**Prosper Loan Portfolio Valuation Report**

Justin Ladrillono   
06/17/2025

**Executive Summary**

**The estimated Net Present Value (NPV) of the "Core Non-Performing" Prosper loan portfolio is $1,721,975.13 as of January 1, 2015, highlighting the portfolio's estimated value under current assumptions.**

This report details the valuation of a historical personal loan portfolio (originated ~2005-2014) from the Prosper platform, focusing on loans that have charged off or defaulted. I created a data pipeline to create a segmented Discounted Cash Flow (DCF) model. This model incorporates granular segmentation based on FICO scores and loan original amounts, applying segment-specific recovery curves that account for loan seasoning.

Sensitivity analysis reveals that the valuation is **most significantly impacted by changes in recovery rates**, followed by the discount rate and collection costs. While the calculated gross recovery rate of ~5.84% is conservative compared to broader industry benchmarks (10-30% for unsecured personal loans pre-2019), this analysis provides a clear, data-driven framework for assessing portfolio value and understanding key risk drivers under various market conditions.

## 1. Introduction & Project Objectives

The primary objective of this project was to establish a fair market valuation for a defined segment of a historical personal loan portfolio. This valuation serves as a foundational step for potential investment decisions, risk assessment, and performance benchmarking for distressed assets. Our core focus was on loans that had already transitioned to a "Chargedoff" or "Defaulted" status, termed the "Core Non-Performing" portfolio.

**Key Deliverables**

* A PostgreSQL database (debt\_portfolio) containing cleaned, transformed, and segmented loan data.
* A functional Discounted Cash Flow (DCF) model to calculate portfolio NPV.
* Comprehensive sensitivity and scenario analyses of the valuation.
* A detailed project report and presentation, designed to convey clear and actionable insights.

## 2. Data Sources & Preparation

The analysis is based on the prosperLoanData.csv dataset, encompassing personal loans originated approximately between 2005 and 2014.

2.1 Database Design

A PostgreSQL database named debt\_portfolio was designed with a structured schema to facilitate robust data processing and analysis:

* **staging\_prosper**: Serves as the initial landing zone for raw imported data, with all columns initially stored as TEXT.
* **portfolio\_accounts\_core**: The central table containing cleaned, filtered (only 'Chargedoff' or 'Defaulted' loans), and extensively feature-engineered data specific to the 'Core Non-Performing' portfolio.
* **score\_risk\_mapping**: A lookup table explicitly mapping FICO score bands to predefined risk multipliers, essential for tailoring recovery rates.
* **recovery\_curve\_parameters**: Stores detailed segment-specific monthly recovery percentages over the entire 84-month forecast horizon.
* **model\_parameters**: A dedicated configuration table housing global assumptions critical to the DCF model.

2.2 Data Cleaning & Feature Engineering

The SQL scripts are the queries I used to transform the raw data from staging\_prosper into a cleaned dataset for analysis, the portfolio\_accounts\_core table. Key transformations included:

* **Targeted Filtering**: Only loans with LoanStatus of 'Chargedoff' or 'Defaulted' and a positive LP\_GrossPrincipalLoss were included, ensuring focus on the relevant distressed assets.
* **Resilient Data Type Casting**: Employed extensive use of TRIM, NULLIF, COALESCE, CAST, and CASE statements to handle problematic text fields (e.g., empty strings, non-numeric characters in numeric columns, various date formats). This ensures data integrity when converting to NUMERIC, INT, or DATE types.
* **Enhanced prosper\_score Handling**: A critical CASE logic was implemented to correctly parse the prosper\_score (which in raw form could contain decimals or non-numeric text). Valid numeric scores are cast to INT, while invalid or empty entries are now correctly assigned NULL, rectifying a previous default-to-0 issue.
* **estimated\_loss & estimated\_return Integration**: These fields, indicative of Prosper's own risk assessments, were carefully processed to ensure proper NUMERIC casting and NULL handling for missing data points, making them available for future detailed analysis.
* **Strategic Feature Engineering**: The following variables were calculated to improve the accuracy of my valuation:
* charge\_off\_date: Precisely determined from the ClosedDate for charged-off/defaulted loans.
* fico\_score\_proxy: Calculated as the midpoint of CreditScoreRangeLower and CreditScoreRangeUpper to provide a consistent FICO representation.
* loan\_age\_at\_charge\_off\_months: Quantifies the loan's age at the point of charge-off.
* months\_since\_charge\_off\_at\_valuation: Measures the elapsed time from charge-off to the valuation date (January 1, 2015), crucial for "seasoning" adjustments.
* **Granular Segmentation**: Loans in portfolio\_accounts\_core were systematically assigned a segment\_id based on a 4x4 matrix, combining distinct fico\_score\_proxy bands (<620, 620-679, 680-719, 720+) with LoanOriginalAmount quartiles. All 16 segments were successfully populated, enabling a more precise valuation.
* **Data Integrity Assurance**: Through iterative diagnostics and refinement, initial data quality issues were successfully addressed, ensuring a clean and reliable dataset for the core model.

## 3. Valuation Methodology: Segmented Discounted Cash Flow (DCF)

The portfolio was valued using a segmented Discounted Cash Flow (DCF) model, a widely accepted approach for valuing assets based on their future cash flow generation.

3.1 Core Components of the DCF Model

* **Core Principle**: The fundamental concept involves discounting projected future net cash flows back to the valuation date to determine their present value.
* **Recovery Base**: LP\_GrossPrincipalLoss serves as the initial principal amount charged off for each loan, forming the base upon which recoveries are calculated.
* **Segment-Specific Recovery Curves**: This is a cornerstone of the model's accuracy. The recovery\_curve\_parameters table stores unique monthly recovery percentages for each of the 16 defined segments over the 84-month forecast horizon. These curves are dynamic, applying a base monthly recovery rate, adjusted by a risk\_multiplier (derived from score\_risk\_mapping based on the loan's FICO band), and incorporate an exponential decay factor to reflect diminishing recovery potential over time.
* **Loan Seasoning**: Crucial for precise forecasting, the model meticulously accounts for each loan's months\_since\_charge\_off\_at\_valuation. Recoveries are only applied from the appropriate point on the segment's recovery curve, accurately reflecting the loan's existing "seasoning" at the valuation date. *(Notably, for this specific portfolio, initial 9 forecast months show zero aggregate recoveries, as all loans contributing significant principal loss were already seasoned by at least 9 months at the valuation date.)*
* **Net Cash Flow Calculation**: For each segment, forecasted monthly gross recoveries are aggregated. From these gross recoveries, a flat base\_case\_collection\_cost\_percentage (25%) is applied to derive the projected net monthly cash flows.
* **Discounting**: These net monthly cash flows are then discounted back to the valuation\_date (January 1, 2015) using a base\_case\_discount\_rate (20% annual), which is converted to a precise monthly rate for period-by-period discounting.
* **Net Present Value (NPV)**: The discounted monthly net cash flows for each segment are summed to yield a segment-specific NPV. These segment NPVs are then aggregated to derive the total portfolio NPV.

### 3.2 Key Assumptions

**Valuation Date**: January 1, 2015.

**Forecast Horizon**: 84 months (7 years).

**Base Case Annual Discount Rate**: 20.00%.

**Base Case Collection Cost Percentage**: 25.00% of gross recoveries.

**Base Monthly Recovery Rate**: 1.0X (adjusted by FICO-based risk multipliers and exponential decay).

## 4. Results: Valuing the Portfolio

### 4.1 Initial Portfolio Metrics

Analysis of the 'Core Non-Performing' portfolio yields the following baseline metrics:

**Total Gross Principal Loss**: $79,806,849.12

**Forecasted Gross Recovery**: $4,659,168.17

**Gross Recovery Rate**: 5.84% (Calculated as Forecasted Gross Recovery / Total Gross Principal Loss)

### 4.2 Base Case Valuation

Applying the segmented DCF model with the base case assumptions:

**Total Portfolio Net Present Value (NPV)**: **$1,721,975.13**



## 5. Sensitivity & Scenario Analysis: Understanding Value Drivers

To provide a comprehensive understanding of the valuation's robustness, a thorough sensitivity and scenario analysis was performed.

### 5.1 Sensitivity Analysis: Identifying Key Drivers

One-way and two-way data tables were constructed to reveal how changes in individual and combined key assumptions impact the Total Portfolio NPV.

* **Key Drivers of NPV Sensitivity (from most to least impactful)**:
  1. **Recovery Adjustment Factor**: This variable demonstrated the most significant influence on the NPV. Even small percentage changes in this factor resulted in the largest swings in the overall valuation. This highlights the paramount importance of accurate recovery forecasting.
  2. **Base Discount Rate**: As a direct driver of present value, the discount rate was the second most influential factor, illustrating the sensitivity to assumed investment returns.
  3. **Collection Cost Percentage**: While impacting the net cash flows, changes in collection costs had a moderate effect on the NPV compared to recovery rates or the discount rate.

5.2 Scenario Analysis: "What If" Outlooks

Three distinct scenarios were defined and analyzed to assess portfolio value under specific plausible economic outlooks, providing a range of potential valuations.

|  |  |  |
| --- | --- | --- |
| **Scenario** | **Portfolio NPV** | **Key Assumptions (Example)** |
| **Downside Scenario** | **$892,815.94** | *DR: 25%, CC: 30%, Rec. Factor: .75X* |
| **Base Scenario** | **$1,721,975.13** | *DR: 20%, CC: 25%, Rec. Factor: 100%* |
| **Upside Scenario** | **$2,820,496.78** | *DR: 15%, CC: 20%, Rec. Factor: 1.25X* |

*A screenshot of a computer screen

AI-generated content may be incorrect.*

## 6. Limitations & Future Work

### 6.1 Limitations

* **Recovery Rate Assumptions**: The model relies on generalized industry benchmarks for recovery rates. Actual Prosper-specific loan characteristics and vintage may differ, potentially leading to discrepancies. The observed portfolio recovery rate (~5.84%) is notably lower than typical industry figures (10-30% for unsecured personal loans), suggesting either a very conservative curve or a uniquely distressed portfolio.
* **Flat Collection Cost**: Assuming a flat percentage for collection costs may not fully capture the nuances of collection efforts, which can vary significantly based on loan age, size, and borrower engagement.
* **Static Discount Rate**: The model employs a fixed discount rate throughout the forecast horizon. In reality, market discount rates can fluctuate over time.
* **Data Vintage**: The valuation is performed as of January 1, 2015, using historical data up to that point. Market conditions and recovery dynamics may have changed considerably in the years since.
* **No "At-Risk" Portfolio Analysis (Optional)**: The current valuation specifically excludes the potential value contributions from loans that are delinquent but have not yet officially charged off.

### 6.2 Areas for Future Improvement

* **Refine Recovery Curve Modeling**: Enhance the precision of recovery forecasts by incorporating more granular borrower attributes (e.g., Debt-to-Income Ratio, Homeowner Status, detailed Prosper loan payment history, comprehensive credit bureau data like Bankcard Utilization) into either the segmentation methodology or the direct derivation of risk adjustments.
* **At-Risk Portfolio Analysis**: Complete Phase VI to estimate the value of potential future charge-offs from currently delinquent loans, thereby providing a more holistic portfolio valuation.
* **Stochastic Modeling**: Introduce advanced techniques such as Monte Carlo simulations to model uncertainty in key variables (e.g., recovery rates, discount rates). This would yield a probabilistic distribution of possible NPV outcomes rather than just discrete scenarios.
* **Tooling Integration & Automation**: Investigate direct API integrations with the PostgreSQL database and Excel/Google Sheets to automate data export, model updates, and even visualization generation, significantly streamlining future analysis.

## 7. Conclusion

To value the “Core Non-Performing” Prosper loan portfolio, I created a segmented discounted cash flow model. The base case NPV of $1.72 million is primarily driven by the assumed recovery rates and the discount rate. While the observed low gross recovery rate highlights the distressed nature of this specific portfolio, the model offers valuable insights into the portfolio's intrinsic value drivers and its sensitivity to key assumptions. The strong foundation laid in data preparation and rigorous financial modeling ensures the model's accuracy and provides a solid basis for all future enhancements and strategic decision-making.