


ARE YOU A DATA SCIENTIST? OR CARE ABOUT DATA QUALITY?

Participate in the 2nd

DATA-THON

Your contribution can strengthen India's data ecosystem.

The **DATA-THON** is an effort to crowdsource insights on the quality of data (survey and program data) using innovative data analytic techniques.

ELIGIBILITY

Data cruncher across fields are invited. Participants include but not limited to:

- Researchers, Statisticians, Epidemiologists, Economists, Data scientists, Big data analysts, Technology and Health Experts

ENROLMENT

Please send your entries in either of the themes to: dataqthon@popcouncil.org

There is no fee for your entry.

Entries must be submitted on or before 15th November, 2019

THEMES OF DATA-THON

1. USE OF MACHINE LEARNING TECHNIQUES TO IDENTIFY DATA QUALITY ISSUES

- In the 1st theme, you can use any publicly available administrative, digital online, and/or survey data.
- Entries beyond identification of missing values or omission errors in the data will be highly encouraged.
- Submit a note (max 2/3 pages) describing data, issue identified, and algorithm used is essential. Also, submit the R/Python/any other machine learning script along with the note.

2. DATA ISSUES IN COMPARISON OF ADMINISTRATIVE AND SURVEY DATA

- In the 2nd theme, identification of inconsistencies (or) insights about challenges, in comparison of publicly available administrative and survey data is essential.
- Submit a note (max 2/3 pages) describing rationale, data sources, methodology, results, issue identified, and possible data quality solutions.
- Use of scientific methodology with sufficient cases for each of the data sources is highly encouraged.

Example data sources: District Level Household Surveys, Annual Health Surveys, Demographic and Health Surveys, HIV Surveillance Surveys, National Sample Surveys, India Human Development Surveys, Health Management Information System Data, and any other publicly available data sources. Some of these data are available at: <https://data.gov.in/> or Harvard Dataverse: <https://dataverse.harvard.edu/>

REWARDS*

Each of the four best entries (two from each theme) will get a scholarship for: Training (or) Conference attendance (or) Course fee (or) Paper publication fee (or) Trending analytic gadget, up to the value of ₹100,000.

TIMELINE

Start Date: 5th October, 2019
End Date: 15th November, 2019

For any queries, please write to dataqthon@popcouncil.org