


Lending Club Case Study



GROUP MEMBER:

1. ISHITAARORA

PROBLEM STATEMENT



- IDENTIFY THESE RISKY LOAN APPLICANTS, THEN SUCH LOANS CAN BE REDUCED THEREBY CUTTING DOWN THE AMOUNT OF CREDIT LOSS.
- UNDERSTAND THE **DRIVING FACTORS (OR DRIVER VARIABLES)** BEHIND LOAN DEFAULT, I.E. THE VARIABLES WHICH ARE STRONG INDICATORS OF DEFAULT.
- UTILISE THIS KNOWLEDGE FOR COMPANY'S PORTFOLIO AND RISK ASSESSMENT.

DATA ANALYSIS PIPELINE

Data Understanding

Remove redundant columns if they don't contribute to data analysis

Data Cleaning

Remove columns & rows with huge percentage of missing values

Data Enhancing

Create Numerical columns from categorical data

Univariate and Segmented Univariate Analysis

Bivariate Analysis

DATA CLEANING

- **DELETE COLUMNS:** DELETE UNNECESSARY COLUMNS WHICH DO NOT ADD ANY INFORMATION FOR DATA ANALYSIS.
- **REMOVE OUTLIERS:** REMOVE HIGH AND LOW VALUES THAT WOULD DISPROPORTIONATELY AFFECT THE RESULT OF YOUR ANALYSIS.
- **MISSING VALUES:** REMOVE COLUMNS AND ROWS WITH TOO HIGH PERCENTAGE OF MISSING VALUES
- **DUPLICATE DATA:** REMOVE IDENTICAL ROWS, REMOVE ROWS WHERE SOME COLUMNS ARE IDENTICAL.
- **FILTER ROWS:** FILTER BY SEGMENT, FILTER BY DATE PERIOD TO GET ONLY THE ROWS RELEVANT TO THE ANALYSIS.
- **CREATED TWO NEW COLUMNS:**
 - Financial Gain Or Loss
 - Categories_Loan_Amount
 - Categories_Annual_Income
 - Categories_Interest_Rate
 - Category_dti

ANALYSIS

1) DURING **UNIVARIATE ANALYSIS** WE HAVE CREATED:

- HISTOGRAMS AND BAR CHARTS TO CHECK OUT THE DISTRIBUTION OF ALL THE DRIVER VARIABLES
- BOX PLOTS TO DETECT THE OUTLIERS

2) DURING **BIVARIATE ANALYSIS**:

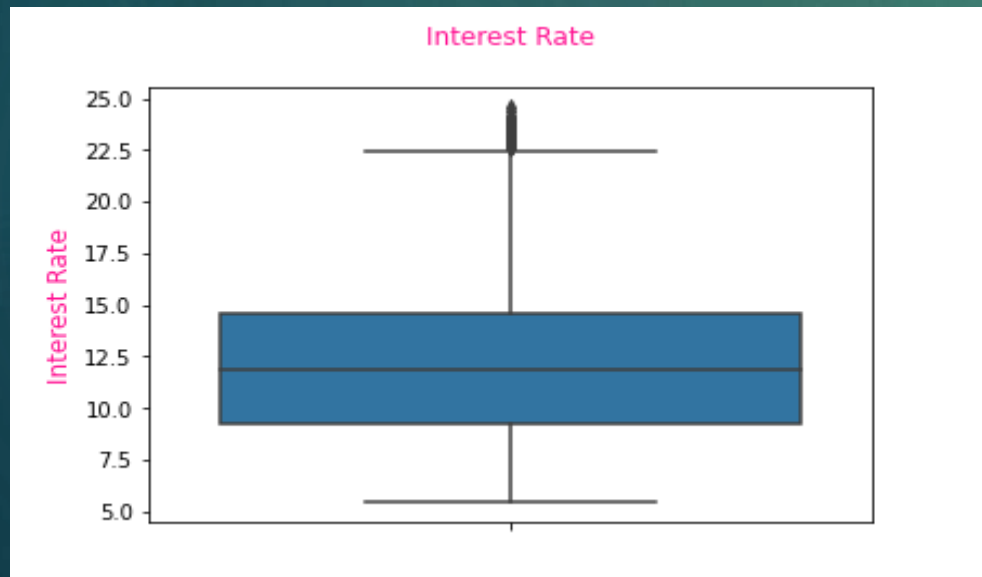
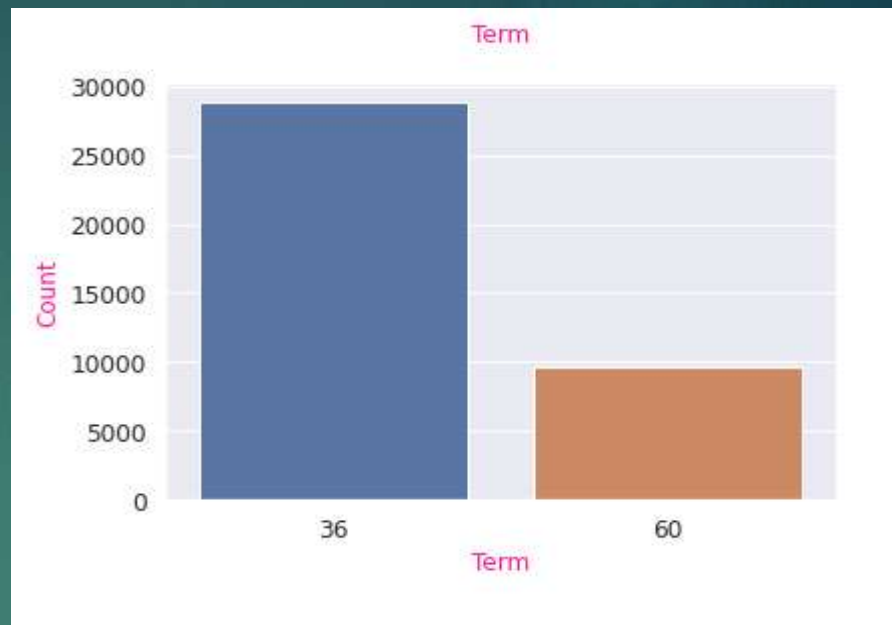
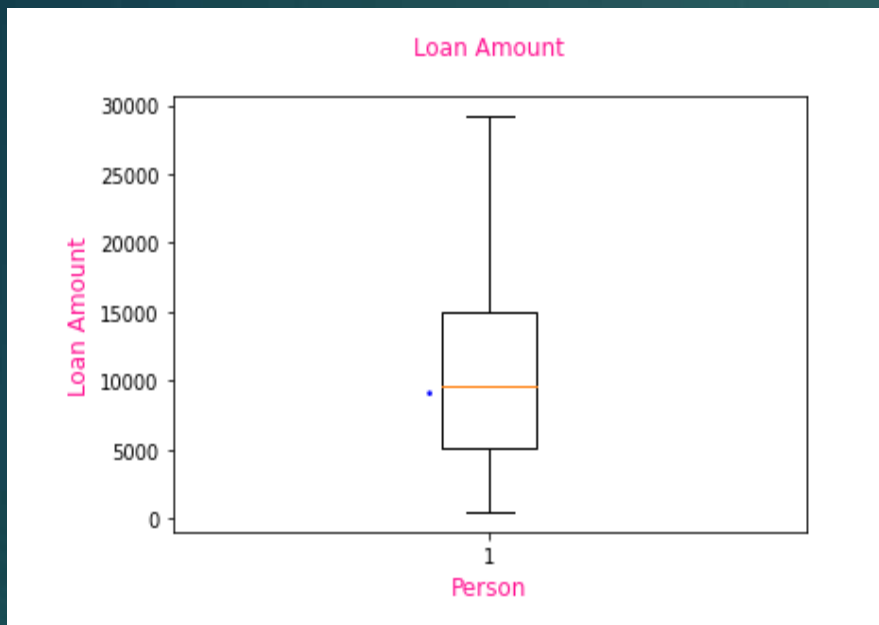
- PERFORMED THE MULTIVARIATE ANALYSIS TO UNDERSTAND HOW DIFFERENT VARIABLES INTERACT WITH EACH OTHER.
 - **TYPES OF ANALYSIS PERFORMED:**

Loan analysis:

- Loan status
- Loan grade
- Loan amount
- Loan Tenure

Profile of loan applicants analysis

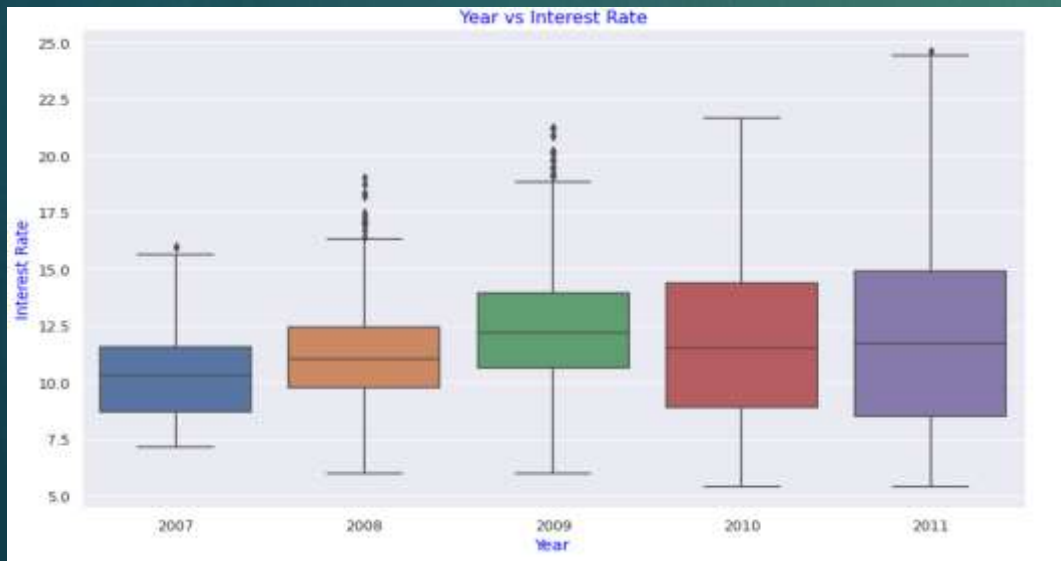
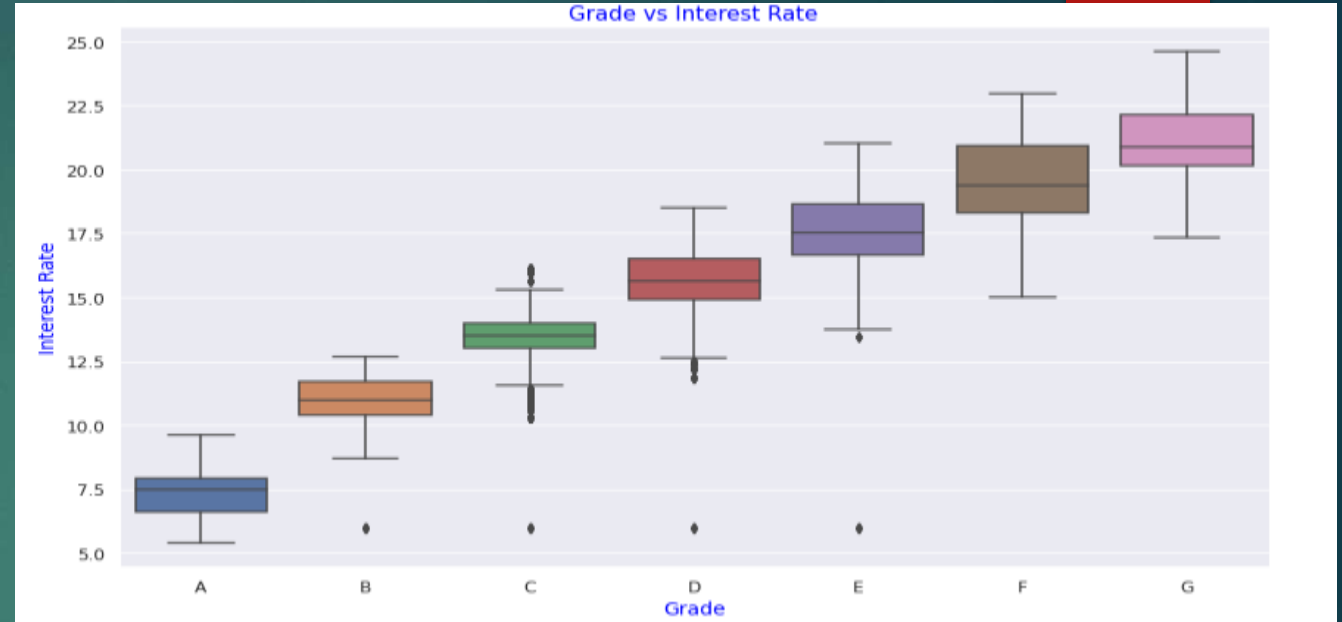
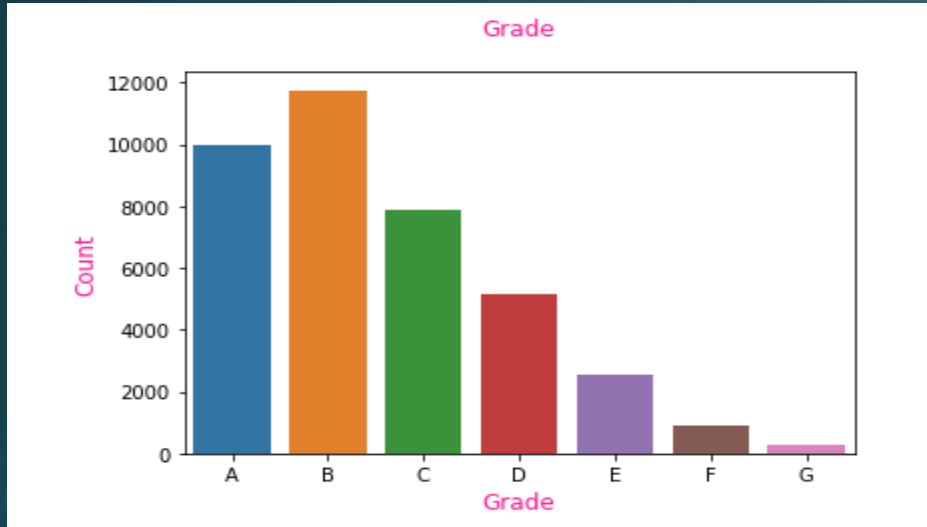
- Work experience
- Income Analysis
- Country/State
- Loan purpose
- Home ownership



OBSERVATIONS:

- THE LOAN AMOUNT VARIES FROM 0 TO 35,000 HAVING MEAN OF 10,000.
- PEOPLE ARE SELECTING 36 MONTHS OF PERIOD MORE THAN 60 MONTHS TIME SPAN FOR FILLING THE LOAN AMOUNT.
- INTEREST RATE WHERE MEAN IS 12.1%

LOAN GRADE ANALYSIS

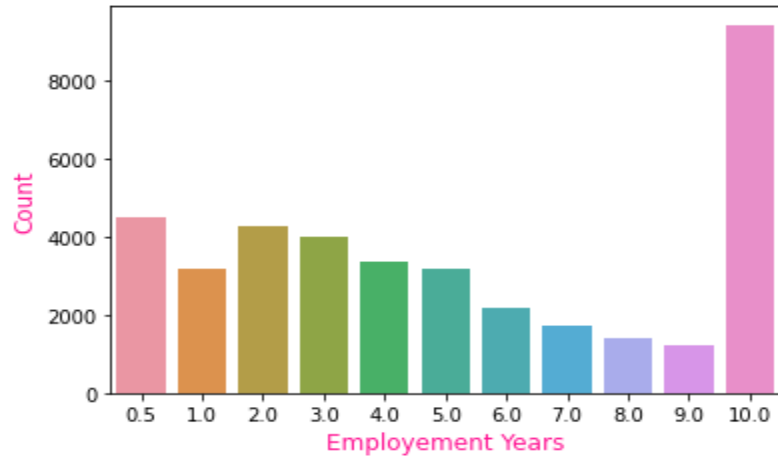


OBSERVATIONS:

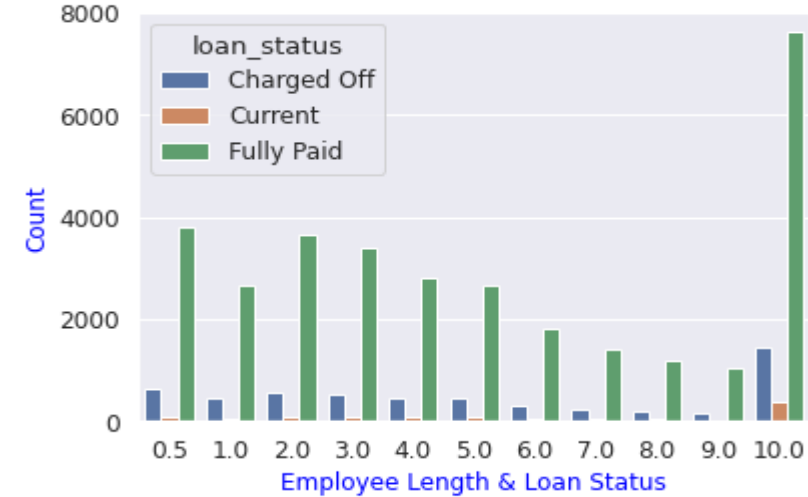
- MOST OF THE LOANS HAVE GRADE OF A AND B. THEREFORE STATING MOST OF THE LOANS ARE HIGH GRADED LOANS.
- HIGH QUALITY LOANS HAVE LOW INTEREST RATE.
- WHICH DEPICTS, HIGHER INTEREST RATE HAVE HIGHER TENDENCY TO DEFAULT THE LOAN.
- INTEREST RATE IS INCREASING BY EACH PASSING YEAR.

LOAN APPLICANTS WORK EXPERIENCE ANALYSIS

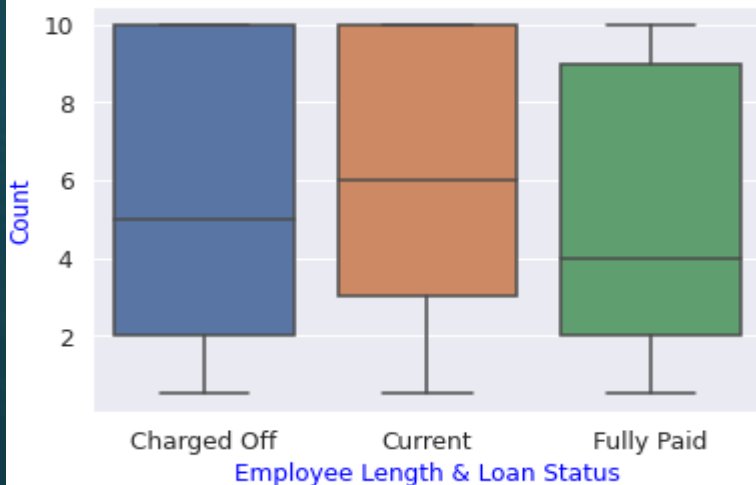
Employment Period



Employee Length And Loan Status



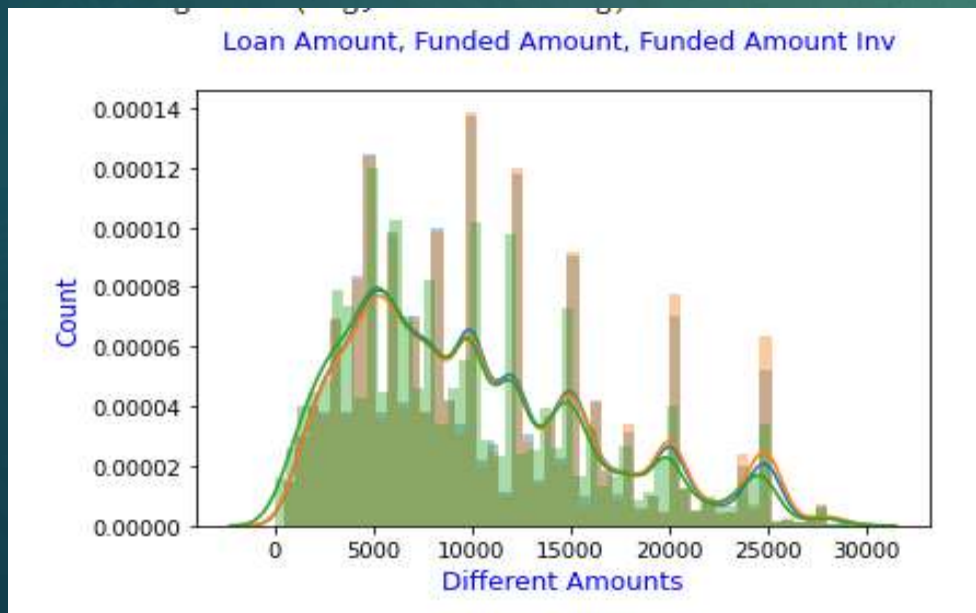
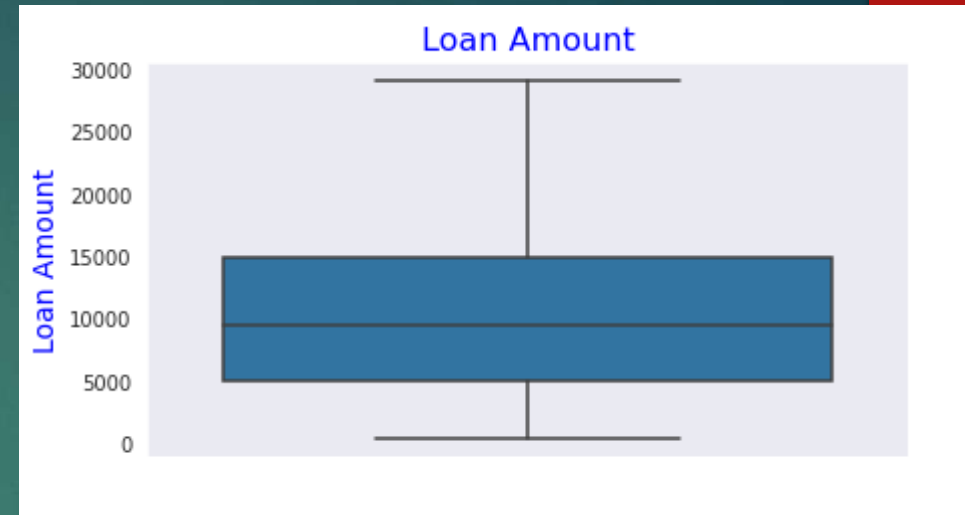
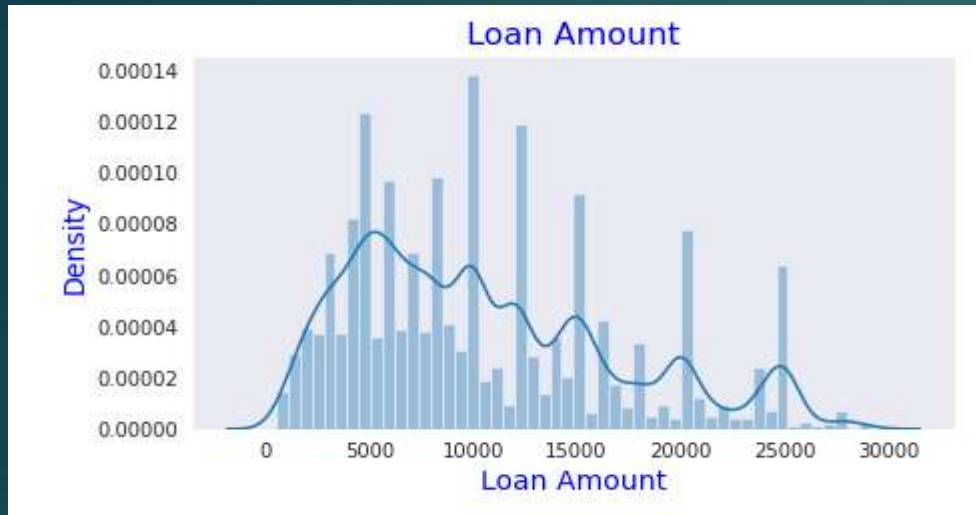
Employee Length And Loan Status



OBSERVATIONS:

- MAJORITY OF EMPLOYEES APPLYING FOR THE LOAN HAVE MORE THAN 10 YEARS OF EXPERIENCE
- TENDENCY OF PERSON TO DEFAULT THE LOAN WITH 10 YEARS OF EXPERIENCE IS ALSO HIGH. SO COMPANY NEED TO BE CAREFUL WHEN GRANTING LOAN.

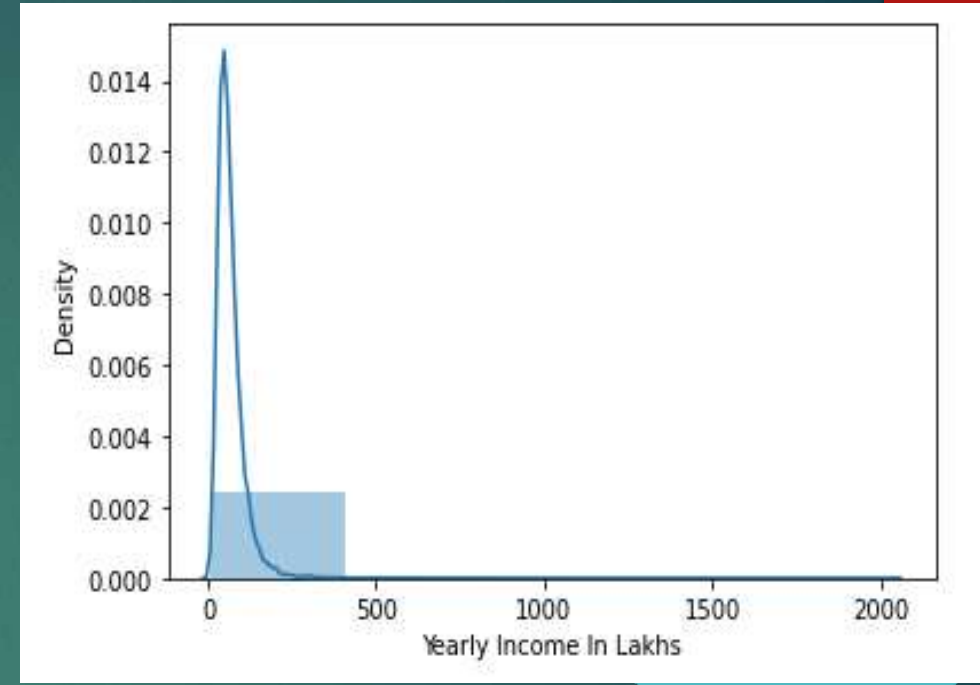
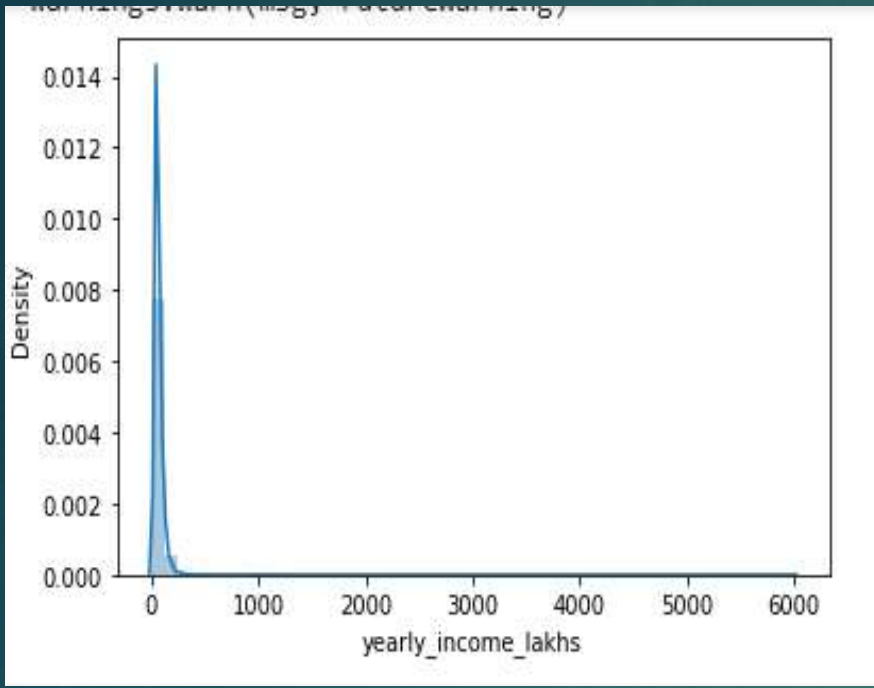
UNIVARIATE ANALYSIS ON LOAN AMOUNT



OBSERVATIONS:

- THE MOST OF THE LOAN AMOUNTS ARE IN RANGE OF 5000- 15000
- PROBABILITY OF PEOPLE WITH LOAN AMOUNT GREATER THAN 7 LAKHS TENDS TO DEFAULT IS MORE THAN THE PEOPLE WITH LESS THAN 7 LAKH
- THESE THREE VARIABLES [LOAN_AMOUNT, FUNDED_AMNT, FUNDED_AMNT_INV] HAVE NEARLY SAME VALUES SO WE CAN TAKE ANY OF THOSE VARIABLES FOR ANALYSIS.

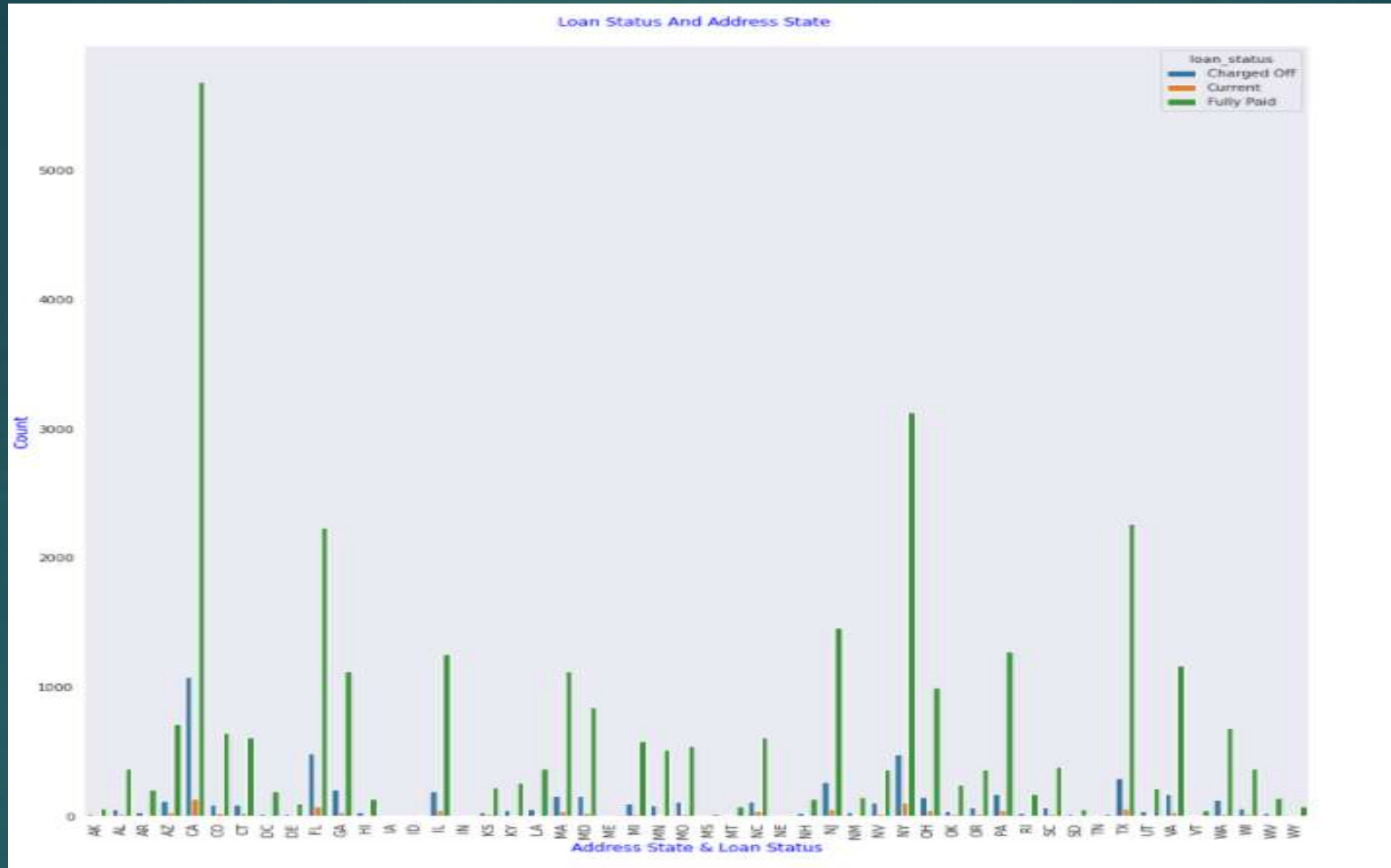
INCOME ANALYSIS FOR THE LOAN APPLICANTS



OBSERVATIONS:

- MOST OF THE APPLICANTS HAVE INCOME BETWEEN 0 TO 5 LAKHS.
- THERE ARE TWO APPLICANTS HAVING INCOME ABOVE 30 LAKHS

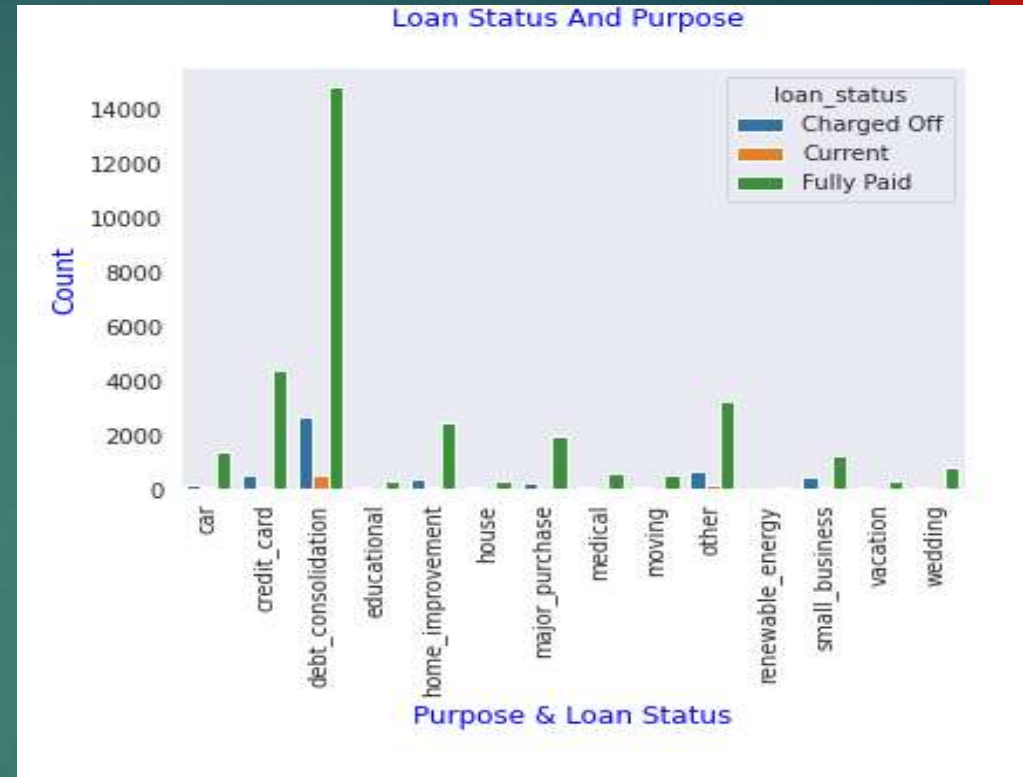
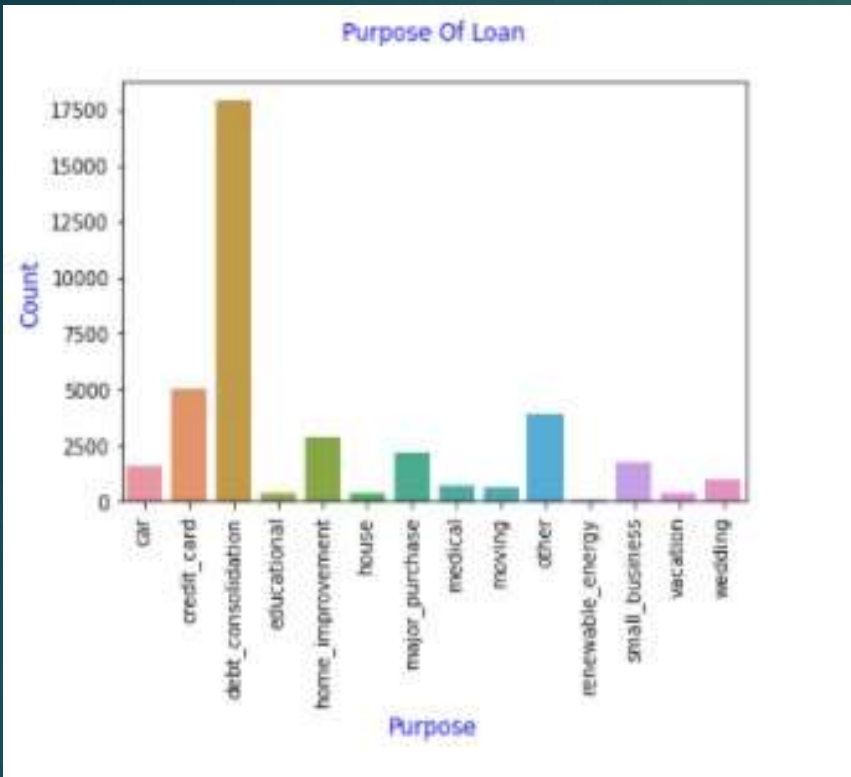
LOAN DEFAULTERS BY STATE ANALYSIS



OBSERVATION:

- APPLICANTS FROM STATE CA HAVE HIGH TENDENCY TO DEFAULT THE LOAN

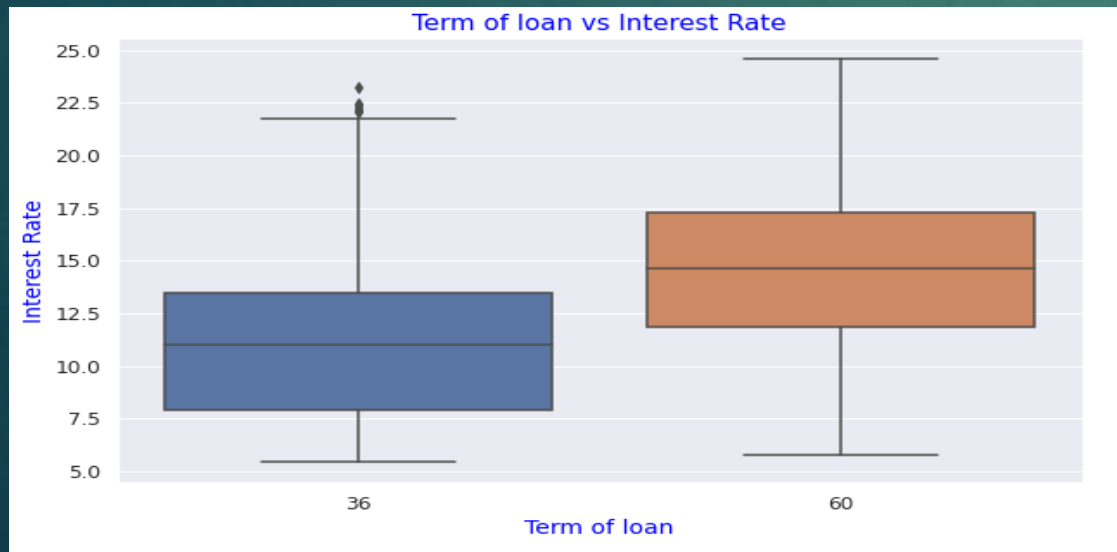
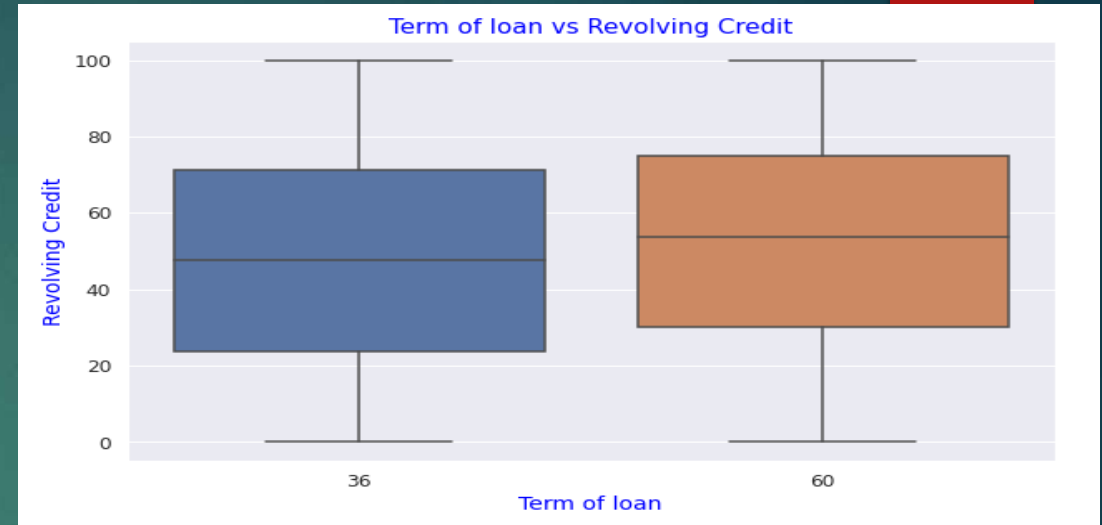
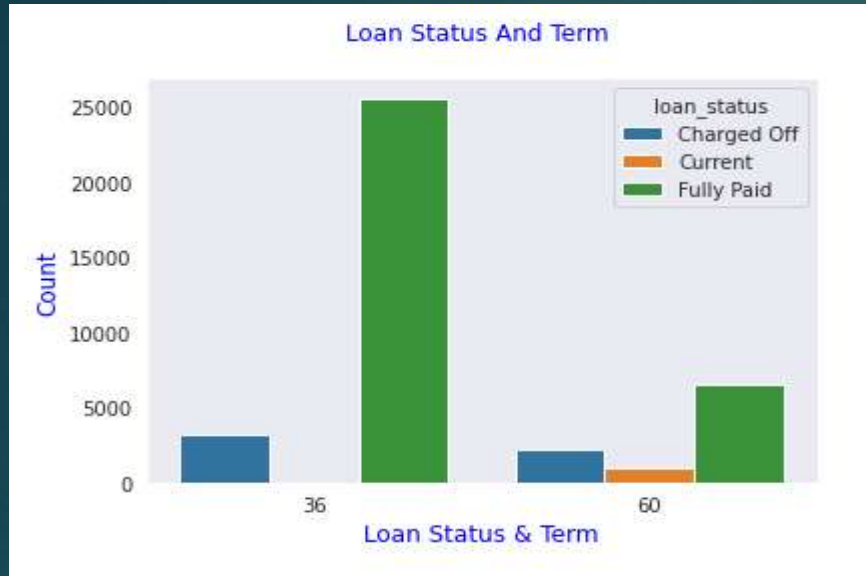
PURPOSE FOR APPLYING LOAN ANALYSIS



OBSERVATIONS:

- MAJORITY OF LOAN IS PROVIDED FOR THE DEBT CONSOLIDATION PURPOSE AND HAS BEEN FULLY PAID.
- DEBT_CONSOLIDATION IS THE MAXIMUM PROVIDED CATEGORY BY THE BORROWER FOR THE LOAN REQUEST.

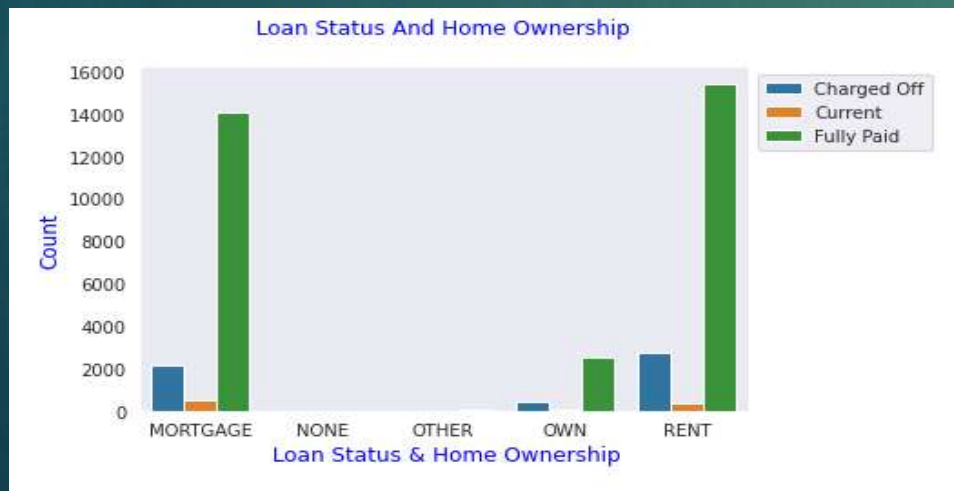
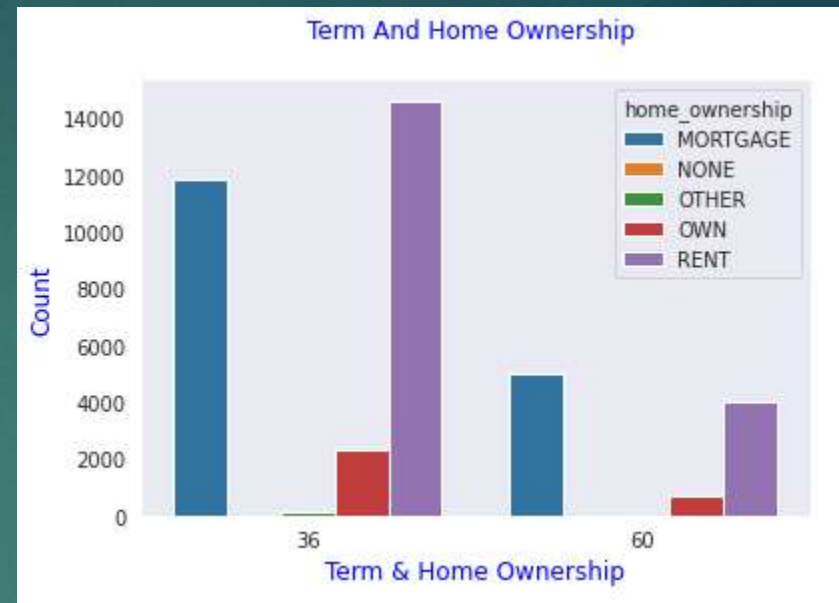
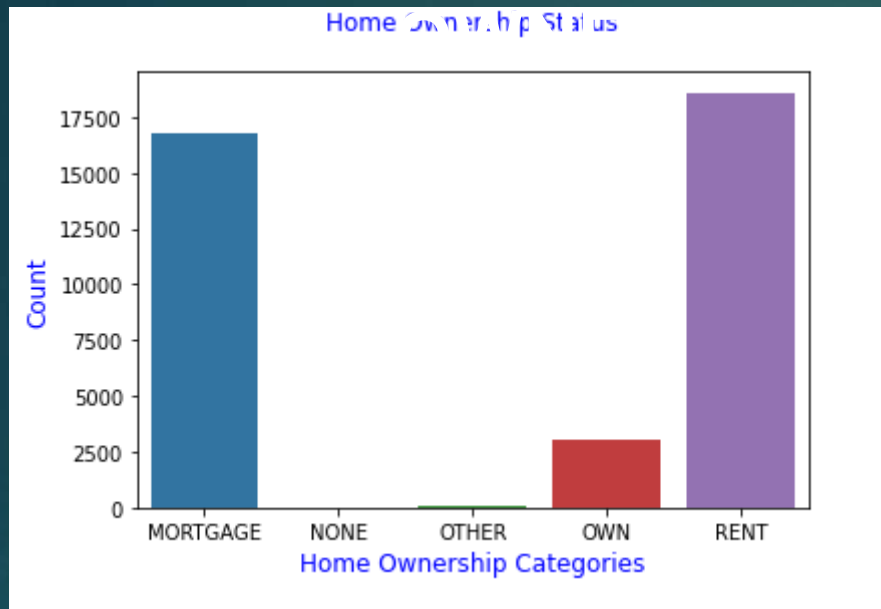
LOAN STATUS VS TERM



OBSERVATIONS:

- LOANS GRANTED FOR 36 MONTHS HAVE SLIGHTLY HIGH TENDENCY TO DEFAULT.
- 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.
- REVOLVING CREDIT IS HIGHER FOR 60 MONTHS LOAN TERM.

LOAN STATUS BY HOME

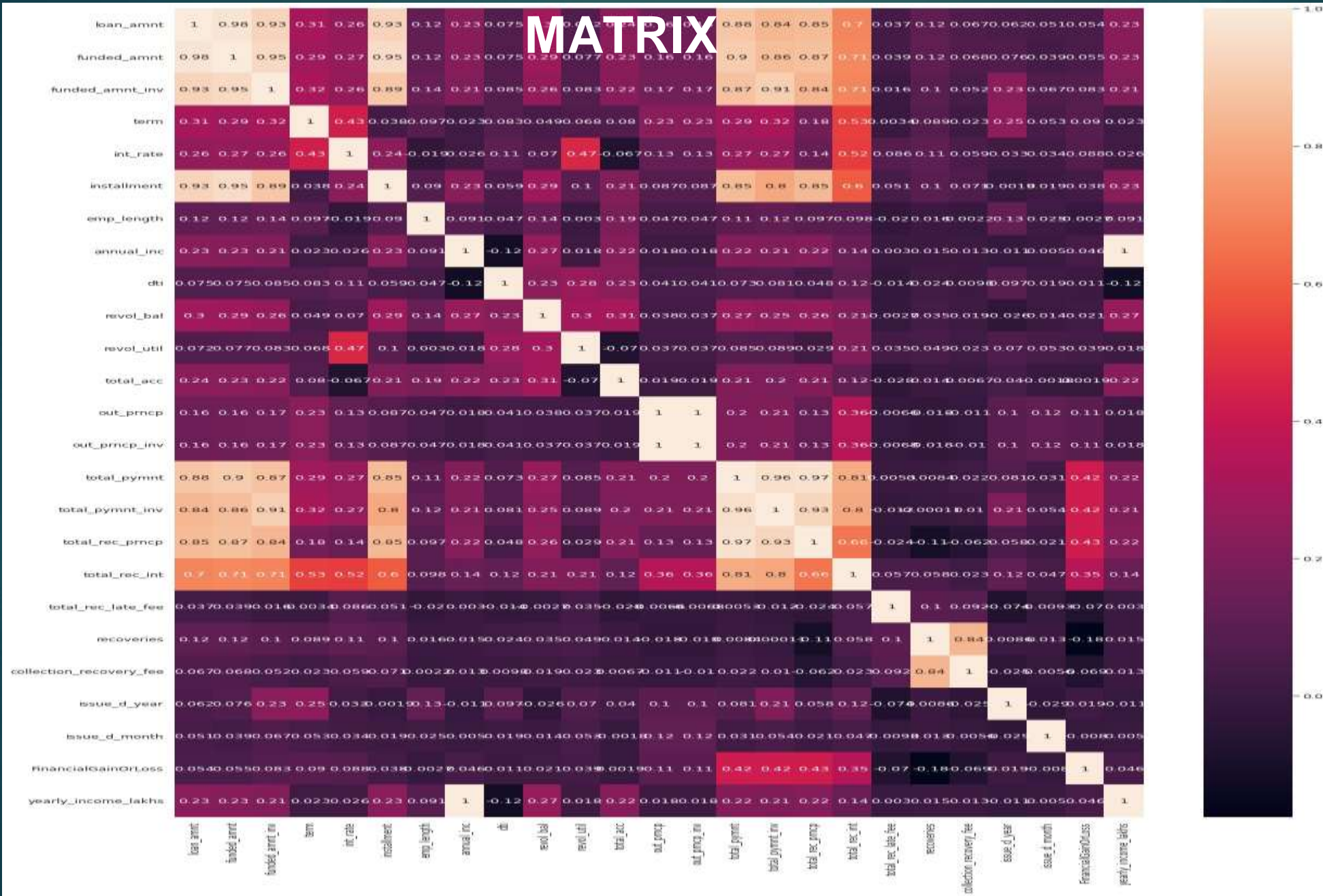


OBSERVATIONS:

- THE ONES WHO HAVE TAKEN LOAN ARE ON RENT AS MENTIONED THEIR STATUS DURING THE TIME OF REGISTRATION.
- LOAN STATUS(FULLY PAID) IS HIGHEST FOR THE OWNERSHIP(MORTGAGE, OWN AND RENT)

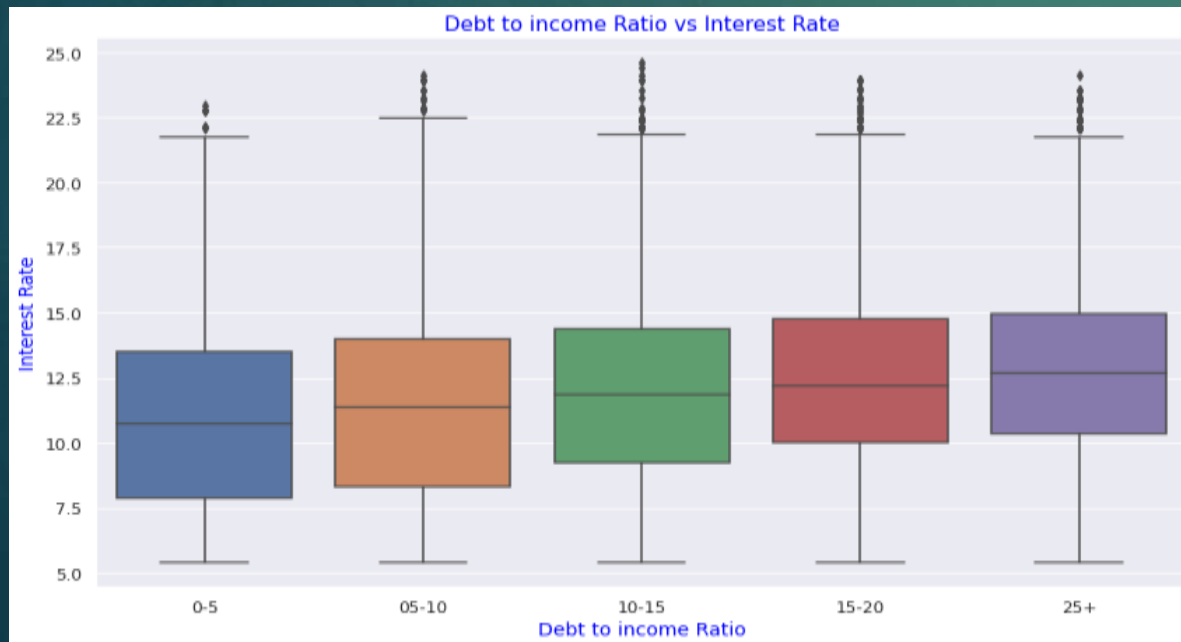
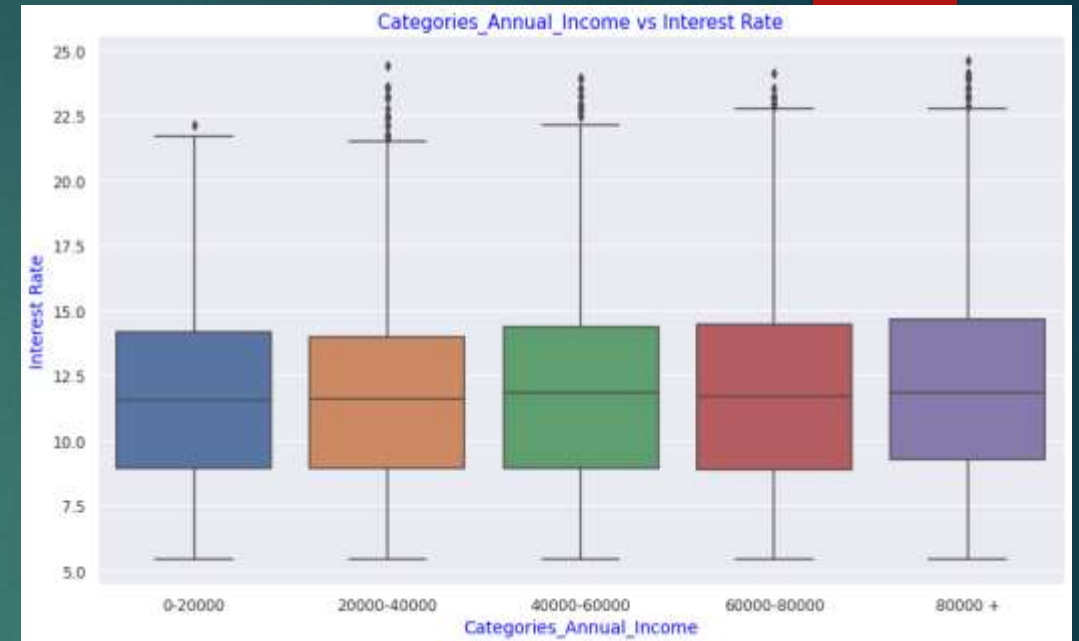
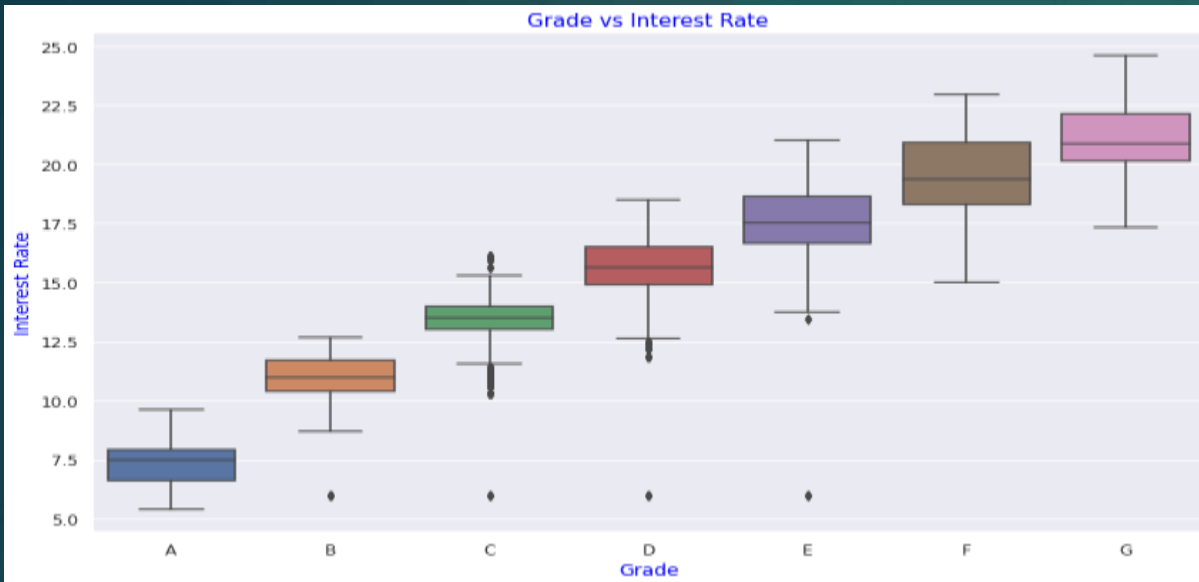
CORRELATION

MATRIX



OBSERVATION:

Loan amount, amount and investor funding are strongly correlated.



OBSERVATIONS:

- THE HIGHER THE BORROWER'S CREDIT GRADE, THE LOWER THE INTEREST RATE OFFERED TO THAT BORROWER ON A LOAN.
- INTEREST RATE IS INCREASING WITH LOAN AMOUNT INCREASE.
- SLIGHT INCREASE IN INTEREST RATE WITH INCREASE IN DTI.

CONCLUSIONS

- LOW GRADE LOANS HAVE HIGH TENDENCY TO DEFAULT. THEREFORE GRADING SYSTEM IS WORKS AS EXPECTED.
- TENDENCY TO DEFAULT IS HIGHEST WHEN THE PURPOSE OF LOAN IS DEBT CONSOLIDATION.
- IF LOAN TENURE IS 36 MONTHS, TENDENCY TO DEFAULT IS HIGHER. 60 MONTHS TENURE IS SLIGHTLY SAFER.
- AVERAGE REVOLVING CREDIT IS HIGHER FOR 60 MONTHS LOAN TERM. MOST OF THE LOANS ISSUED FOR LONGER TERM HAD HIGHER REVOLVING CREDIT FOR REPAYMENT.
- THE ONES GETTING 'CHARGED OFF' HAVE LOWER ANNUAL INCOME THAN THE ONES WHO 'PAID FULLY' FOR EVERY GRADE.
- IF DTI IS LOW ENOUGH YOU MAY GET A LOWER INTEREST RATE. THERE IS SLIGHT INCREASE IN INTEREST RATE WITH INCREASE IN DTI.
- LOANS HAVING HIGHER INTEREST RATE HAVE MORE DEFAULTERS. A THOROUGH BACKGROUND CHECK SHOULD BE DONE IN SUCH CASES.
- LOAN AMOUNT, INVESTOR AMOUNT AND FUNDING AMOUNT ARE STRONGLY CORRELATED. ANNUAL INCOME WITH DTI IS NEGATIVELY CORRELATED.