Assessment cover

Module No:	DALT7016	**Module title:**	Data Visualisation
Assessment title:	Assignment		
Due date and time:		Friday 15th December 13:00	
Estimated total time to be spent on assignment		30 hours	

LEARNING OUTCOMES

On successful completion of this assignment, students will be able to achieve the module following learning outcomes (LOs):
Critically analyse data visualisation approaches with respect to human sensory modalities
Create appropriate visualisations for temporal, dynamic, and high dimensionality data
Devise methodologies for data interaction to facilitate exploratory data analysis

Engineering Council AHEP4 LOs assessed (from S2 2022-23)		**Met? (Y/N)**
M1	Apply a comprehensive knowledge of mathematics, statistics, natural science and engineering principles to the solution of complex problems. Much of the knowledge will be at the forefront of the particular subject of study and informed by a critical awareness of new developments and the wider context of engineering	
M2	Formulate and analyse complex problems to reach substantiated conclusions. This will involve evaluating available data using first principles of mathematics, statistics, natural science and engineering principles, and using engineering judgment to work with information that may be uncertain or incomplete, discussing the limitations of the techniques employed	
M3	Select and apply appropriate computational and analytical techniques to model complex problems, discussing the limitations of the techniques employed	
M17	Communicate effectively on complex engineering matters with technical and non-technical audiences, evaluating the effectiveness of the methods used	

Statement of Compliance *(please tick to sign)*

I declare that the work submitted is my own and that the work I submit is fully in accordance with the University regulations regarding assessments
(www.brookes.ac.uk/uniregulations/current)

DALT7016 Data Visualisation

Assignment Report

Part 1: Context and EDA

Context and Source Description

Figure 1: SDG Goals: https://commons.wikimedia.org/wiki/File:Sustainable_Development_Goals.svg

I have chosen Sustainable Development Data up to 2023 for this assignment. The dataset "Sustainable Development Data up to 2023" sourced from Kaggle. It highlights the annual progress made on the Sustainable Development Goals since their adoption in 2015 by the 193 UN Member States (Sazid The1 2023). The 2023 edition, positioned at the midpoint to 2030, critically evaluates achievements and outlines priorities to rejuvenate and expedite SDG progress. The dataset contains specific information about how well different countries are performing in terms of sustainability and meeting the Sustainable Development Goals (SDGs). It includes scores for each country and illustrates how each country is functioning in terms of certain sustainability goals. This information assists us in understanding and analysing global efforts toward sustainable development, as well as providing a deeper look at how particular countries are developing in this critical and important field.

URL:*https://www.kaggle.com/datasets/sazidthe1/sustainable-development-report?select=sdg_index_2000-2022.csv*

Dataset description and summary statistics

This dataset encompasses a rich array of data, including country names, country codes, years, SDG index scores, and individual scores for each of the 17 SDGs. With a scope spanning 180 countries and a time frame extending from the year 2000 to 2022. The data types of the key columns are appropriately varied, with 'country_code' and 'country' represented as characters, 'year', 'sdg_index_score,' and 'goal_1_score' to 'goal_17_score' as numeric variables.

Table 1: Column name with data types and description

Column Name	Datatype	Description
country_code	char	A unique identifier that links to the primary dataset.
Country	char	The name of the country.
Year	num	The year of the data entry.
sdg_index_score	num	The overall SDG (Sustainable Development Goals) index score of the country.
goal_1_score	num	The score for Goal 1: No Poverty.
goal_2_score	num	The score for Goal 2: Zero Hunger.
goal_3_score	num	The score for Goal 3: Good Health and Wellbeing.
goal_4_score	num	The score for Goal 4: Quality Education.
goal_5_score	num	The score for Goal 5: Gender Equality.
goal_6_score	num	The score for Goal 6: Clean Water and Sanitation.
goal_7_score	num	The score for Goal 7: Affordable and Clean Energy.
goal_8_score	num	The score for Goal 8: Decent Work and Economic Growth.
goal_9_score	num	The score for Goal 9: Industry, Innovation, and Infrastructure
goal_10_score	num	The score for Goal 10: Reduced Inequalities.
goal_11_score	num	The score for Goal 11: Sustainable Cities and Communities.
goal_12_score	num	The score for Goal 12: Responsible Consumption and Production.
goal_13_score	num	The score for Goal 13: Climate Action.
goal_14_score	num	The score for Goal 14: Life Below Water.
goal_15_score	num	The score for Goal 15: Life on Land.
goal_16_score	num	The score for Goal 16: Peace, Justice, and Strong Institutions.
goal_17_score	num	The score for Goal 17: Partnerships for the Goals.

This data was relatively well-organized when I got it, so I did not have to do a ton to reorganize and clean it. I did, however, manipulate the code to get the average of each goal to create an interactive graph. I found that there are no missing values. The absence of missing values in this context suggests that the dataset is well-prepared for exploratory data analysis and visualization.

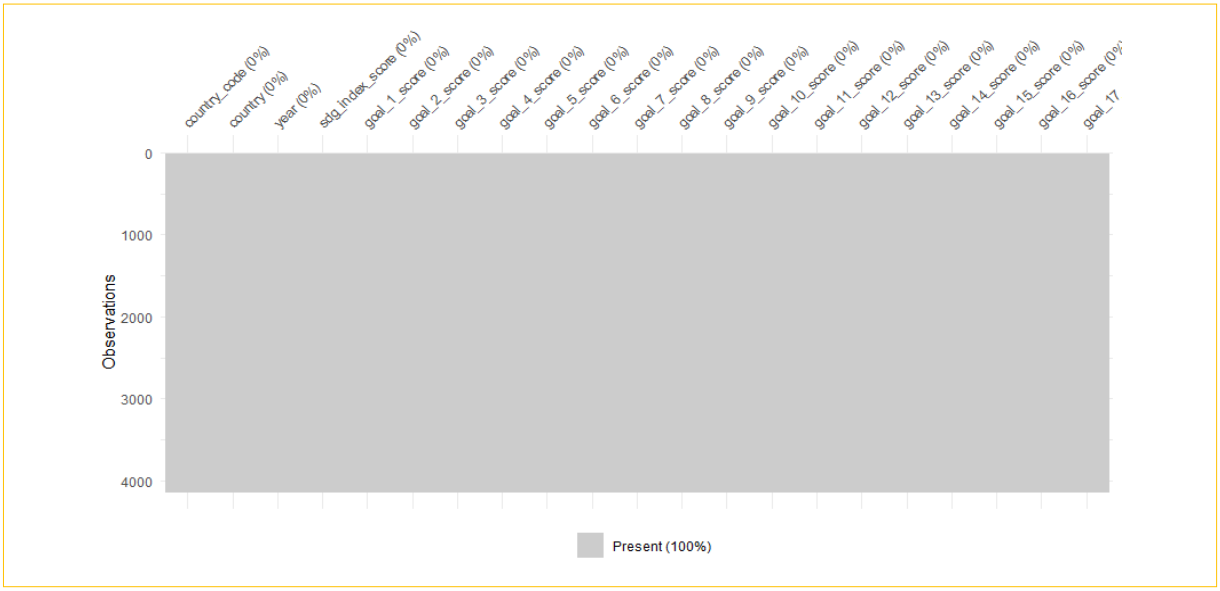

Figure 2 Missing Value

Dataset contains 4,140 rows and 21 columns. The exported CSV file, named data.csv, has a size of 424 KB. These figures provide a quantitative overview of the dataset's dimensions, indicating a moderate-sized dataset. The summary statistics for each variable are as follows:

Table 2: Data Summary

Column Name	Min	Median	Mean	Max
Year	2000	2011	2011	2022
sdg_index_score	36.00	65.10	63.69	86.80
goal_1_score	0.00	81.00	64.82	100.00
goal_2_score	7.70	58.90	57.74	83.40
goal_3_score	5.90	70.20	64.05	97.30
goal_4_score	0.00	80.60	71.99	100.00
goal_5_score	3.5	57.5	55.8	94.0
goal_6_score	11.30	64.40	63.96	95.10
goal_7_score	0.10	64.45	57.34	99.60
goal_8_score	38.40	70.00	69.84	93.60
goal_9_score	0.00	29.70	37.11	99.20
goal_10_score	0.00	53.85	52.37	100.00
goal_11_score	13.80	75.00	69.93	99.90
goal_12_score	32.90	84.80	79.68	99.00
goal_13_score	0.00	90.90	81.93	99.90
goal_14_score	0.00	59.00	48.57	91.20
goal_15_score	26.00	64.40	64.66	97.90
goal_16_score	27.90	61.00	62.19	96.00
goal_17_score	15.10	54.90	55.54	100.00

These statistics show that there is a wide range of variation in the SDG Index Scores and Goal Scores across countries.

Part 2: Design

The visualizations collectively contribute to a detailed understanding of global sustainability efforts and encourage continued collective action towards the SDGs by 2030. The aim of the visualizations is to offer comprehensive insights into the global progress towards Sustainable Development Goals (SDGs) and the specific achievements of individual countries. Through diverse visual representations, the report aims to highlight correlations, trends, and notable accomplishments related to SDGs.

Prospective audience for the visualizations is policymakers, international development organizations, and stakeholders interested in analysing and understanding nation's progress toward certain SDGs. Researchers interested in sustainable development will benefit from comparing goal scores across countries.

In my initial analysis, I am plotting the correlation between a country's overall SDG Index Score and its dedication to achieving Quality Education (Goal 4). The scatter plot, with each point representing a country, shows a consistent trend without outliers. The scatter plot is color-coded by country, with darker shades indicating higher SDG Index Scores and lighter shades representing lower scores. This allows the viewer to easily see how the Quality Education Score varies across countries with different levels of overall SDG progress. The insights from Figure 3 will enhance my visualization, telling a clearer story about top countries' performance in quality education.

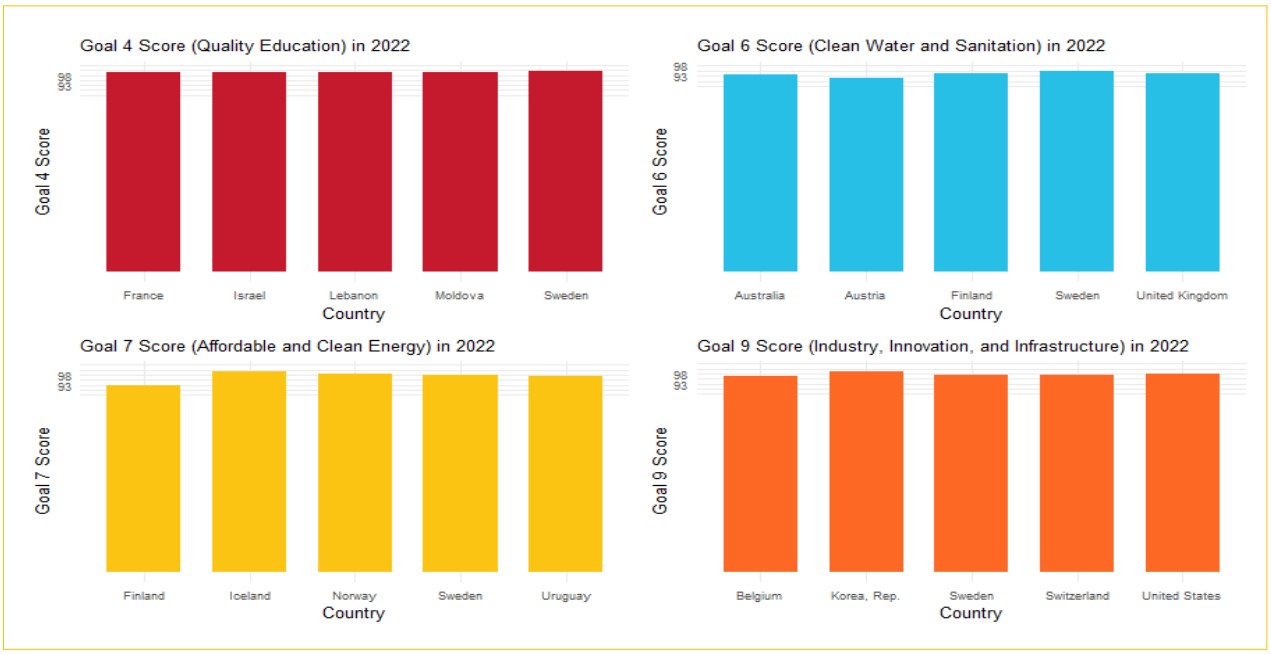

Figure 3: Progress Towards Sustainable Development Goals (SDGs) in 2022: A Comparative Analysis

A bar graph spotlighting sustainability goals like clean water, affordable energy, Quality Education, and innovation guides my final project, showcasing which countries excel in key areas. In my final visualization 1, I will explore how global sustainability progress correlates with a commitment to achieving Quality Education (Goal 4).

This detailed insight into the UK's sustainability trend is a fundamental element in constructing my interactive graph, which will showcase the average scores for each SDG goal worldwide.

By understanding how a specific country like the UK is doing, I can give important background information. This helps provide a complete view of global sustainability trends for all SDG goals. Analysing individual countries adds depth to the interactive graph's story, making it more detailed and informative about the overall global efforts for sustainability.

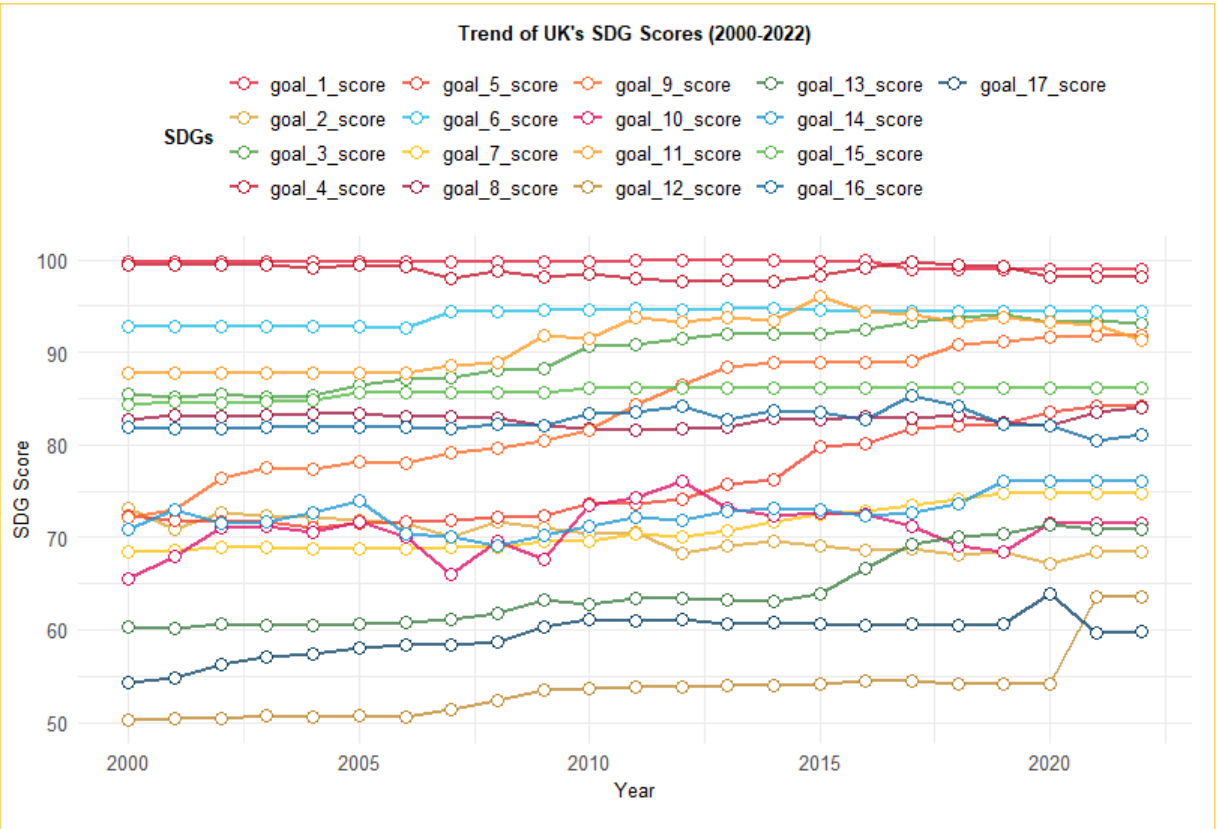

Figure 4: Trend of UK's SDG Score

The planned final visualization serves a dual purpose. Firstly, it will visually present the average scores for each SDG goal, offering a comprehensive overview of global progress in Sustainable Development Goals (SDGs). This allows for easy identification of areas that are thriving and those that may need more attention. Secondly, the graph will be dynamic, featuring an interactive hoverable element. This element will provide additional details such as the title, average score, and the corresponding year for each SDG goal. This interactive feature is tailored to engage a broader audience interested in global sustainability trends. The

intention is to create a visually appealing and user-friendly experience that encourages exploration and understanding of SDG progress worldwide.

Part 3: Final Visualisation

Visualisation 1 Commentary

In crafting the initial scatter plot, I incorporated colours following the UN guideline for goal 4. This thoughtful use of colour enhances the interpretability of the visualization. By concentrating on the connection between a country's SDG Index Score and its dedication to Quality Education, I ensured simplicity and clarity. I used (geom_smooth) to add a smoothed line to the plot. Notably, the visualization maintains a reasonable file size of 238 KB as it is a PDF file, facilitating smooth accessibility. It is tailored to policymakers, presenting a straightforward analysis without unnecessary complexities. The correlation coefficient (r) of 0.862 between the Quality Education Score (Goal 4) and the SDG Index Score suggests a strong positive linear relationship. This line suggests that the relationship between the two measures is not perfectly linear, but it does show a general upward trend. The absence of outliers indicates a consistent trend, suggesting that countries with higher SDG Index Scores also tend to perform better in Quality Education.

Visualisation 2 Commentary

In crafting the interactive graph, I created a custom colour palette inspired by the official hues of each UN (SDG). This intentional design choice aimed at fostering clarity and engagement, particularly for viewers well-versed in UN SDGs (United Nations 2023). For enhanced visual appeal and accessibility, I have used Arial, sans-serif font with a 12px size. Goal names are formatted using sprintf, and the interactive plot is saved as an HTML file using the saveWidget function. The file maintains a reasonable size of 3,827 KB. The hoverinfo feature provides additional information when users interact with the plot. It includes title, year, and average score. The dynamic nature of the graph, coupled with the UN-inspired colours, contributes to both accessibility and visual appeal.

I focused on specific areas of interest, emphasizing positive global progress in the SDGs while highlighting regions requiring attention. Notable strides in SDG 9 (Industry, Innovation, and Infrastructure) are observed, but limited progress in SDG 13 (Climate Action) signals the need for urgent initiatives. Goal 17 demonstrates a positive shift in 2019, highlighting global partnership efforts. The result is a comprehensive visualization that effectively communicates insights, avoids unnecessary repetition, and ensures an impactful presentation of information.

References

United Nations 2023, 'Global Sustainable Development Report 2023', United Nations Sustainable Development Goals, https://sdgs.un.org/

Sazid The1 2023, 'Sustainable Development Report Dataset', Kaggle, https://www.kaggle.com/datasets/sazidthe1/sustainable-development-report