

1. Introduction

The objective of this project is to analyze risk patterns in a synthetic HDFC retail loan dataset and build an interactive Tableau dashboard that replicates real-world credit risk analytics workflows. The project focuses on identifying default trends, understanding high-risk customer segments, evaluating branch and regional performance, and visualizing customer behavior patterns using industry-standard BI techniques.

2. Dataset Description

A synthetic dataset of 1,200 loan customers was generated and preprocessed using Python.

The dataset includes:

- Customer ID
- Region and Branch
- Income
- Loan Amount
- Credit Score
- Utilization Ratio
- High Utilization Flag (0/1)
- Credit Bucket (score ranges)
- Default Flag (0/1)
- Loan Term (months)

Feature Engineering

- **High Utilization Flag:** Utilization > 65%
- **Credit Score Buckets:**
 - 300–579
 - 580–639
 - 640–699
 - 700–749
 - 750+
- **Income-to-Loan Ratio**

These engineered fields enable segmentation and risk profiling.

3. Tools Used

- **Python (Pandas, NumPy):** Data generation, cleaning, feature engineering

- **Tableau Public:** Dashboard creation and visualization
 - **Excel:** Exploratory checks
 - **ReportLab:** Report export
-

4. Visualizations Created in Tableau

A. Default Rate by Region

A bar chart showing the average default percentage across regions.

Insight: Delhi exhibits comparatively higher default rates, followed by Mumbai. This allows risk teams to identify geographic hotspots.

B. Branch Performance

A branch-level default rate comparison.

Insight: HDFC-M2 branch shows the highest default concentration, highlighting areas requiring operational attention.

C. Credit Score × Utilization Segmentation

This is the core risk segmentation visualization.

It shows how default probability varies jointly by:

- Credit Score Bucket
- High Utilization vs Low Utilization

Insights:

- Customers with **low credit score (580–639)** show significantly higher default tendency.
 - **High utilization borrowers default more in every credit bucket**, confirming utilization as a strong risk driver.
-

D. Income vs Loan Amount Scatterplot (Colored by Default)

A scatterplot showing customer behavior based on income and loan size.

Insights:

- Defaulters cluster between **₹40k–₹80k income** and **₹20k–₹60k loan amounts**.
 - High-income customers rarely default, reflecting healthier financial stability.
-

E. Credit Score Distribution with Default Split

A histogram showing distribution of credit scores, segmented by default status.

Insights:

- Most defaults occur in the **580–640** credit score range.
 - Customers with scores above **700** rarely default.
This aligns with real-world credit scoring risk patterns.
-

5. Dashboard Structure

The final Tableau dashboard includes:

Top Row

- Default Rate by Region
- Branch Performance

Middle Row

- Credit Score × Utilization Segmentation (Key Risk Insight Visual)

Bottom Row

- Income vs Loan Scatterplot
- Credit Score Distribution Histogram

Filters Added

- Region
- Branch
- Credit Bucket
- High Utilization
- Default

All filters use **Apply to Worksheets → All Using This Data Source** to ensure full interactivity.

6. Key Findings & Insights

1. **Credit Score is the strongest predictor of default** — especially scores below 640.
2. **Utilization ratio significantly impacts risk** — high-utilization customers default disproportionately more.
3. **Delhi and HDFC-M2 branch show higher risk exposure**, suggesting operational or demographic factors.
4. **Mid-income borrowers (₹40k–₹80k)** show a noticeable default cluster.
5. **Loan amounts between ₹20k–₹60k** appear most susceptible to default.
6. Customers with **credit score >700** show almost **zero default**, consistent with banking risk models.

7. Business Recommendations

- Tighten credit policies for **credit score < 640**.
 - Monitor and control **high-utilization accounts** proactively.
 - Perform deeper review on **Delhi region and high-risk branches**.
 - Introduce **risk-based pricing** for mid-score customers.
 - Strengthen communication and EMIs structuring for at-risk customers.
 - Use utilization alerts for early intervention.
-

8. Conclusion

This project demonstrates a complete risk analytics workflow:

Data Preparation → Feature Engineering → Visualization → Dashboarding → Insights → Recommendations

It effectively mirrors how real financial institutions analyze loan portfolios, uncover risk patterns, and make data-driven portfolio decisions. The final Tableau dashboard provides a powerful interactive tool to explore customer risk profiles and default behavior.