LENDING CLUB CASE STUDY

AKASH VASANI
(EPGP ML C54)

LENDING CLUB

- Largest online loan marketplace, provides loan for various purposes.
- Data set consists of slew of parameters pertaining to past loan applicants including whether they had paid the loan or were charged off.
- Company's aim it to maximise its profit by not providing the loan to the people who are likely to default.
- However, company should proactively provide loan to the applicant who are likely to repay the loan on time for increasing the company's profit.

PROBLEM STATEMENT

- Assisting the company for maximising the profit based on data analysis.
- Based on the data set, finding out various drivers or variables which indicates whether applicant is likely to default or not.
- Utilising the univariate, bivariate and multi-variate analysis for finding out the patterns between different variables and deriving insights from it.
- Reporting various driving factors which can be effectively utilised for risk mitigation.

METHODOLOGY

Following approach is adopted for solving the case study.

- Understanding and Extracting the Data
- Data Cleaning
- Univariate Analysis
- Bivariate Analysis
- Multi-variate Analysis

UNDERSTANDING THE DATA

- Data set consists of loan data pertaining to the past loan issued.
- Data set has 111 columns and around 39 thousand rows.
- Each column represents one variable.
- Loan Status columns contains whether the loan is Current, Fully Paid or Charged Off and this is our target variable.
- Data dictionary is provided for understanding technical meaning of various variables.
- Types of variables:
  - Loan Characteristics: Loan amount, purpose, annual income, etc
  - Demographic Variables: Age, employment status, relationship status, state, etc
  - Behavioural Variables: EMI, delinquency, next payment date, delinquency in past 2 years, etc.

DATA CLEANING

- Removing all the columns which contains null values for all rows. (Columns removed: 54 numbers)
- Removing the columns mentioned in the cols_to_drop, since they are irrelevant for the study. (Columns removed: 28 numbers)
- We are left with 29 columns and 39717 rows.
- Removing the null values in emp_length and revol_util by dropping those rows.
- Filling the null values in pub_rec_bankruptcies with “Not Available”

- Converting the interest rate column to float value.
- Converting the revol_util column to float value.
- Converting the term column to int value.
- Converting the pub_rec_bankruptcies to string value.
- Removing rows for which loan_status is “Current” since we can’t infer whether they would pay the loan or get charged off.

OUTLIERS

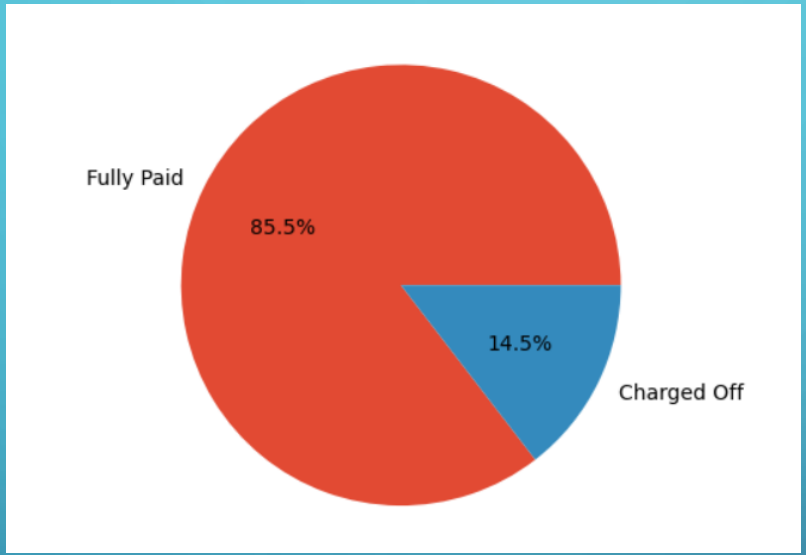
- Annual Income has many outliers as visible in the box plot.
- Removing the rows which has income value greater than 95 percentile value.
- DTI, loan amount, open account does not have significant outliers.

BINNING

- It is the process of converting the continuous variables into categorical for analysis
- Binning has been done for the following variables:
  - Loan Amount
  - Annual Income
  - Instalment
  - DTI
  - Revol Until
  - Interest Rate
  - Total Account

UNIVARIATE ANALYSIS

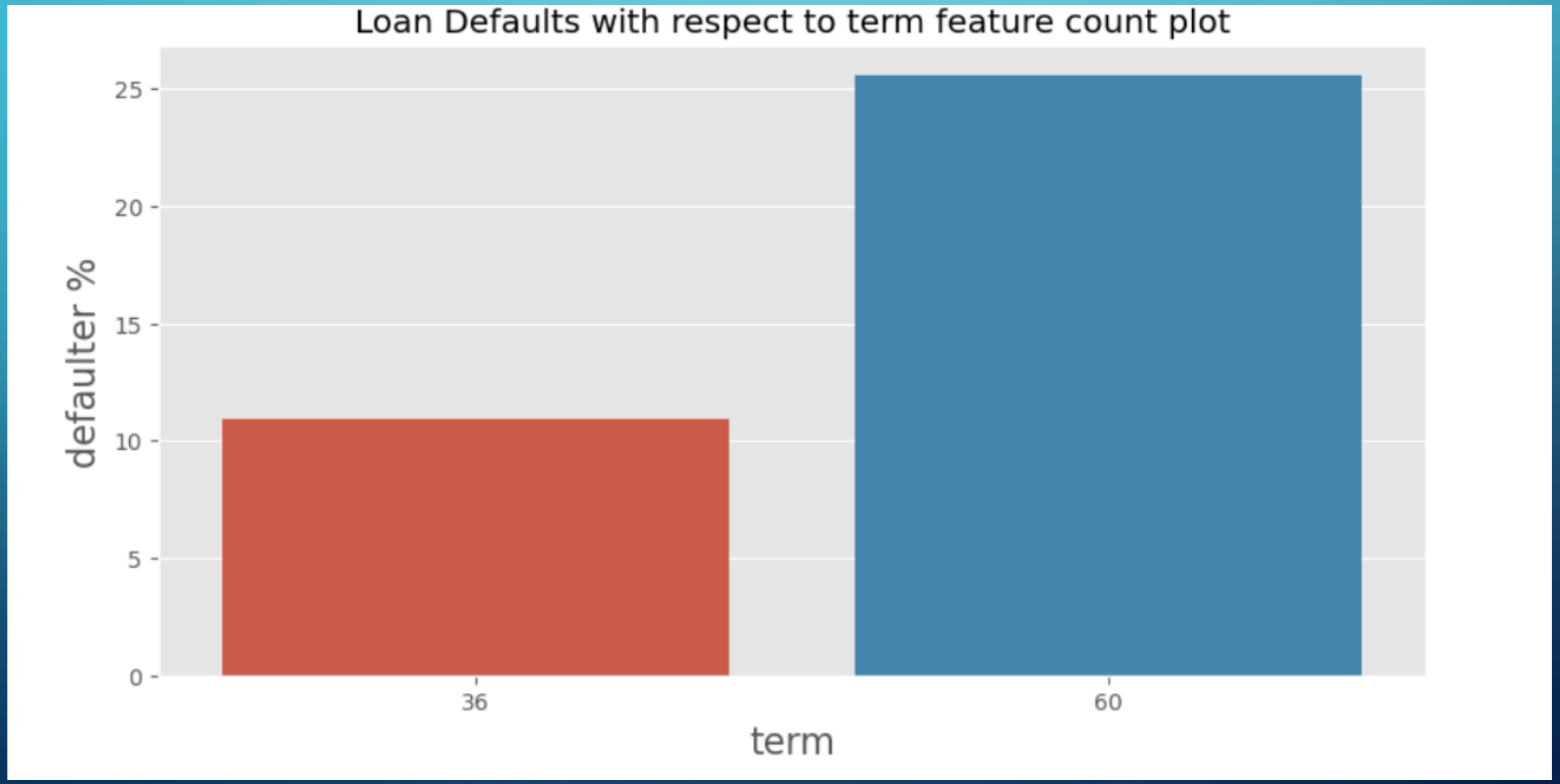
Analysing the influence of single variable on the defaulters percentage.



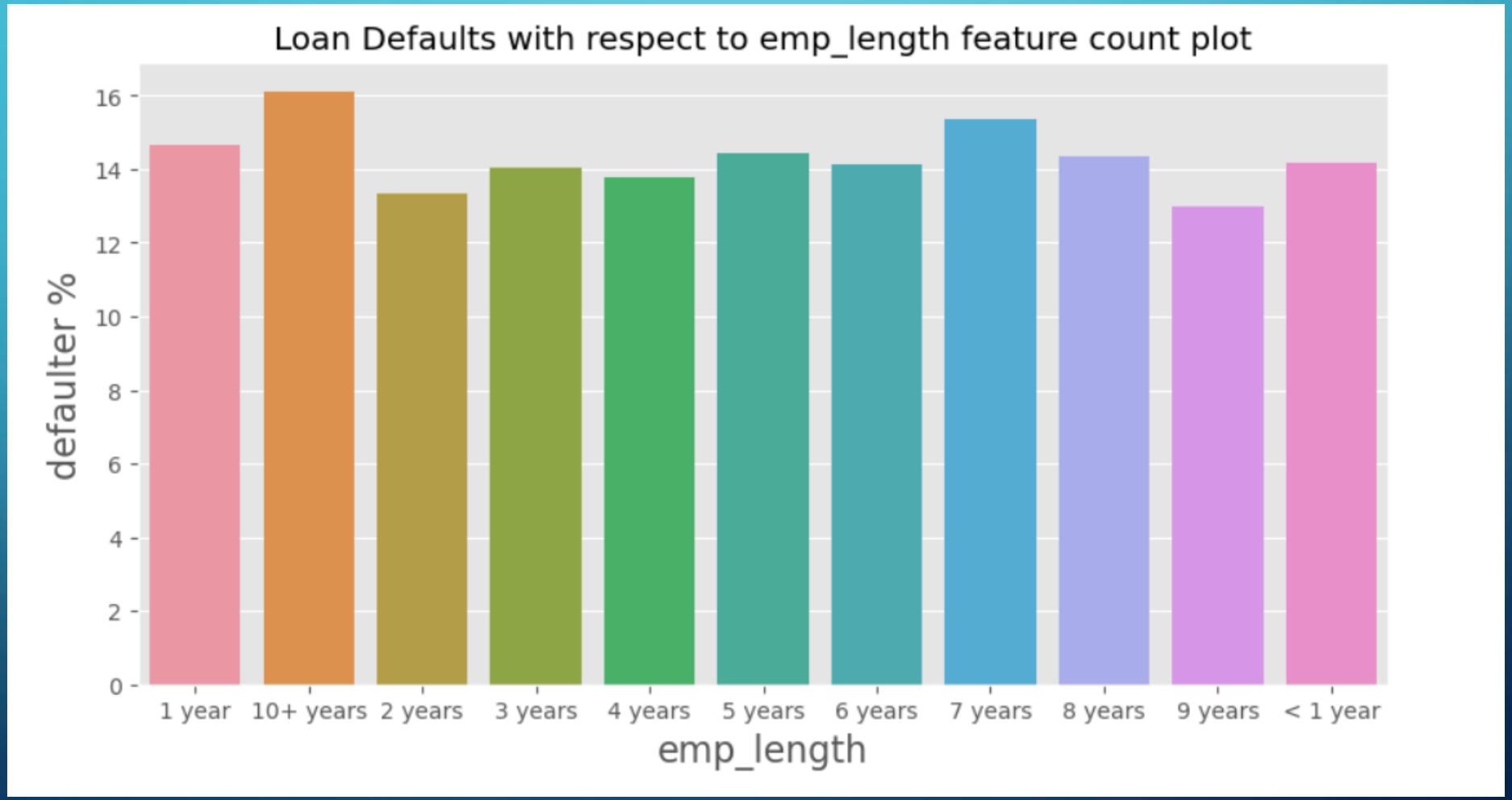
Pie Chart depicting the distribution of the loan status.

14.5% of the applicant are defaulters

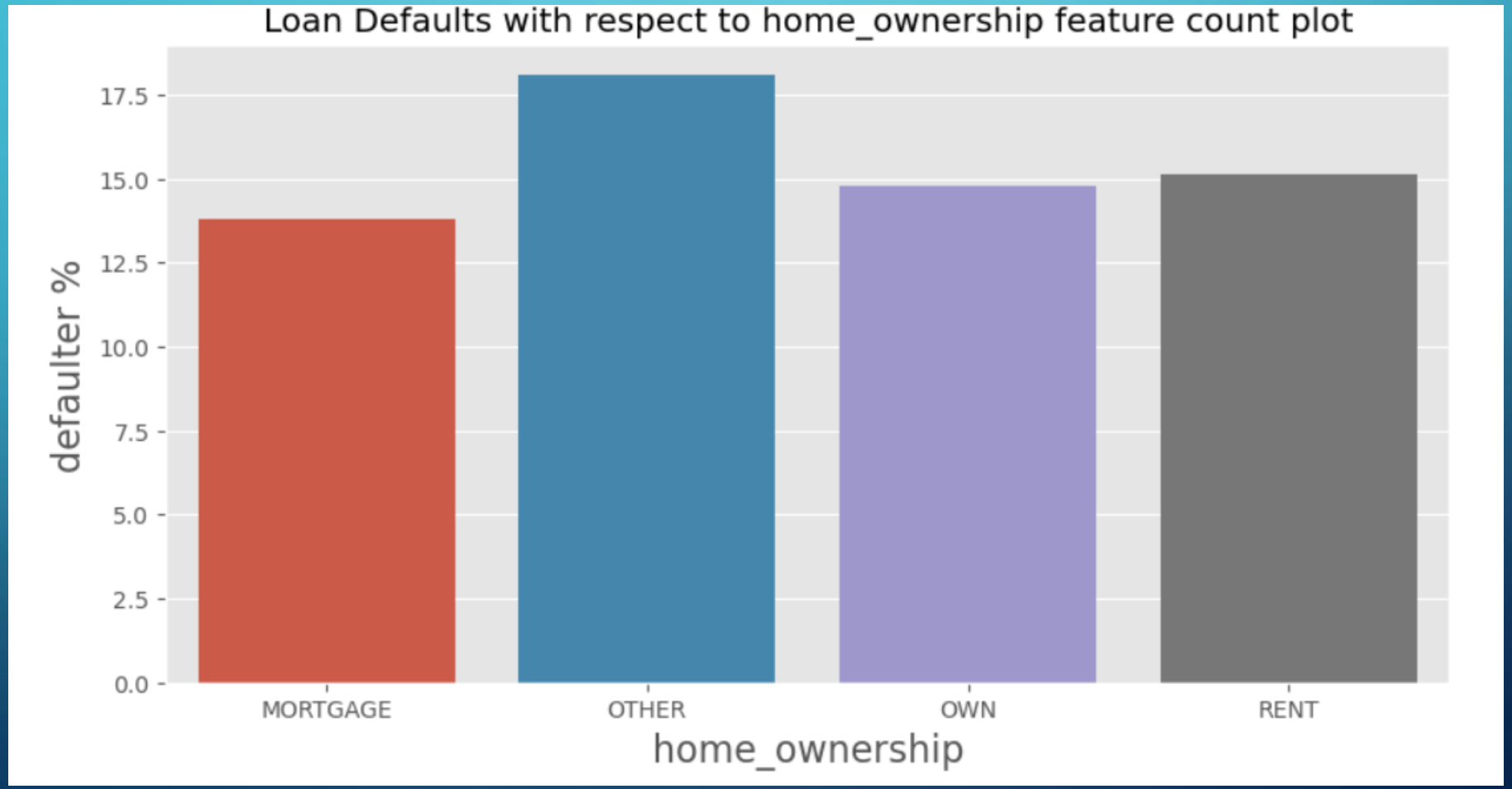
DEFAULTERS PERCENTAGE WITH RESPECT TO THE LOAN TERMS.



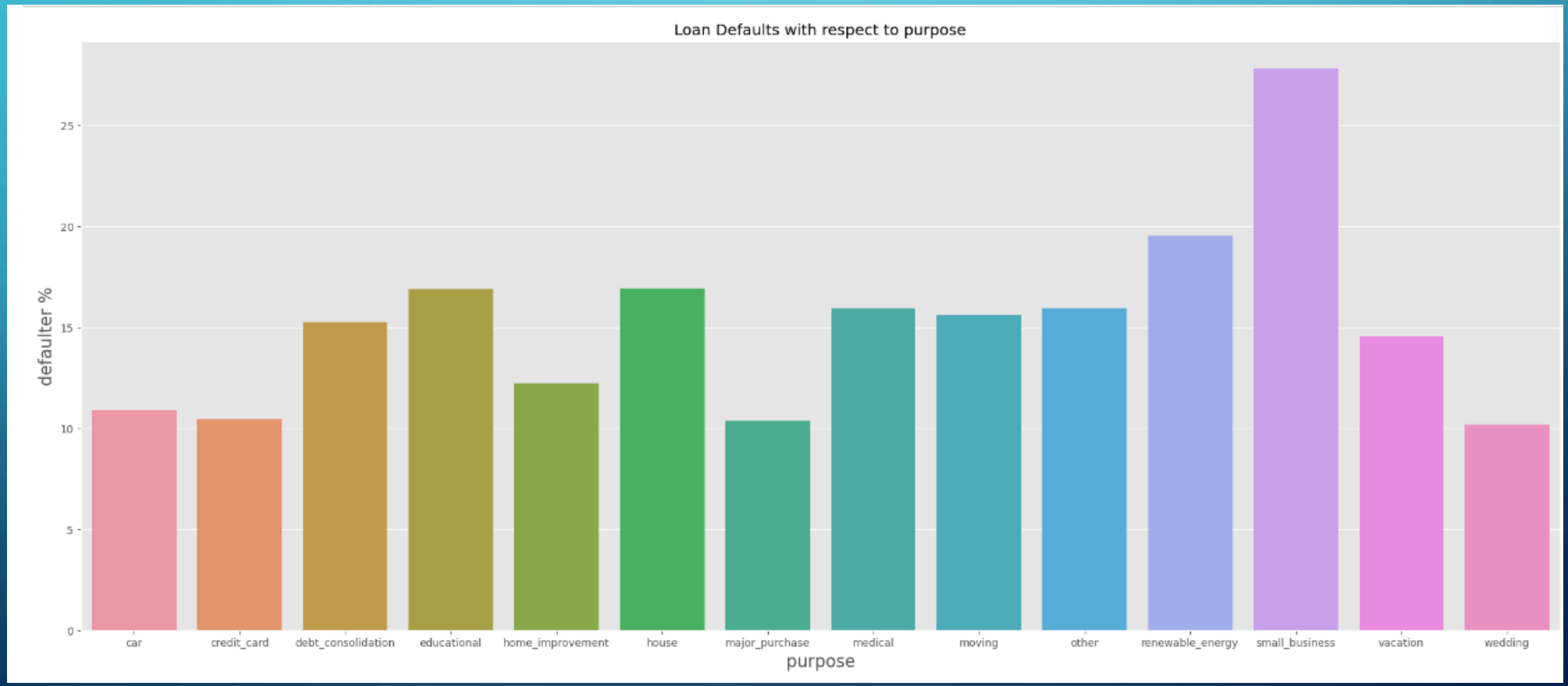
DEFAULTERS PERCENTAGE WITH RESPECT TO THE NUMBER OF YEARS A PERSON HAS WORKED IN ORGANIZATION.



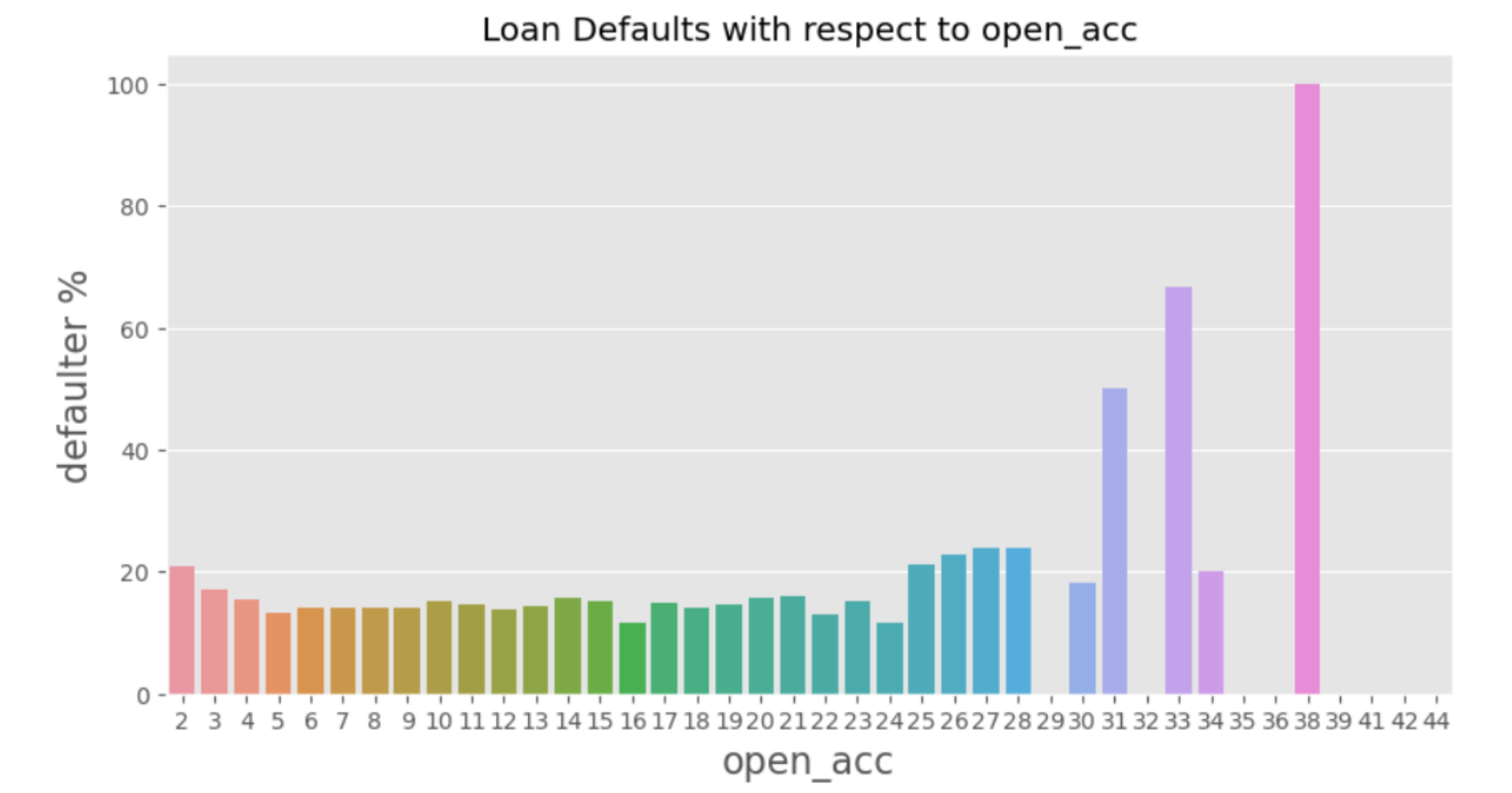
DEFAULTERS PERCENTAGE WITH RESPECT TO THEIR HOME OWNERSHIP.



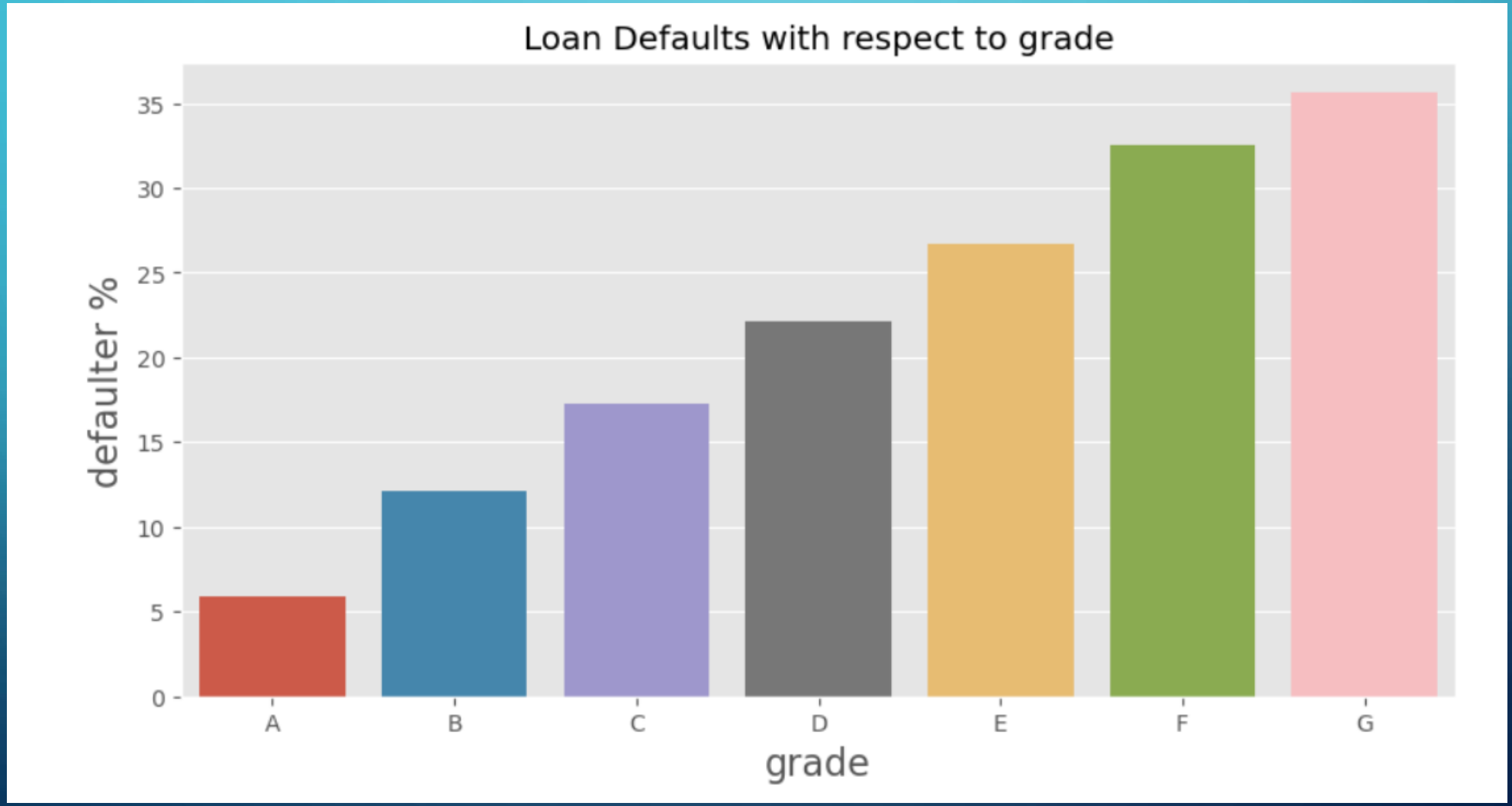
DEFAULTERS PERCENTAGE WITH RESPECT TO FOR WHAT PURPOSE THEY HAD TAKEN LOAN



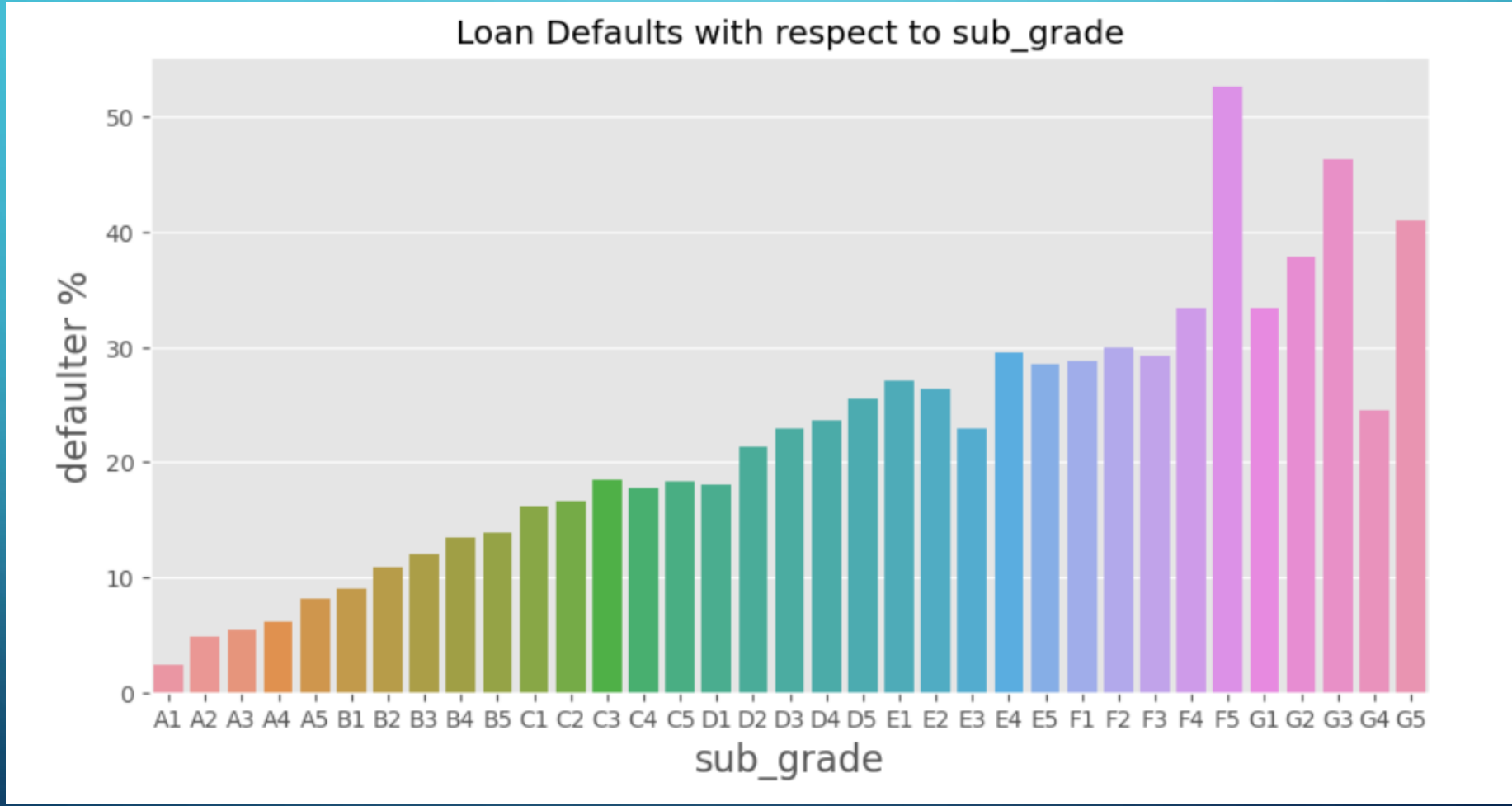
DEFAULTERS PERCENTAGE WITH RESPECT TO OPEN ACCOUNT



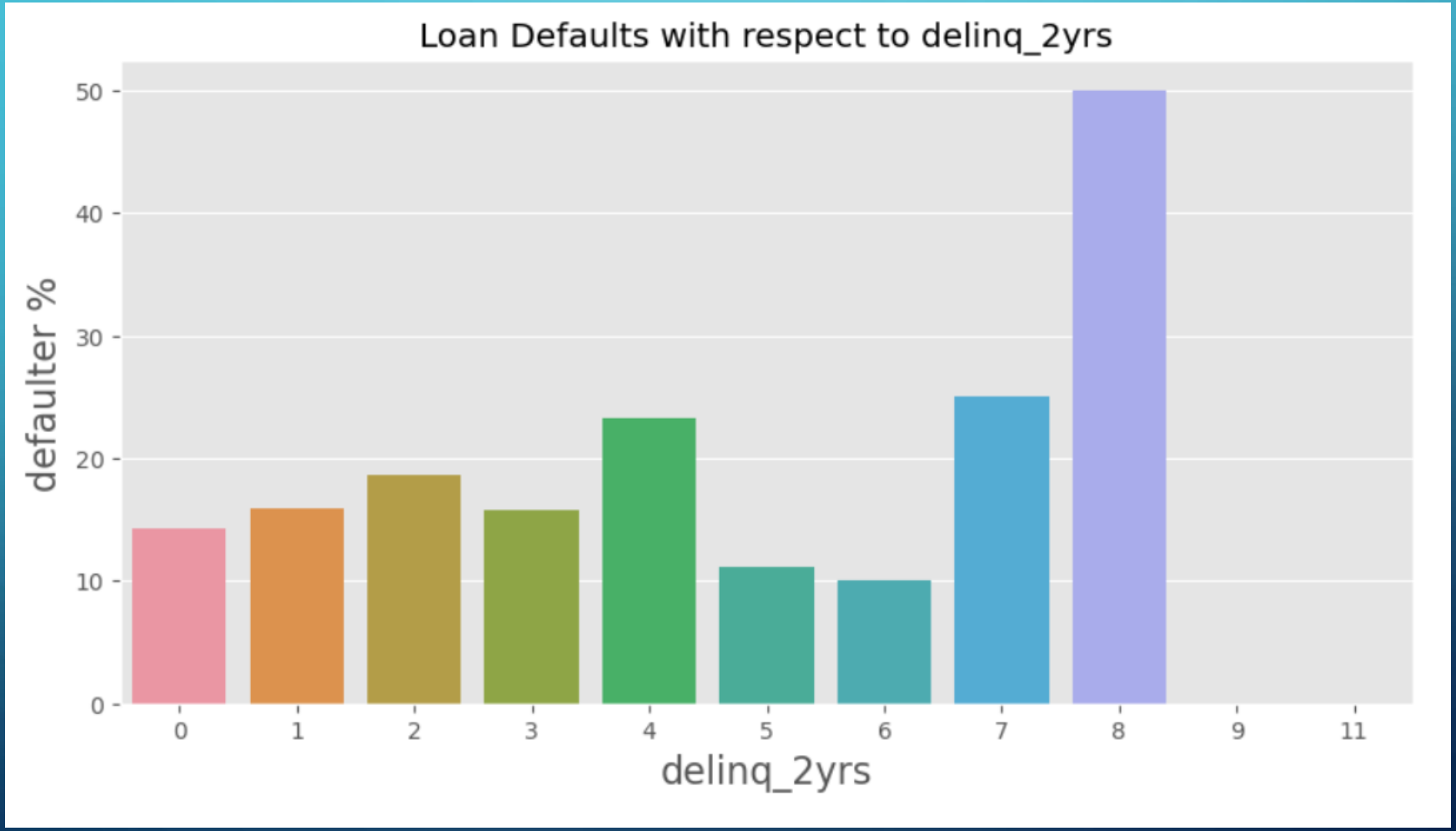
DEFAULTERS PERCENTAGE WITH RESPECT GRADE



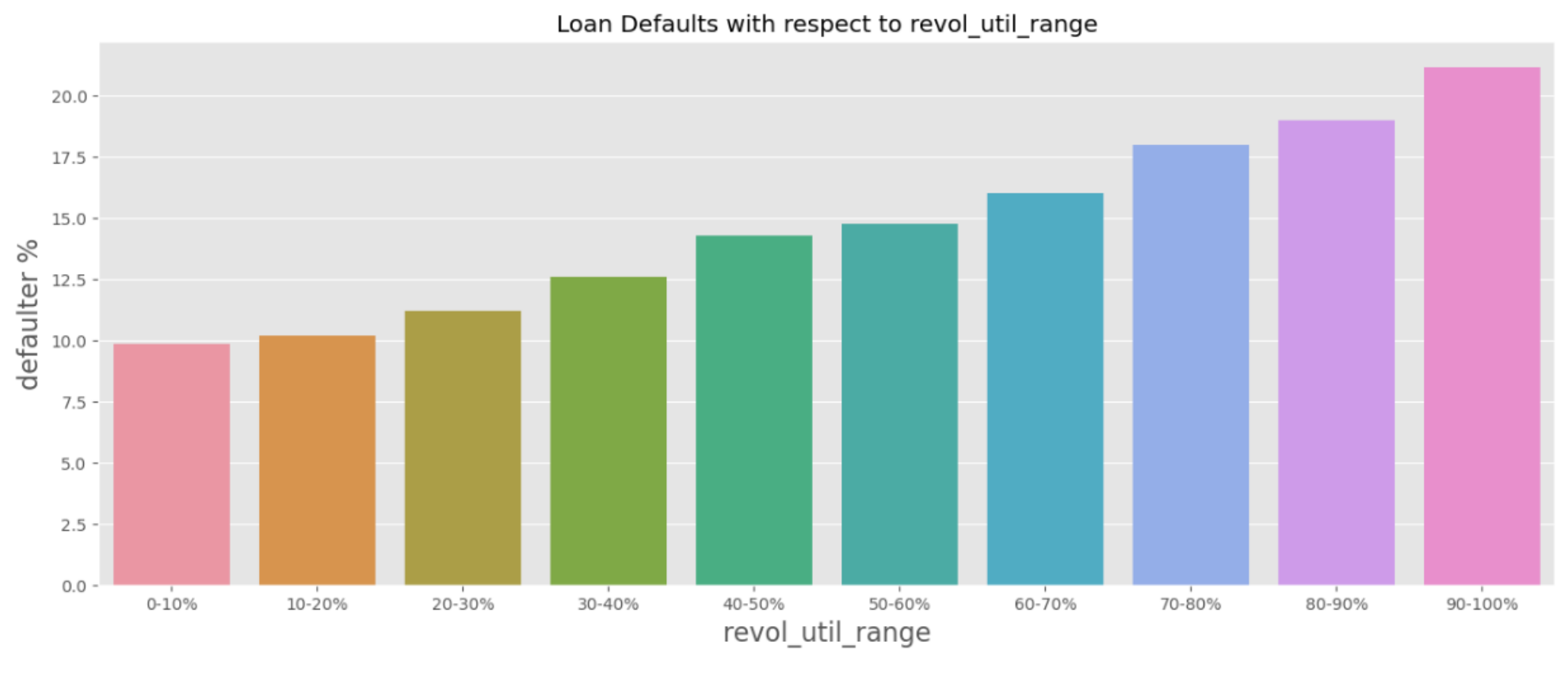
DEFAULTERS PERCENTAGE WITH RESPECT SUB GRADE



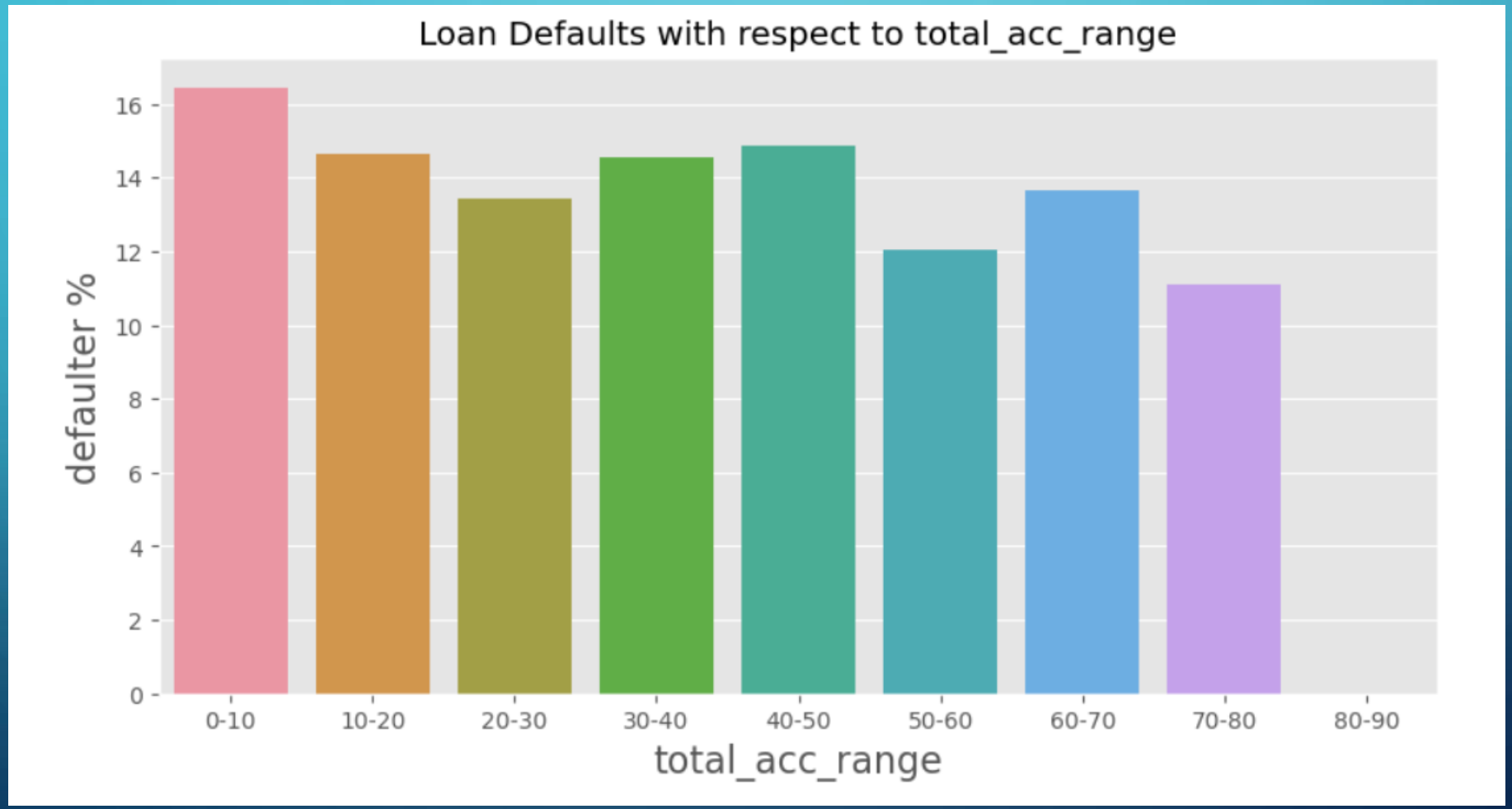
DEFAULTERS PERCENTAGE WITH RESPECT DELINQUENCY



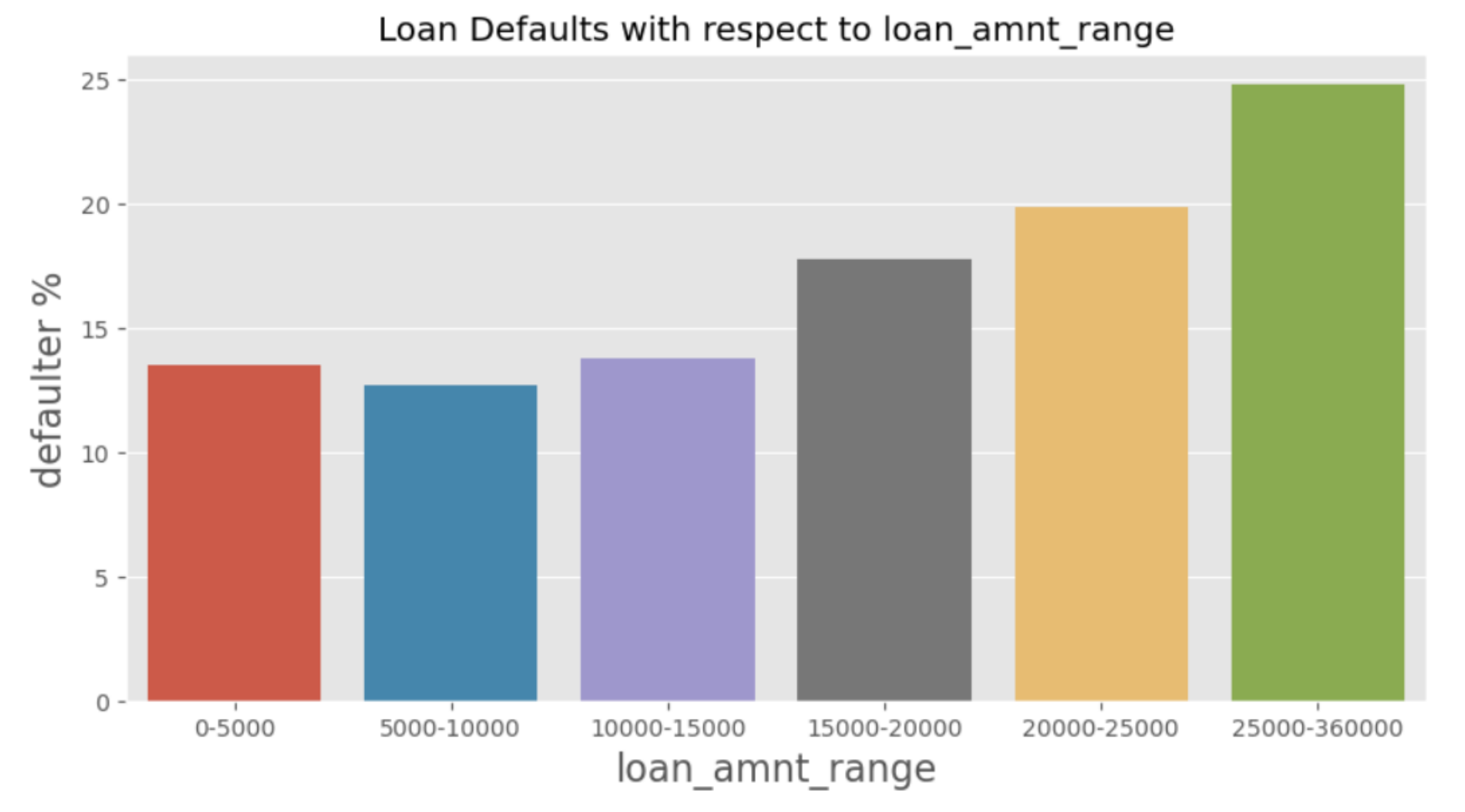
DEFAULTERS PERCENTAGE WITH RESPECT REVOL UNTIL



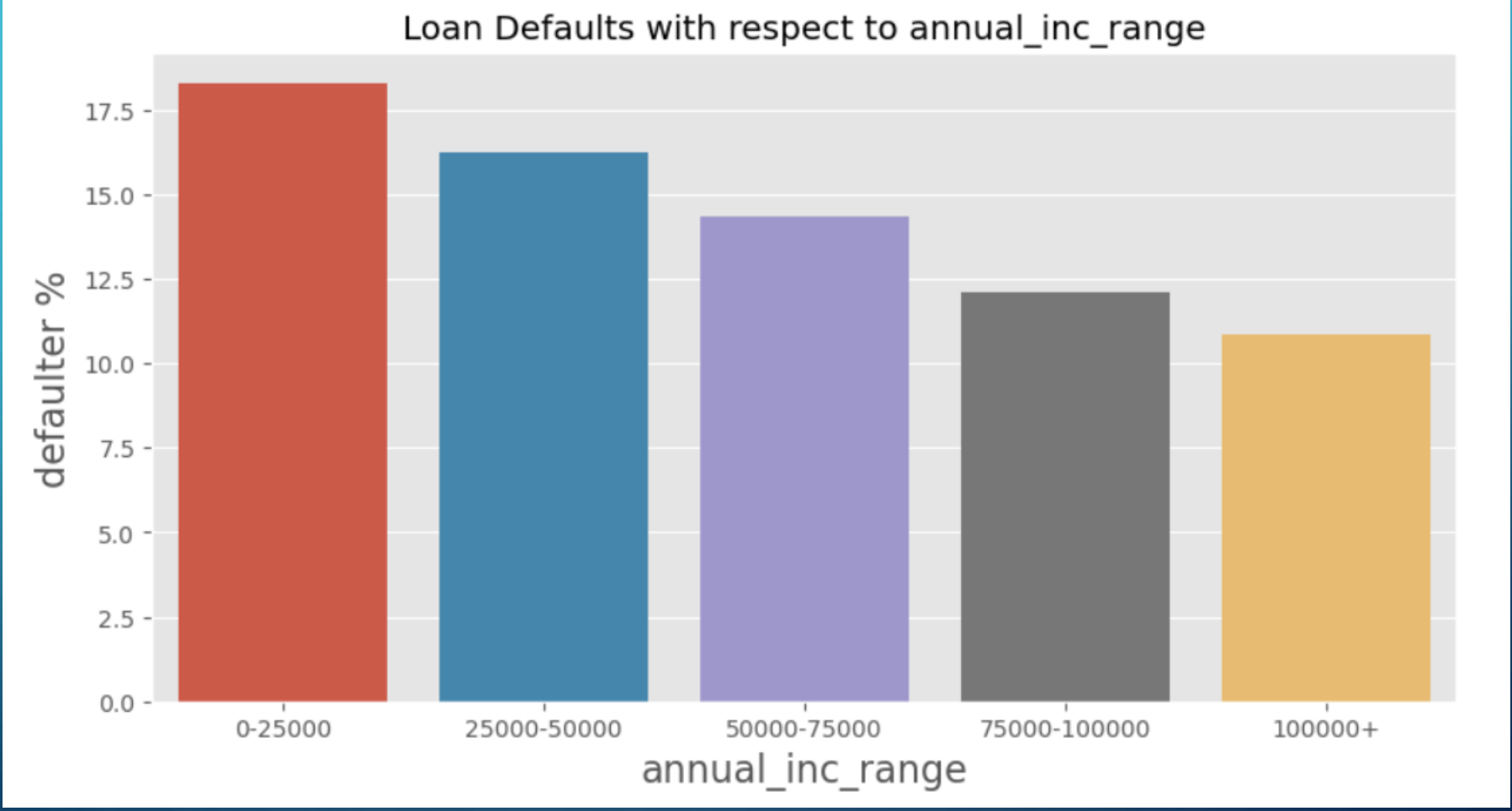
DEFAULTERS PERCENTAGE WITH RESPECT TOTAL ACCOUNT



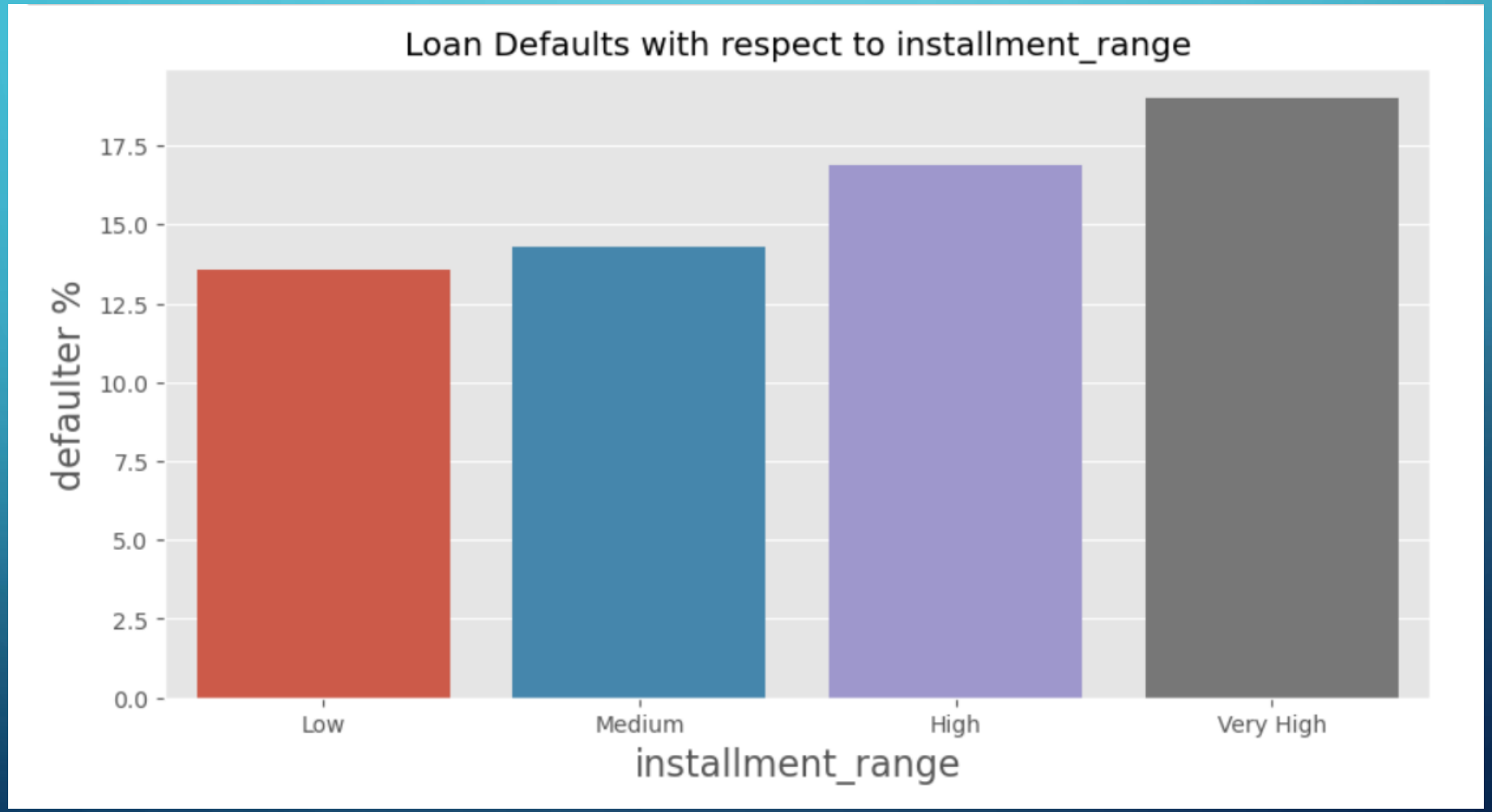
DEFAULTERS PERCENTAGE WITH RESPECT TO LOAN AMOUNT



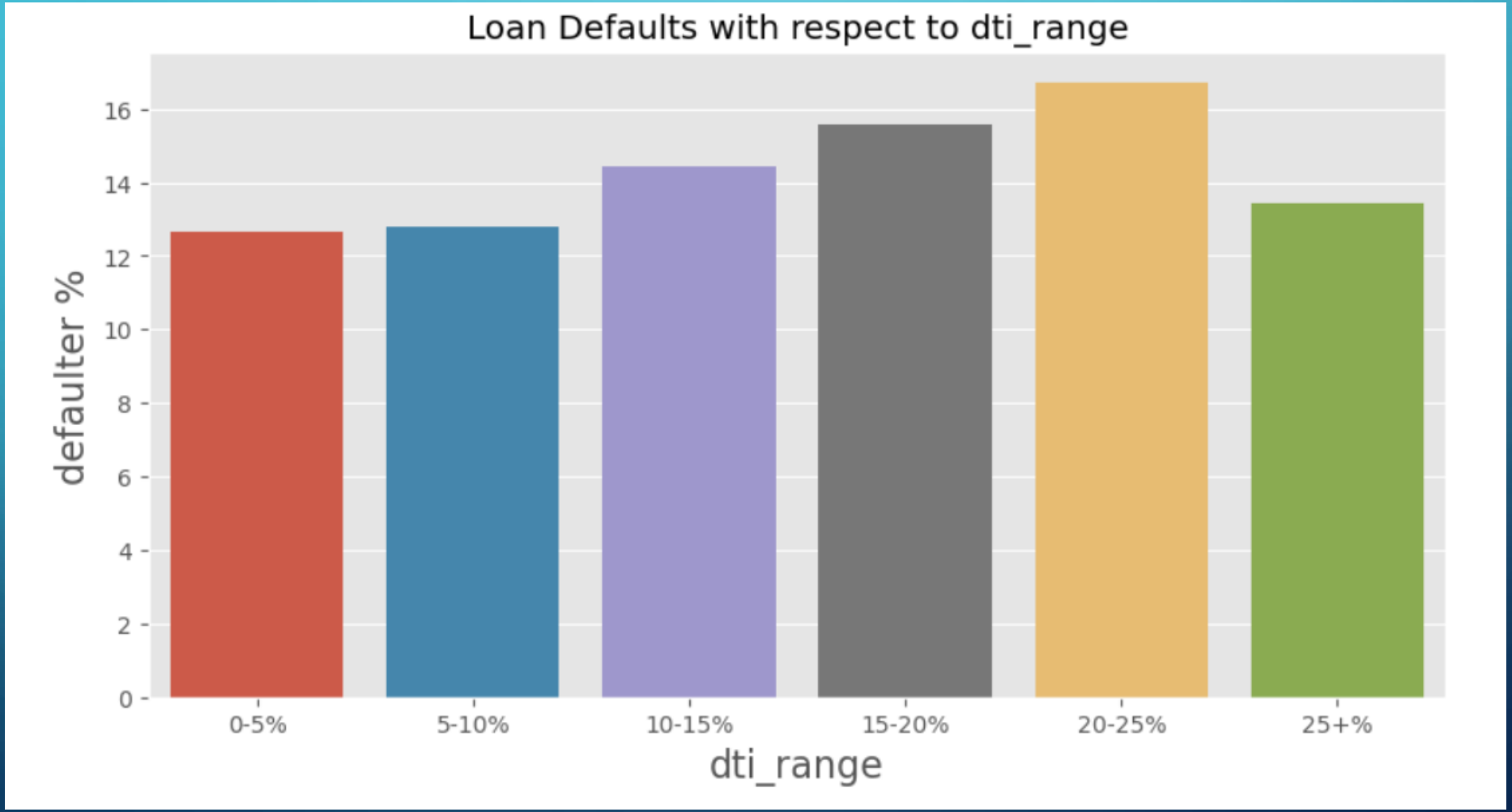
DEFAULTERS PERCENTAGE WITH RESPECT TO ANNUAL INCOME



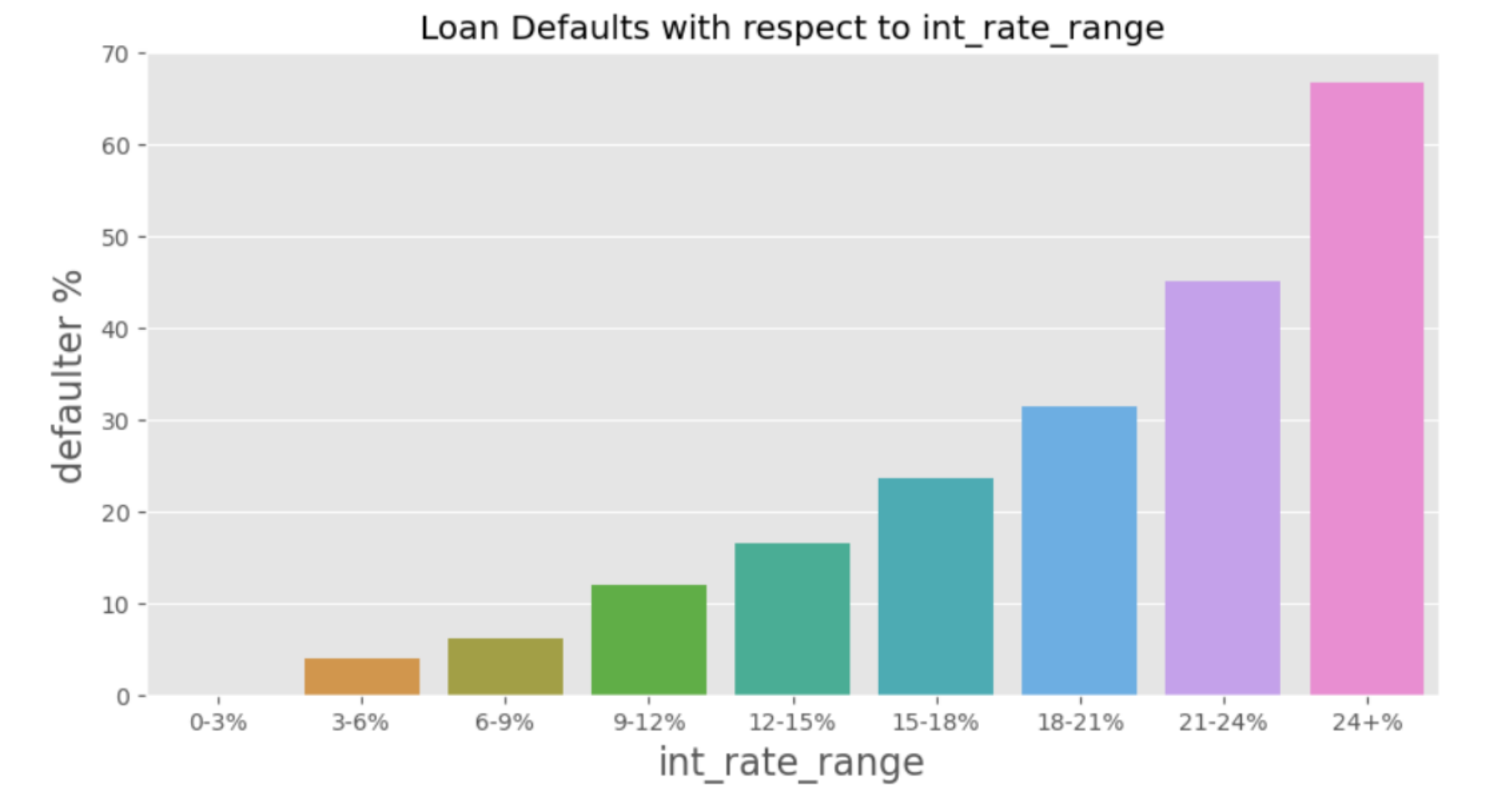
DEFAULTERS PERCENTAGE WITH RESPECT TO INSTALLMENT



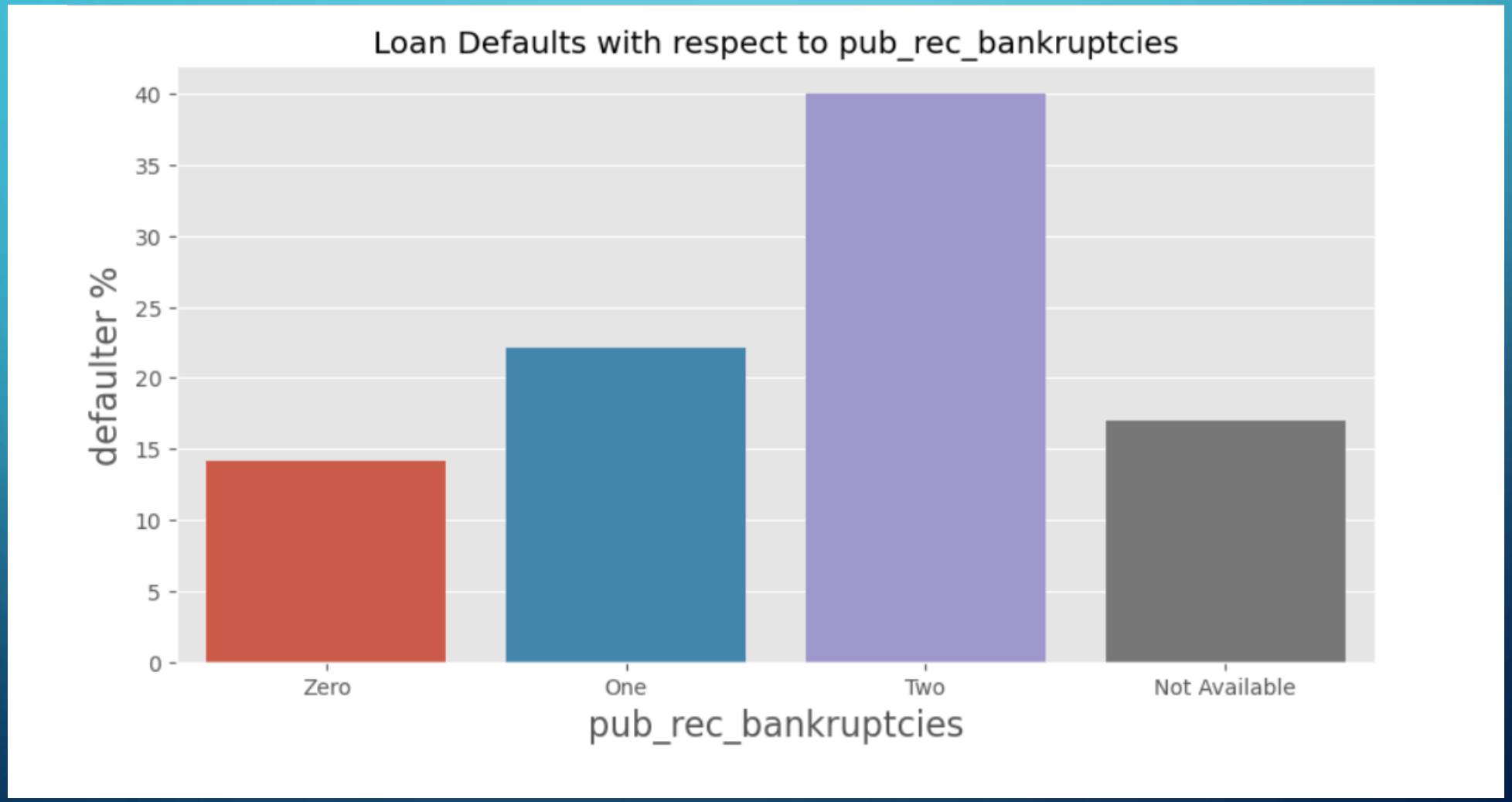
DEFAULTERS PERCENTAGE WITH RESPECT TO DTI



DEFAULTERS PERCENTAGE WITH RESPECT TO INTEREST RATE



DEFAULTERS PERCENTAGE WITH RESPECT TO PUBLIC RECORD OF BANKRUPTCIES



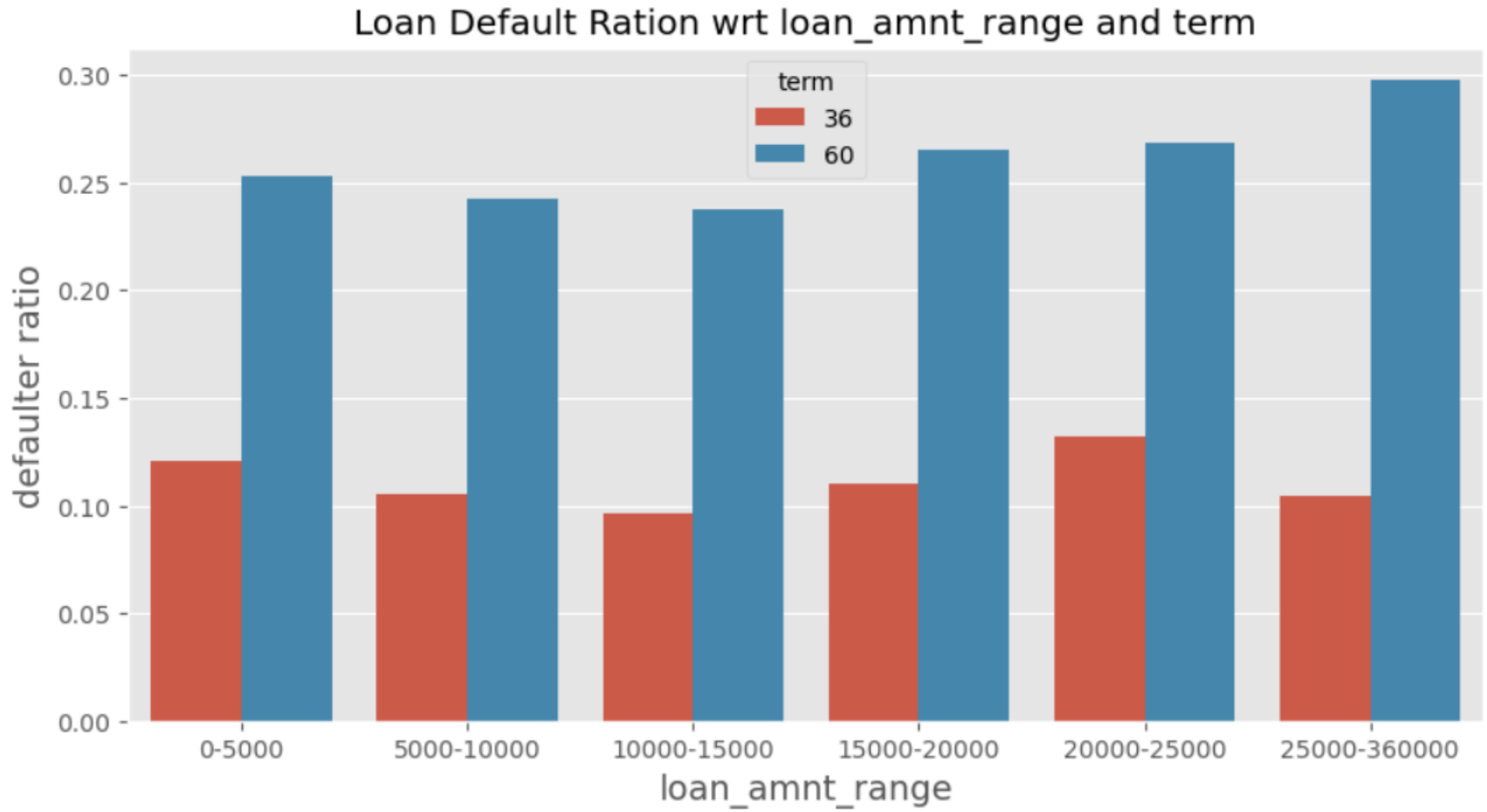
UNIVARIATE ANALYSIS CONCLUSION

1. Loan Terms- As loan term increase chances of defaul increase.
2. Purpose- There is higher chance of default if loan is taken for small businesses.
3. State- There are higher chances of default if loan is taken in the NE state.
4. Grade and Sub Grade- There are higher chances of default as the grade and sub grade increases.
5. Open Account- As the open account increases the chances of default increases.
6. Revol Until- As the revol until percentge increases, the chances of default also increases.
7. Loan Amount- As the loan amount increases, the chances of default increases too.
8. Annual Income- As the annual income of person increases the chances of default decreases.
9. Installement- As the installement increases the chances of default also increases.
10. DTI- As the debt to income ratio increase the chances of default increases.
11. Interest Rate- As the interest rate of the loan increases the chances of default also increases.
12. Public Record of Bankruptcies- As the number of the bankruptcies increases the chances of default also increases.

BIVARIATE ANALYSIS

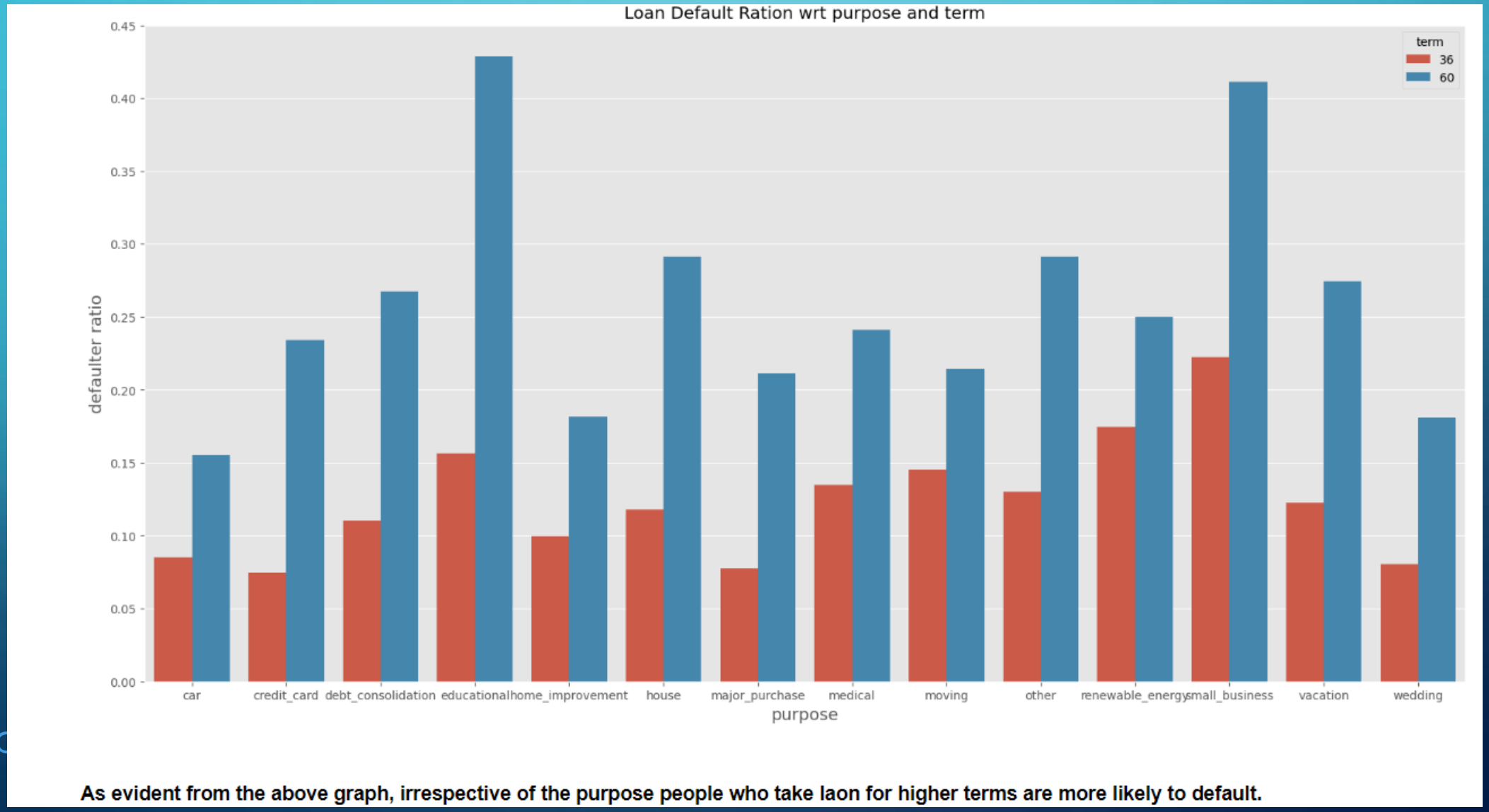
- Two variables are considered and their influence is checked against the defaulters percentage.

DEFAULTERS PERCENTAGE WITH RESPECT TO LOAN AMOUNT AND TERM



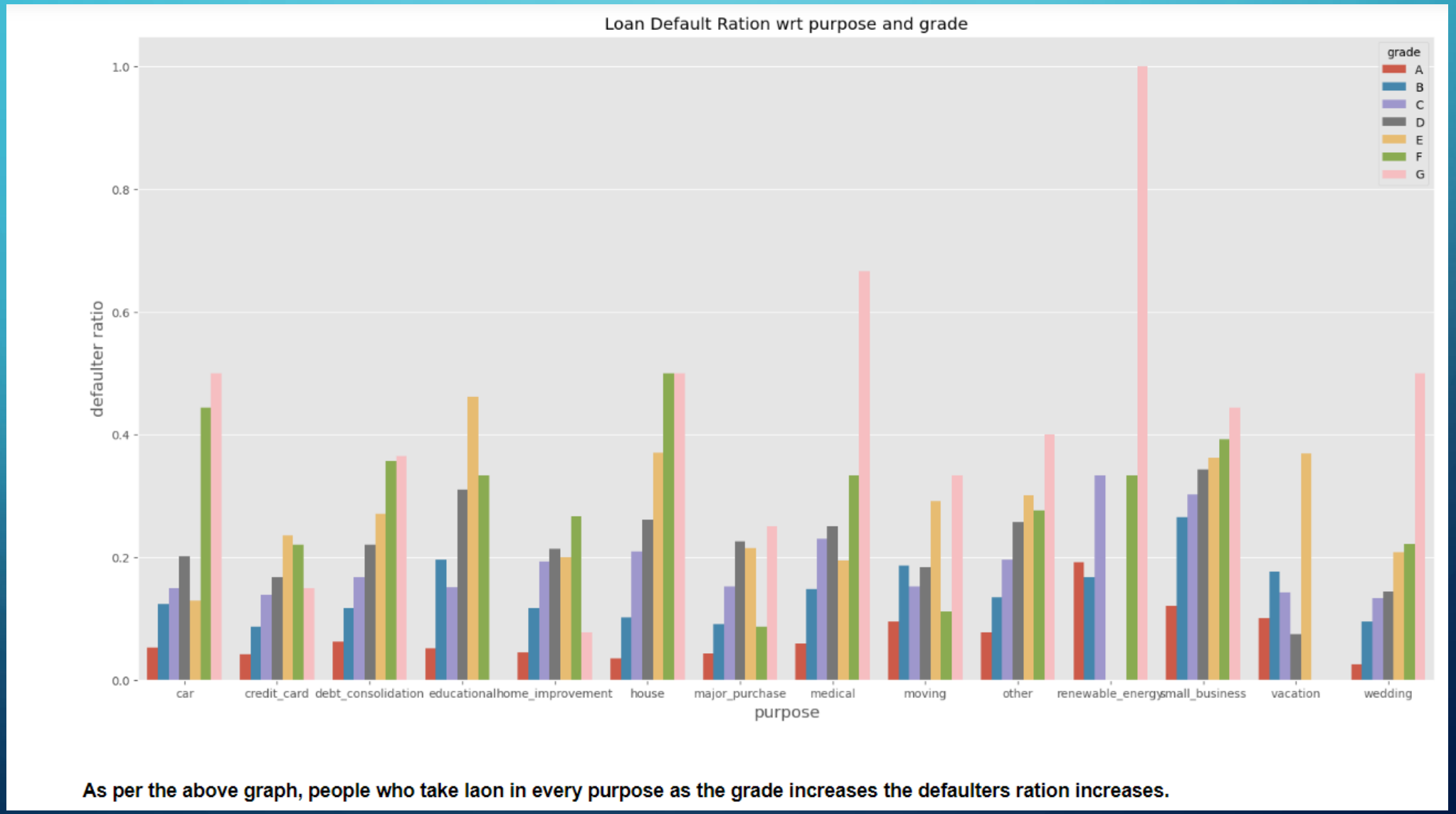
As evident from the above graph, irrespective of the loan amount people who take laon for higher terms are more likely to default.

DEFAULTERS PERCENTAGE WITH RESPECT TO PURPOSE AND TERM



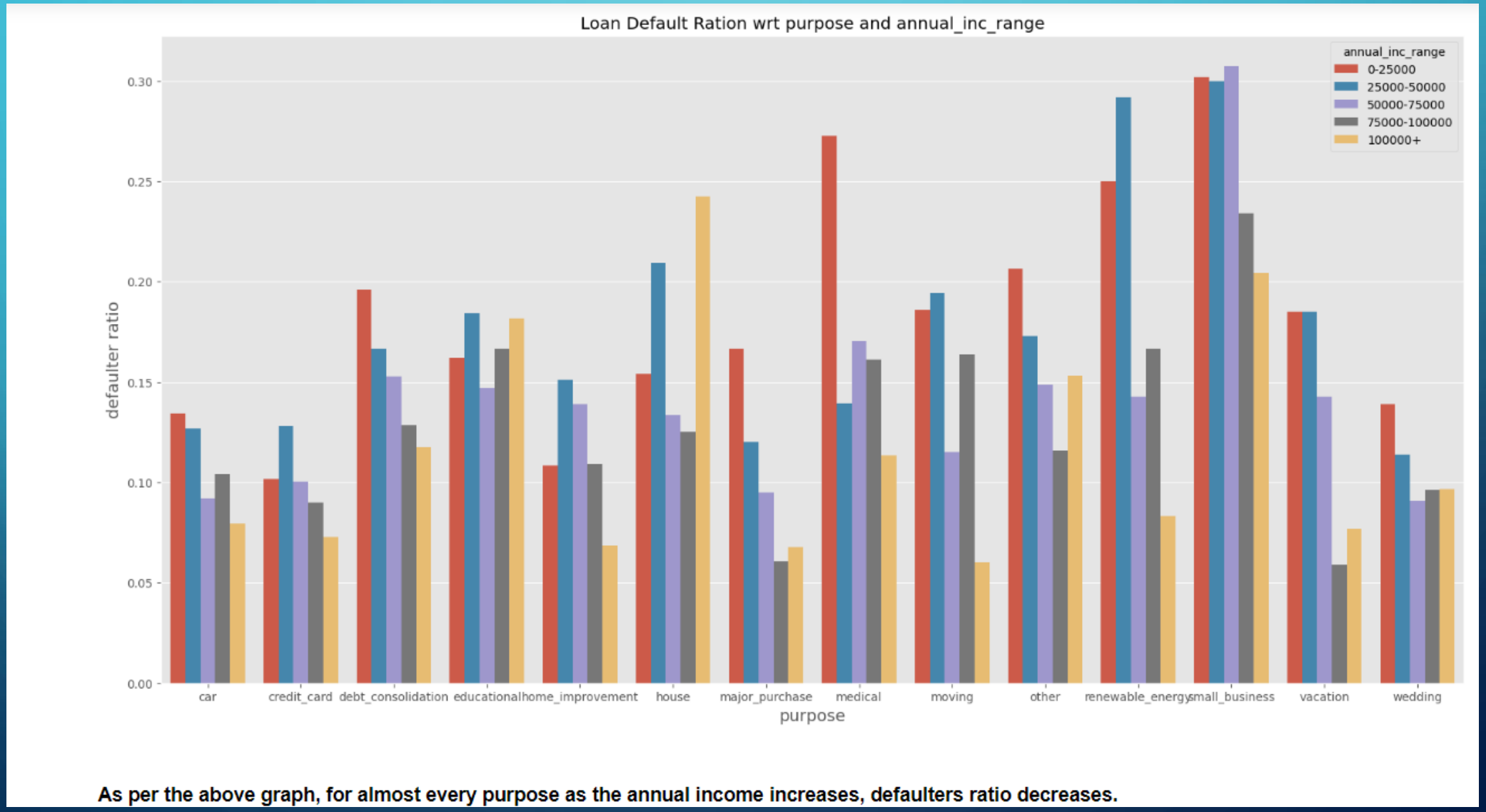
As evident from the above graph, irrespective of the purpose people who take laon for higher terms are more likely to default.

DEFAULTERS PERCENTAGE WITH RESPECT TO PURPOSE AND GRADE



As per the above graph, people who take laon in every purpose as the grade increases the defaulters ration increases.

DEFAULTERS PERCENTAGE WITH RESPECT TO PURPOSE AND ANNUAL INCOME

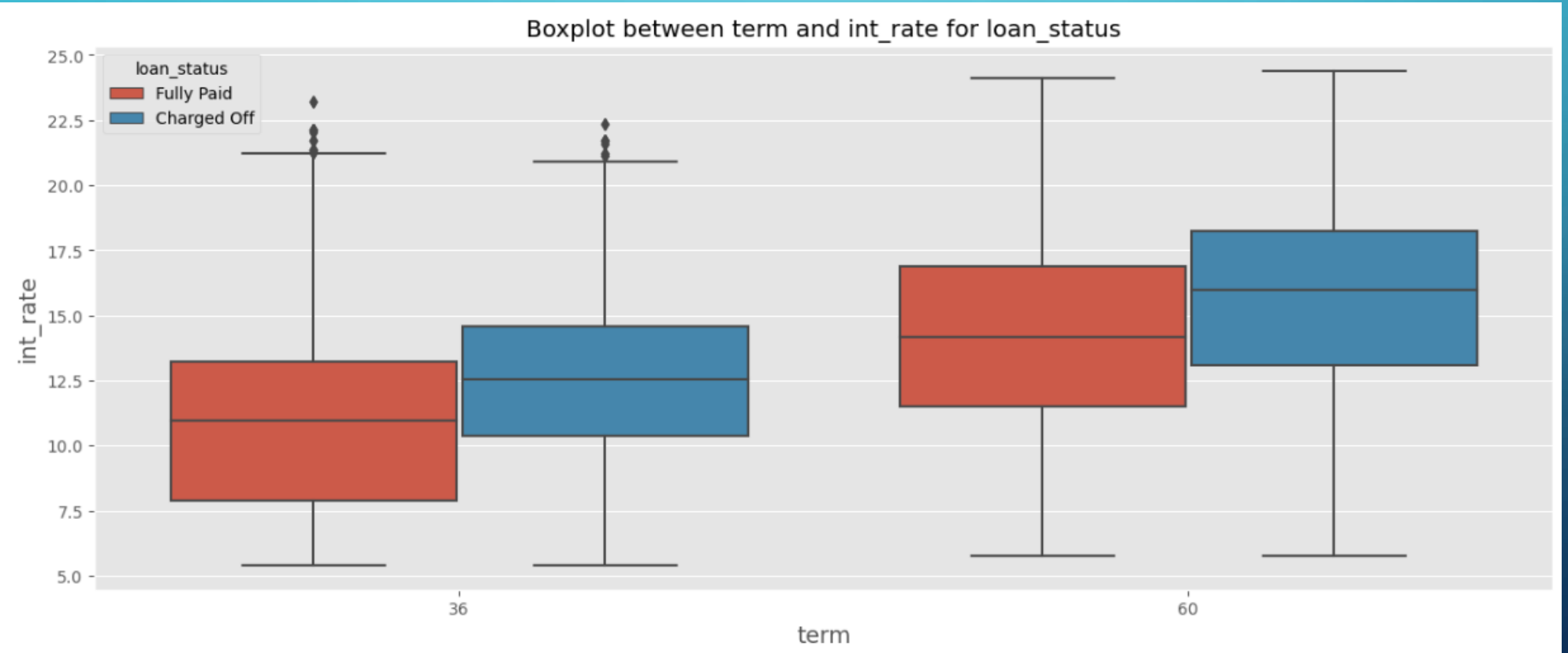


As per the above graph, for almost every purpose as the annual income increases, defaulters ratio decreases.

ANALYSING THE SPREAD OF VARIABLE BOXPLOT

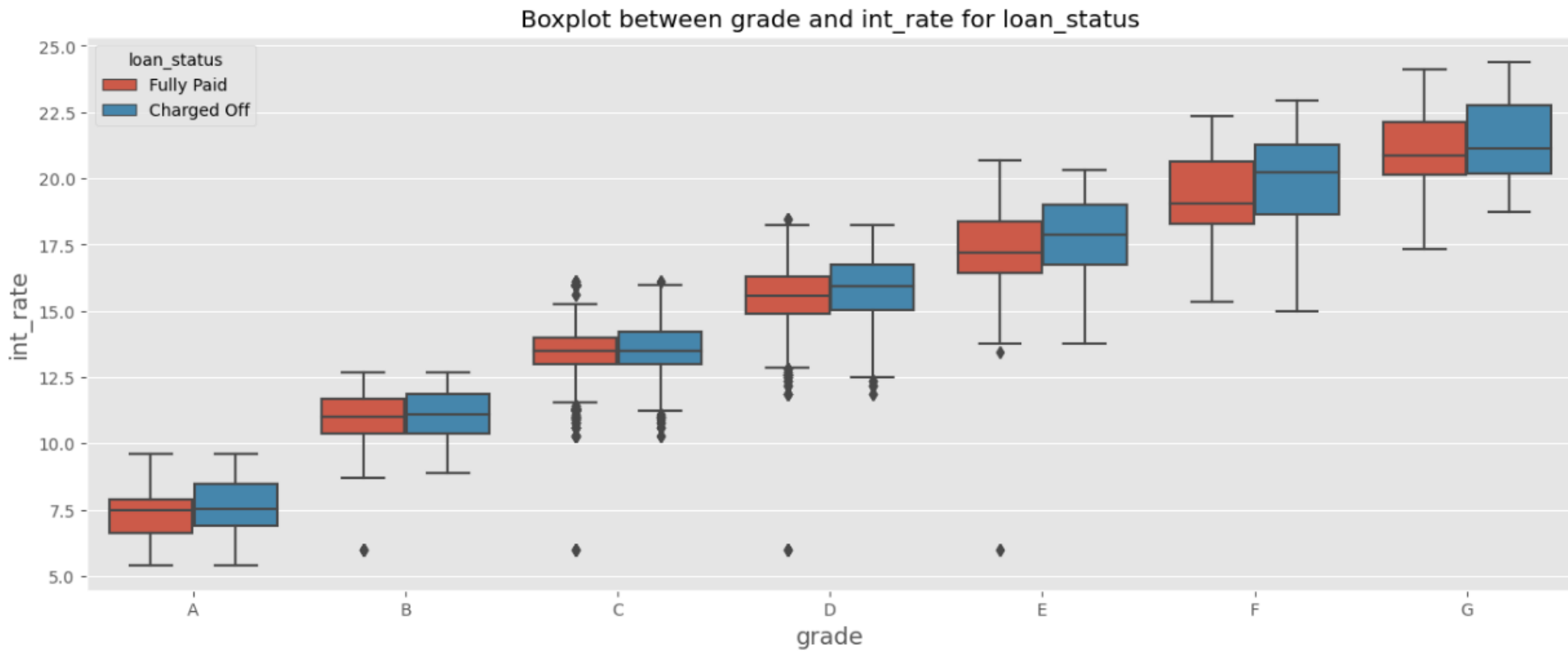
Spread of different variable would be analysed using the box plot

BOX PLOT BETWEEN TERM AND INTEREST RATE FOR LOAN STATUS



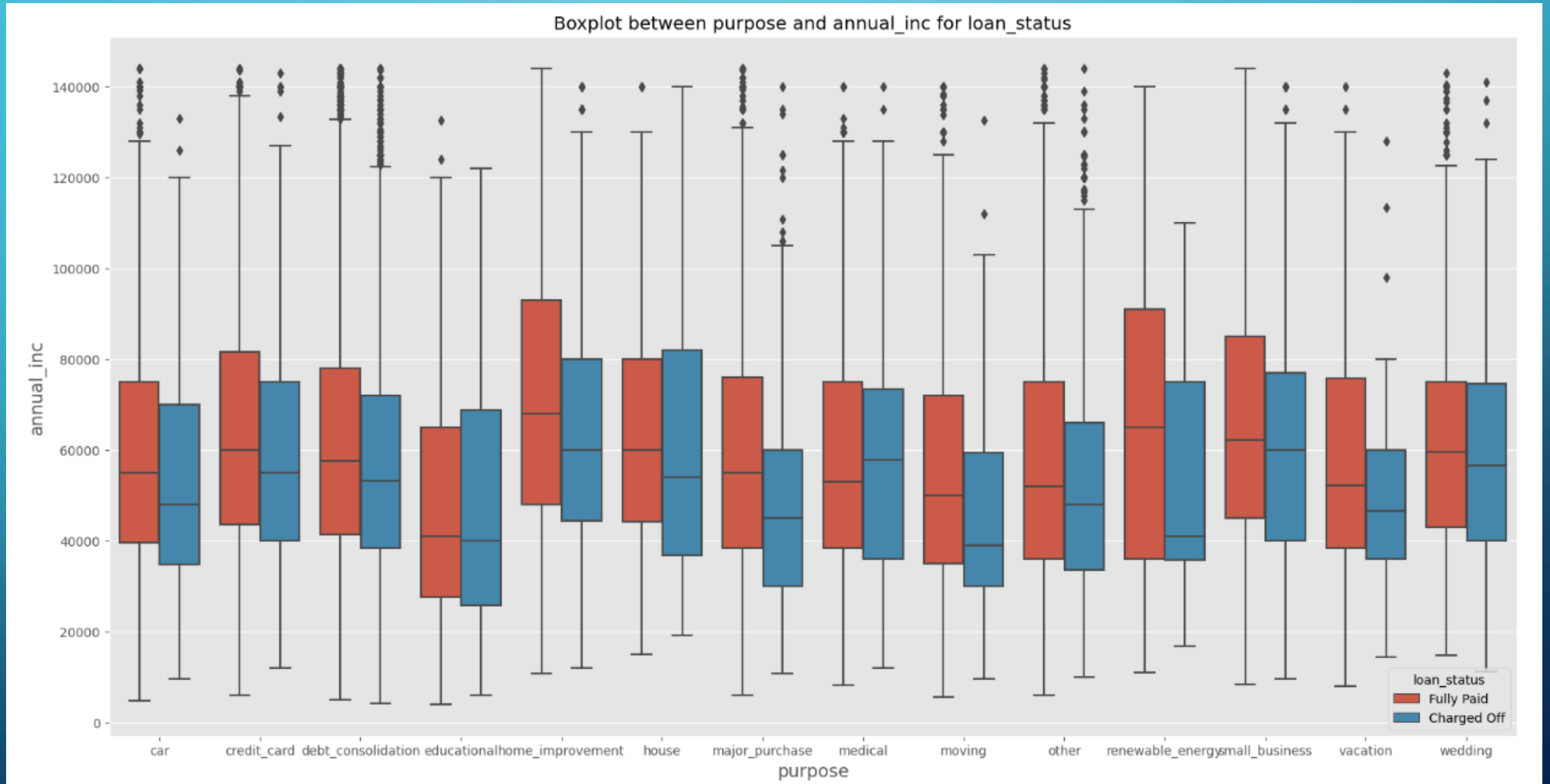
As visible in the above graph, the interest rate is higher for 60 term loan. Moreover, interest rate is also higher for the people who are defaulters.

BOX PLOT BETWEEN GRADE AND INTEREST RATE FOR LOAN STATUS



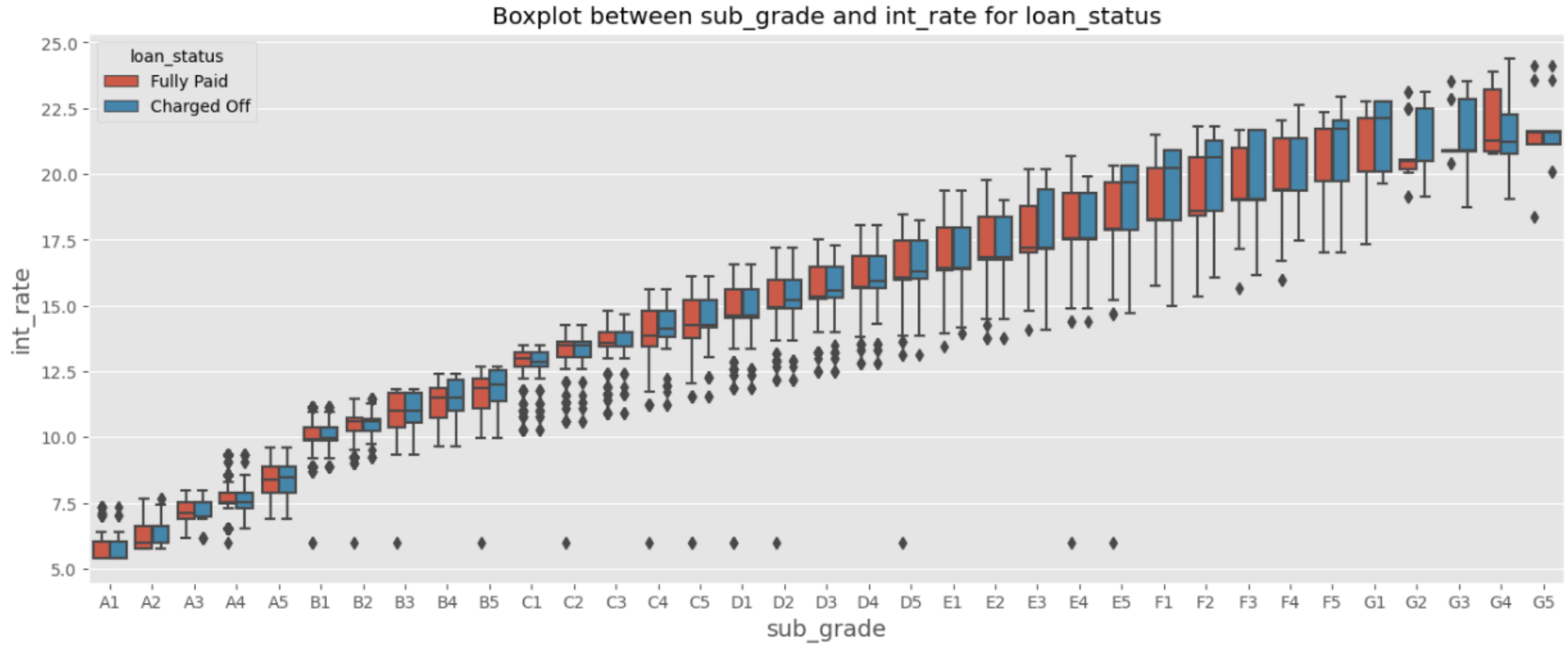
As evident from the above graph, as the grade increase then the interest rate increases. Furthermore, the interest rate for the defaulted person in generally in all the grades.

BOX PLOT BETWEEN PURPOSE AND ANNUAL INCOME FOR LOAN STATUS



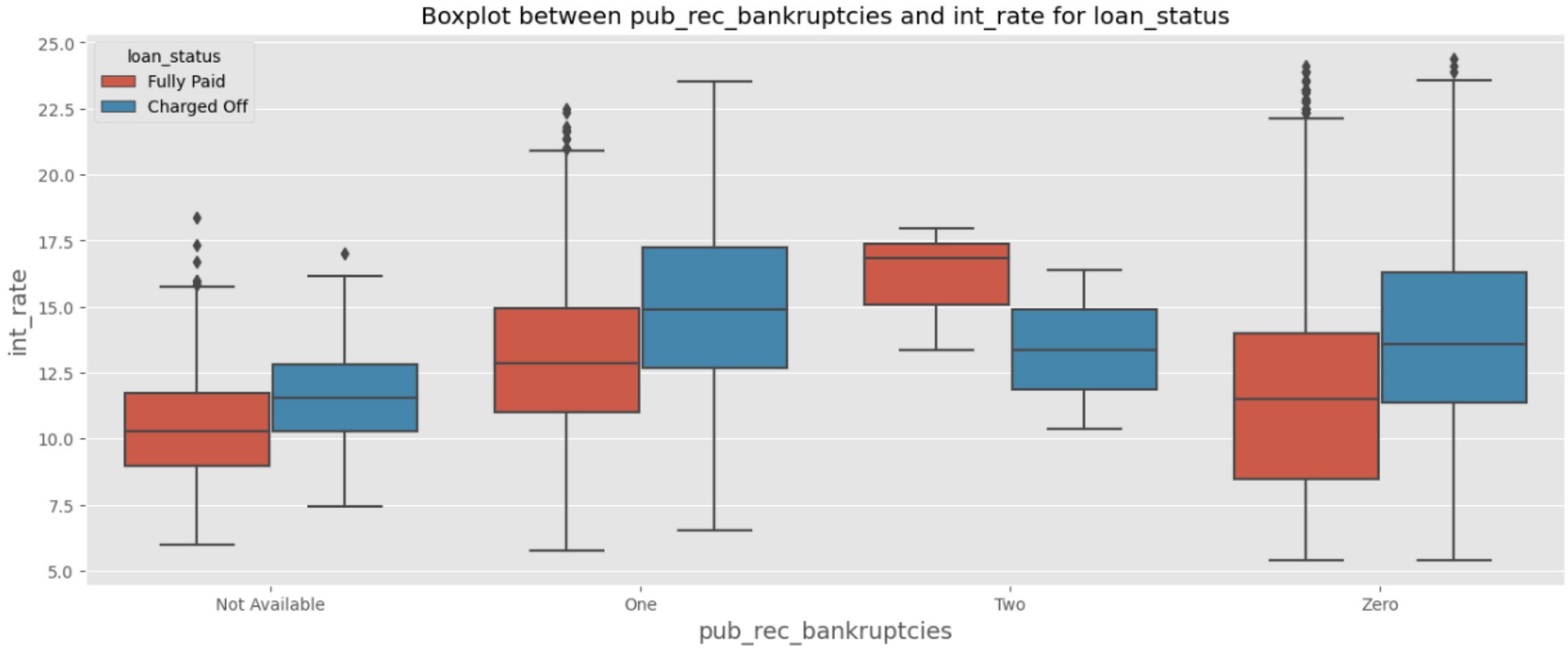
As evident from the above graph, there is not clear pattern visible in the boxplot between the annual income with respect to the purpose for which the loan is taken.

BOX PLOT BETWEEN SUB GRADE AND INTEREST RATE FOR LOAN STATUS



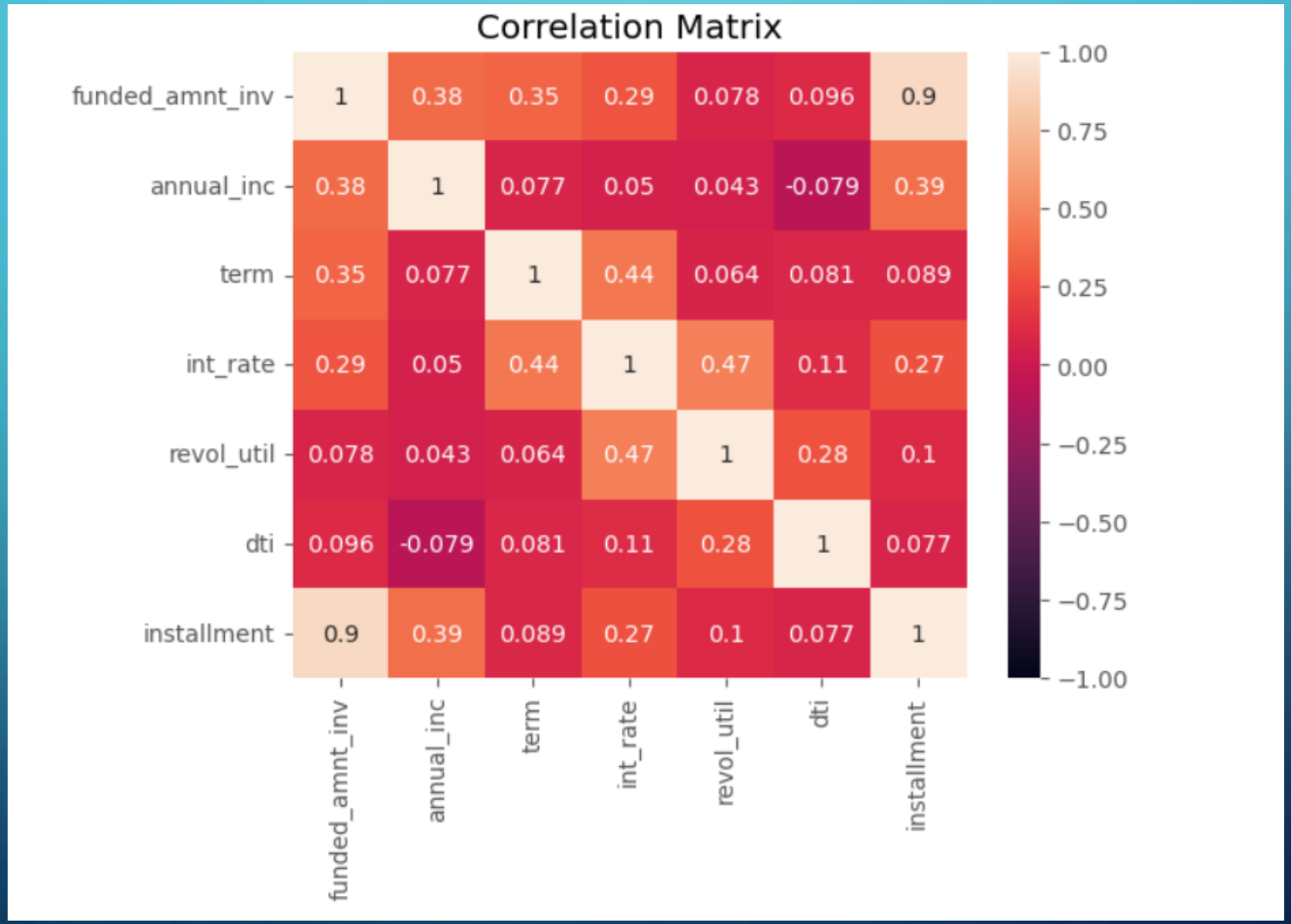
As evident from the above graph, as the level of the sub grade increases the interest rate increases.

BOX PLOT BETWEEN PUBLIC RECORD OF BANKRUPTCIES AND INTEREST RATE FOR LOAN STATUS



As evident from the above graph, as the public record of bankruptcies increase the interest rate also increases.

MULTIVARIATE ANALYSIS- HEAT MAP



CONCLUSION

1. Interest Rate-Higher Interest rate, high chances of default
2. Loan Term- Higher the loan term higher chances of default.
3. Purpose- People who take loan for small businesses have higher chances of default.
4. Annual Income- As the Annaul income increases the chance of default decreases.
5. Revol Until- As the revol until percentge increases, the chances of default also increases.
6. Grade- As the grade increases the chances of default also increases.
7. Installement:- As the installement increases the chances of default also increases.
8. DTI: As the debt to income ratio increases the chances of default also increases.