Insights into Lending Risks

Optimizing Loan Approvals Through Identification Risky Applicant

Amaan Shaikh
Senior Analyst at Tiger Analytics

Table of content

Background

Company:

Largest online loan marketplace facilitating personal and business loans.

Challenge:

High financial losses due to loan defaults.

Focus:

Analyze historical loan data (2007–2011) to identify default risky applicants.

Business Problem

Problem:

Financial losses caused by loan defaults.

Solution:

Need for identifying risky loan applicants during the approval process.

Dataset Overview

loan.csv: Historical loan data (2007–2011).

Data_Directory.xlsx: Variable descriptions.

Key Features: Loan amount, interest rate, annual income, credit history, etc.

loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	v
5000	5000	4975.00	36 months	10.65%	162.87	B	B2	NaN	10+ years	RENT	24000.00	
2500	2500	2500.00	60 months	15.27%	59.83	C	C4	Ryder	< 1 year	RENT	30000.00	
2400	2400	2400.00	36 months	15.96%	84.33	C	C5	NaN	10+ years	RENT	12252.00	
10000	10000	10000.00	36 months	13.49%	339.31	C	C1	AIR RESOURCES BOARD	10+ years	RENT	49200.00	
3000	3000	3000.00	60 months	12.69%	67.79	B	B5	University Medical Group	1 year	RENT	80000.00	

Objectives

Conduct exploratory data analysis (EDA) to:

- ➔ Identify strong indicators of loan default.
- ➔ Highlight potential risky applicants.
- ➔ Provide actionable recommendations.

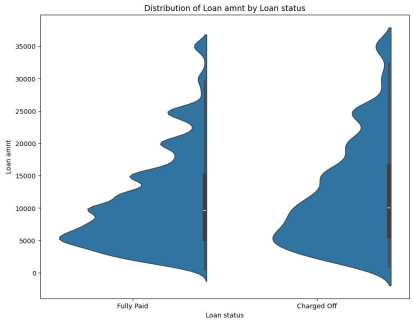
Data Cleaning & Preprocessing

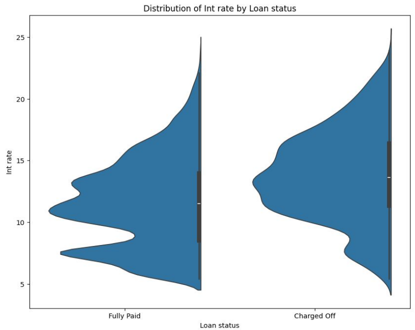
- Handled missing values and outliers.
- Applied log transformations to reduce skewness.
- Engineered features: `credit_history_length`, `time_since_last_payment`, etc.

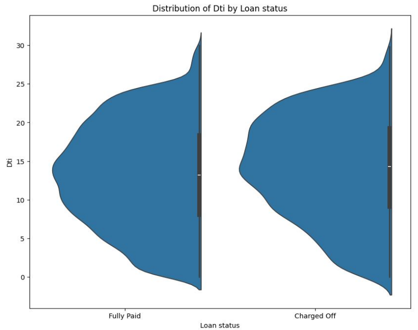
Key Insights –
Numerical Variables (Part-1)

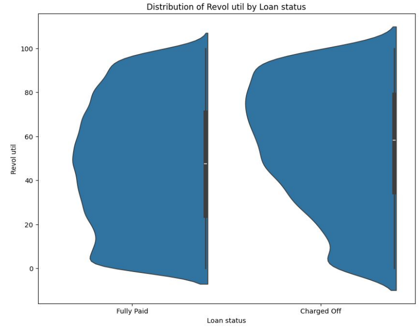
Risky Applicant Indicators:

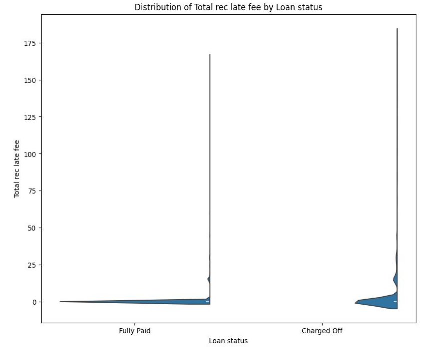
- 🔼 High loan amounts[Loan amnt]
- 🔼 High interest rates [int rate].
- 🔼 High debt-to-income ratio[dti]
- 🔼 High revolving utilization rate[Revol util].
- 🔼 High count of late fee received to date[Total rec late fee]
- 🔼 High day count since last payment[Time since last payment]
- 🔽 Low annual income [Annual inc]
- 🔽 Low short credit history [Credit history length]
- 🔽 Less day count since last credit pull [Time Since last credit pull]

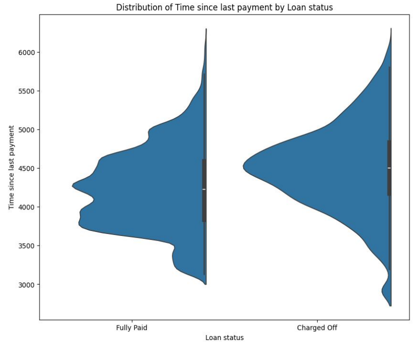


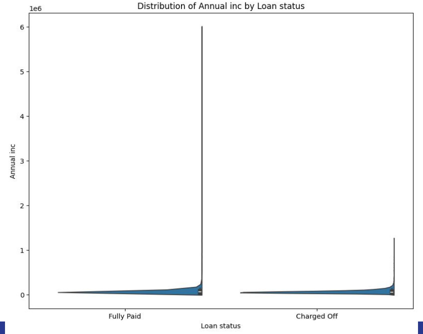


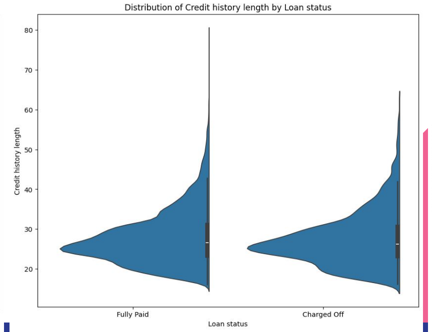


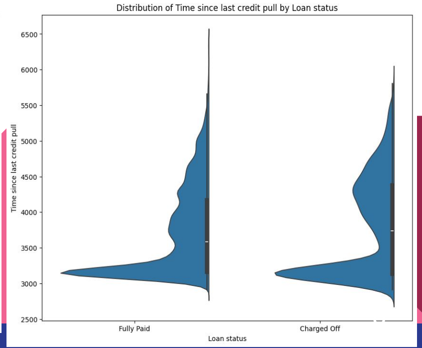








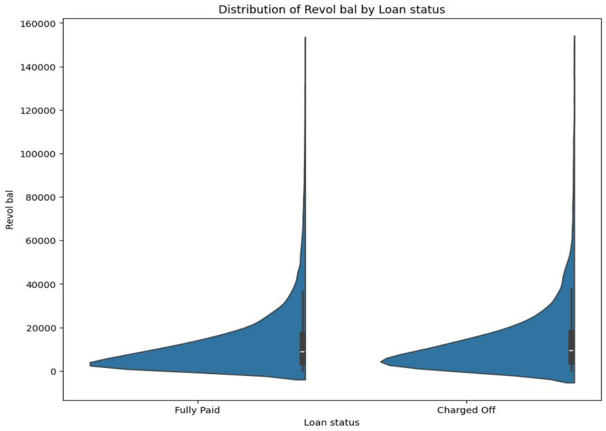


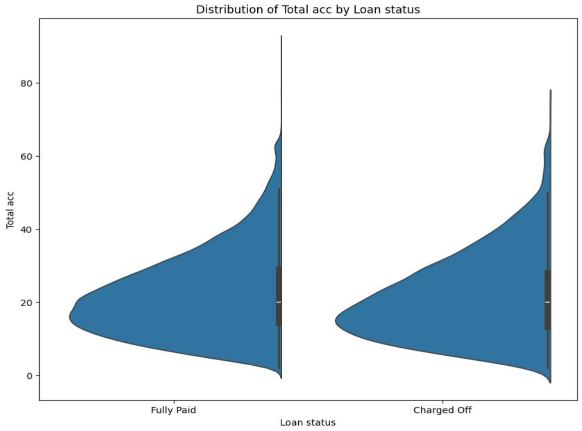


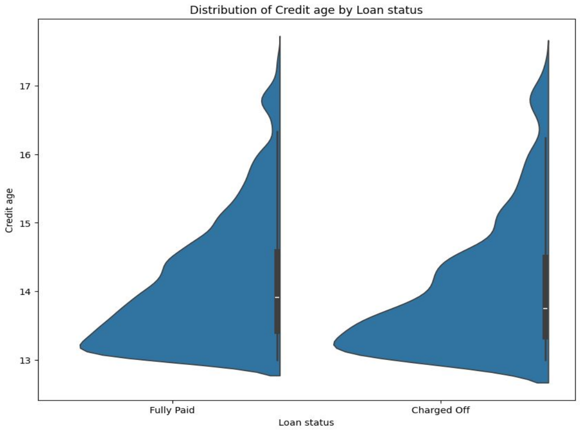
Key Insights –
Numerical Variables (Part-2)

Factors with Minimal or No Impact visible on charts:

★ Total credit revolving balance
★ Total number of credit lines currently in borrower's credit file
★ Total number of years borrower's credit history present



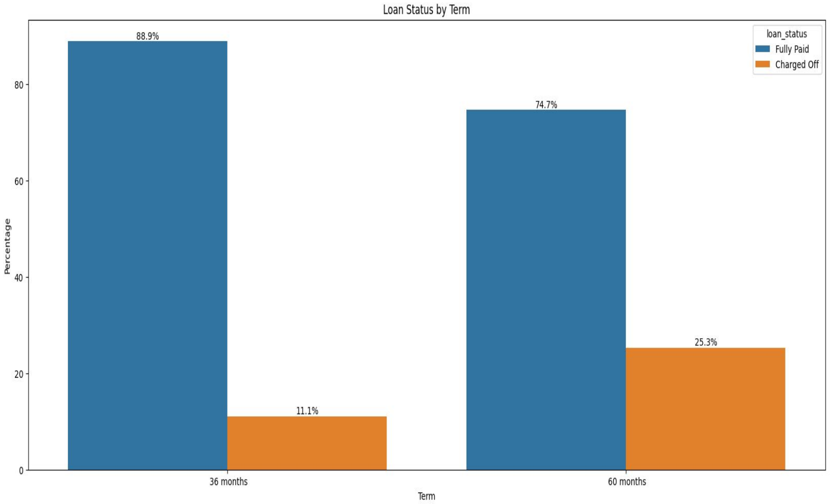


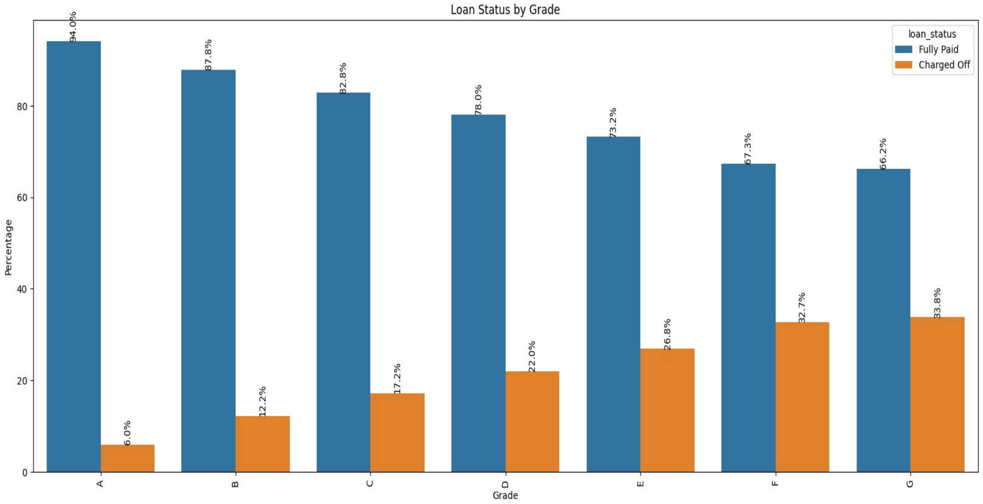


Key Insights –
Categorical Variables (Part-1)

Risky Indicators:

- Loans with a 60-month term have a significantly higher default percentage.
- As grade value increasing, default percentages increase, indicating a strong correlation with risk.

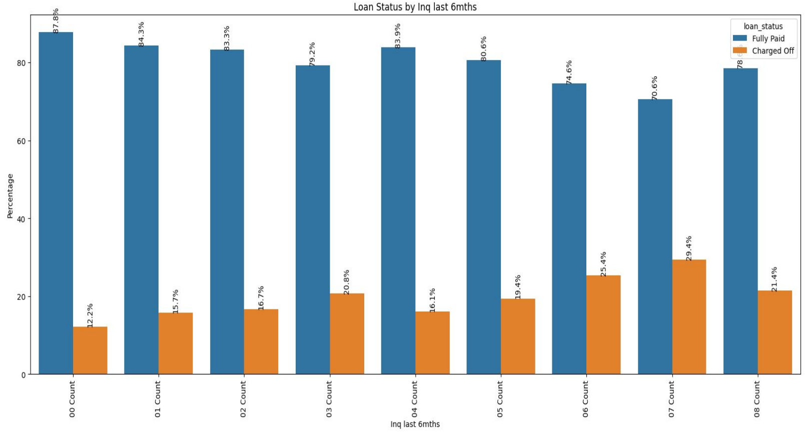


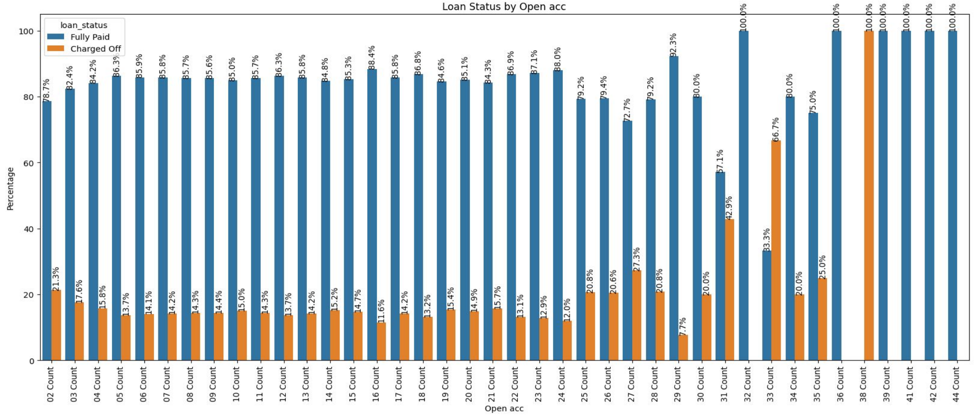


Key Insights –
Categorical Variables (Part-2)

Risky Indicators:

- Higher inquiries in the last 6 months correlate with increased default percentages.
- Higher number of open accounts is associated with increased default risk.

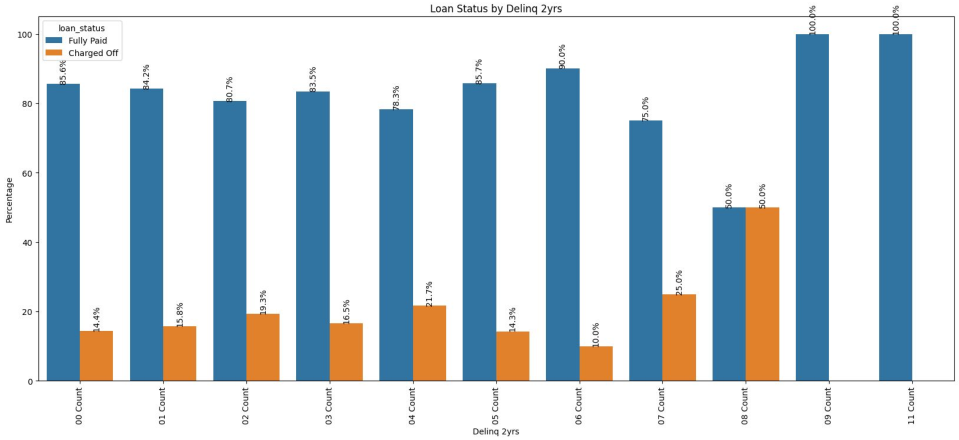


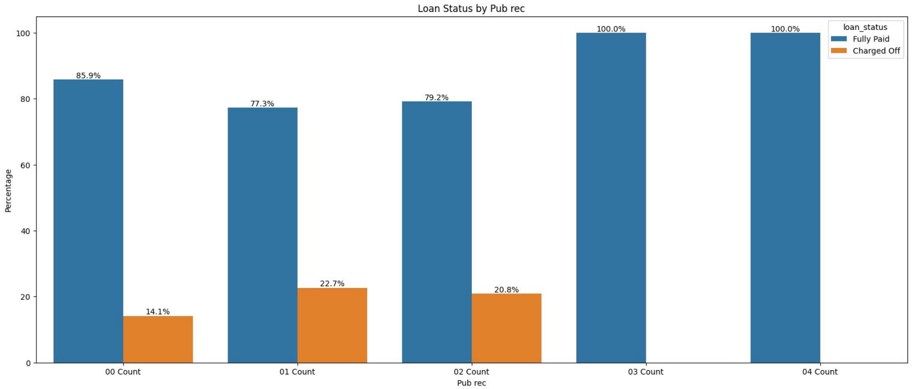


Key Insights –
Categorical Variables (Part-3)

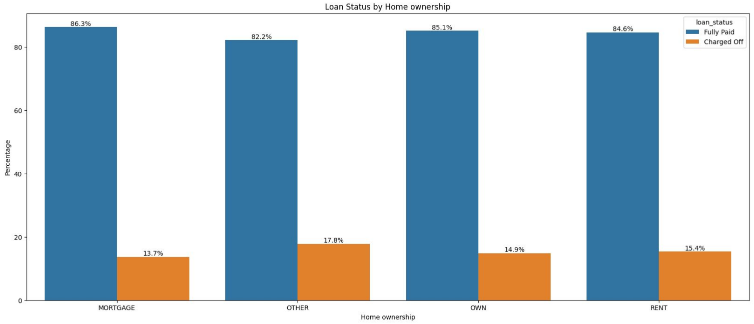
Factors with Minimal Impact visible on charts:

- Generally insignificant, but higher 30+days past due incidences of delinquency counts may indicate potential risk.
- Public records have minimal impact on loan status.



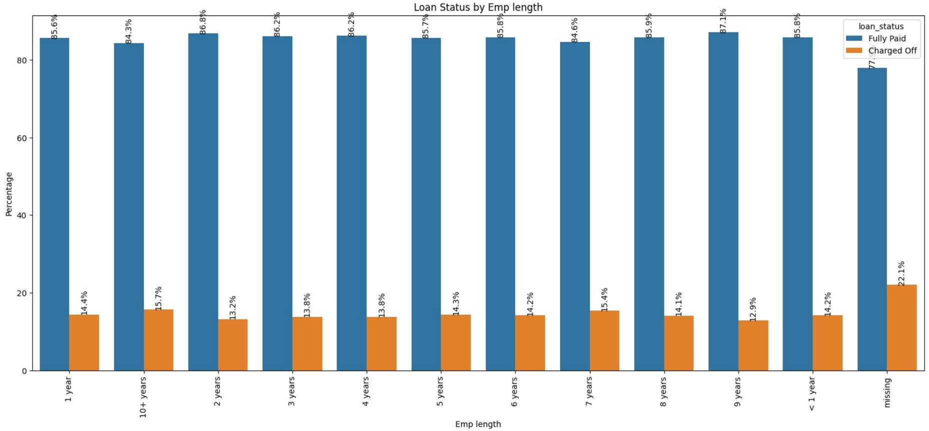


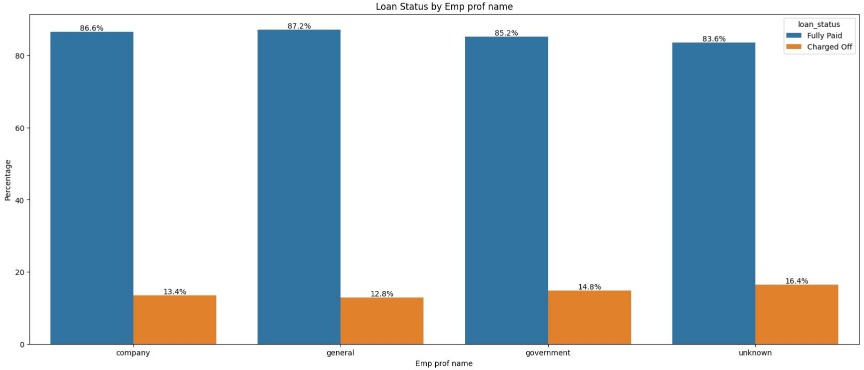
Key Insights –
Categorical Variables (Part-4)



Factors with No Impact visible on charts:

- No significant impact of Employment length, though missing data shows a slight increase in default percentage.
- No notable differences in default percentages based on home ownership.
- Profession names show no strong relationship with default risk.

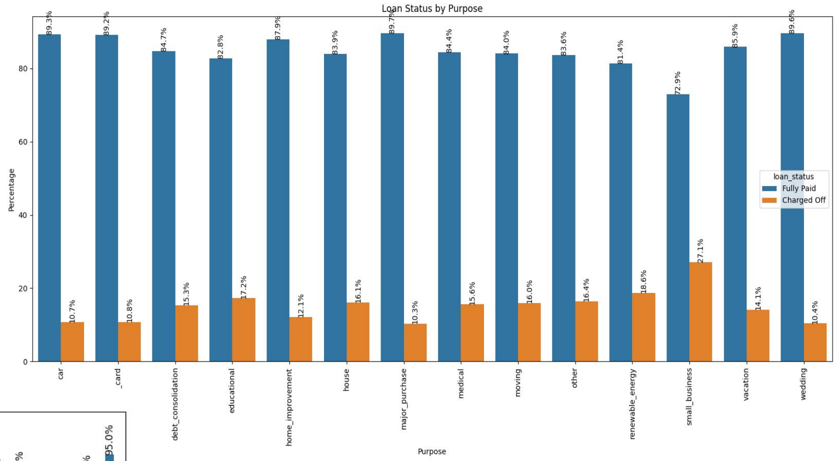


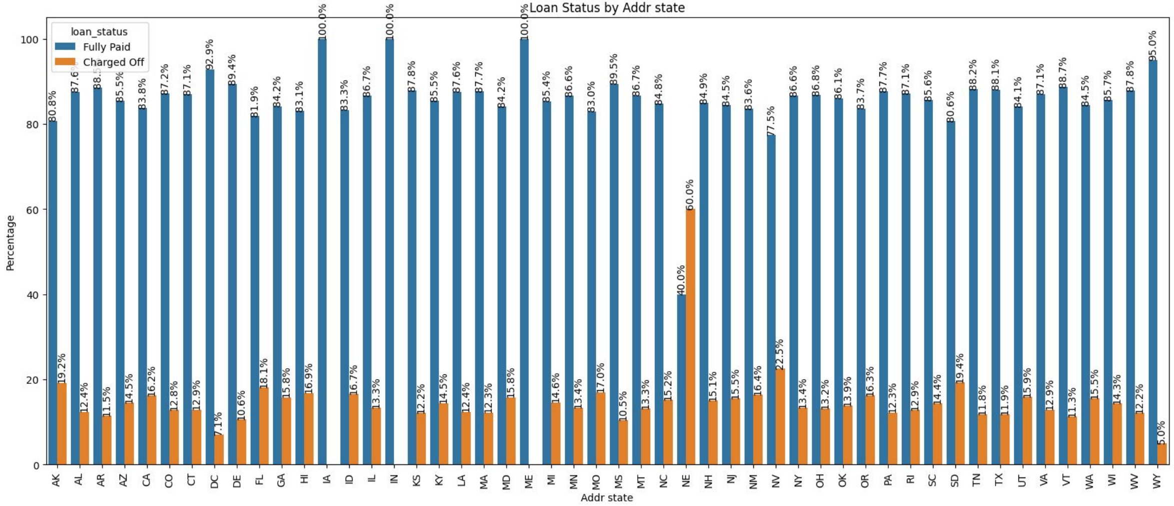


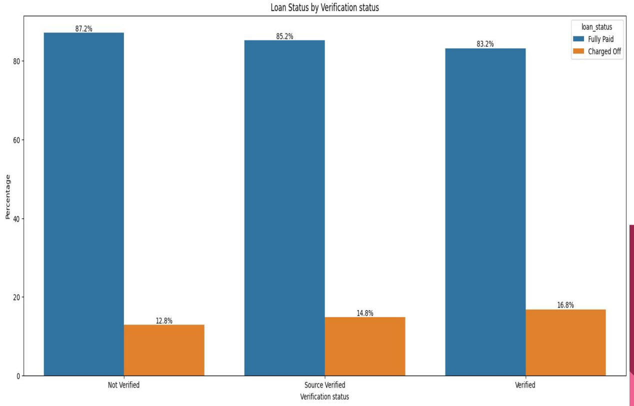
Indirect Impact Factors

Factors with No Direct Impact but Possible Indications:

- Loans with a 60-month term have significantly higher default percentages, indicating a potential need to assess risk more rigorously for longer-term loans.
- Verified applicants show a higher default percentage, suggesting potential flaws in the verification process.
- Loans for small businesses have the highest default rates, indicating a need for stricter background checks for certain purposes.
- Nebraska (NE) shows a 60% default rate, suggesting validation processes vary by state and may require improvement.







Recommendations

- Revise credit approval policies for:
  - **High** loan amounts or interest rates or DTI ratio.
  - **Higher number** of open accounts or inquiries
  - **Short** credit history
  - Longer-term loans
- Strengthen verification processes.
- Strengthen Data collection like, Employee length (Employment experience year)
- Implement stricter checks for:
  - Small business loans.
  - High-risk states like Nebraska.

Challenges & Assumptions

🍕Challenges:

- Balancing outliers vs. noise.
- Interpreting categorical variables effectively.

💡Assumptions:

- Data accurately represents past trends.
- Retained outliers to capture risk-related anomalies.

Conclusion

- ➔ Identified key drivers of loan default.
- ➔ Provided actionable insights to minimize credit risk.
- ➔ Enhanced decision-making for loan approvals.

GitHub Repository

Link: insights-into-lending-risks

Reference: https://github.com/amaan2398/insights-into-lending-risks

Thank You