GRAEMENER LOAN - CASE STUDY

Loan Processing Solutions
Increasing Efficiency, Productivity, and Security with Adeptia

UpGrad iiit-b

By – GUARAV, AMAN, MUKUND

Problem Analysis

Data and Problem Statement:

This data has details about 4 years of loan that has been issued by the bank. They are mainly of three types *Fully Paid off, Current and Defaulted*. The data set contains all sorts of information about customer like annual income, employee length, purpose of loan, bankrupt count etc.

Strategies to solve problem:

- Gather data for the analysis.
- Clean and Format the data for readability.
- Extrapolate the Derived Metrics.
- Filter out the outliers based on Annual Income.
- Filter Data Set on Loan Status.
- Perform Univariate and Bivariate analysis on the Data.
- Plot the results of Analysis and Hypnotize the Root of the Issue.

Data Loading

- Source data have total more than 11dimensions
- Data has been loaded in loan_data data frame

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	verification_status	issue_d	loan_status
	1077501	1296599	5000	5000	4975	36 months	10.65%	162.87	B	B2		10+ years	RENT	24000	Verified	11-Dec	Fully Paid
	1077430	1314167	2500	2500	2500	60 months	15.27%	59.83	C	C4	Ryder	< 1 year	RENT	30000	Source Verified	11-Dec	Charged Off
	1077175	1313524	2400	2400	2400	36 months	15.96%	84.33	C	C5		10+ years	RENT	12252	Not Verified	11-Dec	Fully Paid
	1076863	1277178	10000	10000	10000	36 months	13.49%	339.31	C	C1	AIR RESOURCES	10+ years	RENT	49200	Source Verified	11-Dec	Fully Paid
	1075358	1311748	3000	3000	3000	60 months	12.69%	67.79	B	B5	University Medi	1 year	RENT	80000	Source Verified	11-Dec	Current
	1075269	1311441	5000	5000	5000	36 months	7.90%	156.46	A	A4	Veolia Transpor	3 years	RENT	36000	Source Verified	11-Dec	Fully Paid
	1069639	1304742	7000	7000	7000	60 months	15.96%	170.08	C	C5	Southern Star Ph	8 years	RENT	47004	Not Verified	11-Dec	Fully Paid
	1072053	1288686	3000	3000	3000	36 months	18.64%	109.43	E	E1	MKC Accounting	9 years	RENT	48000	Source Verified	11-Dec	Fully Paid
	1071795	1306957	5600	5600	5600	60 months	21.28%	152.39	F	F2		4 years	OWN	40000	Source Verified	11-Dec	Charged Off
	1071570	1306721	5375	5375	5350	60 months	12.69%	121.45	B	B5	Starbucks	< 1 year	RENT	15000	Verified	11-Dec	Charged Off
	1070078	1305201	6500	6500	6500	60 months	14.65%	153.45	C	C3	Southwest Rural	5 years	OWN	72000	Not Verified	11-Dec	Fully Paid
	1069908	1305008	12000	12000	12000	36 months	12.69%	402.54	B	B5	UCLA	10+ years	OWN	75000	Source Verified	11-Dec	Fully Paid
	1064687	1298717	9000	9000	9000	36 months	13.49%	305.38	C	C1	Va. Dept of Cons	< 1 year	RENT	30000	Source Verified	11-Dec	Charged Off
	1069866	1304956	3000	3000	3000	36 months	9.91%	96.68	B	B1	Target	3 years	RENT	15000	Source Verified	11-Dec	Fully Paid
	1069057	1303503	10000	10000	10000	36 months	10.65%	325.74	B	B2	SFMTA	3 years	RENT	1.00E+05	Source Verified	11-Dec	Charged Off
	1069759	1304871	1000	1000	1000	36 months	16.29%	35.31	D	D1	Internal revenue	< 1 year	RENT	28000	Not Verified	11-Dec	Fully Paid
	1065775	1299699	10000	10000	10000	36 months	15.27%	347.98	C	C4	Chin's Restaura	4 years	RENT	42000	Not Verified	11-Dec	Fully Paid
	1069971	1304884	3600	3600	3600	36 months	6.03%	109.57	A	A1	Duracell	10+ years	MORTGAGE	110000	Not Verified	11-Dec	Fully Paid

Snapshot – of loan source data

Loan Case Study

- ## Loan Applications and Evaluations:

  - Identify "High Risk" loan applicants
  - EDA behind Loan default
  - Risk Minimize and Decision

- ## Loan Data : Charged off + Full Paid + Running

  – Total Loan Applications: 39717

  – Total Charged off + Full Paid : 38577

  – Running Loan :1140

Data Understanding

- **Types of variables**

    a. Customer (applicant) demographics:

    - Employment length, Title, Annual Income, Zip Code

    b.   Loan related information and characteristics

    - Loan Amount Funded Amount, investment, Interest Rate, Loan Status, Loan Grade

    a. Customer behavior (in case loan is granted

    - Credit History, Yearly Balance, Recovery, Type, Purpose

Data Mining

- Replacement of Separator (/ to -)
- Count of NA and Blanks
- Length of Loan Id
- Duplicate Value check : No Duplicate
- Loan Count : 39717
- Date Format Change
- Character Numerical and Factor Conversion
- Comparison and Analysis

Fields of Data Set

- **Loan id: id and member_id**
  - Each new loan is identified by this number
  - Discrete variable since it is like a count of the loan
- **Loan Amount and Interest**: loan_amnt and int_rate
  - The amount that has been given as a loan
  - Continuous variable since it can be fractioned between given intervals
- **Loan Duration (in months):**
  - Repayment period
  - Discrete since it is countable and is restricted to months
- **Loan Payments and Member Income:**
  - Total amount that has been repaid from income except expense
  - Continuous variable since it can be fractioned between given intervals
- **Member Grade and Loan status:**
  - The status of the loan i.e. whether the customer has paid it on time or not.
  - Ordinal categorical/Qualitative variable

Income Amount Box Plot
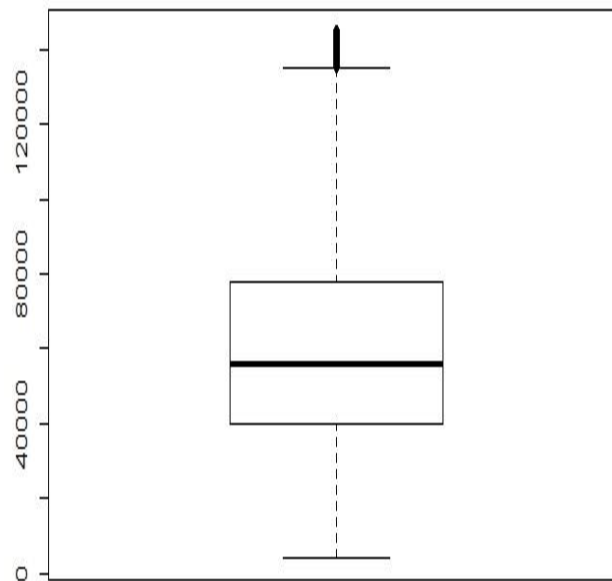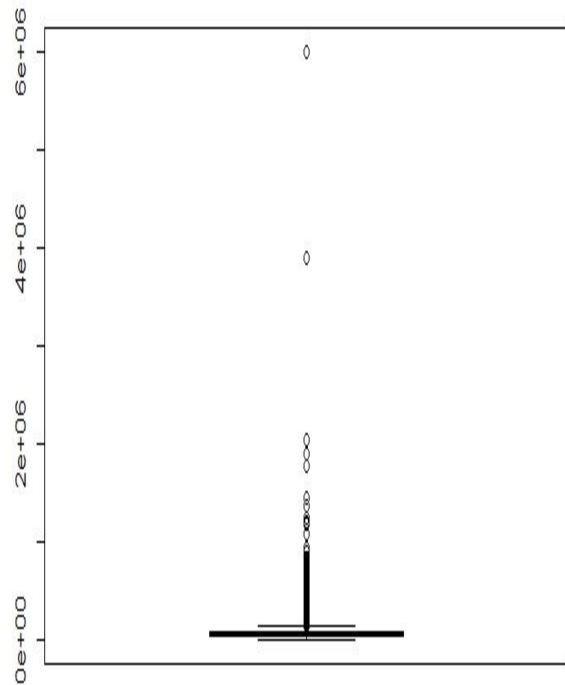
Funded Amount Box Plot
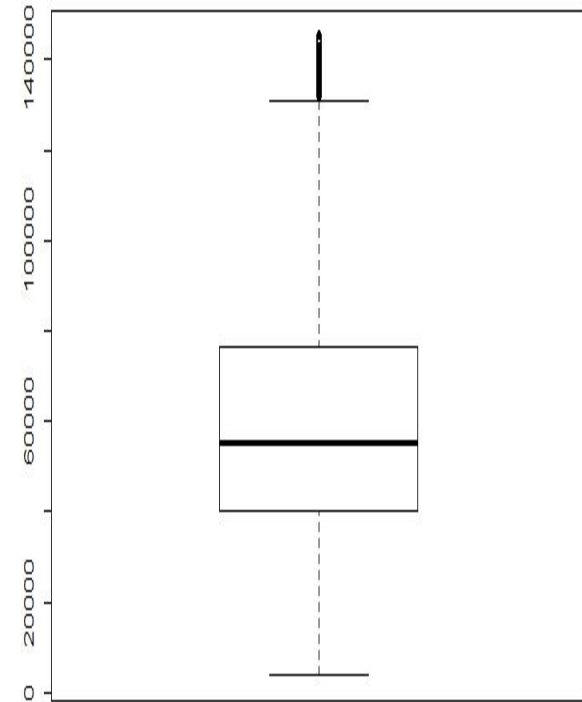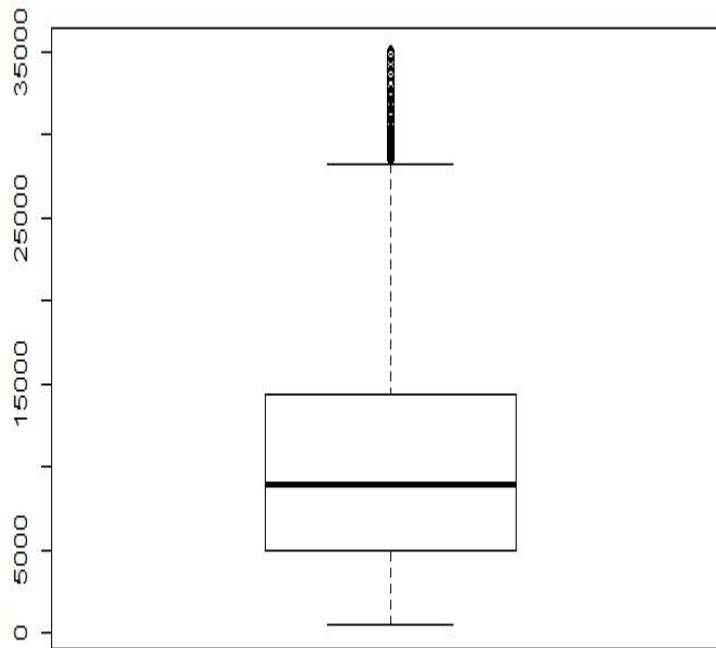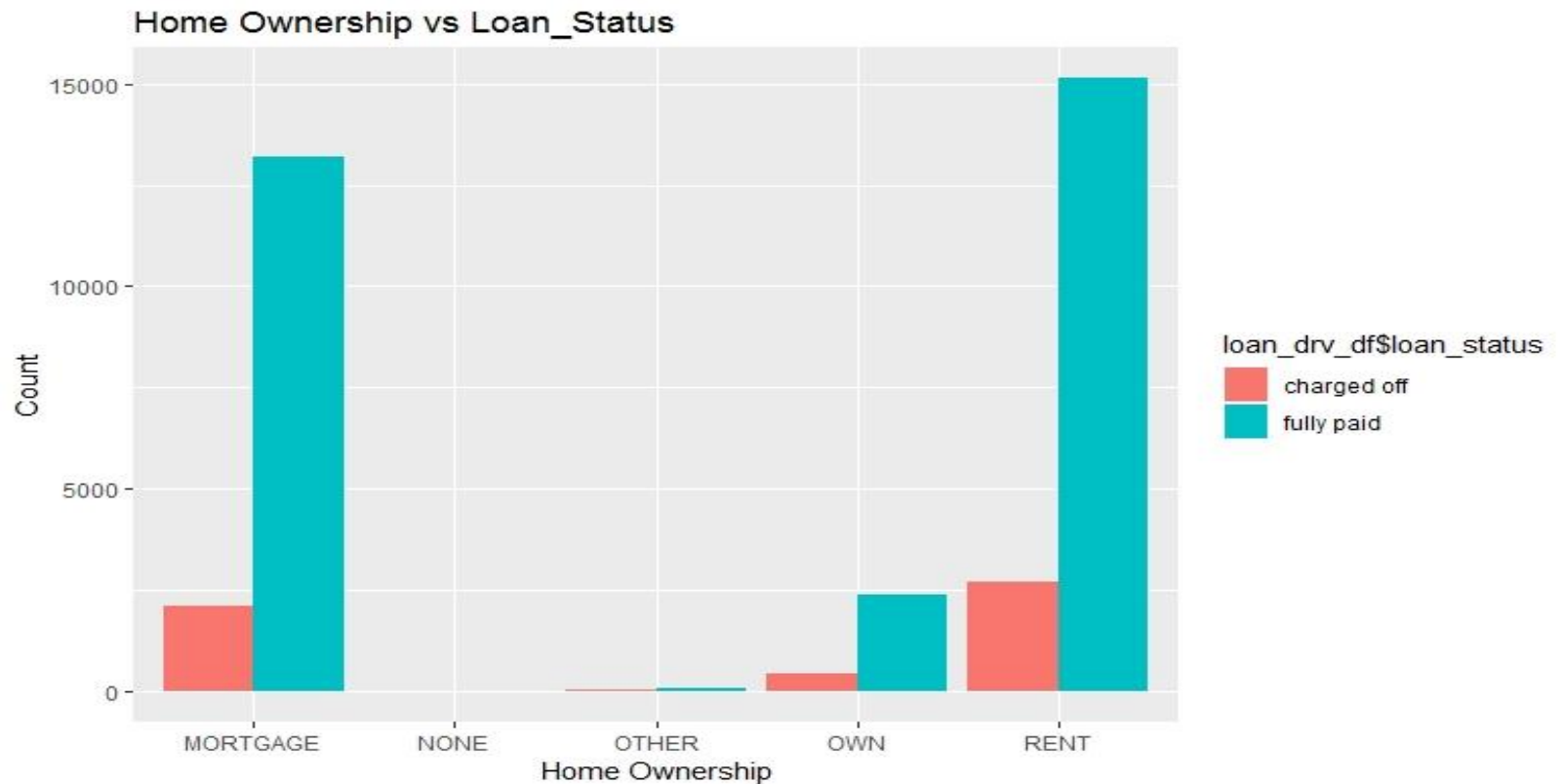
Loan Vs Loan Status
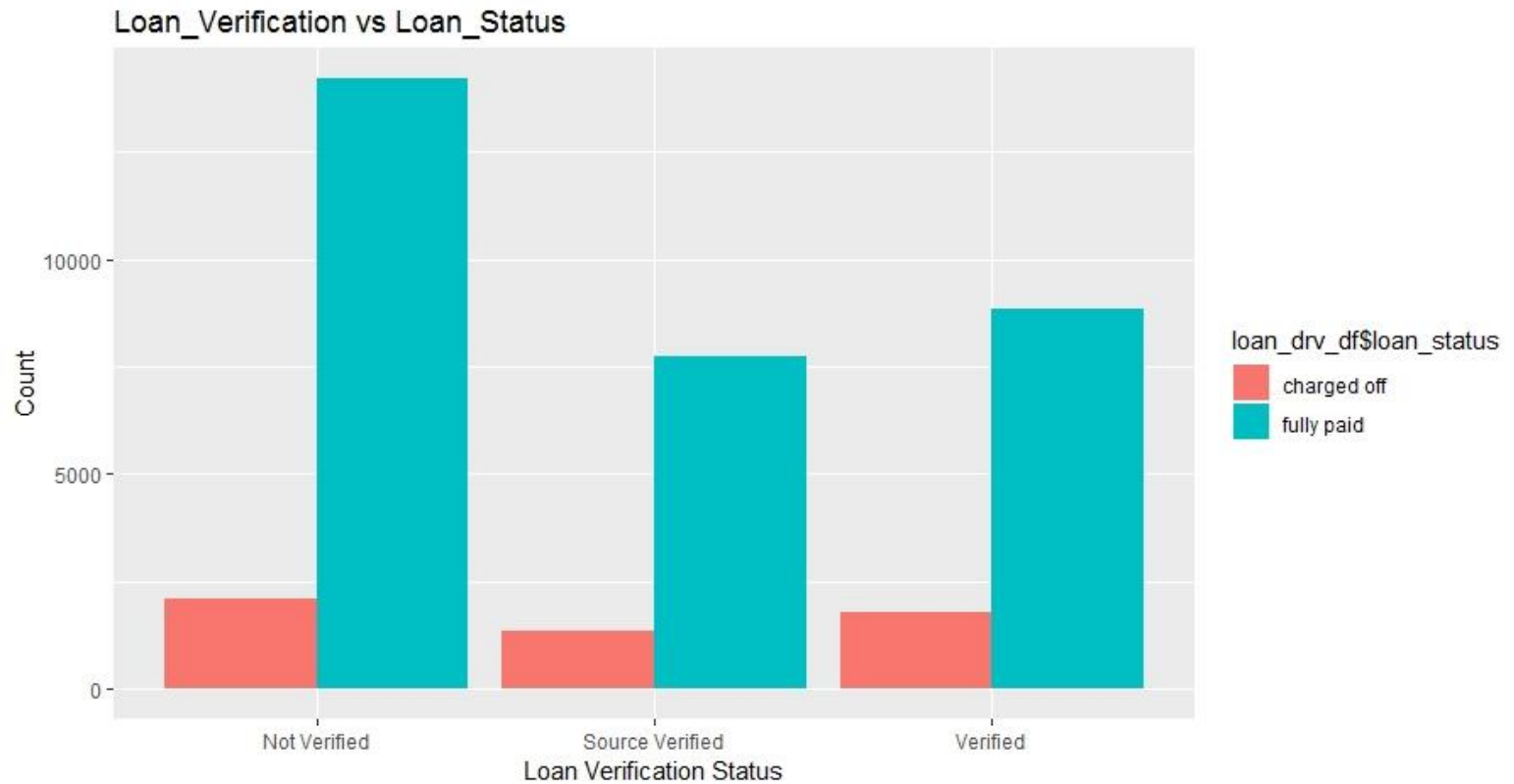
Loan Grade vs Loan Status
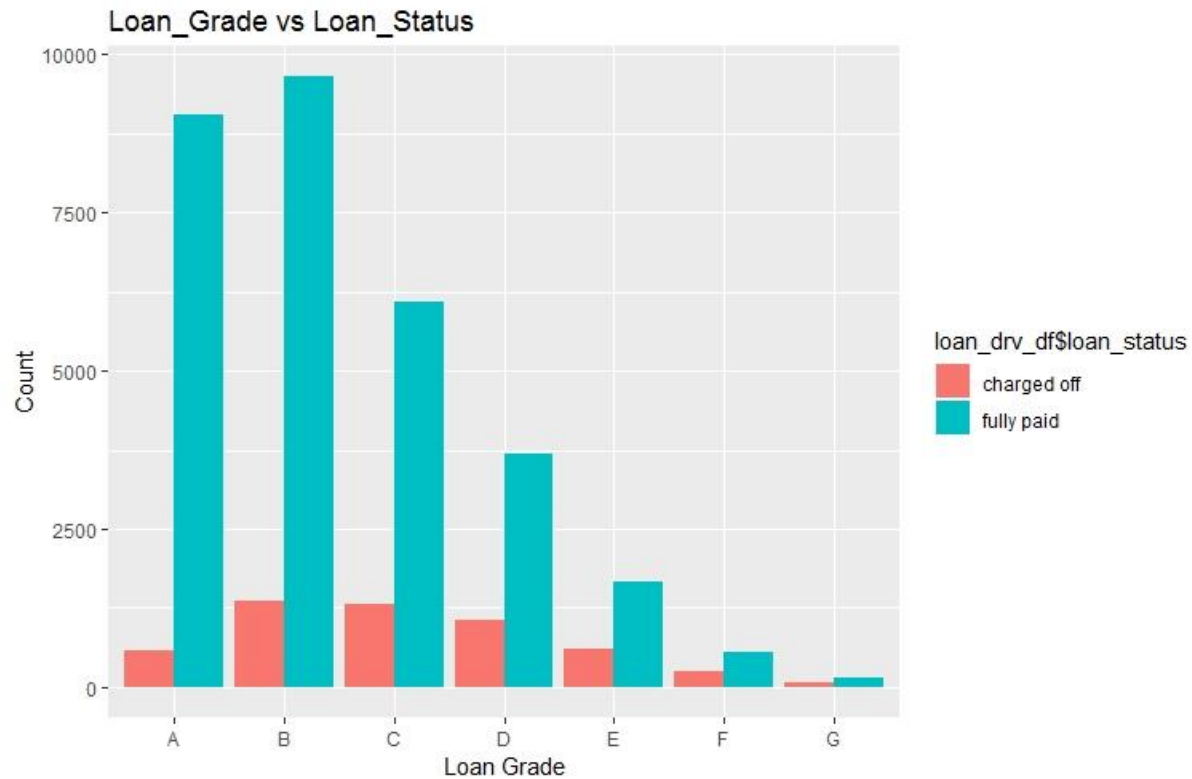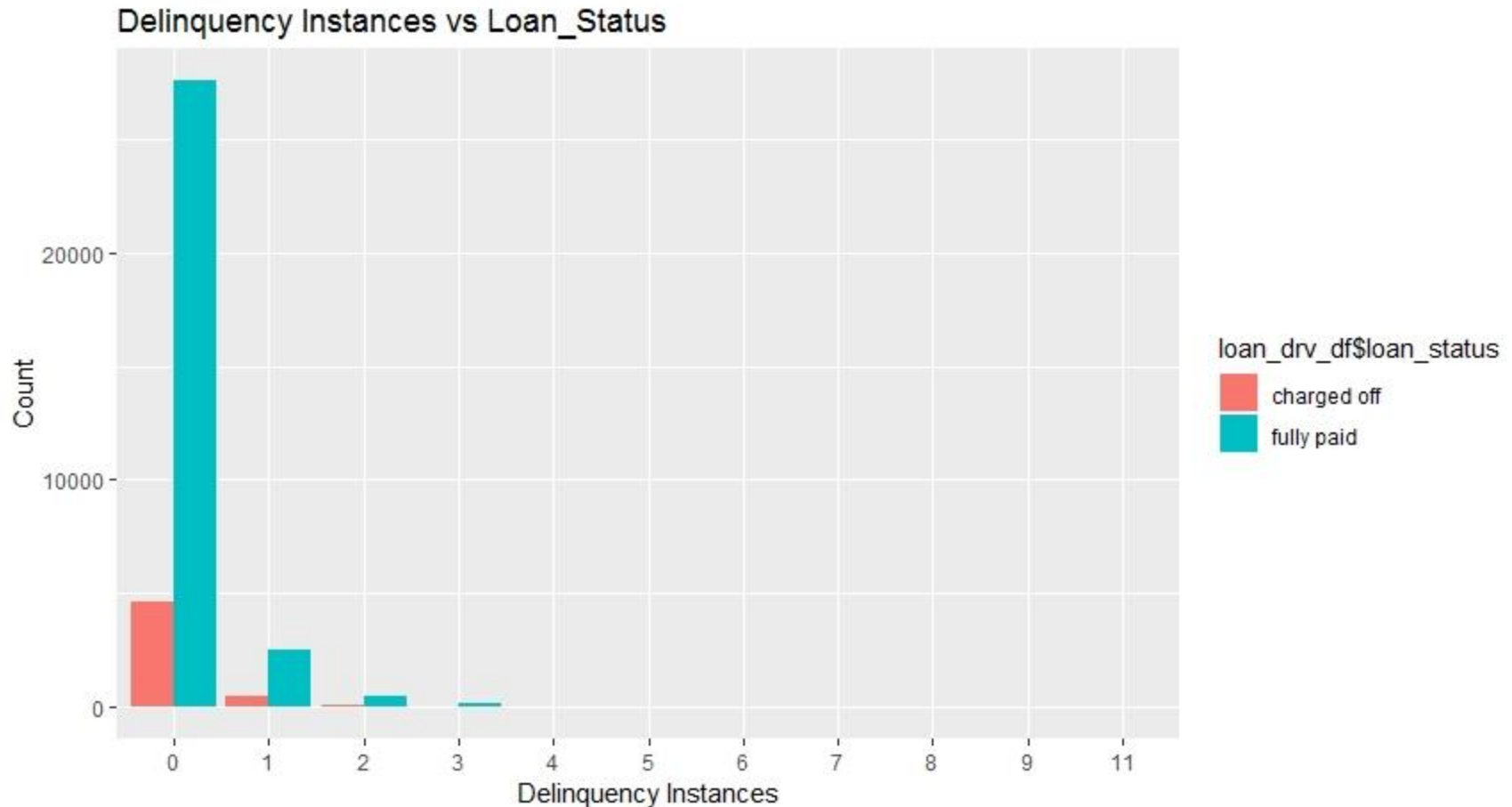
DTI_Level vs Loan_Status
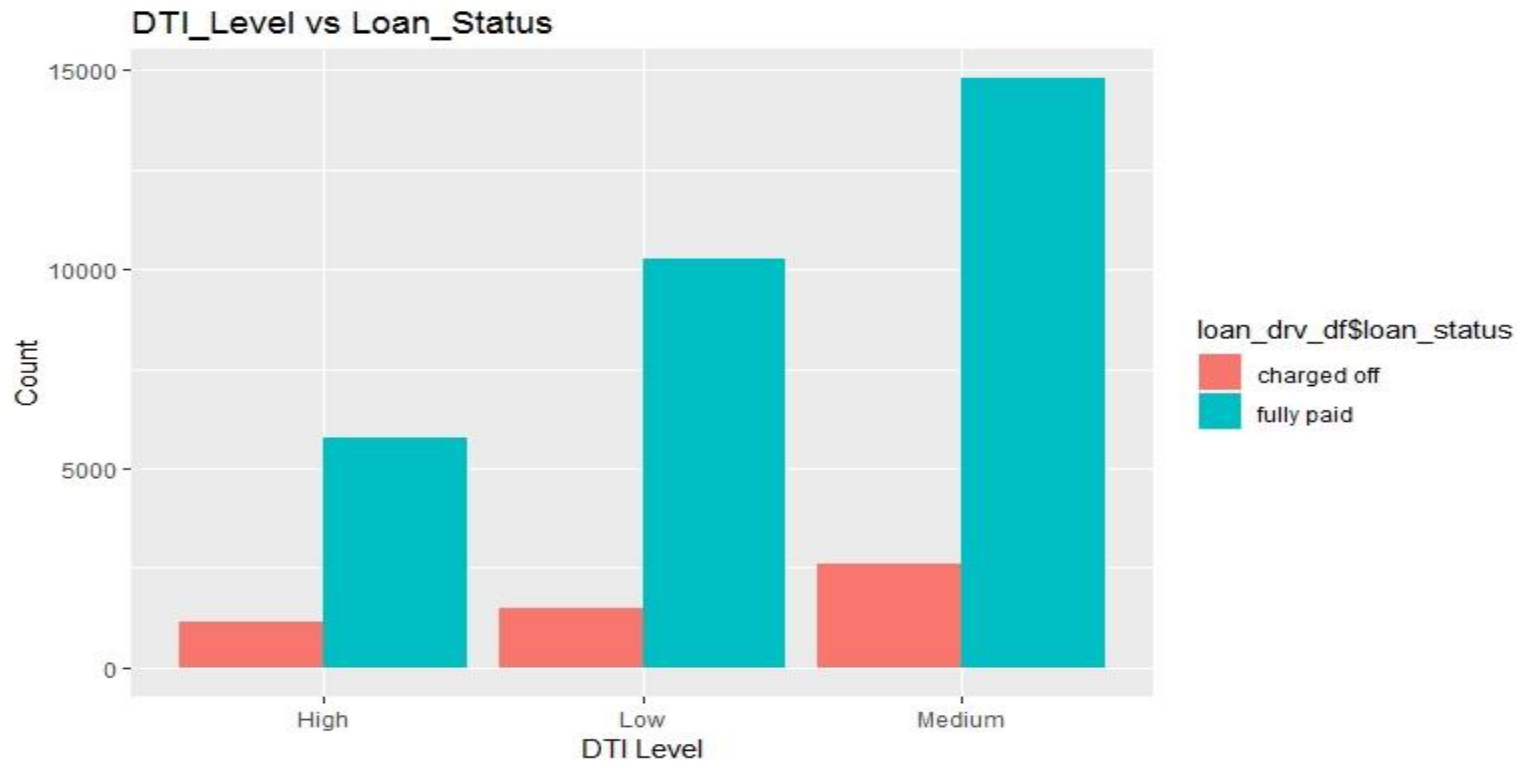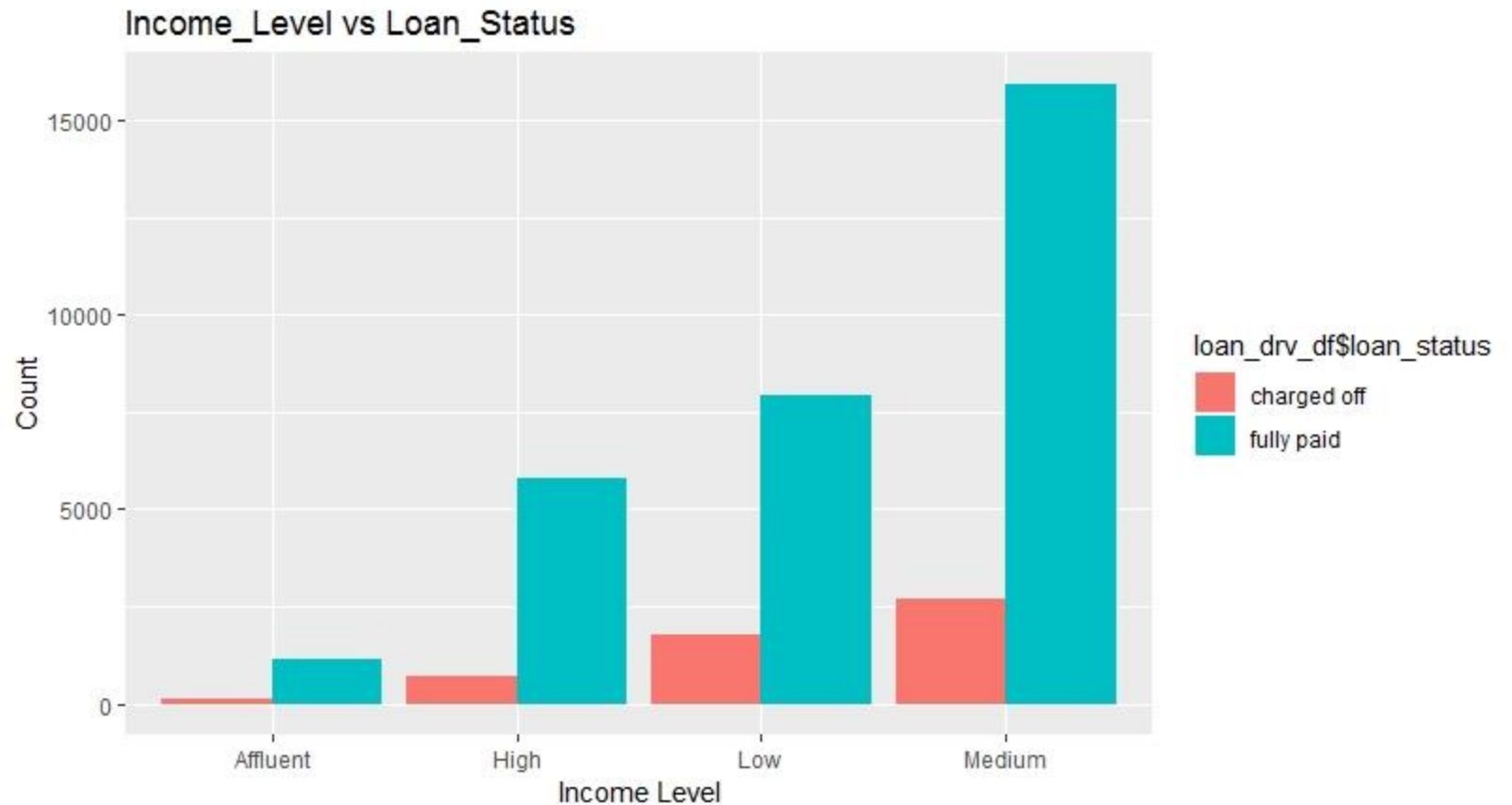
Loan_Purpose vs Loan_Status
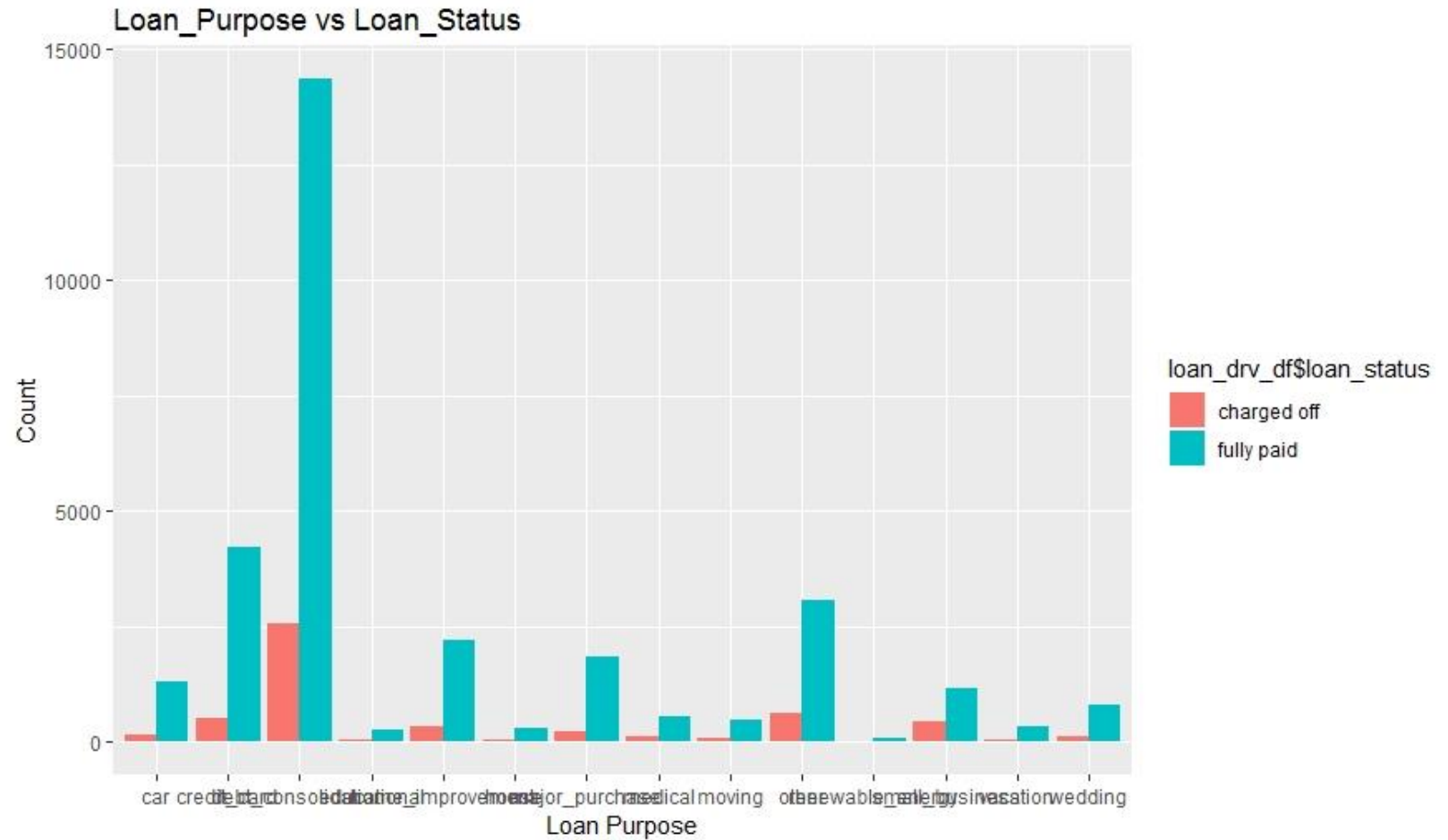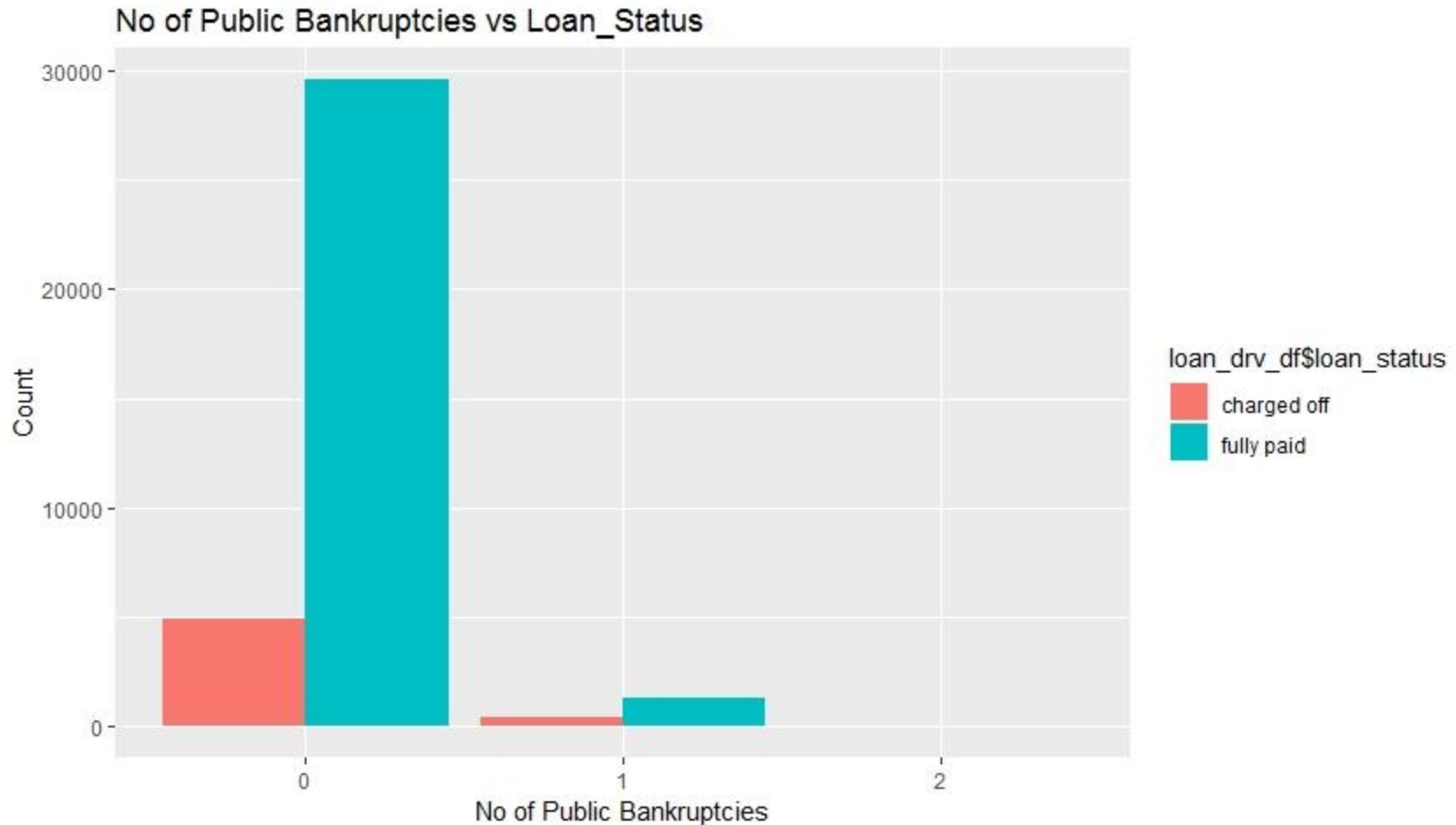
Home Ownership vs Loan_Status
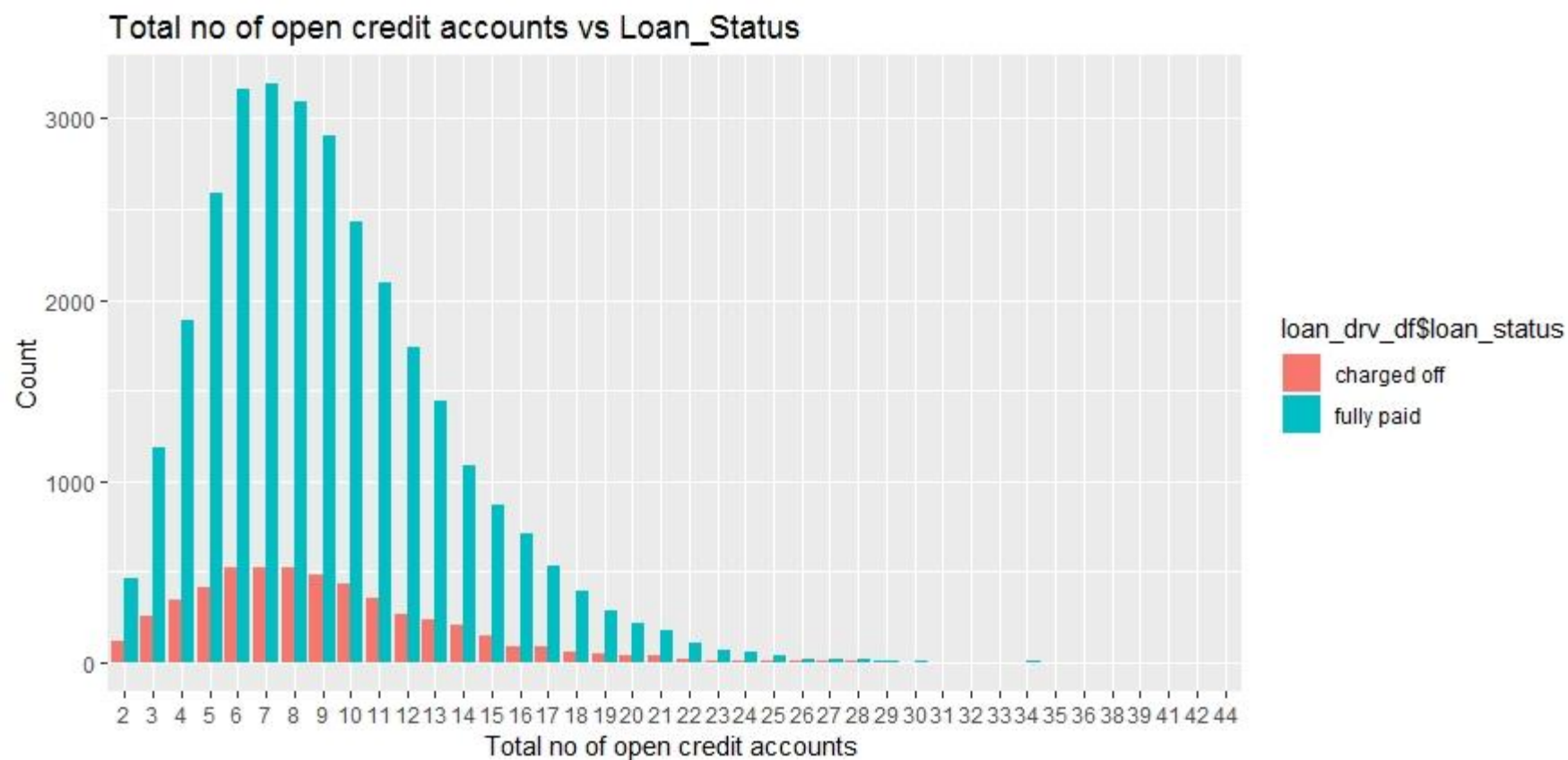
Delinquency instances vs Loan_Status

Income level vs Loan_Status

Public Bankrupt Count vs Loan_Status

Bank Credit vs Loan_Status

Loan Defualted

Employee Service Length (Loan Defaulted)

Loan Defualted
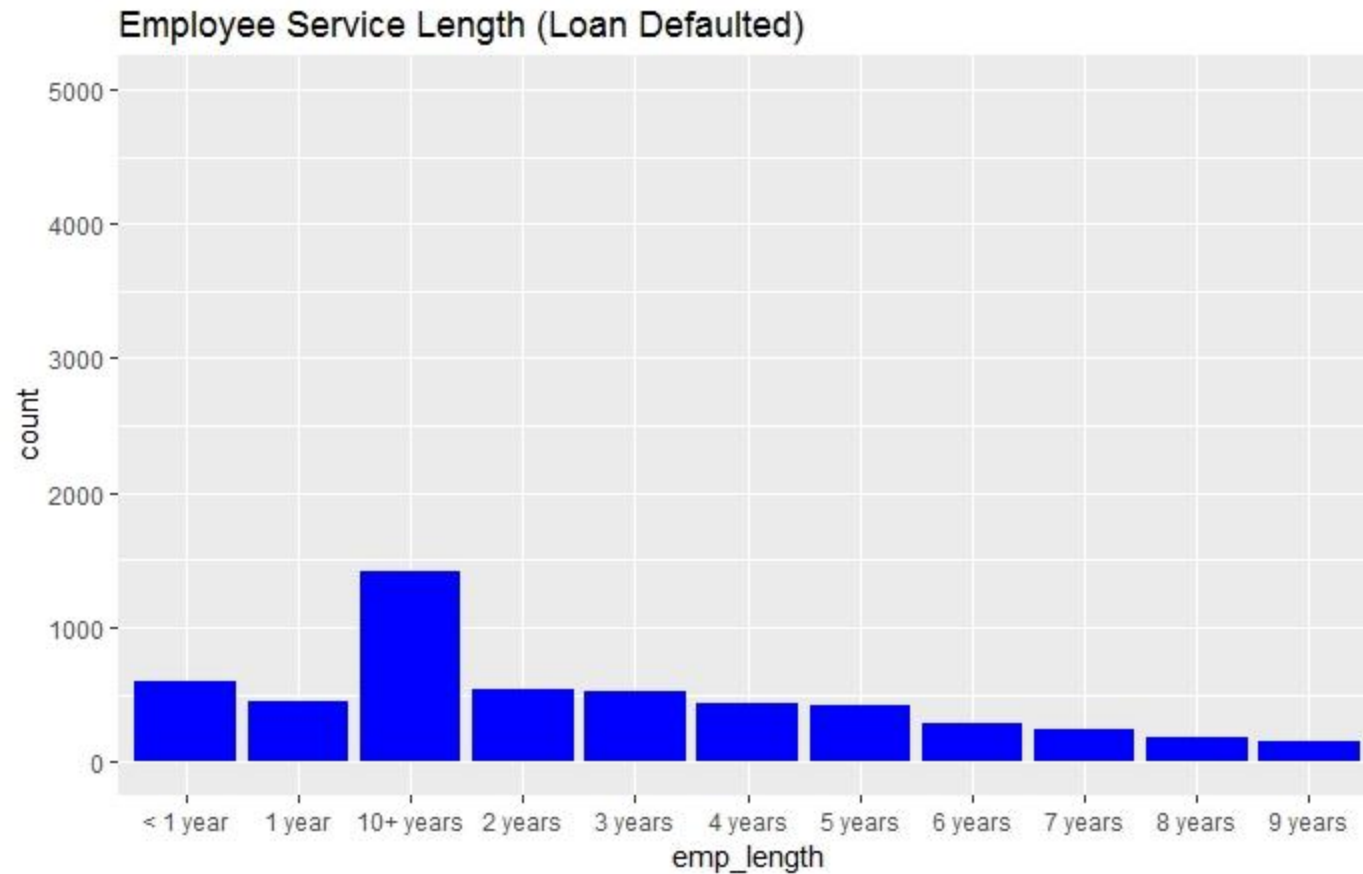
Employee Service Length (Loan Defaulted)

Income vs Loan_Status
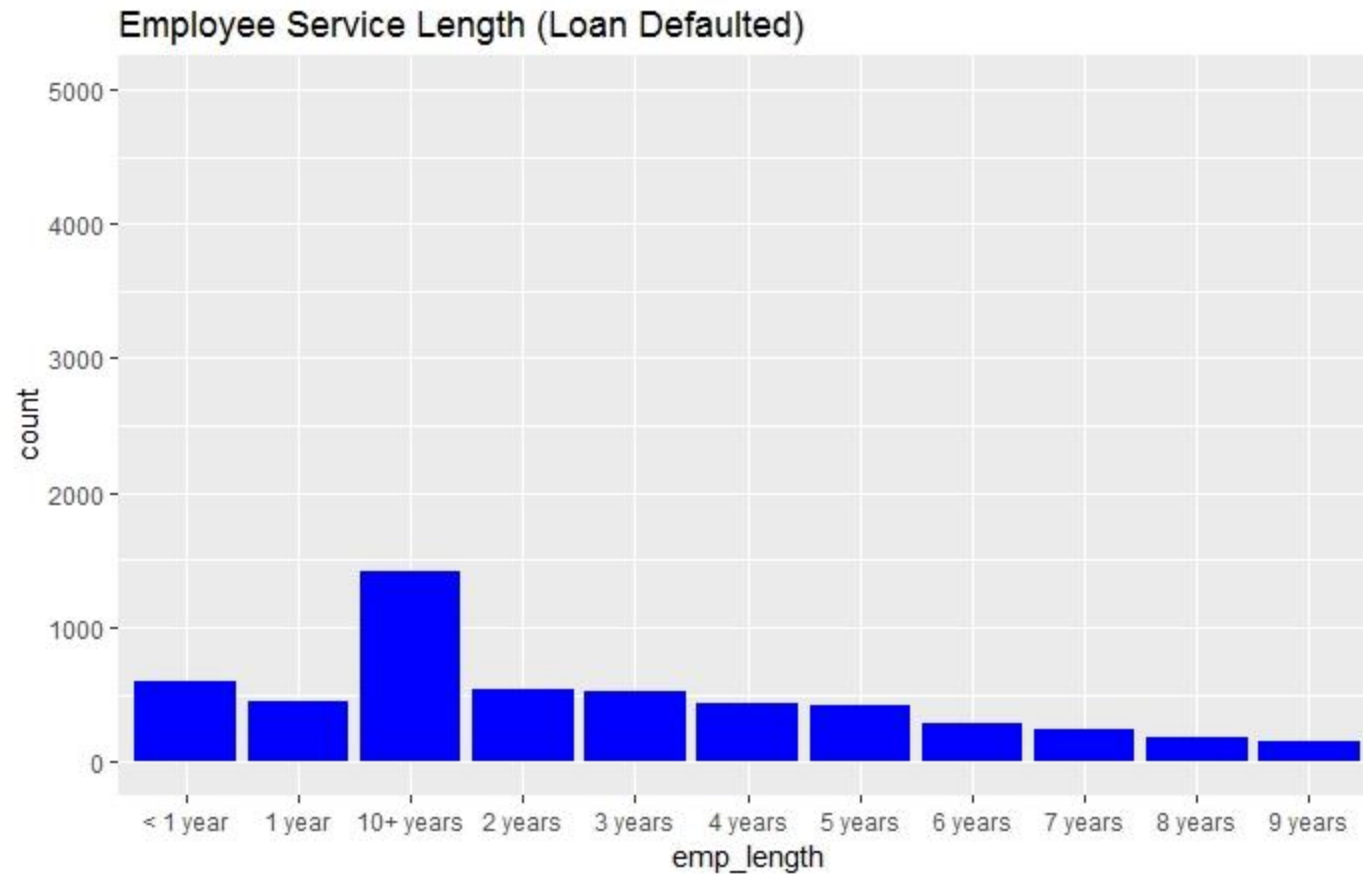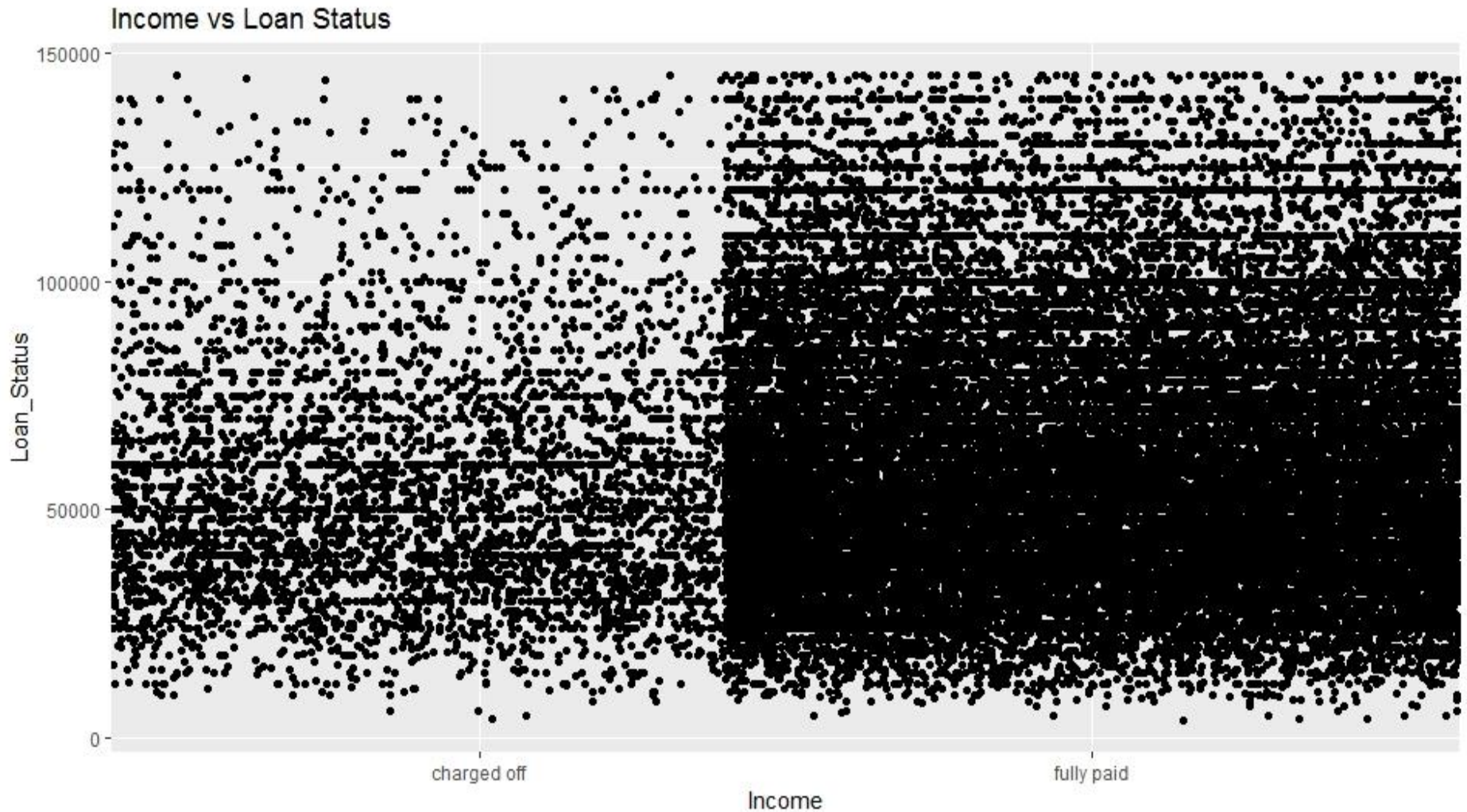

Income vs Loan Status

Conclusions

- After examining annual income and grades, found that all the grades have almost similar distribution.

- *Annual Income, interest rate* and *verification* have bearing on whether an loan will become fully paid or become charged off.

- For current loan disbursal, from grade D to grade G, the percentage of defaults are more. Grades play an important role in loan disbursal and repayment.

- Target based or last minute loan disbursals should be avoided.

Recommendations

- Customer should be classified clearly based on their Credit score, income, frequency of application, purpose of use, loan duration and interest rate and disbursed amount.

- The process of income verification should be more accurate as it is an important parameter to control defaulters.

- Source of income and stability should be verify