

ABHINAV JAJOO

Data Engineer | Turning Data into Business Insights

Email: abhinavajoo49@gmail.com | Phone: +91-9549339414

LinkedIn: linkedin.com/in/abhinav-jajoo-b81b2b203

GitHub: github.com/Abhinav-source2

Portfolio: abhinav-portfolio-khaki.vercel.app

ABOUT

Data Engineer | Building Scalable & Automated Data Systems

Final-year Computer Science student with hands-on experience in building and orchestrating ETL/ELT pipelines using Apache Airflow, Kafka, and Docker. Skilled in data ingestion, transformation, and validation across cloud environments (AWS). Specializing in optimizing SQL workflows and automating reporting processes with Power BI, with a focus on creating reliable, production-ready data pipelines that generate actionable business insights.

EXPERIENCE

Data Analyst Intern

Onsite | May 2024 – July 2024

Tools: Python, SQL, Power BI, Excel, Selenium, BeautifulSoup, Postman

FUDR is a SaaS startup focused on restaurant analytics and customer data optimization.

- Automated data ingestion workflows by scraping 15–20 restaurant and food websites using Python (Selenium, BeautifulSoup, Postman) — reducing manual data collection time by ~20 hours/week.
- Engineered structured data storage formats in collaboration with backend developers, improving query performance and ensuring dashboard compatibility across analytics systems.
- Refactored SQL queries and optimized join operations to reduce dashboard refresh latency by 40%, enabling faster business reporting.
- Consolidated 600+ raw client leads into a cleaned, validated dataset, enhancing the accuracy of sales outreach and trend analysis.
- Designed and deployed Power BI dashboards integrated with automated refresh pipelines, providing near real-time visibility into customer and competitor insights.

KEY PROJECTS

Smart Revenue Optimization & Anomaly Detection Platform (SROAD)

Tools: AWS, Airflow, Kafka, Git Version, Docker, Terraform, Streamlit, LLaMA-based AI Agent

- Designed an end-to-end real-time data pipeline using Apache Kafka to ingest Shopify and synthetic order events, unifying batch and streaming data in an AWS S3 data lake for scalable analytics.
- Built serverless analytics on AWS using Glue and Athena, creating unified views and advanced SQL queries to analyze revenue trends, regional performance, and high-value order anomalies in near real time.
- Developed an interactive Streamlit dashboard integrated with Athena to visualize live sales metrics, trends, and anomaly indicators, enabling fast, data-driven operational insights.
- Implemented an AI-assisted anomaly detection layer combining rule-based logic and a LLaMA-based agent for natural-language explanations, with Docker and Terraform for automated.

Breast Cancer Classification with Grad-CAM GitHub:

[Breast Cancer Project](#)

Tools: EfficientNetB0, TensorFlow, Grad-CAM, Python

- Trained a deep learning model on CBIS-DDSM mammogram dataset to classify images as benign or malignant.
- Achieved 93% test accuracy using EfficientNetB0 with advanced preprocessing and augmentation techniques.
- Applied Grad-CAM visualizations to interpret model predictions and ensure medical explainability.
- Implemented cross-validation to minimize overfitting and improve generalization.

Sentiment Analysis Pipeline & Dashboard GitHub:

Sentiment Pipeline Repo

Tools: Apache Airflow, Docker, Power BI, Python

- Developed a fully automated NLP pipeline to process 150K+ YouTube and Reddit comments for sentiment insights.
- Built ETL workflows in Airflow to extract, clean, and classify text data using pre-trained deep learning models.
- Containerized the pipeline with Docker, ensuring scalability and consistent local-to-production performance.
- Created Power BI dashboards visualizing live sentiment trends and keywords, reducing manual processing by 95%.

TECHNICAL SKILLS

- Languages: Python, SQL, Java, JavaScript.
- Data Engineering: Airflow, Kafka, ETL/ELT, PySpark, Hadoop, Data Modeling
- Cloud & DevOps: AWS (S3, Glue, Lambda, Athena), Docker, Vercel, CI/CD
- Databases: PostgreSQL, MongoDB
- Generative AI & LLM Systems: LangChain, LLaMA, Retrieval-Augmented Generation (RAG), Prompt Engineering, LLM Fine-Tuning, Embeddings, Tokenization, Context Window Optimization
- Machine Learning: TensorFlow, Keras, Transformers, Scikit-Learn, Grad-CAM
- Tools: Power BI, Tableau, Pandas, NumPy, Selenium, BeautifulSoup, Git
- Frontend: React, Tailwind CSS, Three.js

EDUCATION

B.Tech – Computer Science & Engineering

JK LakshmiPat University | 2022–2026

Relevant Coursework: Machine Learning, Deep Learning, Big Data, DBMS, Cloud Computing, Data Structures

CO-CURICULAR ACTIVITIES

- Patent Pending: Arduino-based Multi-Sensory Museum System (Hardware + UX Design)
- Event Coordinator, JKLU Spardha Fest — managed 5+ events with 500+ participants
- 5★ in Java & C on HackerRank