

# Lending Club Case Study

Exploring Patterns in Loan Applicants

Group Members  
Sravanthi Chundi

Srinivas NV

# Introduction

---

**Objective:** Analyze factors influencing loan defaults using exploratory data analysis (EDA).

**Dataset:** It contains the complete loan data for all loans issued through the time period 2007 to 2011.

## **Problem Statement:**

You work for a consumer finance company which specialises in lending various types of loans to urban customers. When the company 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:

If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.  
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

The data given below contains information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

# Approach

Load the data into a data frame.

Drop the columns with null values, single unique value, More than 30% missing values

Standardise the data into correct data type.

Generate multiple graphs based on the data

Generate correlation heat maps

Data  
Cleaning

EDA

Visualization

Insight  
Generation

Create derived matrix for year and month

Univariate Analysis

Bivariate Analysis

Multivariate Analysis

Analyse the Graphs and generate the insights

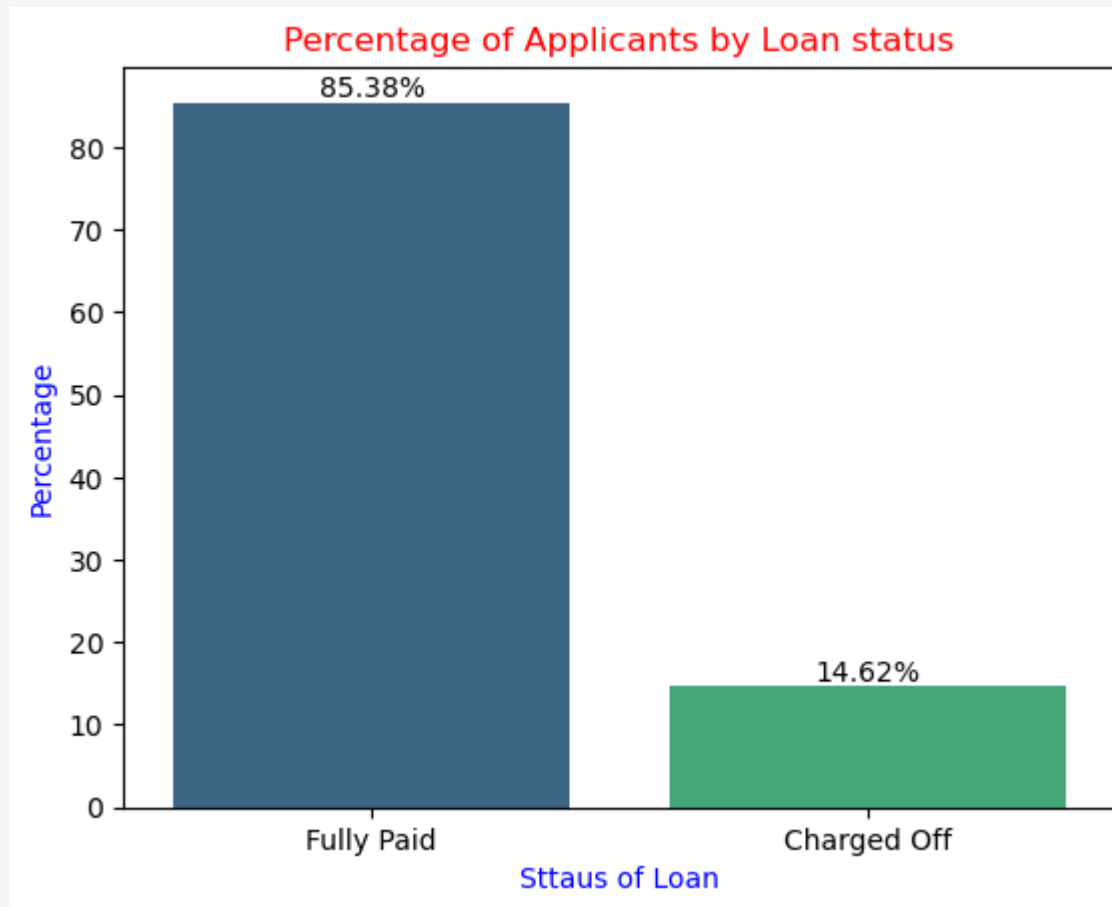
# Data Understanding

---

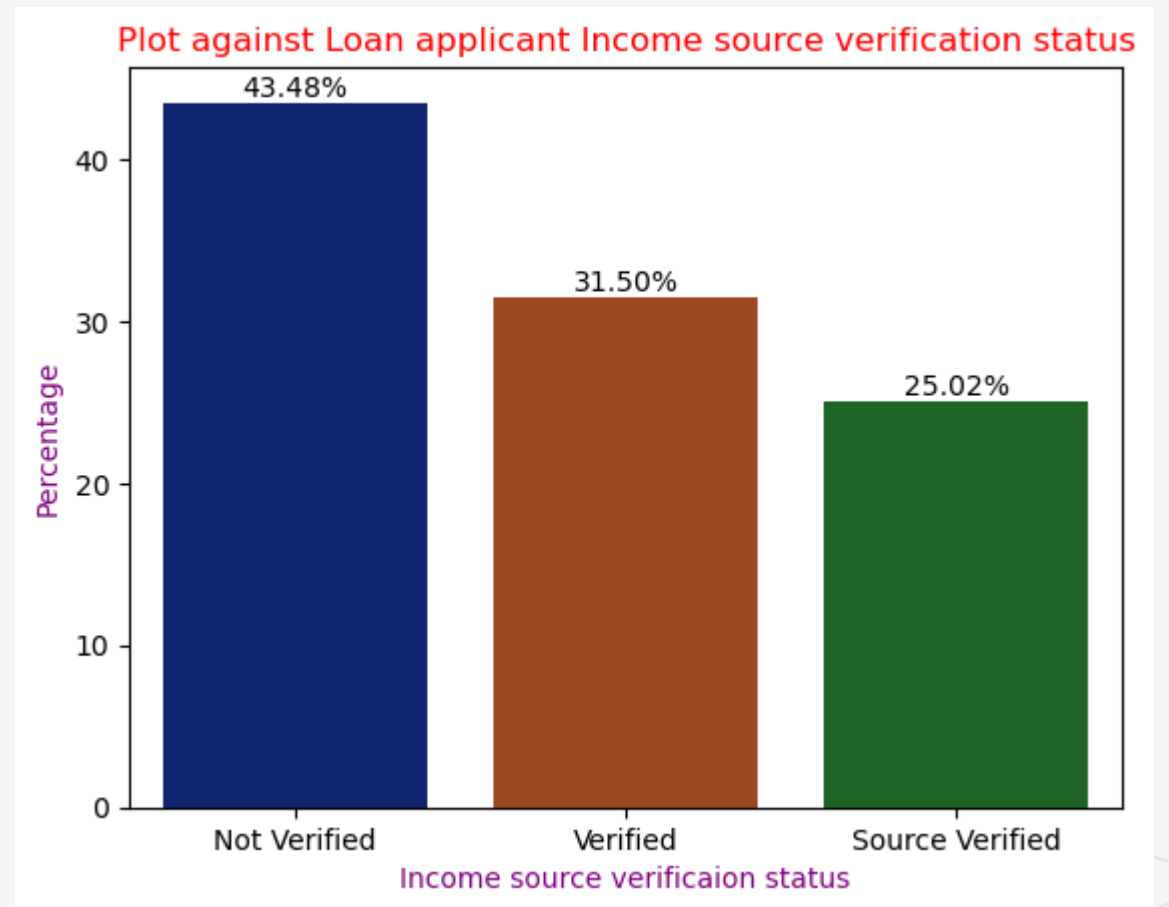
## Key Variables:

- `annual_inc` : The self-reported annual income provided by the borrower during registration.
- `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.
- `loan_amnt` : The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
- `loan_status` : Current status of the loan
- `Grade` : LC assigned loan grade
- `home_ownership` : The home ownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER.
- `verification_status` : Indicates if income was verified by LC, not verified, or if the income source was verified

# Distribution of overall Applicants by Loan status and Income verified status

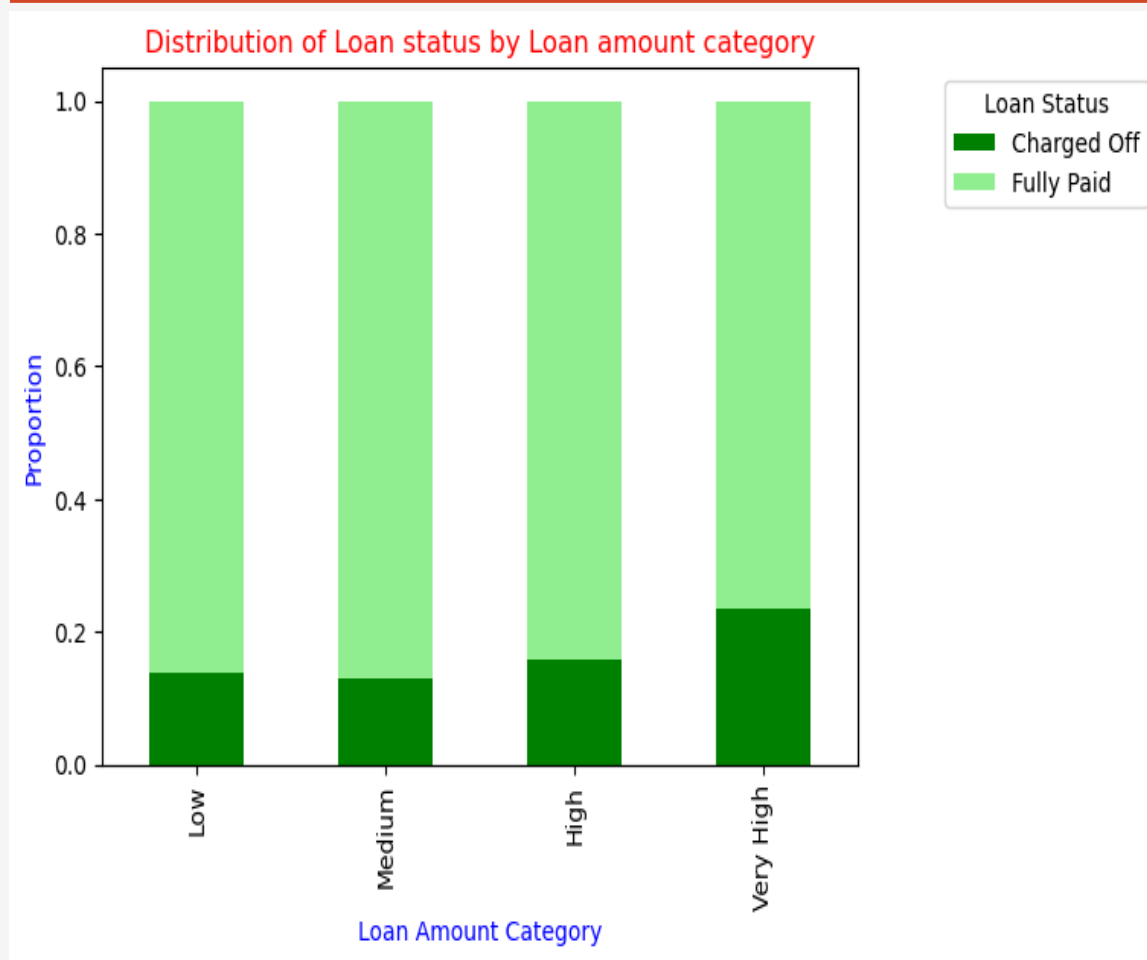


1. 85% of loan applicants paid the loan fully
2. **Nearly 15% of loan applicants are defaulted**, and could not paid the loan

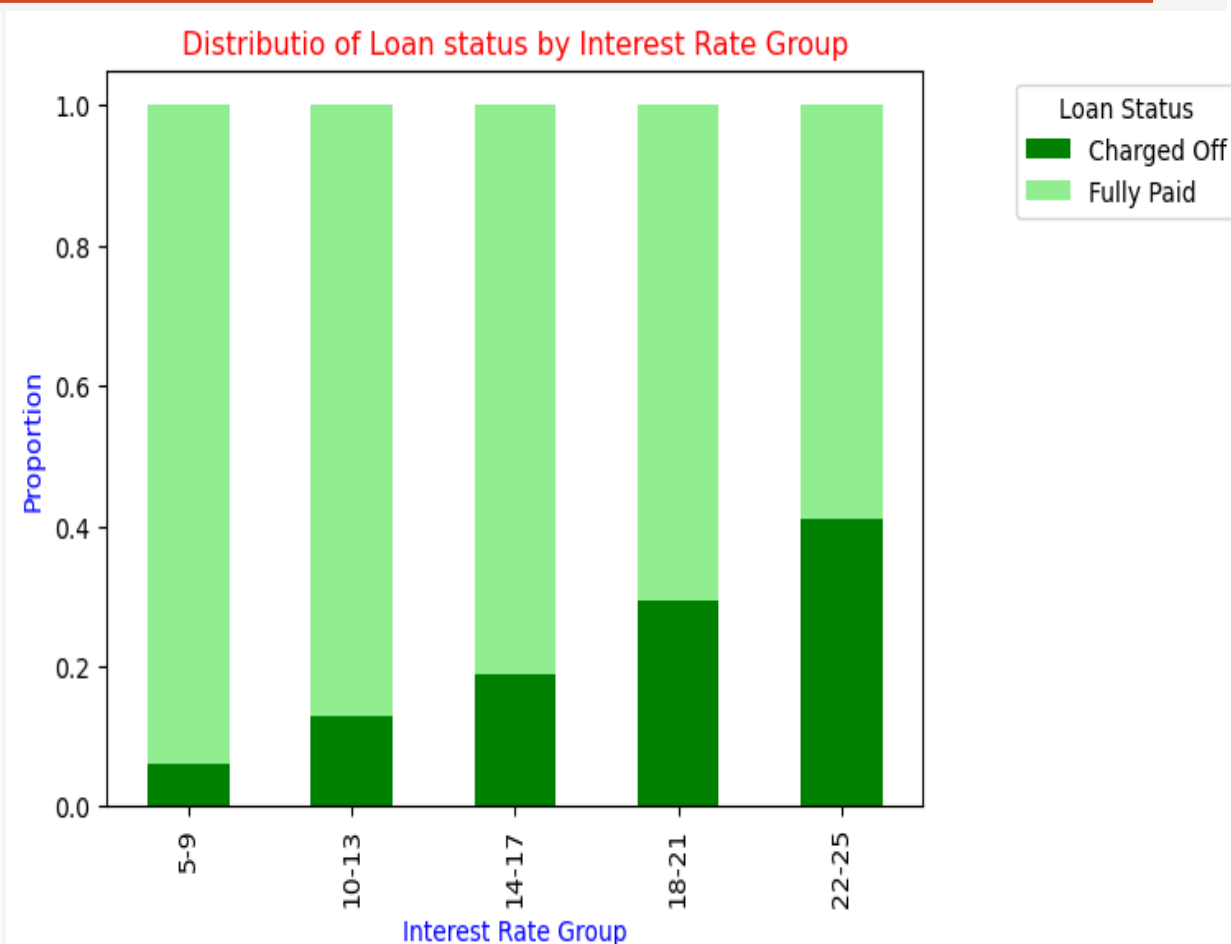


1. Plot shows, most of loan applicants income source is Not verified, it creates a risk

# Loan Status Vs Loan Amount Category

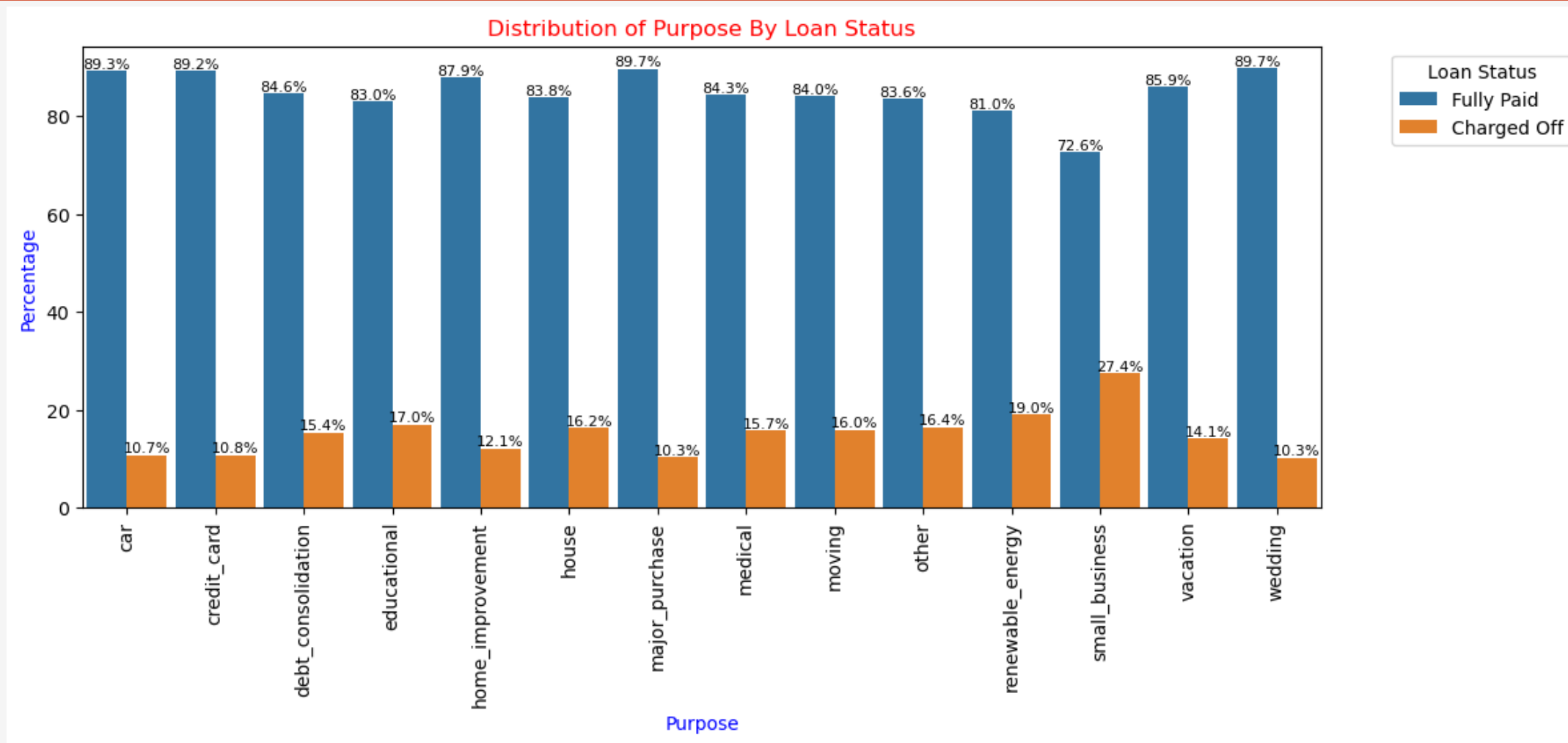


Plot shows, for **Very High** Loan Amounts the **Charged Off** are more.



1. Plot shows - As interest increases, the proportion of **Charged Off** loans increased and at the same time the proportion of **Fully Paid** loans decreased
2. Higher Interest Rate groups 18-21% and 22-25% , loan applicants for **Charged Off** loans increased. It shows higher Interest rate is a cause for "Defaults"

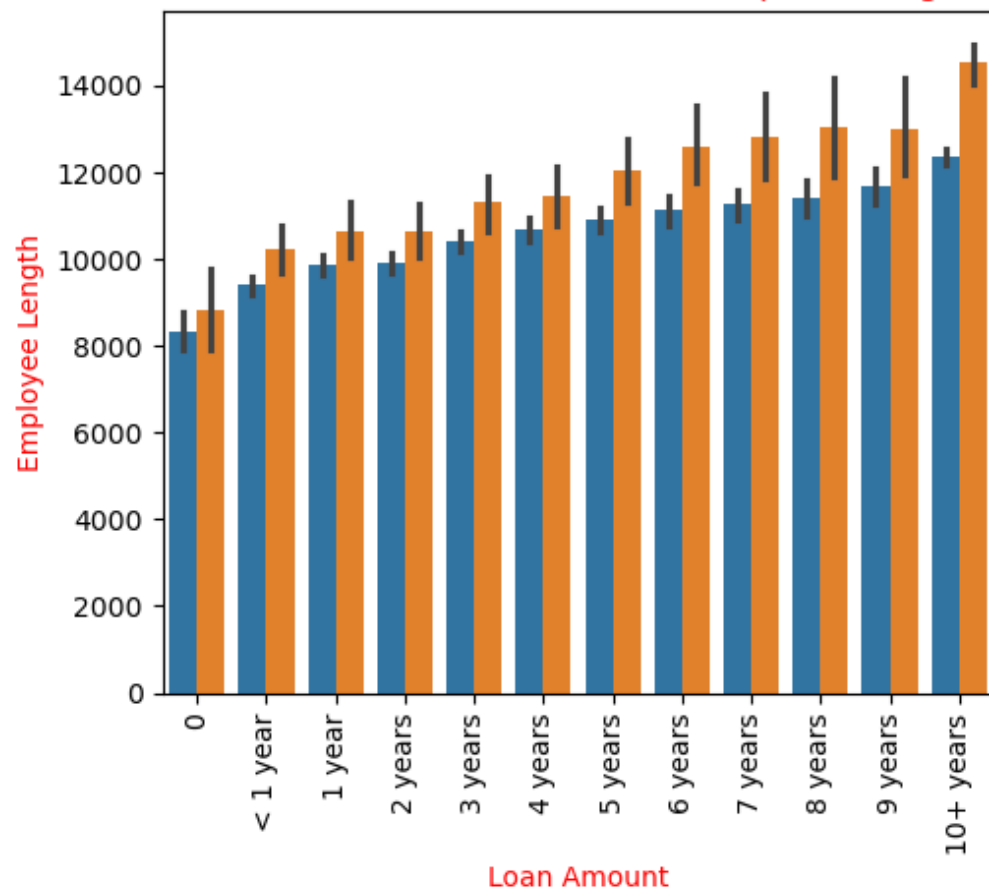
# Loan Purpose by Loan Status



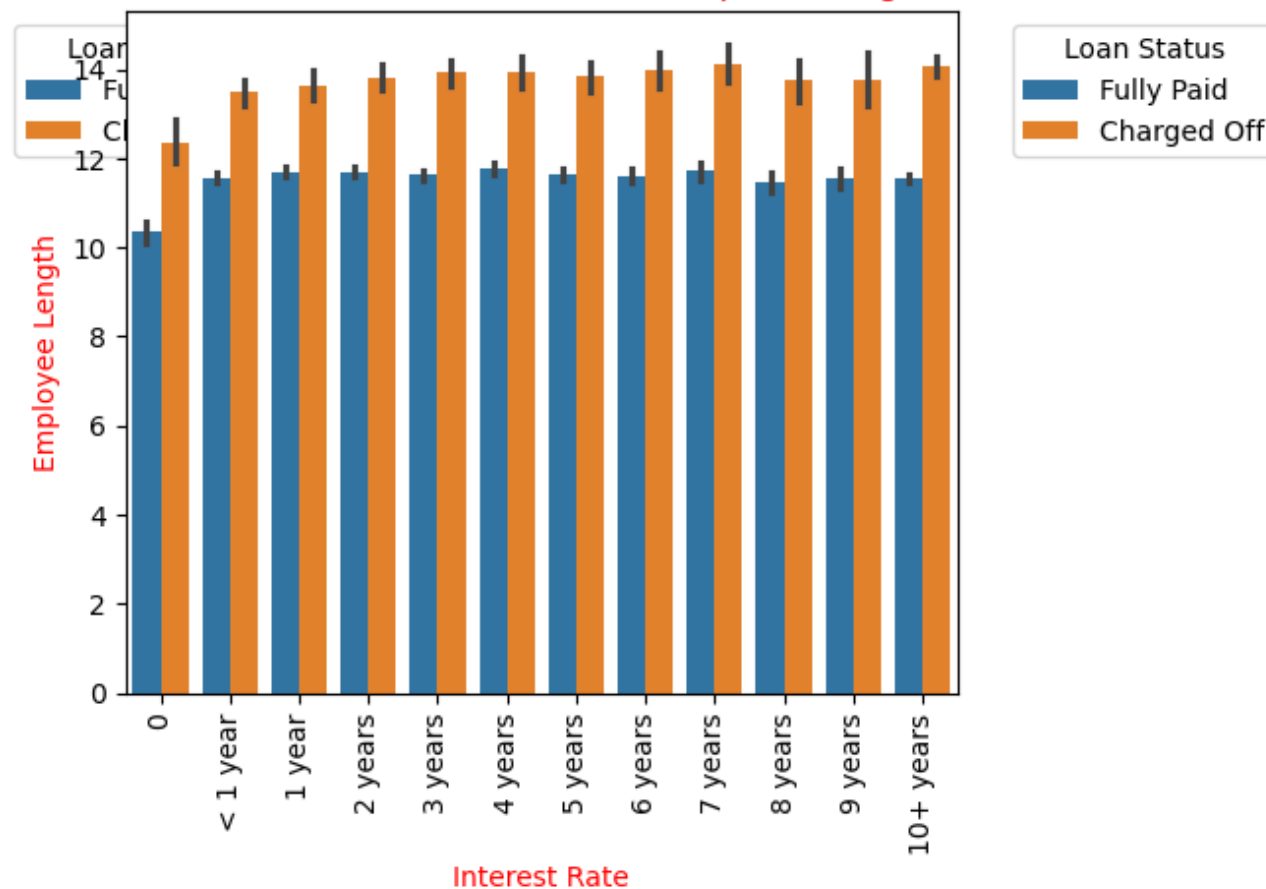
Bar graphs denotes that loans taken for small business, educational have more chances of defaulters

# Loan Amount Vs Emp Length and Interest Rate

Distribution of Loan Amount Vs Employee Length



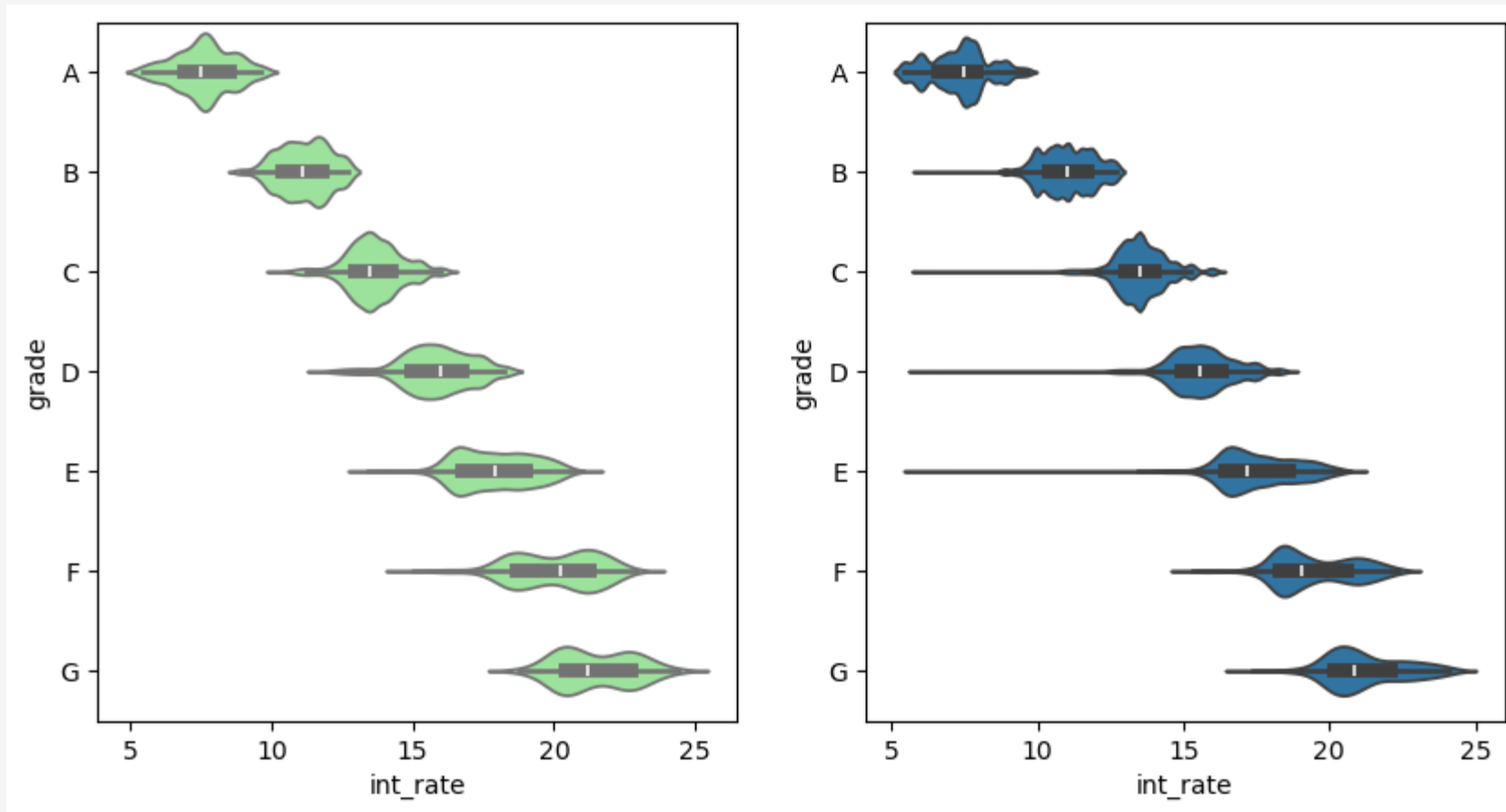
Distribution of Interest Rate Vs Employee Length



1. **Employee Length and Loan Amount:** Applicants with longer employment length have higher loan amounts, then the less employment length.
2. **Employee Length and Interest Rates:** Applicants with shorter employment lengths are having higher interest rates, and "Charged Off" loans generally have higher interest rates than "Fully Paid" loans.

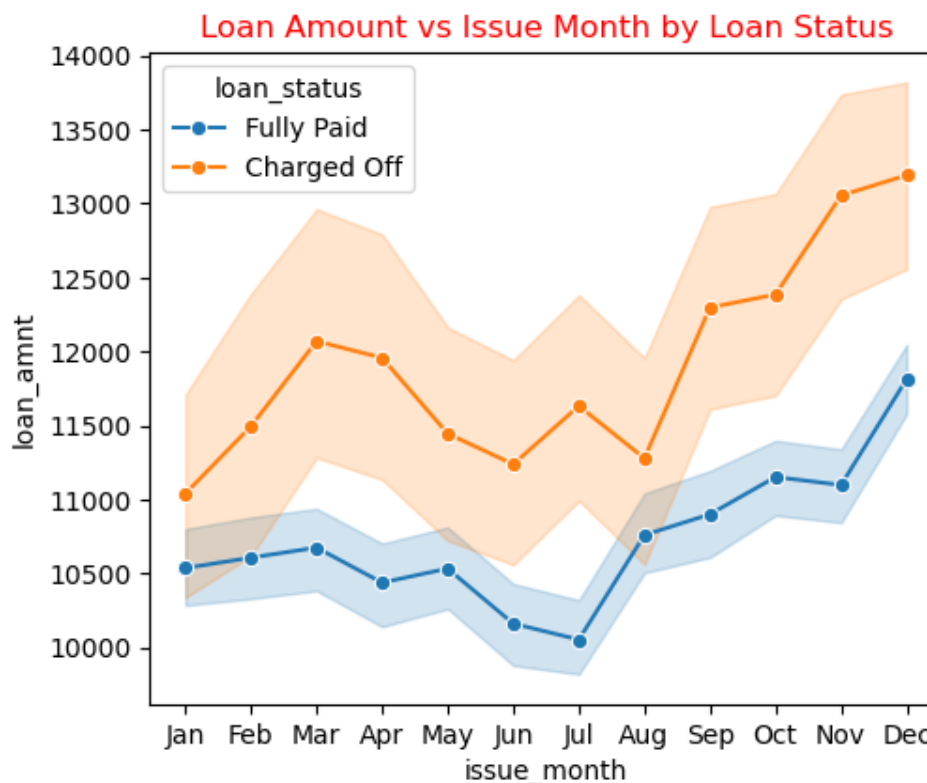
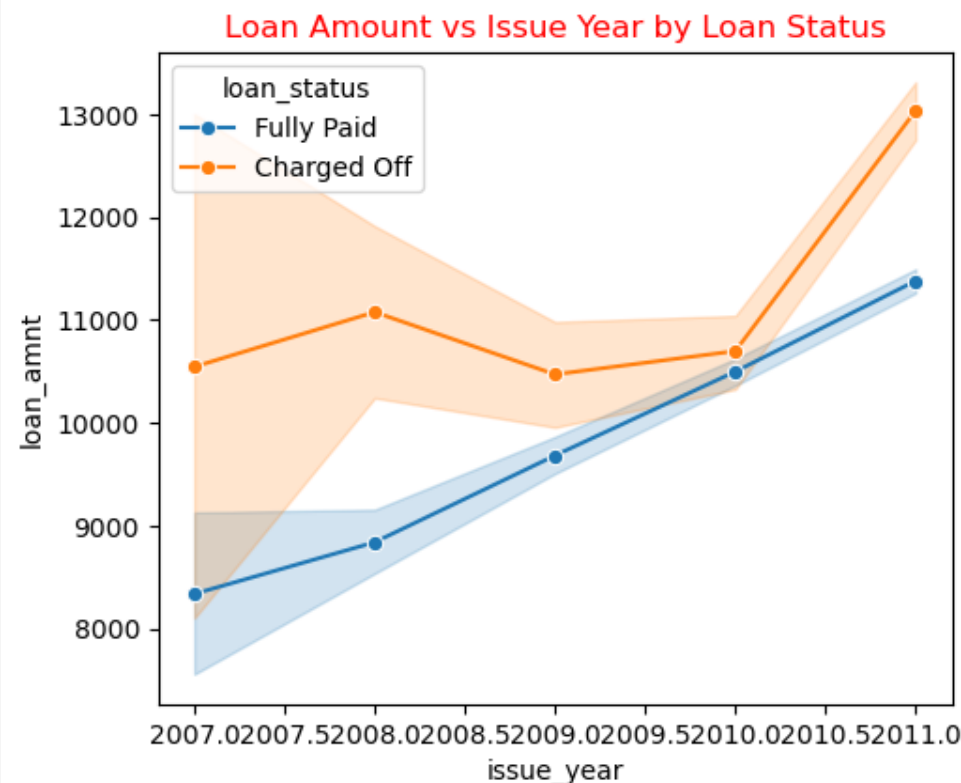


# Interest Rate Vs Grade



1. The interest rates are increasing with each grade.
2. Low grades/ Low Creditworthiness (F and G) are associated with higher interest rates.
3. It seems Interest rates are higher for low grades. Lenders charging higher interest rates for risk loans.

# Loan Amount Vs Issue Year By Loan Status and Loan Amount Vs Issue Month By Loan Status



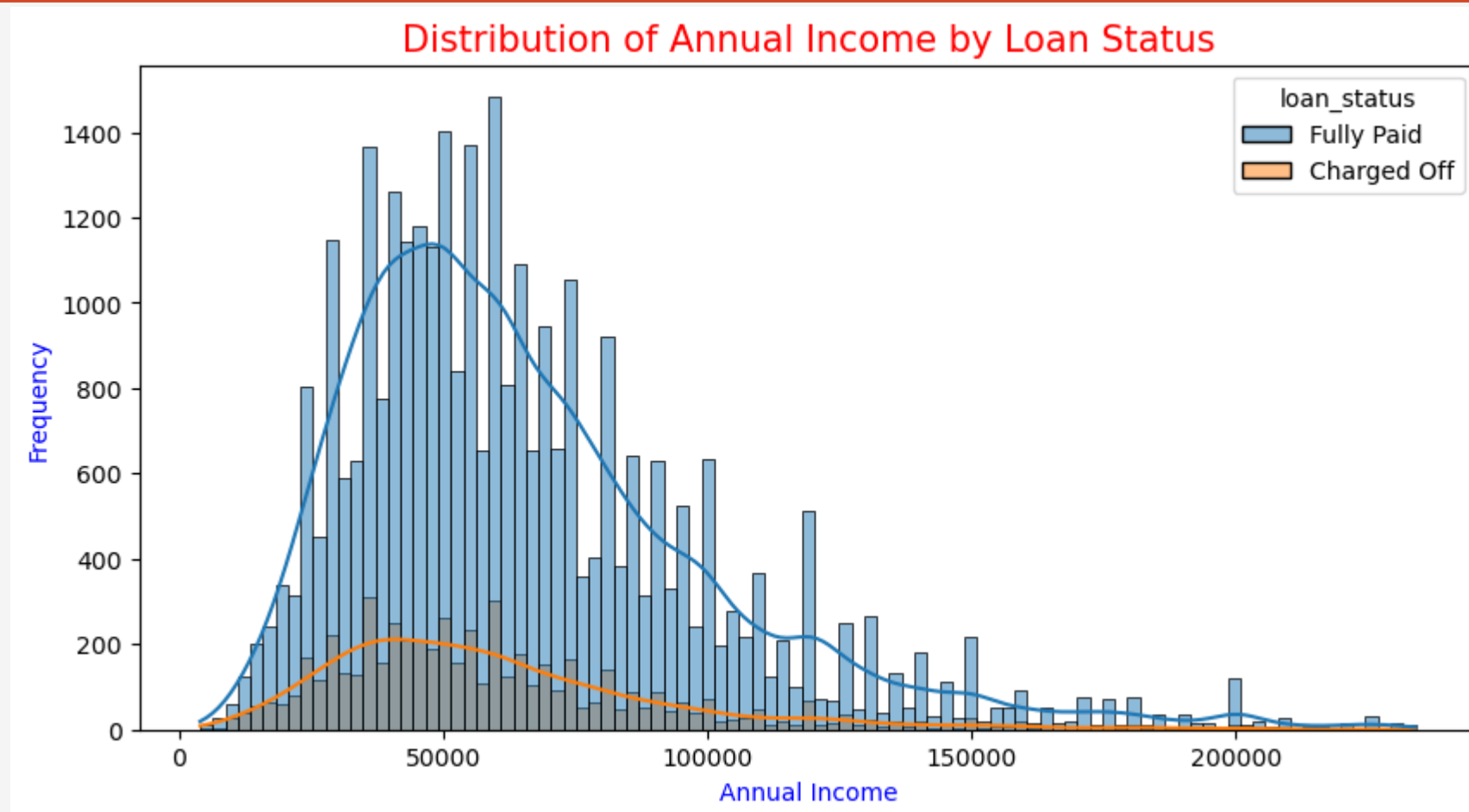
## Loan Amount vs Issue Year

1. Loan Amounts for Charged Off are usually higher when compared to Fully Paid across all the Years
2. For Fully Paid loans the loan amount has increased consistently over years and observed peak in 2011
3. For Charged Off loans, we can see it increased and decreased till 2010 and observed a sudden peak in 2011

## Loan Amount vs Issue Month

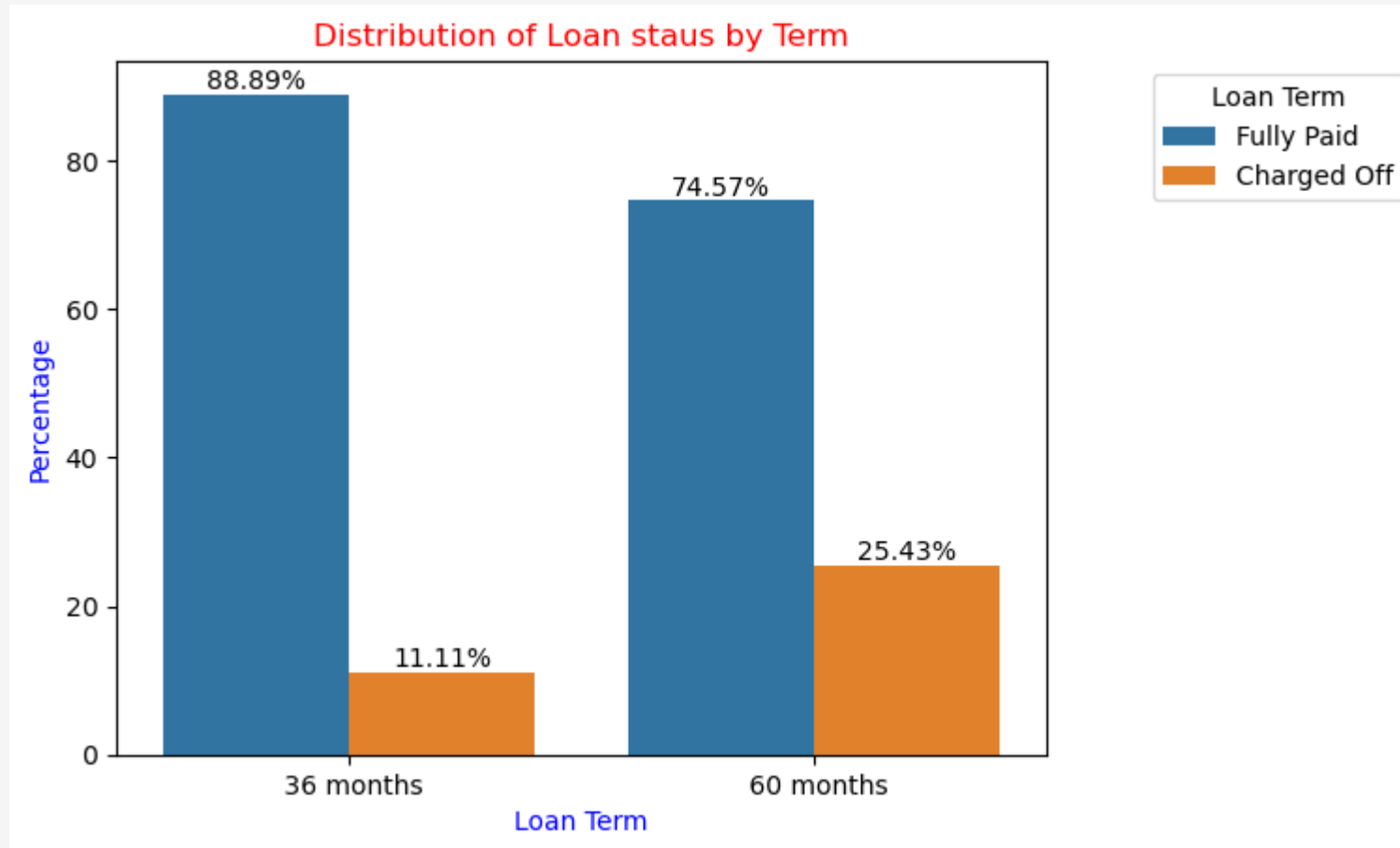
1. Loan Amounts for Charged Off are usually higher when compared to Fully Paid across all Months
2. For "Fully Paid" loans show less variability till for first few months and increased in later months of the year.
3. For Charged Off loans show high variability and increased in later months of the year.

# Annual Income Vs Loan Status



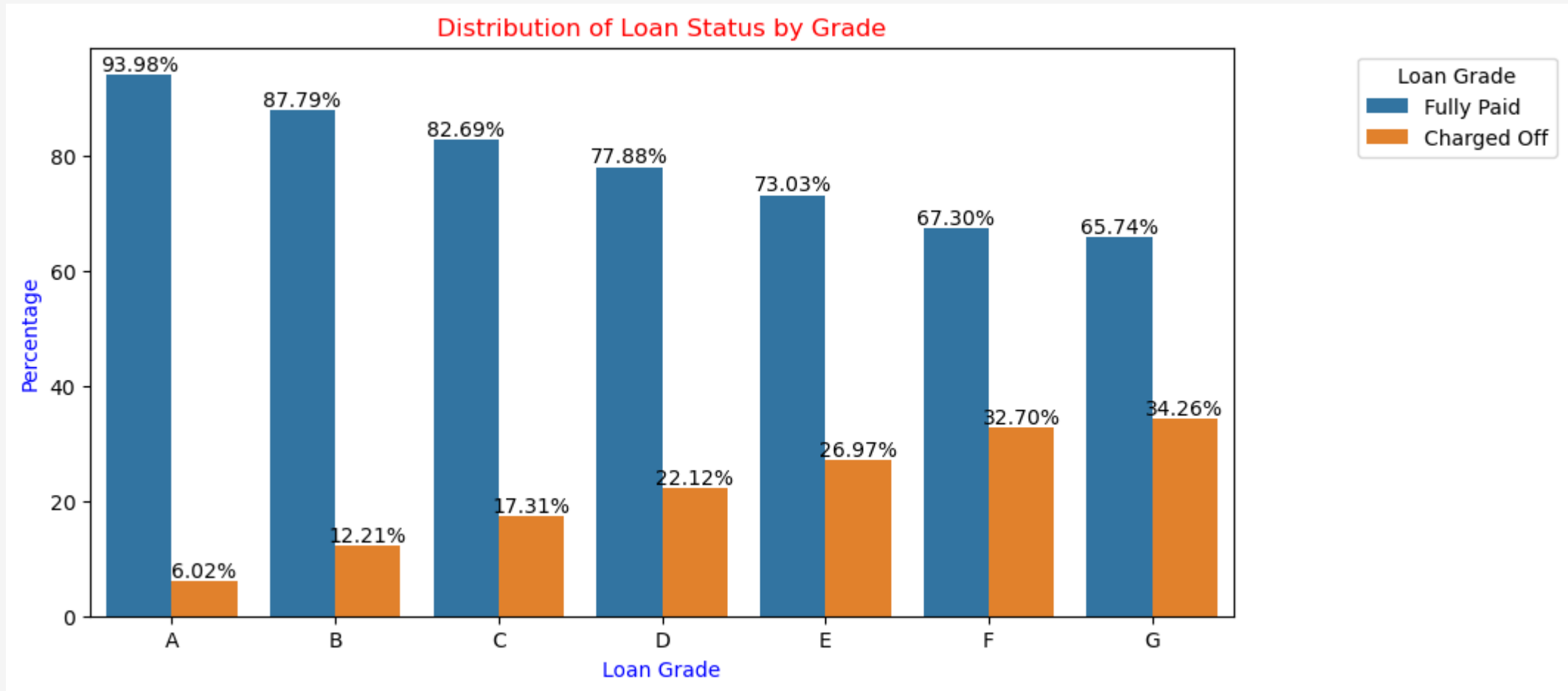
1. Distribution of Annual Income is rightly skewed for both Fully paid and Charged off loan status.
2. Its indicating most of the loan applicants havig low to moderate income and very few have high income
3. **Most Applicatons with lower income have Charged Off loans.** Most of Higher annual income applicants are "Fully paid".

# Distribution of Loan status by Term



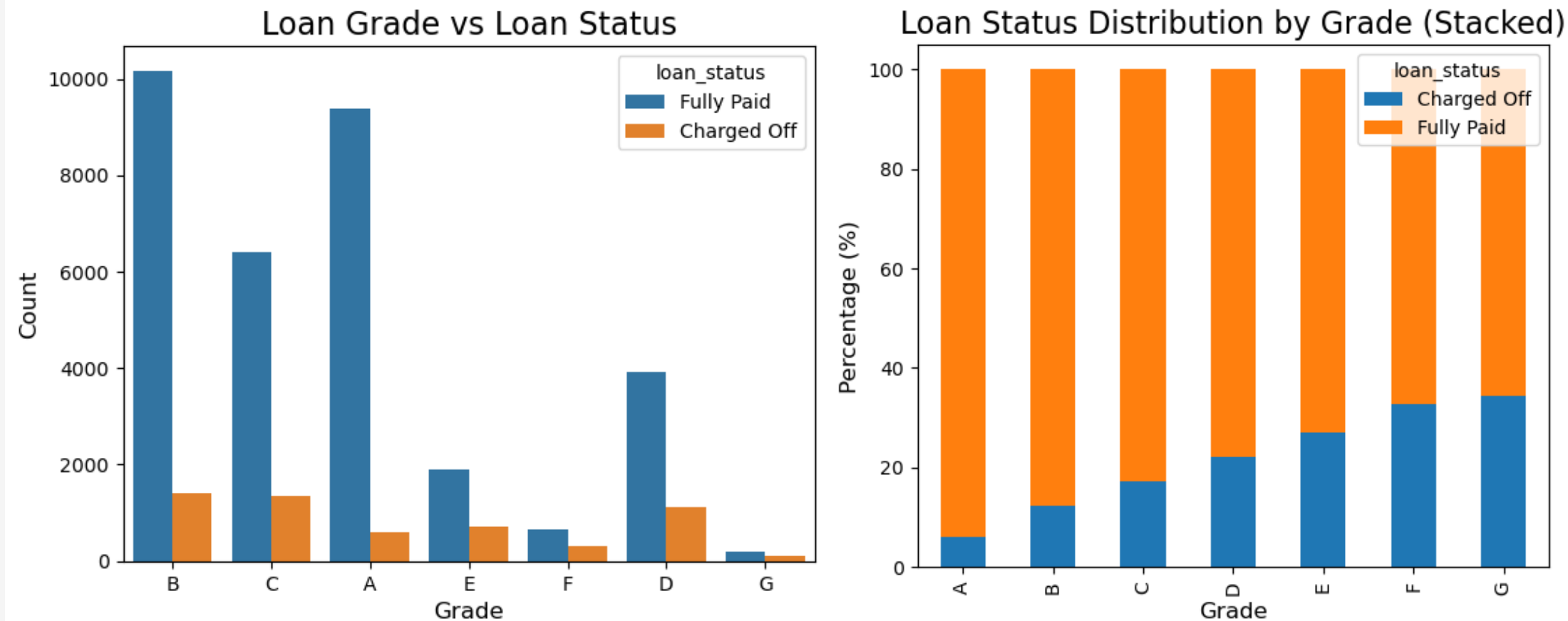
1. Plot shows, The number of **Defaults for Charged of loan are higher in 60 months loan term than 36 months**
2. Fully paid, term 36 months loan apploicants are paying loan more than 60 months term

# Distribution of Loan status by Term



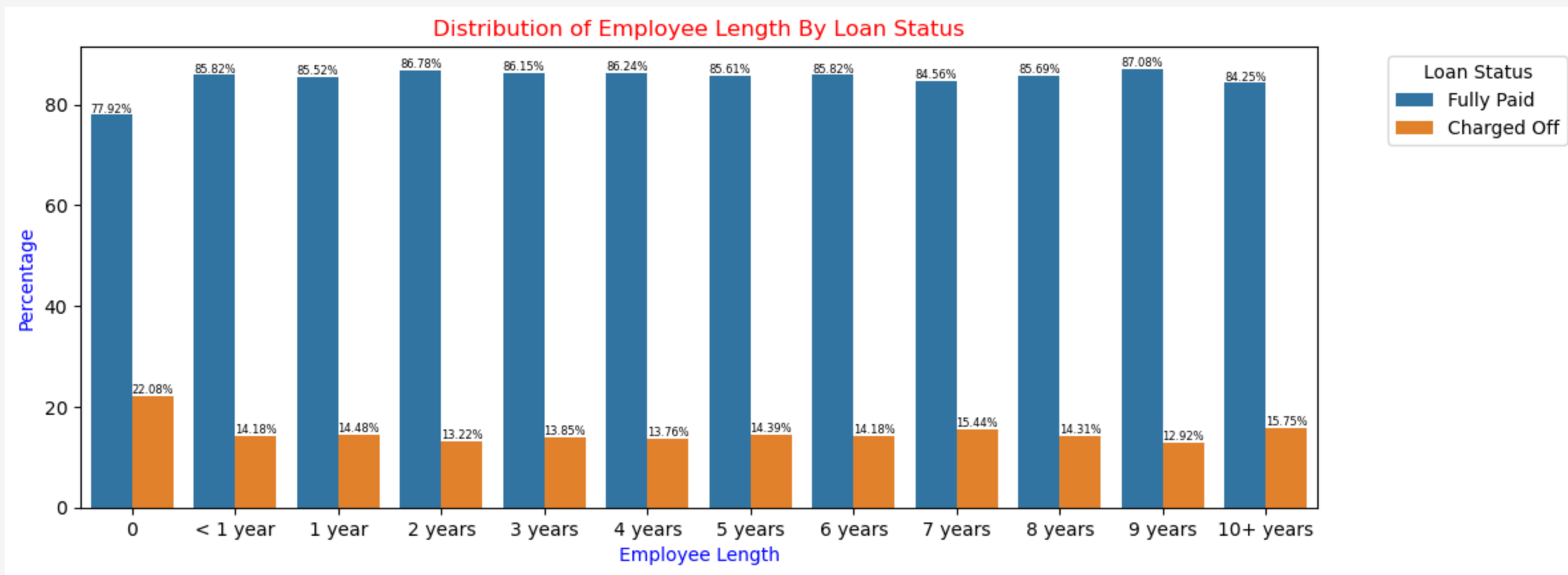
1. Plot shows, The number of **Defaults for Charged of loan are higher in 60 months loan term than 36 months**
2. Fully paid, term 36 months loan apploicants are paying loan more than 60 months term

# Grade Vs Loan Status



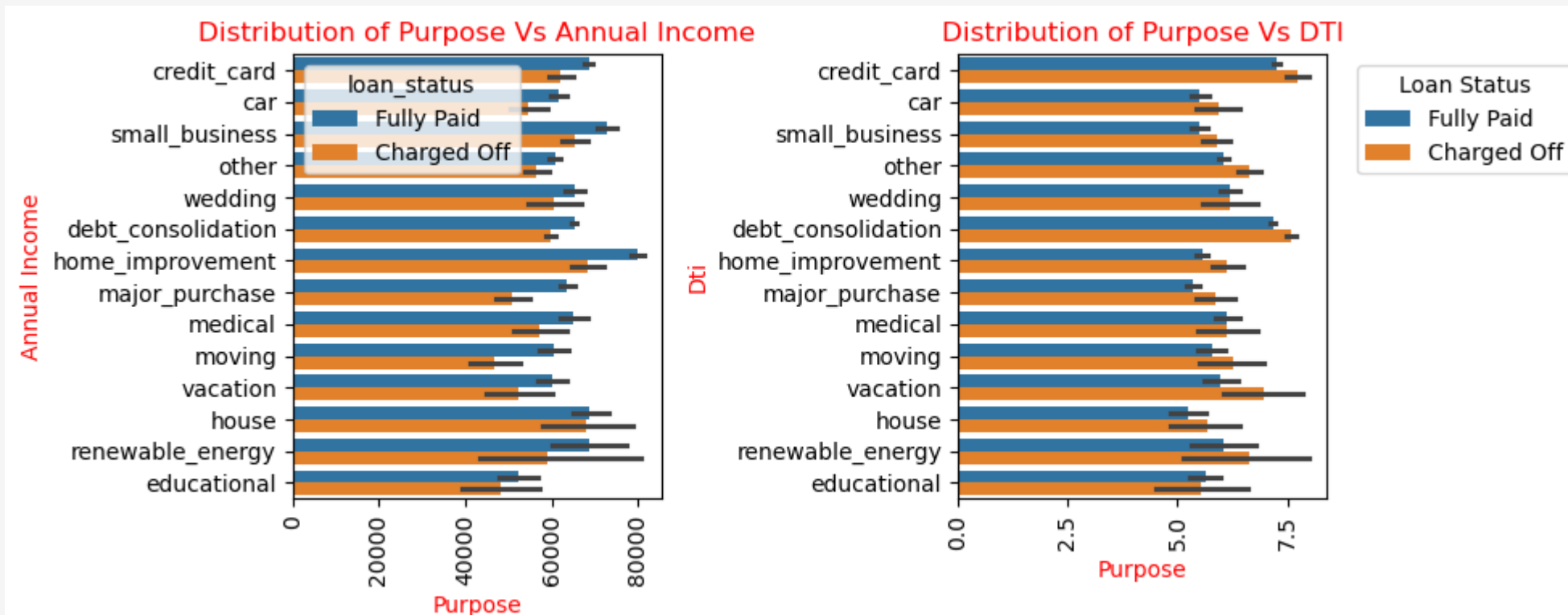
Plot shows that Applicants with Grade A,B,C has high chances of Fully paid. In other words, lower the grades higher the risk of defaulters.

# Employee Length By Loan Status



Graph denotes that Employees with More than 10 years of experience has high chances of getting verified over others with less experience

# Annual Income Vs Purpose and DTI Vs Purpose

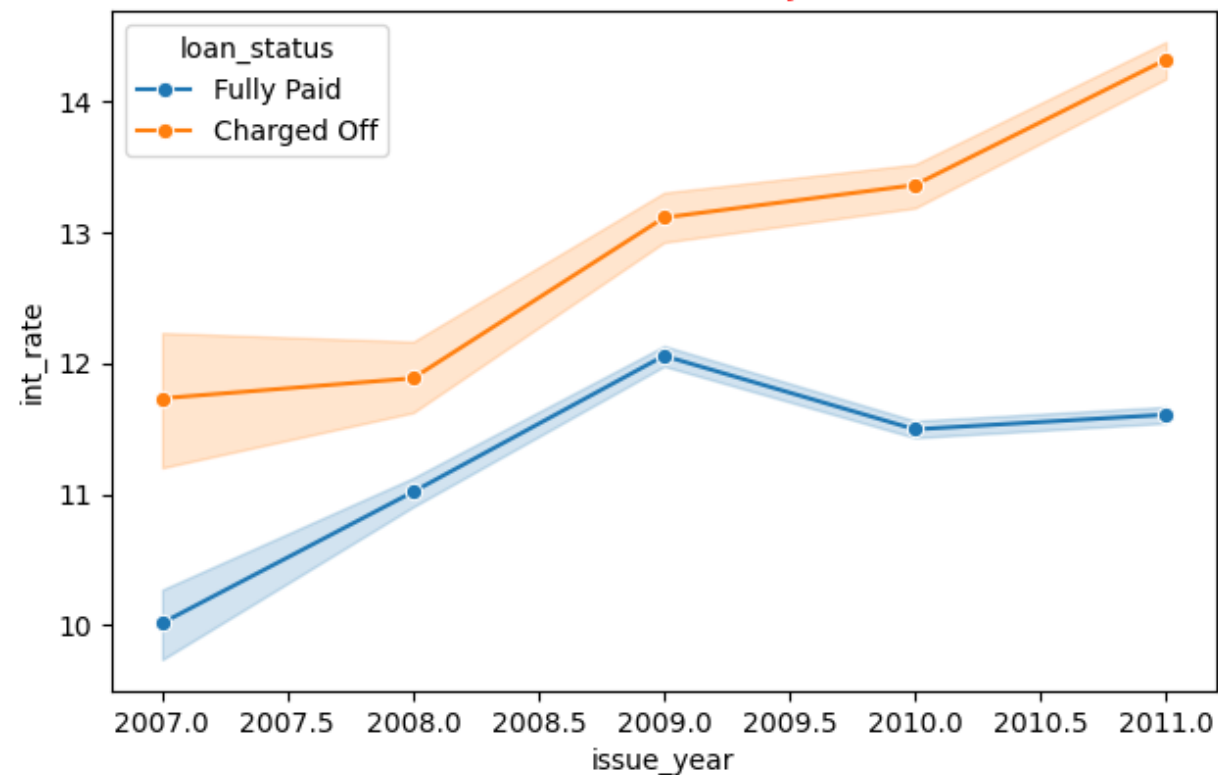


1. **Interest Rate vs. Purpose:** Interest rates for "Charged Off" loans are high when compared to "Fully Paid" For House and Small Business the rates are exceptionally high for "Charged Off" loans.
2. **Loan Amount vs. Purpose:** Loan amounts tend to be higher for loans that were "Charged Off" compared to "Fully Paid"
3. **Annual Income vs. Purpose:** Annual income levels tend to be higher for "Fully Paid" loans compared to "Charged Off" loans for most purposes.
4. **DTI vs. Purpose:** DTI ratios for "Charged Off" loans are generally higher compared to "Fully Paid" loans across most loan purposes.
5. **Summary:** Higher interest rates, larger loan amounts for purposes like "small\_business" and "home\_improvement", higher DTI ratios are more likely to be charged off.

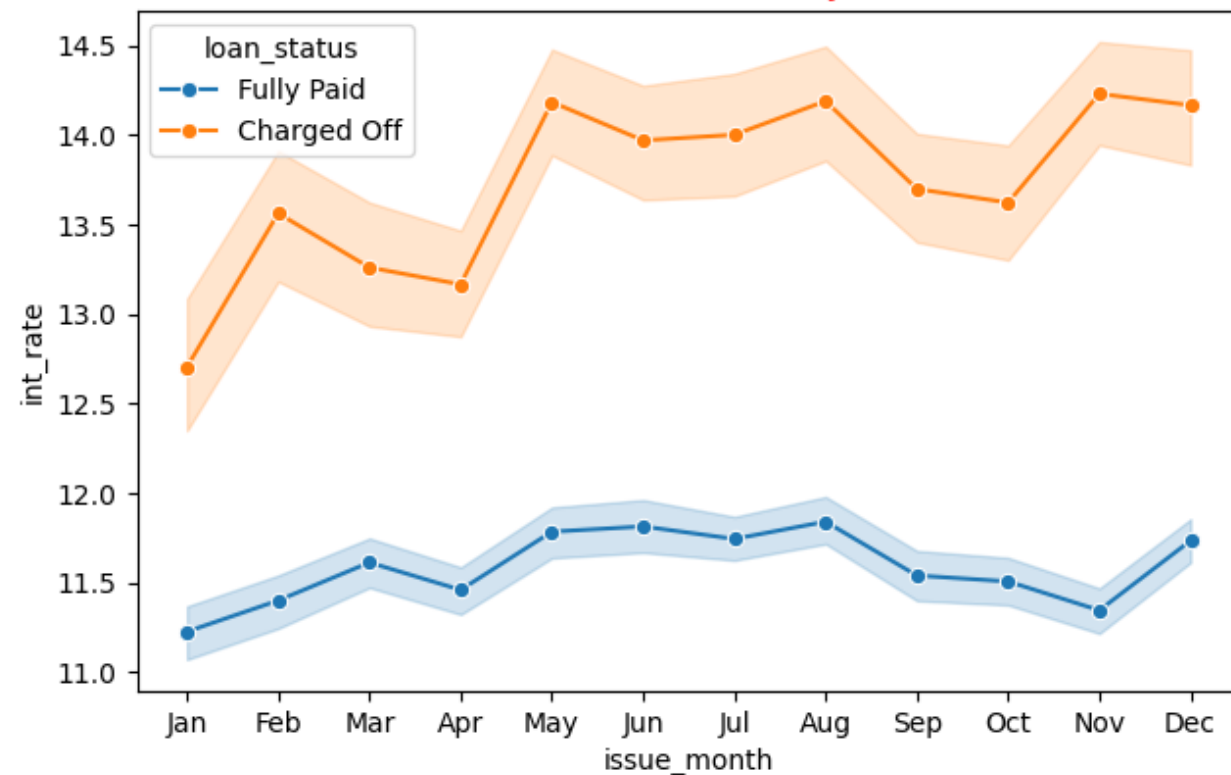


# Interest Rate Vs Issue Year and Interest Rate Vs Issue Month By Loan Status

## Interest Rate vs Issue Year by Loan Status



## Interest Rate vs Issue Month by Loan Status



### Interest Rate vs Issue Year


1. Interest rates for "Charged Off" loans increase significantly over the years, peaking around 2011.
2. Interest rates for "Fully Paid" loans also increase over the years but at a slower rate

### Interest Rate vs Issue Month

1. Interest rates for "Charged Off" loans are generally higher across all months, with noticeable peaks in May, Aug and November.
2. Interest rates for "Fully Paid" loans show less variability and generally lower values across the months

# Key Findings

---

1. **Interest rate** is a strong predictor of loan default risk. Higher interest rates are associated with significantly increased charge-off rates.
  2. **Applicants income** is a critical factor to do Loan payment. Lower-income individuals exhibit higher default rates.
  3. **Loan grade** is a significant determinant of loan risk. Lower-grade loans (D, E, F, G) have a substantially higher defaults compared to higher-grade loans (A, B).
  4. **Debt-to-income (DTI) ratio** is positively correlated with loan default risk. Higher DTI ratios indicate increased chances of charge-off.
  5. **Bankruptcy history** is a strong indicator of elevated loan default risk.
- 

# Observations

---

1. Loan Applicants with shorter employment lengths having higher interest rates are more likely to default on their loans.
  2. Economic conditions can influence loan performance. In year 2011, we saw an increase in both loan volume and default rates
  3. Loan purpose impacts loan performance. Loans for small businesses tend to have higher interest rates and default rates.
  4. Loan grade is a strong indicator of loan performance. Higher-grade loans have lower default rates.
  5. Income level is associated with loan performance. Lower-income Applicants have a higher default compared to higher-income borrowers.
  6. Bankruptcy history is likely correlated with loan default, as indicated by the relationship between grade and charge-off rates
- 