# Lending Club Case Study

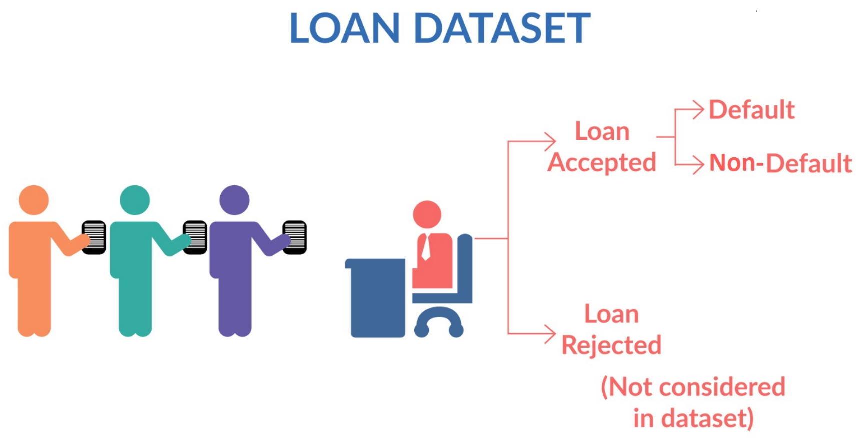
Group members:

- Bharat

# Problem Statement

Given a dataset of loan records having various attributes along with their end status (whether they are fully paid or defaulted). Our objectives are:

1. To identify the risky loan applicants, so that such loans can be reduced thereby cutting down the amount of credit loss
2. To understand the driving factors (or driver variables) behind loan default, i.e., the variables which are strong indicators of default



# Analysis approach

1. Understanding Dataset
(Using given Data-Dictionary)

2. Data Cleaning
(Drop/impute columns, Outlier treatment)

3. Perform EDA
(Univariate, Bivariate analysis etc.)

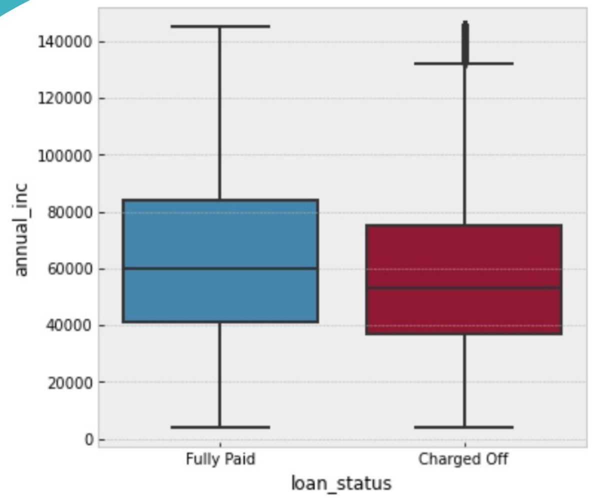
4. Present Results
(Charts, boxplots, histograms etc.)

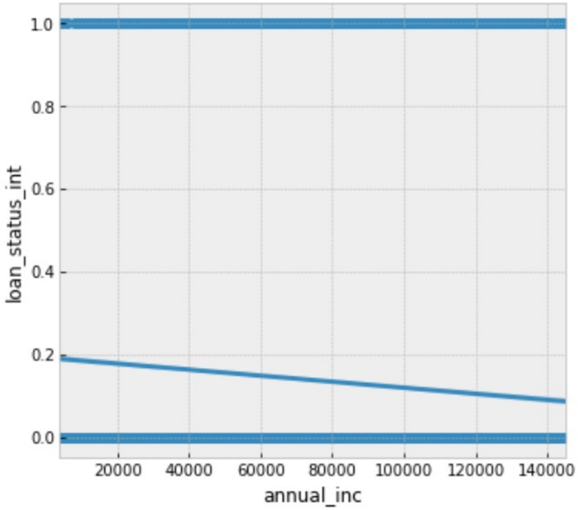
# Data Cleaning Steps

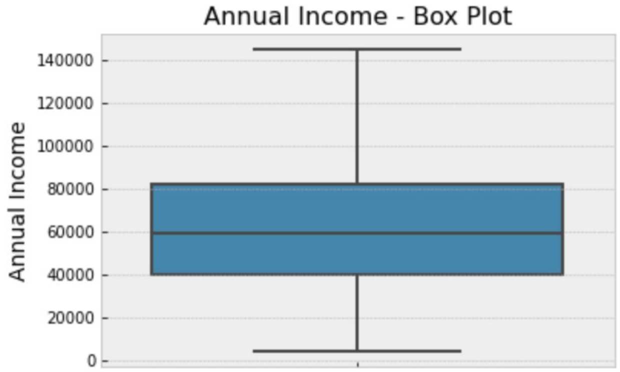
1. Removed member_id and the all-other features which have 30% or more null values
2. Dropped some behavioral columns which are not available at the time of loan application, hence we dropped them. E.g., 'delinq_2yrs', 'earliest_cr_line', 'inq_last_6mths', 'open_acc' etc.
3. We removed all the loan records which are currently running, because we don't know whether they will result in 'Charge-Off' or 'Fully Paid'.
4. Converted int_rate (interest rate) from object to numeric, by removing the '%' symbol.
5. Standardized the emp_length column, by removing the special chars
6. Performed outlier treatment over annual income using formula: (Q3 + 1.5 * IQR)

# 1. Annual Income analysis

- Most of the applicants have annual income in the range of 40k-80k
- Applicants with lower annual income have high chances of doing default on loans as compared to applicants with higher annual income.

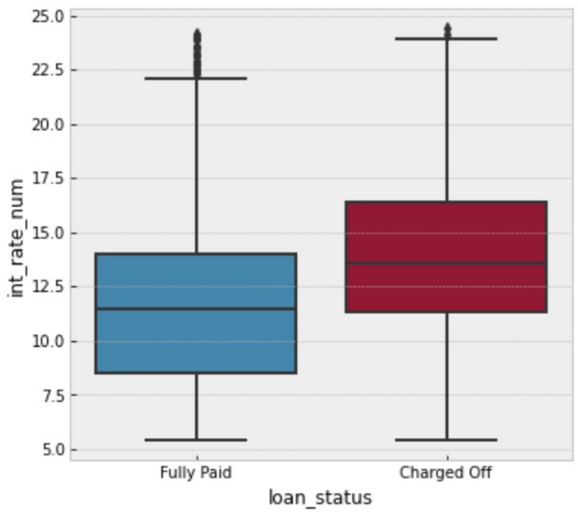


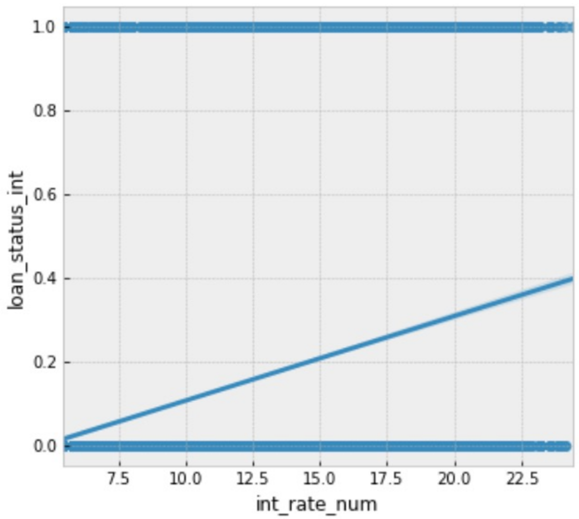



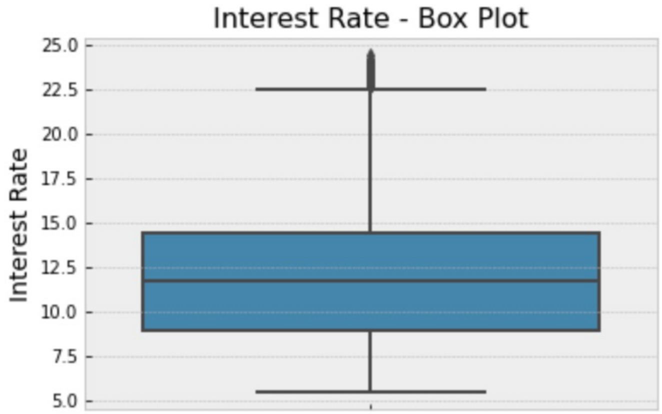



# 2. Interest Rate analysis

- Most of the loans interest rates are in the range of 9% - 14%
- As the interest rate increases, the chances of default also increases
- For the interest rate greater than 15%, the chances of defaults are high.

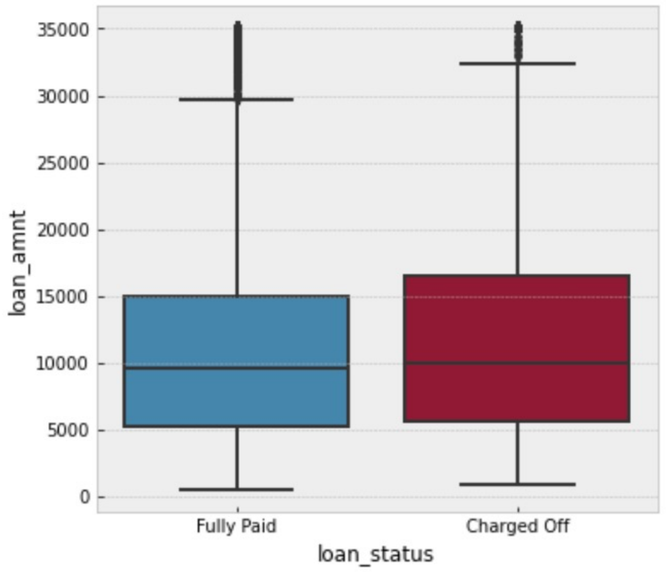


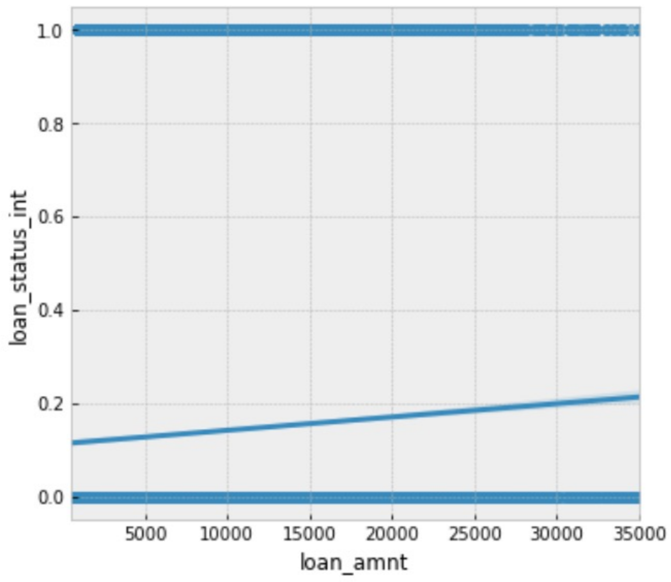



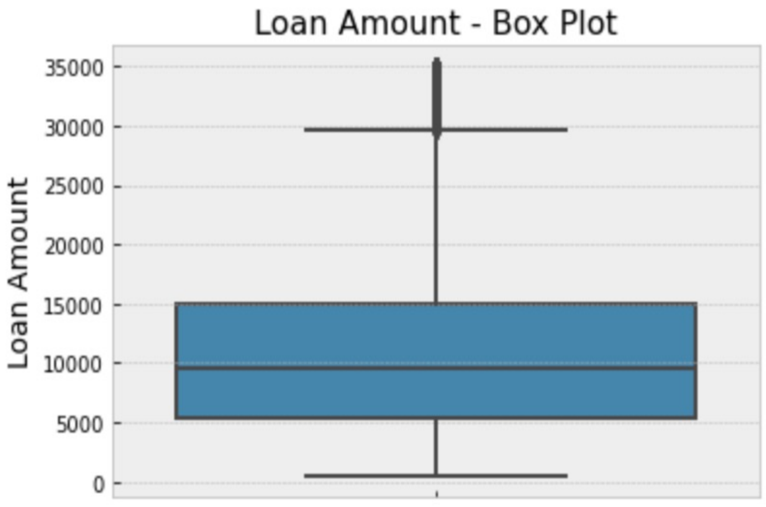



# 3. Loan amount analysis

- Loan amount for most of the loans is between 5000 - 15000
- As you can see in the middle pic below, Loan amount is positively correlated with loan status, That means as the loan amount increases, chances of default also increases.

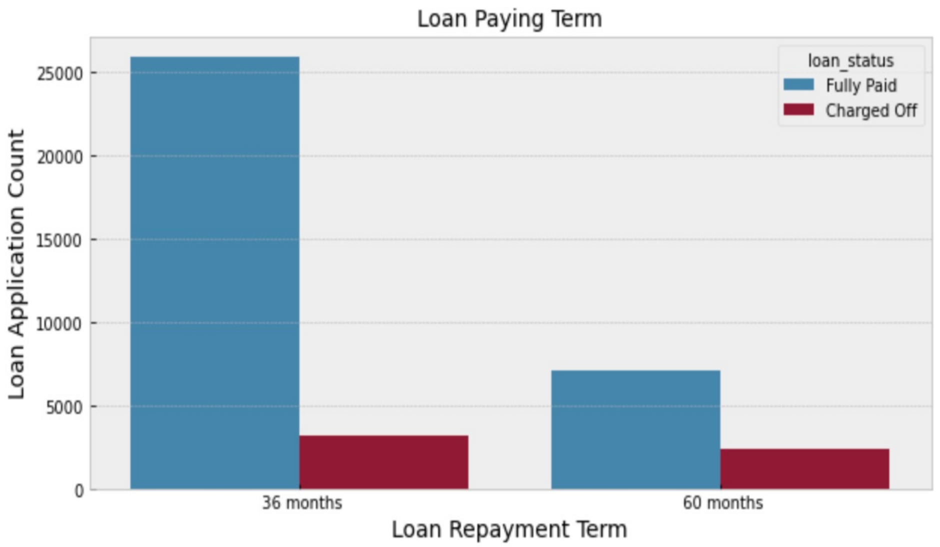


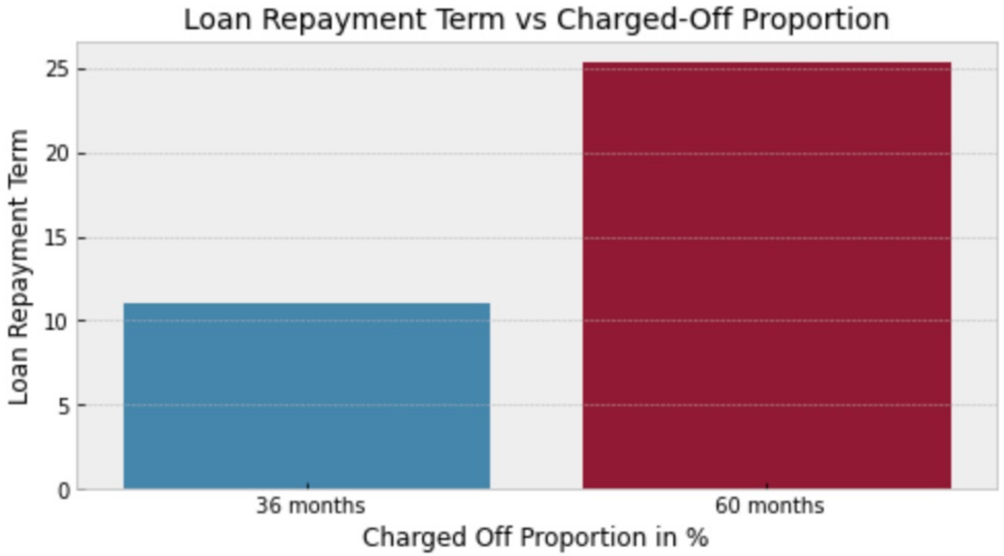



# 4. Loan Repayment Term

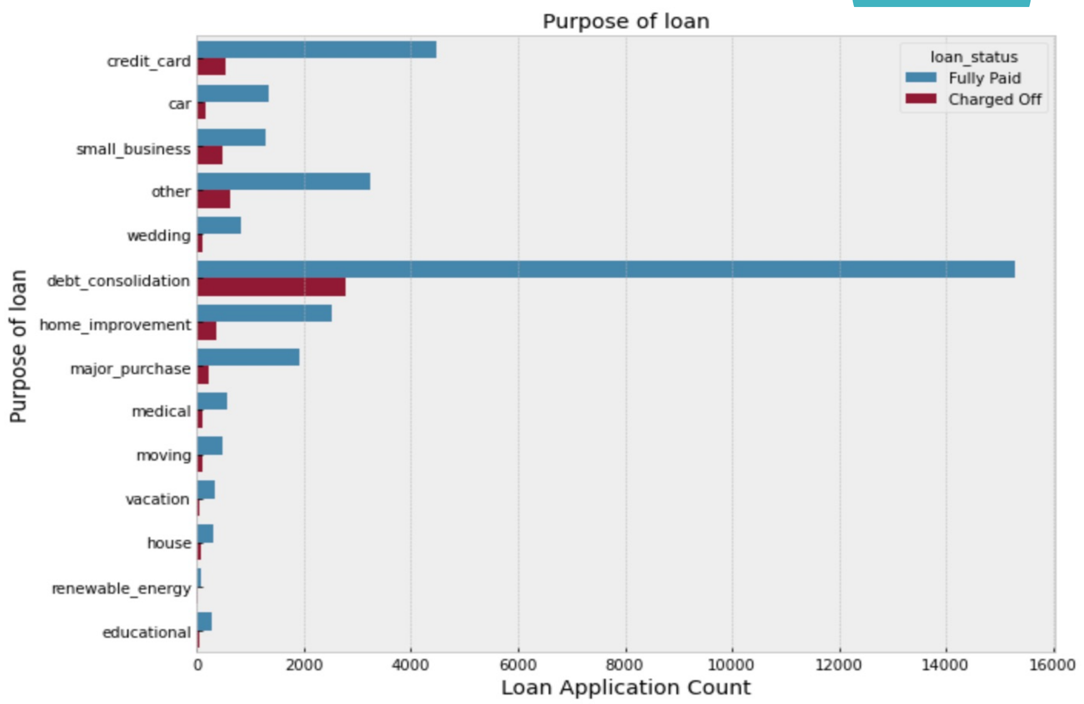
- Most number of loans are having repayment term of 36 months.
- The charged-off proportion is quite high in loans with repayment term of 60 months.

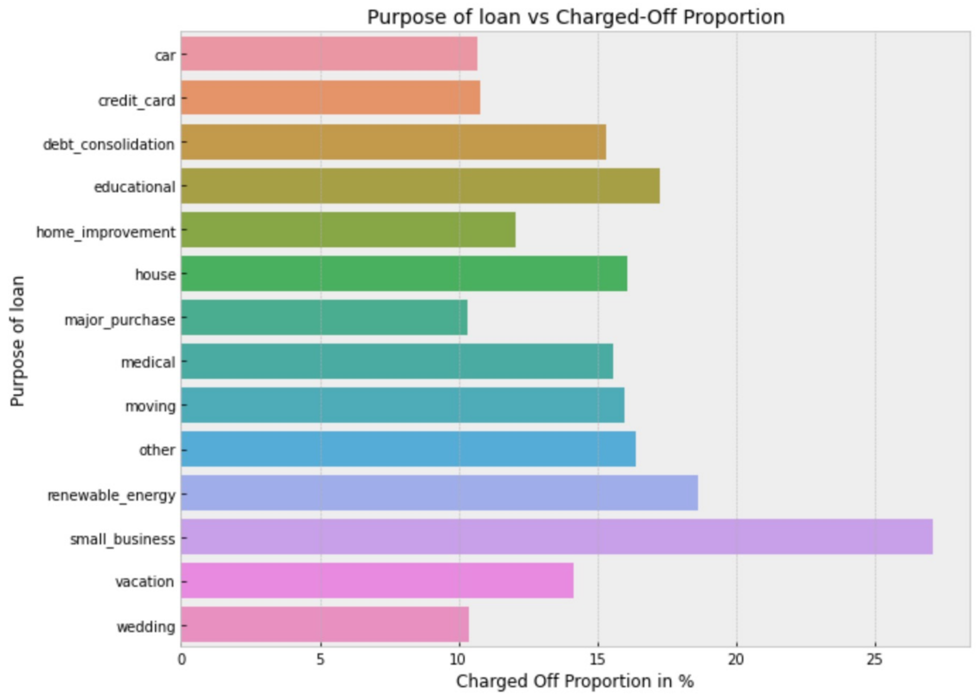


# 5. Purpose of loan

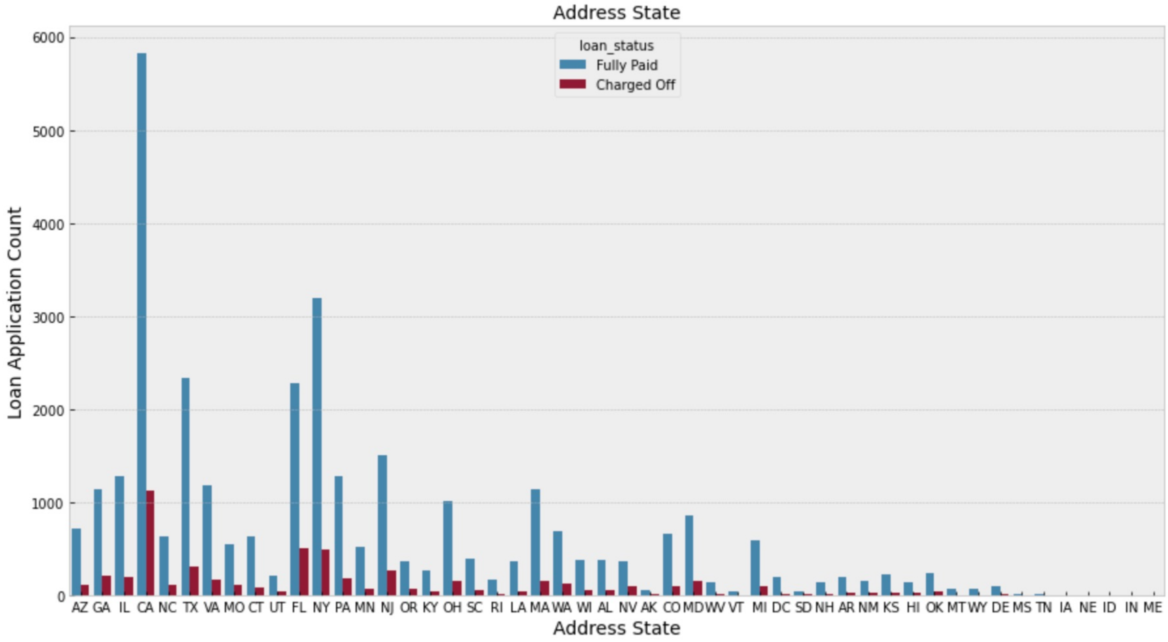
- Most number loans are taken for the purpose of debt consolidation and credit card.
- Whereas the charged-off proportion is quite high for the loans taken by small businesses.

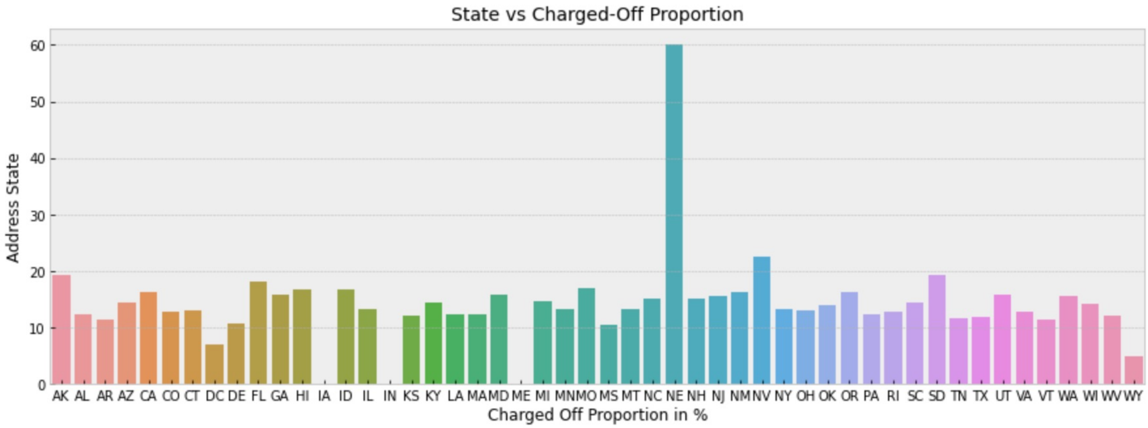


# 6. Address State

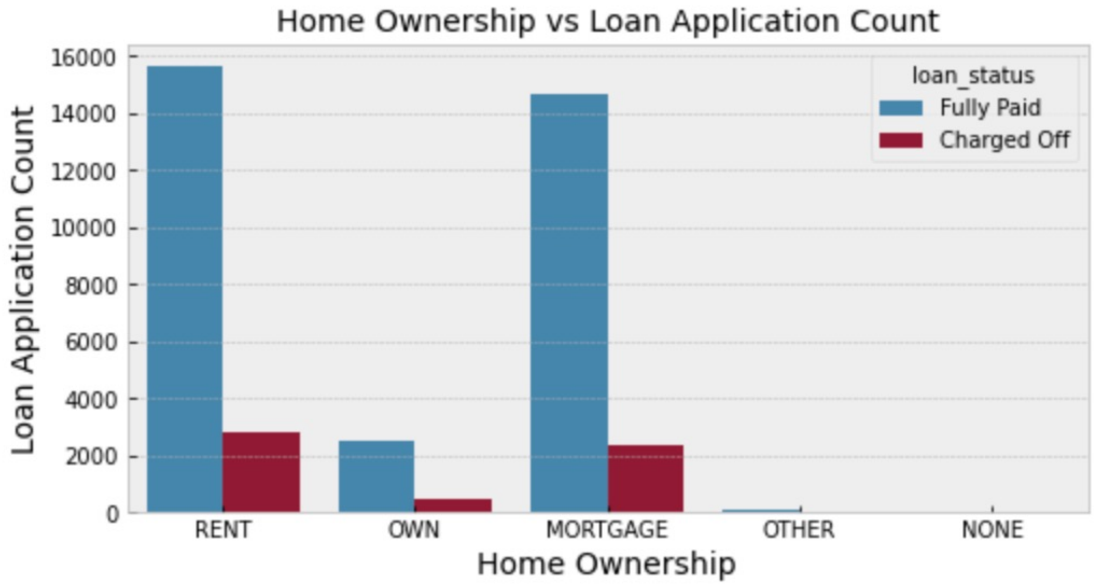
- Most number loan applicants are from State ‘**CA**’, and hence most charge-off are also from ‘**CA**’
- But the state with high charge-off proportion is ‘**NE**’

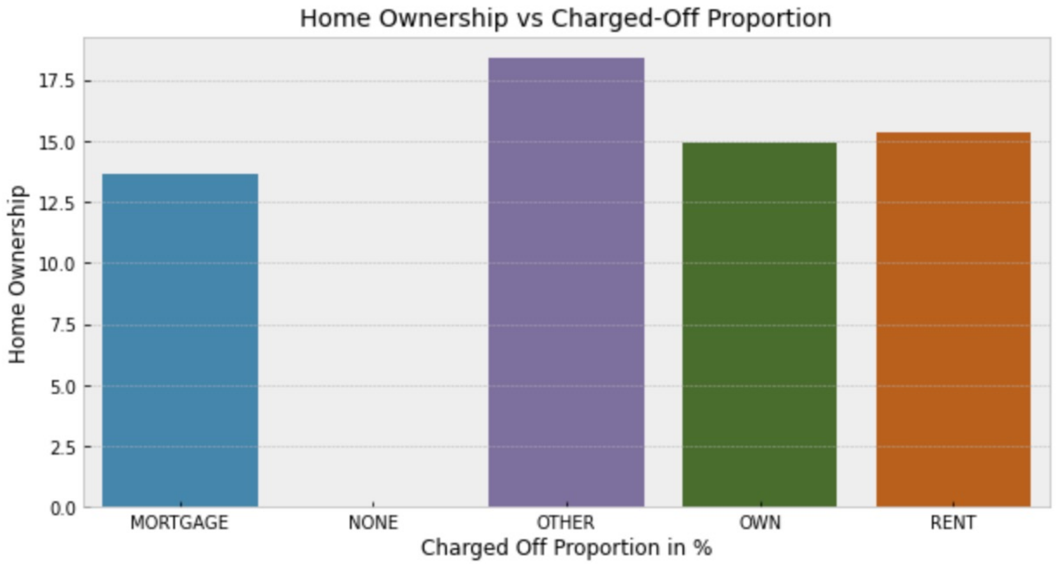


# 7. Home ownership

- There is an equal possibility of applicants doing default on loan repayment, no matter whether they own a home or not. So, we can not use this feature in our predictions.



# Conclusion

1. Applicants with lower annual income have high chances of doing default on loans as compared to applicants with higher annual income.
2. As the interest rate increases the chances of loan default also increases
3. Chances of defaults are high when repayment term length is longer than 36 months
4. Chances of charge-off is very high in case of **small businesses**, so due diligence should be performed while considering those loan applications.
5. Chances of charge-off are pretty high for the applicants from '**NE**' state, so we should carefully investigate their applications.