



Countries Rating

Evgenii Sokolov

April 2023

OUTLINE



- EXECUTIVE SUMMARY.....(3)
- METHODOLOGY.....(4)
- METHODOLOGY in detail*.....(5-9)
- RESULTS.....(10)
- DASHBOARD.....(11)
- FINDINGS.....(12-16)
- CONCLUSION & VALUE.....(17)

*Slides 5-9 outline the study's methodology. Skip if uninterested in details.

EXECUTIVE SUMMARY



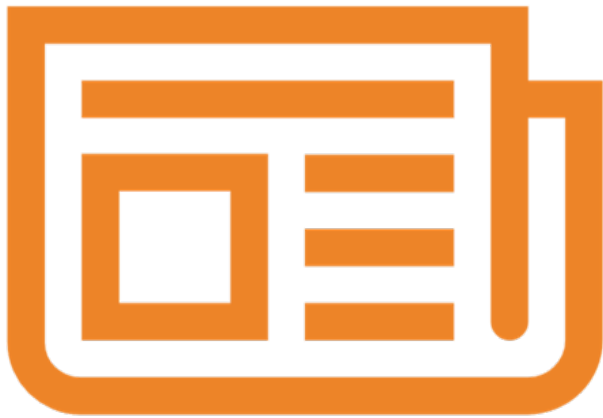
Determining the best country is a complex question that depends on the data used, the criteria for comparison, and personal priorities.

In this project, data has been collected that characterizes the level of development of countries across various aspects of human life.

The data has been prepared for comparison and the construction of a corresponding ranking.

Interactive dashboards have been generated as a result of the analysis, which allow for the exploration of a country's development across different aspects and can help identify the best country for an individual based on their preferences.

METHODOLOGY



- **Data Collection:** The data for this project was sourced from the OECD's* ["How's Life?" study](#), with relevant information selected based on the Better Life Index that focuses on key areas of societal well-being.
- **Data Exploration:** The dataset was explored to identify data types, missing values, data distributions, outliers, and correlations.
- **Data Wrangling:** Missing values were imputed using linear regression models and nearest neighbor methods. Overall Features and a Total Score parameter were created for comparing countries.
- **Data Visualization and Analysis:** The results of the study were visualized and analyzed, including overall country rankings and rankings based on overall and specific features.
- **Building Dashboard:** Dashboards were created to explore overall country rankings, overall and specific features, specific countries, and country comparisons.

Tools: Jupiter Notebook. Python: Pandas, Numpy, Plotly, Sklearn. Power BI: charts, buttons, slicers, sliders, bookmarks, selections, parameters, measures, DAX.

*OCED - [The Organization for Economic Co-operation and Development](#).

METHODOLOGY in detail

Data Collection:*

- For this research, specific indicators of a country's development were selected from the OECD's "How's Life?" study, which allows for clear comparisons of the main spheres of societal life and has been formulated by the OECD as the "**Better Life Index**". These indicators form the basis of our research and have been grouped by the OECD into the following categories: 'Housing', 'Income', 'Jobs', 'Community', 'Education', 'Environment', 'Civic engagement', 'Health', 'Life satisfaction', 'Safety', and 'Work-life balance'. Countries are compared based on these categories. More information about these categories, including the specific indicators, can be found [here](#).
- The dataset containing the Better Life Index data has been loaded as a .csv file at this [link](#).
- **The ranking includes countries** that are members of the OECD or OECD partner countries: Australia, Austria, Belgium, Brazil, Canada, Chile, Colombia, Costa Rica, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Russia, Slovak Republic, Slovenia, South Africa, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States.
- For ease of further manipulation, the dataset has been imported into a Jupyter Notebook and transformed into a **Pandas dataframe**.

Tools: Jupyter Notebook. Python: Pandas: `.read_csv()`.

*For more detailed information, including the code and graphs, please refer to the Jupyter Notebook. Available in [.ipynb](#) and [.pdf](#) formats.

METHODOLOGY in detail

Data Exploration:*

- **The dataset** consists of 2369 rows and 17 columns in its original form, which was transformed into 42 rows and 24 columns with 1008 values. Thus, the dataset contains 41 countries and 24 features (one row - mean value).
- **The data type** of the dataset is float64.
- **Missing values** were identified and counted for each country and feature. The dataset contains a total of 73 missing values, with the highest number of missing values found in the features "Time devoted to leisure and personal care", "Household net wealth" and countries such as South Africa, Brazil, and Colombia.
- **Data distributions** were explored by summary statistics and creating box plots for each feature. A total of 36 outliers were found in the dataset, with the highest number of outliers detected in the "Homicide rate" feature. A loop was created to generate corresponding graphs.
- **Feature correlations** were studied using an interactive heatmap. Scatter plots with trend lines were created to examine the distribution of values for features whose correlation index was greater than 0.7 in absolute value. Features whose correlation index was greater than 0.7 were recorded in a dictionary. Loops were created to generate corresponding graphs and record feature correlations in a dictionary.

Tools: Jupyter Notebook. Python: for. Pandas: `.shape`, `.head()`, `.pivot()`, `.info()`, `.isna()`, `.sum()`, `.sort_values()`, `.describe()`, `.columns`, `.corr()`, `.loc[]`. Plotly: `.box()`, `.imshow()`, `.scatter()`.

*For more detailed information, including the code and graphs, please refer to the Jupyter Notebook. Available in [.ipynb](#) and [.pdf](#) formats.

METHODOLOGY in detail

Data Wrangling:*

- **The missing values were handled** using linear regression models based on features with a correlation index of absolute value greater than 0.7. Each missing value was filled individually based on the not null corresponding features. After filling the missing values, there were only 14 missing values left in columns that did not exhibit significant correlation with other columns or where the corresponding values in columns with significant correlation were absent. To fill the remaining missing values, the method of nearest neighbors was used after normalizing the necessary data using the min-max normalization method. A loop was created to train the linear regression models individually for each missing value, predict the values, and fill them in the dataset. This approach can be applied to any dataset in the future.
- **To prepare the data**, overall features were created for the comparative analysis of countries based on the following features: 'Housing', 'Income', 'Jobs', 'Community', 'Education', 'Environment', 'Civic engagement', 'Health', 'Life satisfaction', 'Safety', and 'Work-life balance'. The normalized values of these features were summed up to create a total score that was then normalized to eliminate any unequal influence of the individual features on the final ranking.

Tools: Jupyter Notebook. Python: `for`, `.items()` `range()`, `len()`, Pandas: `.isna()`, `.any()`, `.loc[]`, `.dropna()`, `.to_frame()`, `.transpose()`, `.empty`, `.tolist()`, `.sum()`, `.sort_values`, `.merge()`. Numpy: `np.around()`. Sklearn: `.LinearRegression()`, `.fit()`, `.predict()`, `MinMaxScaler()`, `.fit_transform()`, `KNNImputer()`, `.inverse_transform()`.

*For more detailed information, including the code and graphs, please refer to the Jupyter Notebook. Available in [.ipynb](#) and [.pdf](#) formats.

METHODOLOGY in detail

Data Visualization and Analysis:*

- **To examine the total rankings of countries**, a bar chart was generated for identifying the top-performing and underperforming countries based on their ratings. The countries with unexpected ratings were further examined by showcasing their overall features in the bar chart. This analysis was performed to discern the areas in which these countries should focus on to enhance the quality of life for their citizens.
- **To compare countries** with similar overall features, a stacked bar chart was created. By comparing such countries, their distinct attributes and characteristics that need improvement were identified.
- **To investigate the specific features** of interest, a bar chart was created to represent the relevant data. This provided a detailed understanding of the factors that contribute to the high or low rating of a particular country, thereby highlighting areas that require improvement.
- The final dataset was stored in a .csv format and is available for viewing [here](#).

Tools: Jupyter Notebook. Python: for. Plotly: .bar(). Pandas: .loc[], .sort_values(), and read_csv(). Loops were also created to generate the charts.

*For more detailed information, including the code and graphs, please refer to the Jupyter Notebook. Available in [.ipynb](#) and [.pdf](#) formats.

METHODOLOGY in detail

Building dashboard:

- The data from the previously obtained .csv file was loaded into Power BI, and a dashboard with **5 pages** was created: overall country ranking, aggregated feature analysis, specific feature analysis, country-specific analysis, and country comparison. **Navigation** through the dashboard is achieved via created buttons, and a quick access button to feature values was also included.
- **The first page** displays the overall ranking. A slider allows for the selection of countries, which is saved for subsequent pages. Countries can be chosen at the user's discretion or by clicking on the provided country group buttons. A slider allows for the selection of the importance of a particular aggregated feature when calculating the country's ranking (using parameters, measures via DAX). The initial weight of each feature is set to 1. When a different weight is selected, the value of each feature is multiplied by the selected weight to calculate the Total Score.
- **The second and third pages** display the country ranking based on specific overall and specific features. Normalized values (0-1) are used for aggregated features, and raw values are used for specific features. Buttons allow for the exploration of specific features of interest.
- **The third and fourth pages** display the country-specific analysis and comparison. A slider allows for the selection of the country of interest. Buttons allow for toggling between the analysis of aggregated and specific features. Normalized values (0-1) are used, with negative specific features turned positive and higher values being better for all features.

Tools: Power BI: stacked bar charts, clustered column charts, text boxes, buttons, slicers, sliders, bookmarks, selections, parameters, measures, DAX.

RESULTS

TOP 10*



Bottom-ranked 10*



*The ranking includes countries that are members of the OECD or OECD partner countries (slide 5).

DASHBOARD



The detailed results of the study can be examined on an interactive dashboard created in Microsoft Power BI. The dashboard includes navigation options, the ability to select countries of interest, and the significance of features used to rank countries. Additionally, users can investigate specific characteristics and compare countries.

Available on [this webpage](#) or within the [this PowerBI file](#).

Individuals may conduct their own research using the dashboard. Herein, we present several examples of potential discoveries that may be revealed.

FINDINGS

Construct a personalized ranking by altering the significance of different features.

The resulting outcomes will be influenced by the degree of importance assigned to each feature.



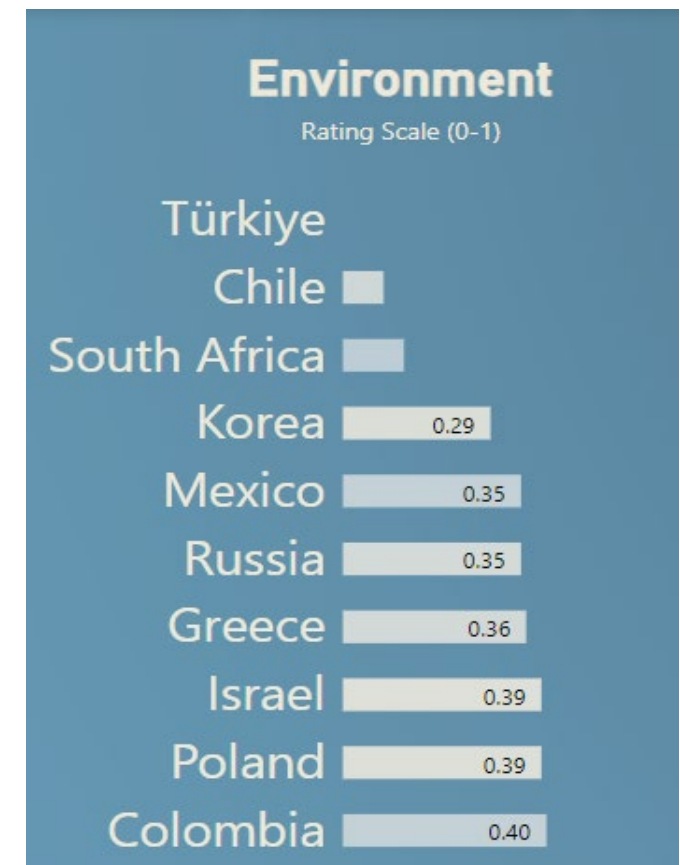
Rate the features by their importance for you:

Housing 2 	Income 5 	Jobs 4
Community 2 	Education 3 	Environment 3
Civic engagement 1 	Health 3 	Life Satisfaction 3
Safety 3 	Work-Life Balance 1 	

FINDINGS

Explore overall features.

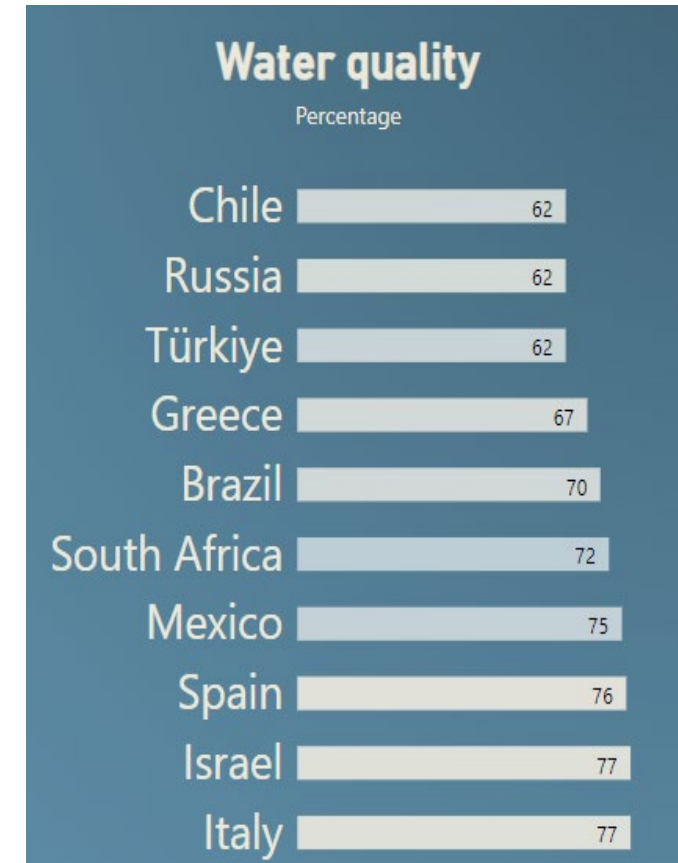
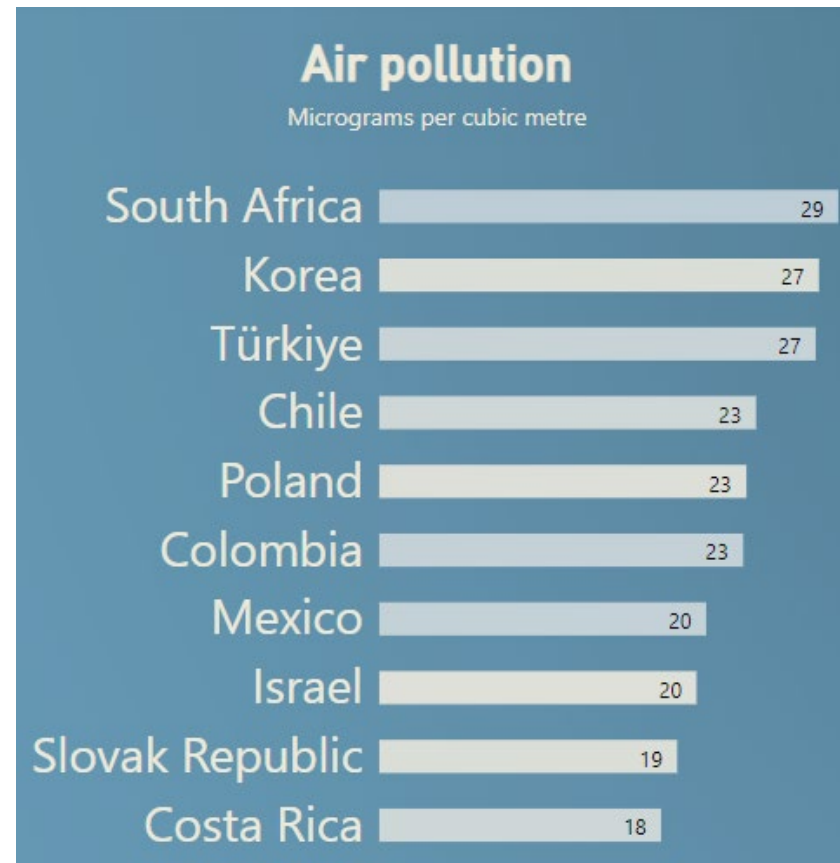
We can examine the overall features to determine which countries have the best and worst environmental conditions.



FINDINGS

Explore specific features.

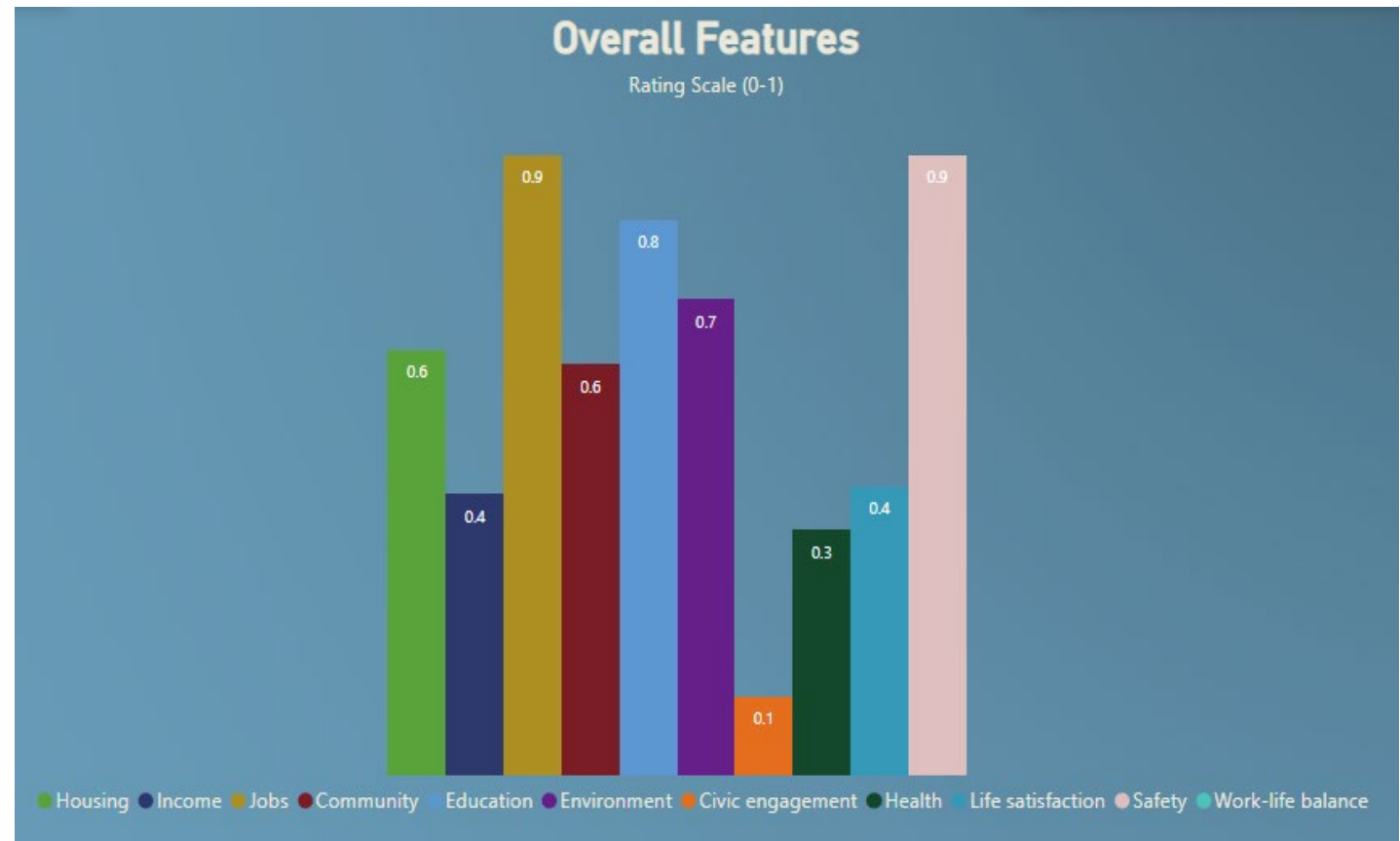
Our investigation will focus on the specific features that contribute to Turkey's poor environmental ranking. Graphical analysis indicates that Turkey has some of the highest air pollution levels at 27 mcg/m³ and one of the lowest water quality satisfaction rates at 62%.



FINDINGS

Explore specific country.

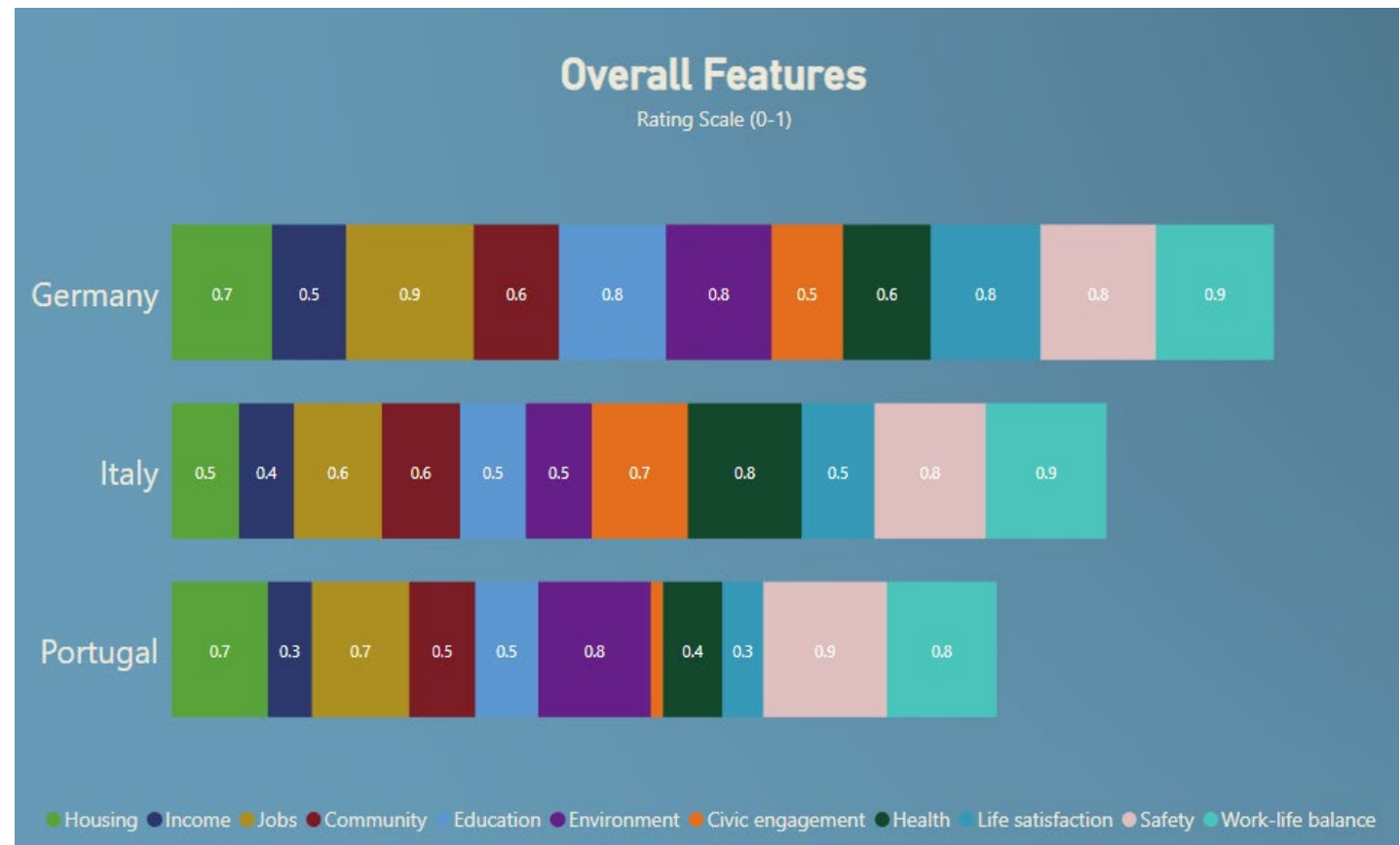
Furthermore, we will perform an in-depth analysis of Japan, a country that received a low overall ranking in previous Total Score reports. Our graphical representation reveals that this is due to Japan's extremely low scores in 'Work-life balance' and 'Civic engagement', as well as low scores in 'Health', 'Life satisfaction,' and 'Income'.



FINDINGS

Compare countries.

We can compare neighboring countries, for example, Germany, Portugal, and Italy. It can be observed that Portugal lacks in indicators such as "Community" and "Civic engagement", while Italy needs to work on such indicators as "Environment" and "Work-life balance". In comparison to these countries, Germany shows high levels across all dimensions, with the lowest scores in the categories of "Income" and "Civic engagement".



CONCLUSION & VALUE



In this study, we examined, prepared, and analyzed OECD data regarding key indicators of societal well-being in various countries. Based on our work, we developed an interactive ranking of countries in which users can find a leader among countries based on their own preferences, as well as explore specific areas in greater detail.

The study materials will be useful to individuals who wish to better understand their country or are planning to relocate. Additionally, this visual representation can be beneficial to officials for a more comprehensive understanding of the processes taking place in their country.

Commence your investigative expedition by accessing the dashboard on [this webpage](#) or within [this PowerBI file](#).