

KHAJA MUJAHIDDIN MOHAMMED

CT USA | +1 347-736-5812 | muhammedshoaib835@gmail.com | [LinkedIn](#) | [Portfolio](#)

Data Scientist and Python Developer experienced in building scalable data pipelines, real-time streaming systems, and cloud-native analytics on AWS, Azure, and Hadoop. Skilled in ETL design with Glue, Lambda, SSIS, PySpark, and SQL; developing ML/NLP models for sentiment analysis, forecasting, and recommendations using Python, R, spaCy, and LSTM. Delivered 20+ Power BI and R Shiny dashboards for 500+ stakeholders, improving decision-making and reducing manual reporting by 40%. Hands-on with Kafka, Kinesis, and Spark for real-time processing and 99.9% uptime.

WORK EXPERIENCE

Data Scientist - Intern | Innover Global INC| Aurora, IL

(Aug 2025 – Present)

- Designed and deployed real-time data pipelines using AWS Kinesis, Glue, and Lambda to process 100K+ events/day with 25% lower latency and 99.9% uptime.
- Automated ETL workflows with PySpark and AWS Glue for ingesting structured/unstructured data, optimizing storage and cloud costs by ~\$5K annually.
- Implemented CI/CD pipelines and security best practices (CodePipeline, IAM, VPC, CloudWatch), reducing deployment time from 2 days to 4 hours while ensuring governance.

Jr. Python Developer | SRIK Consulting Services | Hyderabad, India

(Sep 2020 – July 2023)

- Designed 20+ ETL pipelines using Python, SSIS, and PySpark on Hadoop/HDFS; ingested data from MySQL, SQL Server, Excel, and APIs into OLAP warehouses with star schema modeling and Row-Level Security (RLS) in Power BI for 500+ secure users across finance, HR, and operations
- Automated Power BI dataset refreshes via pandas + SQLAlchemy schedulers, cutting manual reporting by 40% (~10 hrs/analyst/week)
- Developed RESTful APIs using Flask/FastAPI to deliver real-time KPIs and integrate with HR/finance platforms

PROJECTS

Real-Time Customer Segmentation Pipeline | PySpark, Scikit-learn, pandas, NumPy, PCA, Power BI, DAX, Power Query

- Architected end-to-end ETL + ML pipeline using Python, PySpark, Scikit-learn, and K-Means clustering to segment customers by purchasing patterns, lifetime value, and behavior trends
- Performed feature engineering, PCA dimensionality reduction, and automated data ingestion/retraining with Python schedulers. Visualized clusters, trends, and KPIs in interactive Power BI dashboards with DAX

Multimodal Fake News Detection Pipeline | Apache Airflow, Python, Hugging Face Transformers, TensorFlow

- Engineered scalable Airflow-orchestrated data pipeline fusing text (NLP) and metadata features using Hugging Face Transformers, TensorFlow, and Scikit-learn
- Applied preprocessing (tokenization, stop word removal, lemmatization) and achieved 85% accuracy on 25K+ labeled news items. Deployed model to Hugging Face Spaces for real-time inference and public access

E-Commerce Data Warehouse & Dashboard | DAX, Power Query, Advanced Excel, Data Modeling, Data Warehousing

- Designed star schema data model & built centralized data warehouse using Power Query, Excel, & Power BI
- Preprocessed and cleaned 4 years (2017–2020) of sales data to create KPIs: Total Unit Price, Profit, Orders, YoY/MoM Trends, Top/Bottom Products & Customers
- Detected 43% sales drop in 2020 vs. 2019, enabling leadership to implement targeted recovery strategies

SKILLS

- **Languages:** Python (Pandas, NumPy, PySpark), SQL (T-SQL, PL/SQL, MySQL, PostgreSQL), Bash
- **Data Engineering:** ETL/ELT, Apache Spark, Kafka, Hadoop (HDFS, Hive), AWS Glue, SSIS, Airflow
- **Cloud (AWS):** S3, Lambda, Kinesis, SNS, EC2, RDS, IAM, VPC, CloudWatch, Code Pipeline, Code Build
- **Databases:** SQL Server, MySQL, PostgreSQL, MongoDB
- **DevOps & BI:** Docker, Git/GitHub, CI/CD, Jenkins | Power BI (DAX, Power Query, RLS), Tableau, Excel (P.P)
- **Certifications:** AWS Cloud Practitioner (2023), Power BI DAX Modeling (2023), MySQL 5.6 Developer

EDUCATION

University of New Haven, Connecticut, USA

Master of Science: Data Science (GPA-3.49)

May 2025

Sreenidhi Institute of Science and Technology, Hyderabad, India

Bachelor of Technology: Mechanical Engineering (GPA-7.9)

May 2021