



GRAMMENER CASE STUDY

Submitted By : *Srinivasan. G, Naveed.J, Denny.J and Kumar.A*

Date of Submission : *30-December-2018*

Batch : *September - 2018*



Problem Statement

- A Consumer Finance Company lends various types of loans to urban customers
- If the Applicant is likely to repay the loan and if loan is rejected for this applicants then its loss of Business to the company
- If the applicant is not likely to repay the loan and if loan is approved for this applicants then this would lead to Financial loss to the company
- To avoid this, the company wants to know the potential patterns if a person likely to default i.e. not repay altogether or not in time

Methodology Applied

Data Cleaning

- Fix rows, columns, Null and Invalid Values
- Standardize, Transform, Deduplicate

Univariate Analysis

- Analyze one Variable at a time
- Data distribution, Summary Metrics
- Segmented Univariate analysis – Comparison of averages and metrics

Bivariate Analysis

- Analysis on Continuous Variables
- Correlation between variables

Data Analysis & Cleansing

- There are 111 columns or variables in the loan data frame.
- Remove NULL or Missing or Invalid Values
 - *54 columns from the loan has all values NULL is removed*
 - *Remove columns have more than 70 % ONLY has 0 , NA or both*
 - *Removed columns which non-essential for analysis*
- Standardize, Transform and Deduplicate
 - *Transform Month & Year to standard Datetime Format*
 - *Remove redundant key attribute – member_id*
 - *Transform ‘%’ from columns and convert to numeric*
 - *Convert columns which are Non-numeric to Numeric format*
 - *Standardize all string columns/attributes to common (Upper) Case.*
 - *Filter out rows to new data frames based on Loan Status for analysis*

Overall Loan Status

- After Data Cleansing the filtered data frames have ;
 - ***Charged off*** – 5627 rows and 41 columns (14%)
 - ***Full Payment*** – 32950 rows and 41 columns (83%)
 - ***Current*** – 1140 rows and 41 columns (3%)
- The customer & loan attributes that influence the tendency to default the loan leading to 'credit loss' are considered .
 - *The loan status not `Current` (Charged Off & Fully-Paid loan status) .*
 - *For credit loss analysis ,only data associated with charged off loan status.*

Univariate Analysis

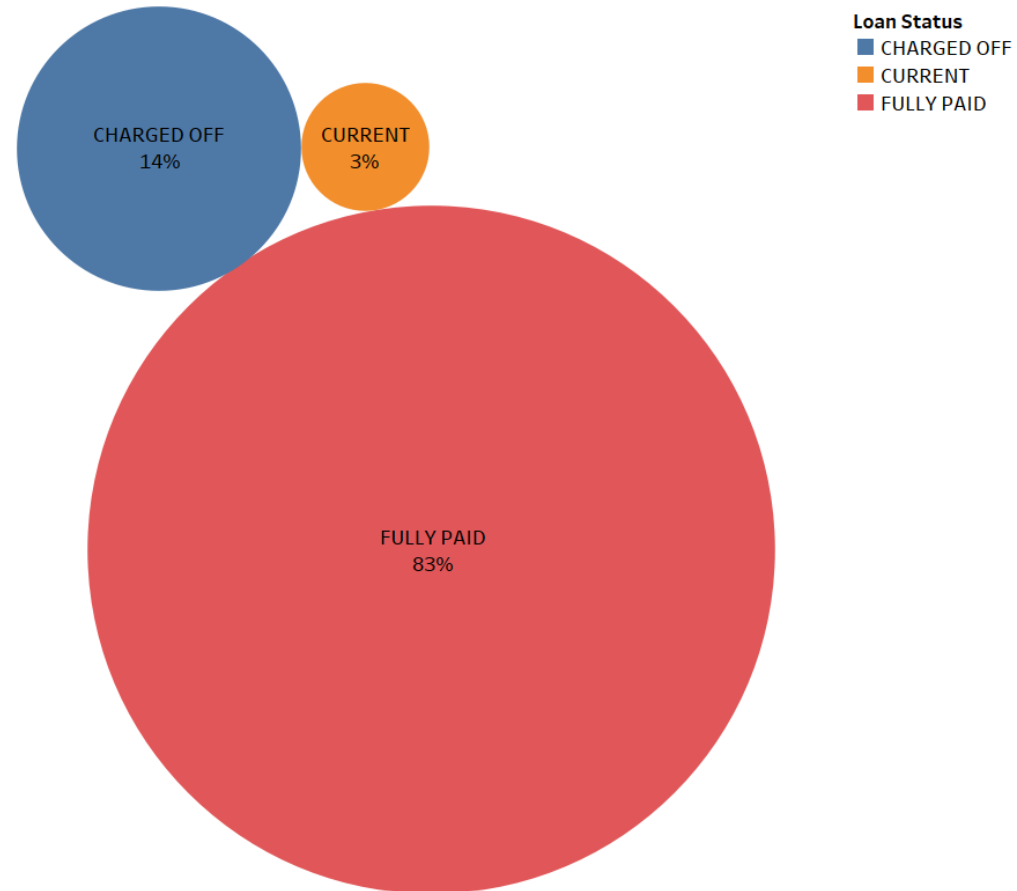
- Overall Loan status and charged off Loans
- Variables against the number of charged off loans such as ;
 - *Categorical Variables (Nominal & Ordinal)*
 - Term, Verification Status, Grade, Sub Grade and Employee Experience
 - Home Ownership and Resident State
 - Purpose
 - *Continuous Variables (Interval)*
 - Loan Amount,
 - Funded Amount , Funded Amount Invested,
 - Interest Rate
 - Instalment, Annual Income, DTI,
 - Revolving Balance and Revolving Utility %
 - Issue Date, Pulled Credit Date, Last Credit Payment Date
- *Refer Section 3.1 Univariate Analysis in Jupyter Notebook for the details*

Bi-Variate Analysis

- It includes the below for the charged off loans ;
 - *Home ownership against Verification Status*
 - *Segmented Annual Income Group (Bins)*
 - For Term
 - Verification Status
 - Segmented Annual Income Group
 - *Interest Rate Group (Bins)*
 - For Term
 - Employee Experience
- Correlation analysis
 - *Loan Amount against Funded Amount for different purpose*
 - *Funded Amount Invested against Instalment for different purpose*
 - *Loan Payment against Loan Payment Invested for different purpose*

Overall Loan Status

Plot-1. Percentage Issued Loans against the Loan Status



Insight-1 :

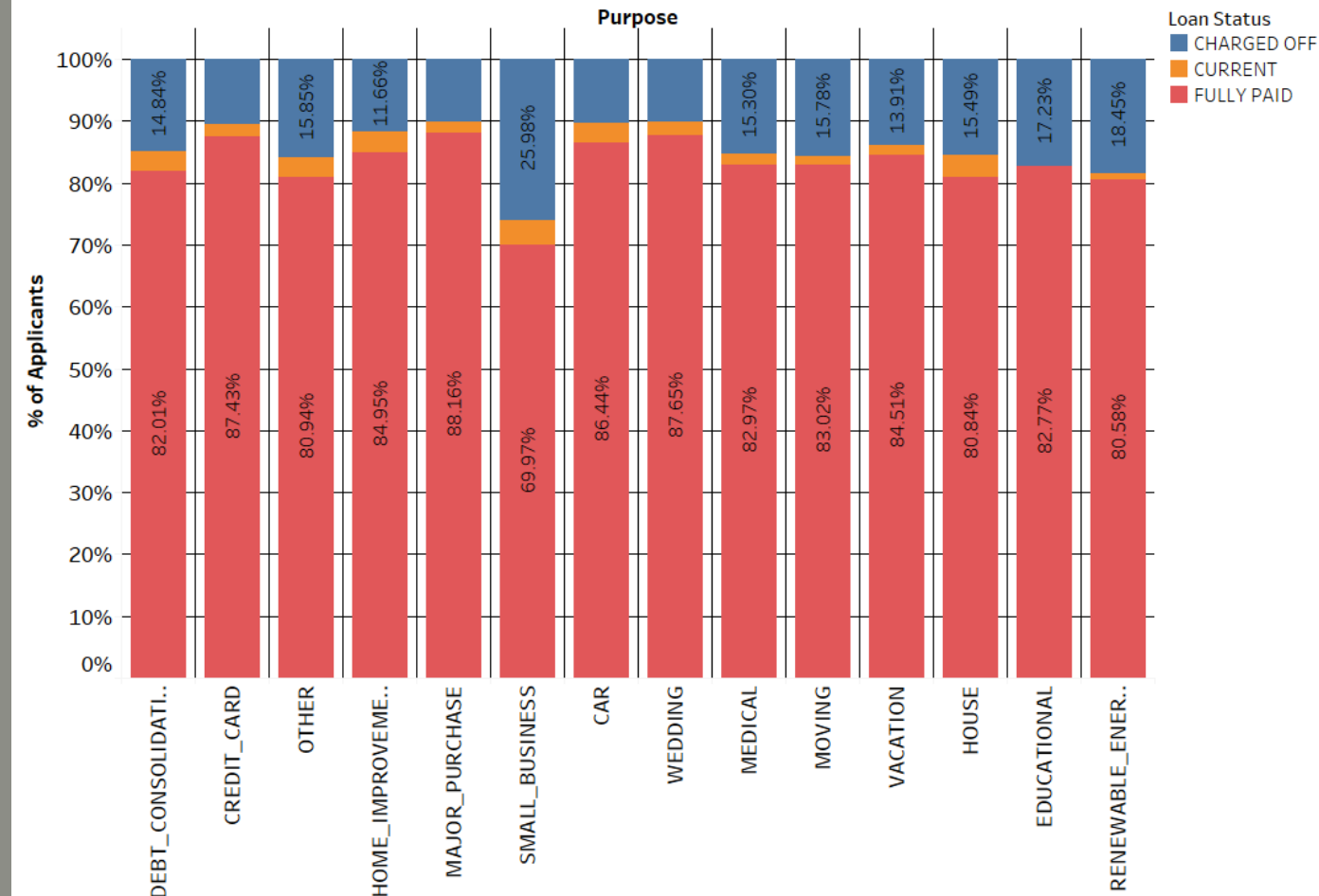
[1] ~ 14% out of all issued loans under consideration have resulted in credit loss.

Loan Purpose versus Loan Status Analysis

Insight-2 :

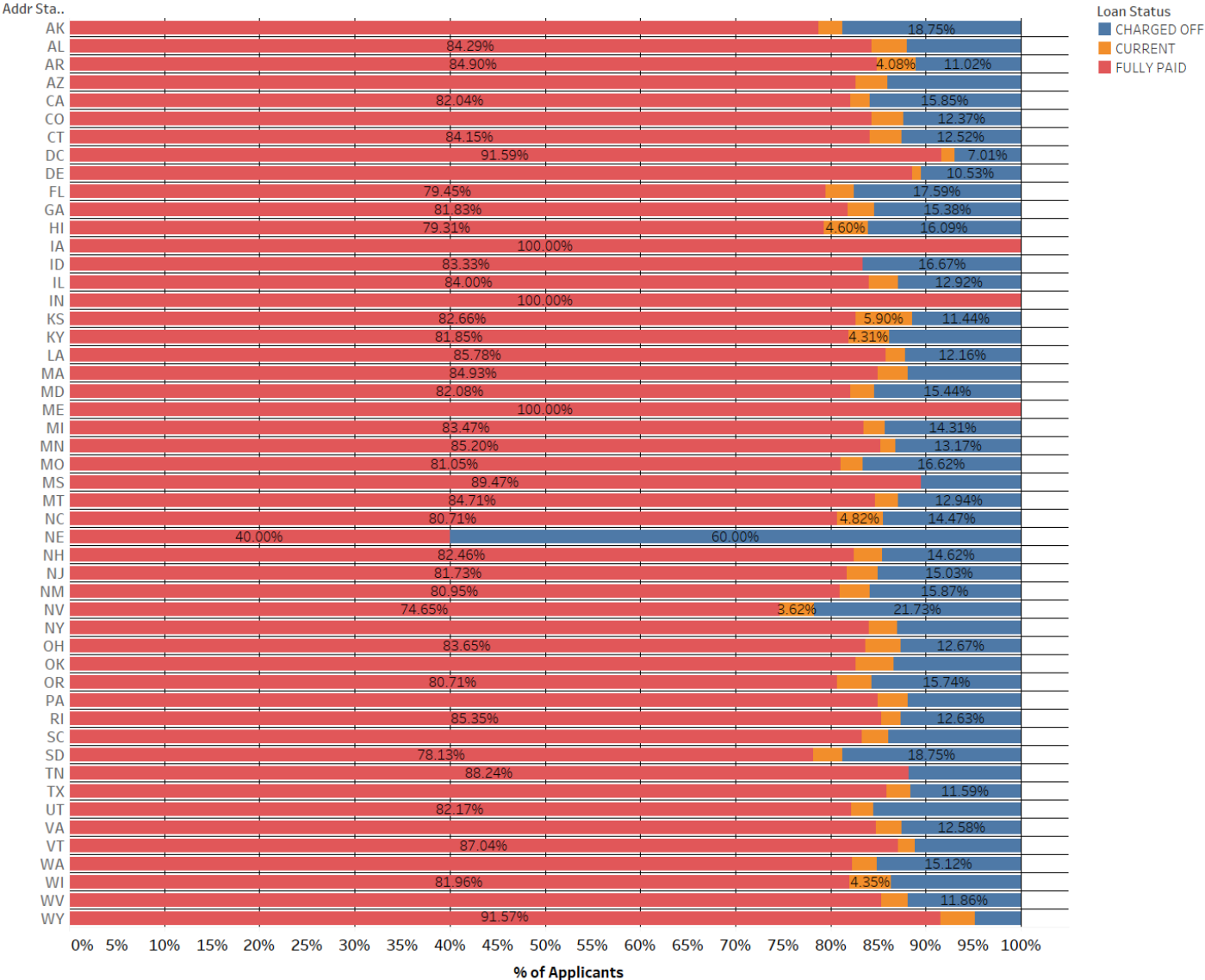
- [1] Although the highest number of loan applications are received from debt consolidation From the plot percentage of defaulter's loan purpose is **highest for small business**,
- [2] 26% of all loans taken from small business result in credit loss to the company.
- [3]

Plot-2. Loan Purpose Versus Loan Status Analysis



Loan Purpose	Charged Off	Current	Fully Paid	Grand Total	Fraud %
SMALL_BUSINESS	475	74	1279	1828	26.0
RENEWABLE_ENERGY	19	1	83	103	18.4
EDUCATIONAL	56		269	325	17.2
OTHER	633	128	3232	3993	15.9
MOVING	92	7	484	583	15.8
HOUSE	59	14	308	381	15.5
MEDICAL	106	12	575	693	15.3
DEBT_CONSOLIDATION	2767	586	15288	18641	14.8
VACATION	53	6	322	381	13.9

Plot-3. Resident State Against Loan Status Analysis



Application Resident State versus Loan Status Analysis

nsight-2 :

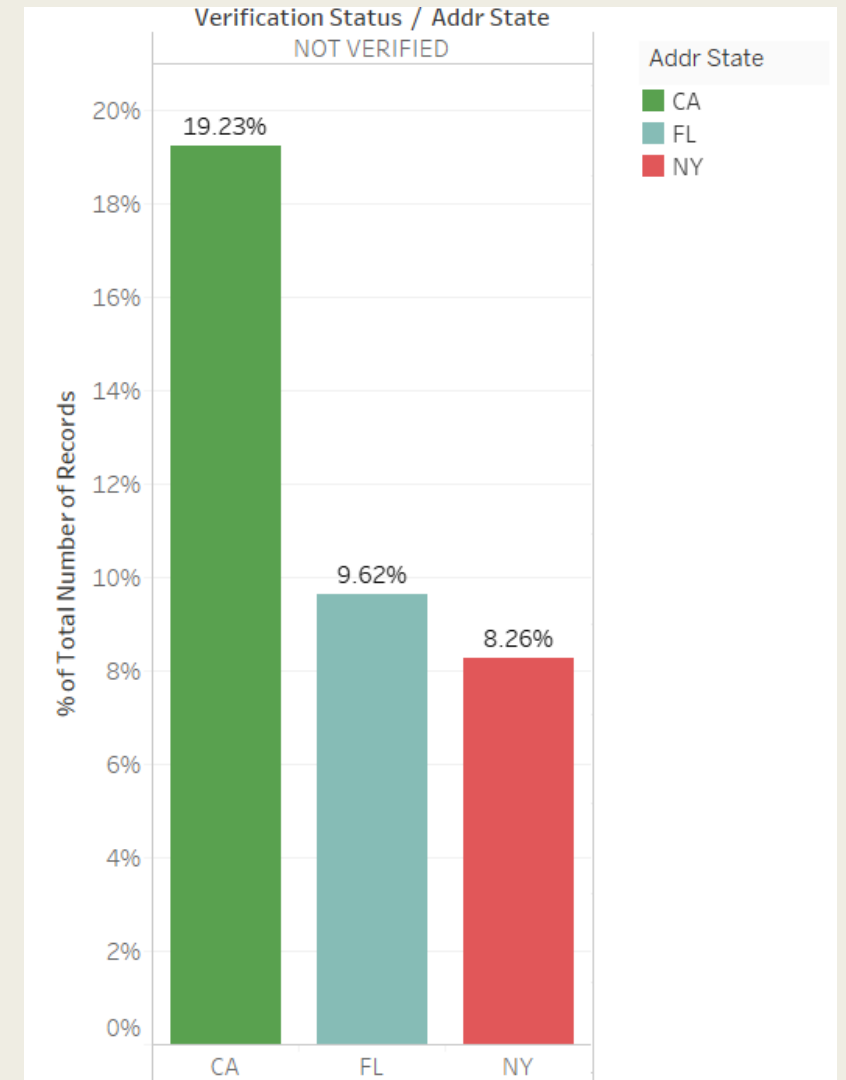
- 1 From the proportionality plot the borrowers belonging to state of Nebraska and Nevada (with 60% and 22.5% respectively) contribute to the highest default percentage. However the sample size for the state of Nevada is only 5. Therefore disregarding it and the next highest state is Nebraska.
- 2] The Highest number of loan applications are received from California, New York, Florida, and Texas. Their respective default rates are 16.2%,13.4%, 18.1% and 11.9%.

3]

State	Fully Paid	Charged Off	Grand Total	Fraud %
CA	5824	1125	6949	16.2
NY	3203	495	3698	13.4
FL	2277	504	2781	18.1
TX	2343	316	2659	11.9
NV	371	108	479	22.5
NE	2	3	5	60.0

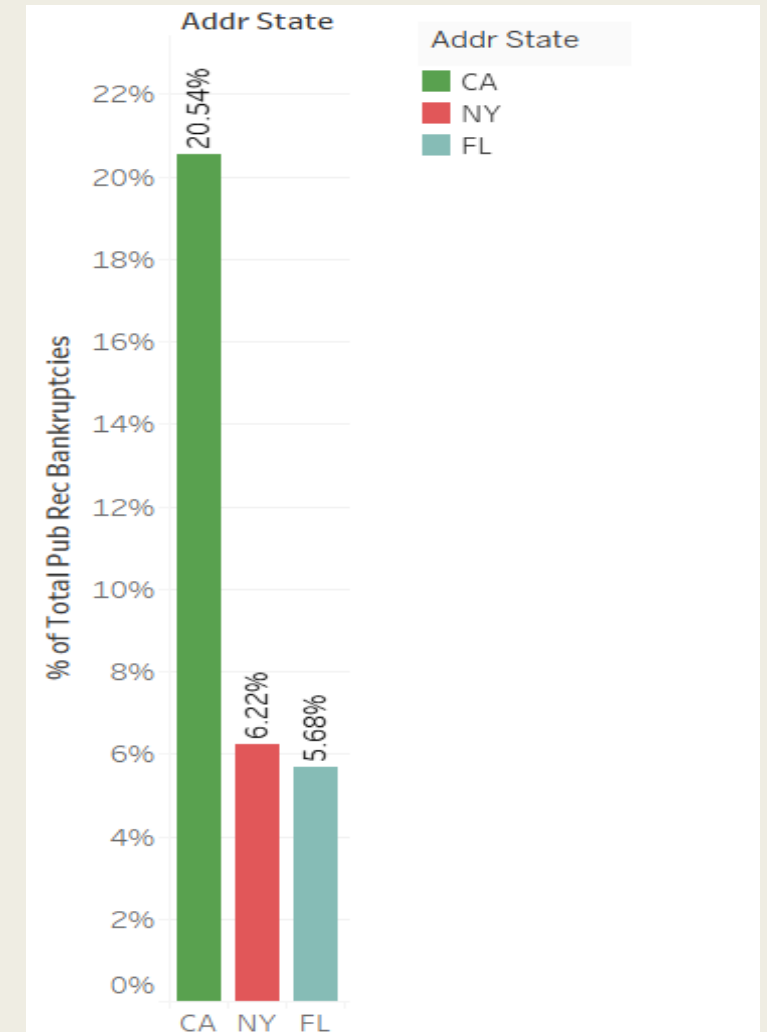
Top 3 States Approved Loan without Verifying

- California has the highest number of borrowers where are provided without verifying the their source of Income at 19%, followed by Florida and New York at 10% and 8% respectively.



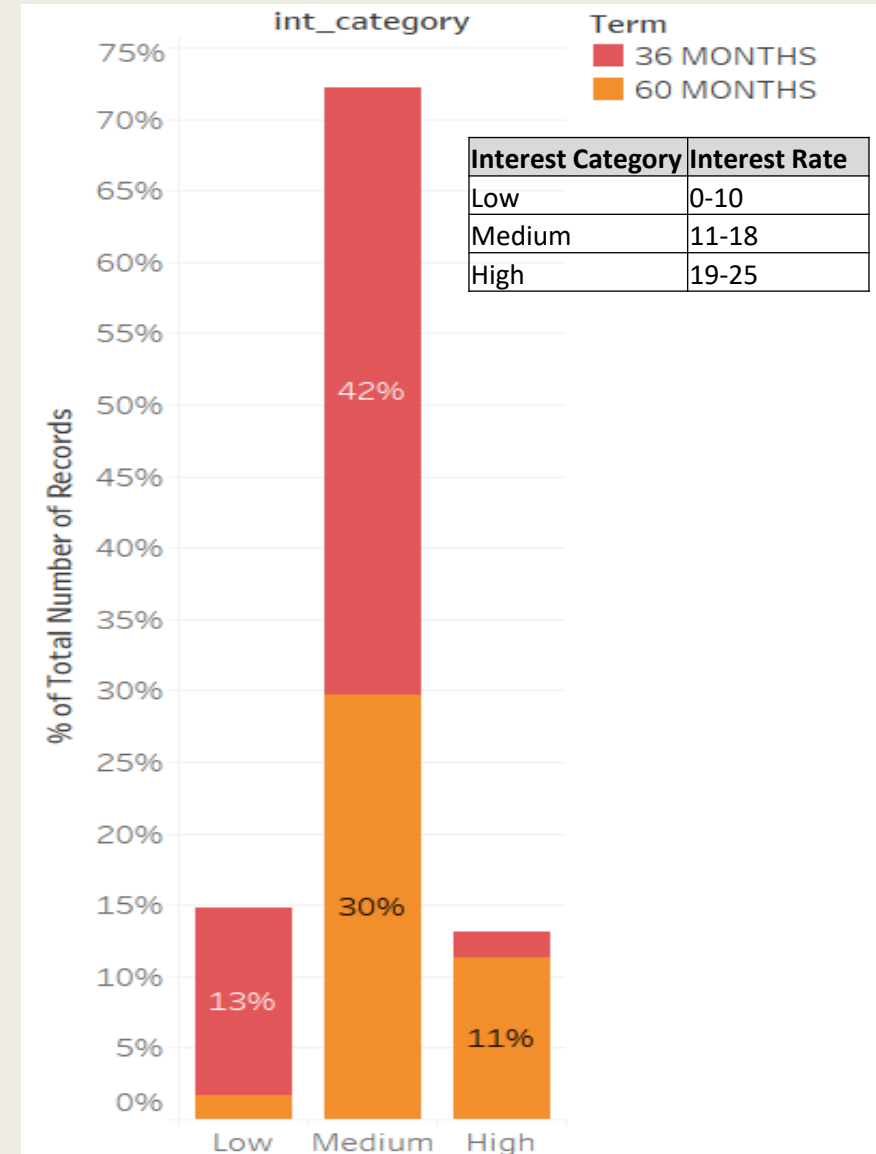
Top 3 States – Pay Installment Late

- Based on the graph we can infer that California, New York followed by Florida still remain the top three states where the installments are paid late.



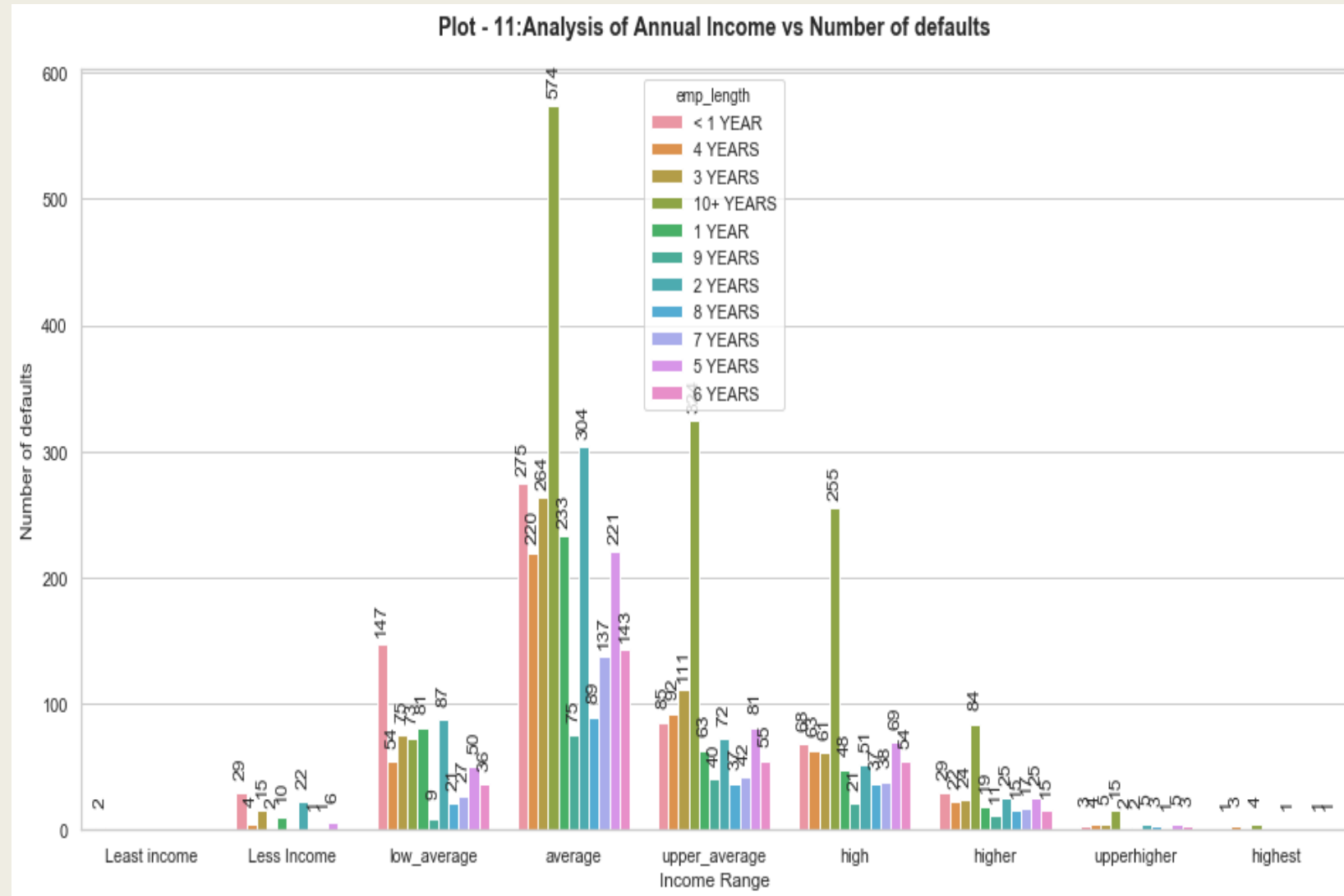
Interest Category Vs Term

- Most of the borrowers falls in the medium category and prefer a term of 36 months.
- At higher interest rates borrowers prefer a term of 60 months rather than 36 months.



Annual Income Vs Employee length

- Loan Applicants in the Average income zone(30000 - 60000) having experience ≥ 10 years are peak Defaulters.
- We further zoom in to find the residential status of these 10+ years of experienced Applicant

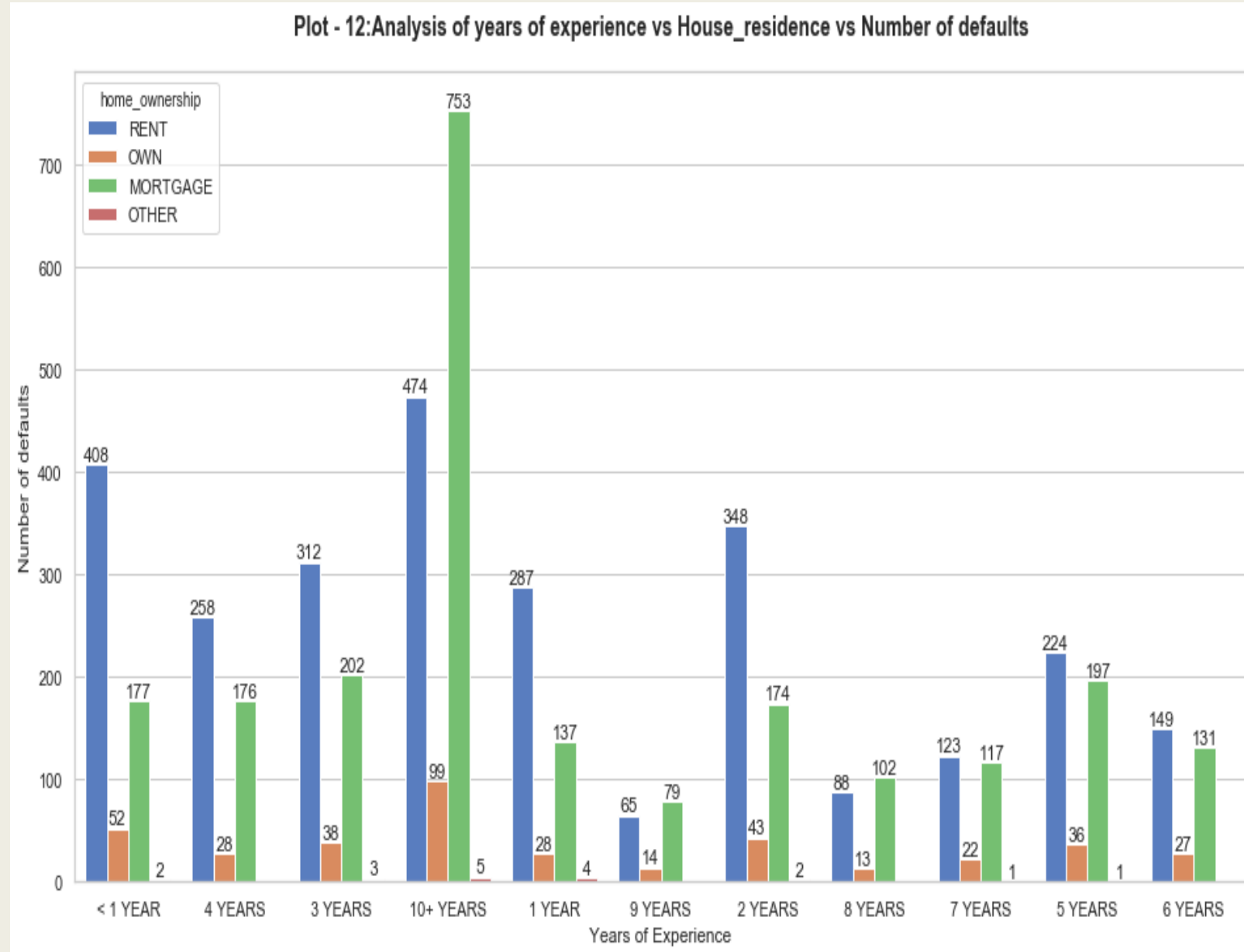


Years of Experience Vs House_residence

- Applicants who do not own a House after 10 or more years of Experience have a high default Count

- Summarizing the 2 slides:

Applicants who could not grow professionally in terms of Annual income in proportion to Work Experience peak the defaulters list



Conclusions

- It has been noticed that 3 states CA, FL and NY are approving Loans without verifying. Almost 40% of Defaulters are from this 3 states. This 3 states should focus on verifying before Approving the Loan. It is also been observed that this 3 state's consumers pay Installment late too
- Nebraska has the maximum number of Defaulters. So this state's loan applications need to be strictly scrutinized
- Loan Applicants who are taking loan for Small Business are the maximum who are Defaulters which is around 27%. More investigation needs to be done on why this Purpose of Loan Consumers are unable to repay the loan
- Loan Applicants who have taken medium interest rate 10-18% are the maximum defaulters. The interest rate needs to be reduced for this consumers
- Loan Applicants who could not grow professionally in terms of Annual income in proportion to to Work Experience and do not own a house peak the defaulters list