# Project Description: Processing United Nations SDG Data

Author: Jan Hrušák

This project processes data on the progress of the United Nations Sustainable Development Goals (SDGs) for various countries, downloaded from the [UN SDG Data Portal](https://unstats.un.org/sdgs/dataportal/analytics/DataAvailability). The dataset provides a percentage progress towards 17 specific SDGs, such as "No Poverty" or "Zero Hunger". The data represent averages for at least two years since 2015.

**Steps Conducted in the Jupyter Notebook:**

1. **Import Libraries**: Pandas library was used for data manipulation and Matplotlib for visualization.
2. **Load and Clean Data**: The dataset was loaded as a CSV file using pd.read\_csv() and first few unnecessary rows were removed.
3. **Inspect Data Structure**: The data structure was inspected using info() method. The dataset included 251 rows and 19 columns.
4. **Rename Columns**: Column names were corrected using the first row of the dataset.
5. **Data Cleaning**: The first (unnecessary) row was removed, the index reset, % symbols stripped from the goal columns values, and those were converted to a float for analysis.
6. **Summary Statistics**: Summary statistics were generated for the goal columns using .describe().
7. **Visualize Data**: Histograms were plotted to see the distribution of values of the SDG goals.
8. **Write to Excel**: Processed data and summary statistics were written to an Excel file with two separate sheets.

For the individual steps and their description in more detail please see the file l63408.ipynb.