Project Report: Credit Risk & Customer Retention Analysis – Predictive Modeling and Insights

Date: 23rd August, 2025
To: Stakeholders
From: Shijin Ramesh, Data Analyst

Objective:

The primary objective of this project was to build a predictive and analytical solution to assist banks and financial firms in predicting loan/credit card default risk, identifying customer churn risk for proactive retention, and providing actionable business insights through an interactive dashboard showing risk, churn, and product penetration.

Tools & Libraries Used:

The project utilized a suite of Python-based tools and libraries for data analysis and predictive modeling, including:

- Python, Pandas, Numpy for data manipulation and numerical operations.
- Matplotlib, Seaborn for data visualization.
- Sklearn for machine learning algorithms.
- Jupyter Notebook for interactive development.
- Snowflake for data management (mentioned in EDA tools).

1. Exploratory Data Analysis (EDA) Summary

The EDA phase involved studying customer demographics, credit profiles, product holdings, transactions, and engagement patterns to understand factors linked to default and churn.

- **Customer Demographics & Financial Profile:**

◦ **Age**: Most customers are between 45 and 70 years old, actively engaged in loans or investments, likely due to increased income with age. Age was not found to be a major driver of default or churn.

◦ **Credit Score**: The distribution is generally healthy (600–750 average). A small group has very high scores (>800), and a small group has very low scores (<500), marking them as high-risk. While average scores were similar between defaulters and non-defaulters, defaulters showed a wider spread, indicating some extremely poor scorers pulling down the group. Customers with credit scores less than 580 defaulted more, but surprisingly, an 18.5% default rate was observed for scores over 750. Credit score showed no impact on churn.

◦ **Credit Utilization**: Most customers use 10% to 40% of their credit, considered healthy. Only a small number use above 60%, signaling financial stress. Both very low (<13%) and moderate

(13-22%) credit utilization patterns can be early indicators of risk, either churn or default. Higher utilization is mildly positive for delinquency.

◦ **Loan Amounts**: Mostly on the smaller side, with a rapid decrease in borrowers as loan size increases. Defaults are more common in medium-ticket loans, possibly due to better underwriting and monitoring for very large loans.

◦ **Fixed Deposit (FD) Investment**: Over 700 customers have fixed deposits of around ₹1,00,000, with fewer holding larger deposits (₹2.5 lakh to ₹20 lakh). FD holdings did not protect against default risk or predict churn. Many customers with higher FDs were identified as potential high-value customers.

◦ **Income Levels**: Most customers earn between ₹20,000–₹1.5 lakh per month, with a smaller group earning ₹2 lakh+ (potential premium customers for cross-selling). Income and expenses are almost perfectly correlated (0.996), suggesting redundancy in using both for modeling.

◦ **Income Bracket**: The largest segments are 6–10 LPA (24.2%) and 3–6 LPA (24.0%). Higher-income customers (15–25 LPA) showed a higher default rate (~17%) than mid-income groups (~13%). Mid-income groups (3–10 LPA) are managing defaults better but are more likely to churn. Low-income (<3 LPA) and high-income (25+ LPA) customers churn less.

◦ **Employment Status**: Salaried customers (52.3%) are the most stable but have the highest default rate (~15%) and churn rate (~17%). Retired and unemployed customers have lower default (~9–10%) and churn rates (unemployed 6%). Students (11.2%) are a young segment for long-term relationships.

◦ **Gender**: The customer base is balanced (50.5% female, 45.4% male, 4.1% other). Gender was not a strong driver of default or churn.

• **Engagement & Behavior:**

◦ **Digital Engagement**: Net Banking is used 5–15 times/year for high-value transactions. Mobile App is used 15–29 times/year by most, with heavy users logging in >40 times. Digital engagement levels did not predict default. Active customers used net banking slightly more (10 vs. 9.6 times), suggesting digital engagement may help reduce churn. However, churned customers used the app slightly more, implying app usage alone doesn't guarantee retention. Service usage (net banking, mobile app) showed very weak correlation with financial risk.

◦ **Complaints Raised**: More than 600 customers never raised a complaint, but some raised up to 3. Non-defaulters raised slightly more complaints, possibly due to issues being resolved faster. Interestingly, customers with the most complaints had the lowest churn rate (7.8%), suggesting effective resolution increases loyalty. Those with 1-2 complaints had the highest churn. Complaints did not strongly link to churn overall.

◦ **Spending Behavior**: Average monthly expense peaks around ₹20,000–₹22,000. Some spend much more (₹50,000–₹2.5 lakh). Retail, travel, and bills spending are highly correlated with total expenses and with each other, implying they tell a similar story. Spending patterns did not strongly link to default or churn.

◦ **Repayment Behavior**: 350+ customers were delinquent once, 200–250 twice, and some 3+ times. Overdue days ranged from 120 to 160 days for some, a serious red flag. Delinquency and overdue days are strongly related (0.74 correlation). Surprisingly, non-defaulters had slightly higher delinquency counts and overdue days, possibly due to resolution of short-term issues. These behaviors were not predictive of churn.

◦ **Loan History**: 39.6% have 1–2 loans, 26.2% have none (new lending opportunities), and 10.3% have 6+ loans (high-risk heavy borrowers). Customers with 3–5 loans (28%) and 6+ loans (23%) have the highest default risk. Multiple loans are a red flag for default. Customers with 3–5 loans churn the most (~20%), while 1–2 loans churn the least (~13%). Loan History is identified as the strongest predictor for both default and churn.

◦ **City Distribution**: Top cities are Ahmedabad, Pune, Hyderabad, Delhi, Kolkata, with Mumbai having lower representation. Default risk is higher in Bengaluru, Kolkata, and Ahmedabad (~16–17%), and lower in Chennai and Pune. Churn is highest in Mumbai and Hyderabad (~18%), and lowest in Delhi, Pune, Ahmedabad.

◦ **Product Ownership & Offers**:

▪ **Loan Type**: Personal loans dominate (50.6% of loan holders). Loan type did not strongly affect default or churn rates.

▪ **Response to Offers**: 56.3% ignored, 24.2% accepted, 19.5% clicked but didn't convert. Marketing offer response did not predict default risk or strongly influence churn.

▪ **Credit Card Ownership**: 67.4% own a credit card. Credit card holders (14.2%) and non-holders (14.1%) had nearly the same default rate. Credit card holders churned more (approx. 17%) than non-holders (15%).

▪ **Insurance**: 39.2% have insurance. Insurance holders had slightly higher default (15.5%) than non-holders (13.3%). Insurance ownership did not make a difference for churn. Overall, products and offers did not strongly change risk patterns.

- **Key Takeaways from EDA:**

◦ **Loan History is the strongest predictor**: Customers with multiple loans are at much higher risk of both default and churn.

◦ **Income is not straightforward**: Middle-income customers (3–10 LPA) manage defaults better but are more likely to churn.

◦ **City-level differences** exist in both default and churn risk.

◦ **Employment Status**: Salaried and self-employed groups drive most defaults and churn due to being the largest customer base.

◦ **Products & Offers**: Credit card ownership, insurance, or offer responses do not strongly predict risk.

◦ **Correlation**: High correlations between income/expenses, spending categories, and delinquency/overdue days were identified, indicating potential multicollinearity for modeling.

◦ **Important Risk Indicators**: Expense-to-Income ratio, credit utilization, delinquency/overdue behavior, and credit score are identified as business-relevant drivers.

2. Predictive Modeling Summary

The modeling phase aimed to build models for default and churn prediction, addressing data challenges.

• **Leakage Check**: The project used 12-month aggregates for features and flags for default/churn as target variables, acceptable for a demo. In real-world scenarios, a T+1 prediction approach is recommended to avoid using future information.

• **Class Imbalance**: The dataset showed an imbalance with 10-20% positive cases (default/churn), which is common. Initial strategies included using class_weight=balanced for logistic regression or tree models to give more attention to minority classes.

• **Feature Engineering**:

◦ **Ordinal Encoding**: Applied to ordinal categories to preserve their natural order.

◦ **Expense to Income Ratio**: Created from AVG_MONTHLY_INCOME and AVG_MONTHLY_EXPENSE to capture financial stress and reduce collinearity. AVG_MONTHLY_INCOME was retained to capture earning capacity.

◦ **TOTAL_CARD_SPEND**: Created by summing AVG_RETAIL_SPEND, AVG_TRAVEL_SPEND, and AVG_BILLS_SPEND to reduce multicollinearity and capture overall spending behavior. For logistic regression, this combined feature was used, while tree-based models could potentially use separate features.

◦ **TOTAL_OVERDUE_DAYS_12M** was removed due to high collinearity with DELINQUENCY_COUNT_12M.

- **Model Performance Evaluation**:
  - **Logistic Regression (Baseline)**: Struggled to separate risky customers. For default, it had recall ~53% but very low precision (~20%). For churn, performance was close to random. This indicated that a simple linear model was insufficient for complex patterns and imbalanced data.
  - **Recommendations for Model Improvement**: Use advanced models (Random Forest, XGBoost), better data balancing, and fine-tune thresholds.
  - **Random Forest (with Adjusted Thresholds)**:
    - **Overfitting**: Initial Random Forest models completely overfit the training data (Train AUC = 1.0), showing no generalization (Test AUC = 0.654 for default, 0.507 for churn).
    - **Default Model**: Achieved 75% recall (catching most defaulters) but with low precision (20%), meaning many false positives. Useful as a risk detector but not for precise decision-making.
    - **Churn Model**: Caught ~72% of churners but with even lower precision (15%), making it unreliable for targeted retention offers.
    - **Conclusion**: Class imbalance was still not handled strongly enough, and the model was overfitting. XGBoost with class weighting was suggested as a next step due to its better handling of imbalance.
  - **XGBoost / SMOTE with Logistic Regression / SMOTE with XGBoost:**
    - **Overall Challenge**: Simple models underfit, while complex ones overfit and default to majority class predictions, mainly due to class imbalance and limited positive samples.
    - **SMOTE with Logistic Regression**: Improved recall (Default: 43%, Churn: 34%) compared to plain Logistic Regression, catching more risky customers, but overall predictive power remained weak with high false alarm rates.
    - **SMOTE with XGBoost**: Performed very strongly on training data but failed to generalize to unseen customers. For default, it caught only 25% of defaulters with high false alarm rates (AUC 0.61). For churn, performance was worse than random guessing (AUC 0.44), showing it completely failed to capture churn signals.
    - **Key Conclusion from Modeling**: Simply rebalancing and using advanced models is not enough. The underlying issue is that the current dataset lacks strong predictive signals for churn and defaults.

3. Overall Conclusion

- **Default Prediction Potential**: The models showed potential for default prediction as an early-warning tool, with recall rates up to 75%. However, low precision means banks should use it for monitoring at-risk customers, with final decisions requiring additional checks.

• **Churn Prediction Limitations**: Churn models did not generalize well with the current features. This highlights a need for more behavioral and temporal data to better understand customer loyalty.

• **Problems Addressed**:

◦ **Customer Risk Segmentation**: Identified groups at higher default risk (e.g., 3–5 loans, lower credit scores, higher delinquency) and found very low/moderate credit utilization correlated with risk.

◦ **Product & Service Gaps**: Customers with multiple loans and high overdue days showed financial stress. Churn was more common in salaried/self-employed groups.

◦ **Engagement & Retention Issues**: Marketing offers did not strongly influence retention. Handling complaints well reduced churn for those with many complaints, while 1-2 complaints correlated with higher churn.

◦ **Behavioral Insights**: Digital usage was generally healthy, but heavy users were not necessarily more loyal. High-value customers were not fully secured from churn/default, emphasizing service quality.

4. Recommendations for Stakeholders

Based on the EDA and predictive modeling results, the following recommendations are provided to enhance risk management and customer retention strategies:

• **Risk Monitoring & Early Warning:**

◦ Closely monitor customers with 3–5 loans, high delinquency counts, and credit scores below 600. These segments are consistently identified as high-risk.

◦ Track under-utilizers (low credit usage) and moderate utilizers (13-22%) for hidden risk signals, as both patterns correlate with churn or default.

◦ Utilize the current default models for monitoring at-risk customers as an early-warning system, while acknowledging the need for additional checks due to precision limitations.

◦ Strengthen collections strategy for customers with repeated delinquencies or high overdue days.

• **Retention Strategy:**

◦ Improve how complaints are handled; early and effective complaint resolution may prevent churn and build loyalty, especially for customers with 1-2 complaints.

◦ Re-design marketing offers; current response patterns do not strongly reduce churn. Focus on personalized offers based on customer behavior rather than broad campaigns.

◦ Proactively engage high-value customers (those with higher incomes and deposits) with exclusive services, loyalty programs, or priority banking to prevent defection, as they are prone to churn if not given premium experiences.

◦ Track churn-prone segments like salaried and self-employed customers with regular health checks on their accounts.

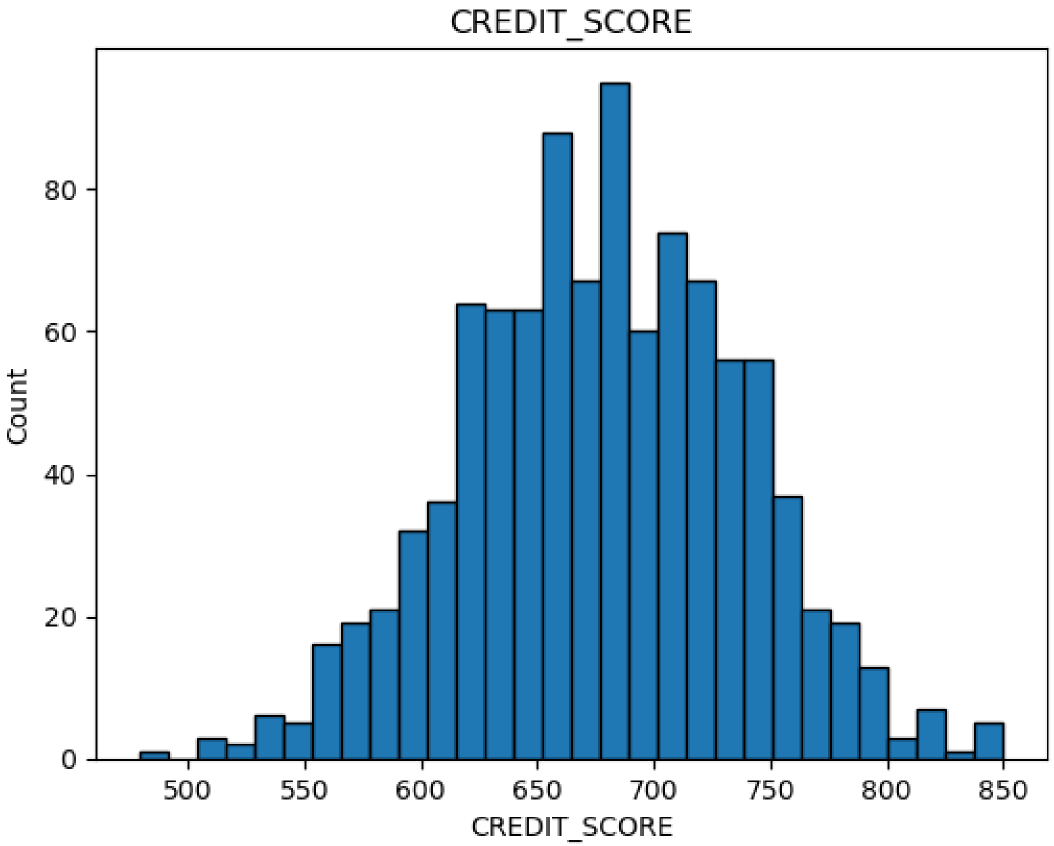
- **Customer Experience Improvements:**

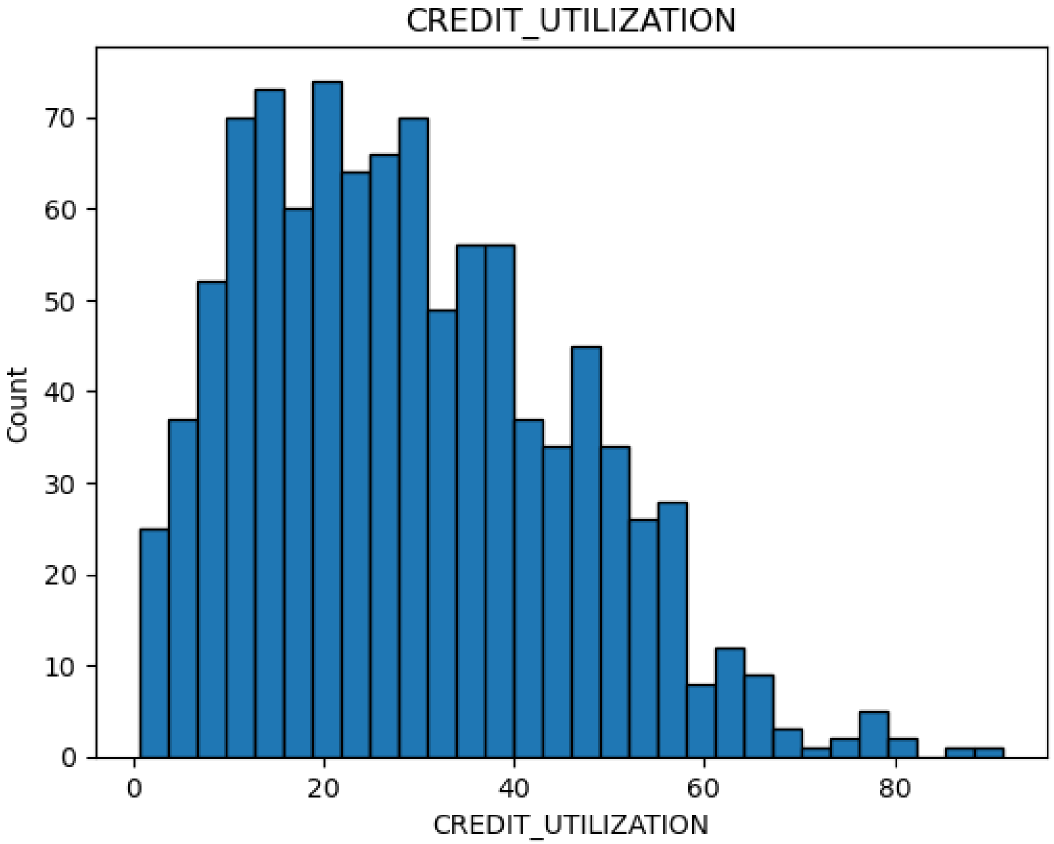
  ◦ Investigate high-expense customers to ensure they are not over-leveraged, and simultaneously offer them tailored financial products.
  ◦ Address city-specific churn issues, particularly in Mumbai and Hyderabad, by investigating local service quality or competition.

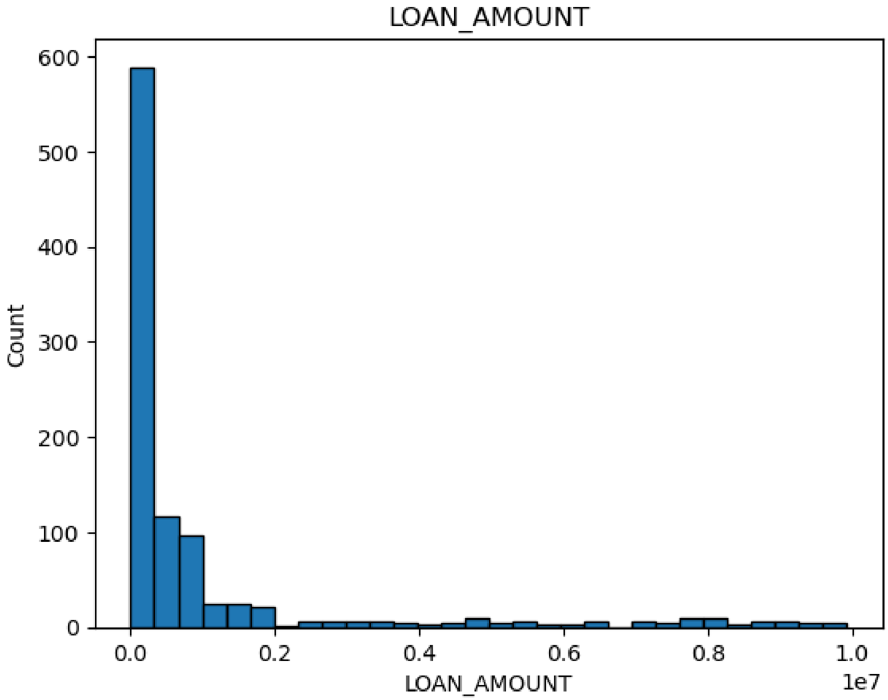
- **Operational & Data Focus for Future Enhancements:**

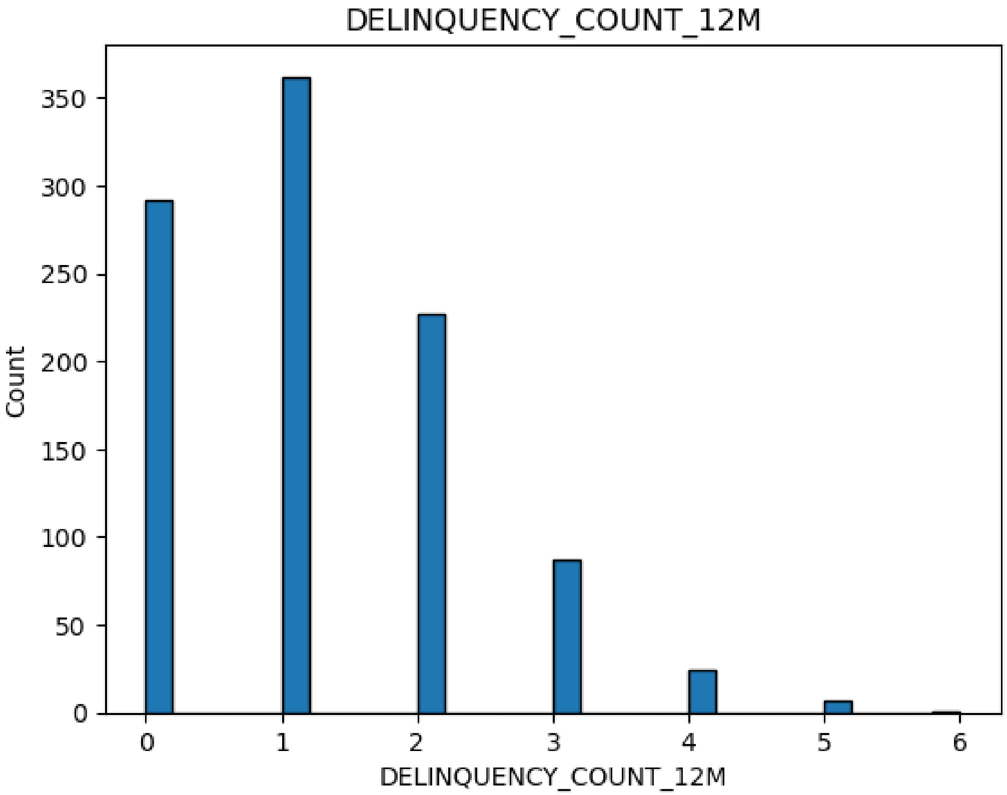
  ◦ Invest in collecting richer behavioral and temporal features for churn prediction, as current features proved insufficient for reliable churn models.
  ◦ Expand analytics with trend-based features for more reliable predictions in the future.
  ◦ Consider city-level strategies to manage default risk, especially in cities like Bengaluru, Kolkata, and Ahmedabad, where default rates are higher.

Visualizations:

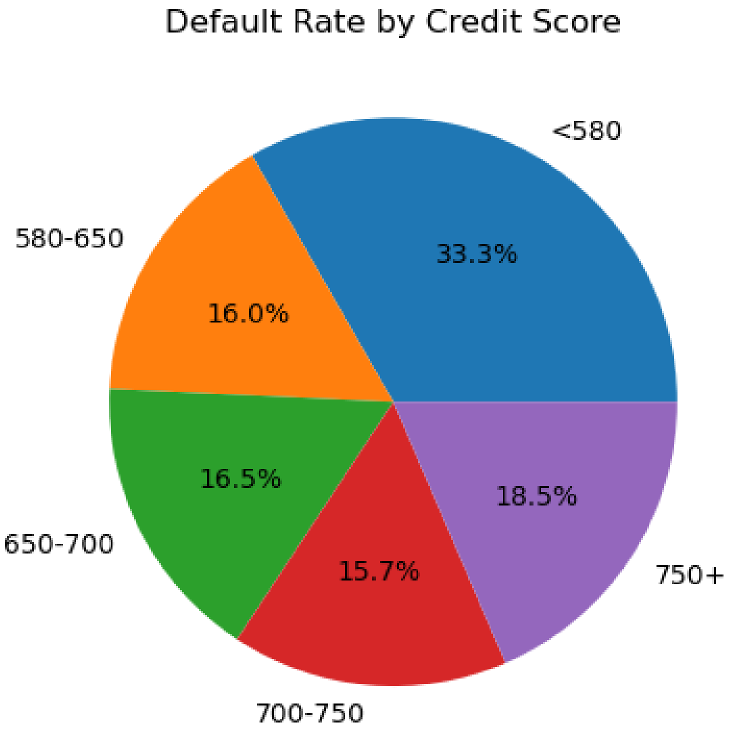



CREDIT_UTILIZATION
Count
0
10
20
30
40
50
60
70
0
20
40
60
80
CREDIT_UTILIZATION


LOAN_AMOUNT
Count
0
100
200
300
400
500
600
0.0
0.2
0.4
0.6
0.8
1.0
1e7
LOAN_AMOUNT

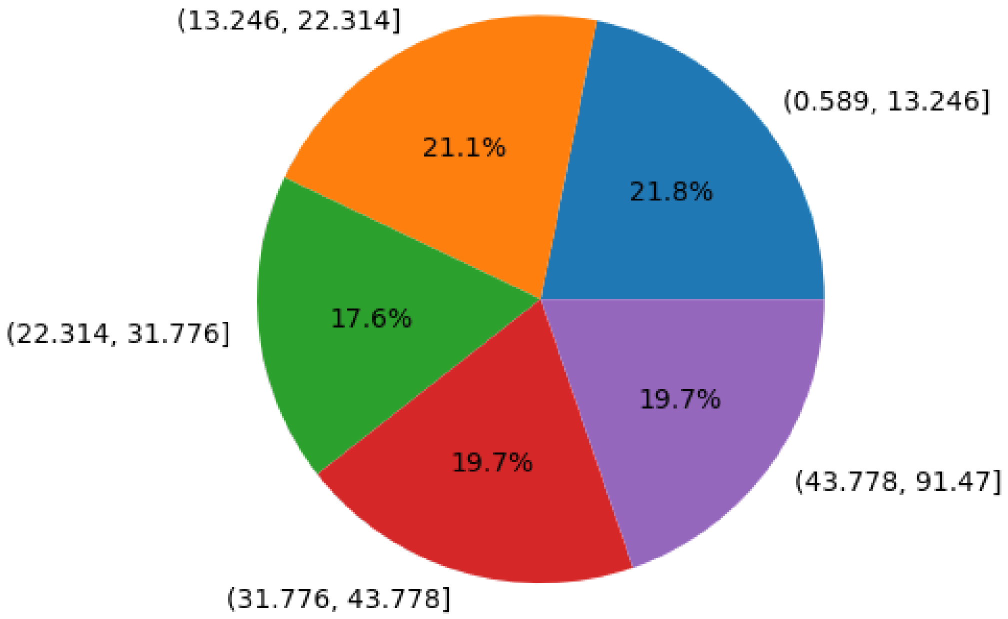


DELINQUENCY_COUNT_12M

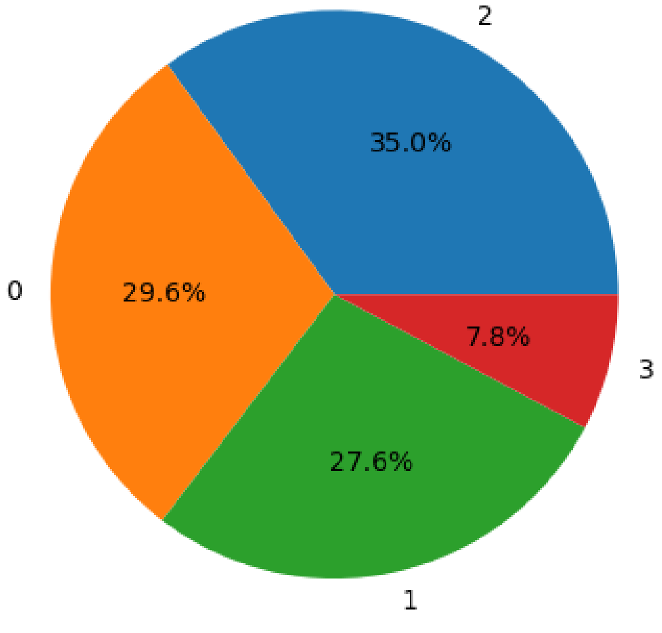


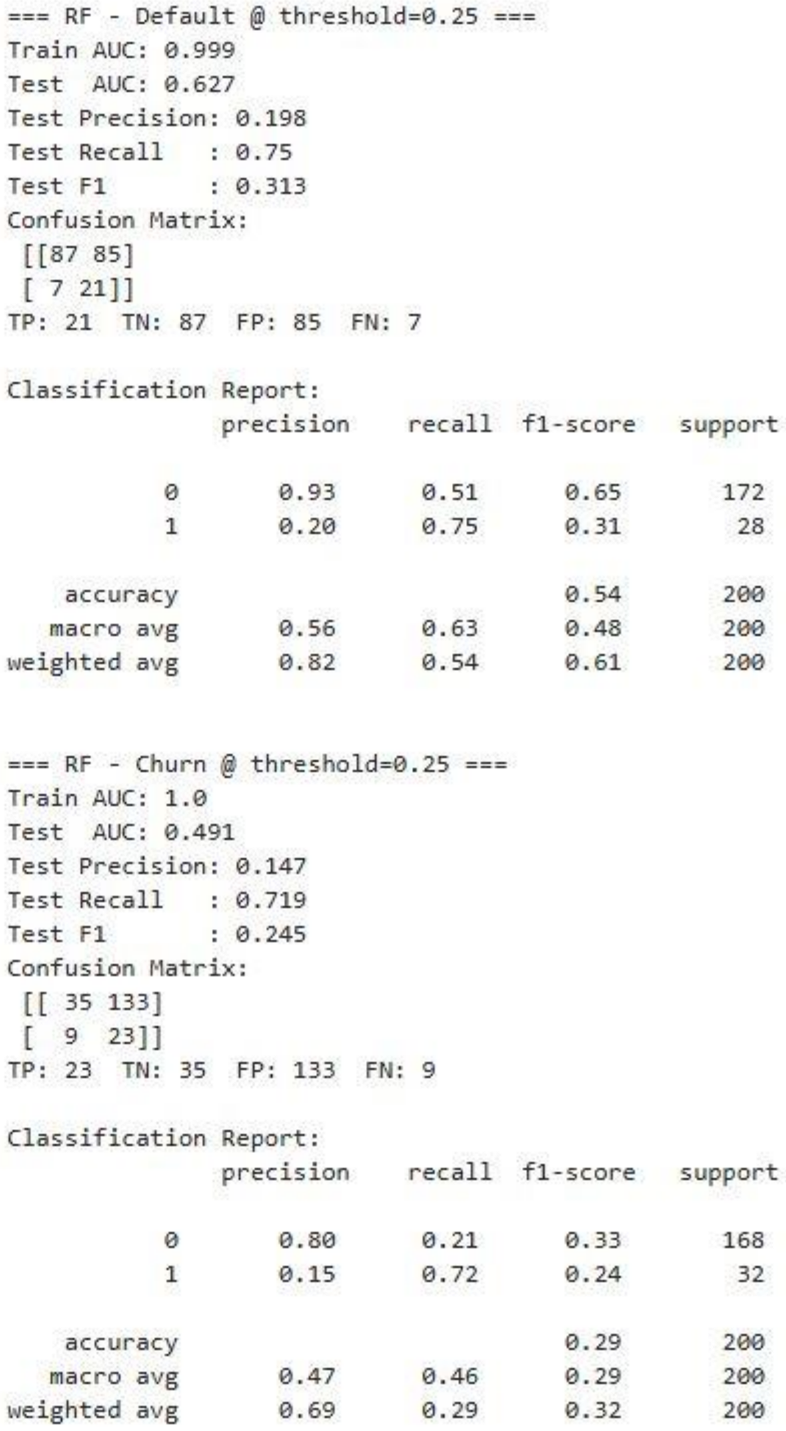
Default Rate by Credit Score

Credit Utilization vs Default



Churn Rate by Complaints Raised





```
=== RF - Default @ threshold=0.25 ===
Train AUC: 0.999
Test  AUC: 0.627
Test Precision: 0.198
Test Recall   : 0.75
Test F1       : 0.313
Confusion Matrix:
 [[87 85]
 [ 7 21]]
TP: 21  TN: 87  FP: 85  FN: 7

Classification Report:
               precision    recall  f1-score   support

           0       0.93      0.51      0.65       172
           1       0.20      0.75      0.31        28

    accuracy                           0.54       200
   macro avg       0.56      0.63      0.48       200
weighted avg       0.82      0.54      0.61       200


=== RF - Churn @ threshold=0.25 ===
Train AUC: 1.0
Test  AUC: 0.491
Test Precision: 0.147
Test Recall   : 0.719
Test F1       : 0.245
Confusion Matrix:
 [[ 35 133]
 [  9  23]]
TP: 23  TN: 35  FP: 133  FN: 9

Classification Report:
               precision    recall  f1-score   support

           0       0.80      0.21      0.33       168
           1       0.15      0.72      0.24        32

    accuracy                           0.29       200
   macro avg       0.47      0.46      0.29       200
weighted avg       0.69      0.29      0.32       200
```

Random Forest with Tuned Threshold