



National University of Computer and Emerging Sciences (FAST–NUCES)

Department of Data Science

Data Analysis and Visualization Semester Project

1 Groups

- **Each group must contain exactly 2 students.** No solo submissions and no groups larger than two will be accepted.
- Register your group members on the google sheet before submission.

2 Deadline

- **Final submission deadline: 30 November 2025 at 11:59 PM.**
- Submissions must be uploaded to the course Google classroom by the above deadline.

3 Report Submission

The PDF report should include:

- The report should be in IEEE standard format.
- Title, group members, and a one-paragraph abstract.
- Problem statement and objectives.
- Data sources and preprocessing steps.
- Methods and implementation details (algorithms, visualization tools, libraries).
- Key results, visualizations, and interpretation.
- Limitations and future work.
- Appendix with additional plots, tables, or code snippets.
- List of references(if used any)

4 Late Policy

- Late submissions will not be accepted.
- If you anticipate problems before the deadline, contact the instructor *before* the deadline with evidence. Last-minute requests will generally not be accepted.

5 Plagiarism Policy

- Plagiarism or copying from other groups (or online sources) is strictly prohibited and will result in disciplinary action. Cite any external code, libraries, or examples you use.
- Each group must clearly state which parts of the work were written by which member in `authors.txt`.
- The instructor reserves the right to run code similarity checks on submissions.
- Use of public libraries and example code is allowed, but you must:
 - (a) Clearly cite the source in the report.
 - (b) Indicate exactly which files/sections are reused or adapted.
 - (c) Ensure final submission is coherent and all members understand and can explain the work.

6 Final Notes

- Remember: Groups must be exactly 2 people, and the submission deadline is 30 November 2025, 11:59 PM (Asia/Karachi).
- Keep backups of your code and final submission. Good luck!

Resilience Under Pressure:

A Data-Driven Framework for Global Disaster Risk, Response, and Recovery Analytics

Scenario

In recent decades, the world has faced a rising wave of natural and man-made disasters from catastrophic floods and wildfires to pandemics and industrial accidents. Each disaster exposes deep vulnerabilities in human systems: governance failures, economic fragility, unequal recovery, and infrastructure weaknesses.

To move beyond reactive response, global organizations are turning to data intelligence systems that can integrate historical data, socio-economic indicators, and environmental information to reveal how societies prepare for, absorb, and recover from disasters.

The **Global Disaster and Humanitarian Response Agency (GDHRA)** has commissioned your team (*Pivotal Data Systems*) to design a prototype analytical platform capable of quantifying and visualizing resilience at national and regional scales. This system will help decision-makers identify which countries are most vulnerable, how fast they recover, and what socio-economic or environmental factors drive that recovery.

However, GDHRA provides no fixed formulas or dashboards, only large, unstructured, publicly available data. It is your job to define what “resilience” means computationally, model it quantitatively, and tell the story visually through Tableau.

Analytical Context

Disaster resilience is multi-dimensional it cannot be directly measured, only inferred. Your task is to build derived indicators using data-driven reasoning. Combine datasets across multiple domains and develop your own computational representation of:

- **Disaster Exposure (E):** Frequency and severity of disasters in a region.
- **Vulnerability (V):** Human and economic losses relative to population or GDP.
- **Adaptive Capacity (A):** Socio-economic and institutional strength for recovery.

You are expected to construct and justify quantitative indices using such relationships.

1. Disaster Impact Index (DII)

$$DII = \frac{(F + A_{pop})}{GDP_{pc}} \times S \quad (1)$$

Where:

- F = fatalities per million people per year
- A_{pop} = affected population (% of total population)
- GDP_{pc} = GDP per capita
- S = severity weight (based on disaster type)

This metric may represent how disruptive disasters are relative to a country's wealth and size.

2. Resilience Recovery Score (RRS)

$$RRS = \frac{GDP_{growth_{post}} - GDP_{growth_{pre}}}{T_{recovery}} + \frac{HDI + Gov_{index}}{2} \quad (2)$$

Where:

- $GDP_{growth_{post}}$ = GDP growth rate after disaster years
- $GDP_{growth_{pre}}$ = GDP growth rate before disaster
- $T_{recovery}$ = years to reach pre-disaster output
- HDI = Human Development Index
- Gov_{index} = governance or institutional stability score

3. Composite Resilience Index (CRI)

$$CRI = \frac{A}{E \times V} \quad (3)$$

Where:

- A = adaptive capacity (e.g., infrastructure, literacy, health)
- E = exposure (disaster frequency \times intensity)
- V = vulnerability (impact per capita or GDP)

Higher CRI indicates greater resilience, the ability to sustain or recover quickly from disaster events.

Note: These are conceptual models. You are encouraged to modify or expand them by integrating other variables such as CO₂ emissions, urban density, aid flows, or inequality indices.

Computational and Analytical Requirements

R1. Data Fusion: Integrate at least three open datasets. Possible sources include:

- EM-DAT Disaster Database: <https://www.emdat.be/>
- World Bank Data: <https://data.worldbank.org/>

- UNDP Human Development Reports: <https://hdr.undp.org/data-center>
- NASA Earth Data (Climate): <https://earthdata.nasa.gov/>
- Our World in Data: <https://ourworldindata.org/natural-disasters>

Align datasets on country–year pairs. Handle missing values analytically.

R2. Feature Engineering: Derive variables such as:

- Annualized disaster frequency (per country)
- Average economic loss per event (normalized by GDP)
- Recovery rate (based on GDP rebound)
- Infrastructure exposure (urbanization × hazard intensity)
- Human cost ratio (fatalities per 100k population)

R3. Model Formulation: Implement one or more derived indices (DII, RRS, CRI, or your own) using appropriate formulae. Clearly explain the mathematical logic and justification.

R4. Comparative Analysis: Your Tableau dashboard must enable exploration across:

- Time (temporal evolution of resilience)
- Geography (cross-country or regional comparison)
- Disaster types (earthquake, flood, drought, etc.)
- Socio-economic groups (developed vs developing countries)

R5. Analytical Storytelling: Insights must emerge visually. Your dashboard should help answer questions like:

- Which nations show high exposure but low vulnerability?
- Do wealthier nations recover faster, or does governance matter more?
- How does resilience evolve alongside climate risk?

Visualization and Design Requirements

You must independently determine which visualization types best represent your analysis. Each visual choice should reflect analytical intent, not just aesthetics.

Suggested elements:

- Geo-maps for risk and resilience gradients.
- Time-series plots for resilience evolution.

- Scatter or bubble plots showing relationships (e.g., GDP vs. RRS).
- Treemaps or Sankey diagrams to depict flow of impact or aid.
- Interactive filters by region, year, or disaster type.

Your final dashboard should consist of at least **three analytical views** and feel exploratory, not static.

Deliverables

1. Interactive Tableau Dashboard:

- Multi-view and interactive with meaningful interrelations.
- Derived metrics and documented formulas.
- Insightful and well-labeled design.

2. Technical Report (6–8 pages):

- Refer to **Report Submission** guild lines.
- Datasets used, preprocessing methods, and data transformations.
- Computed formulas and analytical rationale.
- Justification of visualization choices and insights.

3. Dataset Documentation Sheet:

- Dataset names, URLs, and sample structures.
- Description of derived or cleaned attributes.

Learning Outcomes

The project must satisfy these outcomes:

- Apply data fusion and feature engineering techniques on large public datasets.
- Construct and reason about computational models using real data.
- Translate analytical insights into effective visualization narratives.
- Demonstrate analytical autonomy and creative problem-solving.
- Produce professional, interactive dashboards for complex societal phenomena.