ITU Jupyter Notebooks – A practical tool for analysing mobile phone data for measuring the information society

Mobile Network Operator (MNO) data offers an opportunity to enhance data quality, timeliness, frequency, and granularity of official Information and Communication Technology (ICT) statistics. This initiative by the International Telecommunication Union (ITU) aims to help countries in using mobile phone big data (MPD) to address data gaps and provide detailed and timely sub-national data crucial for informed policymaking.

To support this initiative, ITU has developed a series of Jupyter Notebooks written in PySpark. These Notebooks, based on the ITU Methodological Guide on using MPD for Information Society Statistics, serve as a comprehensive and practical guide for statisticians and data scientists on processing Call Detail Records (CDRs) and Internet Protocol Detail Records (IPDRs). Specifically, the Notebooks was designed to calculate SDG indicator 17.8.1, Percentage of the population using the Internet, normally collected using official household surveys. Due to lack of resources in many developing countries, these data are not available or not collected. The use of MPD will not only address data gaps but will also provide a more disaggregated and timely data. The series includes multiple Notebooks, each focusing on a specific aspect of data processing:

- Setup Instructions: Provides detailed instructions on setting up the technical environment for data processing using PySpark. It includes setup guidelines for both Windows and Linux/Mac systems and instructions for using SageMaker Studio Lab, an AWS cloud environment.
- Processing of Raw Mobile Phone Data: Elaborates on the necessary structure and fields required for processing mobile phone data. This Notebook covers basic checks, removal of duplicate records, exploratory data analysis, and filtering out irrelevant data such as records from robots or tourists.
- Quality Assurance: A critical step to ensure the quality of MPD before generating reliable statistics. This Notebook details multiple quality gates, starting with an initial assessment of raw data to ensure its consistency and validity. Specific metrics such as the percentage of null values and the consistency of subscriber counts are used to detect irregularities or discrepancies.
- Home Algorithm: Determines a subscriber's 'home' location using mobile phone data. This is vital for mapping mobile phone data with reference data like Local Administrative Units (LAU) and population estimates. The Notebook outlines a multi-step logic to identify the most frequent cell locations during anchor times, assigning a 'home cell' to each subscriber.
- Indicator Calculation: Focuses on calculating SDG indicator 17.8.1, the proportion of individuals using the Internet. This is achieved by determining the number of subscribers with at least one IPDR event and dividing it by the total target population. The Notebook also allows for the analysis of Internet usage by technology type and geography, providing a nuanced understanding of digital accessibility.

Examples from pilot projects in countries like Brazil and Indonesia highlight the alignment of MPD results with traditional survey data, demonstrating the reliability of MPD and potential source to complement traditional data sources.

By utilizing ITU's comprehensive guidelines and practical tools, countries can strengthen their ICT statistics frameworks, leading to more informed decision-making and better policy outcomes. The ITU Jupyter Notebooks provide a solid foundation for national stakeholders to implement MPD projects, with technical support available to facilitate successful execution.

For more information, please contact:

Esperanza Magpantay
Senior Statistician
magpantay[at]itu.int

Fredrik Eriksson
Data Scientist
fredrik.eriksson[at]itu.int