



EDA CASE STUDY SUBMISSION

Presented by:

Krishnan Vybhava Srinivasan

Unmesh Kumar

Sudesh Kumar Patel

Nithya V

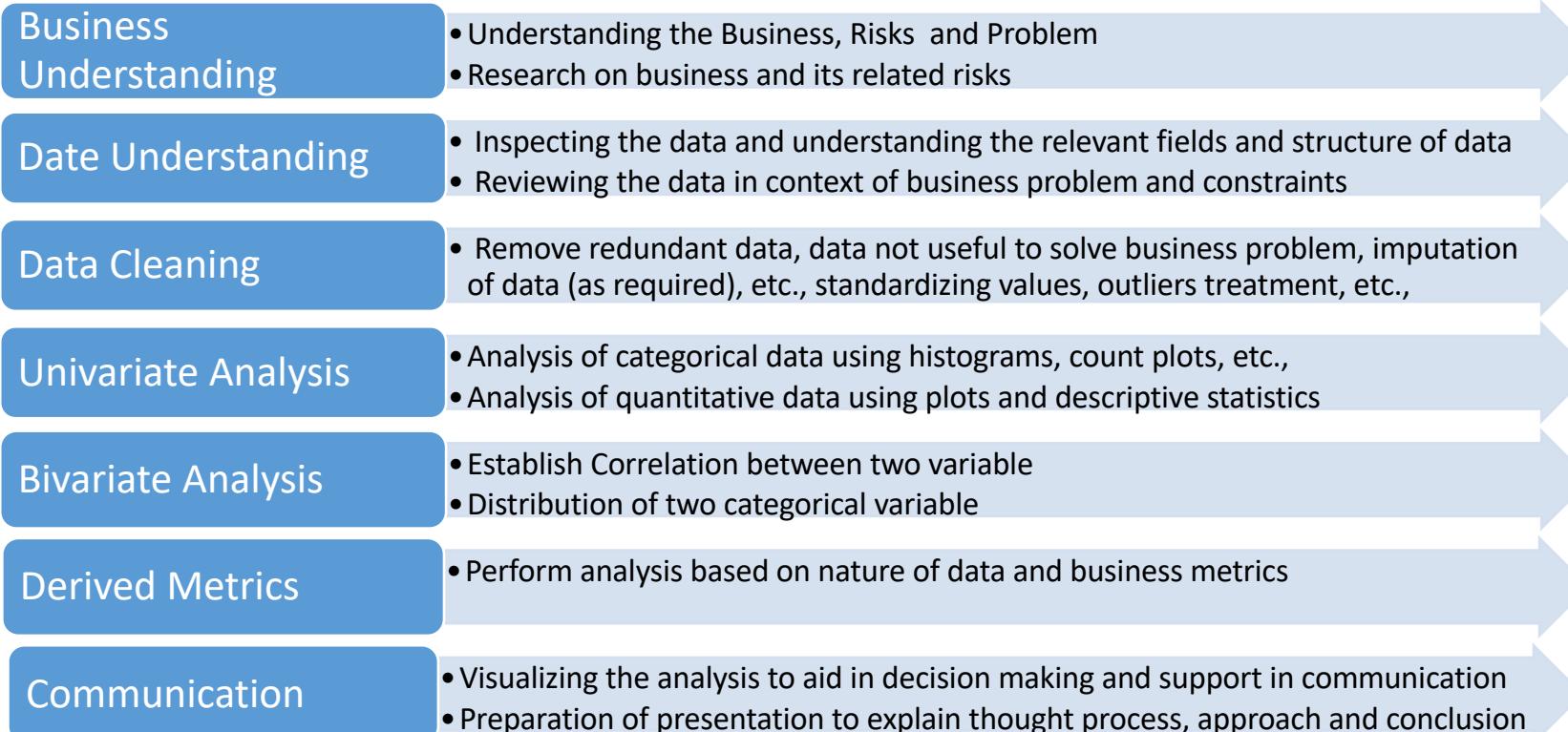
Business Understanding

- About the Business
 - finance company which specialises in lending loans to urban customers;
 - largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures
- Business Risks resulting in Financial Loss
 - Not approving a loan where a customer is likely to repay;
 - Approving a loan where a customer is likely not to repay

My Understanding

Objective	Problem Statement	Data Understanding	Data Analysis Goal
<ul style="list-style-type: none">Understand the driving factors resulting in a loan default	<ul style="list-style-type: none">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	<ul style="list-style-type: none">The data given contains on information about past loanUnderstand how consumer attributes and loan attributes influence the tendency of default	<ul style="list-style-type: none">The aim is to identify patterns which indicate if a person is likely to default,This insight could be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc

Our Approach



Software Tools - Python and Tableau

Data Cleaning

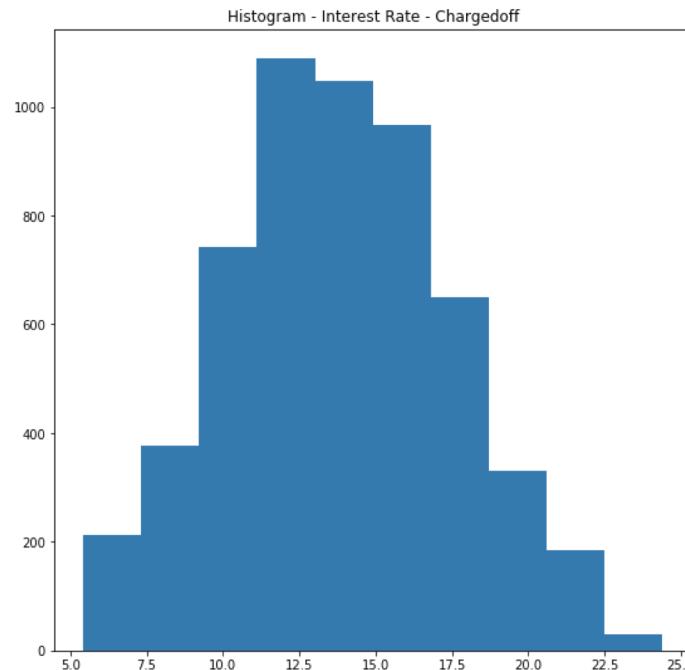
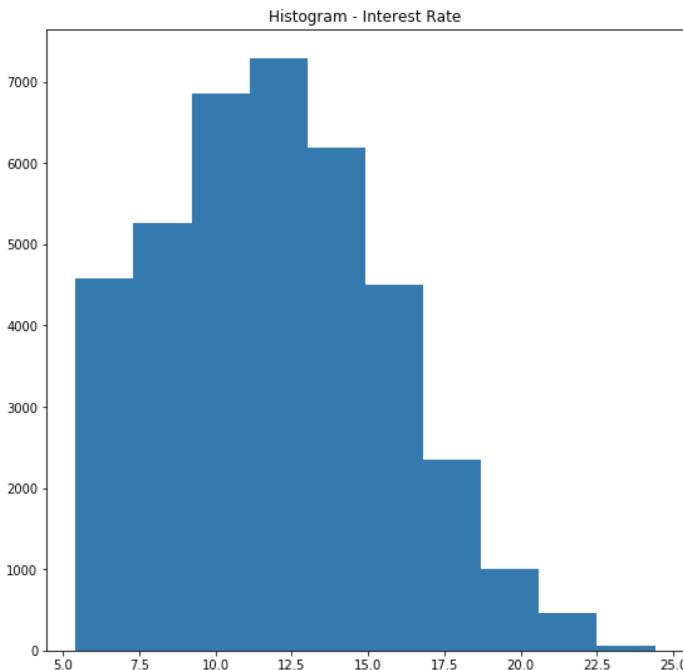
- The original data contained 39717 rows and 111 columns
- 55 columns with only Null or significantly Null values hence were dropped
- 27 columns where not found useful ('0' values) or not relevant to solve the business problem, hence were dropped
- Balance columns assessed during univariate analysis and decided if relevant for developing insights to solve the problem
- Certain columns created imputed columns such date, month, year, income bucket, etc

Customer Attribute	Loan Attribute
grade	loan_amnt
sub_grade	term
emp_title	int_rate
emp_length	installment
home_ownership	issue_d
annual_inc	loan_status
verification_status	total_rec_prncp
purpose	
zip_code	
addr_state	
dti	
delinq_2yrs	
annual_inc_bucket	

Univariate/ Bivariate/ Derived Metrics

- Univariate Analysis
 - Categorical Fields
 - Count Plots on entire data
 - Counts Plots for loan status – charged off
 - Quantitative Fields
 - Hist plots on entire data
 - Hist plots for loan status – charged off
- Bi-Variate Analysis
 - Filtered for Loan Status Charged off
 - Scatter plots

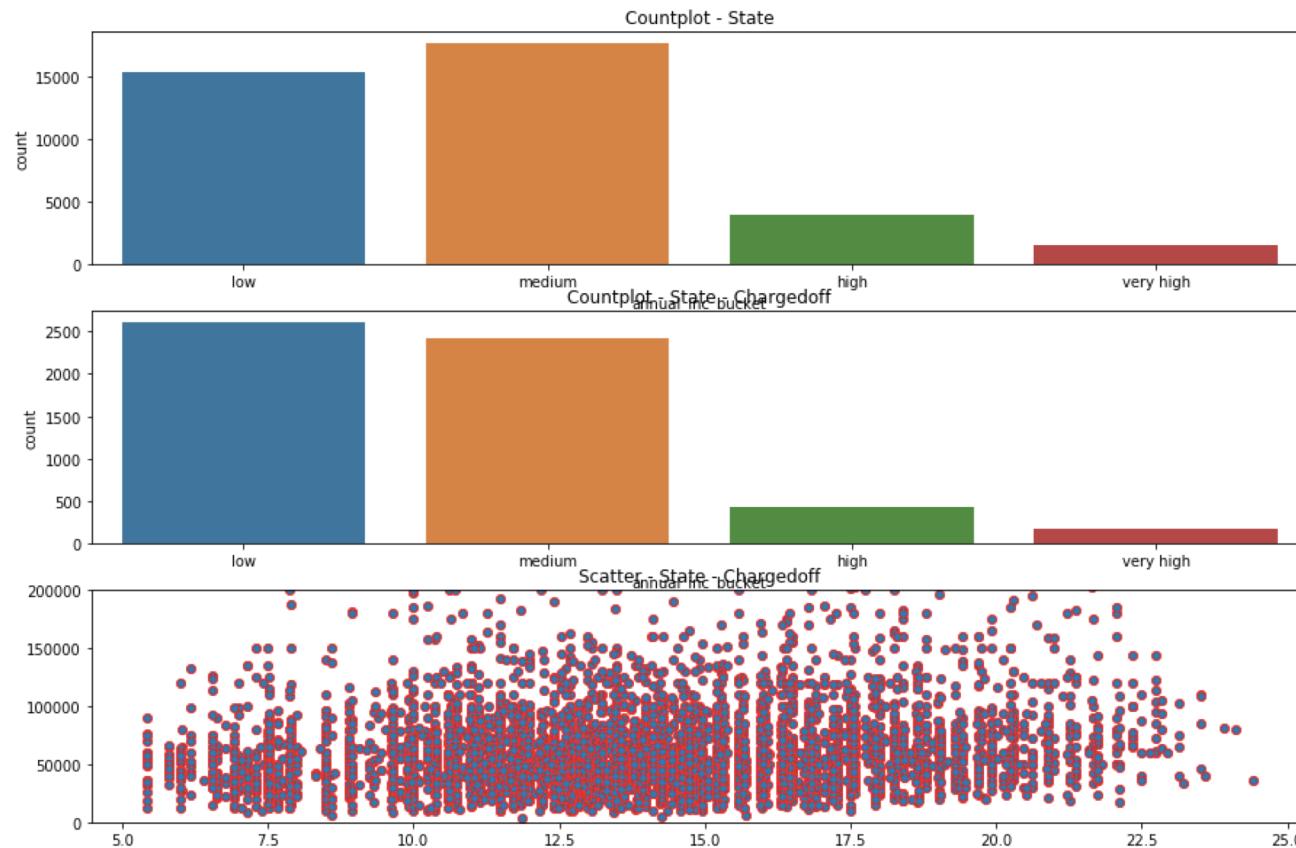
Key Insights – Higher Interest Rate at Higher Risk



Insight – Loans with higher interest rate have higher risk of charge off

Suggestion – Loans with high interest typically are also high risk. Hence there must be adequate precautions and verifications on issue of these loans

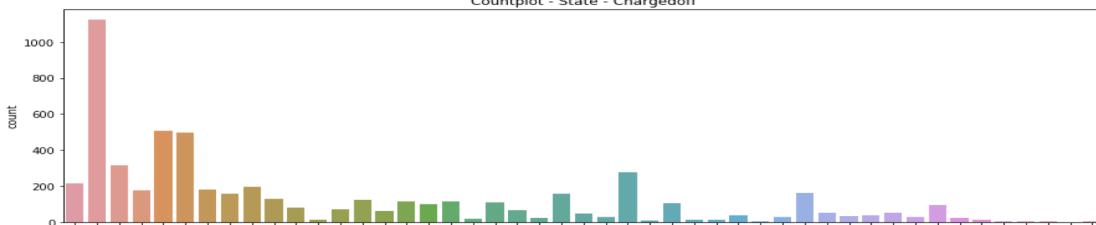
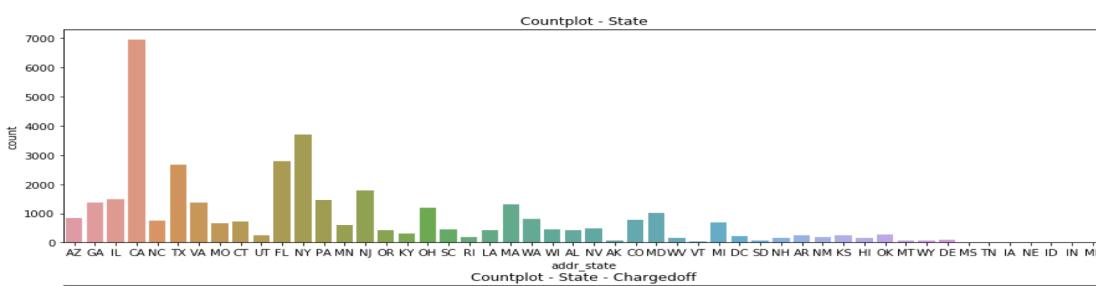
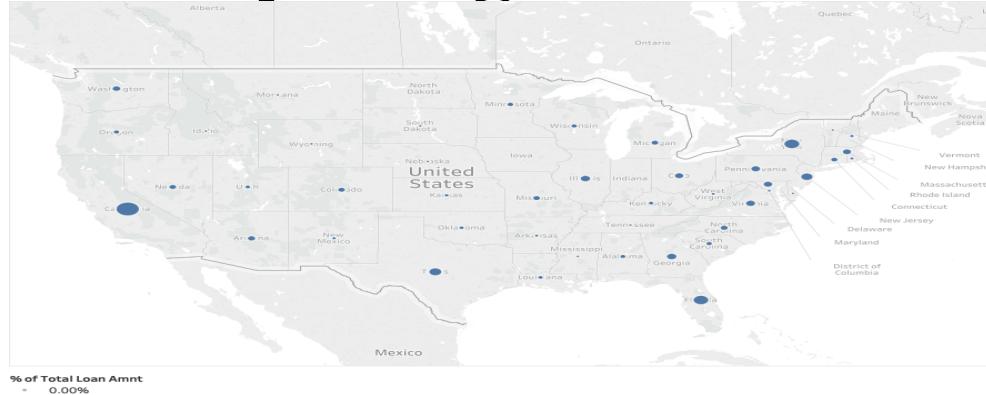
Key Insights – Better Verification of Income reduces risk



Insight – Loans which are not verified have a very high default rate

Suggestion - The Company must implement process to ensure all loans are issued post verification to minimize possibility of issue of loans to potentially defaulting borrowers

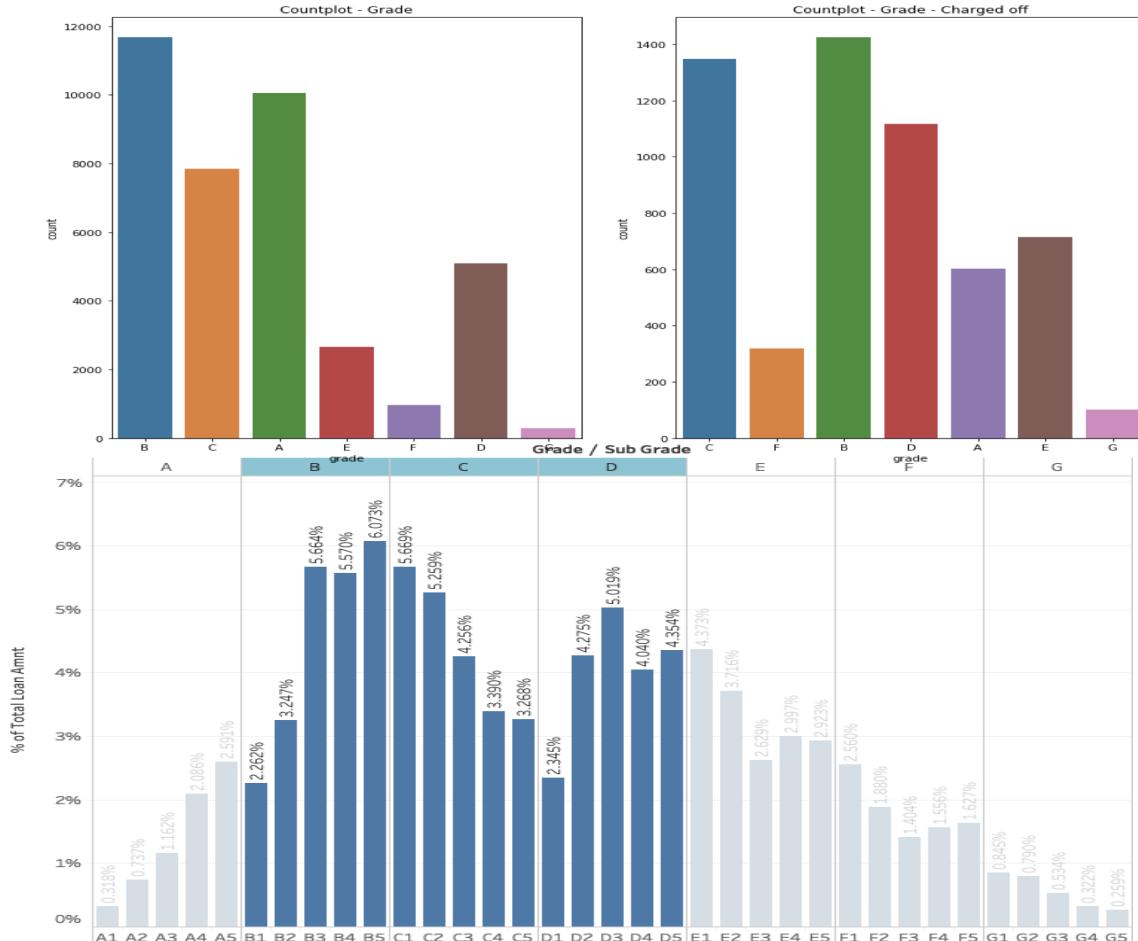
Key Insights – Location Risk



- Insight – We have observed maximum charged off loans have originated from CA, FL and NY

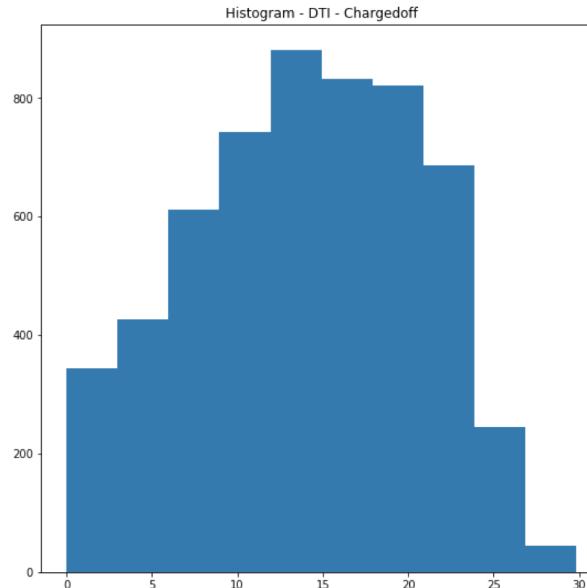
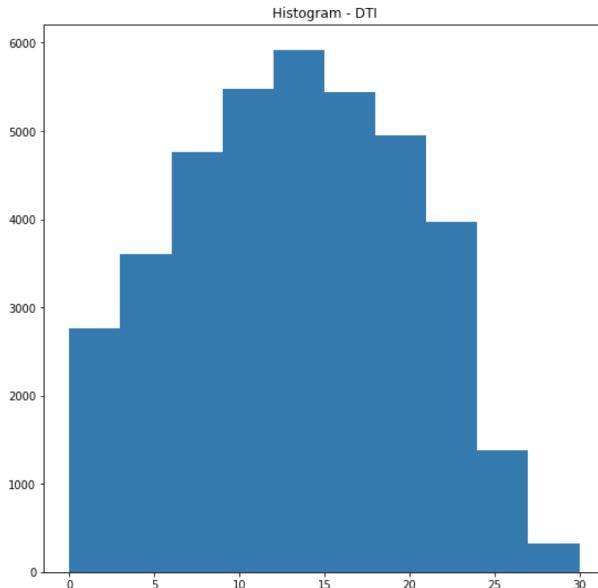
- Suggestion – Loans sanctioned to these residents of these states needs to be evaluated strongly

Key Insights – Grade Risk



- Insight – Credit Rating B, C and D have the highest charge out history. Further sub-rating B3, B4, B5, C1, C2, D3 and D5
- Recommendation – Loans with B, C and D ratings (with sub-rating mentioned above). Need to be issued at higher interest rate and rigorous process of verification

Key Insights – High DTI Risk



Insight – Borrowers with higher debt to income ratio have a higher tendency to default on loan

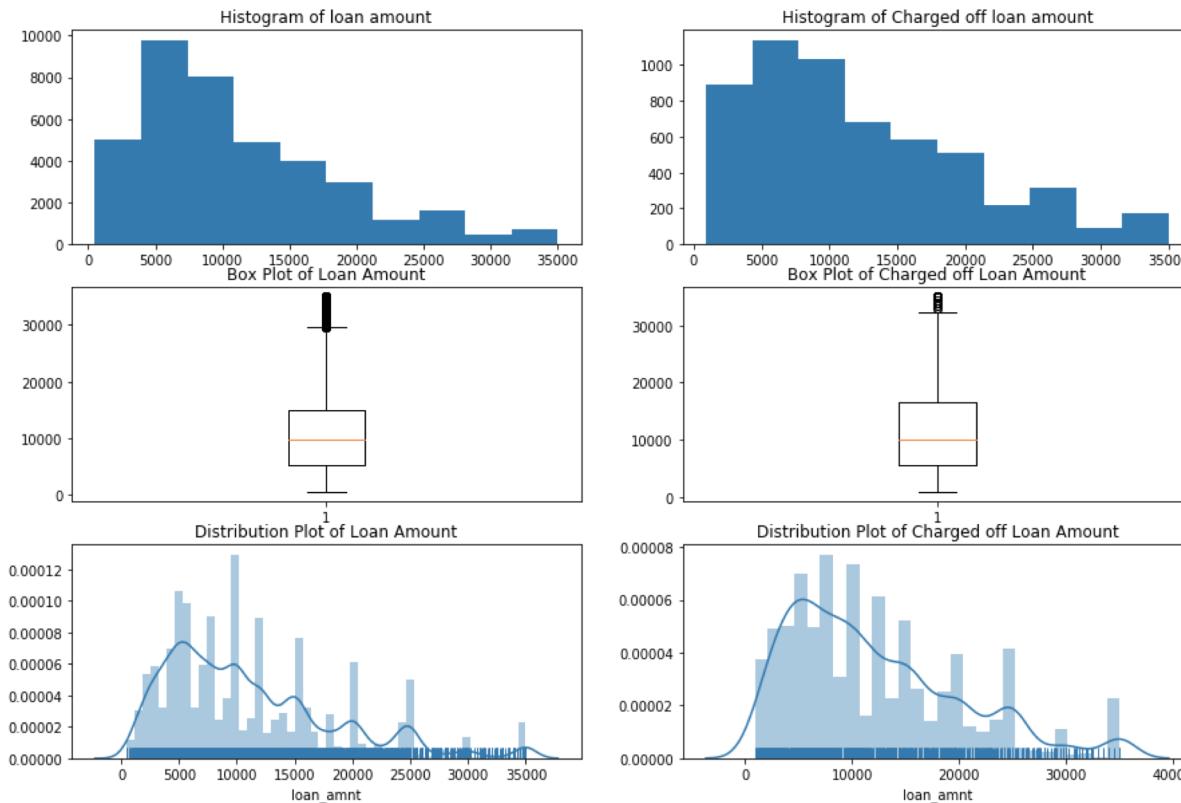
Recommendation – These loans should be backed by higher interest rate as risk of collection remains high,



UpGrad

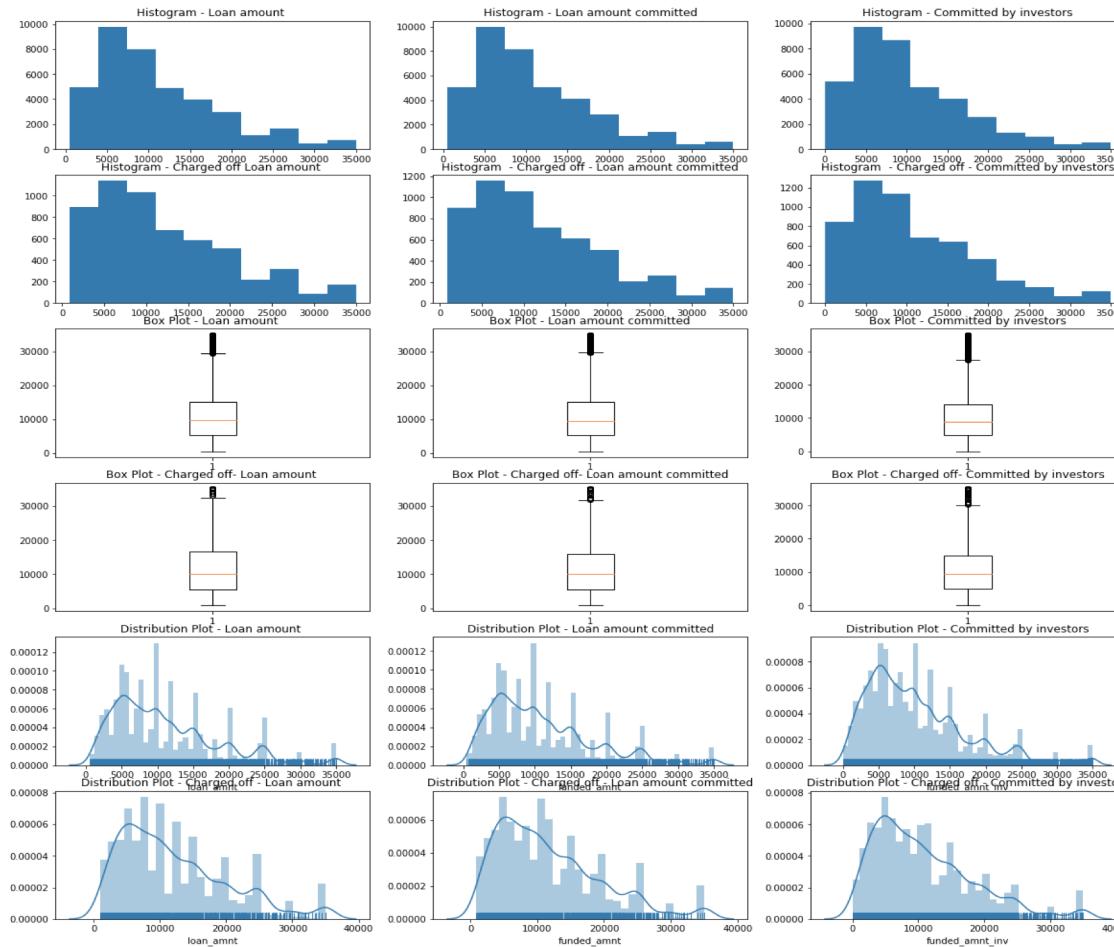
Annexure

Annexure 1 - Analysis of Loan Amount



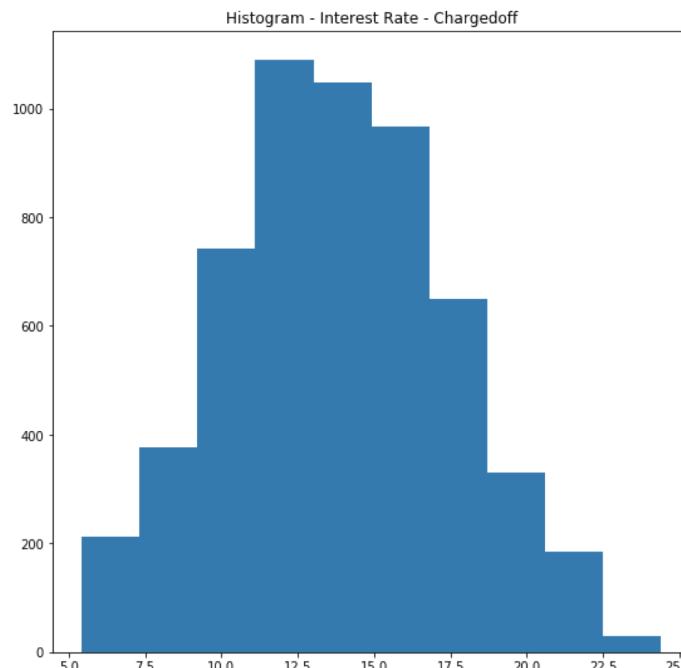
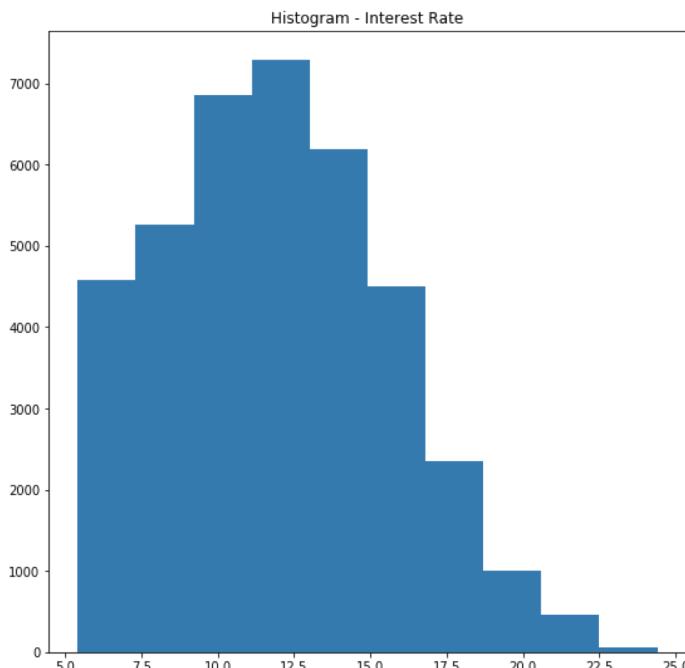
- The distribution of loan amount does not change due to status of loan

Annexure 2 - Analysis of Loan Amount



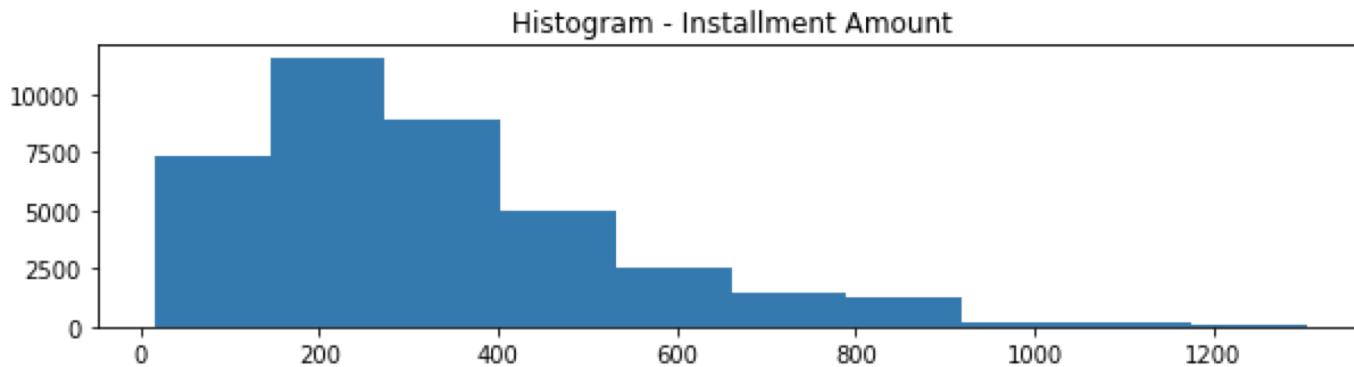
- It can be seen "funded_amnt" and "funded_amnt_inv" fields are very similar to
- given we need to analyse the business problem of bad loans, the committed loans may not be required
- There is not much difference between all loan status and only charged off

Annexure 4 – Interest Rate

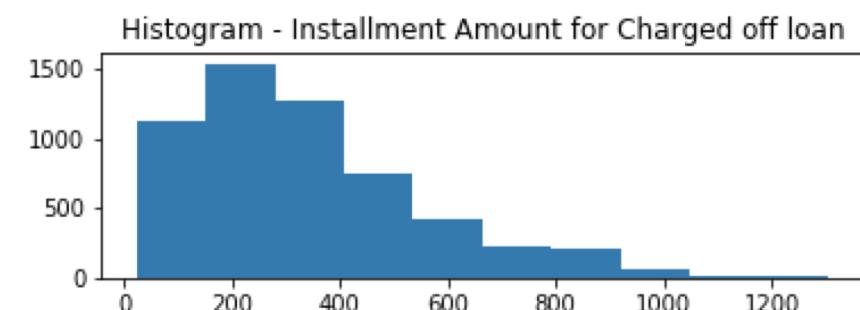


- The interest rates on loans are between 7.5% - 15%
- Charged off loans are higher at 11.5% - 17.5% interest

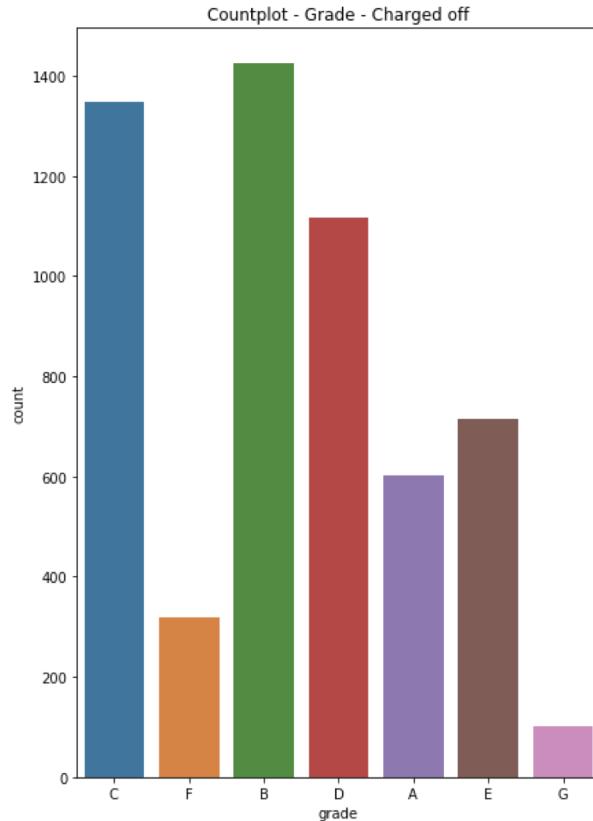
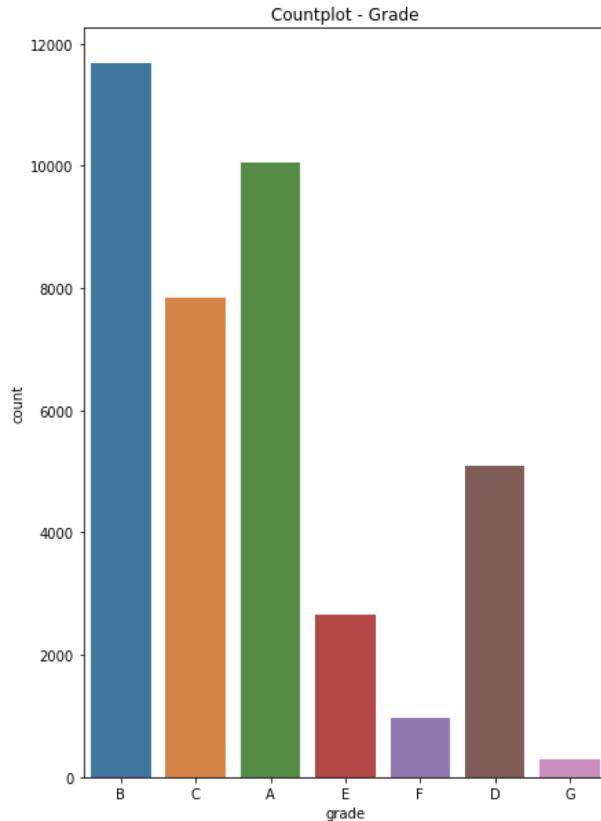
Annexure 5 – Installment Amount



- Installments value range around \$200
- For charged off loans installment values around \$100-\$400
- More charge off is around installment value of \$200

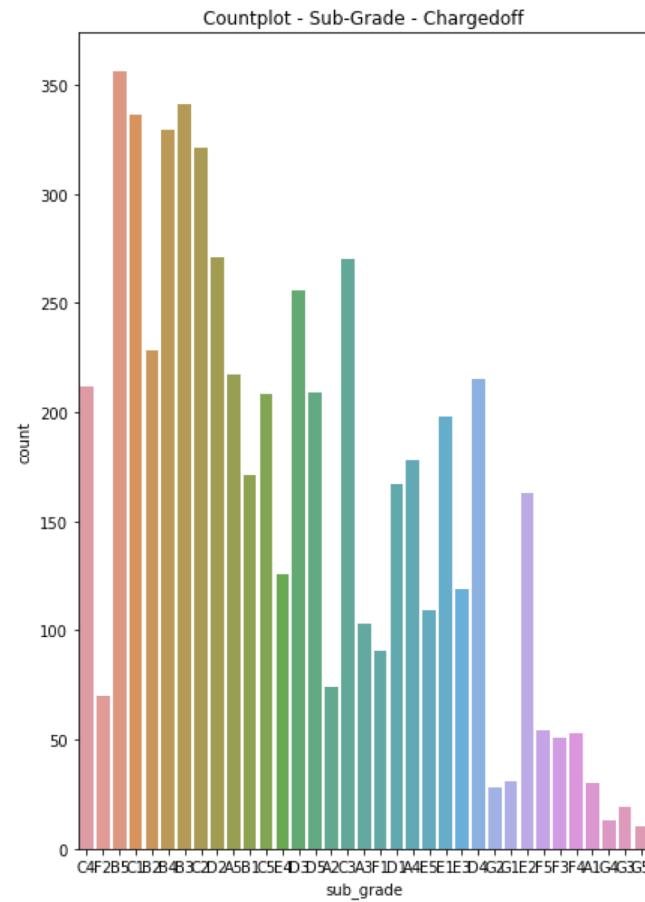
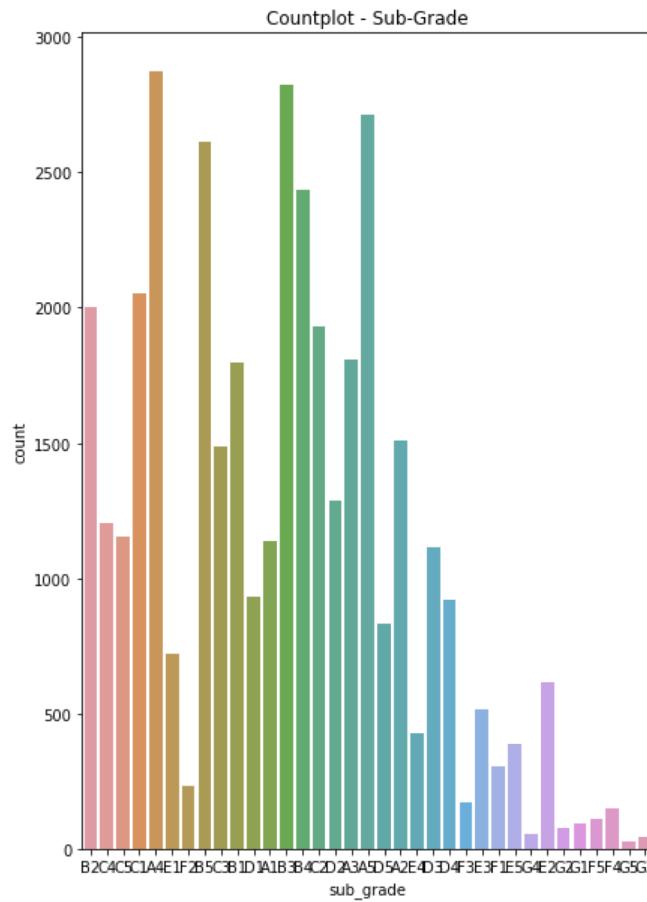


Annexure 6 – Grade



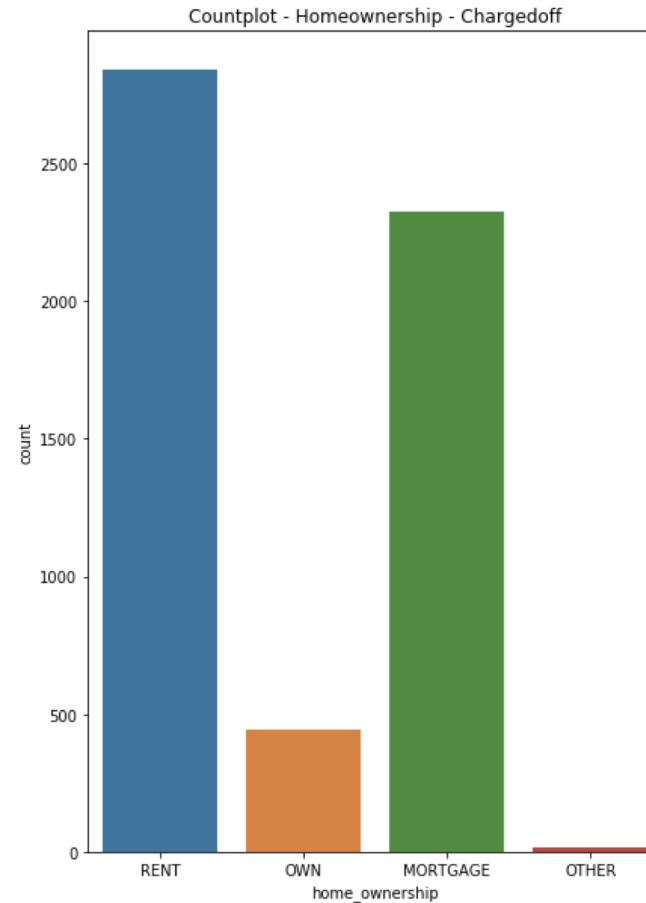
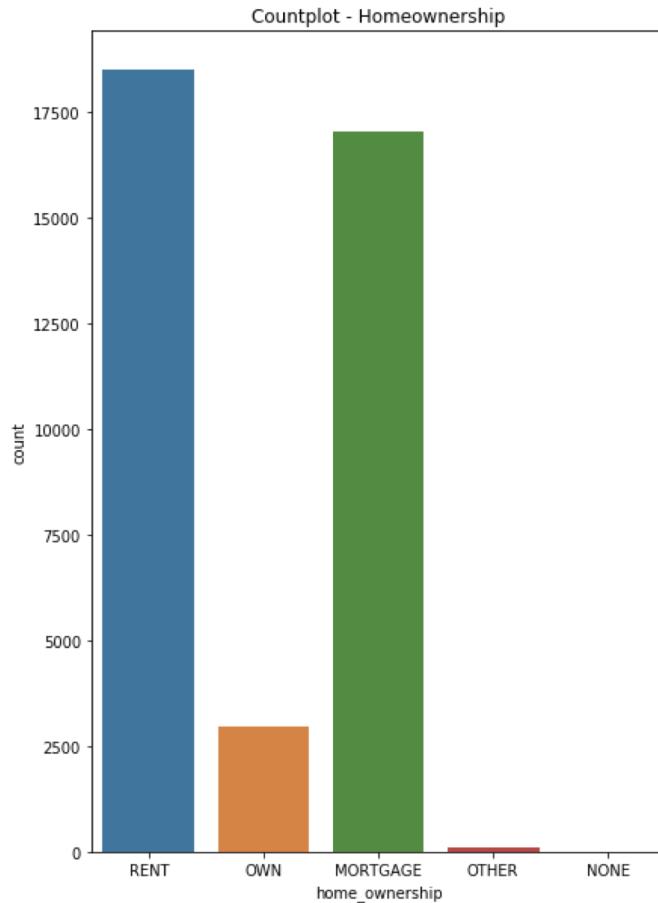
- Most Charge off Loans are B, C and D Category

Annexure 6 – Grade (Cont'd)



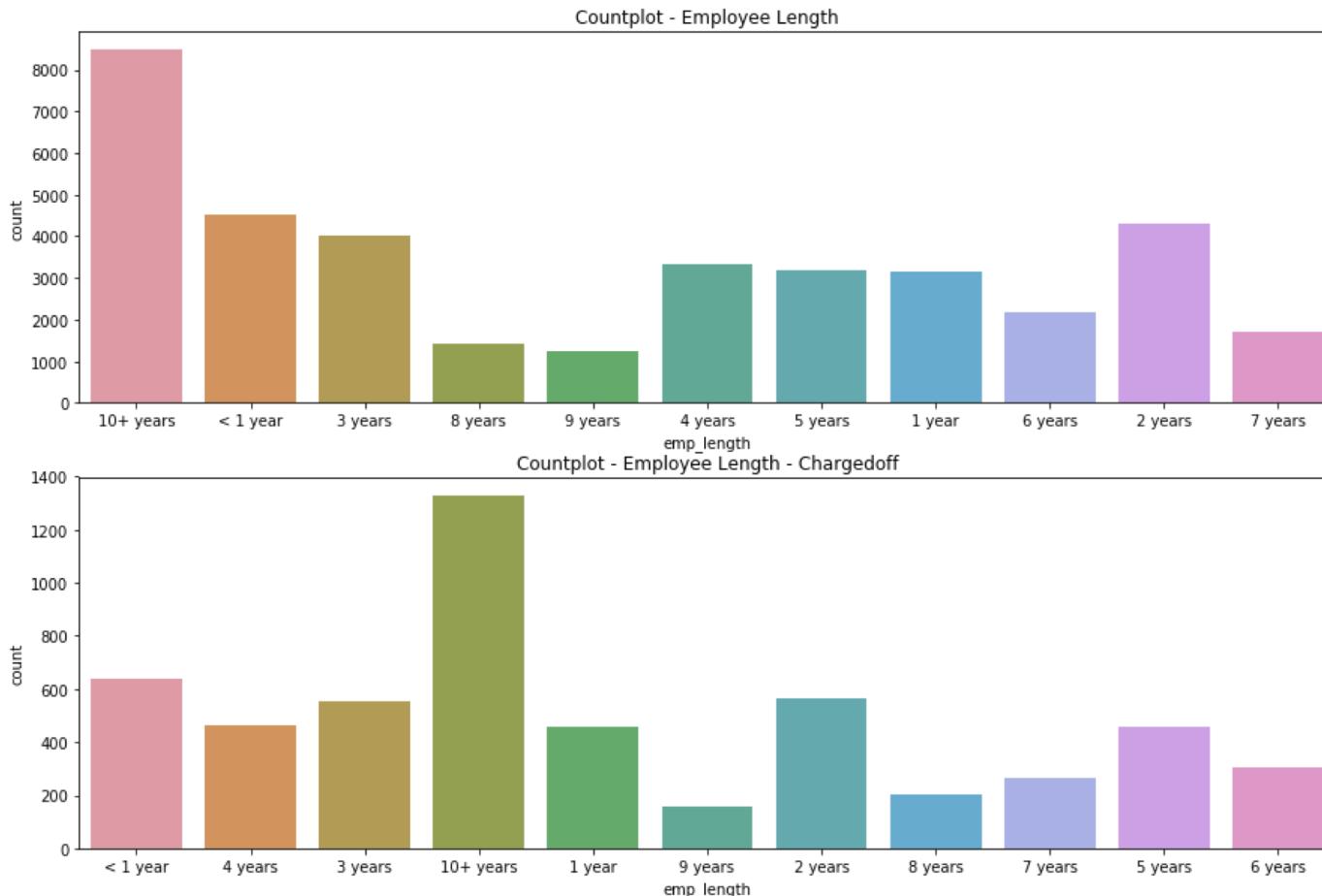
- Most charged off sub-category are:
 - Highest Charge off has occurred in following sub-category:
 - B3, B4 and B5
 - C1, C2

Annexure 7 – Home-ownership



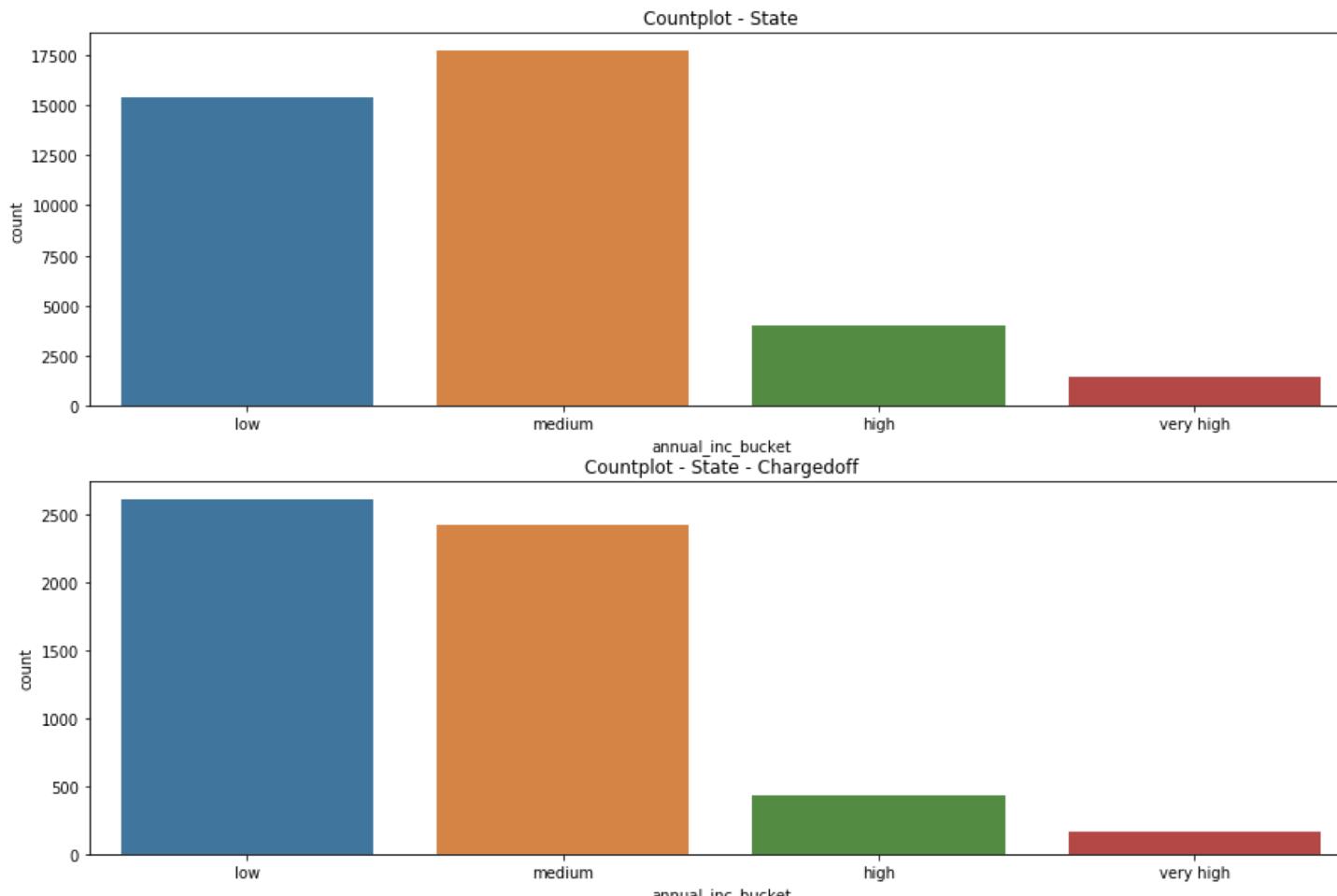
- Most charged off sub-categories are:
 - charge off is much higher in rental and followed by mortgage loans

Annexure 8 – Employment Length



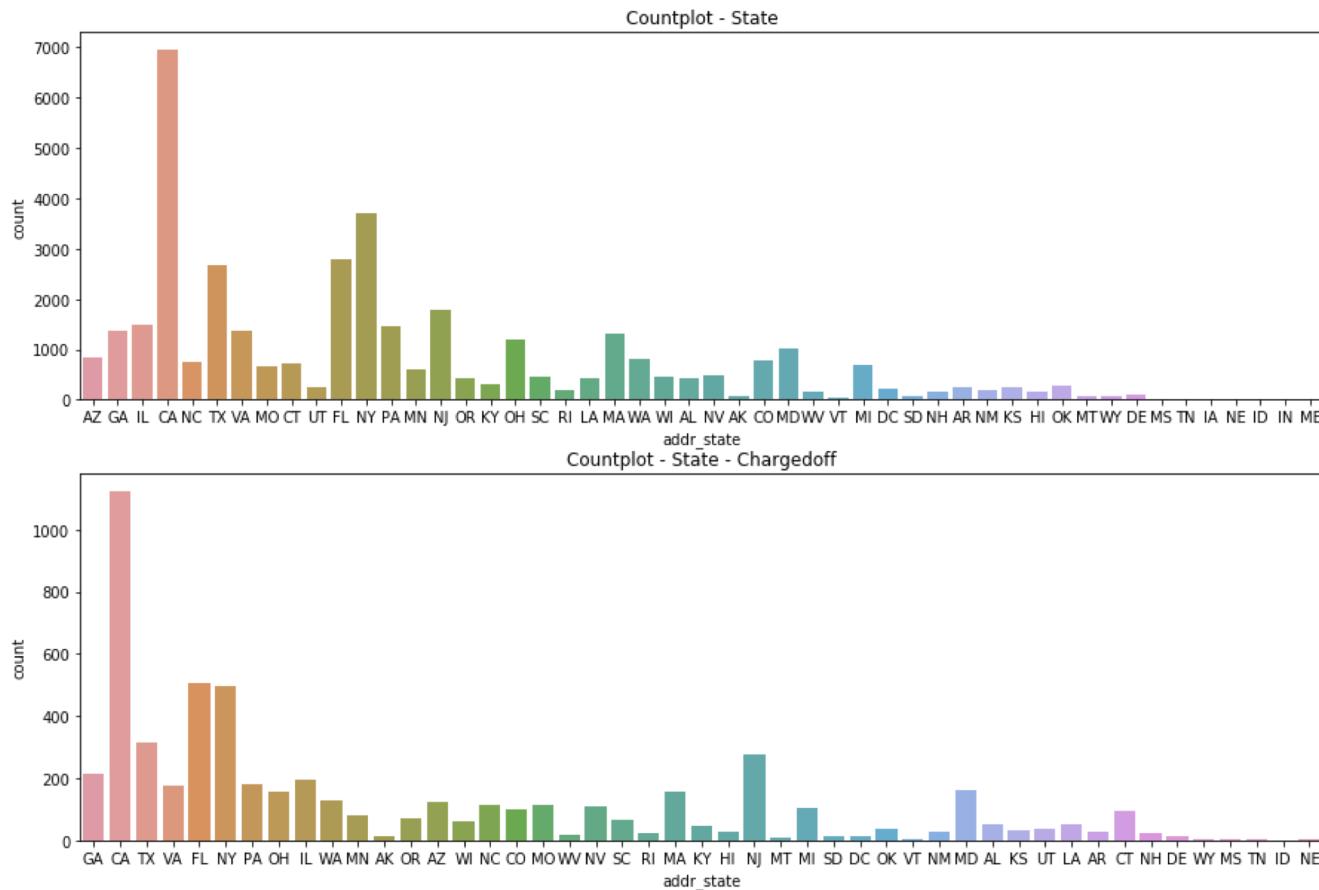
- The maximum loans are given to 10+ year category and also the charge off is maximum in this category

Annexure 9 – Annual Income



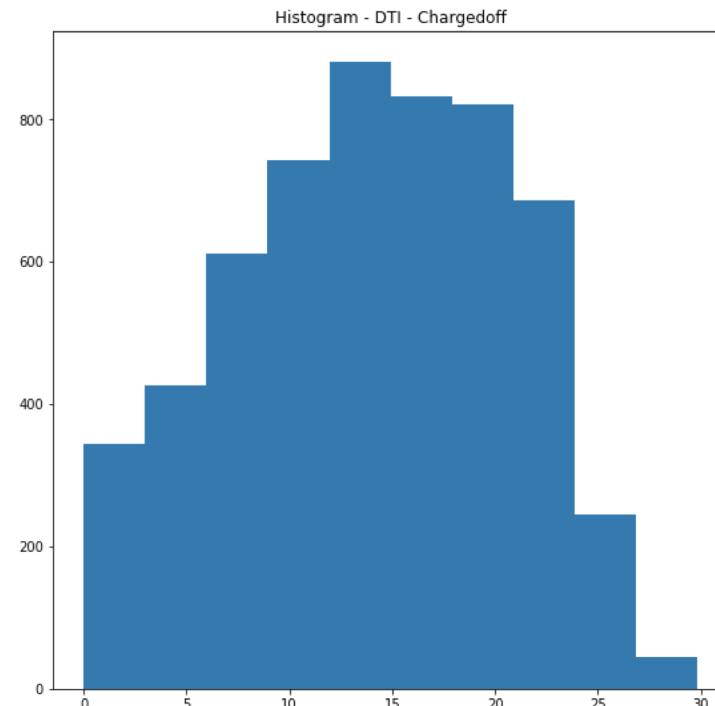
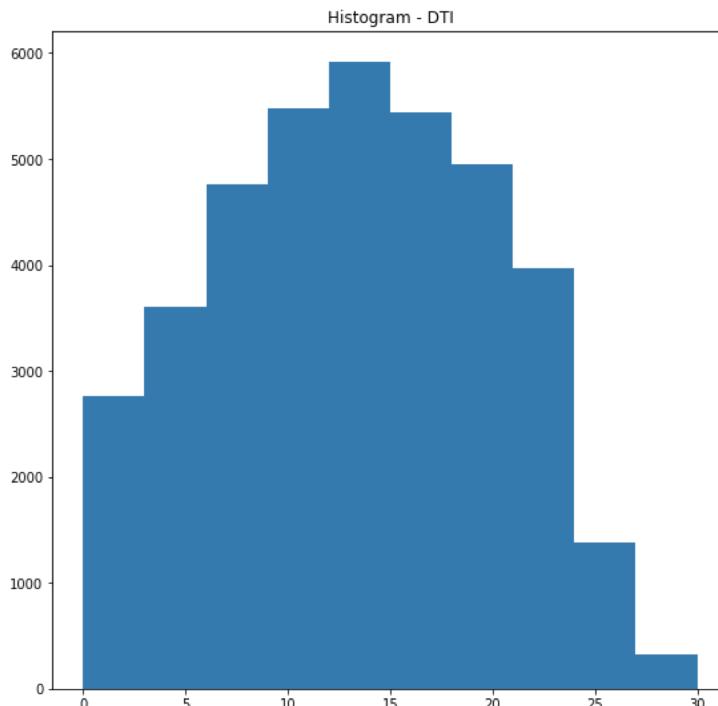
- Maximum charge off are from low income group and closely behind is medium income group
- % charge off from High and very high is very high as compared to loan given

Annexure 10 – State



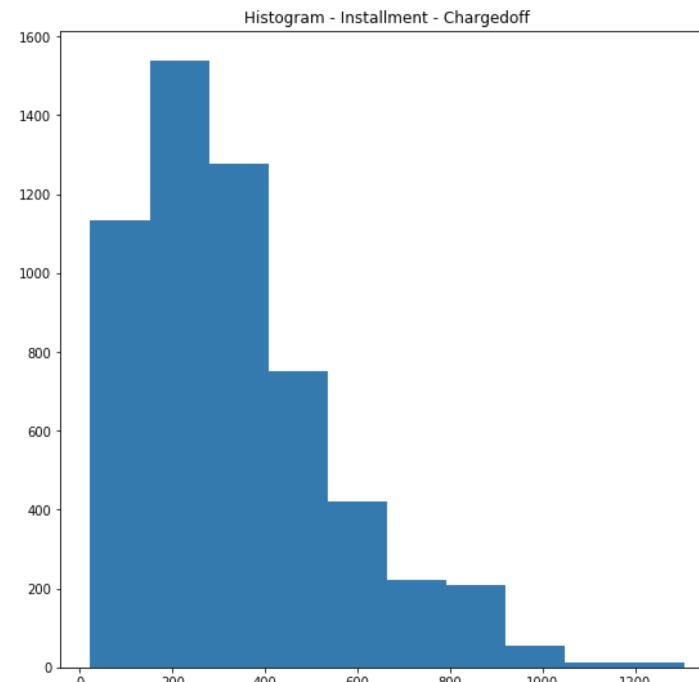
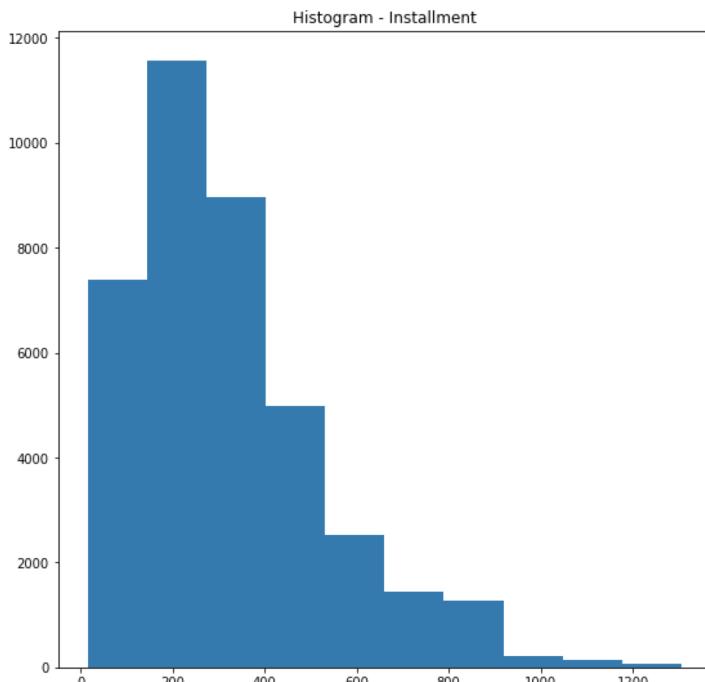
- maximum charge off are from CA, FL and NY

Annexure 11 – Debt Total Income



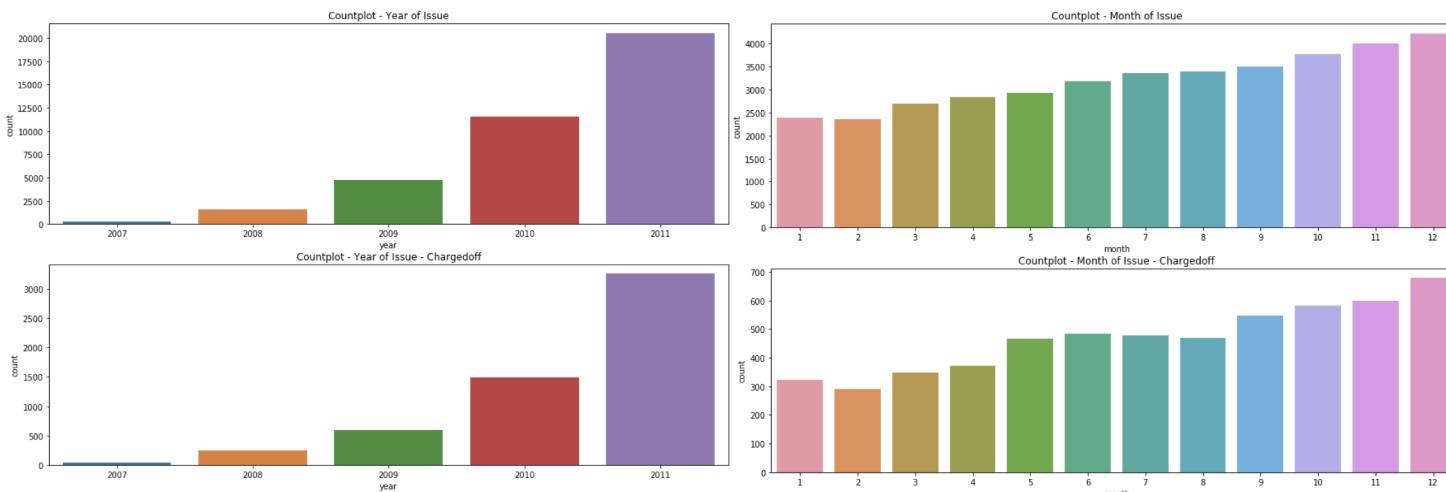
- Charge off is very high where DTI 12.5% - 22.5%

Annexure 12 – Installment Amount



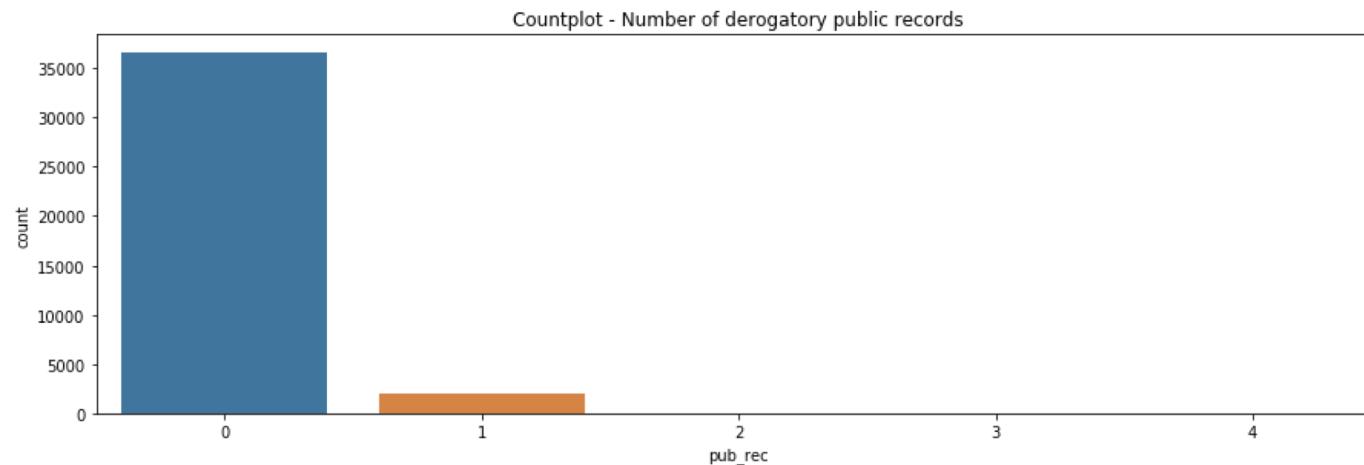
- Charge off is highest when installment is between \$200 - \$400
- This is consistent to loan fully paid

Annexure 13 – Issued Date

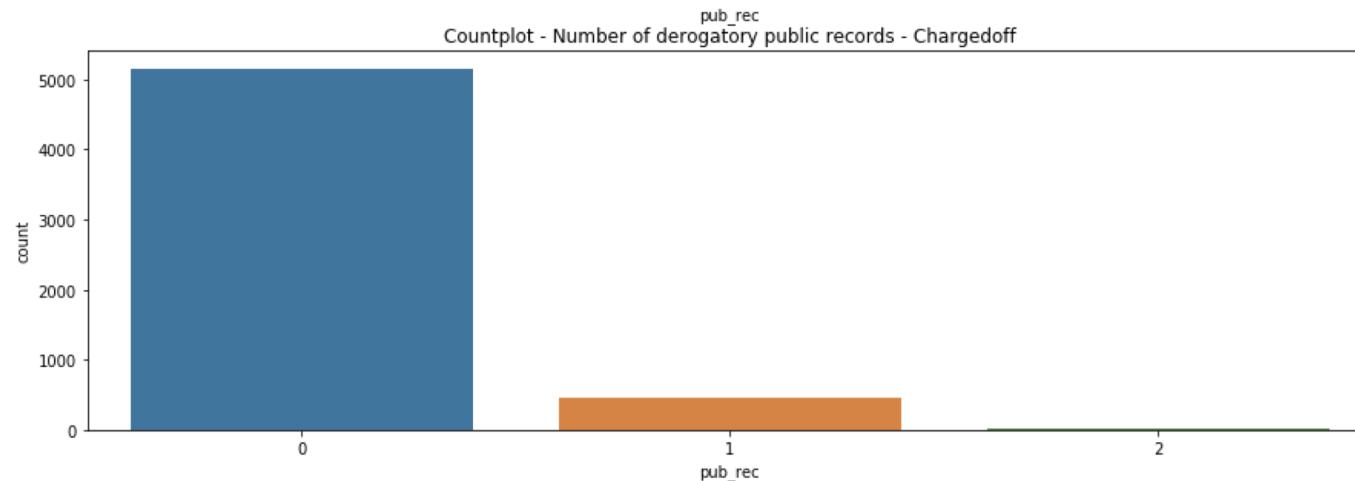


- The trend of loan issue and being charged off (for loan issued in respective months) is similar

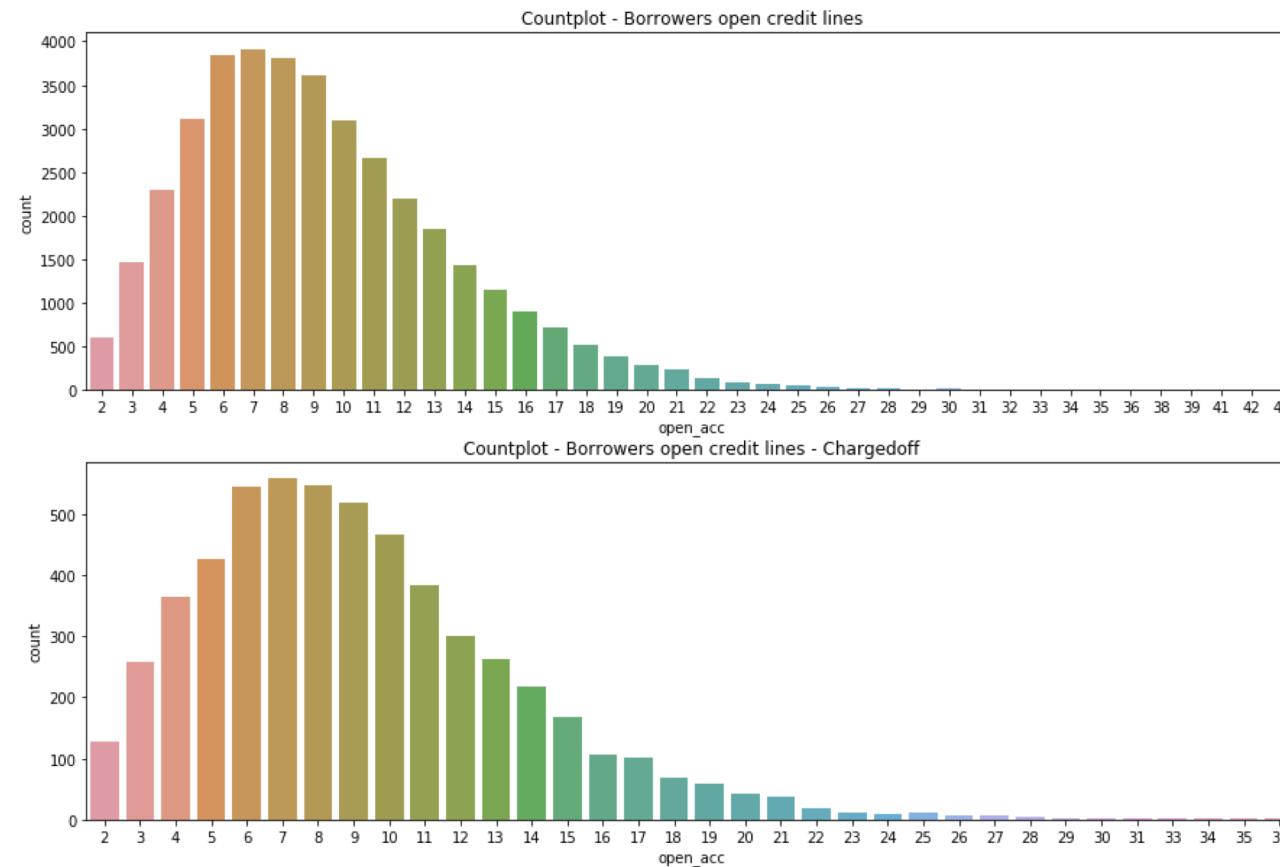
Annexure 14 – Derogatory public records



- Most Values are Nil

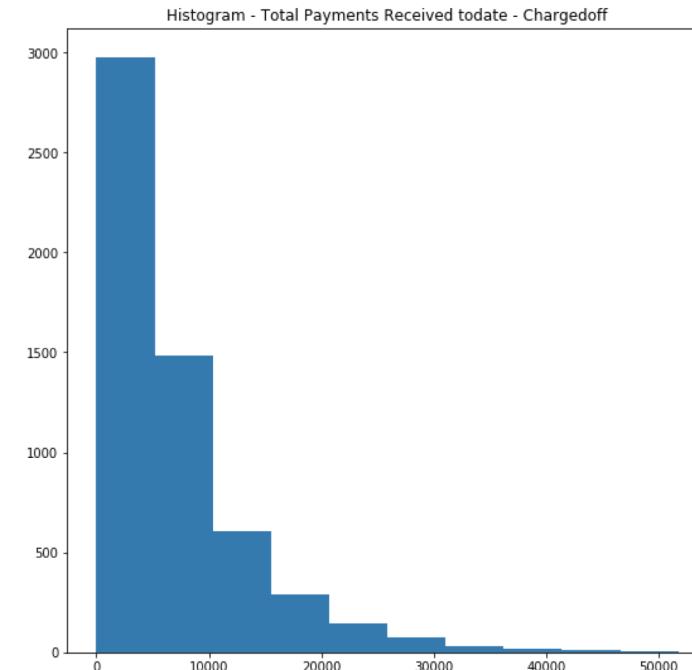
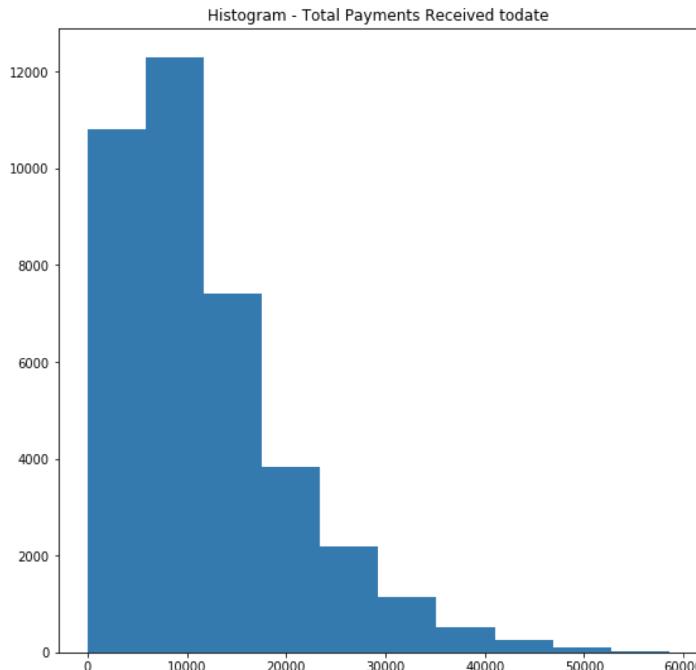


Annexure 15 – Borrowers Open Credit Lines



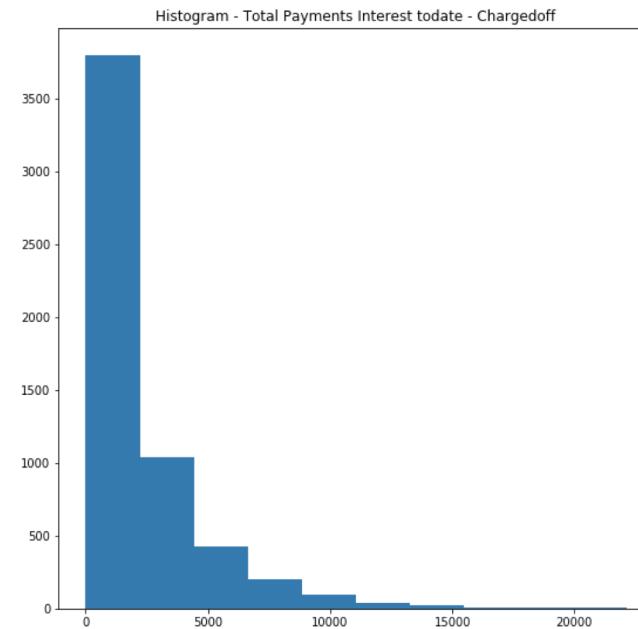
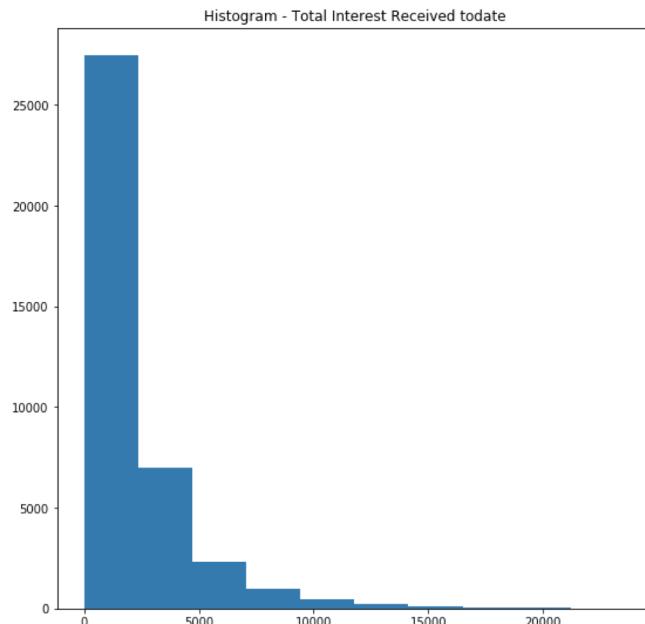
- Not much to differentiate between Charged off loans and total loans

Annexure 16 – Total Payments Received



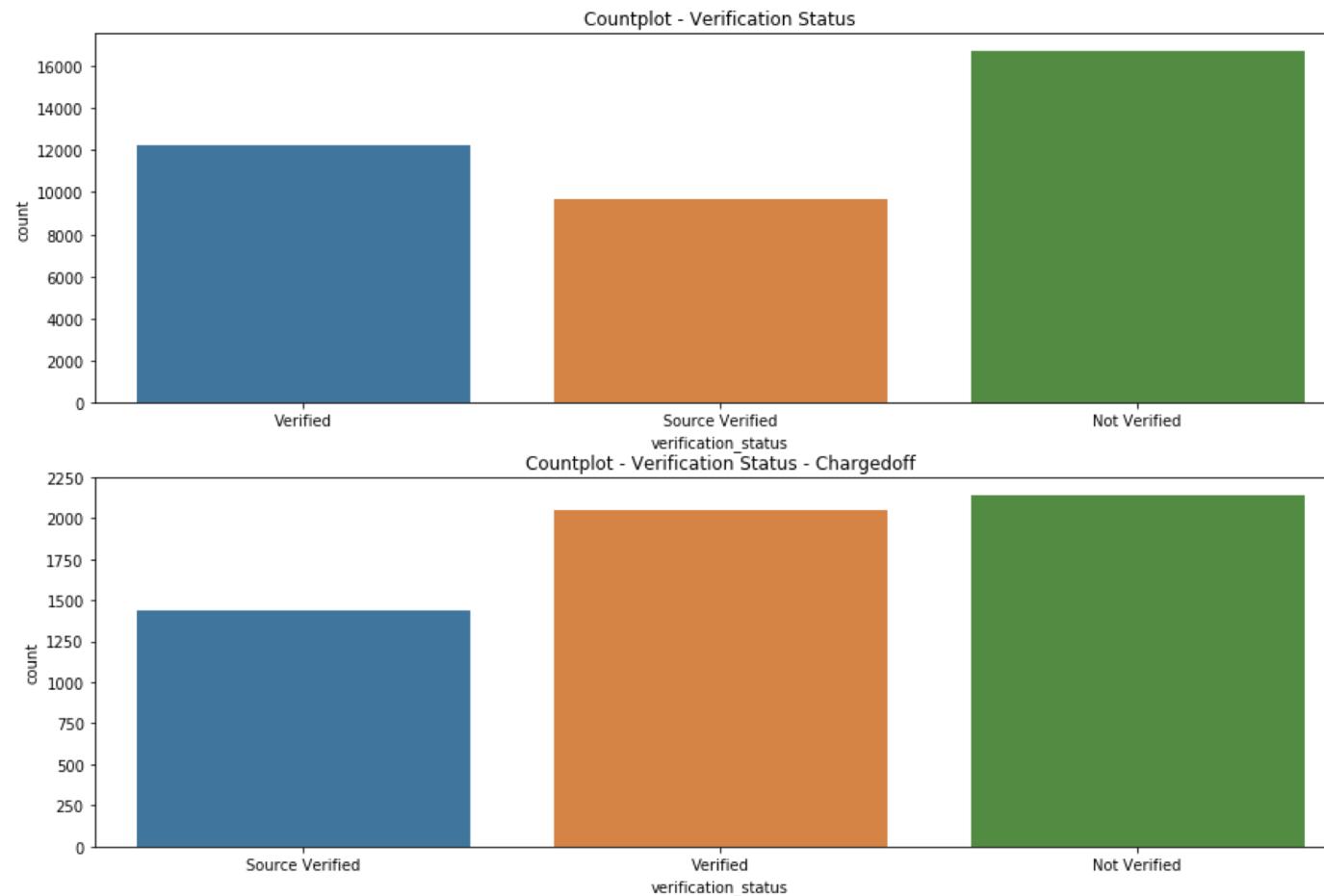
- maximum chargeoff is where total principle payments received is upto date less than \$5000
- Once the principle payment exceeds \$5000 the chances of charge off are low

Annexure 17 – Total Interest Received



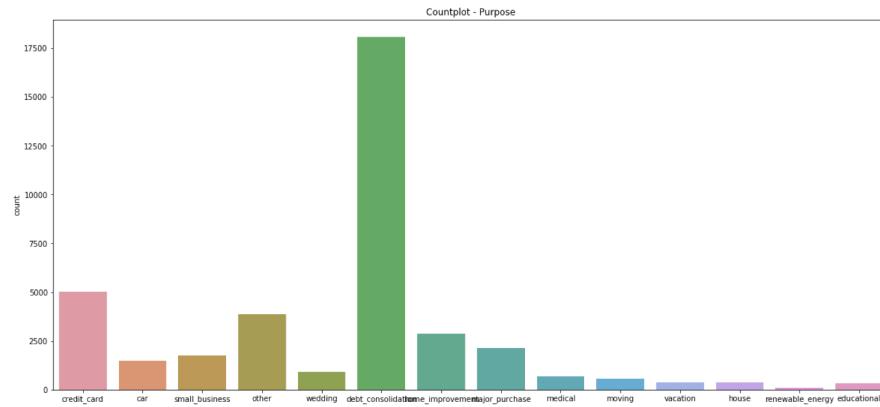
- There is not much difference between all loans given and charged off

Annexure 18 – Total Interest Received

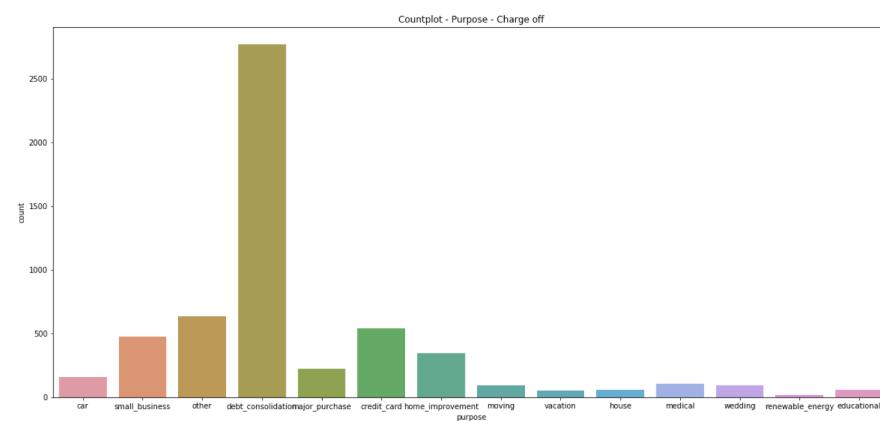


- Charge off is low where source is verified and high where not verified

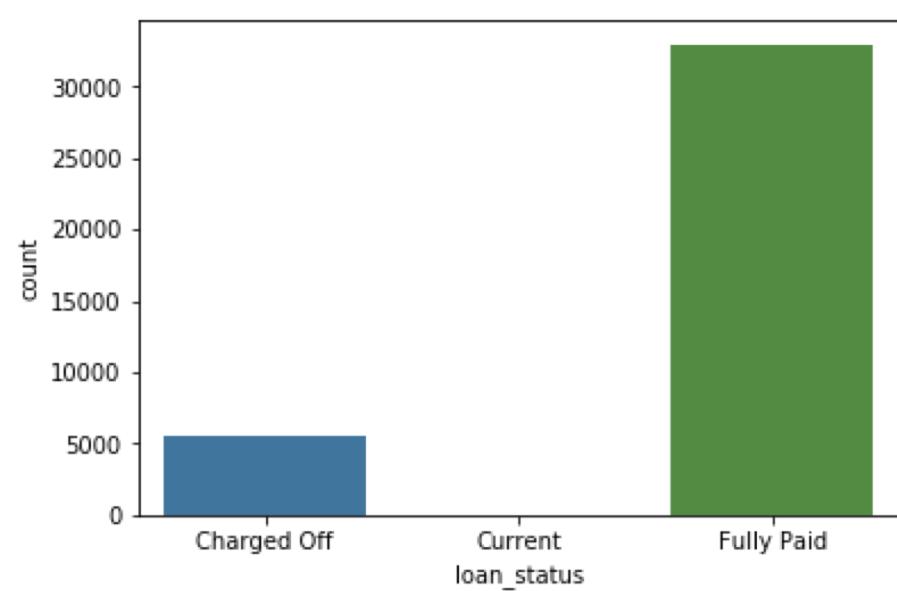
Annexure 19– Purpose



- Debt Consolidation, Credit Cards and Others are main loans being charged off



Annexure – 20 Loan Status



- Charged off loans are about 15% of total loans