Python
+ Power BI

Neironi Germán

Exploratory Data Analysis

I started with a dataset downloaded from this website: Ministerio de Economía

(website) (dataset)

The dataset has: **7827 rows and 14 columns**.
I used **Python, Pandas, Matplolib, Seaborn, CSV, IO, Numpy and Scipy libraries** to analyze this dataset.
I see the hight level information **info()**, and the first fice rows **head()**.
I manipulate the date values to convert into datetime data type.
I check the **nulls values**.
I made som basic **plots**.

With **feature engineering** I made new collumns, now the dataset has 22 columns.

Them I try to answer the project questions.

You can see the notebook here:

(repository)

Project Questions

With this data, I try to answer this 3 questions:

A) **What is the trend regarding the amount of energy generated per year in this dataset?**

B) **On average, which is the month of the year in which the most energy is produced?**

C) **Which provinces have improved their energy production in the last year, compared to the first?**

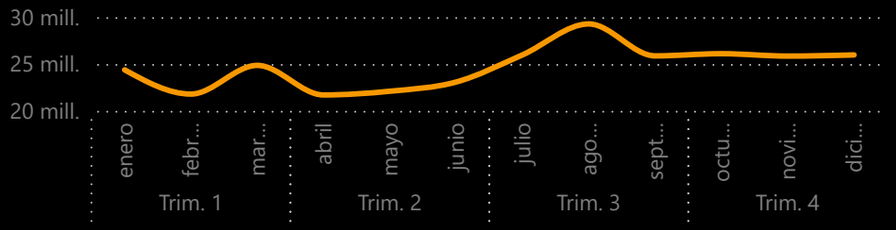
Power Bi

I loaded the dataset in Power BI.
I created new metrics and columns with **DAX.**
I created some visualization that correctly illustrates the different categories and values needed to solve the project question.
I made a **map plot to geolocate** the energy generation.
I made a **line plot** to show the average generation per month.
I made a **dot histogram with trend line** to show the total energy generation by date.
I made a **bar plot** to show the energy generation by each source.
I made a bar plot to show the energy generation difference between the first year and the last year by each province.
I made **4 cards (KPIs)** with summary data.
I made 4 filters **(data segmentation)** to apply over the visualization.

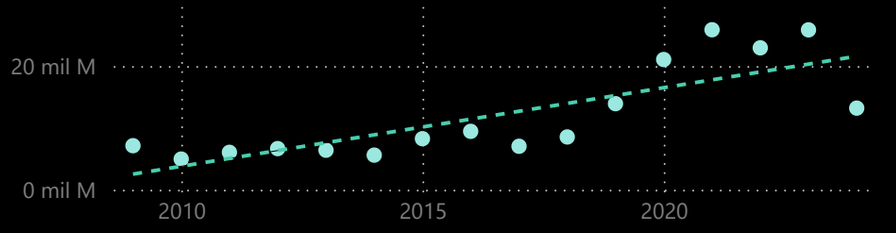
Energy Generation in Argentina

Ubication: Todas | Month: Todas | Energy Source: Todas | Has Improved: Todas

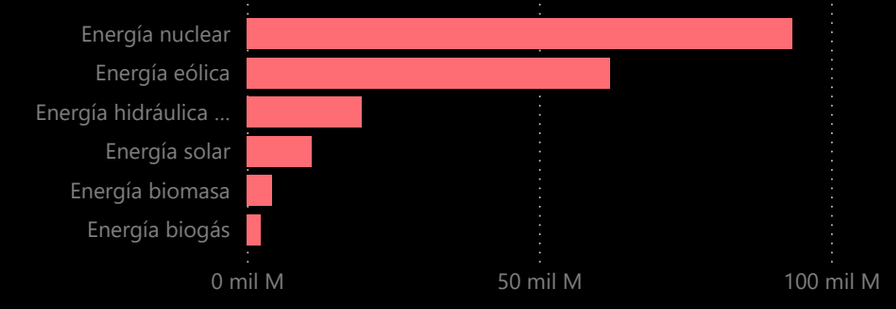
Average generation by month



Total Generation by date



Generation by source



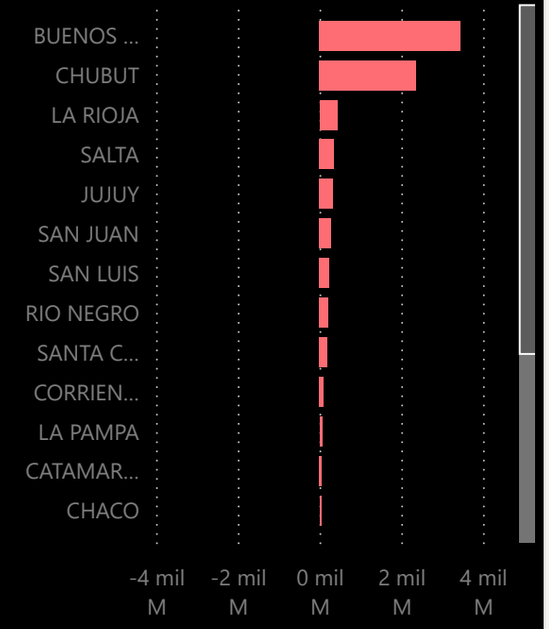
20
Provinces Analyzed

24,62 mill.
Average Genration by Month (MWh)

193 mil M
Total Generation (MWh)

6
Sources

Production Difference by province



Medida rápida

Summary: We can find the answers with Python and Pandas, I we can show them with matplotlib. Finally, we can comunicate this informatión correctly with Power BI.

Question A: What is the trend regarding the amount of energy generated per year in this dataset?

Answer: The trend line shows an average yearly increase of 1376034.16 units of energy.
R-squared value: 0.65.

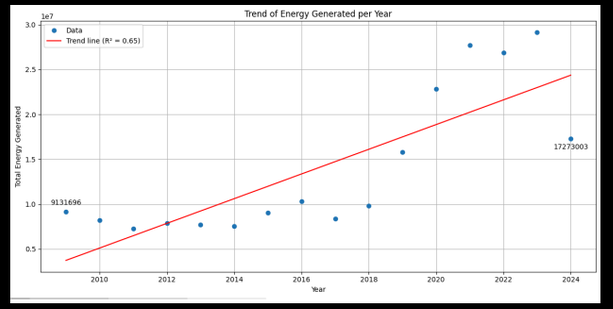
Question C: Which provinces have improved their energy production in the last year, compared to the first?

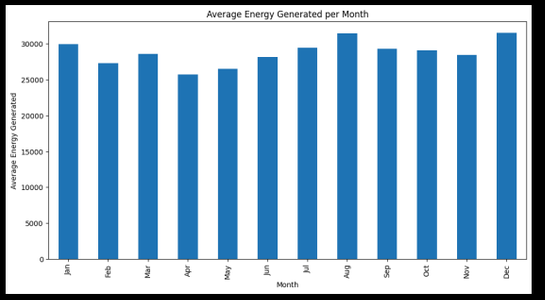
Answer: August, December and January.

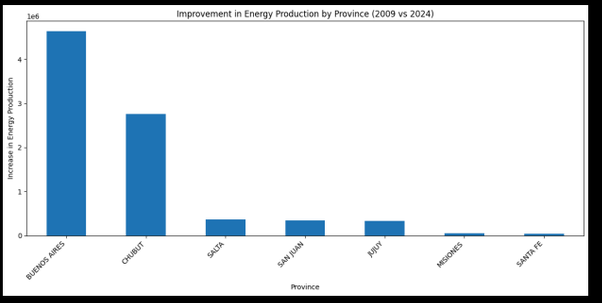
Question C: Which provinces have improved their energy production in the last year, compared to the first?

Answer: Buenos Aires, Chubut and La Rioja has improved the production.

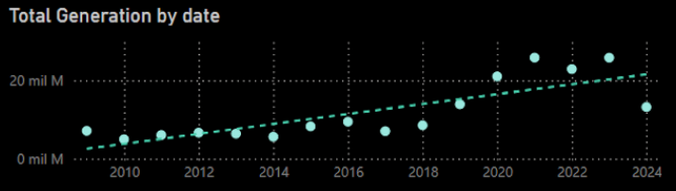
Python

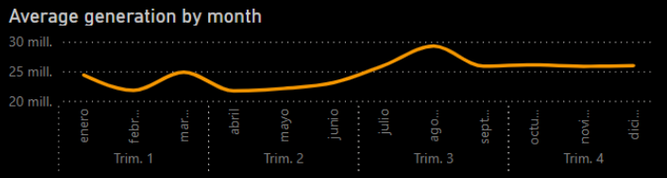


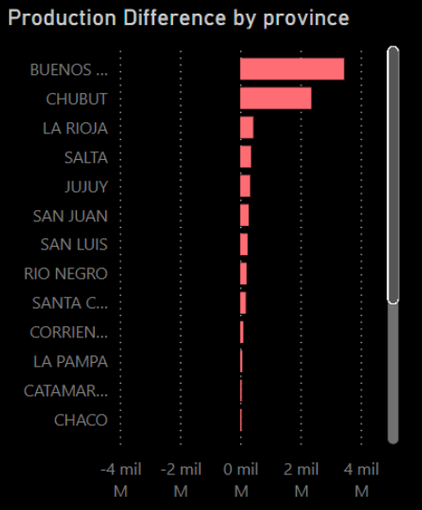




Power BI







Medida rápida

Germán Neironi

Repository Linkedin Website

Medida rápida