LENDING CLUB CASE STUDY

RAJASHEKAR E

KARTHIK VENKATRAMAN

THE PROBLEM

What is the problem?

- Solving this assignment will give you an idea about how real business problems are solved using EDA. In this case study, apart from applying the techniques you have learnt in EDA, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

Who has this problem?

- consumer finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile

Why should this problem be solved?

- To better analyze and provide the loan for the customer which is beneficial to the customer as well as the financial company

How will I know this problem has been solved?

- Solving this assignment will give you an idea about how real business problems are solved using EDA. In this case study, apart from applying the techniques you have learnt in EDA, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers

BUSINESS UNDERSTANDING

- You work for a consumer finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

- 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 data given in the upcoming slide contains the information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

- In this case study, you will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

BUSINESS UNDERSTANDING

- When a person applies for a loan, there are two types of decisions that could be taken by the company:

1. Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:

    1. Fully paid: Applicant has fully paid the loan (the principal and the interest rate)

    2. Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

    3. Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

2. Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

ANALYSIS APPROACH
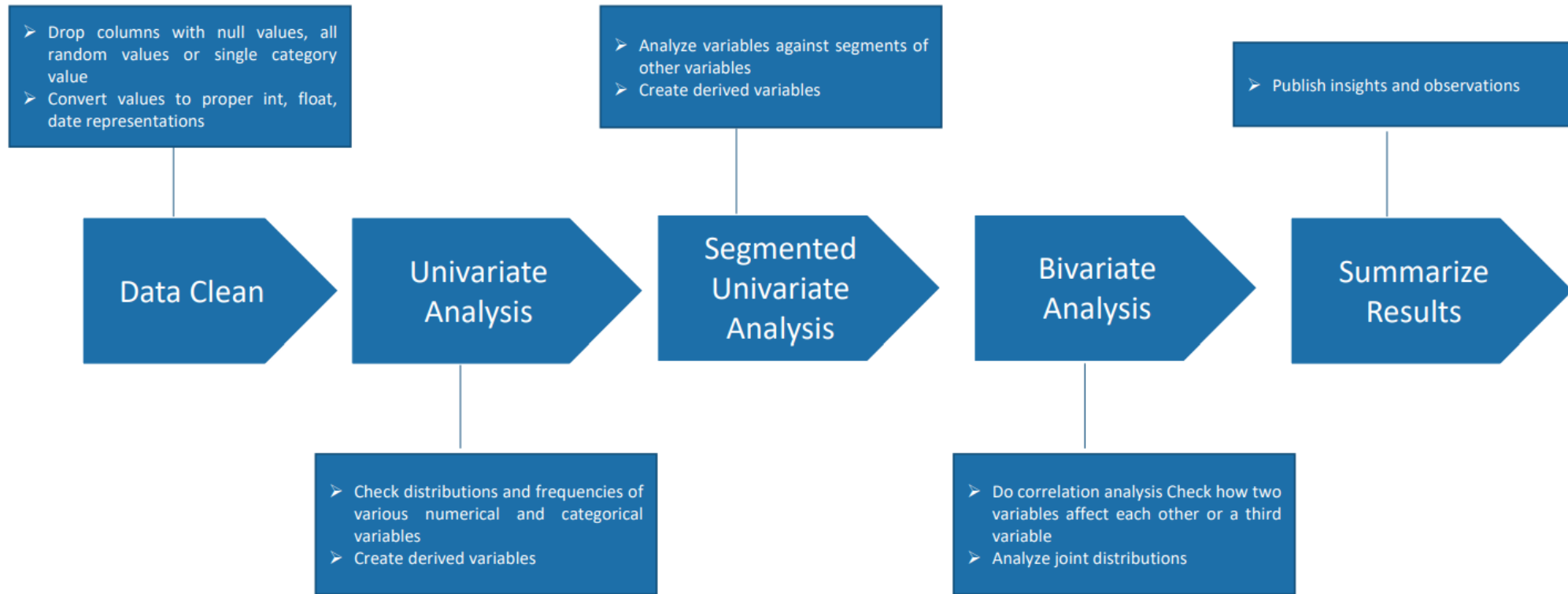
DATA CLEANING STEPS

- Delete columns: Delete unnecessary columns.

- Remove outliers: Remove high and low values that would disproportionately affect the results of your analysis.

- Missing values: Treat missing values with appropriate approach.

- Duplicate data: Remove identical rows, remove rows where some columns are identical.

- Filter rows: Filter by segment, filter by date period to get only the rows relevant to the analysis

ANALYSIS

i. The essence of the whole project is to analyze and understand how consumer attributes and loan attributes are influencing the tendency of defaulting.

ii. We performed data cleaning and preparation on the Loan dataset:
   - ❑ Imputed the NA values for all the variables
   - ❑ Created two new columns:
      - Profit and Loss column
      - Ratio of funded amount and annual income

iii. During univariate analysis we have created:
   - ❑ Histograms and Bar charts to check out the distribution of all the driver variables
   - ❑ Box plots to detect the Outliers
   - ❑ Performed the Multivariate analysis to understand how different variables interact with each other.

ANALYSIS – OVERALL LOAN STATUS

Total Loans :

Approximately 14% of loans are defaulted Any variable that increases percentage of default to higher than 16.5% should be considered a business risk
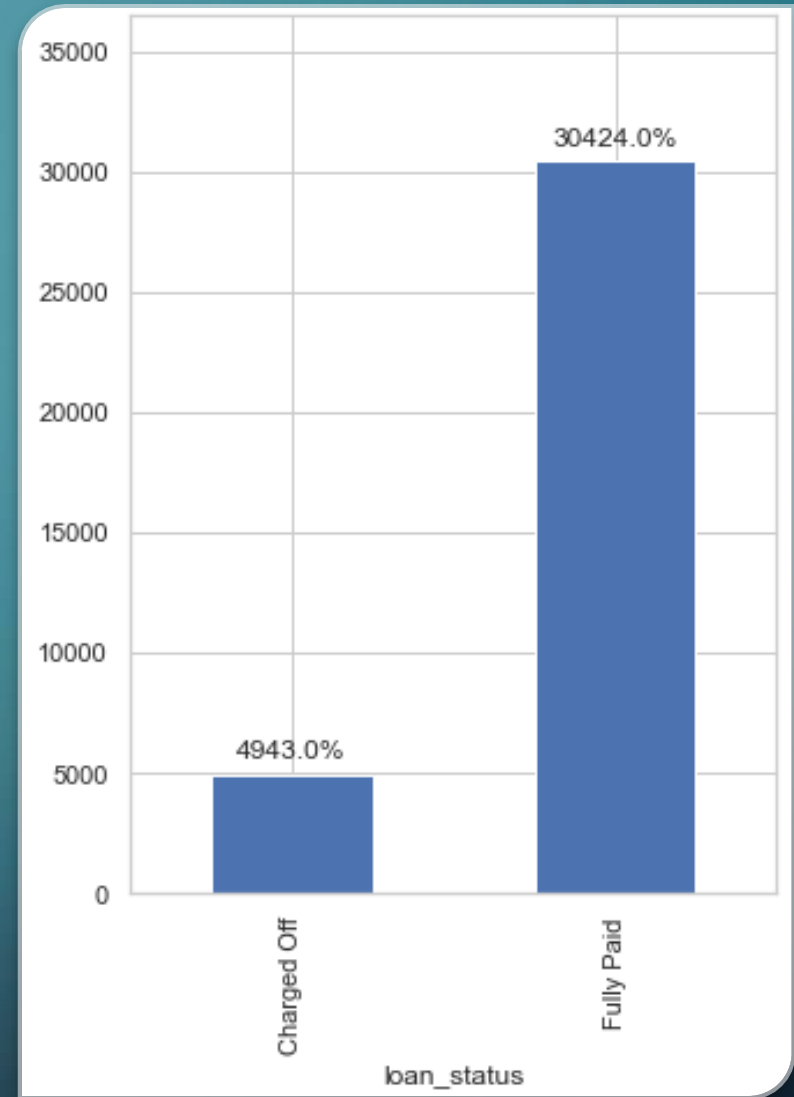
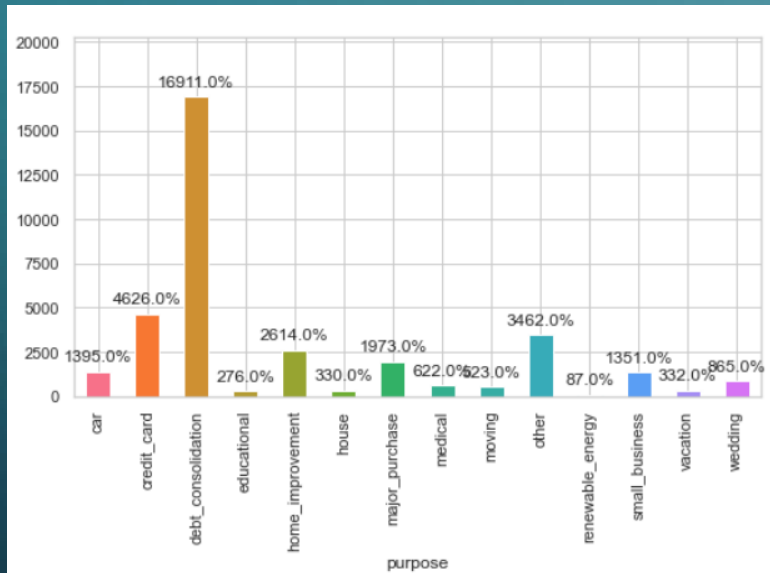# ANALYSIS – OVERALL LOAN STATUS

Total Money Earned :

Lending Club only recovers 57% of the loan amount when loans are defaulted. On fully paid up loans, the company makes 17% profit.
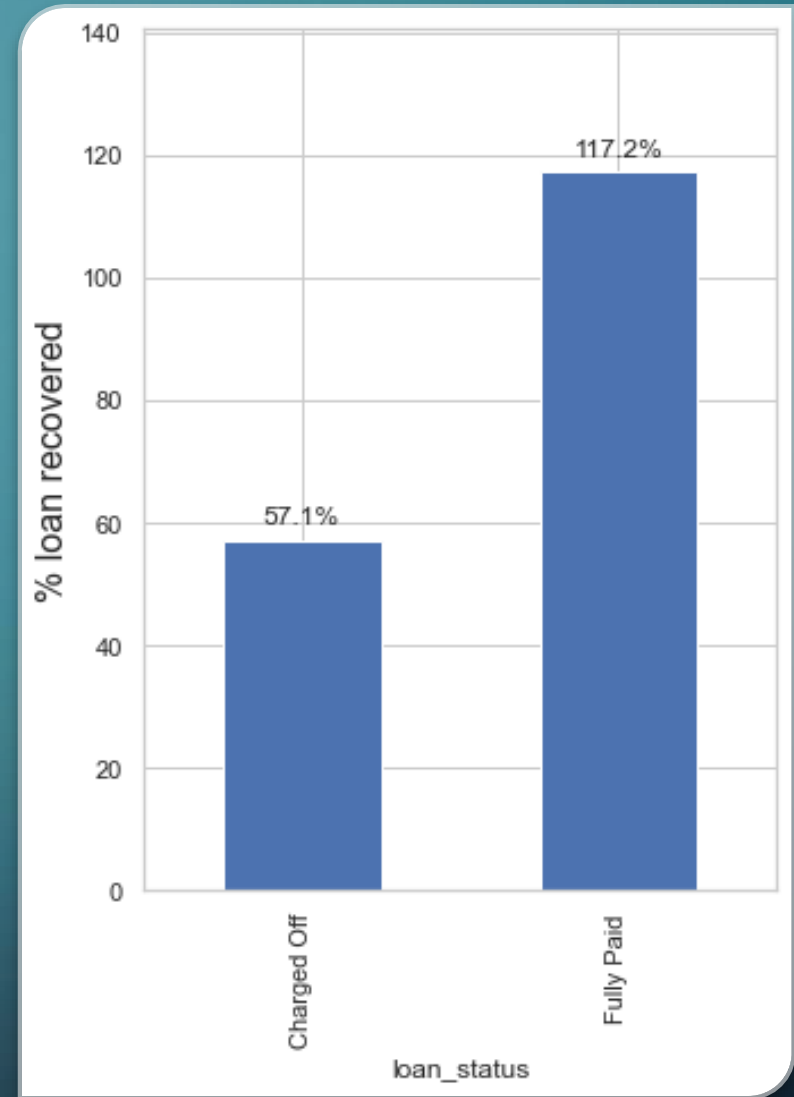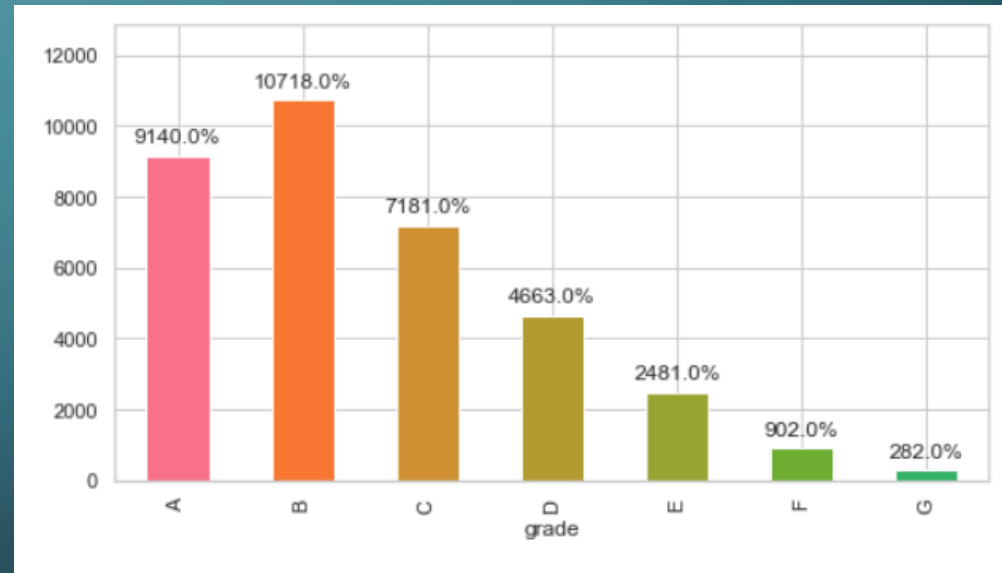
ANALYSIS – UNDERSTANDING LOANS

MAXIMUM NUMBER OF LOANS ARE FOR DEBT CONSOLIDATION, FOLLOWED BY CREDIT CARD


MOST LOANS ARE HIGH QUALITY, WITH A GRADE OF A OR B

ANALYSIS – UNDERSTANDING LOANS CONTINUED

LOWER GRADES HAVE HIGHER INCIDENCE OF DEFAULTS ON LOANS. THE GRADING SYSTEM IS WORKING! LENDING CLUB CHARGES HIGHER INTEREST RATES AS THE GRADE OF LOAN BECOMES WORSE. HOWEVER, AS WE WILL SEE ON NEXT SLIDE - THE DRIVING VARIABLE FOR DEFAULTS IS THE HIGHER INTEREST RATE.
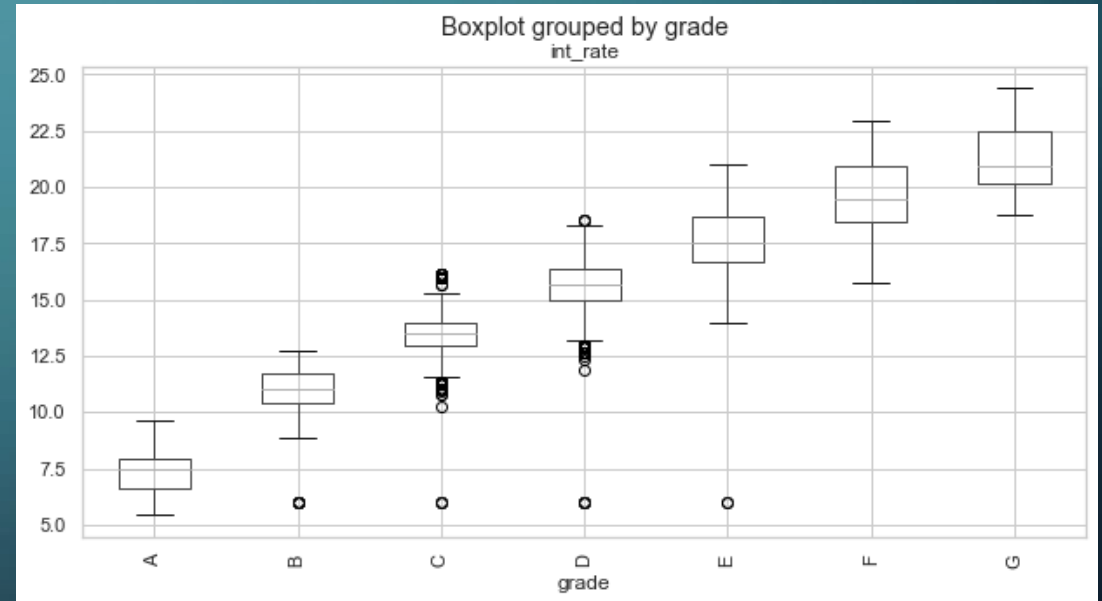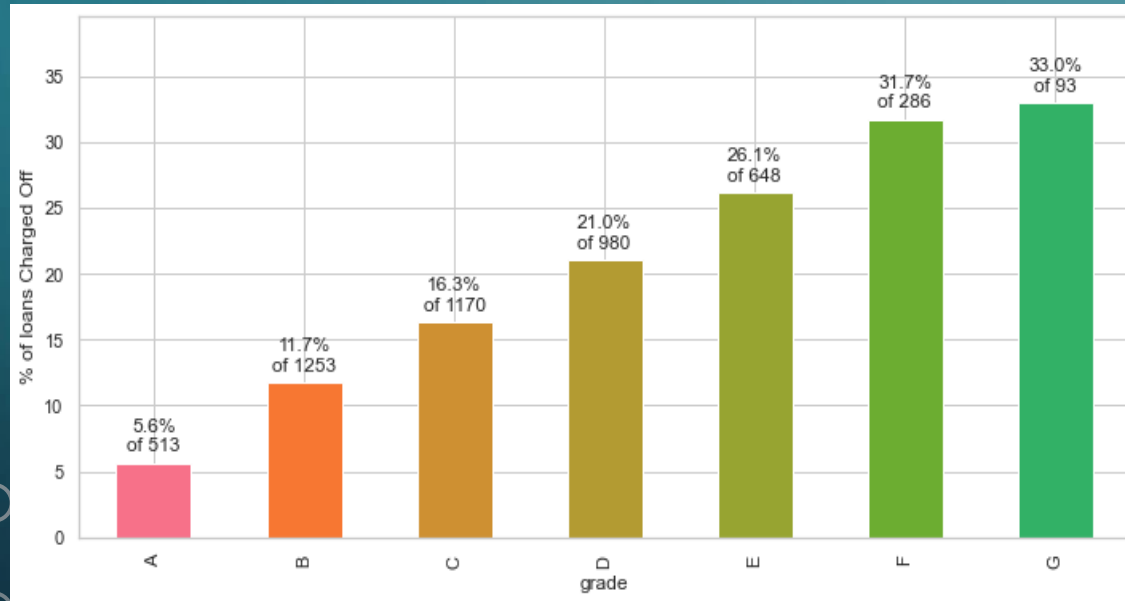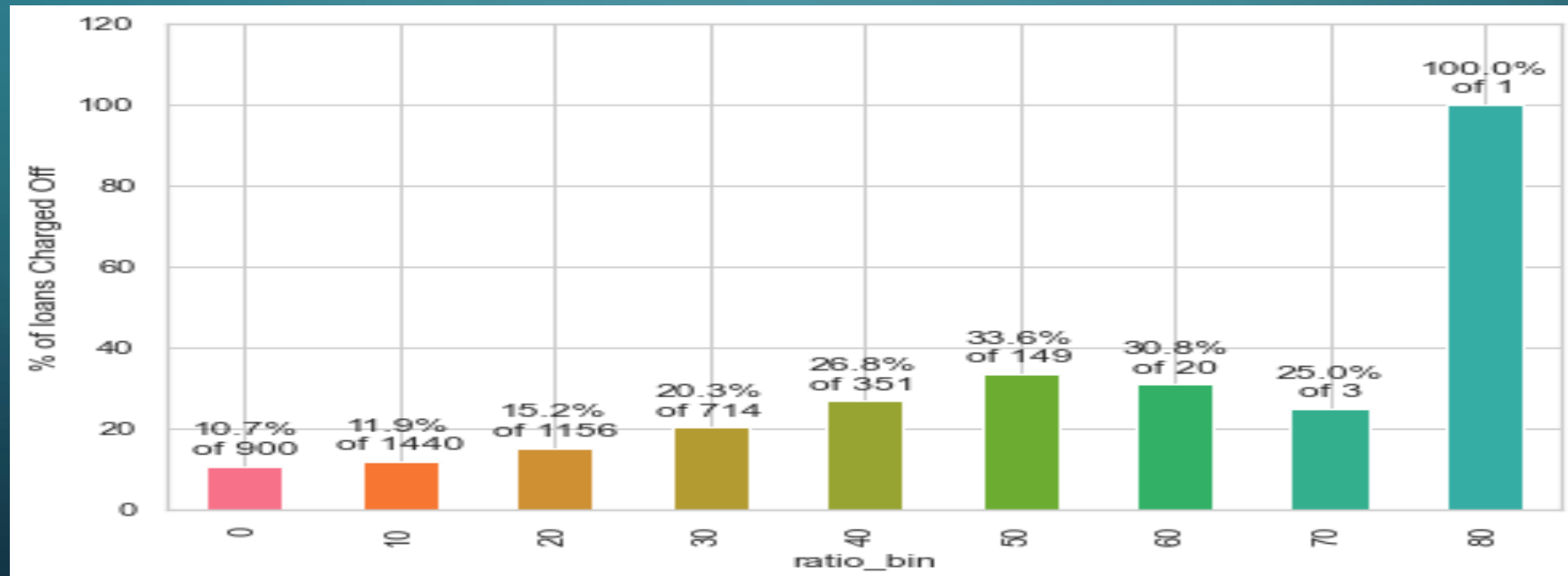
ANALYSIS – DEFAULTS BY BORROWER'S INCOME

BORROWERS HAVING ANNUAL INCOME LESS THAN 20000 DEFAULT ON THEIR LOANS AT MUCH HIGHER RATES. LOAN DEFAULT DECREASES WITH HIGHER ANNUAL INCOME. AS WE WILL SEE ON NEXT SLIDE – THE RATIO OF AMOUNT TO INCOME IS MORE IMPORTANT
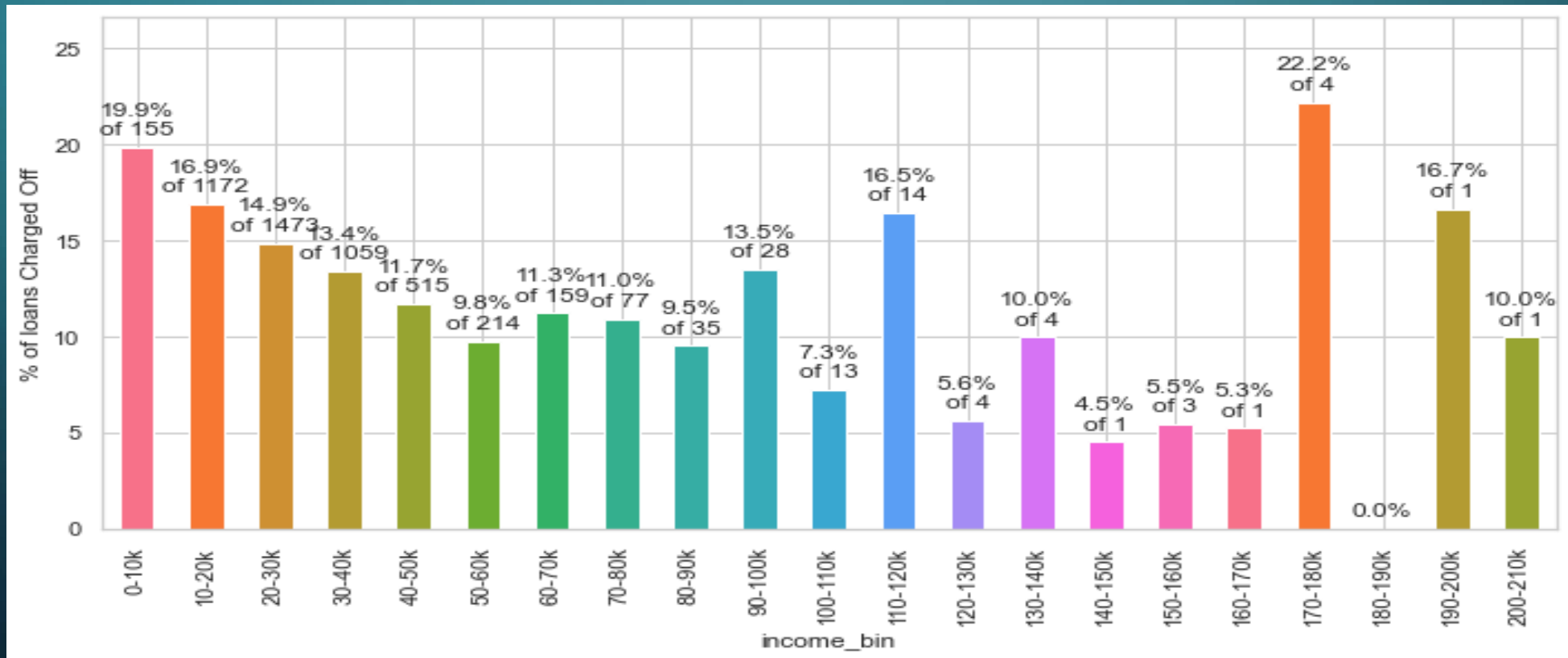
ANALYSIS – DEFAULTS BY RATIO OF AMOUNT TO INCOME

AS LONG AS LOAN AMOUNT IS LESS THAN 20% OF ANNUAL INCOME, DEFAULTS ARE LOW.

LOAN AMOUNTS OF 30% OF ANNUAL INCOME OR HIGHER SEE A HIGH RATE OF DEFAULT
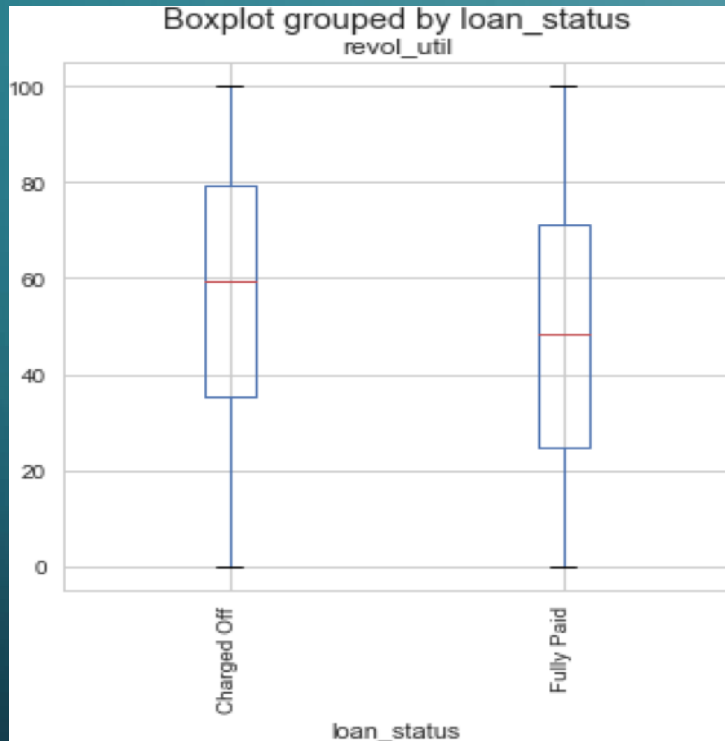
ANALYSIS - DEFAULTS BY RATIO CONTINUED

- We see here that Lending Club has extended high-value loans to people with low income.

- There are many cases of people with income 50000 or less getting loans of 25000 or more.

- This practice should be curtailed.
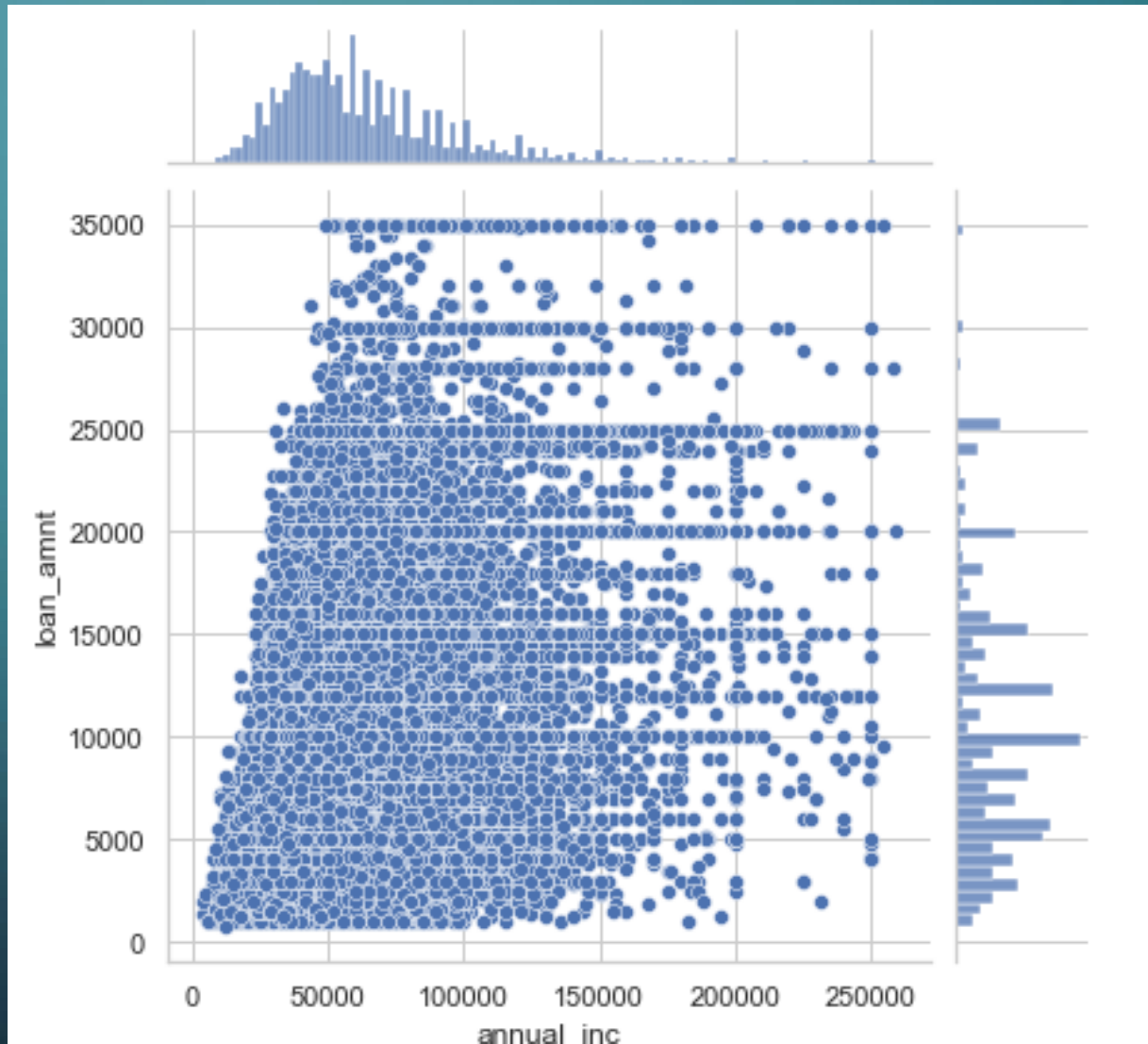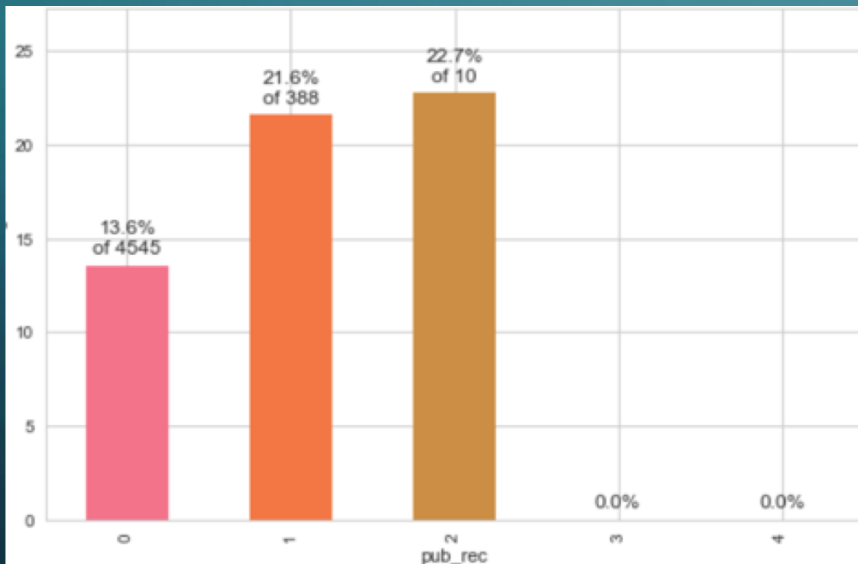
ANALYSIS – DEFAULTS BY REVOLVING LINE UTIL RATE

People with high utilization of Revolving Line of Credit at the time of taking loan default more.

Loans with utilization > 75% are risky.

ANALYSIS – DEFAULTS BY PRIOR BAD RECORD

- 94% have no Public derogatory records.
- Having even 1 derogatory record increases the chances of Charge Off significantly.

- 96% have no bankruptcy record.
- Having even 1 bankruptcy record increases the chances of Charge Off significantly.
- Public Derogatory Record and Public Bankruptcy records have 83% correlation. We can use any one.
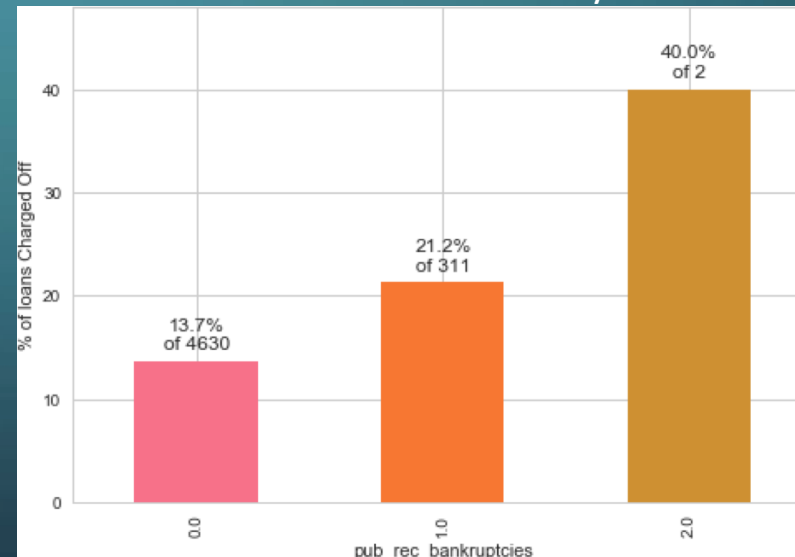
ANALYSIS BY PRIOR BAD RECORD - CONTINUED
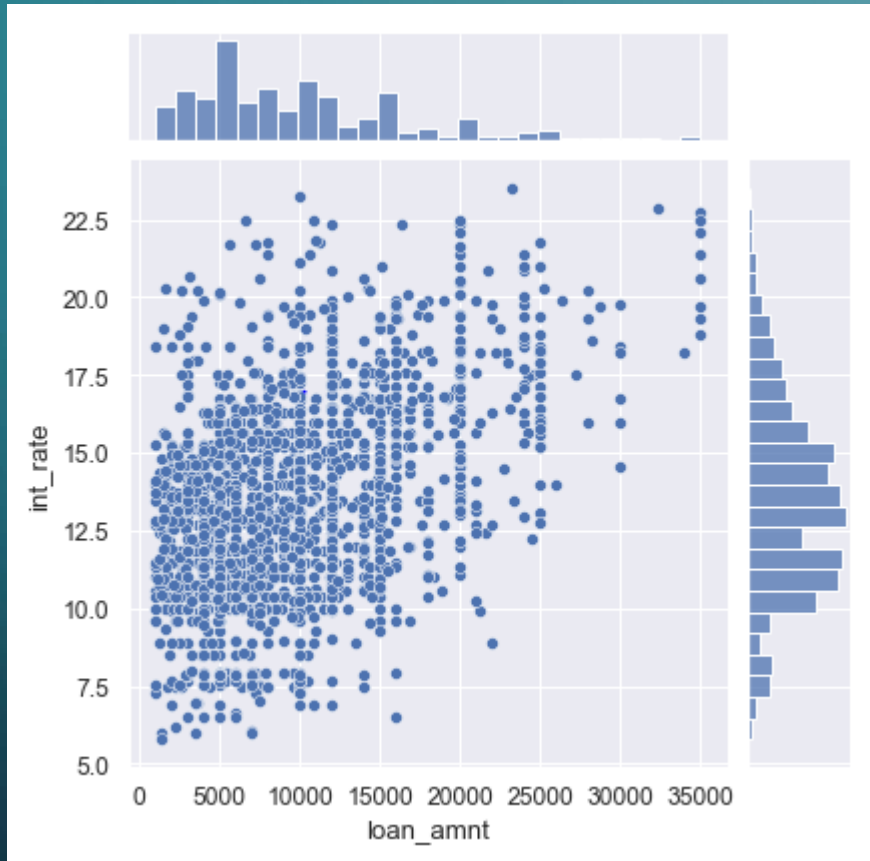
Data of people with >0 bad records

High value loans, as well as low interest loans have been extended to those with prior public derogatory records.

This practice can be stopped to improve business metrics.

APPENDIX: INTERESTING TIDBITS

People who took their first loans just before an economic crisis, like the one in 1980 and the subprime crisis in 2008, have higher rates of default.
Presumably, these are young people who were affected by the economic conditions of their early career.
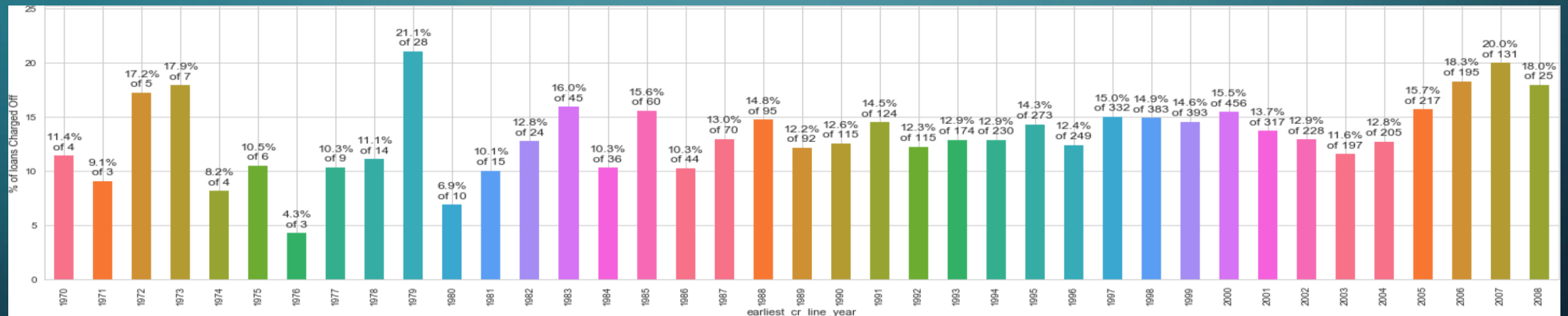
Conclusion

Results

➢ Low grade loans have high tendency to default.
Grading system is working as expected.

➢ Loans having higher interest rate have more defaulters. Check the background of applicant thoroughly if interest rate is high.

➢ Extra scrutiny must be done for the applicants belonging to CA state, as tendency to default is high.

➢ When the purpose is debt consolidation check applicant thoroughly as it has high tendency to default.

Thank You!