

Exercise 6.1

Sourcing Open Data

Hanadi Salim

1. Data Source:

- a. **Summary of the data:** The Sustainable Development Report (SDR) reviews progress made each year on the Sustainable Development Goals since their adoption by the 193 UN Member States in 2015. At the halfway mark to 2030, the Sustainable Development Report 2023 takes stock of progress made and discusses priorities to restore and accelerate SDG progress.
 - b. **Data Source:** The primary dataset was retrieved from the [SDG Transformation Center](#)
 - c. **Collection method:** One third of the data used in the SDG Index comes from outside official statistics, and research, science, NGOs and space-based technologies. Combined with official datasets, often provided by UN custodian agencies.
 - d. **Content Overview:** the data contain two data sets.
- Primary Dataset (sustainable_development_report_2023.csv) includes the following columns:

Column Name	Description
country_code	The country code of the respective country.
country	The name of the country.
region	The region to which the country belongs.
overall_score	The overall sustainability (SDG index) score of the country.
goal_1_score	The score for Goal 1: No Poverty.
goal_2_score	The score for Goal 2: Zero Hunger.
goal_3_score	The score for Goal 3: Good Health and Wellbeing.
goal_4_score	The score for Goal 4: Quality Education.
goal_5_score	The score for Goal 5: Gender Equality.
goal_6_score	The score for Goal 6: Clean Water and Sanitation.
goal_7_score	The score for Goal 7: Affordable and Clean Energy.
goal_8_score	The score for Goal 8: Decent Work and Economic Growth.
goal_9_score	The score for Goal 9: Industry, Innovation and Infrastructure.
goal_10_score	The score for Goal 10: Reduced Inequalities.
goal_11_score	The score for Goal 11: Sustainable Cities and Communities.
goal_12_score	The score for Goal 12: Responsible Consumption and Production.
goal_13_score	The score for Goal 13: Climate Action.
goal_14_score	The score for Goal 14: Life Below Water.

goal_15_score	The score for Goal 15: Life on Land.
goal_16_score	The score for Goal 16: Peace, Justice and Strong Institutions
goal_17_score	The score for Goal 17: Partnerships for the Goals.

- **Supplementary Dataset (sdg_index_2000-2022.csv):**

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

goal_16_score	The score for Goal 16: Peace, Justice and Strong Institutions.
goal_17_score	The score for Goal 17: Partnerships for the Goals.

2. Data Profile:

a. Data Cleaning

Data frame	Column Name	Changes Made	Comment
sustainable_development_report_2023	goal_1_score goal_10_score goal_14_score	Dropped missing values	
	Region	Dropped	To have the same columns for both data sets
	year	Added “year” column	To be in line with the other data set
	Country code Country Year	Change type to category	For data consistency and save memory.
sdg_index_2000-2022	Country code Country Year	Change type to category	For data consistency and save memory.
	sdg_index_score	Rename to “overall_score”	To be in line with the other data set

Note: There were no mixed or duplicated values in both data sets.

b. Data Understanding (basic statistic):

	overall_score	goal_1_score	goal_2_score	goal_3_score	goal_4_score	goal_5_score	goal_6_score	goal_7_score	goal_8_score	goal_9_score
count	4247.000000	4247.000000	4247.000000	4247.000000	4247.000000	4247.000000	4247.000000	4247.000000	4247.000000	4247.000000
mean	63.810287	65.176286	57.794174	64.225798	72.176887	55.985601	64.074543	57.486799	69.912924	37.573592
std	10.788753	35.974981	11.112886	22.328609	26.456557	17.320707	15.349202	21.472004	10.027497	26.062037
min	36.000000	0.000000	7.700000	5.900000	0.000000	3.500000	11.300000	0.100000	38.400000	0.000000
25%	55.108814	31.900000	52.400000	45.200000	57.450000	43.300000	52.400000	41.500000	64.200000	15.800000
50%	65.200000	81.200000	59.000000	70.400000	80.800000	57.600000	64.600000	64.500000	70.100000	30.200000
75%	72.100000	98.600000	65.400000	81.400000	94.500000	68.900000	74.500000	72.100000	76.550000	53.750000
max	86.800000	100.000000	83.401125	97.300000	100.000000	94.021667	95.100000	99.600000	93.600000	99.200000

	goal_10_score	goal_11_score	goal_12_score	goal_13_score	goal_14_score	goal_15_score	goal_16_score	goal_17_score
count	4247.000000	4247.000000	4247.000000	4247.000000	4247.000000	4247.000000	4247.000000	4247.000000
mean	52.613359	70.002384	79.674469	81.978055	49.015592	64.713160	62.192049	55.677464
std	31.177232	18.150558	16.717471	21.271097	27.849283	13.431565	14.610898	13.046680
min	0.000000	13.800000	32.900000	0.000000	0.000000	26.000000	27.900000	15.100000
25%	24.100000	55.610000	69.400000	72.900000	43.653250	55.100000	51.600000	46.150000
50%	53.942500	75.100000	84.800000	90.900000	59.200000	64.500000	61.000000	55.000000
75%	80.200000	84.800000	94.100000	97.123750	67.500000	73.900000	73.500000	65.000000
max	100.000000	99.900000	99.000000	99.925333	91.200000	97.900000	96.000000	100.000000

c. Limitations and ethics: the data doesn't have the volume of the population for each country to know if that has an impact on reaching the goals. I didn't see any ethic related issues in this data set.

3. Project Questions:

- a. *Which country/ countries have the highest overall sustainability development Goals score?*
- b. *What is the overall progress over the years for the Sustainability development Goals?*
- c. *Were there any high or low points over the years?*
- d. *Which goals have improved and which ones have decreased?*
- e. *Where there any correlation between the any of the goals?*