

SEGUNDA PARTE DEL PROYECTO

Dairo Joan Rivas García
Responsable

Raúl Ramos Pollas
Docente

Modelos y Simulación de Sistemas
Materia



**UNIVERSIDAD
DE ANTIOQUIA**

1 8 0 3

Universidad de Antioquia

Facultad de ingeniería

Medellín 2023

Introducción

En esta entrega del proyecto se describe el avance realizado en la evaluación del dataset seleccionado, el cual consta de la información del consumo de energía eléctrica de varias organizaciones mundiales y las cuales se presenta información sobre consumo de energías renovables y generada de compuestos fósiles y químicos.

1. Avance del análisis de Dataset

El Dataset a utilizar contenía una gran cantidad de información de las cuales se obtenía datos desde el año 1990, hasta el 2020, llegando a tener más de 17 mil filas, para lo cual habían años que estos no contenían información relevante para el análisis, de igual forma lo consta de 122 columnas para las cuales hay datos pocos relevantes o que no siguen el objetivo para el cual vamos a diseñar el modelo predictivo.

Por lo cual se realizó una previa limpieza de los datos de tal forma que se escogió desde el año 1998 hasta el año 2020 y de las columnas se seleccionaron 38 columnas relevantes para analizar el consumo de energía renovables de un país.

Las variables seleccionadas fueron:

country,year,energy_cons_change_pct,energy_cons_change_twh,electricity_generation,hydro_electricity,nuclear_electricity,other_renewable_electricity,renewables_electricity,solar_electricity,wind_electricity,energy_per_gdp,energy_per_capita,hydro_share_elec,low_carbon_share_elec,low_carbon_electricity,low_carbon_elec_per_capita,low_carbon_share_energy,low_carbon_energy_per_capita,other_renewables_elec_per_capita,other_renewables_share_elec,other_renewables_share_energy,other_renewables_cons_change_twh,other_renewables_energy_per_capita,per_capita_electricity,population,primary_energy_consumption,renewables_elec_per_capita,renewables_share_elec,renewables_cons_change_pct,renewables_share_energy,renewables_cons_change_twh,renewables_consumption,renewables_energy_per_capita,solar_share_elec,solar_cons_change_pct,solar_share_energy,gdp,

Para las cuales se seleccionaron 9 de estas variables y se volvieron categóricas

Para las cuales se seleccionaron 3 valores. Bajo, medio y alto.

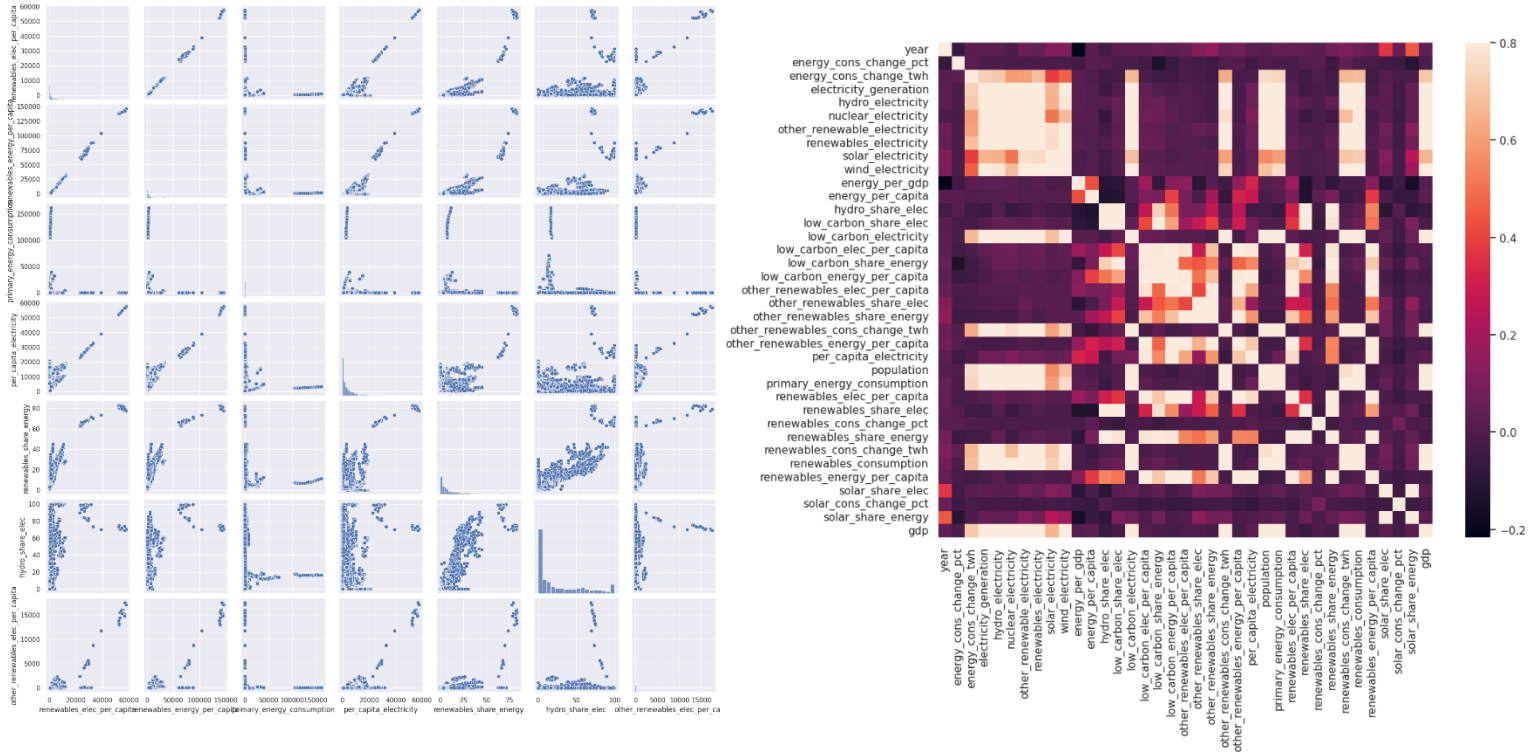
En total el dataset se redujo a 5328 filas y 38 columnas.

2. Exploración del dataset

Se realizó una revisión del dataset en cuantos, a los campos y sus valores, para determinar la correlación existente en cada uno de las columnas.

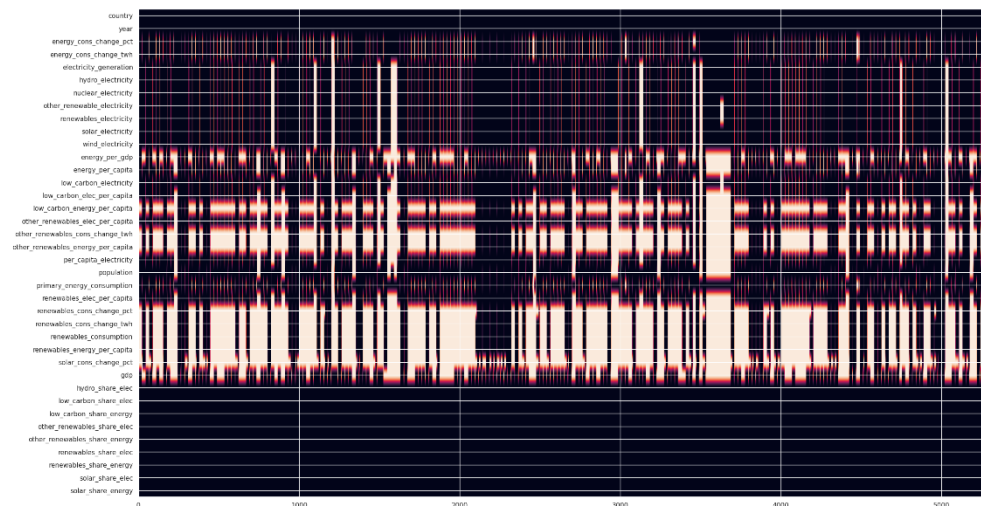
Para lo cual inicialmente se realizó la evaluación de relación entre las variables consideradas más significativas.

Se obtuvo las siguientes matrices de relaciones, en la que se observa sus correlaciones entre cada una de las columnas.



Se realiza una visión general del dataset en cuestión de datos faltantes.

Se obtiene el siguiente cuadro, en la que se observa una gran cantidad de columnas con datos faltantes.

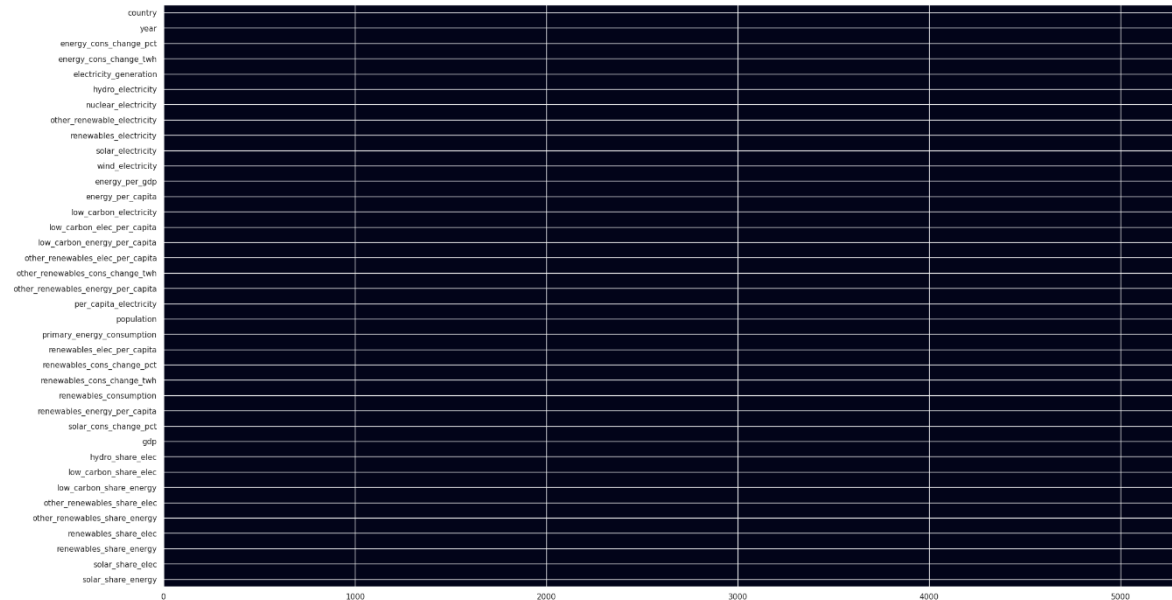


Se realiza una limpieza del data set a través de validaciones de columnas que contengan mas del 40% de sus datos vacíos, serán eliminados.

Para lo cual se eliminaron las siguientes columnas:

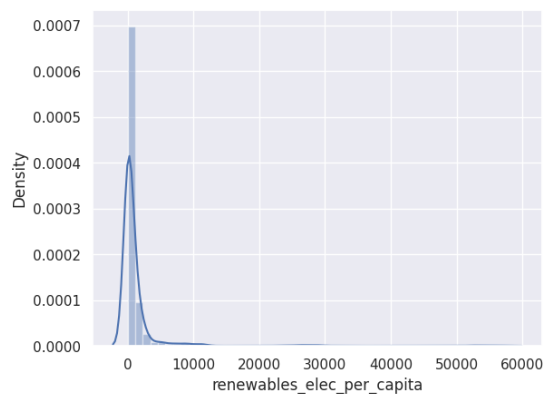
'low_carbon_energy_per_capita' 'other_renewables_cons_change_twh'
'other_renewables_energy_per_capita' 'renewables_cons_change_pct'
'renewables_cons_change_twh' 'renewables_consumption'
'renewables_energy_per_capita' 'solar_cons_change_pct'.

Posteriormente se realiza un llenado de los datos faltantes con un cero, y se obtiene la siguiente visión general de los datos.



3. Evaluación de la variable objetivo "renewables_elec_per_capita"

Consumo per cápita de energía primaria a partir de fuentes renovables (kilovatios-hora). Esta es la variable a examinar para lo cual se revisa su distribución.



Se realiza una transformación logarítmica de la variable objetivo ya que presenta una asimetría muy pronunciada hacia la izquierda

