

Sample Pre-doc Data Task

April 30, 2023

This sample data task will use resource-level scheduled output data from the Electricity Reliability Council of Texas (ERCOT). For the first few questions, use the file `ercot_resource_output.csv`.

1. How many unique values does the variable Resource Name take in the data? the variable QSE?
2. What is a QSE? Do a quick online search for this ERCOT acronym. Provide a brief (1-3 sentences) definition for QSE as used in ERCOT's market for electricity.
3. Find the set of unique QSE/Resource Name pairs. Answer the following questions.
 - (a) Is it ever the case that a single QSE is paired to multiple resource names? What might this indicate about the relationship between QSEs and Resource Names? What are the 10 largest QSEs in terms of the number of unique Resource Names they are paired to in the data?
 - (b) Is it ever the case that a single Resource Name is paired to more than one QSE in the data? For how many Resource Names is this true for? Why might a single Resource Name pair with multiple QSEs in the data? Hint: Look at how pairs change over time
4. Now turn to `resource_type.csv`
 - (a) How many unique, non-missing values does Resource Type take? Can you find definitions for them? (No need to define all of them, just attempt a few)
 - (b) Are there any empty strings in the resource type column? Which resource names are missing their type? Can you guess what the missing values should be? Fill in the missing values with your guesses (you will carry your filled in guesses for the remainder of the data task).
5. Based on the following definitions, use the resource type column to make a "Fuel Type" column. After doing so, merge Fuel Type and Resource Type onto `ercot_resource_output.csv` using Resource Name (you should end up with 6 unique values of Fuel Type).
 - DSL - Other
 - SCGT90 - Natural Gas
 - WIND - Wind
 - PWRSTR - Other

- HYDRO - Other
- CCGT90 - Natural Gas
- PVGR - Solar
- SCLE90 - Natural Gas
- GSREH - Natural Gas
- CCLE90 - Natural Gas
- CLLIG - Coal
- GSSUP - Natural Gas
- NUC - Nuclear
- GSNONR - Natural Gas
- RENEW - Other

6. Plot the following:

- (a) output summed by day
- (b) output summed by hour-of-day (hours 0-23)
- (c) output summed by hour-of-day and by Fuel Type (the variable you defined in 5.)

Discuss the patterns you see in each plot.

- 7. Looking at the plot from 6.(a), does this data look stationary? Using the data summed at the daily level, test for a unit root and interpret the result. Now calculate its first difference and plot it. Does it look stationary?
- 8. Now sum output at the hourly level (day-hour, not hour-of-day). Fit an AR(3) model on electricity output. Do you believe an AR model is a good fit? Why or why not?
- 9. Run the following dummy variable regressions and interpret the coefficients:
 - (a) output regressed on a set of indicator variables for each Fuel Type in the data
 - (b) output regressed on a set of indicator variables for each day of the week (Sun, Mon, Tues, etc.)
 - (c) output regressed on a set of indicator variables for each week in the data

What factors might explain the values of the coefficients you found?