

1. Data Gathering

- a. Go to: [https://www.misoenergy.org/markets-and-operations/real-time--market-data/market-reports/#nt=%2FMarketReportType%3ASummary%2FMarketReportName%3AGeneration%20Fuel%20Mix%20\(xlsx\)](https://www.misoenergy.org/markets-and-operations/real-time--market-data/market-reports/#nt=%2FMarketReportType%3ASummary%2FMarketReportName%3AGeneration%20Fuel%20Mix%20(xlsx))
- b. Based on the information available on the website, write a brief (around 5 sentences) description of the data ("Generation Fuel Mix (xlsx)" dataset) you are about to download.
- c. Create a script to download "Generation Fuel Mix (xlsx)" daily data for period from June 1, 2022 to July 31, 2022
- d. For further analysis use data from "Real-Time State Estimator Generation Fuel Mix Report". Disregard "Day-Ahead Cleared Generation Fuel Mix Report" data.
- e. Aggregate data in a single pandas dataframe. Create [DateTime] column of datatype [datetime64]. Do not take timezones into consideration.
- f. Export dataframe in one .xlsx file

Please note that we would prefer using Python (pandas package) for this step. The objective is to minimize manual work and to create a robust process that can be executed on a regular basis. Completing this step manually will result in a score of zero for this step. This step is meant to test technical skills. However, being practical is key here, so keep it simple.

Deliverables (zipped in single file):

- Data description document (.txt or .docx file)
- Script used to download data (.py or .ipynb file)
- Raw data (.csv or .xlsx format)

2. Data Processing

- a. Load the data downloaded in the previous step only for North, South and Central regions (disregard summarized values across regions)
- b. Transform data into "tall" ("deep") format, so dataframe consists of the following columns:
 - I. [DateTime] – contains timestamp
 - II. Region – containing regions
 - III. [Power source] – containing all available power sources (fuels)
 - IV. [Value, MW] – containing supply (generation) values.
- c. Convert generation values to TW, rename column to [Value, TW]
- d. Keep data only July, 2022. Delete all other rows
- e. Calculate generation mix for July in North region. Use a separate dataframe to output the result of calculation.
- f. Export dataframe created during step e into .csv or .xlsx file.
- g. Assuming that Renewable energy sources include: Hydro, Solar, Wind and Storage. What was the share of Renewable generation in July in North region? Print answer to console.
- h. Optional: visualize dataset created during step e using any of python visualization tools in a way you find most

Please note that we would prefer using Python (pandas package) for this step. The objective is to minimize manual work and to create a robust process that can be executed on a regular basis. This step is meant to test technical skills. However, being practical is key here, so keep it simple.

Deliverable (zipped in single file):

- Script used to process data (.py or .ipynb file)
- Output data (step f) (.csv or .xlsx format)