Lending Club Case study

Group Members:

Sudarshan Govindaraghavan (sudarshan_g@outlook.com)
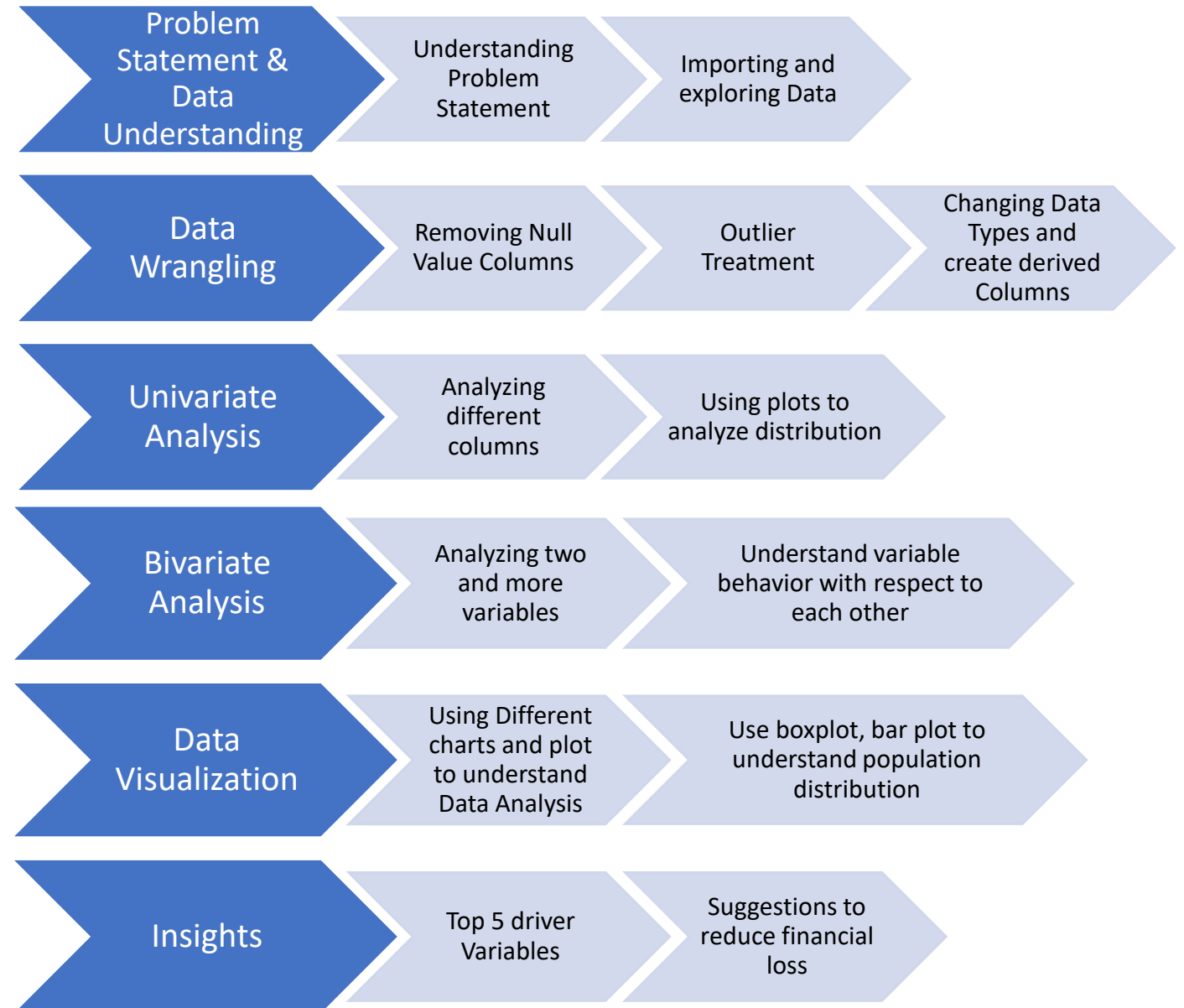
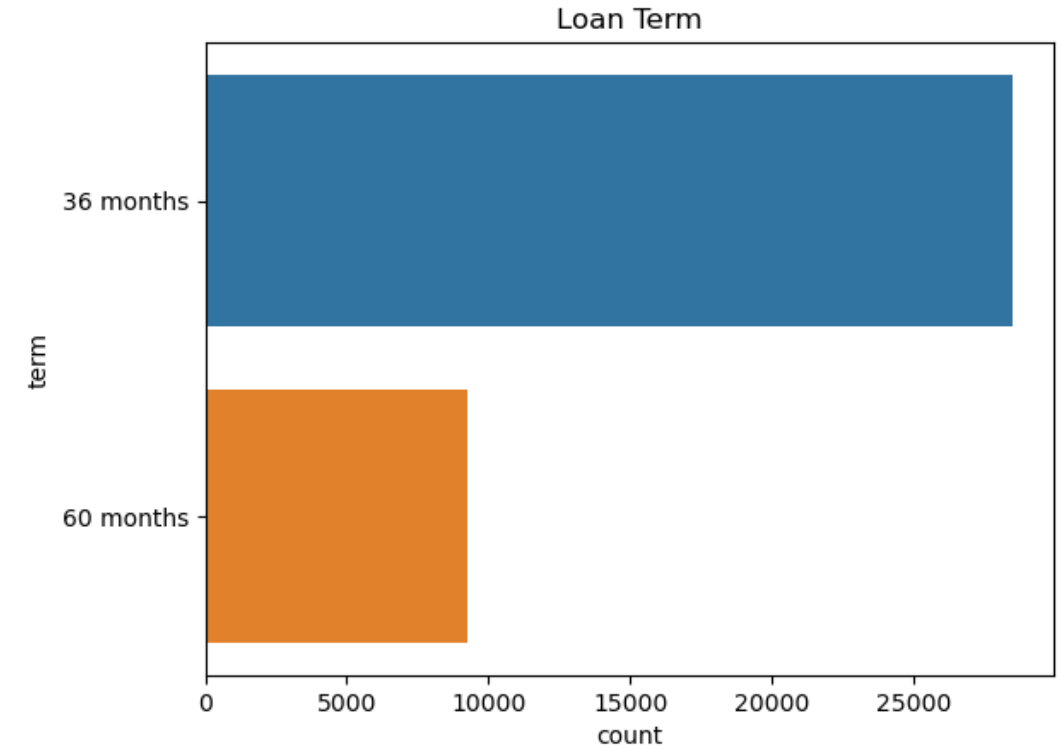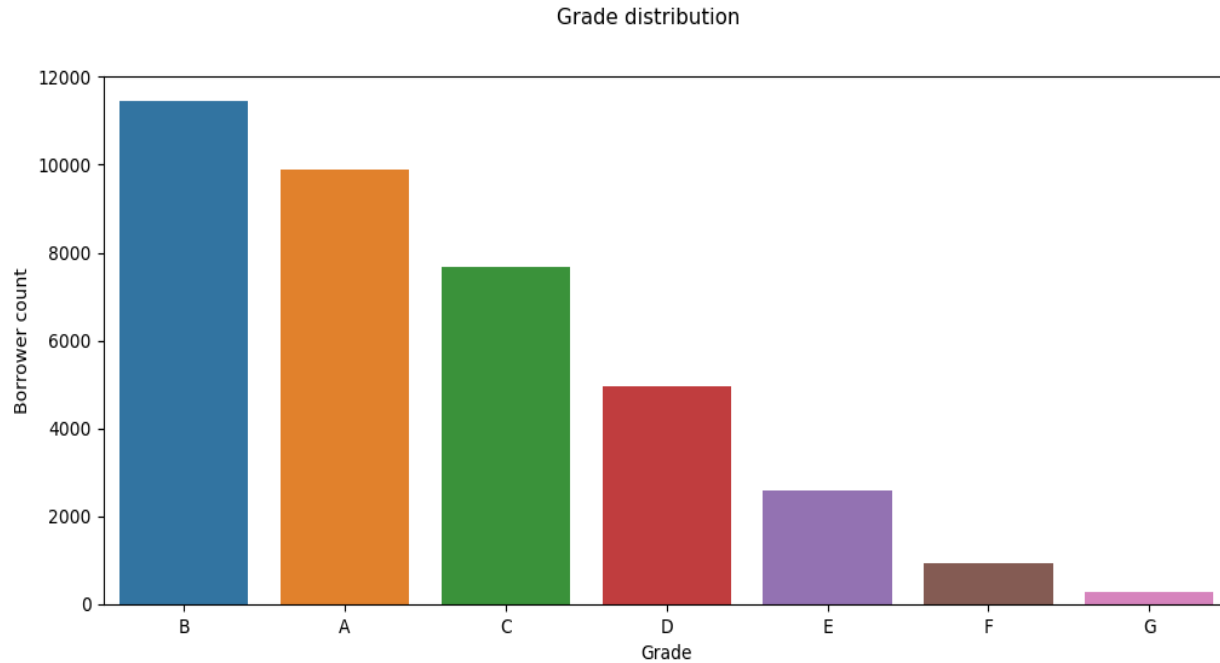**Alpesh Kumar** (kumaralpesh0695@gmail.com)

upGrad

Problem Statement

- Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

- Two **types of risks** are associated with the bank's decision:
  1. If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
  2. If the applicant is **not likely to repay the loan,** i.e., he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

- The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e., the variables which are strong indicators of default.

Problem Solving Methodology

| Problem Statement & Data Understanding | Understanding Problem Statement | Importing and exploring Data | |
Data Wrangling	Removing Null Value Columns	Outlier Treatment	Changing Data Types and create derived Columns
Univariate Analysis	Analyzing different columns	Using plots to analyze distribution	
Bivariate Analysis	Analyzing two and more variables	Understand variable behavior with respect to each other	
Data Visualization	Using Different charts and plot to understand Data Analysis	Use boxplot, bar plot to understand population distribution	
Insights	Top 5 driver Variables	Suggestions to reduce financial loss	

Analysis Insights – Grade & Term

- Borrowers with Grade B has maximum number of loans, followed by Grade A and C
- 75%+ borrowers has a tenure of 36 months

Analysis Insights – Loan Status and Loan amount
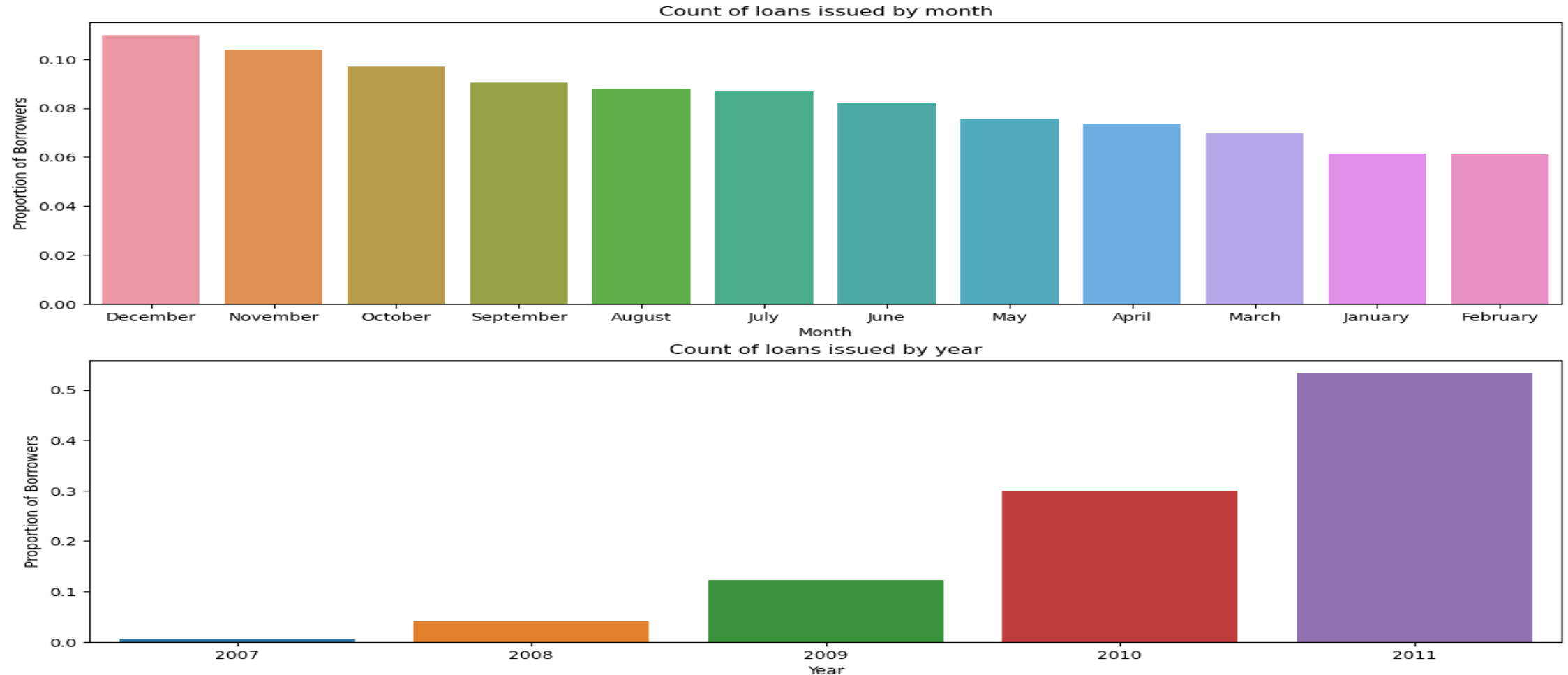

Loan Status distribution
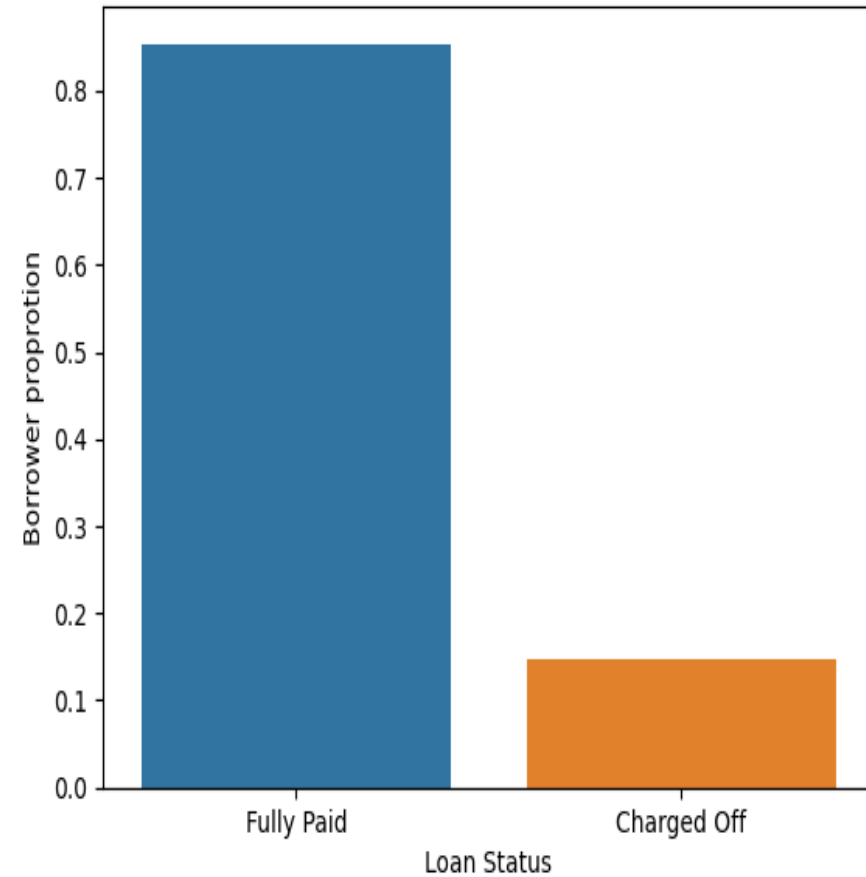
Loan amount distribution
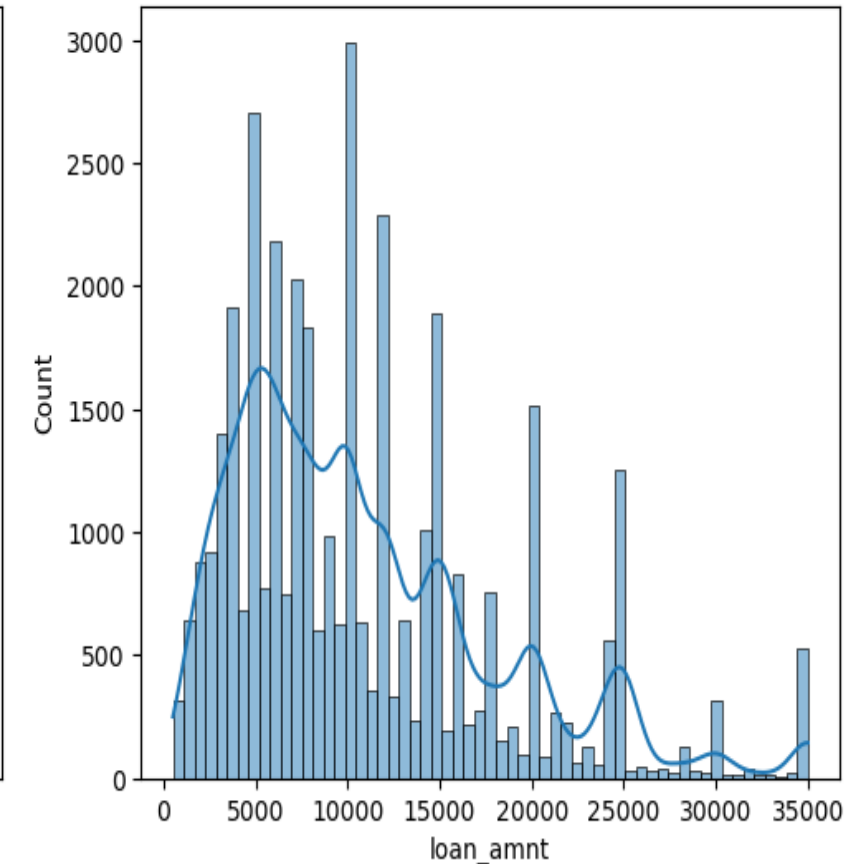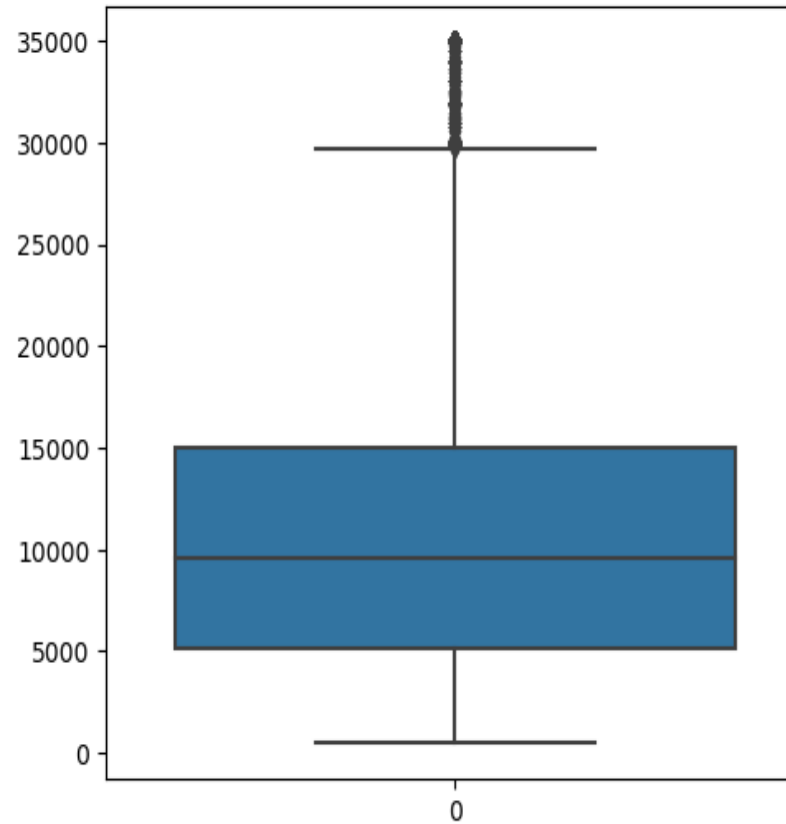
- 85% population has Fully paid the loan amount
- Most of the loan amounts are under 15k, however A spike can be observed in higher loan amount for round off numbers, for e.g., 25k, 30k, 35k etc.

# Analysis Insights – Loan Issued month and year


Loan distribution by issue date

- More loans are given towards the year end compared to first half of the year
- Lending Club has grown exponentially over the period of 2007 to 2011

# Analysis Insights – Verification Status and Purpose

- 43%+ of the borrowers were not verified by the Lending Club. Verification reduce the possibility of a default.
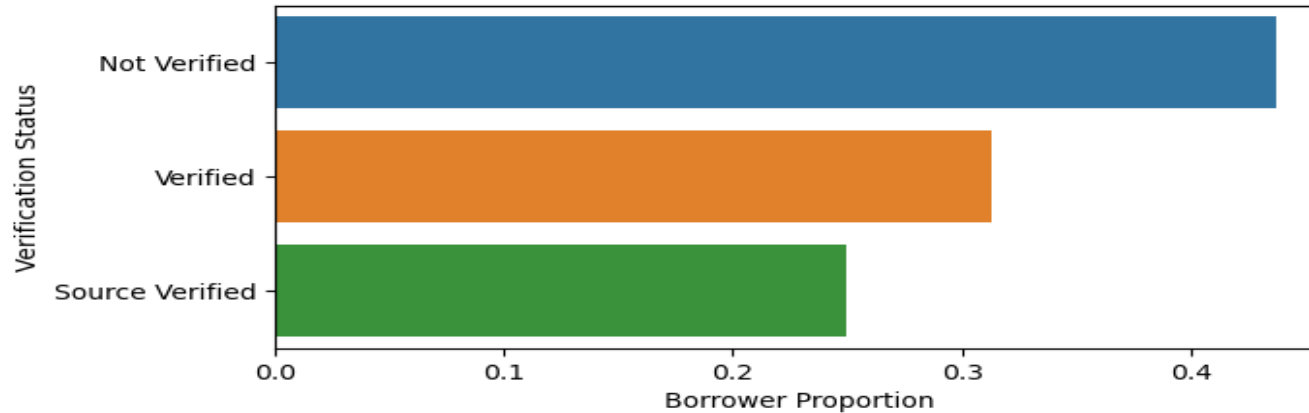
- Most loans are borrowed for the purpose of debt Consolidation, follower by credit card.
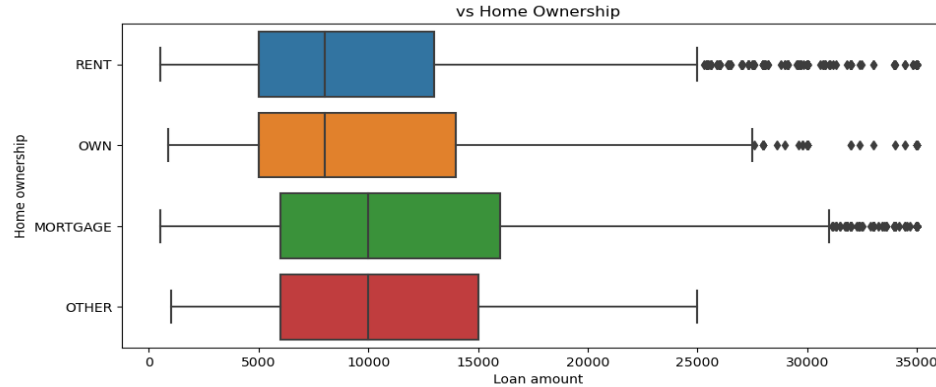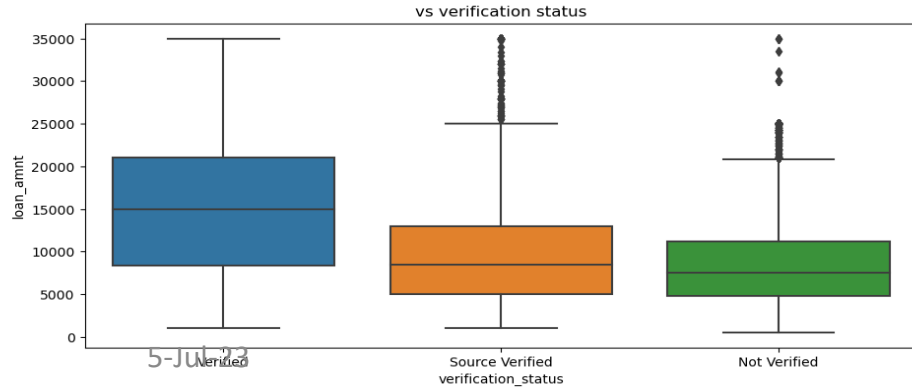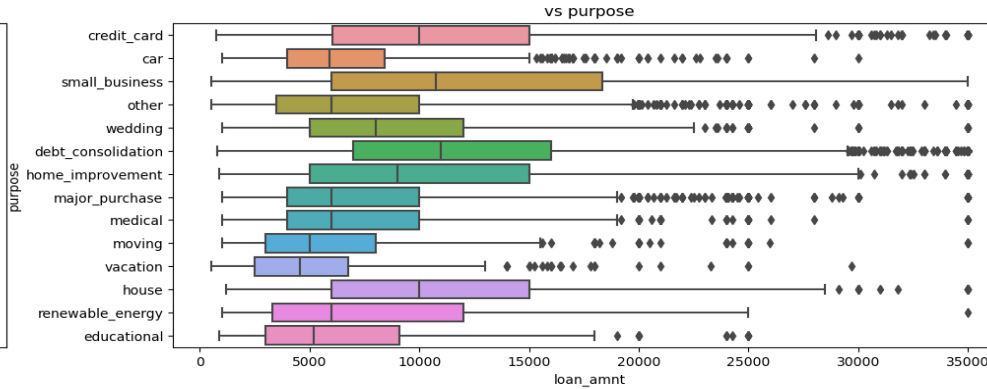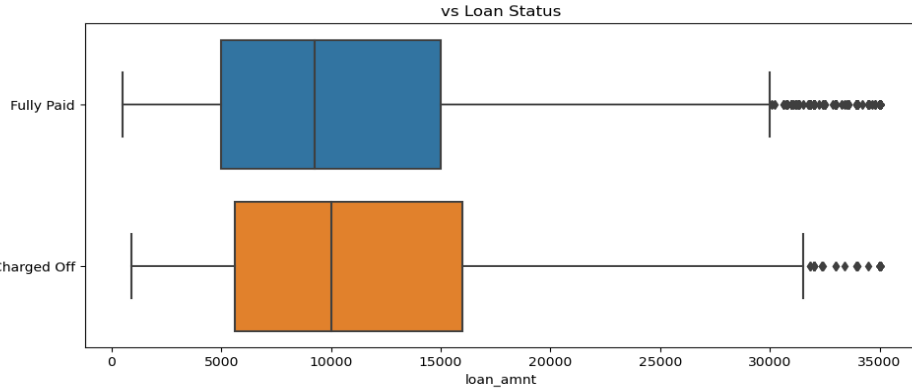 Other major purpose: home improvement, purchase, small business, car
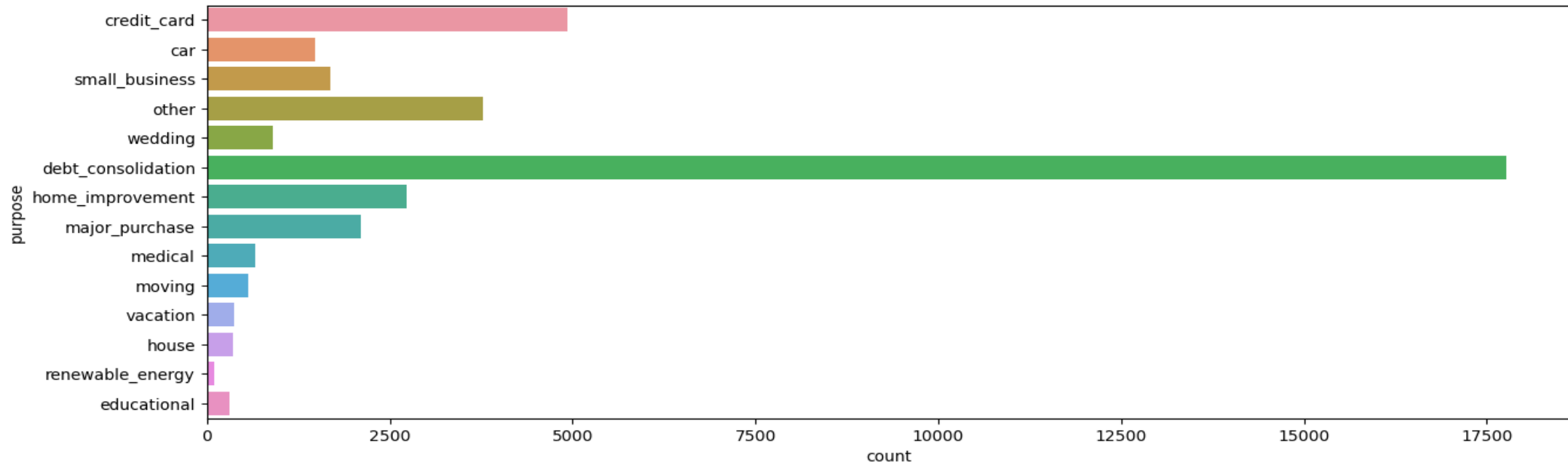
Bivariate Analysis Insights

grade distribution
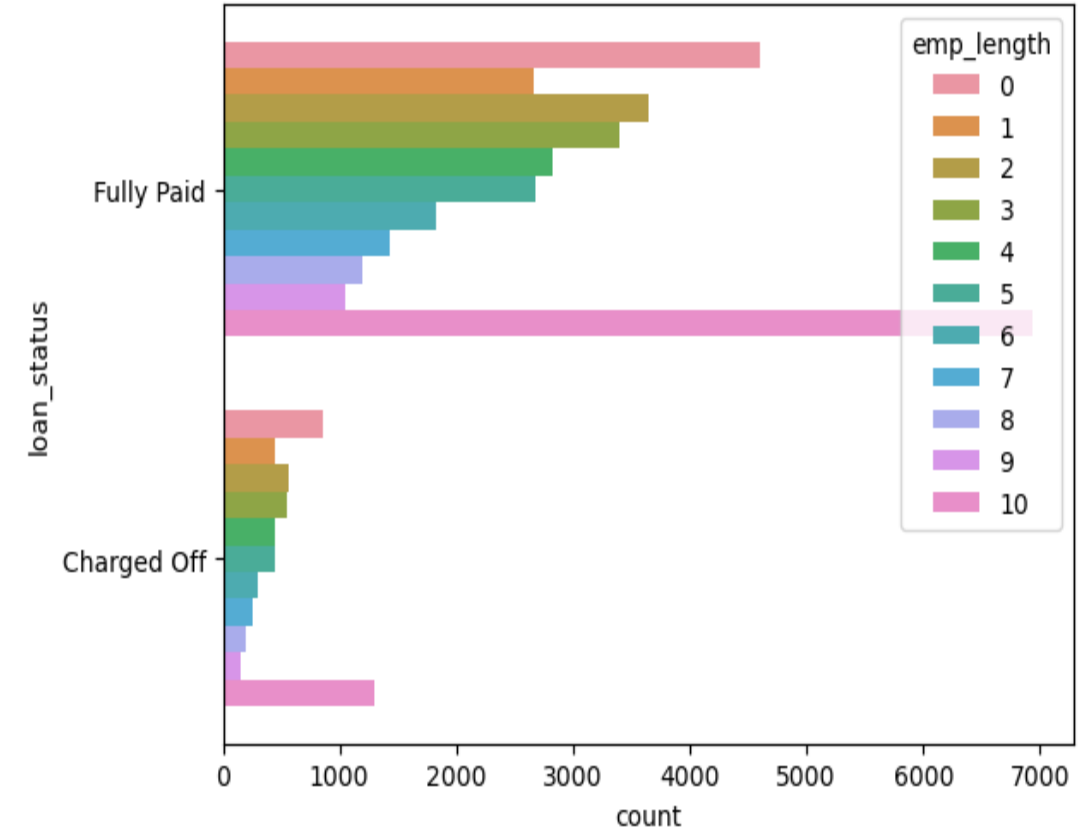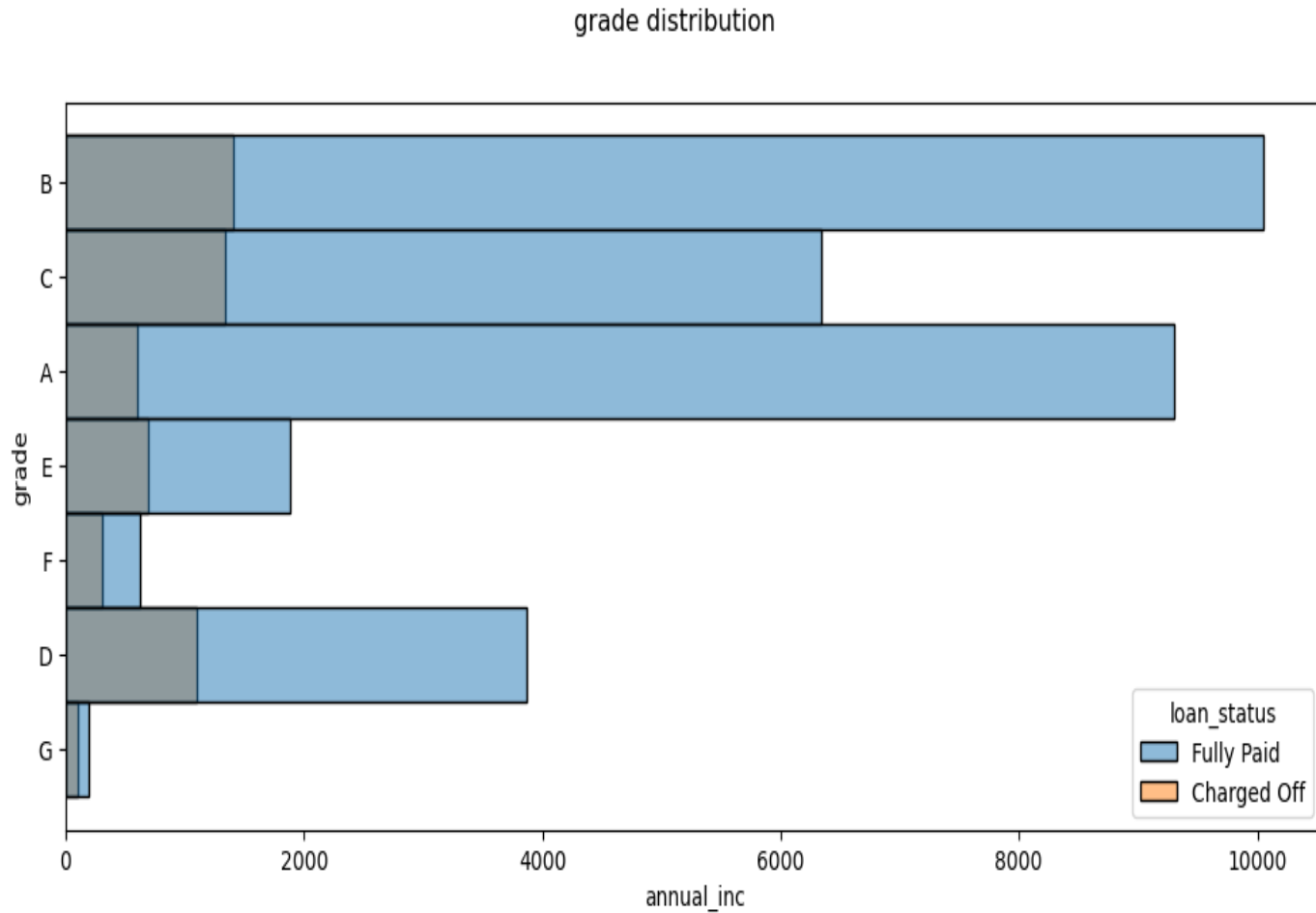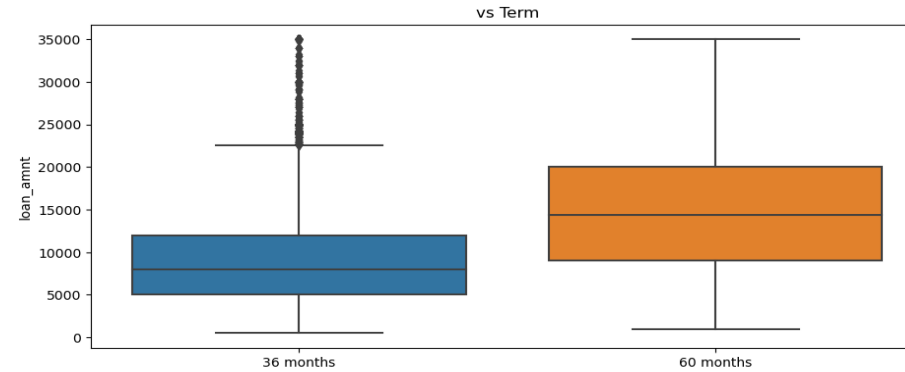
- Grade B has highest Charged Off population despite being the highest Annual Inc
- Most borrowers of Lending Club has employment length of 10 or more years

Analysis Insights – Segmented Univariate Analysis

Loan amount (Segmented Analysis)
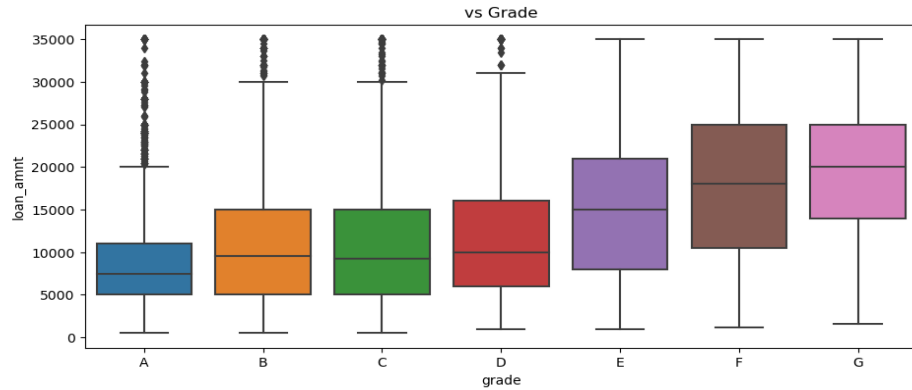
- Grade F has max IQR range

- Higher loan amounts for 60 months term

- 'Charged off' median & 75th percentile is higher than 'Fully Paid'

- All categories has max amount loans

- 'Not Verified' also has some high amount loans

- Mortgage has highest spread of loan amount

Analysis Insights – Segmented Univariate Analysis
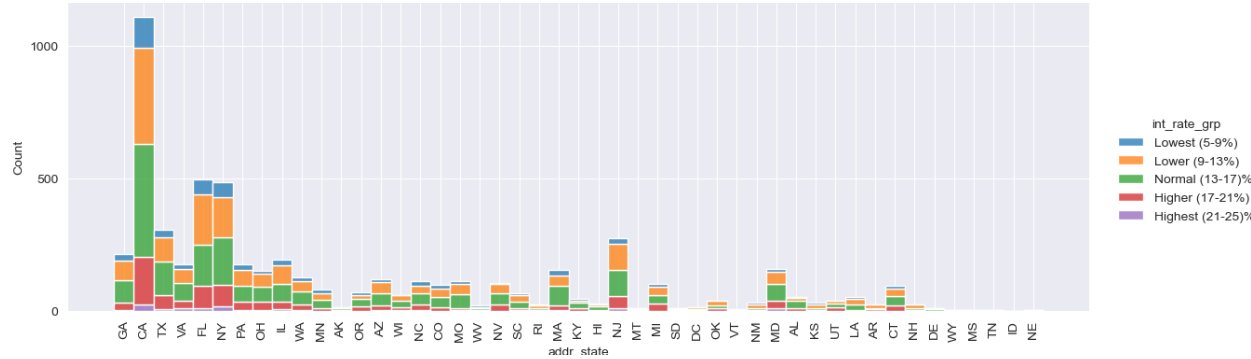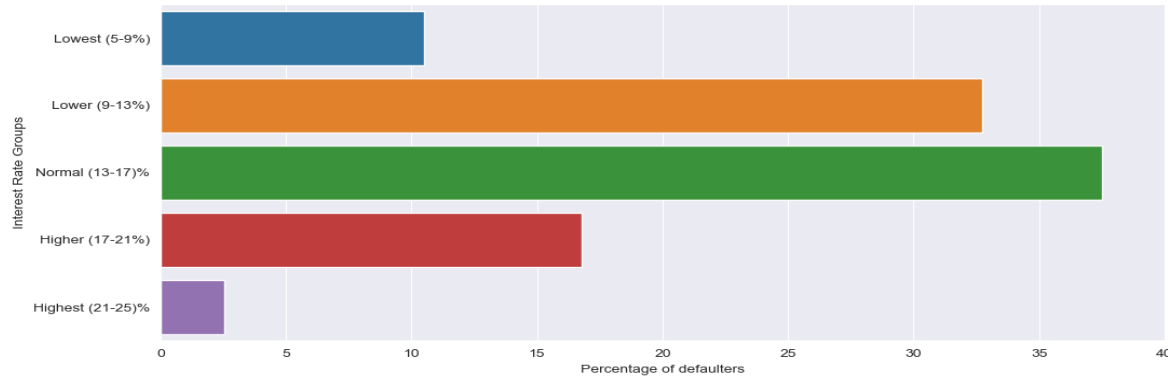

Defaulter's state by Interest Rate Groups

Percentage of defaulters in each interest rate group

Annual Income distribution for defaulters
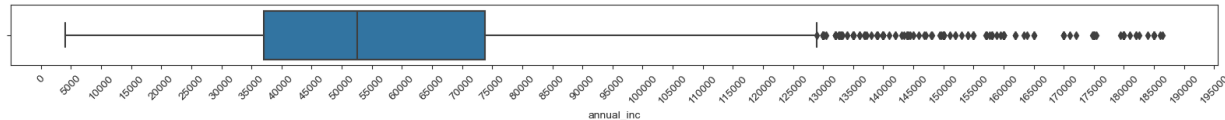
Percentage of defaulters by borrower's state
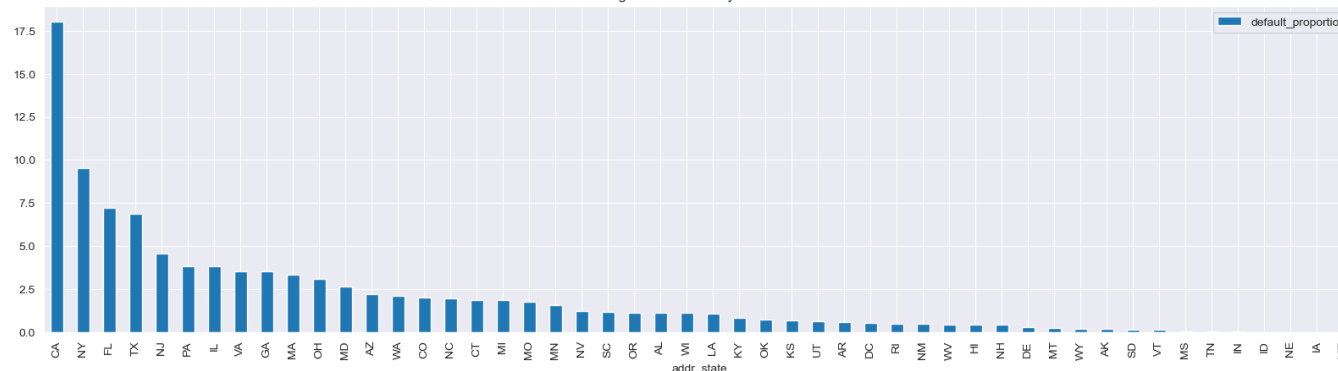
- `CA`, `FL` and `NY` have a higher proportion of borrowers with high interest rate.

- Majority of the defaulters are within `Lower (9-13%)` and `Normal (13-17%)` interest groups.

- `IA`,`IN` and `ME` had no defaulters. Whereas `CA`, `NY`, `FL` and `TX` had the highest proportion of defaulters. Borrowers from `WY` and `AK` and `UT` are likely to default on higher loan amounts but the number of defaulters is less compared to other states.

- Majority of the defaulters have annual income between 37k and 73k.

Bivariate Analysis (Demographic)

Borrower's state by Loan Status
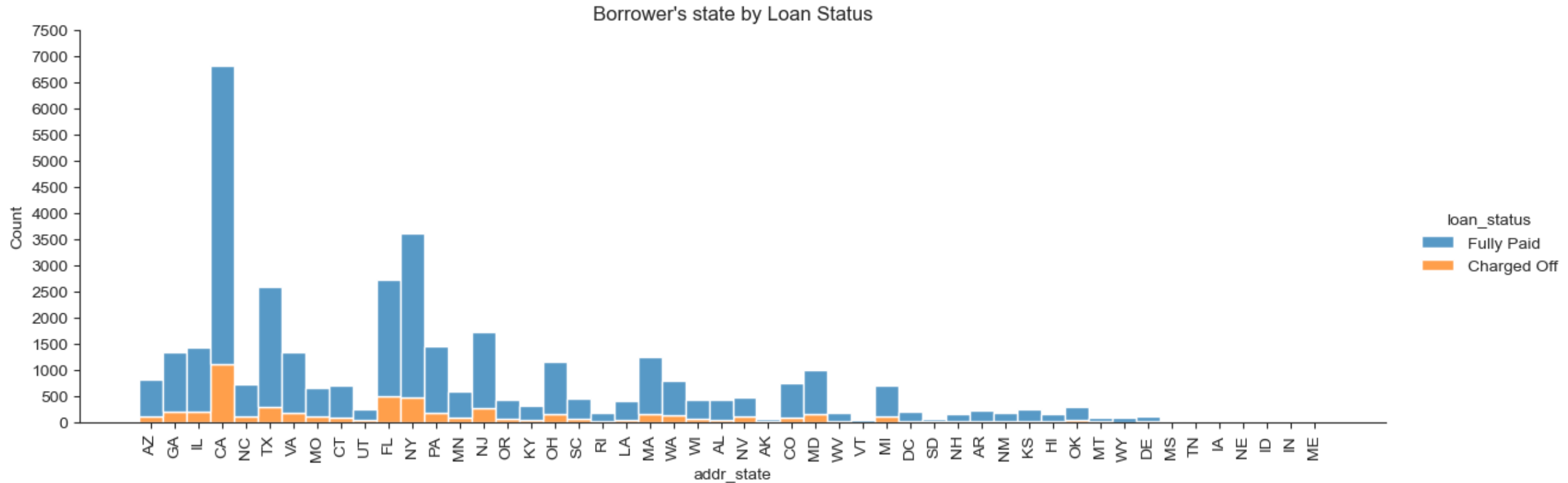
- Borrowers in CA, FL and NY are most likely to default on their loans compared to other states

Analysis Insights –Bivariate Analysis

Borrower's by verification_status
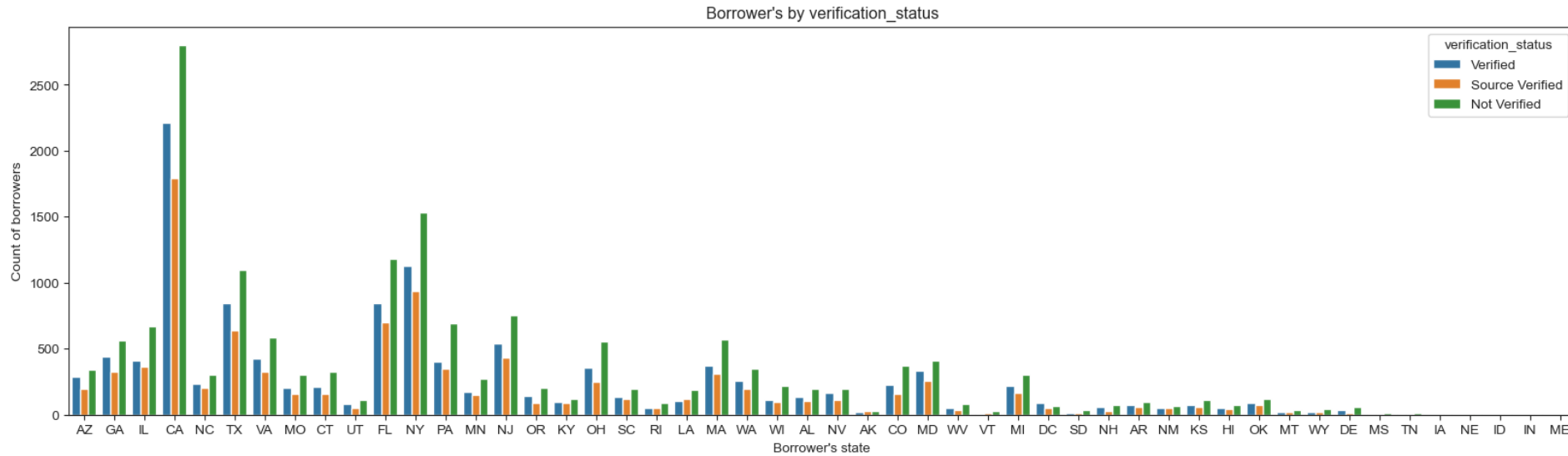
`CA`, `NY`, `FL`, `TX`, `NJ` account for almost half (~ 44.70%) of the not verified borrower count.

Majority of the defaulters are also from these 4 states - `CA`, `NY`, `FL`, `TX`.

Borrower Proportion by Grade & Loan Status
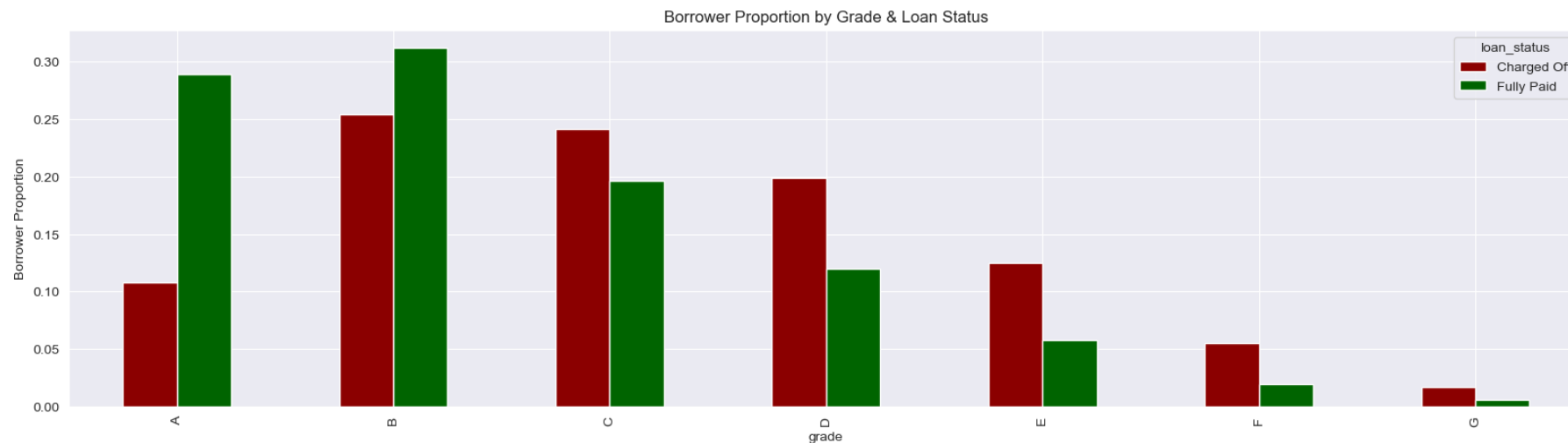
Out of all the borrowers, A large proportion of defaulters are within grades `B`, `C` and `D` compared to other grades.

Analysis – Multivariate Analysis
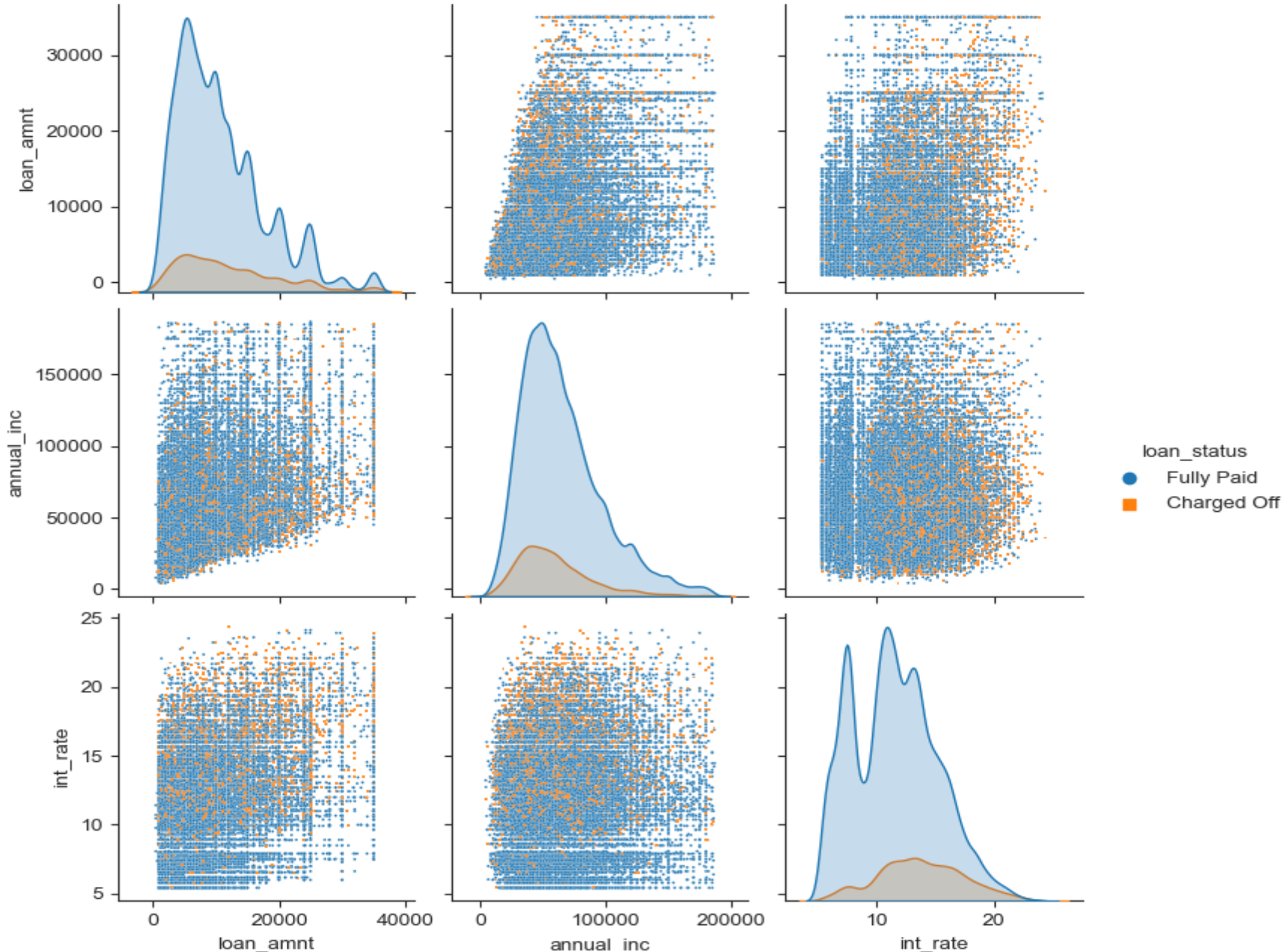
- Analyze Population spread across different variables:
  1. Loan Amount
  2. Annual Incentive
  3. Interest Rate
  4. Loan Status

- Small positive trend observed between increasing Annual Inc and Loan Amount
- High Interest Rates accounts for high Charged Off rate
- Interest Rate increase with loan amount

Conclusion

- Maximum Customers borrow loan for the purpose of Debt Consolidation followed by Credit Card

- Nearly 43% population is not verified by the Lenders Club.

- One in every four person with loan tenure on 60 months defaults on the loan.

Top Driver Variables for loan status are:

✓ Grade: Higher grade leads to higher interest rates & higher loan amounts which increases the probability of a default.

✓ Loan Amount: Defaulters had a higher median loan amount compared to others.

✓ Interest rate: Defaulters were paying a slightly higher interest rate on their loan.

✓ Loan Purpose: Borrowers who mention their purpose as 'Small Business', 'Debt Consolidation' and 'Credit Card' are likely to default more compared to others.

✓ Annual Inc: Borrowers with annual income between 37k and 75k are likely to default more on loan amounts

✓ Loan Tenure: Borrowers with a loan term of 60 months default more compared to those with 36 months

✓ Borrowers State: Borrowers in CA, FL and NY are most likely to default on their loans compared to other states

Recommendations

- Borrowers should be verified by the lending club for the following states - `CA`, `NY` and `FL` as borrowers from these places are likely to default compared to other states.

- Lending Club should evaluate the borrowers for purpose like 'Debt Consolidation', 'Credit_card', 'small_business' and ensure they are verified to reduce default risk.

- Ensure that borrowers with loan grade `C` and `D` are verified and have lower interest rates to limit default risk.

- Minimize lending at a higher term as borrowers with `60 month` defaulted more compared to `36 months`.

- Minimize lending loans with grades higher than `C` or higher loan amounts to people with income in 36k to 73k.

- Limit the interest rate to lower amounts in `CA`, `NY`,`FL`, `TX` to reduce default risk.

Appendix

- Python Notebook has findings at the bottom of each Analysis and the code cells are commented with the approach used

- Data Cleaning is done based on certain assumptions while keeping in mind the different Data Drivers and Business Drivers

- New columns are added for ease of calculations in the refined dataset

- Univariate Analysis is revisited and carried out on different parameters as they are identified in other Analysis Techniques

- Some Analysis like 'Correlation Matrix' are tested in the notebook for understanding how different variable affect each other.

- PPT does not contain all the views, but only the ones that are necessary from Business perspective to derive Actionable Insights

Github Link: https://github.com/sudarshan-g/AIML_C53_Lending_Club