# Lending case study

By: Surya Bandari

# Problem Statement

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

- Now this would require a preliminary understanding of the features that define whether applicant's loan should be accepted or rejected.

# Analysis approach

- Understanding data

Perform Data cleaning

- Remove irrelevant columns
- Remove null columns and rows
- Check for null values
- Imputation of null values
- Finding outliers with box plot

Univariate analysis

- Univariate analysis on Continuous variables
- Univariate analysis on Categorical variables

Segmented and Business driven analysis

- Segmentation based on defaulter across different variables

Bivariate analysis

- Bivariate analysis on Continuous vs Continuous variables
- Bivariate analysis on Categorical vs Continuous variables

Derived matric analysis

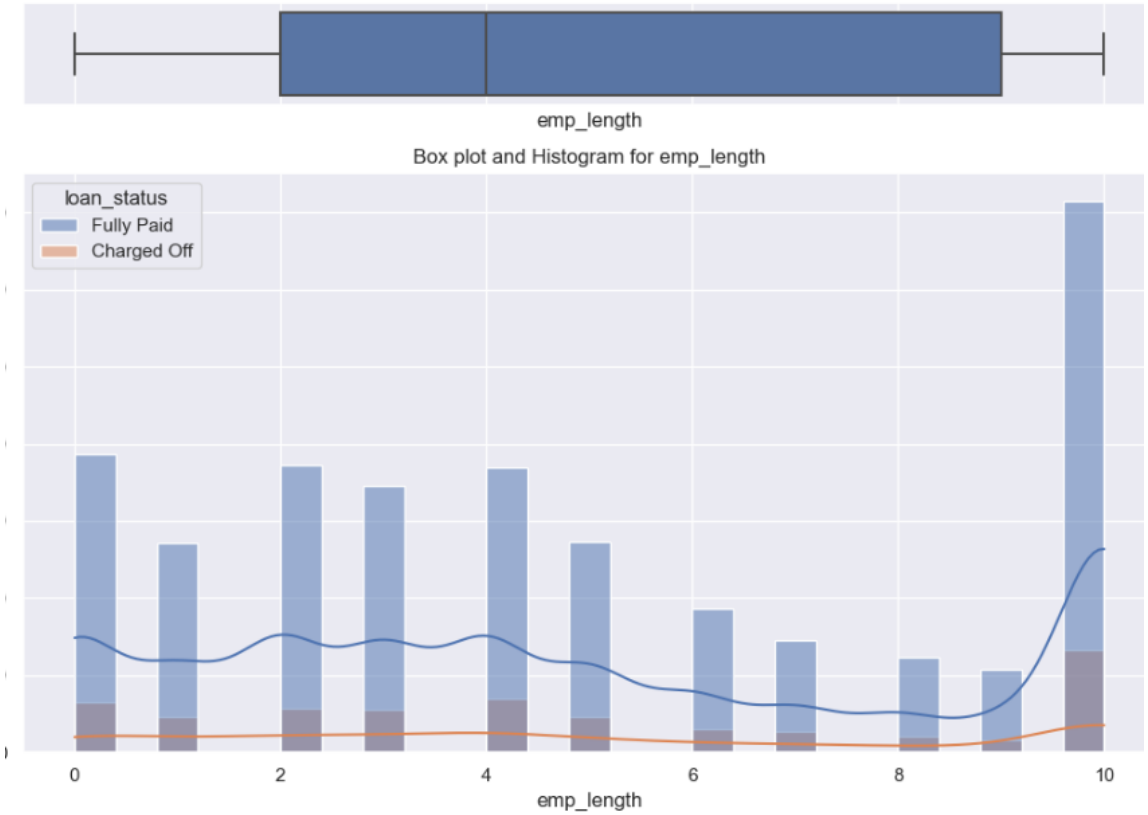
- Type driven analysis
- Data driven analysis

Conclusion

# Data Understanding and Data cleaning

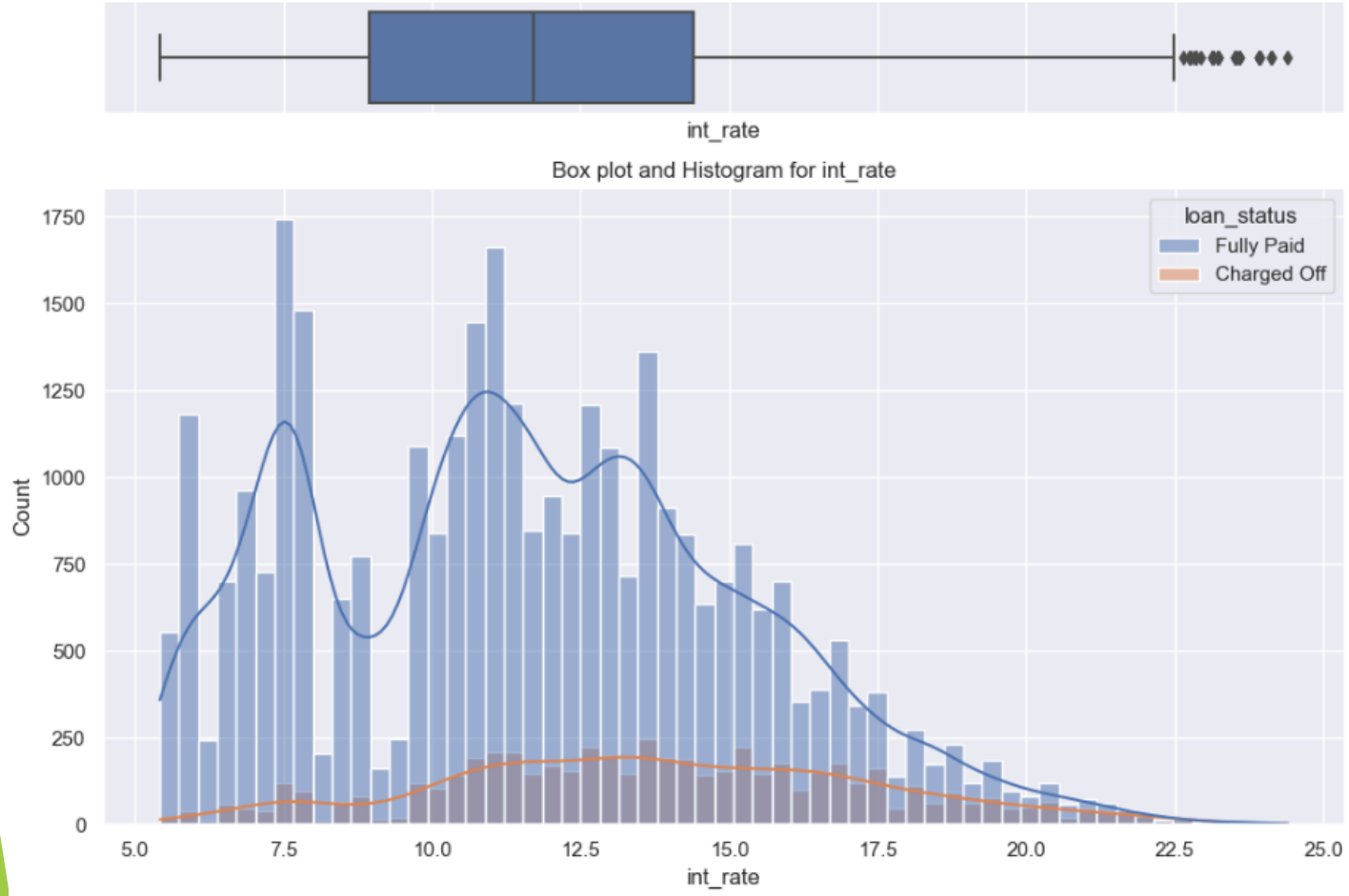
- Analysis performed on the data set with 39717 applicants with 111 attributes
- 111 attributes includes:
  - applicant relevant information
  - Loan characteristics
  - Customer behavior variables
- Through the analysis, most of the columns are dropped to identify the real driving factors(or driver variables) behind loan default
- Used appropriate imputation technique to fill out the missing values.

# Univariate on Tenure or Work experience



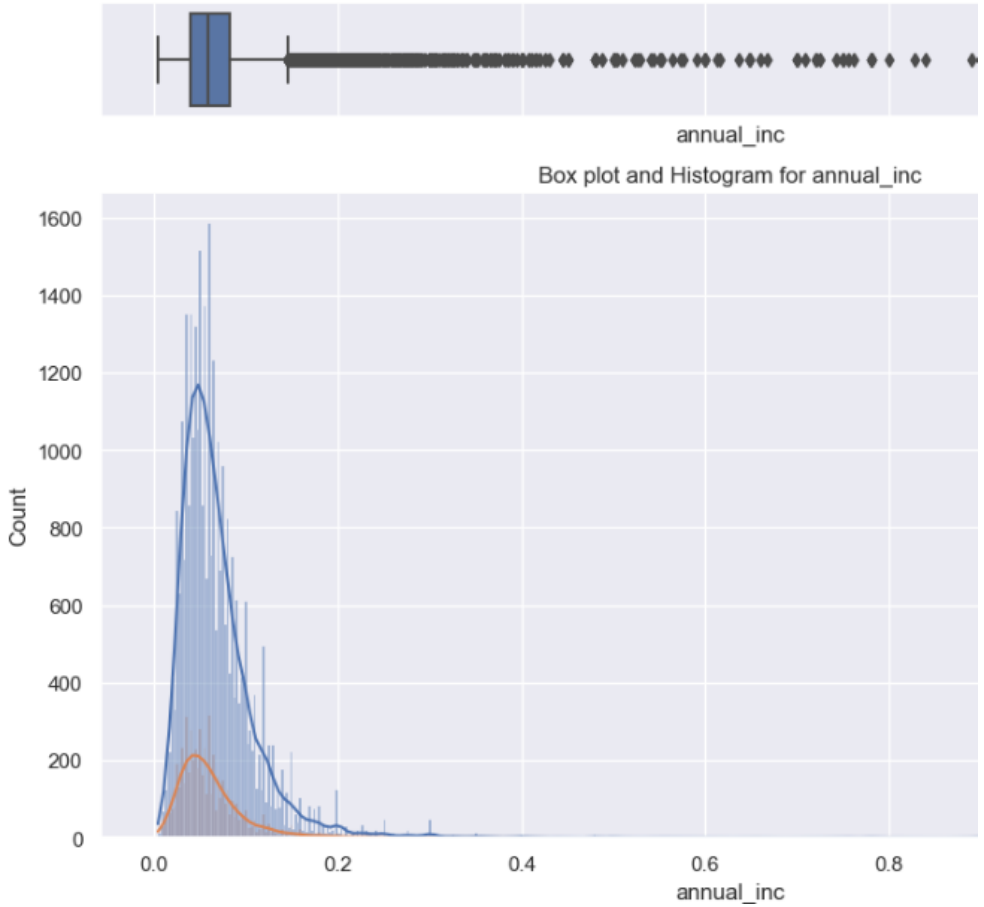

- Most of the loan applicants are having 10 or more years of work experience. So mostly they are in 30s

# Univariate on "Interest rate"



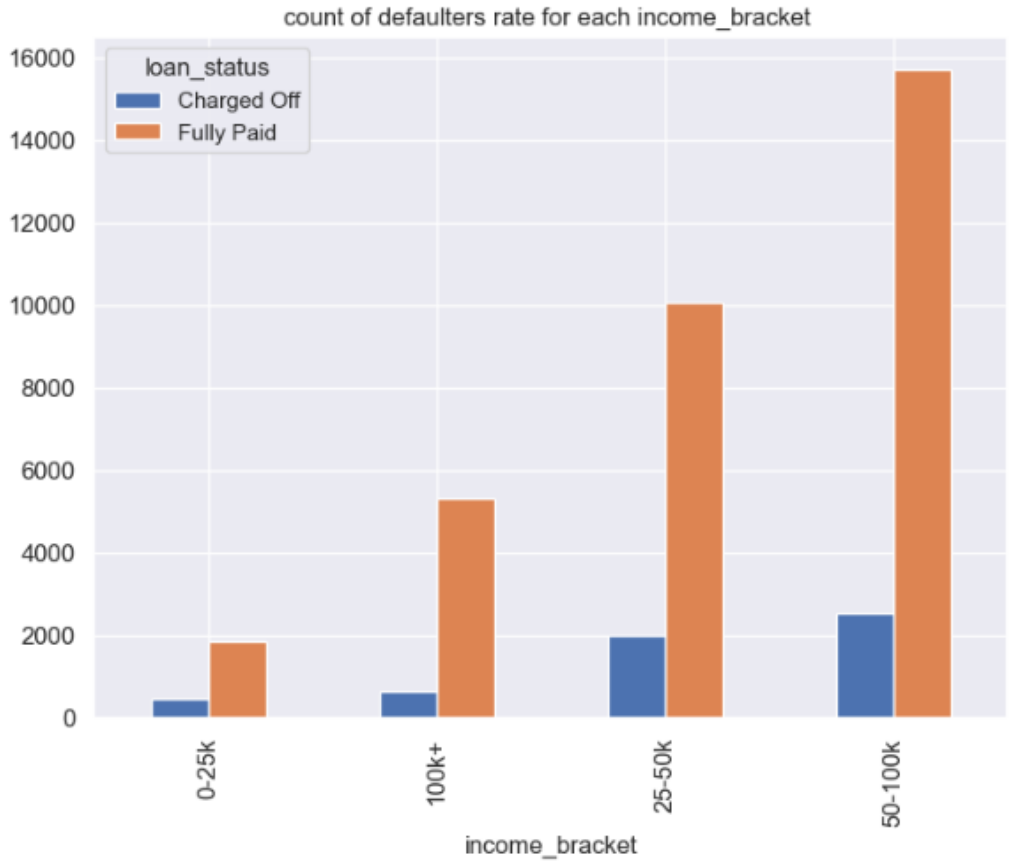

- Interest rate is mostly falls between 10-14.5%

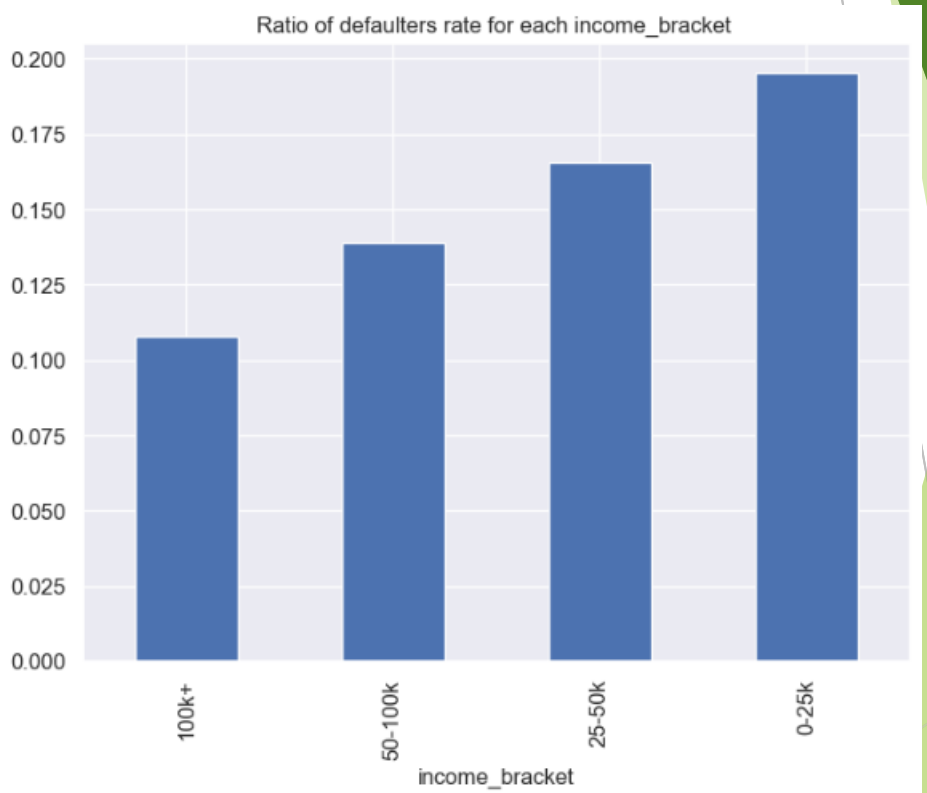
# Univariate on annual income




HIGH PORTION OF THE APPLICANTS ARE HAVING ANNUAL INCOME BETWEEN 50K-100K.

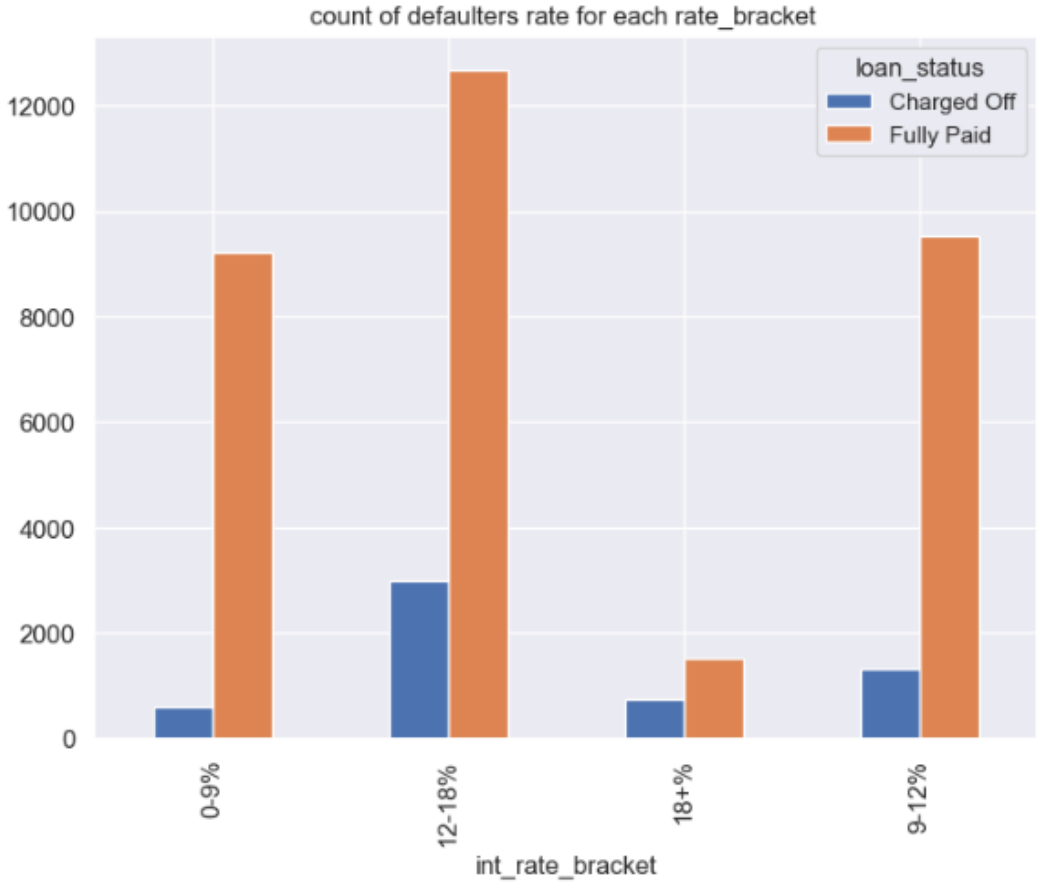
# Applicants defaulted based on Annual income

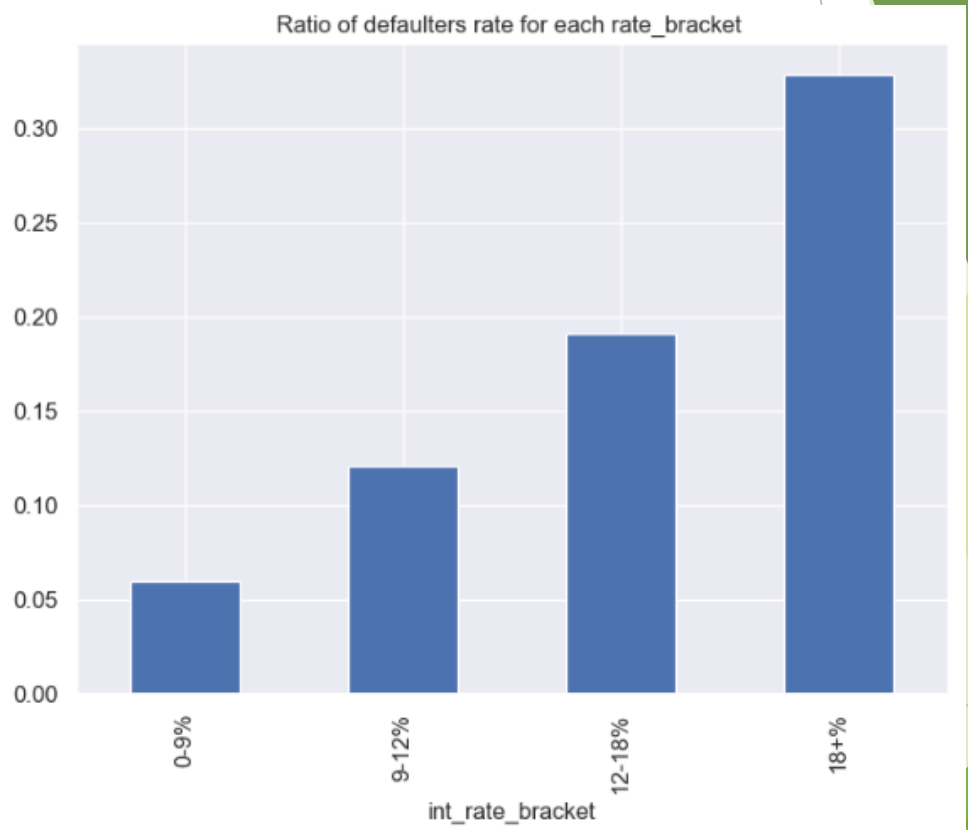



- Most of the defaulters are in 0-25k bracket

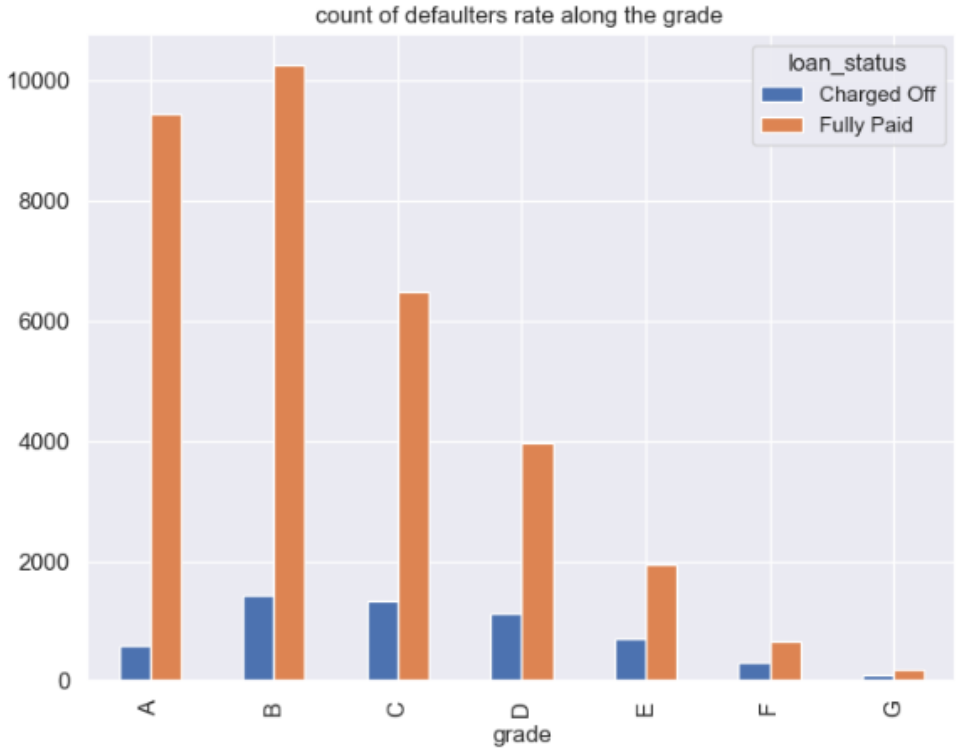
# Applicants defaulted based on Interest rate

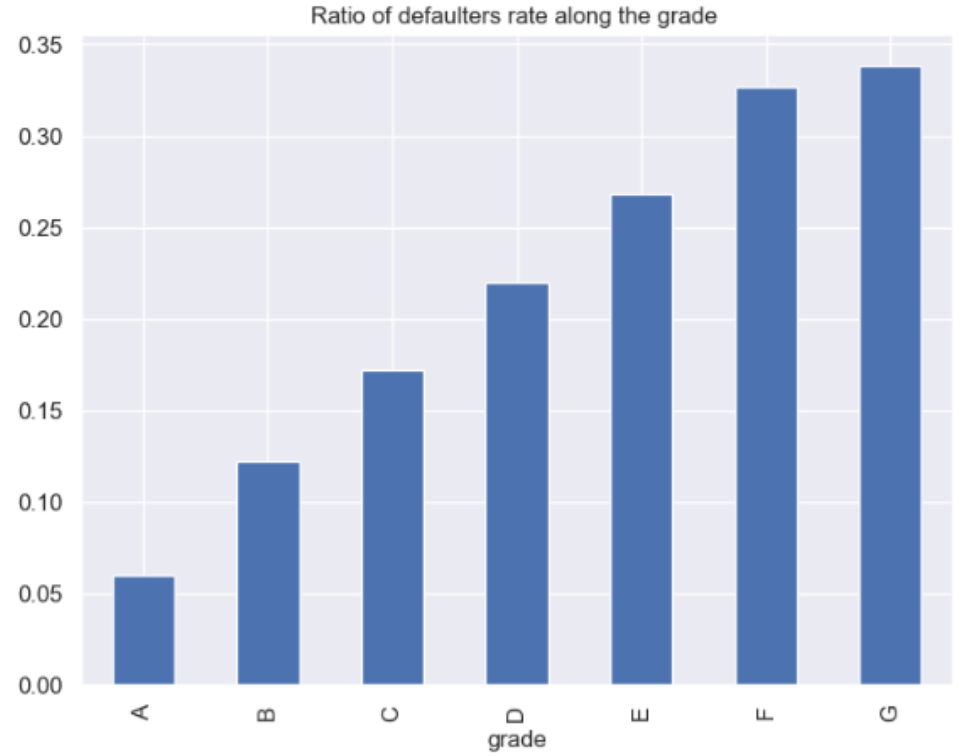



- Clearly there are many defaulters in 18+% interest rate bracket

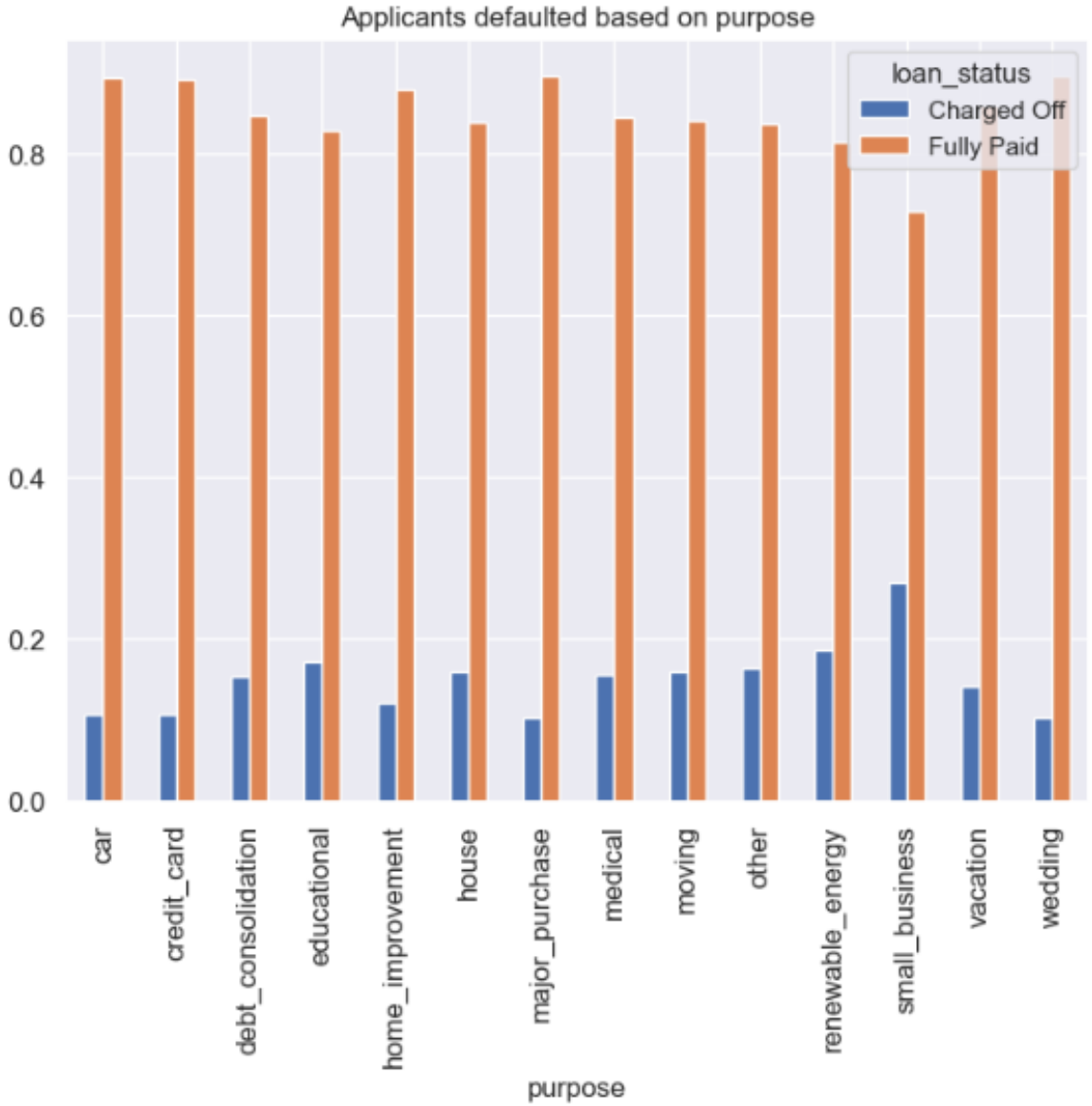
# Applicants defaulted based on Grade assigned



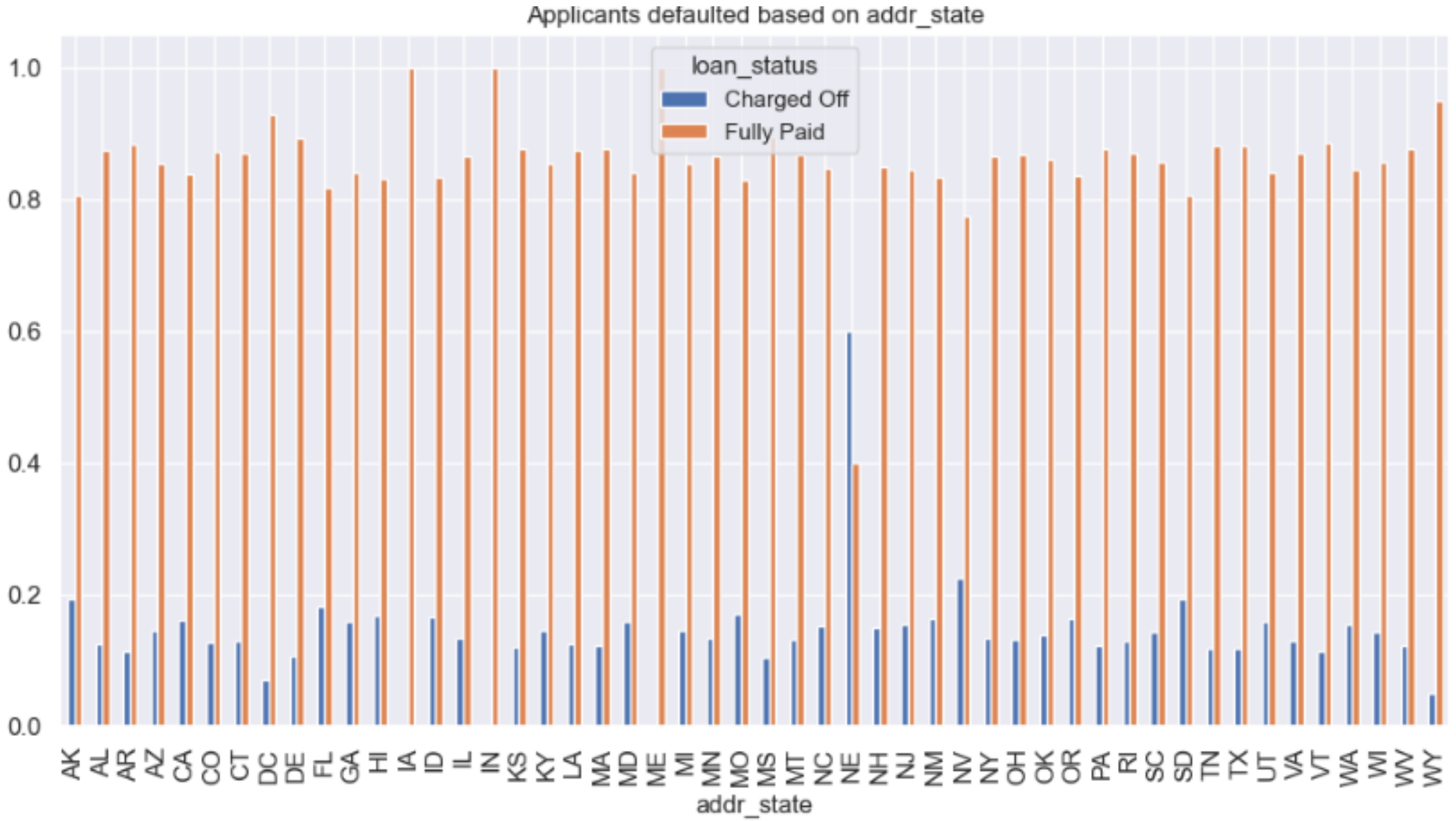

- Defaulters increase with the grade but the count of F and G grade records are very less.

# Applicants defaulted based on purpose stated



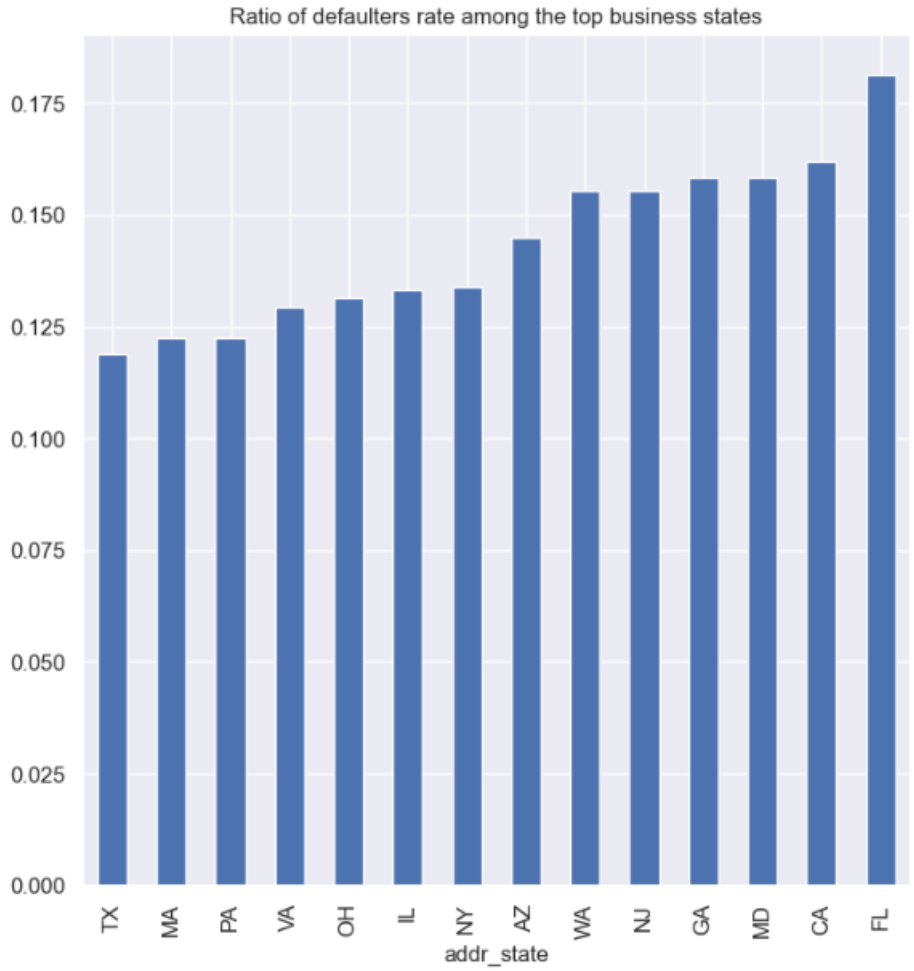

- Ratio of defaulters are more in "small business".

# Applicants defaulted based on Address State



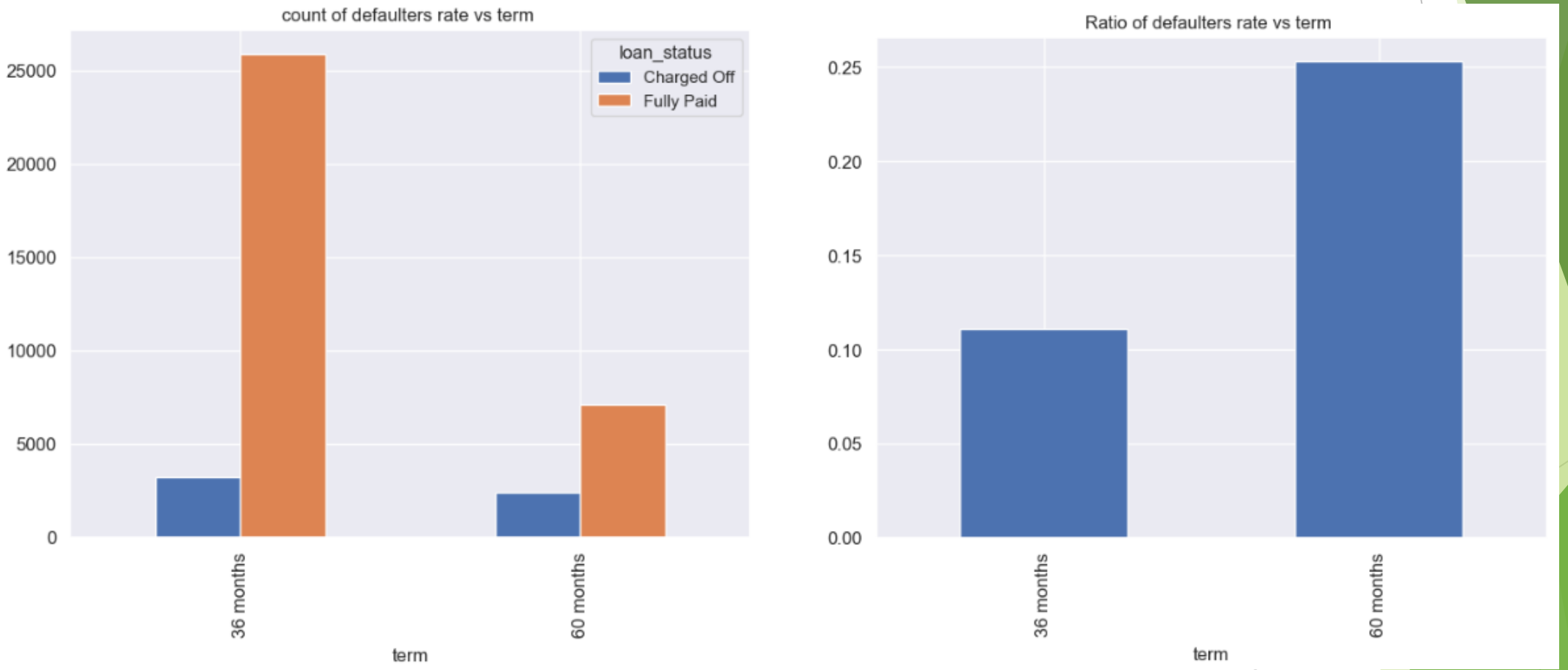

- Initial observation: High defaulters in NE and IA/ME/IN is good business area with zero defaulters. Conti…

# Conti: Applicants defaulted based on Address State



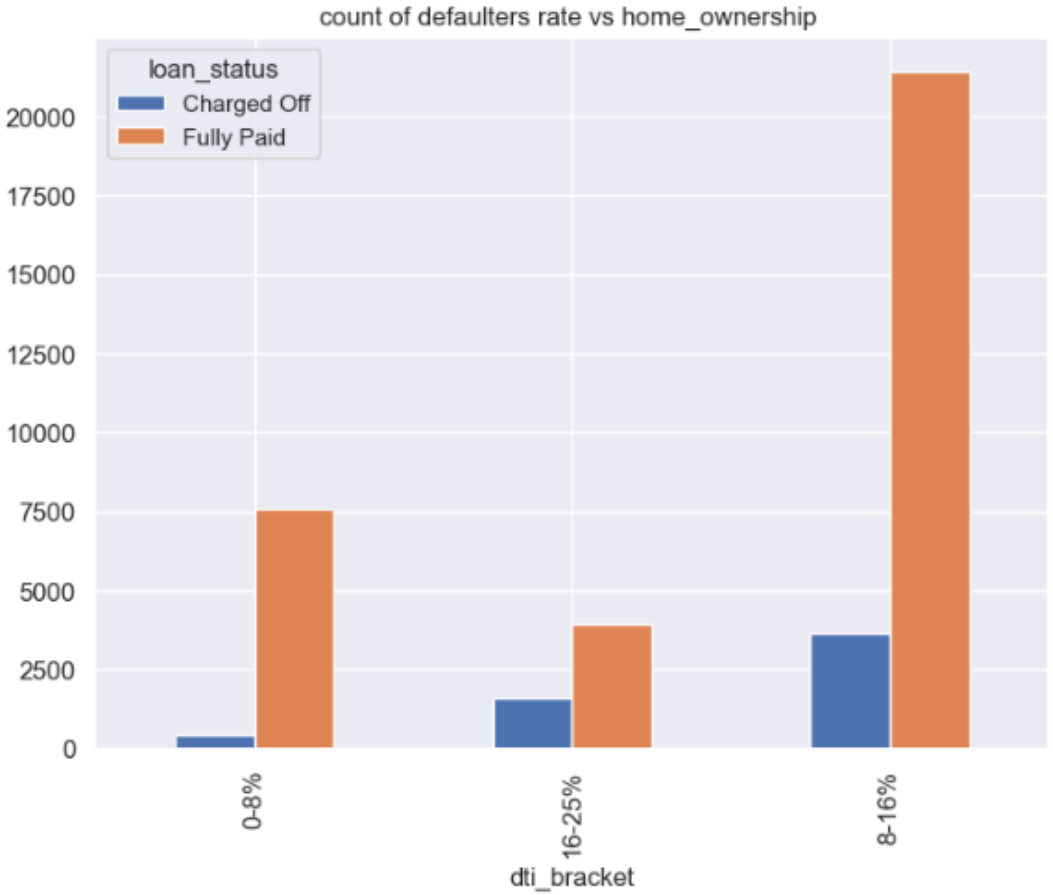

- It is observed that applicants from NE/ IA/ME/IN region are in single digital.
- On continuing further analysis on top State with maximum applicant, it is found that FL is the worst performing state.

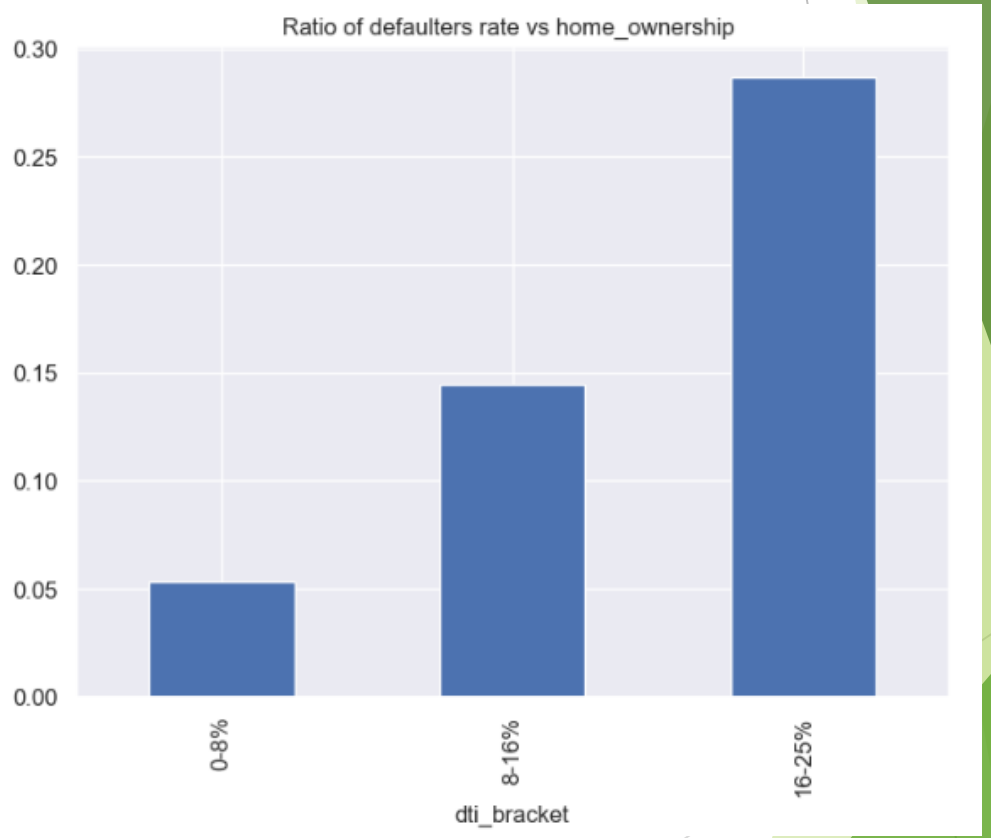
# Applicants defaulted based on Term selected




- Higher default rate among 60 months term

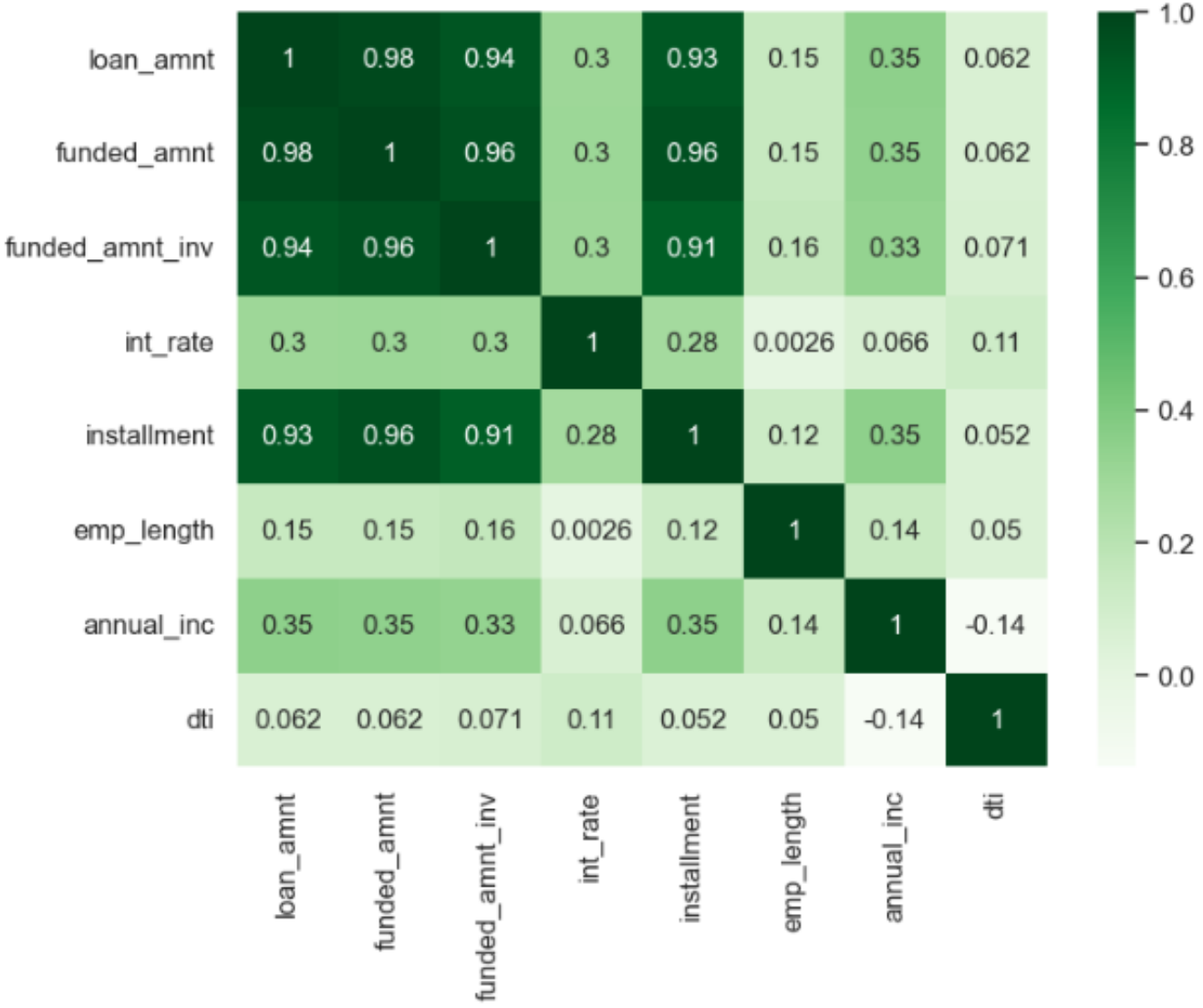
# Applicants defaulted based on dti



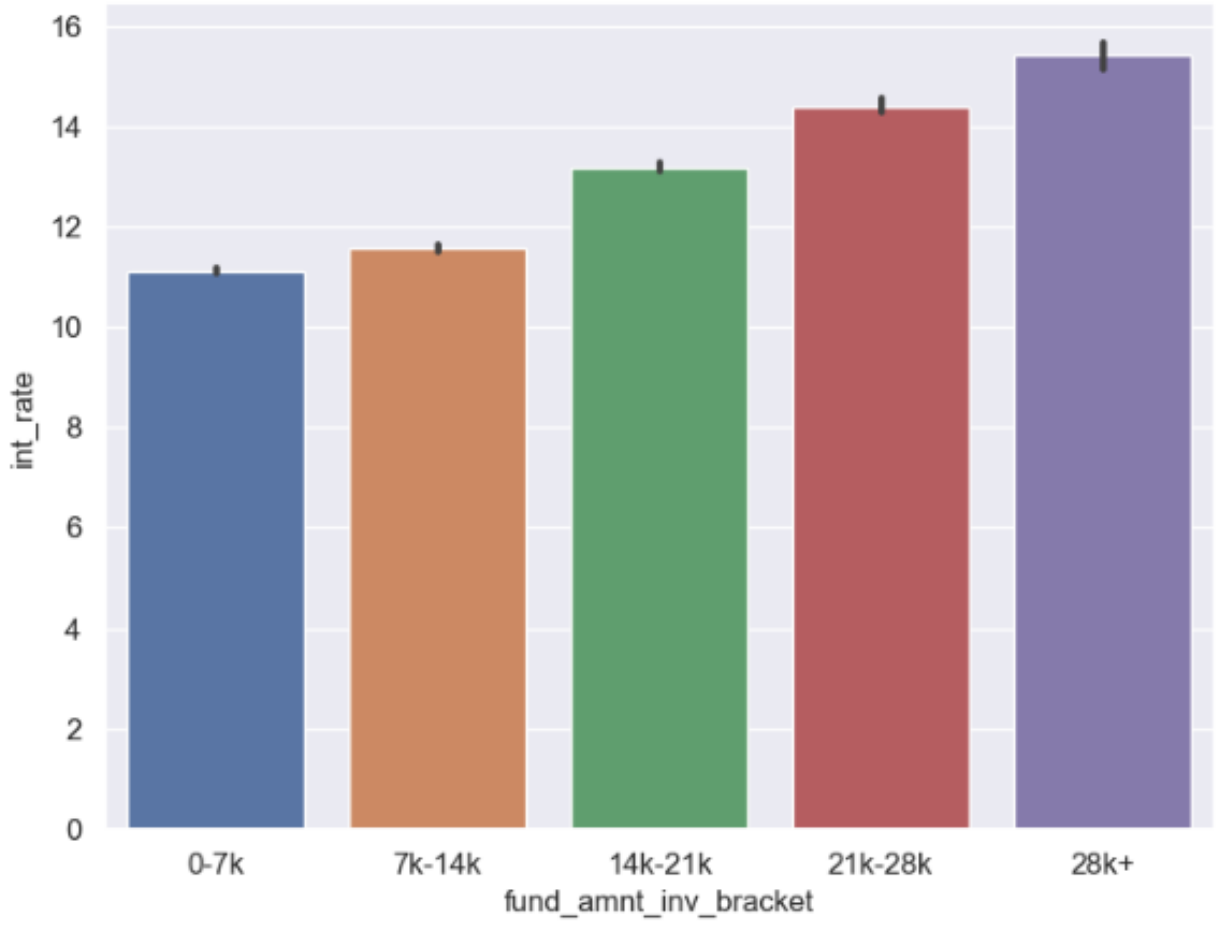

- Defaulters increases with dti score

# Bivariate analysis – between continuous columns



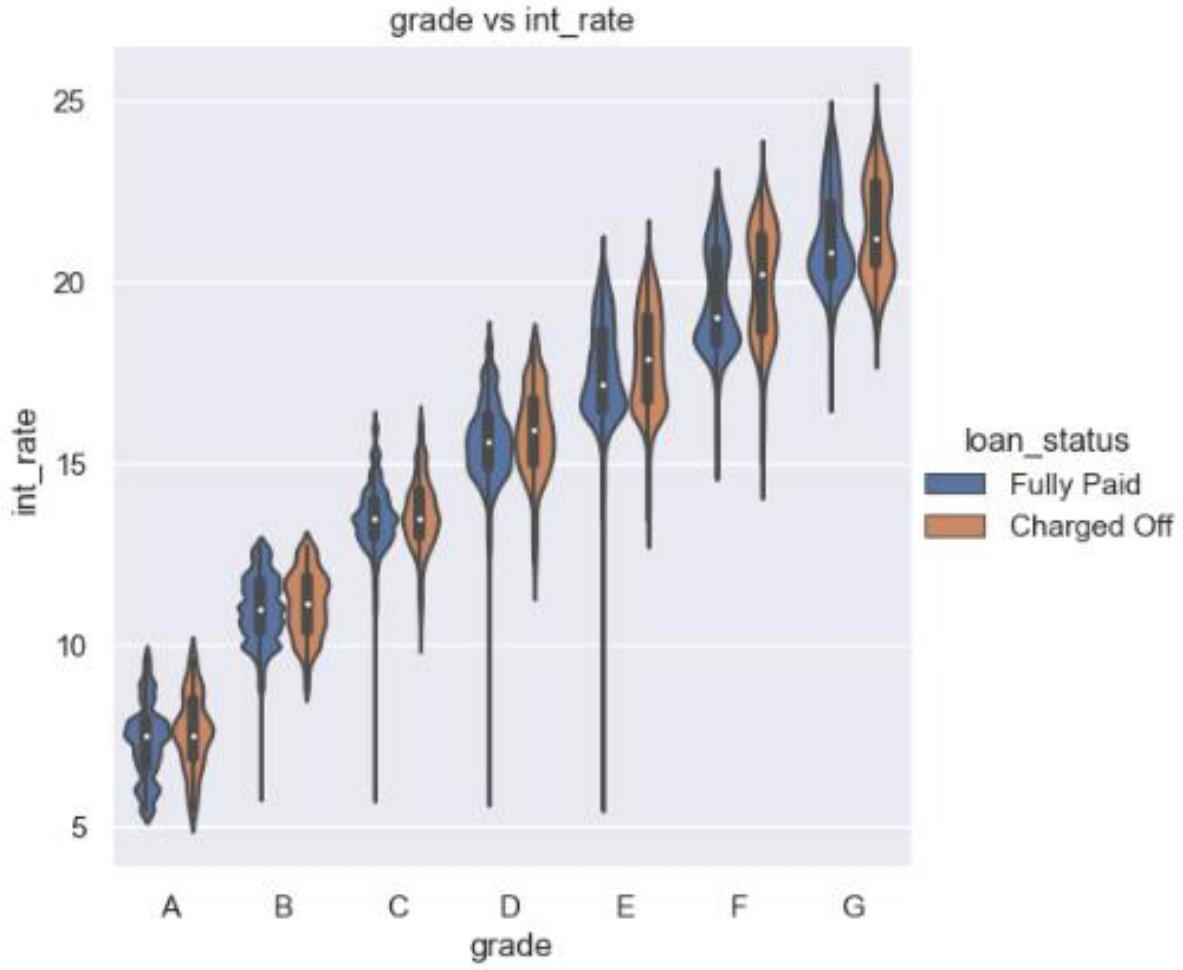

- Installment has string correlation with the amount columns which is obvious.
- Negative co-relation between annual_inc and dti
- Relatively positive co-relation between annual_inc to loan_amnt.
- As observed before, strong correlation between loan amount/funded amount /funded amount invested
- Also 0.3 positive cor-relation of int_rate with loan amnt.

# Bivariate analysis – Loan amount vs Interest rate



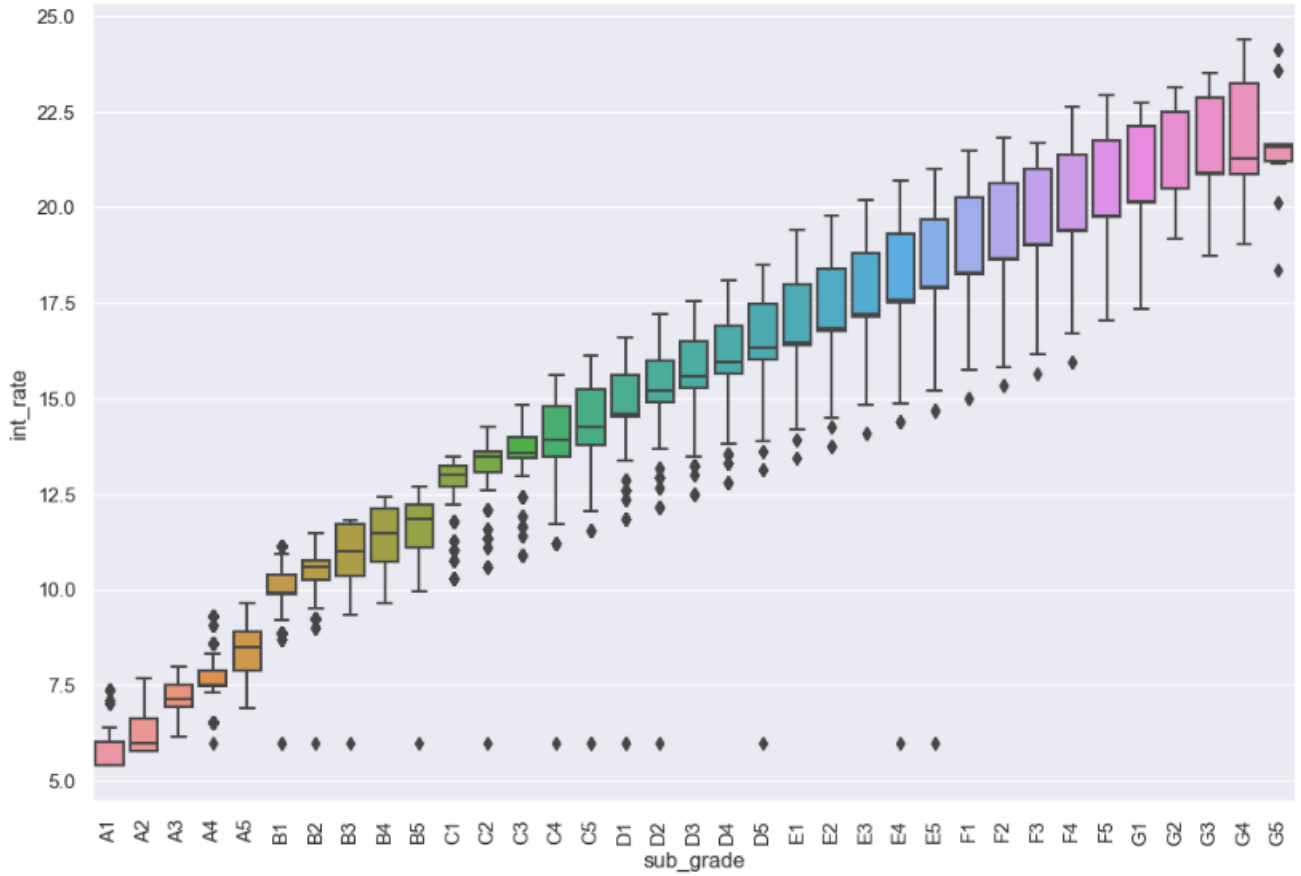

- As already observed in heatmap, the interest rate increase with the loan amount related columns

# Bivariate analysis – grade vs Interest rate



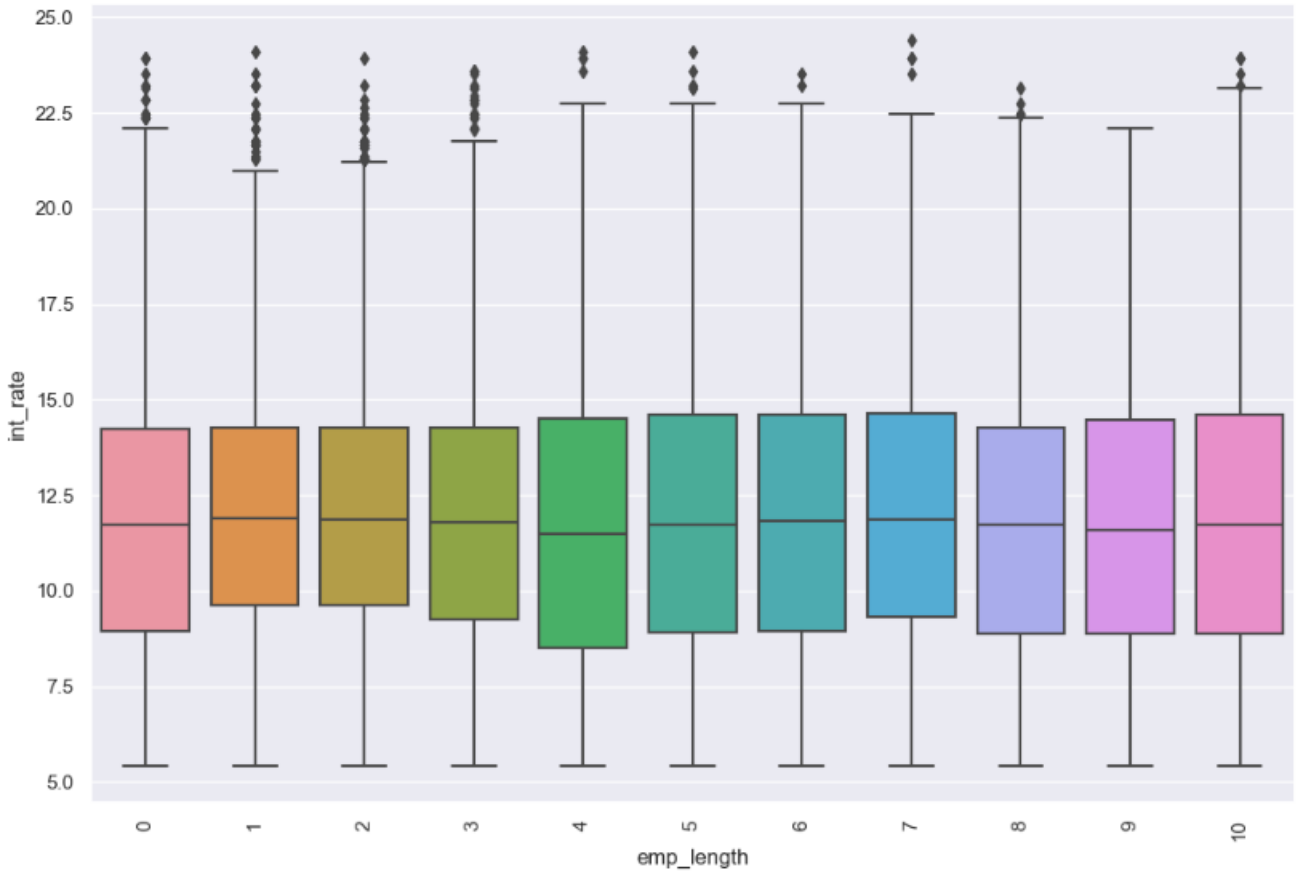

- Interest rate increase with grade A->G

# Bivariate analysis – subgrade vs Interest rate



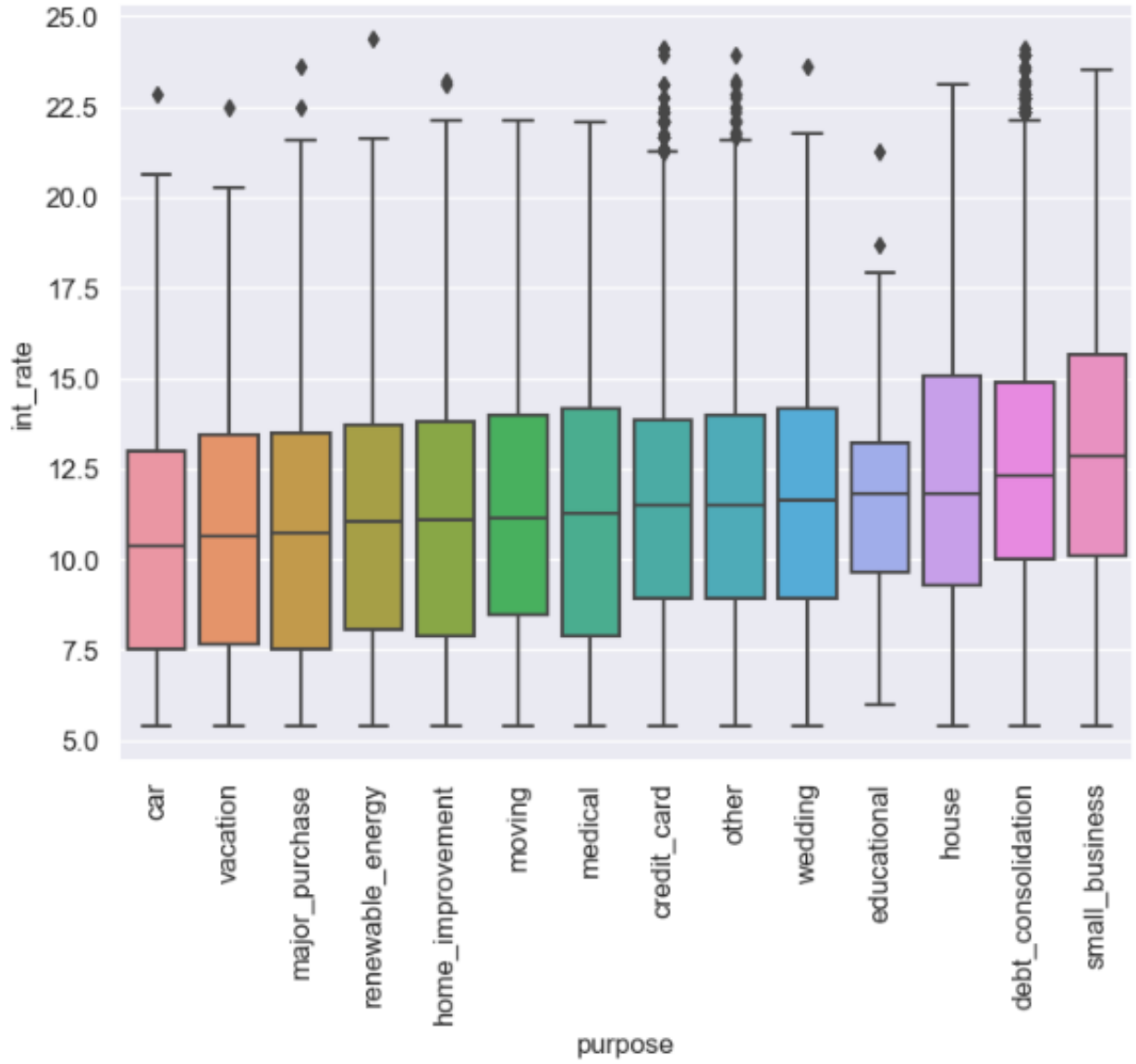

- Same as grade, Interest rate increase with sub grade Ax-->Gx

# Bivariate analysis – tenure vs Interest rate



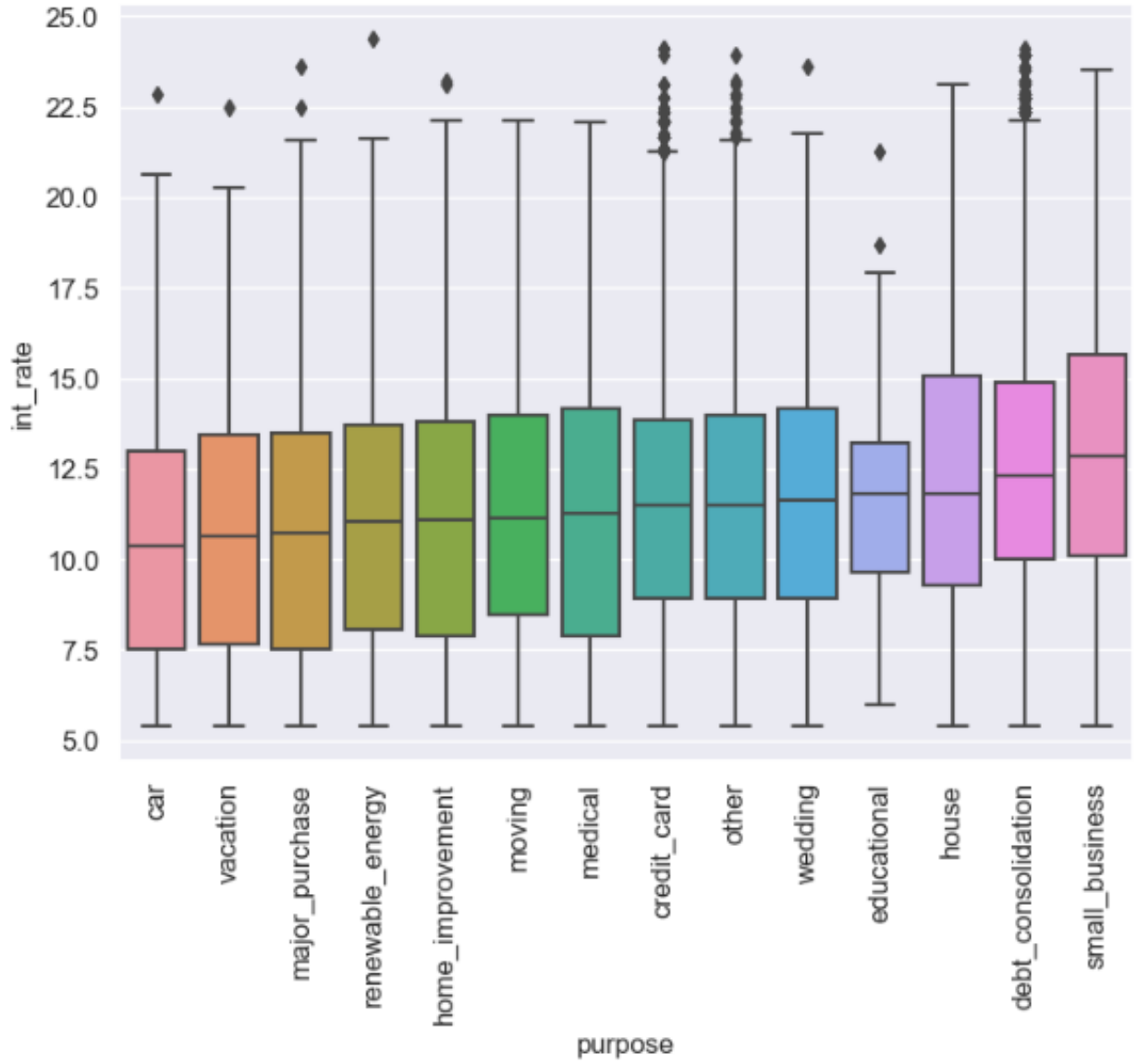

- Interest rate  is mostly same across all the applicant tenure.

# Bivariate analysis – purpose vs Interest rate



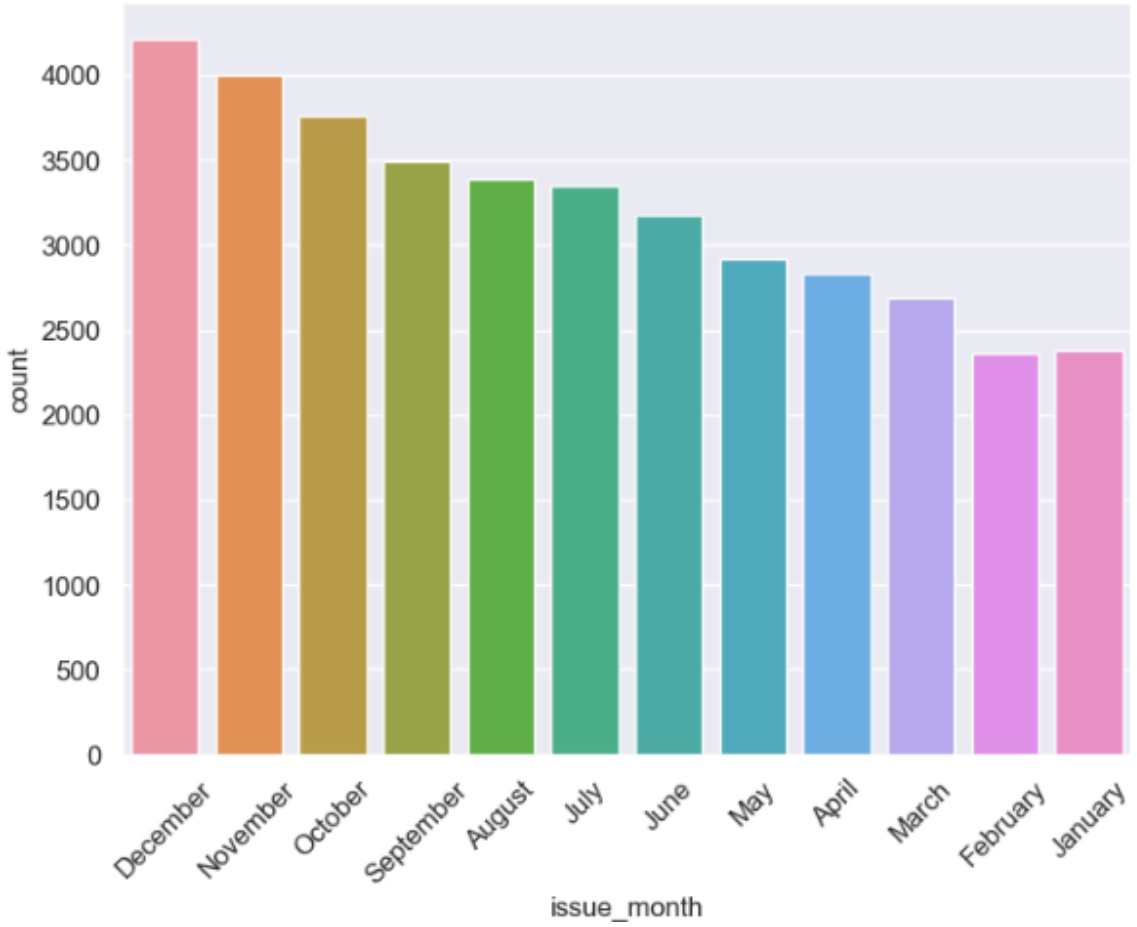

- Clearyly the 'small business' have higher interest rates followed by debt_consolidation

# Bivariate analysis – purpose vs Interest rate



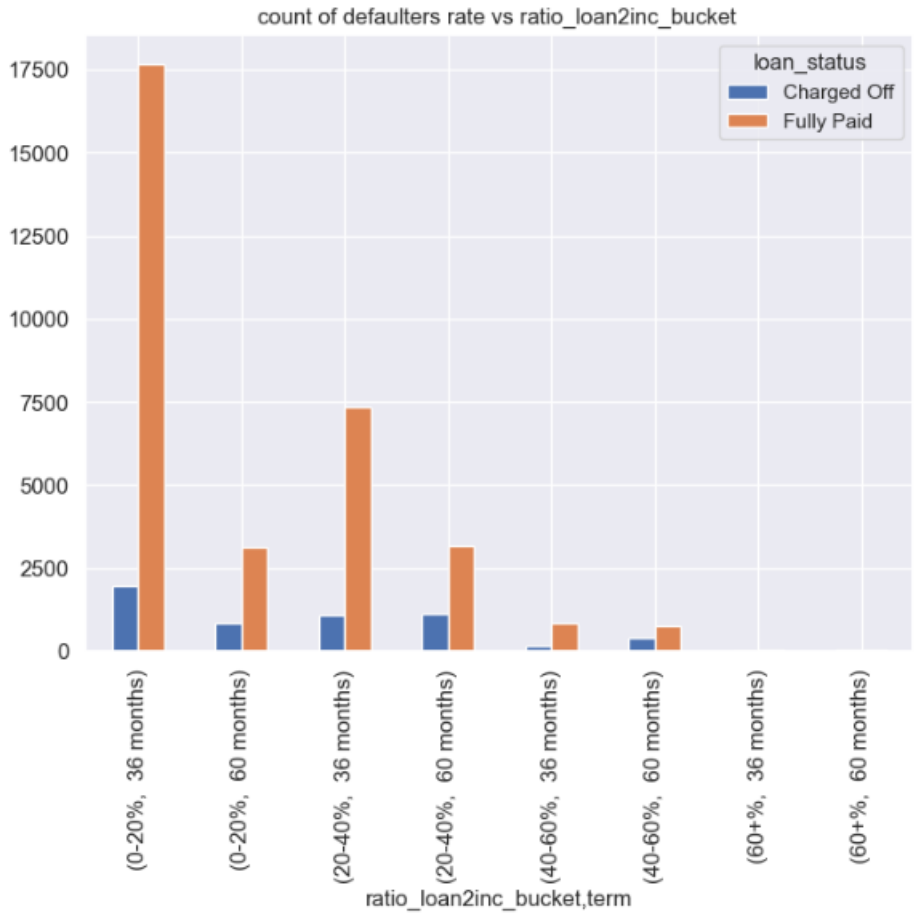

- Clearyly the 'small business' have higher interest rates followed by debt_consolidation

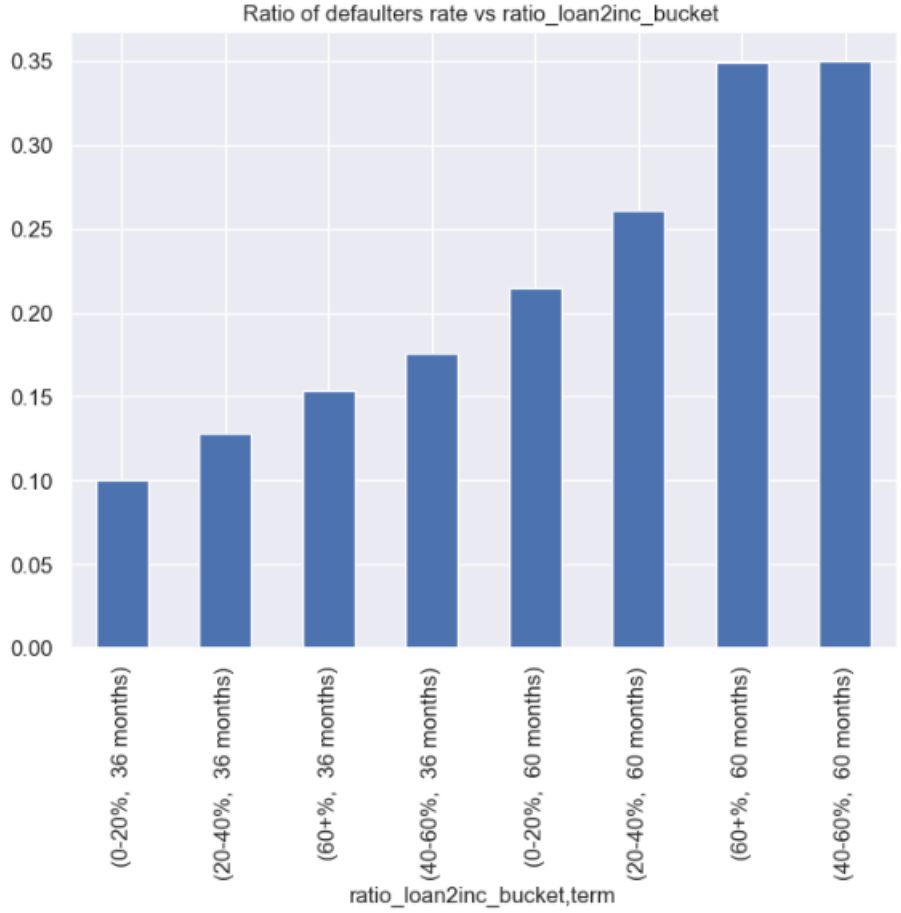
# Derived metrics – Month from issue date




- Application for loan increases towards year end. Could be because of vacation period.

# Derived metrics – Defaulters rate to "Portion of loan amount to annual income"




- Interesting insight that as the ratio of loan amount to annual income increase, the defaulters count increases.

# Conclusions:

- Annual income - Applicant with lesser income tend to default especially in the range of 0-25K. Income with more than 100K are good and 50-100K are moderate.
- Term – Rate of defaulters increase with the term. As the length of period is more, there could be many factors coming with time which could give more chance for defaulting.
- Interest rate - Applicant with interest rate greater than 15% are more likely to default.
- Home ownership – Home ownership is relatively not affecting the defaulters count.
- DTI: Applicant with lower DTI score are preferrable. Customer falling under 16-25% highly risky applicants.
- Grade – It is clear from analysis that with the loan amount, grade increases, which inturn increase the interest rate, so LC should be careful while granting loans with higher grade(A→G)
- Sub-grade: Same as grade
- Address State: Among the top business states(based on maximum applicants), highest defaulters are from FL. Order: FL>CA>MD>GA>NJ>WA>AZ>NY>IL>OH>VA>PA>MA>TX
- Purpose: LC should be careful while granding loan to "small business" have higher defaulters.
- Portion of loan amount to annual income: As per analysis, LC also should make sure the portion of the loan amount doesnot exceed 20-40% of the annual salary irrespective of term.