

# 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: <https://www.linkedin.com/in/eashwar-subramanian-02461a1b3/>

GitHub: <https://github.com/Eashwar-Subramanian>

## SUMMARY

Early-career Data Analyst / Data Scientist with hands-on experience delivering end-to-end data pipelines, analytics, and machine learning workflows. Experienced in collecting, cleaning, and structuring structured and unstructured data to generate decision-ready insights and support evidence-based decision-making. Skilled in building reproducible analytical workflows, monitoring data quality and model behavior over time, and applying statistical and machine learning techniques in real-world, evolving data environments. Known for a rigorous analytical approach, strong collaboration, and translating complex technical outputs into clear insights for cross-functional stakeholders.

## PROFESSIONAL EXPERIENCE

### Data Science Intern

Jan 2026 – Present

Cultural Infusion

Melbourne, Australia

- Design and implement analytical workflows for monitoring public DEI-related signals, including careful data preparation, exploratory analysis, and insight validation across evolving sources.
- Build automated data pipelines to collect and structure public signals (news articles, public statements, online discourse), enabling longitudinal analysis and comparison of cultural and sentiment trends over time.
- Develop text-based analysis workflows to transform unstructured public content into standardized, analyzable datasets, supporting sentiment and thematic pattern analysis across time periods.
- Contribute to early-stage analytics and data product development by shaping data schemas, validation logic, and monitoring outputs for internal decision-support tools.

### Data Science Intern

July 2025 – Nov 2025

Solara Health

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 and content-type handling, improving consistency of downstream datasets for analysis and retrieval.
- 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.
- Built QA + audit tooling (flag logging + CSV exports, citation audit harness, labelled evaluation dataset) and reported quality trends over time using reproducible metrics.

### Data Analyst Intern

Dec 2023 – Feb 2024

PrepInsta Pvt Ltd (Remote)

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 | July 2025 – Nov 2025

- Documented the system design and QA approach (governance controls + audit outputs) and created a repeatable evaluation workflow to track reliability over time.
- Created a reproducible monitoring pipeline (labelled test set + automated evaluation runs) and reported quality benchmarks via RAGAS (faithfulness 69.4%, answer relevancy 54.3%, context precision 92.7%, recall 91.2%).

### Australian Retail Customer Segmentation | June 2025

- Identified a high-value customer segment representing 15% of total revenue using K-Means clustering, translating analytical outputs into targeted marketing recommendations.
- Corrected 72.5% invalid order IDs in 5,000-transaction dataset enabling accurate RFM analysis.

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

- Built SARIMA forecasting system achieving 1.2°C MAE, supporting climate trend analysis for Australian policy and planning use cases.
- Integrated Flask web app with interactive Folium mapping for urban planning stakeholder support.

## TECHNICAL EXPERTISE

Languages: Python, SQL, R

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

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

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

Visualization: Power BI, Tableau, Matplotlib, Excel

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

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

June 2024 – Present

Networking with 50+ analytics professionals | ML trends, Power BI, Python applications

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