LENDING CLUB CASE STUDY SUBMISSION

Group Members:

Nikhil Oommen Jacob
Ragula Srinivas

PROBLEM STATEMENT

A consumer finance company which specialises in lending various types of loans to urban customers, receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

Two types of risks are associated with the bank's decision:
1.If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
2.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 a financial loss for the company.

DATA SET

LOAN DATASET

Loan Accepted → Default
Loan Accepted → Non-Default
Loan Rejected (Not considered in dataset)
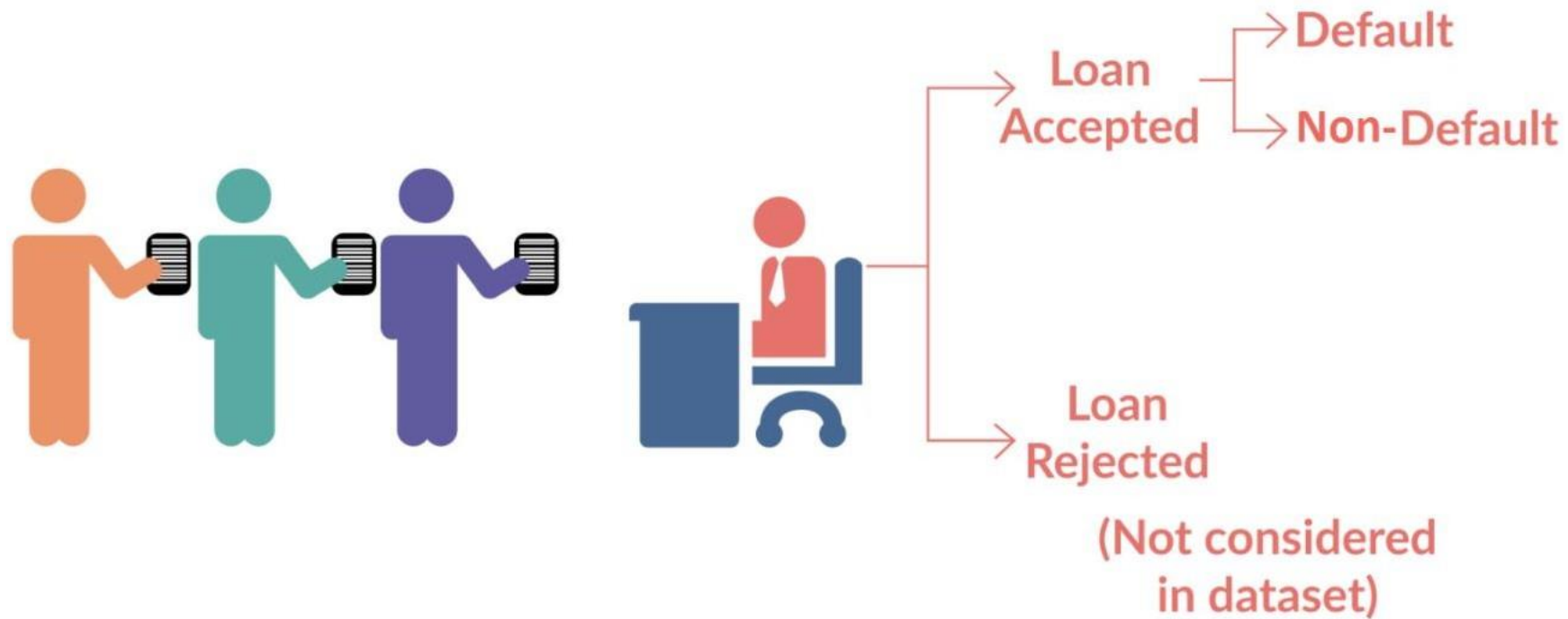
When a person applies for a loan, there are two types of decisions
that could be taken by the company:
Loan accepted: If the company approves the loan, there are 3 possible scenarios
described below:
1.**Fully paid:** Applicant has fully paid the loan (the principal and the interest rate)
2.**Current:** Applicant is in the process of paying the instalments, i.e. the tenure
of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
3.**Charged-off:** Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan.

Loan rejected: The company had rejected the loan (because the candidate does not
meet their requirements etc.).

CASE STUDY OBJECTIVE

Identify the risky loan applicants whom are tend to 'default/charged of'

Understand the driving factors or Driver Variables behind Loan Default

Lending company can utilise this analysis for its portfolio and risk assessment

Problem solving methodology

> Drop columns with null values, all random values or single category value
>Convert values to proper int, float, date representations

>Analyse variables against segments of other variables
> Create derived variables

Publish insights and observations

| Clean Data | Univariate Analysis | Segmented Univariate Analysis | Bivariate Analysis | Summarize Results |

>Check distributions and frequencies of various numerical and categorical variables
> Create derived variables

>Do correlation analysis
Check how two variables affect
each other or a third variable
> Analyse joint distributions

ANALYSIS

The essence of the whole project is to analyze and understand how consumer attributes and loan attributes are influencing the tendency of defaulting.
● We performed **data cleaning and preparation** on the Loan dataset:
○ Imputed the NA values for all the variables
○ Created two new columns:
■ Profit and Loss column
■ Ratio of funded amount and annual income
● During **univariate analysis** we have created:
○ Histograms and Bar charts to check out the distribution of all the driver variables
○ Box plots to detect the Outliers
○Performed the Multivariate analysis to understand how different variables interact with each other.

Loan Status Analysis

1. Maximum no. of loans are applied by people from CA.
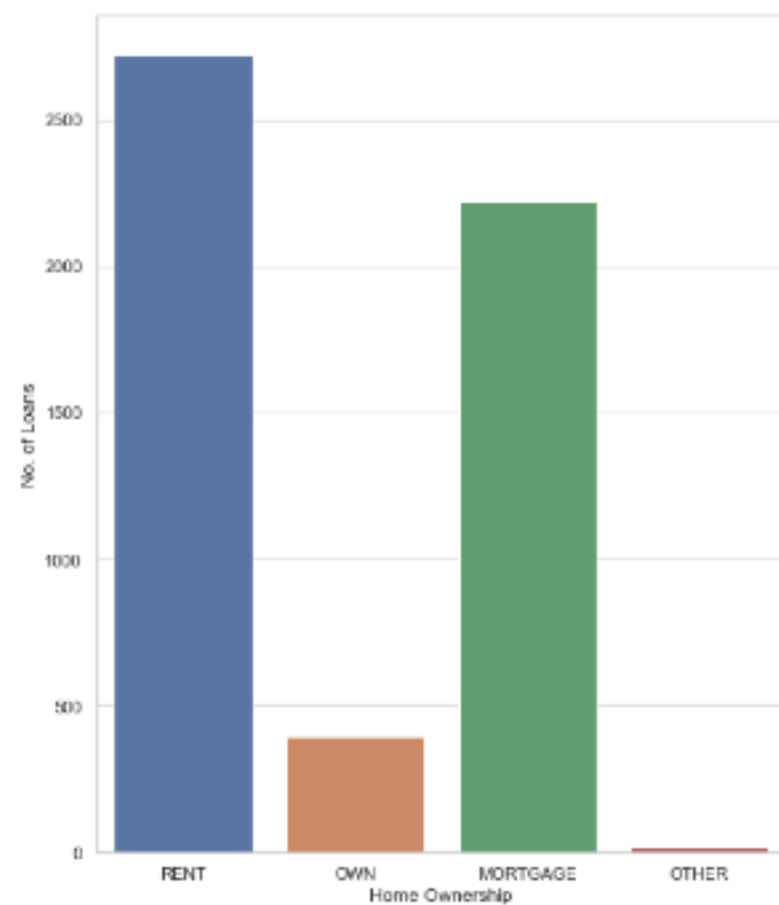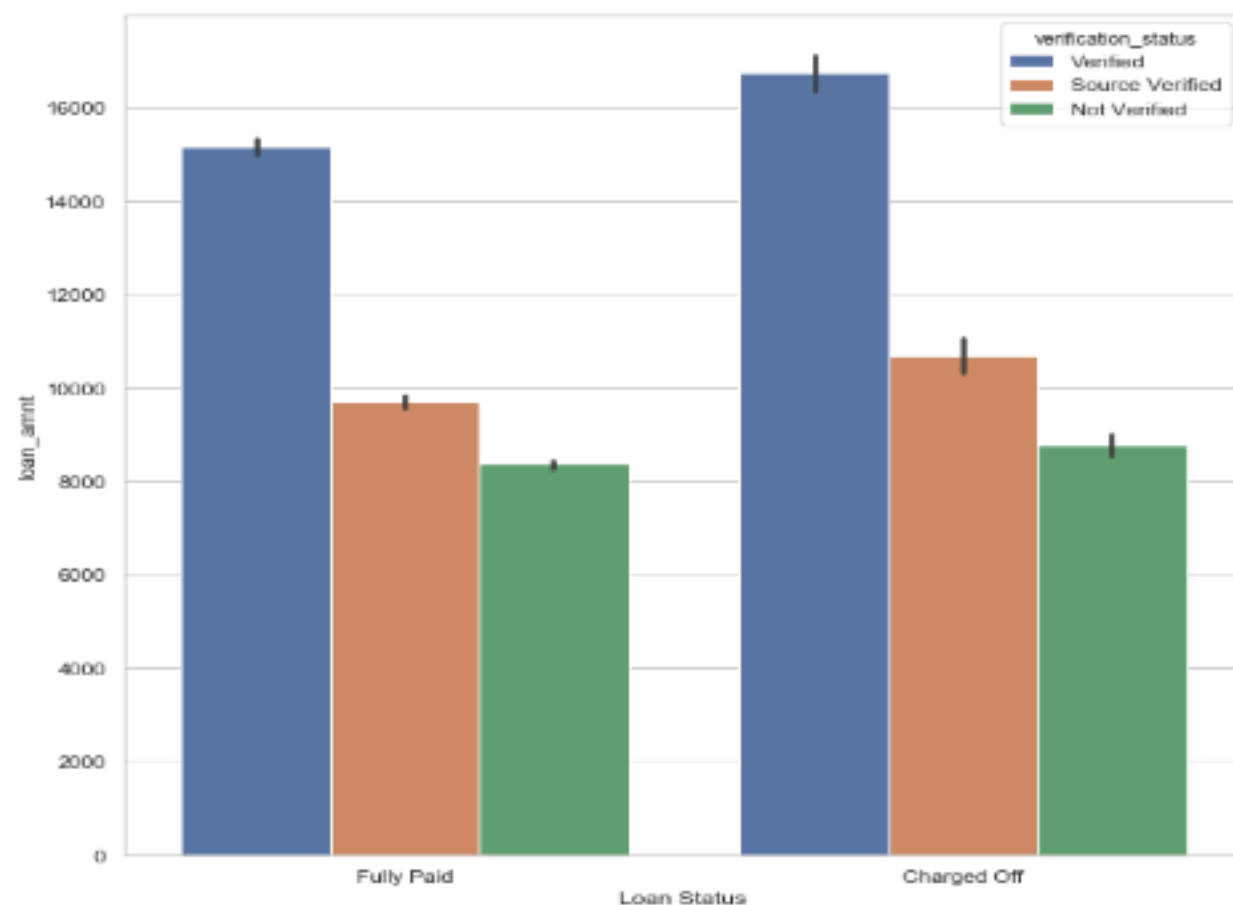2. Maximum no. of loans are charged off for people of NV.
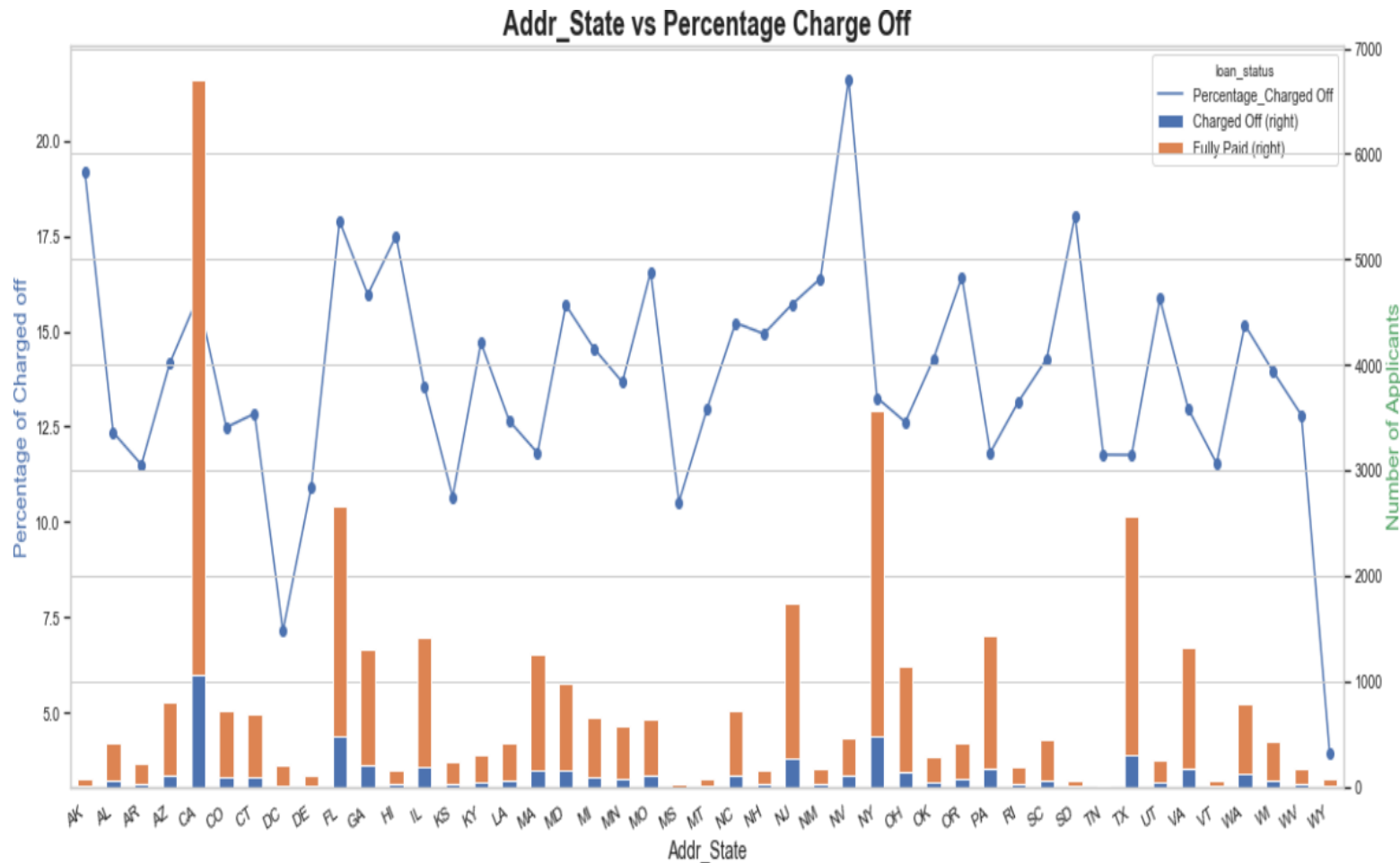
1. More than 30000 loans are fully paid.
2. Around 5000 loans are charged off.


Addr_State vs Percentage Charge Off
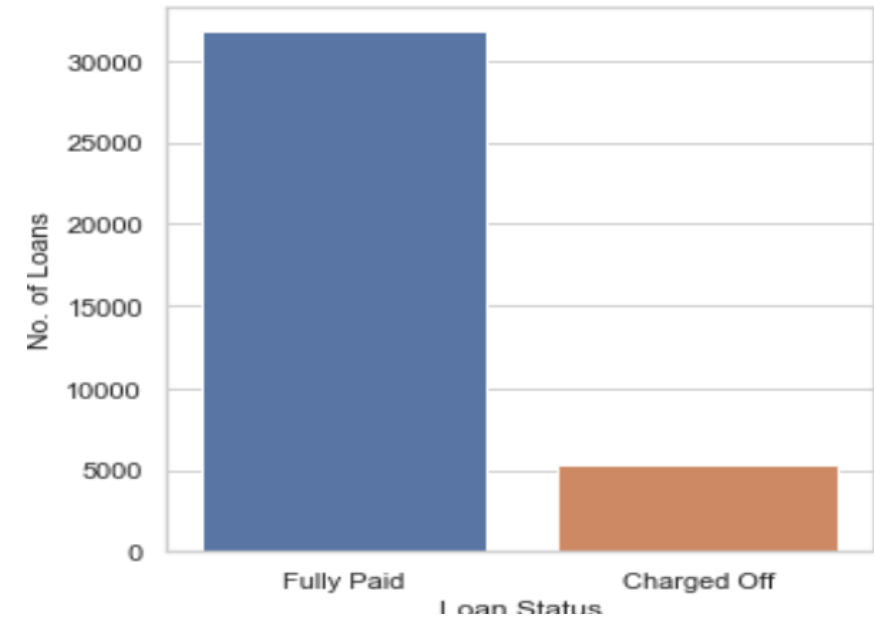
Verification Status & Home Ownership as per Loan Status

1. Max no. of loans are charged off where verification status is **Verified**.
2. Max no. of loans are charged off for people who are living on rent.

Year & Month Wise trends of Loans

1. No. of loans, fully paid and charged off are increasing year-by-year. They are at maximum in the year 2011. This is
a very positive trend for Lending Club as the requirement of loans are increasing every year.

2. The month-wise trend shows that most of the loans are fully paid as well as charged off as the year comes to an end, applicants apply more loans towards the end of year.
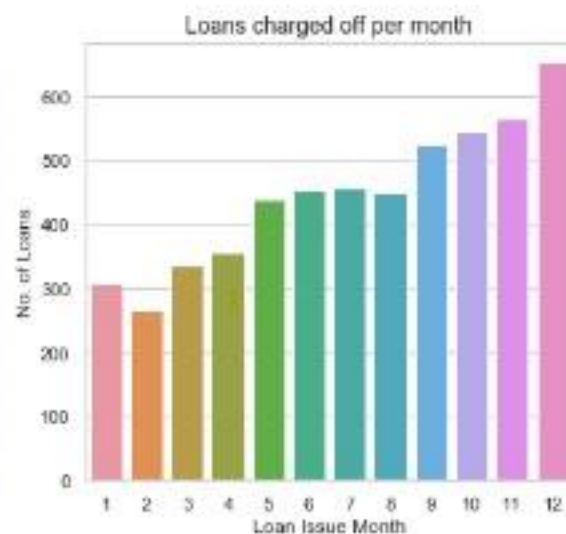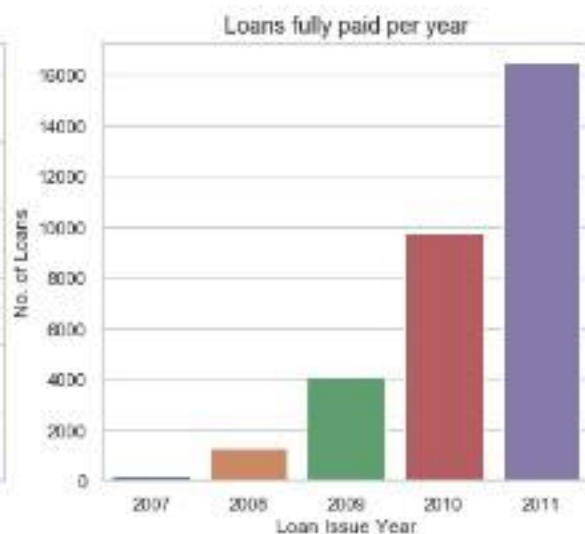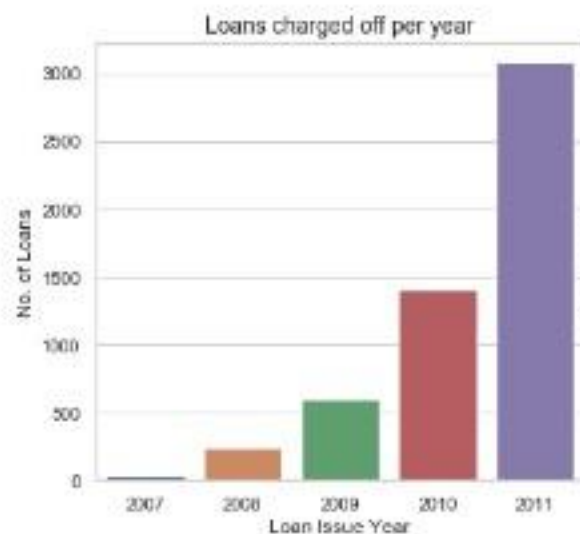
Loan Term Analysis

1. The interest rates are higher for 60 month term.
2. The average loan amount is similar for both the terms for both loan statuses.
3. The average loan amount is higher for the 60 month term.
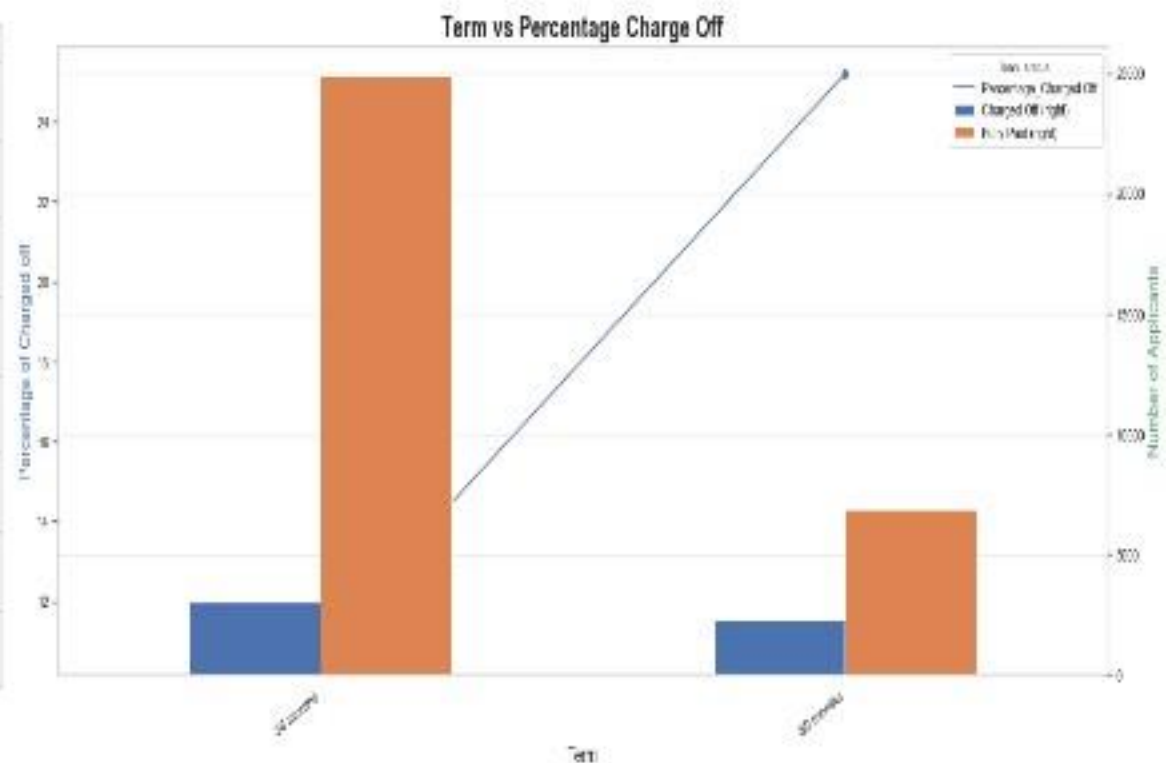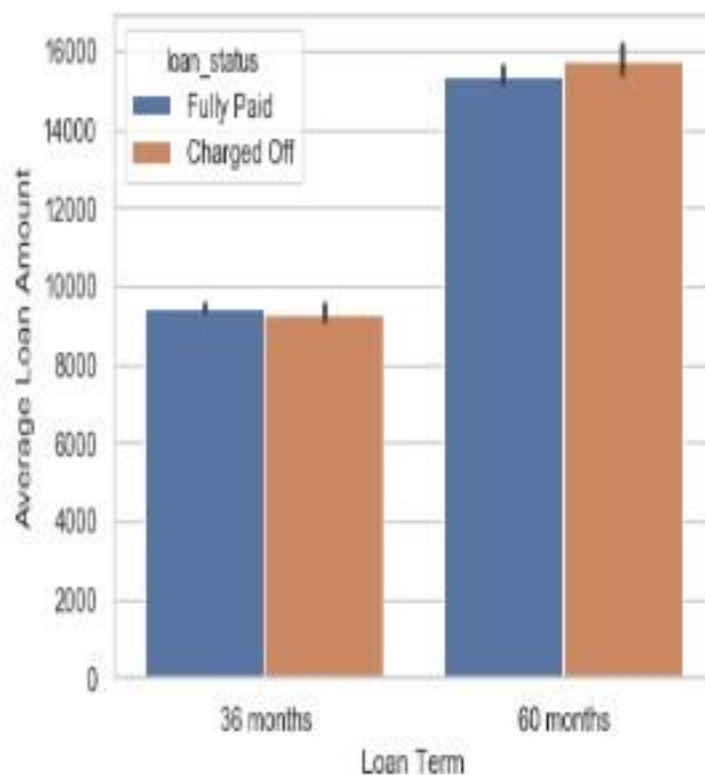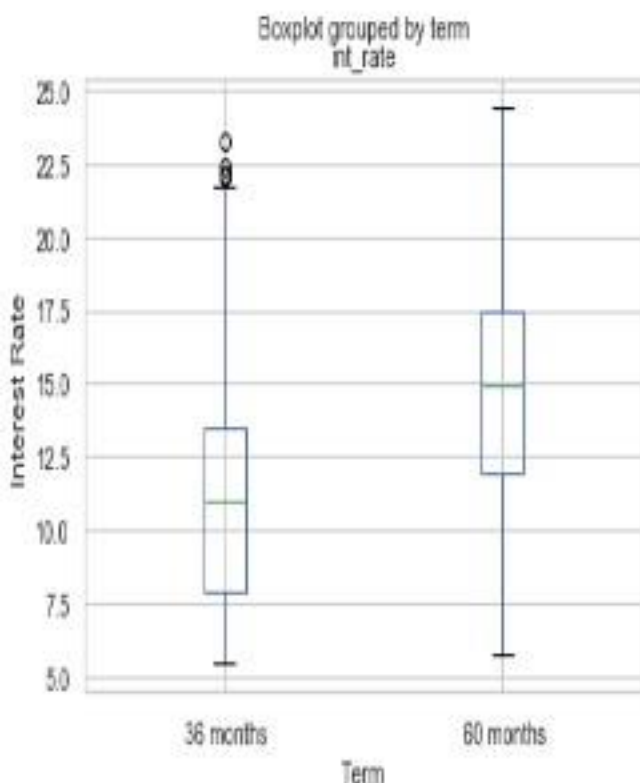
Loan Term Analysis Contd.

1. Average Interest Rate is increasing for charged off and fully paid loans year-wise in 36 month term but after inclusion of 60 months term in 2010, Average Interest Rate decreased significantly for charged off and fully paid loans.
2. 60 month term was introduced in 2010.
3. Average interest rate is higher for 60 month term.
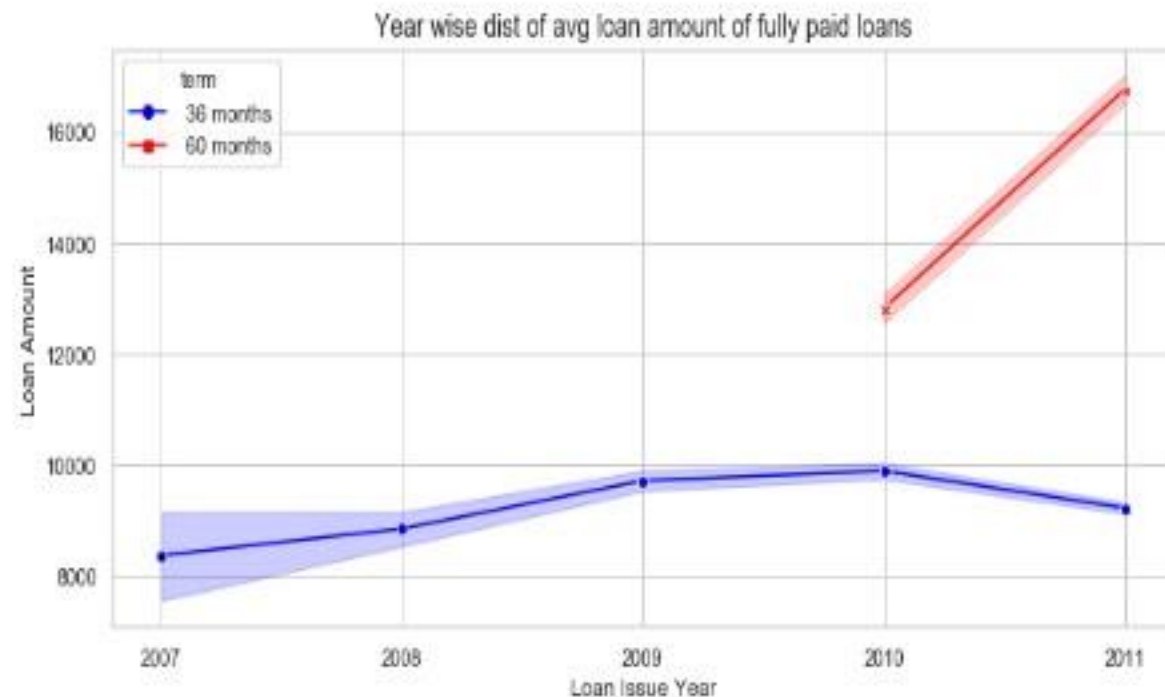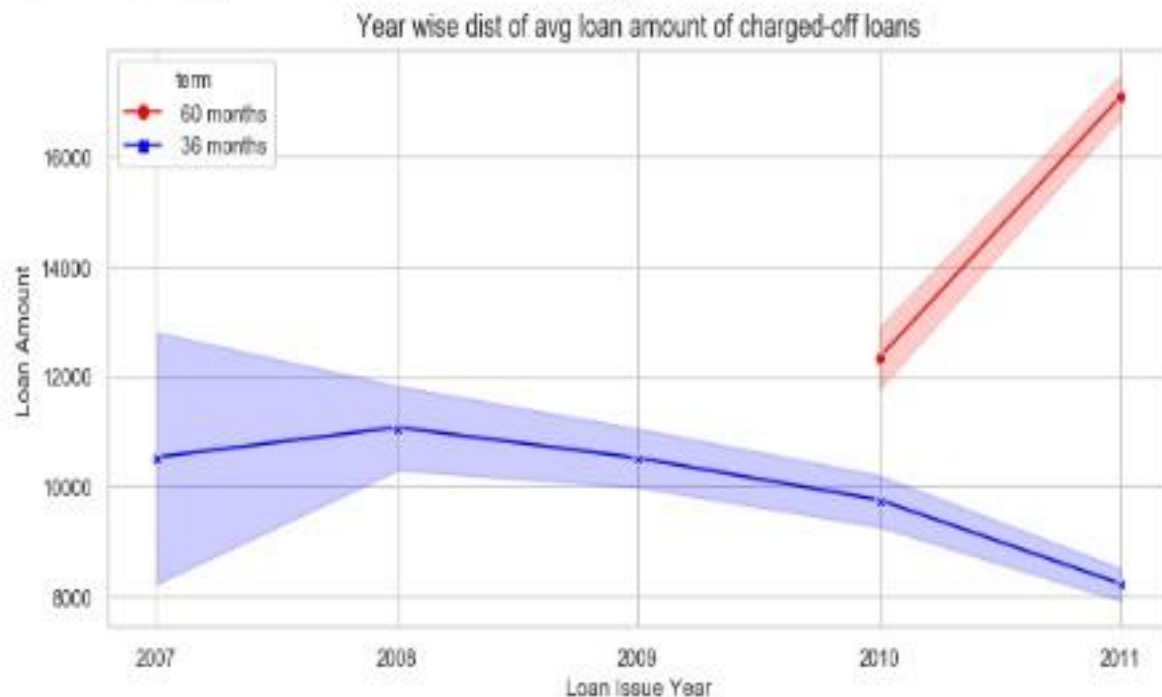
Charged off Percentage as per Loan Purpose
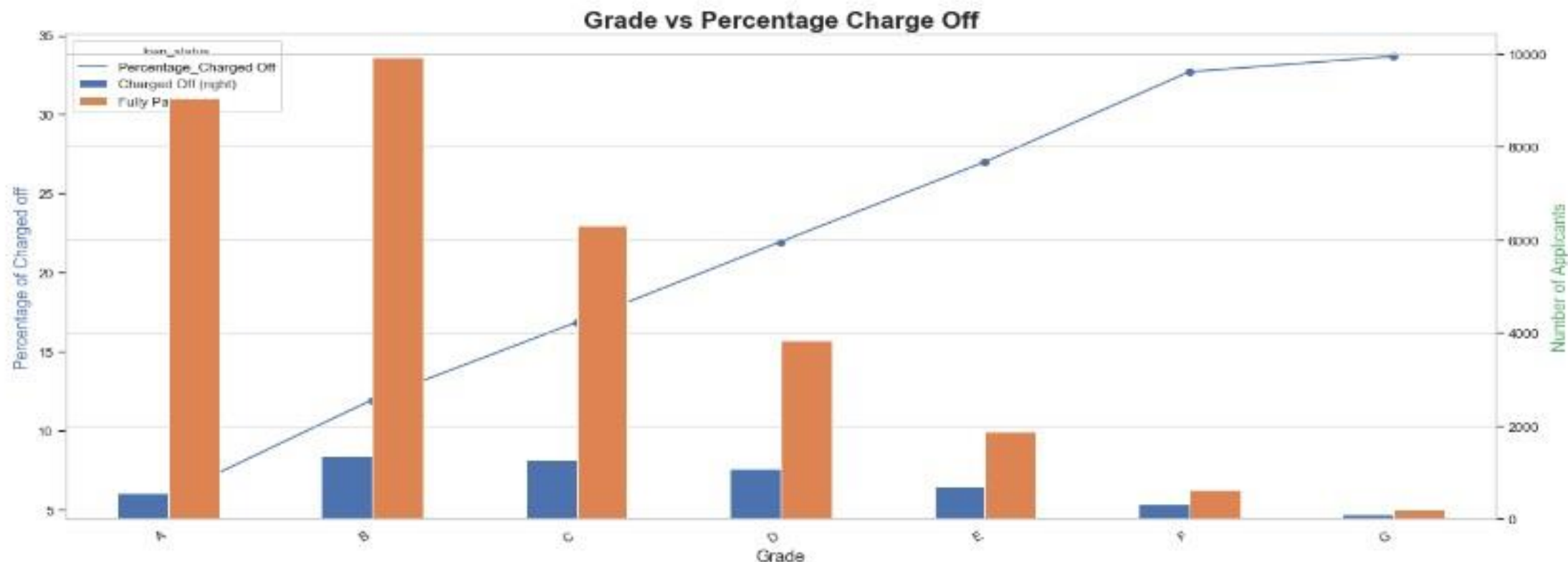
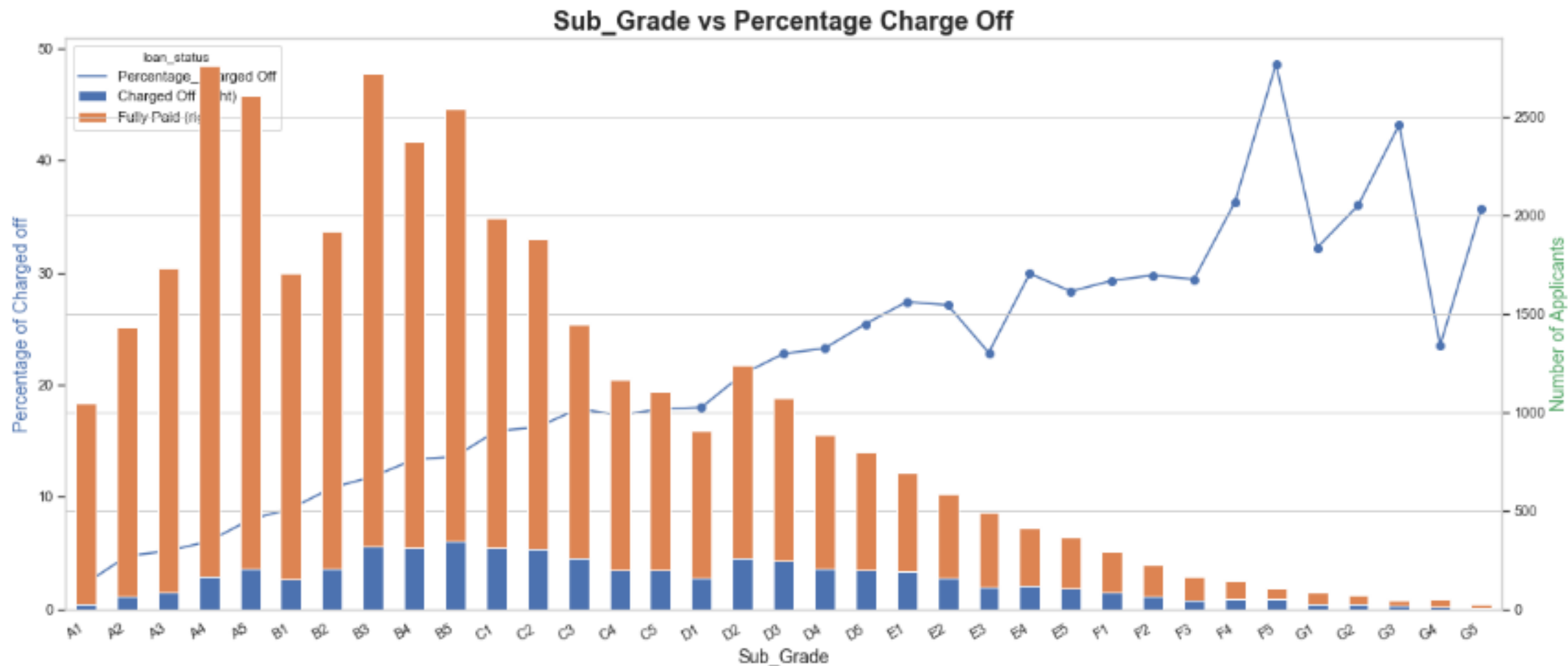Interest Rate is higher for charged off loans across all purposes.
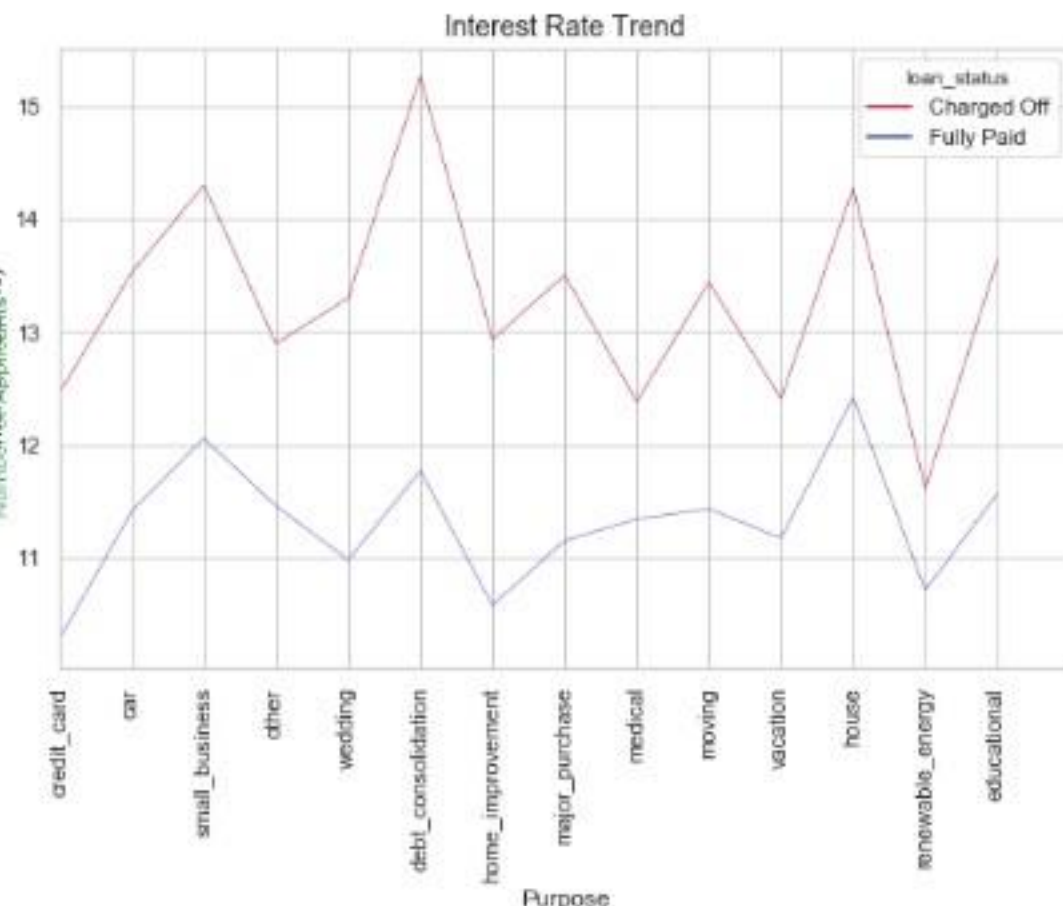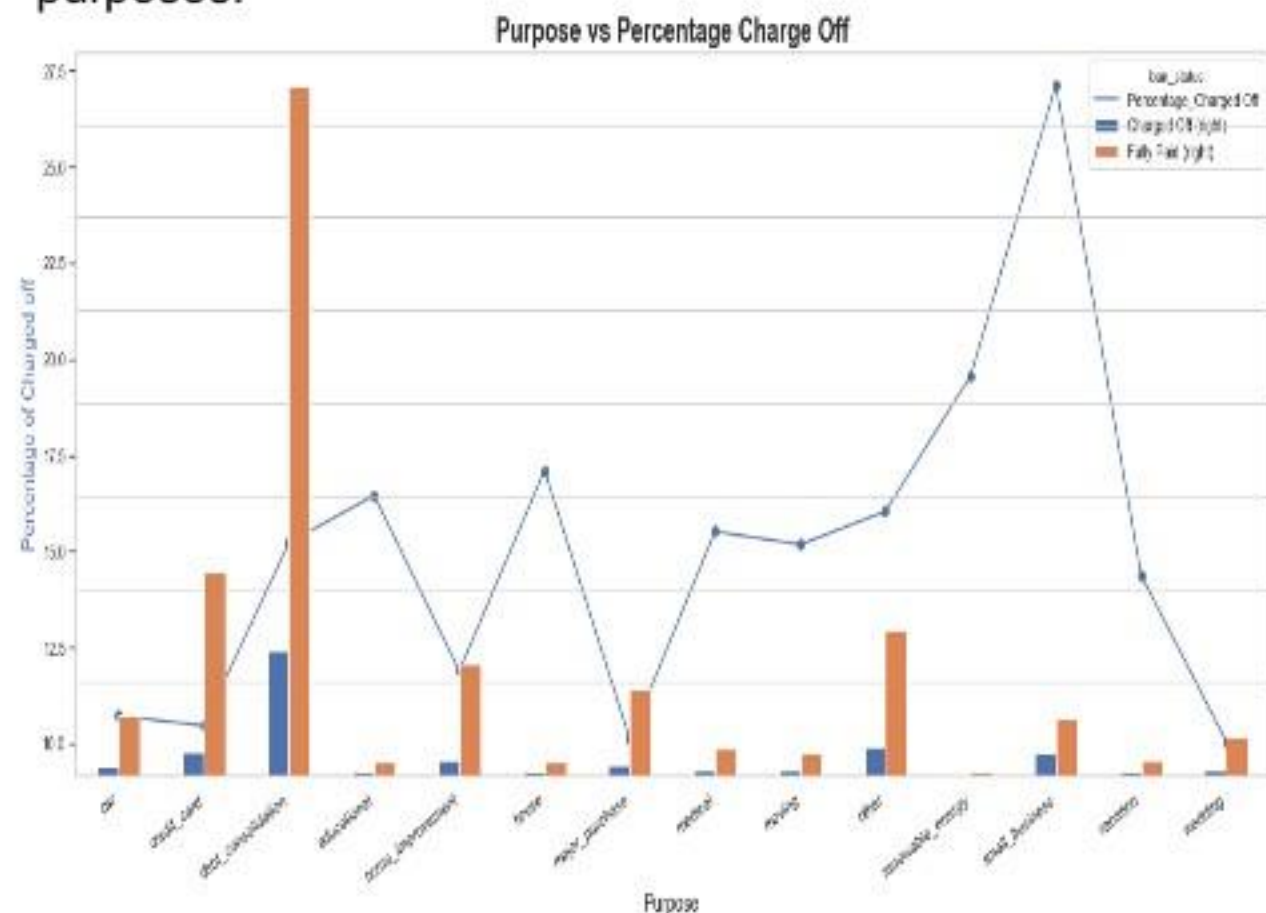
Charged off Percentage as per Grade

1. Grade **A and B** loans are **safe**. Within these the sub-grades **A4 and B3** have the highest number of loan applicants.

2. Grade **E, F, G** loans are less safe as compared to others.

Charged off Percentage as per Sub-Grade

Sub_Grade vs Percentage Charge Off

# Charged off Percentage as per Annual Income Range

1.As the Annual Income is increasing, The % of loan getting charged off is decreasing. It is minimum for income more than 100000 and maximum where annual income less than 25000.
2.Maximum loans are applied by applicants whose annual income is between 25000 and 75000. 3.Max no. of loans are charged off for people having 10+ years of experience.
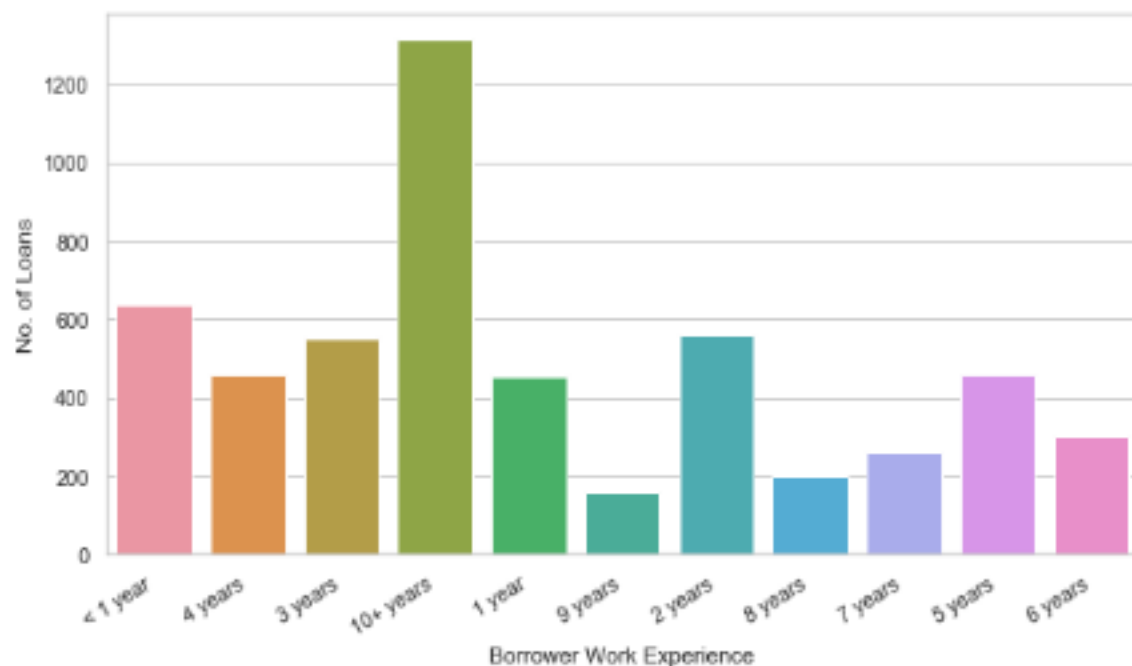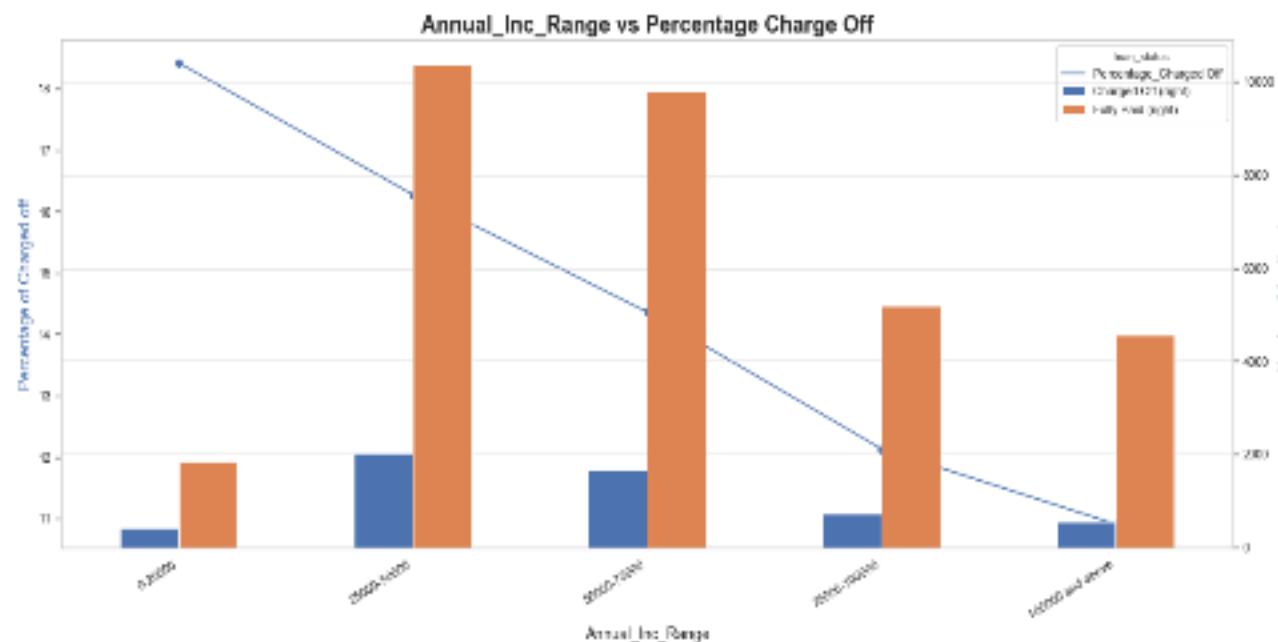
Charged off Percentage as per Loan Interest Rate

1. As the Interest Rate is increasing, The % of loan getting charged off is also increasing. It is minimum for interest rate below 7.5.
2. Maximum loans are applied by applicants where the interest rate is between 10 and 15%.

Int_Rate_Range vs Percentage Charge Off
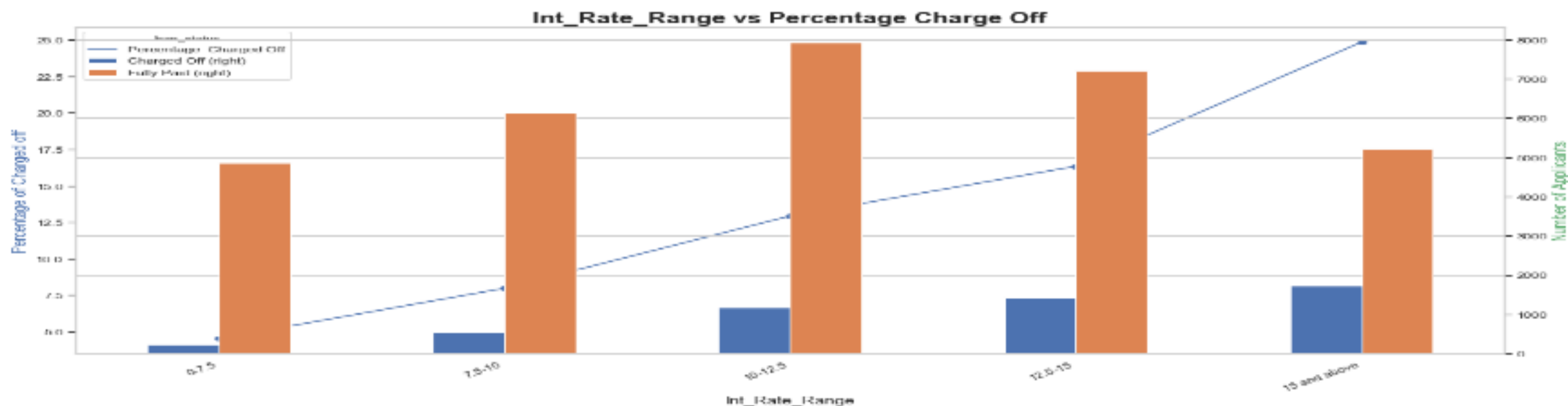
Correlation Heat-Map

1. Loan Amount, Funded Amount, Funded Amount invested and installment are highly correlated with each other.
2. Annual Income is negatively correlated with DTI.
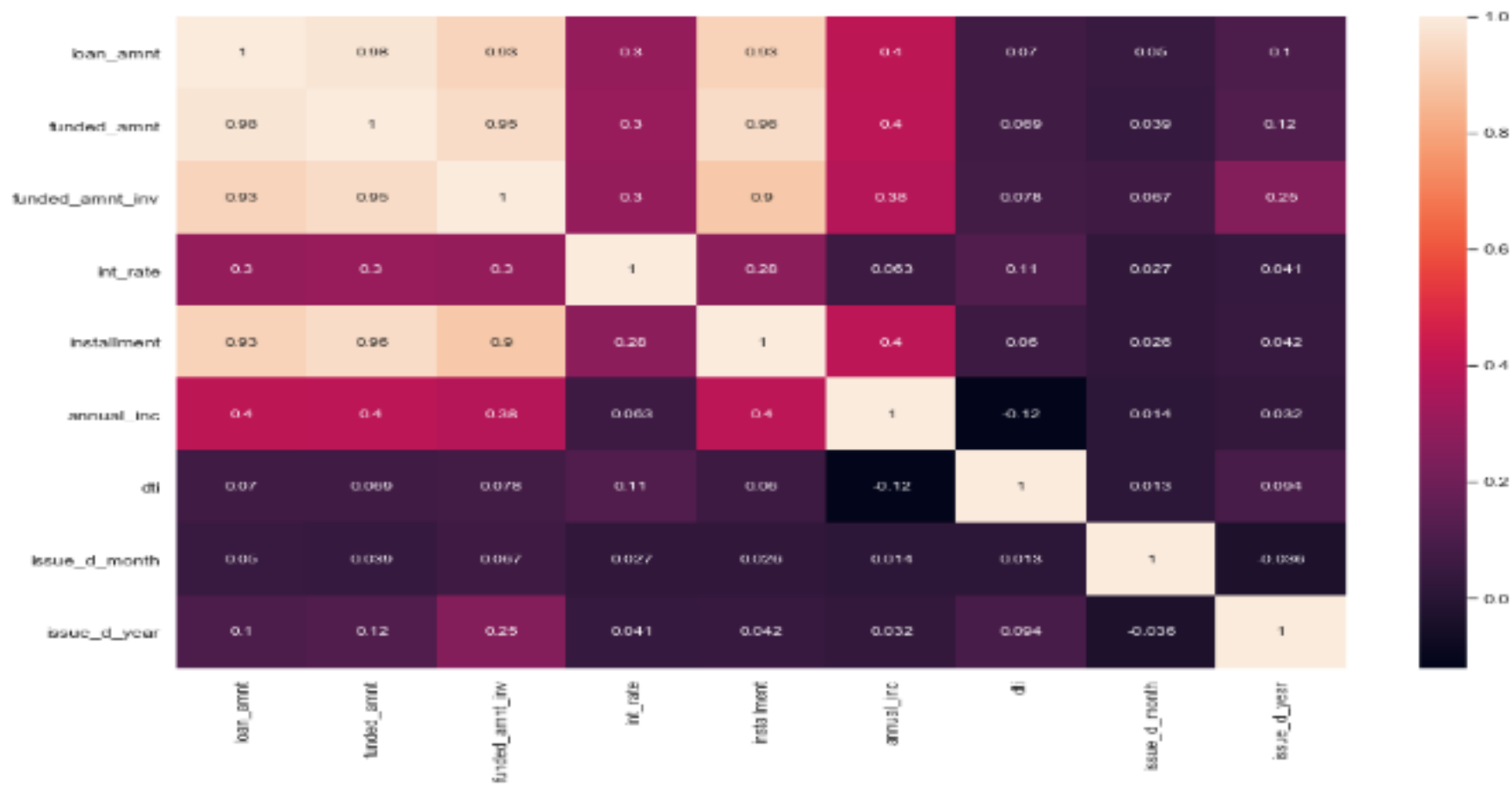
Conclusion

1. Lending Club should be wary of the loans where the purpose is **Small Business** as the percentage of a loan being charged off is maximum **27.1%**. Accepting loans for the purpose of **Weddings, major purchase, car and credit card** is highly recommended.
2. Higher the loan amount, the higher the chances of loan being charged off. Therefore Lending Company should consider accepting loans of lower amount. The maximum no. of loans, **12410** are applied where the **loan amount** ranges between 5000 and 10000. For this range the probability of the the loan getting charged off is very less. Hence the risk factor is low for lending club.
3. Lending Club should consider accepting more loans from applicants whose annual income is **greater than 100000** as their probability of charge off is minimum.
4. Lending Club should consider accepting more loans where interest rate is **less than 7.5%** as their probability of charge off is minimum.
5. Lending Club should consider accepting more loans of grade **A and B**. It should be cautious of loans falling in grades **E,F and G**
6. Lending Club should consider accepting more loans from people who **owns** a house.
7. The number of loan applicants are increasing with each passing year bringing more business to Lending Club which is leading to more loans getting accepted as well as more loans getting charged off. Moreover the last few months of the year are quite critical as most of the loans are either fully paid or charged off at that time.
8. Lending Club should accept more loans for the term of **36 months** as the % of charged off loans is less and the no. of loan applicants are more. Also the avg fully paid loan amount is increasing over the years and avg charged off lon amount is decreasing over the years.