

# **Terms of Reference (ToR) for Data Scientist to Support the Ghana National Households Registry**

## **1. Background**

The Ghana National Households Registry (GNHR) aims to create a comprehensive database of households to facilitate the effective targeting of social protection programs. The GNHR requires the expertise of a Data Scientist to enhance data collection, analysis, and management processes.

## **2. Objectives**

The primary objective of this assignment is to support the GNHR in improving the quality, accuracy, and usability of household data through advanced data science techniques.

## **3. Scope of Work**

The Data Scientist will be responsible for the following tasks:

- **Advanced Data Analytics:** Develop and implement advanced statistical models and machine learning algorithms to analyze data, predict trends, identify bottlenecks and develop effective solutions
- **Data Engineering:** Design and maintain robust data pipelines for efficient data coordination, collection, cleaning, and preprocessing.
- **Algorithm Development:** Create and optimize algorithms for data integration and de-duplication, ensuring seamless merging of registry data internally and with other data sources, including administrative data and detailed mobile call data.
- **Big Data Management:** Utilize big data technologies to handle large datasets and ensure data scalability and performance.
- **Innovative Solutions:** Research and develop innovative data solutions to enhance the GNHR's data capabilities, supporting continuous updates without the need for home visits and interoperability with other databases.
- **Data Validation and Reliability:** Ensure the reliability and validity of integrated data sources to maintain the accuracy of the registry.
- **Interdisciplinary Collaboration:** Work closely with other teams to implement data-driven solutions and support decision-making processes.

## **4. Deliverables**

- **Advanced Analytical Models:** Development and documentation of statistical models and machine learning algorithms.
- **Data Pipelines:** Robust data pipelines for data coordination, collection, cleaning, and preprocessing.
- **Integration Plan:** A comprehensive plan for integrating external data sources with the registry data and internally
- **Validation Reports:** Regular reports on the validation and reliability of integrated data sources.
- **Innovative Data Solutions:** Research findings and implementation of innovative data solutions.

## **5. Duration**

The assignment is expected to last for nine (9) months, with the possibility of extension based on performance and project needs.

## **6. Qualifications**

- Education: Minimum a Masters degree in Data Science, Statistics, Computer Science, or a related field.
- Experience: At least 5 years of experience in data science, with a focus on advanced analytics, machine learning, and big data technologies.
- Skills: Proficiency in programming languages (e.g., Python, R), big data technologies (e.g., Hadoop, Spark), and data visualization tools (e.g., Tableau). Experience with integrating and validating external data sources is highly desirable.

## **7. Reporting**

The Data Scientist will report to the GNHR Head through the GNHR data analyst/Statistical Officer.