# PROJECT PRESENTATION ON LENDING CLUB CASE STUDY BY

## BY

### RADHIKA G
### PRATIBHA SRIVASTAVA
### (MLC 63 BATCH)

# PROBLEM STATEMENT

- To understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators.

- This knowledge shall be utilized by the company for its portfolio and risk assessment.

- The prime focus will be on the assessment of risk where 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 .

- Even though the customer is **likely to repay** , Lending Club rejects the loan.This could also be financial loss.

# ASSUMPTIONS

- As the prime focus is to understand differences in the features between loans that have been fully paid or charged off , we have not considered  loans that are current. So we only keep the loans with status "Fully Paid" or "Charged Off."

- This analysis is carried out on public data based on the loan attributes and consumer attributes .

# APPROACHES

- DATA CLEANING
  - The first step is to check for the null values and filter the data. There were 111 columns and 54 columns were completely NA. So we discarded the columns.
  - In the required columns , the NA rows were either deleted or imputed with median depending on the percentage of null values .
  - The string like '%' ,'months' and 'years' were deleted and variables were converted to numeric.
  - The columns which had only one unique value were dropped from analysis

# APPROACHES…contd

- SEGMENTATION
  - The data was further categorized into numerical , categorical and extra columns.
  - Further  2 new  dataframes were created –for Fully Paid customers and Charged off customers

# APPROACHES…contd

- UNIVARIATE ANALYSIS
  - The univariate analysis was carried out on categorical and numerical data for the 2 newly created data frames and they were compared to each other .This is carried out using boxplot and count plot
- BIVARIATE ANALYSIS
  - New variable labelled 'recovery_percentage' that showed the percentage of recovered amount was created from 'total amount' and 'funded amount' variables.
  - The numerical and categorical variables were analyzed for their relation with the recovery_percentage in the Charged off dataframe.
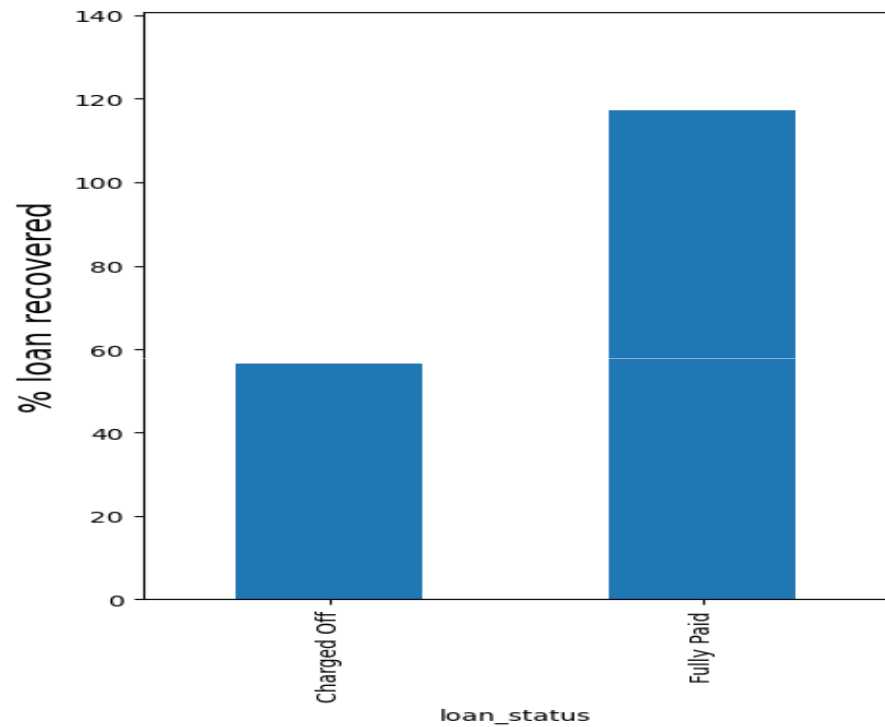  - This is carried out using histograms and boxplots.

# APPROACHES…contd

Loan amount buckets of 5 K were created and compared against the charged off counts.
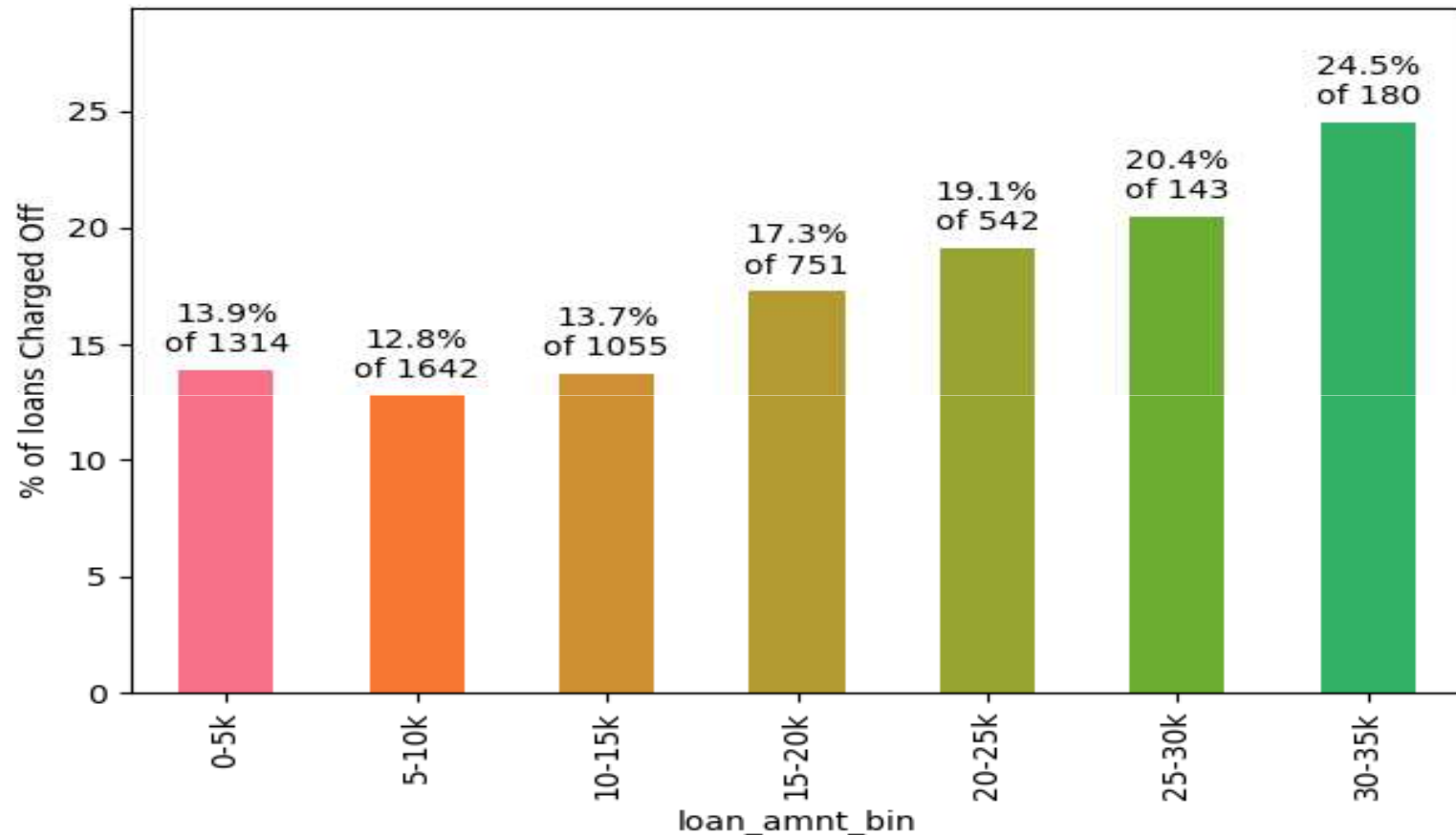
# RESULTS – UNIVARIATE ANALYSIS
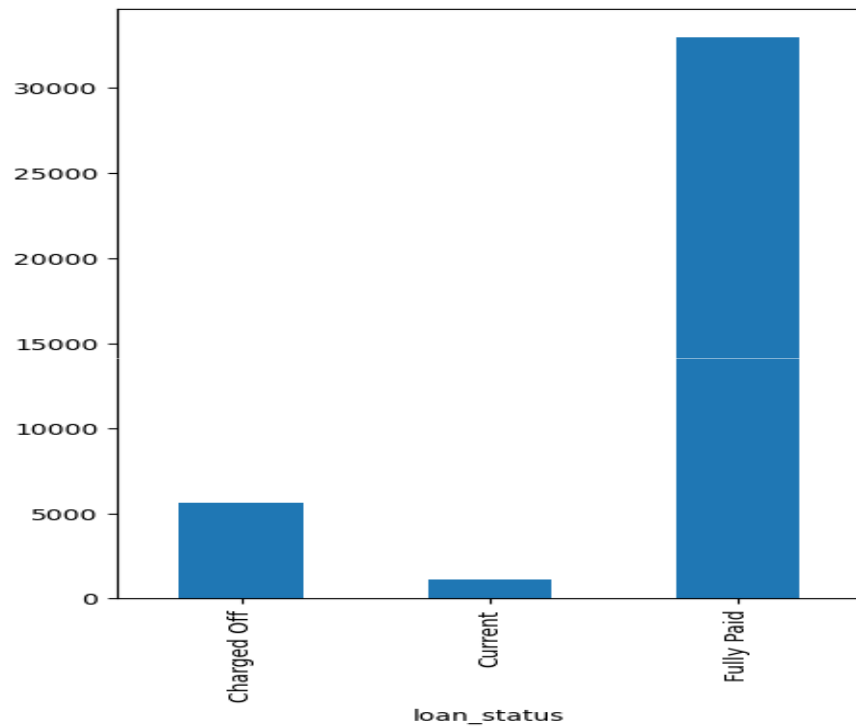


*14.17 % of loans in the dataset are defaulted*



*Lending Club only recovers 57% of the loan amount when loans are defaulted. On fully paid up loans, the company makes 17% profit.*
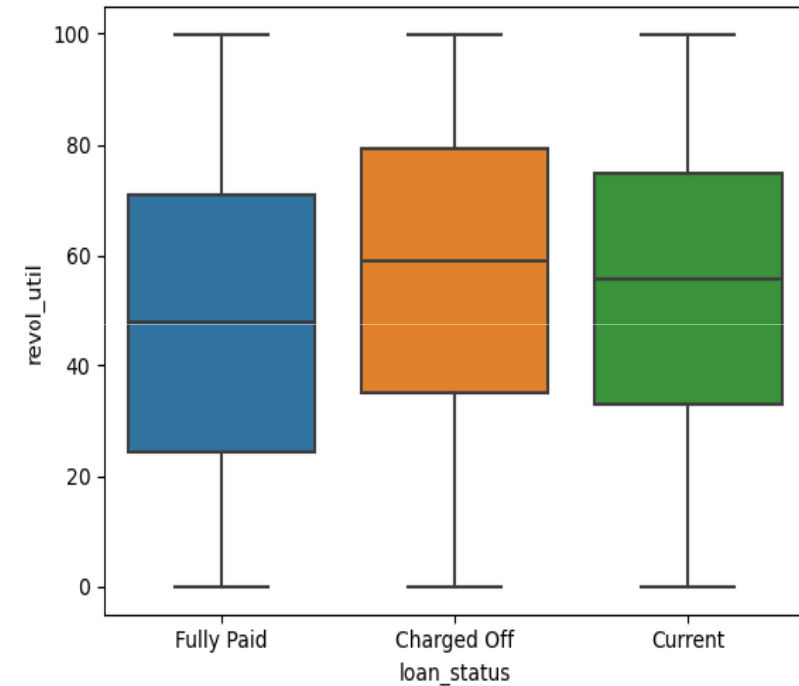
# RESULTS- BIVARIATE ANALYSIS



The % of charged off loans increases substantially as we go up the loan amount buckets. Most loans are below 20000 amount. The higher loans, though lesser in number, carry a substantially higher risk of default.
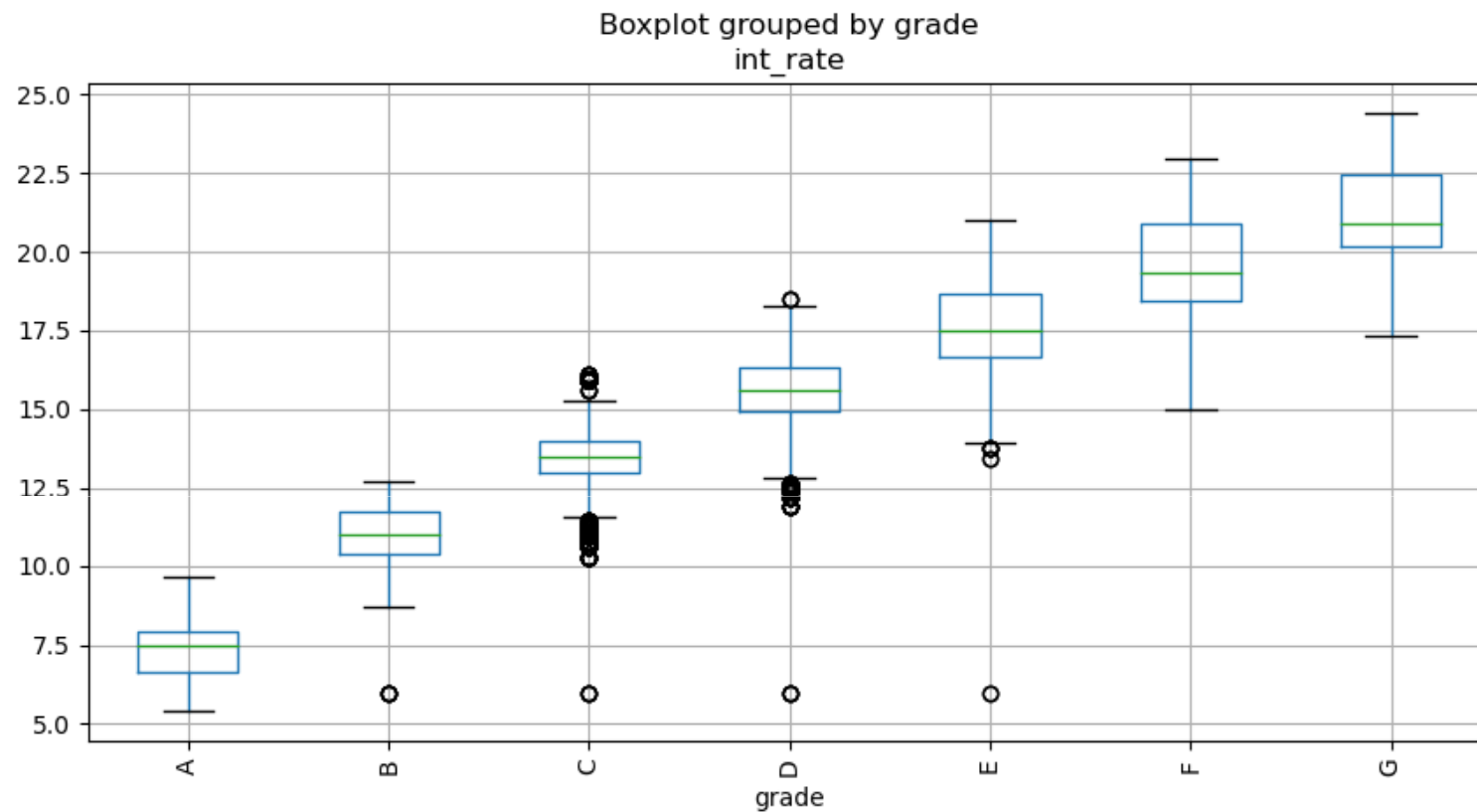
# RESULTS - BIVARIATE ANALYSIS
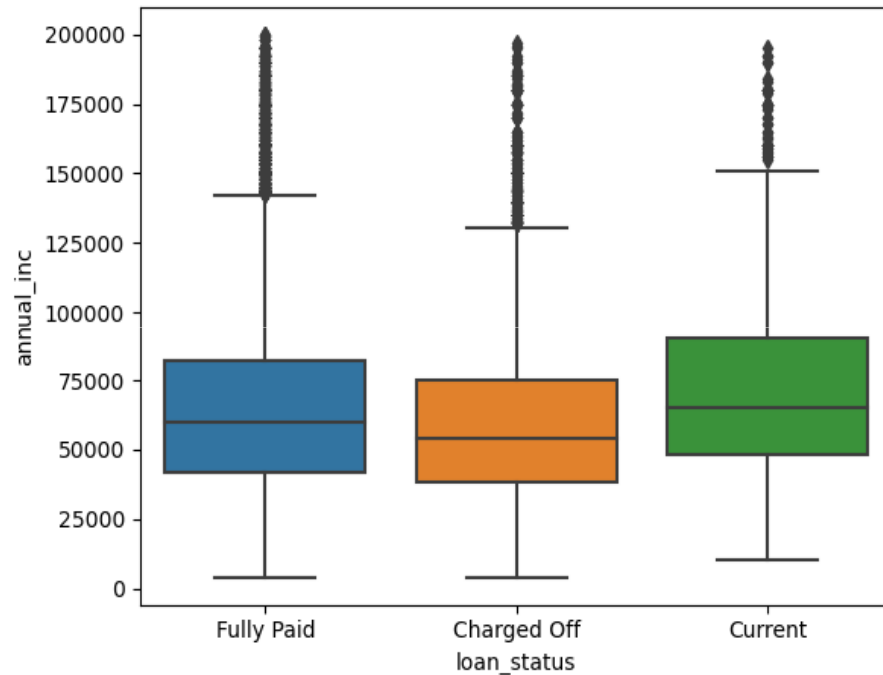


The annual income of ChargedOff customers is slightly lower than that of Fully Paid customers
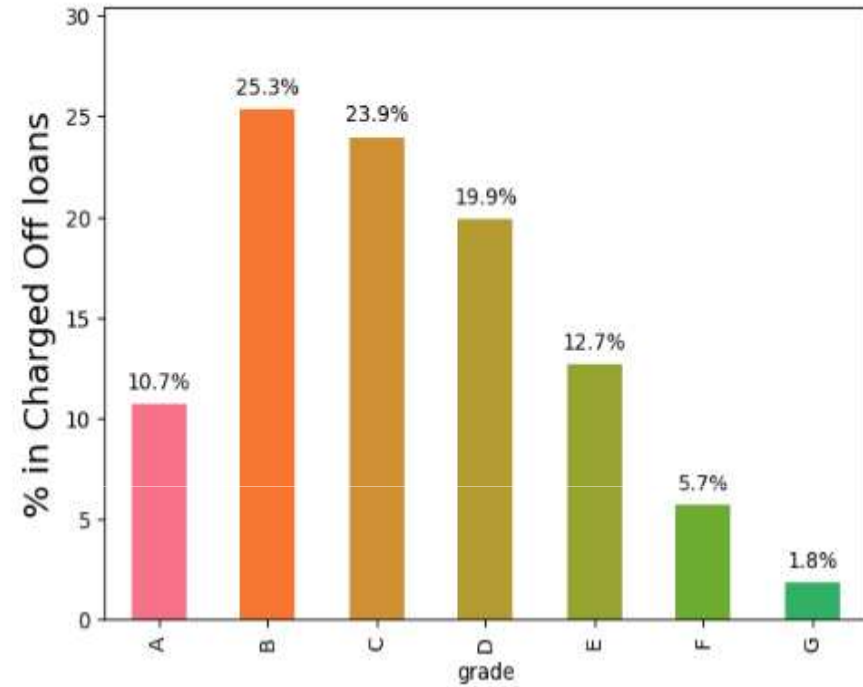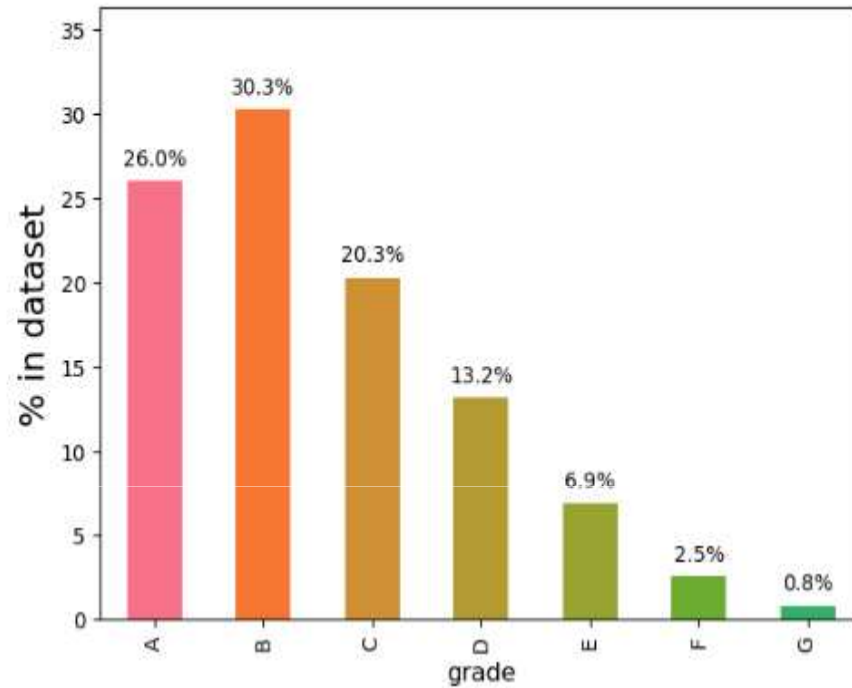
The revol_util of ChargedOff customers is higher than that of Fully Paid customers

# RESULTS OF BIVARIATE ANALYSIS



Boxplot grouped by grade
int_rate

We can observe lending club charging higher interest rates for lower grade loans

# RESULTS OF BIVARIATE ANALYSIS



- Grade A and B loans are safe.
- The percentages in full dataset are much higher than percentages in Charged Off loans.
- Grade D, E, F, G loans are less safe. We should plot grade by percentage Charged Off by category
- Lending Clubs grading system is working well.

# RESULTS OF BIVARIATE ANALYSIS
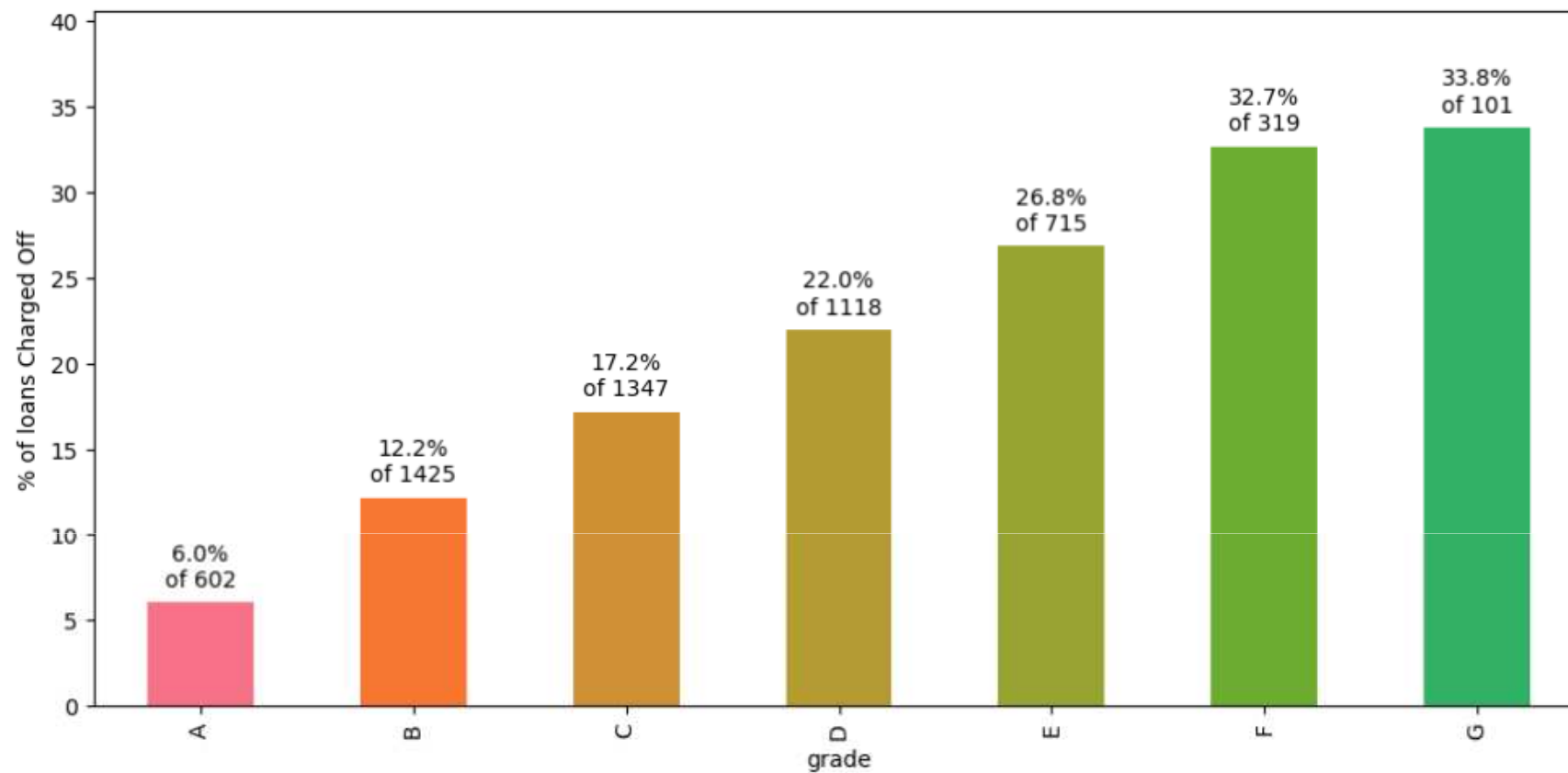


Observation:

Nearly 30% of all loans in Grades F and G see a default. Grade E onwards are risky, and less numerous. Lending Club should either refuse loans or charge high interest rates.

# CONCLUSION

- From the data set we can conclude that the annual income ,revol_util and Grade variables affect the loan_status variable .

- The lower annual income are more likely to default.

- Grade E onwards are risky, and less numerous. Lending Club should either refuse loans or charge high interest rates.

- The revol_util is observed to be higher for Charged_off customers as compared to that of Fully Paid customers.