# Russia-Ukraine War Analytics: Data Profiling Analysis

## Data Overview

You have three interconnected datasets tracking Russian military losses from Ukraine's perspective:

1. **russia\_losses\_equipment.csv** - Daily cumulative equipment losses (tanks, aircraft, etc.)
2. **russia\_losses\_equipment\_correction.csv** - Corrections/adjustments to equipment data
3. **russia\_losses\_personnel.csv** - Daily cumulative personnel casualties

## Data Quality Issues & Discrepancies

### Critical Data Quality Problems:

1. **Inconsistent Date Formatting**
   * Equipment data: MM/DD/YY format
   * Personnel data: MM/DD/YY format
   * Correction data: Mixed MM/DD/YY format
   * Some dates appear to be in future (7/20/25 - likely 2024)
2. **Missing Data Patterns**
   * Equipment corrections show many zero values and negative adjustments
   * Personnel data has consistent "about" qualifier suggesting estimates
   * Some equipment categories have suspicious flat-line periods
3. **Data Integrity Issues**
   * Negative values in corrections data (e.g., -253, -578) indicating retroactive adjustments
   * Large jumps in cumulative values without clear explanation
   * "Greatest losses direction" field inconsistently populated
4. **Structural Inconsistencies**
   * Different column counts across datasets
   * Personnel\* vs personnel columns suggest data versioning issues
   * POW column in personnel data mostly empty

## Business Questions for Client Insights

### Strategic Assessment Questions:

1. **What is the trend of Russian equipment attrition over time?**
   * Track monthly/quarterly loss rates by equipment type
   * Identify acceleration or deceleration patterns
2. **Which equipment categories show highest loss rates?**
   * Compare relative losses: tanks vs aircraft vs artillery
   * Calculate daily/weekly burn rates for each category
3. **How reliable are the casualty estimates?**
   * Analyze correction patterns and frequency
   * Quantify uncertainty ranges in personnel estimates

### Operational Intelligence Questions:

1. **What geographic patterns emerge from "greatest losses direction"?**
   * Map loss concentrations by reported direction
   * Correlate with known battlefield locations
2. **Are there seasonal or temporal patterns in losses?**
   * Weekly/monthly cyclical analysis
   * Correlation with known military operations
3. **What is the equipment-to-personnel loss ratio?**
   * Calculate ratios for different time periods
   * Compare to historical military conflict data

### Data Quality & Methodology Questions:

1. **How frequently are corrections made and why?**
   * Analyze correction patterns and timing
   * Identify most frequently corrected equipment types
2. **What is the confidence level of different data points?**
   * Distinguish between confirmed vs estimated losses
   * Track data source reliability over time