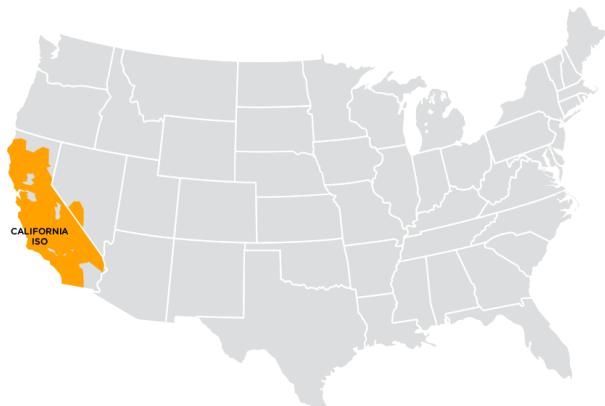# Project Description

The California Independent System Operator (CAISO) is a non-profit Independent System Operator (ISO) serving California. It oversees the operation of California's bulk electric power system, transmission lines, and electricity market generated and transmitted by its member utilities. CAISO is one of the largest ISOs in the world, delivering 300 million megawatt-hours of electricity each year and managing about 80% of California's electric flow.

Each day, the ISO publishes crucial information related to the market. For your task, you will be responsible for predicting the power generated by wind plants in SP15 Hub of the ISO.

- Under the project email, you will find two files with each the name *Wind_Actual.csv* and *Wind_Forecast.csv*:
  - *Wind_Actual* contains realized power generation data by wind plants.
  - *Wind_Forecast* contains forecasted power generation data by wind plants made 1 day prior to the actual power generation. (For example, the value on DATE 2024-03-02 represents wind generation forecasted on 2024-03-01).
  - Both files contain data from 2023-01-01 to 2024-03-15 inclusive
  - Both files contain the columns:
    - DATE: Target date of wind generation, forecasted or actual
    - HE: Hour ending. HE x represents hour from x-1 to x
    - Trading_Hub: Hub being traded, always *SP15* in the data
    - Renewable_Type: Type of renewable energy, always *Wind* in the data
    - MW: Total MW of power/forecasted power generated

- **Goal 1 (Optional): Scrap actual and forecasted wind generation from 2024-03-16 onwards**
  - Please note that the file we provide only contains data up to 2024-03-15, and your final goal is to predict the power generation by wind plants on 2024-04-09. We encourage you to develop a data scraping code to download the most recent data.
  - To do so, go to http://oasis.caiso.com/mrioasis/logon.do, click *SYSTEM DEMAND -> Wind and Solar Forecast*

- In the *Wind and Solar Forecast* tab, input dates from 2024-03-16 onwards and click *apply* to view the most recent data. Please be aware that the data is updated hourly, completing its updates around 4AM each day, when CAISO will finish publishing actual wind generation of the previous day and forecasted wind generation of the next 7 days. The future 7-day forecast will be replaced with the newest forecast when it becomes available. (For example, on 2024-03-02, CAISO will publish the actual wind generation of 2024-03-01 and the forecasted wind generation from 2024-03-03 to 2023-03-09; then, on 2024-03-03, CAISO will publish the actual wind generation of 2024-03-02 and the forecasted wind generation from 2024-03-04 to 2023-03-10, overwriting the forecast from 2024-03-04 to 2024-03-09 made on 2024-03-02).

- Click *Download CSV* to get the data. Below is a preview of the csv file, note that you only need to reference these columns:
  - OPR_DT: Equivalent to the DATE column in the *Wind_Actual* and *Wind_Forecast* file
  - OPR_HR: Equivalent to the HE column in the *Wind_Actual* and *Wind_Forecast* file
  - TRADING_HUB: Keep only TRADING_HUB == *SP15*
  - RENEWABLE_TYPE: Keep only RENEWABLE_TYPE == *Wind*
  - MW
  - MARKET_RUN_ID: There are 5 categories in total. In your task, you only need to reference 2 categories: *DAM* and *ACTUAL* where *DAM* is the forecasted wind generation value and *ACTUAL* is the realized wind generation value

- We recommend you develop a web-scraping tool to automatically load the page, navigate to the *Wind and Solar Forecast* tab, input the dates, and download the CSV files, as data scraping is an essential skill at 5DE. However, we recognize that there might be time constraints for some. Therefore, if you feel unable to complete the auto-scraping process, you can either rely solely on the data we have provided, which should be sufficient for this project, or simply manually download the most recent data.
  - There is more than one way to automatically download the data, some of which are easier than others. We recommend thoroughly exploring the page and the downloading process before beginning to code.
- Given that the project is due on 2024-04-08 12:00PM EST, the most recent actual wind generation data available will be from 2024-04-07. Therefore, if you opt to download data from 2024-03-16

onwards, whether automatically or manually, you have the flexibility to choose the date range up to whatever date you believe is necessary for completing the project. In other words, there is no strict requirement to use the latest data published immediately before the project's due date.

- o Please ensure that your code clearly indicates if you utilize any recent data, specifying the exact dates of the data you have downloaded. Additionally, if you have developed a web-scraping tool, make sure to distinctly indicate this portion of your code.

- **Goal 2: Predict actual power generation by wind plants of SP15 hub on 2024-04-09**
  - o Train a model to predict the 24-hour wind generation for the SP15 hub on 2024-04-09.
  - o Feel free to use either actual data, forecast data, or both for modelling.
  - o Choose any model you believe is appropriate for this task.
  - o While it is not mandatory, you are encouraged to incorporate any steps you find beneficial, such as visualization, data preprocessing, or model tuning processes. Feel free to explore various models. However, please clearly select and specify the model you intend to use for the final prediction, along with the reasons for your choice. Ensure to include necessary and brief comments in your code to explain your decision-making process and methodology.

- Submission Guideline:
  - o **This project should be completed independently without assistance from others.**
  - o This project should be completed using Python, and we strongly recommend utilizing ipynb format so that we can track your intermediate steps. A good place to format ipynb is Google CoLab where you will also have access to their shared GPU if you are using a NN model.
  - o **This project is due by 2024-04-08 11:59PM EST.** To ensure fairness, late submissions will not be accepted.
  - o While prediction accuracy will be considered during our recruitment process, we value more on the cohesion and quality of your code. Therefore, focus more on developing clear, well-structured code rather than stressing excessively over the modeling aspect.
  - o To submit your project, please send us an email with the subject line: *2024-04-09 Project [First Name] [Last Name]*. Attach a zip-file that includes the following:
    - ▪ Your Python code file
    - ▪ The final input CSV file that generates your prediction
    - ▪ The final model that generates your prediction
    - ▪ Your prediction result file
      - • The prediction CSV should have the name *2024-04-09_Pred_[First Name]_[Last Name].csv*, and should follow the same format of the *Wind_Actual* and *Wind_Forecast* file (Same columns with 24 rows representing each HE of 2024-04-09)
  - o If your file size exceeds the maximum attachment size for an email, you can also provide us with a google drive link to your zip-file, make sure to grant access to the zip-file.