```
capture log close
log using "ERCOT_Analysis.log", text replace

*--------------------------------------------------
* ERCOT Electricity Output Analysis
* Author: Anjana Azhuvath
* Date: March 13, 2025
* Objective: Analyze ERCOT resource output data
*--------------------------------------------------


cd "/Users/anjanaraja/Desktop/STATA_for_RA/Utility_Project" // Setting Directory
import delimited "ercot_resource_output.csv" // Importing .csv file

/*******************************************************************************
Question 1:

How many unique values does the variable Resource Name take in the data? the
variable
QSE?
Ans:a.  Number of unique values of resource_name is  1121
                Number of records is  3008438

b.Number of unique values of qse is  194
Number of records is  3008438
*******************************************************************************/

unique resource_name
unique qse

/*************************************************************************************************

Question2:

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.

Ans: ERCOT stands for the Electric Reliability Council of Texas.
QSE or "Qualified Scheduling Entities" submit bids and offers on behalf of
resource entities (REs) or load serving entities (LSEs) such as retail electric providers
(REPs).
*************************************************************************************************/

/*************************************************************************************************
Question 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?

Ans: Number of unique values of resource_name qse is 1127
Number of records is 3008438

(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

Ans: There are 64 resources paired with only one QSE. A vast majority of resources are paired with only one qse and 6 are paired with 2 qse.

**********/

```
*PART A
unique resource_name qse

*PART B

unique resource_name, by(qse) gen(num_resource)
tab num_resource // resource with more than one qse

unique qse, by(resource_name) gen(num_qse)
tab num_qse // Count resource name

sort num_qse sced_time_stamp // change over time

//save "ercot_resource_output.dta"
```

/**********

Question 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)

There are 4 missing values and 15 unique resource_type.

missing resource_name: There are two solar and wind companies
GALLOWAY_SOLAR1

ROSELAND_SOLAR3
SSPURTWO_WIND_1
SWEETWN2_WND24

(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).

**********/

```
clear

import delimited "ercot_resource_types.csv", clear

rename v1 resource_name // renaming variables

rename v2 resource_type

drop in 1/1  // dropped first row which was a variable name

*PART A

codebook resource_type // count of missing and unique
browse if missing(resource_type) // Finding resource names with missing resource types


*PART B

tab resource_type // PVGR is the Solar variable and WIND is the resource_type for Wind Company

replace resource_type = "PVGR" if resource_name == "GALLOWAY_SOLAR1"
replace resource_type ="PVGR" if resource_name == "ROSELAND_SOLAR3"
replace resource_type ="WIND" if resource_name == "SSPURTWO_WIND_1"
replace resource_type ="WIND" if resource_name == "SWEETWN2_WND24"

//save "ercot_resource_type.dta"
```

/**********
QUESTION 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
**********************************************************************************************/
use "ercot_resource_output.dta", clear

merge m:1 resource_name using "ercot_resource_type.dta" // merging data sets

drop _merge

gen fuel_type = "Other" // Other type
replace fuel_type = "Natural Gas" if
inlist(resource_type,"SCGT90","CCGT90","SCLE90","GSREH","CCLE90", "GSSUP",
"GSNONR") // Natural Gas type
replace fuel_type = "Nuclear" if inlist(resource_type,"NUC") // Nuclear Type
replace fuel_type = "Wind" if inlist(resource_type,"WIND") // Wind Type
replace fuel_type = "Coal" if inlist(resource_type,"CLLIG") // Coal Type
replace fuel_type = "Solar" if inlist(resource_type,"PVGR") // Solar Type


/**********************************************************************************************
QUESTION 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.)
**********************************************************************************************/

//sum telemetered_net_output

*PART A
```

```
gen sced_date = substr(sced_time_stamp, 1, 10)
gen date= date(sced_date, "MDY")
format date %td   // subsetting date


preserve
collapse (mean) telemetered_net_output, by(date)
twoway (line telemetered_net_output date, sort), ///
   title("Telemetered Net Output Over Time") ///
   xtitle("Date") ///
   ytitle("Telemetered Net Output")
restore

*PART B

generate sced_hour = real(substr(sced_time_stamp, 12, 2))

preserve
collapse (mean) telemetered_net_output, by(sced_hour)
twoway (line telemetered_net_output sced_hour, sort), ///
   title("Telemetered Net Output by Hour of Day") ///
   xtitle("Hour of Day") ///
   ytitle("Telemetered Net Output")
restore

*PART C

preserve
collapse (mean) telemetered_net_output, by(sced_hour fuel_type)
twoway ///
   (line telemetered_net_output sced_hour if fuel_type == "Coal", sort lcolor(red)) ///
   (line telemetered_net_output sced_hour if fuel_type == "Natural Gas", sort
lcolor(blue)) ///
   (line telemetered_net_output sced_hour if fuel_type == "Nuclear", sort
lcolor(green)) ///
   (line telemetered_net_output sced_hour if fuel_type == "Other", sort lcolor(gray)) ///
   (line telemetered_net_output sced_hour if fuel_type == "Solar", sort lcolor(orange)) ///
   (line telemetered_net_output sced_hour if fuel_type == "Wind", sort lcolor(cyan)), ///
   title("Summed TeleMetered Net Output by Hour and Fuel Type") ///
   xlabel(0(1)23) ///
   xtitle("Hour of Day") ///
   ytitle("Summed Net Output (MW)") ///
   legend(order(1 "Coal" 2 "Natural Gas" 3 "Nuclear" 4 "Other" 5 "Solar" 6 "Wind"))
```

```
restore
/*****************************************************************************************
Question 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?

Ans: Fail to reject null hypothesis of unit root, therefore first difference is needed.
*****************************************************************************************/

preserve
collapse (sum) telemetered_net_output, by(date)
tsset date, daily

dfuller telemetered_net_output, lags(5)

generate firstdiff_output = D.telemetered_net_output

twoway (line firstdiff_output date, sort), ///
   title("First Difference of TeleMetered Net Output") ///
   xtitle("Date") ytitle("Δ Summed Output") ///
   ylabel(, format(%10.0gc))
restore

/*****************************************************************************************

Question 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?
*****************************************************************************************/

gen sced_date_hour = substr(sced_time_stamp, 1, 13) // day+hour
gen sced_dh= sced_date_hour + ":00" // rounded to nearest hour
generate double date_time=clock(sced_dh,"MDY hm" ) //Generate date-hour var
format date_time %tc

//save "Ercot_Merge.dta", replace

preserve
collapse (sum) telemetered_net_output, by(date_time)
tsset date_time
tsline telemetered_net_output // The data appears to be mostly stationary. There isn't
any trend
//pac telemetered_net_output
regress telemetered_net_output L(1/3).telemetered_net_output, robust
estimates store model1
```

```
esttab model1 using regression_output.tex, replace tex booktabs label //saving output to
latex
restore

/**********************************************************************************************
Question 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?
**********************************************************************************************/
clear
use "Ercot_Merge.dta", replace

*PART A
encode fuel_type, generate(fuel_type_dummy)
regress telemetered_net_output i.fuel_type_dummy, robust
estimates store model2

*PART B
gen sced_date_num = date(sced_date, "MDY")
format sced_date_num %td
gen day= dow(sced_date_num) // Estimating day of week

gen week_num = week(sced_date_num) //Estimating week of year
regress telemetered_net_output i.fuel_type_dummy, robust
estimates store model2


ssc install estout
foreach var in fuel_type_dummy day week_num {
   eststo clear
   eststo: regress telemetered_net_output i.`var', robust
}


log close
```