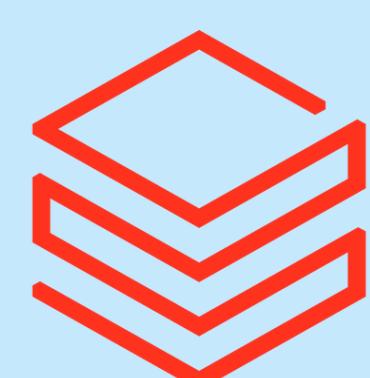


# Humanitarian Data Challenge



## Instructions Packet





Welcome to CMU Data Science Club's datathon in collaboration with the UN and Databricks! All important competition info has been compiled into this packet. Have fun, learn lots, and good luck!

## Challenge Prompt

We provide two challenges. You may **address either or both**.

1. **Smart Beneficiary Targeting Validation** - Analyze the relationship between project budgets and population data to flag statistical outliers in cost-per-beneficiary ratios. Provide a standardized framework which enables the UN to evaluate efficiency across different humanitarian clusters.
2. **Geo-Insight Challenge** - Integrate crisis severity indices with global funding data to determine mismatch(es) between humanitarian need and actual resource allocation, highlighting “forgotten” or “underserved” crises.

## Submission Details

1. Your project code as either:
  - a. Python Notebook (.ipynb) or
  - b. R markdown file (.rmd)
2. Video presentation (5 minutes max)

Submit your final project by submitting both materials to the [Datathon 2026 Submission Form](#) by  
**Sunday, January 25<sup>th</sup>, at 1pm**



# Databricks Signup

Use Databricks to complete your notebook and perform your analysis. Access Databricks through the following link and following the instructions provided:

[Databricks Signup Link.](#)

# Starter Repository

This repository contains the provided UN **datasets**, a starter notebook, and instructions to get started. Setup instructions and schemas for datasets provided are contained in the README.

The starter notebook is *optional* to use. Additionally, your team **must** use **at least one** of the datasets provided. You are also encouraged to use outside datasets as long as they are credibly sourced and cited.

[Starter Repository Link.](#)

# Evaluation

Exploration	Modeling	Impact
<p><b>Descriptive Statistics and Visualization</b></p> <p>Comprehensive summary statistics, varied visualizations, and clear insights extracted from the data.</p>	<p><b>Model Selection and Justification</b></p> <p>Well-justified model choice based on data and problem characteristics, including discussion of alternatives.</p>	<p><b>Humanitarian Relevance and Alignment</b></p> <p>Clear articulation of how the model and insights contribute to UN humanitarian goals; solutions should address specific challenges, provide actionable signals for aid officers, and align closely with the datathon's theme.</p>
<p><b>Data Understanding and Feature Exploration</b></p> <p>In-depth exploration of the relationships in the data or identification of key drivers of trends, justified by data insights.</p>	<p><b>Explainability and Interpretability</b></p> <p>The model is explainable, with clear explanations of how specific projects were flagged as outliers or how specific regions were identified as overlooked.</p>	<p><b>Potential for Real-World Application</b></p> <p>A clear, well-reasoned plan for how UN field teams could implement the tool or results, with thoughtful consideration of stakeholder needs, technical feasibility in low-resource environments, and potential barriers to adoption.</p>
<p><b>Significance of Data in Context</b></p> <p>Demonstrate an understanding of what the data represents in a real-world crisis, such as identifying where missing information might change the conclusion, and showing how your analysis makes aid distribution more transparent.</p>	<p><b>Model Accuracy and Performance</b></p> <p>The model is useful, the performance is measured and interpreted, and final analyses are contextualized using the performance. Model limitations are explained.</p>	