

# Eashwar Subramanian

Data Analyst / Data Scientist | Data Quality & Governance | SQL • Python • Excel • Power BI • Tableau

Melbourne, Australia | eashwars2001@gmail.com | +61 450 332 782 | LinkedIn: linkedin.com/in/eashwar-subramanian-02461a1b3 | GitHub: github.com/Eashwar-Subramanian

## SUMMARY

Early-career Data Analyst / Data Scientist with hands-on experience building end-to-end analytics and ML workflows (data ingestion, QA, evaluation, and reporting). Skilled in collecting, cleaning, and structuring structured + unstructured data to produce decision-ready insights. Strong in reproducible pipelines, data quality validation, and communicating technical outputs clearly to cross-functional stakeholders.

## PROFESSIONAL EXPERIENCE

### Data Science Intern

Cultural Infusion

Jan 2026 – Present

Melbourne, Australia

- Building an automated OSINT-style pipeline (GDELT + website/RSS scraping) to collect and structure public signals related to workplace culture, DEI, compliance, and reputational risk for ongoing monitoring.
- Defining taxonomies/keywords and data schemas; implement validation checks (deduplication, source normalization, timestamping, and quality flags) to support longitudinal trend comparisons.
- Developing NLP workflows to convert unstructured content into analyzable datasets (topic tagging + sentiment/thematic signals) for change detection over time.
- Contributing to internal decision-support outputs (alerts/briefings), including QA sampling and human-in-the-loop review to reduce noise and false positives.

### Data Science Intern

Solara Health

Jul 2025 – Nov 2025

Melbourne, Australia

- Built ingestion pipelines for a healthcare content library (PDF/HTML), including metadata governance and SHA-256 deduplication to improve dataset reliability.
- Developed automated download + parsing workflows with retries, content-type handling, and standardized text outputs to stabilize downstream retrieval datasets.
- Designed PostgreSQL data structures for healthcare content (metadata + embeddings) with indexing patterns, and exposed curated datasets via a FastAPI service with basic automated tests (Pytest) for reliability.
- Implemented tenant-aware access controls using PostgreSQL Row Level Security (RLS) to enforce segregation across multi-hospital deployments.
- Collaborated in an Agile team environment using Jira/Confluence/Bitbucket, contributing to delivery planning, documentation, and iterative improvements.
- Built QA + audit tooling (flag logging + CSV exports, citation audit harness, labelled evaluation dataset) and ran repeatable evaluation reporting (RAGAS + custom checks) to track quality over time.

### Data Analyst Intern

PreInsta Pvt Ltd (Remote)

Dec 2023 – Feb 2024

India

- Delivered 5+ stakeholder dashboards (Tableau, Excel) and performed source-to-dashboard cross-checks to improve metric consistency; increased client engagement by 10%.
- Optimized SQL reporting queries across multiple tables, reducing data retrieval time by 25% for executive reporting.
- Automated reporting workflows with Python (BeautifulSoup) cutting manual effort by 30%.

## PROJECTS

### Solara Healthcare RAG System | Jul 2025 – Nov 2025

- Built a reproducible monitoring workflow (labelled test set + automated evaluation runs) and tracked benchmarks via RAGAS (faithfulness 69.4%, answer relevancy 54.3%, context precision 92.7%, context recall 91.2%).
- Documented system design and QA controls (governance + audit outputs) to support reliable, repeatable evaluation and stakeholder review.

### Australian Retail Customer Segmentation | Jun 2025

- Identified 106 loyal high-value customers generating 15% of revenue (~\$47K) using RFM + K-Means clustering, translating outputs into targeted marketing recommendations.
- Repaired 1,032 corrupted records (72.5% invalid order IDs) in a 5,000-transaction dataset to enable accurate segmentation and reporting.

### Climate Policy Forecasting Dashboard | Jul – Sept 2024

- Built a SARIMA forecasting dashboard achieving 1.2°C MAE for 12-month temperature forecasts to support climate trend analysis and planning use cases.
- Integrated a Flask app with interactive Folium mapping and location search for stakeholder exploration.

## TECHNICAL EXPERTISE

Languages: Python, SQL, R

Cloud/DevOps: AWS (Bedrock, Aurora, S3, Lambda, EC2), Docker, CI/CD, Git

Analytics & ML: NLP (text analytics), forecasting (SARIMA), segmentation (RFM, K-Means), scikit-learn, RAG (LlamaIndex), embeddings, evaluation/monitoring (RAGAS)

Databases: PostgreSQL (pgvector, RLS, indexing), MySQL, DynamoDB

Frameworks / Libraries: FastAPI, Flask, Pytest, Pandas, NumPy, pdfplumber, BeautifulSoup

Other: Data governance (RLS/tenant isolation), metadata management, audit logging, reproducible evaluation/monitoring

## EDUCATION

### Master of Data Science

Royal Melbourne Institute of Technology (RMIT)

GPA: 3.6/4.0 | Dec 2025 | Melbourne, Australia

### Bachelor of Electronics & Communication Engineering

Rajagiri School of Engineering and Technology

First Class Distinction (8.8/10) | Aug 2023 | Kerala, India

## PROFESSIONAL DEVELOPMENT

### Data Science Melbourne Meetup - Active Member

Jun 2024 – Present

Regular participation and networking; discussions on ML trends, Power BI, and applied Python analytics.

Relevant Coursework: Machine Learning • Big Data Processing • Cloud Computing • Advanced Programming • Computer Vision • Probability & Random Processes