LENDING CLUB CASE STUDY

GAURAV KHANNA 03/04/2024

The problem statement

- PROJECT AIMS TO FIND THE DEFAULTERS ANALYSIS

- BACKGROUND OF THE CASE STUDY IS THAT THE COMPANY IS THE LARGEST ONLINE LOAN MARKETPLACE,

FACILITATING PERSONAL LOANS, BUSINESS LOANS, AND FINANCING OF MEDICAL PROCEDURES.

- BORROWERS CAN EASILY ACCESS LOWER INTEREST RATE LOANS THROUGH A FAST ONLINE INTERFACE.

- SO THE DEFAULTS ARE LARGE, AND IT WAS REQUIRED TO IDENTIFY THE DEFAULTERS REASON AND WHAT TO AVOID TO REDUCE DEFAULTERS

The analysis approach briefly

STEP 1 : ANALYZE THE DATASET

STEP 2 : CHECK FOR INFO, SHAPE , DATA TYPES OF THE COLUMN, DESCRIBE

STEP 3 : DO DATA CLEANING

a. DUPLICATE CHECK
b. MISSING VALUES
c. DROP COLUMNS WHERE ALL VALUES ARE NULL ALL, OR UP TO 30%
d. CORRECT THE VALUE FOR THE COLUMNS WHICH HAVE LOW MISSING VALUES BY FINDING MODE OR MEDIAN VALUES AND FILL IT

STEP 4 : CHECK VARIABLE "LOAN STATUS" IN TOTAL LOANS ISSUED

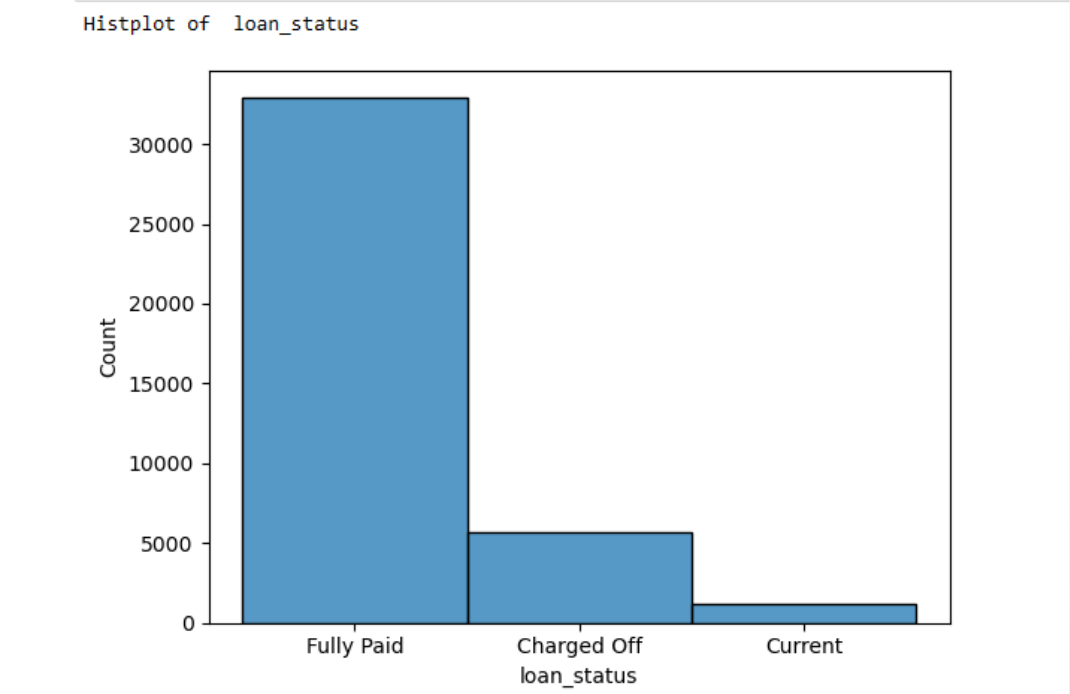
STEP 5 : UNIVARIATE ANALYSIS - NUMERICAL VARIABLES

STEP 6 : BIVARIATE ANALYSIS

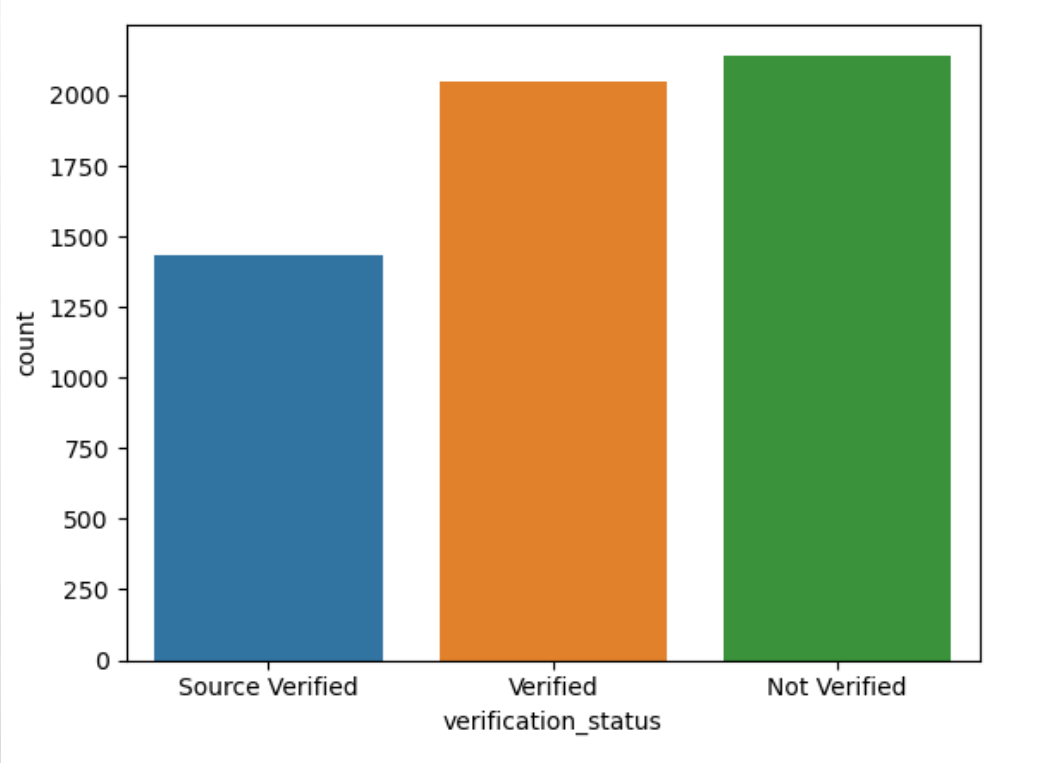
Explain the results of univariate, bivariate analysis etc. in business terms

- ALMOST 14 ~ 15% MEMBERS HAS DEFAULTED THE LOAN
- OUT OF ~5K DEFAULTERS, AROUND 40% MEMBERS WHERE NOT VERIFIED
- FOUND THAT ANNUAL INCOME OF ABOVE 20L IS OUTLIARS'
- 20% OF THE DEFAULTERS ARE FROM "CA" CITY
- 90 % OF DEFAULTER ARE EITHER STAYING ON RENT OR MORTGAGE PROPERTIES
- GRADE TYPE C, B, D ARE THE MOST DEFAULTERS WHICH TOGETHER RESPONSIBLE FOR ALMOST 70% OF THE DEFAULTERS
- MEMBERS WITH ANNUAL INCOME OF 55K ~ 75K ARE MOST LIKELY TO BE DEFAULTERS
- SHORTER TERM MOST LIKELY TO DEFAULT LOAN
- INCASE INTERESTS RATES ARE HIGHER %, CHANCES OF DEFAULT IS HIGHER
- FOR AROUND 80% OF THE DEFAULTS HAVE LOAN INTEREST RATE BETWEEN 11~20%

Histplot of loan status → charged off is around 5K in comparison to fully paid which is more than 30K
almost 14 ~ 15% members has defaulted the loan



Histplot of verification status of defaulters
- out of ~5K defaulters, around 40% members where not verified

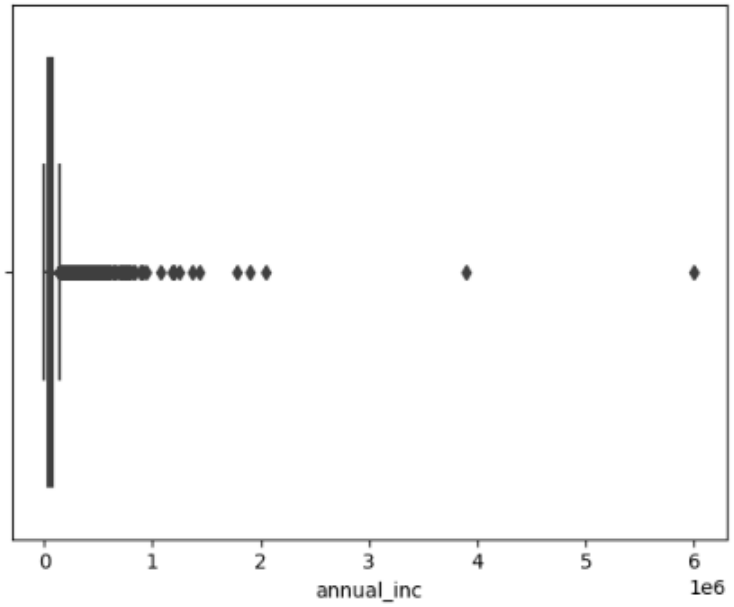


BOXPLOT TO FIND OUT OUTLIERS IN THE ANNUAL INCOME - FOUND THAT ANNUAL INCOME OF ABOVE 20L IS OUTLIARS

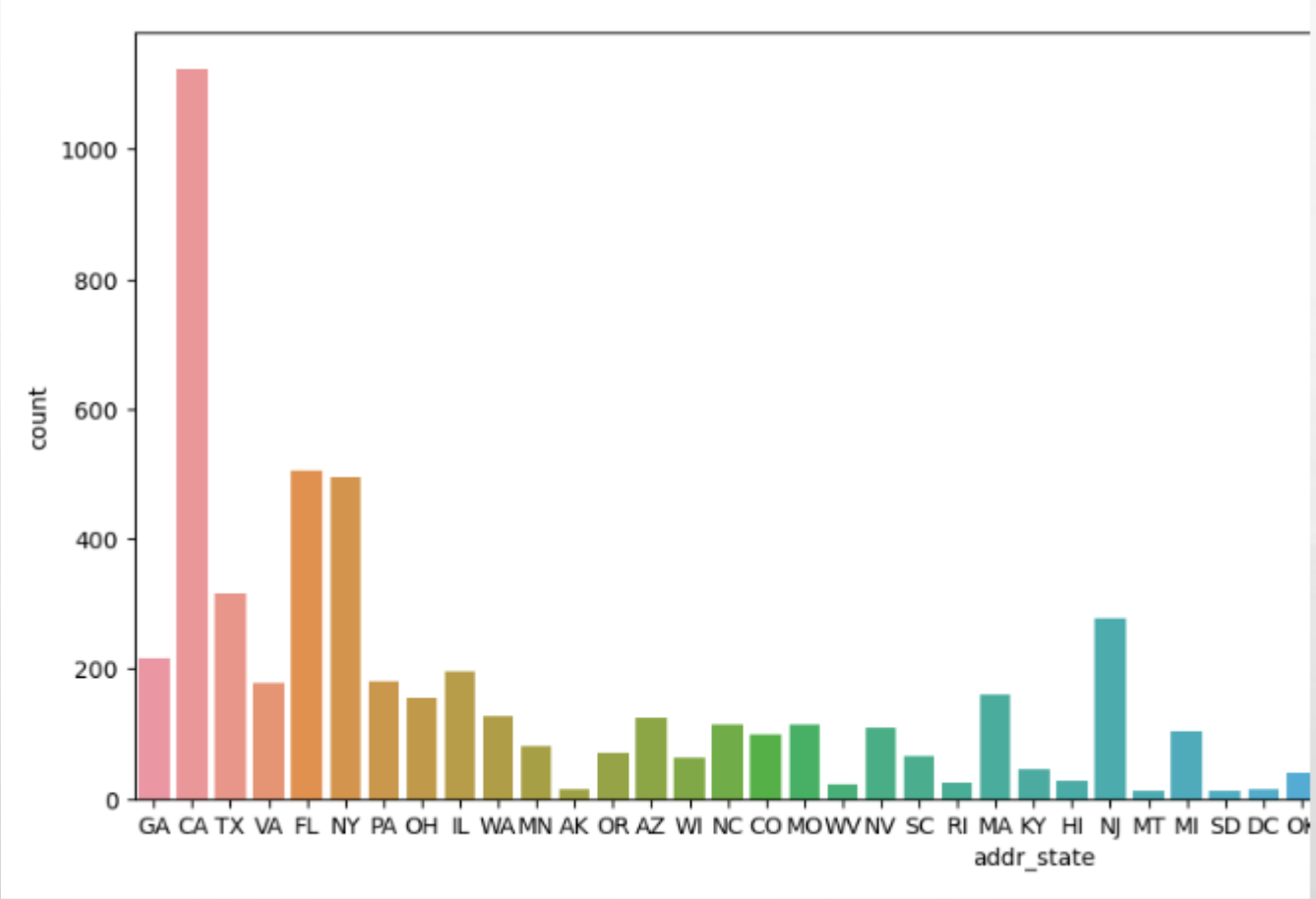
Now we need to check if the salaries have any outliars

use a box plot to find the annual income outliars

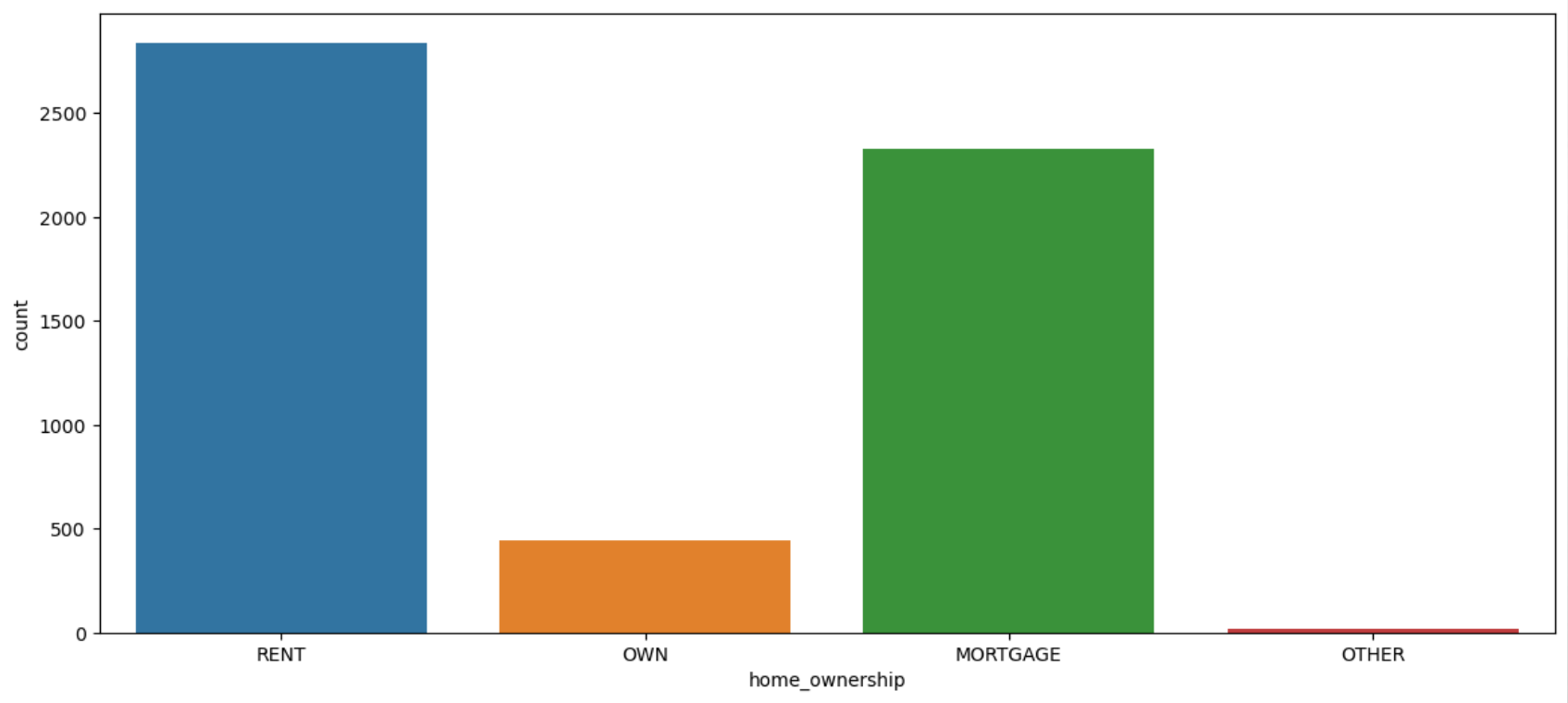
In [459]:
```
sns.boxplot(x=loan_df['annual_inc'])
plt.show()
```



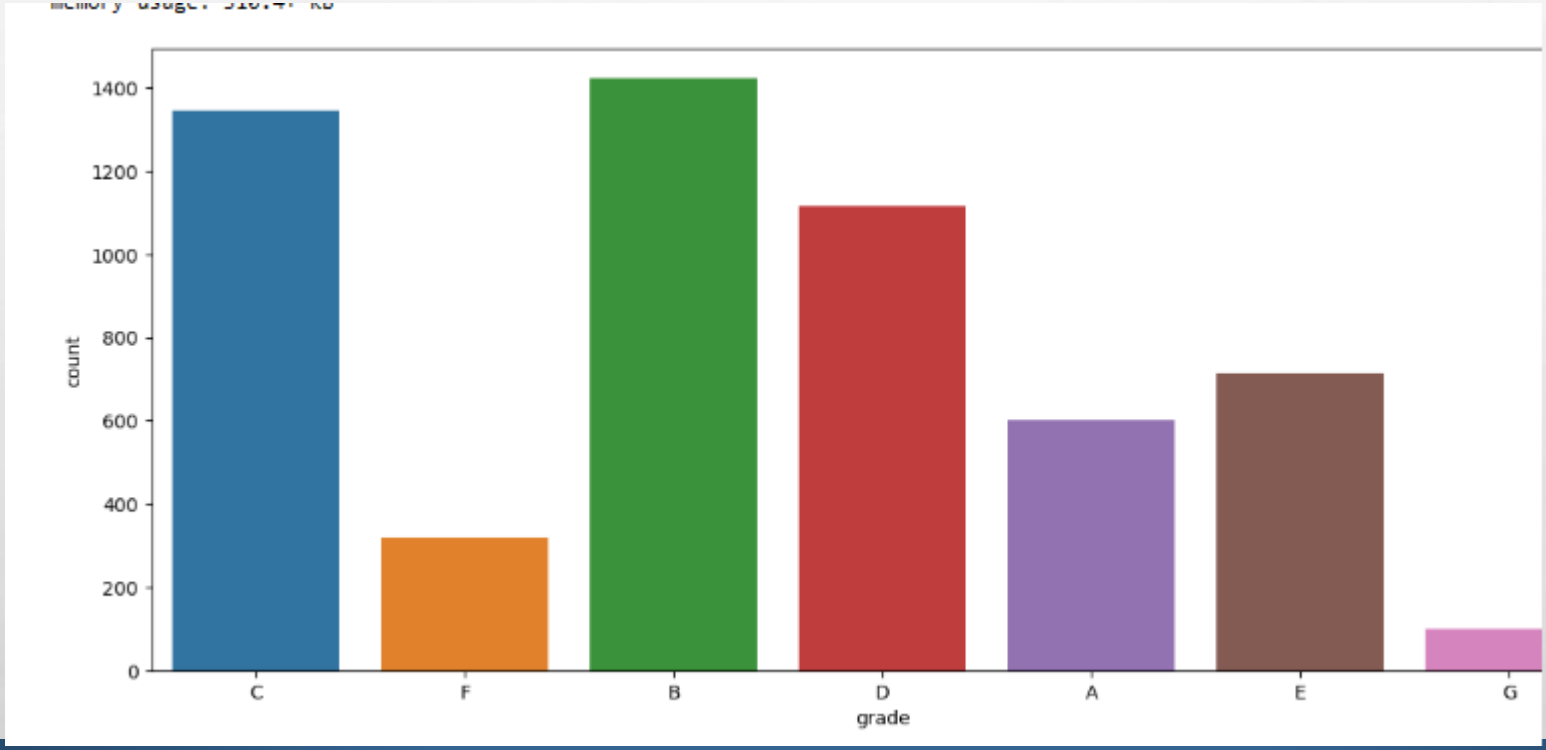
Countplot of the defaults vs address
- 20% of the defaulters are from “CA” city



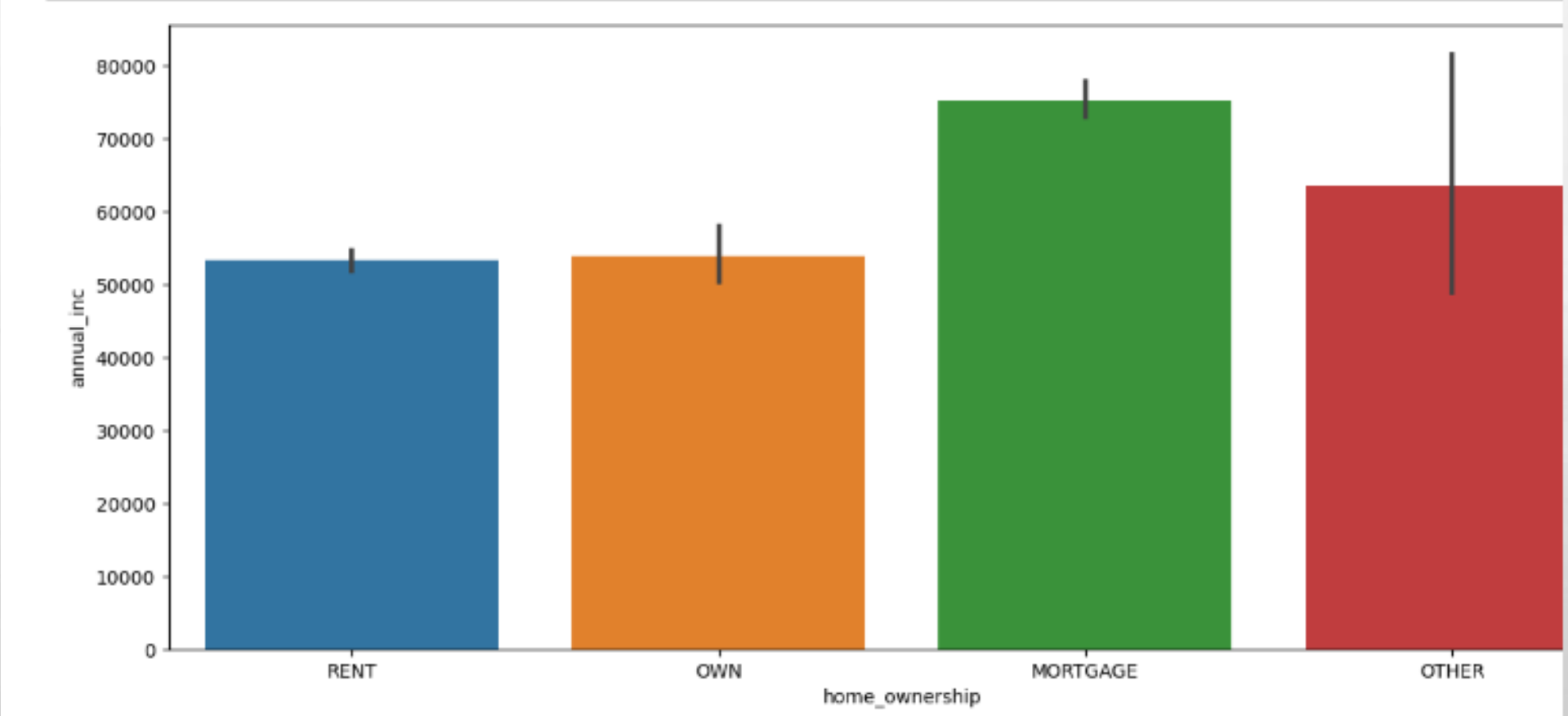
Count plot of defaulters vs home ownership
-only 10% of defaulters are home owners
- rest 90 % of defaulter are either staying on rent or mortgage properties



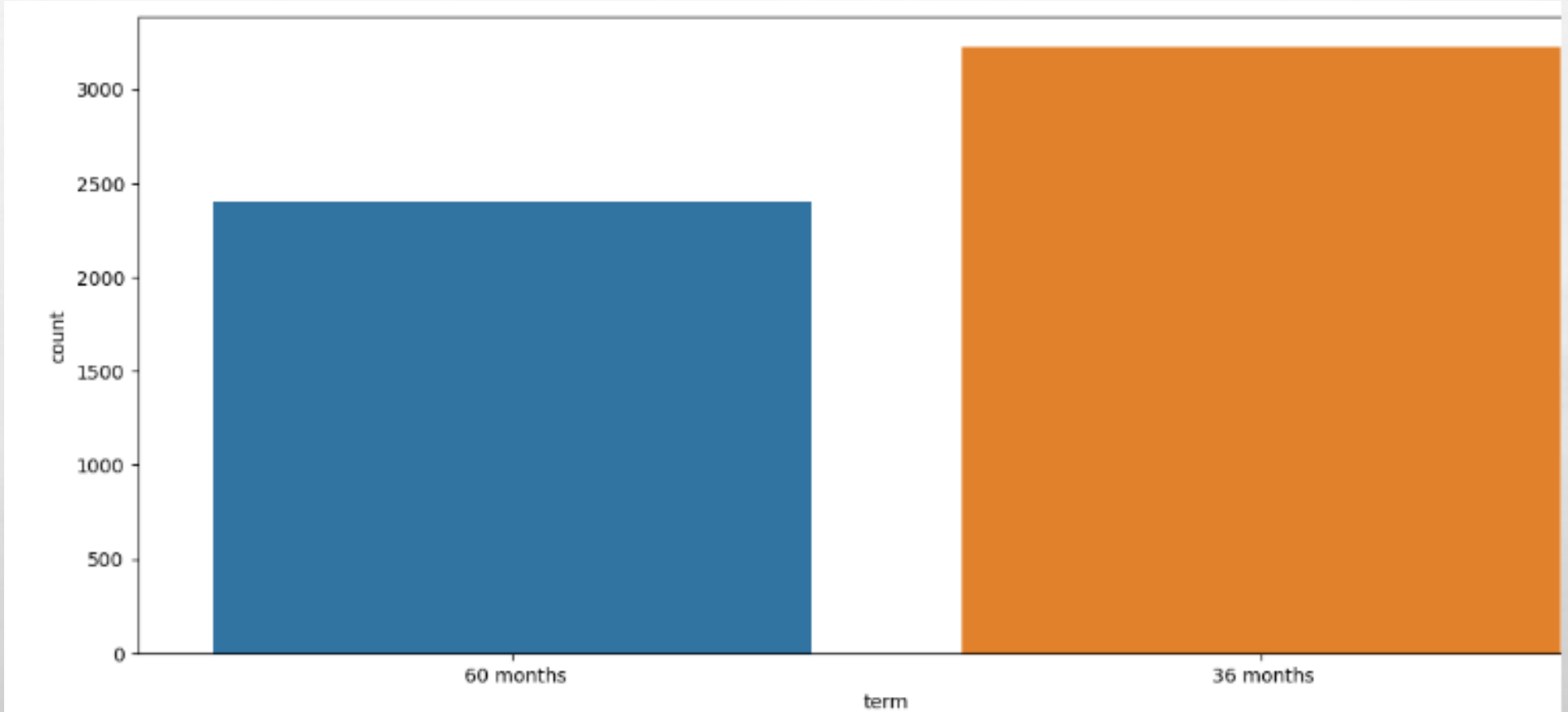
Countplot of defaulters vs grade
- grade type C, B, D are the most defaulters which together responsible for almost 70% of the defaulters



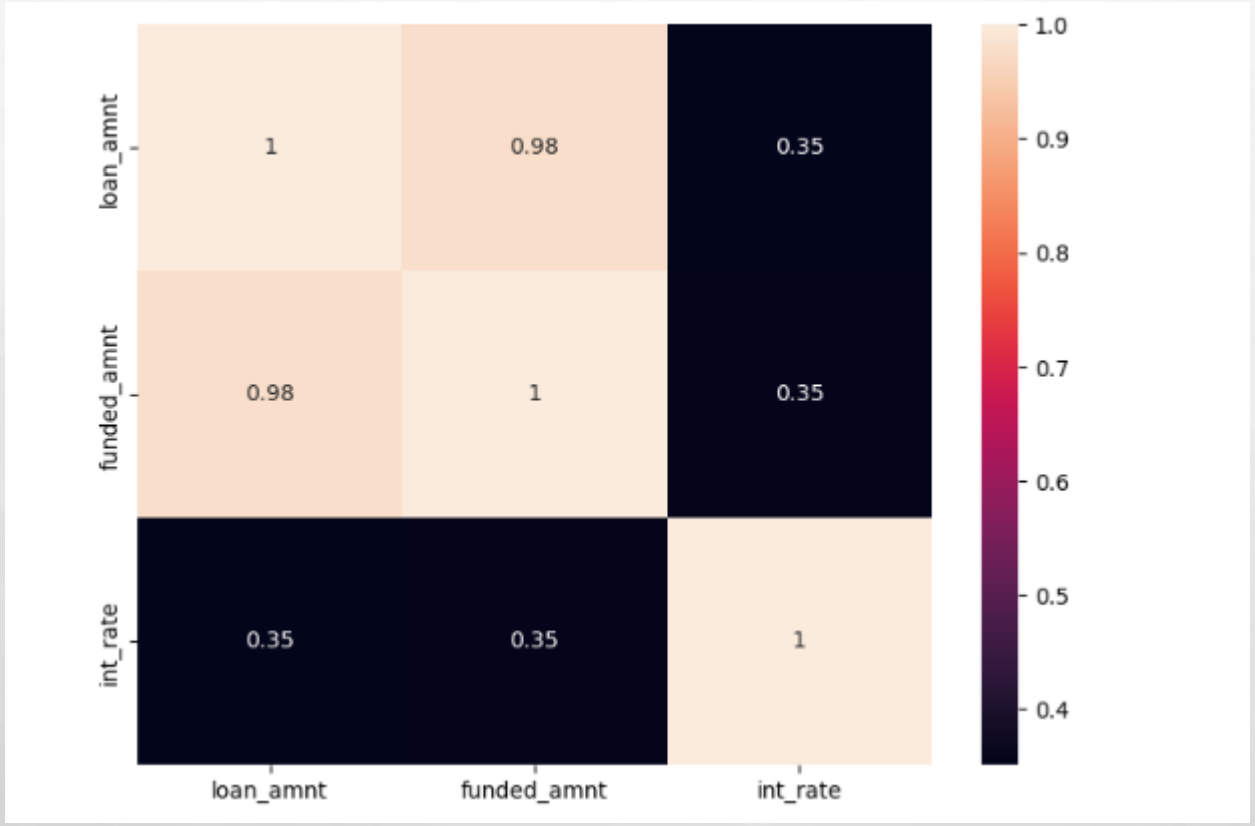
Barplot of defaulters data between ownership vs annual income
- members staying on rent have average income of around 55K
- members mortgage property have annual income of around 75K
- member staying in owned house with average income of around 55K
→ members with annual income of 55K ~ 75K are most likely to be defaulters



Countplot of defaulters vs term - shorter term most likely to default loan

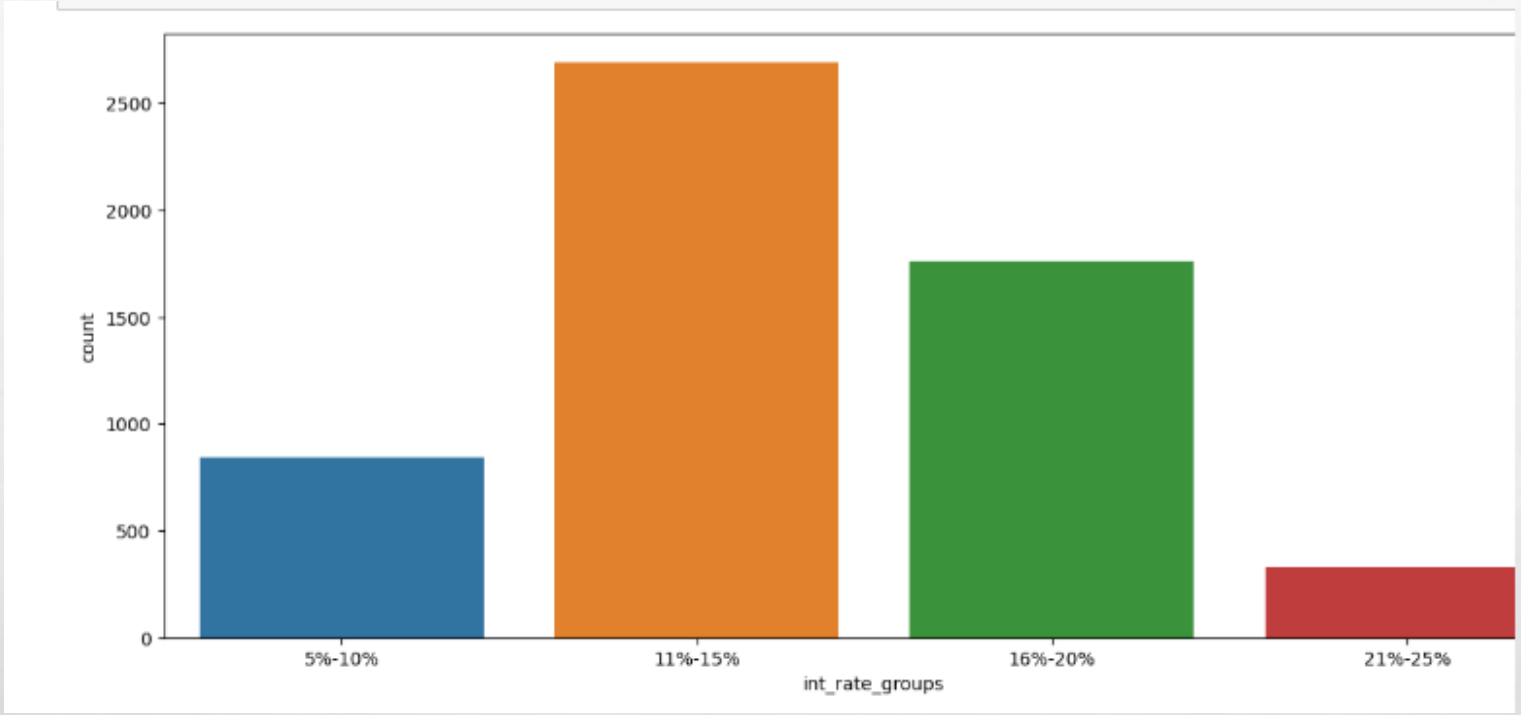


Heatmap of defaulters for 3 points
- Int rate vs funded amnt, loan amount
→Interests rates higher % for defaults



Count plot for defaults vs Interest rate

- For around 80% of the defaults have loan interest rate between 11~20%



Thank you