

**Final Project**



# **CREDIT RISK LOAN**

---

Meida Rahma Al Kariim

ID/X Partners (Rakamin Academy)



# Hi! I'm Meida!

[linkedin.com/in/meida-rahma/](https://www.linkedin.com/in/meida-rahma/)

---

Meida Rahma Al Kariim is a junior data scientist specializing in transforming complex datasets into actionable insights for informed decision-making.

Previous project:  
[Speeding to Victory: A Predictive Modeling Approach in Formula 1](#)

## Technical Skills:

- Data Analysis
- Data Visualization
- Machine Learning

# OUTLINE

**01 BUSINESS & DATA  
UNDERSTANDING**

**02 EXPLORATORY DATA  
ANALYSIS**

**03 DATA  
PREPROCESSING**

**04 MODELING**

**05 CONCLUSION**

**06 BUSINESS  
RECOMMENDATION**

**01.**

**BUSINESS &  
DATA  
UNDERSTANDING**

---



# CREDIT RISK



## CREDIT DECISION MAKING

Assisting banks in assessing the ability of customers to repay loans.

---



## CREDIT RISK MANAGEMENT

Banks can take necessary preventive measures, such as setting credit limits and determining appropriate interest rates.

---



## FINANCIAL HEALTH MAINTENANCE

Assisting banks in maintaining their financial health by preventing significant losses that could impact capital and liquidity.

---



## TRUST AND REPUTATION

Helping banks maintain the trust of customers, investors, and other stakeholders.

# ABOUT DATASET

## 2007-2014

The dataset encompasses various information about borrowers who have applied for loans. These data provide valuable insights into the financial profiles of borrowers, credit history, and loan status.

**75**

COLUMN



**45**

COLUMN

**466.285**

ROWS



**437.710**

ROWS

For more info about data dictionary, [click here](#)



## MISSING VALUE

There is 37 columns that has missing value in it.



## CATEGORIC

18 Columns is categorical.

01

02



## DUPLICATED DATA

There is no duplicated data



## NUMERIC

27 Columns is numerical.

# DATA UNDERSTANDING

04

03

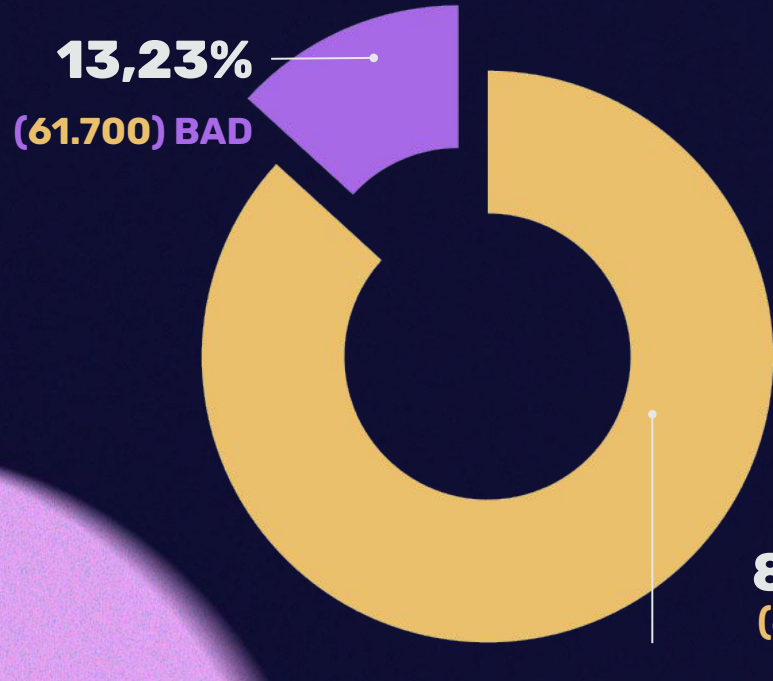
**02.**

# **EXPLORATORY DATA ANALYSIS**

---



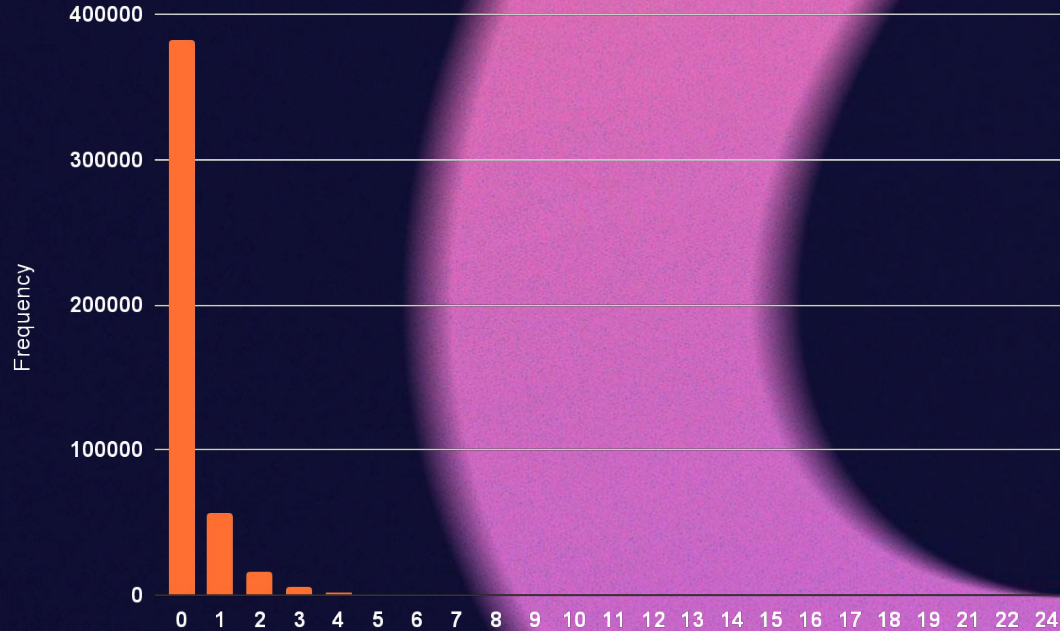
# EDA



## CHECK IMBALANCE

The credit risk ratio is quite imbalanced. Out of 466.285 rows, only 13,23% resulted in bad. Oversampling needs to be performed first to generate balanced data.

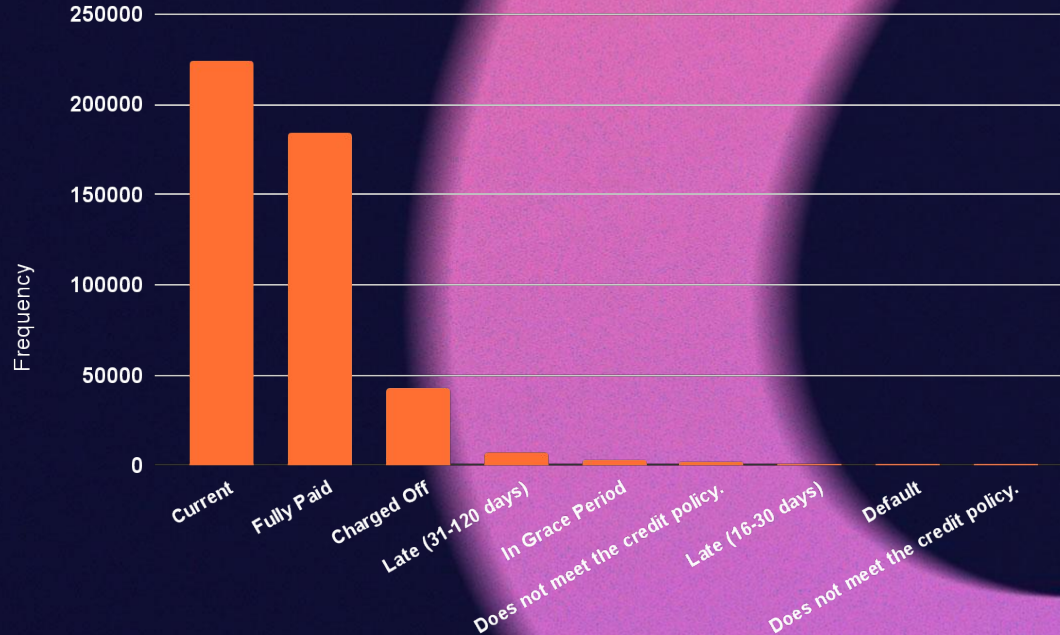
# DELING\_2YRS



DELING\_2YRS which has a late payment history for the past 2 years.

Based on the graph in addition, the majority of customers rarely make late payments. This is a good thing for bank reputation.

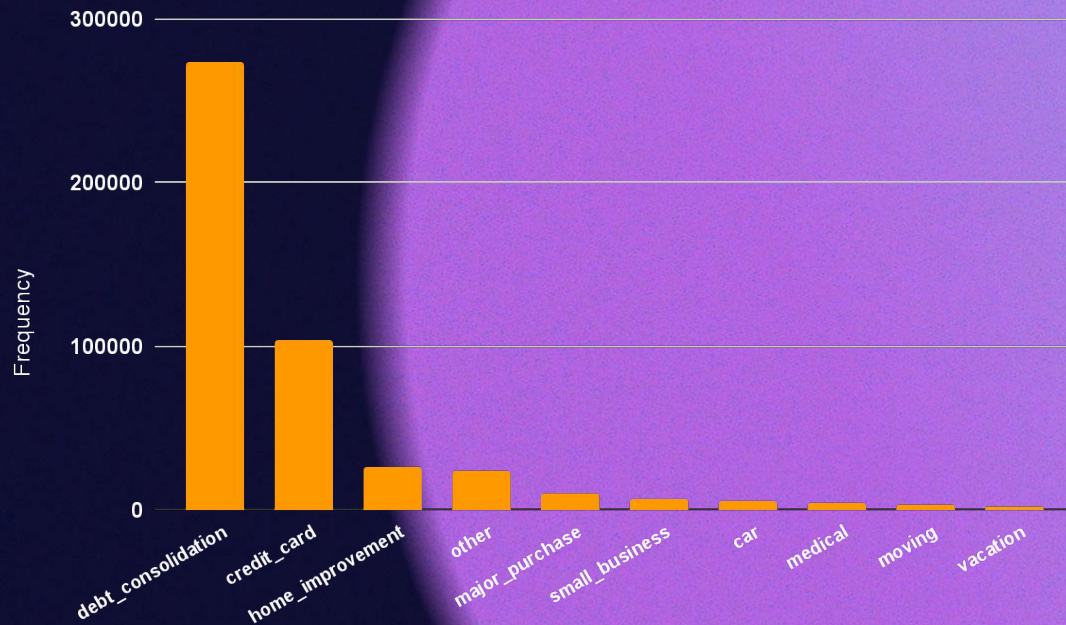
# LOAN STATUS



Based on the graph in addition. The majority of customers are in Current and Fully Paid, this means that many customers can be categorized as Good.



# TOP 10 PURPOSE



NO.	PURPOSE
1	debt_consolidation
2	credit_card
3	home_improvement
4	other
5	major_purchase
6	small_business
7	car
8	medical
9	moving
10	vacation



# TOP COUNTRY

**71.450**

CA - California

**40.242**

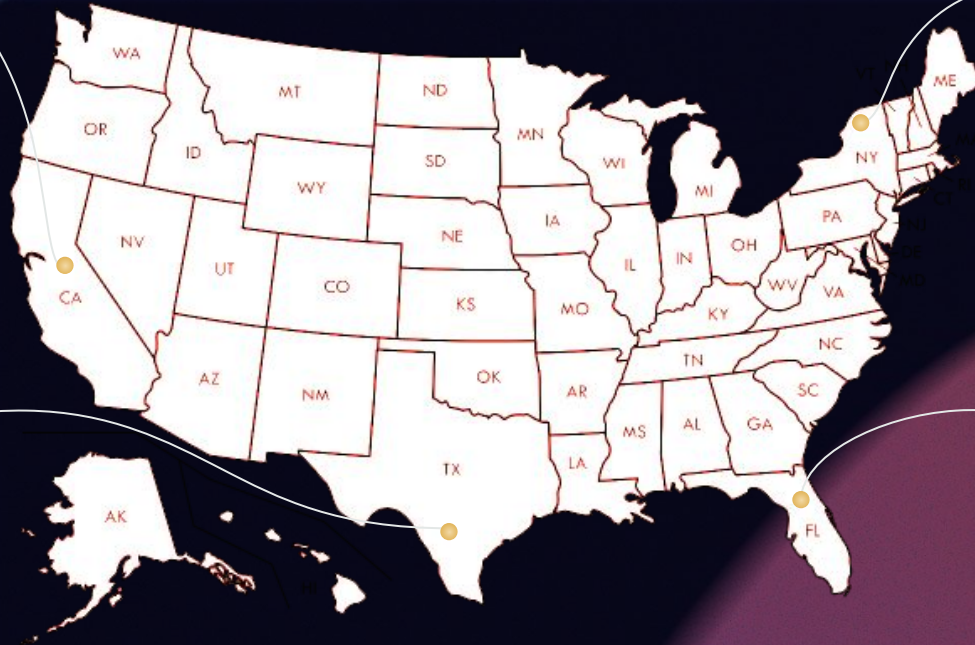
NY - New York

**36.439**

TX - Texas

**31.637**

FL - Florida



The top 4 countries with the highest number of borrowers are all from the United States.



**03.**

**DATA  
PREPROCESSING**

---

# TO DO LIST

**Change data type that matches the contents of the column.**

---

**Train Test Split**

---

**Multicollinearity Study**

**Delete unnecessary columns.**

---

**Handling Imbalance**

---

**Standardization**

**04.**

# **MODEL- ING**

---



# MODELING

Training Data Results:

MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE
Logistic Regression	77.81 %	78,02%	77,47%	78%
Random Forest	94.13 %	90.24 %	98.97 %	94 %
Decision Tree	94,01%	90,67%	98,12%	94%
KNN	95,52%	99,06%	91,92%	95%

Testing Data Results:

MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE
Logistic Regression	77.85 %	78,02%	77,43%	78%
Random Forest	94.14 %	90.26%	98.94%	94%
Decision Tree	94,02%	90,73%	98,02%	94%
KNN	90,73%	96,81%	84,2%	90%

# FINDINGS

## MODELS

## OBSERVATION

**Logistic Regression**

---

The accuracy value of this model is relatively low, primarily due to the presence of outliers in the dataset.

**KNN**

---

The accuracy of this model has decreased by 5%, and the possible cause could be the Euclidean distance not being a suitable metric for the data.

**Decision Tree  
And Random Forest**

---

Based on the accuracy results of both models, Random Forest performs better in predicting credit risk.

**05.**

# **CONCLUSION**

---

# CONCLUSION

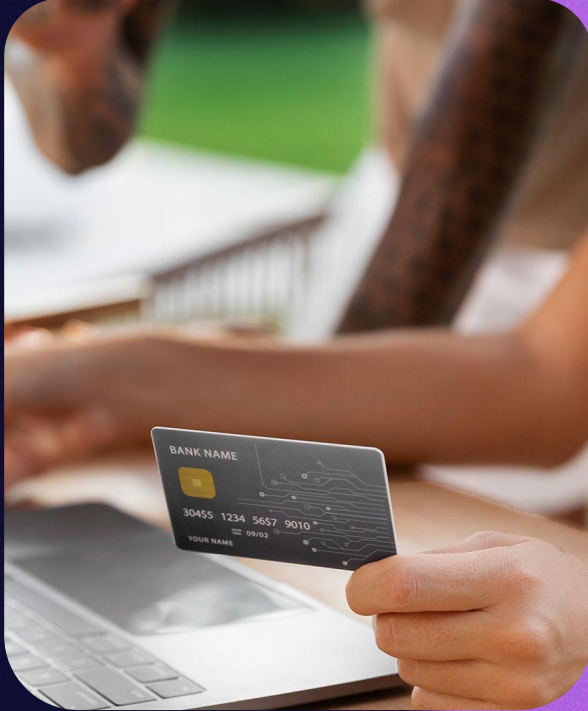
The dataset comprises 466,285 rows and 75 columns containing diverse information about loan applicants. It was observed that 37 columns have missing values. To address this, columns with more than 50% missing values were removed, and for the remaining ones, the missing values were imputed with the median. Following this, data preprocessing steps were undertaken, including adjusting data types, handling imbalanced data, performing Train-Test Split, checking for Multicollinearity, and standardization.

Once the data was prepared, Logistic Regression, Random Forest, Decision Tree, and KNN models were employed. Evaluating the models based on accuracy, as the modeling approach used is classification and the data is balanced, it was concluded that the Random Forest model performed the best in predicting credit risk for customers. Therefore, this model is deemed more adept at credit risk prediction based on the loan dataset.



# **06. BUSINESS RECOMMENDATION**

---



## Appropriate Interest Rate Setting

---

When knowing the credit risk of borrowers, several business recommendations can be provided to manage risk and enhance the financial performance of the company.

Adjust interest rates based on credit risk levels. Borrowers with high credit risk may be subject to higher interest rates to compensate for the risk.

A close-up photograph of a hand with light-colored nail polish pouring several US dollar bills into a clear glass jar. The bills are fanned out as they fall into the jar. The background is dark, and the scene is framed by soft, abstract shapes in shades of pink and purple.

# Continuous Monitoring

Continuously monitor borrowers with high credit risk. Effective monitoring enables the company to promptly identify signs of financial difficulties and take necessary actions.





## Development of Additional Services

For borrowers with high credit risk but still presenting good business potential, consider offering additional services such as insurance or debt recovery programs.



# Financial Education

Provide financial education programs for borrowers with high credit risk to help them improve personal financial management and reduce future credit risks.



# Collaboration with Other Financial Institutions

---

Consider collaborating with other financial institutions, such as credit associations or credit data providers, to gain further insights into borrowers and strengthen credit risk management policies.



# THANKS!



XXX

**Do you have any questions?**

meidarahma1105@gmail.com

+6281284848236

[linkedin.com/in/meida-rahma/](https://www.linkedin.com/in/meida-rahma/)

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon** and infographics & images by **Freepik**