

# SHOURYA THAPLIYAL

Data Engineer | Software Engineer | Melbourne, Victoria (485 Visa)

+61 423 058 511 | shouryat32@gmail.com

linkedin.com/in/shourya-thapliyal | shouryathapliyal.com | github.com/shouryat32/Portfolio

---

## PROFESSIONAL SUMMARY

Data Engineer and Software Engineer with 2+ years production experience and Master's in Data Science. AI-first practitioner using Claude API and LLMs daily. Published researcher shipping end-to-end systems from concept to production at scale. Strong across Python, SQL, C#, cloud platforms (Azure, GCP), ML, and DevOps. Experienced designing data pipelines processing millions of records, building ML systems achieving 0.99+ accuracy, implementing robust testing and CI/CD. Portfolio-driven with published research and shipped production systems.

## EDUCATION

**Master of Data Science – University of Melbourne** | Melbourne, VIC | 2023 – Jan 2026 Specialisation: Big Data Analytics, Machine Learning, Cloud Computing.

**Bachelor of Engineering (Computer Science)** | Nitte Meenakshi Institute of Technology, India | 2016 – 2020 Published: Cardiac Arrhythmia Prediction Using Machine Learning (Springer ERCICA, 2020).

## WORK EXPERIENCE

**Data Engineer – ZeitPass** | Munich, Austria (Remote from Australia) | Apr 2026 – Present

- Design and operate production data platform processing 1,800+ cultural events weekly. Built ETL pipelines with Claude (Anthropic) LLM enrichment, semantic embeddings, pgvector storage supporting sub-second event matching.
- Engineered Python automation handling concurrent LLM requests, rate-limit handling, CLI-based reruns. Implemented 3-stage data quality pipeline with row-level security and comprehensive monitoring.
- Applied software best practices: Git version control, modular design, error handling, observability. Collaborated with stakeholders translating requirements into scalable solutions.

**Software Engineer – Tata Consultancy Services** | Bangalore, India | Jan 2021 – Dec 2022 Commonwealth Bank of Australia (50,000+ customers).

- Engineered 15+ REST APIs (C#/.NET) with zero security breaches, PCI DSS compliance, handling millions of daily transactions. Designed secure, scalable API solutions.
- Implemented test automation (Selenium, C#) achieving 95% code coverage, 60% regression testing reduction. Built CI/CD pipelines and deployment automation.
- Optimized application performance reducing production incidents by 40%. Collaborated across teams on system design and production reliability.

## PROJECTS

---

## Energy Price & Weather Correlation Analysis | Melbourne | Dec 2025 – Mar 2026

- Architected Azure Databricks medallion pipeline processing 26K+ hourly AEMO records with sub-5 minute refresh. Optimized data structures for high-performance analytics workloads.
- Developed three ML models (XGBoost, CatBoost): price forecasting  $R^2=0.89$ , demand prediction  $R^2=0.98$ , spike detection ROC-AUC=0.99. Built comprehensive data preprocessing pipelines.
- Deployed production Power BI dashboard (4 pages) enabling stakeholders to monitor energy market trends and make data-driven decisions. Full ML lifecycle demonstration.

## Phoenix Australia – Health Services Analytics | Melbourne | Mar 2024 – Dec 2025

- Built end-to-end ETL pipeline processing 1,000+ survey categories across 190+ programs (50K+ responses/week via Qualtrics API). Implemented multi-stage data quality and validation frameworks.
- Developed ML classification system (87-93% accuracy) using RoBERTa transformers and DBSCAN clustering for sentiment analysis and theme extraction.
- Deployed production system with 99.9% uptime, automated weekly outputs, processing millions of records. Reduced manual analysis time by 95%.

## TECHNICAL SKILLS

---

**Data Engineering:** Python (Pandas, NumPy, PySpark), SQL optimization, ETL/ELT, Apache Spark, Databricks, Data quality frameworks

**Machine Learning:** XGBoost, CatBoost, Scikit-learn, RoBERTa, Feature engineering, Model deployment

**Cloud Platforms:** Azure (Databricks, Data Lake, Data Factory), GCP, Cloud infrastructure

**Backend & APIs:** REST APIs, C# (.NET), Enterprise system design, 15+ production APIs, PCI DSS compliance

**Testing & DevOps:** Test automation (Selenium, C#), 95% code coverage, CI/CD pipelines, Docker, Git, Linux

**Visualization:** PowerBI, Tableau, Plotly, Matplotlib

**AI & LLMs:** Claude API (Anthropic), GitHub Copilot, Prompt engineering

**Languages:** Python, SQL, C#, R, JavaScript

## CERTIFICATIONS AND PUBLISHED WORK

- 
- Building with the Claude API – Anthropic
  - Google Data Analytics Professional Certificate
  - Data Quality, Management and Governance
  - Cardiac Arrhythmia Prediction (Springer ERCICA, 2020)
  - Portfolio: [github.com/shouryat32/Portfolio](https://github.com/shouryat32/Portfolio)