

# Maxson Stephen Mathew

0455411395 | [msmmj6499@gmail.com](mailto:msmmj6499@gmail.com) | Melbourne

[Linkedin](#) | [Portfolio website](#)

## SUMMARY

Data Analyst with Master of Data Science and hands-on experience developing AI/ML solutions, analytics pipelines, and cloud-based data systems. Proven ability to translate complex business challenges into technical solutions through stakeholder collaboration and data-driven insights. Strong technical foundation in Python, SQL, and modern data platforms (Azure, GCP, AWS) with demonstrated expertise in RAG architectures, predictive modelling, and business intelligence. Experienced in delivering production-ready solutions that drive measurable business outcomes in fast-paced environments.

## SKILLS

**Programming & Analytics:** Python | SQL | R | Statistical Analysis | Machine Learning | Predictive Modelling | RAG Architecture.

**AI & LLM Technologies:** LangChain | Prompt Engineering | Groq API | Document AI | RAG | Semantic Extraction.

**Cloud Platforms:** Microsoft Azure (Fabric, Lakehouse) | Google Cloud Platform (BigQuery) | Databricks.

**Data Engineering:** ETL Pipelines | PySpark | Data Quality Assurance | Data Governance | Distributed Computing.

**Business Intelligence:** Power BI | Tableau | Executive Dashboards | Data Storytelling | Stakeholder Engagement.

## EDUCATION

### Master of Data Science

Feb 2023 - July 2025

*The University of Melbourne, Parkville, Melbourne*

**Relevant Coursework:** Statistical Machine Learning, Database Management, Cloud Computing, Spatial Data Analytics.

- Led analytics projects generating actionable insights through comprehensive stakeholder collaboration.
- Developed expertise in modern data architectures and business intelligence platforms.

### Bachelor of Technology – Computer Science

Aug 2017 - Jul 2021

*Guru Gobind Singh Indraprastha University, Delhi, India*

## PROFESSIONAL EXPERIENCE

### Data Science Intern

Feb 2024 - Oct 2024

**John Holland Group, Melbourne, Australia**

- Applied Social Network Analysis using NetworkX to identify influential personnel across 10,000+ employee network through centrality metrics, supporting strategic workforce planning and organisational design.
- Performed advanced feature engineering using Social Network Analysis metrics and implemented decision tree algorithms for agent-based modelling with Mesa framework, incorporating organisational behaviour rules to predict workforce dynamics.
- Developed scalable data pipelines using PySpark on Databricks for high-volume workforce data processing, implementing distributed computing for network analysis and predictive modelling at enterprise scale.
- Designed analytics solutions using Python, SQL, and Power BI achieving 91% improvement in retention prediction accuracy, translating complex analysis into actionable workforce strategies for executive stakeholders.
- Collaborated closely with business stakeholders across HR, operations, and executive teams to translate complex network analysis findings into actionable workforce strategies and organisational design recommendations.
- Implemented data governance frameworks including quality assurance, metadata management, and compliance standards while providing technical training to 15+ business users.

### Associate System Engineer

Aug 2021 – Dec 2022

**Tata Consultancy Services, Bengaluru, India**

- Provided production support for Deutsche Bank Belgium's banking applications, managing deployments and monitoring application health using SQL, Unix, and Kubernetes.
- Generated operational reports using SQL and Excel, providing stakeholder insights that improved system performance monitoring and strategic decision-making.
- Led technical problem-solving reducing mean time to resolution by 95% while maintaining 99.5% application uptime through distributed systems analysis.

## PROJECTS

---

### AI-Powered Invoice Data Extractor - RAG Application

*Jan 2026 - Feb 2026*

*Personal Portfolio Project | RAG AI framework with LLM*

- Built RAG-based invoice extraction system using LangChain and Llama 3.3 70B, processing diverse PDF formats and outputting structured JSON with vendor, amounts, dates, and line items.
- Implemented prompt engineering techniques achieving 90% extraction accuracy without format-specific rules, demonstrating practical application of semantic AI understanding.
- Deployed production application on Streamlit Cloud integrating PyMuPDF for document processing, Groq API for free LLM inference, and Pandas for data visualisation.
- Solved real business challenge: automated solution could reduce accounts payable manual entry by 80-90%, saving 40+ hours monthly for mid-size organisations.

### End-to-End Data Analytics Solution - NYC Taxi Operations Analytics

*Aug 2025 - Sept 2025*

*Personal Portfolio Project | Modern Data Architecture & Self-Service Analytics*

- Built a scalable data pipeline using Microsoft Fabric Lakehouse implementing data quality practices and automated testing for operational analytics.
- Developed business intelligence solution with Power BI dashboards featuring KPI cards and trend analysis, enabling self-service analytics for decision-making.

## PROFESSIONAL CERTIFICATIONS

---

- Databricks Fundamentals Accreditation
- Databricks AI Agents Fundamentals Accreditation
- Machine Learning, Data Science and Deep Learning with Python - Udemy
- Deloitte Australia - Data Analytics Job Simulation (Forage)