**Tackling challenge of Teenager pregnancy using Data Science**

**Introduction**

Teenage pregnancy represents a significant challenge for the youth of Rwanda, as highlighted by UNFPA (United Nations Population Fund), which reported that 35,123 girls became pregnant in 2022. The consequences of teenage pregnancy are particularly evident in areas such as mental health, maternal mortality, and the prevalence of sexually transmitted diseases among young mothers. These issues contribute to a sense of hopelessness regarding the future, making teenage pregnancy a multifaceted problem that impacts not only the young mothers themselves but also their children, families, and the nation as a whole.

**Data methodology**

We used Demographic Health Survey which have been conducted by different years like 2000,2005,2010,2015 and 2020 dataset, and we used adolescent birth rate per country from UNFPA.

**Data science solution we developed**

We employed Data Handling Systems (DHS) to monitor the socio-economic factors associated with heads of households across all surveys, which are likely to correlate with teenage pregnancy. Additionally, we utilized data visualization techniques for each survey, as well as for the adolescent birth rate by country, to analyze trends in teenage pregnancy. All socio-economic variables and their respective descriptions are detailed in the document we have previously submitted to you.

We have created a solution that serves as a guide for policymakers to evaluate the measures they can implement to reduce teenage pregnancy. This solution employs a model known as the "logistic regression model," where the qualitative dependent variable is the occurrence of pregnancy, represented by binary values 1 for yes and 0 for no. The independent variables consist of various socioeconomic factors related to the head of the household, which may be associated with instances of teenage pregnancy.

**Limitation**

The inadequacy of data concerning teenage pregnancies within the Demographic and Health Surveys (DHS) time series renders it insufficient for predictive analysis. Consequently, the limited data available in the time series impedes the ability to generate forecasts based on the DHS, which ultimately leads to challenges in making accurate predictions in this context.

**Recommendations**

In order to facilitate analysis, we are requesting that the National Institute of Statistics of Rwanda (NISR) provide data specifically related to teenage pregnancy. This information will enable young professionals in the field of statistics to gain more fundamental access to relevant datasets. Furthermore, the strategies proposed by policymakers to address teenage pregnancy should also focus on implementing measures aimed at supporting young females who are at a heightened risk of experiencing pregnancy.