Maven Analytics Challenge 12/2023: <https://mavenanalytics.io/challenges/maven-power-outage-challenge/28>

For the **Maven Power Outage Challenge**, you'll be playing the role of a Senior Analytics Consultant hired by the U.S. Department of Energy (DOE). Here's your project brief:

Electricity outages are a growing concern as we enter an age of unprecedented energy demand and climate disasters.

We have event-level power outage data going back to 2002, but have struggled to make sense of it due to severe issues with the data quality and integrity.

This is where you come in.

We need you to **consolidate and clean up the raw data**, and **create a dashboard or report** to help us understand patterns and trends around outages, quantify their impact on our communities, and identify possible weak points in the grid.

Last but not least, please **explicitly call out any caveats or assumptions** you make in regards to data quality issues or missing values.

--

Think about – what kind of analysis am I performing?

* Time
* Durations
* Locations/region
* Event types
* Event severity (MW loss or people disrupted)

What level of detail will I be looking at (hierarchy)?

* Year/month/day/hour/minute
* Country/state/county/city
* Grid sector
* Individual types/grouped
* All grades of severity or grouped

--

e.g. if you’re only interested in the year, you can just extract the year

- can look at just the state level if you’re only interested in that

--

My goal:

Only look at years (not months/days)

Only look at states

Group worksheets by similar layouts

Look at headings for each of the years and compare

--

I edited using Power Query Editor

Make sure there is a year value

N/a – change to null

Time:

Ongoing is probably a mistake – remove & put as null

When we want to remove stuff, we convert the column values to text

Replace values: replace N/A > null, ongoing > null, evening > 6 p.m., midnight > 12 am, noon > 12 pm

Remove . from times