Credit Risk Analysis for Loan Default Assessment

(A Data Analytics Project using Python, Excel, and Statistical Analysis)

Prepared by:
Jiya Chouksey
B.Tech Computer Science & Engineering
Tools Used:
Python, Pandas, NumPy, Excel, Statistics
Date:
(18 January 2026)

➢ INTRODUCTION

Credit risk analysis plays a crucial role in the financial sector by helping institutions assess the likelihood of loan default by borrowers. Effective risk assessment enables banks and financial organizations to minimize losses while making informed lending decisions. This project focuses on analyzing historical loan and customer data to identify key factors associated with loan default. Using data analytics and statistical techniques, the objective is to understand borrower behavior and highlight risk patterns that can assist financial institutions in improving credit evaluation strategies.

➢ PROBLEM STATEMENT & OBJECTIVES

Problem Statement: Financial institutions face challenges in identifying high-risk borrowers before approving loans. Poor credit assessment can lead to increased default rates and financial losses.
The goal of this project is to analyze loan and borrower data to determine which factors contribute most to loan defaults and to segment customers based on risk levels.
Objectives:

- To analyze the distribution of loan defaults
- To study the relationship between borrower attributes and default risk
- To identify key risk indicators using statistical analysis
- To segment customers into risk categories
- To provide data-driven recommendations for credit risk management

➢ DATASET OVERVIEW

Dataset Description: The dataset used in this project contains **32,581 loan records** with 12 attributes related to borrower demographics, loan characteristics, and credit history.
Kaggle: Credit-Risk-dataset
Key features include:

- Age
- Annual income
- Employment duration
- Loan amount
- Interest rate
- Credit grade
- Loan purpose
- Credit history length
- Past default record
- Loan default status (target variable)
  (The target variable default indicates whether a borrower defaulted on a loan.)

➢ DATA CLEANING & PREPARATION

Data Cleaning and Preparation: Before analysis, the dataset was cleaned and prepared to ensure accuracy and consistency.
Steps Performed:

- Renamed columns for better readability and clarity
- Identified missing values in employment duration and interest rate
- Handled missing values using **median imputation** to avoid skewness

- Verified data consistency and validated value ranges

Some extreme values (such as very high age or employment duration) were observed. Since these values represented a very small proportion of the dataset and the objective was descriptive analysis, they were retained.

➢ EXPLORATORY DATA ANALYSIS

Default Distribution:

Table: Loan Default Distribution

Default Status	Percentage
Non-Defaulter (0)	78.18
Defaulter (1)	21.82

- The dataset shows a typical class imbalance, with approximately 22% of borrowers defaulting on loans.

Income vs Default

Table: Average Annual Income by Default Status

Default Status	Avg Annual Income
Non-Defaulter (0)	70,804
Defaulter (1)	49,126

- Defaulters have a **significantly lower average income**, indicating that income is a strong indicator of repayment capability.

Loan Amount vs Default

Table: Average Loan Amount by Default Status

Default Status	Avg Loan Amount
Non-Defaulter (0)	9237
Defaulter (1)	10851

- Borrowers who defaulted generally took **higher loan amounts**, suggesting increased repayment burden.

Credit History vs Default

Table: Average Credit History Length

Default Status	Avg Credit History (Years)
Non-Defaulter (0)	5.84
Defaulter (1)	5.69

- Defaulters tend to have **slightly shorter credit histories**, indicating limited borrowing experience.

Past Default Impact

Table: Default Rate by Past Default History (%)

Past Default	Non-Defaulter (%)	Defaulter (%)
No	81.61	18.39
Yes	62.19	37.81

- Borrowers with a history of default are **more than twice as likely to default again**, making past behavior a critical risk factor.

Loan Purpose Risk
Table: Default Rate by Loan Purpose (%)

Loan Purpose	Default Rate (%)
Debt Consolidation	28.59
Medical	26.70
Home Improvement	26.10
Personal	19.89
Education	17.22
Venture	14.81

- Certain loan purposes such as **debt consolidation and medical loans** show higher default rates, indicating purpose-based risk variation.

➢ CATEGORICAL RISK ANALYSIS

Past Default Impact:
Borrowers with a history of past defaults showed a **significantly higher probability of default**, making it one of the strongest risk factors.
Loan Purpose Risk:
Loan default rates varied by loan purpose. Certain categories such as debt consolidation and medical loans exhibited higher default risk compared to education or venture loans.

➢ STATISTICAL INSIGHTS

Correlation analysis revealed:

- A **positive correlation** between loan-to-income ratio and default risk
- A **negative correlation** between income and default
- Interest rate showed moderate association with default behavior

These findings suggest that borrower leverage and affordability play a key role in credit risk.

➢ RISK SEGMENTATION

Borrowers were segmented into risk categories based on **loan-to-income ratio**:

- **Low Risk:** Lowest default rate
- **Medium Risk:** Moderate default rate
- **High Risk:** Significantly higher default rate (~40%)

This segmentation provides a practical framework for identifying high-risk applicants.

➢ BUSINESS INSIGHTS & RECOMMENDATIONS

- Borrowers with prior defaults should undergo stricter verification
- High loan-to-income ratio applicants should be closely evaluated
- Longer credit history indicates lower default risk
- Certain loan purposes may require higher interest margins or additional checks

Implementing these insights can help financial institutions reduce default rates and improve loan portfolio quality.

- **CONCLUSION**

  This project demonstrates how data analytics and statistical techniques can be used to assess credit risk effectively. By analyzing borrower and loan characteristics, key risk drivers were identified and practical recommendations were proposed.

  The analysis highlights the importance of income, loan size, credit history, and prior defaults in predicting loan repayment behavior.