**IMAT5314 Project Terms of Reference (ToR)**

**Student Name: Folashikemi Maryam Asani, Olaniyan**

**P-number: P2586104**

**Programme: MSc. Data Analytics**

**Email address: P2586104@my365.dmu.ac.uk**

**Project Title: A Comparative Evaluation of WHO Data Quality Assessment Tool and Unsupervised Anomaly Detection Algorithms for Detecting Data Completeness and Consistency**

**Supervisor: Dr Adebowale Owoseni**

# **Email:** adebowale.owoseni@dmu.ac.uk

# **Phone:** +44 116 366 4553

**Introduction**

Health data quality is indispensable in advancing effective policies and strategies for global health practices. Poor data quality is a considerable challenge to progress. This is even more so during periods of pandemics as containment measures rest heavily on the accuracy of health data. Hence, due to a couple of challenges, the quality of data could be distorted. Administrative errors and other technical difficulties could promote weak data quality. The effectiveness of the data review tool kit becomes critical in ensuring data quality and hence promote best practices in global health. This study gains its motivation from the forgoing. The WHO use of data review tool-kit sets the pace for trusted data quality and informs best practices. Efforts to examine the effectiveness of current tools is critical for the sustainability of global health practices. Thus the drive to evaluate these tools against possible more viable measures is justified.

**Aim/Objectives**

**Aim**

The study aims to identify and implement unsupervised anomaly detection techniques, methods, and strategies for improving data quality assessment in public health.

## **Research Objectives**

1. Evaluate the current methods and techniques used in assessing data quality in public health.
2. Suggest a suitable method for improving the assessment of data quality in public health.

**Hypothesis or** **Research Questions**

1. Do the methods and techniques used in DQA have the standard that can assess data quality in public health?
2. Is there a better method or technique of data quality assessment in public health?

**Resources**

Personal computer

Jupyter Notebook

PyCaret Library

Scikit Learn Library

DMU Library Resources

WHO Data Quality Assessment Toolkit

**Constraints**

Limited availability of public health data

No availability of date and timestamp data

Limited availability of anomaly detection libraries

**Sources of Information**

Data – Pepfar Site Performance data from the fiscal year 2016 to 2021 found in [data.pepfar.gov](https://s3.amazonaws.com/media.data.pepfar.gov/mer_downloadables/SitePerformanceFY2016-FY2021.zip)

FOLASHIKEMI MARYAM OLANIYAN 21 MAY 2021

**Student** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Date**\_\_\_\_\_\_\_\_



**Supervisor** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Date**\_\_\_\_\_\_\_\_

