# KHAJA MUJAHIDDIN MOHAMMED

Connecticut USA | +1 347-736-5812 | muhammedshoaib835@gmail.com | LinkedIn | Portfolio | Github

Entry-Level Data Analyst with a master's in data science (University of New Haven, May 2025). Hands-on experience building real-time AWS data pipelines (Kinesis, Glue, Lambda), Power BI dashboards with DAX, and Python automation (pandas, SQLAlchemy) during a 4-month US internship and full-time analyst role. Delivered 8–10 production dashboards used by 50+ stakeholders and automated reporting workflows that saved ~8 hours/week. I am skilled in SQL, Python, Power BI, ETL processes, and cloud analytics. OPT-eligible (no sponsorship required for 3 years). Actively seeking Junior Data Analyst / BI Analyst roles in CT, NY, NJ, or Remote.

## EDUCATION

**University of New Haven, Connecticut, USA**
**Master of Science:** Data Science (GPA: 3.29)                                        (Aug 2023 – May 2025)
**Sreenidhi Institute of Science and Technology, Hyderabad, India**
**Bachelor of Technology:** Mechanical Engineering (GPA: 7.9)                     (Aug 2017 – May 2021)

## TECHNICAL SKILLS

- **Languages:** Python (pandas, NumPy, scikit-learn), SQL (MySQL, PostgreSQL)
- **Visualization & BI:** Power BI (DAX, Power Query, RLS), Excel (Power Pivot, PivotTables)
- **Cloud & Tools:** AWS (S3, Glue, Lambda, Kinesis – internship projects), Git, Docker basics
- **Databases:** MySQL, PostgreSQL, MongoDB
- **Certifications:** AWS Certified Cloud Practitioner (2023), Power BI Data Modeling with DAX (2023)

## RELEVANT EXPERIENCE

**Data Scientist - Intern | Innover Global INC| Aurora, IL**                         (Aug 2025 – Present)

- Building prototype real-time streaming pipeline (AWS Kinesis → Glue → Lambda → S3) processing 100K+ events/day
- Developed customer segmentation model using K-Means clustering + PCA in PySpark for behavioral insights
- Designed interactive Power BI dashboards with DAX measures, slicers, and drill-throughs for internal stakeholders
- Implemented lifecycle policies and CloudWatch monitoring to optimize storage costs

**Jr. Data Analyst | SRIK Consulting Services |(Started as Intern)|Hyderabad, India**      (Sep2020 – July 2023)

- Created 8–10 production Power BI dashboards for finance & HR teams (50+ users) with Row-Level Security
- Automated monthly dataset refreshes using Python (pandas + SQL Alchemy + Windows Task Scheduler), saving ~8 hours/analyst/week
- Cleaned and transformed data from MySQL, SQL Server, Excel, and REST APIs; built star-schema models
- Wrote Flask APIs to serve real-time KPIs to existing HR/finance web portals

## ACADEMIC AND PERSONAL PROJECTS

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

- End-to-end pipeline: AWS S3 → Glue → PySpark → K-Means → Power BI
- Feature engineering + PCA dimensionality reduction; automated retraining with Python scheduler

**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

**Titanic Survival Prediction (Python)| DAX, Power Query, Advanced Excel, Data Modeling, Data Warehousing**

- Full EDA + feature engineering in pandas; Random Forest model (82% accuracy)
- Jupyter notebook with detailed markdown, visualizations, and model interpretation

**Microsoft Fabric Real-Time Intelligence**

- Built streaming analytics workflow using KQL queries and live Power BI dashboard
- Integrated Eventstream → Lakehouse → real-time visual update

## ADDITIONAL

- OPT eligible (STEM extension) – no sponsorship required for 3 years
- Available immediately | Open to relocation (USA, Remote)