

Loan Default Analysis

Analysis of Lending Club Loan Data

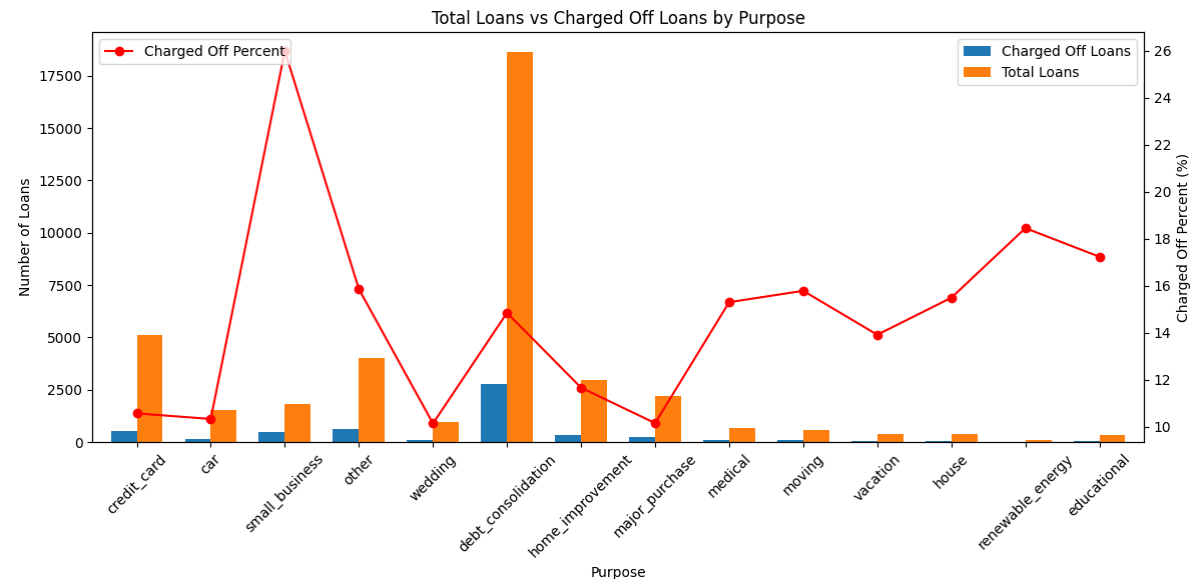
Dataset Overview

- Dataset contains loan information with mixed data types
- Key columns include loan amount, interest rate, employment length, purpose
- Data preprocessing included handling missing values and data type conversions
- Target variable: loan_status (Fully Paid vs Charged Off)
- Focus on identifying factors that predict loan defaults

Loan Purpose vs Default Rates

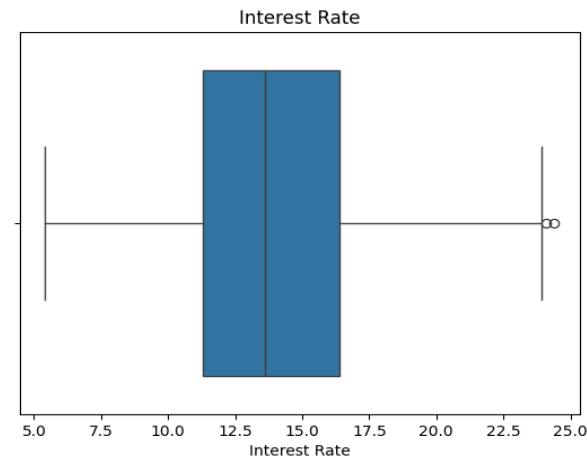
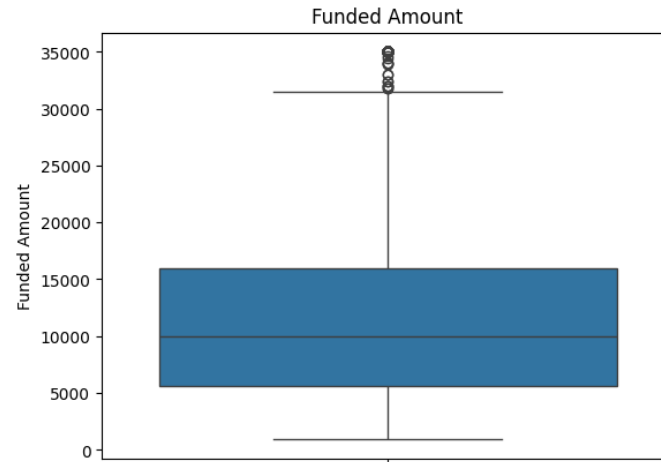
- Key Findings from Purpose Analysis:
- High-Risk Loan Purposes:
 - **Small Business** loans have the **HIGHEST** default rate: 25.98%
 - Renewable Energy loans: 18.45% default rate
 - Educational loans: 17.23% default rate
 - Other high-risk purposes: Moving (15.78%), Medical (15.30%)

- Low-Risk Loan Purposes:
 - Wedding loans: 10.14% default rate
 - Major Purchase: 10.15% default rate
 - Car loans: 10.33% default rate
 - Credit Card: 10.57% default rate



Key Statistical Insights

- Loan Amount Distribution:
 - Mean loan amount: \$11,753.40
 - Standard deviation: \$7,760.52
 - 25th percentile: \$5,575
 - 50th percentile (median): \$10,000
 - 75th percentile: \$16,000
- Interest Rate Distribution:
 - Mean interest rate: 13.82%
 - Standard deviation: 3.65%
 - 25th percentile: 11.31%
 - 50th percentile (median): 13.61%
 - 75th percentile: 16.40%

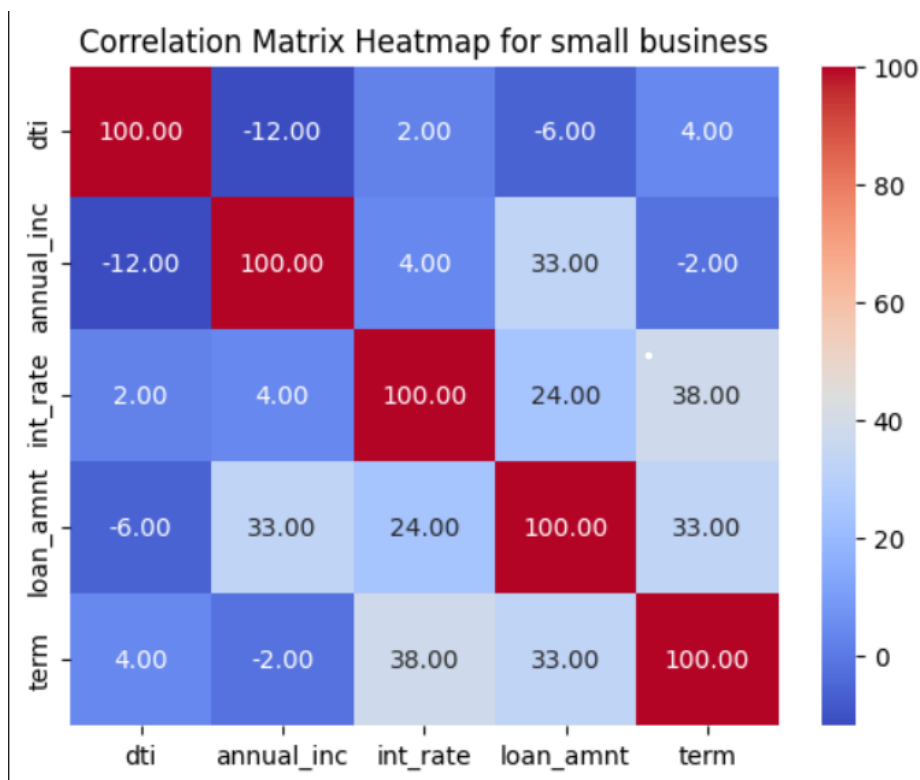


Correlation matrix on Small Business

- Top 3 'Clustered' Variables on charged off customers for 'Small Business' purpose

- Term
- Interest rate
- Loan amount

- Combination of above 3 variables have significant impact in “defaulting”



Data Quality & Processing Steps

Data Processing Pipeline:

1. Data Loading & Inspection:

- Mixed data types detected
- Initial exploration of loan status distributions

2. Missing Value Treatment:

- Removed rows with missing loan_status values
- Handled missing values in numerical columns systematically

3. Feature Engineering:

- Converted interest rate from percentage string to float and removed trailing '%' symbol
- Created binary target variable (0=Fully Paid, 1=Charged Off)
- Selected relevant numerical features for analysis

Business Implications & Recommendations

Risk Assessment Recommendations:

HIGH RISK Categories (Require Enhanced Due Diligence):

- Small Business loans (26% default rate) - Implement stricter criteria
- Renewable Energy & Educational loans (17-18% default)
- Moving & Medical emergency loans (15%+)

Lending Strategy:

- Focus on low-risk segments: Wedding, Car, Major Purchase loans (10%)
- Enhanced screening for borrowers seeking high-risk loan purposes

Conclusions & Next Steps

Key Finding Summary:

- ✓ Small business loans represent highest risk segment (26% default rate)
- ✓ Loan purpose significantly impacts default rates (10% to 26% range)

Actions to be Undertaken:

1. Implement risk-based pricing by loan purpose
2. Enhanced due diligence for high-risk categories