Report

P Sri Akhil
IIT Kharagpur

Problem Statement :

These files contain complete loan data for all loans issued through 2007-2015, including the current loan status (Current, Late, Fully Paid, etc.) and latest payment information.

Objective is to identify the risky loan applicants at the time of loan application so that such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

In other words, to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment. And thus minimise the risk of losing money while lending to customers.

Analysis Approach

Clean data:-

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

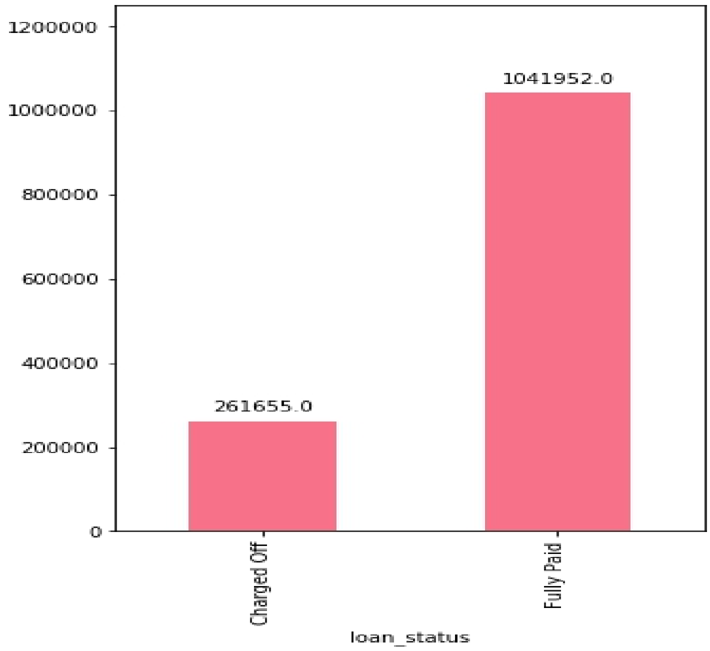
Analyze data :-

-> Check distributions and frequencies of varios numerical and categorical variables.
-> Analyze variables against segments of other variables.
-> Create derived variables.
-> Do correlation analysis.
-> Check how two variables affect each other or a third variable.
-> Analyze joint distributions.

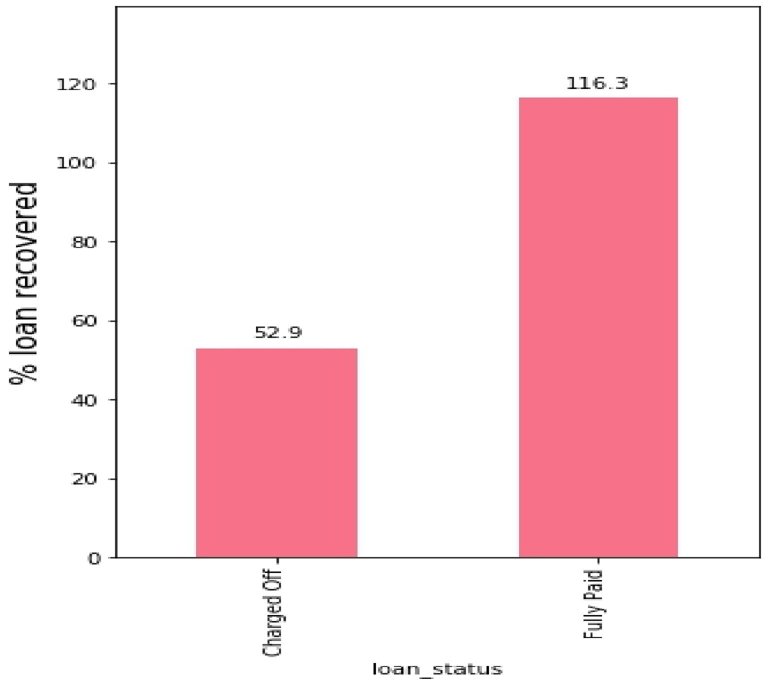
Summarize Results :-

-> Publish insights and observations.

Analysis - Overall Loan Status

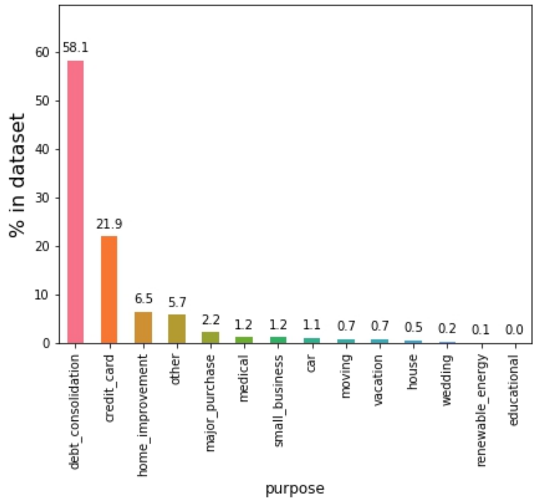


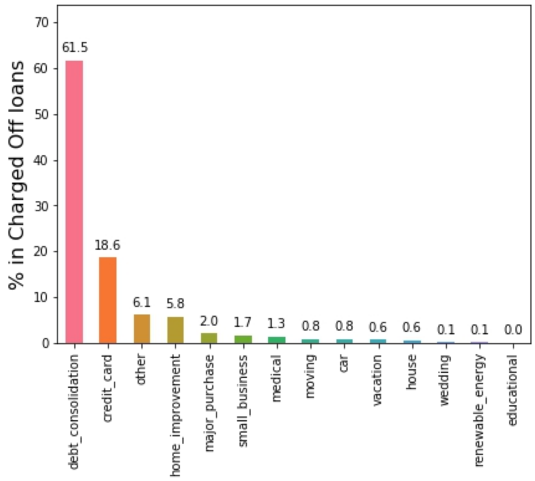
Approximately 21% of the loans are defaulted.



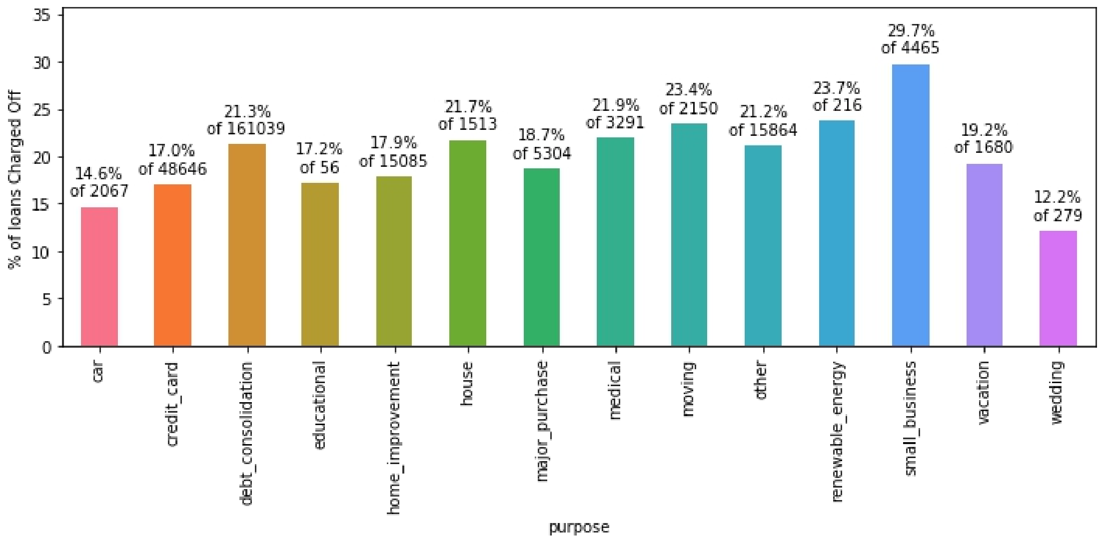
- **Company recovers only 53% of the loan amount when loans are defaulted.**
- **On fully paid loans company makes 16.3% profit.**

Analysis - Purpose of Loans



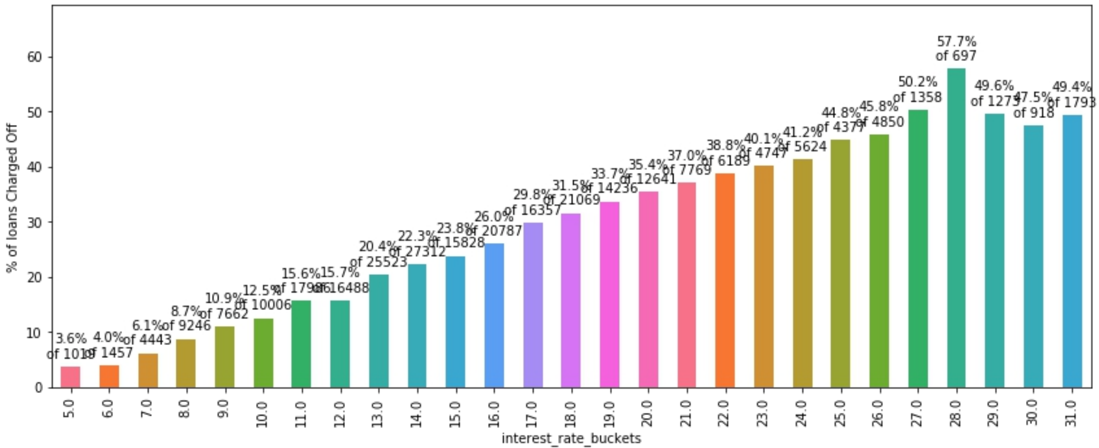


- Max number of loans are of Debt Consolidation followed by credit card.

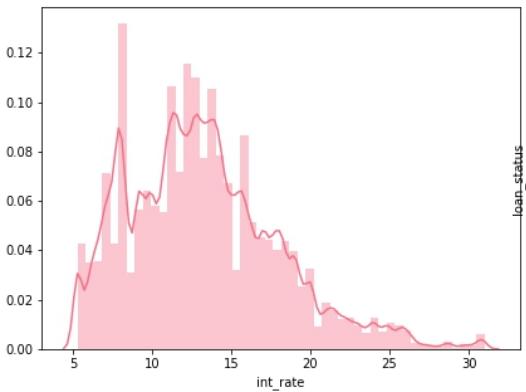


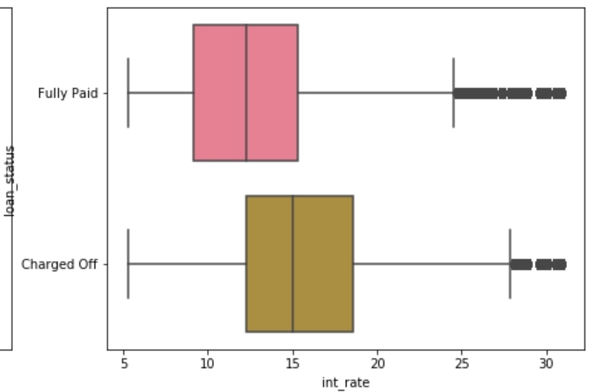
- Small Business is the most risky purpose.
- More than quarter of loans taken for purpose of Small Business see defaults.

Analysis - Interest Rate



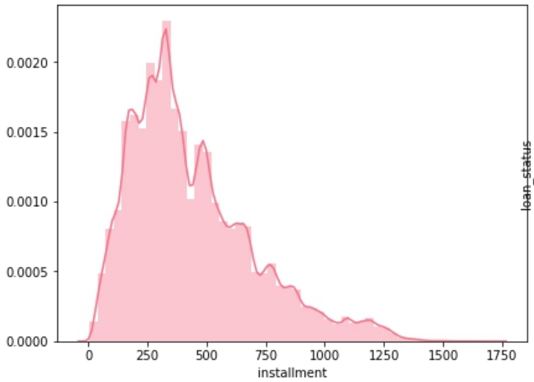
- Percentage of default loans increases monotonically with higher interest rates.

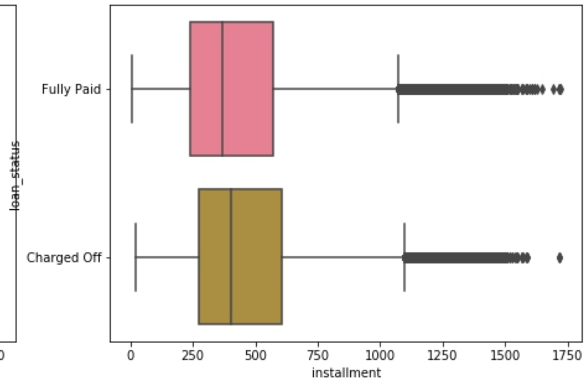




- Overall, the interest rate varies from 5.31% to 30.99%. The interest rate for Charged Off loans appear to be higher than for Fully paid. This is naturally expected. As, the risk increases the rate of interest imposed on the loan also increases.

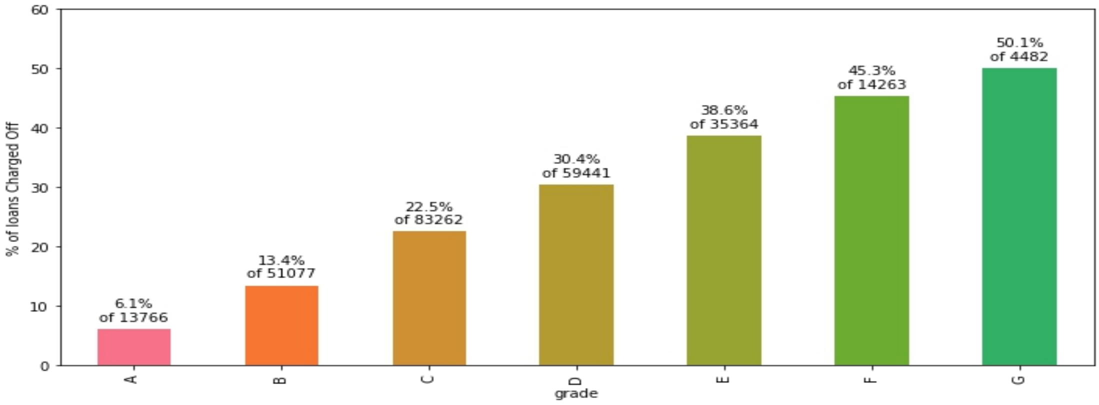
Analysis - Installment



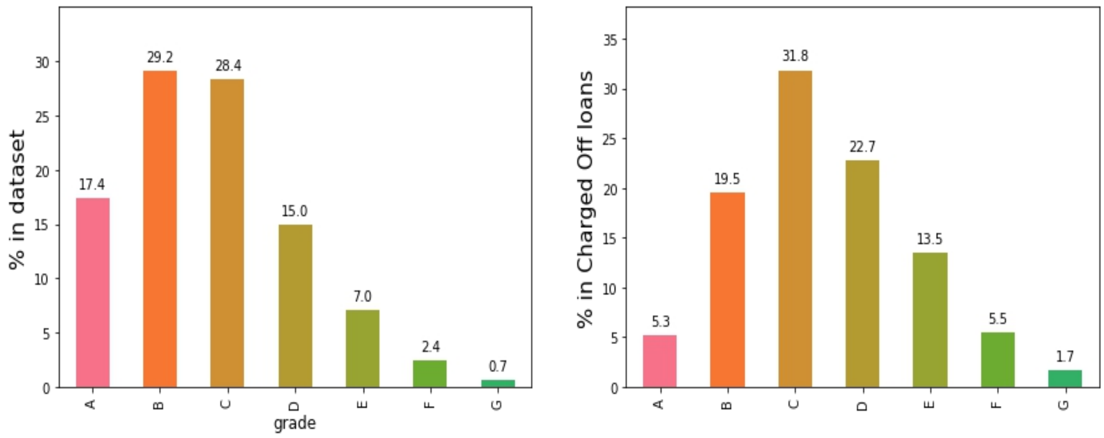


◆ **Loan Installment have high installment on average.**

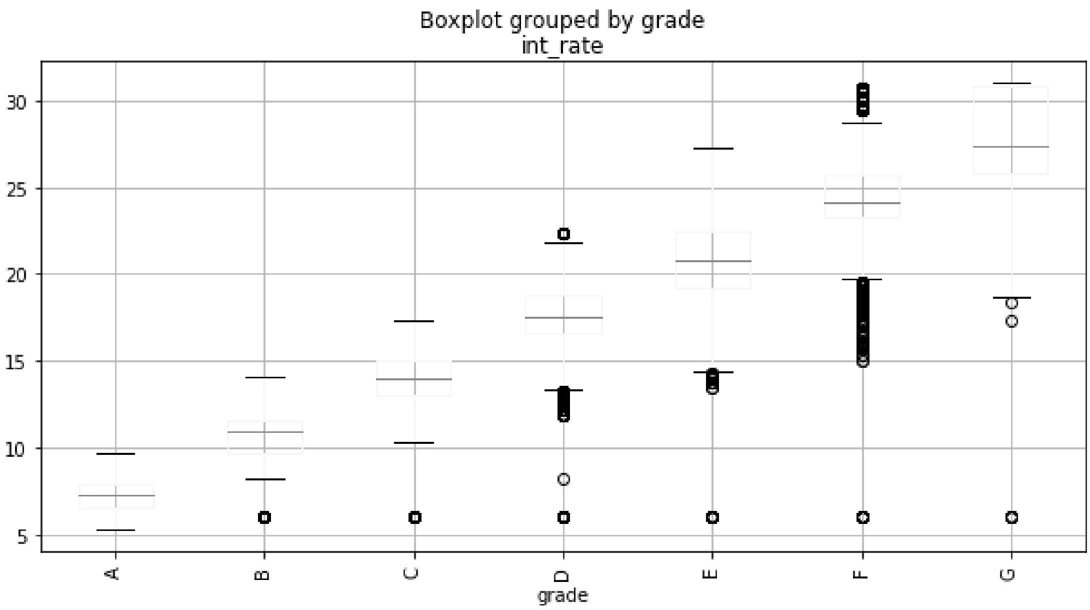
Analysis - Grade



◆ **Lower grades have higher incidence of defaults on loans.**

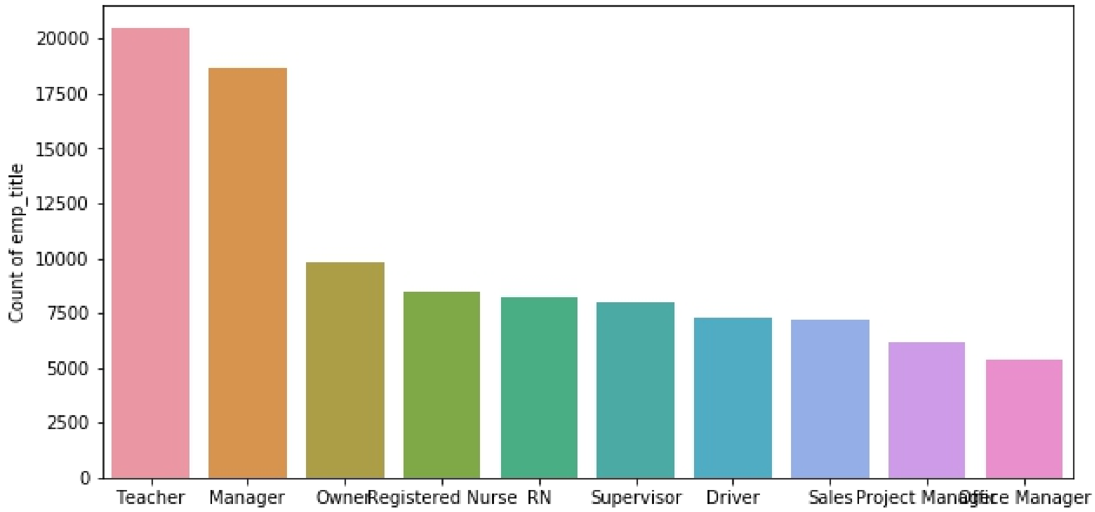


- **Grade A and B loans are safe. The percentages in full dataset are much higher than percentages in Charged Off loans. Grade D, E, F, G loans are less safe.**
- **Grade E onwards are risky, and less numerous.**



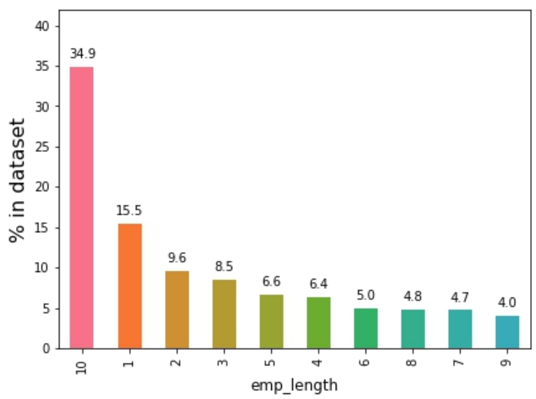
- **Higher rates for lower grade loans.**

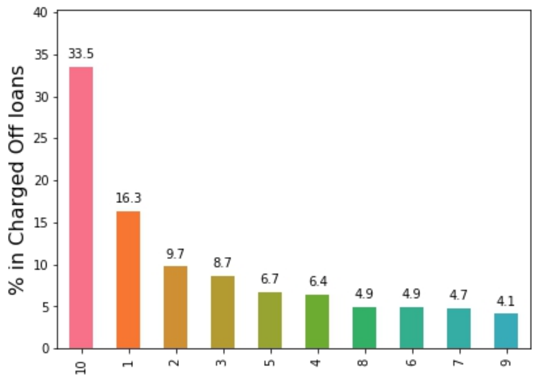
Analysis - emp-title



- Max of the loan takers are Teacher, Manager follwed by Owner and Registered Nurse.

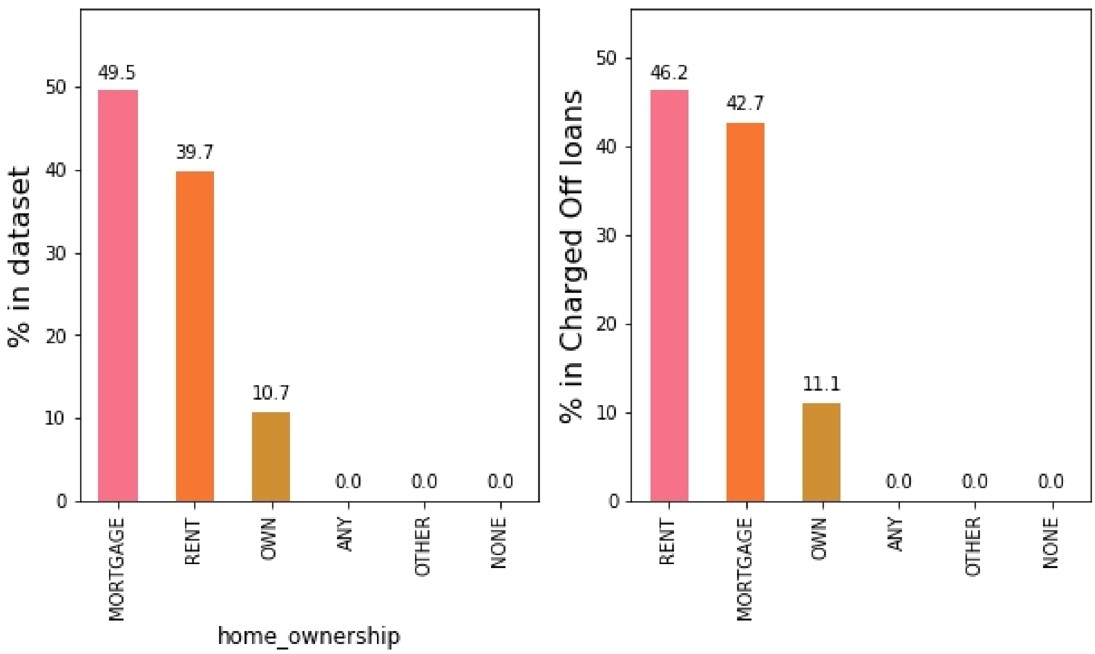
Analysis - emp-length





- 35% of the of loan taker have an employment length of more than 10 years.
- There is no impact of employment length to chance of loan default.

Analysis - home ownership



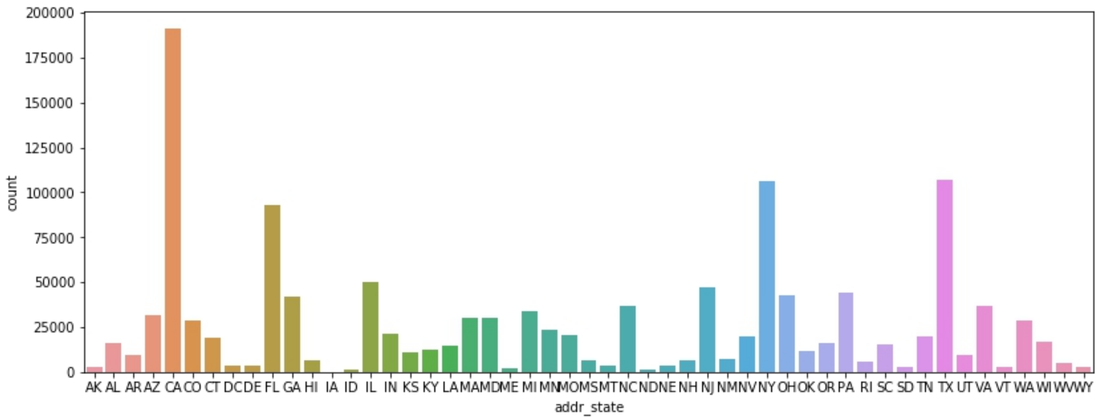
- There is no significant impact of home_ownership and default on loan.

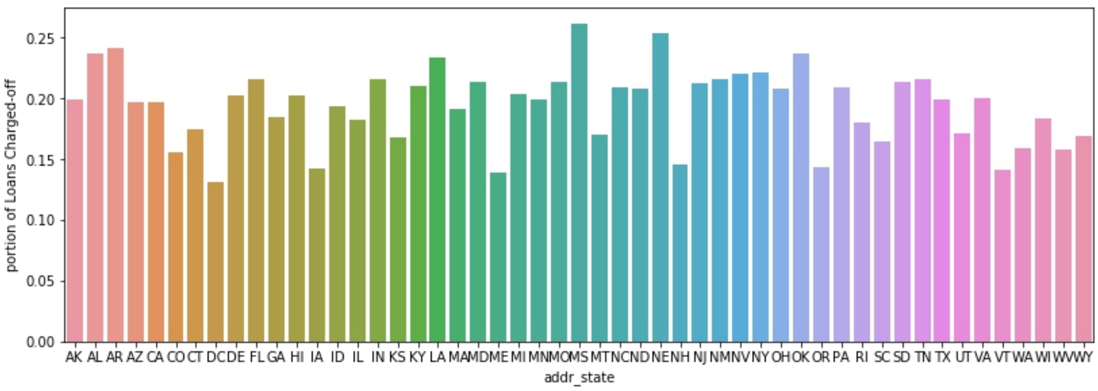
Analysis - Annual-inc

```
count    1.303607e+06
mean     7.615899e+04
std      7.004890e+04
min      0.000000e+00
25%      4.584000e+04
50%      6.500000e+04
75%      9.000000e+04
max      1.099920e+07
Name: annual_inc, dtype: float64
```

- Median annual income is around 65,000. Most people have an annual income less than 90000.

Analysis - addr_state

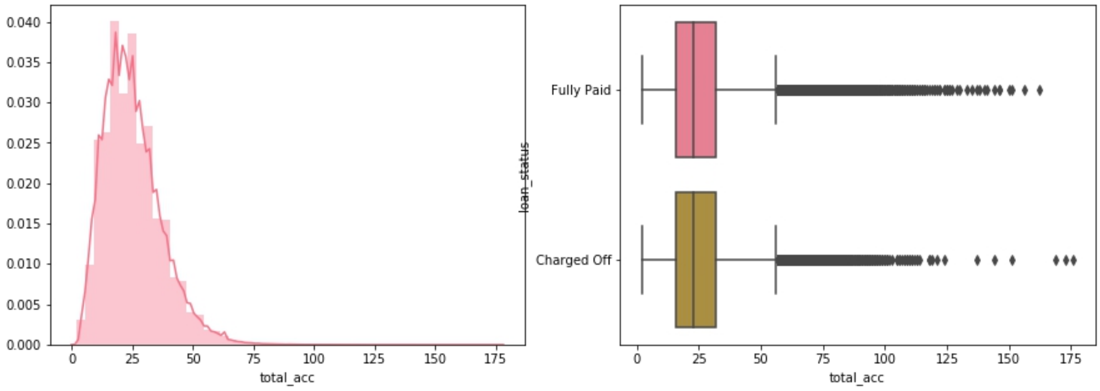




- Maximum loans ~18% are from California following from New york state and Florida .This is to be expected as these are also the three most populous US states.

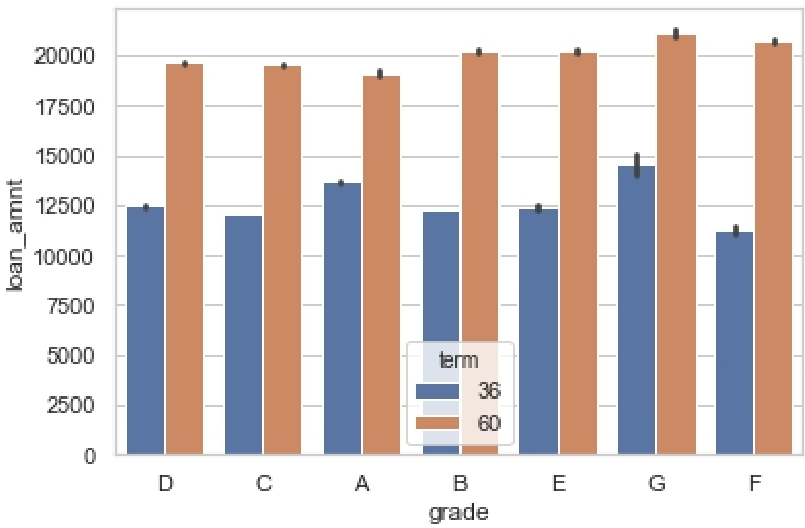
- States with higher Charge Off rates have very low number of loans.

Variable - total_acc



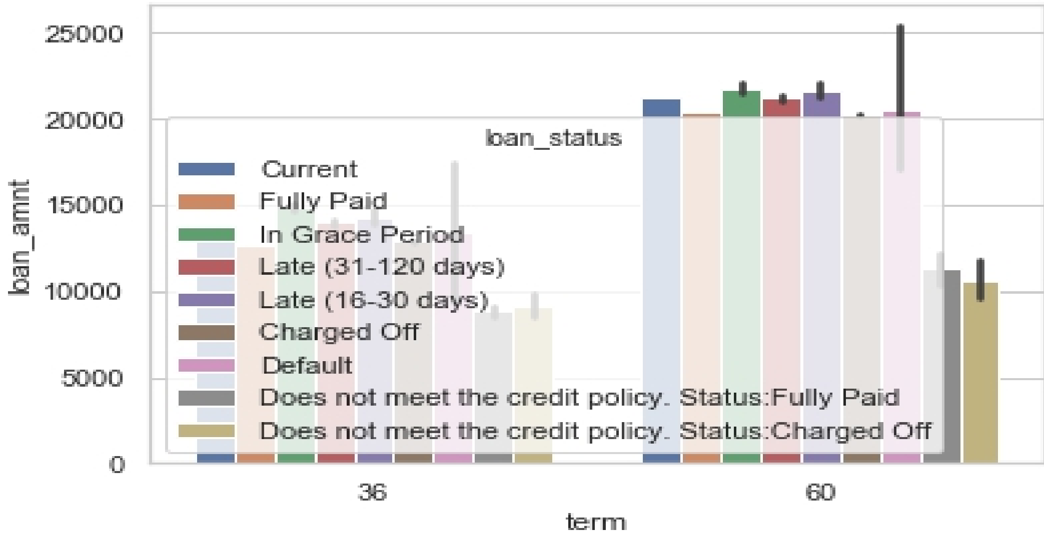
- The are approx. 25 average credit lines for the borrower's.
- There is no big difference between the Charged Off and Fully Paid credit lines.

Loan amount vs Grade vs Term



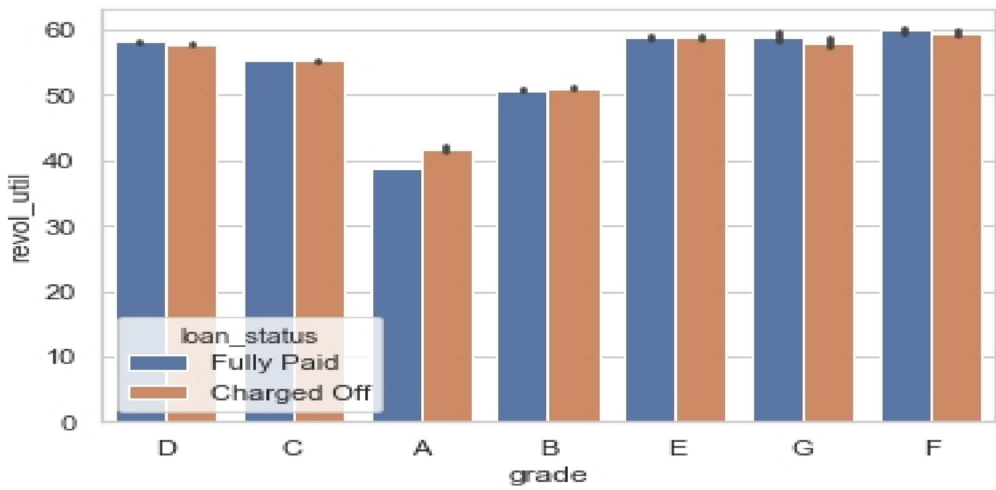
- Higher loan amount are associated with lower grade for longer terms.

Loan amount vs Term vs Loan status



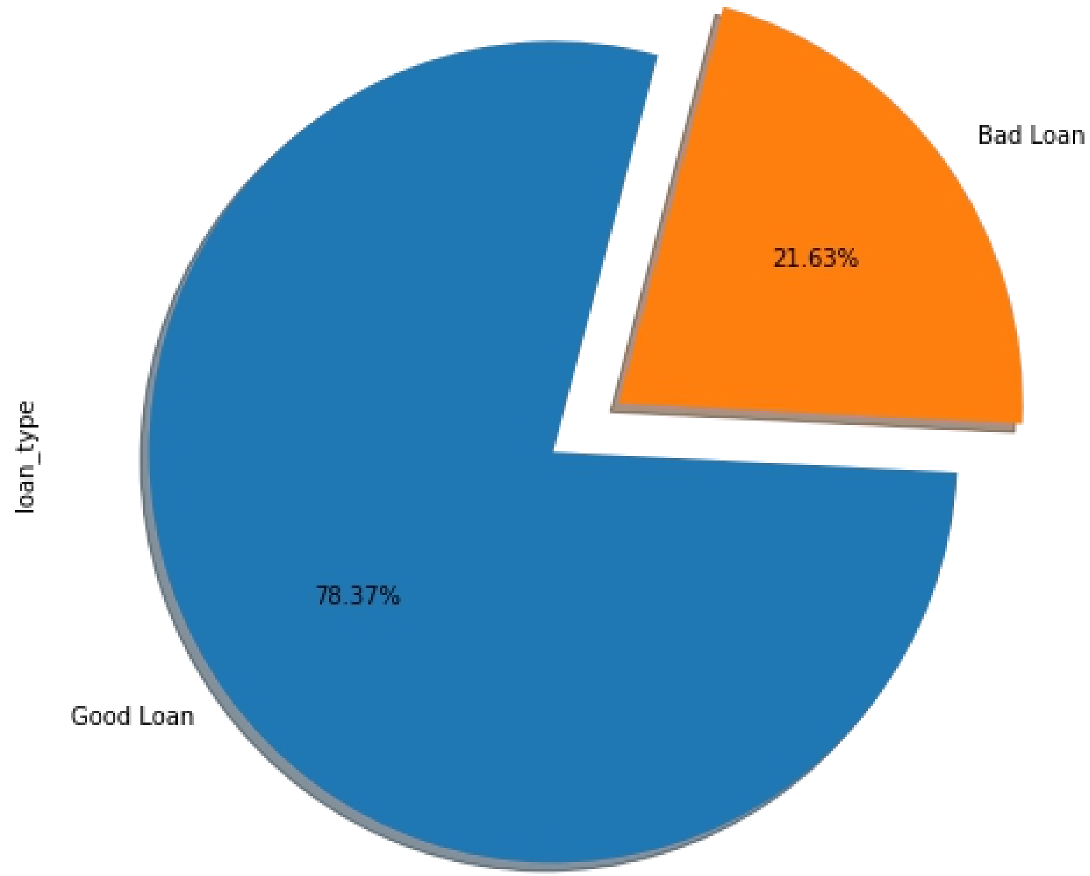
- **Higher loan amount are associated with longer terms and see higher Charge Offs.**
-

revol_util Vs grade Vs Loan Status



- **The revol_util is positivly correlated to the grade.**

Analysis - Bad vs Good loans



- **Bad vs Good loans by the company is 78.37 : 21:63**

Suggestions for company :

- **Reduce number of loans when purpose is small business.**

- **Stop approving loans to people with bad record / high value loans.**

- **Stop approving loans amount of 30% of annual income or higher.**

- **Start charging higher interest rates for loans with dti greater than 20.**