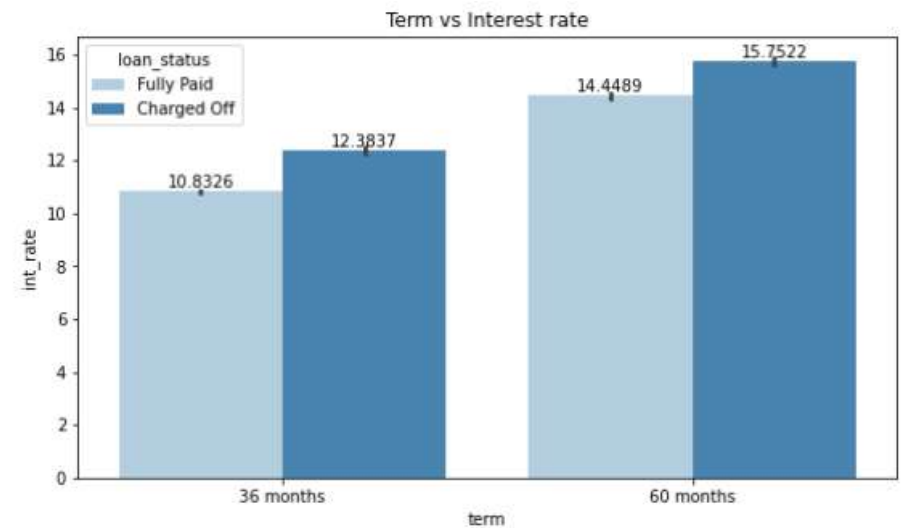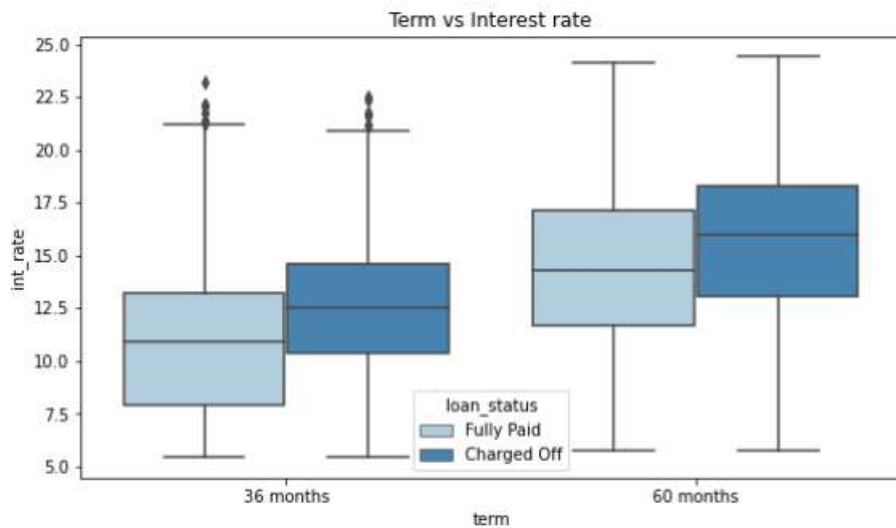

verification_status vs Interest Rate

Grade vs dti

- **'pub_rec_bankruptcies' vs 'loan_status'-** Borrowers who are having 0 record of bankruptcies, but failed to pay the loan, are the risky borrowers.
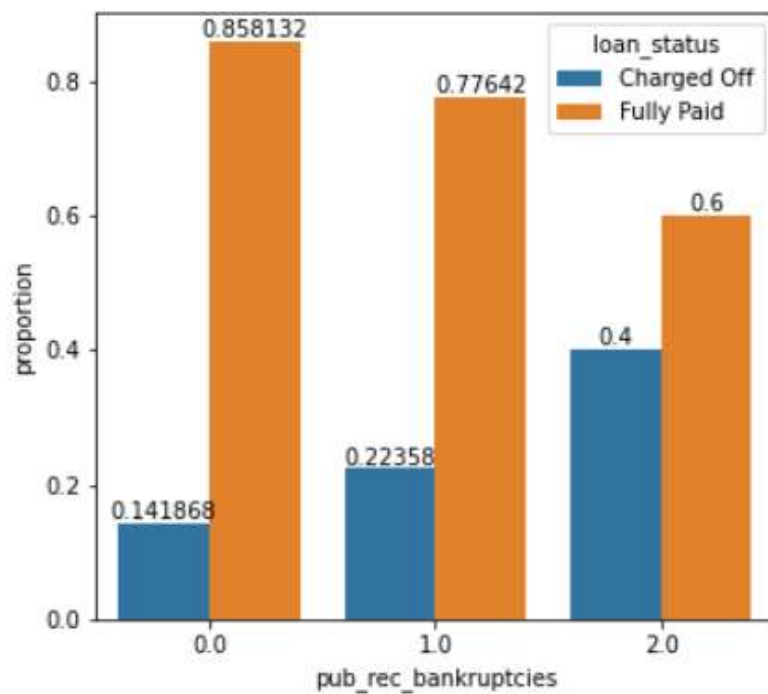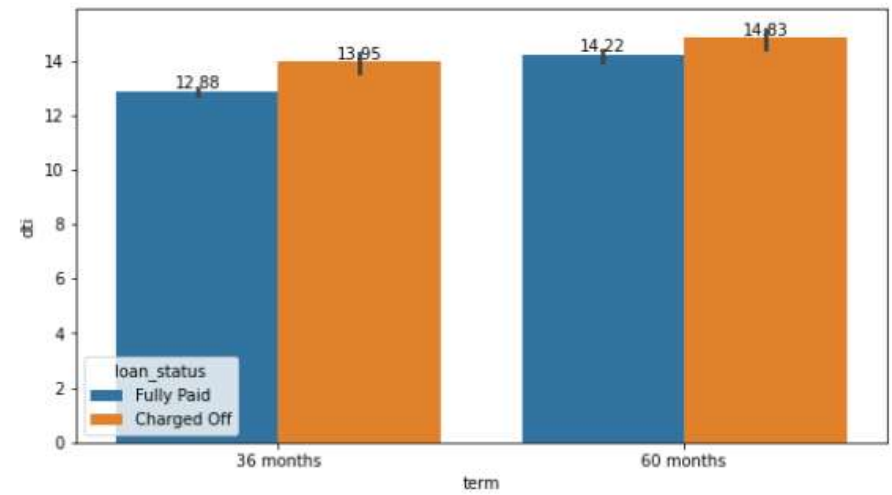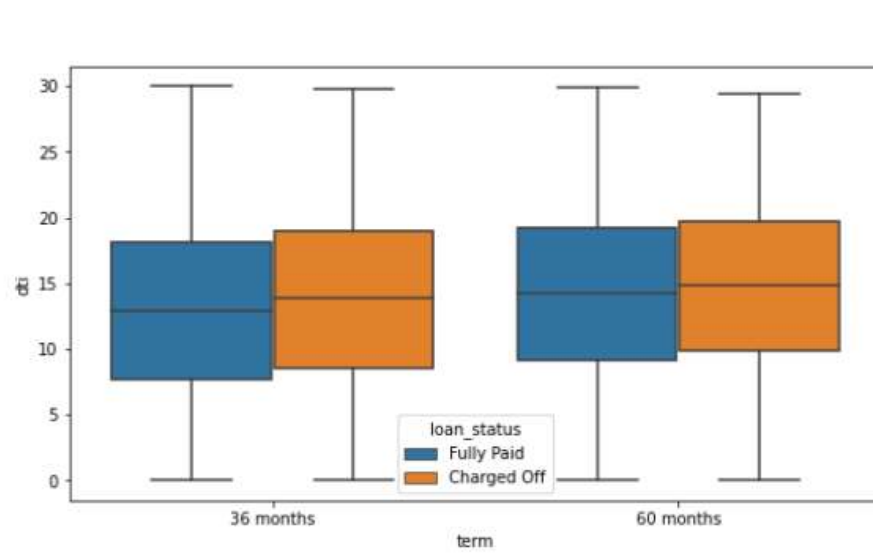
Bivariate Analysis

Term vs Interest rate

- For higher interest rates the default rate is higher in both 36- and 60-months tenure.

Term vs DTI

1. For both the tenure, DTI is high for Charged off loan

Term vs DTI

Loan Amount, Int Rate vs Loan Status

Though for in any emp_length group, loan amount difference is not much high for Fully paid and Charged off loans but interest rate for all emp_length group for charged off loan is high. Irrespective of employment length loans with more interest rates got defaulted more.
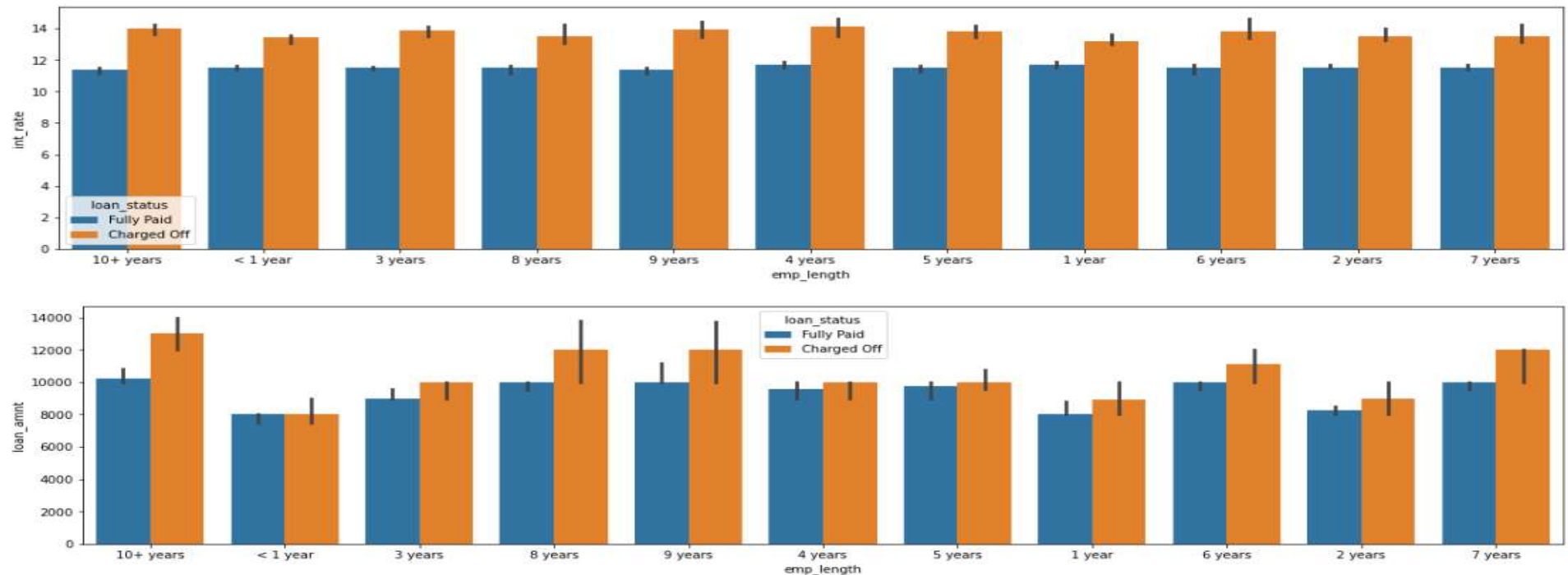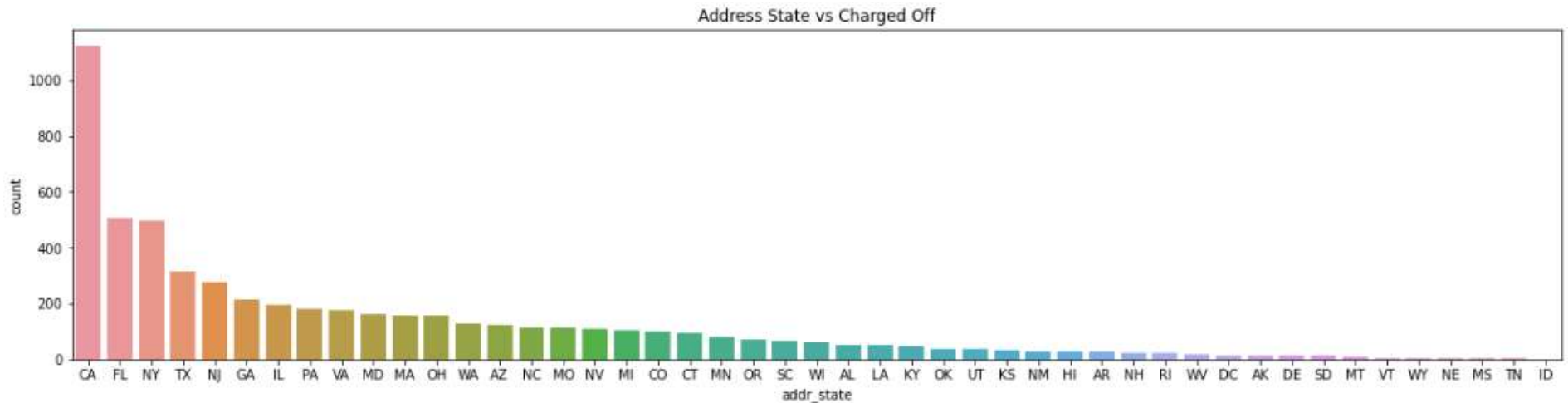
'Home Ownership vs Loan status'

Home Ownership other is having high charged off loan percentage

Home Ownership vs Loan status
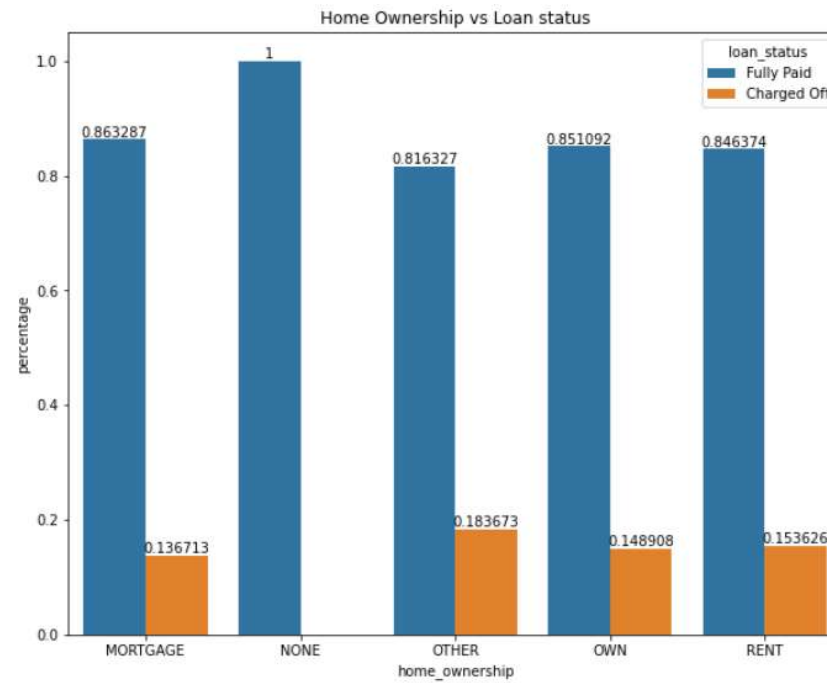
Address State vs Charged Off

- More number of borrowers defaulted in CA , FL and NY states.
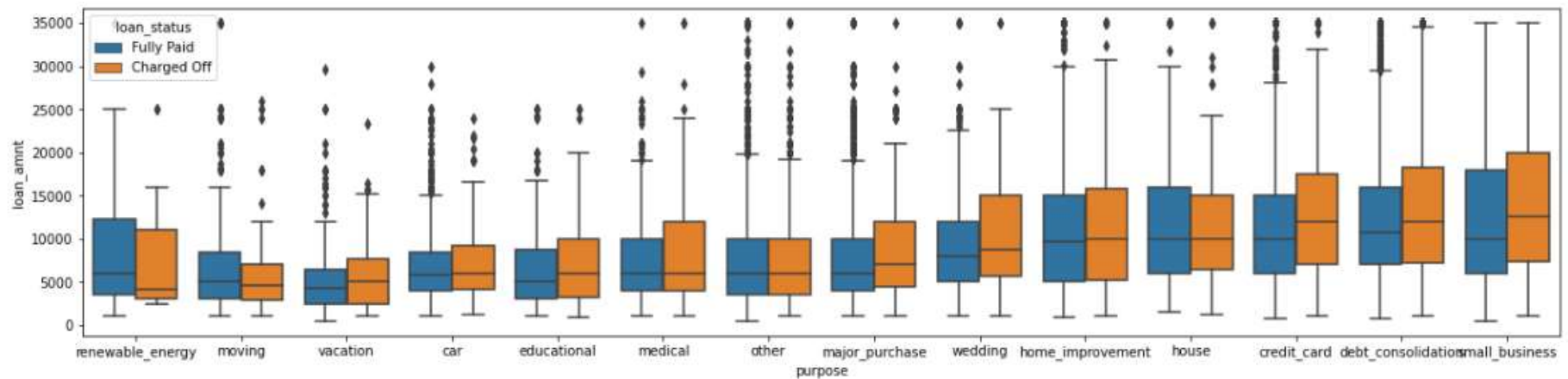

Address State vs Charged Off

Verification Status vs interest rate

- Irrespective of verification status higher interest rates are incurring default of loan. There is slight increase in the dti mean for defaulted laons for all the verification status categories.
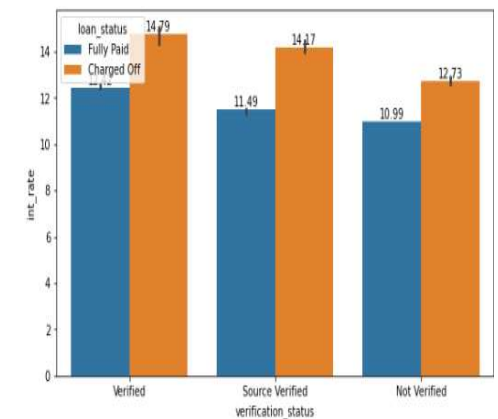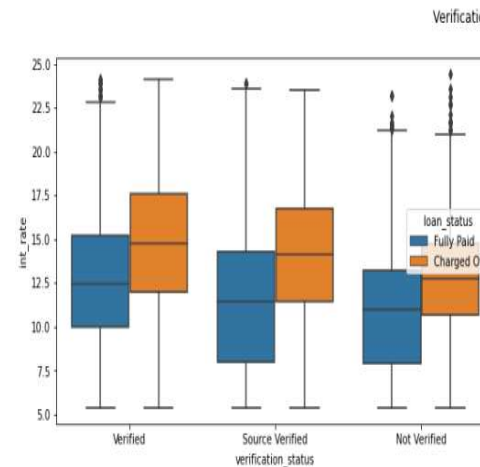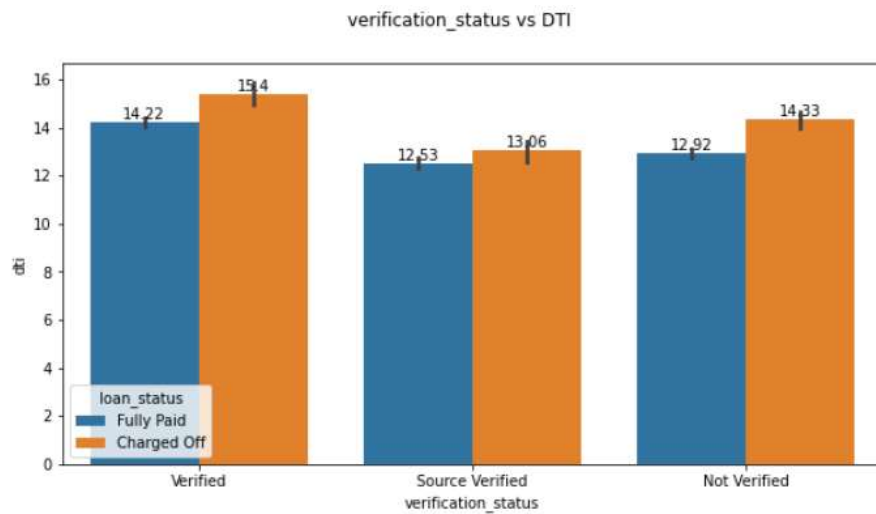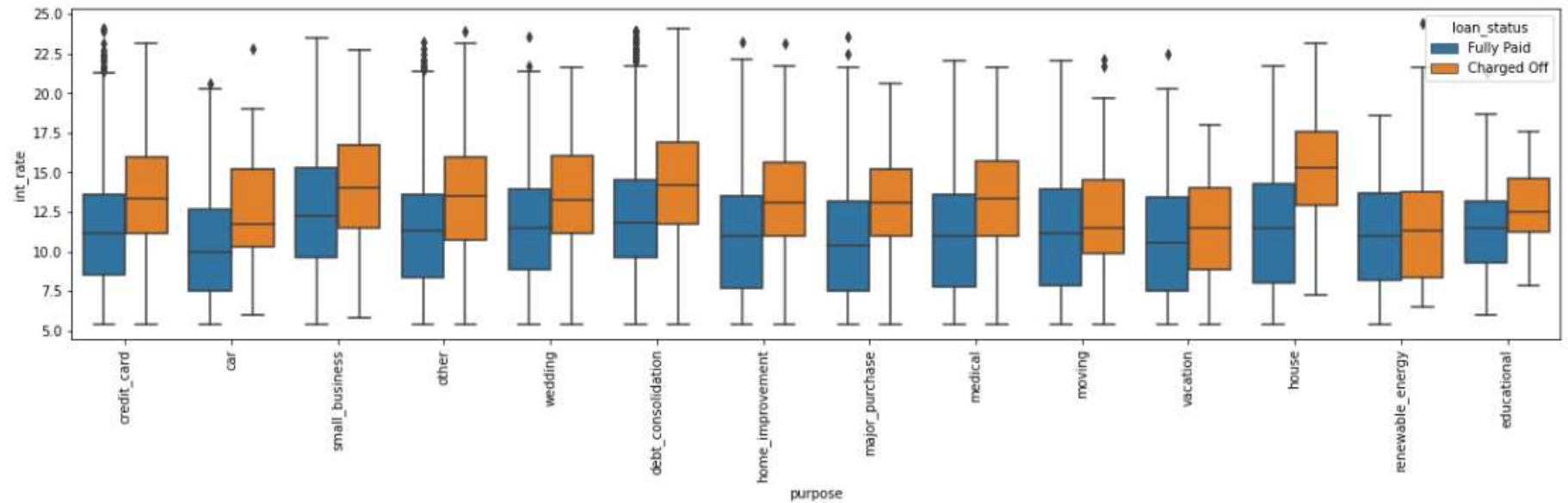
Purpose vs Loan Status

- Small Business has more defaults when the loan amount is also high.

Purpose vs Interest Rate

- Home loans with high interest rates are mostly defaulted. Even small business and debt consolidation has similar observation.

Conclusion

- Lending Club should scale back its high-interest loans with a 60-month repayment term since they are more likely to default.

- Small business loans default more frequently. The loans from Lending Club should cease or be scaled back for them.

- People are more likely to file for bankruptcy if they have more publicly available negative records about them.

- Grades are a reliable indicator of defaulters. Before making loans to Low grade (G to A) borrowers, Lending Club should investigate more information from the borrowers.

- To turn a profit, Lending Club should limit the quantity of loans it extends to residents of California, Florida, and New York.

- Homeowners with mortgages are taking out larger loans and defaulting on the authorised loans. When a loan request exceeds $12,000, Lending Club should stop making loans to this category.