

# Ketan Shukla

---

## Python ETL Developer

San Diego, CA | [resume@ketankshukla.com](mailto:resume@ketankshukla.com) | 619-669-8545 | [LinkedIn](#) | [GitHub](#)

---

## Professional Summary

Python ETL Developer focused on data processing, automation, and API development. Some experience in building ETL pipelines, designing database integrations, and implementing data validation systems through a portfolio of 2 practical projects. Some experience with Python libraries for data manipulation, web scraping, and database operations with a strong commitment to data accuracy and quality. Seeking an entry-level ETL Developer position to leverage technical expertise and passion for data engineering in solving complex business problems.

## Technical Skills

### Data Engineering & ETL

- ▣ Data Processing: Pandas, NumPy, Data Transformation, Data Cleaning
- ▣ Data Validation: Error Handling, Data Quality Checks, Schema Validation

### Data Integration

- ▣ Web Scraping: Requests
- ▣ API Integration: RESTful APIs, API Authentication
- ▣ Document Processing: PyPDF2, JSON/XML Parsing, regex

### Database Technologies

- ▣ SQL: SQLite

### Development & Tools

- ▣ Version Control: Git, GitHub
- ▣ Development Environment: Jupyter Notebook, VS Code
- ▣ Testing: pytest, unittest

# Portfolio Projects

## Log Analysis & Monitoring System

- ✦ Developed a Python-based log analysis system that processes server logs, extracts performance metrics, and identifies potential security threats
- ✦ Implemented regex pattern matching to extract structured data from unstructured logs with 97% accuracy
- ✦ Created an anomaly detection algorithm using statistical methods to identify unusual patterns in server response times
- ✦ Built a notification system using SMTP for alerting on critical issues and performance degradations
- ✦ Designed a data retention policy with automatic archiving of processed logs to optimize storage usage
- ✦ Tech Stack: Python, regex, Pandas, SQLite, smtplib
- ✦ GitHub: [github.com/ketankshukla/log\\_analysis\\_system](https://github.com/ketankshukla/log_analysis_system)

## Data Warehouse ETL Framework

- Engineered a modular ETL framework for transferring data from multiple source systems to a central data warehouse
- Implemented configurable extractors for various data sources including CSV, JSON, XML, and SQL databases
- Developed transformation pipelines with comprehensive data cleaning, normalization, and validation steps
- Created a metadata-driven approach for dynamically generating table schemas and tracking data lineage
- Built a robust error handling system with transaction support to ensure data integrity during loading
- Tech Stack: Python, Pandas, SQLAlchemy, PyYAML, psycopg2
- GitHub: [github.com/ketankshukla/data-warehouse-etl](https://github.com/ketankshukla/data-warehouse-etl)

# Education

## University of California San Diego (UCSD)

- Certificate in Python Development (Completed)
- Certificate in Data Science (Ongoing)
- Core Focus: Python for Data Science, Database Management, Data Visualization