

GRAMENER CASE STUDY SUBMISSION

Group Name:

- 1.Subhanshu Rathi
- 2.Pavan ML
- 3.Ravi Shekhar Rai
- 4.Srikanth K

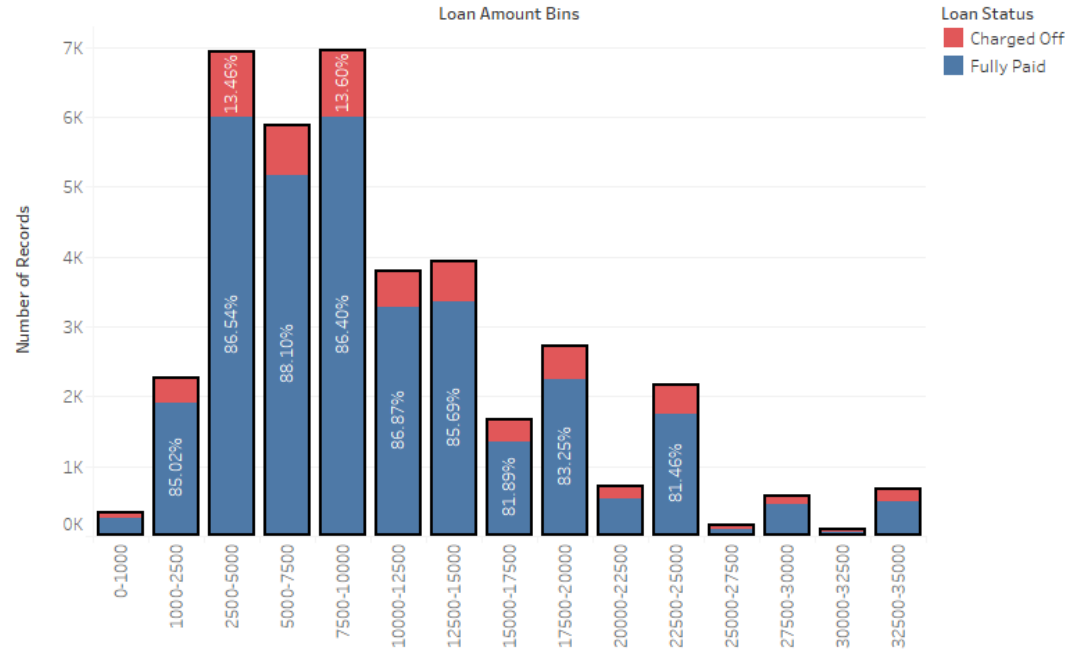
- Case study is to provide the company, Lending club about the factors they have to consider before approving the loans for addressing the following risks involved during their loan processing
 - If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to 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 financial loss for the company
- Following analysis are done on various consumer attributes and loan attributes
 - Univariate
 - Bivariate
 - Multivariate
 - Correlation between continuous variables
- Used Graphs to show the Observations

DATA CLEANSING & DATA PREPARATION

S.No	Category	Activity
1	Date Column Handling	<ul style="list-style-type: none"> Date has been considered to be in YYYY-MM-DD format. Default day (01) added for the 'earliest_cr_line' and 'last_credit_pull_d' columns. <p>Ex: Jan-16 converted to 2016-01-01</p>
2	Single Valued Column	<ul style="list-style-type: none"> Removed the column from analysis where the entire column has single value across the dataset.
3	NA Column Handling	<ul style="list-style-type: none"> Columns having only NA values are removed from analysis. Columns having NA values more than 50% are removed from analysis.
4	' emp_length ' Column Handling	<ul style="list-style-type: none"> emp_length columns values are considered as mentioned below based on inputs from Data Dictionary excel. <ul style="list-style-type: none"> Value is "n/a" or <1 year – Considered as 0 Year. Value is 10+ year – Considered as 10 Year.
5	Redundant Columns Handling	<ul style="list-style-type: none"> Removed the redundant columns from analysis. Redundant columns identified are "sub_grade" and "zip_code".
6	No Value Addition Columns	<ul style="list-style-type: none"> Removed the columns from analysis which adds no value for analysis. Columns identified for this category are "url", "desc", "id", "member_id", "funded_amnt", "funded_amnt_inv", "issue_d", "out_prncp", "out_prncp_inv", "total_pymnt", "total_pymnt_inv", "total_rec_prncp", "total_rec_int", "total_rec_late_fee", "recoveries", "collection_recovery_fee", "last_pymnt_d" and "last_pymnt_amnt".

S.No	Category	Activity														
7	Derived Columns	<ul style="list-style-type: none">• 'annualIncomeBins' column derived from column 'annual_inc' Bin range - "0-5000", "5000-10000", "10000-20000", "20000-30000", "30000-40000", "40000-50000", "50000-60000", "60000-70000", "70000-80000", "80000-90000", "90000-100000", "100000-125000", "125000-150000", "150000-175000", "175000-200000"• 'loanAmountBins' column derived from column 'loan_amnt' Bin range - "0-1000", "1000-2500", "2500-5000", "5000-7500", "7500-10000", "10000-12500", "12500-15000", "15000-17500", "17500-20000", "20000-22500", "22500-25000", "25000-27500", "27500-30000", "30000-32500", "32500-35000"• 'intRateBins' column derived from column 'int_rate' Bin range - "0-5", "5-7.5", "7.5-10", "10-12.5", "12.5-15", "15-17.5", "17.5-20", "20-22.5", "22.5-25"• 'revolUtilRateBins' column derived from column 'revol_util' Bin range - "0-5", "5-10", "10-15", "15-20", "20-25", "25-30", "30-35", "35-40", "40-45", "45-50", "50-55", "55-60", "60-65", "65-70", "70-75", "75-80", "80-85", "85-90", "90-95", "95-100"														
8	NA Values Handling	<ul style="list-style-type: none">• Replaced the values of NA in following columns as mentioned below <table><thead><tr><th>COLUMN NAME</th><th>NA VALUE REPLACED WITH</th></tr><tr><th>=====</th><th>=====</th></tr></thead><tbody><tr><td>emp_title</td><td>Self Employed</td></tr><tr><td>title</td><td>Others</td></tr><tr><td>collections_12_mths_ex_med</td><td>0</td></tr><tr><td>pub_rec_bankruptcies</td><td>0</td></tr><tr><td>tax_liens</td><td>0</td></tr></tbody></table>	COLUMN NAME	NA VALUE REPLACED WITH	=====	=====	emp_title	Self Employed	title	Others	collections_12_mths_ex_med	0	pub_rec_bankruptcies	0	tax_liens	0
COLUMN NAME	NA VALUE REPLACED WITH															
=====	=====															
emp_title	Self Employed															
title	Others															
collections_12_mths_ex_med	0															
pub_rec_bankruptcies	0															
tax_liens	0															

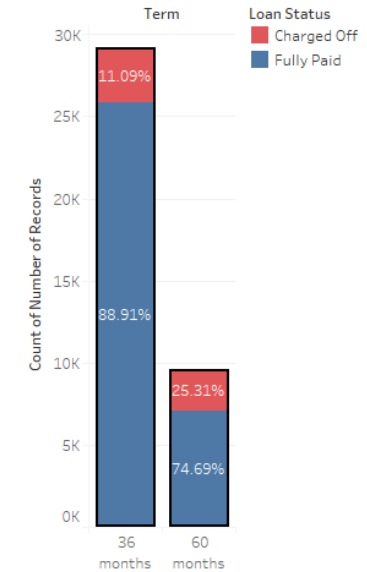
LoanAmountBins



Sum of Number of Records for each Loan Amount Bins. Color shows details about Loan Status. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

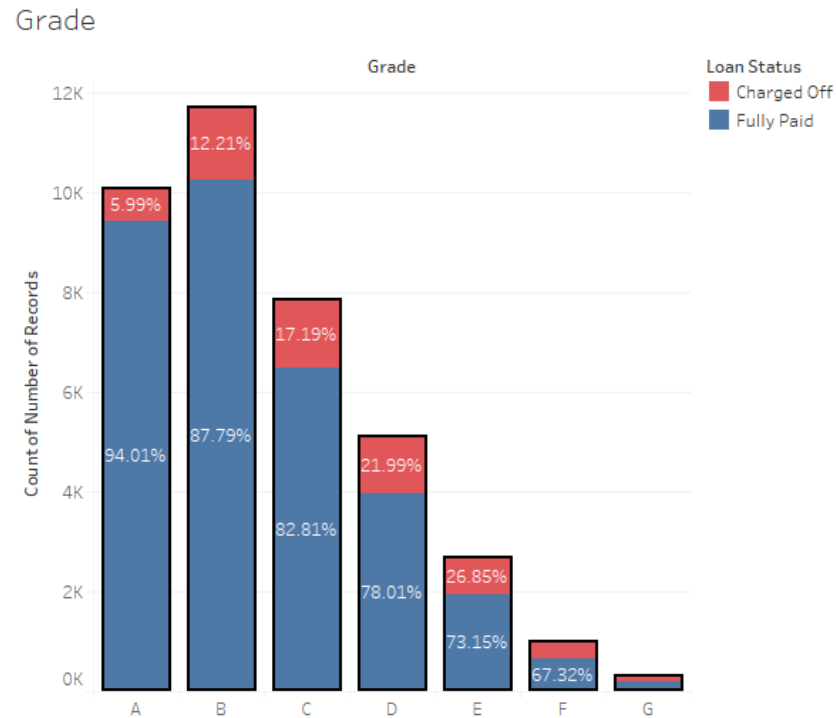
COLUMN	OBSERVATION
loanAmountBins	<ul style="list-style-type: none"> In the bracket 32,500 – 35,000 more number of defaulters observed In the bracket 5000 – 7500 more number of loyal payers observed

Term



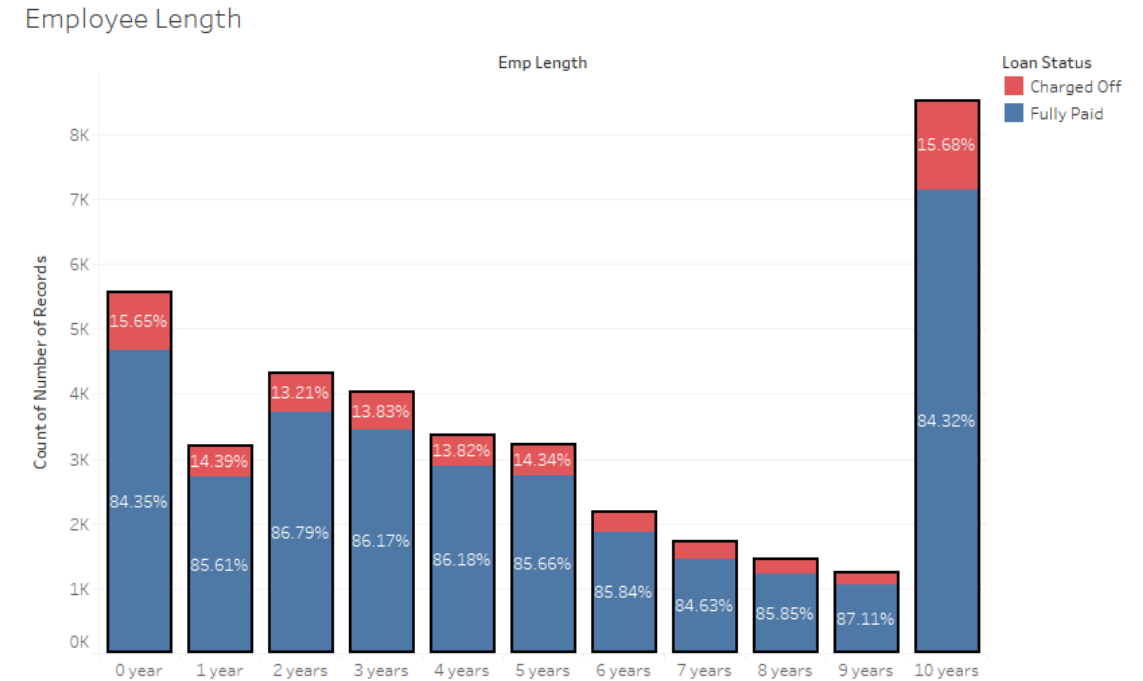
Count of Number of Records for each Term. Color shows details about Loan Status. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

COLUMN	OBSERVATION
term	<ul style="list-style-type: none"> Term 60 months has more defaulters comparatively Term 36 months has more loyal payers comparatively



Count of Number of Records for each Grade. Color shows details about Loan Status. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

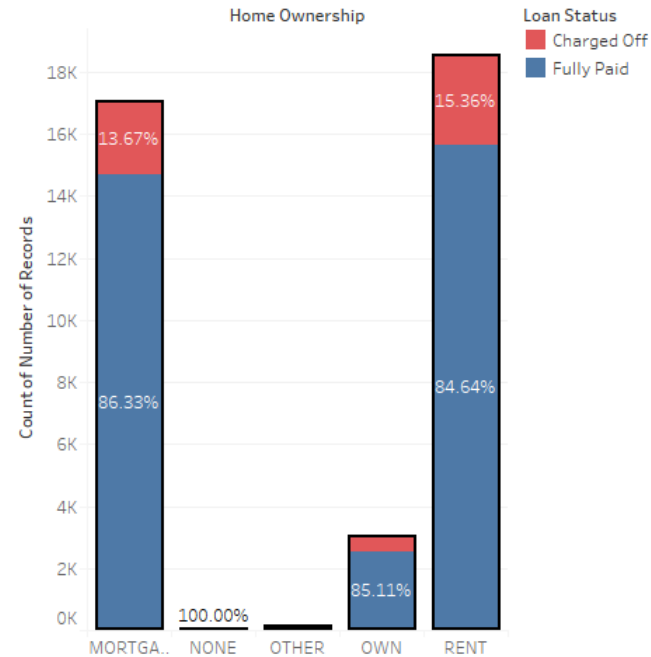
COLUMN	OBSERVATION
Grade	<ul style="list-style-type: none"> Defaulter are increasing from Grades A-G Loyal Payers are decreasing from Grades A-G



Count of Number of Records for each Emp Length. Color shows details about Loan Status. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

COLUMN	OBSERVATION
Employee Length	<ul style="list-style-type: none"> 10 years and less than 1 year applicants are more defaulters comparatively People between 2years to 9 years are almost loyal payers

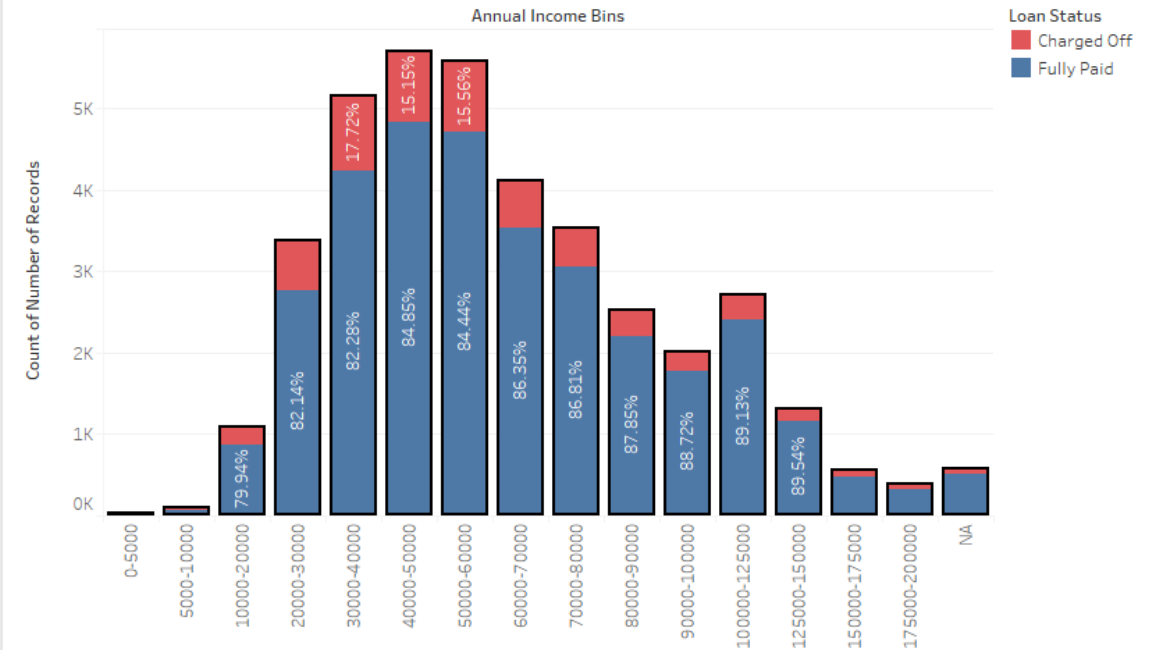
Home Ownership



Count of Number of Records for each Home Ownership. Color shows details about Loan Status. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

COLUMN	OBSERVATION
Home Ownership	<ul style="list-style-type: none"> People having home status as RENT/ MORTGAGE are more defaulters

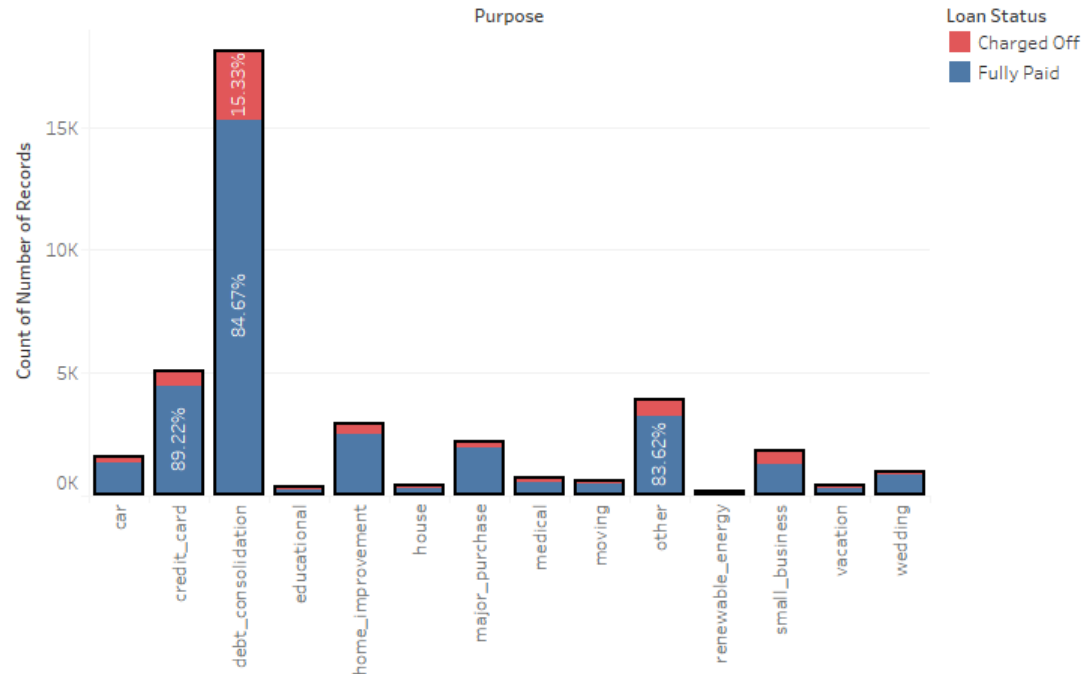
Annual Income Bins



Count of Number of Records for each Annual Income Bins. Color shows details about Loan Status. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

COLUMN	OBSERVATION
Annual Income Bins	<ul style="list-style-type: none"> People having annual incomes in the bracket 20,000 – 40,000 are more defaulters

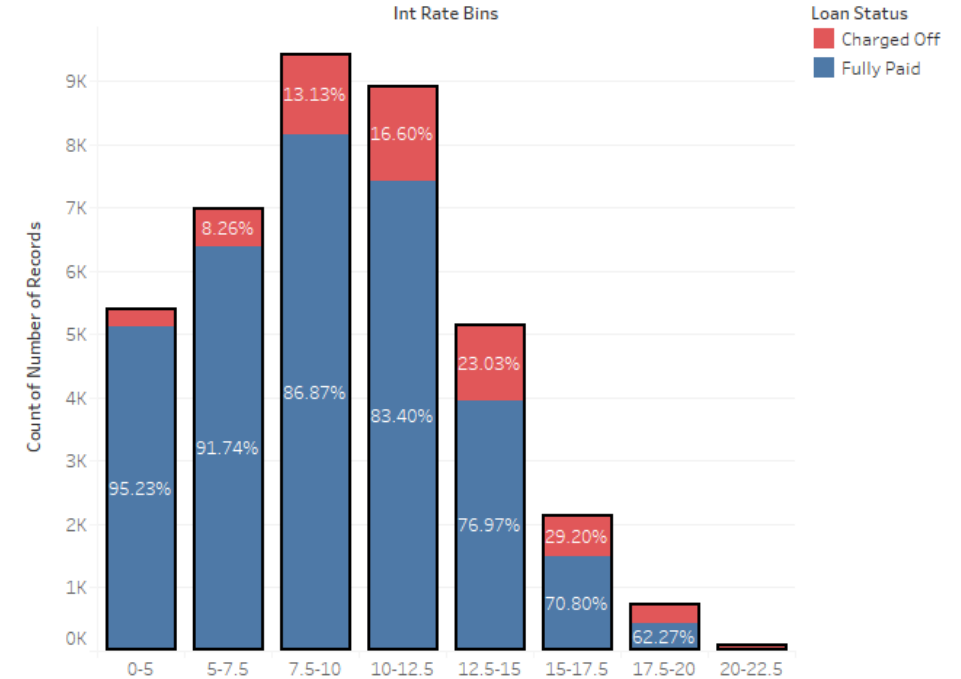
Purpose



Count of Number of Records for each Purpose. Color shows details about Loan Status. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

COLUMN	OBSERVATION
Purpose	<ul style="list-style-type: none"> Debt consolidation are more defaulters

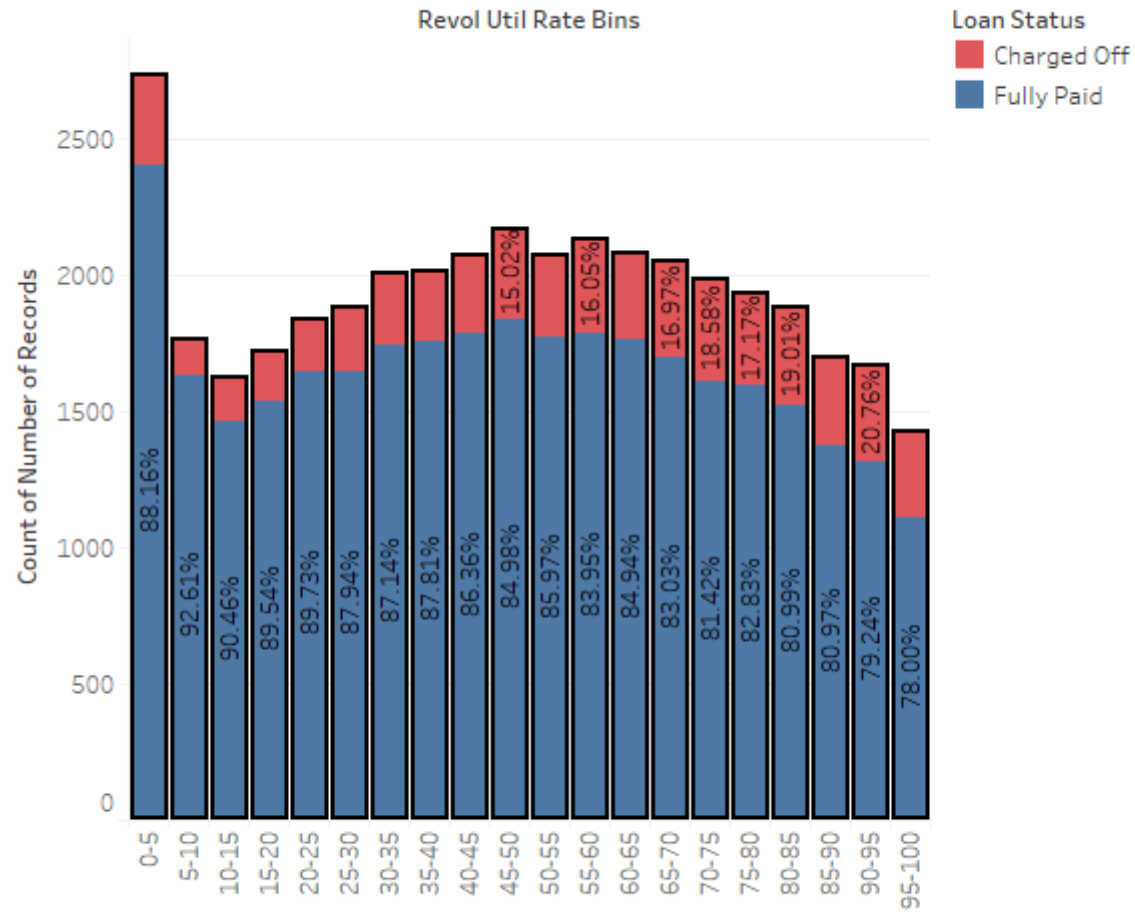
Interest Rate Bins



Count of Number of Records for each Int Rate Bins. Color shows details about Loan Status. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

COLUMN	OBSERVATION
Interest Rate Bins	<ul style="list-style-type: none"> Higher the interest rate more the defaulters

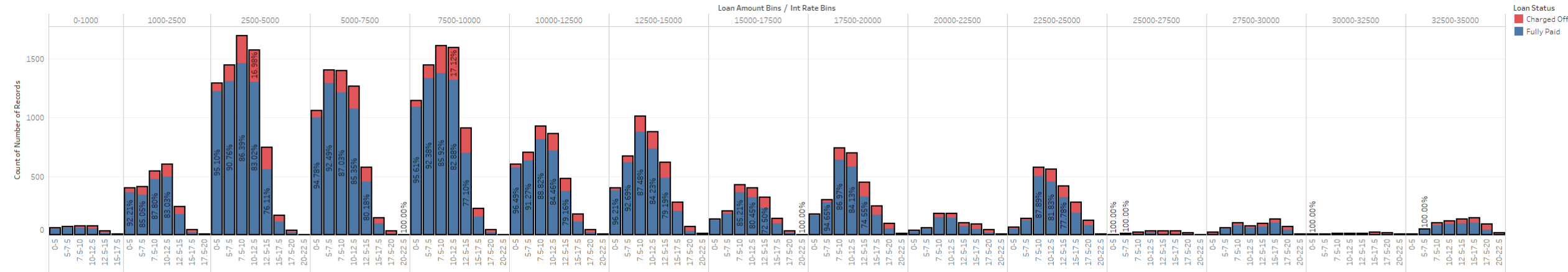
Revol Util Rate Bins



Count of Number of Records for each Revol Util Rate Bins. Color shows details about Loan Status. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

COLUMN	OBSERVATION
Revol Util Rate Bins	<ul style="list-style-type: none"> Higher the Revol Util Rates higher the defaulters

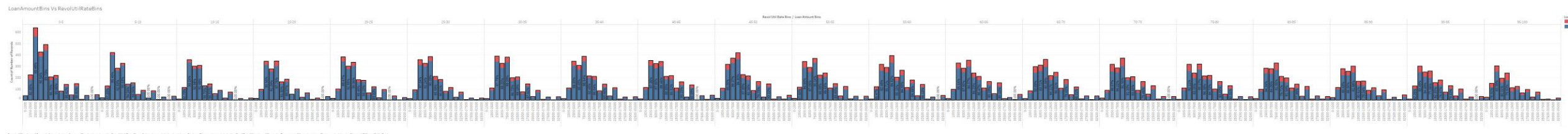
LoanAmountBins Vs InterestBins



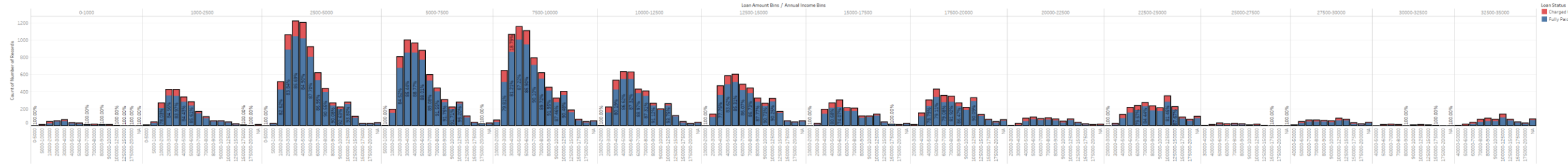
Count of Number of Records for each Int Rate Bins broken down by Loan Amount Bins. Color shows details about Loan Status. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

COLUMN	OBSERVATION
Loan Amount Bins & Interest Rates	<ul style="list-style-type: none"> Higher the interest across all brackets of loan amounts are observed to be more defaulters Lesser the interest across all brackets of loan amount less chances of defaulters

COLUMN	OBSERVATION
Loan Income Bins & Revol Util Rates Bins	<ul style="list-style-type: none"> More the loan amount more the Revol Util rate we see more the defaulters



LoanAmountBins Vs AnnualIncomeBins

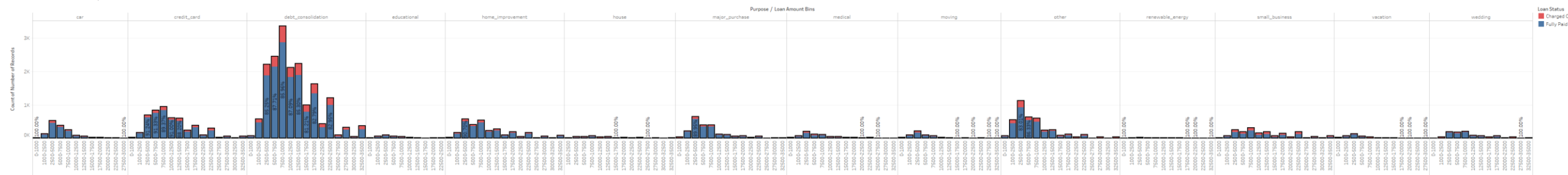


Count of Number of Records for each Annual Income Bins broken down by Loan Amount Bins. Color shows details about Loan Status. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

COLUMN	OBSERVATION
Loan Amount Bins & Annual Income Bins	<ul style="list-style-type: none"> Lesser Annual Income people are observed more defaulters comparatively

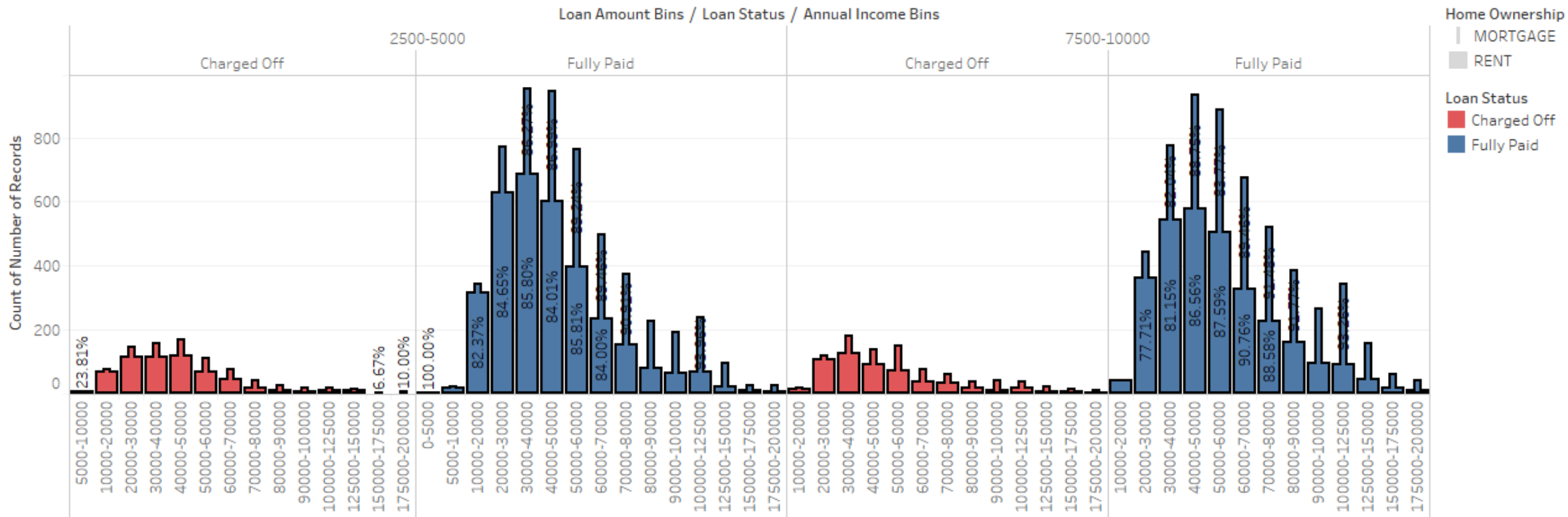
COLUMN	OBSERVATION
Loan Income Bins & Purpose	<ul style="list-style-type: none"> Higher the loan amount in debt consolidation bracket more number of defaulters found.

LoanAmountBins Vs Purpose



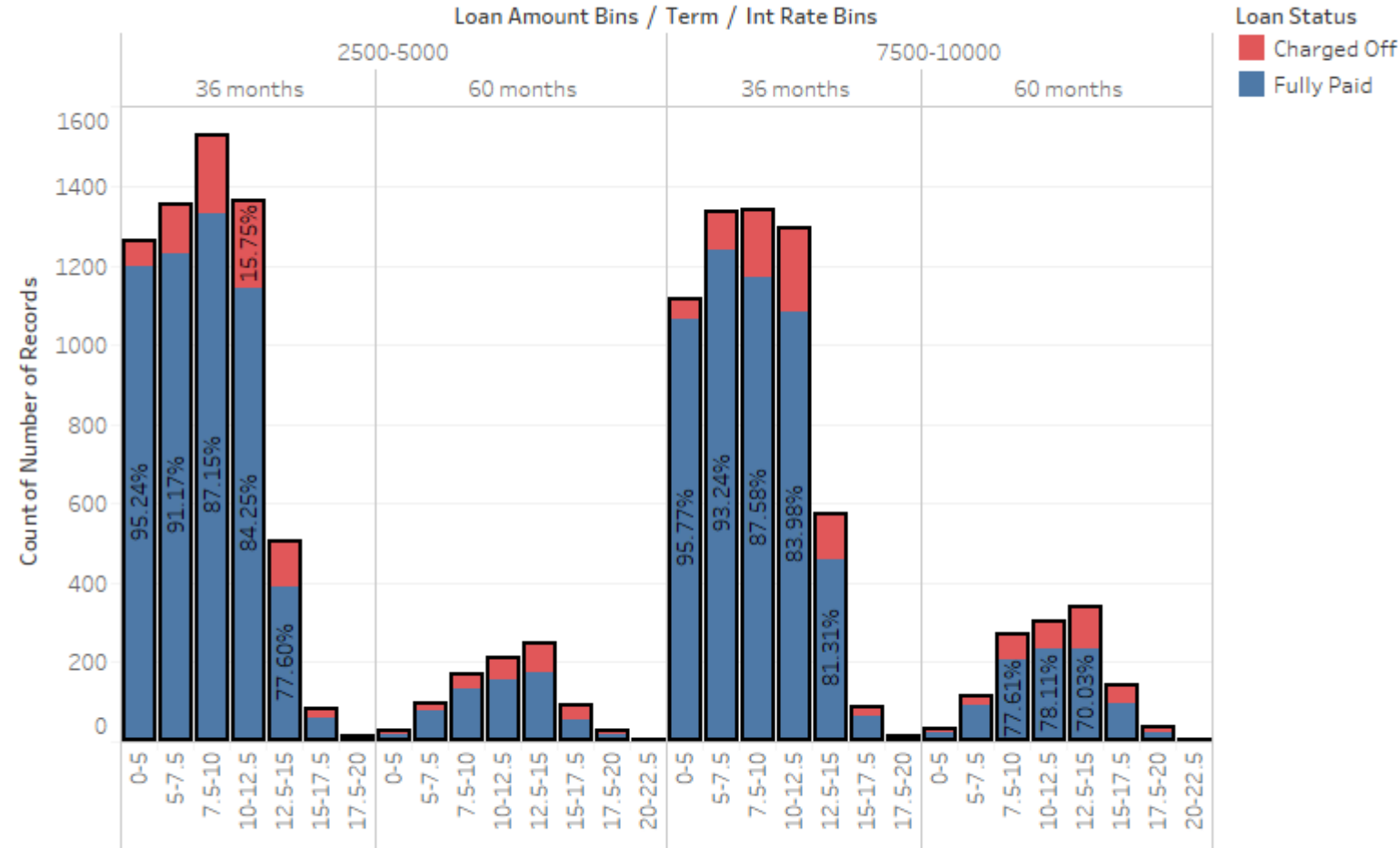
Count of Number of Records for each Loan Amount Bins broken down by Purpose. Color shows details about Loan Status. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

Loan Amount Bins Vs Loan Status Vs Annual Income Bins



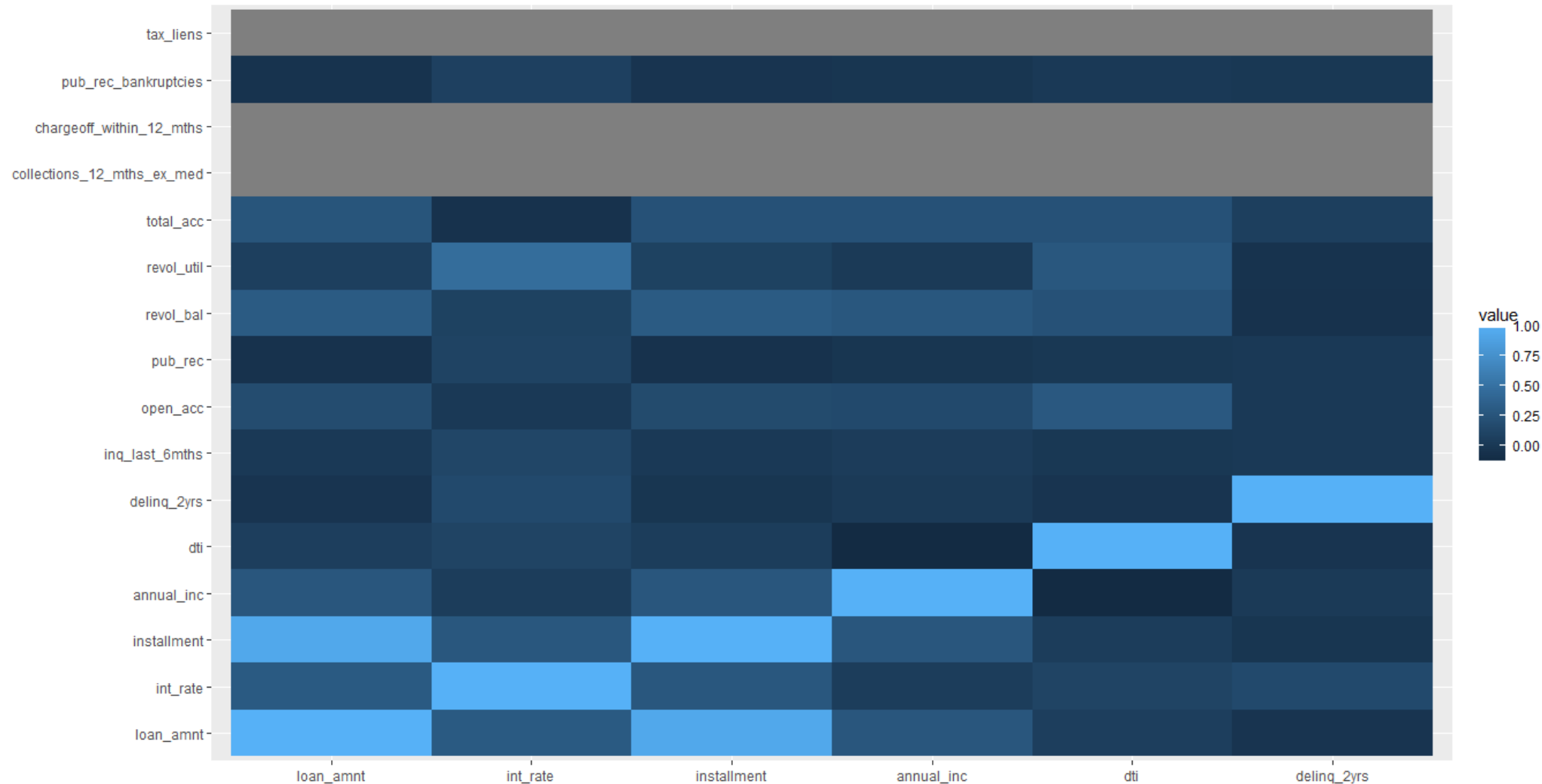
Count of Number of Records for each Annual Income Bins broken down by Loan Amount Bins and Loan Status. Color shows details about Loan Status. Size shows details about Home Ownership. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Amount Bins, Loan Status, Annual Income Bins and Home Ownership. The Loan Amount Bins filter keeps 2500-5000 and 7500-10000. The Loan Status filter keeps Charged Off and Fully Paid. The Annual Income Bins filter keeps 15 members. The Home Ownership filter keeps MORTGAGE and RENT.

Loan Amount Bins Vs Term Vs Interest Rates



Count of Number of Records for each Int Rate Bins broken down by Loan Amount Bins and Term. Color shows details about Loan Status. The marks are labeled by % of Total Number of Records. The view is filtered on Loan Status and Loan Amount Bins. The Loan Status filter keeps Charged Off and Fully Paid. The Loan Amount Bins filter keeps 2500-5000 and 7500-10000.

CORRELATION BETWEEN CONTINUOUS VARIABLES



- Positive correlation exists between “Loan amount” and “installment” followed by “interest rate” and “annual income”
- Negative correlation between “dti” and “annual_income”

CONCLUSIONS

Points observed from BIVARIATE & MULTIVARIATE ANALYSIS	
Annual Income	Higher the annual income the defaulters are increasing
Interest Rate	Higher the interest rate more number of defaulters
Grade of Loan	From grade A-G defaulters increases
Term of Loan	60 months higher the defaulters comparatively
Home Ownership	RENT/ MORTGAGE applicants are more defaulters
Revol Util Rate	Higher the revol Util rates more number of defaulters
Employee Length	People having less than 1 year or 10+ years
Purpose	Debt consolidations is having high more number of defaulters