



GRAMENER CASE STUDY

Kolahalan.B

Approach

DATA IMPORTING

Imported Data using
pandas(pd.read_csv)

DATA CLEANING

- NA's analysis
- Duplicate
- Changing the class of observation.

DATA ANALYSIS

Exploratory data analysis –Question the data to perform:

- Univariate* analysis of both categorical and continuous variables.

DATA VISUALIZATION

1. Using Matplotlib & Seaborn to create graphs that aid in

- Defining the issues
- Analysis (univariate, bivariate)
- Segmentation analysis

Tools Used

- Pandas
- Numpy
- Matplotlib.pyplot
- Seaborn

Formatting and standardizing date time –(issue_d), percentage etc.

Bivariate analysis* –

- Categorical vs categorical
- Categorical vs continuous
- Continues vs continuous

2. Communicate inferences, understanding with supporting analysis and graphs to decision making audience and any larger audience.

Creating derived metrics -

- Charge off amount
- Charge off as a percentage of funded amount

Correlation matrix.

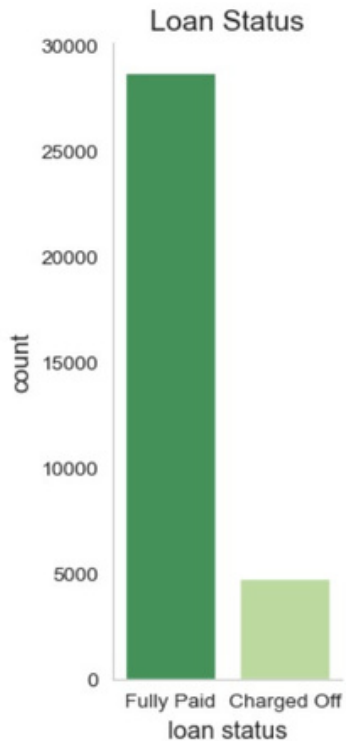
**Univariate analysis, Bivariate analysis includes Derived metrics as well

We are going to work with the loan data

1	LoanStatNew	Description
2	acc_now_delinq	The number of accounts on which the borrower is now delinquent.
3	acc_open_past_24mths	Number of trades opened in past 24 months.
4	addr_state	The state provided by the borrower in the loan application
5	all_util	Balance to credit limit on all trades
6	annual_inc	The self-reported annual income provided by the borrower during registration.
7	annual_inc_joint	The combined self-reported annual income provided by the co-borrowers during registration
8	application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers
9	avg_cur_bal	Average current balance of all accounts
10	bc_open_to_buy	Total open to buy on revolving bankcards.
11	bc_util	Ratio of total current balance to high credit/credit limit for all bankcard accounts.
12	chargeoff_within_12_mths	Number of charge-offs within 12 months
13	collection_recovery_fee	post charge off collection fee
14	collections_12_mths_ex_med	Number of collections in 12 months excluding medical collections
15	delinq_2yrs	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
16	delinq_amnt	The past-due amount owed for the accounts on which the borrower is now delinquent.
17	desc	Loan description provided by the borrower
18	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.
19	dti_joint	A ratio calculated using the co-borrowers' total monthly payments on the total debt obligations, excluding mortgages and the requested LC loan, divided by the co-borrowers' combined self-reported monthly income
20	earliest_cr_line	The month the borrower's earliest reported credit line was opened
21	emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
22	emp_title	The job title supplied by the Borrower when applying for the loan.*
23	fico_range_high	The upper boundary range the borrower's FICO at loan origination belongs to.
24	fico_range_low	The lower boundary range the borrower's FICO at loan origination belongs to.
25	funded_amnt	The total amount committed to that loan at that point in time.
26	funded_amnt_inv	The total amount committed by investors for that loan at that point in time.
27	grade	LC assigned loan grade
28	home_ownership	The home ownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER.
29	id	A unique LC assigned ID for the loan listing.
30	il_util	Ratio of total current balance to high credit/credit limit on all install acct
31	initial_list_status	The initial listing status of the loan. Possible values are – W, F
32	inq_fi	Number of personal finance inquiries
33	inq_last_12m	Number of credit inquiries in past 12 months

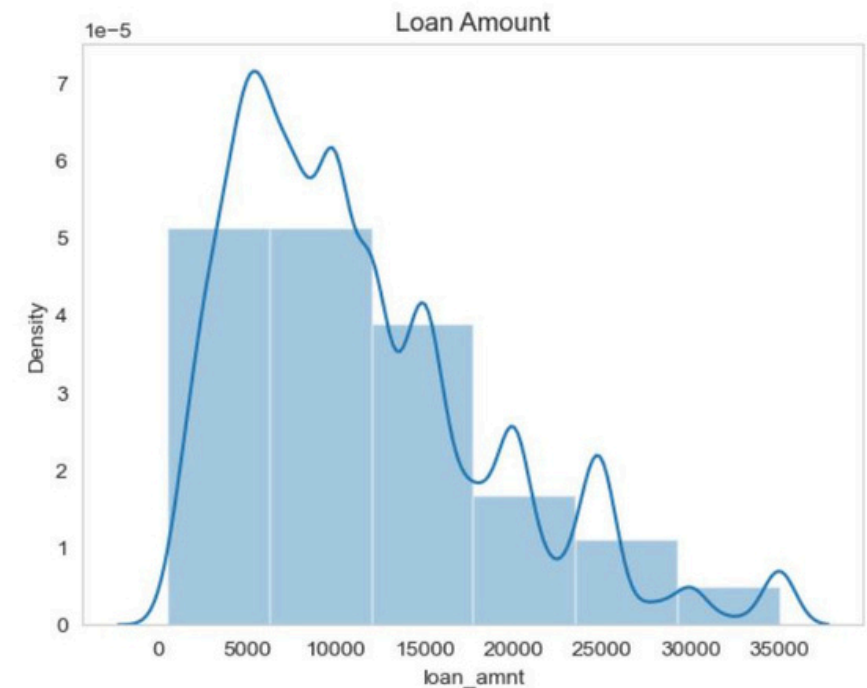
Analysis – Key Graph 1

Univariate Analysis



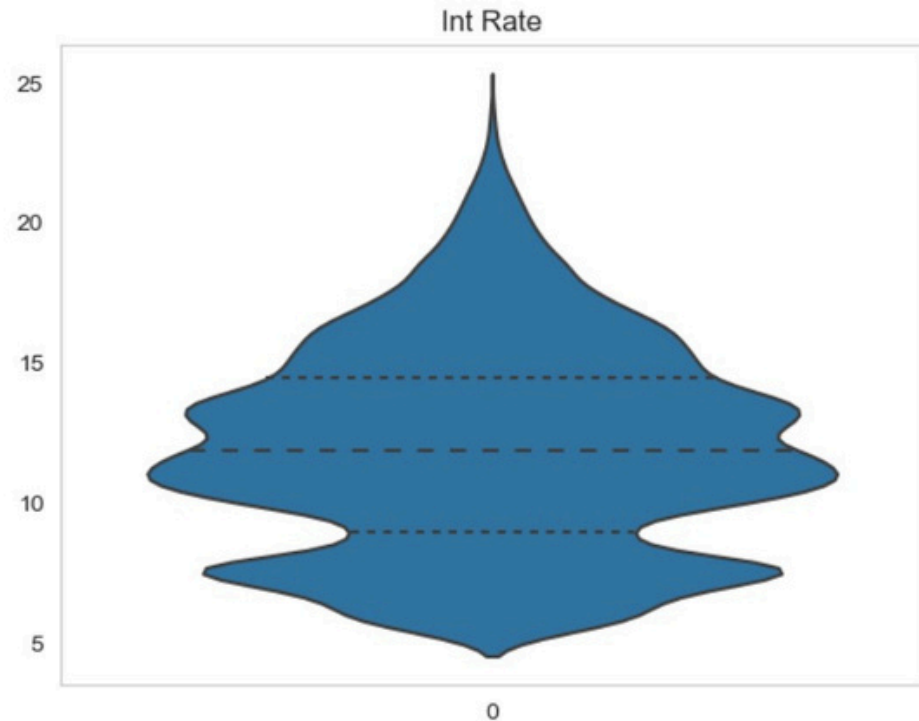
The graph shows that around 5000 people have defaulted/charged off.

Distribution of loan amount

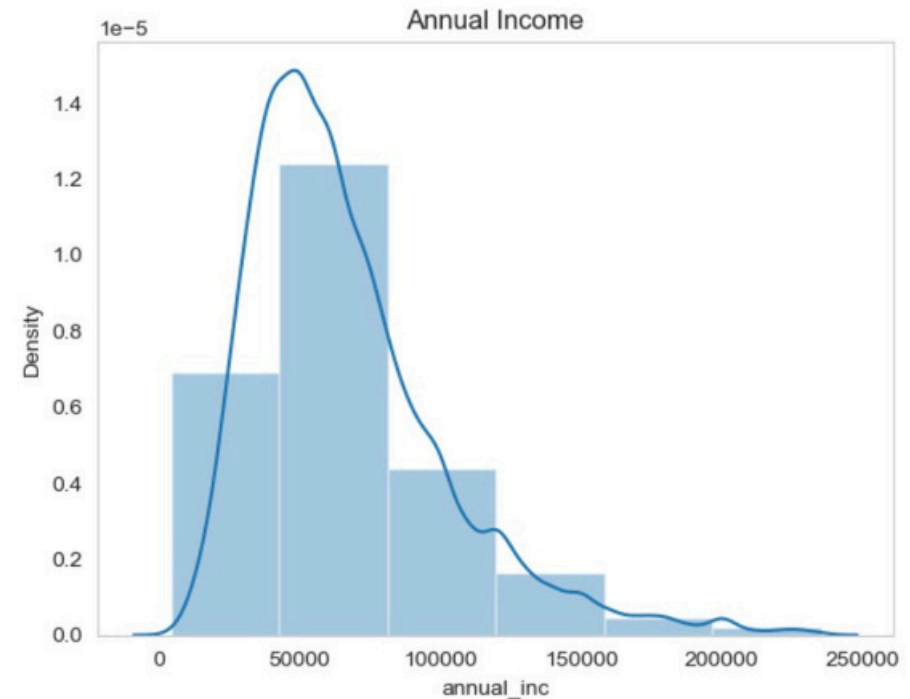


The graph above shows that the loan amount is majorly spread around say 6000 to 18000 approximately

Analysis – Key Graph 2

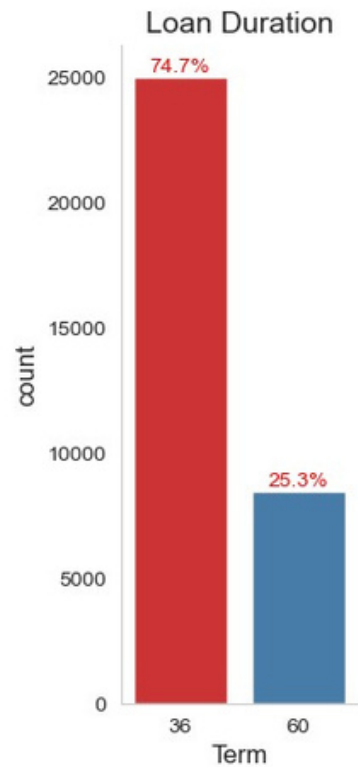


The above graph shows that the interest rate is spread majorly between 8% and 14% approximately

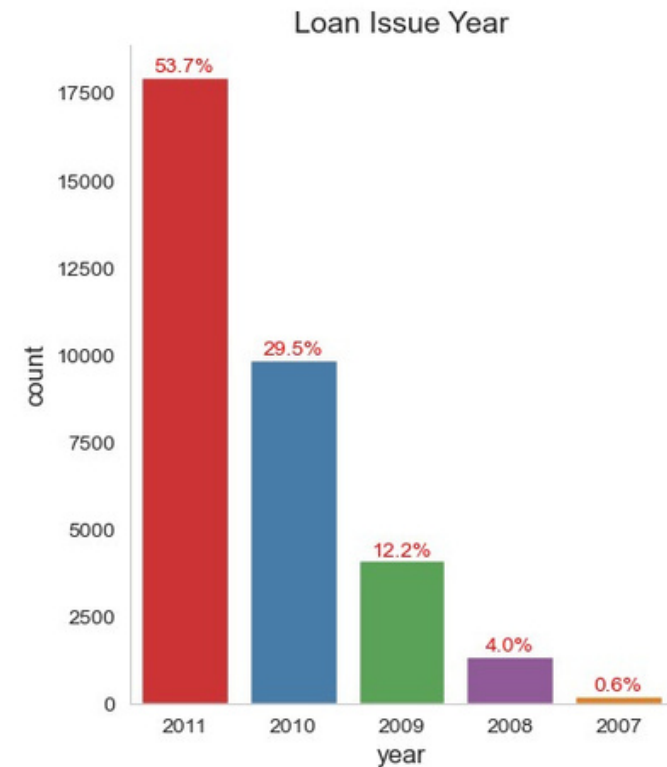


As per the above graph, majority of the applicants have an annual income ranging approximately between 40000 USD to 90000 USD.

Analysis – Key Graph 3

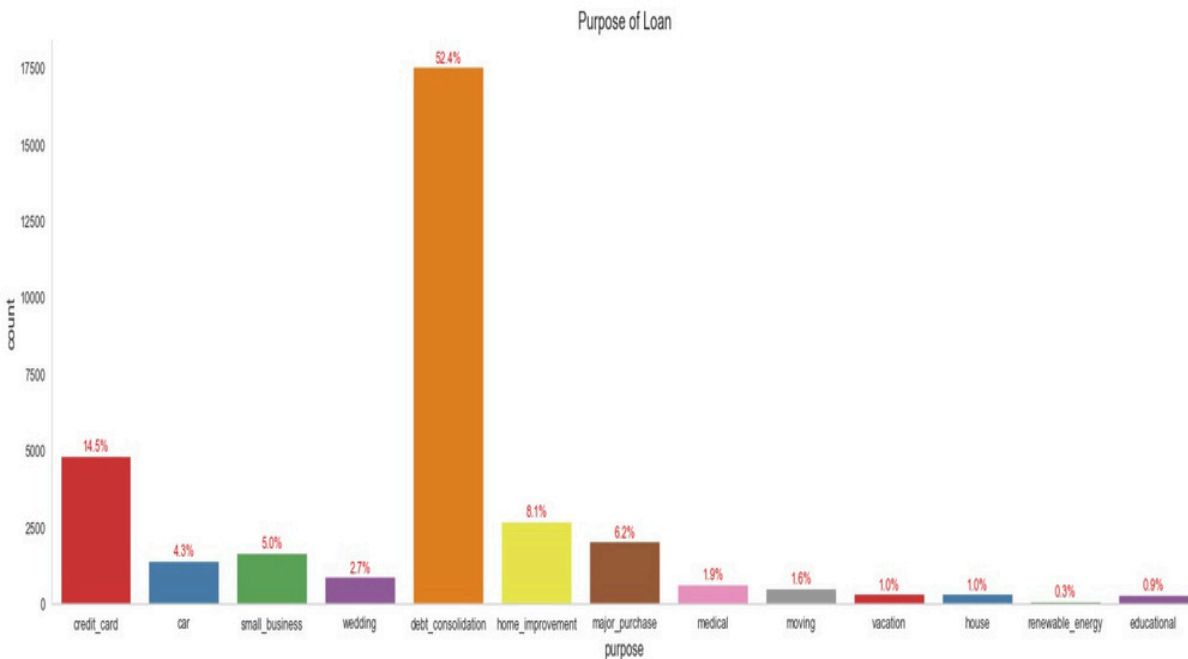


As per the above graph, majority of the applicants have taken the loan duration as 36 months (74.7%)



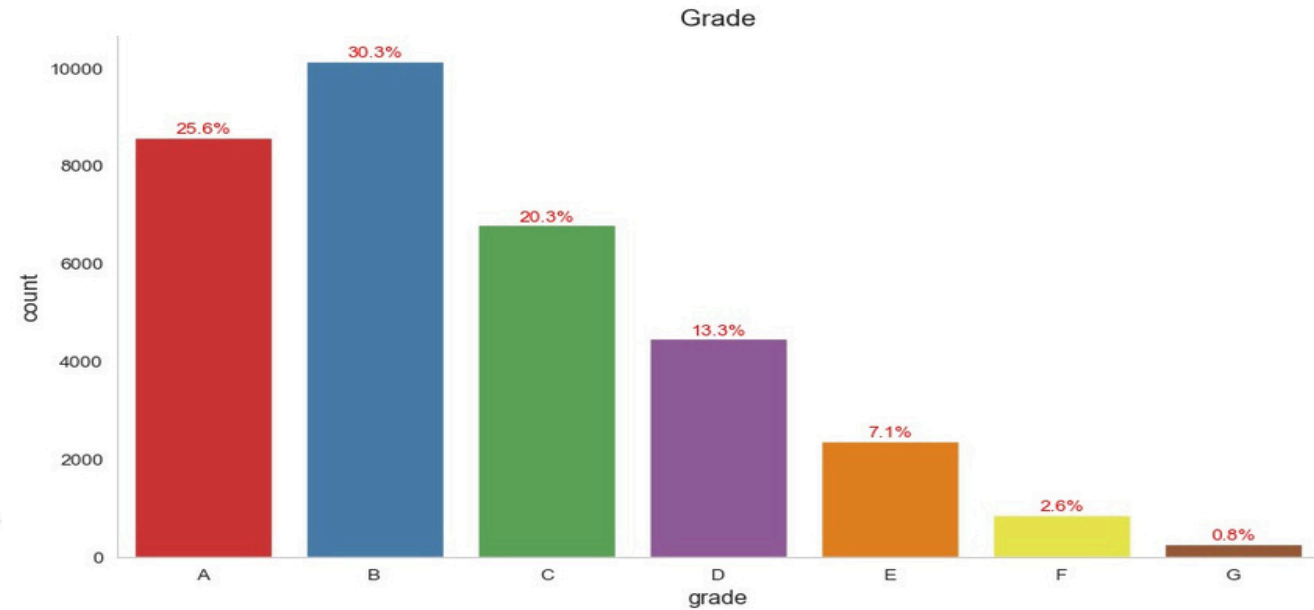
Per the above graph, applicants for loan increased as the years increased. In 2011 the number of applicants for loan was 53.7%. Since the variable issue year does not provide us any direction in the analysis, we would not be using this variable for any further analysis.

Analysis – Key Graph 4



The following are the purposes for which more than 5% applicants have taken loan

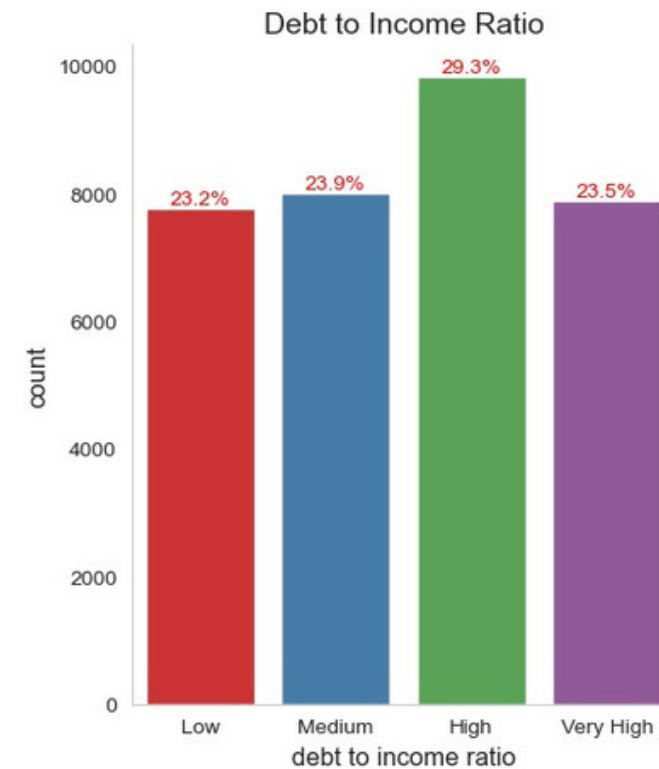
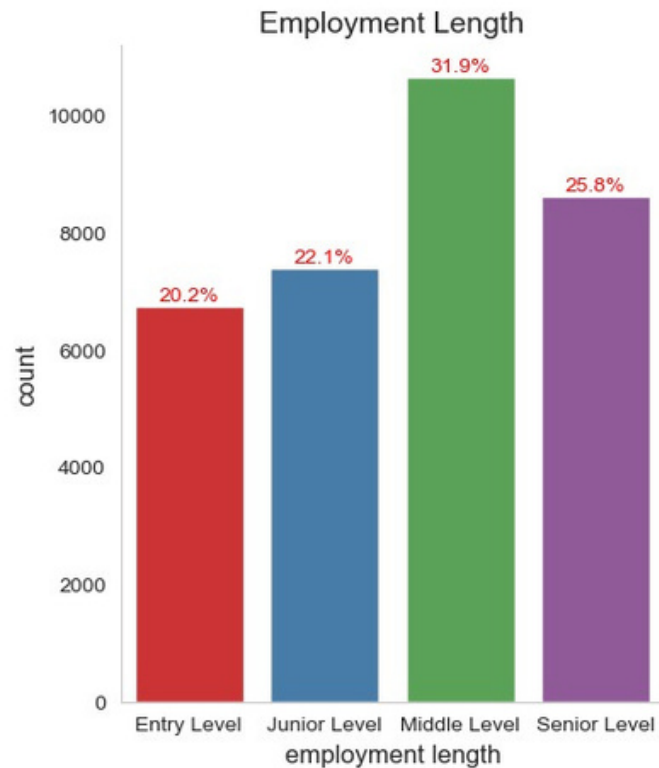
- debt_consolidation - 52.4%
- credit_card - 14.5%
- home_improvement - 8.1%
- major_purchase - 6.2%
- small business - 5.0%



The above graph shows that most of the applicants fall under the grade B(30.3%), followed by A(25.6%) and C(20.3%)

Analysis – Key Graph 5

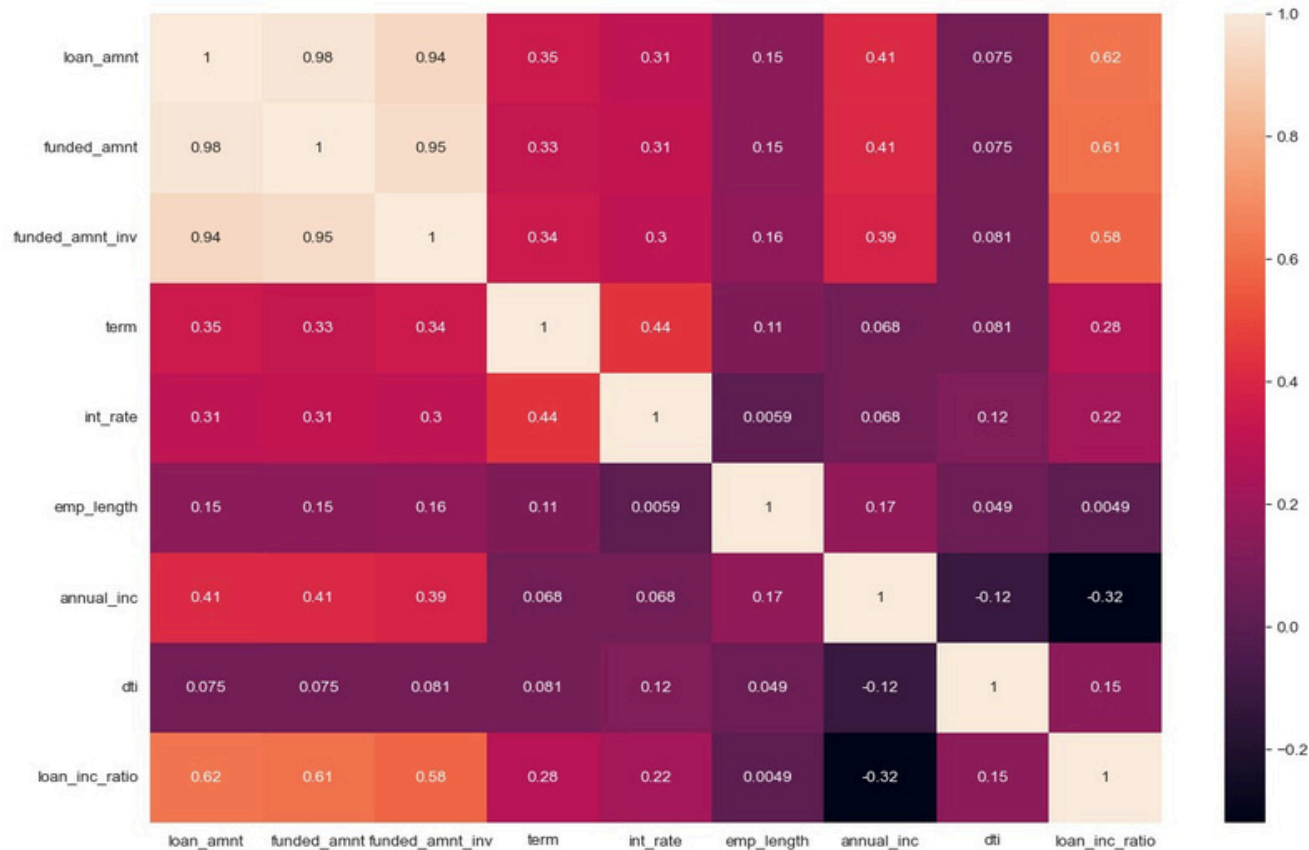
Univariate Analysis



There are more number of loan applicants belonging to the middle level category (31.9%) i.e. between 4 to 8 years

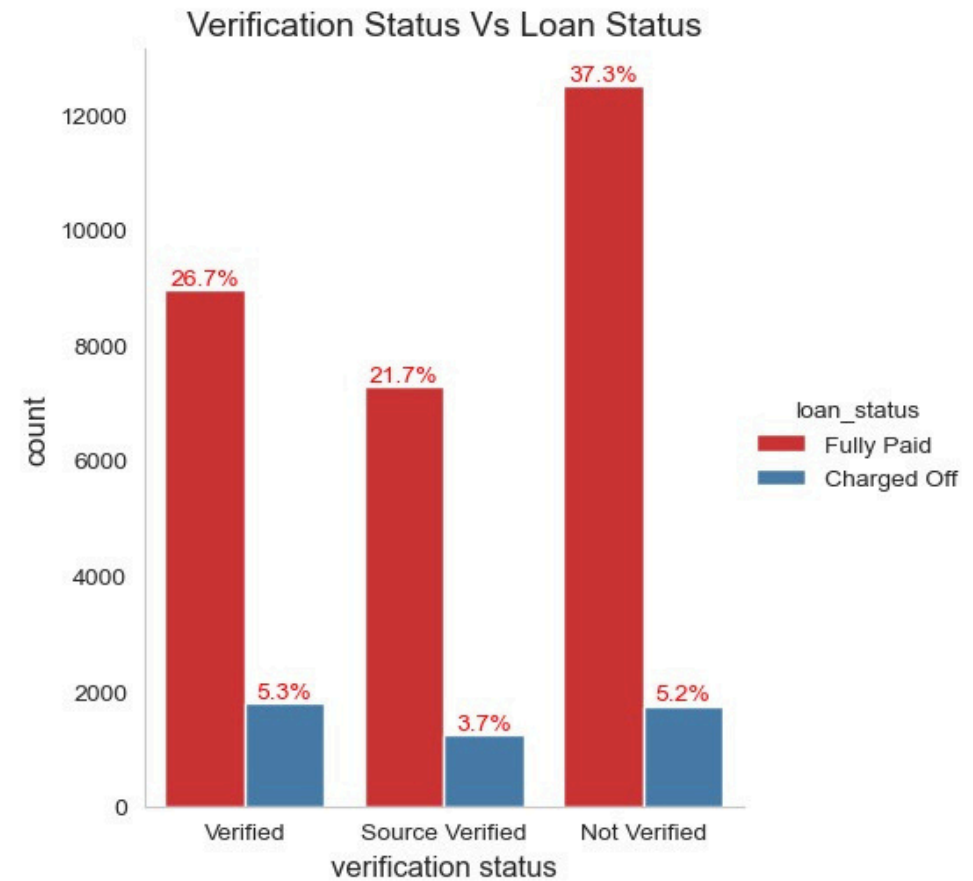
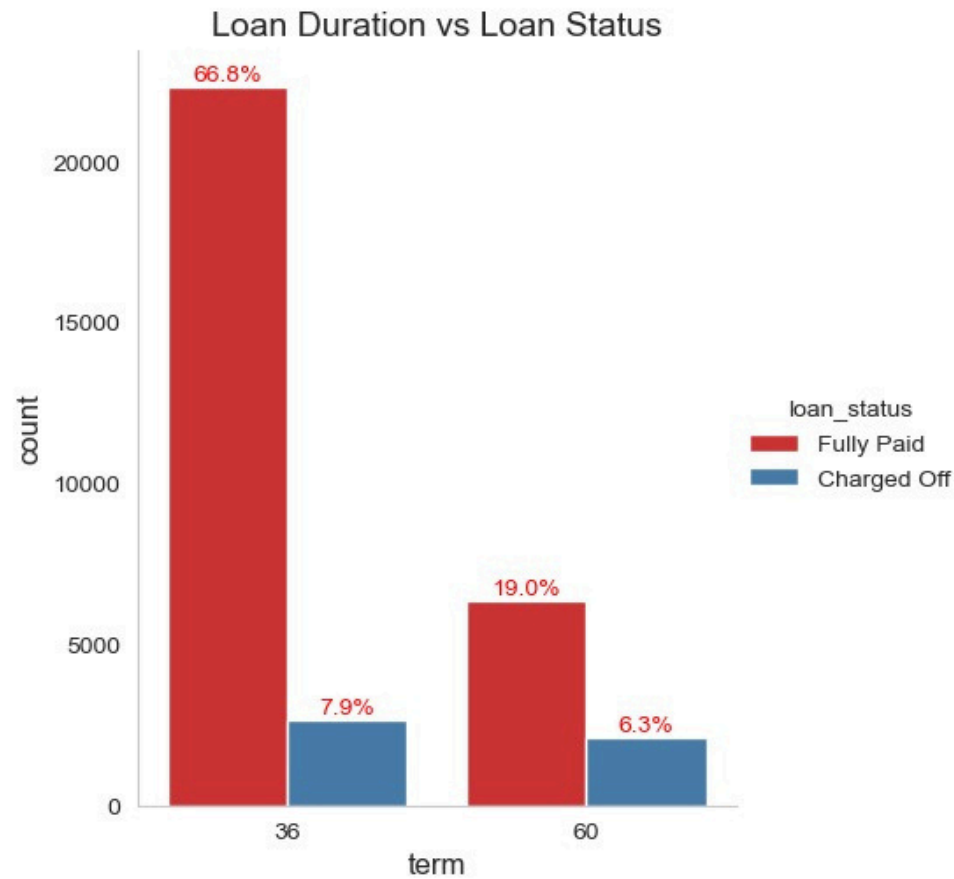
Analysis – Key Graph 6

Bivariate Analysis

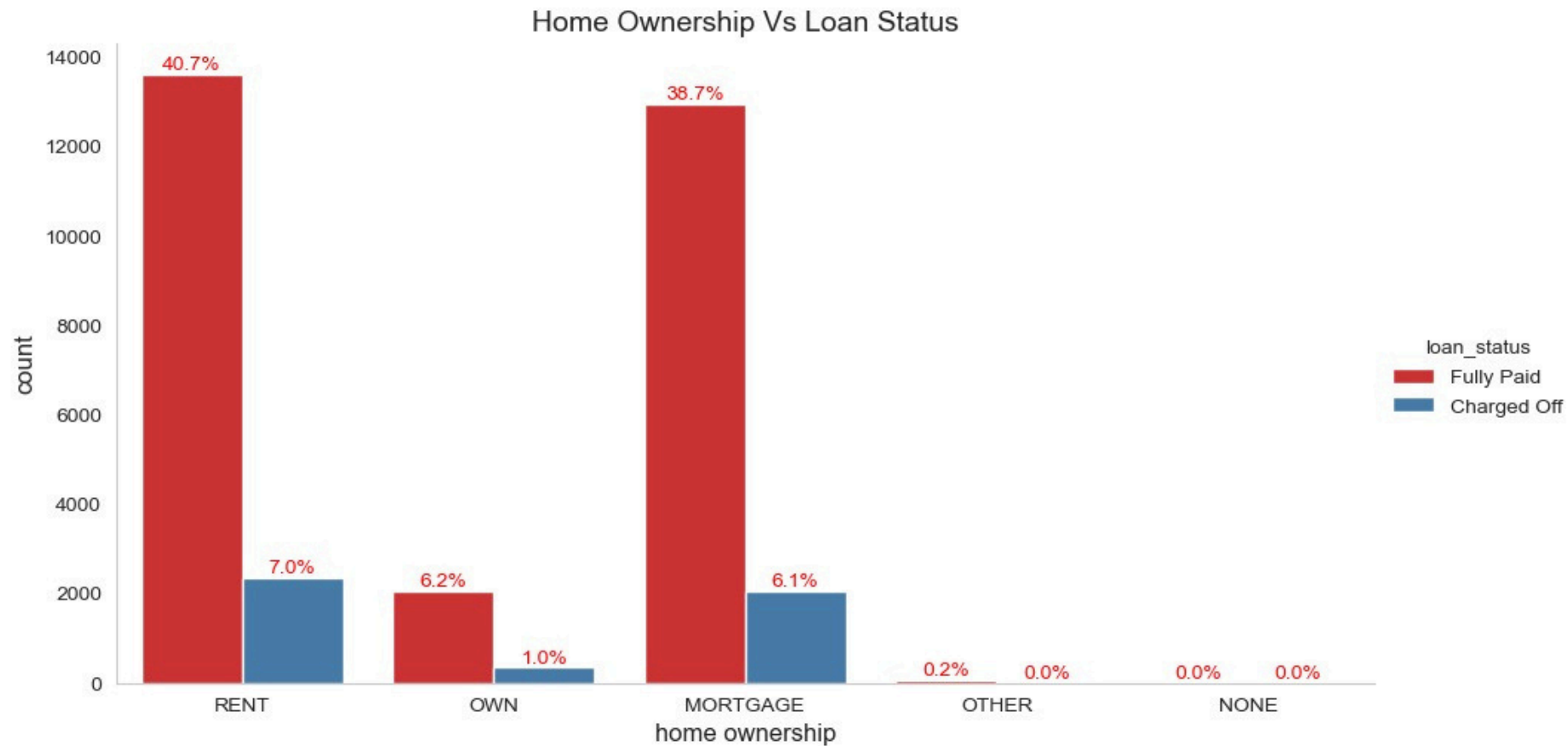


The heat map shows that the loan_amount, funded_amount and funded_amount_inv are very closely correlated. Hence we can safely take any one of the fields from the above 3 fields for our analysis.

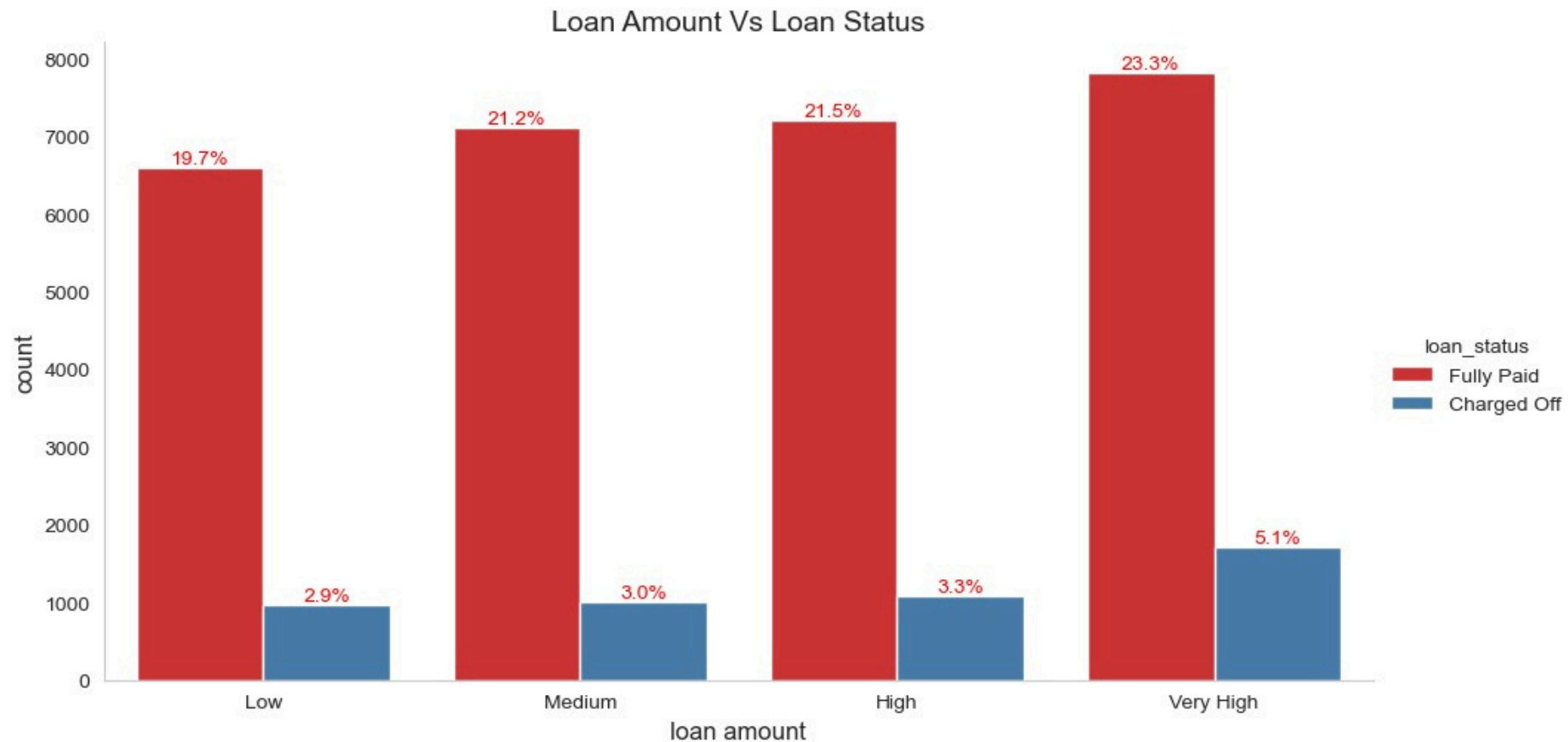
Analysis – Key Graph 7



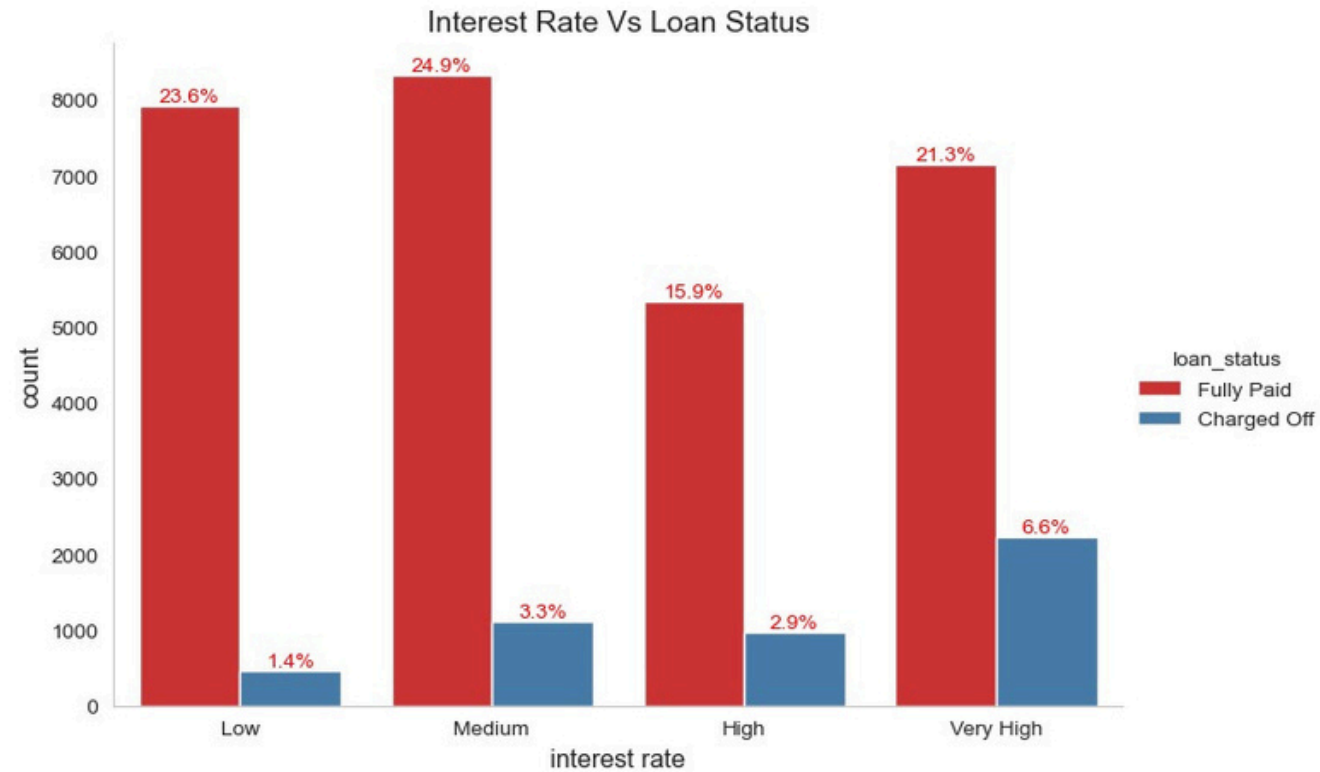
Analysis – Key Graph 8



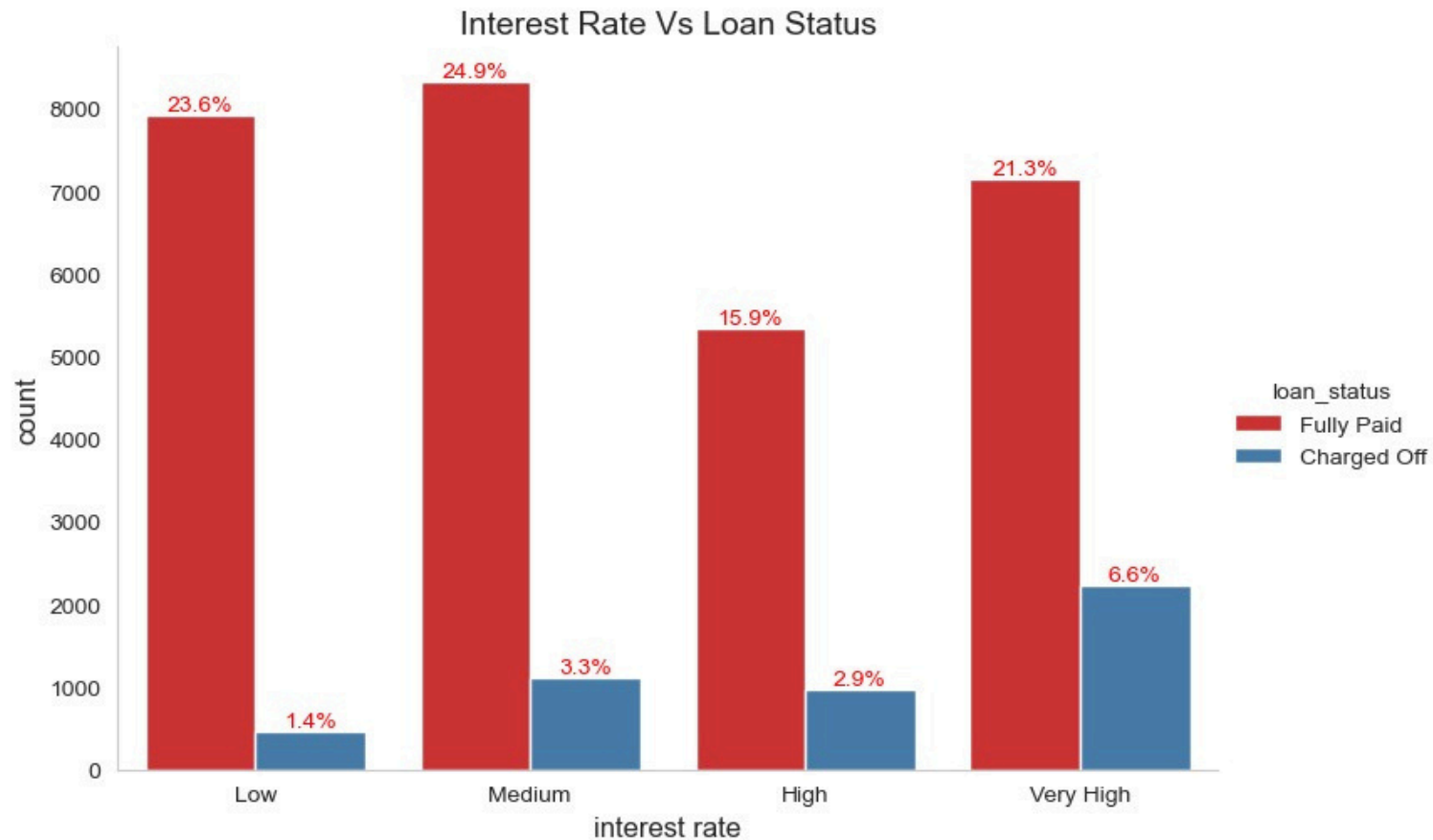
Analysis – Key Graph 9



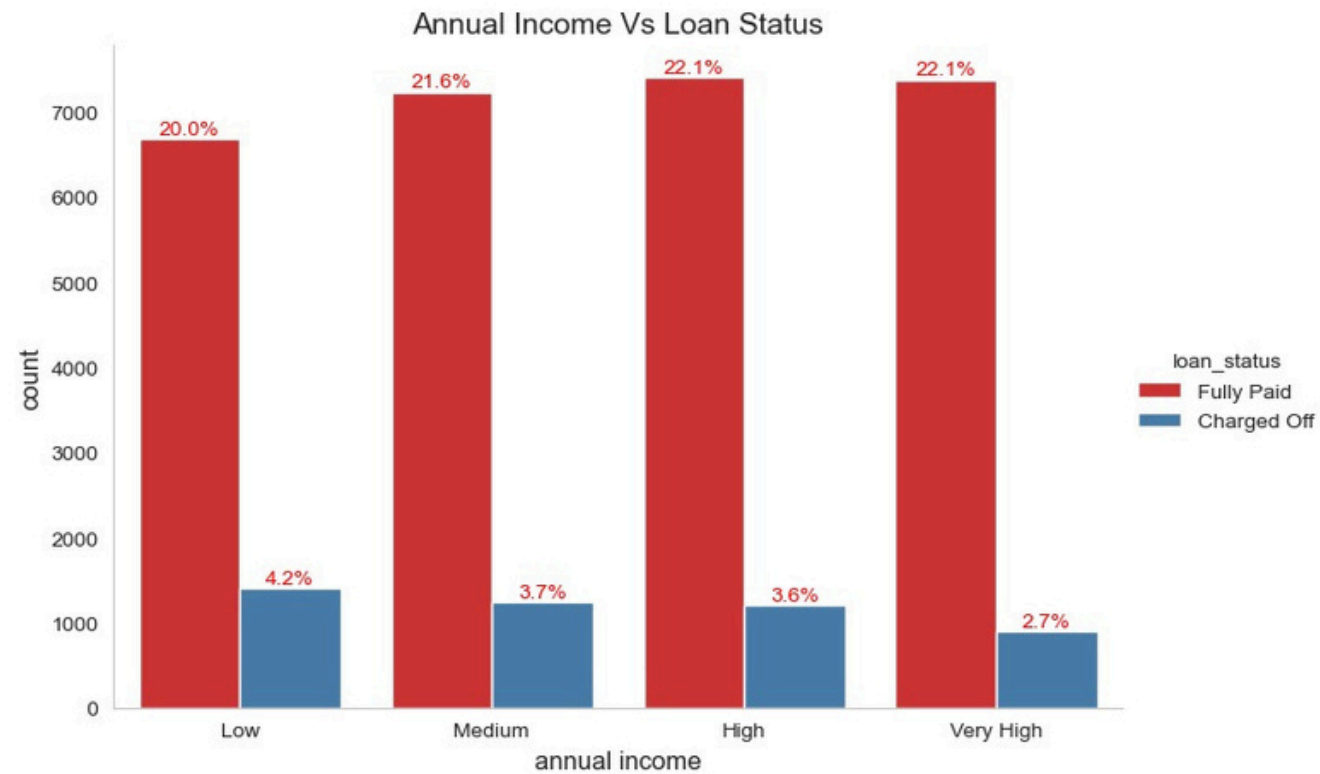
Analysis – Key Graph 10



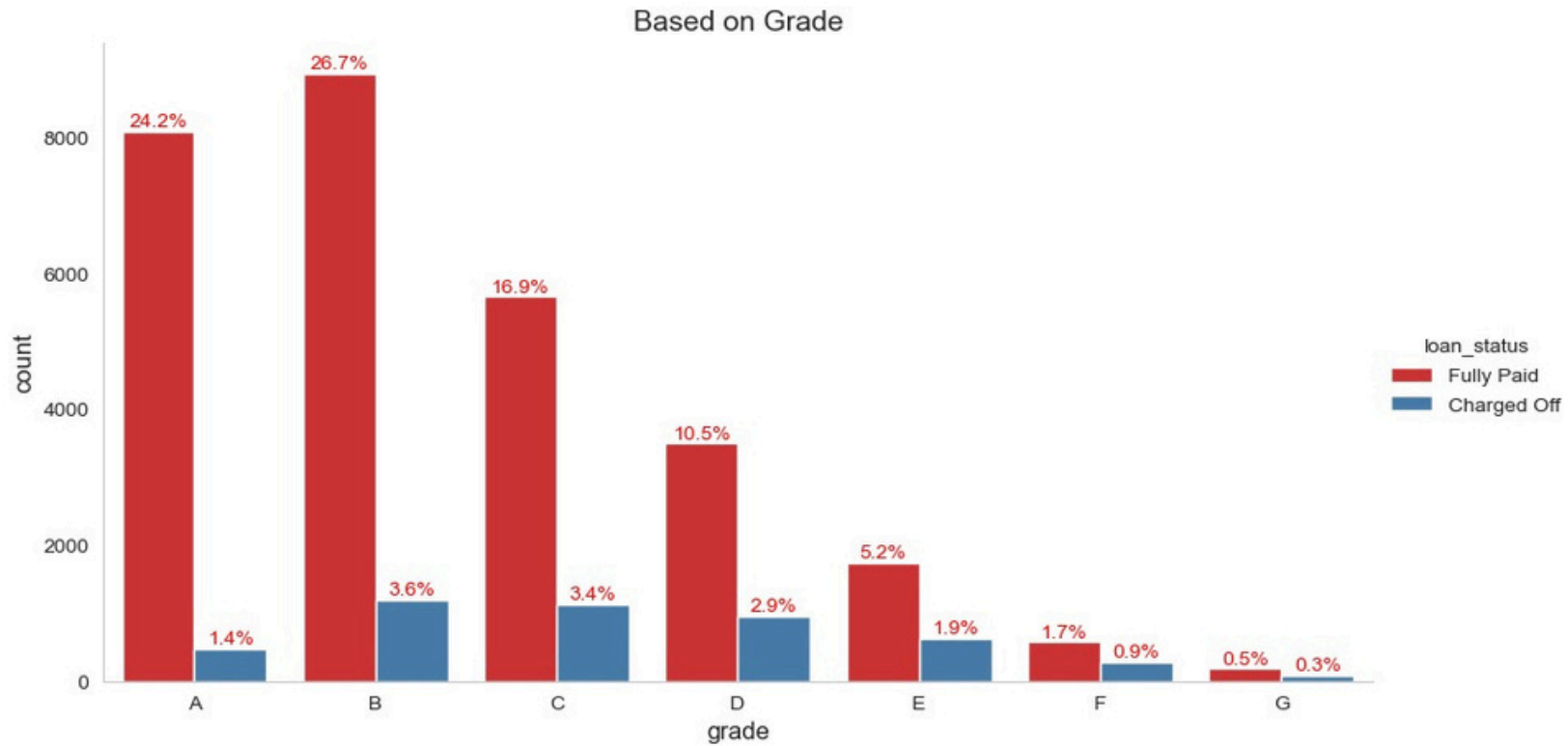
Analysis – Key Graph 11



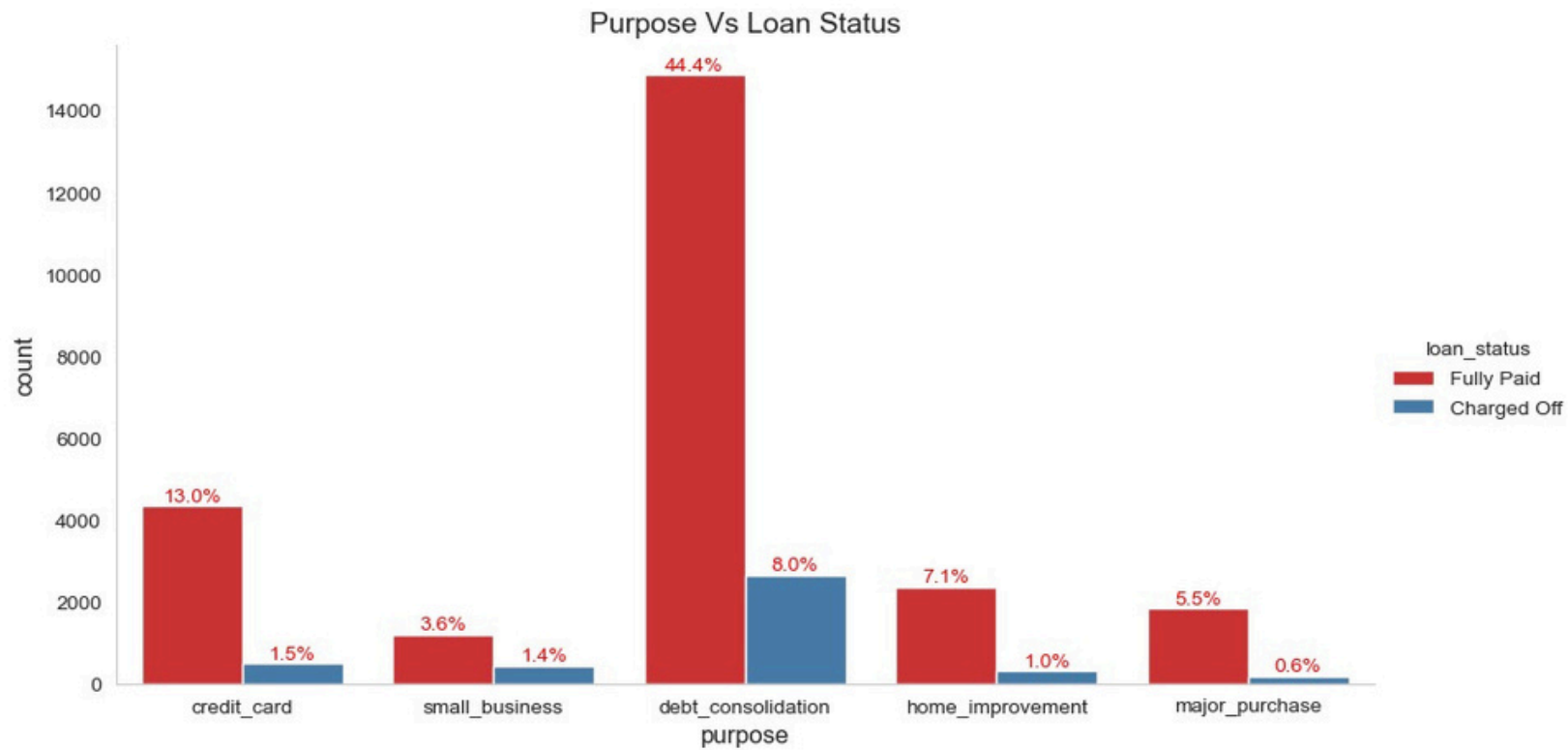
Analysis – Key Graph 12



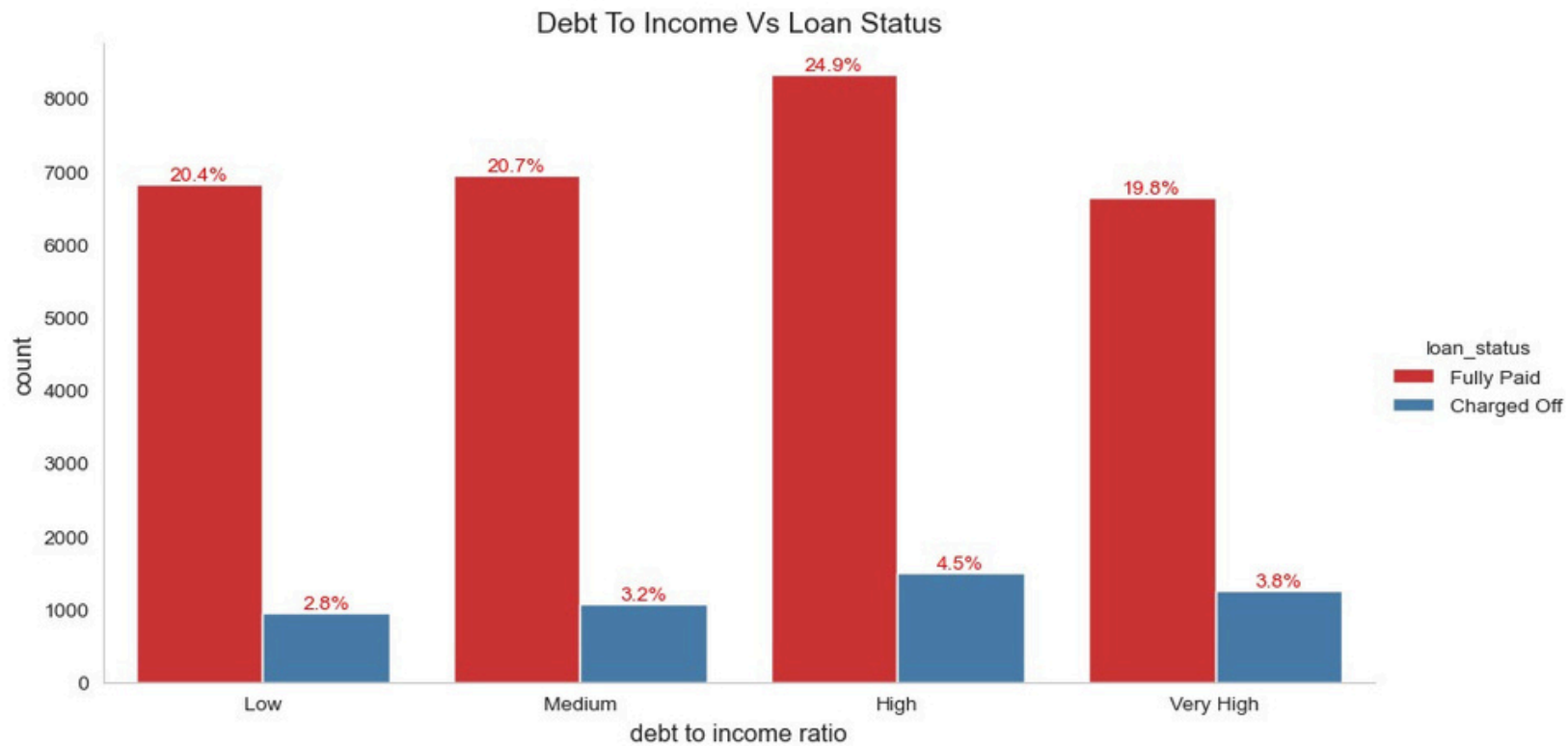
Analysis – Key Graph 13



Analysis – Key Graph 14



Analysis – Key Graph 15



THANK YOU