Bandaru Raju

Boca Raton, FL   [Portfolio Website](#)   [Github](#)   +1 (305)-954-1077   [Linkedin](#)   rajbandaru1201@gmail.com

Data engineer with experience building scalable, end-to-end data pipelines and interactive dashboards using AWS, Databricks, and PostgreSQL. Specialized in fraud detection, ETL automation, and GPT-integrated analytics for real-world impact.

SKILLS

Languages & Libraries: Python, SQL, Pandas, PySpark, Seaborn, Scikit-learn
Cloud & Platforms: AWS (Lambda, RDS, Glue, Redshift), Databricks, Airflow (MWAA), Docker
Databases & Storage: PostgreSQL, MySQL, Delta Lake, Parquet

Tools & Dashboards: Streamlit, GPT API, ETL Pipelines, Data Cleaning, Data Visualization

PROJECTS

RetailX – Fraud & Sales Intelligence Dashboard (3M+ Orders) https://github.com/Rajbandaru9090/Databricks_Fraud_ai

- Built RetailX, an AI-powered dashboard that processed 3M+ grocery orders using PySpark + Delta Lake, enabling **5x** faster query performance through caching, partitioning, and storage optimization.

- Developed a fraud detection system combining rule-based logic with unsupervised ML (Isolation Forest, KMeans), achieving 92% anomaly detection precision and identifying 2,300+ high-risk transactions.

- Designed a GPT-integrated assistant using OpenAI API to support natural language business queries (e.g., "Top 5 categories with high fraud"), enabling AI-driven decision-making for saleSs and fraud analytics.

- Delivered real-time Streamlit dashboards with advanced filtering (region, category, time, fraud level), fraud heatmaps, and 15+ business KPIs including revenue leakage, return hotspots, and churn-prone segments.

Cloud-Based ETL + Dashboard Pipeline (AWS + PostgreSQL)

- Built a cloud-native ETL pipeline that automated ingestion, transformation, and loading of sales data from raw files in AWS S3 into PostgreSQL (RDS) using Pandas-based Python scripts orchestrated with AWS Airflow (MWAA).

- Engineered daily ETL workflows in Airflow with task retries, dependencies, and logging, ensuring robust data refresh and pipeline reliability (99.9%) across environments.

- Cleaned and transformed large datasets (~100K+ rows) including type standardization, missing value imputation, and normalization before persisting to PostgreSQL for downstream analytics.

AI Resume Screener – GPT, Streamlit, Skill Matching Dashboard https://github.com/Rajbandaru9090/ai-resume-screener

- Developed a GPT-powered ATS screener to analyze and score 20+ resumes against job descriptions using skill extraction, semantic matching, and prompt engineering.

- Built a Streamlit dashboard with advanced filtering, heatmaps, and skill match radar charts to visualize matched vs missing resume keywords.

- Enabled batch resume processing and CSV download with fully deployable architecture on Streamlit Cloud, demonstrating real-world AI and data product development.

EDUCATION

Master of Science in Computer Science   **Expected May 2026**
Florida Atlantic University , Boca Raton

Bachelors of Science in Information Technology   **Aug 2020 – May 2024**
Sreenidhi Institute of Science and Technology, Hyderabad, India