Credit Card Fraud Risk Analysis - Q&A Preparation Guide

William Matthews Capstone Project

Core Message Reminders

- **Primary Contribution**: Developed and validated a Bayesian fraud risk assessment framework
- **Data Period**: 2023-2024 analysis demonstrates methodology effectiveness
- **Implementation Approach**: Framework requires current data validation before deployment
- **Key Value**: Transferable analytical methodology, not time-specific findings

---

Anticipated Questions by Category

DATA CURRENCY & RELEVANCE

Q: "Why should we care about 2023-2024 fraud patterns if they might not exist anymore?" A: The value isn't in the specific patterns but in proving the methodology works. Just like testing a medical diagnostic tool on historical patients validates the tool's effectiveness, our 2023-2024 analysis proves this framework can detect fraud risk concentrations. Current implementation would identify today's patterns using the same proven approach.

Q: "How do you know your framework will work with current data?" A: The Bayesian methodology is designed to adapt to whatever fraud patterns exist in the data. The statistical approach doesn't assume specific merchants or states will always be high-risk - it identifies whatever risk concentrations currently exist. Our historical analysis proves the detection capability.

Q: "What if current fraud patterns are completely different?" A: That's exactly what the framework is designed to handle. Fraud patterns evolve constantly, which is why we need a systematic approach rather than static assumptions. The framework will identify current high-risk segments regardless of whether they match our 2023-2024 findings.

METHODOLOGY & STATISTICS

Q: "Can you explain the Bayesian approach in simple terms?" A: Think of restaurant reviews. A restaurant with 1 five-star review shouldn't be ranked above one with 500 four-star reviews. Similarly, a merchant with 2 transactions and 1 fraud case shouldn't be

labeled higher risk than one with 1000 transactions and 30 fraud cases. Bayesian analysis combines the overall fraud rate patterns with each merchant's specific data to give more reliable risk estimates.

Q: "What were your prior assumptions in the Bayesian model?" A: [Note: This wasn't specified in your slides - you may need to add this detail] We used the overall dataset fraud rate as our baseline prior and adjusted based on transaction volume for each merchant to prevent small-sample bias.

Q: "How statistically significant are your merchant risk rankings?" A: The Bayesian approach inherently accounts for statistical uncertainty. Merchants with fewer transactions have wider confidence intervals, which the framework considers when generating risk scores. This prevents overconfidence in rankings based on limited data.

BUSINESS IMPACT & ROI

Q: "How can you justify a $200-400K investment based on old data?" A: The investment isn't in implementing 2023-2024 findings - it's in deploying a validated analytical framework. The cost covers systems integration, staff training, and current data analysis to identify today's risk patterns. We're buying proven methodology, not outdated intelligence.

Q: "What evidence supports your ROI projections?" A: I've removed specific ROI predictions because they require current baseline establishment. The framework provides measurable fraud reduction, but exact percentages depend on current fraud rates and implementation effectiveness. ROI calculation begins from implementation date with current data validation.

Q: "Why not just use existing fraud detection systems?" A: Most systems focus on transaction-level detection. Our framework adds merchant and geographic risk stratification that can guide resource allocation and monitoring intensity. It complements rather than replaces existing detection systems.

TECHNICAL IMPLEMENTATION

Q: "How would this integrate with existing systems?" A: The Bayesian framework outputs risk scores that can feed into existing fraud monitoring systems. It provides merchant risk ratings and geographic risk factors that can adjust transaction monitoring thresholds and alert priorities.

Q: "What data requirements does ongoing implementation have?" A: The framework needs transaction data with merchant information, geographic indicators, and fraud labels. Quarterly updates require minimal additional data collection since it uses the same transaction streams that organizations already process.

Q: "How do you handle data quality issues like the missing merchant names?" A: The framework is designed to work with imperfect data. Missing merchant information gets handled through geographic and transaction pattern analysis. The 92% completeness rate for merchant names in our dataset was sufficient for meaningful pattern detection.

COMPETITIVE ANALYSIS

Q: "How does this compare to commercial fraud detection solutions?" A: Commercial solutions focus on transaction-level scoring. Our contribution is merchant-level and geographic risk stratification using Bayesian methods that account for statistical uncertainty. This provides a complementary layer for resource allocation and monitoring optimization.

Q: "What makes your approach unique?" A: The combination of Bayesian statistical methods with multi-dimensional risk analysis (geographic, merchant, temporal) in a framework designed for continuous updating. Most approaches either focus on single dimensions or don't account for statistical uncertainty in risk rankings.

LIMITATIONS & CHALLENGES

Q: "What are the biggest limitations of your approach?" A: The framework requires sufficient transaction volume for reliable statistical analysis, depends on data quality for geographic mapping, and needs regular updates to maintain relevance. It's most effective as part of a comprehensive fraud prevention strategy rather than a standalone solution.

Q: "How do you handle false positives?" A: The Bayesian approach reduces false positives by avoiding overreaction to small samples. However, the framework focuses on risk stratification rather than binary fraud/not-fraud decisions, so false positive rates depend on how organizations use the risk scores in their monitoring systems.

Key Statistics to Remember

- Dataset: 100,000+ transactions over 12 months
- Geographic coverage: 85.3% successfully mapped to US states
- Data completeness: 100% fraud labels, 92% merchant names
- Top risk state: NY at 1.2% fraud rate (in 2023-2024 data)
- Highest risk merchant: 8.4% fraud rate (Golden Hill Corporation in 2023-2024 data)

Closing Reinforcement Messages

- **Methodology Focus**: "The real value is the validated analytical framework, not the specific 2023-2024 patterns."
- **Implementation Readiness**: "This approach provides a systematic way to identify and respond to current fraud risk concentrations."
- **Continuous Improvement**: "The framework adapts to evolving fraud patterns through regular data updates and pattern recognition."