



Lending Case Study

ANTO & HAREESH

Problem Statement:

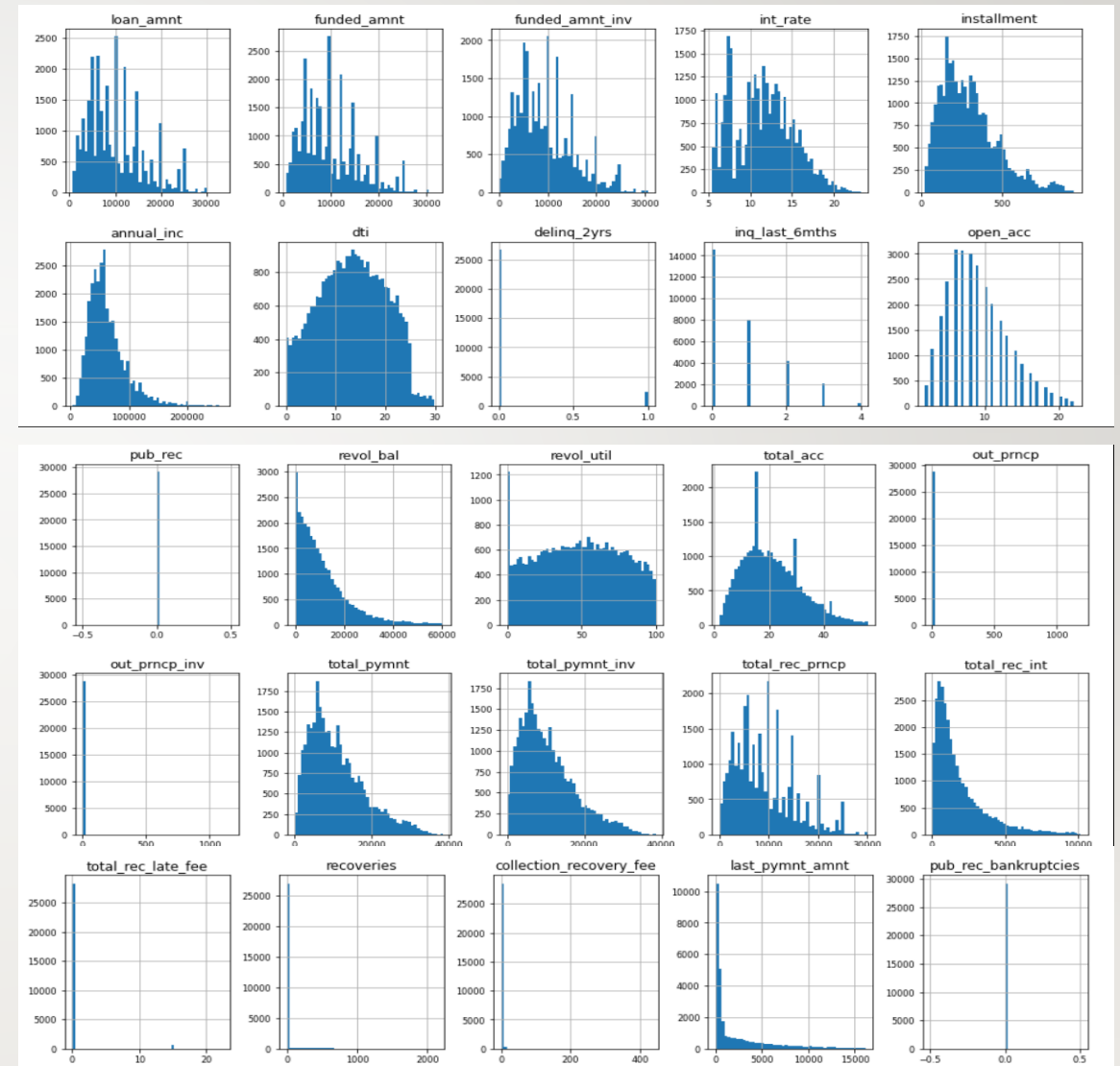
- This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.
- Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who **default** cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

Data variables Observations

- There are total 111 columns & 39,717 rows in the provided data set
- Out of 111 columns below are 3 data types
 - float64(74), int64(13), object(24)
- There are no duplicates in the dataset
- Missing/Null Values
 - 54 columns with 100% null values exists in dataset
 - 3 columns with $\geq 50\%$ null values exists in dataset
 - 10 columns with $< 10\%$ of null values
 - We have dropped Null columns with $\geq 50\%$ & rows with $< 10\%$ of null values

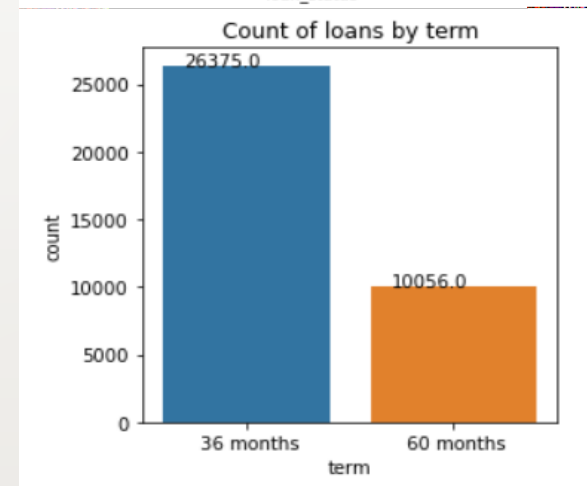
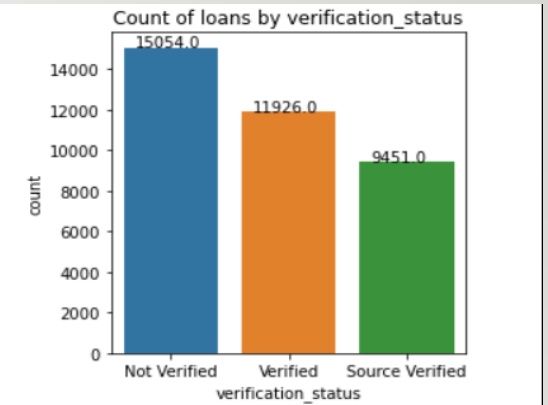
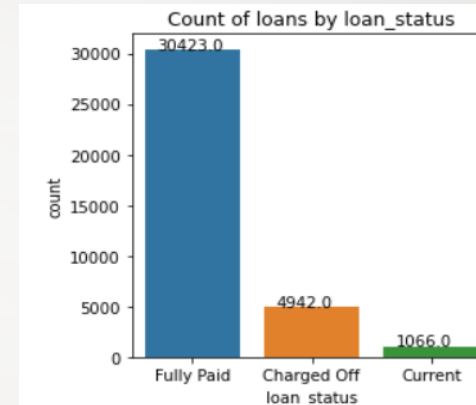
Univariate Analysis

- Most of the variables are positively skewed, it means higher number of data points are having low values
- Variables 'dti', 'revol_util', 'pub_rec_bankruptcies' & 'pub_rec' are normally distributed.
- 'loan_amnt', 'funded_amnt', 'funded_amnt_inv' plot shows that maximum number of members have received loan of amount 10,000
- 'annual_inc' plot shows that maximum number of member have annual income between 50-60k.
- 'int_rate' plot shos that maximum number of loans have interest rate range between 10-15%



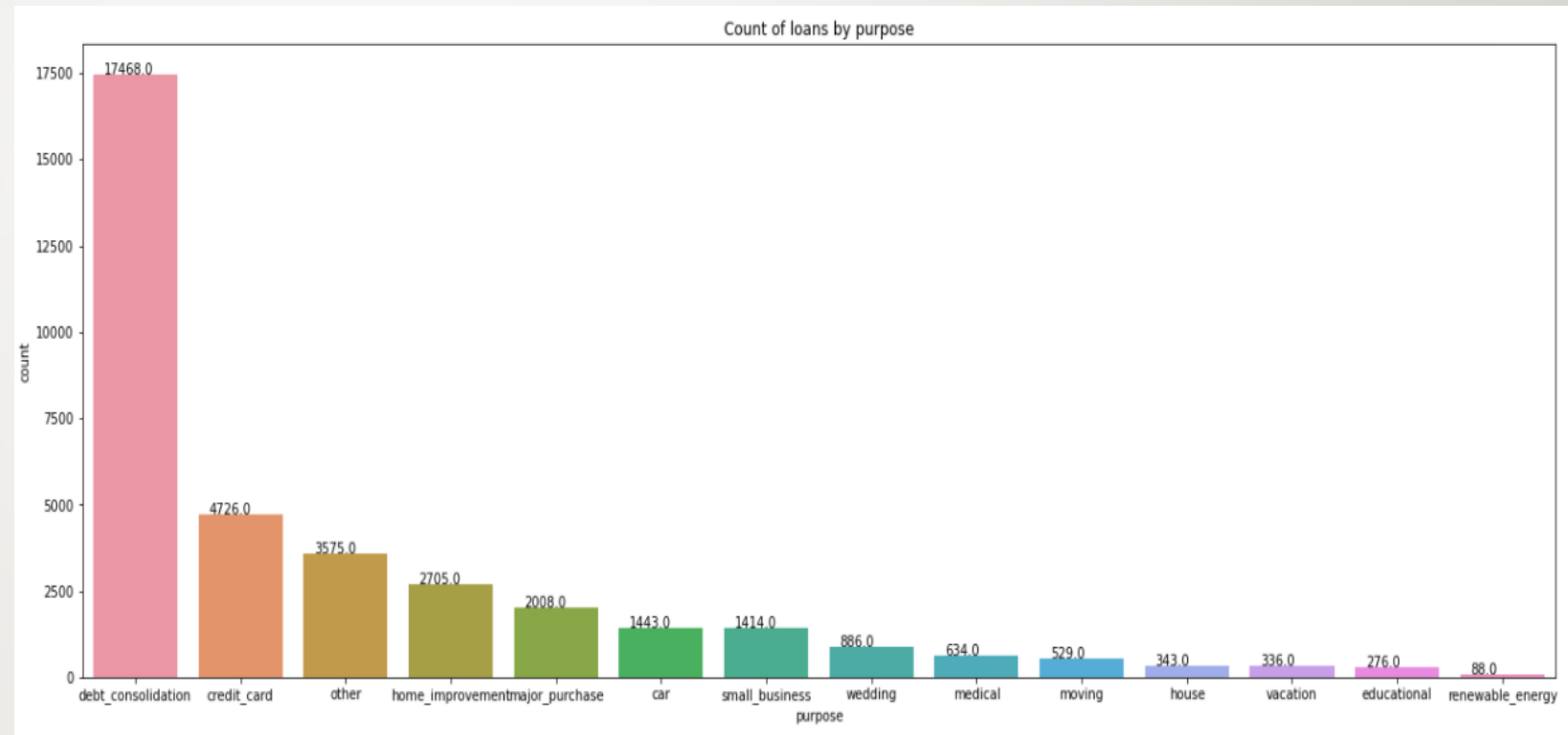
Univariate Analysis

- First plot shows that 84% of loans are fully paid and for 14% of loans applicants has not paid the instalments in due time for a long period of time and for 3% of loans installment are still in progress
- Second graph shows that, 41% of loans are not verified, which gives an indication of high risk.
- Second graph shows that , 33% of loans are verified and 26% of loans are source verified
- Third graph shows that maximum no. of loans have term of 36 months



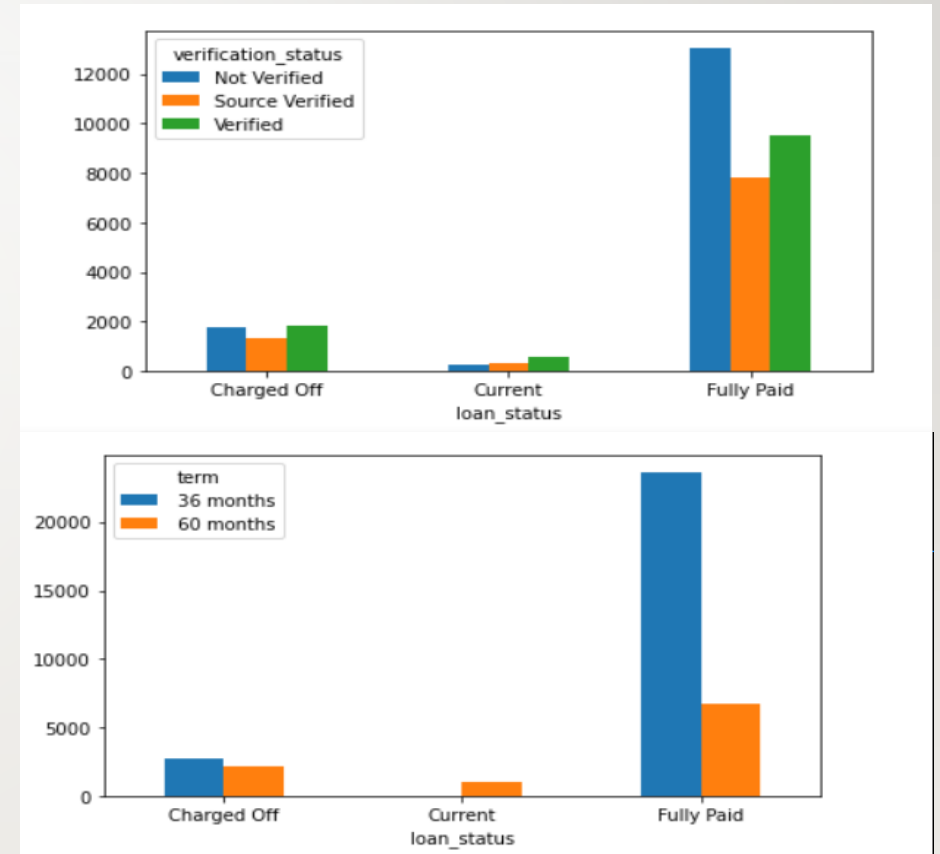
Univariate Analysis

- 48% of loans applications are received under category "Debt_Consolidation"
- 13% applicants have taken loan in the form of credit card



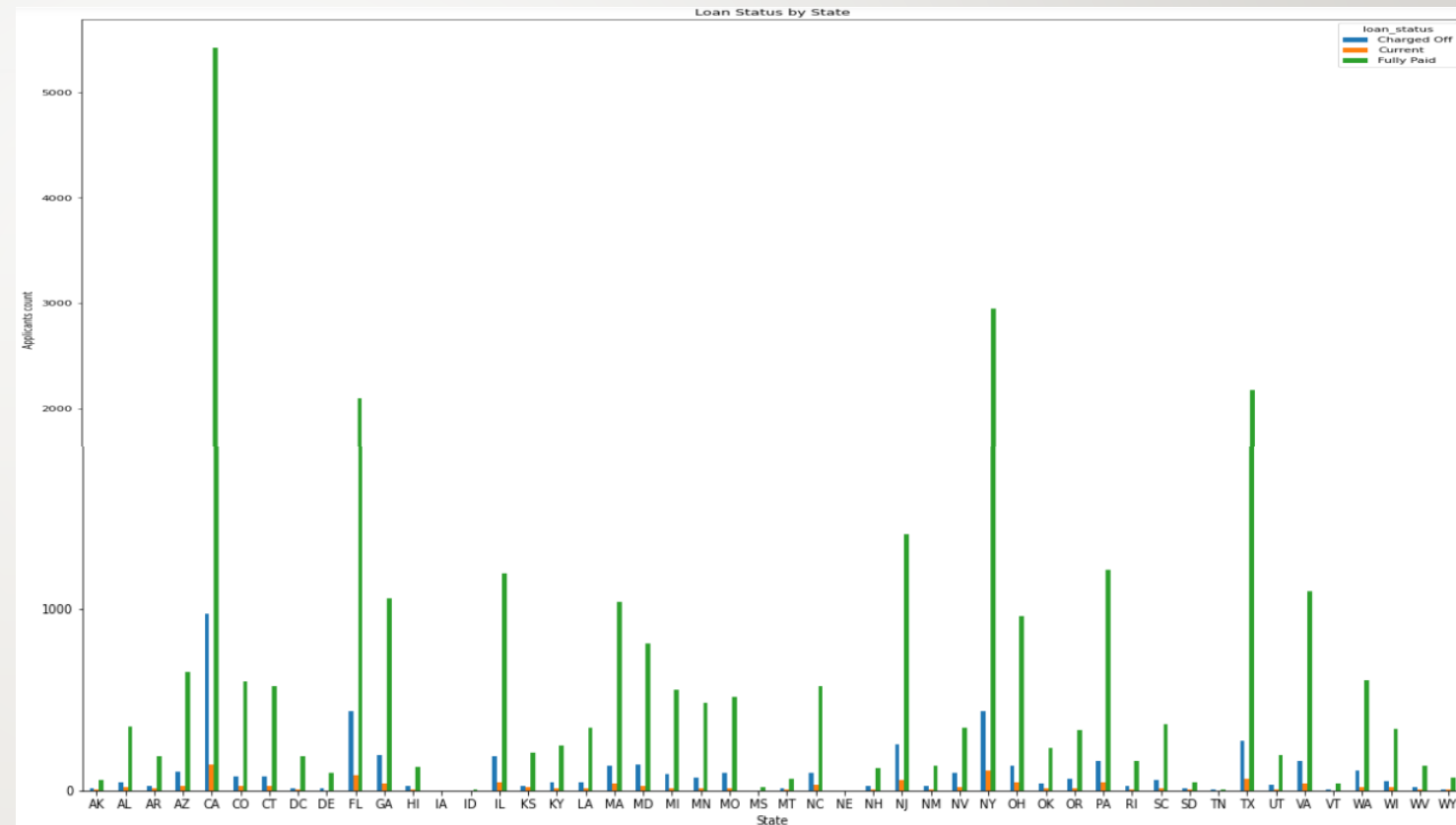
Segmented Univariate Analysis

- First plot shows that ,Larger number of loans under Fully paid category are not verified and maximum number of loans are verified under category for which installments are in progress
- Second plot shows that, loans which are in progress are part of 60 month tenure and maximum number of loans from 36 month tenure are in charged off status.



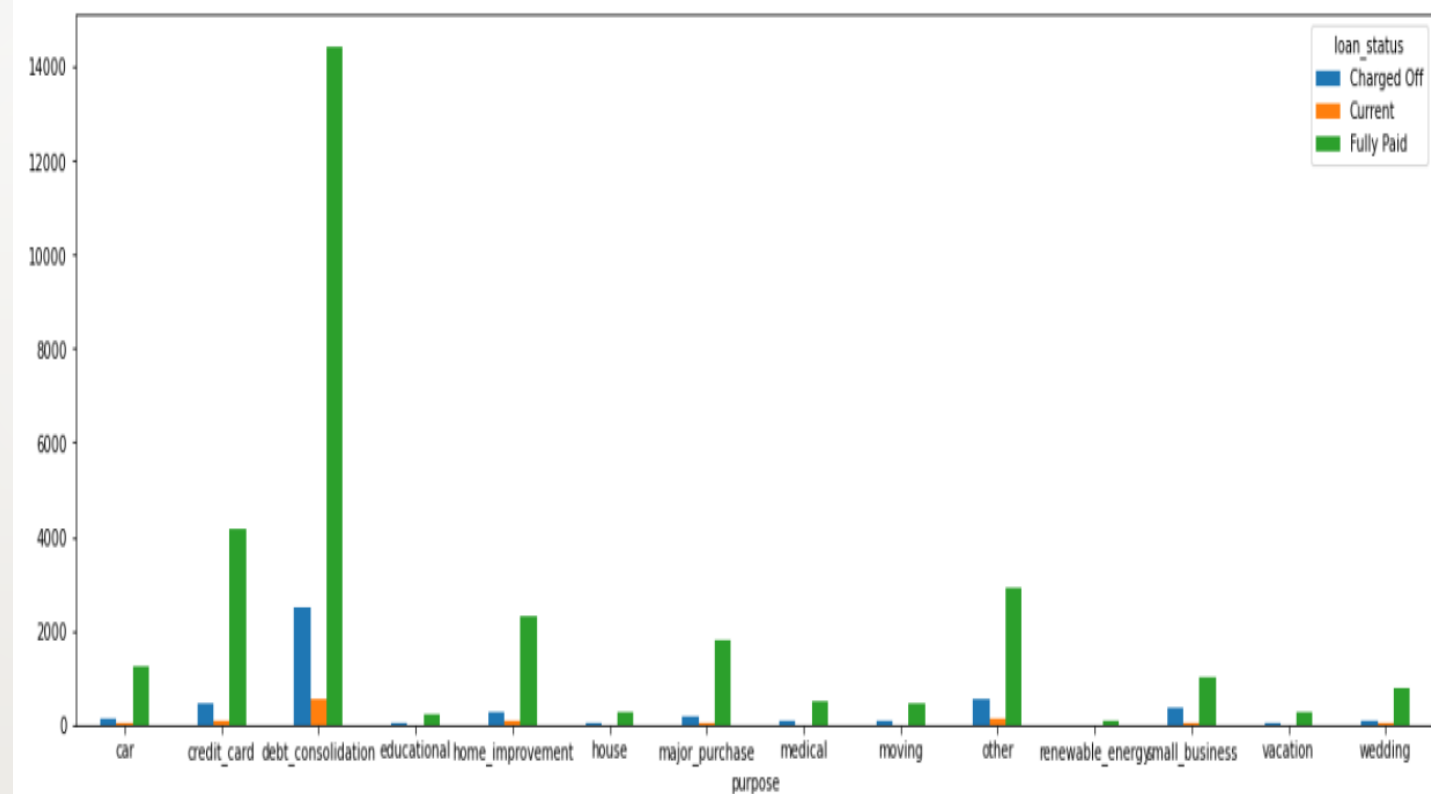
Segmented Univariate Analysis

- Maximum number of loans are fully paid in California
- Another concerning point with California is, large number of loans are in charged off status which indicates that there are high risk applicants in this state.



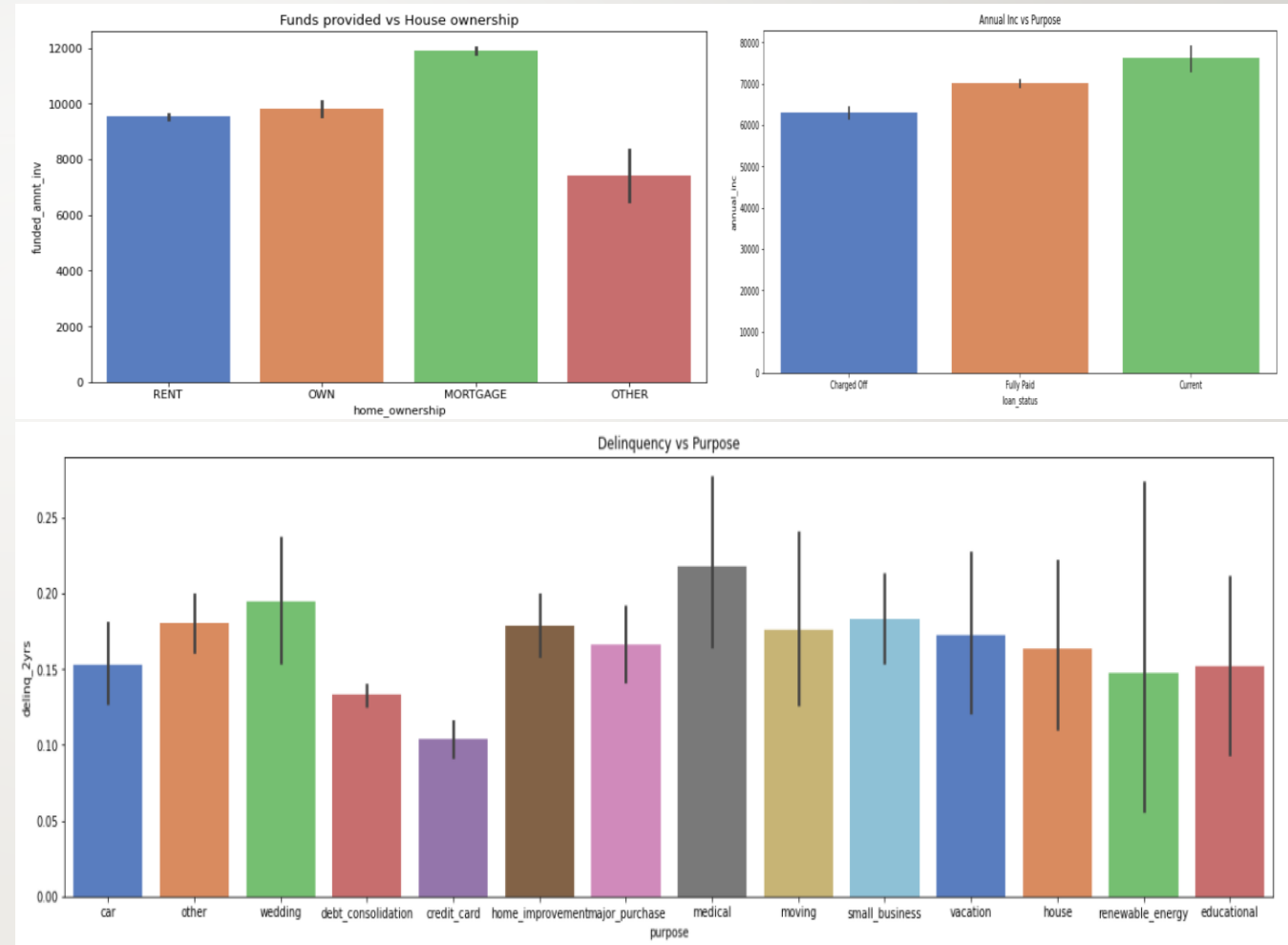
Segmented Univariate Analysis

- Highest defaulters are from "debt consolidation category"
- Credit card applicants have good payment record after "debt consolidation category"



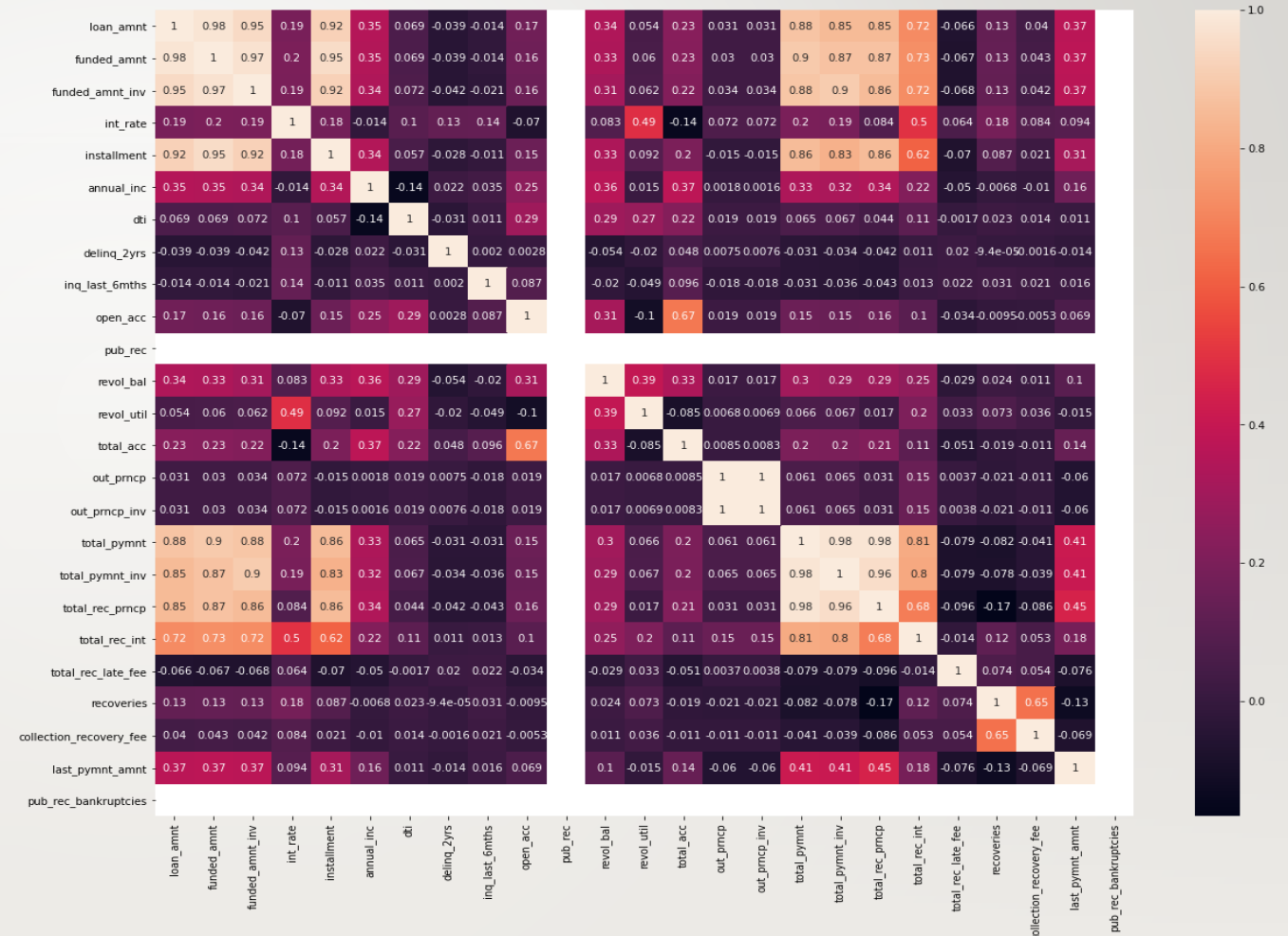
Bivariate Analysis

- Fig1. Applicants with mortgage have taken funds close to 12k & with own house have taken loan close to 10k
- Fig 2, Applicants with annual salary between 60k to 65k are defaulters i.e. charged off
- Fig 3, Highest delay in payment done by applicants who have taken loan under "medical" category & Least delay in payment are done by credit card applicants



Multivariate Analysis

- delinq_2yrs & total_rec_late_fee have least correlation with other variables
- total_rec_prncp has 97% percent correlation with funded_amnt_inv, which means high recovery of principal is resulting in loan fulfillment
- Installment has 86% of correlation with total_rec_prncp, higher installments are helping principal recovery



Recommendation:

- 14% of loans applicants who have has not paid the instalments are from "debt consolidation category". Need assess the risk for these applicants more rigorously
- Credit card applicants have good payment record after “debt consolidation category”, company should consider providing interest rate waiver to these applicants.
- Highest delay in payment done by applicants who have taken loan under "medical" category, company should reconsider providing more installments for these category