Section E-Group: 16

Predictive Analytics for Loan Default Risk

A Comprehensive Data Study

Sector :Banking, Financial Services, and Insurance (BFSI)

By: Param Khodiyar | Ananya soni | Ansh Tomar | Himanshu | Kaavya Gala | Gauri mishra

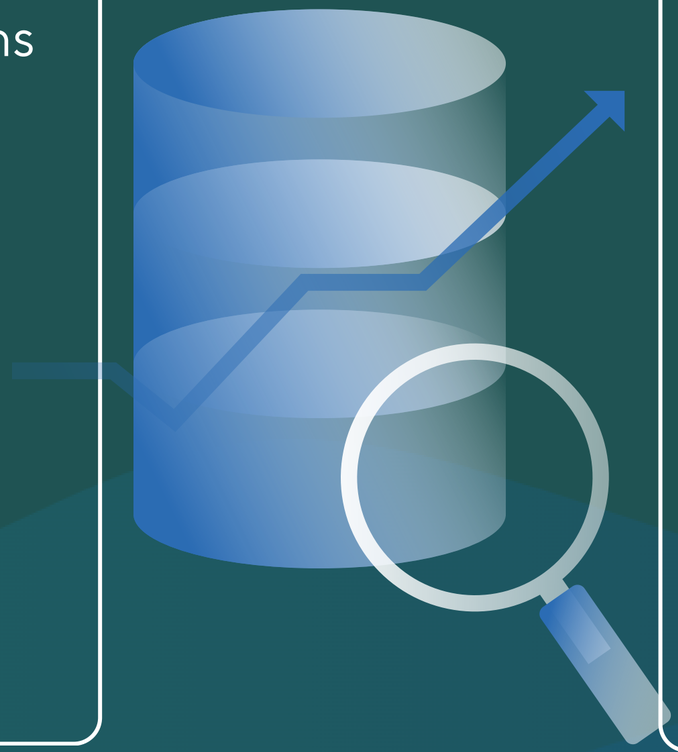
Sector Context:
The High Stakes of Lending

- **Why it matters?**: In the BFSI (Banking, Financial Services, and Insurance) sector, interest income is the primary revenue driver, but Loan Defaults are the primary threat to capital stability.
- **The Challenge**: Aggressive lending grows the portfolio but increases the risk of Non-Performing Assets (NPAs). Conversely, overly cautious lending leads to missed market opportunities.
- **Decision-Maker**: Chief Risk Officer (CRO), Credit Underwriting Managers, and Financial Analysts.

Problem Statement:

Financial institutions face significant losses due to borrower defaults, often driven by complex interactions between debt-to-income ratios , property valuations, and borrower credit history.

How can we identify high-risk applicants early in the application process to minimize credit loss without turning away creditworthy customers?



Objective:

The primary objective of this project is to analyze historical loan data to:

1. Identify Key Risk Drivers: Determine which factors (e.g., Credit Score, LTV, or Loan Purpose) most accurately predict a default.
2. Optimize Underwriting: Support the decision of whether to approve, reject, or adjust interest rates based on a borrower's specific risk profile.
3. Visual Storytelling: Create a dashboard that allows stakeholders to monitor portfolio health across different regions and demographics.

Data Engineering (Source to Sink)

Data Source & Profile

- **Dataset Name:** Loan Default Dataset (Multi-variable Financial Records).
- **Dimensions:**
  Rows: 10,000
  Columns: 35
- **Time Period:** 2019
- **Format:** Structured CSV ready format.

Data Cleaning & Preprocessing

- We replaced technical codes with full names.
- Missing interest rates and Income used the median value.
- Removed extra spaces from text cells.

Data Dictionary

- Target Variable – Status
- Loan Details – Loan Amount, Rate of Interest
- Demographics – Age, Region, Income

KPI & Metrics Framework

KPIs			
	Total Applications 9905	Average Credit Score 701	Average Interest Rate 4.03%
	Default Rate 24.87%	High-Risk Borrower Ratio 1.1	Interest-Only Loan Share 4.39%

Average Rate of Interest: 4.03%

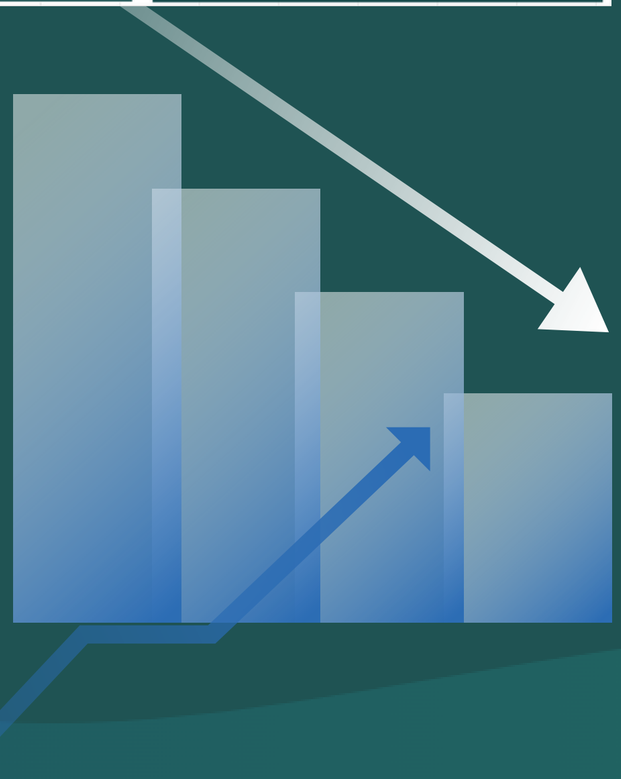
Total Applications: 9,767

Total Loan Amount: $333.9 Million

Default Rate: 24.87%

Average Credit Score: 701

- **Default Rate & Credit Score:** These are our "Early Warning" signals. A 24.87% default rate is high; we monitor these to identify if we are lending to the wrong risk profiles.
- **Total Loan Amount & Applications**: These measure our market reach and scale. They show if the business is expanding or shrinking.
- **Rate of Interest**: This is our primary revenue driver. We analyze this against the Default Rate to see if higher interest rates are leading to more defaults.



EDA

- Regional Risk Concentration: The North-East region shows a staggering 40% default rate for low-score borrowers. This geographical hotspot requires immediate credit policy review.
- The Prime Borrower Trap: High credit score borrowers (750+) accounts for 928 total defaults. High scores alone are not preventing significant losses.
- Collateral Risk Extremes: Loans with LTV ratios over 100 have a 71% default rate. Lending more than the property's value is extremely risky.
- Loan Type Performance: "Type 2" loans show the highest default rate at 33%. This specific loan structure is underperforming compared to others.
- Income Stability Observation: Average income remains flat across all credit bands. Borrower earnings are not increasing alongside their credit scores.
- Co-applicant Lending Power: "EXP" co-applicants secure much higher average loan amounts. Joint applications are driving the largest capital deployments.

Advanced Analysis

Risk Segmentation & Root Cause Analysis

We performed a deep-dive segmentation to find the "Danger Zones" within our data:

Defaults spike sharply when LTV exceeds 100%.

North-East region shows elevated default risk.

Type 2 loans exhibit the highest failure rate.

What New Understanding Did This Provide?

- Credit Scores are Not Enough: High scores (750+) do not protect us if the LTV ratio is too high.
- Geography Matters: Repayment behavior is heavily influenced by the borrower's Region, not just their personal finances.
- Structural Risk: Our current "Type 2" loan structure is fundamentally riskier than other products we offer.

Loan Analysis Dashboard

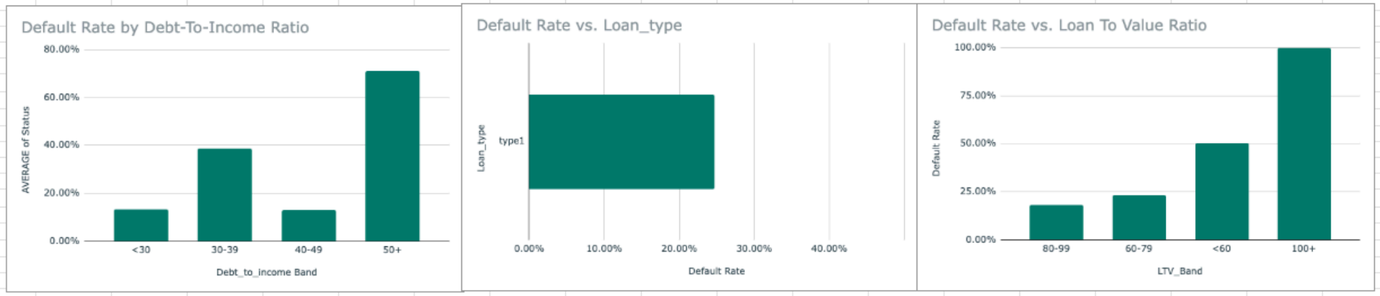
KPIs

Total Applications	Average Credit Score	Average Interest Rate
1780	700	4.16

Default Rate	High-Risk Borrower Ratio	Interest-Only Loan Share
24.87%	1.1	4.20%

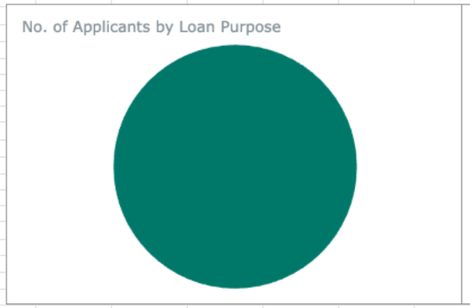
Risk Drivers

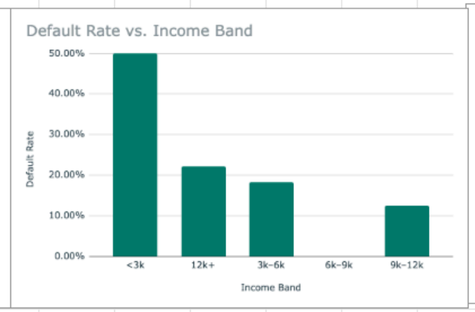
Filters: Loan Type Filer 1 of 3 | Credit Score All

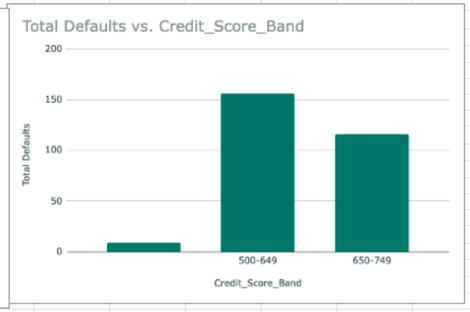


Borrower Profile

Filters: Loan Purpose 2 of 5 | Co-applicant Credit Type All

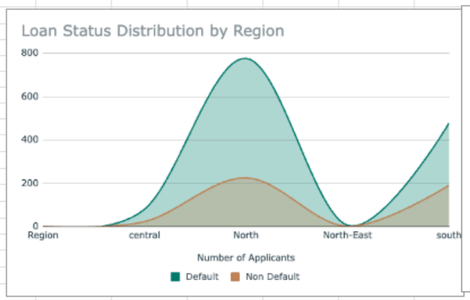


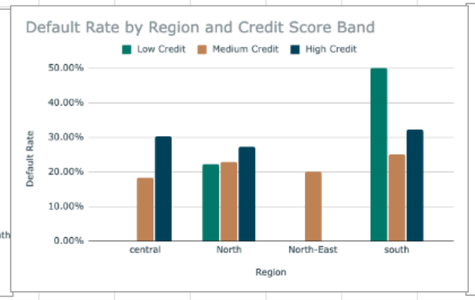


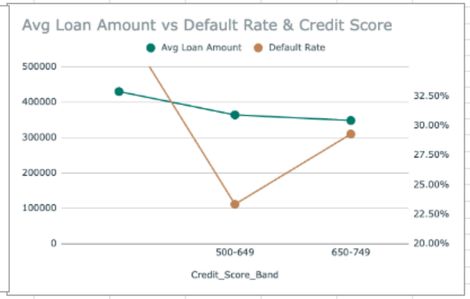


Region Wise Analysis

Filters: Region All | Credit Score All







- Shows key KPIs including Default Rate, Credit Score, Interest Rate, and Risk Ratios.
- Visualizes defaults across LTV, DTI, Loan Type, and Region.
- Interactive filters enable dynamic portfolio analysis.
- Highlights risk patterns and borrower segmentation.

Recommendations

- **Tighten LTV Thresholds:** Stop approving loans with an LTV over 100%, as these currently result in a 71% default rate.
- **Regional Risk Mitigation:** Implement stricter credit requirements for the North-East region, where low-score defaults hit 40%.
- **Re-evaluate "Type 2" Products:** Audit the Type 2 loan structure to identify why its 33% default rate is significantly higher than other types.
- **Enhanced DTI Checks:** For high credit score borrowers (750+), prioritize Debt-to-Income (dtir1) analysis over scores to prevent "Prime" defaults.
- **Co-applicant Limits:** Place a cap on total loan amounts for EXP co-applicant types to reduce high-capital exposure in risky segments.

Impact & Value

IMPACT

"So What?"

- Reduced Credit Loss: By eliminating loans with LTV>100, we can potentially prevent a significant portion of our 24.87% default rate.
- Capital Efficiency: Redirecting funds from high-risk Type 2 loans to safer products ensures better returns on our $333.9M portfolio.
- Targeted Risk Management: Focusing on the North-East region reduces operational costs by identifying high-risk areas before defaults occur.

VALUE

Why should the stakeholder approve this?

- Protect Assets: Implementing these changes protects the principal amount of future loans from "high-score" defaults.
- Data-Driven Growth: Moves the bank from "guessing" to using Pivot Table insights for more accurate lending decisions.
- Profitability: Lowering the default rate directly increases the net profit margin from the 4.03% average interest rate.

Limitations

- **Time-Specific Insights:** The analysis reflects the 2019 economic climate; future iterations could incorporate post-pandemic shifts to test model resilience.
- **Scope of Variables:** While our financial metrics are robust, adding "Employment Tenure" would further sharpen the prediction of borrower stability beyond debt ratios like dtir1.
- **Methodological Precision:** We utilized Median Imputation for rate_of_interest to maintain statistical integrity; future work could use advanced regression to predict these missing values even more accurately.

Next Steps

- Predictive Modeling: Moving from spreadsheet analysis to Machine Learning (like Logistic Regression) to automate default probability scores.
- Dynamic Dashboards: Integrating live data feeds to track how the Default Rate fluctuates in real-time as market interest rates change.
- Granular Segmentation: Collecting more data on Property Location to understand why the North-East region is such a high-risk outlier.