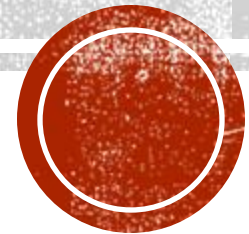# LENDING CLUB CASE STUDY

Group Members

Chethana Manyam

Jenifer Sam

Pallavi Nigam

TABLE OF CONTENT

- Problem Statement

- Data Understanding

- Data Cleaning and Manipulation

- Data Analysis

- Recommendations

PROBLEM STATEMENT: IDENTIFY THE DEFAULTERS / RISK ANALYSIS

Description

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

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
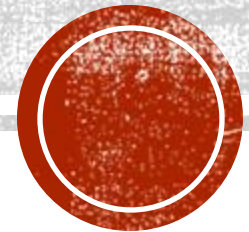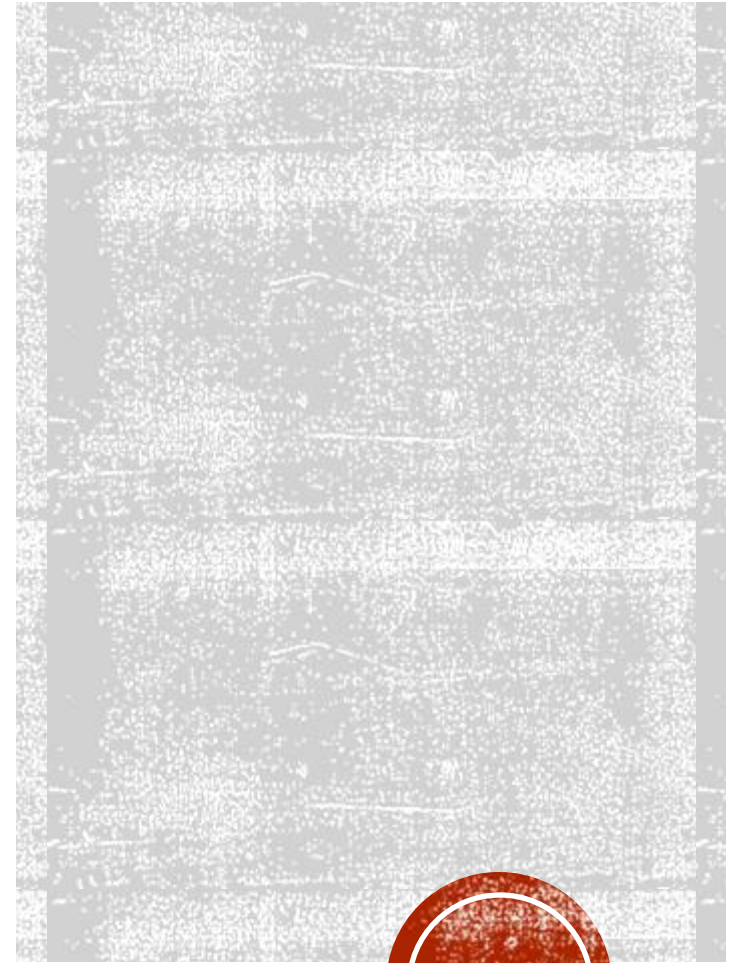
Goal

Identify risky loan applicants. Such loans can be reduced thereby cutting down the amount of credit loss.

https://www.lendingclub.com/

DATA UNDERSTANDING

DATA UNDERSTANDING

Metadata:

Data Source   :Loan.CSV

Data Dictionary :Data_Dictionary.xls (Sheet – LoanStats)

Data Size   :39717 rows and 111 columns of data

Loan Amount related Columns:

loan_amnt : The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.

**funded_amnt:**The total amount committed to that loan at that point in time.

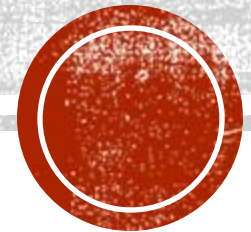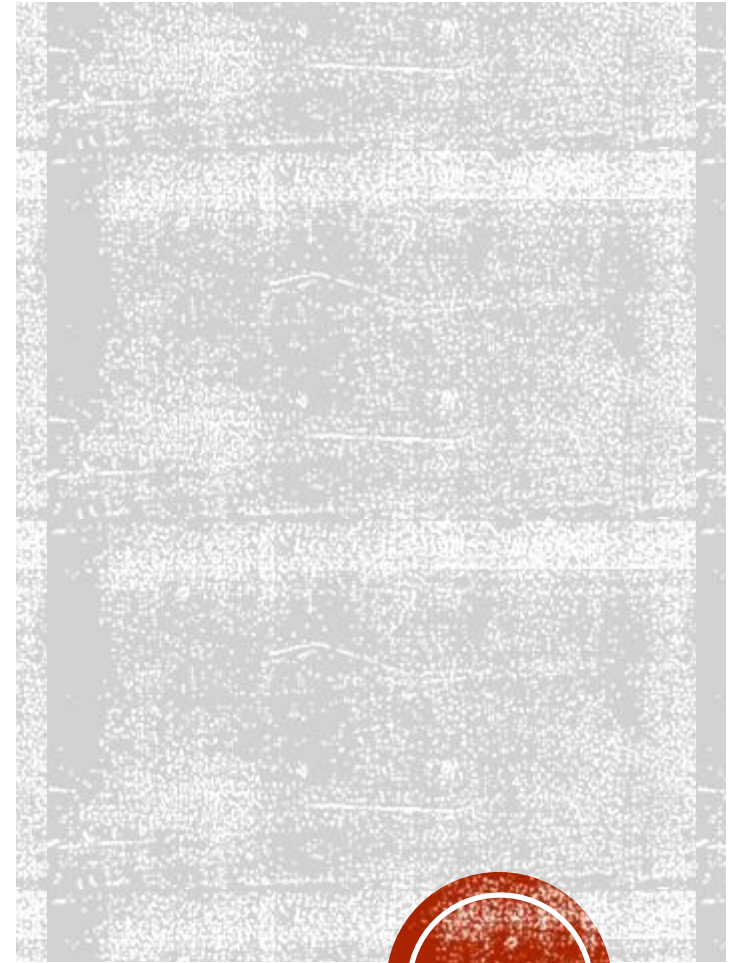**funded_amnt_inv:**The total amount committed by investors for that loan at that point in time.

DATA UNDERSTANDING

Loan_Status(Loan Accepted Data Only):

- **Fully Paid:** Applicant has fully paid the loan

- **Charged-off:** Applicant not paid the Installments in due time (failed to make sufficient payments for an extended period)

- **Current:** Applicant in process of loan payment/Installments(Can ignore this category as we cant predict whether applicant can pay or not pay the loan)

LOAN DATASET

Loan Accepted → Default
            → Non-Default

Loan Rejected
(Not considered in dataset)
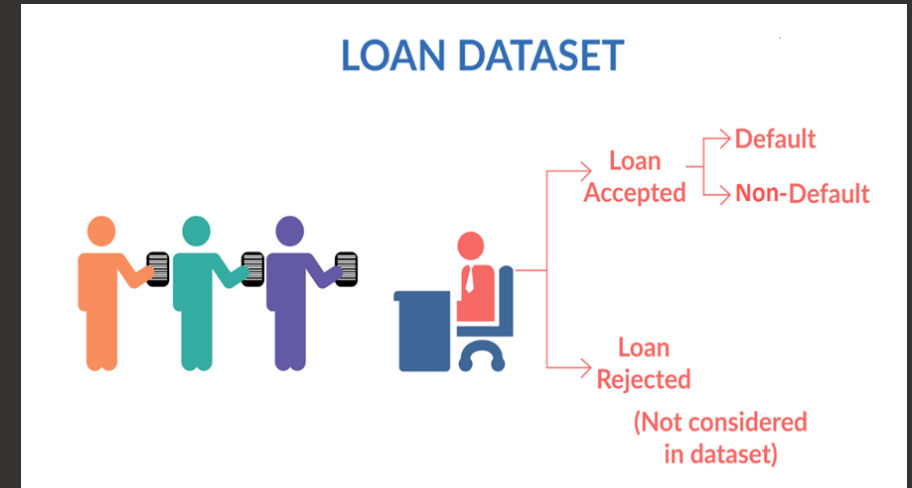
Data Cleaning and Manipulation

DATA CLEANING(COLUMN-WISE)

A. Removing customer variables

Customer variables are those variables that are generated after a loan is approved. These metrics will not be used to determine the outcome of this case study. Therefore, these columns can be dropped.

loan = loan.drop(cust_var_col, axis=1)

Data Size = 39717 rows and 89 columns

*loan – Data frame

- application_type
- collection_recovery_fee
- delinq_2yrs
- earliest_cr_line
- inq_last_6mths
- last_credit_pull_d
- last_pymnt_amnt
- last_pymnt_d
- open_acc
- out_prncp
- out_prncp_inv

- pub_rec
- recoveries
- revol_bal
- revol_util
- total_acc
- total_pymnt
- total_pymnt_inv
- total_rec_int
- total_rec_late_fee
- total_rec_prncp
- url

DATA CLEANING(COLUMN-WISE)

B. Removing columns where all values are null

loan = loan.dropna(axis=1, how=all)

Data Size = 39717 rows and 35 columns

C. Using missingno library, identify columns with more than 50% missing values

Drop columns : mths_since_last_delinq, mths_since_last_record , next_pymnt_d

Data Size = 39717 rows and 32 columns

D. Remove columns with just one value - These wont impact the analysis

Drop Columns : pymnt_plan, initial_list_status, collections_12_mths_ex_med, policy_code, acc_now_delinq, chargeoff_within_12_mths, delinq_amnt, tax_liens

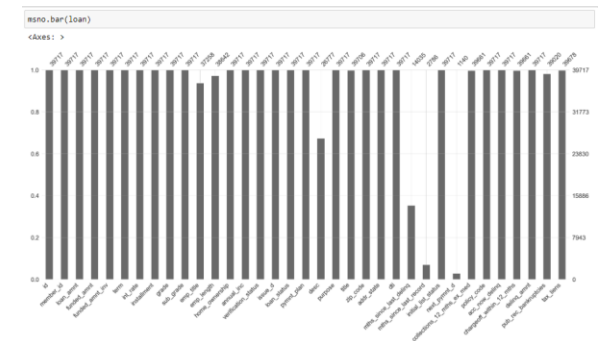Data Size = 39717 rows and 24 columns

E. Remove free text columns

Drop Columns : desc,title

Data Size = 39717 rows and 22 columns

F. Remove columns with all unique values

Drop Columns : id, member_id

Data Size = 39717 rows and 22 columns

DATA CLEANING(ROW-WISE)

A. Detect Missing Values - Propose imputation or cleanup

Very small percentage of missing data – no drop

B. Loan Status: Filter Out Values Needed For Analysis

loan = loan[loan.loan_status != 'Current']

Data Size = 35482 rows and 20 columns

DATA MANIPULATION

A. Detect Missing Values - Propose imputation or cleanup

Very small percentage of missing data – no drop

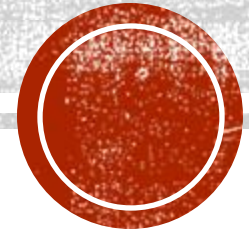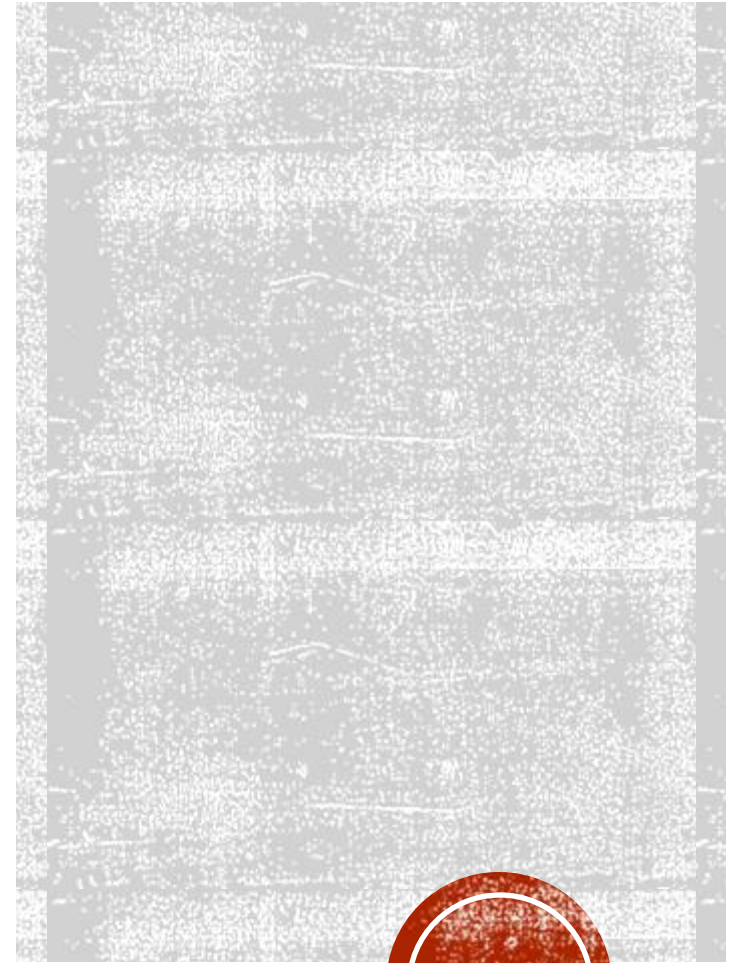**B. Loan Status: Filter Out Values Needed For Analysis**

loan = loan[loan.loan_status != 'Current']

Data Size = 35482 rows and 20 columns

C. Check the correlation and identify the columns(numeric values) which are highly correlated and ignore, if possible, for the further analysis

Correlation between different numeric values
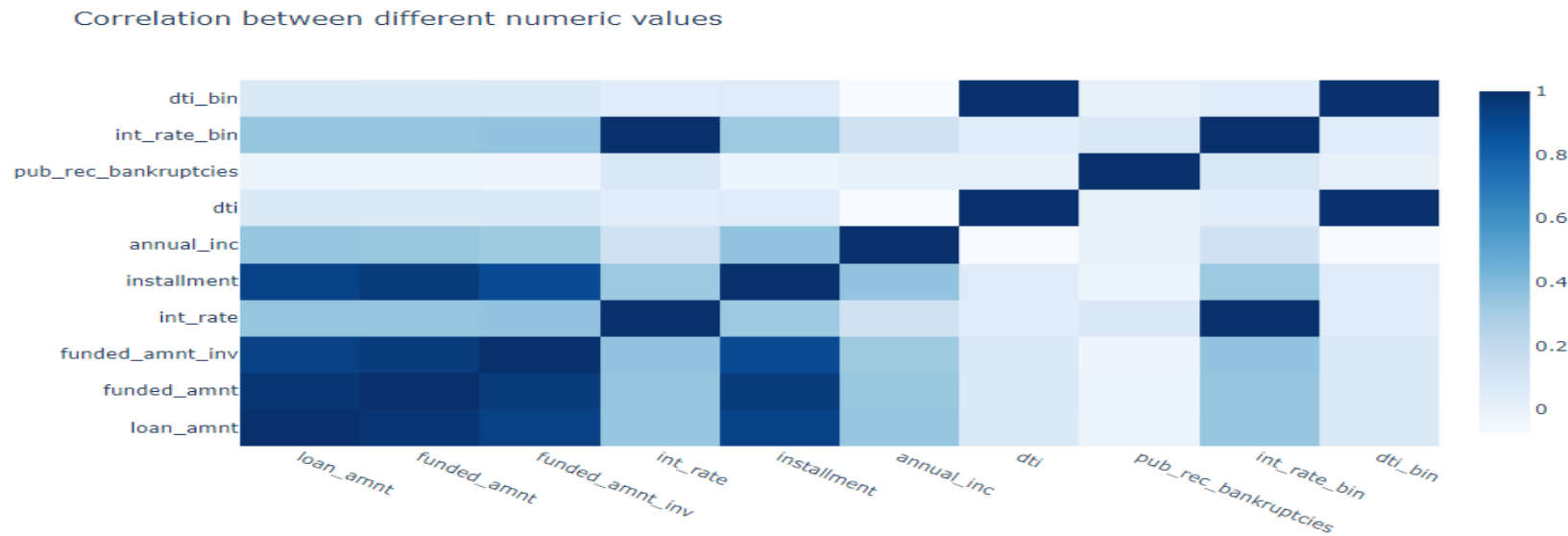
DATA ANALYSIS

DATA UNDERSTANDING

- **loan_amnt:**

  - Loan which has been Requested by the borrower from bank as per past records

  - This is continuous integer value

- **funded_amnt:**

  - Funded amount is the amount which was recommended by lending club with respect to loan applied

  - Funded amount is continuous integer value

  - Funded amount should be less than or equal to loan amount

- **funded_amnt_inv**

  - Loan amount funded by the investor

  - It is float value and can take any float in domain

- **term**

  - Duration in which loan will be payed

  - This can also be seen as instanllments in which loan is payed including loan amount and instrest rate

  - In the lending case study terms has two values

    - 36 Months

    - 60 Months

DATA UNDERSTANDING

- **int_rate**

  - Interest rate on the loan [having value in percentage]

  - Borrower will have to pay actual loan + interest on top of the loan in order to pay off complete loan

  - Some of the factor which could influence interest rate could be economic strength, inflation, government policy, supply and demand, credit risk, and loan period

- **installment**

  - Fixed amount of money that must be repaid with regularly scheduled payments by borrowers and is float value

- **grade**

  - LC assigned loan grade

  - Loan grades are set based on both the borrower's credit profile and the nature of the contract. 'A' grade loans represent   the lowest risk while 'G' grade loans

  - In current case study grade can have values [A,B.C,D, E,F,G]

- **sub_grade**

  - LC assigned loan subgrade

  - With in each loan grade also subgrade assigned to get better risk factors attached with the loans

  - For example, within grade A we have [A1, A2, A3, A4, A5] to better categorize the loan

- **emp_title**

  - The job title supplied by the Borrower when applying for the loan.*

  - Can take any value in String domain

DATA UNDERSTANDING

- **emp_length**

  - Employment length in years.

  - No of Year Employee has done the job

  - Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.

- **home_ownership**

  - The home ownership status provided by the borrower during registration.

  - Our values are RENT, OWN, MORTGAGE, OTHER.

- **annual_inc**

  - Annual Income provided by the borrower during registration.

  - The self-reported by borrower

  - Can take any value in float domain

- **verification_status**

  - Indicates that for the loan if borrowers' income was verified by lending club

  - It can take below three values with corresponding details

    - Verified [Income claimed by borrower is verified by lending club]

    - Source Verified[Income Source is verified]

    - Not verified [Neither income, not income source is verified]

DATA UNDERSTANDING

- **issued**

  - The month which the loan was funded

- **loan_status**

  - Indicates current payoff status for loan

  - Can take below their values

    - Fully Paid - Loan is fully paid by borrower

    - Current - Loan is being paid off

    - Charged off - Borrower could not pay the loan

- **purpose**

  - Indicates the purpose for which laon was applied

  - Below are some of the categories

    - Debt consolidation refers to taking out a new loan or credit card to pay off other existing loans

    - Small Business - Refers to loan take in order to start business etc

- **zip_code** :The first 3 numbers of the zip code provided by the borrower in the loan application.

- **addr_state**:The state provided by the borrower in the loan application

- **dti:**A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.

- **pub_rec_bankruptcies**: Number of public record bankruptcies

DATA ANALYSIS

- loan_status – Dependent Variable

Fully Paid   85.8548

Charged Off   14.1452

Insight:

- 14.14% of loans are being charged off at Lending Club.

- Our goal is to ensure this overall percentage is being reduced.

- Therefore, in our exploratory data analysis, we will focus on metrics that are influencing the percentage of charged off loans.

- That is, technically, every categorical and numeric variable will be viewed as a percentage of charged off loans.

In our analysis, 14.14% is being considered as the threshold amount. Any metric that affects the loan status (charged-off loan percentage) more than this threshold value needs to be considered with caution.

Loan Status Overview in Percentage
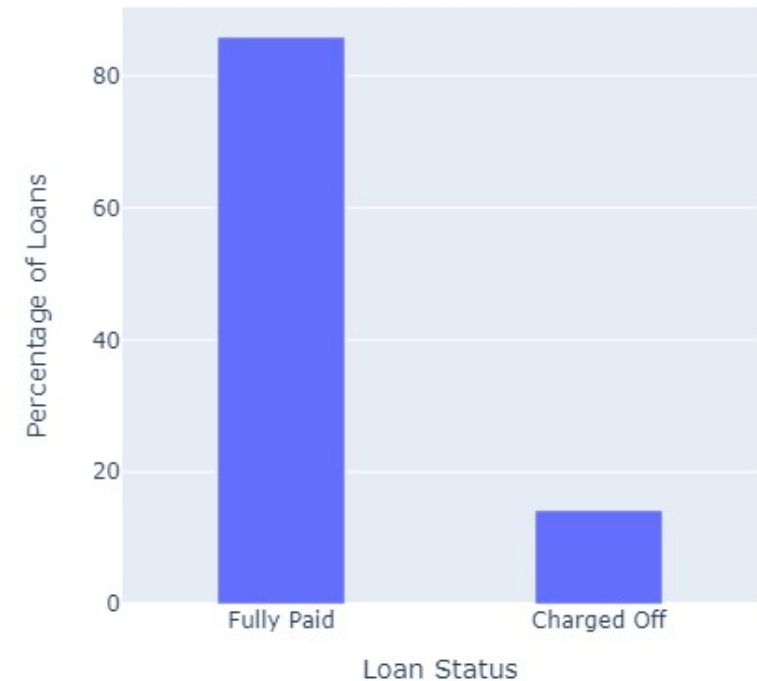
DATA ANALYSIS

Loan Amount

- loan_amnt: Amount requested by the borrower

- funded_amnt: Amount approved/recommended by Lending Club

- funded_amnt_inv: Amount invested by the investor

Basic Data Cleansing

- loan_amnt should be greater than or equal to funded_amnt

- funded_amnt should be greater than or equal to funded_amnt_inv

Insight:

- Maximum loans have been given between 5k-15k.

Distribution of Loan Amount
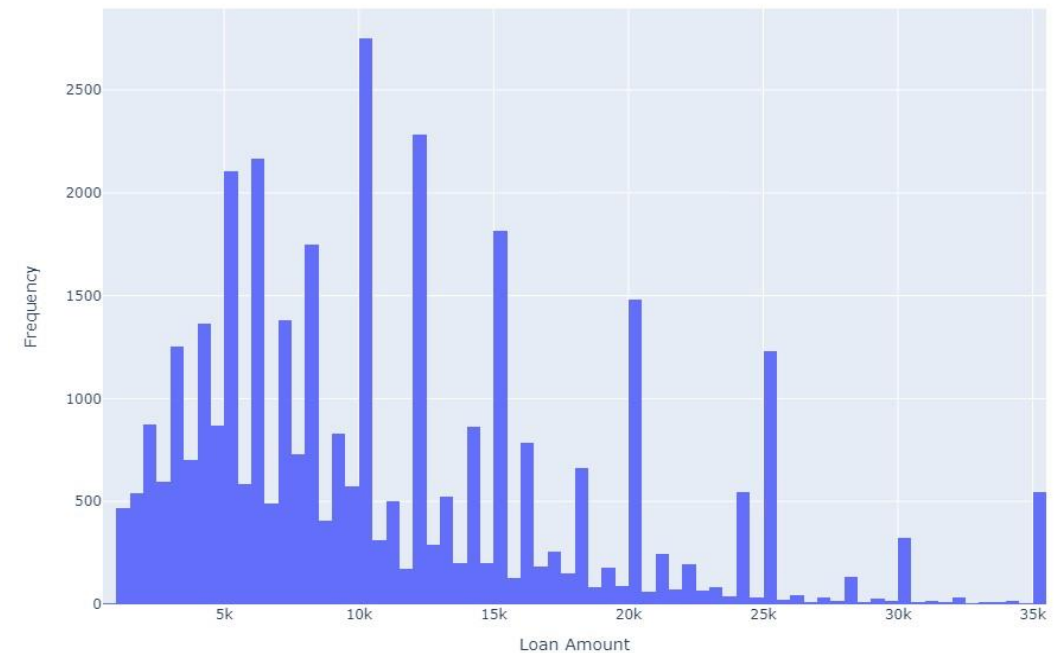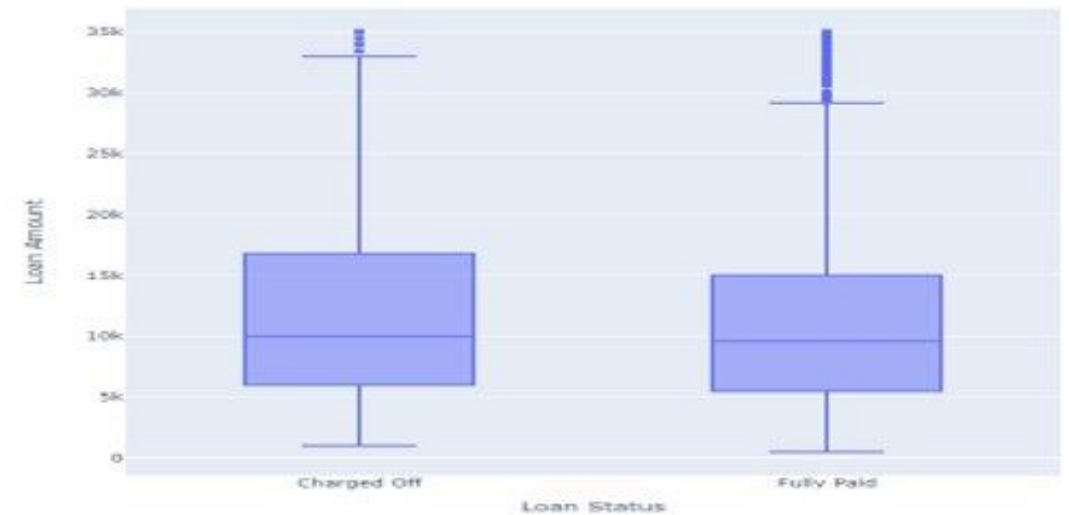
DATA ANALYSIS

Loan status + Loan Amount:

Insights:

- The plot depicting the Charged Off loans is slightly right skewed.

- In comparison to the loans being fully paid, higher amount loans appear to be charged off more.

Insights(Binning):

- Loans of high value (30k-35k) are being charged off the most(24.07%) which means high value loans are high risk.

- Loans graeter than 15k are being charged off more than 15% which is higher than the overall 14.14%.

- Loans between 5000-10000 are being charged off the least(12.42%) and paid off the most.

Spread of Loan Amount Across Various Loan Status

Loan Status Across Various Loan Amount Bins
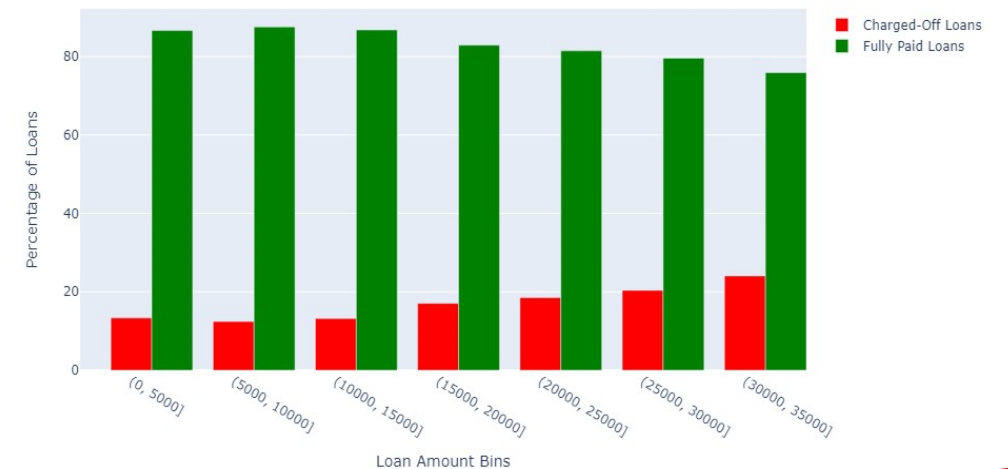
DATA ANALYSIS

Funded Amount Invested - Confidence of Investors:

Insight:

- The investors have had an equally mixed opinion on the borrowers.

- Overall, the investors have had full confidence while issuing 50% of the loans which turns out to be ~17.6k loans.

- Of the given out ~17.6k loans where the investors have had full confidence, 13.81% have been charged-off.

- Likewise, of the given out ~17.8k loans where the investors have not had full confidence, 14.4% have been charged-off.

- The confidence of the investors does not seem to have any direct impact on the repayment of loans.

Number of Loans Over Confidence of Investors
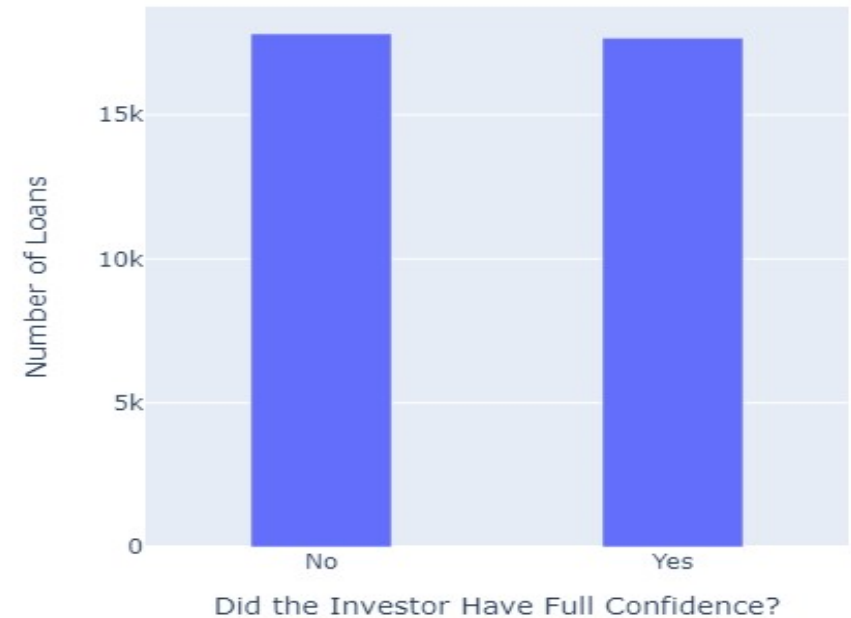
Percentage of Charged-Off Loans Over Confidence of In
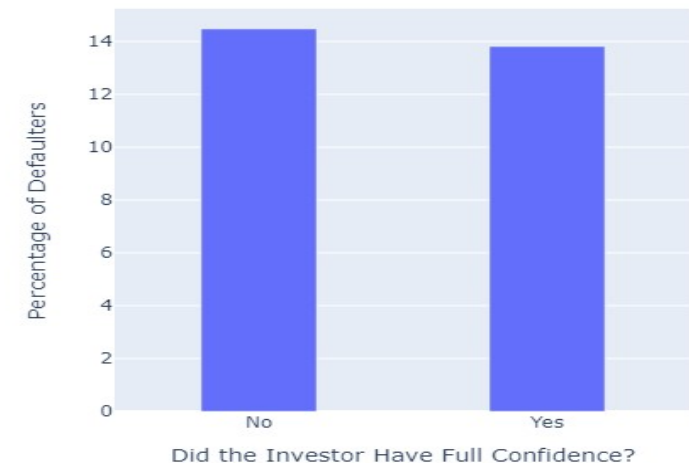
DATA ANALYSIS

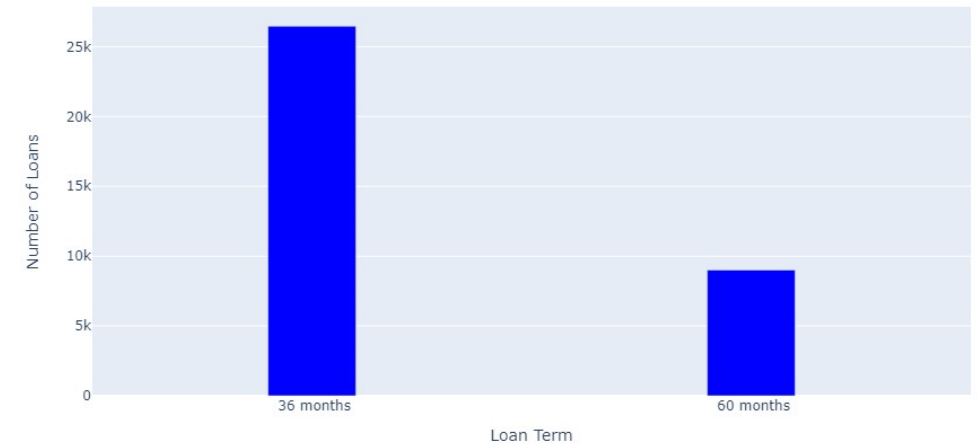## Loan Term:

▪ Short term or Long term

▪ Categoric variable

Insight 1:

▪ Borrowers tend to go for shorter loan term of 36 months.

▪ 26.47k people have opted for a short-term loan which is about 75% of the total loans given
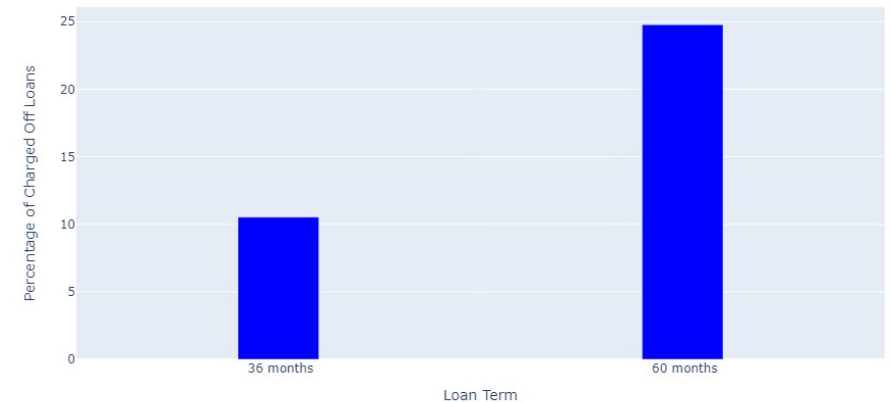
Insight 2:(Charged-off)

▪ Only 10% of loans with a 36-month term are being charged off, i.e., only 10% of 26.47k loans are being charged off.

▪ 24.76% of loans with a 60-month term are being charged off. This is way higher than the threshold value.

▪ Further insights on this metric can be derived during bivariate analysis with other metrics.

Number of Loans Over Loan Terms

Variation in Debtors Over Loan Terms

DATA ANALYSIS

Interest Rate:

- rate at which loans are given(numeric variable)

- Most loans have an interest rate between 7 to 13%.

Insight 1:(Interest Rate + Loan Status)

- The plot displaying charged off loans is right skewed.

- Higher interest rate loans are being charged off more with a median interest rate of 13.8%.

Insight 2:(Interest Rate + Loan Status- Charged-off)

- It is very evident that as interest rate goes up, percentage of charged off loans also increases.

- Loans with interest rate greater than or equal to 12% are being charged off more than 14.5% which is above the threshold value determined for our analysis.

- More than 30% of the loans with interest rate greater than 19% are being charged off which is twice the threshold value.

Spread of Interest Rate Over Loan Status
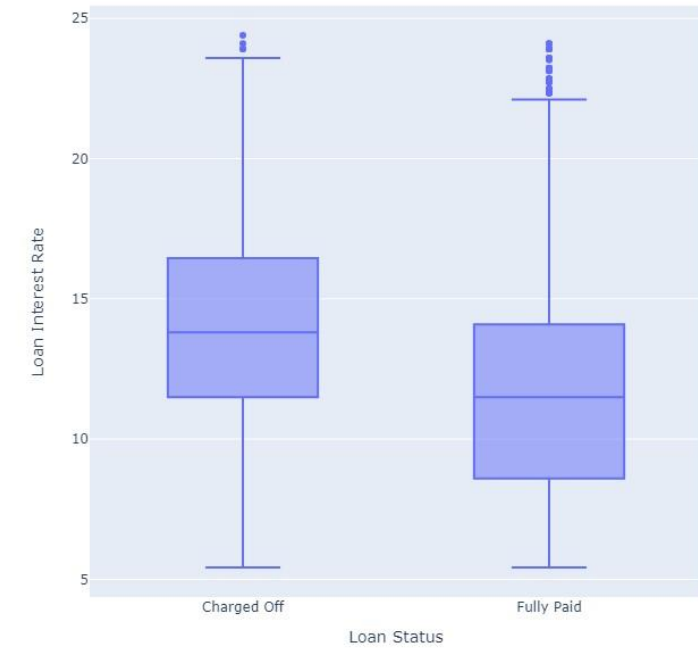
Variation in Debtors Over Loan Terms
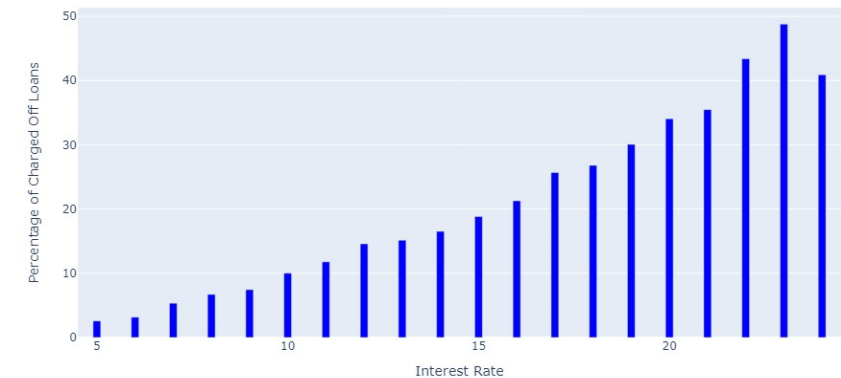
DATA ANALYSIS

Installment:

▪ Monthly amount to be paid by borrower (numeric variable)

▪ Most loans have a monthly installment between 100 to 400.

Insight 1:(Installment + Loan Status)

▪ The plot displaying charged off loans is slightly right skewed.

▪ Number of loans with higher monthly installments being charged off is more with a median installment amount of 296.84.

Insight 2:

▪ Loans with installment amount between 400 and 1000 are being charged off more than 14.5% which is above the threshold value determined for our analysis.

▪ It is interesting to note that for extremely high installment rates, the charge off % is lesser compared to the mid-range.

▪ It is interesting to note that for extremely high installment amount, the charge off % is lesser compared to the mid-range. It is possible that people with higher monthly installment are likeyly to close their loans since they will incur higher penalty when not done in time.

Spread of Monthly Installment Amount Over Loan Status
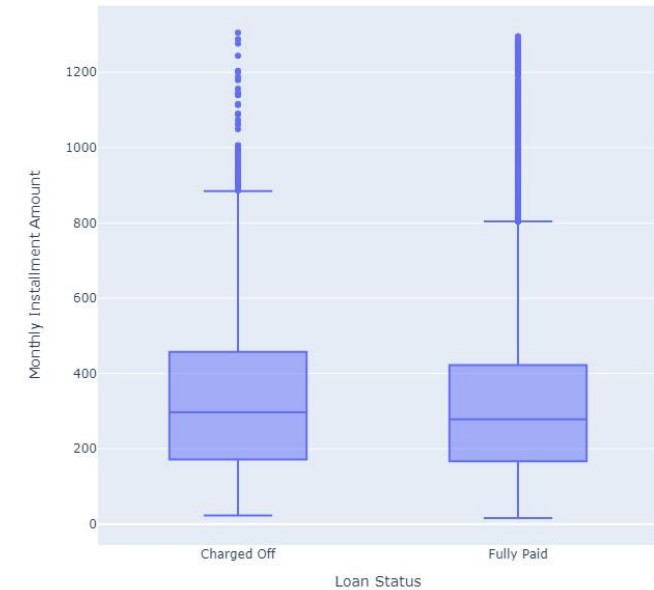
Loan Status Across Various Monthly Installment Amount Bins
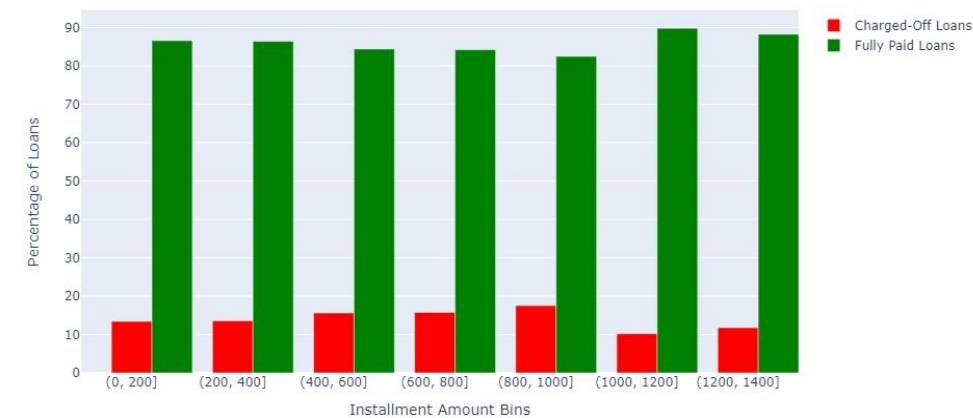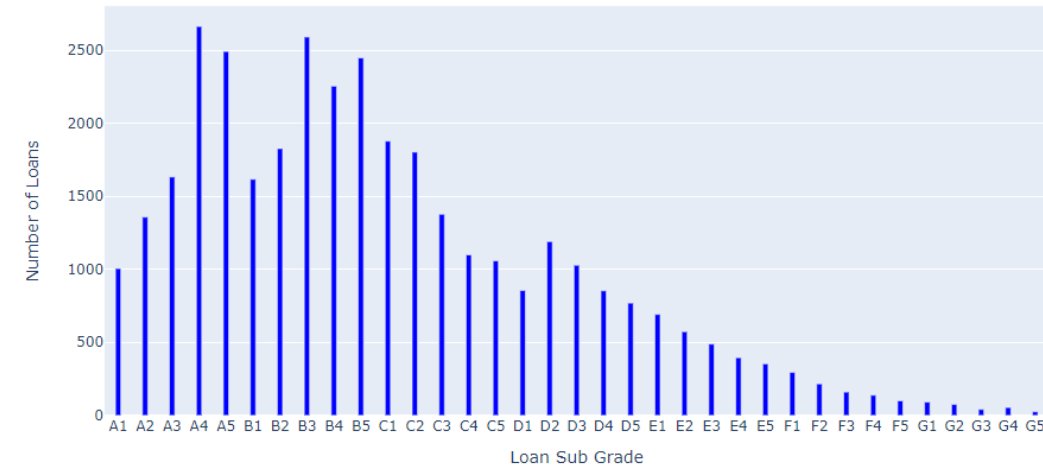
DATA ANALYSIS

Grade/Sub-Grade:

- Sub grade variable gives detailed analysis of grades therefore Sub-grade analysis is considered here.

- Sub-grade graph helps us identify which sub grades have most loans given out.

- ~40% of the loans are given out between A4 and C2.

Insights:(Sub-grade + Loan Status (Charged-off))

- Higher grade loans are high risk loans - they are being charged off more.

- Loans above C grade are being charged off more than 15% which is more than the threshold value.

- Loans in E4 and above have a charge off percentage greater than 28 which is twice more than the set threshold for analysis.

- Since sub-grades give deeper insights, the grade metric can be ignored.

Number of Loans Over Loan Sub Grade


Variation in Debtors Over Loan Sub Grades
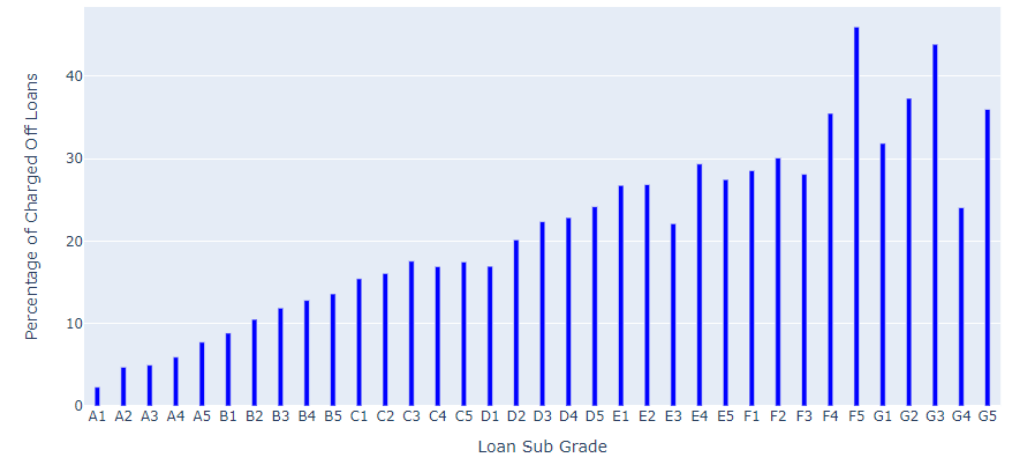
DATA ANALYSIS

Employee Title :

- <1% of borrowers have the same employee title. Grouping borrowers under this metric will not give us any effective analysis.

- It is a free text field. Example, US Army and U.S. Army are two values observed. Although they might mean the same, it is difficult to comprehend how many such different values are to be grouped together.

- Another example of similar kind is that of USAF, US Air Force, United States Air Force. It is unclear if USAF is different or same as US Air Force. Therefore, imputation also cannot be performed.

- Therefore, the employee title metric will be ignored for future analysis

Number of Loans Over Loan Grade
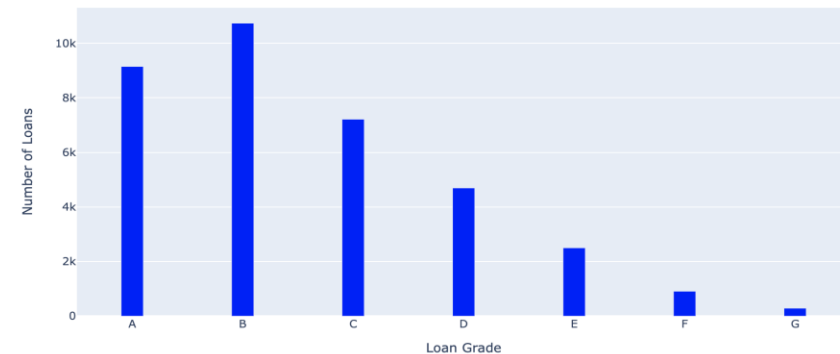
DATA ANALYSIS

Employee Length :

- People with the highest employment tenure tend to apply for more loans. More than 8000 loans which contribute to ~22% of the loans taken are by people with highest employment tenure (>10 years).

- There is no significant difference in the percentage of debtors across employment tenures

- In all the tenures, loans are being charged off at a rate closer to the threshold value taken for analysis (~14%)

- Employee length metric does not have any significant impact on loans being charged off

- We will analyze further with ither variables in bivariate analysis

- It can be observed that less experienced people and people with more than 10+ year experience are taking more loans.

Number of Loans Over Employee Tenure
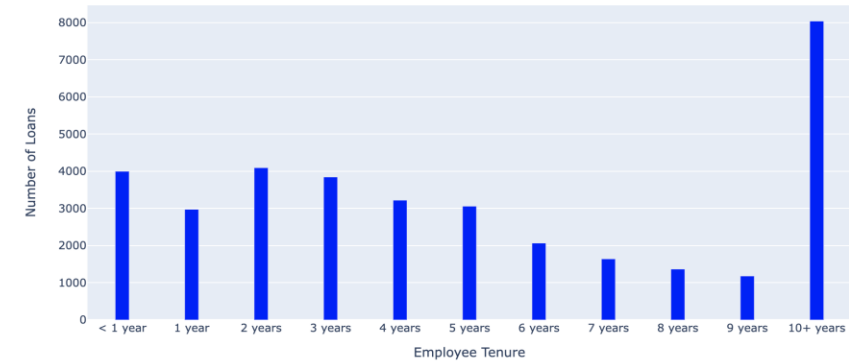
Variation in Debtors Over Employee Tenures

DATA ANALYSIS
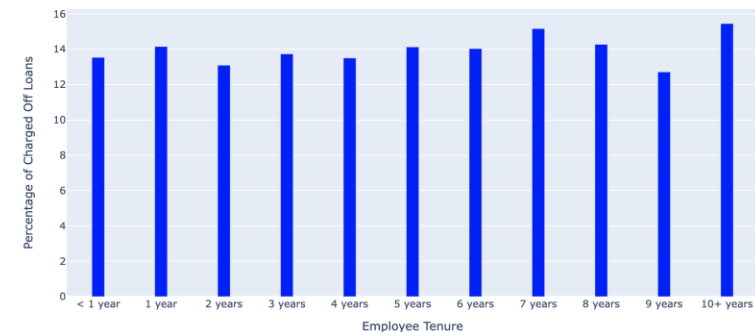
Number of Borrowers In Various Home Ownership Status

Home Ownership:

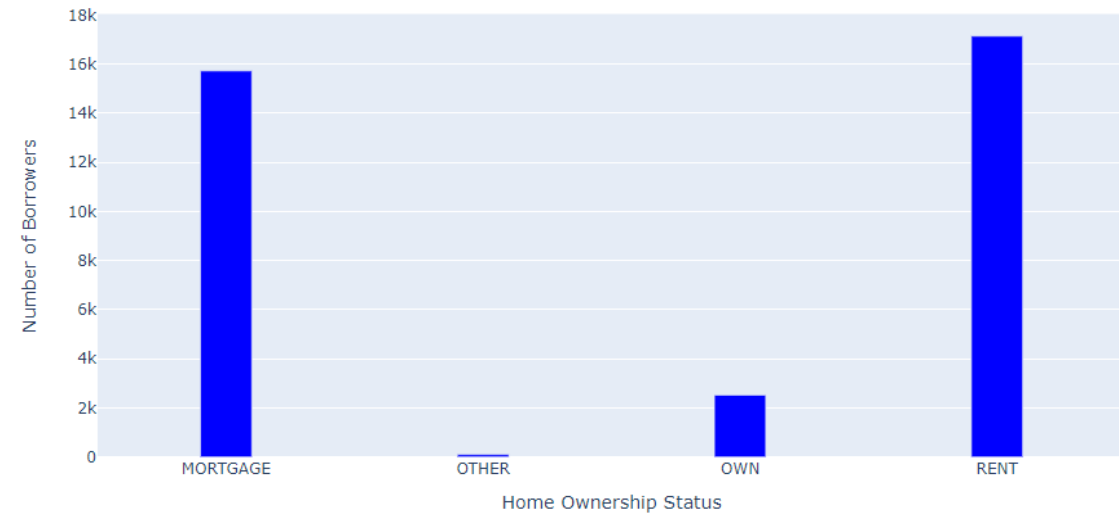▪ Categorical variable with 4 values

Insights:

▪ People on Rent and Mortgage contribute to over 90% of total loans taken (~32k loans)

Insight 2:(Home Ownership + Charge-off Loan Status)

▪ Although we noticed in the previous graph that home ownership category influenced whether a loan was taken or not, once a loan was taken, the rate of charge off doesn't vary much as can be seen in this graph.

▪ Borrowers under the 'other' category where their home ownership status is clearly unknown are charging off the most at 18.75% which is above the threshold value.

▪ Loans can be given out to borrowers whose home ownership status is known. Borrowers under 'other' category need to be dealt with caution.

Variation in Debtors Over Home Ownership Status
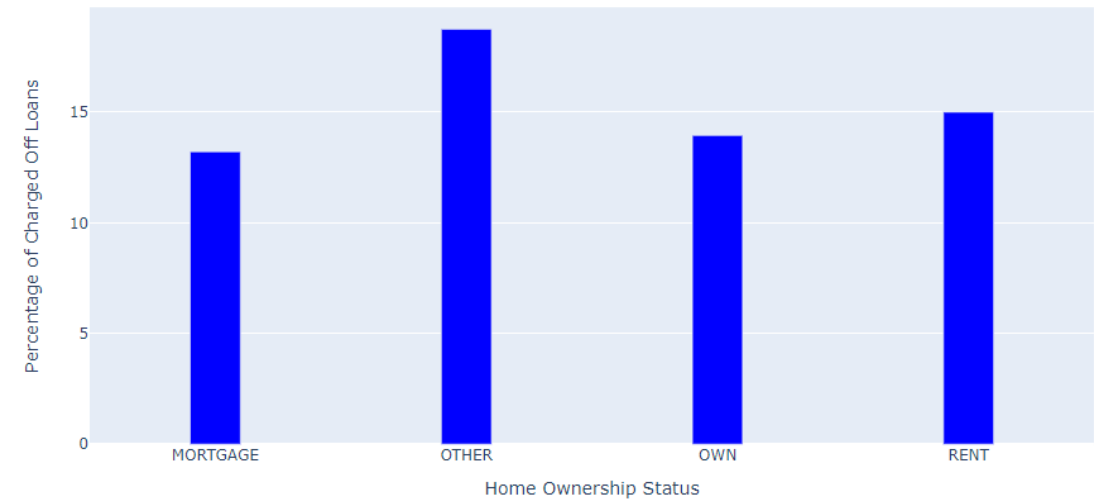
DATA ANALYSIS

Annual Income:

▪ Numeric Variable

▪ 95% of people have a salary of 140k or below. We will consider income from minimum to 140k for our analysis (ignoring outliers)

Insights:

▪ People with income between 40k to 60k seem to be taking the most loans.

▪ As salary increases, people tend to take lesser loans.

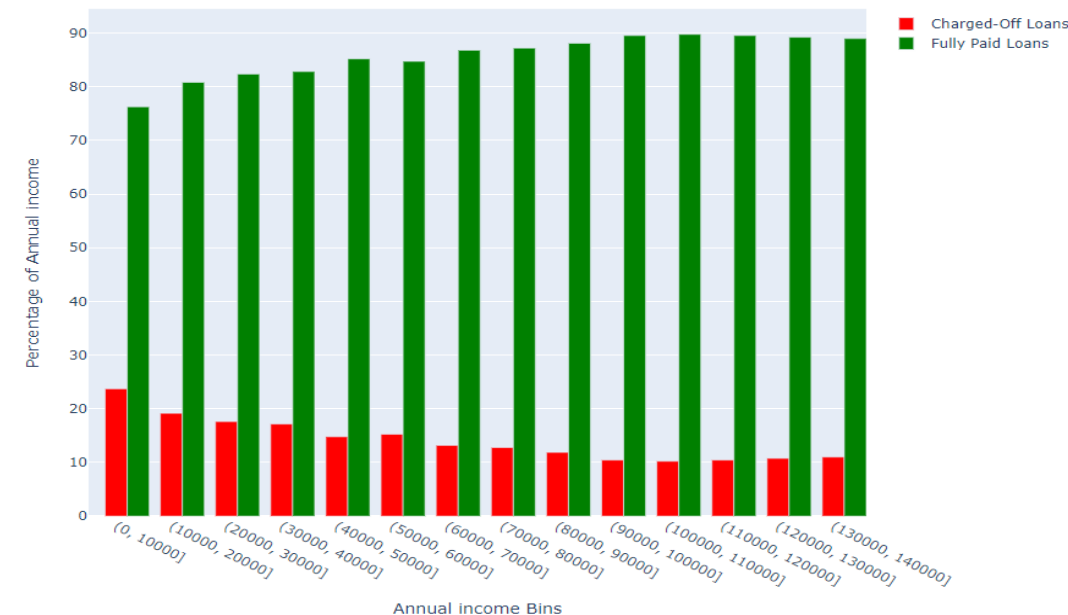▪ Likewise, even people with lower salaries are taking lesser loans.

Insight 2:(Annual Income + Loan Status)

▪ As annual income increases, percentage of borrowers defaulting is reducing.

▪ People with annual income > 60k default less than 15%.

Distribution of Annual Income
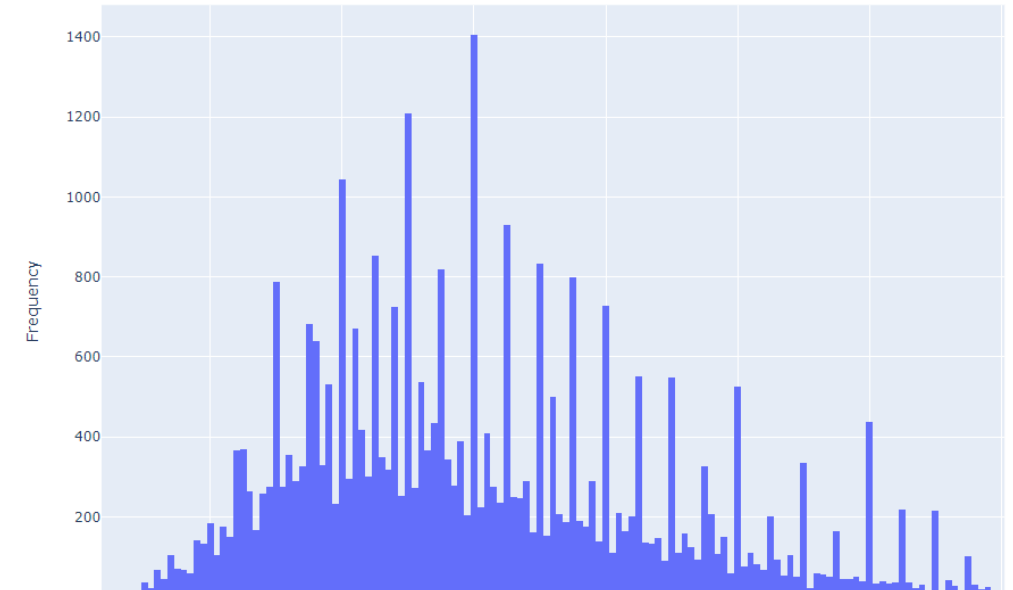
Loan Status Across Various annual income Bins
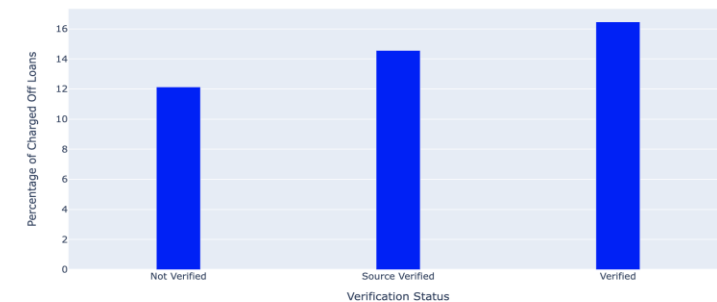
DATA ANALYSIS

Verification Status

- About 40% of the loans are given without any verification while for the remaining 60%, either the income or the source of income is being verified.

- The verification status does not have any considerable impact on loans being charged off.

- In fact, the verified loans seem to be charged off at an average of 15% which is closer to the threshold value.

- We can derive that the verification status of income of a borrower does not impact the payment of loans.


Number of Loans Over Verification Status
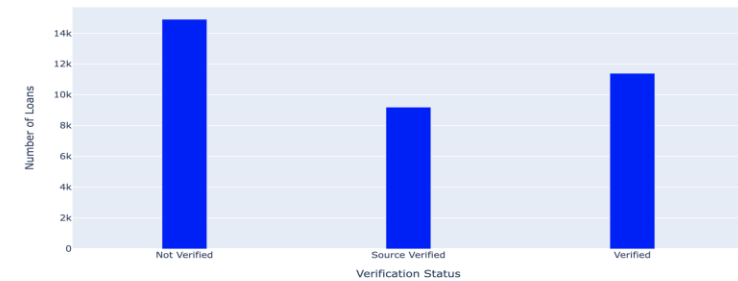
Percentage of Charged Off Loans By Verification Status

DATA ANALYSIS

Month of issue of loan

- Derived from date of issue

- Maximum loans are taken in the months of November and December. One can guess that it is due to the holiday season (Black Friday Sale, Thanksgiving, Christmas) where expenditure is high.

- The number of loans taken drops to almost half in January. One can guess it is due to the holiday season that has come to an end already.

- The month of issue of loan is not paying a significant role on the repayment of loans.

Number of Loans Taken Over Various Months

Percentage of Charged Off Loans By Month of Issue
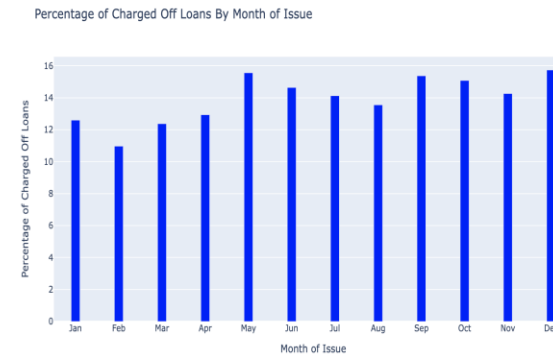
DATA ANALYSIS

Loan Purpose:

▪ Categorical variable

▪ Nearly 50% of the loans are given for debt consolidation followed by credit card purposes.

Insights: (Loan Purpose + Charge off loan status)

▪ 26.7% of all small business borrowers are being charged off, followed closely by renewable energy and educational loans.

▪ Major purchase, wedding and credit card loans are being paid off the most.

▪ Small business loans are defaulting more

Number of Loans Taken For Various Purposes
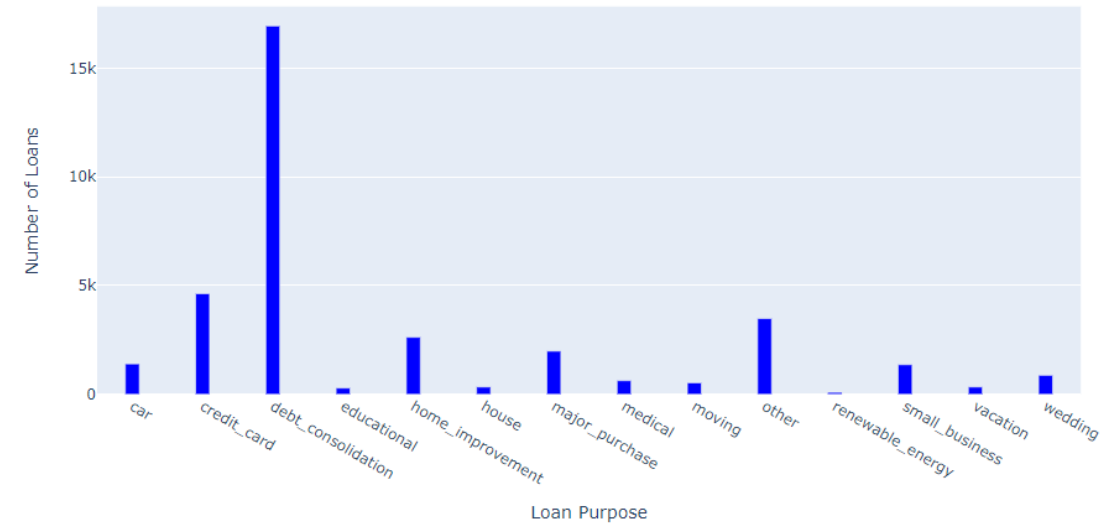
Variation in Charged of Loans Across Various Loan Purposes
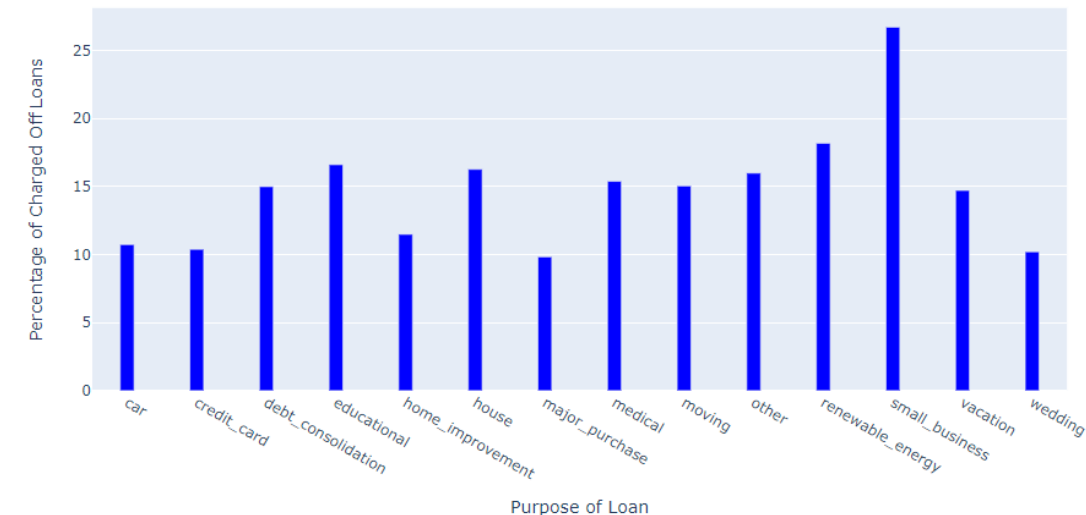
DATA ANALYSIS

State:

- Categorical variable

- Total 49 states

Insights:

- The state of CA has the highest number of loans of 6419 which is close to 20% of all loans, followed by NY

- Top 75% of states of loan issued are being considered for further analysis

- NV, FL, NM default more.

Zip Code:

- It can be seen that the entire 5-digits of the zip code is not visible.

- The first three digits can correspond to the US state. For example, 309xx belongs to Augusta in Georgia.

- Since we have the state information in addr_state column, we will ignore the zip_code metric for analysis.

Number of Loans Taken in Various States
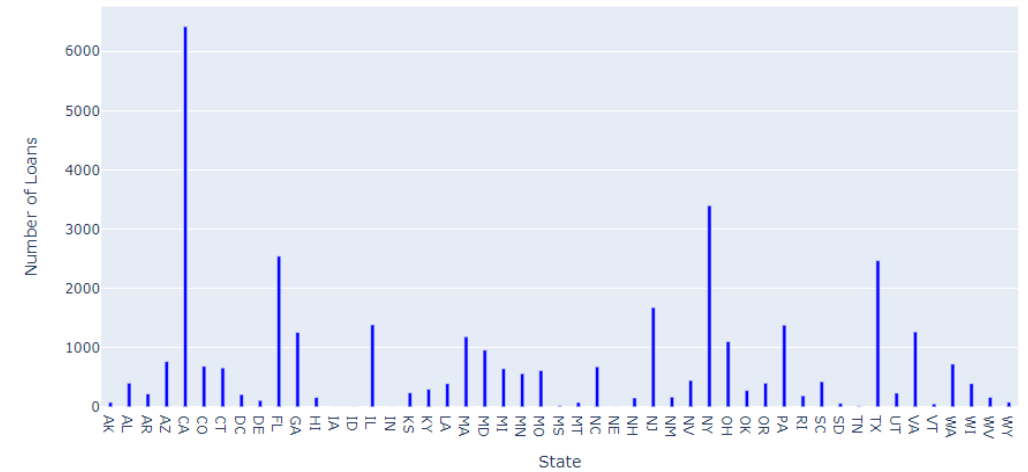
Variation in Charged off Loans Across States
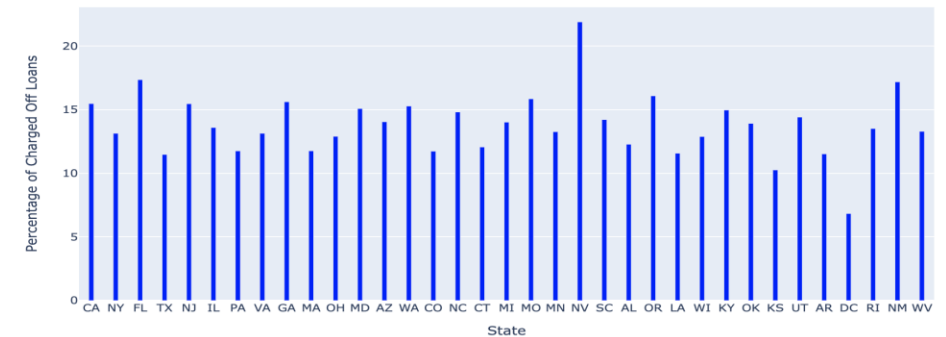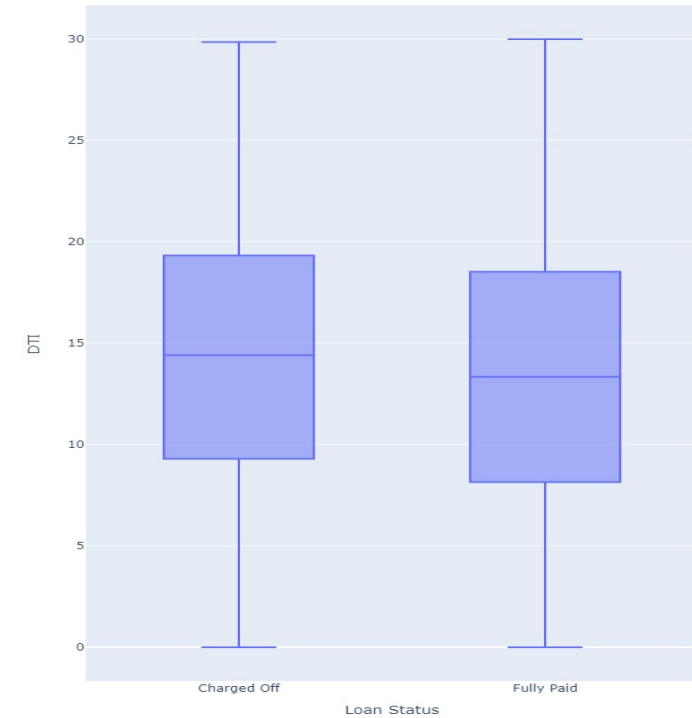
DATA ANALYSIS

DTI

- Numeric variable

- A lower DTI ratio indicates that an individual has a lower level of debt relative to their income, suggesting a lower risk for lenders. On the other hand, a higher DTI ratio suggests that an individual has a higher level of debt in comparison to their income, which may increase the risk for lenders.
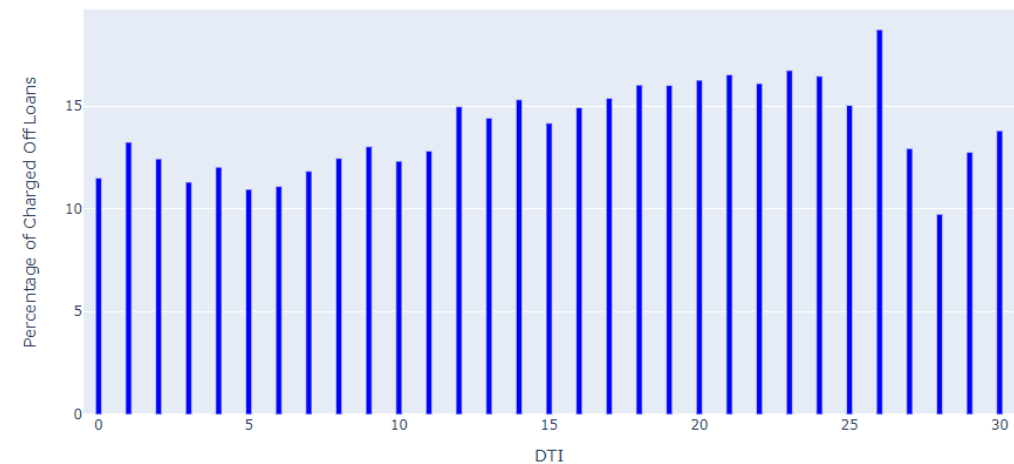
Insights:

- Charged Off Loans are right skewed; i.e., higher the DTI, higher the number of charged off loans with a median dti of 14.41

- Similar to the box plot, we can see that more number of loans are obtained for a DTI between 12 and 15..

- DTI values range from 0 to 30.

- Over 15% loans are being charged off for dti over 15 which is higher than our set threshold value for analysis.

- A downward curve is seen for lower dti and an upward curve is visible for higher dti which confirms that at high dti, risk of charging off is more.

Spread of DTI Across Various Loan Status

Variation in Debtors Over DTI

DATA ANALYSIS

## Public Record Bankruptcies

- Numeric variable (0,1,2)

Insights:

- Borrowers with no record of bankruptcy are given more loans.

- The bar chart is not very clear but the value counts above shows that there are 5 records of borrowers with 2 bankruptcy records.

- People with prior record of bankruptcy tend to charge off again

Number of Loans Taken if Bankruptcy Record Exists
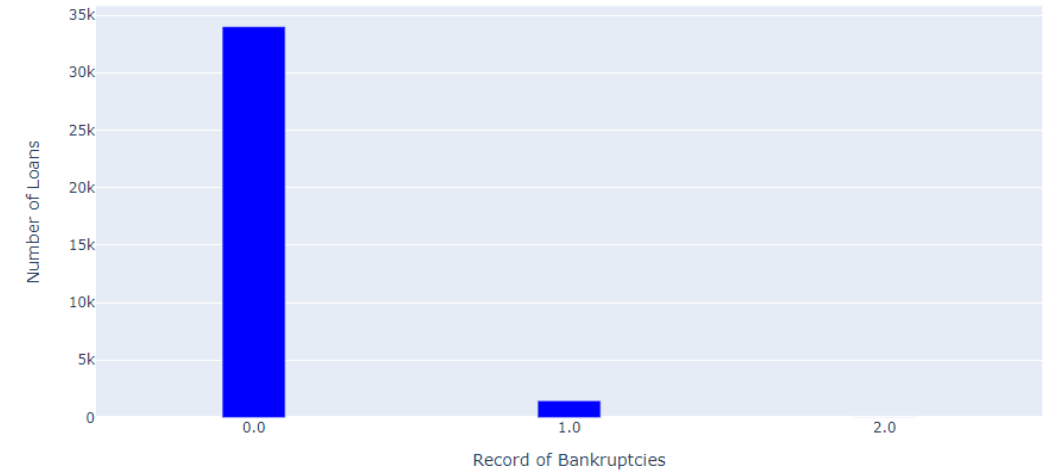
Variation in Charged off Loans if Bankruptcy Record Exists
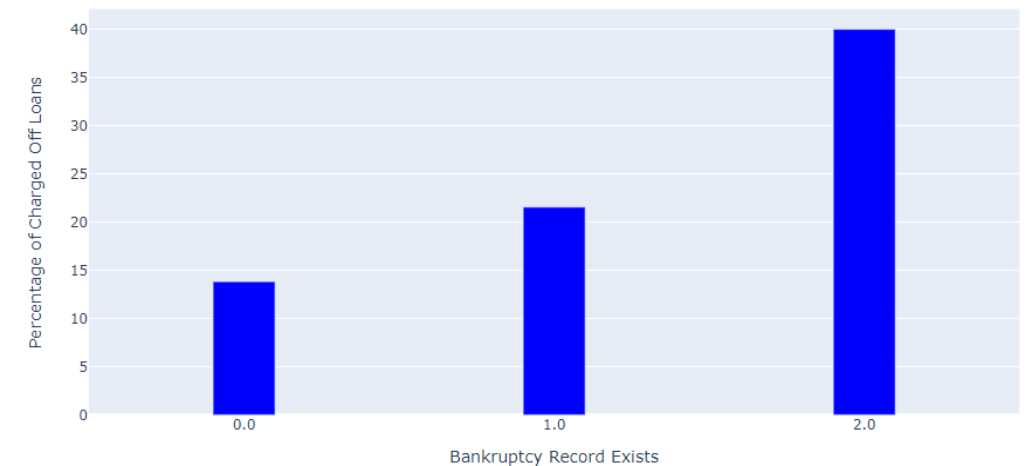
DATA ANALYSIS

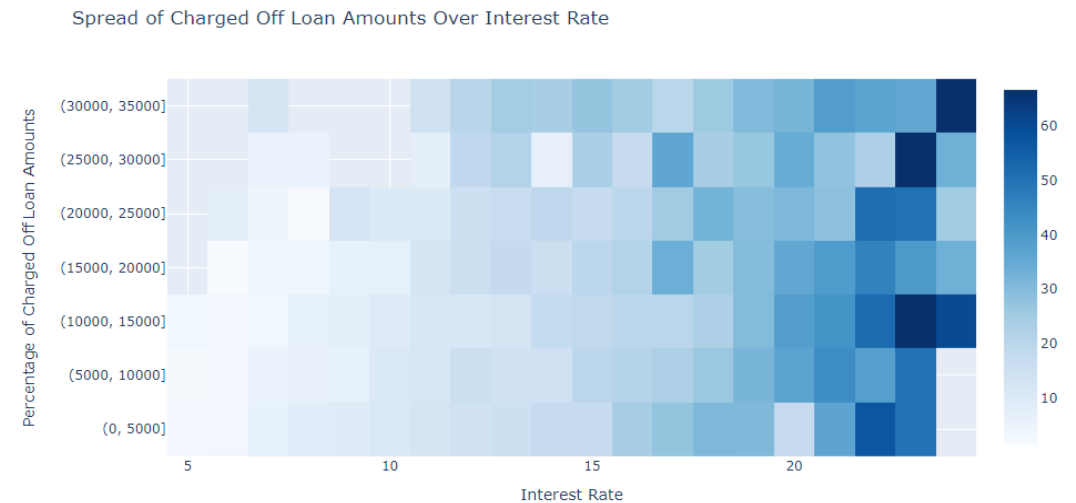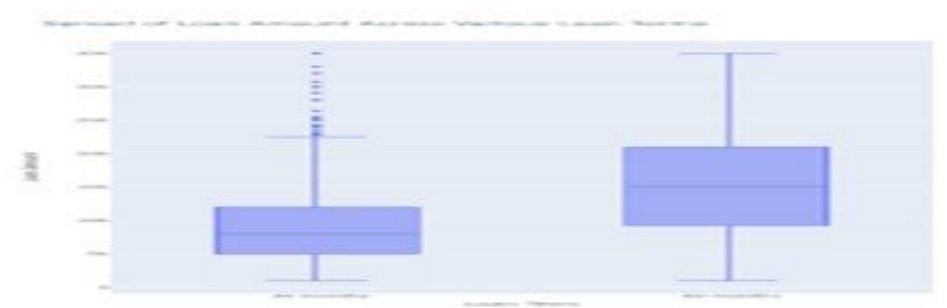## BIVARIATE ANALYSIS ON LOAN AMOUNT WITH OTHER METRICS

Insights:

Loan Amount Vs Loan Term

- larger loan amounts taken for a longer term are being charged off more.
- The graph shows that long term loans of larger loan amounts (>15000) charge off over 25% which is a cause for concern.

Loan Amount Vs Interest Rate

- The number of loans with interest rate greater than 15% is charging off more across all loan amount bins.

Spread of Charged Off Loan Amounts Over Loan Term

Spread of Charged Off Loan Amounts Over Interest Rate

DATA ANALYSIS

BIVARIATE ANALYSIS ON LOAN AMOUNT WITH OTHER METRICS

Insights:

Loan Amount Vs Installment

- Larger loan amounts have larger installment amounts, highly correlated. So, installment amount can be left out..

Distribution of Loan Amount Vs Installment Amounts

Loan Amount Vs Sub-Grades

- For smaller loan amounts, charge off happens at higher risk grades.
- As loan amount increases, charge off is happening starting from relatively lower risk grade.
- Across all loan amounts, however, charge off at high risk grades (F, G) is more.
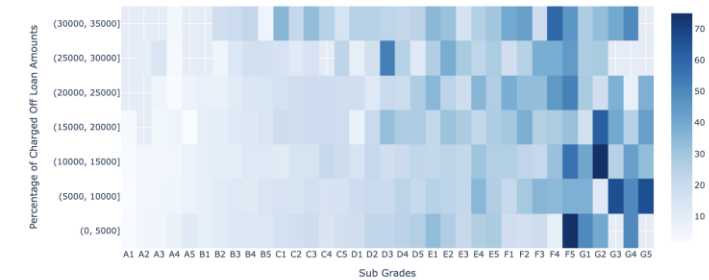

Spread of Charged Off Loan Amounts Over Loan Sub Grades

Loan Amount Vs Home Ownership

- Not much variation observed across the categories

Spread of Charged Off Loan Amounts Over Home Ownership category
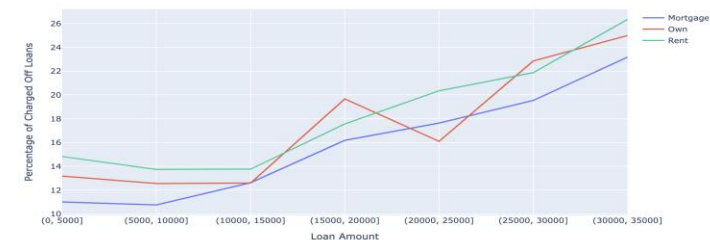
DATA ANALYSIS

BIVARIATE ANALYSIS ON LOAN AMOUNT WITH OTHER METRICS

Insights:

Loan Amount Vs Annual Income
- People with annual income from 20k to 60k taking loan more than 15k are charging off more.

Loan Amount Vs Month of Issue
- Loan amounts given across all months are evenly spread. There is a slight increase in the amount for December.
- This analysis, however, does not give us insight into how it affects the charge off. Therefore, month of issue is not a metric that affects charge off percentage.

Loan Amount Vs Loan Purpose
- From univariate analysis, we know small businesses charge off at a higher percentage. Here, we can see that insight being confirmed with the darker blues.
- In small businesses, loan amounts from 10k to 35k charge off more, especially the higher loan amounts of 25k and above.
- It is interesting to see that the loan amounts between 20k to 25k have a higher charge off percentage across loan purposes.
- A crucial insight here is that higher loan amounts (>25k) are charging off at a significant rate when taken for medical purposes. One can think that it could be due to increased death rates for crucial ailments.

Spread of Charged Off Loan Amounts Over Annual Income
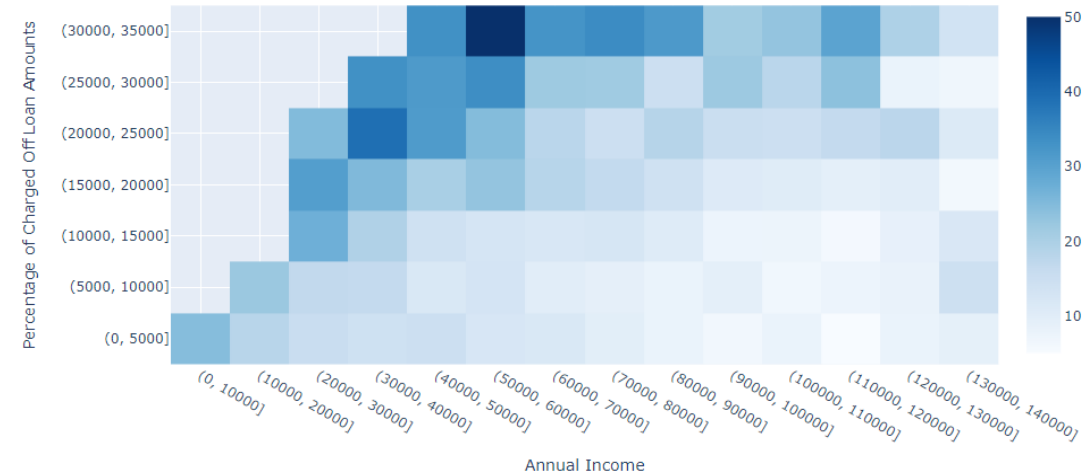
Spread of Charged Off Loan Amounts Over Loan Purpose
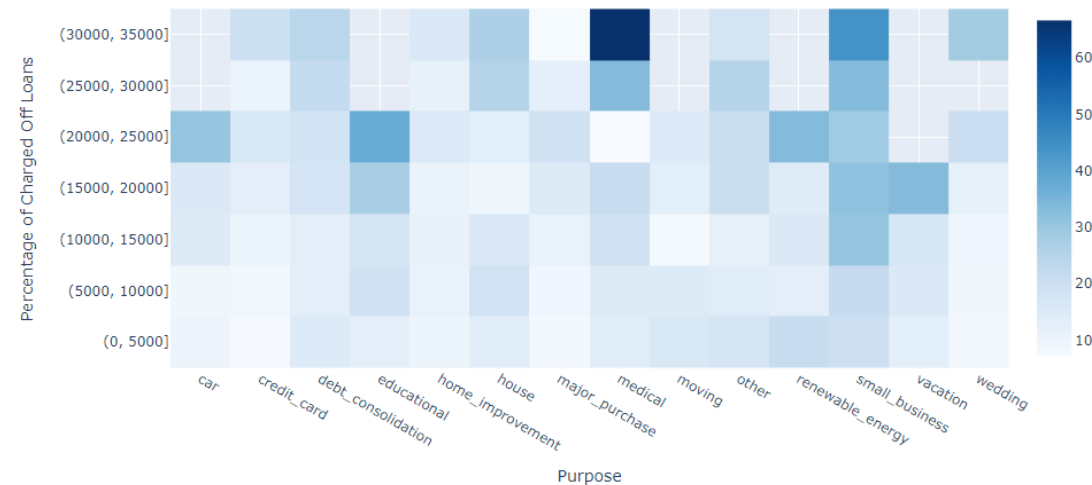
DATA ANALYSIS

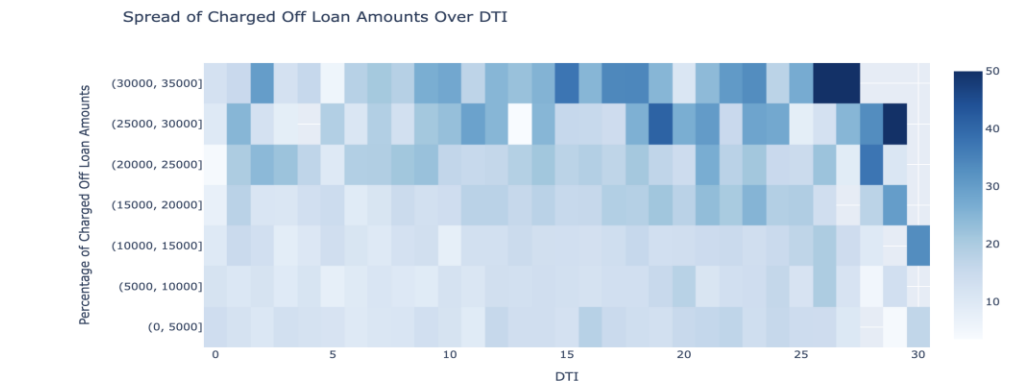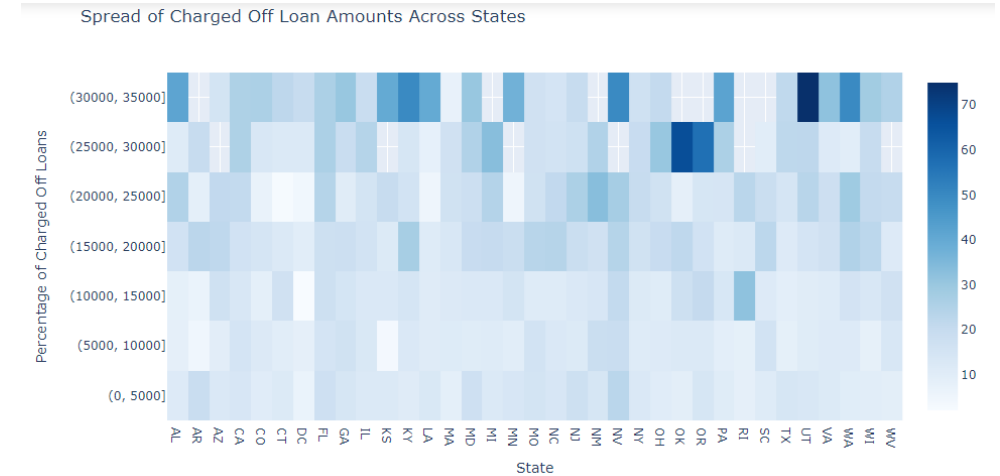## BIVARIATE ANALYSIS ON LOAN AMOUNT WITH OTHER METRICS

Insights:

Loan Amount Vs State

- Very few people take loans greater than 30k in all states as per the outliers in the above graph.
- We know that most loans are given out in CA, NY as per univariate analysis.
- Most charge off happens at NV, FL, NM as per univariate analysis.
- Let's concentrate on these states here.
- CA has more charge off for higher loans amounts (> 20k)
- It is clear that NV is charging off across all loan amounts. Loans amounts for this state have to be given out carefully.
- NY pays off loans well.
- In general, across states, charge off is high for higher loan amounts (>25k)

Loan Amount Vs DTI

- Lending Club gives out lesser loans at higher DTI.= which is good.
- Loan density is high between 10 to 25 dti.
- Smaller loan amounts (<10k) are given out irrespective of the dti.
- Most loans are charged off for higher loan amounts(>20k) for high dti(>16)

Spread of Loan Amount Across States

Spread of Charged Off Loan Amounts Across States

Spread of Charged Off Loan Amounts Over DTI

DATA ANALYSIS

BIVARIATE ANALYSIS ON LOAN AMOUNT WITH OTHER METRICS

Insights:

Loan Amount Vs Bankruptices

- It is good to avoid loans for borrowers with previous bankruptcy records.
- Lending club is already giving out very meagre number of loans to people with prior bankruptcy history.

Spread of Charged Off Loan Amounts Over Bankruptcy Record

DATA ANALYSIS

BIVARIATE ANALYSIS ON LOAN TERM WITH OTHER METRICS

Spread of Interest Rates Over Loan Term
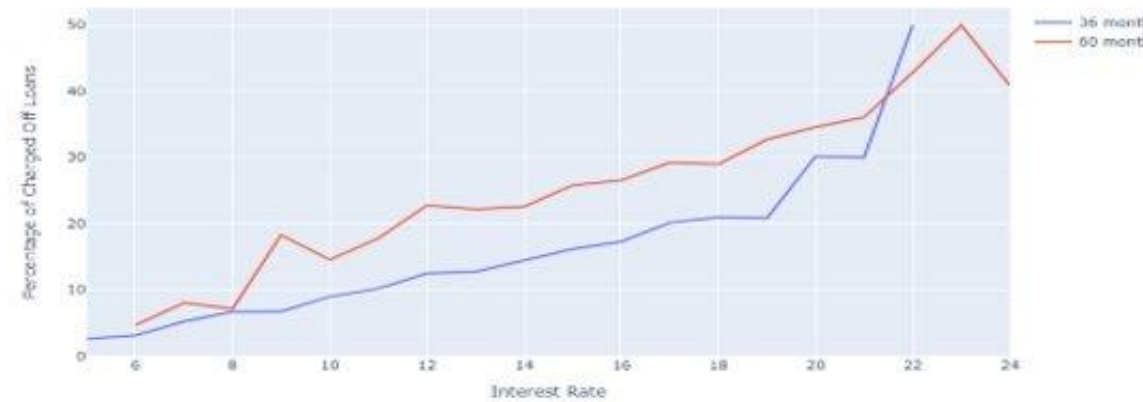
Insights:

Loan Term Vs Interest Rate

- Short term loans are given at relatively lesser interest rates and long-term loans are given at higher interest rates.
- It is clear that long term loans have a higher charge off rate as interest rate increases.
- But after int rate >20, the gap between both loan terms reduces.
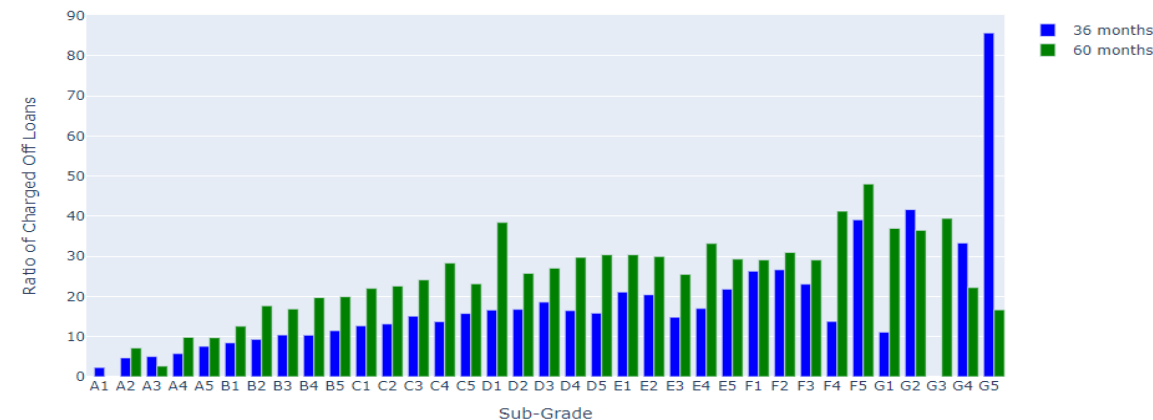

Loan Term Vs Sub-Grades

- Long term loans across all sub grades are charging off at a higher percentage.
- Interesting to note that for G2, G4 and G5, short term loans are charging off more.
- High risk loans have high interest rate. Irrespective of the term, high risk loans (Grade G) charge off more.

Loan Sub-Grades In Each Term

DATA ANALYSIS

BIVARIATE ANALYSIS ON LOAN TERM WITH OTHER METRICS

Insights:

Loan Term Vs Home Ownership
- We have removed 'other' category from analysis as suggested in Univariate analysis.
- We already know that long term loans are charged off more. So, home ownership does not impact loan term.
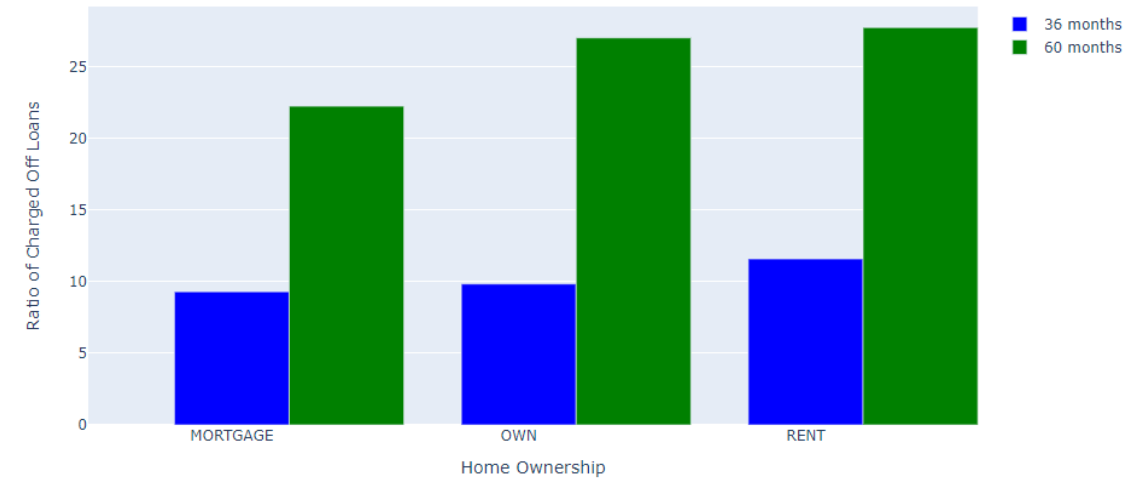
Loan Term Vs Annual Income
- People with income in all brackets take loan across both terms.
- People with income < 20k have a higher charge off percentage.
- Then there is a dip in charge off percentage and then after 50k, the charge off percentage increases until 90k

Loan Term Vs Loan Purpose
- Credit card, debt consolidation, educational, house, major purchase, small business, vacation - for these purpose, long term loans charge off more than twice the percentage when compared to short term loans.
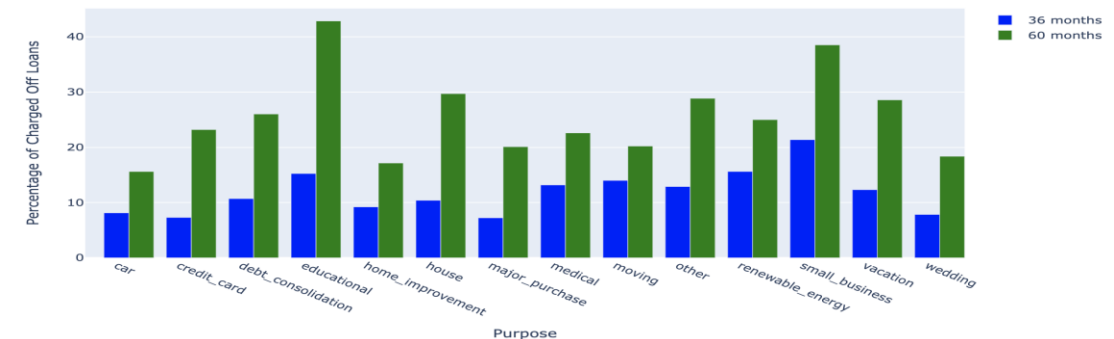

Home Ownership In Each Term

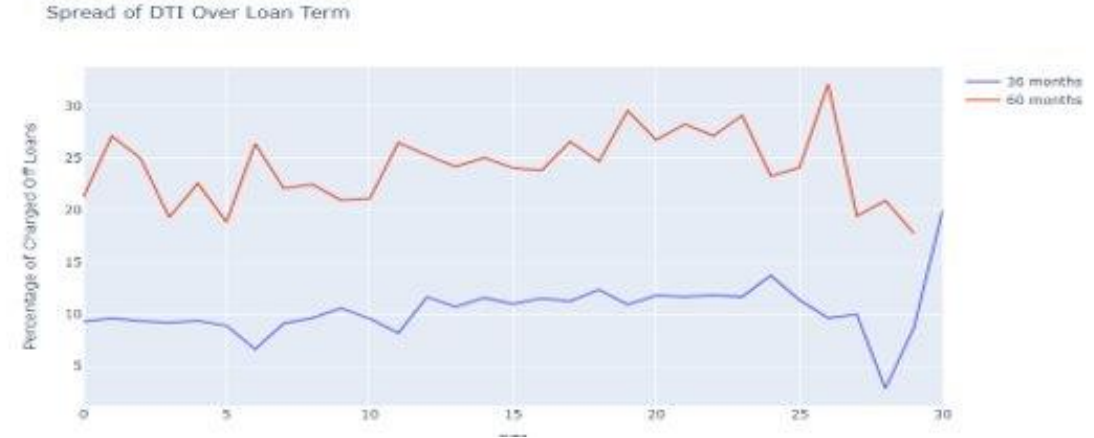Spread of Annual Income Over Loan Term
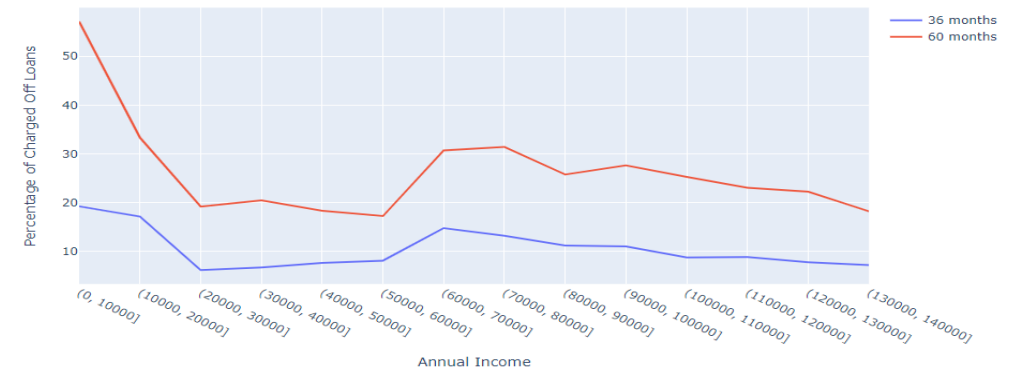
Purpose Vs Term

DATA ANALYSIS

BIVARIATE ANALYSIS ON LOAN TERM WITH OTHER METRICS

Insights:

Loan Term Vs DTI

- People with DTI in all ranges equally take loans in both short term and long term.
- Long term loans charge off more, Long term loans strongly default more in comparison to short term ones.

Spread of DTI Over Loan Term

DATA ANALYSIS

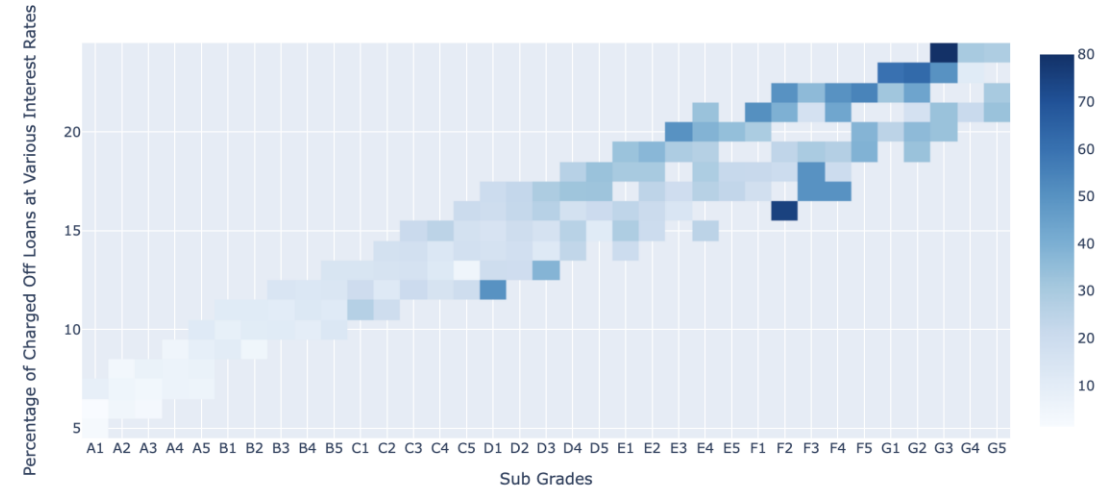## BIVARIATE ANALYSIS ON INTEREST RATE WITH OTHER METRICS

Insights:

Interest Rate Vs SubGrade

- High risk loans are given at high interest rates.
- From univariate analysis of both the metrics individually, we know that high risk loans and high interest rate loans are charged off at a higher rate.
- The correlation between these two metrics is high.
- For further analysis, we can consider just one of these metrics and understand that any impact on interest rate indirectly impacts grades of loans.
- Interest Rate and Grades are highly correlated.
- Higher the risk of the loan (grade: a<b<c<d<e<f<g), hgiher the interest rate at which it is being charged off.
- We can consider just one of these variables for future analysis.

Interest Rate Vs Annual Income

- Interest Rate and Annual Income have an even spread when loans are taken.
- As income increases, lesser loans are taken.
- Irrespective of annual income, at higher interest rate >20%, more default happens.
- For income upto 100000, loans at interest rate from 17% itself is charging off more.

Spread of Charged Off Loans at Various Interest Rates Over Loan Sub Grades

Spread of Charged Off Loan Amounts in Various Annual Income Brackets Over Interest Rate
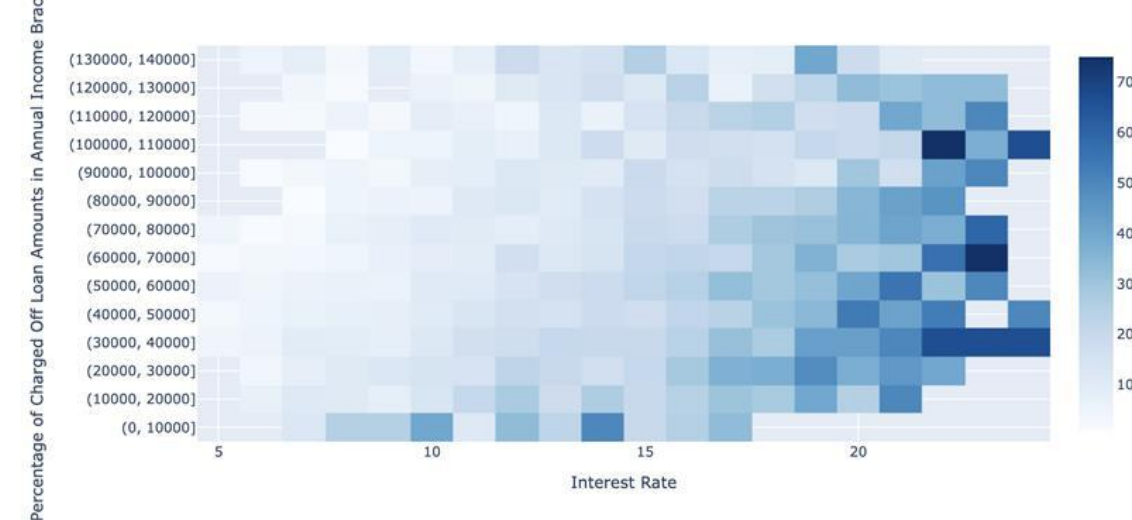
DATA ANALYSIS

BIVARIATE ANALYSIS ON INTEREST RATE WITH OTHER METRICS
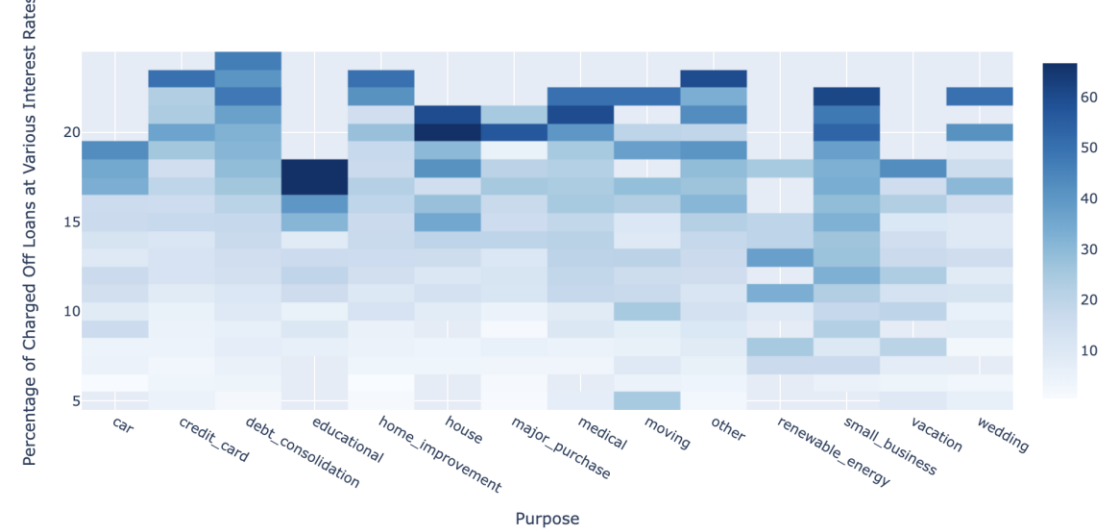
Insights:

Interest Rate Vs Loan Purpose

- High interest rate is a cause for charge off across all purposes.
- We clearly see that small business loans have a darker blue shade across a wider interest rate range indicating a greater number of small business loans being charged off.
- Educational loans at 17-18% ,housing loans at 15-20% have a higher charge off rate in comparison to other loans at this interest rate.
- Loans taken for moving purposes charge off at lower interest rates too.

Interest Rate Vs DTI

- Loans at higher interest rate irrespective of dti are being charged off.
- Likewise, loans at higher dti irrespective of interest rate are being charged off more wherever given.

Spread of Charged Off Loans at Various Interest Rates Over Loan Purpose

Spread of Charged Off Loan Amounts at Various Interest Rates Over DTI
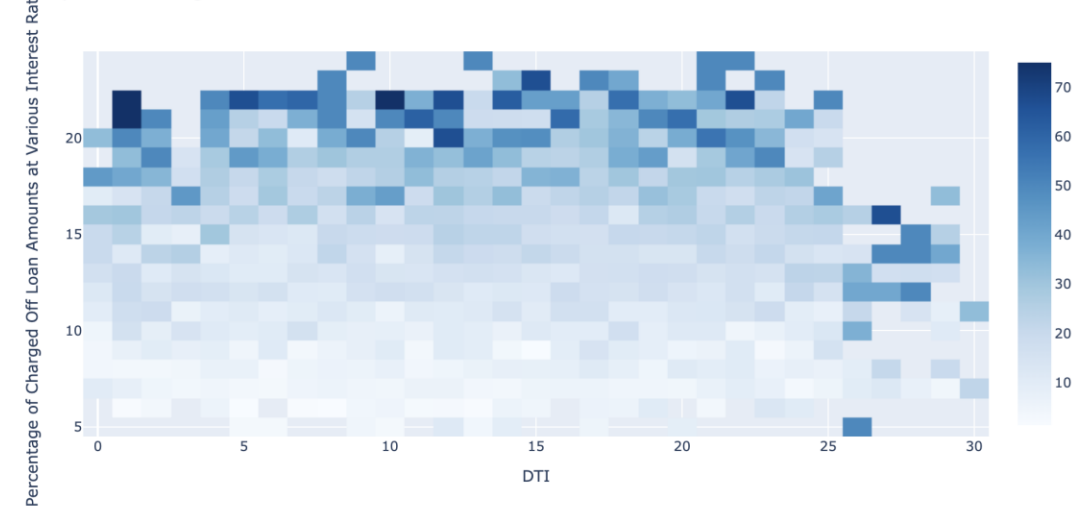
DATA ANALYSIS

BIVARIATE ANALYSIS ON ANNUAL INCOME WITH OTHER METRICS
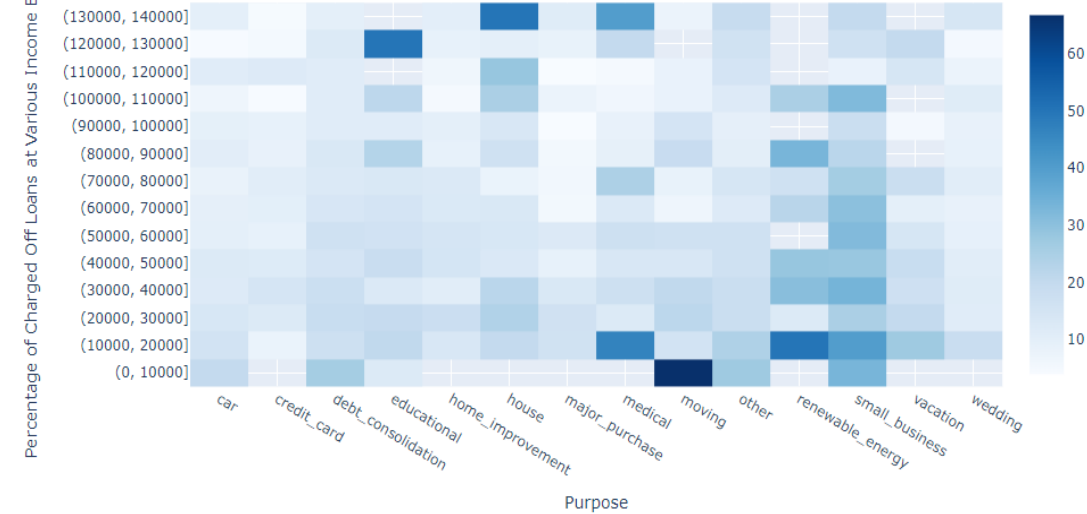
Insights:

Annual Income Vs Purpose

- Borrowers taken loans for various purpose across all income groups.
- More people in the income range above 80k take loans for small business and vacations.
- Annual income does not have any impact on the purpose of loan being charged off.
- However, it is clear that small business and renewable energy loans are being charged off at all income levels.

Annual Income Vs DTI

- Most loans have been given off between the range of 20k and 100k annual income but the dti is well spread.
- Correlation between these two metrics is less.
- We see a slight increase in the blue shade for lower income levels as DTI increases.
- Upto an annual income of 40k with dti above 10 seems to have more charged off loan percentage.

Spread of Charged Off Loans at Various Income Brackets Over Loan Purpose

Spread of Charged Off Loan Amounts at Various Income Brackets Over DTI
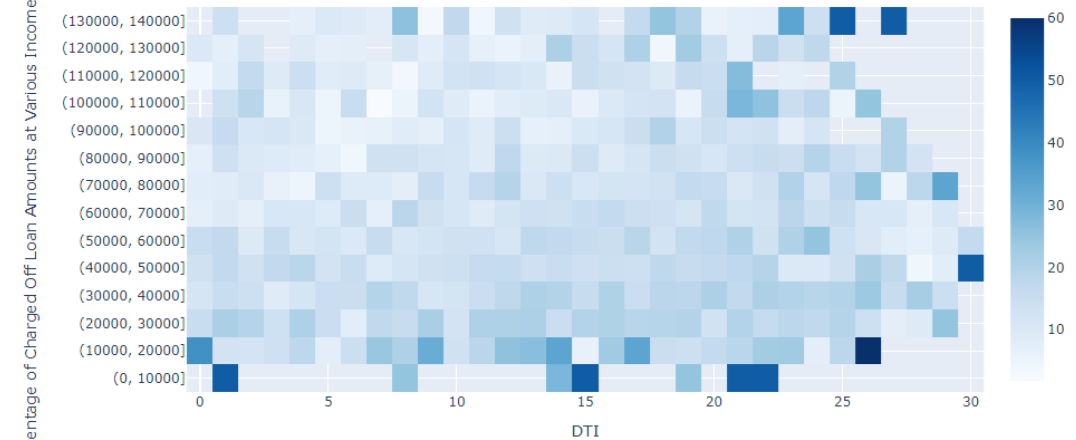
DATA ANALYSIS

## BIVARIATE ANALYSIS WITH OTHER METRICS

Insights:

Loan Purpose Vs DTI

- High risk loans are given for credit card and debt consolidation in comparison to the other loan purposes.
- Small business loans are being charged off at all DTIs.
- There is a clear indication that loans with higher DTIs are being charged off at a higher rate.
- Renewable sources of energy loans are charging off at a high rate across dtis.

Spread of Charged Off Loans for Various DTIs Over Loan Purpose
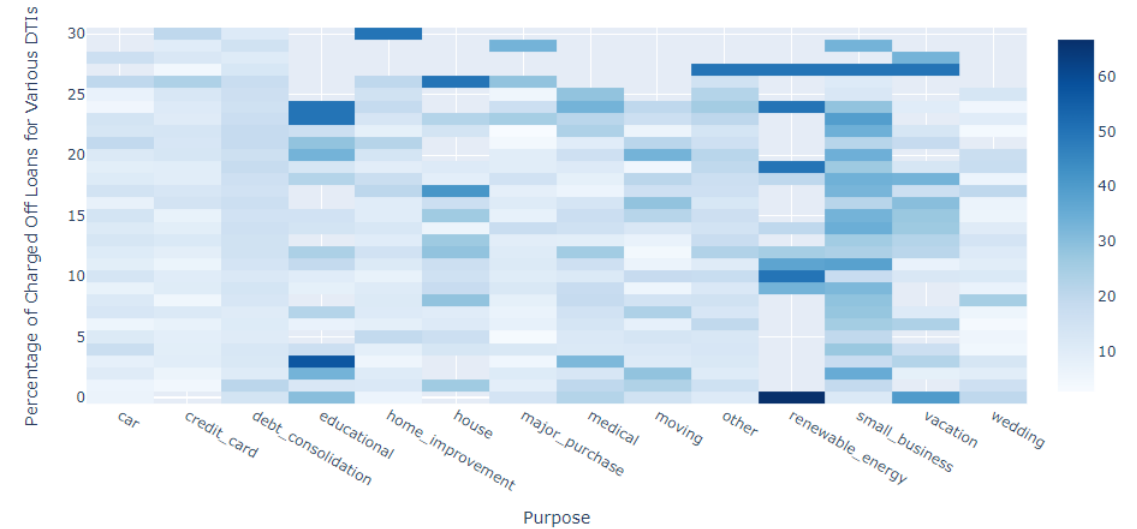
RECOMMENDATION
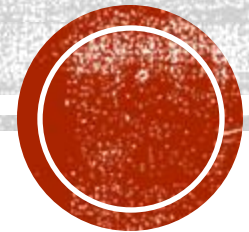
RECOMMENDATION

- Larger loan amounts (> 15000) for 60-month term across high-risk grades (D to G) must be given with the caution.

- In general, small business loans are charging of at a high rate - these loans have to be given out with caution considering these additional variables - Loan amount taken between 10k to 35k, for 60-month term with interest rate > 9%.

- Higher loan amounts (>20k) having higher dti (>16) must be given out with caution.

- Educational and housing loans taken at 60-month term with interest rate >15% must be given out with caution.

- People with annual income between 20k to 60k taking loans more than 15k at higher interest rates (>17%) tend to have higher dti (>10 - these are high risk loans) must be considered with caution.

- High risk loans with grades E,F,G charge off more than 30% and must be given out with caution.

- Loans given out in NV state must be considered with caution.

- Although we do not have sufficient data on borrowers with previous record of bankruptcies, from the available dataset, it can be seen that people with previous record of bankruptcies tend to default again. These loans must be given with less priority.

THANK YOU