

JUNO LI

Data Scientist | Data Analyst

Location: Perth, WA

Email: juno.li.research@gmail.com

Portfolio: jun01ee.github.io

Linkedin: linkedin.com/in/jun01ee

PROFILE SUMMARY

Data professional with more than five years' experience analysing complex datasets, developing statistical models and building reproducible data workflows. Skilled in Python, SQL and R, with practical project experience in Databricks, PySpark, machine learning and automated data pipelines. Experienced in transforming noisy, multi-source data into reliable analytical datasets, investigating data-quality issues and communicating practical insights to technical and non-technical stakeholders.

TECHNICAL SKILLS

Data Analysis: data validation, anomaly investigation, exception analysis, reconciliation, data cleansing, data transformation, quality checks, metadata, documentation, root-cause analysis

Analytics and Reporting: statistical analysis, trend analysis, forecasting, visualisation, KPI and reporting metrics, Power BI dashboard development, DAX measures and calculated columns, matplotlib, Excel

Programming and Data Tools: Python, SQL, R, pandas, NumPy, SciPy, PostgreSQL, REST APIs, GitHub

Data Workflows: reproducible workflows, ETL concepts, scheduled pipelines, API ingestion, structured analytical datasets, workflow logging, automated validation, configuration management

Stakeholder Delivery: requirements clarification, responding to technical queries, written communication, technical documentation, cross-disciplinary collaboration, prioritisation and attention to detail

PROFESSIONAL EXPERIENCE

The University of Western Australia, Perth

Postdoctoral Research Associate

Apr 2026 – Present

Research Assistant, Part-time

Oct 2025 – Feb 2026

Doctoral Researcher

Mar 2022 – Sep 2025

- Built reproducible Python and R workflows for data ingestion, cleansing, transformation, validation, analysis and reporting.
- Investigated data irregularities, documented likely causes and communicated resolution steps to support reliable downstream reporting.
- Prepared and delivered validated analytical outputs, reports and datasets for downstream users and collaborators.
- Applied statistical analysis, anomaly investigation, uncertainty assessment and model validation to identify patterns, risks and potential data-quality issues.
- Produced visual reports, diagnostic plots and written documentation to explain data limitations, assumptions and findings to technical and non-technical stakeholders.
- Managed competing analytical priorities across collaborative projects, maintaining accuracy, documentation and clear communication under changing requirements.
- Re-engineered notebook-based workflows into modular software with improved configuration management, automated validation and version control.

SELECTED PROJECTS

Power BI Movie Financial Performance Dashboard – [GitHub](#)

- Built a multi-page interactive Power BI dashboard with custom report backgrounds, page navigation, active tooltips, slicers and simple DAX measures/calculated columns to support exploratory analysis and stakeholder-oriented reporting.
- Tools: Power BI, DAX, Python, pandas, TMDb API, CSV, Git/GitHub, pytest, data modelling, data validation
- Skills: Power BI reporting, DAX measures, calculated columns, dashboard design, page navigation, report tooltips, data modelling, stakeholder-oriented visualisation

Solar Yield Forecasting MLOps Pipeline – [GitHub](#)

- Built an automated workflow that produces and publishes refreshed seven-day solar-yield forecasts from external weather data.
- **Tools:** Python, Databricks, PySpark, SparkSQL, REST APIs, XGBoost, MLflow
- **Skills:** Forecasting, feature engineering, workflow automation, validation, scheduled inference

Data Engineering Zoomcamp Capstone Project – [GitHub](#)

- Built an end-to-end pipeline that converts raw source data into validated, structured datasets for downstream analysis.
- **Tools:** Python, SQL, PostgreSQL, GCP, GitLab, Google Looker
- **Skills:** ETL, orchestration, data modelling, schema validation, data quality, CI/CD

Large-Scale Astronomical Data Pipeline – [Architecture](#)

- Processed and indexed more than 20TB of archival imaging data during my PhD, enabling efficient discovery of related observations without repeatedly inspecting large source files.
- **Tools:** R, Linux, astronomical image formats, WCS
- **Skills:** Metadata engineering, spatial matching, reconciliation, large-scale processing, data-quality investigation

Spatially Resolved Galaxy Analysis Software – [Architecture](#)

- Re-engineering complex research workflows into reusable and documented software as part of the current postdoctoral role.
- **Tools:** Python, R, Git/GitHub, VS Code, AI coding agents
- **Skills:** Modular software design, configuration management, testing, validation, documentation, human review

ADDITIONAL EXPERIENCE

Quantitative Developer (Intern) – Mai Capital Ltd, Hong Kong

Jun 2019 – Nov 2019

- Built Python and SQL workflows using broker APIs to monitor portfolio positions, performance and compliance breaches in near real time.
- Applied financial risk metrics and stochastic-process concepts to support strategy evaluation and automated controls.

EDUCATION

PhD in Astrophysics – University of Western Australia (ICRAR)

Completed Feb 2026

MPhil in Astrophysics – The University of Hong Kong, Hong Kong

2016–2019

Bachelor of Science (Hons) – The University of Hong Kong, Hong Kong

2012–2016

PROFESSIONAL DEVELOPMENT

Data Engineering Zoomcamp Certificate — DataTalks.Club

Issued Apr 2025

[Google Data Analytics Certificate – Coursera](#)

Issued Sep 2022