**Interview with a Data Scientist: A Glimpse into the World of Applied Analytics**

For this assignment, I interviewed **Umar Zakariyya Muhammad, Ph.D.,** a Geoscience Research Scientist with deep experience in data analysis, Python scripting, GIS applications, and machine learning. Currently working in the energy sector, he combines scientific expertise with advanced data science techniques to solve real-world problems related to subsurface exploration and CO₂ storage.

When asked about his typical day, Dr. Umar explained that it often begins with a review of ongoing model training processes and data pipelines. “I start the day by checking the results from my overnight runs whether that’s a machine learning model predicting subsurface lithofacies or a Python script cleaning large seismic dataset.” Much of his time is spent preparing and wrangling data, which he noted is one of the most time-consuming yet crucial aspects of his work. “You can’t build fair or accurate models without high-quality input.”

Contrary to the assumption that data scientists work only on algorithms, Umar is more of a **full-stack data scientist**. He explained, “I handle everything from data acquisition and preprocessing, to exploratory data analysis, model selection, deployment, and even result communication to stakeholders.” For deployment, he often uses cloud-based platforms and version control tools like GitHub. He emphasizes reproducibility and collaboration in his workflows.

One interesting takeaway from the conversation was the interdisciplinary nature of his role. “I constantly interact with geologists, reservoir engineers, and software developers. My job is to translate domain-specific problems into data science tasks and vice versa.” His background in geoscience has given him a unique advantage in understanding complex spatial-temporal data.

When asked about tools and languages, Umar mentioned Python as his primary language, along with libraries such as **Pandas**, **NumPy**, **Scikit-learn**, and **TensorFlow**. For visualization, he uses **Seaborn**, **Plotly**, and **GIS software** like QGIS. He also mentioned his recent exploration of fairness libraries like **Fairlearn**, particularly for detecting biases in predictive models.

Ethics and fairness, he noted, are becoming integral parts of his work. “It’s no longer enough to make accurate predictions. Stakeholders want to understand the fairness implications, especially in areas like environmental justice and sustainable resource allocation.”

Reflecting on the skills required for success in his role, Umar emphasized critical thinking, domain knowledge, communication, and continuous learning. “The field changes rapidly, so I dedicate time each week to learn something new, whether it's a new ML algorithm or a policy framework around data ethics.”

In conclusion, the interview highlighted the diverse and dynamic nature of a data scientist’s role. Rather than being narrowly specialized, professionals like Umar Zakariyya operate across the full data science stack combining programming, modeling, deployment, and stakeholder communication. The interview was insightful in revealing how data science can bridge the gap between technical innovation and domain-specific impact.

**Source:**  
Personal Interview with Dr. Umar Zakariyya Muhammad, August 2025.