Credit Risk Analysis & Portfolio Insights

Author: Arif Shaikhan

Date: 30/1/2026

Dataset: Kaggle - Bank Loan Status Dataset

1. Problem Statement

Our client seeks to identify high-risk borrowers to reduce defaults while maintaining portfolio size. Using loan-level data, we engineered a transparent risk score and analyzed portfolio exposure to quantify risk concentration and guide credit decision-making.

This project is a Credit Risk Analysis using SQL, where I analyzed loan and borrower data to identify factors driving defaults. Starting from a Kaggle dataset of ~8,000 loans (filtered to 5,083 after cleaning), I built an end-to-end workflow including risk scoring, segmentation into Low/Medium/High-risk bands, and portfolio trend analysis. The project highlights key risk drivers like Credit Score, Debt-to-Income ratio (DTI), and Credit Utilization, producing actionable insights and professional deliverables for portfolio management and credit decision-making.

2. Key Insights

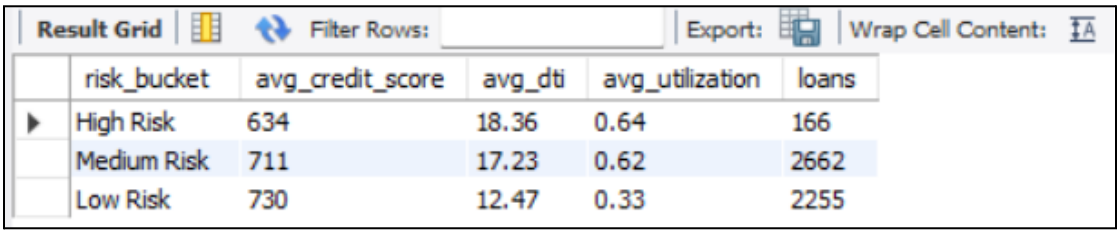
Risk Factor	Observation	Impact on Default
Credit Score	Default rate decreases as credit score increases. Very poor & poor scores contribute disproportionately to defaults.	Strong predictor of credit risk.
Debt-to-Income (DTI)	Borrowers with DTI >35% have the highest default rates.	Primary driver (≈93% contribution).
Credit Utilization	High utilization correlates with higher default probability.	Secondary driver (≈11% contribution).
Employment Tenure	Tenure <2 years = higher default risk.	Moderate predictor
Home Ownership	Renters are riskier than owners.	Minor but relevant driver
Risk Segmentation	The portfolio concentrated in **Medium (52%) & High (44%) risk bands**, contributing most to cumulative defaults. Low-risk loans = 3%.	Indicates portfolio exposure is significant in higher-risk segments

Trend Analysis (Window Functions)	Cumulative defaults accumulate faster in Very High / High-risk bands.	Shows actionable segments for monitoring

Summary: Credit score, DTI, and credit utilization are the **top 3 risk drivers**. Medium & high-risk borrowers account for the majority of defaults, highlighting the portfolio's exposure.

3. SQL Key Outputs

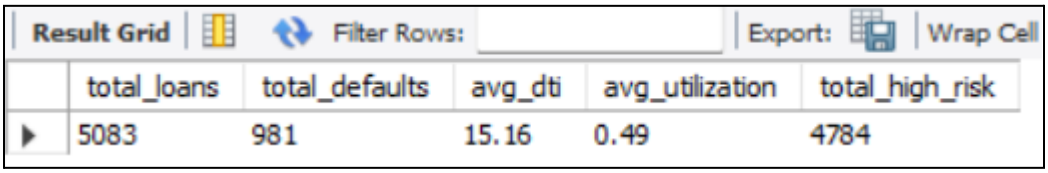
i) Loan-level table with risk metrics (`credit_risk_scoring`)



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

risk_bucket	avg_credit_score	avg_dti	avg_utilization	loans
High Risk	634	18.36	0.64	166
Medium Risk	711	17.23	0.62	2662
Low Risk	730	12.47	0.33	2255

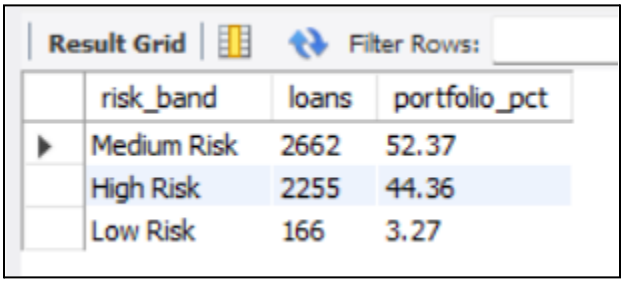
ii) Analytical Base Table (ABT) (`credit_risk_abt`)



Result Grid | Filter Rows: | Export: | Wrap Cell

total_loans	total_defaults	avg_dti	avg_utilization	total_high_risk
5083	981	15.16	0.49	4784

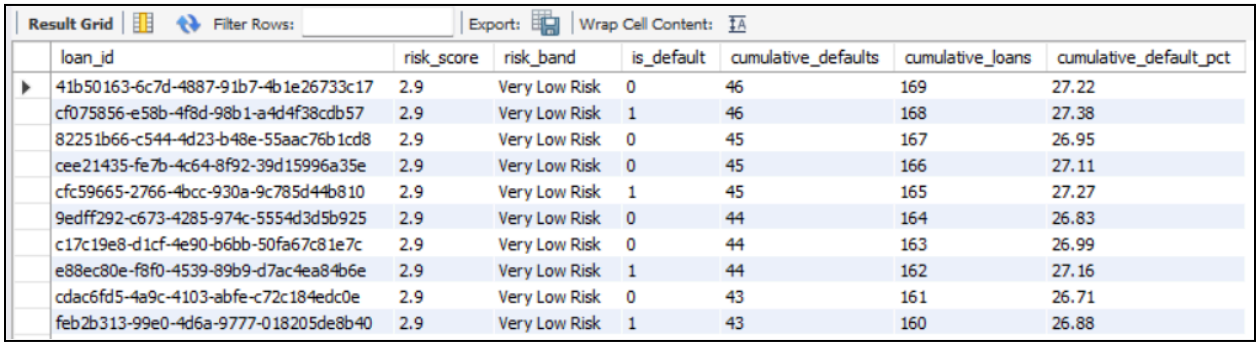
iii) Risk Band Table (`credit_risk_bands`)



Result Grid | Filter Rows:

risk_band	loans	portfolio_pct
Medium Risk	2662	52.37
High Risk	2255	44.36
Low Risk	166	3.27

iv) Cumulative Default Table (`credit_risk_ranked`)



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

loan_id	risk_score	risk_band	is_default	cumulative_defaults	cumulative_loans	cumulative_default_pct
41b50163-6c7d-4887-91b7-4b1e26733c17	2.9	Very Low Risk	0	46	169	27.22
cf075856-e58b-4f8d-98b1-a4d4f38cdb57	2.9	Very Low Risk	1	46	168	27.38
82251b66-c544-4d23-b48e-55aac76b1cd8	2.9	Very Low Risk	0	45	167	26.95
cee21435-fe7b-4c64-8f92-39d15996a35e	2.9	Very Low Risk	0	45	166	27.11
cfc59665-2766-4bcc-930a-9c785d44b810	2.9	Very Low Risk	1	45	165	27.27
9edff292-c673-4285-974c-5554d3d5b925	2.9	Very Low Risk	0	44	164	26.83
c17c19e8-d1cf-4e90-b6bb-50fa67c81e7c	2.9	Very Low Risk	0	44	163	26.99
e88ec80e-f8f0-4539-89b9-d7ac4ea84b6e	2.9	Very Low Risk	1	44	162	27.16
cdac6fd5-4a9c-4103-abfe-c72c184edc0e	2.9	Very Low Risk	0	43	161	26.71
feb2b313-99e0-4d6a-9777-018205de8b40	2.9	Very Low Risk	1	43	160	26.88

4. Recommended Actions

- Review or reject top 10–20% highest-risk loans.
- Implement enhanced monitoring for Medium Risk loans.
- Maintain standard workflow for Low Risk loans.
- Use DTI and utilization thresholds as early warning triggers.
- Future enhancement: incorporate macroeconomic or behavioral data for predictive modeling.

5. Business Impact

Metric	Current	Expected Improvement
Portfolio Default Rate	~19%	Reduce to <16% (by targeting high-risk borrowers)
High-Risk Exposure	44% of portfolio	Focused monitoring could mitigate ~50–60% of potential losses.
Decision-Making Speed	Manual analysis	SQL-based dashboards enable **real-time actionable insights**

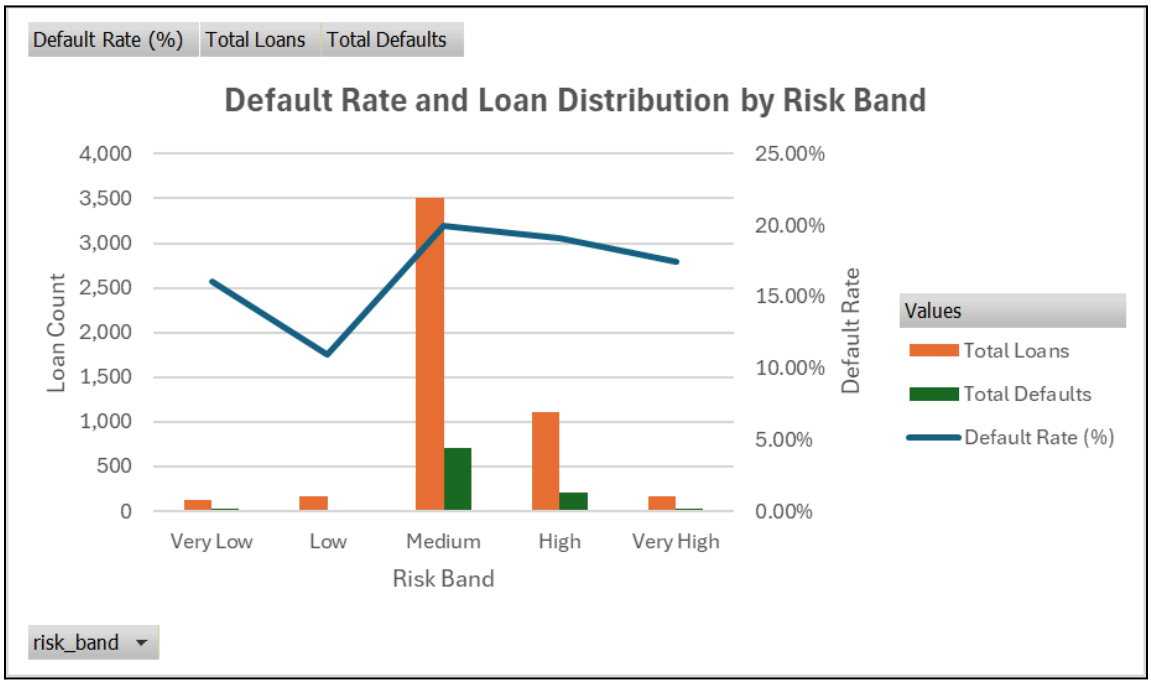
Implementing these measures will **improve risk-adjusted portfolio performance** and strengthen the bank's credit decision process

6. Implementation Plan

- Integrate risk_score and risk bands into loan approval workflow.
- Generate monthly risk band reports with cumulative default metrics.
- Review top drivers periodically; recalibrate score as needed.
- Optionally, extend score with macroeconomic trends or behavioral indicators.

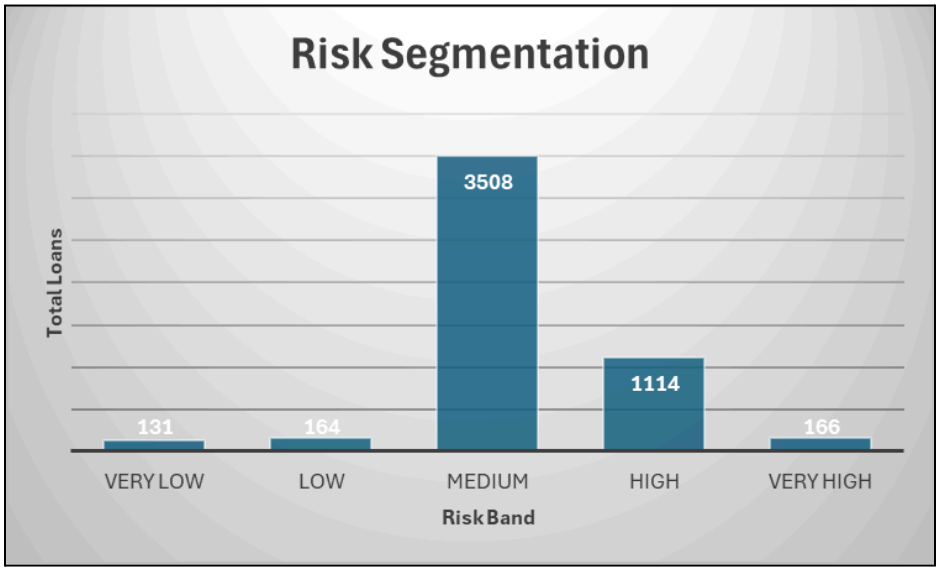
7. Visuals

i) Default Rate and Loan Distribution by Risk Band



The chart shows Default Rate (%), Total Loans, and Total Defaults across risk bands from Very Low to Very High. Most loans are concentrated in the Medium risk band, which also accounts for a large portion of total defaults. Default rates generally increase with risk, confirming that the risk scoring system aligns with actual borrower behavior. The slight decline in the Very High band is due to a smaller number of loans, which introduces some volatility at the extreme tail.

ii) Risk Segmentation



The portfolio is dominated by **Medium-risk borrowers** (3,508 loans), while **High** and **Very High-risk** loans total 1,280, representing significant default exposure. **Low** and **Very Low-risk** loans are rare (295 loans), showing that most of the portfolio carries moderate to high risk. Focused monitoring of these segments is essential for managing defaults.

iii) Top 5 Riskiest Borrowers

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

loan_id	customer_id	risk_score	dti	credit_utilization
f9ad44d2-ec3d-4998-813d-2c11f2c69334	9abdb8b9-9ccf-4a27-89d8-240628ebc3f2	4.3	29.4	0.77
2945b719-a387-494e-9c49-b4e2b500e135	d4f59a9f-98a7-476b-bc80-32cb7a7a94ee	4.3	7.01	0.78
c52938aa-5c99-4e2f-9745-1c6b768cab17	d65be7af-e9fa-41e0-b653-0c81c69d85f8	4.3	16.2	0.79
1cf81c05-be9f-4d19-9a35-6c8d11c56eb5	31b9cc39-5856-4fc9-b6d9-b1692bedcf57	4.3	28.6	0.83
05f70be6-5930-4536-8091-538e5dc2a48f	e0b51878-e75e-4283-8d43-8d5be06b69ec	4.3	12.96	0.85

The top 5 riskiest borrowers all have a **risk score of 4.3**, indicating extremely high default potential. Their **Debt-to-Income ratios range from 7% to 29%**, and **credit utilization** is between 0.77–0.85, showing that both high debt relative to income and high credit usage contribute to elevated risk. These borrowers highlight the types of accounts that require **immediate monitoring or intervention**.

8. Conclusion

The analysis demonstrates that Credit Score, DTI, and Credit Utilization are the strongest drivers of default risk. Medium and High-risk borrowers account for the majority of portfolio exposure and cumulative defaults. Implementing targeted monitoring, early warning triggers, and SQL-driven dashboards will reduce default rates, improve risk-adjusted performance, and support more informed, data-driven credit decisions. Future enhancements could include predictive modeling with macroeconomic and behavioral data to further optimize portfolio management.