

# **LENDING CLUB CASE STUDY**

Indraja Punna

# PROBLEM STATEMENT

The consumer finance company faces a critical decision-making challenge in approving loans, balancing the risk of business loss if a creditworthy applicant is denied and the financial loss if a likely defaulter is approved.

The objective is to analyze past loan applicant data to identify patterns that signal potential defaults, enabling the company to make informed decisions such as denying loans, reducing loan amounts, or adjusting interest rates for risky applicants.

The company aims to optimize its loan approval process by leveraging historical data to minimize both business loss and financial risk associated with loan approvals.



# APPROACH

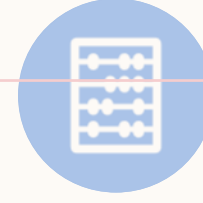
1. Data Cleaning
2. Data Analysis
  - a. Univariate Analysis
  - b. Segmented Analysis
  - c. Bi-variate Analysis
  - d. Multi-variate Analysis
3. Data Insights

# DATA CLEANING



## CLEAN COLUMNS

- ☐ Create a list of unneeded columns and drop them
- ☐ Drop all the columns where mean percentage of null values is greater than 30.
- ☐ Drop all the columns where there is only a single unique value
- ☐ Convert date fields and numeric fields into correct formats
- ☐ Create aging columns from date columns for easy data analysis
- ☐ Drop columns with outliers



## CLEAN ROWS

- ☐ Impute rows for numerical and categorical columns.
- ☐ For numerical columns rows are imputed with median
- ☐ For categorical columns rows are imputed with mode.
- ☐ Removed rows with categories like Other which do not make much sense



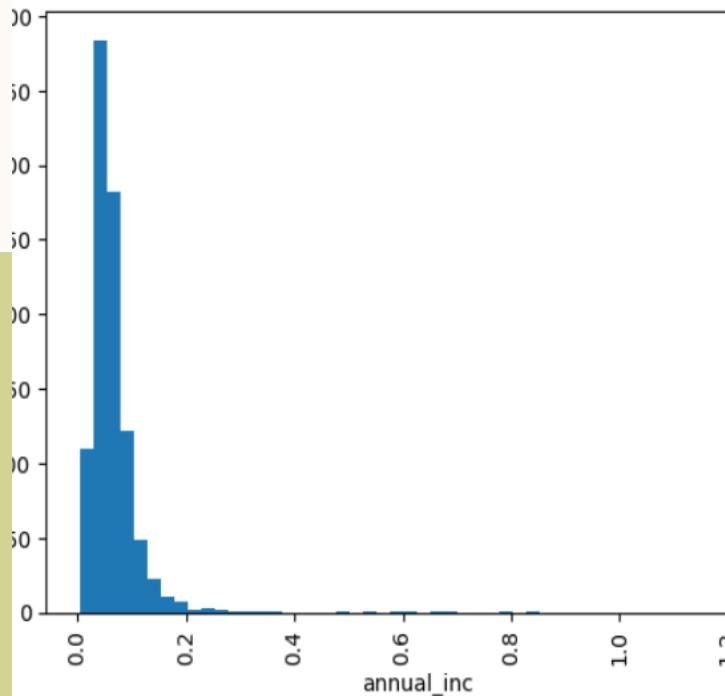
# UNIVARIATE ANALYSIS - APPROACH

- We will be analyzing only the data with the status of loan as charged off to find the insights.
- Categorized each column into either numerical or categorical based on the type of values it has.
- Created a histogram for each column in numerical columns list to derive business insights from it.
- Created a count plot for each column in categorical columns list to derive business insights from it.

# UNIVARIATE ANALYSIS - INSIGHTS

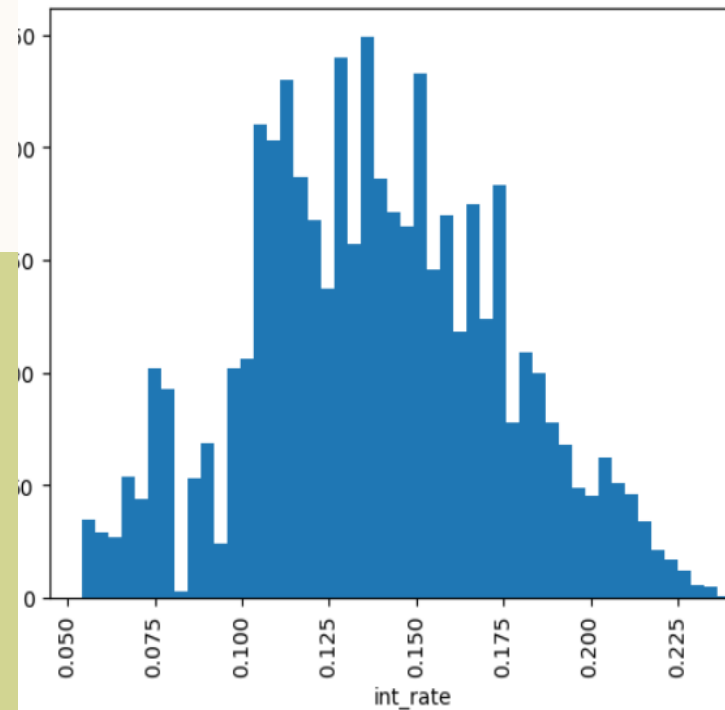
## NUMERICAL COLUMNS

Histogram for annual\_inc



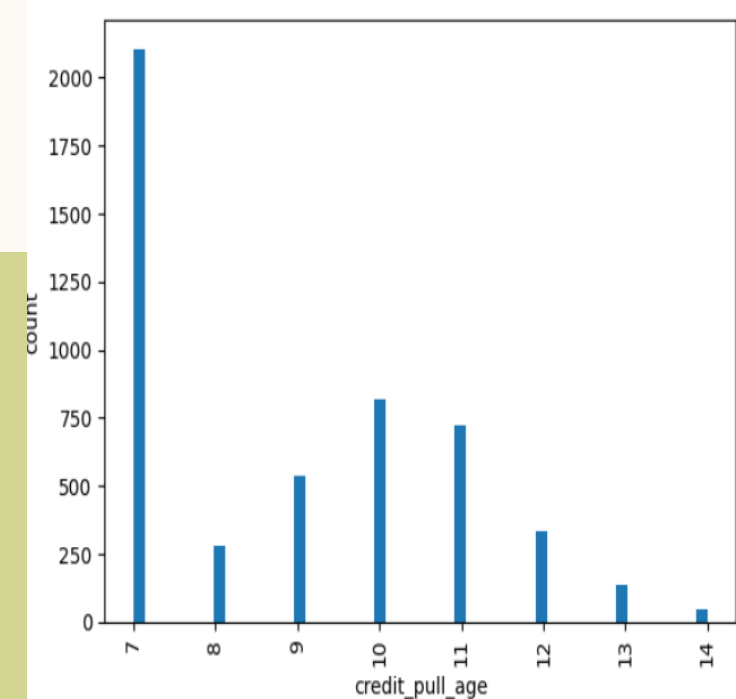
MAJORITY OF DEFAULTERS WHO HAVE  
TAKEN LOAN HAD AN ANNUAL INCREMENT  
BETWEEN 0 TO 1 LAKH

Histogram for int\_rate



MAJORITY OF DEFAULTERS HAVE TAKEN  
INTEREST RATE BETWEEN 10% TO 17%

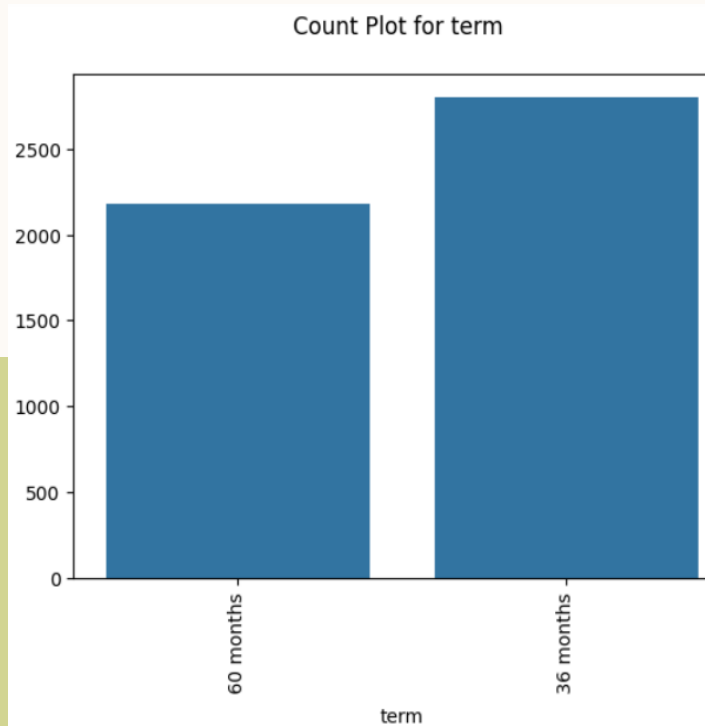
Histogram for credit\_pull\_age



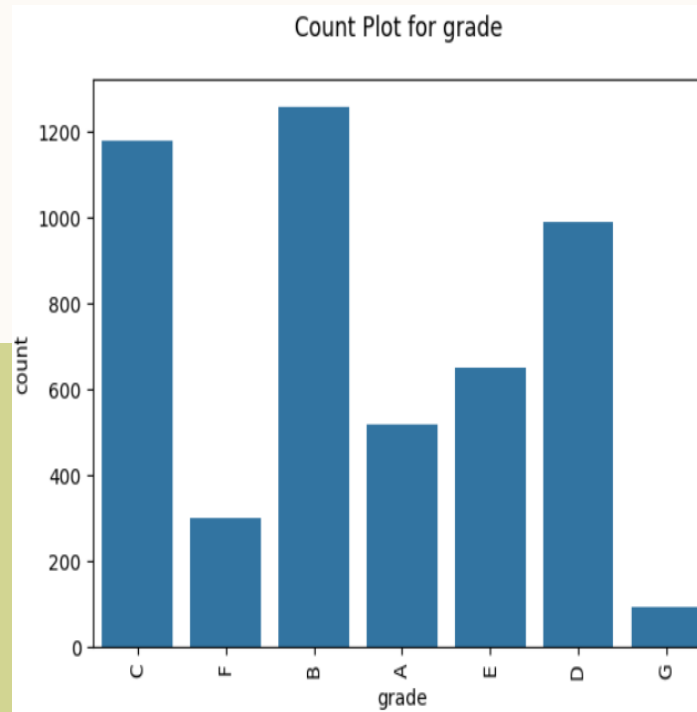
THE LAST CREDIT DATA PULLED FOR  
MAJORITY OF DEFAULTERS IS 7 YEARS AGO

# UNIVARIATE ANALYSIS - INSIGHTS

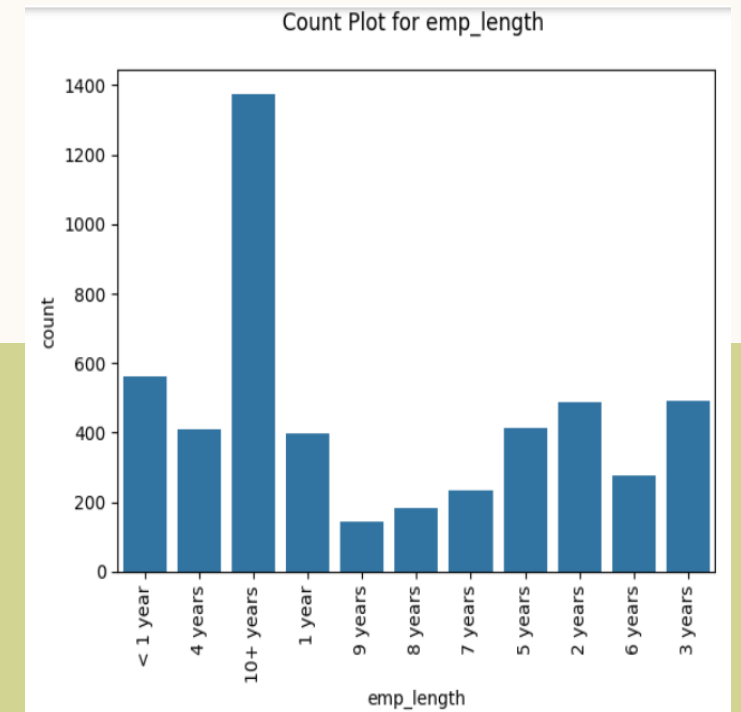
## CATEGORICAL COLUMNS



MAJORITY OF PEOPLE WHO HAVE  
DEFAULTED TOOK IT FOR A TERM OF 36  
MONTHS



LEAST NUMBER OF PEOPLE DEFAULTED IN  
GRADE G

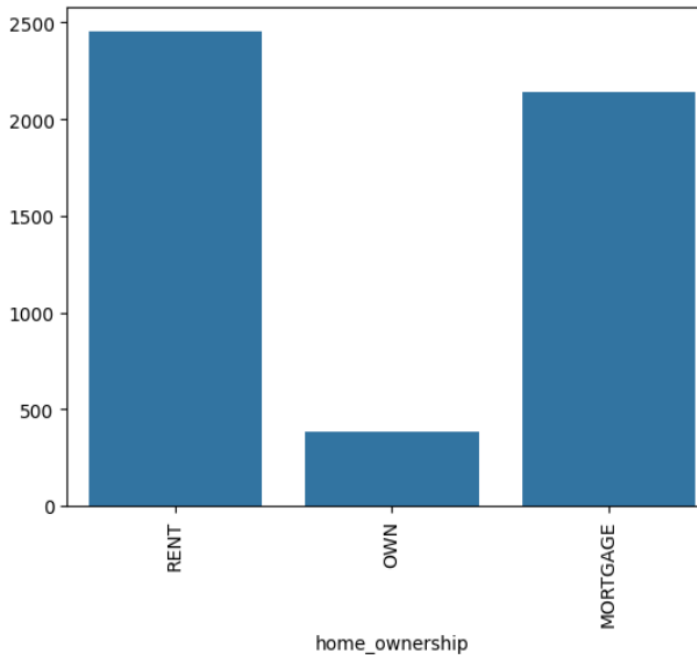


MAJORITY OF PEOPLE WHO HAVE  
DEFAULTED FALL WITH EMPLOYMENT  
DURATION OF 10+YEARS.

# UNIVARIATE ANALYSIS - INSIGHTS

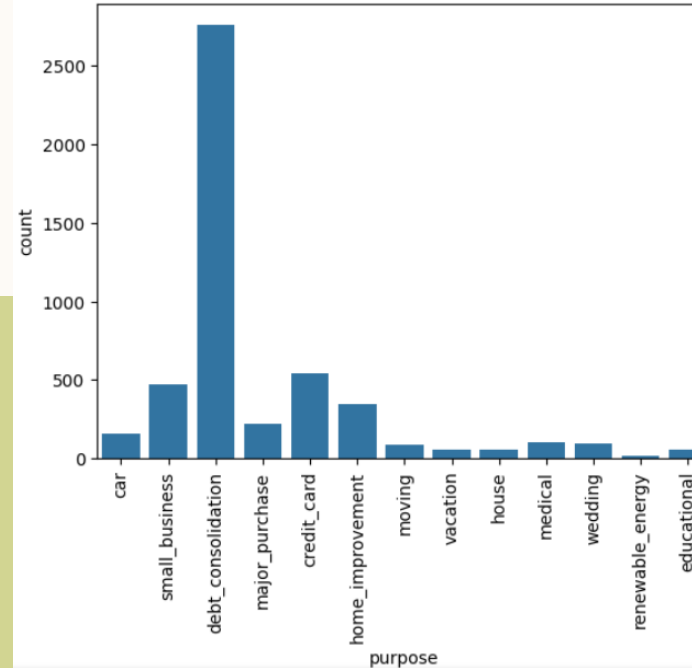
## CATEGORICAL COLUMNS

Count Plot for home\_ownership



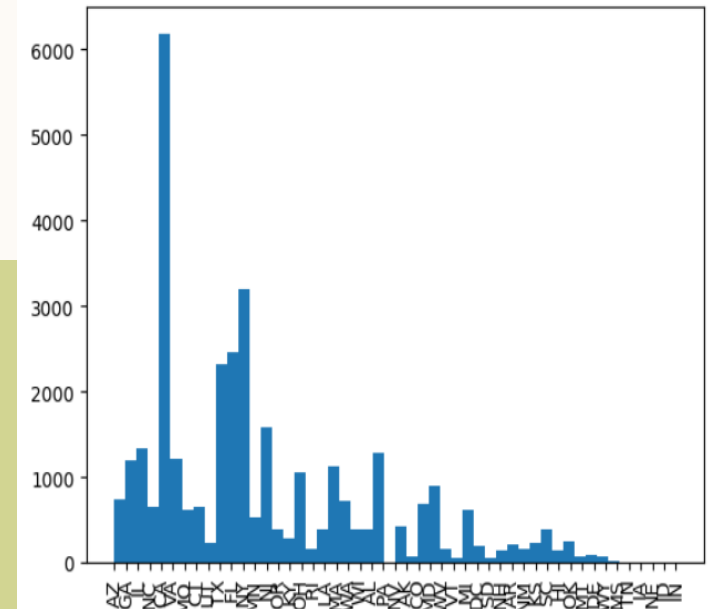
LEAST NUMBER OF PEOPLE WHO HAVE  
DEFAULTED ARE HOME OWNERS. MOST ARE  
RENTERS

Count Plot for purpose



MAJORITY OF PEOPLE WHO DEFAULTED  
HAVE PURPOSE AS DEBT\_CONSOLIDATION

Histogram for addr\_state



MAJORITY OF PEOPLE WHO DEFAULTED  
ARE FROM THE STATE OF CA





# **SEGMENTED ANALYSIS - APPROACH**

- We will be analyzing the data with the status of loan as fully paid and charged off to find the insights.
- Calculated the percentage of people who have defaulted in each categorical variable.

# SEGMENTED ANALYSIS - INSIGHTS

- Percentage of people who defaulted in 60 months term is 25% compared to 36 months term which is 10.8%.
- Percentage of people who defaulted in Grades F and G is higher and grade A is the lowest.
- Percentage of defaulters is higher in people who took loan for small business.
- Percentage of defaulters is higher in state NE compared to others.
- Around 16% of total loan amount is defaulted

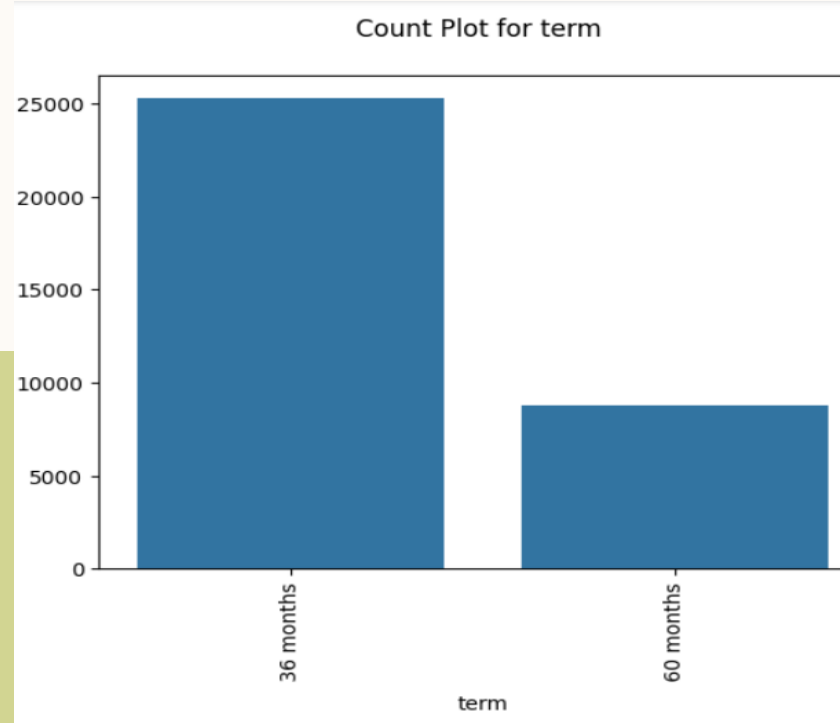


# **BIVARIATE ANALYSIS - APPROACH**

- We will be analyzing the data with the status of loan as fully paid and charged off to find the insights.
- Created a scatter plot for each column in numerical columns vs numerical to derive business insights from it.
- Created a box plot for each column in categorical columns vs numerical columns to derive business insights from it.
- Created a bar plot for each column in categorical columns vs categorical columns to derive business insights from it.

# BIVARIATE ANALYSIS - INSIGHTS

## NUMERICAL VS NUMERICAL

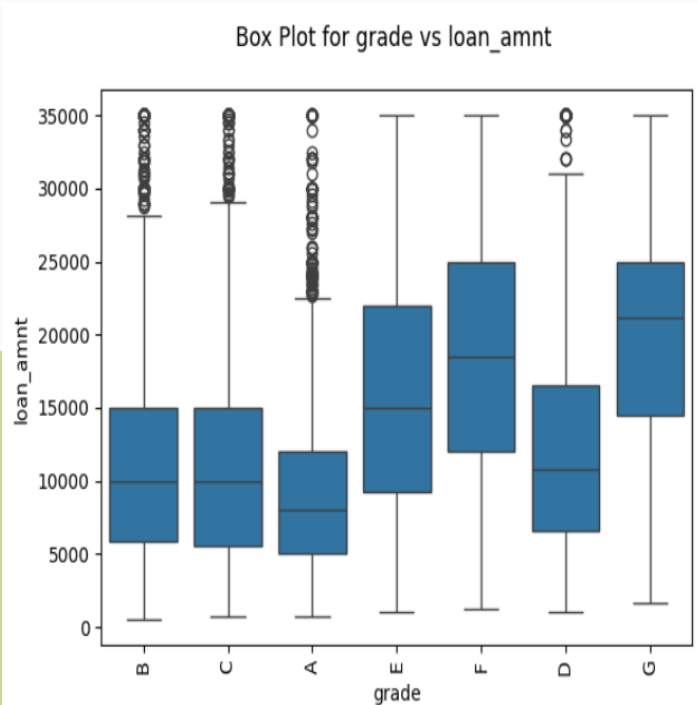


“INSTALLMENT VS LOAN\_AMNT”

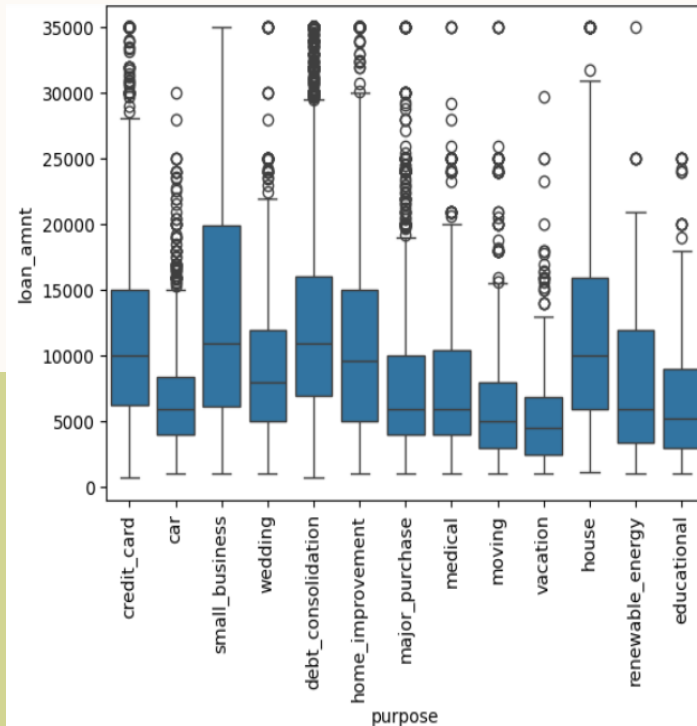
IN MOST OF THE CASES THE HIGHER THE LOAN  
AMOUNT INSTALLMENTS ALSO INCREASE

# BIVARIATE ANALYSIS - INSIGHTS

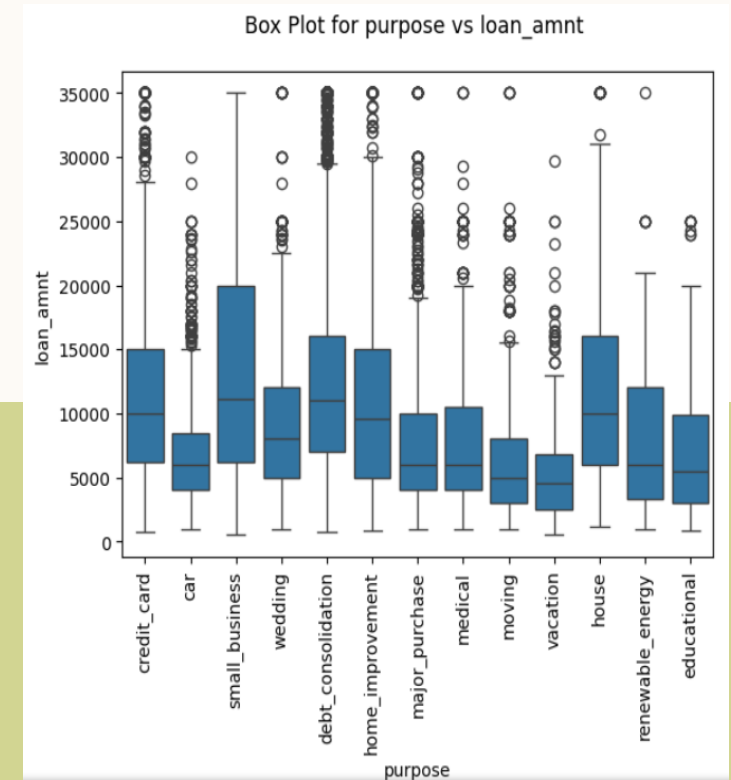
## NUMERICAL VS CATEGORICAL



PEOPLE IN GRADE G HAVE TAKEN MINIMUM  
LOAN GREATER THAN ANY OTHER GRADE



PEOPLE WITH EMPLOYMENT LENGTH 10+  
YEARS HAVE TAKEN HIGHEST LOAN  
AMOUNT

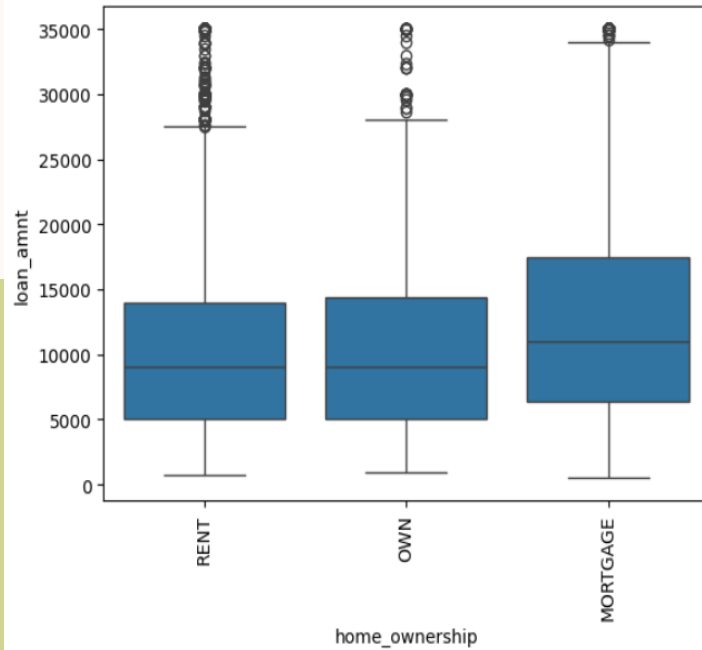


PEOPLE WHO WANTED TO DO SMALL  
BUSINESS HAVE TAKEN HIGHER LOAN  
AMOUNT NEXT COMES HOUSE PURCHASE

# BIVARIATE ANALYSIS - INSIGHTS

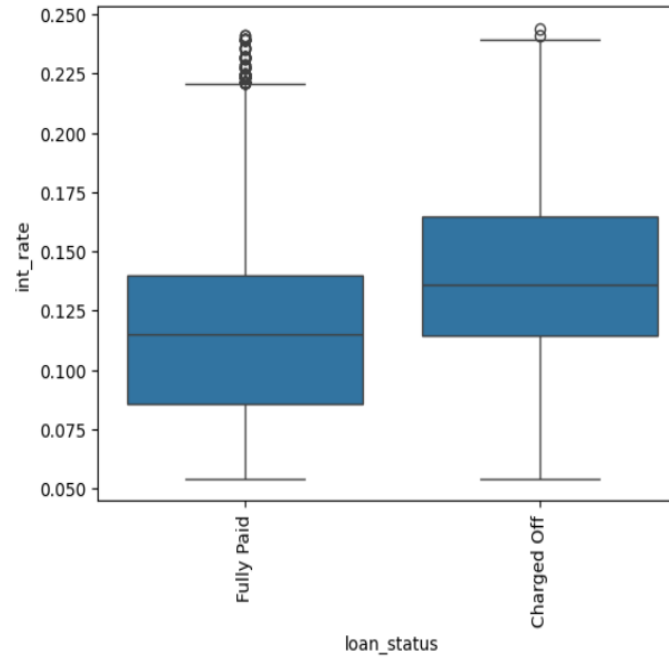
## NUMERICAL VS CATEGORICAL

Box Plot for home\_ownership vs loan\_amnt



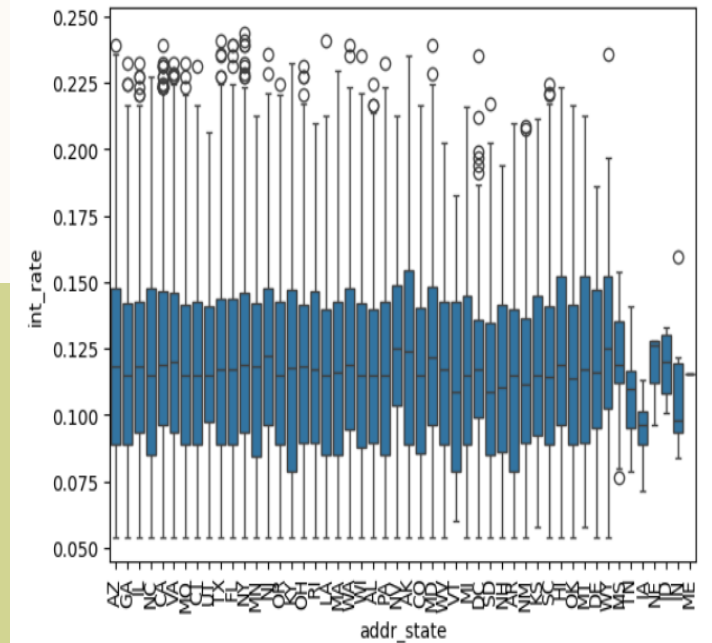
PEOPLE IN MORTGAGE HAVE TAKEN HIGHER  
LOAN AMOUNT

Box Plot for loan\_status vs int\_rate



DEFAULTERS HAVE TAKEN LOANS WITH  
HIGHER INTEREST RATES. HIGHER THAN 50  
PERCENTILE OF FULLY PAID PEOPLE

Box Plot for addr\_state vs int\_rate



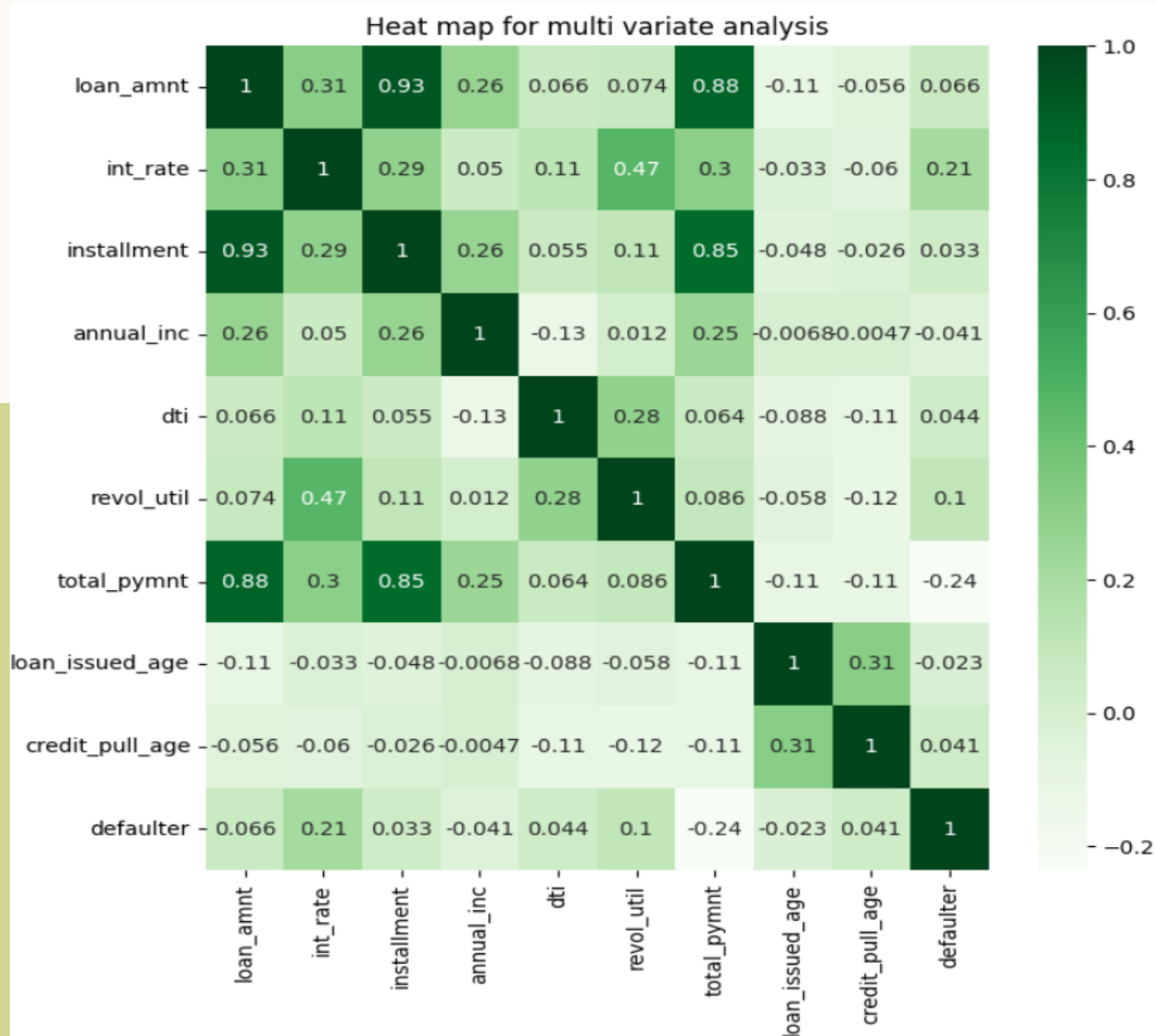
PEOPLE FROM THE STATE OF NE ARE  
PAYING HIGHEST MINIMUM INTEREST RATE



# **MULTI-VARIATE ANALYSIS - APPROACH**

- Created a heatmap with correlation
- Created a new derived field defaulter with values 0 or 1 to include in the heat map.

# MULTIVARIATE ANALYSIS - INSIGHTS



- Annual increment and Defaulter are negatively correlated. This means if a person gets less annual increment he has a higher chance to default
- Interest rate and defaulter are positively correlated. This means there is a high chance of a person being a defaulter if he has taken loan at a higher interest rate.



# FINAL INSIGHTS

- Annual Increment: Exercise caution while approving loans for the people with less annual increments between 0 to 1 lakh as there are majority of defaulters in this range
- Home Ownership: Home Owners default the least. Renters and Mortgagers default the most. Mortgagers take high amount of loans. It is advisable to give loans to have strict scrutiny while giving loans to renters and mortgagers.
- State: People from the state of NE are paying highest minimum interest rate and have higher percentage of defaulters. Be wary when dealing with them.
- Interest Rate: Defaulters have taken loans with higher interest rates. Be wary if a person he is willing to take loan with high interest rate.
- Purpose: Be wary when the person is taking loan for debt consolidation or small business. As majority of people who defaulted took the loan for debt consolidation and percentage of defaulter is higher in people who took loan for small business
- Grade: People in grade G have taken highest minimum loan than any other grades. And percentage of people who defaulted in grades G and F are the highest. Be wary while giving loans to people in grade G

# FINAL INSIGHTS CONT..

- **Term:** Percentage of people who defaulted in 60 months term is 25% compared to 36 months term which is 10.8%. Be cautious while giving loans for 60 months term.
- **Loan Amount:** Around 16% of total loan amount is defaulted. Try to limit the exposure.
- **Employment length:** People with employment length 10+ years have taken highest loan amount. Higher the loan amount installments also rise which can cause a default. Be wary of them.
- **Last Credit Pull Date:** The last credit data pulled for majority of defaulters is 7 years ago. Please make sure credit pull is done at the earliest and give the loan.

In summary, people with less annual increments, who are renters or mortgagers, who take loan for the purpose of small business, who take higher loan amount, with higher term, in higher installments, with higher interest rates tend to default the most.



# **THANK YOU**

Indraja Punna

[Indraja.punna@gmail.com](mailto:Indraja.punna@gmail.com)