# Lending Club Case Study

Deepak Kaithwas
Reena Upadhyay

# Business Understanding

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

# Objective

Company wants to **understand the driving factors behind loan default** (loan_status = 'Charged Off' in the loan data set), i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.
**The business objective is to take a decision whenever they receive a loan application whether to reject or approve based on certain variables.**
As part of this case study, implement EDA technique on a real world problem and understand the insights and present in a business first manner via visualization.
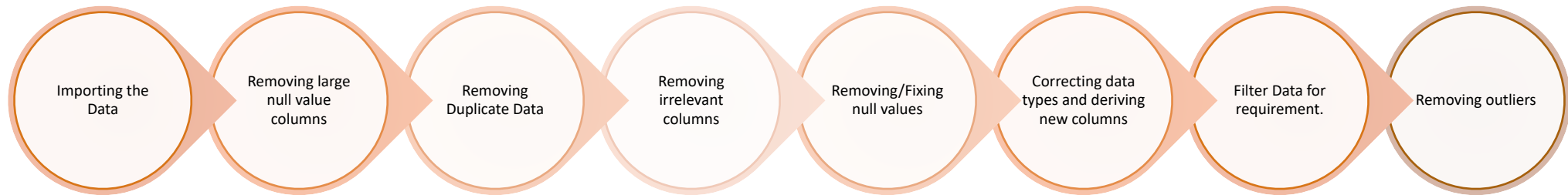
**Benefits of the case study:**
- Gives a idea about how EDA is used in real life business problems.
- It also develops a basic understanding of risk analytics in banking and financial services.
- How the data is used to minimize loss of money while lending it to clients.
- It improves our understating of visualization and what charts to use for real life data.

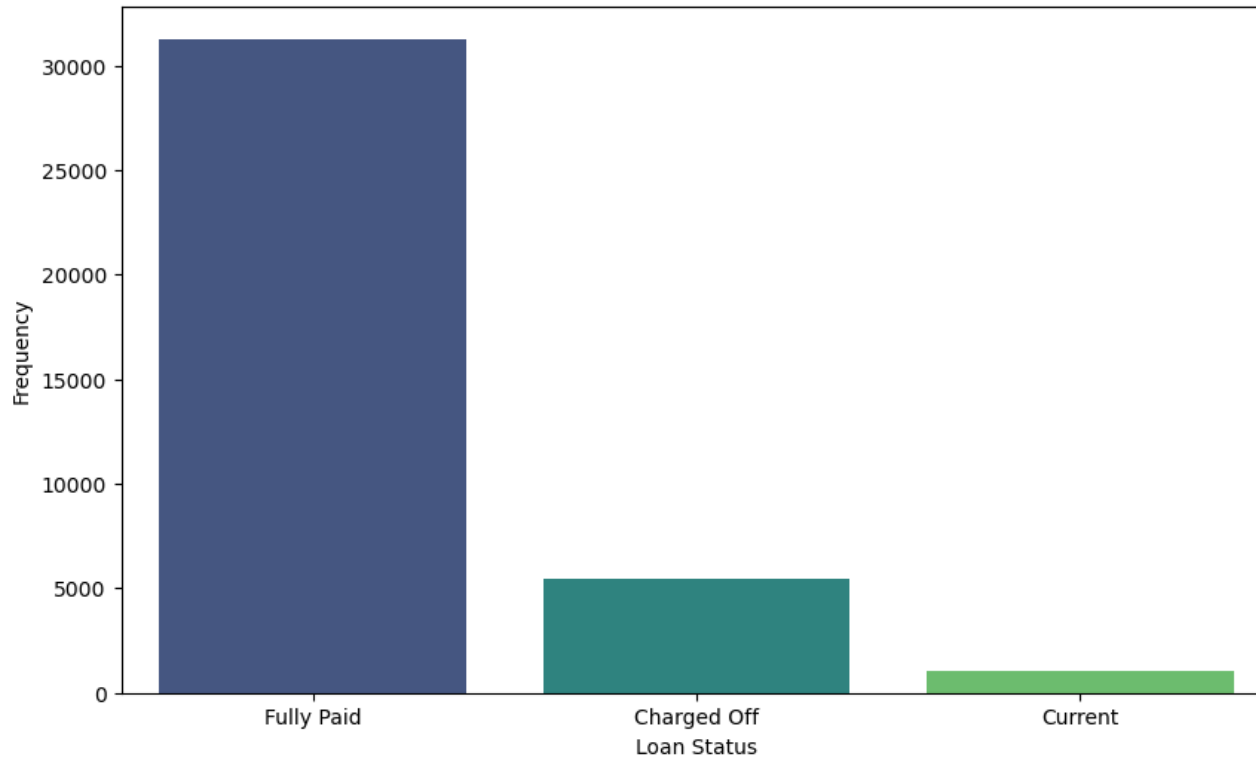# Data Understanding & cleaning

The loan data set contains information about past loan applicants and whether they 'defaulted' or not. Data has details regarding approved loan not the rejected ones. It has 3 status of loan which is Fully Paid, Current and Charged-Off.

**Data Clean-up and preparation process:**

Importing the Data → Removing large null value columns → Removing Duplicate Data → Removing irrelevant columns → Removing/Fixing null values → Correcting data types and deriving new columns → Filter Data for requirement. → Removing outliers

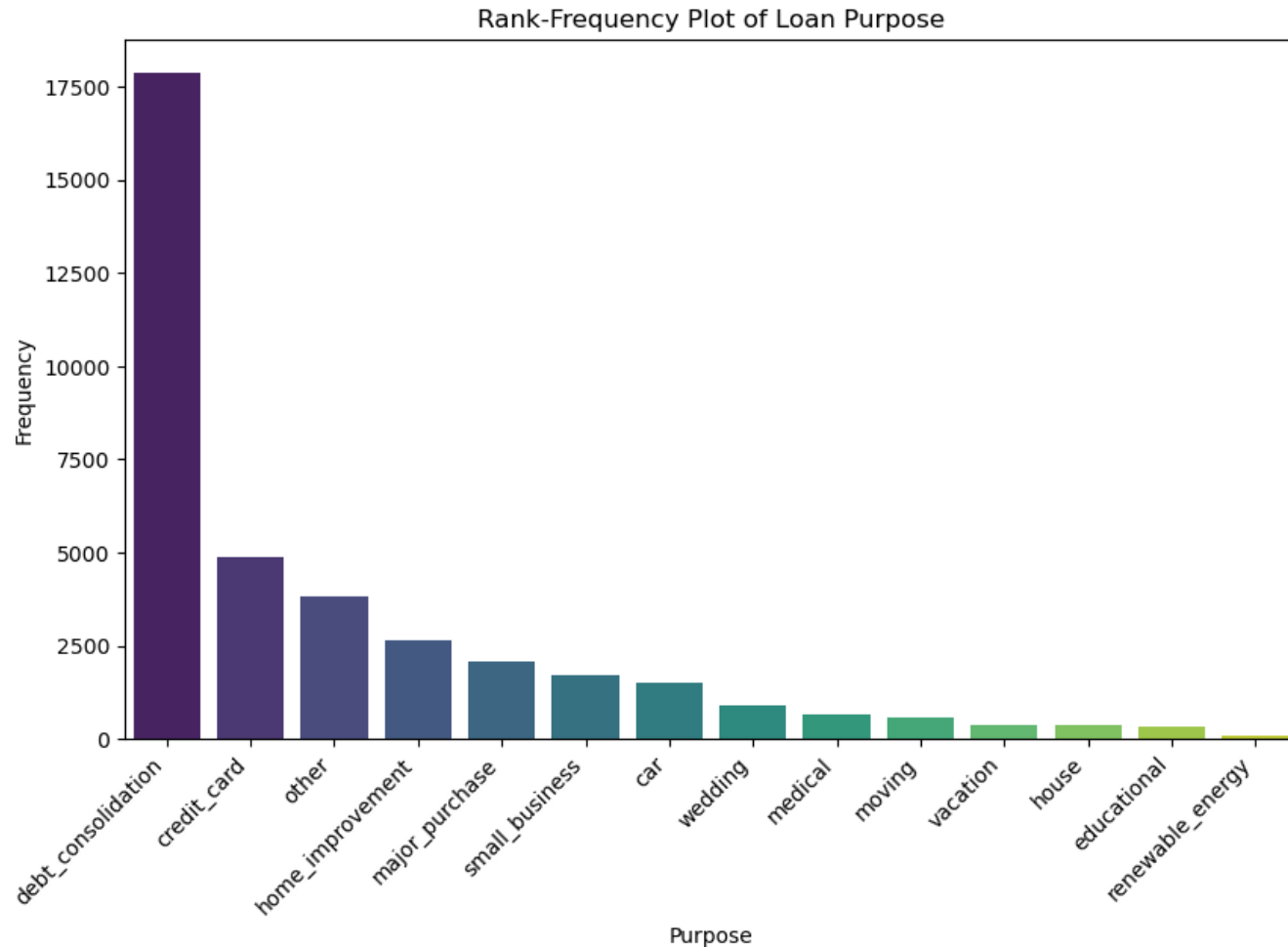# Univariate Analysis - Loan Status



Rank-Frequency Plot of Loan Status

The number of charged off loan is much smaller (14%) compared to total count.
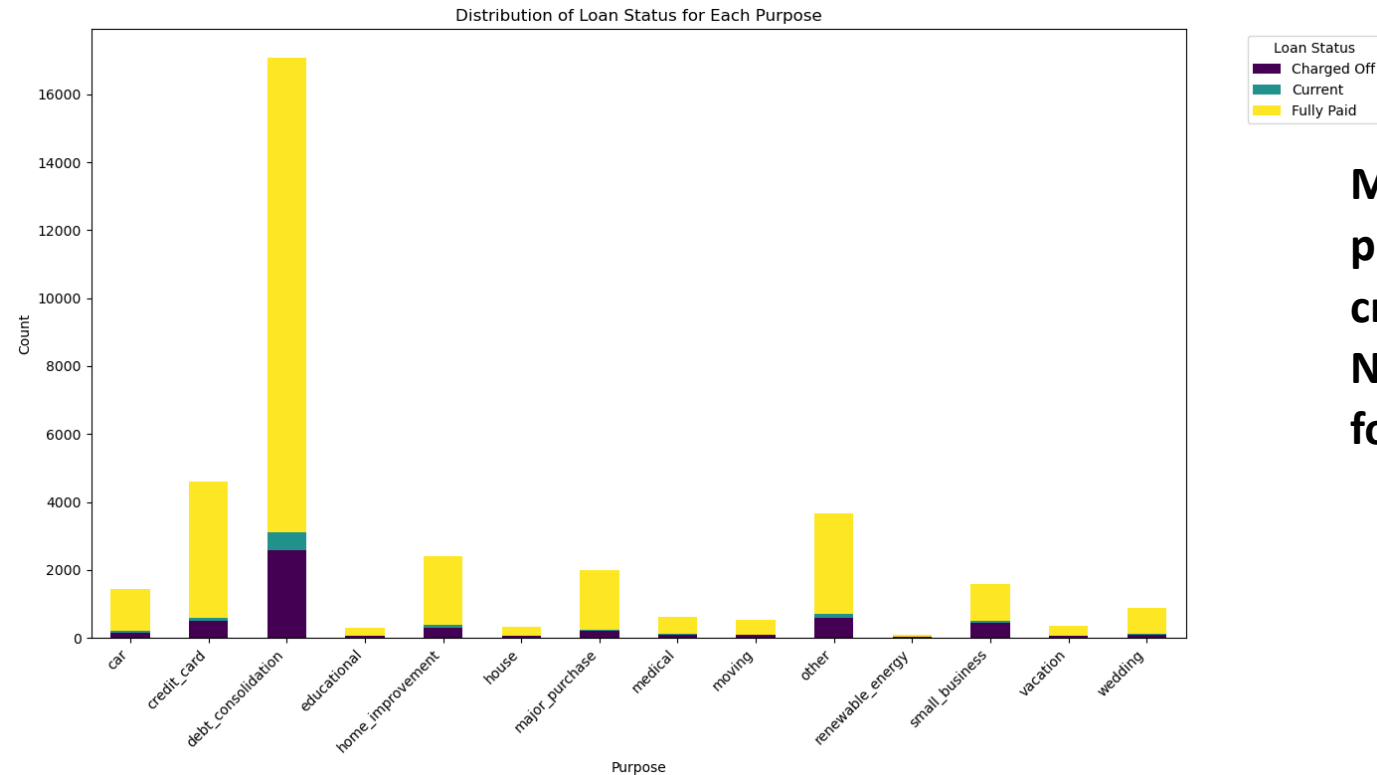
# Univariate Analysis - Loan Purpose



Rank-Frequency Plot of Loan Purpose

**Most of the loans were taken for the purpose of debt consolidation & paying credit card bill.**

# Univariate Analysis - Home Ownership



Rank-Frequency Plot of Home Ownership
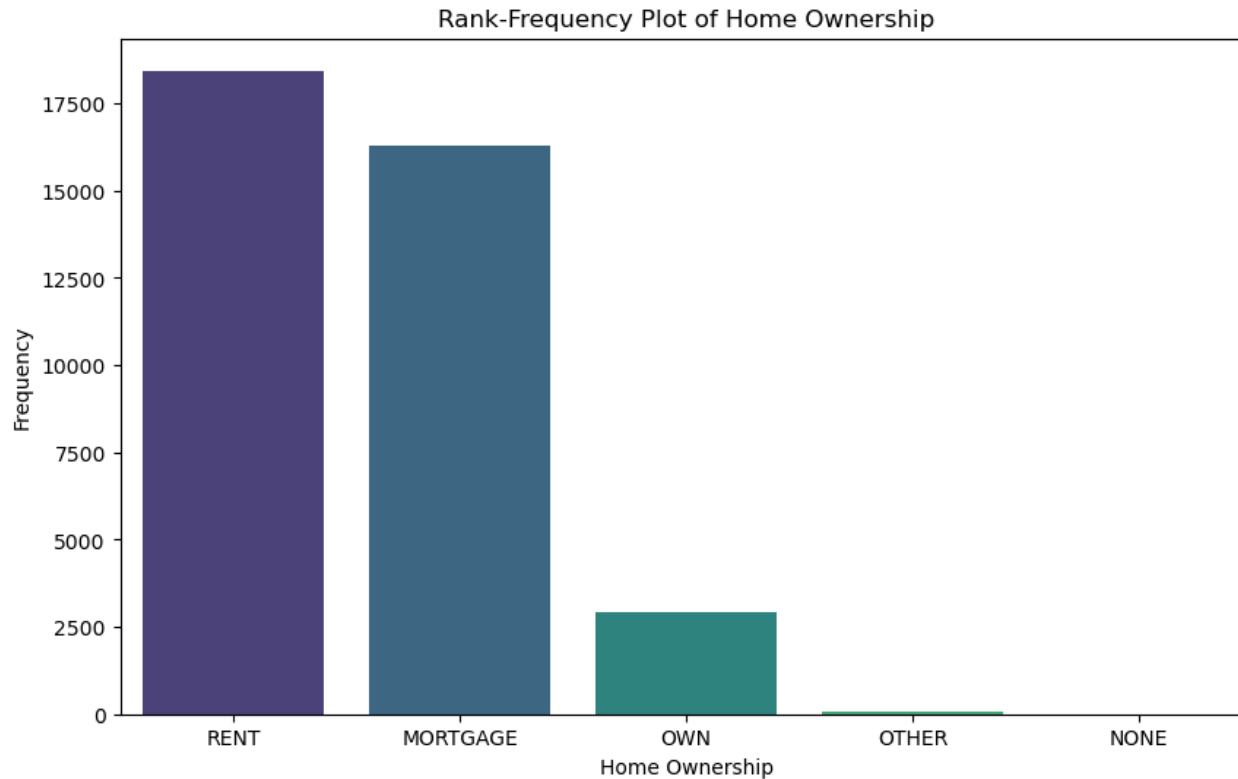
**Most of the loan applicants are from people on rent or Mortgage.**
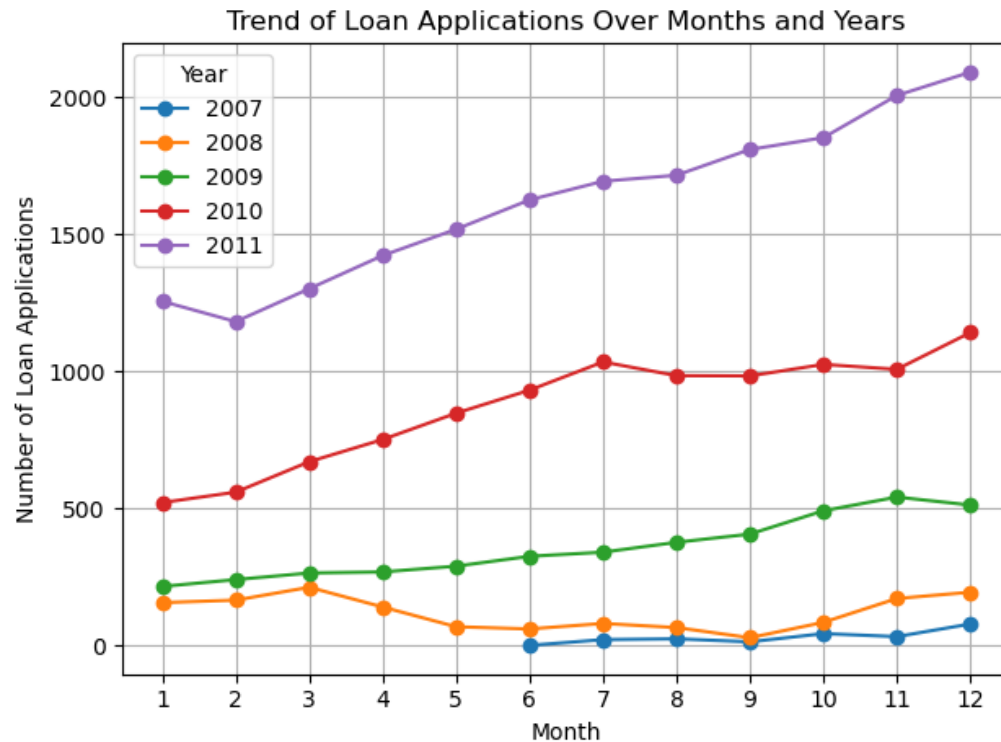
# Distribution of loan status for each purpose



Distribution of Loan Status for Each Purpose

**Most of the loans were taken for the purpose of debt consolidation & paying credit card bill.**
**Number of charged off count also high too for these loans.**

# Univariate Analysis - Ordered categorical variables



Count of loan application is increasing every passing year, so increase in number of loan applications are adding more to number of charged off applications. Number of loans issued in 2008( May-October) got dipped, may be due to Recession.

# Univariate Analysis - Numerical variables
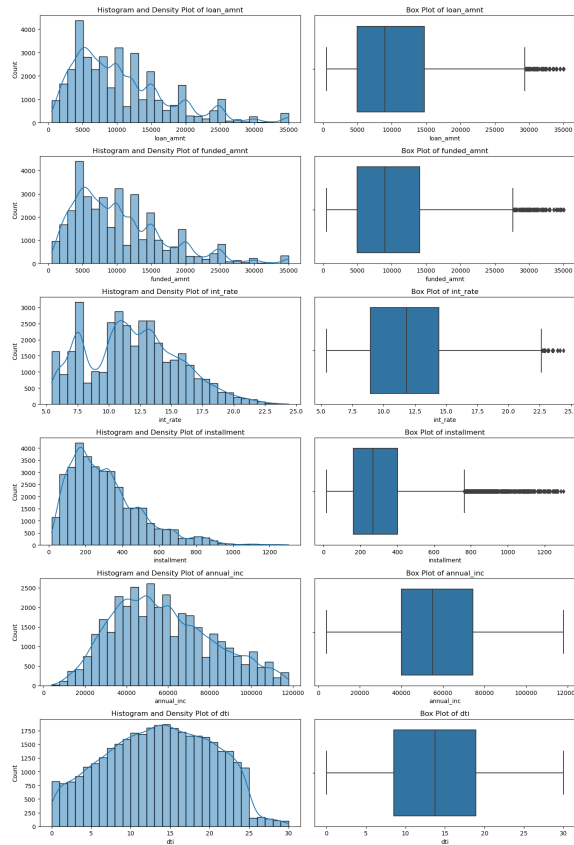
# Bivariate Analysis



Correlation Matrix
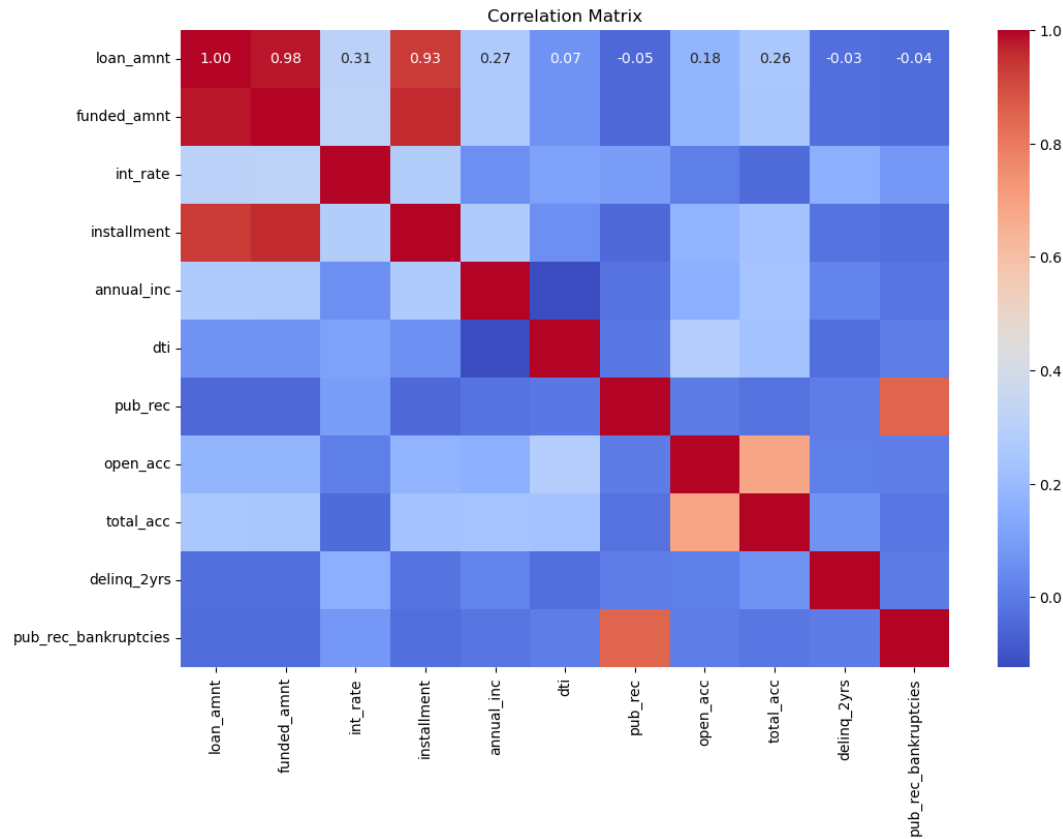
we can clearly see loan_amnt, funded_amnt and installment have huge correlation. These fields are proportional to each other. Next, the public records related fields pub_rec & pub_rec_bankrupcies and number of accounts related fields open_acc & total_acc are correlated.

# Emp length vs Loan status



emp_length vs loan_status
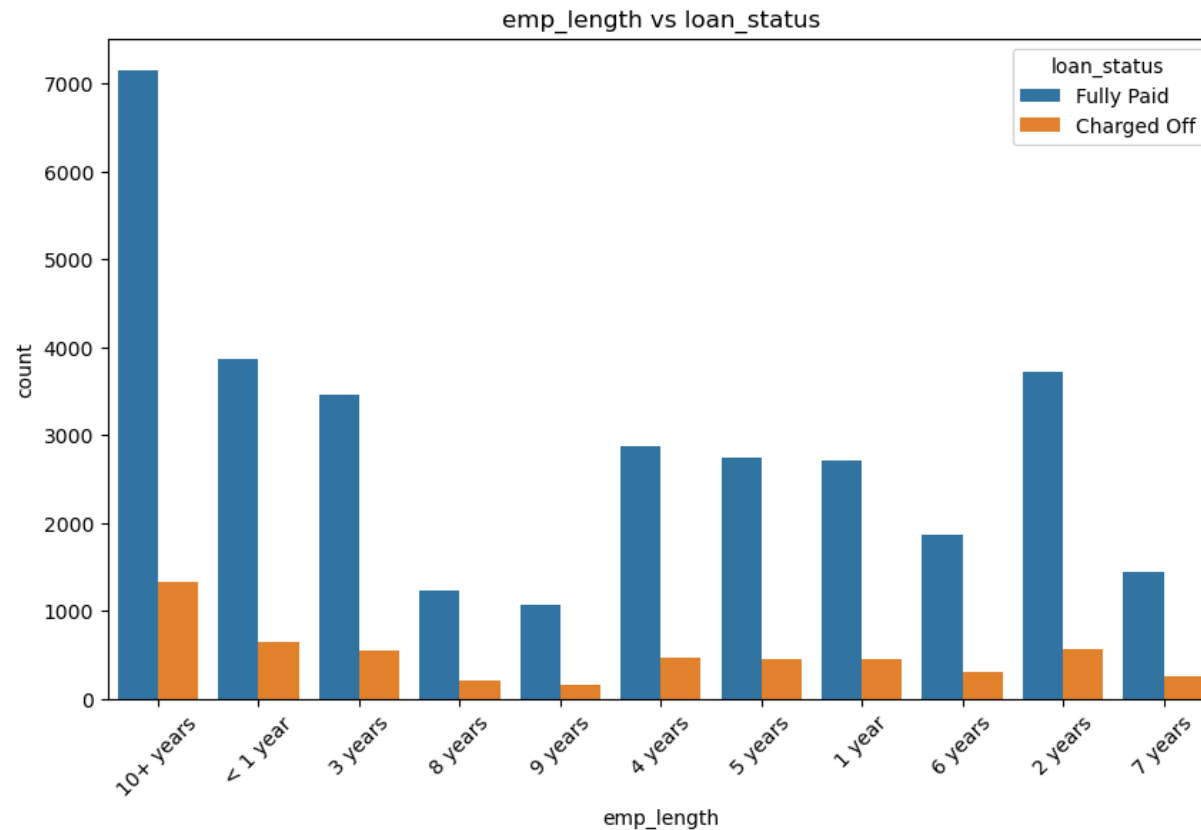
We can see higher default rate among borrowers with shorter employment lengths (e.g., <1 year).
Borrowers with more than 10 years of employment tend to have lower default rates.

# Grade vs Loan status



grade vs loan_status
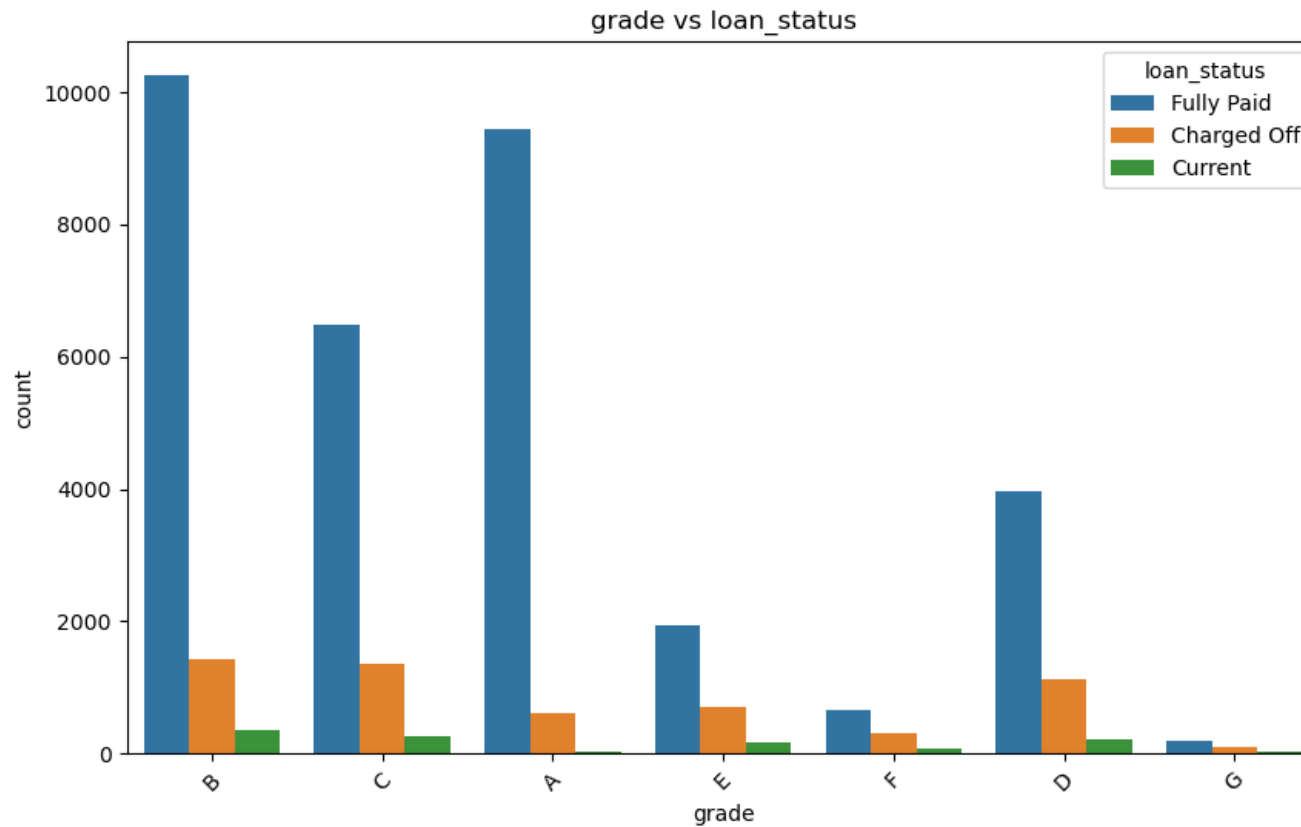
**Higher grade loans (e.g., grade 'A' and 'B') tend to have fewer defaults compared to lower grade loans (e.g., grade 'D', 'E', 'F', and 'G').**

**#We can say credit rating assigned to loan is significant predictor of default risk.**

# Recommendations

Major Driving factor which can be used to predict the chance of defaulting and avoiding Credit Loss:
- DTI
- Grades
- Verification Status
- Annual income
- Pub_rec_bankruptcies

Other considerations for 'defaults' :
- Burrowers having annual income in the range 50000-100000.
- Burrowers having Public Recorded Bankruptcy.
- Burrowers with least grades like E,F,G which indicates high risk.
- Burrowers with very high Debt to Income value.
- Burrowers with working experience 10+ years.