

Table of Contents

1. Title Page
2. Abstract
3. Introduction
4. Literature Review
5. Methodology
6. Results
7. Discussion
8. Conclusion
9. References

Banking Case Analysis: Risk Analysis in Financial Services

Author: Navdeep Maurya

Date: April 6, 2025

Abstract

This project investigates risk analysis in banking and financial services, with a specific focus on identifying high-risk lending patterns and improving credit risk management through data analysis. The objective is to understand how customer-related financial indicators influence loan repayment behavior and how banks can utilize this insight to minimize financial loss.

- **Purpose:** To analyze loan and customer data to identify trends that contribute to loan defaults and to recommend data-driven strategies to minimize risk during the loan approval process.
- **Approach:** Exploratory Data Analysis (EDA) was performed using tools such as Google Colab and Jupyter Notebook. The analysis focused on variables like income, credit score, loan amount, employment type, and purpose of the loan. Additionally, MySQL Server was used to manage the dataset, and Power BI was integrated for dashboard creation and visualization.
- **Findings:** Strong correlations were observed between credit score, income stability, and the likelihood of loan default. Salaried individuals and those with higher credit scores were less likely to default. Certain loan purposes, such as business and weddings, had a higher risk of non-repayment.
- **Outcome:** The findings can be used to refine lending policies, develop credit risk prediction models, and enhance decision-making systems, ultimately contributing to safer and more profitable banking operations.

Introduction

Background Information

In the modern banking environment, financial institutions face growing risks when issuing loans. A major challenge is assessing the creditworthiness of borrowers to minimize the probability of loan default. With increasing digital data availability, banks now rely on data science techniques to make informed lending decisions.

Objectives

The project aims to:

- Analyze customer data to identify high-risk patterns in loan disbursement.
- Understand how factors like income, employment, loan amount, and credit score affect repayment.
- Explore data-driven methods for risk minimization in banking.

Scope

The analysis is limited to the given dataset, representing a sample of customer profiles with attributes related to financial status and loan outcomes. The scope covers data cleaning, exploration, visualization, and basic interpretation, without building predictive models.

Literature Review

Risk analysis in banking has traditionally relied on statistical credit scoring models. However, with the growth of machine learning and business intelligence tools, more dynamic methods are now used. Studies show that historical repayment behavior, income stability, and credit scores are strong indicators of loan performance. Integration of SQL databases with analytical tools like Power BI and Jupyter has become common for real-time dashboards and customer profiling.

Methodology

Data Source and Tools

- Data Source: Customer financial records and loan data (CSV format)
- Tools Used: Jupyter Notebook, Google Colab, MySQL Server, Power BI

Data Processing Steps

1. Loading & Cleaning Data: Handled null values, checked data types, standardized formats.
2. Exploratory Data Analysis:
 - Univariate and bivariate analysis of features like credit score, employment status, income, loan amount.
 - Visualized patterns using histograms, boxplots, correlation matrices, and bar charts.
3. Insights Extraction:

- Identified trends in loan default vs non-default classes.
- Used SQL and Power BI to generate dynamic charts for stakeholder presentation.

Results

Key Observations:

- Credit Score: Customers with lower credit scores had a significantly higher default rate.
- Employment Type: Salaried and government employees were more likely to repay loans on time than self-employed individuals.
- Loan Amount: Higher loan amounts correlated with increased default risk when not backed by higher income.
- Purpose of Loan: Loans for business expansion and weddings showed relatively higher default rates.
- Income Bracket: Default likelihood was lower for customers with a stable and high monthly income.

Discussion

The analysis confirms well-established financial risk patterns. Credit scores, income, and employment type are critical indicators for evaluating loan risk. These findings align with existing theories in financial literature. The use of SQL-based backend with Power BI provided seamless integration and helped in creating actionable dashboards.

From a business standpoint, these insights can be used to:

- Refine eligibility criteria for loans
- Offer personalized loan schemes based on customer risk profile
- Set up early warning systems for potential defaulters

Conclusion

The project successfully demonstrates how data can be used to analyze and minimize risk in lending operations. By understanding the behavioral and financial attributes of borrowers, banks can improve their lending policies and reduce financial losses. Future work may include building predictive models (e.g., logistic regression, decision trees) to automate credit risk scoring and integrate them into real-time systems using APIs and dashboards.

References

1. "Credit Risk Modelling" by Basel Committee on Banking Supervision
2. Articles from Kaggle and Towards Data Science on risk analytics
3. Documentation of Power BI, MySQL, and pandas library
4. Internal Data Source: Banking_data.csv