

Pavan Kumar Regalla

Data Analyst

Jersey City, NJ | +1 917-979-3462 | pavan.launch@gmail.com

[linkedin.com/in/thepavanlkdnxyz](https://www.linkedin.com/in/thepavanlkdnxyz) | github.com/pavanregalla1 | pavanregalla.dev

PROFESSIONAL SUMMARY

Data Analyst with nearly five years across data analytics and engineering. Strong in SQL, Python, Power BI, Tableau and AWS, with experience in KPI reporting, dashboard development, data validation and business analysis — turning reliable data into decision-support reporting.

TECHNICAL SKILLS

Languages and Analytics: SQL, Python, Pandas, PySpark

Data Engineering and Processing: ETL/ELT, data validation and reconciliation, dimensional modeling

Cloud and Data Platforms: AWS, Snowflake, Azure Databricks

Databases and Warehousing: Amazon Redshift, Snowflake, SQL Server, star schema

Business Intelligence and Visualization: Power BI, DAX, Power Query, Tableau, Excel, KPI reporting, data visualization

Orchestration, Quality and Development Tools: Apache Airflow, data-quality checks, Git

PROFESSIONAL EXPERIENCE

Data Analyst | Amazon Web Services | United States *July 2024 – Present*

- Wrote SQL and Python/Pandas analyses over data landed by ETL/ELT pipelines (AWS Glue, S3, EMR) to answer stakeholder questions with validated, reproducible results.
- Built and maintained analysis-ready datasets from Redshift and Snowflake dimensional models, giving stakeholders consistent KPI definitions instead of conflicting spreadsheet numbers.
- Developed recurring dashboards and reports tracking key metrics for business stakeholders, with every measure reconciled back to validated source data.
- Investigated data-quality issues in Airflow-orchestrated batch and near-real-time workflows, tracing discrepancies to source and resolving them before they reached reporting.
- Gathered stakeholder requirements and translated them into dataset and dashboard changes, delivering decision-support reporting on a reliable schedule.

Data Analyst | Deloitte | India *January 2021 – July 2023*

- Analyzed data from multiple source systems ingested through ETL workflows (Python, SQL Server, SSIS, Azure Data Factory), delivering cleansed, validated datasets for management reporting.
- Built reports and dashboards in Power BI on consistent dimensional models, replacing manual spreadsheet hand-offs with repeatable, validated reporting.
- Wrote stored procedures, views and Excel-based reconciliations in SQL Server to validate completeness and accuracy of reporting data.
- Ran data profiling in Azure Databricks to investigate anomalies and data-quality issues, documenting findings so stakeholders could act on them.

TECHNICAL PROJECTS

Retail Sales Analytics — Batch Lakehouse & Power BI Pipeline

github.com/pavanregalla1/retail-sales-analytics

- Modeled 60,600 synthetic raw orders into a gold star schema (56,420 clean fact rows, \$80.96M modeled revenue) with KPI queries for revenue, category and regional performance; 10/10 data-quality checks passing, 2,113 bad rows quarantined, never silently dropped.
- Built dashboard-ready datasets with documented measure definitions so every KPI reconciles to validated source data instead of conflicting spreadsheet numbers.
- Scope: no packaged PBIX file is included; a Power BI file can be built by following the repository's 10-minute guide. Dashboard images are matplotlib renders of the gold data; all data is synthetic.

Real-Time Retail Intelligence — Kafka, Spark & ML Platform

github.com/pavanregalla1/realtime-retail-intelligence

- Analyzed 80,628 synthetic streaming events: fraud detection reached 100% recall on injected scenarios at a 3.02% alert rate, and demand-forecast MAE came in 83.4% below a naïve baseline; 24/24 dbt and pytest checks green.
- Validated alert quality on the gold layer — every flagged event traces to a documented rule or model score, keeping the 3.02% alert rate reviewable rather than a black box.
- Scope: all data is synthetic. The file-replay path is the CI-verified route; the Kafka/Docker streaming route was implemented but not executed in this environment.

Cloud Analytics Knowledge Search — Hybrid Retrieval & Grounded QA

github.com/pavanregalla1/cloud-analytics-knowledge-search

- Built hybrid TF-IDF plus dense (sentence-transformers) retrieval over 300 synthetic enterprise documents with extractive answers, document citations and honest abstention; hybrid recall@1 of 87.5% beats TF-IDF (75.0%) and dense (85.0%) alone on a 40-question evaluation set.
- Scope: corpus and questions are synthetic. Answers are extractive — no LLM is executed; Bedrock, LangChain and Docker are documented integration points, not run here.

CERTIFICATIONS

- AWS Certified Data Engineer – Associate
- Microsoft Certified: Power BI Data Analyst Associate (PL-300)
- Databricks Certified Data Analyst Associate

EDUCATION

- Master of Science in Computer Science | Adelphi University | NY | Sep 2023–May 2025
- Bachelor's Degree | JNTUH | Hyderabad, India | 2018–2022