



Enrichment of informal sector data through collaboration between agencies

Ahmad Najmi ARIFFIN¹ ; Noradilah ADNAN¹

¹ Core Team Big Data Analytics, Department of Statistics, Malaysia

Abstract:

The informal sector plays a key role in the development of the Malaysia economy. In Malaysia current practice, Informal Sector Survey is conducted by the Department of Statistics Malaysia (DOSM) using households' approach to obtain data on the structure of employment in the informal sector. From 2020, DOSM took the initiative to enrich the available data through cooperation among agencies in line with current technological developments. This initiative considers extending the informal sector definition with the involvement of agricultural activities. Although this approach describes valuable data relevant to the informal sector, there are also some challenges encountered in coordinating such data. This study addressed the challenges faced throughout the data integration process and possible solutions to improve the data quality of 11 agencies. Among the challenges faced are different raw data formats from different agencies and incomplete data. The element of data management workflow has been considered in the development of database architecture which includes data profiling, data standardization, data clustering, data cleansing, text matching and data integration. This modern architecture offers fast analytics to explore the data representativeness from DOSM (43.2%) and multiple agencies (56.8%) data sources. To ensure the data is internally consistent, and every data type has the same content and structure, the standardize template are created for operational used. Meanwhile, more than 71.0% incomplete data based 89 variables were resolved by matching the data using entrepreneur or business information at different agencies as well as cross-referencing data from online sales websites and social media platforms. Enrichment of informal sector data through cooperation among agencies provided clear sector informal criteria to help the government plan the appropriate policies accordingly.

Keywords:

informal sector; data consolidation; data enrichment; database; data integration