

DATA DESCRIPTION REPORT

Dictionary

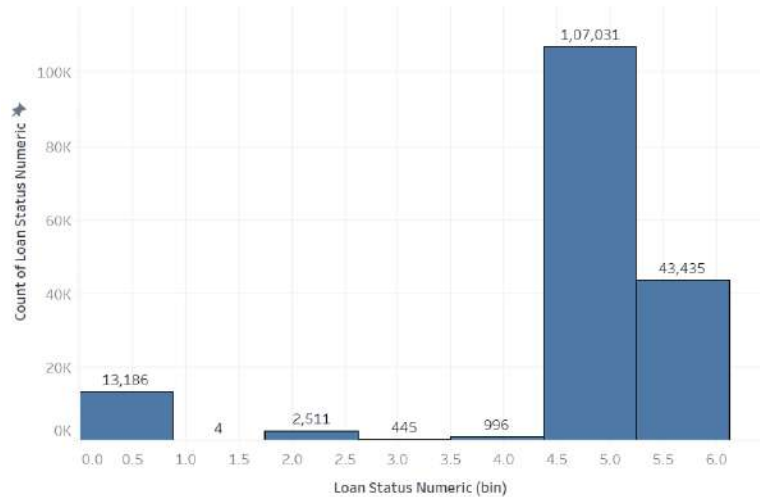
- **Loan status numeric:** It is a rank-ordered numeric version of loan status
- **Loan Amount:** It is the listed amount of the loan applied for by the borrower
- **Term:** It is the number of payments on the loan
- **Int Rate:** It is the interest rate on the loan
- **Emp Length:** It is employment length in years
- **Home Ownership:** It is the homeownership status provided by the borrower
- **Annual Income:** It is the self-reported annual income provided by the borrower
- **Verification Status:** It was income verified by LC, the source, or not verified
- **Months Since Last Deline:** It is the number of months since the borrower's last delinquency
- **Open Account:** It is the number of open credit lines in the borrower's credit file
- **DTI:** It is a ratio calculated using the borrower's total monthly payments on the total debt obligations, excluding mortgages and the requested LC loan, divided by the borrower's combined self-reported monthly income

Univariate Properties

Feature	Variable Type	data Type	Count	Missing	Unique	Min	Q1	Med	Q3	Max	Mode	Mean	Std	Skewness	Kurtosis
Loan Status Numeric	Label	Integer	167608	0	7	0	6	5	5	6	5	4.81	1.53	-2.42	4.95
Loan Amount	Feature	Integer	167608	0	1519	1000	8000	12600	20000	40000	10000	15229.69	9678	0.81	-0.15
Term	Feature	String		0	2	36									
Int Rate	Feature	float	167608	0	217	5.31	9.44	11.99	15.99	30.99	11.99	13.03	5.13	0.93	0.84
Employee Length	Feature	String		12575											
Home Ownership	Feature	String		0	4										
Annual Inc	Feature	float	167608	0	13202	0	46992	66000	95000	110000000	60000	80290	280929	358.56	139866.36
Verification Status	Feature	String		0	3										
Months since last decline	Feature	float	81601	140699	133	0	16	32	50	179	0	34.81	21.9	0.47	-0.65
Open account	Feature	float	167608	0	67	0	8	11	15	101	9	11.65	5.9	1.34	3.64
DTI	Feature	float	167402	206	6248	0	11.9	18	24.9	999	0	19.3	16	26.33	1369.37

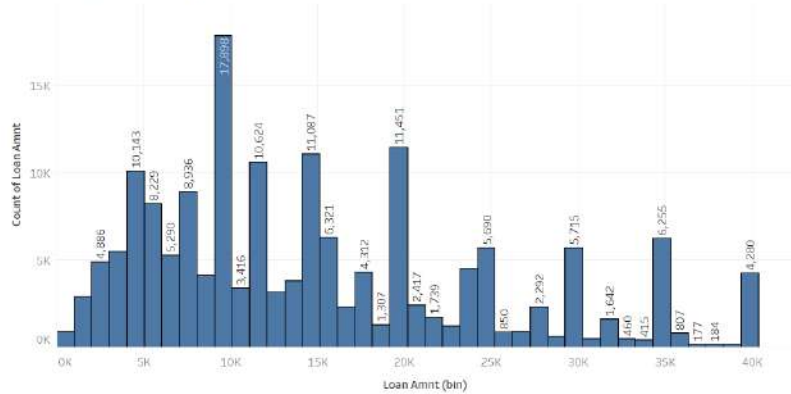
Univariate Visualizations

Histogram - Loan Status Numeric



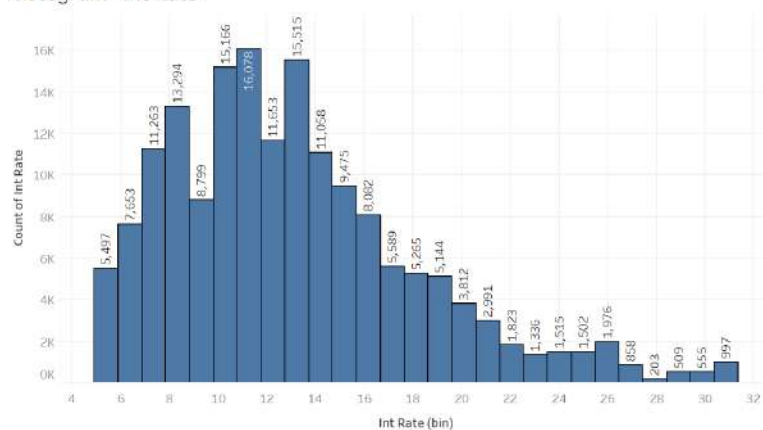
The trend of count of Loan Status Numeric for Loan Status Numeric (bin).

Histogram - Loan Amount



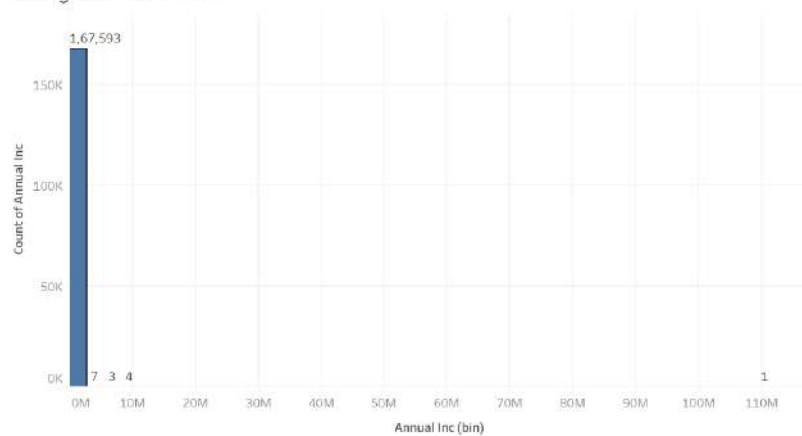
The trend of count of Loan Amnt for Loan Amnt (bin).

Histogram - Int Rate



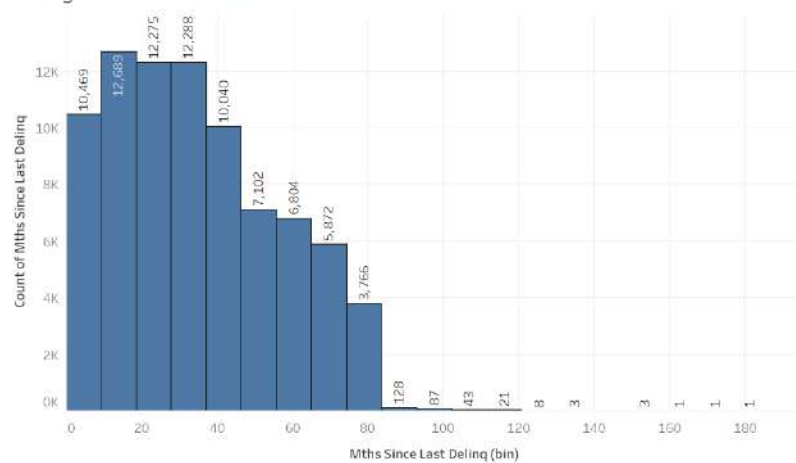
The trend of count of Int Rate for Int Rate (bin).

Histogram - Annual Inc



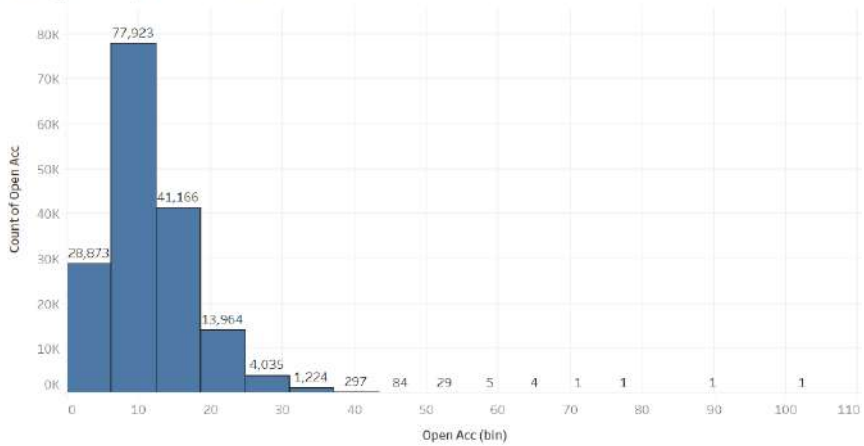
The trend of count of Annual Inc for Annual Inc (bin).

Histogram - Months since last decline



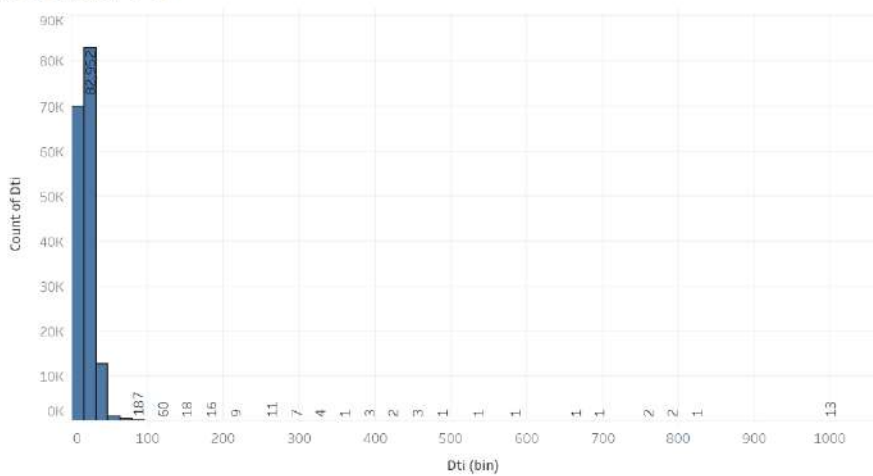
The trend of count of Mths Since Last Delinq for Mths Since Last Delinq (bin).

Histogram - Open account



The trend of count of Open Acc for Open Acc (bin).

Histogram - DTI



The trend of count of DTI for DTI (bin).

DATA EXPLORATION REPORT

This report details the relationship between each potential feature with the label “Loan Status Numeric”

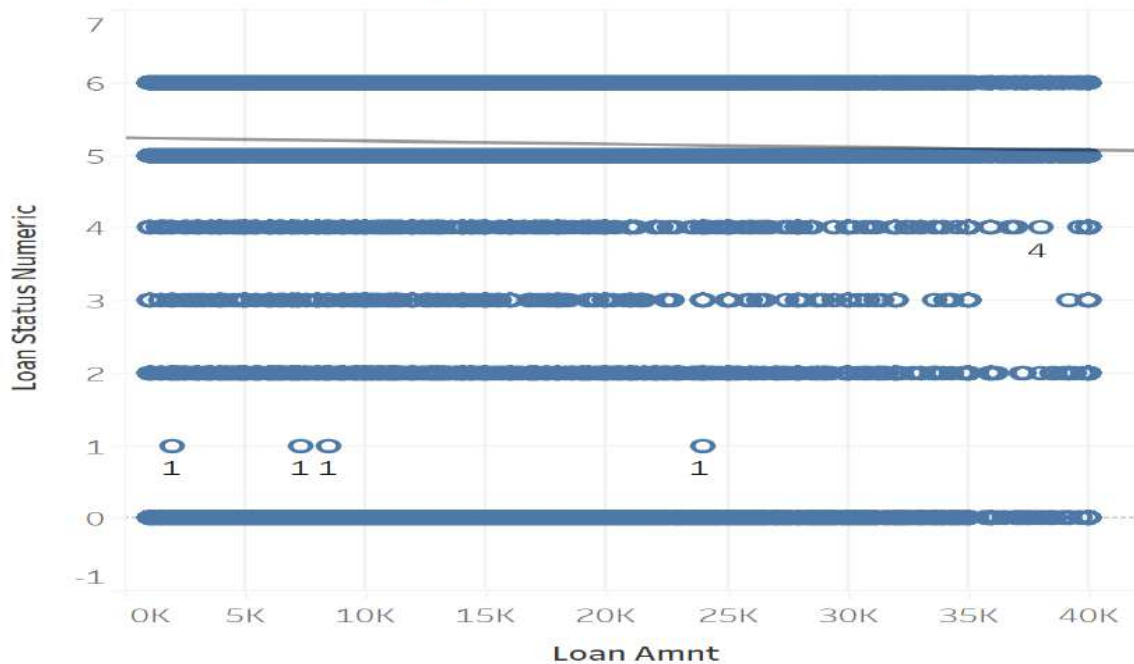
Summary Table

Feature	Analysis	Effect Size	P-value
Loan Amount	R squared	0.0011232	< 0.0001
Term	F - stat	720.54	< 0.00001
Int Rate	R squared	0.0315754	< 0.0001
Home Ownership	F - stat	192.872	< 0.0001
DTI	R squared	0.0023268	< 0.0001

The remainder of the report includes greater details on each relationship. We find that all five features are worth including during the modeling phase.

Loan Amount

Scatter - Loan Amount



Loan Amnt vs. Loan Status Numeric.

R-Squared: 0.0027743

P-value: < 0.0001

Equation: $\text{Loan Status Numeric} = -0.082868 \cdot \ln(\text{Loan Amnt}) + 5.58837$

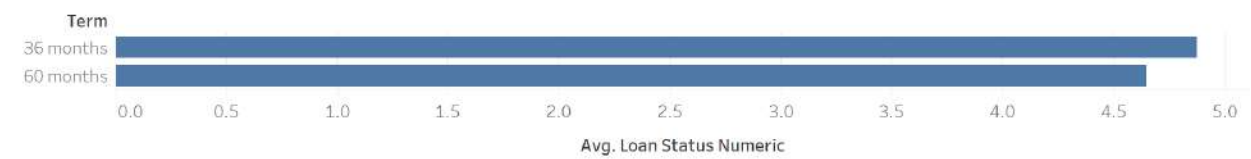
Coefficients

Term	Value	StdErr	t-value	p-value
$\ln(\text{Loan Amnt})$	-0.082868	0.005083	-16.303	< 0.0001
intercept	5.58837	0.0479207	116.617	< 0.0001

A linear trend model is computed for natural log of Loan Status Numeric Given Loan Amnt. The model may be significant at $p \leq 0.05$.

Term

Term - Bar



Average of Loan Status Numeric for each Term.

Mean of 36 Months: 4.8718

Mean of 60 Months: 4.6486

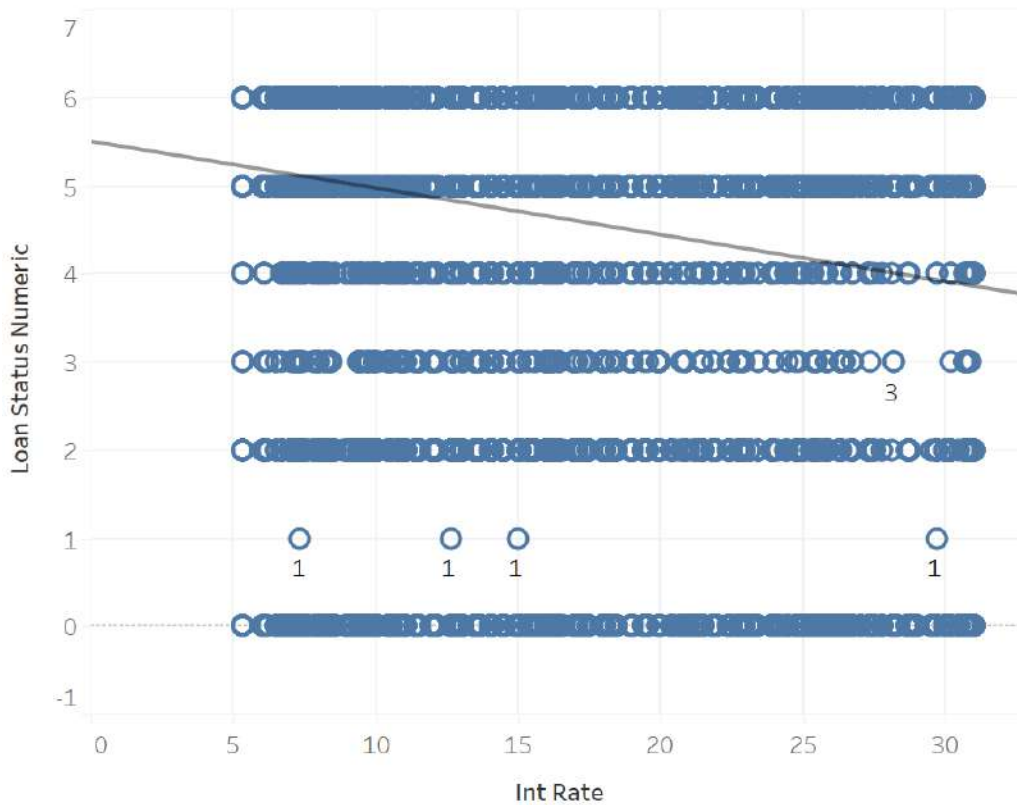
One Way ANOVA F: 720.539420839037

P – value: 2.21643750887698E-158

SUMMARY						
Groups	Count	Sum	Average	Variance		
5	120812	588572	4.871801	2.226341		
5	46794	217528	4.64863	2.602842		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1679.905	1	1679.905	720.5394	2.22E-158	3.841514
Within Groups	390761.2	167604	2.331455			
Total	392441.1	167605				

INT Rate

INT rate - Scatter



Int Rate vs. Loan Status Numeric.

R-Squared: 0.0315754

P-value: < 0.0001

Equation: $\text{Loan Status Numeric} = -0.0530124 \times \text{Int Rate} + 5.50048$

Coefficients

Term	Value	StdErr	t-value	p-value
Int Rate	-0.0530124	0.0007171	-73.9241	< 0.0001
intercept	5.50048	0.0100448	547.594	< 0.0001

A linear trend model is computed for natural log of Loan Status Numeric given Int Rate. The model may be significant at $p \leq 0.05$.

Home Ownership

Home Ownership- Bar



Average of Loan Status Numeric for each Home Ownership.

Mean of Any: 4.7087

Mean of Mortgage: 4.8971

Mean of Own: 4.7992

Mean of Rent: 4.7047

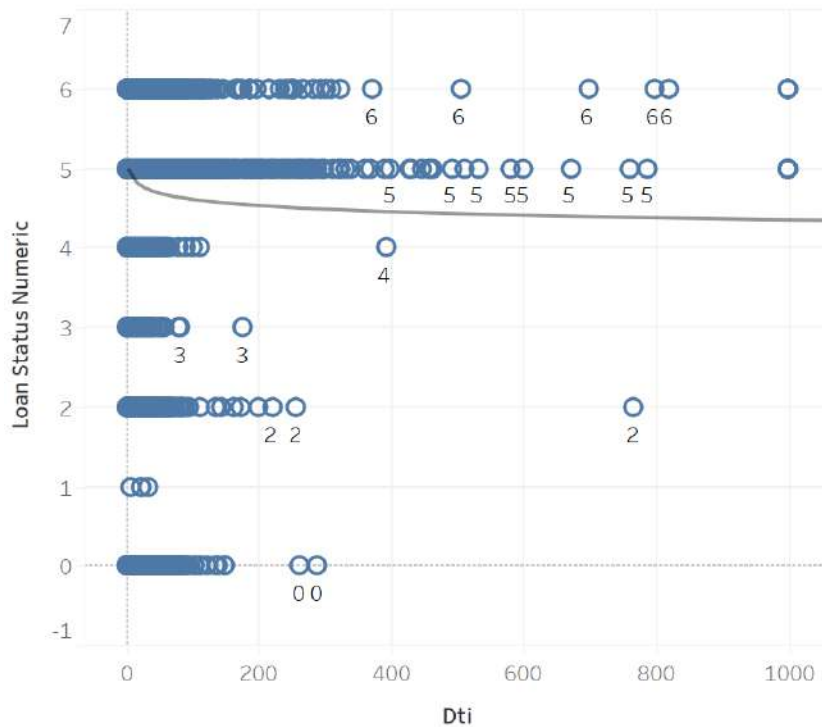
One Way ANOVA F: 192.872066323683

P – value: 7.15507947794915E-125

SUMMARY						
Groups	Count	Sum	Average	Variance		
5	126	593	4.706349	2.369079		
6	81331	398281	4.897038	2.065681		
5	20332	97577	4.799184	2.395803		
6	65815	309637	4.704657	2.645047		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1350.179	3	450.0595	192.8721	7.16E-125	2.604962
Within Groups	391088.1	167600	2.333461			
Total	392438.3	167603				

DTI

DTI - Scatter



Dti vs. Loan Status Numeric.

R-Squared: 0.0023268

P-value: < 0.0001

Equation: $\text{Loan Status Numeric} = -0.111685 \cdot \ln(\text{Dti}) + 5.11962$

Coefficients

Term	Value	StdErr	t-value	p-value
ln (Dti)	-0.111685	0.0056552	-19.7491	< 0.0001
intercept	5.11962	0.016149	317.024	< 0.0001

A linear trend model is computed for Loan Status Numeric given natural log of Dti. The model may be significant at $p \leq 0.05$.