

Fariha Shah

(253) 392-9379 | Fariha_shah@outlook.com | Seattle, WA | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

SUMMARY

Data Engineer with 4+ years of building reliable, scalable data pipelines and analytics platforms. Strengths in Python, SQL, and Spark on AWS to automate data flows, improve data quality, and ship stakeholder-ready datasets and dashboards. Seeking roles where I can own end-to-end data engineering work that turns raw data into a measurable business impact.

EDUCATION

Seattle University, Washington	June 2026
Master of Science in Data Science, GPA: 3.66/4.0	
Dawood University of Engineering and Technology, Pakistan	Dec 2019
Bachelor of Engineering in Petroleum and Natural Gas, GPA: 3.62/4.0	

PROFESSIONAL EXPERIENCE

Data & Operations Assistant Mission Integration, Seattle University Seattle, WA	Nov 2025 – Present
• Automated event workflows using Power Automate, triggering email invitations, confirmations, calendar invites, and response logging upon form submission, significantly reducing manual coordination efforts from daily to hourly .	
• Managed and updated intranet and extranet platforms, ensuring accurate mission integration values presentation, improved accessibility, and consistent content delivery	
• Developed and managed a centralized SharePoint event database to store, validate, and maintain participation data, improving data accuracy and accessibility	
• Integrated SharePoint data with Tableau and built interactive dashboards to analyze event attendance trends, engagement levels, and participation patterns.	
• Collaborated with cross-functional stakeholders to define reporting requirements and align analytics outputs with departmental and organizational goals	
Data Engineer II FLOWHCM Karachi, Pakistan	July 2020 – Aug 2024
• Resolved a critical HR data pipeline failure by building a flexible date parser, adding schema checks, and implementing idempotent upserts to handle inconsistent inputs. Restored reports within one day, improved pipeline reliability, and saved the client hours of manual data work weekly	
• Built a multi-view HR KPI download system integrating 50+ data sources (SQL, REST APIs, custom connectors) with schema validation and deduplication, reducing manual analysis time from 2 weeks to 7–10 days and enhancing dashboard insights	
• Optimized 12 under-performing pipelines and BI queries, fixed join strategies, partitioning, cut ETL runtimes by 40% and reduced computational cost 25%	
• Built a searchable catalog for 100+ datasets and instrumented pipelines with freshness & schema checks, achieving near-perfect on-time loads	
Data Engineer Intern Pakistan Petroleum Limited Karachi, Pakistan	Jan 2020–June 2020
• Built PySpark/Python ETL to ingest time-series telemetry from industrial/field sensors (SCADA/data historian) and business/operations transactions (ERP/finance) into a curated store; eliminated a third-party tool and saved \$85K+/yr.	
• Developed Tableau dashboards for production, downtime, and throughput, reducing manual reporting by 10 hrs/week and improving data freshness from 7 days to 12 hours	
• Added schema validation, deduplication, and job monitoring to harden pipelines; improved reliability of daily summaries for wells and compressor stations.	

SKILLS

Basic: Tableau, Power BI, AWS (Lambda, S3, EMR, Dynamo), Docker, Git, Bash, UNIX, NLP

Intermediate: A/B Testing, Predictive Modeling, Machine Learning, Forecasting, scikit-learn, TensorFlow, PyTorch, Apache Spark, Hadoop, Kafka, PySpark

Proficient: Python, SQL (Postgres, MySQL), R, C#, Angular, .NET, REST APIs, ETL Pipelines, Powerautomate

PROJECTS

Capstone - Costco Wholesale - Retail Price Recommendation Engine	Jun. 2026
Jun. 2026	

- Designed and prototyped a price recommendation engine for fast-moving consumer products by estimating price elasticity of demand using regression-based models and simulated sales data
- Analyzed price elasticity, historical sales trends, and competitor pricing signals to identify products with high price sensitivity versus margin flexibility
- Translated model outputs into interpretable pricing insights aligned with competitive dynamics, margin constraints, and member-value considerations
- Developed an optimization framework to recommend revenue-maximizing prices under realistic business constraints, and proposed an A/B testing strategy to validate pricing changes
- The expected benefits include stronger margin protection, greater confidence in our pricing decisions, and lasting trust from our members, all supported by data-driven pricing strategies

Educational Data Pipeline: ETL and Reporting for Academic Metrics | Python, MySQL, Tableau

- Enhanced university database to visualize student and institution performance over time where previously unable.
- Integrated UOW APIs through Python-based ETL pipelines, with schema/null checks and curated tables; authored SQL (window/CTEs) for KPIs. **Result:** unlocked 25 KPIs and **reduced report prep by 70%.**

Flood Risk Prediction using ML & SARIMA | Asia-wide Climate Data | Python, Scikit-learn, Climate Datasets, Machine Learning

- Built dashboard to identify at risk flood areas in Asia and share them with the larger community for awareness
- Built a predictive framework SARIMA and regression models to predict flooding
- Merged multi-source data (historical flood records, GDP, CO2, satellite zones) and created a geospatial risk map using GeoPandas .