FINAL PROJECT REPORT

Title: Strategic Risk Analysis & Portfolio Optimization

Sector: Financial Services / Peer-to-Peer Lending

Team Details: G-14

Institute / Faculty: Newton School of Technology , Rishihood University

Executive Summary

- **Problem:** The current loan portfolio faces a **1.11% default rate**, with concentrated risk in specific sub-segments (High-risk grades and long-term loans) that threatens overall

profitability.

- **Approach:** We conducted a comprehensive audit of the **LendingClub Q1 2018 dataset** (10,000+ records), utilizing Exploratory Data Analysis (EDA) to identify risk drivers across Credit Grades, Tenure, and Loan Purposes.

- **Key Insights:**
    - **Grade F** loans are the highest risk category with a **10.34% default rate**.
    - **60-month loans** carry significantly higher risk (**1.49% default**) compared to 36-month loans (**0.95%**).
    - **House loans** underperform with a **3.97% default rate**.

- **Key Recommendations:** Implement a strategic exclusion policy capping **Debt-to-Income (DTI) at ≤25% for 60-month loans**.

- **Impact:** Projected **risk reduction from 1.11% to ~0.9%**, safeguarding an estimated **$3.4M in principal**.

Sector & Business Context

- **Sector Overview:** The Peer-to-Peer (P2P) lending market has grown rapidly, offering borrowers alternative financing and investors higher yields. However, it faces inherent credit risks due to unsecured lending.

- **Current Challenges:** Rising default rates in sub-prime segments and the difficulty of accurately pricing risk for long-term (60-month) loans.

- **Why this problem was chosen:** Optimizing the portfolio to balance **Yield (12.43%)** against **Default Risk (1.11%)** is critical for sustainable growth and investor confidence.

Problem Statement & Objectives

- **Formal Problem Definition:** Identify specific cohorts within the loan portfolio that disproportionately contribute to the **1.11% default rate** and develop a strategy to minimize these defaults without significantly eroding the **12.43% average yield**.
- **Project Scope:** Analysis of the **$163,619,225** portfolio from Q1 2018, focusing on credit grade, term, purpose, and DTI.
- **Success Criteria:** A measurable reduction in the projected default rate and identification of at least **$1M+** in principal protection.

Data Description

- **Source:** LendingClub Q1 2018 Dataset.
- **Data Structure:** Structured tabular data (Rows: Loans, Columns: Attributes).

- **Columns Explanation:** Key variables include `loan_status` (Target), `grade`, `sub_grade`, `term`, `purpose`, `dti`, `int_rate`, and `annual_inc`.
- **Data Size: 10,000+ loan records** with **50+ variables**.
- **Data Limitations:** Historical data may not fully predict future economic downturns; limited to Q1 2018 snapshot.

Data Cleaning & Preparation

- **Missing Values Handling:** Imputed or removed records with critical missing fields (e.g., `dti`, `loan_status`).
- **Outlier Treatment:** Analyzed `annual_inc` and `dti` for extreme values that could skew averages.
- **Transformations:** Converted `term` (e.g., " 36 months") to numeric format for correlation analysis.
- **Feature Engineering:** Created `default_flag` (1 for Default/Charged Off, 0 for Fully Paid) to facilitate binary classification analysis.
- **Assumptions:** "Charged Off" and "Default" statuses are treated as loss events.

KPI & Metric Framework

KPI	Definition	Formula	Why it Matters
Default Rate	Percentage of loans that have failed to pay.	(Count of Defaults / Total Loans) * 100	Primary measure of portfolio risk (Current: **1.11%**).
Average Yield	Mean interest rate across the portfolio.	Average(int_rate)	Measure of return on investment (Current: **12.43%**).
Portfolio Value	Total principal amount of all loans.	Sum(loan_amnt)	Context for impact estimation (Current: **$163.6M**).
Loss Given Default	Estimated loss when a default occurs.	Loan Amount * (1 - Recovery Rate)	Critical for financial impact modeling.

Exploratory Data Analysis (EDA)

- **Trend Analysis (Default by Grade):**
  - **Grade A:** 0.37% (Safe)
  - **Grade B:** 0.72%
  - **Grade C:** 1.24%
  - **Grade D:** 2.14%
  - **Grade E:** 2.69%
  - **Grade F: 10.34%** (Highest risk concentration)
  - **Grade G:** 8.33%
  - *Insight:* Grades F and G represent a disproportionate risk relative to their volume.
- **Comparison Analysis (Term):**
  - **36 Months:** 0.95% Default Rate.
  - **60 Months: 1.49% Default Rate**.
  - *Insight:* Longer tenure increases exposure to default risk by **~57%**.

- **Distribution Analysis (Purpose):**

    - **House Loans: 3.97% Default Rate** (Significantly higher than portfolio average of 1.11%).
    - *Insight:* Real estate-linked loans in this dataset underperform standard personal loans .

Advanced Analysis

- **Risk Segmentation:** By cross-referencing **Term (60 months)** with **DTI (>25%)**, we isolated a specific "Toxic" sub-segment responsible for a large portion of losses.
- **Scenario Analysis:**
    - *Scenario A (Status Quo):* Portfolio Default Rate remains at **1.11%**.
    - *Scenario B (Optimization):* Excluding 60-month loans with DTI >25% reduces the Default Rate to **~0.9%**, saving **$3.4M** in principal .

Dashboard Design

- **Dashboard Objective:** Monitor real-time portfolio health, focusing on Default Rate and Yield by Grade and Term.
- **View Structure:**
    - *Summary Card:* Total Value ($163.6M), Default Rate (1.11%).
    - *Charts:* Bar chart for "Default Rate by Grade," Pie chart for "Portfolio by Term."
- **Filters & Drilldowns:** Filter by Grade, Term (36/60), and State to isolate pockets of risk.
- **Screenshots:** [Insert Screenshot of Google Sheets Dashboard here]

Insights Summary

1. **Grade F is the primary risk driver** with a default rate of **10.34%**, 9x higher than the portfolio average.
2. **60-month loans are riskier** (1.49% default) than 36-month loans (0.95%).
3. **House loans perform poorly** (3.97% default), suggesting mispricing or adverse selection.
4. **High DTI correlates with default**, especially in longer-term loans.
5. **Grade A and B are stable anchors**, with default rates under 0.8%.
6. **Yield does not always compensate for risk** in Grade F, where losses (10%+) erode the high interest yield.
7. **Portfolio Volume is healthy ($163M)**, but quality needs optimization.
8. **Strategic exclusion** is more effective than broad rate hikes for risk management.

Recommendations

- **Recommendation 1:** Cap Debt-to-Income (DTI) ratio at **25%** for all **60-month** loan applications.
    - *Insight:* High DTI borrowers on long terms default at significantly higher rates.
    - *Business Impact:* Reduces overall default rate; feasible to implement in underwriting rules.
- **Recommendation 2:** Review pricing or tighten criteria for **Grade F** and **House** loans.
    - *Insight:* These segments have default rates (10.34% and 3.97%) that threaten profitability.
    - *Business Impact:* Improves net margin by avoiding "bad" revenue.

Impact Estimation

- **Reduce Risk:** Implementation of the DTI cap and Grade F review is projected to lower the portfolio default rate from **1.11%** to approximately **0.90%**.
- **Principal Protection:** By avoiding these high-probability defaults, the firm preserves an estimated **$3,400,000** in principal capital per annum (based on the $163M portfolio size).
- **Improve Efficiency:** Automated DTI flags will reduce manual underwriting time for high-risk applications.

Limitations

- **Data Issues:** The dataset is limited to Q1 2018; economic conditions may have shifted.
- **Assumption Risks:** Assumes that past default patterns (e.g., in Grade F) will persist in the future.
- **What cannot be concluded:** We cannot definitively predict the impact of external macro-economic shocks (e.g., recession) on Grade A defaults.

15. Future Scope

- **What more analysis can be done:** Develop a Logistic Regression or Random Forest model to predict individual loan default probabilities.
- **What new data is needed:** Real-time credit bureau data and borrower employment history validation to refine DTI calculations.

Conclusion

The analysis of the **$163M** portfolio identifies specific, actionable levers to improve profitability. By addressing the concentration of risk in **Grade F**, **60-month terms**, and **High DTI** borrowers, the firm can reduce its default rate to **~0.9%** and protect **$3.4M** in capital, achieving a more sustainable balance between risk and reward.

Appendix

- **Data Dictionary:** Link
- **Extra Charts:** Distribution of Loan Amounts, Interest Rate vs. Grade Scatter Plot.

18. Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Abhigya Sachdeva	medium	medium	high	high	low	low	Dashboard lead
Udit Jain	medium	low	medium	medium	low	low	Strategy lead
Yashpal	medium	high	medium	high	low	low	Analysis lead
Karan Chhillar	high	high	low	low	low	high	Data lead
Rishav Dewan	medium	low	low	low	high	medium	Ppt lead
Rishiwant Kumar Maurya	medium	low	medium	low	low	medium	Project lead

Declaration: We confirm that the above contribution details are accurate and verifiable through version history and submitted artifacts.