Python Use Cases

By Dhandapani Yedappalli Krishnamurthi

Sep 3, 2025 2:00 AM GMT-12

1. Employee Data Cleaning & Consolidation

Problem: HR teams receive monthly Excel sheets from different branches with inconsistent column names (`empID` vs `id`, `salary` vs `sal`).
ETL Task:

- **Extract:** Read multiple Excel/CSV files.
- **Transform:** Standardize column names, clean missing data, validate salary as numeric.
- **Load:** Save into a single clean master employee CSV/DB table.
  Python Features: file I/O, loops, `pandas`, functions, exception handling.

---

2. Sales Data Currency Normalization

Problem: An e-commerce site logs sales in multiple currencies (USD, INR, EUR). Management needs everything in INR.
ETL Task:

- **Extract:** Pull daily sales CSVs.
- **Transform:** Call currency API to convert values, handle rounding.
- **Load:** Append into PostgreSQL sales fact table.
  Python Features: API calls (`requests`), dictionaries, loops, `pandas`, error handling.

---

3. Web Server Log Analysis

Problem: A company wants to analyze failed login attempts from Apache logs.
ETL Task:

- **Extract:** Parse `.log` files line by line.
- **Transform:** Filter out `401`/`403` status codes, count attempts by IP.
- **Load:** Save suspicious IPs into a security report CSV.
  Python Features: regex, file handling, loops, string parsing, writing CSV.

---

4. IoT Sensor Data Aggregation

Problem: Sensors generate JSON files every hour with temperature, humidity, pressure. Management needs daily averages.
ETL Task:

- **Extract:** Read JSON files.
- **Transform:** Aggregate per day, handle missing sensor readings.
- **Load:** Store into a time-series database or CSV for reporting.
  Python Features: JSON parsing, loops, `pandas`, datetime handling.

---

5. Customer Feedback Sentiment Pipeline

Problem: Daily customer reviews are collected in text files. The team wants sentiment scores.
ETL Task:

- **Extract:** Read raw `.txt` feedback files.
- **Transform:** Apply NLP sentiment scoring (basic libraries like `textblob` or `nltk`).
- **Load:** Store structured feedback with sentiment (`positive/negative/neutral`) into a CSV.
  Python Features: text processing, external library use, exception handling.

---

6. Retail Store Inventory Sync

Problem: Inventory comes from multiple stores (CSV, XML, API). Need a unified product catalog.
ETL Task:

- **Extract:** Read CSV, XML, and API sources.
- **Transform:** Normalize column mappings (e.g., `item_code` vs `sku`), remove duplicates.
- **Load:** Save unified dataset into SQLite.
  Python Features: XML parsing, JSON parsing, `sqlite3`, pandas merges.

---

7. Healthcare Patient Records Validation

Problem: Patient data has missing DOBs, duplicate IDs, and incorrect phone/email formats.
ETL Task:

- **Extract:** Read CSV/Excel records.
- **Transform:** Validate formats with regex, remove duplicates, flag invalid rows.
- **Load:** Store valid + invalid records separately for review.
  Python Features: regex, conditionals, error handling, writing multiple files.

---

8. Movie Metadata Normalization

Problem: Multiple sources (IMDb, Rotten Tomatoes, TMDB) provide movie data with different formats.
ETL Task:

- **Extract:** Fetch APIs + CSV dumps.
- **Transform:** Standardize field names (`title`, `release_date`, `rating`), convert rating scales to 1–10.
- **Load:** Insert into PostgreSQL/SQLite movie database.
  Python Features: APIs, dictionaries, list comprehensions, pandas.

---

9. Financial Fraud Detection Prep

Problem: Bank transactions dataset has outliers (sudden huge withdrawals). Analysts need flagged data.
ETL Task:

- **Extract:** Read transaction files.
- **Transform:** Apply threshold checks & z-score outlier detection.
- **Load:** Store flagged suspicious records separately.
  Python Features: math functions, conditionals, pandas, exception handling.

---

10. Weather Forecast ETL for Travel Agency

Problem: A travel agency needs 7-day weather forecasts for multiple cities.
ETL Task:

- **Extract:** Pull weather API data (JSON).
- **Transform:** Keep only temp, humidity, rain probability, restructure into tabular format.
- **Load:** Save into SQLite / CSV for dashboarding.
  Python Features: API calls, loops, JSON handling, pandas, scheduling (with `cron` or `schedule` lib).

Share & follow

For any inquiries, please contact:

♟ Person

Email

Website