# LENDING CLUB EDA CASE STUDY

KAPIL BHATT

OBJECTIVES

- Aim of this project is to identify predictors of default so during loan application we can use those variables for loan approval/rejection.

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

- So the company can utilize this knowledge for its portfolio and risk assessment.
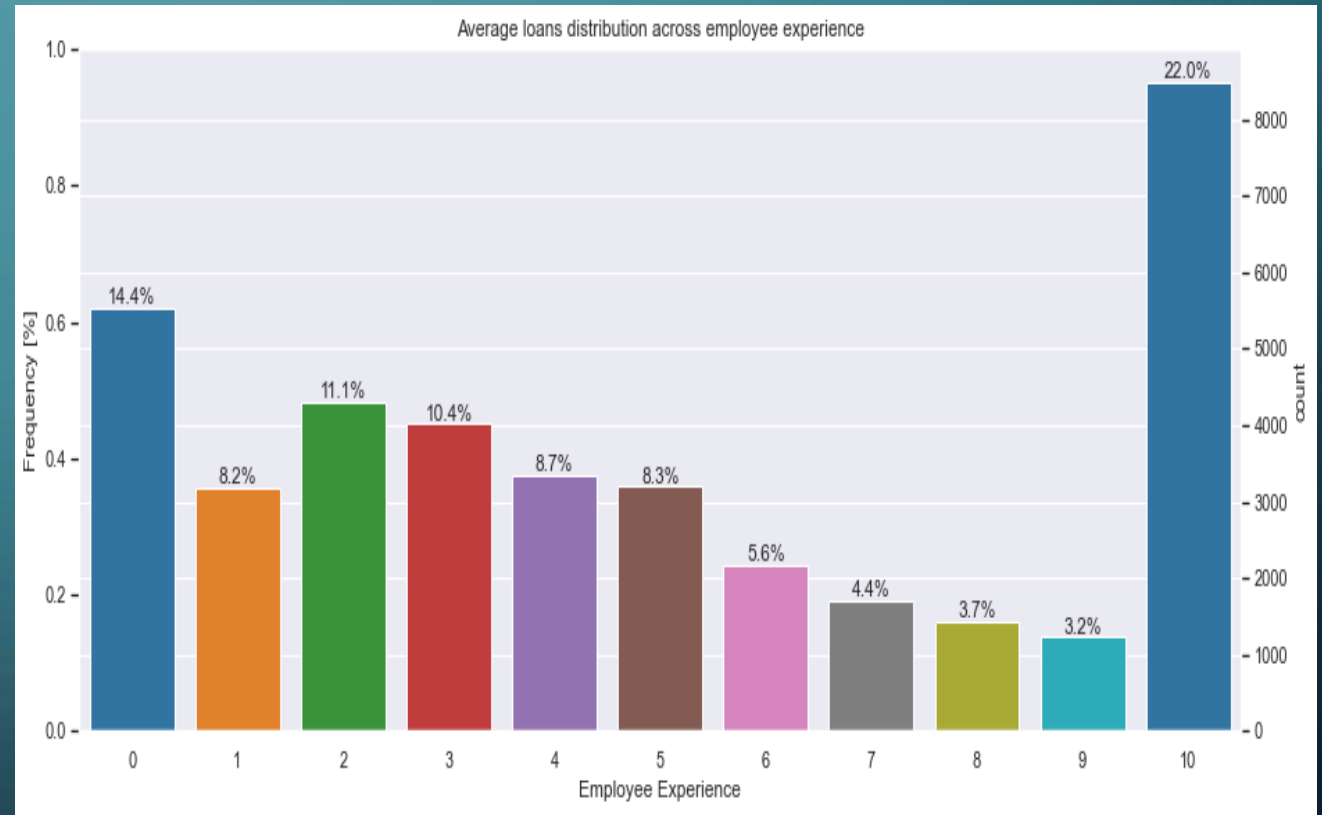
AVERAGE DEFAULT RATE

- Graphs showing the Loan Status between 'Fully Paid' and 'Charged Off' category.

- **Average default rate across all categories is 14.6%.**
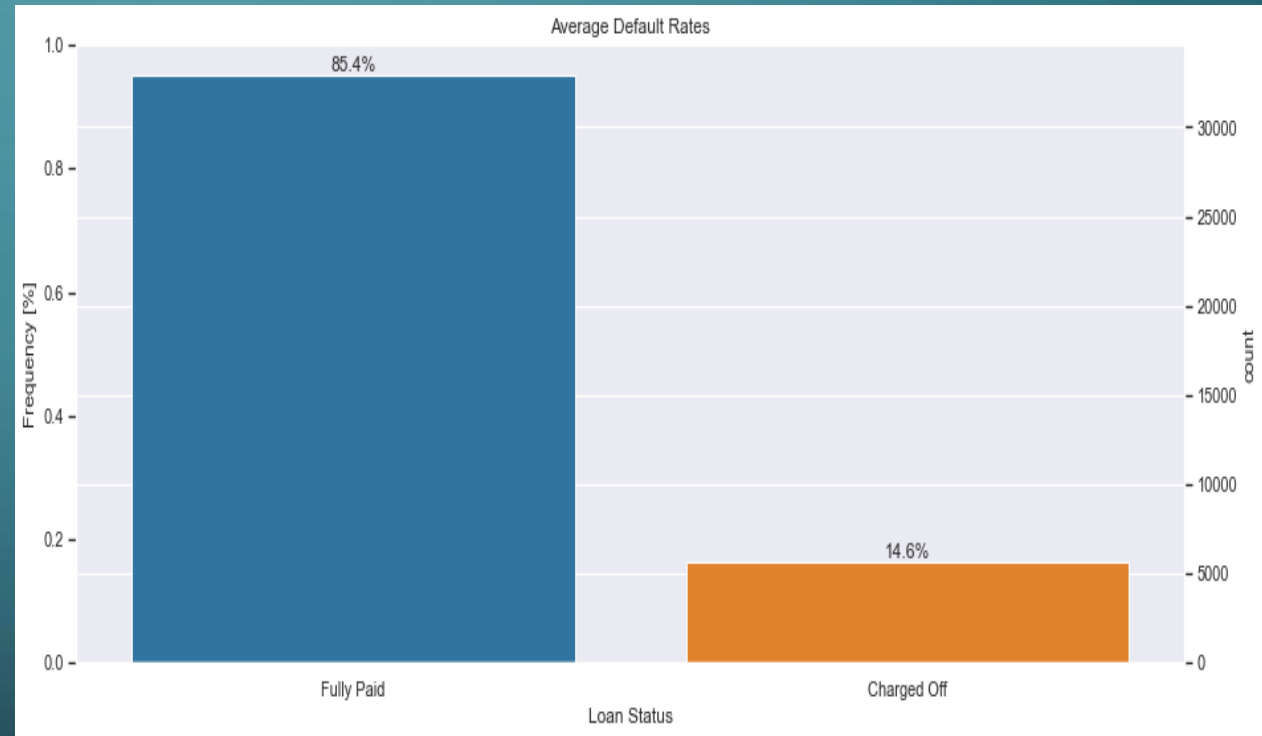
AVERAGE LOANS DISTRIBUTION ACROSS EMPLOYEE EXPERIENCE

- **Peoples having experience of 1 year and 10 years are taking loans more frequently compare to others.**

IMPORTANT VARIABLES FOR ANALYSIS

- **Target variables:** Loan characteristics: Loan amount, interest rate, purpose of loan etc.

- For Loan status having values as 'current' are neither fully paid or defaulted, So we can remove them and only use 'Fully paid' and 'charger off'.

- House ownership and verified source of income affect the analysis so they can give insights about loan defaulting.
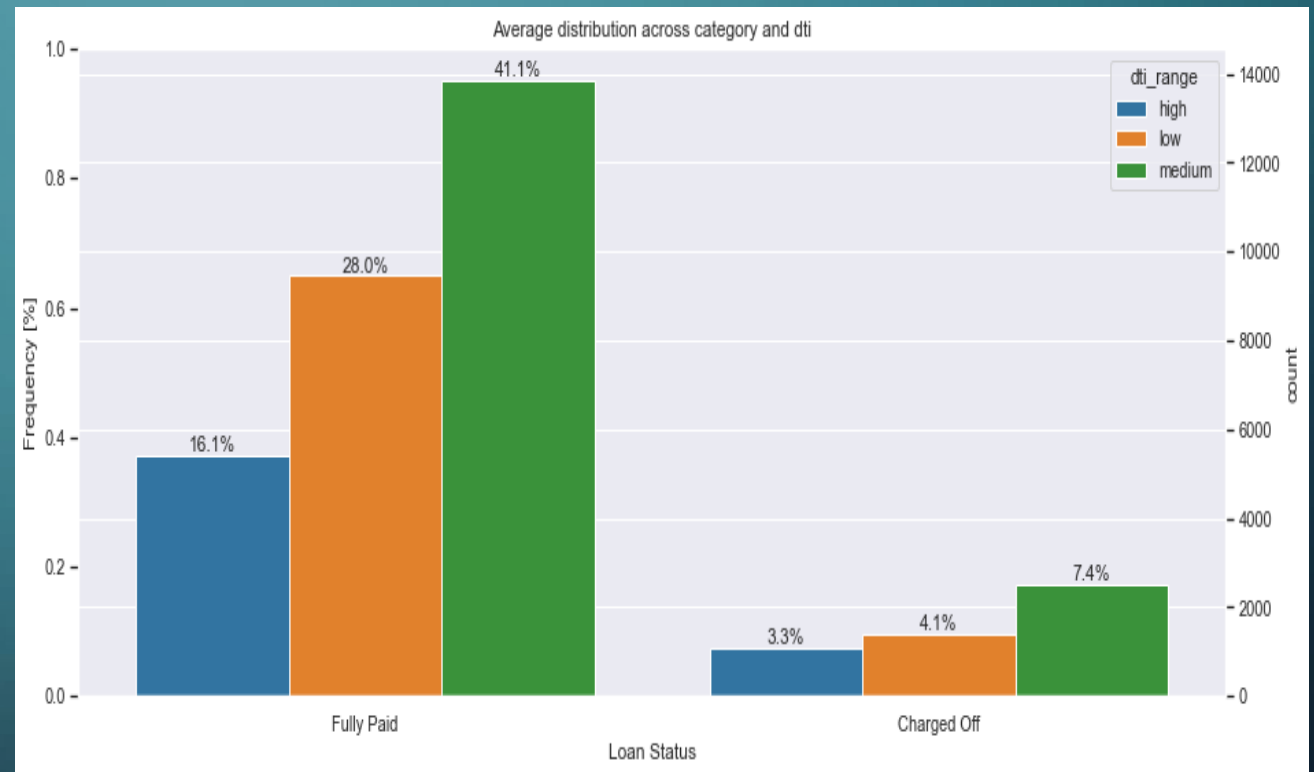
CATEGORICAL UNIVARIATE ANALYSIS

AVERAGE DISTRIBUTION ACROSS CATEGORY AND EXPERIENCE LEVEL

People with less years of experience have high chance of default in each category.
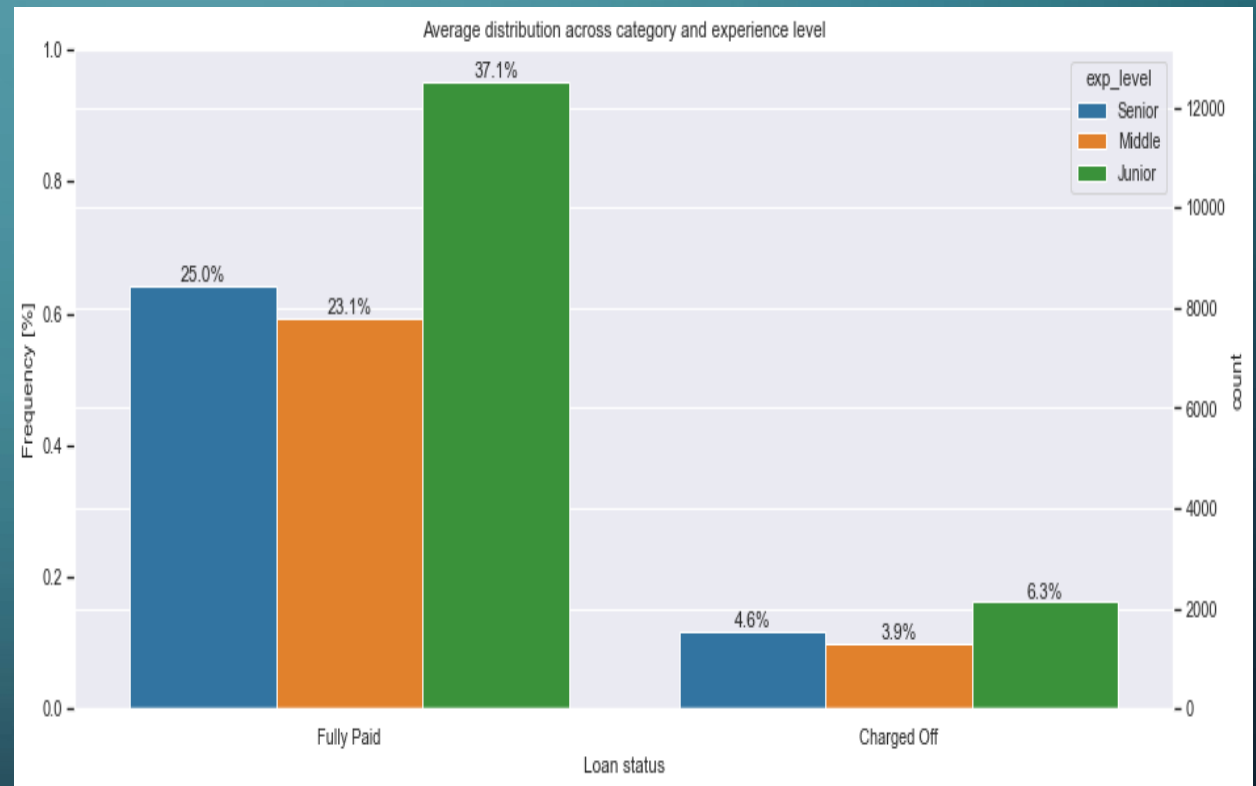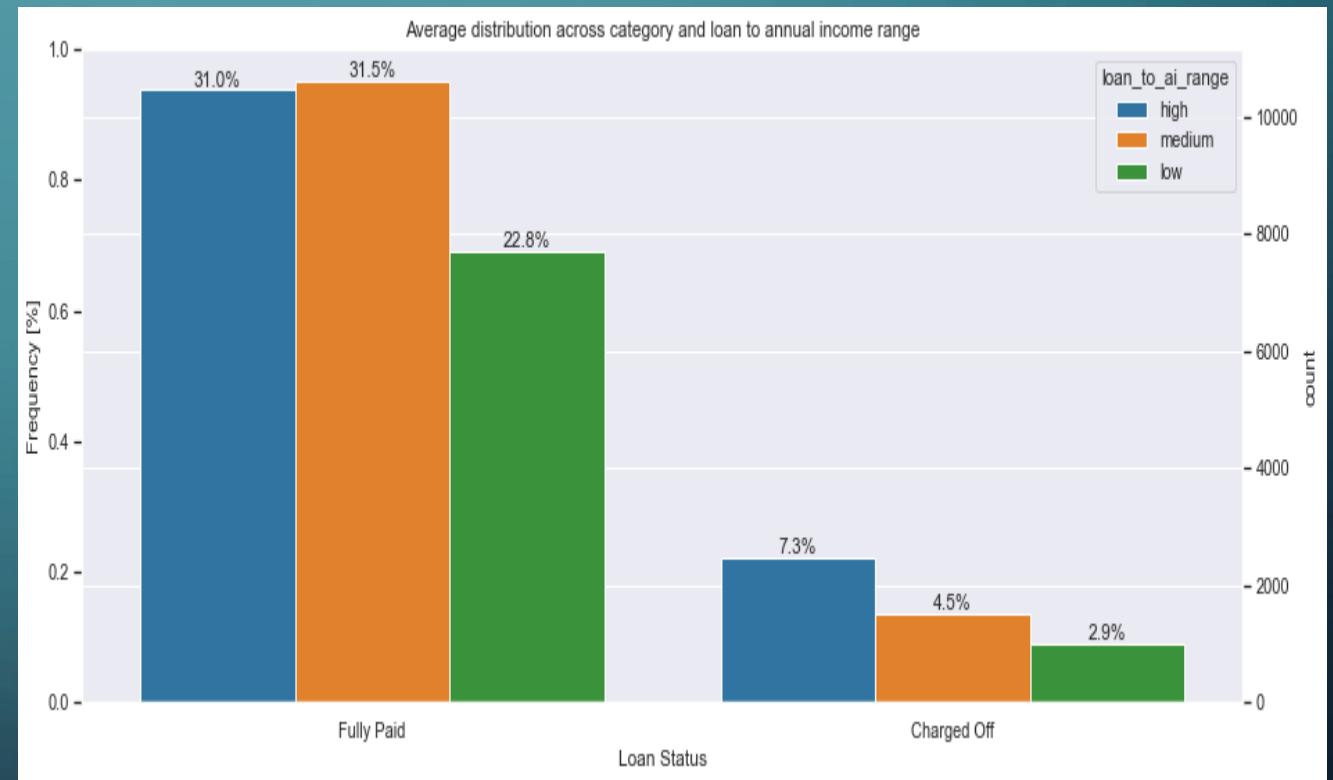
AVERAGE DISTRIBUTION ACROSS CATEGORY AND DTI (DEBT-TO-INCOME)

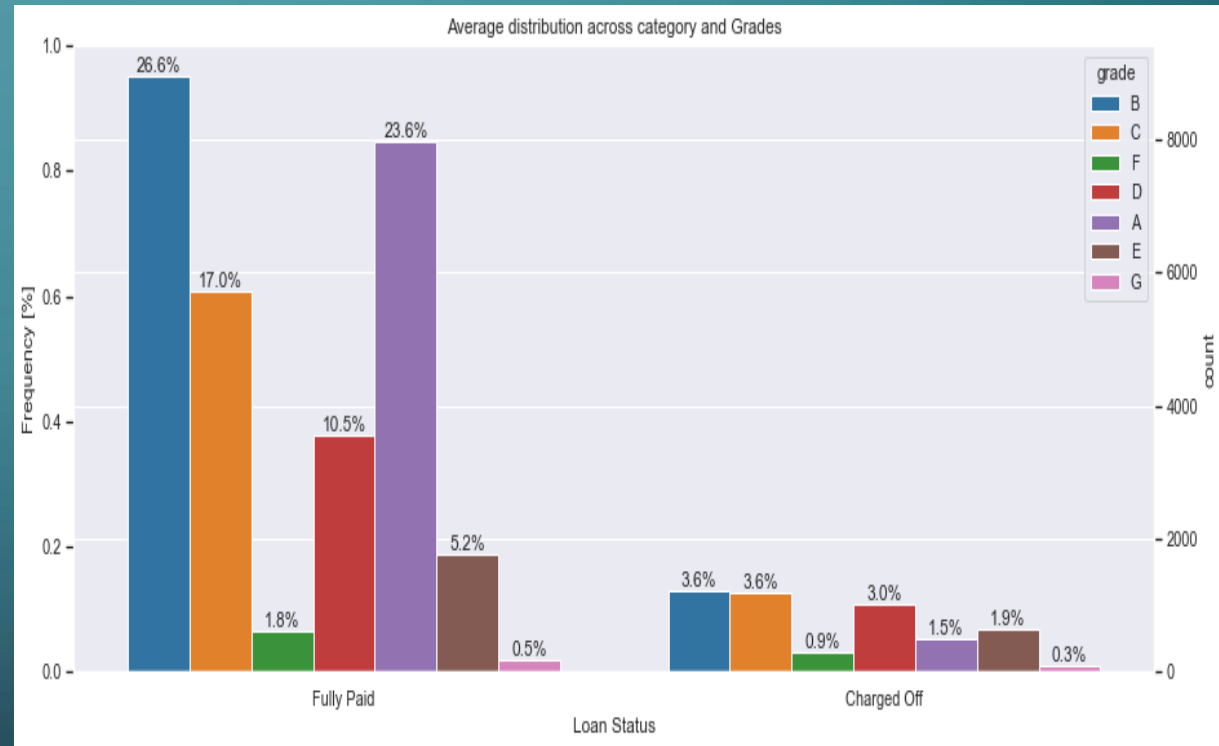People lying in medium dti range have high chances of default as well.

AVERAGE DISTRIBUTION ACROSS CATEGORY AND LOAN TO ANNUAL INCOME RANGE

People who have high loan to annual income ratio are at high risk of default

Average distribution across category and loan to annual income range

AVERAGE DISTRIBUTION ACROSS CATEGORY AND GRADES

Grades B, C and D are at high probability of defaulting

CATEGORICAL BIVARIATE ANALYSIS
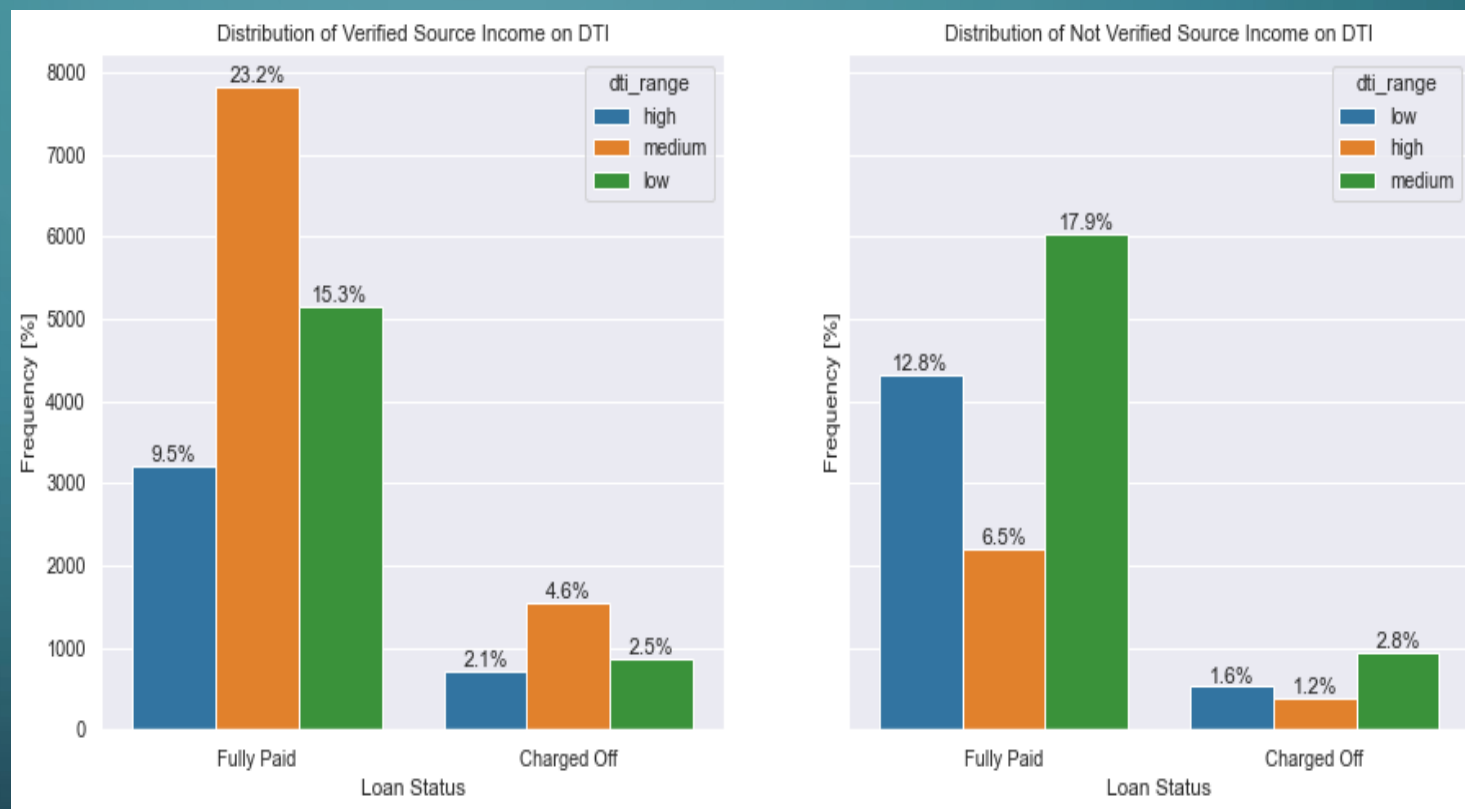
DISTRIBUTION OF LOAN STATUS ACROSS INCOME ON DTI
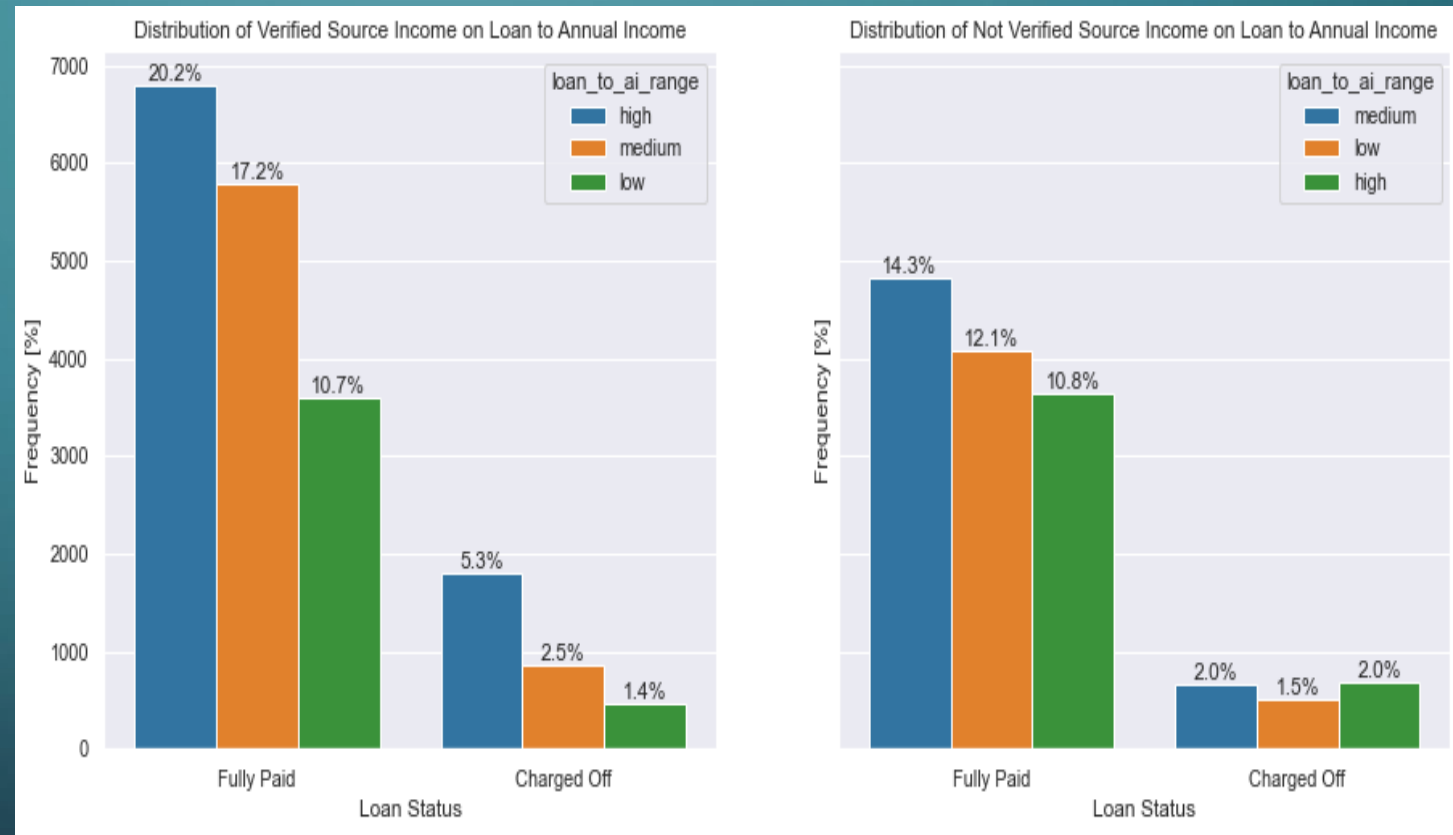
Verified vs Not Verified

- People lying in high DTI range For Verified Source Income have high chances to default.

- People lying in Medium DTI range for Not Verified Source Income have high chances to default.

DISTRIBUTION OF LOAN STATUS ACROSS INCOME ON LOAN TO ANNUAL INCOME
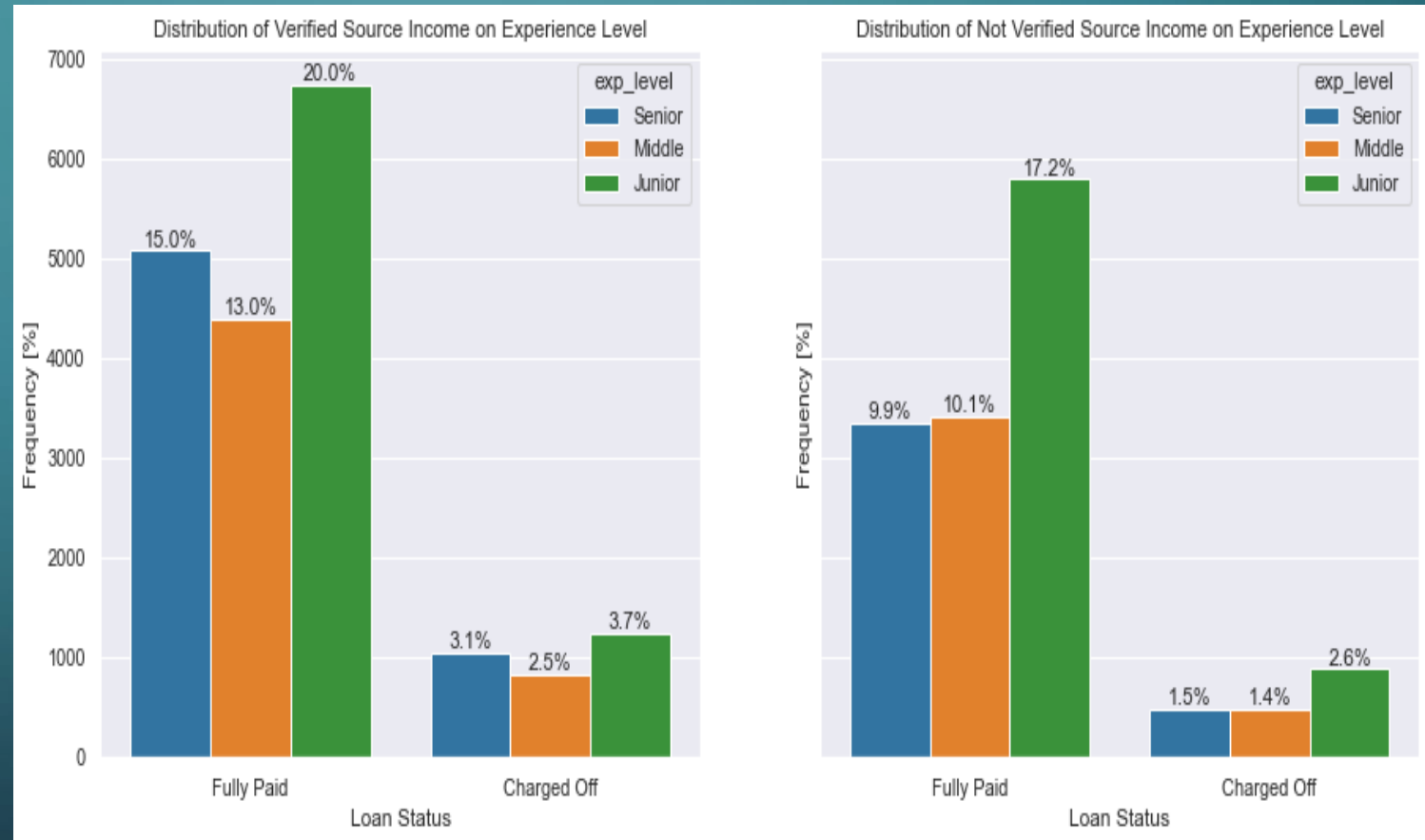
Verified vs Not Verified

- People who have high loan to annual income ratio are at high risk of default for Verified Source Income.

- People who have medium loan to annual income ratio are at high risk of default for Not Verified Source Income.
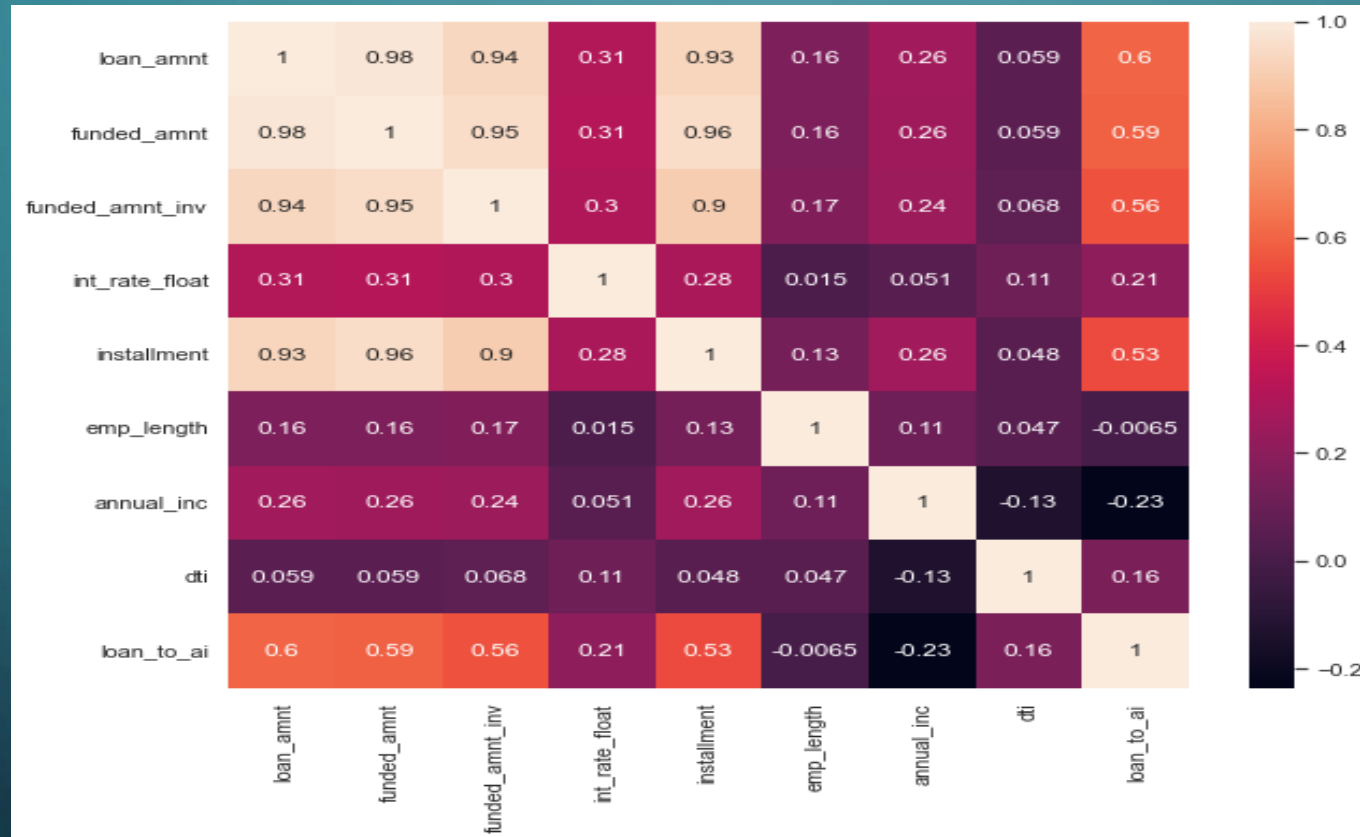

Distribution of Verified Source Income on Loan to Annual Income

Distribution of Not Verified Source Income on Loan to Annual Income

DISTRIBUTION OF LOAN STATUS ACROSS INCOME ON EXPERIENCE LEVEL

Verified vs Not Verified

- People with less experience have high chance of default for both Verified and Non verified Source Income.

CORERELATION OF VARIABLES

CONCLUSION

- Banks should focus more on Senior levels or high experienced person for Rented ownership and middle level for owned assets.

- People who have high loan to annual income ratio are at high risk of default for Verified Source Income.

- Maximum loan applications have been received and hence high is the defaulting probability in these categories:

  1) Debt consolidation 2) Credit card 3) Home improvement 4) Others 5) Major purchase

- Get as much as clients from housing type with owned asset.