Praveen S

Data Engineer - Web Scraping

✉ praveensureshbabu03@gmail.com   📞 +91 9962320073   🌐   *GitHub/Praveen S* | *LinkedIn/Praveen S*

Data Engineer / Web Scraping with 3+ years of experience building automated data extraction and ETL pipelines using Python, SQL, Selenium, BeautifulSoup, and Streamlit. Proven expertise in large-scale web scraping, API integration, data cleaning, and transforming raw data into analytics-ready datasets. Experienced with PostgreSQL and SQL Server, and in building interactive dashboards using Power BI and Streamlit. Strong client-facing skills with a proven track record of delivering SLA-driven data solutions.

Experience

Associate - Data Engineer - Web Scraping - Cognizant
2022 - Present

- **Data Request Management:** Evaluated and fulfilled client data requests, ensuring timely and accurate delivery in compliance with SLAs.
- **Data Extraction & Client Collaboration**: Worked closely with clients to understand data requirements and deliver customized data extractions using **SQL**, **PostgreSQL**, and **Shell scripting** for automated ETL workflows.
- **Workflow Automation & Web Scraping:** Developed automated data extraction and transformation pipelines using **Python (Pandas**, **NumPy)** and **Shell scripting;** built scalable web scraping solutions using **BeautifulSoup**, **Selenium**,including **CAPTCHA** handling and anti-bot workarounds.
- **Advanced SQL & Data Transformation:** Wrote complex SQL queries (joins, CTEs, subqueries) in PostgreSQL and SQL Server to retrieve, transform, validate, and optimize datasets.
- **Data Quality & Normalization:** Designed and implemented data cleaning, validation, and normalization routines, improving data accuracy and downstream analytical reliability.
- **Data Analysis & Dashboarding**: Generated actionable insights using Python and SQL; developed interactive dashboards using Power BI and Streamlit to monitor KPIs, trends, and business metrics.
- **API Integration:** Integrated third-party REST APIs for real-time data extraction and enrichment; utilized HTML and CSS for web-based data access and scraping.
- **Dashboard Automation:** Automated Power BI and Streamlit dashboards, enabling real-time data refreshes and reducing manual reporting effort.
- **Client-Centric Reporting:** Customized datasets and dashboards to align with business goals and stakeholder requirements.
- **Reusable Scraping & Storage Modules:** Built reusable **web scraping**, **CAPTCHA-handling**, and data storage modules, accelerating onboarding of new data sources.

Education

R.M.D.Enigneering College - Anna University
2018-2022

- Bachelor of Engineering in Electrical and Electronics

CGPA: 8.1

Skills

- Web Scraping & Automation :**Selenium, BeautifulSoup, REST APIs, PowerShell, CAPTCHA handling**
- Programming & Data Processing :**Python (Pandas, NumPy, Streamlit), SQL**
- ETL & Data Engineering :**Data Pipelines, Data Cleaning, Transformation, Normalization**
- Databases : **PostgreSQL, MS SQL Server**
- Data Analysis & Visualization :**Power BI, Tableau, Streamlit, Matplotlib, MS Excel**
- Web & App Development :**HTML, CSS, Flask, Bootstrap**
- Statistical & Analytical Methods :**Hypothesis Testing, Pattern Recognition**
- Other Skills :**Client Communication, Stakeholder Collaboration, Requirement Gathering**

Projects

Amazon Hunter – Automated Amazon Product Scraper (Web Scraping to Structured CSV)
- Built a Selenium-based scraper for Amazon India to extract product name, price, and ratings across multiple categories, handling dynamic content, pop-ups, and exporting data to CSV/JSON.

CAT Tracker: Incident Data Classification and Visualization App
- Developed a web-based tool to upload and categorize incident data from CSV/Excel files, generate visual summaries with charts, and export structured reports for analysis.

End-to-End Churn Analysis using Python, Pandas, and Matplotlib
- Analyzed Spotify user data to identify churn trends and behavioral patterns across subscription types and usage habits. Delivered insights on customer retention factors through data visualization and reporting."