

# Predicting Vehicle Loan Defaults in India

India is the 3<sup>rd</sup>-largest Automobile market on  
Earth



“The current size of India’s retail vehicle finance industry is worth \$60 Bn now.”

—Invest India



# Lowering the default rate is worth Billions

Given the 21% default rate we've measured, a reduction to 10% would be worth \$6.6 Billion USD per year. Our target is to predict over 50% of those defaults.

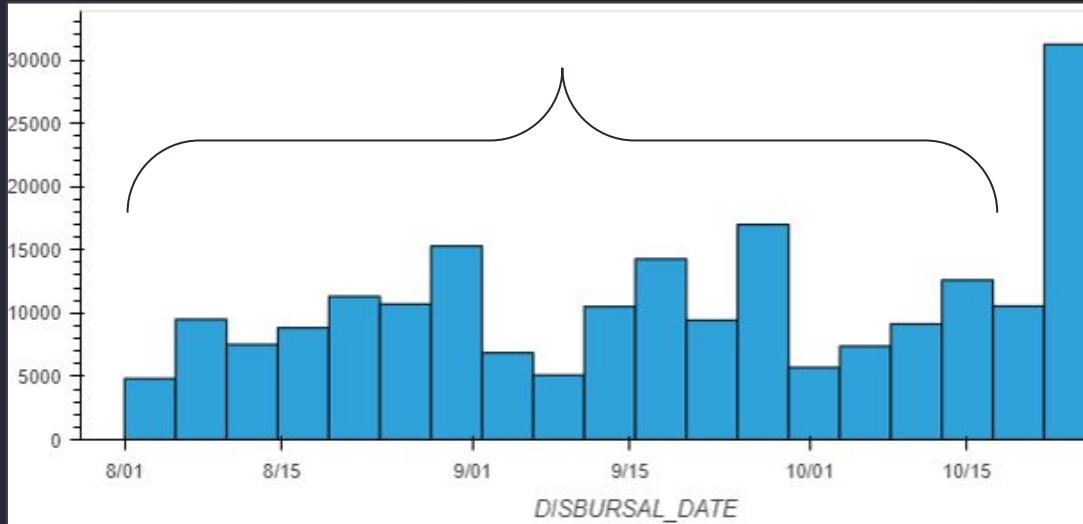


# Diwali

Diwali is a festival of new beginnings, the triumph of good over evil, and light over darkness. It also symbolizes the spiritual victory of light over darkness, good over evil.

This time of year is considered as the most auspicious time of year to make new purchases. One of the major items bought are new cars.

# Peak Months for Vehicle Purchases



Our data is for August-October, which corresponds with peak Indian car buying season leading up to Diwali.

Dhanteras, this day before Diwali is deemed perfect for buying gold, cars, bikes, or any valuable asset. It's an auspicious occasion for buying automobiles.

# Our Dataset

- 41 columns of loan information
- 233,156 loans in the training dataset
- 112,394 loans in the testing dataset
- 'LOAN\_DEFAULT' - OBJECTIVE
- Created Features
  - Credit\_History\_Total\_Month
  - Average\_Acc\_Age\_Total\_Month
  - Customer\_age
  - Amount\_Cost\_Diff is the difference between asset cost and disbursed amount

## Original Features

'UNIQUEID'

'DISBURSED\_AMOUNT'

'ASSET\_COST'

'LTV'

'BRANCH\_ID'

'SUPPLIER\_ID'

'LOAN\_DEFAULT' - OBJECTIVE

'CURRENT\_PINCODE\_ID'

'DATE\_OF\_BIRTH'

'EMPLOYMENT\_TYPE'

'DISBURSAL\_DATE'

'STATE\_ID'

'EMPLOYEE\_CODE\_ID'

'MOBILENO\_AVL\_FLAG'

'AADHAR\_FLAG'

'PAN\_FLAG'

'VOTERID\_FLAG'

'DRIVING\_FLAG'

'PASSPORT\_FLAG'

'PERFORM\_CNS\_SCORE'

'PERFORM\_CNS\_SCORE\_DESCRIPTION'

'PRI\_NO\_OF\_ACCTS'

'PRI\_ACTIVE\_ACCTS'

'PRI\_OVERDUE\_ACCTS'

'PRI\_CURRENT\_BALANCE'

'PRI\_SANCTIONED\_AMOUNT'

'PRI\_DISBURSED\_AMOUNT'

'SEC\_NO\_OF\_ACCTS'

'SEC\_ACTIVE\_ACCTS'

'SEC\_OVERDUE\_ACCTS'

'SEC\_CURRENT\_BALANCE'

'SEC\_SANCTIONED\_AMOUNT'

'SEC\_DISBURSED\_AMOUNT'

'PRIMARY\_INSTAL\_AMT'

'SEC\_INSTAL\_AMT'

'NEW\_ACCTS\_IN\_LAST\_SIX\_MONTHS'

'DELINQUENT\_ACCTS\_IN\_LAST\_SIX\_MONTHS'

'NO\_OF\_INQUIRIES'



## Vehicle Loan Prediction in India

### Loan & Personal Information

Government ID  
Aadhar Card

Government ID  
PAN Card

Unique ID

Vehicle State ID

Date of Birth

Disbursed Amount

Asset Amount

Primary Overdue Amounts

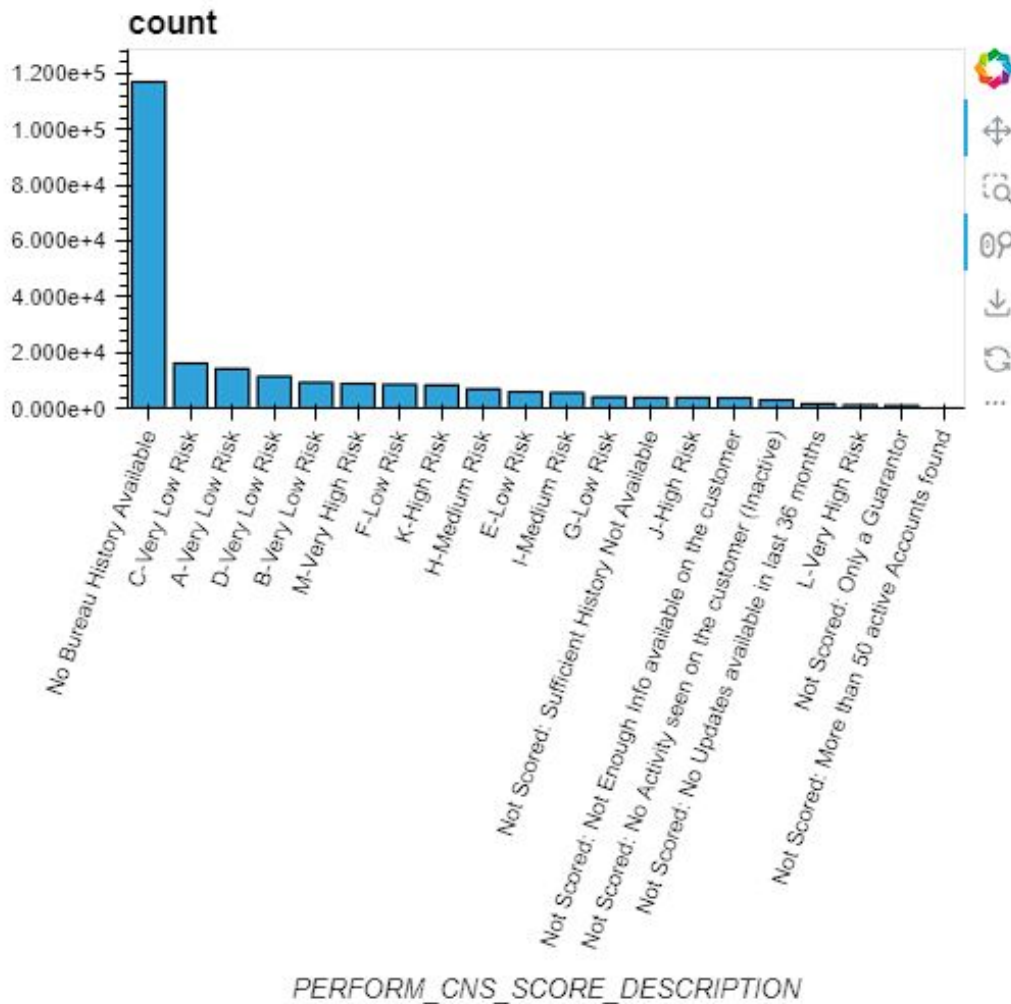
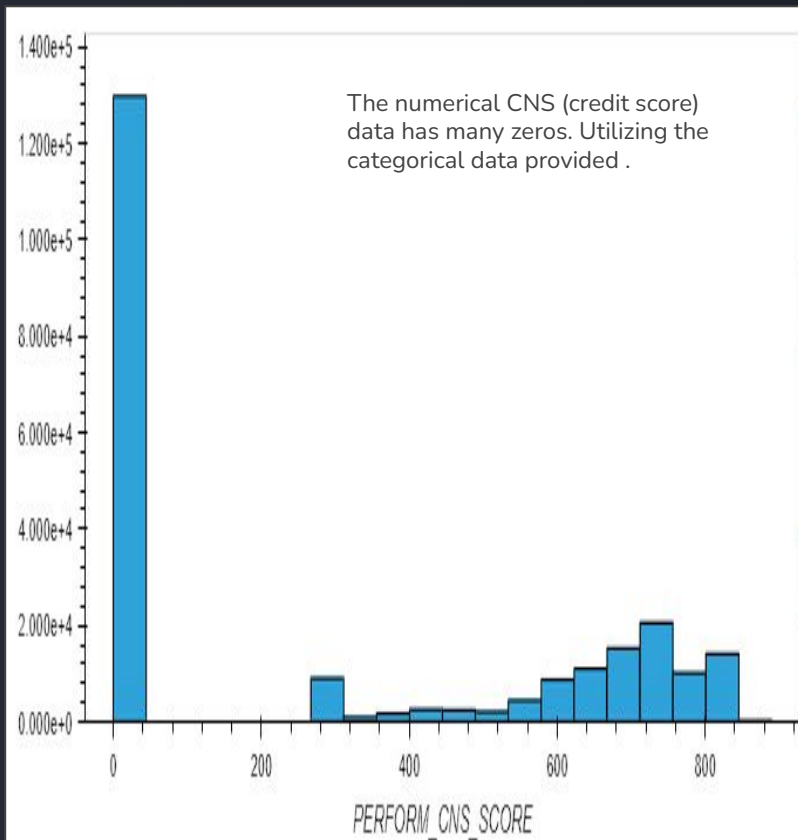
Average Acct. Age

Primary Current Balance

The minimum age of the individual to get a loan 21 years when applying for a loan. The candidate's car loan repayment tenure should not exceed 65 years of age. The applicant should be in their job for 2-3 years.

CIBIL Score Range	Rating
300-499	Poor
500-649	Average
650-749	Good
750-900	Excellent

# Data Exploration





PERFORM\_CNS\_SCORE has 116950 (50.2%) zeros

PRI\_NO\_OF\_ACCTS has 116950 (50.2%) zeros

PRI\_ACTIVE\_ACCTS has 137016 (58.8%) zeros

PRI\_OVERDUE\_ACCTS has 206879 (88.7%) zeros

PRI\_CURRENT\_BALANCE has 141696 (60.8%) zeros

PRI\_SANCTIONED\_AMOUNT has 138096 (59.2%) zeros

PRI\_DISBURSED\_AMOUNT has 138204 (59.3%) zeros

SEC\_NO\_OF\_ACCTS has 227289 (97.5%) zeros

SEC\_ACTIVE\_ACCTS has 229337 (98.4%) zeros

SEC\_OVERDUE\_ACCTS has 231817 (99.4%) zeros

SEC\_CURRENT\_BALANCE has 229790 (98.6%) zeros

SEC\_SANCTIONED\_AMOUNT has 229418 (98.4%) zeros

SEC\_DISBURSED\_AMOUNT has 229450 (98.4%) zeros

PRIMARY\_INSTAL\_AMT has 159517 (68.4%) zeros

SEC\_INSTAL\_AMT has 230937 (99.0%) zeros

NEW\_ACCTS\_IN\_LAST\_SIX\_MONTHS has 181494 (77.8%) zeros

DELINQUENT\_ACCTS\_IN\_LAST\_SIX\_MONTHS has 214959 (92.2%) zeros

NO\_OF\_INQUIRIES has 201961 (86.6%) zeros

## FEATURE ENGINEERING

### Variable Encoding, Data Cleaning & Feature Creation:

- The original column “**CREDIT\_HISTORY\_LENGTH**”, describes how long it has been since the customer's first loan, and is formatted as “Oyrs0mon”.

#### Feature Created:

Feature Name: “**Credit\_History\_Total\_Months**”

Description: Represents the length of the customer's credit history in months.

- The original column “**AVERAGE\_ACCT\_AGE**”, describes the average loan tenure and is formatted as “Oyrs0mon”.

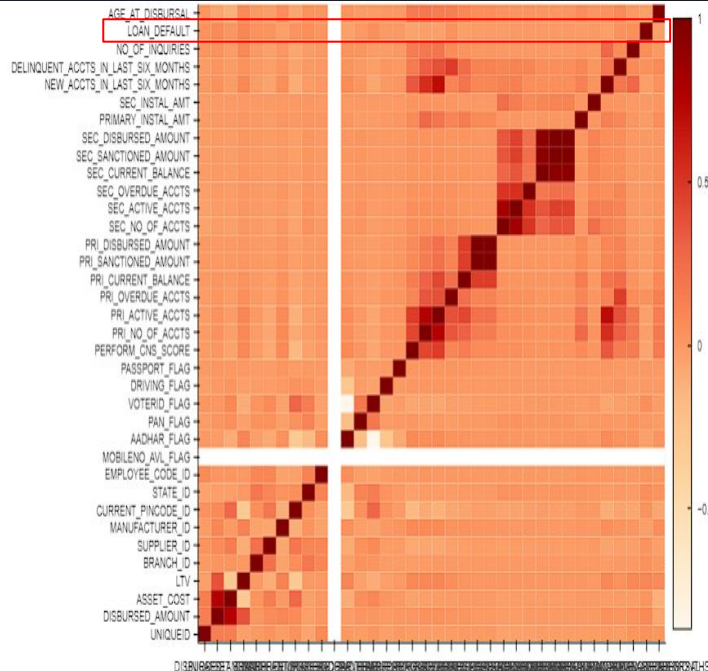
#### Feature Created: “Average\_Acc\_Age\_Total\_Months”

Description: Represents the average length of the customer's loans in months.

- The original columns “**DATE\_OF\_BIRTH**” and “**DISBURSAL\_DATE**”, describe the date the customer was born and the date the customer received the loan. We can determine a customer's age using the date of birth which could be very useful for our model along with the disbursal date. We converted both columns as datetime.

#### Feature Created: “Customer\_age”

Description: Represents the age of the customer at disbursal.



A lot of our data was heavily skewed with zeros due to poor recordkeeping

# Modeling Goals & Initial Results

- Goal is to over-predict which loans will default while minimizing false positives
- Recall was chosen as the primary measure as we are looking for the most true positives and least false negatives

<u>Model Type</u>	<u>Precision</u>	<u>Recall</u>	<u>F1-Score</u>	<u>Accuracy</u>
Logistic Regression	0.14	0.00	0.00	0.78
Random Forest	0.26	0.28	0.27	0.67
Neural Network	0.50	0.01	0.02	0.78

# Final Model

- Numerous attempts to model different layers with additional nodes did not have a significant improvement on Recall
- The most significant impact was by varying the threshold on the final layer's sigmoid function. Initial values at 0.5 were very poor recall, but a threshold of 0.2 ended up balancing recall for positives and negatives.

<u>Model Type</u>	<u>Precision</u>	<u>Recall</u>	<u>F1-Score</u>	<u>Accuracy</u>
Neural Network	0.28	0.69	0.40	0.55

# Indian to International Place Value

Indian Place Value Chart		
PERIODS	FIGURES	DIGITS
CRORES	Ten Crores (TC) 10,00,00,000	9
	Crores (C) 1,00,00,000	8
LAKHS	Ten Lakhs (TL) 10,00,000	7
	Lakhs (L) 1,00,000	6
THOUSANDS	Ten Thousands (TTh) 10,000	5
	Thousands (Th) 1,000	4
ONES	Hundreds (H) 100	3
	Tens (T) 10	2
	Ones (O) 1	1

International Place Value Chart		
PERIODS	PLACES	DIGITS
MILLIONS	Hundred Millions (HM) 100,000,000	9
	Ten Millions (TM) 10,000,000	8
	Millions (M) 1,000,000	7
THOUSANDS	Hundred Thousands (HT) 100,000	6
	Ten Thousands (TTh) 10,000	5
	Thousands (Th) 1,000	4
ONES	Hundreds (H) 100	3
	Tens (T) 10	2
	Ones (O) 1	1

# Closing Remarks

- We plan to extend this work to longer-period data sets and tighten up the prediction of loan defaults.
- We plan to market this model to auto loan providers at a subscription rate.
- Loan providers would subscribe on a tiered basis depending on their loan origination plans:

<u>Loan Origination Goals</u>	<u>% Rate</u>	<u>Subscription Length</u>
U to \$1MM - \$100MM	1.75 % of Yearly Origination	Half-years
\$1MM - \$100MM	1.50 % of Yearly Origination	Half-years
\$100MM - \$1,000MM	1.00 % of Yearly Origination	Full years

- Opex budget for the company: \$1MM/yr primarily in manpower

# Thank you!

Dataset reference:

<https://www.kaggle.com/datasets/avikpaul4u/vehicle-loan-default-prediction?select=test.csv>

Slide template reference:

<https://slidesgo.com/theme/automotive-industry-consulting>

**Indian Place Value Chart**

<https://www.cuemath.com/numbers/indian-place-value-chart/>

CIBIL scores

<https://www.bajajfinserv.in/insights/everything-you-need-to-know-about-your-credit-score>

**Auspicious Dates for Vehicle Purchase in 2023**

[https://www.dildarnagar.com/2023/11/vehicle-purchase-muhurat.html#:~:text=Dhanteras%20\(November%2010%2C%202023\),auspicious%20occasion%20for%20buying%20automobiles.](https://www.dildarnagar.com/2023/11/vehicle-purchase-muhurat.html#:~:text=Dhanteras%20(November%2010%2C%202023),auspicious%20occasion%20for%20buying%20automobiles.)

Diwali

<https://kids.nationalgeographic.com/pages/article/diwali>