Bank Loan Analysis – Project Report

Project Title: Bank Loan Analysis
Prepared By: *Rudra Pratap Singh*
Tools Used: Python, Jupyter Notebook, Pandas, NumPy, Matplotlib, Seaborn
Objective:
To analyze loan applications, funded amounts, repayment behavior, and borrower risk patterns in order to help the bank improve its underwriting strategy, reduce defaults, and optimize portfolio performance.

Executive Summary

This project analyzes the bank's loan portfolio to understand borrower behavior, repayment performance, risk distribution, and revenue impact. The bank receives loan applications from a diverse customer base—varying in state, employment length, income levels, loan purpose, and home ownership.

Through data cleaning, exploration, KPIs, and visual insights, this report provides a complete view of the loan lifecycle—from application to funding and repayment.

Key outputs include:

- Evaluation of **Monthly Trends**, **State-wise performance**, and **Borrower risk segments**
- Identification of **Good Loans vs. Bad Loans**
- Portfolio health metrics such as **Funded Amount**, **Received Amount**, **Average Interest Rate**, and **DTI**
- Actionable recommendations for improving underwriting, customer segmentation, and risk mitigation.

Business Problem / Problem Statement

The bank is receiving thousands of loan applications from customers across different states, income groups, employment backgrounds, and loan purposes. However, the bank does not have a clear understanding of how these loans are performing, which customers are reliable, or where the major risks exist.

This lack of clarity makes it difficult to:

- Identify high-risk borrowers
- Understand repayment and default behavior
- Track monthly performance and profitability
- Detect regional and seasonal trends
- Improve underwriting and loan approval decisions

Therefore, a detailed analysis of the loan portfolio is required to evaluate loan quality, track key KPIs, compare Good vs. Bad loans, and understand borrower behavior. This analysis will help the bank reduce defaults, improve decision-making, and strengthen overall financial performance.

Dataset Description

The dataset contains **38,576 loan records** with details about borrowers, loan amounts, repayment behavior, and loan performance. It includes key information used to analyze Good vs. Bad loans and understand borrower risk.

Key Columns

- **Loan Details:** loan_amount, funded_amount, received_amount, interest_rate, term, issue_month, loan_status
- **Borrower Information:** emp_length, home_ownership, annual_income, dti
- **Loan Purpose:** purpose (e.g., debt consolidation, credit card), state
- **Performance Metrics:** total payment received, profit/loss, default_flag

Summary

This dataset provides all essential information to assess **loan demand**, **risk levels**, **borrower profiles**, and **portfolio performance**, enabling effective Bank Loan Analysis.

Data Cleaning & Preprocessing

The dataset used for this project contains detailed information on customer demographics, loan characteristics, financial indicators, repayment behavior, and loan status. It serves as the foundation for KPI tracking, risk assessment, and visualization-based insights.

To ensure accuracy, consistency, and reliability, the following **data cleaning, transformation, and preparation steps** were performed:

✓ Data Cleaning & Preparation Steps

- **Removal of missing, duplicate, and invalid entries** to ensure data quality and avoid skewed results.
- **Standardization of data formats** including dates, percentages, text categories, and loan terms for uniformity across the dataset.
- **Conversion of financial fields** such as interest rate, debt-to-income ratio (DTI), annual income, and funded amount into consistent numeric formats.
- **Derivation of new time-based features** including *Month*, *Year*, and *Month Name* to support trend analysis and monthly reporting.
- **Classification of loans into performance categories**, specifically:
    - **Good Loans** → Fully Paid
    - **Bad Loans** → Charged Off

- **Filtering out incomplete, irrelevant, or inconsistent records** to maintain dataset integrity.
- **Creation of aggregated datasets** for generating KPIs, trend charts, geographical insights, term-based comparisons, and customer segmentation.

Exploratory Data Analysis (EDA)

(Includes BRD 1 KPIs, Good vs Bad Loan insights, and BRD 2 Visualization Requirements)

A. BRD 1 — KPI Requirements

1. Total Loan Applications

- **Total Loan Applications:** *38,576*
  Insight:
  A high number of applications indicates **strong market demand** and wide customer reach. This provides the bank with a robust pipeline for future lending and potential revenue growth.

2. Total Funded Amount

- **Total Funded Amount:** *$435.76 Million*
  Insight:
  This represents the total capital deployed by the bank. It reflects the **overall exposure to credit risk** and the scale of the bank's lending operations. Higher deployment suggests an active loan market and strong lending activity.

3. Total Amount Received

- **Total Amount Received:** *$473.07 Million*
  Insight:
  The bank has collected **$37.31 Million more** than the funded amount ($473.07M − $435.76M). This indicates strong repayment behavior and good portfolio performance before accounting for charge-offs and operational expenses.

4. Average Interest Rate

- **Average Interest Rate:** *12.05%*
  Insight:
  This interest rate suggests a **moderate-to-high risk profile** among borrowers. Higher interest rates typically compensate for increased credit risk and directly impact portfolio revenue.

5. Average Debt-to-Income Ratio (DTI)

- **Average DTI:** *13.33%*
  Insight:
  A DTI of 13.33% is **low and financially healthy**, indicating that borrowers are not heavily burdened by existing debts. Lower DTI levels are a positive sign and generally correlate with **higher repayment reliability**.

B. BRD 1 — Good Loan vs. Bad Loan KPIs

Good Loans (Fully Paid)

- **Good Loan Applications:** *33,243*
- **Good Loan Funded Amount:** *$370.22M*
- **Good Loan Total Received Amount:** *$435.79M*
- **Good Loan Application Percentage:** *86.18%*
- **Profit Generated:** *$65.56M*

Insight:
The bank's core lending portfolio is **highly profitable**, with over **86% of all loans performing successfully**. These good loans generated a strong profit of **$65.56 Million**, demonstrating the effectiveness of current underwriting for stable customer segments.

However, this profitability is **heavily concentrated** in:

- A **single loan purpose** (Debt Consolidation), and
- A **single state** (California)

This creates a **concentration risk** where any adverse shift in borrower behavior or economic conditions in these segments could significantly impact total revenue. While good loans drive strong performance, this dependency highlights the need to **diversify loan purposes and geographic reach**.

Bad Loans (Charged Off / Defaulted)

- **Bad Loan Applications:** *5,333*
- **Bad Loan Funded Amount:** *$65.53M*
- **Bad Loan Total Received Amount:** *$37.28M*
- **Bad Loan Application Percentage:** *13.82%*
- **Total Loss:** *$28.25M*

Insight:
Bad loans represent **13.82% of the portfolio**, resulting in a **direct loss of $28.25 Million**. While the bank manages to recover more than half of the amount funded to this high-risk segment, the remaining loss substantially reduces net profitability.

The high number of defaults (5,333 loans) highlights:

- Weak borrower profiles such as **low employment stability**
- **Renters** (higher risk than homeowners)
- Overexposure to **high-risk loan purposes** like debt consolidation

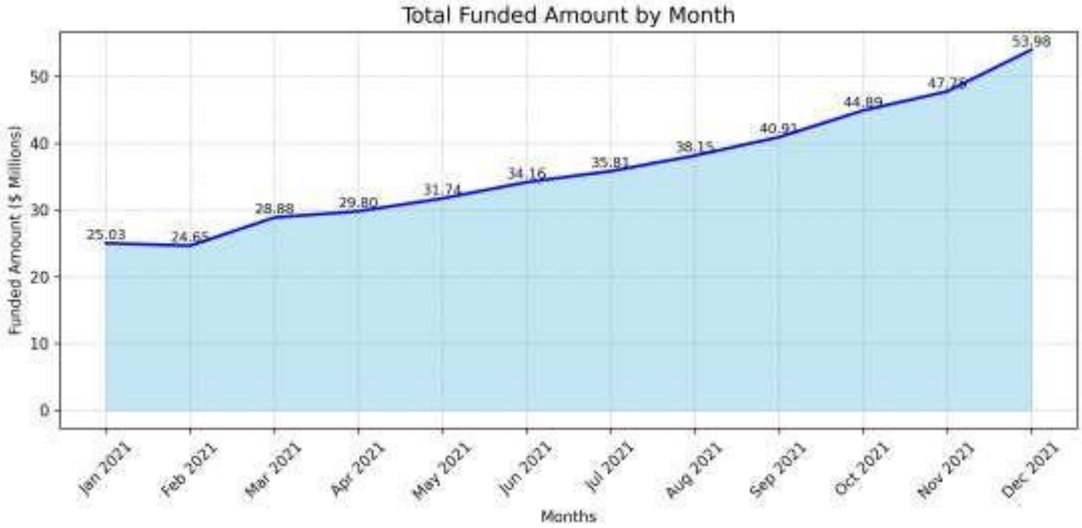
This emphasizes an urgent need for:

- **Stricter underwriting policies**
- **Risk-based pricing adjustments**
- **Higher credit quality requirements** for vulnerable segments

The goal should be to **reduce the default rate**, thereby protecting future profits and improving overall portfolio health.

C. BRD 2 — Visualization Requirements & Chart Insights

1. Total Funded Amount by Month

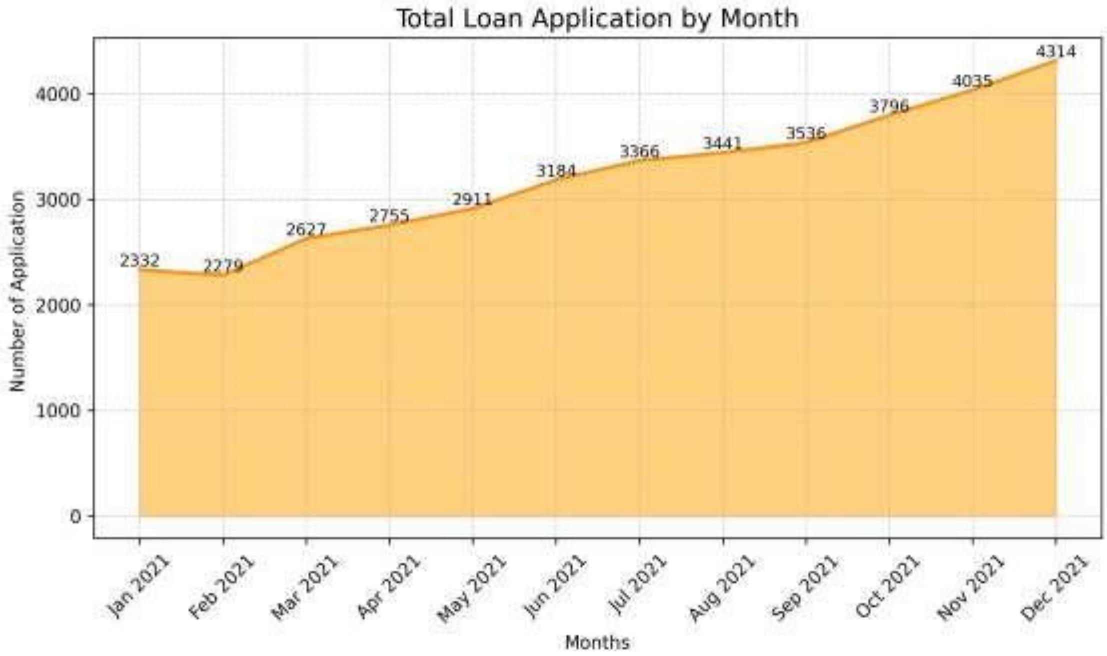


Insight:
Loan funding remains stable throughout the year, with a strong surge in **December**, indicating peak lending activity and the bank's maximum capital deployment during year-end.

2. Total Amount Received by Month

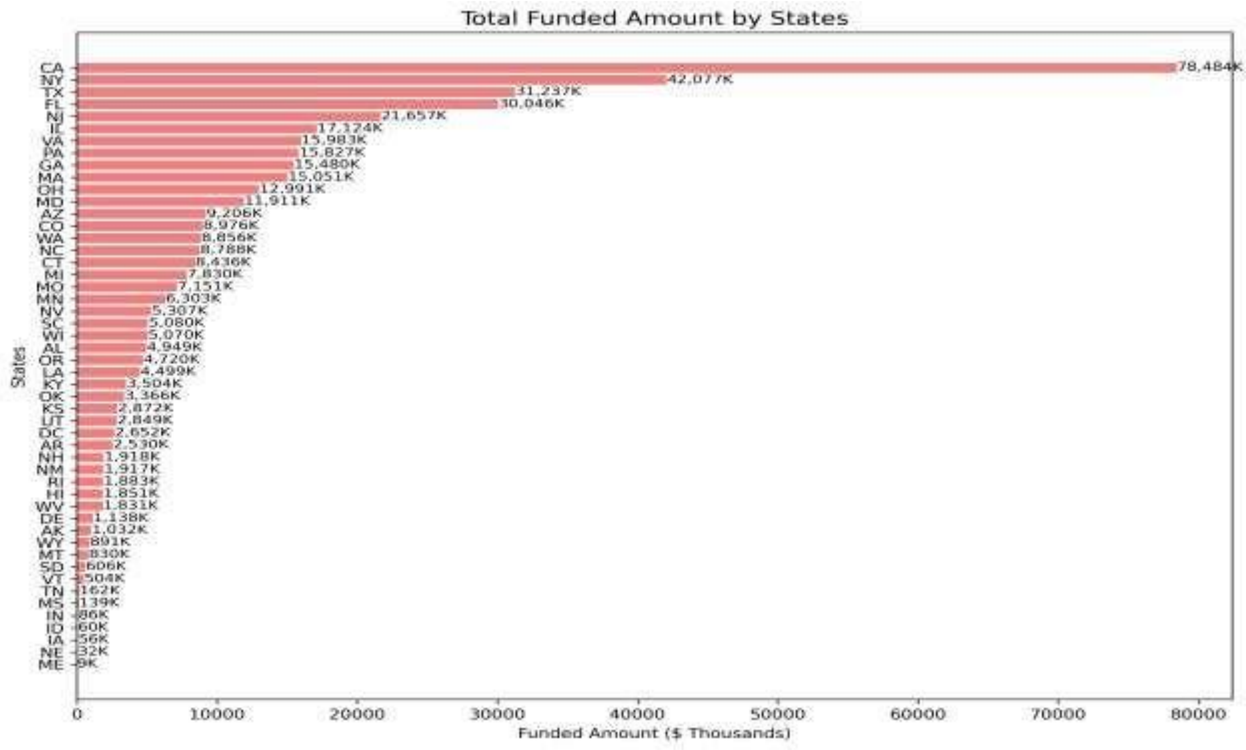
Insight:
Repayments closely track the funded amounts, showing a healthy loan recovery cycle. December records the **highest collection volume**, reflecting strong portfolio performance and repayment discipline.

3. Total Loan Applications by Month



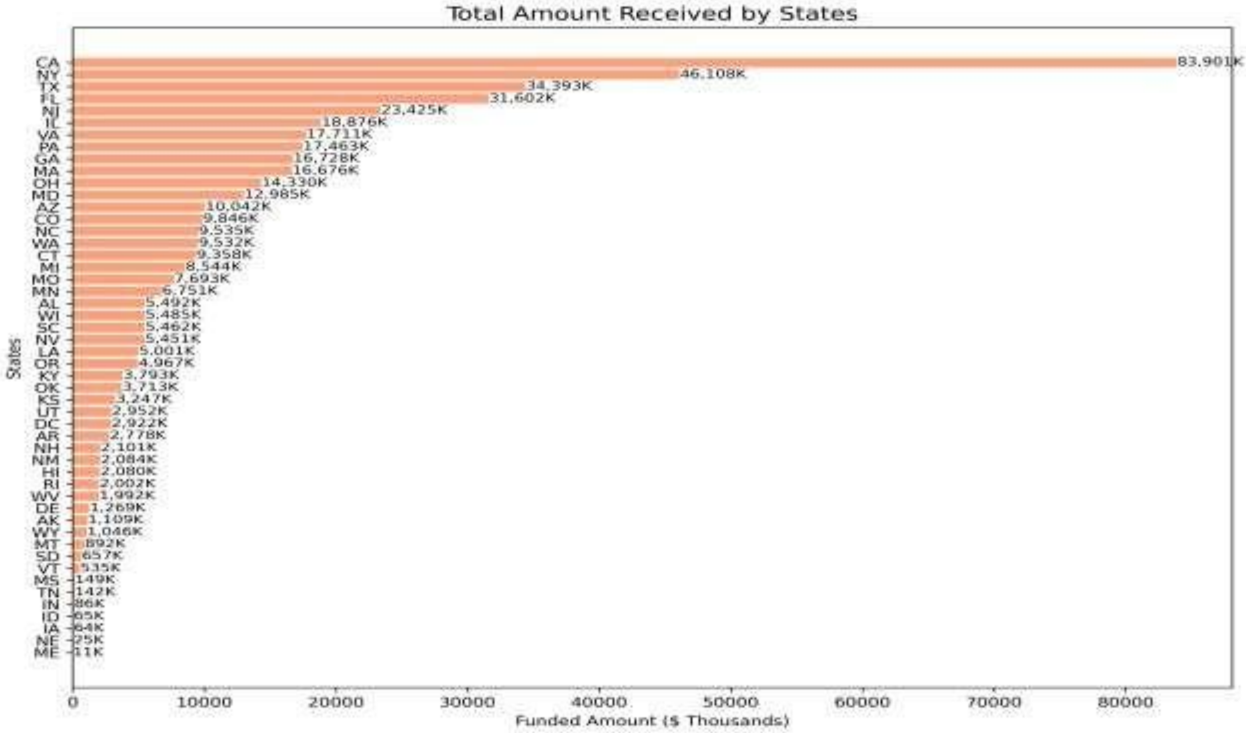
Insight:
Loan application volume remains consistently strong, with a clear spike in **December**. This indicates increased customer demand—possibly due to year-end financial planning, festive spending, or consolidation needs.

4. Total Funded Amount by State



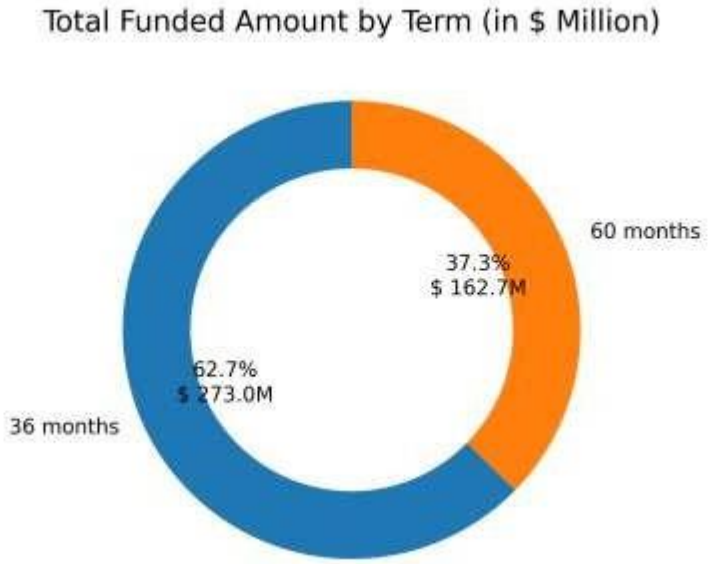
Insight:
The bank's funding is **highly concentrated in California**, making it the primary driver of total funded volume. While profitable, this creates **significant regional concentration risk**.

5. Total Amount Received by State



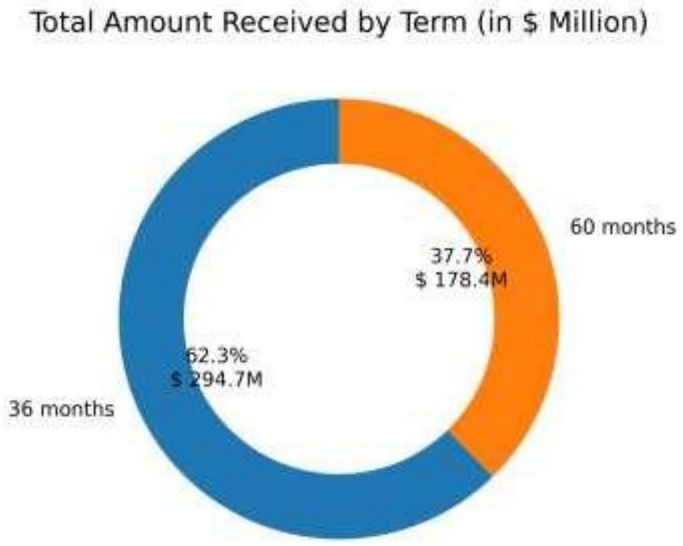
Insight:
California continues to dominate in total collections, indicating strong borrower performance in the region. However, over-reliance on a single state increases exposure to localized economic downturns.

6. Total Funded Amount by Loan Term (36 vs. 60 Months)



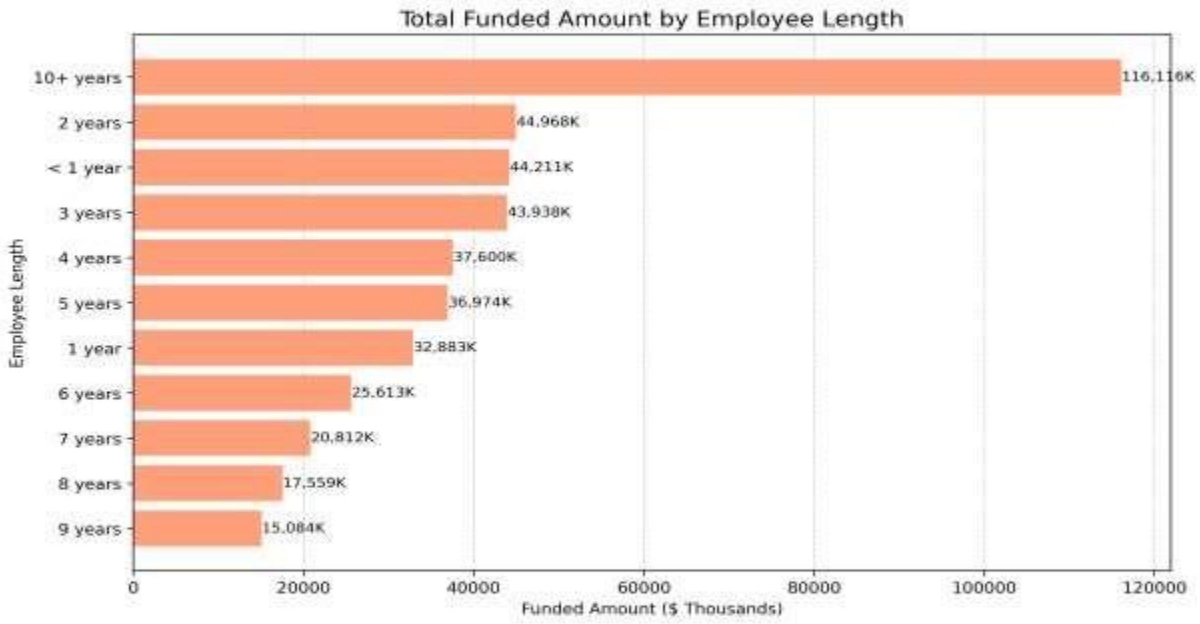
Insight:
Most funding is directed toward **36-month loans**, indicating the bank's preference for shorter-term credit—reducing default risk and improving liquidity.

7. Total Amount Received by Loan Term



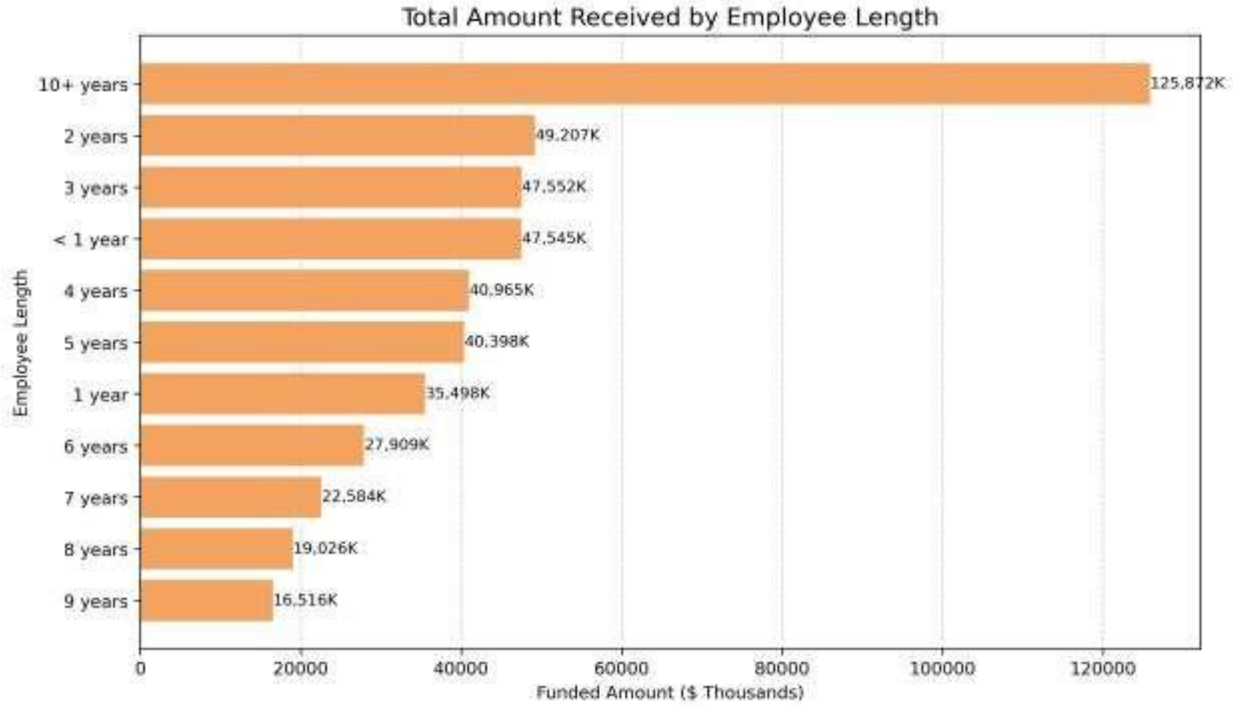
Insight:
Revenue is primarily generated from **36-month loans**, confirming them as the bank's highest-performing and most profitable repayment segment.

8. Total Funded Amount by Employment Length



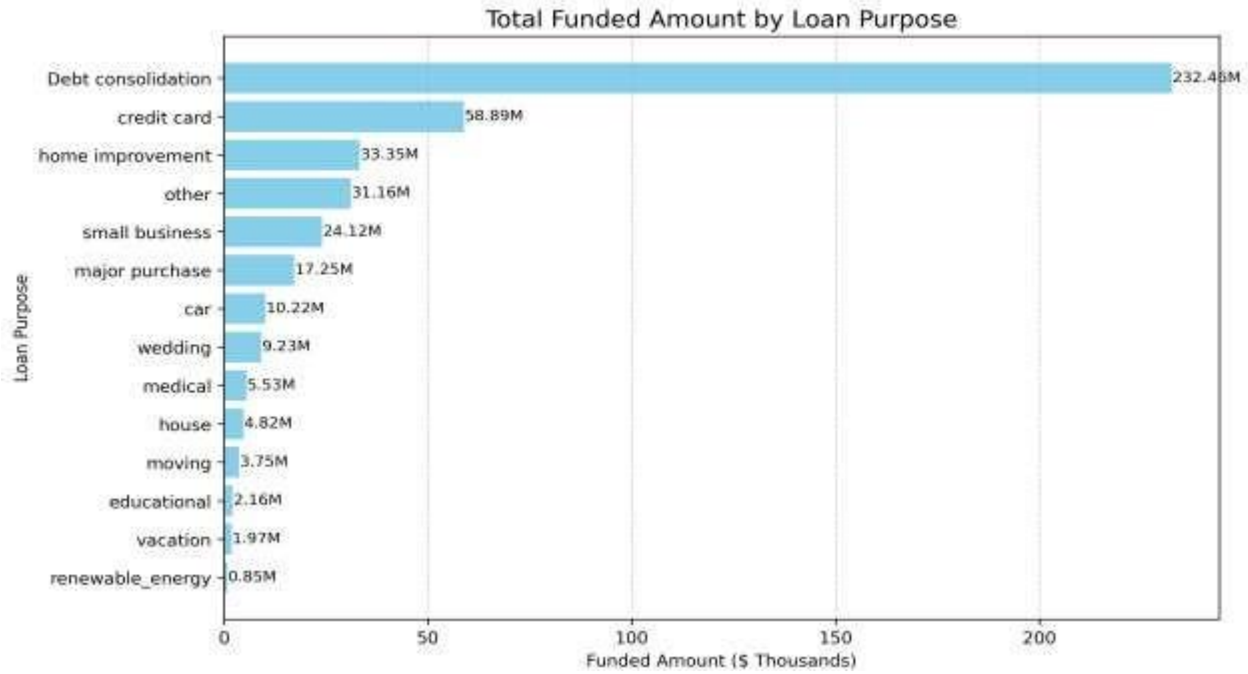
Insight:
Borrowers with **10+ years of employment** receive the highest funding—aligning with lower risk. However, funding is also extended to high-risk groups like **<1-year employees**, suggesting the need for stronger risk filtering.

9. Total Amount Received by Employment Length



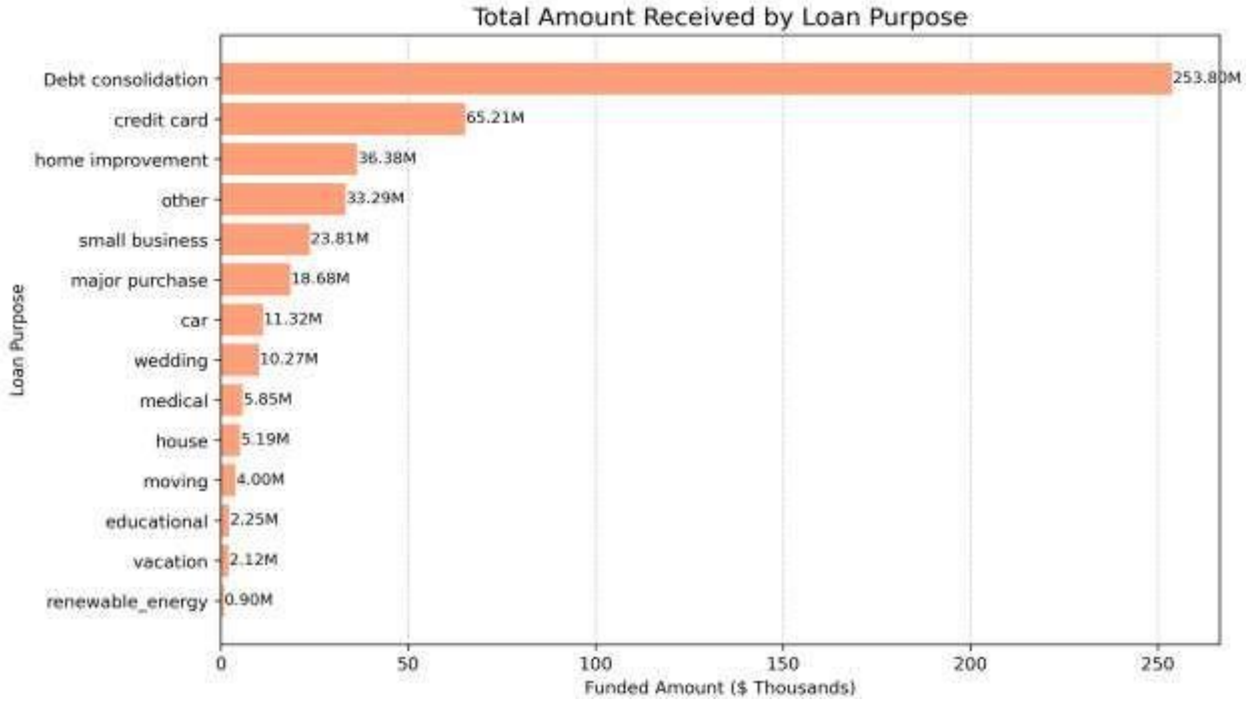
Insight:
Borrowers with **10+ years of work experience** generate the strongest and most reliable repayment inflows, proving them to be the bank's safest and most profitable customer group.

10. Total Funded Amount by Loan Purpose



Insight:
Funding is overwhelmingly concentrated in **Debt Consolidation**, making it the bank's core business line. This creates a **single loan-purpose dependency**, increasing business risk.

11. Total Amount Received by Loan Purpose



Insight:
Debt Consolidation accounts for the largest share of total repayments, but this heavy reliance means revenue is vulnerable if performance within this segment deteriorates.

12. Total Funded Amount by Home Ownership

Insight:
The bank's largest funding is directed toward **Mortgage holders**, the lowest-risk category. However, a significant amount is also extended to **Renters**, indicating notable exposure to higher-risk borrower segments.

13. Total Amount Received by Home Ownership

Insight:
Mortgage holders generate the highest repayment inflows, confirming them as the most dependable borrower group. This suggests home ownership is a strong positive predictor of loan performance.

Key Insights Summary

The Bank Loan Analysis reveals several important findings related to portfolio performance, borrower behavior, and risk exposure. The insights highlight both strengths and vulnerabilities within the bank's lending strategy.

1. Portfolio Performance & Demand Trends

- The bank processed **38,576 loan applications**, reflecting strong customer demand throughout the year.
- Monthly trends show consistent application and funding activity, with a **significant surge in December**, suggesting seasonal or year-end financial behavior.
- Total funded amount reached **$435.76M**, while total amount received amounted to **$473.07M**, generating a positive cashflow margin.

2. Profitability & Repayment Efficiency

- The bank generated an estimated **$37.31M** in gross returns (before losses), indicating efficient repayment behavior across the portfolio.
- Short-term (36-month) loans contribute the highest share of total revenue, confirming the profitability and lower risk of shorter repayment cycles.
- Repayment performance is strongest among **Mortgage holders**, **long-term employees (10+ years)**, and **Debt Consolidation borrowers**, who form the bank's most reliable segments.

3. Good Loan vs. Bad Loan Portfolio Quality

- **Good Loans account for 86.18%** of the portfolio (33,243 loans), generating **$435.79M** in repayments and producing **$65.56M profit**.

- **Bad Loans represent 13.82%** of the portfolio (5,333 loans), resulting in **$28.25M in losses**.
- Although the Good Loan segment drives strong profitability, the Bad Loan segment contributes significant financial drag—indicating the need for stronger risk controls.

4. Geographic & Customer Concentration Risks

- **California (CA)** dominates both funding and repayment activity, making it the bank's primary revenue engine.
  However, this introduces **high regional concentration risk**, exposing the bank to localized economic fluctuations.
- The portfolio is heavily concentrated in **Debt Consolidation loans**, which, although profitable, increases vulnerability if this segment experiences rising default rates.

5. Borrower Profile Insights

- Borrowers with **10+ years of employment** are the most low-risk, generating the highest total repayment amounts.
- Renters and borrowers with **<1 year employment** show elevated risk, indicating areas for underwriting improvement.
- Mortgage holders prove to be the most stable and profitable borrower group, confirming the importance of asset-backed financial behavior.

6. Loan Structure & Term Preferences

- The bank predominantly issues **36-month loans**, which outperform 60-month loans in both repayment volume and reliability.
- Preference for shorter terms supports strong liquidity, lower default risk, and healthier cash flow cycles.

Overall Insight

The bank's lending strategy is profitable and driven by strong customer demand, but significant risks arise from **portfolio concentration**, **high-risk borrower categories**, and **geographic dependency**. Strengthening underwriting policies and diversifying the loan portfolio will be critical for future stability and growth.

Business Recommendations

1. Strengthen Underwriting for High-Risk Segments

The 13.82% bad-loan segment is responsible for **$28.25M in direct losses**, indicating weak risk controls.
Recommendation:

- Implement stricter credit score thresholds for high-risk profiles.
- Introduce additional income verification for borrowers with high DTI (Debt-to-Income).

- Use ML-based risk scoring to flag early-warning patterns.

2. Diversify Loan Portfolio to Reduce Concentration Risk

Current loan performance relies heavily on **Debt Consolidation** and borrowers from **California**, creating geographical and purpose concentration risk.
Recommendation:

- Promote loan products in under-represented states.
- Expand marketing efforts for Home Improvement, Medical, and Personal loans.
- Set internal caps for over-concentrated loan categories.

3. Enhance Recovery Strategy for Charged-Off Loans

While some capital is being recovered, the recovery rate is insufficient to offset the high loss volume.
Recommendation:

- Strengthen collections through data-driven prioritization of recoverable accounts.
- Partner with specialized recovery agencies.
- Implement automated repayment reminders and settlement campaigns.

4. Refine Interest Rate Pricing Based on Risk

A uniform pricing strategy may be underpricing risk in higher-default categories.
Recommendation:

- Use risk-based pricing to ensure high-risk borrowers pay for the added risk.
- Introduce dynamic APR models using credit score + historical performance + loan purpose.

5. Improve Customer Profiling With Behavioral Analytics

Better behavioral understanding can significantly reduce default probability.
Recommendation:

- Track payment patterns, spending behavior, and transaction flows to pre-identify potential defaulters.
- Develop personalized credit-limit and repayment-plan recommendations.

6. Strengthen Early-Delinquency Intervention

Early action can prevent many loans from becoming charge-offs.
Recommendation:

- Trigger automated alerts after 3–5 days of missed payments.
- Offer one-click payment rescheduling options.
- Launch hardship-assistance programs for loan restructuring.

7. Continuously Optimize the Underwriting Model

The current model is profitable, but improvements will reduce loss and increase stability.
Recommendation:

- Evaluate model performance quarterly.
- Incorporate new variables like employment stability, banking history, spending category ratios.
- Monitor key drift indicators to avoid performance degradation.

Conclusion

The loan portfolio demonstrates a **strong core performance**, with over **86% of loans classified as Good Loans** and generating a substantial **$65.56 million profit**. This confirms that the bank's current underwriting framework is fundamentally effective and capable of delivering consistent, large-scale profitability.

However, the presence of **13.82% Bad Loans**—resulting in **$28.25 million in losses**—reveals significant **risk concentration and exposure**, particularly in specific geographies and loan purposes. These losses reduce the overall profitability and highlight the urgent need for a stronger, more diversified risk strategy.

By implementing **stricter underwriting controls**, **expanding the loan mix**, **enhancing collections**, and **leveraging advanced analytics**, the bank can substantially reduce default-related losses while maintaining the strong performance of its good-loan portfolio.

Overall, the analysis reinforces that the business is on a solid foundation but must take targeted actions to improve stability, minimize risk, and unlock long-term, scalable financial growth.