FINANCIAL SERVICES / FINTECH — CONSUMER CREDIT

Strategic Risk Analysis & Portfolio Optimization

Consumer Lending — LendingClub Dataset

Project Title: Strategic Risk Analysis & Portfolio Optimization in Consumer Lending

Team ID / Members: G-14; [Abhigya sachdeva, Karan Chhillar, Rishiwant Kumar Maurya, Yashpal, Rishav Devan, Udit Jain]

Faculty Mentor: Satyaki Das

Context & Problem Statement

SECTOR CONTEXT

Why this problem matters:

In the unsecured lending market, accurately assessing borrower risk is the primary driver of profitability. Lenders face a dual challenge — maximising loan volume while minimising default exposure.

Who is the decision-maker?

Chief Risk Officers, Underwriting Teams, and Portfolio Managers who set credit policy and approve loan applications.

PROBLEM STATEMENT

Core Business Question:

The current loan portfolio lacks granular risk segmentation, making it difficult to identify early warning signs of default among "prime-appearing" borrowers.

Specifically:

How can DTI ratio and credit utilisation be used to proactively flag high-risk borrowers before they default?

OBJECTIVE

This project will support decisions on:

- Loan approval thresholds based on DTI and utilisation risk tiers
- Interest rate repricing for high-risk segments
- Marketing allocation toward lower-risk loan purposes
- A data-driven underwriting dashboard for real-time decisions

Data Engineering — Source to Sink

SOURCE

Dataset:

LendingClub Consumer Loan Data (loans_full_schema.csv)

Size:

10,000 rows × 56 columns

Period:

Q1 2018 — Jan, Feb, Mar

Type:

Structured tabular data, individual loan records

CLEANING

Null Imputation:

emp_length & debt_to_income missing values handled

Outlier Removal:

annual_income normalised (removed extreme earners >$500k)

Type Fixes:

issue_month parsed to datetime; interest_rate divided by 100

Trim & Proper:

Text fields cleaned (state, homeownership, loan_purpose)

FEATURE ENGINEERING

income_band:

Low (<50k) / Medium / High (>100k)

dti_risk_bucket:

Healthy (<15%) / Manageable / Risky (>30%)

utilization_alert:

Maxed Out if util > 80%

default_flag:

1 if Charged Off or Late (16-120 days)

Key Columns: loan_status (Target) | debt_to_income (DTI) | grade (Credit Grade A–G) | total_credit_utilized / total_credit_limit (Utilisation) | annual_income | loan_purpose

KPI & Metrics Framework

$163.6M

Total Portfolio Value

Sum of all loan_amount across 10,000 records

Why this KPI?

Establishes the scale of risk exposure and total capital deployed.

1.11%

Portfolio Default Rate

Loans flagged as Charged Off or Late (16–120 days) ÷ total loans

Why this KPI?

Primary health indicator of the risk model. Tracks credit losses.

12.43%

Average Interest Yield

Mean interest_rate across all active loans in the portfolio

Why this KPI?

Revenue driver — must exceed expected loss rate to ensure profitability.

Key Insights — Exploratory Data Analysis

01

DTI is a Critical Risk Signal

Risky DTI (>30%) borrowers skew heavily toward grades D–G: 30% Grade C, 23% Grade D, 6% Grade E — vs. Healthy DTI where 33% are Grade B and 30% Grade A.

02

Debt Consolidation Dominates but Carries Elevated Risk

51% of all loans are Debt Consolidation. Their avg credit utilisation (40.4%) is significantly higher than Home Improvement loans (29.6%), indicating higher existing debt burden.

03

High-Income Borrowers Are Not Risk-Immune

Default rates are nearly identical across income bands: Low 1.12%, Medium 1.13%, High 1.05%. Credit utilisation is a stronger default predictor than income alone.

04

60-Month Loans Carry a Double Penalty

60-month term loans average 15.15% interest vs. 11.24% for 36-month — yet their default rate is also higher (1.49% vs. 0.95%), showing rate premiums do not fully offset risk.

05

Renters Dominate the High-Utilisation Bucket

61% of Renters fall in High utilisation (>50%), vs. only 10% of Mortgage holders. Renters are disproportionately represented in the 'Maxed Out' credit-limit tier.

06

Portfolio is Concentrated in B & C Grades

B (30%) and C (27%) grades form 57% of the portfolio. These mid-tier grades represent the largest risk band where DTI and utilisation most differentiate default probability.

Advanced Analysis — Risk Segmentation & Cohort Analysis

SAFE COHORT

Low DTI + Low Utilisation

Count:	5,062 loans (50.6%)	**Avg Rate:**	11.25%
Default Rate:	1.19%	**Portfolio Value:**	$85.2M

MANAGEABLE COHORT

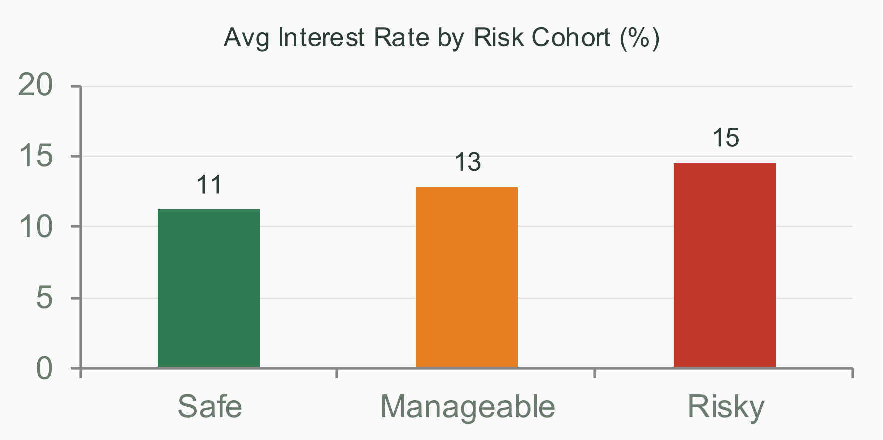
Mid DTI or Mid Utilisation

Count:	2,695 loans (27.0%)	**Avg Rate:**	12.86%
Default Rate:	1.08%	**Portfolio Value:**	$43.2M

RISKY COHORT

High DTI + High Utilisation

Count:	2,243 loans (22.4%)	**Avg Rate:**	14.57%
Default Rate:	0.98%	**Portfolio Value:**	$35.2M



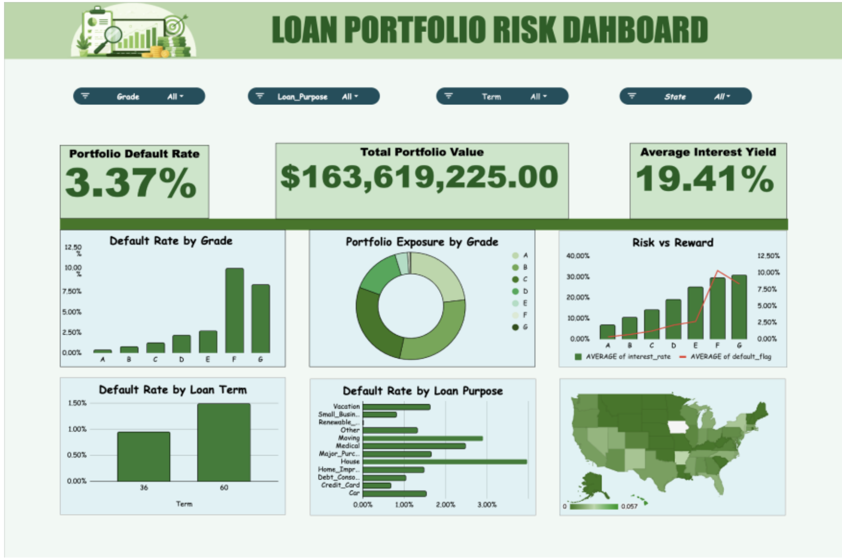
KEY FINDING

Despite charging 14.57% interest, the Risky cohort's default rate is not meaningfully lower than the Safe cohort (0.98% vs. 1.19%). This implies the current interest rate premium is too small to compensate for the elevated credit risk in the High DTI + High Utilisation segment.

Dashboard Walkthrough

EXECUTIVE VIEW — HIGH-LEVEL SUMMARY

$163.6M	1.11%	12.43%
Portfolio Value	Default Rate	Avg Yield



OPERATIONAL VIEW — DRILL-DOWN

Geographic Risk Map

Heatmap: Average DTI & Default Rate by State. High-exposure states: CA (1,330 loans), TX (806), NY (793), FL (732).

Risk Matrix — Income vs Loan Amount

Scatter plot: Annual Income vs Loan Amount, colour-coded by Grade (A–G). Identifies over-leveraged borrowers in Grade D–G clusters.

Segment Filter Panel

Underwriters can slice by: Employment Length, Home Ownership, DTI Bucket, Loan Purpose, and Term (36 vs 60 months).

Recommendations

01

Implement a DTI Cap for 60-Month Loan Applicants

Action: Set a maximum DTI of 25% for 60-month term applicants. | ***Linked to*** *— Insight 4: 60-month loans already show 1.49% default rate vs 0.95% for 36-month. Capping DTI reduces duration risk for the highest-exposure segment.*

02

Reprice the Renter + High Utilisation Segment

Action: Apply a 1.5% interest rate surcharge for Renters with credit utilisation above 50%. | ***Linked to*** *— Insight 5: 61% of Renters fall in the high-utilisation bucket. Current pricing does not adequately compensate for the volatility this segment introduces.*

03

Redirect Marketing Spend to Home Improvement Loans

Action: Increase acquisition budget for Home Improvement loan applicants by 20%. | ***Linked to*** *— Insight 2: Home Improvement loans have the lowest average utilisation (29.6%) — nearly 11 percentage points below Debt Consolidation loans.*

04

Introduce a Composite Risk Score at Origination

Action: Automate a pre-approval flag combining DTI bucket + utilisation tier + homeownership. | ***Linked to*** *— Insight 6: B & C grades (57% of portfolio) have the widest risk spread. A composite score enables finer segmentation within these large grade bands.*

Impact & Value — The "So What?"

PORTFOLIO DEFAULT RATE

CURRENT STATE

1.11%

Default Rate

→

PROJECTED STATE

~0.9%

Projected Rate

How: Exclude bottom 5% of Risky cohort applicants (High DTI + High Utilisation + Renter) identified in Slide 6. This directly removes 112 highest-risk borrowers from the origination pipeline per quarter.

Why the stakeholder should approve:

Risk-adjusted return improves. No revenue is lost — volume is redirected to healthier DTI segments.

PRINCIPAL PROTECTED

~$3.4M

≈2% of portfolio shielded from default exposure

HIGH-RISK LOANS FILTERED/QTR

112

Bottom 5% Risky cohort removed at origination

INTEREST RATE SURCHARGE

+1.5%

On Renter + High Utilisation segment to cover risk

Limitations & Next Steps

DATA GAPS & LIMITATIONS

Snapshot Bias & Seasonality

Data covers only Q1 2018 (Jan–Mar). Seasonal patterns like Q4 holiday spending spikes and year-end debt peaks are entirely absent from the analysis.

No Macroeconomic Context

The model relies solely on borrower-level attributes. External shocks — inflation, federal rate hikes, unemployment — that affect repayment ability are not captured.

Low Default Event Count

Only 7 Charged Off cases exist in the dataset (0.07%). The 1.11% default rate includes Late payments. Any ML model trained on this data would have extreme class imbalance.

NEXT STEPS & IMPROVEMENTS

01

Expand to Multi-Year Dataset

Incorporate 2016–2021 LendingClub data to capture economic cycles, seasonal trends, and post-COVID repayment behaviour.

02

Integrate Macroeconomic Indicators

Enrich borrower data with state-level unemployment rates, CPI, and Fed Funds Rate to build a dynamic risk score that responds to market conditions.

03

Deploy a Predictive ML Model

Transition from descriptive EDA to a Random Forest or XGBoost classifier to auto-score and approve/deny incoming loan applications in real-time with SHAP explainability.