

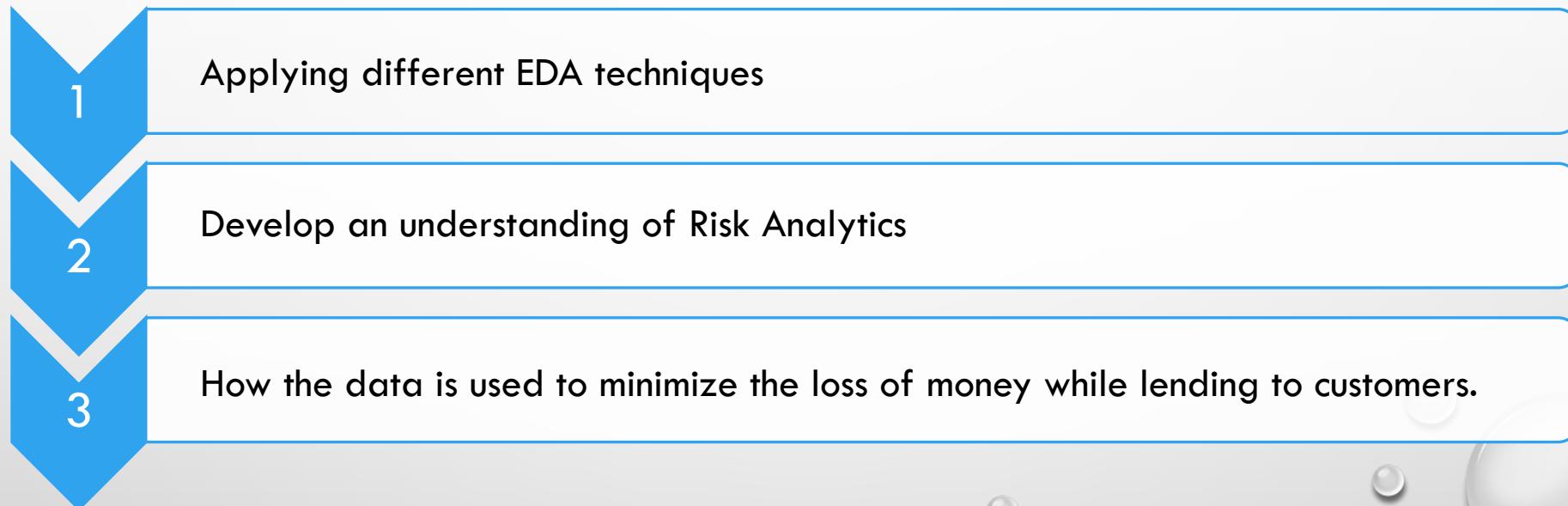
LENDING CLUB CASE STUDY EDA

BY:-

- 1) RAJAT OMPRAKASH MANDANIYAN
- 2) MANISH KONDURI

OBJECTIVE

The aim of this case study is to get an idea of how we are going to work on Lending Club Use Case using EDA.

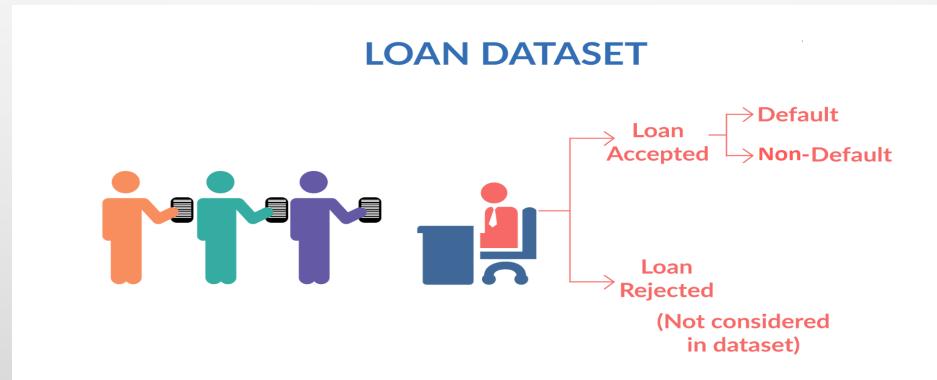
- 
- 1 Applying different EDA techniques
 - 2 Develop an understanding of Risk Analytics
 - 3 How the data is used to minimize the loss of money while lending to customers.

BUSINESS UNDERSTANDING

We are working for Lending club a finance company which specializes in lending various types of loans to urban customers.

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:

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



GOAL OF DATA ANALYSIS:-

Lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). 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.

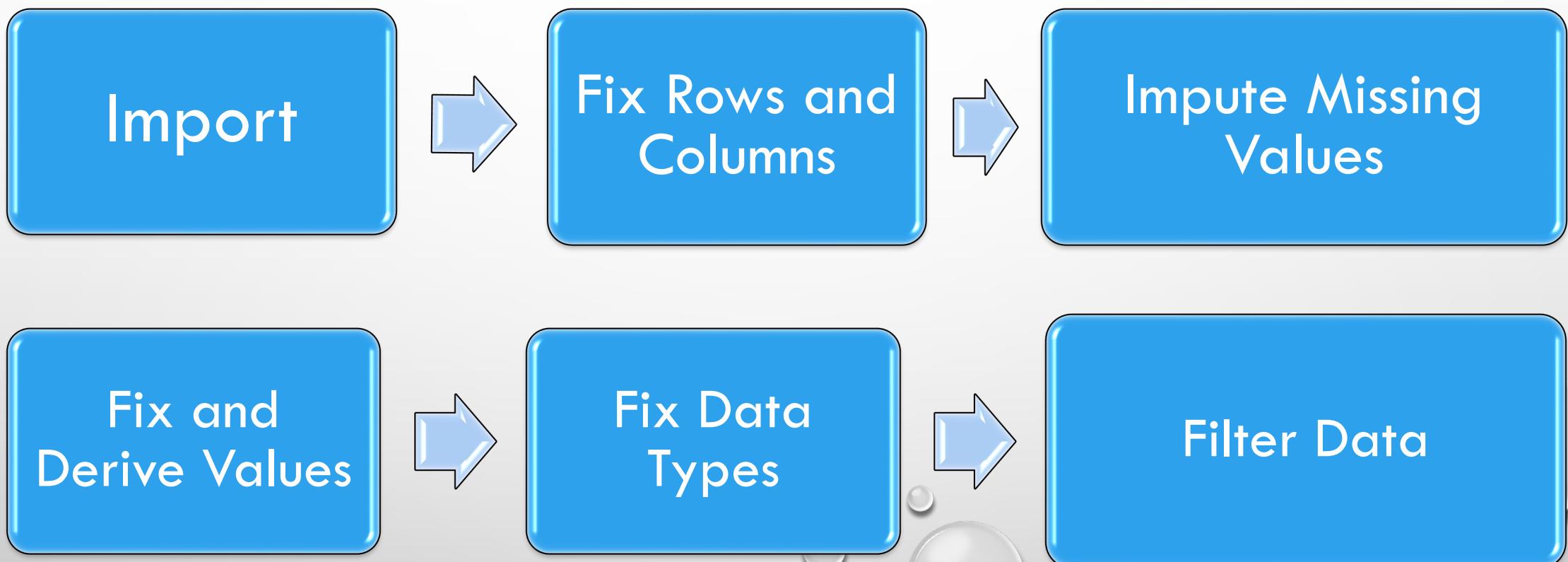
The main objective is to be able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

Perform an analysis 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. .

Major Steps performed in lending club case Study :

- A. Data Cleanup
- B. Derived Metrics
- C. Identifying outliers
- D. Data Visualization (Univariate Analysis and Bivariate Analysis)
- E. Correlation
- F. Observation
- G. Insights

DATA CLEANUP



Data Cleaning

The initial shape of the Loan Data Set provided has “39717” Rows and “111” Columns

Steps Performed to Clean Up Data :-

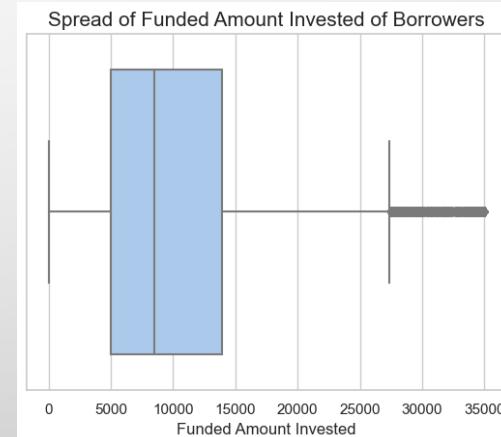
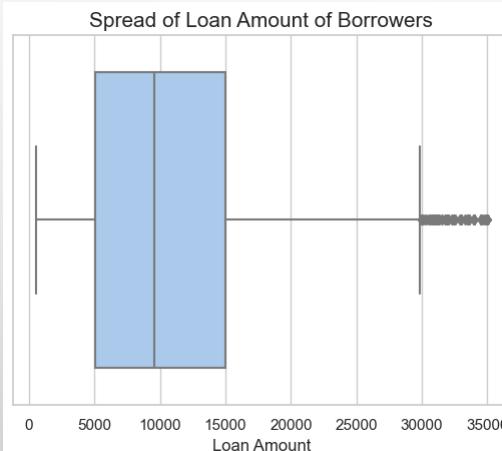
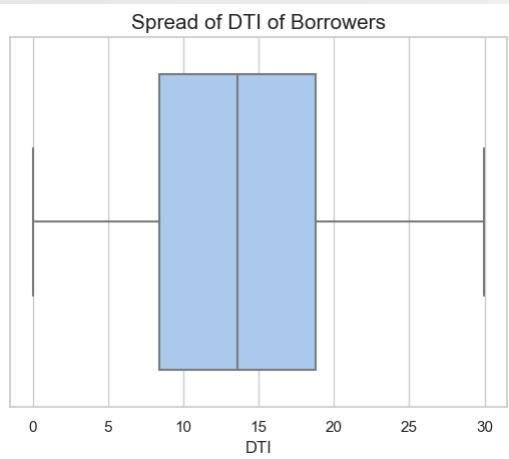
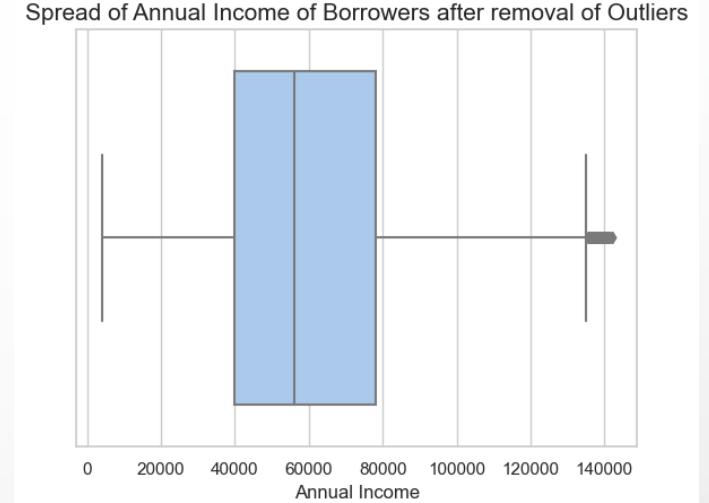
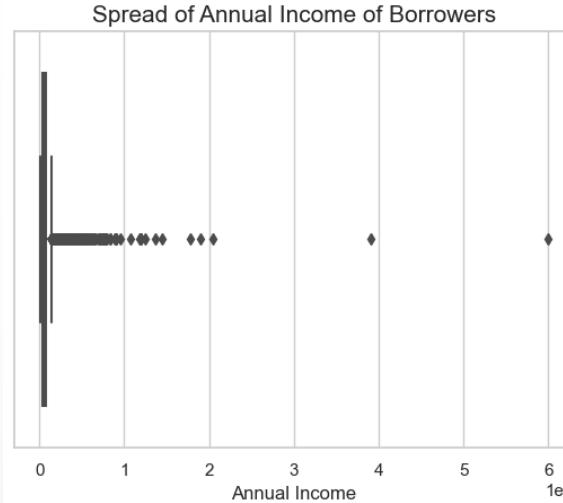
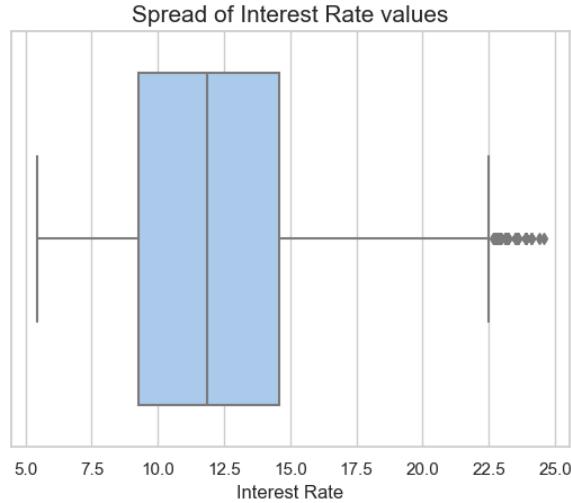
- 1) Removed all the columns which have all values as Null.
- 2) Majority of the Columns have Null 100% Null values, So these columns can't give any insights, So we can drop these columns
- 3) Deleted 57 Columns because they have entire Null values.
- 4) Dropped the columns which doesn't provide any significance.
- 5) Dropped the columns which are Customer Behavioural specific.

Data Filtering

The data provided has so much significance but some of it is stored in incorrect format. Filtered the data to usable standards

Plotting the Columns to find Outliers

Outliers will effect the analysis and will give biased results



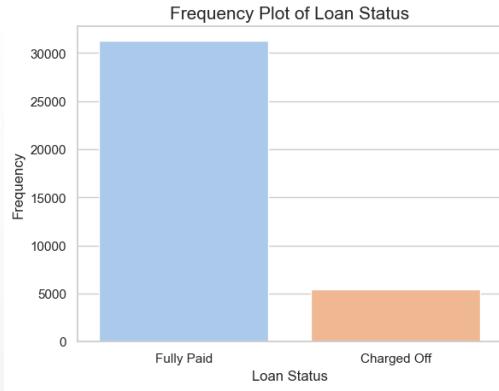
Data Visualisation

Univariate Analysis

- Unordered Categorical Variables :-
 - Loan Status
 - Home Ownership
 - Purpose
 - Verification Status
- Ordered Categorical Variables :-
 - Grade
 - Term
 - Employee Length
 - Derived Columns - Month and Year
- Quantitative Variables :-
 - Loan Amount
 - Funded Amount Invested
 - Interest Rate
 - Annual Income
 - DTI

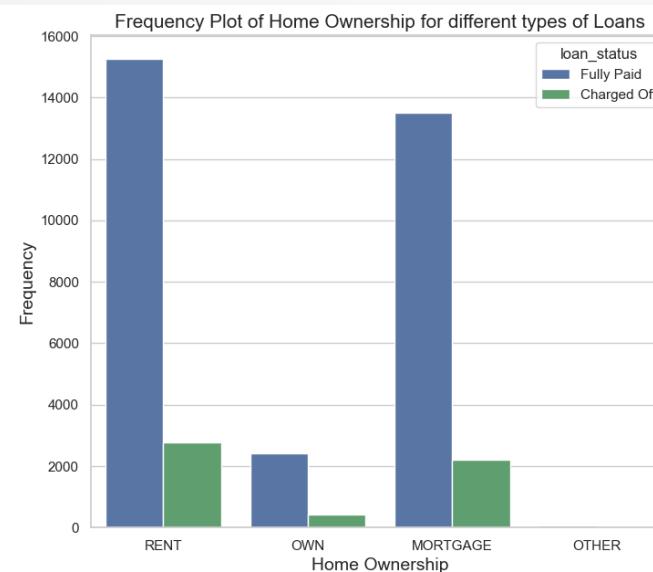
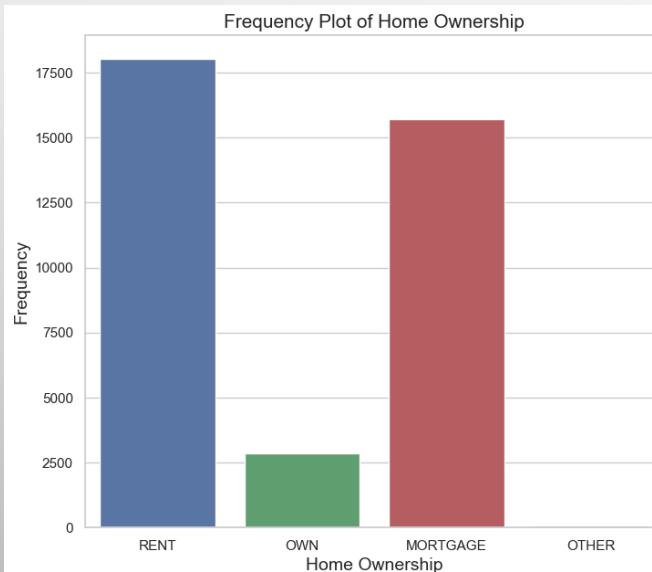
1. Performing Univariate Analysis on Loan Status

- As loan_status is an unordered categorical variable we can use CountPlot to plot this categorical variable

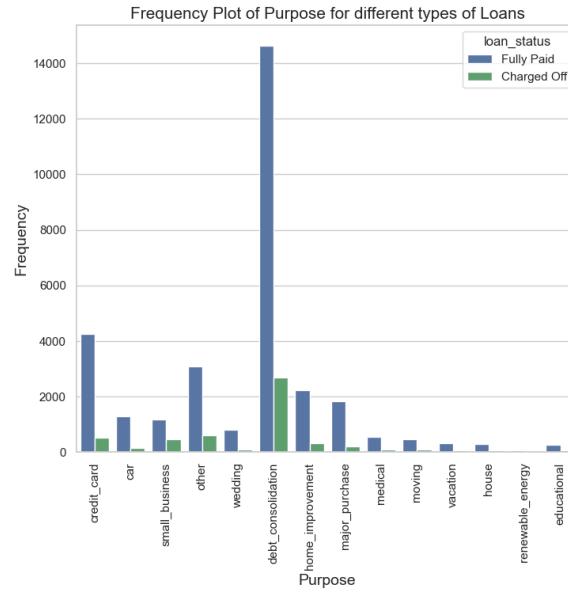
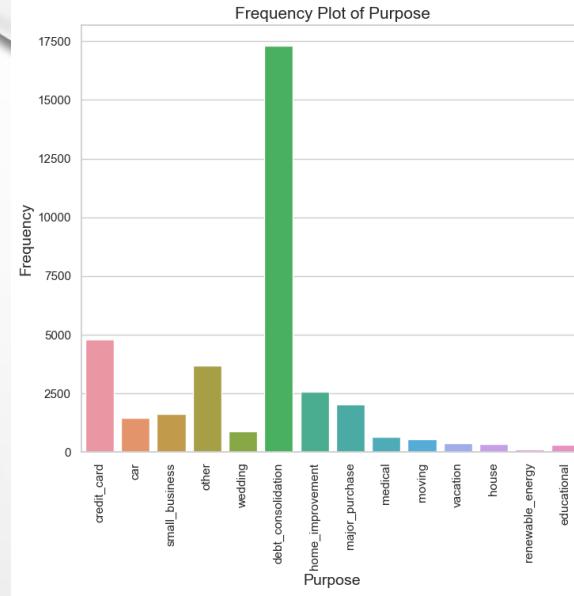


2. Performing Univariate Analysis on Home Ownership

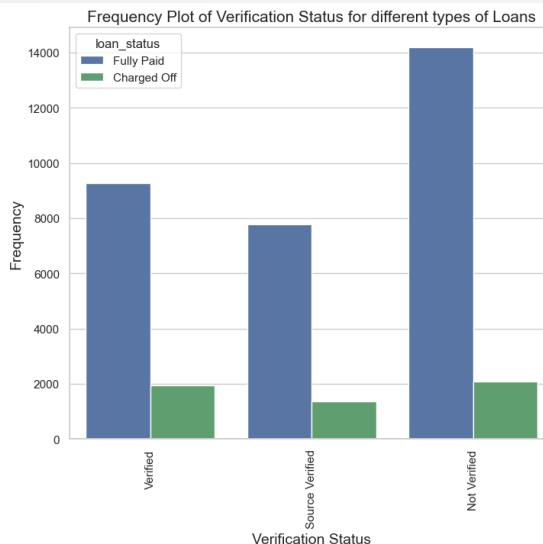
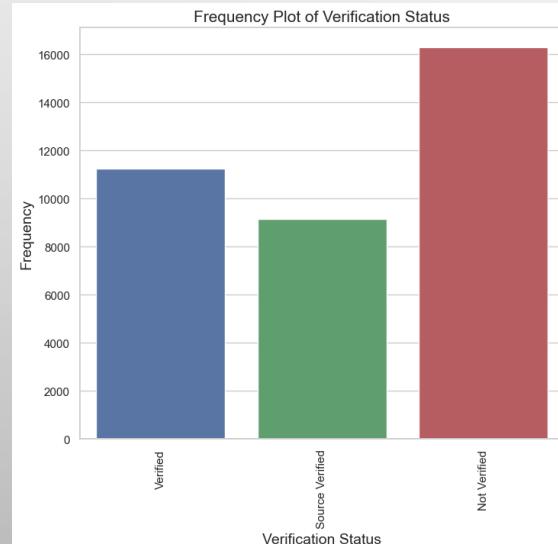
- As home_ownership is an unordered categorical variable we can use CountPlot to plot this categorical variable
- This plot has the combination of Charged Off, Fully Paid, Current loan statuses



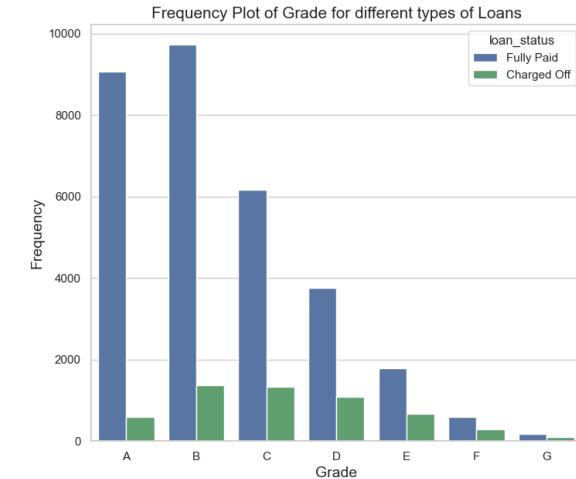
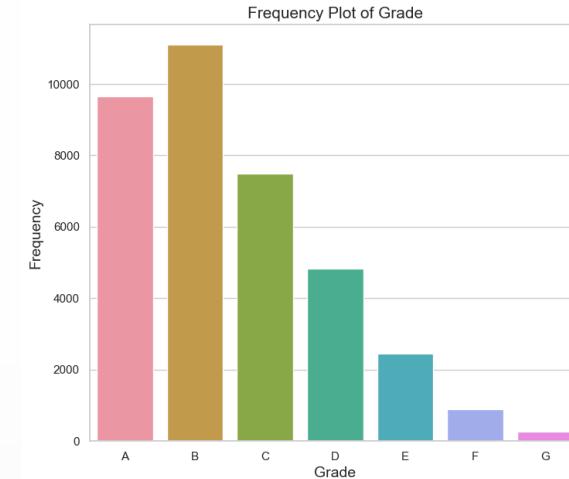
3. Performing Univariate Analysis on purpose



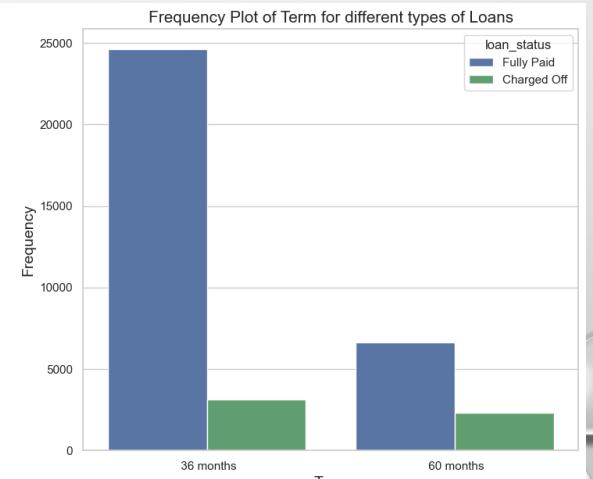
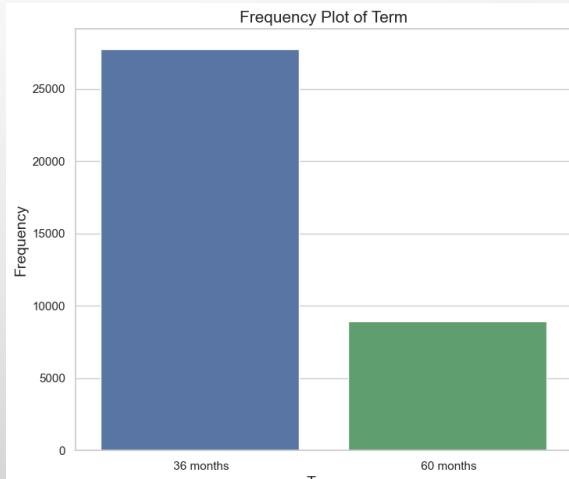
4. Performing Univariate Analysis on verification_status



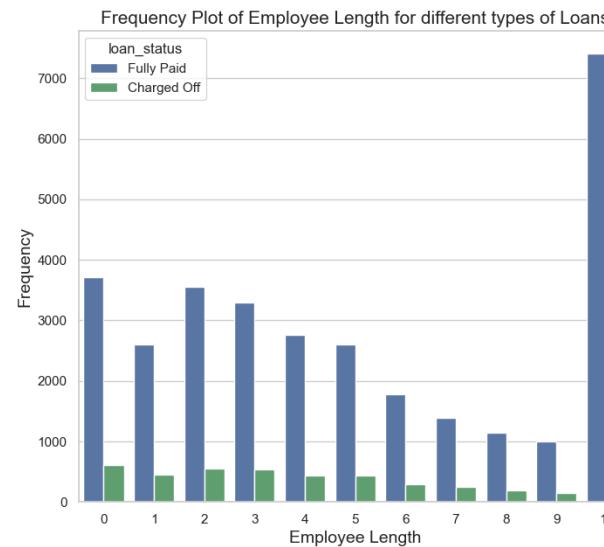
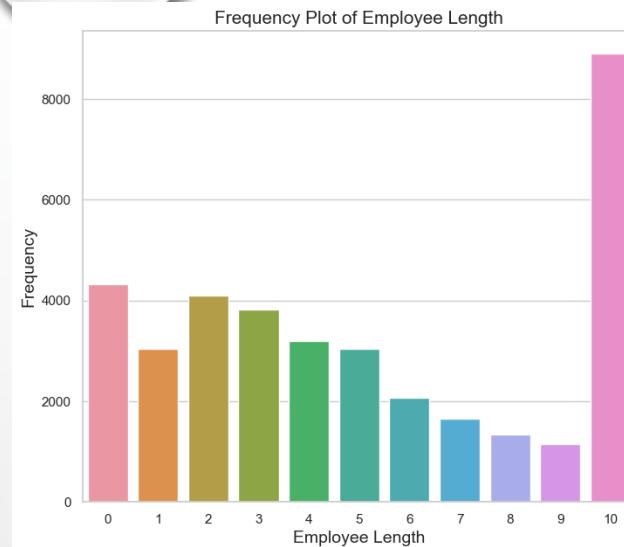
5. Performing Univariate Analysis on grade



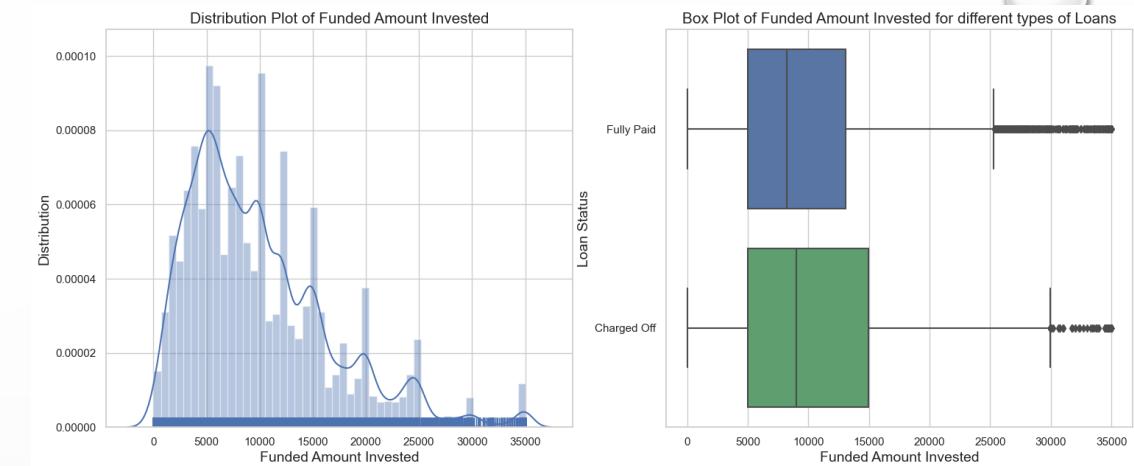
6. Performing Univariate Analysis on term



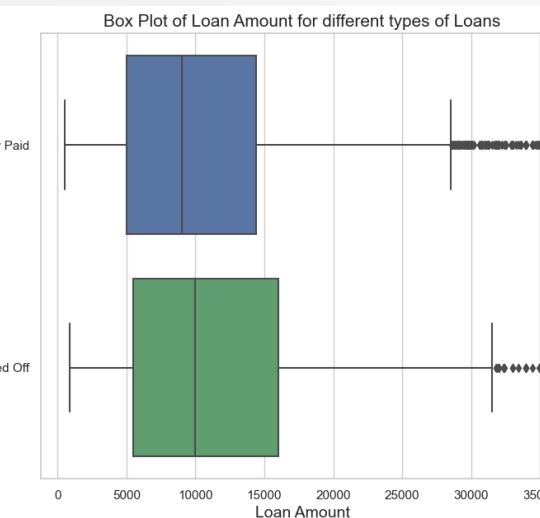
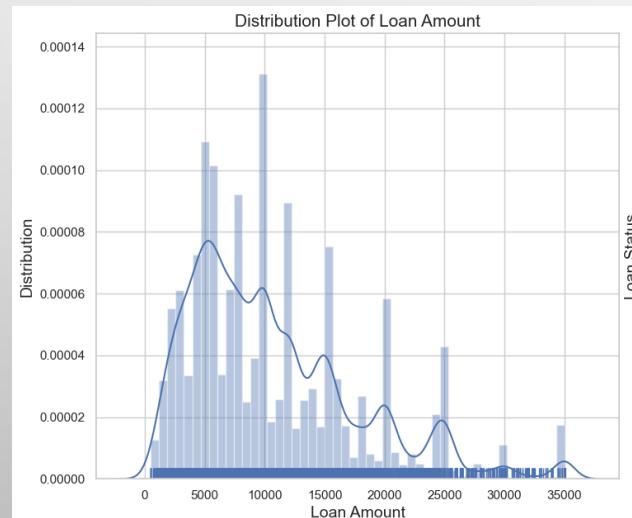
7. Performing Univariate Analysis on emp_length



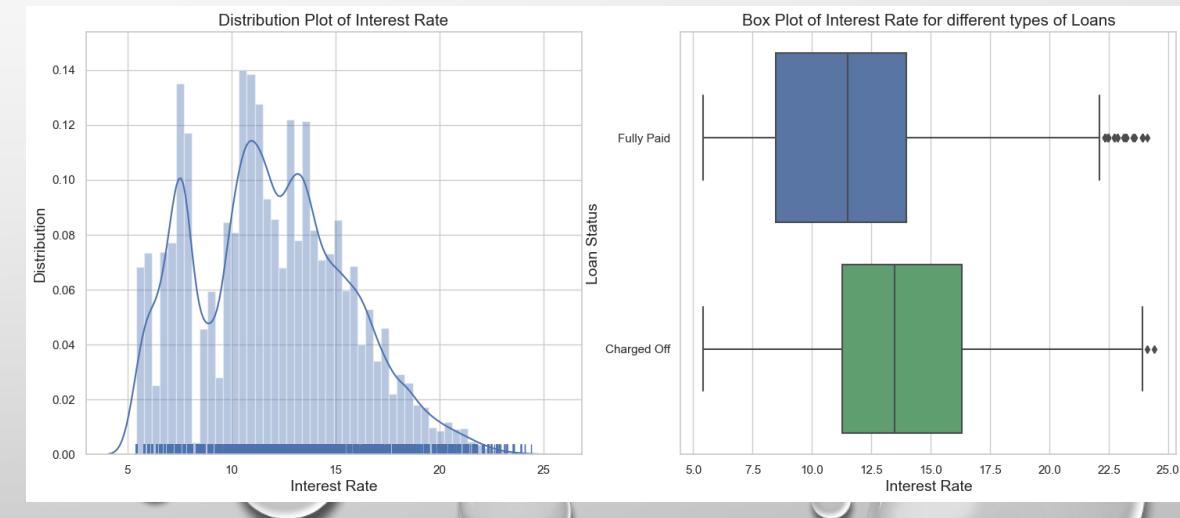
9. Univariate Analysis on Funded Amount Inv



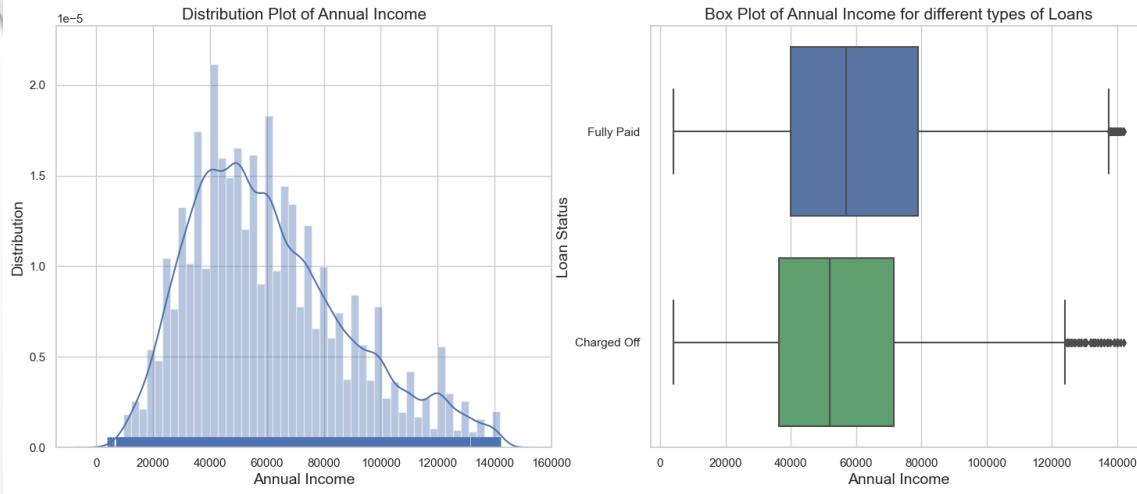
8. Univariate Analysis for Funded Amount Invested



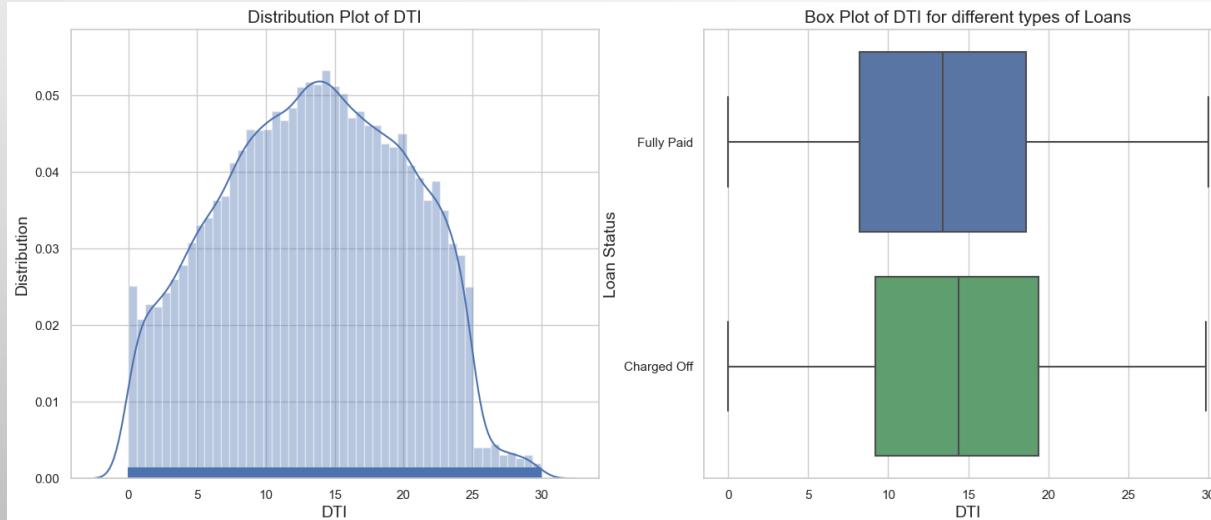
10. Univariate Analysis on Interest Rate



11. Univariate Analysis on Annual Income



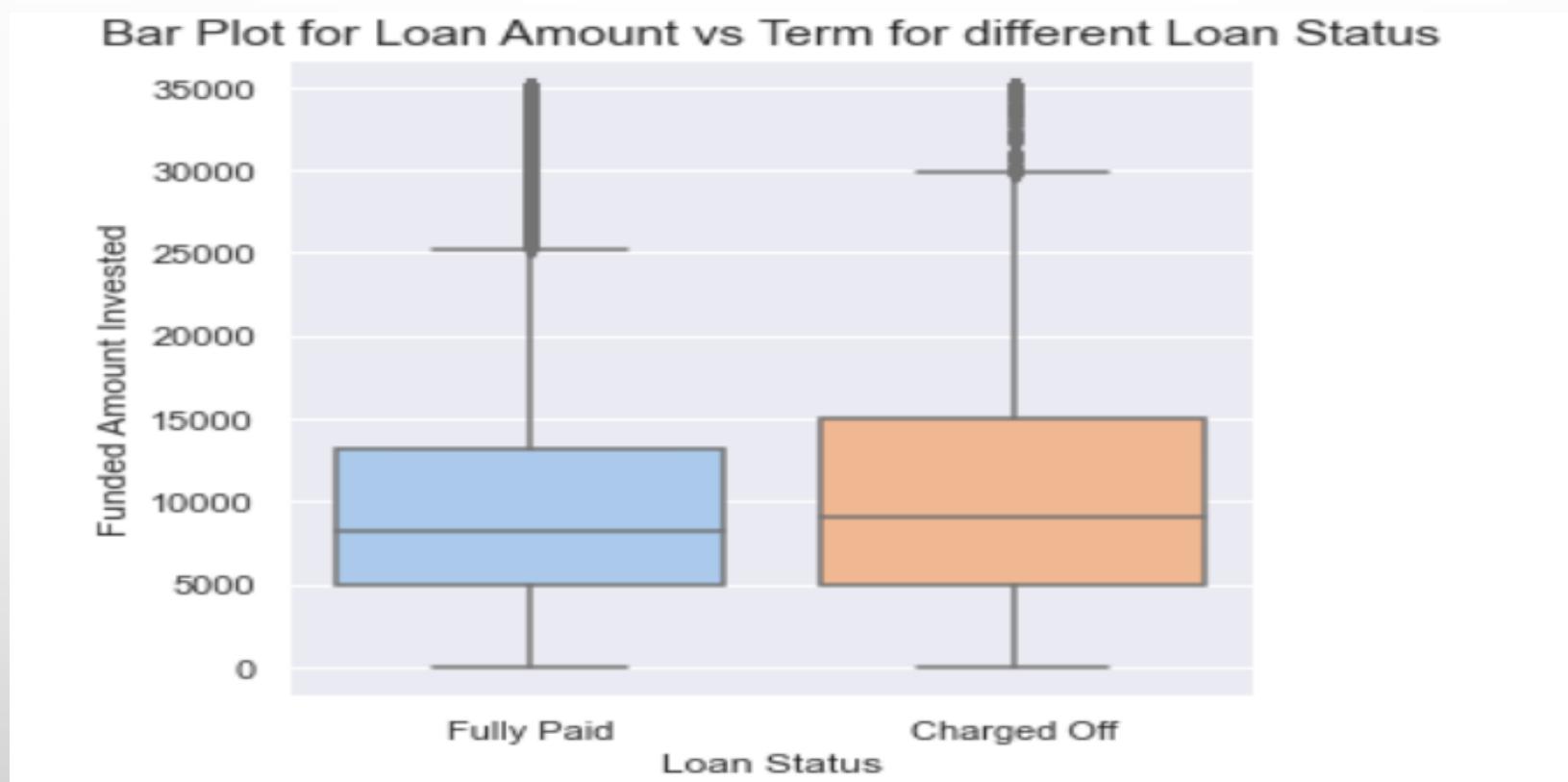
12. Univariate Analysis on DTI



Bivariate Analysis

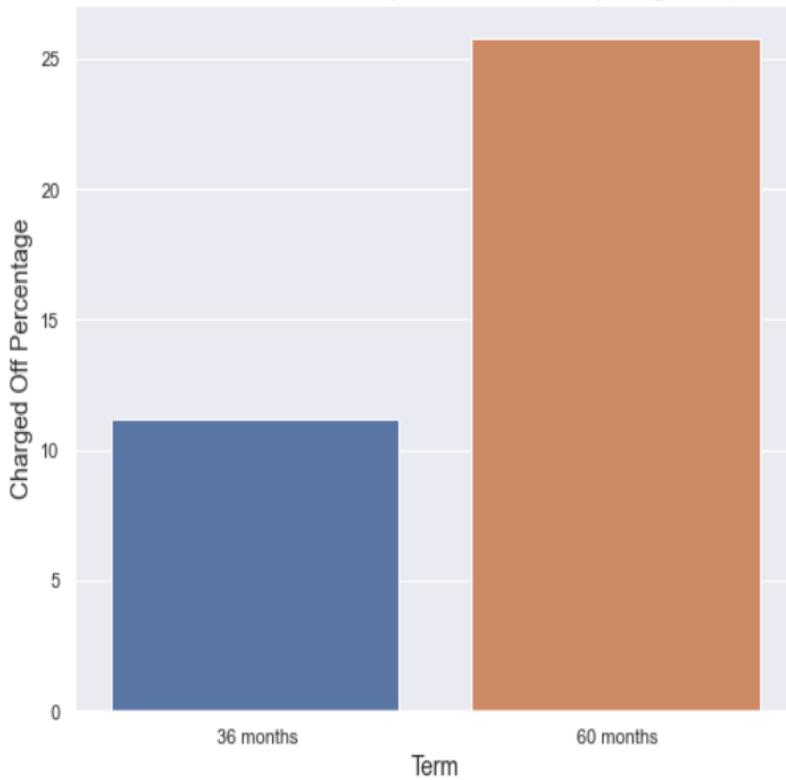
- 1. Funded Amount Invested vs Loan Status
- 2. Funded Amount Invested vs Term
- 3. Funded Amount Invested vs Interest Rate
- 4. Funded Amount Invested vs Annual Income
- 5. Funded Amount Invested vs Purpose
- 6. Funded Amount Invested vs Employee Length
- 7. Funded Amount Invested vs Home Ownership
- 8. Funded Amount Invested vs Month and Year issued
- 9. Funded Amount Invested vs Grade
- 10. Funded Amount Invested vs Verification Status
- 11. Funded Amount Invested vs DTI
- 12. Grade vs Interest Rate
- 13. Annual Income vs Home Ownership
- 14. Annual Income vs Interest Rate
- 15. Home Ownership vs Purpose
- 16. Annual Income vs Purpose

FUNDED AMOUNT INVESTED VS LOAN STATUS

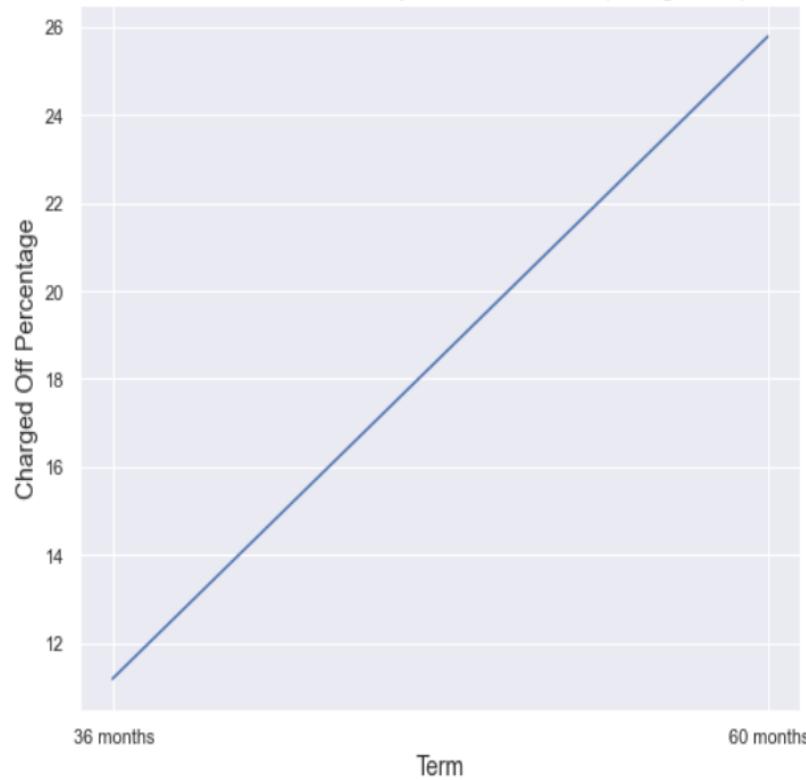


FUNDED AMOUNT INVESTED VS TERM

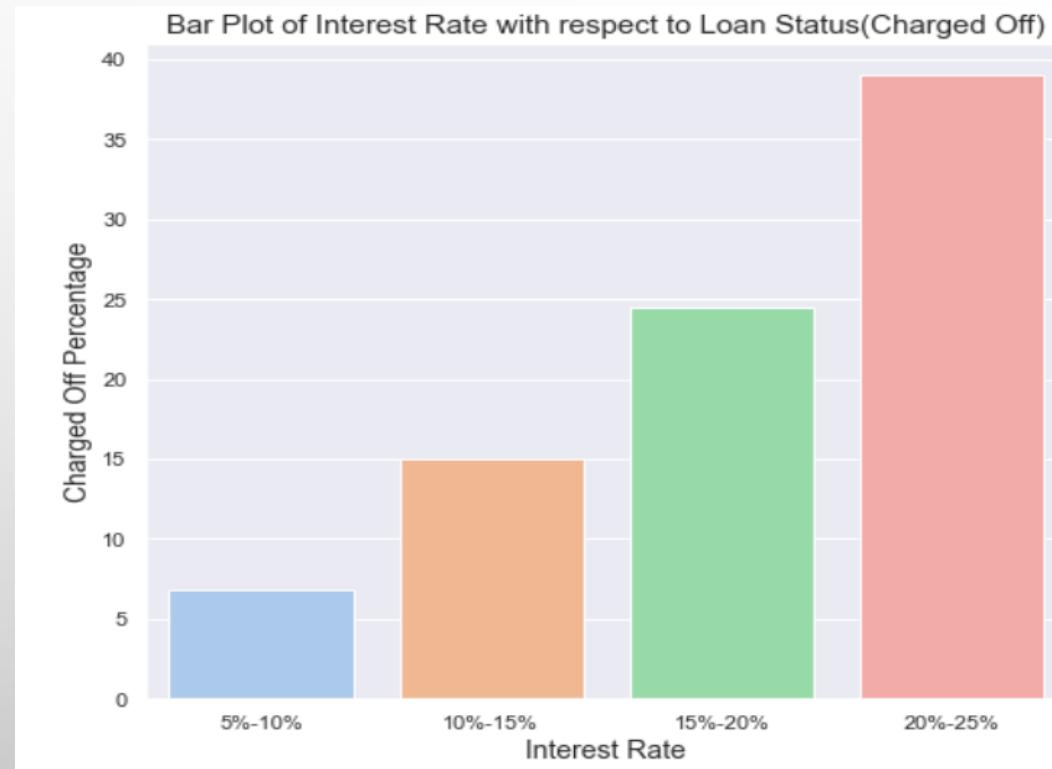
Bar Plot of Term with respect to Loan Status(Charged Off)



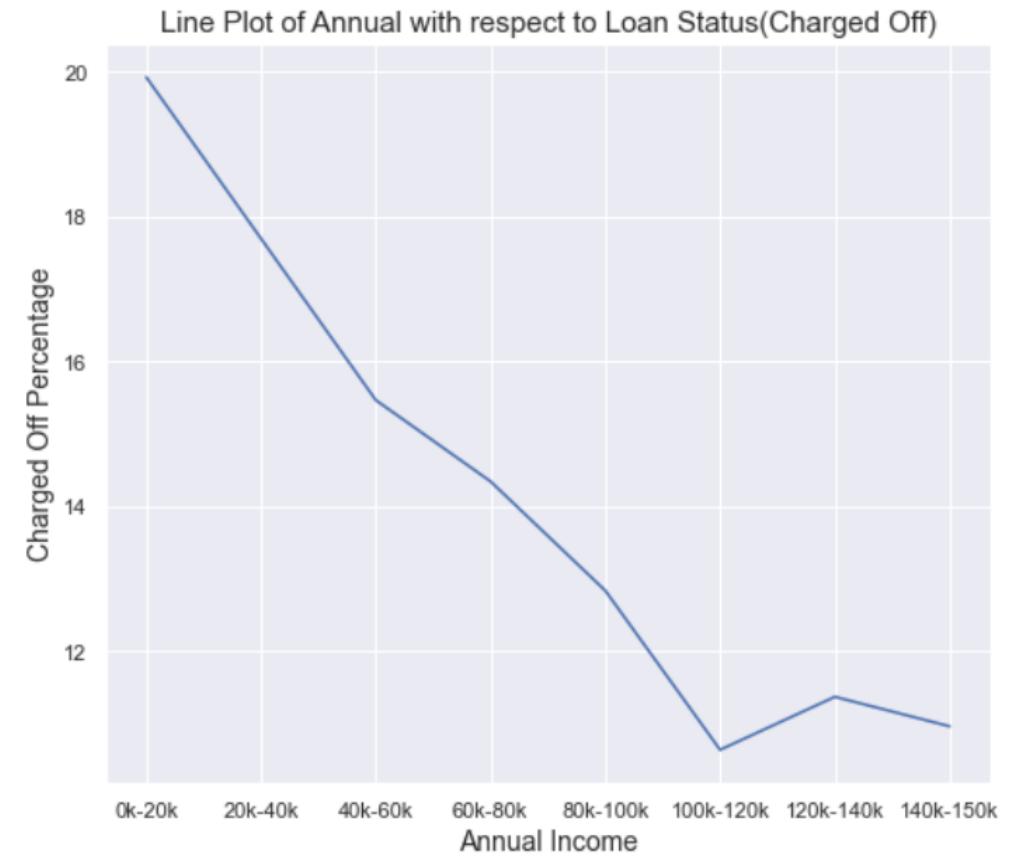
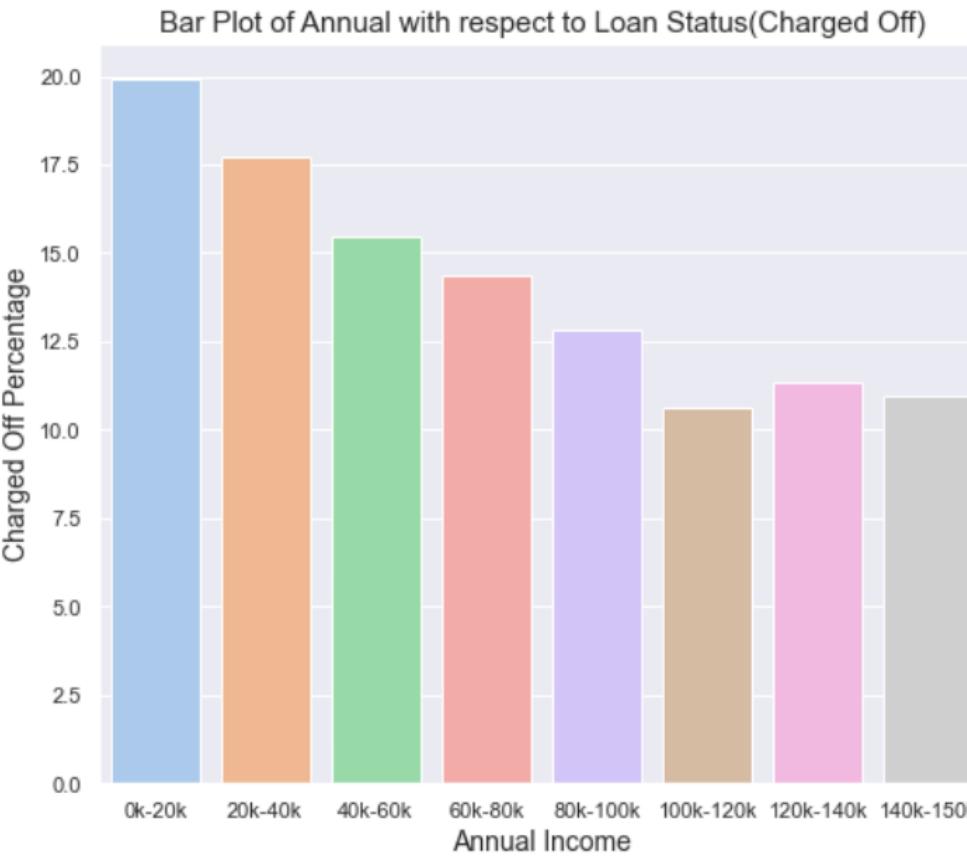
Line Plot of Term with respect to Loan Status(Charged Off)



FUNDED AMOUNT INVESTED VS INTEREST RATE



FUNDED AMOUNT INVESTED VS ANNUAL INCOME



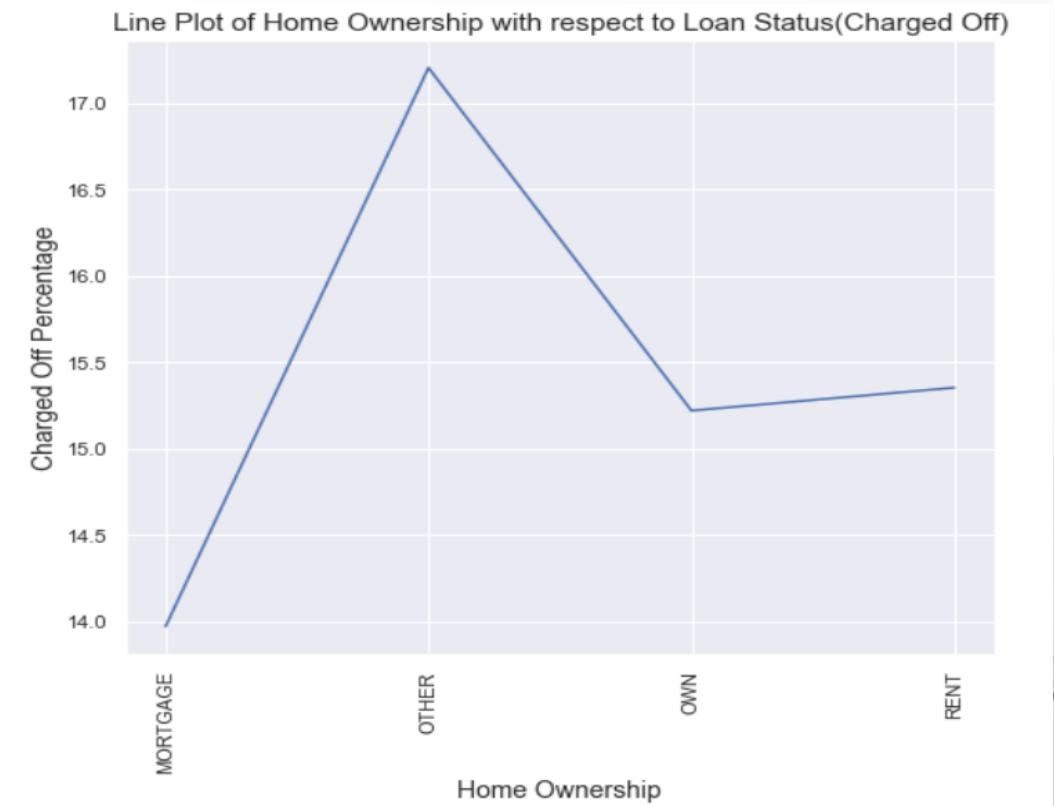
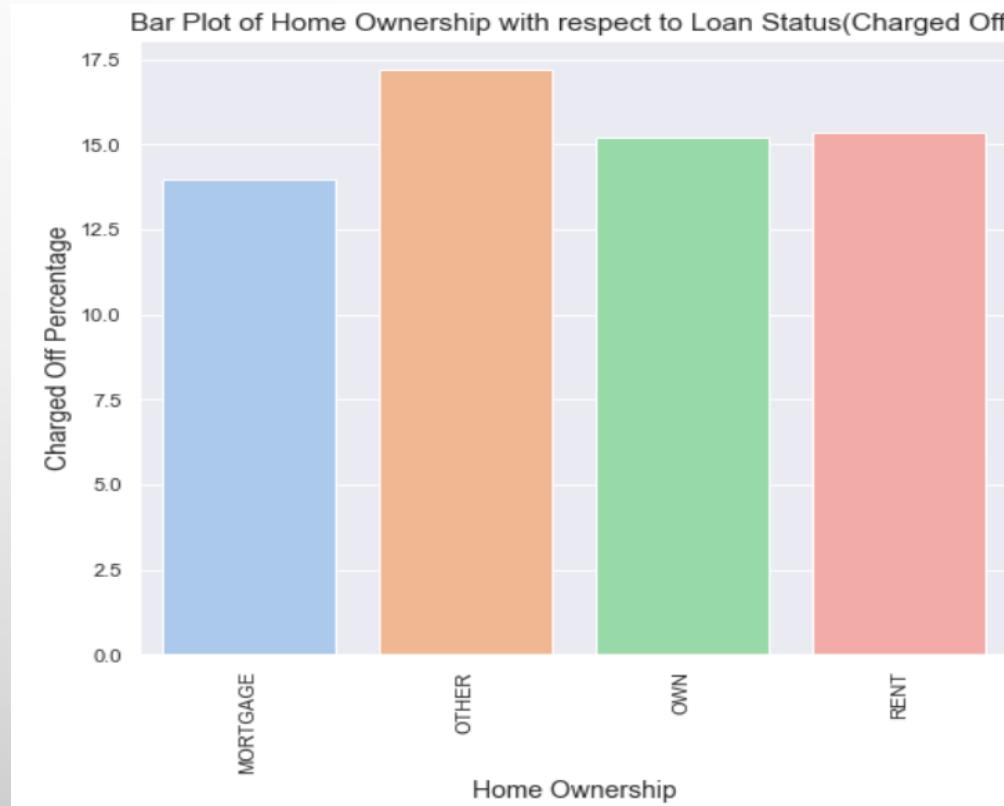
FUNDED AMOUNT INVESTED VS PURPOSE



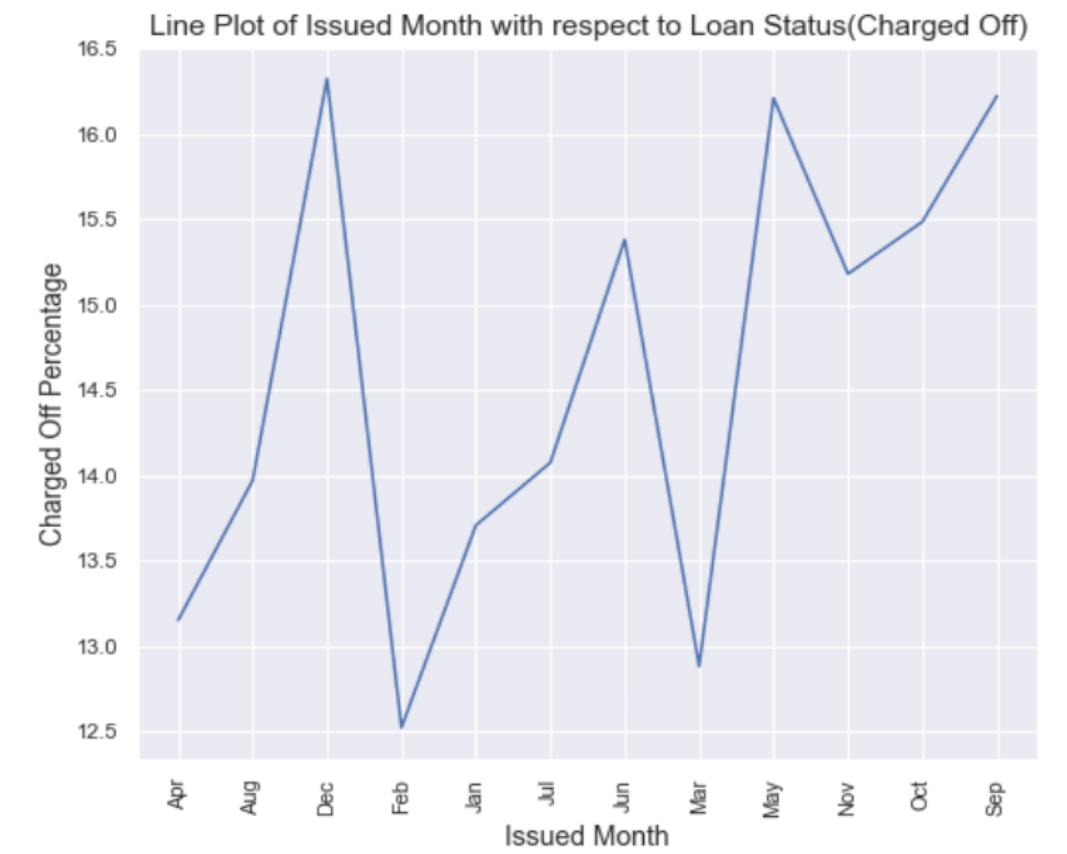
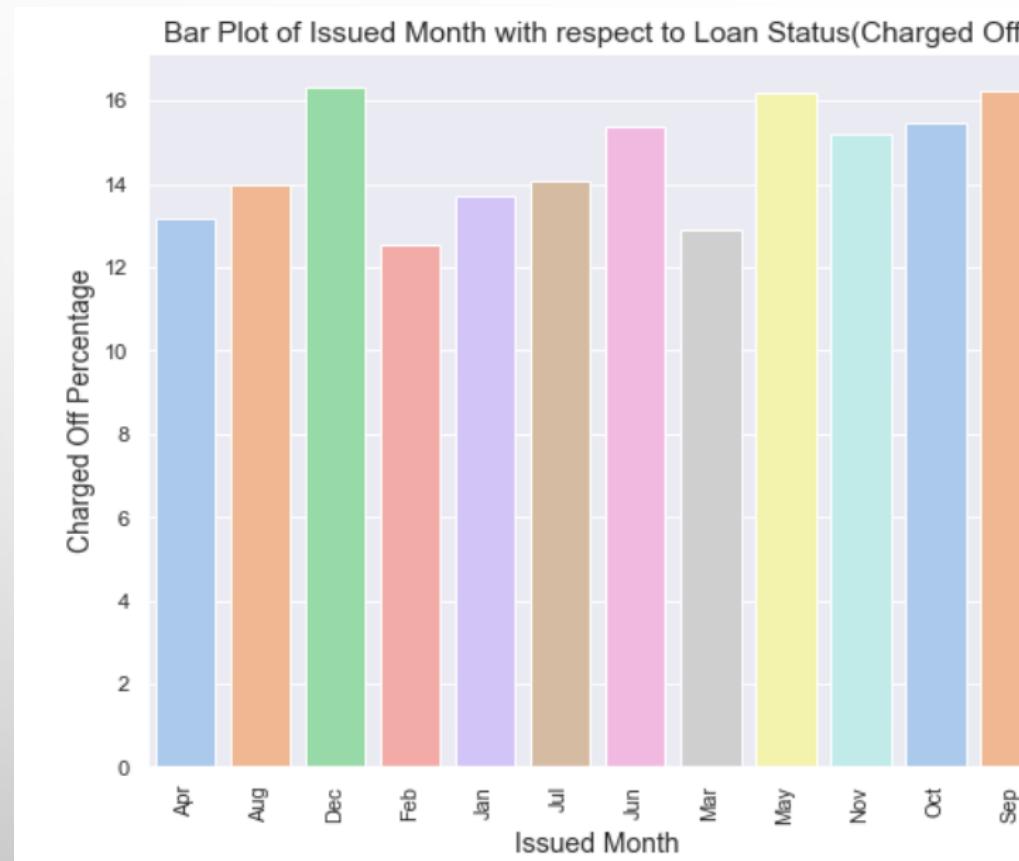
FUNDED AMOUNT INVESTED VS EMPLOYEE EXPERIENCE



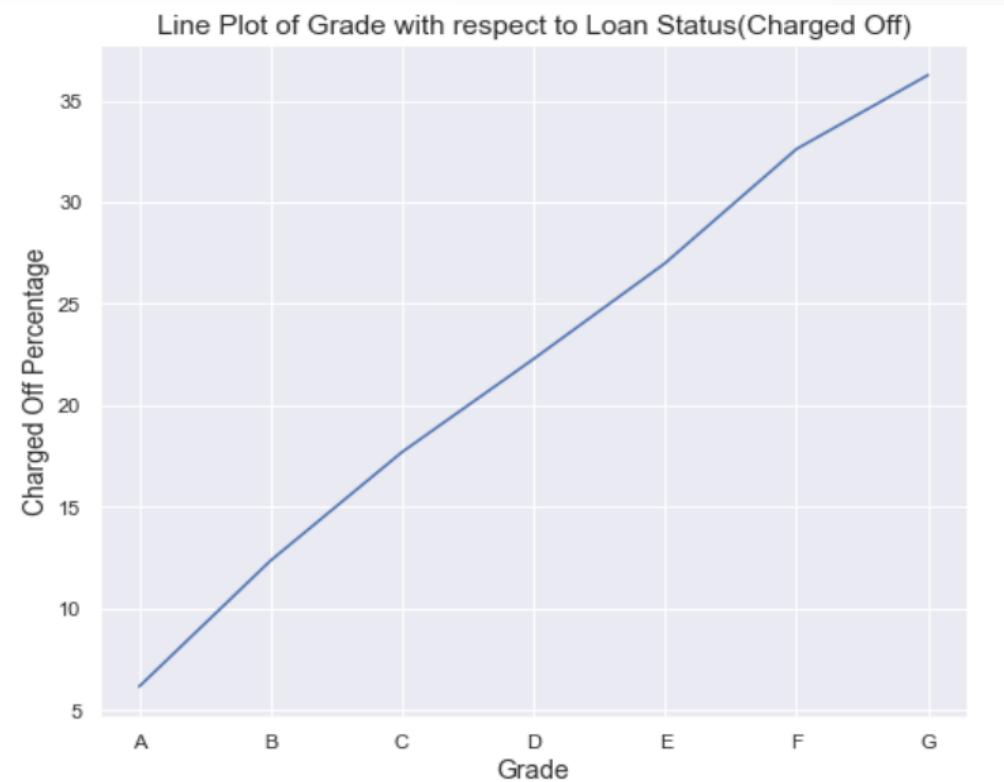
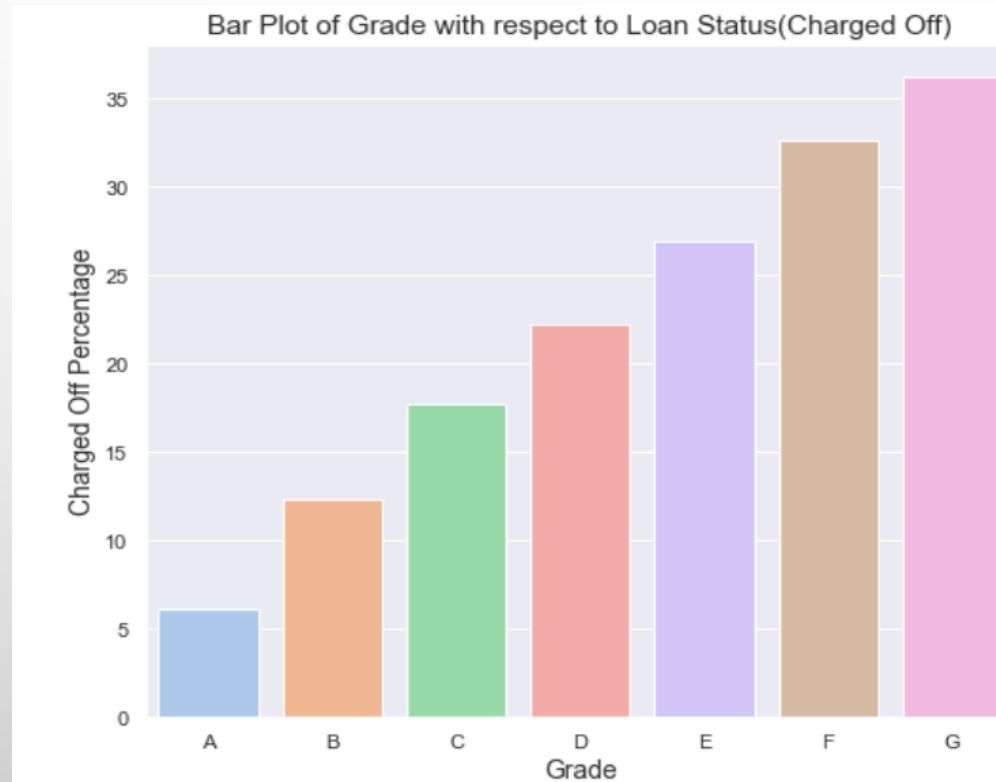
FUNDED AMOUNT INVESTED VS HOME OWNERSHIP



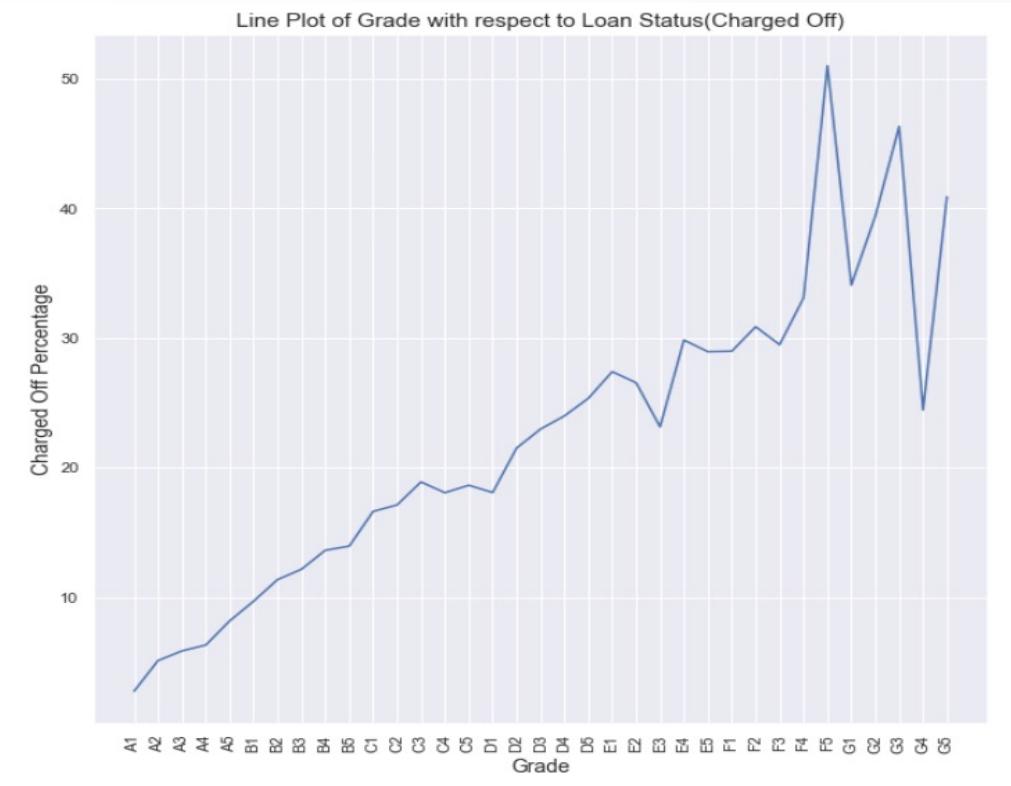
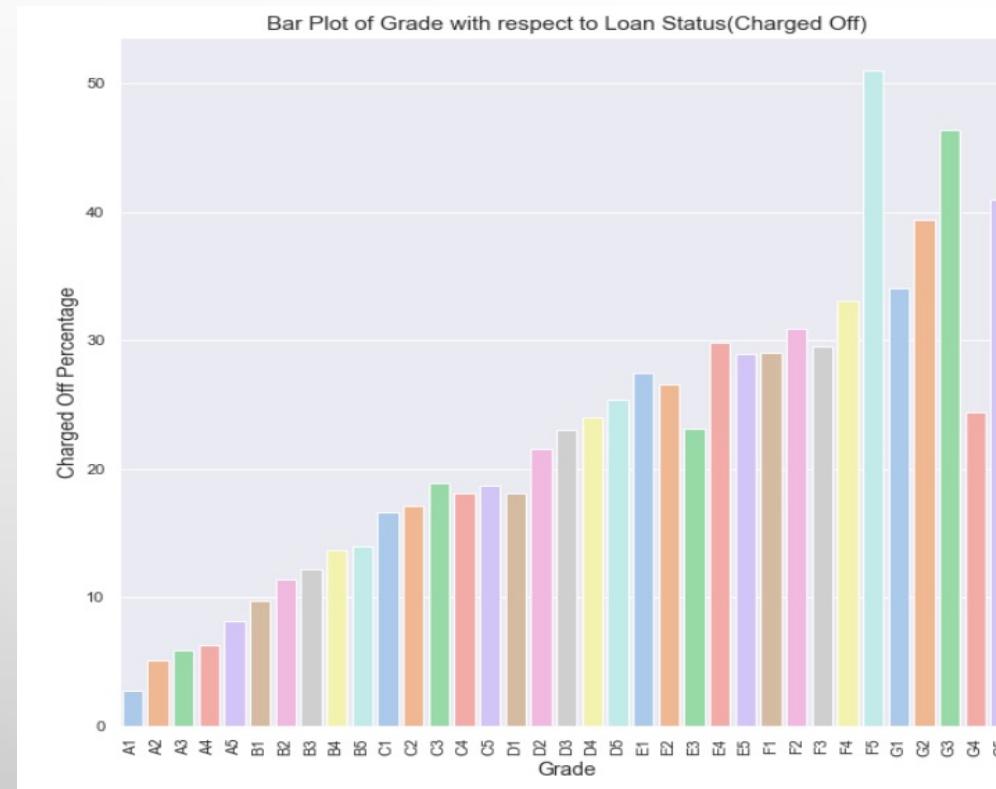
FUNDED AMOUNT INVESTED VS MONTH AND YEAR



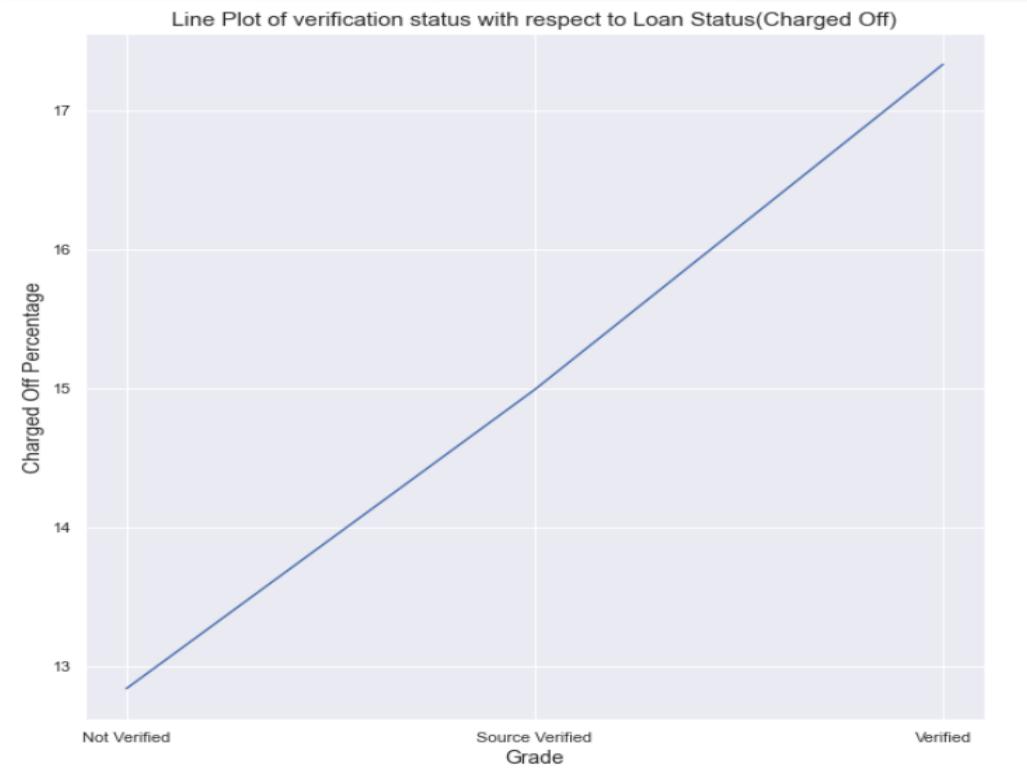
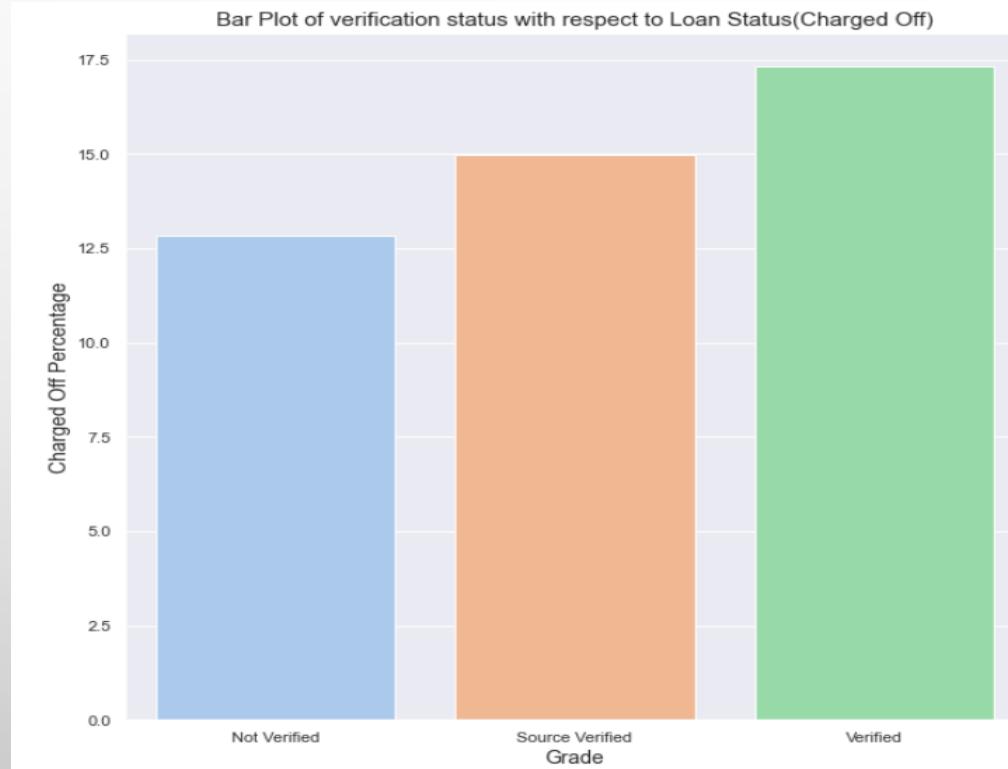
FUNDED AMOUNT INVESTED VS GRADE



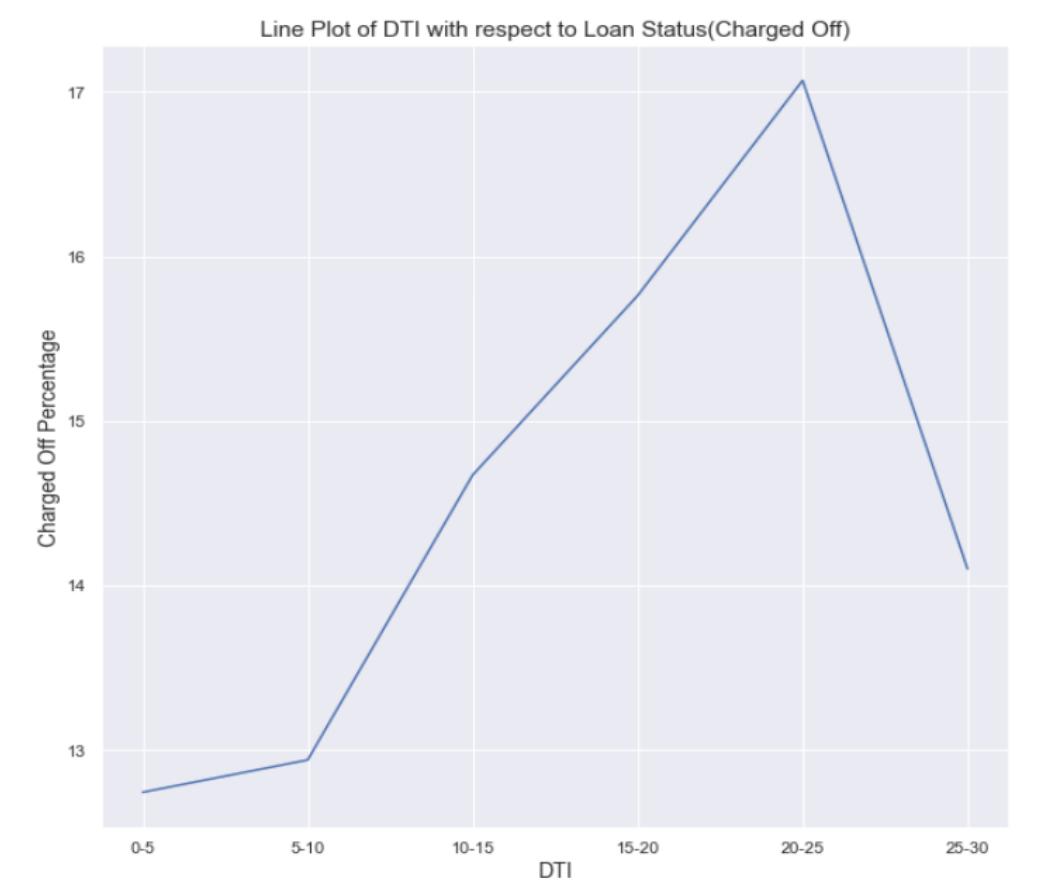
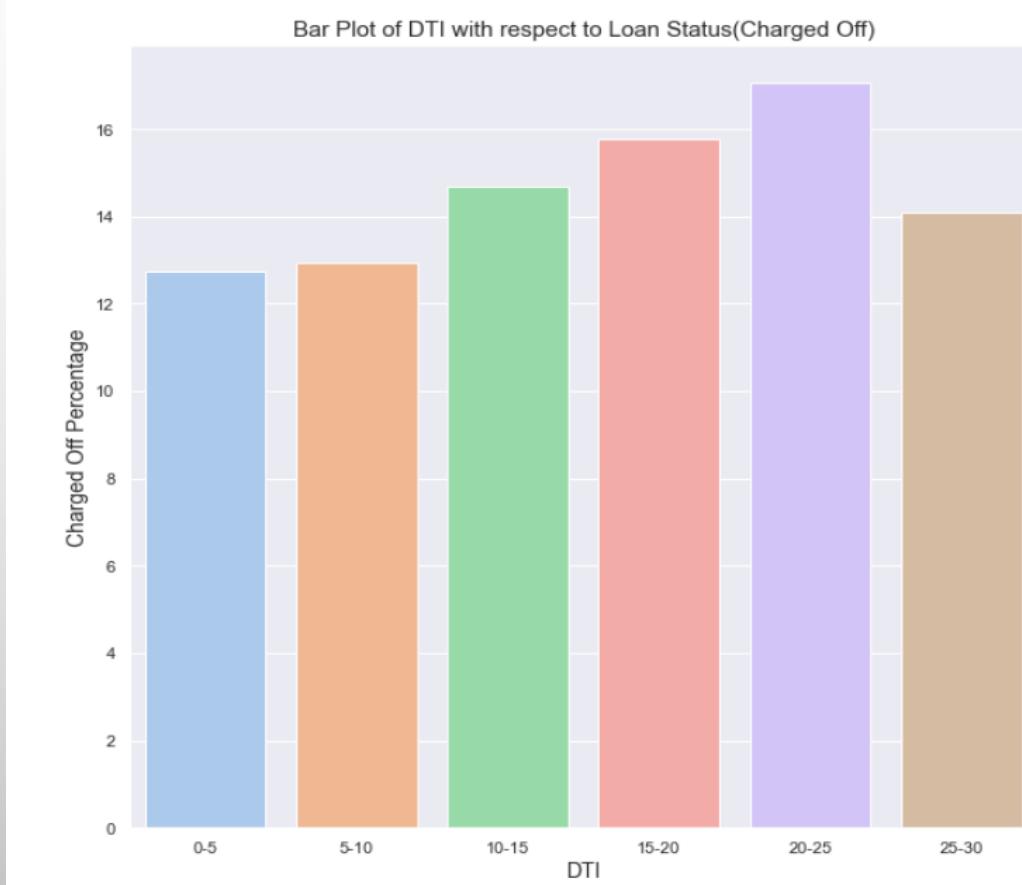
FUNDED AMOUNT INVESTED VS SUB GRADE



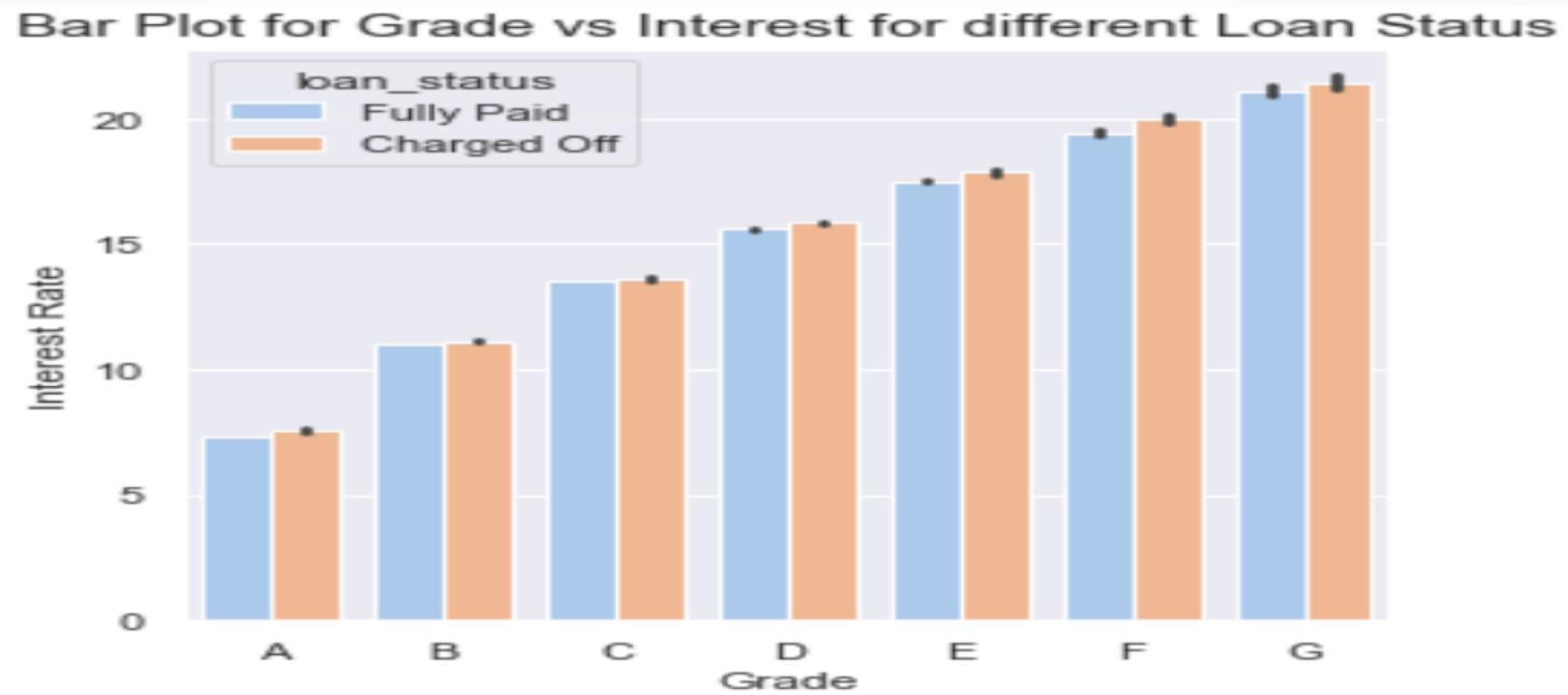
FUNDED AMOUNT INVESTED VS VERIFICATION STATUS



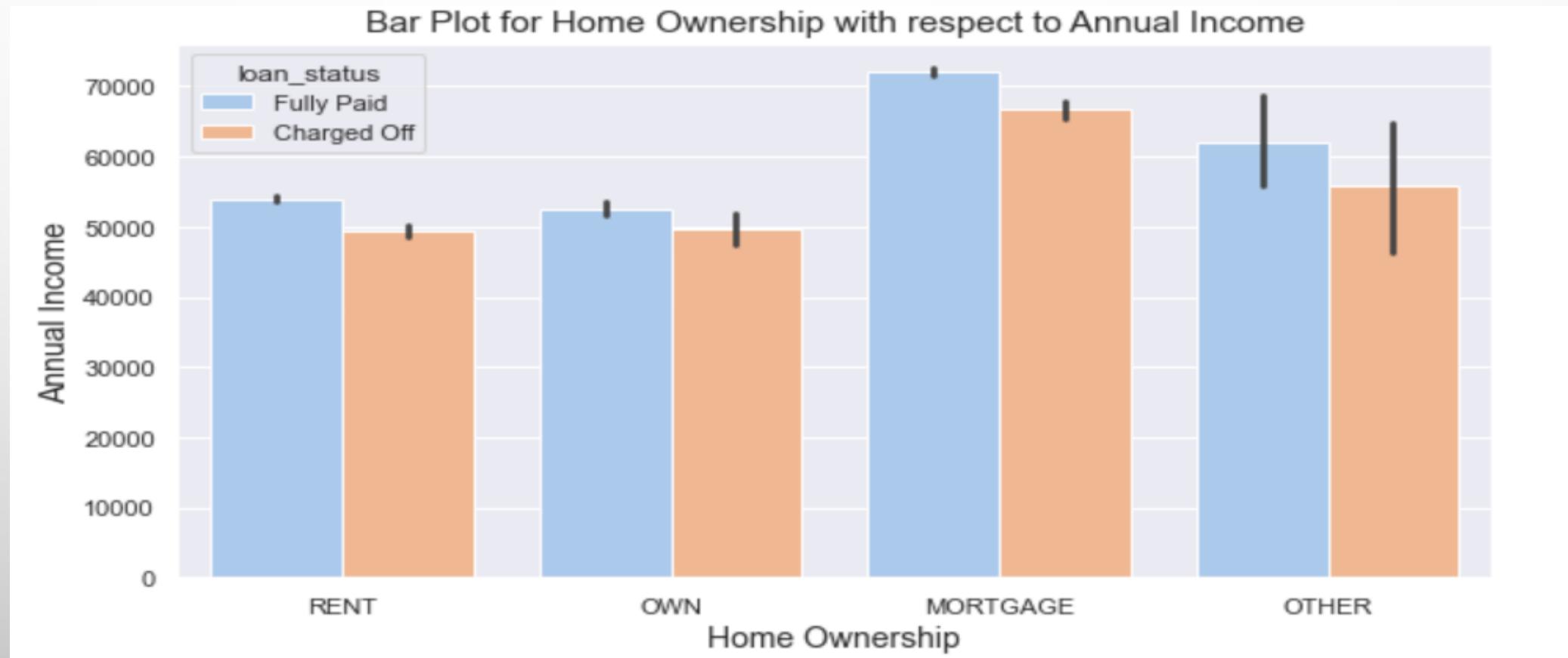
FUNDED AMOUNT INVESTED VS DTI



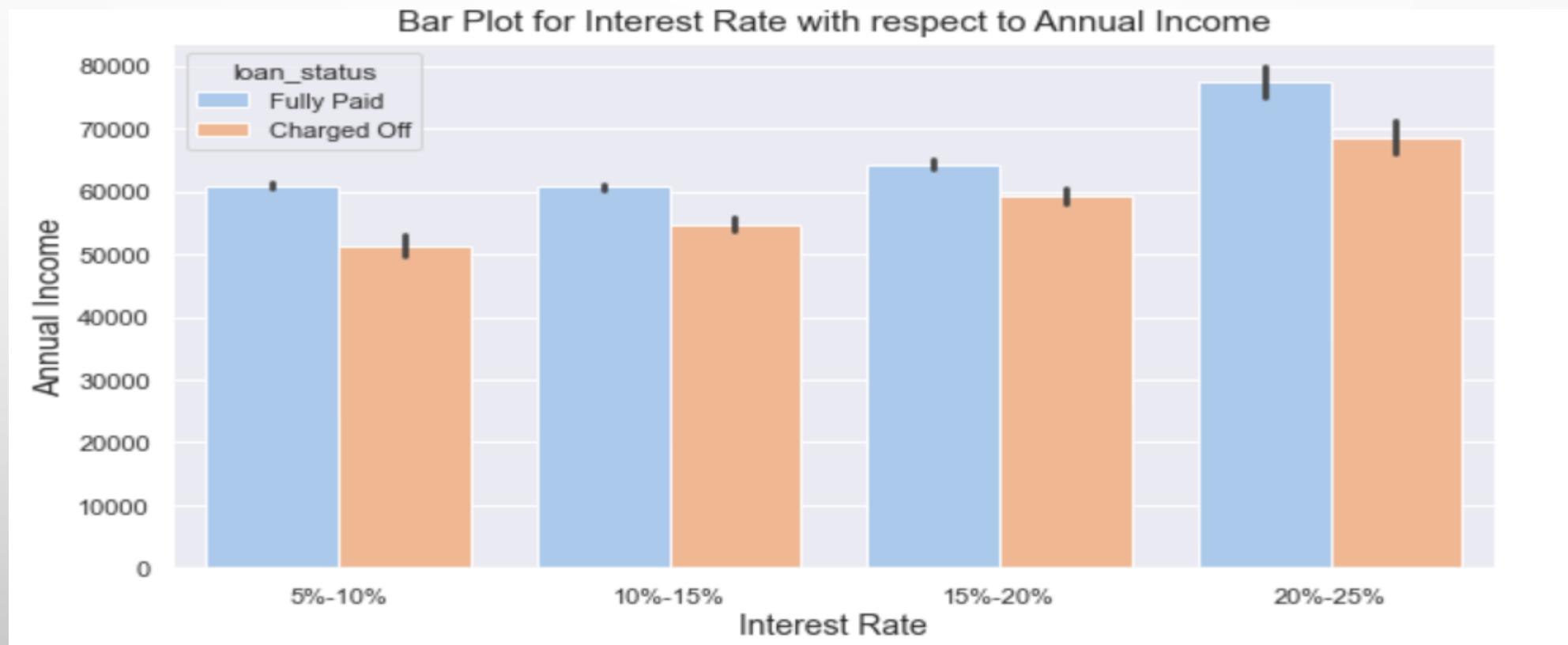
GRADE VS INTEREST



ANNUAL INCOME VS HOME OWNERSHIP



ANNUAL INCOME VS INTEREST RATE



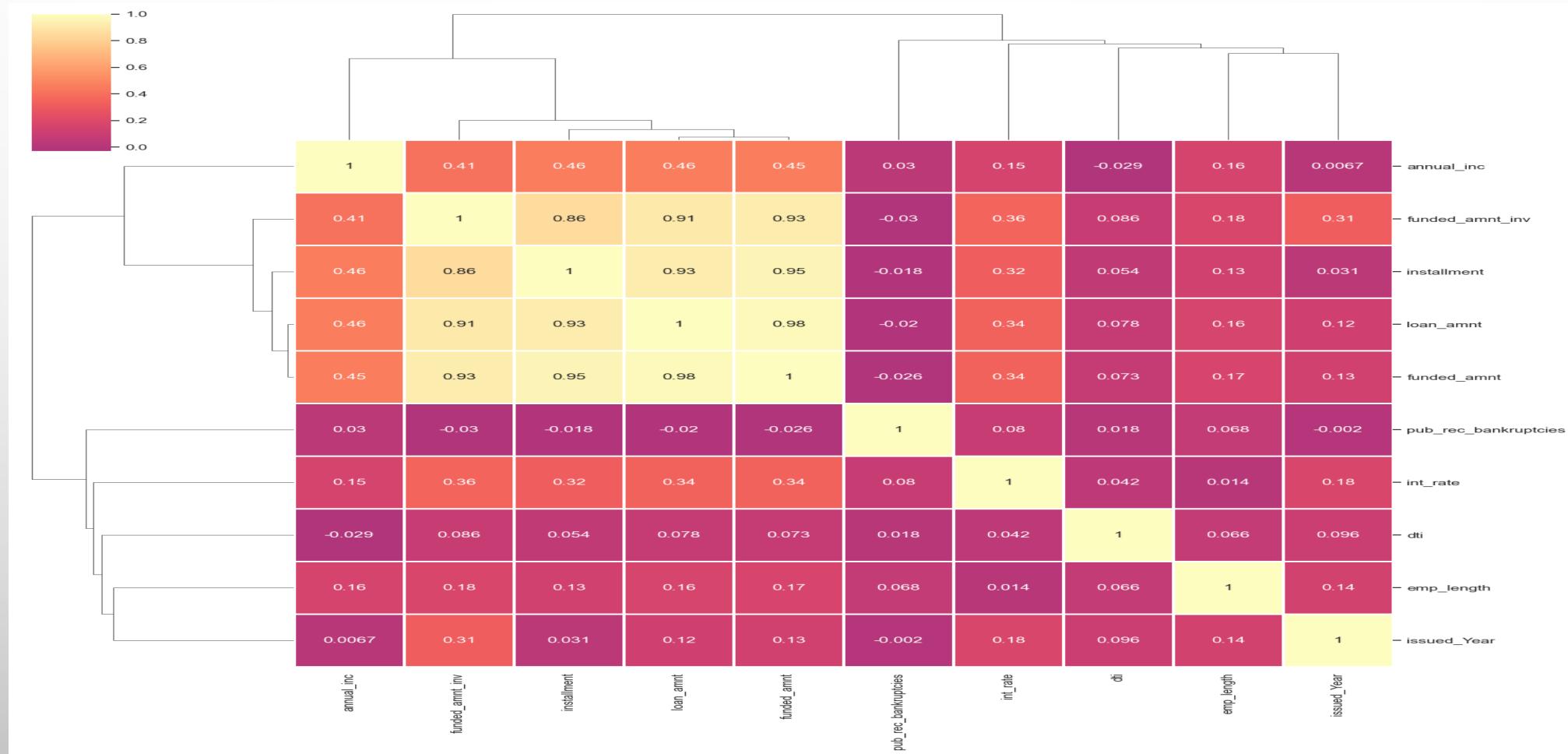
HOME OWNERSHIP VS PURPOSE



ANNUAL INCOME VS PURPOSE



CORRELATION



OBSERVATIONS

- **THESE ARE THE MAJOR VARIABLES THAT HELPED US TO PERFORM DATA ANALYSIS :-**
- 1. LOAN PURPOSE
- 2. EMPLOYMENT LENGTH
- 3. INTEREST RATE
- 4. ANNUAL INCOME 5
- . GRADE AND SUB GRADE
- 6. TERM
- 7. ISSUED MONTH AND YEAR
- 8. VERIFICATION STATUS
- 9. HOME OWNERSHIP
- 10. DTI

INSIGHTS

- LOAN PURPOSE :-
 - SMALL_BUSINESS -> 27.95 %. (OUT OF 1624 LOANS TAKEN BY KEEPING LOAN PURPOSE AS "SMALL_BUSINESS" 454 LOANS CHARGED OFF)
 - RENEWABLE_ENERGY -> 19.14 %. (OUT OF 94 LOANS TAKEN BY KEEPING LOAN PURPOSE AS "RENEWABLE_ENERGY" 18 LOANS CHARGED OFF)
 - EDUCATIONAL -> 17.70 %. (OUT OF 305 LOANS TAKEN BY KEEPING LOAN PURPOSE AS "EDUCATIONAL" 54 LOANS CHARGED OFF)

- Employment Length
 - The charged Off percentage is high for the loans where the Employment length is equal to 10+ years.
 - 10+ years is 16.78 %
 - 7 years is 15.28 %
 - 1 year is 14.64 %
- Interest Rate
 - Charged Off percentage is high for Interest Rates :-
 - 20-25% is 39.03%
 - 15-20% is 24.46%
- Annual Income
 - Charged Off percentage is high for Annual Incomes :-
 - 0-20k is 20%
 - 20k-40k is 18%
 - 40k-60k is 15%

- Verification Status
 - Charged Off percentage is high for the Verification Statuses :-
 - Verified is 17%
 - Source Verified is 15%
 - Not Verified is 13%
- Home Ownership
 - Charged Off Percentage is high for the Home Ownership values :-
 - OTHER is 17%
 - RENT is 15%
 - OWN is 15%
- DTI
 - Charged Off Percentage is high for the DTI values:-
 - 20-25% is 17%
 - 15-20% is 16%

- Grade and Sub Grade
 - Most of the Charged Off applicants belong to Grades G, F, E. For Grades G, F the most charged off sub grades are:-
 - G is G3, G5
 - F is F4, F5
- Term
 - For term, As the Term increases the Charged Off Percentage also Increases.
- Issued Month and Year
 - 2007 is 18%
 - 2011 is 16%
 - 2008 is 15%

INSIGHTS – UNIVARIATE VS BIVARIATE

- Although we got insights from Univariate analysis at the beginning those insight were contradicted when we performed Bivariate analysis Examples :-Grades - The Univariate Analysis gave almost the opposite insights of Bivariate Analysis,
- Term
- Sub Grades,.

CONCLUSION

- LESS IMPACT METRICS :-
 - Lower Annual Income, In the range 20,000 to 40,000.
 - Higher Debt to Income Ratio, Between 15% - 20%.
 - Higher Loan Amount (Above 15600)
 - Loan Purpose - Education
 - Loan Grade and Subgrade - B and C
- HIGH IMPACT METRICS :-
 - Loan Grade and Subgrade - Greater than D
 - Loan Purpose - Small Business, Renewable Energy
 - High Interest Rate - above 15%
 - Lower Annual Income - Below 20,000
- COMBINED IMPACT ABOVE METRICS :-
 - Home Ownership vs Loan Purpose - Other vs moving, car, small_business
 - Annual Income vs Loan Purpose - 0k to 20k it is better not to provide loans for the purposes "medical", "renewable_energy", "small_business"