LENDING CLUB CASE STUDY

EXPLORATORY DATA ANALYSIS

Group Name:

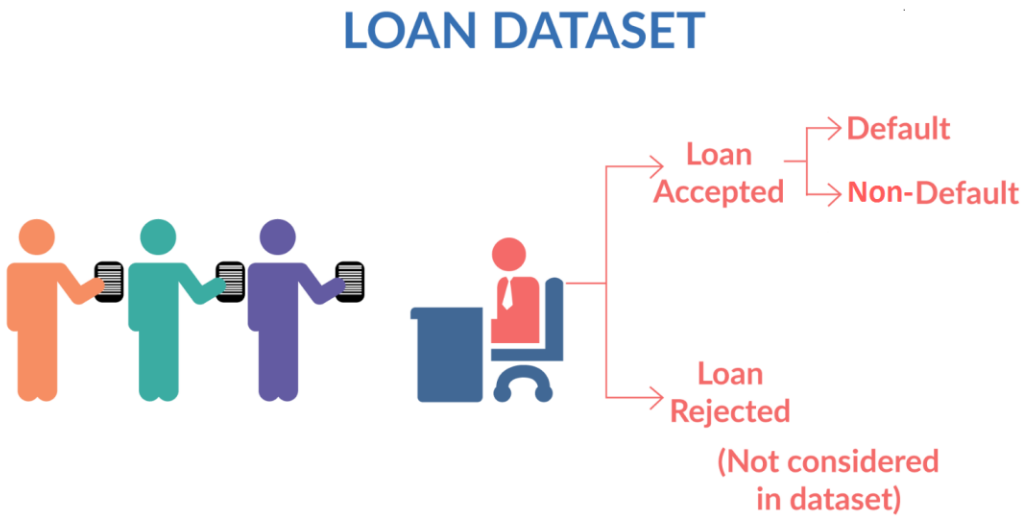
1. Vijay Kumar A.S
2. Harsha Vardhan Siginam
3. Ved Prakash
4. Sachindra Nath Mukherjee

Business Problem

A consumer finance company which specializes in lending various types of loans to urban customers, when receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

1. If the applicant is **likely to repay the loan.**
2. If the applicant is **not likely to repay the loan.**

Thus doing EDA from past records of the applicants, Company needs to identify possible defaulters to minimize the risk factor if an applicant won't repay the loan.



Business Objectives

There are three possible scenarios for loan borrowers:

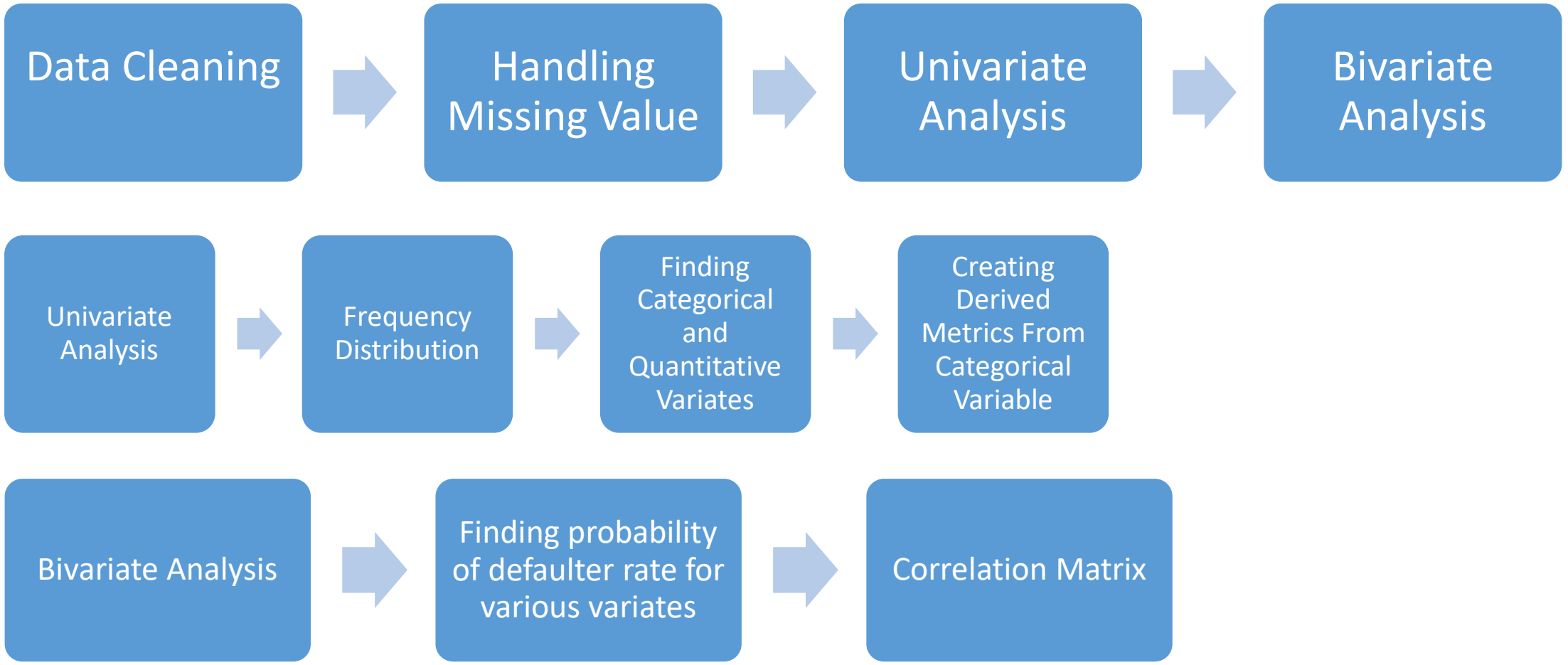
a) The borrower has fully paid the loan.

b) Paying the installment is in process.

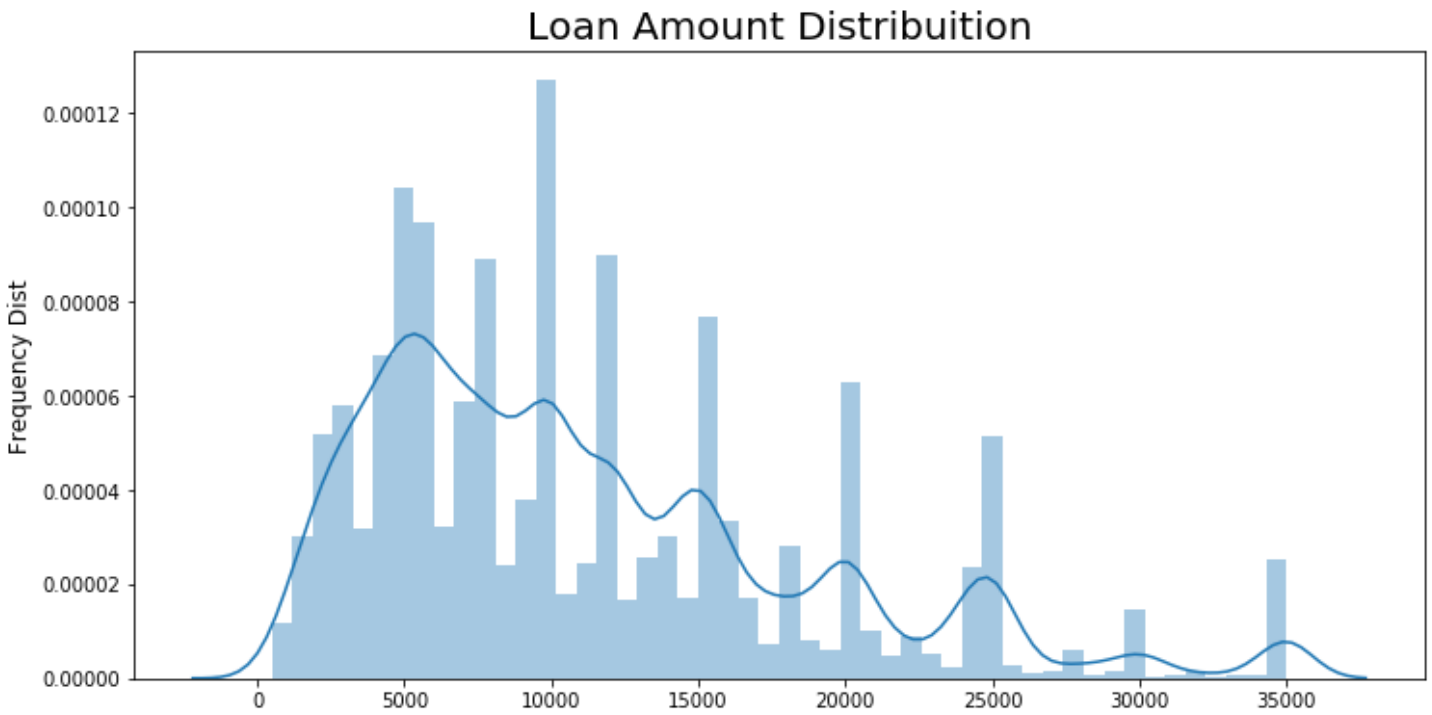
c) The borrower didn’t pay the loan for a long period of time.

Business needs to find out what are driving factors for an applicant which may cause an applicant to be defaulter/Charged-off, thus reducing the probability of financial loss by not giving loans to probable defaulters.

Problem Solving Methodology



Univariate Analysis - Frequency Distribution



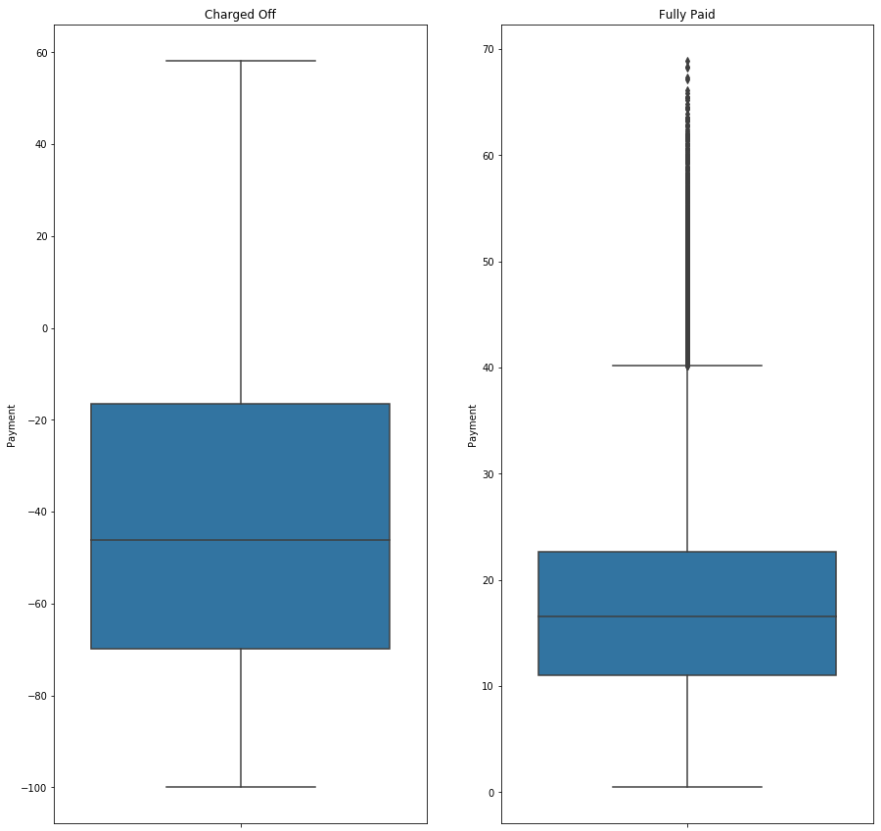
The Mean of Loan Amounts: 11219.44
The Median of Loan Amounts: 10000.00

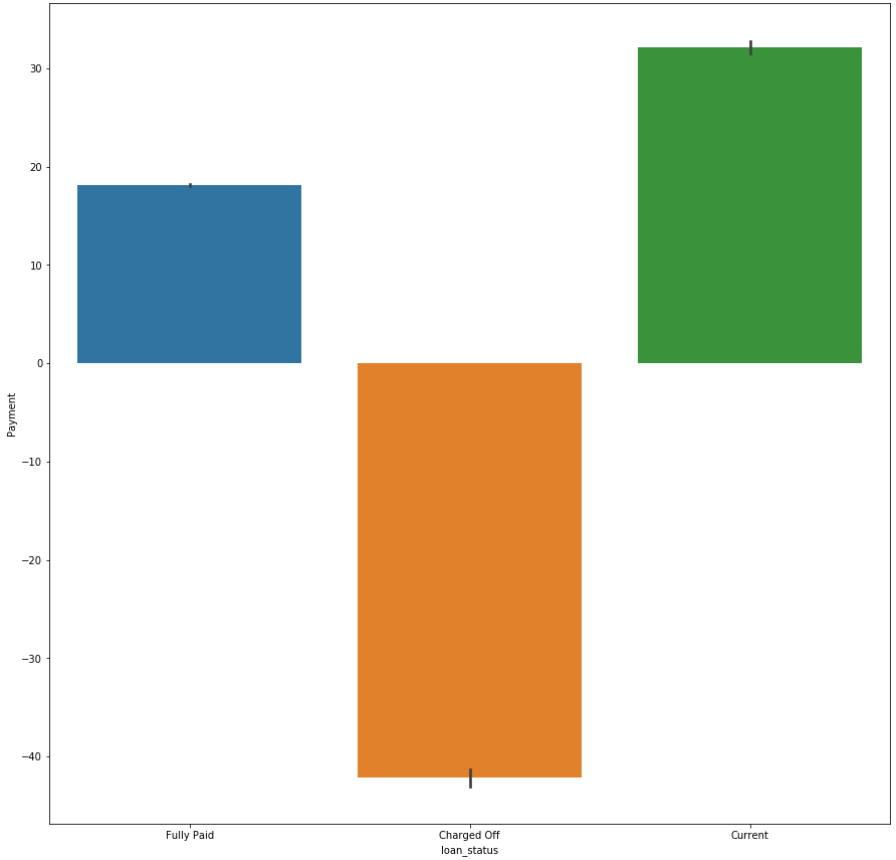
The frequency distribution of loan amount is more for lower amount of loans.

Univariate Analysis - Finding Categorical and Quantitative Variates

Categorical Variates		Quantitative Variates
term		id
int_rate		member_id
grade		loan_amnt
sub_grade		funded_amnt
emp_title		funded_amnt_inv
emp_length		installment
home_ownership		annual_inc
verification_status		dti
issue_d		delinq_2yrs
loan_status		inq_last_6mths
pymnt_plan		open_acc
url		pub_rec
desc		revol_bal
purpose		total_acc
title		out_prncp
zip_code		out_prncp_inv
addr_state		total_pymnt
earliest_cr_line		total_pymnt_inv
revol_util		total_rec_prncp
initial_list_status		total_rec_int
last_pymnt_d		total_rec_late_fee
last_credit_pull_d		recoveries
application_type		collection_recovery_fee
		last_pymnt_amnt
		collections_12_mths_ex_med
		policy_code
		acc_now_delinq
		chargeoff_within_12_mths
		delinq_amnt
		pub_rec_bankruptcies
		tax_liens

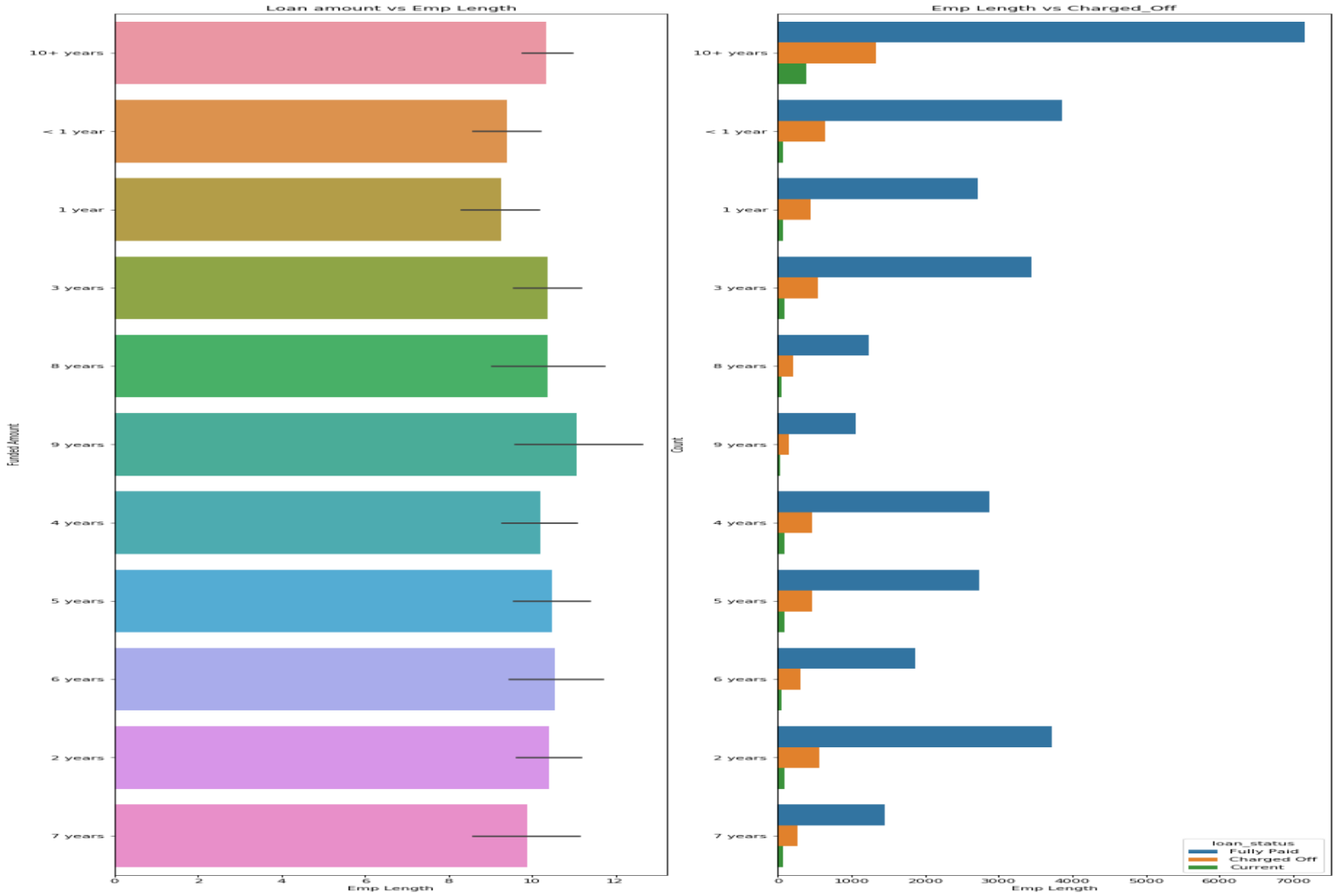
Univariate Analysis - Derived Metrics



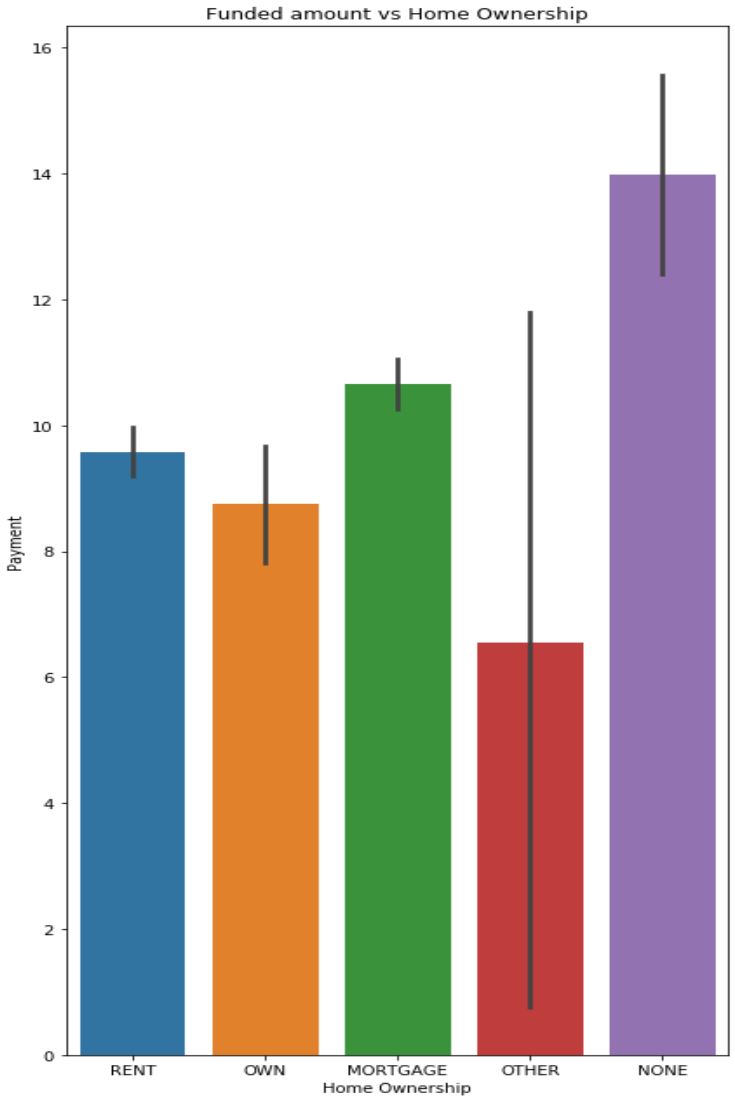


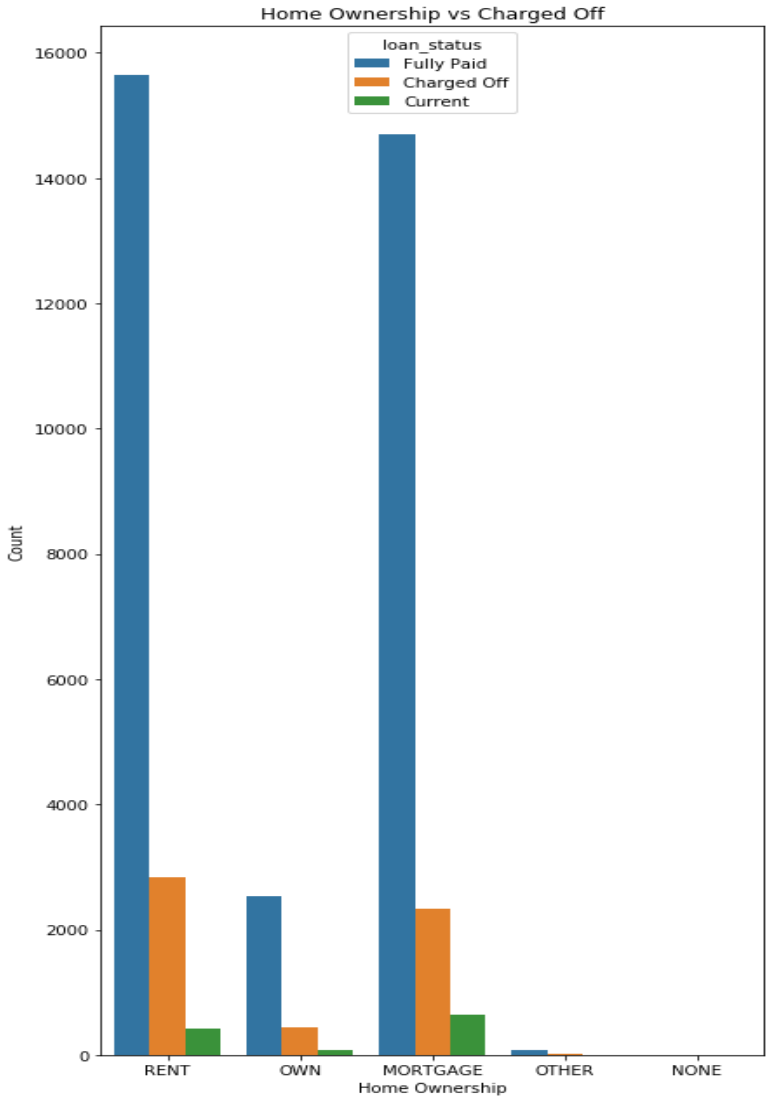
We have Created Derived Metric “Percentage of Payment Made” to represent the categorical variable Loan Status in quantitative variate. Above two plots clearly show us when percentage of payment is negative the loan status is Charged Off.

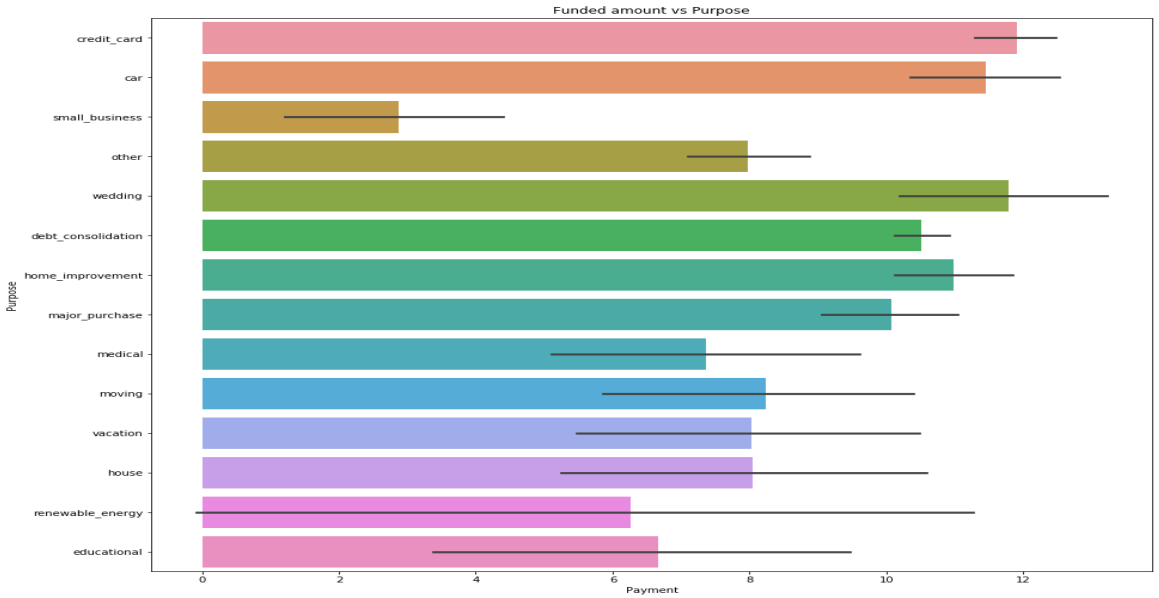
Bivariate Analysis - Relationship between Loan and employee length

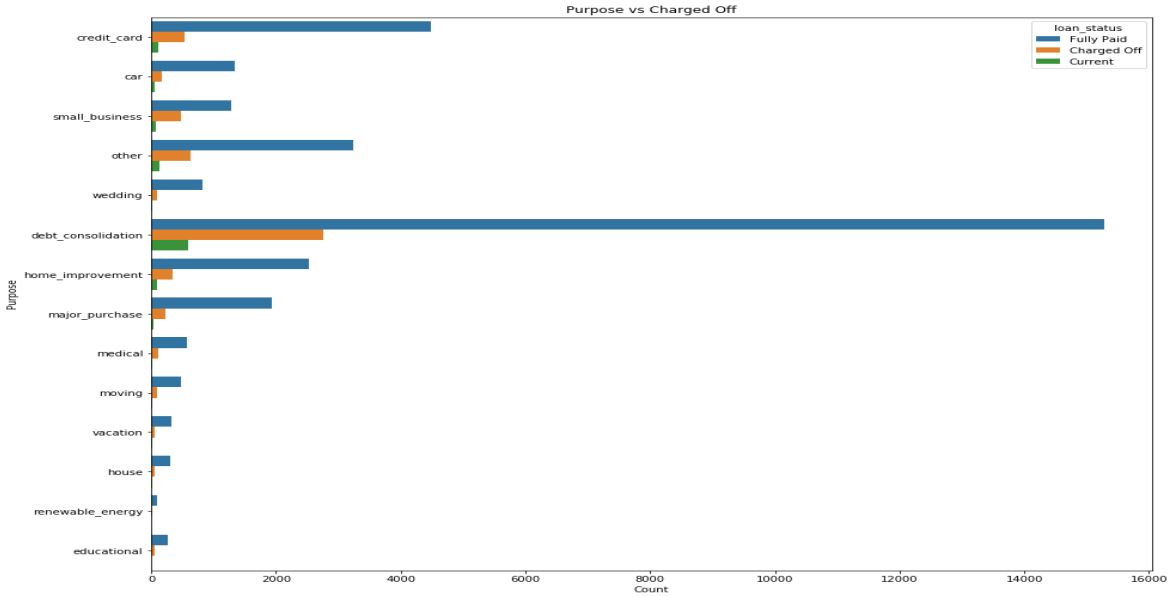


Bivariate Analysis - Relationship between Loan and Home Owner









Bivariate Analysis- Relationship between Loan and Purpose

Bivariate Analysis

emp_length	median	mad	default_rate
10+ years	15.23	16.872745	15
7 years	15.63	16.94271	14.8
1 year	15.32	16.15918	14.1
5 years	15.255	15.78903	14
< 1 year	14.69	16.024387	13.9
6 years	15.55	16.600938	13.8
8 years	14.31	15.1952	13.7
3 years	15.52	15.540073	13.6
4 years	14.81	16.098257	13.4
2 years	15.54	15.32938	12.9
9 years	14.72	15.006902	12.6

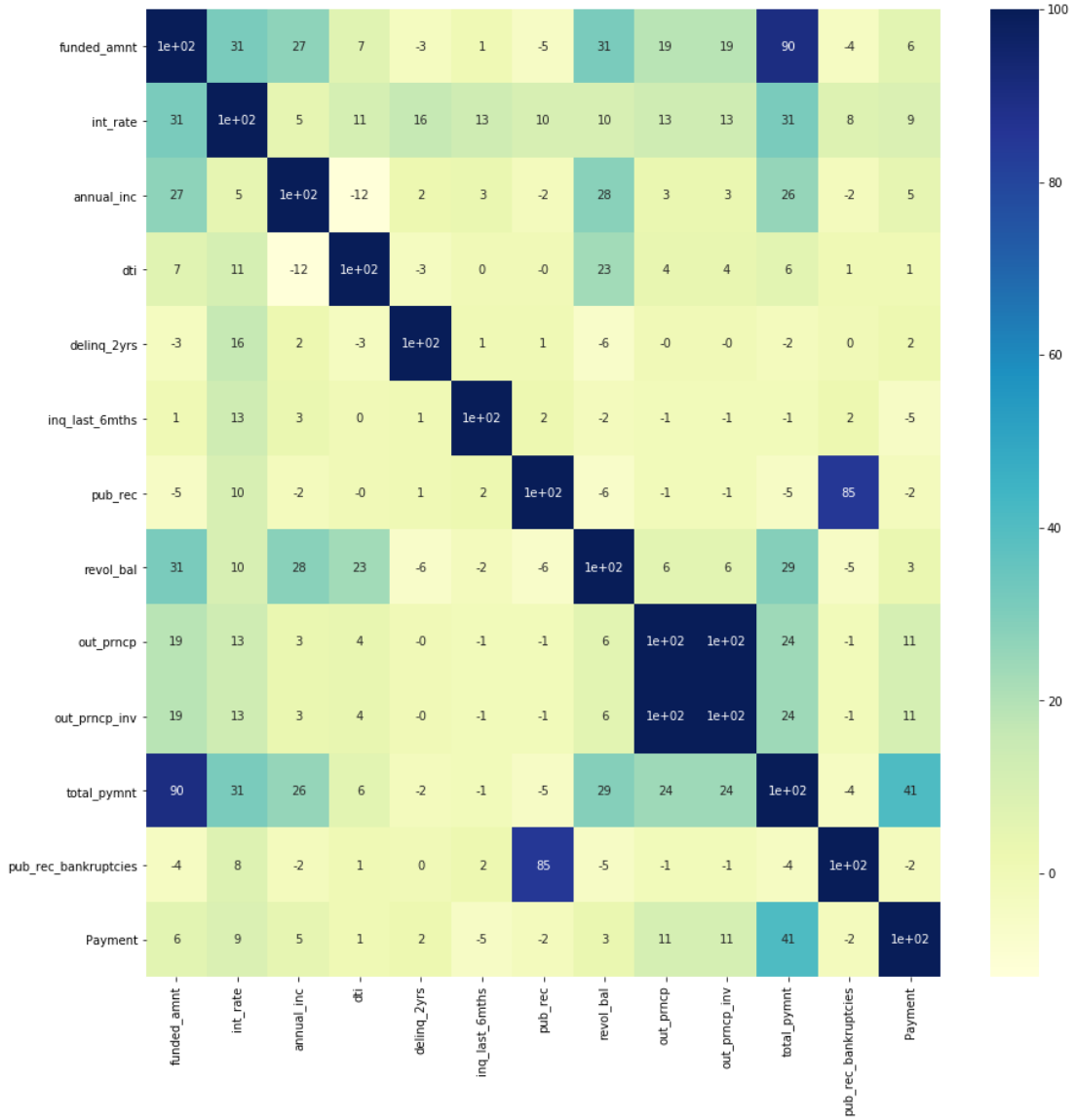
purpose	median	mad	default_rate
small_business	14.075	26.782303	26
renewable_energy	12.08	18.393888	18.4
educational	15.04	16.716205	17.2
other	14.38	17.480899	15.9
moving	13.8	17.018864	15.8
house	12.64	17.129229	15.5
medical	12.64	18.263116	15.3
debt_consolidation	16.01	16.872113	14.8
vacation	12.62	14.78048	13.9
home_improvement	13.86	14.79032	11.7
credit_card	15.43	12.735122	10.6
car	12.81	12.248226	10.3
major_purchase	12.39	12.939577	10.2
wedding	15.54	13.739243	10.1

home_ownership	median	mad	default_rate
MORTGAGE	14.32	15.52787	13.2
OWN	13.755	16.22194	14.5
RENT	15.89	16.90852	15
OTHER	15.1	17.48207	18.4
NONE	13.97	1.055556	NaT

From the above analysis, we can see that

- employees more than 7/10 years of employee length have higher default Rate.
- Loan borrowed for small businesses have very high default probability as the chance of failure is high.
- People living in rented house have more probability for being defaulter.

Bivariate Analysis- Correlation Matrix



Our Recommendation from EDA

Strong Variates:

annual_inc
inq_last_6mths
total_pymnt
total_pymnt_inv
emp_length
home_ownership
loan_status