Gramener Case Study

Submitted by :

1.Aarti Verma

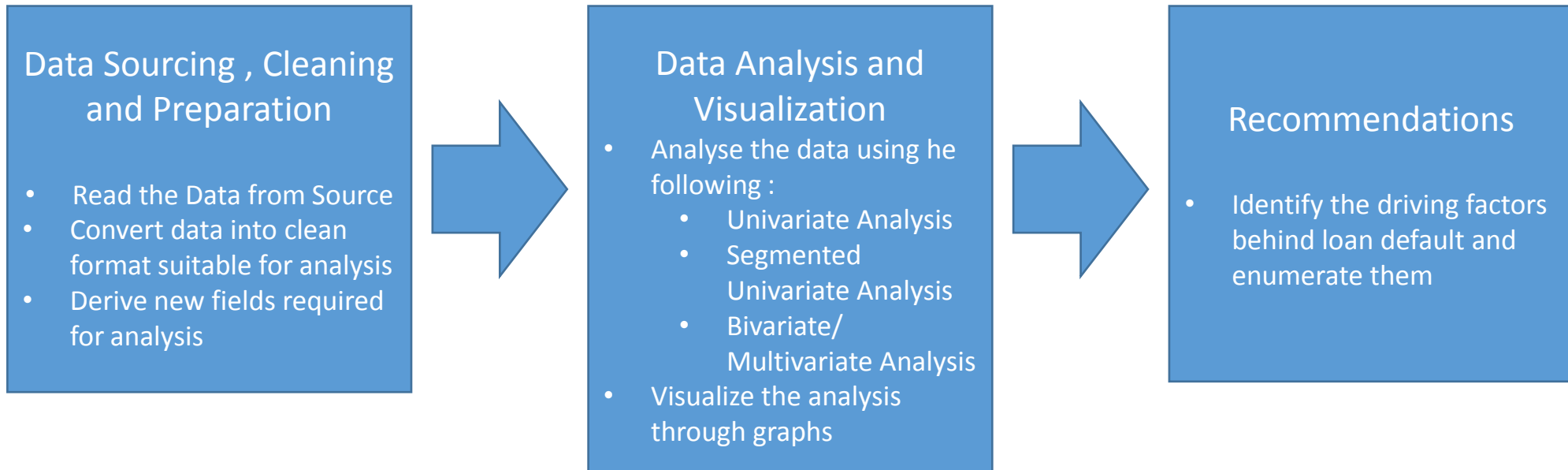2.Deepa Rajesh

3.Swagat Mishra

4.Veera Venkat Raja A

A largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures is looking at how to reduce its credit loss by identifying risky applicants
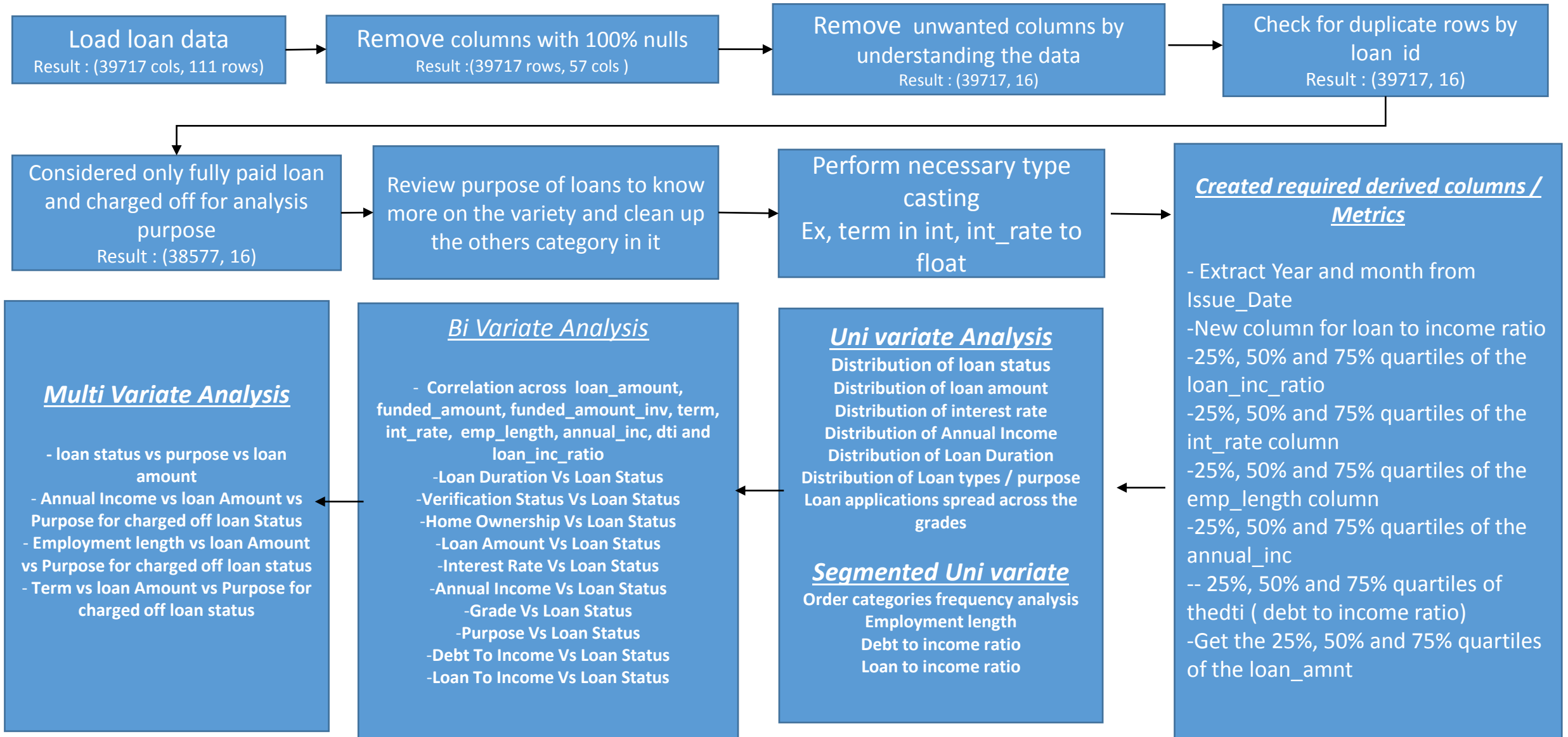
Business Objective :

- Identify the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default..

Strategy:

- Source the data for analysis
- Clean and Prepare the data.
- Derive metrics from the data set.
- Identify the driving factors responsible for loan default.
- Plot graphs to visualize the results of the analysis

Analysis Workflow

```
┌─────────────────────────┐     ┌─────────────────────────┐     ┌─────────────────────────┐     ┌─────────────────────────┐
│   Load loan data        │ ──▶ │ Remove columns with     │ ──▶ │ Remove unwanted columns │ ──▶ │ Check for duplicate rows│
│ Result : (39717 cols,   │     │ 100% nulls              │     │ by understanding the    │     │ by loan id              │
│ 111 rows)               │     │ Result :(39717 rows,    │     │ data                    │     │ Result : (39717, 16)    │
│                         │     │ 57 cols )               │     │ Result : (39717, 16)    │     │                         │
└─────────────────────────┘     └─────────────────────────┘     └─────────────────────────┘     └─────────────────────────┘
```

Load loan data
Result : (39717 cols, 111 rows)

Remove columns with 100% nulls
Result :(39717 rows, 57 cols)

Remove unwanted columns by understanding the data
Result : (39717, 16)

Check for duplicate rows by loan id
Result : (39717, 16)

Considered only fully paid loan and charged off for analysis purpose
Result : (38577, 16)

Review purpose of loans to know more on the variety and clean up the others category in it

Perform necessary type casting
Ex, term in int, int_rate to float

Created required derived columns / Metrics

- Extract Year and month from Issue_Date
- New column for loan to income ratio
- 25%, 50% and 75% quartiles of the loan_inc_ratio
- 25%, 50% and 75% quartiles of the int_rate column
- 25%, 50% and 75% quartiles of the emp_length column
- 25%, 50% and 75% quartiles of the annual_inc
- 25%, 50% and 75% quartiles of thedti (debt to income ratio)
- Get the 25%, 50% and 75% quartiles of the loan_amnt

Multi Variate Analysis

- loan status vs purpose vs loan amount
- Annual Income vs loan Amount vs Purpose for charged off loan Status
- Employment length vs loan Amount vs Purpose for charged off loan status
- Term vs loan Amount vs Purpose for charged off loan status

Bi Variate Analysis

- Correlation across loan_amount, funded_amount, funded_amount_inv, term, int_rate, emp_length, annual_inc, dti and loan_inc_ratio
- Loan Duration Vs Loan Status
- Verification Status Vs Loan Status
- Home Ownership Vs Loan Status
- Loan Amount Vs Loan Status
- Interest Rate Vs Loan Status
- Annual Income Vs Loan Status
- Grade Vs Loan Status
- Purpose Vs Loan Status
- Debt To Income Vs Loan Status
- Loan To Income Vs Loan Status

Uni variate Analysis

Distribution of loan status
Distribution of loan amount
Distribution of interest rate
Distribution of Annual Income
Distribution of Loan Duration
Distribution of Loan types / purpose
Loan applications spread across the grades

Segmented Uni variate

Order categories frequency analysis
Employment length
Debt to income ratio
Loan to income ratio

Following assumptions have been made for categorizing the interest rates, loan amount, annual income, employment length, debt to income ratio and loan to income ration.

For interest rates:

   < 9%   → Low
- 9% – 11%   → Medium
- 12% – 13%   → High
- >= 14%   → Very High

For loan amount:

- < 5400   → Low
- 5400– 9599   → Medium
- 9600– 14999 → High
- >= 15000   → Very High

For debt to income ratio

- < 8   → Low
- 8 – 12   → Medium
- 13 – 18   → High
- >= 19   → Very High

For annual income:

- < 41000   → Low
- 41000 – 59999 → Medium
- 60000 – 82999 → High
- >= 83000   → Very High

For employment length:

- < 2   → Entry Level
- 2 – 3   → Junior Level
- 4 – 8   → Middle Level
- >= 9   → Senior Level

For loan to income ratio

- < 10   → Low
- 10 – 16 → Medium
- 17 – 24 → High
- >= 25   → Very High

Inferences drawn from the below graphs are as follows :

- Around 14.2% of applicants have defaulted on the loan
- The tenure of the loans are for 2 periods - 36 and 60 months, of which 74.7% applicants have taken loan duration of 36 months
- Around 31.9% of applicants belong to the middle level category, i.e., have an employment length of around 4 – 8 years.
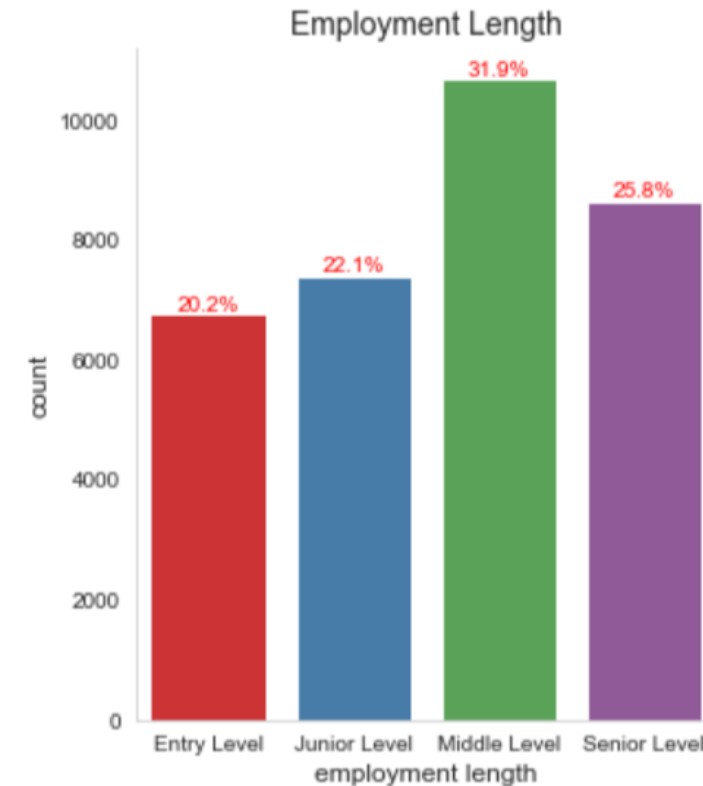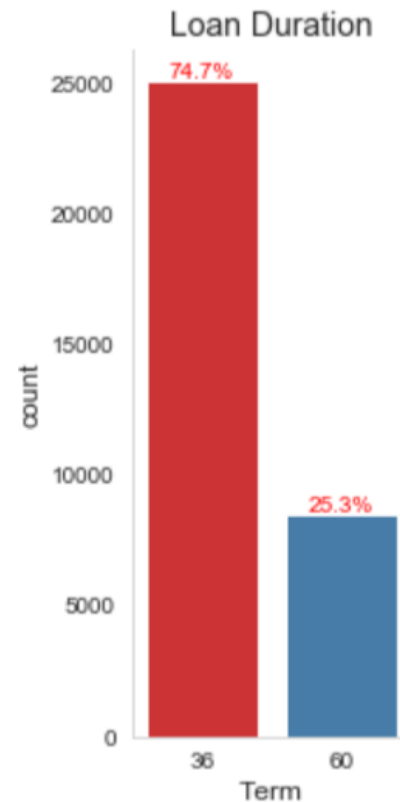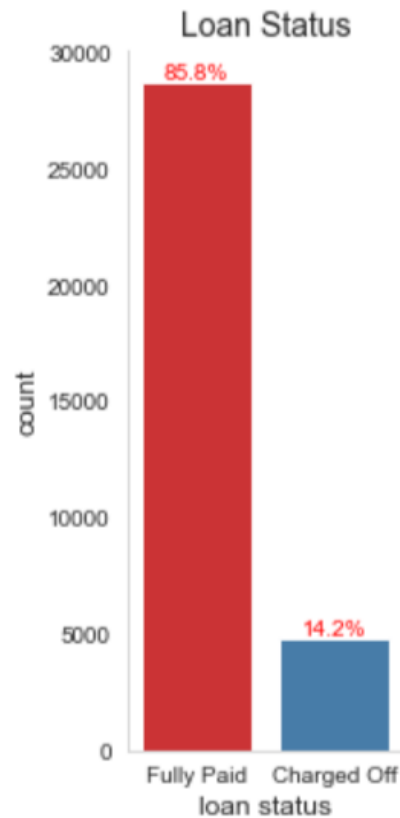
Inferences drawn from the below graphs are as follows :

- The loan amount is majorly spread around say 6000 to 18000 approximately.
- The interest rate is spread majorly between 8% and 14% approximately
- Majority of the applicants have an annual income ranging approximately between 40000 USD to 90000 USD

Analysis Of Purpose of Loan
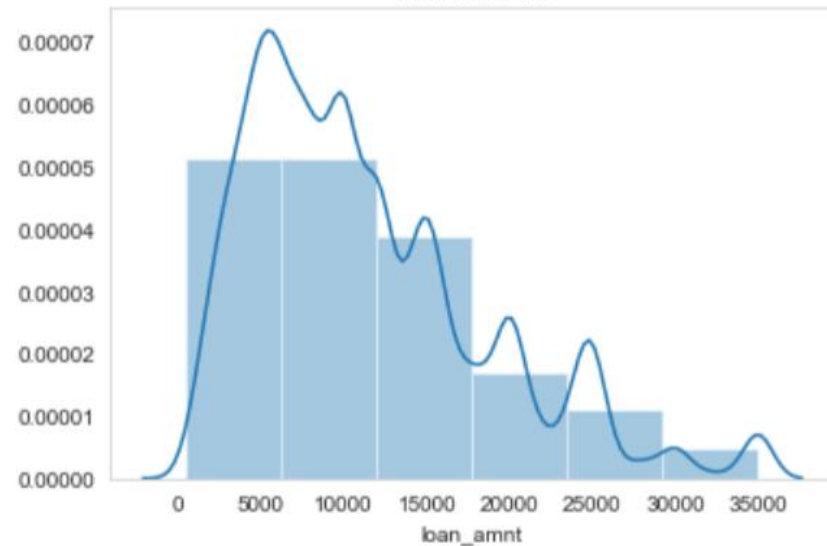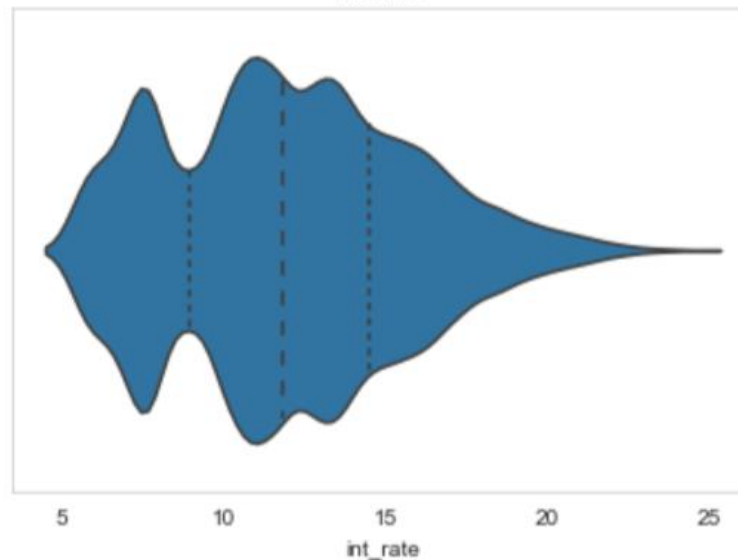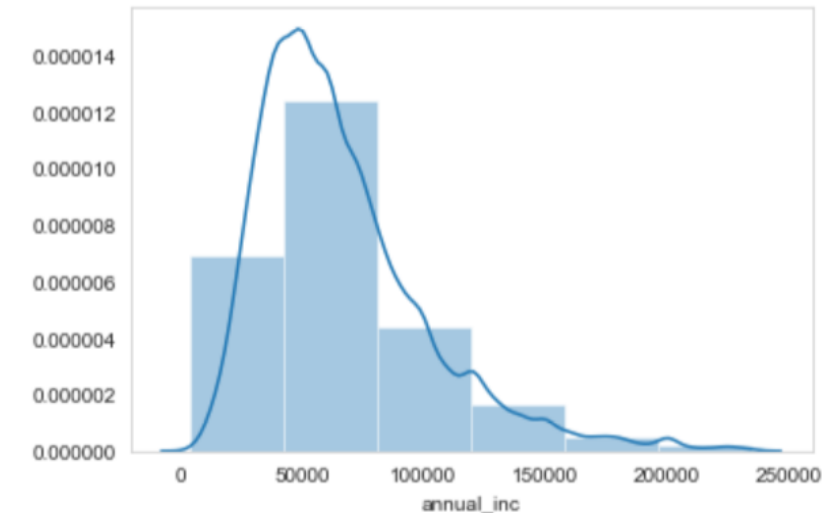
The following are the purposes for which more than 5% applicants have taken loan:

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

Purpose of Loan
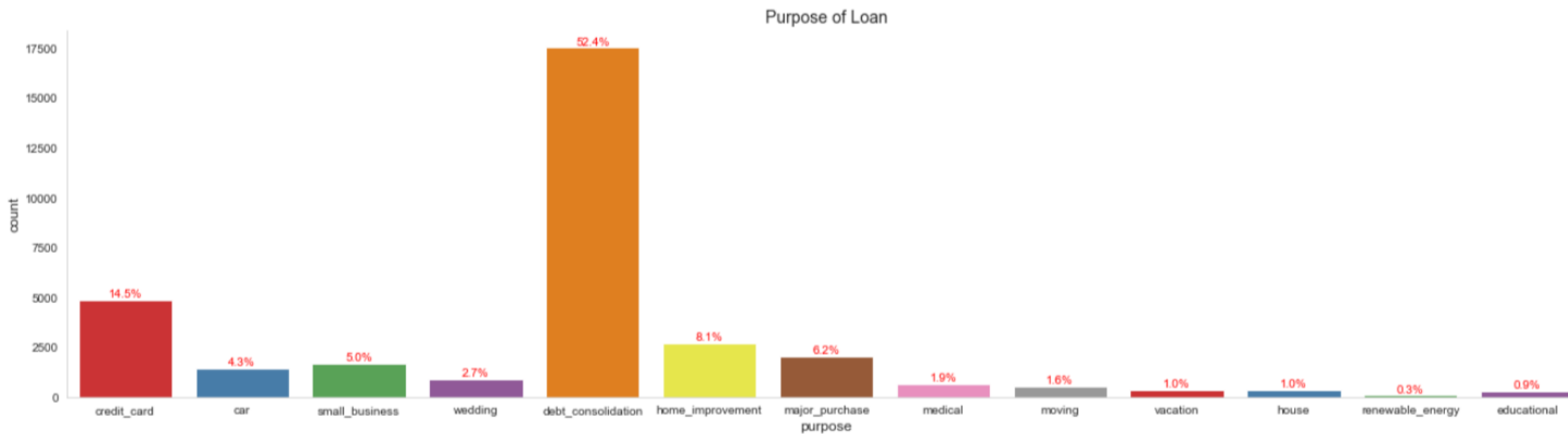
The above graph shows that most of the applicants fall under the grade B(30.3%), followed by A(25.6%) and C(20.3%)
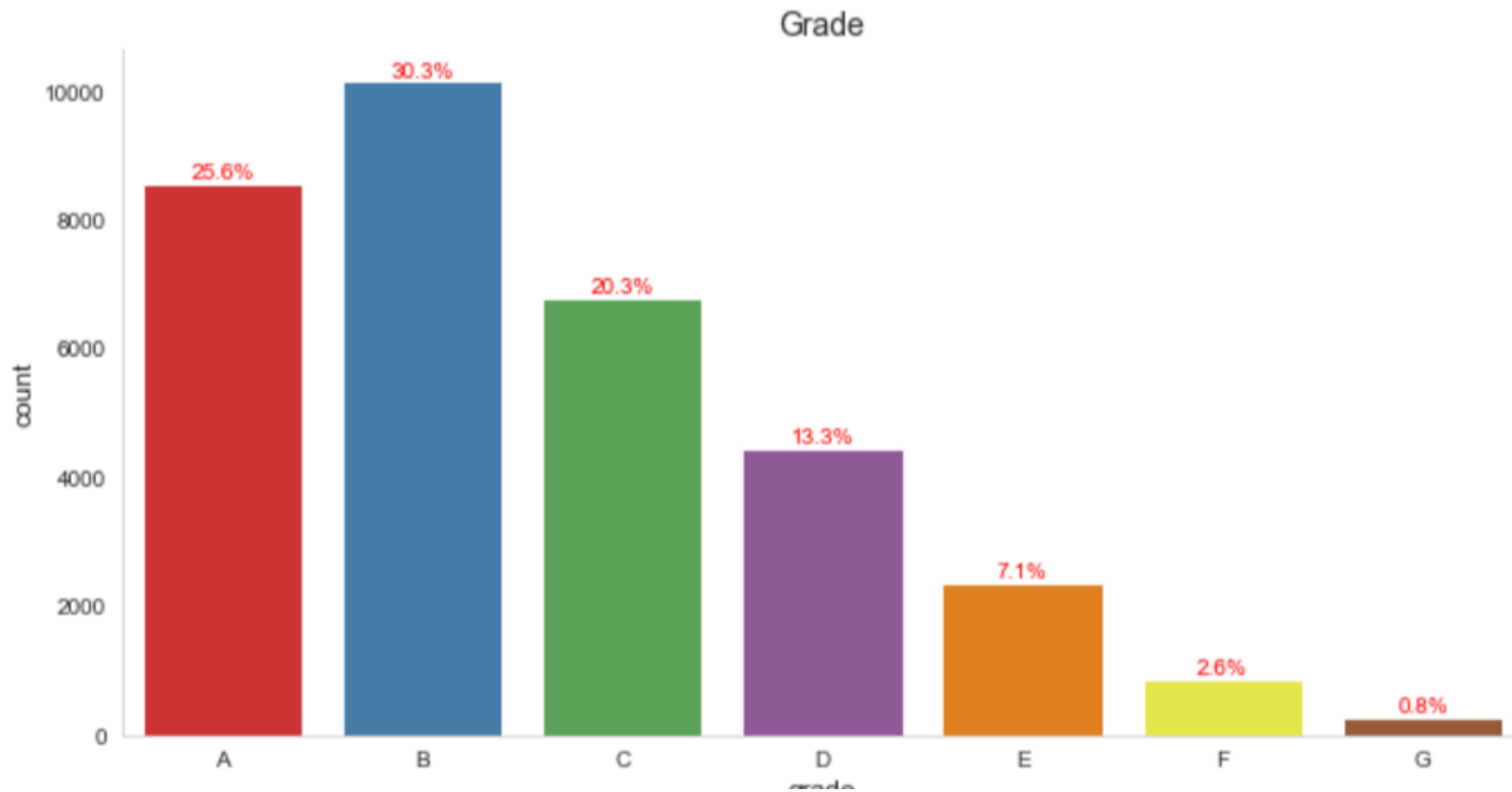
Following can be inferred from the below plots:

- The defaulters are more for 36 month loan duration – 7.9%
- Applicants default when the loan amount is increasing. In this case major defaults happen for Loan amount of 15000 and above

Analysis of Interest Rate Vs Loan Status and Annual Income Vs Loan Status
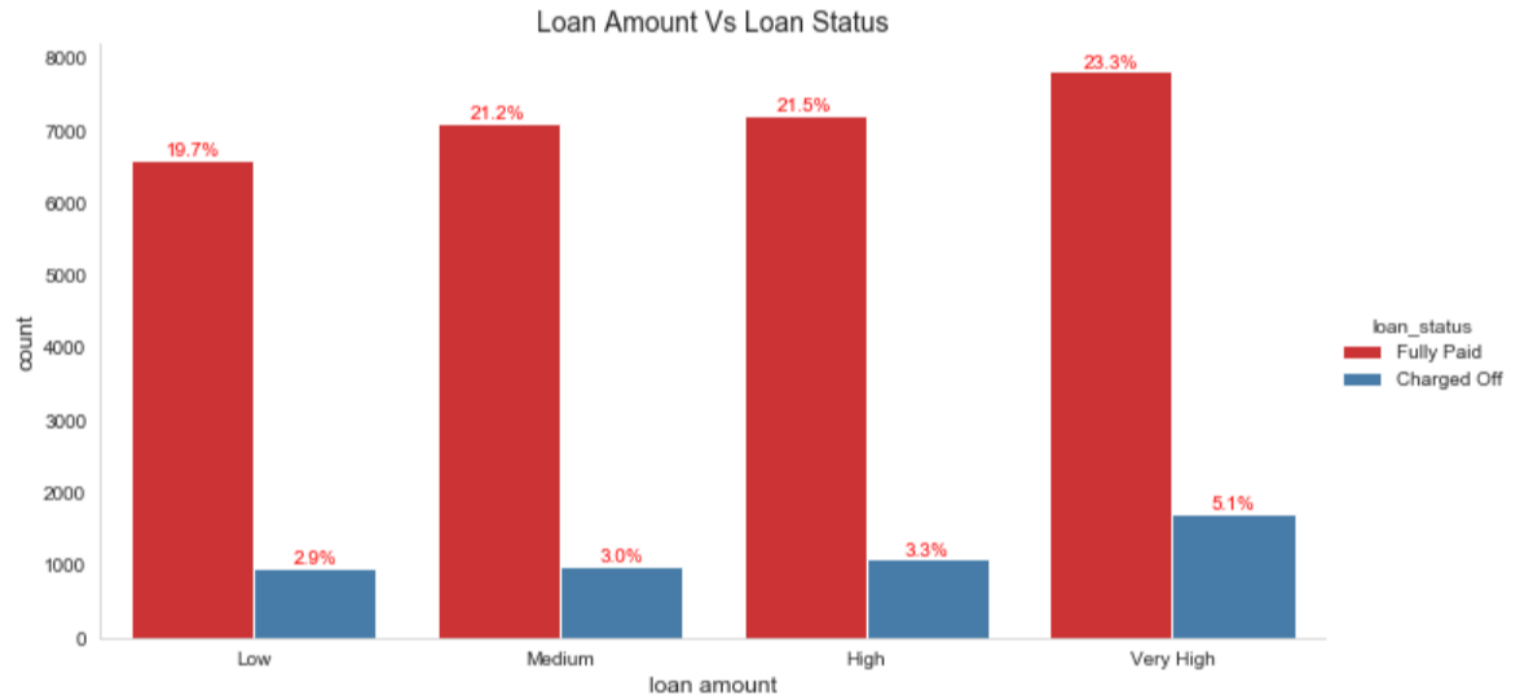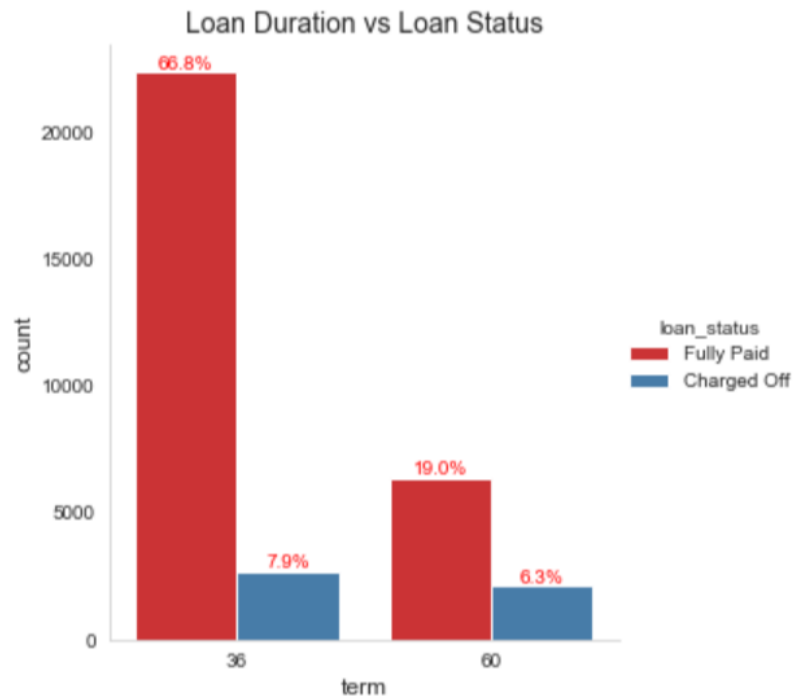
Following can be inferred from the below plots:

- The defaulters are more for interest rates of 14% and above
- Annual Income is inversely proportional to loan default i.e., as the income increases the loan default decreases. Loan default is very high with the low annual income group i.e., less than $41000
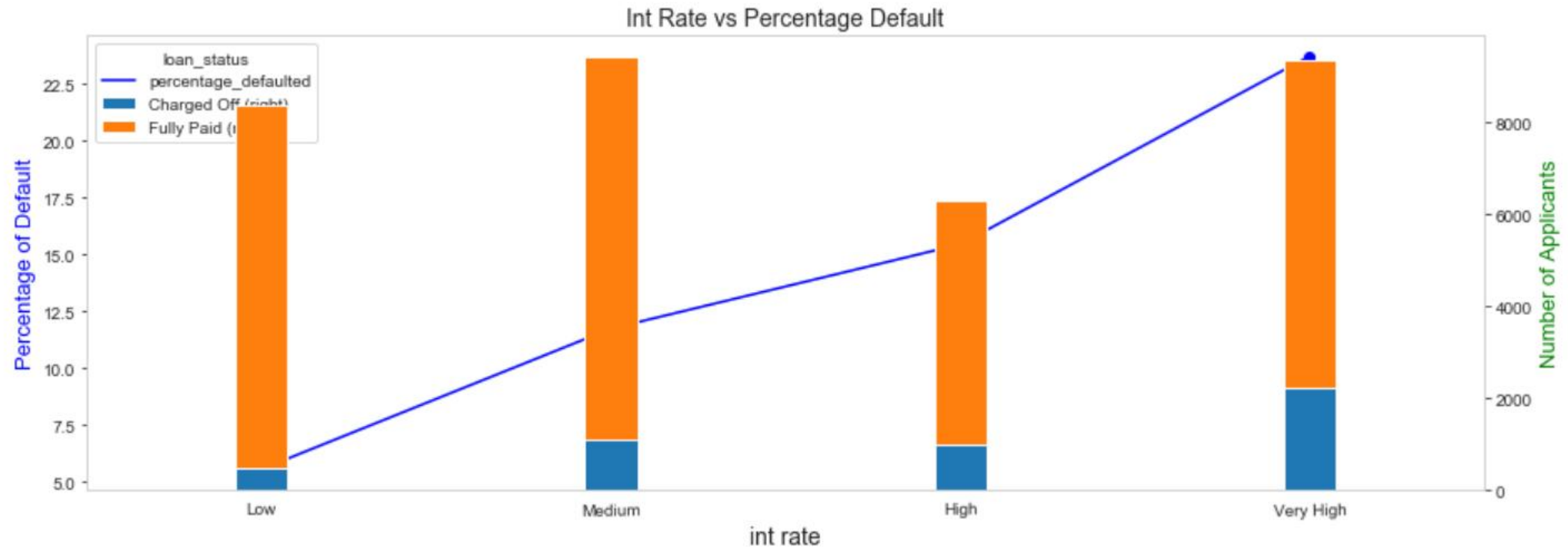
Analysis of probability of applicant default for small business
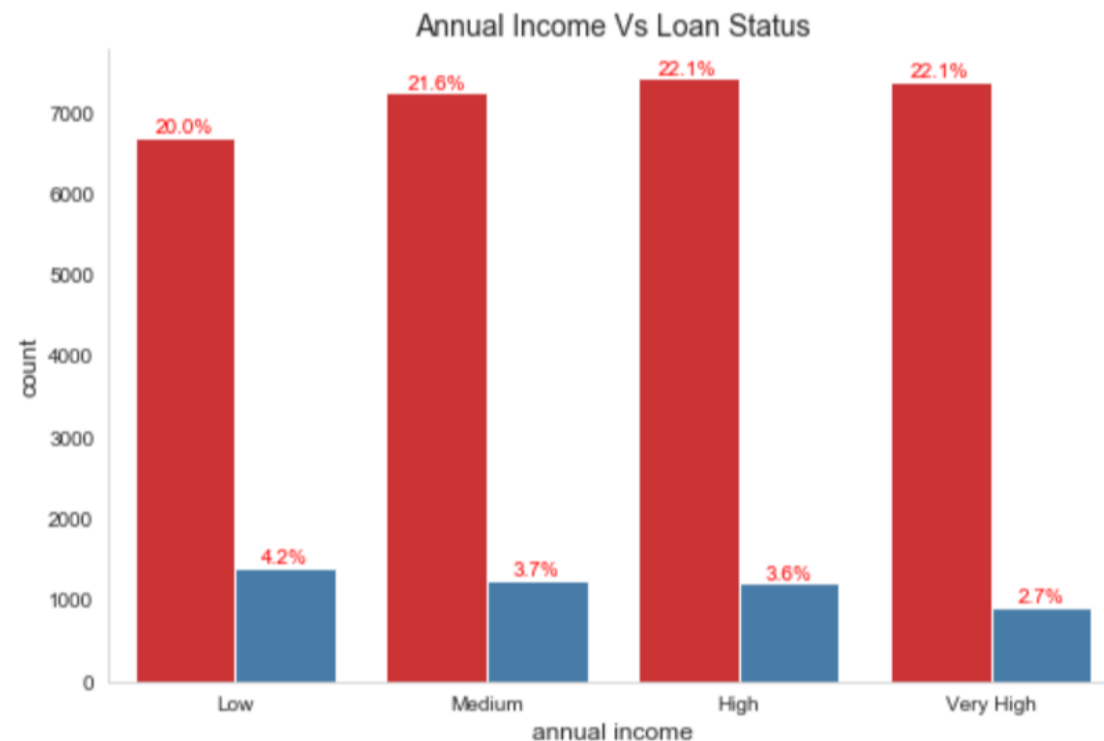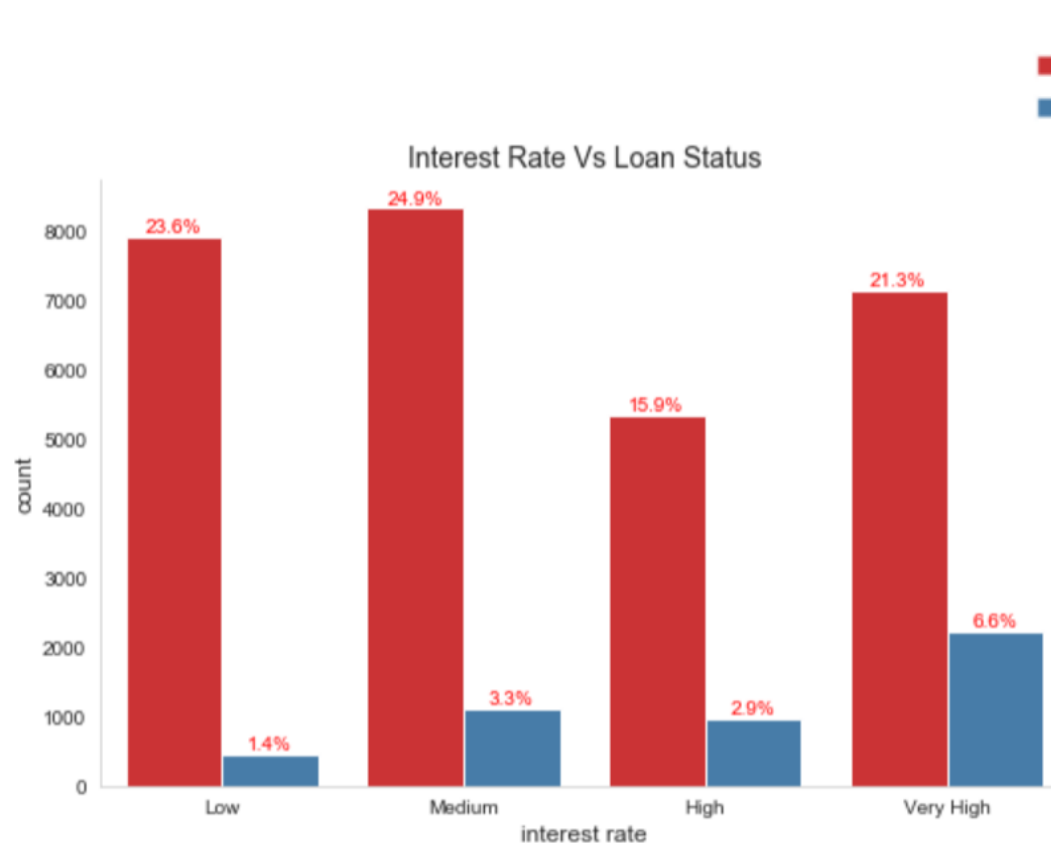
- More number of applicants have taken loan for Small Business followed by Debt consolidation.
- Applicants who have taken loan for small business (27.31%) tend to default more.
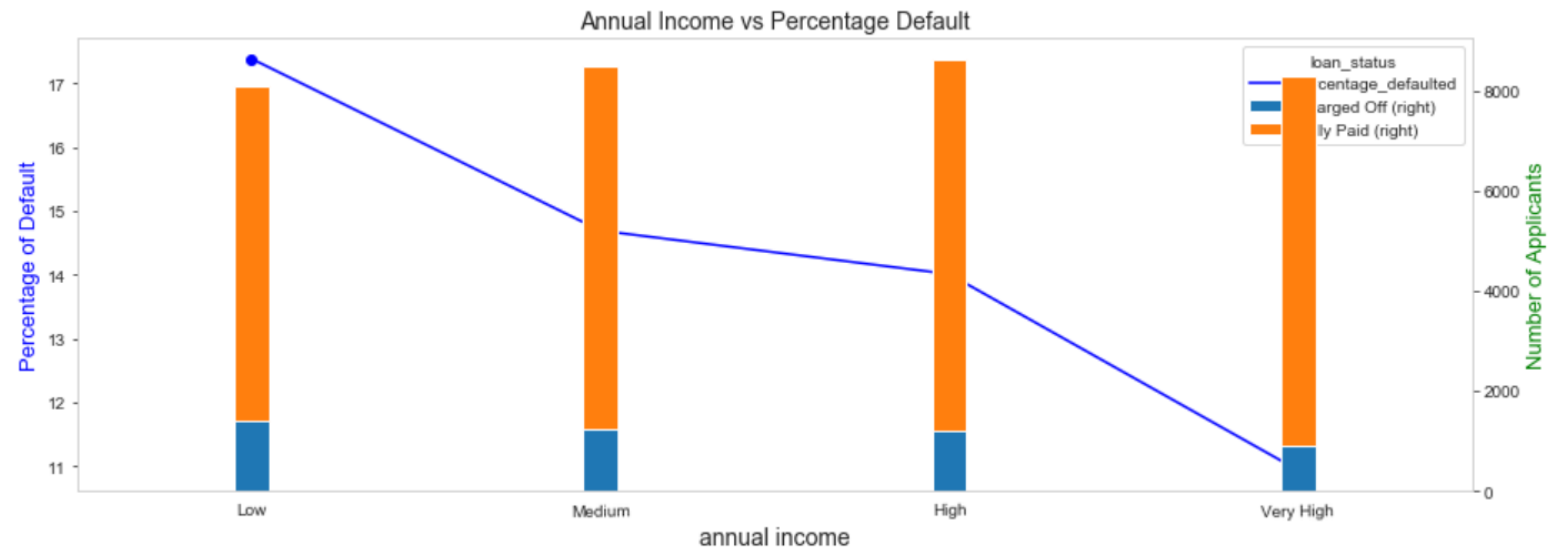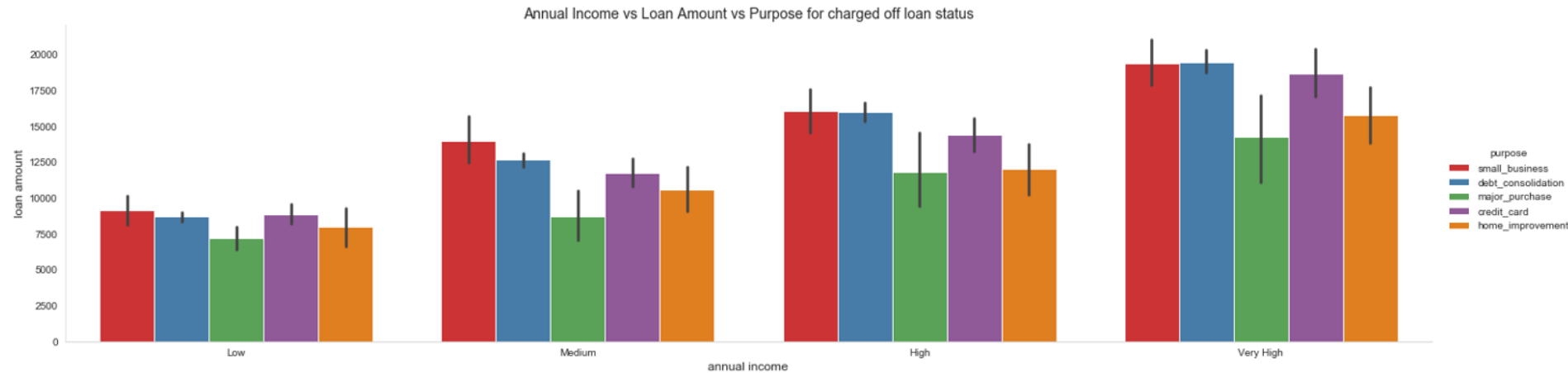

Loan Amount vs Purpose Vs Loan_Status

Purpose vs Percentage Default

Analysis of Term Vs Loan Amount Vs Purpose for Charged off Loans
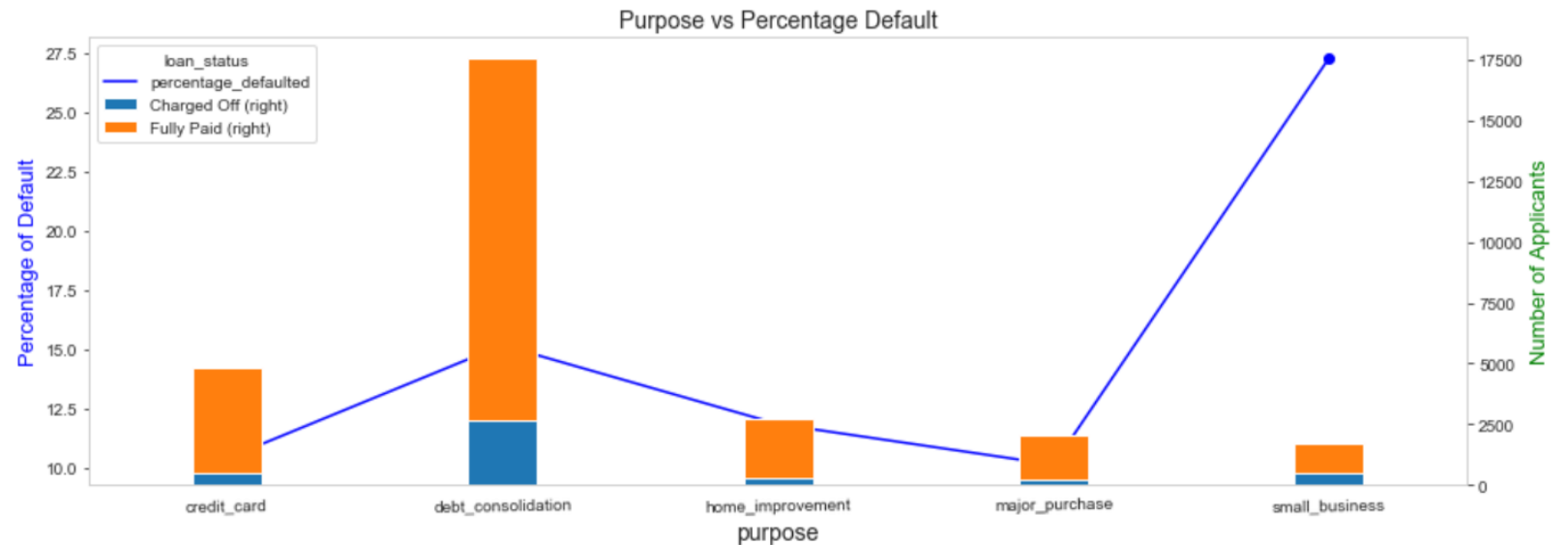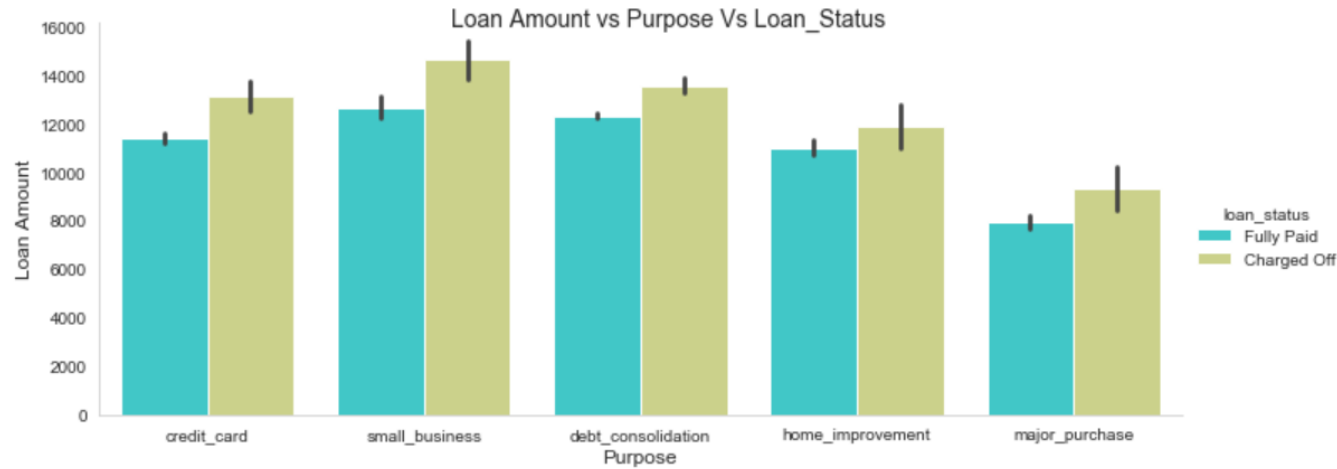
More number of applicants have taken a duration of 60 months period for small business loan.

Term vs Loan Amount vs Purpose for charged off loan status
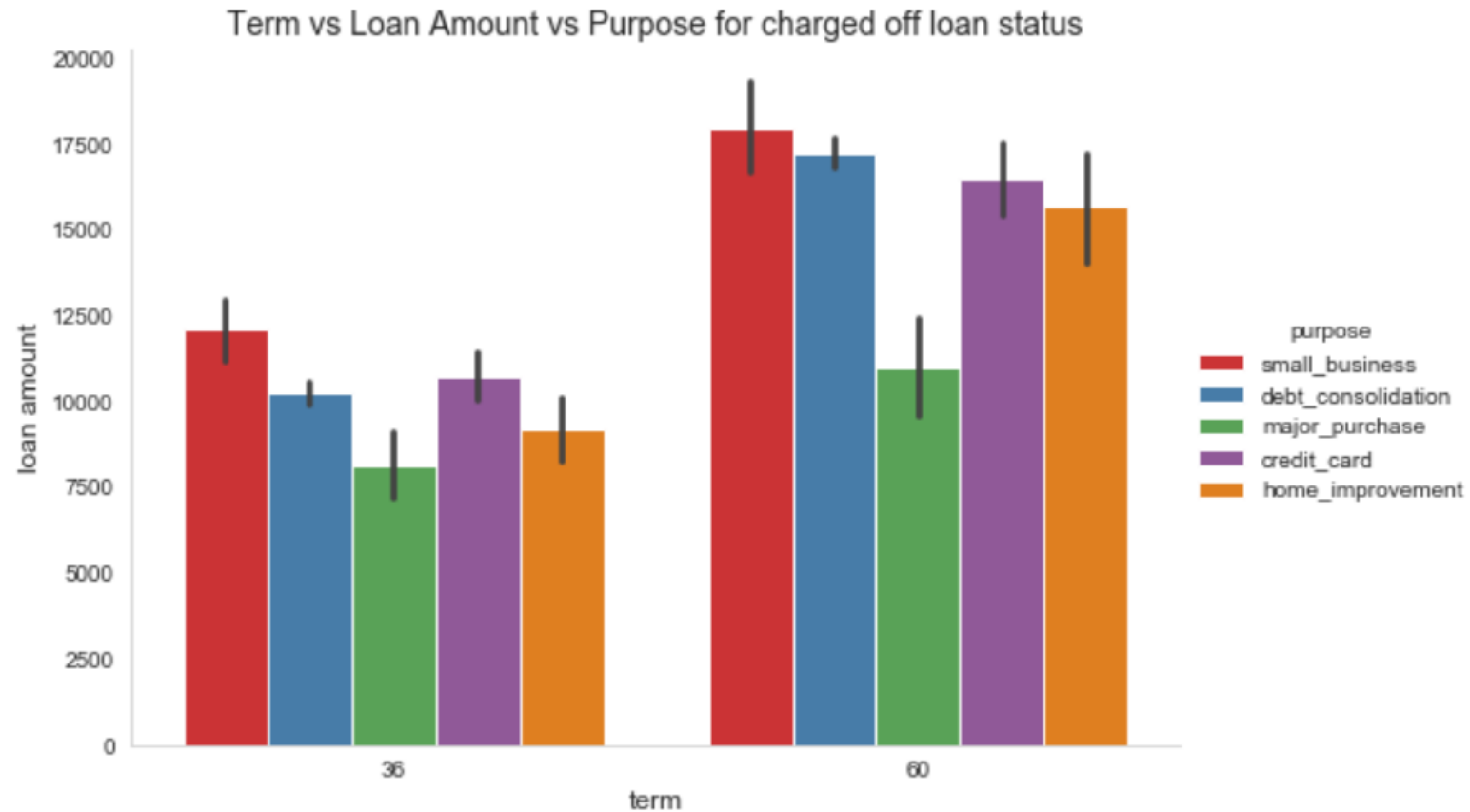
- Higher the interest rate, higher is the loan default
- Applicants who have taken a very high interest (>= 14%) rates are bound to default more.


Int Rate vs Percentage Default

- Number of applicants who have taken loan for small business is more across various annual income group. However, applicants under the category of high income have taken more loan for debt consolidation followed by small business.
- Default rate is highest amongst the low annual income (17.38%) group i.e. whose annual income is < 41000 USD per annum.

Annual Income vs Loan Amount vs Purpose for charged off loan status

Annual Income vs Percentage Default

# Analysis of probability of applicant default based on Employment length

Following can be inferred from the below plots:

- There are more number of applicants who have taken loan for small business across all the employment length group.
- Applicants in the Junior level i.e. with a tenure of around 2 to 4 years are a safe bet for giving loans as they do not seem to be defaulting much.

Analysis of probability of applicant default based on Grade
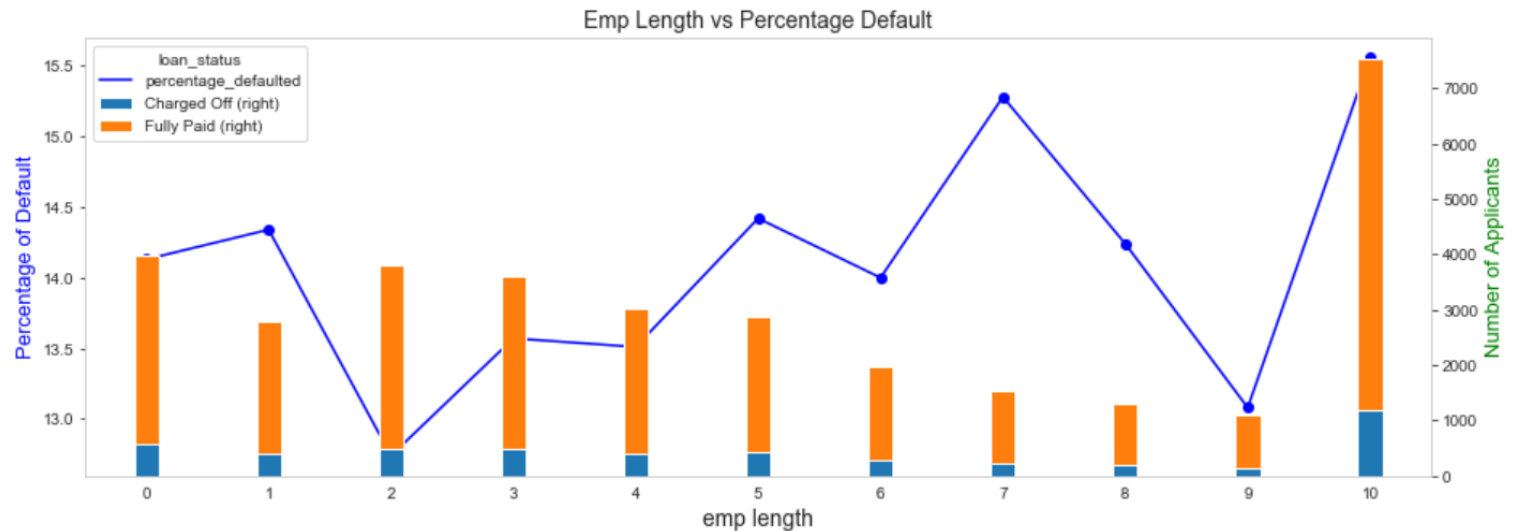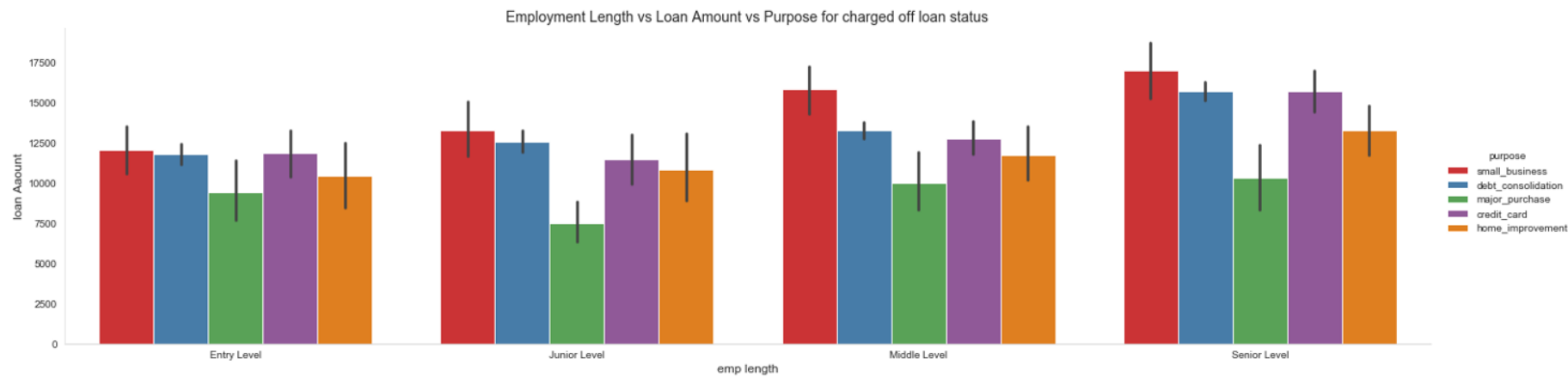
Probability that the applicant would default increases from grade A to G
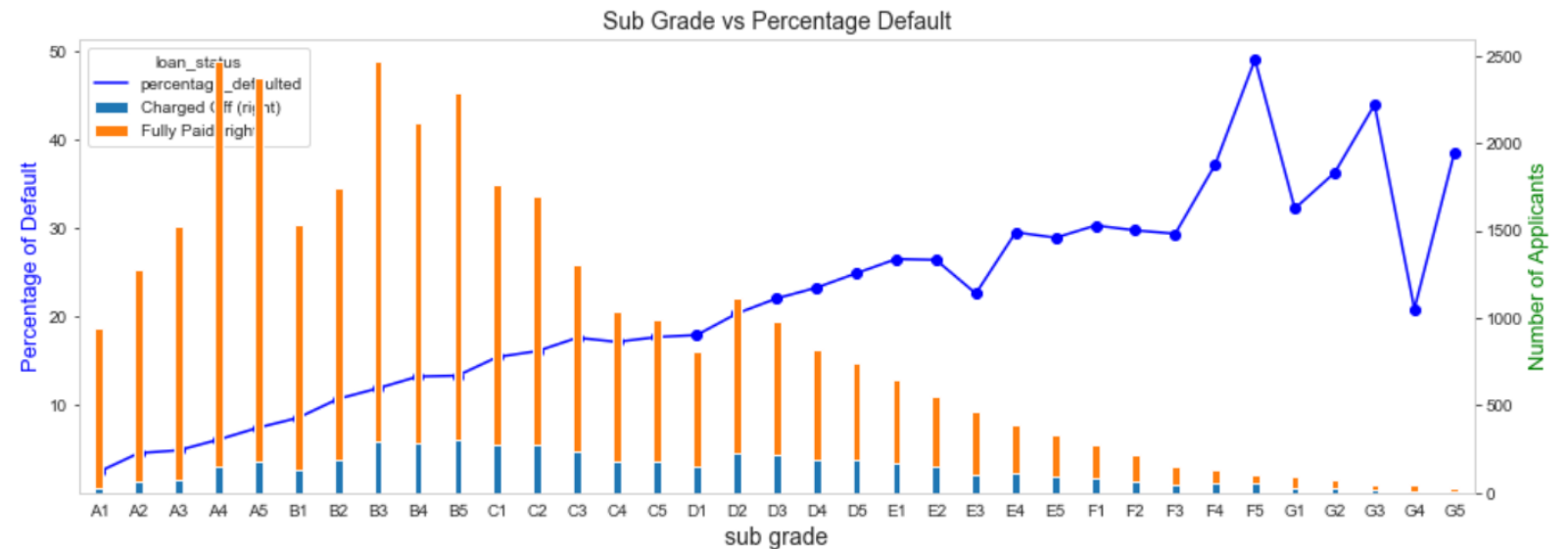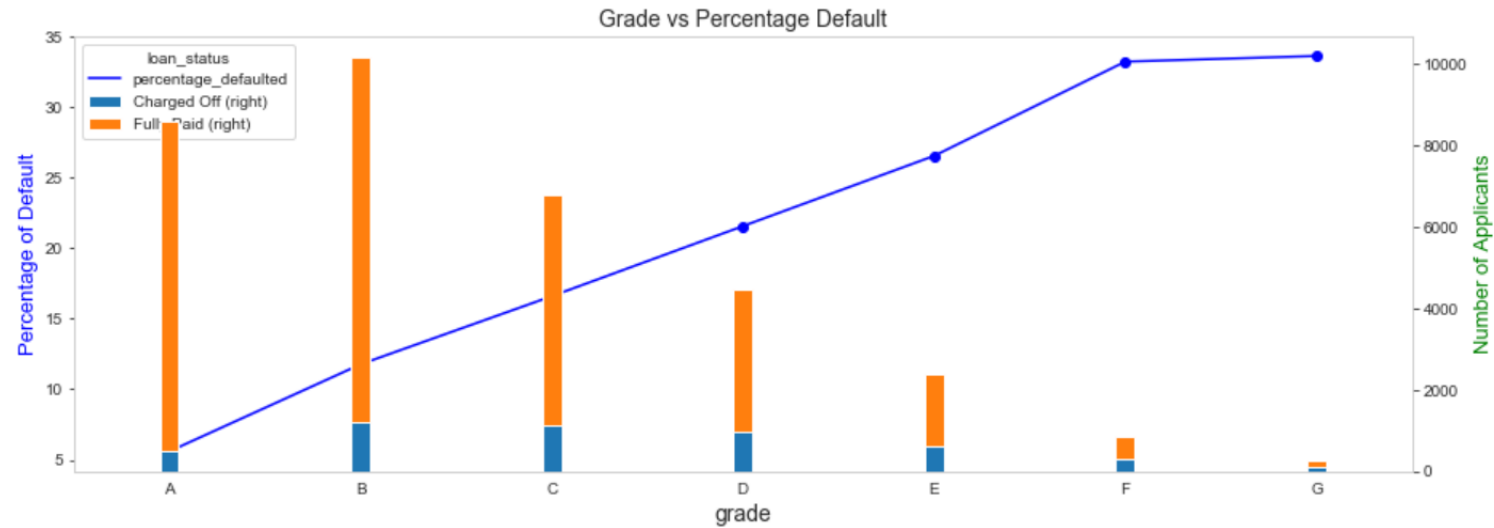
- Following are the main parameters, taken into consideration for arriving at the analysis conclusion. Loan Status and Loan amount have been compared with the below parameters to arrive at the analysis conclusion
    - Purpose for which the loan is taken
    - Annual Income of the loan applicants
    - Loan duration (term)
    - Grade allocated by the LC
    - Employment Length of the loan applicants
    - Interest Rate of the loan amount

- As per the analysis, it can be inferred that applicants who are of low income group and have taken a high interest loan with longer duration for small business have more probability of defaulting. Hence extra care should be taken before lending them loans.