

# Yugesh B

Data Engineer

## CONTACT

Chennai, Tamil Nadu, India

yugeshb26@gmail.com

+91 77083 25738

github.com/yugeshb26

linkedin.com/in/yugesh-b-331279230

## SKILLS

### Languages

Python, Perl, SQL

### Web Scraping & Automation

Selenium, Playwright, Scrapy, BeautifulSoup

### Databases

MySQL, MongoDB, SQLite, ClickHouse

### AI / GenAI

LLM Integration (Gemini, Ollama), RAG-style Grounding, Vector Databases, Prompt Engineering

### Data & Visualization

Pandas, NumPy, Matplotlib, Power BI, ETL Pipelines

### Cloud & Tooling

AWS S3, Git, GitLab

### Practices

Agile, Scrum

## CERTIFICATIONS

Salesforce Certified Administrator (SCA)

Salesforce

Python

LinkedIn Learning

Data Science (Machine Learning)

LinkedIn Learning

## EDUCATION

### BCA, Computer Programming & Applications

Vel-Tech Ranga Sanku Arts College, Chennai

2018 – 2021

## SUMMARY

Data Engineer with 4+ years across web-scraping automation, ETL/data pipelines, and GenAI-powered applications. Started in Perl-based publishing automation, moved through large-scale e-commerce scraping and Bloomberg financial data engineering at TCS, and now builds data infrastructure and AI-assisted tooling at Thurro. Comfortable owning a problem end-to-end — from a raw HTML page or PDF to a production pipeline or a working LLM-backed product.

## EXPERIENCE

### Thurro

Data Engineer · Dec 2025 — Present · Chennai, India

- Building and maintaining data engineering pipelines and analytics infrastructure across internal reporting and data-store systems.
- Working with database and reporting tooling spanning structured storage and analytical querying.

### Tata Consultancy Services

Senior Analyst (Python Data Analyst) — Key Client: Bloomberg · Feb 2025 — Nov 2025 · Chennai, India

- Developed Power BI dashboards to visualize Bloomberg financial datasets, enabling data-driven decisions for stakeholders.
- Automated data cleaning (Python/pandas) and ETL pipelines for Bloomberg market data feeds, improving processing speed by 30%.
- Built Python ETL scripts to extract, transform, and load large-scale financial data into SQL/MongoDB databases.
- Collaborated with Bloomberg's analytics team to optimize data-retrieval queries, reducing report generation time by 25%.
- Designed a web scraping tool (Scrapy/Selenium) to collect supplementary financial data from public sources.

### Lumina Datamatics

May 2022 — Feb 2025 · 2 yrs 10 mos · Chennai, India

#### Software Engineer

May 2024 — Feb 2025

- eBay web scraping: designed end-to-end scraping pipelines (Scrapy, BeautifulSoup, Selenium) against JavaScript-heavy, AJAX-based pages.
- Automated dynamic content handling with rate-limiting/retry mechanisms; cut scraping time by 40% via multi-threaded crawling.
- E-commerce data processing: built Python validation/transformation scripts and integrated REST APIs for real-time sync.

#### Junior Software Engineer

May 2023 — May 2024

- JXpress automated publishing platform: automated XML/HTML transformations using Perl and regex-based text processing.
- Developed Python/MySQL automation for content-classification systems, improving workflow efficiency by 25% across 500+ daily manuscripts.

#### Software Engineering Trainee

May 2022 — May 2023

- Onboarded into production software engineering practices — agile delivery, code review, and on-call support for internal systems.

## SELECTED PROJECTS

### Indian Law AI Chatbot

Streamlit legal Q&A assistant grounded in a 20.7K-entry curated dataset, with multi-key Gemini failover and a self-refreshing case-law pipeline via GitHub Actions.

Python · Streamlit · Google Gemini API · SQLite · Plotly · GitHub Actions

### Automated Document Intelligence Pipeline

Multi-engine scraper (Selenium → Playwright → HTTP) across 229 site configs, loading into a dual MySQL + ClickHouse pipeline with automated S3 archival — ~29K lines of Python.

Python · Selenium · Playwright · MySQL · ClickHouse · AWS S3