# DATA ANALYTICS PROJECT DEFINITION – MRC CONSULTANTS

Madrid, 3 July 2023

# Geographic scope

Latin America and the Caribbean region, and more specifically these countries:

* Brazil.
* Mexico.
* Argentina.

# Project name

Analysis, forecast and visualization of energy matrices and balances.

# Programming languages and tools

The solution will be implemented in R/Python.[[1]](#footnote-0)

The results of the exercise will be summarized in a dashboard using the R Shiny and/or Flexdashboard libraries or their equivalent packages in the Python environment.

Dashboards presented in other platforms like Power BI or Tableau will be also acceptable.

# Scope of work

# Background and rationale

Energy matrices are important tools in the field of energy planning and analysis. They provide a comprehensive overview of energy sources, consumption, and related factors within a particular region or system. They serve various purposes, including:

* Allow policy makers to formulate and implement affective measures.
* Help identify energy inefficiencies.
* Play a crucial role in assessing the environmental impact of different energy sources.

# Approach and methodology

The objective is to assess the evolution of an energy matrix for a specific country or region over time and develop a mid-term projection. In this context, data analysis plays a crucial role by:

* Collecting and organizing the data in a structured format, ensuring its quality and consistency. This step is vital to establish a reliable foundation for further analysis.
* Performing exploratory data analysis (EDA) to gain a preliminary understanding of the data. EDA helps uncover patterns, trends, and relationships within the **energy matrix dataset**. It provides insights into the variables at hand and informs subsequent analytical steps.
* Building indicators to effectively evaluate and analyse the energy matrices. These indicators serve as measurable metrics that reflect specific aspects of the energy system, facilitating meaningful comparisons and assessments.
* Developing a mid-term projection of the energy matrix using suitable machine learning algorithms and econometric techniques. By using historical data, economic and demographic projections, and applying appropriate modelling techniques, it is possible to project future trends in the energy matrix. These projections provide valuable insights into potential changes in energy sources, consumption patterns, and related factors. Decision-makers can utilize this information to formulate effective strategies and policies for the future, aiming for a more sustainable and resilient energy system.
* Constructing visualizations such as Sankey charts and other relevant figures to analyse the energy matrices and present key information.[[2]](#footnote-1) Visual representations enhance the comprehension and communication of complex data, enabling stakeholders to grasp trends, interconnections, and potential areas for improvement more easily.

# Work schedule

Start date: first half of June 2023.

Completion date: first half of September 2023.

Meetings:

* Inception / kick-off meeting
* Progress meetings: approx. every two weeks
* Final meeting: presentation of results

# Datasets and modeling

In particular, we will use the datasets published by OLADE[[3]](#footnote-2) in its SIELAC website:

<https://sielac.olade.org>

The databases contain economic and energy statistics from 1970 to the present day. The analyst should choose among any of these countries:

* Brazil.
* Mexico.
* Argentina.

The key variable to be forecasted is the aggregated demand (consumption) up to 2040, with annual granularity[[4]](#footnote-3).

Once the long-term demand is predicted, then the analyst will determine how to cover the consumption based on historical information provided by OLADE, considering the following assumptions:

* **No additional new electricity generation capacity based on nuclear power.**
* **No additional new electricity generation capacity based on fossil fuels.**
* **Additional new generation capacity will be based mainly on renewable energies (biomass, solar, wind, geothermal, etc.).**

# Results and deliverables

# Project code

The students will provide MRC with the source code and scripts / notebooks developed in the R/Python languages.

# Reporting

The students will provide MRC with the final project report implemented in an R markdown / Quarto notebook, or its equivalent reporting tools in the Python environment.

# Visualization and communication of results

The results of the exercise will be summarized in a dashboard using the R Shiny and/or Flexdashboard libraries or their equivalent packages in the Python environment.

Dashboards presented in other platforms like Power BI or Tableau will be also acceptable.

Ideally, the dashboard should include a series of Sankey charts representing the energy balances per year in the forecasting period (2023-2040), taking as references the following examples:

* <https://sielac.olade.org/WebForms/Reportes/InfogramaBalanceEnergeticoSimplificado.aspx?or=545&ss=2&v=3>
* <https://plotly.com/r/sankey-diagram/>

# Language

Solutions implemented in English or Spanish will be acceptable.

# Contact details

Santiago Fernández (sfernandez@mrc-consultants.com)

Jorge Colomer ([jcolomer@mrc-consultants.com](mailto:jcolomer@mrc-consultants.com))

Repositorio de data  
Repositorio en github

Explorar la forma de descargar los datos de excel automáticamente (web scraping)

Descargar de forma manual los datos de excel  
  
Evaluar la calidad del dato de cada uno de los países y quedarnos con la mejor opción, como se pide en el apartado 5 de este texto.

1. Implementation in R is preferable, although solutions in the Python environments will be acceptable. [↑](#footnote-ref-0)
2. https://plotly.com/r/sankey-diagram/ [↑](#footnote-ref-1)
3. Organización Latinoamericana de Energía [↑](#footnote-ref-2)
4. It’s up to the analyst to choose the methodology for carrying out the long-term forecast of the consumption based on the historic information provided by OLADE’s SIELAC. [↑](#footnote-ref-3)