- **Lending Club Case Study:**
- Facilitator : Pranali Mane
- Team Member: Bitragunta Venkata Sai Vijay Aditya

Lending Club: EDA Case Study

- Problem Statement :

- To Identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.
- To find out driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default

Lending Club: EDA Case Study

- **Data Understanding :**
- All the columns are analyzed and there meaning has been learned from Data dictionary
- Columns having more than 80% Null values are dropped
- Data set contain some Customer behavior variables such as delinq_2yrs, collection_recovery_fee, last_pymnt_d etc. This are post loan approval variables, thus they cannot be used as predictors for credit approval. Hence they are dropped
- Data set contained some Single_valued columns (e.g. pymnt_plan, collections_12_mths_ex_med, policy_code) are also dropped.
- Few columns in dataset does not contribute to analysis, this are identified and delete (e.g. 'id', 'member_id', 'emp_title', 'url' etc)

Lending Club: EDA Case Study

- **Data Understanding :**

  Columns which were considered for Analysis are as follows :

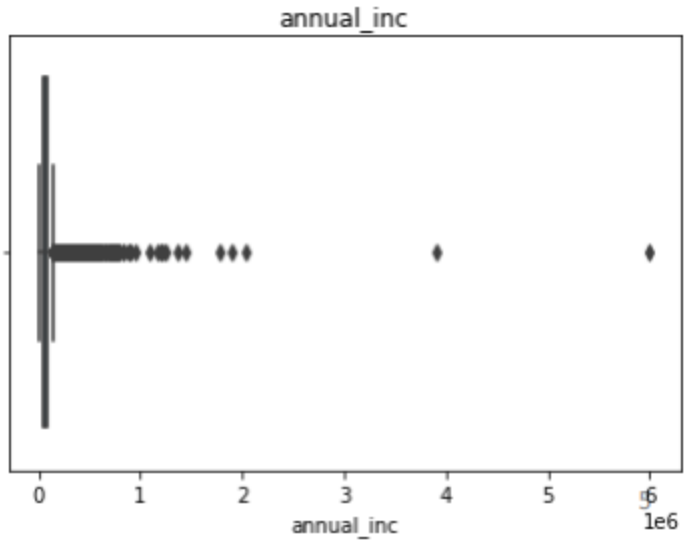
- loan_amnt
- funded_amnt
- funded_amnt_inv
- term
- int_rate
- installment
- grade
- sub_grade
- emp_length
- home_ownership
- annual_inc
- verification_status
- issue_d
- loan_status
- purpose
- addr_state
- dti
- pub_rec_bankruptcies

Lending Club: EDA Case Study

Data Cleaning and Imputation :

- Null values present in columns have been identified and treated wherever necessary
- Data is segmented into Continuous and categorical variables
- Boxplots are plotted for continuous variables to check presence of outliers
- For Simplification purpose, binning is done for few continuous variable

It is observed that, annual_inc column has very uneven distribution due to outliers, hence data beyond 95 percentile is dropped

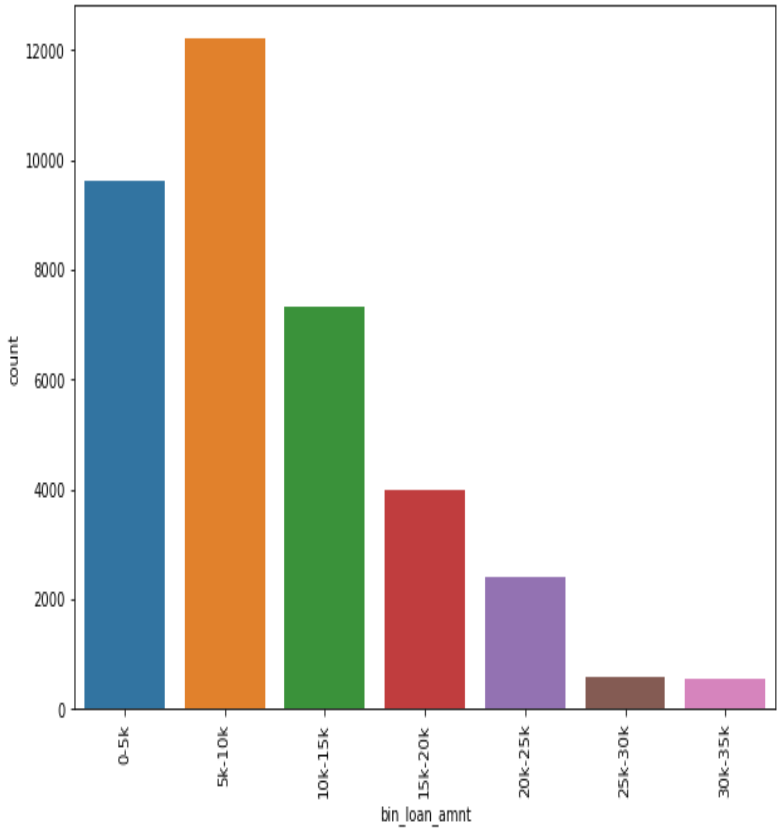


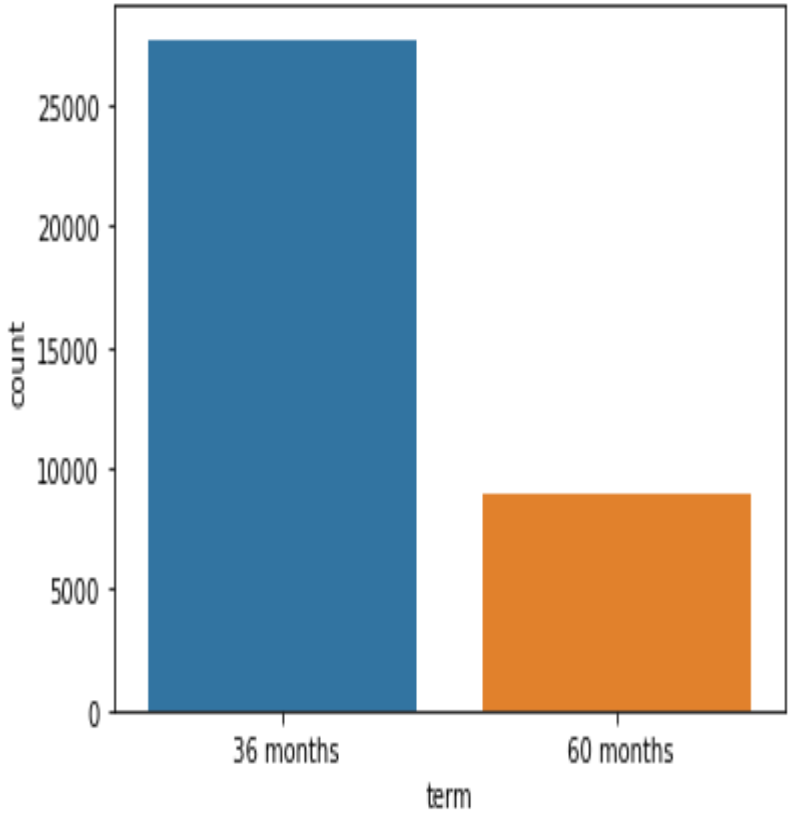
Lending Club: EDA Case Study

- There are high Chances of loan defaulting when
  - Term is of 36 months
  - loan grade is 'B‘
  - Home ownership status is "Rent“
  - Verification status is "Not verified“
  - For "debt consolidation" purpose
  - for applicant belonging to "CA" state
  - Loan issued in last month "december" of the year
  - For loan amount in the range of "5k-10k“
  - having annual income in between 2lac-3lac
  - having installment in between 100-200
  - DTI in range 10-15

Lending Club: EDA Case Study

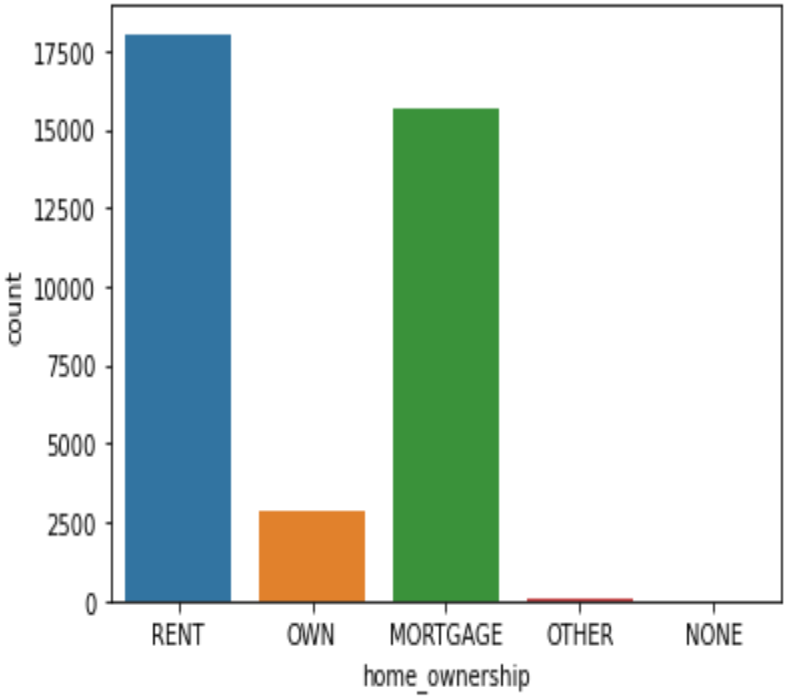
Univariate Analysis

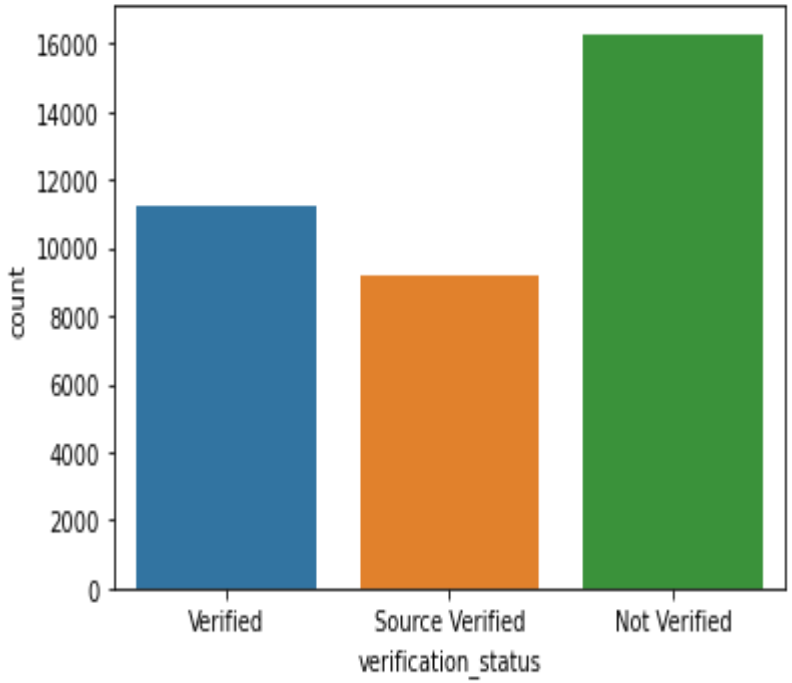




Lending Club: EDA Case Study

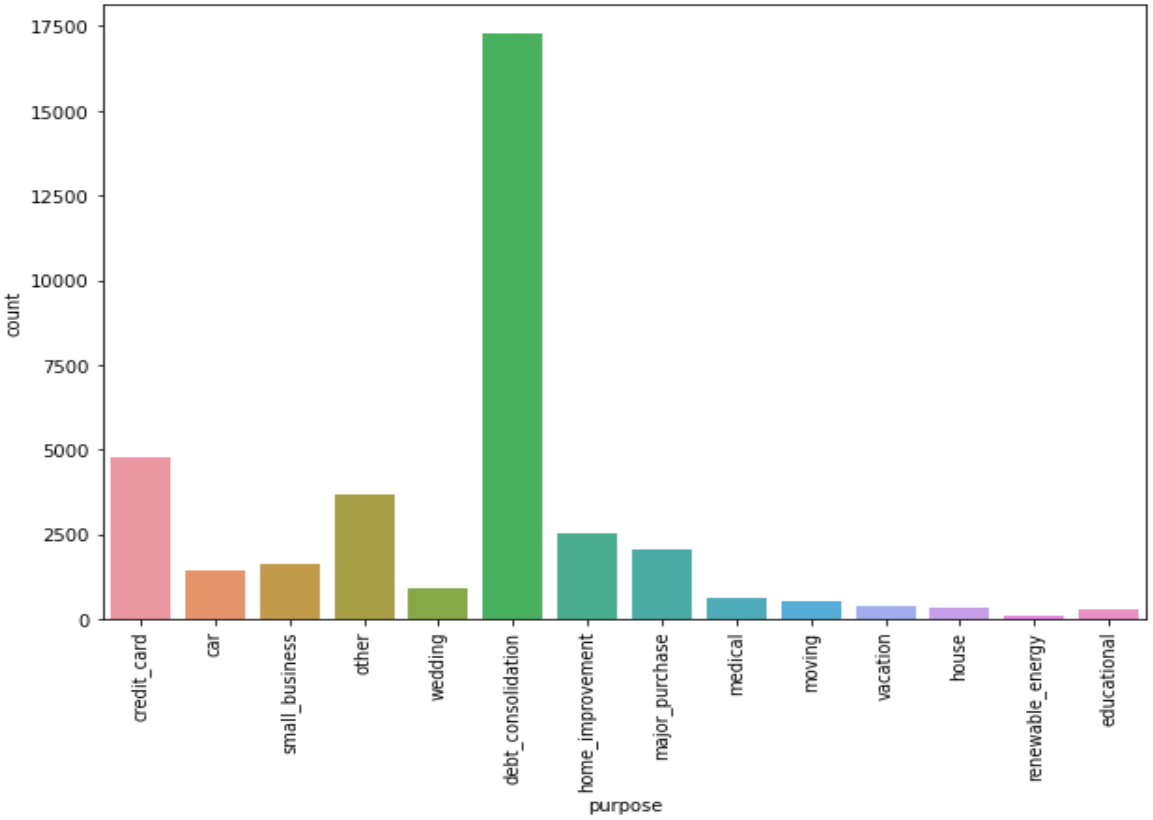
Univariate Analysis





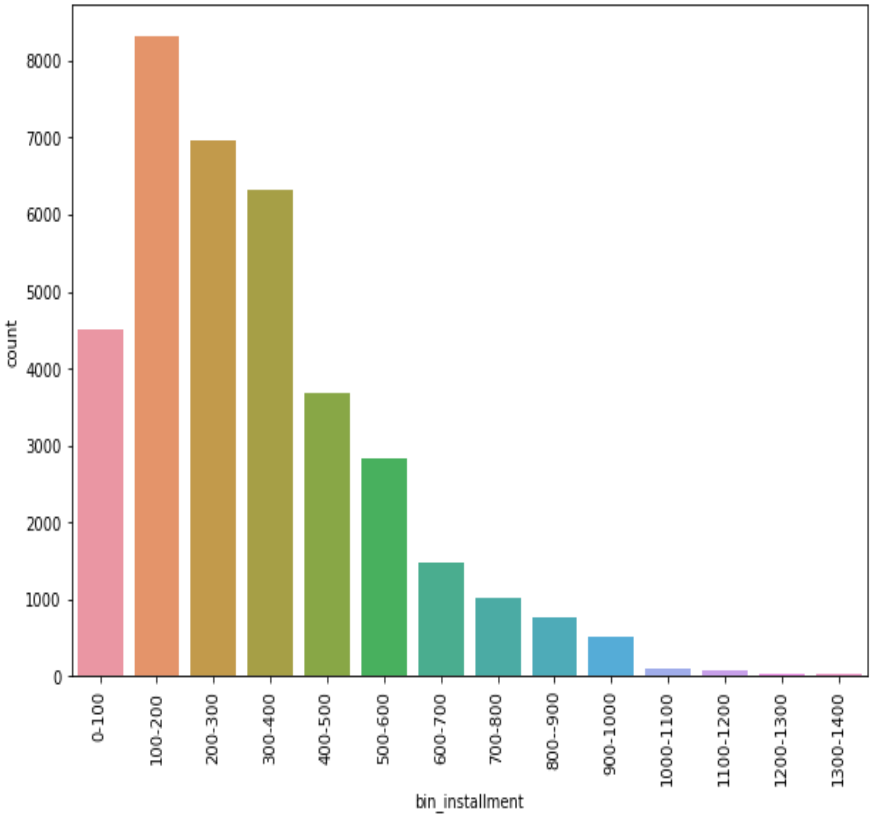
Lending Club: EDA Case Study

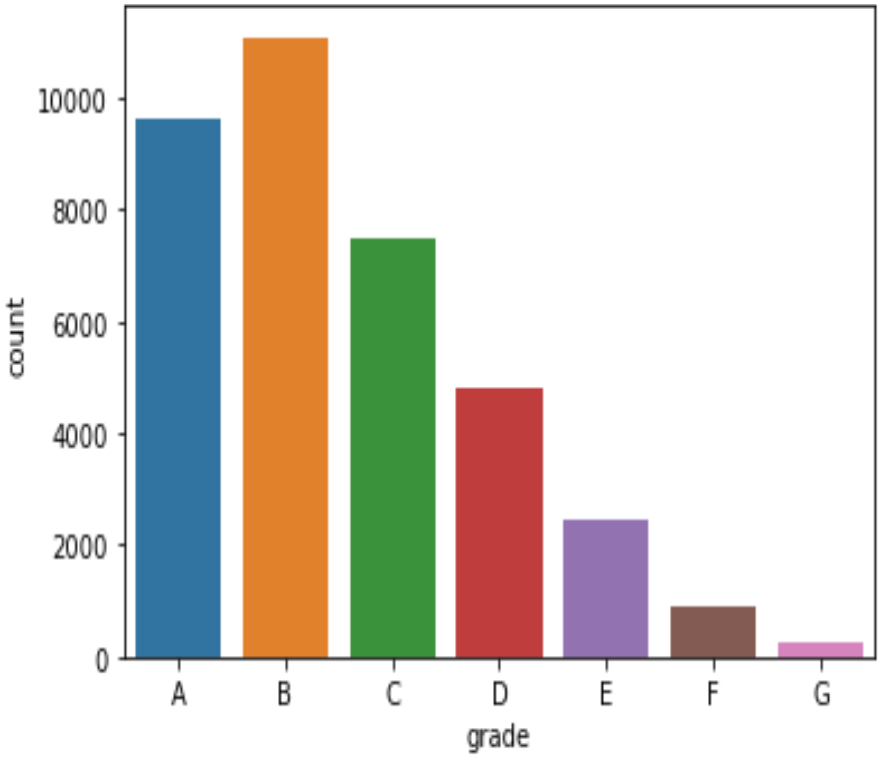
Univariate Analysis



Lending Club: EDA Case Study

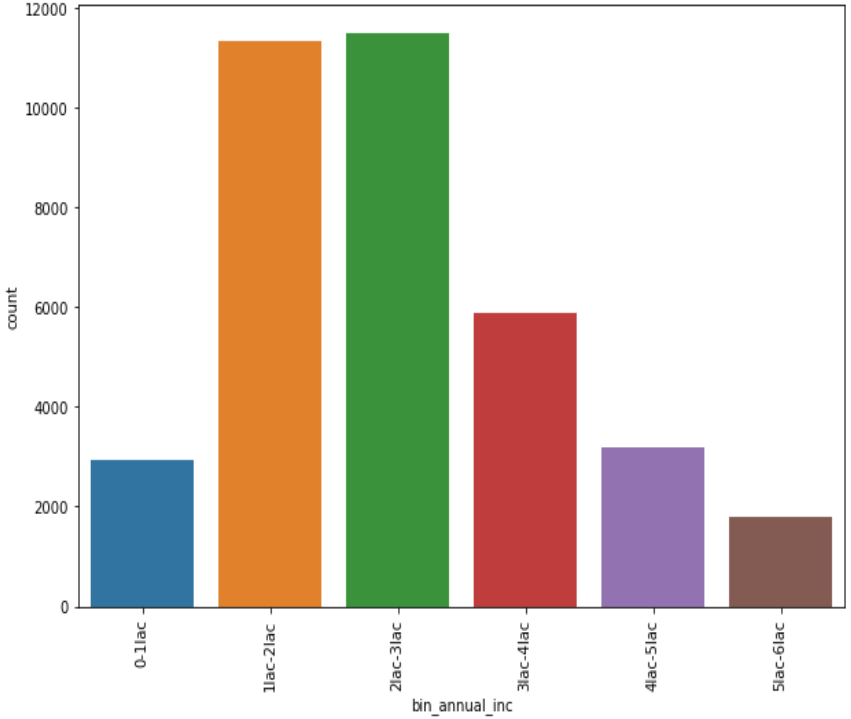
Univariate Analysis

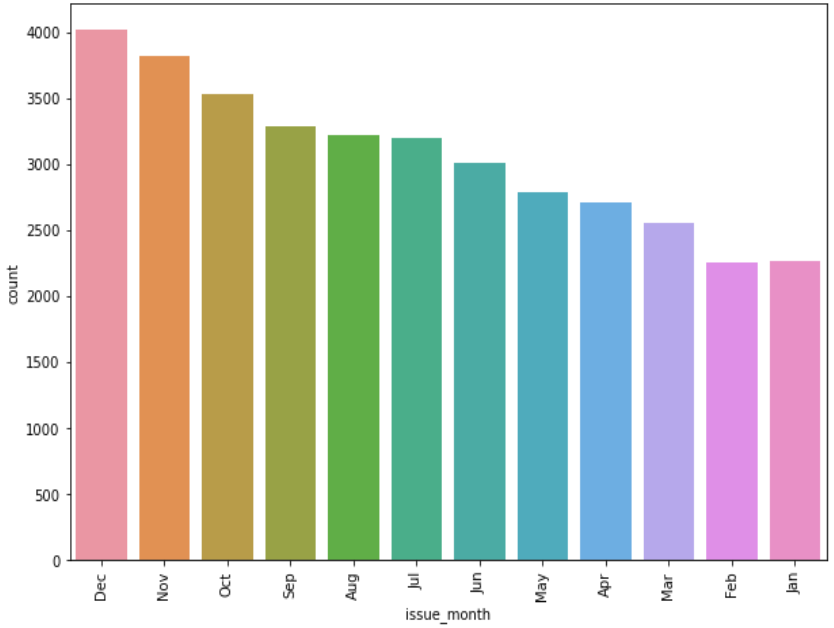




Lending Club: EDA Case Study

Univariate Analysis





Lending Club: EDA Case Study

- **Data Visualization : Bivariate Analysis**

  Below are the inference from the bi-variate analysis

  1) Applicant who choose to have 60 month of term plan have higher loan amount.

  2) Loan Amount increases with grade and so the chances of loan defaulting.

  3) Applicant having home ownership as "Mortgage" applies for more loan and has high chances of defaulting.

  4) Applicants having 10+ years of employment period and loan amount more than 12000, have high chances of loan defaulting.

  5) Interest rate is higher for more 400-500 instalments, and it has high chances of loan defaulting.

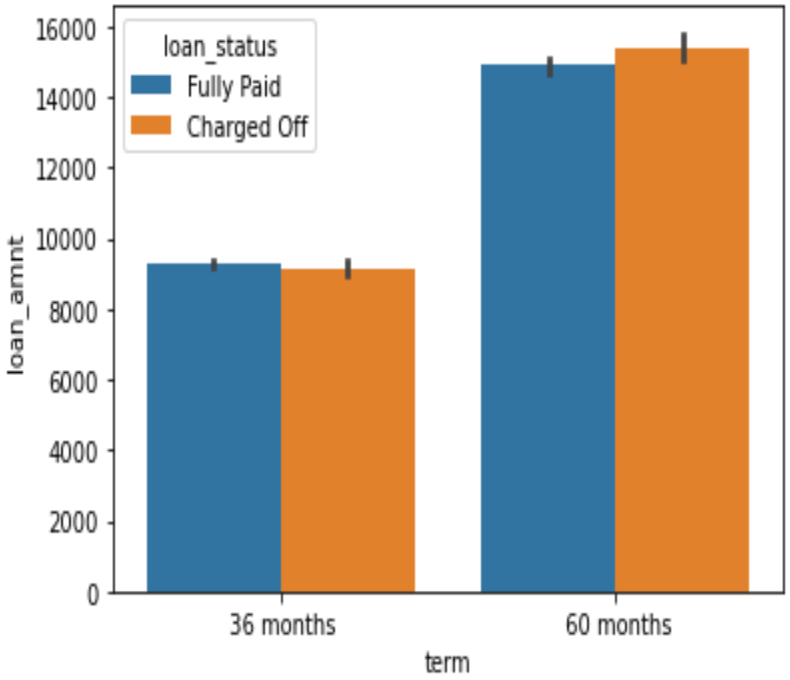
  6) People having annual income greater than 80000, gets high loan.

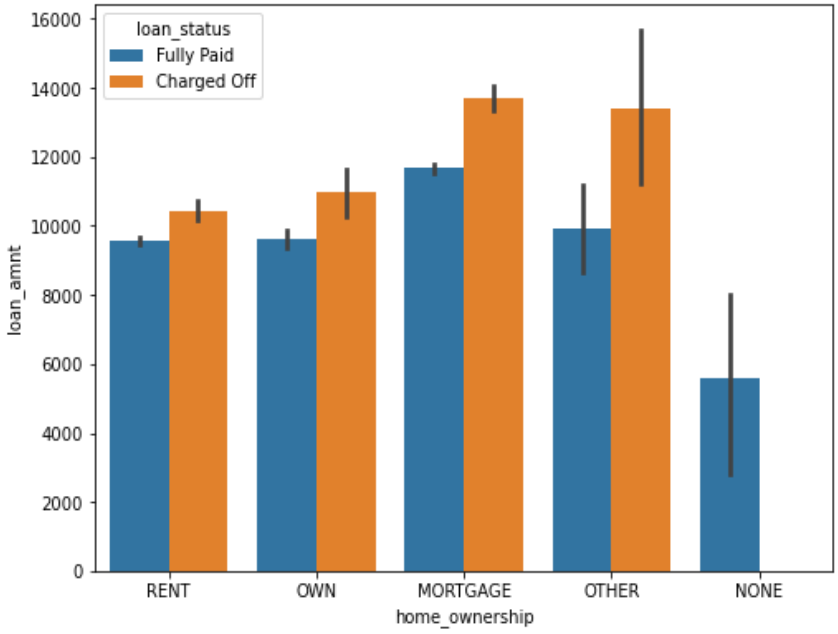
  7) People in rage of 4-6 lac PA choose to have more instalments.

  8) The people who takes the loan for purpose such as home improvement, medical, major purchase, renewable energy and education are generally not defaulters.

Lending Club: EDA Case Study

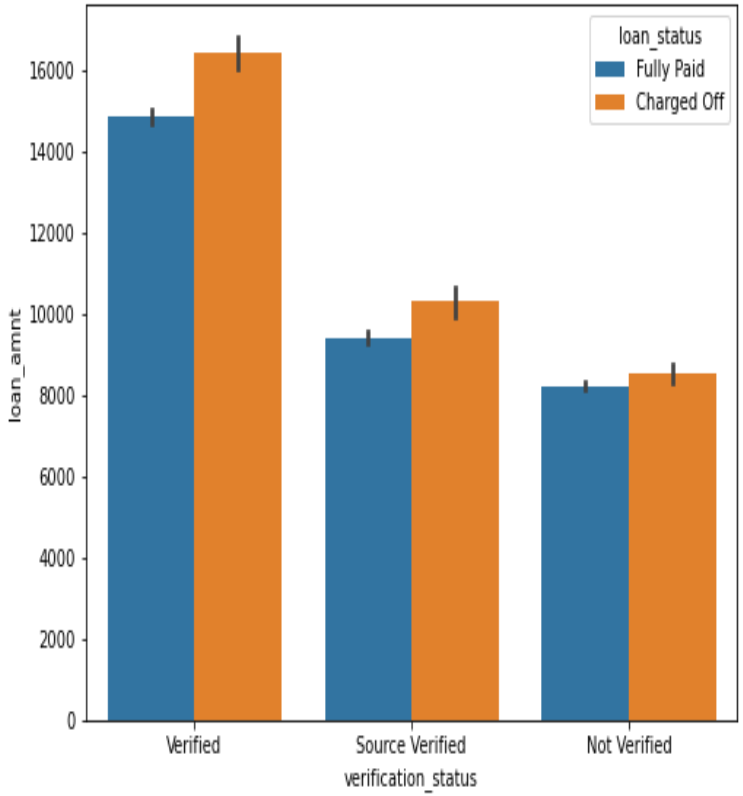
Bivariate Analysis

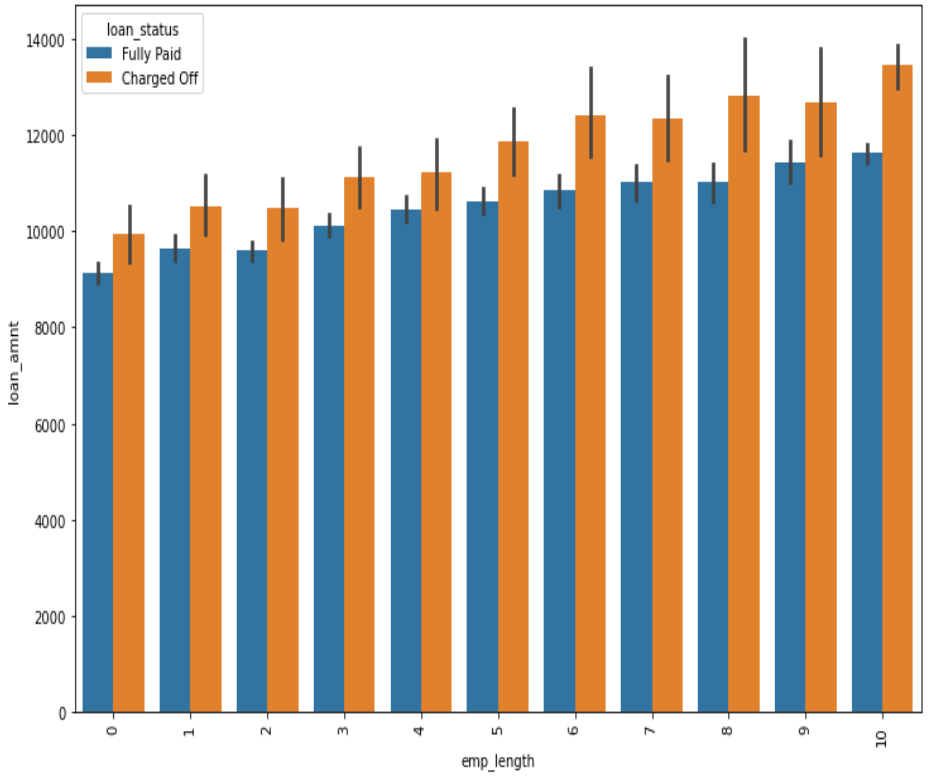




Lending Club: EDA Case Study

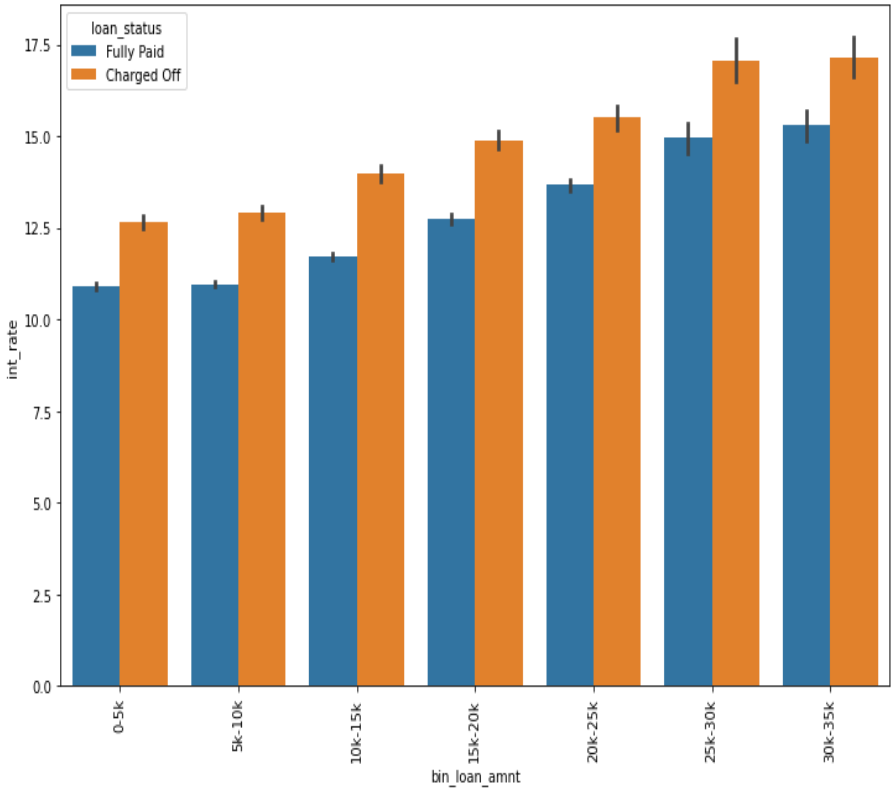
Bivariate Analysis

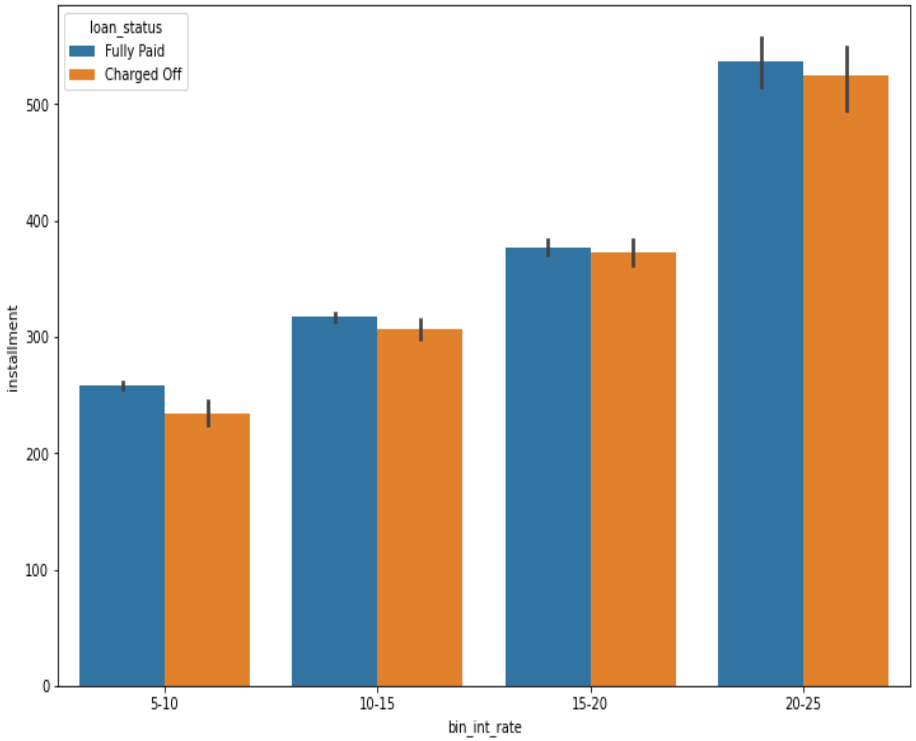




Lending Club: EDA Case Study

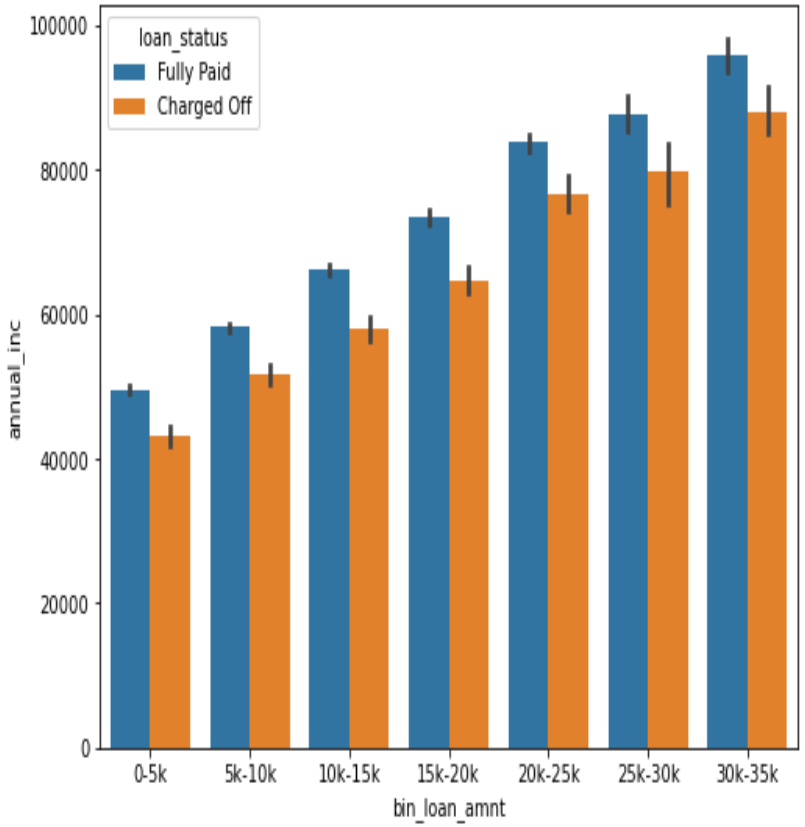
Bivariate Analysis

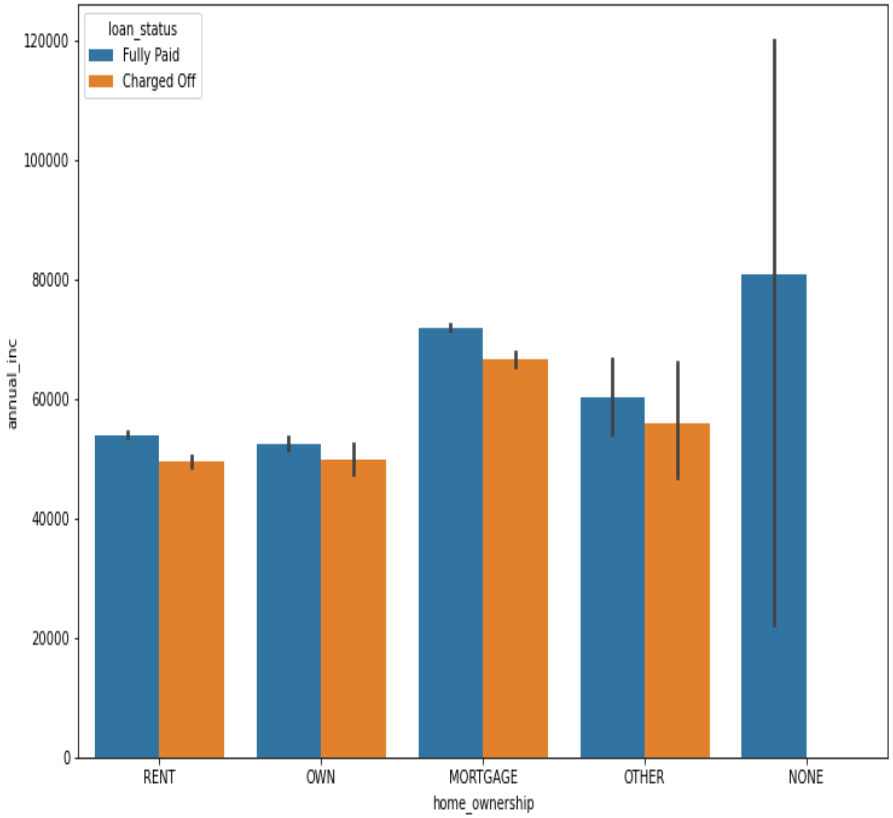




Lending Club: EDA Case Study

Bivariate Analysis





Lending Club: EDA Case Study

- **Key fields** :

  After referring to the data and perform the EDA below are the list of the driving variables which will help to define the loan defaulters

  - loan_amnt
  - term
  - int_rate
  - installment
  - grade
  - emp_length
  - home_ownership
  - annual_inc
  - verification_status
  - purpose

Conclusions / Recommendations

- Annual income is a very strong attribute for taking the loan and the loan amount is higher for higher annual income.
- If the loan tenure is more, the chances to become the defaults is less.
- Higher the grade, lesser is the chance to become defaulters.
- The riskiest purpose for the defaulters are credit card, small business, vacation and debt consolidation.
- Verified customer have less chances to become the defaulters.
- Application who are verified have high loan amount with greater tenure.
- With higher grade, interest rate is higher - People with higher salary tend to take more loan amount as compared to people having less salary.
- Loan should be provided to the application who has greater employment tenure.
- Loan should be provided to the people who own’s the house.

upGrad

Thank You!