LENDING CLUB CASE STUDY

MLC61 AI&ML

ROHINI KANNAPIRAN
BHAGAVANTRAO PATIL

CONTENTS

PROBLEM STATEMENT

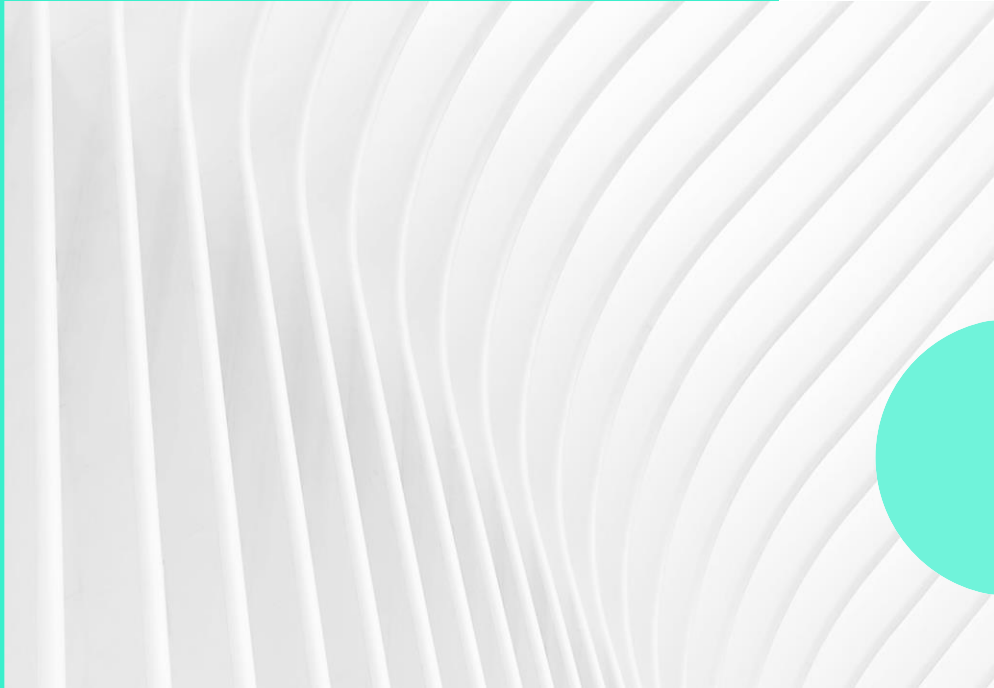Lending club is looking for analysis on the given historical loan dataset to identify potential risks that influences defaulting behavior.

DATA UNDERSTANDING

Exploring the dataset manually gives basic understanding of data as follows :

1. Dataset with 111 columns and historical data from 2007-2011

2. Data not limited to Fully paid or defaulted customers, but also Current customers

3. Data dictionary attached for more business understanding of columns.

DATA CLEANING

- Removing not so impactful geographical attributes.

- Removing columns with most NULL values and single unique values.

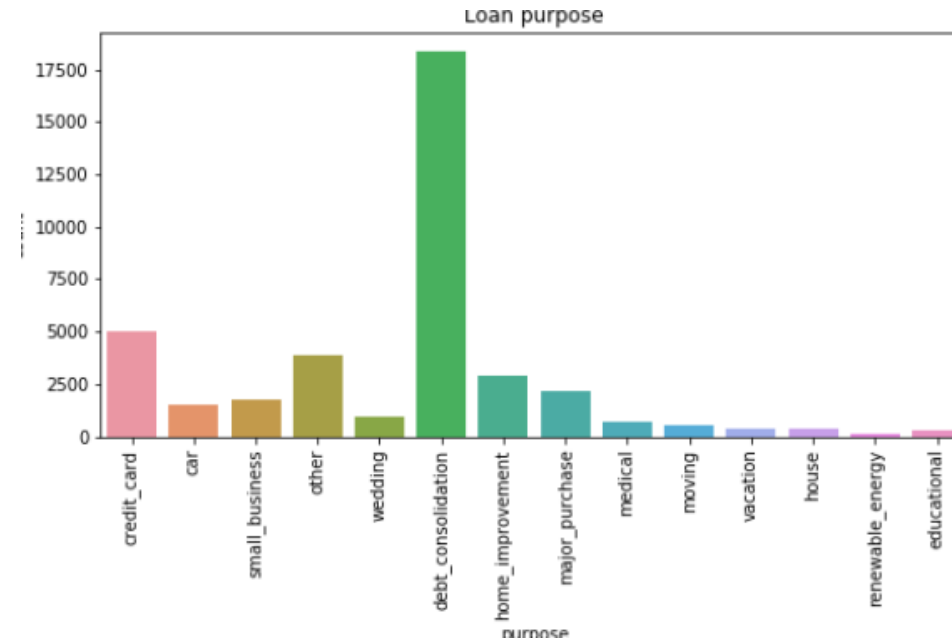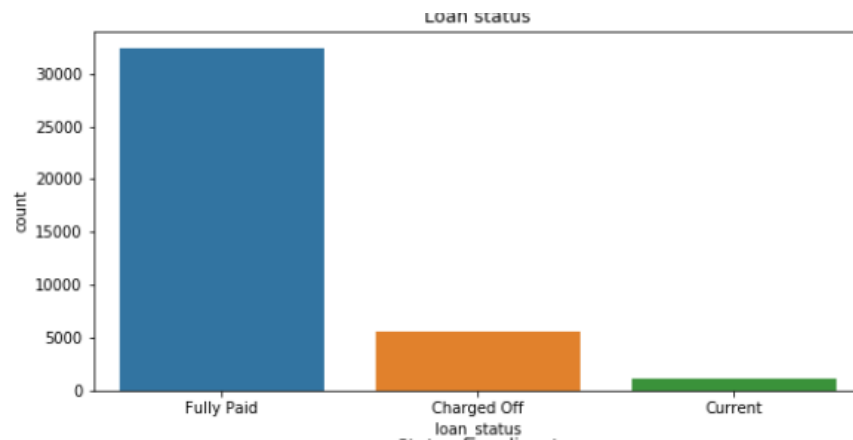- Removing irrelevant columns for the target context.

UNIVARIATE ANALYSIS

- Almost most of the applicants are funded with the requested amount
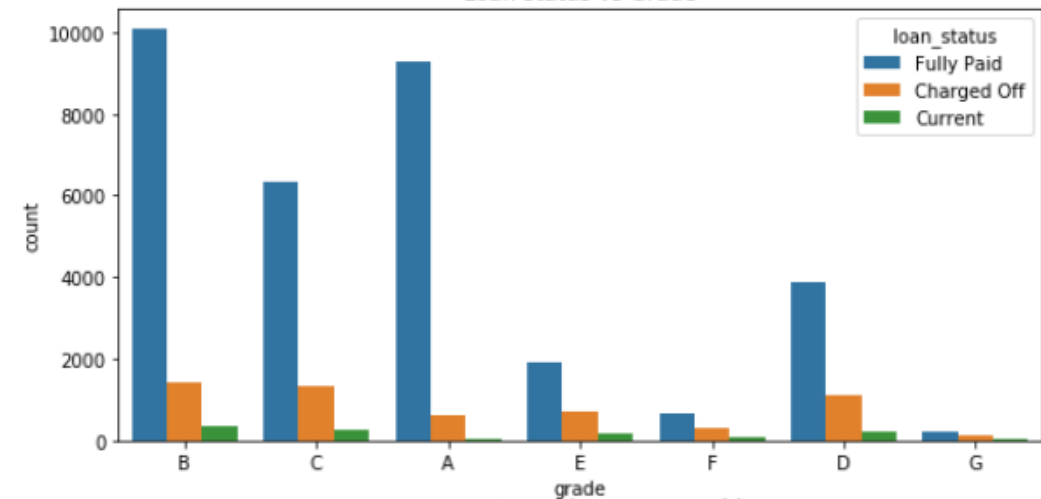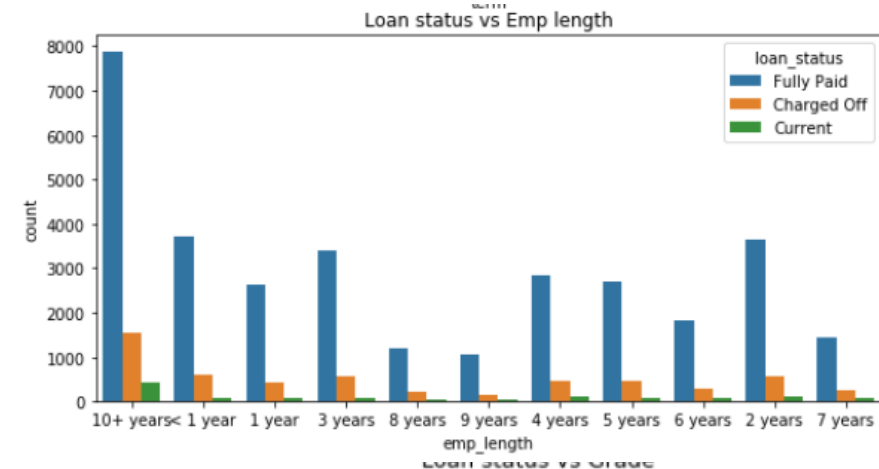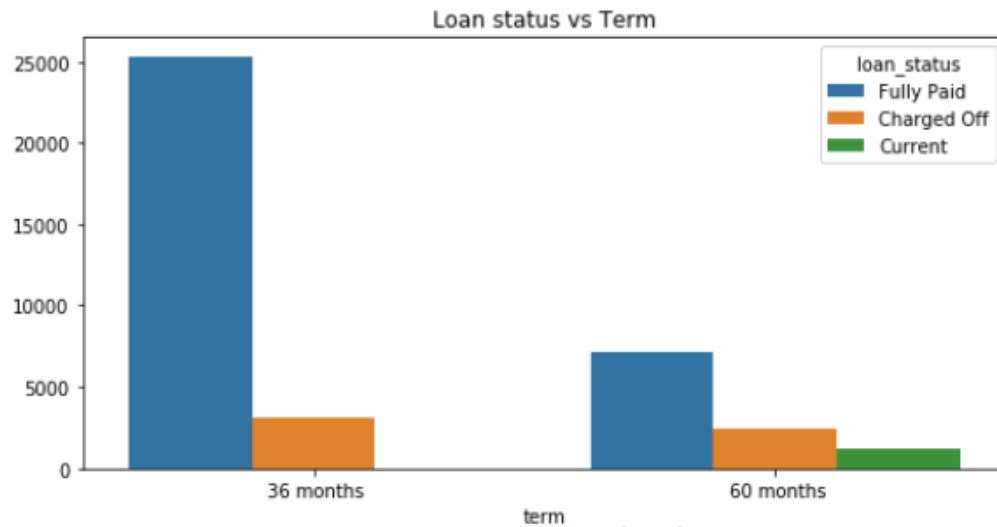- People preferred a shorter term compared to longer ones.
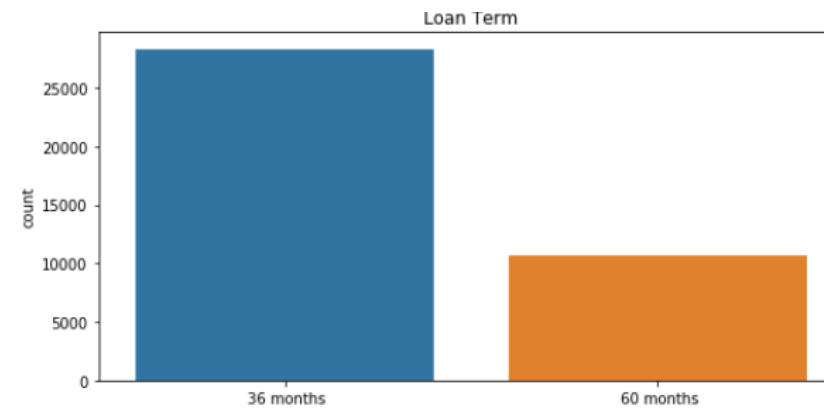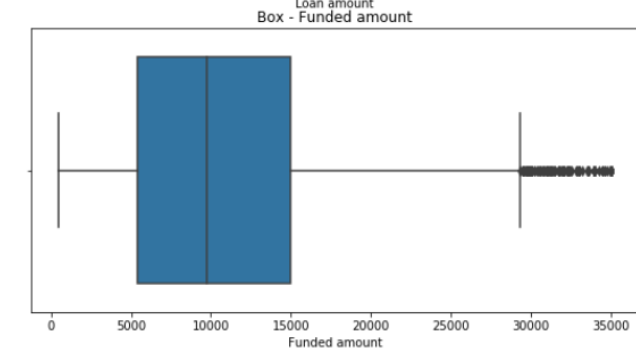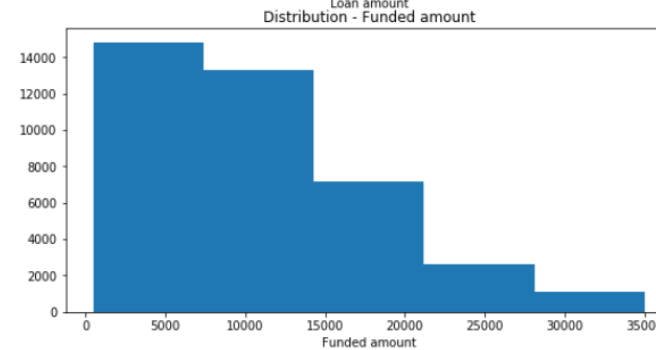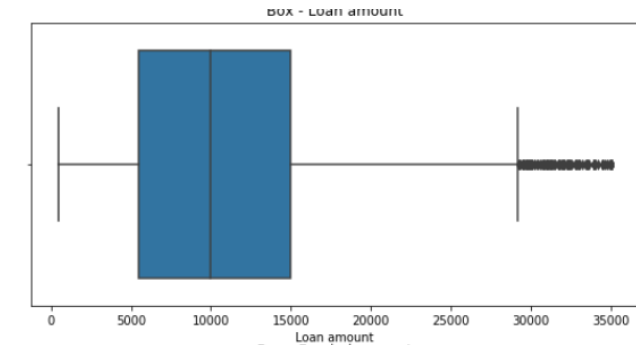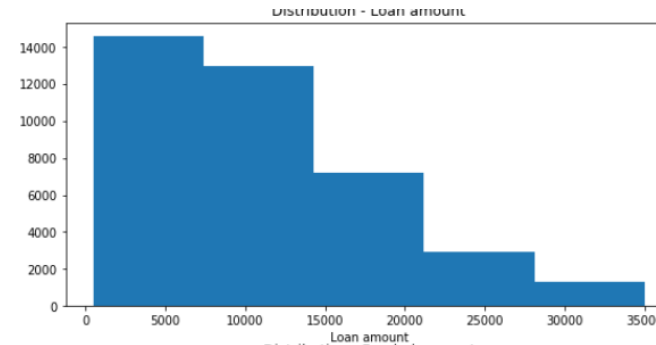
UNIVARIATE ANALYSIS

- Overall charged off proportion is less

- Majority of people wanted to consolidate their debts by taking new loan

LOAN STATUS AGAINST PARAMETERS

- Grades A & B are safer grades

- Fixing term to be minimal looks in favor

- Most people with more than 10years of employment pays back fully.

BIVARIATE ANALYSIS

- Interest rates increase as term increase and proportionate charge off.

- Grades implies a similar pattern to term

- People living in mortgage homes tend buy higher interest rate loans, but high charge off as well. Which supports our previous observation which says people buy more loans for debt consolidation
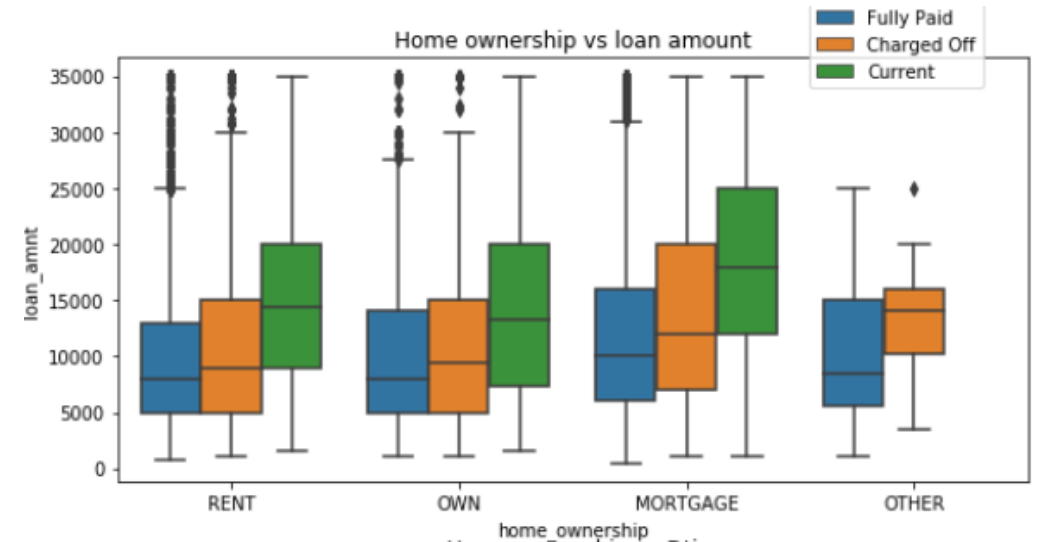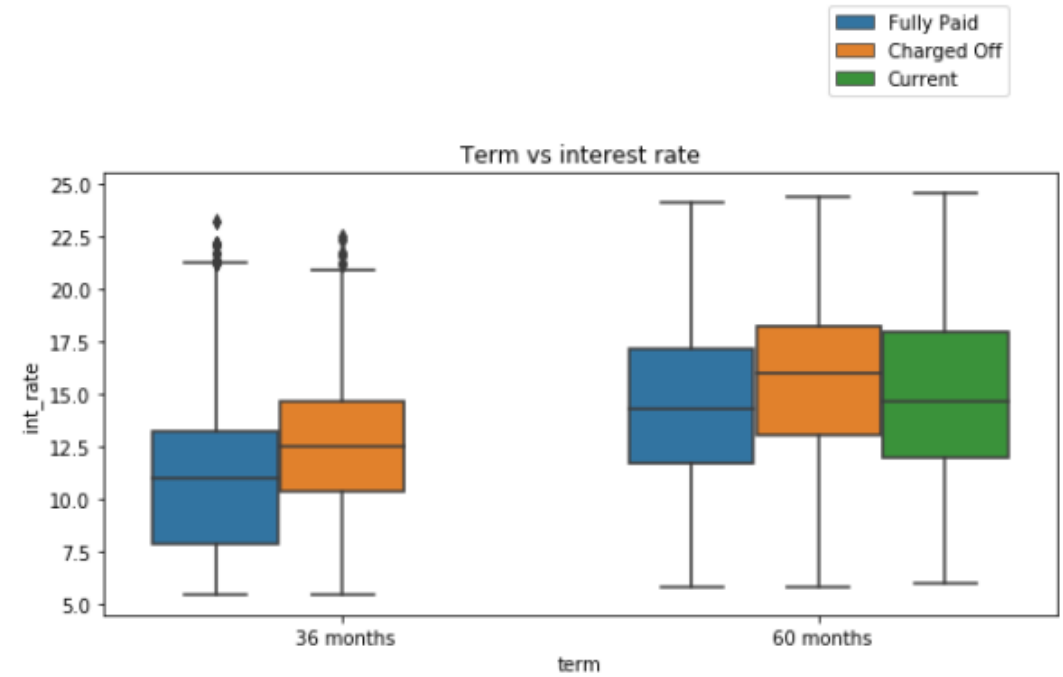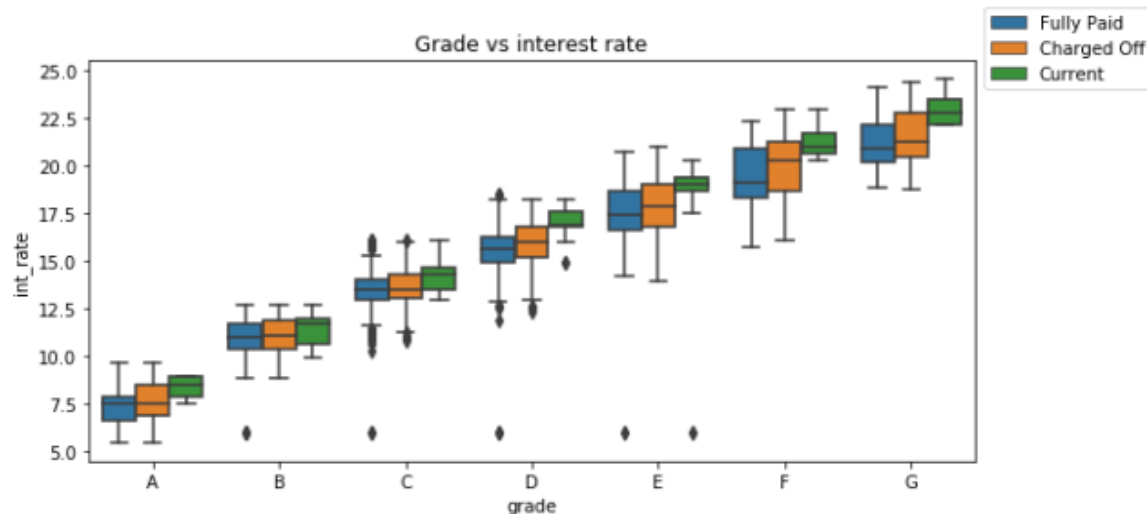
BIVARIATE ANALYSIS

- Verification status gives strange observation verified loans are the ones having high charge off ratio compared to 'Not Verified' ones.
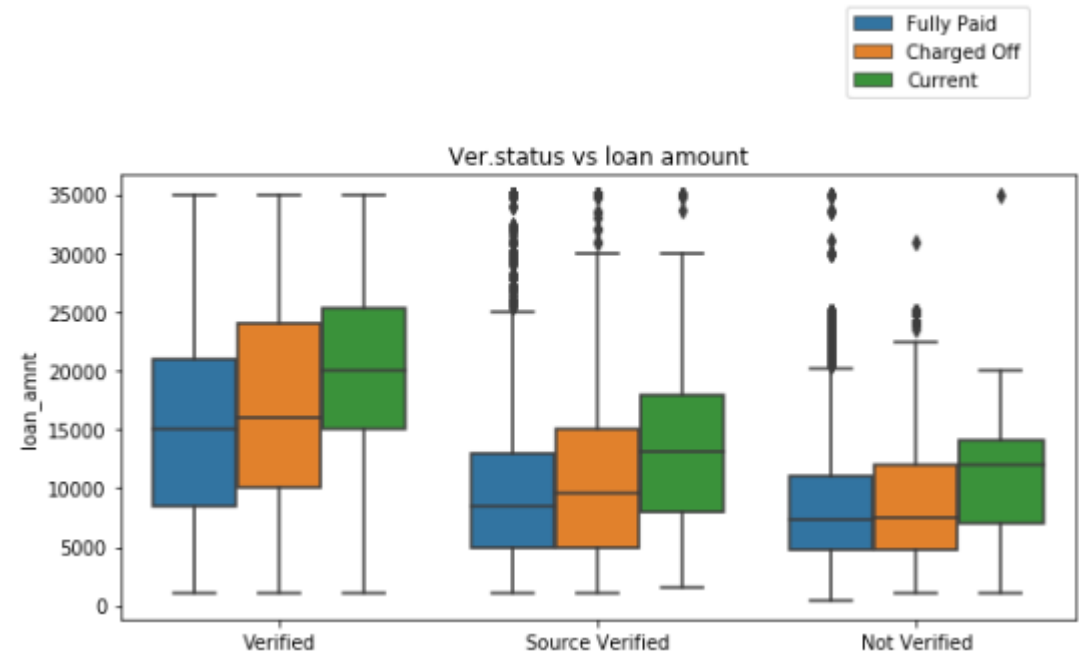
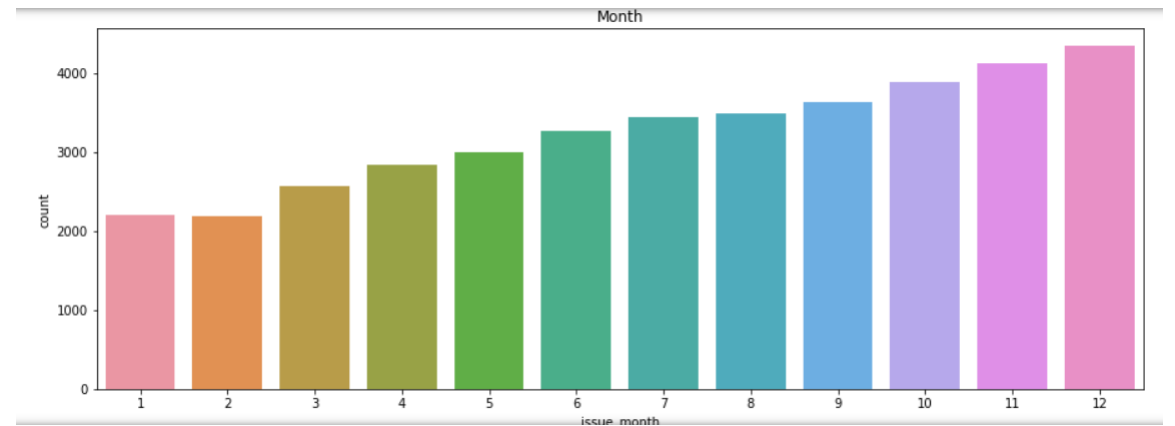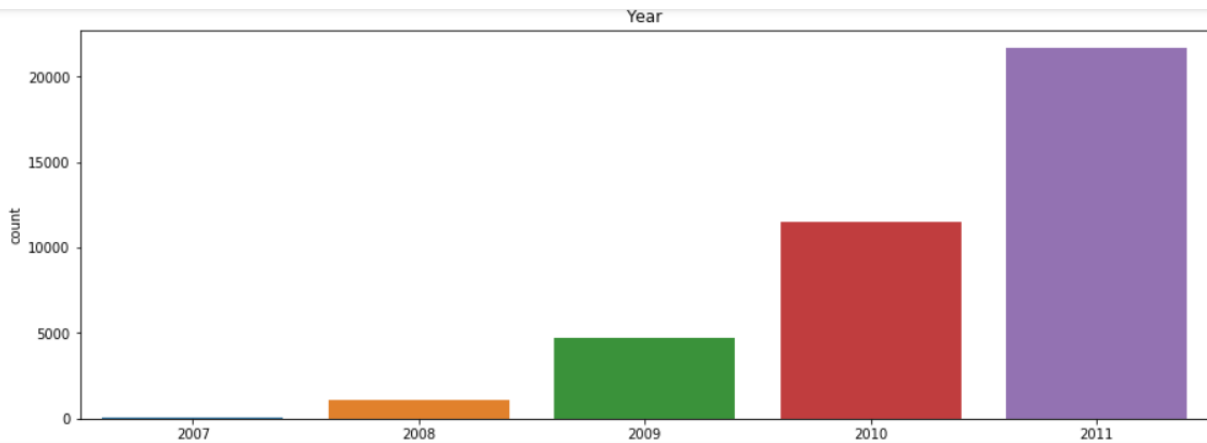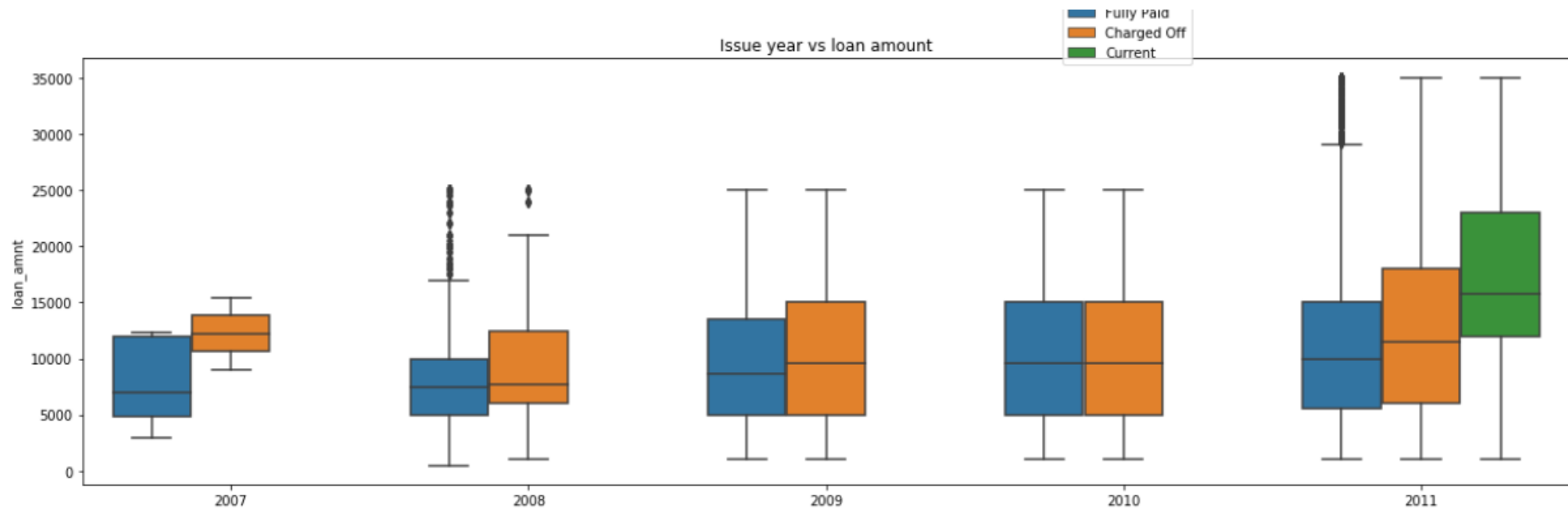# ANALYSIS BASED ON TIMELINE

- Number of applications keep increasing year on year

- Number of applications are more towards end of the year

ANALYSIS BASED ON TIMELINE(CONTD...)

- Though initially charge offs more at higher loan amounts, it was coming down in upcoming years till 2010. And 2011 we could see increase on charge offs again

Issue year vs loan amount

CORRELATION

The overall correlation of the data is as follows

- Loan amount and funded amount are highly correlated to each other

- Installment is the next closely correlated parameter with loan amount

- Pub_rec plays a crucial role in approval of requested amount as well as is highly correlated with 'pub_rec_bankruptcies'

- Total repaid amount also has strong positive relationship with loan amount
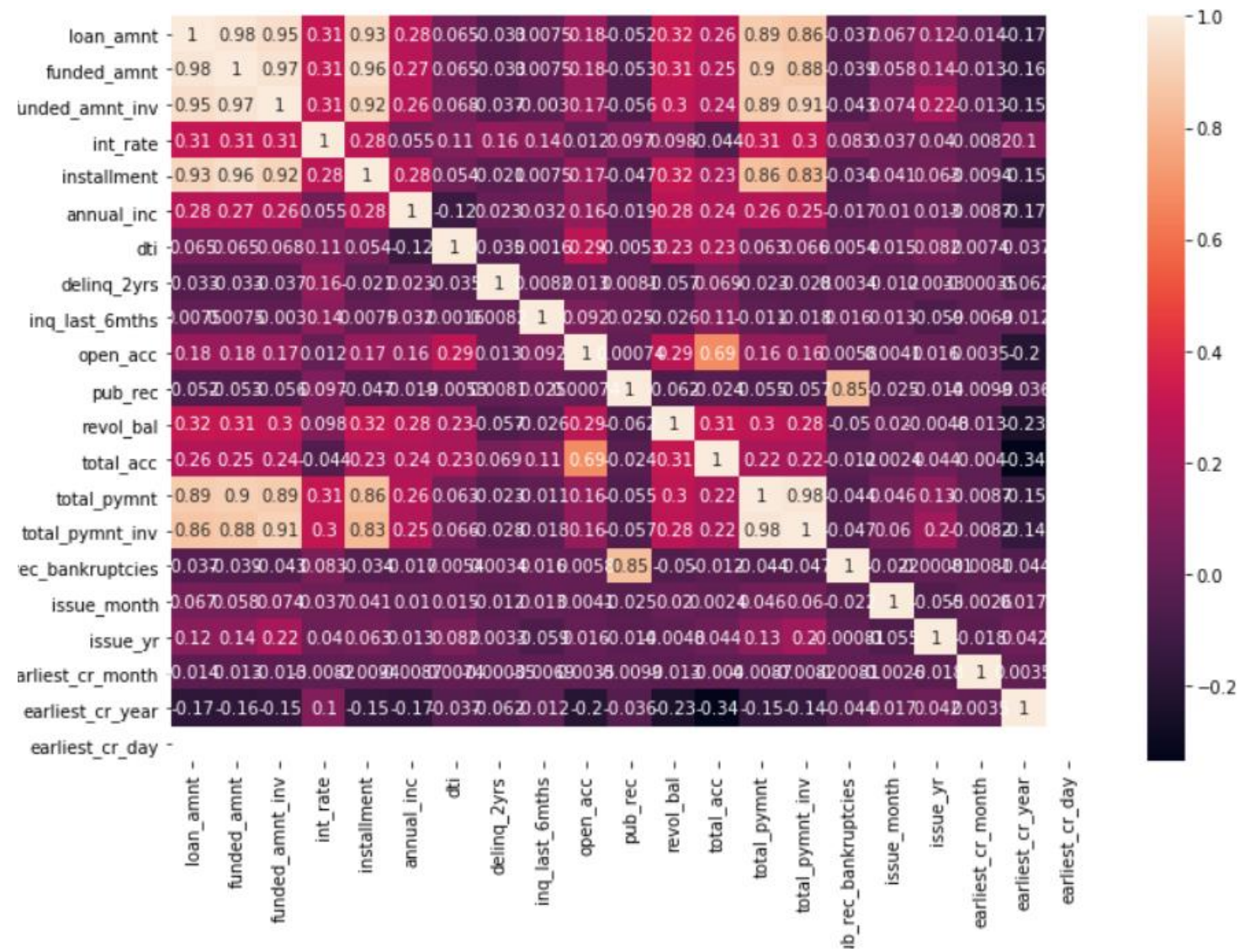
CONCLUSION

Less Risky	More Risky
Loan amount within the range of 5000-1000	Loan amount more than 30,000
Grades A&B	Grades F
Interest rates less than 7%	Interest rates more than 15%
Terms around 36 months. Possibly introduce new term of 24,48 months	Terms equal or more than 60 months are risky

THANK YOU