Is Borrower Defaulter?

The Largest Lending Company

By

1. Gaurav Singh

2. Vipul Narain

3. Pavithran PV

4. Abhinav Singh

Identify Risk of Loan Approval

- Goal
  - Identify Safe & Risky Consumer
  - Reduce Credit Loss
  - Increase Profit
- Approach : Exploratory Data Analysis
  - Load Loan Dataset, Data Cleanup
  - Identify Driving Factors/Variables for Defaulter using Univariate & Bivariate Analysis
- Conclusion
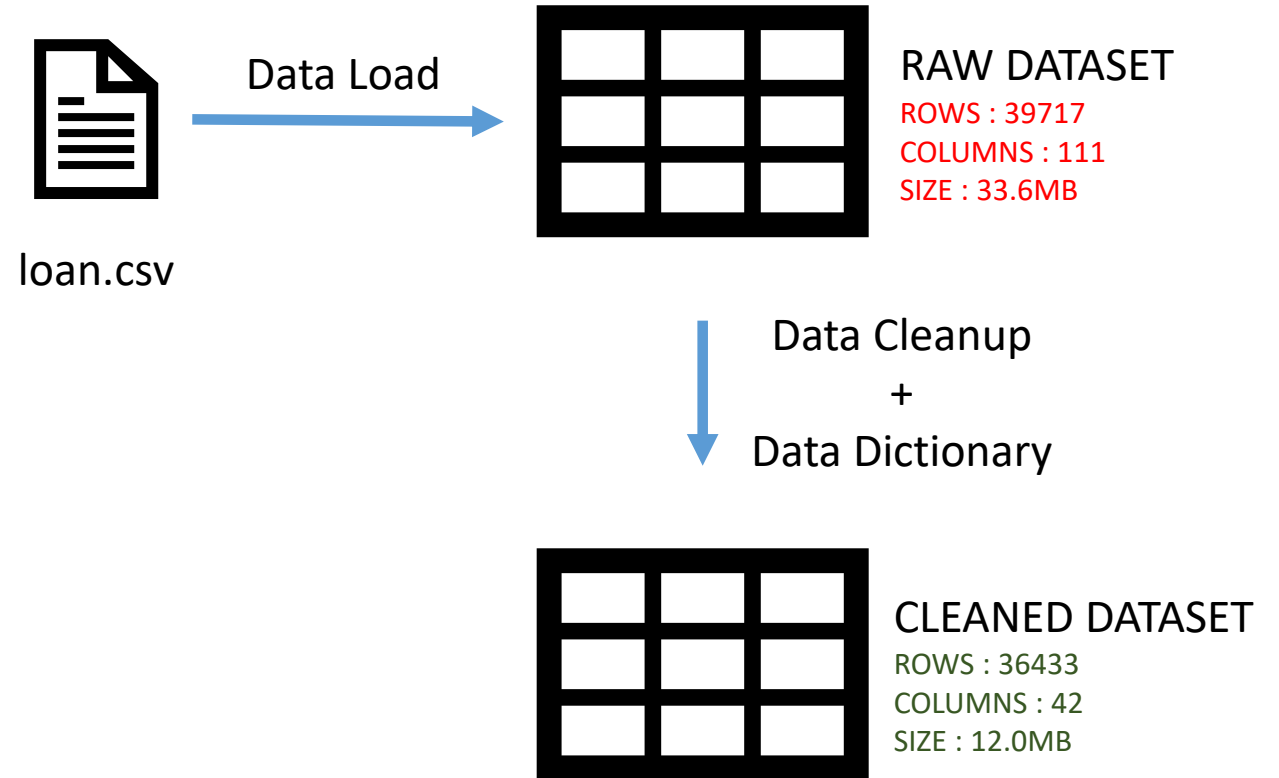 - Driving Factors and their effect on Consumer being Defaulter or not

Goal

- Identify Safe & Risky Consumer
  - Identify the Consumer who did not Repay the the loan and are called **Defaulters**.
  - Use loan data which contains:
    - Charged-Off Consumers : Did not repay loan
    - Fully Paid Consumers : Repaid the loan
    - Current Consumers : Ongoing loan

- Reduce Credit Loss
  - By identifying Defaulters Lending Company can avoid giving loan to them and reduce the Credit Loss.

- Increase Profit
  - Identifying Safe Consumers will help in Lending Company to increase its Profit

Approach : Exploratory Data Analysis

- Loading Loan Dataset
- Data Cleanup & Understanding
  - Drop empty Columns
  - Drop Columns with Single Value
  - Get rid of rows with empty data
  - Standardizing Column Values
    - Lower Case Conversion of all values in a column
    - Removing unnecessary prefix like '%' and 'years' etc.
    - Converting the Date columns to datetime
  - Standardizing Column Names
    - Columns renamed to identify their business meaning

loan.csv

Data Load →

RAW DATASET
ROWS : 39717
COLUMNS : 111
SIZE : 33.6MB

Data Cleanup
+
Data Dictionary

CLEANED DATASET
ROWS : 36433
COLUMNS : 42
SIZE : 12.0MB

EDA : Univariate Categorical Analysis

- Ordered Categorical Variable
  - Job Experience of Borrower
- Unordered Categorical Analysis
  - Status of Home Ownership of Borrower
    - RENT
    - MORTGAGE
    - OWN
    - OTHER
  - Income Verification Status
    - Not Verified
    - Verified
    - Source Verified
  - Loan Status
    - Fully Paid
    - Charged Off
    - Current
  - Number Of Payments(Month) : 36 Months, 60 Months
 - Purpose : The purpose why loan has been taken

Analysis Result(For Categorical Univariates)

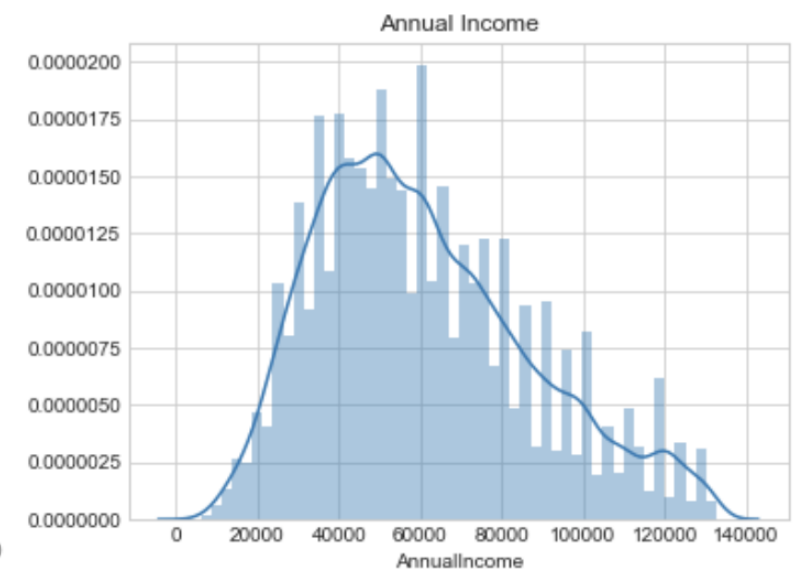| Attribute Name	Analyzed Info
JobExperienceOfBorrower(Years)	Most of the borrowers have experience 10+ Years
HomeOwnershipStatusOfBorrower	Income is 'not verified' for most of the borrowers
loan_status	83.5% borrowers => fully paid the loan 13.6% => Defaulters 2.9% => Current Consumers
NumberOfPayments(Months)	Loan for 36 months => 26377 Borrowers Loan for 60 months => 10056 Borrowers
purpose	Most borrowers took loan for "debt consolidation"

EDA : Univariate Numeric Data Analysis

- Check whether Number Attributes are Normally Distributed or not
- For all the Numeric Attributes Distribution has been drawn
- **Annual Income** column is **Positively Skewed,** i.e. there are Outliers in positive side

- Distribution Plot for relevant numeric columns

- Central Tendency of Numeric Attributes

- Get rid of Outliers
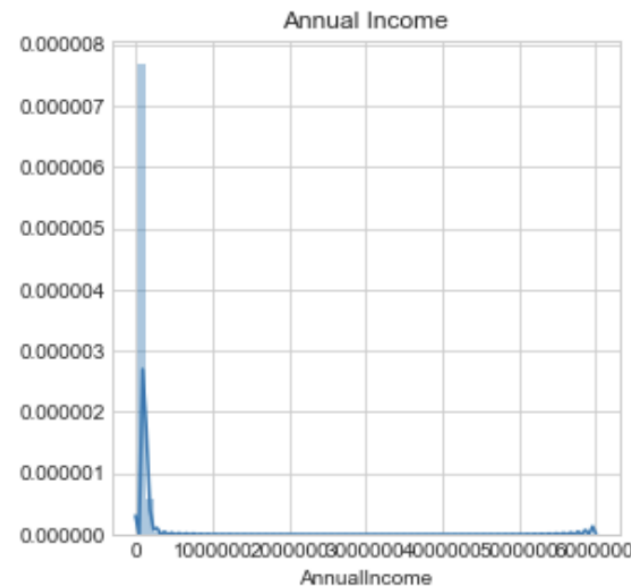

Before Removing Outliers | After Removing Outliers

Analysis Result(For Numeric Univariates)

After removing Outliers from Annual Income attribute, not only Annual Income distribution became Normal but all other Numeric attributes also become much more Normal Distribution

EDA : Segmented Univariate Analysis

- Analyzing Final Loan Amount Taken for Segmented Loan Status

- Analyzing %Debt To Income Ratio for Segmented Loan Status
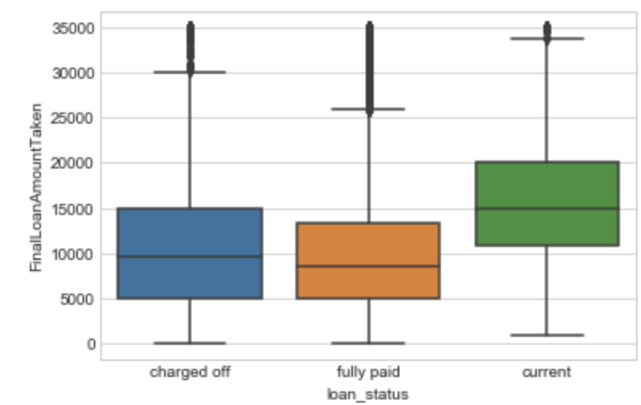
Analysis Result
(Loan Status & Loan Amount Taken)

More the loan amount allocated more risk of consumer being a Defaulter

loan_status	count	mean	std	min	25%	50%	75%	max
charged off	4720.0	10964.274798	7412.819321	0.0	5000.0	9600.000000	15000.0	35000.0
current	970.0	15786.427547	7764.415487	1000.0	10812.5	14975.000000	20000.0	35000.0
fully paid	28550.0	9907.932266	6514.132767	0.0	5000.0	8494.815287	13350.0	35000.0
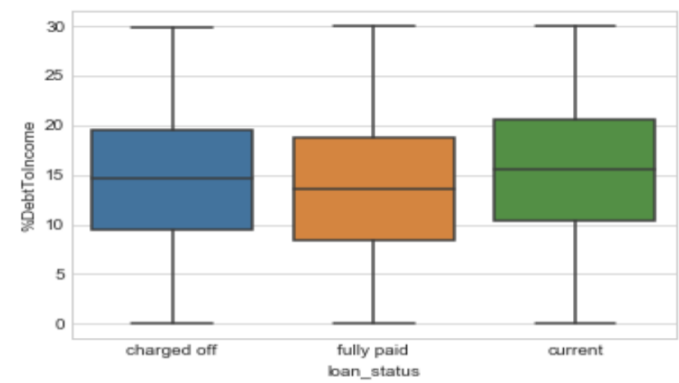
Analysis Result
(Loan Status & % Debt To Income Ratio)

Higher the %DebtToIncome Ratio, Higher is the chance of Loan Default

loan_status	count	mean	std	min	25%	50%	75%	max
charged off	4720.0	14.274388	6.534656	0.0	9.500	14.60	19.540	29.85
current	970.0	15.179876	6.714388	0.0	10.355	15.49	20.485	29.95
fully paid	28550.0	13.478514	6.639302	0.0	8.400	13.60	18.700	29.99

EDA : Segmented Univariate Analysis

- Analysing %InterestRate for Segmented Loan Status

- Analysing Annual Income for Segmented Loan Status
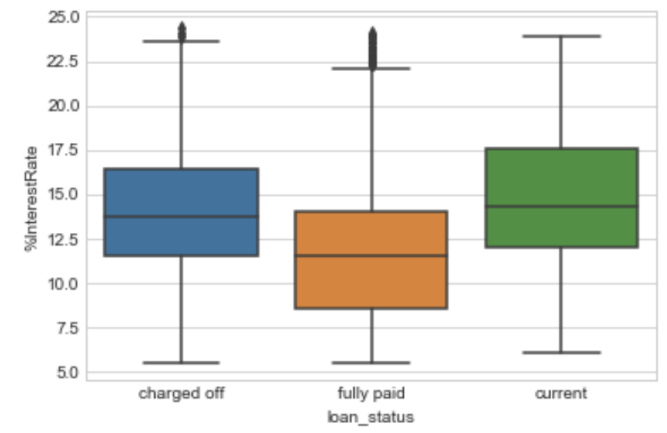
Analysis Result

(Loan Status & % Interest Rate)

Higher the %InterestRate, Higher is the chance of Default

loan_status	count	mean	std	min	25%	50%	75%	max
charged off	4720.0	13.866519	3.625928	5.42	11.48	13.67	16.40	24.40
current	970.0	14.926711	3.518501	6.00	11.99	14.27	17.58	23.91
fully paid	28550.0	11.609616	3.578114	5.42	8.49	11.49	13.99	24.11
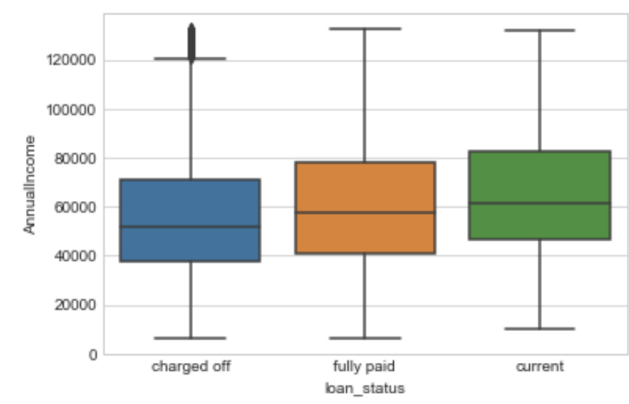
Analysis Result

(Loan Status & Annual Income)

Lower the Annual Income, Higher is the chances of Default

loan_status	count	mean	std	min	25%	50%	75%	max
charged off	4720.0	56540.640415	25348.782314	6000.0	37753.2075	52000.0	71031.0	132725.38
current	970.0	65631.076144	25196.167690	10020.0	46505.2200	61700.0	82500.0	132000.00
fully paid	28550.0	61313.072044	26501.715788	6000.0	40896.0000	57500.0	78000.0	132750.00

UpGrad

1. **Annual Income** is **-ve correlated** to **%DebtToIncome**, *it is understood because %DebtToIncome is inversely proportional to Income*
2. **Annual Income** is Moderately **+ve correlated** to **RevolvingBalance** & **FinalLoanAmountTaken**
3. **'%DebtToIncome'** is Moderately **+ve correlated** to **RevolvingBalance**
4. **'%InterestRate'** is Moderately **+ve correlated** to **FinalLoanAmountTaken**

EDA : Bivariate Analysis-Numeric Variable

- Analysing %InterestRate for Segmented Loan Status

- Analysing Annual Income for Segmented Loan Status

EDA : Bivariate Analysis- Categorical Variable

- loan_status and NumberOfPayments(Months)

- loan_status & JobExperienceOfBorrower(Years)

loan_status	NumberOfPayments(Months)	FinalLoanAmountTaken	AnnualIncome	%InterestRate	%DebtToIncome	RevolvingBalance
charged off	36	8202.50	53462.14	12.39	13.89	11812.97
	60	14398.12	60368.29	15.70	14.75	14099.30
current	60	15786.43	65631.08	14.93	15.18	15194.04
fully paid	36	8801.59	60157.12	10.84	13.22	11656.33
	60	13860.23	65442.58	14.36	14.41	13853.35

Analysis Result
(Loan Status, Loan Amount Taken & Numeric Values)

For companies profit & to avoid Defaulters

- 36 months Term
  - Max Limit of %DebtToIncome = 13.22%
  - Max Interest Rate = 10.84%
- 60 months Term
  - Max Limit of %DebtToIncome = 14.41%
  - Max Interest Rate = 14.36%

Analysis Result
(Loan Status, Job Experience in Years & Numeric Values)

- For any number of years of Job Experience, the Borrower is not a Defaulter if %Interest Rate is in the range of 11.50% to 11.74% but if the the interest rate is high from 13.61% to 14.06% then the Borrower could be defaulter.

Hence for Company's Benifit, for any Job Experience Level, the interest Rate should be in the Range of 11.50% to 11.74%

EDA : Bivariate Analysis- Categorical Variable

- loan_status & HomeOwnershipStatusOfBorrower

loan_status	HomeOwnershipStatusOfBorrower	FinalLoanAmountTaken	AnnualIncome	%InterestRate	%DebtToIncome	RevolvingBalance
charged off	MORTGAGE	12586.04	66364.34	13.79	14.50	16406.18
	OTHER	8954.72	55870.29	13.29	13.51	13981.47
	OWN	10629.05	50601.60	13.96	14.42	11610.06
	RENT	9756.13	49685.36	13.92	14.09	10198.00
current	MORTGAGE	16929.50	72747.72	14.47	15.35	17824.22
	OWN	14201.65	53193.46	14.91	15.04	10795.97
	RENT	14464.59	57835.66	15.57	14.97	12279.46
fully paid	MORTGAGE	10941.68	71025.11	11.26	13.30	15165.11
	OTHER	6949.93	60965.59	11.82	11.13	10835.69
	OWN	9247.90	52754.81	11.51	13.52	10607.56
	RENT	9117.64	54094.30	11.93	13.64	9725.66

Analysis Result
(Loan Status, Home Ownership Status & Numeric Values)

Irrespective whether the Borrower stays in own house or rented house the %Interest Rate should not exceed from the range of 11.26% to 11.93%.

Conclusion : Driver Factors/Variables for Defaulter

- %Interest Rate

- % Debt To Income Ratio

- Annual Income

- Number of Payments (in Month)

High Level Analysis Result

- *Higher the %Interest Rate goes more, chances of Borrower being a Defaulter will increase*

- *Higher the %DebtToIncome Ratio, chances of Borrower being a Defaulter will increase*

- *Lower the Annual Income, chances of Borrower being a Defaulter will increase*

Final Conclusion

If the Lending Company want to have less chances that Borrower will be a Defaulter and in tern reduce their Credit Loss, they have to limit % DebtToIncome ratio and %Interest Rate as follows:

- *In Terms of Number Of Payments by Month:*
  - ***36 months Term***
    - *%DebtToIncome Ratio <= 13.22%*
    - *%Interest Rate <= 10.84%*
  - ***60 months Term***
    - *%DebtToIncome Ratio <= 14.41%*
    - *%Interest Rate <= 14.36%*

By above limitations, chances of having Safe Consumer increases and hence Credit Loss will be reduced.