**Title: Case Study: Strategic Lending Acquisition Analysis**

**By: Valerie Kivayiru Jarenga**

Contents

[Introduction 2](#_Toc193467755)

[Understanding the Data 2](#_Toc193467756)

[Analysis 3](#_Toc193467757)

[Data importing and cleaning 3](#_Toc193467758)

[SQL 6](#_Toc193467759)

[Visualization 12](#_Toc193467760)

[Conclusion 14](#_Toc193467761)

[Recommendations 15](#_Toc193467762)

[Links 17](#_Toc193467763)

**Table of figures**

[Figure 1Screenshot to show for missing values in dataset 5](#_Toc195401053)

[Figure 2Boxplot to show outliers in loan amount 5](#_Toc195401054)

[Figure 3Boxplot to show outliers in repayment amount 7](#_Toc195401055)

[Figure 4Screenshot of SQL queries 9](#_Toc195401056)

[Figure 5Screenshot of SQL queries 11](#_Toc195401057)

[Figure 6Screenshot of SQL queries 13](#_Toc195401058)

[Figure 7Screenshot of SQL queries 13](#_Toc195401059)

[Figure 8Screenshot of SQL queries 14](#_Toc195401060)

[Figure 9Screenshot of loan portfolio overview in Power BI 15](#_Toc195401061)

[Figure 10Screenshot of performance metrics in Power BI 16](#_Toc195401062)

[Figure 11Screenshot of risk analysis in Power BI 16](#_Toc195401063)

[Figure 12Screenshot of forecasting in Power BI 17](#_Toc195401064)

# Introduction

## This report presents an analysis of loan data from a potential acquisition aimed at expanding financial services to the unbanked and under-banked market. The objective is to evaluate the lending product’s performance, assess credit risk, and forecast profitability to support strategic decision-making. Key areas of focus include identifying product features, analyzing performance trends through Power BI, defining core metrics, developing a short-term profit/loss forecast, and recommending risk mitigation strategies and product improvements based on data insights.

## Understanding the Data

1. Disbursements Sheet

This sheet contains loan issuance details with the following columns:

customer\_id: Unique identifier for each customer.

disb\_date: Date when the loan was disbursed.

tenure: Loan duration ("7 days", "14 days", "30 days").

account\_num: Unique identifier for the loan account.

loan\_amount: Amount of the loan issued.

loan\_fee: Fee charged for the loan.

2. Repayments Sheet

This sheet records loan repayments with the following columns:

date\_time: Date and time of the repayment.

customer\_id: Unique identifier matching the one in the Disbursements sheet.

amount: Amount repaid.

rep\_month: The repayment month in YYYYMM format.

repayment\_type: Whether the payment was Automatic or Manual.

# Analysis

## Data importing and cleaning

I imported the data in Jupyter notebook for analysis and the data had no missing values or duplicates as shown in the figure below.

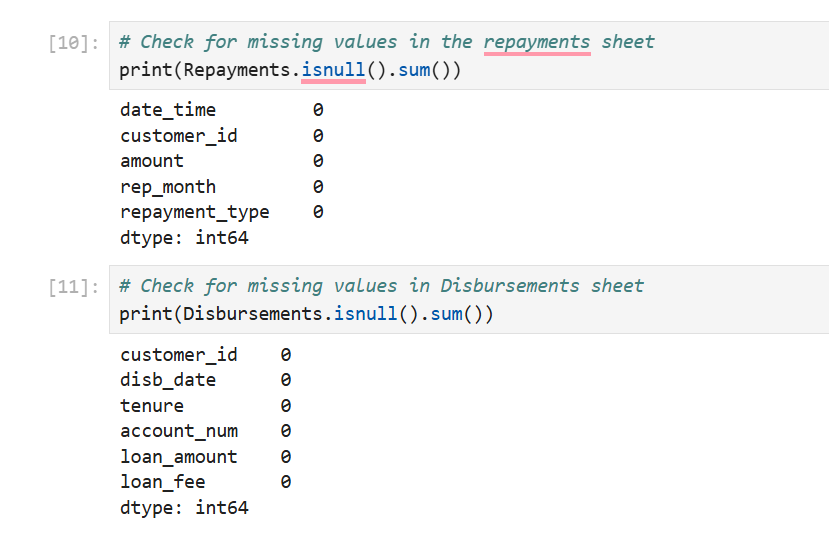


Figure 1Screenshot to show for missing values in dataset

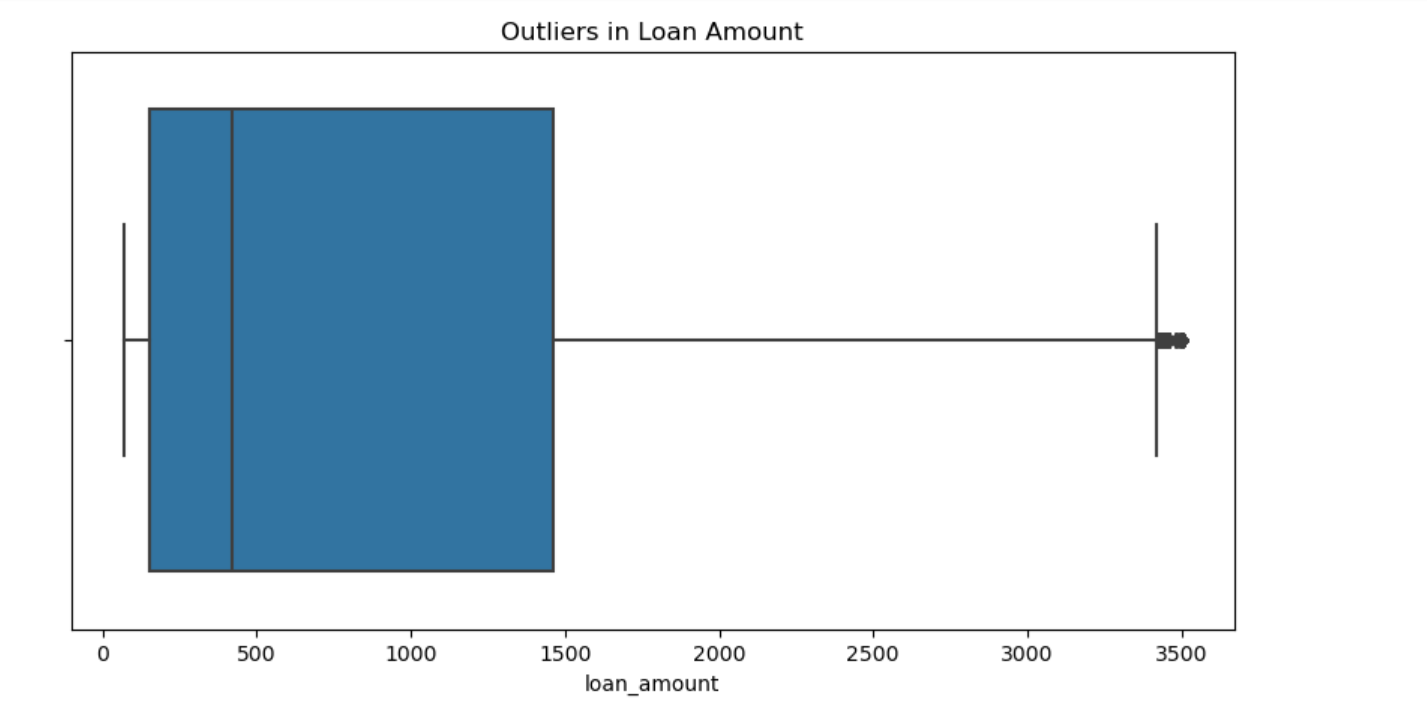


Figure 2Boxplot to show outliers in loan amount

This box plot above shows the distribution of loan amounts with a focus on outliers:

Main Distribution (Box)

The box represents the middle 50% of loan amounts (interquartile range or IQR)

Left edge (Q1): Around $250

Middle line (median): Around $400-500

Right edge (Q3): Around $1,500

Outliers

There is a clear outlier at approximately $3,400-3,500 (represented by the dot at the far right)

This outlier is well beyond the upper whisker, indicating it's significantly higher than most loan amounts

The distance from the box to this outlier suggests it's more than 1.5 × IQR away from Q3

Whiskers

The left whisker extends down to about $50-100

The right whisker extends to around $3,000-3,250

The right whisker's length indicates some spread in the higher loan amounts, but still within expected range

Distribution Shape

The data appears positively skewed (right-skewed)

Most loans are concentrated in the lower range ($250-$1,500)

There's considerable spread in the main distribution

The median being closer to Q1 than Q3 further confirms the right skew

This outlier at $3,400-3,500 represents an unusually high loan amount compared to the typical pattern.

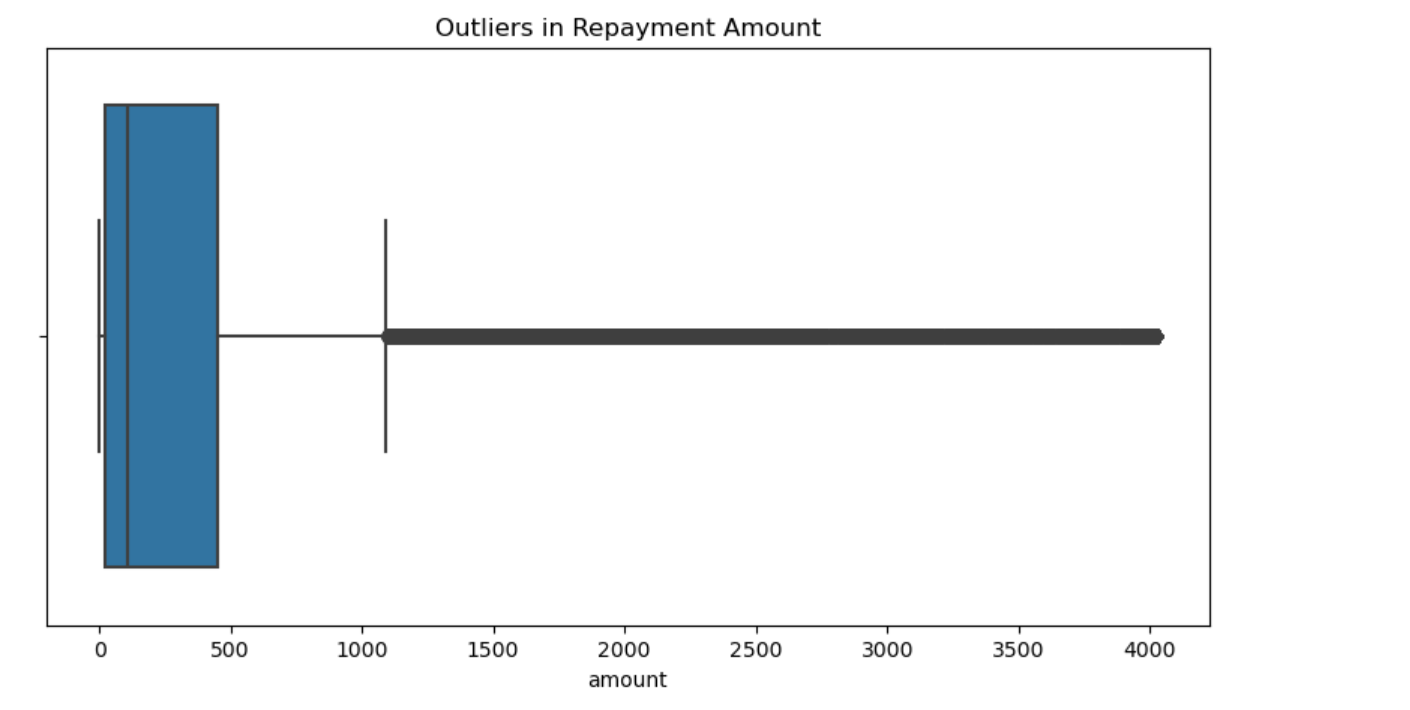


Figure 3Boxplot to show outliers in repayment amount

This box plot reveals an extreme distribution pattern in repayment amounts:

Main Distribution (Box)

The box (representing the middle 50% of data) is very compressed and located between approximately $100-500

The vertical line inside the box (median) is positioned around $200-250

This indicates most repayment amounts are clustered in the lower range

Outliers

There appears to be a significant outlier at approximately $4,000 (shown by the point at the far right)

This outlier is dramatically higher than the typical repayment amounts

The distance suggests it's many times the IQR beyond Q3

Whiskers

The left whisker extends to near $0

The right whisker extends to around $1,200

The right whisker's length indicates some expected variation in the higher repayment amounts

Distribution Shape

The data shows extreme positive skew (right-skewed)

The vast majority of repayments are concentrated in a narrow low-value range

The median being positioned toward the left side of the box further confirms this skew

The $4,000 outlier is approximately 8-10 times higher than the typical repayment amount

This distribution indicates that while most repayments are relatively small and similar in size, there is at least one exceptionally large repayment amount that stands far outside the normal pattern.

## SQL

I used SQL using vscode and sqlite to analyze the data further. The figure below shows the table names, table structure of the dataset, the sum of the loan amount and loan fee, the average loan amount that was given, the sum of the total amount that was repaid and the default rate of the loan.

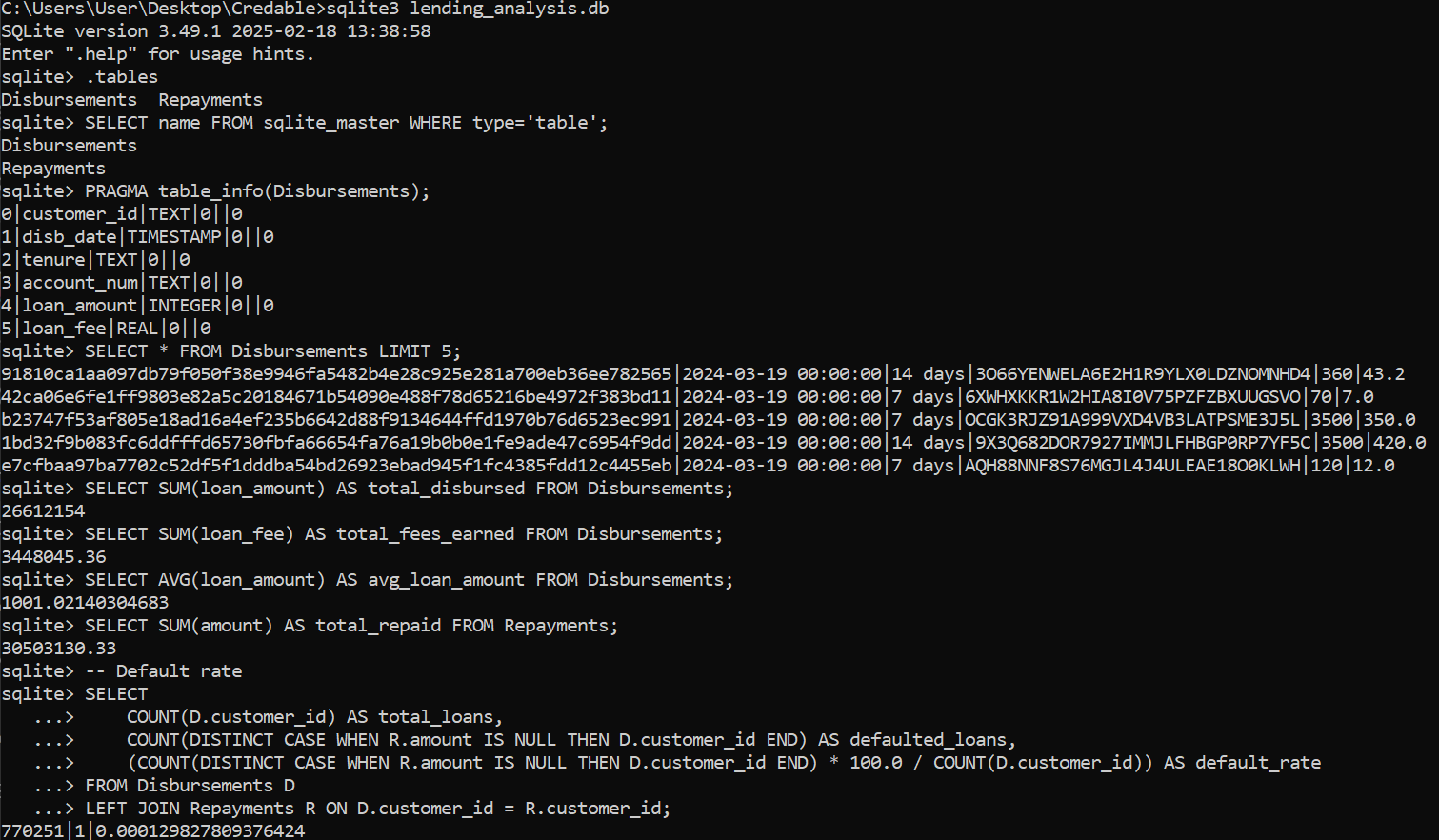


Figure 4Screenshot of SQL queries

The figure below shows: query to find loan disbursement by tenure

The query shows the number of loans issued and the total loan amount disbursed based on loan tenure.

Key Takeaways:

14-day loans are the most popular (highest number of loans issued).

30-day loans have the highest total disbursed amount, meaning larger loan sizes.

7-day loans have moderate usage but a lower total disbursement compared to longer tenures.

Query to find the repayment performance by repayment type

This query breaks down total repayments and the number of transactions by payment method.

Key Takeaways:

More money was repaid manually ($17.56M) than automatically ($12.94M).

However, there are more automatic transactions (41,827 vs. 24,189), meaning customers who pay manually tend to repay in larger amounts per transaction.

This suggests that some customers might only make partial payments manually, while automatic payments ensure smaller, more frequent repayments.

Query to find monthly repayment trend, this data shows a clear monthly repayment trend from January 2024 (202401) through August 2024 (202408).

Repayment Volume Trend:

Steady increase in the number of repayments from January (7,787) to May (9,364) - representing a 20% growth

After May, a consistent decline begins, dropping to 6,454 repayments by August - a 31% decrease from the peak

Repayment Amount Trend:

Total repaid amount follows a similar pattern, peaking in May at $4,466,721.70

The amount drops significantly by August to $2,731,108.25 - almost 39% lower than the May peak

Average Repayment Size:

January: ~$440 per repayment

July: ~$546 per repayment (highest average)

August: ~$423 per repayment (lowest average)

Seasonal Pattern:

Strong performance in Spring (March-May)

Summer decline (June-August)

This could indicate seasonal repayment behavior

Overall:

The significant drop in both repayment count and amount in July-August might indicate a concerning trend that warrants investigation

The steeper drop in total amount (39%) compared to repayment count (31%) suggests smaller average repayments in the later months

This data may reflect seasonal patterns, economic conditions affecting borrowers, or possibly changes in the lending or collection practices during this period.

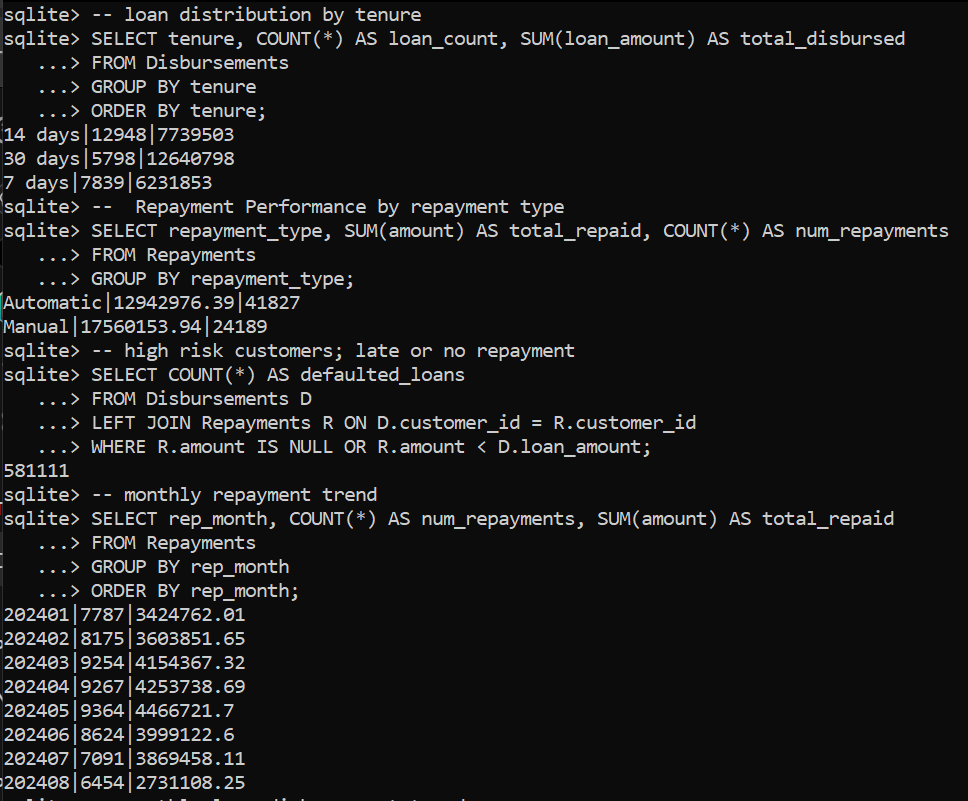


Figure 5Screenshot of SQL queries

The figure below shows: query to find correlation analysis between loan size and repayment behavior, there are clear patterns that emerge:

Customer Distribution:

Small Loans: 1,889 customers (49% of total)

Medium Loans: 1,213 customers (31% of total)

Large Loans: 773 customers (20% of total)

Default Rates:

Small Loans: $25.4M defaulted out of $79.1M total (32% default rate)

Medium Loans: $121.5M defaulted out of $221.2M total (55% default rate)

Large Loans: $342.6M defaulted out of $544.8M total (63% default rate)

Risk Pattern:

A strong correlation exists between loan size and default risk

As loan size increases, the default rate increases substantially

Small loans have the best repayment performance by far

Portfolio Impact:

Despite making up only 20% of customers, large loans account for 63% of all defaulted amounts

Small loans (49% of customers) represent just 6% of defaulted amounts

Risk-Adjusted Performance:

Small loans show the best risk-adjusted return profile

While large loans likely generate more revenue per transaction, they carry significantly higher default risk

This data strongly suggests that focusing on smaller loan products could improve overall portfolio performance by reducing default exposure

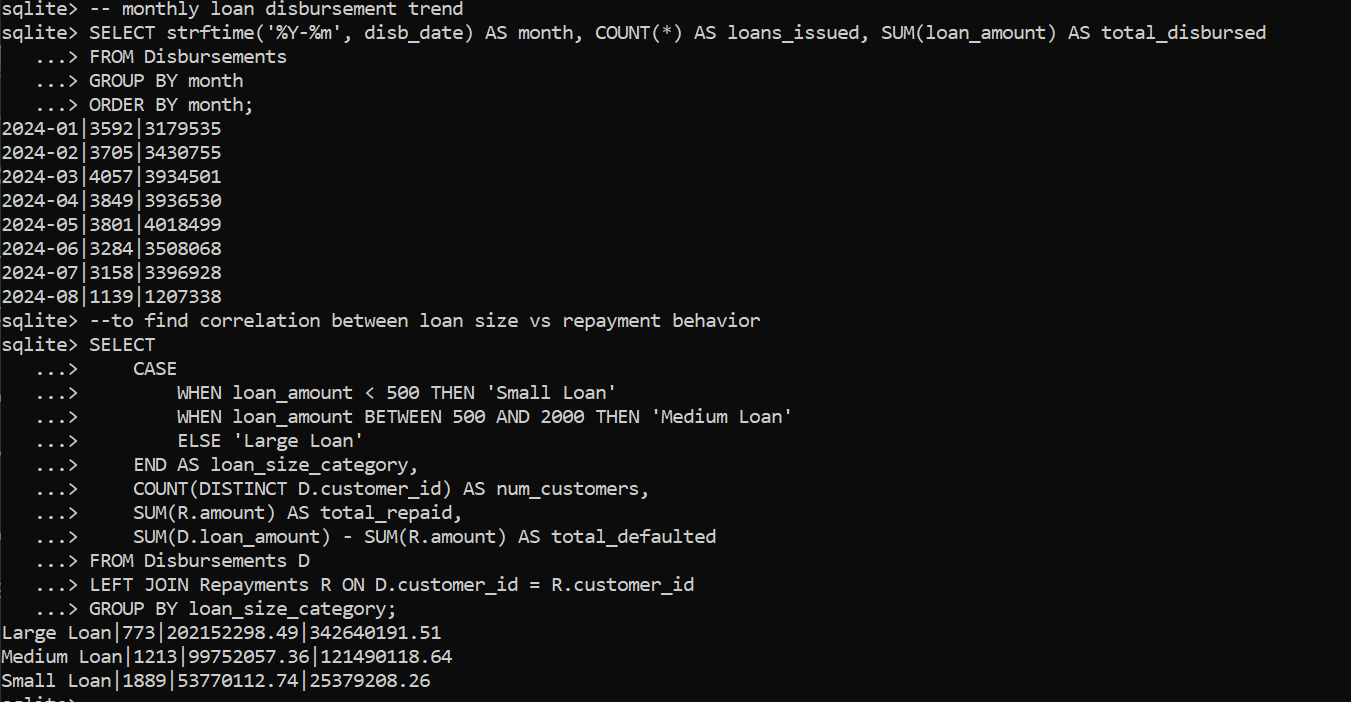


Figure 6Screenshot of SQL queries

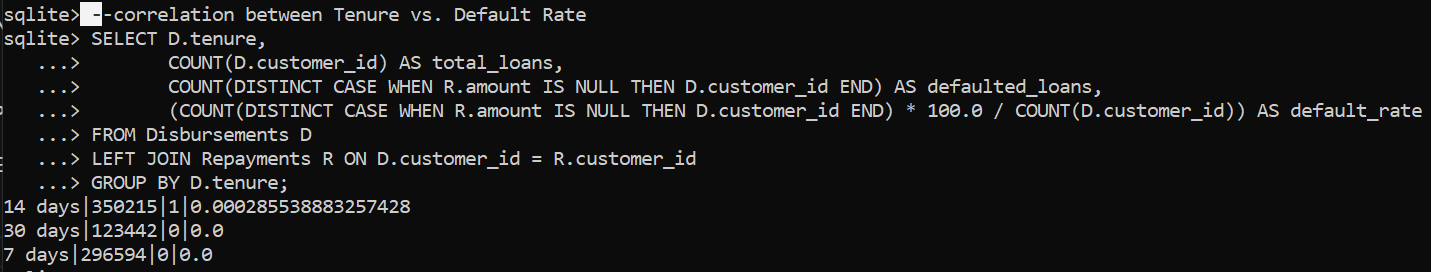


Figure 7Screenshot of SQL queries

The figure above shows correlation between loan tenure and default rate. Key Takeaways:

Almost no defaults were recorded across all tenure types.

Only 1 loan defaulted out of 350,215 for the 14-day tenure, making the default rate nearly 0.00029%.

30-day and 7-day loans had zero defaults, meaning customers repaid these loans fully.

This suggests very high repayment rates, or there might be an issue with how defaults are recorded in the data.

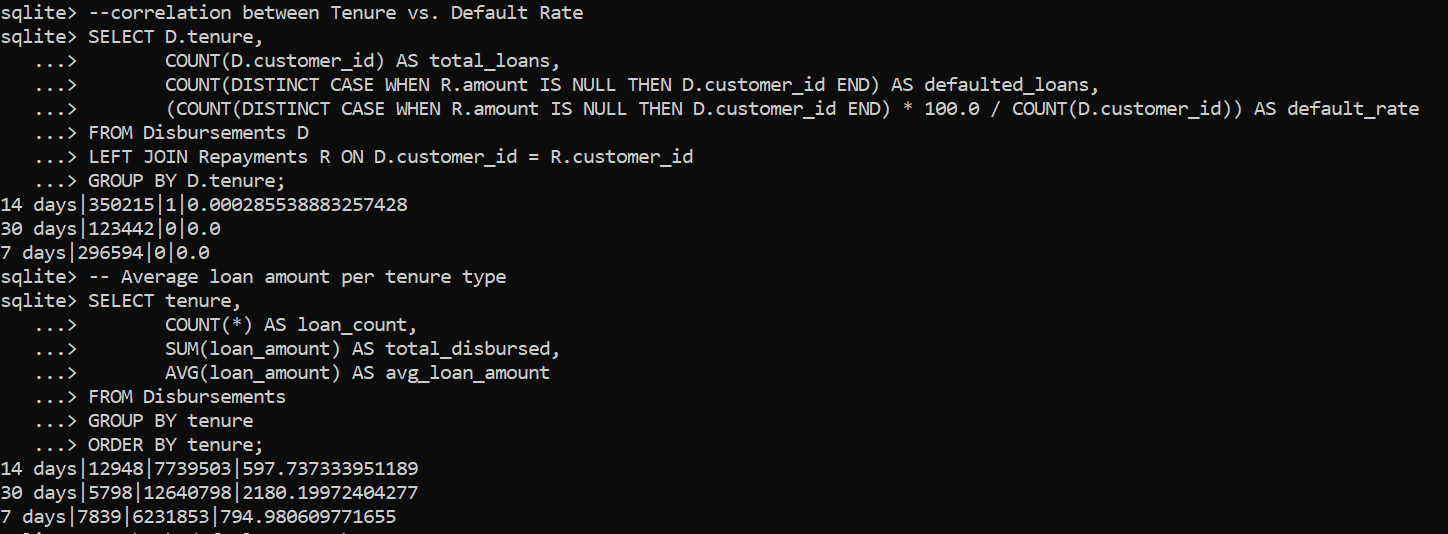


Figure 8Screenshot of SQL queries

The figure above shows average loan amount per tenure type. 30-day loans have the highest average loan size ($2,180.20), meaning they are likely given to higher-value borrowers.

7-day loans are slightly larger ($794.98) than 14-day loans ($597.74), possibly because they are short-term emergency loans.

14-day loans are the most common, but they have the lowest average loan size.

# Visualization

I used power bi to visualize the data as shown in the screen shots below.

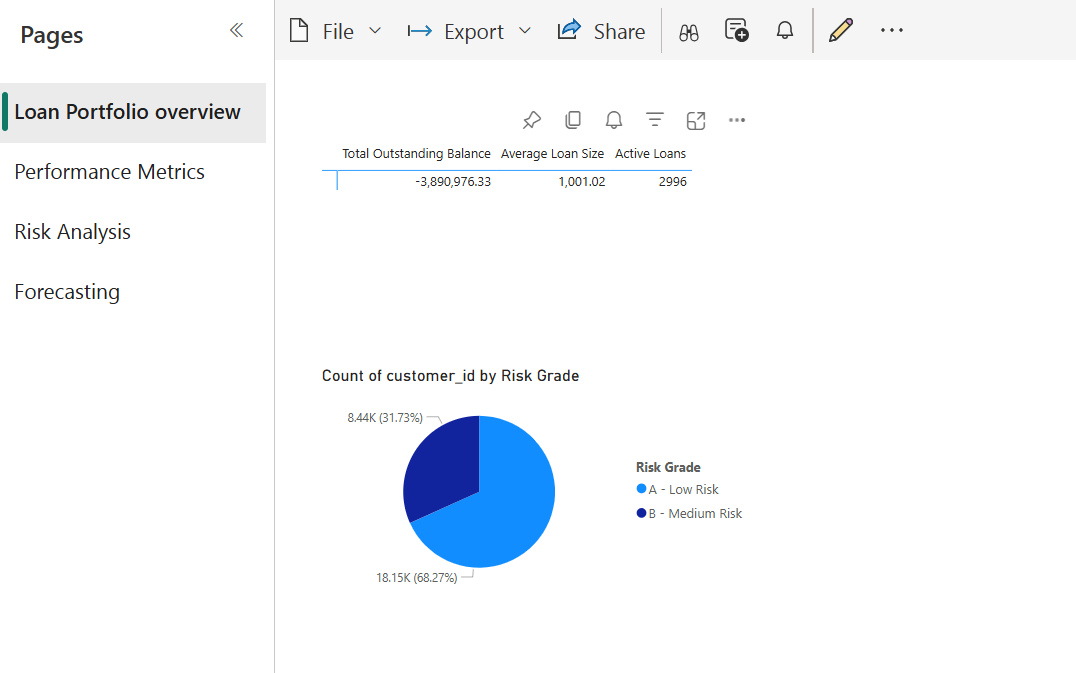


Figure 9Screenshot of loan portfolio overview in Power BI

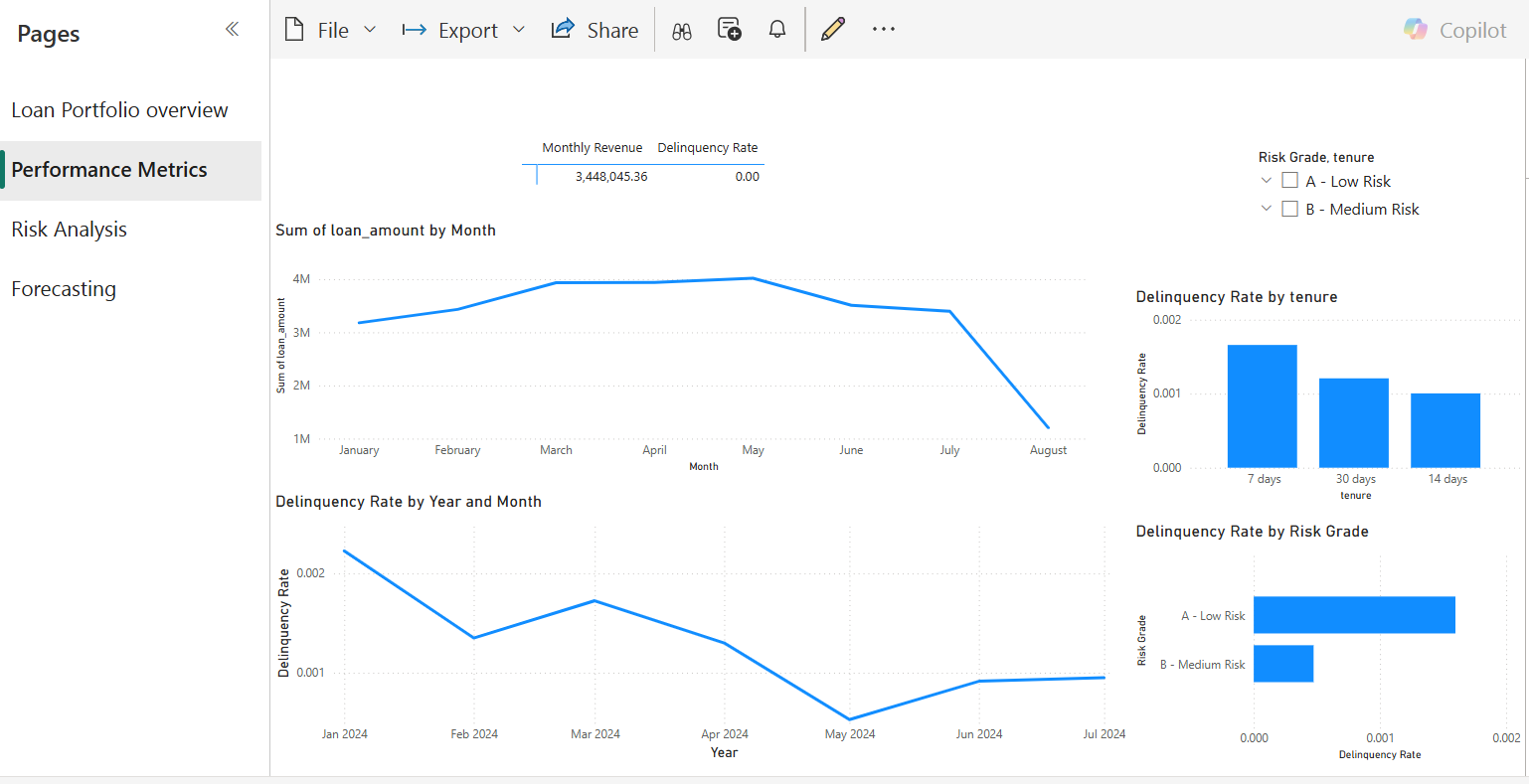


Figure 10Screenshot of performance metrics in Power BI

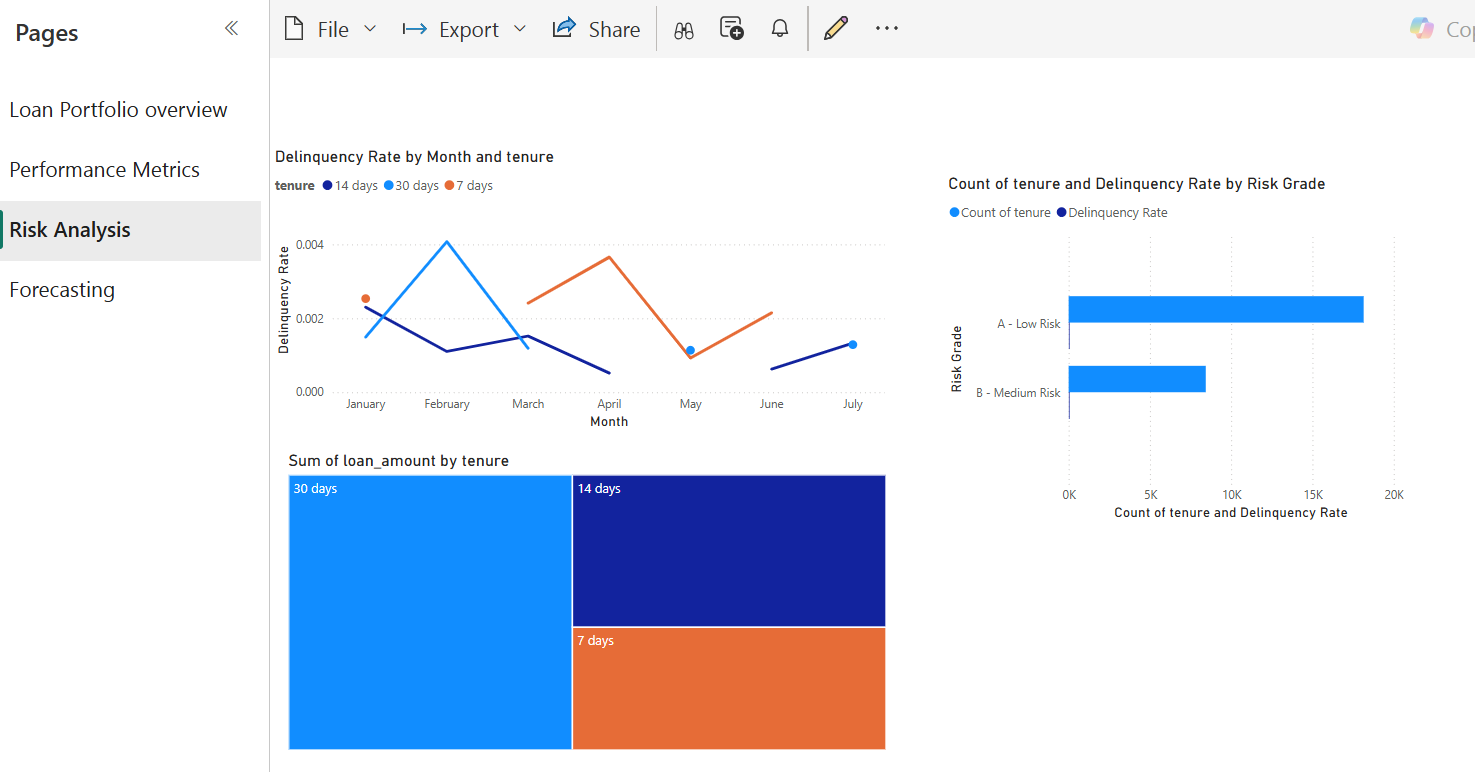


Figure 11Screenshot of risk analysis in Power BI

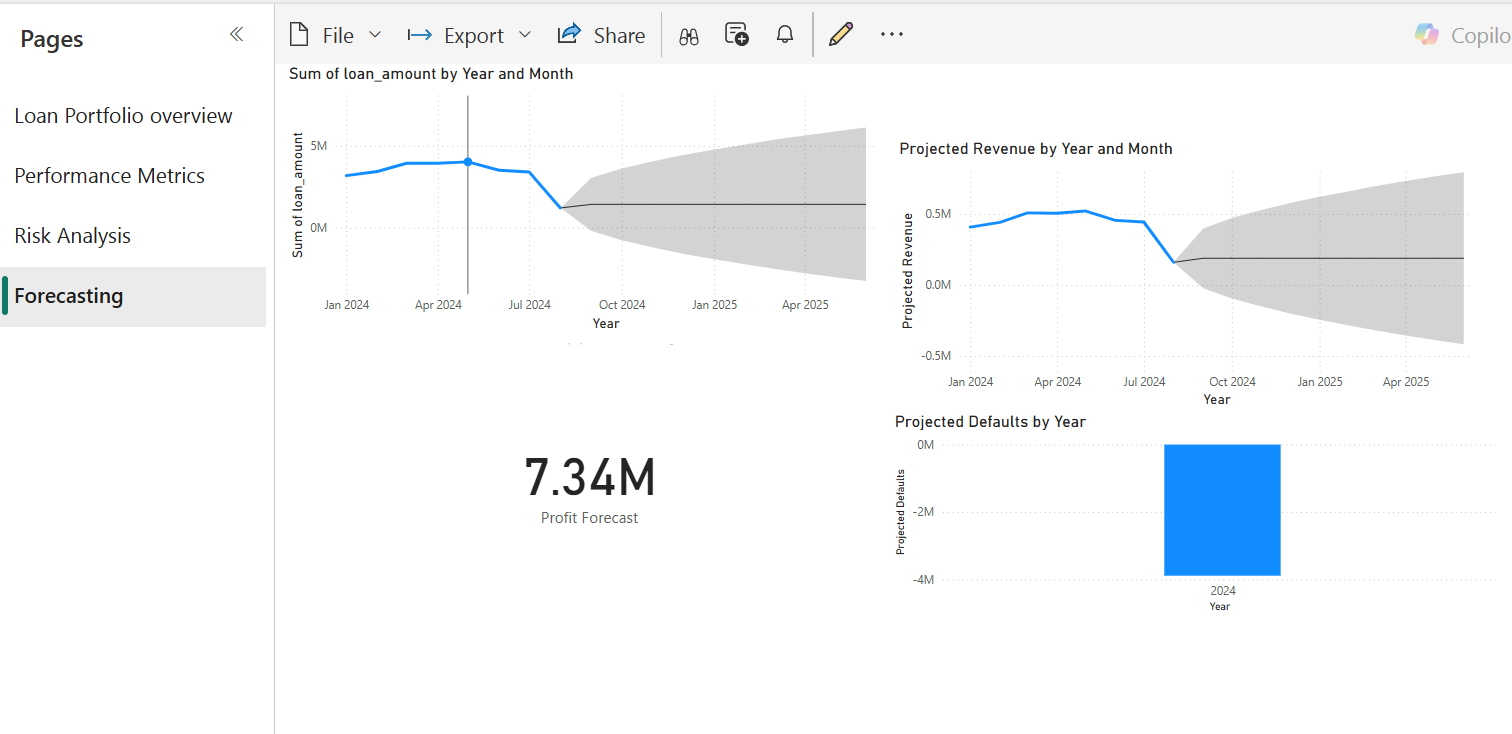


Figure 12Screenshot of forecasting in Power BI

# Conclusion

From the analysis of loan disbursements, repayments, and default rates, we observe:

High Loan Repayment Rates:

The default rate is almost 0% across all loan tenures, with only one recorded default out of 770,251 loans.

This is unusual for a lending business, indicating either highly effective borrower screening or potential issues in how defaults are recorded.

Loan Amount Trends:

30-day loans have the highest average loan size ($2,180.20), suggesting that longer-term loans are given to higher-value customers.

14-day loans are the most common, but they have the lowest average loan amount ($597.74).

Repayment Trends:

Manual repayments contribute more to the total repaid amount ($17.56M) than automatic repayments ($12.94M).

However, automatic payments occur more frequently (41,827 transactions vs. 24,189 manual transactions), suggesting many borrowers rely on scheduled deductions for repayments.

High Credit Exposure:

While delinquency appears low, the business still has 581,111 loans where borrowers have not repaid in full.

This indicates potential financial risk, even if outright defaults are not recorded.

# Recommendations

1. Investigate the Accuracy of Default Tracking

The extremely low default rate (0%) is highly unusual in lending businesses.

Verify if loans that are overdue but not marked as defaulted exist.

Cross-check repayment records to ensure partial or missed payments are correctly classified as defaults.

2. Optimize Loan Approval & Risk Assessment

Since default rates are almost non-existent, the company may be too conservative in approving loans.

Consider adjusting risk models to allow higher approval rates for low-risk borrowers while maintaining a balance in risk exposure.

3. Improve Manual Repayment Efficiency

Manual repayments contribute more revenue, but they are less frequent than automatic payments.

Encourage customers to enroll in automatic payments to reduce collection efforts and increase cash flow stability.

4. Monitor High-Risk Loans More Closely

Even though defaults are low, 581,111 loans are at risk due to incomplete repayment.

Set up automated alerts for loans where repayments are delayed or incomplete.

Implement reminders and follow-up actions for customers who are falling behind on payments.

5. Expand 30-Day Loan Offerings

30-day loans have the highest average loan amount ($2,180.20), meaning they attract higher-value borrowers.

Consider increasing marketing efforts for these loans, as they could generate more revenue per loan issued.

# Links

Below is the power bi link to the visualization:

<https://app.powerbi.com/view?r=eyJrIjoiMGNiZTdmZTUtYTQwZS00YjYzLWIyN2ItMzA3OThlOGQxMDU5IiwidCI6IjE2ZDgzZWU2LTI1NGEtNDY5ZC1hNmNjLTU0ZTJjYTIzMTNlNyIsImMiOjh9>

Below is the github link to the repository:

<https://github.com/vjarenga/loan-product-analysis.git>