Project Requirements

1. Dataset Selection (must be challenging)

- Choose a dataset that contains:
    - At least **10,000 records**.
    - At least **5 different variables**, including:
        - Quantitative (numeric, continuous or discrete)
        - Categorical (nominal or ordinal)
        - Temporal (date/time)
        - Optional: spatial or network data.
- Possible sources: Kaggle, government open data portals, WHO/UN databases, World Bank, GitHub datasets OR any other

2. Data Preparation & Cleaning (Real Challenge)

- Perform **data wrangling**:
    - Handle **missing values** using appropriate strategies.
    - Detect and remove **redundant or derivable attributes**.
    - Deal with **outliers** and justify treatment.
    - Standardize inconsistent categories (e.g., "M", "Male", "m").

- Document **before-and-after transformations** (with screenshots/code snippets).

3. Data Reduction & Transformation

- Apply at least one **data reduction technique**:
    - Aggregation (e.g., monthly averages instead of daily).
    - Sampling (e.g., stratified sampling for large datasets).
    - Dimensionality reduction (e.g., PCA for multivariate data).
- Justify why the reduction was necessary.

4. Visualization Design & Implementation

Out of many, at least **2 visualizations must be interactive** (using Tableau, Power BI, or Python libraries like Plotly/Bokeh).

5. Ethical & Design Considerations

- Explicitly state:
    - How your visualization avoids **misleading representations**.
    - Accessibility features (colorblind-friendly palettes, readable labels).
    - Ethical concerns (privacy, bias, sensitive data).

6. Deliverables

- **Technical Report (15–20 pages):**
    - Dataset description & source.
    - Detailed cleaning, wrangling, and reduction process.
    - Explanation of visualization choices (why this chart, why this encoding).
    - Discussion of insights and conclusions.
    - Reflection on ethics and limitations.
    - Overall message/ Narrative
- **Final Presentation (12–15 minutes):**
    - Introduction of the dataset
    - Demonstrate before/after cleaning.
    - Walk through explanatory data analysis.
    - Highlight 2–3 key insights supported by visuals.
    - Present an overall message/narrative

Evaluation Criteria

- **Data Preparation & Wrangling (20%)**
- **Data Reduction Techniques (15%)**
- **Visualization Variety & Quality (25%)**
- **Analytical Depth & Insights (20%)**
- **Ethics & Design Awareness (10%)**
- **Team Collaboration & Presentation (10%)**

Rubrics Table for Data Visualization Project

Criteria	Weight	Excellent (Full Marks)	Good (75%)	Fair (50%)	Poor (25% or less)
1. Data Preparation & Wrangling	20%	Dataset cleaned thoroughly: missing values handled with justification; redundant attributes removed; outliers treated logically; consistent categories ensured; full documentation with before/after snapshots provided.	Minor issues in cleaning; most justifications provided; some documentation missing.	Basic cleaning done but little justification; transformations not fully clear.	Minimal effort; data left messy; missing justifications/documentation.
2. Data Reduction Techniques	15%	Applied ≥1 advanced technique (aggregation, sampling, PCA, etc.) with strong justification; demonstrated clear improvement in dataset usability.	Applied a reduction technique with some justification; impact partially explained.	Technique applied but not well justified; unclear effect on analysis.	No reduction or inappropriate/unjustified method used.
3. Visualization Variety & Quality	25%	Portfolio includes ≥6 high-quality visuals covering required types; at least 2 are interactive; visuals are clear, accurate, aesthetically pleasing, and insightful.	≥5 visuals with moderate variety; at least 1 interactive; some design/clarity issues.	Limited variety (≤4 visuals); static visuals dominate; unclear or cluttered charts.	Very few visuals (<3) or visuals are misleading/confusing.
4. Analytical Depth & Insights	20%	Visualizations lead to deep, non-trivial insights; clear narrative built; interpretations backed by visuals; connections made to real-world implications.	Insights present but somewhat surface-level; narrative less cohesive.	Descriptive only (restating what chart shows); limited analytical depth.	No meaningful insights; visuals presented without interpretation.
5. Ethics & Design Awareness	10%	Explicit consideration of bias, privacy, and data sensitivity; design follows accessibility guidelines (colorblind-friendly, readable labels); visuals avoid distortion/misleading encoding.	Mentions ethical and design issues but lacks depth; minor accessibility issues.	Very limited discussion of ethics/design; ignores accessibility.	No mention of ethics, bias, or accessibility.
6. Team Collaboration & Presentation	10%	Presentation (12–15 mins) well-organized, engaging, clear narrative; smooth collaboration; all members contribute; strong delivery of key insights.	Clear presentation with some timing/narrative issues; majority contributed.	Uneven team contribution; weak delivery; presentation unclear in places.	Poor coordination; very short or unclear presentation; one or two members dominate.