LENDING CLUB CASE STUDY

USING EXPLORATORY DATA ANALYSIS

By: Divyansh Saxena & Manchikanti Venkat Sreeni

PROBLEM STATEMENT

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

Aim

To minimize credit loss for Lending Club, by identification of risky loan applicants

Analysis

Analyze the Dataset and to understand the driving factors behind loan default

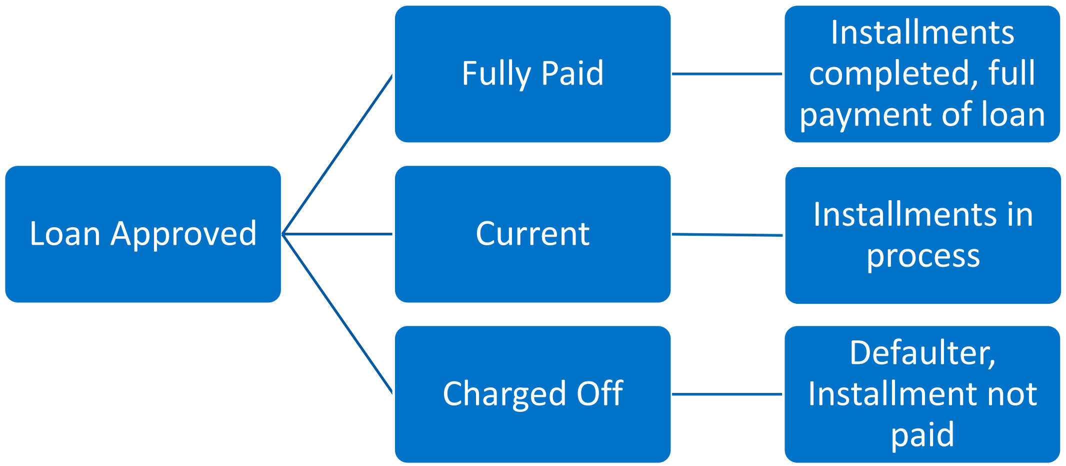
Outcome Expected

Identification of risky applicants using EDA

BUSINESS SCENARIO AND RISK INVOLVED

BUSINESS SCENARIO

- When the company receives a loan application, the company can either accept the loan or reject it based on the applicant's profile.



RISKS ASSOCIATED

Loss of Business

- Not approving loan, If the applicant is likely to repay the loan

Financial Loss

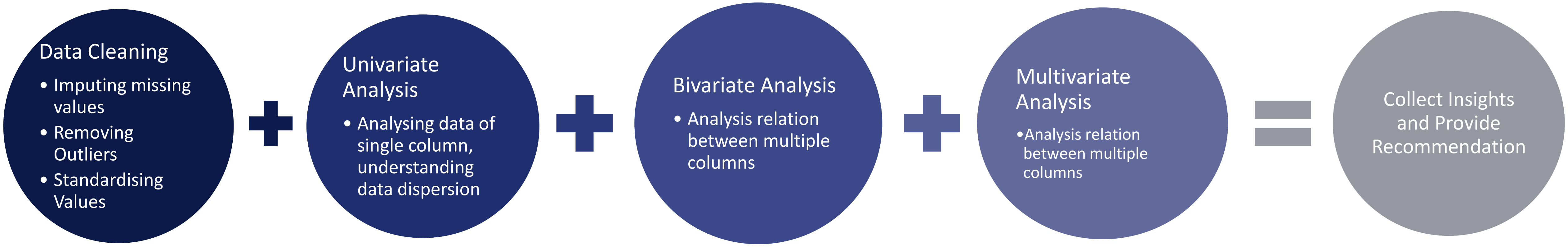
- Approving loan if the applicant is not likely to repay the loan

EXPLORATORY DATA ANALYSIS

An approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.

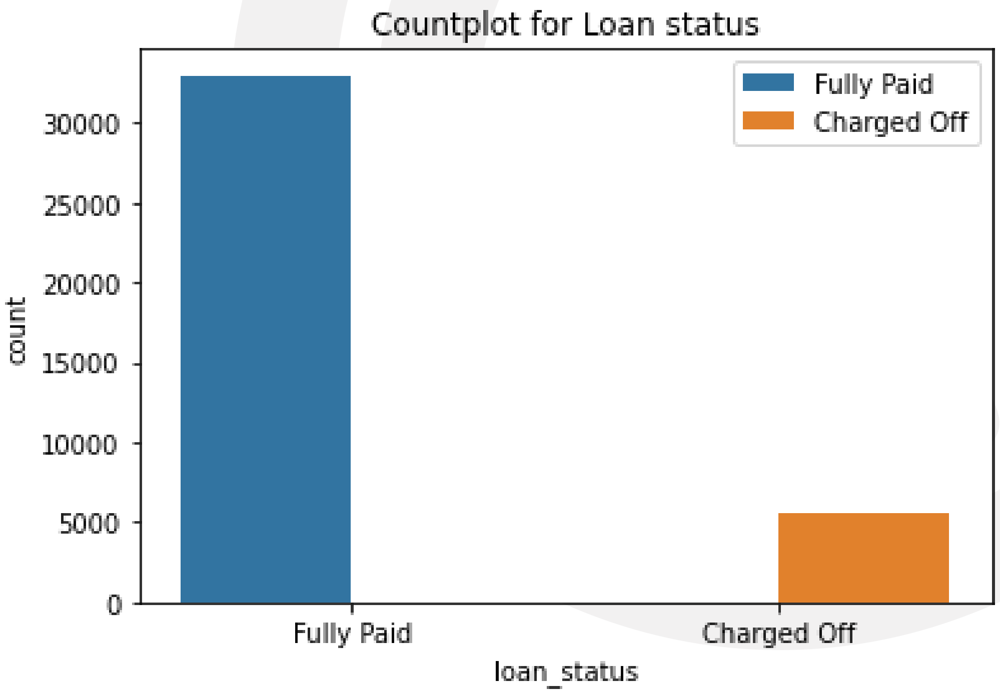
EDA APPROACH

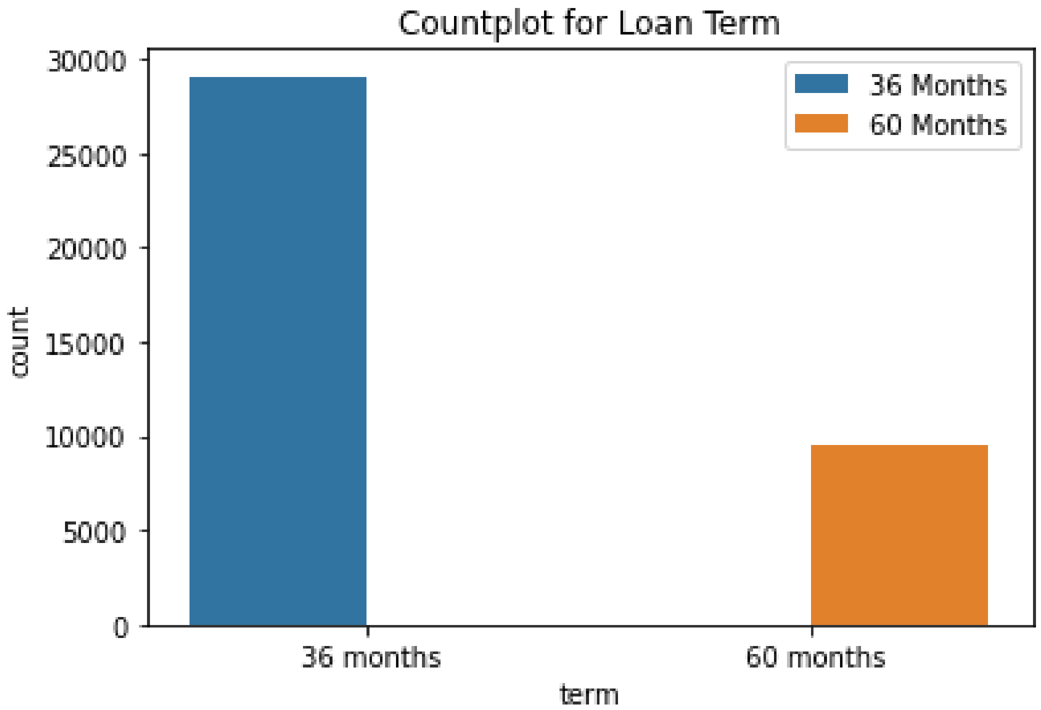
- Data Cleaning: Fixing data, removing outliers, standardising values and imputing missing values
- Univariate Analysis : Statistical analysis technique that focuses on examining and describing a single variable.
- Bivariate Analysis: Analysis is to determine whether there is a relationship between the two variables and to understand the nature and strength of that relationship.
- Multivariate Analysis: Analysis to uncover complex patterns, structures, and trends in the data by considering the joint behavior of multiple variables.
- Using EDA to gain understanding of the data, identify patterns, trends, and relationships, and generate insights that can guide further analysis or decision-making.



ANALYSIS OF LOAN STATUS AND TERM

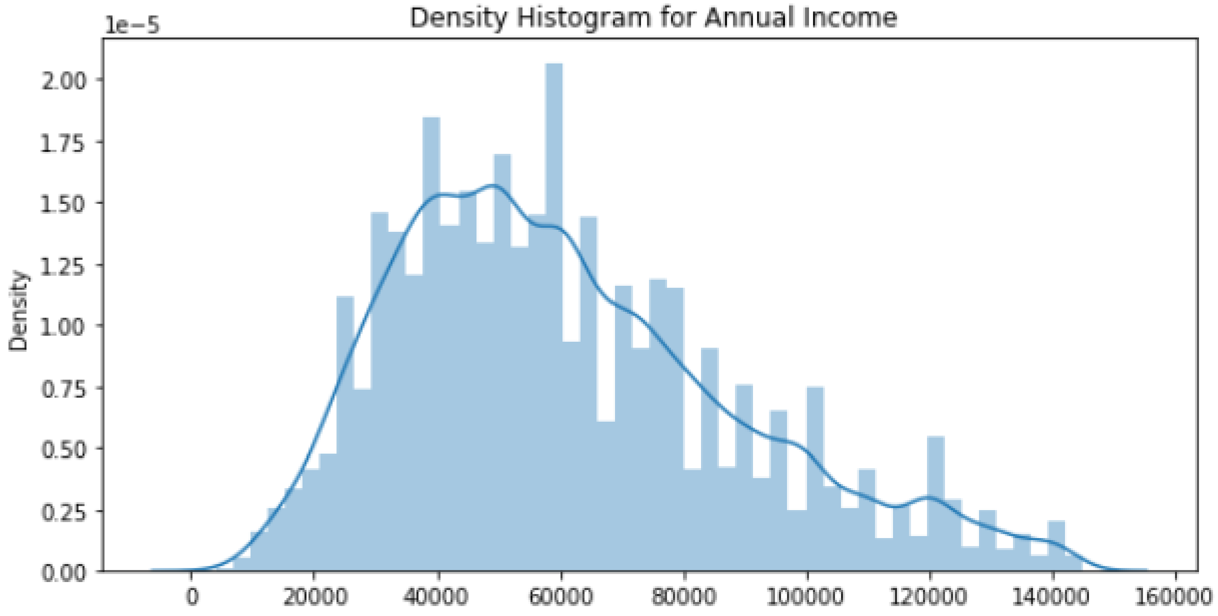
- Majority of borrowers completed the loan in due time.
- Percent of Defaulted Loans – 14.5%
- Percent of Non-defaulted Loans – 85.41%
- Large amount of customers opt for short term loans
- Percentage of short term loans – 75.4%
- Percentage of long term loans – 24.5%

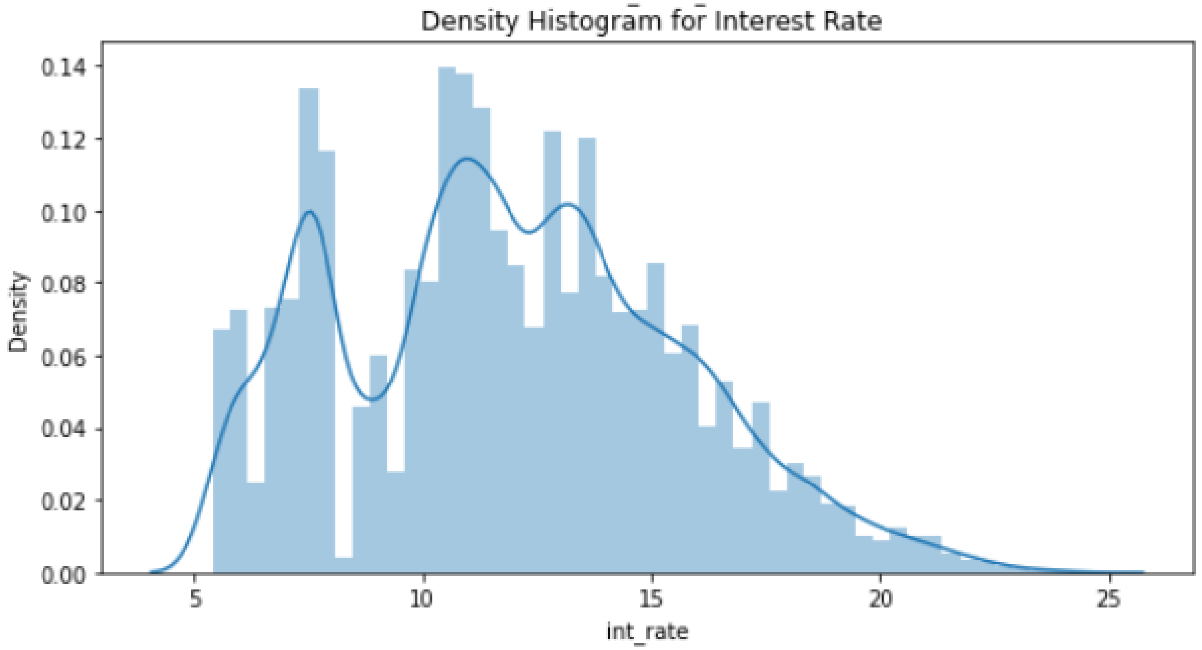




ANALYSIS OF ANNUAL INCOME AND INTEREST RATE

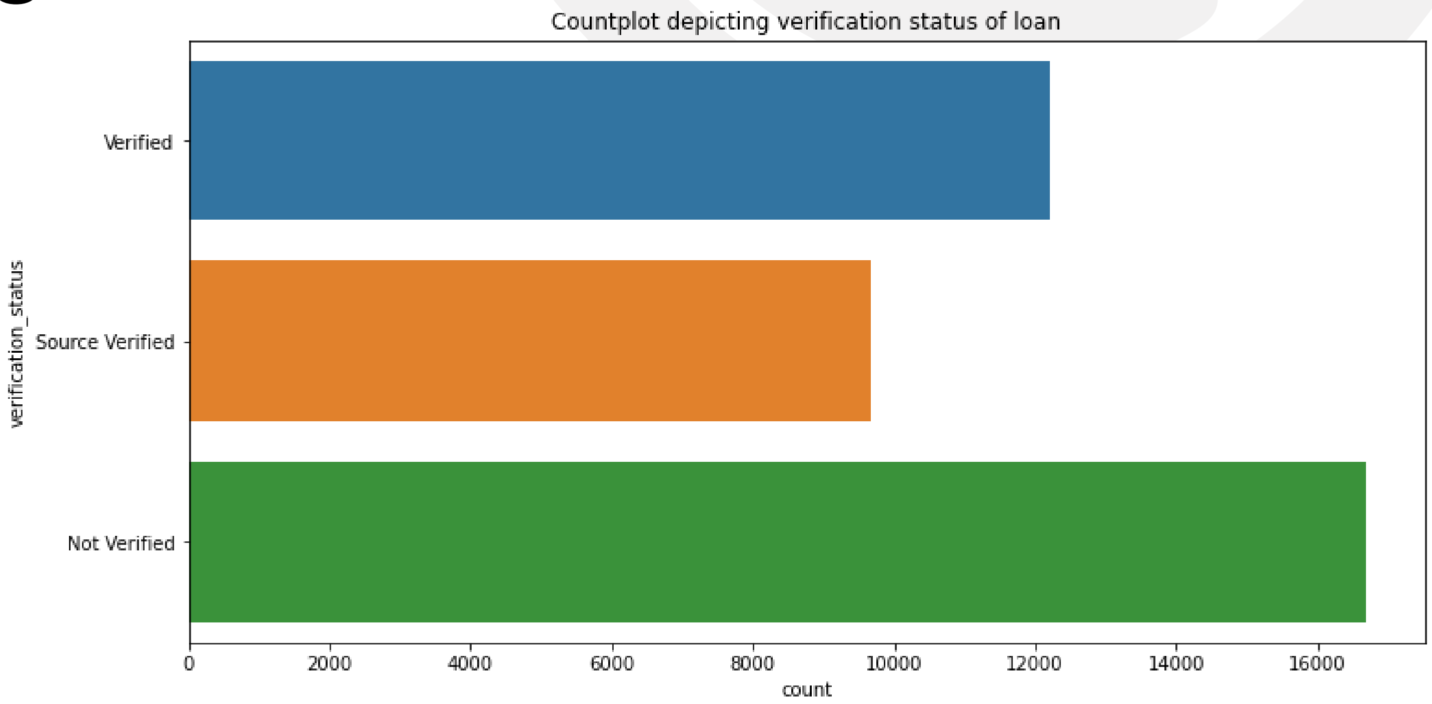
- Annual income of customer lies between 4K to 6K.
- With increase in salary, customer share decreases, indicating that smaller percent of customers have higher salaries
- For Interest Rate, density plot indicates that good amount of loans are provided at loan percent between 6%-8%
- Also, interest rate between 10%-15% loans have high density, indicating that good amount of loans are approved in this range.





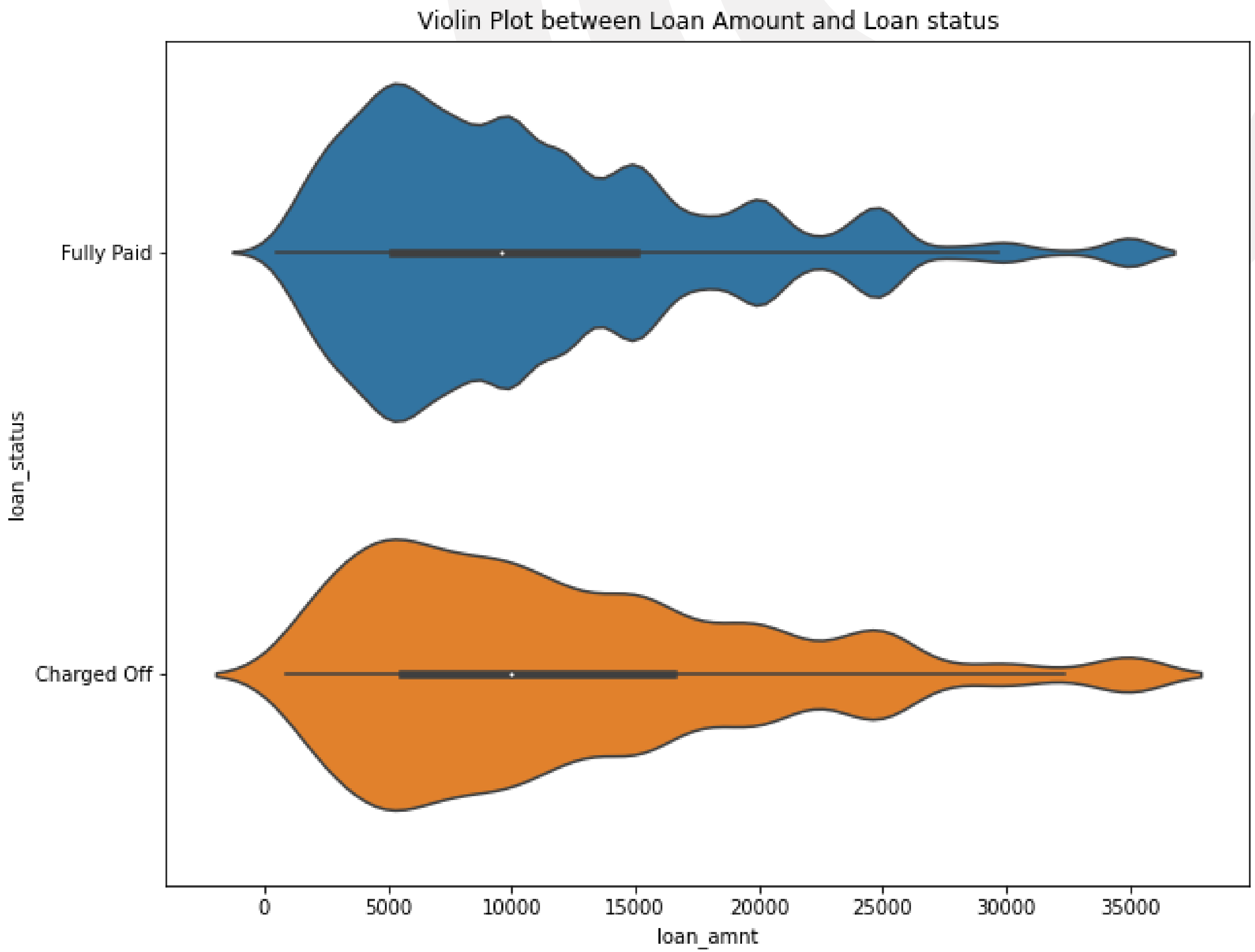
ANALYSIS OF VERIFICATION STATUS

- Large number of sanctioned loans are not verified.
- This can cause increase in loan defaults.
- It's advised to verify customers before approving any loans to minimize credit loss.



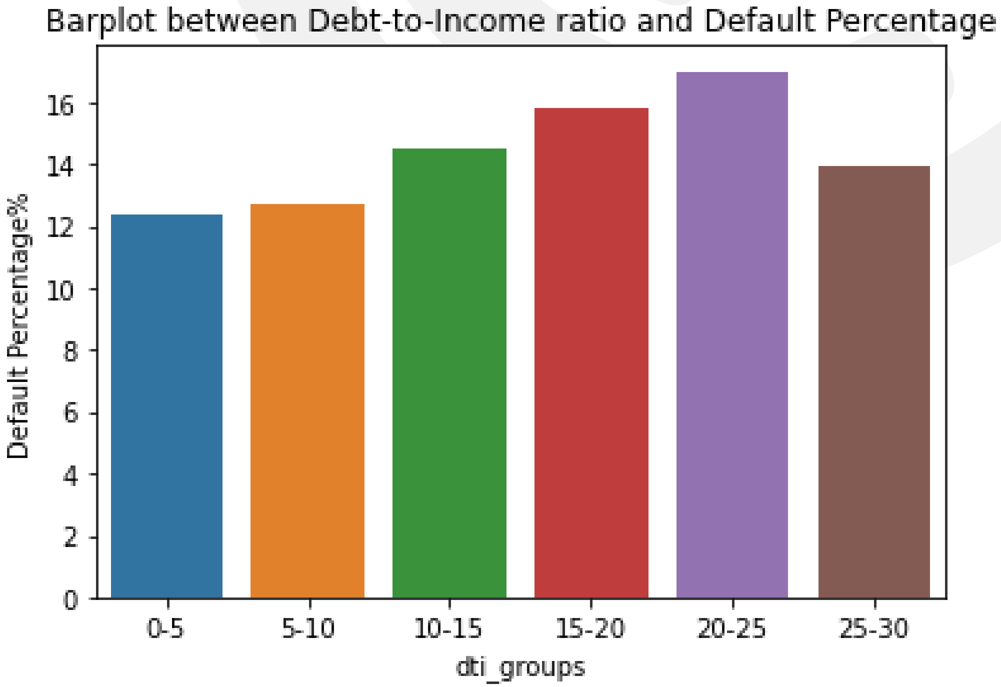
ANALYSIS OF LOAN AMOUNT VS LOAN STATUS

- Most customers apply for smaller loan amounts.
- As the loan amount increases there's a gradual/subtle decrease in density for defaulter(Charged Off).
- The dispersion of loan amount is higher for defaulter.
- This may imply that defaulters may face challenges in managing higher loan amounts, leading to committing fraud and causing loss of business to Lending Club.



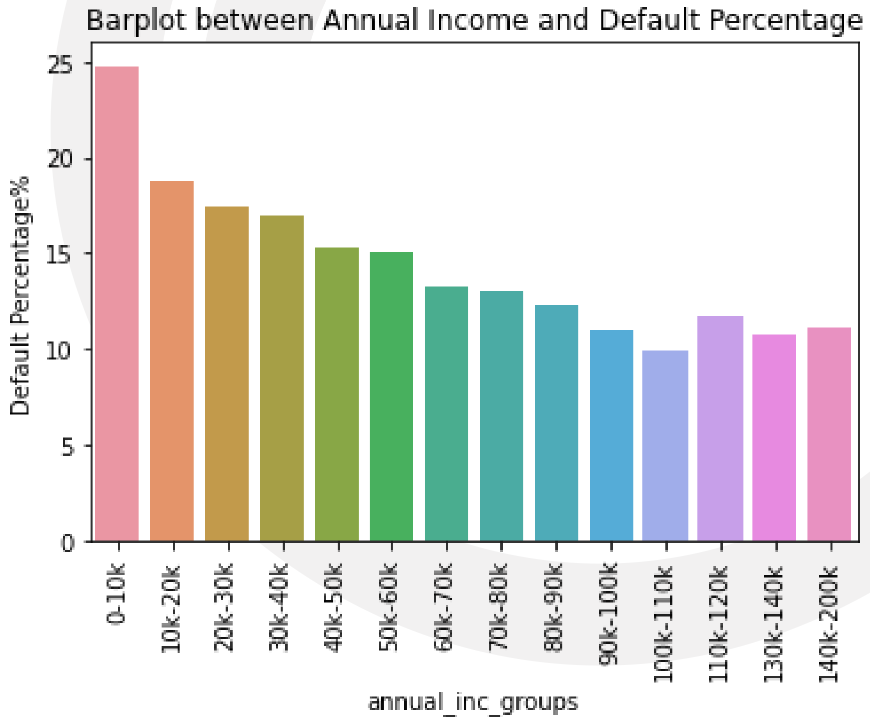
ANALYSIS OF DTI VS LOAN STATUS

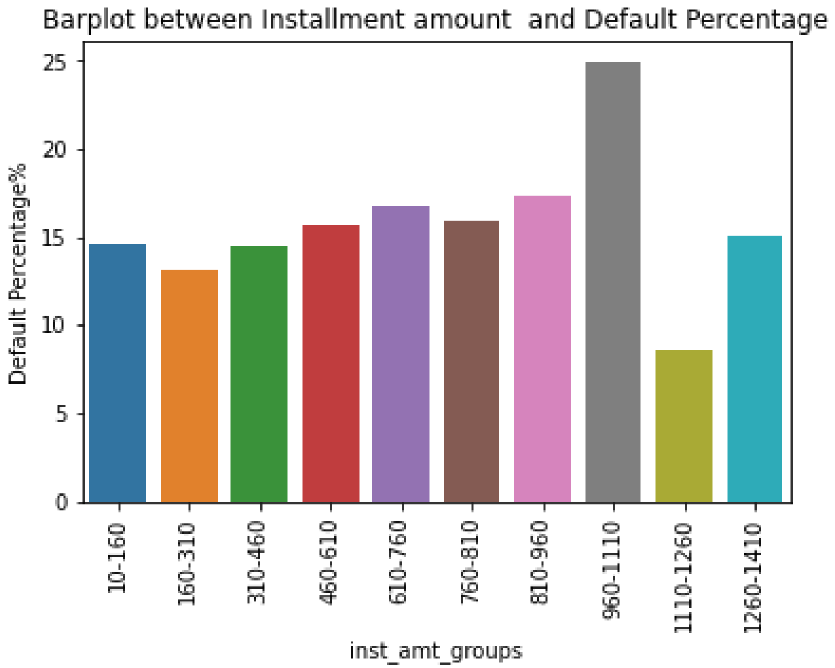
- Dti stands for Debt-to-income ratio, it tells us about the percentage of gross monthly income that is used to pay debt and determines borrowing risk.
- Customers with high debt-to-income ratio will tend to commit fraud, as they already have high debt and they are spending money in other loans as well so there's a high probability of fraud.
- It's a potential risk to provide loan to customers with higher dti.



ANALYSIS OF ANNUAL INCOME AND INSTALLMENTS VS LOAN STATUS

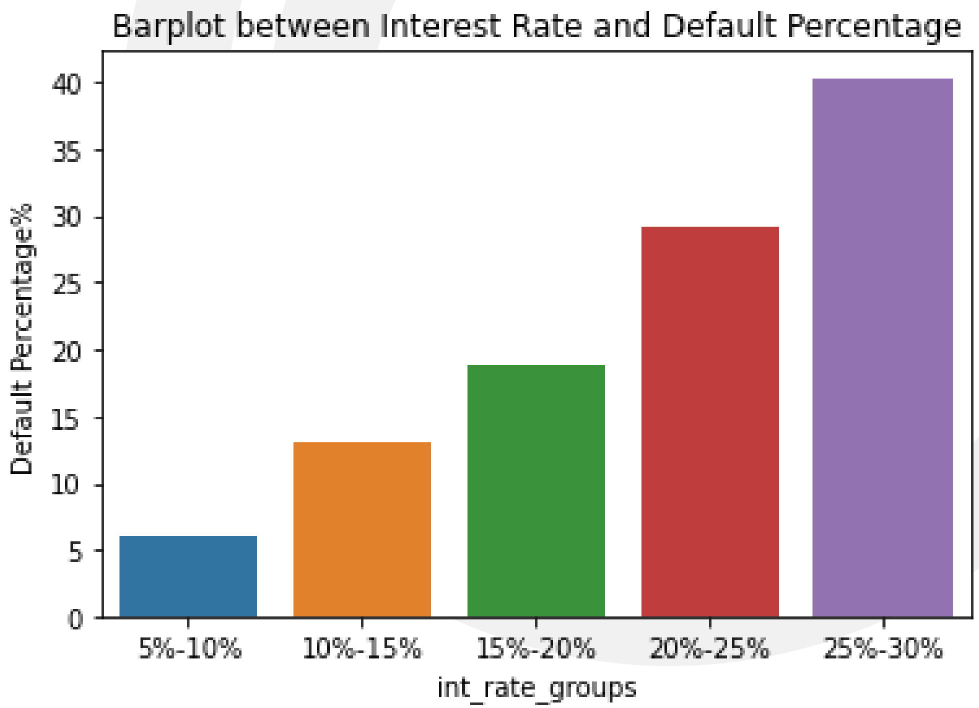
- It can be inferred that customers with lower incomes and high installment amounts customers will tend more towards committing frauds.
- With increase in annual income, dti should reduce gradually too, leading to lesser default percentage
- Large installments amount make it difficult for less incomes to repay loans due.

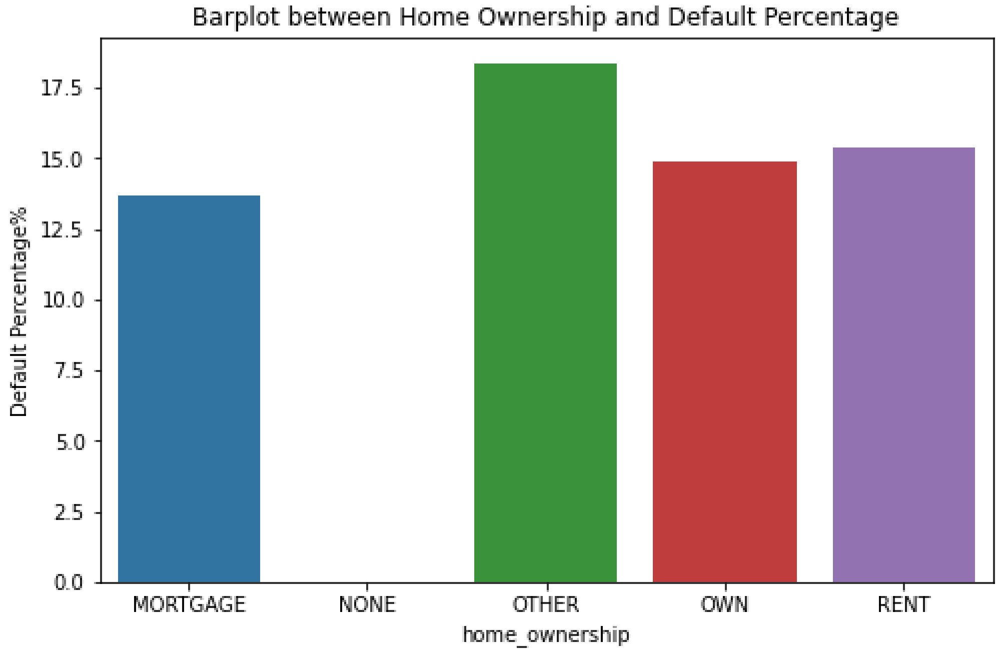




ANALYSIS OF INTEREST RATE AND HOME OWNERSHIP VS LOAN STATUS

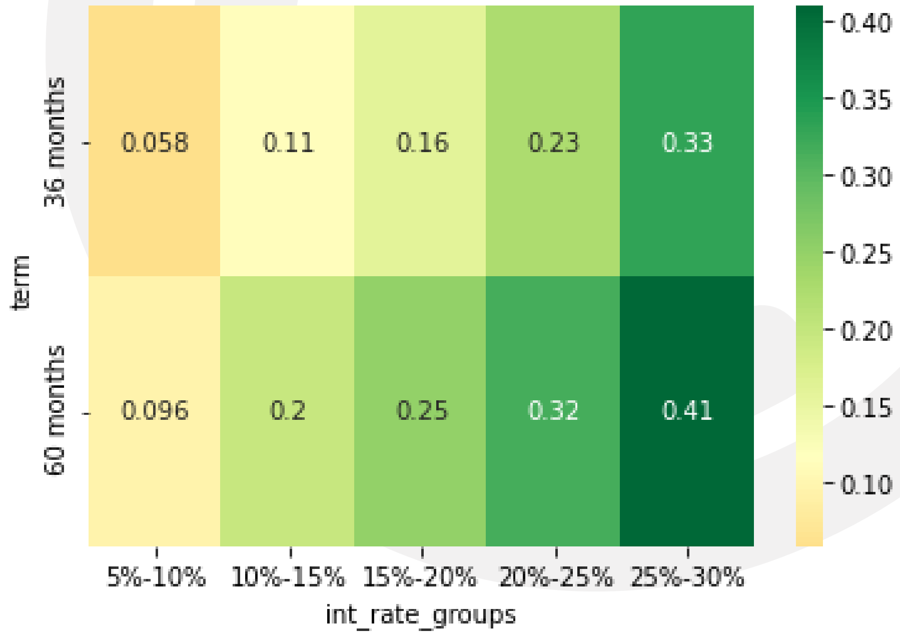
- It can be inferred that customers tend more towards committing frauds as the interest rate increases.
- Customers with home ownership as Others commit more defaults.
- It's recommended to sanction loans at lower interest rate, and reduce loan amount for high interest rate loans.
- For customers owning house are suitable for loans, as compared to other home ownership categories





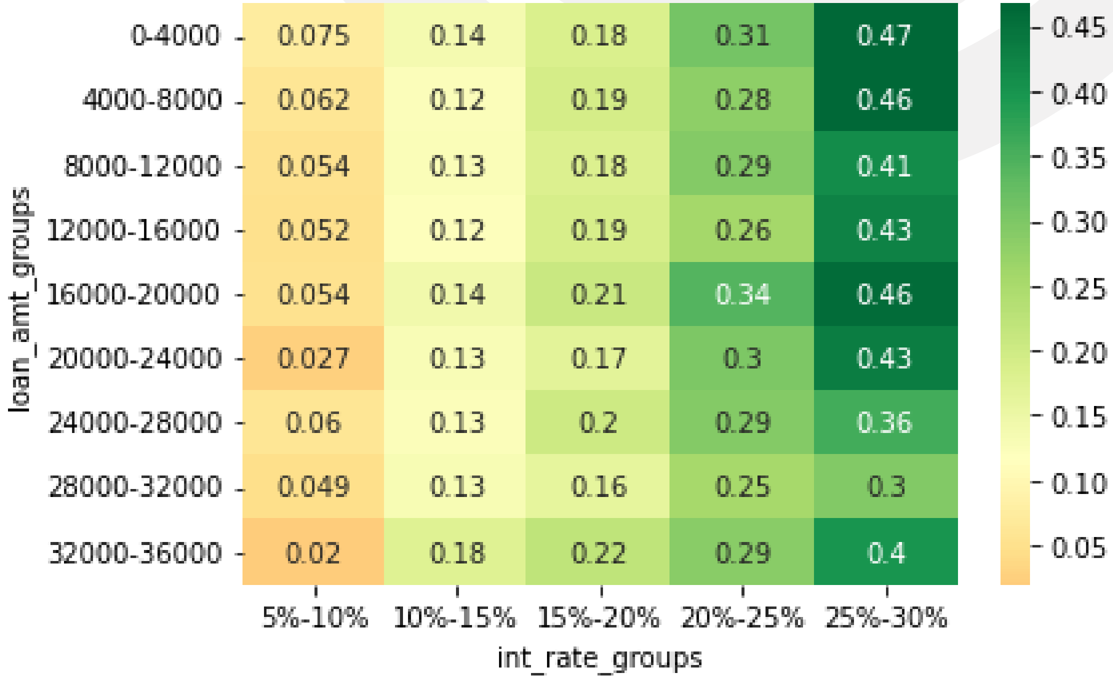
MULTIVARIATE ANALYSIS BETWEEN INTEREST RATE, LOAN TERM, DTI AND LOAN DEFAULTS

- In multivariate analysis, we have used heatmaps, to represent relations between variables on x and y axis, and values represent their behavior with loan status.
- Values closer to '0' represent non defaulted loans, and values closer to '1' represent defaulted loans.
- For loan term, increase in interest rate indicates more loan defaults.
- In 5 year loan term, more loan default occur as compared to 3 year loan period.
- Similarly, higher interest rate implies more defaults within entire dti range.
- As dti increases, loan default also increases. This trend can be seen in each interest rate group



MULTIVARIATE ANALYSIS BETWEEN INTEREST RATE ,LOAN AMOUNT AND LOAN DEFAULTS

- Higher interest rate, for any loan amount results in higher loan defaults.
- Customers will tend to fraud instead of repaying amount, more than they borrowed.
- As loan amount increases, loan default increases too.
- Lending Club should be cautious when approving loan at higher rate of interest.

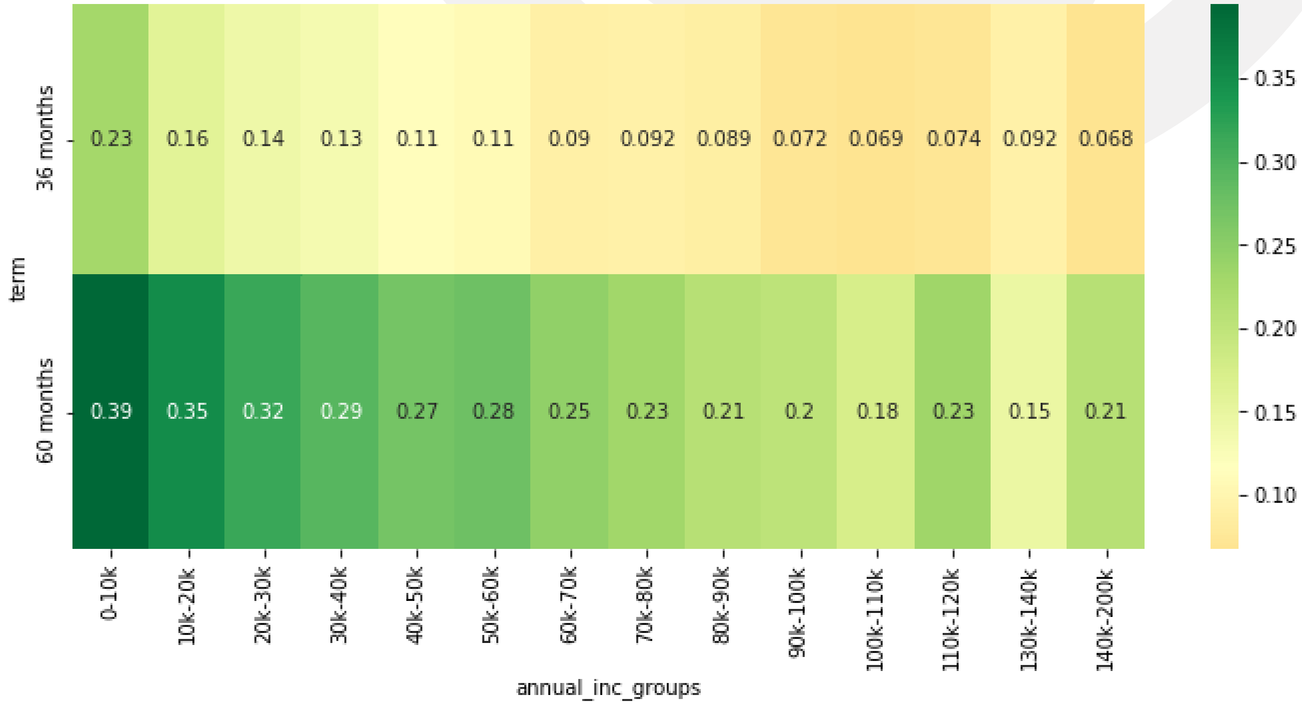


MULTIVARIATE ANALYSIS BETWEEN LOAN AMOUNT, ISSUE YEAR AND LOAN DEFAULTS

- Between 2007 to 2010, maximum loan amount is 25000/-
- Since in 2007 there was an economic crisis in USA, we can see higher loan defaults for higher loan amount as compared to other years.
- In 2008, for higher loan amounts there are more loan defaults, indicating the effects of 2007 crisis.
- In 2011, customer applied for higher loan amounts, maybe because of economic growth, inflation.
- Similar trend as 2008 can be seen in 2011, higher amount of loans have more loan defaults

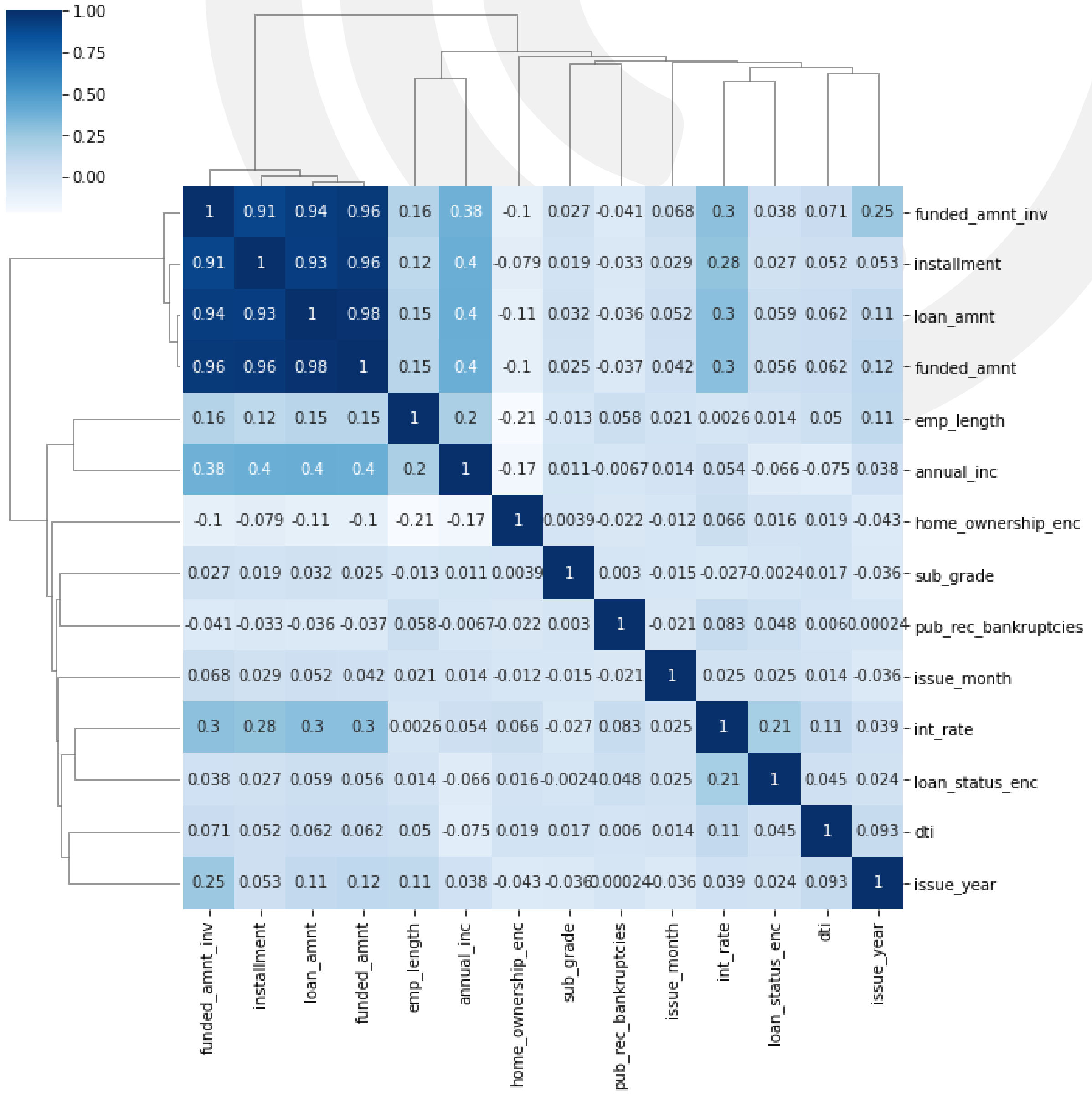
MULTIVARIATE ANALYSIS BETWEEN ANNUAL INCOME, LOAN TERM AND LOAN DEFAULTS

- Annual income and loan term, it can be inferred that customers with lower income are risky.
- Customers falling in lower income segments are committing frauds more for longer loan terms.
- As the annual income increases the loan default decreases for both loan terms.
- 5 year loan term is at more risk than 3 year loan term.
- It's recommended to be more cautious while approving loans for lower income and longer terms.



CORRELATION MATRIX

- Correlation matrix is a statistical technique used to evaluate the relationship between two variables in a data set.
- Values tending towards 1 represent positive correlation, and values closer to -1 represents negative correlation.
- Darker values in plot represent positive correlation and lighter shades represent negative correlation.
- Loan amount, Funded amount and installments show high positive correlation, as loan amount increases installments amount increases too.
- Interest rate also shows high correlation with loan amounts and installments



DRIVING FACTORS FOR LOAN DEFAULTS

CONTINUOUS

- Loan Amount: Loan amount higher than 10000$ has more default percentage
- Interest Rate: Higher interest rates show higher loan defaults.
- Annual Income: Lower annual incomes(up to 10K) shows higher loan defaults (up to 25%)
- Installments: Higher loan installments (between 600$ to 1200$) cause loan defaults.
- DTI: DTI ratio higher than 10 witnesses more loan defaults (14-16%).

CATEGORICAL

- Term: 60 months have more loan defaults in comparison to 36 month term.
- Purpose: Small business owners are more likely to commit frauds.
- Grade: As the grade decreases (from A to G) default percent increases, customers with lower grades and subgrades are a risk.
- Verification Status: Verified customer are more prone to charge off (loan default)

RECOMMENDATIONS

- Lending Club should be cautious before approving any loans above 15000$
- If interest rate is higher, loan defaults are higher, consider approving loans at lower interest rates and low loan amounts.
- Lower annual income and higher DTI are a risk, avoid such applicants to reduce credit loss.
- Small business owners and 60 month loan terms are risk areas, should minimise these applicants.
- Applicants with lower grade and subgrade(D2-G5) should not be considered for loan approvals.
- Verified borrowers are causing more credit loss, update customer verification process.

THANK YOU!

divysaxena.saxena2@gmail.com

manchikantisreenivasulu027@gmail.com

https://github.com/divyanshsaxena7594