# Deriving Insights from Scraped Contract Data Using AI

**Date Prepared:** March 04, 2025
**Purpose:** This document demonstrates how artificial intelligence can derive actionable insights from a CSV dataset of government contract opportunities scraped from sam.gov. The analysis leverages key metrics, rows, and columns to identify patterns, trends, and recommendations.

## Dataset Overview

The CSV contains **49 rows** representing unique contracts, with one row flagged as a failed extraction. The data was scraped on **March 04, 2025**, between 18:46:41 and 19:10:18. Below are the **15 columns** analyzed:

1. **Contract Name**: Title of the contract (e.g., "AVEVA SCADA INTEGRATION").
2. **Notice ID**: Unique identifier (e.g., "W912NW009610").
3. **Department**: Contracting agency (e.g., "DEPT OF DEFENSE").
4. **Contract Link**: URL to the contract page on sam.gov.
5. **Failed Row**: Boolean indicating scrape failure (True/False).
6. **Incomplete Data**: Boolean indicating missing data (True/False).
7. **Total Attachments**: Number of attached files (0-20).
8. **Date Scraped**: Timestamp of data collection (e.g., "2025-03-04 18:46:41").
9. **Contract Number**: Sequential identifier (1-49).
10. **General Published Date**: Latest publication date (e.g., "Mar 04, 2025 03:01 pm CST").
11. **Original Published Date**: Initial publication date (e.g., "Feb 25, 2025 11:44 am CST").
12. **Updated Date Offers Due**: Current deadline (e.g., "Mar 12, 2025 03:00 pm CDT").
13. **Original Date Offers Due**: Initial deadline (e.g., "Mar 12, 2025 03:00 pm CDT").
14. **File Name**: Name of attached files (e.g., "W912NW-25-Q-0046 Questions and Answers.pdf").
15. **File Link**: URL to download attachments.

## Metrics Derived:

- Total Contracts: 49
- Failed Contracts: 1 (Row 49, Contract Number 49)
- Contracts with Missing Data: 1 (same as above)
- Total Attachments: 168 (average ~3.43 per contract)

**Insight**: The dataset is robust, with only one failed row (Contract Number 49, "Error extracting"), indicating a need for improved scraping reliability.

- pattern recognition (e.g., identifying tech trends in contract names).

**1. Department Distribution**

**Columns Analyzed:**

- Department (Column 3)
- Contract Number (Column 9, to count unique contracts)

**Methodology:**

- Grouped rows by Department and counted unique Contract Number values.

**Results:**

- **DEPT OF DEFENSE**: 29 contracts (59.2%)
- **FEDERAL DEPOSIT INSURANCE CORPORATION**: 8 contracts (16.3%)
- **VETERANS AFFAIRS, DEPARTMENT OF**: 3 contracts (6.1%)
- **INTERIOR, DEPARTMENT OF THE**: 3 contracts (6.1%)
- **NATIONAL AERONAUTICS AND SPACE ADMINISTRATION**: 2 contracts (4.1%)
- **SENATE, THE**: 2 contracts (4.1%)
- **UNITED STATES GOVERNMENT PUBLISHING OFFICE**: 1 contract (2.0%)
- **STATE, DEPARTMENT OF**: 1 contract (2.0%)
- **ENERGY, DEPARTMENT OF**: 1 contract (2.0%)
- **TREASURY, DEPARTMENT OF THE**: 1 contract (2.0%)
- **COURT SERVICES AND OFFENDER SUPERVISION AGENCY**: 1 contract (2.0%)

**Insight**: The Department of Defense dominates with 59.2% of contracts (29/49), highlighting a focus on defense-related opportunities. Businesses should prioritize this department for maximum exposure.

**2. Urgency Analysis**

**Columns Analyzed:**

- Updated Date Offers Due (Column 12)
- Contract Name (Column 1)
- Notice ID (Column 2)
- Reference date: March 04, 2025 (from Date Scraped, Column 8)

**Methodology:**

- Compared Updated Date Offers Due to March 04, 2025, to categorize contracts by urgency.

**Results:**

- **Imminent (Due by March 06, 2025)**:
    o Row 5: "7610--VISN 09 wide Drug Information Electronic Database" (36C24925Q0135): Mar 06, 2025, 10:00 am CST (2 days left)
    o Row 4: "SIMULIA Abaqus Renewal FY25" (80TECH25Q0051): Mar 06, 2025, 04:00 pm CST (2 days left)
    o Row 7: "Network Detection and Response (NDR) Tool" (W911SD25QA0030): Mar 06, 2025, 12:00 pm EST (2 days left)
    o Row 10: "FY25 Invenio" (80NSSC25894148Q): Mar 06, 2025, 01:30 pm CST (2 days left)
- **Near-Term (Due by March 11, 2025)**:
    o Row 25: "Brand Name Justification & Bluebeam Software Solicitation" (FA469025Q0010): Mar 04, 2025, 03:00 pm MST (same day/overdue)
    o Row 18: "549 CTS SVTC" (F3G5DA4270A001): Mar 04, 2025, 10:00 am PST (same day/overdue)
    o Row 21: "Call 001: Data Models and Flows..." (FA930021SC001Call001FY25): Mar 10, 2025, 11:00 am PDT (6 days)
    o Row 15: "D--SOFTWARE SUPPORT-earthquake monitoring..." (140G0325Q0019): Mar 10, 2025, 10:00 am PDT (6 days)
    o Row 33: "NetScout Hardware Refresh" (2025-R-032): Mar 10, 2025, 12:00 pm EDT (6 days)
- **Longer-Term (Due after March 11, 2025)**:
    o Row 14: "FDIC's Digicert Subscription Maintenance Requirement" (CORHQ-25-Q-0120): Apr 07, 2025
    o Row 24: "Trellix FireEye (Goods and Support)" (CORHQ-25-Q-0076): Apr 02, 2025

**Insight**: Four contracts are due within 2 days, requiring immediate action. Overdue contracts (e.g., Bluebeam) suggest either scraping latency or expired opportunities—verify via Contract Link.

**3. Complexity Analysis**

**Columns Analyzed:**

- Total Attachments (Column 7)
- Contract Name (Column 1)
- Notice ID (Column 2)

**Methodology:**

- Categorized contracts by Total Attachments to assess documentation complexity.

**Results:**

- **High Complexity (10+ attachments)**:
  - Row 20: "Next Generation (NextGen) Passport Personalization Printers Support" (19AQMM24R0113): 20 attachments
  - Row 46: "MCSC PDSS Support Services" (M6785424R7603): 14 attachments
  - Row 15: "D--SOFTWARE SUPPORT-earthquake monitoring..." (140G0325Q0019): 12 attachments
  - Row 43: "Financial Management Systems Maintenance and Support" (2031JW25Q00002): 10 attachments
- **Moderate Complexity (5-9 attachments)**:
  - Row 18: "549 CTS SVTC" (F3G5DA4270A001): 6 attachments
  - Row 19: "Solicitation for the Defense Commissary Agency..." (HC102825R0012): 5 attachments
- **Low Complexity (0-4 attachments)**:
  - Most contracts, e.g., Row 4: "SIMULIA Abaqus Renewal FY25" (0 attachments)

**Insight**: High attachment counts (e.g., 20 for NextGen Passport) suggest complex requirements or amendments, potentially indicating higher value or scrutiny.

**4. Deadline Shifts**

**Columns Analyzed:**

- Original Date Offers Due (Column 13)
- Updated Date Offers Due (Column 12)
- Contract Name (Column 1)

**Methodology:**

- Compared Original Date Offers Due with Updated Date Offers Due to identify delays.

**Results:**

- **Significant Extensions**:
  - Row 20: "Next Generation (NextGen) Passport...": Jul 02, 2024 → Mar 17, 2025 (~8.5 months)
  - Row 15: "D--SOFTWARE SUPPORT-earthquake monitoring...": Mar 04, 2025 → Mar 10, 2025 (6 days)
  - Row 32: "Enterprise Test Data Management": Jan 31, 2025 → Feb 21, 2025 (21 days)
- **No Change or Minor Adjustments**:
  - Row 1: "AVEVA SCADA INTEGRATION": Mar 12, 2025 (unchanged)

**Insight**: Significant delays (e.g., NextGen Passport) may reflect complexity or procurement challenges, offering opportunities to monitor for updates.

**5. Technology and Service Trends**

**Columns Analyzed:**

- Contract Name (Column 1)
- File Name (Column 14)

**Methodology:**

- Extracted keywords and themes using AI pattern recognition.

**Results:**

- **Software and IT Services**:
    - Row 1: "AVEVA SCADA INTEGRATION" (SCADA systems)
    - Row 3: "Air Force Test and Analysis Tool..." (STARLIMS, HTML5)
    - Row 4: "SIMULIA Abaqus Renewal FY25" (engineering software)
    - Row 11: "Microsoft Unified Support" (enterprise IT)
    - Row 35: "SailPoint Identity IQ Renewal" (identity management)
- **Cybersecurity**:
    - Row 2: "D--JOFOC CyberSafe TrustBroker"
    - Row 7: "Network Detection and Response (NDR) Tool"
    - Row 17: "Fortify Security testing software"
- **Hardware/Infrastructure**:
    - Row 8: "Install Eighty-Two (82) NIPR drops..."
    - Row 12: "CPU" (SPMYM225Q1010)
    - Row 33: "NetScout Hardware Refresh"

**Insight**: Strong demand for IT, software renewals, and cybersecurity aligns with government priorities, suggesting lucrative opportunities in these areas.

**Actionable Recommendations**

1. **Prioritize Urgent Contracts**: Target the four contracts due by March 06, 2025 (Rows 5, 4, 7, 10), especially "7610--VISN 09 wide Drug Information..." (low complexity, 3 attachments).
2. **Focus on DEPT OF DEFENSE**: With 29 contracts (59%), prioritize defense-related IT and infrastructure (e.g., Rows 1, 3, 7).
3. **Monitor High-Complexity Contracts**: "NextGen Passport" (Row 20) and "MCSC PDSS" (Row 46) may yield higher rewards.
4. **Track Delays**: Investigate contracts with extensions (e.g., Row 20) via Contract Link for updates.
5. **Leverage Tech Trends**: Build capabilities in software, cybersecurity, and IT infrastructure to match demand.