

# Lending Club Case Study

By Varun A M and Rahul Kumawat

# Problem Statement

- This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.
- Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who **default** cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.
- The company wants 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.

# Dataset and Approach

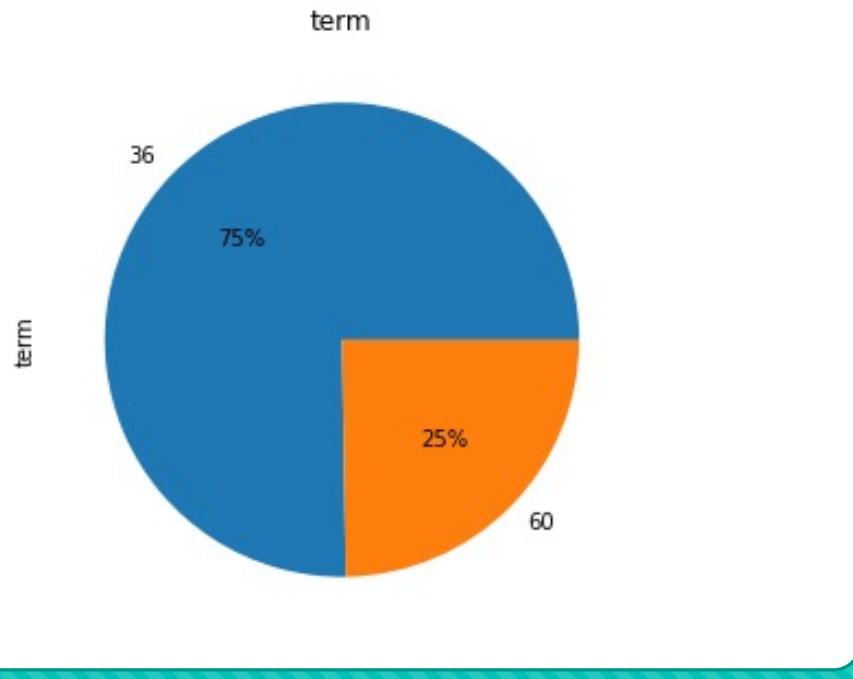
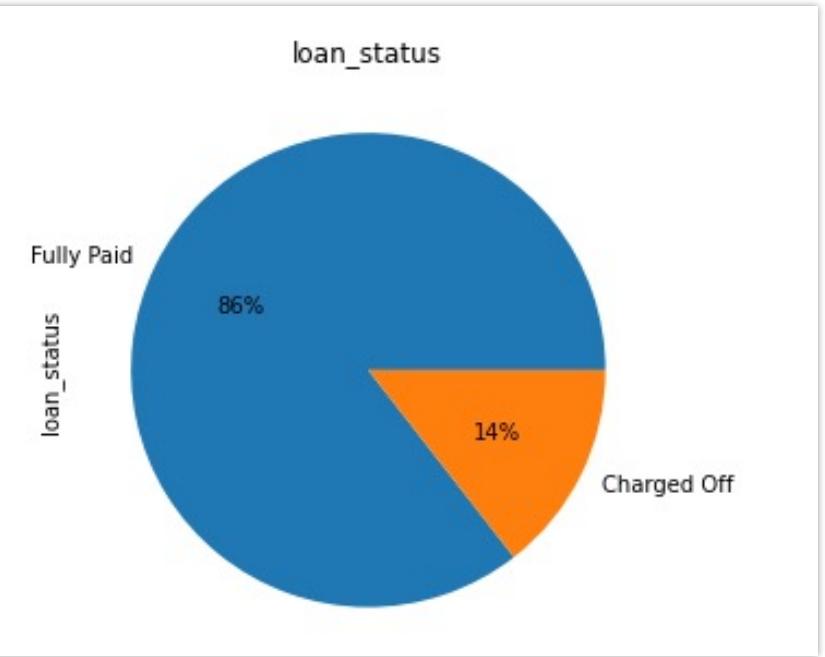
- Loans.csv
  - The file which has all the loan details from 2007 to 2011.
- Data Dictionary
  - Contains details of each column present in the data set.
- Approach
  - Data Cleaning → Sanitisation → Univariate Analysis → Bi-variate Analysis → Multi-variate Analysis

# Data Cleaning

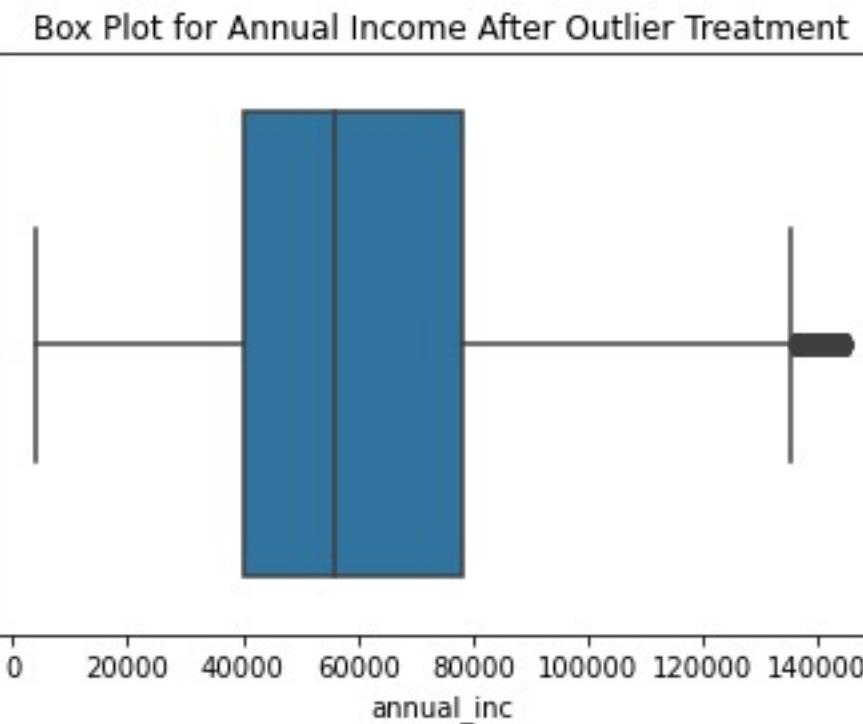
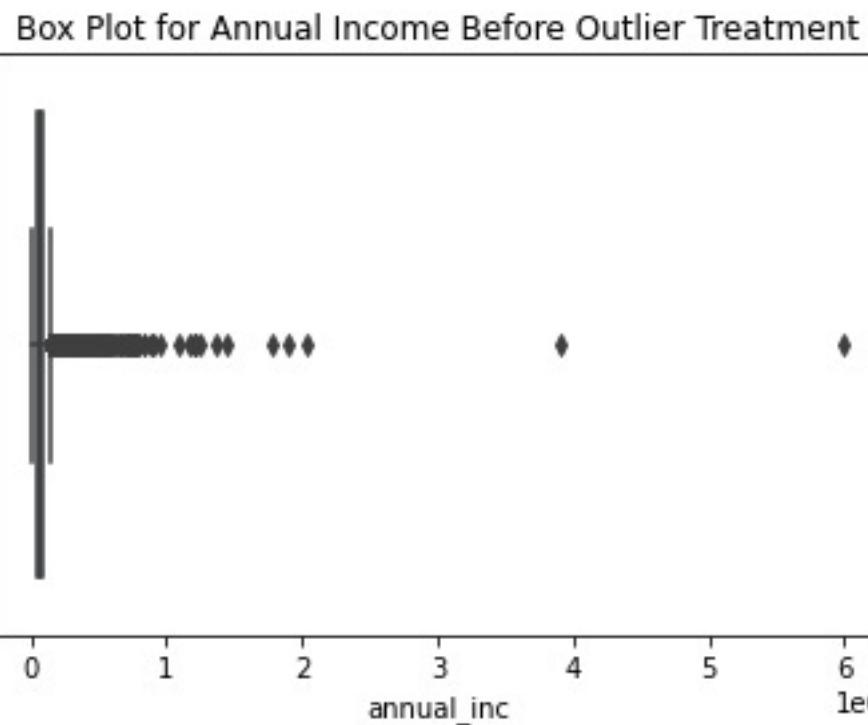
- Dropped all the columns which has complete null values.
- Loan status of "Current" was not necessary. Removed them.
- Filled the missing values in the employee length column with mode.
- Dropped the missing values in the revol-util column.
- Standardized columns like 'int\_rate','emp\_length','revol\_util'
- Wrote a function block to remove the character elements from the string 'emp\_length'.
- Performed sanitization checks like if the installment is lesser than loan\_amount/fund\_amount.

# Univariate Analysis

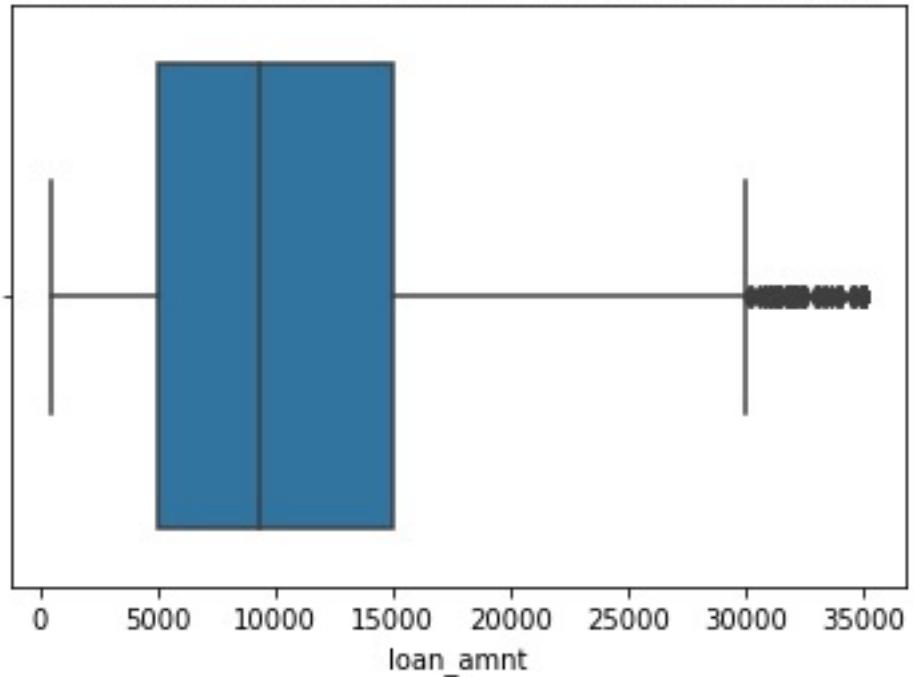
- Considered Grade/Sub-Grade as Ordered variables.
- Considered Purpose/Home-Ownership as Unordered variables.
- Considered Annual income as quantitative variable.
- Removed outliers by calculating IQR for quantitative variables and plotted box plots.
- Graphs are attached in the next slide...



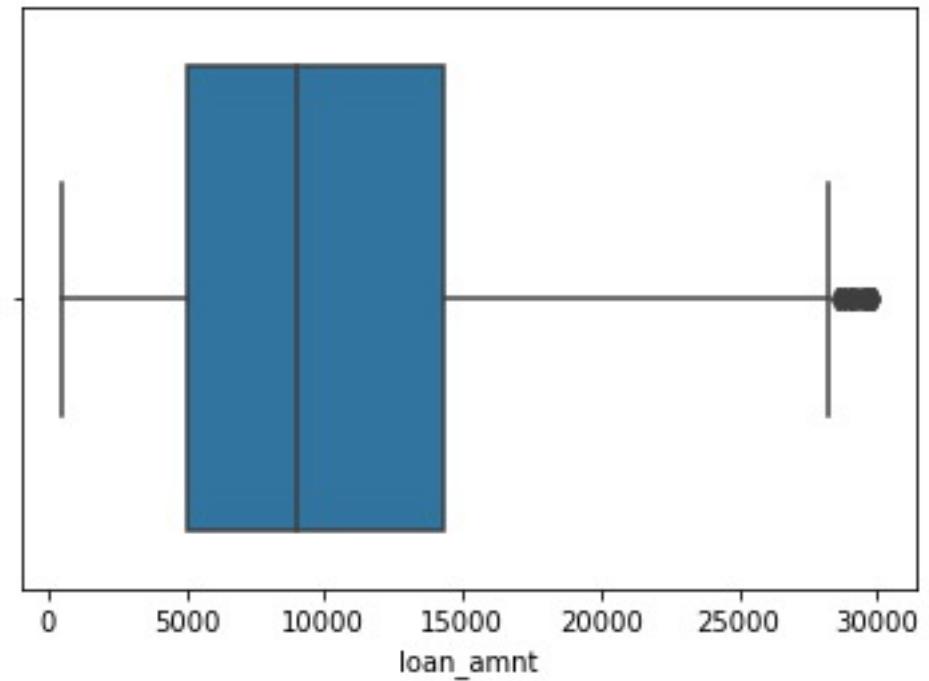
**14% of the loans were Charged Off.**  
**75% of the loans were for term 36 and 25% was for 60.**

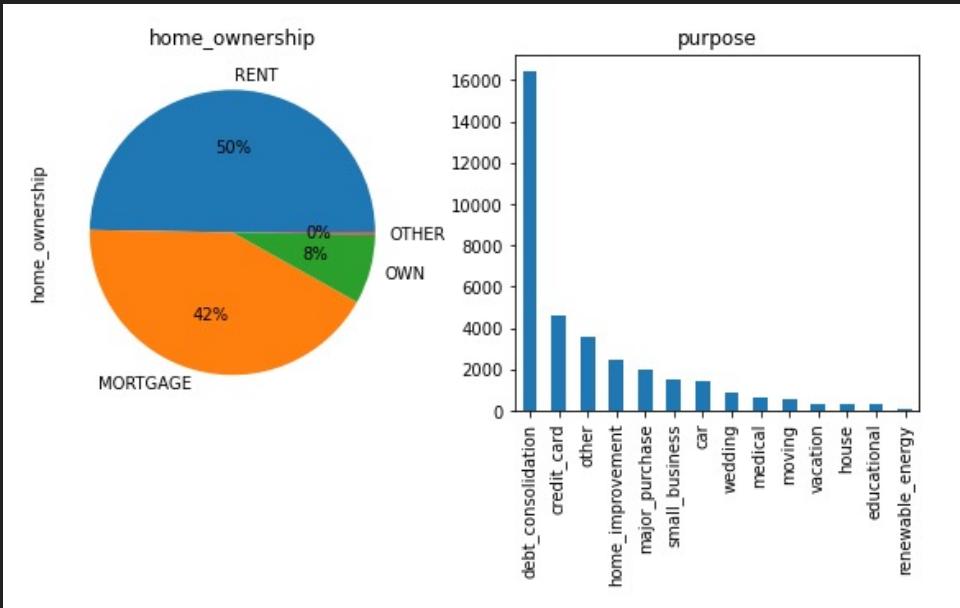


Box Plot for Loan Amount Before Outlier Treatment



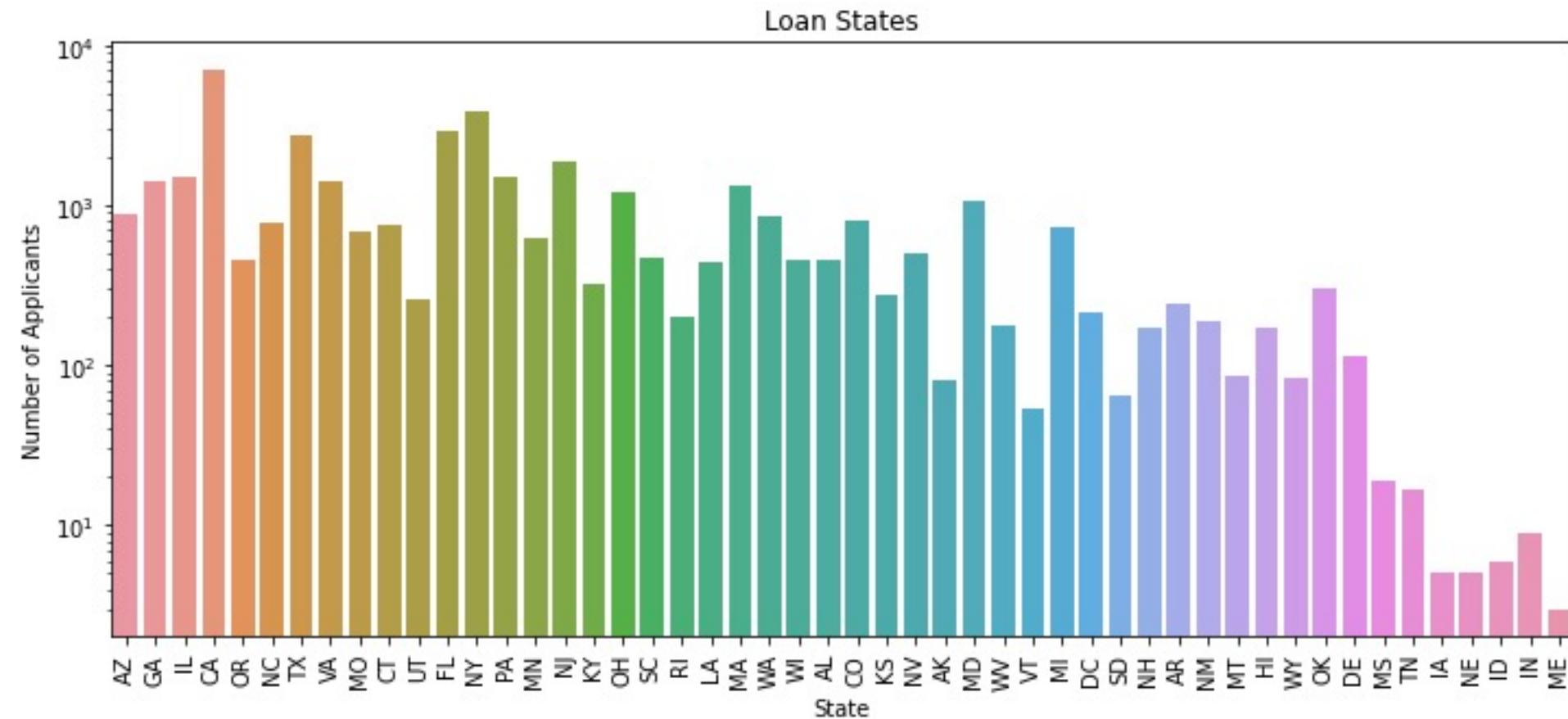
Box Plot for Loan Amount After Outlier Treatment





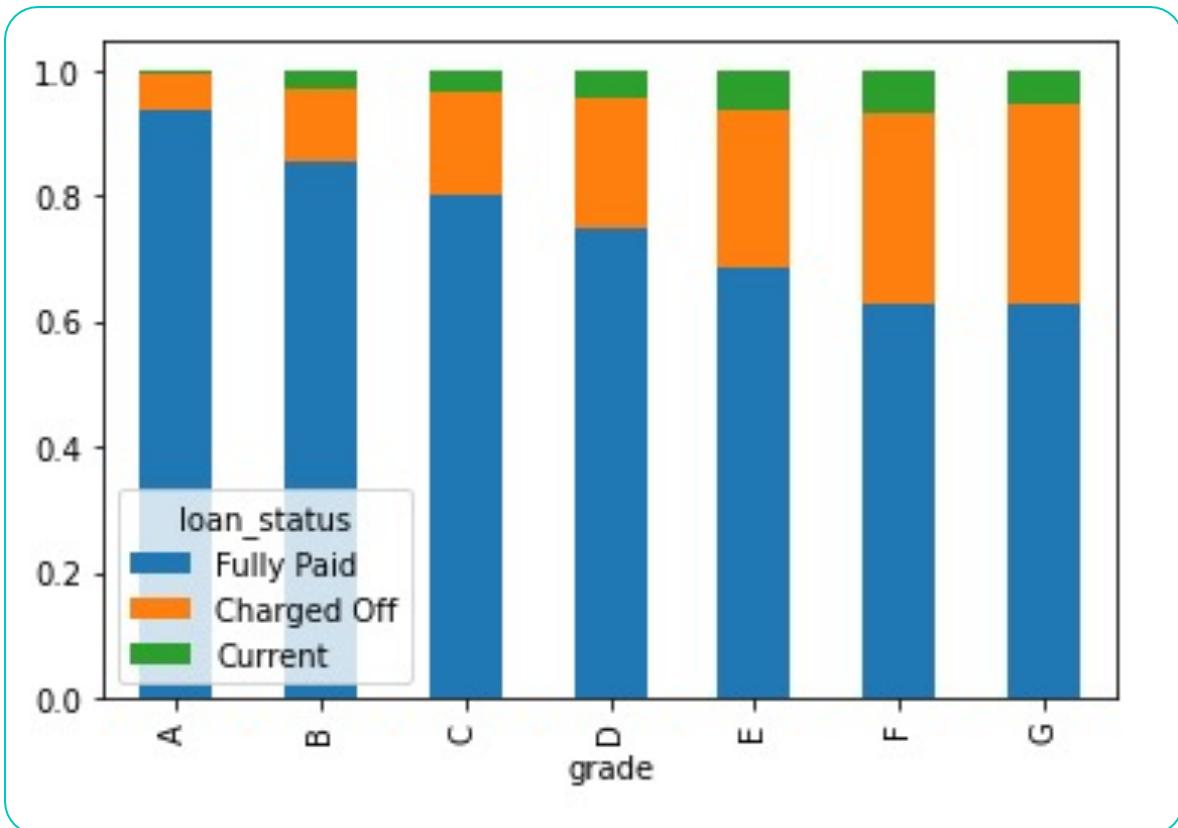
# Home Ownership Purpose Employee Length

- Number of accepted applications decreases with the years of employment indicates the decrease in the need of loans.
- Observed that 10 and above years of experienced people got more loans possibly due to the purchase of new house.
- Customers with rented/mortaged houses are accepted with loans the most and loans are requested mostly for debt consolidation.



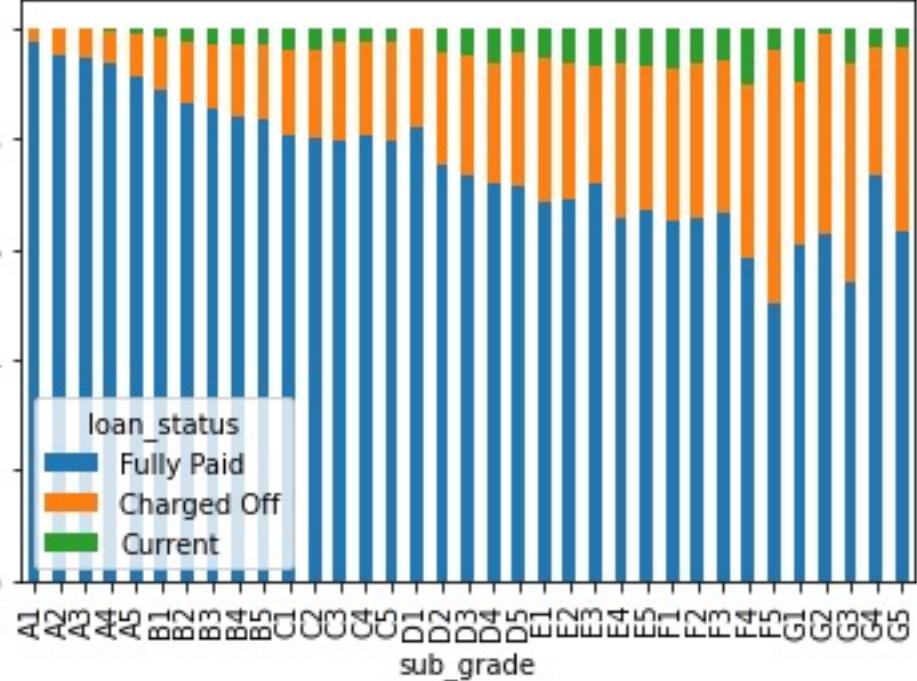
## Loan States

- Observation: From the above plot, we can infer that most loan taking customers are from CA.



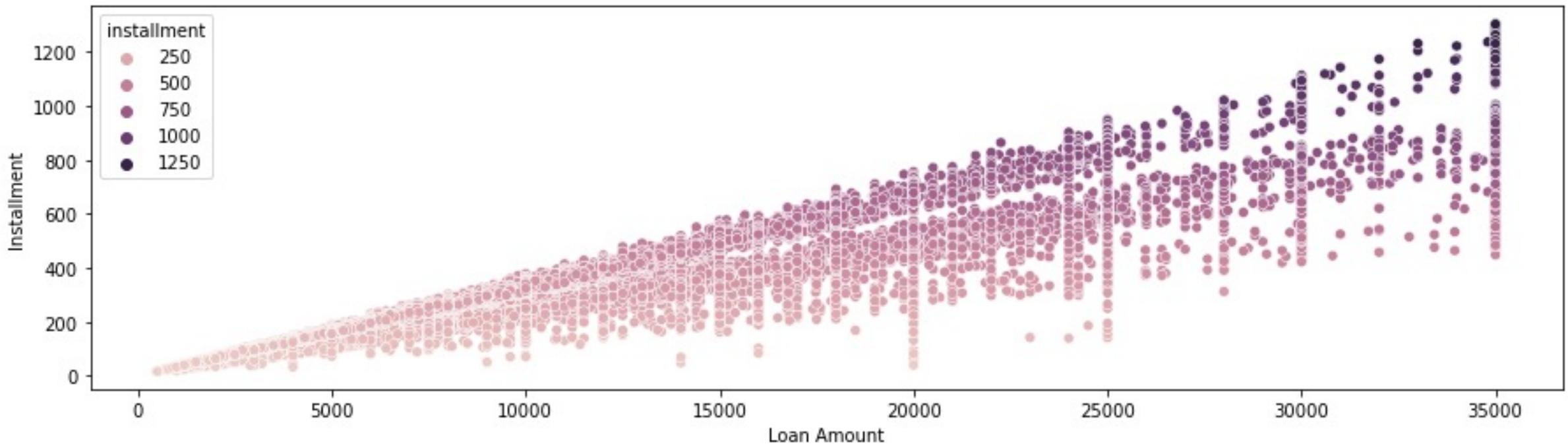
## Bi-variate Analysis

- Observation: From the above plot, we can see that there are more default customers with "B" grade followed by "C", "D", "A", "E", "F", "G"



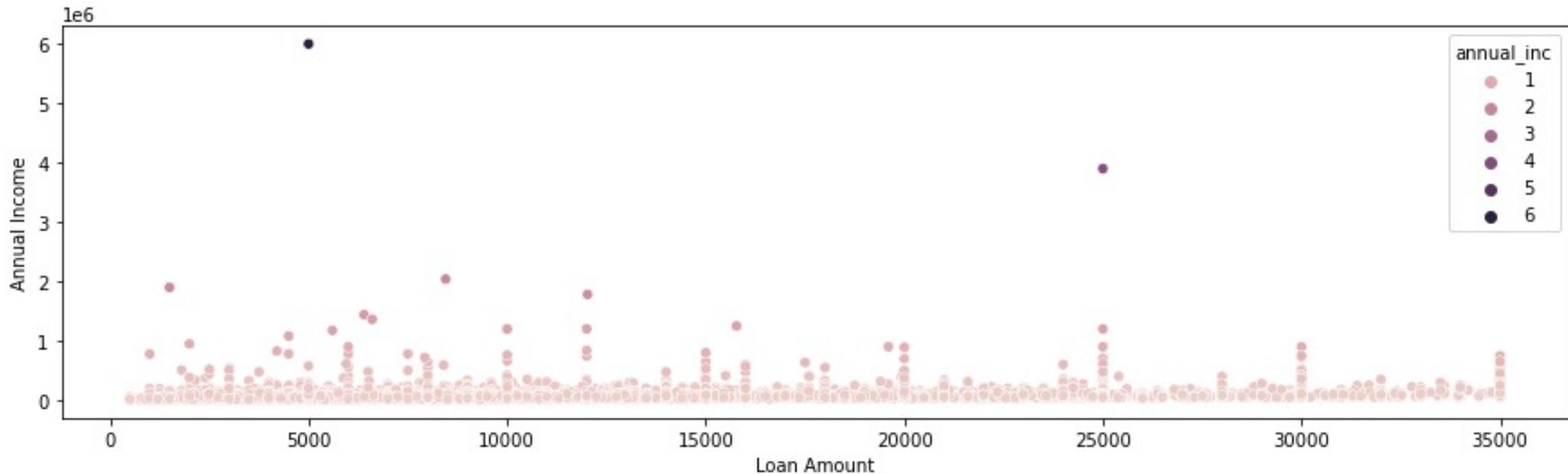
## Bi-variate Analysis

- Observation: From the above plot, it's evident that Charge-off frequency increases with decline in grades



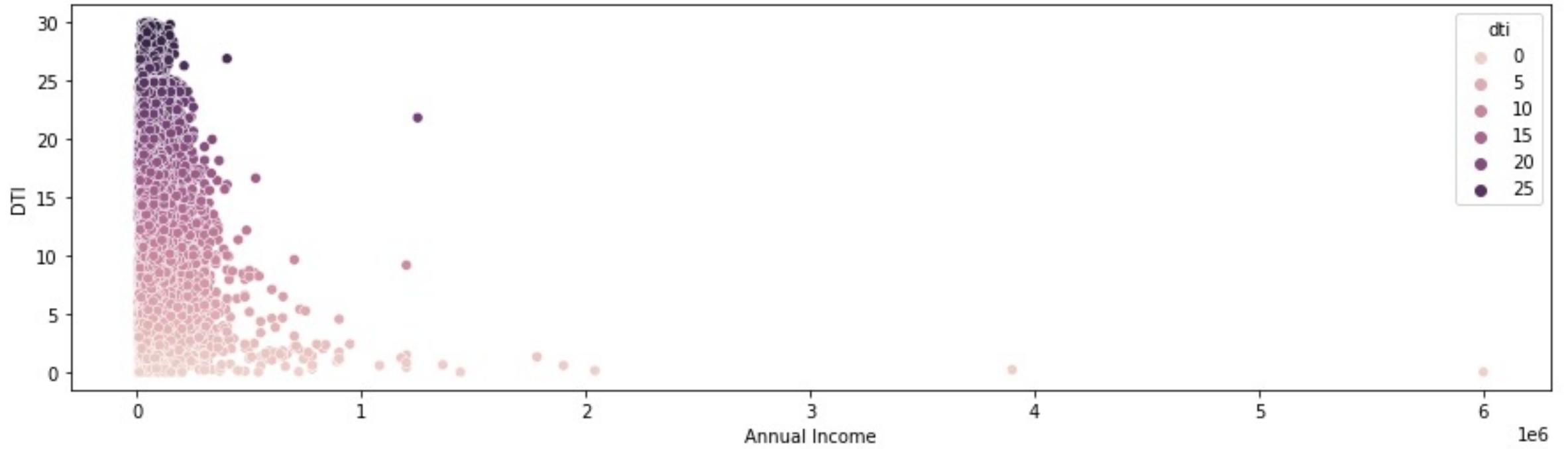
## Bi-Variate Analysis and Multi-Variate Analysis

- Observation: From this plot, we can see that when the loan amount increases , the installment amount also increases along with it which is a positive relation.



## Bi-Variate Analysis and Multi-Variate Analysis

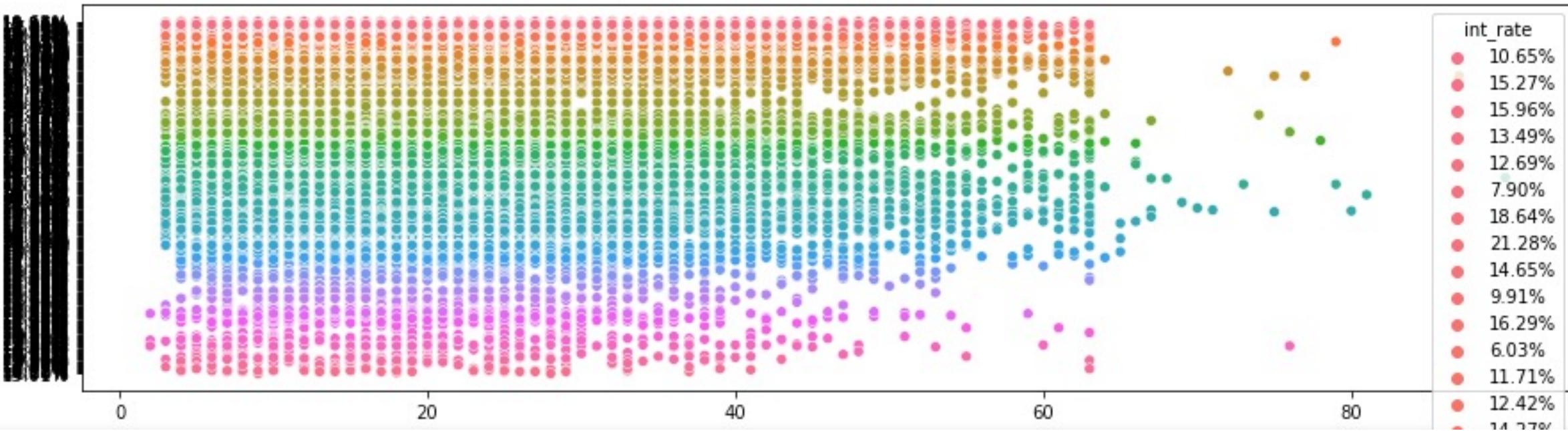
- Observation: From this plot, we can infer that the customer's taking Loan amount is increasing if their annual income is high.



## Bi-Variate Analysis and Multi-Variate Analysis

- Observation: From Above plot we can see if a person annual income increases then his debt ratio(dt) decreases resulting in negative relation

Interest Rate



## Bi-Variate Analysis and Multi-Variate Analysis

- Observation: From the above plot, we can infer that the interest rate decreases then total credit lines increases

# Conclusion

- Below are the parameters which may lead to increase in the default rate.
- Term:
  - It is observed that the charged-off rate increases along with the increase in the term.
  - Suggestion would be to avoid giving long-term loans considering other parameters.
- Annual Income:
  - It is observed that the annual-income is in negative co-relation with the default rate. Lower the annual income, better are the chances of default.
- Interest Rate:
  - It is observed that the default rate increases along with the increase in interest rate.
- Home Ownership
  - It is observed that the default rate is increased when the home ownership is rented.

# Recommendation

- Providing loans to employees with lesser annual income should be avoided.
- Employees with more than a year of experience and less than 10 years of experience can be considered more to avoid defaults.
- Do not increase the interest rate beyond 20%. This leads to more default rate.
- It's better to approve loans only to verified employees.