

# SHISHIR NANGA

■ Reston, VA 20190 ■ 571-866-0967 ■ [shishir.nanga@gmail.com](mailto:shishir.nanga@gmail.com) ■ [www.linkedin.com/in/shishir-nanga](https://www.linkedin.com/in/shishir-nanga) ■ [www.github.com/shishirnanga](https://www.github.com/shishirnanga)

Data Scientist & Engineer skilled in statistical modeling, data engineering, and software development. Proficient in Python, R, SQL, and cloud platforms with experience building ETL pipelines, APIs, and AI-powered applications. Published contributor in applied data science. Adept at translating raw data into scalable systems and actionable insights that drive business, product, and social impact.

## Core Skills

- **Languages Programming:** Python, R, SQL, Java, Git, APIs (Flask/FastAPI)
- **Data Science:** A/B Testing, Causal Inference, Regression, Classification, Clustering, Forecasting, NLP
- **Data Engineering:** ETL/ELT (Airflow, DBT), Kafka, Spark, Delta Lake, Data Warehousing (Snowflake, BigQuery)
- **Visualization:** Tableau, Power BI, ggplot2, Shiny, Plotly, Streamlit
- **Cloud & DevOps:** AWS (S3, Glue, Redshift, Lambda), Azure (Data Factory, Databricks), Docker, CI/CD
- **Certifications:** IBM Data Science, Google Data Analytics, Meta Database Engineering

## Professional Experience

### Rockwoods Inc

Reston, VA

Software and Data Engineer (Part Time)

January 2024 – Current

- Designed Python/SQL/Airflow pipelines processing 20M+ records daily for risk monitoring, fraud detection, and analytics.
- Built REST APIs with FastAPI/Flask to serve processed data into dashboards and reporting systems.
- Deployed workflows on AWS (S3, Glue, Redshift) with Dockerized services for portability and reliability.
- Improved pipeline latency by 25% by optimizing query performance and introducing task parallelization.
- Partnered with analysts to translate business needs into technical solutions, improving reporting accuracy.

### Alpha Omega Integration

Tysons, VA

Cloud Technical Intern

May 2023 – August 2023

- Collected and structured 50K+ online reviews for sentiment analysis using Python (NLTK) and R (tidytext).
- Built sentiment classification models that improved accuracy by 15% after data cleaning and feature engineering.
- Created dashboards (Plotly, ggplot2) visualizing customer sentiment and product perception for leadership.
- Assisted frontend team with React-based UI components, linking visualizations to backend APIs.
- Contributed to internal data documentation and process automation, cutting manual review time by 20%.

### KPMG

Hyderabad, India

Data Analyst Intern

July 2022 – August 2022

- Analyzed large datasets in SQL, R, and Excel to support tax reporting for luxury goods sector clients.
- Automated recurring data checks and reconciliations, reducing manual effort by 30%.
- Built ad-hoc reports highlighting anomalies, enabling managers to flag compliance issues faster.
- Collaborated with cross-functional teams to present findings, gaining exposure to client deliverables and workflows.

## Projects

### AutoAdvisor – AI Product Strategy Assistant

- Built a Streamlit application integrating GPT-4, FAISS, and OCR for A/B test analysis, feedback clustering, and dashboard summarization.
- Engineered a modular RAG pipeline with APIs for data retrieval and GPT-based product recommendations.
- Designed multi-tab layout with clear workflows for experimentation, feedback, and dashboard insights.

### Clean Water Scarcity Monitor

- Designed Python/R ETL workflows to integrate WHO/UN/World Bank datasets on water access, GDP, and resources.
- Built Tableau + ggplot2 dashboards to uncover geographic disparities and risk zones.
- Enabled actionable insights for NGOs tackling school-level water scarcity.

### Real-Time Viewer Behavior Dashboard — Streaming Analytics

- Developed Kafka to PostgreSQL to Airflow pipelines processing 1M+ streaming events for CTR & dwell-time analytics.
- Designed real-time dashboards and APIs enabling content teams to optimize recommendations instantly.
- Applied R-based time-series and survival models to predict churn and engagement decay.

### Customer Segmentation & Churn Analytics – PenFed

- Engineered Power BI dashboards and SQL pipelines tracking deposit and churn trends across millions of members.
- Applied R clustering (k-means, hierarchical) to segment customers; retention strategy informed a 15% churn reduction.

## Education

### George Mason University

Fairfax, VA

BS, Computer Science, Dean's List Candidate

2020-2025

Relevant Coursework: Machine Learning, Data Mining, Statistics, Distributed Systems, Networking, Cloud Computing

## Publications and Writing

- AutoAdvisor: Automating experiment readouts and insight synthesis for product decisions — project write-up, 2025 (repository and methodology notes in GitHub)

## Leadership & Activities

- Vice President, Indian Student Association (GMU): Organized events, managed finances, grew membership 3x.
- District & State Basketball Captain: Led competitive teams, developing resilience & strategic leadership.