2025 Illinois Statistics Datathon

Team Conquistadors
Vinit, Shobha, Anushka, Nagarathna

Our goal was to build a **Predictive Framework** for managing **Credit Lines**

Forecasting customer spending for Q4 2025

Segmenting accounts by credit line eligibility & risk

Recommending personalized credit line increases or no change

What we analysed

- Financials
- 8 Structured Tables
- Credit Behaviors
- Fraud metrics
- Account summaries

Preprocessing the Data

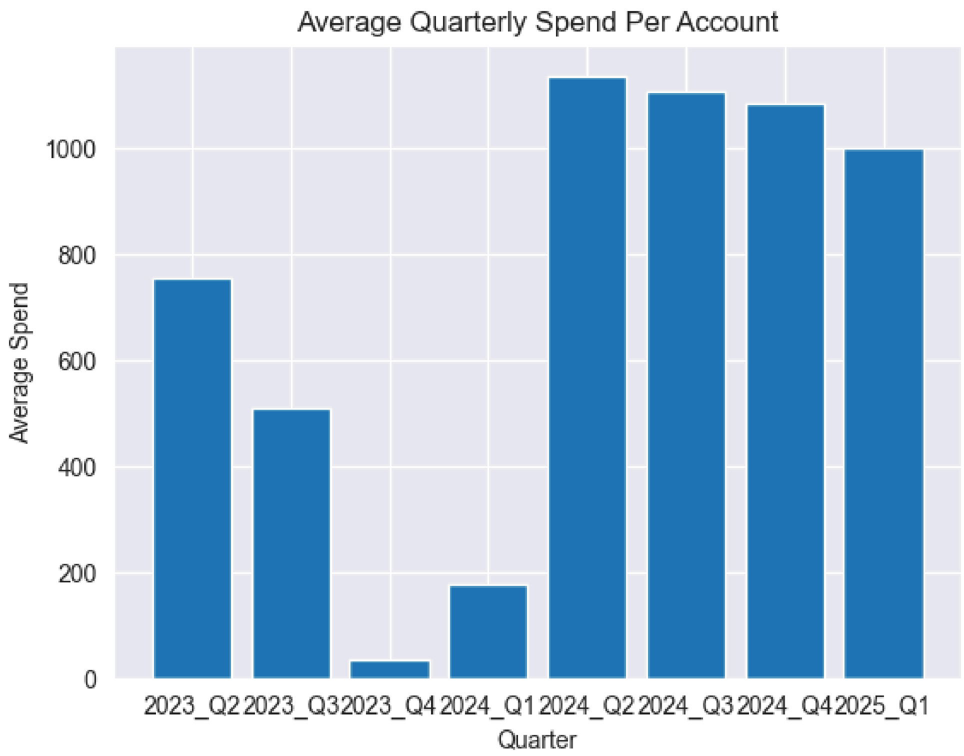
Data Cleaning and Merging

The data consisted of 8 structured tables covering financials, credit behaviour, fraud measures, delinquencies, and account summaries.

- Columns with **>70% missing values** were removed to ensure data quality.
- Tables were combined using primary and foreign keys as stated in the mapping specification.
- Accounts with **null credit line values** were excluded, since they are not eligible for limit adjustments.
- Retained only **active accounts**.

Exploratory Data Analysis

Insights from Historical Data



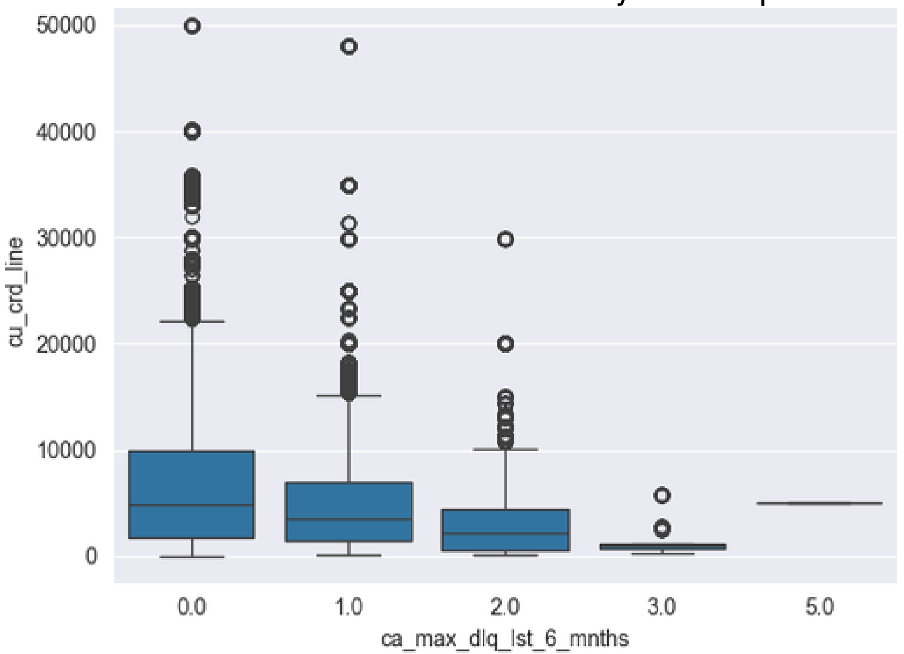
- Examined average quarterly spending patterns(i.e Q2 2023 to Q1 2025)
- Observed correlation between Confidence Level & Credit Line
- Evaluated customer spending across multiple periods to assess repeat behavior and brand loyalty.

synchrony

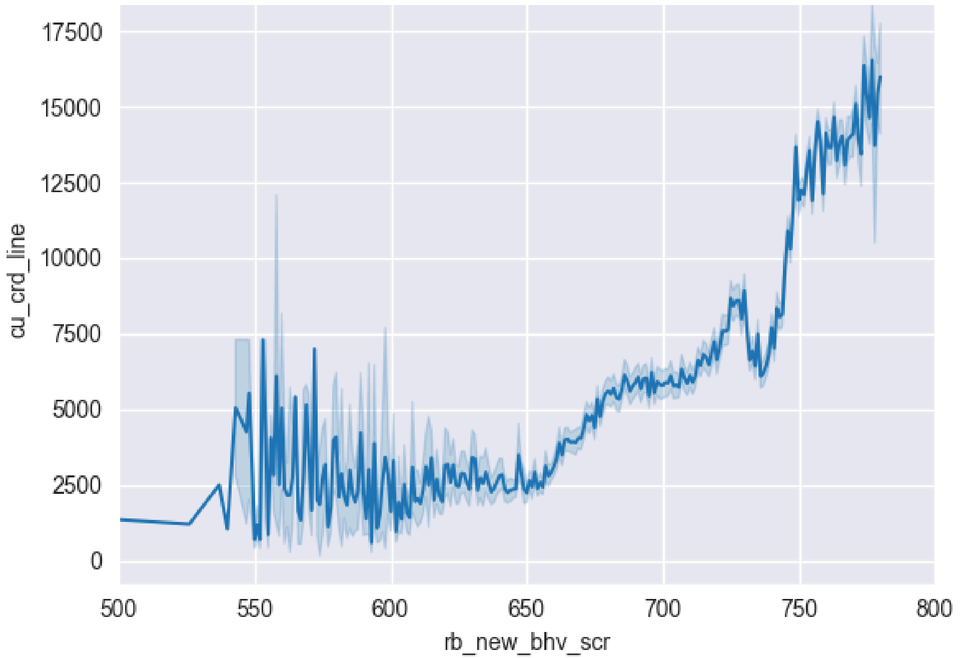
Exploratory Data Analysis

Insights from Historical Data

Credit Line vs Number of Days Delinquent



Credit Line vs Behaviour Score



- Examined correlation between Delinquency, Customer Behaviour Score with Credit Line.
- Observed Negative correlation between Delinquency & Credit Line.
- Observed Positive correlation with Behavioral Score & Credit Line.
- Since there is a visible correlation between these metrics we used them in making predictions for future Q4 credit line adjustments.

synchrony

Feature Engineering

Total Accounts

13,841

Engineered features

43

Features Including

Average Quarterly Spend
(from Q2 2023 to Q1 2025)

Average Quarterly Open-to-Buy (OTB)
(from Q3 2024 to Q1 2025)

Total Net Fraud Amount & Excl Flag
If an account is tagged for exclusion from credit limit changes and an account's total fraud amount

Credit Utilization
(balance / credit limit)

Behavioral Scores
(internal and bureau)

Delinquency metrics:
days past due, NSF counts, returned checks

Objective 1: Predict Q4 2025 Customer Spending

Model Used: GradientBoostingRegressor

Input Features (X): **All behavioral, credit, and historical spending features**

Target Variable (y): **avg_spend_2025_Q4 (predicted)**

Training Data Range: Q2 2023 – Q1 2025

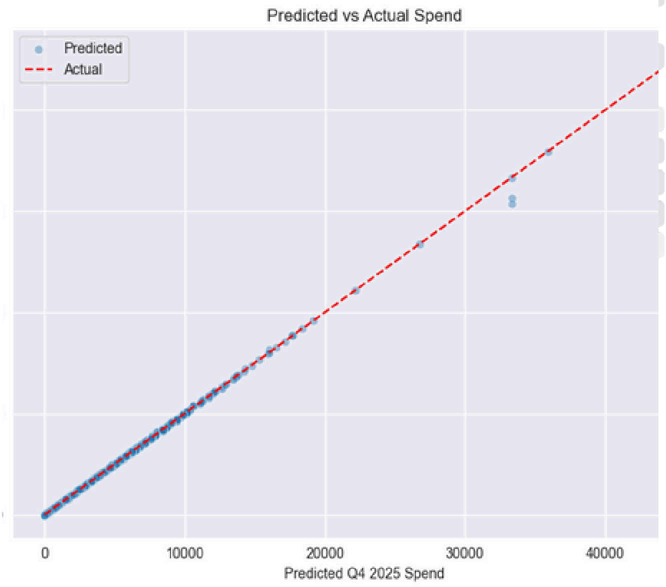
Verification method: Used avg_spend_2024_Q4 as proxy for validation

Model Metrics

RMSE: 64.046

R^2 Score: 0.9985

Result - Predicted vs Actual Q4 2025 Spend



Objective 2: Segment Accounts for Credit Line Eligibility

Step 1: Manual Labeling using Business Logic

1. Eligible – No Risk (Class 0):
• Predicted Q4 spend > $300
• Utilization > 50%
• Bureau Score > 600 and Behavioral Score > 550

2. Eligible – With Risk (Class 1):
• High spenders with good utilization but low credit scores

3. No Increase Required (Class 2):
• Moderate/low spenders or underutilized accounts

4. Non-Performing / High Risk (Class 3): • Fraud > $100 • NSF count > 3
• Days delinquent > 60
• Behavior score < 400
• Returned checks > 3

Step 2: Model Used

- We experimented with multiple models like XGBoost Classifier, RandomForestClassifier is used for the final analysis.
- Due to class imbalance, we implemented SMOTE to address the issue.
- We tested combinations of SMOTE with Random Forest and SMOTE with XGBoost.
- After evaluation, we concluded that SMOTE + XGBoost provided the best performance.

Model Performance

High Accuracy for Majority Classes
6,510 No-Increase accounts correctly predicted
3,894 Eligible-Safe accounts correctly predicted
Only **9** No-Increase accounts wrongly predicted as Eligible-Risk

Rare Class Segmentation Still Effective
70 Eligible-Risk accounts correctly classified
2,788 High-Risk accounts correctly predicted
0 misclassifications of High-Risk customers

Model Performance

Class	Description	Precision	Recall	F-1 Score	Support
0	No Increase Required	0.89	0.87	0.88	1372
1	Eligible – No Risk	0.80	0.82	0.81	824
2	Eligible – With Risk	0.47	0.53	0.50	15
3	Non-Performing / High Risk	1.00	1.00	1.00	558
accuracy				0.88	2769
macro avg		0.79	0.81	0.80	2769
weighted avg		0.88	0.88	0.88	2769

Insights

- Overall Accuracy: 88%
- Excellent at classifying high-risk and low-risk groups eligible for increase
- Slight underperformance in minority class (Class 2) due to class imbalance
- Weighted Average F1 Score: 0.88 confirms general robustness of the model.

synchrony

Objective 3 :Recommend Credit Line Adjustments

To recommend **personalized credit line increases**, we built a **regression-based framework** that balances **spending demand, utilization trends, and risk factors.**

Model
RandomForestRegressor

Applied to
Accounts from Segments 1 (without risk) and 2(with risk)

Features
Current credit line, utilization, spending, behavioral scores

Output
Personalized credit line increase per account

High Risk Check
1 in 173 accounts expected to commit fraud
• 3+ NSF incidents • 2+ returned Checks YTD • 45+ Delinquent or max deliquency in 6 months > 30 • Low behavioral or recent risk scores <400

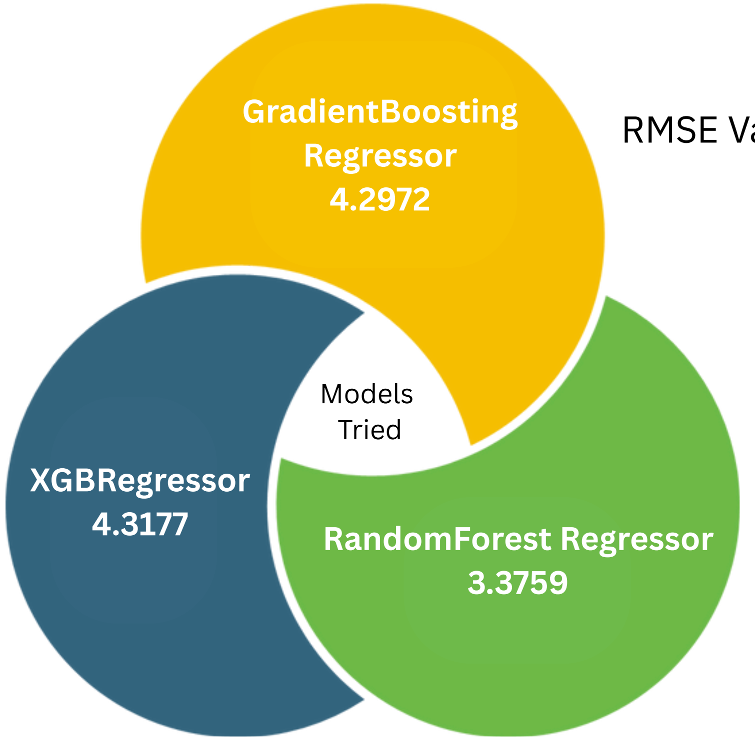
Spending Demand Logic

Spend to Limit Ratio	Credit Line Increase
> 1.2 and utilization > 60%	30%
> 1.0	20%
>0.8	10%
Otherwise	No increase

Step 2:Model Training and Evaluation

Features Used:

- Predicted Q4 spend
- Credit utilization (3M, 6M, current)
- Risk and delinquency signals
- Historical spend and fraud markers
- Available credit and current balance
- Credit line segmentation



RMSE Values of Models

✅ **Final Output**

The best model (RandomForestRegressor)was used to predicted credit line increase amount for all accounts in the eligible segments. The model was off by just **$3.3759** on an average.

This approach allowed us to recommend data-backed, risk-aware credit line increases tailored to each customer's financial profile.

Conclusion

Key Achievements

- Developed a comprehensive predictive framework addressing all three challenge objectives
- Achieved exceptional model performance:
  - Spending Prediction: R^2 of **0.9985** (GBRegressor)
  - Customer Segmentation: **88%** accuracy (XGBClassifier + SMOTE)
  - Credit Line Recommendations: MAE of **$3.38** (RandomForestRegressor)

Business Impact

- Enables data-driven decisions for credit line management
- Identifies high-value customers while mitigating risk exposure
- Provides personalized recommendations balancing growth opportunities with risk management

"Our solution provides Synchrony with a powerful, data-driven framework to optimize credit line decisions, maximize customer value, and minimize financial risk - all built on robust machine learning foundations."