

Hello everyone! So I've been thinking a lot about the features given in the [dataset](#), and as to what each one implies. Some are more obvious, but others are a little trickier to understand (to me anyways!).

I've also come up with two additional features that can be added to the dataset (i.e features derived from a combination of other features).

1. **record_id** : unique id given to identify each record
2. **utility_id_ferc1** : id given by the Federal Energy Regulatory Commission to each utility. (Numbers seem arbitrary.)
3. **report_year** : The year the record was given.
4. **plant_name_ferc1** : name of the energy plant/station. When analysing data involving this feature, be careful of possible mistakes in spelling. For example, `respondent's portion` and `resondent's portion` are the same energy plants, but the latter has a spelling error (the 'p' is missing). Be careful not to treat this as two different plants. There are also some other edge cases you might want to consider. This should be part of the Data cleaning phase.
5. **fuel_type_code_pudl** : the type of fuel (i.e coal, gas, nuclear, oil).
6. **fuel_unit** : the unit of fuel. This would typically be related to the `fuel_type_code_pudl` feature, as the coal fuel type would be measured in ton; the gas fuel type in mcf, etc. It's worth doing more research on what each one means.
7. **fuel_qty_burned** : the quantity of fuel burned.
8. **fuel_mmbtu_per_unit**: the measure of energy per unit. The metric being used here is the one million British Thermal Units (MMBTU), which is the measure of energy content across most industries. This feature measures the MMBTU per unit.
9. **fuel_cost_per_unit_burned** : the fuel cost per unit burned. This is the cost of burning one unit of `fuel_qty_burned`.
10. **fuel_cost_per_unit_delivered** : the cost of fuel delivered per unit. This is the cost of delivering one unit of fuel. (Note: there is no feature indicating the amount of fuel quantity delivered in this dataset.)
11. **fuel_cost_per_mmbtu** : the cost of fuel per mmbtu. This is the cost of one `fuel_mmbtu_per_unit`.

To be sure, there are enough features here to perform data analysis on a significant level, however, a few more can be extracted from the given features. *Here are my derived features:*

12. **fuel_qty_burned_in_mmbtu** : Total fuel burned in MMBTU.

$$\circ \text{ `fuel_qty_burned` } \times \text{ `fuel_mmbtu_per_unit` }$$

13. **total_fuel_cost_burned** : Total cost of fuel burned.

$$\circ \text{ `fuel_qty_burned` } \times \text{ `fuel_cost_per_unit_burned` }$$

Other interesting finds:

$$14. \text{ fuel_mmbtu_per_unit } \times \text{ fuel_cost_per_mmbtu } \approx \text{ fuel_cost_per_unit_burned }$$