Team 7

Project Title: Local Governments Survey (Harmonizing ARISE City Survey + CivicPulse Kansas Sample)

Course Number:CS896
Team Names: Anudeepthi Chirumamilla- Y828R524
Rupali Avhad - d798z925
Teja Sree Mukkapati-Q389W584

1. Introduction

Objective: This project aims to enable comparable measurement of workforce challenges, fiscal condition, and capacity/service preparedness indicators across local governments by harmonizing overlapping survey constructs across two instruments: Civic Pulse (Kansas sample from a national survey) and ARISE (Kansas city survey).

Why harmonize? Although the ideas measured by both surveys are similar, their question formats, coding systems, and item coverage differ. In order to define harmonized variables for analysis and align constructs, a crosswalk is necessary.

2. Data Source
The dataset used in this analysis is:
Rachel_Krause_ARISEData_Combined General Dataset.csv

This dataset contains survey responses from local governments and includes:

1. Workforce challenge (Q8)
2. Fiscal condition (Q119)
3. Grant capacity (Q5)
4. Preparedness program indicators (Q9 series)
5. Vulnerable population support indicators (Q10 series)

Each row represents a respondent, and each column represents a survey item.

3. Final Crosswalk Table(deliverable)
In order to ensure conceptual consistency and facilitate cross-source comparison, the Final Crosswalk Table was created to standardize important variables across the ARISE and CivicPulse datasets. In order to improve the dependability of cross-dataset interpretations, the harmonization procedure concentrated on bringing variable definitions, response scales, and measurement goals into alignment.
Four primary concepts were harmonized:

Concept	Harmonization Summary	Purpose / Interpretation
1.Workforce Challenge	1–5 scale retained (higher = more challenging) across ARISE Q8 and CivicPulse workforce challenge question.	measures staffing capacity and service strain.
2.Fiscal Condition:	1–5 scale (Very Poor–Excellent) harmonized between ARISE Q119 and CivicPulse fiscal condition rating.	Indicator of local government financial health.
3.Emergency Preparedness Capacity	Converted preparedness items to binary indicators (1 = present, 0 = not present); summed for total capacity score.	Measures disaster readiness and response capability.
4.Protection of Vulnerable Populations	Aligned actions for elderly, low-income, and disabled groups with CivicPulse vulnerability items to form support score.	Measures social resilience and equity.

4. Data Cleaning Methodology (Implemented in Code)

Missing Value Handling

The ARISE dataset contained missing values represented in multiple inconsistent formats.

The cleaning pipeline standardized missing data by converting:

- Blank cells and whitespace → NaN
- Textual missing indicators (e.g., NA, N/A, Don't know, Refused) → NaN
- Numeric coded missing values (e.g., 99, 999, -99) → NaN
- This step ensures that all missing values are uniformly treated during analysis.

Additionally, a missingness report was generated to identify:
- Columns with highest missing percentage
- Rows with high levels of non-response

This improves transparency and reproducibility of the cleaning process.

5. Harmonized Variable Construction

Core Harmonized Variables (Likert 1–5)

Based on the crosswalk and cleaned dataset, the following harmonized variables were created:

workforce_challenge_1to5 ← Derived from Q8
fiscal_condition_1to5 ← Derived from Q119
grant_capacity_1to5 ← Derived from Q5
survey_source = "ARISE"

These variables preserve the original 1–5 scale and maintain directional consistency (higher values indicate stronger magnitude of the construct).

Preparedness Capacity Score (Q9 Composite)

Preparedness capacity was operationalized using Q9 program indicators (e.g., early warning systems, evacuation plans, backup infrastructure).

Method used in code:
- Automatically detected all columns starting with "Q9_"
- Cleaned values to ensure valid binary representation (0/1)
- Preserved missing values (NaN) instead of converting them to 0

Computed:

preparedness_score = row-wise sum of Q9 items (min_count=1)

Vulnerable Population Support Score (Q10 Composite) Support for vulnerable groups (e.g., elderly, disabled, low-income populations) was captured using Q10 indicators.

Method used in code:
- Identified all Q10 columns dynamically
- Converted responses to numeric and preserved missing values

Computed:
- vulnerable_support_score = row-wise sum of Q10 items

- This produces a robust composite index representing the breadth of vulnerable population support initiatives.

6. Correlation and Diagnostic Analysis
Correlation analysis was conducted as an optional exploratory diagnostic step to examine relationships among:
- Workforce challenges
- Fiscal condition
- Grant capacity
- Preparedness score
- Vulnerable support score

This step was included for validation and exploratory insight, but it is not the primary focus of the data cleaning deliverable.

7. Limitations
CivicPulse response dataset is not currently available; therefore cross-survey harmonized comparison cannot yet be performed.
Survey items differ in wording and scope, requiring careful harmonization through the crosswalk framework.
Composite indices (Q9/Q10) depend on self-reported program availability and may contain inherent reporting bias.

8. Conclusion
This submission successfully implements a reproducible ARISE data cleaning and harmonization pipeline aligned with the finalized crosswalk table. The workflow standardizes missing values, corrects inconsistent entries, and constructs theoretically grounded composite indices (preparedness and vulnerable support).