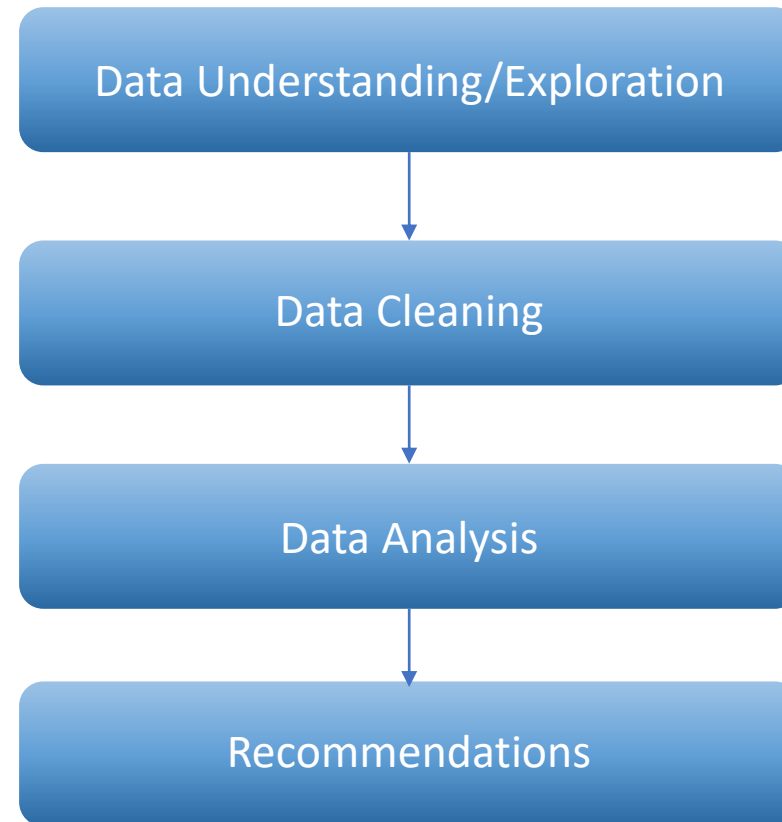# LENDING CLUB EDA ASSIGNMENT

# SUBMISSION

- R PRAVEEN

Objective

To Analyze and identify patterns from loan application data that would indicate if a borrower of loan is likely to default, which in turn may be used for taking corrective actions which includes denying the loan, reducing the amount of loan, lending at a higher interest rate etc..

Analysis Pipeline

```
┌─────────────────────────────────────┐
│   Data Understanding/Exploration     │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│            Data Cleaning             │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│            Data Analysis             │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│          Recommendations             │
└─────────────────────────────────────┘
```

General overview of Data

- It turns out that among the total no of loans accepted about 15% of them are charged off the rest are fully paid.

- Most loan amounts lent are in the range 2K to 16K.

- Among the amounts lent, most of the loans are having interest range between 5% to 10%
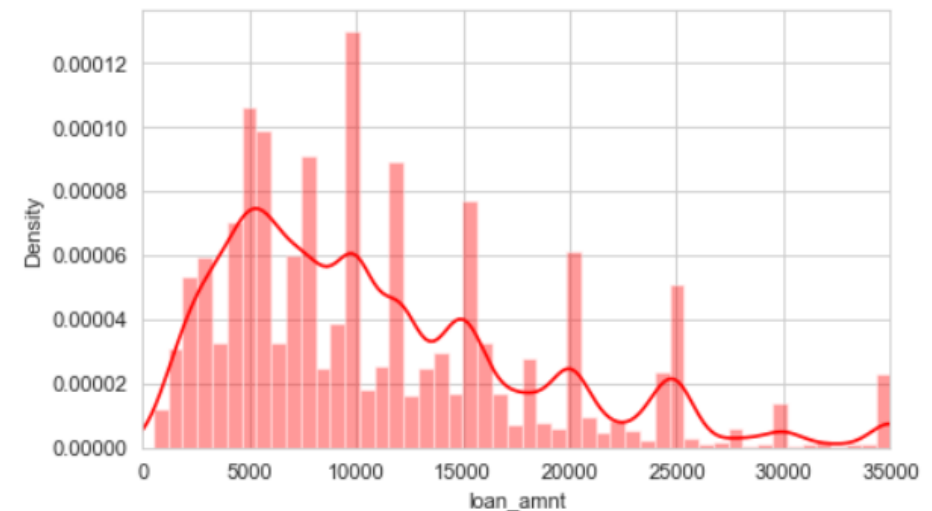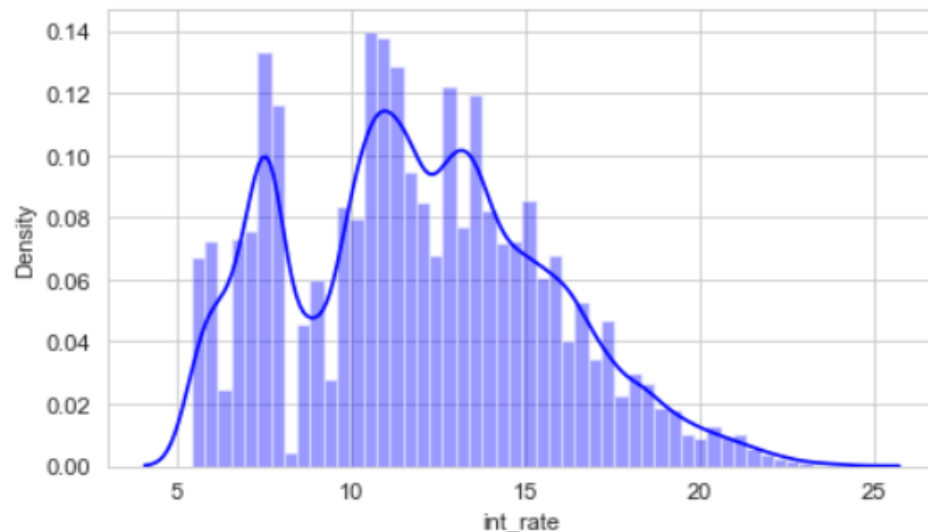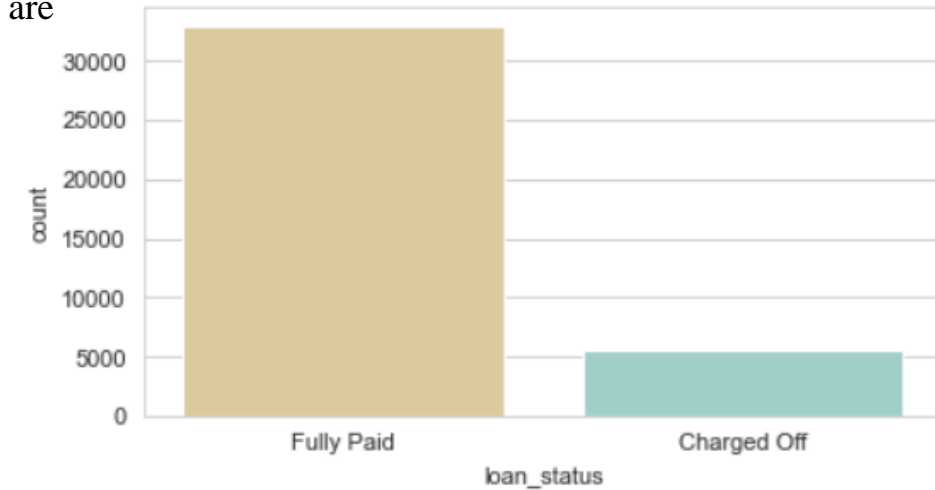
General overview of Data

- Maximum number of loans are for debt consolidation, followed by credit card.

- Most loans are having high grade A and B.

- Majority of the accepted loans are borrowed by borrowers with employment length of either more than 10 yrs or 0-3 years.
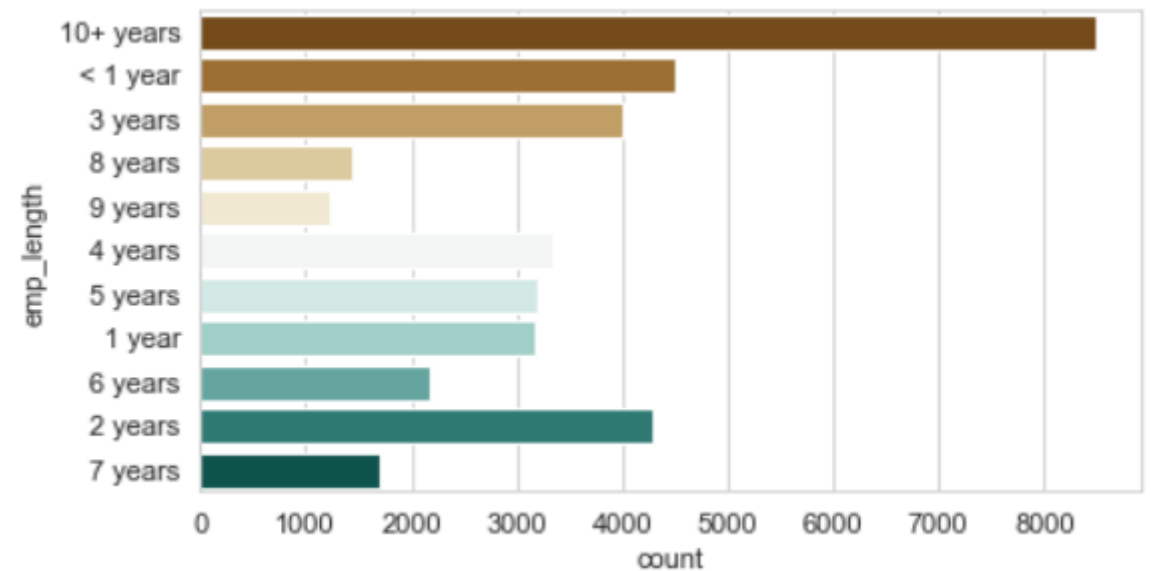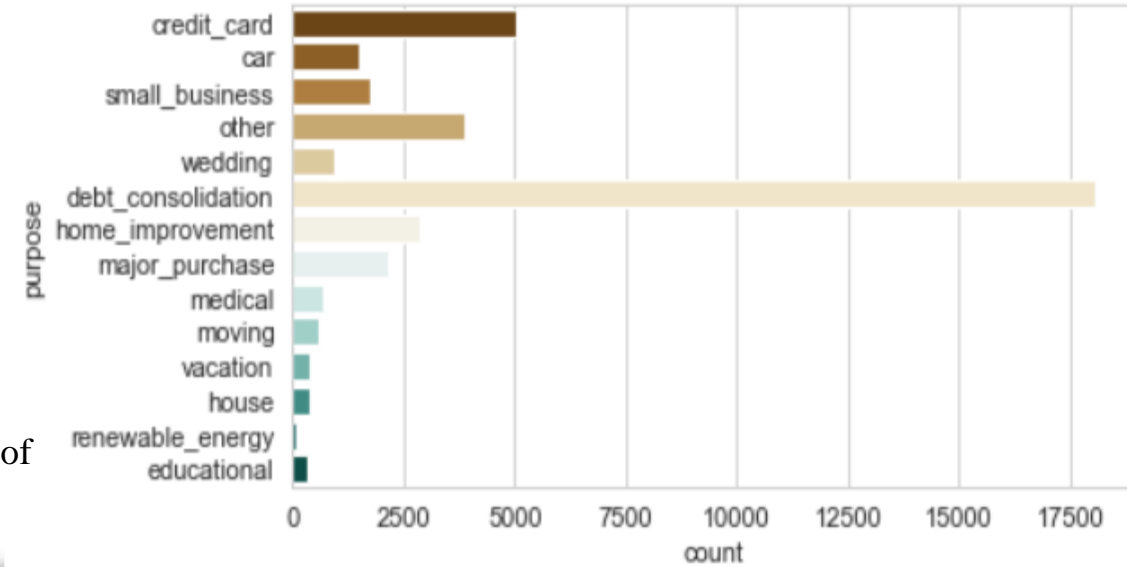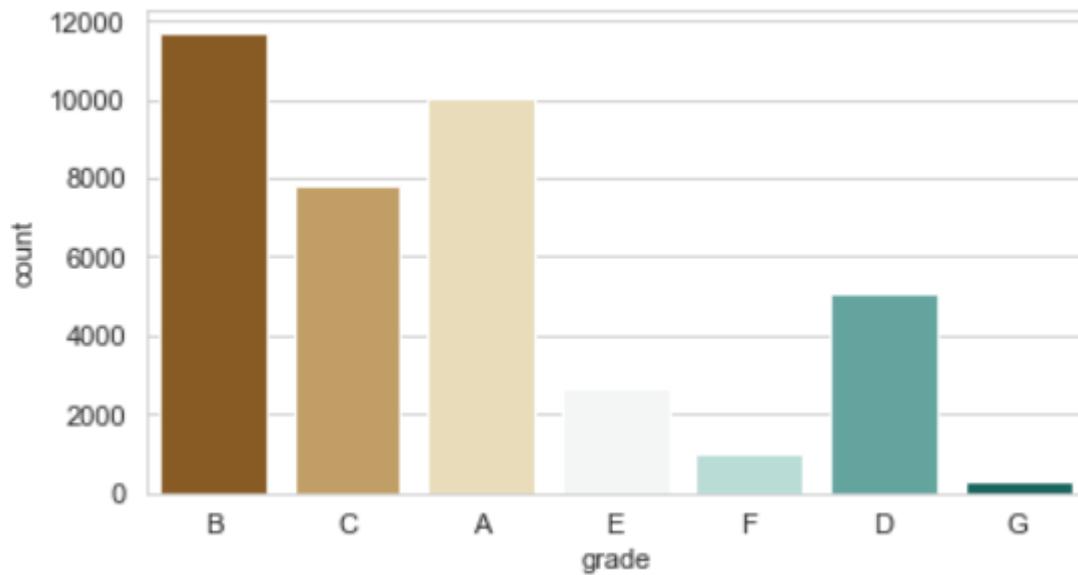
Insights Inferred: Purpose behind loan

- It was observed that Small Businesses have had the highest default rate over other purpose.

- Also it was interesting to note that debt consolidation had highest no of defaulted loans.
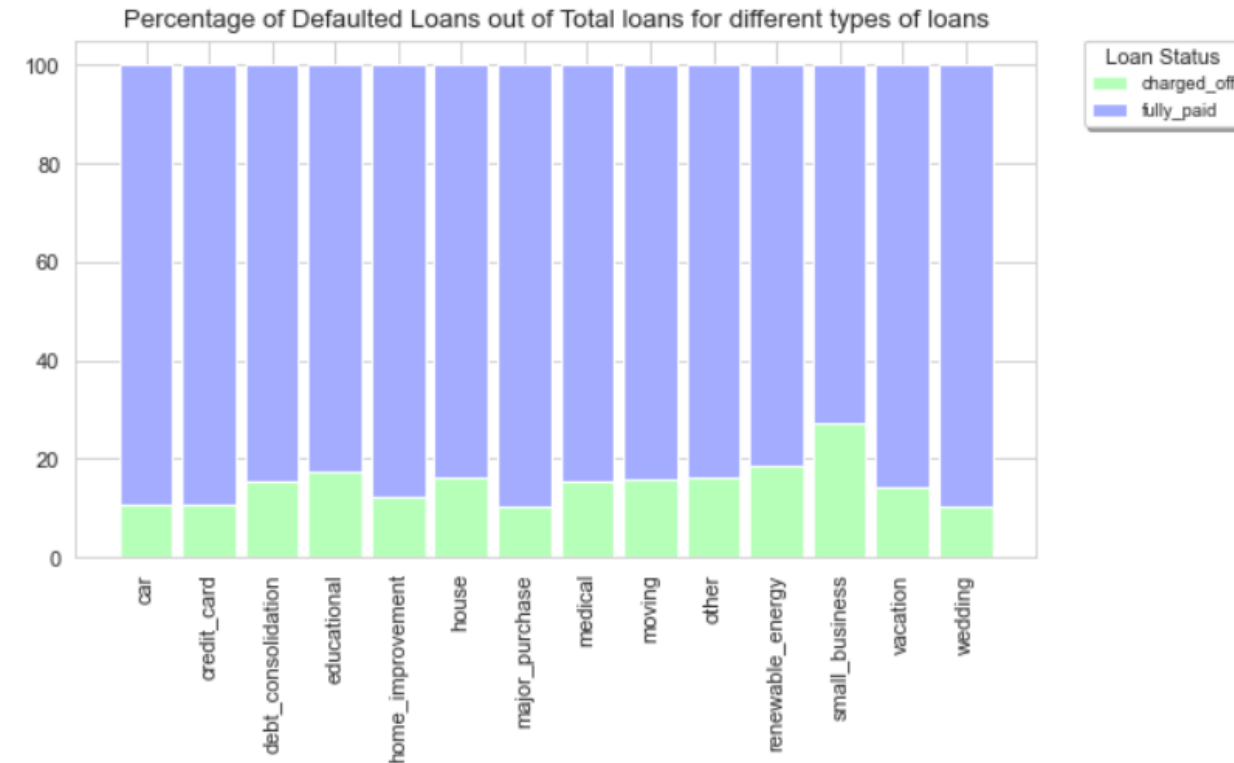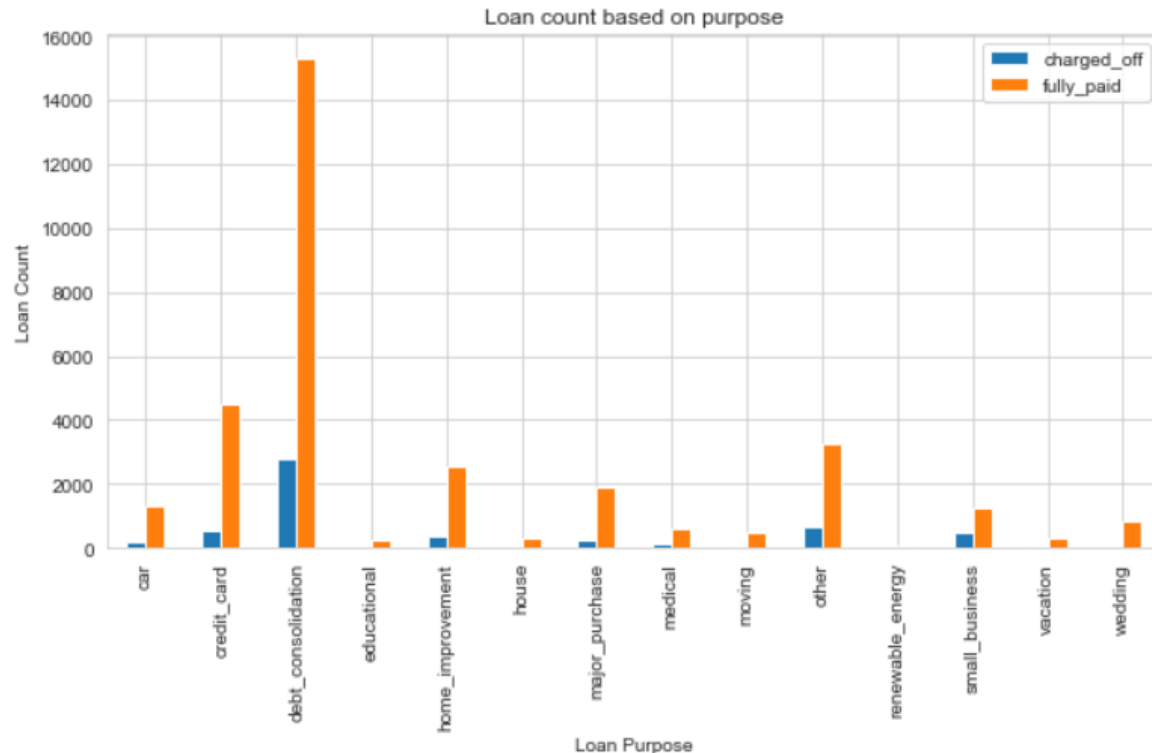
- It turns out that most defaulters are observed in loans that spans over 36 months as compared to loans that spans over 60 months

- And the Loans whose term is over 60 months have default rate that is twice that of default rate of loans whose term is 36 months.

Insights Inferred: Loan Issue Date
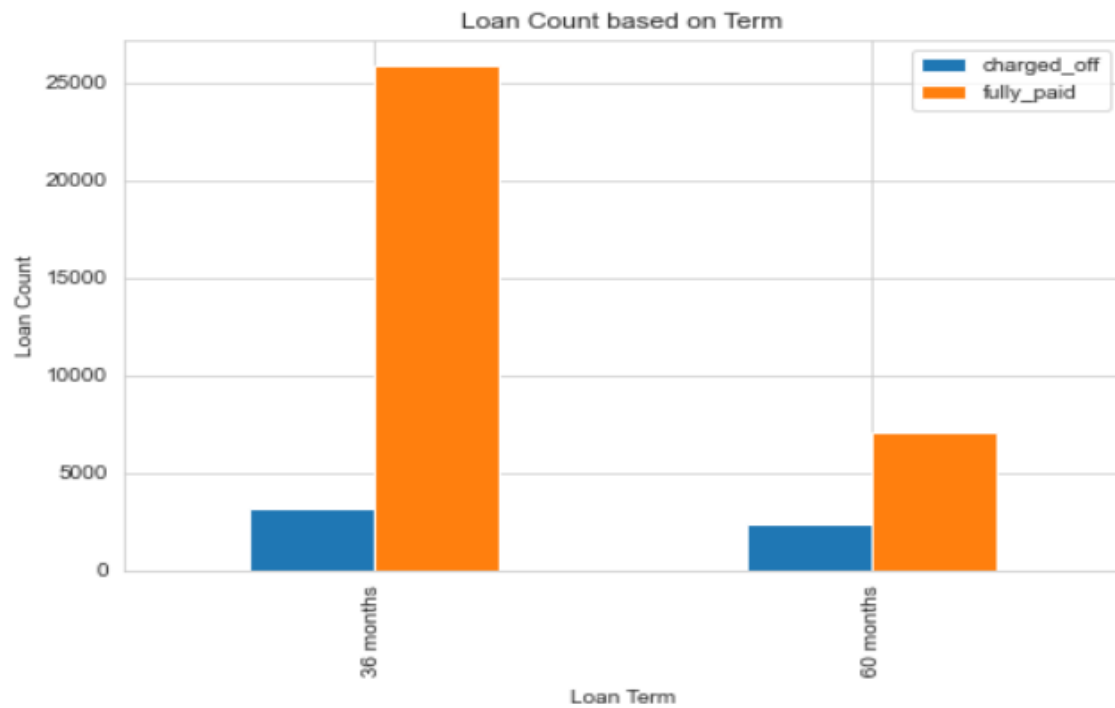
- Its evident that no of Defaulters are increasing every year.

- Its also interesting to note that though defaulters are increasing the default rate was maximum on 2007 followed by 2011
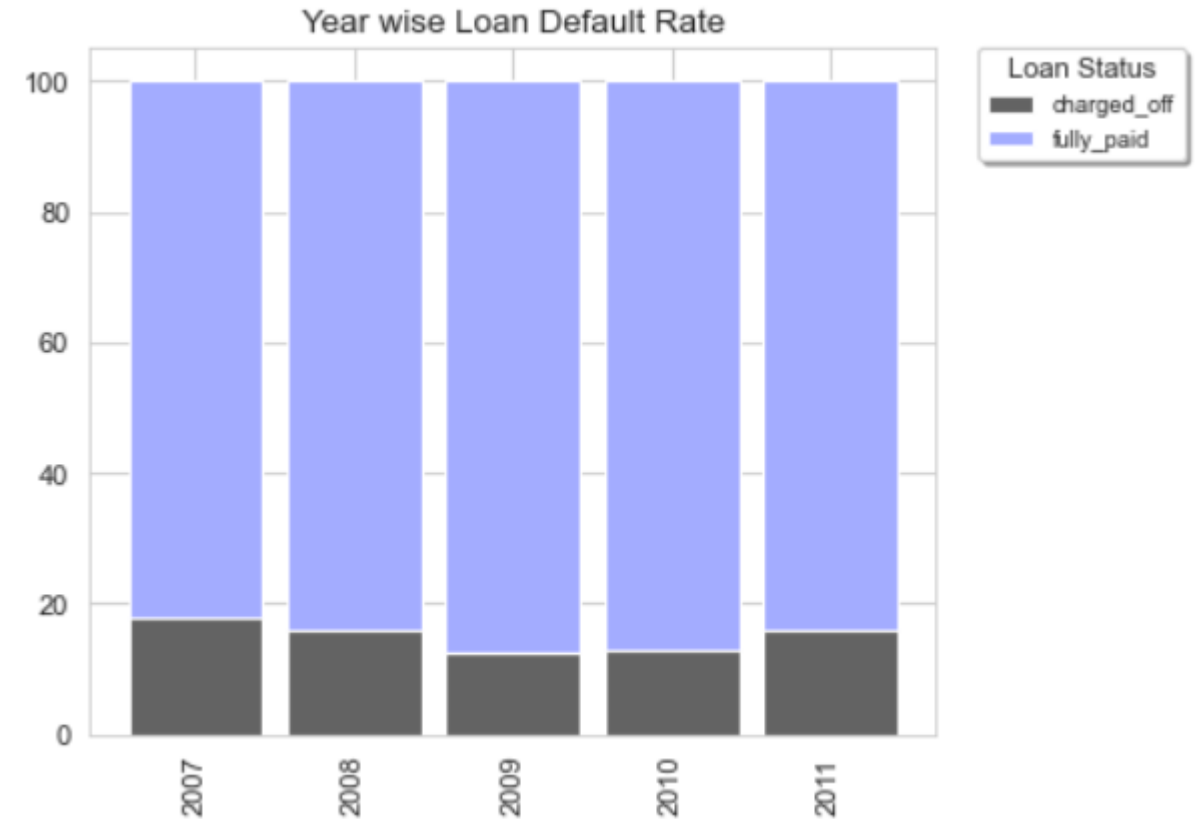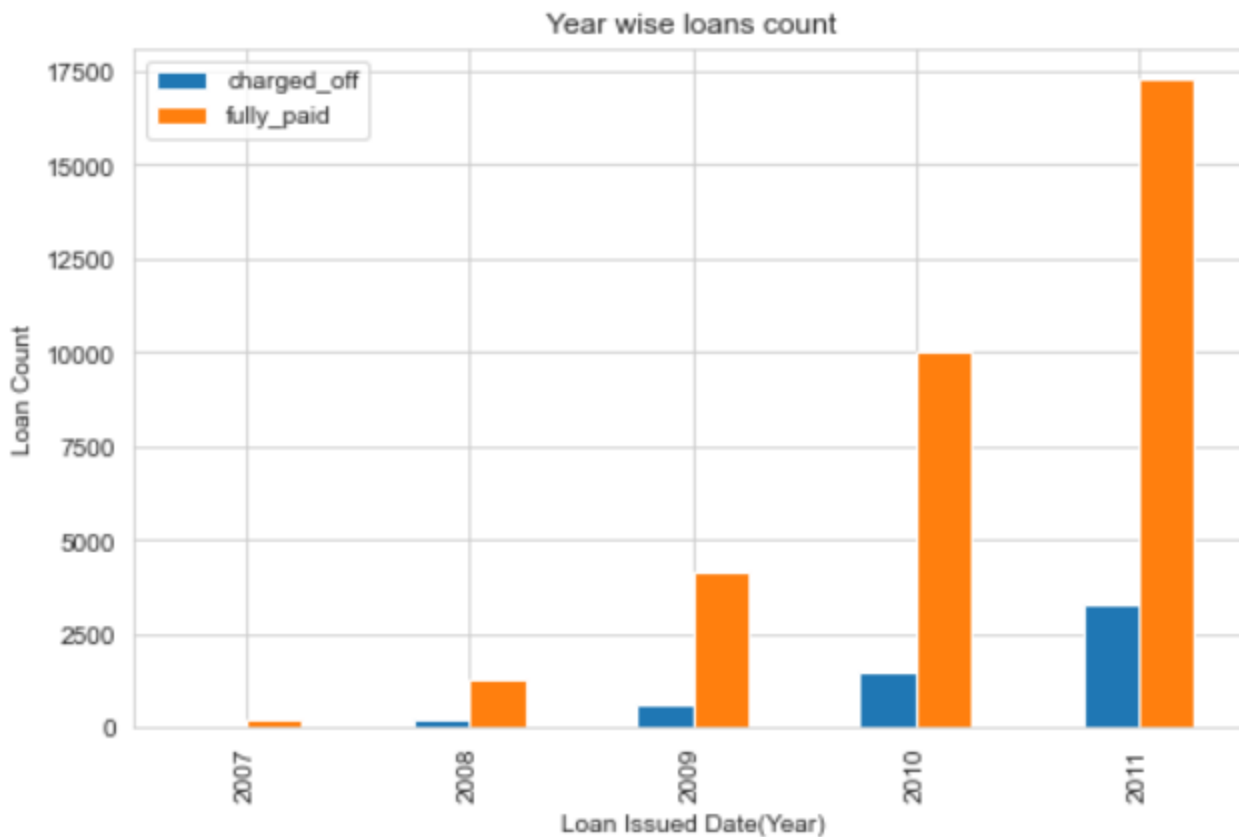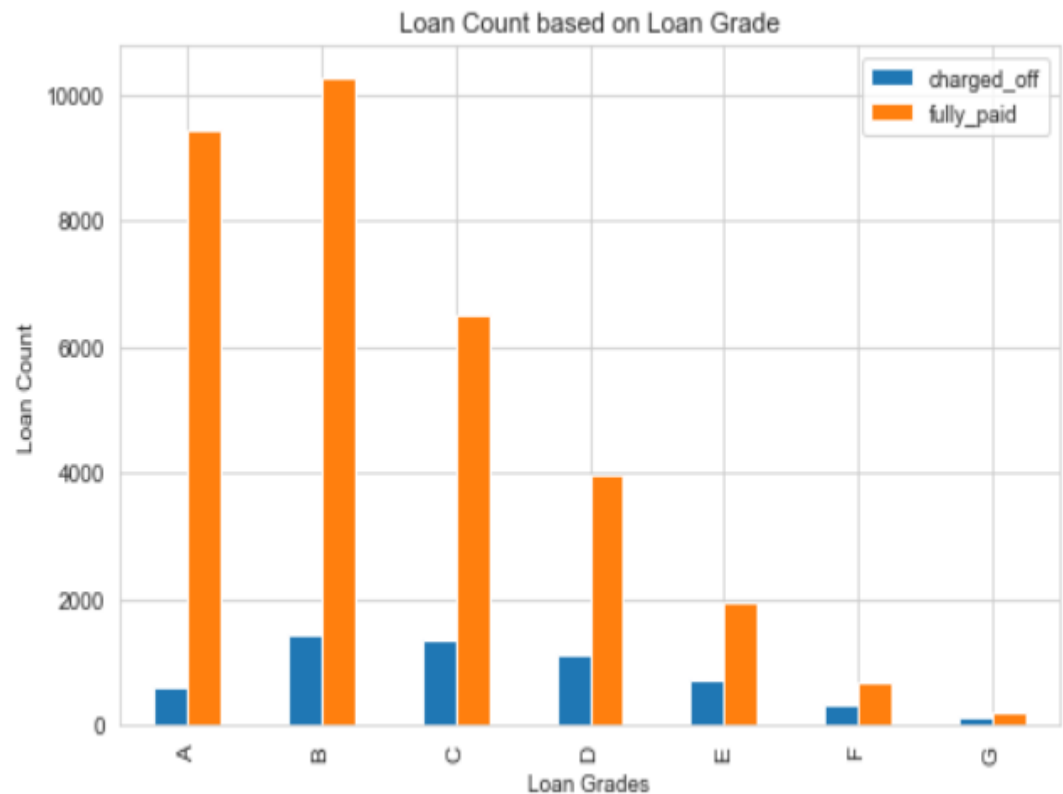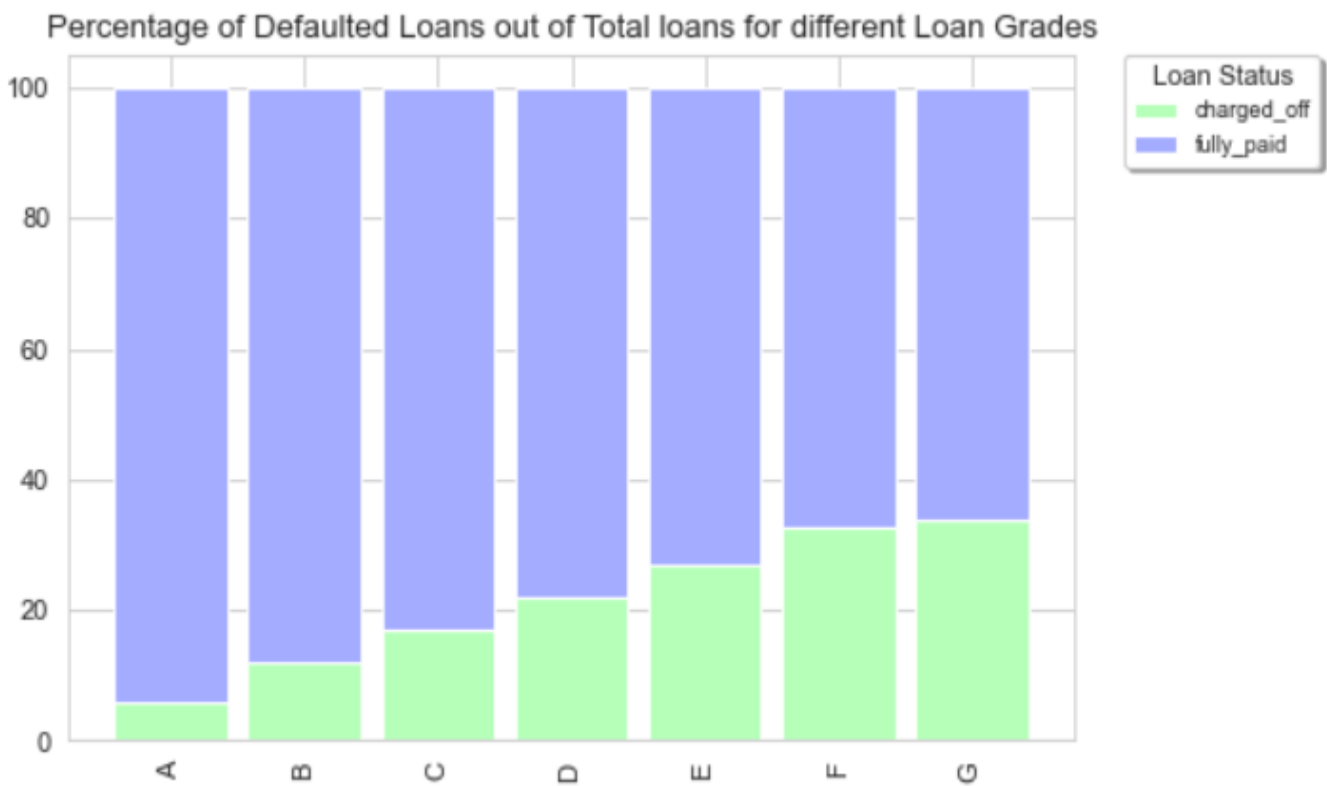
Insights Inferred: Loan Interest Rate

- It was observed that the default rate increases as the rate of interest increases.

- However it turns out that the defaulters are more likely to be observed in loans where interest rate is between 10% and 15%(Medium) compared to low(<10%) and High(>15%).
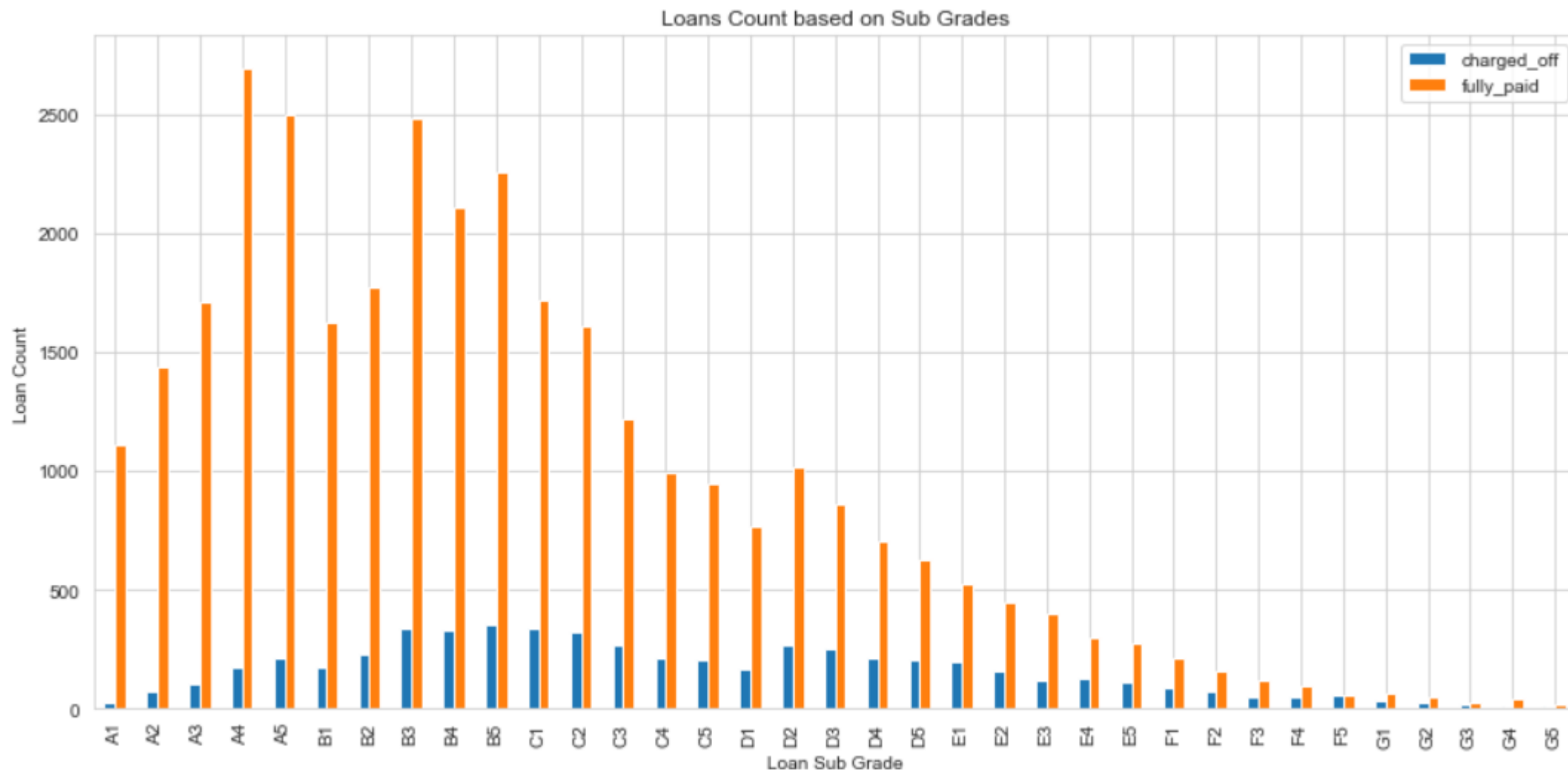
Insights Inferred: Loan Grade
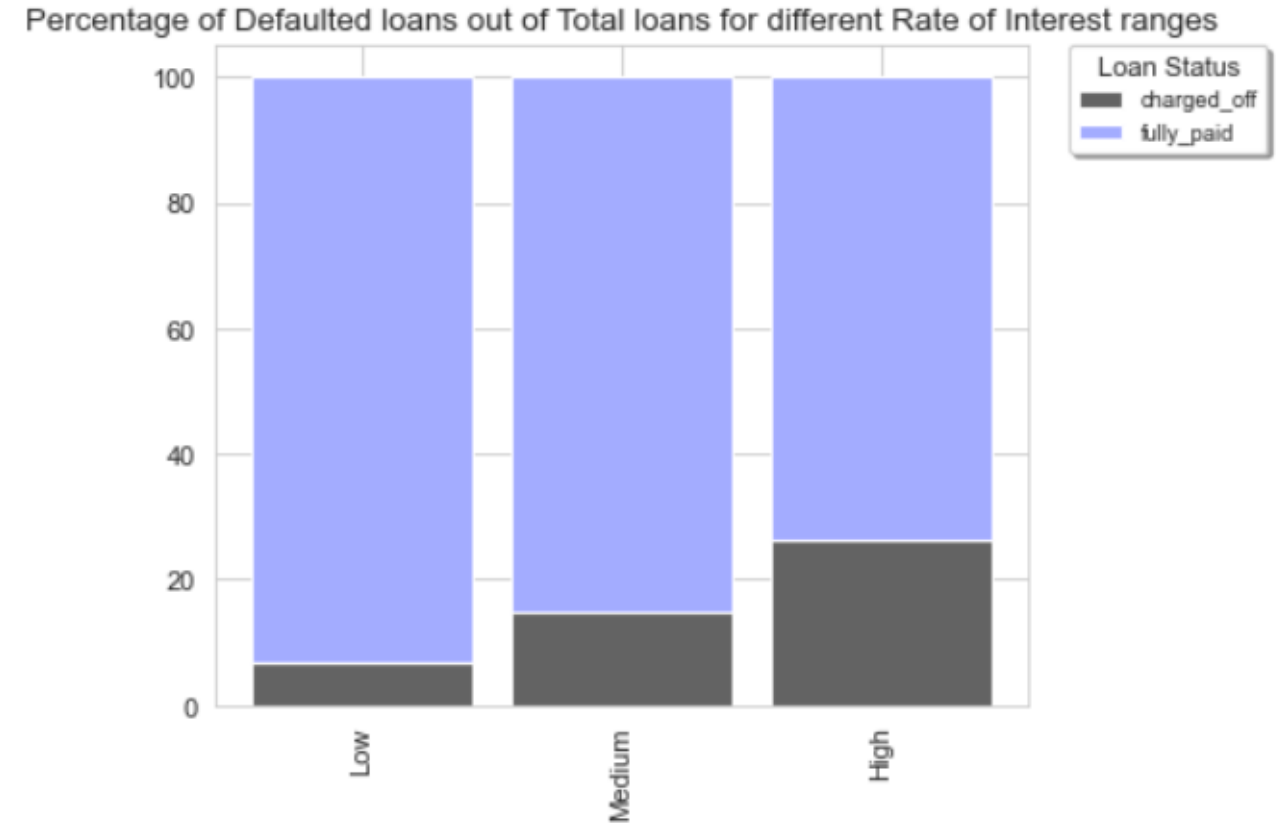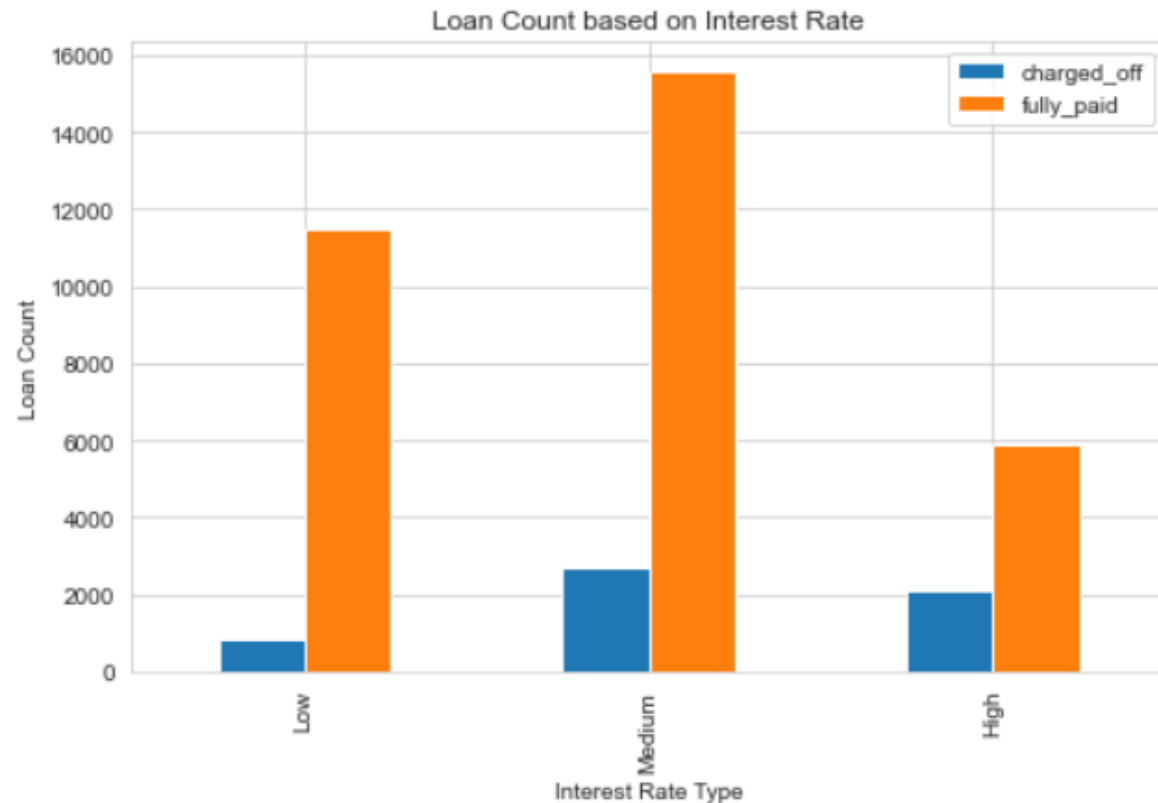
- Default rate in loan increases with decrease in Grade quality(A through B) just as in Rate of Interest.

- Seems that loan applications with grades B, C and D have the highest no of Defaulters.
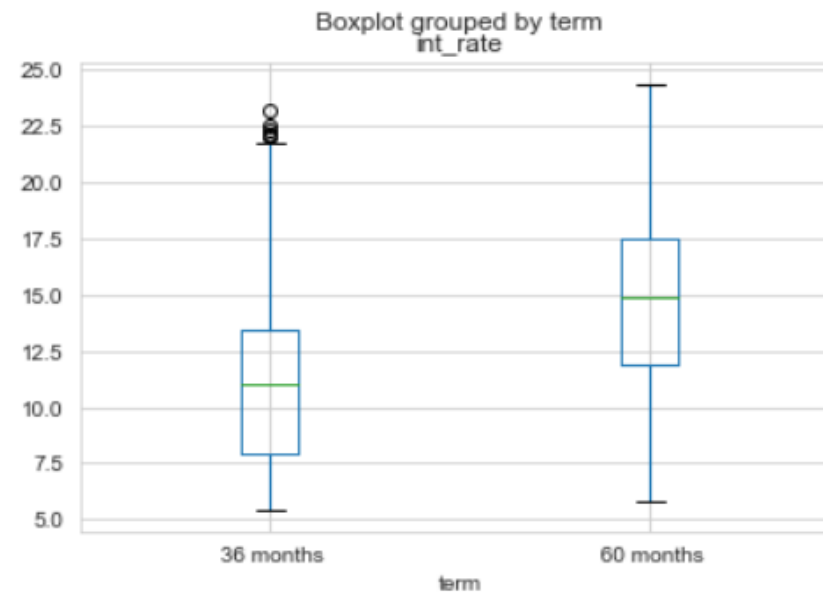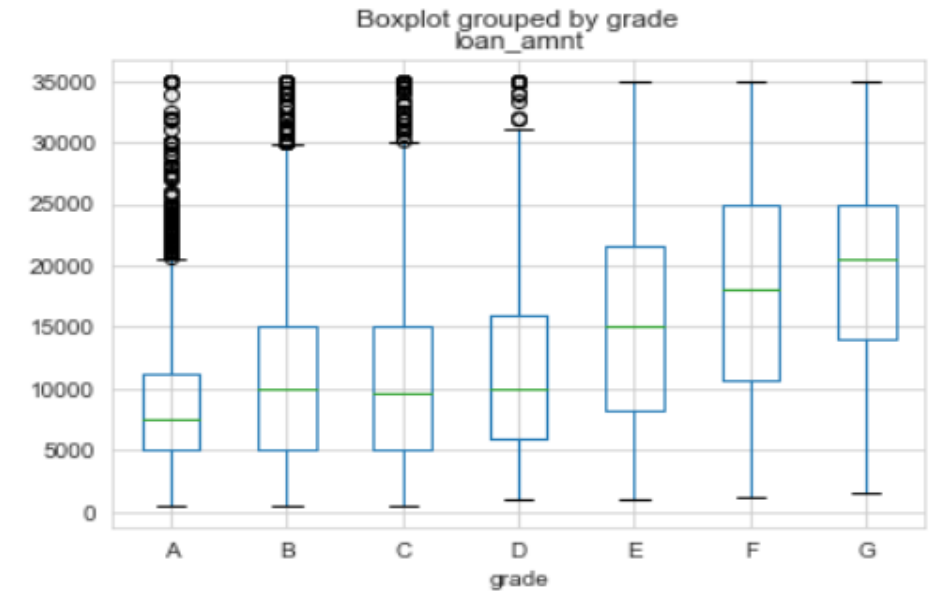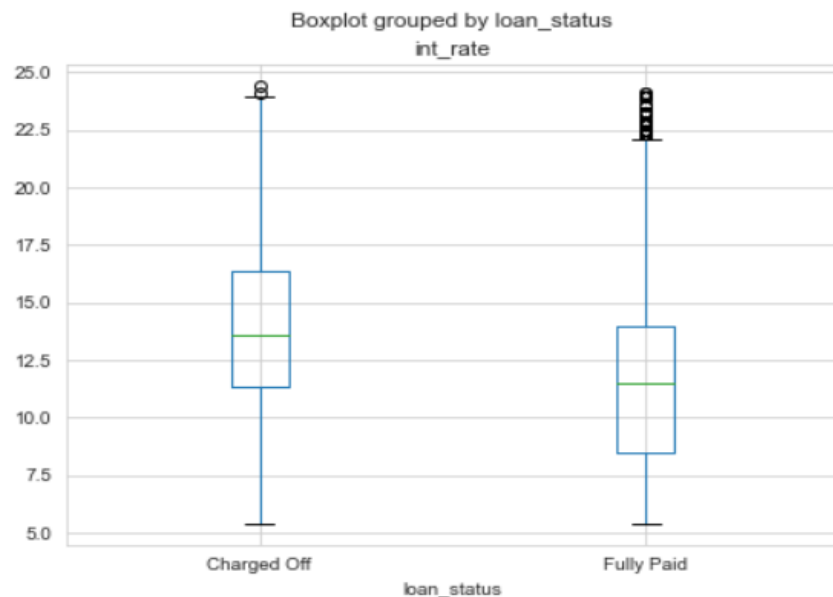

Percentage of Defaulted Loans out of Total loans for different Loan Grades

Loan Count based on Loan Grade

- Most Defaulters were found in loans with sub grades B5, B3, C1, B4, C2, D2 and D3.

Loans Count based on Sub Grades

Insights: Grade

- It seems that apparently lower grade is given to Large loan amounts.

- Loans at a higher interest rate are more likely to be defaulted.

- Seems like large amounts are having longer term and hence the interest associated with them is also high.

Boxplot grouped by grade
loan_amnt

Boxplot grouped by loan_status
int_rate

Boxplot grouped by term
int_rate

- The defaulters here are most likely to be observed in the case where borrowers experience are 10+ years and 1 year or lesser

- As far as default rate is concerned no significant trend is seen as all of them are more or less closer to each other
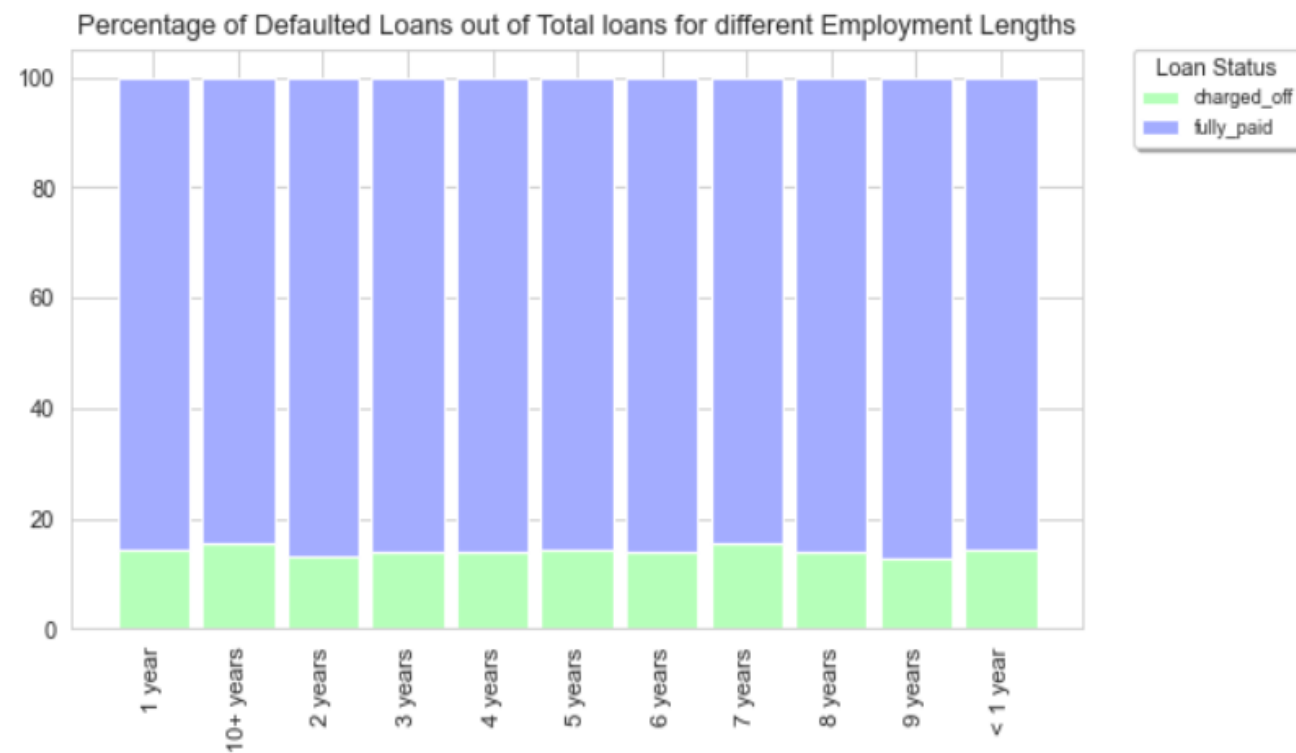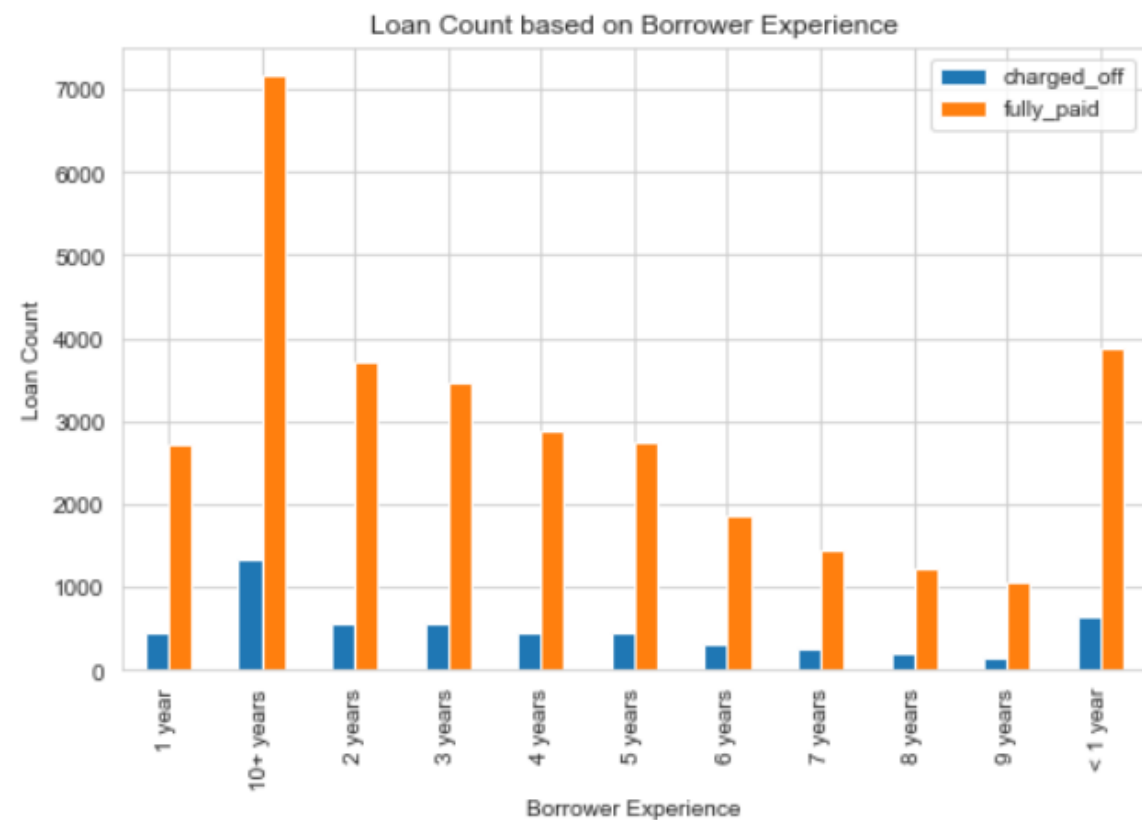
Insights Inferred: Home Ownership

- It was observed that Default rate in home ownership type 'Other' is slightly high(more or less same) over others.

- However it appears that most defaulters are the borrowers with house ownership type mortgage and rent.
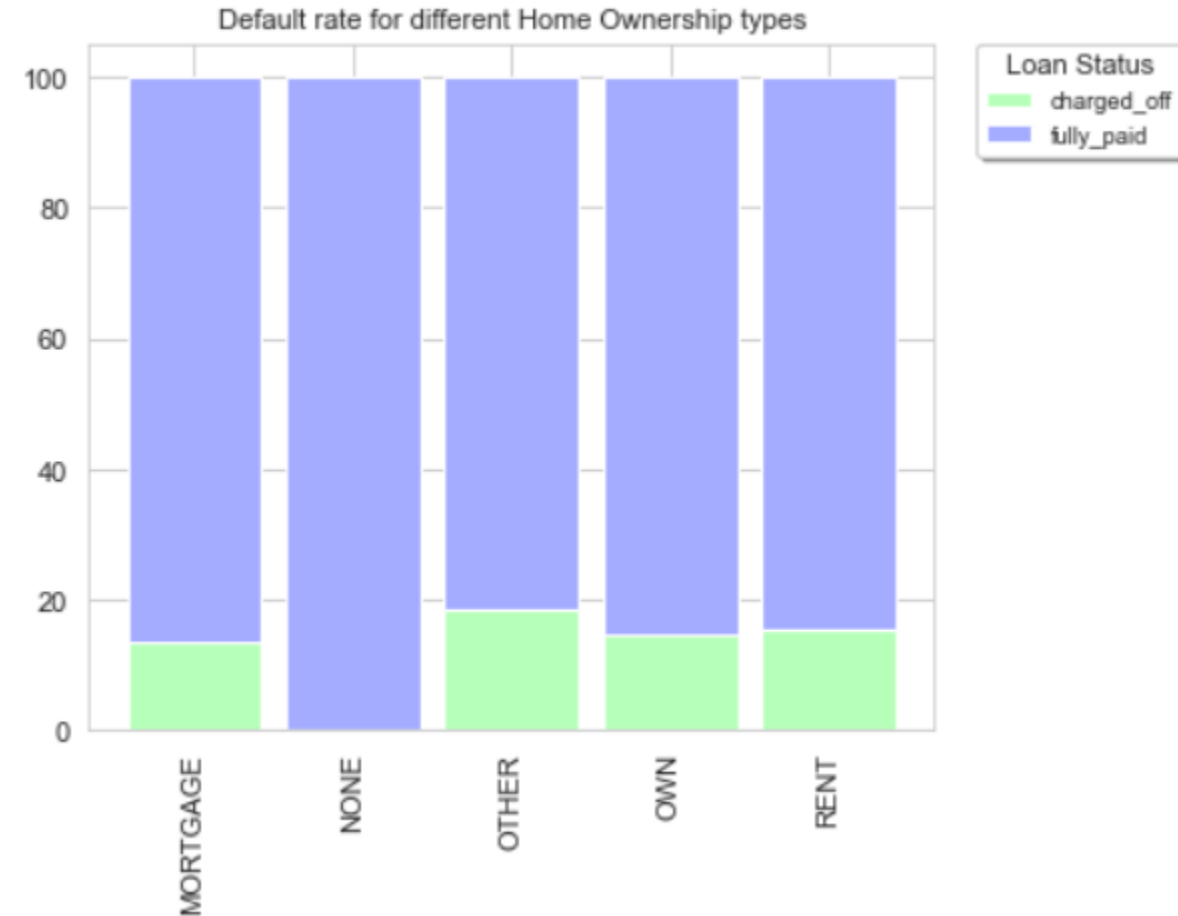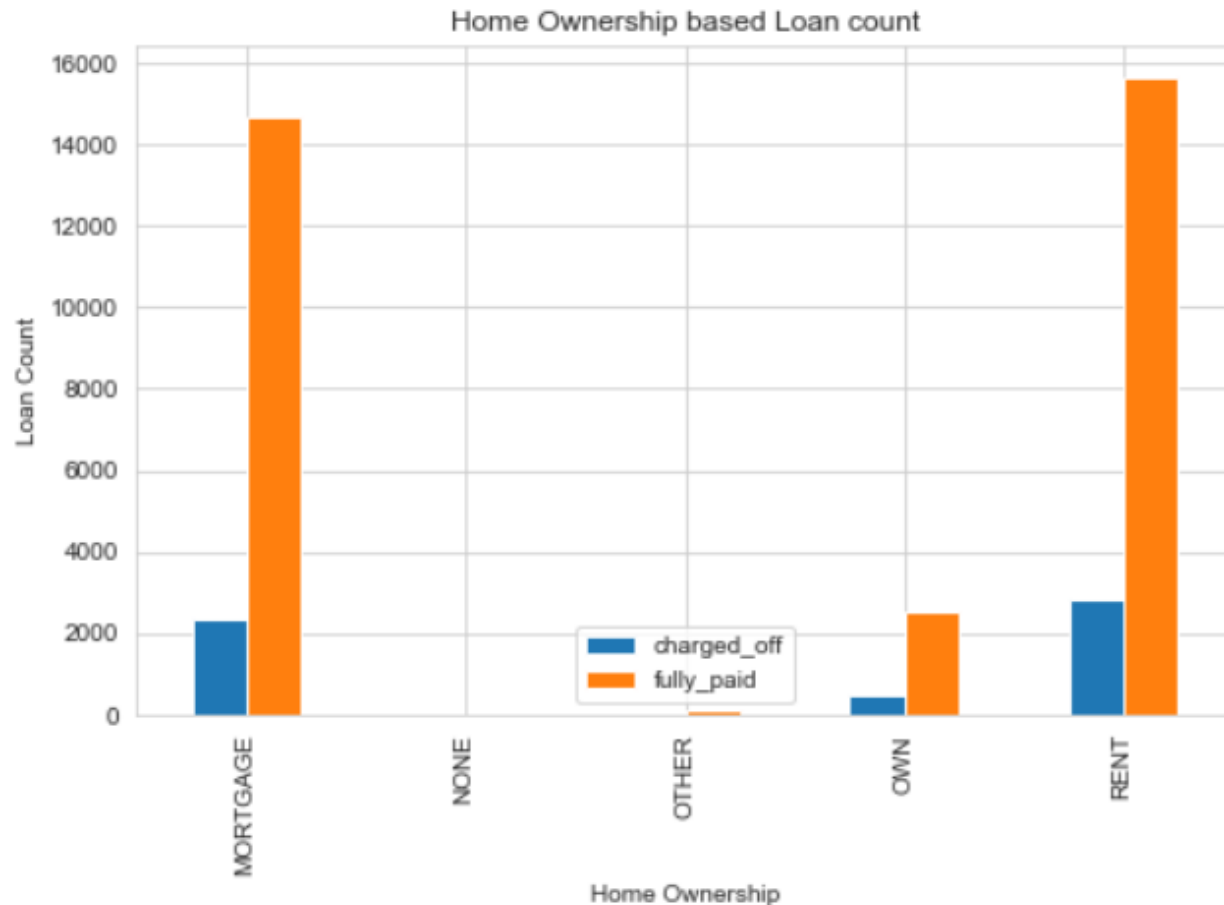
Insights Inferred: Based on State

- Default Rate is maximum in the state NE .

- It turns out that CA has maximum no of Defaulters followed by NY, FL and TX.

State wise Loans count

Percentage of Defaulted Loans out of Total loans across different States

Insights Inferred: DTI
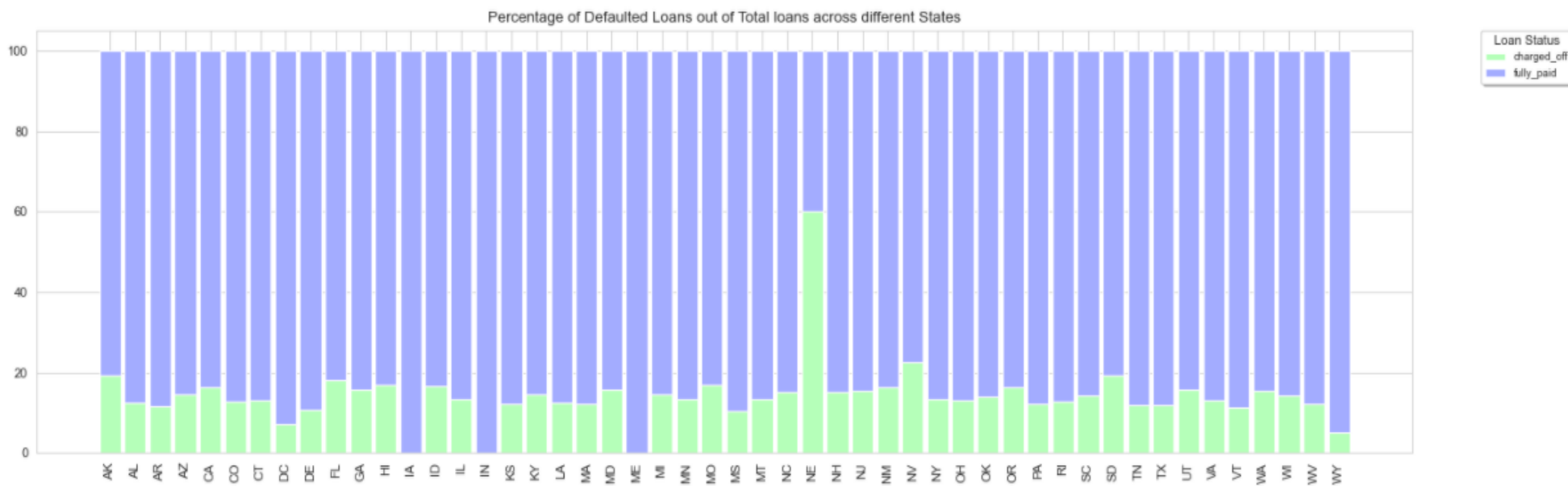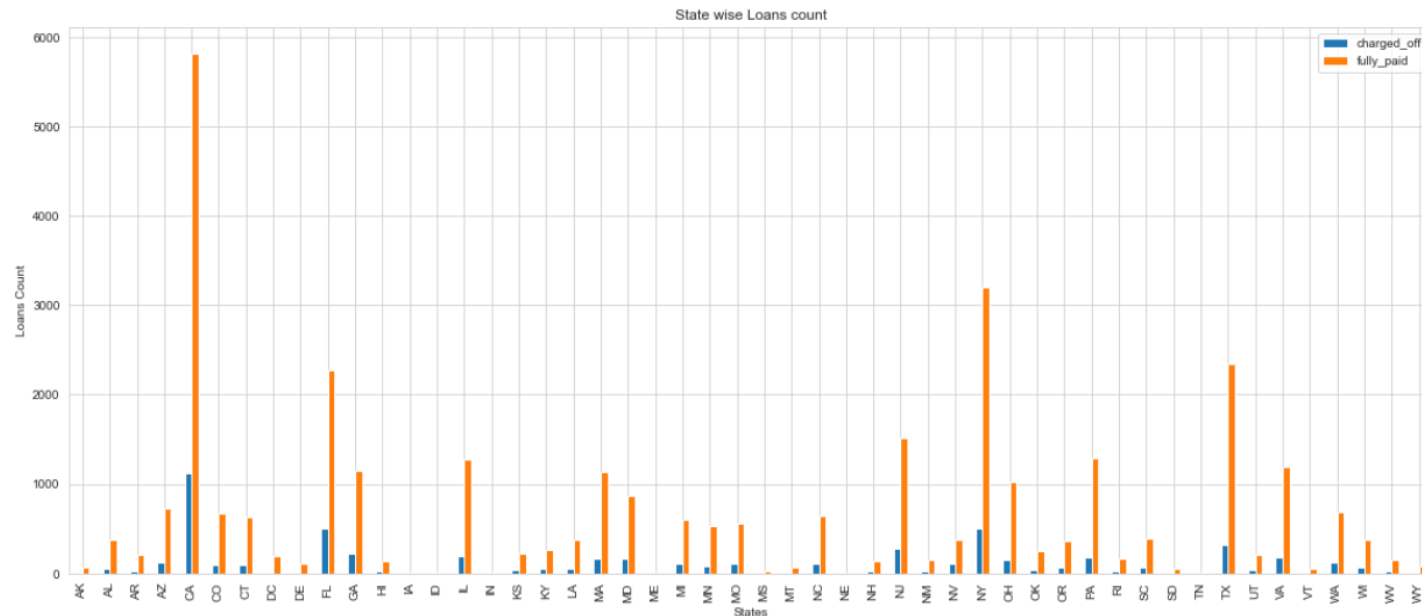
- More no of Defaulters were seen in loans with DTI range between 10 & 15 , 15 & 20 , 20& 25 and 5 & 10 respectively.

- Default rate was slightly high in loans with DTI between 20 & 25 over others.

Loans across different DTI Bins

Percentage of Defaulted Loans out of Total loans for different ranges of DTI
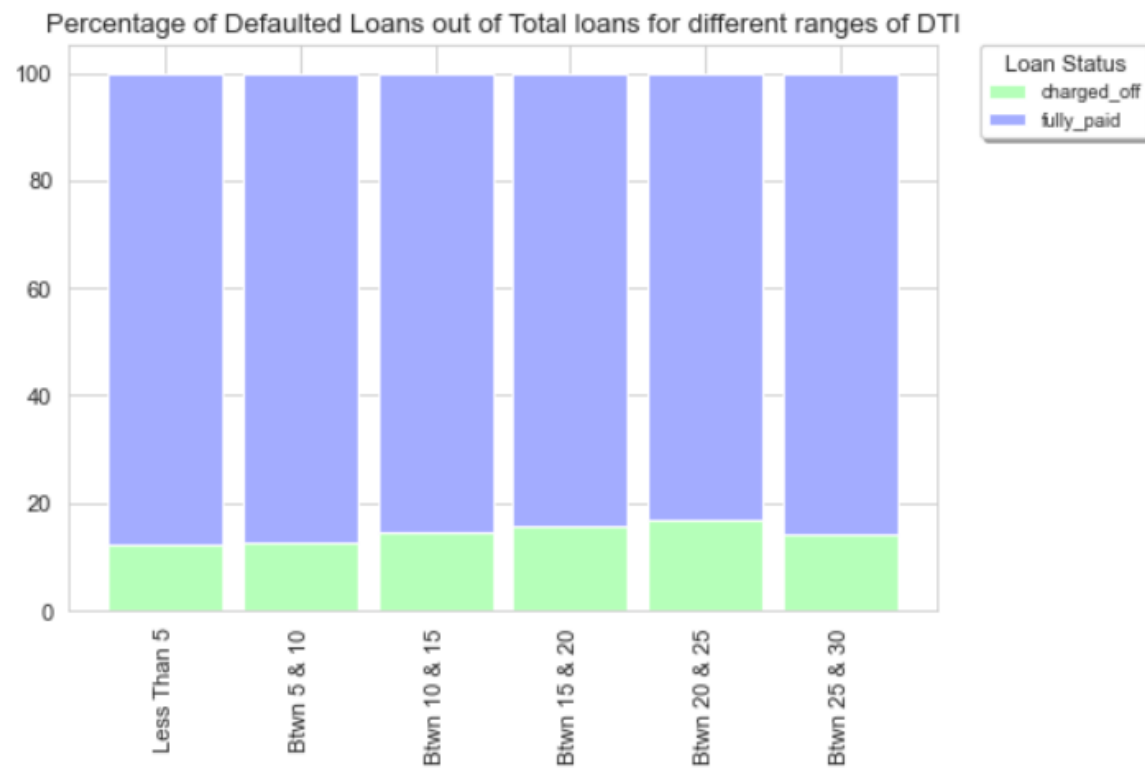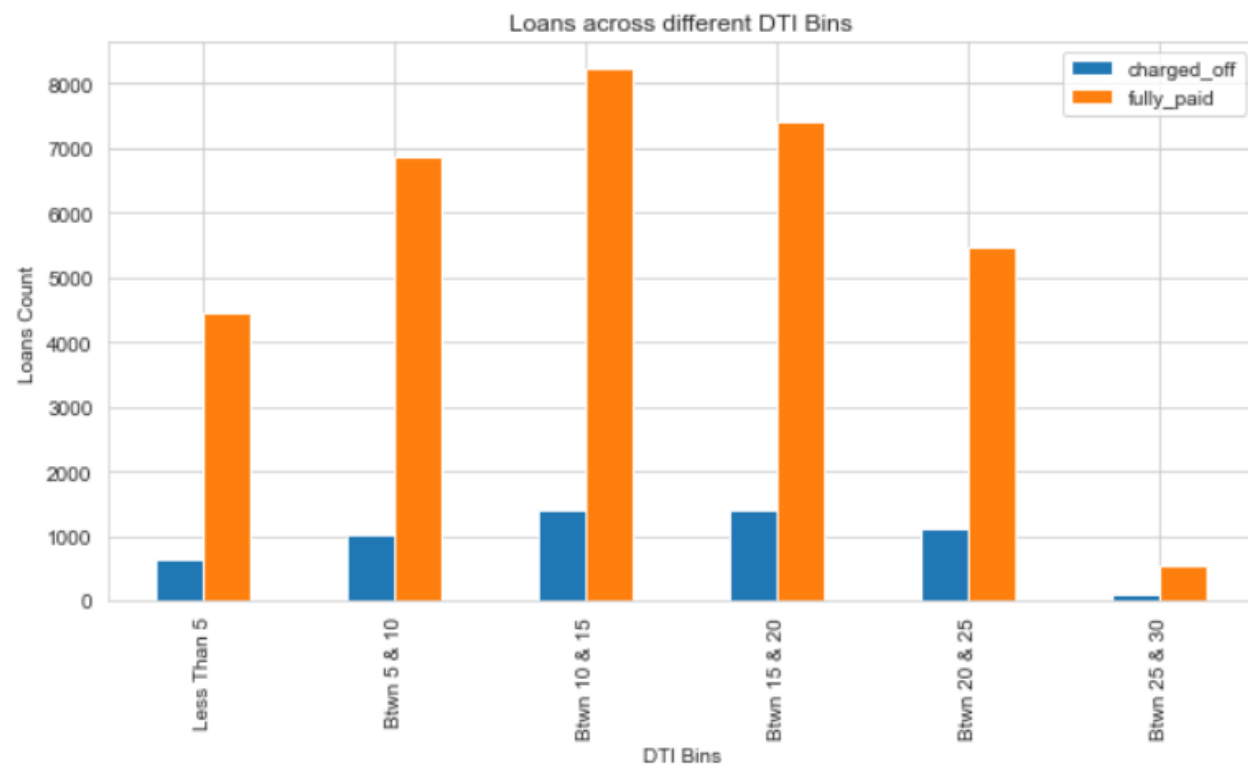
Insights Inferred: Public Records

Loans count based on Public Records
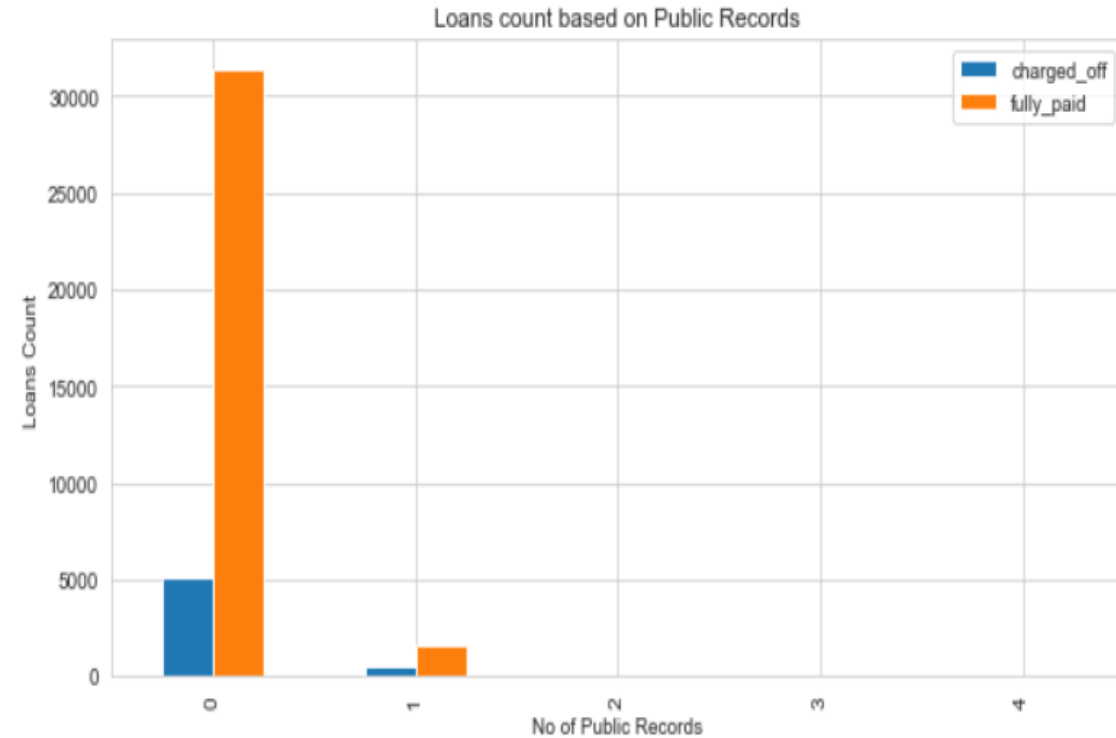
- It turns out that having public derogatory record increases the chances of Charge Off significantly.

Percentage of Defaulted Loans out of Total loans for different Pub_rec Values

Insights Inferred: Annual Income

Loans count based on Annual Income
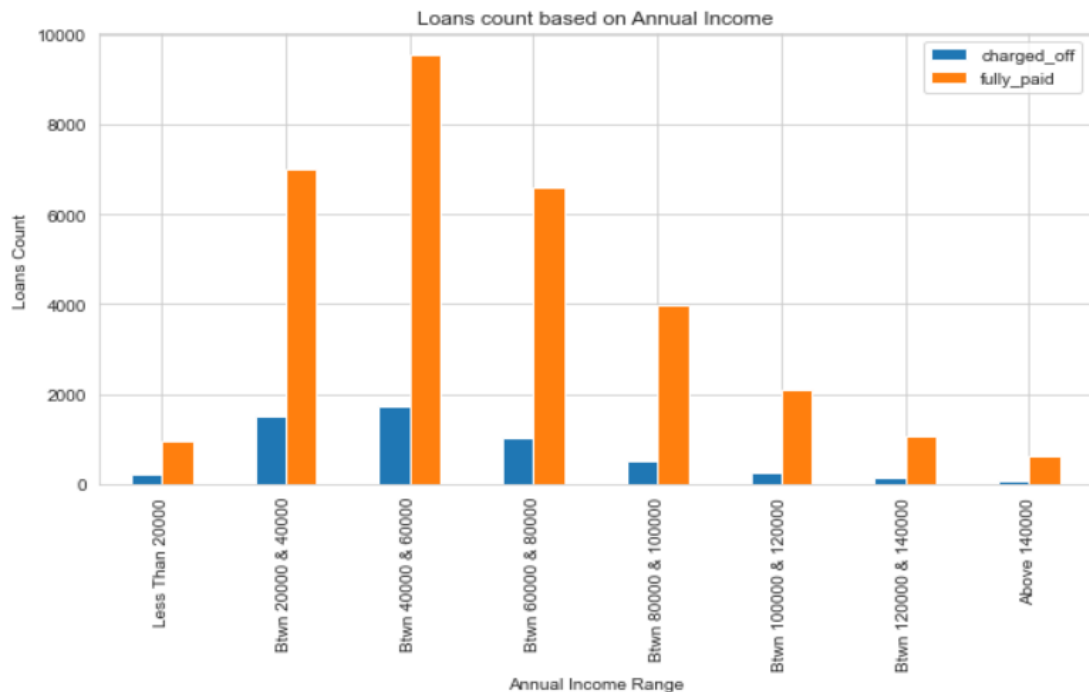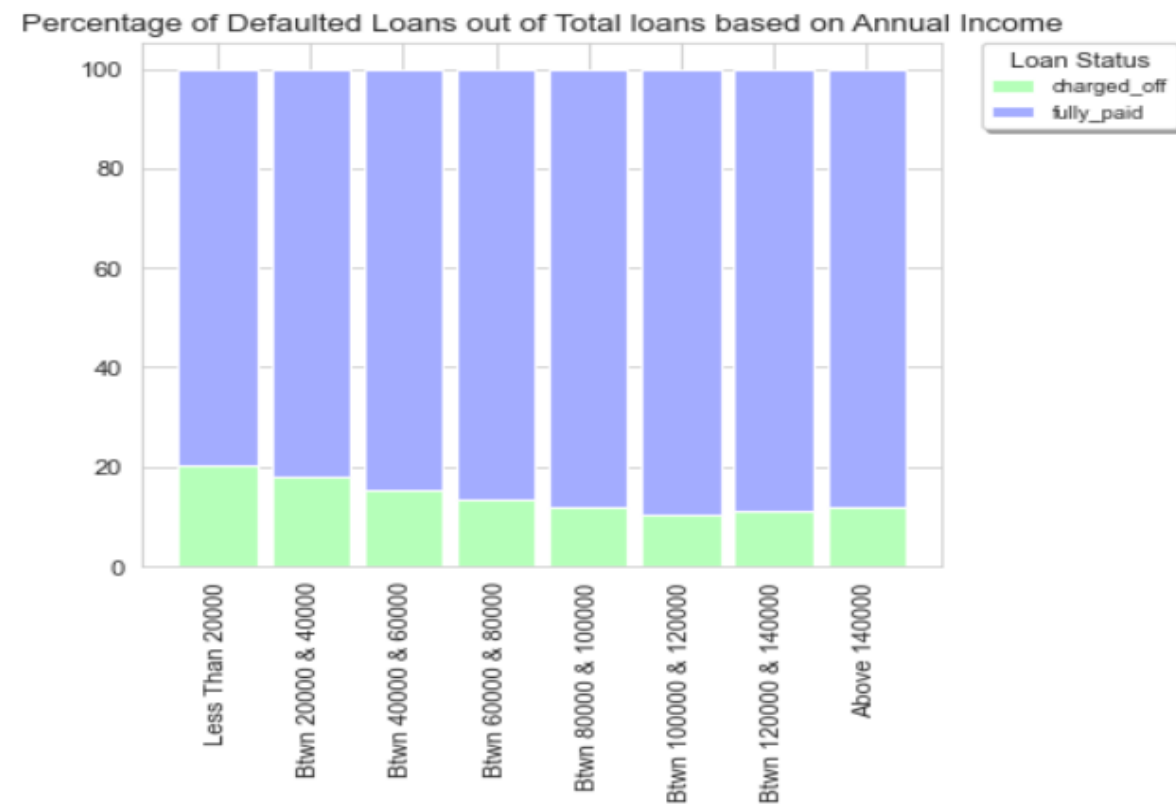
- It seems that default rate decreases with increase in annual income of the borrower.

- Most Defaulters are the Borrowers whose annual income is between 0-20K, 20k-40k and 40K-60K respectively

Percentage of Defaulted Loans out of Total loans based on Annual Income

Features of Interest

- Based on our analysis here are some features that will aid in decision making.
  - Loan Purpose
  - Loan Grade
  - Loan Term
  - Address State
  - Annual Income
  - Home Ownership
  - Public Records

- Its better to double check or if possible reduce the number of approvals where purpose is small business.

- Avoid/Reduce approving loans to people having bad past records.

- As Default rate of loans increase from Grades A to G. Loans with higher grades(towards A) should be considered for approval.

- Based on analysis it turns out that short term loans are most likely to be fulfilled in time. Therefore focusing more on short term loans will be an appropriate decision.

- Borrowers with Rented or Mortgaged house turn out to be defaulters in most cases, therefore its better to avoid such case or should be scrutinised with much more caution.

- People with low income tend to fall into Defaulter list. Therefore annual income is a significant feature to be considered during loan approval.