P2-SEC: Methodology & Quality Assurance

1. Data Source

SEC EDGAR - Official repository for SEC filings and structured XBRL data.

```
GET /submissions/CIK{cik}.json - Company filing history
GET /api/xbrl/companyfacts/CIK{cik}.json - XBRL financial facts
```

Compliance: User-Agent header required. Rate limit: 10 req/sec.

2. Data Model

Table	Description	Key Columns
company_dim	Company master data	cik, ticker, name, sector
filings_dim	SEC filing history	accession, form, filing_date
concept_map	XBRL to canonical mapping	raw_concept, canonical_metric
xbrl_facts	Financial fact values	cik, concept, value, period_end

3. Canonical Metric Mapping

Canonical	XBRL Concepts	Category
revenue	Revenues, RevenueFromContractWithCustomer...	Income Statement
net_income	NetIncomeLoss, ProfitLoss	Income Statement
assets	Assets	Balance Sheet
equity	StockholdersEquity, ...	Balance Sheet
operating_cash_flow	NetCashProvidedByUsedInOperatingActivities	Cash Flow

4. Quality Gates

Gate	Threshold	Logic
Unit Consistency	≥90%	Each concept uses single unit (USD/shares)
Period Logic	≥95%	Valid period types (instant/quarterly/annual)
Coverage	≥80%	Required metrics present per company
Restatement	Info	Detect same metric+period with different values
Outliers	≥85%	Z-score outliers ($|z| > 5$) below threshold

5. KPI Formulas

```
Net Margin = Net Income / Revenue
Gross Margin = Gross Profit / Revenue
```

```
ROA = Net Income / Total Assets
ROE = Net Income / Stockholders Equity
```

6. Run Metrics

Metric	Value
Duration	55.8 seconds
API Calls	80
Raw Facts	1,082,675
Cleaned Facts	1,082,675
Companies	39
Quality Score	96.6%