

Federal Spend Intelligence

USAspending.gov Data Pipeline | Portfolio Demo

TL;DR

Built an **enterprise data pipeline** that ingests federal procurement data from **USAspending.gov** (official government source). Pipeline processes **24.8K contract awards** spanning **FY2004-2027** (multi-year contracts) totaling **\$67.14B**, identifies **7.5K vendors**, calculates **concentration risk metrics**, and produces **audit-ready analytics** with a **97% data quality score**.

\$67.14B

TOTAL SPEND

24.8K

AWARDS

7.5K

VENDORS

97%

QUALITY SCORE

What This Demonstrates

Data Engineering

- Public API integration
- Incremental data loading
- Star schema modeling
- Data quality gates

Analytics Engineering

- HHI concentration analysis
- Spend trend detection
- Vendor risk scoring
- Automated PDF reporting

Key Insight: Vendor Concentration

HHI Index: 75 — Unconcentrated (healthy competition)

The Herfindahl-Hirschman Index measures market concentration. This analysis helps procurement teams identify single-vendor dependency risks and supports vendor diversification strategies.

Quality Gates (All Passing)

✓ **Schema Drift:** 100% ✓ **Freshness:** 95% ✓ **Completeness:** 100% ✓ **Duplicates:** 100% ✓
Value Sanity: 85% ✓ **Referential Integrity:** 100%

Technical Highlights

- **Source:** USAspending.gov API (official, public, verifiable)
 - **Scope:** DoD, HHS, DHS contracts for FY2024
 - **Pipeline:** Python → Star Schema → Quality Gates → KPIs → PDF
 - **Duration:** 156 seconds
-

Report prepared by Mboya Jeffers | MboyaJeffers9@gmail.com

Data Source: USAspending.gov | Generated: 2026-02-01