Data Analysis

Credit Risk & Customer Retention Analysis

Using Data Analytics to Identify Risk & Opportunities

Shijin Ramesh | Data Analyst | August 24, 2025

Executive Summary

- This project analyzes **1,000 bank customers** to understand risk of **loan default** and **customer churn**.
- Approach: Data cleaning, exploratory data analysis (EDA), feature engineering, predictive modeling, and dashboard building in Power BI.
- Outcome: Identified **risk patterns**, built a **risk flagging model**, and provided **recommendations** to stakeholders for credit risk monitoring and customer retention.

Business Problem

- **Default Risk:** Customers failing to repay loans which is a financial loss to bank.
- **Churn Risk:** Customers leaving the bank → loss of revenue, higher acquisition costs.
- **Need:**
  - Early detection of risky customers.
  - Retain valuable customers through data-driven insights.

Dataset Overview

- **Customers Analyzed:** 1,000
- **Features:**
  - Demographics - Age, Gender, Income Bracket, Employment Status, City
  - Credit & Loan - Credit Score, Loan History, Loan Type, Loan Amount, Credit Utilization
  - Engagement - Mobile App Usage, Net Banking Usage, Complaints Raised, Response to Offers
  - Investments - FD Investment, Insurance, Credit Card ownership
- **Targets:**
- Default_Flag (customer defaulted or not)
- Churn_Flag (customer churned or not)

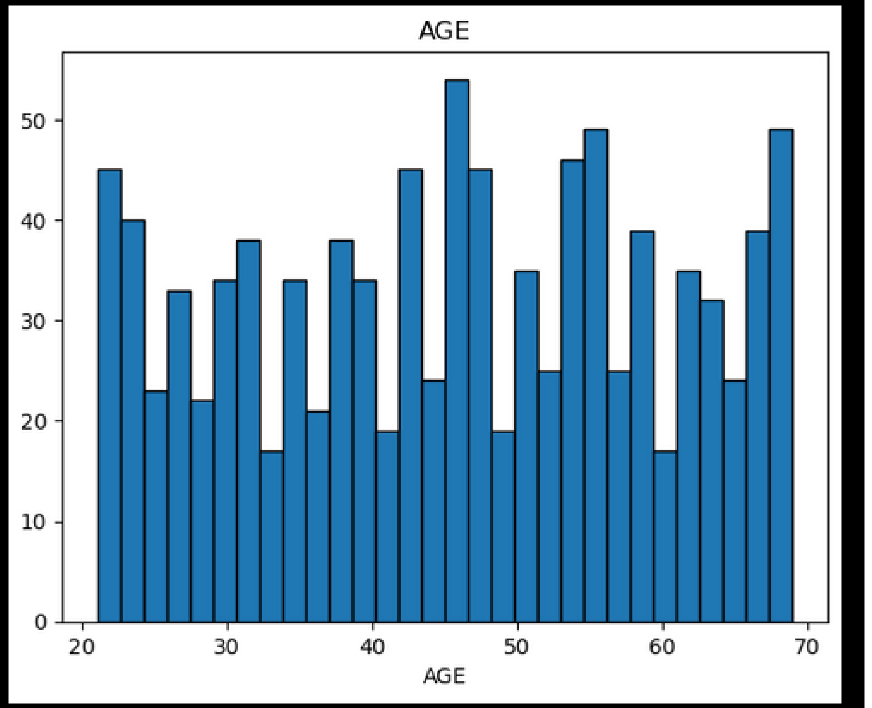
Methodology

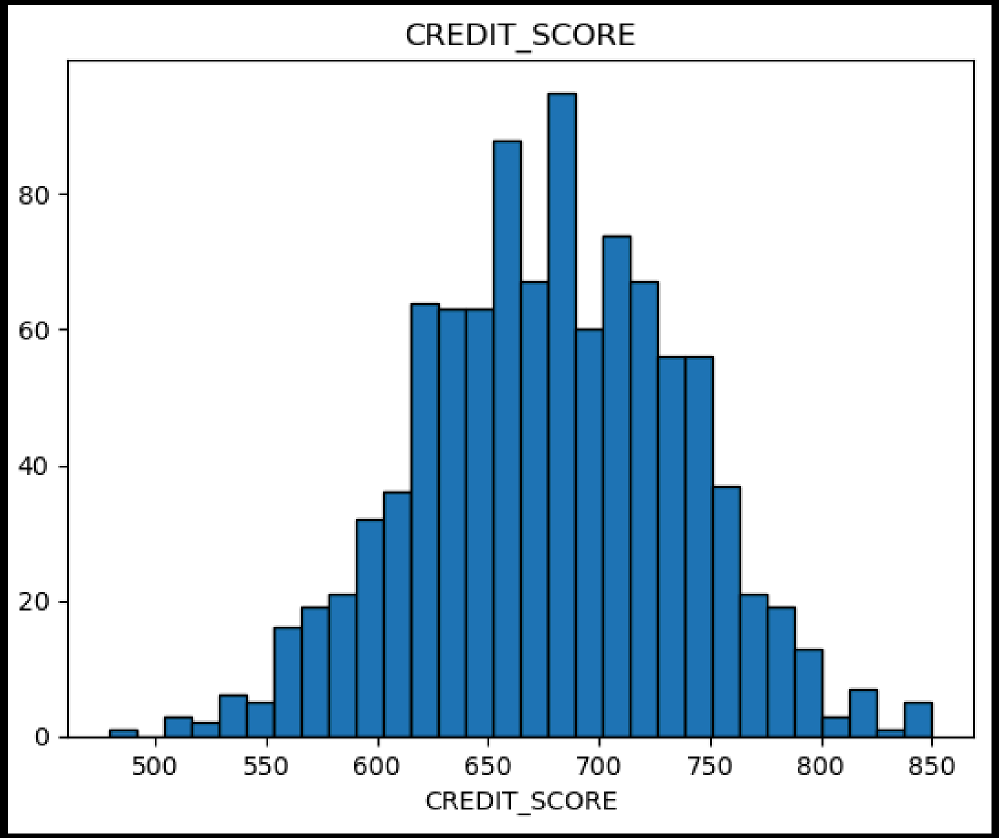
- **EDA:** Distribution analysis, outlier detection, categorical breakdown.
- **Feature Engineering:**
  - Created Expense-to-Income Ratio.
  - Designed High-Risk Flag using rules (Credit Score <600, Utilization >60%, Delinquencies, Loan History).
- **Modeling:** Tested Logistic Regression, Random Forest, XGBoost, SMOTE balancing, threshold tuning.
- **Dashboarding:** Designed interactive Power BI dashboard with KPIs, distributions, and risk insights.

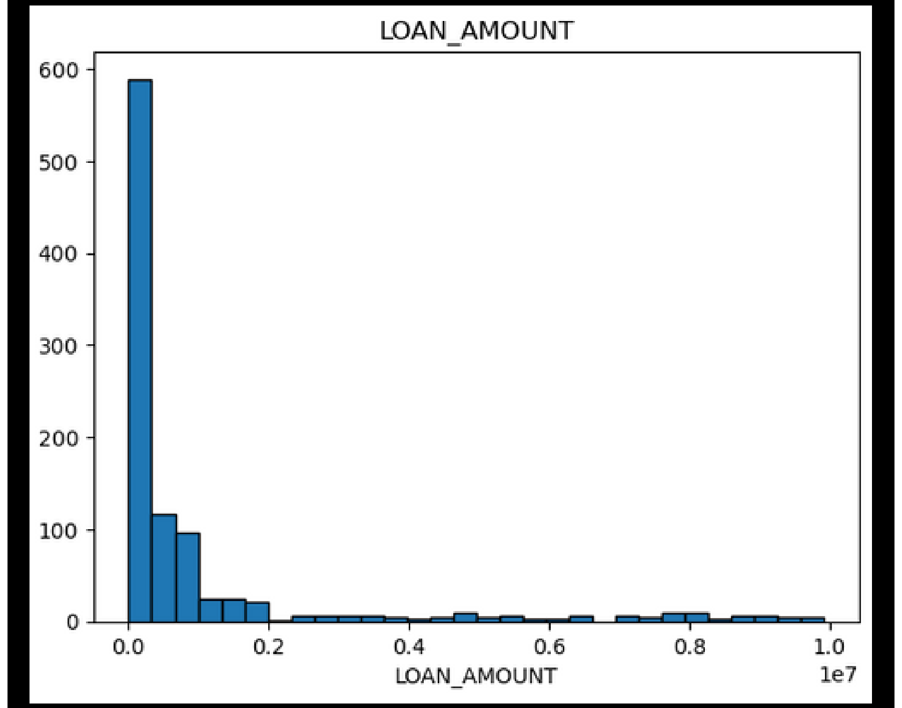
Tools used: Snowflake SQL, Python (Pandas), Power BI

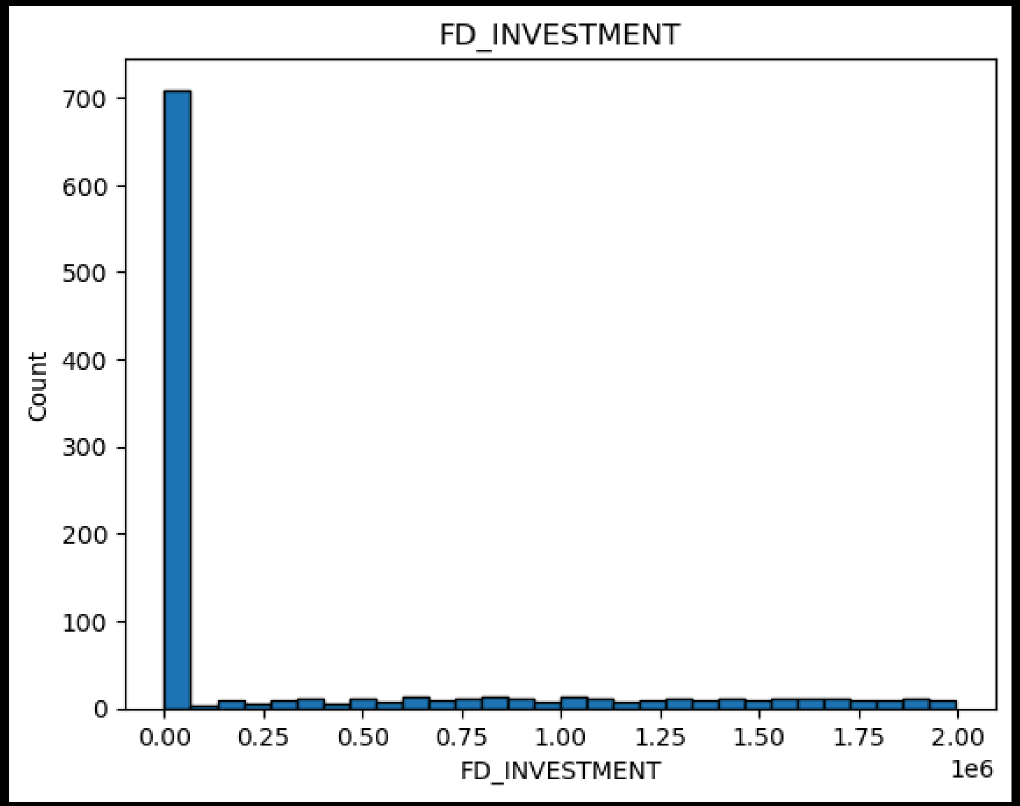
Key EDA Insights (1)

- **Age:** Customer base centered around 45-55 years.
- **Credit Score:** Majority between 600–750; <500 flagged as high risk.
- **Loan Amount:** Skewed, most customers took small loans; very few high-value loans.
- **Fixed Deposits:** 700+ customers kept deposits around 1 lakh; only few invested >10 lakh.



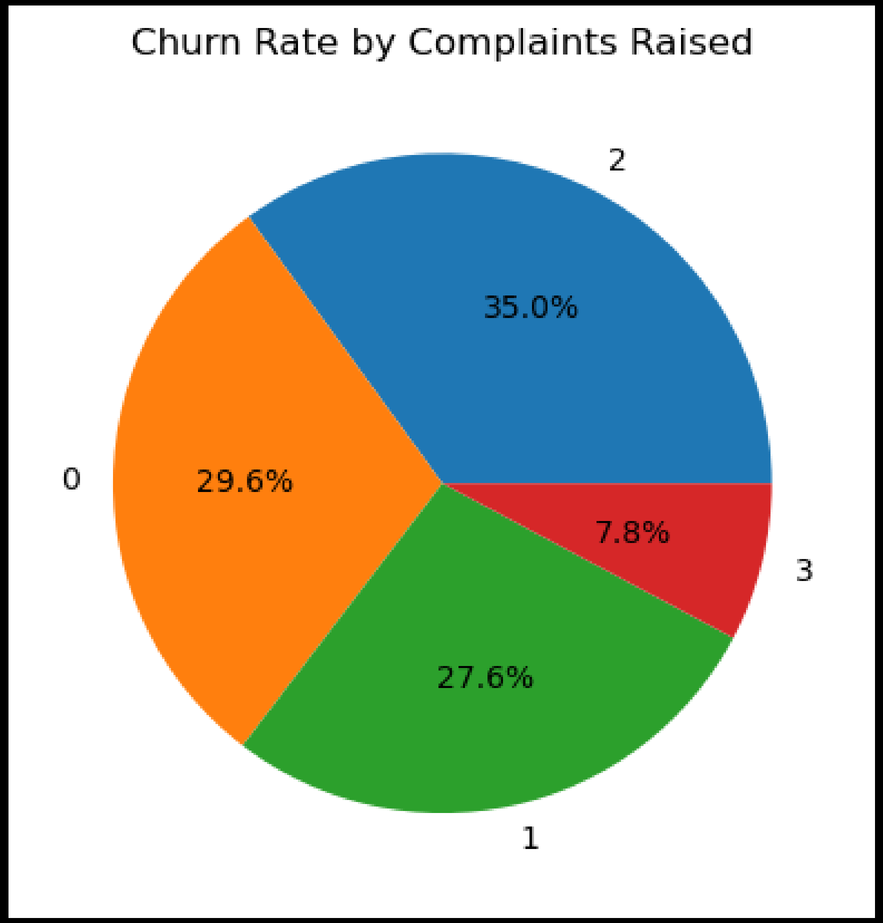


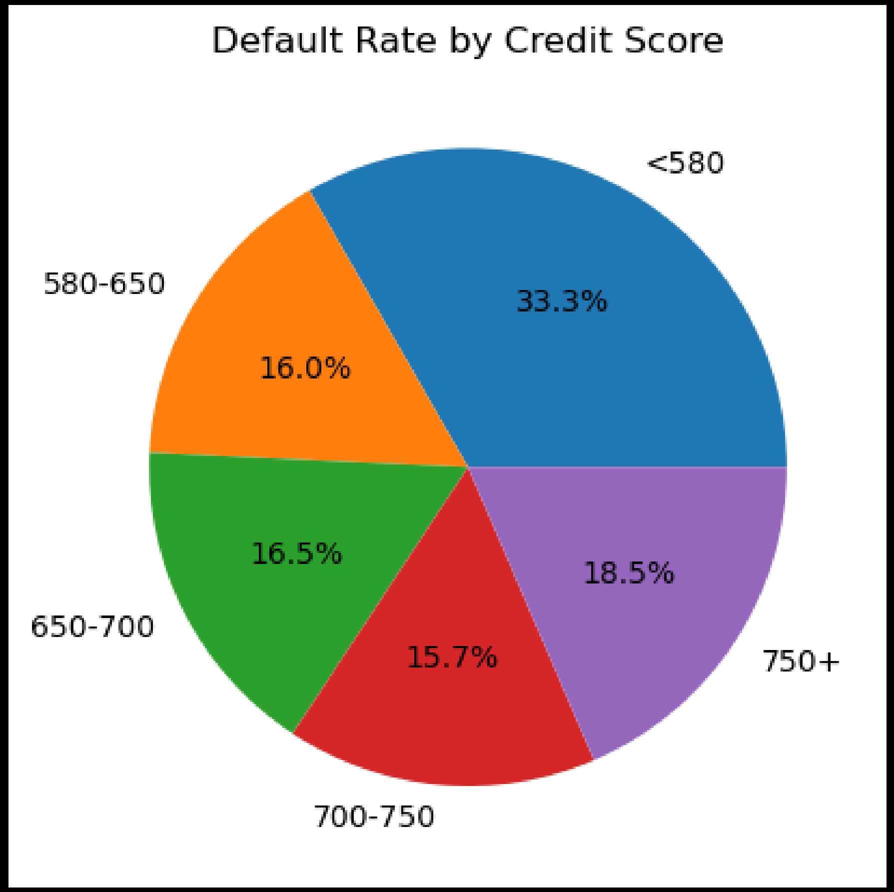


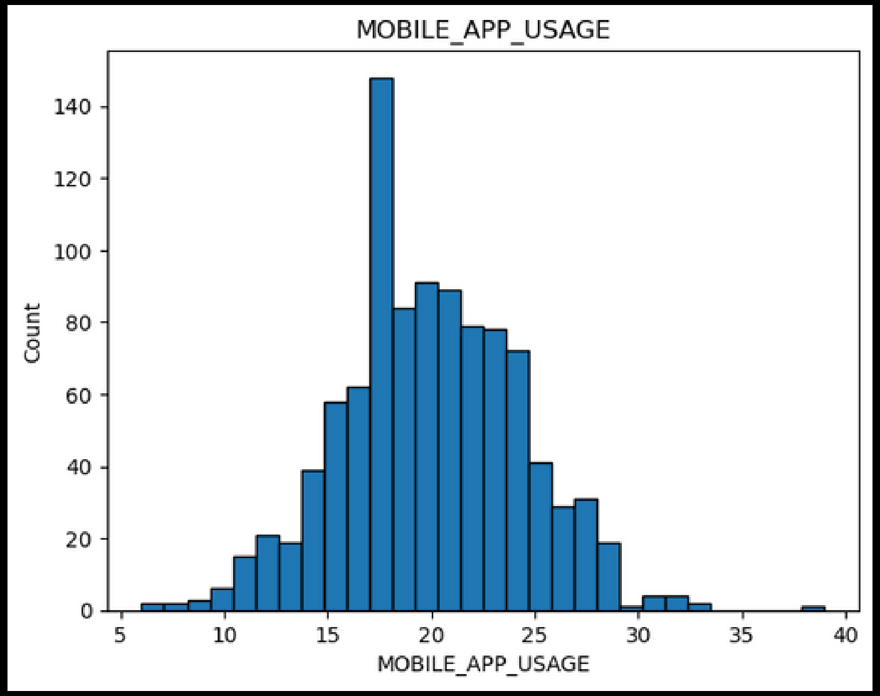


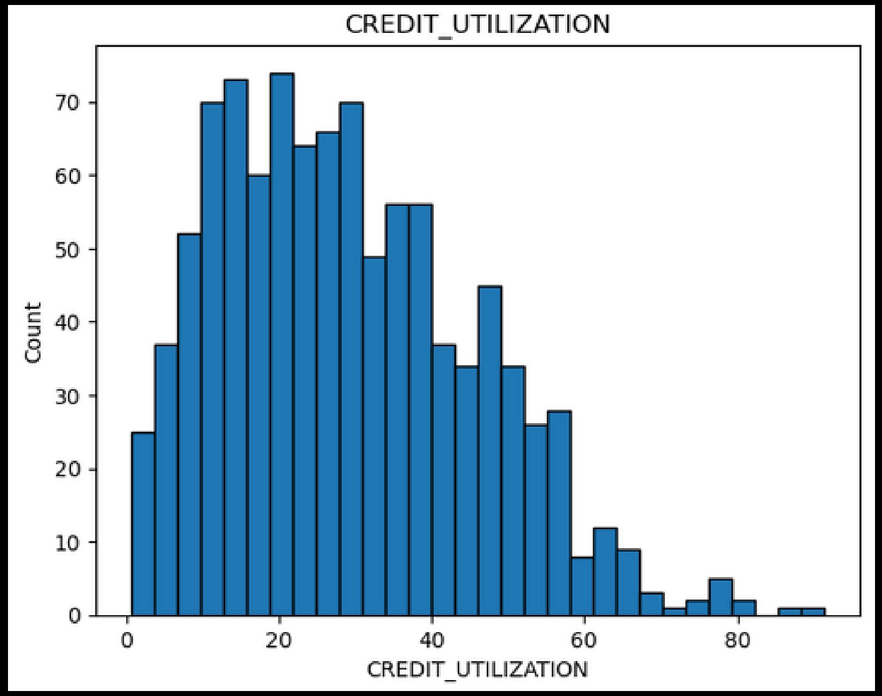
Key EDA Insights (2)

- **Credit Utilization:**
  - Most between 10–40% (healthy).
  - Small group >60% (risky).
- **Complaints:**
  - 600+ customers raised no complaints.
  - Customers with resolved complaints showed more loyalty (lower churn).
- **Engagement:**
  - Mobile app usage normally distributed.
  - Net banking less frequent but used for high-value transactions.



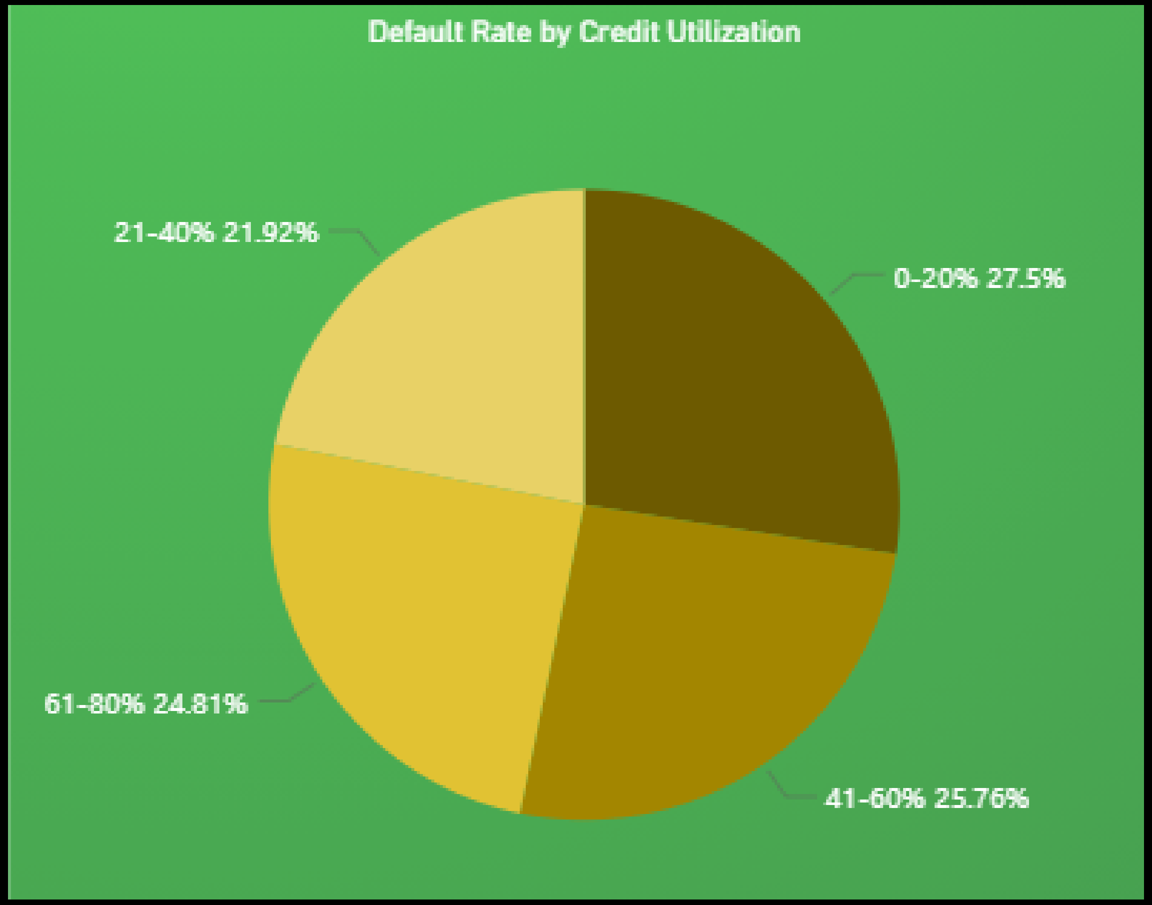


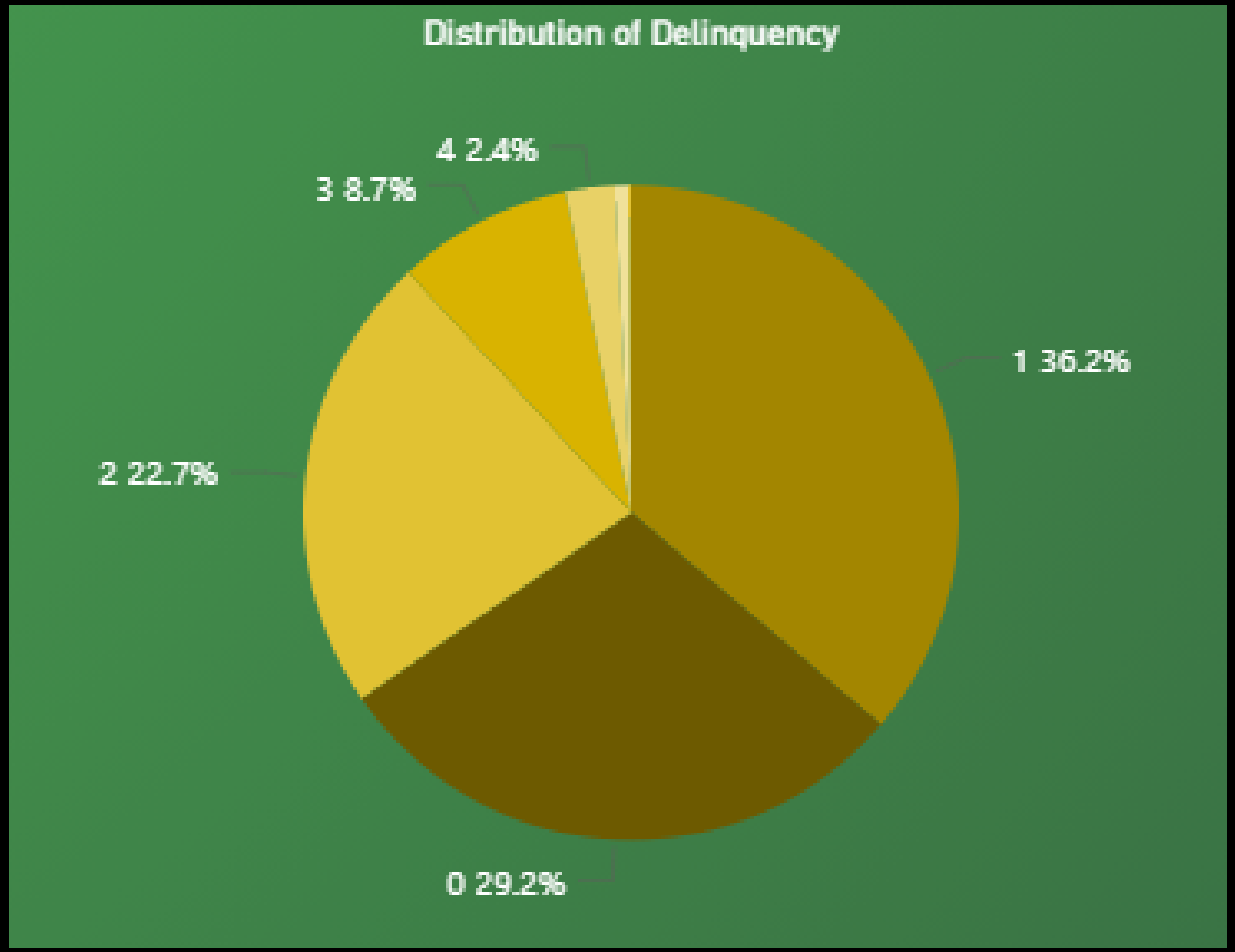


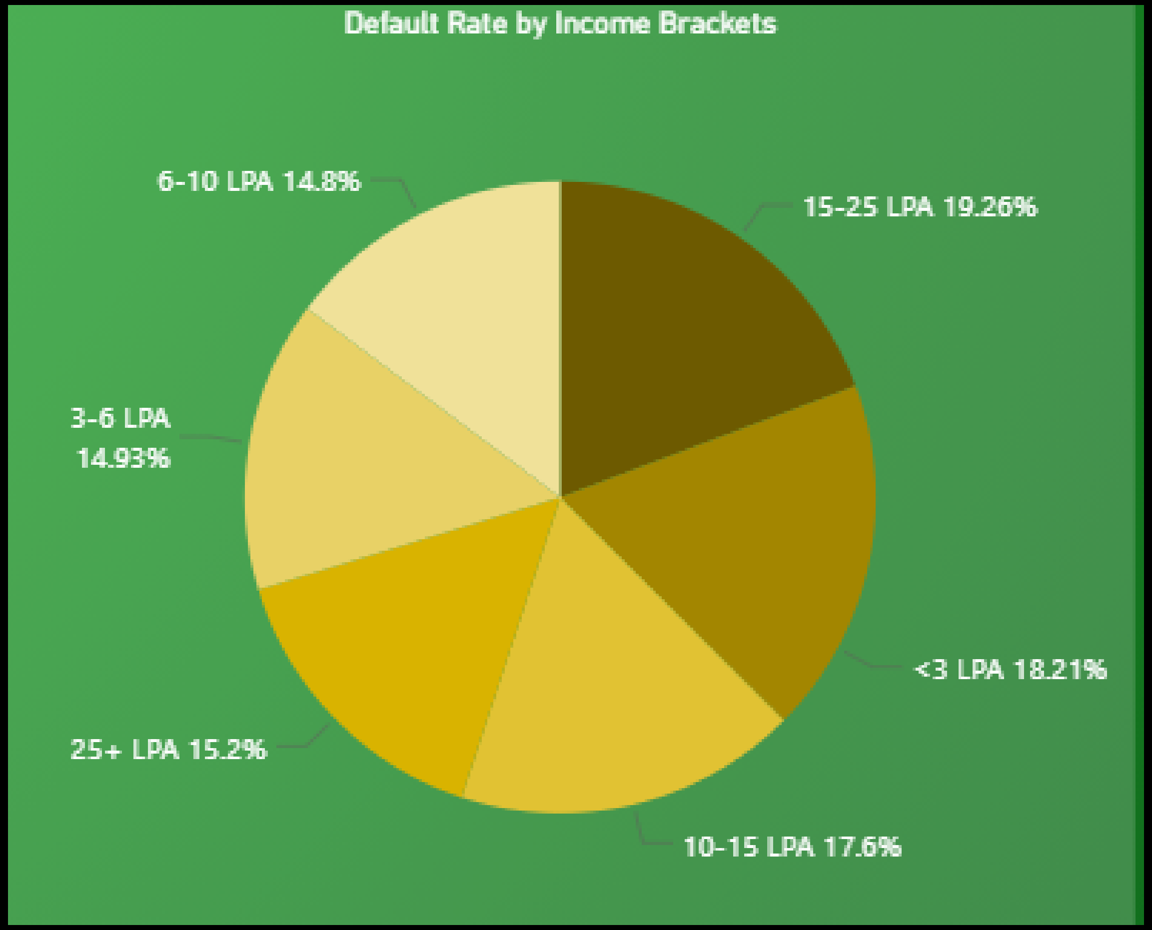


Risk Analysis

- **Loan History:**
  - Customers with less than 3 or 3–6 loans - 33.14% default rate.
  - Customers with 6+ loans → 66.86% default rate.
- **Credit Utilization:** Defaults are more in the utilization band of 0-60%.
- **Delinquencies:** Customers with repeated late payments are high-risk.
- **Overdue Days:** Some extreme cases >120 days are very risky customers.

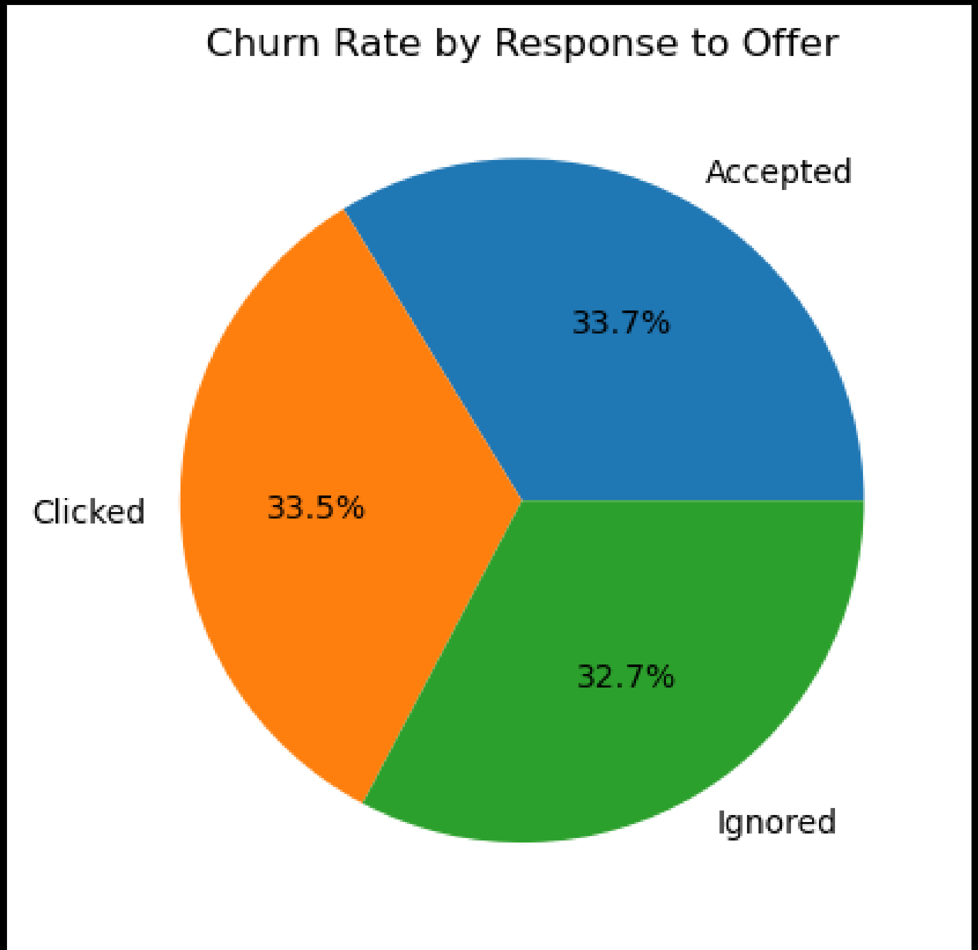


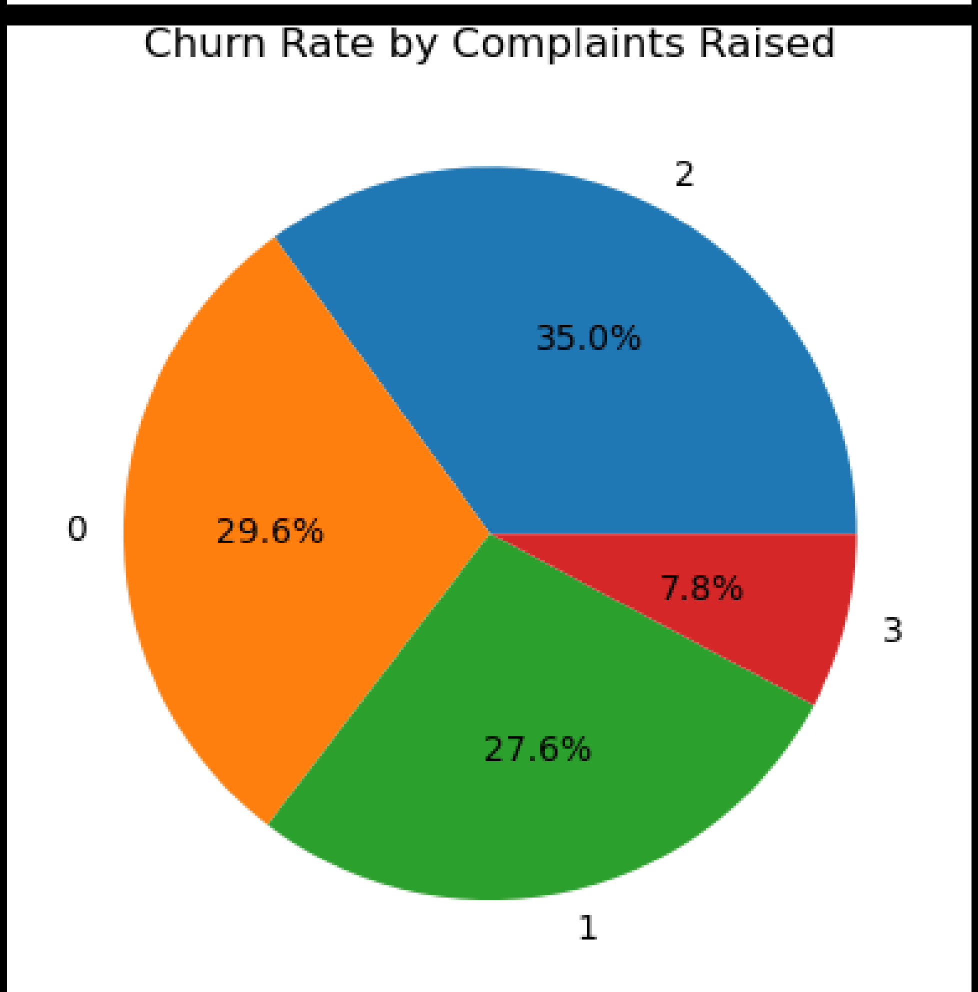




Engagement & Churn

- **Response to Offers:**
  - Customers who accepted/clicked offers had churn ~33%.
  - Those who ignored - churn ~32%.
  - Insight: Churn not strongly impacted by offers.
- **Complaints vs Churn:**
  - 0 complaints - churn rate ~29.6%.
  - 1–2 complaints - higher churn.
  - 3+ complaints - churn dropped to 7.8% (likely because issues were resolved well).
- **Takeaway:** Churn more linked to customer experience quality than offers.





Predictive Modeling

- **Logistic Regression:** Gave baseline AUC ~0.59–0.62; models leaned toward majority class.
- **Random Forest:** Strong recall (75%) after threshold tuning, but weak precision.
- **XGBoost:** Overfitted; test performance weak even with class balancing.
- **SMOTE + Models:** Balanced datasets improved recall but still weak precision.
- **Conclusion:** Models useful as early-warning tools rather than precise predictors.

```
=== 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
```

Power BI Dashboard

[Live Dashboard]

Recommendations

- **Credit Risk:**
  - Monitor customers with Credit Score <600, Utilization >60%, and multiple loans.
  - Prioritize collection strategies for high delinquency customers.
- **Customer Retention:**
  - Improve complaint handling & link resolution speed to loyalty.
  - Segment offers by income/city to improve targeting.
- **Data Strategy:**
  - Collect richer behavioral/temporal data for churn modeling.
  - Track repayment behavior, customer service interactions, and satisfaction.

Business Impact

- **Risk Reduction:** Early identification reduces defaults & financial losses.
- **Customer Loyalty:** Faster complaint resolution improves retention.
- **Cost Savings:** Targeted retention campaigns cheaper than new acquisition.

Conclusion

- EDA uncovered clear risk & engagement patterns.
- Predictive models, while not perfect, work as risk radar systems.
- Dashboard provides some actionable insights.
- Data-driven strategies help reduce defaults & strengthen customer trust.

Appendix

- *Code Link: https://github.com/shijin/CustomerCreditRiskAnalysis*