MAHESH SAI KANDULA

Sydney, NSW • +61 451138 996 • maheshsaikandula@gmail.com • LinkedIn • GitHub

PROFESSIONAL SUMMARY

Master of Data Science graduate (Macquarie University) working with data pipelines end to end — schema design, ETL, data-quality engineering, SQL business marts, ML modelling, and BI dashboards. Comfortable across the full data lifecycle in PostgreSQL, Python, and Tableau, turning messy, large-scale data into trustworthy, decision-ready insight. Seeking a junior data analyst or analytics engineer role in Sydney.

TECHNICAL SKILLS

Languages: Python, SQL
Databases: PostgreSQL
Data & Analytics: pandas, NumPy, ETL, data cleaning & validation, data modelling, statistical analysis, exploratory data analysis
BI & Visualisation: Tableau, PowerBI, matplotlib
Machine Learning: scikit-learn, XGBoost, SHAP (model explainability)
Tools & Cloud: Git, Jupyter, Jira, Excel, AWS (fundamentals)

PROJECTS

PayScope — Digital Payments Risk & Merchant Analytics Platform *PostgreSQL · SQL · Python · Tableau*

- Engineered an end-to-end payments analytics and risk-monitoring platform on PostgreSQL using a layered raw → clean → mart architecture across 3 schemas, processing 100K+ transactions across 10K customers and 1.2K merchants.
- Built a Python (pandas, SQLAlchemy, psycopg2) ETL and data-quality framework that injects, detects, and flags realistic issues (nulls, duplicates, broken foreign keys, invalid codes); invalid records are flagged rather than deleted to preserve a full audit trail, lifting trusted transaction coverage to 69.7%.
- Developed 7 SQL business marts using an aggregate-before-join pattern to prevent grain-mismatch double-counting, powering merchant-performance, chargeback, and customer-risk reporting.
- Designed an explainable rule-based risk engine and behaviour-driven merchant risk segmentation (Critical / High Risk / Monitor / Low Risk), surfacing that source-system risk tiers diverged from actual merchant behaviour.
- Built 3 interactive Tableau dashboards (Transaction Monitoring, Merchant Risk, Chargeback Analysis) with KPI tiles, dual-axis time series, and a risk scatter, connected live to the mart layer.

Sydney Airbnb Guest Satisfaction Classifier *Python · scikit-learn · XGBoost · SHAP*

- Built an end-to-end supervised ML pipeline on Inside Airbnb's Sydney dataset (17,730 listings, Sep 2025 scrape) to classify listings as top-tier vs. standard, defining the target (review_scores_rating ≥ 4.9, ≥5 reviews) by class-balance testing rather than an arbitrary percentile cut.
- Diagnosed a 100%-null price column (verified at both listing and calendar level) and pivoted feature strategy to host- and property-behaviour signals instead of dropping the dataset, documenting the sourcing decision.
- Engineered features from raw fields — host tenure from signup date, amenities parsed into count and binary flags, bathrooms recovered via regex — and excluded leakage columns (review sub-scores, host_is_superhost) to keep the model honest.
- Compared Logistic Regression, Random Forest, and XGBoost, tuning the final model with RandomizedSearchCV (3-fold CV) to reach ROC-AUC 0.771 on held-out test data.
- Used SHAP (TreeExplainer) to explain predictions and identified host portfolio size — not price or amenities — as the strongest driver of guest satisfaction (46.6% top-tier rate for single-listing hosts vs. 16.2% for hosts with 20+ listings), a finding that also explained a misleading instant-bookable pattern in the raw data.

EDUCATION

Master of Data Science, Macquarie University, Sydney — WAM: 76.6 2024 – 2026
B.Tech, Computer Science & Engineering, GITAM University, India — CGPA: 8.75 / 10 2020 – 2024

CERTIFICATIONS

- IBM Data Science Professional Certificate (Coursera)
- Google Data Analytics Professional Certificate (Coursera)