

Credit Card Fraud Risk Analysis (2023–2024)

Capstone Project – William Matthews

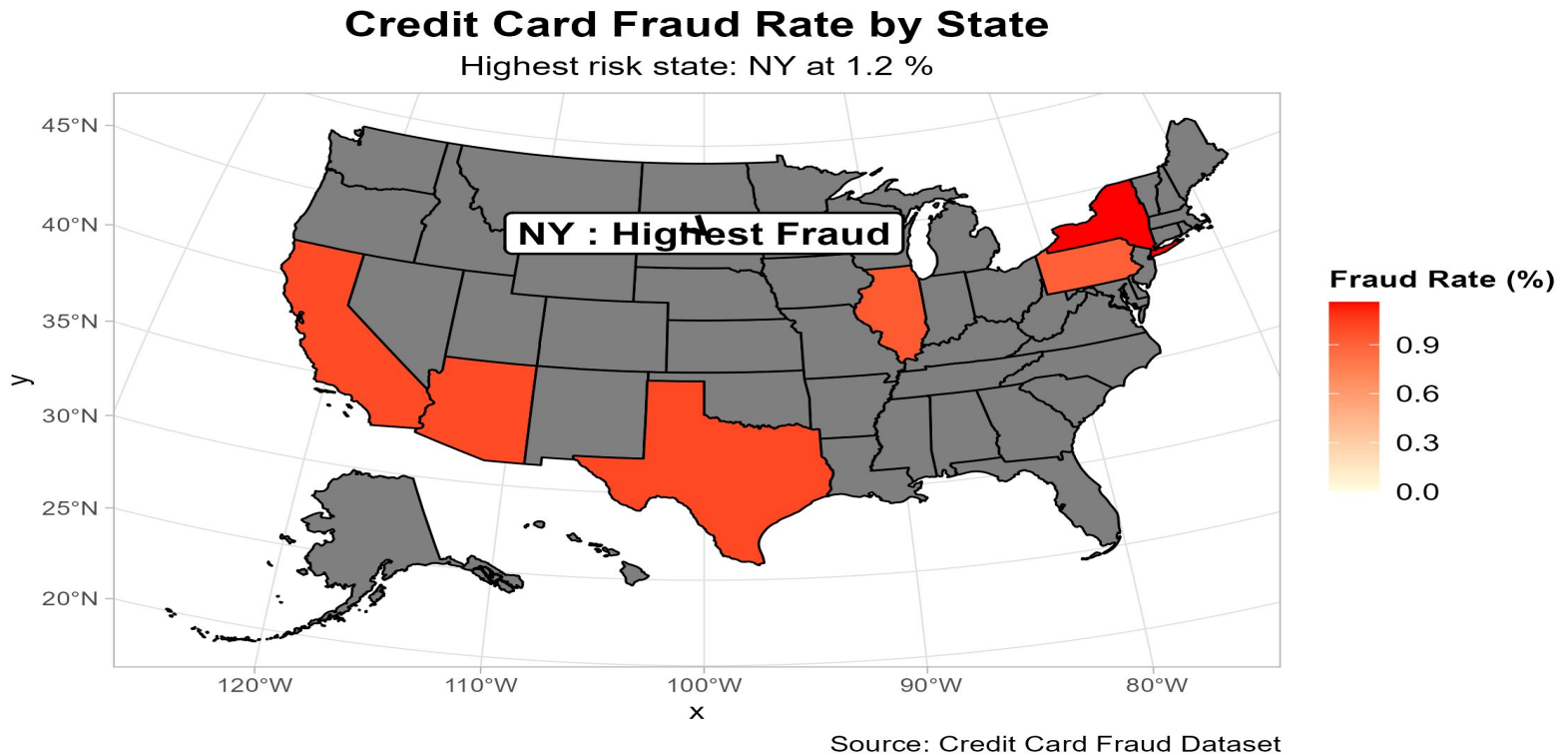
Data Sources & Quality Assessment

- Data Source:
 - Primary Dataset: Credit Card Fraud Dataset (2023-2024)
 - 100,000+ transactions from multiple states and merchants
 - Binary fraud classification (IsFraud: 0/1)
 - Include city-level geographic data, merchant info, transaction amounts
- Data Quality Metrics:
 - Geographic Mapping: 85.3% of transactions mapped to U.S. states
 - Completeness: 100% of fraud labels; 92% for merchant names
 - Time coverage: October 2023-October 2024 (12 months)
- Key Data Challenges:
 - Missing placeholder merchant names in ~8% of records
 - Manual city-to-state mapping required
 - Transaction concentration: 6 primary states

Data Processing Pipeline

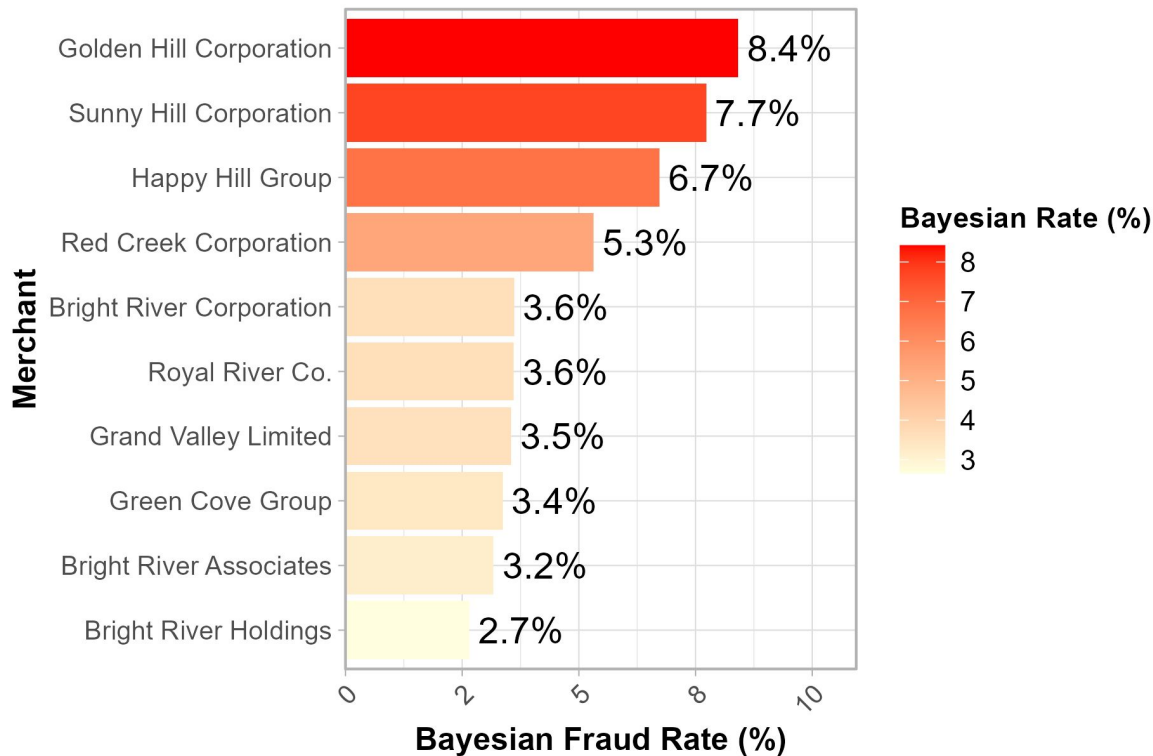
- Raw data import: CSV, city merchant and transaction details
- Data cleaning: Remove duplicates, handle missing values, standardized merchant names
- Geographic mapping: City-to-state lookup table for > 85% transactions
- Feature engineering: Create derived columns (IsFraud, AmountBand, TimeofDay, MerchantType, etc.)
- Validation: Consistency checks, completeness audits, quality logs
- Ready for analysis: Cleaned, mapped, and enriched dataset

Geographic Fraud Hotspots

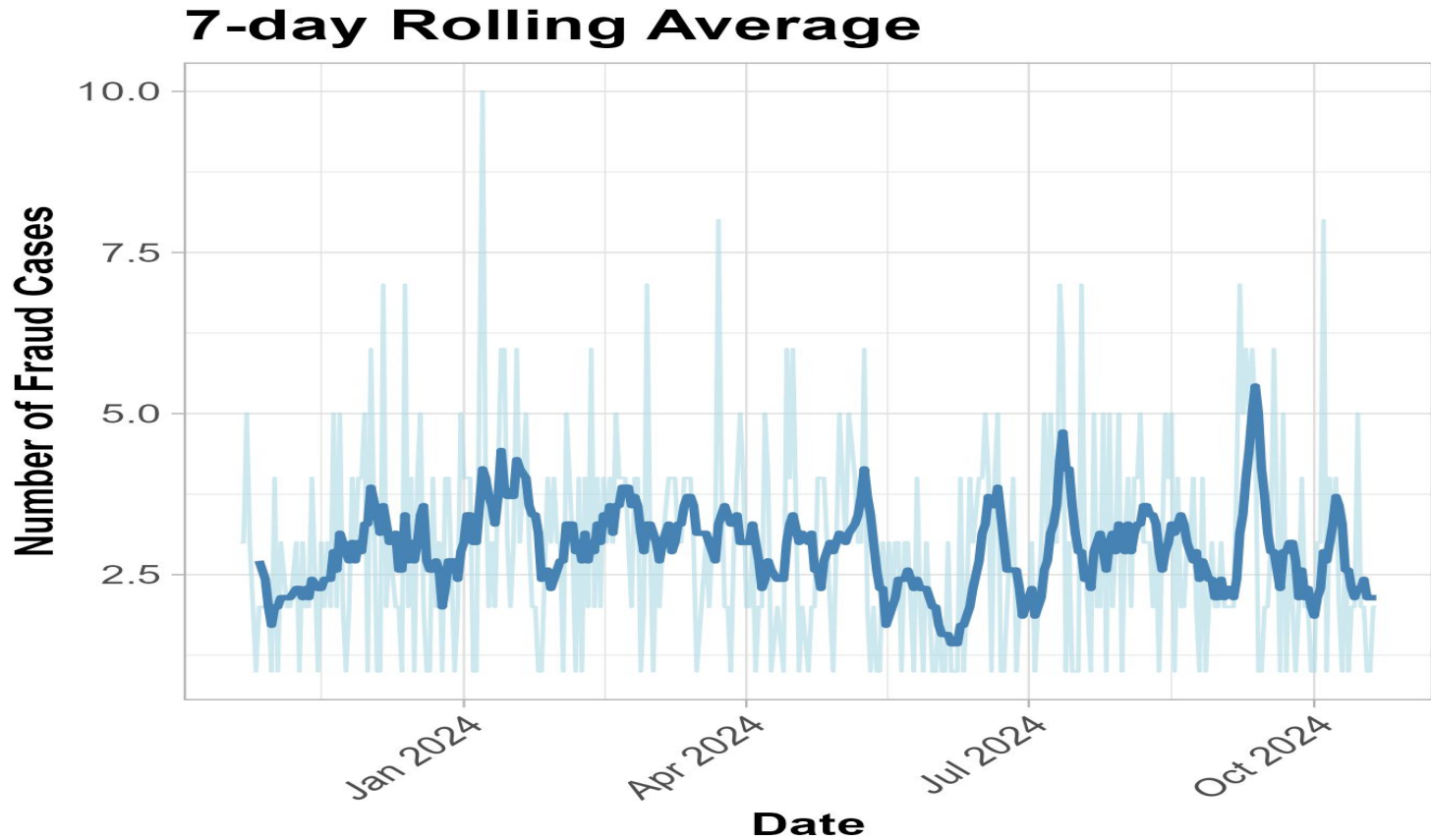


Merchant-Level Risk Analysis

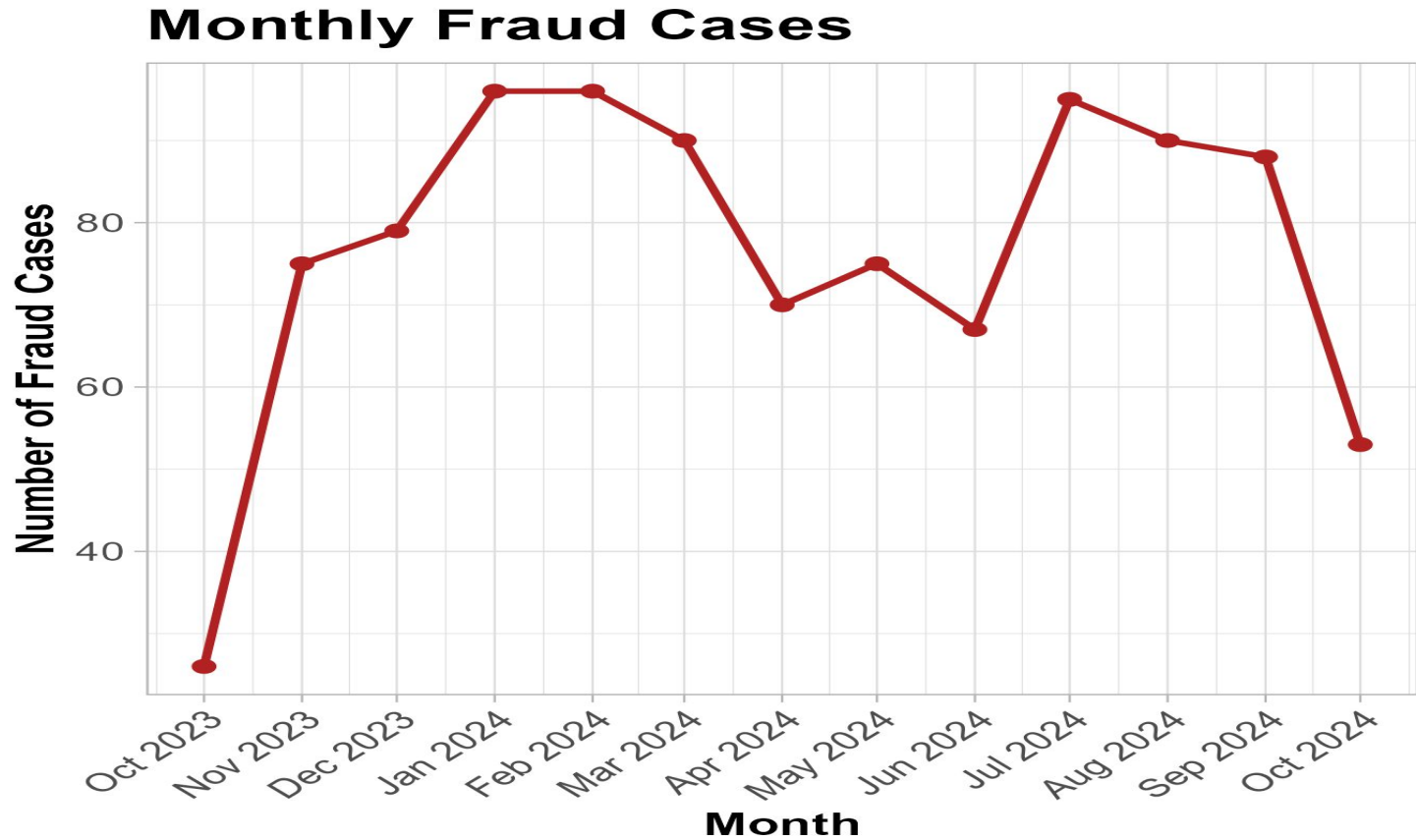
Top 10 Merchants by Bayesian Fraud Rate (%)



Daily Fraud Trends



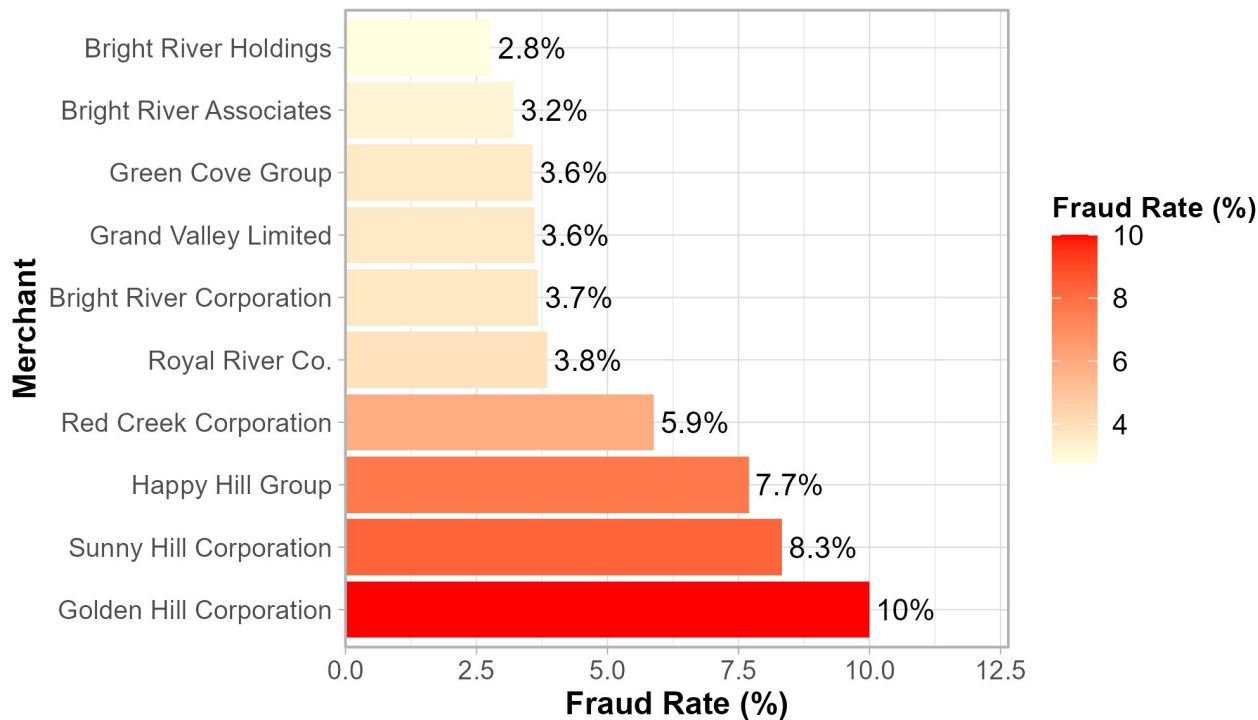
Monthly Fraud Trends



Observed Fraud Rate – Top Merchants

Top 10 Merchants by Fraud Rate in NY

Minimum 10 transactions required for inclusion



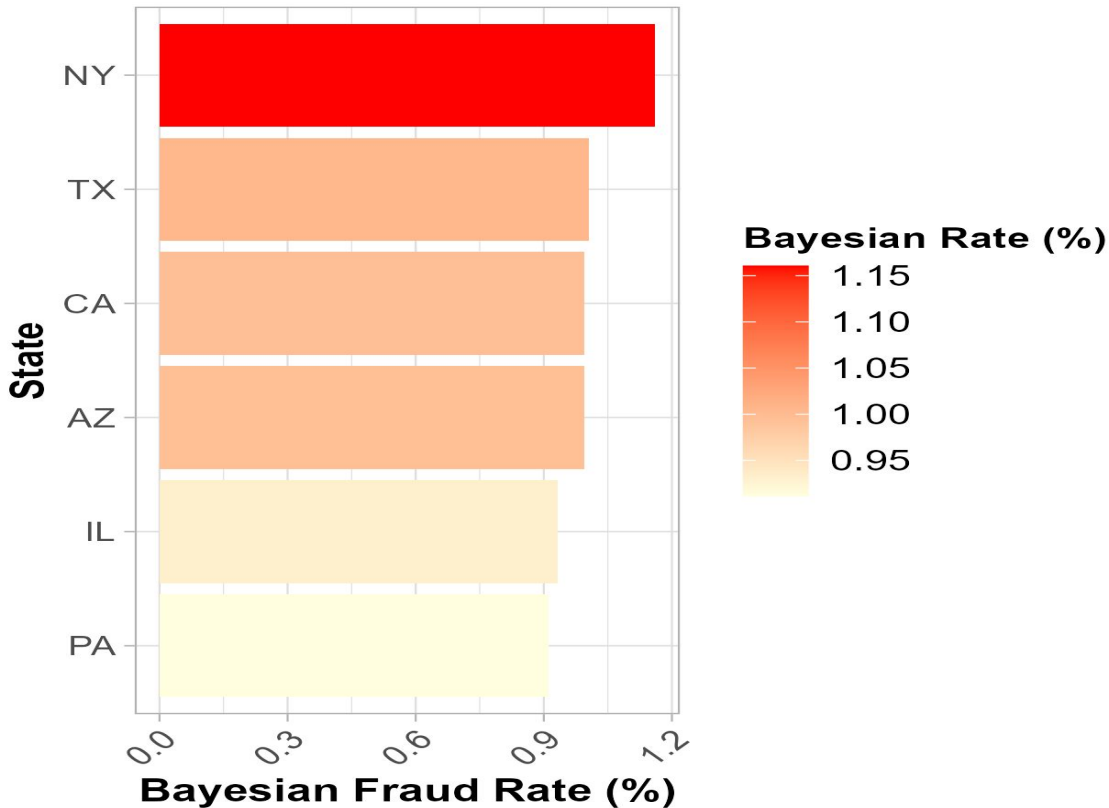
Source: Credit Card Fraud Dataset

Understanding Our Statistical Approach

- **Restaurant analogy:** 1 review vs. 1,000 reviews - which do you trust?
- **Same logic for fraud:** Small merchants shouldn't be unfairly penalized
- **Bayesian analysis:** Combines overall patterns with specific data
- **Result:** More reliable risk estimates for business decisions

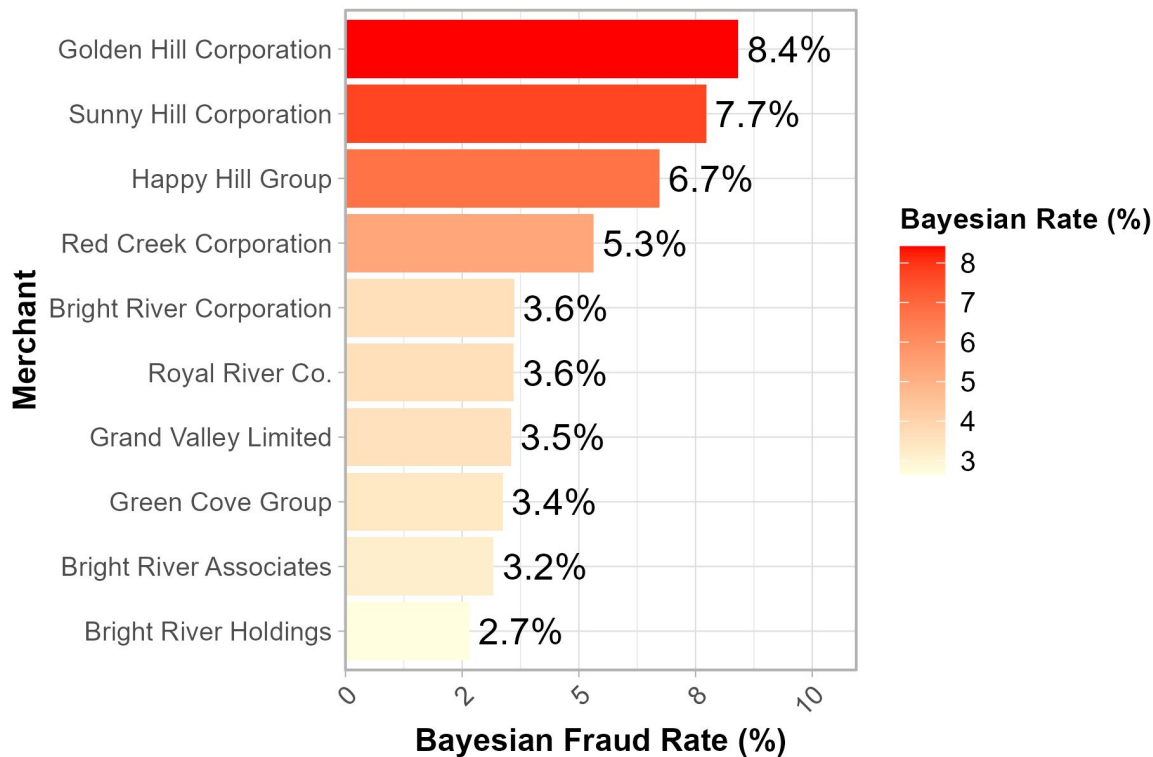
Bayesian Fraud Rate by State

Bayesian Fraud Rate by State (%)



Bayesian Fraud Rate – Top Merchants

Top 10 Merchants by Bayesian Fraud Rate (%)



Key Findings Summary

- **Geographic Risk Concentration:** New York emerges as the highest-risk state with 1.2% fraud rate
- **Merchant-Level Patterns:** Several merchants exhibit fraud rates several times higher than the average
- **Temporal Dynamics:** Monthly patterns show seasonal variation with peaks in early 2024
- **Statistical Validation:** Bayesian analysis confirms underlying risk patterns beyond raw observations
- **Actionable Intelligence:** Clear targets identified for intervention and risk mitigation

Strategic Recommendations

- **Immediate Action:** Enhanced monitoring and controls for New York region (highest priority)
- **Merchant-Specific Interventions:** Targeted fraud prevention for Golden Hill Corporation Limited and top-risk entities
- **Temporal Optimization:** Adjust monitoring intensity based on seasonal fraud patterns
- **Data-Driven Approach:** Implement Bayesian risk scoring for continuous assessment
- **Resource Allocation:** Focus 60% of fraud prevention resources on identified high-risk segments

Return on Investment & Success Metrics

- **Investment:** \$200K-400K total (systems, training, implementation)
- **Target ROI:** 25-40% fraud reduction within 12 months
- **Break-even:** Year 1 through prevented losses
- **Success metrics:** Fraud rates, detection speed, cost savings
- **Long-term value:** 200-300% ROI by Year 2-3