

AI enhanced Application for Automated Data Preparation, Estimation and Report Writing

1. Track:

Data Processing and Analysis

2. Description

Official statistical agencies often work with diverse survey datasets that require extensive preprocessing before analysis. Manual workflows for cleaning and weighting are laborious and error-prone, delaying estimates and reducing reproducibility. An automated, low-code tool can accelerate data readiness and ensure methodological consistency. An AI- augmented web application designed to streamline survey data processing and Analysis results in substantial time and resource saving. Participants will develop configurable modules to clean raw survey inputs, handling missing data, outliers, and rule violations and apply design weights for estimation of population parameters, and generate standardized output reports/templates for official statistical releases.

Participants are to build a prototype of application that ingests raw survey files (CSV/Excel), performs cleaning (imputation, outlier & rule-based checks) through frontier technologies, integrates survey weights, and produces final estimates with margins of error alongside PDF/HTML reports. All functionality should be configurable via a user-friendly interface.

3. Expected Outcomes/Solutions

Participants should build a prototype of an application that:

- Ingests raw survey files (CSV/Excel)
- Performs cleaning (imputation, outlier & rule-based checks)
- Applies design weights
- Produces final estimates with margins of error
- Generates standardized output reports in PDF/HTML format
- Offers a user-friendly, configurable interface

4. Relevance to National Priorities or Ongoing MoSPI Initiatives

This use case supports MoSPI's objective of improving data quality and efficiency through automation and AI integration in data processing. It enhances reproducibility and reduces delays in producing official statistics.



5. Background Resources or Datasets (if available)

- Gold-standard benchmark datasets for accuracy validation
- PDF report templates
- Documentation on survey-weight methodology

6. Key Features Required

Data Input & Configuration:

- CSV/Excel upload
- Schema mapping via UI or JSON config

Cleaning Modules:

- Missing-value imputation (mean, median, KNN)
- Outlier detection (IQR, Z-score, winsorization)
- Rule-based validation (consistency, skip-patterns)

Weight Application:

- Apply design weights
- Compute weighted/unweighted summaries and margins of error

Report Generation:

- Auto-generate reports using templates
- Include workflow logs, diagnostics, and visualizations

User Guidance:

- Tooltips, inline explanations, error-checking alerts

7. Bonus Features / Future Scope

Innovation beyond requirements such as dashboards or audit trails will be given extra credit.

8. Impact Potential

The solution will accelerate survey readiness, reduce errors, and ensure methodological consistency which will ultimately strengthen MoSPI's capacity to deliver high-quality official statistics with speed and accuracy.