**KETAN SHUKLA**

**Python ETL Developer**

San Diego, CA | resume@ketankshukla.com | 619-669-8545  
[LinkedIn](https://www.linkedin.com/in/ketankshukla) | [GitHub](https://github.com/ketankshukla)

**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

**PROJECTS**

**Financial Market ETL Pipeline**

* Designed and implemented a comprehensive ETL pipeline for processing financial market data from multiple sources (CSV, JSON, REST APIs)
* Engineered data transformation components that calculate advanced financial metrics including RSI, MACD, and Bollinger Bands with 99% accuracy
* Created a flexible orchestration system using a task-based architecture with dependency management for reliable pipeline execution
* Developed a robust validation framework to ensure data consistency and completeness across all processing stages
* Implemented both database and CSV export capabilities with configurable retention policies for optimized storage
* Built a command-line interface with comprehensive logging for monitoring pipeline execution and troubleshooting
* **Tech Stack**: Python, Pandas, NumPy, SQLAlchemy, Requests, BeautifulSoup4
* **GitHub**: [github.com/ketankshukla/financial\_market\_etl](https://github.com/ketankshukla/data-warehouse-etl)

**COVID-19 Data Integration ETL Pipeline**

* Engineered a comprehensive Python ETL pipeline that extracts COVID-19 data from multiple sources including CSV files, JSON data, REST APIs, and web scraping
* Implemented data transformation modules with standardization for dates, locations, and missing values, ensuring 100% data consistency across disparate sources
* Designed a flexible orchestration system with task scheduling and dependency management for reliable pipeline execution
* Created robust data validation checks using Great Expectations to ensure data quality and integrity throughout the pipeline
* Developed a unified data loading system that writes to SQLite database with configurable export options to CSV
* Built a mock API server for local testing, enabling development without relying on external services
* Implemented detailed logging and error handling to facilitate troubleshooting and pipeline monitoring
* **Tech Stack**: Python, Pandas, SQLAlchemy, NumPy, Requests, BeautifulSoup4, Great Expectations
* **GitHub**: [github.com/ketankshukla/covid19\_etl](https://github.com/ketankshukla/covid19_etl)

**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)

**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)

**E-commerce Sales ETL Pipeline**

* Developed a comprehensive data pipeline for extracting, transforming, and loading e-commerce sales data from multiple platforms and formats
* Implemented versatile extractors supporting diverse data sources including CSV, JSON, Excel, PDF, SQL databases, XML, FTP/SFTP, and email content
* Created transformation modules to calculate key business metrics including sales trends, customer lifetime value, and inventory turnover
* Designed a flexible orchestration system with task scheduling and dependency management for reliable pipeline execution
* Built robust data validation components to ensure consistency and completeness across all processing stages
* Developed configurable data loaders for database integration and multi-format exports
* Implemented automated report generation capabilities for business intelligence and analysis
* Tech Stack: Python, Pandas, SQLAlchemy, lxml, PyPDF, smtplib, paramiko
* GitHub: [github.com/ketankshukla/ecommerce\_etl](https://github.com/ketankshukla/ecommerce_etl)

**EDUCATION**

**San Diego City College**

* Certificate in Python Development (Completed)
* Certificate in Data Science (Expected June 2025)
* Core Focus: Python for Data Science, Database Management, Data Visualization