

**Title: *Bank Loan Data Analytics: Risk, Repayment & Trends***

**(with MySQL and MS Excel)**

**Name**:Shivika Gupta

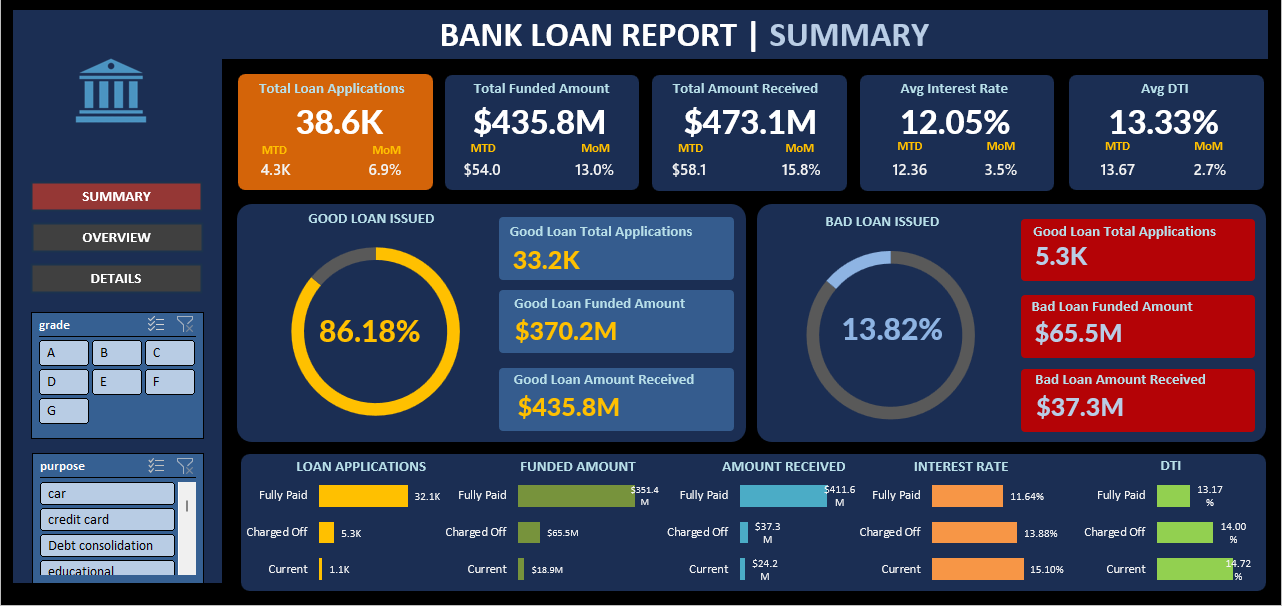
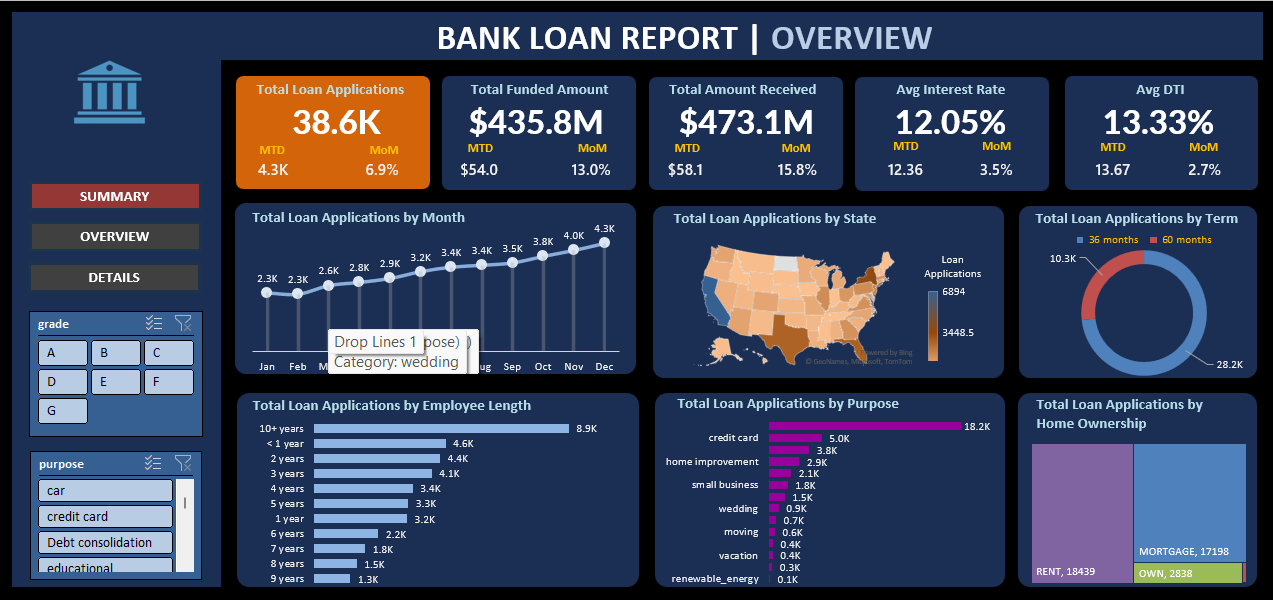
**Institution:**Gokhale Institute of Politics and Economics

**GitHub:** <https://github.com/shivikagupta21>

**Linkedln:** <https://www.linkedin.com/in/shivika-gupta-6a5ab3217/>

**Dataset Size:** 38,576 loan records

**Date:** July 8, 2025

### 

### 

### 

### **Introduction**

This project presents a comprehensive analysis of a bank's loan performance. By leveraging SQL for robust data transformation and MS Excel for dynamic visualization and reporting, this project aims to uncover key business insights related to lending activities, risk assessment and borrower behavior. The report addresses the critical need to monitor and assess the bank's lending activities to make data-driven decisions, track the health of the loan portfolio and identify trends that can inform lending strategies. Bank loans are a crucial financial tool and understanding their performance is essential for both the bank and its borrowers. This dataset allows us to:

* **Assess risk** associated with lending to individuals and businesses, evaluating creditworthiness and predicting default probabilities.
* **Support decision-making** when evaluating loan applications using data-driven insights to approve or deny requests.
* **Manage loan portfolios** effectively, monitoring their health and identifying underperforming loans.
* **Detect fraudulent loan applications** by identifying unusual patterns or inconsistencies.
* Gain **customer insights** into behavior, preferences and needs allowing for tailored loan products and marketing.

### 

### **Research Objective**

The primary objective of this project is to analyze bank loan performance by focusing on:

* **Key Performance Indicators (KPIs):** Calculating and monitoring total loan applications, funded amounts, amounts received, average interest rates and average Debt-to-Income (DTI) ratios including Month-to-Date (MTD) and Month-over-Month (MoM) changes.
* **Good Loan vs. Bad Loan Analysis:** Evaluating the performance of lending activities by distinguishing between 'Good Loans' (Fully Paid, Current) and 'Bad Loans' (Charged Off) including their respective percentages, counts, funded amounts and received amounts.
* **Loan Status Grid View:** Providing a comprehensive overview of lending operations through a grid categorized by 'Loan Status' offering insights into key indicators for different loan states.
* **Trend and Distribution Analysis:** Visualizing key loan related metrics and trends over time (Monthly Trends by Issue Date), by geographic region (Regional Analysis by State), loan term (Loan Term Analysis), employment length (Employee Length Analysis), loan purpose (Loan Purpose Breakdown) and home ownership status (Home Ownership Analysis).
* **Repayment Behavior Analysis:** Delving deeper into borrower repayment patterns by analyzing recovery ratios, segmenting repayers (Fully, Partial, Low) and understanding repayment behavior across different loan grades and purposes.

### **Dataset and Columns Introduction**

**Dataset Overview:**

* **Table Name:** bank\_loan\_data
* **Rows:** 35,577
* **Columns:** 24

**Key Columns and their Descriptions:**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| **id** | Integer | Unique identifier for each loan application or account. |
| **address\_state** | Text | Indicates the borrower’s location. |
| **application\_type** | Text | Specifies if the application is for an individual, business or other. |
| **emp\_length** | Text | Insights into the borrower’s employment stability. |
| **emp\_title** | Text | Specifies the borrower’s occupation or job title. |
| **grade** | Text | Represents a risk classification assigned to the loan. |
| **home\_ownership** | Text | Indicates the borrower’s housing status. |
| **issue\_date** | Date | Marks the loan’s origination date. |
| **last\_credit\_pull\_date** | Date | Records when the borrower’s last credit report was assessed. |
| **last\_payment\_date** | Date | Marks the most recent loan payment received. |
| **loan\_status** | Text | Current state of the loan. |
| **next\_payment\_date** | Date | Estimates the date of next loan payment. |
| **member\_id** | Integer | A unique identifier assigned to each bank customer. |
| **purpose** | Text | Specifies the reason for the loan. |
| **sub\_grade** | Text | Refines the risk assessment within a grade. |
| **term** | Text | Defines the duration of the loan in months. |
| **verification\_status** | Text | Whether the borrower’s financial status has been verified. |
| **annual\_income** | Integer | Borrower’s total yearly earnings. |
| **dti** | Numeric | Borrower’s debt burden relative to income. |
| **installment** | Numeric | Fixed monthly payment amount for loan repayment. |
| **int\_rate** | Numeric | Annual cost of borrowing expressed as a percentage. |
| **loan\_amount** | Numeric | Total borrowed sum, principal amount. |
| **total\_acc** | Integer | Number of accounts of the borrower/ loan applicant. |
| **total\_payment** | Numeric | Total amount received from the borrower. |

### **Theory**

Understanding bank loan processes and data analysis reasons is crucial for interpreting the report:

* **Data Collection:** Banks collect loan data through various channels and processes, including loan applications, credit reports from credit bureaus, internal records of loan transactions, online portals and third-party data sources.
* **Loan Granting Process:** The process begins with a loan application followed by an application review, identity verification, a crucial credit check, income verification, Debt-to-Income (DTI) ratio check, employment verification and collateral assessment. This leads to a comprehensive risk assessment, loan approval or denial, the signing of a loan agreement, disbursement of funds and finally ongoing repayment and monitoring.
* **Reasons for Data Analysis:** Banks analyze loan data for several critical reasons: risk assessment, decision-making, portfolio management, fraud detection, regulatory compliance, customer insights, profitability analysis, market research, credit risk management and customer retention.

**Data Cleaning and Transformation**

Before analysis and dashboarding, the raw data underwent several crucial processing steps using SQL:

1. **Column Names and Data Types Check:** Ensured all columns were correctly typed.
2. **Standardize Date Fields:** Converted issue\_date, last\_payment\_date, last\_credit\_pull\_date and next\_payment\_date to DATE type.
3. **Handle Missing or Null Values:** Checked for NULLs in critical fields like loan\_status and funded\_amount.
4. **Clean emp\_length Column:** Standardized mixed values (e.g., '< 1 year', '10+ years') into a numeric emp\_years column.
5. **Remove Duplicates:** Verified and confirmed no duplicate records based on id.
6. **Trim Whitespace in Text Columns:** Removed leading/trailing whitespace from emp\_title and purpose.
7. **Create Derived Columns:**
   * **Year-Month Column (issue\_month):** Extracted the year and month from issue\_date in 'YYYY-MM' format.
   * **Good/Bad Loan Tag (loan\_type):** Categorized loans as 'Good' (Fully Paid, Current) and 'Bad' (Charged Off)

#### **Key Performance Indicators (KPIs)**

* **Total Loan Applications:** 38.6K
  + MTD: 4.3K
  + MoM Change: 6.9%
* **Total Funded Amount:** $435.8M
  + MTD: $54.0M
  + MoM Change: 13.0%
* **Total Amount Received:** $473.1M
  + MTD: $58.1M
  + MoM Change: 15.8%
* **Average Interest Rate:** 12.05%
  + MTD: 12.36%
  + MoM Change: 3.5%
* **Average Debt-to-Income Ratio (DTI):** 13.33%
  + MTD: 13.67%
  + MoM Change: 2.7%

#### **Loan Performance Analysis: Good vs. Bad Loans**

**Good Loan KPIs:** Loans with a status of 'Fully Paid' and 'Current' are categorized as 'Good Loans'.

* **Good Loan Application Percentage:** 86.18%
* **Good Loan Applications:** 33.2K
* **Good Loan Funded Amount:** $370.2M
* **Good Loan Total Received Amount:** $435.8M

**Bad Loan KPIs:** Loans with a status of 'Charged Off' are categorized as 'Bad Loans'.

* **Bad Loan Application Percentage:** 13.82%
* **Bad Loan Applications:** 5.3K
* **Bad Loan Funded Amount:** $65.5M
* **Bad Loan Total Received Amount:** $37.3M

#### **Loan Status**

| **Loan Status** | **Total Applications** | **Total Funded Amount** | **Total Amount Received** | **Average Interest Rate** | **Average DTI** |
| --- | --- | --- | --- | --- | --- |
| Charged Off | 5,333 | $65,532,225 | $37,284,763 | 13.87% | 14.00% |
| Fully Paid | 32,145 | $351,358,350 | $411,586,256 | 11.64% | 13.16% |
| Current | 1,098 | $18,866,500 | $24,199,914 | 15.09% | 14.72% |

#### **Visual Trends and Key Distributions**

#### The dashboard provides visual insights into critical loan metrics and trends aiding data driven decision making through various chart types.

#### **Monthly Trends:** Loan applications, funded amounts and received amounts generally show an increasing trend from January to July 2021.

#### **Regional Analysis (State):** California (CA) significantly leads in applications, total funded and total received amounts highlighting its strong lending activity.

#### **Loan Term Analysis:** The majority of loan applications, funded amounts and received amounts are concentrated in 36-month terms indicating a preference for shorter loan durations.

#### **Employment Length:** Borrowers with '10+ years' of employment history represent the largest segment for loan applications, funded and received amounts followed by '< 1 year' and '3 years'.

#### **Loan Purpose Breakdown:** "Debt consolidation" is the predominant reason for loans accounting for the highest numbers in applications, funded and received amounts significantly outpacing other purposes like "credit card" and "home improvement".

#### **Home Ownership Analysis:** "RENT" and "MORTGAGE" constitute the largest categories for loan applications, funded and received amounts reflecting the primary housing statuses of borrowers.

#### 

### **Loan Repayment Behavior Analysis**

This section delves into how borrowers repay loans, moving beyond simple default statuses to understand consistency, speed and patterns offering richer behavioral insights for the bank.

Key findings include:

* The **Recovery Ratio** (total payment divided by loan amount) provides insights into actual funds recovered.
* **Average Recovery Ratios by Loan Status** show "Charged Off" loans recover approximately 57.5% , while "Fully Paid" loans recover around 116.5% and "Current" loans about 129.4%.
* Borrowers are segmented by **Repayment Behavior**: 32,992 are "Fully Repaid" (recovery ratio ≥ 0.95) , 3,054 are "Partial Repayers" (recovery ratio between 0.5 and 0.95) and 2,530 are "Low Repayers".
* **Average Recovery Ratios by Loan Grade** indicate that Grade G has the highest average recovery at approximately 113.8%, followed closely by F and E while Grade A has the lowest at about 105.3%.
* **Average Recovery Ratios by Loan Purpose** show "wedding" and "car" loans with the highest average recovery at 112% and 111% respectively while "moving" loans have an average recovery of 108%.

**Additional Loan Insights**

Additional analyses reveal deeper trends and risk factors within the loan portfolio:

* **Top 5 States by Funded Amount:** The top states by funded amount are CA ($78.48M), NY ($42.08M), TX ($31.24M), FL ($30.05M) and NJ ($21.66M).
* **Average Loan Amount by Purpose:** "Small business" loans have the highest average loan amount ($13,582.83), followed by "house" ($13,182.86) and "Debt consolidation" ($12,762.69).
* **Default Rate by Grade:** Grade G has the highest default rate (31.31%) while Grade A has the lowest (5.70%).
* **Highest Risk Subgrades:** The top 5 highest risk subgrades are F5 (46.09%), G3 (39.58%), G2 (35.90%), G5 (33.33%) and F4 (31.90%).
* **Bad Loan Trends Over Time:** The number of charged-off loans increased monthly from January 2021 (309) to July 2021 (454).
* **Top 5 Job Titles Associated with Charged Off Loans:** The top job titles with charged-off loans include unspecified titles (266), US Army (29), Bank of America (24), Walmart (20) and AT&T (16).
* **Average Interest Rate by Home Ownership Status:** "RENT" has the highest average interest rate (12.30%) while "NONE" has the lowest (8.70%).
* **Most Common Loan Purpose Among High DTI Borrowers:** "Debt consolidation" (370 loans) is the most common purpose for borrowers with DTI > 0.25.
* **Loan Grade Distribution for 'Current' Loans Only:** For current loans, Grade B has the highest count (327) followed by C (257) and D (216).

**Strategic insights and Business recommendations**

1. **Risk Management Enhancement:**
   * **Focus on High-Risk Segments:** The analysis clearly identifies grades F and G and subgrades F5, G3, G2, G5 and F4 as having the highest default rates. Implement more stringent credit assessment criteria and potentially higher interest rates for applicants falling into these categories.
   * **Monitor Charged Off Loan Trends:** The trend of bad loans increasing from January to July 2021 (309 to 454) suggests a need for deeper investigation into the factors contributing to this rise. Proactive measures such as earlier intervention for at-risk loans could help mitigate further losses.
   * **Evaluate Employment Titles with High Defaults:** Job titles like "US Army," "Bank of America," "Walmart" and "AT&T" show a notable number of charged off loans. While these might be large employers, understanding specific roles or internal factors contributing to defaults within these groups could inform risk models.
2. **Portfolio Optimization & Product Strategy:**
   * **Diversify Loan Terms:** While 36-month loans are predominant, actively promoting 60-month terms for suitable borrowers could help diversify the portfolio and potentially increase total funded and received amounts over longer periods.
   * **Targeted Marketing for Loan Purposes:** "Debt consolidation" is the most common loan purpose particularly among high DTI borrowers. Develop tailored loan products or financial counseling services specifically for debt consolidation to better serve this large segment while managing associated risks.
   * **Re-evaluate Home Ownership Impact:** Borrowers with 'RENT' home ownership have a slightly higher average interest rate compared to 'OWN' or 'MORTGAGE'. This indicates perceived higher risk. Consider if there are specific segments within renters that are less risky or if targeted educational resources could improve their financial standing.
3. **Operational Efficiency and Customer Retention:**
   * **Regional Focus:** California leads significantly in total funded amounts. This suggests strong market demand. Further investment in resources and marketing in top-performing states like CA, NY, TX and FL could yield higher returns.
   * **Enhance Repayment Support:** For "Low Repayer" and "Partial Repayer" segments (totaling 5,584 borrowers), implement targeted communication strategies, flexible repayment options or financial literacy programs to improve recovery ratios before loans become "Charged Off."
   * **Grade-Based Recovery Strategies:** While higher grades generally show better recovery ratios, even "Good Loans" (Fully Paid, Current) have average recovery ratios above 1.0 (1.165 and 1.294 respectively) indicating that some borrowers pay back more than the principal (due to interest). This highlights the profitability of 'Good Loans'. Conversely, for 'Charged Off' loans with a 0.575 recovery ratio focus on optimizing collection efforts.

**Conclusion**

This comprehensive Bank Loan Report driven by SQL-based data processing and interactive MS Excel dashboards has provided valuable insights into the bank's lending trends, performance by loan status and borrower behavior patterns. Key findings include the significant volume of 'Good Loans', strong overall funded and received amounts and clear trends across various demographic and loan-specific dimensions. The analysis revealed the importance of continuous monitoring of MoM changes in key metrics to quickly identify shifts in performance. Furthermore, the detailed examination of 'Good' vs. 'Bad' loans coupled with repayment behavior analysis offers actionable intelligence for credit risk management. Understanding the default rates across different grades and the most common purposes for high-DTI loans allows for more precise risk modeling and targeted product development. By leveraging these data-driven strategies, the bank can make more informed decisions regarding loan origination, portfolio management, risk mitigation and customer engagement ultimately contributing to sustained growth and profitability in its lending operations.