

# EXECUTIVE SUMMARY

## DATA QUALITY ISSUES & INCONSISTENCIES

### ■ CRITICAL DATA QUALITY FINDINGS FOR POWERPOINT

#### 1. MISSING DATA ISSUES

- **Missing Model Results: 2,213 records (4.8% of dataset)**
  - These applications have no FRAUD\_MODEL\_RESULT assigned
  - Represents a significant coverage gap in fraud screening
  - Impact: 4.8% of applications lack fraud model assessment

#### 2. DATA INCONSISTENCIES

- **Date Sequence Issues:** Applications with illogical date progressions
  - Offer shown before application submitted
  - Offer selected before offer shown
  - Origination before offer selected
- **Fraud Status vs Model Result Misalignment:** Cases where final status contradicts initial model decision

#### 3. MODEL COVERAGE GAPS

- **DIT Model Coverage:** Only runs when applications pass credit policy
- **Kount Model Coverage:** Only runs when applications pass credit policy
- **DNB Model Coverage:** Most comprehensive but still has gaps
- **Result:** Incomplete fraud screening for applications that fail early in process

#### 4. TEMPORAL DATA ANOMALIES

- **March 2025 Fraud Spike:** 2,724 confirmed fraud cases (vs. 51 in February)
  - Represents 97% of all confirmed fraud in dataset
  - Suggests massive fraud attack or delayed detection
  - Model failed to proactively identify this trend

#### 5. SUSPICIOUS VALUE PATTERNS

- **Empty Strings in Critical Fields:** Phone numbers, emails, IP addresses
- **Score Anomalies:** Behavior and device scores outside expected ranges

- **Flag Inconsistencies:** Model flags not aligned with actual data presence

6. FRAUD MODEL PERFORMANCE GAPS

- **Low Model Coverage:** Only 53.7% of data covered by complete fraud assessment
- **Missing Fraud Status:** Some applications lack final fraud determination
- **Incomplete Workflow:** Applications stuck in review states without resolution

■ QUANTIFIED IMPACT

| Issue Category        | Count    | % of Dataset | Business Impact             |
|-----------------------|----------|--------------|-----------------------------|
| Missing Model Results | 2,213    | 4.8%         | Unscreened applications     |
| Date Inconsistencies  | Variable | <1%          | Process integrity issues    |
| Coverage Gaps         | ~20,000  | ~43%         | Incomplete fraud assessment |
| March Fraud Spike     | 2,724    | 5.9%         | Major fraud event missed    |
| Suspicious Values     | Variable | <5%          | Data quality concerns       |

■ RECOMMENDATIONS FOR DATA QUALITY

IMMEDIATE ACTIONS (0-30 days)

1. **Investigate Missing Model Results:** Identify why 2,213 applications lack fraud assessment
2. **Review March 2025 Fraud Event:** Conduct forensic analysis of the fraud spike
3. **Implement Data Validation:** Add checks for date sequence logic

SHORT-TERM IMPROVEMENTS (1-3 months)

1. **Expand Model Coverage:** Ensure all applications receive fraud screening
2. **Fix Flag Inconsistencies:** Align model flags with actual data processing
3. **Implement Real-time Monitoring:** Add alerts for unusual fraud patterns

LONG-TERM STRATEGY (3-12 months)

1. **Enhanced Data Pipeline:** Implement end-to-end data quality monitoring
2. **Predictive Analytics:** Add early warning systems for fraud trends
3. **Model Optimization:** Improve coverage and reduce gaps in fraud detection

■ KEY TAKEAWAYS FOR LEADERSHIP

- **Dataset is usable** but has significant quality issues that impact fraud detection effectiveness
- **Critical gaps exist** in model coverage and data completeness
- **March fraud event** suggests reactive rather than proactive fraud detection

- **Immediate action required** to improve data quality and model coverage
- **ROI opportunity** exists to significantly improve fraud detection through data quality improvements

*Report Generated: August 02, 2025 at 05:12 PM*  
*Source: Fraud Analysis Master Notebook*