

# LENDING CLUB EDA CASE STUDY

Risk analysis to find appropriate applicant to balance risk of loosing business or avoiding financial loss while lending to customers

Date: 9 Mar 2022

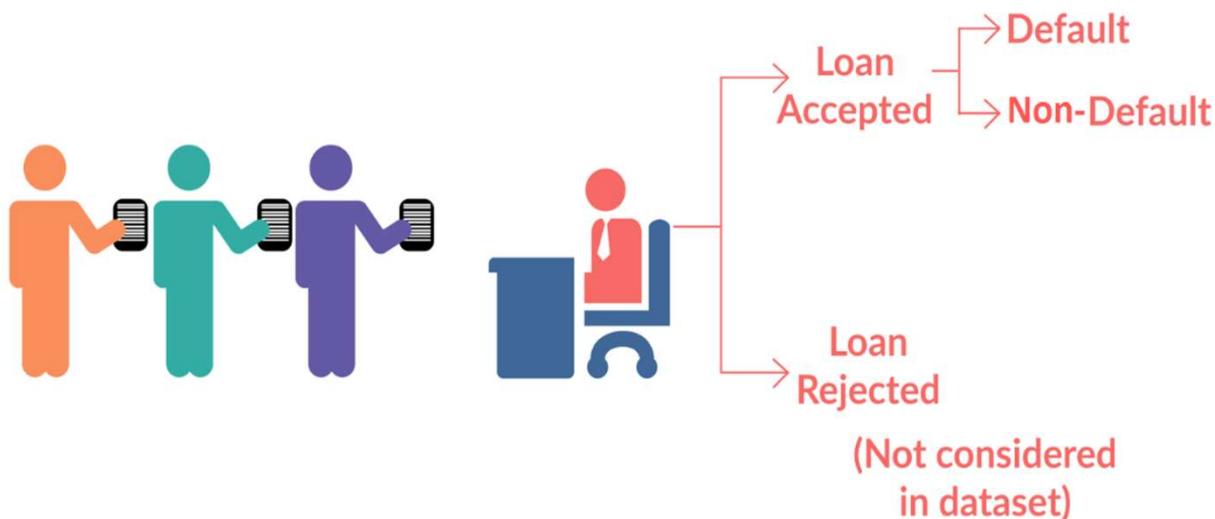
By  
Ravish Upadhyay  
Rajendra Shelar

## BUSINESS

- Finance company specializes in lending various types of loans to urban customers.
- The company decides for loan approval based on the applicant's profile.
  - Two types of risks are associated with the bank's decision:
    - Likely to repay – potential business
    - Unlikely to repay - financial loss
- The company has the market data of the successful borrowers

# LOAN APPROVALS PROCESS

## LOAN DATASET



### Loan accepted

- Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
- Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
- Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

## PROBLEM STATEMENT

- Identify an applicant who can potentially default the loan and help to finalize their decision to approve, partially approve or reject the loan.
- Understand the various driving factors behind the loan which are strong indicators of default. This can be used by the lender for it's portfolio and risk assessment.
- Identify the applicants who has the strong possibility of defaulting the loan and to identify the applicants who can repay the loan

# APPROACH - EXPLORATORY DATA ANALYSIS (EDA)

analyzing data sets to summarize their main characteristics with visual methods

- Data Cleaning / Standardization
  - Cleaning missing values / Imputation
  - Removing Duplicates / Null values etc

## Data Analysis

- Data Modelling / Classification
- Statistical Analysis/Visualization

## Data Assessment

- Mapping data with Problem statement
- Customer patterns



## Recommendation

- Applicants with risk of not repaying the loan

# DATA ASSESSMENT

# Data Assessment

Use of data dictionary to understand the terms, map the data with problem statement

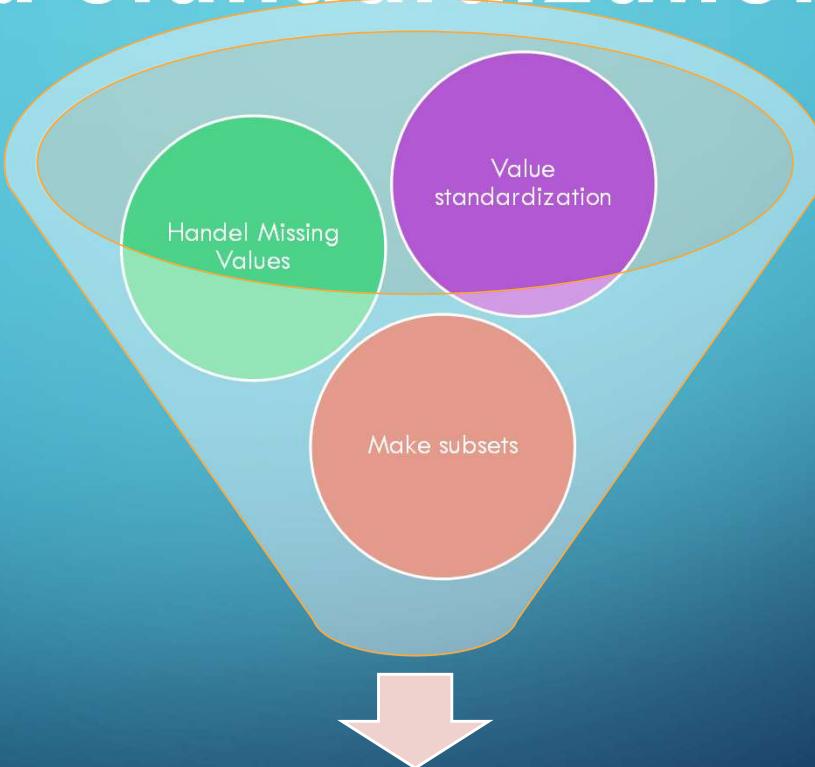
Summary of the key numbers to understand the maximum , minimum loan borrowed vs annual income

Studied the columns to be used for the various analysis to find the potential default applicant

LoanStatNew	Description
acc_now_delinq	The number of accounts on which the borrower is now delinquent.
acc_open_past_24mths	Number of trades opened in past 24 months.
addr_state	The state provided by the borrower in the loan application
all_util	Balance to credit limit on all trades
annual_inc	The self-reported annual income provided by the borrower during registration.
annual_inc_joint	The combined self-reported annual income provided by the co-borrowers during registration
application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers
avg_cur_bal	Average current balance of all accounts
bc_open_to_buy	Total open to buy on revolving bankcards.

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv
count	39,717.00	39,717.00	39,717.00	39,717.00	39,717.00
mean	683,131.91	850,463.56	11,219.44	10,947.71	10,397.45
std	210,694.13	265,678.31	7,456.67	7,187.24	7,128.45
min	54,734.00	70,699.00	500.00	500.00	0.00
25%	516,221.00	666,780.00	5,500.00	5,400.00	5,000.00
50%	665,665.00	850,812.00	10,000.00	9,600.00	8,975.00
75%	837,755.00	1,047,339.00	15,000.00	15,000.00	14,400.00
max	1,077,501.00	1,314,167.00	35,000.00	35,000.00	35,000.00

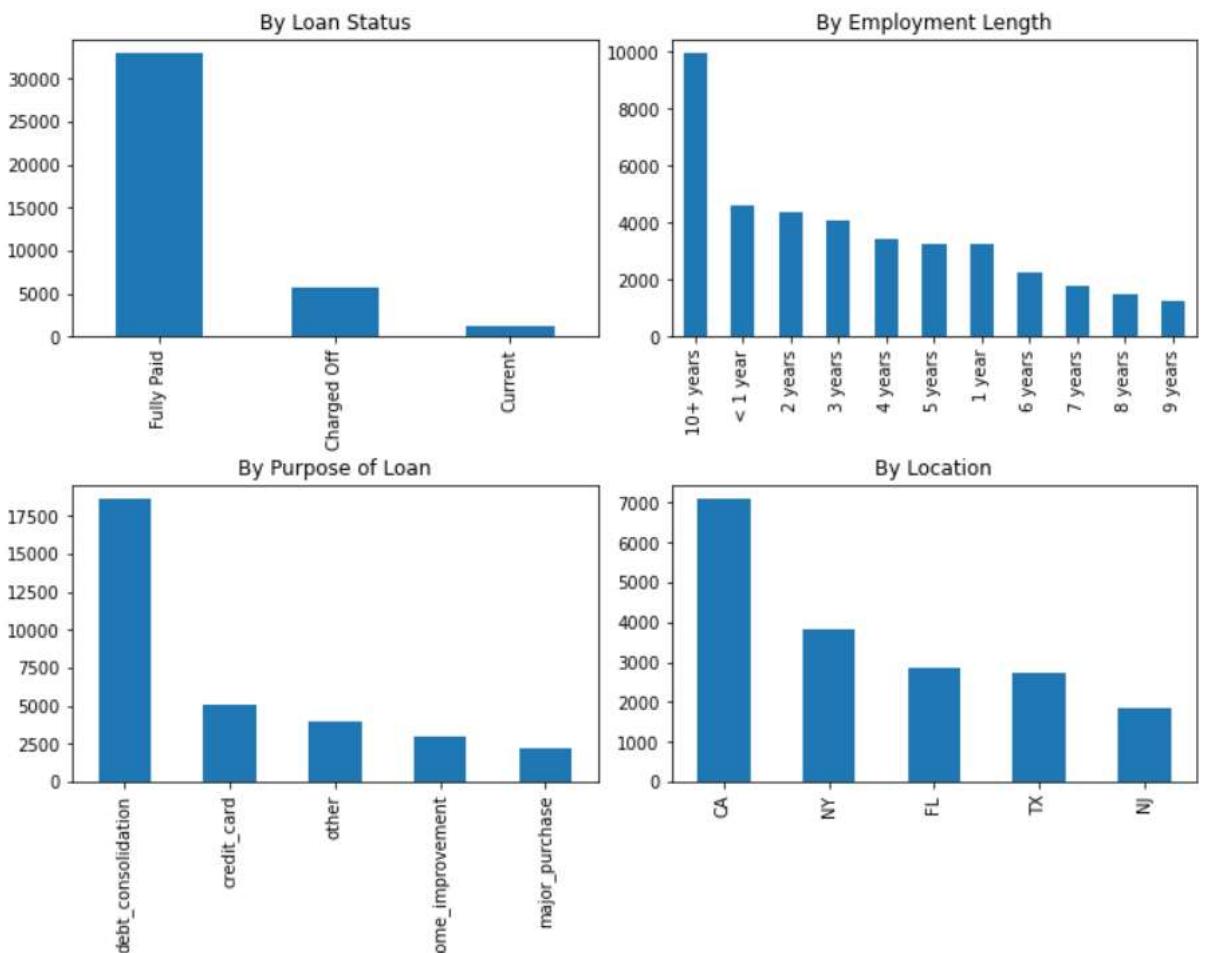
# Data Cleaning and Standardization



Data For Analysis

# DATA ANALYSIS

# Data at Glance

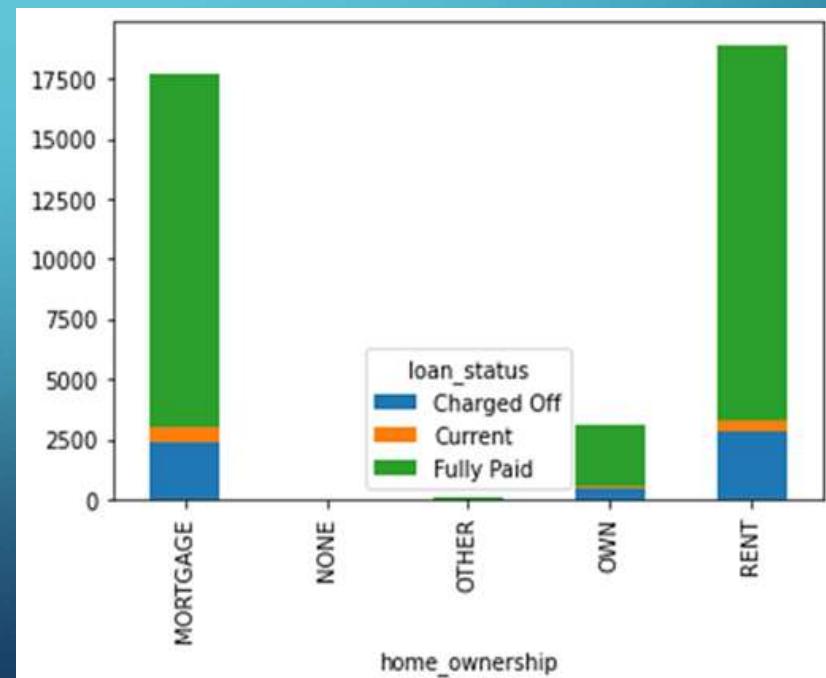
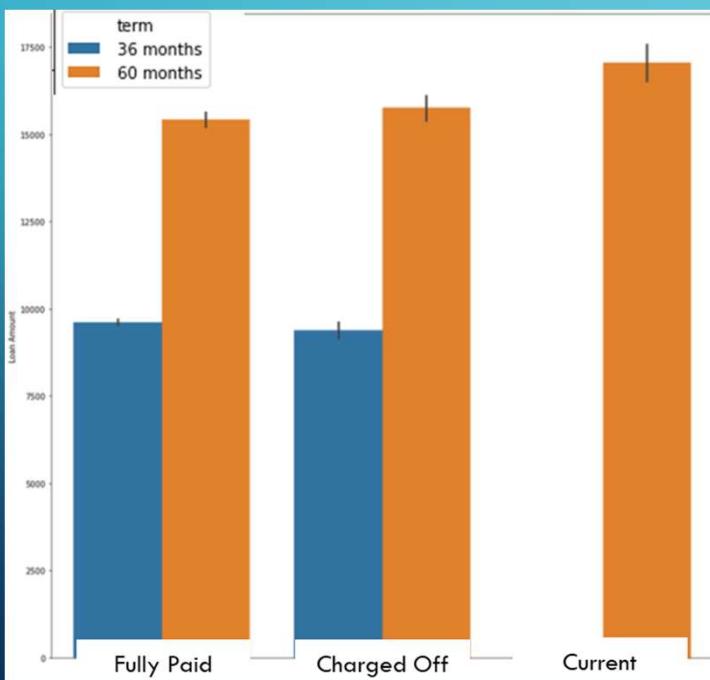


- Fully paid applicants are more
- Higher experience applicants take more loans
- Repaying debt is the reason for most of the applicants
- Maximum applicants are from California

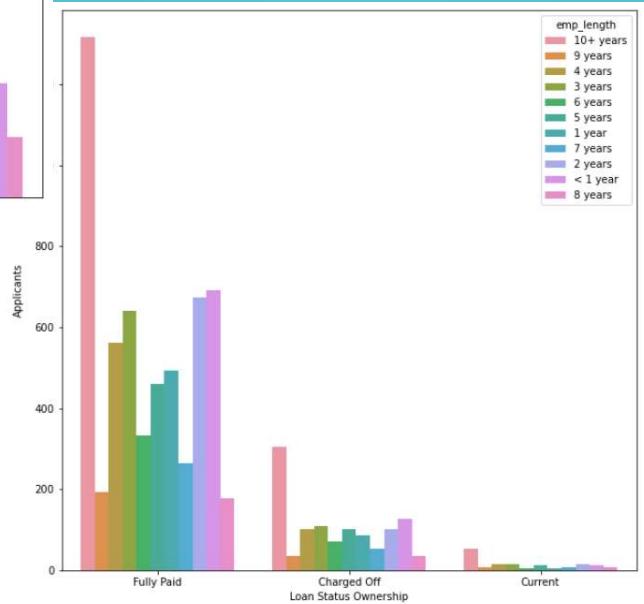
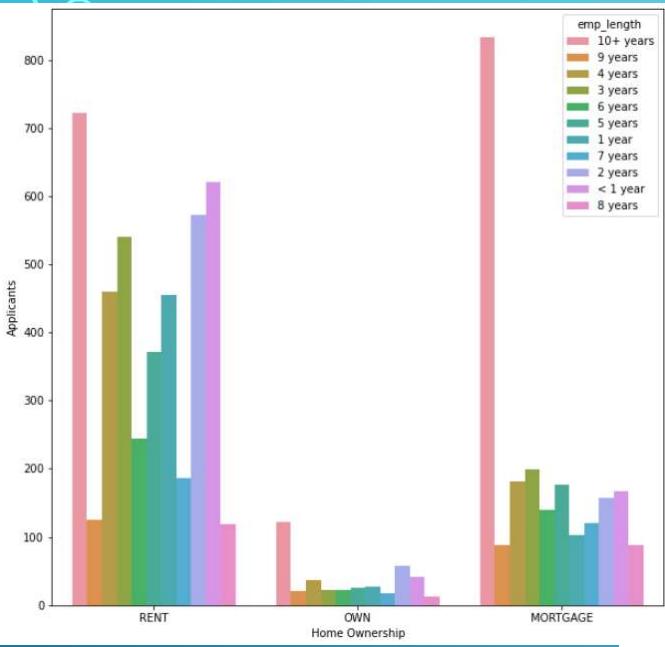
# Data at Glance

Long term loans get charged off more than short term. In general, Long term loans are more popular.

Applicant who stays in rented or mortgage houses are charged off more than applicants with own houses.



# Data at Glance



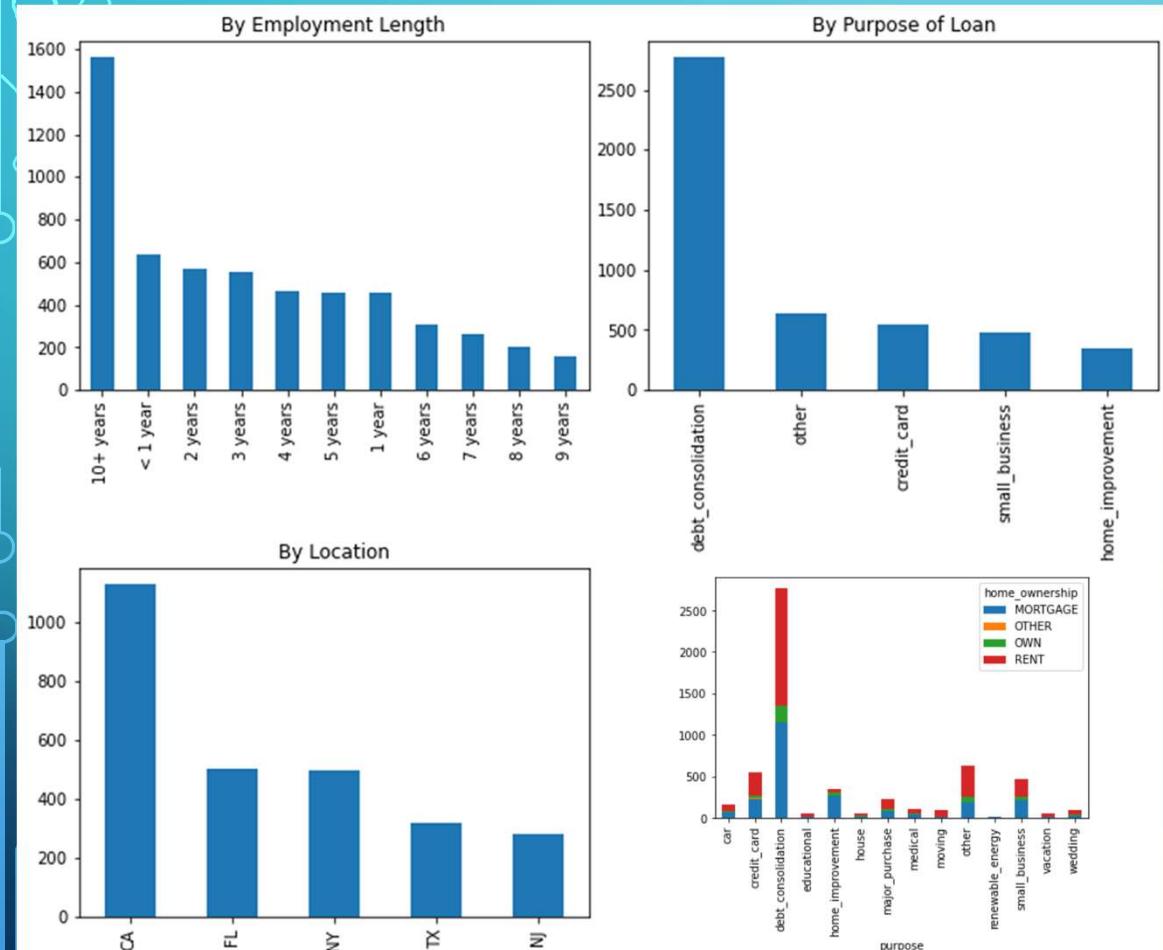
Higher experience applicants are more across various loan status

Applicants are either stay in rented or mortgage house

## ANALYZE CHARGED OFF APPLICANTS

Review the applicant profile to know the reason for the charged off

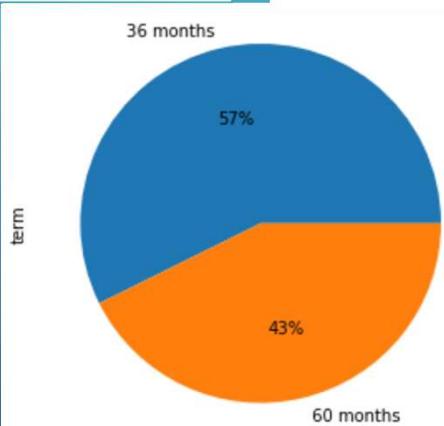
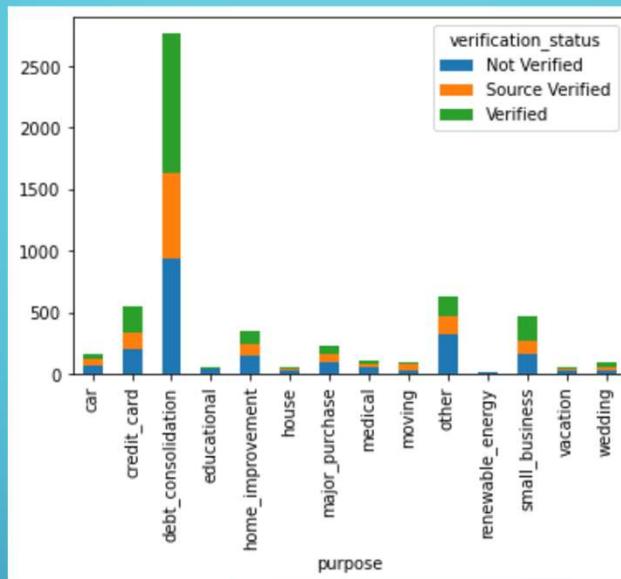
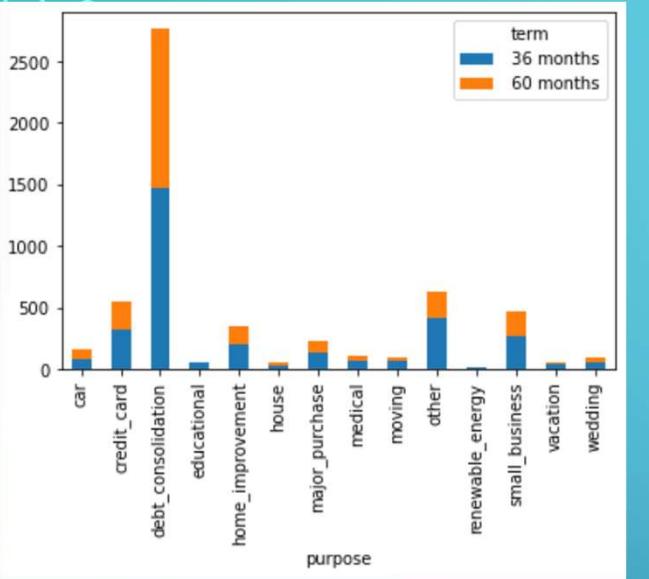
# Charged-Off Loans



- Taken to clear the debt
- Stays either in rented house or mortgage house
- Higher experience
- From California

So we need to analyze these dimensions in the applicants who have fully paid the loan to find their risk level.

# Charged-Off Loans



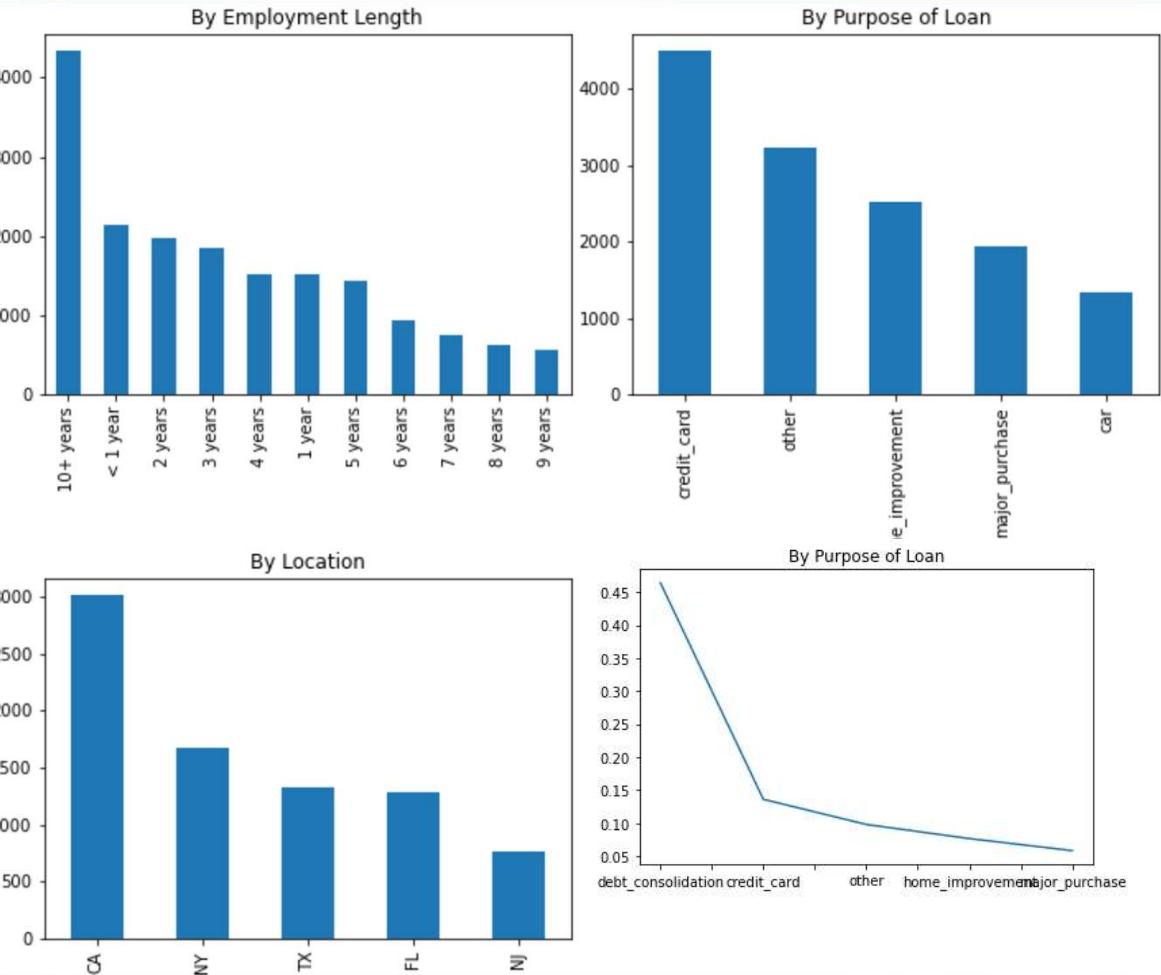
- Term distribution across defaulters is closed to 36 or 60 months hence will not be useful in the assessment.
- Loan term and verification status are evenly spread for debt consolidation

hence these attributes will be ruled out for identifying Risky profile

## ANALYZE FULLY PAID APPLICANTS

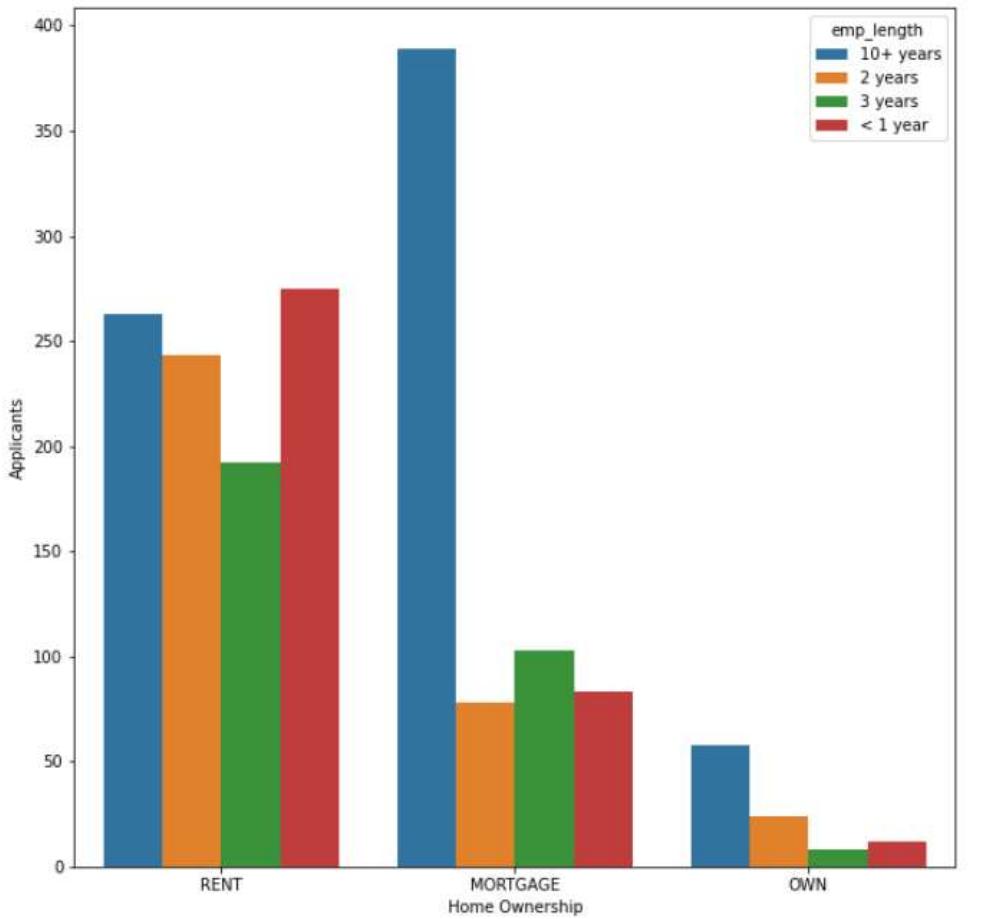
Review the applicant profile to know the reason for the charged off

# General Statistics – univariate analysis



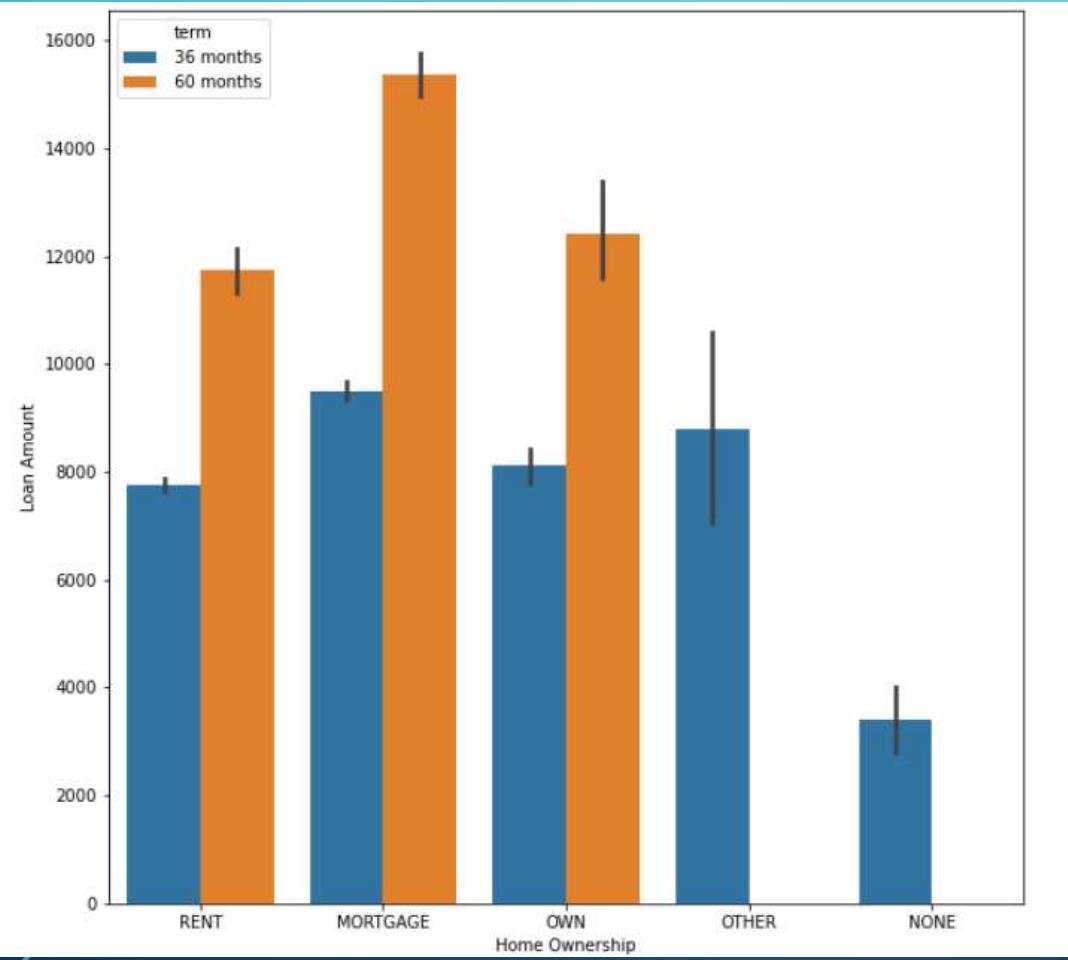
- There are 46% of the applicants have paid the loan fully taken to clear the debt so these are risky applicants, We shall remove it from the dataset.
- Next category is loan taken to credit card payments
- Higher experience applicant are more
- Applicants from California

# General Statistics – bivariate analysis



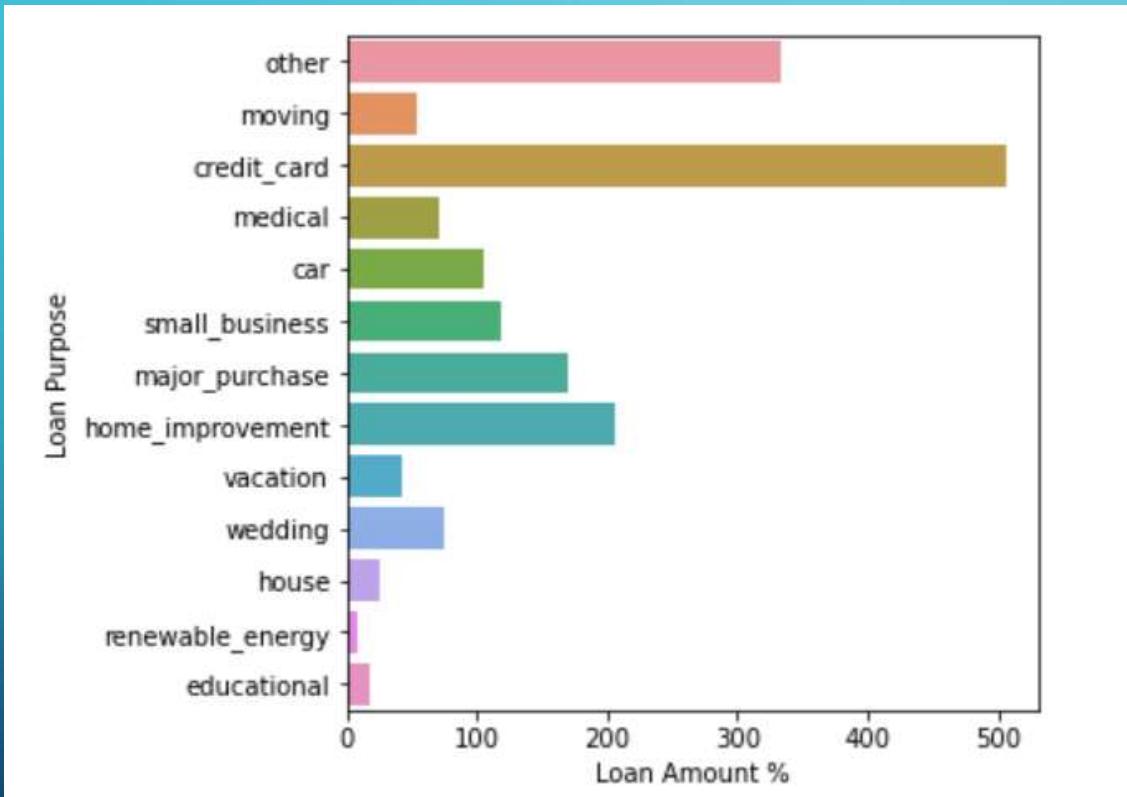
- Higher experience applicants are the most amongst other borrowers who stay in Mortgage house.
- There were not many application having own house

# General Statistics – bivariate analysis



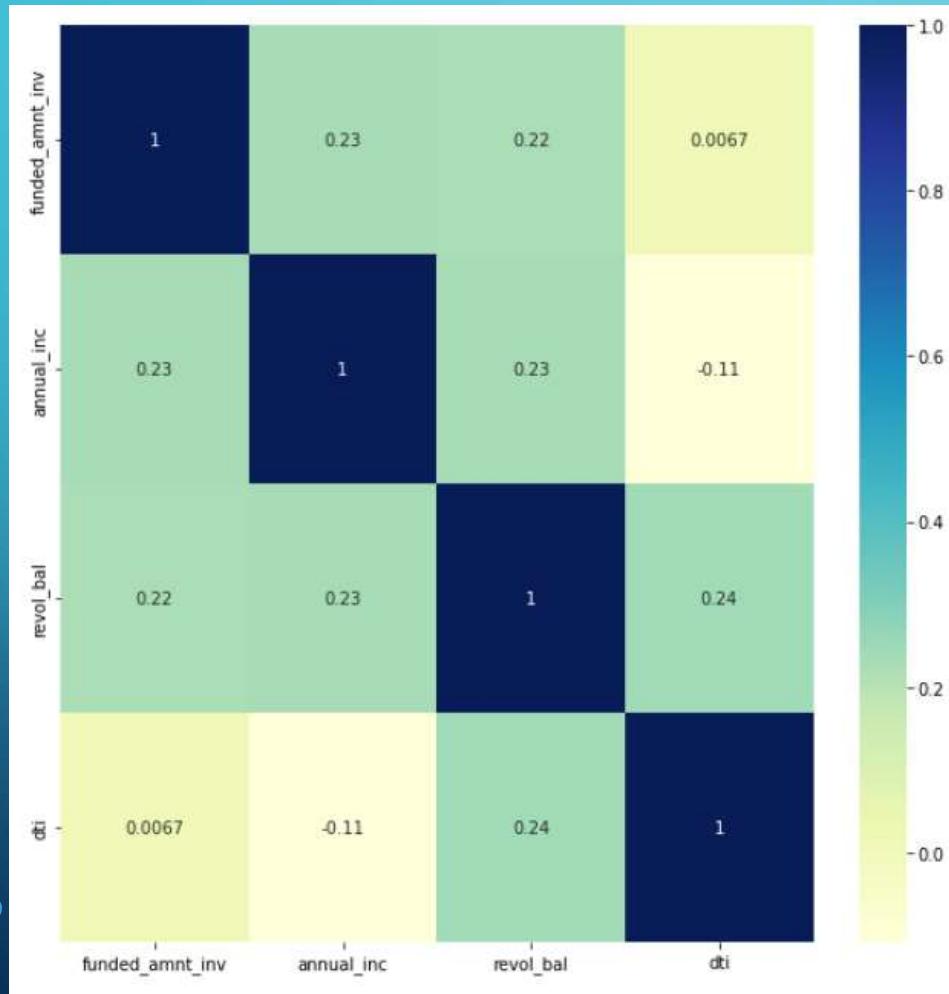
- Applicants  
prefers Long  
term loans

# General Statistics – correlation heatmap



- Annual income is high for the applicant whose loan is fully paid and purpose for the loan was credit card (was the

# General Statistics – correlation heatmap

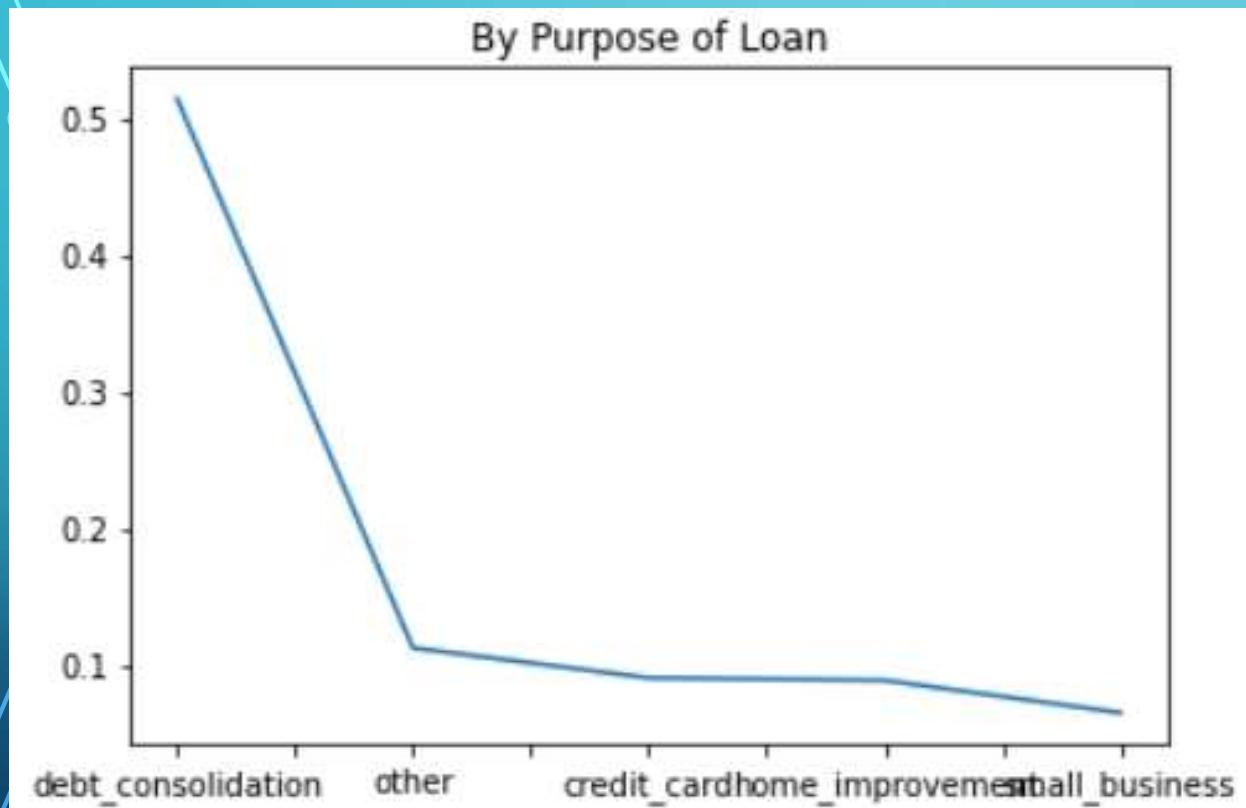


- Annual income and loan amount have positive correlation meaning higher the annual income, higher the loan amount funded

## ANALYZE CURRENT APPLICANTS

Review the applicant profile to know the reason for the loan

# General Statistics – correlation heatmap



- 50% of the applicant with loan status as “Current” borrowed money to pay the debt.

# RECOMMENDATION

Applicants to filter out or flag as high / low level risk candidates

# RISK PROFILE

SN	Risk Profile	Condition	Recommendation
1	High Risk	Applicant with loan status as “Charged Off” or Applicant with loan status as Fully Paid or Current but purpose is Debt Consolidation	Default these applicant and don’t accept the loan application
2	Low Risk	Applicant from CA	Review the application and approve less amount and short-term tenure
3	Low Risk	Applicant with Higher number of years of experience	Review the application and approve less amount and short-term tenure
3	Low Risk	<ul style="list-style-type: none"> <li>Applicants whose annual income is above 82k (i.e., 75%) is outlier</li> <li>Applicants whose annual income is below 4k (i.e., 25%) is outlier</li> </ul>	Review the application and approve less amount and short-term tenure

**THANK YOU**