
LendingClub Case Study

Rohit Aggarwal
Arunkumar Subramanyam

Introduction

Consumer finance company specializing in lending various types of loans to urban customers. Types of risks associated with lending:

1. If applicant is likely to repay the loan, then not approving is a loss of business
2. If applicant is not likely to repay the loan, then approving may lead to a financial loss

Objective: Company wants to understand the driving factors behind loan default. If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.

Data Understanding

Description	Loans issued by the company to its customers		
Rows	39717	Columns	111
Each Row is	Details associated with a particular loan account		
Sampling Method	All the loans issued through the time period 2007 and 2011		

Data Type	Count
Integer	13
Float	74
Others	24

```
# Load the loan dataset
dataset = pd.read_csv("datasets/loan.csv")

# Print shape and basic information about the dataset
print(dataset.shape)
print()
print(dataset.info())
```

```
(39717, 111)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 39717 entries, 0 to 39716
Columns: 111 entries, id to total_il_high_credit_limit
dtypes: float64(74), int64(13), object(24)
memory usage: 33.6+ MB
```

Data Cleaning

- Remove columns where more than 90% of the values are not set
- Remove columns where only one logical value existed within the entire dataset
- Remove columns which don't hold any significance to the analysis

```
# Removing columns where data itself is not significant for analysis
clean_dataset = clean_dataset.drop(["id", "member_id", "title", "emp_title", "desc", "url", "zip_code", "funded_amnt", "funded_amnt_inv",
                                     "sub_grade", "out_prncp_inv", "total_pymnt_inv", "total_rec_prncp", "collection_recovery_fee",
                                     "earliest_cr_line", "total_rec_int", "total_rec_late_fee", "last_credit_pull_d", "inq_last_6mths"],
                                   axis = 1)

# Print top 5 rows of the new dataset
clean_dataset.head()
```

	loan_amnt	term	int_rate	installment	grade	emp_length	home_ownership	annual_inc	verification_status	issue_d	...	pub_rec	revol_bal	revol_util	tot
0	5000	36 months	10.65%	162.87	B	10+ years	RENT	24000.0	Verified	Dec-11	...	0	13648	83.70%	
1	2500	60 months	15.27%	59.83	C	< 1 year	RENT	30000.0	Source Verified	Dec-11	...	0	1687	9.40%	
2	2400	36 months	15.96%	84.33	C	10+ years	RENT	12252.0	Not Verified	Dec-11	...	0	2956	98.50%	
3	10000	36 months	13.49%	339.31	C	10+ years	RENT	49200.0	Source Verified	Dec-11	...	0	5598	21%	
4	3000	60 months	12.69%	67.79	B	1 year	RENT	80000.0	Source Verified	Dec-11	...	0	27783	53.90%	

5 rows x 27 columns

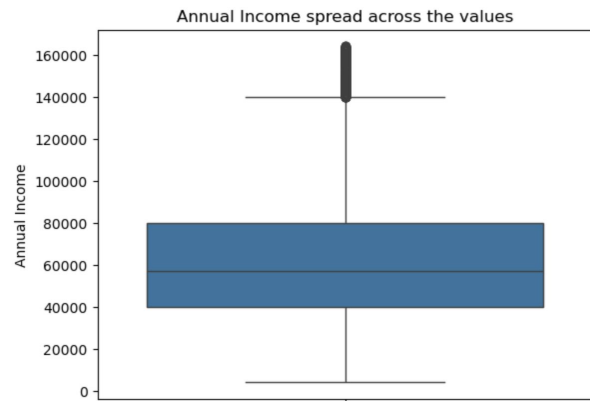
Data Fixing

- Stripping extra text like “months” and “%” from values
- Converting values to appropriate data type and relevant precision
- Imputing values based on data in other columns, or using median/mode
- Separating month and year from date type columns
- Removing outliers from data, found after quartile analysis

```
# Removing all entries where the loan is ongoing as not important for analysis
fixed_dataset = fixed_dataset[fixed_dataset["loan_status"] != "Current"]
fixed_dataset.head()
```

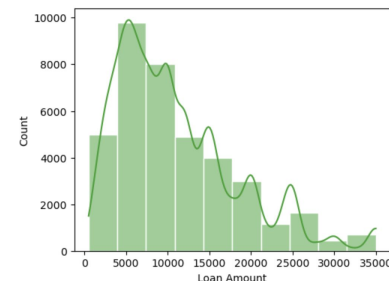
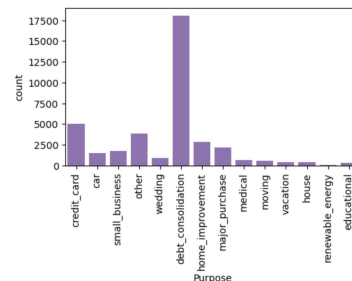
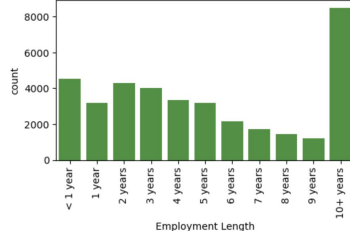
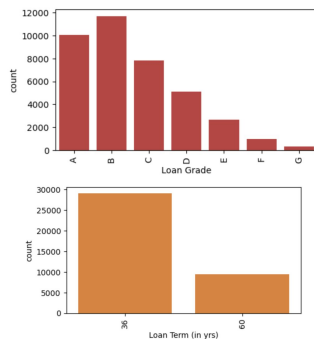
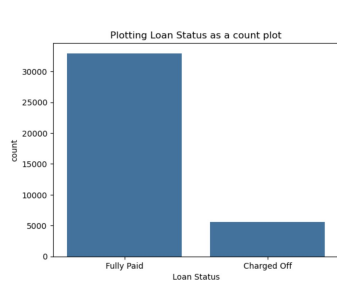
	loan_amnt	term	int_rate	installment	grade	emp_length	home_ownership	annual_inc	verification_status	loan_status	...	total_acc	out_prncp	total_pym
0	5000	36	10.65	162.87	B	10+ years	RENT	24000.0	Verified	Fully Paid	...	9	0.0	5863
1	2500	60	15.27	59.83	C	< 1 year	RENT	30000.0	Source Verified	Charged Off	...	4	0.0	1008
2	2400	36	15.96	84.33	C	10+ years	RENT	12252.0	Not Verified	Fully Paid	...	10	0.0	3005
3	10000	36	13.49	339.31	C	10+ years	RENT	49200.0	Source Verified	Fully Paid	...	37	0.0	12231
5	5000	36	7.90	156.46	A	3 years	RENT	36000.0	Source Verified	Fully Paid	...	12	0.0	5632

5 rows × 29 columns



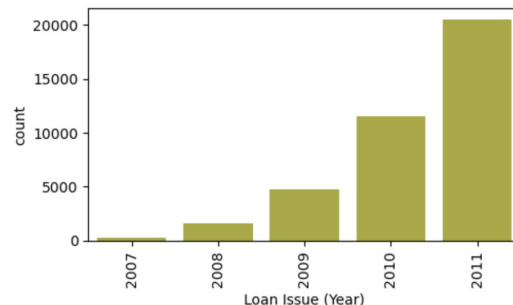
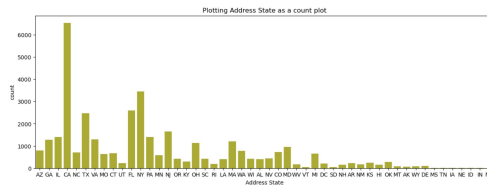
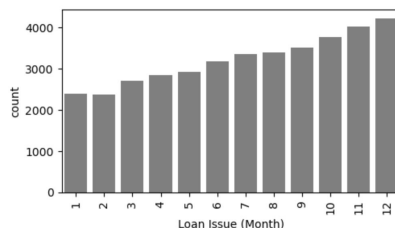
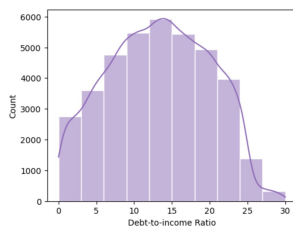
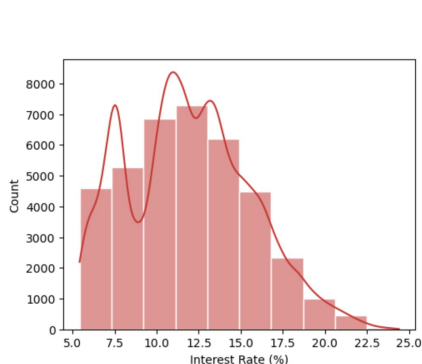
Univariate Analysis

- Most of the loan accounts have already been marked as Fully Paid
- No. of loans with 36 month term are much higher as compared to 60 months
- As the grade goes from A to G, there is a decline in the number of loan accounts
- As the employment length increases from 0 to 9 yrs there is a decline in the number of loan accounts with a steep increase at more than 10 yrs
- Most of the loan accounts are taken for debt_consolidation, with credit_card ranking at 2nd. There is also a substantial number of loans where the purpose is not specified
- No. of loans taken decreases as there is an increase in the loan amount



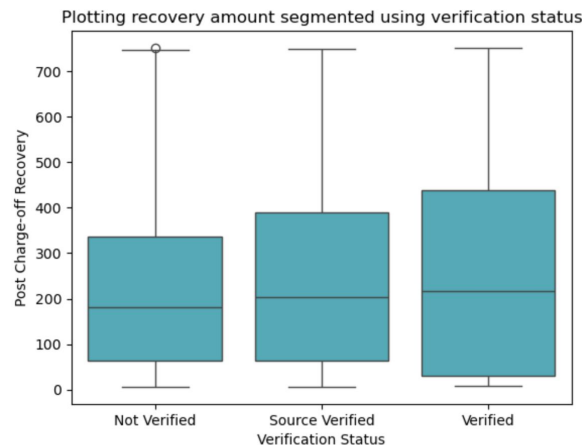
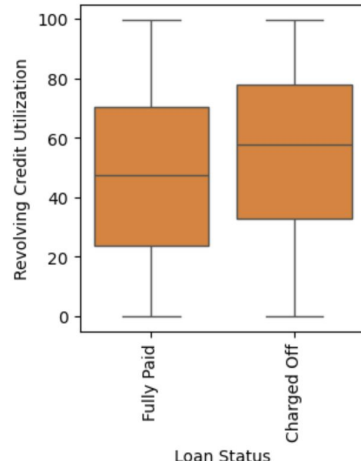
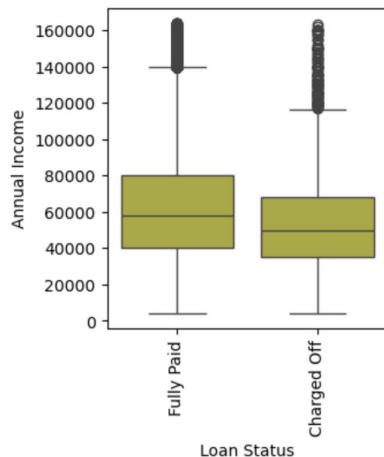
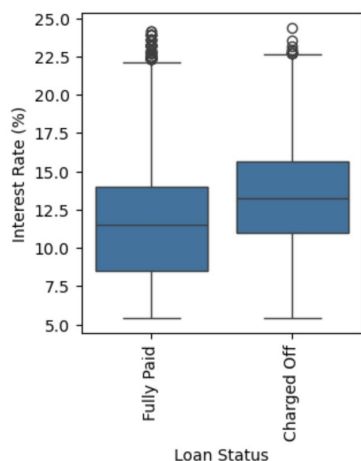
Univariate Analysis (cont...)

- No. of loans taken increases with the interest rate going up until 10-12% after which it starts declining
- No. of loans increases as Debt-to-Income ratio becomes higher peaking at 15, post which it starts declining
- There is a gradual increase in the no. of loans throughout the months. Also, the accounts have consistently increased from 2007 to 2011
- California (CA) tops the list in number of loans followed by New York (NY) and Florida (FL)



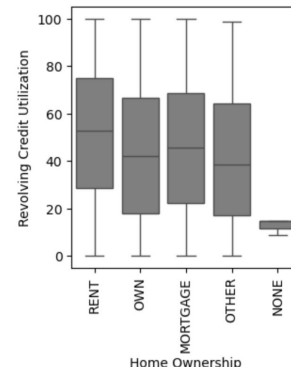
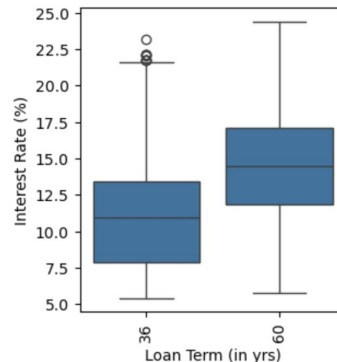
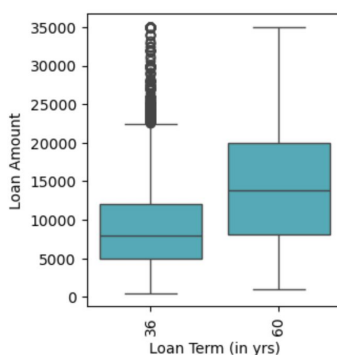
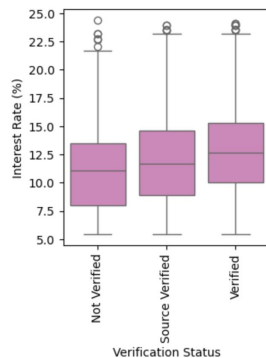
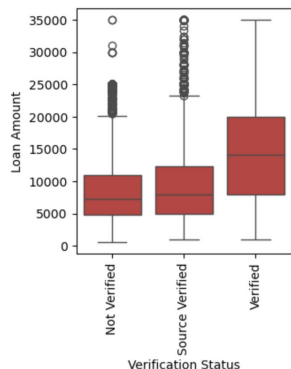
Segmented Univariate Analysis

- Median Interest Rate in case of Charged-Off loan is higher to that of Fully-Paid loan
- Median Annual Income of borrower is lower for Charged-Off loan as compared to Fully-Paid loan, also the revolving credit utilization is higher
- Quartile Recovery amount is the highest where Borrower's income is verified



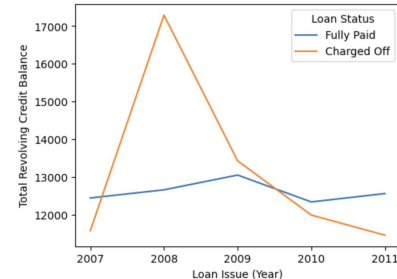
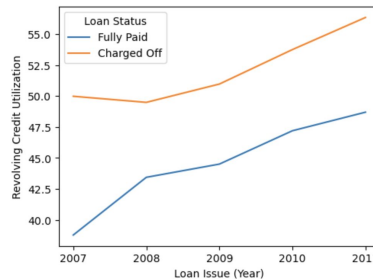
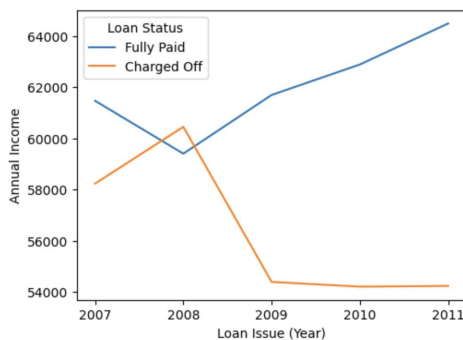
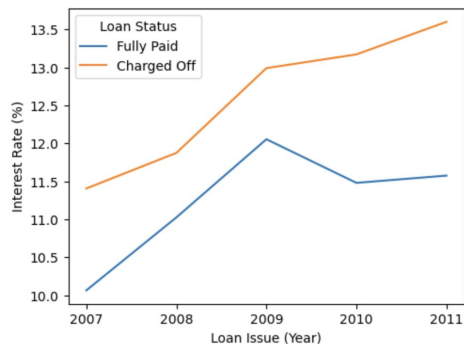
Segmented Univariate Analysis (cont...)

- Borrowers with Verified Income are more likely to get a higher loan, but also at a higher interest rate
- Borrowers take a bigger loan for a longer duration, and also the interest rate are much higher as the loan duration increases
- Borrower's having their own home are likely to have a lower Revolving Credit utilization



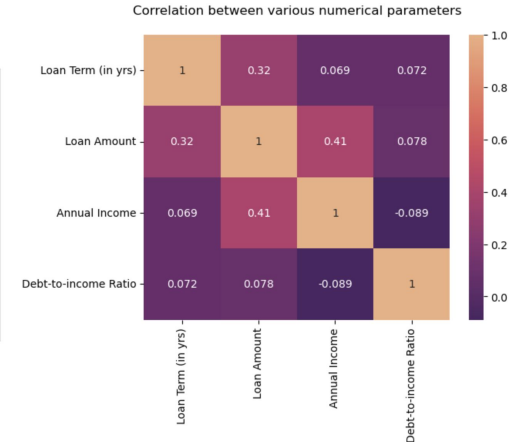
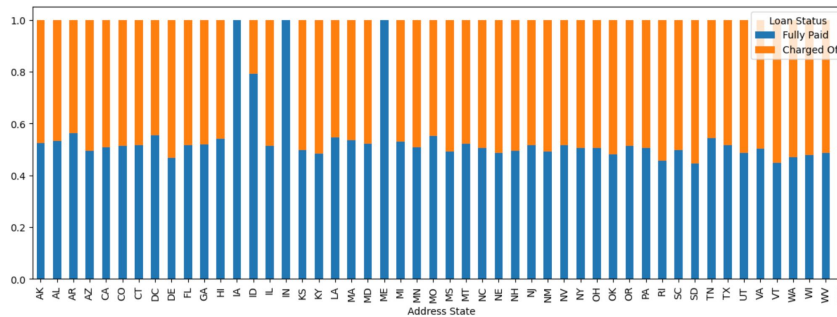
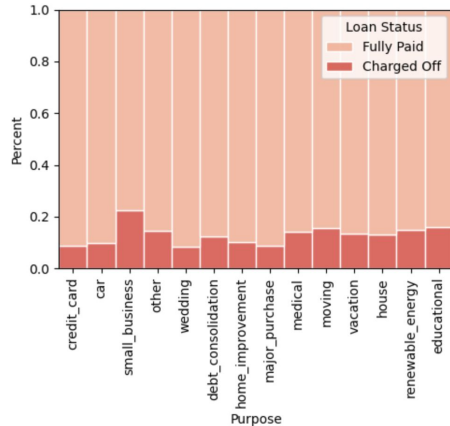
Bivariate Analysis

- Average Interest Rate for Charged-Off loan has seen substantial growth compared to Fully paid loans between 2007 to 2011
- Average Income of Borrower have dropped substantially for Charged-Off loans between 2007 and 2011
- Revolving credit utilization for charged-off loans is higher to that of fully-paid loans, although the credit balance is similar



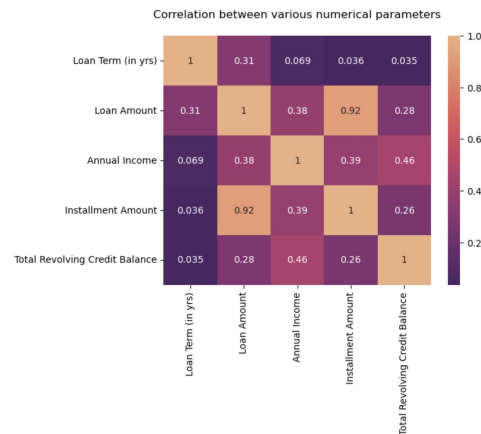
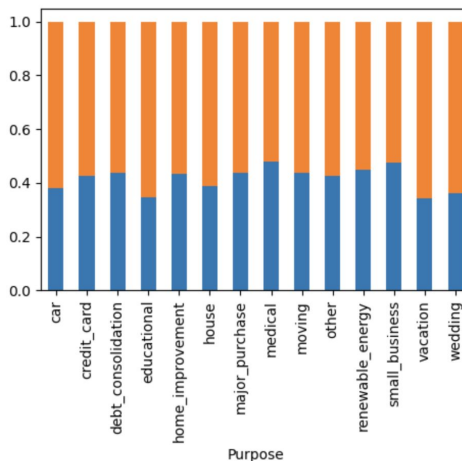
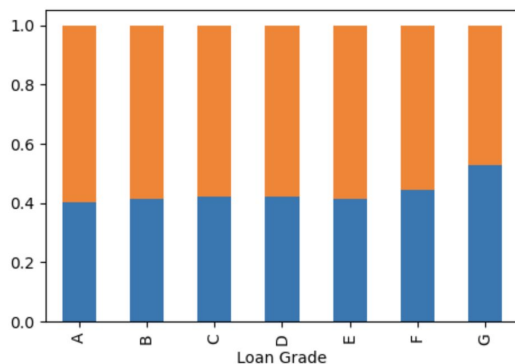
Bivariate Analysis (cont...)

- Small business loans have higher probability of getting charged off while the number of loans getting charged off are highest for debt consolidation
- Safest states to provide loan are Iowa (IA), Indiana (IN) and Maine (ME). Riskiest states would be Rhode Island (RI), South Dakota (SD), Vermont (VT)
- Weak correlation exists between Loan Amount and Loan Term / Annual Income



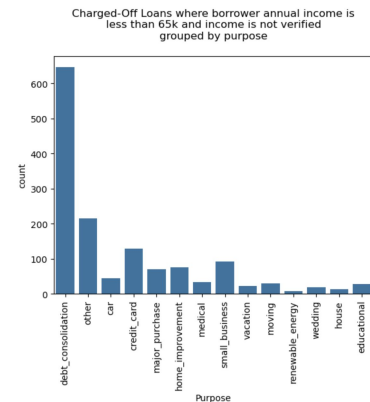
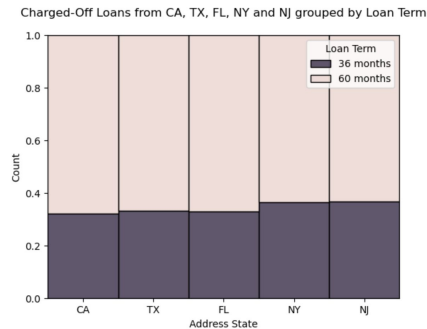
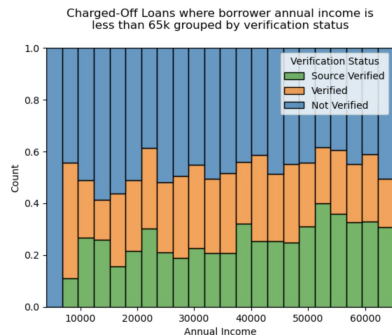
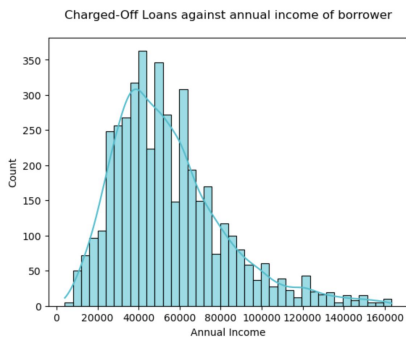
Derived Metrics

- No. of installments paid for charged-off loans is higher for short term loans as Loan Grade goes from A to G
- No. of installments paid for charged-off loans is higher for short term loans taken for medical or small business
- There is a weak correlation between Monthly income and Installment Amount for charged off loans



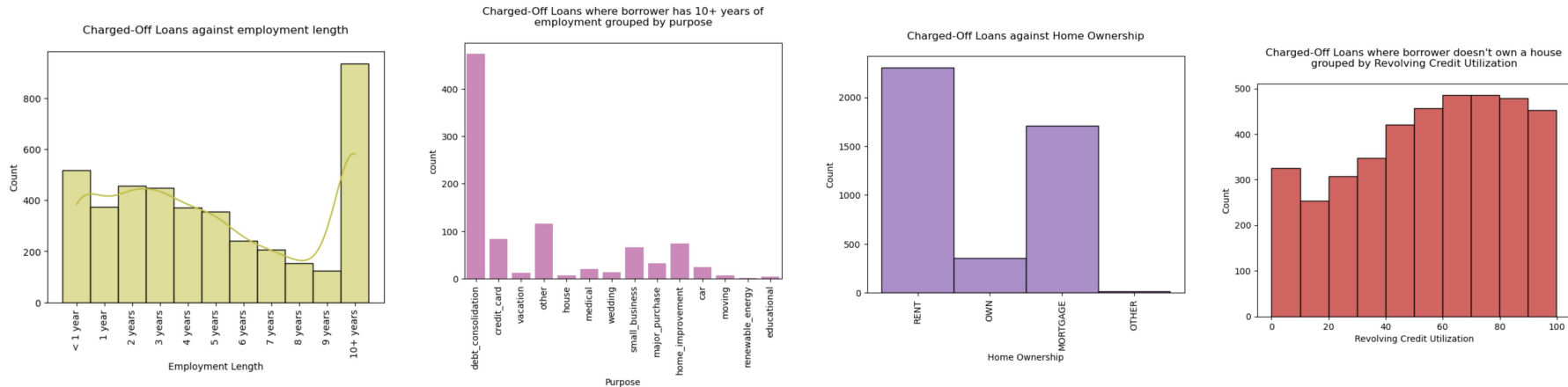
Conclusion

- Loans should be provided to borrowers with higher annual income
- When providing loans to borrowers with lower annual income, verification of income should be made mandatory
- Without verification, loan should not be provided for debt consolidation, small business or credit card. It should not also be provided if the purpose is not specified
- Short term loans should be avoided in California (CA), Florida (FL), New York (NY), NJ (New Jersey) and Texas (TX)



Conclusion (cont...)

- Loans should be provided to borrowers with more than 2 years and less than 10 years of employment length
- For borrower's with more than 10 years employment, avoid loans for Debt Consolidation
- Loan should be provided to borrowers owning their own house, for others Low Revolving Credit Utilization should be considered



Thank
You!