Research Report | September 2025

Risk and Return Dynamics in U.S. Peer-to-Peer Lending

An Empirical Analysis of Lending Club Loan Performance

by
Minh D. Nguyen

Executive Summary

Peer-to-peer (P2P) lending, led by Lending Club in the United States, transformed consumer finance between 2007 and 2014 by offering borrowers direct access to credit and investors a novel fixed-income product. Using SQL-based analysis of over 466,000 loans totaling $6.7 billion in originations, this report examines the interplay of credit risk, borrower characteristics, and investor outcomes during Lending Club's formative years.

Key Findings

- **Strong loan growth with concentrated purposes:** Loan originations expanded sharply from under $10 million in 2007 to over $11 billion by 2014. Debt consolidation and credit card refinancing dominated, accounting for nearly 85% of loan volume.
- **Risk-return stratification across grades:** Default rates rose steadily from 3.4% in Grade A loans to 22.4% in Grade G. Investor returns, however, did not scale linearly—riskier grades offered higher nominal yields but carried heightened volatility and near-zero recoveries.
- **Borrower characteristics matter:** Default rates were materially higher among low-income borrowers (12.3%) compared to high-income borrowers (6.8%). Renters and those with uncertain employment tenure also exhibited greater credit risk.
- **Geographic concentration:** Lending activity was heavily concentrated in California, New York, and Texas, with California alone accounting for $1.0 billion in volume. Default rates varied significantly by state, with some smaller states showing outlier risk profiles.
- **Minimal recovery and partial repayment:** Charged-off loans yielded negligible recoveries (about 0.6%), and overall repayment ratios averaged only 80.6% of principal, highlighting the structural credit risk inherent in the platform.

Contents

Preface

This report was prepared as part of an independent project showcasing the application of SQL-based analytics in financial data. The work combines structured data modeling, query design, and reporting techniques to examine risk and return dynamics in the U.S. peer-to-peer lending market. The intent is both to demonstrate technical capabilities and to provide actionable insights into loan performance patterns.

Introduction

Peer-to-peer (P2P) lending emerged in the mid-2000s as an alternative to traditional bank financing, enabling individuals to borrow and invest through digital platforms. Among these, Lending Club became the largest and most influential marketplace, originating billions in loans across diverse borrower segments. By disintermediating banks, P2P lending offered borrowers access to credit and investors exposure to a new asset class.

With the expansion of this market came questions of credit risk, default behavior, and the sustainability of investor returns. For both lenders and investors, understanding the interplay of borrower characteristics, loan terms, and repayment outcomes is critical to assessing the viability of P2P as a financial product.

This report examines risk and return dynamics in the U.S. P2P lending market using Lending Club data from 2007 to 2014. The analysis is conducted entirely through SQL, leveraging a star-schema data model and a series of structured queries to measure loan performance, borrower profiles, and investor outcomes. The dataset covers more than 460,000 loans, offering a representative view of the early years of the industry.

The study is structured in five parts:

1. **Market Overview** – loan volumes, grade distribution, funding efficiency, and loan purposes.
2. **Loan Performance and Risk** – status distribution, default rates, recoveries, repayments, and outstanding balances.
3. **Borrower Insights** – income groups, employment length, home ownership, and geographic variation.
4. **Time-Series and Seasonality** – loan issuance trends, default dynamics by year, and grade shifts.
5. **Investor and Market Perspectives** – interest rates, realized investor returns, state concentration, and purpose-based lending.

Together, these sections provide a structured view of how credit risk and investor performance evolved during Lending Club's formative years. The findings

highlight both the opportunities and vulnerabilities inherent in marketplace lending, with implications for future credit innovation.

1 Portfolio Overview

The Lending Club portfolio under review consists of approximately **466,285 loans**, representing a total funded volume of **$6.68 billion**. The average loan size is around $\$14,317$, reflecting the relatively small-ticket, retail-oriented nature of the platform's credit exposures.

From a credit quality perspective, the portfolio exhibits a pronounced concentration in the mid-tier grades. Grade B loans constitute the largest segment, accounting for nearly 137 thousand loans with an aggregate value of $\$1.8$ billion, followed closely by Grade C at $\$1.78$ billion across roughly 125 thousand loans. Grade A loans represent a meaningful share as well, with more than 74 thousand facilities totaling just over $1 billion. Lower-rated exposures (Grades D through G) are smaller in volume but remain material, led by Grade D at $\$1.16$ billion. The overall distribution reflects a balanced risk-return profile, skewed toward higher-quality consumer credit while maintaining diversification into higher-yield segments.

At the subgrade level, exposures are more granular but broadly consistent with the overall pattern. Within Grade A, subgrades A4 and A5 dominate, together representing more than $560 million of exposure. In Grade B, subgrades B3 and B4 alone account for nearly $825 million, underscoring investor appetite for the mid-risk tier. Grade C shows an evenly distributed allocation across subgrades C1 to C4, each in the $350-370 million range, while sub-investment-grade segments E, F, and G collectively remain a smaller but deliberate allocation to higher-risk credit, led by E1 at $153 million.

Funding efficiency remains exceptionally strong. The ratio of loans funded to loans issued stands at 99.82%, indicating that nearly all originated loans are fully absorbed by investors. This suggests not only resilient borrower demand but also sustained investor confidence in the asset class.

Loan purpose analysis reveals a pronounced concentration in consumer refinancing. Debt consolidation dominates the book, comprising $\$4.15$ billion or roughly 62% of total portfolio volume, while credit card refinancing contributes

an additional $1.54 billion (23%). Beyond refinancing activity, secondary categories such as home improvement and major purchases are present but comparatively modest in scale, at $369 million and $103 million, respectively. Smaller niches, including medical, wedding, and moving-related loans, remain peripheral to the portfolio.

Taken together, the portfolio demonstrates the defining characteristics of a consumer-lending platform: concentrated exposure to mid-tier credit grades, near-universal funding execution, and a borrower base heavily oriented toward debt refinancing. The risk posture is consistent with a strategy designed to balance quality and yield while sustaining high levels of investor demand.

2 Loan Performance and Risk

The Lending Club loan portfolio exhibits a performance profile reflective of both the scale of issuance and the inherent credit risk of unsecured consumer lending. As of the latest snapshot, the distribution of loan statuses highlights the predominance of active and successfully repaid obligations. Out of more than 466,000 originated loans, roughly 224,000 remain current while 185,000 have been fully repaid. This indicates a substantial proportion of borrowers meeting their contractual obligations. Nevertheless, the credit risk component is visible, with approximately 42,500 loans charged off and 832 categorized as defaulted. Late-stage delinquencies are relatively limited in volume, though they remain a forward indicator of potential credit losses.

Default risk scales meaningfully with credit quality. Loans graded "A" demonstrate a default rate of just over 3%, underscoring the relatively strong underwriting standards at the higher end of the credit spectrum. By contrast, the default rate increases consistently across the grading structure, reaching 12.9% for "D" grade loans and more than 20% for the lowest categories "F" and "G." This steep risk gradient confirms the critical role of credit grading in managing exposure and pricing.

Recovery rates on charged-off loans remain negligible, with less than 1% of outstanding balances ultimately recovered. This underscores the unsecured nature of the asset class and the limited avenues for collection once borrowers default. The repayment ratio, defined as total payments received relative to original principal, stands at approximately 80.6%. While the majority of issued balances have been at least partially repaid, the figure reflects the drag from defaults and charge-offs accumulated over the portfolio's lifetime.

Outstanding principal balances grew significantly during the platform's expansion phase. Virtually nonexistent in the pre-2010 period, outstanding balances rose sharply from 2011 onward, reaching approximately $5.4 million in that year, $32.3 million in 2012, and surging to $370.6 million in 2013. By 2014, total outstanding principal surpassed $1.6 billion, underscoring both the rapid adoption

of the platform and the rising scale of risk carried forward in subsequent vintages.

Taken together, the performance and risk data illustrate a maturing portfolio with clear differentiation by credit quality, low recovery expectations, and material growth in outstanding balances through the early 2010s. The trends highlight the need for rigorous credit grading and disciplined risk-return analysis when assessing unsecured consumer credit at scale.

3 Borrower Insights

Income and Default Risk

Default risk displays a clear inverse relationship with income levels. Borrowers earning above $80,000 annually exhibit the lowest default incidence, with rates averaging 6.8%. Mid-income borrowers, earning between $40,000 and $80,000, default at a higher rate of 9.8%, while low-income borrowers, earning under $40,000, show the greatest vulnerability with a default rate of 12.3%. This income gradient underscores the importance of disposable income and financial resilience in sustaining debt repayment.

Employment Tenure

Employment stability also correlates with loan performance, albeit with nuances. Borrowers with 10 years of employment demonstrate the lowest default rate (8.4%), suggesting that job tenure enhances repayment capacity through income stability. However, borrowers in the early career stages (1–4 years) record default rates around 9.3–9.4%, only marginally better than mid-tenure borrowers (5–7 years at 9.6–10.0%). Interestingly, missing employment data shows elevated risk, with a default rate of 11.3%, indicating possible underreporting or riskier profiles.

Home Ownership

Owner-occupancy status is a significant differentiator in credit performance. Borrowers with a mortgage exhibit default rates of 8.1%, materially below those renting (10.8%) or with other forms of ownership (14.8%). Ownership of a home outright also carries elevated risk at 9.1%, though still better than renters. This pattern suggests that secured housing (mortgaged) signals stronger creditworthiness, while renting is associated with higher repayment uncertainty.

Geographic Variations

Default rates vary considerably across states, reflecting local economic conditions. The highest observed rate is in Nebraska (28.6%) and Nevada (11.7%), while large-volume states such as California (9.9%) and New York (9.8%) track

close to the national average. Florida (10.7%) and Alabama (10.6%) also record above-average stress, consistent with regional credit cycles. Conversely, borrowers in Washington D.C. (5.8%), West Virginia (6.4%), and Mississippi (6.2%) show stronger repayment profiles despite smaller loan bases. Extremely low or zero default rates in states like Maine and Iowa likely reflect limited loan sample sizes rather than structural resilience.

4 Time-Series and Seasonality

Loan origination in the marketplace lending sector expanded at an exponential pace between 2007 and 2014. Quarterly issuance in the earliest years remained modest, with just 24 loans originated in Q2 2007, representing less than $100,000 in total volume. By 2010, the market had reached critical scale: quarterly volumes regularly exceeded $40 million, and annual issuance surpassed $150 million. This acceleration intensified from 2011 onward, with origination nearly doubling year over year. By 2014, the platform was originating over 74,000 loans per quarter, with annual issuance exceeding $3.5 billion in volume, underscoring the rapid institutionalization of peer-to-peer credit.

Default dynamics mirrored the platform's growth trajectory, reflecting both evolving underwriting standards and borrower composition. Early vintages, particularly 2008–2012, were associated with elevated risk, with default rates climbing from 7.5% in 2007 to a peak of 15.2% in 2012. This deterioration likely reflects looser credit underwriting during a period of rapid expansion and a macro backdrop still influenced by post-crisis stress. However, credit performance improved markedly thereafter: 2013 and 2014 vintages show significantly lower defaults, at 11.1% and 6.2%, respectively. This sharp reversal suggests enhanced borrower screening, risk-pricing discipline, and possibly a stronger macroeconomic tailwind.

Shifts in loan grade composition provide additional perspective. In the earliest years (2007–2009), issuance was distributed relatively evenly across grades A through E, with substantial exposure to mid- and lower-tier credit. As institutional capital entered, the grade distribution shifted decisively. From 2010 onward, origination skewed heavily toward higher-quality credit, particularly grades A through C. By 2014, loans in these three categories accounted for nearly 80% of issuance, while grade F and G originations declined to negligible levels. This rebalancing reflects investor preference for stable returns, and the platform's pivot to building credibility with institutional backers.

In conclusion, issuance patterns, default trends, and grade distribution illus-

trate a market moving from experimental beginnings toward institutional maturity. Early growth brought credit deterioration, but over time, lending standards tightened, resulting in larger volumes concentrated in higher-grade borrowers with improved repayment performance.

5 Market Dynamics and Investor Returns

The lending portfolio exhibits a clear and consistent pricing gradient across credit grades, with interest rates increasing in line with risk. Grade A loans, representing the highest-quality borrowers, carry an average interest rate of 7.5 percent, while Grade G loans, at the riskiest end of the spectrum, demand rates approaching 25 percent. This progression reflects standard risk-based pricing practices, balancing borrower creditworthiness against investor-required compensation for default risk.

When analyzing investor returns, the distribution appears less uniform. While Grade A loans deliver a modest but stable return of 5.7 percent, returns in the lower-quality segments display substantial variability. For instance, Grade D loans register outsized returns relative to expectations, suggesting higher volatility and risk concentrations in this category. Similarly, Grades E through G demonstrate inconsistent performance, with realized returns diverging materially from their corresponding interest rate structures. This indicates that while investors are compensated for taking on risk, realized gains may be distorted by default timing, recovery inefficiencies, or portfolio selection effects.

Geographically, loan origination is highly concentrated in a small number of states. California dominates with over $1.0 billion in total loan volume, followed by New York and Texas at approximately $568 million and $551 million, respectively. Together, the top ten states account for the majority of lending activity, underscoring the market's concentration in large, economically significant regions. Such clustering suggests both an opportunity for scale efficiency and a vulnerability to localized economic shocks.

In terms of loan purpose, debt consolidation remains the cornerstone of borrower demand, accounting for more than $4.1 billion, or the majority of total volume. Credit card refinancing follows with $1.5 billion, while all other categories remain comparatively minor. This distribution highlights the platform's role as a refinancing channel rather than a source of entrepreneurial or discretionary financing. The dominance of debt consolidation suggests that the platform is heav-

ily intertwined with household balance-sheet management, particularly for borrowers seeking to restructure or simplify their liabilities.

Overall, the data reflects a market characterized by rational pricing, but also by geographic and functional concentration. For investors, the appeal lies in predictable pricing tiers and large-scale exposure to consumer credit. However, the volatility of returns in non-prime segments and the reliance on debt restructuring demand underline the importance of disciplined risk selection and portfolio diversification.

6 Technical Appendix

6.1 Data Source

This analysis draws upon the Lending Club loan-level dataset, encompassing 466,256 loans originated between 2007 and 2014. The dataset provides a comprehensive view of consumer credit, including borrower demographics, loan characteristics, performance outcomes, and state-level identifiers. Together, these variables allow for a granular assessment of portfolio structure, borrower behavior, and investor performance.

The dataset underwent a structured cleaning process prior to analysis. Records with incomplete or inconsistent values were excluded, and categorical variables such as loan status, grade, and purpose were standardized. Dates were converted into consistent formats to enable time-series aggregation, while loan-level metrics were consolidated into portfolio-level measures. From this base, key indicators such as loan volumes, average loan size, default outcomes, and investor returns were derived.

6.2 Analytical Methodologies

Analytical outputs were constructed using a series of structured SQL queries. Aggregations were performed by loan grade, loan purpose, and state of origination, enabling a multidimensional perspective on both credit supply and risk dynamics. Investor returns were estimated as realized outcomes per loan grade, net of defaults, providing a direct measure of portfolio performance from the investor's perspective. Concentration analyses were employed to assess the distribution of loan volumes by geography and purpose, highlighting areas of market dependence and structural exposure.

The queries executed to generate these insights followed a standardized design. Dedicated views were developed to capture loan volumes, default rates, repayment ratios, and recovery levels. Additional queries aggregated outcomes by grade, state, and purpose, ensuring consistency across the analytical framework. This modular approach allowed for the systematic construction of higher-level

indicators while preserving transparency and reproducibility of results.

It is important to note certain limitations inherent in the analysis. The dataset reflects realized loan performance only through 2018 and does not incorporate subsequent vintages or macroeconomic conditions. Aggregations at the state level may obscure intra-state variation in credit dynamics, while investor return estimates represent averages that may not fully capture reinvestment strategies or portfolio construction techniques. Despite these constraints, the analysis provides a robust foundation for understanding the evolution of peer-to-peer lending, borrower risk characteristics, and the associated implications for investors.

6.3 Methodological Notes

The following definitions underpin the key indicators presented in this report. All calculations were executed within a structured SQL environment leveraging the star-schema data model.

Total Loans, Total Volume, Average Loan Size (1.1)

$$\text{Total Loans} = N \tag{1}$$

$$\text{Total Volume} = \sum_{i=1}^{N} \text{LoanAmount}_i \tag{2}$$

$$\text{Average Loan Size} = \frac{\text{Total Volume}}{N} \tag{3}$$

Funding Efficiency (1.4)

$$\text{Funding Efficiency} = \frac{\sum \text{FundedAmount}}{\sum \text{LoanAmount}} \tag{4}$$

Default Rate (2.2, 3.1–3.4, 4.2)

$$\text{Default Rate} = \frac{\sum 1_{\text{LoanStatus} \in [\text{Default, ChargedOff}]}}{N} \tag{5}$$

where $1_{\{.\}}$ is the indicator function.

Recovery Ratio (2.3)

$$\text{Recovery Ratio} = \frac{\sum \text{Recoveries}}{\sum \text{LoanAmount}} \tag{6}$$

Repayment Ratio (2.4)

$$\text{Repayment Ratio} = \frac{\sum \text{TotalPayments}}{\sum \text{LoanAmount}} \tag{7}$$

Outstanding Principal (2.5)

$$\text{Outstanding Principal}_t = \sum_{i=1}^{N_t} \text{OutstandingPrincipal}_{i,t} \tag{8}$$

where N_t is the number of active loans in year t.

Average Interest Rate (5.1)

$$\text{Average Interest Rate}_{grade} = \frac{1}{N_{grade}} \sum_{i=1}^{N_{grade}} \text{IntRate}_i \tag{9}$$

Average Investor Return (5.2)

$$\text{Investor Return}_i = \frac{\text{TotalPayment}_i - \text{FundedAmountInv}_i}{\text{FundedAmountInv}_i} \tag{10}$$

$$\text{Average Investor Return}_{grade} = \frac{1}{N_{grade}} \sum_{i=1}^{N_{grade}} \text{Investor Return}_i \tag{11}$$

Loan Volume Concentration (5.3, 5.4)

$$\text{Loan Volume}_{group} = \sum_{i}^{group} \text{LoanAmount}_i \tag{12}$$

where $group$ may refer to state or loan purpose.

6.4 SQL Queries

All queries used in this report are available in the project's GitHub repository. The aim will be an easy approach to get actionable insights, moreover, there are also advanced techniques adapted in order to demonstrating the diversity of SQL. Selected, representative queries of each part will be shown below.

Total loans, total volume, average loan size

```
SELECT
    COUNT(*) AS total_loans,
    SUM(loan_amnt) AS total_volume,
    AVG(loan_amnt) AS avg_loan_size
FROM
    star_schema.fact_loans;
```

or with Common Table Expression,

```
WITH loan_base AS (
    SELECT
        loan_amnt
    FROM
        star_schema.fact_loans
)
SELECT
    COUNT(*) AS total_loans,
    SUM(loan_amnt) AS total_volume,
    AVG(loan_amnt) AS avg_loan_size
FROM
    loan_base;
```

Loan performance - Recovery ratio

```
WITH recover AS (
    SELECT
        recoveries,
        loan_amnt
    FROM
```

```
        star_schema.fact_loans
)
SELECT
    SUM(recoveries) / NULLIF (SUM(loan_amnt), 0) AS
       recovery_rate
FROM
    recover;
```

Borrower characteristics - Default rate by state

```
SELECT
    b.addr_state,
    ROUND(AVG(
            CASE WHEN f.loan_status IN ('Default', 'Charged
               Off') THEN
                1
            ELSE
                0
            END), 4) AS default_rate,
    COUNT(*) AS loan_count
FROM
    star_schema.fact_loans f
    JOIN star_schema.dim_borrower b
    ON f.borrower_key = b.borrower_key
GROUP BY
    b.addr_state
ORDER BY
    default_rate DESC;
```

Temporal analysis - Loan grade distribution by year

```
SELECT
    d.year,
    l.grade,
    COUNT(*) AS loan_count
FROM
    star_schema.fact_loans f
```

```
    JOIN star_schema.dim_loan l ON f.loan_key = l.loan_key
    JOIN star_schema.dim_date d ON f.date_key = d.date_key
GROUP BY
    d.year,
    l.grade
ORDER BY
    d.year,
    l.grade;
```

Interest and returns - Average investor return by grade

```
WITH investor_returns AS (
    SELECT
        l.grade,
        (f.total_pymnt - f.funded_amnt_inv) /
        NULLIF (f.funded_amnt_inv, 0) AS return_i
    FROM
        star_schema.fact_loans f
        JOIN star_schema.dim_loan l ON f.loan_key = l.loan_key
)
SELECT
    grade,
    ROUND(AVG(return_i), 4) AS avg_return
FROM
    investor_returns
GROUP BY
    grade
ORDER BY
    grade;
```