

Lending Loan Case Study

Atharv Ganla

Nikhil Mothkuri

Agenda

- Data Analysis
- Data Cleansing
- Univariate and Segmented
Univariate analysis
- Bivariate and multi variate analysis
- Conclusion



Introduction

The loan dataset is provided for past and existing customers. As Learners we have to understand and analyse the data to provide substantial evidence on the factors which can play a major role in delinquency of loans and those loans which are most likely not be paid fully.

Data Analysis

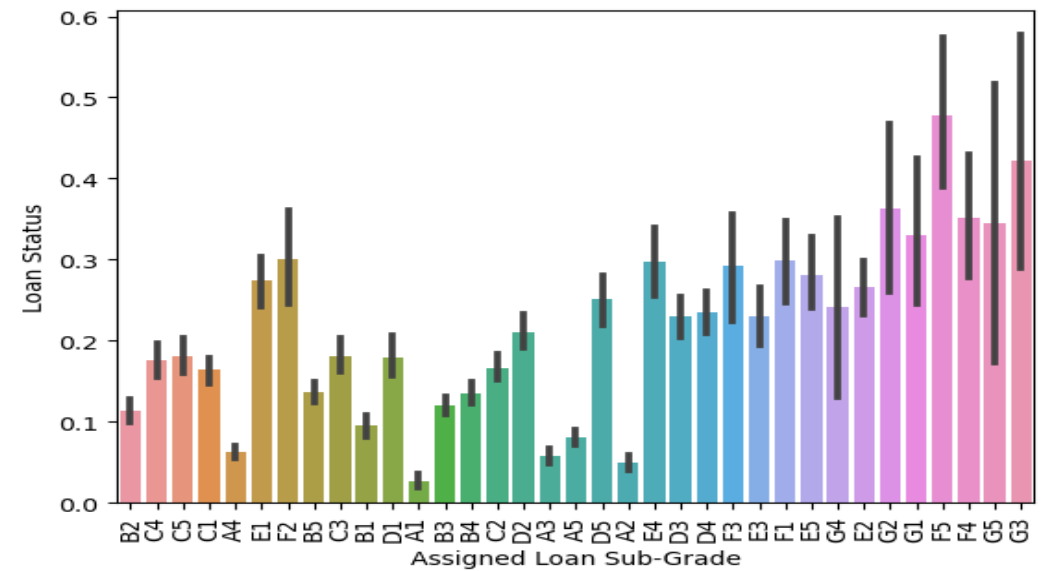
Data Identification and Cleansing

```
import pandas as pd
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np

df = pd.read_csv('loan.csv', low_memory=False)

df.columns
```

```
Index(['id', 'member_id', 'loan_amnt', 'funded_amnt', 'funded_amnt_inv',
      'term', 'int_rate', 'installment', 'grade', 'sub_grade',
      ...,
      'num_tl_90g_dpd_24m', 'num_tl_op_past_12m', 'pct_tl_nvr_dlq',
      'percent_bc_gt_75', 'pub_rec_bankruptcies', 'tax_liens',
      'tot_hi_cred_lim', 'total_bal_ex_mort', 'total_bc_limit',
      'total_il_high_credit_limit'],
      dtype='object', length=111)
```



Data Identification and Cleansing

- 1. A total of 39717 rows and 111 columns are present in the dataframe.
- 2. There are 54 columns with all the rows having null values.
- 3. There are 9 columns with only one single value in it.
- 4. There are columns with a lot of null values and only with few distinct values.
- 5. A total of 47 columns have been renamed.
- 6. 7 columns have been cleansed from objects to strings and int/float.

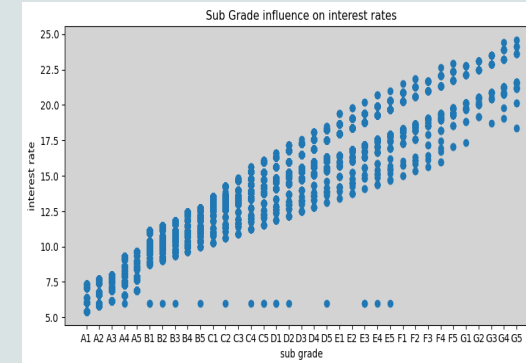
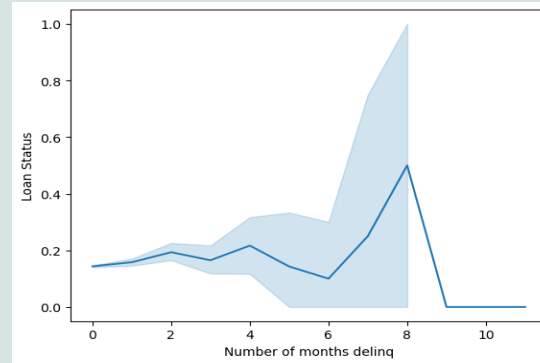
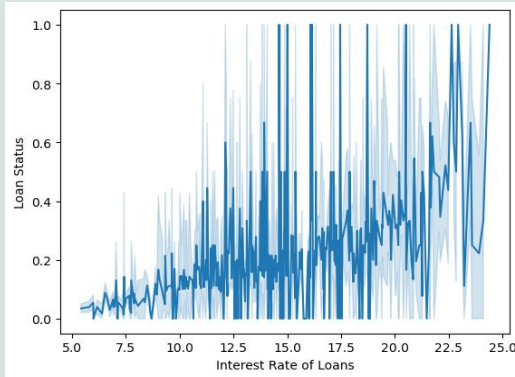
Univariate analysis

1. The univariate analysis revealed interesting insights between the following columns.
 - a. Interest rate (min - 5.42, max - 24.59, heavier above mean)
 - b. Grade (A to G)
 - c. sub grade (Again subgrade by numbers)
 - d. income source verification status (Not verified, verified, Source Verified)
 - e. Loan purpose (debt -consolidation - higher, credit card and other - medium, renewable - lower)
 - f. Loan status (Fully Paid, Current and Charged Off - needed for segmented and bivariate)
 - g. Employment length in years (Values between 0 to 10+ years, most loans given at 4 years)
 - h. Home ownership (Most values for Rent, Mortgage driving the ownership)
 - i. Debt to Income Ratio (Most of the values above the mean value)
2. These are some of the values among other values that we did analysis on using `DataFrame.describe()` which can found in the jupyter notebook

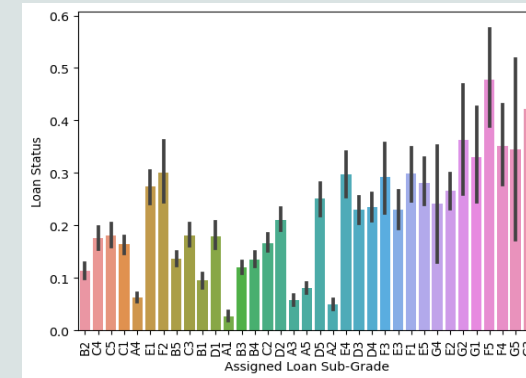
	curr_apprvd_loan_amnt	funded_amnt	total_committed_amnt_by_investors	loan_term_mnth	loan_interest_rate	mnthly_installment	annual_inc	debt_to_income_ratio
count	39717.000000	39717.000000	39717.000000	39717.000000	39717.000000	39717.000000	3.971700e+04	39717.000000
mean	11219.443815	10947.713196	10397.448868	42.418007	12.021177	324.561922	6.896893e+04	13.315130
std	7456.670694	7187.238670	7128.450439	10.622815	3.724825	208.874874	6.379377e+04	6.678594
min	500.000000	500.000000	0.000000	36.000000	5.420000	15.690000	4.000000e+03	0.000000
25%	5500.000000	5400.000000	5000.000000	36.000000	9.250000	167.020000	4.040400e+04	8.170000
50%	10000.000000	9600.000000	8975.000000	36.000000	11.860000	280.220000	5.900000e+04	13.400000
75%	15000.000000	15000.000000	14400.000000	60.000000	14.590000	430.780000	8.230000e+04	18.600000
max	35000.000000	35000.000000	35000.000000	60.000000	24.590000	1305.190000	6.000000e+06	29.990000

8 rows x 25 columns

Univariate segmented analysis

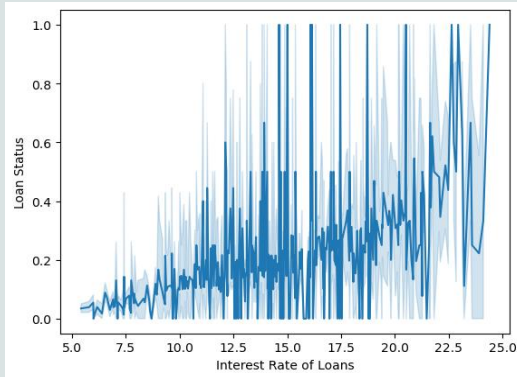


incm_src_verification_status	Not Verified			Source Verified			Verified		
loan_status	Charged Off	Current	Fully Paid	Charged Off	Current	Fully Paid	Charged Off	Current	Fully Paid
assgnd_loan_grade									
A	0.8	0.1	12.2	0.4	0.0	6.0	0.4	0.0	5.5
B	1.6	0.2	11.9	0.9	0.2	6.3	1.1	0.5	7.6
C	1.5	0.2	7.2	0.8	0.2	4.0	1.1	0.3	5.1
D	1.0	0.1	3.8	0.8	0.2	2.5	1.0	0.3	3.7
E	0.4	0.0	1.2	0.5	0.1	1.3	0.9	0.3	2.4
F	0.1	0.0	0.3	0.2	0.0	0.5	0.5	0.1	0.9
G	0.0	0.0	0.1	0.1	0.0	0.1	0.1	0.0	0.3

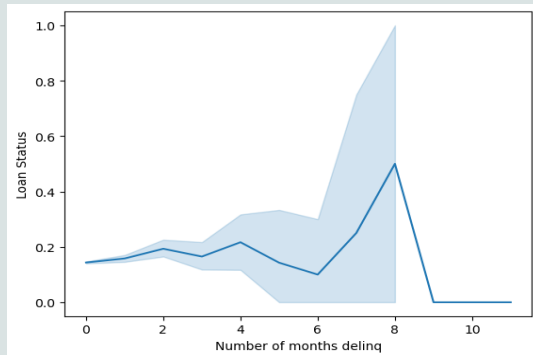


- Segmented univariate analysis is trying to establish a relationship between a single column and another column while segmenting it.
- We have given some relationships that are established between columns

Univariate segmented analysis

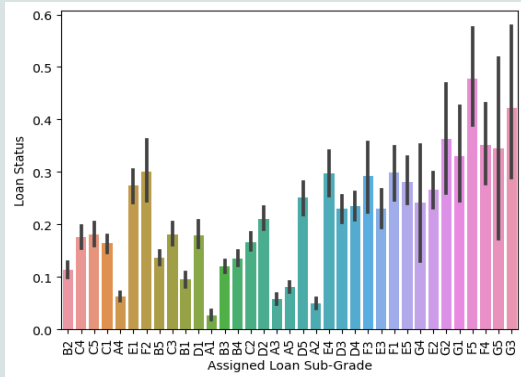


- The interest rate seems to be higher for the loans when we move from current loans to charged off loans

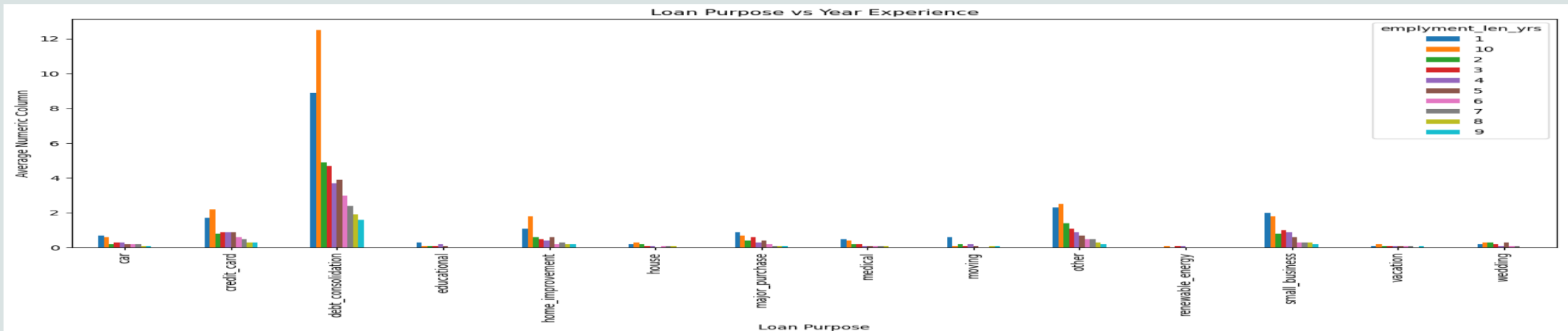


- The number of 30+ months delinq in the last 2 years have between 8 and 10 as we move towards Charged Off loans

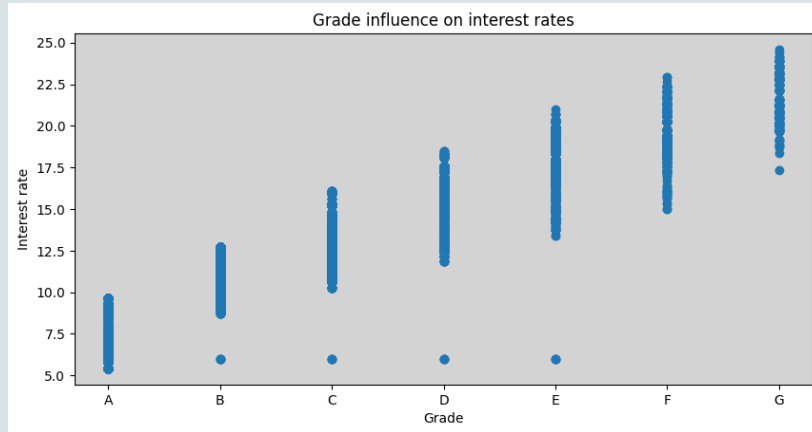
Univariate segmented analysis



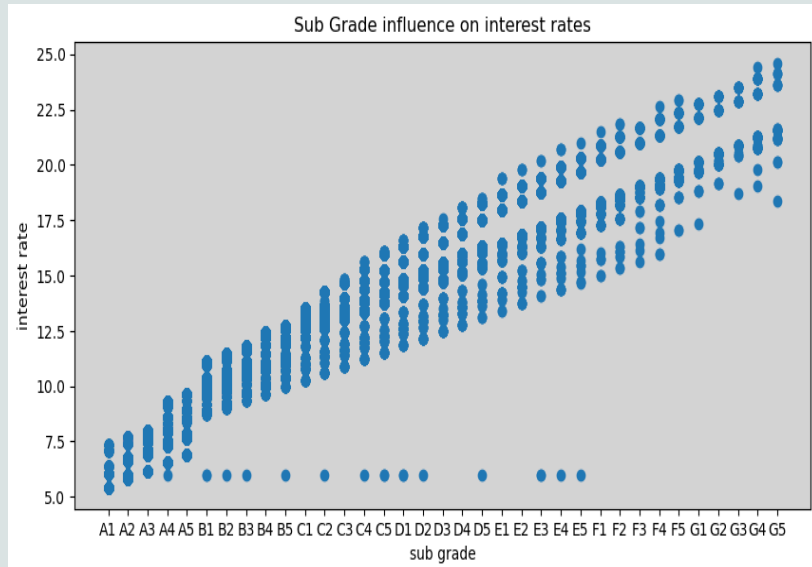
- The interest rates which have higher rate of interest tend to have more Charged Off loans.
- The loan purpose and number of years of experience has a definite relationship. Debt consolidation tops the list with credit card and other come next but with lesser significance



Univariate segmented analysis

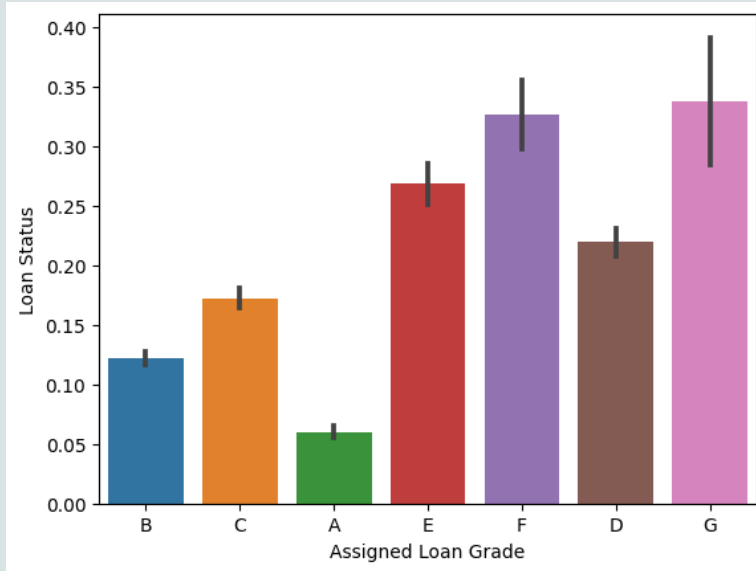


- The loan interest rate is inversely proportional to the assigned loan grade and hence, also the sub-grade.
- Interest Rates increase as the loan grades decrease.

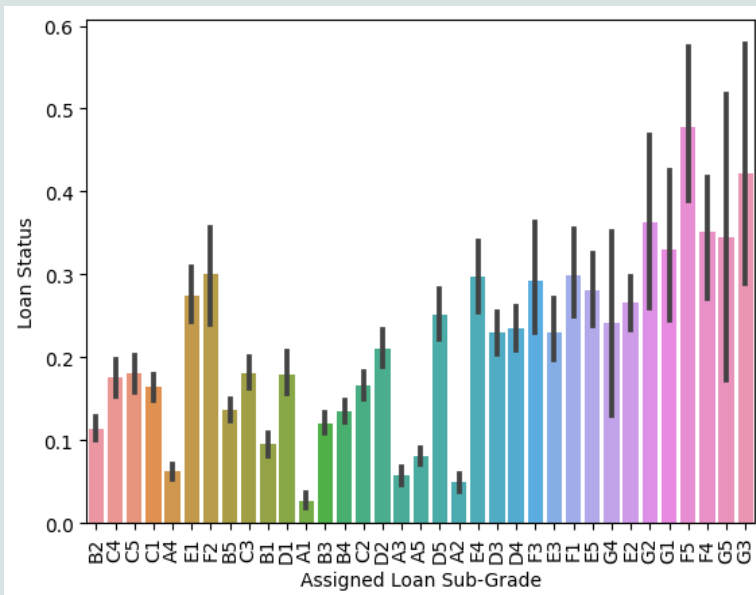


- Looking at the sub-grade divisions, there is not much increment in the interest rate values when compared to the increment between grades.
- This graph shows the almost straight line pattern of increment in the interest rates against the assigned loan sub-grades.

Univariate Segmented Analysis

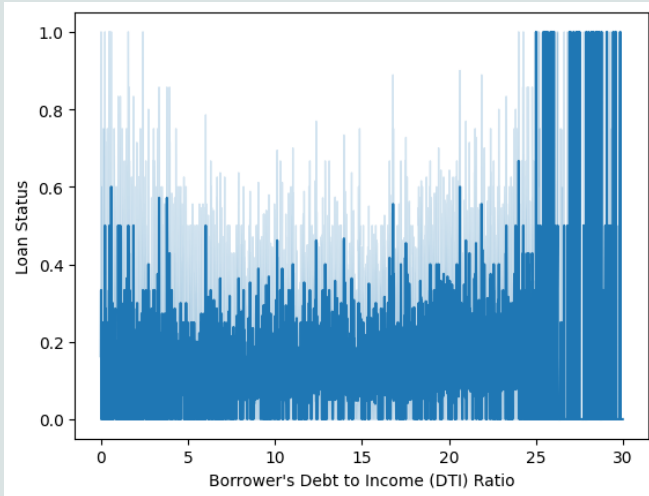


- Lower the assigned grade, higher the tendency of the account defaulting and getting charged-off.
- Both the graphs depict the relation between the assigned grades and the tendency of the borrower of defaulting.

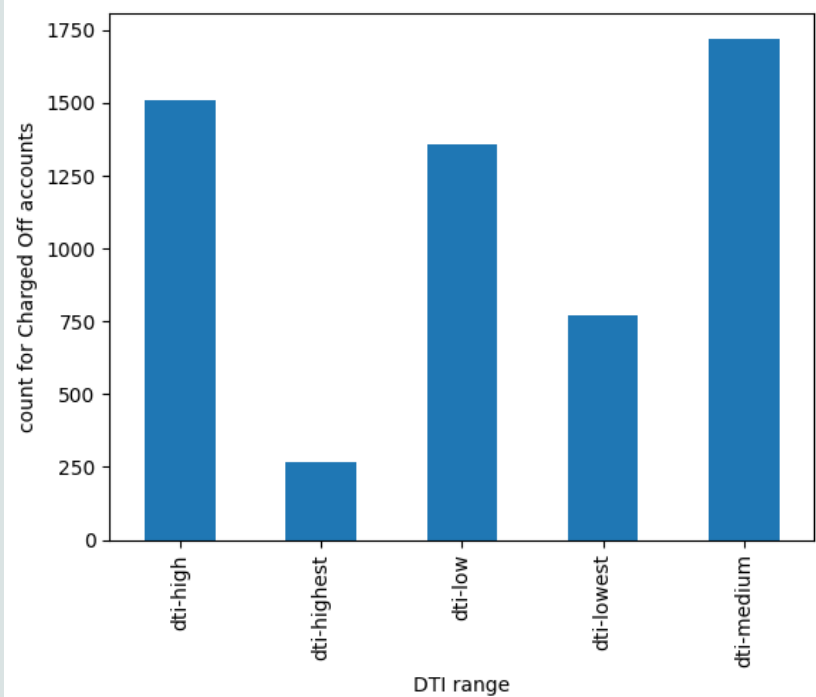


- Lower the assigned sub-grade, higher the tendency of the account defaulting and getting charged-off.

Univariate Segmented Analysis

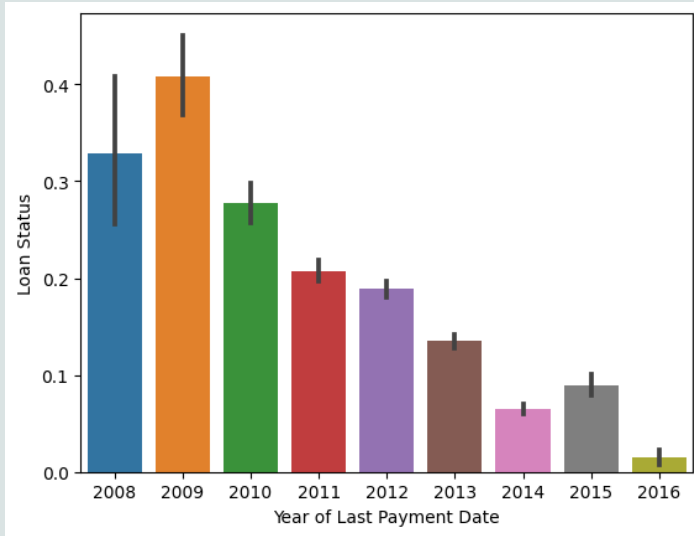


- This lineplot shows that greater the debt-to-income ratio, greater the tendency of the account defaulting and being marked as charged-off.
- The spike in number of accounts getting charged-off can be seen for the debt-to-income ratio values > 25 .

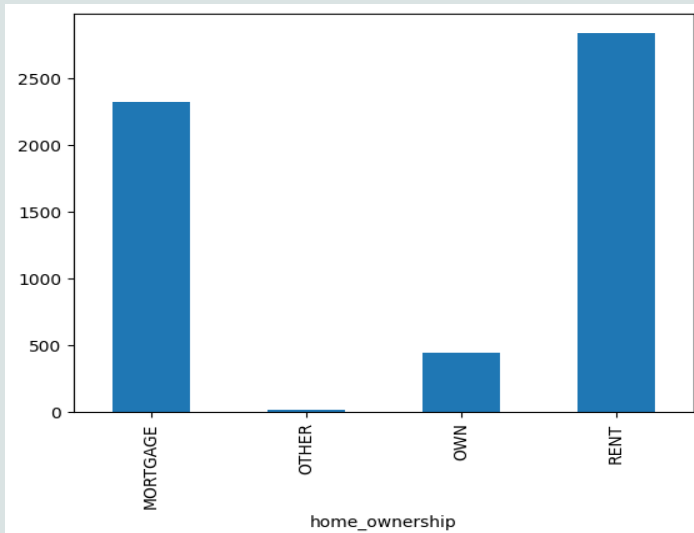


- The graph beside shows the count of Charged-off accounts and the range of debt-to-income ratio they fall under.
- High and Medium DTI marks the majority of the charged-off accounts.

Univariate Segmented Analysis

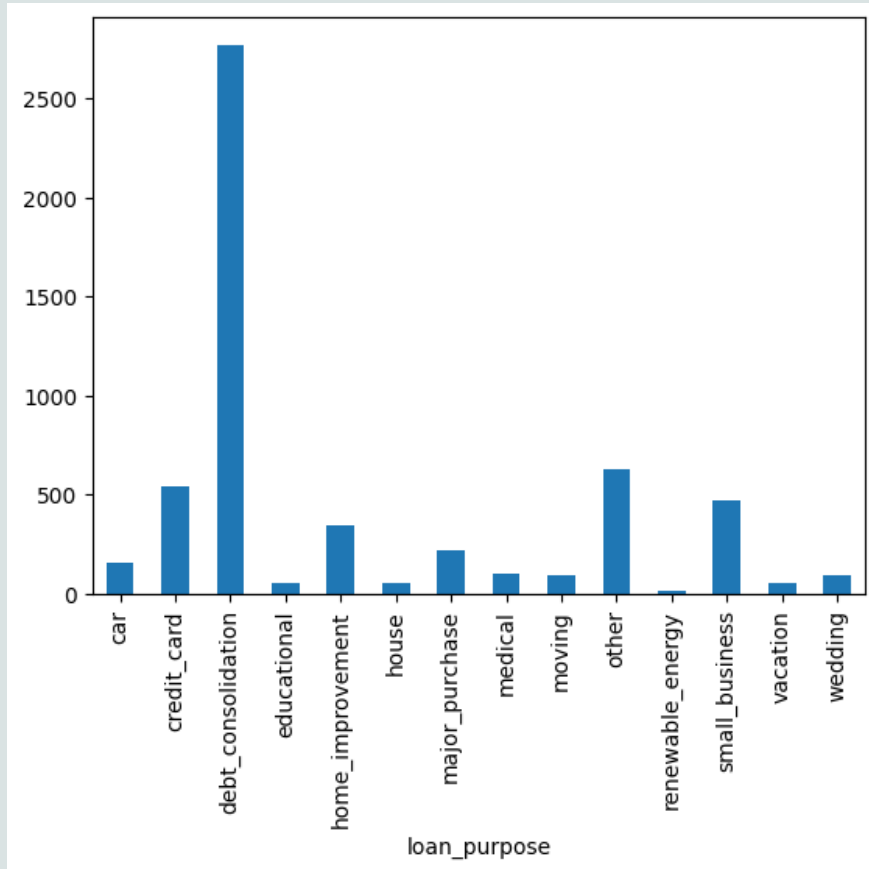


- As you can see in the year-wise graph, there is a steady decrease in accounts being marked as charged-off.
- This goes to say that over the years, there is a decreasing in borrowers going delinquent



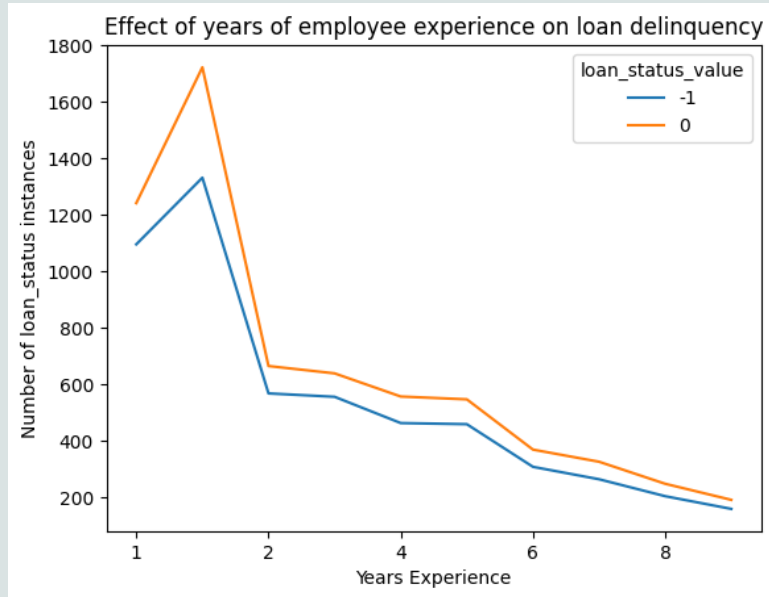
- This graph shows the home ownership segments with the count of accounts that are marked as charged-off
- Majority charged-off accounts belong to borrowers who have mortgaged their houses or are living in rented spaces.
- This goes to show that Home Ownership is a factor that defines charged-off accounts.

Univariate Segmented Analysis



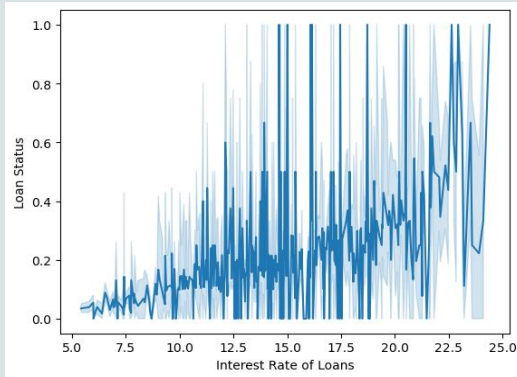
- This graph shows the varied purposes for taking loans, and the count of charged-off accounts for each of the purposes.
- By looking at the graph, you can see that debt consolidation is a major contributor to the number of charged-off accounts.
- This shows that borrowers have existing debts and are now borrowing from the Lending Club to pay off the previous debts.
- In doing so, they are prone to defaulting and thus get charged-off and result in a loss for the Lending Club.

Univariate Segmented Analysis

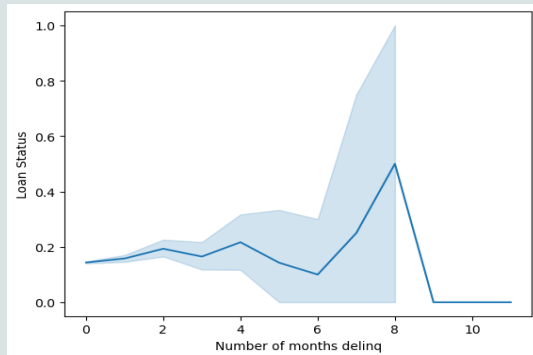


- As you can see more the years of experience, lesser the chance of loan delinquency.
- The blue denotes all the 'Charged-off' accounts while the orange line denotes the 'Current' active loan accounts.

Univariate segmented analysis



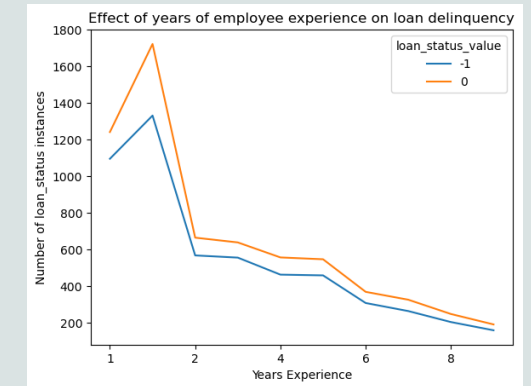
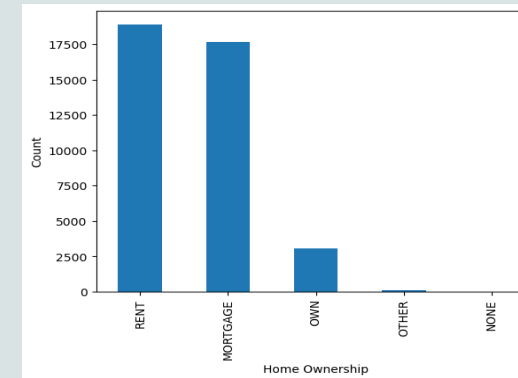
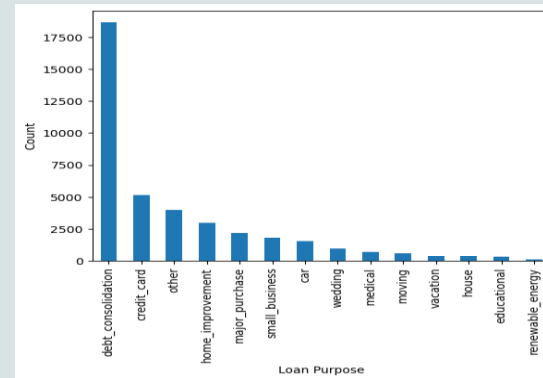
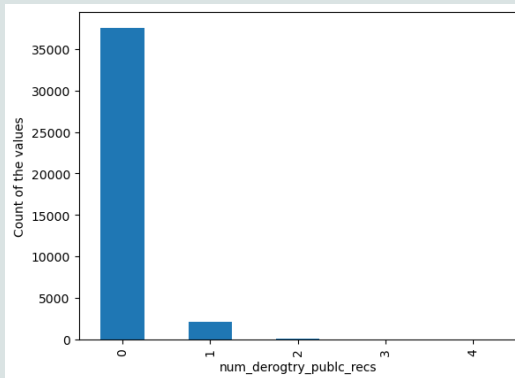
- The interest rate seems to be higher for the loans when we move from current loans to charged off loans



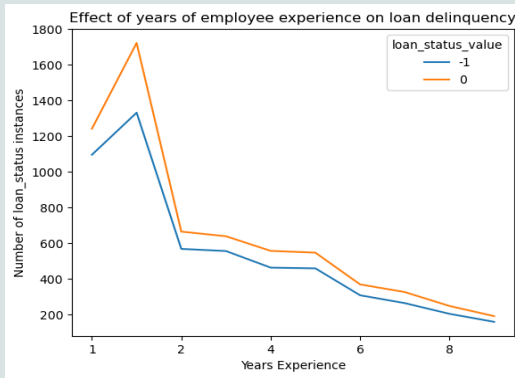
- The number of 30+ months delinq in the last 2 years have between 8 and 10 as we move towards Charged Off loans

Bi and Multi variate analysis

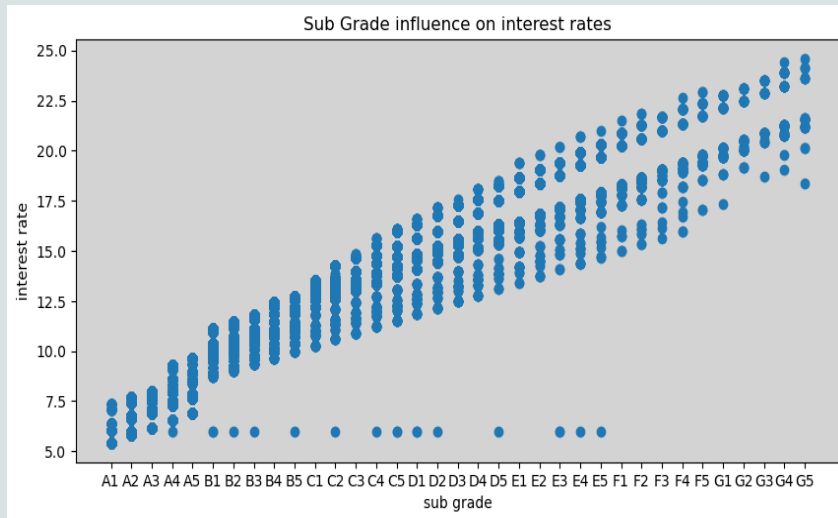
1. Bivariate and multi variate analysis talks about relationships between columns and understanding their correlation. This helps us understand their behavior and influence of each other. Scatter plots and correlations are generally used for this analysis.



Bi and Multi variate analysis

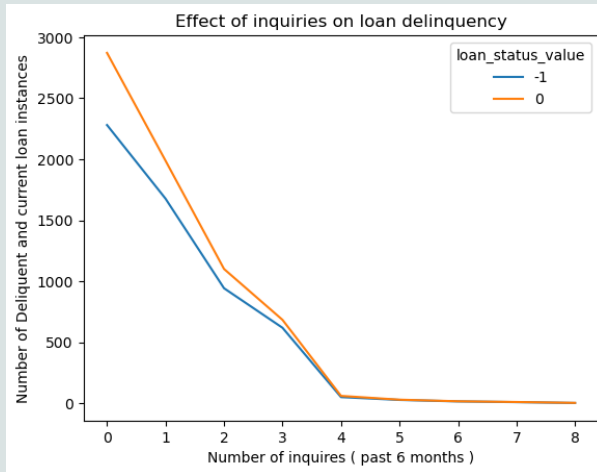


1. This graph correlates the relationship between loan status and the years of experience.
-1 = Charged Off accounts, 0 = current accounts

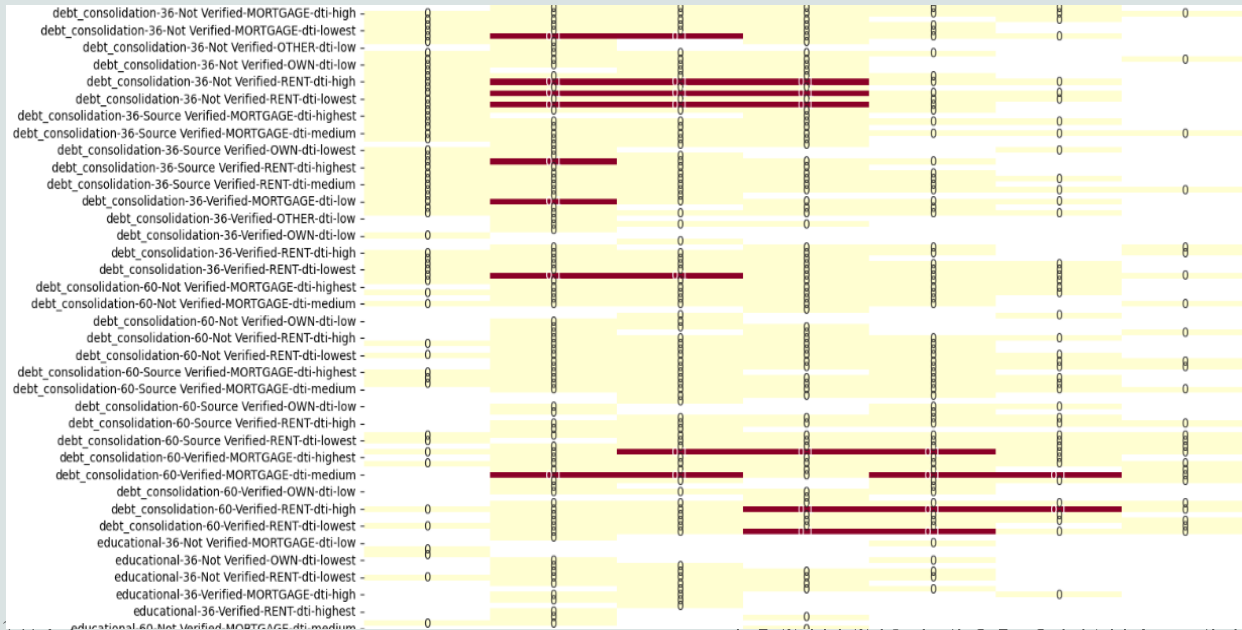


1. The scatter plot gives the relationship between loan grade and interest rate

Bi and Multi variate analysis

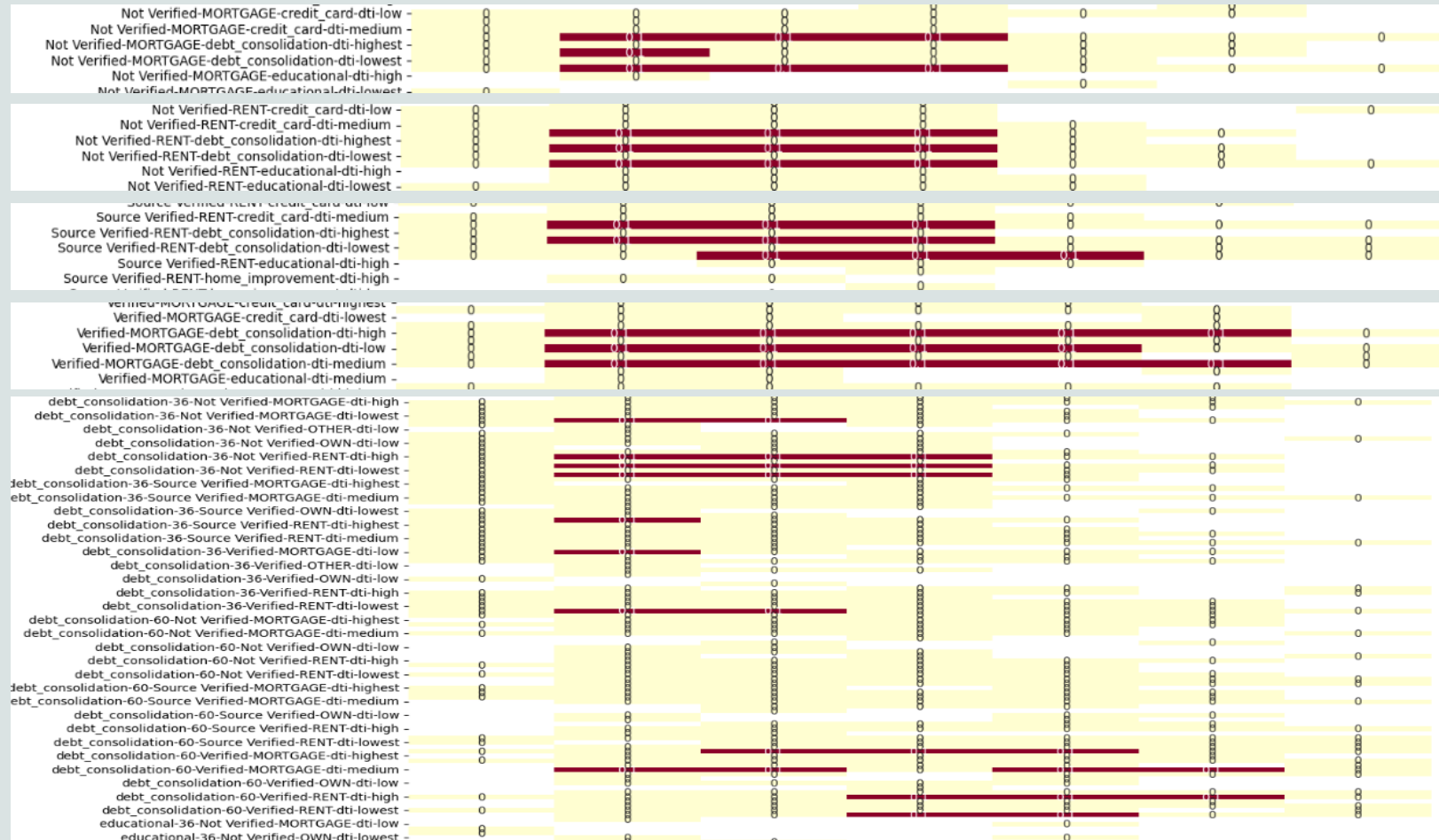


1. This graph talks about the relationship between the delinquent loan instances vs the number of inquiries in the past 6 months



1. The heat map provides how the columns are correlated and some of the ones which are highly correlated. Columns (loan grade, purpose, term, src verification, home ownership, dti)
2. It shows that debt_consolidation, src - not verified, DTI being higher, Rent, Grades - B, C and D are some of the driving factors for Charged Off accounts

Bi and Multi variate analysis



The analysis was done on loan grade, verification status, home ownership, purpose, term and dti. All of them retrieved for Charged Off accounts.

Bi and Multi variate analysis



The analysis was done on loan grade, employment length in years, home ownership, purpose, term and dti. All of them retrieved for Charged Off accounts. This clearly shows a pattern where the charged off accounts have the following -

Employment Length - Low (1 to 2 years)

Loan Grade - B, C and D

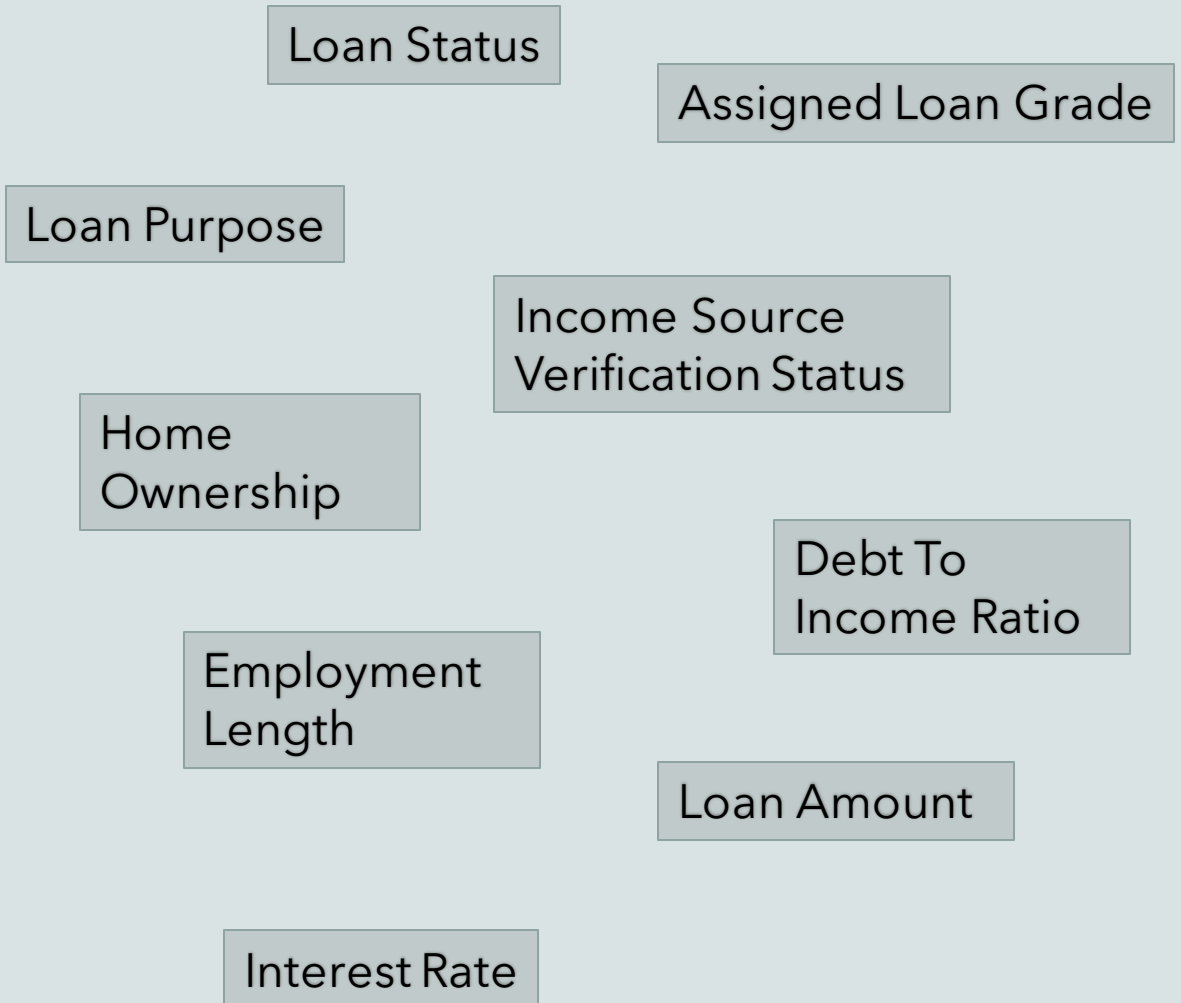
Home Ownership - Rent and Mortgage

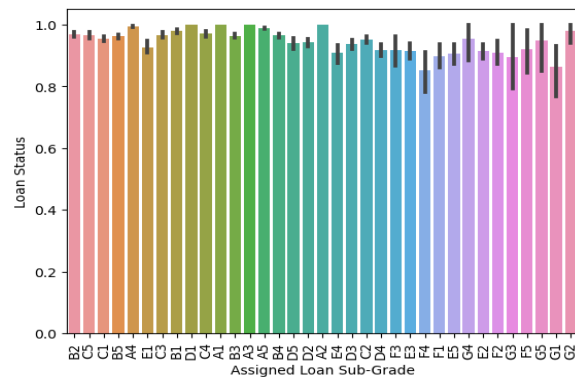
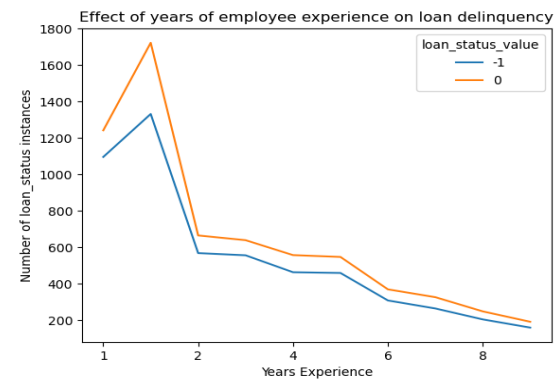
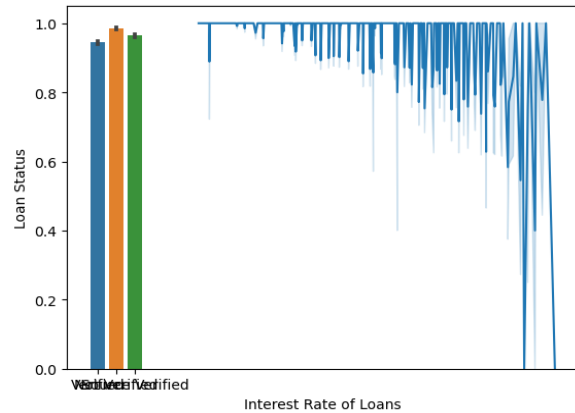
Loan Purpose - Debt-Consolidation and Credit card

DTI - Highest and Medium

The Conclusion

Deciding Factors





Summary

The analysis was done on loan grade, employment length in years, home ownership, purpose, term and dti. All of them retrieved for Charged Off accounts. This clearly shows a pattern where the charged off accounts have the following -

Employment Length - Low (1 to 2 years)

Loan Grade - B, C and D

Home Ownership - Rent and Mortgage

Loan Purpose - Debt-Consolidation and Credit card

DTI - Highest and Medium

Loan Value - The ones which are lowest are getting Charged Off, the others, they are trying to recover.

So, any borrowers, with the following parameters tend to pose a higher risk of delinquency and getting Charged off.

Team



Atharv



Nikhil

Thank you
