

Lending Club Analysis

Description/subtitle/speaker

Speaker name

Organization

2021-01-01

Problem Statement and Analysis Approach

- Problem Statement : To Analyze the Lending club data set with all the information on loan given last few years and understand how some variable impacts or not towards loan success to have a loan decision made for new loans . It is important to understand the parameters which should be considered in taking loan approval decision and if some actions can be taken to secure loan customer
- Analysis Approach :
 - Understand credit funding parameters and factors affecting same through public knowledge and apply same on data set.
 - Use EDA principles to analyse the data set
 - Clean , remove, rename and derive parameters (columns)
 - Use Data visualization methods to understand univariate and Bi variate patterns
 - Describe inferences for parameters impacting credit funding decision.

Factors for Risk analysis of Credit Lending

- 5 C of Credit – Character , Capacity, Collateral , Capital , Conditions
- Credit score factors : Payment history-35% , Amounts owned -30% , Credit history length (15%) , Credit mix(10%), New credits opened (10%).
 - Credit utilization ratio = total revolving credit being used/ total of all revolving credit limits. Using more than 30% of your available credit is not good.
 - Credit mix : car loan student loan , credit card, mortgage and other credit prods , diversified mix and handling helps credit score.
 - The longer the credit history , the higher the credit score.
 - Number of hard enquiries lenders make towards borrower. Too much requests made in short time

| Factors | Loan Data Set parameters | Univariate/Bivariate Analysis | Remarks |
|---|--|--|---|
| Character Factors | | | |
| Personal details – education background / employment history , business references , Domicile | emp_length, emp_title , addr_state, zip_code | <ul style="list-style-type: none"> Check if employment length has impact on loan closure or delinquency ? Check if Addr/zip code has any relation to credit lending based on past behavior | Higher education / employment length is good. No data for educational background and biz references |
| Applicant's Credit history/Credit score? | To be calculated – Derived Metric | | Clean credit history is good. |
| Previous bankruptcy / Past debt payment history / recoveries / tax liens | chargeoff_within_12_mnths, mths_since_last_delinq, delinq_2yrs , pub_rec, pub_rec_bankruptcies, recoveries | | |
| Borrower detail verification | verification status, inq_last_6mths ,mths_since_last_record | | verified_status_joint, inq_last_12m, inq_fi , |
| Capacity Factors | | | |
| Annual Income / savings / additional means of capital | annual_inc , home_ownership | <ul style="list-style-type: none"> Check Annual Income and home ownership impact to failed loans | No data [annual_inc_joint , avg_cur_bal] , |
| Debt to Income Ratio | dti | <ul style="list-style-type: none"> Impact of DTI to failed loans | High debt to income ratio is bad No data [dti_joint] |
| Balance to credit limit ratio / revolving rate utilization rate | revol_bal, revol_util | | Having low ratio improves credit score , revol_util < 30% is good. No data - all_util (needs name change) , bc_util , il_util |
| Mortgage accounts , liabilities | total_acc | | No data [mort_Acc] |

5 C factors and Data set reference

| Factors | Loan Data Set parameters | Univariate/Bivariate Analysis | Remarks |
|---|--|-------------------------------|--|
| Conditions Factors | | | |
| Interest rate of the loan , term payments count , installment | int_rate , term, installment | | |
| Loan amount | loan_amnt, funded_amnt, funded_amnt_inv, out_prncp | | Loan amount |
| Purpose of the loan , description , need, title | purpose , desc , title | | Purpose of the loan , description , need, title |
| loan grade , sub grade | grade , sub_grade | | loan grade |
| Collateral Factors | | | |
| Collateral | home_ownership | | |
| Debt to Income Ratio | dti | | High debt to income ratio is bad No data [dti_joint] |
| Capital Factors | No parameters match | | |
| | | | |

Data Cleaning Topics

- Remove all **Nan** only Columns since it makes no sense
- Remove all **zero** only Columns since it makes no sense to have 0 value
- Remove all **zero and Nan** value only Columns , if it makes no sense 0 or Nan value
- Remove not so useful columns for credit analysis or constant data – [*payment plan, initial_list_status, policy_code, tax_liens, application_type, collections_12_mths_ex_med, url, zip-code.*]
- Clean Columns on Data types : *emp_length(remove years), issue_d, last_pymnt_d, earliest_credit_lines* into datetime mm-yyyy format instead of string.
- Clean Columns on similar values : purpose (collides with title ,desc) ,
- Check for duplicates – if same *member_id or Id* ? If repeat entries , if so why or usable?
- Check for data types and modify them , convert date columns from string to date, change numerical value columns into float or int from string , remove unwanted characters
- Rename columns with meaningful names for easy identification

Biz driven Derived metrics : Credit Score calculation

- Since there is no Credit score given in Dataset , it would probably be a biz driven derived metric.

<https://www.myfico.com/credit-education/whats-in-your-credit-score>

<https://www.cnbc.com/select/this-is-the-most-important-factor-that-determines-your-credit-score/>

1) Payment history (35% of score) = fx { chargeoff_within_12_mths, delinq_2yrs, pub_rec , pub_rec_bankruptcies }

2) Amounts owed (30% of Score) = fx{ dti , revol_util , revol_bal (credit_utilization_ratio) , }

3) Credit History Length (15% of Score) = fx{ earliest_cr_line , loandurationmonths }

4) Credit Mix (10% Score) = fx{opencreditlines}

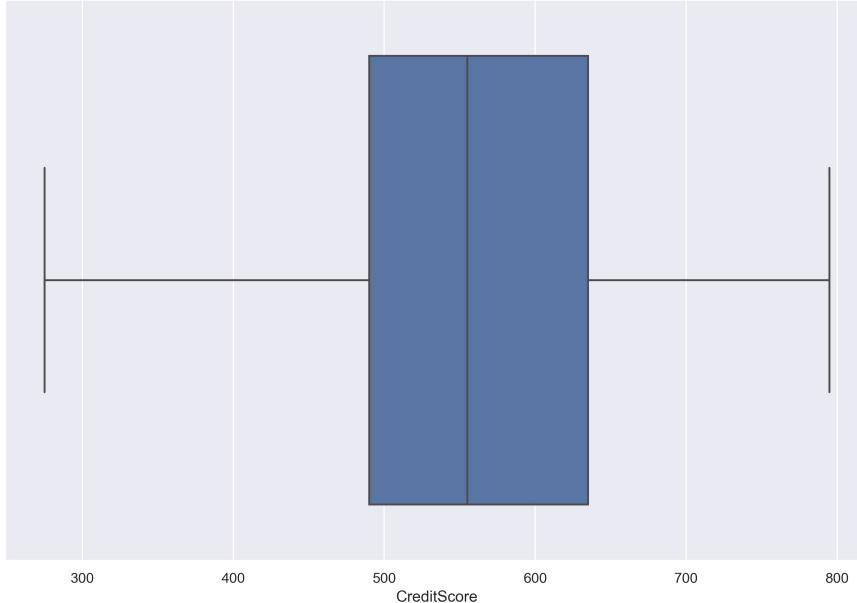
5) New Credits Opened (10% Score) = fx{inq_last_6mths }

Other Derived Metrics

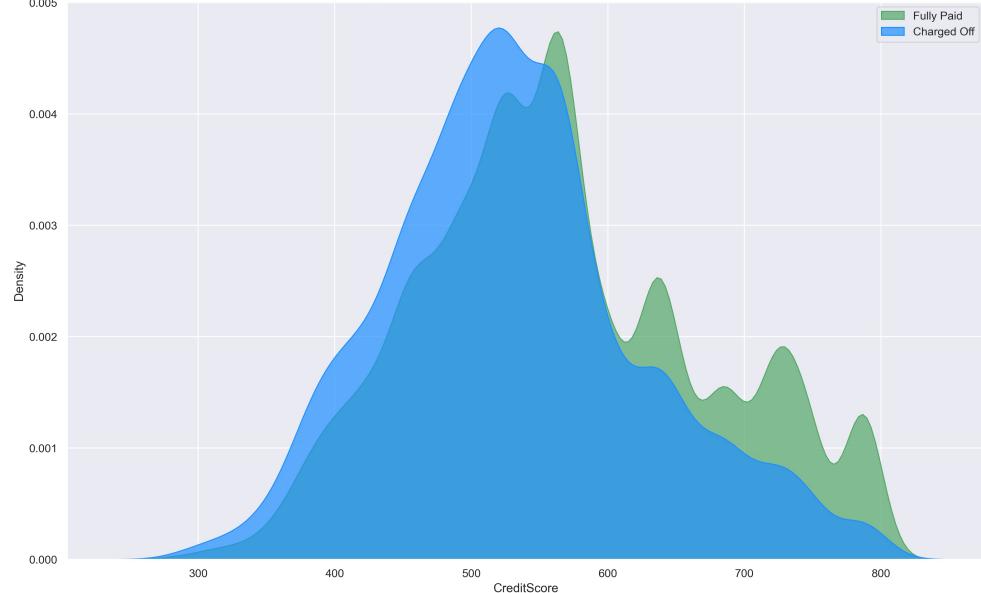
- annual inc buckets for comparison with Loan status / Credit Score.
- creditHistoryLength = earlies_cr_line - current month(Jun16) --> history length in months
- loandurationmonths = total_pymnt / installment
- Total interest paid months = total received interest/ interest
- CreditScoreBucket based on Credit Scores
- Interest Rate Bucket based on Interest rates

Univariate Analysis Charts

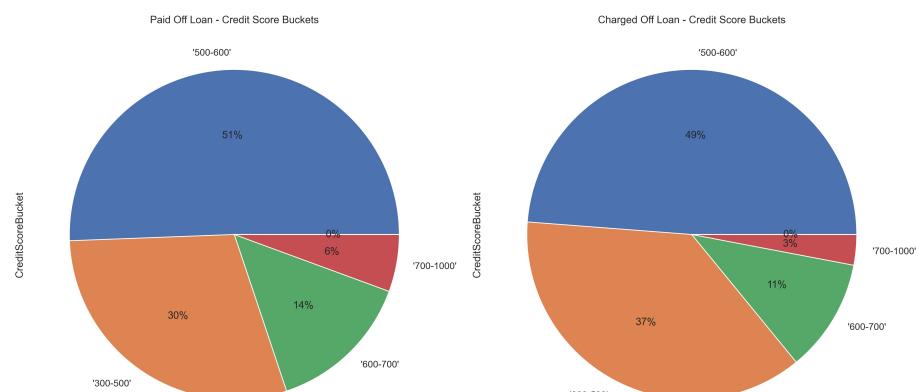
Distribution Analysis of Credit Score



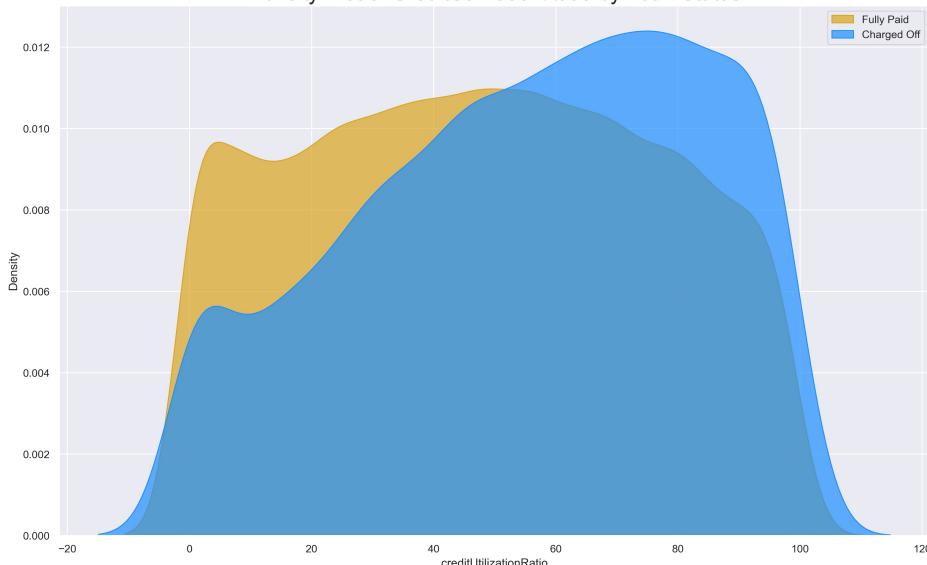
Density Plot of Credit Score by Loan Status



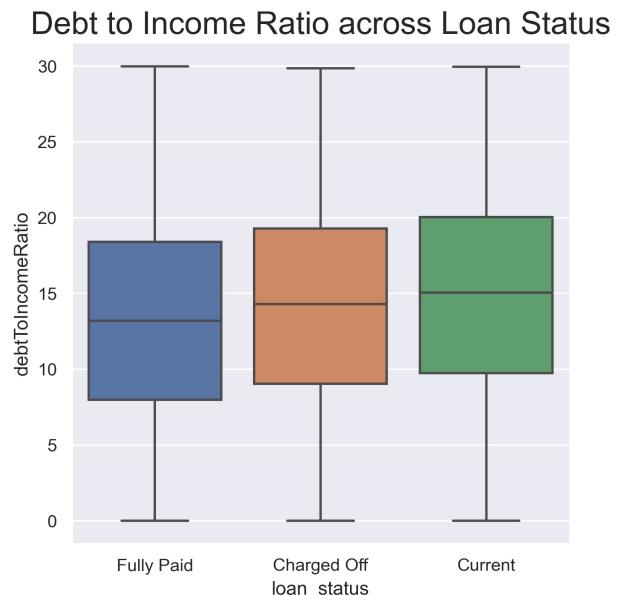
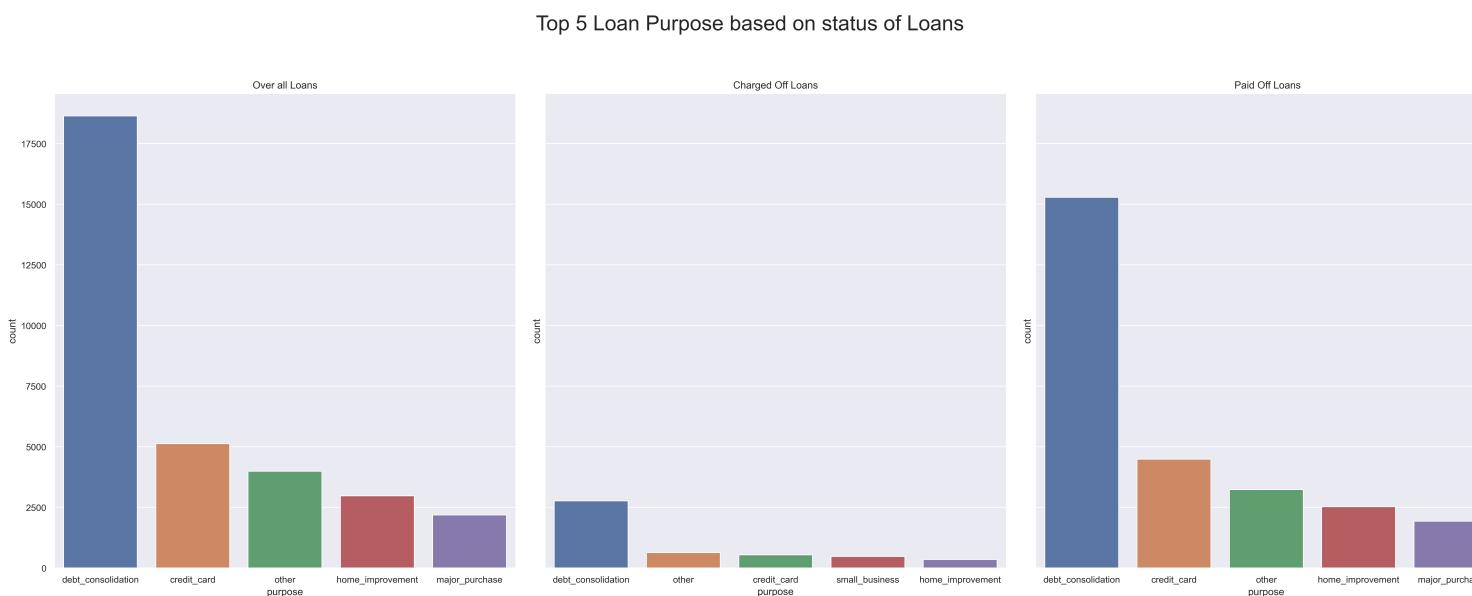
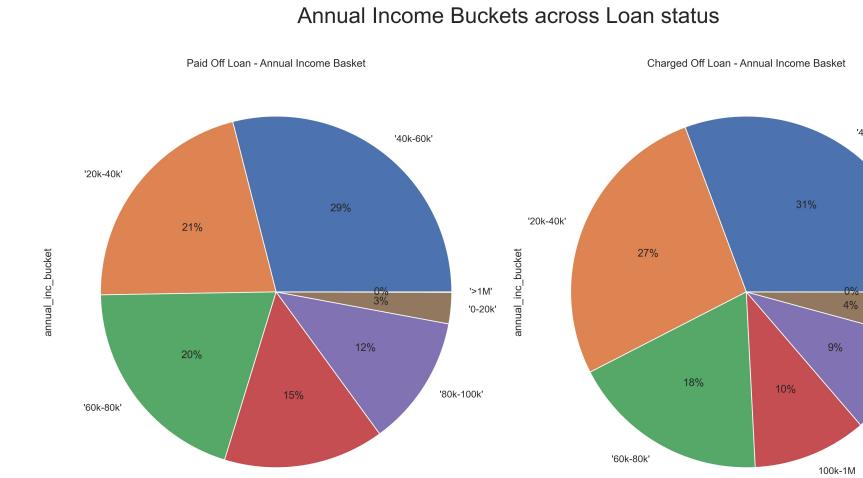
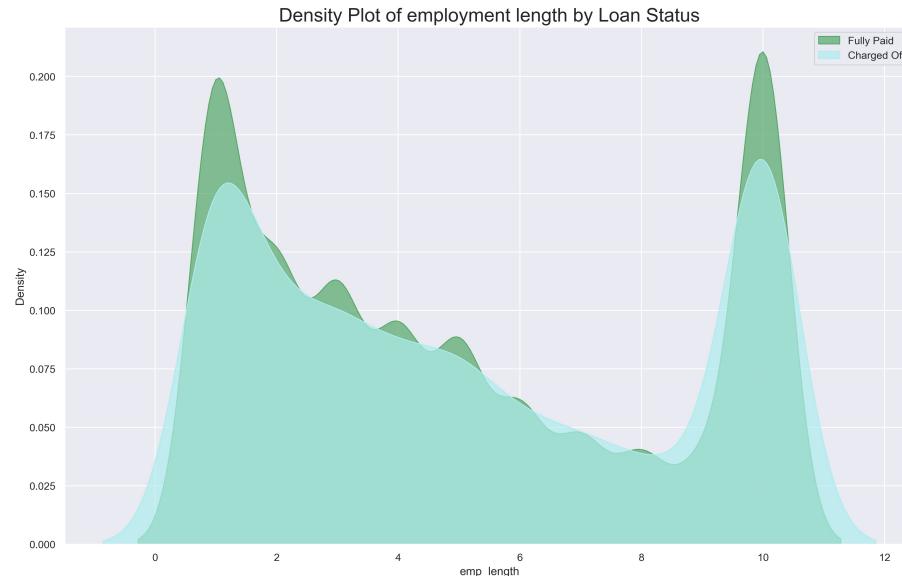
Credit Score Buckets across Loan status



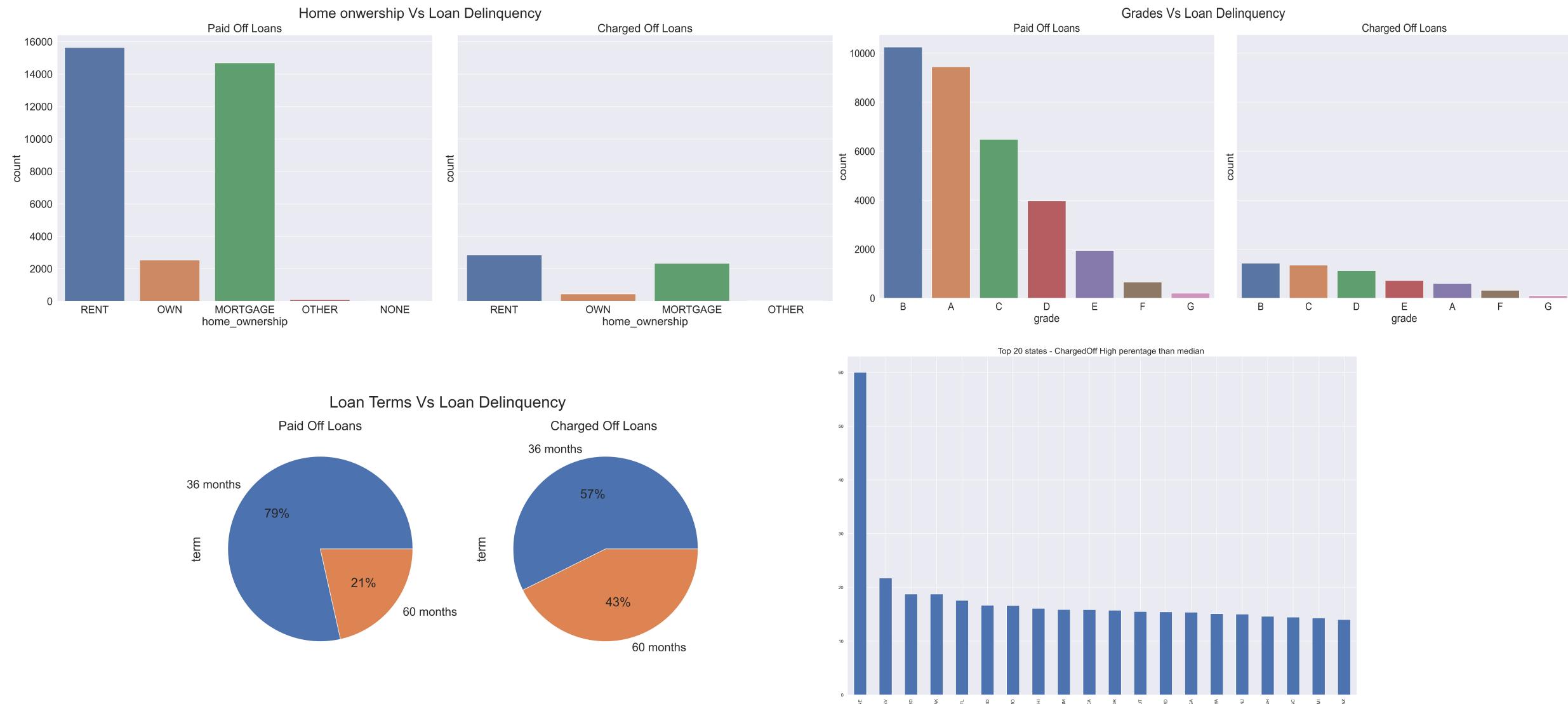
Density Plot of CreditUtilizationRatio by Loan Status



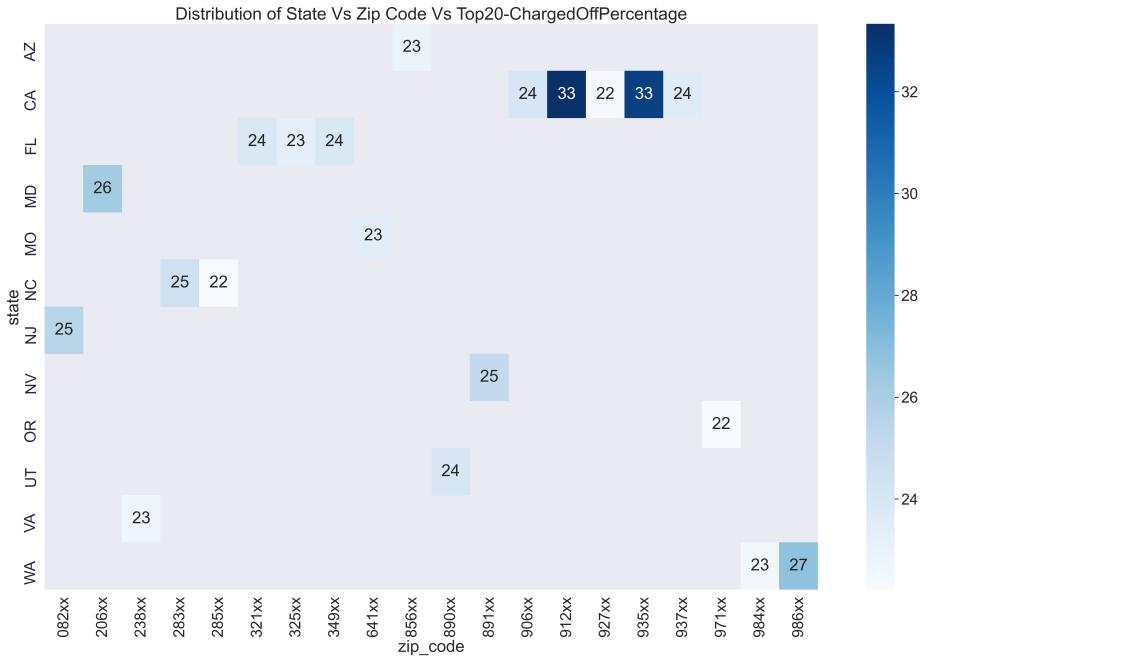
Univariate Analysis Charts



Univariate Analysis Charts



Univariate Analysis Charts

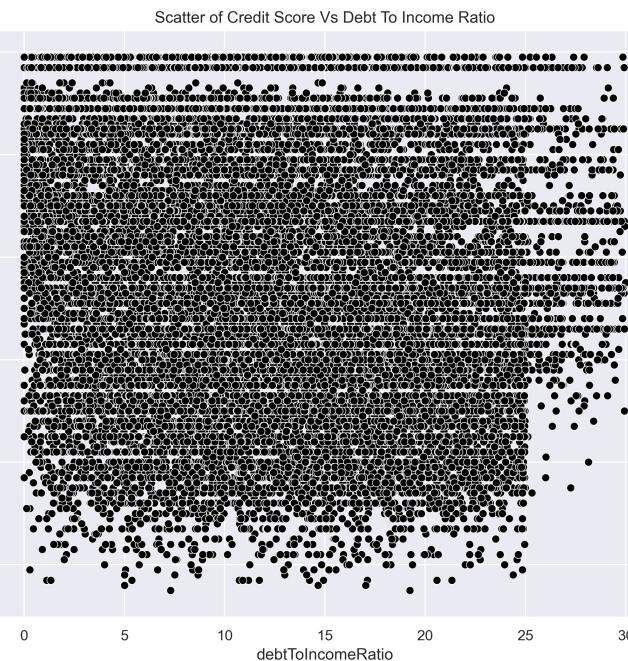


Summary Statements – Univariate Analysis

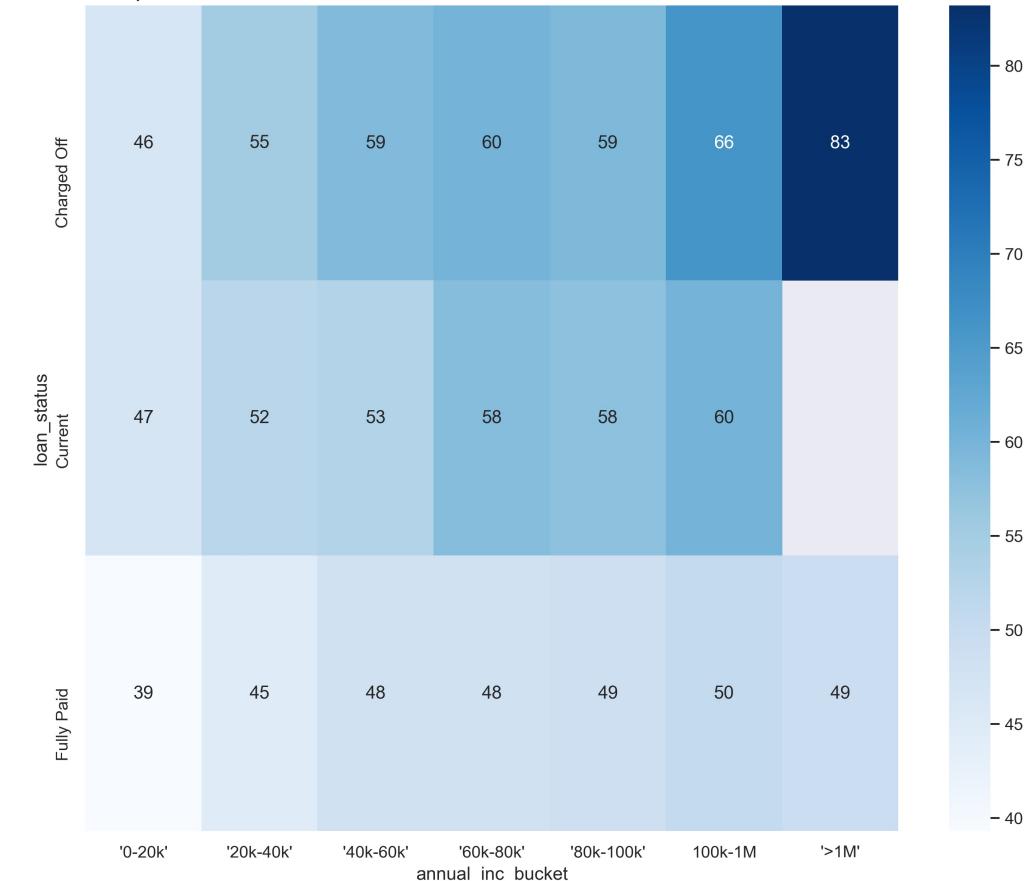
- Credit Score Median is around 530-560 points for all and charged off loans , but there is higher percentage of 300-500 bucket Credit Scores in ChargedOff loan category
- Low Credit Score points bucket is high for charged off loans and although it cannot be said if a 500-600 or 700-100 credit Score will not lead to failed loan , but low credit Score has more probability of chargedOff loan
- Credit Utilization ratio has an impact to loan status the higher the ratio , there are more cases of ChargedOff loans.
- Debt to Income ratio has slight relation to loan success , chargedOff case are of higher ratio value , it shows more debts cause loan delinquency.
- 'Debt consolidation' seems to be prominent purpose/reason for loans and it also contributes to maximum charged off loans but top5 purpose analysis shows no connection to loan success
- Annual income bucket does not have major relation to loan delinquency.
- Employment Length, purpose , home ownership , zip code, grades has minimal impact/ connection to paid or chargedOff loans
- Long Term loans have higher percentage to loan failure compared to short term loans
- There are around 13 states which have high chargedoff Loan percentage (more than 75th percentile of lot) , especially states such as Nebraska is very high – could be red alert state- 3/5 cases !!! , Nevada , South Dakota and alaska being top 3
- There are few zip codes in some states which show high percentage of failed loans above 25% failure.

Bivariate Analysis Charts

Credit Score relationship to different capacity ratios

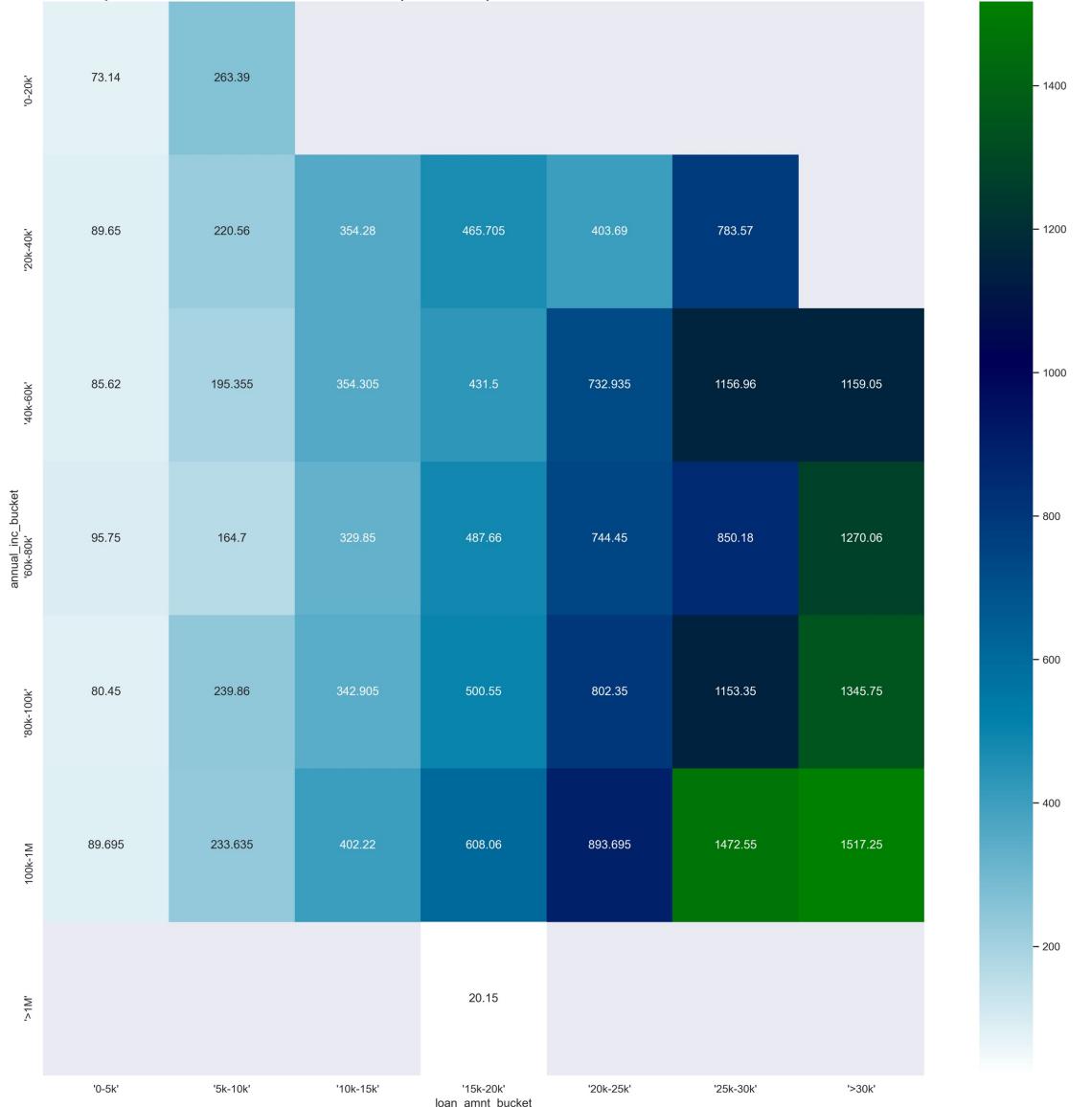


Heat map distribution of Loan Status and Annual Income Bucket on CreditUtilizationRatio

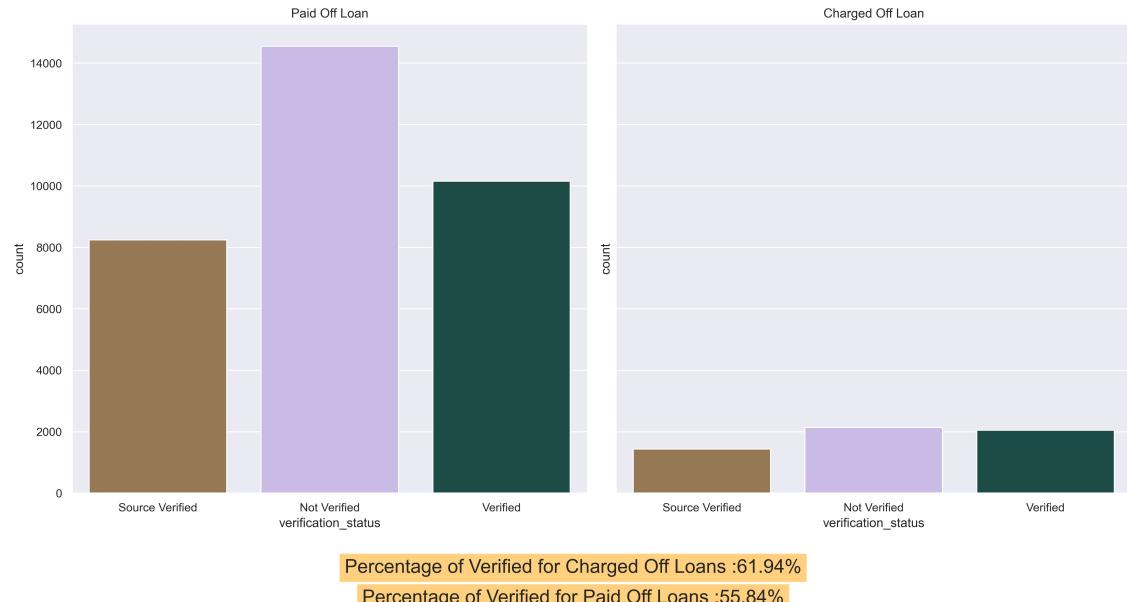


Bivariate Analysis Charts

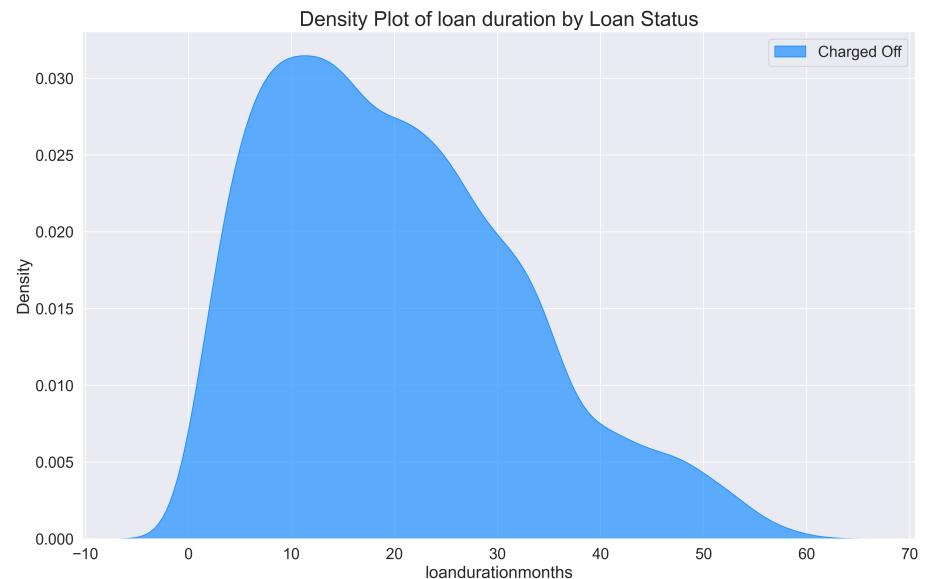
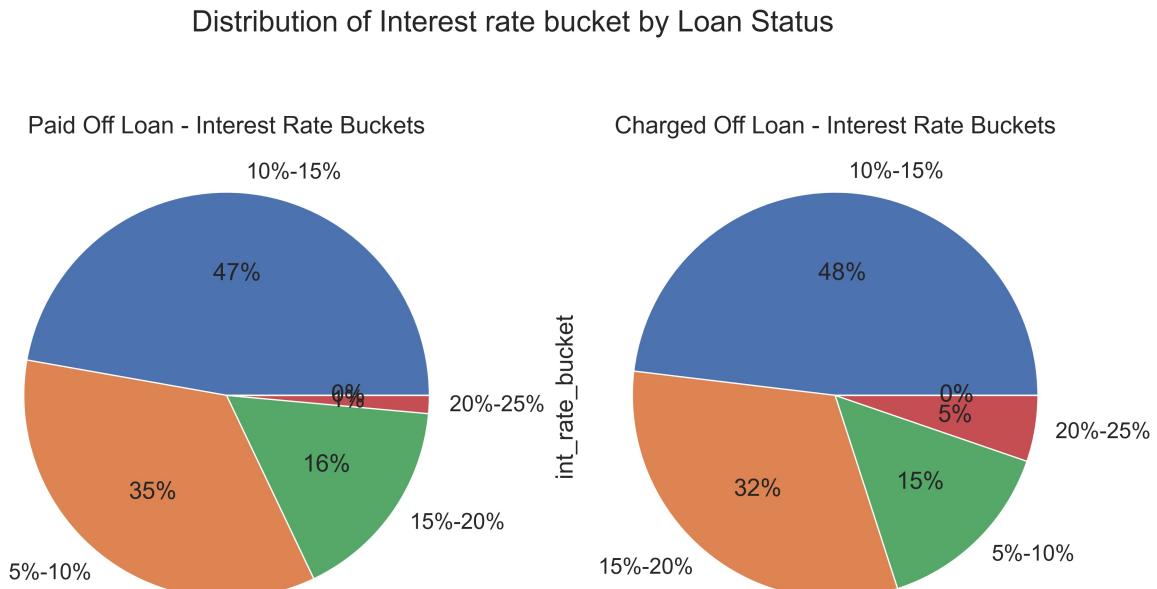
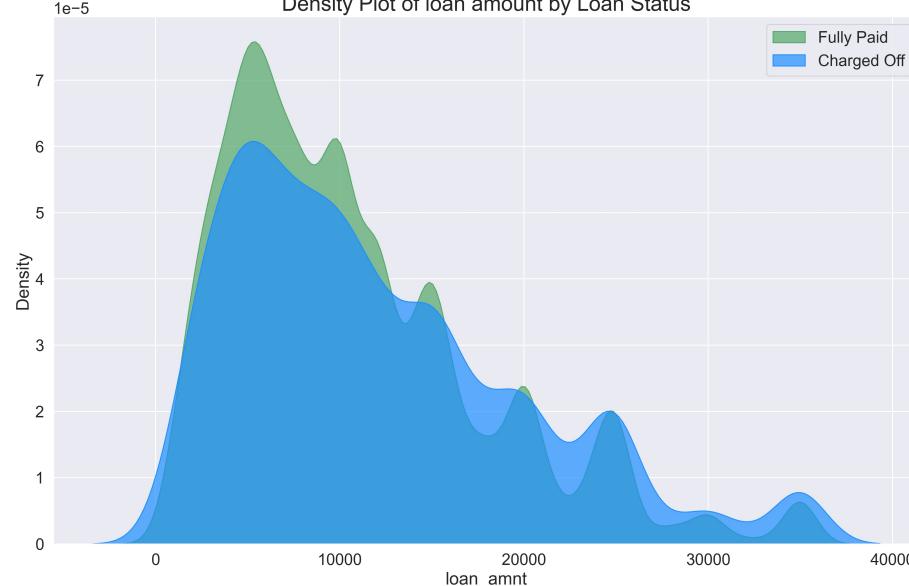
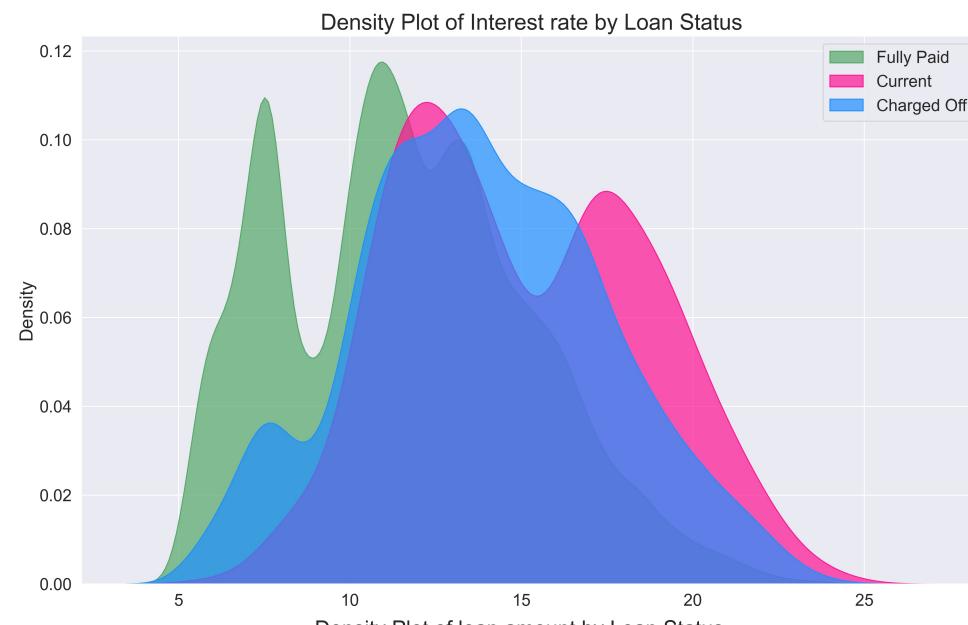
Heat map Distribution of recoveries(Median) on loan amount Vs Annual income Bucket



Verification Status against Loan status



Bivariate Analysis Charts



Summary Statements – Bivariate Analysis

- Credit utilization Ratio has a direct relationship to Credit Score as we know from our own calculation , it also means DTI has no so direct relationship to credit Score, it may be important to see DTI for credit funding, but it has not much effect to Credit Score.
- With Pivot of Credit Utilization Ratio to Annual income bucket and loan status , we see higher the income bucket , for failed loans , the credit utilization ratio is higher , it also could mean higher income bucket with high credit utilization score have more chances for failed loans.
- With Pivot of Annual Income bucket with median recovery amount Vs Loan amount bucket, we see for the same loan amount buckets, recoveries is better with higher annual income buckets.
- Verified Loans have no direct connection to loan failures, both verified and not verified have loan failure cases with almost percentages , infact chargedOff loans have higher percentage of Verification.

Summary Statements – Bivariate Analysis

- Interest Rate ranges are little skewed on lower rates having more paid off loans than higher rates leading to charged Off loans
- 5-15% Interest rate has been more successful for Paid off loans.
- For ChargedOffLoans , we see most of the people have paid instalments max for 10-12 months – almost 35% , so it is a factor to decide getting less damage by planning out interest rate to get maximum out in 1st 12 months.
- There is no correlation of loan amount value Vs Loan success of failure !

Final Summary

- Credit Score and Credit utilization Ration have an important role and decision making factor to grant loans
- Long term and loans for debt consolidations have a skew towards failure chances
- High Loan Interest Rates also have an impactful factor for loan success.
- Certain states and zip codes in them have a considerable factor to decide on loan grant , may be better to watch out more on them or understand why is the case? .
- Although Verification is important , it is no guarantee that loan will succeed if all is well in verification.
- Higher Annual Income bucket is important to consider especially during recoveries.