Lending Club Case Study

Group members:

Deepak Kurup

C. Jagadeeswara Reddy

Abstract

Background:

➢Lending Club is the largest online loan marketplace, facilitating personal loans, business loans and financing of medical procedures. It's largest source of credit loss is lending loans to 'risky' applicants.

Problem Statement:

➢Lending Club wants to identify the risky loan applicants, so that such loans can be reduced thereby cutting down the amount of credit loss.

Our objective is to find the driving factors which could lead to loans getting default.

Problem Solving Methodology

| Data Understanding | → | Data cleaning | → | Univariate Analysis | → | Segmented Univariate Analysis | → | Bivariate Analysis | → | Recommendations |

Data Cleaning

Checkpoint 1

1. Checking the null values in columns.

2. Calculating the percentage of null values.

3. Drop all the columns whose null % is >30

4. We can further clean some more columns by dropping those columns which have only one unique value.

Checkpoint 2

1. Loan status has 3 distinct values - fully paid, charged off and current. So, we remove the rows which are current.

2. Type conversion a attributes or variables into particular data type such as removing percentage and convert it into float, convert to date time format.
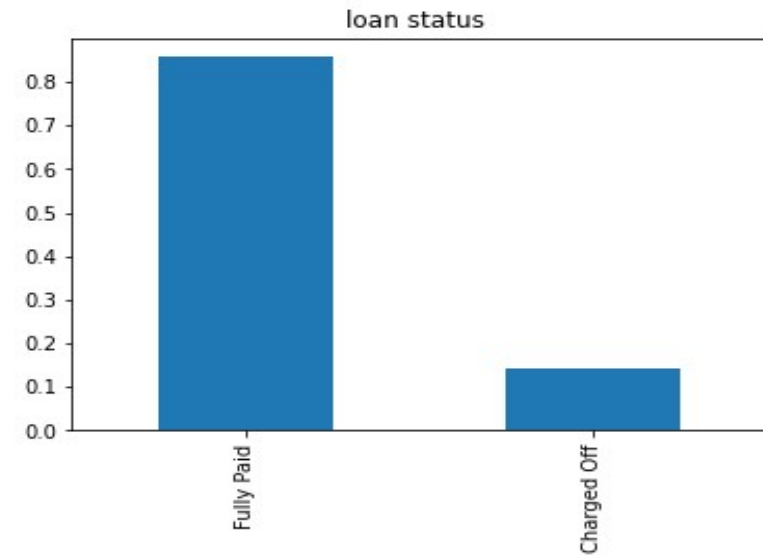
Checkpoint 3

1. Create some additional derived columns to enhance our analysis on the loan information .

2. Assigning numeric values to the category column 'Fully Paid' and 'Charged'.

3. Derived columns for year and month for issue_dt and earliest credit line .
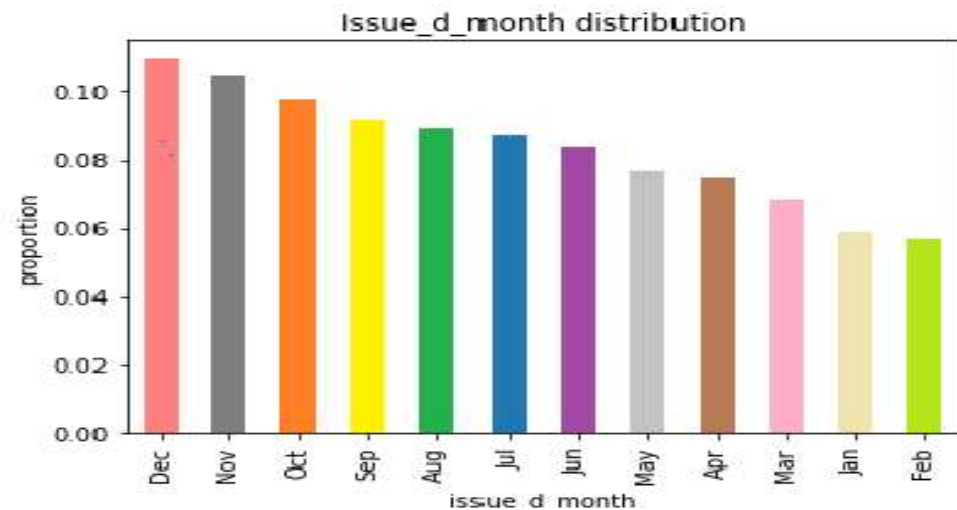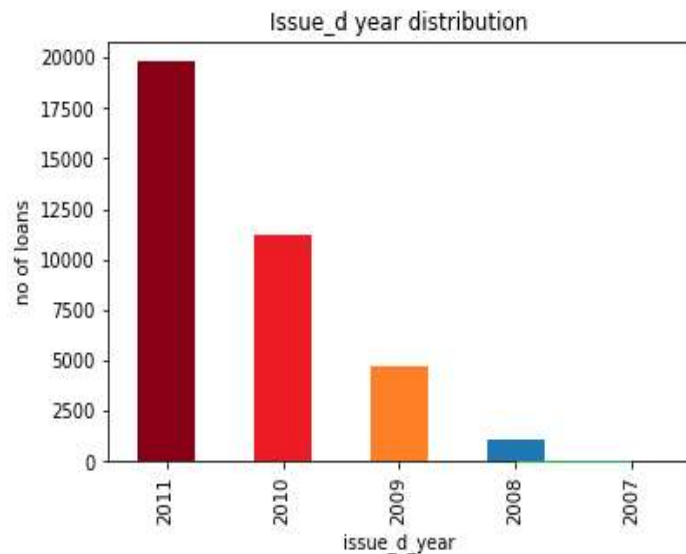
Analysis

Term distribution
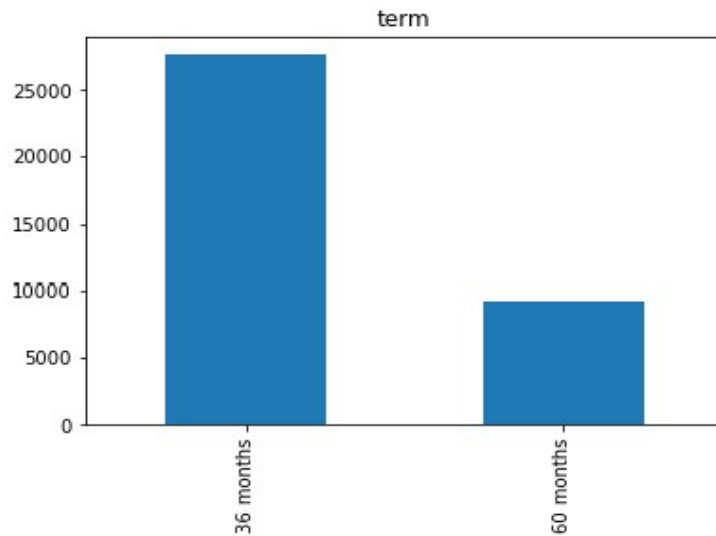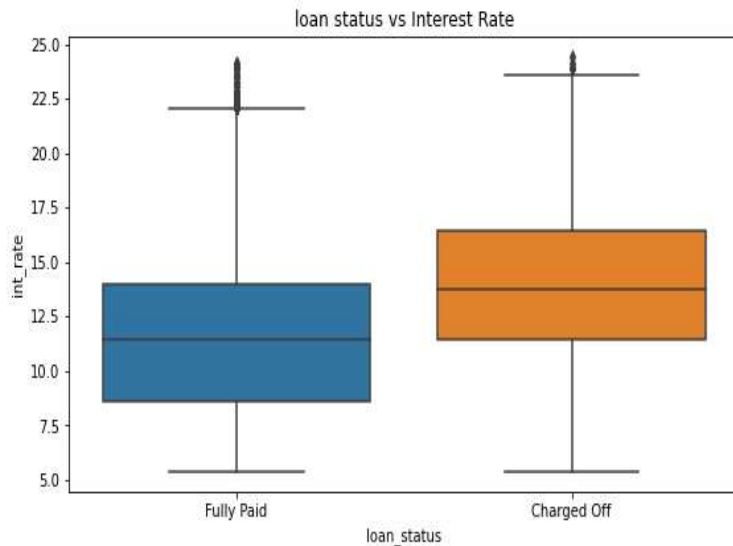
term

Loan status distribution

loan status

- There are only two loan terms 36 and 60 months. Most of the borrowers took loans with 36 months term.
- The charged off borrowers are around 15% and fully paid is around 85% in the given Loan status distribution.

Analysis

Issue_d year distribution

Issue_d_month distribution

- No. of loans 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.
- The month-wise trend shows that applicants apply more loans towards the end of year.

Analysis

loan status vs Interest Rate

- When the loan interest rate is high, there is high chance of loan getting defaulted.
- High interest rates tend to increase the monthly installments against there total annual income.

Analysis

Term vs loan status

Term vs Interest Rate

term vs loan amount
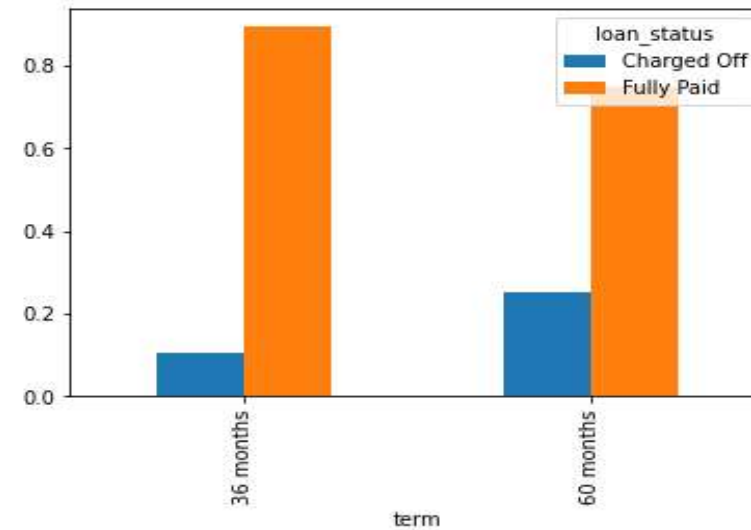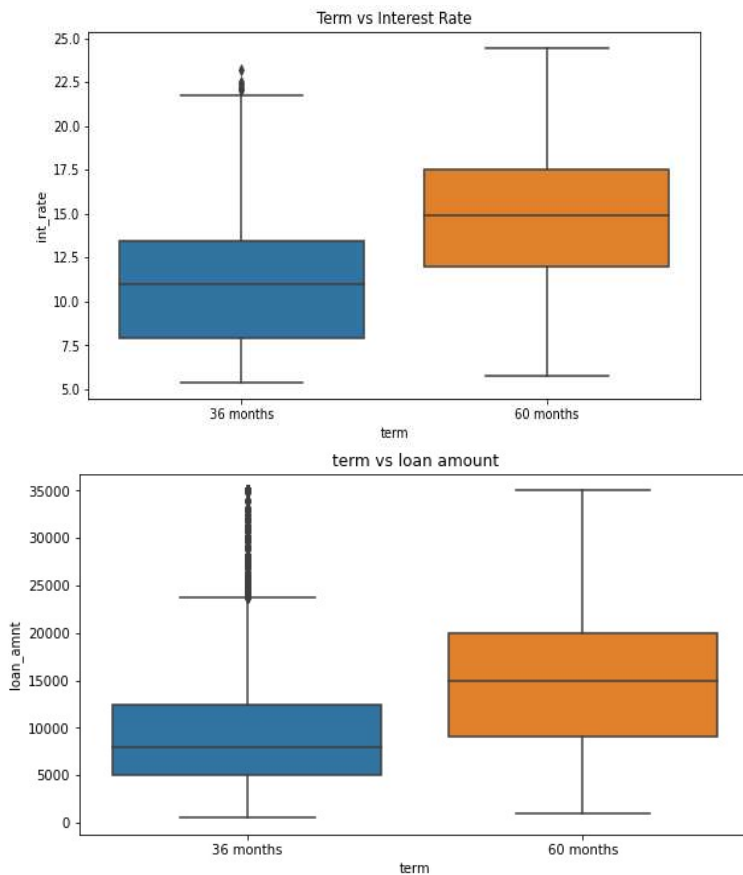
The default rate is high in 60 months tenure because most people took high loan amount with high interest rate in it and they may face difficulties in returning the sum to bank.
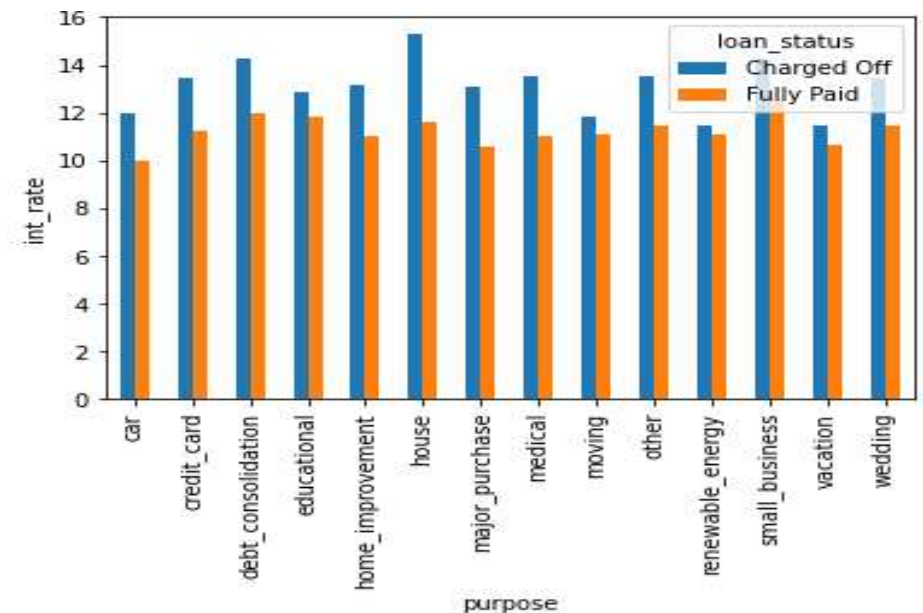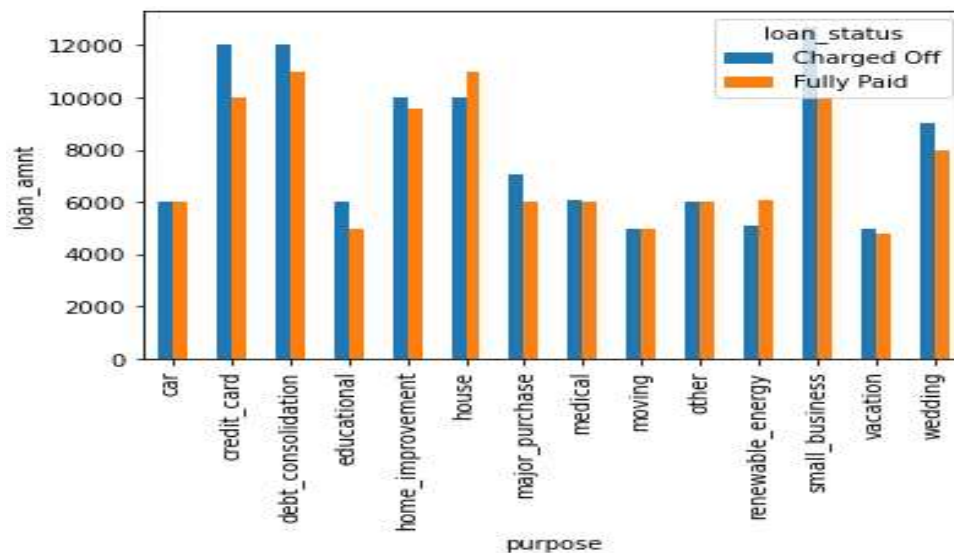
Analysis

grade vs loan amount

Grade vs loan status
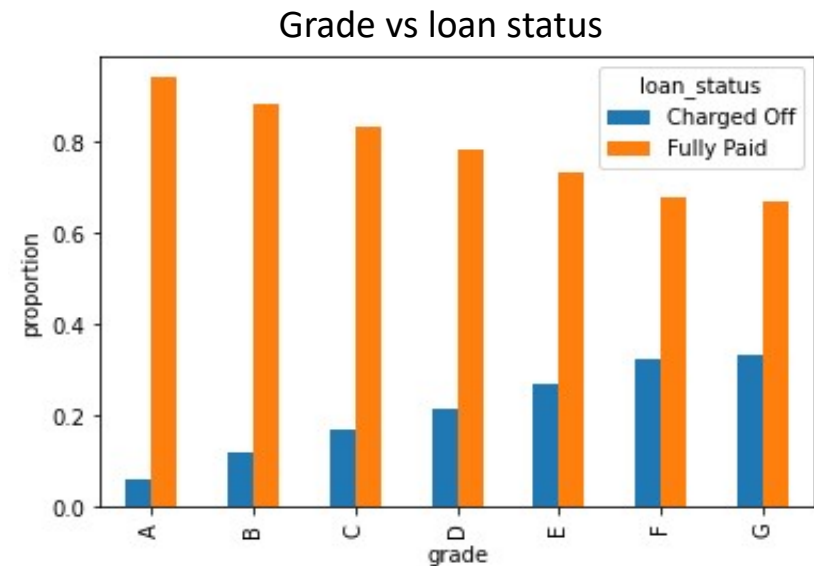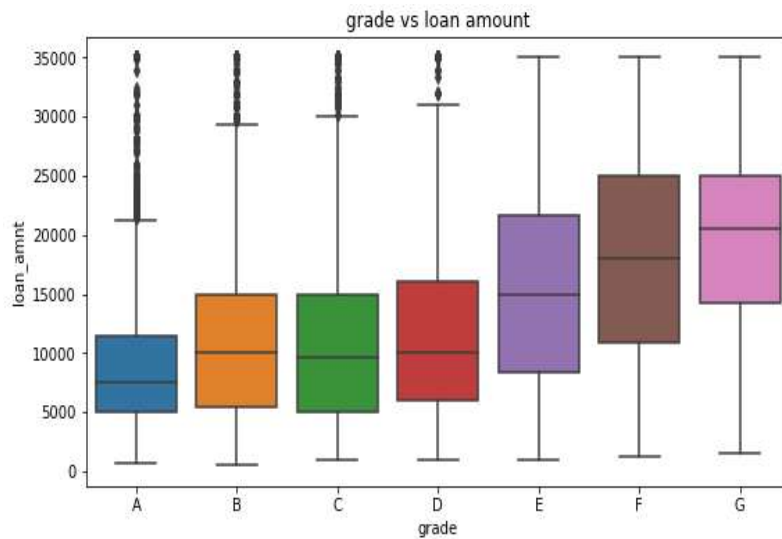
- Grades are very good category to tell the borrower probability of defaulting the loan.
- The Lower grades(E,F,G) have higher chances of defaulting the loan than Higher ones(A,B)
- Also the Lower grades are borrowing more loan amount which might be the cause for loan default.
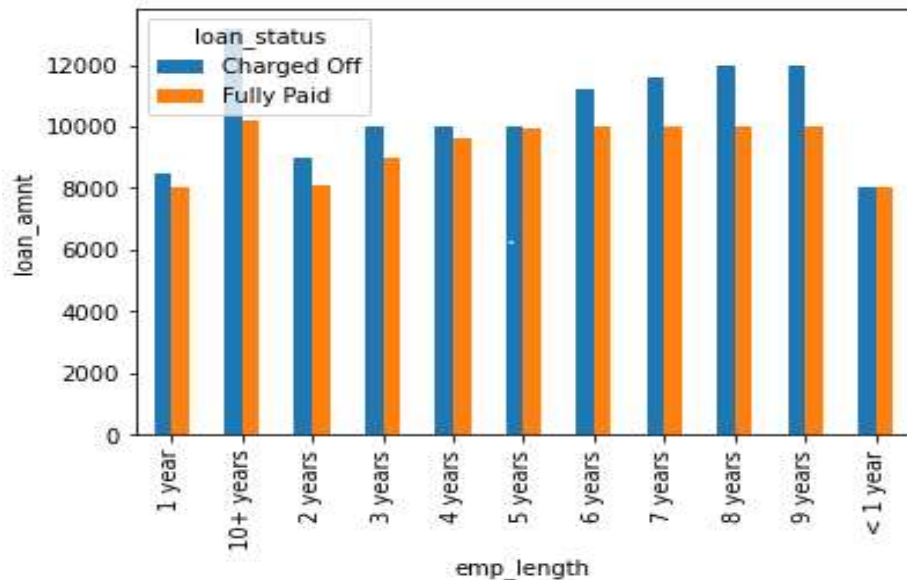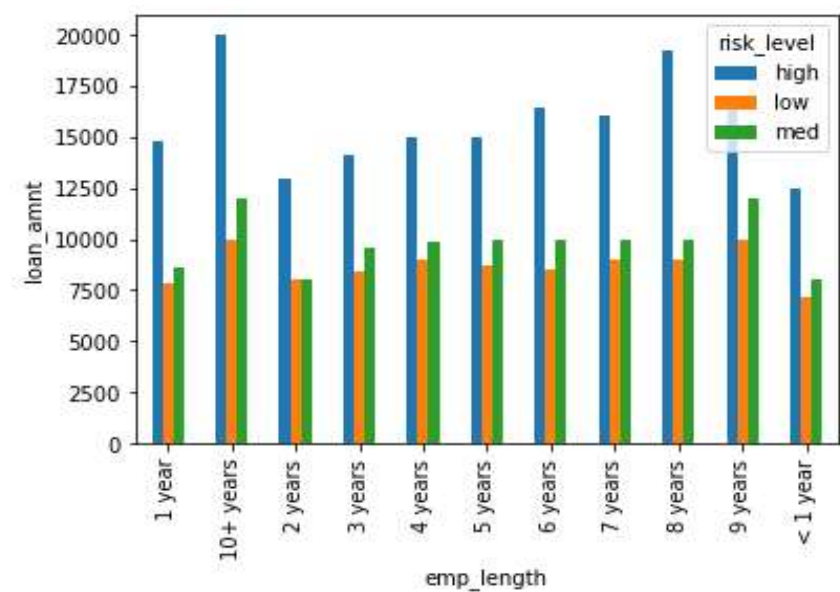
Analysis

Purpose vs loan status

- Some observations based on the above plot to identify high risk candidates
    - Borrower's who took loans for small business purpose
    - Home loans with high interest rate
    - People who borrowed for small business and debt consolidation.

Analysis

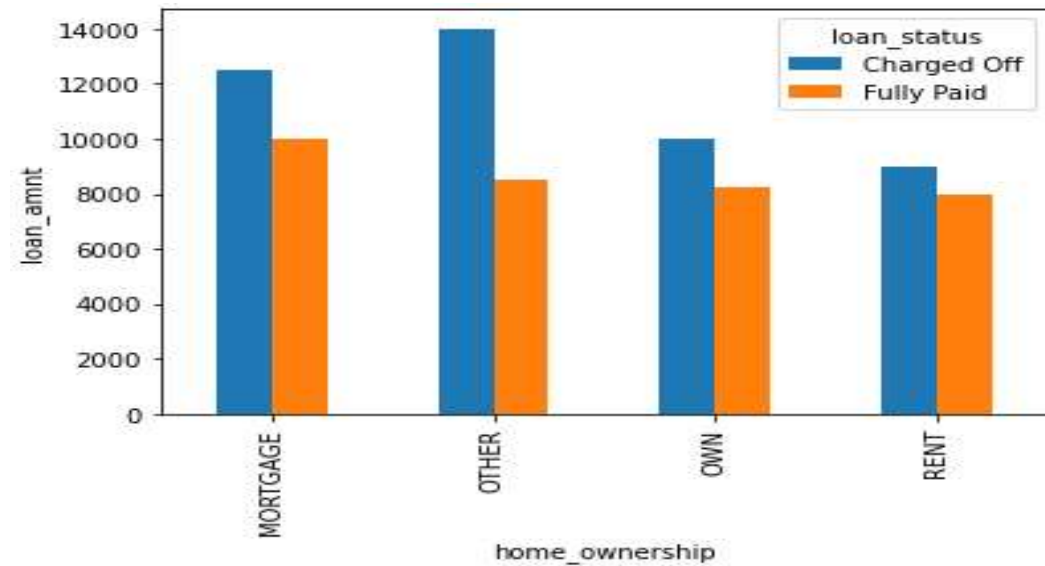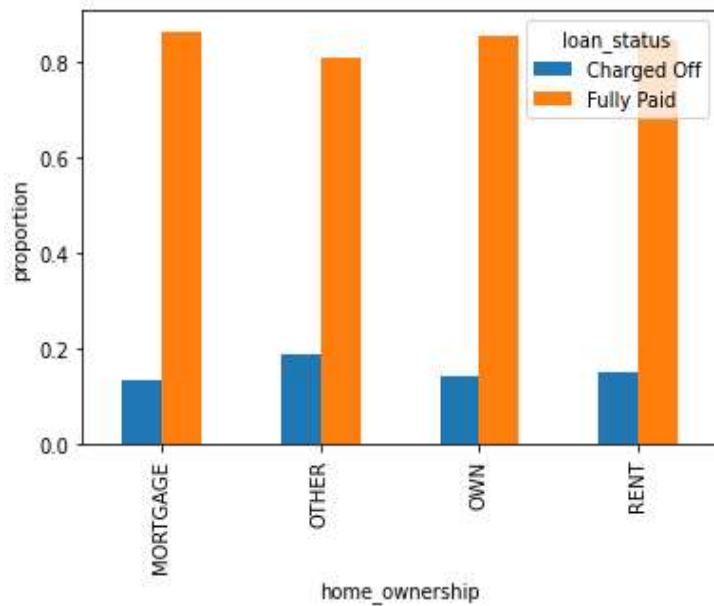**Employment length vs loan status**

Employment length vs risk level

- Borrowers with higher experience having high loan amounts are in high charged off ratio.
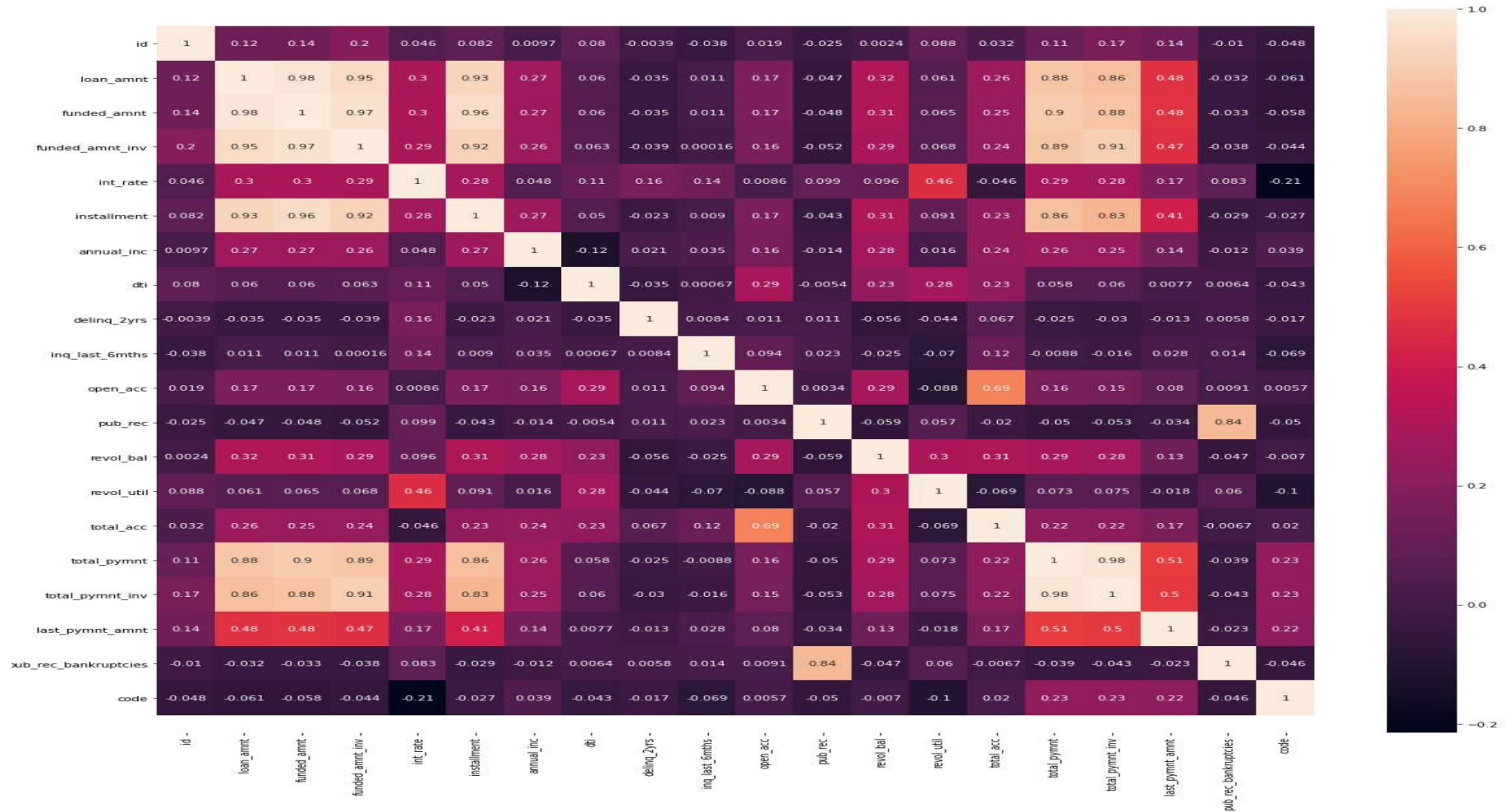- The risk level is high for people with more experience

Analysis

Home ownership vs loan status

- There is bit high proportion of charged off people are recorded in other -home ownership category.
- Borrowers who took higher loan amounts are charged off more than others.

Analysis of Correlation Heat-Map

Observations from heat map

- Loan Amount, Funded Amount, Funded Amount Invested and Installment are highly correlated with each other. Most of the borrowers were approved more than 90% of the loan amount.

- The public derogatory records column is highly (+ve) correlated with public bankruptcies records.   This is a good indicator that majority of the borrowers have no public records or filed for bankruptcy.

Conclusions

- Lending club should reduce the high interest loans for 60 months tenure, they are prone to loan default.
- Grades are good metric for detecting defaulters.
- Lending club should examine more information from borrowers before issuing loans to borrowers with low grade score(G to A).
- People who are not verified based on the verification status and having a loan are risky so LC should improvise on the verification process or reduce loan approvals.
- When the loan interest rate is high, there is high chance of loan getting defaulted.
- Borrower's who took loans for small business purpose have defaulted more. LC should consider these applications more carefully.
- People with mortgages are paying high interest rates and can be risky.
- Borrowers with high experience have more debts so they can be risky.