Using the topic from your CA682 visualisation assignment, apply the Generic Data Analytics Pipeline to describe how the data may have been Gathered, Processed, Analysed, Presented and Preserved. Give a brief description of the activities at each stage (1-2 sentences) and identify any specific tools that you did or would use. If you didn't specifically perform any stage then you can make assumptions or predictions about the actions and tools.

If you didn't complete a visualisation assignment then write about a scenario based on analysing *student feedback to DCU in relation to online teaching in 2020*.

Gathering

Processing

Analysis

Presenting

Preserving

Here are the answers to the assignment questions based on your **"Data Visualization Report: Ireland's Socio-Economic and Environmental Trends (2013–2023)"**:

Q1(a): Generic Data Analytics Pipeline Application

Gathering:
The datasets were sourced from the World Bank Open Data platform and Ireland's Central Statistics Office (CSO). Key indicators included GDP, unemployment, CO2 emissions, population, and life expectancy.
Tools: Web platforms, APIs, and manual downloads.

Processing:
Rigorous pre-processing was conducted, such as handling missing values through interpolation (e.g., life expectancy) and averages (e.g., migration data), data standardization (converting GDP and CO2 emissions into consistent units), and merging all indicators into a single dataset.
Tools: Python libraries like `pandas` for data cleaning and integration.

Analysis:
Trends and correlations were explored, such as the relationship between GDP growth and unemployment decline or between CO2 emission reductions and environmental policies. Outliers and patterns were visualized to understand socio-economic progress.
Tools: Python (`pandas`, `matplotlib`), Tableau for interactive trend analysis.

Presenting:
Data visualizations included multi-axis line graphs, bar charts, and area charts. Dashboards were designed to highlight key insights such as GDP growth, environmental improvements, and population trends.
Tools: Tableau for dashboards, `matplotlib` for static visualizations.

Preserving:
Cleaned data and visualizations were archived for reproducibility and reference. Documentation ensured the analysis process could be revisited.
Tools: Excel for raw data storage, Jupyter Notebooks for process documentation.

Q2: Statistical Models Used

The analysis relied on exploratory data analysis (EDA) to uncover trends and relationships among indicators. Specific techniques included:

- Linear regression for interpolating missing values.
- Correlation analysis to understand relationships between GDP and unemployment, or CO2 emissions and environmental policies.

Q3: Key Findings

1. **Economic Growth**:
   - Ireland's GDP increased steadily over the decade, with an annual growth rate of ~7.94%.
   - Unemployment rates declined significantly, indicating economic recovery and job creation.
2. **Environmental Progress**:
   - CO2 emissions dropped consistently, particularly after the implementation of the 2015 Climate Action Act.
   - 100% electricity access contributed to improved productivity and living standards.
3. **Social Development**:
   - Life expectancy rose from 80 to 82 years, reflecting better healthcare and living conditions.
   - Migration trends (2016–2020) positively impacted population growth and economic activity.

Q4: Improvements and Observations

1. **Broader Data Scope**:
   Including additional indicators such as income distribution or renewable energy usage could provide a more nuanced view of progress.
2. **Visualization Enhancements**:
   Adding interactivity (e.g., zoom, filtering by year) to all visualizations could improve user engagement.
3. **Future Projections**:
   Incorporating predictive models based on historical trends could provide insights into Ireland's future socio-economic landscape.

Let me know if you need further clarifications or enhancements!

.