

Lending Club Loan Report

Executive Summary

This report details the end-to-end data analysis process, from cleaning and standardization to strategic insights. The objective was to transform raw loan data into actionable intelligence, validating risk models and identifying key drivers of business growth.

Phase 1: Data Preparation & Cleaning

To ensure data integrity and enable accurate modeling, the following transformation steps were executed:

- **Standardization**
 - **Term to Numeric:** The term column was converted to numeric format to enable direct mathematical comparisons between 36-month and 60-month loan durations.
 - **Date Formatting:** The issue_d column was standardized to a Month-Year-Day (MYD) format to enable time-series analysis for tracking volume trends over specific periods.
- **Categorization**
 - **Home Ownership:** Grouped into four primary categories (MORTGAGE, OWN, RENT, OTHER) to identify which living situations correlate with higher financial stability.
 - **Verification Status:** Simplified into "VERIFIED" vs "NOT VERIFIED" to allow for a direct comparison of default rates between vetted and non-vetted borrowers.
 - **Annual Income:** Bucketed into "Low" (<\$50k), "Medium" (\$50k–\$100k), and "High" (>\$100k) to determine if specific income thresholds act as a safety net against default.
- **Data Integrity**
 - **Cleaning Variables:** Extra symbols were stripped from the emp_length column to treat employment length as a quantitative variable for risk modeling.

- **Handling Nulls:** Blank entries in emp_title were marked as "NA" to ensure dataset completeness without breaking the analysis.
 - **De-duplication:** Duplicate records were removed to ensure KPI figures like Total Loan Volume were accurate and not double-counted.
-

Phase 2: Key Performance Indicators (KPIs)

The health of the portfolio is monitored through the following high-level metrics:

- **Total Loan Volume:** The overall quantity of loans issued.
 - **Portfolio Risk Rate:** The percentage of the portfolio currently at risk of default.
 - **Average Interest Rate:** The mean interest rate charged across the entire portfolio.
 - **Average Customer Income:** The mean annual earnings of the borrower base.
-

Phase 3: Analytical Findings & Conclusions

1. Risk Prediction by Loan Grade

- **Objective:** To validate the internal grading system and ensure it accurately predicts borrower risk.
- **Result:** Loan Grade proved to be the strongest predictor of risk. Grade A loans are exceptionally safe (~6% default rate), whereas Grade G loans are highly risky (~48% default rate).
- **Conclusion:** The current grading system is highly effective and validated by the data.

2. Business Trajectory (Loan Volume over Time)

- **Objective:** To track the business trajectory and identify growth phases versus slowdowns.
- **Result:** The data shows massive growth between 2012 and 2014, peaking in the 2014–2015 period, followed by a visible slowdown in 2016.
- **Conclusion:** While the early 2010s were "boom" years, recent trends suggest a cooling market, indicating a need for new acquisition strategies.

3. Borrower Motivation (Purpose Analysis)

- **Objective:** To understand customer needs and determine which loan products drive the most volume.
- **Result:** "Debt Consolidation" is the primary driver (over 230,000 loans), followed closely by "Credit Card" refinancing.
- **Conclusion:** These two categories represent the core of the business model and should be the focus of marketing efforts.

4. The Impact of Loan Terms and Income

- **Objective:** To determine if longer terms are dangerous and if high income is a reliable safety net.
- **Result (Term):** Long-term loans (60 months) are nearly **twice as risky** as short-term loans (36 months), with default rates of ~32% versus ~16%.
- **Result (Income):** High earners (>\$100k) have a significantly lower default rate (~15%) compared to low earners (~24%).
- **Conclusion:** Term length is a critical risk factor that should be weighed heavily in underwriting; high income does provide a valid safety buffer.

5. The Verification Paradox

- **Objective:** To test if the administrative effort of verifying income actually results in lower defaults.
- **Result:** Counter-intuitively, "Verified" loans show a **higher** default rate (~22%) than "Not Verified" loans (~15%).
- **Conclusion:** This suggests a "borderline" bias in the process: banks likely only demand verification for borrowers who already look risky on paper, while "rock-solid" borrowers get automatic approval (Not Verified) and perform better.