

# SEC Financial Intelligence

Public Company XBRL Data Pipeline | Portfolio Demo

## TL;DR

Built an **enterprise data pipeline** that extracts standardized financial metrics from **39 Fortune 500 companies** via SEC EDGAR. Pipeline processes **1.08M XBRL facts**, maps them to canonical metrics (Revenue, Net Income, etc.), and produces **investment-grade analytics** with **97% data quality score**.

**1.08M**

XBRL FACTS

**39**

COMPANIES

**4,498**

CONCEPTS

**97%**

QUALITY SCORE

## What This Demonstrates

### Data Engineering

- SEC EDGAR API integration
- XBRL taxonomy parsing
- 1M+ row data processing
- Star schema modeling

### Analytics Engineering

- Financial metric extraction
- Cohort benchmarking
- Data quality gates
- Automated reporting

## Company Coverage

## Sectors Covered:

- Technology: 10 companies
- Consumer: 7 companies
- Financial: 6 companies
- Healthcare: 4 companies
- Communications: 3 companies
- Industrial: 2 companies
- Energy: 2 companies
- Real Estate: 2 companies
- Utilities: 2 companies
- Materials: 1 companies

## Top 5 by Revenue:

- AAPL: \$383.3B
- WMT: \$327.7B
- AMZN: \$291.3B
- MRK: \$179.7B
- XOM: \$176.1B

## Quality Gates (All Passing)

---

✓ **Unit Consistency:** 98%   ✓ **Period Logic:** 100%   ✓ **Coverage:** 99%   ✓ **Restatement:** 83%   ✓  
**Outliers:** 99%

## Technical Highlights

---

- **Source:** SEC EDGAR data.sec.gov (official, public, verifiable)
- **Pipeline:** Python (requests, pandas) → Star Schema → Quality Gates → KPIs → PDF
- **Duration:** 56 seconds for full cohort
- **Rate Limiting:** SEC-compliant (10 req/sec with User-Agent header)

---

Report prepared by Mboya Jeffers | MboyaJeffers9@gmail.com

Data Source: SEC EDGAR (data.sec.gov) | Generated: 2026-02-01